

ESTUDIO UNIFICADO DEL ANALISIS
MULTIVARIANTE MEDIANTE LA UTILIZACION
DE PROYECTORES Y SU APLICACION AL
ESTUDIO - SOCIO - ECONOMICO
DE LA RIOJA

Fernando Gomez Bezares



TESIS

DOCTORAL

TOMO

I



A todos los que han colaborado en
la realización de este trabajo.

Universidad de Deusto (Bilbao)

Facultad de Ciencias Económicas y Empresariales.

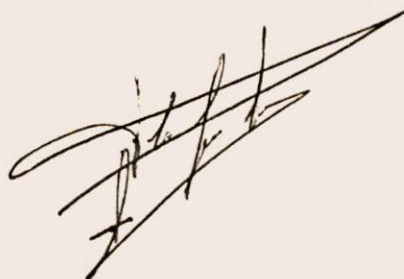
"ESTUDIO UNIFICADO DEL ANALISIS MULTIVARIANTE MEDIANTE LA
UTILIZACION DE PROYECTORES Y SU APLICACION AL ESTUDIO SO
CIO - ECONOMICO DE LA RIOJA"

Tesis Doctoral presentada por D.
Fernando Gómez Bezares, dirigida
por el Profesor de la Universidad
de Deusto Dr. Enrique Chacón Xé
rica.

El Director



El Doctorando



Bilbao, Abril, 1.980.

PRESENTACION.

La tesis que en las siguientes páginas se expone intenta, tal como se puede deducir de su título, un primer estudio teórico del análisis multivariante y una aplicación práctica posterior a la Rioja.

El análisis teórico se encuentra recogido en la primera parte de este trabajo, mientras que la segunda y tercera representan el estudio sociológico y el económico respectivamente, estas dos últimas no pretenden ser completas, sino solo en cuanto sean aplicación directa del análisis multivariante estudiado, y aún en estos casos limitándose a ciertos temas concretos, dada la lógica escasez de medios del autor.

A continuación de las tres partes citadas, añadimos 2 anexos de distinta índole y magnitud. El primero pretende familiarizar de forma breve y concisa al lector, con la historia y geografía de la provincia objeto de nuestro estudio, al objeto de que pueda comprender mejor ciertas consideraciones formuladas en el mismo. El segundo, ciertamente voluminoso, contiene las salidas de ordenador, a partir de los cuales se realiza el trabajo práctico; su consulta ofrecerá la posibilidad de comprobar los resultados que comentaremos e incluso la ampliación del mismo.

Quiero aprovechar estas líneas para manifestar mi gratitud hacia el Doctor Enrique Chacón Xérica S. J., profesor

mío y director de esta tesis, cuya inestimable colaboración ha hecho posible el presente trabajo.

También quiero recordar a los profesores J.L. Narvaiza y J.M. Eguiagaray, que me han orientado acertadamente en los problemas que les he ido planteando, el primero como especialista en análisis cuantitativo y el segundo en estructura y política económica.

Por lo que se refiere al tratamiento de la información en el ordenador, que indudablemente ha resultado largo y complejo, quiero agradecer a la Facultad de Ciencias Económicas y Empresariales de la Universidad de Deusto, al haberme brindado su centro de cálculo para la realización del mismo, así como a los Sres. Ricón y Cubas al haber atendido en el citado centro el proceso de mis datos y también a los perforistas que implantaron la información en fichas.

Pero quizá el mayor esfuerzo realizado haya correspondido a la búsqueda de la información. En el aspecto de los datos económicos, su consecución fue posible gracias al interés que se tomó el Sr, Orbañanos, economista de la Excelentísima Diputación provincial de Logroño y jefe del gabinete de planificación de la misma.

Más difícil resultó la realización de la encuesta, donde a causa de lo ambicioso del proyecto, mi personal esfuerzo resultaba claramente insuficiente. Afortunadamente conté

con la colaboración de un grupo de amigos sin cuya ayuda di
fícilmente se hubiera podido llevar a cabo la citada encuess
ta. Especial mención de entre ellos quiero hacer a J.M. Arriee
ta, que se tomó un especial interés por este trabajo.

Evidentemente sería muy largo el citar aquí a todos
los que de una manera u otra me han ayudado en la ejecución
de esta tesis, pero aprovecho estas líneas para desde aquí
testimoniarles mi agradecimiento.

INDICE GENERAL

	<u>Página</u>
PARTE PRIMERA:	
0.- INTRODUCCION.	10
I.- ESPACIOS VECTORIALES EUCLIDIANOS. CONCEPTO ELEMENTAL DE PROYECTOR. ESQUEMA DE UNIFICA CION.	17
II.- CORRELACION MULTIPLE. DEFINICION DE PROYEC TOR.	23
III.- CORRELACION CANONICA.	31
IV.- COMPONENTES PRINCIPALES	56
- Aplicaciones similares al análisis de componentes principales.	74
- Problema de la medición de los compo- nentes.	89
V.- MODELO GENERAL Y APLICACIONES.	91
- Aplicaciones del modelo.	94
- Análisis discriminante.	95
- Cluster análisis.	119
- Variables cualitativas.	122
- Tablas de contingencia.	128
- Análisis de la varianza.	129
VI.- CONCLUSIONES.	131
VII.- EL PROBLEMA DE LA INFERENCIA.	134
PARTE SEGUNDA:	
0.- INTRODUCCION.	143

I.- RECOGIDA DE LA INFORMACION. BASES DEL TR <u>A</u>	
TAMIENTO.	146
- Encuesta de opinión en La Rioja.	153
- Bases para el tratamiento.	172
II.- TRATAMIENTO DE LA INFORMACION.	173
- Análisis factorial 1.	176
- Análisis factorial 2.	189
- Análisis factorial 3.	200
. Histogramas factores Logroño.	215
. Análisis de nuestras variables.	226
. Cluster análisis.	227
. Interpretación del cluster análisis.	231
. Conclusiones del cluster análisis y an <u>á</u>	
lisis discriminante.	255
. Tablas de contingencia.	259
. Conclusiones del estudio de tablas de	
contingencia.	277
- Estudio de los cruces de variables.	279
- Análisis factorial 4.	301
- Análisis factorial 5.	323
. Interpretación de las puntuaciones.	342
. Creación del I.I., I.P., I.A.	344
. Estudio de nuestra muestra a través de	
los factores.	351
. Conclusiones de este último análisis.	364
. Otras elaboraciones multivariantes.	370
III.- EPILOGO Y CONCLUSIONES.	374

PARTE TERCERA:

0.- INTRODUCCION.	380
I.- OBJETIVOS DEL ESTUDIO.	382
II.- REGIONALIZACION - PROBLEMATICA ECONOMICA.	385
III.- APLICACION DEL ANALISIS MULTIVARIANTE A LA COMARCALIZACION.	397
IV.- PROBLEMATICA DE LA BUSQUEDA DE INFORMACION.	400
- Variables.	409
- Municipios.	422
- Flujos.	430
V.- COMARCALIZACION BASADA EN EL ANALISIS MULTI VARIANTE.	460
- Comarcalización homogénea.	461
- Aproximación a la nodalidad.	539
- Análisis de la calidad de vida.	561
VI.- PROPUESTA DEFINITIVA DE COMARCALIZACION.	576
VII.- RESUMEN GENERAL. APLICACIONES. CONCLUSIONES.	600

APENDICE HISTORICO-GEOGRAFICO:

0.- INTRODUCCION.	617
I.- ETIMOLOGIA.	618
II.- GEOGRAFIA.	621
III.- HISTORIA.	625

APENDICE ORDENADOR:

BANCO DE DATOS 1.	632
BANCO DE DATOS 2.	643
A.F. RIOJA.	678

	<u>Página</u>
A.F. RIOJA 2.	695
NORMALIZACION RIOJA 3.	711
A.F. RIOJA 3.	712
VARIABLES LOGROÑO.	726
VARIABLES RIOJA.	731
CLUSTER.	736
ANALISIS DISCRIMINANTE.	740
CRUCES LOGROÑO.	745
CRUCES RIOJA.	797
NORMALIZACION RIOJA 4.	849
A.F. RIOJA 4.	850
VARIABLES NO INTRODUCIDAS EN EL A.F. 4	878
NORMALIZACION RIOJA 9 (5).	891
A.F. RIOJA 9 (5).	892
DISTRIBUCION FACTORES.	914
VALORES COMARCAS.	938
NORMALIZACION AGRICO.	939
A.F. AGRICO.	940
NORMALIZACION AGRIC. 2.	959
A.F. AGRIC. 2.	960
CLUSTER.	978
NORMALIZACION SERVIC.	984
A.F. SERVIC.	985
NORMALIZACION INDUST.	1.012
A.F. INDUST.	1.013
NORM. CALIDA.	1.025
A.F. CALIDA.	1.026

Página

CONSIDERACION FINAL.

1.045

BIBLIOGRAFIA.

1.049

PARTE PRIMERA

0.- INTRODUCCION.

El objetivo de esta primera parte es la unificación del análisis multivariante, (1), presentando tres casos generales, de los cuales podemos deducir lo fundamental de este análisis.

No cabe ninguna duda de que la amplitud del análisis multivariante es muy grande y que por lo tanto sería excesivo pretender estudiarlo todo dentro de los límites de este trabajo; pero lo que si pretendo es estudiar de forma unificada y coherente las partes fundamentales.

En segundo lugar quiero hacer constar que en principio me voy a circunscribir al análisis de los datos, es decir al análisis descriptivo, pues es mucho más claro a la hora de exponer la unificación; de todas formas para las aplicaciones que haré en las partes sucesivas de este trabajo, tendré que hacer mención en algunos casos a los problemas de inducción.

Sin embargo, y siguiendo en esto a la mayoría de los autores franceses (2), considero que el análisis multiva-

(1).- Anexo página siguiente.

(2).- Por ejemplo Saporta (1.976) y Benzécri (1.973) son los más interesantes. También Bertier P. y Bourroche - J.M. (1.975), Cailliez F. y Pages (1.976) así como otros muchos. Pero no solo son autores Franceses y así tenemos a Tukey J.W. (1.977) y algunos otros.

.- Anexo citado.

La obra clásica de análisis multivariante es la de Anderson (1.958), que aún habiendo sido superada en algunos puntos sigue siendo básica para los estudiosos del tema.

Kendall (1.957), distingue entre:

- Análisis de componentes.
- Análisis factorial.
- Relación funcional.
- Análisis canónico.

La distinción entre componentes y factorial no está claro y lo comentaremos más adelante.

En conjunto este libro es una contribución a la enseñanza del tema. Se basa en su obra anterior (1.948), mejorada con los tratados de Anderson (que fue dada en apuntes antes del 1.958) y Roy (1.957) con lo que se convirtió en el más completo tratado sobre el tema en su día.

Kendall reeditó en 1.961 y posteriormente llegó la obra de Morrison (1.967).

Un trabajo reciente es el de Guigou (1.977) que es bastante interesante. Pero sería extensísimo pretender exponer una bibliografía completa sobre el tema, limitándonos a citar simplemente a unos pocos autores representativos.

riante, se debe presentar al comienzo, como un análisis de datos, dejando la inducción para el estudio de la aplicación concreta de cada técnica, pues cada una de ellas tiene suficientes particularidades como para precisar un estudio separado.

Por todo lo expuesto, insisto en que el estudio se va a referir, mientras no se indique lo contrario a un estudio de los datos, y por lo tanto a una descripción multivariante del problema.

Voy a entender como análisis multivariante, el estudio conjunto de los grupos de variables e individuos. Por lo tanto todo el estudio se basará en las distintas consideraciones sobre la matriz de datos.

Definimos la matriz de datos tal que:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1s} \\ x_{21} & x_{22} & \dots & x_{2s} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{n1} & x_{n2} & \dots & x_{ns} \end{pmatrix}$$

matriz de n individuos y S variables, donde cada fila es un individuo y cada columna es una variable.

Esta matriz la podemos poner en forma vectorial en el espacio de los individuos, es decir n-dimensional.

$$X = (X_{.1}, X_{.2}, \dots, X_{.S}) =$$

$$= \begin{pmatrix} X_{11} \\ X_{21} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{n1} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{n2} \end{pmatrix}, \dots, \begin{pmatrix} X_{1S} \\ X_{2S} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{nS} \end{pmatrix}$$

En este espacio cada eje representa a un individuo y las variables estarán representadas por vectores. P. ej.

$$\text{La variable } X_{.i} = \begin{pmatrix} X_{1i} \\ X_{2i} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{ni} \end{pmatrix}$$

También podemos representar la matriz X en el espacio S-dimensional de las variables, en el que cada individuo es un vector y cada eje una variable.

$$X' = (X_{1.}, X_{2.}, \dots, X_{n.}) =$$

$$= \begin{pmatrix} X_{11} \\ X_{12} \\ \cdot \\ \cdot \\ \cdot \\ X_{1S} \end{pmatrix}, \begin{pmatrix} X_{21} \\ X_{22} \\ \cdot \\ \cdot \\ \cdot \\ X_{2S} \end{pmatrix}, \dots, \begin{pmatrix} X_{n1} \\ X_{n2} \\ \cdot \\ \cdot \\ \cdot \\ X_{nS} \end{pmatrix}$$

Supuestas estas dos divisiones de la matriz de datos, tendremos las siguientes descomposiciones posibles, que dan lugar al análisis multivariante. (Quiero hacer constar aquí de nuevo que este esquema no pretende ser exhaustivo, pero si suficiente para los objetivos del presente estudio). (1)

a) $(X_{.1}), (X_{.2} \dots \dots \dots X_{.S})$. Es decir descomponemos en una variable por un lado y los demás por el otro. Es el caso de la correlación múltiple entre una variable y las demás.

(1).- Hay muchos esquemas de análisis multivariante desde los más sencillos como el exponemos en nuestro trabajo que es similar al de Saporta (1.976) hasta los más complejos como el de Cailliez y Pages (1.976) - (capítulo octavo); también empleado por Pages, Cailliez y Escoufier (1.979) y Koblinsky (1.979).

b) $(X_{.1}, X_{.2}, X_{.3}, \dots, X_{.p}), (X_{.p+1}, X_{.p+2}, \dots, X_{.S})$. Descomponemos en dos grupos de variables, las P primeras y las S-P restantes. Es el caso de la correlación canónica.

Es preciso especificar aquí que si en vez de dos grupos de variables, tuviéramos más, por ejemplo R grupos estaríamos ante el caso de la correlación canónica generalizada, de la cual hay varios modelos según que sea lo que se pretende. No vamos a entrar en ello pues habría que estudiar los distintos modelos y esto se sabe de los límites del presente trabajo (1). De todas formas se puede aplicar lo que se diga sobre la canónica, simplemente generalizándolo a varios grupos de variables.

c) Podemos hacer grupos de variables para cambiar los ejes o disminuir el número de variables. Es el caso del análisis de componentes principales y del análisis factorial. (2)

(1) Un estudio muy reciente sobre el tema ha sido expuesto con claridad por Baigorri (1.979) en su tesis doctoral sobre generalizaciones de la correlación canónica, dirigida por el Dr. Fernández de Troconiz.

Es interesante el trabajo de Kettenring J.R. - (1971) y también el de Carroll J.D. (1.968) entre otros muchos.

Un caso de canónica generalizada perfectamente deducible del esquema que vamos a exponer puede verse en el trabajo de Saporta (1.976) pags. 77 y siguientes.

(2) Hablaremos en su momento de la Bibliografía sobre el tema, baste decir aquí que entre otros muchos autores está como más clásico Harman (1.960 - 1.967 - 1.976) que además contiene abundante bibliografía.

En el caso de que todas o algunas de las variables sean cualitativas, tendremos el análisis de correspondencias (aunque también puede servir para cuantitativas). (1)

d) 1) $(X_1, X_2, \dots, X_p), (X_{p+1}, \dots, X_n)$. En este caso descomponemos en dos grupos de individuos, lo que da lugar al análisis discriminante de individuos en dos grupos y problemas de clasificación y al análisis de la varianza.

2) Si son varios grupos tendremos:

$(X_1, X_2, \dots, X_p), (X_{p+1}, \dots, X_q), \dots, (X_{t+1}, \dots, X_n)$. Lo que da lugar al análisis discriminante con varios grupos y problemas de clasificación y al análisis de la varianza. Y desde otro punto de vista al cluster análisis. (2)

Visto este esquema creo que es interesante el recalar que fundamentalmente nos vamos a mover en el espacio de los individuos, dado que como el estudio se va a basar en el análisis de los grupos de variables y éstas se representan como vectores en el espacio de los individuos, la representación geométrica es más clara.

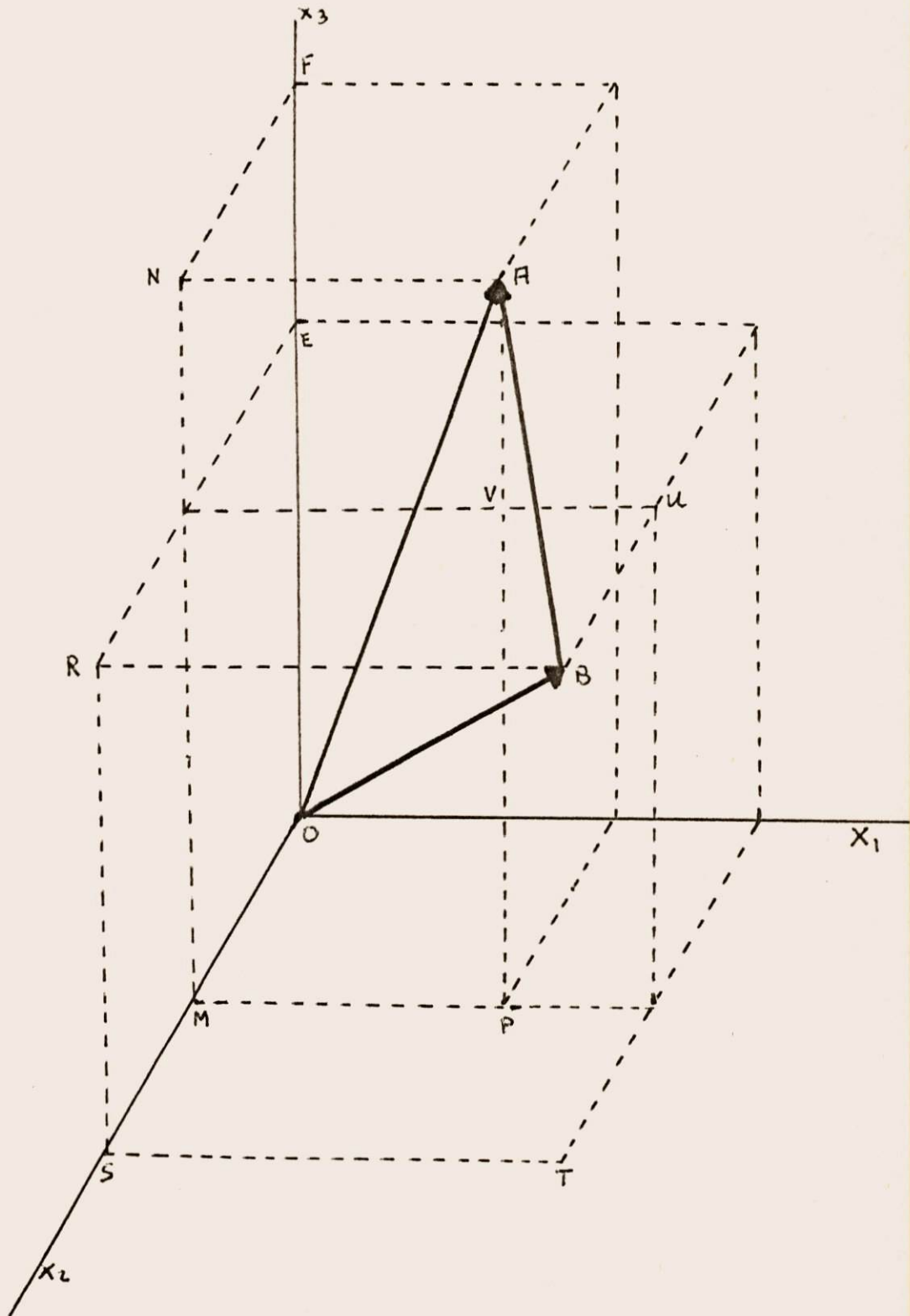
- (1) Benzécri (1.973) es el primero de estos estudios que han sido recogidos posteriormente por diversos autores. Entre otros Saporta (1.976), Jöreskog, Klován y Reyment (1.976), etc.
- (2) Sobre análisis discriminante, cluster, análisis de la varianza etc. veremos la abundante bibliografía existente cuando tratemos específicamente estos temas..

1.- ESPACIOS VECTORIALES EUCLIDIANOS. CONCEPTO ELEMENTAL DE PROYECTOR. ESQUEMA DE UNIFICACION.

En primer lugar quiero aclarar que en general nos vamos a referir a espacios euclidianos. La razón de esto es - doble, en primer lugar por claridad expositiva y en segundo lugar porque la visión en el espacio euclidiano es más intuitiva y sencilla. También y mientras no se especifique lo contrario, la métrica a utilizar en el presente estudio será la Euclidiana. La generalización a otras métricas (Mahalanobis, χ^2 , etc.) no ofrece dificultad, por lo que no la trataremos.

En general, los casos que vamos a tratar, consistirán en, o bien maximizar correlaciones, o bien minimizar distancias entre variables (cuantitativas o cualitativas cuantitativizadas) que serán vectores en el espacio de los individuos. La dimensión de cada vector será la raíz cuadrada de una suma de cuadrados. Para minimizar distancias bastará con minimizar la dimensión del vector diferencia que será a su vez una raíz cuadrada de una suma de cuadrados. Y para minimizar el vector diferencia habrá que hacer que éste sea perpendicular y con ángulo mínimo entre los vectores de partida.

Vamos a ver esto con un ejemplo gráfico de espacios vectoriales euclidianos que nos aclare los conceptos.



Supongamos que tenemos un vector OA y otro OB tal como se ve en el gráfico.

$$OA = \sqrt{OM^2 + MP^2 + MN^2} = \sqrt{x_{21}^2 + x_{11}^2 + x_{31}^2}$$

$$OB = \sqrt{ST^2 + OS^2 + SR^2} = \sqrt{x_{12}^2 + x_{22}^2 + x_{32}^2}$$

$$BA = \sqrt{VU^2 + BU^2 + VA^2} = \sqrt{(x_{11} - x_{12})^2 + (x_{21} - x_{22})^2 + (x_{31} - x_{32})^2}$$

tengamos en cuenta que $BU^2 = SM^2$ y $VA^2 = EF^2$

Los valores que hagan mínimo BA, serán aquellos que hagan que el ángulo AOB sea mínimo y que BA sea perpendicular a OB. (OA se considera fijo).

Para conseguir que sea perpendicular, tenemos que si $\cos AOB = \gamma$; $OB = \gamma \cdot OA$

Supongamos ahora que OA es un vector dado en el espacio de los individuos y que OB es un vector combinación lineal de la matriz X.

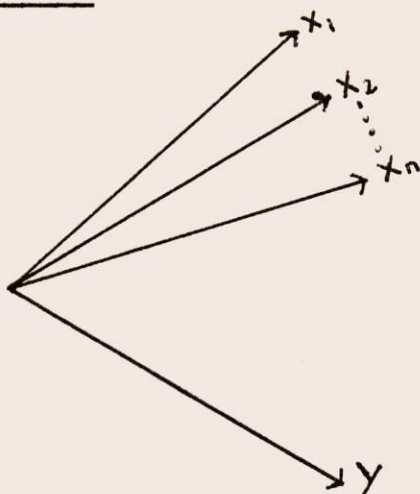
$$OB = X \underline{a}$$

Siendo a un vector que puede variar. Si buscamos el vector OB, que haga mínimo BA, bastará con que BA sea perpendicular a OB, pues OA es fijo y el ángulo mínimo se dará en la perpendicular, ya que OB pertenece al espacio X, que es fijo, y donde no está la variable OA.

Luego bastará hacer $OB = \gamma \cdot OA$, siendo γ , lo que denominamos proyector, (1), que en el caso elemental de dos vectores es el coseno y con grupos de variables toma formas más complejas.

Con todo esto, y considerando que las variables cualitativas (p. ej. en el análisis discriminante la división en grupos es una variable cualitativa), las podremos cuantitativizar o al menos utilizarlas como cuantitativas, (2), tendremos que el análisis multivariante se reduce a tres grandes casos teniendo en cuenta que las variables son vectores en el espacio de los individuos).

Caso 1º



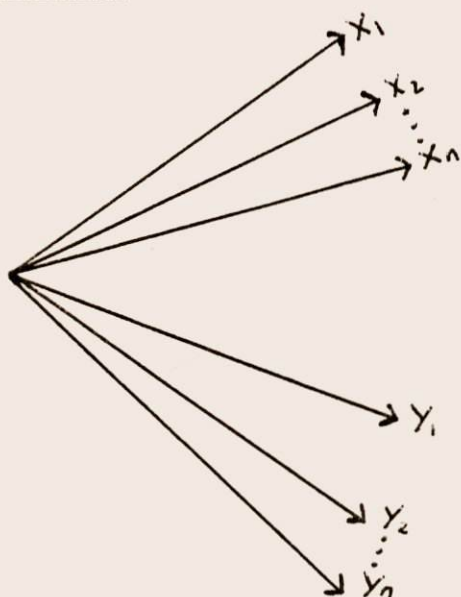
- (1) El concepto de proyector viene matemáticamente estudiado en múltiples tratados. Lo aplica Saporta (1.976) y Pages, Cailliez y Escoufier (1.979) entre otros.

Muy reciente, en este sentido, es el trabajo de Margolis (1.979).

- (2) Estudiaremos posteriormente la aplicación de las variables "dummy" a estos problemas.

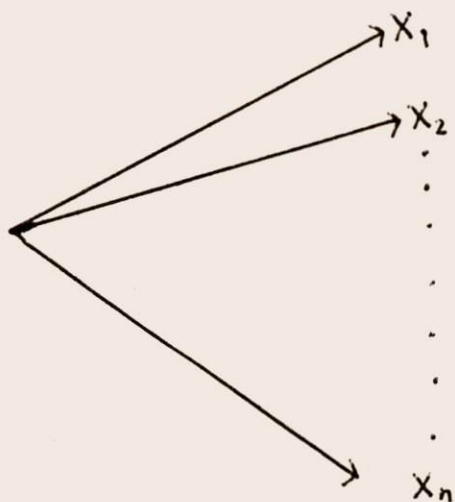
grupo de variables X , representado por una variable desconocida, y una variable conocida Y .

Caso 2º



Dos grupos, cada uno representado por una variable desconocida.

Caso 3º



Un grupo representado por una variable desconocida.

En el primer caso intentaremos minimizar la distancia entre la variable desconocida y la variable Y conocida.

En el segundo caso trataremos de minimizar las distancias entre las dos variables desconocidas (por lo que el caso primero es un caso particular de éste en el que una de las dos variables es conocida).

En el tercer caso impondremos alguna condición a la variable desconocida. (1)

Todo esto lo intentaremos hacer por medio del proyector y para ello pasaremos al siguiente capítulo, donde definiremos el proyector en el caso más sencillo que es el de la correlación múltiple. (2)

Nota: A lo largo de esta primera parte, y si no se indica lo contrario, consideraré las variables normalizadas.

- (1) Este esquema lo completaremos y explicaremos en el capítulo V.
- (2) La correlación múltiple al tratar más de dos variables, veremos que da lugar al proyector de forma matricial, y este proyector será ya aplicable en el futuro a todos los demás casos.

II.- CORRELACION MULTIPLE. DEFINICION DE PROYECTOR. (1)

Sea \underline{a} el vector (matriz columna) de coeficientes a_i . Sea \underline{Y} el vector de valores de la variable Y en los individuos, siendo Y el valor verdadero, mientras y es el calculado.

$$y_j = \sum_i a_i x_{ji} \quad \text{luego } X \underline{a} = \underline{y}$$

En primer lugar, tal como hemos visto en el gráfico, $X \underline{a}$ ha de ser ortogonal a $Y - X \underline{a}$

$$\text{luego } (X \underline{a})'(Y - X \underline{a}) = 0$$

$$a'x'Y = a'x'xa \Rightarrow x'Y = x'xa \Rightarrow \underline{a} = (x'x)^{-1} x'Y$$

Si lo vemos ahora por el método de mínimos cuadrados lo que habrá que hacer es: $\text{Min } (Y - xa)^2$.

$$\begin{aligned} \Phi &= (Y - xa)'(Y - xa) = Y'Y - Y'xa - a'x'y + a'x'xa = \\ &= Y'Y - 2a'x'y + a'x'xa. \end{aligned}$$

derivando e igualando a cero:

$$\frac{d\Phi}{da} = -2x'Y + 2x'xa = 0$$

$$x'xa = x'Y \Rightarrow \underline{a} = (x'x)^{-1} x'Y$$

- (1) El tema de la correlación múltiple está tratado en forma clásica en casi todos los textos de estadística.

Entre los especializados están Johnston (1.975) y Kmenta (1.977) como ejemplos de textos claros y completos.

Los primeros trabajos se deben a Pearson y a Yule.

Luego los resultados son idénticos, por el método de mínimos cuadrados o haciendo que la diferencia sea ortogonal.

tenemos que $\underline{a} = (x'x)^{-1} x'Y$

premultiplicando ambos vectores por X tendremos:

$$X(x'x)^{-1} x'Y = X \underline{a}$$

luego:

$$x(x'x)^{-1} x'Y = \underline{y}$$

es decir que:

(vector calculado) = $x(x'x)^{-1} x'$ (proyector) Y (verda
dera)

relacionando esto con el gráfico

$$OB = \gamma . OA$$

luego el proyector es

$$\gamma = x(x'x)^{-1} x'$$

Veremos que este proyector se puede aplicar a todos los casos a estudiar y que gracias a él se puede unificar su estudio.

Empecemos definiendo el proyector y luego razonaremos su utilización:

Llamamos proyector de un vector Y (N - dimensional) sobre un espacio $p + 1$ - dimensional $(1, X_1, \dots, X_p)$ (de las variables) a la matriz simétrica e idempotente que multiplicada por el vector nos da la proyección ortogonal del mismo sobre el espacio $p + 1$ - dimensional. ($P + 1$ si las variables no tienen promedio cero, si lo tienen sería P).

Entonces el proyector convierte al vector Y en otro vector perteneciente al espacio $P + 1 \leq N$ - dimensional -- ($p + 1 \leq N$; pues si $p + 1 = N$ el problema está determinado y no tiene sentido (1) minimizar las distancias al cuadrado pues se pueden hacer cero) que a su vez es la combinación lineal de los $P + 1$ vectores X , tal que este vector Xa sea perpendicular a $(Y - Xa)$.

Esto resuelve el caso de la correlación múltiple (y el de la simple como caso particular) pero en general se puede aplicar a todo el análisis multivariante pues el proyector lo que hace es proyectar sobre el espacio de X cualquier vector de forma que la proyección se hace perpendicularmente; con este instrumento veremos a lo largo de esta primera parte del trabajo que se pueden resolver los problemas del análisis multivariante.

- (1) En realidad al estar determinado el sistema, solo tiene una solución (si la tiene, es decir si es compatible) y no tiene sentido la minimización, pues solo se tratará de resolver el sistema.

Hemos dicho que el proyector es una matriz simétrica e idempotente, veamos por qué:

Llamemos P al proyector:

$$PY = y$$

hemos dicho que $Y - PY$ ha de ser perpendicular a PY , luego:

$$(PY)'(Y - PY) = 0$$

operando:

$$Y'P'Y - Y'P'PY = 0 \Rightarrow Y'P'Y = Y'P'PY$$

$(Y'P'PY)' = Y'P'PY$ luego es simétrica.

$Y'P'Y$ también tiene que ser simétrica.

luego $Y'P'Y = Y'PY$; premultiplicando por Y

$YY'P'Y = YY'PY$; premultiplicando por $(YY')^{-1}$

$$P'Y = PY \Rightarrow P'YY' = PYY'$$

postmultiplicando por $(YY')^{-1}$

$$\boxed{P' = P} \text{ luego el proyector es una matriz simétrica c.}$$

q.d.

utilizando el mismo proceso si tenemos que:

$$Y'P'Y = Y'P'PY$$

premultiplicando primero por Y y luego por $(YY')^{-1}$

$$P'Y = P'PY$$

postmultiplicando primero por Y' y luego por $(YY')^{-1}$

$$P' = P'P$$

$$P = P'P$$

luego es simétrica e idempotente c.q.d.

Por otro lado se puede ver que es simétrica e idempotente por su propia construcción:

$$(X (X'X)^{-1} X')' = X(X'X)^{-1} X'$$

dado que $(X'X)^{-1}$ es simétrica; luego $X(X'X)^{-1} X'$ es simétrica por otro lado:

$$\begin{aligned} (X (X'X)^{-1} X')' (X (X'X)^{-1} X') &= X(X'X)^{-1} X'X (X'X)^{-1}X' = \\ &= X(X'X)^{-1} X' \end{aligned}$$

luego es también idempotente c.q.d.

Veamos todo lo anterior con un ejemplo sencillo: (el problema es que sólo podemos representar gráficamente tres individuos y entonces el máximo de variables, supuestos con promedio cero, sería dos, luego lo haremos como una correlación simple). (1)

	X	Y
Z ₁	1	-2
Z ₂	0	1
Z ₃	-1	1

$$P = X (X'X)^{-1} X'$$

El sistema es igual que en la correlación múltiple, -

(1) En este caso utilizamos variables centradas.

pues la simple no es más que un caso particular de la múltiple.

$$X = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} ; X' = (1 \ 0 \ -1)$$

$$X'X = 2$$

$$(X'X)^{-1} = 0'5$$

$$X(X'X)^{-1} = \begin{pmatrix} 0'5 \\ 0 \\ -0'5 \end{pmatrix}$$

$$X(X'X)^{-1}X' = \begin{pmatrix} 0'5 & 0 & -0'5 \\ 0 & 0 & 0 \\ -0'5 & 0 & 0'5 \end{pmatrix} = P$$

$$P = P'; \quad P = P'P$$

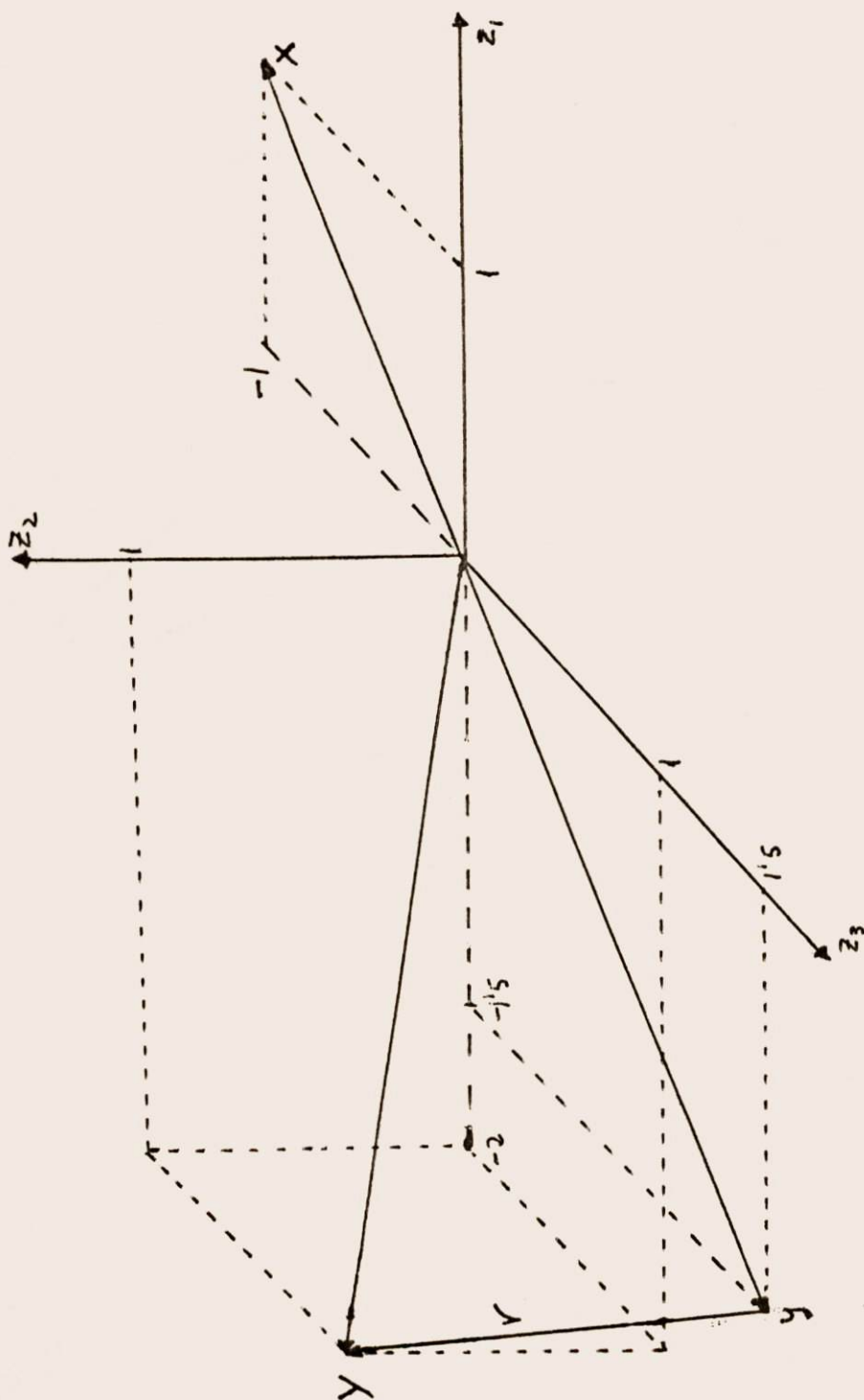
luego es simétrica e idempotente como queríamos comprobar.

$$P.Y = \begin{pmatrix} 0'5 & 0 & -0'5 \\ 0 & 0 & 0 \\ -0'5 & 0 & 0'5 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1'5 \\ 0 \\ 1'5 \end{pmatrix} = X\alpha$$

$$\alpha = (X'X)^{-1}X'Y = (0'5 \ 0 \ -0'5) \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = -1'5$$

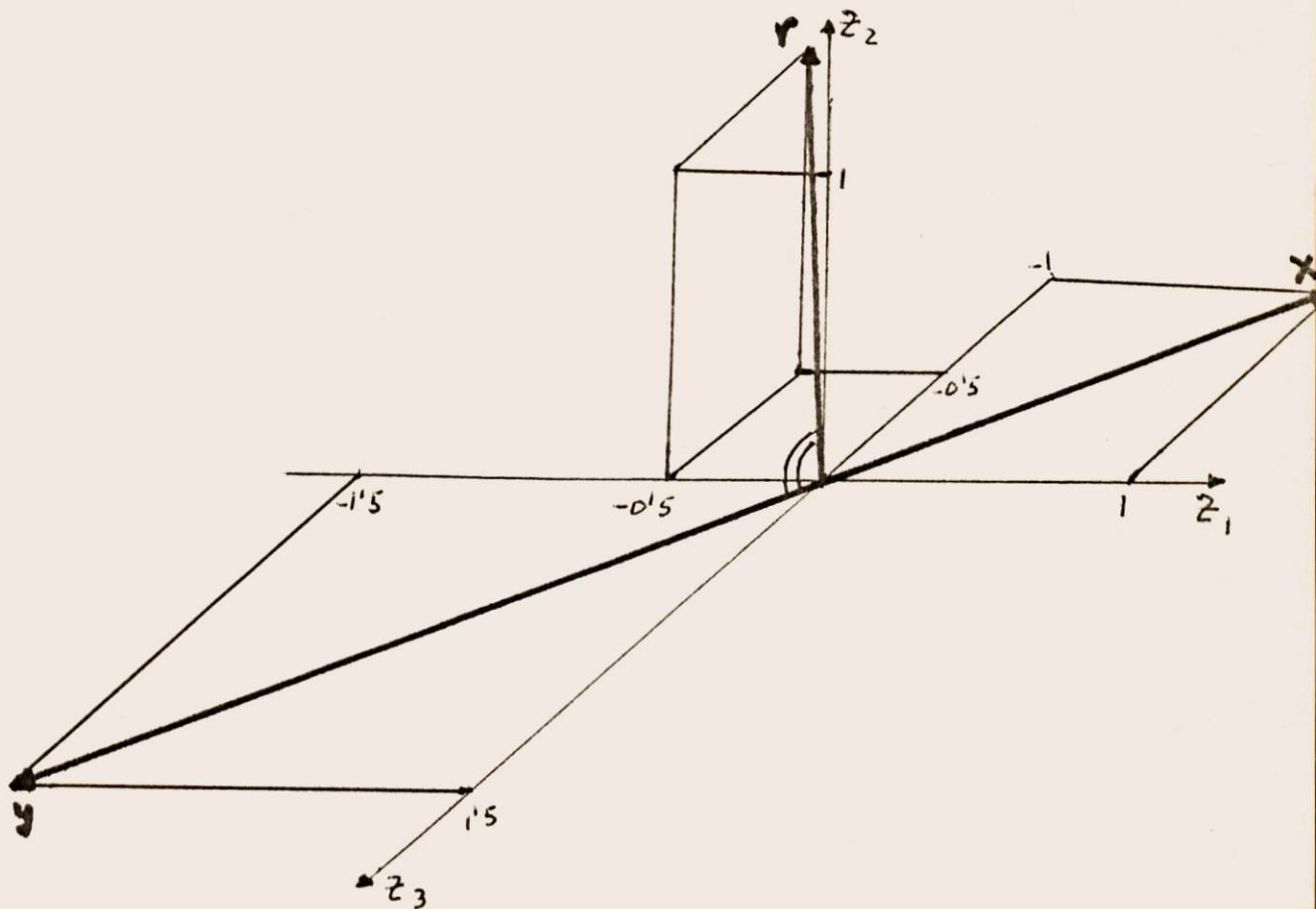
$$Xa = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} (-1'5) = \begin{pmatrix} -1'5 \\ 0 \\ 1'5 \end{pmatrix} = y$$

gráficamente:



$$r = Y - y = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -1'5 \\ 0 \\ 1'5 \end{pmatrix} = \begin{pmatrix} -0'5 \\ 1 \\ -0'5 \end{pmatrix}$$

lo que tiene que suceder es que r sea perpendicular a y . Veamos lo gráficamente trasladando r al centro de coordenadas.



analíticamente:

$$y' \cdot r = (-1'5 \ 0 \ 1'5) \begin{pmatrix} -0'5 \\ 1 \\ -0'5 \end{pmatrix} = 0'75 - 0'75 = 0$$

luego son perpendiculares.

III.- CORRELACION CANONICA.

Aquí tenemos las S variables en dos grupos de p y q variables respectivamente.

En el espacio n -dimensional cada eje es un individuo pero a la vez corresponde a:

- 1) Un conjunto de valores de las P variables tales que:

$$\sum_{i=1}^P a_i X_{1i} = Y_{11}$$

Y_{11} será el valor de la variable canónica $Y_{.1}$ correspondiente al primer dato; e igual se puede hacer con cada individuo en su eje.

- 2) Similarmente con las q variables restantes tendremos que:

$$\sum_{j=1}^q b_j X_{1j} = Y_{12}$$

El vector $Y_{.1}$ será aquel cuyas coordenadas sean:

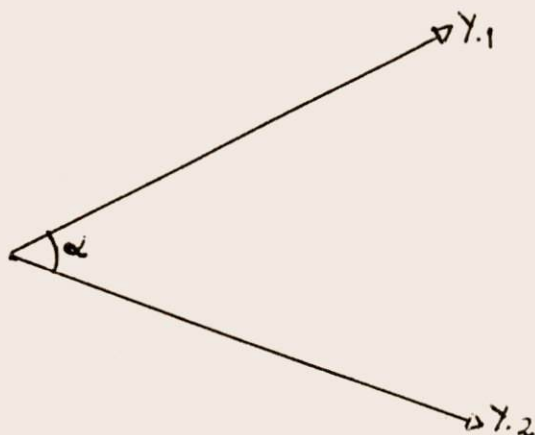
$$\sum_{i=1}^P a_i X_{ri}$$

Por lo tanto:

$$Y_{.1}^2 = \sum_{r=1}^n \left(\sum_{i=1}^P a_i X_{ri} \right)^2 = \sum_{r=1}^n Y_{r1}^2$$

$$Y_2^2 = \sum_{r=1}^n \left(\sum_{j=1}^q b_j X_{rj} \right)^2 = \sum_{r=1}^n Y_{r2}^2$$

Estas dos variables quedarán gráficamente:



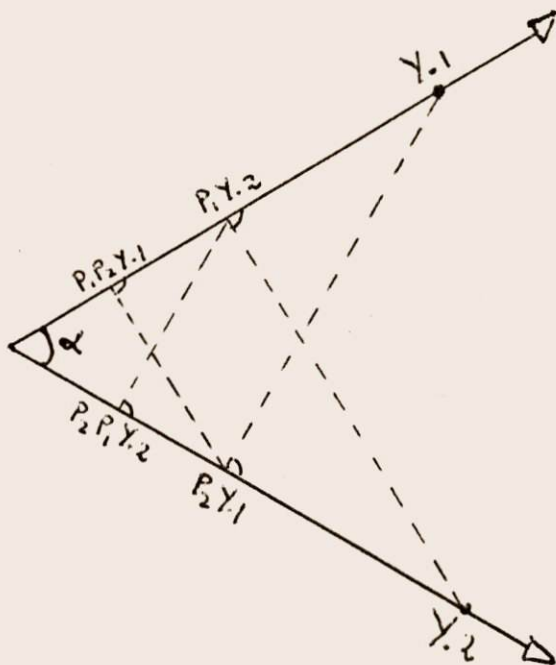
Como las dos son desconocidas, a diferencia de la correlación múltiple donde una era conocida, y lo que queremos hacer es minimizar el ángulo α , tendremos que utilizar el projector doblemente. (1)

Primero hemos de proyectar sobre el espacio de las q variables (q -dimensional) un vector desconocido, que es el $Y_{.1}$. Para ello utilizaremos el proyector que tiene la propiedad de hacerlo ortogonalmente, y como ya hemos visto minimizando el ángulo α . Pero el resultado de esta proyección es desconocido ya que $Y_{.1}$ es una incógnita. Lo que se puede hacer es volver a proyectar nuevamente sobre el espacio P -dimensional y el resultado será $\lambda Y_{.1}$, siendo λ un escalar.

(1) La idea básica de este enfoque se puede ver en Saporta (1.976) pags. 45 y siguientes.

Haciendo el mismo proceso con $Y_{.2}$, es decir, proyectando sobre el espacio P-dimensional, y después el resultado sobre el q-dimensional, el resultado final será $\lambda Y_{.2}$.

Veamos esto de una forma gráfica y sencilla con los vectores $Y_{.1}$ e $Y_{.2}$.



de los vectores $Y_{.i}$ lo que nos interesa es su dirección, ya que su dimensión nos es indiferente dado que el ángulo es realmente lo que queremos minimizar (y α no depende de la dimensión) es por esta causa que la colocación del punto $Y_{.1}$ es completamente arbitraria.

El vector $Y_{.1}$ es desconocido, pero proyectado mediante P_2 en el segundo subespacio (el q-dimensional) y vuelto a proyectar mediante P_1 en el primer subespacio es evidente que:

$$P_1 P_2 Y_{.1} = \lambda Y_{.1} = \lambda X_1 a$$

De la misma manera con el vector $Y_{.2}$ tendremos:

$$P_2 P_1 Y_{.2} = \lambda Y_{.2} = \lambda X_2 b$$

donde P_1 es el proyector que proyecta sobre el subespacio uno es decir:

$$P_1 = X_1 (X_1' X_1)^{-1} X_1'$$

y P_2 es el que proyecta sobre el segundo subespacio, tal que:

$$P_2 = X_2 (X_2' X_2)^{-1} X_2'$$

y el resultado $X_1 a = Y_{.1}$

$$X_2 b = Y_{.2}$$

es tal que dadas las propiedades del proyector, maximiza la correlación canónica entre las p variables del grupo uno y las q variables del grupo dos.

Con lo cual la solución del problema será:

$$X_1 (X_1' X_1)^{-1} X_1' X_2 (X_2' X_2)^{-1} X_2' X_1 a = \lambda X_1 a$$

$$X_2 (X_2' X_2)^{-1} X_2' X_1 (X_1' X_1)^{-1} X_1' X_2 b = \lambda X_2 b$$

donde si λ es un escalar tendremos:

$$(x_1' x_1)^{-1} x_1' x_2 (x_2' x_2)^{-1} x_2' x_1 a = \lambda a$$

$$(x_2' x_2)^{-1} x_2' x_1 (x_1' x_1)^{-1} x_1' x_2 b = \lambda b$$

donde λ es el valor propio y \underline{a} y \underline{b} son los vectores propios soluciones del problema.

Si hacemos M tal que:

$$M = (x_1' x_1)^{-1} x_1' x_2 (x_2' x_2)^{-1} x_2' x_1$$

y N tal que:

$$N = (x_2' x_2)^{-1} x_2' x_1 (x_1' x_1)^{-1} x_1' x_2$$

tendremos que:

$$M a = \lambda a = \lambda I a$$

$$N b = \lambda b = \lambda I b$$

lo que implica que:

$$(M - \lambda I) a = 0$$

$$(N - \lambda I) b = 0$$

que tiene una solución trivial:

$$b = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \end{pmatrix}, \quad a = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \end{pmatrix}$$

para que exista una solución distinta de cero:

$$|M - \lambda I| = 0$$

$$|N - \lambda I| = 0$$

luego la solución consistirá en diagonalizar las matrices M y N y el valor mayor de λ estará asociado a la mayor correlación canónica.

En el gráfico podemos ver que λ tiene el mismo valor en ambos casos por tratarse de triángulos semejantes. Por otro lado cuanto mayor sea λ (que siempre será menor que uno) (1) mayor será la correspondiente a la correlación canónica máxima.

En tercer lugar vimos que el proyector está asociado al coseno del ángulo α (en nuestro caso se trata de un coseno impropio dado que hablamos de más de dos variables), el producto de proyectores lo podemos asociar al $\cos^2 \alpha$, y dado que:

- (1) En realidad se puede dar el valor $= 1$ cuando los dos vectores canónicos coinciden, que sería el caso de correlación perfecta, luego $\lambda \leq 1$.

$$P_1 P_2 Y_{.1} = Y_{.1}$$

$$P_2 P_1 Y_{.2} = Y_{.2}$$

Vemos que λ es equivalente al producto de proyectores luego de alguna forma estará relacionado con el cuadrado del coeficiente de correlación canónica. Vamos a demostrar que es exactamente el coeficiente de correlación canónica al cuadrado, y de esta manera se esclarecerá el significado de λ .

Utilizaremos para ello el método tradicional de obtención de la correlación canónica (1) pero utilizando notación matricial:

Supongamos cuatro variables y n individuos (se puede hacer con más variables sin ningún problema, pero no lo hacemos para que resulte más cómodo).

$$Y = \begin{pmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ \cdot & \cdot \\ \cdot & \cdot \\ X_{1n} & X_{2n} \end{pmatrix} ; \quad X = \begin{pmatrix} X_{31} & X_{41} \\ X_{32} & X_{42} \\ \cdot & \cdot \\ \cdot & \cdot \\ X_{3n} & X_{4n} \end{pmatrix}$$

(suponemos como siempre variables normalizadas).

- (1) El primer modelo de correlación canónica se debe a Hotelling (1.935 y 1.936). Posteriormente ha sido muy tratado en diversidad de estudios.

El trabajo de Hotelling fue desarrollado en términos más simples y con más aplicaciones por Waugh (1.942).

A partir de la formulación de Waugh yo he desarrollado matricialmente el método tradicional para compararlo con el de los proyectores. Puede verse en Chacón (1.955) pags. 232 - 235.

$$C_i = a_1 X_{1i} + a_2 X_{2i} ; C = Y a$$

$$P_i = b_1 X_{3i} + b_2 X_{4i} ; P = X b$$

Llamando R al coeficiente de correlación canónica:

$$R = r_{cp} = C' D_p P = a' y' D_p X b = a' R_{yx} b \quad -1-$$

D_p = matriz diagonal de términos $1/n$

R_{yx} = matriz de coeficientes de correlación tal que:

$$R_{yx} = \begin{pmatrix} r_{x_1 x_3} & r_{x_1 x_4} \\ r_{x_2 x_3} & r_{x_2 x_4} \end{pmatrix}$$

Llamando a $r_{x_i x_j}$, r_{ij} ; tendremos:

$$R = r_{cp} = a_1 b_1 r_{13} + a_1 b_2 r_{14} + a_2 b_1 r_{23} + a_2 b_2 r_{24}$$

$$R_{cc} = a' R_{yy} a = (a_1 \ a_2) \begin{pmatrix} X_{11} \dots \\ X_{21} \dots \end{pmatrix} \begin{pmatrix} 1/n & & \\ & \cdot & \\ & & 1/n \end{pmatrix} \begin{pmatrix} X_{11} & X_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} =$$

$$= a' y' D_p Y a = a_1^2 r_{11} + 2 a_1 a_2 r_{12} + a_2^2 r_{22} = a_1^2 + 2 a_1 a_2 r_{12} + a_2^2 = 1 \quad -2-$$

tiene que ser igual a uno pues se trata de la correlación de la variable canónica consigo misma.

Por la misma razón:

$$R_{pp} = b' R_{xx} b = b' X' D_p X b = b_1^2 + 2 b_1 b_2 r_{34} + b_2^2 = 1 \quad -3-$$

Hagamos ahora un máximo condicionado por el método de Lagrange:

$$\begin{aligned} \Phi &= (1) + \theta_1 (2) + \theta_2 (3) \\ &= a' R_{yx} b + \theta_1 a' R_{yy} a + \theta_2 b' R_{xx} b \end{aligned}$$

Haciendo los derivados parciales e igualando a cero:

$$\frac{\partial \Phi}{\partial a_1} = r_{13} b_1 + r_{14} b_2 + 2 \theta_1 a_1 + 2 \theta_2 a_2 r_{12} = 0$$

$$\frac{\partial \Phi}{\partial a_2} = r_{23} b_1 + r_{24} b_2 + 2 \theta_1 a_2 + 2 \theta_2 a_1 r_{12} = 0$$

luego tenemos que:

$$\begin{bmatrix} r_{13} & r_{14} \\ r_{23} & r_{24} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + 2 \theta_1 \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 0$$

que podíamos haber obtenido directamente utilizando la derivación matricial:

$$\frac{\partial \Phi}{\partial a} = R_{yx} b + 2 \theta_1 R_{yy} a = 0 \quad -4-$$

Haciendo lo mismo respecto a b:

$$\frac{\partial \Phi}{\partial b_1} = r_{13} a_1 + r_{23} a_2 + 2 \vartheta_2 b_1 + 2 \vartheta_2 b_2 r_{34} = 0$$

$$\frac{\partial \Phi}{\partial b_2} = r_{14} a_1 + r_{24} a_2 + 2 \vartheta_2 b_1 r_{34} + 2 \vartheta_2 b_2 = 0$$

esto es lo mismo que:

$$\begin{bmatrix} r_{13} & r_{23} \\ r_{14} & r_{24} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + 2 \vartheta_2 \begin{bmatrix} 1 & r_{34} \\ r_{43} & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = 0$$

o haciéndolo directamente mediante matrices:

$$\frac{\partial \Phi}{\partial b} = R'_{yx} a + 2 \vartheta_2 R_{xx} b = 0 \quad -5-$$

Vamos a hacer ahora una serie de operaciones:

Multiplicando -4- por $R'_{yx} R^{-1}_{yy}$ tendremos:

$$R'_{yx} R^{-1}_{yy} R_{yx} b = -2 \vartheta_1 R'_{yx} a \quad -6-$$

Multiplicando ahora -5- por ϑ_2

$$\vartheta_2 R'_{yx} a = -2 \vartheta_2^2 R_{xx} b \quad -7-$$

Multiplicando -4- por a'

$$a' R_{yx} b = -2 \theta_1 a' R_{yy} a$$

Pero por -2- $a' R_{yy} a = R_{cc} = 1$

Pero por -1- $a' R_{yx} b = r_{cp} = R$

Luego tendremos que $R = -2 \theta_1$

Multiplicando -5- por b'

$$b' R'_{yx} a = -2 \theta_2 b' R_{xx} b$$

Por -3- $b' R_{xx} b = R_{pp} = 1$

$$b' R'_{yx} a = a' R_{yx} b \quad (\text{por ser un escalar})$$

$$a' R_{yx} b = r_{cp} = R \quad (\text{por -1-})$$

luego tendremos que $R = -2 \theta_2$

Entonces con lo visto antes tenemos que:

$$R = -2 \theta_1 = -2 \theta_2 \quad \Rightarrow \quad R = -2 \theta \quad -8-$$

Con -6- y -7- podemos ver que:

$$R'_{yx} R_{yy}^{-1} R_{yx} b = -2 \theta_1 R'_{yx} a$$

$$\hat{\theta}_2 R'_{yx} a = -2 \hat{\theta}_2^2 R_{xx} b$$

Pero por -8- $\hat{\theta}_1 = \hat{\theta}_2 = \theta$

$$R'_{yx} R_{yy}^{-1} R_{yx} b = 4 \theta^2 R_{xx} b$$

Y por -8-

$$\boxed{R_{xx}^{-1} R'_{yx} R_{yy}^{-1} R_{yx} b = R^2 b}$$

Comparando esta fórmula con la obtenida por proyecto res que era:

$$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X b = \lambda b$$

$$(X' X) = N R_{xx} \Rightarrow (X' X)^{-1} = \frac{1}{N} R_{xx}^{-1}$$

$$(X' Y) = (Y' X)' = N R'_{yx}$$

$$(Y' Y) = N R_{yy} \Rightarrow (Y' Y)^{-1} = \frac{1}{N} R_{yy}^{-1}$$

$$(Y' X) = N R_{yx}$$

Luego sustituyendo:

$$R_{xx}^{-1} R'_{yx} R_{yy}^{-1} R_{yx} b = \lambda b$$

de donde se deduce que $\lambda = R^2$

Igual proceso se puede hacer para obtener el valor de a y veríamos que obteníamos el mismo valor de λ , que por su puesto sigue siendo igual a R^2 .

Para terminar si $\lambda = R^2$, nunca puede ser mayor que uno, y habrá que buscar el máximo valor de λ , que será el que maximice el coeficiente de correlación canónica.

Luego hemos demostrado analíticamente lo que intuíamos geométricamente, que resumiendo es lo siguiente:

$$\longrightarrow \lambda = R^2$$

$\longrightarrow \lambda$ tiene el mismo valor para cualquiera de las - dos ecuaciones canónicas.

$\longrightarrow \lambda \leq 1$, y hemos de buscar el λ máximo que será el que maximice la correlación canónica.

Veamos a continuación como la correlación múltiple es un caso particular de la correlación canónica:

EN correlación múltiple tenemos una matriz X y un vector Y e intentamos correlacionar X e Y

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix}$$

El proyector P lo hemos definido ya como:

$$P = X (X' X)^{-1} X'$$

$P Y = X b$ y por lo tanto:

$$X (X' X)^{-1} X' Y = X b \quad b = (X' X)^{-1} X' Y$$

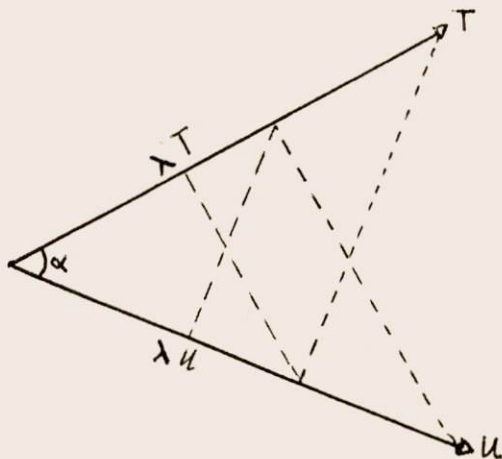
Haciéndolo por canónica:

$$X (X' X)^{-1} X' Y (Y' Y)^{-1} Y' X b = \lambda X b$$

Lo que tiene que suceder es que las b coincidan y en tonces λ será el coeficiente de correlación múltiple al cuadrado, pues siendo λ el coeficiente de correlación canónica al cuadrado equivaldrá al de la múltiple al cuadrado cuando se haga correlación canónica de un grupo de variables con un vector.

Veamos por qué coinciden las b :

En correlación canónica teníamos:

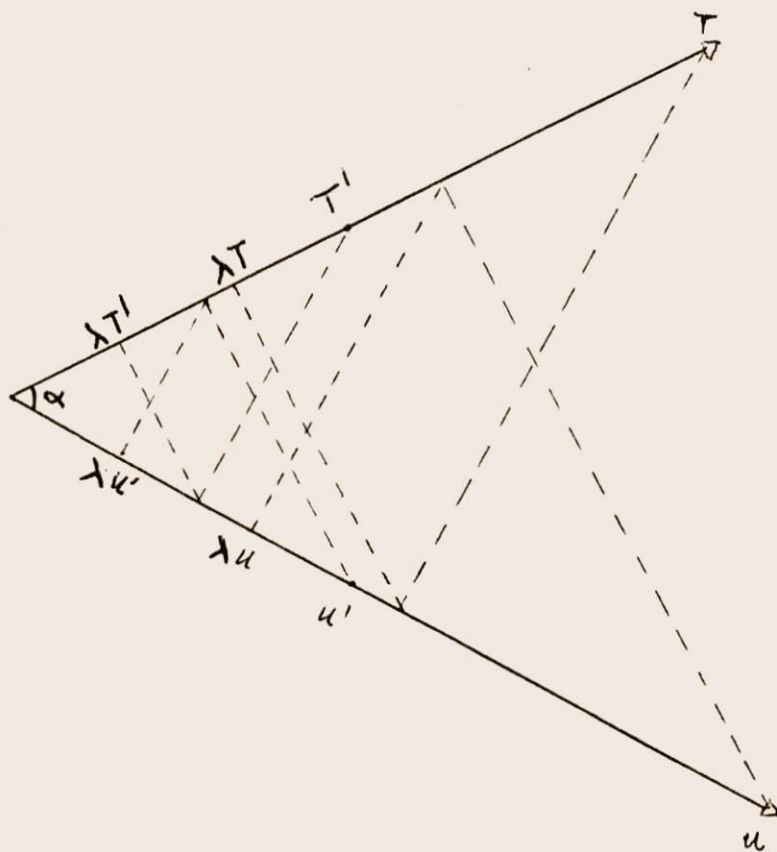


Lo que no podríamos hacer es definir las dimensiones de T y de U, por lo que teníamos infinitas soluciones proporcionales de los vectores a y b. En realidad lo que teníamos es una solución trivial.

$$a = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} ; b = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

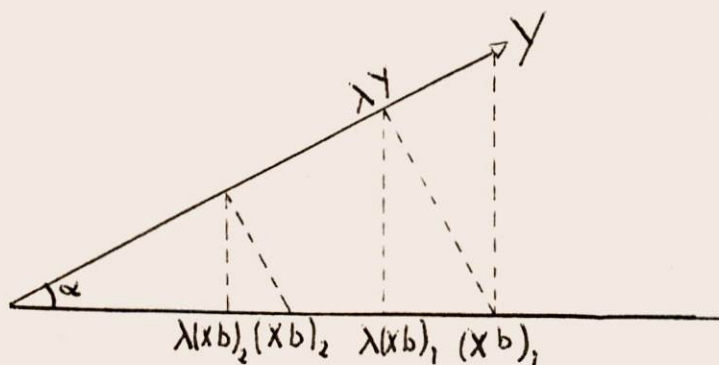
y luego infinitas soluciones proporcionales.

Luego igual de válido es T, U; que T', U'



pues definen el mismo ángulo α .

En el caso que nos ocupa uno de los dos vectores canónicos es fijo y es Y (en realidad habría que hablar de Y a siendo a un escalar pero a no es relevante). Entonces tendremos (haciendo $a = 1$):



$(X b)_1$ y $(X b)_2$ son soluciones, ambas válidas, del problema canónico. Pero $(X b)_1$ es además $P_1 Y$, que es la solución de la correlación múltiple, y por lo tanto coincide la solución de correlación múltiple con una de las infinitas posibles de correlación canónica.

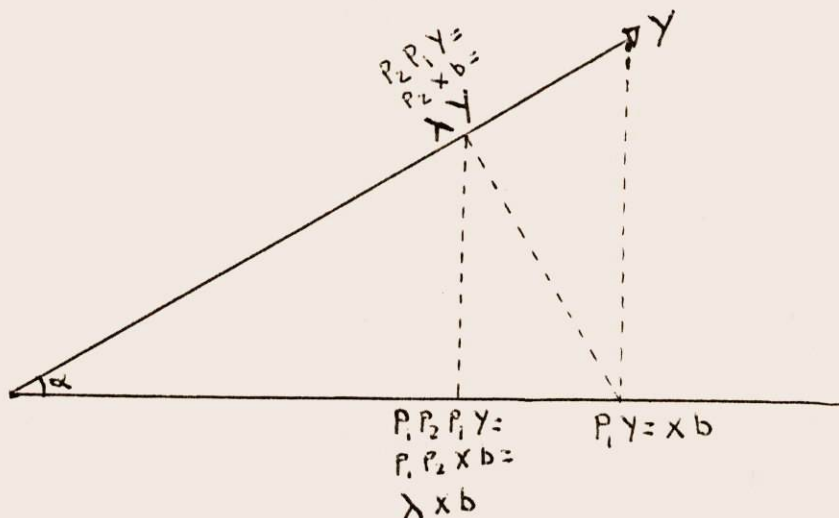
Sustituyendo $b = (X' X)^{-1} X' Y$ en la fórmula de la canónica que era:

$$X (X' X)^{-1} X' Y (Y' Y)^{-1} Y' X b = \lambda X b$$

tendremos que:

$$X (X' X)^{-1} X' Y (Y' Y)^{-1} Y' X (X' X)^{-1} X' Y = \lambda X b$$

Es decir que si proyectamos Y sobre X , el resultado sobre Y y luego otra vez sobre X , nos va a dar $\lambda(X b)$. -
 Viendo esto en el gráfico tenemos:



Con lo visto hasta ahora está claro que las b coinciden y por otro lado hemos expuesto gráficamente como son similares el proceso de correlación múltiple y el de canónica (cuando se trata de un grupo de variables y una variable), gracias a la operatividad del proyector. Vamos a continuación a demostrar que en este caso el coeficiente de correlación canónica al cuadrado $-\lambda-$ es lo mismo que el coeficiente de correlación múltiple al cuadrado.

Tenemos que:

$$(Y' Y)^{-1} Y' X (X' X)^{-1} X' Y a = \lambda a$$

pero a no es relevante, pues es un escalar que en nada modifica la dirección de Y , luego podemos considerarle un valor arbitrario, por ejemplo 1.

$$(Y' Y)^{-1} Y' X (X' X)^{-1} X' Y = \lambda$$

Donde podemos ver que λ es un escalar.

Pero por lo que hemos visto antes:

$$b = (X' X)^{-1} X' Y$$

luego sustituyendo:

$$\underline{(Y' Y)^{-1} Y' X b = \lambda}$$

Si ahora tomamos la fórmula del coeficiente de correlación -R- tal como se expresa en cualquier tratado de estadística -en el caso de datos centrados-. (1)

$$R^2 = \frac{b' X' Y}{Y' Y} = (Y' Y)^{-1} b' X' Y$$

Pero $b' X' Y = Y' X b$ por ser un escalar.

$$\text{Luego } = R^2 = (Y' Y)^{-1} Y' X b$$

Demostrado por lo tanto que la correlación múltiple se puede considerar un caso particular de la correlación canónica, vamos a ver un ejemplo sencillo: (con datos centrados).

(1) Se puede consultar el libro de Johnston (1.975) pag. 138.

Y	1	-1	2	-2
X ₁	-3	1	3	-1
X ₂	0	-2	1	1

$$X' X = \begin{pmatrix} 20 & 0 \\ 0 & 6 \end{pmatrix}$$

$$(X' X)^{-1} = \begin{pmatrix} 0'05 & 0 \\ 0 & 0'1\widehat{66} \end{pmatrix}$$

$$(X' X)^{-1} X' = \begin{pmatrix} -0'15 & 0'05 & 0'15 & -0'05 \\ 0 & -0'3\widehat{3} & 0'1\widehat{66} & 0'1\widehat{66} \end{pmatrix}$$

$$b = (X' X)^{-1} X' Y = \begin{pmatrix} 0'2 \\ 0'3\widehat{3} \end{pmatrix}$$

$$y = 0'2 X_1 + 0'3\widehat{3} X_2$$

Este es el resultado por el método de correlación múltiple, aplicando los proyectores que ya vimos que es equivalente al método de mínimos cuadrados.

El coeficiente de correlación múltiple será:

$$R^2 = \frac{b' X' Y}{Y' Y} = 0'14\widehat{66}$$

$$X' Y = \begin{pmatrix} 4 \\ 2 \end{pmatrix} ; \quad b' X' Y = 1'4\widehat{66}$$

$$Y' Y = 10$$

Veamos ahora por canónica como encontramos la b y R^2 .
La a no la buscamos pues ya hemos comentado que no resulta relevante.

$$(Y' Y)^{-1} = 1/10; \quad (Y' Y)^{-1} Y' = (0'1 \quad -0'1 \quad 0'2 \quad -0'2)$$

$$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' = \begin{pmatrix} 0'02 & -0'02 & 0'04 & -0'04 \\ 0'0\widehat{33} & -0'0\widehat{33} & 0'0\widehat{66} & -0'0\widehat{66} \end{pmatrix}$$

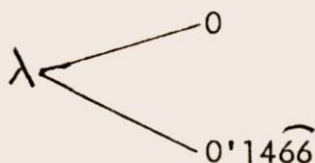
$$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X = \begin{pmatrix} 0'08 & 0'04 \\ 0'1\widehat{33} & 0'0\widehat{66} \end{pmatrix} = M$$

$$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X b = \lambda b$$

luego lo que hay que hacer es diagonalizar M

$$\begin{vmatrix} 0'08 - \lambda & 0'04 \\ 0'1\widehat{33} & 0'0\widehat{66} - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - 0'14\widehat{66} \lambda + 0 = 0$$



Cogemos la mayor y tendremos $\lambda = 0.1466 = R^2$ (de la múltiple), para obtener b:

$$-0.066 \widehat{b}_1 + 0.04 \widehat{b}_2 = 0$$

$$\widehat{b}_2 = \frac{0.066}{0.04} \widehat{b}_1$$

$$\widehat{b}_2 = 1.66 \widehat{b}_1$$

Luego para el vector b tenemos infinitas soluciones proporcionales, una de las cuales coincide con la solución de la correlación múltiple, tal como habíamos visto en el gráfico.

$$\text{si } \widehat{b}_1 = 0.2 \quad \widehat{b}_2 = 0.2 \cdot 1.66 = 0.33$$

como queríamos comprobar.

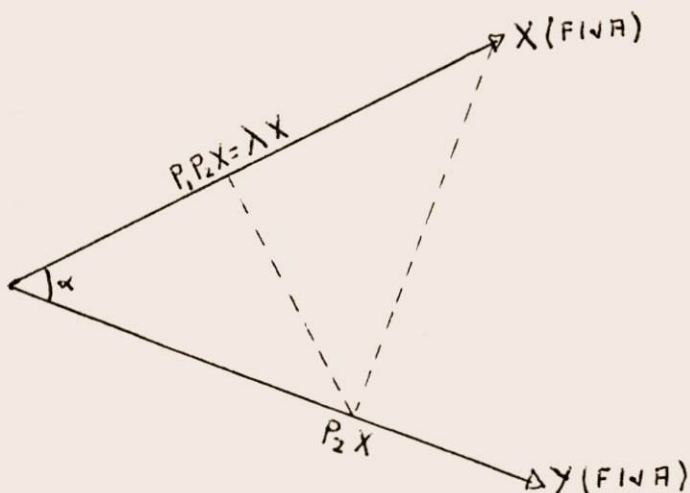
Luego en conclusión el coeficiente de correlación es el mismo haciéndolo por canónico o por múltiple, pues sólo depende del ángulo que forman los vectores.

El vector b, tiene por canónica infinitas soluciones proporcionales pues solo pretende determinar el ángulo. Por correlación múltiple, tiene una sola solución, que coincide

con una de las infinitas de la canónica.

Veamos a modo de simple comentario como la correlación simple es un caso particular de la canónica. (Siendo la múltiple, es evidente que ha de serlo también la simple, pero - de todas formas vamos a verlo brevemente).

El caso planteado gráficamente es:



Este es el planteamiento del caso por correlación ca nónica, donde al igual que antes la dimensión de Y a, la - consideramos tal que:

$$a = 1 ; Y a = Y$$

Por lo tanto lo que nos interesa es X que es lo que proyectamos.

$$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X b = \lambda b$$

Pero $(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X = \lambda$, por tratarse de un escalar.

Tomando como hacíamos antes la fórmula normal del coeficiente de correlación simple -R-. (1)

$$R^2 = \frac{(X' Y)^2}{(X' X) (Y' Y)}$$

Pero $(X' X)^{-1}$, $(X' Y)$, $(Y' Y)^{-1}$, $(Y' X)$, son todos escalares luego podemos hacer:

$$\frac{(X' Y) (Y' X)}{(X' X) (Y' Y)} = R^2$$

Luego como era de esperar el coeficiente de correlación coincide.

El valor de b según la fórmula tradicional es:

$$b = \frac{\sum X Y}{\sum X^2} = \frac{X' Y}{X' X} = (X' X)^{-1} X' Y$$

que es la obtenida para la correlación múltiple.

Si queremos obtener el valor de b mediante la canónica nos resultará imposible ya que:

(1) Se puede consultar el libro de Johnston (1.975) pag. 138.

$(X' X)^{-1} X' Y (Y' Y)^{-1} Y' X = \lambda$ es un escalar y habrá in finitos valores de b. Esto se debe a que el valor de b no influye para nada en el ángulo, al ser este fijo, por estar entre dos vectores fijos.

Luego en conclusión: la correlación simple es un ca so particular de la múltiple a todos los efectos.

La correlación múltiple es un caso particular de la correlación canónica, si atendemos a que el coeficiente de correlación es el mismo por un camino o por otro. Pero con la particularidad de que la correlación canónica nos da co mo solución un infinito número de vectores b, todos propor cionales entre sí, mientras que la correlación múltiple nos da un vector b determinado, que es uno de los infinitos pro porcionales anteriores y que es el que corresponde al caso de la regresión múltiple.

Por otro lado también podemos afirmar que la correla ción canónica es una generalización de la correlación múlti ple.

Luego a efectos prácticos terminamos este capítulo - afirmando que todo el estudio de correlaciones (simple, múl tiple y canónica) se puede resumir en la correlación canóni ca, siendo el proyector el instrumento utilizado en todos - los casos, dado que la única diferencia está en qué es lo

que proyectamos (variable o grupo de variables) y sobre lo que proyectamos (variable o grupo de variables) y siempre con el objetivo de maximizar la correlación. Por eso el caso general es la correlación canónica donde se proyecta grupo sobre grupo.

La generalización de la correlación canónica a varios grupos da lugar a los diversos modelos de correlación canónica generalizada (1). Ya comentamos en la introducción que se salía fuera del plan de estudio que vamos a seguir.

(1) El estudio más reciente sobre correlación canónica generalizada es el de Baigorri (1.979), ya citado en la introducción.

IV.- COMPONENTES PRINCIPALES. (1)

Hemos visto hasta ahora las posibilidades de proyección de grupo o variable, sobre grupo o variable, o dicho de otra manera de variable desconocida o conocida, sobre variable desconocida o conocida; pero siempre se trataba de proyectar un espacio sobre otro espacio representados éstos bien por variables desconocidas o conocidas.

Lo que ahora pretendemos es proyectar un espacio sobre un vector perteneciente a ese espacio.

Recordemos en primer lugar nuestra matriz de datos.

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & \cdot & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & \cdot & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & \cdot & \cdot & \cdot & X_{np} \end{pmatrix}$$

Donde los valores están normalizados.

(1) La literatura sobre análisis factorial (componentes principales, etc.) es muy extensa. Un libro clásico de análisis factorial es el Harman (1.967). Reedición revisada del Harman (1.960), nuevamente editado en 1.976.

Actualmente la literatura sobre el tema es muy amplia, pero considero que como libros más interesantes en el momento actual, a parte del ya citado de Harman, están Jöreskoy, Klovan y Reyment (1.976), Laffitte (1.972), y también podemos citar al Profesor de la Universidad Politécnica de Barcelona Torrens-Ibern (1.972).

X' representará a los individuos en el espacio de las variables. Es en este espacio en el que nos vamos a mover - ahora.

Lo que pretende el análisis de componentes principales es encontrar el vector U , dentro del espacio de las variables, que maximice la varianza. Es decir un vector U que haga que su varianza sea máxima.

Sea un vector U tal que:

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ U_p \end{pmatrix}$$

p -dimensional, por encontrarse en el espacio de las variables.

El problema consiste ahora en proyectar la matriz X' sobre el vector U , de manera que su varianza sea máxima.

En vez de proyectar sobre el vector U , ya que todo es tá normalizado, vamos a proyectar sobre el vector V tal que:

$$V = \frac{U}{\sqrt{U' U}}$$

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ V_p \end{pmatrix}$$

El proyector, sobre el espacio del vector V , será:

$$P_V = V (V' V)^{-1} V'$$

$$\text{Pero } V' V = \frac{U'}{\sqrt{U' U}} \cdot \frac{U}{\sqrt{U' U}} = \frac{U' U}{U' U} = 1$$

$$\text{Luego } P_V = V V'$$

La proyección quedará:

$$P_V X' = N$$

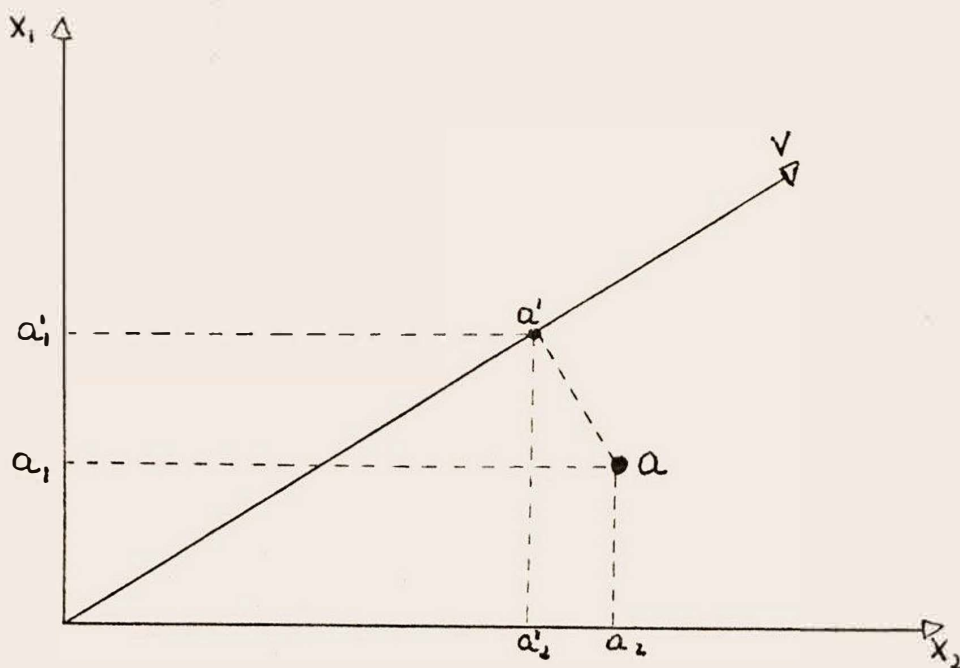
$$V V' X' = N$$

La forma de $V V'$ será:

$$\begin{pmatrix} V_1^2 & V_1 V_2 & \cdot & \cdot & V_1 V_p \\ V_2 V_1 & V_2^2 & \cdot & \cdot & V_2 V_p \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ V_p V_1 & V_p V_2 & \cdot & \cdot & V_p^2 \end{pmatrix}$$

N será una matriz que representa la proyección ortogonal de cada individuo en el vector V , tal que cada individuo viene representado por un vector de p componentes, - cada uno de los cuales es la coordenada en la variable correspondiente del individuo proyectado.

Gráficamente con dos variables sería:



El punto inicial es a , con coordenadas a_1, a_2 ; al proyectar tenemos que las coordenadas del punto a' son a'_1 y a'_2 .

Si llamamos I_{ij} a la coordenada de la proyección del individuo i en la variable j , la matriz N quedará:

$$N = \begin{pmatrix} I_{11} & I_{21} & \cdot & \cdot & \cdot & I_{n1} \\ I_{12} & I_{22} & \cdot & \cdot & \cdot & I_{n2} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ I_{1p} & I_{2p} & & & & I_{np} \end{pmatrix}$$

Siguiendo el gráfico anterior lo que queremos hacer es maximizar $(a')^2 + (b')^2 + \dots + (n')^2$. Siendo a' , b' , \dots , n' las proyecciones de los N individuos.

Pero $(a')^2 = (a'_1)^2 + (a'_2)^2$ y lo mismo pasa con los demás puntos. Por otro lado $a'_1 = I_{11}$ de la matriz N .

Luego lo que hay que maximizar es la traza de $N' N$.

$$N' N = X V V' V V' X' = X V V' X' \text{ pues } V' V = 1$$

Por las propiedades del operador traza:

$$\text{traza } X V V' X' = \text{traza } V' X' X V$$

De todas formas vamos a comprobarlo:

$$X V V' = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & X_{np} \end{pmatrix} \begin{pmatrix} V_1^2 & V_1 V_2 & \cdot & \cdot & V_1 V_p \\ V_2 V_1 & V_2^2 & \cdot & \cdot & V_2 V_p \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ V_p V_1 & V_p V_2 & \cdot & \cdot & V_p^2 \end{pmatrix} =$$

$$\begin{pmatrix} \sum X_{1i} V_i V_1 & \sum X_{1i} V_i V_2 & \cdot & \cdot & \sum X_{1i} V_i V_p \\ \sum X_{2i} V_i V_1 & \sum X_{2i} V_i V_2 & \cdot & \cdot & \sum X_{2i} V_i V_p \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \sum X_{ni} V_i V_1 & \sum X_{ni} V_i V_2 & \cdot & \cdot & \sum X_{ni} V_i V_p \end{pmatrix}$$

$$X V V' X' = X V V' \begin{pmatrix} X_{11} & X_{21} & \cdot & \cdot & X_{n1} \\ X_{12} & X_{22} & \cdot & \cdot & X_{n2} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ X_{1p} & X_{2p} & \cdot & \cdot & X_{np} \end{pmatrix} =$$

$$= \begin{bmatrix} \sum_j X_{1j} V_j & \sum_i X_{1i} V_i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \sum_j X_{2j} V_j & \sum_i X_{2i} V_i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sum_j X_{nj} V_j & \sum_i X_{ni} V_i \end{bmatrix}$$

Luego traza $X V V' X' = \sum_R \left(\sum_i X_{Ri} V_i \right)^2$

Pero $V' X' = \left(\sum V_i X_{1i} \quad \sum V_i X_{2i} \quad \cdot \quad \cdot \quad \cdot \quad \leq V_i X_{ni} \right)$

luego $V' X' X V = \sum_R \left(\sum_i V_i X_{Ri} \right)^2$

Por lo tanto: traza $X V V' X' = V' X' X V$

Pero $V = \frac{U}{\sqrt{U'U}}$ siendo $(U'U)$ un escalar.

Luego $V' X' X V = \frac{U' X' X U}{U' U} = \lambda$ es lo que queremos

maximizar.

derivando matricialmente respecto a U:

$$\frac{d \left(\frac{U' X' X U}{U' U} \right)}{d U} = \frac{(U' U) (2 X' X U) - (U' X' X U) (2 U)}{(U' U)^2} = 0$$

Por lo tanto:
$$\frac{X' X U}{U' U} = \frac{U' X' X U}{(U' U)^2} U$$

operando:
$$X' X U = \frac{U' X' X U}{U' U} U = \lambda U$$

Luego el resultado es: $X' X U = \lambda U = \lambda I U$

Por lo tanto $(X' X - \lambda I) U = 0$; para evitar la solución trivial de $U = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix}$; la solución es diagonalizar la matriz $X' X$. En realidad lo que hay que diagonalizar es $X' D_p X$ siendo $D_p = \begin{bmatrix} 1/n & & \\ & \cdot & \\ & & 1/n \end{bmatrix}$

o lo que es lo mismo $\frac{X' X}{n}$; lo que no afecta a los razonamientos, y así diagonalizamos la matriz de varianzas y covarianzas, que aquí coincide con la matriz de coeficientes de correlación. Pero en general vamos a prescindir de D_p para simplificar el tratamiento.

Una vez obtenidos los valores de λ tomamos el mayor y deducimos el valor de U , con lo que tenemos el primer componente principal. Podíamos repetir el proceso con los demás pero ^{5e}demuestra que las siguientes λ , provenientes de la diagonalización anterior, dan lugar a los siguientes componentes. (1)

(1) Puede verse en Tatsuoka (1.971) pag. 270.

Posteriormente tenemos el problema de la dimensión de U, pero esto tiene sentido a efectos interpretativos, no vamos a entrar pues en el problema, por salirse de los objetivos de este estudio. De todas formas diremos que nosotros en principio la dimensionaremos a uno.

Supuesta la dimensión uno tendremos la ecuación de variables en función de los componentes:

$$\text{Sea } W = \begin{pmatrix} U_{11} & U_{12} & \cdot & \cdot & \cdot & U_{1p} \\ U_{21} & U_{22} & \cdot & \cdot & \cdot & U_{2p} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ U_{p1} & U_{p2} & \cdot & \cdot & \cdot & U_{pp} \end{pmatrix}$$

Suponiendo que $X' X$ es de rango P

$$\text{Sea } Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_p \end{pmatrix} ; \text{ vector de variables.}$$

$$\text{Y llamemos } C = \begin{pmatrix} C_1 \\ C_2 \\ \cdot \\ C_p \end{pmatrix} ; \text{ al vector de componentes.}$$

La relación será:

$$\underline{Z = W \cdot C}$$

Pero al ser las U, ortogonales y de dimensión 1

$$W' W = I \Rightarrow W' = W^{-1}$$

Luego $\underline{C = W' Z}$

Con lo que tenemos los componentes en función de las variables.

Pasando a nuestra matriz concreta que era X tendremos que:

$$K' = W' X' = \begin{pmatrix} C_{11} & C_{21} & \cdot & \cdot & \cdot & C_{n1} \\ C_{12} & C_{22} & \cdot & \cdot & \cdot & C_{n2} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ C_{1p} & C_{2p} & \cdot & \cdot & \cdot & C_{np} \end{pmatrix}$$

por lo tanto $K = X W$; con lo que podemos saber el valor de cada individuo en cada componente.

Esto resulta equivalente a realizar el proceso en el espacio de los individuos, veamos por qué:

En el espacio de los individuos, las variables son vectores y los componentes principales también. Si tenemos todo dimensionado a uno, la dimensión de las variables, que será su varianza, será uno; al igual que los componentes.

Proyectaremos sobre el vector V que ahora tendrá N componentes y será tal que:

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \cdot \\ V_n \end{pmatrix}$$

proveniente de un vector U tal que:

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \cdot \\ U_n \end{pmatrix} ; \text{ donde } V = \frac{U}{\sqrt{U' U}}$$

El proyector $P_V = V (V' V)^{-1} V' = V V'$ igual que antes. Y la proyección quedaría:

$$M = P_V X = V V' X$$

La matriz M, tiene en cada columna el resultado de la proyección de cada variable. Las proyecciones elevadas al cuadrado y sumadas, son la variabilidad del vector. Y eso es lo mismo que la traza de M' M.

Luego lo que hay que maximizar es la traza de M' M o lo que es lo mismo: traza X' V V' V V' X = traza X' V V' X.

Por un razonamiento muy similar al que hemos visto antes, pero teniendo en cuenta que aquí V tiene N componentes, tenemos que:

traza X' V V' X = V' X X' V como se puede comprobar. Además es una propiedad del operador traza como ya hemos comentado.

Luego lo que hay que maximizar es:

$$V' X X' V = \frac{U' X X' U}{U' U} = \lambda$$

$$\frac{d \left(\frac{U' X X' U}{U' U} \right)}{d U} = \frac{(U' U) (2X X' U) - (U' X X' U) (2 U)}{(U' U)^2} = 0$$

operando: $X X' U = \lambda U = \lambda I U$

y como teníamos antes $(X X' - \lambda I) U = 0$

El resultado de la diagonalización de $X X'$ tiene que dar lo mismo que la diagonalización de $X' X$, luego tendremos las mismas λ . (Al igual que antes, ahora habría que diagonalizar $X D_p X'$).

Por otro lado sucede que estas U de N términos son proporcionales a las columnas de la matriz K , vista anteriormente, con lo que la matriz K , da una de las posibles soluciones de U y el modelo queda cerrado.

Llamemos, como anteriormente, C_i , al vector i de la matriz K y U_i al vector U correspondiente de los que antes veíamos de P elementos:

$$C_i = X U_i$$

Tendremos que demostrar que C_i tiene la mayor varianza posible con lo cual será una solución del problema en el espacio de los individuos, es decir, una solución de la U con N términos.

Varianza de $C_i = C_i' D_p C_i = U_i' X' D_p X U_i = \lambda$, que es la mayor posible, c.q.d. (Tengamos en cuenta que en este caso U es de dimensión uno dadas las condiciones del problema.)(1)

(1) Puede verse en Saporta (1.976) pag. 57

Lo que sucede es que normalmente no se agotan todos los componentes y nos quedamos con unos pocos que reúnan la mayor parte de la variabilidad. En este caso W^{-1} no tiene sentido y hay que utilizar otros métodos. Pero en realidad veremos que como $W' W = I$ se puede utilizar algo muy similar. (1)

Hasta aquí hemos visto lo que hemos denominado sistema de componentes principales. Que repito consiste simplemente en diagonalizar la matriz de varianzas y covarianzas. (2)

Un sistema similar es el de factores principales. La diferencia entre ambos no siempre está muy clara. Para algunos autores simplemente se diferencian según si se mantienen los unos en la matriz de coeficientes de correlación o si se sustituyen por los comunales. Esto segundo serían factores principales y lo primero, que matemáticamente es un cambio de ejes, serían componentes principales.

Tatsuoka (3) es más claro y dice que la diferencia entre ambos métodos, a parte de la citada, es que en componen-

- (1) Tengamos en cuenta que aunque W no sea cuadrada, al ser los componentes ortogonales $W'W = I$, si los componentes son de dimensión uno, sino será al menos diagonal.
- (2) Con variables normalizadas, que son las que utilizamos generalmente, esta matriz será la de coeficientes de correlación.
- (3) Tatsuoka (1.971)

tes principales se estudian todos los componentes y si se elimina alguno será porque viene causado por el azar. Factores principales por el contrario elimina todos los factores que no agrupen una determinada variabilidad.

El problema normalmente es estimar esa comunalidad (1), y así tenemos que Kendall y Stuart (2) sugieren con cierta ironía que el análisis factorial es fruto de la desesperación de los Psicólogos por intentar encontrar los factores comunes de la inteligencia. De todas formas considero exagerado esta posición aunque es evidente lo limitado del método.

Nosotros, aún entendiendo que la exposición de Tatsuoka es clara, vamos a limitarnos en lo sucesivo a llamar componentes principales cuando no se estimen las comunaldades y análisis factorial a lo contrario. Dada esta diferenciación, en general son mejores los componentes principales, que evitan los problemas de la comunalidad, si bien este concepto y en consecuencia la utilización del análisis factorial será necesario para saber lo que de común tienen las variables.

(1) Entendemos por comunalidad de una variable la suma de los cuadrados de los coeficientes de los factores comunes. Distinguiéndola de la unicidad que es la contribución del factor único. Definición dada por Harman (1970) pag. 17 -18. De una forma intuitiva es lo que de común tienen las variables.

(2) Kendall y Stuart (1.966), volumen III pags. 306 ss.

En general, lo que denominamos análisis factorial, es un caso particular de la aplicación de la técnica de componentes principales al caso en que sustituimos los unos de la matriz de coeficientes de correlación por las comunales, luego en general, y desde un punto de vista matemático la técnica básica son componentes principales y así la denominaremos en lo sucesivo.

Veamos ahora el problema desde otro punto de vista:

$$X' X = n \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdot & \cdot & \cdot & \sigma_{1p} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & & & & \sigma_p^2 \end{pmatrix}$$

Dejando la n por ser una constante, diremos que $X' X$ es la matriz de varianzas y covarianzas. (1)

$P_v X'$ es lo que antes hemos llamado N y lo que había que maximizar era la traza de $N' N$.

$$N' N = X P_v' P_v X'; \text{ por las propiedades del proyector}$$

$$N' N = X P_v X'$$

(1) Que al tratarse de variables normalizadas, coincide con la de coeficientes de correlación.

