

# Evaluation QCM

SIDIBE Oumou Malika

## Exercice 1

1. Faux . Lorsque les éléments non diagonaux de la matrice de variance-covariance des termes d'erreur sont nuls, on a de l'homoscédasticité, pas de l'hétéroscédasticité.
2. Vrai . Lorsque les éléments non diagonaux de la matrice de variance-covariance des termes d'erreur sont nuls, cela indique qu'il n'y a pas d'autocorrélation entre les termes d'erreur.
3. Faux . Si les éléments diagonaux de la matrice de variance-covariance des termes d'erreur sont égaux à 1, cela indique de l'homoscédasticité, pas de l'hétéroscédasticité.
4. Faux . Une variance ne peut pas être négative ; par contre, une autocorrélation peut être négative, mais cela ne découle pas d'une variance négative.
5. Vrai . L'estimateur des moindres carrés ordinaires est obtenu en minimisant la somme des carrés des résidus.
6. Faux . La statistique de Student associée à une variable explicative est égale au paramètre estimé divisé par l'écart-type estimé du paramètre estimé.
7. Vrai . Minimiser la somme des carrés des erreurs est équivalent à minimiser la distance au carré entre Y et le sous-espace vectoriel engendré par les vecteurs colonnes de la matrice X.
8. Vrai . 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.
9. Faux . Le fait que l'espérance du terme d'erreur est nulle est une hypothèse, tandis que la moyenne des résidus est nulle est un résultat.
10. Faux . La matrice de projection orthogonale utilisée dans les moindres carrés ordinaires est  $X(X'X)^{-1}X'$ .
11. Vrai . Dans le modèle linéaire général, Y est égal à la somme de la valeur estimée de Y et du vecteur des résidus.
12. Vrai . La moyenne de la variable dépendante est égale à la moyenne de la variable estimée de la variable dépendante.
13. Vrai . Un estimateur sans biais de la variance du terme d'erreur est la somme des carrés des résidus divisée par le nombre de degrés de liberté.
14. Faux . L'autocorrélation signifie que les termes d'erreur sont corrélés entre eux dans le temps ou l'espace, pas que leur variance varie selon les individus.
15. Vrai . Minimiser la somme des carrés des erreurs est équivalent à minimiser la distance au carré entre Y et le sous-espace vectoriel engendré par les vecteurs colonnes de la matrice X.
16. Faux . En présence d'autocorrélation, l'estimateur des moindres carrés ordinaires n'est plus à variance minimale.

- 17.** Faux . Rejeter l'hypothèse de base dans un test de Student signifie que la variable explicative est statistiquement significative, pas qu'elle est aléatoire.
- 18.** Vrai . On peut calculer la statistique de Fisher à partir du coefficient de détermination.
- 19.** Faux . Rejeter l'hypothèse de base dans le test de nullité globale de Fisher signifie que certaines variables explicatives influencent la variable dépendante.
- 20.** Vrai . La statistique de Fisher dans le test de nullité globale est nécessairement positive.
- 21.** Vrai . L'estimateur des moindres carrés généralisés s'écrit bien  $\hat{\beta}_{BMG}x = x(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ .
- 22.** Faux . Si la p-value est de 0,1 et que le risque de première espèce accepté est de 1%, on ne rejette pas l'hypothèse de base.
- 23.** Faux . L'estimateur des moindres carrés généralisés est BLUE (Best Linear Unbiased Estimator), contrairement à l'estimateur des moindres carrés quasi-généralisés qui peut ne pas l'être.
- 24.** Faux . Le théorème de Gauss-Markov stipule que l'estimateur des moindres carrés ordinaires est BLUE, pas nécessairement convergent.
- 25.** Vrai . Si la matrice X n'est pas de plein-rang colonne, la matrice  $X'X$  a une valeur propre nulle.
- 26.** Faux . En cas de corrélation forte entre deux variables explicatives, l'estimateur des moindres carrés ordinaires est imprécis (variance élevée), mais il reste non biaisé.
- 27.** Vrai . En cas de corrélation forte entre des variables explicatives, les tests de Student peuvent devenir peu fiables.
- 28.** Faux . Rejeter l'hypothèse de nullité globale de Fisher et non les tests de Student pour les variables explicatives indique plutôt un problème de multi colinéarité ou une autre spécification erronée, pas nécessairement une autocorrélation des résidus.
- 29.** Faux . La constante du modèle représente le salaire moyen des individus qui ne sont ni jeunes ni femmes (Femme = 0 et Jeune = 0).
- 30.** Vrai . Pour corriger de l'autocorrélation, on utilise une matrice P pour rendre les erreurs non corrélées.
- 31.** Faux . L'estimateur des moindres carrés ordinaires reste non biaisé même si  $Var(\varepsilon) = \sigma^2\Omega$  où  $\Omega x \neq xI_n$ .
- 32.** Vrai . En présence d'autocorrélation, l'estimateur des moindres carrés ordinaires reste convergent, bien que non efficient.
- 33.** Faux . Le test de Durbin-Watson, pas Durbin-Wu-Hausman, est utilisé pour tester l'autocorrélation d'ordre 1.
- 34.** Vrai . Le test de White permet de tester l'hétéroscédasticité des résidus.
- 35.** Vrai . La méthode du Condition Index est utilisée pour détecter une presque colinéarité entre les variables explicatives.
- 36.** Faux . Le test de Jarque-Bera permet de tester la normalité des résidus, pas leur convergence.
- 37.** Faux . En cas de multi colinéarité parfaite, l'estimateur des moindres carrés ordinaires n'est pas défini et donc pas BLUE.

