Université du Maine
Année universitaire 2016/2017
L2 Eco-Gestion / L3 Passerelle
Cours d'introduction à l'économétrie
Fiche de travaux dirigés

F. Karamé / S. Blasco / E. Kurtbegu

Exercice 1 : Questions de cours (à revoir après l'introduction et après la conclusion)

1. Quels sont les différents types de données ?
2. Rappelez la notion de corrélation. Une corrélation observée signifie-t-elle toujours quelque chose ?
3. Rappelez la notion de modèle économétrique. Quel est l'apport d'un modèle économétrique relativement à une analyse des corrélations ?
4. Qu'est-ce qu'une perturbation aléatoire ? Quelles sont les implications de cette hypothèse ?
5. Quelles sont les principales hypothèses sous lesquelles l'analyse de la régression est menée, ainsi que leur signification ?
6. Quel est le principe des MCO ? Pourquoi choisir cette méthode plutôt qu'une autre ?
7. Quelles sont les principales propriétés des estimateurs ?

Exercice 2 : Dérivation matricielle

1. Soit une combinaison linéaire $a'z = z'a = \sum_{i=1}^{n} a_i z_i$, avec a et z deux vecteurs colonnes de taille n. Vérifiez dans le cas $n = 3$ que la dérivation par rapport au vecteur de variables z donne :
$$\frac{\delta(a'z)}{\delta z} = \frac{\delta(z'a)}{\delta z} = a$$
2. Soit une forme quadratique $z'Az$ avec A une matrice carrée symétrique de taille n et z un vecteur de taille n. Vérifiez dans le cas $n = 3$ que la dérivation par rapport au vecteur z donne :
$$\frac{\delta(z'Az)}{\delta z} = 2Az$$

Exercice 3 : Linéarité et estimation par les MCO

Parmi les relations suivantes donnant y en fonction de x et parfois de z, quelles sont celles qui peuvent être considérées comme linéaires et être estimées par les MCO ? Justifier et réécrire quand c'est nécessaire l'équation du modèle.

$$y = ax + b \qquad y = x^2 + bx + c \qquad y = ax^2 + b \qquad y = a\ln(x) + 5 \qquad y = ax^2 + bx + c$$

$$y = ab^x c^z \qquad y = a3^x + b \qquad y = a + b^x c^z \qquad y = \frac{c}{1+a\exp(-bx)} \qquad y = \frac{a}{x-1} + b$$

Exercice 4 : Moments

Soit le modèle économétrique $y_t = a + bx_t + \varepsilon_t$, avec ε_t est d'espérance 0 et de variance σ^2.

1. Si x_t est aléatoire, montrez que $\mathbb{E}(y_t|x_t) = a + bx_t$ et $\mathbb{V}(y_t|x_t) = \sigma^2$.
2. Si x_t est non aléatoire, montrez que $\mathbb{E}(y_t) = a + bx_t$ et $\mathbb{V}(y_t) = \sigma^2$.

EXERCICE 5 : Régression linéaire simple
Soit le modèle économétrique suivant : $y_i = a + bx_i + \varepsilon_i$.

1. Utilisez le critère des MCO dans le cadre du modèle de régression linéaire simple pour donner la formule de l'estimateur des paramètres a et b.
2. Montrez que l'estimateur des MCO de b peut aussi s'écrire :

$$\hat{b} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})x_i}$$

3. Rappelez la définition du coefficient de détermination de la régression. Quelle est sa signification ?
4. Quelle est la relation entre le coefficient de détermination de la régression et l'estimateur de la pente de la droite de régression ?
5. Même question avec le coefficient de corrélation linéaire.
6. Réécrire l'estimateur du coefficient de la pente comme une combinaison linéaire des données (vous expliciterez les poids de cette combinaison). Interpréter.

