

Exercices de statistiques descriptives

Série 1: Exercices avec corrections

P. Floquet, J. Gergaud

Exercice 1. On observe les arrivées des clients à un bureau de poste pendant un intervalle de temps donné (10 minutes). En répétant 100 fois cette observation, on obtient les résultats suivants.

Nombre d'arrivées	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	Total
Nombre d'observations	<i>15</i>	<i>25</i>	<i>26</i>	<i>20</i>	<i>7</i>	<i>7</i>	100

1. Représenter graphiquement ces résultats.
2. Calculer la valeur de la moyenne arithmétique, de la médiane, de la variance et de l'écart-type des résultats, du coefficient de variation, de l'amplitude et de l'écart moyen des observations.

Correction

1. Le diagramme en bâtons des fréquences relatives est le suivant :

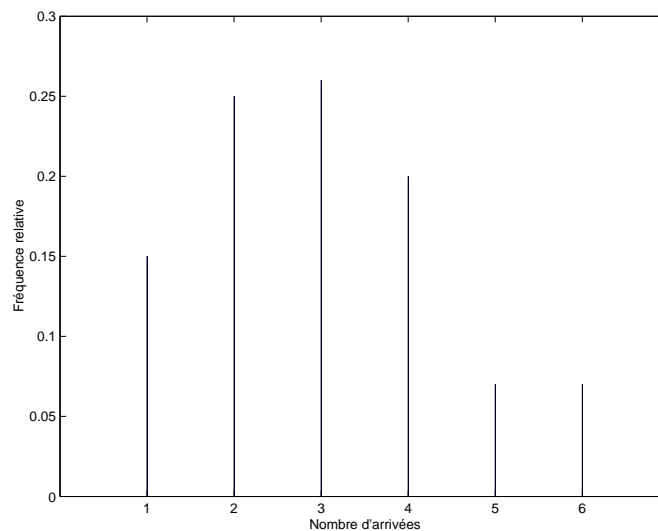
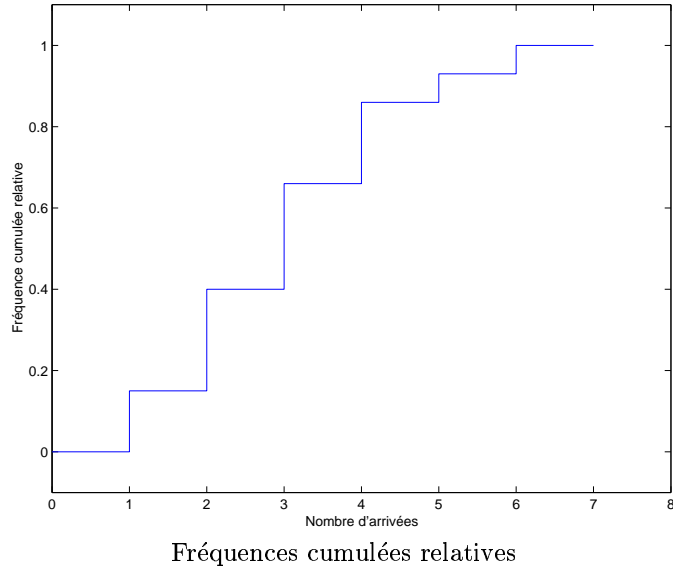


Diagramme en bâtons

Le diagramme des fréquences cumulées relatives est le suivant :



2. La moyenne arithmétique vaut :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i x_i = 3. \text{ La médiane et le mode valent } 3.$$

La variance vaut :

$$s^2 = \frac{1}{n} \left(\sum_{i=1}^n n_i x_i^2 - n \bar{x}^2 \right) = 1,96 \text{ et l'écart-type } s = 1,4. \text{ Le coefficient de variation est égal à } cv = 46,7 \text{ \%.}$$

L'amplitude des observations est de 5, l'écart moyen de

$$e_m = \frac{1}{n} \sum_{i=1}^n n_i |x_i - \bar{x}| = 1,10.$$

Exercice 2. On considère les compositions chimiques de 20 eaux minérales suivantes ¹:

Origines	HCO_3^-	SO_4^-	Cl^-	Ca^+	Mg^+	Na^+
Aix-les-Bains	341	27	3	84	23	2
Beckerish	263	23	9	91	5	3
Cayranne	287	3	5	44	24	23
Chambon	298	9	23	96	6	11
Cristal-Roc	200	15	8	70	2	4
St Cyr	250	5	20	71	6	11
Evian	357	10	2	78	24	5
Ferita	311	14	18	73	18	13
St Hyppolite	256	6	23	86	3	18
Laurier	186	10	16	64	4	9
Ogeu	183	16	44	48	11	31
Ondine	398	218	15	157	35	8
Perrier	348	51	31	140	4	14
Ribes	168	24	8	55	5	9
Spa	110	65	5	4	1	3
Thonon	332	14	8	103	16	5
Veri	196	18	6	58	6	13
Viladreau	59	7	6	16	2	9
Vittel	402	306	15	202	36	3
Volvic	64	7	8	10	6	8

1. Calculer les quartiles Q_1, Q_2, Q_3 , l'écart interquartile et les limites des moustaches pour les variables HCO_3^- et Ca^+ .