Fácil es comprobar que:

$$\text{traza } X P_V X' = \text{traza } P_V X' X$$

Pero $P_V X' X$ es la proyección de $X' X$ sobre V ; luego lo que hay que maximizar es:

$$\text{Max traza } V (V' V)^{-1} V' X' X = \text{traza } V V' X' X$$

Pero se puede comprobar que:

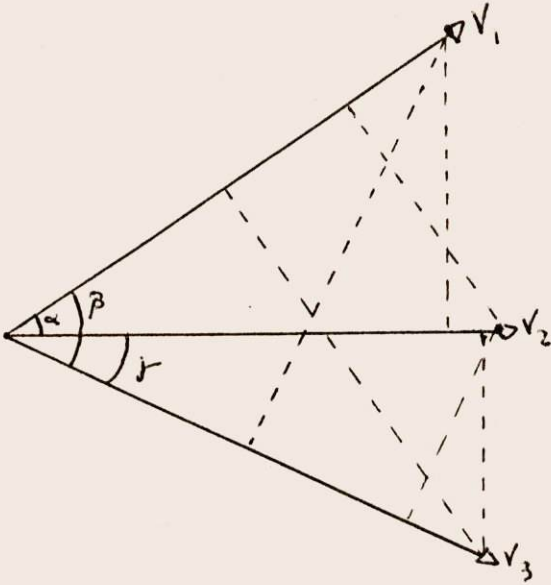
$$\text{traza } V V' X' X = \text{traza } V' X' X V$$

Sustituyendo por U , derivando llegamos como antes a:

$$X' X U = \lambda U$$

El interés de este planteamiento consiste en ver que lo que aquí proyectamos es la matriz de coeficientes de correlación, y por lo tanto cada variable viene representada por su dimensión ($\sigma_i^2 = 1$) y por los coeficientes de correlación con los demás. Pero al ser las variables normalizadas los coeficientes de correlación son los cosenos, que a su vez son las proyecciones de cada variable sobre las demás; luego en definitiva lo que proyectamos es: dentro del espacio de los individuos, las variables en función de las propias variables.

Veámoslo gráficamente:



Las variables vendrán determinadas de la siguiente manera:

$$\begin{aligned} \text{Variable 1} &\rightarrow V_1, V_1 \cos \alpha, V_1 \cos \beta \rightarrow \sigma_1^2, \sigma_{12}, \sigma_{13} \\ \text{Variable 2} &\rightarrow V_2 \cos \alpha, V_2, V_2 \cos \gamma \rightarrow \sigma_{12}, \sigma_2^2, \sigma_{23} \\ \text{Variable 3} &\rightarrow V_3 \cos \beta, V_3 \cos \gamma, V_3 \rightarrow \sigma_{13}, \sigma_{23}, \sigma_3^2 \end{aligned}$$

Es decir que determinando cada variable por la proyección de ella sobre las demás y sobre sí misma tenemos la matriz de varianzas y covarianzas. Al proyectar esta matriz y maximizar la traza de la matriz proyectada ya hemos visto que conseguimos los mismos efectos que con los procedimientos anteriores.

Otro camino más sencillo es crear el componente tal que:

$$Y = X V$$

Lo que hay que maximizar es $Y' Y = V' X' X V$, que es un resultado que ya nos resulta familiar.

En definitiva, hemos visto varios caminos para obtener los componentes principales. Todos ellos nos aportan algo, y sobre todo vemos que el proyector sigue teniendo absoluta vigencia, para resolver este tipo de problemas.

En conclusión hemos visto un nuevo caso que consiste en proyectar un grupo de vectores sobre un vector interno. Nosotros hemos puesto la condición de que la variabilidad - de ese nuevo vector desconocido sea máxima. Siempre habrá que poner una condición, pero cualquiera que sea ésta, el esquema de actuación es el mismo.

APLICACIONES SIMILARES AL ANALISIS DE COMPONENTES PRINCIPALES

La abundante literatura que hace referencia al análisis de componentes principales se engloba generalmente bajo la definición de análisis factorial. Este término es bastante ambiguo dado que cada autor le da un significado distinto. Aquí hemos seguido la diferenciación que parece más clara, entendiendo que la diferencia consiste en si se estiman o no las comunalidades y denominando componentes principales

al método general.

De todas formas hay autores que todavía hacen más distinciones como por ejemplo Jöreskog, Klovan y Reyment en su interesante libro *Geological Factor Analysis*, que especifican dos formas distintas que denominan R-modo y Q-modo. El R-modo sería el que nosotros hemos desarrollado en este capítulo, es decir estudiaría la relación entre las variables. El Q-modo estudiaría la relación entre los individuos. Para este estudio, los autores antes citados proponen el tratar a los individuos como variables y utilizar el coseno como medida de similitud, siguiendo a Imbrie y Purdy (1). Para ello en vez de utilizar la matriz X , la transforman en W :

$$W = D^{-1/2} X$$

$$H = D^{-1/2} X X' D^{-1/2}$$

donde $D = \text{diag. } (X X')$

Posteriormente se proponen otro tipo de medidas de semejanza que en definitiva son maneras de medir la distancia entre individuos.

(1) Imbrie y Purdy (1.962)

Finalmente Benzécri (1) ha desarrollado el que llama análisis factorial de correspondencias, que lo que intenta es obtener las ventajas de todos los planteamientos anteriores. El objetivo es, mediante la utilización de una métrica χ^2 , poder plasmar en los mismos ejes a variables y a individuos.

Su estudio, aun siendo aplicable a variables cuantitativas, trata a fondo las variables cualitativas, creando con ellas una tabla de frecuencias y a partir de ahí y con una métrica de χ^2 realiza una serie de transformaciones y aplica componentes principales.

Luego en definitiva, aplicando diferentes métricas, realizando transformaciones, estudiando bien variables, bien individuos, con variables cuantitativas o cualitativas, lo que se hace al final en todos estos sistemas, llamados genéricamente "análisis factorial" es aplicar el método de componentes principales que ya hemos visto que se puede tratar fácilmente con el proyector.

Nota: Hay otros métodos como el centroide, etc. pero no es objeto de este trabajo el estudiarlos.

- (1) Benzécri (1.973). Este libro en dos tomos, expone el análisis de correspondencias y de él se han inspirado los siguientes. El estudio de Benzécri resulta de difícil lectura y son los estudios basados en él, los que han tenido mayor divulgación.

Ejemplo:

Vamos a ver un ejemplo sencillo (tal que se pueda re presentar gráficamente) de aplicación de componentes princi pales.

En nuestro ejemplo vamos a utilizar variables centra das (no hace falta que estén normalizadas para lo que vamos a hacer, y tampoco multiplicaremos por D_p , por lo que las λ , darán multiplicadas por N).

Definamos la matriz X , tal que:

$$X = \begin{pmatrix} 2 & -1 \\ 0 & -1 \\ -2 & 2 \end{pmatrix}$$

Vamos a hallar el primer componente tal que:

$$V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

Lo que hemos de hacer según hemos visto, será maximi zar la traza de $X V V' X'$.

$$X V = \begin{pmatrix} 2 V_1 - V_2 \\ 0 V_1 - V_2 \\ -2 V_1 + V_2 \end{pmatrix}$$

Lo que habrá que maximizar será:

$$\begin{aligned} \text{Max: } & (2V_1 - V_2)^2 + (-V_2)^2 + (-2V_1 + 2V_2)^2 = \\ \text{Max: } & 4V_1^2 + V_2^2 - 4V_1V_2 + V_2^2 + 4V_1^2 + 4V_2^2 - 8V_1V_2 = \\ \text{Max: } & 8V_1^2 + 6V_2^2 - 12V_1V_2 \end{aligned}$$

Con la condición $V_1^2 + V_2^2 = 1$ (pues si no V_1 y V_2 tenderían a infinito).

Luego planteando el máximo condicionado:

$$\Phi = 8V_1^2 + 6V_2^2 - 12V_1V_2 - \lambda(V_1^2 + V_2^2)$$

$$\frac{\partial \Phi}{\partial V_1} = 16V_1 - 12V_2 - 2\lambda V_1 = 0$$

$$\frac{\partial \Phi}{\partial V_2} = 12V_2 - 12V_1 - 2\lambda V_2 = 0$$

$$\begin{pmatrix} 16 & -12 \\ -12 & 12 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = 2\lambda \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

$$\begin{pmatrix} 8 & -6 \\ -6 & 6 \end{pmatrix} - \lambda I \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = 0$$

Para tener soluciones relevantes:

$$\begin{vmatrix} 8 - \lambda & -6 \\ -6 & 6 - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - 14\lambda + 12 = 0 \quad \lambda_1 = 13'0828$$
$$\lambda_2 = 0'9172$$

$$\lambda_1 + \lambda_2 = 14$$

Para obtener el primer componente tomamos la mayor λ y tendremos:

$$-5'0828 V_1 - 6 V_2 = 0$$

$$-6 V_1 - 7'0828 V_2 = 0$$

Con infinitas soluciones proporcionales a:

$$\begin{pmatrix} -6 \\ 5'0828 \end{pmatrix}$$

Pero V tiene que ser de dimensión uno luego:

$$V_1 = \frac{-6}{\sqrt{(-6)^2 + 5'0828^2}} = -0'7630$$

$$V_2 = \frac{5'0828}{\sqrt{(-6)^2 + 5'0828^2}} = 0'6464$$

Luego el primer componente será:

$$V_1 = \begin{pmatrix} -0'7630 \\ 0'6464 \end{pmatrix}$$

El segundo componente se obtendrá a partir del segundo valor de λ .

$$7'0828 V_1 - 6 V_2 = 0$$

$$-6 V_1 + 5'0828 V_2 = 0$$

Con infinitas soluciones proporcionales a

$$\begin{pmatrix} 6 \\ 7'0828 \end{pmatrix}$$

dimensionando a uno:

$$V_1 = \frac{6}{\sqrt{6^2 + 7'0828^2}} = 0'6464$$

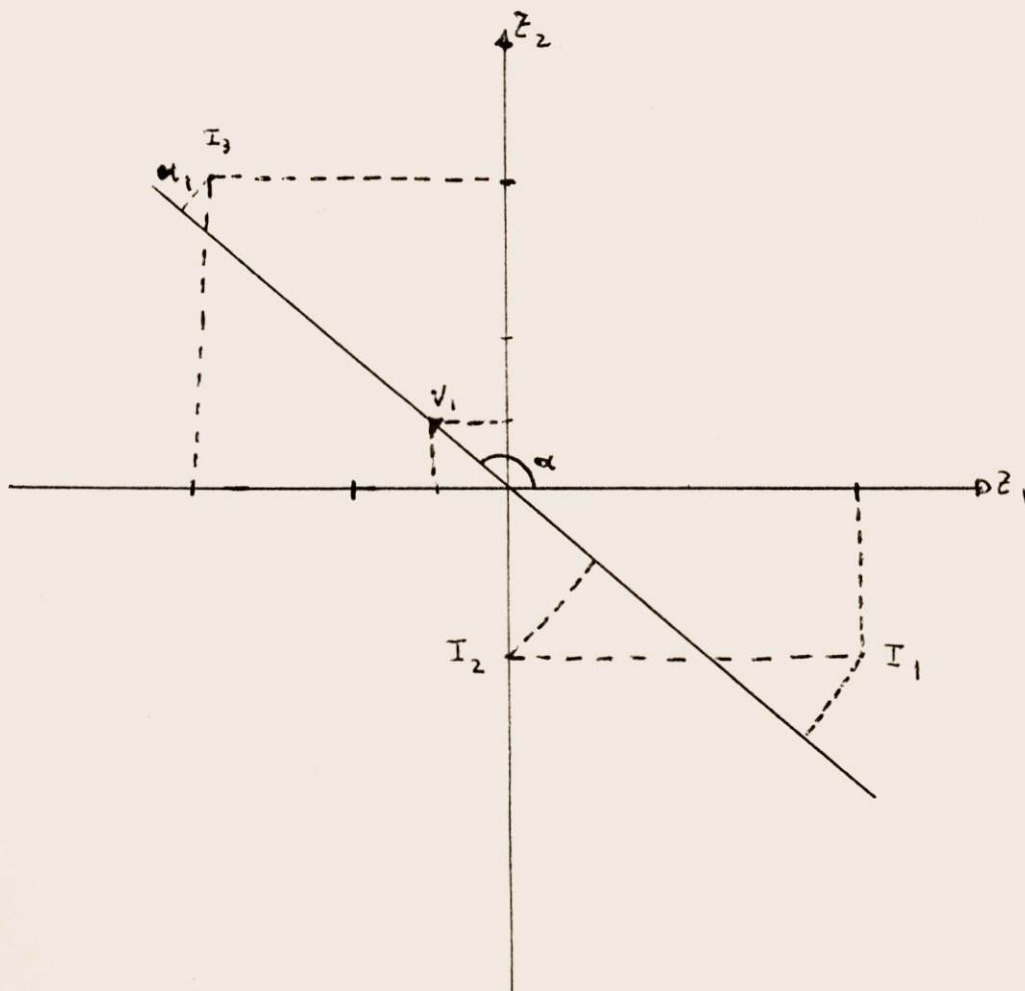
$$V_2 = \frac{7'0828}{\sqrt{6^2 + 7'0828^2}} = 0'7630$$

Luego el segundo componente será:

$$V_2 = \begin{pmatrix} 0'6464 \\ 0'7630 \end{pmatrix}$$

Vamos a estudiar el primer componente en el espacio de las variables:

$$V_1 = \begin{pmatrix} -0'7630 \\ 0'6464 \end{pmatrix}$$



tenemos que $V_1^2 + V_2^2 = 1$

$$\text{Luego } \frac{V_1}{\sqrt{V_1^2 + V_2^2}} = V_1 = \cos \alpha = -0'7630$$

$$V_2 = \sin \alpha = 0'6464$$

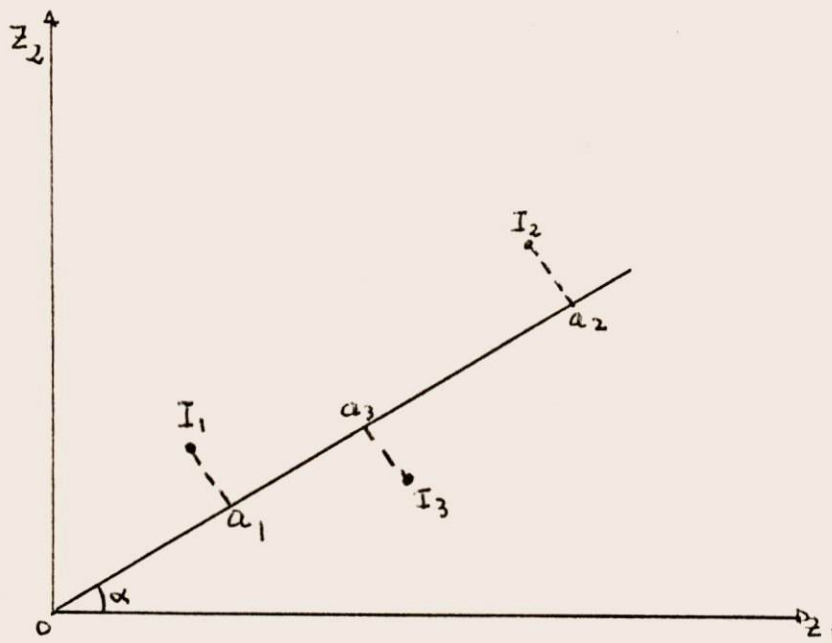
$$V_2/V_1 = \text{tg } \alpha = -0'8472$$

$$\alpha = \text{arc. cos } \alpha = \text{arc. sen } \alpha = \text{arc. tg } \alpha = 139'729$$

Con el ángulo α tenemos definido el vector U_1 (primer componente) que al ser de dimensión uno se convierte en el vector V_1 .

Veamos a continuación en un gráfico lo que estamos haciendo.

Supongamos dos variables Z_1 y Z_2 , con tres individuos tal que su situación en el gráfico sea la siguiente, con su componente principal U :



Lo que hacemos en el método de componentes principales es:

$$\text{Max: } \left[(0 \ a_1)^2 + (0 \ a_2)^2 + (0 \ a_3)^2 \right]$$

Pero por el teorema de Pitágoras:

$$(O I_i)^2 = (O a_i)^2 + (a_i I_i)^2$$

Luego podemos poner el máximo como:

$$\text{Max: } \left[(O I_1)^2 - (a_1 I_1)^2 + (O I_2)^2 - (a_2 I_2)^2 + (O I_3)^2 - (a_3 I_3)^2 \right]$$

Pero $(O I_1)^2 + (O I_2)^2 + (O I_3)^2$ es una constante que podemos sacar del máximo y nos queda:

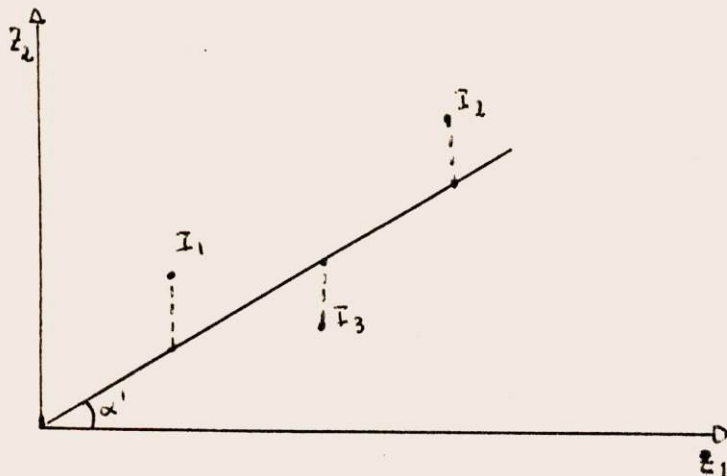
$$\text{Max: } \left[- (a_1 I_1)^2 - (a_2 I_2)^2 - (a_3 I_3)^2 \right] =$$

$$\text{Min: } \left[(a_1 I_1)^2 + (a_2 I_2)^2 + (a_3 I_3)^2 \right]$$

Luego lo que hacemos es también minimizar las distancias al cuadrado de los individuos al componente, medidas - éstas perpendicularmente al componente.

Esto nos hace pensar en la relación entre el sistema de componentes principales y la regresión múltiple. En ésta última, también minimizamos las distancias al cuadrado, la diferencia es que medimos las distancias de forma distinta; en concreto las medimos de forma paralela al eje de la variable explicada. Es decir, si la variable explicada está en ordenadas, mediremos las distancias en ordenadas.

Veámoslo gráficamente:



Luego en ambos casos hacemos mínimos cuadrados, pero midiendo las distancias de distinta forma.

En nuestro ejemplo, podemos hacer la regresión con Z_2 como variable explicada:

$$Y = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} ; \quad X = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}$$

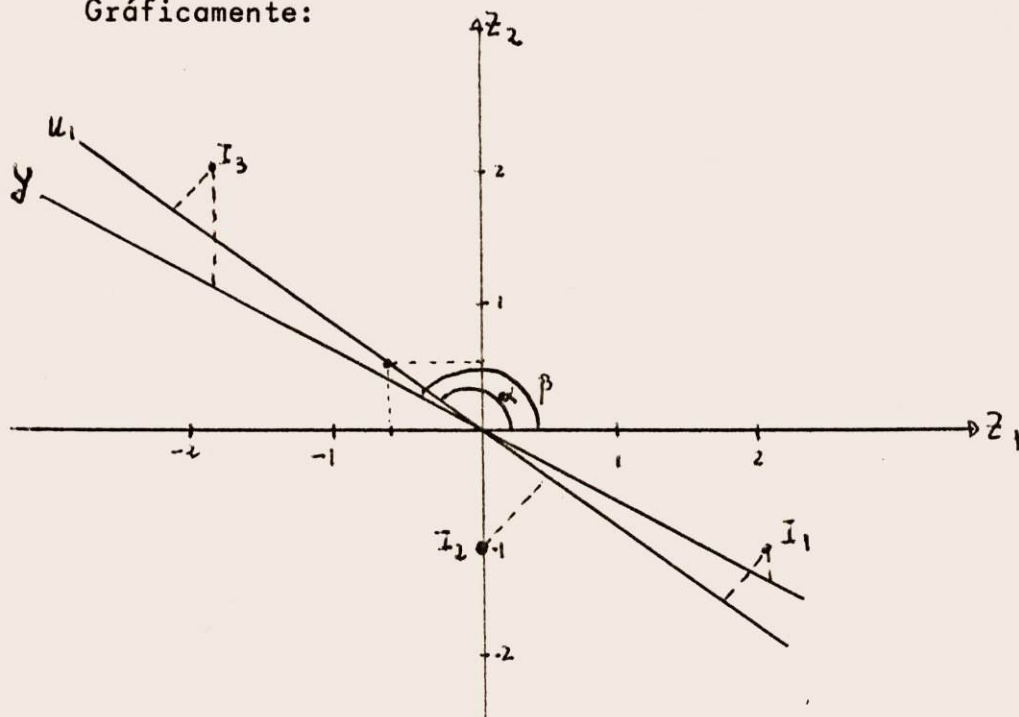
$$X (X' X)^{-1} X' Y = X a$$

$$a = (X' X)^{-1} X' Y$$

$$a = \frac{-6}{8} = -0'75 ; \quad \beta = \text{arc tg } -0'75 = 143'1301$$

$$y = -0'75 X$$

Gráficamente:



Vamos a continuar con nuestro problema, representando ahora los dos componentes en el espacio de las variables.

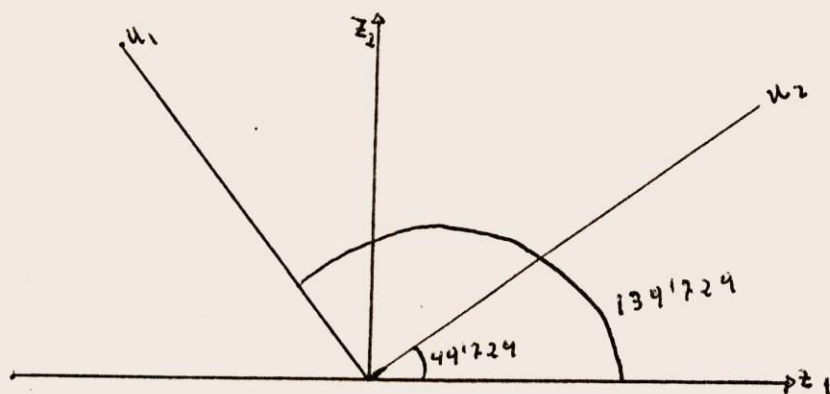
$$V_1 = \begin{pmatrix} -0'7630 \\ 0'6464 \end{pmatrix} ; \quad V_2 = \begin{pmatrix} 0'6464 \\ 0'7630 \end{pmatrix}$$

Vemos que son ortogonales y sabemos que

$$\alpha_1 = 139'729$$

$$\alpha_2 = \arccos 0'6464 = 49'729$$

Gráficamente:



El método que hemos utilizado para obtener los compo nentes principales mediante el máximo condicionado es idéntico al que proviene de diagonalizar $X' X$; pues:

$$X' X = \begin{pmatrix} 8 & -6 \\ -6 & 6 \end{pmatrix}$$

que es justamente la matriz que hemos diagonalizado con an terioridad.

Tenemos ahora el problema de que todavía no hemos ha llado los componentes en el espacio de los individuos. Para ello utilizamos la fórmula:

$$K = X \cdot W$$

$$\begin{pmatrix} 2 & -1 \\ 0 & -1 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} -0'7630 & 0'6464 \\ 0'6464 & 0'7630 \end{pmatrix} = \begin{pmatrix} -2'1724 & 0'5298 \\ -0'6464 & -0'7630 \\ 2'8188 & 0'2332 \end{pmatrix}$$

Estos dos componentes tienen respectivamente dimen sión λ_1 y λ_2 , como tenía que suceder.

Como decíamos en la parte teórica tiene que dar lo mismo diagonalizando $X X'$ (que tendrá las mismas λ que $X' X$).

$$X X' = \begin{pmatrix} 5 & 1 & -6 \\ 1 & 1 & -2 \\ -6 & -2 & 8 \end{pmatrix} \quad \begin{aligned} \lambda_1 &= 13'0828 \\ \lambda_2 &= 0'9172 \\ \lambda_3 &= 0 \end{aligned}$$

Primer vector:

$$\begin{aligned} -8'0828 V_1 + V_2 - 6 V_3 &= 0 \\ V_1 - 12'0828 V_2 - 2 V_3 &= 0 \\ -6 V_1 - 2 V_2 - 5'0828 V_3 &= 0 \end{aligned}$$

$$\text{Solución: (hay infinitas proporcionales)} = \begin{pmatrix} -2'1724 \\ -0'6464 \\ 2'8188 \end{pmatrix}$$

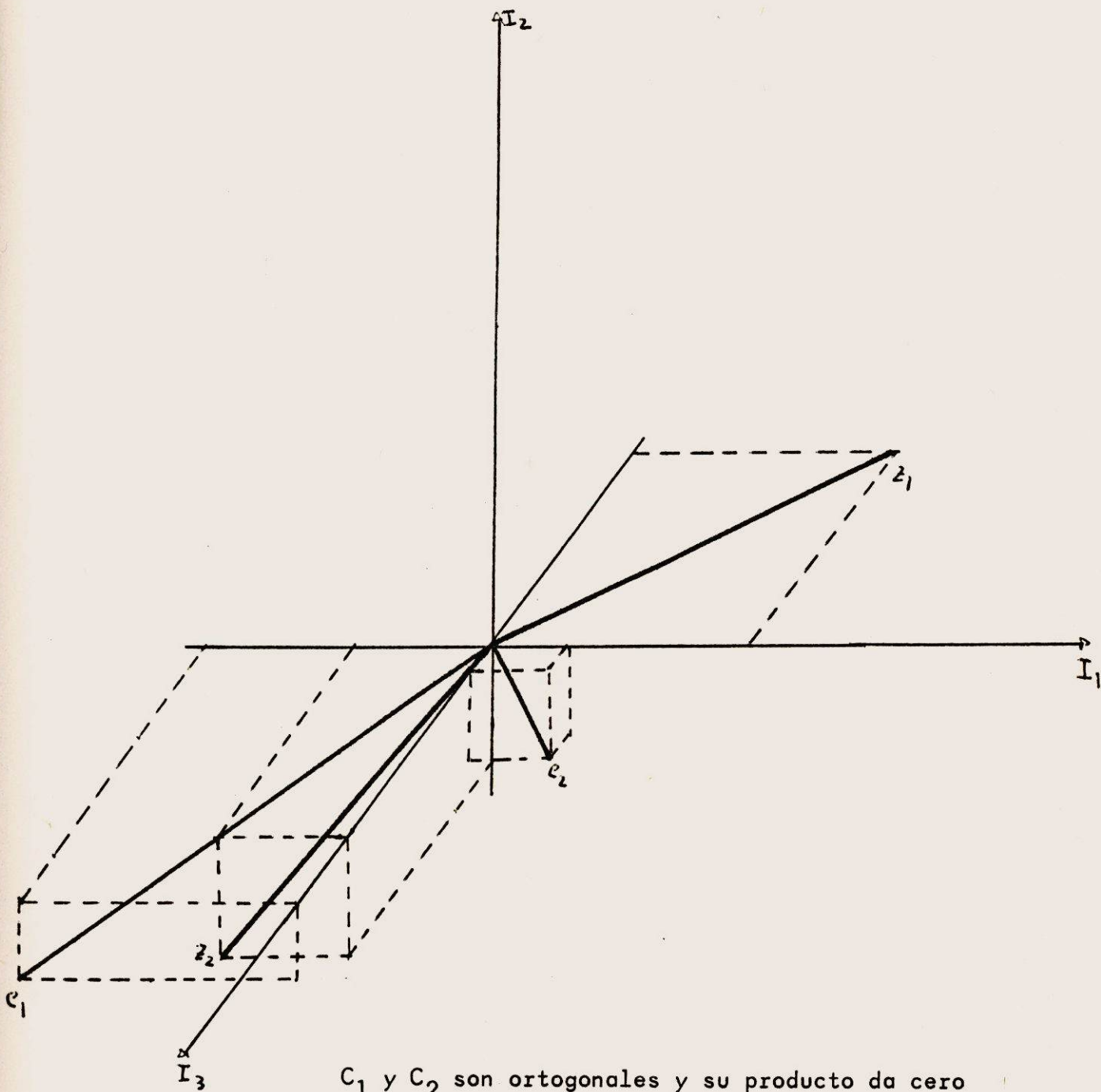
Segundo vector:

$$\begin{aligned} 4'0828 V_1 + V_2 - 6 V_3 &= 0 \\ V_1 + 0'0828 V_2 - 2 V_3 &= 0 \\ -6 V_1 - 2 V_2 + 7'0828 V_3 &= 0 \end{aligned}$$

$$\text{Solución: (hay infinitas proporcionales)} = \begin{pmatrix} 0'5298 \\ -0'7630 \\ 0'2332 \end{pmatrix}$$

Luego hemos comprobado que ambos caminos nos llevan al mismo resultado.

Los vectores de la matriz K, se definen en el espacio de individuos. Vamos a representarlos junto con las variables:



C_1 y C_2 son ortogonales y su producto da cero

$$C' C = 0$$

Con esto terminamos el ejemplo donde hemos plasmado gráfica y analíticamente el resumen del presente capítulo.

PROBLEMA DE LA MEDICION DE LOS COMPONENTES. (1)

La literatura existente, y aquí me voy a referir nuevamente al "Geolical Factor Analysis", por ser uno de los más completos tratados sobre el tema, se preocupa grandemente de lo que llaman "medición de los factores en los individuos", es decir, medición de la cantidad que tiene cada individuo de un factor (nosotros lo denominamos componente, pero ya hemos comentado que la nomenclatura es bastante confusa respecto a esta denominación).

A la vista de lo expuesto en este capítulo, el problema parece de fácil solución, bien mediante W^{-1} , o si W no es cuadrada (es decir, si no hemos agotado todos los componentes posibles) mediante la diagonalización de $X X'$. Esta diagonalización no ofrece excesivas dificultades pues aun siendo de dimensión $n \cdot n$, sus valores propios coinciden con los de $X' X$ de dimensión $p \cdot p$ (y por lo tanto mucho más fácil de diagonalizar), pero con la matriz $(X X' - \lambda I)$ hay que resolver el sistema de ecuaciones y será una matriz $n \cdot n$ lo que representa un problema si tenemos muchos individuos.

(1) Este problema está ampliamente estudiado en los libros de Harman (1.976), Jöreskog, Klován y Reymont (1.976), entre otros.

En realidad todo esto no es necesario pues aunque W^{-1} no tiene sentido si W no es cuadrada, lo que si es cierto es que si los U son de dimensión uno $W' W = I$, y W' es equivalente a la inversa de W .

Así obtenemos los valores de los individuos en los componentes, con varianza igual al valor propio.

Si estos valores, denominados puntuaciones, queremos obtenerlos normalizados, hay muchos sistemas como mínimos cuadrados o máxima verosimilitud, como ya veremos en la segunda parte.

Otro problema es que hemos dicho en la parte teórica que la matriz X debe estar normalizada y en el caso práctico que hemos expuesto, hemos trabajado con variables centradas. Esto lo hemos hecho para simplificar los cálculos. La discusión de este punto es más de aplicación que de estadística teórica, simplemente comentaremos aquí que la normalizar damos igual peso a todas las variables y evitamos los problemas de medida, cosa que en general es más interesante.

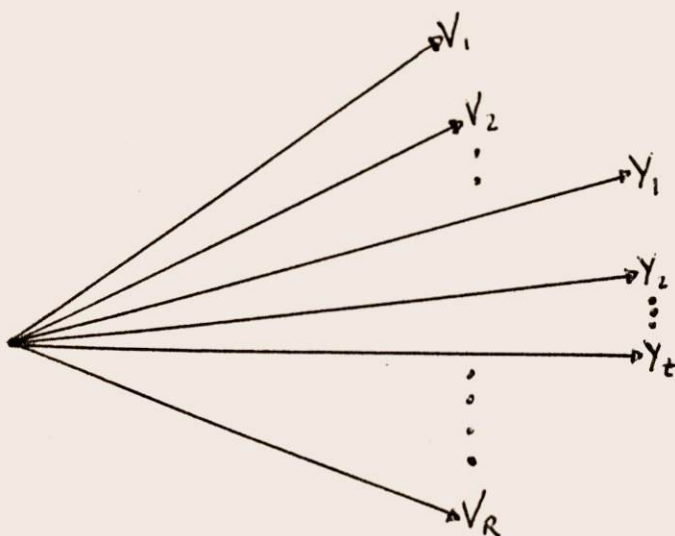
No tratamos aquí el problema de las rotaciones por poderse encontrar en cualquier tratado sobre el tema, de los cuales hemos mencionado ya bastantes.

V.- MODELO GENERAL Y APLICACIONES.

En este capítulo quinto lo que pretendemos es concluir el esquema que citábamos en el capítulo uno y ver como éste es aplicable a diversos problemas de análisis multivariante.

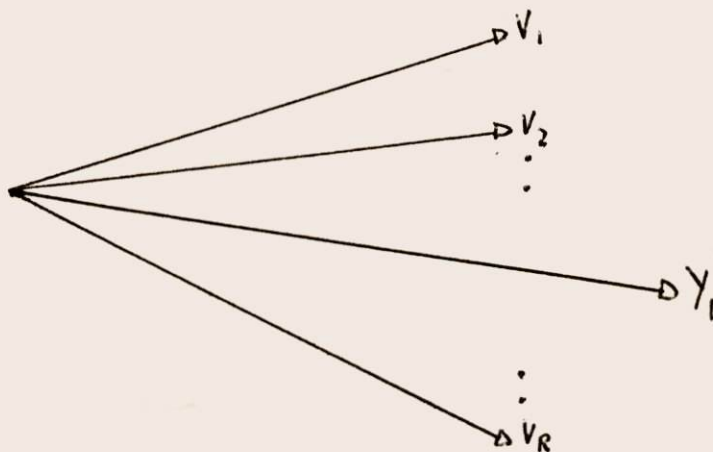
Partimos de que nuestra matriz de datos X como ya comentábamos en el capítulo cero, se puede representar en forma vectorial bien en el espacio de las variables o bien en el espacio de los individuos. Nosotros vamos a estudiar fundamentalmente el caso de que las variables son vectores (aunque ya hemos citado casos en capítulos anteriores donde la representación en el espacio de las variables nos era más cómoda, esto no altera las conclusiones del modelo pues lo fundamental es representar en forma vectorial, bien variables, bien individuos, bien otra cosa).

Dado pues un grupo de vectores obtenidos de la matriz X , este grupo lo representamos por uno o más vectores. Llamemos V_1, V_2, \dots, V_R a los vectores iniciales y Y_1, Y_2, \dots, Y_t a los vectores representantes; el esquema gráfico quedará:



Primer caso:

Cuando el grupo de vectores V es único:



Es lo que hemos llamado grupo de variables conocidas con variable desconocida (el representante Y_1).

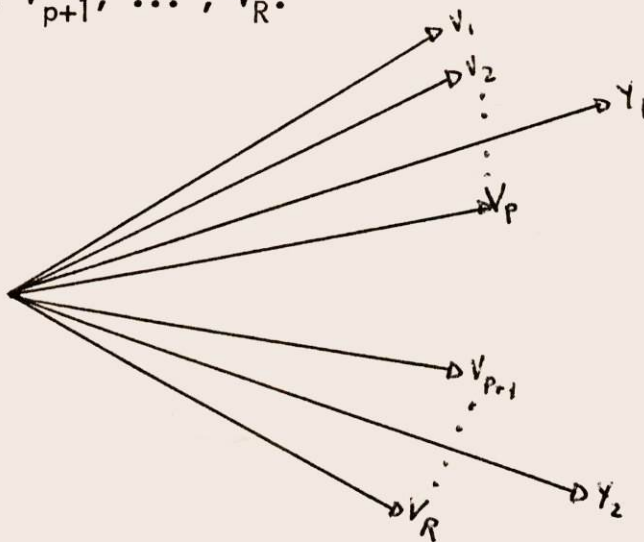
El proyector del subespacio α , lo que hace es projectar vectores de otros subespacios de forma ortogonal sobre el subespacio α . Luego si nosotros proyectamos sobre Y_1 los vectores V_1, \dots, V_R , con alguna condición, pues si no habría infinitos vectores Y_1 posibles, podremos definir el vector Y_1 .

Si la condición es que Y_1 tenga la máxima varianza, tendremos el primer componente principal. Si seguimos proyectando los residuos con la misma condición tendremos el resto de componentes.

Como casos similares a éste tenemos, como ya hemos comentado, el análisis factorial, Q-modo y R-modo, análisis de correspondencia\$, etc.

Segundo caso:

Quando hay dos grupos de vectores $V_1, V_2, \dots, V_p; V_{p+1}, \dots, V_R$.



Denominemos genéricamente V^1 al primer grupo de vectores y V^2 al segundo grupo de vectores.

Proyectando V^1 sobre V^2 y nuevamente sobre V^1 y haciendo una operación simétrica con V^2 obtendremos Y_1 e Y_2 tal que la correlación entre ambos representantes sea máxima.

Es el caso de la correlación canónica. Si el grupo uno es de un solo vector tendremos la correlación múltiple y si ambos son univectoriales tendremos la simple. De similar manera se pueden conseguir las regresiones.

Si hubiera más de dos grupos tendríamos la correlación canónica generalizada, de la cual ya hemos hablado - anteriormente.

APLICACIONES DEL MODELO.

Ya hemos visto con anterioridad que con este modelo podemos resolver los problemas de correlación, regresión y componentes principales y similares.

Pero nos quedan todavía bastantes problemas de análisis multivariante por resolver, entre ellos son los más interesantes:

- Análisis discriminante y problemas de clasificación. Cluster análisis.
- Variables cualitativas y problemas de codificación. Ta-
blas de contigencia.
- . Análisis de la varianza.

Lo que pretendemos ahora es ver que estas técnicas de análisis multivariante se pueden resolver mediante uno de los dos esquemas citados en el modelo.

En primer lugar el análisis discriminante veremos que se puede resolver satisfactoriamente por el esquema uno o por el esquema dos.

Por lo que se refiere a las variables cualitativas, ya vimos que el análisis de correspondencias, perteneciente al esquema uno, resuelve muchos de sus problemas, pero veremos además algunas aplicaciones del esquema dos a este caso, como es la codificación.

En último lugar tenemos el análisis de la varianza que resulta más difícil de encuadrar, pero no por ello - pierde generalidad el modelo como ya veremos posteriormente.

ANALISIS DISCRIMINANTE.

El análisis discriminante lo que pretende es encontrar una $Y = f(X)$, tal que nos sirva para discriminar entre los diversos grupos de individuos, que teníamos establecidos a priori.

$$Y = U_1 X_1 + U_2 X_2 + \dots + U_p X_p$$

Lo que tendremos que maximizar será: (1) $\text{Max} \frac{SS_b(Y)}{SS_w(Y)}$

Siendo:

$$SS_b(Y) = \sum_{i=1}^{i=K} n_i (\bar{Y}_i - \bar{Y})^2$$

$$SS_w(Y) = SS_1(Y) + SS_2(Y) + \dots + SS_k(Y)$$

$$SS_i(Y) = \sum_{j=1}^{j=n_i} (Y_{ji} - \bar{Y}_i)^2$$

Número de individuos = n

Número de variables = p

K grupos de individuos, cada uno de los cuales tiene n_i individuos, tal que:

$$\sum_{i=1}^{i=K} n_i = n$$

$\bar{X}_1 \dots \bar{X}_p \rightarrow$ serán los promedios totales de las p variables.

X_{ij} valor de la variable i en el individuo j.

\bar{X}_{ij} promedio de la variable i en el grupo j.

X_{1ji} valor de la variable 1, en el individuo j, del grupo i.

Y_{ji} valor de la variable Y en el individuo j, del grupo i.

- (1) El tratamiento matemático aquí seguido fue expuesto de forma detallada por Mckeon (1964) y Tatsuoka (1970) así como por Tatsuoka (1971). De forma más profunda pero menos clara está estudiado por Saporta (1976). Seguiremos fundamentalmente a Tatsuoka (1971) pgs. 157 y siguientes.

\bar{Y}_i promedio de la variable Y en el grupo i.
 \bar{Y} promedio global de la variable Y.

Definiendo las matrices T, W_i y B como:

$$T = \begin{pmatrix} \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2 & \sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) & \dots & \sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p) \\ \sum_{j=1}^n (x_{2j} - \bar{x}_2)(x_{1j} - \bar{x}_1) & \sum_{j=1}^n (x_{2j} - \bar{x}_2)^2 & \dots & \sum_{j=1}^n (x_{2j} - \bar{x}_2)(x_{pj} - \bar{x}_p) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{j=1}^n (x_{pj} - \bar{x}_p)(x_{1j} - \bar{x}_1) & \sum_{j=1}^n (x_{pj} - \bar{x}_p)(x_{2j} - \bar{x}_2) & \dots & \sum_{j=1}^n (x_{pj} - \bar{x}_p)^2 \end{pmatrix}$$

Luego la matriz T es la matriz de varianzas y covarianzas multiplicada por n, y sin tener en cuenta los diversos grupos que hay en el colectivo, es decir, tomando el colectivo como un todo.

$$W_i = \begin{pmatrix} \sum_{j=1}^{n_i} (x_{1ji} - \bar{x}_{1i})^2 & \cdot & \cdot & \cdot & \cdot & \sum_{j=1}^{n_i} (x_{1ji} - \bar{x}_{1i})(x_{pji} - \bar{x}_{pi}) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{j=1}^{n_i} (x_{pji} - \bar{x}_{pi})(x_{1ji} - \bar{x}_{1i}) & \cdot & \cdot & \cdot & \cdot & \sum_{j=1}^{n_i} (x_{pji} - \bar{x}_{pi})^2 \end{pmatrix}$$

La matriz W_i será lo mismo que la matriz T pero referido al grupo i. W será la matriz suma tal que:

$$W = \sum W_i$$

$$B = \begin{pmatrix} \sum_{i=1}^k n_i (\bar{X}_{1i} - \bar{X}_1)^2 & \sum_{i=1}^k n_i (\bar{X}_{1i} - \bar{X}_1) (\bar{X}_{2i} - \bar{X}_2) & \dots & \sum_{i=1}^k n_i (\bar{X}_{1i} - \bar{X}_1) (\bar{X}_{pi} - \bar{X}_p) \\ \sum_{i=1}^k n_i (\bar{X}_{2i} - \bar{X}_2) (\bar{X}_{1i} - \bar{X}_1) & \sum_{i=1}^k n_i (\bar{X}_{2i} - \bar{X}_2)^2 & \dots & \sum_{i=1}^k n_i (\bar{X}_{2i} - \bar{X}_2) (\bar{X}_{pi} - \bar{X}_p) \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^k n_i (\bar{X}_{pi} - \bar{X}_p) (\bar{X}_{1i} - \bar{X}_1) & \sum_{i=1}^k n_i (\bar{X}_{pi} - \bar{X}_p) (\bar{X}_{2i} - \bar{X}_2) & \dots & \sum_{i=1}^k n_i (\bar{X}_{pi} - \bar{X}_p)^2 \end{pmatrix}$$

La matriz B tiene por tanto en cuenta el número de individuos de cada grupo y la distancia entre el centroide de cada grupo y el del colectivo.

Tendremos pues:

$$T = \sum W_i + B = W + B$$

Luego lo que pretendemos maximizar es algo semejante

a:

Varianza entre colectivos

Suma de varianzas internas

Ya hemos dicho que:

$$S S_i (Y) = \sum_{j=1}^{n_i} (Y_{ji} - \bar{Y}_i)^2$$

Pero esto es igual que:

$S S_i (Y) = U' W_i U$ como se ve operando en la matriz (1).

$$\begin{aligned} \text{Luego } S S_w (Y) &= S S_1 (Y) + S S_2 (Y) + \dots + S S_k (Y) = \\ &= U' W_1 U + U' W_2 U + \dots + U' W_k U = \\ &= U' (W_1 + W_2 + \dots + W_k) U = U' W U \end{aligned}$$

Por otro lado tenemos:

$$B = (\bar{X} - \bar{\bar{X}})' (\bar{X} - \bar{\bar{X}}), \text{ siendo:}$$

(1)

$$\begin{aligned} U' W_i &= \left[\sum_{R=1}^{R=P} U_R \sum_{j=1}^{n_i} (X_{1ji} - \bar{X}_{1i})(X_{Rji} - \bar{X}_{Ri}) \dots \sum_{R=1}^{R=P} U_R \sum_{j=1}^{n_i} (X_{pji} - \bar{X}_{pi})(X_{pji} - \bar{X}_{Ri}) \right] \\ U' W_i U &= \sum_{\ell=1}^{\ell=P} U_\ell \sum_{R=1}^{R=P} U_R \sum_{j=1}^{n_i} (X_{1ji} - \bar{X}_{1i})(X_{Rji} - \bar{X}_{Ri}) = \\ &= \sum_{\ell=1}^{\ell=P} \sum_{R=1}^{R=P} \sum_{j=1}^{n_i} U_\ell \cdot U_R (X_{1ji} - \bar{X}_{1i})(X_{Rji} - \bar{X}_{Ri}) = \\ &= \sum_{j=1}^{n_i} \sum_{\ell=1}^{\ell=P} (U_\ell X_{1ji} - U_\ell \bar{X}_{1i}) \sum_{R=1}^{R=P} (U_R X_{Rji} - U_R \bar{X}_{Ri}) = \\ &= \sum_{j=1}^{n_i} \left(\sum_{\ell=1}^{\ell=P} U_\ell X_{1ji} - \sum_{\ell=1}^{\ell=P} U_\ell \bar{X}_{1i} \right) \left(\sum_{R=1}^{R=P} U_R X_{Rji} - \sum_{R=1}^{R=P} U_R \bar{X}_{Ri} \right) = \\ &= \sum_{j=1}^{n_i} (Y_{ji} - \bar{Y}_i) (Y_{ji} - \bar{Y}_i) = \sum_{j=1}^{n_i} (Y_{ji} - \bar{Y}_i)^2 \end{aligned}$$

como queríamos demostrar.

$$\bar{X} = \begin{pmatrix} \bar{X}_{11} & \bar{X}_{21} & \cdot & \cdot & \cdot & \bar{X}_{p1} & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot n_1 \text{ veces} \\ \cdot & \cdot & & & & \cdot & \cdot \\ \bar{X}_{12} & \bar{X}_{22} & \cdot & \cdot & \cdot & \bar{X}_{p2} & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot n_2 \text{ veces} \\ \cdot & \cdot & & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{X}_{1k} & \bar{X}_{2k} & \cdot & \cdot & \cdot & \bar{X}_{pk} & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot n_k \text{ veces} \\ \cdot & \cdot & & & & \cdot & \cdot \end{pmatrix}$$

$$\begin{aligned} U'BU &= U' (\bar{X} - \bar{\bar{X}})' (\bar{X} - \bar{\bar{X}}) U = (\bar{X} U - \bar{\bar{X}} U)' (\bar{X} U - \bar{\bar{X}} U) = \\ &= \sum_{i=1}^{i=k} n_i (\bar{Y}_i - \bar{Y})^2 \quad (\text{como se puede ver operando}). \end{aligned}$$

Luego $S S_b (Y) = U' B U$

Con lo que, el cociente a maximizar nos queda:

$$\frac{S S_b (Y)}{S S_w (Y)} = \frac{U' B U}{U' W U} = \lambda$$

Este criterio, aplicado a solo dos grupos, fué utilizado en 1.936 por Fisher.

Denominaremos criterio discriminante al valor de λ . Maximizaremos derivando e igualando a cero:

$$\frac{d\lambda}{dU} = \frac{2[(U' W U)(B U) - (U' B U)(W U)]}{(U' W U)^2} = 0$$

$$\frac{2[B U - \lambda W U]}{U' W U} = 0 \Rightarrow B U = \lambda W U$$

$$(B - \lambda W) U = 0$$

Luego tendremos que: $W^{-1} B U = \lambda U = \lambda I U$

$$(W^{-1} B - \lambda I) U = 0$$

El problema es que W puede ser singular en el caso de que $K > (n - P)$. Pero:

$$T = W + B$$

$$\text{Entonces: } U' T U = U' W U + U' B U$$

$$\text{y en consecuencia: } \text{Max} \frac{U' B U}{U' W U}$$

equivalente a:

$$\text{Max} \frac{U' B U}{U' T U}$$

Luego podemos utilizar la matriz T .

Antes decíamos que:

$$B U = \lambda W U$$

Sumando $(\lambda B U)$ a los dos términos:

$$(1 + \lambda) B U = \lambda (B + W) U$$

$$(1 + \lambda) B U = \lambda T U$$

Y por lo tanto, operando:

$$B U = \frac{\lambda}{1 + \lambda} T U$$

y llamando μ a la siguiente expresión:

$$\mu = \frac{\lambda}{1 + \lambda}$$

$$B U = \mu T U$$

$$T^{-1} B U = \mu U = \mu I U$$

$$(T^{-1} B - \mu I) U = 0$$

Luego si usamos T en vez de W , $\mu = \frac{\lambda}{1 + \lambda}$; donde λ es el criterio discriminante.

Veamos a continuación el caso de dos grupos que es el que estudió Fisher.

La matriz B , con solo dos grupos queda:

$$b_{ij} = n_1 (\bar{X}_{i1} - \bar{\bar{X}}_i) (\bar{X}_{j1} - \bar{\bar{X}}_j) + n_2 (\bar{X}_{i2} - \bar{\bar{X}}_i) (\bar{X}_{j2} - \bar{\bar{X}}_j)$$

$$\bar{\bar{X}}_i = \frac{n_1 \bar{X}_{i1} + n_2 \bar{X}_{i2}}{n_1 + n_2}$$

operando: $b_{ij} = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_{i1} - \bar{X}_{i2}) (\bar{X}_{j1} - \bar{X}_{j2})$

llamamos d' :

$$d' = (\bar{X}_{11} - \bar{X}_{12}, \bar{X}_{21} - \bar{X}_{22}, \dots, \bar{X}_{p1} - \bar{X}_{p2})$$

$$B = \frac{n_1 n_2}{n_1 + n_2} d d'$$

Teníamos que:

$$(T^{-1} B - \mu I) U = 0$$

$$T^{-1} B U = \mu U$$

Sustituyendo B:

$$C T^{-1} (d d') U = \mu U = C (T^{-1} d) (d' U)$$

Siendo $C = \frac{n_1 n_2}{n_1 + n_2}$

Pero $d' U$ es un escalar luego:

$$U = \left(\frac{C (d' U)}{\mu} \right) T^{-1} d$$

$\frac{C (d' U)}{\mu}$ es un valor desconocido, dependiendo del cual hay

infinitas soluciones de U , todas proporcionales, y por lo tanto que solo afectan a la unidad de medida, por lo que lo podemos despreciar.

También podemos usarlo para forzar que U sea un vector de dimensión uno.

Si lo despreciamos tendremos que:

$$U = T^{-1} d$$

Recordando que $B = C d d'$ y que $T^{-1} B U = \mu U$

$$\mu U = T^{-1} B T^{-1} d = T^{-1} C d d' T^{-1} d$$

pero como C es un escalar, podemos cambiarlo de situación y tendremos:

$$\mu U = T^{-1} d C d' T^{-1} d = U \mu$$

$$\text{Pero } U = T^{-1} d$$

$$\text{Luego } \mu = C d' T^{-1} d$$

La distancia de Mahalanobis entre los centros de gravedad de los dos grupos es $d' T^{-1} d$ como se puede ver:(1)

(1) Mahalanobis (1.936) pags. 49 y siguientes.

$$d' T^{-1} d = (\bar{X}_1 - \bar{X}_2)' T^{-1} (\bar{X}_1 - \bar{X}_2) = D^2 \text{ de Mahalanobis.}$$

En consecuencia μ es igual a C por la distancia de Mahalanobis.

Al hacer máximo λ (criterio discriminante), hacemos máximo μ y en consecuencia lo que maximizamos es la distancia de Mahalanobis entre los dos centros de gravedad.

La U obtenida mediante $T^{-1} d$ es la de Fisher.

Volviendo nuevamente al caso general y a la fórmula $\mu U = T^{-1} B U$, podemos hacer las siguientes consideraciones:

En primer lugar que $T^{-1} B$ es una especie de matriz de distancias y codistancias de Mahalanobis. Al hacer $T^{-1} B U = \mu U$, buscamos el eje donde mejor se proyectan estas distancias.

Pero lo que es claro es que al diagonalizar $T^{-1} B$ lo que hacemos es buscar sus vectores propios, es decir, hacemos un análisis de componentes principales de una nube de K centros de gravedad con la métrica de Mahalanobis.

Luego en definitiva el análisis discriminante es un caso particular del análisis de componentes principales.

Pero por otro lado, también se puede enfocar el aná

lisis discriminante como un caso particular de correlación canónica como vamos a ver a continuación:

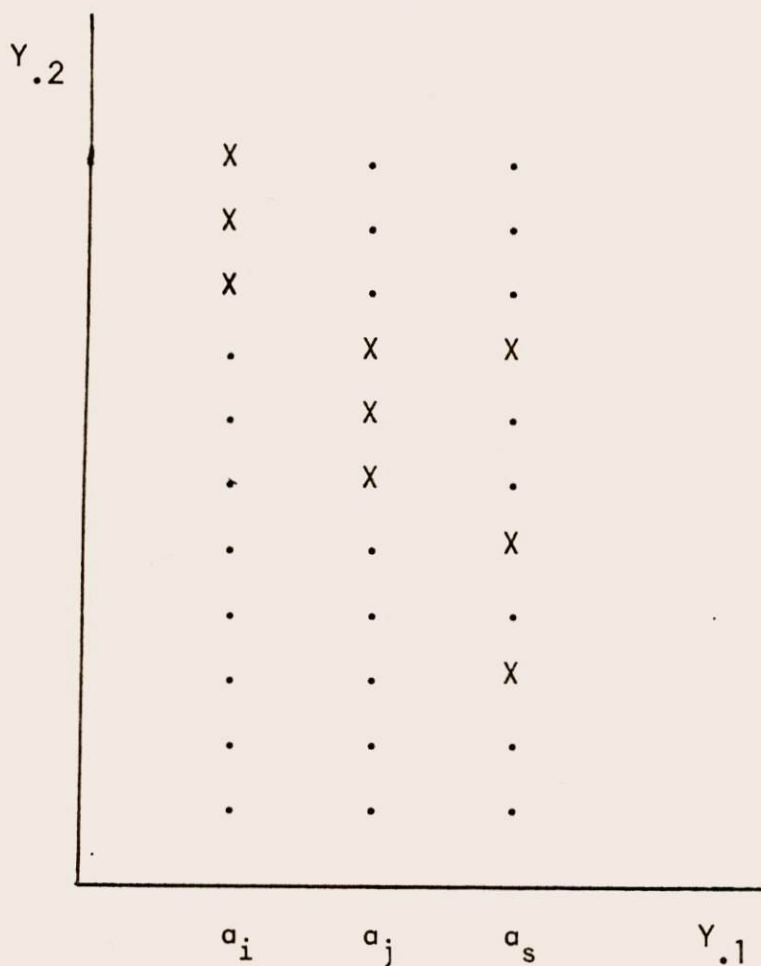
Sean K grupos y P variables con un total de n individuos. Podemos considerar a las P variables como un grupo canónico y crear otro grupo canónico que contenga la idea de la diferenciación en grupos.

Maximizando la correlación canónica entre estos grupos, conseguiremos una forma de discriminación.

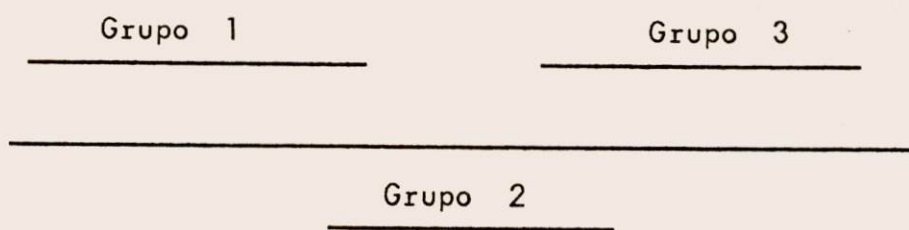
La creación del segundo grupo canónico se basa en la siguiente idea: El decir que un individuo pertenece a un grupo o a otro, es una variable cualitativa, que la podemos cuantitativizar, creando K variables, cada una de las cuales toma valor uno en su grupo correspondiente y valor cero en los demás. Con lo que las matrices de los dos grupos canónicos quedarán:

	Matriz A	Matriz X
$\left. \begin{array}{l} n_1 \\ \vdots \\ n_K \end{array} \right\}$	$\begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 1 \end{pmatrix}$	$\begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n_1,1} & X_{n_1,2} & \cdot & \cdot & X_{n_1,p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n_1+n_2,1} & \cdot & \cdot & \cdot & X_{n_1+n_2,p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n_1+n_2+\dots+n_K,1} & \cdot & \cdot & \cdot & X_{n_1+n_2+\dots+n_K,p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n,1} & \cdot & \cdot & \cdot & X_{n,p} \end{pmatrix}$

Nos quedará un gráfico del estilo de:



Sobre el vector $Y_{.2}$ quedará:



Así obtenemos el primer eje de discriminación, luego si queremos podemos obtener los demás con las siguientes X.

El sistema de correlación canónica coincide con el de componentes principales:

$(X' X)^{-1} = T^{-1}$ Si las variables están normalizadas.

$$X' A = \begin{pmatrix} \sum X_{1j1} & \cdot & \cdot & \cdot & \cdot & \sum X_{1jK} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \sum X_{pj1} & \cdot & \cdot & \cdot & \cdot & \sum X_{pjK} \end{pmatrix}$$

Siendo X_{ijK} el valor de la variable i , en el individuo j , del grupo K .

$$(A' A)^{-1} = \begin{pmatrix} 1/n_1 & & & & & \\ & 1/n_2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & 1/n_K \end{pmatrix}$$

$$A' X = (X' A)'$$

Luego tenemos que $X' A (A' A)^{-1} A' X = B$

Sustituyendo en:

$$(X' X)^{-1} X' A (A' A)^{-1} A' X U = \mu U$$

Tenemos que: $T^{-1} B U = \mu U$ c.q.d.

El análisis discriminante también se puede considerar como una generalización de la regresión múltiple cuando la variable a explicar es cualitativa.

En concreto, en el caso de dos grupos, (1), si hacemos la regresión múltiple donde la variable explicada sea:

$$Y = \begin{pmatrix} 1/n_1 \\ 1/n_1 \\ \cdot \\ 1/n_1 \\ -1/n_2 \\ -1/n_2 \\ \cdot \\ \cdot \\ -1/n_2 \end{pmatrix}$$

Con tantos $1/n$, como elementos tiene el grupo uno y viceversa:

$$U = (X' X)^{-1} X' Y$$

Estando las variables normalizadas:

$$(X' X)^{-1} = T^{-1}$$

$$X' Y = \begin{pmatrix} 1/n_1 \sum X_{1j1} & - & 1/n_2 \sum X_{1j2} \\ 1/n_1 \sum X_{2j1} & - & 1/n_2 \sum X_{2j2} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 1/n_1 \sum X_{pj1} & - & 1/n_2 \sum X_{pj2} \end{pmatrix} = \begin{pmatrix} \bar{X}_{11} & - & \bar{X}_{12} \\ \bar{X}_{21} & - & \bar{X}_{22} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \bar{X}_{p1} & - & \bar{X}_{p2} \end{pmatrix}$$

(1) Estudio similar al propuesto por Fisher (1.936) pags. 184 - 185.

$$\text{Luego } X' Y = d$$

$U = T^{-1} d \Rightarrow$ La solución de Fisher cuando despreciamos el factor de proporcionalidad. $\frac{C d' U}{\mu}$.

$$\text{Pero esto es hacer } \mu = C d' U$$

En correlación múltiple:

$$R = \frac{U' X' Y}{Y' Y}$$

$$\begin{aligned} \text{En nuestro caso } Y'Y &= (1/n_1)^2 n_1 + (-1/n_2)^2 n_2 = 1/n_1 + 1/n_2 = \\ \frac{n_1 + n_2}{n_1 n_2} &= \frac{1}{C} \end{aligned}$$

$$X' Y = d$$

Luego: $R = C U' d = C d' U$; que coincide con la μ de Fisher.

Luego el método de Fisher coincide con una regresión múltiple.

Así tenemos que en conclusión el análisis discriminante es solamente una aplicación de un caso del esquema uno - del modelo general (componentes principales), o del esquema dos (canónica) con lo que queda perfectamente dentro de nues

tro modelo, como ya apuntábamos con anterioridad. (1)

Veamos un ejemplo donde manejemos todos estos conceptos: (utilizaremos datos centrados).

$$X = \begin{pmatrix} 1 & 0 \\ -2 & 1 \\ 1 & 1 \\ 0 & -2 \end{pmatrix}$$

de forma que los individuos uno y dos forman el primer grupo y tres y cuatro el segundo.

$$\bar{X}_1 = 0$$

$$\bar{X}_2 = 0$$

$$\bar{X}_{11} = -0.5 \quad \bar{X}_{12} = 0.5$$

$$\bar{X}_{21} = 0.5 \quad \bar{X}_{22} = -0.5$$

$$T = \begin{pmatrix} 6 & -1 \\ -1 & 6 \end{pmatrix}; \quad B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

- (1) Creo interesante recordar aquí como ha evolucionado hasta la actualidad el análisis de Fisher, es decir hacer un poco de historia del análisis discriminante. Sigo para ello a Dornvlo (1.977), pgs. 1-2,

Fisher empezó en 1936 con el análisis de función lineal discriminante ya citado. En 1939 Welch minimizó la probabilidad total de mala clasificación.

Recientemente tenemos los estudios Enis y Geisser (1970) que estudian el problema desde un punto de vista bayesiano.

Nosotros lo hemos estudiado desde componentes principales y desde la canónica.

$$W_1 = \begin{pmatrix} 4'5 & -1'5 \\ -1'5 & 0'5 \end{pmatrix}; \quad W_2 = \begin{pmatrix} 0'50 & 1'5 \\ 1'5 & 4'5 \end{pmatrix}$$

$$W = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}; \quad T = W + B$$

Como W no es singular, podemos aplicar:

$$W^{-1} B U = \lambda U$$

$$W^{-1} B = \begin{pmatrix} 1/5 & -1/5 \\ -1/5 & 1/5 \end{pmatrix}$$

$$\begin{vmatrix} 1/5 - \lambda & -1/5 \\ -1/5 & 1/5 - \lambda \end{vmatrix} = 0$$

$$\lambda_1 = 2/5$$

$$\lambda_2 = 0$$

Tomamos $\lambda = 2/5$

$$\begin{pmatrix} -1/5 & -1/5 \\ -1/5 & -1/5 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = 0$$

Luego $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ será proporcional a $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Diagonalizando $T^{-1} B$:

$$T^{-1} B = \begin{pmatrix} 6/35 & 1/35 \\ 1/35 & 6/35 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1/7 & -1/7 \\ -1/7 & 1/7 \end{pmatrix}$$

$$\begin{pmatrix} 1/7 - \mu & -1/7 \\ -1/7 & 1/7 - \mu \end{pmatrix} = 0$$

$$\mu_1 = 2/7$$

$$\mu_2 = 0$$

podemos comprobar que $\mu_i = \frac{\lambda_i}{1 + \lambda_i}$

tomando $\mu = 2/7$ $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ será proporcional a $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Por el método de Fisher, pues son solo dos grupos:

$$U = T^{-1} d$$

$$U = \begin{pmatrix} 6/35 & 1/35 \\ 1/35 & 6/35 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/7 \\ 1/7 \end{pmatrix}$$

que es proporcional a $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$$\mu = C d' T^{-1} d = 1 \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} 6/35 & 1/35 \\ 1/35 & 6/35 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \frac{2}{7}$$

que como $C = 1$; coincide con la D^2 de Mahalanobis.

Veamos por canónica:

$$(X' X)^{-1} X' A (A' A)^{-1} A' X U = \mu U$$

siendo $A =$

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Como ya demostramos, coincide con:

$$T^{-1} B U = \mu U$$

U proporcional a

$$\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Veamos el otro vector canónico:

$$X (X' X)^{-1} X' = \frac{1}{35} \begin{pmatrix} 6 & -11 & 7 & -2 \\ -11 & 26 & -7 & -8 \\ 7 & -7 & 14 & -14 \\ -2 & -8 & -14 & 24 \end{pmatrix}$$

$$(A' A)^{-1} A' = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

$$(A' A)^{-1} A' X (X' X)^{-1} X' = 1/70 \begin{pmatrix} -5 & 15 & 0 & -10 \\ 5 & -15 & 0 & 10 \end{pmatrix}$$

$$(A' A)^{-1} A' X (X' X)^{-1} X' A = \begin{pmatrix} 1/7 & -1/7 \\ -1/7 & 1/7 \end{pmatrix}$$

$$\begin{vmatrix} 1/7 - \mu & -1/7 \\ -1/7 & 1/7 - \mu \end{vmatrix} = 0$$

$$\mu_1 = 2/7$$

$$\mu_2 = 0$$

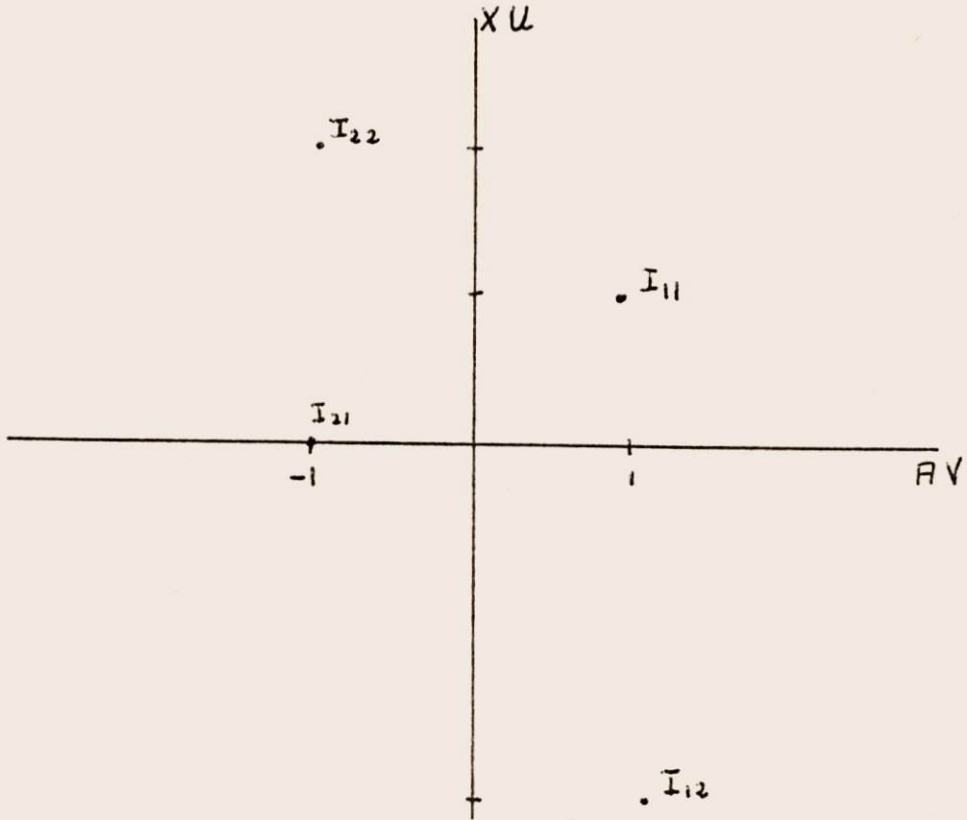
$$(A' A)^{-1} A' X (X' X)^{-1} X' A V = \mu V$$

de donde V será proporcional a $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Por lo tanto:

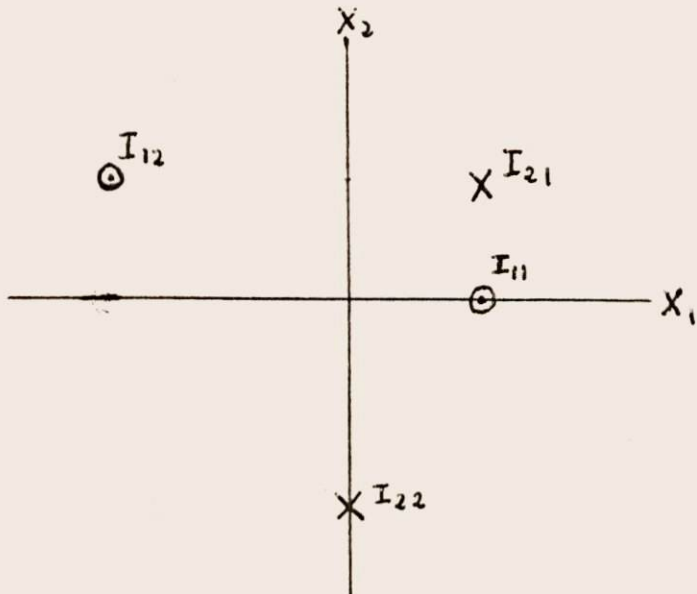
$$X U = \begin{pmatrix} 1 \\ -3 \\ 0 \\ 2 \end{pmatrix}$$

$$A V = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$



Luego vemos que quedan muy mal discriminadas y se mezclan en el vector $X U$, cosa natural dado el valor de μ .

La representación en el espacio original da:



Vemos que son imposibles de discriminar bien.

Veámoslo ahora por regresión múltiple:

$$Y = \begin{pmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{pmatrix}$$

$$U = (X' X)^{-1} X' Y = \begin{pmatrix} -1/7 \\ 1/7 \end{pmatrix} \quad (\text{igual que Fisher})$$

$$R = \frac{U' X' Y}{Y' Y} = \frac{2/7}{1} = 2/7 = \mu$$

Con lo cual hemos visto un sencillo ejemplo, para ver como se pueden emplear las diferentes técnicas de análisis - discriminante.

En la práctica estos sistemas se desdoblan aún más, y aunque no es el objeto del presente trabajo, quiero comentar brevemente que lo más práctico es utilizar el método de Fisher, cuando tenemos dos grupos, e incluso con más grupos para discriminar uno con los demás y así sucesivamente.

El método general, diagonalizando la matriz $T^{-1} B$, tiene interés en el caso de muchos grupos, sobre todo el primer componente, pero los demás son de difícil interpretación.

Por lo que se refiere a los problemas de clasificación, quedan bastante apartados del objetivo de nuestro estudio, de

do que no se refieren al análisis de datos, por otro lado la solución al problema es sencillo, pues se trata de buscar un punto en el eje de discriminante para decidir si un - nuevo punto pertenece, o debe pertenecer, a un grupo o a otro. Este problema se resuelve estadísticamente tratando de minimizar los riesgos.

Muy interesantes son los métodos de programación lineal (1) que se apartan bastante del esquema de Fisher y que al utilizar técnicas bastante distintas, simplemente - las citamos en nuestro estudio.

Estos métodos de programación lineal tienen la ventaja de que no es necesaria la hipótesis de normalidad, ni incluso el que se siga una distribución conocida. También tiene estas cualidades el sistema de regresión logística por - máxima verosimilitud. (2)

CLUSTER ANALISIS.

Esta técnica de agrupamiento lo que pretende es formar grupos homogéneos de individuos. Hay infinidad de métodos para formar estos grupos, basándose normalmente en la

- (1) La ventaja de este método es que no necesita la hipótesis de normalidad. Ha sido estudiado ampliamente por - Chacón (1.975).
- (2) La regresión logística ha sido estudiada en multitud de artículos siendo uno de los más recientes el de S. James Press y Sandra Wilson (1.978).

distancia Euclidiana o en la de Mahalanobis. Los más conocidos son los métodos de encadenamiento, pero hay otra infinidad de métodos, según el criterio que se escoja para ir formando los grupos. No es nuestra intención el explicar, ni aún siquiera enumerar estos métodos, sino simplemente ver - que esto también se puede resolver mediante nuestro modelo.

Dadas las matrices T , W y B tal como definimos anteriormente, T será una matriz conocida, pero W y B serán matrices desconocidas. Lo que se propone es un método denominado "método matricial de cluster análisis", que formando - unos grupos arbitrarios resuelva:

$$T^{-1} B U = \mu U$$

O bien si W es invertible:

$$W^{-1} B U = \lambda U$$

Para cada agrupación se obtienen las λ (o las μ en su caso), y lo que intentaremos será hallar el máximo de la suma de λ , o de su producto, o cualquier otro criterio.

En definitiva nos encontramos en el caso uno de nuestro modelo, es decir en componentes principales, pues lo - que pretendemos es maximizar la suma, o el producto, de los valores propios de la matriz $W^{-1} B$.

Desde luego que este método es muy similar al análisis discriminante, que partiendo de unos grupos dados, los intenta discriminar lo mejor posible, aquí lo que pretendemos es formar los grupos de forma que sean fáciles de discriminar.

Este método, como se puede apreciar es de penosa ejecución, por ello se establecieron otros sistemas abreviados para cuyo estudio se pueden consultar los tratados de cluster análisis (1). Especial mención merece el método de la distancia de Mahalanobis, por su interés tanto teórico como práctico.

Finalmente otro tipo de planteamiento es el denominado Q-modo, que lo que pretende es, así como el R-modo forma grupos de variables (componentes o factores), crear componentes o factores de individuos, tratados estos como vectores (2). Que en definitiva, demuestra que el cluster análisis se puede tratar ampliamente como una técnica derivada del esquema de nuestro modelo.

Por otro lado como el análisis discriminante se puede resolver por canónica, podríamos encuadrar el problema perfectamente en el caso 2 del modelo.

- (1) El cluster y el reconocimiento de patrones van en multitud de casos estudiados conjuntamente. Concretamente en cluster análisis tenemos el Anderbery (1.973). Especialmente claro es el cluster análisis de Narvaiza (1.979).

El libro clásico de reconocimiento de patrones es el de Duda y Hart (1.973).

- (2) El Q-modo está tratado ampliamente en el ya citado Geological Factor Analysis.

VARIABLES CUALITATIVAS:

Ya comentamos que el análisis de correspondencias se ajustaba al caso uno de nuestro modelo.

Veremos ahora algunas aplicaciones del caso segundo a las variables cualitativas. (1)

Supongamos n individuos en los cuales analizamos dos cualidades, como por ejemplo color del pelo y color de los ojos. Supongamos P posibles colores de pelo y q posibles colores de ojos. Podemos crear P variables (una por cada color de pelo) tal que el individuo que tiene el color uno, tenga un uno en la variable uno y cero en las demás y así sucesivamente. También hacemos lo mismo con las q variables representativas del color de los ojos. Con lo cual hemos cuantitizado de alguna manera las variables cualitativas y hemos llegado a una matriz del tipo:

(1) Puede verse en muchos trabajos como en el de Saporta (1.976), página 39.

$$X = \begin{matrix} 1 & 12 & \cdot & \cdot & \cdot & \cdot & P_{12} & \cdot & \cdot & \cdot & \cdot & q \\ 2 & 01 & \cdot & \cdot & \cdot & \cdot & 010 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & 00 & \cdot & .1 & \cdot & \cdot & 000 & \cdot & .1 & \cdot & \cdot & 0 \\ \cdot & 10 & \cdot & \cdot & \cdot & \cdot & 000 & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n & 00 & \cdot & .1 & \cdot & \cdot & 001 & \cdot & \cdot & \cdot & \cdot & 0 \end{matrix}$$

Podemos partir la matriz y llamarle:

$$X = [X_1 \mid X_2]$$

Y operando de la forma:

$$X_1' X_2 = \begin{pmatrix} n_{11} & n_{12} & \cdot & \cdot & \cdot & n_{1q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n_{p1} & n_{p2} & \cdot & \cdot & \cdot & n_{pq} \end{pmatrix}$$

Siendo el dato n_{ij} el número de individuos que tienen la primera cualidad i y la segunda j .

Denominando P_{ij} al cociente n_{ij}/n (pasando a una especie de frecuencias relativas, tendremos:

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdot & \cdot & \cdot & P_{1q} \\ P_{21} & P_{22} & \cdot & \cdot & \cdot & P_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{p1} & P_{p2} & \cdot & \cdot & \cdot & P_{pq} \end{pmatrix}$$

Y llamando D_1 y D_2 a las siguientes matrices:

$$D_1 = \begin{pmatrix} P_{1.} & & & & \\ & P_{2.} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & P_{p.} \end{pmatrix}$$

$$D_2 = \begin{pmatrix} P_{.1} & & & & \\ & P_{.2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & P_{.q} \end{pmatrix}$$

Donde $P_{i.} = \sum_j P_{ij}$ y $P_{.j} = \sum_i P_{ij}$

Dejemos de momento estas matrices D y vamos a intentar la correlación canónica entre X_1 y X_2 .

El objetivo de esta correlación canónica entre variables cualitativas cuantitativizadas, es ver la relación entre el color del pelo (variable uno) y el color de los ojos (variable dos).

Para mantener una uniformidad con lo dicho anteriormente corregimos con la matriz diagonal D_p , tal que:

$$D_p = \begin{pmatrix} 1/n & & & & \\ & 1/n & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1/n \end{pmatrix}$$

Y así tendremos la matriz P, obtenida anteriormente:

$$P = X_1' D_p X_2 = X_1' \begin{pmatrix} 1/n & & & & \\ & 1/n & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1/n \end{pmatrix} X_2$$

$$P = \begin{pmatrix} P_{11} & \cdot & \cdot & \cdot & P_{1q} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ P_{p1} & \cdot & \cdot & \cdot & P_{pq} \end{pmatrix}$$

Apliquemos en consecuencia la fórmula de la canónica corregida con D_p :

$$\text{Proyector 1} = X_1 (X_1' X_1)^{-1} X_1'; \text{ corrigiendo } X_1 (X_1' D_p X_1)^{-1} X_1' D_p$$

$$\text{Proyector 2} = X_2 (X_2' X_2)^{-1} X_2'; \text{ corrigiendo } X_2 (X_2' D_p X_2)^{-1} X_2' D_p$$

Y aplicando las fórmulas:

$$(X_1' D_p X_1)^{-1} X_1' D_p X_2 (X_2' D_p X_2)^{-1} X_2' D_p X_1 a = \lambda a$$

$$(X_2' D_p X_2)^{-1} X_2' D_p X_1 (X_1' D_p X_1)^{-1} X_1' D_p X_2 b = \lambda b$$

Pero resulta que:

$$\begin{array}{l}
 X_1' D_p X_1 = \left[\begin{array}{cccc}
 n_{1.}/n & & & \\
 & n_{2.}/n & & \\
 & & \cdot & \\
 & & & \cdot \\
 & & & & n_{p.}/n
 \end{array} \right] = D_1 \\
 \\
 X_2' D_p X_2 = \left[\begin{array}{cccc}
 n_{.1}/n & & & \\
 & n_{.2}/n & & \\
 & & \cdot & \\
 & & & \cdot \\
 & & & & n_{.q}/n
 \end{array} \right] = D_2
 \end{array}$$

Y sustituyendo en las fórmulas anteriores:

$$D_1^{-1} P D_2^{-1} P' a = \lambda a$$

$$D_2^{-1} P' D_1^{-1} P b = \lambda b$$

De aquí podemos obtener:

$$X_1 a, X_2 b, \text{ y } \lambda = R^2$$

con lo que tenemos el coeficiente de correlación entre las dos variables.

Este planteamiento, bastante relacionado con el análisis de correspondencias, tiene bastantes utilidades a primera vista, y sin lugar a dudas se le irán encontrando más, conforme vaya avanzando su estudio, pues es bastante reciente.

En primer lugar nos da un coeficiente de correlación entre variables cualitativas, no ordinales, de fácil obtención.

En segundo lugar nos permite la clasificación de una variable cualitativa con relación a otra.

En último lugar nos permite la codificación de variables cualitativas.

En general este análisis canónico se puede realizar entre variable cualitativa y cualitativa o cuantitativa, con lo que sus posibilidades son interesantísimas.

Una aplicación de esto es la correlación canónica aplicada al análisis discriminante.

Luego podemos concluir que la correlación canónica nos da grandes posibilidades de estudio de variables cualitativas.

Un caso particular es cuando una de las variables es

dicotómica, pues podemos estudiarlo por regresión múltiple, es una aplicación similar al caso de dos grupos en análisis discriminante por regresión múltiple (pero teniendo en cuenta que al hablar de regresión implica causalidad).

Si ambas variables fueran dicotómicas podríamos aplicar la regresión simple;

En conclusión el estudio de variables cualitativas se puede realizar mediante el caso dos de nuestro modelo (aplicando canónica en general), o mediante el caso uno, (por componentes principales \Rightarrow análisis de correspondencias).

TABLAS DE CONTINGENCIA.

Esto se refiere al caso en el que intentamos estudiar variables cualitativas y su relación entre sí.

Generalmente su estudio se ha basado en la utilización del estadístico $\sum \frac{(O - E)^2}{E}$, que sigue asintóticamente la χ^2 , con lo que estudiamos si se sigue la distribución teórica (1), si esto no se cumple, diremos que hay relación.

- (1) Esta distribución teórica hace referencia al caso de no relación entre las variables.

Pero esto se puede estudiar también mediante la correlación canónica entre variables cualitativas, que nos dirá - si hay o no hay correlación. Además nos da idea de como influyen las distintas variables.

Si una de las variables es dicotómica podemos aplicar la correlación múltiple.

ANALISIS DE LA VARIANZA:

El análisis de la varianza pretende comprobar si los promedios son iguales cuando las varianzas son iguales.

Podríamos dividirlo en ANOVA y MANOVA (1).

Pero en definitiva siempre se trata de una prueba de hipótesis y por lo tanto de inducción estadística con lo que se sale de nuestro objetivo de "análisis de datos".

- (1) La bibliografía sobre el tema es muy amplia. El análisis de la varianza lo trata Chacón (1.955) pgs. 341-398 Tomo 2º.

Recientemente ha sido bastante estudiado y así tenemos a Tatsuoka (1.971) pag. 194 y siguientes que estudia claramente al Manova y el Anova como un caso particular.

De todas formas es perfectamente coherente con nuestro modelo, del cual se puede seguir su estudio, bastaría con aplicar variables dummy.

VI.- CONCLUSIONES.

A lo largo de esta primera parte hemos tratado de unificar el análisis multivariante. Mediante la aplicación del proyector hemos estudiado los dos casos generales del modelo y luego hemos ido viendo como se pueden aplicar estos dos casos a los diversos problemas de análisis multivariante que no estaban específicamente englobados en los dos casos citados.

Con todo esto hemos visto que las actuales técnicas de análisis multivariante son casos particulares de esquemas más generales, con lo que podemos intuir que su desarrollo no ha hecho más que empezar.

Para cada problema concreto, y en base al esquema expuesto, podemos deducir técnicas aplicables a nuestro caso. En efecto, es relativamente sencillo el obtener nuevos sistemas de análisis y aplicaciones, pues el modelo es altamente flexible.

Podríamos de hecho tratar casos nuevos partiendo de la aplicación de las variables dummy, o variando ciertas condiciones dentro del modelo general pero el objeto de este trabajo ha sido crear el modelo, que considero bastante original, tanto en su obtención como en su aplicación, y con ello exponer una base de estudio para futuros análisis y una posibilidad

lidad de obtención matemática de nuevas técnicas.

Podríamos plantear aquí la discusión sobre si se deben de investigar las técnicas y posteriormente buscar las aplicaciones o el proceso debe ser el contrario, es decir, cuando surge el problema ver cuál es la técnica a aplicar y si esta no existe, intentar obtenerla. Considero que este último planteamiento es el correcto pues el análisis - multivariante está fundamentalmente enfocado a su aplicación en ciencias como la economía, la sociología, la psicología, etc, y estas ciencias se encuentran constantemente con problemas nuevos para los cuales nuestras formas de análisis pueden aportar una solución.

Nuestro modelo apunta a esta posibilidad, pues mediante la exposición unificada tenemos la oportunidad de ir deduciendo nuevas técnicas conforme se van planteando las necesidades.

Por otro lado un modelo tan flexible haría penoso, e incluso inútil, el intentar deducir todas las técnicas posibles.

También nuestra unificación nos da base para la aplicación conjunta de estas formas de análisis, es decir, aplicacar estos planteamientos de forma encadenada, como haremos en las partes prácticas.

Por último, quiero comentar que las partes más originales pueden ser:

- Visión geométrica mediante el proyector de como y en qué casos las correlaciones simple o múltiple son un caso particular de la canónica y esquema general de las correlaciones.
- Explicación del análisis de componentes como proyección de variables en los distintos espacios mediante la aplicación del proyector y generalización del mismo para distintos tipos de análisis factorial empleados normalmente.
- Y sobre todo la visión unificada de todas estas técnicas multivariantes mediante la utilización del proyector.

VII.- EL PROBLEMA DE LA INFERENCIA.

Ya en la introducción citábamos que nuestro objetivo era circunscribirnos al análisis de los datos, es decir, no plantearnos los problemas de inferencia estadística. Con el lo suprimíamos la dificultad del salto de la muestra al colectivo.

Este problema no es objeto del presente trabajo, pero creo que éste quedaría incompleto si no hacemos al menos una mención al tema.

En la obra clásica de Análisis Multivariante que es la ya citada de Anderson, tenemos que en su capítulo 14 dice textualmente: "We shall here consider some of the more important and more recent developments in statistical inference - to be applied to multivariate normal distributions", por lo tanto pretende un estudio generalizado y a pesar de tratarse de una obra de 1.958, sus desarrollos veremos que no se han superado de forma notable.

Hablando concretamente de "Factor Analysis", cita el método centroide (hoy poco usado dado el desarrollo de los ordenadores), el de componentes principales y el de máxima verosimilitud. El problema más estudiado respecto a la inferencia es el cálculo del número de factores (o componentes según los autores), que realmente existen.

Este punto concreto está más desarrollado por Kendall y Stuart y por Harman.

La conclusión que podemos sacar del estudio de Anderson es que lo poco estudiado sobre inferencia se basa en la normalidad de las variables originales y si éstas no son normales, poco podemos decir al respecto.

Hay estudios posteriores del mismo Anderson entre los cuales podemos citar el de "Asymptotic theory for principal component analysis". (1)

Kendall y Stuart, en la obra ya citada, capítulo 42, dice que en el estado actual del conocimiento, solo se ha desarrollado el caso de variables normales y citan test como los de "Barlett" y "Lawley y Maxwell" (2) para conocer si el número de factores es satisfactorio.

Más completo, en cuanto al tratamiento del problema, es Harman, en su obra "Modern Factor Analysis". (3)

- (1) Artículo publicado en la revista "The annals of mathematical statistics" 1.963, pas. 122 - 148.
- (2) La obra de Barlett de 1951, que hemos citado se encuentra publicada en "Biometrika", 38, 337.
La de Lawley y Maxwell (1963) describe el método de Lawley (1940) basado en el ratio de máxima verosimilitud y los desarrollos posteriores de otros autores.
- (3) Harman (1967), la primera edición es de 1960, trata ampliamente los autores que han estudiado el problema (pag. 16) los test sobre el número de factores (pgs. 196-197) por el método de minres y (pgs. 219-220) por el de máxima verosimilitud pues en estos métodos no se supone el nº de factores. En factores principales se supone y los test tratan de estudiar si el supuesto es válido.

El libro "Geological Factor Analysis" ya citado, distingue en su capítulo tercero entre el uso descriptivo del análisis factorial (Fixed case) y cuando los datos provienen aleatoriamente de un colectivo (Random case). En este último supuesto comenta que al principio solo se estudió el "Fixed case" y solo posteriormente se estudió el "Random case" que es todavía oscuro.

Este mismo libro expone claramente los intervalos de confianza (pag. 77) para los valores propios, en caso de variables normales.

Torrens-Ibern (1) estudia también el problema y cita diversos test.

También Darmois (2) estudia el problema de la normalidad.

Anderson y Rubin tratan el problema de manera extensa, bajo hipótesis de normalidad, en su artículo "Statistical Inference In Factor Analysis". (3)

- (1) Torrens-Ibern (1972) estudia los test en las pags. 79, 103, 108, 112, 121, 137. También plantea la normalidad en las pags. 31 y 32.
- (2) Darmois (1.955).
- (3) Anderson y Rubin (1.954 - 1.955).

Esta breve alusión bibliográfica, no pretende en absoluto ser exhaustiva, pero si al menos dar una visión de lo que a nuestro juicio es más interesante en esta problemática.

Por lo que se refiere al análisis canónico, nos encontramos en una situación parecida, pues casi todo lo estudiado en inferencia a cerca de él se basa en la hipótesis de normalidad de las variables, a este respecto tenemos el estudio del modelo muestral en el ya citado trabajo de Baigorri (1). De todas formas es el modelo canónico el que más se ha desarrollado en cuanto al estudio del modelo muestral dentro del análisis multivariante, como bien dice Baigorri.

Vista la situación en los dos casos fundamentales que son el modelo canónico y el de componentes principales, es de esperar que las cosas no estén mucho mejor en las aplicaciones concretas.

Una excepción a esta regla la representa el caso de la econometría por haber llevado un camino independiente al del resto de las técnicas multivariantes más desarrolladas. (2)

(1) Baigorri (1.979) pgs. 131 y siguientes.

(2) Puede consultarse al respecto cualquier tratado sobre econometría. El más clásico es el Johnston (1.975) por tratarse de un libro completo y sencillo.

Pero el resto de las técnicas, como por ejemplo el análisis discriminante, se encuentran muy desatendidas en este campo. En la bibliografía por nosotros consultada, - poco se habla de la inducción en los métodos tradicionales de análisis discriminante, pero respecto a los métodos más modernos, como son el de programación lineal (o no lineal) y el de regresión logística, el problema creo que podemos afirmar que ni siquiera se ha planteado de forma matemática.

Lo más que se ha hecho ha sido, si tenemos n individuos, dividir esa muestra en dos grupos n_1 y n_2 , de forma aleatoria y luego comprobar si para los n_2 restantes se cumple bien la discriminación, pero esto no nos dice nada del grado de confianza etc.

Una situación similar tenemos con el cluster análisis y con otras técnicas.

Ante este panorama hemos decidido, al igual que hacíamos en esta primera parte, estudiar fundamentalmente el modelo poblacional en las aplicaciones prácticas.

Esto no representa ningún problema en la parte tercera de este trabajo, pues al tratarse de un estudio sobre los municipios de la provincia, no se plantea ningún problema inductivo y podemos aplicar con toda propiedad las conclusiones de nuestro análisis de datos, es decir, estudiar

do de forma multivariante una población con fines que esta
dísticamente calificaríamos de descriptivos. Esto es posi
ble pues al tratarse de un estudio económico de las carac-
terísticas que en su momento veremos, el estudio se hace -
con todos los datos existentes.

Pero donde nos encontramos con un problema mayor es
en la segunda parte, donde tratamos de hacer un estudio de
tipo sociológico sobre los riojanos. Evidentemente se ha -
de basar en una muestra y en consecuencia nos encontraremos
con el problema de la inferencia.

Por otro lado las variables que vamos a utilizar, di
fícilmente se puede suponer que son normales, como veremos
en su momento, con lo que carecemos de instrumentos estadíss
ticos para intentar la inducción.

La técnica utilizada preferentemente va a ser el análi
sis de componentes principales (análisis factorial según au-
tores), que como hemos comentado carece de estudios de tipo
inductivo fuera de algún test bajo hipótesis de normalidad.

(1)

- (1) Dagnelie (1.975) dice en las páginas 182 y siguientes
que los test aplicables a componentes principales se
basan en las hipótesis de normalidad de las variables,
aunque se pueden aplicar a variables que sean algo -
normales. Pero que en todo caso son muy poco usadas y
se sustituyen por otras reglas.

En la práctica estos test no se usan casi nunca dado que dan lugar a excesivo número de factores y lo que se hace es utilizar reglas prácticas como rechazar los valores propios menores que uno (regla de Kaiser) (1) o incluso coger los factores mientras tengan una fácil interpretación o aporten algo al conocimiento del problema (2), dado que de nada nos sirve conseguir un número elevado de factores y luego no poder interpretarlos.

Hay otras reglas como el método gráfico de Cattell (3) conocido como el "Scree test", que es un refinamiento del de un pionero en este campo, como fue Tucker (4).

Pero en general podemos ver que no existen test realmente científicos fuera de la hipótesis de normalidad, ni fundamentos teóricos aplicables en la práctica. En realidad lo que se suele hacer es estudiar el modelo muestral de forma que prácticamente se trata como poblacional y luego se supone que si la muestra es buena, el resultado puede aplicarse al colectivo.

(1) Kaiser (1.960). 20, pgs. 141 - 151.

(2) Estas ideas las mantiene entre otros Dagnelie (1.975) pag. 182.

Tatsuoka (1.971) pag. 146, afirma que si utilizamos los ratios teóricos nos encontraremos con excesivos factores que nos servirán de muy poco.

(3) Cattell (1.966) critica la regla de Kaiser y propone el "Scree test".

(4) Tucker (1.938) trabajo de interés dada su fecha de publicación.

Evidentemente careceremos de intervalos de confianza, etc, pero si daremos una buena estimación.

Luego en conclusión, y ante la carencia de instrumentos estadísticos adecuados, en nuestra segunda parte trataremos el modelo como poblacional sobre una muestra aleatoria, y desde un punto de vista práctico eso nos servirá como una buena estimación del colectivo.

PARTE SEGUNDA

0.- INTRODUCCION.

En el presente trabajo queremos hacer un estudio sobre el pensamiento riojano. Por lo tanto no vamos a hacer un estudio sociológico completo, sino simplemente un análisis estadístico sobre opiniones, a cerca de diferentes temas, al objeto de tener una idea de lo que piensa la gente de esta región.

En primer lugar creo que esto tiene un indudable interés sociológico, pues opino que es quizá el estudio más serio que se ha hecho sobre opiniones de los Riojanos, a - causa, tanto del rigor con que he intentado llevar la recogida de datos, como de lo completo que ha sido el tratamiento de la información.

En realidad en la Rioja hemos carecido hasta el presente de estudios fiables de opinión, por otro lado nunca se habían aplicado sistemáticamente las técnicas de análisis multivariante al estudio sobre la opinión de los Riojanos, e incluso su aplicación a estudios de opinión en general ha sido hasta el presente, bastante discreta.

Por lo tanto creo que el estudio tiene un gran interés en este momento en que los vientos autonómicos soplan por la Rioja. Este análisis pretende ser imparcial, serán los sociólogos, los políticos, etc. los que habrán de sacar las consecuencias.

