

UOC – ESTUDIS D'INFORMÀTICA

CURS 2019-20 / 2

EXERCICI DE R CORRESPONENT A LA PAC-5

Considerarem de nou el fitxer “PIS_MUN.csv” presentat a la PAC-1 on podem trobar informació sobre el nombre de contractes de lloguer i el lloguer mitjà contractual (mitjana dels preus dels lloguers per contracte) als municipis de Catalunya i als districtes de la ciutat de Barcelona.

Suposem que les observacions corresponen a una mostra.

Exercici-1 (50%)

- Contrasteu a un nivell de confiança del 95% si la mitjana dels preus mitjans dels lloguers als municipis de Tarragona de l'any 2016 és **major** que la mitjana dels preus mitjans dels lloguers als municipis de Lleida d'aquest mateix any 2016. Suposeu que les variables considerades són normals amb la mateixa variància. **Considereu tots els registres en que el lloguer sigui diferent de zero.** Feu copy & paste de la/les instruccions de R i de la sortida del programa.
- Indiqueu les hipòtesis nul·la i alternativa, el p-valor i la conclusió a la que arribeu.
- Quin interval de confiança proporciona el R per a la diferència de mitjanes dels preus mitjans dels lloguers nombre de contractes? Està d'acord amb el resultat trobat en el contrast de l'apartat anterior?

Indicació: Mòdul 8 (Contrast de dues mostres) dels Mòduls d'Estadística. I dels mòduls de R, mireu el mòdul 4 (Distribucions de probabilitat i inferència estadística amb R-Commander), pàgines 37 i 38.

Criteris de puntuació i valoració (sobre 10): a) (4 punts), b) (4 punts), c) (2 punts).

Solució:

- Contrast sobre la diferència de mitjanes.

Primer podem definir els conjunts de dades de Lleida i de Tarragona:

```
> dades_lleida<-subset(dades,PROV=="Lleida" & M2016!=0)
> dades_tarragona<-subset(dades,PROV=="Tarragona" & M2016!=0)
```

Per fer el contrast executem la següent instrucció de R:

```
> t.test(dades_tarragona$M2016,dades_lleida$M2016, alternative
= "greater", var.equal = TRUE, conf.level = 0.95)
```

Obtenim la següent sortida:

Two Sample t-test

```
data:  dades_tarragona$M2016 and dades_lleida$M2016
t = 2.9873, df = 156, p-value = 0.001635
alternative hypothesis: true difference in means is greater
than 0
95 percent confidence interval:
  17.94344      Inf
sample estimates:
mean of x mean of y
380.8882  340.6656
```

Les hipòtesis nul·la i alternativa són:

Hipòtesi nul·la: $H_0: \mu_{Tarragona_M2016} = \mu_{Lleida_M2016}$

Hipòtesi alternativa: $H_1: \mu_{Tarragona_M2016} > \mu_{Lleida_M2016}$

El p-valor = 0.001635 < 0.05 per tant, rebutgem la hipòtesi nul·la i concloem que la mitjana dels preus dels lloguers mitjans als municipis de Tarragona del 2016 és superior a la mitjana en Lleida.

- b) El R ens proporciona l'interval de confiança (17.94344, Infinit). El zero no es troba dintre de l'interval i a més, l'interval correspon a valors positius, per tant, està d'acord amb la conclusió de l'apartat anterior.

Exercici-2 (50%)

Contrasteu a un nivell de confiança del 99% si la proporció de contractes fets als municipis/districtes de Barcelona amb lloguers superiors als 750 euros/mes corresponents a l'any 2018 és **igual** al 25%. **Considereu només registres de cada any on el preu mitjà sigui diferent de zero.** Per respondre seguiu el següent esquema:

- Primer trobeu el nombre contractes fets als municipis/districtes de Barcelona amb lloguers diferents de zero a l'any 2018 i després el nombre contractes fets als municipis/districtes de Barcelona amb lloguers superiors als 750 euros/mes corresponents a l'any 2018.
- Indiqueu les hipòtesis nul·la i alternativa, el p-valor i la conclusió a la que arribeu. Feu copy & paste de la/les instruccions de R i de la sortida del programa.
- Quin interval de confiança del 99% proporciona el R per a la proporció del nombre de contractes fets als municipis/districtes de Barcelona amb lloguers superiors als 750 euros/mes corresponents a l'any 2018? Està d'acord amb el resultat trobat en el contrast de l'apartat anterior?

Indicació: Mireu el mòdul 7 (Contrast d'hipòtesis) dels Mòduls d'Estadística. I dels Mòduls de R, mireu el mòdul 4 (Distribucions de probabilitat i inferència estadística amb R-Commander), pàgina 34.

Criteris de puntuació i valoració (sobre 10): a) (4 punts), b) (4 punts), c) (2 punts).

Solució:

- a) Ara farem un contrast sobre la proporció. Comencem calculant el nombre contractes fets als municipis/districtes de Barcelona amb lloguers diferents de zero a l'any 2018:

```
> dades_bcn<-subset(dades,PROV=="Barcelona" & M2018!=0)
> Ntot2018<-sum(dades_bcn$N2018);
> Ntot2018
[1] 126791
```

I el nombre contractes fets als municipis/districtes de Barcelona amb lloguers superiors als 750 euros/mes corresponents a l'any 2018:

```
> Nsup750<-sum(dades_bcn$N2018[dades_bcn$M2018>750])
> Nsup750
[1] 56291
```

- b) Les hipòtesi nul·la i alternativa son:

Hipòtesi nul·la: $H_0: p_{bcn_N2018_s750} = 0.25$

Hipòtesi alternativa: $H_1: p_{bcn_N2018_s750} \neq 0.25$

Fem el contrast d'hipòtesis seguint la següent sintaxis de R:

```
> prop.test(Nsup750, Ntot2018, 0.25, alternative="two.sided",
conf.level = 0.99)
```

Obtenim la següent sortida:

```
1-sample proportions test with continuity correction
data:  Nsup750 out of Ntot2018, null probability 0.25
X-squared = 25440, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.25
99 percent confidence interval:
0.4403718 0.4475678
sample estimates:
      p
0.4439668
```

En aquest cas el p-valor $< 2.2e-16 < 0.001$, per tant, rebutgem la hipòtesi nul·la i arribem a la conclusió que la proporció és diferent del 25%.

- c) L'interval de confiança del 99% per a la proporció és (0.4403718, 0.4475678). El 0.25 no està dintre de l'interval, per tant, concloem que la proporció no és del 25%.

Observació: en la instrucció `prop.test` es pot incloure l'argument `"correct=FALSE"`. En l'ajuda de `"prop.test"`, es comenta que amb aquesta opció no es té en compte la correcció de continuïtat de Yates. Tal i com es diu en la pàgina 16 del mòdul 5, podem considerar com correctes els dos resultats, que són molt semblants:

```
> prop.test(Nsup750, Ntot2018, 0.25, correct=FALSE, alternative=
"two.sided", conf.level = 0.99)
```

Obtenim la següent sortida:

```
1-sample proportions test without continuity correction
data:  Nsup750 out of Ntot2018, null probability 0.25
X-squared = 25441, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.25
99 percent confidence interval:
0.4403757 0.4475639
sample estimates:
      p
0.4439668
```