UNIVERSITE PARIS 1 – PANTHEON SORBONNE

L3 ECONOMIE

Année 2011-2012

INTRODUCTION A L'ECONOMETRIE

C. DOZ

SUJETS DE TRAVAUX DIRIGES

Fiche de TD n°1

Révisions de Statistique (tests)

EXERCICE 1

Une entreprise fabrique des billes pour roulement à billes, dont le diamètre doit être précisément égal à 5 mm. En pratique, on considère que le diamètre des billes fabriquées est une variable aléatoire X qui suit une loi $\mathcal{N}(m, \sigma^2)$. Le responsable du contrôle qualité prélève au hasard un échantillon de $N = 100$ billes dont il mesure le diamètre en mm avec un instrument de précision. On notera X_n le diamètre de la n^{eme} bille mesurée.

1. Quel estimateur $\hat{m}$ prend-on pour m et quelle est la loi de cet estimateur ?
2. Quel estimateur $\hat{\sigma}^2$ prend-on pour σ^2 et quelle est la loi de cet estimateur ?
3. i) Construire le test de l'hypothèse $H_0 : m = 5$ contre $H_1 : m \neq 5$ au niveau 5%.
   ii) Quelle conclusion tire-t-on si les résultats mesures donnent :

$$\overline{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n = 5,3 \quad \text{et} \quad \frac{1}{N}\sum_{n=1}^{N}(X_n - \overline{X}_N)^2 = 1,3$$

EXERCICE 2

On observe un échantillon de taille n d'une variable aléatoire Y suivant une loi normale d'espérance m et dont la variance est connue : $\sigma^2 = 1$.

1. Quel estimateur proposez-vous pour le paramètre m et pourquoi ? Quelle est la loi de cet estimateur ?
2. On suppose *a priori* dans cette question que $m \leq 1$.
   i) Construire le test au seuil α de l'hypothèse nulle $H_0 : m = 1$ contre l'hypothèse $H_1 : m < 1$.
   ii) Quelle conclusion tirez-vous si $\alpha = 0,05$, $n = 75$, et $\bar{y}_n = 0,85$?
   iii) Même question pour $\alpha = 0,05$, $n = 300$, et $\bar{y}_n = 0,85$.
3. On ne fait plus d'hypothèse *a priori* sur m et on s'intéresse cette fois au test de l'hypothèse $\tilde{H}_0 : m \geq 1$ contre l'hypothèse H_1.
   La région critique a la même forme qu'à la question précédente.
   i) Calculez, en fonction de la vraie valeur inconnue de m, la probabilité $p(m)$ de rejeter $\tilde{H}_0$
   ii) Montrez que $\text{Max}\{p(m), m \geq 1\} = p(1)$.
   iii) En déduire que le test a même région critique que celui de la question précédente.
4. pour $\alpha = 0,05$, calculez la puissance du test pour les valeurs suivantes de m :

$$m = 0,4\ ;\ 0,5\ ;\ 0,7\ ;\ 0,8\ ;\ 0,85\ ;\ 0,9$$

   lorsque $n = 75$, puis lorsque $n = 300$. Commentez.

EXERCICE 3 (facultatif : à ne traiter que si le temps imparti le permet)

On suppose que les notes d'économétrie obtenues à l'examen par les étudiants de L3 sont des variables aléatoires indépendantes qui suivent une loi normale d'espérance m et de variance σ^2.

Le professeur corrige d'abord les copies du groupe de TD n° 1, dont l'effectif est de 35 étudiants, dans le but de voir si son barème est convenable.

Dans la suite, on désigne par n l'effectif de ce groupe de TD ($n = 35$), et par $Y_i, i = 1, \dots, n$ les notes obtenues.