En segundo lugar, el trabajo creo que tiene un gran interés que sobrepasa con mucho su estricto campo de aplicación al colectivo riojano. Este interés lo veo yo en la posibilidad de hacer estudios similares en otras regiones españolas, que contando con mayores medios puedan ser más ambiciosos y completos.

En tercer lugar, me parece que el estudio tiene un gran interés estadístico como ejemplo de aplicación conjunta de las técnicas de análisis multivariante. En realidad esta fue la concepción al empezar el análisis, ver una parte de las inmensas posibilidades del análisis multivariante aplicado a los estudios sociológicos. Es por esto que me he limitado al estudio de las variables que tengan posibilidad de un tratamiento multivariante y solo a una parte de ellas, que considero más interesantes.

En conclusión lo que se pretende es hacer una aplicación conjunta de las técnicas de análisis multivariante a unos datos sobre la opinión de los riojanos, y teniendo siempre en cuenta las limitaciones de medios del autor.

Se nos plantea aquí por lo tanto el problema ya comentado en la primera parte de pasar de los resultados muestrales al colectivo (1). Como ya dijimos, este problema

(1) Dado que este estudio se ha de basar lógicamente en una muestra aleatoria.

ma lo vamos a dejar a un lado y lo que haremos es aplicar el análisis de datos sobre nuestra muestra aleatoria; pero si siguiendo a autores como Dagnelie (1), Tatsuoka (2), y tantos - otros, diremos que el análisis multivariante que nosotros va mos a aplicar (que fundamentalmente va a ser componentes prin cipales) en el cual vamos a mantener, incluso por exceso, la regla de Kaiser (3), tiene validez para sacar conclusiones e sobre el colectivo, aunque careciendo del grado de confianza de las mismas.

En conclusión, y a la vista de las variables que ten dremos, dado el actual desarrollo de la estadística, lo más que podemos hacer es realizar un análisis de datos sobre una muestra aleatoria, desde un punto de vista estrictamente es tadístico. Pero, sin excesivo miedo a equivocarnos, podemos trasladar las conclusiones de la muestra al colectivo, sa biendo que es una buena estimación del mismo. (4)

(1) Dagnelie (1.975) pgs. 182 y siguientes.

(2) Tatsuoka (1.971) pgs. 146 y siguientes.

(3) Kaiser (1.960) pgs. 141 - 151.

(4) Esto por lo que se refiere al análisis de componentes principales, que ya hemos comentado que va a ser el más asiduamente utilizado.

I.- RECOGIDA DE LA INFORMACION. BASES DEL TRATAMIENTO.

Ante el proyecto de estudio esbozado en la introducción, me encontré en primer lugar con la necesidad de plantear una serie de preguntas, es decir, confeccionar una encuesta suficientemente amplia como para poder aplicar un análisis estadístico coherente, y que no fuese excesivamente larga para no cansar al encuestado, que en principio, me hacía un favor al contestar a mis preguntas.

Por otro lado, la encuesta debería tener los suficientes puntos, como para permitir un juicio sobre el pensamiento del encuestado.

Si la encuesta tenía que ser de opinión, no cabe ninguna duda de que había que estudiar la posibilidad de cuantitativizar de alguna manera las opiniones, que en principio son variables cualitativas. Un procedimiento podía haber sido la aplicación de vectores canónicos, tal como hemos estudiado en la primera parte, pero siguiendo el estudio de -- Michael Linden sobre opiniones de candidatos políticos en Suecia (1), decidimos hacer una escala desde totalmente de acuerdo, al cual le dábamos el valor siete, hasta totalmente en desacuerdo, dando entonces el valor uno.

El procedimiento es el siguiente: Se le hace una afirmación al encuestado y éste puntúa su grado de acuerdo con

(1) Linden M. (1.975).

la afirmación en una escala de uno a siete.

El problema es ahora el ver que tipo de afirmaciones hemos de hacer, de manera que obtengamos una mejor información sobre el pensamiento del encuestado.

Para decidir el tipo de preguntas contaba con las expuestas en el citado trabajo de Linden, el problema era que en ese trabajo se pretendía conocer actitudes ante problemas políticos y mi estudio era más ambicioso y no se circunscribía únicamente a la política.

Por otro lado, Linden iba a encuestar a candidatos a unas elecciones, por lo tanto con cierto nivel cultural y con un conocimiento, más o menos importante, sobre problemas de cierta complejidad. Este no era mi caso, pues mis encuestados iban a ser gente normal y desgraciadamente dado el bajo nivel cultural de España en general y de la Rioja en particular, iba a ser difícil que el encuestado respondiera a preguntas complejas.

Ante esta problemática decidí reunir a una serie de personas, de suficientes conocimientos sobre la región y organizar un "Brain Storming", para intentar sacar una serie de temas y preguntas.

Los temas fundamentales que obtuve tras la cuidadosa selección fueron los siguientes:

- Cultural.
- Político - Social.
- Autonomía.
- Campo.
- Religión.
- Familia y mujer.

Con estos temas había que elaborar las preguntas. En primer lugar era preciso no provocar el rechazo del encuestado por lo que las preguntas deberían ser generales y no personales, a parte de que se trata de un estudio de opinión.

En segundo lugar, el encuestado se podía negar a responder una pregunta que considerase excesivamente fuerte, - por lo que las preguntas tenían que ser suaves y no muy comprometidas.

Por último, como ya hemos dicho, tenían que ser preguntas sencillas, para que las entendiese todo el mundo.

En conclusión, elaboramos una lista de 33 preguntas, que resumían de una manera sencilla una parte de la problemática que se pretendía estudiar.

Evidentemente se le pueden poner objeciones a esta - lista, pero creo que a toda lista de preguntas se le pueden objetar cosas.

En esta lista mezclamos las preguntas sobre diferentes temas al objeto de evitar correlaciones extrañas entre las variables, es decir, un individuo puede sentirse condicionado si las preguntas del mismo tema van una detrás de otra. Mezclándolas dábamos mayor libertad al encuestado.

Finalmente añadimos nueve variables, en realidad número de la encuesta y ocho más, de tipo más objetivo, para poder completar nuestro estudio. Estas nueve variables las pusimos al principio de la encuesta, y sobre su puntuación ya hablaremos posteriormente.

El resto de variables, es decir las 33 variables de opinión, se puntuarían de uno a siete como ya hemos comentado con anterioridad.

La encuesta quedó como sigue; en primer lugar veremos como quedó por temas y después como se reordenó mezclando - las preguntas de cada tema:

A) Datos:

- Número de la encuesta.
- Sexo.
- Edad.
- Estado.
- Profesión.
- Clase social.
- Estudios.

- Lugares que conocen.
- Lugar de la encuesta.

B) Cultural:

- Los licenciados están bien preparados para realizar su trabajo.
- El periódico es necesario para estar bien informado.
- Leer libros es perder el tiempo.
- Las películas españolas son buenas porque divierten.
- Habría que aumentar los gastos municipales en actividades culturales aun a costa de reducir otros gastos.

C) Político - Social:

- Hoy en día el divorcio debe ser permitido.
- Los cuarenta años con Franco han sido muy buenos.
- El estado debe nacionalizar las grandes empresas.
- La policía debe emplear mano dura.
- No debe permitirse la huelga.
- Para España una república es mejor que una monarquía.
- El servicio militar no debe ser obligatorio.

D) Autonomía:

- Se debe dejar libertad a las regiones.
- Desde Madrid se controlan demasiado las regiones.
- La Rioja debe estar unida a otras regiones (citar cual).
- La Rioja debe ser autónoma y no debe incluirse en otras regiones.
- Usted se ha considerado siempre riojano.
- La Rioja ha sido menos beneficiada que las regiones que le rodean.

E) Campo:

- Aquí la gente se lo pasa bien.
- Los problemas del campo se arreglarían si hubiese más cooperativas.
- Las cooperativas nunca podrán funcionar bien.
- El dinero del campo debe llevarse a las zonas industriales.
- Esta zona debería ser industrializada.

F) Religión:

- Dios realmente existe.
- Los curas viven muy bien.
- En los momentos difíciles pensar en Dios supone un consuelo.
- Se debe ir a misa.

G) Familia y mujer:

- Los niños no deben saber nada del sexo.
- Los niños pequeños deben opinar.
- No se deben tener niños antes del matrimonio.
- La mujer debe estar en casa cuidando a los niños y esperando al marido.
- Es beneficioso que la mujer trabaje fuera de casa.
- Las mujeres en general son menos inteligentes que los hombres.

Como se puede ver en la redacción de las preguntas, se ha cuidado que no todas vayan en el mismo sentido, es de cir, que un individuo coherente no podrá estar de acuerdo con todas.

Como ya hemos dicho estas preguntas las mezclamos de manera que los del mismo grupo no estén juntas y la ordenación final quedó:

ENCUESTA DE OPINION EN LA RIOJA

PUNTUACION (preguntas 10 - 42)

- 1 totalmente en desacuerdo.
- 2 bastante en desacuerdo.
- 3 simplemente en desacuerdo.
- 4 indiferente.
- 5 simplemente de acuerdo.
- 6 bastante de acuerdo.
- 7 totalmente de acuerdo.

-
- 1) NUMERO.
 - 2) SEXO.
 - 3) EDAD.
 - 4) ESTADO.
 - 5) PROFESION.
 - 6) CLASE SOCIAL.
 - 7) ESTUDIOS.
 - 8) LUGARES QUE CONOCEN.
 - 9) LUGAR DE LA ENCUESTA.

-
- 10) LOS LICENCIADOS (Médicos, abogados, ingenieros, etc.)
están bien preparados para realizar su trabajo.
 - 11) Hoy en día el divorcio debe ser permitido.
 - 12) Se debe dejar libertad a las regiones.
 - 13) Aquí la gente se lo pasa bien.

- 14) Dios realmente existe.
- 15) Los niños no deben saber nada del sexo.
- 16) El periódico es necesario para estar bien informado.
- 17) Los cuarenta años con Franco han sido muy buenos.
- 18) Desde Madrid se controlan demasiado las regiones.
- 19) Los problemas del campo se arreglarían si hubiese más cooperativas.
- 20) Los curas viven muy bien.
- 21) Los niños pequeños deben opinar.
- 22) Leer libros es perder el tiempo.
- 23) El estado debe nacionalizar las grandes empresas.
- 24) La Rioja debe estar unida a otras regiones (citar cual; como por ejemplo: Aragón, Castilla, Navarra, País Vasco, etc.).
- 25) Las cooperativas nunca podrán funcionar bien.
- 26) En los momentos difíciles pensar en Dios supone un con suelo.
- 27) No se deben tener niños antes del matrimonio.
- 28) Las películas españolas son buenas porque divierten.
- 29) La policía debe emplear mano dura.
- 30) La Rioja debe ser autónoma; y no debe incluirse en otras regiones.
- 31) El dinero del campo debe llevarse a las zonas industriales.
- 32) Se debe ir a misa.
- 33) La mujer debe estar en casa, cuidando a los niños y espe rando al marido.
- 34) Habría que aumentar los gastos municipales en actividades culturales aun a costa de disminuir otro tipo de gasto.
- 35) No debe permitirse la huelga.

- 36) Ud. se ha considerado siempre riojano.
- 37) Esta zona debería ser industrializada.
- 38) Es beneficioso que la mujer trabaje fuera de casa.
- 39) Para España una república es mejor que una monarquía.
- 40) La Rioja ha sido menos beneficiada que las regiones que la rodean.
- 41) Las mujeres en general son menos inteligentes que los - hombres.
- 42) El servicio militar no debe ser obligatorio.

Ya comentaremos posteriormente las claves de puntuación, vamos a continuación a estudiar el procedimiento de muestreo.

Se ha pretendido obtener una muestra lo más representativa posible del colectivo a estudiar, que como ya hemos dicho se trata de los habitantes de la Rioja.

En primer lugar, es preciso delimitar exactamente el colectivo:

En el aspecto geográfico La Rioja no es una unidad exacta. Normalmente se suele confundir con la provincia de Logroño, pero esto no es del todo correcto.

Como región natural, La Rioja es una vega a ambas partes del Ebro, que partiendo de las "Conchas de Haro"

termina confundiéndose con la ribera Navarra y con Aragón.

Fundamentalmente esto es la provincia de Logroño o de "La Rioja", (1), pero hay una porción importante de la provincia de Alava que es la conocida "Rioja Alavesa". También hay una porción de la provincia de Navarra e incluso hay zonas lindantes con Burgos de difícil adscripción.

El problema se complica todavía más si consideramos que la parte sur de la provincia es muy accidentada, dado que allí está la Sierra de la Demanda, y es difícil precisar si esto es Rioja o no en el sentido geográfico, ya - que aunque no puede considerarse Vega, si es mayoritariamente cuenca del Ebro, a diferencia de la provincia de Soria que envía mayoritariamente sus aguas al Duero.

Esta es la problemática geográfica, pero desde un punto de vista sociológico y humano, y en base a las apreciaciones de personas con suficientes conocimientos sobre el tema, podemos colegir que despreciando las insignificantes partes de la provincia de Burgos que pueden considerarse como Rioja y haciendo lo propio con ciertas localidades de Navarra, donde por otra parte el Navarrismo es muy superior al sentimiento riojano, La Rioja quedaría circunscrita a la provincia de Logroño y a la Rioja Alavesa con los

(1) Puede consultarse el apéndice histórico-geográfico de la provincia que exponemos al final de este trabajo. De esta manera, el lector, si no está familiarizado con la provincia objeto de nuestro estudio, tendrá ocasión de conocerla algo más.

siguientes matices:

En primer lugar la zona de la Sierra, fundamentalmente la comarca denominada de "Los Cameros", será más o menos Rioja en el sentido geográfico de la palabra, pero humanamente, todos sus habitantes se consideran riojanos, por lo que la incluiremos en nuestro estudio. Tengamos en cuenta además - que en el momento de escribir estas líneas, se está discutiendo a nivel político el cambio de nombre de la provincia de Logroño, por el de Rioja; este cambio ya se ha aceptado en el Senado y va a comenzar ahora los trámites para su definitiva legalización; comento esto para hacer resaltar hasta qué punto toda la provincia de Logroño se identifica con el nombre de "La Rioja", pues la petición de este cambio de nombre ha sido apoyada por todas las fuerzas políticas y por el pueblo en general, es decir, por toda la provincia (Cameros incluido). Luego, en principio, consideraremos como Rioja a toda la provincia de Logroño.

En segundo lugar, y por lo que se refiere a la Rioja Alavesa el problema se complica bastante más. En el momento de realizar la encuesta (verano de 1.977), no existían todavía estatutos de Autonomía, pero ya podía preverse que, al menos las tres provincias vascongadas, llegarían a formar - un ente autonómico. De hecho, hoy ya tenemos refrendado el estatuto de Autonomía para el País Vasco y la Rioja Alavesa, se incluye naturalmente es este estatuto. El problema parte de que los entes autonómicos deben circunscribirse a provinin

cias y que la división en provincias, resulta algo caprichosa en ciertos casos. Sin entrar en discusiones políticas, y sin pretender ningún objetivo partidista ni regionalista, - parece bastante evidente que la Rioja Alavesa no tiene razones ni sociales ni económicas para estar separada del resto de la Rioja. Pero el problema está ahí y sería absurdo cerrar los ojos a la realidad, la Rioja Alavesa se encuentra políticamente enclavada en el País Vasco y social y económicamente vive con Logroño. A este respecto hay un interesantísimo trabajo dirigido por Uriarte (1) que aunque todavía inconcluso comenta esta realidad dual de la Rioja Alavesa.

A la vista de lo expuesto la decisión es difícil, y nos hemos inclinado por incluir a la Rioja Alavesa en el estudio sociológico, pero prescindiremos de ella en el estudio económico que realizaremos en la tercera parte, pues la planificación y dirección económica de esta comarca no va a depender de Logroño sino de los entes vascos correspondientes.

Sin embargo, creo que la Rioja Alavesa humanamente está unida al resto de la Rioja, y no sería completo un estudio sobre la opinión de los riojanos sin tenerla en cuenta.

(1) El citado trabajo se denomina "La reconversión de la vida económica del país vasco" y patrocinado por I.K.E.I. está siendo realizado por el Gabinete de organización y desarrollo. Puede consultarse el estudio sobre los municipios de Alava, concretamente cuando hace mención a la Rioja Alavesa.

Además y volviendo a la discusión sobre el cambio de nombre de la provincia de Logroño, en el pleno del Senado donde se trató el tema, fueron los parlamentarios Alaveses los que pusieron la única objeción, cuestionando no el cambio de nombre, sino que se reconociera el derecho de la Rioja Alavesa a seguir denominándose así. En el mismo sentido se había pronunciado con anterioridad la Diputación Foral de Alava. Datos éstos que confirman el sentimiento riojano de la Rioja Alavesa, Cosa que no impide que tenga hoy o que pueda tener en el futuro un sentimiento vasco, pues nadie puede afirmar que dichos sentimientos sean incompatibles.

Luego en definitiva, el colectivo a estudiar van a ser los habitantes de la Rioja, entendida ésta como provincia de Logroño y Rioja Alavesa.

Siguiendo con criterios fundamentalmente geográficos, hemos decidido dividir esta región en cinco comarcas (1), (sin perjuicio de que en la parte tercera de este es

- (1) Esta división es la tradicional de la provincia (cuatro comarcas) y otra de la Rioja Alavesa. La presente comarcalización se encuentra recogida en multitud de estudios sobre la Rioja, que coinciden en lo fundamental con pequeñas variaciones. Entre otras publicaciones la podemos encontrar en: "Estructura y perspectivas del desarrollo económico de la provincia de Logroño", publicada por la antigua organización sindical y realizada bajo la dirección del consejo económico sindical provincial. (1.970) pg. 279.

tudio se haga una comarcalización mucho más rigurosa en ba
se a criterios económico - geográficos, esta comarcalizacio
n nos es suficiente para el presente análisis).

- 1 Logroño y Rioja Media.
- 2 Rioja Alta.
- 3 Rioja Baja.
- 4 Sierra-Cameros.
- 5 Rioja Alavesa.

A efectos de muestreo el planteamiento ha sido el -
siguiente:

Intentamos conseguir una muestra aleatoria de indi-
viduos de la región a estudiar, para ello lo mejor era con
tar con un censo fiable y actualizado y como en el momento
de realización de la encuesta ya se habían realizado dos -
consultas electorales (referéndun para la Ley de reforma -
política y posteriores elecciones a Cortes del 15 de Junio
de 1.977), consideramos que el instrumento apropiado bien
pudiera ser el censo electoral.

No fue fácil conseguirlo, pues los organismos oficial
es son bastantes reacios a permitir el acceso a sus archiv
os, pero tras gestionar su obtención con el Presidente de
la junta electoral de zona, pudimos obtenerlo. Esto por lo
que se refiere al censo electoral de la provincia de Log -
roño, porque los datos precisos de la provincia de Alava
fueron imposibles de conseguir; pero dada su poca importanci

cia numérica, (escaso peso del número de habitantes de la Rioja Alavesa, respecto al total de la Rioja), pudimos resolver el problema como veremos posteriormente.

Lo que decidimos fue hacer un muestreo sistemático sobre el censo electoral (de cada equis páginas una) y - dentro de cada página utilizábamos un número equiprobable.

Muestreamos así, aproximadamente el 1'5 0/00 del censo electoral.

El muestreo así obtenido es altamente significativo, pues al estar el censo ordenado por municipios, se logró - una muestra prácticamente proporcional al número de habitantes de cada municipio y dentro de cada municipio el individuo a muestrear era absolutamente aleatorio.

Con este sistema, todos los individuos de la provincia tienen la misma probabilidad de salir y además saldrán de forma casi proporcional a los habitantes del municipio.

El problema como ya hemos comentado era la provincia de Alava, en concreto la Rioja Alavesa, pues por problemas administrativos no pudimos llegar al censo.

Lo que hicimos fue sortear pueblos e individuos manteniendo aproximadamente la misma proporción de muestreo.

Una vez terminado este proceso teníamos una lista de 267 individuos a muestrear, repartidos por toda la Rioja. - Estos 267 individuos eran mayores de 18 años (pues solo a partir de esa edad se está incluido en el censo electoral, salvo pocos casos). Luego, nuestro colectivo eran los riojanos de más de 18 años y nuestra muestra era de 267 individuos lo que representaba el 1'5 0/00. (En realidad sería más exacto decir que nuestro colectivo era el censo electoral, pues de todas son conocidas las deficiencias del mismo, aunque para ser sinceros era el mejor instrumento para nuestro estudio, y el único que nos permitía un muestreo relativamente sencillo).

Una vez obtenidos los datos para el muestreo, me encontré con el problema mucho mayor de llevarlo a la práctica. Conté para ello con la inestimable colaboración de un grupo de amigos, que desinteresadamente y con gran cuidado llevaron a cabo la ingrata empresa de ir buscando por pueblos y calles a los individuos concretos.

Puedo asegurar y garantizar la fiabilidad de estas encuestas pues conozco a los encuestadores y se que pusieron el mayor cuidado tanto en la búsqueda de los individuos como en la posterior entrevista. Digo esto, porque estamos demasiado acostumbrados, los que hemos de trabajar con encuestas, a que los encuestadores realicen su trabajo de cualquier manera e incluso muchas veces rellenen las encuestas en su casa. En el presente trabajo, podemos es-

tar tranquilos a este respecto, pues la toma de datos se hizo de forma escrupulosa, encuestando a todos los individuos y tratando el encuestador de no influir lo más mínimo en el encuestado.

Creo que es justo reconocer que sin el cuidado, el esmero y el desinterés de este grupo de jóvenes, hubiera sido imposible que yo solo hubiera concluido la penosa tarea de las 267 encuestas repartidas por toda la Rioja.

Como es natural, muchas veces no se encontró al individuo concreto que había que encuestar según los datos del muestreo, si el individuo existía, se volvía otro día y a otra hora, si esto era posible (tengamos en cuenta - que en pueblos alejados de la capital esto era imposible por lo costoso de los desplazamientos), en otros casos el individuo no existía, bien había muerto, bien por deficiencias del censo nunca había vivido allí o se había trasladado, etc. En todos los casos en los que fue imposible - muestrear al individuo concreto, se buscó a alguien similar, es decir, del mismo barrio, sexo, edad, estado, etc.

En conclusión creo que la tabla de datos es altamente significativa, por lo cuidadoso de su elaboración.

Quizá pueda argumentarse que la muestra es pequeña, pero la realidad es que la mayoría de los estudios sociológicos, se hacen con muestras inferiores al 10/00 y una -

muestra del uno por mil es considerada como altamente signi
ficativa, nosotros hemos ido todavía más lejos y hemos en -
cuestado al 1'5 0/00.

Por supuesto que nos hubiese gustado el tener uan mu
estra más numerosa, pero sin medios económicos y con la co
laboración de un grupo de amigos de los cuales tampoco se
podía abusar, creo que lo conseguido es incluso superior a
lo que cabía esperar a priori.

Por otro lado lo que pretendemos es una aplicación
de análisis multivariante y para nuestros objetivos, los
datos son más que suficientes.

Una vez que hemos analizado la encuesta y que hemos
estudiado el procedimiento de muestreo, creo que podemos -
pasar a ver como se codificaron los datos. La codificación
fue doble:

En primer lugar, cuando el encuestador había termi
nado su trabajo, me entregaba los datos que yo introducía
en unas hojas resumen de encuestas y luego se cambió algo
el código para su entrada en ordenador.

La forma de puntuación fue la siguiente:

- a) Para los datos.
 - 1) El número va del 1 al 267, numerados por comarcas:

Logroño capital: 1 - 100
Logroño comarca: 101 - 109
Rioja Alta: 110 - 169
Rioja Baja: 170 - 243
Cameros Sierra: 244 - 257
Rioja Alavesa: 258 - 267

En el fichero realizado de cara a la utilización del ordenador, no se ha utilizado esta pregunta, pues las fichas ya van numeradas y se ha respetado en su numeración, el código anteriormente citado.

En el fichero de ordenador, se ha introducido en la pregunta 1 el apartado correspondiente a la pregunta 24 de a qué región quiere unirse dándose el siguiente código:

0 - No dice.
1 - A Aragón.
2 - C Castilla.
3 - N Navarra.
4 - P.V. País Vasco.

2) El sexo se puntúa:

M - Masculino - 7
F - Femenino - 1

3) La edad:

Hasta 25 - 1

25 - 35	-	2
35 - 45	-	3
45 - 55	-	4
55 - 65	-	5
65 - 75	-	6
más de 75	-	7

4) Estado:

Soltero	-	1
Casado	-	2
Viudo	-	3

5) Profesión:

Jubilado	}	-	1
Sus labores			
Obrero	-	2	
Agricultor	-	3	
Administrativo	}	-	4
Oficial			
Autónomo	-	5	
Estudiante	-	6	
Profesor	-	7 (EGB, Formación Profesional, etc.)	
Industrial	-	8	
Universitario	-	9 (ejerciendo la profesión)	

6) Clase Social:

Obrera y baja	-	1
Media baja	-	2
Media media	-	3
Media alta	-	4

7) Estudios:

Nada	-	1
Elemental	-	2
Bachiller elemental	-	3
Bachiller superior	-	4
Formación profesional	-	5
Universidad	-	6
Licenciados y Doctores	-	7

8) Lugares que conocen:

Nada	-	2
Provincia	-	3
Región	-	4
Nación	-	5
Extranjero	-	6

9) Lugar de la encuesta:

Logroño - Rioja media - 1

Logroño	-	1
Albelda	-	2
Alberite	-	3
Lardero	-	4
Nalda	-	5
Villamediana	-	6

Rioja Alta - 2

Angunciana	-	7
Bañares	-	8
Briones	-	9

Casalarreina	-	10
Castañares	-	11
Ciuri	-	12
Cuzcurrita	-	13
Grañón	-	14
Haro	-	15
Manzanares de R.	-	16
Ochánduri	-	17
Rodezno	-	18
San Asensio	-	19
Sto. Domingo	-	20
Santurde	-	21
San Vicente S.	-	22
Treviana	-	23
Villalobar de R.	-	24
Villarta Quintana	-	25
Alesón	-	26
Badarán	-	27
Baños	-	28
Bezares	-	29
Cañas	-	30
Cenicero	-	31
Entrena	-	32
Fuenmayor	-	33
Hormilla	-	34
Huércanos	-	35
Medrano	-	36
Nájera	-	37
Navarrete	-	38

Sojuela	-	39
Tricio	-	40
Uruñuela	-	41

Rioja Baja - 3

Aguilar de R. A.	-	42
Alcanadre	-	43
Aldeanueva de E.	-	44
Ausejo	-	45
Autol	-	46
Bergasa	-	47
Calahorra	-	48
Cervera	-	49
Cornago	-	50
Galilea	-	51
Herce	-	52
Igea	-	53
Pradejón	-	54
Quel	-	55
El Redal	-	56
Rincón de Soto	-	57
Tudelilla	-	58
Villar de Arnedo	-	59
Agoncillo	-	60
Arrúbal	-	61
Murillo	-	62
Alfaro	-	63
Arnedo	-	64

Sierra Cameros - 4

Enciso	-	65
Ezcaray	-	66
Pazuengos	-	67
Anguiano	-	68
Laguna	-	69
Lumbreras	-	70
Nieva	-	71
Ribafrecha	-	72
San Millán	-	73
San Román	-	74
Torrecilla	-	75
Villanueva	-	76
Viniegra de Abajo	-	77

Rioja Alavesa - 5

Laguardia	-	78
El Ciego - Lanciego, etc.		79
Oyón	-	80

b) El resto de las variables (10 a la 42) ya hemos comentado que se puntúan de uno a siete.

La codificación del grupo (a) no ha sido arbitraria y vamos a explicarla ahora. Se ha intentado ordenar de alguna manera las variables cualitativas, para así poderlas tratar como cuantitativas.

Este es el caso de las preguntas 5 (profesión) donde se ha intentado ordenar de menor a mayor según el reconocimiento social, esta ordenación es subjetiva pero considero que es válida. Lo mismo sucede con las preguntas 6 (clase social), 7 (estudios), 8 (lugares que conoce).

El resto de las preguntas de este apartado no se han codificado por orden por resultar imposible y tendremos que tener cuidado con su utilización.

En general, va a dar lo mismo puntuar de cero a siete, que de uno a dos, por ejemplo en una dicotomía, dado que vamos a trabajar con variables normalizadas.

BASES PARA EL TRATAMIENTO.

Una vez confeccionado el banco de datos (1), de una forma apropiada para la utilización del ordenador, nos encontramos ante la necesidad de hallar un centro de cálculo con los medios técnicos y humanos suficientes para poder aplicar el análisis multivariante deseado.

Felizmente, la Facultad de Ciencias Económicas y Empresariales de la Universidad de Deusto, donde yo he estudiado, cuenta con un centro de cálculo, que cumple con los mínimos requeridos para este estudio.

Evidentemente no contamos con un ordenador excesivamente potente, pero si suficiente. Además, tenemos unos programas de análisis multivariante que resuelven la mayoría de nuestros problemas. Es cierto que hubo que remodelar algún programa, pero esto se pudo hacer sin dificultad.

De todas formas tuvimos que sufrir las limitaciones de nuestro ordenador, un IBM 1130, que nos impidió en algunos casos, hacer estudios más amplios.

Como el soporte de entrada era la ficha perforada, implantamos nuestra información en 267 fichas y ya estábamos preparados para comenzar el estudio.

(1) Puede verse en el apéndice de ordenador.

II.- TRATAMIENTO DE LA INFORMACION.

Vamos a estudiar a continuación todo el complejo tratamiento que sufrió la información recogida anteriormente. Este estudio se va a exponer exactamente en el mismo orden en que se efectuó, al objeto de que el lector pueda familiarizarse con él más fácilmente.

En primer lugar, quiero explicar que aunque en la primera parte hemos denominado "Análisis de Componentes Principales" al caso en que se diagonalizaba la matriz de coeficientes de correlación, con unos, por lo tanto, en la diagonalización principal, y "Análisis Factorial" al caso en que se estimaban las comunalidades. Dado que como ya he comentado la literatura especializada confunde a menudo estos términos, he decidido llamar siempre análisis factorial cuando se diagonalice la matriz de coeficientes de correlación, bien con unos o bien con comunalidades en la diagonal principal. La causa de este cambio de nomenclatura no es caprichosa; lo que pretendo con él es hablar un lenguaje en el que todos podamos entenderenos. (1)

- (1) Ya hemos comentado que mientras autores como Tatsuoka (1971) mantienen la denominación diferenciada entre - análisis de componentes principales y análisis factorial en la línea expuesta por nosotros en la primera parte, Saporta (1976) lo entiende de otra manera, y la bibliografía anglosajona engloba ambos conceptos en el término genérico de análisis factorial, como puede verse en Harman (1976) y otros, al igual que Jöreskog, Klován y Reymont (1976). Con lo que en las aplicaciones prácticas hablaremos genéricamente de análisis factorial.

Por lo tanto diré que comenzamos nuestro estudio con un análisis factorial.

El programa de nuestro ordenador lo que hace es diagonalizar la matriz de coeficientes de correlación con unos en la diagonal principal (1), después rotar y medir los factores en los individuos, estimando estas mediciones de manera que resulten normalizadas, siempre que introduzcamos las variables normalizadas.

La razón de introducir unos y no comunalidades es que éstos son siempre difíciles de estimar y cuando se estiman el procedimiento suele ser bastante arbitrario. Por otro lado el estimar las comunalidades tiene por objeto estudiar lo que de común tienen las variables y a nosotros nos interesa no solo estudiar lo común, sino el problema de forma completa (2).

Nuestro ordenador para obtener las puntuaciones (3) - utiliza el método de mínimos cuadrados (para tenerlas así normalizadas), haciendo:

- (1) Si mantenemos los unos en la diagonal principal haremos un análisis factorial por componentes principales.
- (2) El estudiar la parte común es el objetivo de los psicólogos que pretenden ver qué es lo común de todas las variables que hacen referencia a la inteligencia, pero en nuestro caso queremos simplemente disminuir el número de variables para estudiarlas más comodamente.

Por otro lado la utilización de la matriz de coeficientes de correlación con unos o comunalidades en la diagonal principal es un problema que no hará variar mucho los resultados Dagnelie (1.975) pag. 219.

- (3) Puede consultarse cualquier tratado sobre el tema, como por ejemplo Harman (1967) pgs. 348 y ss. o Jöreskog, Klován y Reymont (1976) pgs. 68 y ss.

$$F = B' R^{-1} X$$

Siendo B la matriz A que resulta tras la rotación.

Para rotar utiliza el método varimax (1), y carece de programa para rotaciones oblicuas por lo que prescindiremos de las mismas.

(1) Este método puede consultarse en Harman (1.967) pag. 304, entre otros muchos.

ANALISIS FACTORIAL 1.

Para el primer análisis factorial, lo que hicimos fue buscar qué variables podían resultar más interesantes, y cuales, a la vista de los datos, parecía que distinguían mejor a nuestra muestra.

Por limitaciones del programa de ordenador solo podíamos introducir 29 variables.

A continuación veremos cuales fueron las variables introducidas y el número que les asignó el ordenador (el ordenador numera las variables de 1 a 29, sin respetar el número anterior, por lo que para la interpretación es necesaria la siguiente equivalencia):

<u>VARIABLE</u>	<u>Nº DE ORDENADOR</u>
3	1
7	2
8	3
10	4
11	5
12	6
13	7
15	8
17	9
18	10

19	11
20	12
21	13
23	14
24	15
25	16
27	17
28	18
29	19
30	20
31	21
32	22
33	23
34	24
35	25
38	26
39	27
41	28
42	29

Los resultados del ordenador están en las hojas de ordenador nº 678 y ss. bajo la denominación de Análisis Factorial Rioja. Podemos ver que las variables no se normalizan. Esto podría tener problemas, como ya comentamos, al dar más peso a unas variables que a otras, pero esto aquí no sucede pues el programa de ordenador calcula luego la matriz de coeficientes de correlación.

Lo que realmente es más problemático es el valor de las puntuaciones en los individuos, que con nuestro programa de ordenador no son válidas al no partir de variables normalizadas, por lo que no las utilizaremos, aunque si pueden servirnos para una primera idea.

Pasemos a la interpretación de los factores antes de la rotación:

En primer lugar tenemos cuatro factores, esto es a causa de que hemos considerado que se desprecian los factores que expliquen menos del 5% de la varianza total. Con esta condición, nos quedamos con los factores más significativos. (1)

El modelo con cuatro factores explica el 41'516% de la varianza total. Esto pudiera parecernos escaso, pero si consideramos que la matriz de coeficientes de correlación tiene valores bastantes bajos, nuestra apreciación variará.

Tengamos en cuenta que en estudios de este tipo es muy difícil rebasar la cifra del 50% de varianza explicado con unos pocos factores. La causa de esto es que al puntuar de uno a siete, el encuestado comete abundantes errores de apreciación y nos encontramos ante un gran subjetivismo.

(1) $0'05 \cdot 29 = 1'45$ Valor propio mínimo superior al propuesto por Kaiser (1.960).

Estos errores, a menudo sistemáticos, hacen que la matriz de coeficientes de correlación tenga valores bajos y en definitiva la varianza explicada sea baja.

Pero esto no debe desanimarnos pues si comparamos muestras resultados con las del estudio de Linden (1) no son mucho peores y esto es aún más importante si consideramos que Linden encuestó a personas que teóricamente deberían ser más coherentes que los individuos que yo encuesté en razón de su posición social, como ya comentamos (2).

Visto esto pasemos a la interpretación:

Distinguimos entre las variables con el siguiente código:

Variables muy importantes:	coeficiente mayor de	0'65
Variables importantes:	" " "	0'50
Variables secundarias:	" " "	0'35
Variables terciarias:	" " "	0'20

Las variables vendrán definidas por su número real en la encuesta, y así tenemos que antes de la rotación:

Factor 1

Valor propio: 6'71218

(1) Linden (1.975)

(2) Por otra parte las variables utilizadas por Linden son bastante reiterativas como se puede observar en su trabajo.

Porcentaje de explicación: 23'145%

VARIABLES MUY IMPORTANTES: 3, 17, 29, 33, 35

(Damos a cada variable su nombre resumido).

- 3 - Edad.
- 17 - Años con Franco buenos.
- 29 - Policía debe ser dura.
- 33 - La mujer en casa.
- 35 - No a la huelga.

VARIABLES IMPORTANTES:

- 7 - Estudios (negativa).
- 11 - Si al divorcio (negativa).
- 15 - Los niños no deben saber del sexo.
- 27 - No a las relaciones prematrimoniales.
- 28 - Las películas españolas buenas.
- 32 - Sí a la misa.
- 38 - La mujer debe trabajar (negativa).

VARIABLES SECUNDARIAS:

- 8 - Lugares que conoce (negativa).
- 10 - Los licenciados están preparados.
- 12 - Sí a la libertad para las regiones (negativa).
- 13 - La gente se lo pasa bien aquí.
- 39 - Mejor la república (negativa).
- 41 - La mujer poco inteligente.
- 42 - Servicio militar voluntario (negativa).

Variables terciarias:

Son: -18, -21, -23, 25, 31, -34.

No las escribimos para no cansar al lector, dada su poca importancia pues el factor está suficientemente definido, y estas variables solo sirven para corroborar a las anteriores como se puede ver analizando su significado.

Este factor es de clarísima interpretación y podría denominarse reaccionario. Podría identificarse con una mentalidad Franquista y tradicionalista, asociada a las personas de más edad.

Interesante es constatar que antes de la rotación el primer factor es el que más diferencia a la población, esto se puede interpretar como que la muestra se diferencia a lo largo de este factor, es decir, por su mayor o menor progresismo.

Factor 2

Valor propio: 2'06324

Porcentaje de explicación: 7'115%

Variables muy importantes:

- 24 - La Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables importantes: (Carece).

Variables secundarias:

23 - Sí a nacionalizar.

Variables terciarias: (Solo citarlas), 19, 21, 34.

El factor está suficientemente claro y habla de la "No Autonomía". Significa que la autonomía también distingue a nuestra muestra. El sentido no quiere decir nada e igual puede salir un factor autonomía que uno No autono-mía, lo importante es donde se sitúen los individuos en el factor. (1)

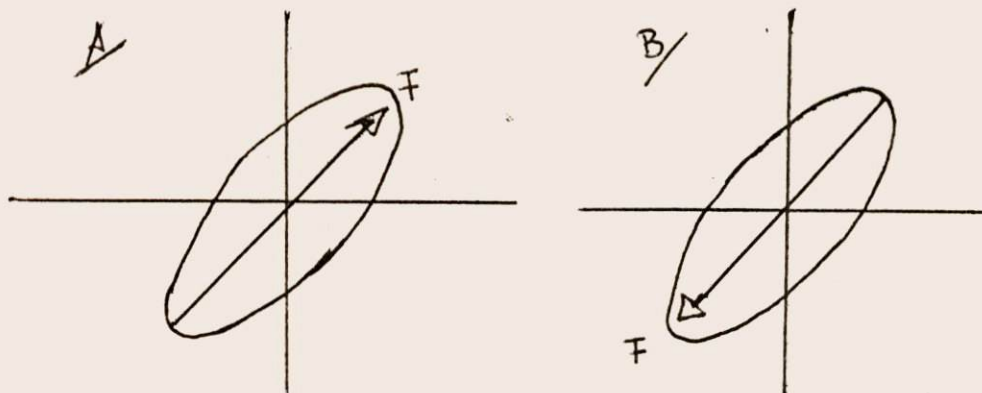
Factor 3

Valor propio: 1'66246

Porcentaje de explicación: 5'73%

Variables muy importantes: (Carece).

(1) El signo de un factor resulta indiferente como podemos ver en el siguiente gráfico:



El factor del gráfico A, es el mismo que el del gráfico B, aunque con signo cambiado. Lo importante es que a lo largo del factor se diferencia la población, el sentido de la misma lo estudiaremos más adelante.

Variables importantes:

- 19 - Sí a las cooperativas.

Variables secundarias:

- 12 - Sí a la libertad para las regiones.
- 18 - Madrid controla demasiado.
- 20 - Los curas viven bien.

Variables terciarias: (Solo citarlas), 10, 13, 21, 23, -24, -25, 28, 30, 34.

Este factor es de difícil interpretación, ya veremos después de rotar si se aclara un poco, en general se identifica con el cooperativismo.

Factor 4

Valor propio: 1'60197

Porcentaje de explicación: 5'524%

Variables muy importantes: (Carece)

Variables importantes:

- 8 - Lugares que conoce (negativa).
- 20 - Los curas viven bien.

Variables secundarias:

- 7 - Estudios (negativa).
- 39 - Mejor la república.

Variables terciarias: (Solo citar), 11, -13, -27, 31, -32, 42.

Este factor, como el anterior, es difícil de definir, lo fundamental pudiera ser incultura con ciertas tendencias inconformistas.

A la vista de estos resultados, no podemos menos que encontrarnos altamente satisfechos, pues en el primer análisis factorial, y aún antes de la rotación los dos factores más explicativos son de clara interpretación.

De todas formas vamos a pasar a la rotación, que nos aclarará las cosas todavía más.

Factores rotados:

Factor 1

Capacidad explicativa: 4'749382 (1)

Porcentaje de explicación: 16'38%

Variables muy importantes:

17 - Años con Franco buenos.

- (1) Como hemos rotado los factores no podemos hablar ya con propiedad de valores propios, en realidad se trata de capacidades explicativas, aunque la suma de ellas tiene que coincidir con la suma de los valores propios.

Variables importantes:

- 3 - Edad.
- 11 - Sí al divorcio (negativa).
- 27 - No a las relaciones prematrimoniales.
- 32 - Sí a la misa.
- 33 - La mujer en casa.
- 39 - Sí a la república (negativa).
- 42 - Servicio militar voluntario (negativa).

Variables secundarias:

- 13 - La gente se lo pasa bien aquí.
- 28 - Las películas españolas buenas.
- 29 - Policía debe ser dura.
- 35 - No a la huelga.
- 38 - La mujer debe trabajar fuera (negativa).
- 41 - La mujer es poco inteligente.

El factor ha bajado un poco en cuanto a su capacidad explicativa respecto al modelo previo a la rotación. Sin embargo sigue siendo altamente significativo y su interpretación es la misma, pudiendo asociarlo a la sociología Franquista-Tradicionalista, con connotaciones de edad madura.

Factor 2

Capacidad explicativa: 2'026892

Porcentaje de explicación: 6'99%

Variables muy importantes:

- 24 - La Rioja unida a otras.

30 - Rioja autónoma (negativa).

Variables importantes: (Carece).

Variables secundarias: (Carece).

Variables terciarias: (Solo citar), -12, 23.

El factor es el mismo que antes de rotar, quizá un po
co más claro, pero el mismo en esencia. Es la No autonomía.

Factor 3

Capacidad explicativa: 2'369801

Porcentaje de explicación: 8'17%

Variables muy importantes:

19 - Sí a las cooperativas.

Variables importantes:

12 - Sí a la libertad para las regiones.

Variables secundarias:

18 - Madrid controla demasiado.

21 - Los niños deben opinar.

23 - Sí a nacionalizar.

25 - Las cooperativas no funcionan (negativa).

34 - Más actos culturales.

Variables terciarias: (Solo citar), -3, 7, -15, 20, -29,

Este factor es bastante más claro de interpretar, es partidario de las cooperativas, y también algo de las nacionalizaciones, como también lo es de mayor libertad regional etc. Lo podíamos denominar social - progresista.

Factor 4

Capacidad explicativa: 2'892966

Porcentaje de explicación: 9'97%

Variables muy importantes: (Carece).

Variables importantes:

- 7 - Estudios (negativa).
- 8 - Lugares que conoce (negativa).
- 20 - Los curas viven bien.
- 28 - Las películas españolas buenas.

Variables secundarias:

- 10 - Los licenciados están preparados.
- 15 - Los niños no deben saber del sexo.
- 31 - Dinero del campo para la industria.
- 33 - La mujer en casa.
- 35 - No a la huelga.
- 41 - La mujer es poco inteligente.

Variables terciarias: (Solo citar), 3, 17, 29, -38.

Este factor varía mucho en cuanto a su contenido, al que tenía antes de la rotación. Es el único que varía sustancialmente.

Su interpretación no es fácil, podemos decir que lo fundamental es la incultura, y luego, una serie de indicadores de esa incultura.

En conclusión, tenemos que el análisis factorial ha resultado bastante bien, han salido cuatro factores, y estos son en general bastante claros. Sin embargo, hemos de realizar más pruebas, para lograr un modelo más significativo y claro.

Las puntuaciones, no las interpretamos, pues como ya comentamos se han hecho a partir de las variables sin normalizar, y en consecuencia, no son significativas dado el programa de nuestro ordenador, de todas formas están en las hojas de ordenador, pues como las varianzas son similares y los promedios también (en las variables) pueden tener algún interés.

ANALISIS FACTORIAL 2.

Las variables introducidas variaron de la siguiente manera:

Se introdujeron la 5 (profesión), 6 (clase social) - por considerarlas interesantes a la vista de los resultados anteriores.

Se quitaron las variables 13 (la gente se lo pasa bien aquí), 21 (los niños deben opinar), 25 (las cooperativas no funcionan), 31 (dinero del campo para la industria), 34 (más actos culturales). A causa de ser poco significativas o de difícil interpretación como ya hemos visto en el estudio anterior.

Los números equivalentes de ordenador quedan:

<u>VARIABLE</u>	<u>Nº DE ORDENADOR</u>
3	1
5	2
6	3
7	4
8	5
10	6
11	7
12	8
15	9

17	10
18	11
19	12
20	13
23	14
24	15
27	16
28	17
29	18
30	19
32	20
33	21
35	22
38	23
39	24
41	25
42	26

Los resultados de ordenador están en la hoja de ordenador nº 695 y siguientes, bajo la denominación de Análisis Factorial Rioja 2.

Las condiciones del programa de ordenador y de la entrada de datos son los mismos que el Análisis Factorial Rioja 1, es decir, rechazo de factores que expliquen menos del 5%, entrada de variables sin normalizar, etc.

El estudio, con cuatro factores explica el 45'915%,

luego hemos mejorado algo, aunque disminuyendo algo el número de variables. Pero si con las variables actuales definimos la mentalidad y la opinión, realmente hemos mejorado.

Comencemos la interpretación antes de la rotación.

Factor 1

Valor propio: 6'7311

Porcentaje de explicación: 25'888%

Variables muy importantes:

- 3 - Edad.
- 17 - Años con Franco buenos.
- 33 - La mujer en casa.
- 35 - No a la huelga.

Variables importantes:

- 5 - Profesión (negativa).
- 7 - Estudios (negativa).
- 11 - Sí al divorcio (negativa).
- 15 - Los niños no deben saber del sexo.
- 27 - No a las relaciones prematrimoniales.
- 28 - Las películas españolas buenas.
- 29 - Policía debe ser dura.
- 32 - Sí a la misa.
- 38 - La mujer debe trabajar fuera (negativa).
- 41 - La mujer es poco inteligente.

Variables secundarias:

- 8 - Lugares que conoce (negativa).
- 10 - Los licenciados están preparados.
- 12 - Sí a la libertad para las regiones (negativa).
- 39 - Mejor la república (negativa).
- 42 - Servicio militar voluntario (negativa).

Variables terciarias: (Solo citar), -6, -18.

El factor es de muy clara interpretación, aun antes de la rotación. Es muy similar al que vimos en el primer análisis factorial, pero a causa de las variables que hemos introducido, hay que tener en cuenta que el tradicionalismo se ha identificado, ya no solo con la edad madura, sino también con el bajo nivel cultural. Respecto al bajo nivel profesional, hay que tener en cuenta que al estar los jubilados en el punto más bajo de la escala, y al tener connotaciones de edad madura este factor, una cosa puede condicionar a la otra (1). Pero de lo que no cabe duda es del bajo nivel cultural, no nos referimos a términos absolutos, sino lo que es más grave dado el bajo nivel cultural de la Rioja, al nivel relativo.

(1) De hecho el coeficiente de correlación entre estas dos variables es del -0'45.

Factor 2

Valor propio: 1'99505

Porcentaje de explicación: 7'674%

Variables muy importantes:

24 - La Rioja unida a otras (negativa).

30 - Rioja autónoma.

Variables importantes: (Carece)

Variables secundarias:

23 - Sí a nacionalizar (negativa).

El factor está muy claro, significa autonomía, es de de cir, lo contrario a su significado en el análisis precedente. Esto nos confirma que el signo de un factor no quiere decir nada.

También se puede observar, como en el precedente, una correlación negativa entre la autonomía y las nacionalizaciones. No es muy grande, por lo que no la consideramos importante.

Factor 3

Valor propio: 1'75271

Porcentaje de explicación: 6'741%

Variables muy importantes: (Carece)

Variables importantes:

- 6 - Clase social.
- 20 - Los curas viven bien (negativa).

Variables secundarias:

- 7 - Estudios.
- 8 - Lugares que conoce.
- 39 - Mejor la república (negativa).

Variables terciarias (Solo citar): 5, -10, -19, 27,
-28, 32, -42.

Este factor es ya de difícil interpretación, puede identificarse con la burguesía, con ciertas connotaciones clericales, que se aumentan a la vista de las terciarias.

Factor 4

Valor propio: 1'45918

Porcentaje de explicación: 5'612%

Variables muy importantes: (Carece)

Variables importantes:

- 12 - Sí a la libertad para las regiones.
- 19 - Sí a las cooperativas.

Variables secundarias:

- 18 - Madrid controla demasiado.
- 23 - Sí a nacionalizar.

Variables terciarias (Solo citar): 7, 8, -11, 27, -39.

Este factor también es muy difícil de interpretar, pero parece asociado a un antientralismo, e incluso a un anticapitalismo (sí a las cooperativas y a las nacionalizaciones de grandes empresas).

En conclusión de este estudio, podemos decir que el factor uno, explica más que entre todos los demás, luego si gue siendo el factor que discrimina realmente a la muestra que estamos estudiando.

Por lo demás, parece ser que la situación ha mejorado respecto al otro análisis. De todas formas quizá tras la ro tación la interpretación esté todavía más clara.

Factores rotados:

Factor 1

Capacidad explicativa: 5'153995

Porcentaje de explicación: 19'82%

Variables muy importantes:

- 11 - Sí al divorcio (negativa).
- 17 - Los años con Franco buenos.
- 27 - No a las relaciones prematrimoniales.
- 32 - Sí a la misa.

Variables importantes:

- 3 - Edad.
- 29 - Policía debe ser dura.
- 33 - La mujer en casa.
- 35 - No a la huelga.
- 38 - La mujer debe trabajar fuera (negativa).
- 39 - Mejor la república (negativa).
- 42 - Servicio militar voluntario (negativa).

Variables secundarias:

- 5 - Profesión (negativa).
- 10 - Los licenciados están preparados.
- 15 - Los niños no deben saber del sexo.
- 28 - Las películas españolas buenas.
- 41 - Las mujeres son poco inteligentes.

Variables terciarias (Solo citar): -7, -12.

El factor sigue siendo de tradicionalismo, pero ha haciendo más énfasis en variables de contenido religioso, y en cierta forma contra la liberación de la mujer. Sin embargo, ha perdido peso la edad y todavía más el nivel cultural.

Factor 2

Capacidad explicativa: 2'019399

Porcentaje de explicación: 7'77%

Variables muy importantes:

- 24 - La Rioja unida a otras (negativa).
- 30 - Rioja autónoma.

Variables importantes: (Carece)

Variables secundarias: (Carece)

Variables terciarias (Solo citar): -23.

El factor se ha hecho clarísimo, es "autonomía". Quiero hacer constar aquí que la alta correlación negativa entre las variables 24 y 30 no debe hacer pensar al lector que son la misma variable, pero de signo contrario, son en realidad variables semejantes, pero llenas de matices diferenciadores, que se hacen patentes en los distintos grados de autonomía.

Factor 3

Capacidad explicativa: 2'908199

Porcentaje de explicación: 11'18%

Variables muy importantes:

- 7 - Estudios.

Variables importantes:

- 5 - Profesión.
- 6 - Clase social.
- 8 - Lugares que conoce.
- 20 - Los curas viven bien (negativa).

Variables secundarias:

- 10 - Los licenciados están preparados (negativa).
- 28 - Las películas españolas buenas (negativa).
- 30 - La mujer en casa (negativa).

Variables terciarias (Solo citar): -3, -15, -17, -29,
-35, 38, -41.

Este factor está ya muy claro y lo podíamos denominar burguesía culta.

Factor 4

Capacidad explicativa: 1'855637

Porcentaje de explicación: 7'14%

Variables muy importantes:

- 19 - Sí a las cooperativas.

Variables importantes:

- 12 - Sí a la libertad para las regiones.
- 18 - Madrid controla demasiado.

Variables secundarias:

- 20 - Los curas viven bien.
- 23 - Sí a las nacionalizaciones.
- 29 - La policía debe ser dura (negativa).

Variables terciarias (solo citar): -3, 7, -15.

Este factor sigue siendo de difícil interpretación, de todas formas se ve un marcado carácter cooperativista, completado con un regionalismo.

La conclusión de este análisis es que los factores son bastante claros y en general muy coherentes, pero toda vía hace falta el calibrar un poco más las preguntas a introducir.

Las puntuaciones no las interpretamos por las razones ya expuestas anteriormente.

Por todo lo dicho, pasamos a un siguiente análisis factorial, donde esperamos mejorar aún más nuestros resultados.

ANALISIS FACTORIAL 3.

En este análisis factorial hemos introducido varias modificaciones, como son:

En primer lugar hemos normalizado las variables, con lo que las puntuaciones ya van a ser interpretables.

En segundo lugar, solo nos vamos a referir a los 100 individuos muestreados en Logroño capital. El objetivo de esta reducción es analizar separadamente la muestra de Logroño, y el poder hacer aplicaciones estadísticas con las puntuaciones. Si mantuviéramos los 267 datos iniciales, dada la capacidad de nuestro ordenador, resultaría imposible realizar ciertos estudios que luego veremos.

También hemos modificado las variables, suprimiendo la variable edad, dado que introducía en el factor uno, una connotación de madurez, y nos interesa ver como queda ahora sin esa variable.

Los números equivalentes de ordenador, quedan:

<u>VARIABLE</u>	<u>Nº DE ORDENADOR</u>
5	1
6	2
7	3

8	4
10	5
11	6
12	7
15	8
17	9
18	10
19	11
20	12
23	13
24	14
27	15
28	16
29	17
30	18
32	19
33	20
35	21
38	22
39	23
41	24
42	25

Los resultados de ordenador están en la hoja de ordenador nº 712 y siguientes, bajo la denominación de Análisis Factorial Rioja 3.

Las condiciones del programa de ordenador, salvo las puntualizaciones ya hechas, siguen siendo las mismas.

El estudio con cuatro factores, explica el 51'019% de la varianza total, lo que debe interpretarse como todo un éxito, pues rebasar la frontera del 50% en estudios de este estilo es realmente difícil.

La explicación de este incremento creo que no se debe a la supresión de la variable 3, sino que su causa fundamental bien puede ser el tratarse de un estudio circunscrito a Logroño capital. Tengamos en cuenta que la gente de la capital normalmente vive más enterada de los problemas y tiene formada una opinión más sólida, lo que se traduce en una mayor coherencia en las respuestas, que la conseguida en tre la gente de los pueblos.

Pasemos ya directamente a la interpretación de los factores antes de la rotación:

Factor 1

Valor propio: 6'81446

Porcentaje de explicación: 27'257%

Variables muy importantes:

- 7 - Estudios.
- 11 - Sí al divorcio.
- 17 - Años con Franco buenos (negativa).
- 33 - La mujer en casa (negativa).
- 35 - No a la huelga (negativa).

Variables importantes:

- 5 - Profesión.
- 15 - Los niños no deben saber del sexo (negativa).
- 27 - No a las relaciones prematrimoniales (negativa).
- 28 - Las películas españolas buenas (negativa).
- 29 - La policía debe ser dura (negativa).
- 32 - Sí a la misa (negativa).
- 38 - La mujer debe trabajar fuera.
- 41 - La mujer es poco inteligente (negativa).
- 42 - El servicio militar voluntario.

Variables secundarias:

- 10 - Los licenciados están preparados (negativa).
- 12 - Sí a la libertad para las regiones.
- 18 - Madrid controla demasiado.
- 39 - La república es mejor.

Variables terciarias (solo citar): 6, 8, 19, 23, -24, 30.

El factor está muy claro y es el contrario al que veíamos en los análisis precedentes. Recordamos que el signo es arbitrario, y lo podemos denominar como factor liberal-progresista.

Este es nuevamente el factor más explicativo, luego - tanto con la muestra a nivel provincial, como a nivel simplemente de la capital, el tradicionalismo o progresismo es lo que más diferencia a la población.

Este factor, como es lógico, al haber suprimido la va

riable 3, ya no tiene la connotación de edad. Por otro lado, sigue teniendo una fuerte influencia cultural.

Factor 2

Valor propio: 2'35764

Porcentaje de explicación: 9'431%

Variables muy importantes: (Carece)

Variables importantes:

- 20 - Los curas viven bien.
- 24 - La Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables secundarias:

- 6 - Clase social (negativa).
- 7 - Estudios (negativa).
- 8 - Lugares que conoce (negativa).
- 19 - Sí a las cooperativas.
- 39 - Mejor una república.

Variables terciarias (solo citar): -5, 10, 11, 23, -27, 28.
-17.

Este es un factor que esperamos se aclare en la rotación, pues ahora es de difícil interpretación, podemos decir que tiene una tendencia antiautonomista y bajo nivel cultural a parte de un cierto inconformismo.

Factor 3

Valor propio: 1'96286

Porcentaje de explicación: 7'851%

Variables muy importantes:

- 24 - La Rioja unida a otras.
- 30 - La Rioja autónoma (negativa).

Variables importantes:

- 20 - Los curas viven bien (negativa).

Variables secundarias:

- 12 - Sí a la libertad para las regiones (negativa).

Variables terciarias (Solo citar): 5, 6, 7, 8, 23.

Factor no excesivamente claro, corresponde en lo esencial al factor antiautonomista que en los estudios precedentes era el segundo y aquí ha pasado a tercer lugar, pero se ha mezclado con variables extrañas que dificultan su clasificación.

Factor 4

Valor propio: 1'61999

Porcentaje de explicación: 6'48%

Variables muy importantes: (Carece)

Variables importantes:

- 19 - Sí a las cooperativas.

Variables secundarias:

- 12 - Sí a la libertad para las regiones.
- 15 - Los niños no deben saber nada del sexo (negativa).
- 23 - Sí a las nacionalizaciones.
- 27 - No a las relaciones prematrimoniales.
- 39 - Mejor una república (negativa).

Variables terciarias (Solo citar): 5, 7, 10, -11, 17,
18, 32.

Nuevamente nos encontramos ante un factor poco claro. Realmente hay un importante cooperativismo, pero mezclado con cierto liberalismo por un lado y tradicionalismo por otro. Es peremos que se aclare en la rotación.

A la vista de estos resultados, quizá sea este el caso en que la rotación es más necesaria, pues a excepción del factor uno, los demás están bastante confusos. Vamos a ver - los resultados tras haber rotado, para comprobar si se aclaran las cosas.

Factores rotados:

Factor 1

Capacidad explicativa: 4'8701

Porcentaje de explicación: 19'48%

Variables muy importantes:

- 11 - Sí al divorcio.
- 17 - Los años con Franco buenos (*negativa*).
- 27 - No a las relaciones prematrimoniales (negativa).
- 35 - No a la huelga (negativa).

Variables importantes:

- 32 - Sí a la misa (negativa).
- 33 - La mujer en casa (negativa).
- 38 - La mujer debe trabajar fuera.
- 39 - Mejor una república.
- 42 - Servicio militar voluntario.

Variables secundarias:

- 10 - Los licenciados están preparados (negativa).
- 28 - Las películas españolas buenas (negativa).
- 29 - La policía debe ser dura (negativa).

Variables terciarias (Solo citar): 5, 7, -15, 18, 20, -41.

Este factor ha quedado completamente definido. Respecto a su contenido antes de rotar, podemos observar que han desaparecido bastante las connotaciones culturales, con lo que podemos considerarlo como plenamente liberal-progresista. Es este un factor plenamente coherente y del que podemos sacar mucho partido.

Factor 2

Capacidad explicativa: 3'0608

Porcentaje de explicación: 12'24%

Variables muy importantes:

- 6 - Clase social (negativa).
- 7 - Estudios (negativa).
- 20 - Los curas viven bien.

Variables importantes:

- 5 - Profesión (negativa).
- 8 - Lugares que conoce (negativa).
- 28 - Las películas españolas buenas.

Variables secundarias:

- 10 - Los licenciados están preparados.

Variables terciarias (Solo citar): 33, 35, 41, -42.

Este factor es ya bastante claro y se puede definir como estatus social bajo, con incultura. En definitiva, es un factor que incluye todas las variables asociadas con un estatus bajo y lo podemos denominar incultura en general.

Factor 3

Capacidad explicativa: 2'2040

Porcentaje de explicación: 8'82%

Variables muy importantes:

- 24 - La Rioja unida.
- 30 - Rioja autónoma (negativa).

Variables importantes: (Carece)

Variables secundarias:

- 12 - Sí a la libertad para las regiones (negativa).

Variables terciarias (Solo citar): 23, 39, -42.

Este factor está ya muy claro también y es el factor de la no autonomía.

Factor 4

Capacidad explicativa: 2'6193

Porcentaje de explicación: 10'48%

Variables muy importantes:

- 19 - Sí a las cooperativas.

Variables importantes:

- 12 - Sí a la libertad para las regiones.
- 15 - Los niños no deben saber del sexo (negativa).
- 18 - Madrid controla demasiado.
- 23 - Sí a las nacionalizaciones.

Variabes secundarias:

- 7 - Estudios.

29 - La policía debe ser dura (negativa).

Variables terciarias (Solo citar): 5, -17, -33, -35,
38, -41.

Este factor está quizá menos claro que los anteriores, pero si lo analizamos detenidamente, vemos que aparte de su importante connotación cooperativista, hay un sentimiento regionalista y algo liberal, a parte de ser favorable a las nacionalizaciones. Podríamos definirlo como opuesto al capitalismo y al centralismo, o simplemente, como cooperativista-regionalista, que puede ser más concreto y explicativo.

A la vista de los resultados de la rotación, creo que debemos darnos por satisfechos. En los dos primeros análisis factoriales hemos ido haciendo pruebas y creo que en este - tercero hemos conseguido el resultado apetecido, luego podemos pasar a la interpretación de las puntuaciones y a la utilización del análisis factorial para sacar consecuencias sobre la muestra.

En primer lugar, vamos a resumir los resultados de la rotación citando a los factores por orden en su capacidad explicativa:

<u>Factor</u>	<u>Porcentaje de explicación</u>	<u>Denominación</u>
1	19'48%	Liberal-progresista
2	12'24%	Incultura
4	10'48%	Cooperativista-regionalista
3	8'82%	Antiautonomista

La denominación no pretende incluir todo lo que quiere decir el factor, sino hacer un resumen para poder utilizarlo con mayor facilidad.

A la hora de interpretar las puntuaciones (que ahora si son significativas dado que hemos introducido las variables normalizadas), nos encontramos con el problema de que los factores también están normalizados, es decir, que tienen promedio cero y desviación típica uno, con lo cual nos dicen bastante poco de la situación de nuestra muestra.

Por ejemplo, si tomamos el factor uno, liberal-progresista, al estar normalizadas las puntuaciones, sabremos muy poco sobre si nuestra muestra es liberal o reaccionaria.

Para resolver este problema se me ocurrió la creación de un individuo que denominaré "individuo indiferente", que lo que hacía era contestar a todas las preguntas como indiferente. Es decir, supuesto un individuo indiferente en todas las preguntas, mediante la ecuación de obtención de las puntuaciones, veremos cual es su valor en cada factor, y así - tendremos un elemento de comparación.

Se nos plantea en primer lugar el problema de las preguntas:

- 5 - Profesión.
- 6 - Clase social.
- 7 - Estudios.
- 8 - Lugares que conoce.

que al ser de valor relativo, y sin un contenido absoluto, no tiene sentido el individuo indiferente a causa de esto, estas preguntas no las tendremos en cuenta para el cálculo del individuo citado.

Por esta misma razón, el factor dos, incultura, no lo vamos a estudiar respecto al individuo indiferente, pues éste no tiene sentido en tal factor.

En segundo lugar será preciso normalizar las respuestas del individuo indiferente, es decir, restarles el promedio y dividir por la desviación típica de cada variable, para así poder calcular el valor del factor, ya que todas las variables han entrado normalizadas en este análisis.

Comencemos pues, con la creación de este individuo:

<u>Variable</u> (según nº de orden)	<u>Contestación</u>	<u>Normalización</u>
1	(no tiene sentido)	
2	(no tiene sentido)	
3	(no tiene sentido)	
4	(no tiene sentido)	
5	4 . .	-0'53
6	4 . .	0'00
7	4 . .	-0'99
8	4 . .	0'66
9	4 . .	-0'34
10	4 . .	-1'34

11	4	.	.	-0'86
12	4	.	.	-0'31
13	4	.	.	-0'66
14	4	.	.	0'22
15	4	.	.	-0'72
16	4	.	.	0'72
17	4	.	.	0'23
18	4	.	.	-0'36
19	4	.	.	-0'79
20	4	.	.	-0'02
21	4	.	.	-0'02
22	4	.	.	-0'28
23	4	.	.	0'27
24	4	.	.	0'58
25	4	.	.	0'11

Con estos valores, introduciéndolos en la ecuación que da las puntuaciones en los factores, tal como viene - en las hojas de ordenador tenemos:

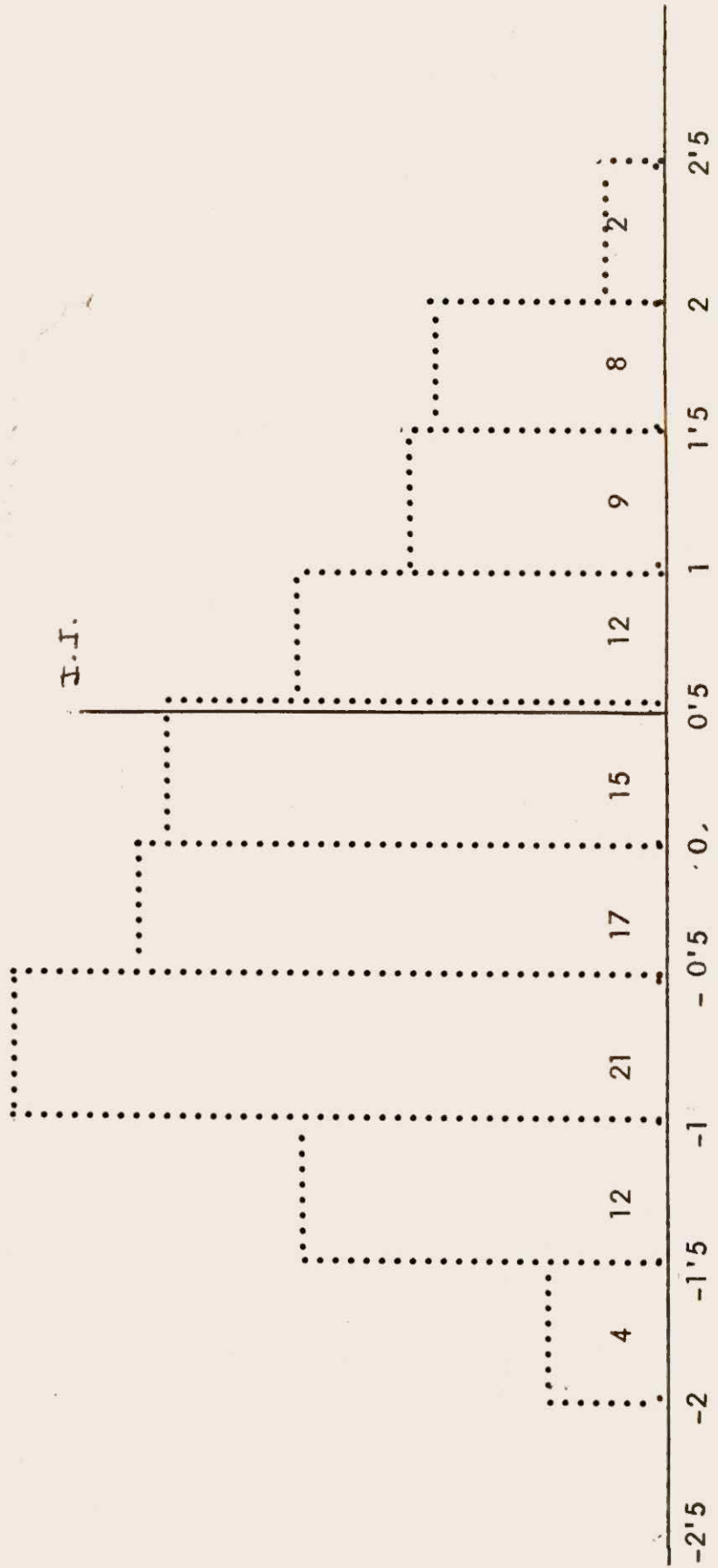
<u>Factor</u>	<u>Valor del individuo indiferente</u>
1	0'49
2	no tiene sentido
3	0'48
4	-1'25

Como el ordenador no nos ha proporcionado los histogramas de los factores los obtendremos normalmente, a par

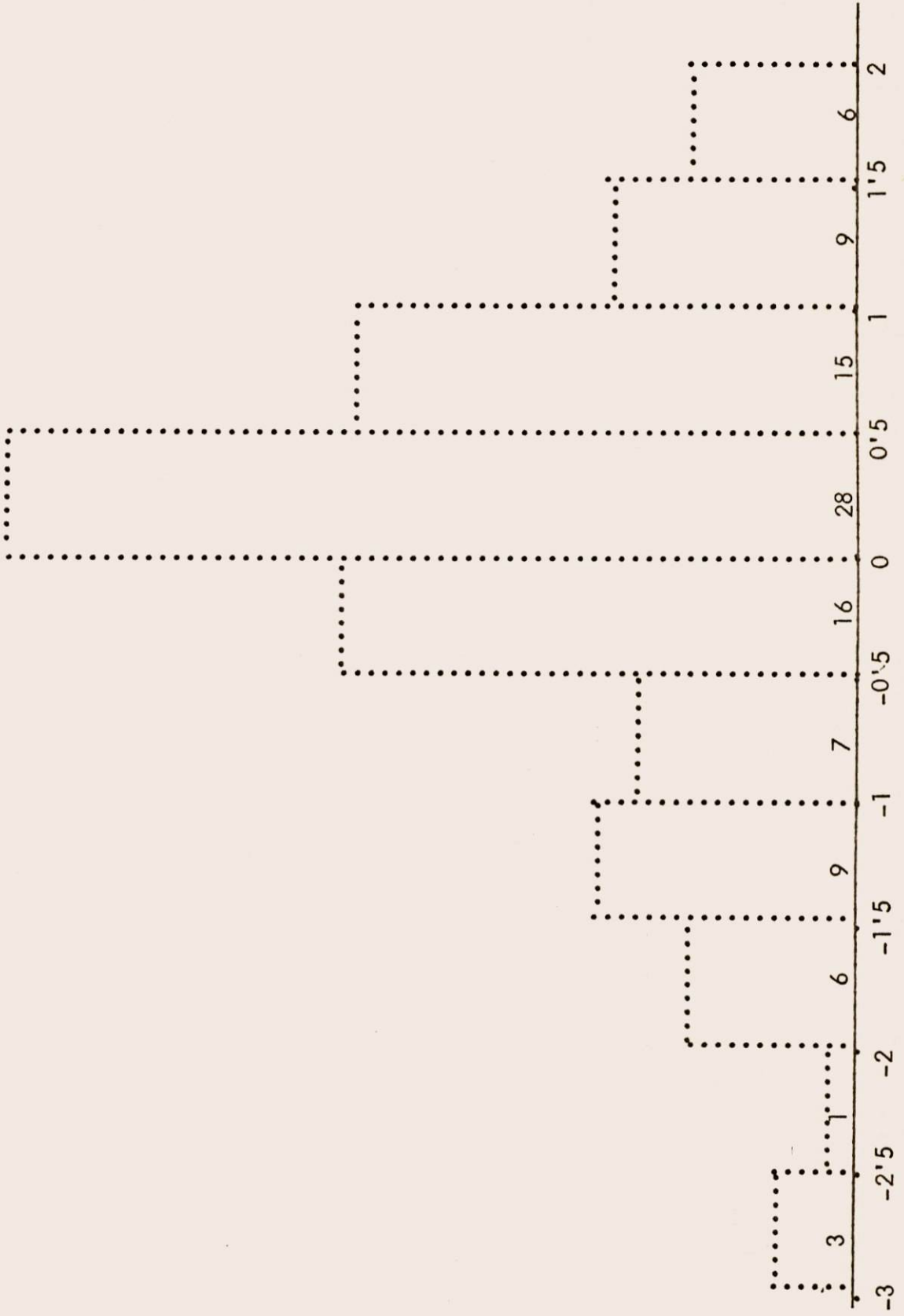
tir de las puntuaciones:

Veámoslas a continuación, marcado ya el individuo
indiferente:

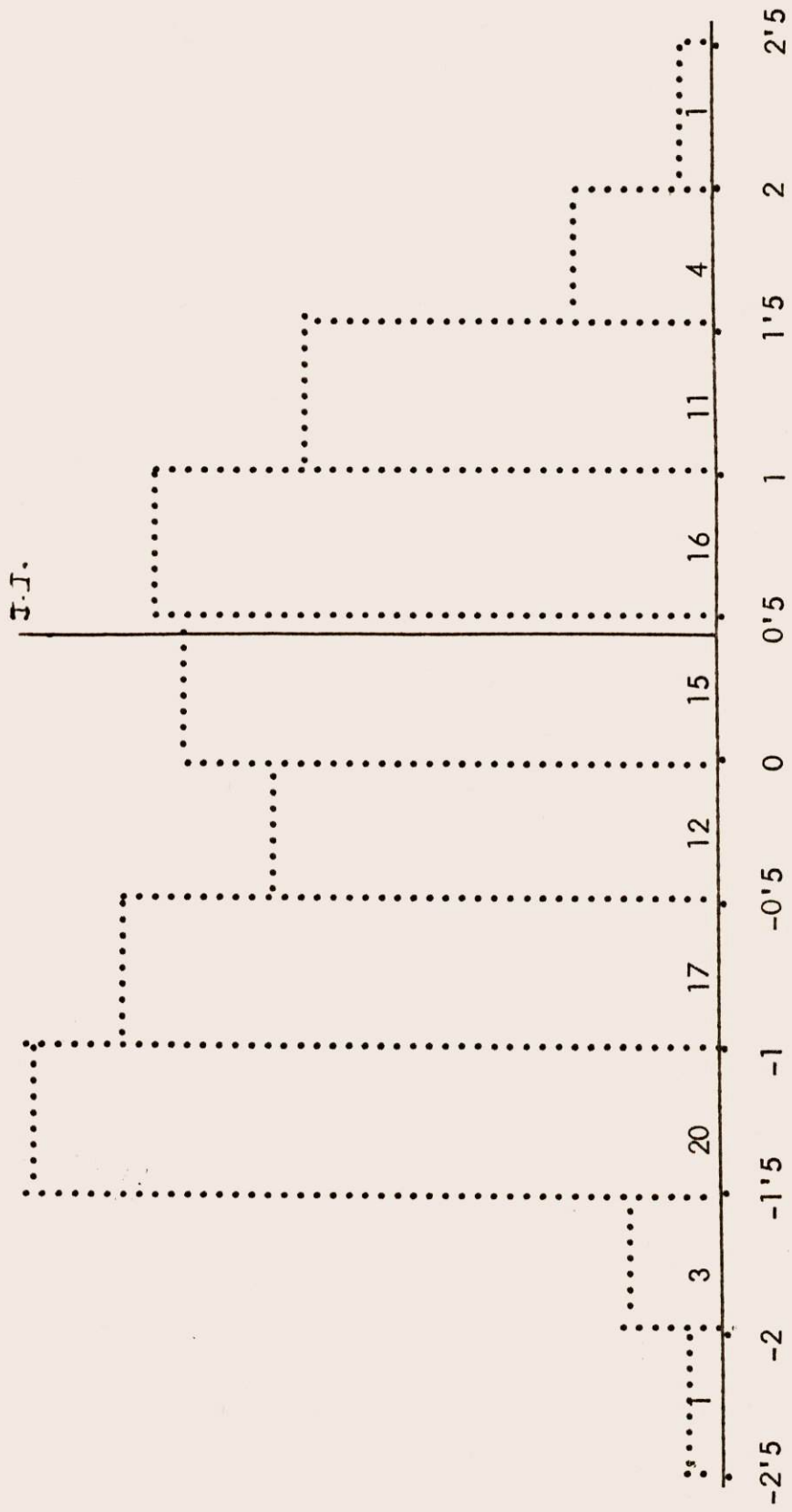
F₁



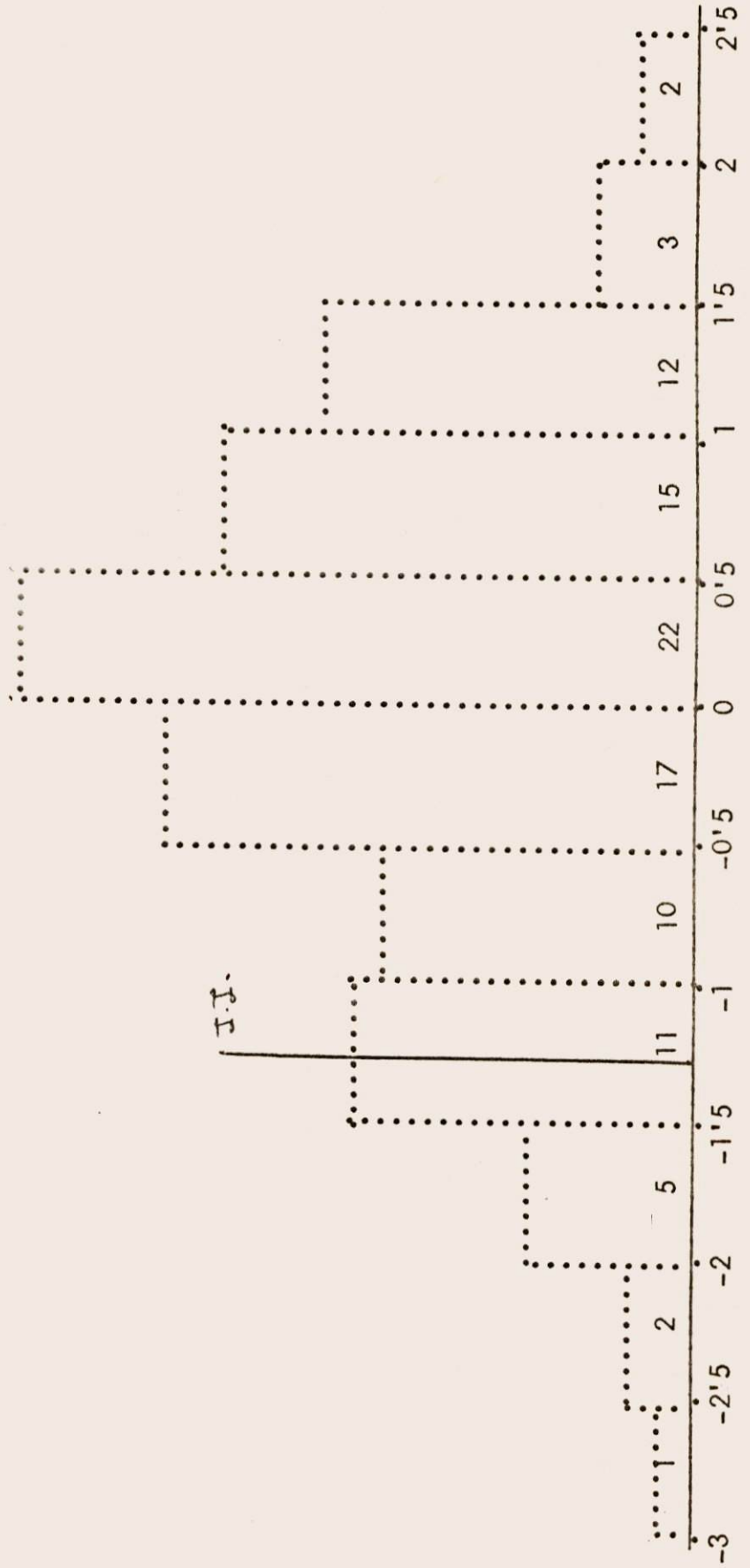
F_2



F₃



F₄



Vamos a comentar los histogramas:

Factor 1

Este factor es el denominado liberal-progresista, el individuo indiferente toma el valor 0'49.

En primer lugar, a la vista del histograma y de la situación del individuo indiferente (I.I.), vemos que éste resultaría relativamente liberal-progresista en nuestra muestra, luego en consecuencia ésta tiene una tendencia tradicionalista.

Por otro lado, la distribución resulta asimétrica con cola a la derecha, y por lo tanto, hay un grupo numeroso, fuertemente tradicionalista, y otro, menos numeroso cuanto mayor es el grado de liberalidad, a gran distancia ideológica del primero.

Lógicamente este estudio se podría completar, ajustando la distribución de la muestra a una conocida y realizando estudios de pruebas de hipótesis, etc. A parte sería interesante estudiar analíticamente los parámetros de asimetría, curtosis y otros estudios estadísticos más o menos complejos. El problema es que nos alargaríamos demasiado, saliéndonos de los planteamientos del presente estudio, pero la puerta queda abierta para investigaciones posteriores.

Factor 2

Este es el factor de incultura, y el I.I. no toma valor aquí por no tener sentido tal y como hemos comentado.

De todas formas puede tener interés la observación - del histograma. En primer lugar, hagamos notar que dadas las premisas de nuestro estudio, la cultura o incultura resulta relativa.

La distribución apunta una pequeña tendencia bimodal y una importante asimetría a la izquierda. Esto quiere decir que hay dos grupos, uno muy numeroso relativamente inculto y un pequeño grupo relativamente culto.

Factor 3

Este es el factor denominado antiautonomista, en el cual el I.I. toma el valor 0'48.

Por lo tanto si el I.I. es relativamente antiautonomista respecto a la muestra. Significa que ésta es de tendencia autonomista.

Por la forma de la distribución, bastante bimodal, vemos que hay dos grupos, uno autonomista y otro cercano al individuo indiferente. En conclusión, nuestra muestra es mayoritariamente autonomista, con un importante grupo

indiferente y muy escasa gente en contra.

Factor 4

Este factor es el denominado cooperativista-regionalista.

El valor de I.I. es de -1'25. Luego la situación es bastante clara. Vemos que el I.I. es, en el plano relativo, claramente anticooperativista-regionalista, luego la muestra es de tendencia favorable al cooperativismo-regionalismo.

En conclusión hemos analizado los cuatro factores, y excepto en el segundo, por las razones expuestas, el individuo indiferente se ha mostrado como un poderoso instrumento de análisis que salva mucha de la indeterminación inherente a los factores a causa de resultados normalizados.

Contraste del Individuo Indiferente.

Lo que pretendemos ahora es comprobar que los resultados obtenidos mediante la utilización del individuo indiferente son coherentes con los valores tomados por las variables. Para ello estudiaremos los promedios y desviaciones típicas de las variables más influyentes en cada factor, acompañadas de su signo correspondiente:

Factor 1

Variables más influyentes:

- 11 - Sí al divorcio.
- 17 - Años con Franco buenos (negativa).
- 27 - No a las relaciones prematrimoniales (negativa).
- 35 - No a la huelga (negativa).

<u>Variables</u>	<u>Promedio</u>	<u>Desviación Típica</u>
11	3'99	2'40
-17	4'62	1'81
-27	5'57	2'18
-35	4'06	2'39

En las variables 11 y 35 estamos prácticamente en el promedio. Pero la 17 nos da algo de Franquismo y la 27 nos da claramente contra las relaciones prematrimoniales.

Luego en conclusión, los resultados del individuo in diferente son coherentes dado el resultado de un cierto tradicionalismo.

Creo que es importante el hacer notar aquí que todo el estudio se basa en considerar la contestación de indiferencia, es decir, puntuar un 4, como punto de comparación. Evidentemente aquí entramos en un juicio de valor de difícil solución. Parece claro que para muchos, habrá preguntas polémicas como la que hace referencia al divorcio, o a las relaciones prematrimoniales, o tantas otras, donde según

la apreciación de cada cual, el punto de comparación no de be ser el indiferente, pues la indiferencia puede considerarse como una cierta toma de posición.

Soy consciente de lo difícil del problema y como ya comenté al principio de esta segunda parte, es mi intención ser lo más imparcial que pueda, dejando los juicios de va lor para las personas competentes (moralistas, políticos, etc.), y limitándome a lo que es mi estudio, es decir estu dio cuantitativo.

El lector puede argüir que en ciertos casos la imparcialidad es imposible, incluso podríamos citar la frase evan gélica de "Todo el que no está conmigo está contra mí", haciendo referencia a casos en que la indiferencia ya significa una toma de posición.

Pero aun con todo, creo que la única solución es con siderar la indiferencia como punto de comparación, siendo - consciente de sus limitaciones, para no incurrir en el error aun peor, de que el cuantitativista refleje sus opiniones - personales en el estudio.

Creo que esta pequeña discusión puede aclarar los - problemas de la utilización del individuo indiferente, y a la vez aumentar su interés dado que resulta necesario.

Factor 2

Variables más influyentes:

- 6 - Clase social (negativa).
- 7 - Estudios (negativa).
- 20 - Los curas viven bien.

<u>Variables</u>	<u>Promedio</u>	<u>Desviación Típica</u>
-6	2'31	0'83
-7	2'98	1'70
20	4'59	1'91

Aquí no podemos hablar de individuo indiferente, pero lo que si podemos hacer viendo los valores de las variables es deducir que los estudios y la clase social son medios, sin ser excepcionales ni en un sentido ni en otro, en comparación con el resto de España.

Factor 3

Variables más influyentes:

- 24 - Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

<u>Variables</u>	<u>Promedio</u>	<u>Desviación Típica</u>
24	3'55	2'02
-30	4'71	1'99

Se ve que, confirmando las apreciaciones del individuo indiferente, el promedio está con la autonomía.

Factor 4

Variables más influyentes:

19 - Sí a las cooperativas.

<u>Variables</u>	<u>Promedio</u>	<u>Desviación Típica</u>
19	5'18	1'37

Luego nuestra muestra está por el cooperativismo.

En conclusión, el individuo indiferente, como técnica de comparación se ha manifestado como coherente y hemos comentado sus limitaciones.

ANALISIS DE NUESTRAS VARIABLES.

A la vista de lo estudiado hasta el momento, puede ser interesante el contar con un análisis más detallado de los valores de nuestras variables.

Para ello puede consultarse la hoja de ordenador nº 726 y siguientes, donde tenemos las frecuencias de todas las variables, según el número de la variable en la encuesta y de acuerdo con la codificación que ya citamos, en cada intervalo, y - hecho esto tanto para Logroño capital como para toda la muestra.

Con todos estos datos podemos ver, en primer lugar, que el individuo indiferente sigue resultando coherente. Esto lo - veríamos estudiando variable por variable, pero no lo escribo aquí para que no resulte excesivamente pesado.

En segundo lugar, vemos la dificultad de realizar pruebas de hiótesis y estudios similares en nuestras variables - pues son muchas de carácter bimodal, a parte de ser discretas. Además todas son difíciles de ajustar a una distribución conocida.

CLUSTER ANALISIS.

El cluster análisis, o análisis de conglomerados, ya comentamos en la primera parte que pretende formar grupos - de carácter homogéneo.

Lo que vamos a hacer aquí es agrupar a los individuos de la muestra, para ver que conclusiones podemos sacar de los citados agrupamientos.

Es evidente, y ya lo citamos, que hay innumerables sis temas de agrupación. La mayoría de ellos muy caprichosos, y nos decantamos por el método matricial como método científico y coherente con nuestro modelo.

El problema es que el ordenador con el cuento carece de programa para la aplicación del método matricial y el lec tor puede comprender lo laborioso de la creación de un progra ma dado lo complejo del método.

Ante esta situación y dado que contaba con programas de encadenamiento, decidí utilizar al método de encadenamiento completo, que se basa en lo siguiente: (1)

(1) Puede verse en cualquier tratado de cluster análisis, co mo por ejemplo en los apuntes de Narvaiza (1.979).

Se comienza considerando a los n puntos del colectivo como grupos de un individuo.

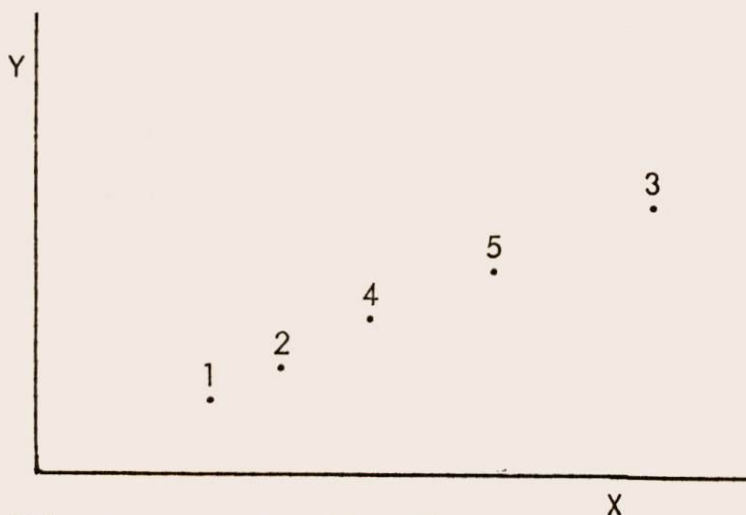
Creamos entre todos los puntos una matriz de similitud, que será la distancia Euclidiana entre los puntos en el espacio de las variables.

Dentro de esa matriz creamos un grupo con los dos puntos más próximos y en la fila y columna correspondiente a este nuevo grupo ponemos la distancia mayor a los demás puntos (la mayor de las correspondientes a los dos puntos unidos).

Luego repetimos este proceso hasta formar un solo grupo.

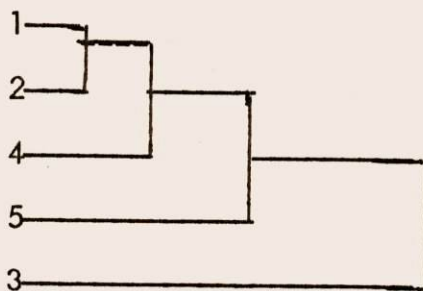
La diferencia en el encadenamiento simple consiste en que éste al formar la nueva matriz de similitud toma el menor valor en vez del mayor. (1)

Gráficamente sería lo siguiente:



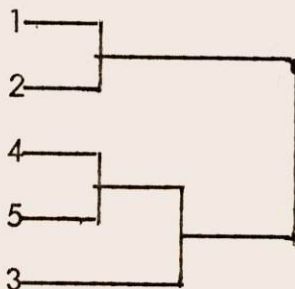
(1) Ver Narvaiza (1.979).

En encadenamiento simple el árbol sería:



Con lo que el punto 1 se uniría antes con el 5 que con el tres, es decir, se unen en formaciones alargadas.

En encadenamiento completo por el contrario, el árbol quedaría:



Es decir, forma grupos redondeados y en definitiva más homogéneos.

Estos encadenamientos forman parte de los denominados métodos jerárquicos.

Nosotros nos inclinamos por la utilización del encadenamiento completo para lograr una mayor homogeneidad.

Una vez elegido el método, tenemos que considerar en qué espacio nos vamos a mover. Podríamos utilizar el espacio de las variables, bien con todas ellas o bien en forma reducida con un grupo.

Pero he considerado como más interesante, para unificar todo el estudio, utilizar los factores obtenidos anteriormente. Esto por un lado supone una pérdida de información, a la vez que un aumento de la comodidad operativa, al tener un concentrada esa información en cuatro factores. Además ya hemos justificado que nuestros factores resultan altamente significativos y clasificadores.

Por lo tanto, pasamos a aplicar el encadenamiento completo en el espacio de los factores.

INTERPRETACION DEL CLUSTER ANALISIS:

El cluster análisis se encuentra en la hoja de ordenador nº 736 y siguientes. Que da lugar al árbol de la página siguiente, donde se ve como quedan formados los cuatro grupos.

Respecto a la formación de los grupos el problema es realmente dicícil pues no existe un criterio fijo para tomar la decisión de cortar por un punto.

