

**Évaluation Statistique
Éléments de théorie**

Exercice 1. QCM (Vrai/Faux)

(+0,25 point si bonne réponse, 0 point si pas de réponse, -0,25 point si mauvaise réponse)

Question 1. Lorsque les éléments non diagonaux de la matrice de variance-covariance des termes d'erreur sont nuls, on a de l'hétéroscédasticité.

Question 2. Lorsque les éléments non diagonaux de la matrice de variance-covariance des termes d'erreur sont nuls, on a de la non-autocorrélation.

Question 3. Si les éléments diagonaux de la matrice de variance-covariance des termes d'erreur sont égaux à 1, alors il y a de l'hétéroscédasticité.

Question 4. Lorsqu'une variance est négative alors on a de l'autocorrélation négative.

Question 5. L'estimateur des moindres carrés ordinaires est obtenu en minimisant la somme des carrés des résidus.

Question 6. La statistique de Student associée à une variable explicative est égale au paramètre estimé associé à cette variable explicative plus par l'écart-type estimé du paramètre estimé.

Question 7. Minimiser la somme des carrés des erreurs est équivalent à minimiser la distance au carré entre Y (le vecteur des valeurs observées concernant la variable dépendante) et le sous-espace vectoriel engendré par les vecteurs colonnes de la matrice X .

Question 8. On peut trouver la valeur estimée de Y en la multipliant par un projecteur orthogonal sur le sous-espace vectoriel engendré par les vecteurs colonnes de la matrice X .

Question 9. Le fait que l'espérance du terme d'erreur est nulle est un résultat alors que le fait que la moyenne des résidus est nulle est une hypothèse.

Question 10. La matrice de projection orthogonale utilisée dans les moindres carrés ordinaires est $X(X'X)^{-1}X'$.

Question 11. Dans le modèle linéaire général, Y (le vecteur des valeurs observées concernant la variable dépendante) est égal à la somme de la valeur estimée de Y et du vecteur des résidus.

Question 12. La moyenne de la variable dépendante est égale à la moyenne de la variable estimée de la variable dépendante.

Question 13. Un estimateur sans biais de la variance du terme d'erreur est égal à la somme des carrés des résidus divisée par le nombre de degré de liberté.

Question 14. L'autocorrélation veut dire que la variance des termes d'erreur varie selon les individus.

Question 15. Minimiser la somme des carrés des erreurs est équivalent à minimiser la distance au carré entre Y (le vecteur des valeurs observées concernant la variable dépendante) et le sous-espace vectoriel engendré par les vecteurs colonnes de la matrice X .

Question 16. En présence d'autocorrélation, l'estimateur des moindres carrés ordinaires reste à variance minimale.

Question 17. Lorsqu'on rejette l'hypothèse de base dans un test de Student, cela veut que la variable explicative associée est aléatoire.

Question 18. On peut calculer la statistique de Fisher à partir du coefficient de détermination.

Question 19. Lorsqu'on rejette l'hypothèse de base dans le test de nullité globale de Fisher, cela veut dire qu'aucune variable explicative du modèle n'influence la variable dépendante.

Question 20. La statistique de Fisher dans le test de nullité globale est nécessairement positive.

Question 21. L'estimateur des moindres carrés généralisés s'écrit $\hat{\beta}_{MCG} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y$.

Question 22. Dans un test de nullité globale de Fisher, si la p-value est de 0,1 et que le risque de première espèce que l'on est prêt à accepter est de 1% alors on rejette l'hypothèse de base.

Question 23. La différence principale entre l'estimateur des moindres carrés généralisés et l'estimateur des moindres carrés quasi-généralisés est que l'estimateur des moindres carrés quasi-généralisés est BLUE.

Question 24. Le théorème de Gauss-Markov dit que l'estimateur des moindres carrés ordinaires est convergent.

Question 25. Si la matrice X n'est pas de plein-rang colonne alors la matrice $X'X$ admet une valeur propre nulle.

Question 26. En cas de corrélation forte entre deux variables explicatives du modèle, l'estimateur des moindres carrés ordinaires est biaisé.

Question 27. En cas de corrélation forte entre des variables explicatives du modèle, on ne peut faire confiance aux tests de Student.

Question 28. Lorsque je fais une régression linéaire et que je rejette l'hypothèse de nullité globale de Fisher sans rejeter, dans les tests de Student, la nullité des paramètres associés aux variables explicatives, alors il y a suspicion d'autocorrélation des résidus.

Question 29. Supposons une équation de salaire avec comme variable explicative, la variable Femme (égale à 1 si l'individu est une femme et 0 sinon), la variable Jeune (égale à 1 si l'individu est Jeune et 0 sinon) et le croisement de Femme et de Jeune, alors la constante du modèle représente le salaire moyen des jeunes femmes.

