

## 240ST014 - Anàlisi de Dades de Transport i Logística

Unitat responsable: 240 - ETSEIB - Escola Tècnica Superior d'Enginyeria Industrial de Barcelona  
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa  
Curs: 2018  
Titulació: MÀSTER UNIVERSITARI EN CADENA DE SUBMINISTRAMENT, TRANSPORT I MOBILITAT (Pla 2014). (Unitat docent Obligatòria)  
MÀSTER UNIVERSITARI EN ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (Pla 2013). (Unitat docent Optativa)  
MÀSTER UNIVERSITARI EN ENGINYERIA INDUSTRIAL (Pla 2014). (Unitat docent Optativa)  
Crèdits ECTS: 5 Idiomes docència: Castellà, Anglès

### Professorat

Responsable: Montero Mercadé, Lúdia

### Requisits

Els estudiants han de tenir coneixements previs d'anàlisi matemàtica i àlgebra per tal d'assimilar els conceptes de probabilitat, distribucions de variables aleatòries univariants, sèries numèriques, notació matricial, funcions reals d'una o més dimensions, derivació i integració.

Els estudiants han de saber programar en pseudocodi o algun llenguatge d'alt nivell.

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

Específiques:

CESCTM2. Dissenyar procediments de recollida de dades de transport que tinguin en compte la seva especificitat, saber aplicar les tècniques adequades per tractar-les, analitzar-les i extreure conclusions pel seu ús adequat als models que les requereixin.

CESCTM3. Dissenyar i realitzar estudis d'anàlisi de la demanda, modelar la demanda i la seva estructuració segons els diferents models de transport.

CETM2. Comprensió i capacitat de quantificació de les variables fonamentals de sistemes de transport i de mobilitat que determinen la seguretat, la qualitat i la sostenibilitat de les infraestructures de transport i optimització del funcionament d'aquests sistemes.

## 240ST014 - Anàlisi de Dades de Transport i Logística

### Metodologies docents

L'aprenentatge del curs es compon de tres fases diferents:

1. Adquisició de coneixements específics a través de l'estudi de la literatura i material proporcionat pels professors.
2. L'adquisició d'habilitats en tècniques específiques d'anàlisi de dades, explotació de la informació i modelatge estadístic.
3. Integració dels coneixements, habilitats i competències (genèriques i específiques) per la resolució de casos pràctics curts.

Les classes de teoria s'exposen els fonaments de les metodologies i tècniques de l'assignatura.

Les classes de laboratori serveixen per aprendre l'ús de tècniques específiques per a la resolució de problemes en l'anàlisi i modelatge de dades estadístiques mitjançant l'ús d'eines informàtiques adequades, en aquest sentit, els estudiants primer han de seguir i prendre notes sobre l'anàlisi realitzada pel professor i després resoldre en hores d'autoaprenentatge un cas d'estudi curt similar, que es centra en els continguts del bloc actual tot seguint les pautes descrites en un qüestionari inclòs en la descripció de les sessions de laboratori. El cas d'estudi curt contingut ha de ser resolt d'acord amb el qüestionari en no més d'una setmana o quan indica el professor durant la sessió de laboratori. L'avaluació formativa es durà a terme facilitant la correcció abans de la propera sessió de laboratori, on s'inclourà una discussió sobre els problemes comuns trobats pel professor en els primers 20 minuts.

Els casos d'estudi curts són desenvolupats pels estudiants en grups durant hores d'autoaprenentatge, serveixen per posar en pràctica els coneixements, habilitats i competències en la resolució dels casos tots ells relacionats amb Logística, Transport i Mobilitat. Programari R és l'eina seleccionada per a l'anàlisi de dades i el modelatge. Programaris professionals comuns en el sector (TransCAD, EMME4, VISSUM) es relacionen amb les eines R presentades segons la disponibilitat de llicències docents a la UPC.

### Objectius d'aprenentatge de l'assignatura

Aprendre a fer un informe sobre la qualitat de les dades (perfil de les dades que falten, detecció de valors singulars i bivariada). Recuperació de dades faltants.

