

200606 - AMD - Anàlisi Multivariant de Dades

Unitat responsable: 200 - FME - Facultat de Matemàtiques i Estadística
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa
1004 - UB - Universitat de Barcelona
Curs: 2019
Titulació: MÀSTER UNIVERSITARI EN ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (Pla 2013). (Unitat docent Optativa)
Crèdits ECTS: 5 Idiomes docència: Castellà, Anglès

Professorat

Responsable: JAN GRAFFELMAN
Altres: Segon quadrimestre:
JAN GRAFFELMAN - A
FERRAN REVERTER COMES - A
MIQUEL SALICRÚ PAGES - A

Capacitats prèvies

1. El curs pressuposa coneixements d'àlgebra lineal: diagonalització de matrius simètriques, projecció de vectors, derivació vectorial de funcions lineals i quadràtiques.
2. També cal haver fet un curs d'inferència estadística pel que fa a les proves univariants clàssiques (t d'Student, F de Fisher).

Competències de la titulació a les quals contribueix l'assignatura

Específiques:

1. CE-3. Capacitat per a formular, analitzar i validar models aplicables a problemes d'índole pràctica. Capacitat de seleccionar el mètode i/o tècnica estadística o d'investigació operativa més adequada per aplicar aquest model a cada situació o problema concret.
2. CE-6. Capacitat per a fer servir el software més adequat per a realitzar els càlculs necessaris a la resolució d'un problema.
3. CE-9. Capacitat per a implementar algoritmes d'estadística i investigació operativa.
5. CE-2. Capacitat per a dominar la terminologia pròpia d'algun àmbit en el que sigui necessària l'aplicació de models i mètodes estadístics o d'investigació operativa per a resoldre problemes reals.
6. CE-4. Capacitat de fer servir els diferents procediments d'inferència per a respondre preguntes, identificant les propietats dels diferents mètodes d'estimació i els seus avantatges i inconvenients, adaptats a una situació concreta i en un context específic.

Transversals:

4. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.
7. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

200606 - AMD - Anàlisi Multivariant de Dades

Metodologies docents

Idioma: la primera part (50%) es farà en Anglès, i la segona part (50%) es farà en Castellà.

Teoria: es fan classes magistrals seguint el temari d'acord amb la temporalització entregada a començament del curs.

Problemes: s'utilitzen per fixar els conceptes teòrics dins de la classe de teoria. Al llarg del curs es demanarà el lliurament de problemes per part dels estudiants.

Pràctiques: Es tracta d'utilitzar les facilitats de la programació matricial per fer un anàlisi multivariant utilitzant jocs de dades multivariants. Les pràctiques s'avaluen. El programari utilitzat és R. Les pràctiques es fan individualment.

Treball: Els estudiants han de fer l'anàlisi d'una base de dades amb els mètodes dels cursos i fer una exposició oral dels resultats per tot el grup. Cal redactar i lliurar un informe de l'anàlisi realitzat. El treball es fa en grups de 3 a 4 estudiants.

Objectius d'aprenentatge de l'assignatura

L'estudiant que supera l'assignatura ha de ser capaç de:

1. Reconèixer la naturalesa multivariant d'una base de dades.
2. Enunciar el guany d'un enfocament multivariant, respecte al tradicional univariable.
3. Enunciar els objectius dels mètodes multivariants més utilitzats (ACP, AC, Anàlisi Factorial, Escalament multidimensional, MANOVA, AD, etc.)
4. Identificar el mètode multivariant més adient per a un conjunt de dades concret.
5. Implementar els mètodes bàsics de l'anàlisi multivariant en llenguatge matricial amb el programari R.
6. Aplicar l'estadística descriptiva multivariant a un conjunt de variables.
7. Aplicar els principals mètodes de reducció de la dimensionalitat.
8. Aplicar les transformacions necessàries per un determinat anàlisi (escollir la mètrica)
9. Realitzar la visualització multivariant amb programari estadístic.
10. Interpretar les representacions visuals (biplots) de les dades multivariants.
11. Enunciar la distribució normal multivariant i les seves propietats.
12. Enunciar la definició de les proves estadístiques multivariants bàsiques.
13. Aplicar les proves d'hipòtesis multivariant més freqüents, sobre el vector de mitjanes i sobre la matriu de covariàncies.
14. Aplicar l'anàlisi discriminant lineal i quadràtic a dades corresponents a diferents poblacions d'individus, obtenint les funcions discriminants sota la hipòtesi de normalitat multivariable i realitzar l'assignació d'individus anònims.
15. Enunciar els mètodes bàsics de creació de grups.
16. Aplicar els algorismes per a la creació de grups.
17. Interpretar els resultats dels mètodes multivariants més utilitzats.
18. Aplicar l'anàlisi factorial i extreure els factors comuns a unes variables.
19. Aplicar l'anàlisi de mesures repetides, de perfils i la MANOVA de dos factors.

