

Disseny dExperiments

Codi: 104862
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503852 Estadística Aplicada	OB	2	1

Professor/a de contacte

Nom: Mercè Farré Cervelló
Correu electrònic: Merce.Farre@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

Alejandro Lerer Gornatti

Prerequisits

Fonaments d'estadística descriptiva i inferencial i de probabilitats, així com coneixer els rudiments de programació amb el llenguatge R.

Objectius

L'objectiu del curs és l'estudi de la modelització i l'anàlisi de dades mitjançant la teoria dels Models Lineals, així com les aplicacions a diversos àmbits (economia, salut, enginyeria, i ciències en general). Els mètodes i tècniques s'introdueixen en base a exemples i es treballen a partir de la resolució dels problemes proposats i de pràctiques d'ordinador pensades per ser executades amb el llenguatge R. En primer lloc, es presenta el model de regressió simple perquè té nombroses aplicacions i perquè és un bon pròleg per a la comprensió del model múltiple. El model de regressió múltiple, expressat matricialment i incloent algunes variants (polinomial, amb interaccions, utilitzant variables regressores fictícies, etc.), constitueix la segona part del curs. En tots els procediments de modelització s'analitzen l'ajust i l'especificació correcta del model, la satisfacció de les hipòtesis, la detecció de dades "especials"(anòmales i influents), i s'estudien possibles solucions quan es detecten anomalies.

Competències

- Analitzar dades mitjançant l'aplicació de mètodes i tècniques estadístiques, treballant amb dades de diverses tipologies.
- Aplicar l'esperit crític i el rigor per validar o refutar arguments tant propis com d'altres persones.
- Avaluar de manera crítica i amb criteris de qualitat el treball realitzat.
- Dissenyar un estudi estadístic o de recerca operativa per a la resolució d'un problema real.
- Formular hipòtesis estadístiques i desenvolupar estratègies per confirmar-les o refutar-les.
- Interpretar resultats, extreure conclusions i elaborar informes tècnics en el camp de l'estadística.
- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.

- Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'índole social, científica o ètica.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.
- Seleccionar els models o tècniques estadístiques per aplicar-los en estudis i problemes reals, així com conèixer-ne les eines de validació.
- Seleccionar i aplicar procediments més apropiats per a la modelització estadística i l'anàlisi de dades complexes.
- Utilitzar correctament un ampli espectre del programari i llenguatges de programació estadístiques, escollint el més apropiat per a cada anàlisi i ser capaç d'adaptar-lo a noves necessitats.
- Utilitzar eficaçment la bibliografia i els recursos electrònics per obtenir informació.

Resultats d'aprenentatge

1. Analitzar dades mitjançant el model d'anàlisi de la variància d'un o diversos factors.
2. Analitzar dades mitjançant el model de regressió lineal.
3. Analitzar dades mitjançant tècniques d'inferència utilitzant programari estadístic.
4. Analitzar els residus d'un model estadístic.
5. Aplicar l'esperit crític i el rigor per validar o refutar arguments, tant propis com d'altres.
6. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
7. Comparar el grau d'ajustament entre diversos models estadístics.
8. Establir les hipòtesis experimentals de la modelització.
9. Extreure conclusions de l'adequació dels models amb la utilització i la interpretació correcta d'indicadors i gràfics.
10. Fer servir gràfics de visualització de l'ajustament i de l'adequació del model.
11. Identificar distribucions de les respostes amb l'anàlisi de residus.
12. Identificar fonts de biaix en l'obtenció de la informació.
13. Identificar la presència d'interacció entre variables mitjançant gràfics de mitjanes i interaccions.
14. Identificar les etapes en els problemes de modelització.
15. Identificar les suposicions estadístiques associades a cada procediment.
16. Identificar les variables resposta, explicatives i de control.
17. Mesurar el grau d'ajustament d'un model estadístic.
18. Modificar lleugerament el programari existent si el model estadístic proposat ho requereix.
19. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
20. Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
21. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'índole social, científica o ètica.
22. Seleccionar les variables explicatives rellevants.
23. Sintetitzar i interpretar els resultats dels models lineals clàssics, generalitzats i no lineals en funció de l'objectiu de l'estudi.
24. Utilitzar eficaçment bibliografia i recursos electrònics per obtenir informació.
25. Utilitzar programari estadístic divers per ajustar i validar models lineals i les seves generalitzacions.
26. Validar els models utilitzats mitjançant tècniques d'inferència adequades.

Continguts

1. El model de egressió lineal simple

- Introducció als models de regressió i passos previs en la regressió simple: Exploració de les dades.

- La regressió lineal simple: Model, hipòtesis i paràmetres.
- Estimació puntual dels paràmetres del model: Mètode de mínims quadrats. Els estimadors de màxima versemblança.
- Inferència sobre els paràmetres del model sota les hipòtesis de Gauss-Markov: Intervals i tests.
- Interval de confiança per a la resposta mitjana i interval de predicció de noves observacions. Inferències simultànies. Bandes de confiança i de predicció.
- Anàlisi de la variància (ANOVA) del model de regressió simple.
- Diagnòstics del model: Avaluació gràfica de la linealitat i de la verificació de les hipòtesis del model mitjançant l'anàlisi dels residus. Test de manca d'ajust lineal.
- Dades anòmales o influents.