1. Le professeur estime m par $\hat{m} = \overline{Y}$ et σ^2 par $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$.
   i) Quelle propriété particulière ces estimateurs possèdent-ils ?
   ii) Quelle est la loi de $\hat{m}$?
   iii) Quelle est la loi de $\hat{\sigma}^2$?
2. Le professeur s'intéresse au test de l'hypothèse $H_0 : m \leq 10$ contre l'hypothèse $H_1 : m > 10$.
   i) Quelle est l'intention de ce professeur en faisant ce choix pour H_0 et H_1 ?
   ii) Comment construit-on le test au niveau $\alpha = 0,05$?
   iii) Les valeurs obtenues sont $\overline{Y} = 10,53$ et $\hat{\sigma}^2 = 3,24$. Quelle est la conclusion du test ? Que va faire le professeur ?

Fiche de TD n°2
Modèle linéaire et introduction des MCO

EXERCICE 1 :

Dans les relations suivantes, y, x , K et L sont des variables pour lesquelles on dispose d'observations $y_1, \dots y_N$, $x_1, \dots x_N$, etc.. et a, b, α, β sont des paramètres dont les vraies valeurs sont inconnues.
Parmi ces relations, quelles sont celles qui peuvent conduire à un modèle linéaire, éventuellement après un changement de variable ou un changement de paramètres ?
Vous écrirez alors le modèle linéaire concerné, et vous préciserez quels sont les paramètres de ce modèle.

1. $y = a + bx^2$
2. $y = a + \frac{b}{x}$
3. $y = a + \frac{x}{b}$
4. $y = a + \frac{1}{b+x}$
5. $y = AK^{\alpha}L^{\beta}$
6. $y = AK^{\alpha}L^{1-\alpha}$
7. $y = \frac{a}{1+x^b}$

EXERCICE 2 :

On considère un échantillon $y_1, \dots y_N$ de variables aléatoires i.i.d. d'espérance m.
On associe à cet échantillon le modèle linéaire $y_n = m + \varepsilon_n$ dans lequel la constante est la seule variable explicative, m est le paramètre inconnu à estimer, et ε_n est le résidu.
Résoudre le problème de m.c.o. associé à ce modèle, c'est-à-dire chercher $\hat{m}$ réalisant :

$$\underset{m \in \mathbb{R}}{\text{Min}} \sum_{n=1}^{N} (y_n - m)^2$$

On vérifiera que l'extremum trouvé est bien un minimum.

EXERCICE 3 :

N.B. Cet exercice a un but purement pédagogique : ne jamais estimer de modèle avec 10 observations !

On dispose des observations suivantes de deux variables y et x, et on veut estimer par m.c.o. le modèle $y_n = a + bx_n + \varepsilon_n$.

x	3	1,5	1	3	1	2,5	0,5	-1	0,5	2
y	-4	-1	-2	-5,5	-1	-5	0,5	2	3	-4

1. Représenter le nuage de points.
2. Calculer $\hat{a}$ et $\hat{b}$.
3. Représenter la droite de régression sur le même graphique que le nuage de points.
4. Calculer les valeurs ajustées $\hat{y}_n$, les résidus estimés $\hat{\varepsilon}_n$, ainsi que le R^2 de cette régression. Commenter.

Fiche de TD n°3

Régression simple : propriétés des MCO

EXERCICE 1 :

On considère les ventes trimestrielles de micro-ordinateurs sur 5 ans. On note $q_t, t = 1, \ldots, 20$ les ventes en volume, au cours de ces 20 trimestres et $p_t, t = 1, \ldots, 20$ les prix relatifs des micro-ordinateurs pendant la même période (le prix relatif est le rapport entre l'indice des prix des micro-ordinateurs et l'indice des prix à la consommation).