Question 30. Pour corriger de l'autocorrélation, il faut multiplier le modèle initial $Y=X\beta + \varepsilon$, par une matrice P, permettant au nouveau terme d'erreur Pe , d'avoir une matrice de variance-covariance avec tous ses éléments non-diagonaux nuls.

Question 31. Soit le modèle linéaire général $Y=X\beta + \varepsilon$, avec $\text{Var}(\varepsilon) = \sigma^2\Omega$ où $\Omega \neq I_n$ la matrice identité. Alors l'estimateur des moindres carrés ordinaires est biaisé.

Question 32. En présence d'autocorrélation, l'estimateur des moindres carrés ordinaires reste convergent.

Question 33. Pour tester l'autocorrélation d'ordre 1, on peut utiliser le test de Durbin-Wu-Hausman.

Question 34. Le test de White permet de tester l'hétéroscédasticité des résidus.

Question 35. La méthode du Condition Index permet de détecter un problème de presque colinéarité entre des variables explicatives du modèle.

Question 36. Le test de Jarque-Bera permet de détecter la convergence des résidus.

Question 37. En cas de multicollinéarité parfaite, l'estimateur des moindres carrés ordinaires reste BLUE.

Question 38. On dispose de données pour n individus où : C est la consommation et R est le revenu disponible. Soit le modèle suivant (1) : $C_i = \beta_0 + \beta_1 R_i + \varepsilon_i$, $i = 1$ à n ; avec $\text{Var}(\varepsilon_i) = \sigma^2 \log R_i$ et $\text{Cov}(\varepsilon_i, \varepsilon_{i'}) = 0$, $\forall i \neq i'$. Alors le modèle (1) viole l'hypothèse de normalité des termes d'erreur.

Question 39. Une distribution qui suit une loi normale correspond à un Skewness=0 et un Kurtosis=3.

Question 40. Soit l'estimation suivante :

$$\text{Budget}_i = 19,44 + 0,018X_{1i} - 0,284X_{2i} + 1,343X_{3i} + 6,331X_{4i} \quad i = 1 \text{ à } 2000 ;$$

(3,406) (0,006) (0,457) (0,259) (3,029)

Au seuil de 5%, la valeur dans la table de student est 1,96. Selon Zeina, la variable X_1 n'est pas explicative.

Question 41. Dans un modèle logit simple, au moins l'une des variables explicatives est binaire.

Question 42. Dans un modèle probit simple, la variable dépendante est binaire.

Question 43. Le vecteur des coefficients estimés dans un modèle logit simple est solution d'un programme de maximisation d'une fonction de vraisemblance.

Question 44. Les coefficients estimés d'une régression logistique simple s'interprètent comme des élasticités.

Question 45. Dans un modèle logit, le terme aléatoire suit une loi normale.

Question 46. Dans un modèle probit, le terme aléatoire suit une loi logistique.

Question 47. Les coefficients estimés des régressions logit et probit peuvent parfois avoir des signes opposés.

Question 48. Dans un modèle logit, la significativité d'une variable explicative se fait via un test dit de Thomas.

Question 49. Dans la pratique, l'estimation des coefficients estimés d'un modèle logit, peut se faire par l'algorithme de Newton-Raphson.

Question 50. Le AIC d'une régression permet de voir si la variable dépendante est vraiment qualitative.

Question 51. La principale différence entre un modèle logit multinomial et un modèle simple réside dans leurs termes d'erreur qui sont différents.

Question 52. Le signe d'une Log Vraisemblance n'est pas déterminé a priori.

Question 53. Le modèle de régression multinomiale est basé sur la théorie du choix probabiliste.

Question 54. Dans le modèle de choix probabiliste, le concept important est que la rationalité de l'agent est probabiliste.

Question 55. Les modèles traditionnels respectent l'IIA.

Question 56. Dans un modèle de régression multinomiale ordonnée, il y a autant de paramètres estimés que de modalités à la variable dépendante.

Question 57. Il y a trois types de modèles de régression multinomiale non-ordonnée.

Question 58. Le pseudo-R2 de McFadden peut être négatif.

Exercice 2.

On dispose de données (fictives) concernant 1500 individus et comportant trois types d'informations, à savoir : la prise de vitamines de type 1 (**vita1**) et de type 2 (**vita2**) du 1^{er} janvier de l'année en cours au 31 décembre de la même année, ainsi que la hausse de taille observée (**taille**) sur cette période.

Question 1. Compléter (en justifiant rigoureusement) la sortie logicielle ci-dessous.

Question 2. Que conclue le test de nullité globale de Fisher au seuil de 1%.

Question 3. Que concluent les tests de nullité des paramètres au seuil de 1%.

Question 4. Interpréter le résultat de l'estimation. Vous semble-t-elle satisfaisante ?

