

Anàlisi multivariant

Curs 2007-2008

Joan Valls i Marsal

Programa

1. Introducció. Dos exemples introductoris: “la despesa publicitària a Europa” i “les feines d’assessorament del Servei d’estadística de la UAB”.
2. Anàlisi de components principals
 - Les dades. Objectius de l’estudi.
 - La transformació de les dades.
 - El núvol d’individus.
 - El núvol de variables.
 - L’ajust del núvol d’individus.
 - L’ajust del núvol de variables.
 - Dualitat i fórmules de transició en ACP.
 - Ajudes a la interpretació.
 - Exemples.
3. Anàlisi de correspondències
 - Les dades, les notacions, la hipòtesi d’independència.
 - Els objectius.
 - La transformació de les dades en perfils.
 - La similitud entre perfils: la distància de khi-quadrat.
 - Els núvols de perfils-fila i de perfils-columna.
 - L’ajust dels núvols de perfils-fils i perfils-columna.
 - La dualitat. L’estadístic de khi-quadrat i la inèrcia dels dos núvols.
 - Exemples.
4. Anàlisi de correspondències múltiples.
 - Dades i notacions. La codificació condensada. La taula disjuntiva complerta. La hipertaula de contingència. La taula de Burt.
 - Objectius. Estudi dels individus. Estudi de les variables. Estudi de les modalitats.
 - Anàlisi de correspondències d’una taula disjuntiva complerta.
 - Anàlisi de correspondències d’una taula de Burt.
 - Consideracions.
 - Exemples.

5. Tècniques de classificació.
 - Introducció.
 - Mesures de proximitat entre objectes. Similaritats, dissimilaritats, distància euclídea, distància de Mahalanobis, distància Manhattan, ...
 - Mesures de proximitats entre grups. Single linkage o veí més proper, Complete linkage o veí més llunyà, Average linkage, ...
 - Mètodes jeràrquics de classificació
 - Altres mètodes de classificació.
 - Combinació dels mètodes factorials amb algorismes de classificació.
 - Exemples.
6. Anàlisi discriminant.
 - Introducció.
 - La funció discriminant lineal.
 - Poblacions normals i poblacions no normals.
 - Anàlisi discriminant per a més de dos grups.
 - Exemples.
7. Introducció a altres tècniques multivariants.
 - Multidimensional scaling.
 - Arbres de regressió.
 - Support vector machine.
 - Xarxes neurals.
 - Anàlisi de dades textuais.
 - Exemples.

Bibliografia

- Escofier, B. i Pagès, J. Anàlisis factoriales simples y múltiples. Objetivos, métodos e interpretación. Servicio editorial Universidad País Vasco. 1990.
- Jobson, J.D. Applied multivariate data analysis, Volume I and II. Springer. 1992.
- Ripley, B.D. Pattern Recognition and Neural Networks. Cambridge University Press. 1996.
- Hastie, T. I Tibshirani, R. The elements of statistical learning: data mining, inference and prediction. Springer. 2001.
- Greenacre, M.J, Correspondence Analysis in Practice. Academic Press, London. 1993
- Aluja, T. i Morineau, A. Aprender de los datos: el análisis de componentes principales. Una aproximación desde el data mining. EUB ediciones. 1999.
- Dalgaard, P. Introductory Statistics with R. Springer. 2002.
- Venables, W.N. i D.M. Smith. An introduction to R. www.R-project.org. 2003.
- Everitt, B. I S. Rabe-Hesketh. Analyzing medical data using S-PLUS. Springer. 2002.

Pràctiques

Es realitzaran dues hores setmanals de pràctiques. S'emprarà, principalment, el paquets estadístics SPAD i R.

Projecte d'anàlisi de dades

Els alumnes hauran de cercar un conjunt de dades de temàtica lliure, amb l'objectiu doble d'aplicar les eines de l'anàlisi multivariant de dades i de presentar correctament els resultats obtinguts en un informe i extreure'n conclusions. Per a avaluar-lo es tindrà en compte:

- 1) Aplicació correcta de les tècniques estadístiques.
- 2) Originalitat i interès del problema i la seva resolució.
- 3) Claredat i brevetat en la presentació i redacció de resultats. Es penalitzaran aquells informes que siguin innecessàriament llargs (més de 15 pàgines).

Existirà la possibilitat d'una avaluació preliminar.

Avaluació

60% examen + 20% projecte d'anàlisi de dades + 20% entrega pràctiques /exercicis

L'examen inclourà una part pràctica que s'haurà de resoldre amb l'ajut del paquet estadístic R i SPAD.

Consultes

A concertar: joan.valls@iconcologia.net

Cal que els alumnes que segueixin l'assignatura virtualment es posin en contacte amb el professor. Regularment s'aniran penjant de la pàgina web de l'assignatura materials diversos.

Horari i calendari

Les classes de teoria es realitzaran els divendres de 15:00 a 18:00. Les classes de pràctiques amb ordinador es realitzaran els dilluns de 18:00 a 20:00. Les classes de problemes es realitzaran els dilluns de 17:00 a 18:00.