1. On étudie le modèle (M1) : $\ln q_t = a + b \ln p_t + \varepsilon_t$
   i) Comment s'interprètent économiquement les coefficient a et b dans ce modèle ?
   ii) Les observations conduisent aux résultats suivants :

   $$\sum_{t=1}^{20} \ln p_t = -5,15, \quad \sum_{t=1}^{20} \ln q_t = 30,89$$

   $$\sum_{t=1}^{20} (\ln p_t)^2 = 1,91, \quad \sum_{t=1}^{20} (\ln q_t)^2 = 72,53, \quad \sum_{t=1}^{20} (\ln p_t)(\ln q_t) = -10,41$$

   Calculer les estimations obtenues pour a et b en appliquant la méthode des m.c.o.
   iii) Calculer le coefficient de détermination du modèle.
2. On considère maintenant le modèle (M2) : $\ln(p_t q_t) = c + d \ln p_t + u_t$
   i) Que peut-on dire de c, de d, des u_t ?
   ii) Calculer les estimateurs de c et d obtenues en appliquant la méthode des m.c.o. ainsi que les estimations associées.
   Vérifier que : $\hat{c} = \hat{a}$ et que $\hat{d} = \hat{b} + 1$.
   iii) Que peut-on dire des résidus estimés dans ce modèle ?
   iv) Que peut-on dire du coefficient de détermination dans ce modèle ?
3. On s'intéresse maintenant au modèle (M3) : $q_t = \alpha + \beta p_t + v_t$.
   Les observations conduisent aux résultats suivants :

   $$\sum_{t=1}^{20} p_t = 15,68, \quad \sum_{t=1}^{20} q_t = 135,58$$

   $$\sum_{t=1}^{20} p_t^2 = 12,64, \quad \sum_{t=1}^{20} q_t^2 = 956,18, \quad \sum_{t=1}^{20} p_t q_t = 102,86$$

   i) Calculer les estimations de α et β obtenues par la méthode des m.c.o.
   ii) Comparer avec celles qui avaient été obtenues pour a et b. Commenter.
   iii) Peut-on dire quelque chose au sujet des résidus estimés et du coefficient de détermination de ce modèle si on les compare à ceux de la question 1.

EXERCICE 2 :

On considère le modèle : $y_n = a + bx_n + \varepsilon_n$ pour lequel on dispose de N observations.

i) On note $\hat{a}$ et $\hat{b}$ les estimateurs des m.c.o. de a et b. Rappeler les formules de $\hat{a}$ et $\hat{b}$.

ii) On note $\overline{x}$ et $\overline{y}$ les moyennes empiriques de x et y dans l'échantillon, et on note $\widetilde{y}_n = y_n - \overline{y}$ et $\widetilde{x}_n = x_n - \overline{x}$.
On considère le modèle sans terme constant : $\widetilde{y}_n = \beta \widetilde{x}_n + u_n$ et note $\hat{\beta}$ l'estimateur des m.c.o. de β dans ce modèle.
Rappeler la formule de calcul de $\hat{\beta}$ et vérifier que $\hat{\beta} = \hat{b}$.

iii) Même question si l'on considère le modèle $y_n = \gamma \widetilde{x}_n + v_n$.

iv) Quelle conclusion pratique tirez-vous de ces résultats ?

Fiche de TD n°4

Régression simple : propriétés des MCO, intervalles de confiance dans le modèle gaussien

EXERCICE 1 :

On considère le modèle : $y_n = a + bx_n + \varepsilon_n$ pour lequel on dispose de N observations.

On note : R^2 le coefficient de détermination obtenu lorsqu'on estime ce modèle par m.c.o. et r_{xy} le coefficient de corrélation empirique entre x et y.

i) Montrer que $r_{xy}^2 = R^2$.

ii) Comment s'interprète une valeur élevée obtenue pour R^2 ?

EXERCICE 2 :

On estime par m.c.o. le modèle $y_n = a + bx_n + \varepsilon_n$.

1. On suppose que les résidus ε_n vérifient les hypothèses de base : rappeler ces hypothèses.
2. Rappeler l'expression de $V\hat{a}$ et $V\hat{b}$.
3. Que peut-on dire de $\hat{a}$ et de $\hat{b}$ si la variance empirique des x_n est très grande ?
4. Que peut-on dire de $\hat{a}$ et de $\hat{b}$ si la variance empirique des x_n est très petite ?
5. Pourquoi dit-on que, pour estimer un modèle, il faut avoir "suffisamment de variabilité dans les données" ?
6. Que se passerait-il si les x_n avaient tous la même valeur ? Cela vous parait-il logique ?