Aprendre a utilitzar i interpretar els conceptes fonamentals de la probabilitat i l'estadística des d'un punt de vista pràctic quan s'utilitza el programari estadístic R: esdeveniment aleatori, població, mostra, variable aleatòria, variables aleatòries contínues i discretes comuns. Punt i estimació de l'interval. Inferència estadística computacional.

Aprendre a analitzar les bases de dades, incloent la descripció univariant numèrica i gràfica, amb eines per descripció bivariant i multivariant. Determinació de les característiques significatives dels grups d'individus.

Aprendre a fer un perfil per a una variable de resposta, ja sigui quantitativa o qualitativa. Selecció de característiques rellevants.

Aprendre els principis bàsics de la Classificació: tècniques de classificació jeràrquica i veïns K més proper. Realitzar i validar una proposta de classificació utilitzant el programari R.

Saber fer el modelatge de les respostes numèriques: formulació, estimació i interpretació de models estadístics utilitzant el programari R.

Conèixer els indicadors per la comparació de models generals de regressió i selecció del millor: estadístics per la bondat de l'ajust ( $R^2$ , F-Test per als models niats, AIC, BIC, etc)

Saber fer la diagnosi dels models lineals generals amb finalitats predictives: valors atípics i dades influents.

Saber fer el modelatge per opcions-discrets amb els models lineals generalitzats : formulació, estimació i interpretació de models estadístics utilitzant el programari R.

Conèixer els indicadors per la comparació de models per resposta binària i selecció del millor: estadístics per la bondat de l'ajust (Pearson  $\chi^2$ , Test desviació per als models niats, AIC, BIC, etc)

Saber fer la diagnosi dels models lineals generalitzats per la resposta binària amb finalitats predictives.

Saber aplicar els models lineals generalitzats de resposta binària a l'elecció modal entre parells de zones de transport (ZAT). Models per dades agregades vs dades desagregades.

Conèixer els principis bàsics de la teoria del mostreig: estimacions per punt i estimacions per interval. Aprendre a calcular errors relatius i errors absoluts per a estimadors de mitjanes, totals i proporcions en el mostreig aleatori i el mostreig estratificat.

## 240ST014 - Anàlisi de Dades de Transport i Logística

### Hores totals de dedicació de l'estudiantat

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

## 240ST014 - Anàlisi de Dades de Transport i Logística

### Continguts

<p><b>Bloc 1. Introducció a l'Anàlisi de Dades en Transport i Logística</b></p>	<p>Dedicació: 6h</p> <p>Grup gran/Teoria: 2h Grup mitjà/Pràctiques: 1h Aprentatge autònom: 3h</p>
<p><b>Descripció:</b> Introducció a les col·leccions de dades comunes i enquestes en Logística, Transport i Mobilitat: enquestes domiciliàries, enquestes origen-destinació, enquestes d'acordonament, les enquestes de preferències revelades i declarades. La recopilació de dades de trànsit: es sensors de bucle inductiu i noves tecnologies (dades Bluetooth, dades de sensors magnètics sense fils, etc.)</p> <p><b>Activitats vinculades:</b> Sessió Presencial més Introducció al R al Laboratori</p> <p><b>Objectius específics:</b> L'objectiu de l'assignatura és proporcionar als estudiants els coneixements i habilitats per fer front a anàlisi exploratòria de dades i les necessitats d'extracció de dades de les organitzacions i la pràctica professional en el camp del Transport i la Logística. És a dir, per prendre avantatge de les dades emmagatzemades pels organismes relacionats i integrar-les en els sistemes automàtics per ajudar a la presa de decisions i les operacions de tràfic i gestió. La idea subjacent és que les dades són un tresor per als entes implicats i a través de la seva exploració es converteixen en informació. El curs es desenvolupa sobre la base de l'estudi de casos. Es divideix en quatre àrees: Anàlisi Exploratori de Dades i Interpretació-Descripció de Resum, Inferència Estadística Computacional, Modelatge i Predicció-Eines. Disseny de qüestionaris i Presa de Mostres de Disseny.</p> <p>L'assignatura dona una sòlida formació en les tècniques per administrar, analitzar, modelar i extreure coneixement a partir dels conjunts de dades massius actuals, bases de dades, Internet, ..., així com en les tècniques per explotar aquest coneixement en el sector.</p>	
<p><b>Bloc 2. Anàlisi Exploratori de Dades</b></p>	<p>Dedicació: 17h 30m</p> <p>Grup gran/Teoria: 3h Grup mitjà/Pràctiques: 1h 30m Aprentatge autònom: 13h</p>
<p><b>Descripció:</b> Anàlisi exploratòria de dades: eines numèriques i gràfiques univariants/bivariants (característiques quantitatives i qualitatives). Les dades faltants: el perfil i la seva recuperació. La detecció de valors atípics univariants i bivariants. Mesures d'associació de dades multivariants (coeficients de correlació de Pearson / Spearman). Exemple de dades massius: . Comptatges de trànsit (recuperació de dades faltants, detecció de valors atípics, filtrat)</p> <p><b>Objectius específics:</b> Aprendre a fer un informe sobre la qualitat de les dades (perfil de les dades que falten, detecció de les altres univariants i bivariada). Falta de recuperació. Aprendre a analitzar les bases de dades, incloent la descripció univariant numèrica i gràfica, bivariada i multivariada amb les eines del R. Determinació de les característiques significatives dels grups d'individus.</p>	

