Statistique descriptive

Remita Mohamed Riad

Université Badji Mokhtar, Faculté des Sciences Médicales, Département de Médecine

2012-2013

Il n'existe pas de définition précise de la statistique mais, pour nous, nous allons utiliser la définition suivante :
On appelle statistique l'ensemble des méthodes ou techniques permettant d'analyser ou des traiter des ensembles d'observations que nous appellerons données.
Les méthodes utilisées relèvent des mathématiques et font largement appel à l'outil informatique pour leur mise en œuvre.
Remarque : Il ne faut pas confondre la statistique qui est la science qui vient d'être définie et une statistique qui est un ensemble de données chiffrées sur un sujet précis.

Nous distinguerons trois phases importantes dans l'évolution de la statistique :

- De l'antiquité et jusqu'à la fin du 19ème siècle, la statistique est restée principalement un ensemble de techniques de dénombrement.
- De la fin du 19ème siècle aux années 1960 s'est construit la statistique mathématique surtout grâce à l'école anglaise (K. Pearson, W. Gosset (Student), R. Fisher, J. Neyman, ...).

Nous distinguerons trois phases importantes dans l'évolution de la statistique :

- De l'antiquité et jusqu'à la fin du 19ème siècle, la statistique est restée principalement un ensemble de techniques de dénombrement.
- De la fin du 19ème siècle aux années 1960 s'est construit la statistique mathématique surtout grâce à l'école anglaise (K. Pearson, W. Gosset (Student), R. Fisher, J. Neyman, ...).
- Depuis les années 1960, et avec le développement des outils informatique et graphique, la statistique a connu un développement considérable.

- **Population :** ensemble concerné par une étude statistique (noté Ω).
- **Individu** : (unité statistique) : tout élément de la population (noté $\omega \in \Omega$).
- **Echantillon** : sous ensemble de la population sur lequel sont réalisées les observations.
- **Enquête** : Opération consistant à observer (mesurer, questionner, ...) l'ensemble des individus d'un échantillon.
- **Recensement** : enquête dans laquelle l'échantillon observé est la population toute entière (enquête exhaustive).
- **Sondage** : enquête dans laquelle l'échantillon observé est un sous ensemble strict de la population (enquête non exhaustive).

- **Caractère** : c'est une caractéristique définie sur la population et observée sur l'échantillon. Le caractère peut être :
  - Qualitatif nominal (sexe, profession, situation familliale, ...)
  - Qualitatif ordinal (grade militaire, grade dans l'enseignement supérieur, ...)
  - Quantitatif discret (nombre d'enfants, nombre de chambre dans un appartement, ...)
  - Quantitatif continu (taille, âge, vitesse, poids, taux, ...)
- **Modalités** : les différentes valeurs prises par chaque caractère.

$\boxed{\textbf{Tableaux de données}}$

Synthétiser

$\boxed{\textbf{Tableaux statistiques}}$

Visualiser

$\boxed{\textbf{Représentations graphiques}}$

Résumer

$\boxed{\textbf{Résumés numériques}}$

Exemple 1. L'étude de la situation familliale de 30 employés d'une entreprise est résumée dans le tableau suivant :

01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**
C	M	C	C	D	V	M	M	C	D
11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
C	M	C	C	M	D	C	C	M	M
21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
C	C	M	C	V	C	C	M	C	C

Exemple 2. On étudie le nombre d'enfants dans une famille dans une cité habitée par 100 familles. On a obtenu les résultats suivants :

01	**02**	**03**	**04**	**..................**	**98**	**99**	**100**
2	0	4	2	....	6	3	1

Exemple 3. On étudie le poids en Kg de 150 nouveaux né dans une maternité. On a obtenu les résultats suivants :

01	**02**	**03**	**04**	**..................**	**148**	**149**	**150**
2.356	3.102	2.212	4.125	....	3.256	3.894	2.352

Exemple 1. On reprend l'exemple sur la situation familliale de 30 employés

X Situation familliale	Effectif
C	16
M	9
D	3
V	2
Total	30

X Situation familliale	Effectif	Fréquence
C	16	0.5333
M	9	0.3000
D	3	0.1000
V	2	0.0667
Total	30	1

