Économétrie 1, M1 MAEF
22/06/2016, Durée : deux heures
Documents et téléphones portables ne sont pas autorisés.
L'usage de la calculatrice est autorisé.

Exercice 1. Régression linéaire

Description : Les données sont un échantillon aléatoire de dossiers de reventes de maisons à partir des fichiers utilisés par des agents immobiliers comme une base d'information.

Nombre de cas : 45
Les noms de variables :
PRICE (P) : prix de vente
SQFT (S) : feet carrés de surface habitable
AGE (A) : âge de la maison
FEATS (F) : nombre des 11 fonctions (lave-vaisselle, réfrigérateur, micro-ondes, broyeur, machine à laver, interphone, lucarne(s), compacteur, sèche-linge, matériel pour handicapé, accès à la télévision par câble)
TAX (T) : taxe annuelle

Vous êtes chargé(e) d'une étude visant à fournir un modèle de la taxe annuelle en fonction des variables sur la propriété de la maison (P, S, A et F).

Nous considérons le modèle à quatre variables explicatives. Les résultats sont présentés dans le tableau suivant.

```
Call:
lm(formula = TAX ~ PRICE + SQFT + AGE + FEATS, data = home)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  94.56842  106.99075   0.884   0.3820
PRICE         0.36113    0.04674   7.726 1.87e-09
SQFT          0.34683    0.03522   9.847 3.01e-12
AGE          -3.51796    1.37729  -2.554   0.0146
FEATS        20.05833   10.13692   1.979   0.0548

Residual standard error: 138 on 40 degrees of freedom
Multiple R-squared:  0.8233,Adjusted R-squared:  0.8056
F-statistic: 46.58 on 4 and 40 DF,  p-value: 1.546e-14
```

1. Réécrivez le modèle en utilisant les résultats de régression.
2. Le modèle est il globalement significatif ?
3. Les coefficient ont-ils un signe conforme à vos attentes?
4. Testez la position du coefficient de PRICE par rapport à 0.3 avec le niveau de test 5%.
5. Y a t il une variable qui doit être éliminée ? Laquelle ?
6. Présentez un algorithme qui permet d'obtenir un modèle ne contenant que les variables significatives.

Exercice 2. Valeurs extrêmes

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. issu de la loi $Unif[0,1]$. C'est à dire la densité commune de X_i est $f(x) = \mathbb{1}_{[0,1]}(x)$ et la fonction de répartition commune de X_i est $F(x) = x$, pour $x \in [0,1]$. Étant donnés $X_1, \ldots, X_n$, on définit deux nouvelles variables aléatoires

$$U = \min(X_1, \ldots, X_n) \quad \text{et} \quad V = \max(X_1, \ldots, X_n),$$

qui sont respectivement le minimum et le maximum de l'échantillon i.i.d..

1. En notant $G(u)$ et $H(v)$ les fonctions de répartition des variables aléatoires U et V, montrer que

$$G(u) = 1 - (1-u)^n \quad \text{et} \quad H(v) = v^n.$$

2. En déduire les densités de U et V.
3. Montrer que pour $u \leq v$, $\mathbb{P}(U \leq u, V \leq v) = \mathbb{P}(V \leq v) - \mathbb{P}(U > u, V \leq v)$.
4. Montrer que pour $u \leq v$, $\mathbb{P}(U > u, V \leq v) = (v-u)^n$.
5. En déduire la fonction de répartition du couple de variables aléatoires (U, V) est la suivante

$$F_{(U,V)}(u,v) = v^n - (v-u)^n \mathbb{1}_{\{0 \leq u \leq v \leq 1\}}(u,v).$$

6. En déduire la densité jointe du couple de variables aléatoires (U, V).

Questions de cours

1. En utilisant la méthode de la transformée inverse, écrivez le code R qui permet de générer 100 nombres aléatoires de loi Fréchet de paramètre 1. On rappelle que la variable X a la loi Fréchet de paramètre $\alpha > 0$, si pour $x > 0$,

$$\mathbb{P}(X \leq x) = \exp(-x^{-\alpha}).$$

2. Que fait le programme suivant ? Pouvez vous prédire la valeur de `t` ?

```
x=runif(5000)
m=rep(0,100)
for(i in 1:100)m[i]=mean(x[((i-1)*50+1):(i*50)])
t=sum(abs((m-1/2)/sqrt(1/12/50))>qnorm(0.975))
```