## 240ST014 - Anàlisi de Dades de Transport i Logística

<p><b>Bloc 3. Inferència Estadística Computacional</b></p>	<p>Dedicació: 24h</p> <p>Grup gran/Teoria: 2h Grup mitjà/Pràctiques: 6h Aprentatge autònom: 16h</p>
<p><b>Descripció:</b> Elements estadístics bàsics utilitzats en el transport i la logística: distribucions univariants comuns (binomial, multinomial, de Poisson, exponencial, Weibull, gamma, (log) logística (log) normal, etc), amb èmfasi en els moments i paràmetres característics (localització, escala i forma ). Identificació. Inferència estadística computacional per mitjanes, variàncies i proporcions i les diferències segons grups: tests paramètrics i no paramètrics (Chi2, Anderson-Darling, Wilcoxon, Kruskal-Wallis, Barlett, etc).</p> <p><b>Objectius específics:</b> Aprendre a utilitzar i interpretar els conceptes fonamentals de la probabilitat i l'estadística des d'un punt de vista pràctic quan s'utilitza R: esdeveniment aleatori, població, mostra, variable aleatòria, variables aleatòries contínues i discretes comuns. Punt i estimació de l'interval. Inferència estadística computacional. Ajust de la distribució de probabilitat per una recollida univariant de dades.</p>	
<p><b>Bloc 4. Modelatge Estadístic per Regressió</b></p>	<p>Dedicació: 24h 30m</p> <p>Grup gran/Teoria: 5h 30m Grup mitjà/Pràctiques: 5h Aprentatge autònom: 14h</p>
<p><b>Descripció:</b> Modelatge mitjançant models de regressió múltiple. Estimació per mínims quadrats. Propietats. La transformació de les variables. Les eines de diagnòstic: residuals, dades influents i valors atípics. Model lineal general: la forma d'introduir les variables qualitatives com a variables explicatives - definició de variables fictícies. Efectes principals i les interaccions entre els factors i amb les covariables: interpretació i validació dels models. Prova F de Fisher per comparar models encaixats. Exemple: Modelatge del nombre de viatges en punta de matí entre zones de transport (ZAT) amb el mètode lm() de R i la tria del millor model amb el mètode step () del programari R.</p> <p><b>Objectius específics:</b> Saber fer modelatge de les respostes numèriques: formulació, estimació i interpretació de models estadístics utilitzant el programari R. Saber fer la comparació de models i selecció del millor: estadístics de bondat de l'ajust (R2, F-Test per als models niats, AIC, BIC, etc) Saber fer la diagnosi i validació dels models lineals generals: valors atípics i dades influents. Saber fer la predicció de característiques numèriques explicades mitjançant un model lineal general. Saber aplicar als models de generació / atracció per a les zones de transport (ZAT): formulació, estimació, validació, predicció i interpretació.</p>	