X Situation familliale	Effectif	Fréquence	Pourcentage
C	16	0.5333	53.33
M	9	0.3000	30
D	3	0.1000	10
V	2	0.0667	6.67
Total	30	1	100

Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

X Nombre d'enfants	Effectif
0	11
1	16
2	21
3	25
4	17
5	8
6 et plus	2
Total	100

Tableau statistique

Variable discrète (Caractère quantitatif discret)

X Nombre d'enfants	Effectif	Fréquence
0	11	0.11
1	16	0.16
2	21	0.21
3	25	0.25
4	17	0.17
5	8	0.08
6 et plus	2	0.02
Total	100	1

Tableau statistique

Variable discrète (Caractère quantitatif discret)

X Nombre d'enfants	Effectif	Fréquence	%
0	11	0.11	11
1	16	0.16	16
2	21	0.21	21
3	25	0.25	25
4	17	0.17	17
5	8	0.08	8
6 et plus	2	0.02	2
Total	100	1	100

Tableau statistique

Variable discrète (Caractère quantitatif discret)

X Nombre d'enfants	Effectif	Fréquence	%	Effectif cumulé croissant
0	11	0.11	11	11
1	16	0.16	16	27
2	21	0.21	21	48
3	25	0.25	25	73
4	17	0.17	17	90
5	8	0.08	8	98
6 et plus	2	0.02	2	100
Total	100	1	100	—

X Nombre d'enfants	Effectif	Fréquence	%	Effectif cumulé croissant	Effectif cumulé décroissant
0	11	0.11	11	11	100
1	16	0.16	16	27	89
2	21	0.21	21	48	73
3	25	0.25	25	73	52
4	17	0.17	17	90	27
5	8	0.08	8	98	10
6 et plus	2	0.02	2	100	2
Total	100	1	100	–	–

Soit X le caractère étudié

Soit X le caractère étudié,
n la taille de l'échantillon

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X

Tableau statistique

Variable discrète (Caractère quantitatif discret)

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X,
$x_1, \cdots, x_k$ les modalités d'effectifs respectifs $n_1, \cdots, n_k$.

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X,
$x_1, \cdots, x_k$ les modalités d'effectifs respectifs $n_1, \cdots, n_k$.
Fréquence de la modalité x_i est $f_i = \frac{n_i}{n}$.

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X,
$x_1, \cdots, x_k$ les modalités d'effectifs respectifs $n_1, \cdots, n_k$.
Fréquence de la modalité x_i est $f_i = \frac{n_i}{n}$.
Pourcentage de la modalité x_i est $p_i = \frac{n_i}{n} \times 100$.

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X,
$x_1, \cdots, x_k$ les modalités d'effectifs respectifs $n_1, \cdots, n_k$.
Fréquence de la modalité x_i est $f_i = \frac{n_i}{n}$.
Pourcentage de la modalité x_i est $p_i = \frac{n_i}{n} \times 100$.
Effectif cumulé croissant à la modalité x_i est
$n_{croissant} = n_1 + n_2 + \cdots + n_i$.

Soit X le caractère étudié,
n la taille de l'échantillon,
k le nombre de modalités du caractère X,
$x_1, \cdots, x_k$ les modalités d'effectifs respectifs $n_1, \cdots, n_k$.
Fréquence de la modalité x_i est $f_i = \frac{n_i}{n}$.
Pourcentage de la modalité x_i est $p_i = \frac{n_i}{n} \times 100$.
Effectif cumulé croissant à la modalité x_i est
$n_{croissant} = n_1 + n_2 + \cdots + n_i$.
Effectif cumulé décroissant à la modalité x_i est
$n_{décroissant} = n_k + n_{k-1} + \cdots + n_i$.

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

On ordonne les données par ordre croissant et on calcule la longueur de la série statistique :

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.
On ordonne les données par ordre croissant et on calcule la longueur de la série statistique :

$$\textbf{étendue} = e = x_n - x_1$$

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.
On ordonne les données par ordre croissant et on calcule la longueur de la série statistique :

$$\textbf{étendue} = e = x_n - x_1$$

x_1 est la plus petite observation

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.
On ordonne les données par ordre croissant et on calcule la longueur de la série statistique :

$$\textbf{étendue} = e = x_n - x_1$$

x_1 est la plus petite observation
x_n est la plus grande observation

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.
On ordonne les données par ordre croissant et on calcule la longueur de la série statistique :

$$\textbf{étendue} = e = x_n - x_1$$

x_1 est la plus petite observation
x_n est la plus grande observation
n étant le nombre d'observations.