EXERCICE 3 :

On considère un modèle de la forme

$$y_n = a + bx_n + \varepsilon_n$$

dans lequel on suppose que les résidus sont indépendants et suivent une loi $\mathcal{N}(0, \sigma^2)$.

Ce modèle est estimé par la méthode des m.c.o. sur un échantillon de $N = 24$ observations.

Dans cet échantillon, les observations conduisent aux valeurs suivantes pour les moyennes empiriques, écart-types empiriques et covariance empirique :

$$\overline{y} = -0.9676, \qquad \overline{x} = 0.1126$$

$$s_y = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})^2} = 0.10413, \quad s_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2} = 0.2296 \quad \text{Cov}_{emp}(x, y) = -0.01846$$

1. Calculez $\hat{b}$ et $\hat{a}$.
2. Calculez le coefficient de détermination du modèle. Commentez.
3. Calculez $\hat{\sigma}^2$
4. i) Construisez un intervalle de confiance à 95% pour le paramètre b.
   ii) Construisez un intervalle de confiance à 99% pour le paramètre b.

Fiche de TD n°5

Régression simple : intervalles de confiance, tests, EMV et prévision dans le modèle gaussien

EXERCICE 1 :

On reprend les données de l'exercice 1 de la fiche 3, et on suppose de plus que les ε_n sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

1. Donnez les intervalles de confiance à 95% pour les paramètres a, b, c, d. Commentez.
2. i) Effectuez au seuil 5% les tests de significativité de ces paramètres.
   ii) Quelles sont les "*p-values*" associées aux valeurs obtenues pour les statistiques de test ?
   iii) A quel test sur le paramètre b le test de significativité de d est-il équivalent ? Quelle en est l'interprétation économique ?
   iv) Comment mèneriez-vous ce test directement en utilisant l'estimateur de b obtenu dans le modèle (M1) ?

EXERCICE 2 :

On reprend les données de l'exercice 3 de la fiche n°4.

Construire un intervalle de prévision pour y lorsque $x = 0,12$:
- au niveau de confiance 95%
- au niveau de confiance 99%

Comparer les deux intervalles de prévisions obtenus et commenter.

EXERCICE 3 (facultatif) :

On considère le modèle $y_n = a + bx_n + \varepsilon_n$, et on suppose que les ε_n sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

1. Quelle est la loi de y_n pour n donné ?
2. Vérifier que les y_n sont des v.a.r. indépendantes mais pas i.i.d.
3. On note $\theta = (a, b, \sigma^2)'$. Ecrire la vraisemblance du modèle : $l(y_1, \dots, y_n, \theta)$
4. Calculer l'estimateur du maximum de vraisemblance de θ. Commenter.

Fiche de TD n°6

Régression simple : révision et introduction des notations matricielles

EXERCICE 1 : exercice de révision, inspiré d'un exercice de l'examen 2009-2010

On considère le modèle

$$y_n = a + bx_n + \varepsilon_n$$

dans lequel on suppose que les résidus sont indépendants et suivent une loi $\mathcal{N}(0, \sigma^2)$.

Ce modèle est estimé par la méthode des m.c.o. sur un échantillon de $N = 30$ observations.

Dans cet échantillon, les observations conduisent aux valeurs suivantes pour les moyennes empiriques, écart-types empiriques et covariance empirique :

$$\overline{y} = 3,09 \qquad \overline{x} = 7,01 \qquad \mathrm{Cov}_{emp}(x,y) = \frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})(x_n - \overline{x}) = 79,45$$

$$s_y = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})^2} = 18,29 \quad s_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2} = 4,47$$

1. Calculez $\hat{b}$ et $\hat{a}$.
2. Calculez $\hat{\sigma}^2$.
3. Menez, au seuil de 5 %, les tests de significativité des paramètres du modèle.
4. Menez, au seuil de 5 %, le test de l'hypothèse $H_0 : b = 3$ contre l'hypothèse $H_1 : b \neq 3$.
5. Menez, au seuil de 5 %, le test de l'hypothèse $H_0 : b \leq 3$ contre l'hypothèse $H_1 : b > 3$.
6. Construisez un intervalle de prévision au niveau de confiance 95% pour y_{N+1} lorsque $x_{N+1} = 9,5$.

