

QCM Econométrie

QCM 1 : Régression linéaire simple

Q1) En matière de régression linéaire, laquelle des alternatives suivantes ne comprend que des synonymes de 'variable dépendante' ?

- (i) Le régressant
- (ii) Le régresseur
- (iii) La variable expliquée
- (iv) La variable explicative
- (v) La variable de contrôle θ

(ii) et (iv) seulement.

(i) et (iii) seulement.

(i), (ii) et (iii) seulement.

(i), (ii), (iii) et (v).

Q2) En matière de régression linéaire, laquelle des alternatives suivantes ne comprend que des synonymes de 'variable indépendante' ?

- (i) Le régresseur
- (ii) La variable de réponse=
- (iii) La variable causale
- (iv) La variable d'effet

☐ (ii) et (iv) seulement.

☒ **(i) et (iii) seulement.**

☐ (i), (ii) et (iii) seulement.

☐ (i), (ii), (iii) et (iv).

Q3) Laquelle des propositions suivantes est vraie concernant le modèle classique de régression?

☒ **y a une distribution de probabilité.**

☐ x a une distribution de probabilité.

☐ Le terme d'erreur peut être corrélé avec x.

☐ Pour qu'un modèle soit adéquat, l'erreur estimée (e) doit être nulle pour chaque donnée observée.

Q4) Laquelle des propositions suivantes est vraie concernant l'estimation par MCO?

☐ MCO minimise la somme des distances verticales entre les données observées et la droite de régression.

☐ MCO minimise la somme des distances horizontales entre données observées et la droite de moyenne.

☐ MCO minimise la somme des carrés des distances verticales entre les données observées et la droite de moyenne.

☒ **MCO minimise la somme des carrés des distances verticales entre les données observées et la droite de régression.**

Q5) Dans un modèle classique de régression, le résidu est défini comme:

- ☐ La différence entre la valeur actuelle, y , et la moyenne, y -barre.
- ☐ La différence entre la valeur ajustée, y -chapeau, et la moyenne, y -barre.
- ☒ La différence entre la valeur actuelle, y , et la valeur ajustée, y -chapeau.
- ☐ Le carré de la différence entre la valeur actuelle, y , et la valeur ajustée, y -chapeau.

Q6) Laquelle des propositions suivantes décrit le mieux la représentation algébrique de la droite de régression ajustée (ou fonction de régression de l'échantillon)?

- ☐ $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t$
- ☐ $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t + u_t$
- ☒ $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$
- ☐ $\hat{y}_t = \alpha + \beta\hat{x}_t + \hat{u}_t$

Q7) Laquelle des propositions suivantes concernant la population et l'échantillon est fausse?

- ☐ La population est l'ensemble complet des éléments d'intérêt.
- ☐ La population peut être infinie.
- ☒ En théorie, l'échantillon peut être plus grand que la population
- ☐ Un échantillon aléatoire est un échantillon pour lequel chaque unité individuelle issue de la population a une chance égale d'être tirée.

Q8) Laquelle des propositions suivantes concernant les fonctions de régression de la population (FRP) et de l'échantillon (FRE) est vraie?

- ☐ La FRP représente le modèle estimé.
- ☐ La FRP est utilisée pour induire les valeurs probables de la FRE.
- ☐ Une comparaison de la FRE et de la FRP peut aider à évaluer la qualité du modèle
- ☒ La FRP est une description du processus de génération des données (PGD).

Q9) Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle? (e indique l'exponentielle)

- (i) $y_t = \alpha + \beta x_t^2 + u_t$
- (ii) $y_t = \alpha + \beta e^{x_t} + u_t$
- (iii) $\ln(y_t) = \alpha + \beta x_t + u_t$
- (iv) $\ln(y_t) = \alpha + \beta \ln(x_t) + u_t$

θ (i) seulement.

θ (i) et (ii) seulement.

θ (i), (ii) et (iii) seulement.

☒ θ (i), (ii), (iii) et (iv).

Q10) Laquelle des conditions suivantes est nécessaire pour rendre possible l'estimation par MCO?

- ☐ Le modèle doit être linéaire par rapport aux variables
- ☐ Le modèle doit être linéaire par rapport aux variables et aux paramètres.
- ☒ Le modèle doit être linéaire par rapport aux paramètres.
- ☐ Le modèle doit être linéaire par rapport aux résidus

Q11) Laquelle des hypothèses suivantes n'est pas une hypothèse du modèle classique de régression?

- ☐ Toute variable explicative n'est pas corrélée avec le terme d'erreur.
- ☐ Le terme d'erreur a une moyenne nulle.
- ☒ La variable dépendante n'est pas corrélée avec le terme d'erreur.
- ☐ Il n'y a pas de corrélation sérielle dans le terme d'erreur pour des données temporelles.

Q12) Laquelle des expressions suivantes est équivalente à dire que 'x est non stochastique'?

- ☐ La variable explicative est partiellement aléatoire.
- ☒ La variable explicative est fixe d'échantillon en échantillon.
- ☐ La variable explicative est corrélée avec le terme d'erreur.
- ☐ Le terme d'erreur a toujours une valeur constante.

Q13) Laquelle des propositions suivantes définit le mieux un estimateur des MCO ?

- ☐ Il comprend les valeurs numériques obtenues par estimation des MCO.
- ☒ C'est une formule qui, appliquée aux données, donne une estimation par les MCO des paramètres du modèle.
- ☐ Il est égal à plusieurs estimations.
- ☐ C'est l'ensemble de toutes les données utilisées pour estimer le modèle de régression linéaire.

Q14) Considérons l'estimateur MCO de l'écart-type du coefficient de la pente de la droite de régression. Quelle(s) affirmation(s) est (sont) vraie(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est négativement relié à la variance de la variable explicative.
 - (iii) Il est négativement relié à la taille de l'échantillon.
 - (iv) Plus il est faible, plus le paramètre est estimé avec précision.
- ☐ (i) seulement.
 - ☐ (i) et (iii) seulement.
 - ☐ (i), (ii) et (iii) seulement.
 - ☒ (i), (ii), (iii) et (iv).

Q15) Soit $y = a + bx + \varepsilon$. Dans ce modèle de régression, la pente de la droite mesure :

- ☐ la variation de y en unités pour une variation donnée de 1% de x .
- ☐ l'élasticité de y par rapport à x .
- ☐ la valeur attendue de y lorsque la pente est multipliée par une valeur donnée de x .
- ☒ la variation de y en unité pour une variation d'une unité de x

Q16) Soit $y_i = a + b x_i + \varepsilon_i$. Dans ce modèle de régression,

- ☐ la valeur de l'ordonnée à l'origine est généralement négligeable.
- ☐ $E(y_i | x_i) = a + b x_i$ représente la fonction de régression de la population.
- ☐ la valeur absolue de la pente est généralement comprise entre -1 et +1.
- ☒ $y_i = a + b x_i$ correspond à la fonction de régression de l'échantillon

Q17) $E(u_i | x_i) = 0$ signifie que :

- ☒ la distribution de l'erreur affiche une moyenne nulle, quelles que soient les valeurs prises par x_i .
- ☐ la moyenne du ratio entre l'erreur et la variable explicative est égale à zéro, quelles que soient leurs valeurs.
- ☐ les résidus sont indépendants des valeurs prises par la variable explicative.
- ☐ la moyenne de x_i dans la population est beaucoup plus élevée que celle de u_i .

Q18) Dans lequel des modèles suivants, l'interprétation de l'ordonnée à l'origine est-elle la plus pertinente ?

- ☒ le taux de chômage par le taux de croissance du PIB ajusté pour l'inflation.
- ☐ la demande de café en fonction de son prix.
- ☐ la moyenne des notes obtenues à un examen cantonal en fonction du nombre d'étudiants par enseignant.
- ☐ le poids en fonction de la taille des individus.

Q19) Le coefficient de la pente traduit un effet « important » de X sur Y lorsque :

- ☐ sa valeur est significativement plus élevée que l'ordonnée à l'origine.
- ☒ sa valeur est économiquement représentative, que cette valeur soit faible ou élevée.
- ☐ la p-valeur est inférieure à 5%.
- ☐ sa valeur est élevée en valeur absolue.

QCM2 : Régression linéaire multiple

Q1) Sous la forme matricielle du modèle de RLM suivante, quelles sont les dimensions de e ? $Y = Xb + e$

☐ 1×1

☐ $n \times K$

☒ $n \times 1$

☐ $1 \times n$

Q2) Sous la même forme matricielle qu'à la question précédente, quelles sont les dimensions de ϵ ?

☒ 1×1

☐ $n \times K$

☐ $n \times 1$

☐ $1 \times n$

Q3) Sous la même forme matricielle qu'à la question précédente, quelle sont les dimensions de b ?

☐ 1×1

☐ $n \times K$

☐ $n \times 1$

☒ $K \times 1$

Q4) Vous disposez des niveaux de consommation (C) et de revenu (R) de 100 ménages en 2015. Vous construisez le modèle suivant : $\ln(C_i) = a + b\ln(R_i) + \epsilon_i$.

☐ L'ordonnée à l'origine mesure la consommation du ménage médian de votre échantillon.

☒ La pente mesure l'élasticité de la consommation par rapport au revenu, soit la propension marginale à consommer ou encore $\% \Delta C / \% \Delta R$.

☐ La pente, $\Delta C / \Delta R$, mesure l'évolution de la consommation en unité pour une variation d'une unité du revenu.

☐ L'ordonnée à l'origine mesure la consommation moyenne observée dans votre échantillon.

Q5) Soit $\ln(\text{salary}_i) = 6,46 + 0,000594 \text{ profits}_i$, où salary_i est le salaire du PDG mesuré en milliers de dollars et profits_i est le profit de son entreprise mesuré en millions de dollars.

☐ L'ordonnée à l'origine n'a pas de signification économique étant donné que le log est appliqué à la variable dépendante.

☐ L'ordonnée à l'origine signifie que le salaire est égal à 6 460 dollars lorsque l'entreprise ne réalise aucun profit.

☒ La pente signifie que le salaire augmente d'environ 5,94% lorsque le profit augmente de 100 millions de dollars.

☐ La pente signifie que le salaire augmente d'environ 1% lorsque le profit augmente de 594 dollars

Q6) Soit $salary_i = -415,11 + 177,15 \ln(sales_i)$, où $salary_i$ est le salaire du PDG mesuré en milliers de dollars et $sales_i$ est le chiffre d'affaires mesuré en millions de dollars.

θ Le salaire chutera de 415 mille dollars si la firme ne réalise aucun chiffre d'affaires.

θ Le salaire augmente de 177 mille dollars environ lorsque le chiffre d'affaire augmente d'1 million de dollars.

θ Le salaire augmente de 177 mille dollars environ lorsque le chiffre d'affaire double.

θ Le salaire diminue de 177 mille dollars environ lorsque le chiffre d'affaire diminue de 1%.

Q7) Soit $\ln(salary_i) = 4,96 + 0,22 \ln(sales_i)$, où $salary_i$ est le salaire du PDG mesuré en milliers de dollars et $sales_i$ est le chiffre d'affaires mesuré en millions de dollars.

θ Le salaire estimé d'un PDG dont la firme ne réalise aucun chiffre d'affaires est égal à 143 mille dollars environ.

θ Le salaire d'un PDG augmente de 22 mille dollars environ lorsque le chiffre d'affaires augmente de 100%.

θ Le salaire d'un PDG augmente de 22% environ lorsque le chiffre d'affaires augmente de 1%.

θ L'élasticité du salaire au chiffre d'affaires est égale à 0,22 environ

Q8) Soit $\ln(salary_i) = 4,68 + 0,26 \ln(mktval_i)$, où $salary_i$ est le salaire du PDG mesuré en milliers de dollars et $mktval_i$ est la capitalisation boursière de son entreprise mesurée en millions de dollars. Quelle affirmation parmi les suivantes est fausse ?

θ Le salaire estimé d'un PDG dont l'entreprise affiche une capitalisation boursière nulle ne peut pas être calculé.

θ Le salaire estimé d'un PDG augmente de 0,26% environ lorsque la capitalisation boursière augmente de 1%.

θ Le salaire estimé d'un PDG diminue de 6,5% environ lorsque la capitalisation boursière chute d'un quart.

θ Le salaire estimé d'un PDG dont l'entreprise affiche une capitalisation boursière de 1 milliard de dollars, est égal à 265 mille dollars environ

QCM3 : Hypothèses du modèle

Q1) Un estimateur des MCO est BLUE quand:

- ☐ Il est sans biais.
- ☐ Il est le moins biaisé de tous les estimateurs à variance minimale.
- ☒ il dispose de la variance minimale parmi tous les autres estimateurs linéaires sans biais.
- ☐ Il a une variance hétéroscédastique.

Q2) Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est PAS une implication?

- ☐ La probabilité que l'estimation soit fortement éloignée de la valeur vraie du paramètre est minimisée.
- ☐ L'estimateur est efficient.
- ☐ Un tel estimateur est qualifié de BLUE s'il est linéaire, sans biais et affiche une variance minimale par rapport aux autres estimateurs linéaires sans biais.
- ☒ Un tel estimateur sera toujours sans biais.

Q3) Un estimateur des MCO est consistant quand:

- ☐ En moyenne, il est sans biais.
- ☐ Il est sans biais mais ne dispose pas de la variance minimale.
- ☐ Les estimations des paramètres qu'il délivre sont égales à la vraie valeur des paramètres, quelle que soit la taille de l'échantillon.
- ☒ Les estimations qu'il délivre convergent vers la valeur vraie des paramètres au fur et à mesure que la taille de l'échantillon augmente.

Q4) Parmi les hypothèses suivantes, quelles sont celles qui sont requises pour obtenir un estimateur sans biais?

- (i) Il n'y a pas de multicolinéarité entre les variables explicatives.
- (ii) Il n'y a pas de colinéarité parfaite entre les variables explicatives.
- (iii) L'erreur conditionnelle est nulle en moyenne.
- (iv) La variance du terme d'erreur doit être constante.

- ☐ (i) et (iv).
- ☐ (ii) et (iii).
- ☒ (i), (ii) et (iii).
- ☐ (i), (ii), (iii) et (iv).

Q5) Parmi les hypothèses suivantes, quelles sont celles qui sont requises pour obtenir un estimateur sans biais?

- (i) Il n'y a pas de multicolinéarité entre les variables explicatives.
- (ii) Il n'y a pas de colinéarité parfaite entre les variables explicatives.
- (iii) L'erreur conditionnelle est nulle en moyenne.
- (iv) La variance du terme d'erreur doit être constante.

☐ (i) et (iv).

☐ (ii) et (iii).

☐ (i), (ii) et (iii).

☒ (i), (ii), (iii) et (iv).

Q6) Une multicolinéarité élevée surgit lorsque:

☐ Au minimum une variable explicative est fortement corrélée avec le terme d'erreur.

☒ Au minimum deux variables explicatives sont fortement corrélées entre elles.

☐ Au minimum une variable explicative est fortement corrélée avec la variable dépendante.

☐ Au minimum deux variables explicatives sont parfaitement corrélées entre elles.

Q7) Parmi les affirmations suivantes, laquelle n'est PAS une indication de multicolinéarité?

☐ La significativité des variables incluses dans la régression changent de manière substantielle lorsqu'une variable est ajoutée ou retirée.

☒ Les écart-types estimés des coefficients ne varient pas de manière substantielle lorsqu'une variable est ajoutée ou retirée.

☐ Le R^2 ne change pas de manière substantielle lorsqu'une variable est ajoutée ou retirée, cela étant d'autant plus vrai que le R^2 est élevé.

☐ Il y a des variables non significatives ou marginalement significatives.

Q8) Parmi les affirmations suivantes, laquelle n'est PAS un remède contre la multicolinéarité?

☒ Mesurer chacune des variables colinéaires en logarithme plutôt qu'en niveau.

☐ Utiliser un plus grand nombre d'observations.

☐ Supprimer une des variables colinéaires.

☐ Utiliser l'analyse en composante principale.

Q9) Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

- ☐ le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
- ☒ la variable omise est fortement corrélée avec les variables incluses;
- ☐ la variance des variables incluses est grande;
- ☐ la taille de l'échantillon est grande.

Q10) Lorsqu'une variable pertinente est omise, lesquelles des affirmations suivantes sont vraies?

- (i) Les estimations des coefficients sont biaisées pour autant que la variable omise soit corrélée avec les variables incluses dans le modèle.
- (ii) Les estimations des écart-types sont biaisées pour les variables incluses et corrélées avec la variable omise.
- (iii) Si la variable omise n'est pas corrélée avec les variables incluses, les estimations de leurs coefficients sont biaisées et ne sont pas consistantes.
- (iv) Quel que soit le degré de corrélation entre la variable omise et les variables incluses, l'estimation de la constante est biaisée et n'est pas consistante.

☐ (i) et (ii).

☐ (i) et (iii).

☒ (i) et (iv).

☐ (i), (ii), (iii) et (iv).

Q11) Si une variable superflue est présente dans la régression,

- ☐ les estimateurs des MCO ne sont plus BLUE.
- ☐ les écart-types des estimateurs des MCO diminuent.
- ☒ le risque de multicollinéarité est plus élevé.
- ☐ le nombre de degrés de liberté augmente.

Q12) La multicollinéarité :

- ☐ implique la violation d'une des hypothèses de Gauss-Markov.
- ☒ élargit les intervalles de confiance.
- ☐ réduit la variance des paramètres estimés.
- ☐ ne peut pas avoir comme conséquence d'éloigner l'estimation d'un paramètre (obtenue à partir d'un échantillon donné) de sa valeur vraie.

QCM4 : Tests d'inférence statistique

Q1) La régression linéaire multivariée suivante est estimée sur base de 27 observations mensuelles:

$$y_t = \alpha + \beta x_{2t} + \gamma x_{3t} + u_t$$

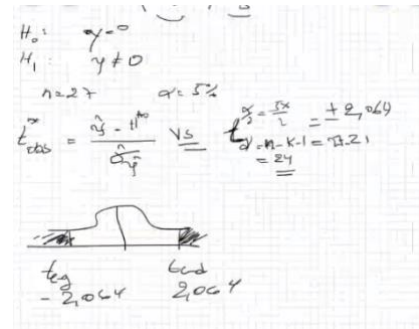
Quelle valeur critique doit être utilisée pour tester bilatéralement, à un seuil de 5 %, l'hypothèse nulle selon laquelle $\gamma=0$?

☐ 1.64

☐ 1.71

☒ 2.06

☐ 1.96



Q2) Si nous voulons tester $H_0: \beta=0$ de la même manière et sur base de la même régression qu'à la question précédente, la valeur critique à utiliser sera:

☐ plus élevée et issue de la distribution de Fisher-Snedecor.

☐ identique mais plus imprécise.

☐ moins élevée et issue de la distribution t de Student.

☒ identique et issue de la même distribution statistique.

Q3) Sur base de $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$ et de 30 observations, on obtient :

$$(X'X)^{-1} = \begin{bmatrix} 0.3 & 0.2 & -0.1 \\ 0.2 & 0.6 & -0.3 \\ -0.1 & -0.3 & 0.4 \end{bmatrix}, (X'y) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, e'e = 0.6$$

Quelle est l'estimation de a_2 ?

☒ 0.01

☐ 0.2

☐ -0.09

☐ 0.4

Q4) En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_1 ?

☐ 0.6

☐ 0.013

☒ 0.115

☐ 0.022

Q5) En se basant sur les informations données à la question 3, quelle est la valeur du t^* (ou t de Student empirique) résultant d'un 't-ratio test' sur la constante?

- ☒ -1.10
- ☐ -0.09
- ☐ 0.3
- ☐ 0.6

Q6) Dans un modèle de RLM, supposons qu'un t-ratio test sur la constante ne soit pas significatif. Quelle conclusion en tirez-vous?

- ☐ Il faut recalculer la statistique du test.
- ☒ Il faut garder la constante.
- ☐ Il faut supprimer la constante et estimer à nouveau le modèle.
- ☐ La ligne de régression passe par l'origine.

Q7) Laquelle des affirmations suivantes concernant les tests classiques d'hypothèse est fausse?

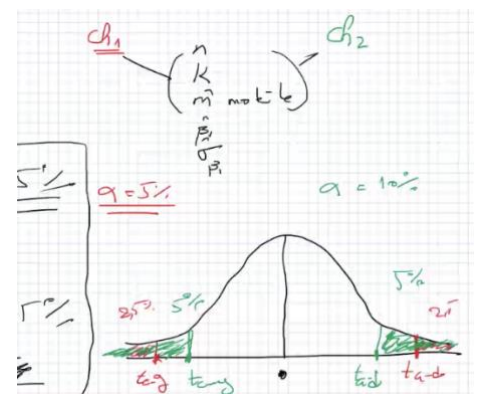
- ☐ L'hypothèse nulle est l'affirmation qui est testée.
- ☐ Les approches du test du coefficient (ou de significativité) et de l'intervalle de confiance conduisent toujours aux mêmes décisions.
- ☒ Si l'hypothèse nulle est rejetée, l'hypothèse alternative est acceptée.
- ☐ Les tests d'hypothèses sont utilisés pour induire les paramètres de la population à partir des paramètres de l'échantillon.

Q8) Supposons que le test du coefficient soit conduit à un seuil de 5%. Quelle(s) affirmation(s) est (sont) vraie(s)?

- (i) Le seuil de significativité est égal à la taille du test.
 - (ii) Le seuil de significativité est égal à la puissance du test.
 - (iii) 2,5% de la distribution totale sera comprise dans chacune des deux zones de rejet, s'il s'agit d'un test bilatéral.
 - (iv) 5% de la distribution totale sera comprise dans chacune des deux zones de rejet, s'il s'agit d'un test bilatéral.
- ☐ (ii) et (iv) seulement.
 - ☒ (i) et (iii) seulement.
 - ☐ (i), (ii) et (iii) seulement.
 - ☐ (i), (ii), (iii) et (iv).

Q9) Deux chercheurs utilisent le même modèle et les mêmes données. Ils obtiennent les mêmes estimations des coefficients et des écart-types. Dans un test bilatéral, le chercheur 1 utilise un seuil de 5% alors que le chercheur 2 utilise un seuil de 10%. Laquelle des affirmations suivantes est correcte?

- ☐ Le chercheur 2 utilisera une valeur critique plus élevée.
- ☒ Le chercheur 2 aura une probabilité d'erreur de type I plus grande.
- ☐ L'hypothèse nulle sera plus facilement rejetée par le chercheur 1.
- ☐ Les deux chercheurs aboutiront aux mêmes décisions.



Q10) Sur base de 100 observations, on obtient (avec les écart-types estimés entre parenthèses):

$$\hat{y}_i = -1.3 - 0.3x_i$$

(2.1) (0.1)

Testons, à un seuil de 5%, si la valeur 'vraie' du coefficient de la pente est plus grande ou égale à -1, contre l'hypothèse alternative qu'elle est inférieure à -1. Quelle est la conclusion?

- ☐ H0 est rejetée.
- ☒ H0 n'est pas rejetée.
- ☐ Il est impossible de conclure.
- ☐ Le test est erroné: il aurait dû être bilatéral.

10) $n = 100$ $\alpha = 5\%$

$$\hat{Y} = -1.3 - 0.3x$$

(2.1) (0.1)

$H_0: \alpha_1 \geq -1$

$H_1: \alpha_1 < -1$

$$t_{obs} = \frac{\hat{\alpha}_1 - \alpha_1^*}{\hat{\sigma}_{\hat{\alpha}_1}} = \frac{-0.3 + 1}{0.12} = \frac{0.7}{0.12} = 5.83$$

$t_{crit} = t_{\alpha/2, n-1} = t_{0.025, 99} = 1.98$

$t_{obs} = 5.83 > 1.98$

Q11) Considérons une situation identique à la question 10, mais effectuons un test bilatéral. A quelle conclusion aboutit-on?

- ☐ On utilise les mêmes valeurs critiques et on rejette l'H0.
- ☒ Même si les valeurs critiques changent et sont plus élevées en valeur absolue, on rejette l'H0.
- ☐ On utilise les mêmes valeurs critiques, et on ne rejette toujours pas H0.
- ☐ Comme les valeurs critiques changent, il est impossible de conclure.

