

Statistiques : Enoncé du TP 1 Régression linéaire

L'exercice utilise des indices annuels de production agricole (2e ligne) et de production industrielle (3e ligne) relevés entre les années 1944 et 1961. On se propose de rechercher une éventuelle relation linéaire entre ces deux indices par la méthode des moindres carrés ordinaires.

1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961
100	60	76	74	90	93	102	98	103	110	117	118	112	115	116	121	134	130
100	50	84	99	113	122	128	143	145	146	159	172	188	204	213	220	242	254

1. Saisie des données : Après avoir lancé Scilab, saisir ces deux suites d'indices sous la forme de deux vecteurs ligne à 18 composantes en tapant les instructions :

```
agri=[100,60, . . . . , 134,130]  
indu=[100,50, . . . . ,242,254]
```

Vérifier votre saisie en tapant par exemple `agri(7)` ou `indu(14)`.

2. Tracé du nuage de points et de son centre de gravité : L'instruction

```
for i=1 :18, plot2d(agri(i),indu(i),-4), end
```

permet de tracer le nuage de points correspondant dans le plan (agri X indu), dans une fenêtre graphique numérotée 0 par défaut. Minimiser cette figure en bas d'écran (mais ne la supprimer pas!) puis, revenant à la fenêtre de travail de Scilab, calculer les coordonnées du centre de gravité du nuage par

```
xbarre=sum(agri)/18;ybarre=sum(indu)/18
```

puis ajouter ce point à votre graphique par

```
plot2d(xbarre,ybarre,-2)
```

enfin ajouter un titre (et le nom des axes) au graphique par

```
xtitle(['Le nuage et son centre de gravité'],'agri','indu')
```

3. Gestion des figures : Pour sauvegarder la figure précédente et celles que vous réaliserez ensuite, ouvrir un éditeur de votre choix (OpenOffice, Word, ...) et dans le menu **Fichier** de la fenêtre graphique, copier dans le presse papier **EnhMetafile** et coller le résultat dans l'éditeur. Il devrait y avoir 4 graphiques dans cet éditeur en fin de séance. Par la commande

```
xset("window",1);
```

ouvrez une nouvelle fenêtre graphique par défaut, numérotée 1 (et non 0), dans laquelle seront réalisés les prochains tracés que vous ferez, aussi longtemps qu'une nouvelle instruction de ce type, avec un numéro différent de 1 ne vient pas changer de fenêtre par défaut.

4. Calcul de la droite de régression : La liste des instructions suivantes permet de calculer la pente et l'ordonnée à l'origine de la droite des moindres carrés :

```
for i=1 :18,prodxy(i)=(agri(i)-xbarre)*(indu(i)-ybarre),end  
for i=1 :18,prodxx(i)=(agri(i)-xbarre)^2,end  
for i=1 :18,prody(i)=(indu(i)-ybarre)^2,end  
achapeau=sum(prodxy)/sum(prodxx) bchapeau=ybarre-achapeau*xbarre
```

5. Tracé de la droite de régression : Avant de tracer la droite dans le graphique 1, y tracer comme dans le graphique 0 le nuage et son centre de gravité en exécutant à nouveau les 2 instructions correspondantes. Puis la tracer au moyen des instructions suivantes :

```
function y=f(x);y=achapeau*x+bchapeau;endfunction  
fplot2d(min(agri) :1 :max(agri),f)
```

6. Ajout de la deuxième droite de régression : Reprendre les deux questions précédentes, calcul puis tracé, pour la droite de régression de agri sur indu (et non plus de indu sur agri). Une fois la figure obtenue comportant les deux droites, lui ajouter un titre et le nom des axes puis la sauvegarder dans votre éditeur à coté de la figure 0.

7. Niveau de corrélation linéaire du nuage : Si l'on désigne par R_{deux} le carré du coefficient de corrélation linéaire, calculer sa valeur en utilisant les vecteurs `prodxy`, `prodxx`, et `prody` introduits précédemment.

8. Etude des résidus : Définir un vecteur $res(i), i = 1, \dots, 18$ dont les composantes sont les résidus (écart entre les valeurs observées et les valeurs prédites). L'instruction

```
xset("window",2);for i=1 :18 plot2d(i,res(i),-1),end
```

permet de tracer les résidus dans une troisième fenêtre graphique. Ajouter à cette figure, l'axe horizontal (qu'on pourra voir comme le graphe d'une fonction nulle), puis un titre et le nom des axes, puis la sauvegarder sur la même feuille que les deux premières.

Pour tracer un histogramme des résidus, on pourra utiliser l'instruction

```
xset("window",3);r=[-45 :10 :45];histplot(r,res)
```

9. Etude de la dispersion du nuage : A l'aide des formules données en cours, calculer les dispersions DA, DT et DR et vérifier la formule $R_{deux} = DR/DT$ pour ce jeu de données.

10. Valeurs observées/valeurs prédites : En ne retenant que le premier modèle (celui de la première droite), calculer pour les années 1947 et 1950, les indices de production industrielle prédits par le modèle et les comparer aux indices observés.

11. Sensibilité aux valeurs extrêmes : On dit que la position de la droite de régression est très sensible par rapport aux valeurs aberrantes. Faites une expérience pour illustrer cette affirmation en changeant l'un des indices observés.