

Contrast de dues mostres

Selecció d'activitats resoltes

Jose Fco. Martínez Boscá, Arnau Mir Torres, Lluís M. Pla Aragonés,
Àngel J. Gil Estallo (autors) i Àngel A. Juan (editor)

Introducció

En aquest mòdul es pretén calcular i interpretar els contrastos sobre la diferència de mitjanes i proporcions entre dues poblacions que permetin prendre decisions sobre quina població hem de tenir en compte en relació amb l'altra.

A més de calcular intervals de confiança (rang de valors en el qual s'espera trobar un determinat paràmetre de la població), es farà el que denominarem *prova d'hipòtesi* al voltant d'una afirmació sobre un paràmetre de la població. Per a posar de manifest les seves aplicacions en la vida real, donarem exemples d'activitats en l'àmbit econòmic empresarial i en l'informàtic.

Fins ara, havíem utilitzat una sola mostra aleatòria, comparant la seva mitjana amb un valor suposat de la mitjana poblacional, és a dir, ens plantejàvem si era possible que una mostra amb una mitjana donada pogués provenir d'una població amb la mitjana proposada. En aquest cas, estendrem la idea anterior a dues mostres, preguntant-nos si les mitjanes de totes dues són iguals o no, és a dir, el plantejament serà raonar si és possible que les dues mitjanes mostrals puguin provenir de dues poblacions idèntiques.

Exemple

Per exemple, en una empresa informàtica es desitja mesurar l'eficiència de dos servidors web. Per a això, mesuren el temps d'espera del client entre la petició que fa i la resposta que li dona el servidor. Es pot demanar:

- Contrasteu si la variabilitat del temps d'espera és més gran en el servidor A que en el B. Considereu $\alpha = 0,1$. Trobeu el p -valor del contrast.
- Contrasteu si podem considerar que el servidor A és menys eficient que el servidor B. Considereu $\alpha = 0,1$. Trobeu el p -valor del contrast.
- Trobeu un interval de confiança per a la diferència de temps d'espera entre tots dos servidors al 99% de confiança.
- Direm que el temps d'espera és acceptable si és més petit que 9 mil·lisegons. Podem dir que la proporció de peticions amb temps d'espera acceptable és diferent per als dos servidors? Considereu $\alpha = 0,1$. Trobeu el p -valor del contrast.

Diferència entre mostres independents i dependents

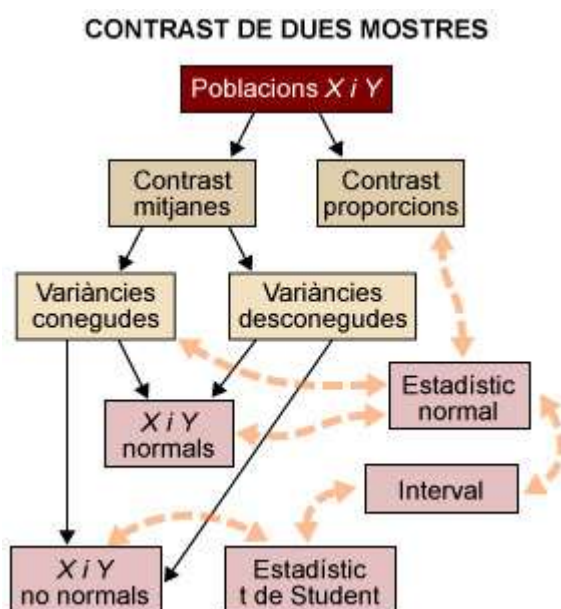
Dues mostres són independents o dependents entre si, en funció de si les observacions de les mostres s'han obtingut dels mateixos individus o objectes o no. Si totes dues mostres s'obtenen de diferents individus, màquines, empreses, objectes, etc., no tenen res en comú, fet que provoca que totes dues siguin **independents**. No obstant això, si les observacions o valors de les mostres s'obtenen dels mateixos individus, empreses, agents, etc., direm

que hi ha alguna cosa en comú en aquestes mostres, i per això seran mostres **dependents o no independents**.