2. El model de regressió lineal múltiple

- Passos previs en la regressió múltiple: Exploració de les dades amb eines de visualització multidimensional.
- Expressió matricial del model i dels estimadors. Interpretació dels coeficients del model lineal múltiple.
- Lleis dels estimadors dels coeficients, de les prediccions i dels residus: Aplicació de les propietats de les matrius idempotents.
- Inferència en el model lineal múltiple. Anova del model.
- Test de "lligadures" per resoldre restriccions lineals sobre els coeficients: El principi de la variabilitat incremental.
- Discussió de les hipòtesis del model lineal: Anàlisi dels residus. Transformacions de Box-Cox i altres.
- El problema de la multi-colinealitat entre variables regressores: Detecció i solucions.
- Variables fictícies en regressió (dummies): Interpretació dels coeficients i aplicacions.
- Selecció de variables en un model lineal: L'estadístic Cp de Mallows, la validació creuada de models i els mètodes de selecció automàtica per passos.

Metodologia

L'assignatura té dues hores setmanals de teoria i problemes on es presenten i motiven les eines i mètodes de models lineals i es fa un treball analític. Es proporcionaran llistes de problemes al llarg del curs, que es lliuraran. A més, hi ha dues hores setmanals de sessions pràctiques que es duran a terme en una aula d'informàtica. En aquestes sessions, els estudiants practiquen els mètodes introduïts a teoria a través de la implementació en el llenguatge de programació R. Al final de cada sessió, es proposaran algunes tasques per lliurar. A més dels lliuraments, l'estudiant també realitzarà altre treball autònom consistent en recerca bibliogràfica i preparació d'exàmens.

El material del curs (notes de teoria, llistes de problemes i tasques d'ordinador) estarà disponible a l'aula *moodle*.

La perspectiva de gènere va més enllà dels continguts dels cursos, ja que implica també una revisió de les metodologies i les interaccions entre els estudiants i els professors, tant dins com fora de l'aula. En aquest sentit, les metodologies participatives d'ensenyament que donen lloc a un entorn d'igualtat, menys jeràrquiques a l'aula, evitant exemples estereotipats en el gènere i el vocabulari sexista, solen ser més favorables a la plena integració i participació de les alumnes. Per això, es farà aquesta aplicació efectiva durant el curs.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de pràctiques	26	1,04	2, 3, 4, 7, 9, 10, 11, 13, 15, 18, 22, 24, 25
Classes de teoria	26	1,04	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 22, 26
Tipus: Autònomes			
Estudi i consultes	36	1,44	5, 6, 24
Resolució d'exercicis amb R	32	1,28	1, 2, 3, 4, 7, 9, 11, 12, 13, 17, 20, 21, 22, 23, 25, 26
Resolució de problemes	18	0,72	1, 2, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 23, 26

Avaluació

PR: Lliurament dels exercicis teòrics i pràctics (amb R). Valoració màxima de PR: 2 punts. Aquesta part no és recuperable.

P1: Prova parcial de regressió simple (teoria, exercicis, i pràctiques). Valoració màxima de P1: 3 punts.

P2: Prova parcial de regressió múltiple (teoria, exercicis i pràctiques). Valoració màxima de P2: 5 punts.

La nota de curs es calcularà: **NC = PR + P1 + P2**. L'aprobat per curs requereix que NC sigui igual o més gran que 5 i que les notes de cada parcial siguin més grans o iguals que 3.5 (sobre 10).

Al final del semestre es farà un examen de recuperació que serà una prova de síntesi, PS, (teoria, exercicis i pràctiques) dels continguts de tot el curs amb una puntuació màxima de 8 punts, pels alumnes que no hagin aprovat per curs o vulguin millorar la nota. Només es podran presentar a la prova de síntesi els estudiants que hagin participat a 2/3 parts de les activitats d'avaluació continuada.

La nota final dels presentats a la prova de síntesi es calcularà: **NF = PR + max(PS, P1+P2)**.

Las matrícules d'honor que eventualment s'atorguin a partir de la NC no es retiraran fins i tot si un altre estudiant obté una nota superior després de la PS.

Atenció: "Sense perjudici d'altres mesures disciplinàries que s'estimin oportunes, i d'acord amb la normativa acadèmica vigent, es qualificaran amb un zero les irregularitats comeses per l'estudiant que puguin conduir a una variació de la qualificació d'un acte d'avaluació. Per tant, plagiar, copiar o deixar copiar una pràctica o qualsevol altra activitat d'avaluació implicarà suspendre-la amb un zero i no es podrà recuperar en el mateix curs acadèmic. Si aquesta activitat té una nota mínima associada, aleshores l'assignatura quedarà suspesa."

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen final	80% (recupera els dos parcials)	4	0,16	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 23, 25, 26

Lliurament de tasques (problemes i pràctiques resoltes)	20%	0	0	5, 6, 9, 18, 19, 20, 21, 24, 25
Primer parcial	30%	4	0,16	2, 3, 4, 8, 9, 10, 11, 13, 15, 16, 17, 25, 26
Segon parcial	50%	4	0,16	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 22, 23, 25, 26

Bibliografia

Montgomery, D. Peck, A. Vining, G.; Introduction to Linear Regression Analysis. Wiley, 2001.

Clarke, B.R.; Linear Models: The Theory and Applications of Analysis of variance. Wiley, 2008.

Christopher Hay-Jahans; An R Companion to Linear Statistical Models. Chapman and Hall, 2012.

Fox, J. and Weisberg, S.; An R Companion to Applied Regression. Sage Publications, 2nd edition, 2011.

Peña, D.; Regresión y diseño de Experimentos. Alianza Editorial (Manuales de Ciencias Sociales), 2002.

Bibliografia complementària:

Sen, A., Srivastava, M.; Regression Analysis: Theory, Methods and Applications. Springer, 1990.

Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman; .Applied Linear Models. Irwin (4th edition), 1996.

Faraway, J.; Linear Models with R. Chapman&Hall/CRC (2nd ed), 2014.

Rao, C. R., Toutenburg, H., Shalabh, Heumann, C; Linear Models and generalizations. Springer, 2008.