EXERCICE 2 :

On reprend les données de l'exercice 3 de la fiche n°2 :

x	3	1,5	1	3	1	2,5	0,5	-1	0,5	2
y	-4	-1	-2	-5,5	-1	-5	0,5	2	3	-4

1. Ecrire le modèle sous sa forme matricielle, qu'on notera $y = X\beta + \varepsilon$ (vous préciserez ce que désignent y, X, β et ε ici).
2. Calculer $X'y$, $X'X$, $(X'X)^{-1}$, puis calculer $\hat{\beta}$.
   Vérifier qu'on retrouve les résultats obtenus précédemment.
3. Calculer $\hat{y}$ et $\hat{\varepsilon}$. Vérifier qu'on retrouve les résultats obtenus précédemment.

Fiche de TD n°7

Introduction de la régression multiple.
Propriétés des vecteurs aléatoires

EXERCICE 1 :

1. On considère une équation de salaire de la forme :
$$\ln w_n = b_0 + b_1 indF_n + b_2 etud_n + \varepsilon_n \qquad \text{(M1)}$$
dans laquelle :
– $indF_n$ est une variable prenant la valeur 1 si l'individu n est une femme et 0 sinon
– $etud_n$ est le nombre d'années d'études effectuées par l'individu n.
Comment doivent s'interpréter les différents coefficients du modèle ?
2. On définit la variable $indH_n$ prenant la valeur 1 si l'individu n est un homme et 0 sinon. Peut-on estimer le modèle (M2) défini par : $\ln w_n = a_0+a_1 indF_n+a_2 indH_n+a_3 etud_n+\varepsilon_n$? Pourquoi ?
3. On considère maintenant le modèle (M3) suivant :
$$\ln w_n = c_1 indF_n + c_2 indH_n + c_3 etud_n + \varepsilon_n$$
Peut-on estimer ce modèle ? Pourquoi ? Précisez comment s'interprètent les coefficients de ce nouveau modèle.
4. Quelles sont les relations entre les paramètres du modèle (M1) et ceux du modèle (M3) ?
5. Montrez que les estimateurs des m.c.o. des paramètres du modèle (M1) et du modèle (M3) sont reliés par les mêmes relations que celles qui lient les paramètres de ces modèles (on utilisera la définition des estimateurs des m.c.o.).
6. On considère maintenant le modèle (M4) suivant :
$$\ln w_n = d_0 + d_1 indF_n + d_2 etud_n + d_3 indF_n * etud_n + \varepsilon_n$$
(on a désigné par $*$ le produit des 2 variables concernées).
Comment s'interprète le paramètre d_3 ? En vous inspirant de la question 3, proposez une formulation équivalente de ce modèle.

EXERCICE 2 :

On considère un vecteur aléatoire $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ vérifiant
$$\mathbb{E}Y = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \text{ et } \mathbb{V}Y = \begin{pmatrix} 1 & 1/2 & -1/2 \\ 1/2 & 2 & 1 \\ -1/2 & 1 & 2 \end{pmatrix}.$$
On définit $Z = \begin{pmatrix} 2Y_1 + Y_2 \\ Y_1 - Y_3 \end{pmatrix}$.

Calculez $\mathbb{E}Z$ et $\mathbb{V}Z$:
i) en calculant ces matrices terme à terme
ii) en utilisant les formules matricielles vues en cours.