- 38.** Faux . Le modèle ne viole pas nécessairement l'hypothèse de normalité des termes d'erreur ; il viole l'homoscédasticité.
- 39.** Vrai . Une distribution normale a un skewness de 0 et un kurtosis de 3.
- 40.** Vrai . Selon la valeur de Student (1,96) et le coefficient de  $X_1$  car  $\frac{0,018}{0,006}x = x3$ ,  $X_1$  est significatif au seuil de 5%.
- 41.** Faux . Dans un modèle logit simple, les variables explicatives peuvent être continues ou binaires.
- 42.** Vrai . Dans un modèle probit simple, la variable dépendante est binaire.
- 43.** Vrai . Les coefficients estimés dans un modèle logit simple sont obtenus par maximisation de la fonction de vraisemblance.
- 44.** Faux . Les coefficients d'une régression logistique s'interprètent comme des log-odds, pas des élasticités.
- 45.** Faux . Dans un modèle logit, le terme aléatoire suit une distribution logistique, pas normale.
- 46.** Faux . Dans un modèle probit, le terme aléatoire suit une loi normale.
- 47.** Vrai . Les coefficients estimés des modèles logit et probit peuvent avoir des signes opposés en raison des différentes distributions des termes aléatoires.
- 48.** Faux . La significativité dans un modèle logit est testée généralement via un test de Wald, pas un test de Thomas.
- 49.** Vrai . L'algorithme de Newton-Raphson est couramment utilisé pour estimer les coefficients d'un modèle logit.
- 50.** Faux . Le AIC (Akaike Information Criterion) évalue la qualité du modèle, pas la nature qualitative de la variable dépendante.
- 51.** Faux . La différence entre un modèle logit multinomial et simple réside principalement dans la structure de la dépendance des termes d'erreur.
- 52.** Vrai . Le signe de la log-vraisemblance n'est pas déterminé a priori car il dépend de la fonction de vraisemblance et des données.
- 53.** Vrai . Le modèle de régression multinomiale est basé sur la théorie du choix probabiliste.
- 54.** Vrai . Dans la théorie du choix probabiliste, la rationalité de l'agent est modélisée de manière probabiliste.
- 55.** Faux . Les modèles traditionnels ne respectent pas nécessairement l'hypothèse d'indépendance des alternatives non pertinentes (IIA).
- 56.** Faux . Dans un modèle de régression multinomiale ordonnée, il y a généralement moins de paramètres que de modalités à la variable dépendante.
- 57.** Vrai . Il y a trois types principaux de modèles de régression multinomiale non-ordonnée : logit multinomial, probit multinomial et logit conditionnel.
- 58.** Faux . Le pseudo-R<sup>2</sup> de McFadden ne peut pas être négatif car il est basé sur les valeurs de la log-vraisemblance, qui sont toujours positives.

## Exercice 2

### Question 1 : Compléter la sortie logicielle

Pour compléter la sortie logicielle, nous devons calculer les degrés de liberté, les sommes des carrés et la moyenne des carrés, ainsi que les valeurs de F et  $Pr > F$ . Voici la sortie complétée :

Dependent Variable : taille					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	<b>2</b>	29.60595	<b>14.802975</b>	<b>4.11</b>	0.0166
Error	<b>1497</b>	<b>5384.39405</b>	3.60200		
Corrected Total	1499	<b>5414.00000</b>			
	Root MSE	<b>1.89737</b>	R-Square	<b>0.0055</b>	
	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	<b>18.03</b>	<.0001
vita1	1	<b>0.01117</b>	0.00425	2.63	0.0087
vita2	1	-0.01040	0.00860	-1.21	0.2262
...					

