

PROGRAMA D'ANALISI DE DADES

CURS 1990-1991

FACULTAT DE CIENCIES POLITIQUES I DE SOCIOLOGIA

UNIVERSITAT AUTONOMA DE BARCELONA

PROGRAMA DE ANALISIS DE DATOS

Prof. Carlos LOZARES

TEMARIO GENERAL

I. COMPONENTES MATEMATICOS

Tema 1. Introducción. Cálculo algebraico y matricial.

II. DEL BIVARIADO AL MULTIVARIADO

Tema 2. Del análisis bivariado al multivariado.

Tema 3. El análisis multivariado.

III. ANALISIS FACTORIALES

Tema 4. Análisis de componentes principales.

Tema 5. Análisis de correspondencias.

Tema 6. Análisis discriminante.

IV. LA CLASIFICACION

Tema 7. Tipologías.

V. INTRODUCCION AL ANALISIS DE ESTRUCTURAS CAUSALES

Tema 8. Análisis causal. Estructuras causales.

VI. ANALISIS DIACRONICO EN SOCIOLOGIA

Tema 9. Introducción al análisis diacrónico.

VII. ANALISIS CUANTITATIVO DE CONTENIDO

Tema 10. Introducción al análisis frecuencial de contenido.

Lexicometría.

ANALISIS DE DATOS Curso 1989-1990. Carlos Lozares.

1.-Planteamiento y objetivos generales

El curso de análisis de datos se plantea como una continuación de la perspectiva cuantitativa inicialmente explorada en el curso de Técnicas de investigación. Para dar coherencia a dicha continuidad y comprender mejor el salto al multivariado se insiste en el paso del análisis bivariado al trivariado.

Con todo dos elementos son fundamentales para el seguimiento y mejor aprovechamiento del curso. El primero es el conocimiento y manejo de los instrumentos matemáticos imprescindibles. Dicha información se dará integrada en el curso como introducción al mismo. El segundo es la familiaridad con el soporte informático que se hará a lo largo del curso como prácticas del mismo con ayuda de colaboradores.

El programa contiene los métodos de análisis de datos que mas interés presentan para el sociólogo. Su variedad y extensión hacen imposible que puedan ser tratados con igual profundidad. Por ello de algunos se dará simplemente una introducción informativa. Otros serán tratados con mas minuciosidad y extensión. Dentro de estos últimos están, además de la introducción matemática comentada, los referentes a la regresión simple y múltiple, el análisis de varianza, el análisis de componentes principales , análisis de correspondencias y cluster. Dentro de los primeros estarán el análisis de covarianza, el logit, el análisis discriminante, los modelos de estructuras causales y las series temporales.

Se pondrá un interés particular en que el alumno posea los elementos suficientes para que pueda prolongar personalmente el conocimiento de aquellos métodos y técnicas que no hayan podido ser tratados con la profundidad requerida.

2.- La evaluación de los alumnos.

Dada la naturaleza del contenido y la poca afinidad que se tiene con los procedimientos matemáticos por parte de los estudiantes de sociología, el seguimiento permanente y la asistencia regular al curso aparecen imprescindible. Esto permite acercarse mas a la idea de evaluación continua.

La evaluación tendrá dos niveles:

-los ejercicios que se propondrán para cada uno de los temas y que deberán ser realizados y entregados por los alumnos de manera regular,

-una investigación propia del alumno a partir de datos secundarios o provenientes de otras investigaciones realizadas ,o

con las que ha tenido relación, por el alumno , pero siempre que tenga la motivación suficiente para realizarla; sobre dichos datos debe existir un modelo de análisis que se puede ir elaborando a lo largo del curso a medida que los distintos métodos van apareciendo.

-el examen que al menos tendrá dos momentos parciales durante el curso.

3El programa del curso.

TEMARIO GENERAL

PARTE I:COMPONENTES MATEMATICOS

Tema 1.Introducción.Elementos del cálculo algebraico y matricial.

PARTE II:DEL BIVARIADO AL MULTIVARIADO

Tema 2.Del análisis bivariado al trivariado.
Tema 3.El análisis multivariado.

PARTE III:LA CLASIFICACION

Tema 4.Tipologías.

PARTE IV:ANALISIS FACTORIALES

Tema 5.Análisis de componentes principales.
Tema 6.Análisis de correspondencias.
Tema 7.Objetivos y planteamiento general del análisis discriminante.

PARTE V:INTRODUCCION AL ANALISIS DE ESTRUCTURAS CAUSALES.

Tema 8 .Objetivos y planteamiento general del análisis causal.
Estructuras causales.

PARTE VI:ANALISIS DIACRONICO EN SOCIOLOGIA

Tema 9.Introducción general al análisis diacrónico.