EXERCICE 6 : Second exercice de l'introduction du cours
Soit le modèle économétrique suivant : $y_i = a + b\, l_i + \varepsilon_i$ avec y_i la note (sur 10) et l_i le nombre d'heures travaillées par semaine par un individu i. Les données sont fournies par le tableau suivant :

l	y
2	4
3	7
1	3
5	9
9	17

1. Comment s'interprète l'équation estimée ?
2. Calculer les estimations de a et de b par la méthode des MCO en utilisant les formules démontrées à l'exercice 5.
3. Même question en utilisant les formules matricielles du cours (dont vous repréciserez les différentes matrices, leur contenu et leur format). Finalement que préférez-vous ?
4. Calculez la variance des estimateurs.
5. Calculer le coefficient de détermination du modèle et le coefficient de détermination corrigé. Commenter.
6. Quelle critique peut-on malgré tout faire à ce modèle ?
7. Supposons que l'on multiplie y par 2 et l par 12. Interprétez cette transformation en termes de changements d'unités des variables. Quelles sont les conséquences en ce qui concerne les estimations ?

EXERCICE 7 : Le modèle sans constante
Soit le modèle économétrique sans constante suivant : $y_i = b\, l_i + \varepsilon_i$

1. Quelles implications du modèle linéaire simple ne sont plus vérifiées du fait de cette spécification ?
2. Trouvez l'estimateur des MCO de b. Est-il identique à celui des MCO dans le modèle standard ?
3. On propose deux estimateurs de b : $\hat{b}_1 = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$ et $\frac{\bar{y}}{\bar{x}}$. Montrez qu'ils sont sans biais.
4. Donner l'expression de la variance de ces deux estimateurs et déterminer lequel possède la plus petite variance. Quel intérêt y a-t-il à utiliser l'estimateur qui a la plus petite variance ?
5. Montrer que le coefficient de détermination calculé dans le cadre de ce modèle peut être négatif alors que celui calculé dans le cadre du modèle standard ne le peut pas.

EXERCICE 8 : Effet non-linéaire dans un modèle linéaire
On cherche à expliquer la consommation des ménages. Pour cela, on dispose d'un échantillon de 5 000 ménages dont on observe les dépenses de consommation, le niveau de revenu et le nombre d'enfants à une date donnée. Le modèle économétrique formulé est le suivant :

$$C_i = \alpha + \beta R_i + \delta nb\_enfants_i + \varepsilon_i$$

1. Quelles sont les implications de cette formulation en ce qui concerne le nombre d'enfant sur la consommation ?
2. Comment introduire des effets différenciés selon le plus ou moins grand nombre d'enfants ? Quelle est alors la nouvelle forme du modèle ?
3. Quelle méthode proposez-vous pour estimer un tel modèle ?

EXERCICE 9 : Effet non-linéaire dans un modèle linéaire
Soit un modèle permettant d'expliquer la consommation d'essence ($depenses$) en fonction de l'âge de l'individu (age), d'une variable indicatrice qui prend la valeur 1 lorsque l'individu est propriétaire de son logement et 0 sinon ($proprio$) et enfin du revenu de l'individu ($revenu$). On estime par la méthode des MCO le modèle suivant :

$$depenses_i = a_0 + a_1 age_i + a_2 proprio_i + a_3 revenu_i + a_4 revenu_i^2 + \varepsilon_i$$

Les résultats sont rassemblés dans le tableau suivant :

	Estimation	Ecart-type
$constante$	-237,15	199,35
age	-3,08	5,51
$proprio$	27,94	82,92
$revenu$	234,35	80,37
$revenu^2$	-15,00	7,47

1. Interprétez les paramètres estimés.
2. Représentez graphiquement l'effet du revenu sur les dépenses pour les propriétaires et les non propriétaires. Que cherche-t-on à capter avec cette variable ?

EXERCICE 10 : Gauss-Markov
On définit l'estimateur BLUE comme le meilleur estimateur linéaire non biaisé.
1. Préciser les propriétés d'un estimateur BLUE.
2. Construire l'estimateur BLUE de la pente de la droite de régression dans le modèle linéaire simple.
3. Montrez que l'estimateur BLUE est identique à celui des MCO.
4. Même question dans le cas général.

EXERCICE 11 : Convergences
1. Démontrez que l'estimateur des MCO est convergent en probabilités.
2. En déduire qu'il est convergent en loi.

