

## Anàlisi de Dades Categòriques

2013/2014

Codi: 103208

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2501919 Estadística Aplicada	OB	3	1

### Professor de contacte

Nom: Joan Valls Marsal

Correu electrònic: Joan.Valls@uab.cat

### Utilització d'idiomes

Llengua vehicular majoritària: català (cat)

Algun grup íntegre en anglès: No

Algun grup íntegre en català: Sí

Algun grup íntegre en espanyol: No

### Prerequisits

Inferència Estadística, Estadística Descriptiva

### Objectius

Conèixer i aprofundir en els tècniques estadístiques específiques per a dades categòriques

### Competències

- Estadística Aplicada
- Analitzar les dades mitjançant l'aplicació de mètodes i tècniques estadístiques i treballar amb dades qualitatives i quantitatives.
- Demostrar iniciativa i inquietud per tal d'actualitzar els coneixements.
- Reconèixer els avantatges i els inconvenients dels procediments estudiats.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Utilitzar correctament una bona part del programari estadístic i de recerca operativa existent, escollir el més apropiat per a cada anàlisi estadística i ser capaç d'adaptar-lo a les noves necessitats.

### Resultats d'aprenentatge

1. Analitzar dades mitjançant el model de regressió logística.
2. Demostrar iniciativa i inquietud per tal d'actualitzar els coneixements.
3. Identificar distribucions de dades temporals i categòriques.
4. Identificar les suposicions estadístiques associades a cada procediment avançat.
5. Identificar, emprar i interpretar els criteris per avaluar el grau de compliment dels requisits necessaris per aplicar cada procediment avançat.
6. Utilitzar els residus per veure l'adequació d'un o altre model (regressió logística, sèries, etc.)
7. Utilitzar programari estadístic per a l'anàlisi de dades categòriques.

### Continguts

1. Introducció
  1. Breu història de l'anàlisi de dades categòriques
  2. Tipus de dades categòriques

2. Descripció estadística de taules de contingència bidimensionals
  1. Taules de contingència
  2. Independència
  3. Comparació de proporcions: diferència de proporcions, risc relatiu, odds ratio (raó d'avantatges). Relació entre la raó d'avantatges i el risc relatiu.
  4. Mesures d'associació ordinals: Gamma de Goodman i Kruskal, Q de Yule, Tau-b de Kendall, d de Sommers.
  5. Mesures d'associació nominals: Tau de Goodman i Kruskal o coeficient de concentració, coeficient d'incertesa.
  6. Mesures d'acord. Kappa de Cohen.
  7. Anàlisi de correspondències simples. Distància entre files i columnes. Representació en biplots.
3. Inferència per a taules de contingència bidimensionals
  1. Distribucions mostrals: distribució de Poisson, distribució multinomial, distribució multinomial independent. Estudis prospectius i retrospectius.
  2. Funcions de versemblança i estimació màxim versemblant.
  3. Sobre dispersió
  4. Testos de bondat d'ajust: per a una multinomial, exemple de Mendel, bondat d'ajust amb estimació de les freqüències esperades.
  5. Testos d'independència: test de la khi-quadrat de Pearson, test de la raó de versemblances, invariància de la khi-quadrat per ordenacions de categories, particions de la khi-quadrat.
  6. Interval de confiança per a mostres grans: estimació de la raó d'avantatges, diferència de proporcions i risc relatiu.
  7. Testos exactes per a mostres petites. El test exacte de Fisher. Altres testos.
4. La paradoxa de Simpson
  1. Estructura de taula per a tres dimensions.
  2. Associació parcial i marginal
  3. Exemple de la pena de mort
5. Models de regressió logística
  1. El model de regressió logística simple. Ajust del model. Testos per a la significació dels coeficients. Altres mètodes d'estimació
  2. El model de regressió logística múltiple. Ajust del model. Testos per a la significació del model. Altres mètodes d'estimació.
  3. Interpretació dels coeficients dels models de regressió logística. Una variable explicativa dicotòmica. Una variable explicativa politòmica. Una variable explicativa contínua. Combinacions multivariants. Interaccions i confusió. Estimació de odds-ratios en presència d'interaccions.
  4. Estratègies per a la construcció del model. Selecció de variables. Mètode stepwise i mètode best subset.
  5. Avaluació de l'ajust del model. Mesures de bondat de l'ajust: khi-quadrat de pearson i deviança, el test de Hosmer-Lemeshow i altres.

## Metodologia

Classes teòriques i pràctiques amb R

## Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes teòriques i pràctiques	130	5,2	1, 2, 3, 4, 5, 6, 7

## Avaluació

Es realitzaran dos exàmens parcials de teoria i un examen parcial pràctic de R (data a determinar) i també hi hauran tres entregues de problemes, de forma que es podrà avaluar l'assignatura de forma contínua. Per

aquells que no hagin superat l'avaluació hi haurà un examen final de recuperació. Per al còmput de la qualificació final es tindrà en compte els percentatges següents:

40% exàmens teòrics + 20% examen pràctic + 20% projecte d'anàlisi de dades + 20% entrega problemes

### Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
exàmens parcials i treballs	100%	20	0,8	1, 2, 3, 4, 5, 6, 7

### Bibliografia

Agresti, A. Categorical data analysis. Wiley. 1990.

Hosmer, D.W. i S. Lemeshow. Applied Logistic Regression. Wiley. 1989.

Simonoff, J.S. Analyzing categorical data. Springer. 2003.

Greenacre, M.J, Correspondence Analysis in Practice. Academic Press, London. 1993

R Development Core Team (2003). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>.

Dalgaard, P. Introductory Statistics with R. Springer. 2002.

Venables, W.N. i D.M. Smith. An introduction to R. [www.R-project.org](http://www.R-project.org). 2003.

Everitt, B. I S. Rabe-Hesketh. Analyzing medical data using S-PLUS. Springer. 2002.