Exercices de statistiques descriptives
Série 2: Exercices avec indications

P. Floquet, J. Gergaud

Exercice 1. *Le tableau suivant donne les revenus imposables des Français en 1970.*

Classes	*Fréquences relatives*
[2500;5000[	*0.0067*
[5000;10000[	*0.3018*
[10000;15000[	*0.2750*
[15000;20000[	*0.1709*
[20000;30000[	*0.1445*
[30000;50000[	*0.0701*
[50000;70000[	*0.0166*
[70000;100000[	*0.0081*
[100000;200000[	*0.0051*
[200000;400000[	*0.0010*

1. *tracer l'histogramme de ces données pour les revenus allant de 0 à 7000.*

Indications Attention les intervalles de classes ne sont pas constants.

Exercice 2. *On désire tester n produits. On fait appel à 2 goûteurs et on leur demande de classer ces n produits. Nous avons donc à notre disposition une série statistique double:*

$$\begin{array}{c} x_1, x_2, \dots, x_n \\ y_1, y_2, \dots, y_n \end{array}$$

avec:

$$\{x_1, x_2, \dots, x_n\} = \{y_1, y_2, \dots, y_n\} = \{1, 2, \dots, n\}$$

On appelle coefficient de Spearman le coefficient de corrélation linéaire:

$$r_s = \frac{cov(x,y)}{s_x s_y}$$

1. *Montrer que*

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

où $d_i = x_i - y_i$.

2. *Que signifie $r_s = 1, r_s = -1, r_s = 0$?*

Indication

1. On rappelle que la somme des n premiers entiers est égale à $n(n+1)/2$ et que la somme des carrés des n premiers entiers est égale à $n(n+1)(2n+1)/6$.

   On calculera SCE_x en fonction de n

Exercice 3. *Dans une solution aqueuse contenant un polluant, on plonge un solide adsorbant (charbon actif sous forme de tissu) qui " capture " une partie des molécules de la substance polluante. Au bout d'un certain temps, le système est à l'équilibre: Chaque point d'équilibre est caractérisé par la concentration à l'équilibre C_e et la quantité de polluant adsorbé par unité de masse de charbon actif, q_e. A une température donnée, on peut mesurer différents points sur une courbe (C_e, q_e) dite isotherme d'adsorption. Le tableau suivant fournit l'isotherme d'adsorption de l'aniline à 25 °C (Faur-Brasquet, 1998).*

C_e *(mg/l)*	*72*	*57,7*	*38,5*	*21,3*	*13,1*	*6,9*	*3,9*	*1,2*
q_e *(mg/g)*	*232,5*	*211*	*192*	*163,4*	*136,7*	*116,3*	*96,2*	*61,9*

Etudier la liaison entre q_e *et* C_e *en supposant que les incertitudes expérimentales sur* C_e *sont négligeables devant celles sur* q_e*. On fera le graphique en " nuage de points " des valeurs de* q_e *en fonction des valeurs de* C_e*. On étudiera ensuite les deux modèles suivants :*

Modèle de Langmuir $q_e = \frac{q_m b C_e}{1+bC_e}$

Modèle de Freundlich $q_e = KC_e^{1/n}$

Pour chacun des deux modèles, on estimera les paramètres du modèle (q_m *et* b*,* K *et* n*) par régression linéaire simple sur des variables " modifiées ".*

Conclure sur " l'adéquation " des 2 modèles proposés.

Indications

On prendra pour le modèle de Langmuir les variables $x_L = 1/C_e$ et $y_L = 1/q_e$ et pour le modèle de Freundlich les variables $x_F = \ln C_e$ et $y_F = \ln q_e$. On donne

$$\sum_i x_{Li} = 1.4151 \quad \sum_i y_{Li} = 0,0628 \quad \sum_i x_{Li}^2 = 0.7904 \quad \sum_i x_{Li} y_L i = 0,0185$$

$$\sum_i x_{Fi} = 21,0887 \quad \sum_i y_{Fi} = 39,5204 \quad \sum_i x_{Fi}^2 = 69,6526 \quad \sum_i x_{Fi} y_{Fi} = 108,5925$$