Q12) Sur base du modèle estimé à la question 10, quel est l'intervalle de confiance à 95% pour le coefficient d'intersection?

- ☐ (-4.79, 2.19)
- ☐ (-4.16, 4.16)
- ☐ (-1.98, 1.98)
- ☒ (-5.46, 2.86)

12) $IC = \hat{\beta}_1 \pm t_{\alpha/2, n-1} \hat{\sigma}_{\hat{\beta}_1}$

$$IC = -1.3 \pm 1.98 \times 2.1$$

$$= [-5.46, 2.86]$$

Q13) Parmi les définitions suivantes, quelle est celle qui définit le mieux un intervalle de confiance à 99%?

- ☒ Dans 99 cas sur 100, suite à des tirages répétés, l'intervalle de confiance contiendra la valeur vraie du paramètre.
- ☐ Dans 99 cas sur 100, suite à des tirages répétés, l'intervalle de confiance contiendra l'estimation du paramètre.
- ☐ Dans 99 cas sur 100, suite à des tirages répétés, l'hypothèse nulle ne sera pas rejetée alors qu'elle est fausse.
- ☐ Dans 99 cas sur 100, suite à des tirages répétés, l'hypothèse nulle sera rejetée.

Q14) Laquelle des affirmations suivantes décrit le mieux une erreur de type II (beta)?

- ☐ C'est la probabilité de rejeter une hypothèse nulle alors qu'elle est, en réalité, vraie.
- ☐ C'est égal à la puissance du test.
- ☐ C'est équivalent à la taille du test.
- ☒ C'est la probabilité de ne pas rejeter une hypothèse nulle qui est, en réalité, fausse.

Q15) Considérons une augmentation de la taille du test (p.e. de 5% à 10%). Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs ?

- ☐ La probabilité de l'erreur de type I diminue.
- ☒ La probabilité de l'erreur de type II diminue.
- ☐ Le critère de rejet devient plus strict.
- ☐ L'hypothèse nulle sera rejetée moins souvent.

Q16) Supposons que la probabilité critique d'une statistique est égale à 0.08. Quelle(s) affirmation(s) est (sont) vraie(s) ?

(i) Si la taille du test était de 8% exactement, nous serions indifférents entre le rejet et le non rejet de H_0 .

(ii) H_0 serait rejetée si la taille du test était de 10%.

(iii) H_0 ne serait pas rejetée si la taille du test était de 1%.

(iv) H_0 serait rejetée si la taille du test était de 1%.

- ☐ (ii) et (iv) seulement.
- ☐ (i) et (iii) seulement.
- ☒ (i), (ii) et (iii) seulement.
- ☐ (i), (ii), (iii) et (iv).

Q17) Sur base de 1000 observations et de la spécification suivante, $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$, quel est un modèle contraint possible pour tester $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$?

- ☐ $y_t = \beta_0 + \beta_2(x_{2t} - x_{1t}) + \beta_3 x_{3t} + u_t$
- ☐ $(y_t - x_{2t}) = \beta_0 + \beta_1(x_{1t} + x_{2t}) + \beta_3 x_{3t} + u_t$
- ☒ $(y_t - x_{1t}) = \beta_0 + \beta_2(x_{2t} - x_{1t}) + \beta_3 x_{3t} + u_t$
- ☐ $(y_t - x_{1t}) = \beta_0 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

FR ? : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$
 ~~H_0~~ : $\beta_1 + \beta_2 = 1 \Leftrightarrow \beta_1 = 1 - \beta_2$
 H_1 : $\beta_1 + \beta_2 \neq 1$

$Y = \beta_0 + (1 - \beta_2) X_1 + \beta_2 X_2 + \beta_3 X_3 + u$
 $Y = \beta_0 + X_1 - \beta_2 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$
 $Y - X_1 = \beta_0 + \beta_2 (X_2 - X_1) + \beta_3 X_3 + u$ (MC)

Q18) Sur base du modèle spécifié à la question précédente, quelle(s) hypothèse(s) peu(ven)t être testée(s) à l'aide d'un F-test ?

(i) $H_0: \beta_3 = -\beta_1$ vs $H_1: \beta_3 \neq -\beta_1$

(ii) $H_0: \beta_1 * \beta_2 = 1$ vs $H_1: \beta_1 * \beta_2 \neq 1$

(iii) $H_0: \beta_1 \geq 1$ vs $H_1: \beta_1 < 1$

(iv) $H_0: \beta_1 = (\beta_3 - \beta_2)$ ET $\beta_2 = 1$ vs $H_1: \beta_1 \neq (\beta_3 - \beta_2)$ ET $\beta_2 \neq 1$

- ☒ (i) seulement.
- ☐ (i) et (iii) seulement.
- ☐ (i) et (iv) seulement.
- ☐ (i), (ii), (iii) et (iv).

Q19) Si nous voulons tester l'hypothèse nulle stipulée à la question 17, quelle serait la valeur critique à utiliser à un seuil de confiance de 5%?

- ☐ 254
- ☐ 253
- ☐ 3,00
- ☒ 3,84

Q20) Après avoir estimé les modèles contraint et non contraint relatifs à l'hypothèse nulle stipulée à la question 17, on obtient les deux sommes du carré des résidus suivantes: 30,2 et 28,1. Quelle est la valeur de F* (ou Fisher empirique)?

- ☐ 37,2
- ☐ -37,2
- ☒ 74,4
- ☐ -74,4

Q21) Quelle serait l'hypothèse nulle pour 'le' test de Fisher (test de signification globale) concernant le modèle spécifié à la question 17?

- ☒ $\beta_1 = 0$ ET $\beta_2 = 0$ ET $\beta_3 = 0$
- ☐ $\beta_1 = 0$ OU $\beta_2 = 0$ OU $\beta_3 = 0$
- ☐ $\beta_0 = 0$ ET $\beta_1 = 0$ ET $\beta_2 = 0$ ET $\beta_3 = 0$
- ☐ $\beta_0 = 0$ OU $\beta_1 = 0$ OU $\beta_2 = 0$ OU $\beta_3 = 0$

Q22) Quelle est la relation entre la distribution t de Student et la distribution de FisherSnedecor?

- ☐ Si $Z \sim F(1,r)$, alors $Z \sim t(r)$.
- ☐ Si $Z \sim t_2(r)$, $Z \sim F(r,1)$.
- ☒ Si $Z \sim F(1,r)$, alors $Z \sim t_2(r)$.
- ☐ Il manque des informations pour déterminer la nature exacte de la relation.

Q23) Laquelle des affirmations suivantes est toujours vraie concernant la somme du carré des résidus?

- ☐ $SCR_{nc} > SCR_c$
- ☐ $SCR_{nc} \geq SCR_c$
- ☐ $SCR_c > SCR_{nc}$
- ☒ $SCR_c \geq SCR_{nc}$

Q24) Que mesurent les statistiques de qualité d'ajustement (goodness-of-fit statistics), comme le R^2 ?

- ☐ La manière dont la FRP s'ajuste aux données observées.
- ☐ La qualité d'ajustement entre la FRE et la FRP.
- ☐ La manière dont la FRE s'ajuste à la FRP.
- ☒ La qualité d'ajustement de la FRE par rapport aux données observées.

Q25) Quelle est la définition la plus appropriée du R^2 ?

- ☐ Le R^2 mesure la proportion de la variation totale de la variable dépendante qui est expliquée par le modèle.
- ☐ Le R^2 mesure la corrélation entre les valeurs ajustées et les valeurs résiduelles de la variable dépendante.
- ☒ Le R^2 mesure la proportion de la variation totale de la variable dépendante par rapport à sa moyenne, qui est expliquée par le modèle.
- ☐ Le R^2 mesure la corrélation entre les valeurs ajustées et la moyenne de la variable dépendante.

Q26) Si la valeur du R^2 est égale à zéro, lesquelles des affirmations suivantes sont correctes?

- (i) Tous les coefficients, à l'exclusion de la constante, sont nuls.
 - (ii) La ligne d'ajustement est horizontale par rapport à chacune des variables explicatives.
 - (iii) Le modèle de régression n'explique rien de la variabilité totale de la variable dépendante par rapport à sa moyenne.
 - (iv) Tous les coefficients, dont la constante, sont nuls.
- ☐ (ii) et (iv).
 - ☐ (i) et (iii).
 - ☒ (i), (ii) et (iii).
 - ☐ (i), (ii), (iii) et (iv).

Q27) Si la valeur du R^2 est égale à 1, lesquelles des affirmations suivantes sont correctes?

- (i) Toutes les valeurs observées se trouvent sur la ligne de régression.
 - (ii) Toutes les valeurs résiduelles sont nulles.
 - (iii) Toute la variabilité de la variable dépendante par rapport à sa moyenne est expliquée par le modèle.
 - (iv) La droite d'ajustement est horizontale par rapport à toutes les variables explicatives du modèle.
- ☐ (ii) et (iv).
 - ☐ (i) et (iii).
 - ☒ (i), (ii) et (iii).
 - ☐ (i), (ii), (iii) et (iv).

Q28) Considérons les deux modèles suivants.

Modèle I: $y_t = \beta_0 + \beta_1 x_{1t} + u_t$

Modèle II: $y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \varepsilon_t$

Lesquelles de ces affirmations sont correctes?

- (i) Le modèle II n'affichera jamais un R^2 inférieur à celui du modèle I.
- (ii) Le modèle II n'affichera jamais un R^2 ajusté inférieur à celui du modèle I.
- (iii) Les deux modèles auront un R^2 identique si le coefficient estimé de α_2 est égal à zéro dans la population.
- (iv) Les deux modèles auront un R^2 ajusté identique si l'estimation de α_2 est égal à zéro dans la population.

☐ (ii) et (iv). ☒ (i) et (iii). ☐ (i), (ii) et (iii). ☐ (i), (ii), (iii) et (iv).

Q29) Sur base des deux modèles spécifiés à la question précédente, supposons que le modèle II affiche à la fois un R^2 plus élevé et un R^2 ajusté moins élevé que le modèle I. Laquelle de ces affirmations est la plus plausible?

- ☐ L'estimation de α_2 est égal à zéro.
- ☒ L'estimation de α_2 est différent de zéro mais de manière non significative.
- ☐ Les variables x_1 et x_2 sont fortement corrélées.
- ☐ Ce résultat est impossible puisque il y a moins de degrés de liberté dans le modèle II.

Q30) Sur base des deux modèles spécifiés à la question 28, supposons que les modèles I et II affichent un R^2 identique. Laquelle de ces affirmations est correcte?

- ☐ Les deux modèles ont également un R^2 ajusté identique.
- ☐ Le modèle II a un R^2 ajusté plus élevé.
- ☒ Le modèle II a un R^2 ajusté moins élevé.
- ☐ Il est impossible de savoir quel modèle affiche le R^2 le plus élevé sans connaître la taille de l'échantillon.

Q31) Quelle est la relation, s'il y en a, entre la distribution normale et la distribution t de Student ?

- ☐ Une distribution t sans degré de liberté est normale.
- ☐ Une distribution t avec un degré de liberté est normale.
- ☒ Une distribution t avec des degrés infinis de liberté est normale.
- ☐ Il n'existe aucune relation entre ces deux distributions.

Q32) Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont correctes?

- (i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est rejetée à un seuil de confiance de 10%.
- (ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.
- (iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.
- (iv) Le coefficient estimé de cette variable affiche un écart-type nécessairement petit.

☒ (i) et (ii). ☐ (i) et (iii). ☐ (i), (ii) et (iv). ☐ (i), (ii), (iii) et (iv).

Q33) Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 + a_2 \geq 3$ versus $H_1: a_1 + a_2$ lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

- ☐ $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + u_t$
- ☒ $y = a_0 + (a_1 + a_2)x_1 + a_2(x_2 - x_1) + a_3 x_3 + u_t$
- ☐ $y = a_0 + (a_1 + a_2)x_1 + (a_2 - a_2)(x_2 - x_1) + (a_3 - a_2)(x_3 - x_1) + u_t$
- ☐ $y = (a_0 - a_2 x_1) + (a_1 + a_2)x_1 + a_2 x_2 + a_3 x_3 + u_t$

Q34) Le test spécifié à la question précédente peut être effectué à l'aide:

- ☐ de la distribution de Fisher-Snedecor.
- ☒ de la distribution t de Student.
- ☐ des distributions t de Student et de Fisher-Snedecor
- ☐ impossible à déterminer.

Q35) Sur base du modèle spécifié à la question 33 et en supposant une hypothèse alternative bilatérale, quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

- (i) $a_1 + a_2 = 3$.
- (ii) $a_1 = 1$ et $a_2 = 0$.
- (iii) $a_1 = 2$ et $a_2 = 6$.
- (iv) $a_2 = 0$.
- ☐ aucune. ☐ (ii), (iii) et (iv). ☒ (i), (ii) et (iii). ☐ (i), (ii), (iii) et (iv).

Q36) Sur base du modèle spécifié à la question 33 et en supposant une hypothèse alternative bilatérale, lesquelles des hypothèses nulles suivantes peuvent être testée(s) à l'aide d'un F-test effectué à partir des R^2 des modèles contraint et non contraint?

- (i) $a_1 + a_2 = 1$.
- (ii) $a_1 + a_2 = 0$.
- (iii) $a_1 = 1$ et $a_2 = 0$.
- (iv) $a_1 = 0$ et $a_2 = 0$.
- ☐ (i), (ii), (iii) et (iv). ☐ (i), (ii) et (iv). ☐ (iii) et (iv). ☒ (ii) et (iv)

Q37) Considérons la régression en séries temporelles suivante, effectuée pour chaque fond

d'investissement j: $R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{mt} - R_{ft}) + \varepsilon_{jt}$

où R_{jt} , R_{mt} , R_{ft} sont respectivement les rendements de l'actif j, du portefeuille de marché, et de l'actif sans risque. Lorsqu'un fond d'investissement fait 'mieux que le marché', quelle condition doit être respectée?

- ☐ β_j est positif et statistiquement différent de zéro.
- ☐ α_j est positif et statistiquement supérieur à 1.
- ☒ α_j est positif et statistiquement différent de zéro.
- ☐ α_j est positif et statistiquement différent de zéro et β_j est positif et statistiquement différent de 1.

Q38) Supposons que 100 fonds d'investissement soient testés sur base de la régression de la question précédente. Si un test bilatéral à 5% est utilisé, que concluez-vous si 2 fonds font significativement 'mieux que le marché'?

- ☐ Le marché est inefficent.
- ☐ Le marché ne montre pas de preuve d'inefficience.
- ☐ On ne peut conclure car la valeur de la statistique t^* pour chaque firme j n'est pas donnée.
- ☐ On ne peut conclure car les coûts de transaction ne sont pas pris en compte.

QCM5 : Variables binaires

Q1) Un chercheur veut tester, à un seuil de 5%, s'il y a eu un effet « guerre en Irak » en 2003. Il utilise une dummy, D_t . Il communique les résultats suivants, en précisant que les ratios de Student sont

$$y_t = 2340,4 + 23,5x_{1t} + 0,3x_{2t} - 120,56D_t + e_t$$

entre parenthèses ($n=18$) : (4,5) (2,2) (2,9) (5,8)

Une erreur s'est-elle glissée dans les résultats?

☐ La constante doit être négative.

☐ Il n'y a pas d'erreur.

☐ La constante ne peut pas être significative.

☒ Le t-ratio de la dummy doit être négatif.

Q2) Sur base de la question précédente et de l'erreur éventuelle faite par le chercheur, peut-on effectivement parler d'un effet « guerre en Irak »?

☐ Non, le t^* ratio n'est pas supérieur à la valeur critique, égale à 6,66.

☐ Oui, le coefficient de la dummy est plus élevé que les deux autres coefficients relatifs aux variables explicatives.

☐ Non, le t^* ratio est inférieur à la valeur critique de -2,14.

☒ Oui, la valeur absolue du t^* ratio est supérieur à la valeur critique de 2,14.

Q3)

$$MAS_i = 8,5 + 0,3BACx_{1i} - 1,2D_i + e_i$$

(4,5) (7,1) (-2,3)

où $n = 60$; MAS = note moyenne obtenue à la fin de la maîtrise; BAC = note moyenne obtenue à la fin du bac universitaire; D = dummy lié au genre (1=homme). Quelle affirmation est vraie ?

☐ La constante représente la note qu'un homme recevrait à la fin de sa maîtrise s'il avait une note de zéro à la fin de son BAC.

☐ La constante représente la note qu'un homme recevrait à la fin de sa maîtrise s'il avait une note de 8,5 à la fin de son BAC.

☒ La constante représente la note qu'une femme recevrait à la fin de sa maîtrise si elle avait une note de zéro à la fin de son BAC.

☐ La constante représente la note qu'une femme recevrait à la fin de sa maîtrise si elle avait une note de 8,5 à la fin de son BAC.

Q4) Sur base de la question précédente, quelle affirmation est vraie?

☐ La variable BAC est la plus importante, à la fois sur les plans 'économique' et statistique.

☐ La dummy est la plus importante, à la fois sur les plans 'économique' et statistique.

☐ La dummy est la plus importante sur le plan statistique.

☒ La dummy est la plus importante sur le plan 'économique'.

$$TEST_i = 1028,10 - 45,09 female_i - 169,81 black_i + 62,31 female_i \cdot black_i + e_i$$

Q5) (6,29) (4,29) (2,3) (18,15)

Les écart-types sont donnés entre parenthèses (n = 4,137). Un chercheur veut tester si les deux termes 'black' et 'female.black' (composé de deux dummies) peuvent être supprimées du modèle. Il estime le modèle contraint sur base d'un sous-échantillon de 3,548 observations. Quelle affirmation est vraie ?

θ Il ne peut pas tester cette hypothèse car il doit garder le même échantillon pour comparer les modèles contraint et non-contraint.

θ Il ne peut pas tester cette hypothèse car elle inclut une dummy d'interaction.

θ Il ne peut pas tester cette hypothèse car il doit effectuer le test sur les trois dummies simultanément.

θ Cela n'a aucun intérêt puisque les deux dummies sont individuellement non-significatives

Q6) Sur base de la question précédente, quelle est l'estimation de la différence de résultat au test entre les femmes 'noires' et les femmes 'blanches'?

θ -107.50

θ -169.81

θ 62.31

θ -152,09

Q7) Sur base de la question 5, que pourriez-vous faire pour tester si la différence de résultat au test entre femmes 'noires' et 'blanches' est significative ?

θ Effectuer un F-test sur les coefficients de female et black.

θ Tester, sur base d'un F-test, si les coefficients de female, black, female.black sont simultanément nuls.

θ Créer une dummy représentant les femmes 'noires', de telle sorte que: $TEST = \alpha + \beta \text{ femblack}_i + \varepsilon_i$

θ En considérant les femmes 'blanches' comme groupe de référence, créer trois dummies représentant les trois autres combinaisons possibles entre sexe et appartenance ethnique, de telle sorte que : $TEST_i = \alpha + \beta \text{ femblack}_i + \delta \text{ menblack}_i + \gamma \text{ menwhite}_i + \varepsilon_i$

Q8) $\log(\text{wage})_i = \beta_0 + \beta_1 \text{ train}_i + u_i$

où train est une dummy (1, si le travailleur participe à un programme de remise à niveau ; 0, autrement). Imaginer que le terme d'erreur inclut la variable ability (aptitude innée du travailleur, impossible à observer). Si les travailleurs qui ont une moins grande aptitude innée participent davantage au programme de remise à niveau, que pouvons-nous dire de l'estimateur des MCO de β_1 ?

θ Il est sans biais car la corrélation entre ability et $\log(\text{wage})$ est nulle.

θ Il est biaisé 'vers le haut' (surévalué ou 'upward biased').

θ Il est biaisé 'vers le bas' (sousévalué ou 'downward biased').

θ Il n'y a pas de données suffisantes.

Q9) Lequel des modèles suivants permet de tester si l'emploi de drogue a un effet sur le salaire, tout en autorisant cet effet à varier selon le sexe de l'individu (drug = nombre de joints fumés par mois)?

- ☐ $\log(\text{wage})_i = \beta_0 + \beta_1 \text{ drug}_i + \beta_3 \text{ female}_i + u_i$
- ☐ $\log(\text{wage})_i = \beta_0 + \beta_1 \text{ drug}_i + \beta_2 \text{ drug}_i \cdot \text{female}_i + u_i$
- ☐ $\log(\text{wage})_i = \beta_0 + \beta_1 \text{ female}_i + \beta_2 \text{ drug}_i \cdot \text{female}_i + u_i$
- ☐ $\log(\text{wage})_i = \beta_0 + \beta_1 \text{ drug}_i \cdot \text{female}_i + u_i$

Q10) Le terme d'interaction inclus dans la régression de la question précédente:

θ permet de faire varier le point d'intersection de la droite de régression de $\log(\text{wage})$ sur drug .

θ correspond à l'interaction entre deux dummies.

θ ne sera pas significatif car ses deux composantes sont corrélées.

θ permet de faire varier la pente de la droite de régression de $\log(\text{wage})$ sur drug .

Q11) Le modèle de probabilité linéaire:

θ ne peut pas inclure de variable binaire comme variable explicative.

θ ne peut pas être estimé par la technique des MCO.

θ délivre des estimateurs biaisés.

θ délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs proches de la moyenne.

Q12) Sur base de données journalières, un analyste financier spécifie le modèle suivant : $r_t = a_0 + a_1D1_t + a_2D2_t + a_3D3_t + a_4D4_t + \varepsilon_t$ (r_t = rendements; $D1 / D2 / D3 / D4 = 1$ pour les lundis / mardis / mercredis / jeudis, 0 autrement). Quelle est la signification de la constante ?

θ Il s'agit de la moyenne des rendements pour les vendredis.

θ Il s'agit de la moyenne des rendements pour les lundis.

θ Il s'agit de l'écart des rendements pour les vendredis par rapport à la moyenne des rendements pour la semaine.

θ Il s'agit de l'écart des rendements pour les lundis par rapport à la moyenne des rendements pour la semaine

Q13) Un analyste financier désire tester si les rendements enregistrés au cours de décembre ont un effet particulier sur le salaire des PDG. (Autrement dit, s'il existe un effet saisonnier / de calendrier lié aux rendements de décembre). Quel(s) modèle(s) peut-il estimer ($D = 1$ si décembre, 0 autrement) ?

i. $\text{salary}_t = a_0 + a_1D_t + a_2r_t + a_3D_t r_t + \varepsilon_t$

ii. $\text{salary}_t = a_0 + a_1D_t + a_2r_t + \varepsilon_t$

iii. $\text{salary}_t = a_0 + a_2r_t + a_3D_t r_t + \varepsilon_t$

iv. $\text{salary}_t = a_0 + a_2r_t + \varepsilon_t$

θ (i) seulement.

θ (i) et (ii) seulement.

θ (i), (ii) et (iii) seulement.

θ (i), (ii), (iii) et (iv).

QCM6 : Modèles de choix discret

Q1) Quelle(s) proposition(s) concernant les modèles logit et probit est (sont) vraie(s) ?

- i. Ils sont basés sur des fonctions grâce auxquelles les probabilités sont comprises entre 0 et 1
- ii. L'utilisation du modèle logit peut conduire à estimer un trop grand nombre de probabilités égales à 0 ou 1
- iii. Dans le modèle Probit, la fonction de répartition de la loi normale est identique à celle de la loi logistique
- iv. Contrairement au modèle probit, l'effet marginal d'une variable explicative dans le modèle logit est mesuré par l'estimation du paramètre de cette variable, multipliée par un « facteur d'échelle »

☒ (i) seulement. ☐ (i) et (ii) seulement. ☐ (i), (ii) et (iii) seulement. ☐ (i), (ii), (iii) et (iv).