Suposem que volem comparar els beneficis empresarials del sector de les telecomunicacions entre els anys 2001 i 2002. Per a això, podem prendre una mostra aleatòria formada per 50 empreses de telecomunicacions de tot el país i mesurar els seus beneficis en l'any 2001. A continuació, per a poder comparar els beneficis del sector amb l'any 2002, es pren una altra mostra aleatòria diferent amb unes altres 30 empreses de telecomunicacions i analitzem els seus beneficis en l'any 2002. En aquest cas es tracta de mostres independents, atès que les observacions de totes dues mostres es prenen de diferents individus, en aquest cas de diferents empreses. No obstant això, si en l'any 2002 observem els beneficis de les mateixes 50 empreses de telecomunicacions de la mostra de l'any 2001, ens trobem davant de mostres dependents o aparellades.

Encara que el contrast sobre variàncies és del mòdul següent, en farem algun per a saber si són iguals o diferents. Així podem decidir quin tipus de contrast portarem a terme sobre les mitjanes i l'estadístic corresponent.

Mapa conceptual



Activitats Resoltes - Contrast d'hipòtesis dues mostres

Activitats Resoltes: Contrast d'hipòtesis dues mostres.

Fitxers necessaris per realitzar les activitats:

- ActR08TATB.csv
- vendes_pac1_P_15_3.csv
- vendes_pac1_P_15_4.csv

A les primeres files dels fitxers consten les descripcions de les variables i al moment d'importar-los a R cal tenir en compte que aquestes línies no contenen dades (s'aconsella usar *skip*).

En altres activitats en les quals el nombre d'observacions és petit se suggereix copiar-enganxar les dades o la instrucció que s'usa per introduir-los en R.

Activitat 1. Eficiència de servidors Web.

Comparació de mitjanes. Comparació de proporcions.

En una empresa informàtica es desitja mesurar l'eficiència de dos servidors Web. Per a això, mesuren el temps d'espera del client entre la petició que aquest fa i la resposta que li dona el servidor.

Els temps d'espera (en mil · lisecons) de tots dos servidors (TA i TB) per 50 peticions es troben en el fitxer ActR08TATB.csv.

- a) Contrasteu si podem considerar que el servidor A és menys eficient que el servidor B. Preneu $\alpha = 0.01$. Trobeu el p-valor del contrast.
- b) Trobeu un interval de confiança per a la diferència de temps d'espera entre tots dos servidors al 99% de confiança. Supposeu que les variàncies són iguals.
- c) Direm que el temps d'espera és acceptable si és menor que 9 mil · lisecons. Podem dir que la proporció de peticions amb temps d'espera acceptable és diferent per als dos servidors? Preneu $\alpha = 0.05$. Trobeu el p-valor del contrast. Indicació: Per fer aquest problema, calculeu una nova variable amb R que valgui 1 si el temps d'espera és menor que 9 mil · lisecons i 0 en cas contrari. Per calcular aquesta variable, podem utilitzar les instruccions *for* i *if* de R.

Solució

a) Primer llegim el fitxer

```
ActR081<-read.table("ActR08TATB.csv", sep=";" , header=TRUE)
```

Hem de fer un contrast per comparar dues mitjanes. Atès que l'enunciat ens pregunta “si el servidor A és menys eficient que el servidor B”, considerant que un servidor és menys eficient si és més lent, llavors hem de contrastar si la mitjana del temps d’espera del servidor A és major que la mitjana del temps d’espera del servidor B. Així doncs, hem de considerar una hipòtesi alternativa unilateral. Hem de considerar el cas de no normalitat però amb mostres grans (superior a 30 observacions). Les hipòtesis nul·la i alternativa són $H_0 : \mu_A - \mu_B = 0$, $H_1 : \mu_A - \mu_B > 0$ i fixem $\alpha = 0.01$. L'estadístic de contrast és

$$z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

i la distribució és una $N(0, 1)$. Procedim de la següent forma:

```
z<-(mean(ActR081$TA)-mean(ActR081$TB))/  
  sqrt((var(ActR081$TA)/length(ActR081$TA)) +  
        (var(ActR081$TB)/length(ActR081$TB)))  
z
```

```
## [1] 3.713408
```

i el p-valor és $P(Z > 3.713408)$:

```
pnorm(z, lower.tail=FALSE)
```

```
## [1] 0.0001022434
```