Pour construire le tableau statistique on doit regrouper les données dans des intervalles qu'appelle **classes**, pour cela on doit d'abords déterminer les nombre de classes qui sera défini par

Pour construire le tableau statistique on doit regrouper les données dans des intervalles qu'appelle **classes**, pour cela on doit d'abords déterminer les nombre de classes qui sera défini par

$$n_c = 1 + \frac{10}{3}\log_{10} n.$$

Pour construire le tableau statistique on doit regrouper les données dans des intervalles qu'appelle **classes**, pour cela on doit d'abords déterminer les nombre de classes qui sera défini par

$$n_c = 1 + \frac{10}{3}\log_{10} n.$$

Par la suite on détermine la longueur des classes qu'on appelle **amplitude** telle que

Pour construire le tableau statistique on doit regrouper les données dans des intervalles qu'appelle **classes**, pour cela on doit d'abords déterminer les nombre de classes qui sera défini par

$$n_c = 1 + \frac{10}{3} \log_{10} n.$$

Par la suite on détermine la longueur des classes qu'on appelle **amplitude** telle que

$$a = \frac{e}{n_c}$$

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.
alors
$$e = 4.593 - 2.212 = 2.381$$

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.
alors

$$e = 4.593 - 2.212 = 2.381$$

$$n_c = 1 + \frac{10}{3} \log_{10} 150 = 8.2536$$

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.
alors

$$e = 4.593 - 2.212 = 2.381$$

$$n_c = 1 + \frac{10}{3}\log_{10} 150 = 8.2536 \approx 8$$

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.
alors

$$e = 4.593 - 2.212 = 2.381$$

$$n_c = 1 + \frac{10}{3}\log_{10} 150 = 8.2536 \approx 8$$

$$a = \frac{e}{n_c} = \frac{2.381}{8} \approx 0.2976$$

Tableau statistique

Variable continue (Caractère quantitatif continu)

Pour notre exemple on a $x_1 = 2.212$ et $x_n = 4.593$.
alors

$$e = 4.593 - 2.212 = 2.381$$

$$n_c = 1 + \frac{10}{3} \log_{10} 150 = 8.2536 \approx 8$$

$$a = \frac{e}{n_c} = \frac{2.381}{8} \approx 0.2976 \approx 0.3$$

Tableau statistique

Variable continue (Caractère quantitatif continu)

X Poids des nouveaux né	Centre
$]2.2; 2.5]$	2.35
$]2.5; 2.8]$	2.65
$]2.8; 3.1]$	2.95
$]3.1; 3.4]$	3.25
$]3.4; 3.7]$	3.55
$]3.7; 4.0]$	3.85
$]4.0; 4.3]$	4.15
$]4.3; 4.6]$	4.45
Total	—

Tableau statistique

Variable continue (Caractère quantitatif continu)

X Poids des nouveaux né	Centre	Effectif
$]2.2; 2.5]$	2.35	5
$]2.5; 2.8]$	2.65	11
$]2.8; 3.1]$	2.95	21
$]3.1; 3.4]$	3.25	39
$]3.4; 3.7]$	3.55	35
$]3.7; 4.0]$	3.85	20
$]4.0; 4.3]$	4.15	13
$]4.3; 4.6]$	4.45	6
Total	—	150

Tableau statistique

Variable continue (Caractère quantitatif continu)

X Poids des nouveaux né	Centre	Effectif	Fréquence	%
$]2.2; 2.5]$	2.35	5	0.0333	3.33
$]2.5; 2.8]$	2.65	11	0.0733	7.33
$]2.8; 3.1]$	2.95	21	0.1400	14
$]3.1; 3.4]$	3.25	39	0.2600	26
$]3.4; 3.7]$	3.55	35	0.2333	23.33
$]3.7; 4.0]$	3.85	20	0.1333	13.33
$]4.0; 4.3]$	4.15	13	0.0867	8.67
$]4.3; 4.6]$	4.45	6	0.0400	4
Total	—	150	1	100