Q2) En vous appuyant sur l'estimation suivante du modèle logit, déterminer la probabilité qu'un étudiant réussisse son année ($Y = 1$) si la note d'économétrie (X) est égale à 10 ?

$$\hat{P}(Y_i = 1 | X_i) = G(-1.49185 + 0.281554X_i)$$

☒ 79% ☐ 0.79% ☐ 17% ☐ 1.32%

```
logit_pchap <- plogis(-1.49185+0.281554*10)
logit_pchap
```

Q3) Vous estimez un modèle logit qui comprend une constante et deux variables explicatives. Les estimations des paramètres sont les suivantes: $\beta_{\text{chapeau 0}} = 0,1$, $\beta_{\text{chapeau 1}} = 0,3$, $\beta_{\text{chapeau 2}} = -0,2$. La moyenne des variables explicatives est: $\bar{X}_1 = 0.5$ et $\bar{X}_2 = 0.8$. Une augmentation de deux unités de X_2 conduira à une variation de la probabilité d'observer $Y = 1$, égale à environ:

- ☐ + 1.8, soit une augmentation de 180 pourcents
- ☐ 0.11, soit une hausse de 11 pourcents
- ☐ - 0.05, soit une baisse de 5 points de pourcentage
- ☒ - 0.10, soit une baisse de 10 points de pourcentage

Q4) A la question précédente, quelle méthode est utilisée pour estimer l'effet marginal de la variable explicative sur la probabilité d'observer $Y=1$?

- ☐ La méthode des moindres carrés
- ☒ L'« EMPM » ou « effet marginal (évalué) au point moyen »
- ☐ L'« EMM » ou « effet marginal moyen »
- ☐ La méthode « backward » ou « general-tospecific »

Q5) Pour la question 3, était-il possible d'estimer l'effet marginal de la variable explicative X_2 sur la probabilité d'observer $Y=1$ sans information sur les moyennes de variables ?

- ☐ Non, car il est impossible de calculer l'alternative, soit l'EMM
- ☒ Non, sauf en calculant l'EMM ou en estimant la fonction de répartition pour deux valeurs données de X_1 et X_2 .

- ☐ Oui, il est égal à -10 points de % également
- ☐ Oui, il est égal à -17 points de %

Q6) Considérez deux étudiants identiques à tous points de vue. La seule différence est que l'un a reçu la note de 10 sur 20 à son cours d'économétrie et que l'autre a obtenu 14. La moyenne des notes est égale à 11. En vous appuyant sur l'estimation suivante du modèle PROBIT, quelle est la différence exacte de probabilité de réussite de leur année universitaire ($Y = 1$) ?

$$\hat{P}(Y_i = 1 | X_i) = G(-1.49185 + 0.281554X_i)$$

```
#QUESTION 06
probit_pchap1 <- pnorm(-1.49185+0.281554*10)
probit_pchap2 <- pnorm(-1.49185+0.281554*14)
probit_pchap2-probit_pchap1
```

- ☐ 8.6p% ☐ 12.4p% ☐ 11p% ☐ 1.12p%

Q7) Sur la base du modèle spécifique dont les résultats de la régression sont affichés au slide suivant), quelle est l'estimation de la probabilité P pour un projet dont :

- ☐ Le pourcentage de réussite dans la catégorie du projet est égal à 64%
- ☐ Le pourcentage de l'objectif de récolte financé au terme de la première semaine est égal à 10%
- ☐ Le nombre de commentaires venant des internautes est égal à 10
- ☐ Le nombre d'actualités déposées par le porteur du projet est égal à 5

Calculez la probabilité et vérifiez si l'estimation de cette probabilité est absurde ou cohérente.

Oui, 2.62

```
Call:
lm(formula = SUCCES ~ POURC_SUCCES_CAT + POURC_FINANCE_1S + COMMENTAIRES)

Residuals:
    Min       1Q   Median       3Q      Max
-0.80280 -0.26176 -0.09451  0.29934  0.82192

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.236450   0.138270  -1.710  0.08935 .
POURC_SUCCES_CAT  0.008099   0.002811   2.882  0.00454 **
POURC_FINANCE_1S  0.010349   0.001517   6.820 2.16e-10 ***
COMMENTAIRES    0.022360   0.002891   7.735 1.47e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3546 on 148 degrees of freedom
(8 observations deleted due to missingness)
Multiple R-squared:  0.5071,    Adjusted R-squared:  0.4971
F-statistic: 50.76 on 3 and 148 DF,  p-value: < 2.2e-16
```

Q8) Utilisez les données FP.xlsx (décrites dans le fichier FP_DES.pdf) et le code CROWDFUNDING_LOGIT_PROBIT.R pour estimer un LOGIT visant à régresser la variable SUCCES sur les deux variables POURC_SUCCES_CAT et VIDEO. Quelle affirmation suivante est vraie ?

- ☐ Le pourcentage de prédictions correctes est égal à 33% et le pourcentage de "faux positifs" est égal à 24%.
- ☐ Le pourcentage de prédictions correctes est égal à 59% et le pourcentage de "faux positifs" est égal à 28%.
- ☐ Le pourcentage de prédictions correctes est 46 points de pourcentage plus élevé que le

pourcentage de prédictions erronées ; quant au pourcentage de "faux positifs, il est égal au pourcentage de "faux négatifs".

☐ Le pourcentage de prédictions correctes est égal à 41% et le pourcentage de "faux positifs" est égal à 39%

Q9) Sur base des mêmes données et code, utiliser un LOGIT pour régresser SUCCES sur POURC_SUCCES_CAT, VIDEO, POURC_FINANCE_1S, CONTRIBUTEURS_1S, ACTUS, COMMENTAIRES, et PARTAGES. Cela revient à ajouter des variables observables après le lancement de la campagne de financement au modèle LOGIT spécifique que nous avons étudié précédemment et qui ne contenait que des variables significatives, observables au moment du lancement de la campagne. Si vous suivez une approche du général au spécifique avec un seuil de 5%, quelle affirmation parmi les suivantes est vraie ?

☐ Le modèle final (ou spécifique) sera identique au modèle initial (ou général) puisque toutes les p-valeurs sont, dès le départ, supérieures à 5%.

☐ Le modèle final (ou spécifique) ne contiendra plus que trois variables : POURC_FINANCE_1S, COMMENTAIRES et POURC_SUCCES_CAT.

☐ Le modèle final (ou spécifique) ne contiendra que deux variables : POURC_FINANCE_1S et VIDEO.

☐ L'ordre de suppression des variables sera : PARTAGES, CONTRIBUTEURS_1S, VIDEO, ACTUS, POURC_SUCCES_CAT

Q10) Utilisez les mêmes données et code. À l'aide d'un LOGIT, régressez SUCCES sur POURC_FINANCE_1S et COMMENTAIRES. L'estimation de l'effet marginal moyen (EMM) de la variable COMMENTAIRES sur la probabilité de réussite d'une campagne de financement (typique) est égale à :

☐ 0,075, ce qui signifie que la probabilité de réussite d'une campagne de financement participatif augmente de 7,5% lorsque le nombre de commentaires augmente de 1%.

☐ 0,054, ce qui signifie que la probabilité de réussite d'une campagne de financement participatif augmente de 0,05 point de % lorsqu'il y a un commentaire supplémentaire laissé par un internaute.

☐ 0,280, ce qui signifie que la probabilité de réussite d'une campagne de financement participatif augmente de 28 points de % lorsqu'il y a un commentaire supplémentaire laissé par un internaute.

☐ 0,021, ce qui signifie que la probabilité de réussite d'une campagne de financement participatif augmente de 2,1 points de % lorsqu'il y a un commentaire supplémentaire laissé par un internaute.

Q11) Utilisez les mêmes données et code. À l'aide d'un LOGIT, régressez SUCCES sur POURC_FINANCE_1S et CONTRIBUTEURS_1S. Quel est l'effet marginal de la variable CONTRIBUTEURS_1S sur la probabilité de réussite d'une campagne de financement ATYPIQUE, dans le sens où les valeurs de POURC_FINANCE_1S et CONTRIBUTEURS_1S sont 50% plus élevées que leur moyenne dans l'échantillon, soit égales à 21 et 11 respectivement ?

☐ L'effet marginal sera égal à 0,10, soit une valeur toujours plus élevée que dans le cas d'une campagne de financement typique.

☐ L'effet marginal sera identique à l'effet marginal moyen (EMM), calculé précédemment pour une campagne typique.

☐ L'effet marginal sera égal à 0,011, soit une valeur toujours moins élevée que dans le cas d'une campagne de financement typique.

☐ L'effet marginal sera égal à 0,975, soit la moyenne des probabilités de réussite calculées à partir

de la fonction G que l'on doit évaluer lorsque CONTRIBUTEURS_1S = 11 et CONTRIBUTEURS_1S = 12, toutes choses égales par ailleurs.

Q12) Le modèle de probabilité linéaire :

- ☐ ne peut pas inclure de variable binaire comme variable explicative.
- ☐ ne peut pas être estimé par la technique des MCO.
- ☒ délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs proches de la moyenne.
- ☐ délivre des estimateurs biaisés des coefficients des variables explicatives.

Q13) Quelle(s) proposition(s) concernant les modèles logit et probit est (sont) FAUSSE(s) ?

- i. Ils sont basés sur des fonctions linéaires sophistiquées, grâce auxquelles les probabilités sont toujours comprises entre 0 et 1.
- ii. L'avantage du probit par rapport au logit est qu'il est plus précis lorsqu'il s'agit d'étudier des cas atypiques de la population.
- iii. L'utilisation du modèle LOGIT est particulièrement appropriée lorsque la variable Y affiche un grand nombre de valeurs égales à 0.
- iv. Contrairement au modèle PROBIT, l'effet marginal d'une variable explicative dans le modèle LOGIT est mesuré par l'estimation du paramètre de cette variable, multipliée par un « facteur d'échelle ».

- ☐ i and iii. ☐ ii and iv. ☐ i, iii, iv. ☒ i, ii, iii, iv.

Q14) Considérez cette fonction de régression de l'échantillon provenant d'un modèle LOGIT.

$$\hat{P}(Y_i = 1 | X_i) = G(-1.49185 + 0.281554X_i)$$

Pour répondre au QCM suivant, vous pouvez éventuellement vous servir de ce fichier LOGIT_QCM3.xlsx. Sinon, utilisez R : c'est encore plus facile ;-) Sur base de l'estimation du modèle LOGIT ci-dessus, quelle est la probabilité qu'un étudiant réussisse son année ($Y = 1$) si sa note d'économétrie (X) est égale à 10 ?

- ☒ 79% ☐ 17% ☐ 132% ☐ 1.32%

Q15) Utilisez la base de données PMTFM.xlsx et estimez la régression suivante en R à l'aide des trois modèles de choix binaire (MPL, LOGIT et PROBIT). Ce modèle vise à expliquer la participation au marché du travail des femmes mariées ou vivant en concubinage.

$$inlf_i = \beta_1 + \beta_2 nwifeinc_i + \beta_3 educ_i + \beta_4 exper_i + \beta_5 age_i + \beta_6 kidslt6_i + \varepsilon_i$$

où inlf ("in the labor force") = 1 si l'épouse ou la concubine a occupé un emploi durant l'année (ne fût-ce que quelques jours), nwifeinc (non-wife income) = revenus du ménage provenant de l'époux ou du concubin, educ = le nombre d'années d'instruction de l'épouse ou la concubine (à partir de la première année primaire incluse), exper = le nombre d'années d'expérience de l'épouse ou la concubine sur le marché du travail, age = l'âge de l'épouse ou la concubine, kidslt6 ("kids lower than 6 years old") = nombre d'enfants âgés de moins de 6 ans.

Vous désirez étudier les décisions prises par une catégorie bien particulière de femmes mariées ou vivant en concubinage. Vous considérez les valeurs suivantes : nwifeinc = 20, educ = 12, exper = 15, age = 30, kidslt6 = 3. Quelle(s) proposition(s) est (sont) vraie(s) ?

- i. Dans le MPL, si un quatrième enfant vient au monde, l'estimation de la probabilité de travailler, soit $P^*(inlf=1|X)$, sera égale à - 14,9%, toutes choses égales par ailleurs.
- ii. Dans le PROBIT, l'estimation de l'effet marginal lié à la naissance d'un quatrième enfant sur la probabilité d'être au travail, calculé par variation de la fonction G avec $\Delta kidslt6=1$ et $c6=3$, est égal à -9 points de pourcentage (p%) environ, contre environ -27,5 p% dans le MPL, toutes choses égales par ailleurs.
- iii. Dans le PROBIT, l'effet marginal calculé au point moyen (EMPM) d'un quatrième enfant est égal à environ -27 p%, toutes choses égales par ailleurs ; cette estimation pourrait convenir si l'analyse portait sur des femmes représentatives de l'échantillon.
- iv. L'utilisation des pseudo R-carrés et du pourcentage de prédictions correctes se justifie davantage dans le cas du PROBIT que du LOGIT, étant donné l'utilisation de la fonction de répartition de la loi Normale dans le cas du PROBIT.

☒ i et ii ☐ i, ii, iii, iv ☐ i and iii ☐ ii and iii

QCM7 : Autocorrélation

Q1) Les erreurs sont sériellement corrélées lorsque :

θ la variable dépendante est corrélée aux variables explicatives.

θ des variables explicatives sont corrélées entre elles.

θ la variable dépendante sous forme décalée est utilisée comme variable explicative.

θ les erreurs sont linéairement interdépendantes.

Q2) Supposez que vous ne teniez pas compte de la présence de la corrélation sérielle dans les erreurs. Quelle(s) en est (sont) la (les) conséquence(s) ?

(i) L'estimateur des MCO est biaisé.

(ii) L'estimateur des MCO est inconsistant.

(iii) L'estimateur n'est pas BLUE.

(iv) Les tests d'inférence statistique sont invalides.

θ (i), (ii), (iii) et (iv). θ (ii) et (iii) seulement. θ (iii) et (iv) seulement. θ (iii) seulement.

Q3) Supposez que vous ne teniez pas compte de la présence de la corrélation sérielle dans les erreurs. Quelle(s) en est (sont) la (les) conséquence(s) ?

(i) Les estimations des paramètres sont invalides.

(ii) Les tests d'hypothèses peuvent mener à des conclusions erronées.

(iii) Les écart-types estimés peuvent être inappropriés.

(iv) Les prévisions faites à partir du modèle sont biaisées.

θ (i), (ii), (iii) et (iv). θ (ii) et (iii) seulement. θ (iii) et (iv) seulement. θ (iii) seulement.

Q4) Sur un graphique représentant les erreurs au cours du temps, une autocorrélation négative est caractérisée par :

θ une séquence en cycle.

θ une séquence alternative.

θ l'absence de séquence.

θ des résidus proches de zéro.

Q5) Quel test peut être valablement utilisé pour tester la présence d'autocorrélation d'ordre 1 lorsque la variable dépendante sous forme décalée est utilisée comme variable explicative?

θ Test de Durbin-Watson.

θ AR(1) t-ratio test.

θ AR(1) t-ratio test de Durbin.

θ Test de Breusch-Pagan.

Q6) Quel test peut être utilisé pour tester la présence d'autocorrélation d'ordre 3 ?

θ Test de Breusch-Godfrey.

θ Test de Durbin-Watson augmenté

θ Test de Breusch-Pagan adapté à l'autocorrélation d'un ordre supérieur ou égal à 1.

θ Test de Yule-Walker.

Q7) Si les résidus souffrent d'une forte autocorrélation négative, quelle est la valeur la plus probable de la statistique de Durbin-Watson ?

θ Proche de 4. θ Proche de 2. θ Proche de 1. θ Proche de 0.

Q8) Si la statistique de Durbin-Watson (DW^*) est égale à 0, quelle sera la valeur approximative du coefficient d'autocorrélation d'ordre 1 ?

☒ +1. ☐ 2. ☐ -1. ☐ -1 ou +1.

Q9) Si la statistique de Durbin-Watson (DW^*) est égale à 0, quelle est la nature de l'autocorrélation ?

☐ Négative. ☒ Positive. ☐ Neutre. ☐ Impossible à déterminer.

Q10) Considérer le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$ ($t=1,2,\dots,50$). Un chercheur effectue le test de Durbin-Watson à un seuil de 5%. Si $DW^*=1.55$, quelle est la conclusion du chercheur?

☐ Les erreurs sont positivement corrélées.

☐ Les erreurs sont négativement corrélées.

☐ Les erreurs ne sont pas corrélées.

☒ On ne peut pas conclure.

Q11) Considérer le même modèle qu'à la question 10. Le chercheur dispose maintenant de 100 observations et obtient $DW^*=1.85$. A un seuil de 5%, quelle est sa conclusion?

☐ Les erreurs sont positivement corrélées.

☐ Les erreurs sont négativement corrélées.

☒ Les erreurs ne sont pas corrélées.

☐ On ne peut pas conclure.

Q12) Considérer le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$. Si vous désirez tester la présence d'autocorrélation dans les résidus sur base d'une régression auxiliaire, laquelle devez-vous utiliser parmi les suivantes ?

☐ $e_t^2 = b_0 + b_1 e_{t-1} + v_t$

☐ $e_t^2 = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{1t}^2 + b_4 x_{2t}^2 + b_5 x_{1t} x_{2t} + v_t$

☒ $e_t = b_0 + b_1 e_{t-1} + v_t$

☐ $e_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{1t}^2 + b_4 x_{2t}^2 + b_5 x_{1t} x_{2t} + v_t$

Q13) Considérer le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$. Si vous désirez tester la présence d'autocorrélation dans les résidus, tout en suspectant que l'exogénéité stricte ne tient pas, quelle régression auxiliaire devez-vous choisir parmi les suivantes ?

☐ $e_t = b_0 + b_1 e_{t-1} + v_t$

☐ $e_t = b_0 + b_1 e_{t-1} + b_2 e_{t-1}^2 + v_t$

☒ $e_t = b_0 + b_1 e_{t-1} + b_2 x_{1t} + b_3 x_{2t} + v_t$

☐ $e_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{1t}^2 + b_4 x_{2t}^2 + b_5 x_{1t} x_{2t} + v_t$

Q14) A quel test peut correspondre la régression auxiliaire que vous avez sélectionnée à la question précédente (en supposant que vous avez choisie la bonne) ?

☒ Au test de Breusch-Godfrey.

☐ Au AR(1) t-ratio test.

☐ Au test de Durbin-Watson.

☐ Au test de Yule-Walker.

Q15) Parmi les solutions suivantes, lesquelles peuvent remédier au problème de corrélation sérielle?

- (i) Prendre les variables en logarithme, plutôt qu'en niveau.
- (ii) Estimer un modèle en différence première (au lieu de l'estimer en niveau).
- (iii) Ajouter des variables sous forme décalée.
- (iv) Utiliser des dummies pour 'supprimer' l'effet lié aux observations 'aberrantes' (outliers).

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. θ (i) et (ii) seulement. θ (ii) et (iii) seulement.

Q16) Quelles conditions doivent être respectées pour que le test de Durbin-Watson soit valide ?

- (i) Les variables explicatives doivent être strictement exogènes.
- (ii) La régression doit contenir une constante.
- (iii) La régression ne peut contenir, comme variable explicative, la variable dépendante sous forme décalée.
- (iv) La régression ne peut contenir aucune variable explicative sous forme décalée.

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. θ (ii) et (iii) seulement. θ (i) et (ii) seulement.

Q17) Quelles caractéristiques peuvent introduire de l'autocorrélation dans les erreurs ?

- (i) Une lenteur de réaction de la variable dépendante aux changements de valeurs des variables explicatives.
- (ii) Une surréaction de la variable dépendante aux changements de valeurs des variables explicatives.
- (iii) Des variables explicatives, elles-même autocorrélées.
- (iv) La présence d'observations 'aberrantes' (outliers).

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. θ (ii) et (iii) seulement. θ (i) et (ii) seulement.

Q18) Si la variable dépendante est stationnaire mais hautement persistante, le fait de l'introduire sous forme décalée en tant que régresseur conduit à :

θ des estimateurs des MCO biaisés mais consistants.

θ des estimateurs des MCO biaisés et inconsistants.

θ des estimateurs des MCO sans biais mais inconsistants.

θ des estimateurs des MCO sans biais, consistants mais inefficients.

Q19) Si la variable dépendante est stationnaire et faiblement dépendante, le fait de l'introduire sous forme décalée en tant que régresseur :

θ ne conduit pas nécessairement à des estimateurs inconsistants.

θ ne conduit pas nécessairement à des estimateurs biaisés.

θ garantit l'efficacité des estimateurs.

θ supprime toute forme d'autocorrélation dans les erreurs.

Q20) Quel terme caractérise le processus de génération de données suivant ? $y_t = \mu + 1.5 y_{t-1} + \varepsilon_t$

θ Une marche aléatoire avec dérive.

θ Un processus stationnaire mais hautement persistant.

θ Un processus stationnaire avec tendance.

θ Un processus non-stationnaire et explosif.

Q21) Dans un processus stationnaire, les chocs :

θ se réduisent peu à peu, jusqu'à disparaître.

θ persistent indéfiniment.

θ croissent de manière exponentielle.

θ sont inexistants.

Q22) Quelle terme caractérise le processus de génération de données suivant ? $y_t = \mu + \lambda t + \varepsilon_t$

θ Une marche aléatoire avec dérive.

θ Un processus à racine unitaire.

θ Un processus stationnaire avec tendance (déterministe).

θ Un processus non-stationnaire et explosif.

Q23) Quel terme caractérise la plupart des séries temporelles relatives aux prix d'actifs financiers ?

θ Un processus 'i.i.d.' (distribué de manière identique et indépendante), complètement aléatoire.

θ Un marche aléatoire avec dérive.

θ Un processus stationnaire avec tendance (déterministe).

θ Un processus non-stationnaire et explosif.

Q24) Un processus z_t est faiblement dépendant si :

θ $\text{Corr}(z_t, z_{t+s}) \rightarrow 0$ lorsque $t \rightarrow \infty$.

θ $\text{Corr}(z_t, z_{t+s}) = \rho_s$, pour tout t .

θ $\text{Corr}(z_t, z_{t+s}) \rightarrow 0$ lorsque $s \rightarrow \infty$.

θ $\text{Corr}(z_t, z_{t+s}) = \rho_s$, pour tout s .

Q25) Imaginez un modèle en séries temporelles linéaire par rapport à ses paramètres, dont les variables explicatives ne sont pas parfaitement colinéaires. Les paramètres d'un tel modèle sont sans biais si, en outre, :

θ les variables explicatives sont strictement exogènes.

θ les variables explicatives sont strictement exogènes, et les erreurs ne sont pas sériellement corrélées.

θ les variables explicatives sont strictement exogènes, et les erreurs ne sont pas sériellement corrélées, de manière stricte.

θ les erreurs ne sont pas sériellement corrélées, de manière stricte.

Q26) Outre la linéarité du modèle par rapport aux paramètres et l'absence de colinéarité parfaite, les paramètres d'un modèle en séries temporelles sont consistants si :

θ les variables sont contemporanément exogènes.

θ les variables sont stationnaires et faiblement persistantes.

θ les variables sont contemporanément exogènes et stationnaires.

θ les variables sont strictement exogènes.

QCM8 : Hétéroscédasticité

Q1) Les erreurs sont homoscédastiques si :

θ leur variance ne varie pas en fonction des valeurs prises par les variables explicatives (au cours du temps).

θ elles sont nulle en moyenne.

θ elles sont normalement distribuées.

θ elles ne souffrent pas de corrélation sérielle.

Q2) Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?

θ Nullité de l'erreur (conditionnelle) en moyenne.

θ Absence de colinéarité parfaite.

θ Homoscédasticité des erreurs.

θ Linéarité du modèle par rapport à ses paramètres.

Q3) Parmi les propositions suivantes, laquelle est FAUSSE? Un estimateur efficient :

θ N'est PAS nécessairement consistant.

θ est nécessairement sans biais.

θ est nécessairement à variance minimale, par rapport n'importe quel autre estimateur linéaire.

θ N'est PAS nécessairement BLUE.

Q4) Quel est le lien entre un estimateur des MCO efficient et un estimateur des MCO sans biais ?

θ Absence de biais \Rightarrow efficience.

θ Efficience \Rightarrow absence de biais.

θ Efficience \Leftrightarrow absence de biais.

θ Aucune des affirmations précédentes n'est vraie.

Q5) Quelle est le lien entre l'efficience et la consistance d'un estimateur des MCO ?

θ Efficience \Rightarrow consistance.

θ Consistance \Rightarrow efficience.

θ Consistance \Leftrightarrow efficience.