Tenim que el p-valor és pràcticament 0 de manera que rebutgem la hipòtesi nul·la en favor de la hipòtesi alternativa i concloem que les mitjanes de temps d’espera són diferents en cada servidor. Com que el p-valor és menor que el nivell de significació, rebutgem la hipòtesi nul·la i donem per bona l’alternativa; el servidor A és menys eficient que el servidor B.

b) Hem de fer un contrast bilateral:

```
t.test(ActR081$TA, ActR081$TB, var.equal=TRUE, mu=0, conf.level=0.99)
```

```
##  
## Two Sample t-test  
##  
## data: ActR081$TA and ActR081$TB  
## t = 3.7134, df = 98, p-value = 0.0003399  
## alternative hypothesis: true difference in means is not equal to 0  
## 99 percent confidence interval:
```

```
## 0.3020033 1.7623967
## sample estimates:
## mean of x mean of y
## 9.9350 8.9028
```

Per tant, l'interval és (0.302, 1.762).

- c) El primer que s'ha de fer és calcular les dues noves variables (una per a cada servidor, s'ensenyen els deu primers valors de cadascuna):

```
for (i in 1:length(ActR081$TA))
  { if(ActR081$TA[i] < 9) {ActR081$PA[i]<-1}
    else {ActR081$PA[i]<-0}
  }
ActR081$PA[1:10]
```

```
## [1] 0 0 0 0 1 0 0 0 0 0
```

i per al servidor B:

```
for (i in 1:length(ActR081$TB))
  { if(ActR081$TB[i] < 9) {ActR081$PB[i]<-1}
    else {ActR081$PB[i]<-0}
  }
ActR081$PB[1:9]
```

```
## [1] 1 0 1 0 1 1 0 1 1
```

Per saber la proporció es pot fer el següent, ja que la proporció es pot veure com la mitjana en una població binària:

```
summary(ActR081$PA)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00   0.00   0.12   0.00   1.00
```

```
summary(ActR081$PB)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00   1.00   0.62   1.00   1.00
```

o fer-ho directament

```
pa<-length(ActR081$PA[ActR081$PA==1])/length(ActR081$PA)
pb<-length(ActR081$PB[ActR081$PB==1])/length(ActR081$PB)
pa
```

```
## [1] 0.12
```

```
pb
```

```
## [1] 0.62
```

Per tant les proporcions mostrals són 0.12 per al servidor A i 0.62 per al servidor B. Haurem de fer ara un contrast bilateral sobre diferència de proporcions amb nivell de significació 0.05. Les hipòtesis són: $H_0 : p_A = p_B, H_1 : p_A \neq p_B$. Primer calculem l'estimació de la proporció poblacional comuna:

```
pcom<-(length(ActR081$TA)*0.12 +length(ActR081$TB)*0.62)/
      (length(ActR081$TA)+length(ActR081$TB))
pcom
```

```
## [1] 0.37
```

Ara l'error estàndard:

```
sp <- sqrt(pcom*(1-pcom)*(1/length(ActR081$TA) + 1/length(ActR081$TB)))
sp
```

```
## [1] 0.09656086
```

I finalment l'estadístic de contrast:

```
z <- (pa-pb)/sp
z
```

```
## [1] -5.178081
```

la p de la qual-valor és $p = 2P(Z > |z|)$:

```
2*pnorm(abs(z),lower.tail=FALSE)
```

```
## [1] 2.241794e-07
```

Com que el p-valor (0.0000) és menor que el nivell de significació (0.05), rebutgem la hipòtesi nul·la i concloem que la proporció de peticions amb temps d'espera acceptable és diferent per als dos servidors.

Activitat 2. Cintes magnètiques

Contrast de mitjanes.

En una empresa s'utilitzen per fer les còpies de seguretat cintes magnètiques de dos tipus diferents A i B. La qualitat de les cintes depèn del nombre de partícules magnètiques (NPM) per micra quadrada en la superfície de la cinta, de manera que quantes més partícules, més qualitat té la cinta. Per estudiar la qualitat de les cintes, s'han pres algunes mesures del NPM de 10 cintes de tipus A i 7 cintes de tipus B i es va obtenir: $x_A = 122.23, x_B = 124.2857, s_A^2 = 337.938$ i $s_B^2 = 472.231$. Suposant que les variàncies del NPM dels dos tipus de cintes són iguals, podem pensar que els dos tipus de cintes tenen, en mitjana, la mateixa qualitat? (Preneu $\alpha = 0.05$). Trobeu el p-valor del contrast.

