

Licence APE-EGE

16 Décembre 2005

Durée : 2h

Polycopiés de cours et Calculatrice autorisée

Numéro Etudiant :

Examen d'Econométrie

Exercice 1 On envisage un modèle linéaire expliquant la variable Chom en fonction de la seule variable Infl. On tente une modélisation linéaire du type :

$$(Chom)_i = \beta_0 + \beta_1(Infl)_i + \varepsilon_i, \quad i = 1, \dots, 41$$

Question : Pourquoi les paramètres β_0 et β_1 ne sont pas calculables ?

Réponse _____

Fin

Question : On s'intéresse tout naturellement à l'estimation des paramètres β_0 et β_1 . Déterminez les estimations obtenues par la méthode des moindres carrés.

On rappelle à titre indicatif que

- $var(\mathbf{x}) = \overline{x^2} - \bar{x}^2$
- $cov(\mathbf{x}, \mathbf{y}) = \overline{x \times y} - \bar{x} \times \bar{y}$.

```
1 | > mean(Infl)
2 | [1] 5.385366
3 | > mean(Infl^2)
4 | [1] 44.03463
5 | > mean(Chom)
6 | [1] 6.487805
7 | > mean(Chom^2)
8 | [1] 57.51659
9 | > mean(Infl*Chom)
10| [1] 31.02854
```

Réponse _____

_____ Fin

Question : *Déterminez le coefficient de corrélation linéaire entre $Infl$ et $Chom$ et donnez-en une interprétation.*

Réponse _____

_____ Fin

Question : *Retrouvez les résultats des deux questions précédentes. Analysez brièvement les résultats obtenus.*

```
1 | > summary(lm(Chom~Infl))
2 |
3 | Call:
4 | lm(formula = Chom ~ Infl)
5 |
```

```

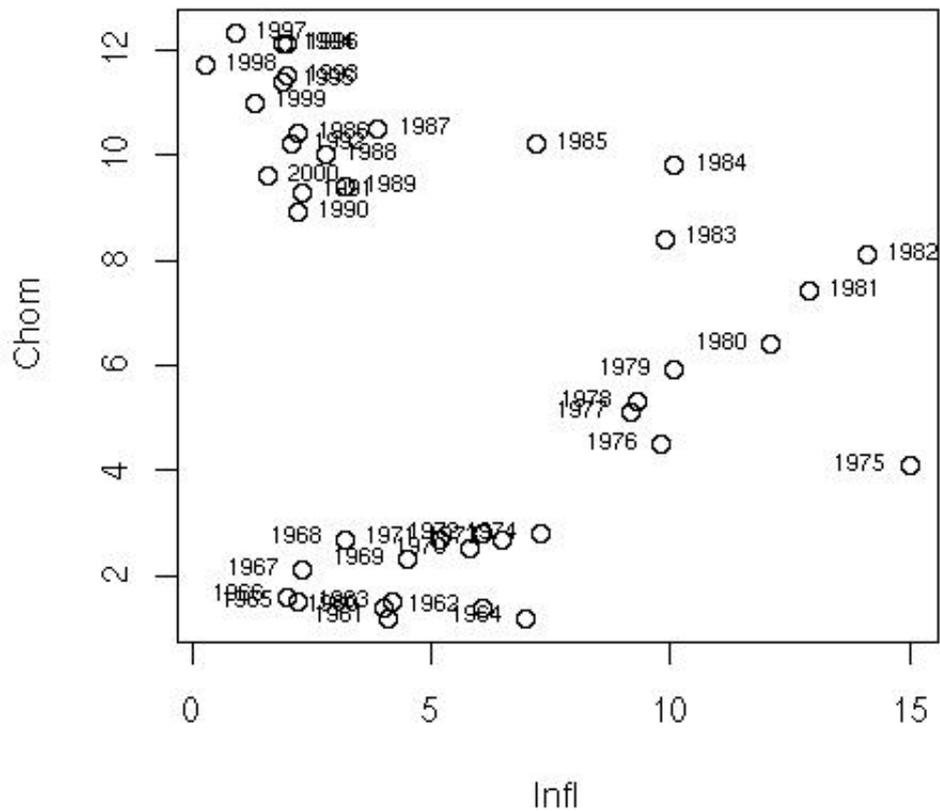
6 Residuals:
7   Min      1Q  Median      3Q      Max
8 -5.816 -3.880  1.584  3.449  4.731
9
10 Coefficients:
11             Estimate Std. Error t value Pr(>|t|)
12 (Intercept)   7.8888     1.0403   7.583 3.43e-09 ***
13 Infl          -0.2601     0.1568  -1.659   0.105
14 ---
15 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
16
17 Residual standard error: 3.892 on 39 degrees of freedom
18 Multiple R-Squared:  0.06595,    Adjusted R-squared:  0.042
19 F-statistic: 2.754 on 1 and 39 DF,  p-value: 0.1050
20
21 > sqrt(0.066)
22 [1] 0.2569047