θ Aucune des affirmations précédentes n'est vraie

Q6) Parmi les propositions suivantes, laquelle est FAUSSE? Un estimateur BLUE :

θ est nécessairement sans biais.

θ est nécessairement à variance minimale, par rapport aux autres estimateurs linéaires sans biais.

θ est nécessairement consistant.

θ est nécessairement efficient.

Q7) Parmi les propositions suivantes, laquelle est FAUSSE? La propriété BLUE requiert :

θ l'absence de biais.

θ l'homoscédasticité des erreurs.

θ la linéarité du modèle par rapport à ses paramètres.

θ la normalité des erreurs.

Q8) Quel est le lien entre la consistance et la propriété BLUE de l'estimateur des MCO ?

θ Consistance \Rightarrow BLUE.

θ BLUE \Rightarrow consistance.

θ BLUE \Leftrightarrow consistance.

θ Aucune des affirmations précédentes n'est vraie

Q9) Sur base de la théorie asymptotique, la validité des tests statistiques requiert :

θ des estimateurs sans biais.

θ l'absence de multicollinéarité entre les variables explicatives.

θ la normalité des erreurs.

θ l'homoscédasticité des erreurs.

Q10) Considérer le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$. Un chercheur veut effectuer un test de White. Quelle est la régression auxiliaire la plus appropriée ?

☐ $e_t^2 = b_0 + b_1 e_{t-1} + v_t$

☒ $e_t^2 = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{1t}^2 + b_4 x_{2t}^2 + b_5 x_{1t} x_{2t} + v_t$

☐ $e_t = b_0 + b_1 e_{t-1} + v_t$

☐ $e_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{1t}^2 + b_4 x_{2t}^2 + b_5 x_{1t} x_{2t} + v_t$

Q11) Sur base du modèle précédent, estimé sur 96 observations, quelle serait la valeur CRITIQUE à utiliser dans la table de Fisher-Snedecor pour effectuer le test de White à un seuil de 5%?

θ 3.10 θ 2.71 θ 2.47 θ 2.32

Q12) Sur base du modèle précédent, estimé sur 96 observations, quelle serait la valeur CRITIQUE à utiliser dans la table du Chi-carré pour effectuer le test de White à un seuil de 5% ?

θ 11.07 θ 9.49 θ 7.81 θ 5.99

Q13) Supposez que vous ne teniez pas compte de la présence d'hétéroscédasticité dans les erreurs. Quelle(s) en est (sont) la (les) conséquence(s) ?

(i) L'estimateur des MCO est biaisé.

(ii) L'estimateur des MCO est inconsistant.

(iii) L'estimateur n'est pas BLUE.

(iv) Les tests d'inférence statistique sont invalides.

θ (i), (ii), (iii) et (iv). θ (ii) et (iii) seulement. θ (iii) et (iv) seulement. θ (iii) seulement.

Q14) Supposez que vous ne teniez pas compte de la présence d'hétéroscédasticité dans les erreurs. Quelle(s) en est (sont) la (les) conséquence(s) ?

(i) Les estimations des paramètres sont invalides.

(ii) Les tests d'hypothèses peuvent mener à des conclusions erronées.

(iii) Les écart-types estimés peuvent être inappropriés.

(iv) Les prévisions faites à partir du modèle sont biaisées.

θ (i), (ii), (iii) et (iv). θ (ii) et (iii) seulement. θ (iii) et (iv) seulement. θ (iii) seulement.

Q15) Parmi les solutions suivantes, lesquelles peuvent remédier au problème d'hétéroscédasticité ?

(i) Prendre les variables en logarithme, plutôt qu'en niveau.

(ii) Utiliser des écart-types ajustés de manière appropriée (robustes).

(iii) Utiliser la méthode des moindres carrés généralisés faisables.

(iv) Ajouter la variable dépendante sous forme décalée en tant que variable explicative.

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. θ (ii) et (iii) seulement. θ (i) et (ii) seulement.

Q16) Parmi les tests suivants, lesquels sont des tests d'hétéroscédasticité ?

- (i) Test de White.
- (ii) Test de Breusch-Pagan.
- (iii) Méthode des écart-types robustes de White.
- (iv) Test de Breusch-Godfrey.

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. θ (ii) et (iii) seulement. **θ (i) et (ii) seulement.**

Q17) Pour autant que l'on dispose d'un nombre suffisamment élevé de données, on peut utiliser les écart-types robustes (HCSE) :

- (i) car ils délivrent des estimations sans biais et consistantes des valeurs 'vraies' des variances-covariances des paramètres.
- (ii) sans se préoccuper (nécessairement) de la présence d'hétéroscédasticité dans les erreurs.
- (iii) car ils sont asymptotiquement efficaces.
- (iv) sauf si les erreurs sont homoscedastiques.

θ (i), (ii), (iii) et (iv). θ (i), (ii) et (iii) seulement. **θ (ii) et (iii) seulement.** θ (i) et (ii) seulement.

Q18) Supposez que la variance de l'erreur d'un modèle est donnée par $V(\varepsilon_i) = \sigma^2 Z_i^8$, où Z_i est une variable (pouvant être une variable explicative du modèle). Pour rendre le terme d'erreur homoscedastique, il est nécessaire de :

θ diviser l'équation de la variance par Z_i^4 OU de diviser chaque observation du modèle par Z_i^8 (en veillant à remplacer la constante par une variable égale à $1/Z_i^8$)

θ diviser l'équation de la variance par Z_i^4 ET de diviser chaque observation du modèle par Z_i^8 (en veillant à remplacer la constante par une variable égale à $1/Z_i^8$) θ diviser l'équation de la variance par Z_i^4

θ diviser chaque observation du modèle par Z_i^4 (en veillant à remplacer la constante par une variable égale à $1/Z_i^4$)

Q19) Supposez que la variance de l'erreur d'un modèle est donnée par $V(\varepsilon_i) = \sigma^2 Z_i^{0.5}$, où Z_i est une variable (pouvant être une variable explicative du modèle). Pour rendre le terme d'erreur homoscedastique, il est nécessaire de :

θ diviser chaque observation du modèle par $Z_i^{0.5}$ (en veillant à remplacer la constante par une variable égale à $1/Z_i^{0.5}$).

θ diviser chaque observation du modèle par $Z_i^{1/4}$ (en veillant à remplacer la constante par une variable égale à $1/Z_i^{0.25}$).

θ diviser chaque observation du modèle par Z_i^2 (en veillant à remplacer la constante par une variable égale à $1/Z_i^2$).

θ diviser chaque observation du modèle par Z_i^4 (en veillant à remplacer la constante par une variable égale à $1/Z_i^4$).

Q20) Considérer le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$. Un chercheur veut effectuer un test de Breusch-Pagan. Quelle est la régression auxiliaire la plus appropriée ?

- ☐ $e_t^2 = b_0 + b_1 e_{t-1} + v_t$
- ☐ $e_t^2 = b_0 + b_1 x_{1t} + b_2 x_{2t} + v_t$
- ☐ $e_t = b_0 + b_1 e_{t-1} + v_t$
- ☐ $e_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + v_t$

QCM9 : Selection

Q1) En tant qu'outil de sélection de modèles, quelles sont les désavantages du R^2 ajusté?

- (i) Le R^2 ajusté ne peut être utilisé que pour comparer des modèles nonemboîtés.
- (ii) Le R^2 ajusté conduit souvent à sélectionner des modèles qui ne sont pas parcimonieux.
- (iii) Le R^2 ne peut pas être utilisé pour comparer des modèles dont la variable dépendante est la même mais la (ou les) variable(s) explicative(s) diffère(nt).
- (iv) Le R^2 ajusté ne peut pas être utilisé, en tant que tel, pour comparer des modèles dont la variable expliquée est la même mais apparaît sous une forme différente.

☐ (ii) et (iv) seulement. ☐ (i) et (ii) seulement. ☐ (i), (ii) et (iii) seulement. ☐ (i), (ii), (iii) et (iv)

Q2) Par rapport à la méthode 'specific-to-general' améliorée (stepwise), quelle est un désavantage de l'approche 'general-to-specific' (backward) à la Hendry?

☐ Certaines variables peuvent être exclues dès la première étape, ce qui peut biaiser l'estimation des coefficients.

☐ Le modèle final peut manquer d'interprétation théorique.

☐ Le modèle peut être statistiquement inapproprié.

☐ Si le modèle est mal spécifié, toutes les étapes ultérieures seront invalides.

Q3) Placer dans le bon ordre les étapes suivantes dans le but d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Estimer le modèle.

(ii) Effectuer des tests d'hypothèse sur les coefficients en fonction de considérations théoriques. (iii)

Supprimer les variables statistiquement superflues.

(iv) Conduire des tests de diagnostic sur les résidus observés.

☐ (i), ensuite (ii), ensuite (iii), ensuite (iv).

☐ (i), ensuite (iv), ensuite (ii), ensuite (iii).

☐ (i), ensuite (iv), ensuite (iii), ensuite (ii).

☐ (i), ensuite (iii), ensuite (ii), ensuite (iv).

Q4) Un modèle surdimensionné (overparameterised) est un modèle qui:

☐ inclut trop de variables;

☐ inclut le moins de variables possible pour expliquer les données;

☐ est bien spécifié.

☐ ne s'ajuste pas assez aux données.

Q5) Un modèle parcimonieux est un modèle qui:

☐ inclut trop de variables;

☐ inclut le moins de variables possible pour expliquer les données;

☐ est bien spécifié.

☐ n'explique pas assez les données.

Q6) Parmi les tests suivants, le(s)quel(s) peu(ven)t servir à choisir la forme fonctionnelle la plus appropriée?

- i. Le RESET;
- ii. Le test de Chow fonctionnellement ajusté;
- iii. Le test de Davidson-MacKinnon.
- iv. Le test de Breusch-Godfrey.

θ (i) et (ii) seulement. θ (i) et (iii) seulement. θ (i), (ii) et (iii) seulement. θ (i), (ii), (iii) et (iv).

Q7) Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + \varepsilon$. Dans quel(s) cas ce choix est une erreur (en supposant que vous connaissez la FRP)?

- i. La FRP est $y = a_0 + a_1 x_1^2 + \varepsilon$
- ii. La FRP est $\log(y) = a_0 + a_1 \log(x_1) + \varepsilon$
- iii. La FRP est $y = a_0 + a_1 1/x_1 + \varepsilon$
- iv. La FRP est $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$.

- ☐ (i) seulement.
- ☐ (i) et (ii) seulement.
- ☐ (i), (ii) et (iii) seulement.
- ☐ (i), (ii), (iii) et (iv).

Q8) Un chercheur opte pour la forme fonctionnelle linéaire, $y = a_0 + a_1 x_1 + \varepsilon$. Dans quel(s) cas l'hypothèse nulle du RESET sera rejetée (en supposant que vous connaissez la FRP) ?

- i. La FRP est $y = a_0 + a_1 x_1^2 + \varepsilon$
- ii. La FRP est $\log(y) = a_0 + a_1 \log(x_1) + \varepsilon$
- iii. La FRP est $y = a_0 + a_1 1/x_1 + \varepsilon$
- iv. La FRP est $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$.

- ☐ (i) seulement.
- ☐ (i) et (ii) seulement.
- ☐ (i), (ii) et (iii) seulement.
- ☐ (i), (ii), (iii) et (iv).

Q9) Un chercheur opte pour la forme fonctionnelle linéaire, $y_0 = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Parmi les modèles suivants, lequel correspond le mieux au modèle non-constraint du RESET?

- ☐ $y_1 = a_0 + a_1 x_1 + a_2 x_2 + a_3 \hat{y}_0^2 + a_4 \hat{y}_0^3 + v$
- ☐ $y_1^2 = a_0 + a_1 x_1 + a_2 x_2 + a_3 \hat{y}_0^2 + a_4 \hat{y}_0^3 + v$
- ☐ $y_0^2 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_1^3 + a_5 x_1 x_2 + v$
- ☐ $y_0 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_1^3 + a_5 x_1 x_2 + v$

Q10) Un chercheur vérifie la stabilité structurelle du modèle suivant : $y = a_0 + a_1 x_{t1} + a_1 x_{t2} + \varepsilon$. Le nombre total d'observations est égal à 200. Deux sous-périodes sont définies : la première comprend les 100 premières observations et la seconde comprend les 100 dernières. Quelle sera la SCR du modèle non contraint?

☐ La SCR obtenue à partir du modèle estimé sur toutes les données.

☐ La SCR obtenue à partir du modèle estimé sur la première souspériode.

☐ La SCR obtenue à partir du modèle estimé sur la seconde souspériode.

☐ La somme des SCR obtenues à partir des deux modèles estimés sur chacune des deux sous-périodes.

Q11) Sur base de la question 10, quelle sera la statistique de Fisher (F^*) du test de Chow si les SCR des trois régressions sont 156.4, 76.2, et 61.9 ?

☐ 4.3 ☐ 7.6 ☐ 5.3 ☐ 8.6

Q12) Sur base des questions 10 et 11, quelle est la valeur critique à utiliser sachant que le seuil de confiance est de 5%?

☐ 2.6 ☐ 8.5 ☐ 1.3 ☐ 9.2

Test n° 2

Q1) Considérez le FRE suivant : (profit chapeau) = $-26.95 + 0.27$ (salary chapeau).

La variable dépendante est mesurée en millions de dollar et la variable indépendante est mesurée en milliers de dollar. Si le salaire d'un PDG d'une entreprise cotée aux USA augmente de 1M\$, l'estimation du profit :

- Augm de 0.27 millions \$
- Augm de 270 millions \$
- Augm de 270 mille \$
- Augm de 253.05 milliers \$

Q2) Considérez le FRE suivant : (profit chapeau) = $-26.95 + 0.27$ (salary chapeau).

La variable dépendante est mesurée en millions de dollar et la variable indépendante est mesurée en milliers de dollar. L'ordonnée à l'origine signifie que :

- Le profit moyen des entreprises dans l'échantillon est négatif et égal à 27millions de \$
- Le modèle est absurde puisque aucun PDG ne gagne un salaire égal à -26.95 milliers \$
- La perte est égale à 27million de \$ pour des entreprises dont le PDG ne gagne aucun revenu
- La perte est égale à 27millions \$ pour des entreprises dont le PDG gagne 100 fois moins

Q3) Considérez la FRE suivante : (salary chapeau) = $747 + 0.57$ (profit chapeau).

La variable dépendante est mesurée en milliers de dollar et la variable indépendante est mesurée en millions de dollar. L'ordonnée à l'origine signifie que :

- Le modèle est absurde puisqu'aucune entreprise ne peut compter sur un PDG qui accepterait de travailler pour 747\$
- Le salaire estimé d'un PDG est égal à 747 mille \$ à condition que son entreprise réalise un profit minimal de 0.57millions \$
- Le salaire estimé est égal à 747 mille \$ pour un PDG dont l'entreprise ne réalise aucun profit
- Le salaire estimé est égal à 747 millions de \$ pour un PDG dont l'entreprise réalise 0.57milliard de \$

Q4) Considérez la FRE suivante : (salary chapeau) = $747 + 0.57$ (profit chapeau).

La variable dépendante est mesurée en milliers de dollar et la variable indépendante est mesurée en millions de dollar. Si le profit d'une entreprise cotée aux USA augmente de 10M \$, le salaire annuel estimé du PDG :

- Augm de 0.57millier \$
- Augm de 5700\$
- Diminue de 5700000\$
- Augm de 5700000\$

Q5) Considérez le modèle de régression suivant : $Y = a + bX^g + e$, où a, b, et c sont des paramètres et e, le terme d'erreur.

- Tous les paramètres du modèle ne pourront pas être estimés à l'aide des MCO
- Ce modèle est incohérent car il y a plus de paramètres que de variables
- Ce modèle ne peut pas être estimé par les MCO car le paramètre g est utilisé comme exposant de la variable X
- Ce modèle est incohérent car les paramètres g et b ne pourront pas faire l'objet d'une estimation distincte

Q6) Dans lequel des modèles suivants, l'interprétation de l'ordonnée à l'origine est-elle la plus pertinente ?

- Le poids en fonction de la taille des individus
- La moyenne des notes obtenues à un examen cantonal en fonction du nombre d'étudiants par enseignant
- La demande de café en fonction de son prix
- Le taux de chômage en fonction le taux de croissance du PIB ajusté pour l'inflation

Q7) Considérons l'estimateur MCO de l'écart-type (ou de la variance) du coefficient de la pente de la droite de régression. Quelle(s) affirmations sont vrai ?

- Il est positivement relié à la variance des résidus
- Il est négativement relié à la variance de la variable explicative
- Il est négativement relié à la taille de l'échantillon
- Plus il est faible, plus le paramètre est estimé avec précision

Q8) Par rapport à la covariance, l'avantage de la corrélation est qu'elle :

- Est comprise entre 0 et 1
- Est sensible aux unités de mesure des variables
- Ne dépend pas des unités de mesure des variables
- Est nécessaire pour calculer le coefficient de régression de Y sur X

Q9) Par rapport au coefficient de corrélation entre 2 variables, le coefficient de régression :

- Est plus précis car il repose sur un modèle
- Est symétrique et ne présuppose aucune causalité particulière
- Est moins précis car il ne repose pas sur un modèle
- Est asymétrique et repose sur une hypothèse sur le sens de causalité

Q10) Si « var.equal=FALSE » est utilisé en R pour tester l'égalité entre deux moyennes, la variable BIN prend la valeur 1 lorsque :

- Rejette H_0 , soit les variances des 2 populations sont égales
- Il est impossible de conclure sans connaître le seuil de significativité
- Ne rejette pas H_0 , soit les variances des 2 populations sont égales
- Ne rejette pas H_0 , soit les moyennes dans les 2 populations sont égales

Test n° 3

Q1) Un étudiant m'a envoyé ce message : « j'ai importé le fichier CYTISE_CENTREE dans R et j'ai utilisé le code suivant : « `M1=lm(c$POURC_TB_SANTE~c$POURC_DIP_ENS_SUPP_C)` ». Puis, « `summary(M1)` ». j'ai obtenu ceci : »

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    30.46182    0.16625  183.23  <2e-16 ***
c$POURC_DIP_ENS_SUPP_C  0.46085    0.02566   17.96  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.691 on 260 degrees of freedom
Multiple R-squared:  0.5536,    Adjusted R-squared:  0.5519
F-statistic: 322.5 on 1 and 260 DF,  p-value: < 2.2e-16
```

Quelle est la bonne interprétation des résultats sur le plan statistique ?

- Le rejet des deux hypothèses nulles selon lesquelles chacun des paramètres du modèle est nul dans la population conduirait à commettre une erreur de type I dans plus de 0.1% des cas
- Il y a rejet de l'hypothèse nulle selon laquelle la pente est égale à zéro dans la population puisque les seuils de significativité (alpha) traditionnellement utilisés sont tous supérieurs à la p-valeur
- Il y a rejet de l'hypothèse nulle selon laquelle la constante est égale à zéro dans la population puisque l'estimation de la constante à 5% est supérieure à la p-valeur
- Il est impossible de conclure étant donné la forte différence entre les p-valeurs et les statistiques du Student empirique

Q2) Dans un modèle surdimensionné, il n'y a pas de biais d'omission car :

- Corrélation entre la variable superflue et la variable requise dans la FRP est nulle
- La multicollinéarité entre la variable superflue et la variable requise ne viole pas les hypothèses BLUE.
- L'estimateur de la variable superflue est égale à zéro dans la population
- Le coefficient de régression de la variable superflue sur la variable requise est égal à zéro dans la population

Q3) Un étudiant m'a envoyé ce message : « j'ai importé le fichier CYTISE_CENTREE dans R et j'ai utilisé le code suivant : « M1=lm(c\$POURC_TB_SANTE~c\$POURC_DIP_ENS_SUPP_C) ». Puis, « summary(M1) ». j'ai obtenu ceci : »

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    30.46182    0.16625   183.23  <2e-16 ***
c$POURC_DIP_ENS_SUPP_C  0.46085    0.02566    17.96  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.691 on 260 degrees of freedom
Multiple R-squared:  0.5536,    Adjusted R-squared:  0.5519
F-statistic: 322.5 on 1 and 260 DF,  p-value: < 2.2e-16
```

Quelle interprétation donnez-vous à l'ordonnée à l'origine dans ce modèle ?

- Il s'agit de l'estimation du % de personnes en très bonne santé pour des communes dans lesquels il n'y a pas de diplômés de l'enseignement supérieur, ceteris paribus.
- Il s'agit du % de personnes en très bonne santé observé en moyenne dans les communes reprises dans l'échantillon
- Elle n'en a aucune car une variable explicative ne peut pas prendre une valeur nulle dans un tel modèle
- Il s'agit de l'estimation du % de personnes en très bonne santé pour des communes dans lesquels le % de diplômés de l'enseignement supérieur est égal à la moyenne observée dans l'échantillon

Q4) La PRECISION avec laquelle les coefficients d'un modèle sont estimés sera :

- D'autant plus élevée que les variables explicatives prennent des valeurs éloignées de leur moyenne
- D'autant plus élevée que la variable dépendante affiche une variance élevée
- D'autant plus faible que le nombre d'observations augmente
- D'autant plus élevée que la variance de l'erreur est élevée

Q5) Un étudiant m'a envoyé ce message : « j'ai importé le fichier CYTISE_CENTREE dans R et j'ai utilisé le code suivant : « $M1 = \text{lm}(c\$POURC_TB_SANTE \sim c\$POURC_DIP_ENS_SUPP_C)$ ». Puis, « `summary(M1)` ». j'ai obtenu ceci : »