PARTE VII:ANALISIS CUANTITATIVO DE CONTENIDO

Tema 10.Introducción general al análisis frecuencial de contenido.Lexicometría.

TEMARIO DETALLADO

PARTE I: COMPONENTES MATEMATICOS

Tema 1. Introducción. Elementos del cálculo algebraico y matricial.

1. Datos multivariados y espacios vectoriales.

1.1. Datos multivariados en la perspectiva distributiva.
Tipos de medidas. Relaciones entre variables. Tipos de análisis posibles.: confirmatorio, relaciones supuestas y exploratorio, descriptivo, correlacional.

1.2. Matriz de datos. El dato. Matriz de datos. Procesos de transformación . Tipos de matrices de datos.

1.3. Análisis multivariado.

2. Vectores y espacios vectoriales.

2.1. Definición

2.2. Operaciones

2.3. Espacios vectoriales.

3. Cálculo matricial.

3.1. Definiciones y tipos de matrices. Matrices y vectores.

3.2. Operaciones con matrices.

3.3. Casos particulares de operaciones con matrices.

4. Base de un espacio vectorial

5. Matriz inversa

5.1. Determinante de una matriz.

5.2. Matriz inversa, propiedades y cálculo

5.3. Aplicaciones.

6. Sistemas lineales.

6.1. Modos generales de resolver los sistemas lineales

6.2. Sistemas homogéneos

7. Vectores y valores propios.

7.1. Valores propios. Propiedades

7.2. Vectores propios

7.3. Métodos de cálculo de valores y vectores propios.

8. Funciones cuadráticas. Transformaciones.

9. Máximos y mínimos de funciones cuadráticas.

10. Ejercicios.

Bibliografia para el tema 1.:

- Bennet S., Bower D., 1976,
"An introduction to multivariate techniques for social and
behavioural sciences"
The Macmillan Press LTD M.London .(etc)
- Laforgue H., 1981, pg. 1-59.
"Analyse multivariée"
Etudes vivantes. Montreal.
- Lebart L., Morineau A., Fénelon J.P., 1979. pg.480-490.
"Traitement des données statistiques"
Dunod. París.
- Lefevre J., 1980, pg.1-90.
"Introduction aux analyses statistiques multidimensionnelles".
Masson. París. (Etc.)
- Manly B. F. J., 1986, pg.17-25.
"Multivariate Statistical Methods"
Chapman and Hall. London. N.Y.
- Namboodiri K., 1984, pg.1-96.
"Matrix Algebra. An introduction."
Sage Publication. B.H.,London, N.D.

PARTE II: DEL BIVARIADO AL MULTIVARIADO

Tema 2: Del análisis bivariado al trivariado.

A. Modelos matemáticos, modelos lineales.

B. Análisis bivariado

1. Repaso de indices de asociación de variables, nominales, y ordinales. Ejercicios.
2. Expresión matricial de la varianza y covarianza en dos variables. Matriz de desviaciones típicas. Matriz de coeficientes de regresión. Matriz de coeficientes de correlación. Ejercicios.
3. Análisis de varianza . ANOVA.
 - 3.1. El modelo de análisis. Tipos de variables intervintes.
 - 3.2. Análisis de varianza con un solo factor,
- efecto fijo, modelo aleatorio.

- efecto fijo, por bloques
- efecto fijo, medidas repetidas.

3.3. Expresión matricial.

3.4. Ejercicios

4. Regresión simple y correlación.

4.1. El modelo y los supuestos.

4.2. Los parámetros. Coeficientes de regresión.

4.3. Varianza de la regresión, significación.

4.4. Coeficiente de correlación, coeficiente de determinación.

4.5. Predicción e intervalos de confianza.

4.6. Validación del modelo.

4.7. Ejercicios.

5. Objetivos y planteamiento general del análisis de covarianza. ANCOVA

6. Objetivos y planteamiento general del logit.

7. Análisis de regresión con variables ficticias.

C. Relación entre tres variables

1. Cruce de tres variables de tipo nominal. Modelos y análisis. Ejercicios.

2. Matrices de varianza, covarianza y correlación. Ejercicios

3. Análisis de varianza con dos factores. ANOVA.

3.1. El modelo

3.2. Modelo con dos factores,

-de efecto fijo

-de efecto fijo y de clasificación jerárquica

-de efecto fijo y medidas repetidas.