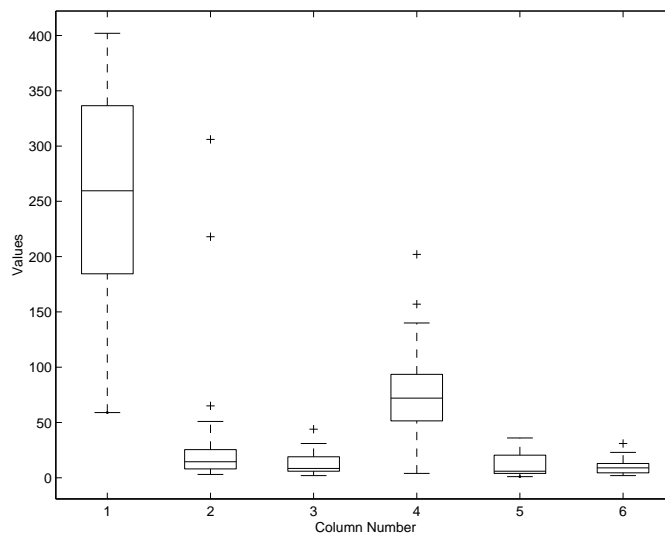
¹Les données proviennent de l'ouvrage de R. Tomassone, C. Dervin, J.P. Masson, " Biométrie, modélisation de phénomènes biologiques", page114.

2. Représenter les boîtes à moustaches pour les variables HCO_3^- et Ca^{+} .

Correction

	HCO_3^-	SO_4^-	Cl^-	Ca^+	Mg^+	Na^+
Q_1	184.5	8.0	6.0	51.5	4.0	4.5
Médianes	259.5	14.5	8.5	72.0	6.0	9.0
Q_3	336.5	25.5	19.0	93.5	20.5	13.0
1. Ecarts interquartiles	152.0	17.5	13.0	42.0	16.5	8.5
$Q_1 - 1,5(Q_3 - Q_1)$	-43.50	-18.25	-13.50	-11.50	-20.75	-8.25
$Q_3 + 1,5(Q_3 - Q_1)$	564.50	51.75	38.50	156.50	45.25	25.75
Moustaches inférieures	59	3	2	4	1	2
Moustaches supérieures	402	51	31	140	36	23

2.



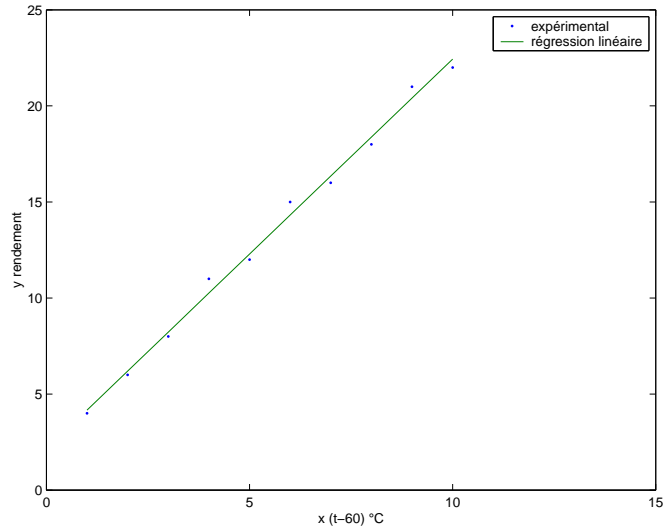
Boîtes à moustache

Exercice 3. On a monté une série d'expériences dans une unité pilote en vue d'étudier l'influence de la température sur le rendement d'une réaction chimique sous une pression donnée. Les données recueillies sont les suivantes (x est la température $t - 60^\circ C$; y est le rendement en %) :

x	1	2	3	4	5	6	7	8	9	10
y	4	6	8	11	12	15	16	18	21	22

Etudier la liaison entre y et x . On fera le graphique en " nuage de points " des valeurs de y en fonction des valeurs de x . On construira la droite de régression de y en x et on donnera la valeur du coefficient de corrélation linéaire $r(X, Y)$. Pour calculer la valeur des deux coefficients de la droite de régression, on se servira des formules dans le cas d'une variable explicative et des relations matricielles (cas de plus d'une variable explicative).

Correction



Droite de régression

$$n = 10;$$

$$\sum_i x_i = 55; \quad \sum_i x_i^2 = 385;$$

$$\sum_i y_i = 133; \quad \sum_i y_i^2 = 2111;$$

$$\sum_i x_i y_i = 899$$

D'où

$$\sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}$$

$$\sum_i (y_i - \bar{y})^2 = \sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n}$$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n}$$

$$\sum_i (x_i - \bar{x})^2 = 82,5$$

$$\sum_i (y_i - \bar{y})^2 = 342,1$$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) = 167,5$$

Donc la droite de régression des moindres carrés de y en x est $y = \hat{\beta}_0 + \hat{\beta}_1 x$

Avec

$$\hat{\beta}_0 = 2,1333 ; \quad \hat{\beta}_1 = 2,0303 ; \quad r(x,y) = 0,9970.$$

Si on applique directement les relations matricielles, on obtient :

$$y = \begin{pmatrix} 4 \\ 6 \\ \vdots \\ 22 \end{pmatrix}; \quad X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 10 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 10 & 55 \\ 55 & 285 \end{pmatrix}; \quad X^T y = \begin{pmatrix} 133 \\ 899 \end{pmatrix}$$

L'équation normale est alors

$$10\beta_0 + 55\beta_1 = 133$$

$$55\beta_0 + 385\beta_1 = 899$$

d'où la solution