Examen TISD - Master 1 I.M.
Vendredi 13 Janvier 2011

Durée 2 h. Calculatrice et formulaire (2 copies doubles) autorisés.
Tous les tests seront faits au seuil 5%. On rappellera les hypothèses nulle et alternative, on donnera la valeur de la statistique de test et la p-valeur.

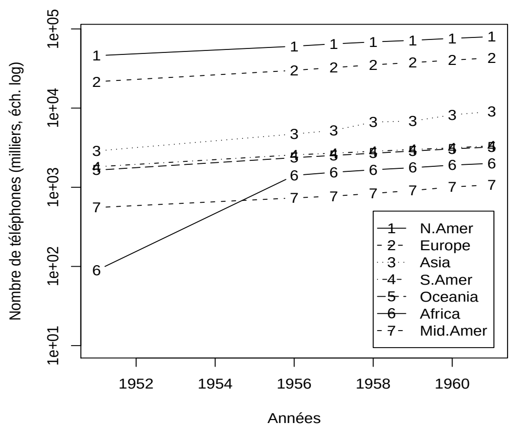
Tran Viet Chi, chi.tran@univ-lille1.fr, bureau 316 (bâtiment M3).

Exercice 1 (Application du cours : régression Probit)
On suppose qu'on a n individus indépendants caractérisés par les variables (X_i, Y_i), toutes deux à valeurs dans $\{0,1\}$ et vérifiant :

$$Y_i = \mathbf{1}_{\{Y_i^*>0\}} \text{ où } Y_i^* = aX_i + b + \varepsilon_i$$

où les ε_i sont des variables Gaussiennes $\mathcal{N}(0,\sigma^2)$ indépendantes, avec σ supposé connu. Calculer les estimateurs du maximum de vraisemblance de a et b. On introduira entre autres la fonction de répartition Φ de la loi $\mathcal{N}(0,1)$.

Exercice 2 (Moshi Moshi ?)
On s'intéresse au nombre de téléphones par continents entre 1951 et 1961. Les chiffres de la figure suivante sont en milliers :



On note $Y_{i,t}$ le log du nombre de téléphones dans le continent $i \in \{1,\dots,7\}$ au temps $t \in \{1951, 1956, \dots, 1961\}$ (les années 1952 à 1955 ne sont pas observées). On a 7 années d'observations pour chacun des 7 continents, ce qui fait 49 lignes dans la base de données. Chaque ligne correspond à un couple (i,t), les variables renseignées sont le log du nombre de téléphones $Y_{i,t}$, le continent i (variable qualitative) et l'année t.
On veut choisir entre 2 modélisations possibles, (1) et (2). Soit on considère que chaque pays suit un modèle qui lui est propre, et

$$Y_{i,t} = \alpha_i + \gamma \mathbf{1}_{\{\text{i=Afrique, t=1951}\}} + \beta_i \times t + \varepsilon_{i,t}, \quad i \in \{1,\dots,7\},\ t \in \{1951,\dots,1961\} \quad (1)$$

où $\varepsilon_{i,t}$ est un résidu qu'on supposera indépendant et de loi $\mathcal{N}(0,\sigma^2)$. Soit on considère que les coefficients β_i sont les mêmes et égaux à β quel que soit le continent :

$$Y_{i,t} = \alpha_i + \gamma \mathbf{1}_{\{\text{i=Afrique, t=1951}\}} + \beta \times t + \eta_{i,t} \quad i \in \{1,\dots,7\},\ t \in \{1951,\dots,1961\} \quad (2)$$

où $\eta_{i,t}$ est un résidu indépendant de loi $\mathcal{N}(0, \rho^2)$.
Dans la suite, on utilisera les notations $T = ^t(1951, 1956, \dots, 1961)$, $e = ^t(1, \dots, 1) \in \mathbb{R}^7$, $Y_i = ^t(Y_{i,1951}, Y_{i,1956}, \dots, Y_{i,1961})$, $\varepsilon_i = ^t(\varepsilon_{i,1951}, \dots, \varepsilon_{i,1961})$ et $\eta_i = ^t(\eta_{i,1951}, \dots, \eta_{i,1961})$.

1. Pouvez-vous expliquer pourquoi dans les régressions (1) et (2), on a inclus l'indicatrice $\mathbf{1}_{\{\text{i=Afrique, t=1951}\}}$?

2. Ecrire matriciellement le système contraint (2) sous la forme

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_7 \end{pmatrix} = \text{Matrice} \times \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_7 \\ \beta \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_7 \end{pmatrix} + \gamma \mathbf{1}_{\{\text{i=Afrique, t=1951}\}}$$

où Matrice est une matrice blocs qu'il faut écrire en fonction de e et T. Ecrire de même matriciellement le système non-contraint (1) avec des matrices blocs constituées des Y_i, ε_i, e et T.

Etude du modèle (2)

3. Pour faire la régression (2) sous **SAS**, on inclut dans la régression des indicatrices qui valent 1 respectivement quand le contient est l'Amérique du Nord (`N_Amer`), l'Asie (`Asia`), l'Amérique du Sud (`S_Amer`), l'Océanie (`Oceani`), l'Afrique (`Africa`), l'Amérique Centrale (`Mid_Am`). Pourquoi n'y a-t-il pas d'indicatrice pour l'Europe ?

