



FACULTÉ DE DROIT
ÉCONOMIE & GESTION

**Sujet d'examen - 1^{ère} session
Semestres 2-4-6-8
Année universitaire 2025-2026**

Intitulé de l'épreuve : Econométrie des Variables Qualitatives

Nom de l'enseignant : TOKPAVI Sessi

Mention / Spécialité / Parcours : Master 1 ESA

Année : 2025-2026

Durée de l'épreuve : 3 heures

Documents autorisés : Aucun

Matériels autorisés : Calculatrice

P1/

SUJET

Université d'Orléans - Master 1 ESA
Cours d'Econométrie des Variables Qualitatives

Sessi Tokpavi

Examen Session 1 - AVRIL 2026 - 3h00

Exercice 1: (9 points):

Soit z_1 un vecteur de variables explicatives, z_2 une variable continue et d_1 une variable dichotomique.

Question 1 (1 point): Dans le modèle suivant:

$$\Pr(y = 1 | z_1, z_2) = \Phi(z_1\delta_1 + z_2\gamma_1 + z_2^2\gamma_2), \quad (1)$$

quel est l'effet marginal de z_2 sur la probabilité que $y = 1$?

Question 2 (2 points): Dans le modèle suivant:

$$\Pr(y = 1 | z_1, z_2, d_1) = \Phi(z_1\delta_1 + z_2\gamma_1 + d_1\gamma_2 + z_2d_1\gamma_3), \quad (2)$$

(i) trouvez l'effet marginal de z_2 sur la probabilité que $y = 1$. (ii) trouvez l'effet marginal de d_1 sur la probabilité que $y = 1$.

On considère à présent une modélisation pour la variable qualitative dichotomique y_i . La spécification du modèle est la suivante:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{sinon,} \end{cases} \quad (3)$$

avec y_i^* une variable latente fonction de deux variables exogènes $x_{1,i}$ et $x_{2,i}$, soit:

$$y_i^* = \alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \varepsilon_i. \quad (4)$$

Les perturbations du modèle sont supposées i.i.d. et suivent une distribution logistique de moyenne nulle et de variance égale à 1.

Question 3 (2 points): Donnez l'expression du vecteur gradient ($G_{L_i}(\alpha)$) pour un individu i , en fonction des variables exogènes et des paramètres $\alpha = (\alpha_1, \alpha_2)'$.

Question 4 (3 points): Vous ne disposez que de façon partielle des résultats de l'estimation, à savoir $\hat{\alpha}_1 = 3$. Mais vous savez néanmoins que pour un individu de l'échantillon d'étude, $x_{1,i}$, $x_{2,i}$ et la première composante de $G_{L_i}(\alpha)$ sont respectivement égales à -1.5 , 2 et 1.4938 . On suppose que pour cet individu $y_i = 0$. Donnez la valeur estimée du paramètre α_2 .

Question 5 (1 point): Donnez trois raisons qui vont en défaveur de la spécification linéaire:

$$y_i = \alpha_1 x_{1,i} + \alpha_2 x_{2,i} + u_i. \quad (5)$$

Exercice 2: (6 points):

Soit l'exemple simplifié du choix de la filière d'étude après l'obtention du BAC en fonction de caractéristiques exogènes. Soit y une variable polytomique prenant quatre modalités résumant l'ensemble des choix de filières possibles (on évacue les problèmes d'exhaustivité):

$$y_i = \begin{cases} 1 & \text{si Economie/Droit} \\ 2 & \text{si Langues} \\ 3 & \text{si Maths/Physiques/Informatique} \\ 4 & \text{si Prépa,} \end{cases} \quad (6)$$

$i = 1, \dots, N$. Les variables exogènes disponibles pour modéliser le choix d'orientation dans une filière donnée sont: l'âge à l'obtention du BAC (x_1), le sexe (x_2) qui prend la valeur 1 si homme et 0 sinon (modèle sans constante). Les coefficients à estimer sont respectivement α_1 et α_2 .

Question 1 (2 points): On prend comme modalité de référence ($y_i = 4$). (i) Donnez l'expression des probabilités $P_j = \Pr(y_i = j)$, $j = 1, 2, 3, 4$. (ii) En

déduire l'expression de la Log-Vraisemblance pour un échantillon de taille N .

