

Anàlisi Multivariant

Codi: 100122
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2500149 Matemàtiques	OT	4	0
2501919 Estadística Aplicada	OB	3	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

Prerequisits

És imprescindible un curs previ d'Àlgebra Lineal, així com cursos bàsics de Probabilitats i d'Inferència Estadística. També se suposa que sap utilitzar amb agilitat el programari R.

Objectius

La majoria de dades recollides són multivariants, és a dir, per a una mateixa unitat experimental (que pot ser un objecte de naturalesa complexa) observem de manera simultània els valors de diverses variables. L'Anàlisi Multivariant tracta els mètodes que són més apropiats per a descriure, explorar i modelitzar aquestes dades vectorials, així com per a aplicar-hi la inferència estadística.

La necessitat de processar gran quantitat de dades amb moltes variables de naturalesa diversa, amb la voluntat de reduir la informació que no sigui rellevant i de descobrir patrons d'associació entre variables, així com l'interès en la classificació supervisada o autònoma de casos, han fet desenvolupar recentment un gran nombre de procediments que se situen en l'escenari multivariant.

L'assignatura es planteja com un primer contacte de l'estudiant amb el món de l'anomenat "aprenentatge estadístic" (*statistical learning*), per tal que compregui la potència i aplicabilitat, i alhora les limitacions, dels mètodes, alguns dels quals es basen en idees heurístiques molt intuïtives. L'enfocament de l'assignatura és eminentment aplicat pel que fa al treball amb dades usant el potencial del programari lliure R, acompanyat del rigor i generalitat adequats en la definició del models teòrics i els corresponents mètodes d'anàlisi i validació de resultats.

Competències

Matemàtiques

- Davant de situacions reals amb un nivell mig de complexitat, demanar i analitzar dades i informació rellevants, proposar i validar models utilitzant eines matemàtiques adequades per a, finalment, obtenir conclusions
- Formular hipòtesis i imaginar estratègies per confirmar-les o refutar-les.
- Utilitzar aplicacions informàtiques d'anàlisi estadística, càlcul numèric i simbòlic, visualització gràfica, optimització o altres per experimentar en Matemàtiques i resoldre problemes

Estadística Aplicada

- Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
- Dissenyar un estudi estadístic o de recerca operativa per a la resolució d'un problema real.
- Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes
- Identificar els models estadístics i de recerca operativa més adequats per a cada context i que permetin la presa de decisions.
- 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. Caracteritzar grups homogenis d'individus mitjançant l'anàlisi multivariant.
2. Demostrar que es té un pensament lògic, un raonament estructurat i capacitat de síntesi.
3. Emprar gràfics de resum de dades multivariades i d'evolució temporal.
4. Emprar índexs de resum de dades multivariants, sèries temporals i totes les altres tècniques avançades.
5. Expressar i aplicar rigorosament els coneixements adquirits en la resolució de problemes.
6. Identificar les etapes en els problemes que requereixen tecnologies avançades.
7. Projectar un estudi a partir de metodologies multivariants i/o una mineria de dades per resoldre un problema contextualitzat en la realitat experimental.
8. Reconèixer la necessitat d'emprar mètodes multivariants en lloc de bivariants.
9. Utilitzar programari estadístic per als diversos mètodes d'anàlisi multivariant.

Continguts

Dades multivariants: Aprenentatge estadístic i reducció de la dimensió.

- L'aprenentatge (*statistical learning*) supervisat i no-supervisat. Classificació dels mètodes multivariants. Exemples.
- Vectors aleatoris. El vector d'esperances i les matrius de covariàncies i correlacions. Propietats.
- Dades multivariants. El vector de mitjanes i les matrius de covariància i correlació mostrals. Estimadors de màxima versemblança en el cas Gaussià.
- Elements bàsics del càlcul matricial. Les descomposicions espectral (SD) i en valors singulars (SVD).
- Teorema fonamental de maximització de formes quadràtiques.

Mètodes factorials I : Anàlisi de components principals (PCA).

- Introducció al PCA: Definició de components. El resultat fonamental.
- Criteris per decidir el nombre de components: Les Components principals.
- Representació de les variables i de les dades individuals. Estandarditzacions.
- Anàlisi de les matrius per files i per columnes.
- Interpretació geomètrica de les components principals.

Mètodes factorials II: Anàlisi factorial (FA).

- El model factorial. Comunalitats i especificitats.
- Teorema fonamental: Descomposició de la matriu de covariàncies.
- Discussió de l'existència i la unicitat del model factorial.
- Rotacions.
- Mètodes d'estimació dels paràmetres del model.
- Estimació o predicció de les puntuacions factorials (*scores*).
- Interpretació dels resultats. Comparació de PCA i FA.

Els mètodes factorials III: Escalament multidimensional (MDS) i anàlisi de correspondències (CA).

- Objectius i tipus d'escalament multidimensional.
- Escalament multidimensional clàssic i mètric.