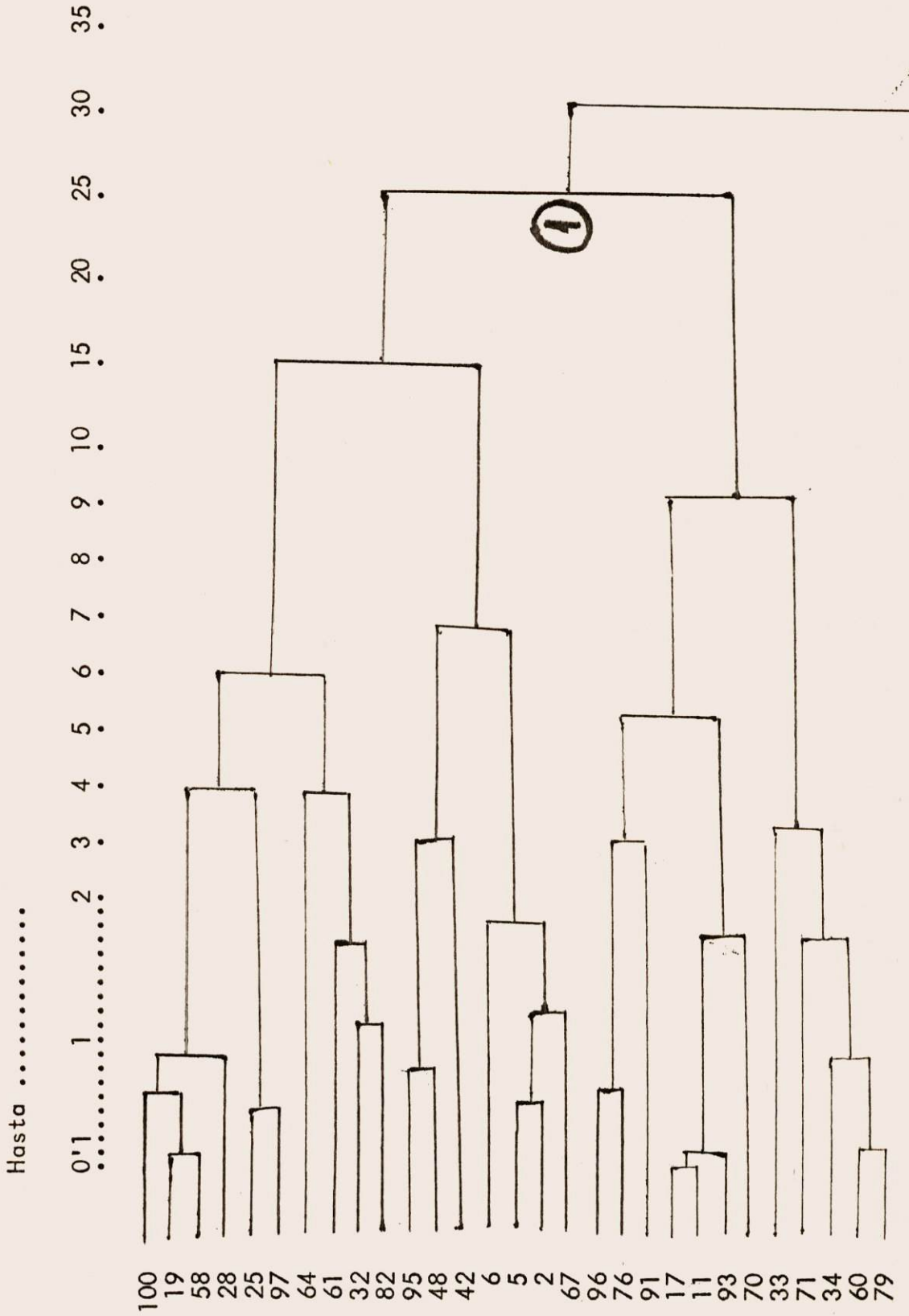
Es evidente que tanto el tener un solo grupo, como el tener muchos no resulta interesante. Aquí lo que hemos hecho es ver en qué punto teníamos que dividir para conseguir tres o cuatro grupos, dado que no había un salto espectacular, como para considerar un punto claro de división.

Los cuatro grupos, están como ya hemos dicho, explicitados en el gráfico del cluster, y podemos ver que los grupos 3 y 2 pueden formar un solo cluster, con lo que tendríamos solo tres grupos.

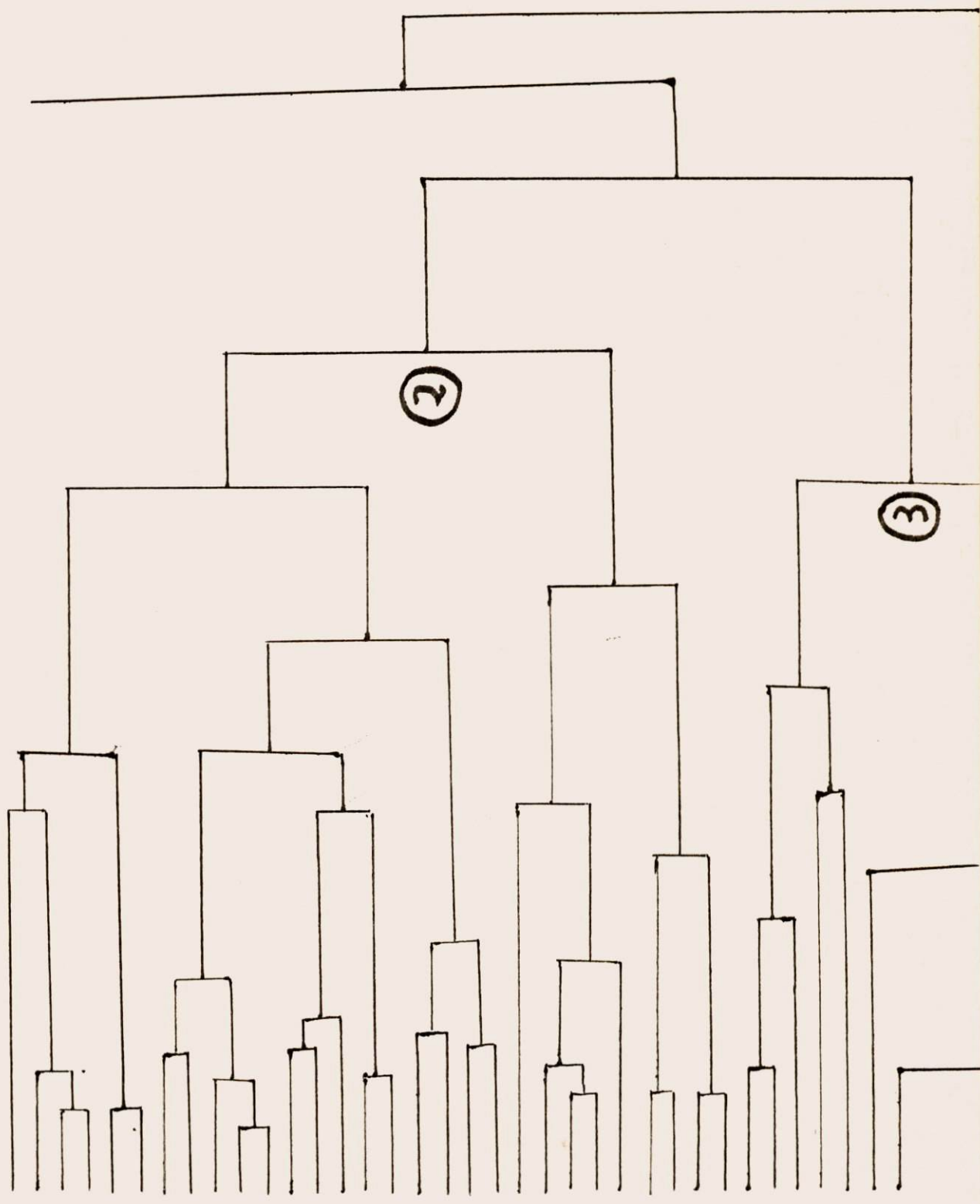
En principio nos inclinamos por cuatro grupos pues tres son pocos.

Vistas así las cosas, veamos los datos referentes a los cuatro clusteres:

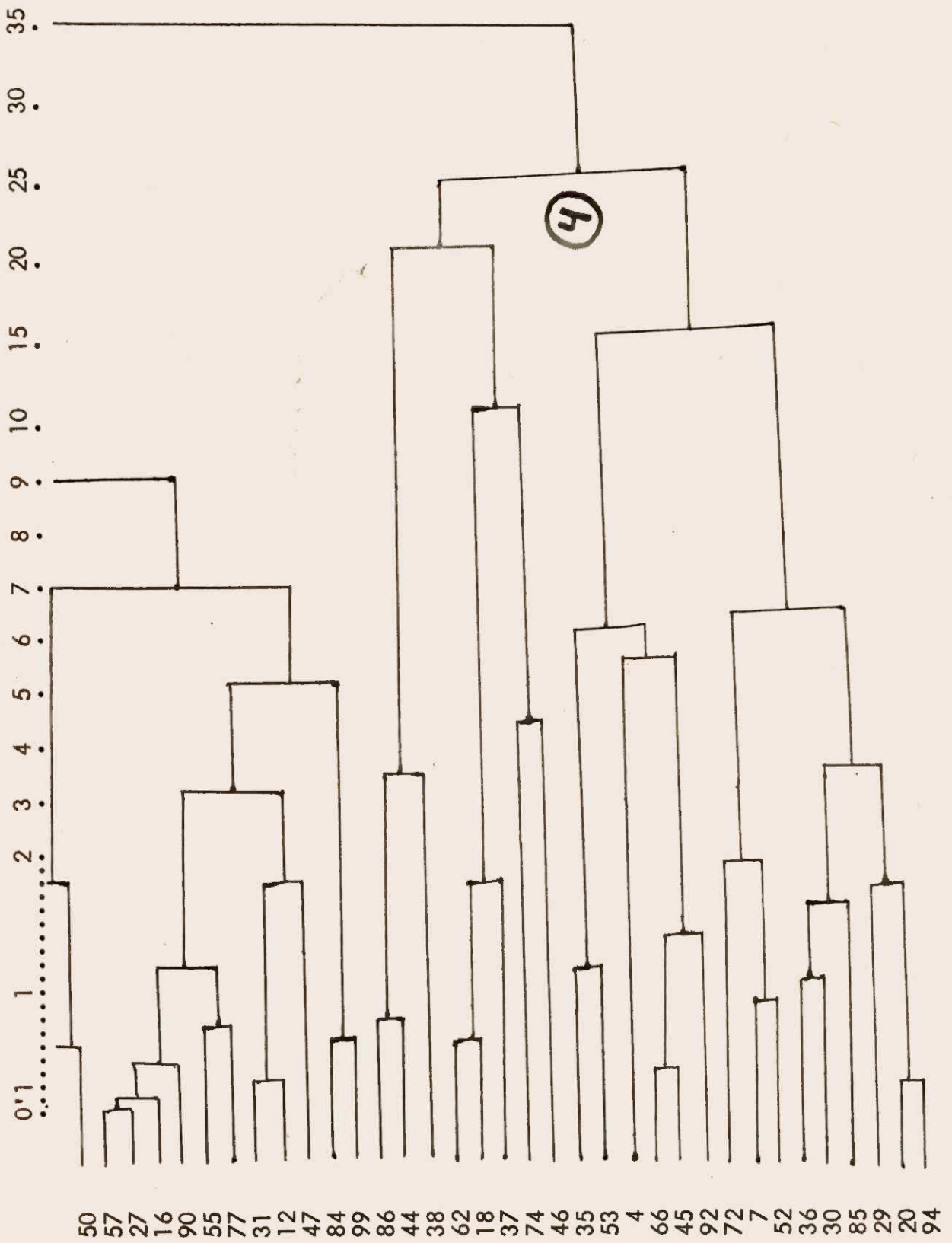
RIOJA - 3 CLUSTER LOGROÑO



0.1
.....
1
.....
2
.....
3
.....
4
.....
5
.....
6
.....
7
.....
8
.....
9
.....
10
.....
15
.....
20
.....
25
.....
30
.....
35



- 98
- 51
- 49
- 54
- 8
- 39
- 87
- 24
- 14
- 23
- 26
- 22
- 13
- 1
- 69
- 40
- 59
- 56
- 15
- 68
- 88
- 78
- 3
- 75
- 81
- 80
- 63
- 10
- 83
- 89
- 9
- 43
- 41
- 65
- 73
- 21



Cluster 1

número de datos: $n_1 = 29$

<u>Factores</u>	<u>Media de los factores en el cluster</u>	<u>Desviación típica de los factores en el cluster</u>
1	0'616559	1'126783
2	-0'073682	0'559435
3	-0'426804	0'873274
4	-0'851891	0'982160

Variable nº 3 (edad): media en el cluster = 3'10
desviación típica = 1'9336

Variable nº 2 (sexo): masculino 7 → 24%
femenino 22 → 76%

Variable nº 4 (estado): casado 13 → 45%
soltero 11 → 38%
viudo 5 → 17%

Proporciones representadas sobre el total de la muestra de Logroño:

Masculino - 17%
Femenino - 37%
Casado - 22%
Soltero - 42%
Viudo - 31%

Datos del cluster respecto al total: 29%

Cluster 2

número de datos: $n_2 = 29$

<u>Factores</u>	<u>Media de los factores en el cluster</u>	<u>Desviación típica de los factores en el cluster</u>
1	-0'494957	0'697560
2	0'744340	0'530158
3	-0'237874	0'786620
4	-0'014389	0'699166

Variable nº 3: media en el cluster = 4'66
 desviación típica = 1'6961

Variable nº 2: Masculino 13 → 45%
 Femenino 16 → 55%

Variable nº 4: Casado 17 → 59%
 Soltero 4 → 14%
 Viudo 8 → 28%

Proporciones representadas sobre el total de la muestra de Logroño:

Masculino - 32%
 Femenino - 27%
 Casado - 29%
 Soltero - 15%
 Viudo - 50%

Datos del cluster respecto al total: 29%

Cluster 3

número de datos: $n_3 = 19$

<u>Factores</u>	<u>Media de los factores en el cluster</u>	<u>Desviación típica de los factores en el cluster</u>
1	0'127087	0'720946
2	0'619840	0'568894
3	1'116300	0'562087
4	0'547466	0'627827

Variable nº 3: Media en el cluster = 2'89

Desviación típica = 1'3289

Variable nº 2: Masculino 7 → 37%

Femenino 12 → 63%

Variable nº 4 Casado 15 → 79%

Soltero 4 → 21%

Viudo 0 → 0%

Proporciones representadas sobre el total de la muestra de Logroño:

Masculino - 17%

Femenino - 20%

Casado - 26%

Soltero - 15%

Viudo - 0%

Datos del cluster respecto al total: 19%

Cluster 4

número de datos: $n_4 = 23$

<u>Factores</u>	<u>Media de los factores en el cluster</u>	<u>Desviación típica de los factores en el cluster</u>
1	-0'258310	0'988388
2	-1'357651	0'726547
3	-0'084064	1'063068
4	0'640032	0'853599

Variable nº 3: Media del cluster = 2'91

Desviación típica = 1'8069

Variable nº 2: Masculino 14 → 61%

Femenino 9 → 39%

Variable nº 4: Casado 13 → 56%

Soltero 7 → 30%

Viudo 3 → 13%

Proporciones representadas sobre el total de la muestra de Logroño:

Masculino - 34%

Femenino - 15%

Casado - 22%

Soltero - 27%

Viudo - 19%

Datos del cluster respecto al total: 23%

Con esto tenemos las medias y desviaciones típicas de los factores y de la variable 3 (edad) en el cluster. Además tenemos en nº de hombres y mujeres, así como de casados, solteros y viudos en el cluster. A parte del tanto por ciento que representan, tanto sobre el propio cluster como sobre el total de la muestra de Logroño.

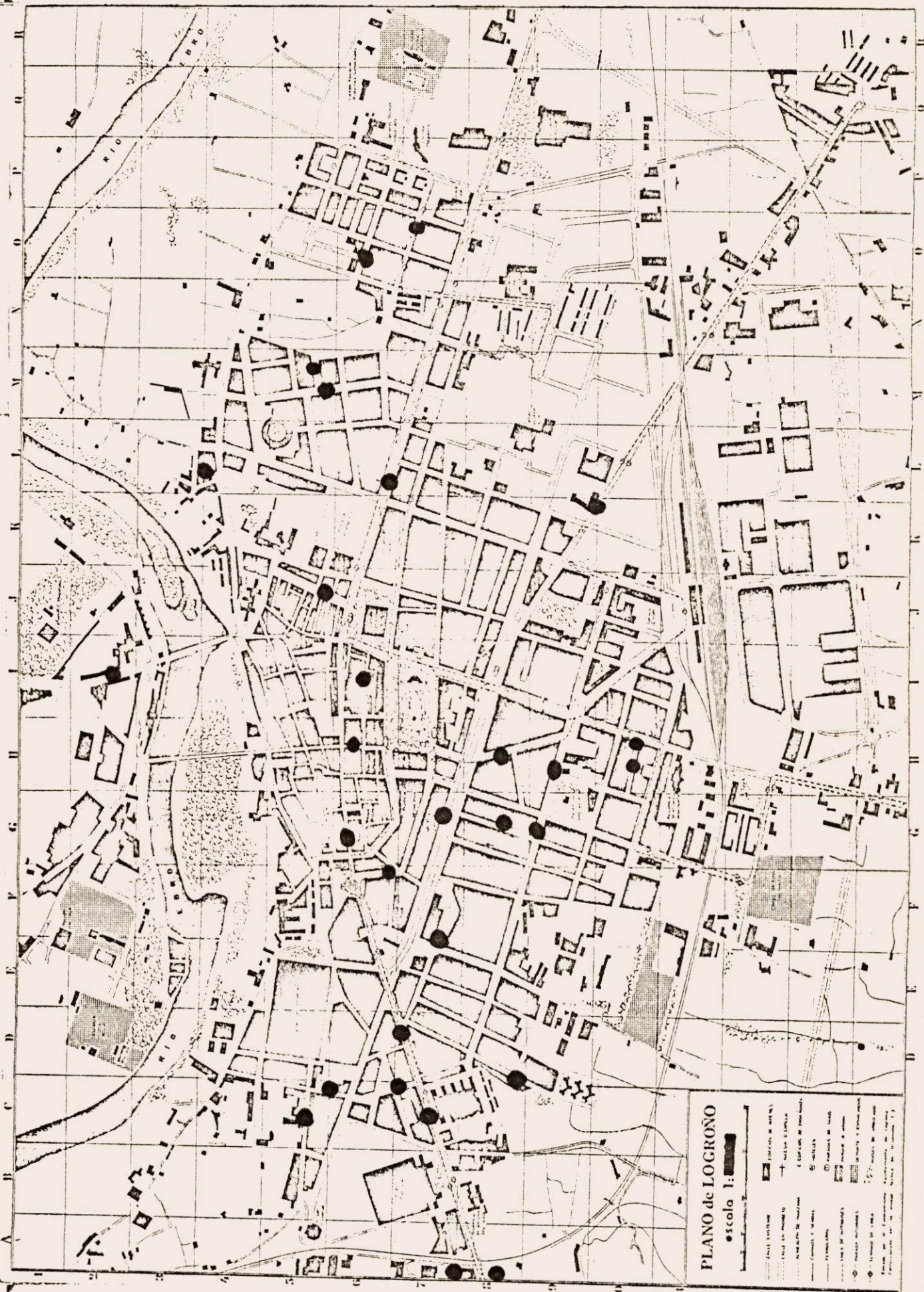
Para completar el análisis vamos a dar los datos - referentes al total de la muestra de Logroño:

Los factores tienen promedio cero y desviación típica uno.

Por otro lado las proporciones son las siguientes:

Masculino	-	41%
Femenino	-	59%
Casado	-	58%
Soltero	-	26%
Viudo	-	16%

Por último estudiaremos la ubicación de los encuestados en el plano de Logroño, como puede verse a continuación:



PLANO de LOGROÑO

escala 1: [redacted]

- [Symbol] Calle pavimentada
- [Symbol] Calle sin pavimento
- [Symbol] Avenida de anchura variable
- [Symbol] Camino y vereda
- [Symbol] Ferrocarril
- [Symbol] Línea de tranvías
- [Symbol] Puente
- [Symbol] Estación de ferrocarril
- [Symbol] Estación de tranvías
- [Symbol] Estación de autobuses
- [Symbol] Estación de taxis
- [Symbol] Estación de bicicletas
- [Symbol] Estación de coches
- [Symbol] Estación de motocicletas
- [Symbol] Estación de camiones
- [Symbol] Estación de autobuses
- [Symbol] Estación de taxis
- [Symbol] Estación de bicicletas
- [Symbol] Estación de coches
- [Symbol] Estación de motocicletas
- [Symbol] Estación de camiones

(2)



9



PLANO de LOCONO
escala 1:

- Calle (Street)
- Calle de Promoción (Promotion Street)
- Calle de Servicio (Service Street)
- Calle de Anillo (Ring Street)
- Calle de Circunvalación (Circumferential Street)
- Calle de Acceso (Access Street)
- Calle de Salida (Exit Street)
- Calle de Entrada (Entrance Street)
- Calle de Salida y Entrada (Exit and Entrance Street)
- Calle de Servicio y Anillo (Service and Ring Street)
- Calle de Servicio y Salida (Service and Exit Street)
- Calle de Servicio y Entrada (Service and Entrance Street)
- Calle de Servicio y Salida y Entrada (Service, Exit, and Entrance Street)
- Calle de Servicio y Anillo y Salida (Service, Ring, and Exit Street)
- Calle de Servicio y Anillo y Entrada (Service, Ring, and Entrance Street)
- Calle de Servicio y Anillo y Salida y Entrada (Service, Ring, Exit, and Entrance Street)

COMENTARIOS SOBRE LA UBICACION EN EL PLANO.

Hemos utilizado los siguientes símbolos:

Cluster 1	→	●	mapa 1
Cluster 2	→	●	mapa 2
Cluster 3	→	●	mapa 3
Cluster 4	→	●	mapa 4

Con lo que podemos ver cual es la vivienda (lugar aproximado) de cada uno de los encuestados. Al hacer las encuestas teníamos las direcciones de los encuestados y lo que hemos hecho es plasmar esas direcciones en el plano.

Esto puede ser un instrumento interesante de cara a la interpretación de los clusteres, pues la zona donde vive una persona puede decir mucho a un buen conocedor de la ciudad.

Lo que he hecho es reunir a un grupo de personas, que fueran de Logroño y con criterio suficiente y conocimientos adecuados, para poder interpretar la ubicación en el plano de los encuestados.

Pero tampoco podemos sobre valorar este dato, y no debe extrañarnos que en el mismo grupo aparezca una vivienda cara y otra humilde. Tengamos en cuenta, que en una vivienda cara no viven solo los dueños, sino por ejemplo, también el servicio doméstico. Otro caso puede ser que en una zona cara

haya casas viejas de renta ridícula, y tantos otros casos que se le puedan ocurrir al lector.

En conclusión, tendremos en cuenta este dato, pero guiándonos fundamentalmente por la tendencia mayoritaria del cluster.

ANALISIS DE LOS CLUSTERES. (resumen de datos de factores)Cluster 1

<u>Factores</u>	<u>Promedio</u>	<u>I.I.</u>
1	0'616559	0'49
2	-0'073682	--
3	-0'426804	0'48
4	-0'851891	-1'25

Cluster 2

<u>Factores</u>	<u>Promedio</u>	<u>I.I.</u>
1	-0'494957	0'49
2	0'744340	--
3	-0'237894	0'48
4	-0'014389	-1'25

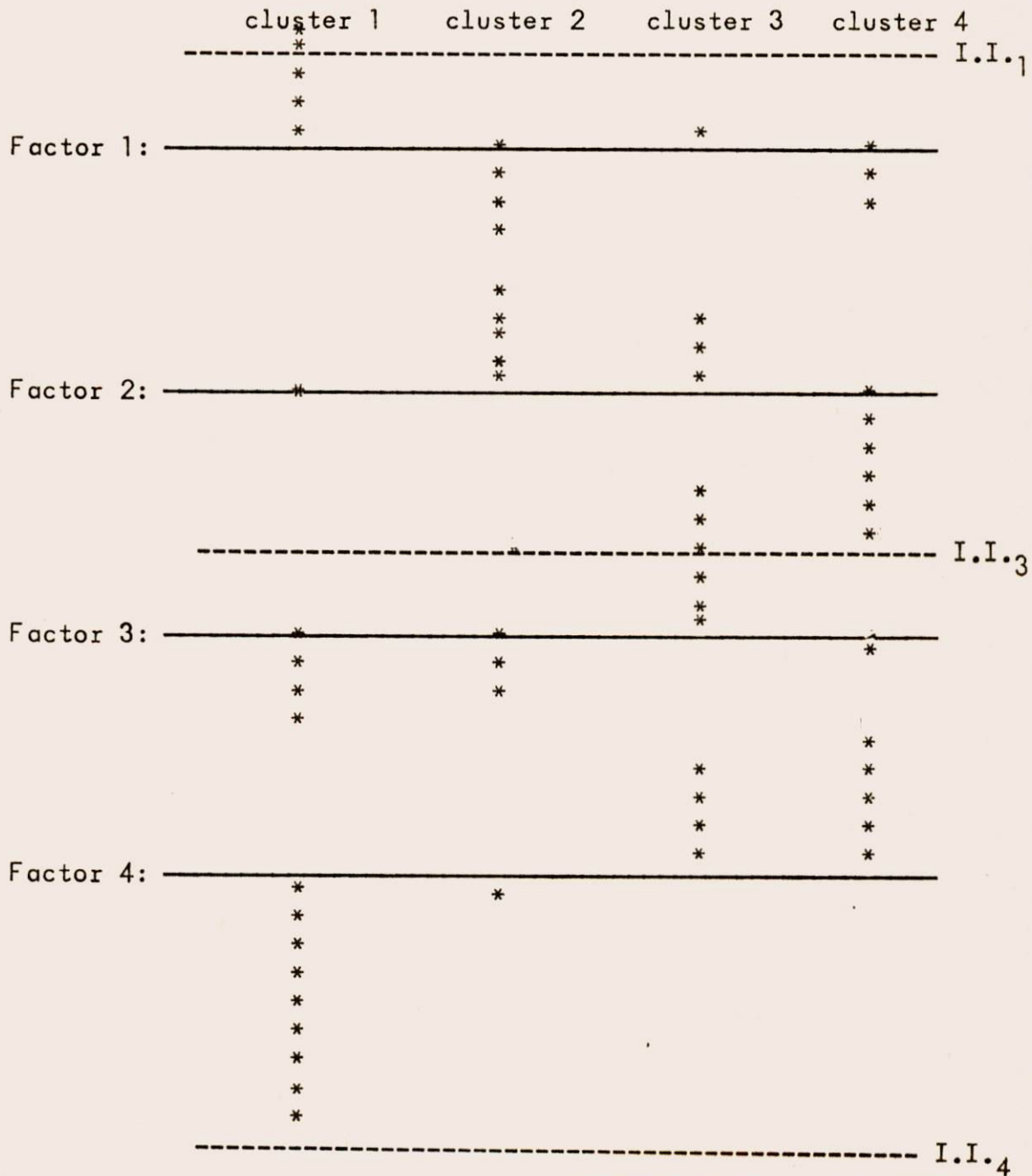
Cluster 3

<u>Factores</u>	<u>Promedio</u>	<u>I.I.</u>
1	0'127087	0'49
2	0'619840	--
3	1'116300	0'48
4	0'547466	-1'25

Cluster 4

<u>Factores</u>	<u>Promedio</u>	<u>I.I.</u>
1	-0'258310	0'49
2	-1'357651	--
3	-0'084064	0'48
4	0'640032	-1'25

REPRESENTACION GRAFICA.



INTERPRETACION.

Vamos a intentar, con los datos que hemos resumido en las páginas anteriores, dar un sentido a cada cluster, es decir, resumir la información, viendo las características de cada grupo.

Cluster 1:

Es el cluster más progresista (dado el promedio del factor uno), y el único que puede calificarse como tal, dado que su promedio supera al individuo indiferente. Lo cierto es que supera por muy poco al individuo indiferente y dado el alto valor de su desviación típica hay que concluir que muchos de sus componentes están por debajo del valor del I.I. Luego en términos referidos al I.I. podríamos calificar a este grupo como moderadamente liberal-progresista. Pero si lo comparamos con el resto de la muestra de Logroño, habría que decir que es fuertemente liberal-progresista.

La cultura y clase social (factor dos) hay que considerarlas como las medias de la muestra, sin demasiada dispensión.

El factor tres (antiautonomista), nos da el valor negativo más fuerte, y muy lejos del I.I. luego por lo tanto es el grupo más autonomista.

Respecto al factor cuatro (cooperativista-regionalista), tienen el valor negativo más fuerte. De todas formas es

tán todavía muy por encima del I.I. y por lo tanto, hemos de pensar que son cooperativistas y regionalistas, pero me nos que el resto de la muestra.

La edad media del cluster está por los 40 años, con alta desviación típica.

Hay un claro exceso de mujeres (76%) y con solo el 29% de los individuos, contiene un 37% de las mujeres del total de la muestra.

También hay que resaltar que están en este grupo 42% de los solteros de toda la muestra y el 31% de los viudos en detrimento de un menor número de casados.

Mirando el plano, vemos que hay una gran dispersión geográfica en este cluster, pero de todas formas se puede ver que hay una tendencia a la vivienda barata y a la zona de expansión de Logroño, predominando edificios de relativamente reciente construcción.

Resumiendo este grupo, sus características fundamentales son:

En primer lugar, liberales y autonomistas, con muchas mujeres y pocos casados, y por lo tanto, con menos obligaciones familiares, con clase y cultura media, que se refleja en la situación en el plano, es decir, con tendencia a la vivien

da barata, pero decente, con una edad media de alrededor de 40 años (pero con alta dispersión) y desconfían algo de las cooperativas, aunque menos que el I.I.

Cluster 2:

Respecto al factor uno, es el grupo más reaccionario, si la media de la muestra ya es reaccionaria, estos son destacados en ese sentido.

Este fuerte antiliberalismo, se junta con gran incultura y baja clase social (dado su alto valor en el factor dos). Además estos dos factores tienen una desviación típica relativamente baja, sobre todo el segundo, lo que da su poca dispersión.

En lo que se refiere al factor tres, son bastante autonomistas.

En el factor cuatro, son algo cooperativistas y regionalistas, pues están en la media de la muestra, y mucho más altos que el I.I.

La edad media es bastante alta (alrededor de 55 años).

Hombres y mujeres van bastante a la par y lo que se no

ta es un gran exceso de viudas y una escasez de solteros, cosa normal a causa de la edad avanzada del grupo.

En cuanto a la ubicación geográfica, se ve una tendencia a la vivienda humilde, teniendo en cuenta que la mayoría de los individuos del grupo habitan en casas periféricas y -de deficiente construcción.

Esto es coherente con la baja clase social apreciada.

En conclusión podemos decir que se trata de un grupo donde los individuos son mayoritariamente reaccionarios, de baja clase social y cultura, de edad avanzada, con un consiguiente exceso de viudas; por otro lado tienen tendencia autonomista y algo cooperativistas y sus viviendas son humil-des.

Creo que el grupo queda claramente definido con lo -que hemos dicho hasta aquí.

Cluster 3:

En lo que se refiere al factor uno, hay un ligero liberalismo si lo comparamos con la media de la muestra, pero son reaccionarios comparados con el I.I. luego los podemos calificar como moderadamente reaccionarios.

En el factor dos tienen un valor alto, lo que quiere decir incultura y baja clase social, con poca dispersión.

En el factor tres, dan el valor positivo más alto, superando incluso al I.I. la dispersión también es bastante baja, luego en conclusión, es el grupo más antiautonomista.

En el factor cuatro, vemos que son altamente cooperativistas.

La edad media es más bien baja, unos 38 años, lo que produce una ausencia de viudos (0%), sin embargo es curiosa la gran cantidad de casados (79%) que representan el 26% de todos los casados de la muestra, cuando el cluster solo tiene el 19% de los datos, pues es el más reducido. También hay bastantes mujeres (63%).

En conclusión de estos últimos datos podemos ver que es un grupo donde predominan los casados más bien jóvenes.

En cuanto a la ubicación en el plano es bastante dispersa con tendencia a la vivienda humilde, cosa lógica si tenemos en cuenta su baja clase socio-cultural.

En resumen, tenemos un grupo moderadamente reaccionario, de baja clase socio-cultural, antiautonomista, altamente cooperativistas, más bien jóvenes y casados.

Cluster 4 :

Este es un grupo bastante reaccionario, más que la media, y mucho más que el I.I.

Es característico el bajo valor de su media en el factor dos (incultura), es decir, es un grupo de clase alta (socio-cultural). Es el grupo de clase más alta y muy lejos de la media.

En el factor tres, por el contrario, están en la media y, por lo tanto, no se destacan por él, pero de todas formas están bastante por debajo del I.I. y los pudiéramos calificar de moderadamente autonomistas, aunque con una gran dispersión.

En el factor cuatro, vemos que son los más cooperativistas.

En cuanto a la edad, son relativamente jóvenes (39 - años), sin que por ello falten los viudos (13%), y aunque - hay predominio de casados (56%), están en la línea del promedio.

Sí es de destacar la gran cantidad de hombres (61%), muy por encima de la media.

La ubicación geográfica se corresponde perfectamente con la elevada clase del grupo y se sitúa preferentemente en

la zona centro de Logroño, sobre todo en las zonas más caras o en las calles céntricas tradicionales.

En conclusión, los rasgos más destacables son su elevada clase socio-cultural, que se corresponde con las viviendas, bastante reaccionarios, moderadamente autonomistas y - fuertemente cooperativistas. También son relativamente jóvenes y con predominio masculino.

CONCLUSIONES DEL CLUSTER ANALISIS Y ANALISIS DISCRIMINANTE

El resultado del estudio del cluster análisis creo que podemos considerarlo como satisfactorio, dado que hemos conseguido cuatro grupos bastante definidos.

Por otro lado, introduciendo para la realización del cluster análisis, tan sólo los cuatro factores, hemos visto que variables, que no estaban introducidas en el análisis, - como edad, estado y ubicación en el plano, han resultado coherentes con los resultados del análisis.

En conclusión, dados los cuatro grupos, sería interesante estudiar el análisis discriminante y así resolveríamos futuras clasificaciones.

Es decir, dado un individuo nuevo, podemos, mediante la ecuación de transformación, lograr su puntuación en los factores a partir de sus valores en las variables. Con estas puntuaciones y el análisis discriminante correspondiente podemos estudiar a qué grupo debería unirse con mayor probabilidad.

Realizando este análisis por el método de Fisher (1),

- (1) Utilizamos el método de Fisher pues es para el único que tenemos programa de ordenador. El problema es que hay que suponer la normalidad de las variables, pero como vamos a utilizar los factores, estos se puede suponer que son algo normales, por otro lado, los resultados son satisfactorios.

tenemos su resultado en las hojas de ordenador nº 740 y ss. En cada cluster el número de ordenador del individuo, es el mismo que el que le corresponde en el cluster según la ordenación primera.

La D^2 de Mahalanobis = 402'29. Recordemos que el método de Fisher lo que pretende es maximizar la D^2 de Mahalanobis, luego hemos de ver si ésta es suficientemente grande (1).

Este estadístico sigue la χ^2 con $K = m (g - 1)$ grados de libertad. (2)

- (1) En realidad nuestro programa de ordenador obtiene 4 hiperplanos (dado que se trata de discriminar 4 grupos), cada uno de los cuales maximiza la D cuadrado de Mahalanobis entre el centroide de un grupo y el centroide total. Para discriminar dos grupos habría que calcular la intersección de sus dos hiperplanos correspondientes.

La D^2 utilizada es:

$$V = \sum_{i=1}^m \sum_{j=1}^m d_{ij} \sum_{k=1}^g n_k (X_{ik} - X_i) (X_{jk} - X_j)$$

g - grupos

m - variables

d_{ij} - elemento de la inversa de la matriz de varianzas y covarianzas.

- (2) La utilización de este test implica una prueba de hipótesis, que aún saliéndose del esquema general del análisis de datos, nos indica que lo que se ha hecho es razonable.

El test utilizado se encuentra propuesto IBM - Application Program 1130 (1.970), página 39; que es el manual utilizado para la programación de nuestro ordenador.

En nuestro caso $K = 4 \cdot 3 = 12$

$$\chi^2_{0'001, K=12} = 32'909$$

Luego como $32'909 < 402'29$, vemos que se rechaza la hipótesis de que los clusteres son iguales, luego podemos afirmar que son distintos. (1)

Esto es lógico, ya que los grupos provienen de un cluster análisis.

Aquí tenemos la ventaja de no haber utilizado el método matricial de cluster análisis, pues de haberlo hecho hubiera tenido escaso interés el análisis discriminante, pues sería esencialmente lo mismo, como ya demostramos en la primera parte.

- (1) Discutiendo el test utilizado, hemos de afirmar que en el caso de dos grupos la D cuadrado de Mahalanobis sigue la T cuadrado de Hotelling que se convierte fácilmente en una F de Fisher con $(p, m - p - 1)$ grados de libertad; en el caso de muchos datos este estadístico se aproxima asintóticamente a la χ^2 con p grados de libertad (siendo p , el número de variables y m el de individuos), puede verse en Chacón (1.955), pag. 419; también en Anderson (1.958) pag. 109; Bock (1.975) pag. 400 y otros. (esto es válido si la hipótesis nula es que la diferencia de los promedios de los colectivos es cero, si no entraríamos en la F y χ^2 descentradas).

En el caso de más de dos grupos Tatsuoka (1.971) pag. 164, expone que D cuadrado sigue aproximadamente χ^2 con $m(g - 1)$ grados de libertad, también está en Bock (1.975) pags. 405 - 406. Este es el mismo test del ya citado manual IBM (A.P. 1130) pag. 39.

Sin embargo, al haber utilizado un método de encadenamiento (concretamente el encadenamiento completo), el análisis discriminante sí tiene interés y nos confirma que los grupos están bien hechos.

Por otro lado, el análisis discriminante nos soluciona el problema de la clasificación, como ya hemos comentado, según las puntuaciones en la encuesta, podemos clasificar a un nuevo individuo.

Por último, nos proporciona la lista de los individuos, viendo si están bien clasificados o no en su cluster, y vemos que mayoritariamente lo están.

TABLAS DE CONTINGENCIA: (1)

Bajo este epígrafe nos referimos al estudio de dos grupos de variables cualitativas. Ya comentamos en la primera parte que este tipo de estudios pueden realizarse mediante la utilización de la correlación canónica.

Evidentemente este sería el camino, pero carecemos de un programa adecuado para hacer el estudio de forma operativa pues aun teniendo un programa de correlación canónica en nuestro ordenador, no poseemos los encadenamientos precisos para hacer el estudio completo. Sin estos encadenamientos, el trabajo de ordenador es muy penoso pues hay que estar haciendo constantes entradas de información.

Ante esta situación, y sin tener tampoco ningún otro programa adecuado para la utilización de otro método de estudio de las tablas de contingencia, me decidí por realizarlo manualmente, al menos lo fundamental.

Para la realización manual, la correlación canónica resultaba imposible y me decidí por utilizar el método de χ^2 aunque éste también es bastante pesado.

- (1) En este apartado, y en el siguiente (cruces de variables) vamos a realizar algunas pruebas de hipótesis, que aun saliéndose algo del planteamiento general de análisis de los datos nos pueden dar una idea de posible inducción y ampliación del estudio.

Lo que voy a estudiar es la influencia de las variables

- 2 - Sexo.
- 3 - Edad.
- 4 - Estado.

En la formación de los clusteres, utilizando el estadístico $\chi^2 = \sum \frac{(O - E)^2}{E}$ y en situaciones dicotómicas.

Luego por lo tanto, voy a ver si sigue la distribución teórica, si rechazo que la siga, acepto que hay relación entre la variable y la formación de los grupos.

Creo que es importante aquí hacer una puntualización: La utilización de este test de la χ^2 se ha popularizado mucho, sobre todo entre los sociólogos, etc. pero se aplica muchas veces sin verdadero conocimiento de lo que se está haciendo. Tengamos en cuenta que el estadístico:

$$\sum \frac{(O - E)^2}{E}$$

sigue asintóticamente la χ^2 , luego si los E son muy pequeños el test es de muy discutible validez. Podemos considerar que si $E > 5$ en cada intervalo, el test es válido. (1)

Vamos a ir estudiando la influencia de cada un de las tres variables en la formación de los grupos:

(1) Como puede verse en Kendall y Stuart (1.966), vol. II, pag. 457.

$$\chi^2 = \sum \frac{(O - E)^2}{E} ; K = (p - 1) (q - 1)$$

Variable 2 - sexo.

Datos observados:

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Masculino	7	13	7	14	41
Femenino	22	16	12	9	59
	29	29	19	23	

Datos estimados: (1)

11'89	11'89	7'79	9'43
17'11	17'11	11'21	13'57

$$\chi^2 \text{ experimental} = 7'47$$

$$\chi^2 \text{ teórica} = 7'815 (\alpha = 0'05) (K = 3)$$

$$7'815 > 7'47$$

Luego aceptamos que sigue la teórica, y rechazamos que haya correlación.

(1) Son las correspondientes a la distribución teórica.

Variable 2 - estado.

Datos observados:

	1	2	3	4	
casado	13	17	15	13	58
no casado	16	12	4	10	42
	29	29	19	23	

Datos estimados:

16'82	16'82	11'02	13'34
12'18	12'18	7'98	9'66

$$\chi_e^2 = 5'51$$

$$\chi_{e'os}^2 = 7'815$$

Luego aceptamos que sigue la teórica, y rechazamos que haya correlación.

En la variable estado, como ya hemos dicho que vamos a hacer un estudio dicotómico, hemos dividido entre casados y no casados, veamos otras divisiones posibles.

Datos observados:

	1	2	3	4	
Soltero	11	4	4	7	26
No soltero	18	25	15	16	74
	29	29	19	23	

Datos estimados:

7'54	7'54	4'94	5'98
21'46	21'46	14'06	17'02

$$\chi^2_e = 4'87$$

$$\chi^2_{0.05} = 7'815$$

Luego aceptamos que sigue la teórica, y rechazamos que haya correlación.

Aquí nos hemos encontrado con el problema de que los estimados del grupo tres, solteros son $4'94 < 5$. En este caso tiene poca importancia pues es prácticamente cinco, pero luego nos encontramos con casos en los que el problema es mayor.

De todas formas seguiremos con el estudio, sabiendo que si los estimados son muy pequeños, la fiabilidad del análisis disminuye.

Datos observados:

	1	2	3	4	
Viudo	5	8	0	3	16
No viudo	24	21	19	20	84
	29	29	19	23	

Datos estimados:

4'64	4'64	3'04	3'68
24'36	24'36	15'96	19'32

$$\chi_e^2 = 6'70$$

$$\chi_{0.05}^2 = 7'815$$

Luego aceptamos la teórica.

Variable 3 - edad (la edad se puntúa de uno a siete).

Datos observados:

	1	2	3	4	
1	9	1	2	7	19
Resto	20	28	17	16	81
	29	29	19	23	

Datos estimados:

5'51	5'51	3'61	4'37
23'49	23'49	15'39	18'63

$$\chi^2_e = 10'13$$

$$\chi^2_{0.05} = 7'815$$

Se rechaza la teórica, hay correlación.

Datos observados:

	1	2	3	4	
1, 2	12	4	9	10	35
Resto	17	25	10	13	65
	29	29	19	23	

Datos estimados:

10'15	10'15	6'65	8'05
18'85	18'85	12'35	14'95

$$\chi^2_e = 8'26$$

$$\chi^2_{0.05} = 7'815$$

Se rechaza la teórica, hay correlación.

Datos observados:

	1	2	3	4	
1, 2, 3	18	7	13	14	52
Resto	11	22	6	9	48
	29	29	19	23	

Datos estimados:

15'08	15'08	9'88	11'96
13'92	13'92	9'12	11'04

$$\chi^2_e = 12'97$$

$$\chi^2_{0'05} = 7'815$$

Se rechaza la teórica, hay correlación.

Datos observados:

	1	2	3	4	
1, 2, 3, 4	22	14	17	19	72
Resto	7	15	2	4	28
	29	29	19	23	

Datos estimados:

20'28	20'88	13'68	16'56
8'12	8'12	5'32	6'44

$$\chi^2_e = 12'47$$

$$\chi^2_{0.05} = 7'815$$

Se rechaza la teórica, hay correlación.

Datos observados:

	1	2	3	4	
1, 2, 3, 4, 5	25	18	18	21	82
Resto	4	11	1	2	18
	29	29	19	23	

Datos estimados:

23'78	23'78	15'58	18'86
5'22	5'22	3'42	4'14

$$\chi^2_e = 11'59$$

$$\chi^2_{0.05} = 7'815$$

Se rechaza la teórica, hay correlación.

Datos observados:

	1	2	3	4	
1, 2, 3, 4, 5, 6	27	25	19	22	93
Resto	2	4	0	1	7
	29	29	19	23	

Datos estimados:

26'97	26'97	17'67	21'39
2'03	2'03	1'33	1'61

$$\chi^2_e = 3'73$$

$$\chi^2_{005} = 7'815$$

Sigue la teórica.

En conclusión de esta primera parte del estudio de las tablas de contingencia, vemos que la mejor división es la que se hace con edad, dividiendo entre 1, 2, 3 y 4, 5, 6, 7, es decir, mayor o menor de 45 años, con un $\chi^2 = 12'97$. Por otro lado, no tenemos problemas con los estimados en este caso, que siempre son mayores que 5.

Una vez que sabemos que la edad, en un intervalo determinado, realiza la mejor división, pasamos a estudiar dentro de cada uno de los dos grupos de edad discriminados, cuales son las variables que discriminan mejor. Así conseguiremos un estudio completo de los cuatro grupos en relación con las variables citadas.

Segunda Subdivisión.

a) Grupo 1 (edad 1, 2, 3)

Variable 2 - sexo.

Datos observados:

	1	2	3	4	
Masculino	6	2	4	10	22
Femenino	12	5	9	4	30
	18	7	13	14	

Datos estimados:

7'62	2'96	5'50	5'92
10'38	4'04	7'50	8'08

$$\chi^2_e = 6'71$$

$$\chi^2_{0'05} = 7'815$$

Luego sigue la teórica

Variable 4 - estado.

Dado que en este grupo de edad no hay ningún viudo en nuestra muestra, solo estudiaremos solteros y casados.

Datos observados:

	1	2	3	4	
Soltero	10	2	4	7	23
Casado	8	5	9	7	29
	18	7	13	14	

Datos estimados:

7'96	3'10	5'75	6'19
10'04	3'90	7'25	7'81

$$\chi^2_e = 2'78$$

$$\chi^2_{0'05} = 7'815$$

Sigue la teórica.

Variable 3 - edad.

Estamos en el grupo de edad 1, 2, 3, luego solo haremos dicotomías con esas edades.

	1	2	3	4	
1	9	1	2	7	19
2, 3	9	6	11	7	33
	18	7	13	14	

Datos estimados:

6'58	2'56	4'75	5'12
11'42	4'44	8'25	8'88

$$\chi_e^2 = 6'5$$

$$\chi_{obs}^2 = 7'815$$

Sigue la teórica.

Datos observados:

	1	2	3	4	
1, 2	12	4	9	10	35
Resto	6	3	4	4	17
	18	7	13	14	

Datos estimados:

12'12	4'71	8'75	9'42
5'88	2'29	4'25	4'58

$$\chi_e^2 = 0'46$$

$$\chi_{obs}^2 = 7'815$$

Luego sigue la teórica.

b) Grupo 2 (edad, 4, 5, 6, 7)

Variable 2 - sexo.

Datos observados:

	1	2	3	4	
Masculino	1	11	3	4	19
Femenino	10	11	3	5	29
	11	22	6	9	

Datos estimados:

4'35	8'71	2'38	3'56
6'65	13'29	3'63	5'44

$$\chi^2_e = 5'64$$

$$\chi^2_{D'W} = 7'815$$

Luego sigue la teórica.

Variable 4 - estado.

Datos observados:

	1	2	3	4	
Casado	5	12	6	6	29
No casado	6	10	0	3	19
	11	22	6	9	

Datos estimados:

6'65	13'29	3'63	5'44
4'35	8'71	2'38	3'56

$$\chi^2_e = 5'4$$

$$\chi^2_{\text{obs}} = 7'815$$

Luego sigue la teórica.

Datos observados:

	1	2	3	4	
Soltero	1	2	0	0	3
No soltero	10	20	6	9	45
	11	22	6	9	

Datos estimados:

0'69	1'38	0'38	0'56
10'31	20'63	5'65	8'44

$$\chi^2_e = 1'45$$

$$\chi^2_{\text{obs}} = 7'815$$

Luego sigue la teórica.

Datos observados:

	1	2	3	4	
Viudo	5	8	0	3	16
No viudo	6	14	6	6	32
	11	22	6	9	

Datos estimados:

3'67	7'33	2	3
7'33	14'67	4	6

$$\chi_e^2 = 3'82$$

$$\chi_{0'05}^2 = 7'815$$

Luego sigue la teórica.

Variable 3 = edad.

Datos observados:

	1	2	3	4	
4	4	7	4	5	20
5, 6, 7	7	15	2	4	28
	11	22	6	9	

Datos estimados:

4'58	9'17	2'5	3'75
6'42	12'83	3'5	5'25

$$\chi^2_e = 3'26$$

$$\chi^2_{0'05} = 7'815$$

Luego sigue la teórica.

Datos observados:

	1	2	3	4	
4, 5	7	11	5	7	30
6, 7	4	11	1	2	18
	11	22	6	9	

Datos estimados:

6'88	13'75	3'75	5'63
4'13	8'25	2'25	3'38

$$\chi^2_e = 3'48$$

$$\chi^2_{0'05} = 7'815$$

Luego sigue la teórica.

Datos observados:

	1	2	3	4	
4, 5, 6	9	18	6	8	41
7	2	4	0	1	7
	11	22	6	9	

Datos estimados:

9'40	18'79	5'13	7'69
1'60	3'21	0'88	1'31

$$\chi_e^2 = 1'45$$

$$\chi_{obs}^2 = 7'815$$

Luego sigue la teórica.

CONCLUSIONES DEL ESTUDIO DE TABLAS DE CONTINGENCIA:

En resumen, lo que hemos hecho es estudiar la relación entre los clústeres y las variables 2, 3 y 4. Dicotomizando estos de forma lógica para que el estudio fuera más completo.

Hemos encontrado que la dicotomía más significativa es la edad en la división 1, 2, 3 - 4, 5, 6, 7; quedando el cuadro:

	1	2	3	4
1, 2, 3	18	7	13	14
4, 5, 6, 7	11	22	6	9

$$\chi^2_e = 12'97$$

Después hemos cogido cada grupo y hemos repetido el proceso y no ha resultado ninguna división significativa.

Podríamos concluir aquí el estudio, pero considero que todavía podemos aumentar el error α . Por ejemplo con un $\alpha = 0'1$.

$$\chi^2_{0'1} = 6'251$$

Aceptaríamos el sexo como división válida en el grupo 1, 2, 3; pues es esa división teníamos:

$$\chi_e^2 = 6'71$$

Si aumentásemos α al 20%:

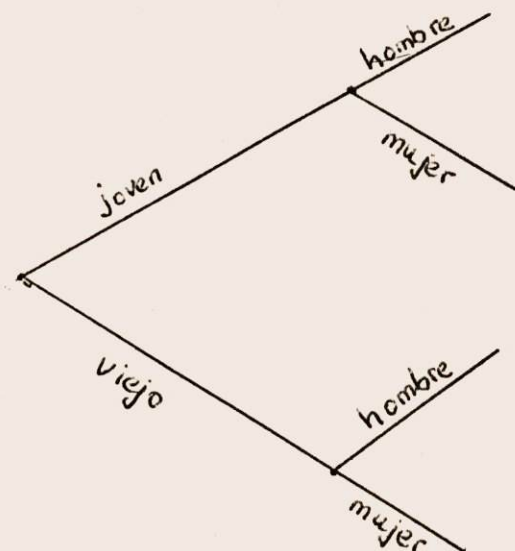
$$\chi_{0.2}^2 = 4'642$$

esto nos permitiría aceptar el sexo en el grupo 4, 5, 6, 7.

Pues en esa división teníamos:

$$\chi_e^2 = 5'64$$

Luego el árbol de discriminación quedaría:



joven < 45 años.

viejo > 45 años.

Evidentemente se podría completar el estudio con otras variables como lugar de residencia, etc. o haciendo tricotomías, etc. pero sin ordenador considero que es suficiente llegar hasta donde hemos llegado.

<u>Nº variable</u>								<u>Nº variable</u>
17	30
17	32
17	35
17	39
18	19
18	23
19	23
19	30
24	30
27	29
27	32
27	35
27	39
29	32
29	35
29	39
32	35
32	39
35	39

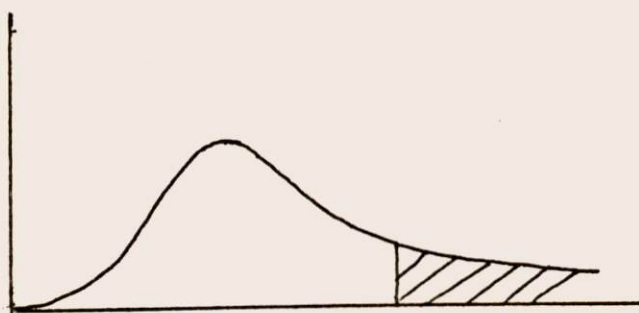
Podemos considerar que son los cruces más interesantes, según lo que ya sabemos gracias a todo el estudio realizado hasta el momento.

Por otro lado, el ordenador nos da los valores de χ^2 correspondientes a cada cruce, en una aplicación similar a

la realizada en las tablas de contingencia. El problema es que dado el número de intervalos y de datos, muy pocas de estas χ^2 se pueden considerar significativas. Tengamos en cuenta que nosotros exigimos que las estimadas sean al menos cinco en cada intervalo. El ordenador nos dice si hay algún intervalo con estimación menor que uno, lo que viene a decirnos que ya es muy poco significativo.

En la mayoría de los cruces con 267 datos, y en todos los de 100 datos, el ordenador nos dice que hay uno o más - valores estimados menores que uno, con lo que el valor de la comparación de la χ^2 teórica con la experimental es muy bajo.

Por lo tanto nosotros no utilizamos el test de χ^2 en nuestro estudio de cruces, excepto cuando sea significativo, (1), pero de todas formas vamos a dar a continuación los valores correspondientes a la χ^2 teórica con los grados de libertad usados en el presente estudio, para que el lector pueda hacer las comparaciones que crea interesantes.



- (1) En este caso vamos a dar como significativos los valores de χ^2 correspondientes a cuadros donde todos los estimados sean al menos mayores que uno, que es el dato que - proporciona el ordenador.

<u>Grados de libertad</u>	<u>$\chi^2_{0.05}$</u>	<u>$\chi^2_{0.01}$</u>
12	21	26'2
18	28'9	34'8
24	36'4	43
30	43'8	50'9
32	46'2	53'5
36	51	58'6
48	65'2	73'7

La utilidad de estos cruces es que el analista pueda ver rápidamente como se relacionan las variables más interesantes y qué frecuencias hay en los intervalos.

Sería pesado y poco interesante el estudiar exhaustivamente cruce por cruce. De todas formas vamos a ver un breve resumen de cada uno.

Cruce 3 - 11

Variable 3: edad.

Variable 11: sí al divorcio.

Rioja: se ve cierta relación, la mayoría está contra el divorcio, pero entre los más jóvenes hay bastante gente de acuerdo.

Logroño: la mayoría va hacia el acuerdo con el divorcio, fundamentalmente hasta los 45 años.

Cruce 3 - 17

Variable 17: años con Franco buenos.

Rioja: es uno de los pocos casos en que no hay valores estimados menores que uno, por lo tanto aquí si que puede tener algún interés la utilización de la χ^2 .

$$\chi^2_{exp.} = 123'4341 \quad K = 36$$

Vemos que en todo caso es mayor que la teórica, por lo tanto, hay correlación, pues se rechaza que siga la teórica. La mayoría están con Franco, siendo clara la influencia de la edad, entre los jóvenes hay más desacuerdo, no así entre los mayores.

Logroño: la mayoría está con Franco, sobre todo a partir de los 35 años, entre los jóvenes la mayoría está en desacuerdo.

Cruce 3 - 27

Variable 27: No a las relaciones prematrimoniales.

Rioja: la mayoría está de acuerdo con la afirmación, las pocas discrepancias se dan entre los jóvenes, pero incluso entre estos la mayoría está de acuerdo.

Logroño: similar al total de La Rioja.

Cruce 3 - 29

Variable 29: Policía debe ser dura.

Rioja: la mayoría está en desacuerdo con la afirmación.
Los viejos no tanto como los jóvenes que lo están mucho más.

Logroño: similar.

Cruce 3 - 30

Variable 30: Rioja autónoma.

Rioja: es otro de los casos en que el ordenador nos dice que no hay valores estimados menores que uno.

$$\chi^2_e = 53'7370 \quad K = 36$$

Con $\alpha = 0'05$, se rechaza que no hay correlación.

La mayoría son autonomistas, sobre todo los jóvenes.

Logroño: similar sin tanta distinción de edades.

Cruce 3 - 32

Variable 32: Sí a la misa.

Rioja: la mayoría opinan que sí, las posturas contrarias, aun siendo minoritarias, se acumulan en los menores de 25 años.

Logroño: similar.

Cruce 3 - 35

Variable 35: No a la huelga.

Rioja: en el conjunto están equilibrados los favorables y los contrarios. Pero se observa que hasta los 35 años hay una mayoría favorable a la huelga y a partir de esa edad hay mayoría en contra.

Logroño: similar.

Cruce 3 - 39

Variable 39: Mejor una república.

Rioja: domina la indiferencia, que aquí habría que calificar con toda seguridad como falta de opinión. Por lo de más hay más monárquicos que republicanos, y de estos últimos, los pocos que hay son muy jóvenes.

Logroño: similar.

Cruce 5 - 6

Variable 5: Profesión.

Variable 6: Clase social.

Rioja: fuerte dominio de las profesiones bajas a causa del gran contingente de jubilados y amas de casa (sus labores) considerados como punto inferior de la escala. También dominan ampliamente las clases medias y bajas. La relación entre las dos variables se ve importante, pero distorionada por lo comentado de amas de casa y jubilados.

Logroño: similar.

Cruce 5 - 7

Variable 7: Estudios.

Rioja: se ve una gran relación entre ambos. El 71'1% no llegan a bachiller elemental, lo que confirma la idea del bajo nivel cultural. Las profesiones altas son las únicas - con una mayor cualificación.

Logroño: similar, pero el porcentaje hasta bachiller elemental baja al 66%.

Cruce 5 - 8

Variable 8: Lugares que conoce.

Rioja: se ve relación. El 53% ha viajado por la nación y el 3% no ha salido de la provincia. De ese tres por ciento, todos menos uno son del nivel profesional más bajo.

Logroño: similar.

Cruce 5 - 17

Variable 17: Años con Franco buenos.

Rioja: la mayoría está con Franco, pero se ve una relación por estar más en desacuerdo los estudiantes y profesionales universitarios.

Logroño: hay menos relación.

Cruce 6 - 7

Variable 6: Clase social.

Variable 7: Estudios.

Rioja: en la clase media-media, el 61% no ha pasado de estudios elementales, en la media-baja, este porcentaje sube al 85% y en la baja al 95%. En la clase media-alta, por el contrario, el 50% ha llegado a la universidad.

Logroño: similar.

Cruce 6 - 8

Variable 8: Lugares que conoce,

Rioja: el 77% de la clase media-media, ha salido por lo menos a la nación, frente al 60% de la media-baja y el 54% de la baja. Sin embargo ha salido el 100% de la media-alta.

Logroño: similar.

Cruce 6 - 17

Variable 17: Años con Franco buenos.

Rioja: ya hemos comentado que la mayoría está con Franco. Solo en la clase media-alta, esta mayoría se reduce al 50%. Pero no se ve mucha relación.

Logroño: la relación es menor.

Cruce 6 - 30

Variable 30: Rioja autónoma.

Rioja: la mayoría son autonomistas con independencia de la clase social.

Logroño: similar.

Cruce 7 - 8

Variable 7: Estudios.

Variable 8: Lugares que conoce.

Rioja: de los que no tienen estudios, nadie ha salido al extranjero, y solo el 9% de los que tienen estudios elementales. De los universitarios por el contrario, han salido al extranjero el 31% y todos han salido por la nación.

Logroño: similar.

Cruce 7 - 17

Variable 17: Años con Franco buenos.

Rioja: la mayoría está con Franco, y la falta de estudios ya la hemos comentado. Hasta los bachilleres Franco ha sido muy bueno, los universitarios no están tan de acuerdo.

Logroño: similar.

Cruce 7 - 30

Variable 30: Rioja autónoma.

Rioja: mayoría autonomista sin distinción de estudios.

Logroño: similar.

Cruce 8 - 17

Variable 8: Lugares que conoce.

Variable 17: Años con Franco buenos.

Rioja: mayoría con Franco, sin que importe mucho lo que han viajado.

Logroño: similar.

Cruce 11 - 17

Variable 11: Sí al divorcio.

Rioja: la mayoría están contra el divorcio y con Franco. Los no franquistas son los más partidarios del divorcio.

Logroño: similar.

Cruce 11 - 27

Variable 27: No a las relaciones prematrimoniales.

Rioja: la mayoría están con la 27 y en contra de la 11. Entre los partidarios del divorcio, la mitad están a favor de las relaciones prematrimoniales.

Logroño: similar, quizá con menos relación.

Cruce 11 - 29

Variable 29: Policía debe ser dura.

Rioja: la mayoría cree que no se debe emplear mano dura. Los muy partidarios del divorcio, están más contra la policía dura, los antidivorcistas están en su mayoría por una policía dura.

Logroño: similar.

Cruce 11 - 32

Variable 32: Sí a la misa.

Rioja: la mayoría están a favor de la misa y contra el divorcio. Solo los muy divorcistas están contra la misa y de forma leve.

Logroño: similar.

Cruce 11 - 35

Variable 35: No a la huelga.

Rioja: hay una pequeña mayoría anti-huelga; se ve relación, pues los antidivorcistas son anti-huelga y viceversa.

Logroño: similar.

Cruce 11 - 39

Variable 39: Mejor una república.

Rioja: ya comentamos que hay mayoría indiferente, pe

ro por lo demás hay más monárquicos. Los que están contra el divorcio son monárquicos. Solo los muy divorcistas son republicanos.

Logroño: similar, pero también son republicanos los bastante divorcistas.

Cruce 12 - 18

Variable 12: Sí a la libertad para las regiones.

Variable 18: Desde Madrid se controla demasiado.

Rioja: la mayoría está de acuerdo con ambas preguntas.

Logroño: similar.

Cruce 12 - 19

Variable 19: Sí a las cooperativas.

Rioja: la mayoría está con las cooperativas, pero de forma más clara que en la pregunta 12.

Logroño: similar.

Cruce 12 - 23

Variable 23: Sí a las nacionalizaciones.

Rioja: la mayoría está con las nacionalizaciones, pero hay un 25% de indiferentes. La mayor parte de la población se agrupa en valores altos de ambas variables.

Logroño: similar.

Cruce 17 - 19

Variable 17: Años con Franco buenos.

Variable 19: Sí a las cooperativas.

Rioja: No hay relación.

Logroño: Tampoco.

Cruce 17 - 24

Variable 24: Rioja unida.

Rioja: no hay relación. Es otro de los casos en que es válido el criterio de χ^2 , pues el ordenador nos dice que no hay valores estimados menores que uno.

$$\chi_e^2 = 41'4438 \quad K = 36$$

$$\chi_{0.05}^2 = 51$$

$$k=36$$

Luego se acepta que sigue la teórica.

Logroño: similar, pero sin ser válido utilizar χ^2 .

Cruce 17 - 27

Variable 27: No a las relaciones prematrimoniales.

Rioja: la mayoría de la población está con Franco y contra las relaciones prematrimoniales. Incluso entre los antifranquistas hay una leve mayoría a favor de la pregunta 27.

Logroño: similar.

Cruce 17 - 29

Variable 29: Policía debe ser dura.

Rioja: empate en la pregunta 29. Los muy franquistas están a favor de la 29, los antifranquistas en contra. Los demás empatados.

Logroño: alguna mayoría contra la policía dura, se ve relación con la pregunta 17, pero menos. Hay mayor tendencia contra la policía dura, con mayor independencia de la opinión sobre Franco.

Cruce 17 - 30

Variable 30: Rioja autónoma.

Rioja: mayoría franquista y autonomista. Más autonomistas entre los antifranquistas.

Es otro de los casos donde es significativa χ^2 .

$$\chi^2_e = 55.7570 \quad K = 36$$

$$\chi^2_{0.05} = 51$$

k:36

Luego se rechaza que siga la teórica, hay correlación.

Logroño: similar, pero χ^2 no es significativa.

Cruce 17 - 32

Variable 32: Sí a la misa.

Rioja: solo los muy antifranquistas están contra la mi
sa.

Logroño: los antifranquistas están empatados en la -
pregunta 32.

Cruce 17 - 35

Variable 35: No a la huelga.

Rioja: mayoría franquista y ligera mayoría anti-huell
ga. Los antifranquistas están con la huelga, solo los franq
uistas claros están totalmente contra la huelga.

Logroño: similar, pero en la pregunta 35 hay empate
y también menor relación.

Cruce 17 - 39

Variable 39: Mejor una república.

Rioja: se ve relación, cuanto más a favor de Franco,
más monárquicos. Los antifranquistas son republicanos.

Logroño: entre los antifranquistas hay empate en la 39.

Cruce 18 - 19

Variable 18: Madrid controla demasiado.

Variable 19: Sí a las cooperativas.

Rioja: la mayoría de acuerdo en ambas preguntas. Son
más cooperativistas los que opinan que Madrid controla demaq

siado, se ve la relación.

Logroño: similar, pero menos relación al ser casi to
dos cooperativistas.

Cruce 18 - 23

Variable 23: Sí a las nacionalizaciones.

Rioja: la mayoría están de acuerdo con ambas preguntas.
Pero los que opinan que se controla demasiado, tienen también
mayor proporción de acuerdo con las nacionalizaciones.

Logroño: hay menos relación.

Cruce 19 - 23

Variable 19: Sí a las cooperativas.

Rioja: mayoría de acuerdo con ambas preguntas; habiendo
más cooperativistas entre los nacionalizadores.

Logroño: similar, pero más acusado.

Cruce 19 - 30

Variable 30: Rioja autónoma.

Rioja y Logroño: hay poca o nula relación.

Cruce 24 - 30

Variable 24: Rioja unida a otras.

Rioja: es uno de los casos en que χ^2 es significativa.

$$\chi^2_e = 480'1622 \quad K = 36$$

$$\chi^2_{0'05} = 51 \\ k=36$$

$$\chi^2_{0'01} = 58'6 \\ k=36$$

Luego hay gran relación, como ya sabíamos por los es tudios anteriores. La mayoría es autonomista.

Logroño: similar.

Cruce 27 - 29

Variable 27: No a las relaciones prematrimoniales.

Variable 29: Policía debe ser dura.

Rioja: los que están contra la 27, lo están también contra la 29. Los que están con la 27, empatan en la 29.

Logroño: similar.

Cruce 27 - 32

Variable 32: Sí a la misa.

Rioja y Logroño: existe la lógica relación.

Cruce 27 - 35

Variable 35: No a la huelga.

Rioja: los que están contra la 27, están también con tra la 35. Los que están con la 27, en una leve mayoría es-

tán con la 35.

Logroño: similar.

Cruce 27 - 39

Variable 39: Mejor una república.

Rioja: los que están con la 27, son monárquicos. Los que están contra la 27, son levemente republicanos.

Logroño: similar.

Cruce 29 - 32

Variable 29: Policía debe ser dura.

Variable 32: Sí a la misa.

Rioja: los que están con la 29, están con la 32. Los que están contra la 29, están en parte contra la 32.

Logroño: similar.

Cruce 29 - 35

Variable 35: No a la huelga.

Rioja: los que están con la 29, lo están con la 35. So lo los muy en contra de la 29, están a favor de la huelga.

Logroño: similar, pero con más gente a favor de la huelga.

Cruce 29 - 39

Variable 39: Mejor una república.

Rioja: los que están con la 29, son monárquicos. Los que están totalmente contra la 29, son republicanos. En los demás dominan los monárquicos.

Logroño: similar.

Cruce 32 - 35

Variable 32: Sí a la misa.

Variable 35: No a la huelga.

Rioja: los que están totalmente a favor de la 32, es tán a favor de la 35. Los que están levemente a favor de la 32, están en una leve mayoría con la 35. Los demás contra - la 35.

Logroño: similar.

Cruce 32 - 39

Variable 39: Mejor una república.

Rioja: los que están con la 32, son monárquicos. Solo los que están claramente contra la 32, son levemente republicanos.

Logroño: similar.

Cruce 35 - 39

Variable 35: No a la huelga.

Rioja: los que están con la 35, son monárquicos y solamente los que están claramente contra la 35 son levemente republicanos.

Logroño: similar.

Terminamos aquí este breve resumen, que como ya hemos indicado, solamente pretende que el lector tenga una idea de los cruces de variables. Un estudio más completo nos alargaría tremendamente y nos sacaría de nuestros objetivos.

ANALISIS FACTORIAL 4.

Una vez que hemos dado por concluido el estudio sobre Logroño, creo que ha llegado el momento de realizar un estudio completo sobre toda La Rioja. Para ello y con todo el conocimiento que hemos acumulado, realizaremos primero el análisis factorial 4, con los 267 datos y después haremos el definitivo análisis factorial 5, que será el que estudiaremos.

Respecto al tercer análisis factorial, vamos a realizar los siguientes cambios en las variables:

Introducimos las variables 14 (Dios existe), 26 (Dios es un consuelo), 37 (sí a la industrialización) y 40 (Rioja discriminada).

Con lo que llegamos nuevamente al tope de las 29 variables que nos permite nuestro programa de ordenador.

La razón de estas introducciones es estudiar para el total de La Rioja el factor religioso y más problemas industriales y regionalistas.

Tengamos en cuenta que al menos a priori, parece lógico que el factor religioso debe aparecer con estas dos nuevas variables (14 y 26) y que el problema de la industrialización es interesante en una zona donde la agricultura es fundamental.

En último lugar, sobre la discriminación de la Rioja, habría mucho que comentar, pues aunque en principio no se debería hablar de ella en una zona relativamente rica, sobre todo en el aspecto agrícola, hay que tener en cuenta que el estar muy cerca de provincias tradicionalmente forales como son Alava y Navarra, ha perjudicado notablemente a la industria Riojana (excepción hecha de la Rioja Alavesa), que en buena parte ha emigrado en busca de ventajas fiscales.

Ante la introducción de las nuevas variables, los números equivalentes de ordenador quedan:

<u>VARIABLE</u>	<u>Nº DE ORDENADOR</u>
5	1
6	2
7	3
8	4
10	5
11	6
12	7
14	8
15	9
17	10
18	11
19	12
20	13
23	14
24	15

26	16
27	17
28	18
29	19
30	20
32	21
33	22
35	23
37	24
38	25
39	26
40	27
41	28
42	29

Los resultados de ordenador, están en las hojas de ordenador nº 850 y ss. bajo la denominación de Análisis Factorial Rioja 4.

Las condiciones del programa de ordenador son similares a las del anterior análisis factorial, es decir, entrada de variables normalizadas para poder interpretar las puntuaciones, etc.

Pero con la innovación de rechazar los factores que expliquen menos del 3'45% (1), esto se debe a que así tendremos más factores que poniendo el 5%, lo que nos proporcionará más elementos de juicio.

Como vemos en las hojas de ordenador, logramos que se explique el 61'537% de la varianza total con nueve factores. Este incremento de explicación se debe fundamentalmente a tener más factores.

Vamos a ver la interpretación de los factores antes de la rotación:

Factor 1

Valor propio: 6'60773

Porcentaje de explicación: 22'785%

Variables muy importantes:

- 17 - Años con Franco buenos (negativa).
- 29 - Policía debe ser dura (negativa).
- 33 - La mujer en casa (negativa).
- 35 - No a la huelga (negativa).

Variables importantes:

- 5 - Profesión.
- 7 - Estudios.
- 11 - Sí al divorcio.

(1) $0'0345 \cdot 29 = 1'0005$; con lo que mantenemos el mínimo exigido por Kaiser (1.960).

- 26 - Dios es un consuelo (negativa).
- 27 - No a las relaciones prematrimoniales (negativa).
- 28 - Películas españolas buenas (negativa).
- 32 - Sí a la misa (negativa).
- 38 - La mujer debe trabajar fuera.

Variables secundarias:

- 8 - Lugares que conoce.
- 10 - Los licenciados están preparados (negativa).
- 12 - Sí a la libertad a las regiones.
- 14 - Dios existe (negativa).
- 15 - Los niños no deben saber del sexo (negativa).
- 39 - Mejor una república.
- 41 - Las mujeres son poco inteligentes (negativa).
- 42 - Servicio militar voluntario.

Variables terciarias (solo citar): 6, 18.

Sigue siendo muy similar este factor a los obtenidos en los anteriores análisis antes de la rotación, pero con algunas novedades interesantes, entre las cuales creo que es preciso destacar dos, fundamentalmente. En primer lugar, la tendencia cultural del factor, reflejada directamente por las preguntas 5 y 7 e indirectamente, por las 28, 10 e incluso la 41.

Por otro lado, la introducción de preguntas de marcado carácter religioso, como son la 26 y la 14, indican que este factor se identifica con una cierta arreligiosidad, -

aunque hemos de tener en cuenta que el valor negativo de la pregunta 14 (Dios realmente existe), es bastante bajo y por ello está encuadrada como variable de tipo secundario.

En conclusión, este factor es el de siempre, es decir, el "liberal-progresista", pero con importantes matices culturales y algunos arreligiosos.

Factor 2

Valor propio: 2'08824

Porcentaje de explicación: 7'201%

Variables muy importantes:

- 24 - La Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables importantes: (Carece)

Variables secundarias:

- 6 - Clase social.
- 14 - Dios existe.
- 20 - Los curas viven bien (negativa).

Variables terciarias (solo citar): 7, 8, 23, 26, 32, -37.

Este factor sigue siendo el "antiautonomista" de siempre, pero con ciertas connotaciones religiosas y de estatus social que esperamos se aclaren tras la rotación.

Factor 3

Valor propio: 1'93921

Porcentaje de explicación: 6'687%

Variables muy importantes: (Carece)

Variables importantes:

- 24 - Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables secundarias:

- 8 - Lugares que conoce (negativa)
- 14 - Dios existe (negativa).
- 20 - Los curas viven bien.
- 39 - Mejor una república.

Variables terciarias (solo citar): -6, -7, 23, -26, -32.

Factor parecido en cierta forma al anterior, pero con menos peso proporcional del antiautonomismo y signo contrario en las connotaciones religiosas y de estatus socio-cultural, a parte de la aparición de un cierto republicanismo. En realidad es de muy difícil interpretación y esperemos que se aclare tras la rotación.

Factor 4

Valor propio: 1'6528

Porcentaje de explicación: 5'699%

Variables muy importantes:

- 19 - Sí a las cooperativas.

Variables importantes: (Carece)

Variables secundarias:

- 12 - Sí a la libertad para las regiones.
- 18 - Madrid controla demasiado.
- 20 - Los curas viven bien.
- 23 - Sí a las nacionalizaciones.
- 37 - Sí a la industrialización.

Variables terciarias (solo citar): 10, 14, 26, 28.

Fundamentalmente es un factor de cooperativismo, con connotaciones regionalistas. También está a favor de las nacionalizaciones y la industrialización. A parte de estar con la pregunta 20. Veremos como queda tras la rotación.

Factor 5

Valor propio: 1'20944

Porcentaje de explicación: 4'170%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

- 40 - Rioja discriminada.
- 41 - La mujer es poco inteligente (negativa).

Variables terciarias (en este caso vamos a ponerlos con significado incluido para intentar interpretar el factor):

- 10 - Los licenciados están preparados.
- 11 - Sí al divorcio.
- 12 - Sí a la libertad para las regiones (negativa).
- 14 - Dios existe.
- 23 - Sí a las nacionalizaciones (negativa).
- 26 - Dios es un consuelo.
- 32 - Sí a la misa.
- 33 - La mujer en casa (negativa).
- 37 - Sí a la industrialización.
- 38 - La mujer debe trabajar fuera.

Este factor es de muy difícil interpretación, pues - carece de variables de importancia, de todas formas, en una primera aproximación lo pudiéramos calificar de riojanista, feminista y religioso, fundamentalmente, pero esperemos que se aclare tras la rotación.

Factor 6

Valor propio: 1'17743

Porcentaje de explicación: 4'061%

Variables muy importantes: (Carece)

Variables importantes:

- 40 - La Rioja discriminada (negativa).

Variables secundarias:

42 - Servicio militar voluntario.

Variables terciarias (solo citar): -5, -6, 11, -12, -18,
-20, -37.

Este factor tiene parecidos problemas que los anteriores. Podemos calificarlo fundamentalmente de No riojanista y partidario del servicio militar voluntario. Esperemos que la cosa mejore tras la rotación.

Factor 7

Valor propio: 1'09101

Porcentaje de explicación: 3'762%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

6 - Clase social (negativa).

39 - Mejor una república (negativa).

Variables terciarias: *(Incluimos su significado)*

7 - Estudios (negativa).

15 - Los niños no deben saber del sexo (negativa).

19 - Sí a las cooperativas.

23 - Sí a las nacionalizaciones (negativa).

29 - La policía debe ser dura (negativa).

- 35 - No a la huelga (negativa).
- 42 - Servicio militar voluntario (negativa).

Nuevamente nos encontramos con otro factor de difícil interpretación, que esperemos mejore tras la rotación. Podría mos resumirlo como clase social baja, monárquico y algo libe ral.

Factor 8

Valor propio: 1'06105

Porcentaje de explicación: 3'658%

Variables muy importantes: (Carece)

Variables importantes:

- 37 - Sí a la industrialización.

Variables secundarias:

- 40 - Rioja discriminada (negativa).

Variables terciarias (solo citar): 6, -12, -15, 38, -42.

El caso es similar a los anteriores en cuanto a su di ficultad. Lo podemos calificar de industrialista, no rioja - nista.

(Al decir riojanista no pretendemos hacer ningún juicio de valor, es una forma de expresar que considera que la Rioja ha sido discriminada).

Factor 9

Valor propio: 1'01887

Porcentaje de explicación: 3'514%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

12 - Sí a la libertad para las regiones (negativa).

18 - Madrid controla demasiasdo.

Variables terciarias: (incluimos su significado)

10 - Los licenciados están preparados (negativa).

20 - Los curas viven bien (negativa).

23 - Sí a las nacionalizaciones.

28 - Las películas españolas buenas (negativa).

33 - La mujer debe estar en casa.

37 - Sí a la industrialización.

Este factor es ya completamente incongruente, y esperemos que mejore al rotar. La causa de esta situación es - que ya son factores residuales de muy poco valor propio.

En conclusión, el factor uno, que es antes de la rotación el que mejor discrimina a la población, es muy similar a los estudios anteriores, pero con las connotaciones culturales y arreligiosas citadas.

Por otro lado, al tener muchos factores ha aumentado el

porcentaje de explicación, pero también los últimos son de muy difícil interpretación. La rotación es totalmente necesaria y veremos que resultados nos depara.

Factores rotados:

Factor 1

Capacidad explicativa: 3'74593

Porcentaje de explicación: 12'917%

Variables muy importantes:

- 11 - Sí al divorcio.
- 33 - La mujer en casa (negativa).
- 38 - La mujer debe trabajar fuera.

Variables importantes:

- 15 - Los niños no deben saber del sexo (negativa).
- 17 - Los años con Franco buenos (negativa).
- 29 - La policía debe ser dura (negativa).
- 35 - No a la huelga (negativa).
- 41 - La mujer es poco inteligente (negativa).

Variables secundarias:

- 27 - No a las relaciones prematrimoniales (negativa).

Variables terciarias (solo citar): 5, 7, 8, -26, -28,
-32, 39, 42.

Este factor es enormemente claro, por un lado es fuertemente feminista y en segundo grado, bastante liberal. Luego lo podríamos denominar: "feminista-liberal".

Factor 2

Capacidad explicativa: 2'03116

Porcentaje de explicación: 7'004%

Variables muy importantes:

- 24 - La Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables importantes: (Carece)

Variables secundarias: (Carece)

Variables terciarias (solo citar): 23, -42.

Este factor es también muy claro, es el "antiautonomista" prácticamente puro.

Factor 3

Capacidad explicativa: 2'12831

Porcentaje de explicación: 7'339%

Variables muy importantes:

- 6 - Clase social (negativa).
- 7 - Estudios (negativa).

Variables importantes:

- 5 - Profesión (negativa).

Variables secundarias:

- 8 - Lugares que conoce (negativa).

Variables terciarias (solo citar): 10, -12, 28, 33, -38,
41.

Este factor ha cambiado mucho al rotar y ahora es clarísimo y lo podríamos definir como "bajo estatus socio-cultural".

Factor 4

Capacidad explicativa: 1'30935

Porcentaje de explicación: 4'517%

Variables muy importantes:

- 20 - Los curas viven bien.

Variables importantes:

- 12 - Sí a la libertad para las regiones.

Variables secundarias:

- 10 - Los licenciados están preparados.

Variables terciarias (solo citar): 19, 28, -32.

Es este un factor de muy difícil interpretación, pues la pregunta "los curas viven bien", es difícil de concretar en cuanto a su significado. Podíamos interpretarla como anti

clericalismo, y quizá sea ese uno de los componentes importantes; pero en la situación actual donde los curas tienen unos ingresos que en muchos casos son casi miserables, también habrá que reconocer un desconocimiento de la situación muy amenudo relacionado con incultura.

Por otro lado, esta afirmación conlleva a su vez un cierto inconformismo, pues el clero está en muchas ocasiones identificado con el poder, todavía más en ambientes rurales.

Este último aspecto se relacionaría más con la pregunta 12, mientras que la incultura se reflejaría en cierta forma en la 10. Tengamos en cuenta que el considerar que - un licenciado, por el hecho de serlo, está bien preparado, indica un desconocimiento de la universidad actual.

En definitiva, el factor, aún siendo confuso, lo podemos calificar como indicador de cierto anticlericalismo, cierta incultura, y cierto inconformismo.

Factor 5

Capacidad explicativa: 2'96264

Porcentaje de explicación: 10'216%

Variables muy importantes:

14 - Dios existe.

26 - Dios es un consuelo.

VARIABLES IMPORTANTES:

- 17 - Los años con Franco buenos.
- 27 - No a las relaciones prematrimoniales.
- 32 - Sí a la misa.

VARIABLES SECUNDARIAS:

- 5 - Profesión (negativa).
- 10 - Los licenciados están preparados.
- 28 - Las películas españolas buenas.

VARIABLES TERCIARIAS (SOLO CITAR): -11, -20, 29, 33,
35, -39.

El factor es bastante claro. Lo podríamos denominar fundamentalmente religioso, pero son importantes los matices tradicionalistas y los representativos de falta de cultura.

Luego este factor lo denominaríamos "religioso-tradicionalista".

Factor 6

Capacidad explicativa: 1'17479

Porcentaje de explicación: 4'051%

VARIABLES MUY IMPORTANTES:

- 40 - La Rioja discriminada (negativa).

Variables importantes: (Carece)

Variables secundarias: (Carece)

Variables terciarias (solo citar): 19, 29.

Este factor, prácticamente coincide con la variable 40, pero con signo contrario. Esto quiere decir que la variable 40, no forma grupo con otras variables y, por lo tanto, que no está correlacionada con otras.

A la vista de esto, habrá que suprimir la variable 40 en el próximo análisis factorial. (1)

Esta es la ventaja de tener muchos factores, así podemos eliminar variables que están solas. Sin embargo, tantos factores nos crearían problemas a la hora de interpretar resultados.

Factor 7

Capacidad explicativa: 1'479

Porcentaje de explicación: 5'1%

Variables muy importantes:

- 39 - Mejor una república. (negativa).
- 42 - Servicio militar voluntario (negativa).

(1) Pues de interesarnos esta variable, la estudiaríamos concretamente y no como factor.

Variables importantes:

23 - Sí a las nacionalizaciones (negativa).

Variables secundarias: (Carece)

Variables terciarias (solo citar): -11.

Este es un factor muy interesante y que anteriormente no nos había salido en ningún otro análisis. He aquí otra de las ventajas de tener muchos factores.

Este factor une un sentimiento monárquico, con un militarismo (entendido como tal la necesidad de mantener el servicio militar obligatorio), e incluso con una aversión a las nacionalizaciones. Creo que podemos decir que este factor incluye, en primer lugar, el militarismo monárquico y después, un sentimiento capitalista, y en conclusión, lo podemos definir como: "monárquico, militarista, capitalista". Que es una ideología bastante coherente.

Factor 8

Capacidad explicativa: 1'31196

Porcentaje de explicación: 4'524%

Variables muy importantes:

37 - Sí a la industrialización.

Variables importantes: (Carece)

Variables secundarias:

- 8 - Lugares que conoce (negativa).

Variables terciarias (solo citar): -12, 18, 20, 28, 38.

Nos encontramos en un caso parecido al del factor seis. La variable de industrialización, casi comprende todo el factor. Por lo tanto, esta variable habrá que suprimirla en futuros análisis, dado que no forma ningún grupo.

De todas formas en este factor de industrialización, tiene algún interés la variable 8 (lugares que conoce), lo que, dado su signo negativo, nos indica que no se ha viajado. Podemos apreciar pues la relación entre lo que se ha viajado y el deseo de industrialización, que es negativa.

Factor 9

Capacidad explicativa: 1'69969

Porcentaje de explicación: 5'861%

Variables muy importantes: (Carece)

Variables importantes:

- 18 - Madrid controla demasiado.
- 19 - Sí a las cooperativas.

Variables secundarias:

- 23 - Sí a las nacionalizaciones.

Variables terciarias (solo citar): 7, 8, -10, 12, -15,
-17, -29, -35.

Este factor une el rechazo al excesivo centralismo y el cooperativismo agrario. Por otro lado, acepta las nacionalizaciones. Este conjunto de ideas no son tan dispares si las analizamos detenidamente; en primer lugar, tenemos un rechazo al capitalismo, no frontal por el tono de las afirmaciones, pero sí desde el punto de vista que acepta, e incluso prefiere, otras soluciones distintas a la empresa capitalista tradicional.

En segundo lugar tenemos un rechazo al centralismo. El factor lo podíamos denominar cooperativista o anticapitalista, por un lado, y anticeutralista por otro.

Creo que es interesante que el lector se de cuenta que estas dos ideas tienen en común una idea anterior, que es el deseo de que el poder decisorio se acerque al pueblo (que no se decida en Madrid, que se decida en cooperativa), o que por lo menos se reste poder a unos pocos (nacionalizando las grandes empresas).

Considero que podemos dar por interpretado este análisis factorial, tanto en su matriz original, como en la rotada. Las conclusiones que podemos sacar son fundamentalmente, que al haber conseguido más factores y al introducir nuevas variables, hemos ampliado nuestro conocimiento sobre la muestra, pero que hay factores de difícil interpretación, o que

expli

explican muy poco (en algunos casos son prácticamente de una variable). Por otro lado, son demasiados factores como para hacer un estudio coherente y completo.

Por lo tanto, aún siendo perfectamente válidas las puntuaciones, no vamos a interpretarlas. Lo que vamos a hacer es un nuevo análisis factorial en el que depuremos la situación actual.

ANALISIS FACTORIAL 5.

Con todos los conocimientos acumulados hasta el momento, fundamentalmente con la experiencia del estudio completo sobre Logroño capital y el análisis factorial 4 sobre La Rioja en términos globales, podemos intentar ya un estudio completo sobre La Rioja.

Este estudio no se podrá hacer de manera tan exhaustiva como el de Logroño capital por limitaciones que comentaremos en su momento, pero si lo haremos tan completo como podamos.

En primer lugar, creo que es interesante destacar que la utilización del que denominábamos individuo indiferente (I.I.), nos resolvió muchos de los problemas de indeterminación al tener los factores promedio cero, pero todavía podemos conseguir mejores resultados.

En los análisis realizados hasta ahora, el método para obtener las puntuaciones fue el que comentábamos al comienzo del capítulo dos de esta segunda parte. Este método es perfectamente válido y tiene todas las ventajas inherentes al manejo de factores normalizados. Pero para este análisis, que pretendo sea el definitivo, vamos a utilizar el método citado por Jöreskog, Klovan y Reymont (1), que consiste en la

(1) Jöreskog, Klovan y Reymont (1.976), pags. 68 y ss.

obtención de las puntuaciones con varianza igual al valor propio de cada factor, esto se hace mediante la formulación:

$$F = T' U' X$$

donde las U, vectores propios, están normalizadas, (este es el método que utilizamos en la exposición teórica, teniendo en cuenta además que realizaremos una rotación).

De esta manera, dado que los factores tienen una varianza igual a su valor propio correspondiente, tendremos importantes ventajas como veremos. (1)

Por otro lado, decidimos suprimir las variables: 20 (- los curas viven bien), a causa de su difícil interpretación, 37 (sí a la industrialización) y 40 (La Rioja discriminada); estas dos últimas por las causas expuestas en el anterior análisis factorial.

Con esto, la numeración equivalente de ordenador queda:

<u>VARIABLE</u>	<u>Nº DE ORDENADOR</u>
5	1
6	2
7	3
8	4
10	5
11	6
12	7

(1) En realidad tras la rotación, la varianza será igual a la capacidad explicativa.

14	8
15	9
17	10
18	11
19	12
23	13
24	14
26	15
27	16
28	17
29	18
30	19
32	20
33	21
35	22
38	23
39	24
41	25
42	26

Los resultados de ordenador, están en las hojas de ordenador nº 892 y ss. bajo la denominación de Análisis - Factorial Rioja 9 (5).

En cuanto a las condiciones del programa de ordenador, son similares a las de otros programas (entrada de variables normalizadas, etc.). Excepto la modificación ya citada para la obtención de las puntuaciones y que el valor propio míni-

mo lo dejamos en un 4%, la causa de esto es que el 3'45% del análisis anterior se ha demostrado que es bajo y el 5% nos impedía el estudio de ciertos factores que pueden resultar interesantes.

Como vemos en las hojas de ordenador, con seis factores, conseguimos explicar el 54'636%, resultado que es bastante satisfactorio.

Veamos a continuación la interpretación de los factores antes de la rotación.

Factor 1

Valor propio: 6'6

Porcentaje de explicación: 25'384%

Variables muy importantes:

- 17 - Años con Franco buenos (negativa).
- 33 - La mujer debe estar en casa (negativa).
- 35 - No a la huelga (negativa).

Variables importantes:

- 5 - Profesión.
- 7 - Estudios.
- 11 - Sí al divorcio.
- 26 - Dios es un consuelo (negativa).
- 27 - No a las relaciones prematrimoniales (negativa).
- 28 - Las películas españolas buenas (negativa).

- 29 - La policía debe ser dura (negativa).
- 32 - Sí a la misa (negativa).
- 38 - La mujer debe trabajar fuera.

Variables secundarias:

- 8 - Lugares que conoce.
- 10 - Los licenciados están preparados (negativa).
- 12 - Sí a la libertad para las regiones.
- 14 - Dios existe (negativa).
- 15 - Los niños no deben saber del sexo (negativa).
- 39 - Mejor una república.
- 41 - Las mujeres son poco inteligentes (negativa).
- 42 - Servicio militar voluntario.

Variables terciarias (solo citar): 6, 18.

Este factor sigue siendo similar al de anteriores análisis factoriales, pero tiene ciertas peculiaridades que creo que hay que destacar:

En primer lugar su semejanza con el correspondiente al análisis factorial 4, pues prácticamente nos encontramos ante el mismo factor, cosa totalmente lógica dado que las variables son fundamentalmente las mismas en ambos análisis factoriales.

Por otro lado, quiero comentar aquí que la identificación que hace este factor del liberalismo y la arreligiosidad (al igual que sucedía en al análisis factorial 4), no debe -

llevarnos a engaño y hacernos pensar que la identificación es clara entre ambos conceptos. Tengamos en cuenta que la variable 26 (Dios es un consuelo) que aparece como variable importante y negativa en este factor, es muy posible - que lo que haga sea identificar el liberalismo-progresismo con la huida del concepto de Dios como consuelo para problemas humanos, es decir, rechazo del concepto utilitarista - de Dios.

Por otro lado, respecto a la pregunta 14 (Dios existe) que aparece como variable secundaria y negativa, dado que el promedio de esta variable es 6'42322, el más alto, - es decir, el que suscita un mayor grado de acuerdo entre la muestra, y su desviación típica es de 1'14685, que es una - de las más bajas, y por lo tanto, una de las preguntas de menor dispersión, hay que deducir que el creer en Dios es bastante mayoritario entre la muestra. El que el coeficiente de la variable 14 sea negativo, dado que las variables están normalizadas al entrar en el programa de análisis factorial, quiere decir que este factor se identifica algo con la duda, o la creencia no muy firme, más que con la no creencia en - Dios.

Todo lo que hemos comentado respecto a la variable 14, se puede aplicar en gran parte también a la 26.

En conclusión, es un factor liberal-progresista, con importantes matizaciones culturales y relativa arreligiosidad.

Factor 2

Valor propio: 2'01251

Porcentaje de explicación: 7'741%

Variables muy importantes:

- 24 - Rioja unida a otras.
- 30 - Rioja autónoma (negativa).

Variables importantes: (Carece)

Variables secundarias:

- 23 - Sí a las nacionalizaciones.

Variables terciarias (solo citar): 19, 26.

Es un factor muy similar al que ha resultado en análisis precedentes, por lo tanto, no lo comentamos más. Es el factor de la "No autonomía".

Factor 3

Valor propio: 1'85711

Porcentaje de explicación: 7'142%

Variables muy importantes: (Carece)

Variables importantes:

- 14 - Dios existe.

Variables secundarias:

- 6 - Clase social.

- 7 - Estudios.
- 8 - Lugares que conoce.
- 26 - Dios es un consuelo.

Variables terciarias (solo citar): 12, 18, 19, -24, 27,
30, 32, -39.

Este factor no lo podemos considerar como completamente definido, será tras la rotación cuando probablemente sea más claro en cuanto a su significado.

De todas formas hay dos puntos que aparecen como fundamentales. En primer lugar existe una religiosidad (entendiendo como tal el estar de acuerdo con las variables 14 y -26). Por otro lado, esto se identifica con un elevado estatus socio cultural, representado éste por las variables 6, 7 e incluso por la 8. Para terminar, podemos observar que entre las variables terciarias están la -24 y la 30, lo que da un cierto autonomismo.

Luego, en definitiva, podemos calificar este factor - como religioso, estatus alto y quizá algo de autonomía.

Factor 4

Valor propio: 1'47342

Porcentaje de explicación: 5'667%

Variables muy importantes: (Carece)

Variables importantes:

- 19 - Sí a las cooperativas.