Tableau statistique

Variable continue (Caractère quantitatif continu)

X Poids des nouveaux né	Centre	Effectif	Fréquence	%	Effectif cumulé croissant
$]2.2; 2.5]$	2.35	5	0.0333	3.33	5
$]2.5; 2.8]$	2.65	11	0.0733	7.33	16
$]2.8; 3.1]$	2.95	21	0.1400	14	37
$]3.1; 3.4]$	3.25	39	0.2600	26	76
$]3.4; 3.7]$	3.55	35	0.2333	23.33	111
$]3.7; 4.0]$	3.85	20	0.1333	13.33	131
$]4.0; 4.3]$	4.15	13	0.0867	8.67	144
$]4.3; 4.6]$	4.45	6	0.0400	4	150
Total	—	150	1	100	—

Tableau statistique

Variable continue (Caractère quantitatif continu)

X Poids des nouveaux né	Centre	Effectif	Fréquence	%	Effectif cumulé croissant	Effectif cumulé décroissan
$]2.2; 2.5]$	2.35	5	0.0333	3.33	5	150
$]2.5; 2.8]$	2.65	11	0.0733	7.33	16	145
$]2.8; 3.1]$	2.95	21	0.1400	14	37	134
$]3.1; 3.4]$	3.25	39	0.2600	26	76	113
$]3.4; 3.7]$	3.55	35	0.2333	23.33	111	74
$]3.7; 4.0]$	3.85	20	0.1333	13.33	131	39
$]4.0; 4.3]$	4.15	13	0.0867	8.67	144	19
$]4.3; 4.6]$	4.45	6	0.0400	4	150	6
Total	—	150	1	100	—	—

Exemple 1. On reprend l'exemple sur la situation familliale de 30 employés

Exemple 1. On reprend l'exemple sur la situation familliale de 30 employés

Diagramme à bandes

Exemple 1. On reprend l'exemple sur la situation familliale de 30 employés

Diagramme à bandes

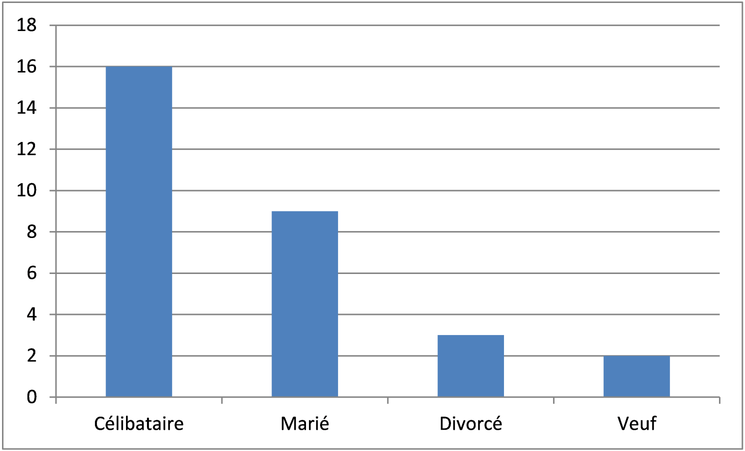


Diagramme à bandes

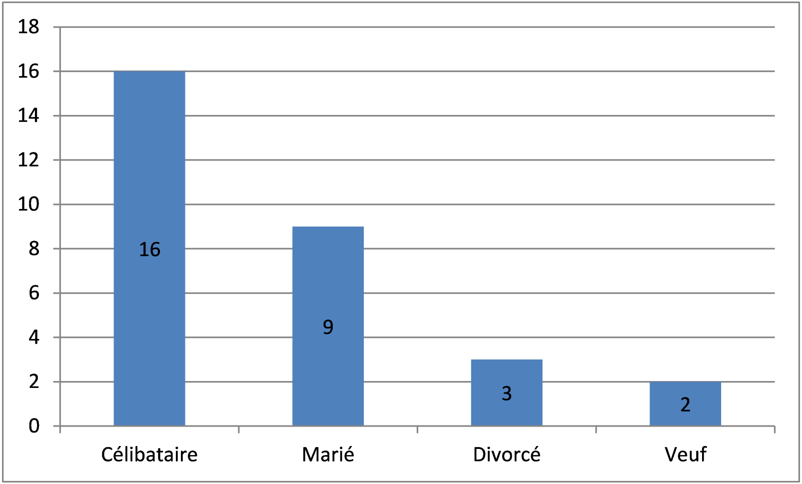


Diagramme circulaire

Diagramme circulaire

Secteur pour "Célibataire"