## 240ST014 - Anàlisi de Dades de Transport i Logística

<p><b>Bloc 5. Models per Resposta Binària</b></p>	<p>Dedicació: 18h</p> <p>Grup gran/Teoria: 4h Grup mitjà/Pràctiques: 2h Aprentatge autònom: 12h</p>
<p><b>Descripció:</b> Modelatge de dades discretes binàries a través de models de regressió generalitzada: funció d'enllaç, l'estimació ML, les propietats, diagnosi i validació de models i la seva interpretació. Prediccions de respostes binàries: capacitat predictiva. Area ROC. Test de deviancia per comparar models encaixats. Estudi de cas de la selecció de mode de transport entre els modes públics i privats segons característiques individuals amb glm() en R.</p> <p><b>Objectius específics:</b> Conèixer el modelatge pels models lineals generalitzats discrets per resposta binària: formulació, estimació, validació i interpretació de models estadístics. Saber calcular, interpretar i fer prediccions amb models per resposta discreta binària utilitzant el programari R. Saber fer diagnosi i validació pels models de resposta binària en R. Saber comparar models per selecció binària i fer la tria del millor. Saber calcular i interpretar els estadístics de Bondat d'ajust i els tests associats (Pearson X2, Test desviància, AIC, BIC, etc) Saber fer Aplicació a l'elecció modal entre parells de zones de transport (ZAT). Entendre les diferències i semblances entre Models Agregats vs Desagregats.</p>	
<p><b>Bloc 6. Introducció al Mostreig Estadístic</b></p>	<p>Dedicació: 12h</p> <p>Grup gran/Teoria: 4h Aprentatge autònom: 8h</p>
<p><b>Descripció:</b> Introducció a la teoria de mostreig: mostreig aleatori i el mostreig estratificat. Punt i estimacions d'interval per als mitjans, totals i proporcions en un mostreig aleatori. Selecció de la mida de la mostra per satisfer errors absoluts / relatius a un mostreig aleatori i estratificat. Exemple: Dimensionament d'una enquesta de mobilitat domiciliària.</p> <p><b>Objectius específics:</b> Conèixer els principis bàsics de la teoria del mostreig: Punt i estimacions d'interval. Apreneu com calcular relativa vs estimacions d'error absoluts per a mitjans, totals i proporcions en el mostreig aleatori i el mostreig estratificat. Aprendre com calcular errors relatius i absoluts per a estimacions de mitjanes, totals i proporcions en el mostreig aleatori i el mostreig estratificat.</p>	

## 240ST014 - Anàlisi de Dades de Transport i Logística

<p><b>Bloc 7. Introducció a la Minería de Dades</b></p>	<p>Dedicació: 12h</p> <p>Grup gran/Teoria: 2h Grup mitjà/Pràctiques: 2h Aprentatge autònom: 8h</p>
<p><b>Descripció:</b> Mineria de dades de dades massius: mètodes útils per a la Logística i el Transport. Classificació: segmentació de la població d'una àrea d'estudi - Classificació jeràrquica amb components R. Principi com una eina per a la reducció de la dimensionalitat. Exemple: Enquesta de satisfacció dels usuaris del transport d'una xarxa d'autobusos</p> <p><b>Objectius específics:</b> Dissenyar com convertir les dades en informació que és d'utilitat per a la presa de decisions. Saber calcular Perfils en R . Conèixer estratègies per la reducció de la dimensionalitat. Saber fer i interpretar la Classificació Jeràrquica no Dirigida usant R. Saber fer i interpretar la Classificació Dirigida (K-Means) usant R.</p>	
<p><b>Avaluació: Parcial i Final</b></p>	<p>Dedicació: 11h</p> <p>Grup gran/Teoria: 5h Aprentatge autònom: 6h</p>
<p><b>Descripció:</b> Quiz i Examen Final</p> <p><b>Activitats vinculades:</b> El Quiz o Examen Parcial té lloc durant la setmana de parcials, a meitat de quadrimestre, que programa l'ETSEIB. La data depèn de la programació de l'ETSEIB no la trien ni professors ni estudiants.</p>	