Variables secundarias:

- 6 - Clase social (negativa).
- 23 - Sí a las nacionalizaciones.

Variables terciarias (solo citar): -5, -7, 10, 12, 18,
28, 39, 42.

Este factor en principio, es cooperativista, pero luego vemos que se mezclan gran cantidad de variables de difícil interpretación; en primer lugar, la clase social baja y el apoyo a las nacionalizaciones, pero por otro lado, son interesantes en este caso las variables terciarias que vamos a estudiar a continuación:

- 5 - Profesión (negativa).
- 7 - Estudios (negativa).
- 10 - Los licenciados están preparados.
- 12 - Sí a la libertad para las regiones.
- 18 - Madrid controla demasiado.
- 28 - Las películas españolas son buenas.
- 39 - Mejor una república.
- 42 - Servicio militar libre.

Vemos en conclusión la existencia, como comentábamos en el anterior análisis factorial, de un deseo de que los poderes de decisión se acerquen al pueblo y dejen los círculos privilegiados, mezclado en este caso con un estatus socio-cull

tural bajo, e incluso con un inconformismo reflejado en las preguntas 39 y 42.

Luego, en definitiva, tenemos en este factor una serie de conceptos, que aún siendo difíciles de resumir en una palabra, podemos tratar de esquematizar:

- a) Las decisiones económicas se deben acercar al pueblo, bien mediante las cooperativas, bien al menos quitándoselas a los grupos poderosos, nacionalizando.
- b) Las decisiones políticas hay que acercarlas, intentando evitar el centralismo.
- c) Estatus socio-cultural bajo.
- d) Inconformismo que se traduce en preferir la república y no estar con el servicio militar obligatorio.

Por supuesto que todo esto no podemos afirmarlo con rotundidad, pues ya hemos comentado que los valores que toman las variables al entrar en el análisis factorial están normalizados.

De todas formas, creo que este resumen de ideas, puede aclarar el sentido del factor, que será de más fácil interpretación tras la rotación.

Factor 5

Valor propio: 1'17406

Porcentaje de explicación: 4'516%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

- 38 - La mujer debe trabajar fuera.
- 41 - La mujer es poco inteligente (negativa).

Variables terciarias: (en este caso las comentamos por estar el factor poco definido hasta el momento, dado que carece de variables de importancia).

- 10 - Los licenciados están preparados.
- 11 - Sí al divorcio.
- 12 - Sí a la libertad para las regiones (negativa).
- 14 - Dios existe.
- 18 - Madrid controla demasiado (negativa).
- 23 - Sí a las nacionalizaciones (negativa).
- 26 - Dios es un consuelo.
- 32 - Sí a la misa.
- 33 - La mujer debe estar en casa (negativa).
- 42 - Servicio militar libre.

Nuevamente nos encontramos ante un factor que mezcla muchos conceptos y que esperemos que se concrete al rotar.

Tal como lo tenemos en este momento, hay un fuerte componente de feminismo, que es lo fundamental, pero hay también importantes connotaciones religiosas, centralistas e incluso un rechazo a las nacionalizaciones, luego el factor hay que considerarlo bastante confuso de momento, aunque indudablemente tiene su lógica interna.

Factor 6

Valor propio: 1'08845

Porcentaje de explicación: 4'186%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

- 6 - Clase social.
- 23 - Sí a las nacionalizaciones.
- 39 - Mejor una república.
- 42 - Servicio militar voluntario.

Variables terciarias: (vemos su significado)

- 7 - Estudios.
- 15 - Los niños no deben saber del sexo.
- 19 - Sí a las cooperativas (negativa).
- 35 - No a la huelga.

Factor muy confuso que representa una forma de pensar un tanto extraña, podríamos buscarle su lógica, pero en este

caso resulta especialmente difícil y quizá sea mejor esperar a los resultados de la rotación.

En conclusión de este análisis factorial, antes de rotar, hemos de decir que el factor uno, que es el que mejor discrimina a la población, sigue siendo muy similar a estudios anteriores.

Por lo demás nos hemos encontrado con abundantes casos de factores de difícil interpretación, vamos a ver lo que sucede al rotar.

Factores rotados:

Factor 1

Capacidad explicativa: 3'02588

Porcentaje de explicación: 11'638%

Variables muy importantes:

- 14 - Dios existe (negativa).
- 26 - Dios es un consuelo (negativa).
- 32 - Sí a la misa (negativa).

Variables importantes:

- 17 - Años con Franco buenos (negativa).
- 27 - No a las relaciones prematrimoniales (negativa).

Variables secundarias:

- 10 - Los licenciados están preparados (negativa).

28 - Películas españolas buenas (negativa)

Variables terciarias (solo citar): 5, 11, -29, -33,
-35, 39.

Este factor ha cambiado bastante al rotar, pero es un factor muy claro y realmente interesante. Lo podemos identificar con el antirreligioso (aunque como ya hemos comentado, será una antirreligiosidad relativa, como todo lo demás al ser variables normalizadas, pero más en este caso, dada la religiosidad que se aprecia al estudiar las variables concretas).

También tiene matices antitradicionalistas y culturales. Pero fundamentalmente es el factor antirreligioso.

Factor 2

Capacidad explicativa: 2'01994

Porcentaje de explicación: 7'769%

Variables muy importantes:

24 - La Rioja unida a otras.

30 - Rioja autónoma (negativa).

Variables importantes: (Carece)

Variables secundarias: (Carece)

Variables terciarias (solo citar): 23, -42.

Este es el factor antiautonomista y no merece en principio mayor comentario, dado que es prácticamente igual que en estudios precedentes.

De todas formas, quiero hacer constar aquí que, aparte de lo que ya comentamos de que las variables 24 y 30 tienen sus propios matices diferenciadores, se podría especular sobre la conveniencia de quitarlas del estudio, dado que apenas se relacionan con las demás. Esto es cierto, pero las mantenemos para estudiar el factor antiautonomista que como decíamos al principio del estudio, es un punto de gran vigencia hoy en día.

Factor 3

Capacidad explicativa: 2'36002

Porcentaje de explicación: 9'077%

Variables muy importantes:

- 6 - Clase social.
- 7 - Estudios.

Variables importantes:

- 5 - Profesión.

Variables secundarias:

- 8 - Lugares que conoce.
- 10 - Los licenciados están preparados (negativa).
- 28 - Las películas españolas buenas (negativa).

Variables terciarias (solo citar): 12, -33, 38, -41.

Este factor ha quedado también muy claro tras la rotación. Es el factor que podíamos denominar de estatus socio-cultural. Las variables 5, 6, 7 e incluso la 8, no ofrecen lugar a dudas, pero las variables 10 y 28, negativas, también indican cierta cultura como ya hemos comentado.

Factor 4

Capacidad explicativa: 1'74044

Porcentaje de explicación: 6'694%

Variables muy importantes:

19 - Sí a las cooperativas.

Variables importantes:

12 - Sí a la libertad para las regiones.

18 - Madrid controla demasiado.

Variables secundarias:

23 - Sí a nacionalizar.

Variables terciarias (solo citar): 7, -15, -29, -35.

Este factor resulta bastante claro. Une el cooperativismo al regionalismo e incluso a las nacionalizaciones. Ya hemos comentado la lógica de esta unión. Lo denominaremos - como cooperativista-regionalista.

Interesante resulta el resaltar que tanto en este análisis factorial como en los anteriores, ha quedado normalmente dividido por un lado el autonomista (antiautonomista en el caso del factor dos), y por otro, el sentimiento regionalista (en este factor cuatro unido al cooperativismo). Esto significa que los sentimientos descentralizadores son independientes del considerar que La Rioja debe ser autónoma sola o si bien se debe unir a otro ente autónomo.

Factor 5

Capacidad explicativa: 3'60594

Porcentaje de explicación: 13'869%

Variables muy importantes:

33 - La mujer debe estar en casa (negativa).

Variables importantes:

11 - Sí al divorcio.

15 - Los niños no deben saber del sexo (negativa).

17 - Años con Franco buenos (negativa).

29 - La policía debe ser dura (negativa).

35 - No a la huelga (negativa).

38 - La mujer debe trabajar fuera de casa.

41 - La mujer es poco inteligente (negativa).

Variables secundarias:

27 - No a las relaciones prematrimoniales (negativa).

Variables terciarias (solo citar): 5, 7, 8, -10, -28,
-32, 39, 42.

En primer lugar, es importante el constatar que tras la rotación, éste se ha convertido en el factor más importante. Es el que tiene mayor capacidad explicativa y, en consecuencia, explica un mayor porcentaje de la varianza total.

Por otro lado, es un factor de fácil interpretación. Fundamentalmente coincide con el liberal-progresista de estudios anteriores pero con un marcado acento en favor de la mujer. Por lo tanto, lo denominaremos liberal-feminista.

Factor 6

Capacidad explicativa: 1'45158

Porcentaje de explicación: 5'583%

Variables muy importantes:

39 - Mejor una república.

Variables importantes:

42 - Servicio militar voluntario.

Variables secundarias:

23 - Sí a las nacionalizaciones.

Variables terciarias (solo citar): -8, 11, -32, -41.

Nuevamente nos encontramos con un factor bastante claro que por orden de importancia es: republicano, antimilitarista y partidario de nacionalizar las grandes empresas. Es esta una forma de pensar bastante coherente y resumiremos - llamándole republicano-antimilitarista.

En conclusión, hemos logrado seis factores muy interesantes, que resumiendo son:

<u>Factor</u>	<u>Significado</u>	<u>Porcentaje de explicación</u>
1	Antirreligioso	11'638%
2	Antiautonomista	7'769%
3	Estatus socio-cultural	9'077%
4	Cooperativista-regionalista	6'694%
5	Liberal-feminista	13'869%
6	Republicano-antimilitarista	5'583%

Estos factores son muy claros y además entre todos se puede lograr una importante explicación, por lo tanto, pasamos a la interpretación de las puntuaciones.

INTERPRETACION DE LAS PUNTUACIONES.

En primer lugar hemos de recordar que con el nuevo programa, las puntuaciones van a dar con promedio cero y con varianza en cada factor igual a su capacidad explicativa, como comentábamos al principio de este análisis factorial. La ventaja principal de este planteamiento consiste en que al no estar normalizados los factores, las relaciones entre los indivíduos son más reales.

Por otro lado ya comentábamos que el I.I. nos resolvía la indeterminación inherente al hecho de tener los factores promedio cero. Pero aquí vamos a completar este análisis con la introducción del individuo Pro-factor (I.P.) y del indivíduo Antifactor (I.A.).

Voy a explicar a continuación qué significan el I.P. y el I.A. y qué sentido tienen.

El factor puede estar más o menos disperso y tiene sentido el estudiar esta dispersión. Para ello creo los I.P. y I.A. que serían unos individuos tales que el primero (I.P.) está en todo de acuerdo con el factor y el segundo (I.A.) está en todo en contra. El I.P. contestará con la puntuación máxima en las variables positivas en el factor y con la mínima en las negativas. El I.A. por su parte contestará con la mínima puntuación posible en las variables positivas y con la máxima en las negativas. Veremos que el I.P. y el I.A. junto con el I.I. nos darán grandes posibilidades de análisis de los resultados del factor y así habremos salvado la indetermi-

ción inherente a los factores en los estudios de opinión. (1)

- (1) La utilización y el concepto del I.I., I.A. e I.P. no los he visto en ningún libro especializado, por lo que los considero bastante originales.

CREACION DEL I.I., I.P., I.A.

Comencemos creando el I.I.; en primer lugar hay que hacer constar aquí que aunque ya comentamos que las variables 5, 6, 7 y 8, que tienen de número de ordenador en este análisis factorial los números 1, 2, 3 y 4, no tenían sentido de cara al individuo indiferente, en este caso si que las vamos a hacer intervenir para que sea coherente la comparación con el I.P. y el I.A.

Lo que queremos es que el I.P. sea simétrico al I.A. respecto al I.I. para ello crearemos el I.I. en los puntos medios de las variables. Tengamos en cuenta que la creación de estos individuos se hace mediante la aplicación de la ecuación que liga los factores con las variables, es decir, mediante una combinación lineal y por lo tanto si el I.I. se crea en los puntos medios y el I.P. y I.A. en los extremos, estos dos últimos han de ser simétricos respecto al I.I.

Por lo tanto crearemos el I.I. en los puntos medios de las variables, es decir, contestará 4 en las preguntas de opinión al igual que sucedía en la creación del I.I. en el Análisis Factorial tres. Pero en las primeras cuatro preguntas corresponderá el punto medio, cuyo significado es el siguiente:

<u>Variable</u>	<u>Nº ordenador</u>	<u>Significado del punto medio</u>
5	1	Profesión .. Autónomo
6	2	Clase social .. Entre MM y MB
7	3	Estudios .. Bachiller
8	4	Lugares que conoce .. La Región

Por lo tanto este individuo indiferente significa una persona de una profesión media (autónomo) de clase entre media-media y media-baja, con estudios de bachiller y que conoce la región, es decir, de datos objetivos medios y que contesta indiferente en todas las preguntas.

Esta pequeña modificación del I.I. no va a distorsionar los resultados y nos ofrecerá ventajas en la interpretación.

Para crear el individuo indiferente normalizamos los puntos medios de las variables:

Sustituyendo estos valores en la ecuación correspondiente tendremos los valores del I.I. en los distintos factores.

	<u>Valor del I.I.</u>
Factor 1	2'7731
Factor 2	0'3550
Factor 3	0'8641
Factor 4	-1'6965
Factor 5	-0'2475
Factor 6	0'3099

Luego ya tenemos el I.I., éste tiene todas las limitaciones expuestas al crear el correspondiente al análisis factorial tres, y además tiene la imprecisión de haber dado un valor a las cuatro primeras variables, pues siempre es discutible que los valores que se han dado sean realmente representativos de un punto de comparación. Pero lo que es evidente es que resulta un punto de comparación y bastará saber lo que significa ese punto para que la comparación sea buena.

Por otro lado los valores que se han dado a las cuatro primeras variables son bastante significativos en cuanto a lo que puede ser un punto de comparación, no la media de nuestra muestra, pero si un nivel medio, que no es ni alto ni bajo; lo podíamos denominar "medio objetivo".

Una vez que tenemos el I.I. vamos a crear el I.P. y el I.A. para ello habrá que normalizar los puntos extremos de ca da variable y después sustituiremos en la ecuación correspondiente bien el extremo superior (si la variable es positiva) o el inferior (si es negativa) para la creación del I.P. y al revés para el I.A.

Veamos los extremos de las variables normalizadas:

<u>Variable</u>	<u>Extremo superior normalizado</u>	<u>Extremo inferior normalizado</u>
1 . . .	3'24 . . .	-0'80
2 . . .	2'16 . . .	-1'30
3 . . .	2'98 . . .	-1'11
4 . . .	1'47 . . .	-2'69
5 . . .	1'30 . . .	-2'89
6 . . .	1'36 . . .	-1'11
7 . . .	0'97 . . .	-2'89
8 . . .	0'50 . . .	-4'73
9 . . .	1'74 . . .	-0'98
10 . . .	1'10 . . .	-1'95
11 . . .	1'17 . . .	-2'88
12 . . .	1'16 . . .	-2'95
13 . . .	1'19 . . .	-2'08
14 . . .	1'62 . . .	-1'26
15 . . .	0'73 . . .	-3'23
16 . . .	0'66 . . .	-2'43
17 . . .	1'52 . . .	-1'51
18 . . .	1'46 . . .	-1'2
19 . . .	1'13 . . .	-1'75
20 . . .	0'85 . . .	-2'29
21 . . .	1'03 . . .	-1'42
22 . . .	1'15 . . .	-1'30
23 . . .	1'36 . . .	-1'50
24 . . .	1'98 . . .	-1'39
25 . . .	1'73 . . .	-0'91
26 . . .	1'35 . . .	-1'17

Con lo que tenemos que aplicando en las ecuaciones tal y como hemos comentado, el I.P., I.I. y I.A. quedan: (repetimos en este cuadro el resultado ya obtenido anteriormente del I.I. para tener un resumen completo). (1)

<u>Factor</u>	<u>I.P.</u>	<u>I.I.</u>	<u>I.A.</u>
1	9'2673	2'7731	-3'7172
2	4'5834	0'3550	-3'8768
3	7'1197	0'8641	-5'4030
4	4'4110	-1'6965	-7'8018
5	5'5588	-0'2475	-6'0472
6	6'0967	0'3099	-5'4807

A continuación vamos a ir estudiando factor por factor, con la ayuda de nuestros tres individuos.

El ordenador nos ha proporcionado en este caso los histogramas de los factores gracias a un encadenamiento de programa que hemos realizado, estos histogramas están en las hojas de ordenador nº 914 y siguientes.

(1) Si operamos veremos que el I.I. no queda exactamente en el centro entre el I.P. y el I.A. la razón de esto es que al haber realizado todas las operaciones ~~manu~~almente, por no tener el ordenador programa aplicable a causa de ser original la idea de la creación de estos tres individuos, la exactitud total es imposible, aunque si da prácticamente exacto.

ESTUDIO DE NUESTRA MUESTRA A TRAVES DE LOS FACTORES.

Factor 1 (antirreligioso)

Vamos a calcular el punto medio entre el I.P. y el I.I. y también entre el I.I. y el I.A. así tendremos:

- 9'2673 - I.P. Bastante antirreligioso.
- 6'0202 - Algo antirreligioso.
- 2'7731 - I.I. Algo religioso.
- 0'47205 - Bastante religioso.
- 3'7172 - I.A.

Luego, dado el histograma, vemos cuantos individuos hay en cada una de las cuatro partes en que hemos dividido el factor. Para ello habrá que dividir el efectivo del intervalo en el cual esté la división de manera proporcional así:

Intervalo 44: 5'999 . . . 6'0202 . . . 6'199
 efectivo = 0
 Luego no hay que dividir.

Intervalo 27: 2'599 . a . 2'7731 . b . 2'799
 efectivo = 4

corresponde al subintervalo a ... 3'482 \simeq 3

corresponde al subintervalo b ... 0'518 \simeq 1

Intervalo 11: -0'599 .. a . -0'47205 . b . -0'399
 efectivo = 10

a ... 6'3475 \simeq 6

b ... 3'6525 \simeq 4

Luego en conclusión nos queda:

Bastante religioso: 124 + 6 = 130

Algo religioso: (240 - 134) + 4 + 3 = 113

Algo antirreligioso: 1 + (266 - 244) = 23

Bastante antirreligioso: (267 - 266) = 1

Pasando a tantos por ciento:

Bastante religioso: 49%

Algo religioso: 42%

Algo antirreligioso: 9%

Bastante antirreligioso: 0%

Estos resultados lógicamente son discutibles, en primer lugar por la creación de los I.P., I.I. e I.A., que aún no siendo arbitraria tienen las limitaciones que comentábamos respecto al I.I. en el análisis factorial tres y que son

también aplicables a los dos nuevos. Tengamos en cuenta que el estar completamente de acuerdo (puntuando un 7) en una pregunta puede no resultar tan extremista como el hacerlo en otra. Pero siempre es difícil como comentábamos buscar la pauta de comparación.

Las dos divisiones hechas en los puntos medios no causan problemas por tratarse de combinaciones lineales las ecuaciones de partida.

En último lugar quiero aclarar que el concepto de ser bastante religioso, por ejemplo, es relativo al tono de las preguntas y al significado de los individuos de comparación (I.P., I.I., I.A.)

Factor 2 (antiautonomista)

4'5834 - I.P.	Bastante antiautonomista.
2'4692 -	Algo antiautonomista.
0'3550 - I.I.	Algo autonomista.
-1'7626 -	Bastante autonomista.
-3'8768 - I.A.	

Haciendo como en el factor anterior, a la vista de los histogramas tendremos:

Intervalo 28: 2'400 . ^a . 2'4692 . ^b . 2'599

efectivo = 6

a ... 2'076 \simeq 2

b ... 3'894 \simeq 4

Intervalo 17: 0'200 . ^a . 0'3550 . ^b . 0'400

efectivo = 11

a ... 8'525 \simeq 9

b ... 2'475 \simeq 2

Intervalo 7: $-1'799$. a . $-1'7626$. b . $-1'599$

efectivo = 20

a ... $3'64 \cong 4$

b ... $16'36 \cong 16$

Luego en conclusión nos queda:

Bastante autonomista: $20 + 4 = 24$

Algo autonomista: $16 + (141 - 40) + 9 = 126$

Algo antiautonomista: $2 + (257 - 152) + 2 = 109$

Bastante antiautonomista: $4 + (267 - 263) = 8$

Pasando a tantos por ciento:

Bastante autonomista: 9%

Algo autonomista: 47%

Algo antiautonomista: 41%

Bastante antiautonomista: 3%

En este factor se pueden hacer las mismas salvedades que en el anterior respecto al significado de los tantos por ciento finales, pero quizá es más claro por el tono de las preguntas que lo forman.

En los factores que siguen también se pueden hacer idénticas consideraciones, que daremos por supuestas, comentaremos todo esto en las conclusiones.

Factor 3 (estatus socio-cultural)

En el caso de este factor creo que es preciso comentar antes de nada la cierta arbitrariedad del I.I. que ya hemos citado con anterioridad, de todas formas si es interesante - dado que es el punto medio de I.P. y I.A., siendo estos últimos de claro significado. Haremos un análisis similar a los anteriores, pero el lector debe ser consciente de esta salvedad.

7'1197 - I.P.

Estatus bastante alto.

3'9919 -

Estatus algo alto.

0'8641 - I.I.

Estatus algo bajo.

-2'2752 -

Estatus bastante bajo.

-5'4030 - I.A.

Intervalo 5: -2'398 . ^a . -2'2752 . ^b . -2'198

efectivo = 3

a ... 1'842 \simeq 2

b ... 1'158 \simeq 1

Intervalo 21: 0'801 . a . 0'8641 . b . 1'001

efectivo = 3

a ... 0'9465 \simeq 1

b ... 2'0535 \simeq 2

Intervalo 36: 3'800 . a . 3'9919 . b . 4'000

efectivo = 3

a ... 2'8785 \simeq 3

b ... 0'1215 \simeq 0

Luego en conclusión queda:

Estatus bastante bajo: $7 + 2 = 9$

Estatus algo bajo: $1 + (204 - 10) + 1 = 196$

Estatus algo alto: $2 + (262 - 207) + 3 = 60$

Estatus bastante alto: $(267 - 265) = 2$

Pasando a tantos por ciento:

Estatus bastante bajo: 3%

Estatus algo bajo: 73%

Estatus algo alto: 23%

Estatus bastante alto: 1%

Factor 4 (cooperativista-regionalista)

- 4'4110 - I.P. Bastante cooperativista-regionalista.
 1'35725 - Algo cooperativista-regionalista.
 -1'6965 - I.I. Algo anticooperativista-regionalista.
 -4'74805 - Bastante anticooperativista-regionalista.
 -7'8018 - I.A.

Intervalo 0: $-\infty$. a . -4'74805 . b . -4'199

efectivo = 0

a ... 0

b ... 0

Intervalo 13: -1'798 . a . -1'6965 . b . -1'598

efectivo = 8

a ... 4'06 \simeq 4

b ... 3'94 \simeq 4

Intervalo 28: 1'201 . a . 1'35725 . b . 1'401

efectivo = 9

a ... 7'03125 \simeq 7

b ... 1'96875 \simeq 2

Luego en conclusión queda:

Bastante anti coop. reg.: 0

Algo anti coop. reg. : $29 + 4 = 33$

Algo coop. reg. : $4 + (220 - 37) + 7 = 194$

Bastante coop. reg.: $2 + (267 - 229) = 40$

Pasando a tantos por ciento:

Bastante anticoop. reg.: 0%

Algo anticoop. reg.: 12%

Algo coop. reg.: 73%

Bastante coop. reg.: 15%

Factor 5 (liberal-feminista)

- 5'5588 - I.P.
Bastante liberal-feminista.
- 2'65565 -
Algo liberal-feminista.
- 0'2475 - I.I.
Algo antiliberal-feminista.
- 3'14405 -
Bastante antiliberal-feminista.
- 6'0475 - I.A.

Intervalo 4: -3'199 . ^a . -3'14405 . ^b . -2'999
 efectivo = 3
 a ... 0'82425 \approx 1
 b ... 2'17575 \approx 2

Intervalo 18: -0'399 . ^a . -0'2475 . ^b . -0'199
 efectivo = 10
 a ... 7'575 \approx 8
 b ... 2'425 \approx 2

Intervalo 33: 2'600 . ^a . 2'65565 . ^b . 2'800
 efectivo = 2
 a ... 0'5565 \approx 1
 b ... 1'4435 \approx 1

Luego en conclusión queda:

Bastante antiliberal-feminista: $4 + 1 = 5$

Algo antiliberal-feminista: $2 + (124 - 7) + 8 = 127$

Algo liberal-feminista: $2 + (238 - 134) + 1 = 107$

Bastante liberal-feminista: $1 + (267 - 240) = 28$

Pasando a tantos por ciento:

Bastante antiliberal-feminista: 2%

Algo antiliberal-feminista: 48%

Algo liberal-feminista: 40%

Bastante liberal-feminista: 10%

Factor 6 (republicano-antimilitarista)

- 6'0967 - I.P.
Bastante rep. antimil.
- 3'2033 -
Algo rep. antimil.
- 0'3099 - I.I.
Algo monárquico militarista.
- 2'5873 -
Bastante monárquico militarista.
- 5'4807 - I.A.

Intervalo 0: $-\infty$. ^a . -2'5873 . ^b . -2'400

efectivo = 0

a ... 0

b ... 0

Intervalo 14: 0'200 . ^a . 0'3099 . ^b . 0'400

efectivo = 19

a ... 10'4405 \simeq 10

b ... 8'5595 \simeq 9

Intervalo 29: 3'199 . ^a . 3'2033 . ^b . 3'399

efectivo = 1

a ... 0'0215 \simeq 0

b ... 0'9785 \simeq 1

Luego en conclusión queda:

Bastante mon. mil.: 0

Algo mon. mil.: $157 + 10 = 167$

Algo rep. antimil.: $9 + (265 - 176) + 0 = 98$

Bastante rep. antimil.: $1 + (267 - 266) = 2$

Pasando a tantos por ciento:

Bastante mon. mil: 0%

Algo mon. mil: 62%

Algo rep. antimil: 37%

Bastante rep. antimil: 1%

CONCLUSIONES DE ESTE ULTIMO ANALISIS.

Vamos a intentar a continuación sacar una serie de conclusiones sobre la opinión de la muestra que analizamos.

Para ello quiero recalcar dos puntos fundamentales:

a) Todo esto se basa en la encuesta que realicé en el verano de 1.977, es decir, recién estrenado el nuevo sistema de mocrático y en aquel momento la gente estaba algo confusa, por lo tanto es muy posible que de realizarla hoy los resultados fueran diferentes.

b) El análisis que voy a exponer a continuación se basa en la situación del I.I. y por lo tanto, ya he comentado a lo largo de este trabajo su cierta relatividad, pero también he expuesto que es el mejor instrumento de análisis con el que contamos. A causa de esa cierta relatividad cuando decimos que hay tantas personas, por ejemplo liberales, hay que entender que el concepto de liberalidad es el asociado al to no de las preguntas y al significado del I.I. y así pasa en todos los factores.

Factor 1

Hechas estas dos puntualizaciones y a la vista de los histogramas y de todo el trabajo previo podemos decir que nuestra muestra es religiosa dado que prácticamente hay un 0%

de bastante antirreligiosos y el 49% son bastantes religiosos creo que a la vista de los resultados expuestos en el estudio de la muestra por el factor uno no cabe ninguna duda de esta religiosidad.

Por otro lado, según los resultados del ordenador el histograma del factor uno tiene la moda en -1'311 que corresponde al grupo de las bastantes religiosas y lógicamente es asimétrica con cola a la derecha (asimétrica = 1'171) y algo leptocúrtica (curtosis = 0'921), la media (media aritmética = 0) corresponde a los algo religiosos, la mediana (-0'410) también.

En conclusión, podemos afirmar la fuerte religiosidad de la muestra, completamente acorde con los resultados de las variables.

Factor 2

Este factor es el antiautonomista, y vemos una ligera tendencia al autonomismo (bastante 9%, algo 47%) frente al antiautonomista (bastante 3%, algo 41%), pero de todas formas las dos tendencias están bastante igualadas.

La media aritmética (que es cero) corresponde a los algo autonomistas, igual que la mediana (-0'112) y la moda (-1'550) pero en orden creciente de autonomismo, lo que nos indica que la distribución es asimétrica con cola a la dereu

cha, cosa que se confirma dado que la asimetría es positiva (0'129). Tengamos en cuenta que el factor es antiautonomista.

Podemos ver también que es platicúrtica (curtosis = -1'243).

En conclusión, hay una tendencia a la autonomía, pero no demasiado fuerte, dentro de nuestra muestra.

Factor 3

Este factor, como ya hemos dicho es bastante relativo y todo se basa en la comparación con el I.I. que difícilmente puede ser objetivo; de todas formas, y supuesta esta premisa, vamos a estudiarlo:

Recordemos que es el factor de estatus socio-cultural y tenemos que colegir del estudio lo bajo de ese estatus en La Rioja, dado el punto de comparación.

La distribución es algo asimétrica (asimetría = 0'786) con cola a la derecha y un poco leptocúrtica (curtosis = 0'358). La moda (-0'724), mediana (-0'271) y media (0), dan un estatus algo bajo, como era de esperar.

Luego en conclusión, tenemos que hablar de incultura y bajo estatus social, ambos relativos al punto de referencia.

Factor 4

Este es el factor cooperativista-regionalista, y vemos que nuestra muestra es muy favorable a este factor pues el 73% son algo cooperativistas-regionalistas y no hay nadie que es te bastante en contra.

Hay cierta asimetría con cola a la izquierda (asimetría = -0'330) y algo platicúrtica (curtosis = -0'222).

La moda (0'883), mediana (0'116) y media (0), corresponden a algo cooperativistas-regionalistas.

En conclusión, hay una fuerte tendencia favorable a este factor en la muestra estudiada.

Factor 5

Este es el factor liberal-feminista y la posición de la muestra respecto a este factor es bastante equilibrada:

A favor	{	Bastante	-	10%
		Algo	-	40%
En contra	{	Bastante	-	2%
		Algo	-	48%

La mediana (-0'209), prácticamente coincide con el I.I. (-0'2475), la media (0), estaría con los algo a favor

y la moda (-0'492) con los algo en contra.

Por otro lado, hay algo de asimetría con cola a la derecha (asimetría = 0'231) y es platicúrtica (curtosis = -0'936).

En conclusión, nuestra muestra está equilibrada en este factor que por otro lado es el más explicativo tras la rotación.

Factor 6

Lo hemos denominado republicano-antimilitarista. Es de destacar la práctica ausencia de valores extremos, lo que indica a mi juicio que mucha gente dentro de nuestra muestra carece de opinión fundada en estos temas.

Hay una asimetría con cola a la derecha (asimetría = 0'362) y es algo platicúrtica (curtosis = -0'342) lo que indica la ausencia de valores extremos.

La mediana (-0'113), moda (-0'333) y media (0), se sitúan en algo en contra del factor.

En conclusión la muestra en general careciendo de valores extremos, está más por la monarquía y el ejército que en contra; pero este tema discrimina poco a nuestros encuestados.

Terminamos este breve análisis diciendo que los valores de las variables significativas corroboran totalmente el resultado de los factores.

Creo que es interesante precisar que este análisis se podría ampliar bastante, pero opino que mi trabajo, para mantener la imparcialidad que propugné al principio de esta segunda parte, debe terminar poco más allá de la exposición de los datos y deben ser los políticos, sociólogos, etc. los - que saquen las consecuencias de esos resultados.

OTRAS ELABORACIONES MULTIVARIANTES.

Terminado el estudio de Análisis Factorial sobre el total de la muestra, podríamos hacer aquí análisis similares a los que hicimos en el estudio sobre la muestra representativa de Logroño capital.

Me refiero a realizar cluster análisis, análisis discriminante, estudios por χ^2 dentro de los clústeres, etc.

Esto no es posible pues con 267 datos el trabajo de ordenador para hacer un cluster, dada la capacidad de nuestro equipo, es excesivo, además de que la manipulación posterior de los datos sería casi imposible dada la carencia de programas apropiados (recuerde el lector lo laborioso del análisis con 100 datos).

Por otro lado, como ya hemos visto el camino a seguir con 100 datos, sería repetir el proceso y no tendría excesivo interés.

Por lo tanto lo único que vamos a hacer es ver como quedan los factores, con sus medios y desviaciones típicas correspondientes, en cada una de las cinco comarcas estudiadas al principio:

- 1 - Rioja media.
- 2 - Rioja alta.

- 3 - Rioja Baja.
- 4 - Sierra - Cameros.
- 5 - Rioja Alavesa.

Los resultados están en las hojas de ordenador nº 938 y ss. Lo que vamos a intentar ver si a la vista de nuestros seis factores, hay diferencias significativas entre las cinco comarcas.

Comarca 1 (Rioja Media)

(En primer lugar quiero recordar que estamos haciendo análisis de los datos, luego no voy a realizar pruebas de hipótesis).

Los factores 1, 2, 4 y 6 están prácticamente en la media. El factor tres (estatus) es algo más alto que la media y el 5 (liberal-feminista) también, pero ambos por muy estrecho margen.

En conclusión esta primera comarca, integrada fundamentalmente por Logroño capital, no se destaca fundamentalmente de la media. Solo se puede apreciar ligeramente por encima de la media el estatus socio-cultural y el liberalismo-feminismo, ambos lógicos por tratarse de la capital.

Comarca 2 (Rioja Alta)

El factor uno está exactamente en la media y muy cerca

lo está el 6, así como también bastante cerca tenemos al 2 y 5. Algo por debajo están el 3 (estatus) y el 4 (cooperativista-regionalista) pero casi inapreciablemente, luego podemos concluir que esta comarca está con la media, teniendo como dato más destacable el estar bajo la media en el factor 4. Cosa natural por ser una zona agrícola donde se desconfía bastante de las cooperativas.

Comarca 3 (Rioja Baja)

Los factores 1, 3, 4, 5 y 6 están prácticamente en la media, el factor 2 da un valor ligeramente positivo, luego es la zona más antiautonomista.

Comarca 4 (Sierra Cameros)

[Esta comarca y la siguiente (Rioja Alavesa) son las que tienen valores más significativos, vamos a analizarlos, pero tengamos en cuenta que estos valores pueden deberse al azar, a causa de ser pocos los datos, pues hay población escasa, de todas formas y siguiendo el principio de "análisis de los datos", los estudiaremos como los anteriores].

En los factores 2 y 4 está prácticamente en la media. El 3 (estatus) y el 6 (republicano-antimilitarista) son algo negativos. El 1 (antirreligioso) y el 5 (liberal-feminista) son bastante negativos, luego es una comarca que se caracteri

za por su religiosidad y antiliberalismo.

Comarca 5 (Rioja Alavesa)

Los factores 1, 4 y 6 están prácticamente en la media. El factor 2 (antiautonomista) es fuertemente negativo, luego es una comarca muy autonomista, esto quizá se deba a la influencia del autonomismo vasco.

El factor 3 es negativo, luego el estatus es bajo y el 5 positivo, luego son liberales.

Para terminar quiero hacer dos apreciaciones:

- a) Al decir que en un factor se está con la media, quiero - decir que se está con la media de la región. La situación de ésta respecto al I.I. ya ha sido estudiada con anterioridad.

- b) Me he referido constantemente al análisis de datos y cuando hablo de una comarca, me refiero a la parte de la muestra tomada en esa comarca.

III.- EPILOGO Y CONCLUSIONES.

Quiero terminar esta parte tratando dos cuestiones que considero importantes:

- Posible continuación del estudio.
- Conclusiones.

POSIBLE CONTINUACION DEL ESTUDIO:

A lo largo del presente análisis hemos estudiado una parte de la encuesta.

Hemos ido estudiando las variables que parecían más - apropiadas para cada parte del análisis, y las que considerá bamos más interesantes.

Evidentemente el lector puede considerar que hay otras variables dentro de la encuesta que tendrían mayor interés - que las estudiadas, pero lo que creo que podemos garantizar es que la elección no ha sido al azar, sino fruto de un laborioso análisis.

Por otro lado, las variables tratadas, lo han sido me diante determinadas técnicas de análisis multivariante y se podrían aplicar muchos más, incluso más interesantes desde - determinados puntos de vista, pero creo que el camino está -

abierto y el lector puede ampliar el análisis según sus pro
pios objetivos.

También quiero dejar sentado aquí que este estudio se
ría solo una parte de un estudio sociológico más amplio, pero
esto ya es tarea de los sociólogos y considero que como apli
cación de análisis multivariante es suficientemente completo.

Evidentemente hay dos caminos claros de ampliación del
presente análisis. En primer lugar, podríamos, a partir de -
los factores que poseemos, estimar las comunalidades y reali
zar un estudio por factores principales.

En segundo lugar, y quizá esto fuera lo más interesan
te, intentar un estudio inductivo, que partiendo de que el -
actual estudio sobre la muestra es una buena estimación, in
tentarse ampliaciones inductivas a base de pruebas de hipóte
sis, etc.

De hecho al final del estudio del análisis factorial
tres, ya hemos comentado e incluso realizado algunas pruebas
de hipótesis que completaban y ampliaban el análisis y el ca
mino queda abierto en ese sentido, pero dado por un lado la
limitación de medios y por otro el volumen que está adquirien
do el presente trabajo, he decidido terminar el estudio, que
por otro lado podría ampliarse casi indefinidamente.

CONCLUSIONES:

Las conclusiones del presente estudio son muy variadas y el expresar todas equivaldría en buena parte a transcribirlas de nuevo, de todas formas podemos resumirlas de la siguiente forma:

a) Desde un punto de vista sociológico hemos logrado conocer de manera mucho más profunda a la muestra estudiada, que a su vez nos sirve para hacer una buena estimación del colectivo. No voy a repetir aquí las conclusiones de los diversos análisis factoriales (sobre todo del tres y del cinco que son los más interesantes), el lector puede consultar allí los diversos resultados. (1)

b) En cuanto al análisis estadístico podemos observar:

1.- El análisis multivariante es un poderoso instrumento de estudio sociológico.

2.- Concretamente, la técnica que más hemos utilizado que es el análisis factorial, ha quedado completada con nuestras aportaciones del I.I., I.P., I.A., que hacen que este análisis resuelva gran parte de sus indeterminaciones, y sea más útil en el futuro, para la aplicación a temas de opinión.

(1) Anexo en página siguiente.

3.- El resto de técnicas multivariantes pueden ser de interésante aplicación, y de hecho hemos utilizado varias, dando idea clara de la posibilidad de encadenamiento de las mismas.

4.- El problema de la inducción, aún siendo grave, y ante el actual desarrollo de la estadística, hay que obviarlo y considerar que si la muestra es aleatoria y amplia, la estimación es buena. (1).

5.- De todas formas se pueden hacer consideraciones tales como pruebas de hipótesis sobre los resultados de la elaboración multivariante, como hacemos al final del análisis factorial tercero.

(1) Dado nuestro tipo de variables y las limitaciones de los test, incluso en condiciones de normalidad.

ANEXO CITADO:

De todas formas podemos resumir los resultados sociológicos en:

1) Logroño: (A.F. 3).

- Factores: - cierto tradicionalismo.
- culturalmente vemos dos grupos, uno muy numeroso relativamente inculto y otro pequeño más culto.

- tendencia autonomista.
- fuerte cooperativismo-regionalismo.

- Cluster análisis: estudiamos las características más importantes de cada cluster.

- Análisis discriminante: conseguimos diferenciar los clústeres con posibilidades de clasificación de nuevos individuos.

- Tablas de contingencia aplicadas al cluster análisis: Nos dan que la edad es la variable que más influye en la formación de los clústeres y en definitiva en la ideología, mientras que el sexo va en segundo lugar.

2) Tablas de contingencia:

- Cruces de variables para La Rioja y para Logroño, viendo las relaciones, etc.

3) Rioja (total): (A.F. 5)

- Factores:
 - gran religiosidad.
 - ligero autonomismo.
 - socioculturalmente similar a Logroño.
 - fuerte cooperativismo-regionalismo.
 - equilibrio en el factor liberal-feminista.
 - valores muy centrados en el repub-antimil.

- Comarcas Riojanas: pocas diferencias.

PARTE TERCERA

0.- INTRODUCCION.

En esta tercera parte pretendemos realizar una aplicación del análisis multivariante a la problemática económica.

Tenemos que repatir, por lo tanto, al igual que dijimos al comienzo de la segunda parte, que no se trata de un estudio económico completo, sino por el contrario, de una aplicación del análisis multivariante a los problemas económicos.

Por otro lado, las posibilidades de aplicación del análisis multivariante a estos problemas son muy extensos y solo vamos a tratar algunos de ellos ciñéndonos a una problemática concreta que va a ser la camarcalización de la Rioja, como ya comentaremos posteriormente.

En cuanto al ámbito del estudio ya adelantábamos al comentar el problema en la parte segunda, cuando tratábamos el análisis sociológico, que en temas económicos hemos de aceptar la división política actual de España y que hemos de ceñirnos a la provincia o al ente autonómico.

La situación autonómica de La Rioja es problemática en el momento de escribir estas líneas, pero parece que hay una tendencia a conseguir que la provincia sea un ente autonómico sin unirse con otras.

Por lo tanto La Rioja, como provincia, es en este momento un objetivo de inmenso interés a la hora de intentar una comarcalización, que evidentemente será necesaria para la futura planificación del ente autonómico, si éste llega a ser una realidad.

I.- OBJETIVOS DEL ESTUDIO.

A la hora de aplicar un análisis multivariante al primer problema con el que nos encontramos, es la necesidad de datos que hagan referencia al colectivo y problemática a estudiar.

En nuestro caso el análisis multivariante lo vamos a aplicar a la provincia riojana y creo que no es necesario insistir en el interés de los estudios económicos en una provincia que en gran parte carece de ellos, más aún cuando se está planteando el problema de su autonomía.

Luego ya tenemos definido el colectivo (la provincia de La Rioja) y el tema (la economía) que vamos a estudiar.

Pero no cabe duda que el decir que vamos a estudiar la economía riojana es altamente difuso y hemos de concretar mucho más.

En primer lugar vamos a estudiar problemas susceptibles de un tratamiento multivariante, con lo que dejamos fuera una serie de temas como contabilidad regional, tablas input-output, etc, (por otro lado las tablas input-output de la economía riojana y sus cuentas regionales han sido ampliamente estudiadas y con muchos más medios que los que yo poseo

por la Caja de Ahorros de Zaragoza, Aragón y Rioja (1)).

Dentro de los problemas que podemos estudiar por técnicas multivariantes, hay todavía una gran cantidad de temas a tratar, y el intentar tratarlos todos sería incluso tratar solo un resumen de ellos se saldría completamente de los límites de este trabajo.

El lector puede imaginarse lo laborioso que resultaría el tratamiento económico de los modelos fundamentales de teoría económica para el caso de la economía riojana.

Por otro lado, creo que dada la escasez de datos a nivel provincial y la absoluta carencia de ellos a niveles más desagregados, exceptuando los pocos datos existentes a nivel municipal, harían dificilísima la aplicación de los modelos. Y todo esto sin entrar en la dudosa fiabilidad de la mayoría de estos datos. (2)

En definitiva, y por todo lo dicho hasta ahora nos vamos a centrar en un problema concreto de la economía regional riojana, que nos sirva como aplicación del análisis multiva-

(1) Caja de Ahorros de Zaragoza, Aragón y Rioja (1.978).

(2) Sin embargo resulta indudable que el estudio se puede ampliar por ese camino, pues las técnicas multivariantes - tienen inmensas posibilidades en este campo; simplemente como ejemplo podemos citar la creación de modelos de emigración, de diferencias de renta, de desarrollo, etc.

riante, y que va a ser concretamente el de la comarcalización. (1)

- (1) Consideramos que esta aplicación resulta suficientemente interesante, sin perjuicio de que haya otras que también lo sean.

Sobre la importancia de la regionalización en general podemos citar a Valeriano Muñoz, en su artículo en Información Comercial Española (1.977), pag. 107. - También dan mucha importancia al tema Marchal (1.964) y García Enterría (1.976), así como Velasco Barroetabeña, en el Boletín de Estudios Económicos (1.972), entre otros muchos autores.

II.- REGIONALIZACION - PROBLEMATICA ECONOMICA.

Dentro de los estudios de regionalización hay diversos autores, pero yo voy a seguir fundamentalmente a Richardson, que es uno de los más importantes.

Evidentemente no es objetivo de este trabajo el hacer un estudio exhaustivo de la teoría económica que trata el tema, que por otra parte es muy extensa (1), pero lo que sí es claro es que hay que hacer un breve planteamiento del problema.

Richardson en su libro de economía regional, plantea, en primer lugar, la problemática de la localización a la luz de la teoría económica. Posteriormente habla de la economía urbana y después de la economía regional que es la parte que a nosotros más nos interesa.

En concreto lo que nosotros queremos es estudiar el concepto de región.

En realidad lo que vamos a hacer es comarcalizar una región, pero los planteamientos son similares al caso de regionalizar un estado, solo que a un nivel inferior.

(1) La bibliografía existente sobre economía regional es ciertamente amplia, pero el tema que a nosotros nos ocupa, que es el de la regionalización, es decir, el estudio de los criterios para crear la región, es bastante repetitivo en la mayoría de los autores. Puede resumirse en lo expuesto por Richardson (1975) pags. 239 y ss. que coincide fundamentalmente con las ideas de Boudeville (1969) pag. 105. Algo distinto es el esquema citado por Tamames (1968) pgs. 20 y ss. que comentaremos más adelante.

Siguiendo a Richardson (1) tenemos que el criterio de regionalización más interesante no existe a nivel teórico. Existen diversos criterios, pero el mejor dependerá en cada caso de dos factores fundamentales:

- Objetivos de la regionalización.
- Características concretas de la Nación a estudiar.

Podemos resumir los diversos criterios de regionalización en los siguientes puntos:

a) Divisiones administrativas:

Hace referencia al caso, muy corriente por otra parte, de que ante la inexistencia de datos a otro nivel, tengamos que conformarnos con la regionalización administrativa (provincias, etc.) a la hora de hacer nuestros estudios.

b) Región homogénea:

Es la idea de la región natural. Lo que se pretende es crear regiones que tengan un fuerte factor de homogeneidad.

La dificultad estriba en cuál debe ser ese factor de homogeneidad. Podemos hablar desde características topográficas, recursos naturales, climas, etc. hasta variables no eco

(1) Richardson (1.976) pgs. 239 y ss. Este autor, aunque plantea los criterios que veremos a continuación, suele procurar evitar el problema de la regionalización.

nómicas como actitudes sociales, ambiente político-histórico, etc. (1)

Otra dificultad se presenta a la hora de establecer las fronteras entre las diversas regiones, pues amenudo las zonas limítrofes se parecen en unos casos a una región y en otros a la otra.

Un criterio interesante de homogeneidad es la renta per cápita, pero no siempre es posible obtener datos sobre el tema a nivel suficientemente desagregado. Otros criterios son la base económica de exportación, propuesto por North (2) las experiencias cíclicas de Vining (3), etc.

De todas formas el criterio de homogeneidad es difícil de aplicar de forma absoluta y hay que concluir que siempre dentro de una región homogénea habrá fuertes heterogeneidades como las producidas entre las áreas urbanas y rurales, etc.

c) Región Nodal:

Lo que pretende es crear la región alrededor de un nodo o polo, por lo que también se denominan regiones polarizada

(1) Carreras (1977) en la pg. 69, propone como criterios de homogeneidad, la agricultura, la renta per cápita, características étnicas, etc. o combinaciones de ellas.

(2) Este artículo de North (1955) refleja un planteamiento esencialmente Kegnesiano tal como expone Saenz de Buruga (1977) pag. 51. Y suscitó una conocida polémica con Tiebout (1956).

(3) Vining (1949) pgs. 89 - 104.

das. Es decir, alrededor de un centro hay una serie de núcleos que viven más o menos estrechamente con él.

La región nodal tiene por lo tanto en cuenta el factor distancia de forma explícita y se compone de zonas heterogéneas interrelacionadas funcionalmente.

Por lo tanto, si seguimos el criterio de nodalidad, nos fijaremos más en cual es el centro (o centros) de la región que en cuales son sus fronteras. (1)

La nodalidad o polaridad pone su énfasis en el estudio de los flujos económicos y por lo tanto podemos estudiarlos como vectores con un módulo (intensidad), una dirección (entre que puntos) y un sentido (de un punto a otro).

Estos flujos pueden ser de muy diversos tipos, y así tenemos mercancías, tráfico, teléfono, estudiantes, etc. etc.

Hay tres tipos especialmente interesantes de organización espacial que acentúan el carácter nodal:

- (1) Este criterio mantenido por Richardson (1975) pg. 244, implica, como dice el propio Richardson, la dificultad existente luego a la hora de fijar las fronteras de la región nodal. Para solucionar esto Sampedro (1964) pg. 12, propone el "método de las zonas de dispersión". LLa mando así a las áreas que rechazan a un lado y a otro la actividad económica hacia opuestos centros de atracción.

- Orden jerárquico de los centros urbanos.
- Regiones metropolitanas.
- Areas de influencia metropolitana.

Normalmente el concepto de nodalidad lleva a una fuerte interrelación, no solo dentro de la región sino también - entre regiones (ya hemos hecho referencia al orden jerárquico de los centros urbanos) con lo que resulta a veces casi imposible delimitar las fronteras entre una región y otra.

Por otro lado dada la cantidad de flujos posibles creo que es interesante hacer hincapié en los flujos de servicios que son los más normalmente organizados a nivel regional.

d) Región - Plan:

Hacer referencia este tipo de región al caso en que se establece la misma con vistas a una planificación posterior. Por lo tanto, normalmente se trata de alguno de los tipos vistos anteriormente. En cada caso habrá que buscar el concepto de región que más se adapte a la idea de planificación que nosotros tenemos a priori.

Con mucha frecuencia se tiene que acomodar a la existencia de unas regiones administrativas por exigirlo así la ordenación política del estado, o bien, por la simple carencia de datos a otros niveles.

En general se acepta que la unidad de planificación debe ser siempre la región nodal y en particular la región metropolitana. (1)

Una vez que hemos estudiado los diversos tipos de regionalización posible, vamos a intentar hacer un resumen - práctico del problema.

Primero nos hemos de plantear cual va a ser el número y tamaño de regiones que deseamos. Este es un fuerte factor de ambigüedad a la hora de tratar el problema. Lo más lógico parece ser el crear una jerarquía de regiones, primero unas de primer orden, y luego ir dividiendo cada una en otras de segundo orden y así sucesivamente. Este es el sistema de regiones, subregiones, comarcas, etc. (2)

En este sentido parece que se estructurará la organización política de España a partir de la constitución de 1.978, donde tenemos en primer lugar la Nación, luego los entes autonomicos (según las nacionalidades o regiones), después van - las provincias y todavía sin estructurar serían posibles las comarcas dentro de cada provincia. En esta línea y para el - caso concreto del país vasco, tenemos el estudio realizado -

- (1) Carreras (1.977) pg. 69, indica que la región nodal tiene especial interés para los problemas microeconómicos y la homogénea para los macroeconómicos.
- (2) El sistema jerárquico es especialmente típico de las regiones polarizadas como dice Boudeville (1968) pg. 27.

por el gabinete de investigación "Organización y Desarrollo" dirigido por Uriarte. (1)

Por otro lado el tipo de regionalización a ~~de~~ decidir de penderá fundamentalmente de qué es lo que estamos regionalizando y para qué.

Hemos planteado, por lo tanto, los conceptos fundamentales que desde un punto de vista económico es preciso abordar a la hora de plantearnos una regionalización.

Pero el problema de la regionalización es claramente interdisciplinar, no son solo los economistas, sino también los sociólogos, geógrafos, demógrafos y planificadores los que han de dar su opinión sobre el tema.

Ante la complejidad del problema y siendo conscientes de que cada regionalización debe de atender fundamentalmente a su objetivo y a donde se va a aplicar, hemos decidido crear nuestro propio modelo de regionalización.

Nuestro objetivo es comarcalizar la provincia de La - Rioja, por lo tanto, nos encontramos de partida con una divi sión administrativa que nos viene impuesta que es la provincia (posible ente autónomo en el futuro).

(1) O. y D. (1.980).

Por otro lado a la hora de comarcalizar nuestra unidad de partida va a ser el municipio, que es de donde tenemos datos, luego aquí tenemos una nueva división administrativa que nos viene dada.

Luego, tanto en el espacio a comarcalizar (provincia), como la unidad básica de comarcalización (municipio), son conceptos dados a priori.

La razón de esto es evidente, pues si queremos que nuestra comarcalización tenga alguna utilidad, ha de partir de unidades político-administrativas que serán las que hay que comarcalizar. Por otro lado, el municipio es la única unidad, inferior a la provincia, de donde se pueden obtener datos en nuestro caso.

Con estas dos premisas de partida hemos de decidirnos por la regionalización nodal o la homogénea (1), a la hora de establecer la comarcalización.

- (1) Dado que estos son los dos tipos fundamentales y a los cuales van a parar todos los autores. Así tenemos que Tamames(1968) pgs. 23 y ss. expone cinco métodos distintos: el de Vidal de la Blache (1910), y el de Reilly (1929) que son fundamentalmente nodales; combinando este último con áreas administrativas o regiones históricas Tamames propone el tercer método; luego estarían el homogéneo y el ya citado de "zonas de dispersión" del profesor Sampedro (1964). En definitiva se manejan los conceptos de homogeneidad y nodalidad, bien puros, bien mezclados con regiones administrativas, etc. y en nuestro caso las limitaciones impuestas por las regiones administrativas, ya las hemos tenido en cuenta.

Dijimos que para establecer el criterio óptimo había que basarse en dos datos:

- Qué vamos a comarcalizar.
- Para qué.

En primer lugar hemos de plantearnos, por lo tanto, el objetivo de nuestra comarcalización y éste ha de ser el de crear unas unidades que tengan dos características fundamentales:

- Cierta homogeneidad interna.
- Fuerte relación.

Esto viene dado porque la homogeneidad es interesante a la hora de un estudio de los problemas de cada comarca y a la hora de planificar su desarrollo de una forma armónica. No es que esto sea necesario, y de hecho muchas regiones de planificación son altamente heterogéneas, pero veremos que en el caso concreto de La Rioja este planteamiento resulta específicamente interesante.

En cuanto a la relación interna dentro de la comarca, parece evidente que esto es conveniente para poder potenciar los centros nodales, comarcalizando servicios y disminuyendo la dependencia de los pequeños o medianos núcleos de población respecto a la capital.

En segundo lugar hemos dicho que es importante saber que es lo que comarcalizamos. En nuestro caso nos referimos a la provincia riojana, así hemos de fijarnos en sus especiales características.

Es un hecho que el peso porcentual del sector agrícola va descendiendo tanto en España como en el resto de los países en desarrollo. La Rioja no es una excepción a esta ley, pero de todas formas la agricultura tiene un peso mayor que en las provincias que le rodean, sobre todo si lo comparamos con el país vasco, que es con el que tiene mayor relación.

Por otro lado, a parte del sector agrícola, tienen en nuestra provincia un peso muy alto las industrias derivadas de la agricultura, y así tenemos, por ejemplo, el caso de las conservas agrícolas y la industria de transformación del vino.

Es por todo lo dicho que considero que en un desarrollo armónico de la Rioja, la agricultura, bien directamente bien de forma indirecta, ha de jugar un papel fundamental.

Ante esta premisa creo que a la hora de comarcalizar hemos de tener en cuenta la homogeneidad agrícola de las zonas a estudiar. Esto nos dará una base de futuro desarrollo común y una posibilidad de estudio unificado de los problemas de cada comarca.

De todas formas, siguiendo a Richardson (1), hemos de afirmar que la comarcalización homogénea basada en datos agrícolas, no siempre es correcta. Así tenemos los deficientes - resultados que obtuvieron Bogue y Beale (2) en su regionalización de los Estados Unidos.

Pero nuestro caso es bastante diferente dada la dependencia riojana respecto a su agricultura.

Por otro lado carecemos de otro tipo de datos (como - podrían ser renta per cápita, por municipios, etc.) que nos permitan otro tipo de homogeneidad.

En último lugar introduciremos también el concepto de nodalidad, tan importante a la hora de comarcalizar servicios etc..

En conclusión tenemos que el objetivo de nuestra comarcalización lo establecemos en que sea útil para los planificadores del desarrollo y equilibrio provincial, y que dadas las características de la Rioja nos inclinamos por un sis

(1) Richardson (1.975), pag. 241.

(2) Bogue y Beale (1.961), intentaron la regionalización homogénea basada fundamentalmente en la homogeneidad agrícola.

tema mixto Homogeneidad - Nodalidad, que es el que vamos a aplicar. (1)

Por lo tanto ya hemos definido lo que vamos a hacer, nos resta ver cómo vamos a aplicar en nuestro caso las técnicas de análisis multivariante.

(1) Este sistema que vamos a establecer puede considerarse bastante original como veremos en el desarrollo del mismo.

III.- APLICACIONES DEL ANALISIS MULTIVARIANTE A LA COMARCALIZACION.

En primer lugar, creo que es importante reseñar aquí - que en este caso, dado que vamos a comarcalizar una provincia a partir de sus municipios, no existe ningún problema de inferencia estadística, dado que nos vamos a limitar a un análisis de datos.

Hecha esta salvedad estadística, diremos que dentro - del análisis multivariante, el análisis factorial ha sido el más usado para la comarcalización, sobre todo para la homogénea (1), pero también para la nodal (2). Nosotros lo que vamos a hacer es utilizar criterios mixtos basándonos en el - análisis factorial y en el cluster análisis.

En primer lugar crearemos comarcas homogéneas según - una serie de factores provenientes de las variables que consigamos.

En segundo lugar, y según otras variables aplicables al concepto de centro nodal, crearemos los factores correspondientes, que darán lugar a los nodos o polos de atracción.

(1) Richardson (1.976), pag. 113, comenta este hecho.

(2) Como puede verse en Aznar (1.974), pag. 112 y ss.

Estudiaremos a continuación el área de influencia de cada polo y las relaciones de los polos entre sí.

En último lugar crearemos nuestras comarcas con el criterio mixto ya citado.

En definitiva vamos a utilizar intensamente el análisis multivariante pero, como ya veremos, será preciso el completarlo con otros casos, como flujos, etc.

Los flujos también son susceptibles de un tratamiento estrictamente estadístico, pero lo que sucede es que sería necesario tenerlos en forma rigurosamente cuantitativa.

A causa de esta limitación, el único flujo que se puede utilizar normalmente es el de llamados telefónicos, pero con la creciente automatización de la red, incluso esta variable se está perdiendo. Una aplicación a este caso es el estudio por cadenas de Markov, hecho por el antiguo ministro de planificación del desarrollo. (1)

Pero en la provincia de Logroño carecemos de datos sobre intensidad telefónica entre los municipios, tal como nos informaron en la Compañía Telefónica (2), a parte de lo dudoso

(1) Este trabajo fue publicado por el propio Ministerio en 1.973 y después por sus autores Russinés Torregrosa y Pascual Güell (1974). Utiliza las llamadas telefónicas entre provincias.

(2) Esto es a causa de la automatización de la red dentro de la provincia, que lleva ya bastantes años.

Esta automatización ha hecho perder este dato, y no solo en España, sino también en Francia como comenta Boudeville (1963), pag. 30

so de este estimador de la relación entre municipios. (1)

El resto de datos de flujos tampoco los tenemos en forma cuantitativa, a excepción de la intensidad de tráfico por carretera, pero incluso este dato lo tenemos de forma incompleta dado que los aforos no están controlados en todas las carreteras como nos informaron en la Dirección Provincial de Carreteras. (2)

Luego, aunque sería interesante, los flujos los tendremos que introducir prescindiendo de los métodos estadísticos.

Lo mismo que decimos de los flujos, podríamos decir de otros muchos conceptos, que la penuria de datos fiables nos impide introducir.

- (1) De todas formas cuando se podría obtener era uno de los mejores a causa de su cuantificación.
- (2) Por otro lado los citados aforos solo nos dicen los coches que pasan, pero nada de su dirección y destino.

IV.- PROBLEMATICA DE LA BUSQUEDA DE INFORMACION.

Ante el planteamiento ya expuesto nos enfrentábamos con la imperiosa necesidad de encontrar información referida a municipios y la situación fue realmente desalentadora pues la carencia de datos a todos los niveles, pero especialmente a nivel municipal, era impresionante.

Los organismos y publicaciones consultados para confeccionar nuestro banco de datos fueron los siguientes:

- Diputación Provincial:
 - Fichero del Gabinete de Planificación.
 - Resumen del estudio socio-económico de la provincia (Iberplan).
 - Plan general de ordenación urbana de la comarca de Logroño.
 - Informes previos al IV plan de desarrollo (1.974).
 - Estudio de áreas deprimidas (Sierra - Cameros).
 - Mapas y datos diversos.

- Cámara Oficial de Comercio e Industria de Logroño:
 - Memoria descriptiva del desarrollo económico de la provincia de Logroño (Diciembre 1.977).
 - Licencias fiscales por ramas (1.975).

- Instituto Nacional de Estadística (entre otras muchas publicaciones):
 - Reseña estadística provincial.
 - Censo de población (1.970).
 - Censos de la población, de la vivienda y de los edificios en España (provincia de Logroño) (1,970).
 - Nomenclator de las ciudades, villas, lugares, aldeas y demás entidades de población (1.970).
 - Anuario estadístico.

- Información de la Antigua Organización Sindical:
 - Estructura y perspectivas de desarrollo económico de la provincia de Logroño (1.970).
 - Otras publicaciones.

- Cesin:
 - Tablas básicas de estructura socio-económica.

- Cesie.
- Caja de Ahorros y Monte de Piedad de Zaragoza, Aragón y Rioja:
 - Tablas input-output y cuentas regionales de la economía riojana.

- Guía - anuario de datos comerciales publicada por el heraldo de Aragón con datos de las Cámaras de Comercio e Industria.
- Instituto de estudios de administración local, dependiente en la actualidad del Ministerio de Administración Territorial:
 - Publicación de datos por municipios (1.974).
- Ministerio de Agricultura (datos sobre cultivos).
- Ministerio de Industria.
- Instituto Nacional de Previsión.
- Ministerio de Obras Públicas y Urbanismo, Dirección Provincial de Carreteras:
 - Datos sobre intensidades de tráfico (1.975).
- Oficina de Turismo:
 - Guía hotelera (1.978).
- Tribunal tutelar de menores (Inspección):
 - Recaudación del impuesto sobre espectáculos (1.977).
- Colegio Oficial de Médicos:
 - Médicos en ejercicio (1.978).
- Delegaciones de Ministerios en Logroño.
- Ayuntamientos de la provincia.

Una vez analizada toda la información, concluimos que la fuente más adecuada para nuestro estudio era el fichero - del Gabinete de Planificación de la Diputación Provincial de Logroño.

Este fichero contiene gran cantidad de información, de bastante fiabilidad (ya comentaremos algunas excepciones) y sobre todo tiene para nosotros la enorme ventaja de encontrarse desagregada a nivel municipal.

Tengamos en cuenta que la mayor parte de la información consultada solo tiene datos a nivel provincial o a lo sumo comarcal (1), pero difícilmente nos encontramos con datos a nivel municipal, que ya hemos dicho que son los que nos interesan. A causa de esto, gran parte de la información consultada solo nos ha servido para contrastar los datos - agregados y ver la coherencia o no de las fuentes utilizadas.

Este fichero básico lo he completado con datos de - otras fuentes (Cámara de Comercio e Industria, etc.) pero ha habido dos datos que ha sido imposible conseguir:

- El primero ha sido el referente a población activa, que aunque en el fichero constaba para algunos municipios, carecíamos de él para las ciudades más importantes (Logroño, etc.) a causa de no tener mecanizado el censo. Hay

(1) Respecto a los datos a nivel comarcal, es la formada por Logroño capital y los pueblos que le rodean la que posee mayor información.

estimaciones de este dato a nivel provincial, pero no municipal, por otro lado, algunas estimaciones municipales consultadas se contradicen fuertemente entre sí, con lo que la fiabilidad de este dato es muy dudosa. Tengamos en cuenta que - en la provincia de Logroño, sobre todo en las zonas donde hay industria, es muy normal que haya trabajadores en la industria que a la vez tienen sus tierras, con lo que a veces es difícil encuadrarles en la agricultura o en la industria.

Otra fuente de distorsión es el trabajo femenino en la agricultura, pues es difícil decir si una mujer trabaja o no fuera de casa cuando lleva el hogar familiar y además colabora en las tareas agrícolas tanto o más que el hombre. Un caso similar sucede con los menores de edad, en multitud de sitios donde ayudan en la explotación familiar.

Dado que los datos muchas veces provienen de apreciaciones distintas, se producen grandes contradicciones, lo que hacen que la información no sea fiable a nivel municipal, - por lo que prescindiremos de ese dato.

- Otro dato de imposible consecución ha sido el de -
trabajadores por sectores en la industria, dado que el I.N.P. no ha podido facilitárnoslo (1). Incluso no se puede conocer el número de industrias por sectores, dado que la -
única estimación es la de la Cámara de Comercio e Industria y el propio economista de la Cámara nos reconoció la poca -
fiabilidad de la misma.

(1) Anexo página siguiente.

(1) A este respecto quiero comentar que el servicio de esta
dística del I.N.P. en Logroño nos informó de que era po
sible conocer los datos a nivel agregado para toda la provinci
cia, pero a causa de la falta de mecanización de este servici
o, carecían de los datos por municipios.

También nos hemos informado que el I.N.E. está elabora
ndo un censo industrial de España, pero puestos en contact
o con el citado Instituto, hemos podido saber que en 1.980
todavía está en fase de preparación y que solo se han publica
do los datos provisionales y a nivel agregado nacional. -
Por otro lado nos han informado que su publicación definitiva
va y la disponibilidad de los datos a nivel municipal han -
de tardar todavía bastante.

Ante la aparente imposibilidad de conocer la importan
cia y dimensión de los distintos sectores por municipios, de
cidí utilizar como único dato fiable y conocido las licencias
fiscales por ramas, aún siendo conscientes de sus limitaciones.

Con todo lo dicho, conseguimos la información que a -
continuación detallamos con la fuente y el año a que se refiere
re. Hemos intentado conseguir los últimos datos disponibles -
antes del proceso de la información que se realizó en los primi
meros meses de 1.979.

- Población activa (está introducida en el banco de datos,
para los municipios que la hemos podido conseguir, aunque
como ya hemos comentado no la utilizaremos por faltarnos
algunos) - Diputación 1.975.
- Superficie pavimentada de vías públicas - Diputación 1.975.
- Viviendas con agua corriente - Diputación 1.975.
- Alcantarillado y necesidades - Diputación 1.975.
- Datos sobre alumbrado - Diputación 1.975.
- Camas y hospitales - Diputación 1.975.
- Estudiantes de E.G.B. - Diputación 1.975.
- Centros escolares - Diputación 1.975.
- Centros de deporte - Diputación 1.975.
- Museos y bibliotecas - Diputación 1.975.
- Habitantes - Padrón 1.975.
- Extensión del término municipal - Diputación.
- Altura - Diputación.

- Hectáreas de regadío - Diputación 1.978.
- Hectáreas de cultivo - Diputación 1.978.
- Hectáreas de prados - Diputación 1.978.
- Hectáreas forestales - Diputación 1.978.
- Hectáreas de otras superficies - Diputación 1.978.
- Hectáreas de cereales - Diputación 1.978.
- Hectáreas de leguminosas - Diputación 1.978.
- Hectáreas de patata - Diputación 1.978.
- Hectáreas de cultivos industriales - Diputación 1.978.
- Hectáreas de cultivos forrajeros - Diputación 1.978.
- Hectáreas de hortalizas - Diputación 1.978.
- Hectáreas de frutales - Diputación 1.978.
- Hectáreas de almendro - Diputación 1.978.
- Hectáreas de viñedo - Diputación 1.978.
- Hectáreas de olivar y otros - Diputación 1.978.
- Cabezas de ganado - Diputación 1.974.
 - Ovino.
 - Bovino.
 - Caprino.
 - Cerda.
 - Trabajo.
- Ingresos municipales - Diputación 1.975.
- Cajas de ahorros - Cámara de Comercio 1.977.
- Bancos - Cámara de Comercio 1.977.
- Crecimiento demográfico - Diputación con datos hasta 1.975.
- Licencias fiscales de comercio al por menor y por mayor -
Listado de la licencia fiscal 1.975.
- Habitaciones en hoteles - Guía hotelera 1.978.

- Médicos en ejercicio - Colegio Oficial de Médicos de Logroño 1.978.
- Recaudación del Impuesto sobre Espectáculos - Inspección del citado impuesto 1.977.
- Servicios y Equipamientos Municipales - Diputación 1.975.
- Centros de atracción - Diputación 1.975.
- Intensidad del tráfico por las carreteras provinciales - Dirección provincial de carreteras 1.975.

Podemos ver que la mayor parte de nuestros datos provienen de la Diputación provincial, aunque han sido comparados y perfeccionados con otras fuentes en determinados casos. Por otro lado, los datos se refieren a 1.975, en su mayor parte, dado que es el año del padrón municipal y hasta 1.980 no tenemos un censo nuevo (recordemos que los censos se hacen los años terminados en cero y los padrones los terminados en cinco).

No todas las variables han sido susceptibles de un tratamiento por ordenador por no estar suficientemente cuantitativizadas, de todas formas, se ha intentado hacer lo posible para introducir el mayor número de variables.

También es preciso tener en cuenta que muchas variables no nos interesan en términos absolutos, sino relativos, por lo que se han pasado a tantos por ciento, ya explicaremos posteriormente las razones de todo esto.

Pasemos a ver como quedó nuestro cuadro de variables para el tratamiento mecanizado (las que no son susceptibles de este tipo de tratamiento, las introduciremos posteriormente).

VARIABLES:

- 1 - Porcentaje de población activa sobre población total.
- 2 - Porcentaje de población activa en la agricultura sobre población activa.
- 3 - Porcentaje de población activa en la industria sobre población activa.
- 4 - Porcentaje de población activa en los servicios sobre población activa.
- 5 - Porcentaje de superficie pavimentada de vías públicas sobre superficie total de vías públicas.
- 6 - Porcentaje de viviendas con agua corriente.
- 7 - Porcentaje de alcantarillado existente sobre alcantarillado existente más necesidades (según estimación de la Diputación).
- 8 - Porcentaje de viviendas familiares con alumbrado.
- 9 - Porcentaje de vías públicas con alumbrado.
- 10 - Camas por cada mil habitantes.
- 11 - Estudiantes de E.G.B. por cada cien habitantes.
- 12 - Centros de deporte (campos, piscinas y frontones).
- 13 - Centros escolares de E.G.B.

14 - Museos y Bibliotecas.

15 - Número de hospitales.

16 - Comercio:

- Supermercado.

- Mercado público.

- Papelería.

- Librería.

- Tejidos y confecciones.

- Calzado.

- Electrodomésticos.

- Muebles.

- Farmacias.

- Venta de automóviles y motocicletas.

- Venta de tractores y maquinaria agrícola.

- Distribuidor de butano.

17 - Servicios comerciales:

- Gestoría administrativa.

- Agencia de publicidad.

- Auto escuela.

- Bancos.

- Cajas de ahorro.

18 - Servicios profesionales:

- Practicante.

- Médico.

- Dentista.

- Veterinario.

- Notario.

- Abogado.

- Procurador.
 - Juez de 1ª Instancia.
 - Arquitecto.
- 19 - Otros servicios:
- Tintorerías.
 - Servicio fotográfico.
 - Pompas fúnebres.
 - Oficina de correos.
 - Central de teléfonos.
 - Oficina de telégrafos.
 - Cooperativas agrícolas.
 - Cuartelillo Policía Municipal.
 - Puesto Guardia Civil.
- 20 - Equipamiento urbano:
- Escuela primaria.
 - Instituto de enseñanza media.
 - Centro de enseñanza profesional.
 - Biblioteca pública.
 - Casa de socorro o urgencia.
 - Hospital.
 - Asilo.
 - Parque de bomberos.
 - Cine.
 - Clubs o salas de fiesta.
- 21 - Equipamiento religioso:
- Iglesia Parroquial.
 - Basílicas o Ermitas.
 - Monasterios.
 - Ordenes religiosas.

- 22 - Otros equipamientos:
 - Matadero municipal.
 - Edificios y locales destinados a la práctica de deportes.
 - Parques y jardines.
 - Médicos especialistas.
 - Asistencia a partos.

- 23 - Transportes:
 - Autobuses de línea.
 - Estación de ferrocarril.
 - Accesos por carretera:
 - Carreteras del estado.
 - Carreteras provinciales.
 - Caminos vecinales.
 - Otros caminos.

- 24 - Habitantes.
- 25 - Extensión del término municipal (km²).
- 26 - Altura (m).
- 27 - Porcentaje de tierras de regadío.
- 28 - Porcentaje de tierras de cultivo.
- 29 - Porcentaje de prados.
- 30 - Porcentaje de terreno forestal.
- 31 - Porcentaje de otras superficies.
- 32 - Porcentaje de hectáreas de cereales.
- 33 - Porcentaje de hectáreas de leguminosas.
- 34 - Porcentaje de hectáreas de patata.
- 35 - Porcentaje de hectáreas de cultivos industriales.
- 36 - Porcentaje de hectáreas de cultivos forrajeros.

- 37 - Porcentaje de hectáreas de hortalizas.
- 38 - Porcentaje de hectáreas de frutales.
- 39 - Porcentaje de hectáreas de almendro.
- 40 - Porcentaje de hectáreas de viñedo.
- 41 - Porcentaje de hectáreas de olivar y otros.
- 42 - Capital ganadero (en millones de pesetas).
- 43 - Porcentaje de Bovino (en pesetas).
- 44 - Porcentaje de Ovino (en pesetas).
- 45 - Porcentaje de Caprino (en pesetas).
- 46 - Porcentaje de Cerda (en pesetas).
- 47 - Porcentaje de animales de trabajo (en pesetas).
- 48 - Ingresos municipales (en miles de pesetas).
- 49 - Cajas de Ahorro.
- 50 - Bancos.
- 51 - Porcentaje de crecimiento demográfico en la década 60-70
- 52 - Porcentaje de crecimiento demográfico en el quinquenio 70-75
- 53 - Licencias fiscales de comercio al por mayor.
- 54 - Licencias fiscales de comercio al por menor.
- 55 - Licencias fiscales de fabricación - Rama 1.
- 56 - Licencias fiscales de fabricación - Rama 2.
- 57 - Licencias fiscales de fabricación - Rama 3.
- 58 - Licencias fiscales de fabricación - Rama 4.
- 59 - Licencias fiscales de fabricación - Rama 5.
- 60 - Licencias fiscales de fabricación - Rama 6.
- 61 - Licencias fiscales de fabricación - Rama 7.
- 62 - Licencias fiscales de fabricación - Rama 8.
- 63 - Licencias fiscales de fabricación - Rama 9.
- 64 - Licencias fiscales de artesanía (total).

- 65 - Habitaciones en hoteles.
- 66 - Médicos en ejercicio.
- 67 - Recaudación del impuesto sobre espectáculos (en miles de pesetas).

VARIABLES DERIVADAS:

- 42/24 - Capital ganadero per cápita (en millones de ptas.).
- 66/24 - Médicos por habitante.
- 67/24 - Recaudación del impuesto sobre espectáculos por ha bitante (en miles de pesetas).
- 24/25 - Habitantes por Km².
- 43' - Capital bovino per cápita (en decenas de miles de ptas.).
- 44' - Capital ovino per cápita (en decenas de miles de ptas.).
- 45' - Capital caprino per cápita (en decenas de miles de ptas.).
- 46' - Capital cerda per cápita (en decenas de miles de ptas.).
- 47' - Capital de trabajo per cápita (en decenas de miles de pt)

En conclusión tenemos estas variables para poder utilizar el análisis multivariante y vamos a estudiarlas detenidamente.

En primer lugar, tengo que decir que las variables - unas veces están expresadas en porcentajes, otras en términos absolutos y otras per cápita. La causa de esto es que yo introduje las variables en la forma que consideraba más conveniente para nuestro tratamiento, como veremos a continuación,

dado que los datos iniciales estaban todos en forma absoluta. Por otro lado, la introducción en tantos por ciento resulta más cómoda de cara al ordenador. Ya hemos comentado que el soporte de entrada en nuestro ordenador es la ficha perforada y en consecuencia, su perforación resulta laboriosa, por lo tanto, era deseable disminuir al mínimo este penoso trabajo y las cifras en tanto por ciento son más manipulables que en términos absolutos.

Por otro lado, como al entrar en el ordenador, según los programas que he utilizado, las variables se normalizan, la unidad de medida pierde importancia y podemos estudiar variables medidas de distinta forma, y de manera conjunta.

A continuación vamos a tratar el porqué de la forma de medición de las variables que presentan peculiaridades y la forma de puntuación de las que ofrezcan alguna dificultad.

De las variables 1 a 9, no existe ningún problema pues son porcentajes usuales.

La variable 10, se da en tanto por mil, dado que lo importante es el grado de asistencia a los vecinos en este caso. Es decir, si tiene un déficit o un superavit.

La variable 11, nos da una idea del grado de juventud y de escolarización, podemos decir que es una idea del futuro del municipio, si no se produjera emigración.

Las variables 12, 13, 14 y 15, están dadas en términos absolutos y nos indican el potencial del municipio en esos campos.

Las variables, 16, 17, 18, 19, 20, 21 y 22, nos indican la carencia o abundancia de servicios en general. Podíamos haber calculado el potencial de cada servicio, pero el problema es que carecíamos de datos y solo sabíamos si existía o no un determinado servicio en un municipio y no su cuantía. Es por esto y dada la división en variables que proponía el Gabinete de Planificación de la Diputación Provincial de Logroño, que decidimos el sumar los servicios de cada variable según si existían o no en el municipio, puntuando con un uno su existencia y con cero su carencia.

Hubiera sido más interesante crear una variable para cada servicio y así tendríamos variables dicotómicas. Pero dada la capacidad de nuestro ordenador, hubiera sido imposible el tratarlos.

La variable 23, y de acuerdo con los criterios de los expertos en el tema y conocedores de la provincia, se puntuó de la siguiente manera:

- Autobuses de línea: 2 puntos.
- Estación de ferrocarril: 5 puntos.
- Carreteras del estado: 4 puntos.
- Carreteras provinciales: 3 puntos.

- Caminos vecinales y locales: 2 puntos.
- Otros caminos: 1 punto.

Así teníamos el potencial de las comunicaciones de ca
da municipio.

Las variables 24, 25 y 26, no ofrecen dificultad.

Las variables 27 a la 41, están dadas en tantos por -
ciento para evitar que los municipios de gran extensión fue-
sen superiores en todo a los de pequeña extensión. Con el -
criterio de porcentajes ponemos a todos en una situación si
milar.

La variable 42 tiene también peculiaridades. Los datos
que poseíamos venían dados en cabezas de ganado y es eviden-
te que no podemos comparar las vacas, cabras, etc. Ante este
problema hemos decidido transformar las cabezas de ganado en
pesetas y así poder sumarlas.

Para esta transformación el procedimiento no fue sen-
cillo dado que una vaca, por ejemplo, tiene cotizaciones muy
dispares según si es buena o mala y los datos que poseía no
decían nada al respecto.

Ante esta situación consideré que lo mejor era buscar
los precios medios de la cotización del ganado en la época de
confección del banco de datos (finales de 1.978). Para ello
consulté las cotizaciones en los mercados de ganado en la

provincia y en sus alrededores y los precios que consideré fueron los siguientes:

- Vaca: 60.000 pesetas.
- Oveja: 4.000 pesetas.
- Cabra: 4.000 pesetas.
- Cerdo: 7.000 pesetas.
- Trabajo: 50.000 pesetas.

Las variables 43, 44, 45, 46 y 47, nos indican las proporciones que en capital guardan los distintos tipos de ganado.

Las variables 48, 49 y 50, no merecen comentario especial.

Las variables 51 y 52, nos dan el crecimiento (normalmente negativo) de la población de los municipios en la década de los 60 y en el quinquenio 70 - 75, que es hasta donde tenemos datos, pues fue la fecha de elaboración del último padrón.

Las variables 53 a la 64, nos indican el número de licencias fiscales de comercio (tanto al por mayor como al por menor), fabricación (según las nueve ramas de la licencia fiscal que especificaremos a continuación) y de artesanía (en -totales, esta última, por no aumentar más el número de variables, que como ya hemos comentado dificultaría enormemente -la utilización de nuestro ordenador).

Veamos pues, como decíamos, las ramas de la licencia fiscal:

- Rama 1 - Alimentación.
- Rama 2 - Industria textil.
- Rama 3 - Madera, corcho, papel y artes gráficas.
- Rama 4 - Piel, calzado y cuacho.
- Rama 5 - Industrias químicas.
- Rama 6 - Industrias de la construcción, vidrio y cerámica.
- Rama 7 - Industrias metalúrgicas.
- Rama 8 - Energías eléctricas y mecánicas, gas ciudad y agua.
- Rama 9 - Actividades diversas.

En el banco de datos lo que haremos es introducir el número de licencias por ramas, etc. En realidad esto nos da cierta idea del potencial de cada municipio según especialidades. Lo que no cabe duda es que el dato no es del todo significativo, pero ya hemos comentado que es el único dato disponible para nosotros, ya que no nos han proporcionado ninguno mejor.

Las variables 65 y 66, no merecen especial comentario, pues nos dan el potencial municipal, en términos absolutos, en sus distintas especificaciones.

La variable 67, tiene especial interés, dado que nos indica una faceta importante del potencial de atracción del municipio, como son los espectáculos.

El resto de variables derivadas, se calcularon a partir de las anteriores, con idea de dar unos porcentajes per cápita que consideramos interesantes (todas son per cápita, a excepción de la 24/25 que nos da habitantes por Km²). Especial mención merecen las variables 43', 44', 45', 46' y 47', que nos dan el capital ganadero per cápita en los distintos tipos de ganado; al realizar el análisis multivariante comentaremos el interés peculiar de estas variables.

Una vez analizadas las variables, nos encontramos con que estos se refieren a los 180 municipios que componen la provincia de Logroño.

En primer lugar hemos de comentar que en 1.975 solo quedaban 175 de los 180 municipios, pues cinco habían sido anexionados, a causa de esto, los municipios van a tener - una doble numeración, la del fichero de la Diputación que en adelante denominaremos oficial y que va del 1 al 180 y la muestra del ordenador que va del 1 al 175.

Los cinco municipios anexionados carecen de datos - en las hojas del banco de datos y son:

<u>Municipio</u>	<u>Número</u>
Montalvo de Cameros	95
Poyales	114
La Santa	131

<u>Municipio</u>	<u>Número</u>
Turruncun	156
Zenzano	179

Damos a continuación la lista de los municipios de la provincia con su número oficial, su número de ordenador y la comarca a la que pertenecen oficialmente, según la Diputación (esta comarcalización está obsoleta y ni la propia Diputación la utiliza en la actualidad, por lo que prescindiremos de ella, pero la damos a título ilustrativo), también está la extensión del término municipal en Km² y la altura en metros.

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
1	1	Abalos	Haro	18'03	583
2	2	Agoncillo	Logroño	34'14	346
3	3	Aguilar R.A.	Cervera	54'26	637
4	4	Ajamil	Logroño	66'15	1.045
5	5	Albelda I.	Logroño	22'93	542
6	6	Alberite	Logroño	20'16	452
7	7	Alcanadre	Logroño	30'94	346
8	8	Aldeanueva E.	Calahorra (1)	39'01	343
9	9	Alesanco	Nájera	17'31	568
10	10	Aleson	Nájera	6'41	576
11	11	Alfaro	Calahorra	193'36	302
12	12	Almarza C.	Logroño	28'00	1.079
13	13	Anguciana	Haro	5'05	472
14	14	Anguiano	Nájera	90'63	663
15	15	Arenzana A.	Nájera	8'21	543
16	16	Arenzana Ar.	Nájera	5'96	598
17	17	Arnedillo	Arnedo	48'23	654
18	18	Arnedo	Arnedo	69'27	550
19	19	Arrubal	Logroño	7'52	357
20	20	Ausejo	Logroño	56'69	556
21	21	Autol	Calahorra	85'30	458
22	22	Azofra	Nájera	11'76	559
23	23	Badaran	Nájera	20'47	624
24	24	Bañares	Sº Domingo	29'68	591
25	25	Baños Rioja	Sº Domingo	9'21	550
26	26	Baños Rio T.	Nájera	17'50	575
27	27	Berceo	Nájera	5'80	727

(1) Entiéndase Calahorra - Alfaro.

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
28	28	Bergasa	Arnedo	26'94	645
29	29	Bergasillas B.	Arnedo	9'68	843
30	30	Bezares	Nájera	4'50	717
31	31	Bobadilla	Nájera	4'69	586
32	32	Brieva C.	Nájera	46'00	963
33	33	Briñas	Haro	2'40	454
34	34	Briones	Haro	37'79	501
35	35	Cabezón C.	Logroño	11'99	916
36	36	Calahorra	Calahorra	94'41	358
37	37	Camprovin	Nájera	20'60	673
38	38	Canales S.	Nájera	40'00	1.027
39	39	Canillas R.T.	Nájera	3'52	630
40	40	Cañas	Nájera	9'84	657
41	41	Cardenas	Nájera	4'02	566
42	42	Casalarreina	Haro	8'08	500
43	43	Castañares R.	Haro	10'94	546
44	44	Castroviejo	Logroño	20'70	955
45	45	Celórigo	Haro	12'32	790
46	46	Cenicero	Logroño	31'81	436
47	47	Cervera R.A.	Cervera	151'89	543
48	48	Cidamon	Sº Domingo	15'41	583
49	49	Cihuri	Haro	9'71	486
50	50	Cirueña	Sº Domingo	12'17	752
51	51	Clavijo	Logroño	19'53	872
52	52	Cordovin	Nájera	4'61	594
53	53	Corera	Logroño	8'21	522
54	54	Cornago	Cervera	79'02	757

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
55	55	Corporales	Sº Domingo	8'38	736
56	56	Cuzcurrita T.	Haro	19'18	519
57	57	Daroca R.	Logroño	11'22	726
58	58	Enciso	Arnedo	33'11	813
59	59	Entrena	Logroño	20'69	558
60	60	Estollo	Nájera	16'08	751
61	61	Ezcaray	Sº Domingo	142'69	813
62	62	Foncea	Haro	22'65	698
63	63	Fonzaleche	Haro	16'93	562
64	64	Fuenmayor	Logroño	33'95	434
65	65	Galbarruli	Haro	15'40	652
66	66	Galilea	Logroño	10'03	563
67	67	Gallinero C.	Torrecilla C.	11'28	1.055
68	68	Gimileo	Haro	4'00	482
69	69	Grañón	Sº Domingo	30'92	723
70	70	Grávalos	Cervera	30'91	765
71	71	Haro	Haro	40'32	479
72	72	Herce	Arnedo	17'08	592
73	73	Herramelluri	Haro	10'85	570
74	74	Hervias	Sº Domingo	13'95	655
75	75	Hormilla	Nájera	16'00	521
76	76	Hormilleja	Nájera	7'35	495
77	77	Hornillos C.	Logroño	11'98	1.052
78	78	Hornos M.	Logroño	7'48	684
79	79	Huércanos	Nájera	18'73	516
80	80	Igea	Cervera	56'07	546
81	81	Jalón C.	Torrecilla C.	8'41	875

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
82	82	Laguna C.	Torrecilla C.	41'54	1.042
83	83	Lagunilla J.	Logroño	22'92	645
84	84	Lardero	Logroño	20'26	435
85	85	Ledesma C.	Nájera	12'13	748
86	86	Leiva	Sº Domingo	12'60	587
87	87	Leza R.L.	Logroño	11'21	567
88	88	Logroño	Logroño	77'44	384
89	89	Lumbreras	Torrecilla C.	141'99	1.183
90	90	Manjarres	Nájera	6'13	629
91	91	Mansilla	Nájera	148'14	898
92	92	Manzanares R.	Sº Domingo	17'84	805
93	93	Matute	Nájera	25'61	680
94	94	Medrano	Logroño	7'55	598
95	--	Montalvo C.	Ane.S.Román C.	4'64	1.073
96	95	Munilla	Arnedo	27'13	791
97	96	Murillo R.L.	Logroño	46'16	407
98	97	Muro Aguas	Arnedo	31'09	891
99	98	Muro Cameros	Torrecilla C.	15'93	1.115
100	99	Nájera	Nájera	39'87	485
101	100	Nalada	Logroño	24'63	623
102	101	Navajun	Cervera	16'37	924
103	102	Navarrete	Logroño	28'32	512
104	103	Nestares	Torrecilla C.	21'63	860
105	104	Nieva C.	Torrecilla C.	41'83	1.021
106	105	Ocón	Arnedo	60'81	653
107	106	Ochanduri	Haro	11'64	556
108	107	Ojacastro	Sº Domingo	44'29	792

<u>Nº.</u>	<u>Nº</u>				
<u>of.</u>	<u>ord.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
109	108	Ollauri	Haro	2'57	493
110	109	Ortigosa	Torrecilla C.	35'29	1.069
111	110	Pazuengos	Sº Domingo	25'16	1.162
112	111	Pedroso	Nájera	19'46	773
113	112	Pinillos	Torrecilla C.	11'82	1.026
114	---	Poyales	Anex. Arnedo	36'92	954
115	113	Prodejón	Calahorra	31'40	362
116	114	Pradillo	Torrecilla C.	10'23	880
117	115	Prejano	Arnedo	42'11	708
118	116	Quel	Arnedo	55'98	490
119	117	Rabanera	Torrecilla C.	13'87	965
120	118	Rasillo (El)	Torrecilla C.	13'69	1.101
121	119	Redal (El)	Logroño	8'33	528
122	120	Ribafrecha	Logroño	34'61	498
123	121	Rincón de Soto	Calahorra	19'58	285
124	122	Robres C.	Logroño	35'81	722
125	123	Rodezno	Haro	14'36	547
126	124	Sajazarra	Haro	13'72	520
127	125	San Asensio	Haro	32'25	535
128	126	San Millán C.	Nájera	40'72	728
129	127	San Millán Y.	Haro	10'62	656
130	128	San Román C.	Torrecilla C.	35'77	849
131	---	Santa (La)	Anex. Munilla	27'00	1.179
132	129	Santa Coloma	Nájera	20'40	763
133	130	Santa Engracia J.	Logroño	85'92	648
134	131	Santa Eulalia B.	Arnedo	8'43	619
135	132	Sº Domingo C.	Sº Domingo	39'91	638

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
136	133	San Torcuato	Sº Domingo	10'84	599
137	134	Santurde	Sº Domingo	15'38	713
138	135	Santurdejo	Sº Domingo	18'28	780
139	136	San Vicente S.	Haro	48'33	528
140	137	Sojuela	Logroño	15'28	668
141	138	Sorzano	Logroño	10'09	718
142	139	Sotes	Logroño	14'78	672
143	140	Soto Cameros	Torrecilla C.	30'71	717
144	141	Terroba	Torrecilla C.	8'81	791
145	142	Tirgo	Haro	9'00	523
146	143	Tobía	Nájera	34'73	680
147	144	Tormantos	Sº Domingo	10'99	609
148	145	Torrecilla C.	Torrecilla C.	30'27	774
149	146	Torrecilla A.	Nájera	4'34	617
150	147	Torre C.	Torrecilla C.	11'61	1.154
151	148	Torremontalvo	Haro	8'28	429
152	149	Treviana	Haro	34'65	587
153	150	Trevijano	Torrecilla C.	8'48	1.012
154	151	Tricio	Nájera	6'32	564
155	152	Tudelilla	Calahorra	18'97	548
156	---	Turruncun	Anex. Arnedo	16'48	896
157	153	Uruñuela	Nájera	10'46	499
158	154	Valdemadera	Cervera	13'92	971
159	155	Valgañón	Sº Domingo	31'66	968
160	156	Ventosa	Nájera	9'37	654
161	157	Ventrosa	Nájera	72'83	1.003
162	158	Viguera	Logroño	56'07	689

<u>Nº.</u>	<u>Nº.</u>	<u>Municipio</u>	<u>Comarca</u>	<u>Extensión</u>	<u>Altura</u>
163	159	Villalba R.	Haro	8'94	579
164	160	Villalobar R.	Sº Domingo	10'88	590
165	161	Villamediana I.	Logroño	20'55	443
166	162	Villanueva C.	Torrecilla C.	19'95	914
167	163	Villar Arnedo (El)	Calahorra	18'22	429
168	164	Villar T.	Nájera	12'01	766
169	165	Villarejo	Nájera	6'37	834
170	166	Villarroya	Nájera	11'84	925
171	167	Villarta Q.	Sº Domingo	24'66	743
172	168	Villavelayo	Nájera	40'00	964
173	169	Villaverde R.	Nájera	5'80	823
174	170	Villoslada C.	Torrecilla C.	94'56	1.072
175	171	Viniegra A.	Nájera	65'50	881
176	172	Viniegra Ar.	Nájera	38'47	1.182
177	173	Zarratón	Haro	18'63	563
178	174	Zarzosa	Arnedo	18'35	981
179	---	Zenzano	An. Lagunilla J.	11'42	1.001
180	175	Zorraquín	Sº Domingo	6'35	862

Una vez que tenemos el listado de variables y el de pueblos, podemos pasar a ver el banco de datos que está en las hojas de ordenador nº 643 y ss.

En estas hojas están las variables de cada municipio, y éste viene representado por dos números, en la primera columna el nº de ordenador (ORD.) que será el que utilizemos en lo sucesivo, mientras no se diga lo contrario, y en la segunda columna el número del fichero de la diputación (NO.).

