

Part de R de la PAC2. Primavera 2020

Seguint amb l'estudi de les dades del fitxer PIS_MUN.csv on apareix la informació sobre el nombre de contractes de lloguer i el lloguer mitjà contractual (mitjana dels preus dels lloguers per contracte) en els municipis de Catalunya i en els districtes de la ciutat de Barcelona. Concretament trobareu les variables.

COD = codi postal del municipi (o "08019" en cas dels districtes de Barcelona),

MUNDIS = Nom del municipi o districte de la ciutat de Barcelona,

MUN = Nom del municipi o "Barcelona" en el cas dels seus districtes,

PROV = Província a la que pertany el municipi/districte.

NAny = Nombre de contractes signats el corresponent any al municipi/districte (on Any pot ser 2018, 2017, 2016 o 2015),

MAny = Lloguer mitjà mensual dels contractes signats el corresponent any al municipi/districte.

Ara heu de respondre a les següents preguntes en un fitxer pdf que s'ha de lliurar en aquesta activitat.

1. Trobeu la taula que indiqui la quantitat de municipis/districtes de cada província on es van signar més de 5 contractes l'any 2018. A partir d'aquesta taula calculeu: (20 punts)

1.1 La probabilitat que un municipi/districte agafat a l'atzar on es van signar més de 5 contractes l'any 2018 sigui de la província de Barcelona.

1.2 La probabilitat que un municipi/districte agafat a l'atzar on es van signar més de 5 contractes l'any 2018 sigui de la província de Lleida o Tarragona.

2. Estudiem ara el nombre de contractes signats en 2018 distingint per províncies. Codificarem la variable N2018 en quatre categories: pocs, regular, bastants i molts. La primera correspondrà al nombre de contractes entre el seu mínim (inclòs) i Q1(inclòs), la segona entre Q1 (no inclòs) i Q2 (inclòs), la tercera entre Q2 (no inclòs) i Q3 (inclòs) i la quarta entre Q3 (no inclòs) i el màxim (inclòs), considerant sempre intervals de la forma (A,B] excepte el primer, que serà de la forma [Mínim,Q1].

1. Crear la variable NombreContractes_N2018 amb aquestes especificacions i mostreu els 10 primers valors. (10 punts)
2. Trobar la taula de contingència de les dues variables NombreContractes_N2018 i PROV. (10 punts)
3. La probabilitat que un municipi/districte tingui "molts" contractes signats. (10 punts)
4. La probabilitat que un municipi/districte tingui "molts" contractes i sigui de la província de Girona. (10 punts)
5. La probabilitat que un municipi/districte sigui de la província de Lleida sabent que té "pocs" contractes signats. (10 punts)
6. La probabilitat que en un municipi/districte hi hagi "bastants" contractes signats sabent que la província és Tarragona. (10 punts)
7. Els successos "Província de Girona" i "molts contractes", són independents? Per què? (20 punts)

Us pot ser útil la funció *cut* amb “breaks” i “labels” per codificar i *table* per tabular les dades. Mireu les activitats resoltes del mòdul 3.

Indicar les fórmules i calcular les probabilitats manualment a partir de la taula de contingència que ens dona R.

SOLUCIÓ

1. Taula:

```
> TaulaProv<-table(dades$PROV[dades$N2018>5])
> TaulaProv
```

Barcelona	Girona	Lleida	Tarragona	Total
254	141	79	98	572

1.1. La probabilitat que un municipi/districte agafat a l'atzar on es van signar més de 5 contractes l'any 2018 sigui de la província de Barcelona.

$$P(\text{Barcelona})=254/572 = 0.444$$

1.2. La probabilitat que un municipi/districte agafat a l'atzar on es van signar més de 5 contractes l'any 2018 sigui de la província de Lleida o Tarragona.

$$P(\text{Lleida o Tarragona})=(79+98)/572 = 177/572 = 0.309$$

2. Primer Codifiquem i mostrem els 10 primers registres per comprovar el resultat

2.1 Creem la variable NombreContractes_N2018

```
> summary(dades$N2018)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	2.00	10.00	174.10	60.25	11510.00

```
>NombreContractes_N2018<-cut(dades$N2018,breaks=c(0,2,10,60.25,11510),
labels=c("pocs","regular","bastants","molts"),include.lowest=TRUE)
```

```
> head(NombreContractes_N2018)
```

```
[1] pocs  molts  regular molts  regular bastants
Levels: pocs regular bastant molts
```

2.2 Taula de contingència:

```
> Taula_cont<-table(dades$PROV, NombreContractes_N2018)
> rownames(Taula_cont)<-c("Barcelona", "Girona", "Lleida", "Tarragona")
> Taula_cont
```

NombreContractes_N2018

	pocs	regular	bastant	molts
Barcelona	41	45	83	150
Girona	44	69	71	37
Lleida	110	69	34	18
Tarragona	59	50	42	33

2.3. La probabilitat que un municipi/districte tingui “molts” contractes signats. (10 punts)

$$P(\text{molts}) = 238/955 = 0.2492.$$

2.4 La probabilitat que un municipi/districte tingui “molts” contractes i sigui de la província de Girona.

$$P(\text{Girona} \cap \text{molts}) = 37/955 = 0.0387.$$

2.5 La probabilitat que la província sigui Lleida sabent que hi ha pocs contractes signats.

$$P(\text{Lleida} | \text{pocs}) = 110/254 = 0.4331.$$

2.6 La probabilitat que hi hagen bastant contractes signats sabent que la província és Tarragona.

$$P(\text{Bastant} | \text{Tarragona}) = 42/184 = 0.2283.$$

2.7 Els successos "Província de Girona" i "molts contractes", són independents? Per què?

No, ja que, $P(\text{Girona} \cap \text{molts}) = 37/956 = 0.0387$.

$$P(\text{Girona}) = 221/955 = 0.2314. \quad P(\text{molts}) = 238/955 = 0.2492.$$

$$P(\text{Girona} \cap \text{molts}) = 0.0387 \neq P(\text{Girona}) * P(\text{molts}) = 0.0577.$$