

STATISTIQUES (STA01) – EXAMEN

ENSTA, Université Paris-Saclay
Année 2025 – 2026

Vendredi 14 novembre 2025
10h – 12h (tiers-temps 10h – 12h30)

Une feuille A4 recto manuscrite est autorisée en tant que support. La calculatrice est autorisée.

Avant de commencer :

- *Le sujet comporte trois exercices. Il est demandé de les traiter dans l'ordre sur la copie.*
- *Des résultats pourront être admis d'une question sur l'autre, à la condition de l'écrire clairement.*
- *Un barème indicatif est donné :*
 - Exercice 1 (questions de cours) : 6 points,*
 - Exercice 2 : 9 points,*
 - Exercice 3 (QCM) : 5 points.*

Exercice 1 – Questions de cours 6 points)

Ces questions sont indépendantes. Répondre à chaque question en faisant les démonstrations demandées. On prendra soin de bien rédiger.

1. Énoncer et redémontrer la décomposition biais-variance du risque quadratique, pour l'estimation d'un paramètre θ en dimension $k \geq 1$. Si vous le souhaitez, on pourra accompagner la démonstration d'un dessin.
2. Soient X_1, \dots, X_n des variables aléatoires i.i.d., de moyenne μ , de variance finie $\sigma^2 > 0$, toutes deux inconnues. On note \bar{X} la moyenne empirique et $\hat{\sigma}^2$ l'estimateur non biaisé de la variance. Démontrer que $\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\hat{\sigma}^2}}$ converge en loi lorsque $n \rightarrow \infty$ vers une loi à préciser.
3. Dans le modèle linéaire $Y = X\theta + \varepsilon$, supposé identifiable, rappeler sans justification la forme explicite de l'estimateur des moindres carrés $\hat{\theta}$ en fonction de X et Y . Puis, redémontrer qu'il est sans biais et calculer sa matrice de covariance en fonction de X et σ^2 .

Exercice 2 – Modèle de Cobb-Douglas en économétrie 9 points)

Dans tout cet exercice, on pourra se référer au tableau de quantiles donnée à la fin de l'énoncé.

En 1928, les économistes Paul Douglas et Charles Cobb proposent un modèle de fonction de production, aujourd'hui connu sous le nom de modèle de Cobb-Douglas. Ce modèle décrit la relation entre la valeur produite V_i d'une entreprise (par exemple, en euros), son capital K_i (en euros) et sa force de travail L_i (nombre d'heures travaillées). Il suppose que la fonction de production de ces entreprises suit la forme suivante :

$$V_i = \lambda L_i^\beta K_i^\gamma,$$

où λ, β, γ sont des paramètres réels avec $\lambda > 0$. En passant au logarithme, tout cela se réécrit :

$$\log V_i = \alpha + \beta \log L_i + \gamma \log K_i$$

avec $\alpha = \log \lambda$. Le modèle linéaire associé au modèle de Cobb-Douglas est :

$$\log V_i = \alpha + \beta \log L_i + \gamma \log K_i + \varepsilon_i, \quad (1)$$

où les ε_i sont supposées i.i.d. de loi $\mathcal{N}(0, \sigma^2)$ où $\sigma^2 > 0$ est inconnu.

Dans cet exercice, on suppose qu'on observe les triplets (V_i, L_i, K_i) pour n entreprises et qu'ils suivent le modèle (1).

1. Écrire le modèle (1) sous sa forme matricielle $Y = X\theta + \varepsilon$ avec $\theta = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$ en précisant Y et X en fonction des V_i, L_i et K_i .

Pour les données récoltées auprès de 1658 entreprises, on a appliqué la méthode des moindres carrés et on a obtenu les résultats suivants :

$$\hat{\theta} = \begin{pmatrix} 3.136 \\ 0.738 \\ 0.282 \end{pmatrix} \quad \text{et} \quad \text{SCR}(\hat{\theta}) = \|Y - X\hat{\theta}\|^2 = 148.27.$$

On donne aussi :

$$(X^T X)^{-1} = \begin{bmatrix} 0.0288 & 0.0012 & -0.0034 \\ 0.0012 & 0.0016 & -0.0010 \\ -0.0034 & -0.0010 & 0.0009 \end{bmatrix}$$

2. Rappeler la forme de $\hat{\sigma}^2$, estimateur sans biais de σ^2 . Calculer ici sa valeur approchée.
3. Donner un intervalle de confiance au niveau 95% pour le coefficient β . On donnera d'abord sa forme générale avant l'application numérique.

On dit que les rendements sont :

- *constants* si, lorsqu'une entreprise double sa force de travail et double son capital, toutes choses égales par ailleurs, sa production double exactement ;
 - *croissants* si, lorsqu'une entreprise double sa force de travail et double son capital, toutes choses égales par ailleurs, sa production fait plus que doubler.
4. Traduire sur les coefficients α, β, γ les hypothèses de rendements constants ou croissants dans le cadre du modèle.
 5. On veut tester H_0 : les rendements sont constants, contre H_1 : les rendements sont croissants. Construire un test de niveau d'erreur 5% répondant à la question. Faire l'application numérique. Conclure.
 6. On a obtenu $\sum_{i=1}^n Y_i^2 = 16650.41$ et $\sum_{i=1}^n Y_i = 5199.49$. Calculer le coefficient du R^2 du modèle. Commenter.
 7. Le modèle étudié ici permet-il de mettre en lumière le lien de causalité entre la force de travail, le capital et la production ? Commenter.