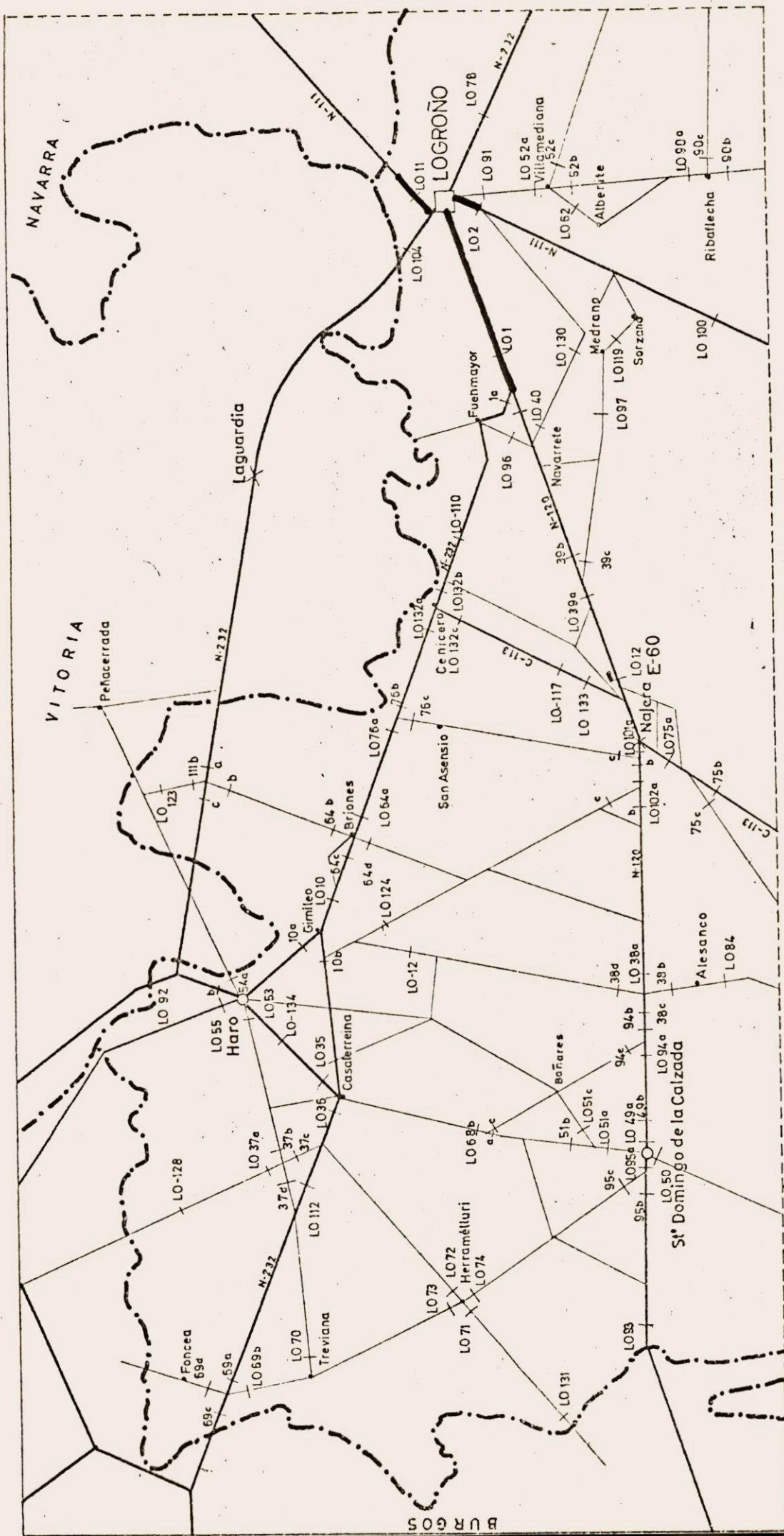
(Hemos comentado que en algunos municipios carecemos de datos para la variable población activa, en esos casos el ordenador ha rellenado con ceros).

Por otro lado, tenemos la información que por no ser susceptible de un tratamiento matemático, es preciso introducir de forma más subjetiva. Esta información es fundamentalmente la referida a flujos, pues en la mayor parte de los casos los tenemos sin cuantitativizar, y tendrán especial importancia a la hora de introducir la nodalidad o polaridad.

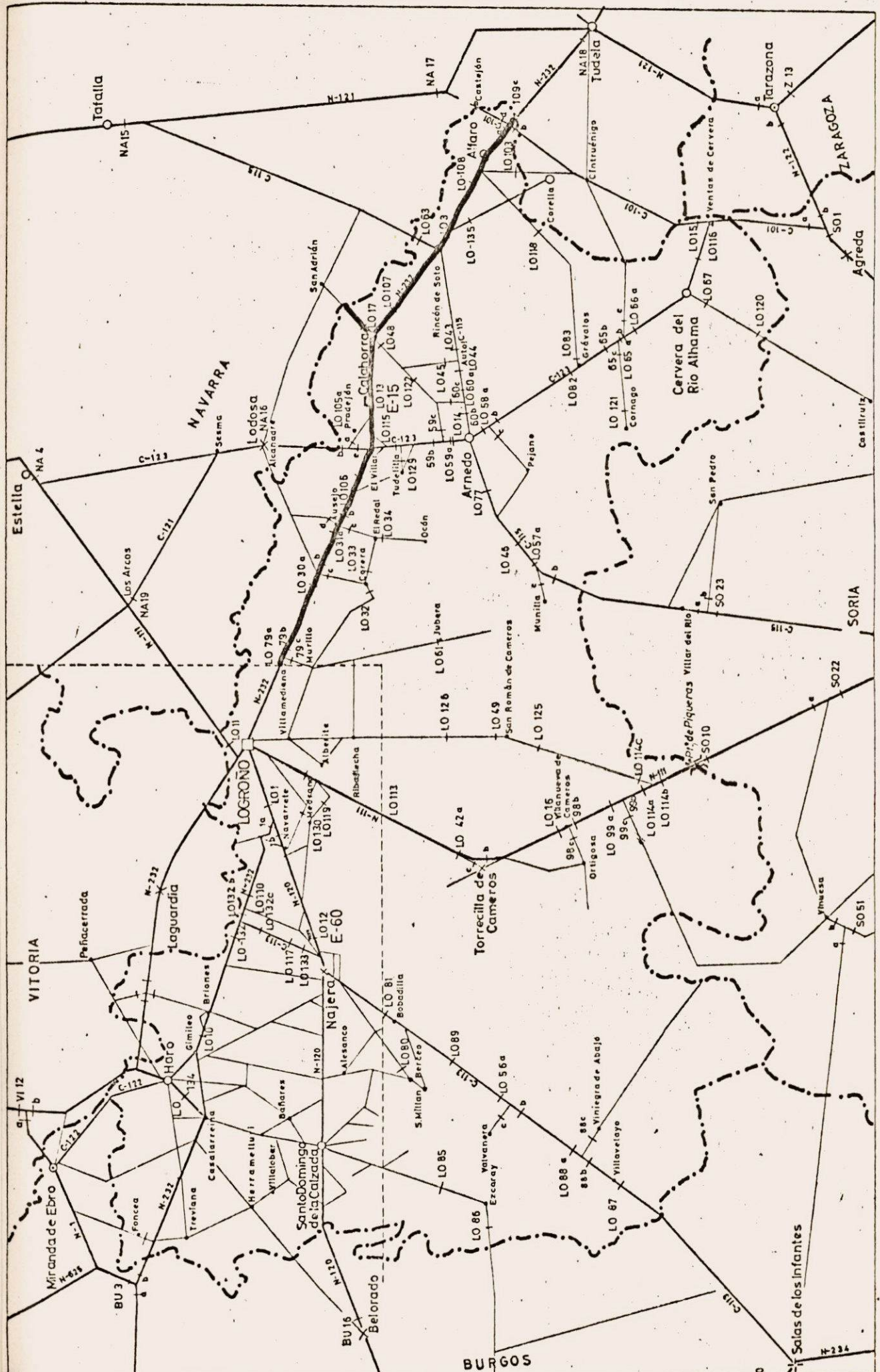
Dentro de este apartado podemos hacer dos grandes grupos:

A.- Flujos por carretera, obtenidos en la dirección provincial de carreteras y que se especifican en los gráficos siguientes de la forma que a continuación se explica:

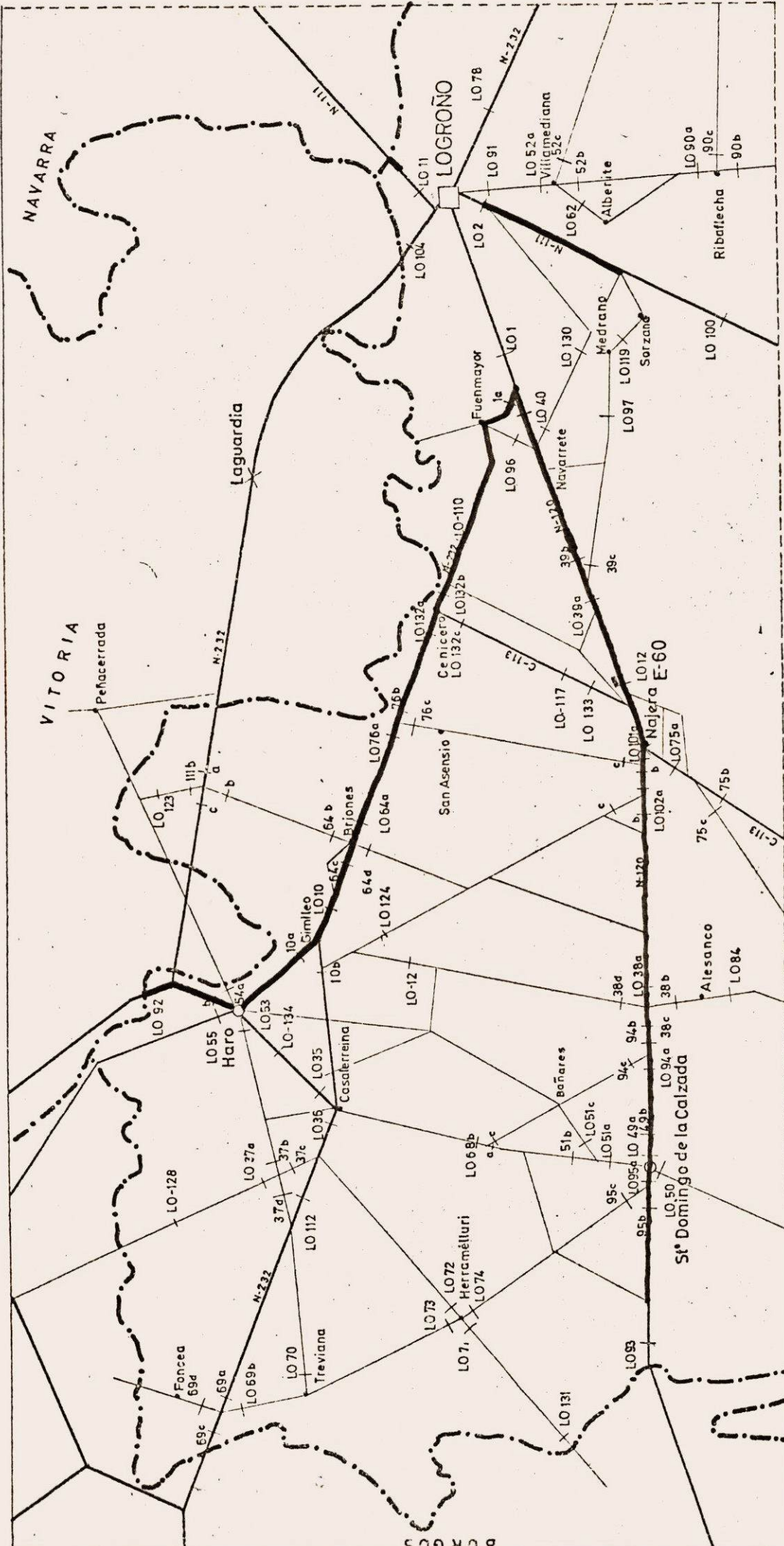
<u>MAPA</u>	<u>I. M. D.</u>
A	———— 5.000 a 9.999
B	———— 2.000 a 4.999
C	———— 1.000 a 1.999
D	———— 500 a 999
E	———— 250 a 499
F	———— 100 a 249
G	———— menos de 100



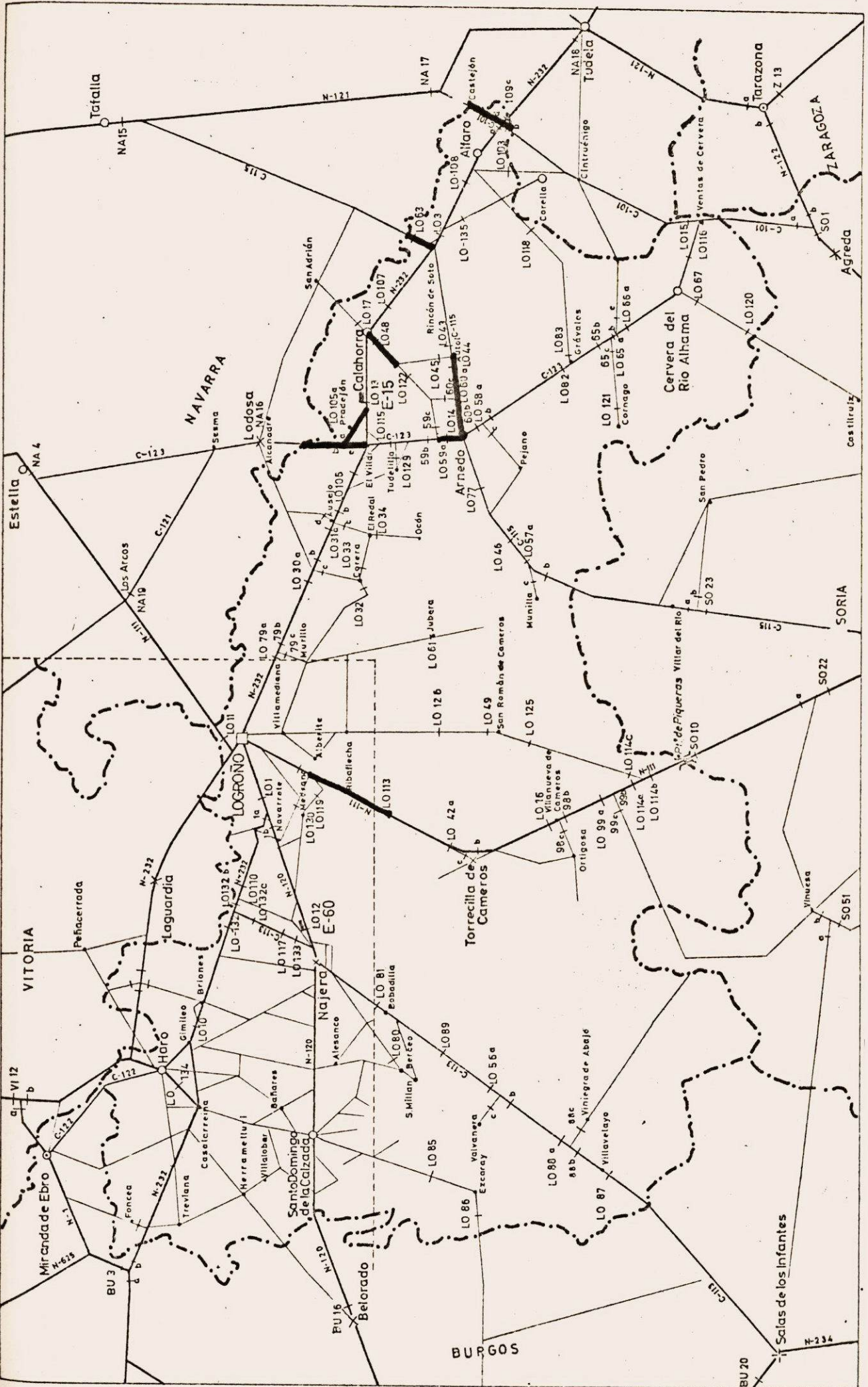
FI



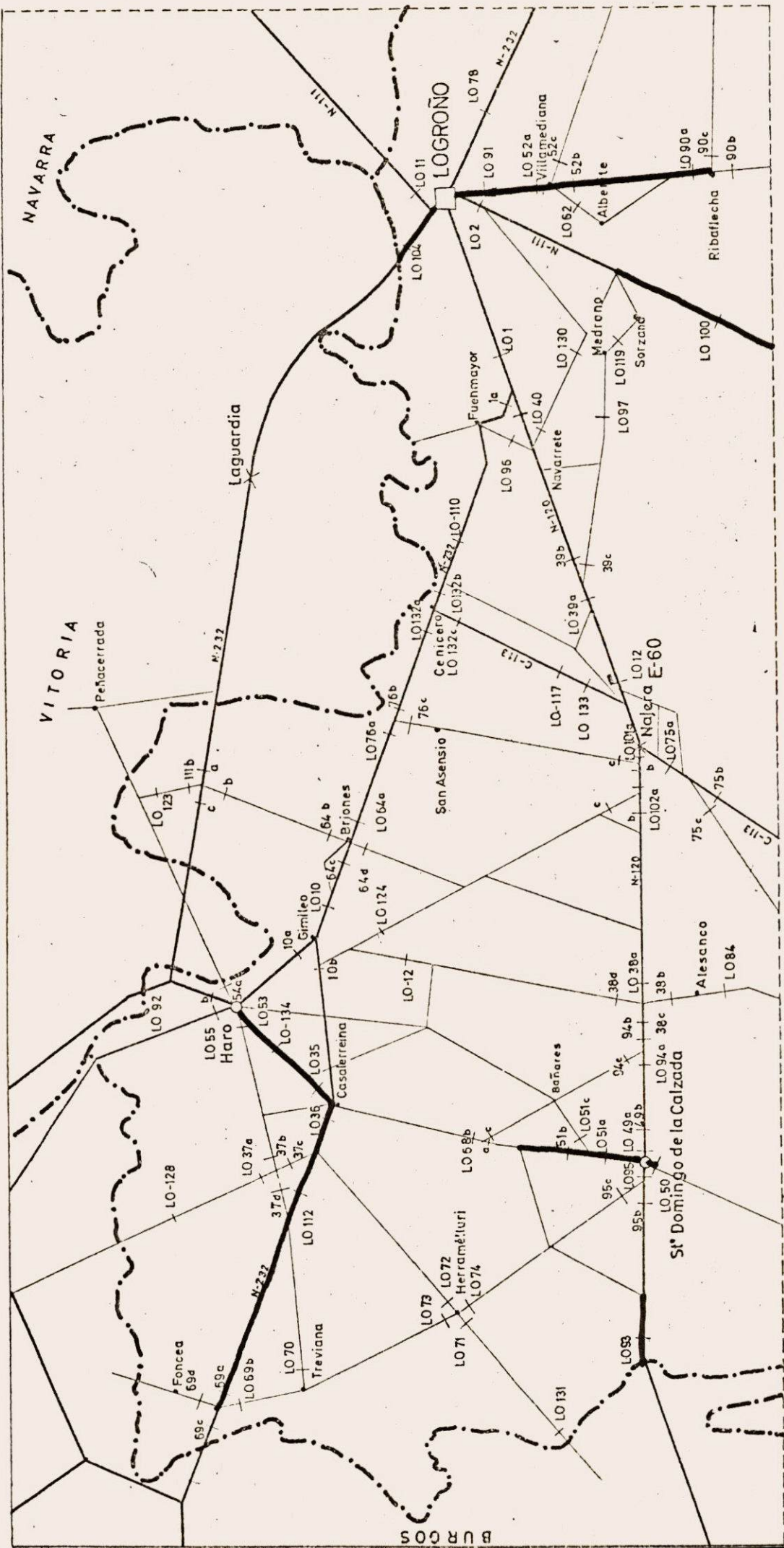
B.



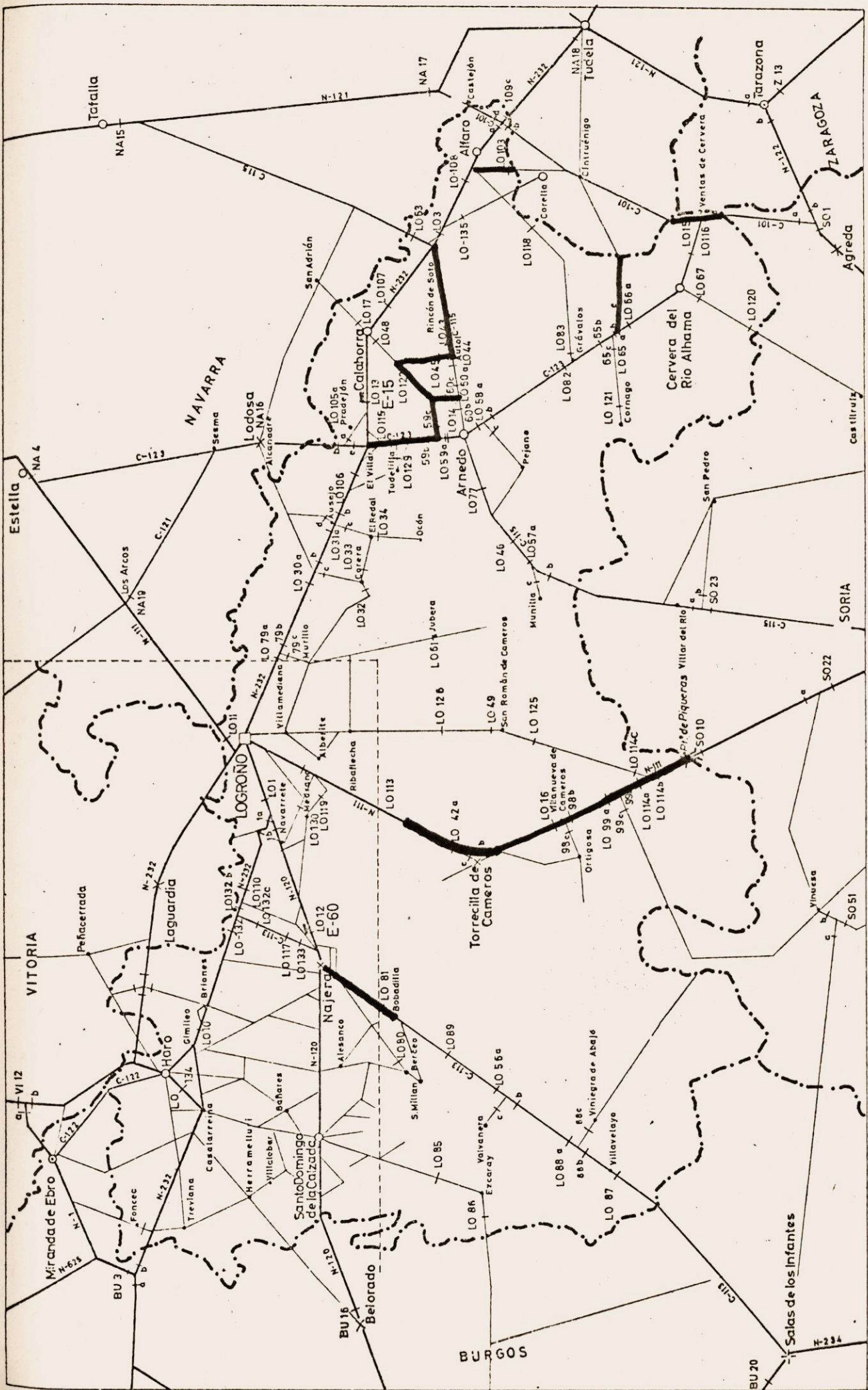
B₂



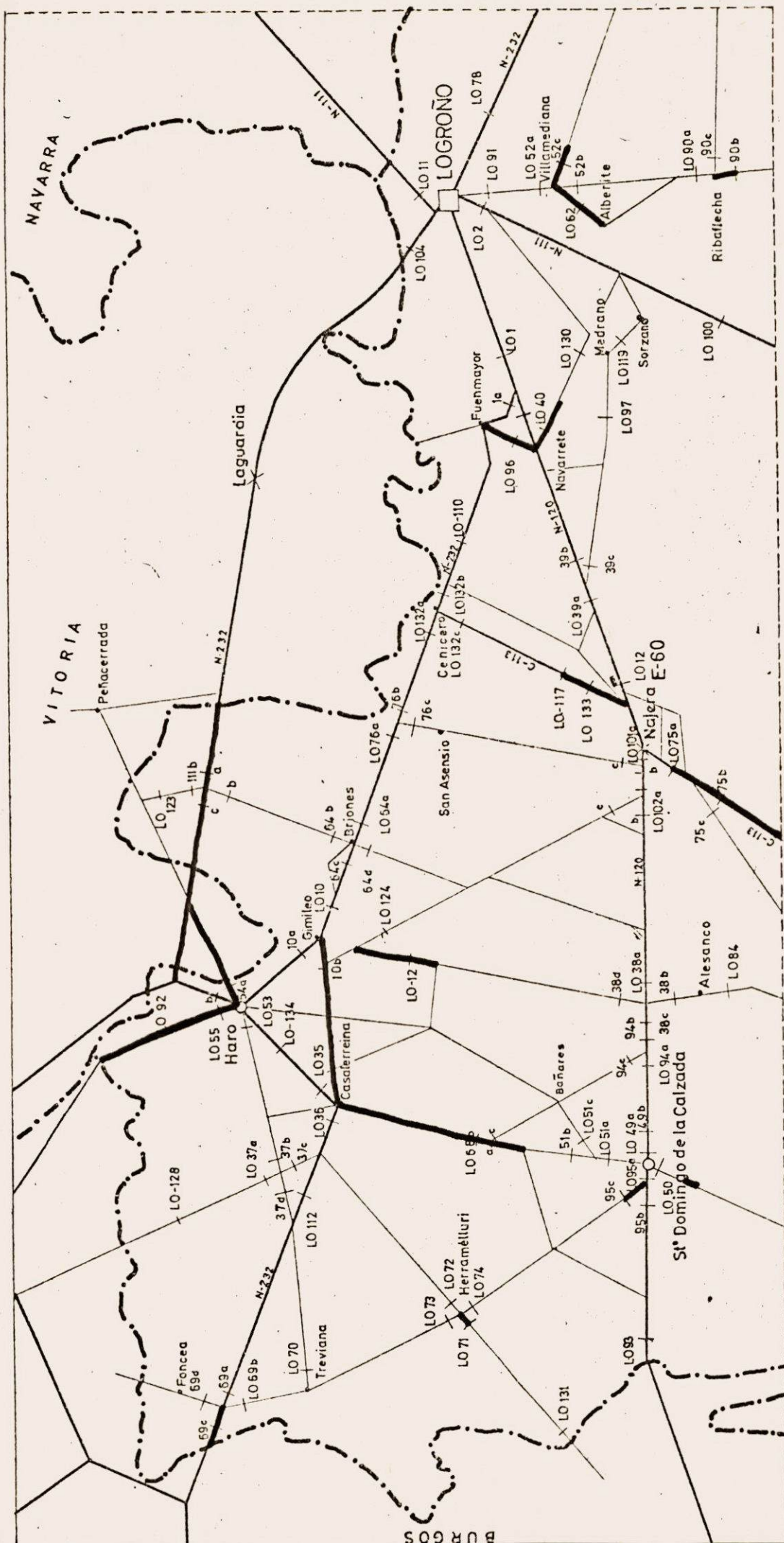
15



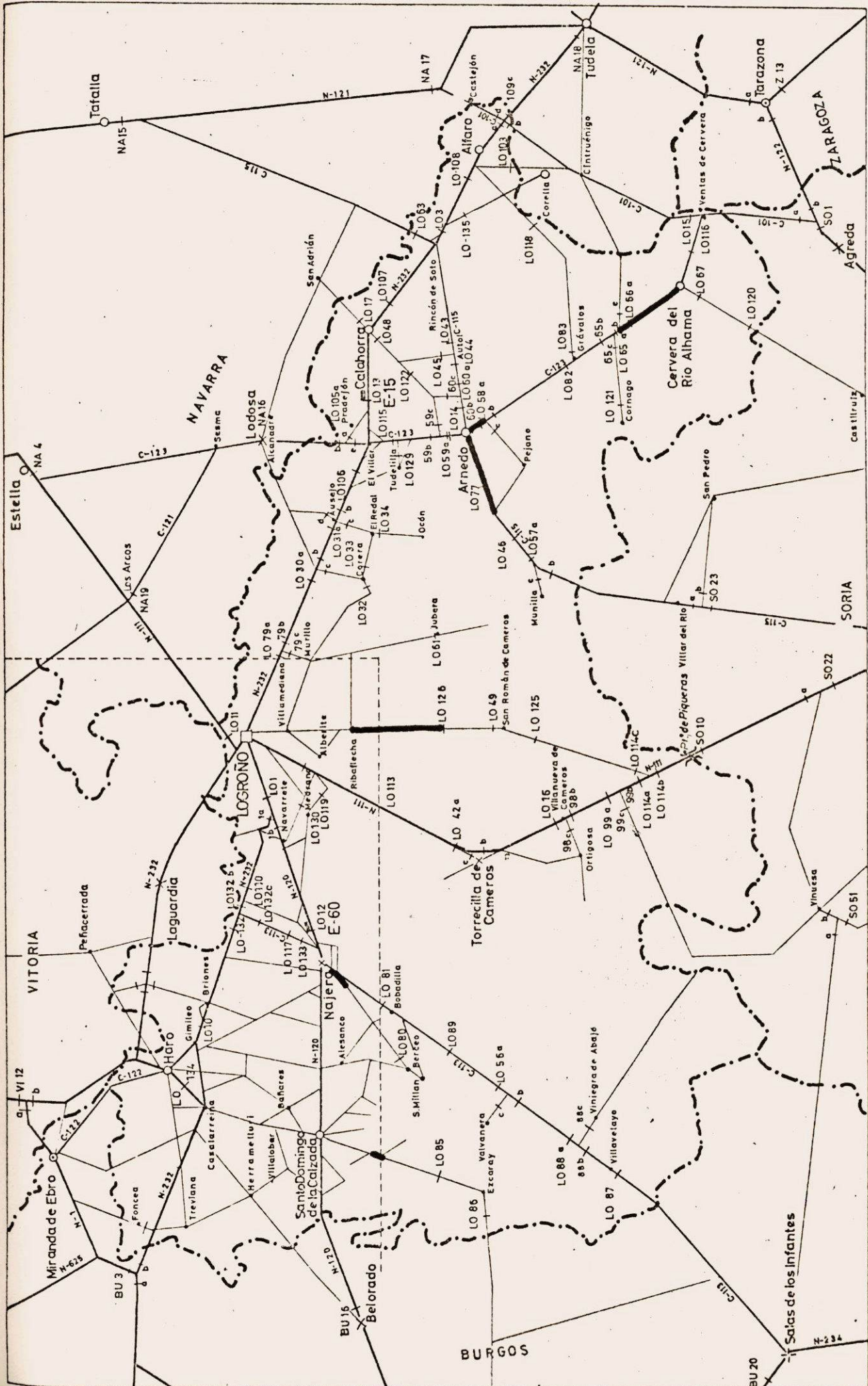
102



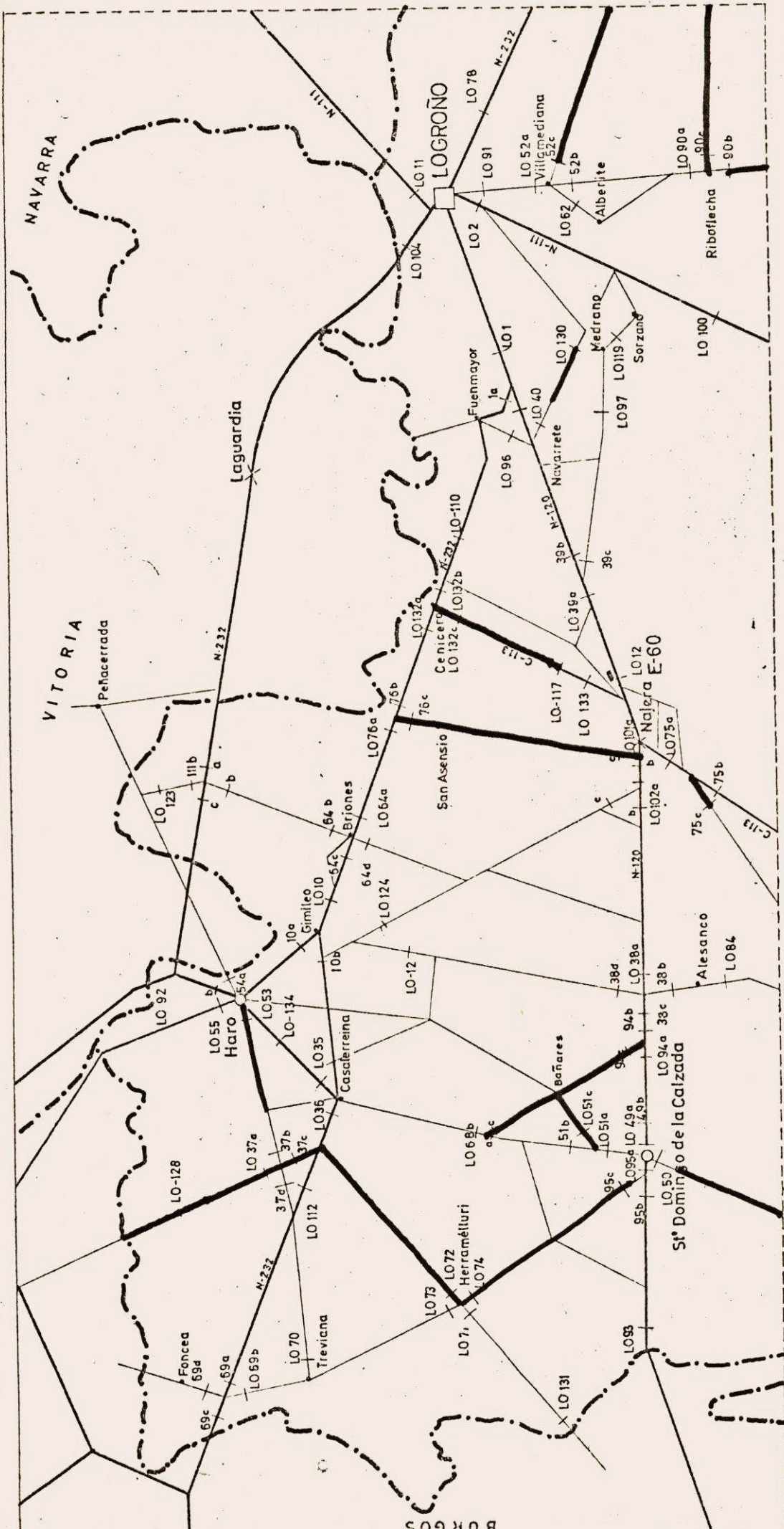
D.



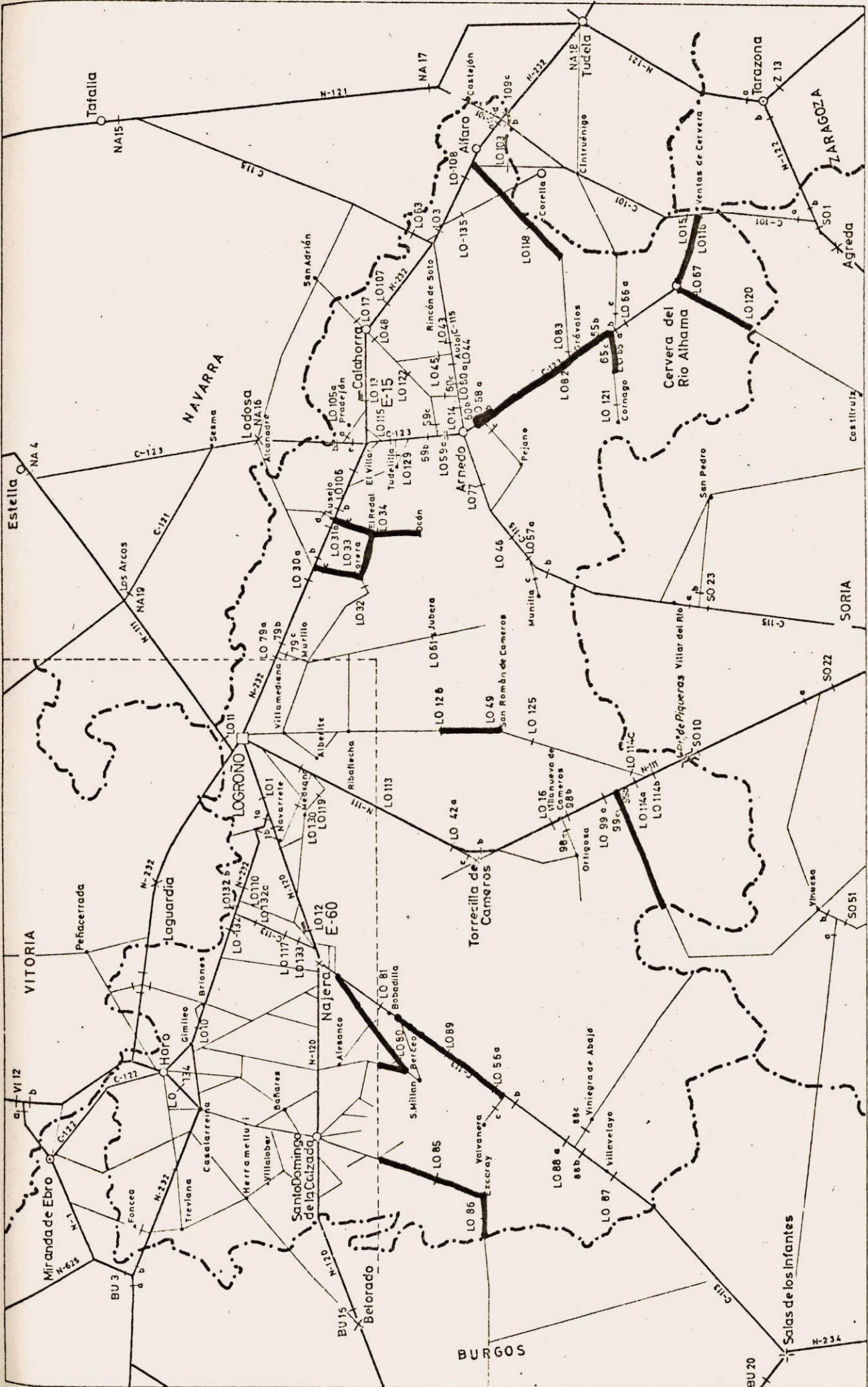
D₂



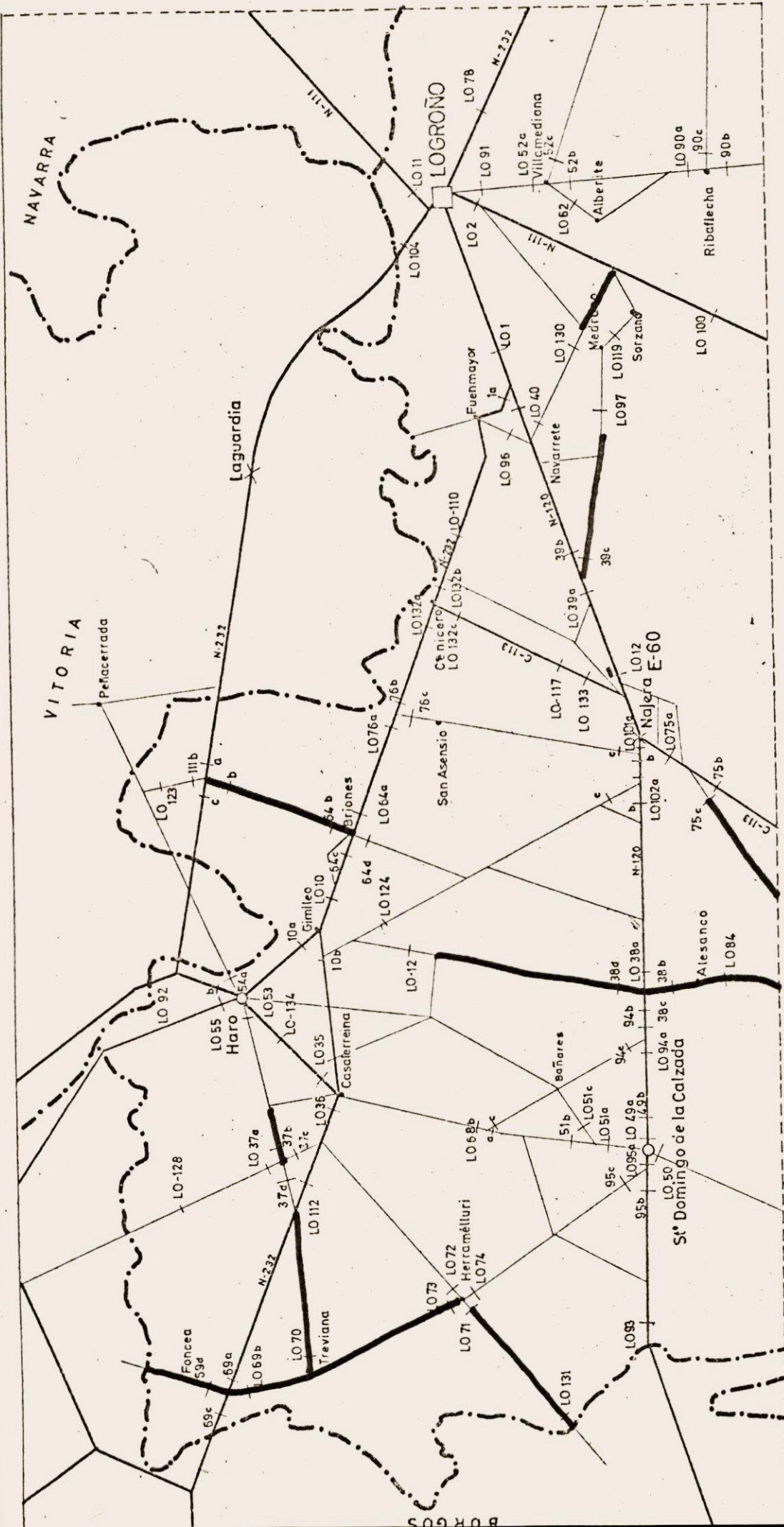
E,



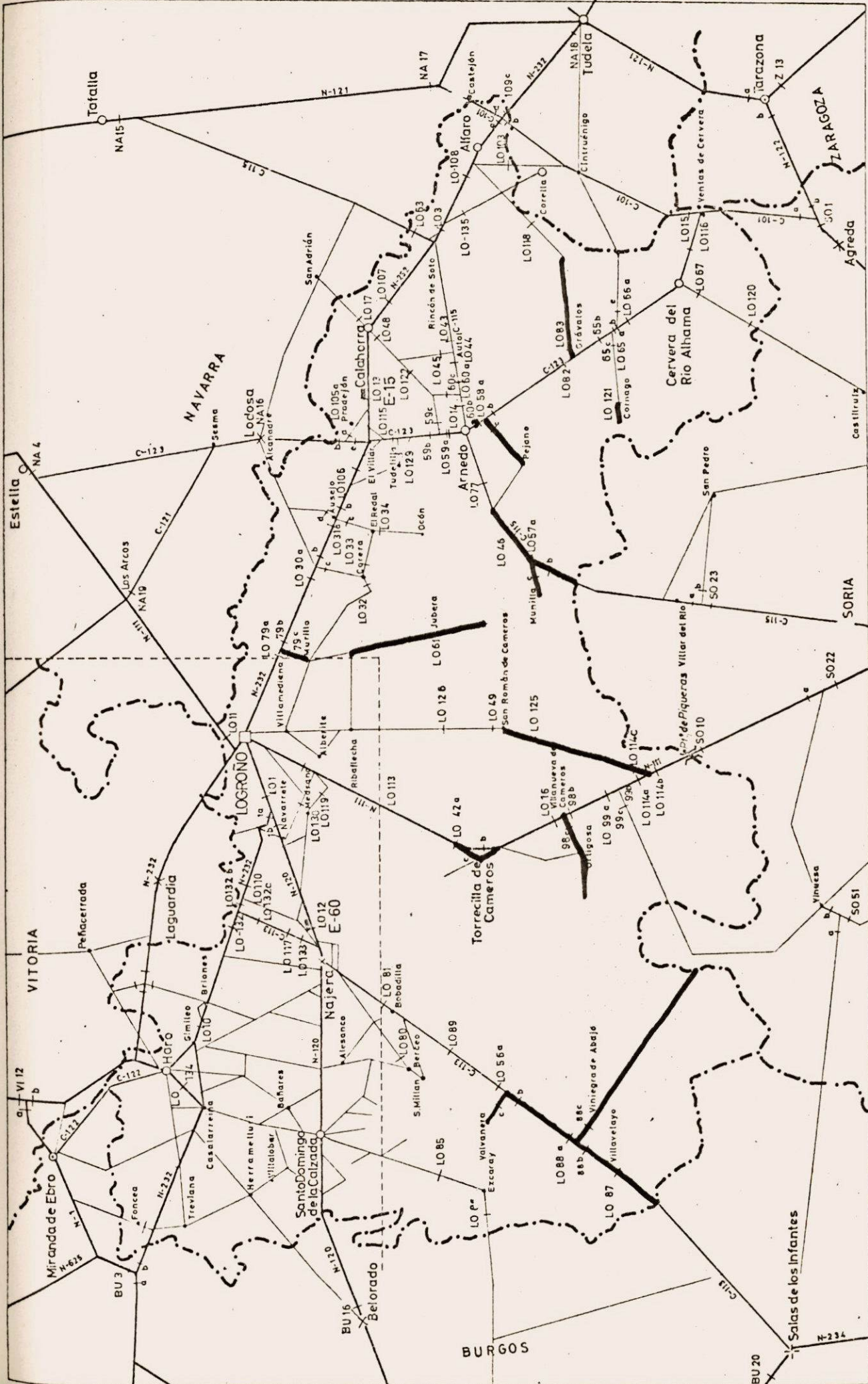
157



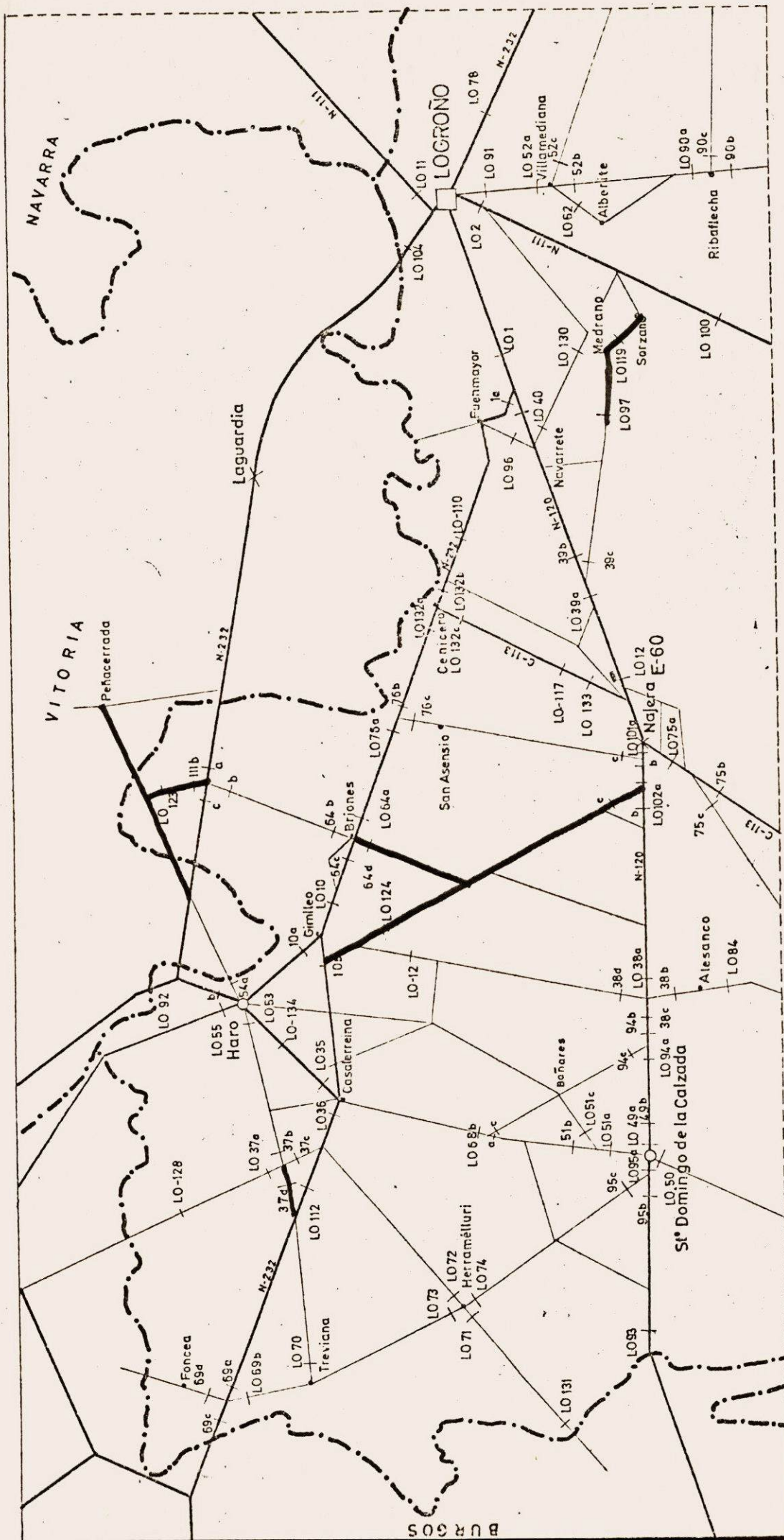
17



F₂



51



62

B.- Flujos de personas desde los municipios carentes de servicios, hacia los que poseen según datos obtenidos por medio de los estudios de la Diputación Provincial.

Para el establecimiento de estos flujos contábamos con la información de los servicios que poseía cada municipio, y en caso de carencia de determinado servicio, sabíamos a donde se dirigían para la satisfacción del mismo.

En cada municipio podíamos conocer cual era el polo o polos de atracción para satisfacer los servicios de los cuales carecía, o bien, si poseía todos o casi todos los servicios, decíamos que no tenía un polo de atracción.

Normalmente un municipio tiene varios polos de atracción, según de qué servicios se trate, y los polos pueden tener mayor o menor entidad, entonces jerarquizamos los polos según la siguiente escala:

Cada municipio tendrá un centro (o varios) de atracción principal, que denominamos centro principal, (a no ser que - él mismo sea el centro).

También puede tener otros centros a donde acuden en - busca de determinados servicios (varios), y si este centro - tiene suficiente entidad (es decir, actúa como centro principal respecto a otros municipios y presta bastantes servicios

al municipio en cuestión), le llamamos centro secundario, si no le llamremos "otros".

En las páginas que siguen, tenemos esa información resumida con la siguiente clave:

- N° del municipio, el oficial de 1 a 180.
- C.P. Centro principal.
- C.S. Centro secundario.
- O. Otros.

1 - ABALOS

C.P. Haro

Logroño

C.S. -

O. Laguardia

2 - AGONCILLO

C.P. Logroño

C.S. -

O. Murillo

3 - AGUILAR DEL RIO ALHAMA

C.P. Logroño

C.S. Tudela

O. Cervera

Calahorra

4 - AJAMIL

C.P. Logroño

C.S. -

O. San Román

5 - ALBELDA

C.P. Logroño

C.S. -

O. -

6 - ALBERITE

C.P. Logroño

C.S. -

O. -

7 - ALCANADRE

C.P. Logroño

C.S. Calahorra

O. Lodosa

8 - ALDEANUEVA DE EBRO

C.P. Calahorra

C.S. Alfaro

O. -

9 - ALESANCO

C.P. Nájera

C.S. Logroño

O. -

10 - ALESON

C.P. Nájera

C.S. Logroño

O. Huércanos

11 - ALFARO

C.P. -

C.S. Calahorra

Logroño

O. -

12 - ALMARZA

C.P. Logroño

C.S. Torrecilla en Cameros

O. -

13 - ANGUCIANA

C.P. Haro
C.S. Logroño
O. -

14 - ANGUIANO

C.P. Logroño
C.S. Nájera
O. Baños de Rio Tobía

15 - ARENZANA DE ABAJO

C.P. Nájera
C.S. Logroño
O. -

16 - ARENZANA DE ARRIBA

C.P. Nájera
Logroño
C.S. -
O. Tricio
Arenzana de Abajo

17 - ARNEDILLO

C.P. Arnedo
C.S. Calahorra
O. -

18 - ARNEDO

C.P. -
C.S. Calahorra
Logroño
O. -

19 - ARRUBAL

C.P. Logroño
C.S. -
O. Murillo

20 - AUSEJO

C.P. Logroño
Calahorra
C.S. -
O. -

21 - AUTOL

C.P. Calahorra
C.S. Arnedo
O. -

22 - AZOFRA

C.P. Nájera
C.S. Logroño
O. Alesanco

23 - BADARAN

C.P. Logroño
Nájera
C.S. -
O. -

24 - BAÑARES

C.P. Santo Domingo
C.S. Logroño
O. Haro

25 - BAÑOS DE RIOJA

C.P. Haro

C.S. -

O. Castañares

31 - BOBADILLA

C.P. Logroño

C.S. Nájera

O. Baños de Rio Tobia

26 - BAÑOS DE RIO TOBIA

C.P. Logroño

C.S. Nájera

O. -

32 - BRIEVA DE CAMEROS

C.P. Logroño

C.S. Nájera

O. -

27 - BERCEO

C.P. Logroño

C.S. -

O. Badaran

33 - BRIÑAS

C.P. Haro

C.S. -

O. -

28 - BERGASA

C.P. Arnedo

C.S. Calahorra

Logroño

O. -

34 - BRIONES

C.P. Haro

C.S. Logroño

O. -

29 - BERGASILLAS BAJERA

C.P. Arnedo

C.S. Calahorra

O. Bergasa

35 - CABEZON DE CAMEROS

C.P. Logroño

C.S. -

O. San Román de Cameros

30 - BEZARES

C.P. Nájera

C.S. Logroño

O. -

36 - CALAHORRA

C.P. -

C.S. -

O. -

37 - CAMPROVIN

C.P. Nájera

C.S. Logroño

O. Baños de Rio Tobia

43 - CASTAÑARES DE RIOJA

C.P. Haro

C.S. Logroño

O. Casalarreina

38 - CANALES DE LA SIERRA

C.P. Logroño

C.S. Nájera

O. -

44 - CASTROVIEJO

C.P. Nájera

Logroño

C.S. -

O. -

39 - CANILLAS DE RIO TUERTO

C.P. Nájera

C.S. Logroño

O. Alesanco

45 - CELORIGO

C.P. Miranda

C.S. Haro

Logroño

O. Sajazarra

40 - CAÑAS

C.P. Logroño

C.S. Nájera

O. Alesanco

46 - CENICERO

C.P. Logroño

C.S. -

O. Nájera

41 - CARDENAS

C.P. Nájera

C.S. Logroño

O. Badarán

47 - CERVERA DEL RIO ALHAMA

C.P. Logroño

C.S. Tudela

Calahorra

O. -

42 - CASALARREINA

C.P. Haro

C.S. Logroño

O. -

48 - CIDAMON

C.P. Haro
C.S. Logroño
O. Bañares

49 - CIHURI

C.P. Haro
C.S. -
O. -

50 - CIRUEÑA

C.P. Santo Domingo
C.S. Logroño
O. -

51 - CLAVIJO

C.P. Logroño
C.S. -
O. Alberite

52 - CORDOVIN

C.P. Nájera
C.S. Logroño
O. -

53 - CORERA

C.P. Logroño
C.S. Arnedo
Calahorra
O. -

54 - CORNAGO

C.P. Arnedo
C.S. Calahorra
Logroño
Cervera
O. -

55 - CORPORALES

C.P. Santo Domingo
C.S. Logroño
O. Grañón

56 - CUZCURRITA DE RIO TIRON

C.P. Haro
C.S. Logroño
O. Santo Domingo

57 - DAROCA DE RIOJA

C.P. Logroño
C.S. -
O. Medrano

58 - ENCISO

C.P. Arnedo
C.S. Logroño
O. Calahorra

59 - ENTRENA

C.P. Logroño
C.S. -
O. -

60 - ESTOLLO

C.P. Nájera
Logroño
C.S. -
O. Santo Domingo
Badarán

61 - EZCARAY

C.P. Santo Domingo
Logroño
C.S. -
O. -

62 - FONCEA

C.P. Haro
C.S. Miranda
Logroño
O. -

63 - FONCALECHE

C.P. Haro
C.S. -
O. Logroño

64 - FUENMAYOR

C.P. Logroño
C.S. -
O. -

65 - GALBARRULI

C.P. Miranda
C.S. Haro
O. Sajazarra
Logroño

66 - GALILEA

C.P. Logroño
C.S. Calahorra
O. -

67 - GALLINERO DE CAMEROS

C.P. Logroño
C.S. Villanueva
Ortigosa
O. -

68 - GIMILEO

C.P. Haro
C.S. Logroño
O. -

69 - GRAÑON

C.P. Santo Domingo
C.S. Logroño
O. -

70 - GRAVALOS

C.P. Arnedo
C.S. Logroño
O. Alfaro
Calahorra

71 - HARO

C.P. -
C.S. -
O. -

72 - HERCE

C.P. Arnedo
C.S. Logroño
Calahorra
O. -

73 - HERRANELLURI

C.P. Santo Domingo
Logroño
C.S. Haro
O. -

74 - HERVIAS

C.P. Santo Domingo
C.S. Logroño
O. Haro

75 - HORMILLA

C.P. Nájera
C.S. Logroño
O. -

76 - HORMILLEJA

C.P. Nájera
C.S. Logroño
O. Haro

77 - HORNILLOS DE CAMEROS

C.P. Logroño
C.S. -
O. San Román

78 - HORNOS DE MONCALVO

C.P. Logroño
C.S. -
O. Navarrete

79 - HUERCANOS

C.P. Logroño
Nájera
C.S. -
O. -

80 - IGEA

C.P. Logroño
Calahorra
C.S. -
O. -

81 - JALON DE CAMEROS

C.P. Logroño
C.S. -
O. San Román

82 - LAGUNA DE CAMEROS

C.P. Logroño
C.S. -
O. -

83 - LAGUNILLA DE JEBERA

C.P. Logroño
C.S. -
O. Murillo

84 - LARDERO

C.P. Logroño

C.S. -

O. -

85 - LEDESMA

C.P. Logroño

Nájera

C.S. -

O. Baños de Rio Tobia

86 - LEIVA

C.P. Santo Domingo

C.S. Logroño

O. Haro

87 - LEZA DE RIO LEZA

C.P. Logroño

C.S. -

O. Ribafrecha

88 - LOGROÑO

C.P. -

C.S. -

O. -

89 - LUMBRERAS

C.P. Logroño

C.S. -

O. -

90 - MANJARRES

C.P. Nájera

C.S. Logroño

O. Tricio

91 - MANSILLA

C.P. Logroño

Nájera

C.S. -

O. -

92 - MANZANARES DE RIOJA

C.P. Santo Domingo

C.S. Logroño

O. -

93 - MATUTE

C.P. Logroño

Nájera

C.S. -

O. -

94 - MEDRANO

C.P. Logroño

C.S. -

O. -

95 - MONTALVO DE CAMEROS (Anexionado)

96 - MUNILLA

C.P. Arnedo

C.S. Logroño

O. Calahorra

97 - MURILLO DE RIO LEZA

C.P. Logroño

C.S. -

O. -

98 - MURO DE AGUAS

C.P. Arnedo

C.S. -

O. Cornago

Logroño

99 - MURO DE CAMEROS

C.P. Logroño

C.S. -

O. San Román

100 - NAJERA

C.P. -

C.S. -

O. -

101 - NALDA

C.P. Logroño

C.S. -

O. -

102 - NAVAJUN

C.P. Aguilar

Logroño

C.S. Cervera

Alfaro

O. -

103 - NAVARRETE

C.P. Logroño

C.S. -

O. Nájera

104 - NESTARES

C.P. Logroño

C.S. Torrecilla en Cameros

O. -

105 - NIEVA

C.P. Logroño

C.S. -

O. Torrecilla en Cameros

106 - OCON

C.P. Logroño

C.S. -

O. Calahorra

Arnedo

107 - OCHANDURI

C.P. Haro

C.S. Santo Domingo

Logroño

O. Casalarreina

108 - OJACASTRO

C.P. Santo Domingo

C.S. -

O. Ezcaray

Logroño

109 - OLLAURI

C.P. Haro

C.S. Logroño

O. -

110 - ORTIGOSA

C.P. Logroño

C.S. -

O. -

111 - PAZUENGOS

C.P. Santo Domingo

C.S. Logroño

O. Haro

112 - PEDROSO

C.P. Logroño

C.S. Nájera

O. Baños de Rio Tobia

113 - PINILLOS

C.P. Logroño

C.S. Torrecilla en Cameros

O. -

- 114 - POYALES (Anexionado) 122 - RIBAFRECHA
C.P. Logroño
C.S. -
O. -
- 115 - PRADEJON
C.P. Calahorra
C.S. Logroño
O. -
- 116 - PRADILLO
C.P. Logroño
C.S. Torrecilla en Cameros
O. Ortigosa
Villanueva
- 117 - PREJANO
C.P. Arnedo
C.S. -
O. Logroño
- 118 - QUEL
C.P. Calahorra
Logroño
Arnedo
C.S. -
O. -
- 119 - RABANERA
C.P. Logroño
C.S. -
O. San Román
- 120 - RASILLO (EL)
C.P. Logroño
C.S. -
O. Ortigosa
- 121 - REDAL (EL)
C.P. Logroño
C.S. -
O. Arnedo
Calahorra
Corera
- 123 - RINCON DE SOTO
C.P. Logroño
Calahorra
C.S. Alfaro
O. -
- 124 - ROBRES DEL CASTILLO
C.P. Logroño
C.S. -
O. -
- 125 - RODEZNO
C.P. Haro
C.S. Logroño
O. -
- 126 - SAJAZARRA
C.P. Haro
C.S. Miranda
O. Logroño
- 127 - SAN ASENSIO
C.P. Logroño
C.S. Haro
O. Nájera
- 128 - SAN MILLAN DE LA COGOLLA
C.P. Logroño
C.S. Nájera
O. -

- 129 - SAN MILLAN DE YECORA
 C.P. Haro
 C.S. Logroño
 O. Treviana
- 130 - SAN ROMAN DE CAMEROS
 C.P. Logroño
 C.S. -
 O. Laguna
- 131 - SANTA (LA) (Anexionado)
- 132 - SANTA COLOMA
 C.P. Nájera
 Logroño
 C.S. -
 O. Hornilla
 Manjarres
- 133 - SANTA ENGRACIA DE JUBERA
 C.P. Logroño
 C.S. -
 O. Murillo
- 134 - SANTA EULALIA BAJERA
 C.P. Arnedo
 C.S. Logroño
 Calahorra
 O. -
- 135 - SANTO DOMINGO DE LA CALZ.
 C.P. -
 C.S. Logroño
 Haro
 O. -
- 136 - SAN TORCUATO
 C.P. Santo Domingo
 Haro
 C.S. Logroño
 O. -
- 137 - SANTURDE
 C.P. Santo Domingo
 C.S. Nájera
 Haro
 Logroño
 O. -
- 138 - SANTURDEJO
 C.P. Santo Domingo
 C.S. Haro
 Logroño
 O. Ezcaray
- 139 - SAN VICENTE DE LA SONSIERRA
 C.P. Haro
 Logroño
 C.S. -
 O. Laguardia
 Briones
- 140 - SOJUELA
 C.P. Logroño
 C.S. -
 O. Entrena
 Medrano
 Nalda
- 141 - SORZANO
 C.P. Logroño
 C.S. -
 O. Nalda

142 - SOTES

C.P. Logroño
C.S. -
O. Navarrete

143 - SOTO EN CAMEROS

C.P. Logroño
C.S. -
O. Ribafrecha

144 - TERROBA

C.P. Logroño
C.S. -
O. San Román
Ortigosa
Laguna

145 - TIRGO

C.P. Haro
C.S. Logroño
O. Casañarreina
Cuzcurrita
Miranda

146 - TOBIA

C.P. Logroño
Nájera
C.S. -
O. Matute
Ortigosa

147 - TORMANTOS

C.P. Santo Domingo
C.S. Logroño
Haro
O. Leiva
Grañón

148 - TORRECILLA EN CAMEROS

C.P. Logroño
C.S. -
O. -

149 - TORRECILLA S/ ALESANCO

C.P. Nájera
C.S. Logroño
O. Alesanco

150 - TORRE EN CAMEROS

C.P. Logroño
C.S. -
O. San Román
Ortigosa
Muro
Jalón
Laguna

151 - TORREMONTALVO

C.P. Logroño
C.S. Cenicero
O. Nájera

152 - TREVIANA

C.P. Haro
C.S. Santo Domingo
Logroño
O. Casalarreina

153 - TREVIJANO

C.P. Logroño
C.S. -
O. Ribafrecha
San Román
Ortigosa

- 154 - TRICIO
 C.P. Nájera
 C.S. ^{LOGROÑO} -
 O. -
- 155 - TUDELILLA
 C.P. Logroño
 Arnedo
 Calahorra
 C.S. -
 O. -
- 156 - TURRUNCUN (Anexionado)
- 157 - URUÑUELA
 C.P. Nájera
 Logroño
 C.S. -
 O. -
- 158 - VALDEMADERA
 C.P. Aguilar
 Cervera
 C.S. Alfaro
 Calahorra
 Logroño
 O. Tudela
- 159 - VALGAÑÓN
 C.P. Santo Domingo
 Ezcaray
 Logroño
 C.S. Haro
 O. -
- 160 - VENTOSA
 C.P. Logroño
 C.S. -
 O. Sotés
 Manjarres
 Navarrete
- 161 - VENTROSA
 C.P. Logroño
 Nájera
 C.S. -
 O. Viniegra de Abajo
 Villavelayo
 Ortigosa
- 162 - VIGUERA
 C.P. Logroño
 C.S. -
 O. Sorzano
 Nalda
- 163 - VILLALBA DE RIOJA
 C.P. Haro
 C.S. -
 O. -
- 164 - VILLALOBAR DE RIOJA
 C.P. Haro
 Santo Domingo
 C.S. Logroño
 O. Grañón
- 165 - VILLAMEDIANA
 C.P. Logroño
 C.S. --
 O. -
- 166 - VILLANUEVA DE CAMEROS
 C.P. Logroño
 C.S. Torrecilla en Cameros
 O. Ortigosa
- 167 - VILLAR DE ARNEDO (EL)
 C.P. Calahorra
 Logroño
 Arnedo
 C.S. -
 O. Pradeión

168 - VILLAR DE TORRE

C.P. Santo Domingo
Nájera
Logroño
Haro

C.S. -

O. Alesanco

169 - VILLAREJO

C.P. Santo Domingo
Logroño

C.S. Nájera
Haro

O. Villar de Torre

170 - VILLARROYA

C.P. Arnedo
Logroño

C.S. Calahorra
O. Cornago
Grávalos

171 - VILLARTA QUINTANA

C.P. Santo Domingo
C.S. Logroño
Haro

O. Grañón

172 - VILLAVELAYO

C.P. Logroño

C.S. -

O. -

173 - VILLAVERDE DE RIOJA

C.P. Nájera
C.S. Logroño
O. Matute
Baños

174 - VILLOSLADA

C.P. Logroño
C.S. -
O. -

175 - VINIEGRA DE ABAJO

C.P. Logroño
Nájera

C.S. -

O. Anguiano
Baños

176 - VINIEGRA DE ARRIBA

C.P. Logroño
Nájera

C.S. -

O. Viniegra de Abajo
Ortigosa
Villavelayo

177 - ZARRATON

C.P. Haro
C.S. Logroño
O. Bañares

178 - ZARZOSA

C.P. Arnedo
C.S. Logroño
Calahorra

O. Arnedillo
Herce
Enciso

179 - ZENZANO (Anexionado)

180 - ZORRAQUIN

C.P. Ezcaray
Santo Domingo
C.S. Haro
Logroño

O. -

V.- COMARCALIZACION BASADA EN EL ANALISIS MULTIVARIANTE.

Lo que vamos a intentar a partir de este momento es la realización de la comarcalización homogénea y su posterior - aproximación a la nodalidad. Para ello utilizaremos siempre que nos sea posible el análisis multivariante, aunque lo completaremos con los datos de flujos que poseemos y que no son susceptibles, dado lo escaso de su cuantitativización, del - tratamiento citado, tal y como estudiaremos en su momento.

La comarcalización homogénea se va a basar en los datos agrícolas (1) (entendemos aquí agrícolas y ganaderos), - que mediante el análisis factorial (entendido éste tal y como explicamos en la segunda parte de este trabajo), reduciremos a factores, con los cuales realizaremos un cluster análisis que contrastaremos con la situación geográfica al objeto de ver si es coherente, como ya comentaremos.

Para el análisis factorial mantendremos los programas utilizados en los últimos estudios de la segunda parte, que ya vimos que eran los más interesantes.

Pasamos por lo tanto, al proceso de análisis multiva-
riante.

(1) Que son los que poseemos, de todas formas comentaremos al final de este trabajo la coherencia y posible ampliación del estudio así realizado.

44	18
45	19
46	20
47	21
42/24	22

Podemos observar que ~~no se~~ han introducido las variables de ganadería per cápita, en sus distintas clases y solo el total per cápita. Ya estudiaremos las consecuencias de esto y veremos las razones de su no inclusión.

El ordenador rechazará los factores que expliquen menos del 5% antes de la rotación.

El código a emplear será:

- Variables muy importantes: Coeficiente mayor del 0'9.
- Variables importantes: Coeficiente mayor del 0'75.
- Variables secundarias: Coeficiente mayor del 0'50.
- Variables terciarias: Coeficiente mayor del 0'30.

Podemos observar que hemos cambiado la escala respecto a la segunda parte. Esto es perfectamente lógico pues al tratarse de datos reales, las correlaciones alcanzadas son mucho mayores que en una encuesta de opinión.

En este primer análisis factorial, con cinco factores,

hemos logrado explicar el 63'624% de la varianza total, dato que hemos de calificar de bajo y que trataremos de arreglar en posteriores estudios.

El lector ha de tener en cuenta que aunque este porcentaje de explicación lo calificaríamos de alto en la segunda parte de este trabajo (estudio de las encuestas), hemos de calificarlo de bajo en esta tercera parte, pues los datos son muy diferentes. Antes teníamos como material de trabajo unas encuestas donde la puntuación es muy subjetiva y por lo tanto expuesta a abundantes errores. Por el contrario en esta tercera parte, nuestra base son datos objetivos y, por lo tanto, menos expuestos al error.

Esta disminución del error es la que se debe reflejar en un aumento del porcentaje explicativo del estudio.

Pasemos, sin más preámbulos, a la interpretación de la matriz factorial antes de la rotación. Los datos de la salida de ordenador están en las hojas de ordenador nº 940 y ss. bajo el nombre de Análisis Factorial Agrico.

Factor 1

Valor propio: 6'06605

Porcentaje de explicación: 27'57%

Variables muy importantes:

28 - Cultivos.

Variables importantes:

- 26 - Altura (negativa).
- 32 - Cereales.

Variables secundarias:

- 27 - Regadío.
- 29 - Prados (negativa).
- 30 - Forestal (negativa).
- 31 - Otras superficies (negativa).
- 34 - Patata.
- 35 - Industriales (se entiende cultivos industriales).
- 40 - Viñedo.
- 42/24 - Capital ganadero per cápita (negativa).

Variables terciarias:

- 37 - Hortalizas.
- 43 - Bovino (negativa).
- 45 - Caprino (negativa).
- 46 - Cerda.
- 47 - Trabajo (se entiende ganado de trabajo).

Este factor, que es el más importante antes de la rotación, es el que más discrimina al colectivo y su interpretación, aunque no del todo clara, tampoco parece muy difícil.

Creo que es interesante antes de pasar a esta interpretación, y de forma que sirva como aclaración para las interpretaciones subsiguientes, que dado que no soy un experto en

temas agrícolas, ciertas explicaciones pueden resultar erróneas al menos en parte, pero he procurado asesorarme en lo posible por personas competentes en la materia para reducir estos errores al mínimo.

Pasando a la interpretación, podemos decir que este factor hace especial referencia a los cultivos y dentro de estos los cereales. También está relacionado negativamente con la altura y, en consecuencia, con las formas de explotación agrícola y ganadera relacionadas con ella, como son en ganadería el ganado Bovino y Caprino y en agricultura - las explotaciones forestales, etc.

Por el contrario, se relaciona positivamente con las explotaciones agrícolas típicas del llano e incluso de la vega, como son las hortalizas, etc.

En conclusión es un factor agrícola del llano fundamentalmente.

Factor 2

Valor propio: 2'82271

Porcentaje de explicación: 12'83%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

34 - Patata (negativa).

- 39 - Almendro.
- 41 - Olivar.
- 47 - Trabajo.

Variables terciarias:

- 32 - Cereales (negativa).
- 35 - Industriales (negativa).
- 37 - Hortalizas.
- 38 - Frutales.
- 43 - Bovino (negativa).
- 44 - Ovino.
- 45 - Caprino.
- 46 - Cerda (negativa).

Este factor es difícil de interpretar, sobre todo en lo referente al ganado, por las razones que comentaremos al final.

De todas formas, centrándonos en el apartado agrícola vemos una preponderancia del árbol (almendro, olivo, frutales) en detrimento del cereal, patata y cultivos industriales. - También hay una relación con las hortalizas, pero hemos de concluir que el factor está mal definido.

Factor 3

Valor propio: 2'3453

Porcentaje de explicación: 10'66%

VARIABLES MUY IMPORTANTES: (CARECE)

VARIABLES IMPORTANTES: (CARECE)

VARIABLES SECUNDARIAS:

- 27 - Regadío.
- 36 - Forrajeras.
- 43 - Bovino.
- 44 - Ovino (negativa).

VARIABLES TERCIARIAS:

- 33 - Legumbres.
- 37 - Hortalizas.
- 38 - Frutales.
- 47 - Trabajo (negativa).

Es este un factor que resulta más claro que el anterior. Vemos que está asociado con cultivos forrajeros y regadío, también tenemos lógicamente legumbres, hortalizas y frutales.

En cuanto a la ganadería parece lógico que el ganado bovino esté asociado a los cultivos forrajeros. Tengamos en cuenta que este ganado puede estar asociado, bien a prados y pastizales si vive en libertad o a cultivos forrajeros en el caso de que se encuentre estabulado.

También es normal la carencia de ganado ovino y de trabajo en este tipo de tierras. Respecto a este último tenga-

mos en cuenta que en las tierras fértiles los animales de trabajo han sido ya sustituidos casi completamente por el tractor.

Factor 4

Valor propio: 1'48092

Porcentaje de explicación: 6'73%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

29 - Prados.

40 - Viñedo.

42 - Trabajo.

42/24 - Capital ganadero per cápita.

Variables terciarias:

46 - Cerda (negativa).

Este factor es nuevamente de difícil interpretación, podemos resumirlo con las variables que lo definen, tal como las hemos expuesto y esperemos que se aclare tras la rotación.

Factor 5

Valor propio: 1'28235

Porcentaje de explicación: 5'83%

VARIABLES MUY IMPORTANTES: (Carece)

VARIABLES IMPORTANTES: (Carece)

VARIABLES SECUNDARIAS: (Carece)

VARIABLES TERCIARIAS:

- 29 - Prados (negativa).
- 31 - Otras superficies.
- 40 - Viñedo.
- 44 - Ovino (negativa).
- 45 - Caprino (negativa).
- 42/24 - Capital ganadero per cápita (negativa).

Este factor es bastante claro a pesar de lo poco importantes que son las variables que lo definen (solo terciarias).

Vemos que está relacionado con la escasez de ganado y más concretamente con la falta de ovino y caprino, también faltan los prados. Sin embargo hay viñedo y otras superficies.

A la vista de los presentes resultados, no excesivamente alentadores, pasamos a la rotación en espera de que la consistencia de los factores aumente en lo que se refiere a su interpretación.

Factores rotados:

Factor 1

Capacidad explicativa: 4'6486

Porcentaje de explicación: 21'13%

Variables muy importantes: (Carece)

Variables importantes:

- 28 - Cultivos.
- 31 - Otras superficies (negativa).
- 32 - Cereales.

Variables secundarias:

- 26 - Altura (negativa).
- 30 - Forestal (negativa).
- 34 - Patata.
- 35 - Industriales.
- 46 - Cerda.

Variables terciarias:

- 27 - Regadío.
- 40 - Viñedo.
- 45 - Caprino (negativa).

Vemos que este factor, que sigue siendo el que más ex
plica tras la rotación, es bastante similar a como era antes
de rotar.

Fundamentalmente hace referencia a los cultivos, y denen
tro de estos a cereales, patata, etc. En lo referente a ganaa

dería, parece que no destaca, salvándose el ganado de cerda como más abundante.

Factor 2

Capacidad explicativa: 2'112

Porcentaje de explicación: 9'6%

Variables muy importantes: (Carece)

Variables importantes:

43 - Bovino (negativa).

44 - Ovino.

Variables secundarias:

45 - Caprino.

Variables terciarias:

33 - Leguminosas (negativa).

39 - Almendro.

Es este el factor menos explicativo, tras la rotación. Parece un factor indudablemente ganadero, haciendo especial referencia a ovino y caprino y ausencia de bovino.

Factor 3

Capacidad explicativa: 2'7192

Porcentaje de explicación: 12'36%

Variables muy importantes: (Carece)

Variables importantes:

37 - Hortalizas.

Variables secundarias:

27 - Regadío.

36 - Forrajeras.

38 - Frutales.

Variables terciarias:

26 - Altura (negativa).

39 - Almendro.

41 - Olivar.

Vemos que este es un factor de tipo agrícola bastante claro. La variable fundamental son las hortalizas, acompañada de regadío, forrajeros y frutales. Que son los cultivos - típicos de la vega.

Factor 4

Capacidad explicativa: 2'3518

Porcentaje de explicación: 10'69%

Variables muy importantes: (Carece)

Variables importantes:

40 - Viñedo.

47 - Trabajo.

Variables secundarias: (Carece)

Variables terciarias:

- 26 - Altura (negativa).
- 30 - Forestal (negativa).
- 39 - Almendro.
- 41 - Olivar.
- 43 - Bovino (negativa).

Es este un factor que se asocia fundamentalmente al viñedo y a los animales de trabajo que van correlacionados.

Factor 5

Capacidad explicativa: 2'1626

Porcentaje de explicación: 9'83%

Variables muy importantes: (Carece)

Variables importantes:

- 29 - Prados (negativa).
- 42/24 - Capital ganadero per cápita (negativa).

Variables secundarias: (Carece)

Variables terciarias:

- 26 - Altura (negativa).
- 43 - Bovino (negativa).
- 46 - Cerda.

Este factor puede resumirse como la ausencia de ganado y de prados.

En conclusión, a la vista de este análisis factorial, podemos decir que tras la rotación, la interpretación se ha hecho más fácil, pero de todas formas nos encontramos todavía con problemas serios que se podían resumir a mi juicio en lo siguiente:

En primer lugar, los terrenos al estar calculados en tantos por ciento sobre el total, hace que si un municipio - tiene mucho cereal, no pueda tener mucho prado, por ejemplo, pero esto es lógico pues al ser los terrenos un bien limitado, lógicamente lo podemos dedicar o a unos casos o a otros, pero no a todos a la vez. Luego el que si un factor tiene correlación positiva con unos cultivos, tenga negativa con otros, es cosa razonable.

Sin embargo, no pasa así con la ganadería. Un municipio puede perfectamente tener mucho ganado de todo tipo, y tal como tenemos calculadas las variables estamos forzando a que si un municipio tiene un tanto por ciento alto de ganado bovino, por ejemplo, ha de tener, en consecuencia, un tanto por ciento bajo en los demás tipos de ganado y esto distorsiona el análisis como hemos podido comprobar.

En consecuencia, hemos de transformar las variables - que hacen referencia a la ganadería y eso es lo que haremos en nuestro siguiente análisis factorial.

Análisis Factorial 2.

Lo que haremos en este análisis que se encuentra en las hojas de ordenador nº 960 y ss. con el nombre de Análisis Factorial 2, será, manteniendo todas las condiciones de ordenador ya citadas,, transformar algunas variables para - que nos den factores que realmente definan la semejanza, o falta de semejanza, entre municipios, según las características agrícolas y ganaderas que los definan, a partir de los datos que poseemos.

En consecuencia, lo que vamos a hacer es suprimir en primer lugar las variables:

- 33 - Leguminosas.
- 39 - Almendro.
- 41 - Olivar y otros. (1)

Por no aportar demasiado en el anterior análisis.

También vamos a transformar las variables 43, 44, 45, 46 y 47, en las 43', 44', 45', 46' y 47'. Esto quiere decir

- (1) En el pasado el olivar estaba asociado con el viñedo y por lo tanto, tenía interés, pero en la actualidad esta asociación ha desaparecido. La existencia de esta asociación puede comprobarse en Riqueza Nacional de España, Universidad Comercial de Deusto (1.968), tomo II página 146, aunque en esta época ya era pequeña. En - nuestro estudio el coeficiente de correlación entre estas variables es de 0'13406 y, por lo tanto, pequeño.

que pasamos de tanto por ciento de un tipo de ganado en el capital ganadero municipal a capital ganadero per cápita - de un determinado tipo de ganado.

El ordenador ha generado estas variables de la siguiente forma:

- 43' = 43 . 42/24
- 44' = 44 . 42/24
- 45' = 45 . 42/24
- 46' = 46 . 42/24
- 47' = 47 . 42/24

La razón de esta transformación, ya comentada, podemos resumirla en que las variables 43 a la 47, tal como estaban concebidas en el banco de datos inicial, distorsionaban el análisis.

Si simplemente hubiéramos multiplicado por la variable 42 para tener los resultados en cifras absolutas, los municipios grandes tendrían gran ventaja sobre los pequeños y por ello hemos dividido también por la 24, para que al ser per cápita se resuelva la dificultad antedicha.

Con todo lo que hemos comentado, la tabla de equivalencias entre la numeración nuestra y la del ordenador quedará:

<u>Nº VARIABLE</u>									<u>Nº ORDENADOR</u>
26	1
27	2
28	3
29	4
30	5
31	6
32	7
34	8
35	9
36	10
37	11
38	12
40	13
42/24.	14
43'	15
44'	16
45'	17
46'	18
47'	19

El modelo resuelto por ordenador, con cuatro factores, explica el 71'864% de la varianza total, el resultado, por lo tanto, ha mejorado mucho respecto al anterior y podemos considerarnos satisfechos en lo que a esto respecta.

Pasemos ya directamente a la interpretación antes de

la rotación (manteniendo la misma nomenclatura que en el anterior análisis).

Factor 1

Valor propio: 6'4715

Porcentaje de explicación: 34'06%

Variables muy importantes: (Carece)

Variables importantes:

- 28 - Cultivos (negativa).
- 29 - Prados.
- 42/24 - Capital ganadero per cápita.
- 43' - Bovino per cápita;

Variables secundarias:

- 26 - Altura.
- 27 - Regadío (negativa).
- 30 - Forestal.
- 31 - Otras superficies.
- 32 - Cereales (negativa).
- 44' - Ovino per cápita.
- 45' - Caprino per cápita.
- 47' - Trabajo per cápita.

Variables terciarias:

- 34 - Patata (negativa).
- 35 - Industriales (negativa).

37 - Hortalizas (negativa).

40 - Viñedo (negativa).

El factor es plenamente coherente y se puede definir como factor ganadero, pero en el sentido de ganado libre y no estabulado como da a entender la abundancia de prados y la falta de énfasis en los forrajeros. Por otro lado, es el factor más explicativo y el que mejor diferencia al colectivo, es decir, por un lado las tierras altas y ganaderas y por otro las tierras bajas y agrícolas.

Factor 2

Valor propio: 3'6719

Porcentaje de explicación: 19'33%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

28 - Cultivos.

31 - Otras superficies (negativa).

44' - Ovino per cápita.

45' - Caprino per cápita.

47' - Trabajo per cápita.

Variables terciarias:

26 - Altura (negativa).

27 - Regadío.

- 30 - Forestal (negativa).
- 32 - Cereales.
- 34 - Patata.
- 35 - Industriales.
- 40 - Viñedo.
- 42/24 - Capital ganadero per cápita.
- 43' - Bovino per cápita.

Este factor, puede resultar algo confuso en una primera aproximación, pero estudiándolo de forma detenida, vemos que asocia el ganado (ovino, caprino y trabajo, fundamentalmente) con la agricultura (cultivos). Podemos concluir que es un factor agrícola-ganadero, con ciertas especializaciones característico de zonas no muy altas.

Factor 3

Valor propio: 2'2541

Porcentaje de explicación: 11'86%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias:

- 27 - Regadío.
- 36 - Forrajeras.
- 37 - Hortalizas.
- 38 - Frutales.

Variables terciarias:

- 32 - Cereales (negativa).
- 34 - Patata (negativa).
- 35 - Industriales (negativa).

Este es un factor no excesivamente definido, pero si expresa con claridad un tipo de especialización agrícola con forrajeras, hortalizas y frutales en regadío y carencia de cereales, patata e industriales.

Factor 4

Valor propio: 1'2567

Porcentaje de explicación: 6'61%

Variables muy importantes: (Carece)

Variables importantes:

- 46' - Cerda per cápita.

Variables secundarias: (Carece)

Variables terciarias:

- 34 - Patata.
- 36 - Forrajeras.
- 40 - Viñedo (negativa).

Este factor es fundamentalmente ganado de cerda, con algunas connotaciones agrícolas de patata y forrajeras, con ausencia de viñedo.

Terminado este estudio previo a la rotación, hemos de admitir que los resultados han mejorado mucho respecto al anterior análisis factorial, gracias a las transformaciones - realizadas en las variables.

Pasamos a continuación a estudiar los resultados tras la rotación.

Factores rotados:

Factor 1

Capacidad explicativa: 5'0768

Porcentaje de explicación: 26'72%

Variables muy importantes:

42/24 - Capital ganadero per cápita.

44' - Ovino per cápita.

45' - Caprino per cápita.

47' - Trabajo per cápita.

Variables importantes:

29 - Prados.

43' - Bovino per cápita.

Variables secundarias: (Carece)

Variables terciarias: (Carece)

Es este un factor muy claro de ganadería, fundamental

mente con ovino, caprino y trabajo, y, en segundo lugar, bovino y prados. Podemos definirlo claramente como factor ganadero.

Factor 2

Capacidad explicativa: 4'6189

Porcentaje de explicación: 24'31%

Variables muy importantes:

28 - Cultivos.

Variables importantes:

31 - Otras superficies (negativa).

32 - Cereales.

Variables secundarias:

26 - Altura (negativa).

30 - Forestal (negativa)

34 - Patata.

35 - Industriales.

Variables terciarias:

27 - Regadío.

29 - Prados (negativa).

40 - Viñedo.

Es este también un factor muy claro de cultivos agrícolas, en concreto parecen más importantes para este factor los

cereales, patata, industriales y viñedo, por este orden. Lo denominaremos factor cultivos.

Factor 3

Capacidad explicativa: 2'5251

Porcentaje de explicación: 13'29%

Variables muy importantes: (Carece)

Variables importantes:

37 - Hortalizas.

Variables secundarias:

27 - Regadío.

36 - Forrajeras.

38 - Frutales.

Variables terciarias:

26 - Altura (negativa).

Este tercer factor es también bastante claro y coherente y podemos denominarlo genéricamente como huerta.

Factor 4

Capacidad explicativa: 1'4326

Porcentaje de explicación: 7'54%

Variables muy importantes: (Carece)

VARIABLES IMPORTANTES:

46' - Cerda per cápita.

VARIABLES SECUNDARIAS:

40 - Viñedo (negativa).

VARIABLES TERCIARIAS:

26 - Altura.

Este factor es similar a lo que era antes de rotar y se puede resumir en ganado de cerda fundamentalmente, con connotaciones de carencia de viñedo y altura positiva.

En conclusión, tenemos que tras la rotación nos encontramos con unos factores coherentes e interesantes. El único que pudiéramos ser poco útil a priori es el cuarto, pero nos interesa mantenerlo de cara a la comarcalización.

Cluster Análisis.

Como ya comentábamos con anterioridad, a partir de los cuatro factores agrícolas y ganaderos, vamos a intentar una comarcalización homogénea de La Rioja.

Recordemos que nos basábamos en este tipo de homogeneidad a causa de:

- Las características concretas de La Rioja.
- La carencia de otros datos.

Lo que vamos a realizar es un cluster análisis por el sistema de encadenamiento completo, por las razones que damos en la segunda parte de este trabajo.

Por otro lado podría parecernos interesante que dado que unos factores explican mucho más que otros convendría ponderar el peso de los factores a la hora de realizar el cluster, pero en este caso no es necesario, debido a que los factores tienen ya una varianza igual a su capacidad explicativa y en definitiva entran ya ponderados.

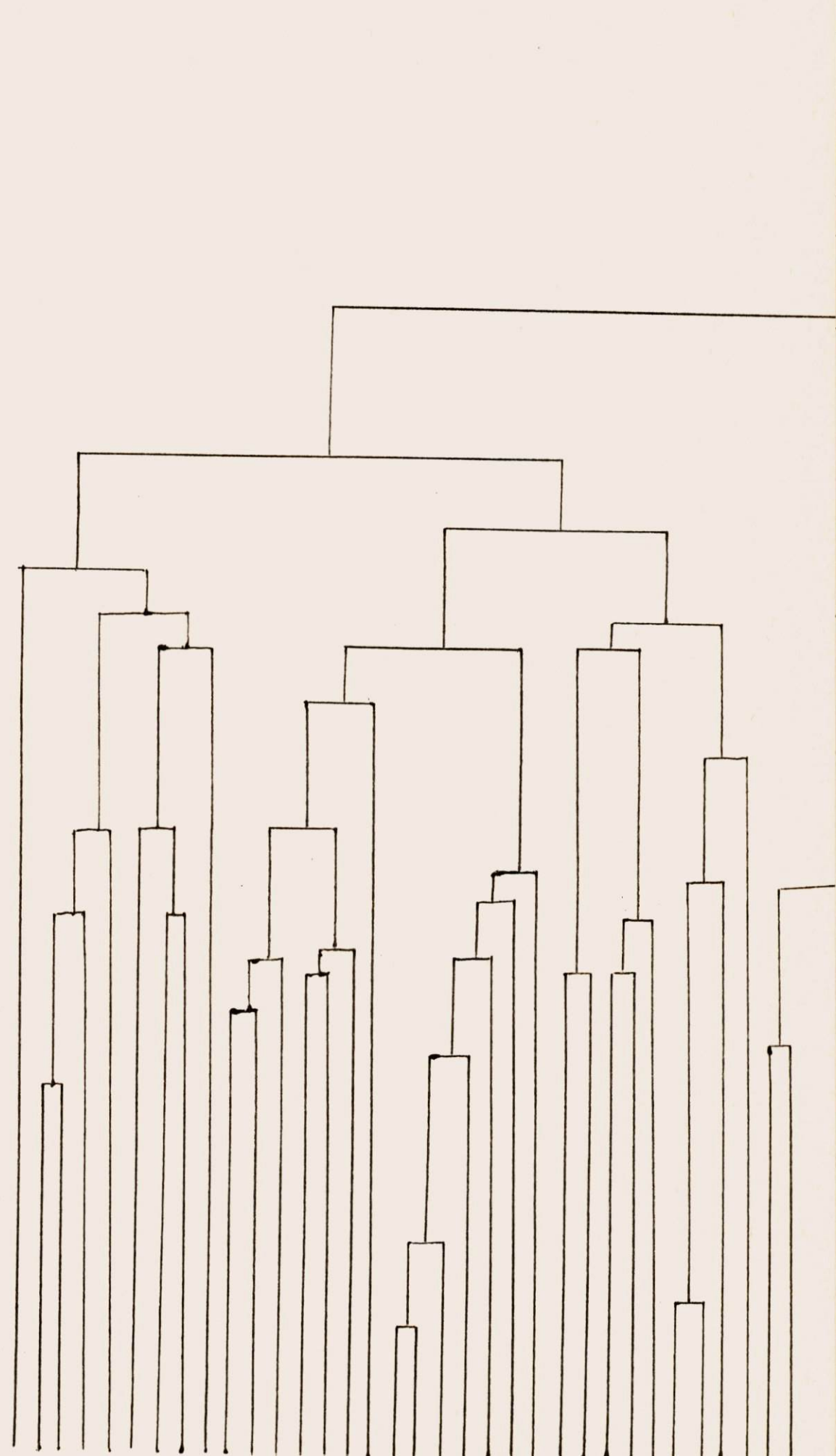
En consecuencia, pasemos a comentar los resultados del cluster que se encuentran en las hojas de ordenador nº 978 y siguientes.

Veamos a continuación el grafo del cluster.

AGRIC - 2 CLUSTER

Hasta.....