### Sistema de qualificació

L'avaluació del curs integra les tres fases del procés d'aprenentatge: coneixements, habilitats i competències.

El coneixement és avaluat per un qüestionari o prova parcial i l'examen final (puntuacions F1 i F2), al mig i darrera setmana del curs.

Les habilitats i competències s'avaluen a partir del lliurament de les m pràctiques ( $m > 1$ ) amb base en els casos d'estudi curts i relacionats amb els continguts del curs. Cadascun dels blocs, excepte el primer, podria implicar una pràctica que els alumnes realitzaran en grup (com a màxim 3 persones). La mitjana de les m puntuacions facilita la puntuació L. Els estudiants han de quantificar les hores dedicades a resoldre cada pràctica i lliurar-la segons indica la corresponent Tasca a ATENEA. Es donarà RetroAlimentació per a l'avaluació formativa per part del professor en no més de 10 dies, abans de la següent sessió de laboratori on es discutiran els problemes i errors comuns en els primers 20 minuts.

La nota final s'obté donant pesos a les anteriors puntuacions:  $\text{Nota Final} = 0.65F + 0.35L$ . On F es Max (F2,  $0.3F1 + 0.7F2$ ).

## 240ST014 - Anàlisi de Dades de Transport i Logística

### Normes de realització de les activitats

Els estudiants podran dur calculadora, taules estadístiques, etc. No es permet dur les resolucions d'exàmens d'anys anteriors, però sí es poden dur les resolucions dels casos d'estudis disponibles al web docent de l'assignatura.

### Bibliografia

#### Bàsica:

Washington, S.P. ; Karlaftis, M.G. ; Mannering, F.L. Statistical and Econometric methods for transportation data analysis. 2nd. Boca Raton: Chapman and Hall, 2011. ISBN 9781420082852.

Dalgaard, Peter. Introductory Statistics with R [en línia]. 2nd ed. New York: Springer, 2008 [Consulta: 05/10/2017]. Disponible a: <<http://dx.doi.org/10.1007/978-0-387-79054-1>>. ISBN 9780387790534.

Clairin, Rémy ; Brion, Philippe. Manual de Muestreo. Madrid: La Muralla, 2001. ISBN 8471337118.

Fox, John. Applied Regression Analysis and Generalized Linear Models. 2nd ed. Los Angeles: SAGE, 2008. ISBN 9780761930426.

Fox, John ; Weisber, Sanford. An R Companion to Applied Regression. 2nd ed. Thousands Oaks: SAGE, 2002. ISBN 9781412975148.

Ortúzar S., Juan de Dios; Willumsen, Luis G. Modelling transport. 4th ed. Chichester: John Wiley & Sons, 2011. ISBN 9780470760390.

#### Altres recursos:

Pàgina web del curs:

- Planificació de l'assignatura
- Notes relacionades amb els blocs de continguts i diapositives usades en les classes teòriques.
- Descripció de les sessions pràctiques, els qüestionaris per a cada bloc i casos d'estudi.
- Casos d'Estudi: les dades (format MS-Excel i R) i descripció del context i la variable objectiu / s.
- Directrius per als casos d'estudi que es presenten en forma d'una llista de preguntes de guia per a l'anàlisi.
- Exàmens Finals i qüestionaris dels cursos anteriors.

#### Enllaç web

Web Docent ADTL

<http://www-eio.upc.es/teaching/adtl/>

#### Material informàtic

ATENEA - Tasques

ATENEA disposarà de Tasques pel lliurament dels casos d'estudi pràctics