Hores totals de dedicació de l'estudiantat

Dedicació total: 125h	Hores grup gran:	30h	24.00%
	Hores grup mitjà:	0h	0.00%
	Hores grup petit:	15h	12.00%
	Hores activitats dirigides:	0h	0.00%
	Hores aprenentatge autònom:	80h	64.00%

200606 - AMD - Anàlisi Multivariant de Dades

Continguts

<p>Estadística Descriptiva Multivariant</p>	<p>Dedicació: 61h</p> <p>Grup gran/Teoria: 15h Grup mitjà/Pràctiques: 6h Aprentatge autònom: 40h</p>
<p>Descripció:</p> <ol style="list-style-type: none"> 1. Introducció i conceptes bàsics. Repas de Àlgebra lineal. Geometria de la mostra. Núvols de punts en R^p i R^n. Concepte de mètrica. Mesures de variabilitat. Projecció M-ortogonal. Descomposició en valors y vectors propis. Descomposició en valors singulars generalitzada. Representacions gràfiques: el biplot. 2. Anàlisi de components principals (ACP). Definició dels components. Propietats. ACP basat en la matriu de covariàncies i en la matriu de correlacions. Biplots. Bondat de la representació. 3. Escalament multidimensional. Distàncies i mètriques. Representació euclidiana d'una matriu de distàncies. Descomposició espectral associada. Bondat de la representació. 4. Anàlisi de correspondències simple. Taules de contingència. Perfils fila i perfils columna. Inercia i estadístic chi-quadrat. Biplots. 5. Anàlisi de correspondències múltiple (ACM). ACM basat en la matriu de Burt. ACM basat en la matriu de variables indicadores. Inercies ajustades. Representacions gràfiques. 6. Anàlisi factorial. El model factorial. Factors comuns i específics. Mètodes d'estimació: anàlisi factorial principal i màxima versemblança. Representacions gràfiques. 7. Anàlisi de correlacions canòniques. Funció objectiva. Correlacions canòniques, variables canòniques i pesos canònics. Relació amb altres mètodes. Biplots. <p>Activitats vinculades: Varies pràctiques, exercicis i el treball.</p> <p>Objectius específics: Realitzar l'anàlisi descriptiu gràfic i numèric d'una taula de dades multivariant, tant per taules de dades quantitatives com per taules de dades categoriques.</p>	
<p>Inferència Estadística Multivariant</p>	<p>Dedicació: 29h</p> <p>Grup gran/Teoria: 9h Aprentatge autònom: 20h</p>
<p>Descripció:</p> <p>La distribució normal multivariant. Estadístics mostrals. Prova de la raó de versemblança. Proves sobre la matriu de covariàncies. Prova de la unió de la intersecció. T2 de Hotelling. Proves sobre el vector de mitjanes. Anàlisi de mesures repetides. Anàlisi de perfils. Comparació de diverses mitjanes. La lambda de Wilks. El model MANOVA amb un i mes factors.</p> <p>Activitats vinculades: Pràctiques i problemes.</p> <p>Objectius específics: Realitzar inferència estadística de naturalesa multivariant.</p>	

200606 - AMD - Anàlisi Multivariant de Dades

<p>Classificació i obtenció de grups</p>	<p>Dedicació: 32h Grup gran/Teoria: 7h 30m Grup mitjà/Pràctiques: 4h 30m Aprentatge autònom: 20h</p>
<p>Descripció:</p> <ol style="list-style-type: none"> 1. Anàlisi discriminant. Anàlisi discriminant paramètrica. Funcions discriminants. Anàlisi discriminant lineal i anàlisi discriminant quadràtica. 2. Anàlisi de conglomerats. Distàncies i similitud. Algorismes. Mètodes jeràrquics i Mètodes de partició. Dendrograma. Propietat ultramètrica. Criteri de Ward. <p>Activitats vinculades: Pràctiques i problemes.</p> <p>Objectius específics: Aplicar l'anàlisi discriminant i l'anàlisi clúster i interpretar els seus resultats.</p>	

Sistema de qualificació

L'avaluació es farà mitjançant una ponderació de diferents elements. Hi haurà dos examens, un examen parcial sobre la primera meitat del curs i un examen final sobre la segona meitat. Les pràctiques s'avaluen i la seva mitjana és la nota de pràctiques. Els problemes també s'avaluen i la seva mitjana és la nota de problemes. Al llarg del curs els estudiants han de fer un treball en el qual s'analitzen dades multivariantes amb les tècniques del curs. Del treball cal fer un informe escrit. La nota s'obté a partir de la qualificació dels exàmens, pràctiques, problemes i el treball. La ponderació de les diferents parts de l'avaluació és el següent: examen parcial primera part (35%), examen final segona part (35% si només la segona part, 70% si inclou també la primera part), pràctiques laboratori i problemes (15%), treball (15%). Els alumnes que hagin aprovat el primer examen no cal que es presentin per la matèria de la primera part a l'examen final.

200606 - AMD - Anàlisi Multivariant de Dades

Bibliografia

Bàsica:

- Aluja, T.; Morineau, A. Aprender de los datos: el análisis de componentes principales. EUB, 1999.
- Johnson, R. A.; Wichern, D.W. Applied multivariate statistical analysis. 6th ed. Prentice Hall, 2007.
- Krzanowski, W. J. Principles of multivariate analysis: a user's perspective. Rev. ed. Oxford University Press, 2000.
- Lebart, L.; Morineau, A.; Piron, M. Statistique exploratoire multidimensionnelle. 2e éd. Dunod, 1997.
- Peña Sánchez de Rivera, D. Análisis de datos multivariantes [en línia]. McGraw-Hill, 2002 Disponible a: http://www.ingebook.com/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203.

Complementària:

- Cuadras, C. M. Métodos de análisis multivariante. 2ª ed. PPU, 1991.
- Dillon, W. R.; Goldstein, M. Multivariate analysis methods and applications. John Wiley and Sons, 1984.
- Mardia, K. V.; Kent, J.T.; Bibby, J.M. Multivariate analysis. Academic Press, 1979.
- Morrison, D. F. Multivariate statistical methods. 3rd ed. McGraw-Hill, 1990.
- Volle, Michel. Analyse des données. 3e éd. Economica, 1985.
- Everitt, Brian. An R and S-PLUS companion to multivariate analysis [en línia]. London: Springer, 2005 Disponible a: <http://dx.doi.org/10.1007/b138954>. ISBN 1852338822.

Altres recursos:

- Material informàtic
 - Lecture slides
 - Transparències