EXERCICE 12 : Projecteurs orthogonaux
1. Dans le cas du modèle de régression linéaire multiple $y = X\beta + \varepsilon$, retrouvez la formule du projecteur sur le plan des variables explicatives X, noté P_X, et sur le plan qui lui est orthogonal, noté M_X.
2. Redémontrez les propriétés des projecteurs.
3. Explicitez le projecteur orthogonal sur la constante notée e_T. A quelle opération bien connue correspond la projection avec M_{e_T} ?
4. Quelle est la loi suivie par une forme quadratique d'un vecteur *iid* normal dont la matrice caractéristique est un projecteur orthogonal de rang r ?

EXERCICE 13 : Théorème de Frish & Waugh et modèles transformés par projection orthogonale
Soit le modèle classique $y = X\beta + \varepsilon$ que l'on partitionne en fonction de deux sous-ensembles de variables explicatives $X = (X_1|X_2)$:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

1. Quel est le programme à résoudre pour trouver les estimations de β_1 et β_2 ?
2. Démontrez que le système d'équations normales s'écrit :

$$\begin{pmatrix} X_1^{'}X_1 & X_1^{'}X_2 \\ X_2^{'}X_1 & X_2^{'}X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} X_1^{'}y \\ X_2^{'}y \end{pmatrix}$$

3. Repartez de cette expression pour déduire le théorème de Frish & Waugh et l'expression de l'estimateur de β_2.
4. En utilisant un projecteur orthogonal (que vous préciserez) et ses propriétés, retrouvez ce résultat de manière directe.
5. Appliquez au cas particulier où la partition consiste à séparer la constante des autres régresseurs $X = (e_T|Z)$: $y = e_T\,\alpha + Z\,\beta + \varepsilon$.
6. Calculer les projecteurs orthogonaux associés à cette partition.
7. En déduire que $\hat{\beta} = (\underline{Z}^{'}\underline{Z})^{'}\underline{Z}^{'}\underline{y}$ où le soulignement signifie $\underline{y} = y - e_T\bar{y}$.
8. En déduire $(\underline{Z}^{'}\underline{Z})^{'}\underline{Z}^{'}y$ (y n'est pas souligné).

EXERCICE 14 : Le modèle sans variable explicative
Soit un modèle avec simplement une constante.

1. Donnez l'expression de l'estimateur de la constante par les MCO.
2. Quelle sont les implications en ce qui concerne l'équation de la variance ?

EXERCICE 15 : Interprétation des paramètres associés aux variables indicatrices
Soit le modèle $y_i = a + bX_i + \varepsilon_i$ où y_i est le salaire perçu par l'individu i et X_i une classe de diplômes : si i est bachelier, alors $X_i = 1$, sinon $X_i = 0$. On dispose de n observations au total pour notre échantillon, n_0 individus ne possédant pas le bac et n_1 individus le possédant ($n_0 + n_1 = n$).

1. A partir d'une simple analyse en espérance conditionnelle sur les valeurs de X_i, montrez que $\hat{a} = \bar{y}_0$ et que $\hat{b} = \bar{y}_1 - \bar{y}_0$, avec $\bar{y}_0$ et $\bar{y}_1$ respectivement le salaire moyen des non bacheliers et des bacheliers.
2. Comment tester l'hypothèse que le salaire est indépendant du diplôme ?
3. Retrouvez les résultats de la question 1 grâce aux formules des estimateurs des MCO en régression linéaire simple.
4. Retrouvez ces résultats en ordonnant les observations de manière à ce que les n_0 observations pour lesquelles la variable indicatrice X_i prend la valeur 0 soient rangées en premier. Dans ce cas, la matrice des variables explicatives s'écrit :

$$X = \begin{pmatrix} e_{n_0} & 0_{n_0} \\ e_{n_1} & e_{n_1} \end{pmatrix}$$

   avec e_{n_0} un vecteur colonne de 1 de taille n_0, e_{n_1} de taille n_1 et 0_{n_0} un vecteur colonne de n_0 lignes de 0.
5. Même question en écrivant le modèle sans constante $y_i = cW_i + dZ_i + \varepsilon_i$, avec

$$W = \begin{pmatrix} e_{n_0} \\ 0_{n_1} \end{pmatrix} \text{ et } Z = \begin{pmatrix} 0_{n_0} \\ e_{n_1} \end{pmatrix}$$