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   30.46182    0.16625   183.23  <2e-16 ***
c$POURC_DIP_ENS_SUP_C  0.46085    0.02566    17.96  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.691 on 260 degrees of freedom
Multiple R-squared:  0.5536,    Adjusted R-squared:  0.5519
F-statistic: 322.5 on 1 and 260 DF,  p-value: < 2.2e-16
```

Quelle interprétation donnez-vous à la pente de ce modèle ?

- L'élasticité entre le % de personnes en très bonne santé et le % de citoyens diplômés de l'enseignement supérieur augmente de 10 points (de %)
- Lorsque le % de citoyens diplômés de l'enseignement supérieur augmente de 0.46% dans une commune, alors le % attendu de personnes en très bonne santé progresse de 1% toutes choses égales par ailleurs
- Lorsque le % de citoyens diplômés de l'enseignement supérieur augmente de 1% dans une commune, alors le % attendu de personnes en très bonne santé progresse de 0.46% toutes choses égales par ailleurs
- Lorsque le % de citoyens diplômés de l'enseignement supérieur augmente de 10 points (de %) dans une commune, alors le % de personnes en très bonne santé progresse de 4.6 points (de %)

Q6) Vous régressez le pourcentage de personnes se déclarant en très bonne santé, soit `POURC_TB_SANTE`, sur le `LOG NEPERIEN` du pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur, soit `LOG(POURC_DIP_ENS_SUP)`, en tenant bien compte de toutes les autres variables explicatives importantes. L'estimation du coefficient est égale à 3,06. Quelle est la variation estimée du pourcentage de personnes se déclarant en très bonne santé si le pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur vient à doubler ?

- Lorsque le pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur double, le pourcentage de personnes se déclarant en très bonne santé augmente de 6.12 points de %, toutes choses égales par ailleurs
- Lorsque le pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur double, le pourcentage de personnes se déclarant en très bonne santé augmente de 3,06 points de %, toutes choses égales par ailleurs
- Il est impossible de le savoir sans connaître le pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur dans la commune en question
- Lorsque le pourcentage de personnes disposant d'un diplôme de l'enseignement supérieur double, le pourcentage de personnes se déclarant en très bonne santé augmente de 0.0612 points de %, toutes choses égales par ailleurs

Q7) Un estimateur efficace est désirable car :

- La précision avec laquelle il permet d'estimer les coefficients d'un modèle est la plus élevée que l'on puisse obtenir par rapport aux autres estimateurs
- L'incertitude avec laquelle il permet d'estimer les coefficients d'un modèle est très proche de zéro
- Son biais disparaît lorsque la taille de l'échantillon tend vers l'infini
- Il consomme moins de degrés de liberté en excluant les variables superflues

Q8) la multicolinéarité intervient :

- Lorsqu'une variable omise est corrélée avec la variable explicative présente dans le modèle
- Lorsque deux variables explicatives sont fortement corrélées entre elles
- Lorsque la variable dépendante et une variable explicative sont fortement corrélées
- Lorsqu'une variable superflue est présente dans le modèle

Q9) Vous estimez l'effet (propre/partiel/ceteris paribus) du niveau d'éducation (mesurée par les années d'études depuis le secondaire) sur le salaire. Vous avez pris en compte toutes les variables importantes pour expliquer le salaire, hormis les aptitudes intellectuelles des individus acquises dès la naissance (dûes à la chance). Dès lors, le biais d'omission introduit dans l'estimateur :

- De la variable « éducation » sera vraisemblablement négatif
- De la variable « éducation » sera vraisemblablement positif
- De la variable « aptitudes » sera vraisemblablement nul
- De la variable « aptitudes » sera vraisemblablement positif et très élevé

Q10) Le biais d'omission introduit dans l'estimateur d'un coefficient est égal à :

- La différence entre l'estimateur du coefficient dans le modèle sous-dimensionné et la covariance entre les variables incluse et omise de ce modèle
- La différence entre l'espérance de l'estimateur du coefficient dans le modèle sous-dimensionné et l'estimateur de ce coefficient dans le modèle correctement dimensionné
- L'estimateur de ce coefficient dans le modèle correctement dimensionné multiplié par le coefficient de régression de la variable présente dans le modèle sur celle qui a été omise
- L'estimateur du coefficient dans le modèle sousdimensionné plus le coefficient de régression de la variable incluse sur la variable omise

Test n° 4

Q1) Quelle valeur allez-vous obtenir en utilisant cette instruction en R ? `qf(0.95,2,173)`

3.05

Q2) Si vous rejetez H_0 à un seuil de significativité de 5%, quelle est la réponse la plus probable ?

- Vous rejetez H_0 à 1%
- Vous prenez le risque de commettre une erreur de type II dans 95% des cas
- La p-valeur est inférieure à 4%
- Vous rejetez H_0 à 10%

Q3) Considérons les deux modèles suivants :

$$\text{Modèle I: } y_t = \beta_0 + \beta_1 x_{1t} + u_t$$

$$\text{Modèle II: } y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \varepsilon_t$$

Lesquelles de ces affirmations sont correctes ?

- Le modèle II n'affichera jamais un R^2 inférieur à celui du modèle I
- Le modèle II n'affichera jamais un R^2 ajusté inférieur à celui du modèle I
- Les deux modèles auront un R^2 identique si le coefficient estimé de α_2 est égal à zéro dans la population
- Les deux modèles auront un R^2 ajusté identique si l'estimation de α_2 est égale à zéro dans la population

Q4) Laquelle des affirmations suivantes est toujours vraie concernant la somme du carré des résidus ?

- ☐ $SCR_{nc} > SCR_c$
- ☐ $SCR_{nc} \geq SCR_c$
- ☐ $SCR_c > SCR_{nc}$
- ☒ $SCR_c \geq SCR_{nc}$

Q5) Considérons les deux modèles suivants :

$$\text{Modèle I: } y_t = \beta_0 + \beta_1 x_{1t} + u_t$$

$$\text{Modèle II: } y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \varepsilon_t$$

Supposons que le modèle II affiche à la fois un R^2 plus élevé et un R^2 ajusté moins élevée que le modèle I. Laquelle de ces affirmations est la plus plausible ?

- Ce résultat est impossible puisqu'il y a moins de degrés de liberté dans le modèle II.
- L'estimation de α_2 est égale à zéro
- Les variables X_1 et X_2 sont fortement corrélées.
- L'estimation de α_2 est différente de zéro mais de manière non significative

Q6) Sur base de 100 observations, on obtient (avec les écart-types estimés entre parenthèse) :

$$\hat{y}_t = -1.3 - 0.3 x_t$$

(2.1) (0.1)

Testons, à un seuil de 5%, si la valeur 'vraie' du coefficient de la pente est égale ou plus grande que -1, contre l'hypothèse alternative qu'elle est inférieure à -1. Quelle est la conclusion ? (Attention : -1 et pas +1)

- Il est impossible de conclure
- H_0 est rejetée
- H_0 n'est pas rejetée
- Le test est erroné : il aurait dû être bilatéral

Q7) Une variable superflue :

- Affichera une p-value faible car elle n'est pas plausible
- Affichera un t^* élevé en valeur absolue sauf dans un test unilatéral
- Aura un coefficient dont la valeur apparaîtra faible relativement à l'imprécision avec laquelle il aura été estimé
- Affichera une p-value élevée car il y a un risque faible de commettre une erreur de type I (en rejetant H_0 selon laquelle la valeur du coefficient de cette variable est égale à zéro dans la population)

Q8) Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont toujours correctes ?

- L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est rejetée à un seuil de confiance de 10%
- Le t-ratio test indique que t^* (calculé en valeur absolue) est plus grand que la valeur critique (cad le 95^{ème} percentile) provenant de la distribution de t de Student
- La valeur absolue du coefficient estimé de cette variable est élevée
- Le coefficient estimé de cette variable affiche un écart-type petit

Q9) Laquelle des affirmations suivantes décrit le mieux la puissance d'un test ?

- C'est la probabilité de rejeter une hypothèse nulle alors qu'elle est, en réalité, vraie
- C'est la probabilité de ne pas rejeter une hypothèse nulle alors qu'elle est, en réalité, fausse
- C'est la probabilité de rejeter une hypothèse nulle qui est fausse
- C'est équivalent à la taille du test

Q10) Supposons que le test du coefficient soit conduit à un seuil de 5%. Quelles affirmations sont vraies ?

- Le seuil de significativité est égal à la taille du test
- Le seuil de significativité est égal à la puissance du test
- 2.5% de la distribution totale sera comprise dans chacune des deux zones de rejet, s'il s'agit d'un test bilatéral
- 5% de la distribution totale sera comprise dans la zone de rejet, s'il s'agit d'un test unilatéral, que cette zone soit située dans l'extrémité gauche ou droite de la fonction de densité

Test n° 5

Q1) Dans la régression ci-dessous, quelle est la différence de salaire entre :

- Des personnes dont la beauté physique est considérée comme standard, et
- Des personnes dont la beauté physique est considérée comme étant en dessous de la moyenne mais qui ont 5ans d'éducation en plus,

Toutes choses étant égales par ailleurs ?

```
278 'pour tout le monde'
279 RLM <- lm(bd$lwage ~ bd$belavg + bd$abvavg + bd$educ + bd$exper + bd$expersq + bd$black)
280 summary(RLM)
281
282 'pour les hommes'
283
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.314868	0.085930	3.66	0.00026 ***
bd\$belavg	-0.186020	0.046121	-4.03	0.0005831573 ***
bd\$abvavg	-0.012881	0.033382	-0.39	0.69967
bd\$educ	0.066812	0.005830	11.46	< 0.000000000000002 ***
bd\$exper	0.047501	0.004779	9.94	< 0.000000000000002 ***
bd\$expersq	-0.000680	0.000107	-6.33	0.00000000034 ***
bd\$black	-0.144617	0.056725	-2.55	0.01091 *

- Aucune des réponses proposées

- 48%
- 0.48%
- 0,48\$

Q2) Effectuer ce type d'opérations :

```
educadj = b$educ - mean(b$educ)
experadj = b$exper - mean(b$exper)
tenureadj = b$tenure - mean(b$tenure)
```

- Est utile lorsque la variable dépendante n'a pas de sens
- Est utilisée essentiellement lorsqu'il y a une variable explicative binaire
- Est utile lorsqu'une ou plusieurs variables explicatives sont incluses sous forme quadratique
- Peut permettre de diminuer l'imprécision avec laquelle la constante va être estimée

Q3) Sur base de données journalière, un analyste financier spécifie le modèle suivant :

$r_t = a_0 + a_1 D1_t + a_2 D2_t + a_3 D3_t + a_4 D4_t + \epsilon_t$ (r_t = rendement ; $D1/D2/D3/D4 = 1$ pour les lundis/mardis/mercredis/jeudis, 0 autrement). Quelle est la signification de la constante ?

- Il s'agit de l'écart des rendements pour les vendredis par rapport à la moyenne des rendements pour la semaine
- Il s'agit de la moyenne des rendements pour les lundis
- Il s'agit de l'écart des rendements pour les lundis par rapport à la moyenne des rendements pour la semaine
- Il s'agit de la moyenne des rendements pour les vendredis

$$Q4) \quad TEST_i = 1028,10 - 45,09 \text{female}_i - 169,81\text{black}_i + 62,31\text{female}_i.\text{black}_i + e_i$$

(6,29) (4,29) (2,3) (18,15)

Les écart-types sont donnés entre parenthèse (n=4,137). Un chercheur veut tester si les deux termes 'black' et 'female.black' (composé de deux dummies) peuvent être supprimées du modèle. Quelle affirmation est vraie ?

- Il peut les supprimer sans effectuer de test supplémentaire puisqu'elles sont non significatives
- Il ne peut pas tester cette hypothèse car elle inclut une dummy d'interaction
- Il ne peut pas tester cette hypothèse car il doit effectuer le test sur les trois dummies simultanément.
- **Il doit réaliser un test de Fisher mais on ne pourra pas de toute façon supprimer la dummy 'female'.**

$$Q5) \quad TEST_i = 1028,10 - 45,09 \text{female}_i - 169,81\text{black}_i + 62,31\text{female}_i.\text{black}_i + e_i$$

(6,29) (4,29) (2,3) (18,15)

Quelle est l'estimation de la différence de résultat au test entre les femmes 'noires' et les femmes 'blanche' ?

- 169.81
- **-107.50**
- -152.09
- 62.31

Q6) Dans la régression linéaire ci-dessous, quelle est la différence de probabilité de succès entre deux campagnes de financement participatif actives dans le secteur de la musique sachant que le seconde affiche :

- Un pourcentage de financement après 1 semaine de 10 points plus élevés,
- 6 commentaires en moins ?

```
Call:
lm(formula = SUCCES ~ POURC_SUCCES_CAT + POURC_FINANCE_1S + COMMENTAIRES)

Residuals:
    Min       1Q   Median       3Q      Max
-0.80280 -0.26176 -0.09451  0.29934  0.82192

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.236450   0.138270  -1.710  0.08935 .
POURC_SUCCES_CAT  0.008099   0.002811   2.882  0.00454 **
POURC_FINANCE_1S  0.010349   0.001517   6.820 2.16e-10 ***
COMMENTAIRES    0.022360   0.002891   7.735 1.47e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3546 on 148 degrees of freedom
(8 observations deleted due to missingness)
Multiple R-squared:  0.5071,    Adjusted R-squared:  0.4971
F-statistic: 50.76 on 3 and 148 DF,  p-value: < 2.2e-16
```

- 3 pourcents en moins, pour la seconde campagne
- **3 points de % en plus, pour la 1ere campagne**
- Aucune des réponses proposées
- Impossible de le calculer sans connaître le taux de succès dans la catégorie « musique »

Q7) Le modèle de probabilité linéaire :

- Délivre des estimateurs biaisés
- Délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs proches de la moyenne
- Ne peut être estimée par la technique des MCO
- Ne peut pas inclure de variable binaire comme variable explicative

Q8) Dans la régression ci-dessous, que signifie la constante ?

```
260 RLM <- lm(b$wage ~ b$female + b$educ + b$exper + b$tenure)
261 summary(RLM)
262
263 #C5EX1c
264 educadj = b$educ - mean(b$educ)
265
262:1 (Top Level) ↕
```

Console Terminal x Jobs x

/cloud/project/ ↗

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.5679	0.7246	-2.16	0.031 *
b\$female	-1.8109	0.2648	-6.84	0.00000000023 ***
b\$educ	0.5715	0.0493	11.58	< 0.0000000000002 ***
b\$exper	0.0254	0.0116	2.20	0.029 *
b\$tenure	0.1410	0.0212	6.66	0.00000000068 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- C'est le cas où la personne qui a un niveau d'éducation nulle, une expérience nulle et une ancienneté nulle est un homme, puisque le coefficient de $w_{\$female}$ est négatif
- C'est le salaire estimé pour des personnes du genre masculin, qui n'ont aucune éducation, aucune expérience et aucune ancienneté
- C'est le cas où la personne qui a un niveau d'éducation faible, une expérience faible et une ancienneté faible est un homme, puisque le coefficient de $w_{\$female}$ est négatif
- Elle n'a aucun sens puisque le coefficient de female est plus proche de -1 que de 0, d'autant plus qu'un niveau d'éducation égal à 0.57 est absurde

Examen Juin 2016 HD QCM :

1. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
 - a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. Un tel estimateur sera sans biais.

2. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est égale ou supérieure à 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
 - a. L'hypothèse nulle n'est pas rejetée.
 - b. L'hypothèse nulle est rejetée.
 - c. Il est impossible de conclure.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

3. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
 - a. L'hypothèse nulle est rejetée.
 - b. L'hypothèse nulle n'est pas rejetée.
 - c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

4. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
 - a. La probabilité de l'erreur de type I diminue.
 - b. La probabilité de l'erreur de type II augmente.
 - c. Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue, en valeur absolue.
 - d. L'hypothèse nulle sera rejetée moins souvent.

5. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

a. -0,05.

b. 0,03.

c. 0,01.

d. -0,03.

6. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
- 0,057.
 - 0,003.
 - 0,632.
 - 0,094.
7. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?
- 3,33.
 - 0,05.
 - 0,18.
 - 0,32.
8. Si une variable superflue est présente dans la régression,
- les estimateurs des MCO ne sont plus BLUE ;
 - la variance de l'erreur diminue ;
 - le risque de multicollinéarité est plus élevé ;
 - le nombre de degrés de liberté augmente.
9. Soit $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ ($i=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?
- On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.
10. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

- a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
- b. la variable omise est faiblement corrélée avec les variables incluses;
- c. la taille de l'échantillon est grande ;
- d. la variance des variables incluses est faible.

11. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

12. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / $DH = 1$ pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

13. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

14. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

- a. (i), ensuite (ii), ensuite (iii), ensuite (iv).
- b. (iv), ensuite (iii), ensuite (ii), ensuite (i).
- c. (iv), ensuite (ii), ensuite (iii), ensuite (i).
- d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

15. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :

(i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.

(ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.

(iii) $y = a_0 + a_1 x_1 + \varepsilon$.

(iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.

a. (i) seulement.

b. (i) et (ii) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

16. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$

a. 0,95.

b. 0,05.

c. 0,10.

d. Aucune des affirmations précédentes n'est vraie.

17. La multicolinéarité :

a. réduit les intervalles de confiance.

b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.

c. augmente la variance des paramètres estimés.

d. implique la violation d'une des hypothèses de Gauss-Markov.

18. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?

a. Nullité de l'erreur (conditionnelle) en moyenne.

b. Absence de colinéarité parfaite.

c. Homoscédasticité des erreurs.

d. Linéarité du modèle par rapport à ses paramètres.

19. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

(i) Il est positivement relié à la variance des résidus.

(ii) Il est positivement relié à la variance de x_i .

(iii) Il mesure le degré d'imprécision dans l'estimation du paramètre a_j .

(iv) Il est positivement relié à la taille de l'échantillon.

a. (i), (ii) et (iv) seulement.

b. (i) et (iii) seulement.

c. (ii) et (iv) seulement.

d. (ii) seulement.

20. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

- (i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.
- (ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.
- (iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.
- (iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

21. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

a. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

b. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

c. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

d. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

22. Le test spécifié à la question précédente peut être effectué à l'aide:

- a. de la distribution de Fisher-Snedecor.
- b. des distributions t de Student et de Fisher-Snedecor
- c. de la distribution t de Student.
- d. impossible à déterminer.

23. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

- (i) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$
- (ii) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$
- (iii) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.
- (iv) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

a. (i) et (ii).

b. (ii) et (iii).

c. (iii)

d. (iv)

24. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :

- a. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
- b. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.
- c. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
- d. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée

25. De quelles données s'agit-il ?

- a. De séries temporelles
- b. De données de panel
- c. De données empilées
- d. De données en coupe transversale.

pays	années	PIB_PPA	LN_PIB_PPA	AJRSet	malfal94	latitude
1	2010	12241.44	9.41	4.36	0	27.90
2	2010	6904.56	8.84	5.63	1	12.60
3	2010	39048.15	10.57	2.15	0	27.00
4	2010	2409.55	7.79	4.27	0.158	24.30
5	2011	4716.98	8.46	4.94	0.74	16.20
6	2011	16959.84	9.74	4.36	0.0704	8.10
7	2011	17644.98	9.78	4.26	0	33.30
8	2011	49803.49	10.82	2.71	0	37.80
9	2012	1334.10	7.20	5.63	1	0.90
10	2012	712.01	6.57	5.48	1	0.00
11	2012	10799.95	9.29	4.14	0	34.20
12	2012	29609.11	10.30	4.44	0	10.80

26. Laquelle de ces propositions est fausse ? Une expérience contrôlée randomisée :

- a. requiert deux groupes : un groupe de contrôle et un groupe de traitement.
- b. consiste à donner une chance égale à chaque sujet de l'expérience de faire partie d'un des deux groupes.
- c. permet précisément de ne pas devoir vérifier si les sujets qui prennent part à l'expérience sont similaires par ailleurs.
- d. permet de tester l'efficacité d'une décision politique ou économique, à condition que les deux groupes soient similaires par ailleurs.

27. Que mesure l'erreur de type II sachant que l'hypothèse nulle est celle d'innocence ?

- a. Le risque d'envoyer un innocent en prison.
- b. Le risque de laisser en liberté un coupable.
- c. Le risque d'envoyer un coupable en prison trop souvent.
- d. Le risque de laisser un innocent trop souvent en liberté.

Partie II

Analyse de résultats

I. Considérez que le porteur du projet a réussi sa campagne de financement lorsqu'il est parvenu à dépasser son objectif de collecte de fonds, soit « Ratio » = « atteint » (en euros) / « objectif » (en euros) ≥ 1 . Pour commencer à étudier le modèle de RLM, envisagez cinq variables explicatives :

1. l'importance de l'objectif fixé par le porteur de projet (en logarithme) ;
2. le nombre d'amis sur Facebook (en logarithme) ;
3. la possibilité de visionner une vidéo sur le projet (1 = oui ; 0 sinon) ;
4. le nombre d'actualités insérées par le porteur du projet sur la page web ;
5. le soin apporté à la description du projet (il s'agit d'un indicateur subjectif allant de 0 à 3, la valeur 3 représentant les projets les plus soignés).

Le modèle suivant est spécifié :

$$Ratio_i = \beta_0 + \beta_1 \ln objectif_i + \beta_2 \ln FBF_i + \beta_3 video_i + \beta_4 actu_i + \beta_5 soin_i + \varepsilon$$

Root MSE	0.50943	R-Square	0.2695
Dependent Mean	0.66492	Adj R-Sq	0.2302
Coeff Var	76.61502		

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	0.83336	0.41275	2.02	0.0464
Inobjectif	-0.14681	0.05227	-2.81	0.0061
lnFBF	0.06244	0.03397	1.84	0.0693
Video	0.24052	0.10809	2.23	0.0285

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Actu	0.03936	0.01041	3.78	0.0003
Soin	0.20810	0.09226	2.26	0.0264

(28) Considérez deux projets identiques, aux seules différences près que le second projet dispose d'une vidéo et d'un objectif plus élevé de 50%, soit 1000 euros de plus. Quel sera la différence estimée au niveau des ratios et des montants récoltés entre les deux projets ?

- Cette différence sera égale à 16,71 points de pourcentage en faveur du projet 2 mais il est impossible d'estimer la différence en euros.
- Cette différence sera égale à 16,71 points de pourcentage en faveur du projet 2, soit 1671,15 euros récoltés de plus pour le projet 2.
- Il est impossible de l'estimer étant donné les données du problème.
- Il est impossible de l'estimer étant donné que la différence estimée entre les deux ratios est inférieure à 1.

(29) Etant donné les résultats de votre régression sur les plans économique et statistique, que décidez-vous de faire ?

- Vous conservez le modèle tel quel, étant donné le R carré qui affiche une valeur suffisamment élevée.
- Vous ne conservez que les variables « Actu » et « Inobjectif » si vous décidez d'utiliser un seuil de significativité égal à 1% pour minimiser le risque de commettre une erreur de type I.
- Vous éliminez la variable « InFBF » car elle affiche une p-valeur supérieure à 5%, ce qui correspond à un seuil de significativité standard.**
- Vous éliminez la variable « Actu » car elle affiche un coefficient proche de 0 et ne revêt donc pas d'importance sur le plan économique.

II. Un vin chilien vient de sortir sur le marché. Le directeur commercial d'une grande chaîne spécialisée dans la vente de produits alcoolisés voudrait connaître l'espace d'exposition qu'il conviendrait de réserver à ce nouveau vin durant le premier mois de commercialisation. Le directeur commercial dispose d'un échantillon relatif aux ventes (« Sales », en euros) générées lors du premier mois pour 47 produits similaires et à l'espace d'exposition (« DisplayFeet », en mètre de rayon) qui leur a été réservé. Il sait aussi que les produits déjà commercialisés depuis plus d'un mois génèrent en moyenne 60 € par mois et par mètre de rayon.

(30) Compléter le tableau suivant en remplissant les cases vides sachant que la variable dépendante est le chiffre d'affaires (« sales »), en euros.

		0.815
Root MSE	41.30820	R-Square3
		0.811
Dependent Mean	268.13000	Adj R-Sq 2

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	83.56026	14.41344	A	<.0001
lDisplayFee t	Log en base e de DisplayFeet1	B		9.83391	14.10	<.0001

a. A = 1204,39 et B = 1,43.

b. A = 5,80 et B = 138,66

- c. A est impossible à calculer car le nombre de degrés de libertés n'est pas connu, et B est impossible à calculer car la FRP n'est pas connue.
- d. Aucune des réponses précédentes n'est vraie.

(31) Donner une interprétation économique au coefficient de 'Intercept'.

- a. Il s'agit du chiffre d'affaires moyen (en euros) attendu sur le premier mois de vente, tous produits confondus, lorsque l'espace d'exposition est égal à 1 mètre.
- b. Il s'agit du chiffre d'affaires (en euros) attendu lorsque l'espace d'exposition augmente de B/100 %.
- c. Ce coefficient n'a pas de sens économique car $\log(0)$ n'existe pas.
- d. Il s'agit des coûts fixes à supporter, c'est-à-dire lorsque l'espace d'exposition est égal à zéro.

(32) Déterminer l'espace d'exposition que le directeur commercial devrait consacrer à ce nouveau vin durant le premier mois.

- a. Environ 3 mètres.
- b. Environ 2 mètres.
- c. Impossible à calculer : l'espace d'exposition désiré par les clients n'est pas connu.
- d. Environ 139 mètres.

(33) Sur base des résultats ci-dessous, donner une interprétation économique du coefficient de INVDisF sachant que : $INVDisF = \frac{1}{DisplayFeet}$; $DisplayFeet = 10$, la dérivée de $1/X$ est égale à $-\frac{1}{X^2}$; la variable dépendante est « Sales », en euros.

		0.826
Root MSE	40.04298	R-Square5
		0.822
Dependent Mean	268.13000	Adj R-Sq 6

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept						
t	Intercept1	376	69522	9.43945	39.91	<.0001
INVDisF		1	-329.7042	122.51988	-14.64	

- Etant donné son signe négatif, ce coefficient n'a pas de sens et le modèle devrait être rejeté.
- Le coefficient traduit l'augmentation du chiffre d'affaires attendu lorsque l'espace d'exposition consacré au nouveau produit passe de 1 mètre à 2 mètres.
- Inclure l'inverse d'une variable explicative dans un modèle est un contre-sens : le logarithme népérien est toujours préférable.
- Le coefficient traduit la baisse du chiffre d'affaires attendu lorsque l'espace d'exposition augmente d'un mètre.

(34) Donner une indication de la p -valeur relative au coefficient de INVDisF dans la case vide.