Solució

Hem de fer un contrast d'igualtat de mitjanes de dues mostres amb variàncies iguals. El contrast d'hipòtesi és: $H_0 : m_A = m_B$, $H_1 : m_A \neq m_B$. L'estadístic de contrast val

$$t = \frac{\bar{x}_A - \bar{x}_B}{s\sqrt{\frac{1}{10} + \frac{1}{7}}}$$

on \bar{x}_A i \bar{x}_B són les mitjanes dels dos tipus de cintes que valen $\bar{x}_A = 122.23$, $\bar{x}_B = 124.2857$, i s és la desviació típica comuna

$$s = \sqrt{\frac{9s_A^2 + 6s_B^2}{15}} = 19.79$$

Els càlculs són

```
scom<-sqrt((9*337.938+6*472.231)/15)
scom
```

```
## [1] 19.79028
```

```
t<-(122.23-124.2857)/(scom*sqrt((1/10)+(1/7)))
t
```

```
## [1] -0.2107814
```

El p-valor és

```
2*pt(-0.2107814,df=15)
```

```
## [1] 0.8358945
```

I el valor crític

```
qt(0.025,df=15,lower.tail=FALSE)
```

```
## [1] 2.13145
```

i per tant acceptem la hipòtesi nul·la i concloem que els dos tipus de cintes tenen, en mitjana la mateixa qualitat.

Activitat 3

Comparació de mitjanes

Volem comparar la rapidesa de dues impressores en imprimir una fotografia. Per a això escollim 10 fotos a l'atzar i les imprimim en les dues impressores, observant quant temps triga cadascuna d'elles a imprimir cada foto. Els resultats obtinguts apareixen en la següent taula:

Impressora A	47.2	72.54	59	46.0	74.1	72.5	46.9	78.6	42	63.3
Impressora B	46.1	82.8	63.6	43.9	76.4	82.5	50.4	77.2	40	69.8

Suposant que el temps d'impressió d'una foto es distribueix normalment i que les variàncies poblacionals són desconegudes i iguals.

- Calculeu un interval de confiança per a la mitjana de cadascuna de les dues poblacions al nivell de confiança del 90%. Comentar els resultats.
- Calculeu un interval de confiança per a la diferència de mitjanes. Utilitzant aquest interval contrasteu la hipòtesi que les mitjanes en els dos grups no són diferents (nivell de confiança del 90%).
- Plantegeu les hipòtesis a contrastar per a la diferència de mitjanes i la fórmula de l'estadístic del contrast així com el seu valor i la seva distribució.
- A partir del criteri del p-valor, a quina conclusió s'arriba? Quin error d'equivocar-ens hauríem d'estar disposats a assumir?
- Quin hauria estat el resultat del contrast si suposem que, en lloc d'utilitzar les mateixes fotos per a l'estudi, s'han utilitzat 10 per a una impressora i altres 10 diferents per a l'altra? Llavors, planteja les hipòtesis a contrastar sobre la igualtat de les mitjanes i la conclusió a la qual s'arribaria amb el criteri del p-valor. Compara els resultats amb els de l'apartat anterior

Solució:

- Introduïm les dades:

```
ia<-c(47.2 , 72.54 , 59 , 46.0 , 74.1 , 72.5 , 46.9 , 78.6 , 42 , 63.3)
ib<-c(46.1 , 82.8 , 63.6 , 43.9 , 76.4 , 82.5 , 50.4 , 77.2 , 40 , 69.8)
```

Atès que desconexim la variància de la població, utilitzarem la *t*-Student amb el següent test:

```
t.test(ia, alternative='two.sided', mu=0.0, conf.level=.90)
```

```
##
## One Sample t-test
##
## data:  ia
## t = 13.753, df = 9, p-value = 2.391e-07
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  52.18832 68.23968
## sample estimates:
```

```
## mean of x
##      60.214

t.test(ib, alternative='two.sided', mu=0.0, conf.level=.90)

##
## One Sample t-test
##
## data:  ib
## t = 11.916, df = 9, p-value = 8.177e-07
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  53.53638 73.00362
## sample estimates:
## mean of x
##      63.27
```

Si ens fixem en els dos intervals de confiança, aquests se superposen. A més, les mitjanes mostrals són també molt semblants. Això fa pensar que aquestes mitjanes poblacionals, és a dir, les mitjanes de les impressores, poden ser iguals.

