

PAC 4 INFÈRENCIA ESTADÍSTICA II: CONTRAST D'HIPÒTESIS

Grau en Ciència de Dades Aplicada

Curs 2020/21

El nom i cognoms de l'estudiant s'han d'indicar clarament a la primera pàgina de l'informe.

L'informe final es lliurarà únicament en format PDF i a l'aula Moodle.

Exercici 1 (4 punts)

En les següents situacions heu d'indicar el tipus d'anàlisi o de proves estadístiques que cal utilitzar en cadascun d'ells.

- a) Es vol provar un nou medicament que es creu que és més eficaç per al control del colesterol. Per aquest motiu es dissenya un assaig clínic y s'assignen 100 pacients al grup del nou tractament i 100 al grup del tractament habitual. Es vol veure si al final de l'estudi ha millorat el percentatge d'individus que tenen ben controlat el nivell de colesterol. Indica el test que caldria fer, l'estadístic de contrast i el valor crític amb una confiança del 95%.
- b) Es passa una prova de coneixements d'estadística a un grup de 100 professors d'institut. Posteriorment, s'efectua una formació específica on-line de caràcter voluntari. Al finalitzar la formació es torna a passar la prova als mateixos alumnes. Indica el test que caldria fer per veure si ha augmentat el percentatge de coneixement, l'estadístic de contrast i el valor crític amb una confiança del 95%.
- c) Una empresa vol contrastar si dos proveïdors de pistons li donarien el mateix resultat. L'equip de compres selecciona una mostra de 50 pistons del proveïdor A i 50 del proveïdor B i els sotmet a un test d'estrès. Vol contrastar si hi ha diferències entre el percentatge de pistons que el superen en cada proveïdor. Indica el test que caldria fer, l'estadístic de contrast i el valor crític amb una confiança del 95%.

Exercici 2 (6 punts)

En aquest exercici farem un anàlisi amb dades reals. S'ha simplificat la base de dades i us la proporcionem en diversos formats: STATA (dta), EXCEL (xls) i Text pla separat per comes (csv).

Les dades

La demora entre el començament dels símptomes i l'ingrés hospitalari és un factor que determina la mortalitat de l'infart agut de miocardi (IAM). S'estudien 426 individus que acudeixen al servei d'urgències de 5 hospitals per dolor toràcic, recollint el temps entre els primers símptomes i l'arribada a l'hospital i una sèrie de variables sociodemogràfiques. S'està interessat en estimar el retard prehospitalari i determinar algunes variables associades.

Les dades les podeu trobar en els fitxers de Stata demora.dta, d'Excel demora.xls i de text pla separat per comes demora.csv

Les variables de la base de dades són:

DEMORA Minuts des de l'inici dels símptomes fins arribar a l'hospital

HOSPITAL

EDAD

SEXO 0 Home 1 Dona

NACIONAL 0 Espanyol 1 Estranger

MEDIO1 Lloc inici símptomes. 0 Domicili 1 Fora del domicili

DISTANC Distància en isòcrones (0,1,2,3, ...)

REMITIDO Remès per un sanitari 0 Si 1 No

NSE Nivell socioeconòmic 0 Baix 1 Mitjà 2 Alt

COHABIT Convivents 0 Viu en companyia 1 Viu sol

CARDIOP Diagnòstic previ de cardiopatia 0 No 1 Si

DOLOR Nivell de dolor 0 intens 1 moderat 2 lleuger

NOCHE Aparició nocturna 0 Dia 1 Nit

AMBULAN Acudir a l'hospital en ambulància 0 Ambulància 1 Altres mitjans

Les preguntes

Comencem llegint la base de dades, per exemple:

```
library(foreign)
```

```
dades_demora<-read.dta("demora.dta")
```

- **a) Hem llegit que de mitjana els individus triguen 5 hores (300 minuts) des de l'inici dels símptomes fins arribar a l'hospital. Feu un test d'hipòtesis amb la nostra mostra per decidir si aquesta informació és certa amb una confiança del 95%.**
- **b) Feu un test, amb una confiança del 95%, per decidir si la demora és diferent segons si l'aparició dels símptomes es produeixen durant el dia o durant la nit.**
- **c) Feu un test, amb una confiança del 95%, per decidir si la demora és diferent segons el sexe del pacient.**