$1 - \alpha$ \diagdown loi	$\mathcal{T}(k)$ avec $k \geq 1000$	$\mathcal{N}(0, 1)$
0.90	1.282	1.282
0.95	1.645	1.645
0.975	1.960	1.960

Tableau de quantiles d'ordre $1 - \alpha$ pour différentes lois.

Exercice 3 – QCM 5 points)

Ces questions n'ont aucun lien entre elles. Indiquer sur la copie, sans justification, l'unique bonne réponse à chaque question. Chaque bonne réponse rapporte 1 point, une mauvaise réponse ne rapporte ni n'enlève aucun point.

1. On veut tester H_0 contre H_1 . Supposons qu'on dispose d'un test de niveau d'erreur α , dont la zone de rejet s'écrit $\{T > -1.5\}$ pour T une statistique de test. Avec les données dont je dispose, T prend la valeur -0.8 .
 - A) Je conserve H_0 , avec une probabilité d'erreur d'au moins $1 - \alpha$.
 - B) Je rejette H_0 , avec une probabilité d'erreur que je ne contrôle pas.
 - C) La statistique de test T est toujours plus petite sous H_0 que sous H_1 .
 - D) Je m'expose à un risque de première espèce.
2. On se place dans un modèle où X_1, \dots, X_n sont des variables aléatoires i.i.d. réelles, de loi \mathbb{P}_θ de densité $x \mapsto f(x - \theta)$ par rapport à la mesure de Lebesgue sur \mathbb{R} , où f est une densité connue sur \mathbb{R} , et $\theta \in \mathbb{R}$ est un paramètre inconnu. On suppose que ce modèle est régulier et on note \mathbb{E}_θ l'espérance sous \mathbb{P}_θ . Dans ce modèle, l'information de Fisher :
 - A) est une variable aléatoire qui dépend de θ .
 - B) ne dépend pas de θ .
 - C) vaut $n \times \mathbb{E}_\theta[(\log f(X_1 - \theta))^2]$.
 - D) vaut $n \times \mathbb{E}_\theta \left[-\frac{f'(X_1 - \theta)}{f(X_1 - \theta)} \right]$.
3. Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires réelles strictement positives telle que $\sqrt{n}(Z_n - e) \rightarrow \mathcal{N}(0, 4)$, en loi, lorsque n tend vers ∞ . Alors, lorsque n tend vers ∞ , $\sqrt{n}(\log Z_n - 1)$ tend en loi vers :
 - A) $\mathcal{N}(0, 4e)$
 - B) $\mathcal{N}(0, 4)$
 - C) $\mathcal{N}(0, 4/e)$
 - D) $\mathcal{N}(0, 4/e^2)$

4. On se place dans un modèle où X_1, \dots, X_n sont des variables aléatoires i.i.d. de loi $\mathcal{N}(\mu, 1)$ où $\mu \in \mathbb{R}$ est un paramètre inconnu. On note \bar{X} la moyenne empirique des X_i et q_β le quantile d'ordre β de la loi $\mathcal{N}(0, 1)$. On veut tester : $H_0 : \mu = 0$ contre $H_1 : \mu = 1$. Soit $\alpha \in]0, 1[$. Le test de région de rejet R est uniformément plus puissant de niveau d'erreur α :

A) lorsque $R = \left\{ \exp \left(-\frac{1}{2} \sum_{i=1}^n (X_i - 1)^2 + \frac{1}{2} \sum_{i=1}^n X_i^2 \right) > nq_{1-\alpha}^2 \right\}$.

B) lorsque $R = \left\{ \sum_{i=1}^n (X_i - 1)^2 - \sum_{i=1}^n X_i^2 < q_{1-\alpha}^2/n \right\}$.

C) lorsque $R = \left\{ \bar{X} > q_{1-\alpha} \right\}$.

D) lorsque $R = \left\{ \bar{X} > q_{1-\alpha}/\sqrt{n} \right\}$.

5. Soit $Z \sim \mathcal{N}(0_d, I_d)$, et $A \in \mathbb{R}^{d \times d}$ une matrice de rang r . On définit la variable aléatoire Y de la façon suivante :

$$Y := Z - \arg \min_{u \in \mathbb{R}^d : Au = AZ} \|u\|^2.$$

A) Y et $Z - Y$ sont indépendants et $\mathbb{E}[\|Y\|^2] = d - r$.

B) Y et $Z - Y$ sont indépendants et $\mathbb{E}[\|Y\|^2] = r$.

C) Z et $Z - Y$ sont indépendants et $\mathbb{E}[\|Y\|^2] = d - r$.

D) Z et $Z - Y$ sont indépendants et $\mathbb{E}[\|Y\|^2] = r$.

Fin du sujet.