

## 240615 - Introducció a Data Science

Unitat responsable: 240 - ETSEIB - Escola Tècnica Superior d'Enginyeria Industrial de Barcelona  
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa  
Curs: 2019  
Titulació: GRAU EN ENGINYERIA EN TECNOLOGIES INDUSTRIALS (Pla 2010). (Unitat docent Optativa)  
GRAU EN ENGINYERIA DE MATERIALS (Pla 2010). (Unitat docent Optativa)  
GRAU EN ENGINYERIA QUÍMICA (Pla 2010). (Unitat docent Optativa)  
Crèdits ECTS: 4,5 Idiomes docència: Anglès

### Professorat

Responsable: JOSEP GINEBRA  
Altres: JOSEP GINEBRA

### Horari d'atenció

Horari: Dilluns i dimecres de 11:30 a 13:30

### Capacitats prèvies

Haver fet l'assignatura d'Estadística

### Competències de la titulació a les quals contribueix l'assignatura

Específiques:

1. Coneixements bàsics sobre l'ús i programació d'ordinadors, sistemes operatius, bases de dades i programes informàtics amb aplicacions en enginyeria.
2. Coneixements i capacitats per organitzar i gestionar projectes. Conèixer l'estructura organitzativa i les funcions d'una oficina de projectes.

Transversals:

3. COMUNICACIÓ EFICAÇ ORAL I ESCRITA: Comunicar-se de forma oral i escrita amb altres persones sobre els resultats de l'aprenentatge, de l'elaboració del pensament i de la presa de decisions; participar en debats sobre temes de la pròpia especialitat.
4. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com un membre més, o realitzant 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.

### Metodologies docents

Totes les classes son pràctiques, a davant de l'ordinador. L'anàlisi de les dades es farà amb MINITAB i amb R. Cada setmana es proposaran petits exercicis d'anàlisi de dades per fer a casa. Hi haurà un treball final d'assignatura.

### Objectius d'aprenentatge de l'assignatura

A l'acabar l'assignatura l'estudiant ha de ser capaç d'identificar situacions en les que es profitos anàlitzar dades, d'identificar el model o mètode d'anàlisi més adient per a les seves dades, de construir un model que sintetzi la informació a les dades i permeti fer prediccions, de reduir la dimensionalitat i visualitzar dades multivariants, de implementar algoritmes de classificació supervisada i no supervisada, i d'avaluar la qualitat dels resultats obtinguts.



## 240615 - Introducció a Data Science

### Hores totals de dedicació de l'estudiantat

Dedicació total: 112h 30m	Hores grup gran:	0h	0.00%
	Hores grup mitjà:	45h	40.00%
	Hores grup petit:	0h	0.00%
	Hores activitats dirigides:	0h	0.00%
	Hores aprenentatge autònom:	67h 30m	60.00%

## 240615 - Introducció a Data Science

### Continguts

<p>Tema 1: Introducció</p>	<p>Dedicació: 3h 30m</p> <p>Grup gran/Teoria: 1h 30m Activitats dirigides: 1h Aprentatge autònom: 1h</p>
<p>Descripció:</p> <p>1.- Tipus de problemes: Associació, predicció i classificació. 2.- Tipus d'eines: Models estadístics i anàlisi multivariant.</p>	
<p>Tema 2: Models lineals per a resposta continua</p>	<p>Dedicació: 30h</p> <p>Grup gran/Teoria: 6h Grup petit/Laboratori: 6h Activitats dirigides: 6h Aprentatge autònom: 12h</p>
<p>Descripció:</p> <p>1.- Model lineal normal teòric i ajustat. 2.- Ajust del model; mínims quadrats i regressió robusta. 3.- Taula ANOVA i mesura de bondat de l'ajust. 4.- Inferència sobre els paràmetres del model. 5.- Predicció. 6.- Validació del model. 7.- Selecció del model. 8.- Validació creuada i proves de falta d'ajust. 9.- Interpretació del model; Biaix, colinearitat i causalitat. 10.- Us de variables explicatives categòriques. 11.- Comparació de mitjanes. 12.- Anàlisi de dissenys factorials a dos nivells.</p>	
<p>Tema 3: Models no-lineals per a resposta continua</p>	<p>Dedicació: 6h</p> <p>Grup gran/Teoria: 1h 30m Grup petit/Laboratori: 1h 30m Aprentatge autònom: 3h</p>
<p>Descripció:</p> <p>1.- Model no-lineal normal teòric i ajustat. 2.- Ajust del model. 3.- Inferència. 4.- Validació del model.</p>	
<p>Tema 4: Models per a resposta categòrica i discreta</p>	<p>Dedicació: 22h 30m</p> <p>Grup gran/Teoria: 4h 30m Grup petit/Laboratori: 4h 30m Activitats dirigides: 4h 30m Aprentatge autònom: 9h</p>
<p>Descripció:</p> <p>1.- Model lineal generalitzat. 2.- Model per a comptetjos. 3.- Model per a resposta binària. 4.- Ajust. 5.- Inferència. 6.- Validació. 7.- Predicció. 8.- Interpretació del model. 9.- Taules de contingència i models per a resposta politòmica.</p>	