- b) Per calcular un interval de confiança per a la diferència de mitjanes amb dades aparellades fem

```
t.test(ia, ib, alternative='two.sided', conf.level=.90, paired=TRUE)

##
## Paired t-test
##
## data:  ia and ib
## t = -2.0278, df = 9, p-value = 0.07319
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -5.8185811 -0.2934189
## sample estimates:
## mean of the differences
##                  -3.056
```

El valor zero no està en l'interval de confiança, llavors podem concloure que les mitjanes són diferents. També ens podem fixar en el p-valor que és $0.073 < 0.1$, per tant rebutgem la H_0 .

- c) Es realitza un contrast de mitjanes en dues mostres aparellades: $H_0 : d_{A-B} = 0$, $H_1 : d_{A-B} \neq 0$ amb d_{A-B} representant la diferència mitjana del temps d'impressió d'una foto utilitzant una impressora i l'altra. L'enunciat diu que les variàncies poblacionals són desconegudes però iguals. L'estadístic de contrast

$$t = \frac{\bar{x}_{A-B}}{\frac{s}{\sqrt{n}}}$$

és una t de Student amb 9 graus de llibertat, on s és la desviació típica de les diferències. El valor de l'estadístic de contrast per a aquest problema val $t = -2.0278$.

- d) El p-valor val 0.073, com és més petit que 0.10 rebutgem la hipòtesi nul·la, per tant acceptem que les mitjanes no són iguals. L'error d'equivocar-nos en rebutjar la hipòtesi d'igualtat de mitjanes és de 0.073.
- e) Per fer un contrast per a la diferència de mitjanes amb dues mostres independents fem

```
t.test(ia, ib, alternative='two.sided', conf.level=.95)

##
##  Welch Two Sample t-test
##
## data:  ia and ib
## t = -0.44405, df = 17.369, p-value = 0.6625
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -17.55249  11.44049
## sample estimates:
## mean of x mean of y
##    60.214    63.270
```

El valor zero està en l'interval de confiança, llavors podem concloure que les mitjanes no són diferents. També ens podem fixar en el p-valor que és molt superior al nivell de significació, 10%, per tant acceptem la H_0 .

Les hipòtesis del contrast són $H_0 : \mu_A - \mu_B = 0$, $H_1 : \mu_A - \mu_B \neq 0$ on apareixen el temps mitjà d'impressió d'una foto usant la impressora A i el temps mitjà d'impressió d'una foto usant la impressora B. L'estadístic de contrast és

$$t = \frac{\bar{x}_A - \bar{x}_B}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

que segueix una distribució t de Student amb 18 graus de llibertat, on s és la desviació típica comuna:

$$s = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}.$$

El valor de l'estadístic de contrast per a aquest problema val $t = -0.444$. El p-valor val 0.663, com és més gran que 0.10 no podem rebutjar la hipòtesi nul·la, per tant acceptem la igualtat de mitjanes. Així la conclusió no és la mateixa que l'obtinguda amb dades aparellades, de fet és totalment diferent, abans hem rebutjat la hipòtesi nul·la i ara l'hem acceptada. També podem observar que l'error d'equivocar ens en rebutjar la hipòtesi d'igualtat de mitjanes (diferències igual a 0) és de 0.663 2 molt superior al 0.073 anterior.

Activitat 4.

Contrast de mitjanes.