EXERCICE 3 :

On considère les deux modèles suivants :

$$\begin{aligned} y_n &= b_0 + b_1 x_{1n} + b_2 x_{2n} + u_n & (M1) \\ y_n &= b_0 + b_1 x_{1n} + b_2 x_{2n} + b_3 x_{3n} + v_n & (M2) \end{aligned}$$

Ces deux modèles sont estimés par m.c.o. sur N observations et on note :

- pour le modèle $(M1)$:
  - $\hat{u}$ le vecteur des résidus estimés
  - $SSR_1 = \sum_{n=1}^{N} \hat{u}_n^2$ la somme des carrés des résidus estimés
  - R_1^2 le coefficient de détermination associé
- pour le modèle $(M2)$:
  - $\hat{v}$ le vecteur des résidus estimés
  - $SSR_2 = \sum_{n=1}^{N} \hat{v}_n^2$ la somme des carrés des résidus estimés
  - R_2^2 le coefficient de détermination associé

1. Ecrire, pour $i = 1, 2$ la relation entre R_i^2 et SSR_i.
2. Ecrire les problèmes de m.c.o. associés aux modèles $(M1)$ et $(M2)$.
3. En déduire que $SSR_1 \geq SSR_2$, puis que $R_1^2 \leq R_2^2$. Commenter ce résultat et l'énoncer dans le cas où le nombre de variables explicatives est quelconque.
4. Etablir une formule générale permettant de calculer le R^2 ajusté d'une régression (c-à-d la quantité qui a été notée $\overline{R}^2$ dans le cours) en fonction du R^2 de la même régression.
5. On note $\overline{R}_1^2$ et $\overline{R}_2^2$ les R^2 ajustés des deux régressions associées à $(M1)$ et $(M2)$.
   Que peut-on dire de $\overline{R}_1^2$ et $\overline{R}_2^2$? Commenter.

Fiche de TD n°8

Propriétés de la régression multiple.
Lois des estimateurs des MCO

EXERCICE 1 :

Dans une étude économétrique aux Etats-Unis sur un échantillon de 4 000 salariés, les auteurs ont modélisé le salaire horaire moyen de chaque individu pendant l'année 1998 en fonction de son niveau de diplôme, de son sexe et de son âge.

Les variables explicatives retenues sont les suivantes :

- *Educ* est une variable indicatrice qui prend la valeur 1 si l'individu a fait des études supérieures et 0 sinon
- *Female* est une variable indicatrice qui prend la valeur 1 si l'individu est une femme et 0 sinon
- *Age* est l'âge de l'individu

Les résultats obtenus sont les suivants :

Regressor	Model (1)	Model (2)
Educ	5.46	5.48
Female	-2.64	-2.62
Age		0.29
Intercept	12.69	4.40
SER	6.27	6.22
R^2	0.176	0.190
$\overline{R}^2$		

Remarque : SER signifie "standard error" et désigne ce qui a été noté $\hat{\sigma}$ dans le cours.

1. Ecrire explicitement les deux modèles $(M1)$ et $(M2)$ qui sont estimés ici.
2. Expliquer soigneusement comment on doit interpréter les coefficients qui apparaissent dans chacun de ces modèles.
3. Commenter les signes des coefficients estimés, et comparer les valeurs obtenues dans les deux modèles.
4. Commenter les valeurs des R^2 obtenues dans les deux modèles en liaison avec ce qui a été vu dans l'exercice 2 de la fiche 8.
5. Quel test faudrait-il faire pour tester une hypothèse de différence salariale selon le sexe dans le cadre des modèles étudiés ? Le tableau ci-dessus vous permet-il de faire ce test ?
6. Calculer les $\overline{R}^2$ pour les deux régressions. Commenter.

EXERCICE 2 :

On considère un modèle de la forme $y_n = b_0 + b_1 x_{1n} + b_2 x_{2n} + \varepsilon_n$ et on suppose que les ε_n sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

On note $\beta = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ et $\hat{\beta} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}$ l'estimateur des MCO de β.

