

PAC 5:

Probabilitat i Estadística

Curs 2020/21

El nom i cognoms de l'estudiant han d'indicar clarament a la primera pàgina de l'informe. L'informe final es lliurarà únicament en format PDF ia l'aula Moodle.

Considerarem el fitxer "PIS_MUN2.csv" En aquest fitxer hi trobareu 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. Concretament trobareu les variables.

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

PROV = Província la 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.

Suposem que les observacions corresponen a una mostra.

EXERCICI 1 (60%)

- Feu amb R el diagrama de dispersió del núvol de punts de la variable (X) valor del lloguer mitjà mensual dels contractes signats l'any 2015 als municipis/districtes i la variable (Y) valor del lloguer mitjà mensual dels contractes signats l'any 2018 als municipis/districtes. Què trobeu d'estrany en aquesta gràfica?
- Per a no tenir en compte aquestes observacions anòmales trobades anteriorment, crearem dues noves variables M_{2015_bis} i M_{2018_bis} en les quals només conservarem els registres que tenen mitjana de lloguers superiors a zero **tots dos anys**. Feu de nou el diagrama de dispersió i a més, ara afegiu la recta de regressió. Calculeu amb R la recta de regressió de la variable M_{2018_bis} en funció de la variable M_{2015_bis} . Heu de donar el pendent i l'ordenada a l'origen. Interpreteu el valor del pendent obtingut.
- Quin és el valor del coeficient de determinació? I el valor del coeficient de correlació? Que podeu dir sobre la bondat de l'ajust?
- Estimeu el lloguer mitjà en el 2018 en un municipi que en el 2015 va tenir un lloguer mitjà de 500 euros/mes.
- Volem fer un contrast sobre el pendent de la recta de regressió per saber si la variable M_{2015_nou} és explicativa. Indiqueu les hipòtesis nul·la i alternativa, el p-valor i la conclusió a que arribeu.

EXERCICI 2 (20%)

Considerem ara les variables M_2015_bis2, M_2016_bis2, M_2017_bis2 i M_2018_bis2 en els quals només conservem els registres que tenen totes les mitjanes de lloguers superiors a zero.

- a) Calculeu amb R el model de regressió múltiple per explicar la variable M_2018_bis2 en funció de totes les altres variables. Heu de donar explícitament el model.
- b) Quina variabilitat de la variable M_2018_bis2 queda explicada per aquest model.
- c) Indiqueu si algun dels coeficients del model de regressió múltiple no és significatiu.

EXERCICI 3 (20%)

Feu un ANOVA amb R per estudiar si hi ha diferències entre les mitjanes dels lloguers de l'any 2018, M_2018_bis2, de les 4 províncies (Barcelona, Lleida, Girona i Tarragona). A quina conclusió podem arribar?