

Introduction à l'économétrie  
Examen final  
Durée : 2h

Décembre 2025

L'objet de cet examen est d'évaluer vos aptitudes à utiliser R, à estimer quelques régressions linéaires et à bien interpréter les résultats de ces estimations. Certaines questions théoriques peuvent être traitées sans code. *A fortiori*, elles attendent en général une réponse qualitative, plutôt qu'un calcul fastidieux.

Pour les questions qui exigent de votre part de coder, toute syntaxe R cohérente et qui aboutit au résultat attendu est acceptée.

Vos réponses prendront la forme d'un script R ou d'un fichier Rmarkdown bien structuré, que vous enverrez à 11h à pierre.pora[at]ensae.fr. L'objet de votre email sera [Introduction à l'économétrie] Examen final et votre fichier en pièce jointe sera intitulé NOM\_Prenom.

## Exercice 1 – Manipulation de données et statistiques descriptives

On utilisera les données CPS1985 du package **AER**, issues du Current Population Survey américain.

1. Charger les données dans R et les convertir en **data.table**.
2. Conserver uniquement les individus âgés de 25 à 55 ans.
3. Créer une variable indicatrice **female** valant 1 pour les femmes et 0 pour les hommes.
4. Calculer, pour chaque sexe :
  - le salaire horaire moyen (**wage**) ;
  - l'écart-type du salaire horaire ;
  - le nombre d'observations.
5. Calculer, pour chaque niveau d'éducation, le salaire horaire moyen.
6. Représenter graphiquement le salaire horaire moyen en fonction du niveau d'éducation. *La mise en forme de ce graphique n'a pas à être parfaite.*

On attend une utilisation claire des opérations par groupe.

## Exercice 2 – Estimation de régressions

On s'intéresse au lien entre salaire, éducation et genre.

1. Estimer par MCO la régression suivante :

$$\text{wage} = \alpha + \beta_1 \text{female} + \beta_2 \text{education} + \epsilon$$

2. Comment interpréter le coefficient  $\beta_2$  ?
3. Construire l'intervalle de confiance à 95 % de chacun des coefficients en utilisant la matrice de variance-covariance appropriée.
4. Expliquer le choix de cette matrice de variance-covariance. Quelle hypothèse fait la fonction `1m` par défaut ? Pourquoi ce choix est-il discutable ici ?
5. Tester au seuil de 5% l'hypothèse nulle selon laquelle le coefficient  $\beta_1$  est nul.
6. Ajouter une interaction entre le sexe et l'éducation dans la régression.
7. Comment interpréter le nouveau coefficient d'éducation ?
8. Peut-on exprimer l'ancien coefficient  $\beta_2$  en fonction des deux nouveaux coefficients ?

## Exercice 3 – Données de panel

On s'intéresse au lien entre état de santé et offre de travail. On utilisera les données `LaborSupply` du package `plm`.

1. Que représente la variable `id` ? Quel est l'intérêt de sa présence pour l'utilisation de ces données ?
2. Est-il possible d'estimer la régression :  $\lnhr_{it} = \beta_1\text{year}_{it} + \beta_2\text{age}_{it} + \alpha_i + \epsilon_{it}$ , où  $\alpha_i$  représente un effet fixe propre à chaque individu ? Pourquoi ?
3. Estimer la régression  $\lnhr_{it} = \gamma\text{disab}_{it} + \lambda_i + \nu_{it}$  en utilisant l'estimateur *within* pour les régressions avec des effets fixes.
4. Comment s'interprète le coefficient  $\gamma$  ? Quel est l'intérêt de la spécification de la variable dépendante en logarithme ? Sur quelle sous-population porte ce coefficient ? Quelle est la part de cette sous-population dans la population totale ?
5. Construire l'intervalle de confiance à 95% sur le coefficient  $\gamma$ . Quelle est la matrice de variance-covariance à utiliser ici ? Pourquoi ?
6. Quelle serait une stratégie alternative pour estimer le coefficient  $\gamma$  ? Mettez-là en œuvre. Dans quel cas coïnciderait-t-elle exactement avec l'estimateur *within* ?
7. Quelle serait la différence avec la régression  $\lnhr_{it} = \eta\text{disab}_{it} + \mu_t + \omega_{it}$  ? Comment s'interprète le coefficient  $\eta$  dans le cas présent ?
8. Quelle est la  $p$ -valeur sur le coefficient  $\eta$  avec un test de Student ? Quelle est la matrice de variance-covariance à utiliser ici ? Pourquoi ?