```

Réponse _____

Fin

Question : *Sur le graphique ci-dessous reportez la droite ajustée (même approximativement) et illustrez la notion de valeur ajustée et de résidu.*



Question : *Peut-on penser au vu des données que la variable $Infl$ apporte de l'information pour expliquer la variable $Chom$ (indication : fournir la p -valeur associée puis conclure.)*

Réponse _____

Fin

Question : *Quelle variable nommée X dans la sortie suivante a été introduite ? Comparez les résultats obtenus avec ceux du précédent modèle. Représentez sur le graphique précédent ce nouveau modèle. Peut-on alors penser que $Infl$ a un pouvoir explicatif sur $Chom$?*

```
1 |> summary(lm(Chom~Infl*X))
```

```

2 |
3 | Call:
4 | lm(formula = Chom ~ Infl * X)
5 |
6 | Residuals:
7 |      Min       1Q   Median       3Q      Max
8 | -1.99808 -0.59426  0.09453  0.71012  1.61847
9 |
10 | Coefficients:
11 |             Estimate Std. Error t value Pr(>|t|)
12 | (Intercept)  11.24871    0.32782  34.314 < 2e-16 ***
13 | Infl         -0.24380    0.05581  -4.369 9.72e-05 ***
14 | X            -10.71149    0.58361 -18.354 < 2e-16 ***
15 | Infl:X        0.61452    0.08663   7.094 2.13e-08 ***
16 | ---
17 | Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18 |
19 | Residual standard error: 1.015 on 37 degrees of freedom
20 | Multiple R-Squared:  0.9398,    Adjusted R-squared:  0.9349
21 | F-statistic: 192.4 on 3 and 37 DF,  p-value: < 2.2e-16
22 |

```

Réponse

Fin

Exercice 2

L'étude portera sur l'analyse du prix d'une maison en fonction de ses différentes caractéristiques, et sera basée sur le jeu de données `maison` de taille $n = 150$ présenté à la fin du devoir. Voici très brièvement la description des variables associées :

- `PRIX` : prix des maisons.
- `SURFACE` : surface des maisons.
- `HECTARES` : surfaces des terrains associés.
- `PIECES` : nombre de pièces des maisons.
- `BAINS` : nombre de salle de bains.

On envisage un modèle linéaire multiple expliquant la variable `PRIX` en fonction de tous les régresseurs du jeu de données.

Question : *A la vue de la matrice de corrélation ci-après, quels sont les régresseurs qui vous semblent être les plus explicatifs ?*

```
1 > cor(log(maison))
2           PRIX  SURFACE  HECTARES  PIECES  BAINS
3 PRIX      1.0000000 0.7746669 0.5374595 0.6612338 0.7089474
4 SURFACE  0.7746669 1.0000000 0.3118338 0.8542915 0.7856913
5 HECTARES 0.5374595 0.3118338 1.0000000 0.1961679 0.3525309
6 PIECES   0.6612338 0.8542915 0.1961679 1.0000000 0.6502278
7 BAINS    0.7089474 0.7856913 0.3525309 0.6502278 1.0000000
```

Réponse _____

Fin

Question : *Interprétez la sortie ci-dessous, en particulier les p -valeurs des tests de significativité locales, le R^2 .*

```
1 > summary(lm(log(PRIX)~log(SURFACE)+log(HECTARES)+log(PIECES)+log(BAINS)))
2
3 Call:
4 lm(formula = log(PRIX) ~ log(SURFACE) + log(HECTARES) + log(PIECES) +
5     log(BAINS))
6
7 Residuals:
8     Min       1Q   Median       3Q      Max
9 -0.428593 -0.089897  0.005171  0.082209  0.366439
```

```

10
11 Coefficients:
12           Estimate Std. Error t value Pr(>|t|)
13 (Intercept)  9.00815    0.47661  18.901 < 2e-16 ***
14 log(SURFACE)  0.35710    0.08354   4.275 3.45e-05 ***
15 log(HECTARES) 0.09231    0.01419   6.504 1.19e-09 ***
16 log(PIECES)   0.11086    0.10312   1.075  0.2842
17 log(BAINS)    0.11738    0.04817   2.437  0.0160 *
18 ---
19 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
20
21 Residual standard error: 0.1419 on 145 degrees of freedom
22 Multiple R-Squared:  0.7108,    Adjusted R-squared:  0.7029
23 F-statistic: 89.11 on 4 and 145 DF,  p-value: < 2.2e-16
24