- Escalament multidimensional no- mètric.
- Distàncies, proximitats i dissimilaritats.
- Dades categòriques: Distància *khi-quadrat* i altres dissimilaritats
- Anàlisi de correspondències: Perfils i inèrcies.
- Descomposicions de la inèrcia.
- Representacions gràfiques i interpretació dels resultats de CA.

Anàlisi de clústers (CLA).

- Presentació comparativa de diversos enfocaments: Exemples.
- Anàlisi i validació dels resultats de la classificació.
- Classificació jeràrquica: Funcions d'enllaç.
- Mètodes basats en centres: El mètode de les K- mitjanes.
- Mètodes basats en models: Esperança i maximització (EM).

Elements d'inferència multivariant

- El test de la raó de versemblança.
- Tests per a vectors d'esperances.
- Tests per a matrius de covariàncies.
- El MANOVA com una generalització de l' ANOVA.

Anàlisi Discriminant (DA) i altres mètodes supervisats.

- Objectius de l'anàlisi discriminant: Criteris de discriminació.
- Anàlisi discriminant en un model normal.
- Anàlisi discriminant canònica de Fisher.
- Regressió de mínims quadrats parcials (PLS).
- Altres mètodes.

Metodologia

Classes de teoria, on s'exposaran detalladament i es discutiran els continguts en base a exemples d'aplicació convenientment interpretats. Els resums dels apunts es penjaran al campus virtual. S'encoratjarà la revisió i ampliació de continguts amb la bibliografia bàsica i complementària.

Classes pràctiques amb software estadístic R. Els guions de les pràctiques, enunciats de problemes, i l'altre material auxiliar es penjaran al Campus Virtual. A les classes pràctiques es proposaran exercicis d'ampliació que caldrà resoldre autònomament.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de pràctiques	28	1,12	1, 2, 3, 4, 6, 7, 8, 9
Classes de teoria	28	1,12	3, 4, 6, 8
Tipus: Supervisades			
Estudi de casos	12	0,48	2, 3, 6, 8, 9
Tipus: Autònomes			
Estudi	40	1,6	3, 4, 6, 8, 9

Presentació d'informes de pràctiques i problemes	30	1,2	1, 2, 3, 4, 5, 6, 7, 8, 9
--	----	-----	---------------------------

Avaluació

La nota de curs (NC) es calcularà a partir de les entregues d'exercicis i pràctiques (10%), i de les qualificacions de proves parcials (P1 i P2), tant de teoria (45%) com de pràctiques (45%). Excepte pels alumnes virtuals, l'assistència a pràctiques és obligatòria per a tenir nota de curs. Detall de la nota de curs:

$$NC=0,2 \cdot P1t+0,2 \cdot P2p+0,25 \cdot P2t+0,25 \cdot P2p+0,10 \cdot Lli$$

on P1t i P1p corresponen a la nota del primer parcial, parts de teoria - problemes i de pràctiques respectivament, i el mateix pel segon parcial, i Lli és la nota dels lliuraments durant tot el semestre.

Si NC és igual o superior a 5, l'alumne haurà superat l'assignatura, sense necessitat d'anar a l'examen de recuperació. L'examen final de recuperació (F) serà una prova de síntesi de tot el curs i tindrà una part teòrica i una de pràctica, que faran mitjana. Si un alumne vol millorar la nota de curs, es pot presentar a la recuperació i la seva nota final es calcularà amb la fórmula:

$$NF=Max(F , 0,50 \cdot NC+0,50 \cdot F)$$

No obstant, el professor es reserva el dret a decidir les matrícules d'honor abans de l'examen de recuperació.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Entregues d'informes de pràctiques i problemes.	10%	4	0,16	1, 2, 3, 4, 5, 6, 7, 8, 9
Exàmens parcials teòrics	45%	4	0,16	1, 2, 3, 4, 5, 8, 9
Proves pràctiques	45%	4	0,16	1, 2, 3, 4, 6, 7, 8, 9

Bibliografia

Bibliografia bàsica

Everitt, B., Hothorn, T. : An introduction to Applied Multivariate Analysis with R. Springer. 2011 (Llibre específic per a les pràctiques que conté breus resums de teoria).

Härdle, W., Simar, L.: Applied Multivariate Statistical Analysis. Springer. 2007.

Peña, D.: Análisis de datos multivariantes. McGraw Hill. 2002.

Rencher, A., Christensen, W.: Methods of Multivariate Analysis. Wiley Series in Probability and Mathematical Statistics, 2012.

Bibliografia complementària.

Cuadras, C.: Nuevos Métodos de Análisis Multivariante (web), 2014

Greenacre, M.: La pràctica del anàlisi de correspondències. Fundacion BBA, 2003

James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. Springer, 2014

Mardia, K.V, Kent, J.T., Bibby, J.M.: Multivariate Analysis. Academic Press, 2003.

Rencher, A.: Multivariate Statistical Inference and Applications. John Wiley & Sons, 1998.

Little book of R for Multivariate Analysis. <https://little-book-of-r-for-multivariate-analysis.r...> 2015.