4. Les résultats de cette régression sont présentés en Annexe 1. Certains résultats ont été perdus. Pouvez-vous recalculer le R^2, ρ^2 et la statistique de Student pour le coefficient associé à l'Amérique Centrale ? (valeurs grisées dans l'Annexe 1)

5. Commenter la significativité et la qualité de la régression.

6. Commenter la significativité des coefficients. Que cela signifie-t-il par rapport à l'Europe ?

Test d'homogénéité dans le modèle (1). On voudrait tester l'égalité des β_i, $i \in \{1, \dots, 7\}$. Il ne s'agit pas tout à fait d'un test de Chow, puisqu'on permet aux constantes de conserver des valeurs différentes, il s'agit d'un cas particulier de *modèle de panel* dit *à effets fixes*.

7. Dans le modèle non contraint (1), le paramètre est $(\gamma, \alpha_1, \dots, \alpha_7, \beta_1, \dots, \beta_7)$. Le test qui nous intéresse est : "$H_0 : \forall i \in \{2, \dots, 7\}$, $\beta_i = \beta_1$" contre l'hypothèse alternative "Deux des β_i au moins sont distincts". Ecrire la matrice R associée à la contrainte définie dans H_0. Quel est son rang ? En déduire la loi asymptotique de la statistique de test sous H_0. Et sous l'hypothèse alternative ?

8. A partir des sorties de l'Annexe 2, dire si on peut considérer que l'augmentation du nombre de téléphones est homogène sur l'ensemble des continents.

9. Quels sont les continents pour lesquels l'évolution de Y est la plus différente ? Que dire si on les exclut de l'échantillon ?

Estimateurs MCO pour un panel à effets fixes. On se propose ici d'étudier un peu plus précisément les estimateurs des moindres carrés pour une régression du type de (2) :

$$Y_{i,t} = \alpha_i + \beta \times t + u_{i,t} \tag{3}$$

où les $u_{i,t}$ sont des résidus $\mathcal{N}(0, v^2)$ indépendants. On notera $u_i = ^t(u_{i,1951}, u_{i,1956}, \dots, u_{i,1961})$.

10. Montrer que la somme des carrés des résidus $S = \sum_{i,t} u_{i,t}^2$ peut se réécrire comme $\sum_{i=1}^{7} \|u_i\|^2$.

11. En écrivant u_i en fonction de Y_i, e, T et des coefficients α_i et β, montrer que la minimisation de S nous amène à choisir :

$$\widehat{\alpha}_i = \bar{Y}_i - \widehat{\beta}\bar{T}$$

où $\bar{Y}_i$ est la moyenne des composantes de Y_i et $\bar{T}$ la moyenne des composantes de T.

12. On introduit l'opérateur *within* : $W = Id_{\mathbb{R}^7} - e\ {}^t e/7$. Calculer WY_i en fonction des $Y_{i,t}$. Quelle est la fonction de l'opérateur W ?

13. Montrer que $WY_i = \beta WT + Wu_i$. Que peut-on dire des Wu_i ?

14. En déduire que l'estimateur des moindres carré de β est :

$$\widehat{\beta} = \frac{\sum_{i=1}^{7} {}^tTWY_i}{7\ {}^tTWT}$$

15. Montrer que cet estimateur est aussi celui obtenu si on fait la régression MCO sur :

$$Y_{i,t} - \bar{Y}_i = \beta\big(T - \bar{T}\big) + w_{i,t}$$

où $w_{i,t}$ sont des résidus indépendants.

Exercice 3 (Modèle AR(1))
On considère une série temporelle stationnaire au second ordre, $(X_t)_{t\in\mathbb{Z}}$, suivant le modèle AR(1) suivant :

$$X_t = aX_{t-1} + \varepsilon_t$$

où $0 < a < 1$ et où les $(\varepsilon_t)_{t\in\mathbb{Z}}$ sont des variables Gaussiennes $\mathcal{N}(0, \sigma^2)$ i.i.d. On notera L l'opérateur retard : $LX_t = X_{t-1}$.

1. Donner le polynôme $A(X)$ tel que $A(L)X_t = \varepsilon_t$. Quel est le degré et la racine de ce polynôme ?

2. Soit $\mu = \mathbb{E}(X_t)$. Calculer μ.

3. On note $\gamma(h)$, pour $h \in \mathbb{Z}$ la fonction d'auto-covariance de $(X_t)_{t\in\mathbb{Z}}$. Ré-établir les équations de Yule-Walker.

4. En déduire $\gamma(h)$ en fonction de h, σ^2 et a. Quel est le comportement de $\gamma(h)$ quand h tend vers l'infini ?

5. En quoi le résultat précédent diffère-t-il de ce qui est connu sur les processus MA(q) ?

6. Déduire de la question 4 un estimateur de a en fonction de la variance empirique S^2 des X_t et de σ^2.