Considerem el fitxer de R “vendes_pac1_P_15_3.csv”. tenint en compte que el separador decimal és la coma “,” i el de camps és el “;”. Les variables del fitxer són:

- m2: superfície
 - Ubi: ubicació (1 Centre ciutat, 2 Centre comercial, 3 Carrer per als vianants, 4 Barris, 5 Extraradi)
 - PreuAm2: preu del lloguer per m2 abans de fer reformes
 - PreuDm2: preu del lloguer per m2 després de fer reformes
 - AugmentFact: Augment de la facturació durant l’últim any
- a) Definir les variables “Preu del lloguer dels pisos abans de fer reformes amb menys de 300 m2” i “Preu del lloguer dels pisos abans de fer reforma amb més de 300 m2”. Anomeneu-les Preu.Lloguer.Abans.Menys.300 i Preu.Lloguer.Abans.Mas.300 respectivament.
- b) Contrastar a un nivell de confiança del 95% si el preu del lloguer abans de fer reformes dels pisos amb menys de 300 m2 és major que el preu del lloguer abans de fer reforma dels pisos amb més de 300 m2. Suposar que les variables considerades són normals amb la mateixa variància.
- c) Contrastar si el preu del lloguer dels pisos després de fer reformes és el mateix que el preu de lloguer abans de fer-les.
- d) Trobar un interval de confiança al 90% de confiança de la diferència del preu dels pisos abans i després de les reformes. Segons aquest últim interval, diríeu que el preu dels pisos ha pujat o no després de fer reformes? Raonar la resposta.

Solució

- a) Primer carreguem les dades:

```
ActR084<-read.table("vendes_pac1_P_15_3.csv", sep=";", dec=",", header=TRUE)
```

Les variables seran les següents:

```
Preu.Lloguer.Abans.Menys.300 = ActR084$PreuAm2[ActR084$m2 <300]  
Preu.Lloguer.Abans.Mas.300 = ActR084$PreuAm2[ActR084$m2 > 300]
```

- b) El test serà

```
t.test(Preu.Lloguer.Abans.Menys.300,Preu.Lloguer.Abans.Mas.300,  
       alternative = "greater",var.equal=TRUE)
```

```
##  
## Two Sample t-test  
##
```

```
## data: Preu.Lloguer.Abans.Menys.300 and Preu.Lloguer.Abans.Mas.300
## t = 0.85844, df = 97, p-value = 0.1964
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.3814182      Inf
## sample estimates:
## mean of x mean of y
## 11.80435 11.39623
```

Com el p-valor és major que el nivell de significació 0.05, concloem que no tenim suficients indicis per afirmar que el preu abans de fer reformes és major per als pisos amb menys de 300 m2.

c) El contrast demanat és:

```
Preu.Lloguer.Abans = ActR084$PreuAm2
Preu.Lloguer.Despues = ActR084$PreuDm2
t.test(Preu.Lloguer.Abans,Preu.Lloguer.Despues,paired =TRUE)

##
## Paired t-test
##
## data: Preu.Lloguer.Abans and Preu.Lloguer.Despues
## t = 0.66342, df = 98, p-value = 0.5086
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1810264 0.3628446
## sample estimates:
## mean of the differences
## 0.09090909
```

El p-valor també és més gran que el nivell de significació. Per tant, tampoc tenim suficients indicis per afirmar que el preu després de fer reformes és major per als pisos amb menys de 300 m2.

d) L'interval demanat és:

```
Dif.preu=ActR084$PreuDm2 - ActR084$PreuAm2
t.test(Dif.preu)

##
## One Sample t-test
##
## data: Dif.preu
## t = -0.66342, df = 98, p-value = 0.5086
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.3628446 0.1810264
```

```
## sample estimates:  
##   mean of x  
## -0.09090909
```

Com que el valor 0 està dins de l'interval, podem dir que no tenim suficients indicis per poder afirmar que el preu dels pisos ha pujat.

Activitat 5

Contrast de mitjanes.

Considerem el fitxer de R “vendes_pac1_P_15_4.csv” en el qual el separador decimal és la coma “,” i el de camps és “;”. Les variables del fitxer són:

