

Pràctica resposta recomptes – Enunciat

Objectius breus

- 1) Saber quins són els factors explicatius del nombre d'execucions de problemes d'*e-status* per un alumne de la FIB a l'assignatura de Probabilitat i Estadística (PE).
- 2) Establir la relació que hi ha entre els factors:
 - a. *nationality* (Spanish/Foreign): nacionalitat de l'estudiant
 - b. *group* (Morning/Afternoon): Si l'estudiant està matriculat en un grup de matí o tarda
 - c. *repeater* (No/Yes): Si l'estudiant és repetidor

Pel primer punt, haureu de testejar quin dels models explicats durant l'assignatura (poisson, quasipoisson,...) s'escau més per explicar la variable resposta. Pel segon punt, s'hauria d'emprar un model log-lineal per establir la relació entre les tres variables a través de les freqüències observades a la taula de contingència.

Descripció

Els estudiants de l'assignatura de PE de la Facultat d'Informàtica de Barcelona (FIB) tenen la possibilitat de preparar millor els exàmens parcials i finals mitjançant una plataforma de problemes anomenada ***e-status*** que permet resoldre enunciats de problemes de forma interactiva amb dades diferents a cada execució. Els professors recomanem l'ús d'aquesta plataforma perquè creiem que reforça l'aprenentatge adquirit a classe. En aquesta pràctica, s'analitzarà el **nombre d'execucions de problemes** fets pels alumnes en aquesta plataforma en funció d'una sèrie de factors explicatius (nacionalitat, grup, etc...). S'entén per nombre d'execucions el número d'intents totals de fer algun problema durant el quadrimestre independentment de que es resolgui bé o malament.

Dades

S'entregarà 1 conjunt de dades amb els 228 alumnes matriculats a l'assignatura de PE de la FIB durant un quadrimestre específic. Hi haurà 8 variables al joc de dades que es descriuen a continuació.

Variables

Variables explicatives

id: identificador de l'alumne

repeater: si l'alumne és repetidor (No/Yes)

n_subject: nombre d'assignatures que cursa l'alumne a banda de "Probabilitat i Estadística"

group: Si l'estudiant està matriculat en un grup de matí o tarda (Morning/Afternoon)

nationality: si l'alumne és espanyol o estranger

n_test: nombre de proves presencials a classe que l'alumne va realitzar sobre un total de 12.

note: nota que va treure l'alumne de l'assignatura en una escala de 0 a 10

Variable resposta:

y: nombre de problemes d'estatus que va realitzar l'alumne durant tot el quadrimestre

Tasques

A la pràctica heu de desenvolupar, com a mínim, les següents tasques:

1. Llegir les dades (*data_status.txt*)
2. Fer una descriptiva
 - a. Univariada de les variables explicatives i la resposta
 - b. Bivariada de les variables explicatives amb la resposta.
3. Fer depuració de les dades si s'escau: transformacions, agrupacions de categories, etc...

Objectiu 1

4. Ajustar algun/s dels models vistos a la sessió de teoria per recomptes (poisson, quasi-poisson, binomial negarita, zero-inflats) per explicar la variable resposta número d'execucions de problemes en funció de la resta de variables explicatives (es pot incloure variables addicionals creades a la fase de depuració). A més:
 - a. Es pot estudiar incloure alguna interacció (si es creu convenient).
 - b. S'ha de fer un procés de selecció de variables (si es creu convenient)
 - c. S'ha de validar el model

Objectiu 2

5. Ajustar un model log-lineal (**suposeu el total global fixat**) per veure la relació entre els factors:
 - a. *nationality* (Spanish/Foreign)
 - b. *group* (Morning/Afternoon)
 - c. *repeater* (No/Yes).
6. Alguns consells són:
 - a. Ajusteu els models començant des del més simple (independència total)
 - b. Compareu amb el model final escollit, els efectius esperats vs. els observats
 - c. Interpreteu el resultats.

Avaluació

Es valoraran, entre altres, els següents aspectes:

1. Interpretació dels resultats.
2. Criteris de selecció dels models.
3. Presentació i claredat dels resultats.

Data d'entrega

La data límit per realitzar l'entrega serà el dia **23/12/18 a les 23:55** a través del campus.

Documentació i format de la documentació a entregar

1. **Una descripció del treball realitzat. NOTA: Aquesta descripció ha de ser auto-explicativa en el sentit que el document per si sol s'ha de poder entendre sense recórrer a consultar el script de R.** El document ha d'incloure de forma obligatòria: 1) si es va crear alguna variable derivada o si es va realitzar alguna transformació sobre les variables ja existents; 2) el procés de selecció dels models finals per a cada objectiu; 3) l'expressió del model final per a cada objectiu; 4) La validació dels models finals; i 5) una interpretació global del resultat. Extensió aproximada: 5 a 10 pàgines (es recomana ometre la informació no rellevant i passar-la a l'apèndix). **Format: .docx/pdf/html**

2. **El codi** utilitzat per a realitzar aquesta part amb comentaris explicatius. Format: **.R o .Rmd**

Normes de format

1. Els noms complets dels components del grup han d'estar tant dins del fitxer que conté la descripció del treball com dins del codi d'anàlisi.
2. El format del nom dels dos fitxers com de la carpeta comprimida que els contingui ha de ser cognom1_cognom2.ext. P.ex: garcia_lopez.html, garcia_lopez.Rmd i garcia_lopez.zip
3. Només es farà una entrega per grup en una carpeta comprimida que contingui els dos fitxers.

L'incompliment d'aquestes normes comportarà una penalització a la nota.