## 240615 - Introducció a Data Science

<p>Tema 5: Models de sèries temporals</p>	<p>Dedicació: 13h</p> <p>Grup gran/Teoria: 3h Grup petit/Laboratori: 3h Activitats dirigides: 3h Aprentatge autònom: 4h</p>
<p>Descripció: 1.- Descripció d'una sèrie temporal. 2.- Models AR. 3.- Models MA. 4.- Models ARIMA. 5.- Models ARIMA estacionals.</p>	
<p>Tema 6: Visualització de dades multivariants (Reducció de la dimensionalitat)</p>	<p>Dedicació: 6h</p> <p>Grup gran/Teoria: 1h 30m Grup petit/Laboratori: 1h 30m Aprentatge autònom: 3h</p>
<p>Descripció: 1.- Anàlisi de components principals. 2.- Anàlisi de correspondències.</p>	
<p>Tema 7: Anàlisi clúster (Classificació no supervisada)</p>	<p>Dedicació: 6h 30m</p> <p>Grup gran/Teoria: 1h 30m Grup petit/Laboratori: 1h 30m Activitats dirigides: 1h 30m Aprentatge autònom: 2h</p>
<p>Descripció: 1.- Mètodes jeràrquics. 2.- Mètodes de partició (Algoritmes k-means). 3.- Anàlisi cluster per variables.</p>	
<p>Tema 8: Anàlisi discriminant (Classificació supervisada)</p>	<p>Dedicació: 8h 30m</p> <p>Grup gran/Teoria: 1h 30m Grup petit/Laboratori: 1h 30m Activitats dirigides: 1h 30m Aprentatge autònom: 4h</p>
<p>Descripció: 1.- Discriminació lineal. 2.- Discriminació quadràtica. 3.- Discriminació logística.</p>	

## 240615 - Introducció a Data Science

Tema 9: Model de regressió i classificació no paramètrica

Dedicació: 4h 30m

Grup gran/Teoria: 1h 30m

Grup petit/Laboratori: 1h 30m

Aprentatge autònom: 1h 30m

Descripció:

1.- Suavitzats locals. 2.- Veïns mes propers. 3.- Models additius. 4.- Classificació i regressió en arbre. 5.- Xarxes neuronals.

### Sistema de qualificació

Hi haurà un examen parcial fet a casa i un examen final fet a classe.

Nota Assignat. = 0,1 Exercicis + 0,3 Projecte + 0,1 Examen Parcial + 0,5 Examen Final

## 240615 - Introducció a Data Science

### Bibliografia

#### Bàsica:

Hastie, Trevor; Tibshirani, Robert; Friedman, J. The elements of statistical learning: data mining, inference and prediction [en línia]. 2nd ed. New York: Springer Verlag, 2009 [Consulta: 14/06/2019]. Disponible a: <<http://dx.doi.org/10.1007/978-0-387-84858-7>>. ISBN 9780387848570.

Peña, Daniel. Regresión y diseño de experimentos. Madrid: Alianza, 2002. ISBN 9788420693897.

Venables, William N; Ripley, B.D. Modern Applied Statistics with S. 4th ed. New York: Springer Verlag, 2003. ISBN 0387954570.

Peña, Daniel. Análisis de datos multivariantes [en línia]. Madrid: McGrawHill, 2008 [Consulta: 28/01/2019]. Disponible a: <[http://www.ingebok.com/ib/NPcd/IB\\_BooksVis?cod\\_primaria=1000187&codigo\\_libro=4203](http://www.ingebok.com/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203)>. ISBN 9788448136109.

Weisberg, Sanford. Applied Linear Regression. 3rd ed. New York: Wiley, 2005. ISBN 0471663794.

Everitt, B.S.; Dunn, G. Applied Multivariate Data Analysis [en línia]. 2nd. New York: Wiley, 2010 [Consulta: 27/06/2014]. Disponible a: <<http://onlinelibrary.wiley.com/book/10.1002/9781118887486>>. ISBN 9781118887486.

Greenacre, Michael J. Correspondence Analysis in Practice. 2nd Ed. Boca Raton: Chapman and Hall, 2007. ISBN 9781584886167.

James, Gareth [et al.]. An Introduction to statistical learning : with applications in R. New York: Springer Verlag, 2013. ISBN 9781461471370.

Gareth, James [et al.]. An Introduction to statistical learning : with applications in R. New York: Springer Verlag, 2013. ISBN 9781461471370.

#### Complementària:

Clarke, Bertrand; Fokoue, Ernest; Zhang, Hao Helen. Principles and Theory for Data Mining and Machine Learning. Berlin: Springer Verlag, 2009. ISBN 9780387981345.

Dobson, Annette J. An Introduction to Generalized Linear Models. 3rd ed. Boca Raton: Chapman Hall, 2008. ISBN 9781584889502.

Johnson, Richard; Wichern, Dean. Applied multivariate statistical analysis. 6th ed. Englewood Cliffs, N.J.: Pearson, 2007. ISBN 9780131877153.

Peña, Daniel. Análisis de series temporales. Madrid: Alianza, 2005. ISBN 8420691283.

Wakefield, Jon. Bayesian and frequentist regression methods. New York: Springer Verlag, 2013. ISBN 9781441909244.