PAC4 (Pràctica)

Josep Andreu Miralles

22/11/2020

output: pdf_document: latex_engine: xelatex —

EXERCICI 1

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 i 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 online 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%.

Solució

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 i 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%.

Hem de realitzar un contrast sobre proporcions i comprobar si amb el nou medicament hi ha hagut un increment en la proporció de pacients amb el colesterol ben controlat. Així doncs, s'ha d'establir un hipòtesi alternativa unilateral (una cua) tenint en compte que en tamanys mostrals grans es segueix una distribució normal del teorema del límit central (aproximació binomial a distribució normal).

p1: Proporció de malalts que tenen ben controlat el nivell de colesterol amb el nou medicament. p2: Proporció de malalts que tenen ben controlat el nivell de colesterol amb el tractament habitual. $\bar{p}$: Estimació de la proporció poblacional comuna.

Hipòtesi nula la proporció de pacients es manté invariable:

$$H_1 : p_1 = p_2$$

Contra la hipòtesi alternativa (unilateral) en que la proporció de pacients que tenen controlat el colesterol es superior amb el nou medicament que no amb el tractament habitual.

$$H_1 : p_1 > p_2$$

L'estadístic de contrast és:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{((\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

L'error estàndard és:

$$s_p = \sqrt{((\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2}))}$$

El valor crític amb nivell de significació igual a 0,05 (una cua) segons les taules de distribució normal és:

$$z_{\alpha=0.05} = 1.645$$

b) Es passa una prova de coneixements d'estadística a un grup de 100 alumnes d'institut. Posteriorment, s'efectua una formació específica online 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%.

Hem de realitzar un contrast de dades aparellades, ja que els individus objecte de les dues observacions (abans i desprès de la formació) són els mateixos, tenim doncs dues observacions per cada individu. Hem de comprobar que hi ha hagut una millora en el coneixement, per la qual cosa es tracta d'establir un hipòtesi alternativa unilateral. Es desconeix la variança per la qual cossa s'assimila a una disrtribució t de Student.

$\bar{x}$: Mitjana de la diferència de les mostres. n: Tamany de la mostra. Igual a 100.

Hipòtesi nula no hi ha hagut millora en el coneixement desprès de la formació:

$$H_0 : \mu = 0$$

Contra la hipòtesi alternativa (unilateral) en que hi ha hagut una millora en el coneixement desprès de la formació.

$$H_1 : \mu > 0$$

L'estadístic de contrast és:

$$\frac{\bar{x} + \mu_0}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} + 0}{\frac{s}{\sqrt{100}}} = \frac{10\bar{x}}{s}$$

L'error estàndard és:

$$\frac{s}{\sqrt{n}} = \frac{s}{\sqrt{100}} = \frac{s}{10}$$

El valor crític amb nivell de significació igual a 0,05 (una cua) i n–1=99 graus de llibertat segons les taules de distribució t de Student és:

$$t_{99/\alpha=0.05} = 1,6604$$

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%.

Hem de realitzar un contrast sobre proporcions i comprobar si el percentatge que superen el test d'estrès es diferent segons el proveidor. Així doncs, s'ha d'establir un hipòtesi alternativa bilateral (dues cues) tenint en compte que en tamanys mostrals grans es segueix una distribució normal del teorema del límit central (aproximació binomial a distribució normal).

p1: Proporció de peces proporcionades pel proveïdor A que superen el test d'estrès. p2: Proporció de peces proporcionades pel proveïdor B que superen el test d'estrès. $\bar{p}$: Estimació de la proporció poblacional comuna.

Hipòtesi nula la proporció de peces defectuoses és la mateixa en les dues mostres:

$$H_1 : p_1 = p_2$$

Contra la hipòtesi alternativa (bilateral) en que la proporció de peces defectuoses de les dues mostres no és igual.

$$H_1 : p_1 \neq p_2$$

L'estadístic de contrast és:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{((\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

L'error estàndard és:

$$s_p = \sqrt{((\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2}))}$$

El valor crític amb nivell de significació igual a 0,05 (dues cues) segons les taules de distribució normal és :

$$z_{\alpha/2=0.025} = 1.96$$

EXERCICI 2

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.

- 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.

Solució

- **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%.**

Suposem la hipòtesi nula on la mitjana de la demora és igual a 300 i la hipòtesi alternativa en que la mitjana de la demora és diferent a 300.

```
library (foreign)
dades_demora<-read.dta("demora.dta")
demora<-dades_demora$demora
t.test(dades_demora$demora,getOption="na.action", alternative="two.sided",mu=300,
       conf.level=0.95)
```

```
##
##  One Sample t-test
##
## data:  dades_demora$demora
## t = 2.3547, df = 382, p-value = 0.01904
## alternative hypothesis: true mean is not equal to 300
## 95 percent confidence interval:
##  313.4344 449.4116
## sample estimates:
## mean of x
##   381.423
```

Com que el p-valor (0.01904) és més petit que el nivell de significació 0.05 rebutgem que la mitjana és igual a 300. L'interval de confiança al 95% per la mitjana és (313.4344, 449.4116) i com 300 està fora de l'interval rebutgem també la hipòtesi nul · la. Això està d'acord amb la conclusió que hem obtingu també el p-valor.

- **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.**

Suposem la hipòtesi nula on les mitjanes de la demora segons sigui de dia o de nit són iguals i la hipòtesi alternativa on les mitjanes segons sigui de dia o de nit són diferents.

```
demora_dia=dades_demora$demora[dades_demora$noche == "Dia"]
demora_noche=dades_demora$demora[dades_demora$noche == "Noche"]
t.test(demora_dia,demora_noche,getOption="na.action",alternative="two.sided",var.equal=TRUE,
       conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  demora_dia and demora_noche
## t = -7.2005, df = 381, p-value = 3.225e-12
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -682.4182 -389.6678
## sample estimates:
## mean of x mean of y
##  244.2632  780.3061
```

El p-valor és més petit que el nivell de significació (p-valor=3.225e-12 < 0.05), rebutgem doncs la pòtesi nula. Podem afirmar que la demora és diferent si apareixen el símptomes de dia o de nit.

- ### c) Feu un test, amb una confiança del 95%, per decidir si la demora és diferent segons el sexe del pacient.

Suposem la hipòtesi nula on les mitjanes de la demora segons els pacients siguin homes o dones són iguals i la hipòtesi alternativa on les mitjanes segons els pacients siguin homes o dones són diferents.

```
demora_home=dades_demora$demora[dades_demora$sexo == "Hombre"]
demora_dona=dades_demora$demora[dades_demora$sexo == "Mujer"]
t.test(demora_home,demora_dona,getOption="na.action",alternative="two.sided",
       var.equal=TRUE,conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  demora_home and demora_dona
## t = 0.94597, df = 381, p-value = 0.3448
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -78.86374 225.10957
## sample estimates:
## mean of x mean of y
##  401.6606  328.5377
```

El p-valor és més gran que el nivell de significació (p-valor=0.3448 > 0.05), no rebutgem doncs la hipòtesi nula. No tenim indicis suficients per afirmar que el temps de demora per a pacients homes sigui diferent al temps de demora per a pacients dones.