6. Interprétez la signification des coefficients selon les différentes spécifications.

EXERCICE 16 : Spécification des modèles avec variables indicatrices et hypothèse H_3
La consommation C_i d'un individu i est supposée dépendre de son revenu, de son patrimoine W_i et de son âge age_i,

1. Construisez les 3 variables qualitatives définies comme suit :

$$I_{1,i} = \begin{cases} 1 & \text{si l'âge de } i \text{ est inférieur à 30 ans} \\ 0 & \text{sinon} \end{cases}$$

$$I_{2,i} = \begin{cases} 1 & \text{si l'âge de } i \text{ est compris entre 30 et 60 ans} \\ 0 & \text{sinon} \end{cases}$$

$$I_{3,i} = \begin{cases} 1 & \text{si l'âge de } i \text{ est supérieur à 60 ans} \\ 0 & \text{sinon} \end{cases}$$

2. Ecrivez le modèle en fonction de la variable âge. Quelle est la limite de cette spécification ? Comment la contourner ?
3. Ecrivez le modèle linéaire dans lequel les effets du revenu et de la richesse sur la consommation sont indépendants de l'effet de l'âge sur la consommation. Quelles sont les implications de cette spécification ? Leurs limites ?
4. Comment testeriez-vous l'hypothèse que l'âge n'a pas d'effet sur la consommation ?
5. Ecrivez le modèle linéaire dans lequel seul l'effet de la richesse sur la consommation dépend de l'âge de l'individu, c'est-à-dire que les effets de l'âge et du revenu sur la consommation sont indépendants.
6. Que pensez-vous du modèle suivant ?

$$C_i = \beta_1 R_i + \gamma_1 I_{1,i} + \gamma_2 I_{2,i} + \gamma_3 I_{3,i} + \beta_2 I_{1,i} R_i + \beta_3 I_{2,i} R_i + \beta_4 I_{3,i} R_i + \beta_5 W_i + \varepsilon_i$$

EXERCICE 17 : Estimation par la méthode du maximum de vraisemblance
Soit le modèle $y = X\beta + \varepsilon$ avec $\varepsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. On veut estimer le modèle par la méthode du maximum de vraisemblance.

1. Ecrire la densité de probabilités de y_t. De quels paramètres inconnus dépend-elle ?
2. Compléter l'expression du critère du maximum de vraisemblance : $\{?\} = \arg\max_{?} \prod_{t=1}^{T} \phi(y_t)$, où $\phi(.)$ est la densité de y_t.
3. Ré-écrire le critère à maximiser de manière plus simple.
4. En déduire l'expression des estimateurs.

EXERCICE 18 : Intervalles de confiance et tests sur les paramètres du premier ordre

1. Fournir un intervalle de confiance pour $\hat{\beta}_j = 6$ et $\sigma_{\hat{\beta}_j} = 2$.
2. Sachant que la taille de l'échantillon est égale à 28 et qu'on a estimé 3 paramètres dans le modèle, fournir un intervalle de confiance pour $\hat{\beta}_j = 6$ et $\widehat{\sigma}_{\hat{\beta}_j} = 2$. Comparer au résultat obtenu à la question précédente.
3. On a toujours $\hat{\beta}_j = 6$ et $\widehat{\sigma}_{\hat{\beta}_j} = 2$. Tester si $\beta_j = 0$.
4. Montrez qu'on peut accepter l'hypothèse nulle d'un test bilatéral avec une probabilité de risque de première espèce α en observant que la valeur supposée du paramètre appartient à l'intervalle de confiance du paramètre testé au niveau de confiance de $(1 - \alpha)$.

EXERCICE 19 : Intervalle de confiance pour la variance des perturbations

1. Fournir un intervalle de confiance pour la variance des perturbations aléatoires, σ^2.
2. Application numérique : $T = 50$, $k = 10$, $\widehat{\sigma^2} = 0.5$.

EXERCICE 20 : Distributions des statistiques de tests joints
Soit $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X))$. Soient C et c respectivement une matrice non aléatoire de format $r \times k$ et un vecteur non aléatoire de format $r \times 1$. Soit l'hypothèse nulle à tester : $H_0 : C\beta = c$.