- m2: superfície
 - Ubi: ubicació (1 Centre ciutat, 2 Centre comercial, 3 Carrer per als vianants, 4 Barris, 5 Extraradi)
 - PreuAm2: preu del lloguer per m2 abans de fer reformes
 - PreuDm2: preu del lloguer per m2 després de fer reformes
 - AugmentFact: Augment de la facturació durant l'últim any
- a) Definir les variables “Preu del lloguer dels locals comercials situats al centre de la ciutat abans de fer reformes” i “Preu del lloguer dels locals comercials situats als barris abans de fer reformes”. Crideu-les Preu.Lloguer.Abans.Centre i Preu.Lloguer.Abans.Barris, respectivament.
- b) Contrastar a un nivell de confiança del 95% si el preu del lloguer abans de fer reformes dels locals comercials situats al centre de la ciutat és major que el preu del lloguer abans de fer reforma dels locals comercials situats als barris. Suposar que les variables considerades són normals amb la mateixa variància.
- c) Fer el mateix que als apartats a) i b) però ara en lloc de considerar el preu del lloguer abans de les reformes, considerar el preu del lloguer després de les reformes. O sigui, contrastar a un nivell de confiança del 95% si el preu del lloguer després de fer reformes dels locals comercials situats al centre de la ciutat és major que el preu del lloguer després de fer reforma dels locals comercials situats als barris. Suposar que les variables considerades són normals amb la mateixa variància. Crideu a les noves variables Preu.Lloguer.Després.Centre i Preu.Lloguer.Després.Barris.
- d) Trobar un interval de confiança al 95% de confiança de la diferència del preu dels locals comercials situats al centre de la ciutat abans i després de les reformes. Segons aquest últim interval, diríeu que el preu dels locals comercials ha pujat o no després de fer reformes en els locals comercials situats al centre de la ciutat? Raonar la resposta.

Solució

- a) Primer carreguem les dades:

```
ActR085<-read.table("vendes_pac1_P_15_4.csv", sep=";" ,dec=",", header=TRUE)
```

Les variables seran les següents:

```
Preu.Lloguer.Abans.Centre = ActR085$PreuAm2[ActR085$Ubi==1]  
Preu.Lloguer.Abans.Barris = ActR085$PreuAm2[ActR085$Ubi==4]
```

- b) El contrast és

```
t.test(Preu.Lloguer.Abans.Centre,Preu.Lloguer.Abans.Barris,  
       alternative = "greater",var.equal=TRUE)
```

```
##  
## Two Sample t-test  
##  
## data: Preu.Lloguer.Abans.Centre and Preu.Lloguer.Abans.Barris  
## t = 0.54363, df = 37, p-value = 0.295  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## -1.004955 Inf  
## sample estimates:  
## mean of x mean of y  
## 11.36667 10.88889
```

Com que el p-valor és major que el nivell de significació 0.05, concloem que no tenim suficients indicis per afirmar que el preu del lloguer dels locals comercials abans de fer reformes situats al centre de la ciutat sigui major que el preu del lloguer dels locals comercials abans de fer reformes situats als barris.

- c) Les variables i el contrast seran les següents

```
Preu.Lloguer.Després.Centre = ActR085$PreuDm2[ActR085$Ubi==1]  
Preu.Lloguer.Després.Barris = ActR085$PreuDm2[ActR085$Ubi==4]  
t.test(Preu.Lloguer.Després.Centre,Preu.Lloguer.Després.Barris)
```

```
##  
## Welch Two Sample t-test  
##  
## data: Preu.Lloguer.Després.Centre and Preu.Lloguer.Després.Barris  
## t = 1.173, df = 12.989, p-value = 0.2619  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.9729606 3.2840717  
## sample estimates:
```



```
## mean of x mean of y
## 11.60000 10.44444
```

Com que el p-valor és major que el nivell de significació 0.05, concloem que no tenim suficients indicis per afirmar que el preu del lloguer dels locals comercials després de fer reformes situats al centre de la ciutat sigui major que el preu del lloguer dels locals comercials després de fer reformes situats als barris.

d) L'interval demanat és:

```
t.test(Preu.Lloguer.Abans.Centre,Preu.Lloguer.Després.Centre,
       conf.level = 0.95,paired=TRUE)
```

```
##
## Paired t-test
##
## data: Preu.Lloguer.Abans.Centre and Preu.Lloguer.Després.Centre
## t = -0.89323, df = 29, p-value = 0.3791
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.7675947 0.3009281
## sample estimates:
## mean of the differences
## -0.2333333
```

Com que el valor 0 està dins de l'interval, podem dir que no tenim suficients indicis per poder afirmar que el preu dels locals comercials situats al centre de la ciutat hagi canviat.