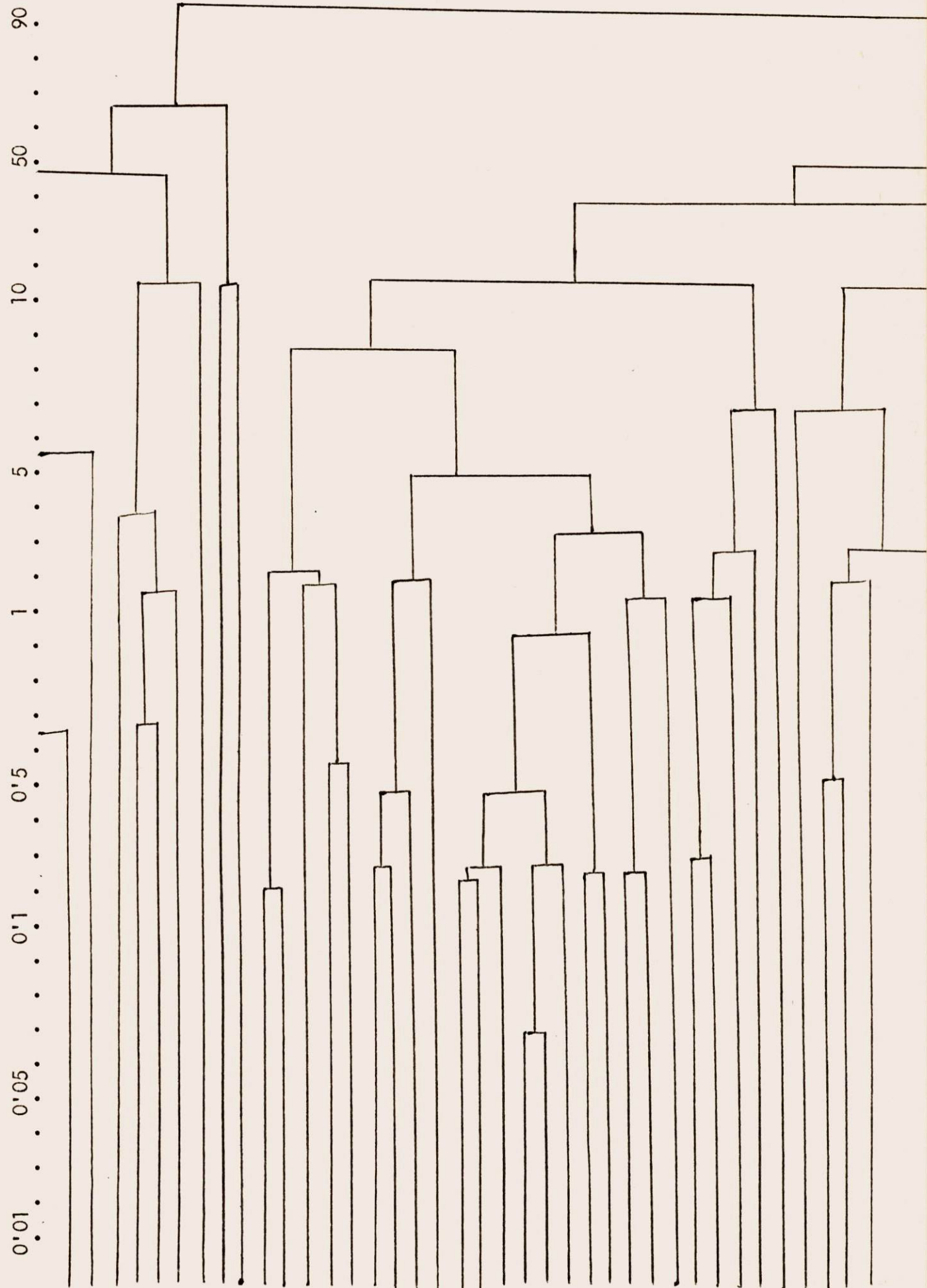
0'01 0'05 0'1 0'15 1 5 10 50 90 .



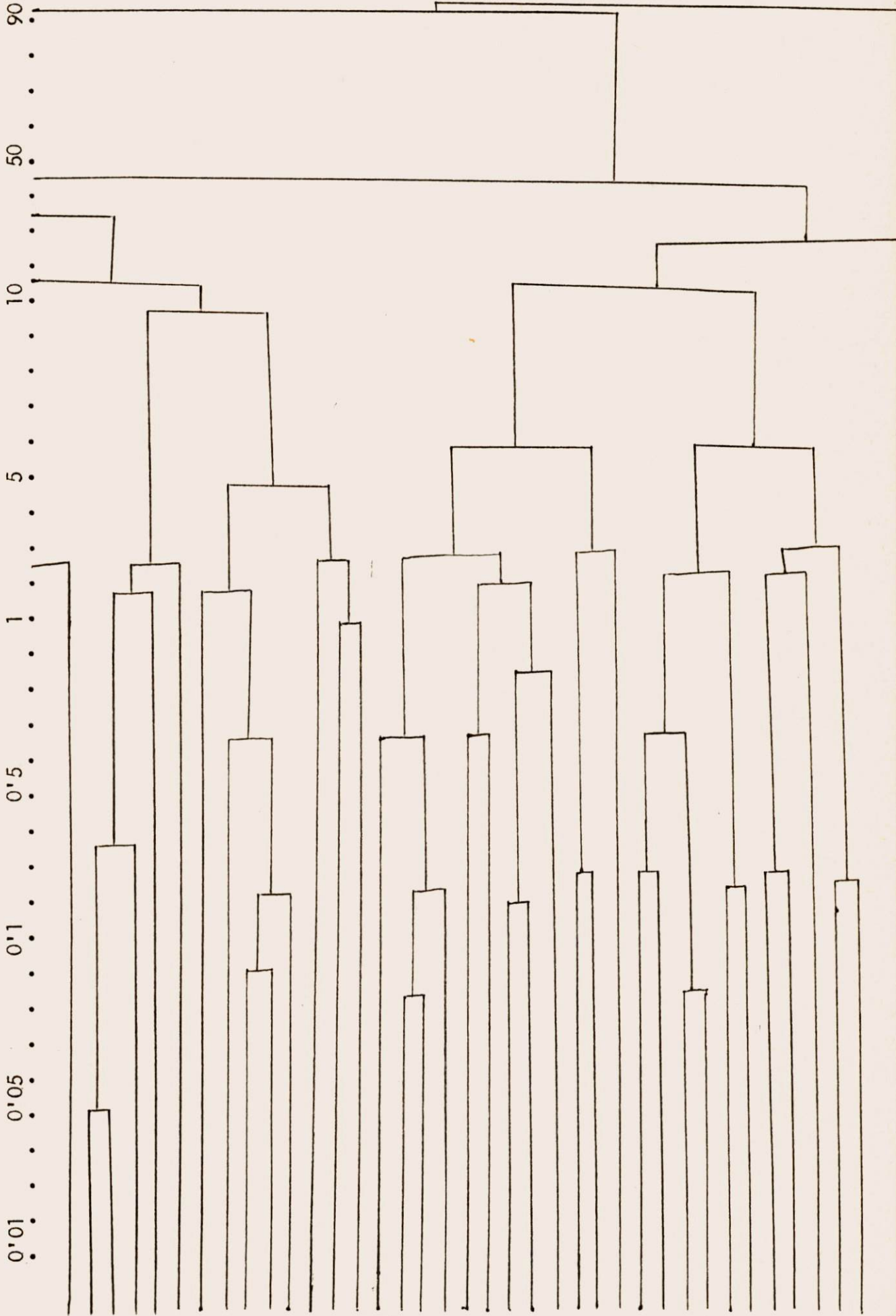
0'01
0'05
0'1
0'5
1
5
10
50
90

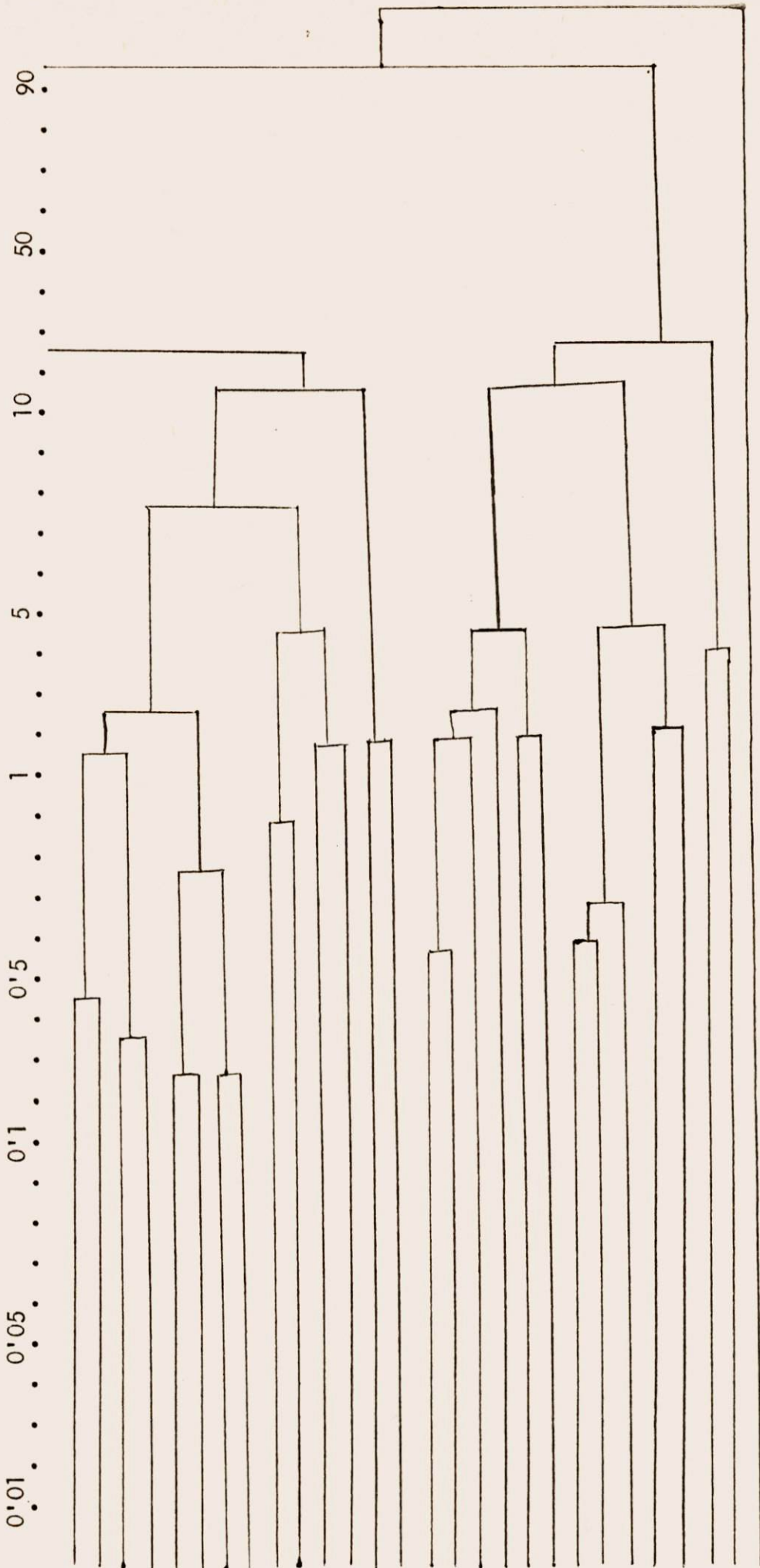
157
38
91
168
77
170
35
109
155
112
110
98
89
117
67
4
44
131
87
72
3
130
83
29
105
70
37
51
65
45
28
62
30
60
69
67
35
34





78 92 74 72 32 47 12 65 82 73 42 53 51 13 46 52 79 22 42 71 49 08 23 56 76 34 75 68 48 33 74 9 23 39 64 69 50 55





148
47
139
2
20
66
96
10
19
02
16
11
31
26
61
64
88
6
59
36
00
5
18
80
19
21
84
03

A partir del grafo ya expuesto, donde consta el número del municipio (según la numeración de ordenador) y cuando y con quién se va uniendo formando los clústeres, hemos de crear grupos apropiados para nuestro análisis.

Veamos una primera propuesta de corte del grafo, de tal manera que logramos 13 grupos; nos quedará una lista de municipios como la que veremos a continuación. En ella se ha mantenido el mismo orden que en el grafo cluster, y se ha añadido al número del municipio (según la numeración de ordenador) el nombre del mismo para su mejor identificación. También se ha puesto entre paréntesis la comarca oficial a la que pertenece según el fichero de la Diputación.

A continuación cada grupo se ha denominado con una letra, estas van desde la A hasta la M.

Veamos como queda:

- 175 - Zorraquín (Stº. D.)
- 143 - Tobia (Naj.)
- 140 - Soto de C. (Torr.)
- 118 - Rasillo (Torr.)
- 114 - Pradillo (Torr.)
- 129 - Sta. Col. (Naj.)
- 95 - Munilla (Arn.)
- 57 - Daroca (Log.)
- 128 - Román de C. (Torr.)
- 166 - Villarroya (Arn.)
- 85 - Ledesma de la C. (Naj.)
- 141 - Terroba (Torr.)
- 162 - Villanueva de C. (Torr.)
- 61 - Ezcaray (Sto. D.)
- 150 - Trevijano (Torr.)
- 104 - Nieva de C. (Torr.)
- 158 - Viguera (Log.)
- 58 - Enciso (Arn.)
- 107 - Ojacastro (Sto. D.)
- 145 - Torrecilla en C. (Torr.)
- 14 - Anguiano (Naj.)
- 115 - Prejano (Arn.)
- 17 - Arnedillo (Arn.)
- 154 - Valdemadera (Cerv.)
- 97 - Muro de Aguas (Arn.)
- 122 - Robres del C. (Log.)
- 111 - Pedroso (Naj.)
- 101 - Navajun (Cerv.)

- 137 - Sajuela (Log.)
- 54 - Cornago (Cerv.)
- 93 - Matute (Naj.)
- 126 - S. Millán C. (Naj.)

-
- 171 - Viniegra de Ab. (Naj.)
 - 81 - Jalón de C. (Torr.)
 - 157 - Ventrosa (Naj.)
 - 38 - Canal. Si. (Naj.)
 - 91 - Mansilla (Naj.)
 - 168 - Villavelayo (Naj.)
 - 77 - Hornillos C. (Log.)
 - 170 - Villoslada C. (Torr.)
 - 35 - Cabezón C. (Log.)
 - 109 - Ortigosa (Torr.)
 - 155 - Valganón (St. D.)
 - 112 - Pinillos (Torr.)
 - 110 - Pazuengos (Sto. D.)
 - 98 - Muro de C. (Torr.)
 - 89 - Lumbreras (Torr.)
 - 117 - Rabanera (Torr.)
 - 67 - Gallinero (Torr.)
 - 4 - Ajamil (Log.)
 - 44 - Castroviejo (Log.)

-
- 131 - St. Eul. Baj. (Arn.)
 - 87 - Leza (Log.)
 - 72 - Herce (Arn.)

- 3 - Aguilar R. Al. (Cerv.)
 - 130 - Sta. Engr. Jub. (Log.)
 - 83 - Lagunilla Jub. (Log.)
 - 29 - Bergasillas Bj. (Arn.)
 - 105 - Ocón (Arn.)
 - 70 - Grávalos (Cerv.)
 - 37 - Camprovín (Naj.)
 - 51 - Clavijo (Log.)
 - 65 - Galbarruñi (Haro)
 - 45 - Celórigo (Haro)
 - 28 - Bergasa (Arn.)
 - 62 - Foncea (Haro)
 - 30 - Bezares (Naj.)
 - 60 - Estollo (Naj.)
-

- 169 - Villaverde de R. (Naj.)
 - 167 - Villarta Quin. (Sto. D.)
 - 135 - Santurdejo (Sto. D.)
 - 134 - Santurde (Sto. D.)
 - 78 - Hornos del Monc. (Log.)
 - 92 - Manzan. Rioja (Sto. D.)
-

- 174 - Zarzosa (Arn.)
 - 172 - Viniegra de Arr. (Naj.)
 - 32 - Brieva de C. (Naj.)
 - 147 - Torre. en C. (Torr.)
 - 12 - Almarza de C. (Log.)
-

- 165 - Villarejo (Naj.)
 - 82 - Laguna de C. (Torr.)
-

- 173 - Zarratón (Haro)
 - 42 - Casalarreina (Haro)
 - 153 - Uruñuela (Naj.)
 - 151 - Tricio (Naj.)
 - 13 - Anguciana (Haro)
 - 146 - Torr. Al. (Naj.)
 - 52 - Cordovín (Naj.)
 - 79 - Huércanos (Naj.)
 - 22 - Azofra (Naj.)
 - 142 - Tirgo (Haro)
 - 71 - Haro (Haro)
 - 49 - Cihuri (Haro)
 - 108 - Ollauri (Haro)
 - 23 - Badarán (Naj.)
 - 56 - Cuzcurrita (Haro)
 - 76 - Hormilleja (Naj.)
 - 34 - Briones (Haro)
 - 75 - Hormilla (Naj.)
 - 68 - Gimileo (Haro)
 - 48 - Cidamón (Sto. D.)
 - 133 - San Torcuato (Sto. D.)
 - 74 - Hervias (Sto. D.)
 - 9 - Alesanco (Naj.)
 - 123 - Rodezno (Haro)
 - 39 - Canillas (Naj.)
-

- 164 - Villar de Torre (Naj.)
 - 69 - Grañón (Sto. D.)
 - 50 - Cirueña (Sto. D.)
 - 55 - Corporales (Sto. D.)
 - 40 - Cañas (Naj.)
 - 160 - Villalobar (Sto. D.)
 - 73 - Herramelluri (Haro)
 - 86 - Leiva (Sto. D.)
 - 24 - Bañares (Sto. D.)
 - 43 - Castañares de R. (Haro)
 - 149 - Treviana (Haro)
 - 144 - Tormantos (Sto. D.)
 - 132 - Sto. Domingo (Sto. D.)
 - 27 - Berceo (Naj.)
 - 25 - Baños de R. (Sto. D.)
 - 138 - Sorzano (Log.)
 - 106 - Ochanduri (Haro)
 - 90 - Manjarres (Naj.)
-
- 163 - Villar de Arn. (Calahorra)
 - 152 - Tudelilla (Calahorra)
 - 15 - Arenzana de Ab. (Naj.)
 - 20 - Ausejo (Log.)
 - 113 - Pradejón (Cal.)
 - 53 - Corera (Log.)
 - 99 - Nájera (Naj.)
 - 94 - Medrano (Log.)
 - 8 - Aldeanueva de E. (Cal.)

- 125 - San Asensio (Haro)
 - 41 - Cárdenas (Naj.)
 - 46 - Cenicero (Log.)
 - 159 - Villalva de R. (Haro)
 - 16 - Arenzana de Arr. (Naj.)
 - 124 - Sajazarra (Haro)
 - 63 - Fonzaleche (Haro)
 - 156 - Ventosa (Naj.)
 - 127 - San Millán Y. (Haro)
 - 136 - S. Vicente de la S. (Haro)
 - 1 - Abalos (Haro)
 - 33 - Briñas (Haro)
 - 21 - Autol (Cal.)
 - 7 - Alcanadre (Log.)
-

- 148 - Torremont. (Haro)
 - 47 - Cervera (Cerv.)
 - 139 - Sotes (log.)
 - 2 - Agoncillo (Log.)
 - 120 - Ribafrecha (Log.)
 - 66 - Galilea (Log.)
 - 96 - Murillo (Log.)
 - 10 - Aleson (Naj.)
 - 119 - El Redal (Log.)
 - 102 - Navarrete (Log.)
 - 116 - Quel (Arn.)
 - 11 - Alfaro (Cal.)
 - 31 - Bobadilla (Naj.)
 - 26 - Baños de R. T. (Naj.)
-

- 161 - Villamediana (Log.)
 - 64 - Fuenmayor (Log.)
 - 88 - Logroño (Log.)
 - 6 - Alberite (Log.)
 - 59 - Entrena (Log.)
 - 36 - Calahorra (Cal.)
 - 100 - Nalda (Log.)
 - 5 - Albelda (Log.)
 - 18 - Arnedo (Arn.)
 - 80 - Igea (Cer.)
 - 19 - Arrubal (Log.)
-

- 121 - Rincón (Cal.)
 - 84 - Lardero (Log.)
-

- 103 - Nestares (Torr.)

Veamos a continuación como van quedando los diversos mapas de cada grupo.

Dado que son trece grupos, la realidad es que resultaría muy confuso representarlos todos en el mismo mapa, a parte de la dificultad de utilizar distintos colores, pues sería imposible realizar fotocopias.

Ante este problema hemos decidido utilizar un mapa para cada grupo, coloreando de rayado la zona correspondiente a cada cluster en su mapa.

Los grupos los podemos denominar de la siguiente forma:

<u>GRUPO</u>	<u>DENOMINACION</u>
A	Sierra Media
B	Sierra Alta
C	Sierra Baja
D	Rioja Alta junto a Sierra
E	Sierra muy alta
F	Otras Sierras
G	Rioja Alta (Haro)
H	Rioja Alta (Santo Domingo)
I	Ribera uno
J	Ribera dos
K	Ribera tres
L	Ribera cuatro
M	Nestares

Antes de pasar a los mapas citados, creo que es importante hacer una advertencia respecto a las numeraciones:

La numeración utilizada por nosotros ha sido la que va del uno a ciento setenta y cinco, es decir, la de ordenador, pero los mapas van del uno al ciento ochenta y tres, que es la numeración existente en los mapas actuales de división provincial en términos municipales. Esta numeración tampoco coincide con la oficial del uno al ciento ochenta.

Para pasar de los 183 municipios a los 175, las ane-xiones realizadas son las siguientes:

- | | | |
|------------------------------------|---|----------------------|
| - Larriba | a | Ajamil |
| - Luezas | a | Soto de Cameros |
| - Montalvo de C. | a | San Román de Cameros |
| - La Santa | a | Munilla |
| - Poyales | a | Enciso |
| - Santa M ^{ra} de Cameros | a | San Román de Cameros |
| - Turruncún | a | Arnedo |
| - Zenzano | a | Lagunilla |

Adjuntamos a continuación una tabla de equivalencias de las tres numeraciones.

El lector puede preguntarse el por qué de no utilizar siempre la misma, la razón es que los mapas que utiliza la Diputación tienen una numeración, el ordenador solo utiliza los municipios actualmente existentes (175) y los datos tomados -

en el fichero de la Diputación tienen 180 números.

Nota: Posteriormente a la toma de datos, Trevijano se ha unido a Soto, pero prescindiremos de esta unión, por no haber sido introducida en el fichero.

Nº mapa	Nº ofi.	Nº ord.	Nº mapa	Nº ofi.	Nº ord.
1	1	1	28	28	28
2	2	2	29	29	29
3	3	3	30	30	30
4	4	4	31	31	31
5	5	5	32	32	32
6	6	6	33	33	33
7	7	7	34	34	34
8	8	8	35	35	35
9	9	9	36	36	36
10	10	10	37	37	37
11	11	11	38	38	38
12	12	12	39	39	39
13	13	13	40	40	40
14	14	14	41	41	41
15	15	15	42	42	42
16	16	16	43	43	43
17	17	17	44	44	44
18	18	18	45	45	45
19	19	19	46	46	46
20	20	20	47	47	47
21	21	21	48	48	48
22	22	22	49	49	49
23	23	23	50	50	50
24	24	24	51	51	51
25	25	25	52	52	52
26	26	26	53	53	53
27	27	27	54	54	54

Nº mapa	Nº ofi.	Nº ord.	Nº mapa	Nº ofi.	Nº ord.
55	55	55	79	79	79
56	56	56	80	80	80
57	57	57	81	81	81
58	58	58	82	82	82
59	59	59	83	83	83
60	60	60	84	84	84
61	61	61	85	(1)	(1)
62	62	62	86	85	85
63	63	63	87	86	86
64	64	64	88	87	87
65	65	65	89	88	88
66	66	66	90	(2)	(2)
67	67	67	91	89	89
68	68	68	92	90	90
69	69	69	93	91	91
70	70	70	94	92	92
71	71	71	95	93	93
72	72	72	96	94	94
73	73	73	97	95	(3)
74	74	74	98	96	95
75	75	75	99	97	96
76	76	76	100	98	97
77	77	77	101	99	98
78	78	78	102	100	99

(1) Larriba se une a Ajamil.

(2) Luezas se une a Soto de Cameros.

(3) Montalvo de C. se une a San Román de C.

Nº mapa	Nº ofi.	Nº ord.	Nº mapa	Nº ofi.	Nº ord.
103	101	100	127	125	123
104	102	101	128	126	124
105	103	102	129	127	125
106	104	103	130	128	126
107	105	104	131	129	127
108	106	105	132	130	128
109	107	106	133	131	(2)
110	108	107	134	132	129
111	109	108	135	133	130
112	110	109	136	134	131
113	111	110	137	(3)	(3)
114	112	111	138	135	132
115	113	112	139	136	133
116	114	(1)	140	137	134
117	115	113	141	138	135
118	116	114	142	139	136
119	117	115	143	140	137
120	118	116	144	141	138
121	119	117	145	142	139
122	120	118	146	143	140
123	121	119	147	144	141
124	122	120	148	145	142
125	123	121	149	146	143
126	124	122	150	147	144

(1) Poyales se une a Arnedo.

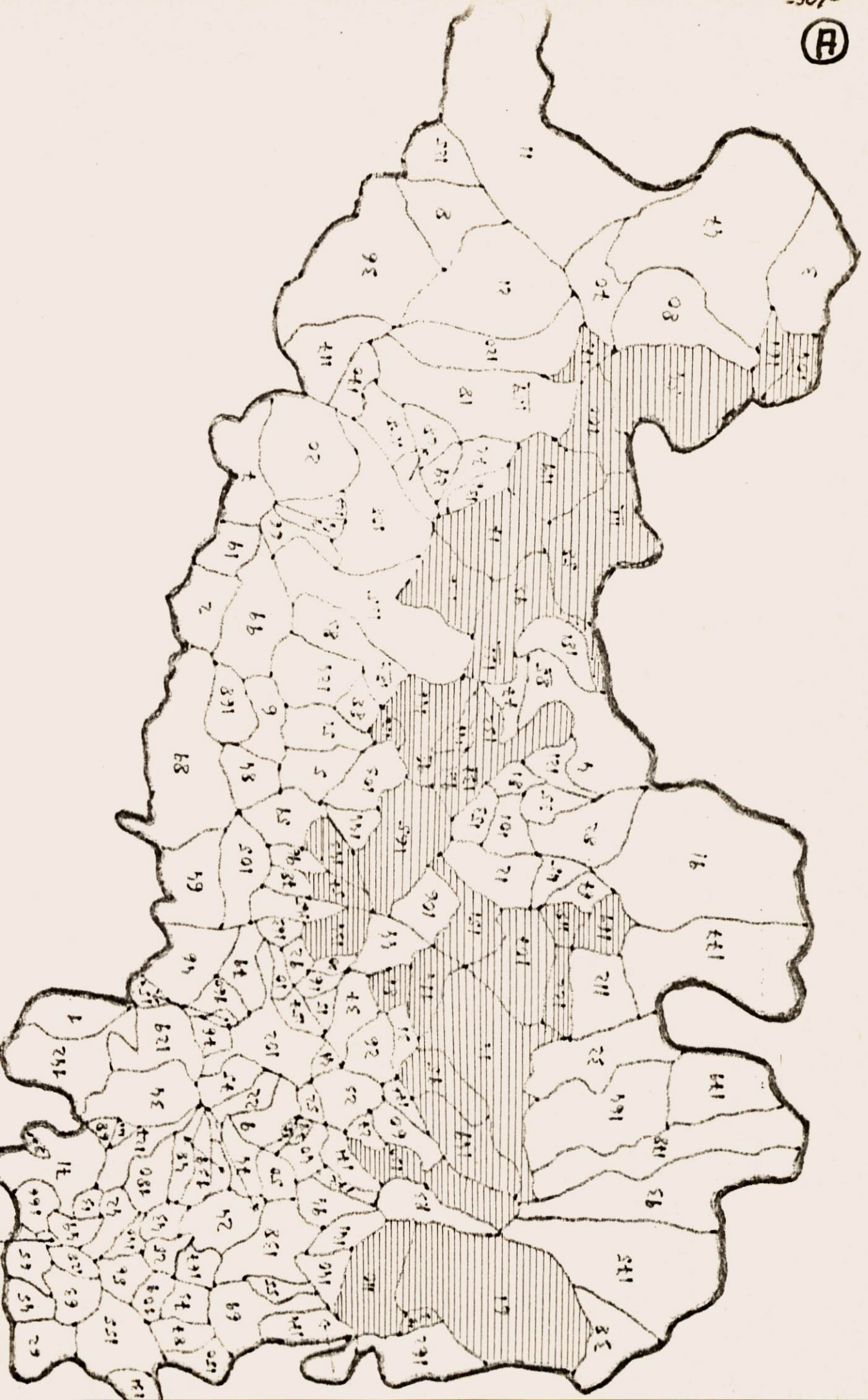
(2) La Santa se une a Munilla.

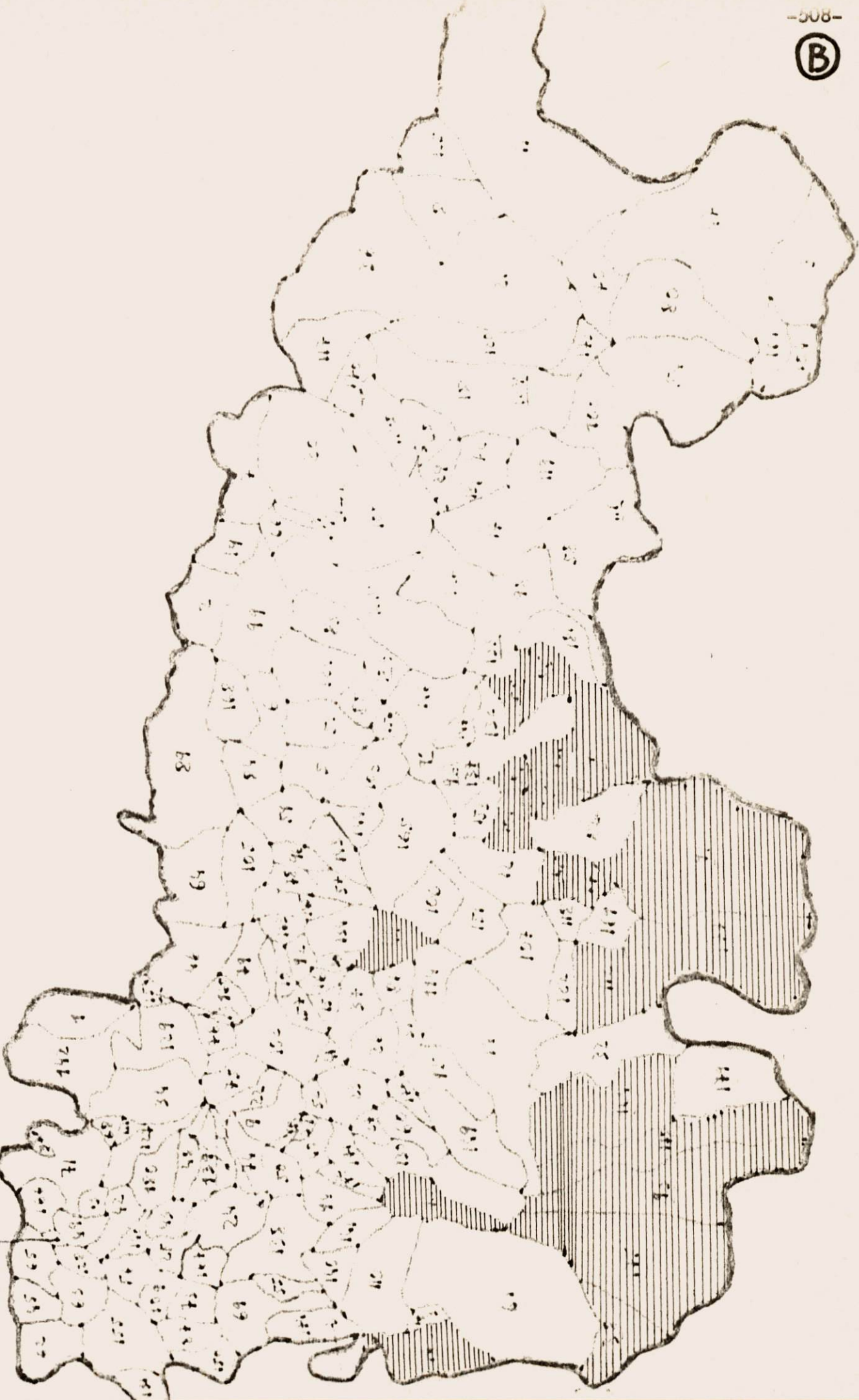
(3) Santa M^{ra} de Cameros se une a San Román de C.

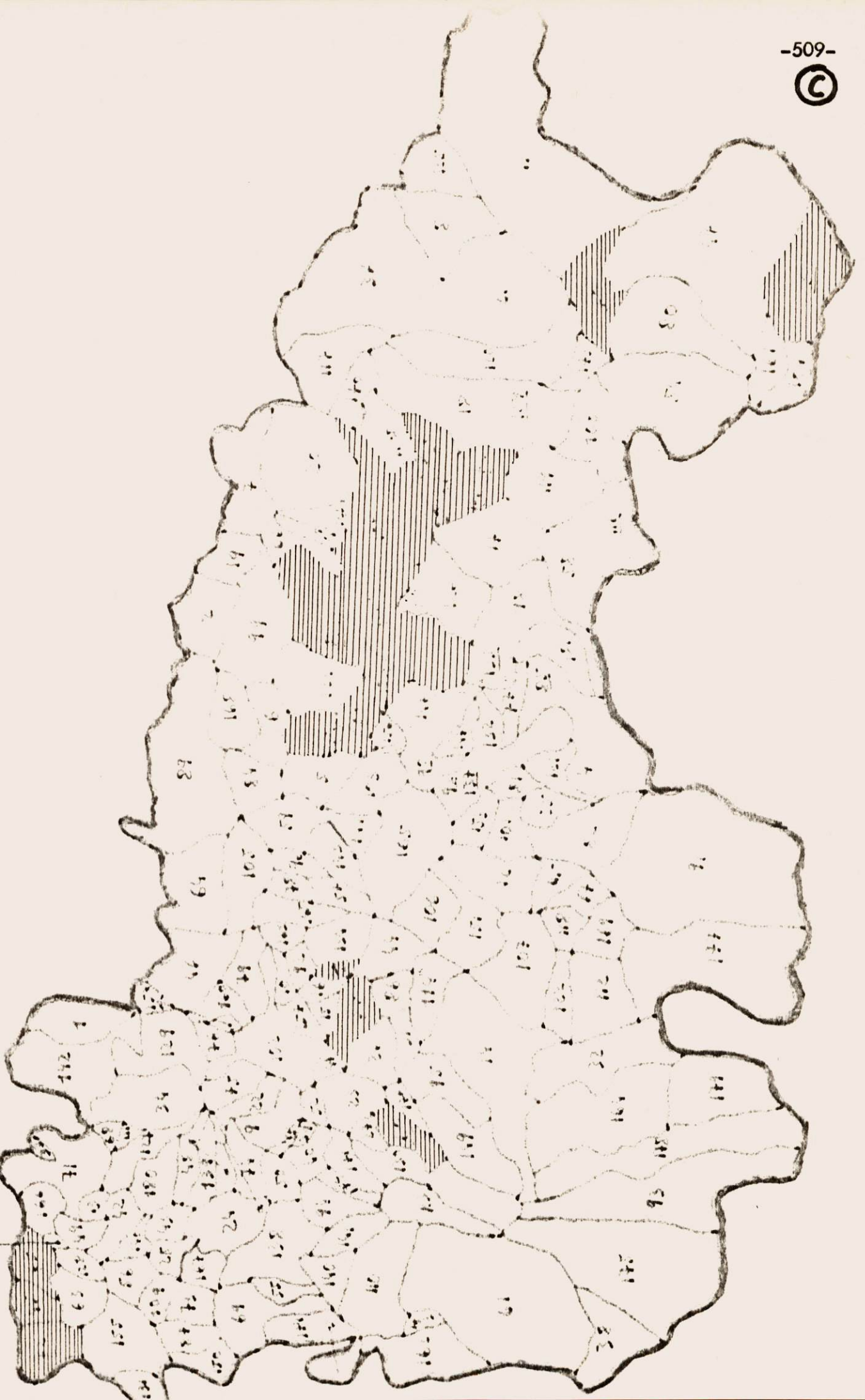
Nº mapa	Nº ofi.	Nº ord.	Nº mapa	Nº ofi.	Nº ord.
151	148	145	175	172	168
152	149	146	176	173	169
153	150	147	177	174	170
154	151	148	178	175	171
155	152	149	179	176	172
156	153	150	180	177	173
157	154	151	181	178	174
158	155	152	182	179	(2)
159	156	(1)	183	180	175
160	157	153			
161	158	154			
162	159	155			
163	160	156			
164	161	157			
165	162	158			
166	163	159			
167	164	160			
168	165	161			
169	166	162			
170	167	163			
171	168	164			
172	169	165			
173	170	166			
174	171	167			

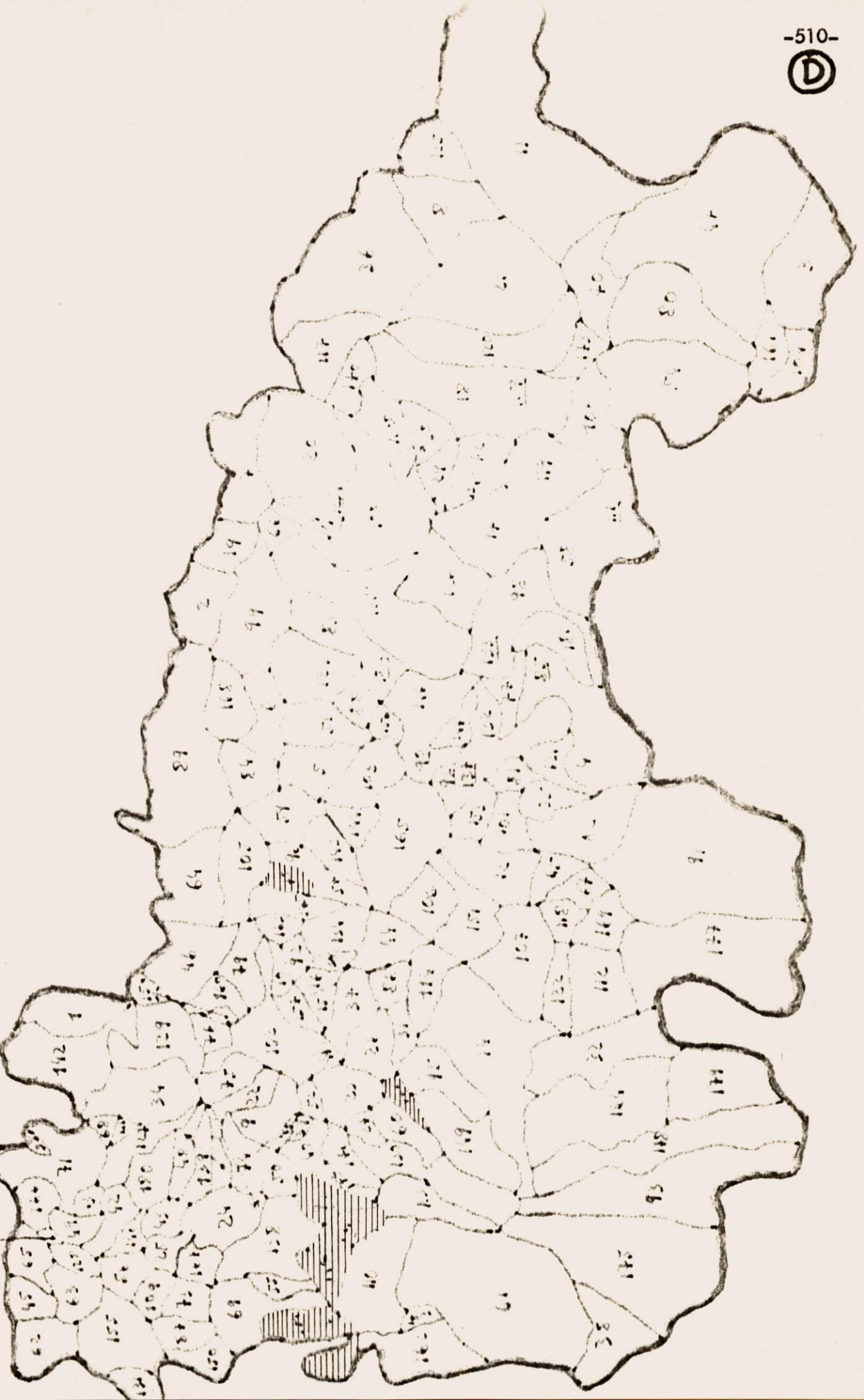
(1) Turruncún se une a Arnedo.

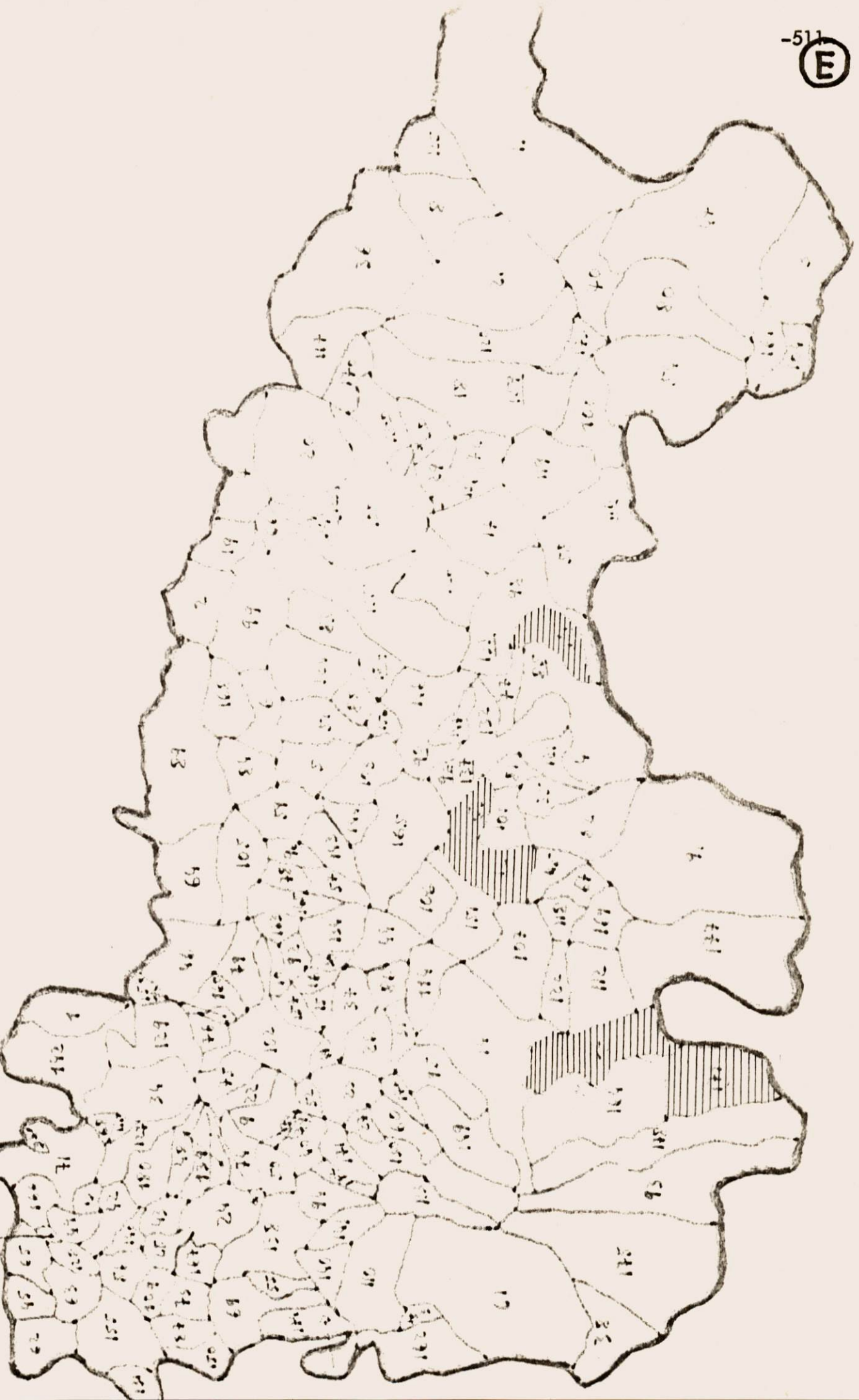
(2) Zenzano se une a Lagunilla de Jubera.

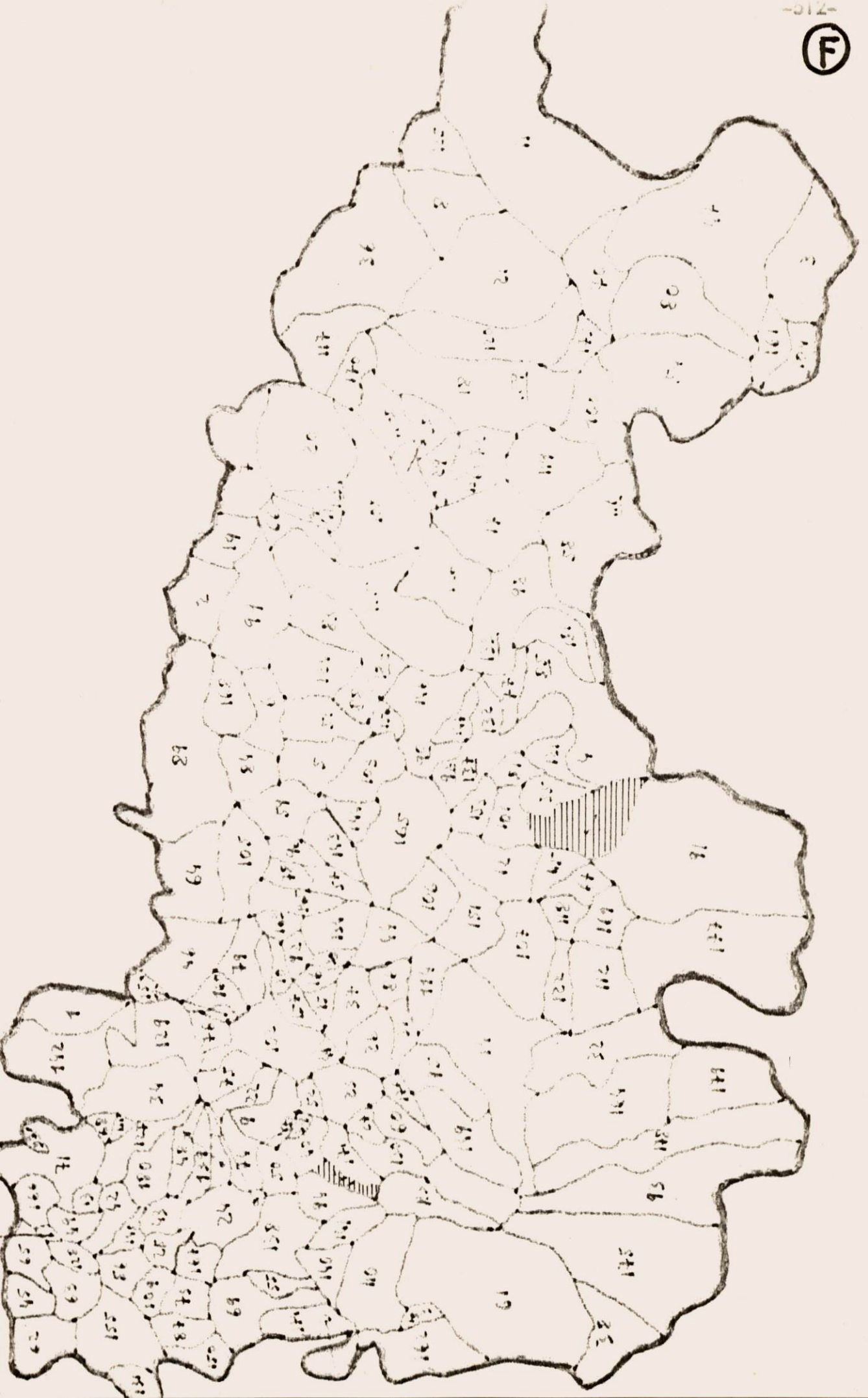


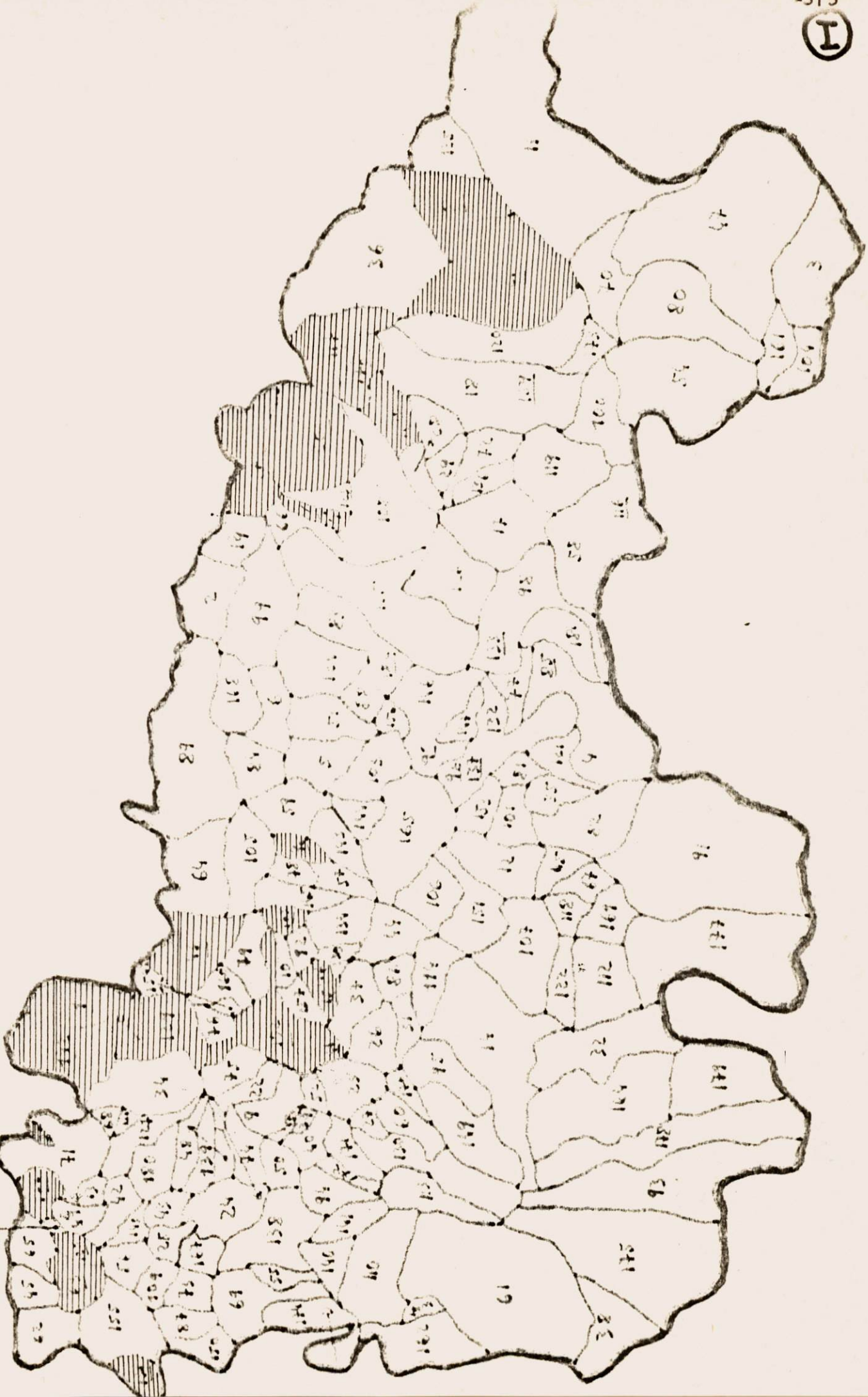


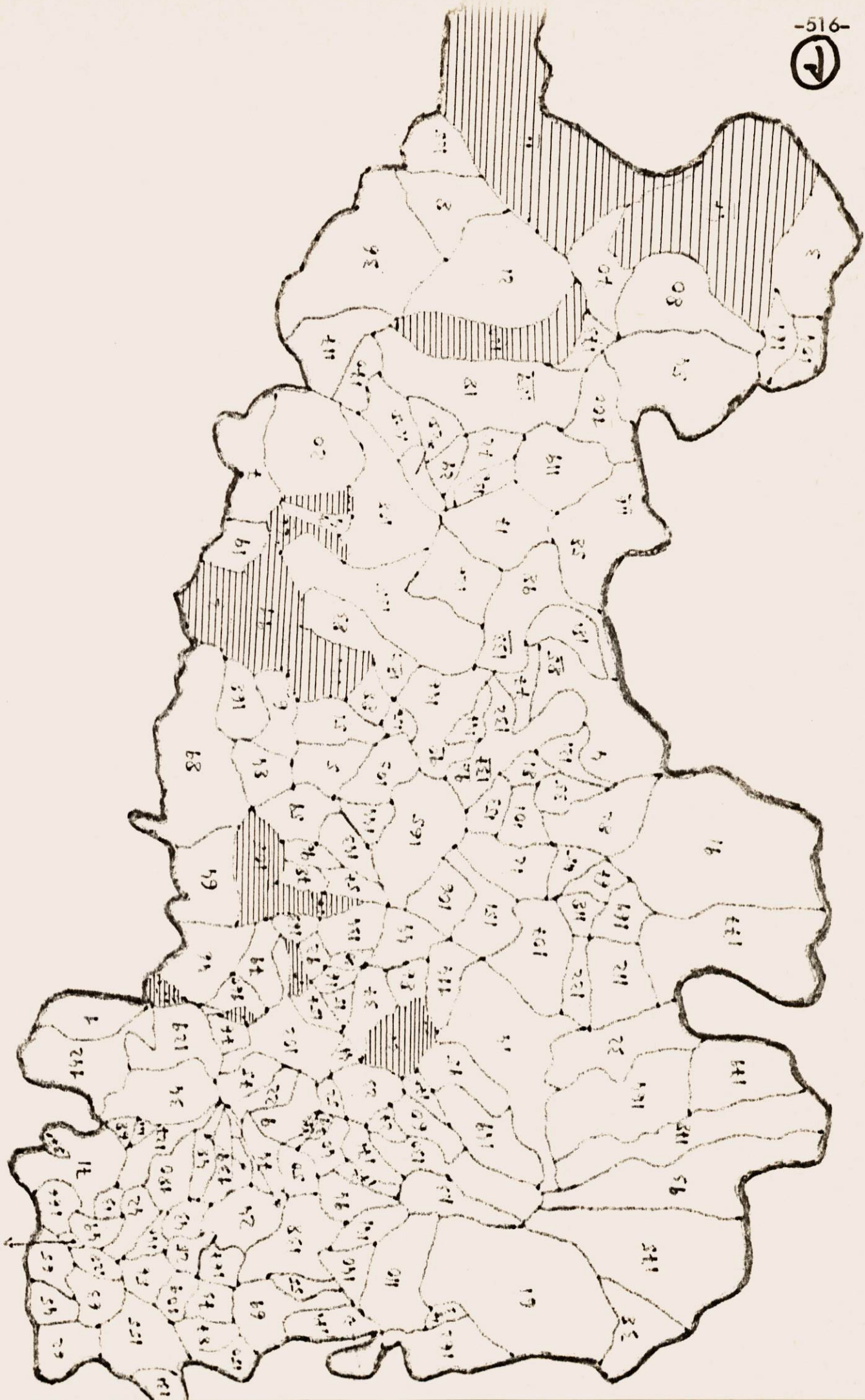


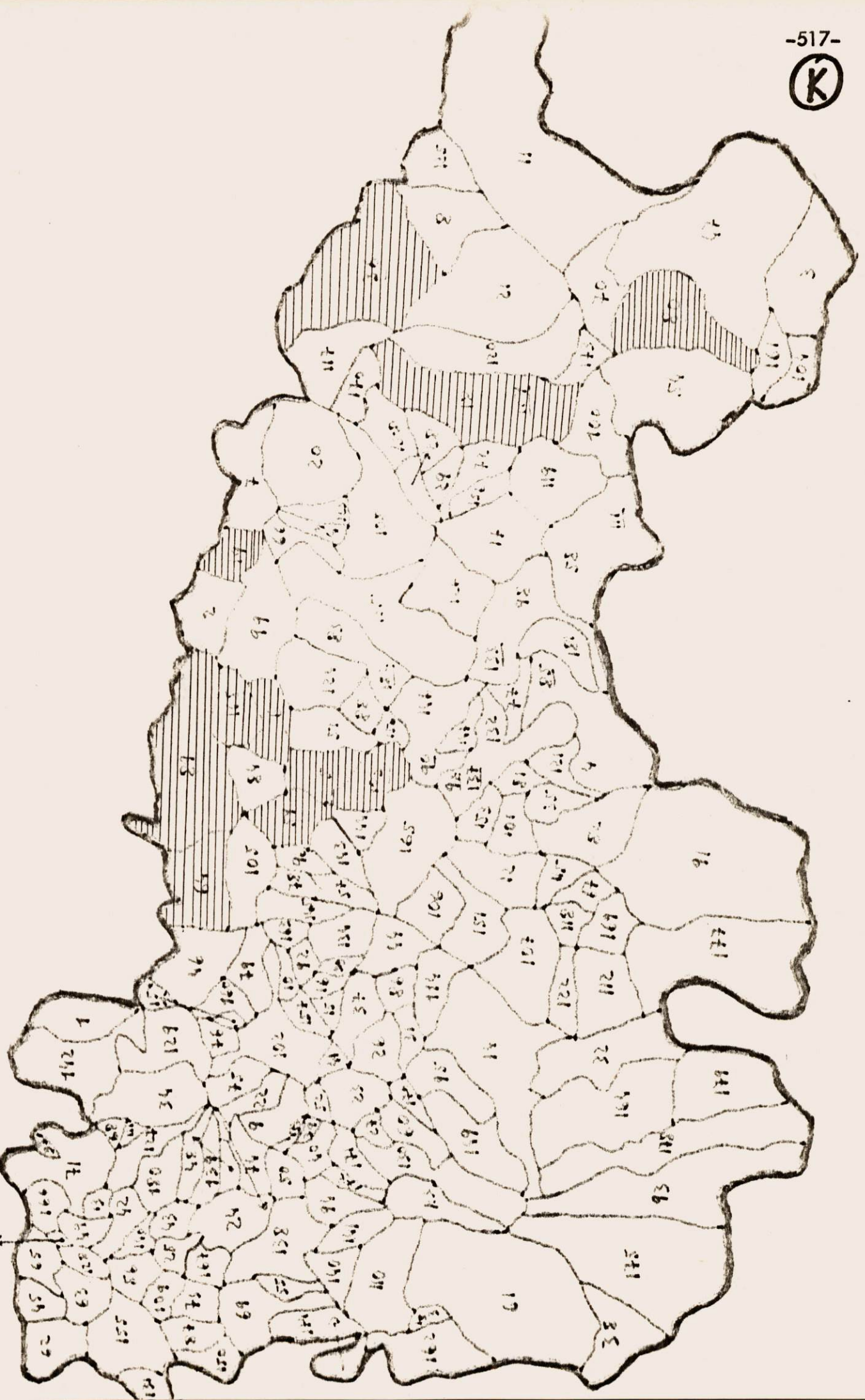


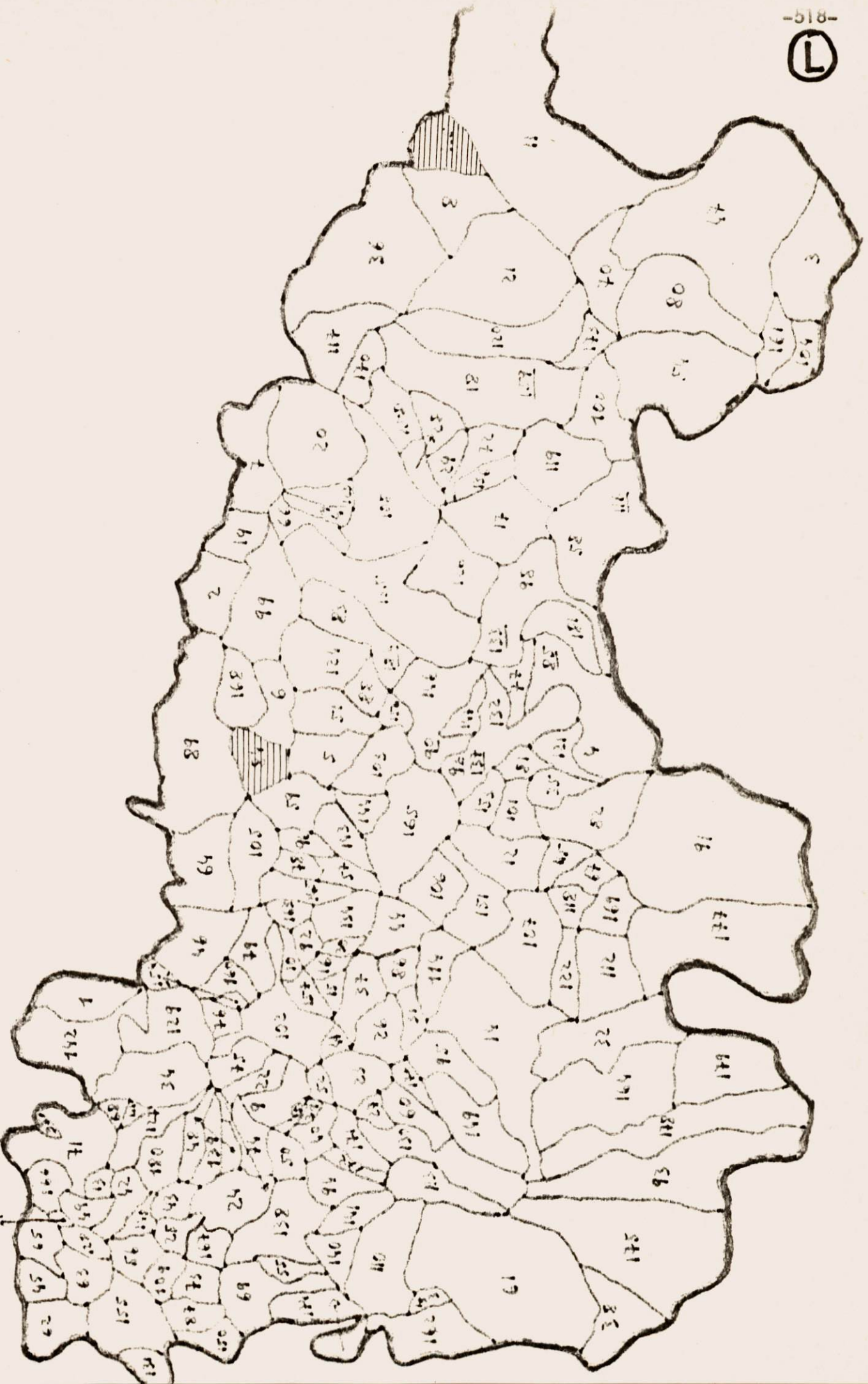


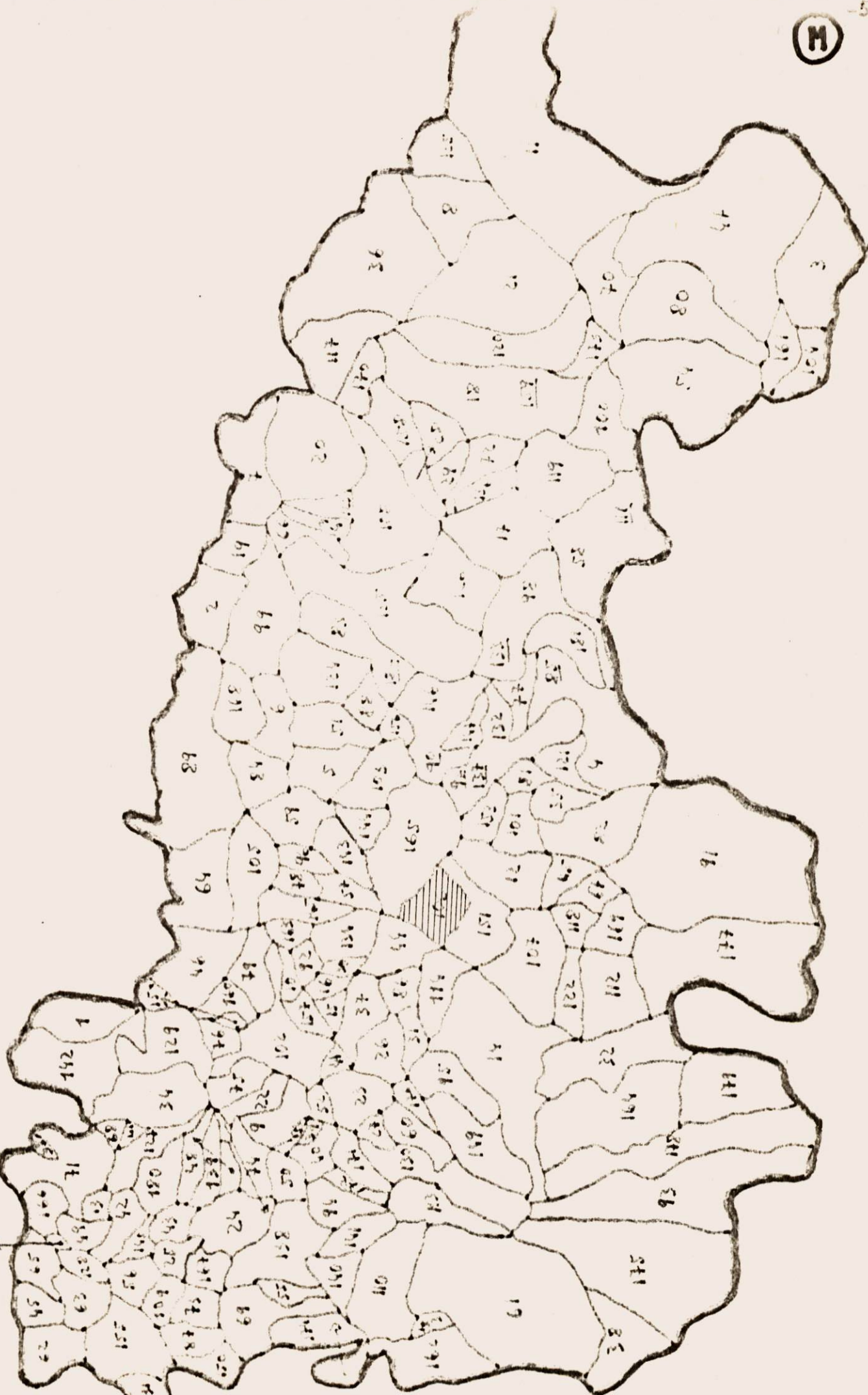












En conclusión de esta división podemos afirmar que hay bastante coherencia en el cluster, más aún si tenemos en cuenta las dos premisas siguientes:

a) Para hacer el cluster no se ha introducido la matriz de distancias y esto puede producir como consecuencia - que se unan pueblos muy distantes entre sí.

En nuestro estudio esto no ha sucedido con excesiva frecuencia, hecho del que debemos congratularnos y cuando ha sucedido es perfectamente explicable, como veremos en el siguiente apartado.

Sobre las razones de no introducir la matriz de distancias en el cluster, creo que se pueden resumir diciendo que si damos un peso alto a las distancias, el resto de las variables apenas si sirven para nada, si por el contrario se lo damos bajo, lo que no sirven son las distancias, en consecuencia, nos encontramos ante una manifiesta arbitrariedad a la hora de ponderar, por lo que he decidido no introducirla.

Además, el problema, que tiene gran importancia no sólo en planteamientos económicos, sino también geológicos, ha sido estudiado por los ingenieros de minas y de hecho no han llegado a conclusiones satisfactorias.

b) Los casos en que dos municipios del mismo grupo se encuentran alejados pueden deberse a que:

1) En realidad forman el mismo grupo, pues las condi ciones físicas son similares y las económicas también y el - hecho de que se encuentren alejados indica que a la hora de comarcalizar, a parte de la homogeneidad, habrá que tener en cuenta otras variables, entre las cuales la distancia es una de las más importantes.

2) Forman distinto grupo y su inclusión en el mismo grupo se debe a la imperfección de nuestras variables. Tenga mos en cuenta que productos como el espárrago, la lechuga, - etc. están todos entre los productos de huerta y evidentemente son productos distintos, pero tenemos la limitación pro - pia de lo agregado de nuestras variables en el campo agrícola.

Otro caso que se puede dar es que dos municipios simi lares, pertenezcan a grupos distintos, es el caso del municipio 103 - Nestares, que no se une a ningún otro. En este ca so el problema se debe a que la variable capital ganadero per cápita, a causa de lo poco poblado de este municipio y de su abundante ganadería, se dispara hacia arriba de forma poco - significativa. En otros casos, que sería monótono el anali - zar uno por uno, se producen situaciones similares al manejar nuestras variables, pero esos problemas se producen a menudo al intentar cuantitativizar de forma absoluta.

En último lugar quiero hacer constar aquí que muchas inexactitudes del estudio bien pudiesen deberse a lo impre ciso de los datos en muchos casos, sin que con ello quiera

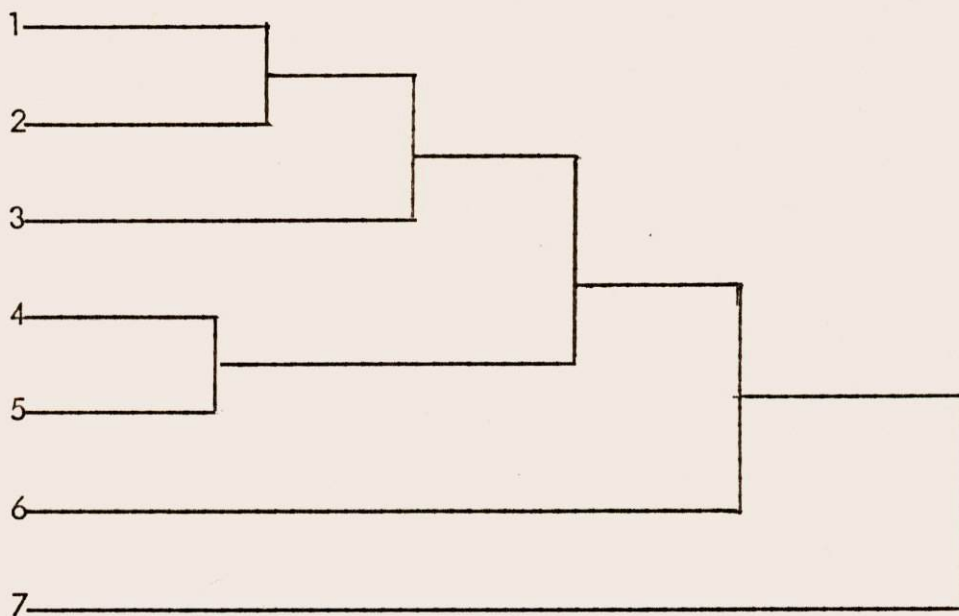
culpar a la Diputación Provincial de Logroño que ha sido mi principal suministrador, dado que con la penuria de medios que esta Institución ha tenido hasta ahora, hemos de admirar la cantidad de datos que posee, pudiendo ser envidiada por la mayoría de las Diputaciones españolas.

Vamos a continuación a intentar reducir el número de grupos a siete, cortando en el grafo un poco más arriba (concretamente en el 40).

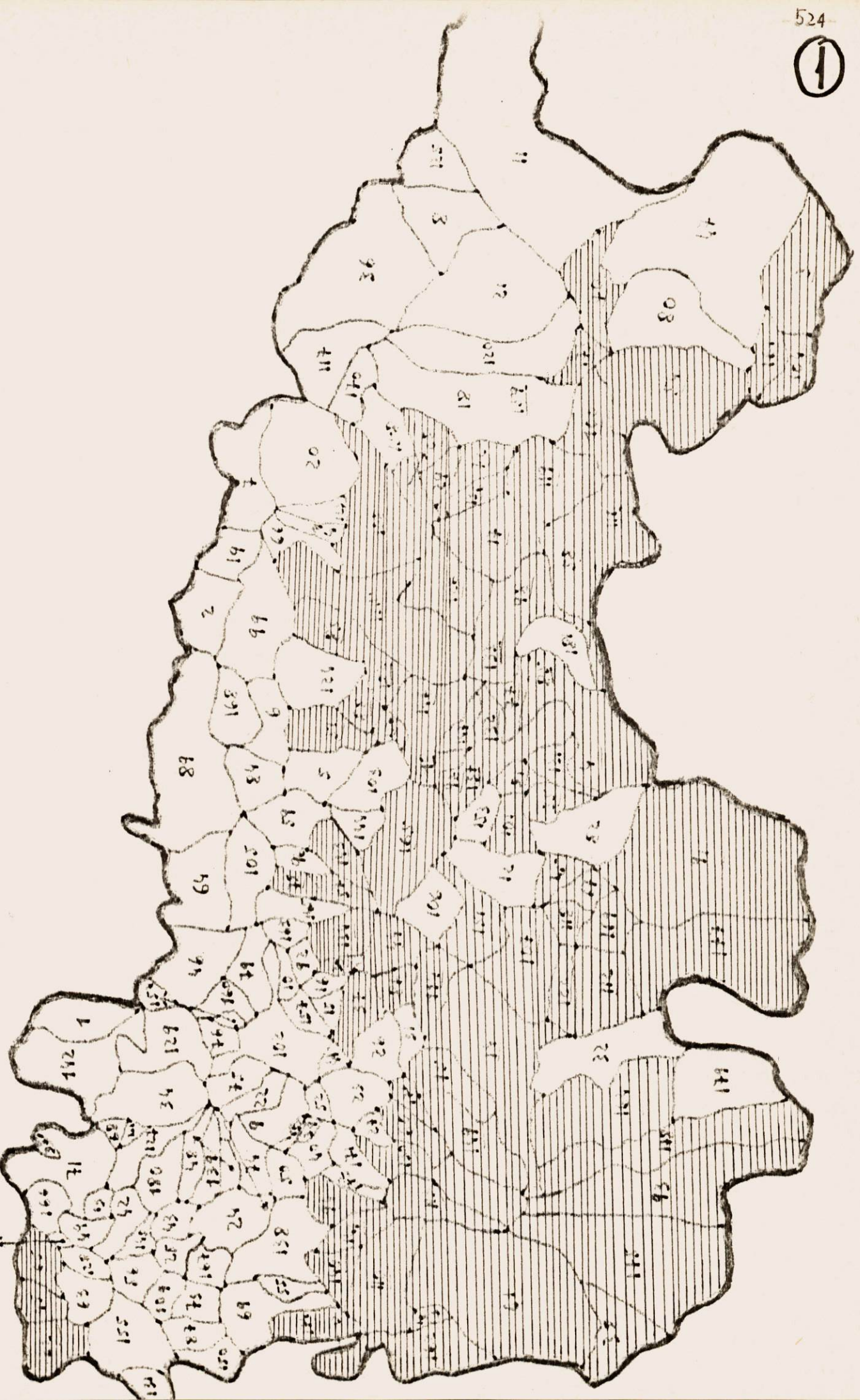
Veamos mediante un esquema como quedarían los siete grupos citados, y como se seguirían uniendo según el grafo cluster.

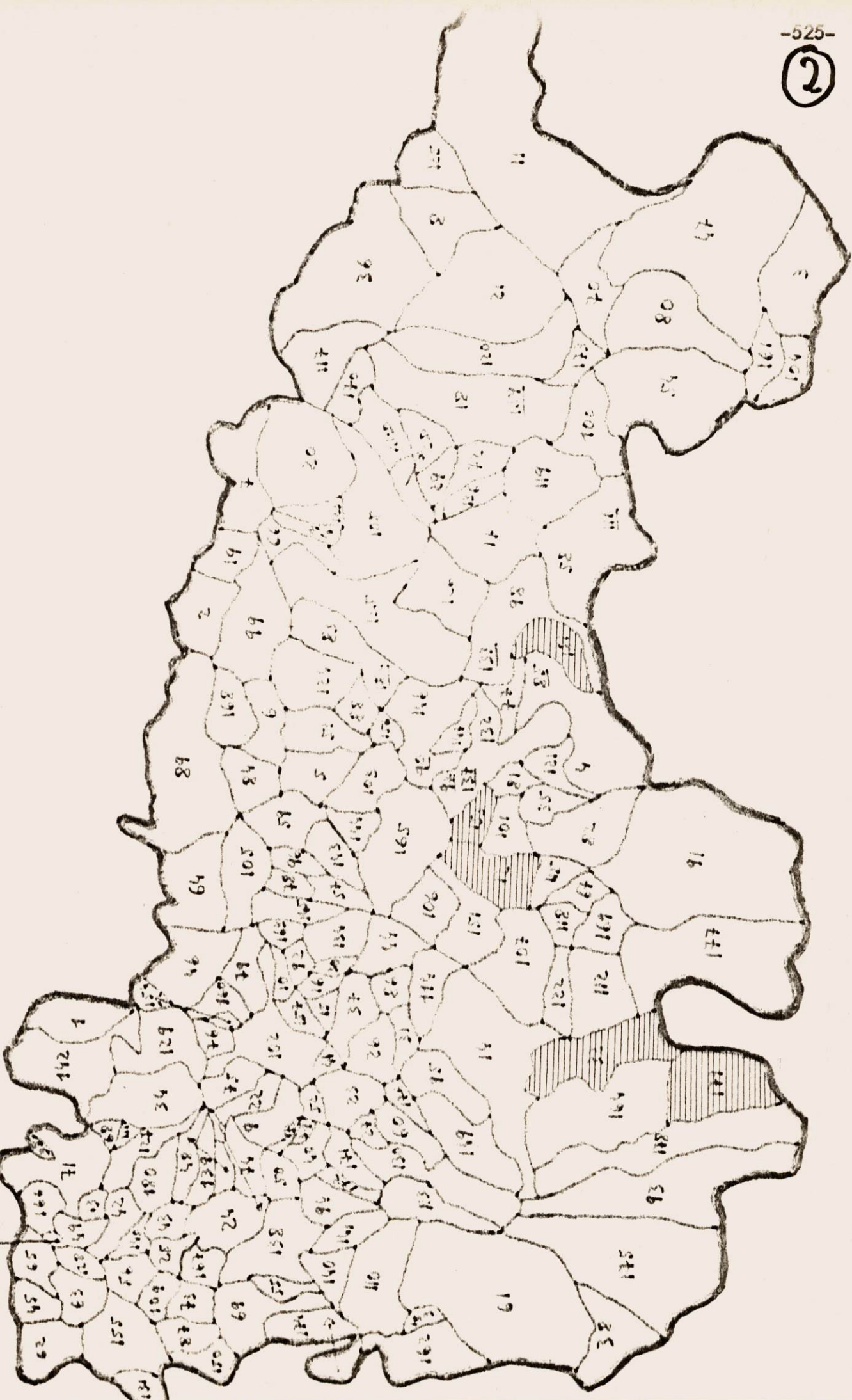
<u>Cluster</u>	<u>Grupos que comprende</u>
1	A, B, C, D
2	E
3	F
4	G, H
5	I, J
6	K, L
7	M

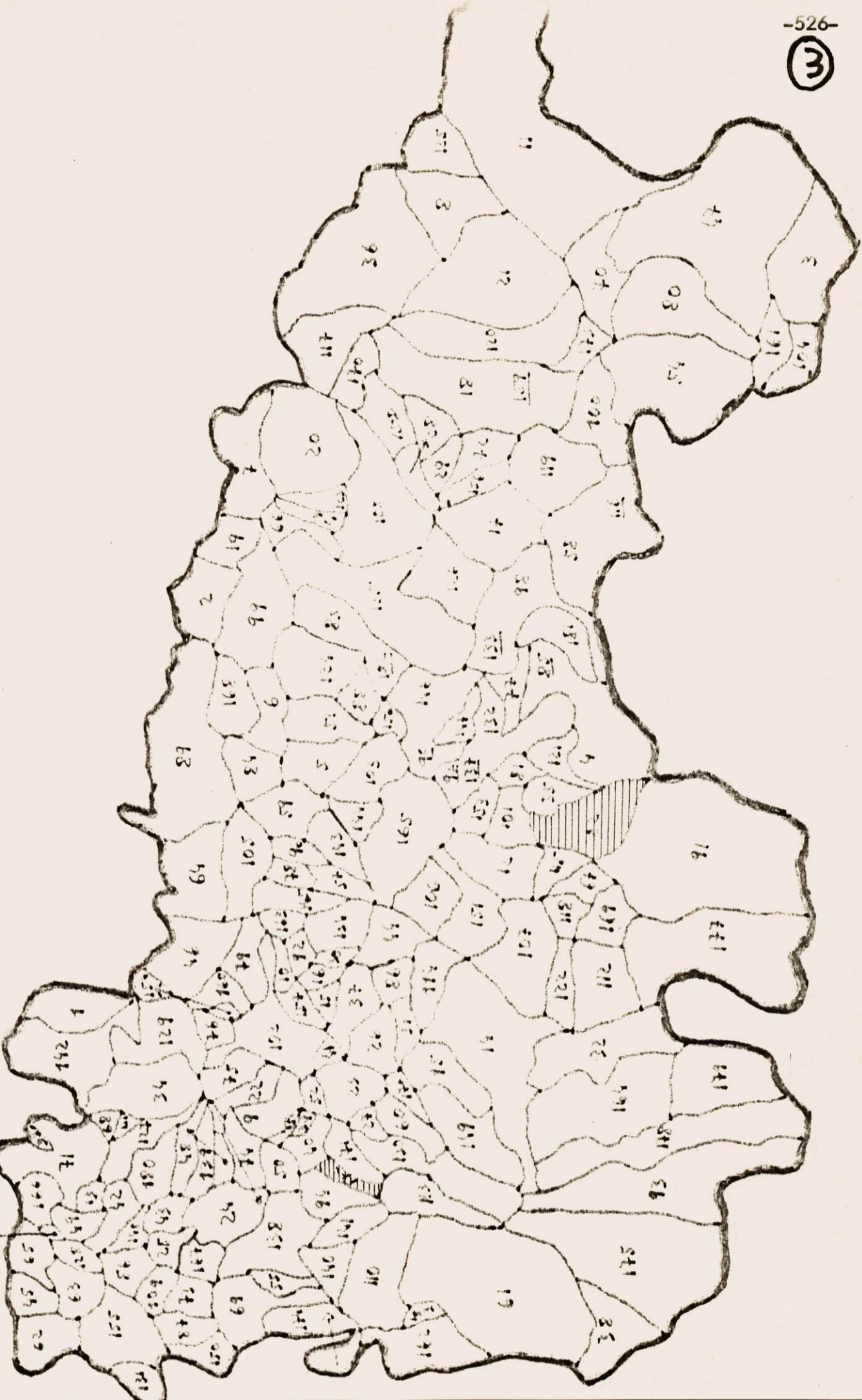
Uniones posteriores:



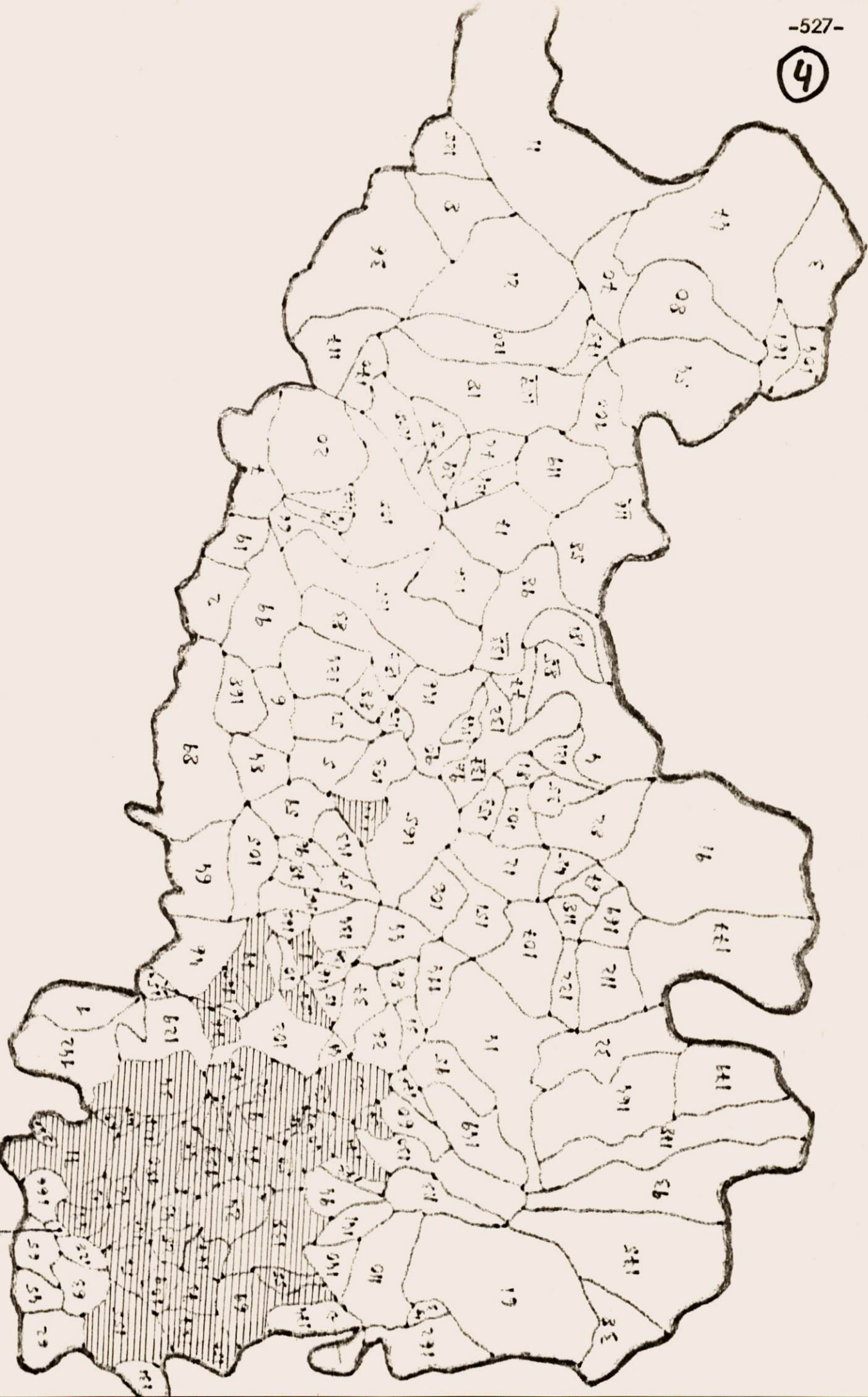
Los siete mapas correspondientes quedarían:

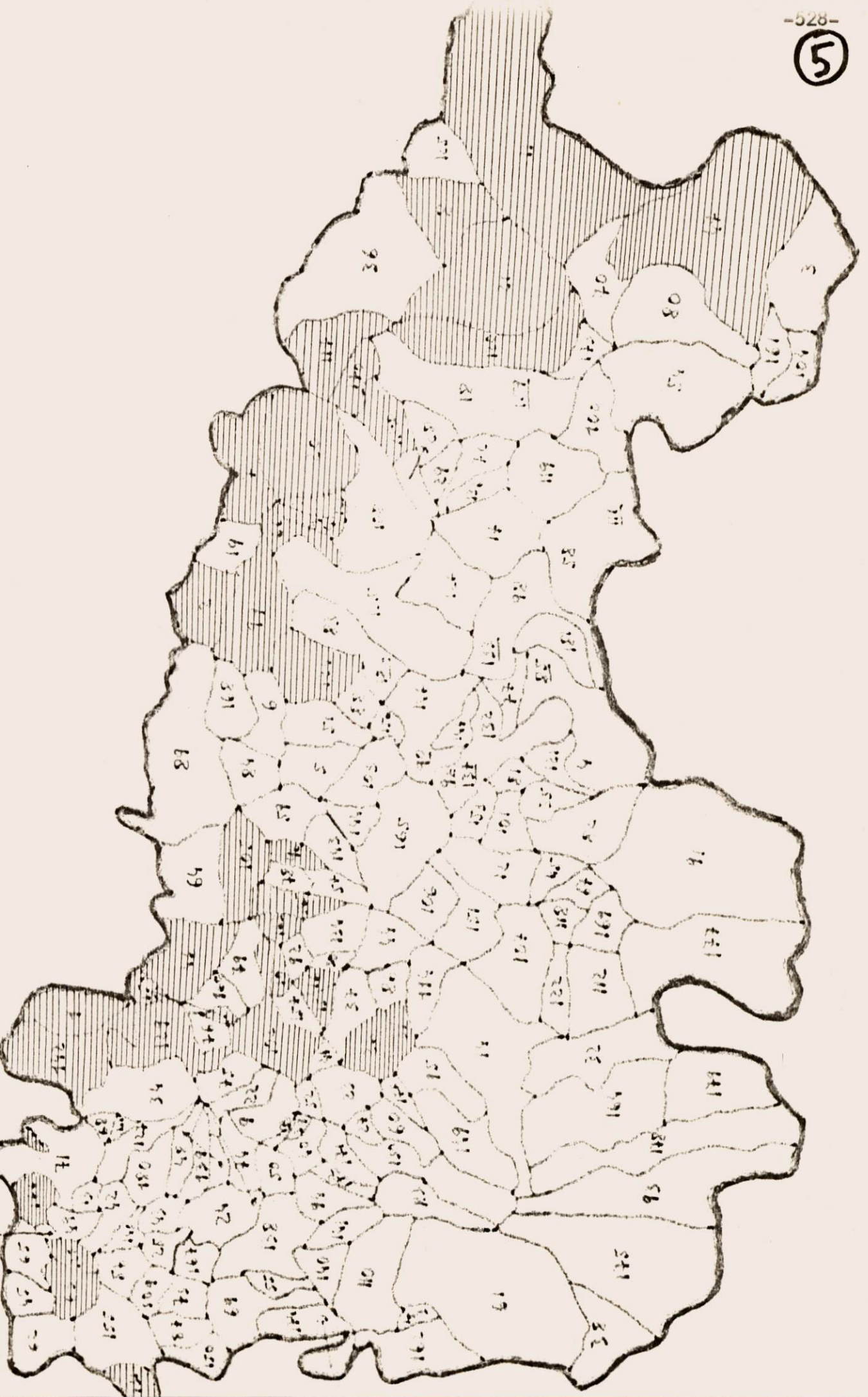




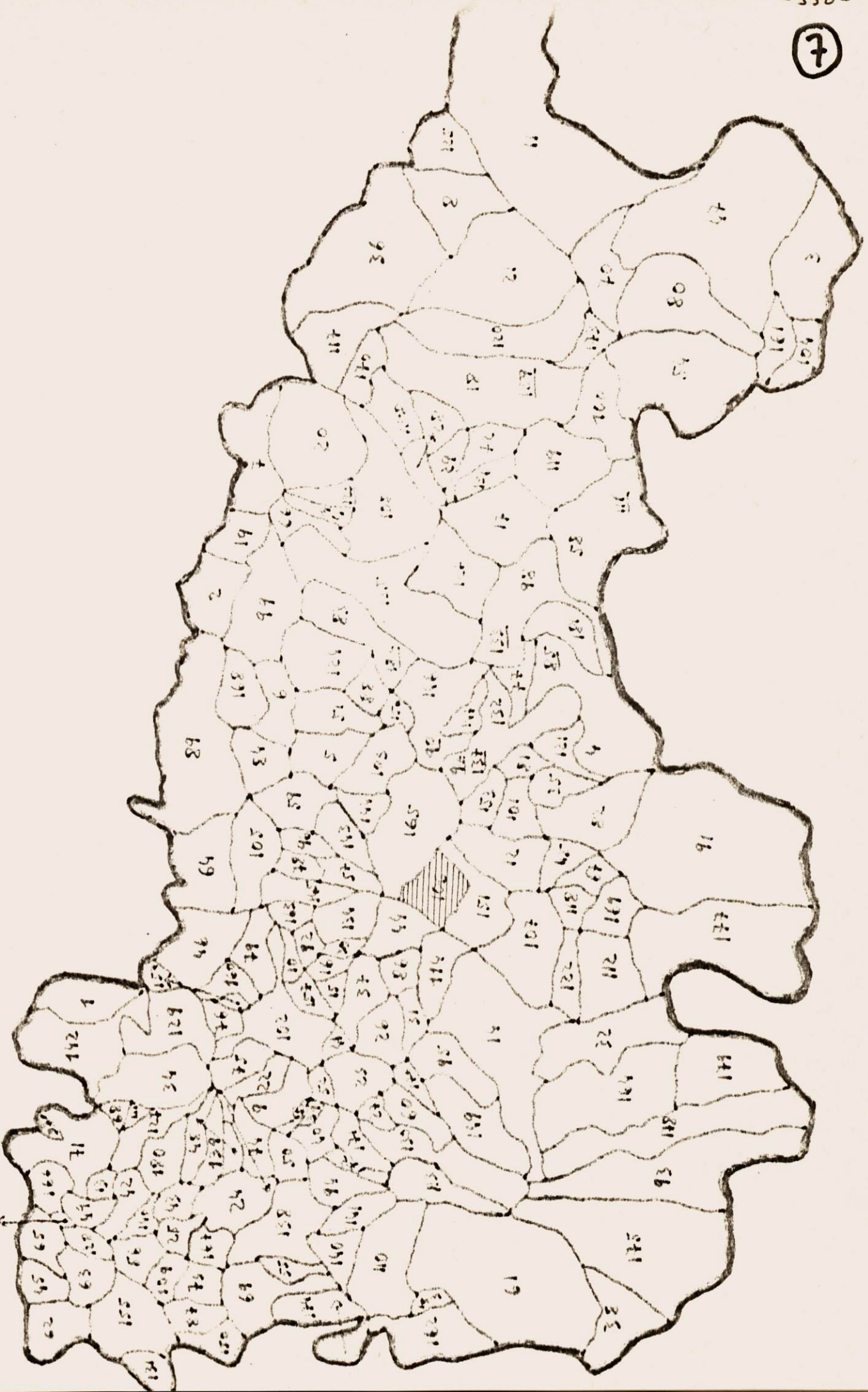


4





7



Con lo visto hasta ahora tenemos dos posibles divisiones en zonas homogéneas, la de trece grupos y la de siete.