1. Interpréter r. Pourquoi s'intéresser à $C\hat{\beta} - c$?

2. Quelle est la loi suivie par la variable aléatoire $C\hat{\beta} - c$?
3. Quelles en sont l'espérance et la variance ? Le format ?
4. Quelle est la loi suivie par la forme quadratique construite à partir de cette variable aléatoire ? Quels en sont les degrés de liberté ?
5. Pourquoi cette statistique est-elle intéressante ? Quel est son usage ?
6. Comment rendre cette statistique opérationnelle ?

EXERCICE 21 : Hypothèse nulle et modèle sous contrainte

1. Supposons la fonction de production suivante :

$$y_t = a + \alpha\, l_t + \beta\, k_t + \varepsilon_t$$

   Traduire l'hypothèse de rendements constants en fonction des paramètres du modèle et réécrire le modèle sous contrainte(s).

2. Soit l'équation de salaire suivante :

$$w_t = a + \alpha_0\, p_t + \alpha_1\, p_{t-1} + \alpha_2\, p_{t-2} + \gamma\, Z_t + \varepsilon_t$$

   avec Z les autres variables explicatives du modèle. Traduire l'hypothèse de parfaite indexation des salaires sur les prix en fonction des paramètres du modèle et réécrire le modèle sous contrainte(s).

3. Soit le modèle de régression multiple suivant :

$$y_t = a + b_1\, x_{1,t} + b_2\, x_{2,t} + b_3\, x_{3,t} + b_4\, x_{4,t} + \varepsilon_t$$

   Ecrire le modèle sous l'hypothèse $b_1 = b_2$ et $b_3 = -b_4$.

4. Soit le modèle de régression multiple suivant :

$$y_t = b_0 + b_1\, x_{1,t} + b_2\, x_{2,t} + \cdots + b_{k-1}\, x_{k-1,t} + \varepsilon_t$$

   Traduire l'hypothèse de non pertinence de la régression en fonction des paramètres du modèle et réécrire le modèle sous contrainte(s).

EXERCICE 22 : Entrainement au partiel n°1

Partie 1

Soit une fonction de production de Cobb-Douglas avec Q_t le produit et les facteurs de production capital K_t et travail L_t à la date t :

$$Q_t = AL_t^{\alpha} K_t^{\beta}$$

1. Ce modèle est-il linéaire ? Quelle transformation lui appliquer pour le rendre estimable par les MCO ?

   L'estimation par les MCO sur un échantillon d'observations trimestrielles allant de 1970 :1 à 2001 :2 a fourni les résultats suivants :

$$\hat{q}_t = 11.5 + 0.39 k_t + 0.69 l_t$$

   On a également $X = (e_T|k|l)$ avec k et l les variables prises en logarithme et e_T un vecteur de taille T contenant uniquement des 1 :

$$(X'X)^{-1} = \begin{pmatrix} 0.04 & 0.5 & 0.1 \\ 0.5 & 0.14 & -0.6 \\ 0.1 & -0.6 & 0.1 \end{pmatrix} \quad \text{et} \quad \hat{u}'\hat{u} = 0.33$$

2. Quelle hypothèse majeure a-t-on faite pour estimer ce modèle ? Pourquoi ? Quelles en sont les implications ?
3. Interpréter les coefficients estimés.
4. Fournir une estimation non biaisée de la variance des perturbations. Précisez l'intérêt d'une estimation non biaisée. Connaissez-vous un autre estimateur ? Précisez ses avantages et ses inconvénients.

5. En déduire une estimation de la variance des estimateurs des MCO.
6. Fournir une estimation par intervalles de confiance (au niveau de confiance de 95%) pour les paramètres associés à k et l.
7. Tester la nullité de ces paramètres au seuil de significativité de 5%.
8. La variance de la variable endogène est égale à 0.1068. En déduire la valeur du coefficient de détermination de la régression.
9. Tester la significativité globale de la régression (indication : pour cela, vous partirez de la statistique de Fisher associée aux modèles général et contraint et démontrerez qu'elle peut se réécrire uniquement comme une fonction du coefficient de détermination de la régression non contrainte).
10. Tester l'égalité entre les coefficients de k et de l. Interpréter économiquement. Conclure. (indication : vous préciserez (i) le principe du test réalisé en terme de modèles contraint et non contraint, (ii) la distribution suivie par la statistique utilisée, les degrés de liberté, ...(iii) la valeur de la statistique de test ainsi que la conclusion du test).