### Explications

- DF (degrés de liberté) :

Le modèle : nombre de variables explicatives (2).

Erreur : nombre total d'observations - nombre de paramètres estimés (1500 - 3 = 1497).

Total corrigé : nombre total d'observations - 1 (1500 - 1 = 1499).

- Sum of Squares (somme des carrés):

Le modèle : calculée à partir de la somme des carrés des résidus expliqués par le modèle.

Erreur : somme des carrés totale - somme des carrés du modèle.

Total corrigé : somme des carrés totale des observations.

- Mean Square (moyenne des carrés) :

Le modèle : somme des carrés du modèle / degrés de liberté du modèle.

Erreur : somme des carrés de l'erreur / degrés de liberté de l'erreur.

- F Value et  $Pr > F$  :

Valeur de F : moyenne des carrés du modèle / moyenne des carrés de l'erreur.

$Pr > F$  : valeur p associée au test F.

**Question 2** : Conclusion du test de nullité globale de Fisher au seuil de 1%

Le test de nullité globale de Fisher a une valeur p de 0.0166, ce qui est supérieur au seuil de 1%. Donc, nous ne rejetons pas l'hypothèse nulle, ce qui signifie que les variables explicatives (vita1 et vita2) n'ont pas un effet statistiquement significatif sur la variable dépendante (taille) au seuil de 1%.

**Question 3** : Conclusion des tests de nullité des paramètres au seuil de 1%

**vita1** : La valeur p est de 0.0087, ce qui est inférieur au seuil de 1%. Nous rejetons l'hypothèse nulle et concluons que vita1 a un effet statistiquement significatif sur la taille.

**vita2** : La valeur p est de 0.2262, ce qui est supérieur au seuil de 1%. Nous ne rejetons pas l'hypothèse nulle et concluons que vita2 n'a pas un effet statistiquement significatif sur la taille.

**Question 4** : Interprétation de l'estimation

Le modèle estime que la prise de vitamines de type 1 (vita1) a un effet positif significatif sur la hausse de la taille, alors que la prise de vitamines de type 2 (vita2) n'a pas d'effet significatif. L'ajustement du modèle (R-carré ajusté de 0.0041) est très faible, indiquant que les variables explicatives expliquent très peu de la variation dans la variable dépendante. Le modèle semble donc peu satisfaisant pour expliquer la hausse de taille.

## Exercice 3 : Autocorrélation d'ordre 1

**Question 1** : Commenter l'estimation

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

- L'intercept est significatif (p-value < 0.001), ce qui signifie que la constante du modèle a un effet statistiquement significatif sur Y.

- La variable X1 n'est pas significative (p-value = 0.179), ce qui signifie que X1 n'a pas un effet statistiquement significatif sur Y.

- R-carré ajusté est très faible (0.009477), indiquant que le modèle explique très peu de la variance de Y.

**Question 2 : Hétéroscédasticité ou homoscedasticité ?**

Sans le graphe, il est difficile de déterminer visuellement l'hétéroscédasticité. Cependant, en général, on regarde si les résidus montrent une dispersion constante ou non autour de la ligne de régression.

**Question 3 : Hypothèse de base du test de Breusch-Pagan**

L'hypothèse de base du test de Breusch-Pagan est que les résidus ont une variance constante (homoscédasticité).

**Question 4 : Résultat du test de Breusch-Pagan**

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

La valeur p est très élevée (0.9871), ce qui signifie que nous ne rejetons pas l'hypothèse nulle d'homoscédasticité. Les résidus ont donc une variance constante.

**Question 5 : Autocorrélation des résidus ?**

Oui, au vu de ce graphe, il y a des indications d'autocorrélation des résidus. Les motifs observés suggèrent que les erreurs d'une observation sont corrélées avec les erreurs des observations adjacentes.

**Question 6 : Hypothèse de base du test de Durbin-Watson**

L'hypothèse de base du test de Durbin-Watson est qu'il n'y a pas d'autocorrélation des résidus (autocorrélation d'ordre 0).

**Question 7 : Résultat du test de Durbin-Watson**

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

La valeur du test DW est très faible (0.39904), et la p-value est très petite, indiquant une forte autocorrélation positive des résidus.

**Question 8 : Comparaison avec l'estimation de la question 1)**

Coefficients :

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

- Les coefficients sont légèrement différents, avec des erreurs standards plus petites dans la nouvelle estimation.

- La nouvelle estimation de X1 est significative (p-value = 0.0001), contrairement à l'estimation initiale.

- Ceci indique que la correction pour l'autocorrélation a amélioré l'estimation des coefficients.