

Anàlisi de Dades Categòriques

Codi: 103208
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2501919 Estadística Aplicada	OB	3	1

Professor/a de contacte

Nom: Anna Espinal Berenguer

Correu electrònic: Anna.Espinal@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: català (cat)

Grup íntegre en anglès: No

Grup íntegre en català: Sí

Grup íntegre en espanyol: No

Equip docent

Joan Valls Marsal

Prerequisits

Inferència Estadística, Estadística Descriptiva

Objectius

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

Competències

- 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.
- Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
- Desenvolupar estratègies d'aprenentatge autònom.
- Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
- Desenvolupar un pensament i un raonament crítics i saber comunicar-los de manera efectiva, tant en les llengües pròpies com en una tercera llengua.
- Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes
- Identificar la utilitat i la potencialitat de l'estadística en les diferents àrees de coneixement i saber aplicar-la adequadament per a extreure'n conclusions rellevants.
- Interpretar resultats, extreure conclusions i elaborar informes tècnics.
- Reconèixer els avantatges i els inconvenients dels procediments estudiats.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Utilitzar bibliografia o eines d'Internet específiques de l'estadística i d'altres ciències afins, tant en llengua anglesa com en la llengua pròpia.
- 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. Caracteritzar grups homogenis d'individus mitjançant l'anàlisi multivariant.
3. Conèixer l'existència de recursos específics per a l'ús i aprenentatge del programari estadístic.
4. Conèixer programari estadístic per programar funcions i procediments avançats.
5. Demostrar iniciativa i inquietud per tal d'actualitzar els coneixements.
6. Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
7. Desenvolupar estratègies d'aprenentatge autònom.
8. Desenvolupar l'interès pel coneixement i la interpretació de fenòmens susceptibles de quantificació.
9. Desenvolupar un pensament i un raonament crítics i saber comunicar-los de manera efectiva, tant en les llengües pròpies com en una tercera llengua.
10. Emprar índexs de resum de dades multivariants, sèries temporals i totes les altres tècniques avançades.
11. Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes.
12. Extreure conclusions rellevants a problemes aplicats, mitjançant l'aplicació de mètodes estadístics avançats.
13. Fer consultes bibliogràfiques en l'àmbit dels mètodes avançats, també en llengua anglesa.
14. Identificar distribucions de dades temporals i categòriques.
15. Identificar l'existència de recursos específics d'aplicació de metodologies estadístiques i practicar-hi.
16. Identificar l'existència de recursos específics per a l'ús i l'aprenentatge del programari estadístic i familiaritzar-s'hi.
17. Identificar les característiques metodològiques especials en l'anàlisi estadística segons les diferents àrees d'aplicació.
18. Identificar les diferents àrees d'aplicació més comunes per a cada metodologia avançada.
19. Identificar les etapes en els problemes que requereixen tecnologies avançades.
20. Identificar les suposicions estadístiques associades a cada procediment avançat.
21. Identificar, emprar i interpretar els criteris per avaluar el grau de compliment dels requisits necessaris per aplicar cada procediment avançat.
22. Interpretar resultats amb metodologies avançades i extreure'n conclusions.
23. Redactar l'informe tècnic a partir d'una anàlisi estadística avançada.
24. Utilitzar els residus per veure l'adequació d'un o altre model (regressió logística, sèries, etc.)
25. Utilitzar programari estadístic per a l'anàlisi de dades categòriques.
26. Utilitzar programari estadístic per als diversos mètodes d'anàlisi multivariant.

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 tests.
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, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26

Avaluació

Es realitzaran dos exàmens parcials de teoria i pràctica de R (data a determinar) i també s'haurà d'elaborar un petit projecte d'anàlisi de dades, 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.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen, primer parcial	30%	5	0,2	3, 7, 9, 11, 12, 14, 18, 20, 22, 25, 26
Examen, segon parcial	40%	10	0,4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26
Treball d'Anàlisi de dades	30%	5	0,2	1, 2, 4, 6, 11, 12, 19, 20, 21, 22, 23, 24, 25

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. 2003.

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