Partie 2 : test d'hypothèses

On cherche à présent à tester l'hypothèse de rendements constants dans la fonction de production.

11. Formuler l'hypothèse à partir du modèle le plus général.
12. Réécrire le modèle sous contraintes.
13. Ecrire ces contraintes sous la forme matricielle $C\beta = c$.
14. Décrire de manière littéraire et synthétique les différentes possibilités pour tester cette hypothèse.
15. Tester le modèle général contre le modèle contraint. La somme des carrés des résidus dans le modèle contraint est de 0.345. Conclure. (indication : vous préciserez (i) le principe du test réalisé en terme de modèles contraint et non contraint, (ii) la distribution suivie par la statistique utilisée, les degrés de liberté, ...(iii) la valeur de la statistique de test ainsi que la conclusion du test).
16. Aux vues des résultats, quel modèle faut-il retenir ?

Partie 3 : modèles emboités et sélection de modèles

On cherche à présent à estimer une fonction de production dite translog, dont la spécification est la suivante :

$$q_t = \beta_1 + \beta_2 k_t + \beta_3 l_t + \frac{\beta_4}{2} l_t^2 + \frac{\beta_5}{2} k_t^2 + \beta_6 k_t l_t + u_t$$

17. Pourquoi ce modèle peut-il être estimé par les MCO ?

L'estimation fournit les résultats suivants :

$$\hat{q}_t = 10 + 0.3k_t + 0.7l_t + \frac{1}{2} l_t^2 + \frac{1}{2} k_t^2 + 2k_t l_t$$

18. La somme des carrés des résidus de ce modèle est égale à 0.20. Tester ce modèle contre le modèle précédent (vous préciserez le modèle général et le modèle contraint), en détaillant les étapes de votre démarche.
19. Conclure en faveur d'un modèle. Commenter rapidement.

Partie 4 : variables indicatrices temporelles

On cherche à présent à modéliser l'effet des chocs pétroliers dans l'équation. On introduit dans le modèle une variable indicatrice I_t valant 1 en 1974 et 1979 et 0 pour les autres années.

20. Quelle est l'implication en termes de paramètres du modèle de cette spécification ? (indication : vous calculerez l'espérance de q_t conditionnellement aux différentes valeurs possibles de la variable I_t).
21. Par rapport à un modèle où on utiliserait 2 variables indicatrices (une valant 1 en 1974, 0 sinon et une valant 1 en 1979 et 0 sinon), quelle est l'hypothèse faite dans un modèle où on n'utilise qu'une seule variable indicatrice pour les 2 chocs pétroliers ?

EXERCICE 23 : Premier exercice d'introduction (spécification du modèle et questions 1 à 3) et entrainement au partiel n°2 (Stéphane Plazza)
On cherche à estimer un modèle de prix hédoniques. Ce type de modèles est utilisé pour donner une valeur à des actifs réels, particulièrement en immobilier, en supposant que la valeur du bien dépend d'un certain nombre de caractéristiques observables apportant de l'utilité ou de la désutilité au consommateur. L'idée sous-jacente étant évidemment d'utiliser le modèle une fois estimé pour déterminer de façon automatique le loyer de tout appartement dont on connaîtrait les caractéristiques. Ici le modèle considéré porte sur la détermination du niveau de loyer d'un échantillon de 13 378 appartements de la région de Québec pour l'année 1990. Il est estimé par les MCO.

