Ecole Normale Supérieure
FIMFA – Statistique 2011-2012

Feuille d'exercices n°6.

Exercice 1. Soit R le revenu d'un individu. On veut analyser la différence de revenu selon le sexe. Pour ceci, on considère le modèle

$$R = \mu + \beta s + \varepsilon$$

où s est une variable qui vaut 1 si l'individu est un homme et 0 si c'est une femme, et ε est une variable gaussienne centrée de variance connue σ^2. Les paramètres (μ, β) sont inconnus. On observe des réalisations $(R_i, s_i)_{1\leq i\leq n}$ i.i.d. suivant le modèle précédent. Dans la suite, $(s_1, ..., s_n)$ est considérée comme déterministe.

1. Comment s'interprète le coefficient β ? Que se passe-t-il si $\beta = 0$, $\beta < 0$?
2. Donner une condition sur $(s_1, ..., s_n)$ pour que l'estimateur du maximum de vraisemblance de (μ, β) soit bien défini et préciser alors son expression et sa loi.
3. Proposer un intervalle de confiance pour β. En déduire un test $H_0 : \beta = 0$ contre $H_1 : \beta \neq 0$ de niveau α.
4. On veut tester $H_0 : \beta \leq 0$ contre $H_1 : \beta > 0$. Proposer un test de niveau α et le comparer au test de la question précédente.

Exercice 2. Nous disposons pour n entreprises, $i = 1, ..., n$, des valeurs du capital K_i, de la valeur ajoutée VA_i et de la quantité de travail L_i. Nous supposons que la fonction de production F de ces entreprises est du type Cobb-Douglas, c-à-d

$$VA_i = F(L_i, K_i) = e^{\alpha} L_i^{\lambda} K_i^{\gamma}.$$

Les paramètres α, λ, γ sont inconnus.

1. Ecrire un modèle de régression linéaire associé lorsque les erreurs sont supposées gaussiennes indépendantes, centrées et de variance commune σ^2 (inconnue). On pourra écrire ce modèle sous la forme

   $$Y = X\beta + E.$$

   Rappeler alors l'expression matricielle de l'estimateur des moindre carrés ordinaires $\hat{\beta}$ de β et préciser sa matrice de variance-covariance.

   Dans la suite, on utilisera les données suivantes : pour 1658 entreprises, on a obtenu à l'aide de l'estimateur MCO de β, la formule de régression

   $$\log VA_i = 3.136 + 0.738 \log L_i + 0.282 \log K_i$$

et SCR(=somme des carrés des résidus)= 148.27. Nous donnons également

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

2. Donner un estimateur $\hat{\sigma}^2$ sans biais de σ^2 et un estimateur sans biais de $\mathrm{Var}(\hat{\beta})$.
3. Donner un intervalle de confiance de niveau 95% pour les paramètres α et λ.
4. Tester l'hypothèse $H_0 : \gamma = 0$ contre $H_1 : \gamma > 0$.
5. On souhaite à présent tester l'hypothèse selon laquelle la fonction de production F est à rendements d'échelle constants, c-à-d

$$F(\mu L, \mu K) = \mu F(L, K), \quad \forall \mu \in \mathbb{R}_+.$$

   Proposer un test de niveau 5% pour H_0 : "rendements d'échelle constants" contre H_1 : "rendements d'échelle croissants."
6. Nous nous intéressons à la fonction de production dans trois secteurs

   T07 : minerais et métaux ferreux, première transformation de l'acier
   T15B : fabrication de biens d'équipement ménager
   T18 : industries textiles et habillement.

   Ces trois secteurs sont-ils régis par la même fonction de production ? Pour le savoir, nous avons estimé par la méthode des moindres carrés ordinaires les paramètres de l'équation $\log VA = \alpha + \lambda \log L + \gamma \log K$ séparément sur chacun des trois secteurs, puis sur l'ensemble des entreprises. Nous avons obtenus les résultats suivants :

   T07 : $n = 40$, SCR= 2.57
   T15B : $n = 24$, SCR= 1.17
   T18 : $n = 238$, SCR= 17.8
   Ensemble : $n = 302$, SCR= 41.12

   Pour un niveau de 5%, tester l'hypothèse selon laquelle la fonction de production est la même dans les trois secteurs.

Exercice 3. Nous avons estimé le paramètre vectoriel inconnu θ_1 du modèle de régression linéaire suivant

$$Y = W\theta_1 + U$$

où U est un vecteur gaussien centré de matrice de variance-covariance $\sigma_U^2 I_n$, alors que le vrai modèle est

$$Y = W\theta_1 + Z\theta_2 + V$$

avec V un vecteur gaussien centré de matrice de variance-covariance $\sigma_V^2 I_n$. On notera p_1 (resp. p_2) la dimension du vecteur θ_1 (resp. θ_2) et on supposera que

$$\mathrm{rang}(X) := \mathrm{rang}(W, Z) = p = p_1 + p_2.$$

1. Calculer $\mathbb{E}\big[\hat{\theta}_1\big]$, où $\hat{\theta}_1$ est l'estimateur des moindres carrés ordinaires du paramètre θ_1 du modèle estimé.
2. Plus généralement, considérons le cas d'un modèle de régression linéaire

$$Y = X\beta + E,$$

avec E vecteur gaussien centré de matrice de variance-covariance $\sigma_E^2 I_n$, sous la contrainte $R\beta = r$, où le vecteur r et la matrice R sont connus.

(a) Dans le cas du problème d'omission de variables mentionné ci-dessus, écrire le modèle sous contrainte.

(b) Dans le cadre général d'un modèle sous contrainte, si on note $\hat{\beta}$ l'estimateur MCO de β sans la contrainte et $\hat{\beta}^c$ l'estimateur MCO de β en considérant la contrainte, montrer que

$$\hat{\beta}^c = \hat{\beta} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r - R\hat{\beta}).$$

(c) Calculer $\mathbb{E}\big[\hat{\beta}^c\big]$.

(d) Montrer que $\mathrm{Var}(\hat{\beta}^c) = \sigma_E^2\Big((X'X)^{-1} - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big)$. Comparer $\mathrm{Var}(\hat{\beta}^c)$ et $\mathrm{Var}(\hat{\beta})$.

(e) Interpréter ce résultat dans le cas d'omission de variables.