Dependent Variable: taille					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	a	29.60595	e	f	0.0166
Error	b	c	3.60200		
Corrected Total	1499	d			
Root MSE		g	R-Square	h	
Dependent Mean		2.39933	Adj R-Sq	0.0041	
Coeff Var		79.10087			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.27839	0.12633	k	<.0001
vital1	1	i	0.00425	2.63	0.0087
vital2	1	-0.01040	j	-1.21	0.2262

Exercice 3 : Autocorrélation d'ordre 1

On dispose de données temporelles avec une variable dépendante Y et sur une variable explicative X1.

1) Commenter l'estimation ci-dessous :

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.06845	0.37239	5.555	3.05e-07 ***
X1	0.08322	0.06148	1.354	0.179

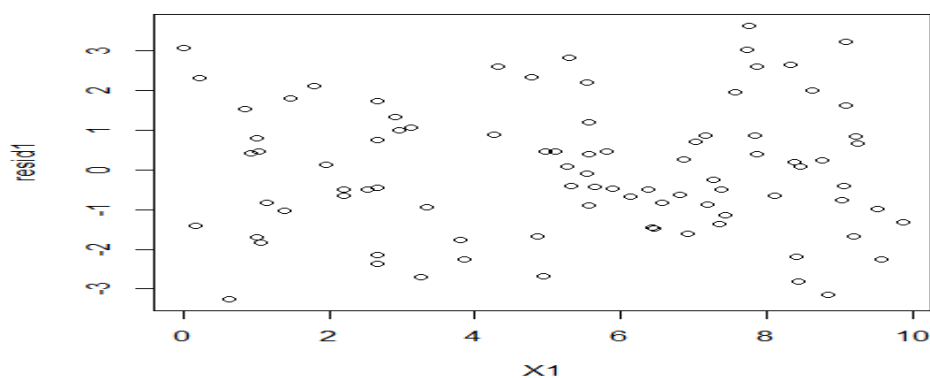
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.632 on 86 degrees of freedom

Multiple R-squared: 0.02086, Adjusted R-squared: 0.009477

F-statistic: 1.832 on 1 and 86 DF, p-value: 0.1794

2) Le graphe ci-dessous décrit les résidus en fonction de X1. Au vu du graphe, y a-t-il hétéroscédasticité ou homoscedasticité ?



3) Quelle est l'hypothèse de base du test de Breusch-Pagan ?

4) Le test de Breusch-Pagan donne le résultat ci-dessous :

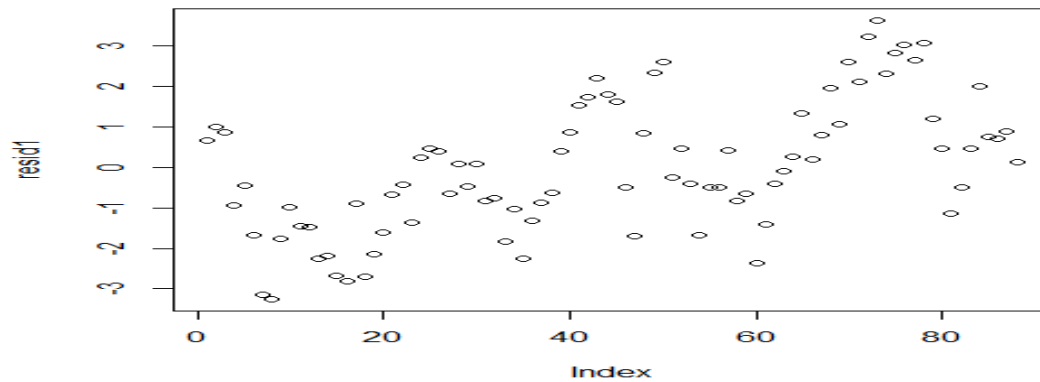
studentized Breusch-Pagan test

data: reg1

BP = 0.00026223, df = 1, p-value = 0.9871

Que concluez-vous quant à hétéroscédasticité ou homoscedasticité ?

5) Le graphe ci-dessous donne la distribution des résidus. Au vu de ce graphe, y a-t-il autocorrélation des résidus ?



6) Quelle est l'hypothèse de base du test de Durbin et Watson ?

7) Le test de Durbin et Watson donne le résultat ci-dessous :

Durbin-Watson test

data: reg1

DW = 0.39904, p-value < 2.2e-16

Que concluez-vous quant à l'existence de l'autocorrélation ?

8) L'estimation en moindres carrés généralisés donne :

Coefficients:

	<i>Value</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
<i>(Intercept)</i>	1.9475699	0.5495166	3.544151	6e-04
<i>X1</i>	0.1123282	0.0272659	4.119724	1e-04

Commentez cette estimation, notamment en la comparant avec l'estimation de la question 1).