```

Réponse _____

_____ Fin

Question : *Rappelez les effets indésirables sur les tests de significativité locale s'il y a colinéarité entre les régresseurs.*

Réponse _____

_____ Fin

Question : *Rappelez la définition du VIF, et son interprétation générale. Ensuite, interprétez-les quant au jeu de données étudié.*

```

1 > vif(lm(log(PRIX)~log(SURFACE)+log(HECTARES)+log(PIECES)+log(BAINS)))
2   log(SURFACE) log(HECTARES)   log(PIECES)   log(BAINS)
3         5.715257         1.166638         3.783690         2.707968

```

Réponse _____

Fin

Question : (*Relation avec la matrice de corrélation*) Justifier l'ordre de grandeur des VIFs des covariables $\log(\text{SURFACE})$ et $\log(\text{PIECES})$ en utilisant l'instruction suivante.

```
1 > 1/(1-(0.854291515)^2)
2 [1] 3.701154
```

Réponse _____

Fin

Question : Quelle est la stratégie à adopter pour soigner la colinéarité ? Après une seule étape voici les résultats du `summary(lm(...))` présentés sous une forme spécialement adaptée aux notations du cours. Complétez les 3 **cases manquantes** en vous aidant des indications (à la suite du tableau). Justifiez qu'il n'est pas nécessaire d'effectuer d'étape supplémentaire. Précisez l'équation du modèle sélectionné et analysez brièvement les sorties.

$\mathbf{x}^{(j)}$	$\widehat{\beta}_j(\mathbf{y} \mathbf{x})$	$\widehat{\sigma}_{\widehat{\beta}_j}(\mathbf{y} \mathbf{x})$	$\widehat{\delta}_{\beta_j,0}(\mathbf{y} \mathbf{x})$	p -valeur bilatérale
1	8.73297	0.40225	21.71048	$\simeq 0$
$\log(\text{SURFACE})$	0.42316			$\simeq 0$
$\log(\text{HECTARES})$	0.09028	0.01408	6.41437	$\simeq 0$
	0.11529	0.04816	2.39405	0.01793

Indications : $R^2 = 70.85\%$, $\widehat{\sigma}_\varepsilon(\mathbf{y}|\mathbf{x}) \simeq 0.142$ et $VIF_{\log(\text{SURFACE})} \simeq 2.6226$.

Réponse _____

Fin

Question : *Etant donnée la modélisation adoptée, complétez la phrase ci-dessous : lorsque SURFACE _____ de 10%, on peut s'attendre à ce que PRIX _____ de ____.*

Question : *A partir de cette question, nous ne considérerons que le modèle final. Peut-on montrer au vu des données que le paramètre $\beta_1 < 0.5$ au seuil de 5% ?*

1 | $> (0.42316 - (0.5)) / 0.05662$
2 | [1] -1.357118

Réponse _____

Fin

Question : *A partir de l'instruction R suivante, que peut-on avancer au vu des données comme assertion(s) d'intérêt au seuil 5% ?*

```
1 | > pnorm((0.09028-(0.05))/0.01408)
2 | [1] 0.9978871
```

Réponse _____

Fin

Question : *Déduisez de l'instruction ci-dessous un intervalle de confiance à 95% pour le paramètre β_2 et interprétez-le (via l'approche expérimentale). Quelle relation y-a-t-il entre cet intervalle et le test de significativité locale du paramètre β_2 ?*

```
1 | > 0.09028+c(-1,1)*qnorm(0.975)*0.01408
2 | [1] 0.06268371 0.11787629
```

Réponse

Fin

Jeu de données :

```
1 > maison
2     PRIX SURFACE HECTARES PIECES BAINS
3 1  179000   3060   0.7500     8  2.0
4 2  126500   1600   0.2600     8  1.5
5 3  134500   2000   0.7000     8  1.0
6 4  125000   1300   0.6500     5  1.0
7 5  142000   2000   0.7500     9  1.5
8 6  164000   1956   0.5000     8  2.5
9 7  146000   2400   0.4000     7  2.5
10 8  129000   1200   0.3300     6  1.0
11 9  141900   1632   3.0000     6  3.0
12 ...
13 141 121900   1300   0.7800     6  1.0
14 142 126000   1232   0.3140     6  2.0
15 143 164900   1980   0.7000     8  2.5
16 144 172000   2100   1.0000     8  2.5
17 145 100000   1338   0.1200     6  1.0
18 146 129900   1070   1.6900     5  1.0
19 147 110000   1289   0.2500     6  1.0
20 148 131000   1066   0.3300     5  1.0
21 149 107000   1100   0.1700     5  1.0
22 150 165900   1840   1.1620     8  2.0
```