- Il est impossible de déterminer la p -valeur sans l'utilisation d'un tableur.
- La p -valeur sera inférieure à -0,01, soit -1%.
- La p -valeur sera négative puisque le t^* est négatif.
- La p -valeur confirmera qu'il est approprié de rejeter l'hypothèse nulle selon laquelle la valeur du coefficient de INVDisF est nulle, à un seuil de 5% dans un test bilatéral.

III. Dans l'étude de Des Rosiers et Thériault (1996), la variable expliquée est le prix de location d'appartement (en dollars canadiens, 1990). L'échantillon contient 13,378 observations. Le résultat de leur étude est synthétisé de la manière suivante :

Variable	Coefficient	t-ratio	<i>A priori</i> sign expected
Intercept	282.21	56.09	+
LnAGE	-53.10	-59.71	-
NBROOMS	48.47	104.81	+
AREABYRM	3.97	29.99	+
ELEVATOR	88.51	45.04	+
BASEMENT	-15.90	-11.32	-
OUTPARK	7.17	7.07	+
INDPARK	73.76	31.25	+
NOLEASE	-16.99	-7.62	-
LnDISTCBD	5.84	4.60	-
SINGLPAR	-4.27	-38.88	-
DSHOPCNTR	-10.04	-5.97	-
VACDIFF1	0.29	5.98	-

Notes: Adjusted R^2 = 0.651; regression F -statistic = 2082.27. Source: Des Rosiers and Thériault (1996). Reprinted with permission of the American Real Estate Society.

Les variables explicatives comprennent : LnAGE (log de l'âge de l'appartement), NBROOMS (nombre de chambres à coucher), AREABYRM (surface moyenne en m², par pièce), ELEVATOR (1=ascenseur; 0=autrement); BASEMENT (1=au rez-de-chaussée; 0=autrement); OUT-/INDPARK (nombre de places de parking à l'extérieur / intérieur); NOLEASE (1=pas de bail; 0=autrement) ; LnDISTCBD (distance en miles par rapport au centre d'affaires le plus proche).

(35) Quelle interprétation donnez-vous au coefficient de LnDISTCBD ? L'estimation du coefficient de LnDISTCBD indique que :

- le loyer d'un appartement augmentera de 2,92% si la distance entre l'appartement et le centre d'affaires le plus proche augmente de 50 miles, *ceteris paribus* ;
- le loyer d'un appartement augmentera de 2,92 dollars canadiens si la distance entre l'appartement et le centre d'affaires le plus proche augmente de 50 %, *ceteris paribus* ;
- le loyer d'un appartement augmentera de 5,84 dollars canadiens si la distance entre l'appartement et le centre d'affaires le plus proche augmente de 1 mile , *ceteris paribus* ;
- le loyer d'un appartement augmentera de 5,84% si la distance entre l'appartement et le centre d'affaires le plus proche augmente de 1%, *ceteris paribus*.

(36) Deux appartements A et B ont chacun une superficie totale de 100 m² et quatre pièces. Il sont exactement identiques par ailleurs. Quelle est la variation attendue du prix si, en prévision de la naissance d'un enfant, une paroi est ajoutée dans l'appartement A ?

- La différence du loyer entre les deux appartements (*ceteris paribus*) sera égale à 48,47 dollars canadiens, en faveur de l'appartement A.
- La différence du loyer entre les deux appartements (*ceteris paribus*) sera égale à 310,83 dollars canadiens, en faveur de l'appartement A.
- La différence du loyer entre les deux appartements (*ceteris paribus*) sera égale à 28,62 dollars canadiens, en faveur de l'appartement A.
- La différence du loyer entre les deux appartements (*ceteris paribus*) sera égale à 330,68 dollars canadiens, en faveur de l'appartement A.

IV. Partons de l'équation estimée suivante.

$$\begin{aligned} \log(\text{salary}) &= 4.30 + .288 \log(\text{sales}) + .0167 \text{roe} - .226 \text{rosneg} \\ &\quad (0.29) \quad (.034) \quad (.0040) \quad (.109) \end{aligned}$$

$n = 209, R^2 = .297, \bar{R}^2 = .286.$

Salary = salaire perçu en 1990 par le PDG de la firme (milliers de dollar) ; *rosneg* = dummy égale à 1 si *ros* < 0 et égale à 0 si *ros* ≥ 0 (*ros* = rendement de l'action entre 1988 et 1990) ; *sales* = ventes effectuées en 1990 (millions de dollar) ; *roe* = moyenne du rendement sur fonds propres entre 1988 et 1990 (%). Les écart-types sont entre parenthèses.

(37) Quelle est l'interprétation du coefficient de *rosneg* sur le plan économique ?

- a. Une baisse de 1 point de pourcentage du rendement de l'action entre 1988 et 1990 conduit à une baisse attendue du salaire du PDG de 22,6 milliers de dollars.
- b. Une augmentation de 1 point de pourcentage du rendement de l'action entre 1988 et 1990 conduira à une baisse attendue du salaire de 0,226 % du salaire de PDG.
- c. Une entreprise dont le rendement de l'action a été négatif entre 1988 et 1990 verra le salaire de son PDG diminuer de 22,6 milliers de dollars.
- d. Une entreprise dont le rendement de l'action a été négatif entre 1988 et 1990 verra le salaire de son PDG diminuer de 22,6%.

(38) Quelle est l'importance de *rosneg* sur le plan statistique ?

- a. Le coefficient de cette variable n'est pas statistiquement différent de zéro dans la population, étant donné le faible écart entre le R^2 et le \bar{R}^2 . La variable est donc superflue.
- b. Le coefficient de cette variable ne peut pas être considérée comme différent de zéro dans la population, même à un seuil de 10% dans un test bilatéral. La variable est donc superflue.
- c. La significativité du coefficient de cette variable dans un test bilatéral implique que la variable *rosneg* est pertinente et ne doit pas être retirée du modèle, à un seuil de 5% et 10% mais pas à un seuil de 1%.
- d. L'hypothèse nulle selon laquelle la valeur du coefficient de cette variable dans la population est égale à zéro, est rejetée à un seuil de 1% mais pas à 5%. La variable est donc pertinente.

V. Soit $\log(wage) = a_0 + a_1 educ + a_2 exper + a_3 tenure + a_4 married + a_5 black + a_6 south + a_7 urban + \varepsilon$, où $wage$ = salaire mensuel en dollar ; $educ$ = année(s) d'éducation ; $exper$ = année(s) d'expérience de travail ; $tenure$ = ancienneté (avec l'employeur actuel) ; $married$ = 1 si marié(e) ; $black$ = 1 si d'origine africaine ; $south$ = 1 si vit dans le Sud ; $urban$ = 1 si vit dans une 'zone urbaine' ; $lwage = \ln(wage)$.

Dans la régression suivante, la variable 'interaction' permet de rendre l'effet causal du niveau d'éducation sur le salaire dépendant de 'la couleur de la peau'.

Dependent Variable: lwage						
Number of Observations Read			935			
		Parameter	Standard			
Variable	Label	DF	Estimate	Error	t Value	Pr > t
Intercept	Intercept	1	5.375	0.11470	46.86	<.0001
educ	educ	1	0.067	0.00643	10.44	<.0001
exper	exper	1	0.014	0.00319	4.33	<.0001
tenure	tenure	1	0.012	0.00245	4.81	<.0001
married	married	1	0.199	0.03905	5.09	<.0001
black	black	1	0.095	0.25540	0.37	0.7106
south	south	1	-0.089	0.02628	-3.40	0.0007
urban	urban	1	0.184	0.02695	6.82	<.0001
interaction	educ*black	1	-0.023	0.02018	-1.12	0.2626

(39) Quelle est la signification économique du coefficient de la variable 'interaction' ?

- Entre deux personnes qui partagent les mêmes caractéristiques par ailleurs (*ceteris paribus*), celle qui est afro-américaine verra son salaire diminuer de 2,3% pour toute année d'éducation supplémentaire par rapport à celle qui est caucasienne.
- Il mesure la discrimination raciale en % du salaire, quel que soit le niveau d'éducation.
- Il mesure le 'rendement de l'éducation' en %, quel que soit « l'origine ethnique ».
- Une telle variable n'a pas de sens, étant donné la multicolinéarité qu'elle introduit.

(40) Au vu de leur significativité statistique, que devrions-nous faire avec les variables 'interaction' et 'black'?

- Il faut les supprimer, car leur p -valeur est supérieure à 5%.

- b. Il faut les conserver, étant donné la significativité de la variable *educ* qui intervient dans le calcul de la variable 'interaction'.
- c. Il est impossible de le déterminer sans effectuer un test de Fisher impliquant les deux coefficients de ces variables.
- d. Il faut les supprimer, étant donné la forte multicolinéarité qu'elles introduisent.

QCM examen

AVRIL 2013

1.b / 2.a / 3.d / 4.a / 5.b / 6.c / 7.b / 8.c / 9.a / 10.c / 11.c / 12.c / 13.d / 14.b / 15.d / 16.d / 17.b / 18.c / 19.a / 20.b / 21.c / 22.a / 22.d / 23.c / 24.c / 25.c / 26.c / 27.c / 28.b / 29.c / 30.a

1. Laquelle des propositions suivantes ne correspond pas à la représentation algébrique de la fonction de régression de l'échantillon ? (u_i caractérise le terme d'erreur non-observé)
 - a. $y_i = \hat{\alpha} + \hat{\beta}x_i + \hat{u}_i$
 - b. $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i + u_i$
 - c. $y_i = \hat{\alpha} + \hat{\beta}x_i + (y_i - \hat{y}_i)$
 - d. $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$

2. Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle ? (u_i caractérise le terme d'erreur non-observé ; e représente l'exponentielle ; les lettres grecques indiquent les paramètres du modèle)
 - (i) $y_i = \alpha + \beta(z_i / x_i)$
 - (ii) $y_i = \alpha x_i^\beta z_i^\gamma u_i$
 - (iii) $y_i = e^\alpha x_i^\beta e^{u_i}$
 - (iv) $y_i = \alpha + \beta x_i^\delta + u_i$
 - a. (i), (ii) et (iii) seulement
 - b. (ii) et (iii) seulement
 - c. (iii) seulement
 - d. Tous

3. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?

- a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. **Un tel estimateur sera sans biais.**
4. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
- a. **L'hypothèse nulle n'est pas rejetée.**
 - b. L'hypothèse nulle est rejetée.
 - c. Il est impossible de conclure.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
5. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
- a. L'hypothèse nulle est rejetée.
 - b. **L'hypothèse nulle n'est pas rejetée.**
 - c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
6. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
- a. La probabilité de l'erreur de type I diminue.
 - b. La probabilité de l'erreur de type II augmente.
 - c. **Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue.**
 - d. L'hypothèse nulle sera rejetée moins souvent.
7. Supposons que la consommation de 100 ménages soit disponible pour l'année 2006.
- a. Données de panel.
 - b. **Données en coupe transversale.**
 - c. Données temporelles.
 - d. Impossible de conclure.

8. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- a. -0,05.
 - b. 0,03.
 - c. 0,01.**
 - d. -0,03.
9. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
- a. 0,057.**
 - b. 0,003.
 - c. 0,632.
 - d. 0,094.
10. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?
- a. 3,33.
 - b. 0,05.
 - c. 0,18.**
 - d. 0,32.
11. Si une variable superflue est présente dans la régression,
- a. les estimateurs des MCO ne sont plus BLUE ;
 - b. la variance de l'erreur diminue ;
 - c. le risque de multicollinéarité est plus élevé ;**
 - d. le nombre de degrés de liberté augmente.
12. Soit $y_t = \beta_0 + \beta_1x_{t1} + \beta_2x_{t2} + u_t$ ($t=1,2,...,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des

résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?

a. On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.

b. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.

c. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.

d. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.

13. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;

b. la variable omise est faiblement corrélée avec les variables incluses;

c. la taille de l'échantillon est grande ;

d. la variance des variables incluses est faible.

14. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

15. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / $DH = 1$ pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

16. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

17. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

a. (i), ensuite (ii), ensuite (iii), ensuite (iv).

b. (iv), ensuite (iii), ensuite (ii), ensuite (i).

c. (iv), ensuite (ii), ensuite (iii), ensuite (i).

d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

18. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :

(i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.

(ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.

(iii) $y = a_0 + a_1 x_1 + \varepsilon$.

(iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.

a. (i) seulement.

b. (i) et (ii) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

19. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. **0,95.**
 - b. 0,05.
 - c. 0,10.
 - d. Aucune des affirmations précédentes n'est vraie.

20. Une distribution leptokurtique :

- a. dispose d'extrémités plus épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- b. dispose d'extrémités plus épaisses et est plus 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.**
- c. dispose d'extrémités moins épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- d. dispose d'extrémités plus épaisses et est asymétrique.

21. La multicolinéarité :

- a. réduit les intervalles de confiance.
- b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
- c. augmente la variance des paramètres estimés.**
- d. implique la violation d'une des hypothèses de Gauss-Markov.

22. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?

- a. Nullité de l'erreur (conditionnelle) en moyenne.**
- b. Absence de colinéarité parfaite.
- c. Homoscédasticité des erreurs.
- d. Linéarité du modèle par rapport à ses paramètres.

23. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est positivement relié à la variance de x_i .
 - (iii) Il donne une mesure de précision de l'estimation du paramètre a_1 .
 - (iv) Il est positivement relié à la taille de l'échantillon.
- a. (i), (ii) et (iv) seulement.
 - b. (i) et (iii) seulement.
 - c. (ii) et (iv) seulement.
 - d. (ii) seulement.

24. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

25. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

e. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

f. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

g. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

h. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

26. Le test spécifié à la question précédente peut être effectué à l'aide:

e. de la distribution de Fisher-Snedecor.

f. des distributions t de Student et de Fisher-Snedecor

g. de la distribution t de Student.

h. impossible à déterminer.

27. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(v) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(vi) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(vii) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(viii) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

e. (i) et (ii).

f. (ii) et (iii).

g. (iii)

h. (iv)

28. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :
- e. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
 - f. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.
 - g. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
 - h. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée
29. Les données de cohorte:
- a. sont plus générales que les données de panel
 - b. regroupent plusieurs jeux de données en coupe transversale, qui ont été prélevées à différents moments temporels
 - c. impliquent que la taille de l'échantillon ainsi que les unités comprises dans l'échantillon ne varient pas au cours du temps.
 - d. portent sur des individus ou des entreprises mais pas sur des objets.
30. Quel est le lien entre convergence et la consistance d'un estimateur des MCO ?
- a. Consistance \Rightarrow convergence.
 - b. convergence \Leftrightarrow consistance.
 - c. Convergence \Rightarrow consistance.
 - d. Aucune des affirmations précédentes n'est vraie.

AOUT 2013

1.b / 2.a / 3.d / 4.a / 5.b / 6.c / 7.b / 8.c / 9.a / 10.c / 11.c / 12.c / 13.d / 14.b / 15.d / 16.d / 17.b / 18.c / 19.a / 20.b / 21.c / 22.a / 22.d / 23.c / 24.c / 25.c / 26.c / 27.c / 28.b / 29.c / 30.a

1. Laquelle des propositions suivantes ne correspond pas à la représentation algébrique de la fonction de régression de l'échantillon ? (u_t caractérise le terme d'erreur non-observé)
- a. $y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t$
 - b. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t + u_t$
 - c. $y_t = \hat{\alpha} + \hat{\beta}x_t + (y_t - \hat{y}_t)$
 - d. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$

2. Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle ? (u_i caractérise le terme d'erreur non-observé ; e représente l'exponentielle ; les lettres grecques indiquent les paramètres du modèle)
- (i) $y_i = \alpha + \beta(z_i / x_i)$
- (ii) $y_i = \alpha x_i^\beta z_i^\gamma u_i$
- (iii) $y_i = e^\alpha x_i^\beta e^{u_i}$
- (iv) $y_i = \alpha + \beta x_i^\delta + u_i$
- a. (i), (ii) et (iii) seulement
- b. (ii) et (iii) seulement
- c. (iii) seulement
- d. Tous
3. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
- a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
- b. l'estimateur est efficient
- c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
- d. Un tel estimateur sera sans biais.
4. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6 x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
- a. L'hypothèse nulle n'est pas rejetée.
- b. L'hypothèse nulle est rejetée.
- c. Il est impossible de conclure.
- d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
5. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
- a. L'hypothèse nulle est rejetée.
- b. L'hypothèse nulle n'est pas rejetée.
- c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.

- d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
6. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
- La probabilité de l'erreur de type I diminue.
 - La probabilité de l'erreur de type II augmente.
 - Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue.
 - L'hypothèse nulle sera rejetée moins souvent.
7. Supposons que la consommation de 100 ménages soit disponible pour l'année 2006.
- Données de panel.
 - Données en coupe transversale.
 - Données temporelles.
 - Impossible de conclure.

8. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- 0,05.
 - 0,03.
 - 0,01.
 - 0,03.
9. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
- 0,057.
 - 0,003.
 - 0,632.
 - 0,094.

10. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?
- 3,33.
 - 0,05.
 - 0,18.
 - 0,32.
11. Si une variable superflue est présente dans la régression,
- les estimateurs des MCO ne sont plus BLUE ;
 - la variance de l'erreur diminue ;
 - le risque de multicolinéarité est plus élevé ;
 - le nombre de degrés de liberté augmente.
12. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?
- On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.
13. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:
- le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
 - la variable omise est faiblement corrélée avec les variables incluses;
 - la taille de l'échantillon est grande ;
 - la variance des variables incluses est faible.

14. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?
- (i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.
 - (ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.
 - (iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.
 - (iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.
- a. (i) seulement.
 - b. (i), (ii) et (iv) seulement.
 - c. (i), (ii) et (iii) seulement.
 - d. (i), (ii), (iii) et (iv).
15. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :
 $v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / DH = 1 pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?
- a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.
 - b. Il s'agit de la moyenne des ventes sur toute l'année.
 - c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.
 - d. Aucune des réponses précédentes.
16. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :
- a. n'a pas de variable indicatrice comme variable dépendante.
 - b. ne peut pas être estimé par la technique des moindres carrés généralisés.
 - c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.
 - d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.
17. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.
- (i) Supprimer les variables statistiquement superflues.
 - (ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.
 - (iii) Conduire des tests de diagnostic sur les résidus observés.
 - (iv) Estimer le modèle.

- a. (i), ensuite (ii), ensuite (iii), ensuite (iv).
- b. (iv), ensuite (iii), ensuite (ii), ensuite (i).
- c. (iv), ensuite (ii), ensuite (iii), ensuite (i).
- d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

18. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :

- (i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.
 - (ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.
 - (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
 - (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.
- a. (i) seulement.
 - b. (i) et (ii) seulement.
 - c. (i), (ii) et (iii) seulement.
 - d. (i), (ii), (iii) et (iv).

19. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$

- a. 0,95.
- b. 0,05.
- c. 0,10.
- d. Aucune des affirmations précédentes n'est vraie.

20. Une distribution leptokurtique :

- a. dispose d'extrémités plus épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- b. dispose d'extrémités plus épaisses et est plus 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- c. dispose d'extrémités moins épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- d. dispose d'extrémités plus épaisses et est asymétrique.

21. La multicolinéarité :

- a. réduit les intervalles de confiance.
- b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
- c. augmente la variance des paramètres estimés.
- d. implique la violation d'une des hypothèses de Gauss-Markov.

22. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?

- a. Nullité de l'erreur (conditionnelle) en moyenne.
- b. Absence de colinéarité parfaite.
- c. Homoscédasticité des erreurs.
- d. Linéarité du modèle par rapport à ses paramètres.

23. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est positivement relié à la variance de x_i .
 - (iii) Il donne une mesure de précision de l'estimation du paramètre a_1 .
 - (iv) Il est positivement relié à la taille de l'échantillon.
- a. (i), (ii) et (iv) seulement.
 - b. (i) et (iii) seulement.
 - c. (ii) et (iv) seulement.
 - d. (ii) seulement.

24. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

25. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

i. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

j. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

k. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

l. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

26. Le test spécifié à la question précédente peut être effectué à l'aide:

i. de la distribution de Fisher-Snedecor.

j. des distributions t de Student et de Fisher-Snedecor

k. de la distribution t de Student.

l. impossible à déterminer.

27. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(ix) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(x) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(xi) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(xii) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

i. (i) et (ii).

j. (ii) et (iii).

k. (iii)

l. (iv)

28. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :
- i. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
 - j. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.
 - k. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
 - l. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée
29. Les données de cohorte:
- a. sont plus générales que les données de panel
 - b. regroupent plusieurs jeux de données en coupe transversale, qui ont été prélevées à différents moments temporels
 - c. impliquent que la taille de l'échantillon ainsi que les unités comprises dans l'échantillon ne varient pas au cours du temps.
 - d. portent sur des individus ou des entreprises mais pas sur des objets.
30. Quel est le lien entre convergence et la consistance d'un estimateur des MCO ?
- a. Consistance \Rightarrow convergence.
 - b. convergence \Leftrightarrow consistance.
 - c. Convergence \Rightarrow consistance.
 - d. Aucune des affirmations précédentes n'est vraie.

AVRIL 2016

1D/2A/3B/4C/5C/6A/7C/8C/9C/10D/11B/12D/13D/14B/15C/16A/17C/18A/19C/20C/21C/22C/23C/24B/25C/26C/27B/28A/29C/30B/31A/32B/33B/34D/65B/36C/37D/38C/39A/40C

28. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
- a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. Un tel estimateur sera sans biais.

29. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est égale ou supérieure à 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
- a. L'hypothèse nulle n'est pas rejetée.
- b. L'hypothèse nulle est rejetée.
- c. Il est impossible de conclure.
- d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
30. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
- a. L'hypothèse nulle est rejetée.
- b. L'hypothèse nulle n'est pas rejetée.
- c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
- d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
31. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
- a. La probabilité de l'erreur de type I diminue.
- b. La probabilité de l'erreur de type II augmente.
- c. Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue, en valeur absolue.
- d. L'hypothèse nulle sera rejetée moins souvent.

32. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- a. -0,05.
- b. 0,03.
- c. 0,01.
- d. -0,03.

33. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
- a. 0,057.
 - b. 0,003.
 - c. 0,632.
 - d. 0,094.
34. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?
- a. 3,33.
 - b. 0,05.
 - c. 0,18.
 - d. 0,32.
35. Si une variable superflue est présente dans la régression,
- a. les estimateurs des MCO ne sont plus BLUE ;
 - b. la variance de l'erreur diminue ;
 - c. le risque de multicolinéarité est plus élevé ;
 - d. le nombre de degrés de liberté augmente.
36. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?
- a. On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.
 - b. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
 - c. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.
 - d. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.
37. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

- a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
- b. la variable omise est faiblement corrélée avec les variables incluses;
- c. la taille de l'échantillon est grande ;
- d. la variance des variables incluses est faible.

38. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

39. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t \quad (v_t = \text{ventes}; DP / DA / DH = 1 \text{ pour le } 1^{\text{er}} \text{ trimestre} / 3^{\text{ème}} \text{ trimestre} / 4^{\text{ème}} \text{ trimestre}, 0 \text{ autrement}).$$

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

40. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

41. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

- a. (i), ensuite (ii), ensuite (iii), ensuite (iv).
- b. (iv), ensuite (iii), ensuite (ii), ensuite (i).
- c. (iv), ensuite (ii), ensuite (iii), ensuite (i).
- d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

42. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :
- (i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.
 - (ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.
 - (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
 - (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.
- a. (i) seulement.
 - b. (i) et (ii) seulement.
 - c. (i), (ii) et (iii) seulement.
 - d. (i), (ii), (iii) et (iv).
43. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. 0,95.
 - b. 0,05.
 - c. 0,10.
 - d. Aucune des affirmations précédentes n'est vraie.
44. La multicollinéarité :
- a. réduit les intervalles de confiance.
 - b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
 - c. augmente la variance des paramètres estimés.
 - d. implique la violation d'une des hypothèses de Gauss-Markov.
45. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?
- a. Nullité de l'erreur (conditionnelle) en moyenne.
 - b. Absence de colinéarité parfaite.
 - c. Homoscédasticité des erreurs.
 - d. Linéarité du modèle par rapport à ses paramètres.
46. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est positivement relié à la variance de x_i .
 - (iii) Il mesure le degré d'imprécision dans l'estimation du paramètre a_i .
 - (iv) Il est positivement relié à la taille de l'échantillon.
- a. (i), (ii) et (iv) seulement.
 - b. (i) et (iii) seulement.
 - c. (ii) et (iv) seulement.
 - d. (ii) seulement.

47. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

48. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

m. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

n. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

o. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

p. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

49. Le test spécifié à la question précédente peut être effectué à l'aide:

m. de la distribution de Fisher-Snedecor.

n. des distributions t de Student et de Fisher-Snedecor

o. de la distribution t de Student.

p. impossible à déterminer.

50. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(xiii) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(xiv) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(xv) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(xvi) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

m. (i) et (ii).

n. (ii) et (iii).

o. (iii)

p. (iv)

51. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :

- m. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
- n. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.
- o. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
- p. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée

52. De quelles données s'agit-il ?

- a. De séries temporelles
- b. De données de panel
- c. De données empilées
- d. De données en coupe transversale.

pays	années	PIB_PPA	LN_PIB_PPA	AJRSet	malfal94	latitude
1	2010	12241.44	9.41	4.36	0	27.90
2	2010	6904.56	8.84	5.63	1	12.60
3	2010	39048.15	10.57	2.15	0	27.00
4	2010	2409.55	7.79	4.27	0.158	24.30
5	2011	4716.98	8.46	4.94	0.74	16.20
6	2011	16959.84	9.74	4.36	0.0704	8.10
7	2011	17644.98	9.78	4.26	0	33.30
8	2011	49803.49	10.82	2.71	0	37.80
9	2012	1334.10	7.20	5.63	1	0.90
10	2012	712.01	6.57	5.48	1	0.00
11	2012	10799.95	9.29	4.14	0	34.20
12	2012	29609.11	10.30	4.44	0	10.80

53. Laquelle de ces propositions est fausse ? Une expérience contrôlée randomisée :

- e. requiert deux groupes : un groupe de contrôle et un groupe de traitement.
- f. consiste à donner une chance égale à chaque sujet de l'expérience de faire partie d'un des deux groupes.
- g. permet précisément de ne pas devoir vérifier si les sujets qui prennent part à l'expérience sont similaires par ailleurs.
- h. permet de tester l'efficacité d'une décision politique ou économique, à condition que les deux groupes soient similaires par ailleurs.

54. Que mesure l'erreur de type II sachant que l'hypothèse nulle est celle d'innocence ?

- a. Le risque d'envoyer un innocent en prison.
- b. Le risque de laisser en liberté un coupable.
- c. Le risque d'envoyer un coupable en prison trop souvent.
- d. Le risque de laisser un innocent trop souvent en liberté.

JUIN 2016

Réponses

1D/2A/3B/4C/5C/6A/7C/8C/9C/10D/11B/12D/13D/14B/15C/16A/17C/18A/19C/20C/21C/22C/23C/24B/25C/26C/27B/28A/29C/30B/31A/32B/33B/34D/65B/36C/37D/38C/39A/40C

1. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
 - a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. Un tel estimateur sera sans biais.

2. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6 x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est égale ou supérieure à 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
 - a. L'hypothèse nulle n'est pas rejetée.
 - b. L'hypothèse nulle est rejetée.
 - c. Il est impossible de conclure.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

3. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
 - a. L'hypothèse nulle est rejetée.
 - b. L'hypothèse nulle n'est pas rejetée.
 - c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

4. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
 - a. La probabilité de l'erreur de type I diminue.
 - b. La probabilité de l'erreur de type II augmente.
 - c. Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue, en valeur absolue.
 - d. L'hypothèse nulle sera rejetée moins souvent.

5. Sur base de $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- a. -0,05.
- b. 0,03.
- c. 0,01.
- d. -0,03.

6. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
- 0,057.
 - 0,003.
 - 0,632.
 - 0,094.
7. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?
- 3,33.
 - 0,05.
 - 0,18.
 - 0,32.
8. Si une variable superflue est présente dans la régression,
- les estimateurs des MCO ne sont plus BLUE ;
 - la variance de l'erreur diminue ;
 - le risque de multicolinéarité est plus élevé ;
 - le nombre de degrés de liberté augmente.
9. Soit $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ ($i=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?
- On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.
10. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

- a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
- b. la variable omise est faiblement corrélée avec les variables incluses;
- c. la taille de l'échantillon est grande ;
- d. la variance des variables incluses est faible.

11. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

12. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / $DH = 1$ pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

13. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

14. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

- a. (i), ensuite (ii), ensuite (iii), ensuite (iv).
- b. (iv), ensuite (iii), ensuite (ii), ensuite (i).
- c. (iv), ensuite (ii), ensuite (iii), ensuite (i).
- d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

15. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :
- (i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.
 - (ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.
 - (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
 - (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.
- a. (i) seulement.
 - b. (i) et (ii) seulement.
 - c. (i), (ii) et (iii) seulement.
 - d. (i), (ii), (iii) et (iv).
16. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. 0,95.
 - b. 0,05.
 - c. 0,10.
 - d. Aucune des affirmations précédentes n'est vraie.
17. La multicollinéarité :
- a. réduit les intervalles de confiance.
 - b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
 - c. augmente la variance des paramètres estimés.
 - d. implique la violation d'une des hypothèses de Gauss-Markov.
18. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?
- a. Nullité de l'erreur (conditionnelle) en moyenne.
 - b. Absence de colinéarité parfaite.
 - c. Homoscédasticité des erreurs.
 - d. Linéarité du modèle par rapport à ses paramètres.
19. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est positivement relié à la variance de x_i .
 - (iii) Il mesure le degré d'imprécision dans l'estimation du paramètre a_i .
 - (iv) Il est positivement relié à la taille de l'échantillon.
- a. (i), (ii) et (iv) seulement.
 - b. (i) et (iii) seulement.
 - c. (ii) et (iv) seulement.
 - d. (ii) seulement.

20. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

21. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

q. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

r. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

s. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

t. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

22. Le test spécifié à la question précédente peut être effectué à l'aide:

q. de la distribution de Fisher-Snedecor.

r. des distributions t de Student et de Fisher-Snedecor

s. de la distribution t de Student.

t. impossible à déterminer.

23. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(xvii) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(xviii) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(xix) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(xx) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

q. (i) et (ii).

r. (ii) et (iii).

s. (iii)

t. (iv)

24. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :

- q. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
- r. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.
- s. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
- t. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée

25. De quelles données s'agit-il ?

- a. De séries temporelles
- b. De données de panel
- c. De données empilées
- d. De données en coupe transversale.

pays	années	PIB_PPA	LN_PIB_PPA	AJRSet	malfal94	latitude
1	2010	12241.44	9.41	4.36	0	27.90
2	2010	6904.56	8.84	5.63	1	12.60
3	2010	39048.15	10.57	2.15	0	27.00
4	2010	2409.55	7.79	4.27	0.158	24.30
5	2011	4716.98	8.46	4.94	0.74	16.20
6	2011	16959.84	9.74	4.36	0.0704	8.10
7	2011	17644.98	9.78	4.26	0	33.30
8	2011	49803.49	10.82	2.71	0	37.80
9	2012	1334.10	7.20	5.63	1	0.90
10	2012	712.01	6.57	5.48	1	0.00
11	2012	10799.95	9.29	4.14	0	34.20
12	2012	29609.11	10.30	4.44	0	10.80

26. Laquelle de ces propositions est fausse ? Une expérience contrôlée randomisée :

- i. requiert deux groupes : un groupe de contrôle et un groupe de traitement.
- j. consiste à donner une chance égale à chaque sujet de l'expérience de faire partie d'un des deux groupes.
- k. permet précisément de ne pas devoir vérifier si les sujets qui prennent part à l'expérience sont similaires par ailleurs.
- l. permet de tester l'efficacité d'une décision politique ou économique, à condition que les deux groupes soient similaires par ailleurs.

27. Que mesure l'erreur de type II sachant que l'hypothèse nulle est celle d'innocence ?

- a. Le risque d'envoyer un innocent en prison.
- b. Le risque de laisser en liberté un coupable.
- c. Le risque d'envoyer un coupable en prison trop souvent.
- d. Le risque de laisser un innocent trop souvent en liberté.

AOUT 2016

1d / 2b / 3a / 4b / 5d / 6d / 7d / 8c / 9c / 10d / 11B / 12D / 13D /

14B / 15C / 16A / 17C / 18A / 19C / 20C / 21C / 22C / 23C / 24B / 25C / 26C / 27B / 28A /

29C / 30B / 31A / 32B / 33B / 34D / 35B / 36C / 37D / 38C / 39A / 40C

1. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
 - a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. Un tel estimateur sera sans biais.**

2. Sur base de 31 observations, on obtient : $\hat{y}_i = 0.5 - 1.5 x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 5%, si la valeur 'vraie' du coefficient de la pente est égale ou supérieure à 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
 - a. L'hypothèse nulle n'est pas rejetée.
 - b. L'hypothèse nulle est rejetée.**
 - c. Il est impossible de conclure.
 - d. Le test est erroné : l'hypothèse alternative n'est pas correctement spécifiée.

3. Considérons une situation identique à la question précédente, sauf que le test d'hypothèse est bilatéral. A quelle conclusion aboutit-on?
 - a. L'hypothèse nulle est rejetée.**
 - b. L'hypothèse nulle n'est pas rejetée.
 - c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

4. Considérons une diminution de la taille du test bilatéral précédent, de 5% à 1%. Quelle en est la conclusion ?
 - a. L'hypothèse nulle est rejetée et la probabilité de commettre une erreur de type I augmente.
 - b. L'hypothèse nulle n'est pas rejetée mais la probabilité de commettre une erreur de type II augmente.**
 - c. L'hypothèse nulle est rejetée et la probabilité de commettre une erreur de type I diminue.
 - d. Il est impossible de conclure, d'autant plus que les valeurs critiques varient.

5. Sur base de $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$ et de 50 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.1 & 0.3 & -0.1 \\ 0.3 & 0.2 & -0.2 \\ -0.1 & -0.2 & 0.5 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ -0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.9$$

Quelle est l'estimation de a_2 ?

- a. -0,05.
- b. -0,03.
- c. 0,09.
- d. -0,13.**

6. En se basant sur les informations données à la question précédente, quelle est l'estimation arrondie de l'écart-type de a_2 ?

- a. 0,044.
- b. 0,062.
- c. 0,010.
- d. 0,098.**

7. En se basant sur les valeurs exactes obtenues aux deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?

- a. 7.83.
- b. 0,48.
- c. 13,58.
- d. 1,33.**

8. Si une variable superflue est présente dans la régression,

- a. les estimateurs des MCO ne sont plus BLUE ;
- b. la variance de l'erreur diminue ;
- c. le risque de multicollinéarité est plus élevé ;**
- d. le nombre de degrés de liberté augmente.

9. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?

a. On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.

b. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.

c. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.

d. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.

10. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;

b. la variable omise est faiblement corrélée avec les variables incluses;

c. la taille de l'échantillon est grande ;

d. la variance des variables incluses est faible.

11. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

12. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / $DH = 1$ pour le 1^{er} trimestre / 2^{ème} trimestre / 3^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 4^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

13. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

14. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

a. (i), ensuite (ii), ensuite (iii), ensuite (iv).

b. (iv), ensuite (iii), ensuite (ii), ensuite (i).

c. (iv), ensuite (ii), ensuite (iii), ensuite (i).

d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

15. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :
- (i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.
 - (ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.
 - (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
 - (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.
- a. (i) seulement.
 - b. (i) et (ii) seulement.
 - c. (i), (ii) et (iii) seulement.**
 - d. (i), (ii), (iii) et (iv).
16. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. 0,95.**
 - b. 0,05.
 - c. 0,10.
 - d. Aucune des affirmations précédentes n'est vraie.
17. La multicollinéarité :
- a. réduit les intervalles de confiance.
 - b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
 - c. augmente la variance des paramètres estimés.**
 - d. implique la violation d'une des hypothèses de Gauss-Markov.
18. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?
- a. Nullité de l'erreur (conditionnelle) en moyenne.**
 - b. Absence de colinéarité parfaite.
 - c. Homoscédasticité des erreurs.
 - d. Linéarité du modèle par rapport à ses paramètres.
19. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

(i) Il est positivement relié à la variance des résidus.

(ii) Il est positivement relié à la variance de x_i .

(iii) Il mesure le degré d'imprécision dans l'estimation du paramètre a_i .

(iv) Il est positivement relié à la taille de l'échantillon.

a. (i), (ii) et (iv) seulement.

b. (i) et (iii) seulement.

c. (ii) et (iv) seulement.

d. (ii) seulement.

20. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

21. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

u. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

v. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

w. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

x. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

22. Le test spécifié à la question précédente peut être effectué à l'aide:

u. de la distribution de Fisher-Snedecor.

v. des distributions t de Student et de Fisher-Snedecor

w. de la distribution t de Student.

x. impossible à déterminer.

23. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(xxi) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(xxii) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(xxiii) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(xxiv) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

u. (i) et (ii).

v. (ii) et (iii).

w. (iii)

x. (iv)

24. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :

- u. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique
- v. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.**
- w. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs
- x. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée

25. De quelles données s'agit-il ?

- a. De séries temporelles
- b. De données de panel
- c. De données empilées**
- d. De données en coupe transversale.

pays	années	PIB_PPA	LN_PIB_PPA	AJRSet	malfal94	latitude
1	2010	12241.44	9.41	4.36	0	27.90
2	2010	6904.56	8.84	5.63	1	12.60
3	2010	39048.15	10.57	2.15	0	27.00
4	2010	2409.55	7.79	4.27	0.158	24.30
5	2011	4716.98	8.46	4.94	0.74	16.20
6	2011	16959.84	9.74	4.36	0.0704	8.10
7	2011	17644.98	9.78	4.26	0	33.30
8	2011	49803.49	10.82	2.71	0	37.80
9	2012	1334.10	7.20	5.63	1	0.90
10	2012	712.01	6.57	5.48	1	0.00
11	2012	10799.95	9.29	4.14	0	34.20
12	2012	29609.11	10.30	4.44	0	10.80

26. Laquelle de ces propositions est fausse ? Une expérience contrôlée randomisée :

- m. requiert deux groupes : un groupe de contrôle et un groupe de traitement.
- n. consiste à donner une chance égale à chaque sujet de l'expérience de faire partie d'un des deux groupes.
- o. permet précisément de ne pas devoir vérifier si les sujets qui prennent part à l'expérience sont similaires par ailleurs.**
- p. permet de tester l'efficacité d'une décision politique ou économique, à condition que les deux groupes soient similaires par ailleurs.

27. Que mesure l'erreur de type II sachant que l'hypothèse nulle est celle d'innocence ?

- a. Le risque d'envoyer un innocent en prison.
- b. Le risque de laisser en liberté un coupable.**
- c. Le risque d'envoyer un coupable en prison trop souvent.
- d. Le risque de laisser un innocent trop souvent en liberté.

MAI 2013

1. Laquelle des propositions suivantes ne correspond pas à la représentation algébrique de la fonction de régression de l'échantillon ? (u_t caractérise le terme d'erreur non-observé)

a. $y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t$

b. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t + u_t$

c. $y_t = \hat{\alpha} + \hat{\beta}x_t + (y_t - \hat{y}_t)$

d. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$

2. Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle ? (u_i caractérise le terme d'erreur non-observé ; e représente l'exponentielle ; les lettres grecques indiquent les paramètres du modèle)

(i) $y_i = \alpha + \beta(z_i / x_i)$

(ii) $y_i = \alpha x_i^\beta z_i^\gamma u_i$

(iii) $y_i = e^\alpha x_i^\beta e^{u_i}$

(iv) $y_i = \alpha + \beta x_i^\delta + u_i$

a. (i), (ii) et (iii) seulement

b. (ii) et (iii) seulement

c. (iii) seulement

d. Tous

3. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?

a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.

b. l'estimateur est efficient

c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.

d. Un tel estimateur sera sans biais.

4. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?

a. L'hypothèse nulle n'est pas rejetée.

b. L'hypothèse nulle est rejetée.

c. Il est impossible de conclure.

d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

5. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
- L'hypothèse nulle est rejetée.
 - L'hypothèse nulle n'est pas rejetée.**
 - Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
6. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
- La probabilité de l'erreur de type I diminue.
 - La probabilité de l'erreur de type II augmente.
 - Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue.**
 - L'hypothèse nulle sera rejetée moins souvent.
7. Supposons que la consommation de 100 ménages soit disponible pour l'année 2006.
- Données de panel.
 - Données en coupe transversale.**
 - Données temporelles.
 - Impossible de conclure.

8. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- 0,05.
 - 0,03.
 - 0,01.**
 - 0,03.
9. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?

a. **0,057.**

b. 0,003.

c. 0,632.

d. 0,094.

10. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?

a. 3,33.

b. 0,05.

c. **0,18.**

d. 0,32.

11. Si une variable superflue est présente dans la régression,

a. les estimateurs des MCO ne sont plus BLUE ;

b. la variance de l'erreur diminue ;

c. **le risque de multicollinéarité est plus élevé ;**

d. le nombre de degrés de liberté augmente.

12. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?

a. On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.

b. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.

c. **On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.**

d. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.

13. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:

a. le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;

b. la variable omise est faiblement corrélée avec les variables incluses;

c. la taille de l'échantillon est grande ;

d. la variance des variables incluses est faible.

14. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + a_3D_tv_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1D_t + a_2v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2v_t + D_t\varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2v_t + a_3D_tv_t + \varepsilon_t$.

a. (i) seulement.

b. (i), (ii) et (iv) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

15. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$v_t = a_0 + a_1DP_t + a_2DA_t + a_3DH_t + \varepsilon_t$ (v_t = ventes; DP / DA / $DH = 1$ pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.

b. Il s'agit de la moyenne des ventes sur toute l'année.

c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.

d. Aucune des réponses précédentes.

16. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

a. n'a pas de variable indicatrice comme variable dépendante.

b. ne peut pas être estimé par la technique des moindres carrés généralisés.

c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.

d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.

17. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.

(i) Supprimer les variables statistiquement superflues.

(ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.

(iii) Conduire des tests de diagnostic sur les résidus observés.

(iv) Estimer le modèle.

a. (i), ensuite (ii), ensuite (iii), ensuite (iv).

b. (iv), ensuite (iii), ensuite (ii), ensuite (i).

c. (iv), ensuite (ii), ensuite (iii), ensuite (i).

d. (i), ensuite (iii), ensuite (ii), ensuite (iv).

18. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :

(i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.

(ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.

(iii) $y = a_0 + a_1 x_1 + \varepsilon$.

(iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.

a. (i) seulement.

b. (i) et (ii) seulement.

c. (i), (ii) et (iii) seulement.

d. (i), (ii), (iii) et (iv).

19. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$

a. 0,95.

b. 0,05.

c. 0,10.

d. Aucune des affirmations précédentes n'est vraie.

20. Une distribution leptokurtique :

- a. dispose d'extrémités plus épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- b. dispose d'extrémités plus épaisses et est plus 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.**
- c. dispose d'extrémités moins épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- d. dispose d'extrémités plus épaisses et est asymétrique.

21. La multicolinéarité :

- a. réduit les intervalles de confiance.
- b. permet une estimation plus fiable des paramètres (obtenus à partir d'un échantillon donné) du modèle.
- c. augmente la variance des paramètres estimés.**
- d. implique la violation d'une des hypothèses de Gauss-Markov.

22. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?

- a. Nullité de l'erreur (conditionnelle) en moyenne.**
- b. Absence de colinéarité parfaite.
- c. Homoscédasticité des erreurs.
- d. Linéarité du modèle par rapport à ses paramètres.

23. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

- (i) Il est positivement relié à la variance des résidus.
 - (ii) Il est positivement relié à la variance de x_i .
 - (iii) Il donne une mesure de précision de l'estimation du paramètre a_1 .
 - (iv) Il est positivement relié à la taille de l'échantillon.
- a. (i), (ii) et (iv) seulement.
 - b. (i) et (iii) seulement.
 - c. (ii) et (iv) seulement.**
 - d. (ii) seulement.

24. Lorsqu'une variable est 'significative à 10%', lesquelles des affirmations suivantes sont *fausses*?

(i) L'hypothèse nulle selon laquelle la valeur vraie du coefficient de cette variable est égale à zéro, est non rejetée à un seuil de confiance de 10%.

(ii) Le t-ratio test indique que le t^* (calculé en valeur absolue) est plus grand que la valeur critique (c.à.d. le 95ème percentile) provenant de la distribution t de Student.

(iii) La valeur absolue du coefficient estimé de cette variable est nécessairement élevée.

(iv) Le coefficient estimé de cette variable affiche un écart-type petit relativement à 10%.

a. (i) et (ii).

b. (i) et (iii).

c. (i), (iii) et (iv).

d. (i), (ii), (iii) et (iv).

25. Imaginez le modèle spécifié suivant: $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + a_3x_{3t} + u_t$. Si vous voulez tester $H_0: a_1 - a_2 = 2$ versus $H_1: a_1 - a_2 > 2$, lequel des modèles spécifiés ci-dessous vous permettra d'effectuer ce test le plus facilement (sans devoir recourir à la covariance entre a_1 et a_2)?

y. $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + u_t$

z. $y = a_0 + (a_1 - a_2)x_1 + (a_2 + a_2)(x_2 + x_1) + (a_3 + a_2)(x_3 + x_1) + u_t$

aa. $y = a_0 + (a_1 - a_2)x_1 + a_2(x_2 + x_1) + a_3x_3 + u_t$

bb. $y = (a_0 + a_2x_1) + (a_1 - a_2)x_1 + a_2x_2 + a_3x_3 + u_t$

26. Le test spécifié à la question précédente peut être effectué à l'aide:

y. de la distribution de Fisher-Snedecor.

z. des distributions t de Student et de Fisher-Snedecor

aa. de la distribution t de Student.

bb. impossible à déterminer.

27. Quelle(s) hypothèse(s) nulle(s) parmi les suivantes peu(ven)t être testée(s) à l'aide d'un F-test?

(xxv) $H_0: a_1 + a_2 = 2$ versus $H_1: a_1 + a_2 < 2$

(xxvi) $H_0: a_1 = 1$ et $a_2 = 0$ versus $H_1: a_1 \neq 1$ et $a_2 \neq 0$

(xxvii) $H_0: a_1/a_2 = 2$ et $a_2 = 6$ versus $H_1: a_1/a_2 \neq 2$ ou $a_2 \neq 6$.

(xxviii) $H_0: a_2a_2 = 0$ versus $H_1: a_2a_2 \neq 0$

y. (i) et (ii).

z. (ii) et (iii).

aa. (iii)

bb. (iv)

28. Du point de vue épistémologique, il est préférable de dire qu'une hypothèse n'est pas rejetée plutôt que de dire qu'elle est acceptée car :

y. aucune conclusion fiable ne peut être tirée de l'utilisation de l'outil économétrique

z. le non-rejet d'une hypothèse testée n'implique pas que l'hypothèse alternative est nécessairement et toujours fausse.

aa. la probabilité de trouver un cygne noir est trop faible pour accepter l'hypothèse nulle selon laquelle tous les cygnes sont noirs

bb. l'absence de preuve de culpabilité signifie que l'hypothèse (nulle) d'innocence est acceptée

29. Les données de cohorte:

a. sont plus générales que les données de panel

b. regroupent plusieurs jeux de données en coupe transversale, qui ont été prélevées à différents moments temporels

c. impliquent que la taille de l'échantillon ainsi que les unités comprises dans l'échantillon ne varient pas au cours du temps.

d. portent sur des individus ou des entreprises mais pas sur des objets.

30. Quel est le lien entre convergence et la consistance d'un estimateur des MCO ?

a. Consistance \Rightarrow convergence.

b. convergence \Leftrightarrow consistance.

c. Convergence \Rightarrow consistance.

d. Aucune des affirmations précédentes n'est vraie.