Question 2 (1 point): Donnez l'expression de l'effet marginal d'une variation unitaire de l'âge, sur la probabilité qu'un individu choisisse la filière "Langues".

Question 3 (1 point): Donnez l'expression de l'effet marginal du sexe sur la probabilité qu'un individu choisisse la filière "Economie/Droit".

Question 4 (2 points): Notons qu'il est parfois plus simple, surtout pour l'interprétation, de présenter le modèle en terme de cote ou (odds-ratio), soit $\frac{P_{j,i}}{P_{4,i}}$, $j = 1, 2, 3$. Donnez l'expression (i) de la cote $\frac{P_{j,i}}{P_{4,i}} \forall j = 1, 2, 3$, (ii) de la cote $\frac{P_{j,i}}{P_{4,i}}$ suite à une augmentation unitaire de l'âge, en fonction de la cote initiale.

Exercice 3: (5 points):

On considère un modèle tobit tel que:

$$y_i^* = \alpha + \beta x_i + \varepsilon_i, \quad \forall i = 1, \dots, N \quad (7)$$

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0, \end{cases} \quad (8)$$

où y_i^* désigne une variable latente, x_i désigne une variable dichotomique prenant deux valeurs $\{0; 1\}$, α et β désignent deux réels. On suppose que les erreurs ε_i sont *i.i.d.* et suivent une loi $N(0, \sigma_\varepsilon^2)$. On dispose d'un échantillon de 15 observations des variables x_i et y_i (dont 12 sont censurées), réparties de la façon suivante:

| Nombre d'observations | y_i | x_i |
|-----------------------|------------------|-----------|
| 9 observations | $y_i = 0$ | $x_i = 1$ |
| 3 observations | $y_i = 0$ | $x_i = 0$ |
| 1 observation | $y_i = \sqrt{4}$ | $x_i = 0$ |
| 1 observation | $y_i = \sqrt{2}$ | $x_i = 0$ |
| 1 observation | $y_i = 1$ | $x_i = 1$ |

Question 1 (1 point): Déterminez en fonction des paramètres α , β et σ_ε , la probabilité d'apparition de l'événement de censure, $y_i = 0$.

Question 2 (2 points): Montrez que la fonction de log-vraisemblance de ce modèle tobit peut s'écrire sous la forme:

$$\begin{aligned} \text{Log}L(y; \alpha, \beta, \sigma_\varepsilon^2) &= 9 \log \left[1 - \Phi \left(\frac{\alpha + \beta}{\sigma_\varepsilon} \right) \right] + 3 \log \left[1 - \Phi \left(\frac{\alpha}{\sigma_\varepsilon} \right) \right] - \frac{3}{2} \log (\sigma_\varepsilon^2) \\ &\quad - \frac{3}{2} \log (2\pi) - \frac{1}{2\sigma_\varepsilon^2} \left[(\alpha + \sqrt{4})^2 + (\alpha + \sqrt{2})^2 + (1 + \alpha - \beta)^2 \right], \end{aligned}$$

avec $\Phi(\cdot)$ la fonction de répartition de la loi $N(0,1)$.

Question 3 (2 points): On se propose à présent d'estimer les paramètres α , β et σ_ε par la méthode d'Heckman en deux étapes. Soit z_i une variable dichotomique telle que:

$$z_i = \begin{cases} 1 & \text{si } y_i > 0 \\ 0 & \text{sinon.} \end{cases} \quad (9)$$

Montrez que la vraisemblance du modèle probit associé à z_i , ne dépend que des ratios $\alpha/\sigma_\varepsilon$ et β/σ_ε et s'écrit:

$$\begin{aligned} \text{Log}L(z; \alpha/\sigma_\varepsilon, \beta/\sigma_\varepsilon) &= 2 \log \left[\Phi \left(\frac{\alpha}{\sigma_\varepsilon} \right) \right] + \log \left[\Phi \left(\frac{\alpha + \beta}{\sigma_\varepsilon} \right) \right] \\ &\quad + 9 \log \left[1 - \Phi \left(\frac{\alpha + \beta}{\sigma_\varepsilon} \right) \right] + 3 \log \left[1 - \Phi \left(\frac{\alpha}{\sigma_\varepsilon} \right) \right]. \end{aligned}$$