Esta última, la de siete, va a ser la que básicamente vamos a utilizar y además la comentaremos a continuación, pero ya veremos que la de trece también nos va a ser útil para algunas subdivisiones.

De los siete grupos vemos que el grupo uno es el más numerosos, a parte de muy homogéneo, comprende fundamentalmente la comarca de Sierra - Cameros, además de un pico en el - Noroeste, pero tiene dentro de su comarca municipios que no están dentro del grupo y que deberían pertenecer a él. Es el caso de los grupos 2, 3 y 7; respecto al grupo 7 ya hemos comentado sus peculiaridades con antelación. Por lo que se refiere a los grupos 2 y 3, nos encontramos en una situación similar al tratarse de pueblos de la Sierra Alta, fuertemente ganaderos, aunque el caso no sea tan destacado como el del - grupo siete.

En conclusión, creo que no es ningún error el crear el Grupo Homogéneo I, con los clústeres 1, 2, 3 y 7.

El grupo 4, es otro grupo, fuertemente homogéneo, al cual no es preciso añadirle, ni quitarle nada, formaremos - con ~~él~~ Grupo Homogéneo II.

Los otros dos grupos restantes, que son el 5 y el 6, tienen a priori una interpretación más complicada, pero profundizando un poco en ellos, vemos, consultando las puntuaciones del análisis factorial, que mientras el grupo cinco hace más énfasis en los cultivos de cereal, etc. el grupo 6 tiene más importancia en huerta y regadío, aunque ambos discurren por las vegas de los ríos.

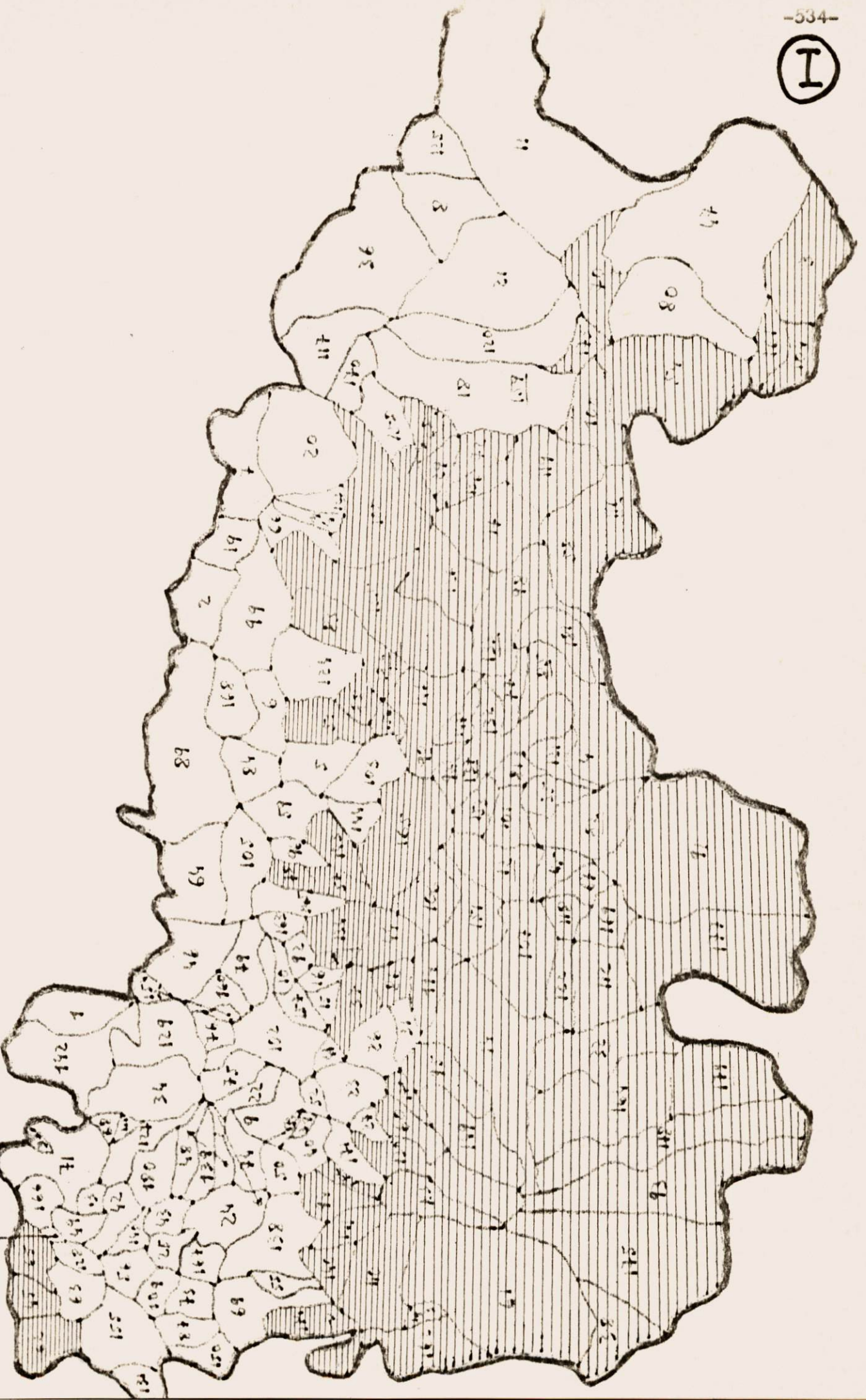
Luego tendremos dos nuevos grupos homogéneos, el III y el IV.

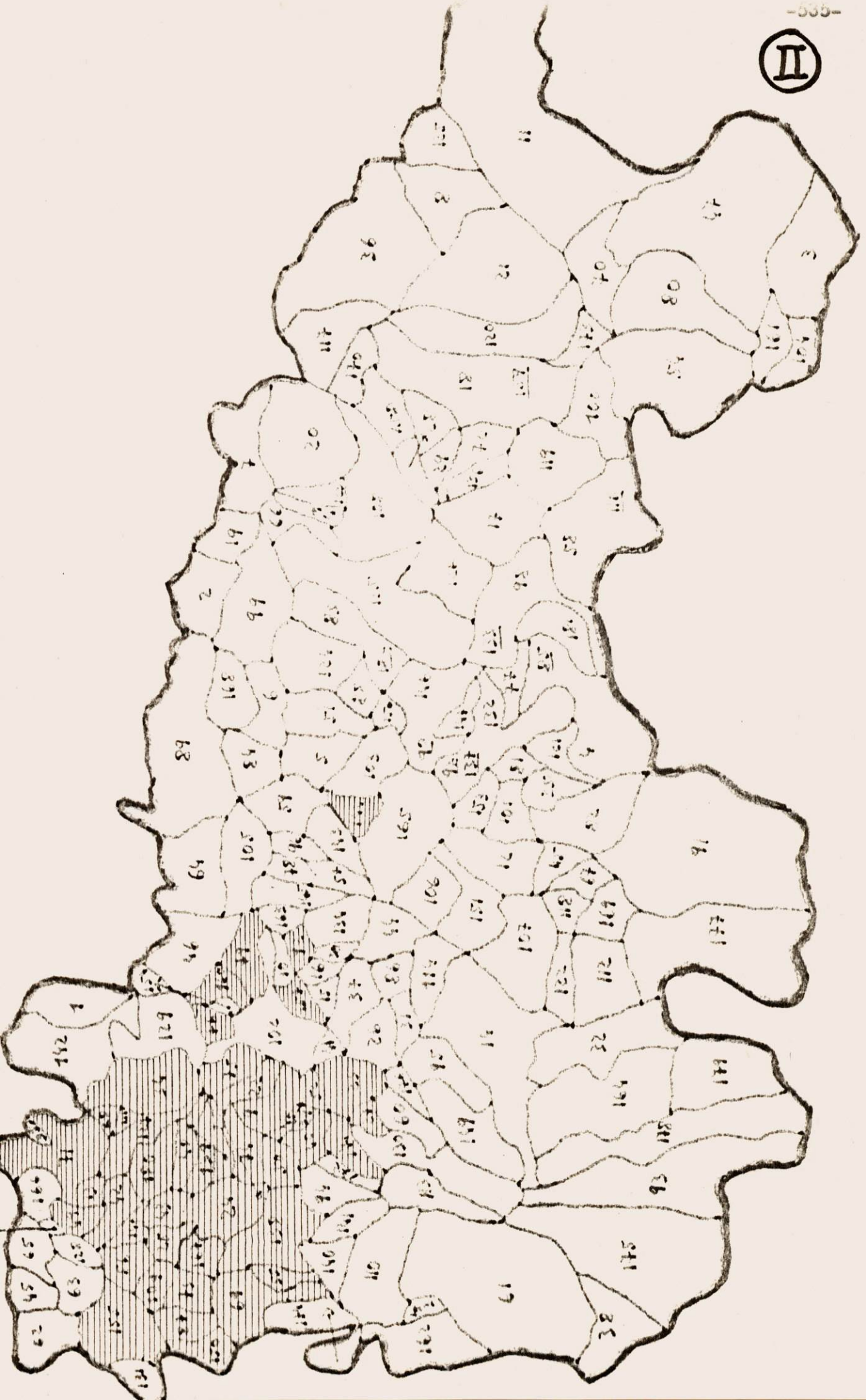
En resumen podemos decir que tenemos:

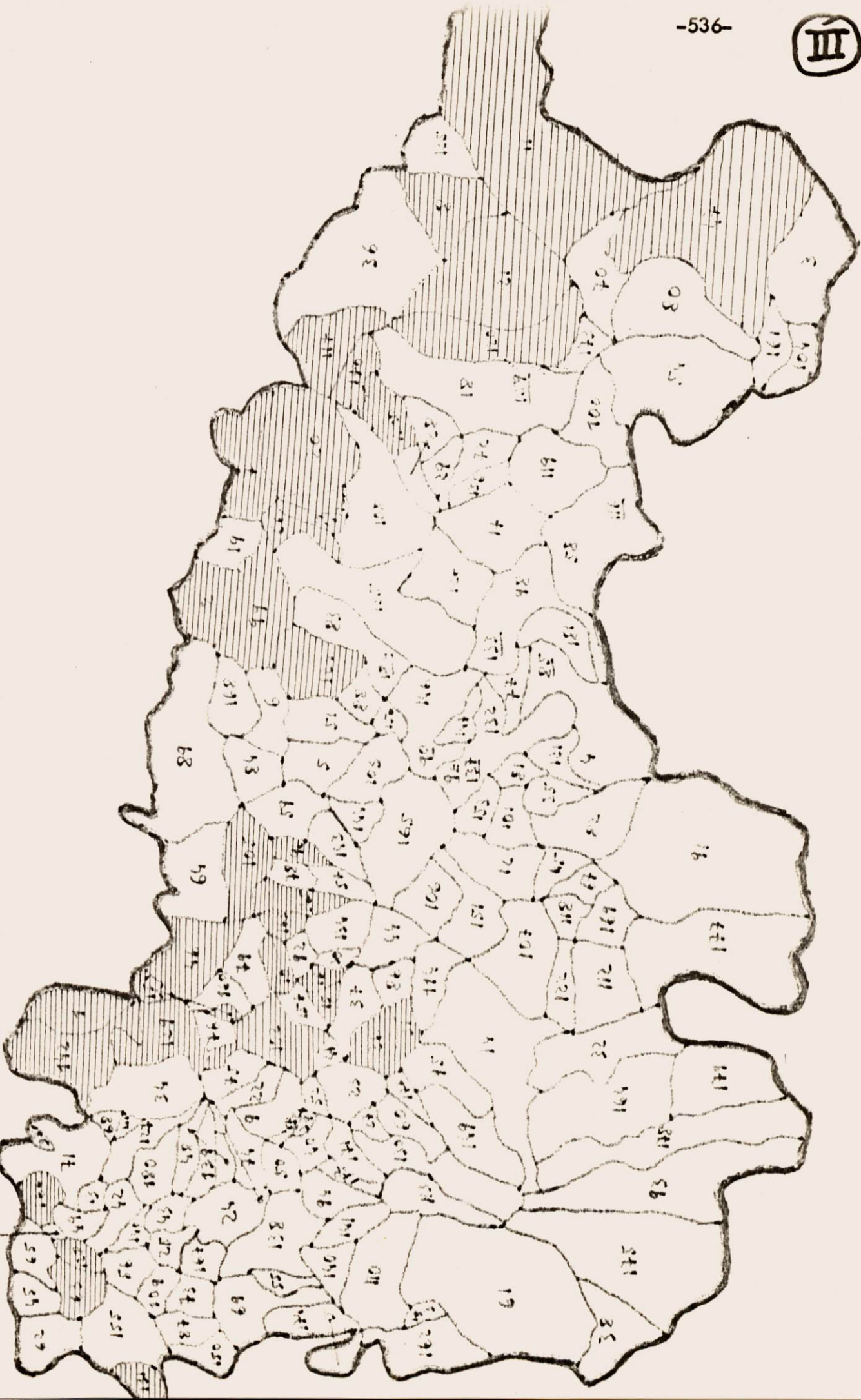
- 1) Una primera comarcalización homogénea con 13 grupos que nos ha servido de base para conseguir los otros y que además utilizaremos posteriormente.
- 2) Una segunda con 7 grupos, donde terminaría el análisis -cuantitativo propiamente dicho.
- 3) Una tercera comarcalización homogénea con 4 grupos, que hemos conseguido a base del estudio del cluster y de las variables, introduciendo ciertas apreciaciones personales.

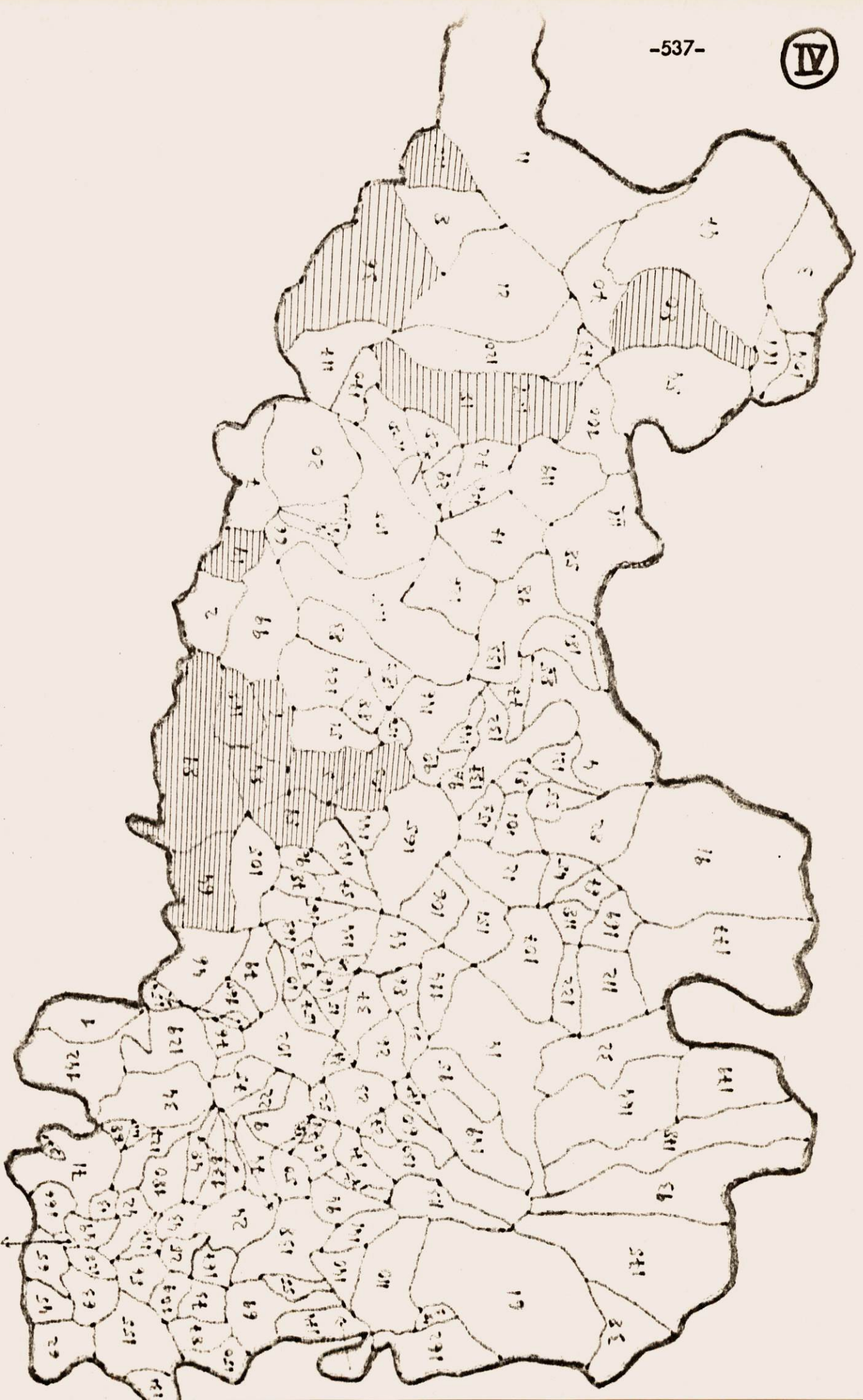
Esta última, de la cual veremos los mapas a continuación, podemos resumirla brevemente como:

(I)









Podemos observar que hemos conseguido cuatro grupos homogéneos realmente interesantes y coherentes, y además poseemos otras dos divisiones más, la de siete y la de trece, que también son muy convincentes.

Realmente los resultados no pueden ser más alentadores, pues introduciendo las variables citadas hemos conseguido unos grupos homogéneos realmente claros y que si los contrastamos con la observación directa de la provincia son reales y representativos de la situación de la economía en general y de la agricultura en particular en el área objeto de nuestro estudio.

APROXIMACION A LA NODALIDAD

A continuación, y tal como ya habíamos expuesto, vamos a intentar crear unos factores de servicios para definir los centros nodales.

Análisis factorial 3

Manteniendo las condiciones de los programas precedentes, vamos a estudiar los factores de servicios, lo denominamos Análisis Factorial Servic en las hojas de ordenador, donde tienen el nº 985 y ss.

Introducimos en este análisis factorial las variables definitorias de servicios que podemos clasificar en dos bloques bien definidos:

1) Variables que indican potencial cuantitativo de servicios que a su vez pueden ser o no ser per cápita, como es el caso:

a) Per cápita:

10 - Camas por cada 1.000 habitantes.

b) En términos absolutos:

12 - Centros de deporte.

13 - Centros escolares de E.G.B.

14 - Museos y bibliotecas.

- 15 - Hospitales.
- 49 - Cajas de Ahorro.
- 50 - Bancos.
- 53 - Licencias al por mayor.
- 54 - Licencias al por menor.
- 65 - Habitaciones hoteleras.
- 66 - Médicos.
- 67 - Recaudación del impuesto sobre espectáculos.

Unas de estas variables indican potencial real y otras son indicadores del potencial posible, como también lo son:

- 24 - Habitantes.
- 48 - Ingresos municipales.

2) Variables que indican potencial cualitativo, es decir, la existencia o no de servicios sin indicar su cuantía, como son:

- 16 - Comercio.
- 17 - Servicios comerciales.
- 18 - Servicios profesionales.
- 19 - Otros servicios.
- 20 - Equipamiento urbano.
- 22 - Otros equipamientos.
- 23 - Transportes.

Con este conjunto de variables la equivalencia con la numeración de ordenador queda:

<u>Nº VARIABLE</u>									<u>Nº ORDENADOR</u>
10	1
12	2
13	3
14	4
15	5
16	6
17	7
18	8
19	9
20	10
22	11
23	12
24	13
48	14
49	15
50	16
53	17
54	18
65	19
66	20
67	21

Con esto hemos conseguido que con un modelo de tan so lo dos factores se explica el 83'92% de la variabilidad to - tal, resultado altamente significativo y que habla de la gran correlación existente entre las variables introducidas.

Hemos de aclarar aquí que se podrían introducir otras muchas variables de indudable interés, pero hemos de comprender que a nivel municipal la información es escasa y resultaría difícil encontrar más variables de las utilizadas, que simultáneamente tuvieran una mínima fiabilidad. Por otro lado, creo que con las variables que procesamos en este estudio tenemos un conocimiento suficiente dadas las características de la provincia y el hecho incuestionable de que la mayoría de las restantes variables que hipotéticamente introduciríamos, tendrían alta correlación con las presentes.

Pasamos por lo tanto a la interpretación del análisis previo a la rotación (manteniendo la misma nomenclatura que en los anteriores).

Factor 1

Valor propio: 14'2205

Porcentaje de explicación: 67'22%

Variables muy importantes:

- 13 - Centros escolares.
- 15 - Hospitales.
- 24 - Habitantes.
- 48 - Ingresos municipales.
- 50 - Bancos.
- 53 - Licencias fiscales al por mayor.
- 54 - Licencias fiscales al por menor.
- 65 - Habitaciones en hoteles.

67 - Recaudación del impuesto sobre espectáculos.

Variables importantes:

10 - Camas por mil habitantes.

12 - Centros de deporte.

49 - Cajas de Ahorro.

66 - Médicos.

Variables secundarias:

14 - Museos y bibliotecas.

16 - Comercio.

17 - Servicios comerciales.

18 - Servicios profesionales.

19 - Otros servicios.

20 - Equipamiento urbano.

22 - Otros equipamientos.

23 - Transportes.

Variables terciarias: (Carece)

En conclusión, vemos que es un factor claro de potencialidad de servicios, pero haciendo más hincapié en los servicios medidos en forma cuantitativa. Esto es lógico dada su mayor correlación entre sí. Luego denominaremos a este factor como potencialidad de servicios, fundamentalmente en términos cuantitativos.

Este es evidentemente el factor que más explica y el que mejor discrimina a la población.

Factor 2

Valor propio: 3'4034

Porcentaje de explicación: 16'21%

Variables muy importantes: (Carece)

Variables importantes : (Carece)

Variables secundarias:

- 16 - Comercio.
- 17 - Servicios comerciales.
- 18 - Servicios profesionales.
- 19 - Otros servicios.
- 20 - Equipamiento urbano.
- 22 - Otros equipamientos.
- 23 - Transportes.

Variables terciarias:

- 15 - Hospitales (negativa).
- 53 - Licencias fiscales al por mayor (negativa).
- 66 - Médicos (negativa).

Este factor en principio indica el potencial de servi
cios en forma cualitativa y tenemos el caso curioso de que -
las variables terciarias dan negativas. Esto puede parecer en
principio una contradicción que esperamos desaparezca al efec
tuar la rotación. De todas formas tampoco es tan contradicto
rio y podemos imaginarnos la situación en el espacio n-dimen
sional donde se forman los factores.

Factores rotados:

Factor 1

Capacidad explicativa: 11'0733

Porcentaje de explicación: 52'73%

Variables muy importantes:

- 13 - Centros escolares.
- 15 - Hospitales.
- 24 - Habitantes.
- 48 - Ingresos municipales.
- 53 - Licencias fiscales al por mayor.
- 54 - Licencias fiscales al por menor.
- 66 - Médicos.
- 67 - Recaudación del impuesto sobre espectáculos.

Variables importantes:

- 10 - Camas.
- 12 - Centros de deporte.
- 50 - Bancos.
- 65 - Habitaciones en hoteles.

Variables secundarias:

- 49 - Cajas de Ahorro.

Variables terciarias:

- 14 - Museos y bibliotecas.

Este es un factor clarísimo de potencial cuantitativo de servicios, que no merece mayor comentario.

Factor 2

Capacidad explicativa: 6'5499

Porcentaje de explicación: 31'19%

Variables muy importantes: (Carece)

Variables importantes:

- 16 - Comercio.
- 17 - Servicios Comerciales.
- 18 - Servicios profesionales.
- 19 - Otros servicios.
- 20 - Equipamiento urbano.
- 22 - Otros equipamientos.
- 23 - Transportes.

Variables secundarias:

- 14 - Museos y bibliotecas.
- 49 - Cajas de Ahorro.

Variables terciarias:

- 12 - Centros de deporte.
- 13 - Centros escolares.
- 50 - Bancos.
- 65 - Habitaciones en hoteles.

Evidentemente este factor se caracteriza por definir el potencial de servicios, pero con una preponderancia de las

variables cualitativas.

Luego en definitiva tenemos dos factores que resumen nuestra información:

- Factor 1 - "Potencial cuantitativo de servicios". Nos dice qué cantidad de servicios ofrece un municipio determinado.
- Factor 2 - "Potencial cualitativo de servicios". Nos dice si existen o no servicios en un municipio, pero no nos habla de su cantidad.

Las puntuaciones obtenidas por el ordenador son muy interesantes y estudiaremos su utilización con posterioridad.

En las hojas de ordenador nº1004 y ss. tenemos el mismo análisis factorial que hemos estudiado pero con la restricción de solo poder sacar un factor, esto se debe a nuestro interés en tener las puntuaciones por municipios con un solo - factor que aglutine toda la información. Como es natural coinside con el valor del factor uno antes de la rotación en el estudio anterior, por lo que no lo interpretaremos, pero las puntuaciones si pueden sernos de utilidad posteriormente.

Quiero hacer ahora una pequeña aclaración sobre la aparición de asteriscos en la desviación típica de la variable - 48 (14 según la numeración de ordenador).

Estos asteriscos, en la nomenclatura informática, son indicativos de infinito, eso quiere decir que la desviación típica de la variable 48 no cabe en el campo de salida que tiene asignado según el programa de ordenador.

Esto no debe preocuparnos puea a la hora de operar el ordenador lo tiene en cuenta en su valor real, luego los cálculos no se verán afectados por esta incidencia.

CLASIFICACION DE LOS MUNICIPIOS SEGUN SU POTENCIAL DE SERVICIOS.

A la vista de las puntuaciones del análisis factorial anterior (tanto con dos factores, como con uno). Vamos a calcular el potencial de servicios de cada municipio.

Establecemos las siguientes categorías:

- Municipio de 1^ª - Puntuación mayor que 10
- Municipio de 2^ª - Puntuación mayor que 5
- Municipio de 3^ª - Puntuación mayor que 2
- Municipio de 4^ª - Puntuación mayor que 1

(Esta escala ha salido de la contrastación de los resultados de nuestro análisis con los datos reales que poseemos).

En el factor uno, tenemos : (Potencial cuantitativo de servicios).

Municipios de 1^ª categoría:

- 88 - Logroño.

Municipios de 2^ª categoría:

- 36 - Calahorra.

Municipios de 3^ª categoría:

- 11 - Alfaro.

- 18 - Arnedo.
- 71 - Haro.
- 132 - Santo Domingo.

Municipios de 4ª categoría:

- 61 - Ezcaray.
- 99 - Nájera.
- 145 - Torrecilla en Cameros.

En el factor dos, tenemos: (Potencial cualitativo de servicios).

Municipios de 1ª categoría:

- 36 - Calahorra.
- 88 - Logroño.

Municipios de 2ª categoría:

- 8 - Aldeanueva de Ebro.
- 11 - Alfaro.
- 18 - Arnedo.
- 47 - Cervera.
- 71 - Haro.
- 99 - Nájera.
- 132 - Santo Domingo.

Municipios de 3ª categoría: (Solo citar) 5, 21, 42, 46, 59, 61, 64, 96, 113, 121, 125, 163.

Municipios de 4ª categoría: (Solo citar), 3, 6, 7, 14, 20, 26, 34, 38, 69, 84, 100, 102, 109, 116, 120, 136, 145, 161, 170, 171.

Ahora lo que podemos hacer es efectuar un análisis si milar con el factor único y sus puntuaciones. Recordemos que este factor nos indica el potencial total de servicios, ha - ciendo especial hincapié en las variables cuantitativas, cosa lógica dada la imprecisión de las cualitativas.

Factor único:

Municipios de 1ª categoría: "Centros de servicios de 1ª"

36 - Calahorra.

88 - Logroño.

Municipios de 2ª categoría: "Centros de servicios de 2ª"

11 - Alfaro.

18 - Arnedo.

71 - Haro.

99 - Nájera.

132 - Santo Domingo.

Municipios de 3ª categoría: "Centros de servicios de 3ª"

5 - Albelda.

8 - Aldeanueva.

47 - Cervera.

61 - Ezcaray.

64 - Fuenmayor.

- 121 - Rincón de Soto.
- 145 - Torrecilla en Cameros.

Municipios de 4ª categoría: "Centros de servicios de 4ª"

- 21 - Autol.
- 42 - Casalarreina.
- 46 - Cenicero.
- 59 - Entrena.
- 96 - Murillo.
- 102 - Navarrete.
- 113 - Pradejón.
- 125 - San Asensio.
- 163 - El Villar de Arnedo.

En conclusión, lo que hemos hecho es clasificar a los centros de servicios en función del factor único que es el - que más varianza explica.

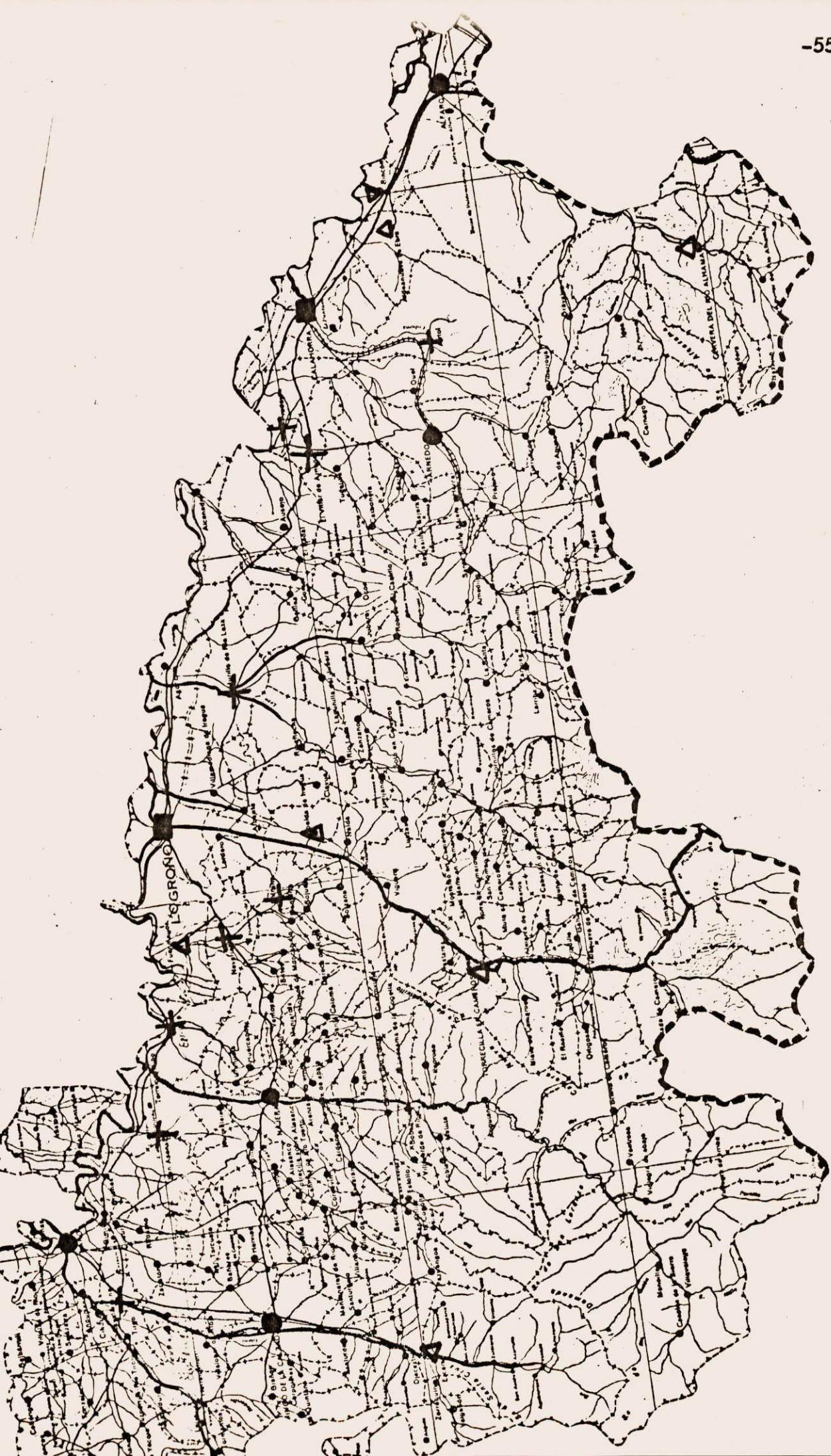
Las clasificaciones según el factor uno y el dos del modelo de dos factores nos pueden ser útiles en la comarcalización final, pero la idea básica partirá de la clasificación mediante el factor único.

Se puede discutir este criterio diciendo que el factor único se parece más al factor uno del modelo de dos factores que al dos y que éste último por hablar de servicios en sentido cualitativo (su existencia o no) es más interesante para la creación de nodos.

Pero esto no es cierto pues, en primer lugar, las variables fundamentales del factor dos son bastante imprecisas y en segundo lugar, y por poner un ejemplo, a la hora de considerar un nodo no solo es interesante que tenga médico, sino además - cuantos tiene, para poder elegir entre especialidades, etc.

A continuación vemos un mapa con los centros de servicios según el siguiente código:

- Centros de servicios de 1ª.
- Centros de servicios de 2ª.
- △ Centros de servicios de 3ª.
- + Centros de servicios de 4ª.



Análisis Factorial 4

En este análisis factorial vamos a intentar unir las variables industriales y comerciales, pretendemos, en consecuencia, medir el potencial de actividad mercantil (industrial y comercial) de cada municipio, con las variables que poseemos.

El análisis factorial correspondiente se encuentra en las hojas de ordenador nº1013 y ss., bajo el nombre de Análisis Factorial Indust.

La clave de equivalencia del número real de las variables con la nomenclatura de ordenador es:

<u>Nº VARIABLE</u>	<u>Nº ORDENADOR</u>
49	1
50	2
53	3
54	4
55	5
56	6
57	7
58	8
59	9
60	10
61	11
62	12

63	13
64	14
65	15
66	16
67	17

El resultado del modelo, con las mismas condiciones de los anteriores, es un único factor que explica el 84'31% de la varianza total.

Por lo tanto este único factor es altamente explicativo y al ser uno solo, no tiene sentido la rotación.

El que un solo factor explique tan bien el conjunto de variables, nos da a entender la alta correlación existente entre el conjunto de actividades mercantiles estudiadas.

Veamos la interpretación (manteniendo la nomenclatura de los anteriores):

Factor Único

Valor propio: 14'3325

Porcentaje de explicación: 84'31%

Variables muy importantes:

- 50 - Bancos.
- 53 - Licencias fiscales al por mayor.
- 54 - Licencias fiscales al por menor.

- 56 - L. F. Rama 2 Fabricación.
- 57 - L. F. Rama 3 Fabricación.
- 59 - L. F. Rama 5 Fabricación.
- 61 - L. F. Rama 7 Fabricación.
- 63 - L. F. Rama 9 Fabricación.
- 64 - L. F. Artesanía.
- 65 - Habitaciones en hoteles.
- 66 - Médicos.
- 67 - Recaudación impuesto sobre espectáculos.

Variables importantes:

- 49 - Cajas de Ahorro.
- 60 - L. F. Rama 6 Fabricación.

Variables secundarias:

- 55 - L. F. Rama 1 Fabricación.
- 58 - L. F. Rama 4 Fabricación.
- 62 - L. F. Rama 8 Fabricación.

Variables terciarias: (Carece)

El factor, sin más explicaciones, se ve que es clara
mente un factor industrial y comercial.

CLASIFICACION DE LOS MUNICIPIOS SEGUN SU POTENCIAL INDUSTRIAL
Y COMERCIAL.

A continuación vamos a clasificar los municipios en base a las puntuaciones obtenidas en el factor industrial y comercial (I.C.)

La clave, basada en la clasificación de centros de servicios, pero algo más ampliada, será:

- Municipios I.C. de 1ª - Puntuación mayor que 10
- Municipios I.C. de 2ª - Puntuación mayor que 5
- Municipios I.C. de 3ª - Puntuación mayor que 3
- Municipios I.C. de 4ª - Puntuación mayor que 1
- Municipios I.C. de 5ª - Puntuación mayor que 0'5

Municipios I.C. de 1ª: "Centros I.C. de 1ª"

- 88 - Logroño.

Municipios I.C. de 2ª: "Centros I.C. de 2ª"

- 18 - Arnedo.
- 36 - Calahorra.
- 71 - Haro.

Municipios I.C. de 3ª: "Centros I.C. de 3ª"

- 11 - Alfaro.
- 99 - Nájera.
- 132 - Santo Domingo.






Municipios I.C. de 4ª: "Centros I.C. de 4ª"

- 47 - Cervera.
- 61 - Ezcaray.
- 64 - Fuenmayor.
- 102 - Navarrete.

Municipios I.C. de 5ª: "Centros I.C. de 5ª"

- 5 - Albelda.
- 14 - Anguiano.
- 21 - Autol.
- 26 - Baños de Rio Tobia.
- 46 - Cenicero.
- 121 - Rincón de Soto.
- 161 - Villamediana.

Con lo que tenemos los centros industriales y comerciales, que los podemos ver en el siguiente mapa con la clave:

-  Centro I.C. de 1ª.
-  Centro I.C. de 2ª.
-  Centro I.C. de 3ª.
-  Centro I.C. de 4ª.
-  Centro I.C. de 5ª.



ANALISIS DE LA CALIDAD DE VIDA

Análisis Factorial 5

Una vez que tenemos la comarcalización homogénea y los estudios correspondientes a centros de servicios e industriales y comerciales, pasamos a realizar un estudio complementario de municipios insuficientemente dotados para ser considerados como habitables en la década de los setenta.

Hemos denominado a este estudio "Análisis de la Calidad de Vida", y lo que pretende es el estudio de las variables indicativas de la citada calidad de vida.

Este estudio lo abordamos mediante un análisis factorial que nos resume la información utilizada, basada ésta en el banco de datos que poseemos.

La utilización de los datos que obtengamos, nos permitirá un conocimiento más exhaustivo de las áreas deprimidas, que deberán tener una consideración especial a la hora de la comarcalización, para intentar que las futuras planificaciones se ocupen de manera especial de las mencionadas áreas.

El análisis factorial correspondiente se encuentra en las hojas de ordenador nº 1026 y ss. bajo la denominación de Análisis Factorial Calida. y manteniendo las condiciones de programación de los anteriores.

Las variables utilizadas y su clave de equivalencia para la numeración de ordenador es:

<u>Nº VARIABLE</u>									<u>Nº ORDENADOR</u>
5	1
6	2
7	3
8	4
9	5
10	6
11	7
12	8
13	9
14	10
15	11
16	12
17	13
18	14
19	15
20	16
21	17
22	18
23	19
51	20
52	21

El modelo factorial resultante, explica con cuatro factores el 67'23% de la variabilidad, lo que no representa un

resultado excesivamente bueno, pero tengamos en cuenta la imprecisión de muchas de las variables utilizadas y el hecho - de que algunos datos nos inducen a dudar de la fiabilidad de las mismas.

Esta falta de fiabilidad afecta principalmente a las primeras variables de este análisis (que no han sido introducidas en estudios anteriores), dado que contrastada su información con otras fuentes, los resultados han sido dispares.

Pero el resto de fuentes solo ofrecen informaciones - parciales por lo que ante la carencia de información solvente y completa, hemos de atenernos a la que poseemos, que por - otro lado creo que es aceptablemente buena.

Pasamos a analizar los factores antes de la rotación (manteniendo la nomenclatura de los anteriores):

Factor 1

Valor propio: 8'5254

Porcentaje de explicación: 40'60%

Variables muy importantes: (Carece)

Variables importantes:

- 12 - Centros de deporte.
- 16 - Comercio.
- 17 - Servicios comerciales.

- 18 - Servicios profesionales.
- 19 - Otros servicios.
- 20 - Equipamiento urbano.
- 22 - Otros equipamientos.
- 23 - Transportes.

Variables secundarias:

- 6 - Agua corriente.
- 10 - Camas.
- 13 - Centros escolares.
- 14 - Museos y bibliotecas.
- 15 - Hospitales.
- 21, - Equipamiento religioso.
- 51 - Crecimiento demográfico (60 - 70).

Variables terciarias:

- 5 - Pavimentación.
- 7 - Alcantarillado.
- 8 - Alumbrado en viviendas.

Este factor es un indicador de la calidad de vida, pero haciendo hincapié en la potencia cualitativa de servicios (que aquí es la fundamental para medir la habitabilidad).

Pero hay gran cantidad de variables algo revueltas, por lo que esperamos que al rotar la interpretación queda más clara.

De todas formas es un factor bastante explicativo, el que más del modelo, dado que es el primero y, por lo tanto, el que mejor discrimina a la población, en nuestro caso a los municipios.

Factor 2

Valor propio: 2'7905

Porcentaje de explicación: 13'29%

VARIABLES MUY IMPORTANTES: (Carece)

VARIABLES IMPORTANTES: (Carece)

VARIABLES SECUNDARIAS:

- 6 - Agua corriente. (negativa).
- 7 - Alcantarillado (negativa).
- 8 - Alumbrado en viviendas (negativa).
- 10 - Camas.
- 13 - Centros escolares.
- 15 - Hospitales.

VARIABLES TERCIARIAS:

- 5 - Pavimentación (negativa).
- 9 - Alumbrado de vías públicas (negativa).
- 12 - Centros de deporte.
- 21 - Equipamiento religioso.

Este factor es de difícil interpretación y esperamos que al rotar sea más asequible.

Tal como está ahora lo hemos de definir como factor que asocia la carencia de infraestructura básica (alumbrado, agua corriente, etc.) con la posesión de otros servicios (hospitales, escuelas, etc).

Factor 3

Valor propio: 1'5403

Porcentaje de explicación: 7'33%

Variables muy importantes: (Carece)

Variables importantes: (Carece)

Variables secundarias: (Carece)

Variables terciarias:

- 6 - Agua corriente (negativa).
- 7 - Alcantarillado (negativa).
- 8 - Alumbrado en viviendas (negativa).
- 10 - Camas (negativa).
- 15 - Hospitales (negativa).
- 16 - Comercio.

Nos encontramos nuevamente ante un caso difícil pues este factor asocia la carencia de cierta infraestructura y ciertos servicios con la situación de tener comercio. Hemos de volver a decir que la rotación tendrá que aclararnos esta situación.

Factor 4

Valor propio: 1'2626

Porcentaje de explicación: 6'01%

VARIABLES MUY IMPORTANTES: (Carece)

VARIABLES IMPORTANTES: (Carece)

VARIABLES SECUNDARIAS:

52 - Crecimiento demográfico (70 - 75).

VARIABLES TERCIARIAS:

7 - Alcantarillado (negativa).

9 - Alumbrado de vías públicas.

11 - Estudiantes de E.G.B. por cada 100 habitantes (negativa).

51 - Crecimiento demográfico (60 - 70).

Este factor, algo más claro, lo hemos de asociar al dinamismo demográfico, fundamentalmente en el quinquenio 70-75. Pero por lo demás es confuso.

En conclusión hemos de pasar a analizar los resultados tras la rotación, dado que los factores antes de rotar son en su mayoría algo contradictorios.

Factores rotados:

Factor 1

Capacidad explicativa: 3'948

Porcentaje de explicación: 18'80%

Variables muy importantes:

15 - Hospitales.

Variables importantes:

10 - Camas.

12 - Centros de deporte.

13 - Centros escolares.

Variables secundarias:

21 - Equipamiento religioso.

Variables terciarias:

14 - Museos y bibliotecas.

Este factor fundamentalmente indica la dotación médica y también la deportiva y educativa.

Factor 2

Capacidad explicativa: 2'9019

Porcentaje de explicación: 12'39%

Variables muy importantes: (Carece)

Variables importantes:

6 - Agua corriente (negativa).

7 - Alcantarillado (negativa).

Variables secundarias:

- 5 - Pavimentación (negativa).

Variables terciarias:

- 8 - Alumbrado en viviendas (negativa).

Este factor es también bastante claro y hace referencia a la carencia de infraestructura básica (agua, alcantarillado, etc.).

Factor 3

Capacidad explicativa: 5'8632

Porcentaje de explicación: 27'92%

Variables muy importantes: (Carece)

Variables importantes:

- 16 - Comercio.
- 17 - Servicios comerciales.
- 18 - Servicios profesionales.
- 19 - Otros servicios.
- 20 - Equipamiento urbano.
- 23 - Transportes.

Variables secundarias:

- 14 - Museos y bibliotecas.
- 22 - Otros equipamientos.
- 51 - Crecimiento demográfico (60 - 70).

Variables terciarias:

- 12 - Centros de deportes.
- 13 - Centros escolares.

Vemos que este factor, que es el más explicativo tras la rotación, hace especial referencia al potencial cualitativo de servicios, es decir al hecho de poseer o no los servicios, que es lo interesante de cara a la calidad de vida. Podemos denominarlo genéricamente factor servicios.

Factor 4

Capacidad explicativa: 1'7052

Porcentaje de explicación: 8'12%

Variables muy importantes: (Carece)

Variables importantes:

- 52 - Crecimiento demográfico (70 - 75).

Variables secundarias:

- 8 - Alumbrado en viviendas.
- 9 - Alumbrado en vías públicas.
- 51 - Crecimiento demográfico (60 - 70).

Variables terciarias: (Carece)

Este factor lo hemos de asociar al crecimiento demográfico, sobre todo en el quinquenio 70 - 75 y al alumbrado.

Vemos que tras la rotación los factores han quedado bastante coherentes y que nos pueden ser útiles para saber la calidad de vida de los municipios, para ello utilizaremos las puntuaciones.

MUNICIPIOS CON CALIDAD DE VIDA MUY BAJA

Para su obtención, el criterio ha sido buscar los muni cipios que estaban muy mal dotados en tres o cuatro de los - factores. Para decir que estaba muy mal dotado, el criterio ha sido tener una puntuación inferior a -1 en los factores - uno, tres y cuatro (que tienen sentido positivo) y superior a 1 en el factor dos (que tiene sentido negativo).

Municipios mal dotados en tres factores:

- 4 - Ajamil.
- 29 - Bergasillas Bajera.
- 35 - Cabezón de Cameros.
- 44 - Castroviejo.
- 101 - Navajún.
- 110 - Pazuengos.
- 117 - Rabanera.
- 122 - Robres del Castillo.
- 130 - Santa Engracia de Jubera.
- 141 - Terroba.
- 147 - Torre en Cameros.
- 148 - Torremontalvo.
- 150 - Trevijano.
- 154 - Valdemadera.
- 166 - Villarroya.
- 173 - Zarratón.
- 174 - Zarzosa.

Municipios mal dotados en los cuatro factores:

87 - Leza de Rio Leza.

Veamos a continuación un mapa donde se señala con X los municipios citados y con (X) el 87.



A la vista del mapa podemos observar como la mayoría de los municipios citados se encuentran en los llamados Cameros Viejos y, en general, en las zonas montañosas salvo raras excepciones.

Todo esto habrá que tenerlo en cuenta a la hora de planificar el futuro de la provincia y en nuestro caso al comarcalizar hemos de procurar que estas zonas insuficientemente dotadas se beneficien de algún tipo de centro nodal que propondremos crear.

Pero toda esta problemática la trataremos en el siguiente capítulo. Para terminar, baste reseñar aquí que como comprobación de nuestros resultados, hemos consultado a la Diputación, donde nos han informado de la similitud de nuestras conclusiones con el estudio de zonas deprimidas de la provincia realizado por la citada entidad.

VI.- PROPUESTA DEFINITIVA DE COMARCALIZACION.

A la vista de lo hasta ahora estudiado nos encontramos capacitados para proponer nuestra comarcalización que compararemos con otras realizadas por diversas entidades.

Para nuestra comarcalización vamos a tener en cuenta tres objetivos fundamentales (sin que su orden de enumeración indique en modo alguno prioridad).

a) Agrupar zonas homogéneas con nuestro criterio Agrícola-Ganadero. Creo que no es preciso repetir aquí lo ya ex-puesto sobre la validez de este criterio, pero lo que si quiero recalcar es que nunca lo consideraremos como un criterio - absoluto, sino que por el contrario, habrá que contrastarlo - con la consecución de los dos objetivos siguientes.

b) Realizar la agrupación alrededor de centros de servicios de la mayor categoría posible, fijándonos también en el potencial industrial y comercial de esos centros, lo que puede ser la base del asentamiento de futuros servicios a la vez que son puntos de demanda de mano de obra.

c) Tener en cuenta los flujos que se produzcan entre los distintos municipios, lo que indicará la fuerza de atracción nodal que ejerce un centro de servicios sobre su entorno.

Estos flujos los podemos medir de dos maneras posibles en base a los datos que poseemos, y así veremos por un lado a donde se acude desde cada municipio para satisfacer los servicios, según los datos de la Diputación. Por otro lado, estudiaremos los mapas de intensidad media diaria (I.M.D.) del Ministerio de Obras Públicas, de donde podremos extraer información, tanto de la cantidad de relación entre los municipios (de la que la I.M.D. es un buen estimador), como de la posibilidad de esa relación, según si existe vía de comunicación practicable o no.

Un objetivo complementario a estos tres será la agrupación de zonas deprimidas alrededor de centros de servicios existentes o potenciables, de cara a reducir el creciente des poblamiento y empobrecimiento de las mismas.

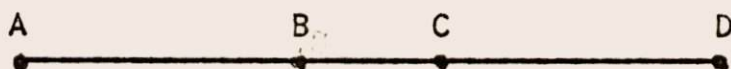
El lector podrá apreciar la dificultad intrínseca al presente planteamiento, donde con facilidad pueden entrar en conflicto dos o más objetivos, pero vamos a intentar ir las conciliando, de manera que consigamos el resultado apetecido.

En primer lugar, y de acuerdo con los objetivos dos y tres y a la vista del sistema de comunicaciones de la provincia, en el cual tanto su geografía física como la humana han tenido mucho que ver, queda grandemente dificultada la relación entre las partes orientales y occidentales de la sierra, lo que hace que quede dividida esta zona homogénea en diversas subzonas, lo que pone en contradicción el objetivo uno con los dos restantes.

Dificultades como esta nos vamos a encontrar a menudo lo que nos debe hacer ser más flexibles en nuestros planteamientos, e intentar coordinar en vez de contraponer los objetivos.

Un caso similar nos va a suceder entre el objetivo tres y el dos, pues hay municipios, que estando cerca de un centro de servicio, se dirigen a otro de mayor entidad por las razones que vemos a continuación:

Supongamos que el municipio A es un centro de servicios de máxima categoría y que B, es un centro de categoría suficiente, siendo C y D dos municipios, carentes de casi todos los servicios y cuya situación geográfica es:



Resulta lógico pensar que los habitantes de C se dirigirán a B para satisfacer sus necesidades, si no pueden hacerlo en el propio C, pero los habitantes de D, una vez que han de recorrer un largo camino, fácilmente seguirán hasta A donde los servicios ofrecidos son de mejor calidad que en B.

Por poner un ejemplo a nivel nacional, podemos suponer que la Universidad de Valladolid atrae a la gente de la región castellana, pero una persona de La Coruña, si tiene que escoger entre la Universidad de Valladolid y la de Madrid, fácilmente se irá a ésta última, aún estando más lejos, dado

que una vez que se desplaza se irá a la que supone más comple
ta.

Este hecho, bastante evidente, pone en contradicción el objetivo de adaptarnos a los flujos reales y el mantener una lógica con el objetivo dos, es decir, agrupar alrededor de centros de servicios. Según el objetivo tres, el municipio D se uniría al A, según el dos a B, y en este caso esta solución parece la más lógica.

Finalmente habrá casos en los cuales no exista centro de servicios, ni tampoco flujos de importancia entre municipios contiguos y sin embargo sea conveniente el crear una comarca y un centro dentro de ella que será preciso potenciar.

En conclusión, y siguiendo un criterio de flexibilidad vamos a intentar coordinar los distintos objetivos.

La primera limitación, ya comentada, es la estructura de comunicaciones de la provincia, y a ella nos tendremos que adaptar.

Esta red la podemos dividir en dos grandes grupos:

- Ferrocarril: La provincia solo cuenta con el tramo Miranda - Castejón, que la atraviesa de Este a Oeste por el Norte, de forma paralela al Ebro, pasando por Haro, Logroño, Calahorra y Alfaro, como poblaciones más importantes.

- Carretera: Como puede verse en el mapa que se adjunta, las comunicaciones son fáciles a lo largo del Ebro, pero la entrada en la sierra (parte Sur de la provincia) ha de hacerse a lo largo de los valles de los rios, siendo muy difícil y muchas veces imposible el paso de valle a valle.

Esta configuración de las comunicaciones nos obliga a establecer que las comarcas han de ir de Norte a Sur, y no de Este a Oeste que es donde se da la mayor homogeneidad.

En la actualidad hay un proyecto en ejecución que trata de abrir una carretera que una la sierra en la dirección Este - Oeste, pero si bien han comenzado algunas obras del mismo y se están terminando otras, no es todavía una realidad y se enfrenta con dos dificultades importantes:

Por un lado lo difícil del trazado y lo agreste de la geografía por la que ha de pasar la carretera, lo que se traduce en mayores costes de ejecución de la obra y en el peligro de intransitabilidad durante parte del invierno.

En segundo lugar, dado lo despoblado de la zona, el rendimiento económico de la citada vía de comunicación sería bastante bajo. Indudablemente no nos debemos dejar guiar exclusivamente por criterios económicos y a los riojanos nos alegraría mucho que se realizara la citada carretera, pero si somos realistas hemos de pensar que tardará en terminarse si es que se termina completamente.

Ante esta situación, lo que proponemos es un orden jerarquizado de:

- Zonas.
- Comarcas.
- Subcomarcas.

Donde sea al nivel de las comarcas y más aún al de sub comarcas donde la homogeneidad sea mayor.

Comencemos por el establecimiento de las zonas. Para ello hemos de encontrar unos centros de servicios de mucha categoría como son:

Centros de 1ª:

- Logroño.
- Calahorra.

Centros de 2ª:

- Haro.
- Arnedo.
- Alfaro.
- Santo Domingo.
- Nájera.

Evidentemente, si a cada centro le asignamos una zona, tendríamos demasiadas zonas, luego vamos a complementar la información actual con la categoría industrial y comercial de los centros citados.

- | | |
|-----------------|----|
| - Logroño | 1ª |
| - Calahorra | 2ª |
| - Haro | 2ª |
| - Arnedo | 2ª |
| - Alfaro | 3ª |
| - Santo Domingo | 3ª |
| - Nájera | 3ª |

Luego tras la primera selección nos quedan cuatro centros: Logroño, Calahorra, Haro y Arnedo. Dada la proximidad de Calahorra y Arnedo y la mayor importancia del primer municipio, rechazamos Arnedo y nos quedan tres centros de zona:

- Logroño.
- Calahorra.
- Haro.

Tendremos en consecuencia:

- Zona Oeste: Centro Haro.
- Zona Centro: Centro Logroño.
- Zona Este: Centro Calahorra.

Ahora ya en cada zona, y a la vista de los objetivos mencionados al comienzo de este capítulo, trataremos de crear las comarcas y subcomarcas correspondientes.

El criterio ha sido buscar en cada zona los centros de servicios existentes (C.S.) y los industriales y comerciales (I.C.), ver el área de influencia de cada centro y crear alrededor del mismo una comarca, para calcular este área de influencia nos hemos fijado fundamentalmente en los flujos (de servicios y por carretera) y en los casos dudosos en la homogeneidad.

Una vez que se crea la comarca, si esta no resulta homogénea (como sucede en muchos casos), se crean subcomarcas

homogéneas, atendiendo fundamentalmente a los estudios del - cluster análisis, pero sin descuidar las comunicaciones, que deben ser fáciles.

Estas subcomarcas, deben ser dotadas de centros de rango inferior, que si no existen, por lo menos nos encontraremos con potencialidad para su creación.

El objetivo de este sistema es doble. Por un lado pretendemos que la comarcalización sea útil de cara a crear centros de servicios (escuelas, asistencia sanitaria, servicios varios, etc.), cerca de los consumidores y, por otro lado, - para la planificación; con nuestro sistema mixto (homogeneidad -nodalidad), podemos intentar conseguir simultáneamente ambos objetivos.

La necesidad de los centros nodales de cara a la creación y potenciación de centros de servicios parece evidente.

En cuanto a la necesidad de cierta homogeneidad, también la considero necesaria, pues si no, las zonas pobres, incluidas en comarcas ricas, donde no se estudien sus problemas específicos, seguirán siempre siendo la zona pobre de la comarca. Sin embargo al unir estas zonas en subcomarcas, tienen la posibilidad de afrontar sus problemas y exigir una planificación más acorde con sus necesidades.

Por otro lado la potenciación de centros, actualmente inexistentes, en zonas deprimidas colaborará a frenar el desde

poblamiento y ~~re~~esión de estas zonas.

Pasamos a continuación a la comarcalización de la zona Oeste, cuyo centro es Haro.

La comarcalización propuesta está en el mapa de la página siguiente que explicaremos a continuación.

Zona Oeste

Centro: Haro.

Esta zona tiene tres municipios importantes:

<u>Municipio</u>	<u>C.S.</u>	<u>I.C.</u>
Haro	2ª	2ª
Nájera	2ª	3ª
S. Domingo	2ª	3ª

Comprende la denominada Rioja Alta y las cuencas del Najerilla, Oja y Tirón.

Tendrá tres comarcas:

Comarca de Haro:

Está formada por la zona de influencia nodal de Haro y por otro lado resulta bastante homogénea dado que se constituye fundamentalmente con el cluster G y parte del H (que son bastante similares) y algunas zonas aisladas de otros - clústeres.

Comarca de Nájera:

Está formada en su mayor parte por la cuenca del Najerilla y vemos dos zonas claramente diferenciadas, como son - el Alto y Bajo Najerilla.

El Bajo Najerilla, sin ser excesivamente homogéneo, si

tiene bastante en común, dado que es una zona de agricultura interesante. En el centro se halla Nájera, que es la capital de la comarca.

El Alto Najerilla, es una zona bastante homogénea, a la vez que bastante pobre y en clara recesión. Es por ello - que hemos decidido crear una subcomarca del Alto Najerilla - con centro en Anguiano, que es el único municipio de relativa importancia y que a la vez está bastante bien comunicado.

El problema es que Anguiano se encuentra muy desplazado del centro de la subcomarca, pero es la única solución posible, ya que al Sur del Alto Najerilla está muy despoblado.

Esta subcomarca habrá de ser objeto de especial atención, dado que su potencial actual es mínimo.

Comarca de Santo Domingo:

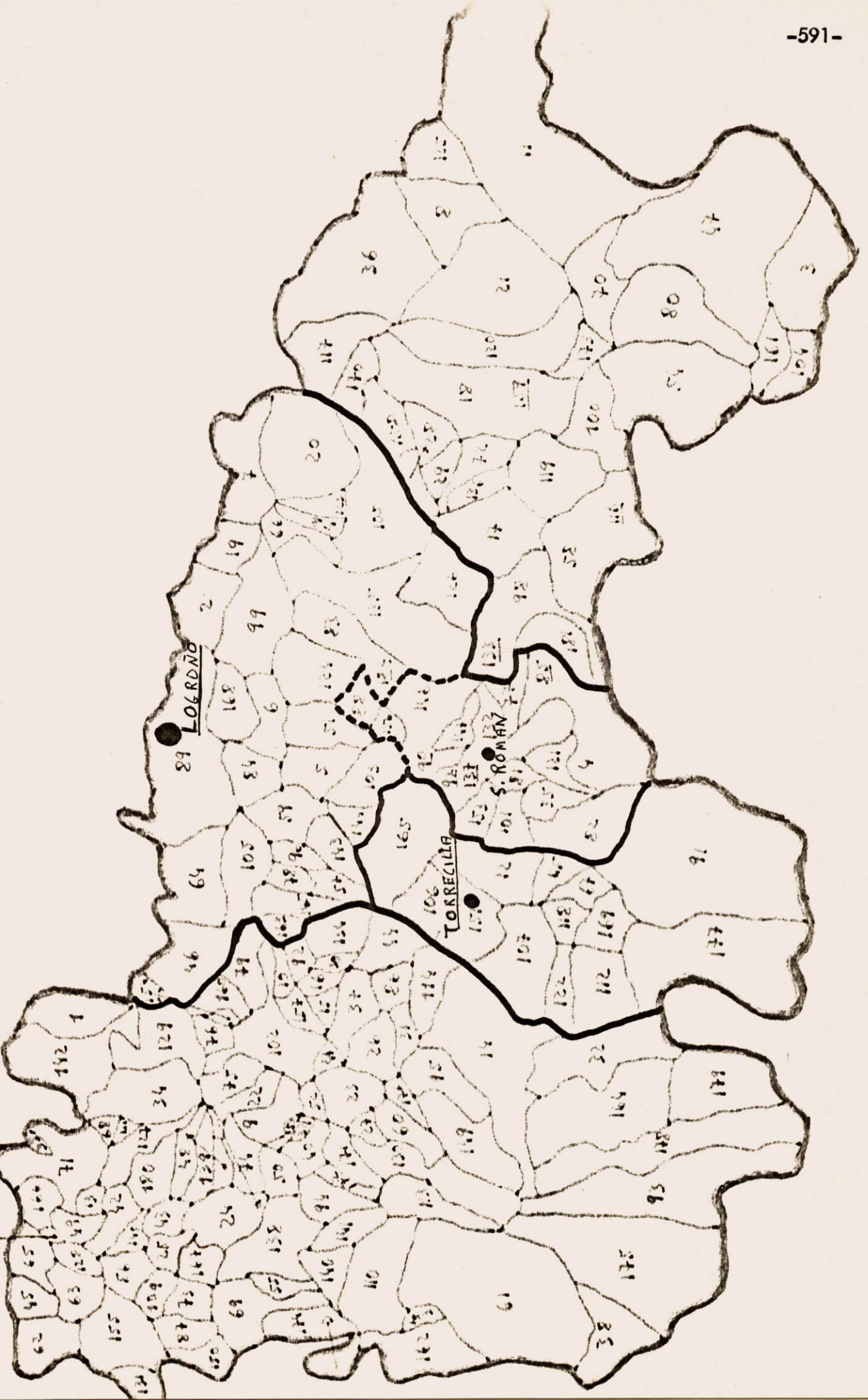
Se forma esta comarca con la zona de influencia de Santo Domingo, que coincide en gran parte con los clusteres D y H y con la cuenca alta del río Oja.

Nos encontramos aquí con un problema similar al anterior, pues mientras los alrededores de Santo Domingo son zonas agrícolas de cierta importancia, la parte Sur de la comarca la hemos de encuadrar entre las tierras altas de la sierra.

La ventaja es que en esa parte Sur tenemos a Ezcaray,

que por su tradición como pueblo de veraneo y su actual desarrollo a causa de la estación invernal de Valdezcaray, mantiene servicios y puede ser un centro aceptable sin excesivo esfuerzo. Es por ello que no dudamos en crear en esta parte sur de la comarca la subcomarca del Alto Oja, con centro en Ezcaray.

Pasamos a continuación a la comarcalización de la zona centro, cuyo centro es Logroño.



Zona Centro

Centro: Logroño.

Esta zona tiene como ciudad más importante Logroño que es C.S. de 1ª categoría e I.C. de 1ª también y se extiende - por la denominada Rioja Central y las cuencas del Iregua y del Leza.

A parte de Logroño, carece de centros de servicios importantes, pues no tiene ninguno de 2ª.

De tercera tenemos:

- Albelda.
- Fuenmayor.
- Torrecilla.

Las dos primeras carecen de interés a la hora de comarcalizar, pues se encuentran muy cerca de Logroño, solo Torrecilla tiene un interés real como centro comarcal

Los centros de servicios de cuarta categoría se encuentran también muy cerca de Logroño, así como todos los centros industriales y comerciales, que por otro lado, son de escasa entidad.

A la vista de esto, solo podemos hablar de dos comarcacas, la de Logroño y la de Cameros Nuevo, con centro en Torrecilla.

Comarca de Logroño:

Según nuestro criterio de nodalidades es muy difícil conocer cual es el verdadero alcance de la comarca de Logroño, pues al ser la capital de la provincia y además la ciudad más importante con bastante diferencia, ejerce un fuerte potencial de atracción sobre toda la provincia. Por ello ha sido necesario acotar un poco su influencia siguiendo criterios geográficos y de comunicaciones.

Dado que la comarca de Torrecilla va a incluir el Cameros Nuevo (1), desde un punto de vista de homogeneidad (2). La comarca de Logroño tiene dos partes claramente diferenciadas, la parte agrícola cercana al Ebro (3) y a la desembocadura del Iregua y del Leza, y la parte montañosa de las cuencas altas del Leza y su afluente el Jubera.

Sería interesante a la vista del estudio del cluster análisis, crear una subcomarca Leza - Jubera, pero las comunicaciones entre ambas cuencas son prácticamente imposibles.

- (1) Se entiende por Cameros Nuevo, el alto Iregua.
- (2) En realidad, todos los Cameros son homogéneos, pero por problemas de comunicaciones, los Cameros nuevos, no pueden unirse con los viejos.
- (3) En esta parte de la comarca se encuentra el polo de desarrollo de Logroño, que está plenamente incluido en la comarca de Logroño.

Por ello, y ante la tremenda influencia nodal de Logroño sobre la cuenca del Jubera, hemos decidido crear una subcomarca en el alto Leza, donde nos encontramos con un problema similar a lo que sucedía en la subcomarca del alto Najerilla, pues aquí tampoco tenemos un centro relativamente importante.

De hecho se puede observar en las hojas que recogen - los flujos que San Román ejerce una importante atracción nodal sobre esta zona, aún sin tener excesivo potencial, y puede ser el centro de esta subcomarca, que será preciso potenciar de cara a evitar su continua recesión, pues ya observamos que esta subcomarca, también denominada de los Cameros Viejos, tiene la mayor parte de los municipios con baja calidad de vida.

Por lo que se refiere a la cuenca del Jubera, desde un punto de vista homogéneo, habría que unirla a la subcomarca del alto Leza, pero sus comunicaciones son muy malas, y prácticamente hay que ir hasta Logroño para volver a subir por el Leza. De todas formas se podría arbitrar un sistema, como ya comentaremos, donde la dependencia nodal dependa de Logroño, pero a la hora de planificar se incluya en la subcomarca del Alto Leza.

Comarca de Torrecilla:

Incluye los Cameros con comunicación posible con Torrecilla, que coinciden con los denominados Cameros Nuevos, o con la comarca del Alto Iregua.

A la hora de fijar los límites de esta comarca nos hemos fijado en las comunicaciones a Este y Oeste y en la homogeneidad para fijar el límite con la comarca de Logroño en el Sur. Esto se ha debido a que las actuales comunicaciones son fáciles entre la comarca de Torrecilla y Logroño, y siendo Torrecilla un municipio en franca recesión, su potencial de atracción nodal es más bien escaso y si nos fijásemos en ello, toda la comarca habría de incluirse en Logroño. Esto no es lógico, dado que los Cameros Nuevos tienen potencial suficiente para tener una comarca y Torrecilla para ser un centro de servicios, luego será otra comarca a potenciar.



Zona Este:

Centro: Calahorra.

Comprende la Rioja Baja y las cuencas del Cidacos y Alhama (este último a su paso por la provincia).

Desde un punto de vista nodal contiene bastantes centros importantes, entre ellos podemos destacar:

<u>Municipio</u>	<u>C.S.</u>	<u>I.C.</u>
Calahorra	1ª	2ª
Arnedo	2ª	2ª
Alfaro	2ª	3ª
Cervera	3ª	4ª
Rincón	3ª	5ª
Aldeanueva	3ª	-

Desde un punto de vista homogéneo, hay dos zonas fundamentales: Las vegas de los rios y la parte montañosa.

A la hora de elegir los centros nodales parece que Calahorra, Arnedo y Alfaro, deberían ser los municipios que polarizasen la zona, pero dada la proximidad de Calahorra y Alfaro, y debido al mayor potencial de la primera de estas ciudades, por un lado, y a las deficientes comunicaciones de la zona del Alhama, cuyo centro natural es Cervera, hemos decidido que los centros sean:

- Calahorra - Alfaro. Comarca bicéfala con mayor importancia naturalmente de Calahorra.
- Arnedo.
- Cervera.

Comarca de Calahorra - Alfaro:

Es una comarca altamente homogénea que recoge la influencia polar de estos dos importantes centros.

Comprende fundamentalmente la ribera del Ebro, extendiéndose también por la del Cidacos.

Se han incluido en esta comarca las zonas de influencia dudosa que son homogéneas con el resto.

Comarca de Arnedo:

Se extiende fundamentalmente por la parte montañosa - de la zona que tiene influencia de Arnedo, pero al igual que en la anterior, en las zonas de influencia compartida se ha atendido más a la homogeneidad, tratando de incluir en esta comarca las zonas montañosas y en la de Calahorra las vegas de los rios.

Arnedo, concretamente como municipio, es el único que tiene características de vega más que de montaña dentro de esta comarca, pero parece que es claro que hay que incluirlo - en esta comarca y no sacarlo de ella pues desde un punto de vista nodal su influencia sobre las partes montañosas es evidente.

Comarca de Cervera:

Esta comarca se caracteriza por sus malas comunicaciones con el resto de la zona, lo que hace que Tudela sea un centro de atracción nodal importante.

Hemos incluido aquí los municipios que dependen de Cervera de manera más o menos intensa y en los casos dudosos las zonas homogéneas.

El problema es que Cervera ha de ser potenciada mucho más, pues sino esta comarca quedará dividida entre las otras dos, hecho que ya hubiera sucedido de no ser por el mal estado de las comunicaciones.

VII.- RESUMEN GENERAL. APLICACIONES. CONCLUSIONES.

Comenzaremos resumiendo brevemente los criterios de comarcalización seguidos, de manera que nos sirvan, junto con el esquema general obtenido, para estudiar las aplicaciones del presente estudio.

El criterio general ha sido mixto: Homogeneidad - Nodalidad, y por lo tanto, implícito al criterio de nodalidad hay un componente geográfico y comunicacional.

Así, para la división en zonas, hemos elegido tres zonas suficientemente representativas con sus nodos o polos correspondientes. Estas zonas se han creado siguiendo los si-guientes puntos:

- Geográfico.
- Comunicacional.
- Influencia nodal, no solo del centro principal de la zona, sino también de los centros comarcales.

Cada zona se divide en comarcas, atendiendo a la influencia nodal de los centros comarcales, lo que implica una posibilidad e incluso una existencia de comunicación, y en los puntos limítrofes se ha atendido a la homogeneidad.

En el caso de comarcas con fuertes heterogeneidades,

se ha estudiado la creación de subcomarcas homogéneas, si es to es posible, bien alrededor de nodos existentes, bien potenciando nodos posibles. Para la creación de estas subcomarcas se ha atendido también la posibilidad de comunicación con el centro nodal.

Vemos pues que hemos optado por un sistema jerarquizado de divisiones, que creo que es el que mejor se adapta, tanto a la estructura de la provincia, como a la aplicación de nuestra sistemática.

El resultado que hemos comentado en el capítulo precedente se ha logrado pues, mediante un continuo proceso de - feed-back, es decir, se ha ido comarcalizando en zonas, comarcas y subcomarcas y volviendo en cada paso al anterior, - retocando los puntos conflictivos y llegando a un resultado que creo que es satisfactorio, pues hemos obtenido divisiones de gran homogeneidad en general y de gran sentido nodal.

El esquema general de comarcalización podemos resumirlo en:

- . Provincia: La Rioja - Capital Logroño.
- Zona Oeste - Centro Haro.
 - . Comarca de Haro.
 - . Comarca de Nájera.
 - Subcomarca del Alto Najerilla: Centro Anguiano.
 - . Comarca de Santo Domingo.
 - Subcomarca del Alto Oja: Centro Ezcaray.

- Zona Centro - Centro Logroño.
 - . Comarca de Logroño.
 - Subcomarca del Alto Leza: Centro San Román.
 - Especial trato al área del Jubera.
 - . Comarca de Torrecilla.

- Zona Este - Centro Calahorra.
 - . Comarca de Calahorra - Alfaro.
 - . Comarca de Arnedo.
 - . Comarca de Cervera.

Creo que es preciso reseñar aquí al decir especial trato al área del Jubera, nos referimos a la zona homogénea del Alto Jubera que al carecer de ningún tipo de centro, y ante la imposibilidad de crearlo, dada la poca población del área e incluso los pocos municipios, además de sus malas comunicaciones, hemos optado por introducirlo en la comarca de Logroño, pues a causa de sus deficientes comunicaciones no puede estar con el Alto Leza, que es su zona homogénea natural.

En consecuencia, es el caso más claro de falta de homogeneidad y corre el peligro de quedar olvidada, como hasta - ahora. Por ello proponemos como posible solución, que aunque nodalmente dependa de Logroño, se le aplique la planificación correspondiente a la subcomarca del Alto Leza, por ser muy - similares sus características.

Un hecho, ya comentado, que solucionaría en gran parte

los problemas de la sierra, y que cambiaría muchos supuestos de nuestra comarcalización, sobre todo por lo que se refiere a la nodalidad, sería la carretera de Cameros que los atravesaría de Este a Oeste, pero ya hemos hablado de las dificultades de este proyecto para que se complete definitivamente.

En cuanto a la utilidad de este sistema de comarcalización, para el caso concreto de la provincia riojana, creo que resulta bastante convincente.

En primer lugar, gracias a que hemos respetado escrupulosamente los componentes comunicacionales, cada área (zona, comarca o subcomarca) puede acudir a su centro sin dificultad y además mientras los nodos zonales o comarcales son nodos existentes, los nodos subcomarcales son potenciables y en definitiva, pueden servirnos, tanto unos como otros, aunque siempre los primeros en mayor medida, para la comarcalización de servicios de todo tipo, a diversos niveles y con un orden jerárquico establecido.

Evidentemente no es el objetivo de este trabajo el realizar un proyecto de planificación de servicios, pues sería necesaria la colaboración de especialistas en los temas específicos, pensemos por ejemplo en una comarcalización de los servicios de la Seguridad Social o de los centros escolares, que evidentemente es necesaria, pero de lo cual solo puedo hablar con los conocimientos del usuario, y deben ser los especialistas en los distintos temas los que establezcan

la magnitud, tanto cuantitativa como cualitativa de los centros a los diversos niveles.

Lo que quiero recalcar aquí es que esta característica nodal y jerárquica de nuestra comarcalización, permite - grandes posibilidades en los temas comentados y en tantos - otros que se podían citar, dada su gran flexibilidad.

Pero a parte del componente nodal, contamos con la - característica de la homogeneidad. Esto considero que puede ser muy interesante de cara a la planificación del desarrollo, pues las áreas estudiadas tienen en general, una problemática bastante común. Este hecho lo podemos corroborar con la observación directa de las diferentes áreas y nos daremos cuenta que realmente tanto sus posibilidades como sus carencias son muy uniformes dentro de cada área.

En concreto, estudiaremos posteriormente el caso de las áreas deprimidas y veremos la coherencia de nuestra comarcalización.

Ya en la idea de la creación de las subcomarcas, está implícita la necesidad de una potenciación de determinados centros que sirvan como motor de desarrollo del área y en general una aspiración de desarrollo equilibrado de todas las áreas.

Solamente en el caso del área del Jubera, hemos decidi

dido la imposibilidad de crear ahí una subcomarca y ya hemos apuntado la necesidad de una atención especial.

En general, podemos observar que aun insistiendo en ~~que~~ la labor de planificación es posterior a este trabajo, ya podemos dar las pautas a seguir bajo el presupuesto de dar prioridad a las zonas menos favorecidas.

Estas áreas se encuentran fundamentalmente en la parte montañosa del Sur de la provincia, que genéricamente denominamos sierra. En ellas será preciso potenciar el nodo si este no tiene fuerza, o incluso crearlo, y potenciar los municipios.

Las áreas de la sierra son:

- Subcomarca del Alto Oja:
 - Potenciar Ezcaray.
 - Potenciar área.
- Subcomarca del Alto Najerilla:
 - Crear y potenciar un centro en Anguiano.
 - Potenciar área.
- Comarca de Torrecilla:
 - Potenciar Torrecilla.
 - Potenciar área.

- Subcomarca del Alto Leza:
 - Crear y potenciar un centro en San Román.
 - Potenciar área.

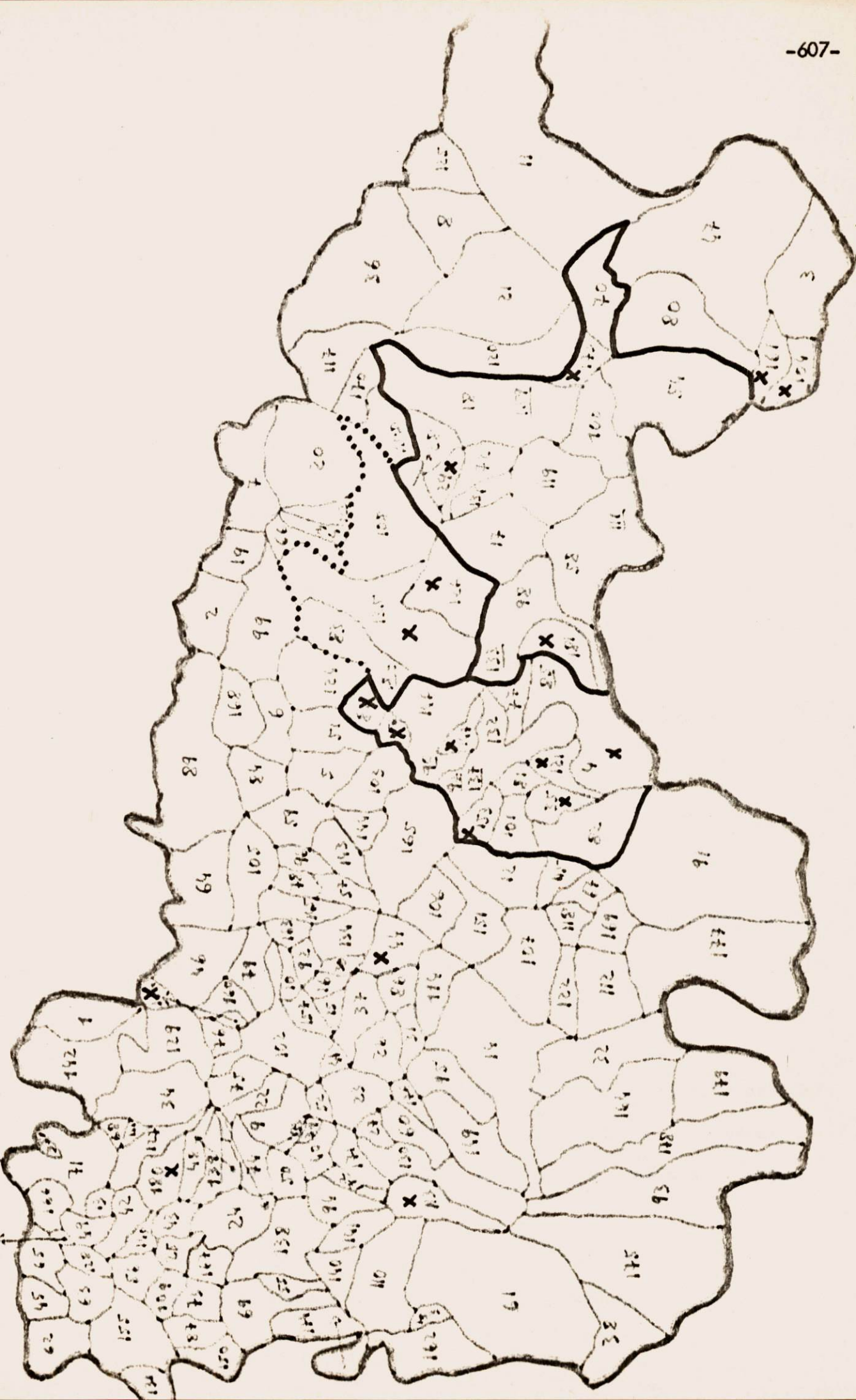
- Area del Alto Jubera:
 - Potenciar área.
 - No creemos posible el crear un centro.

- Comarca de Arnedo:
 - Potenciar área.
 - El centro, Arnedo, es potente ya.

- Comarca de Cervera:
 - Potenciar Cervera.
 - Potenciar área.

Especial atención habrá que conceder a los municipios que se encuentran con estándares de vida realmente bajos y - donde nuestro estudio sobre la calidad de vida ha puntuado - muy negativamente.

Estas son las zonas más deprimidas de la provincia, y podemos ver por el siguiente mapa, que su ubicación es curiosamente coincidente en determinadas áreas.



En el mapa precedente podemos observar los 18 municipios, que podemos considerar que carecen de los servicios más elementales, encontrándose muy por debajo de la media provincial.

Curiosamente de estos 18 municipios, 7 se encuentran en la subcomarca del Alto Leza. Si consideramos que esta subcomarca tiene 13 municipios en la actualidad, nos daremos cuenta que el 53'8% de los municipios de esta área tienen muy baja calidad de vida.

En el área del Jubera, 2 de los 4 municipios existentes se encuentran en la misma situación, es decir el 50%.

Tres más tenemos en la comarca de Arnedo y dos en la de Cervera, en una zona que por homogeneidad debería estar unida a Arnedo, pero al carecer de comunicaciones ha de unirse a Cervera.

Estos dos municipios de Cervera, no los hemos estudiado especialmente a causa de su poco peso específico, siendo lo normal que se integren en otros municipios.

El resto (cuatro) de los municipios con baja calidad de vida se encuentran bastante repartidos y, en general, prácticamente ya no existen pues apenas quedan habitantes.

Luego en definitiva, podemos concluir que entre el Leza y el Jubera están concentrados muchos municipios depri

midos y que nuestra comarcalización, al haber tenido en cuenta la homogeneidad, los ha agrupado en áreas concretas, de tal forma que puedan ser objeto de un tratamiento unificado.

El resto de los municipios deprimidos (salvo los cuatro dispersos ya citados), se concentran entre Arnedo y Cervera (de los de Cervera ya hemos hablado) y pueden ser objeto de un tratamiento unificado también.

En conclusión, la subcomarca del Alto Leza, el área del Jubera, la comarca de Arnedo, y el Sur de la comarca de Cervera (esta última la que menos), han de ser objeto de una atención muy especial.

CONCLUSIONES.

El estudio aquí presentado creo que tiene un interés múltiple:

A) En primer lugar, el tratamiento cuantitativo de los datos reduce la aparición de subjetivismos, y aunque es tos siempre pueden darse de alguna manera, los reduce al mí nimo.

B) En segundo lugar, el tratamiento cuantitativo se ha aplicado de una manera multivalente, al tener en cuenta no sólo los conceptos fundamentales de nodalidad y homogenei dad, sino incluso también los componentes industrial y comer cial y de calidad de vida, lo que lo hace en su conjunto de una elevada originalidad.

C) El énfasis hecho en el criterio mixto, bastante - original por otra parte en cuanto a su utilización, nos produce evidentes ventajas ya estudiadas de cara a la planifica ción.

D) El planteamiento de un sistema de comarcalización jerarquizada nos ha permitido la utilización de criterio mix to sin excesivos problemas, a parte de las ventajas evidentes de los sistemas comarcales jerarquizados.

E) Esta comarcalización puede ser la base del comien zo de una planificación coherente para toda la provincia -

atendiendo a las áreas deprimidas de forma especial, y en general comarcalizando servicios y potenciando áreas.

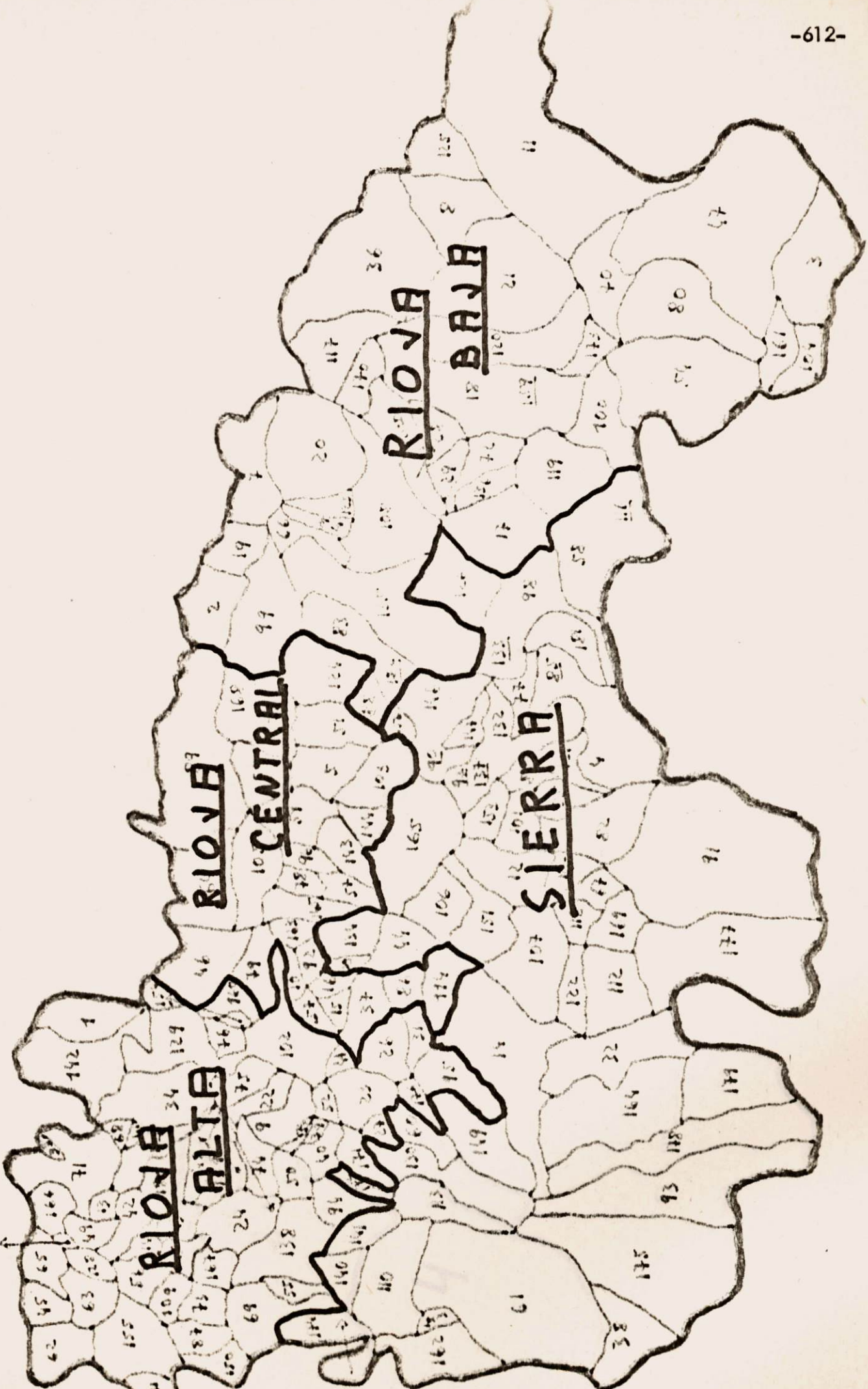
F) En comparación con otros estudios realizados en la provincia, el nuestro tiene evidentes ventajas.

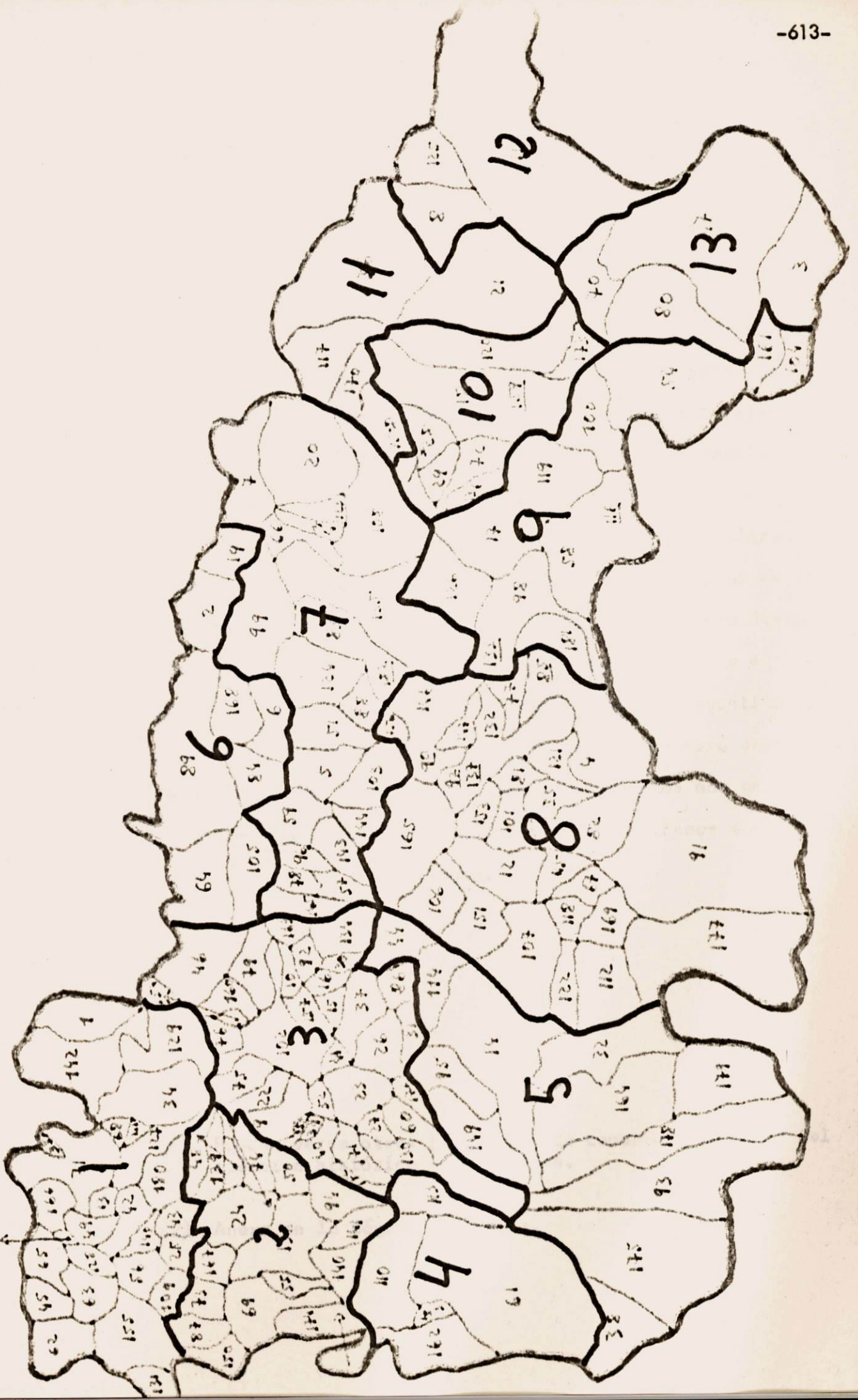
No voy a ser exhaustivo en la cita de los estudios de comarcalización realizados en La Rioja, pues en general sus defectos y conclusiones son similares. Como ejemplo podemos ver el ya citado de la Diputación (cada municipio está adscrito a una comarca), el recogido por la antigua Organización Sindical (1), que se basa en crear zonas supuestamente homogéneas y el reciente estudio de la Diputación para comarcalizar en zonas homogéneas (2).

Vemos en la siguiente página, los mapas correspondientes a estos dos estudios, primero el de la Organización Sindical y luego el de la Diputación.

(1) Consejo Económico Provincial Sindical de Logroño (1,970) página 279.

(2) Gabinete de planificación de la Diputación de Logroño.





En general estos dos estudios adolecen de defectos si milares a los del resto de las comarcalizaciones realizadas en la provincia (1) y que podemos resumir en:

- Apreciación subjetiva de la homogeneidad.
- Carencia de visión nodal.

Con lo que las comarcas obtenidas son de poca utilidad. Nuestro estudio, por el contrario, creo que supera estos problemas y consigue un modelo mucho más válido y general.

G) Por último, considero que nuestra comarcalización puede servir como pauta para el estudio de provincias de si milares características a las de la Rioja, o incluso distintas, adoptando convenientemente los supuestos (2), y en general puede ser la base para un estudio realmente cuantitativo y riguroso de los problemas económicos, que se verá acrecentado en su eficacia conforme dispongamos de datos más amplios y fiables, pues en la carencia de los mismos hemos encontrado nuestras mayores limitaciones.

(1) Pueden verse otras propuestas de comarcalización en el apéndice histórico - geográfico.

(2) Anexo en la página siguiente.

Anexo citado:

Con esto quiero decir que, como al hablar de homogeneidad en La Rioja, nos hemos referido a los aspectos agrícolas, dadas por un lado las características de la región que estudiamos y por otro la ausencia de otros datos, en regiones donde de podesen datos de rentas p.c. por municipios, etc. este criterio u otros pueden ser interesantes.

Por otro lado, en regiones muy industrializadas puede ser interesante utilizar como criterio complementario, la especialización industrial, pero nosotros a parte de carecer de datos, podemos deducir, a la vista de los resultados del estudio de las licencias fiscales, que hay escasez de especialización, al menos al nivel de agregación que representan las licencias, y la industrialización es más un elemento distintivo de determinados municipios, que un factor de homogeneidad.

Puede haber también otros criterios interesantes, pero no es objeto de este trabajo estudiarlos.