JUIN 2013

1. Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle ? (u_i caractérise le terme d'erreur non-observé ; e représente l'exponentielle ; les lettres grecques indiquent les paramètres du modèle)
 - (i) $y_i = \alpha + \beta(z_i / x_i) + \gamma(w_i / v_i)$
 - (ii) $y_i = \alpha x_i^\beta z_i^\gamma w_i^\delta u_i$
 - (iii) $y_i = e^\alpha x_i^\beta e^{u_i}$
 - (iv) $y_i = \alpha + \beta x_i^\delta + u_i$
 - a. (i), (ii) et (iii) seulement
 - b. (ii) et (iii) seulement
 - c. (iii) seulement
 - d. Tous
2. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
 - a. la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - b. l'estimateur est efficient
 - c. Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - d. Un tel estimateur sera sans biais.
3. Sur base de 27 observations, on obtient : $\hat{y}_i = -1.6 + 0.9 x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
 - a. L'hypothèse nulle n'est pas rejetée.
 - b. L'hypothèse nulle est rejetée.
 - c. Il est impossible de conclure.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

4. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
 - a. L'hypothèse nulle est rejetée.
 - b. L'hypothèse nulle n'est pas rejetée.
 - c. Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - d. Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.

5. Considérons une augmentation de la taille du test, de 1% à 5%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
 - a. La probabilité de l'erreur de type I diminue.
 - b. La probabilité de l'erreur de type II augmente.
 - c. Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue.
 - d. L'hypothèse nulle sera rejetée moins souvent.

6. Si la consommation de 100 ménages ayant tous subi les ravages du tremblement de terre de 1994 est à la fois disponible pour les années 1994, 2000, et 2006, il s'agit de :
 - a. données de cohorte ;
 - b. données en coupe transversale ;
 - c. données temporelles ;
 - d. données mixtes et indéterminées.

7. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 38 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.1 & 0.3 & -0.1 \\ 0.3 & 0.2 & -0.2 \\ -0.1 & -0.2 & 0.5 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

- a. 0,05.
 - b. -0,11.
 - c. 0,07.
 - d. -0,05.
-
8. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?
 - a. 0,059.

- b. 0,041.
- c. 0,093.
- d. 0,003.

9. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?

- a. 1,21.
- b. -1,88.
- c. 0,76.
- d. -1,21.

10. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?

- a. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
- b. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.

c. On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.

- d. On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.

10 Si une variable superflue est présente dans la régression,

- a. le nombre de degrés de liberté augmente ;
- b. les estimateurs des MCO ne sont plus BLUE ;
- c. la variance de l'erreur diminue ;

d. le risque de multicollinéarité est plus élevé.

11. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?

(i) $\text{profit}_t = a_0 + a_1 D_t + a_2 v_t + a_3 D_t v_t + \varepsilon_t$.

(ii) $\text{profit}_t = a_0 + a_1 D_t + a_2 v_t + \varepsilon_t$.

(iii) $\text{profit}_t = a_0 + a_2 v_t + D_t \varepsilon_t$.

(iv) $\text{profit}_t = a_0 + a_2 v_t + a_3 D_t v_t + \varepsilon_t$.

- a. (i), (ii) et (iv) seulement.
- b. (i), (ii) et (iii) seulement.
- c. (i) seulement.
- d. (i), (ii), (iii) et (iv).

12. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :

$$v_t = a_0 + a_1 DP_t + a_2 DA_t + a_3 DH_t + \varepsilon_t \quad (v_t = \text{ventes}; DP / DA / DH = 1 \text{ pour le } 1^{\text{er}} \text{ trimestre} / 3^{\text{ème}} \text{ trimestre} / 4^{\text{ème}} \text{ trimestre}, 0 \text{ autrement}).$$

- Quelle est la signification de la constante ?
- a. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.
 - b. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.
 - c. Il s'agit de la moyenne des ventes sur toute l'année.
 - d. Aucune des réponses précédentes.

13. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :

- a. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.**
- b. ne peut pas être estimé par la technique des moindres carrés généralisés.
- c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.
- d. n'a pas de variable indicatrice comme variable dépendante.

14. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :

- (i) $y = a_0 + a_1 x_1^2 + a_2 x_2 + \varepsilon$.
- (ii) $y = a_0 + a_1 1/x_1 + a_2 x_2 + \varepsilon$.
- (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
- (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.

- a. (i), (ii), (iii) et (iv).
- b. (i), (ii) et (iii) seulement.**
- c. (i) et (ii) seulement.
- d. (i) seulement.

15. Un chercheur opte pour la forme fonctionnelle, $y_0 = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \varepsilon$ ($n=20$). Sachant que le modèle non-contraint utilisé par le chercheur pour effectuer le RESET est $y_1 = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4\hat{y}_0^2 + a_5\hat{y}_0^3 + v$, quelle est la conclusion du test à un seuil de confiance de 5% si $R_1^2 = 0,339$ et $R_0^2 = 0,301$?
- a. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;17) > 0,33$.
 - b. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;14) > 0,40$.**
 - c. L'hypothèse nulle n'est pas rejetée car $F_{crit}(3;14) > 0,46$.
 - d. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;14) > 0,25$.
16. Un chercheur vérifie la stabilité structurelle du modèle $y_{t0} = a_0 + a_1x_{t1} + a_2x_{t2} + a_3x_{t3} + \varepsilon_t$ ($n=654$). A un seuil de confiance de 5%, quelle est la conclusion du test de Chow si les SCR de deux régressions sont 0,189 et 0,082 ?
- a. On ne peut tirer aucune conclusion car on ne connaît pas la SCR de la troisième régression.**
 - b. L'hypothèse nulle n'est pas rejetée car $F_{crit}(3;648) > 281.85$.
 - c. L'hypothèse nulle n'est pas rejetée car $F_{crit}(4;646) > 210.74$.
 - d. L'hypothèse nulle n'est pas rejetée car $F_{crit}(4;646) > 17.04$.
17. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. 0,95.**
 - b. 0,05.
 - c. 0,10.
 - d. 0.90.
18. Une distribution leptokurtique :
- a. dispose d'extrémités moins épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
 - b. dispose d'extrémités plus épaisses et est plus 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.**
 - c. dispose d'extrémités plus épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
 - d. dispose d'extrémités plus épaisses et est asymétrique.
19. Sous l'hypothèse nulle de Jarque-Bera, la distribution normale a :

- a. une skewness égale à 1 et un kurtosis de 3.
 - b. Aucune autre affirmation n'est vraie.**
 - c. une skewness égale à 1 et un 'excess kurtosis' nul.
 - d. une skewness égale à 1 et un kurtosis nul.
20. Laquelle des affirmations suivante est vraie ? La normalité des erreurs est requise pour garantir:
- a. l'absence de biais des estimateurs des MCO.
 - b. la validité des tests d'inférence statistique en échantillon fini.**
 - c. l'efficacité des estimateurs des MCO.
 - d. la consistance des estimateurs des MCO.
21. Sur base de la théorie asymptotique, la validité des tests statistiques requiert :
- a. l'absence de multicolinéarité entre les variables explicatives.
 - b. des estimateurs biaisés.
 - c. la normalité des erreurs.
 - d. l'absence d'hétéroscédasticité dans les erreurs.**
22. Pour obtenir un estimateur efficace, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?
- a. Homoscédasticité des erreurs.
 - b. Linéarité du modèle par rapport à ses paramètres.
 - c. Nullité de l'erreur (conditionnelle) en moyenne.**
 - d. Absence de colinéarité parfaite.
23. Quel est le lien entre l'efficacité et la consistance d'un estimateur des MCO ?
- a. Consistance \Rightarrow efficacité.
 - b. Consistance \Leftrightarrow efficacité.
 - c. Efficacité \Rightarrow consistance.
 - d. Aucune des affirmations précédentes n'est vraie.**

24. Supposez que vous ne teniez pas compte de la présence de corrélation sérielle dans les erreurs. Quelle(s) en est (sont) la (les) conséquence(s) ?

- (i) Les écart-types estimés sont inappropriés.
- (ii) L'estimateur n'est pas BLUE.
- (iii) Les tests d'inférence statistique sont invalides.
- (iv) Les prévisions faites à partir du modèle sont biaisées.
- a. (i) et (ii) seulement.
- b. (i) seulement.
- c. (i), (ii), (iii) et (iv).
- d. (i), (ii) et (iii) seulement.

25. Considérer le modèle $y_{i0} = a_0 + a_1x_{i1} + a_2x_{i2} + a_3x_{i3} + \varepsilon_i$. Un chercheur veut effectuer un test de White. Quelle est la régression auxiliaire la plus appropriée ?

- a. $e_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + b_4x_{i1}^2 + b_5x_{i2}^2 + b_6x_{i3}^2 + v_i$
- b. $e_i^2 = b_0 + b_1\hat{y}_i + b_2\hat{y}_i^2 + v_i$.
- c. $e_i^2 = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i1}^2 + b_4x_{i2}^2 + v_i$.
- d. $e_i^2 = b_0 + b_1e_{i-1} + v_i$.

26. Quel terme caractérise le processus de génération de données suivant ?

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

- a. **Un processus non-stationnaire et explosif.**
- b. Une marche aléatoire avec dérive.
- c. Un processus stationnaire mais hautement persistant.
- d. Un bruit blanc.

27. Pour autant que l'on dispose d'un nombre suffisamment élevé de données, on peut utiliser les écart-types robustes (HCSE) :

- (i) car ils délivrent des estimations *sans biais* et consistantes des valeurs 'vraies' des variances-covariances des paramètres.
- (ii) sans se préoccuper (nécessairement) de la présence d'hétéroscédasticité dans les erreurs.
- (iii) car ils sont asymptotiquement efficaces.
- (iv) sauf si les erreurs sont homoscedastiques.
- a. (i), (ii), (iii) et (iv).
- b. (i), (ii) et (iii) seulement.
- c. **(ii) et (iii) seulement.**

d. (i) et (ii) seulement.

28. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

(i) Il est positivement relié à la variance des résidus.

(ii) Il est positivement relié à la variance de x_i .

(iii) Il donne une mesure de précision de l'estimation du paramètre a_1 .

(iv) Il est positivement relié à la taille de l'échantillon.

a. (i), (ii) et (iv) seulement.

b. (i) et (iii) seulement.

c. (ii) et (iv) seulement.

d. (ii) seulement.

JUIN 2012

1. Laquelle des propositions suivantes ne correspond pas à la représentation algébrique de la fonction de régression de l'échantillon ? (u_t caractérise le terme d'erreur non-observé)

a. $y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t$

b. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t + u_t$

c. $y_t = \hat{\alpha} + \hat{\beta}x_t + (y_t - \hat{y}_t)$

d. $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$

2. Quel(s) modèle(s) peu(ven)t être estimé(s) par MCO, après transformation éventuelle ? (u_i caractérise le terme d'erreur non-observé ; e représente l'exponentielle ; les lettres grecques indiquent les paramètres du modèle)

(i) $y_i = \alpha + \beta(z_i / x_i)$

(ii) $y_i = \alpha x_i^\beta z_i^\gamma u_i$

(iii) $y_i = e^\alpha x_i^\beta e^{u_i}$

(iv) $y_i = \alpha + \beta x_i^\delta + u_i$

a. (i), (ii) et (iii) seulement

b. (ii) et (iii) seulement

c. (iii) seulement

- d. Tous
3. Si on dit qu'un estimateur a une variance minimale, quelle affirmation suivante n'en est pas une implication?
- la probabilité que l'estimation soit *fortement* éloignée de la valeur du paramètre obtenu en moyenne est *minimisée*.
 - l'estimateur est efficient
 - Un tel estimateur est qualifié de BLUE si, par ailleurs, il est linéaire et sans biais.
 - Un tel estimateur sera sans biais.**
4. Sur base de 27 observations, on obtient : $\hat{y}_i = -2.6 + 0.6 x_i$. Les écart-types estimés sont donnés entre parenthèses. Testons, à un seuil de 1%, si la valeur 'vraie' du coefficient de la pente est 1, contre l'hypothèse alternative qu'elle est inférieure à 1. Quelle est la conclusion?
- L'hypothèse nulle n'est pas rejetée.**
 - L'hypothèse nulle est rejetée.
 - Il est impossible de conclure.
 - Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
5. Considérons une situation identique à la question précédente, sauf que le seuil de significativité est de 5% et que l'hypothèse alternative stipule que la valeur 'vraie' du coefficient de la pente est supérieure à 1. A quelle conclusion aboutit-on?
- L'hypothèse nulle est rejetée.
 - L'hypothèse nulle n'est pas rejetée.**
 - Il est impossible de conclure, d'autant plus que les valeurs critiques changent.
 - Le test est erroné: l'hypothèse alternative n'est pas correctement spécifiée.
6. Considérons une augmentation de la taille du test, de 5% à 10%. Quelle pourrait en être une conséquence, toutes choses restant égales par ailleurs?
- La probabilité de l'erreur de type I diminue.
 - La probabilité de l'erreur de type II augmente.
 - Le critère de rejet devient moins strict : autrement dit, la valeur critique du test diminue.**
 - L'hypothèse nulle sera rejetée moins souvent.
7. Supposons que la consommation de 100 ménages soit disponible pour l'année 2006.
- Données de panel.

b. Données en coupe transversale.

c. Données temporelles.

d. Impossible de conclure.

8. Sur base de $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$ et de 40 observations, on obtient :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.3 & 0.3 & -0.1 \\ 0.3 & 0.4 & -0.2 \\ -0.1 & -0.2 & 0.2 \end{bmatrix}, (\mathbf{X}'\mathbf{y}) = \begin{bmatrix} -0.5 \\ 0.4 \\ 0.2 \end{bmatrix}, \mathbf{e}'\mathbf{e} = 0.6$$

Quelle est l'estimation de a_2 ?

a. -0,05.

b. 0,03.

c. 0,01.

d. -0,03.

9. En se basant sur les informations données à la question précédente, quelle est l'estimation de l'écart-type de a_2 ?

a. **0,057.**

b. 0,003.

c. 0,632.

d. 0,094.

10. En se basant sur les deux questions précédentes, quelle est la valeur du 't-ratio' pour a_2 ?

a. 3,33.

b. 0,05.

c. 0,18.

d. 0,32.

11. Si une variable superflue est présente dans la régression,

a. les estimateurs des MCO ne sont plus BLUE ;

b. la variance de l'erreur diminue ;

c. le risque de multicollinéarité est plus élevé ;

d. le nombre de degrés de liberté augmente.

12. Soit $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ ($t=1,2,\dots,20$). Après avoir estimé les modèles contraint et non contraint relatifs à $H_0: \beta_1 + \beta_2 = 1$ vs $H_1: \beta_1 + \beta_2 \neq 1$, on obtient les deux sommes du carré des résidus suivantes: 21,4 et 19,2. Quelle est la conclusion du test d'hypothèse à un seuil de confiance de 5% ?
- On rejette l'hypothèse nulle car la statistique du test, égale à 4,35, est supérieure à la valeur critique, égale à 1,95.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,35.
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 1,95, est inférieure à la valeur critique, égale à 4,45.**
 - On ne rejette pas l'hypothèse nulle car la statistique du test, égale à 2,29, est inférieure à la valeur critique, égale à 4,35.
13. Lorsqu'une variable pertinente est omise, les estimations des coefficients sont d'autant plus biaisées que:
- le pouvoir explicatif (ou l'effet partiel) de la variable omise sur la variable dépendante est faible;
 - la variable omise est faiblement corrélée avec les variables incluses;
 - la taille de l'échantillon est grande ;
 - la variance des variables incluses est faible.**
14. Un directeur financier désire tester si les ventes (v_t) effectuées durant les soldes ont un effet particulier sur le profit de la société. Quel(s) modèle(s) peut-il estimer ($D = 1$ si ventes durant les soldes, 0 autrement) ?
- $\text{profit}_t = a_0 + a_1 D_t + a_2 v_t + a_3 D_t v_t + \varepsilon_t$.
 - $\text{profit}_t = a_0 + a_1 D_t + a_2 v_t + \varepsilon_t$.
 - $\text{profit}_t = a_0 + a_2 v_t + D_t \varepsilon_t$.
 - $\text{profit}_t = a_0 + a_2 v_t + a_3 D_t v_t + \varepsilon_t$.
- (i) seulement.
 - (i), (ii) et (iv) seulement.**
 - (i), (ii) et (iii) seulement.
 - (i), (ii), (iii) et (iv).
15. Sur base de données trimestrielles, un analyste spécifie le modèle suivant :
- $$v_t = a_0 + a_1 DP_t + a_2 DA_t + a_3 DH_t + \varepsilon_t$$
- (v_t = ventes; DP / DA / DH = 1 pour le 1^{er} trimestre / 3^{ème} trimestre / 4^{ème} trimestre, 0 autrement). Quelle est la signification de la constante ?

- a. Il s'agit de la moyenne des ventes pour le 1^{er} trimestre.
- b. Il s'agit de la moyenne des ventes sur toute l'année.
- c. Il s'agit de l'écart entre les ventes pour le 2^{ème} trimestre et la moyenne des ventes pour toute l'année.
- d. Aucune des réponses précédentes.**

16. Laquelle des affirmations suivante est vraie ? Le modèle de probabilité linéaire :
- a. n'a pas de variable indicatrice comme variable dépendante.
 - b. ne peut pas être estimé par la technique des moindres carrés généralisés.
 - c. délivre des prédictions cohérentes lorsque les variables explicatives prennent des valeurs éloignées de la moyenne.
 - d. ne bénéficie pas d'estimateurs des MCO qui sont 'BLUE'.**
17. Placer dans le bon ordre quelques-unes des étapes qui permettent d'appliquer une méthode de construction de modèle statistiquement valide.
- (i) Supprimer les variables statistiquement superflues.
 - (ii) Si nécessaire, estimer le modèle en appliquant les méthodes de correction visant à 'blanchir' (ou nettoyer) le terme d'erreur.
 - (iii) Conduire des tests de diagnostic sur les résidus observés.
 - (iv) Estimer le modèle.
- a. (i), ensuite (ii), ensuite (iii), ensuite (iv).
 - b. (iv), ensuite (iii), ensuite (ii), ensuite (i).**
 - c. (iv), ensuite (ii), ensuite (iii), ensuite (i).
 - d. (i), ensuite (iii), ensuite (ii), ensuite (iv).
18. Un chercheur opte pour le modèle suivant : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$. Le modèle est mal spécifié si la FRP est :
- (i) $y = a_0 + a_1 x_1^2 + a_1 x_2 + \varepsilon$.
 - (ii) $y = a_0 + a_1 1/x_1 + a_1 x_2 + \varepsilon$.
 - (iii) $y = a_0 + a_1 x_1 + \varepsilon$.
 - (iv) $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$.
- a. (i) seulement.
 - b. (i) et (ii) seulement.
 - c. (i), (ii) et (iii) seulement.**
 - d. (i), (ii), (iii) et (iv).
19. Un chercheur opte pour la forme fonctionnelle, $y_0 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \varepsilon$ ($n=20$). Sachant que le modèle non-contraint utilisé par le chercheur pour effectuer le RESET est $y_1 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 \hat{y}_0^2 + a_5 \hat{y}_0^3 + v$, quelle est la conclusion du test à un seuil de confiance de 5% si $R_1^2 = 0,339$ et $R_0^2 = 0,301$?

- a. L'hypothèse nulle n'est pas rejetée car $F_{crit}(3;14) > 0,46$.
- b. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;14) > 0,40$.**
- c. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;17) > 0,33$.
- d. L'hypothèse nulle n'est pas rejetée car $F_{crit}(2;14) > 0,25$.
20. Un chercheur vérifie la stabilité structurelle du modèle $y_{t0} = a_0 + a_1x_{t1} + a_2x_{t2} + a_3x_{t3} + \varepsilon_t$ (n=654). A un seuil de confiance de 5%, quelle est la conclusion du test de Chow si les SCR de deux régressions sont 0,189 et 0,082 ?
- a. On ne peut tirer aucune conclusion car on ne connaît pas la SCR de la troisième régression.**
- b. L'hypothèse nulle n'est pas rejetée car $F_{crit}(4;646) > 210.74$.
- c. L'hypothèse nulle n'est pas rejetée car $F_{crit}(3;648) > 281.85$.
- d. L'hypothèse nulle n'est pas rejetée car $F_{crit}(4;646) > 17.04$.
21. Considérer une variable aléatoire centrée réduite X tel que : $X \sim t(\infty)$. Dans ce cas, $\text{Prob}(X < 1.645) =$
- a. 0,95.**
- b. 0,05.
- c. 0,10.
- d. Aucune des affirmations précédentes n'est vraie.
22. Une distribution leptokurtique :
- a. dispose d'extrémités plus épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- b. dispose d'extrémités plus épaisses et est plus 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.**
- c. dispose d'extrémités moins épaisses et est moins 'pointue' autour de la moyenne qu'une distribution normale de même moyenne et de même variance.
- d. dispose d'extrémités plus épaisses et est asymétrique.
23. Sous l'hypothèse nulle de Jarque-Bera, la distribution normale a :
- a. une skewness égale à 1 et un 'excess kurtosis' nul.
- b. une skewness égale à 1 et un kurtosis nul.
- c. une skewness égale à 1 et un kurtosis de 3.
- d. Aucune des affirmations précédentes n'est vraie.**

24. Laquelle des affirmations suivante est vraie ? La normalité des erreurs est requise pour garantir:
- a. l'absence de biais des estimateurs des MCO.
 - b. la consistance des estimateurs des MCO.
 - c. l'efficacité des estimateurs des MCO.
 - d. la validité des tests d'inférence statistique en échantillon fini.**
25. Sur base de la théorie asymptotique, la validité des tests statistiques requiert :
- a. des estimateurs biaisés.
 - b. l'absence de multicolinéarité entre les variables explicatives.
 - c. l'absence d'hétéroscédasticité dans les erreurs.**
 - d. la normalité des erreurs.
26. Considérer le modèle : $y_0 = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \varepsilon$ (n=30). Sur base de la distribution de Fisher-Snedecor et à un seuil de 5%, quelle est la valeur critique à utiliser si l'on veut effectuer un test de White ? (N.B. *La régression auxiliaire est celle de la version parcimonieuse du test, ne contenant que les valeurs ajustées*)
- a. 3,37**
 - b. 2,21
 - c. 4,23
 - d. 2,27
57. Pour obtenir un estimateur efficient, quelle hypothèse de Gauss-Markov n'est PAS nécessaire?
- a. Nullité de l'erreur (conditionnelle) en moyenne.**
 - b. Absence de colinéarité parfaite.
 - c. Homoscédasticité des erreurs.
 - d. Linéarité du modèle par rapport à ses paramètres.
27. Pour autant que l'on dispose d'un nombre suffisamment élevé de données, on peut utiliser les écart-types robustes (HCSE) :
- (i) car ils délivrent des estimations sans biais et consistantes des valeurs 'vraies' des variances-covariances des paramètres.
 - (ii) sans se préoccuper (nécessairement) de la présence d'hétéroscédasticité dans les erreurs.

(iii) car ils sont asymptotiquement efficaces.

(iv) sauf si les erreurs sont homoscedastiques.

e. (ii), (iii) et (iv).

f. (i), (ii) et (iii) seulement.

g. (ii) et (iii) seulement.

h. (i) et (iv) seulement.

28. Soit $y_i = a_0 + a_1 x_i + \varepsilon_i$. Considérons l'estimateur MCO de l'écart-type de \hat{a}_1 . Quelle(s) affirmation(s) est (sont) fausse(s)?

(i) Il est positivement relié à la variance des résidus.

(ii) Il est positivement relié à la variance de x_i .

(iii) Il donne une mesure de précision de l'estimation du paramètre a_1 .

(iv) Il est positivement relié à la taille de l'échantillon.

a. (i), (ii) et (iv) seulement.

b. (i) et (iii) seulement.

c. (ii) et (iv) seulement.

d. (ii) seulement.

29. Considérer le modèle $y_{i0} = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \varepsilon_i$. Un chercheur veut effectuer un test de White. Quelle est la régression auxiliaire la plus appropriée ?

a. $e_i^2 = b_0 + b_1 e_{i-1} + v_i$.

b. $e_i^2 = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1}^2 + b_4 x_{i2}^2 + v_i$.

c. $e_i^2 = b_0 + b_1 \hat{y}_i + b_2 \hat{y}_i^2 + v_i$.

d. $e_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i1}^2 + b_5 x_{i2}^2 + b_6 x_{i3}^2 + v_i$