3.3. Expresión matricial.

3.4. Ejercicios de interpretación.

4. Regresión múltiple, dos variables independientes.

4.1. El modelo y los supuestos.

4.2. Los parámetros: coeficientes de regresión parcial no estandarizados y estandarizados, coeficientes de correlación parcial

4.3. Varianza de la regresión. Significación.

4.4. Coeficiente de regresión múltiple.

4.5. Predicción y validación del modelo.

4.6. Ejercicios de interpretación.

5. Planteamiento general del ANCOVA, LOGIT y var. ficticias.

Bibliografía para el tema 2

Domenech J.M., Riba M.D., 1985

"Métodos estadísticos. Modelo lineal de regresión"
Herder. Barcelona.

García Ferrando M., 1984

"Socioestadística. Introducción a la estadística en sociología"
CIS. Madrid.

Glass G.V., Stanley J.C., (1970), 1974

"Métodos estadísticos aplicados a las ciencias sociales"
Prentice/Hall International. Englewood Cliffs. [Etc]

Laforgue H., 1981

"Analyse multivariée"
Etudes Vivantes. Montreal.

Lebart L., Morineau A., Fénelon J., 1979.

"Traitement de données statistiques"

Dunod. París.

Lefebvre J., 1980

"Introduction aux analyses statistiques multidimensionnelles"
Masson. París.

San Martín Castellanos R., Pardo Merion A., 1989

"Psicoestadística"
Pirámide . Madrid.

Stevens J., 1986

"Applied multivariate statistics for the social sciences"
Lawrence Erlbaum Associates. Hillsdale. NJ. London.

Visauta Vinacua B, Batalle Descals P., 1986

"Métodos estadísticos aplicados"
Tomas I, II
PPU. Barcelona.

Wonnacott H., Wonnacott R.J., (1972), 1981

"Fundamentos de estadística para la administración y economía"
Linusa. Mexico.

Tema 3 :Panorama general del análisis multivariado

1. En qué consiste al análisis multivariado.

2. Los criterios de clasificación del anál. multivariado:

- tipos de medidas y escalas
- naturaleza de las variables empleadas
- las matrices de procedencia

- los espacios de atributos referenciales
- índices utilizados
- los objetivos perseguidos

3. La clasificación.

- con finalidad más bien explicativa lo que supone una relación de dependencia entre variables
- con finalidad más bien descriptiva y de dimensionalización lo que implica la interdependencia entre variables
- con finalidad más bien descriptiva y clasificatoria lo que también supone la interdependencia entre variables.

Bibliografía para los temas 3,4,5,6 :

Aldenderfer M.S., Blashfield R.K., 1984.
Cluster Analysis T44.
Beverly Hills. Sage Publications.

Bennett S., Bowker D., 1976.
An introduction to multivariate techniques for Social and Behavioural Sciences.
London. The Macmillan Press.

Benzecri J.P., 1982.
L'Analyse des données. Vol 1: *La taxonomie*.
" " . Vol 2: *L'Analyse de correspondances*.
París . Dunod.

Bertier P., Bourouche J.M., 1975.
Analyse des données multidimensionnelles.
París. PUF.

Bourouche J-M, Saporta G., 1980
L'analyse des données
París. Que sais-je?. PUF.

Cibois Ph., 1983.
L'Analyse factorielle.
París. PUF. Que sais-je?.

Comrey A.L., 1985
Manual de análisis factorial
Madrid. Cátedra.

Cornejo J.M., 1988.
Técnicas de investigación social.
Barcelona. PPU.

Cuadras C.M., 1981.
Métodos de análisis multivariados.
Barcelona. Ed. Universitarias.

- Chandon J.L., Pinson S., 1981.
Analyse typologique.
Paris. Masson.
- De Lagarde J., 1983
Initiation à l'analyse des données
Paris .Dunod.
- Diday E., Lemaire J.L., Fouget J., Testu F., 1982.
Eléments d'analyse de données.
Paris. Dunod.
- García Ferrando M., 1984
Socioestadística
Madrid. CIS.
- Gendre F., 1976.
L'Analyse statistique multivariables.
Genève. Librairie Doz.
- Guigou J-L., 1977.
Méthodes multidimensionnelles.
Paris. Dunod.
- Hope K., 1978.
Methods of multivariate analyse.
London.Univ.of London Press Ltd.
- Kim J., Mueller Ch..W., 1978a.
Introduction to factor analysis. T13.
Beverly Hills. Sage Publication.
- Kim J., Mueller Ch..W., 1978b.
Factor Analysis. T14.
Beverly Hills. Sage Publication.
- Laforgue H., 1981
Analyse multivariée.
Etudes Vivantes. Montreal.
- Lebart L., Morineau A., Fénelon J., 1979.
Traitemennt de données statistiques.
Dunod. Paris.
- Lefebvre J., 1980
Introduction aux analyses statistiques multidimensionnelles.
Masson. Paris.
- Manly B.F.Y., 1986.
Multivariate Statistical Methods.
London. Chapman and Hall.