Variable	Définition	Estimation	Student
Prix	Prix du loyer (par mois, charges incluses, en euros)		
Constante	-	282.21	56.09
Age	Age du bien	-53.10	-59.71
Nb_chambres	Nombre de chambres à coucher	48.47	104.81
Surface	Surface moyenne par chambre (en m^2)	3.97	29.99
Ascenseur	= 1 si présence d'un ascenseur, 0 sinon	88.51	45.04
RDC	= 1 si en rez-de-chaussée, 0 sinon	-15.90	-11.32
Parking_ext	Nombre de places de parking en extérieur	7.17	7.07
Parking_int	Nombre de places de parking en intérieur	73.76	31.25
Distance1	Distance au centre-ville (en km)	5.84	4.60
Distance2	Distance au centre commercial le plus proche (en km)	-10.04	-5.97
R^2	0.65		

1. Commentez les valeurs estimées et leur signe. Comment interpréter les paramètres associés aux variables indicatrices Ascenseur et RDC ? Les estimations vous paraissent-elles conformes à l'attente ?
2. Quelle est la limite de ce modèle ? Vous vous focaliserez plus particulièrement sur la variable « nombre de chambres à coucher » pour illustrer votre réponse.
3. Donnez la formule pour calculer le prix du loyer d'un appartement de 10 ans, disposant de 3 chambres à coucher, d'une surface moyenne par chambre de $10.5m^2$, au 3ième étage avec ascenseur, sachant que l'immeuble ne dispose d'aucune place de parking en extérieur et de 10 places intérieures, et qu'il est situé à 1 km du centre-ville et à 5 km du centre commercial le plus proche.
4. Discutez la significativité de chaque paramètre (après avoir rappelé l'hypothèse nulle, la statistique de test, la loi suivie et ses degrés de liberté, la valeur critique et la règle de décision du test).
5. Rappelez la définition du R^2 et le commentez.
6. En déduire la statistique de significativité globale de cette régression (vous rappellerez l'hypothèse nulle, les modèles contraint et non contraint, la statistique de test, la loi suivie et ses degrés de liberté, la valeur critique, la règle de décision et la conclusion du test).
7. Tester que l'impact sur le loyer du nombre des places de parking intérieures et extérieures est le même. (vous préciserez l'hypothèse nulle, les modèles contraint et non contraint, la statistique de test, la loi suivie et ses degrés de liberté, la valeur critique et la règle de décision du test).
8. Ecrire H_0 sous la forme matricielle $C\beta - c = 0$.
9. La statistique de test est égale à $2\,544$. Conclure concernant le modèle à retenir.

EXERCICE 24 : Pour aller plus loin : levée de l'hypothèse H_4
Soit le modèle standard $y = X\beta + \varepsilon$ tel que $\mathbb{E}(\varepsilon) = 0$ et $\mathbb{V}(\varepsilon) = \sigma^2\Omega \neq \sigma^2 I_T$, une matrice symétrique définie positive.

1. Quel est le problème ici ? Distinguez et commentez les différents cas possibles.
2. Montrez que l'estimateur des MCO est sans biais.
3. Donnez l'expression de la 'vraie' matrice de variance-covariance de $\hat{\beta}$. Quel(s) problème(s) risque-t-on de rencontrer si l'on calcule $\mathbb{V}(\hat{\beta})$ comme $\sigma^2(X'X)^{-1}$?

Soit le critère des Moindres Carrés Généralisés (MCG) : $\hat{\beta}_{MCG} = \arg\max_{\beta}(y - X\beta)'\Omega^{-1}(y - X\beta)$.

4. Discutez en fonction du critère des MCO.
5. Montrez que $\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$.

   Soit le modèle transformé $Py = PX\beta + P\varepsilon$ avec $\Omega^{-1} = P'P$ et P sa décomposée de Cholesky.

6. Quelle est l'interprétation de P ?
7. Montrez que $P\Omega P' = I_T$.
8. Calculez l'estimateur des MCO du modèle transformé et montrez que c'est l'estimateur des MCG.
9. Montrez que la matrice de variances-covariances des perturbations du modèle transformé vérifie à nouveau l'hypothèse traditionnelle H_4.
10. Quel problème rencontre-t-on pour appliquer les MCG dans la pratique ?

EXERCICE 25 : Pour aller plus loin : levée de l'hypothèse H_2
Soit le modèle simple sans constante suivant :

$$y_t = b\, x_t^* + \varepsilon_t$$

avec x_t^* une variable non directement observable. A la place, on se propose d'utiliser x_t tel que : $x_t = x_t^* + e_t$.