1. Rappelez l'écriture du modèle et l'expression de $\hat{\beta}$
2. En utilisant ce qui a été vu en cours, montrez que $\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (X'X)^{-1}\right)$
3. On suppose que $(X'X)^{-1} = \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & 2 & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{4} & 1 \end{pmatrix}$

   i) Quelle est la loi de $\hat{b}_0$? Quelle est la loi de $\hat{b}_2$?

   ii) Quelle est la loi du vecteur aléatoire $Z = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}$?

   iii) Quelle est la loi de la variable aléatoire réelle $2\hat{b}_1 - \hat{b}_2$?

Fiche de TD n°9

Régression multiple.
Tests sur les coefficients et intervallles de prévision

EXERCICE 1 :

On cherche à savoir dans quelle mesure le montant des loyers dans une ville universitaire est influencé par la présence de la population étudiante et on étudie pour cela un échantillon de 64 villes universitaires.

On note *loy* le logarithme du loyer moyen au m^2 des appartements dans chaque ville, *pop* le logarithme de la population totale de chaque ville, *revmoy* le logarithme du revenu moyen des ménages de la ville et *pctstu* le pourcentage d'étudiants dans la population de la ville.
On estime le modèle suivant :

$$loy = b_0 + b_1\ pop + b_2\ revmoy + b_3\ pctstu + \varepsilon$$

et on suppose que les résidus sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

1. i) Indiquez de façon précise comment s'interprètent les différents coefficients du modèle
   ii) Que pensez-vous des signes attendus pour b_1, b_2 et b_3 ?
2. Les résultats obtenus sont les suivants (en-dessous de chaque coefficient estimé, l'écart-type estimé correspondant figure entre parenthèses) :

$$\begin{array}{rcl} \widehat{loy} & = & \underset{(0.844)}{0.043} + \underset{(0.035)}{0.072}\ pop + \underset{(0.081)}{0.507}\ revmoy + \underset{(0.0017)}{0.0056}\ pctstu \\ R^2 & = & 0.458 \end{array}$$

   i) Testez la significativité des différents coefficients au seuil 5 %.
   ii) Quel est l'impact attendu sur les loyers, toutes choses égales par ailleurs, si la population de la ville augmente de 2 % (vous préciserez ce que l'on entend par "toutes choses égales par ailleurs") ?
   iii) Formulez l'hypothèse nulle selon laquelle le pourcentage d'étudiants dans la population n'a pas d'effet sur le loyer moyen au m^2, toutes choses égales par ailleurs.
   Comment formuleriez-vous l'hypothèse alternative, compte-tenu du signe attendu pour le coefficient correspondant ? Menez le test associé au seuil 5%. Qu'en concluez-vous ?

EXERCICE 2 :

On considère le modèle linéaire suivant : $y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + \varepsilon_n$ dans lequel on suppose que les résidus sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

Ce modèle est estimé sur $N = 25$ observations, et on obtient les résultats suivants :

$\hat{\beta} = \begin{pmatrix} 0.8 \\ 1.2 \\ 0.3 \end{pmatrix}$ et la matrice de variance-covariance estimée de $\hat{\beta}$ est

$$\hat{V}\hat{\beta} = \hat{\sigma}^2 (X'X)^{-1} = \begin{pmatrix} 0.13 & -0.11 & -0.013 \\ -0.11 & 0.31 & 0.026 \\ -0.013 & 0.026 & 0.0126 \end{pmatrix}$$

1. Testez la significativité des différents coefficients au seuil 5 %.
2. Testez, au seuil 5 %, l'hypothèse $H_0 : b_2 \geq 0.5$ contre l'hypothèse $H_1 : b_2 < 0.5$.
3. i) Quelle est la loi de $\hat{b}_1 - \hat{b}_2$?
   ii) Calculez la variance estimée de $\hat{b}_1 - \hat{b}_2$.
   iii) Construisez le test de l'hypothèse $H_0 : b_1 - b_2 = 1$ contre l'hypothèse $H_1 : b_1 - b_2 \neq 1$ et effectuez ce test au seuil 5%.
4. On suppose connus : $x_{26,1} = 1$ et $x_{26,2} = -1$. Construire un intervalle de prévision au niveau 95% pour y_{26}.