



Manipulation de données

Exercice 1:

On considère les données suivantes :

menage	individu	sexe	depense
1	1	Homme	20000
1	2	Femme	15000
1	3	Homme	40000
2	1	Femme	60000
2	2	Femme	10000
3	1	Homme	25000
4	1	Homme	30000
4	2	Femme	20000
4	3	Homme	15000
5	1	Femme	70000

1. Créer l'environnement de travail "DossTravail" faisant référence à votre dossier de travail. Créer un dossier "Exo1" dans cette environnement.
2. Entrer les données du tableau dans Stata. Sauvegarder les données lues dans une base "InfoMenage.dta" dans le dossier "Exo1".
3. Calculer la dépense moyenne par sexe.
4. A l'aide de la commande "egenrate" centrer et réduire la variable dépense, puis générer une variable x contenant un numéro différent pour chaque modalité de la variable sexe.
5. créer une base ménage contenant pour chaque observation le numéro du ménage et la dépense totale. Sauvegarder dans une base "DepenseMenage.dta". Effacer la mémoire.

Exercice 2:

On considère les mêmes données que dans l'exercice 1.

1. Charger la base "InfoMenage.dta".
2. Lister:
 - (a) les 6 premières données de la base;
 - (b) les 6 premières dépenses de la base.
3. Décrire:
 - (a) la base;
 - (b) les dépenses de la base.
4. Restructurer les données en format wide. Restructurer les données en format long. Normalement vous constatez des données manquantes. Dans ce cas utiliser "drop" pour éliminer ces mauvaises observations.
5. Faire une description plus détaillée des dépenses.
6. Faire une analyse croisée des données avec les variables "menage" et "depense".
7. Afficher la matrice de corrélation entre les variables "menage" et "depense".
8. Tracer l'histogramme des "depense".
9. Tracer la courbe des "depense" en fonction des "menage".
10. Afficher la matrice de corrélation entre toutes les variables.

Statistiques, économétrie et graphiques

Exercice 3:

1. Créer l'environnement de travail "DossTravail" faisant référence à votre dossier de travail. Créer un dossier "Exo1" dans cette environnement.
2. Définir une variable "DossData" faisant référence au dossier contenant les bases de données sur lesquelles travailler.
3. Copier la base "timeseries_ppi.dta" depuis le "\$DossData" vers "\$DossTravail".
4. Charger la base "timeseries_ppi.dta" depuis le répertoire "\$DossTravail".
5. Créer un fichier dofile pour y mettre toutes les commandes qui vont suivre.
6. Analyser les données en utilisant list, describe et summarize(su).
7. Écrire la commande qui permet de définir :
 - la variable ypiib = gdp : la variable d'intérêt qui représente le produit intérieur brut
 - la variable temps = t : la fréquence de temps à laquelle sont enregistrées les données
8. Écrire la commande qui permet de donner les statistiques descriptives détaillées de la variable ypiib.
9. Écrire la commande qui permet de faire la représentation graphique de la variable ypiib en fonction de la variable temps.

10. Écrire la commande qui permet de faire le test de Dickey-Fuller sur la stationnarité de la variable ypb en incluant une tendance dans le test.
11. Calculer les autocorrélations totales sur la valeur à niveau, sur la valeur différenciée.
12. Calculer les autocorrélations partielles sur la valeur à niveau, sur la valeur différenciée.
13. Estimer les modèles arima (1,0,0), (2,0,0), (1,0,1), (0,1,1), (2,1,3).

Régression linéaire et non linéaire

Exercice 4:

On considère des informations de valeurs immobilières en fonction des variables explicatives: superficies utile, bureaux, entrées et âge.

Superficie utile (x1)	Bureaux (x2)	Entrées (x3)	Age (x4)	Valeur immobilière (y)
2310	2	2	20	142000
2333	2	2	12	144000
2356	3	1.5	33	151000
2379	3	2	43	150000
2402	2	3	53	139000
2425	4	2	23	169000
2448	2	1.5	99	126000
2471	2	2	34	142900
2494	3	3	23	163000
2517	4	4	55	169000
2540	2	3	22	149000
2612	4	3	23	
2574	3	3.5	30	
3120	5	4	50	

1. Saisir les données.
2. Faire la regression de la variable dépendante y en fonction des variables explicatives x1, x2, x3 et x4.
3. Prédire les valeurs immobilières manquantes.

Exercice 4:

On considère des informations de valeurs immobilières en fonction des variables explicatives: superficies utile, bureaux, entrées et âge.

x	y
-10	5.329190259
-9	4.164759769
-8	11.61482388
-7	4.497516367
-6	11.14326286
-5	8.425721127
-4	17.74669844
-3	16.89998326
-2	22.84311893
-1	32.8087403
0	42.8666155
1	39.0084578
2	49.95429101
3	53.86797608
4	47.58443844
5	51.80471201
6	53.53105708
7	57.13742719
8	52.01842044
9	51.1135302
10	57.92977362

1. Saisir les données.
2. Faire la regression non linéaire de la variable dépendante y en fonction de variables explicative x en utilisant une fonction logistique $y = \frac{b}{(c + \exp(-ax))}$.
3. Générer les predictions ypred en définissant la fonction logistique avec les paramètres a, b, c déterminés précédemment.
4. Tracer le graphique de la courbe ypred en fonction de x et des points (x,y).