1. Ecrire le modèle que l'on va finalement estimer. Expliciter la structure de la perturbation aléatoire de ce modèle (qu'on notera u_t) en fonction de ε_t et de e_t.
2. Calculer $\mathbb{E}(x_t u_t)$. Commenter.
3. Mettre le modèle sous forme matricielle générique $y = X\beta + u$. Donner le programme des MCO à résoudre et sa solution générale. Appliquer au modèle que l'on traite ici.
4. Vers quoi converge $\frac{X'u}{T}$ ici ? En déduire que l'estimateur des MCO $\hat{b}$ est asymptotiquement biaisé.

   Soit Z un ensemble de variables dites instrumentales (tel que $dim(Z) \geq dim(X)$). On définit le critère des variables instrumentales comme : $\hat{\beta}_{IV} = \arg\max_{\beta}(y - X\beta)'P_Z(y - X\beta)$ avec P_Z le projecteur orthogonal sur le plan engendré par les variables Z.

5. Rappelez l'expression de P_Z.
6. Montrez que l'estimateur des variables instrumentales $\hat{\beta}_{IV} = (X'P_ZX)^{-1}X'P_Zy$.
7. Comment transformer le modèle original pour obtenir une estimation équivalente par MCO ?
8. Montrer que si $dim(Z) = dim(X)$, alors $\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$.
9. En utilisant les propriétés des projecteurs orthogonaux, montrez dans le cas général qu'il est nécessaire de projeter les explicatives X sur l'espace des Z mais pas nécessairement la variable à expliquer y.
10. Montrez que l'estimateur peut se réécrire : $\hat{\beta}_{IV} = \beta + (X'P_ZX)^{-1}X'P_Zu$.

    On admet à présent les convergences en probabilités suivantes :

$$\frac{Z'Z}{T} \xrightarrow{p} Q_{ZZ}$$
$$\frac{X'Z}{T} \xrightarrow{p} Q_{XZ}$$
$$\frac{Z'u}{T} \xrightarrow{p} 0$$

    (Attention : Q_{ZZ} est une matrice symétrique par construction alors que Q_{XZ} ne l'est pas).

11. En utilisant la définition du projecteur orthogonal P_Z et ses propriétés, montrer que $\hat{\beta}_{IV}$ est convergent en probabilité. En déduire la façon de choisir les variables Z.
12. On admet à présent que $\frac{Z'u}{\sqrt{T}} \xrightarrow{L} \mathcal{N}(0, \sigma^2 Q_{ZZ})$. Montrer que :

$$\sqrt{T}(\hat{\beta}_{IV} - \beta) \xrightarrow{L} \mathcal{N}(0, \sigma^2(Q_{XZ}Q_{ZZ}^{-1}Q_{XZ}')^{-1})$$

- Si $X \sim N(\mu, \sigma^2)$, alors $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right)$, $E(X) = \mu$ et $Var(X) = \sigma^2$.
- On note quelquefois U la v. a. gaussienne centrée-réduite et Φ sa fonction de répartition : $U \sim \mathcal{N}(0, 1)$.
- La table qui suit donne les valeurs de la fonction de répartition empirique de la loi normale centrée réduite $\Phi(x)$ pour les valeurs de x positives.
- Pour les valeurs négatives de x, on utilisera la relation $\Phi(x) = 1 - \Phi(-x)$

$\Phi(x) = P(X \leq x)$ où $X \sim N(0, 1)$ et $x = x_1 + x_2$

	x_2									
x_1	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.50	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.60	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.70	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.80	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.90	1	1	1	1	1	1	1	1	1	1

$t_{\nu,\alpha}$								
	α							
ν	0.6	0.75	0.9	0.95	0.975	0.99	0.995	0.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.373
1000	0.253	0.675	1.282	1.646	1.962	2.330	2.581	3.300

Loi de Fisher

(Valeurs de x telles que $Prob(F_{p,q} \leq x) = 95\%$)

$q \backslash p$	1	2	3	4	5	6	7	8	9	10	12
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75