Robert C., 1989.
Analyse descriptive multivariée.
París. Flammarion.

Sanchez Carrión J.J., (ed) 1984.
Introducción a las técnicas de análisis multivariante
aplicadas a las ciencias sociales
Madrid. CIS.

Takenchi K. y alt., 1982.
The foundations of multivariate analysis
Nueva Delhi. [Etc]. Wiley Eastern Limited.

Volle M., 1985.
Analyse des données.
París. Economica.

III METODOS DE CLASIFICACION.

Tema 4: Técnicas multivariables: La clasificación.

- 1.-Introducción. Generalidades. La clasificación en diversas ciencias. Reflexiones morfológicas y metodológicas.
- 2.-La tipología y clasificación en sociología.
 - 2.1.Clasificación y tipologías.
 - 2.2.Criterios de clasificación en sociología.
 - 2.3.Alcance y contenidos de la clasificación de tipologías: realistas, analíticas, tipo ideal, el tipo concreto.
- 3.-Tipo concreto y su construcción.
 - 3.1.Tipo concreto.
 - 3.2.Técnicas empíricas de construcción.
- 4.-Particiones.
 - 4.1.Partición.
 - 4.2.Orden de una partición, cadena, numeración y contabilidad de las particiones.
 - 4.3.Ultramétrica de una partición.
- 5.-Clasificación automática.
 - 5.1.Generalidades y nociones. Elecciones a realizar.
 - 5.2.Cadena binaria.
 - 5.3.Criterios de proximidad.
 - 5.3.1.Definición de distancias.
 - 5.3.2.Tipos de distancias.
 - 5.3.2.1.Variables cuantitativas.
 - a) Individuos
 - b) Variables.
 - 5.3.2.2.Tablas binarias.
 - a) Individuos

- b) Variables.
- 5.3.2.3. Tablas nominales.
 - a) Individuos
 - b) Variables.
- 5.3.2.4. Ordinales.
 - a) Individuos.
 - b) Variables.
- 5.4. Estrategia de clasificación. Medida de distancia entre individuos/grupos. Algoritmos de agrupación. Índices de agregación.
 - 5.4.1. Distancia max., mín., media.
 - 5.4.1.1. Lazo mínimo.
 - 5.4.1.2. Lazo máximo.
 - 5.4.1.3. Lazo medio.
 - 5.4.2. Distancia entre centroides.
 - 5.4.2.1. Definición de centroíde.
 - 5.4.2.2. Criterios de agrupamiento. Centroíde del agrupamiento.
 - 5.4.3. Inercia interclase/ inercia intraclase.
 - 5.4.3.1. Definición de inercia.
 - 5.4.3.2. Índice de aumento mínimo de inercia intraclase.
 - 5.4.3.3. Índice de aumento mínimo de varianza.
 - 5.4.4. Wilian y Lambert.
 - 5.4.5. El método de K-medias.
- 6.- Clasificación no jerárquica.
 - 6.1. Agregación sobre centros móviles. Centroídes. K-medias.
 - 6.2. Inercia interclase/inercia intraclase.
 - 6.3. Nubes dinámicas.
- 7.- Clasificación jerárquica.
 - 7.1. Ascendente. Aglomerativa.
 - 7.1.1. Pérdida de inercia. Ward.
 - 7.1.2. Aglomeración según diversos tipos de salto (algoritmos)
 - 7.2. Descendente. Divisiva. Wilian y Lambert.
- 8.- Ejercicios e interpretaciones.

IV METODOS FACTORIALES

Tema 5: Análisis de Componentes Principales.

- 1.- Introducción matemática. Parámetros.
- 2.- Naturaleza de las componentes principales. Inercia. Criterios en la elección del número de ejes.
- 3.- El peso de las variables y la estructuración del espacio.
- 4.- Interpretación de los ejes.

5.-Ejemplos e interpretaciones.

Tema 6: Análisis de correspondencias.

1.-Introducción matemática.

2.-ACP y Análisis de Correspondencias.

3.-Número de ejes . Criterios.

4.-Interpretación de los ejes.

5.-Ejemplos e interpretaciones.

Tema 7: Objetivos y planteamiento general del análisis discriminante.

PARTE V:INTRODUCCION AL ANALISIS DE ESTRUCTURAS CAUSALES.

**Tema 8: Objetivos y planteamiento general del análisis causal.
Estructuras causales.**

PARTE VI:ANALISIS DIACRONICO EN SOCIOLOGIA

Tema 9:Introducción general al análisis diacrónico.

PARTE VII:ANALISIS CUANTITATIVO DE CONTENIDO

**Tema 10:Introducción general al análisis frecuencial de
contenido.Lexicometría.**