Diagramme circulaire

Secteur pour "Célibataire"

$$\theta_C \longrightarrow 16$$
$$360 \longrightarrow 30$$

Diagramme circulaire

Secteur pour "Célibataire"

$$\begin{matrix} \theta_C \longrightarrow 16 \\ 360 \longrightarrow 30 \end{matrix} \Longrightarrow \theta_C = \frac{16 \times 360}{30} = 192^\circ$$

Diagramme circulaire

Secteur pour "Célibataire"

$$\begin{array}{c} \theta_C \longrightarrow 16 \\ 360 \longrightarrow 30 \end{array} \Longrightarrow \theta_C = \frac{16 \times 360}{30} = 192^\circ$$

Secteur pour "Marié"

Diagramme circulaire

Secteur pour "Célibataire"

$$\begin{matrix} \theta_C \longrightarrow 16 \\ 360 \longrightarrow 30 \end{matrix} \quad \Longrightarrow \theta_C = \frac{16 \times 360}{30} = 192^\circ$$

Secteur pour "Marié"

$$\begin{matrix} \theta_M \longrightarrow 9 \\ 360 \longrightarrow 30 \end{matrix} \quad \Longrightarrow \theta_M = \frac{9 \times 360}{30} = 108^\circ$$

Diagramme circulaire

Diagramme circulaire

Secteur pour "Divorcé"

Diagramme circulaire

Secteur pour "Divorcé"

$$\begin{array}{l} \theta_D \longrightarrow 3 \\ 360 \longrightarrow 30 \end{array} \Longrightarrow \theta_D = \frac{3 \times 360}{30} = 36^\circ$$

Diagramme circulaire

Secteur pour "Divorcé"

$$\begin{array}{r} \theta_D \longrightarrow 3 \\ 360 \longrightarrow 30 \end{array} \Longrightarrow \theta_D = \frac{3 \times 360}{30} = 36^\circ$$

Secteur pour "Veuf"

Diagramme circulaire

Secteur pour "Divorcé"

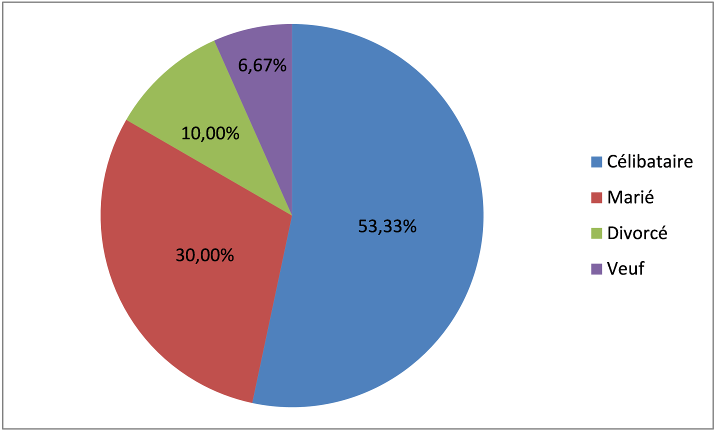
$$\begin{matrix} \theta_D \longrightarrow 3 \\ 360 \longrightarrow 30 \end{matrix} \quad \Longrightarrow \theta_D = \frac{3 \times 360}{30} = 36^\circ$$

Secteur pour "Veuf"

$$\begin{matrix} \theta_V \longrightarrow 2 \\ 360 \longrightarrow 30 \end{matrix} \quad \Longrightarrow \theta_V = \frac{2 \times 360}{30} = 24^\circ$$

Représentation graphique

Caractère qualitatif (distribution des effectifs)



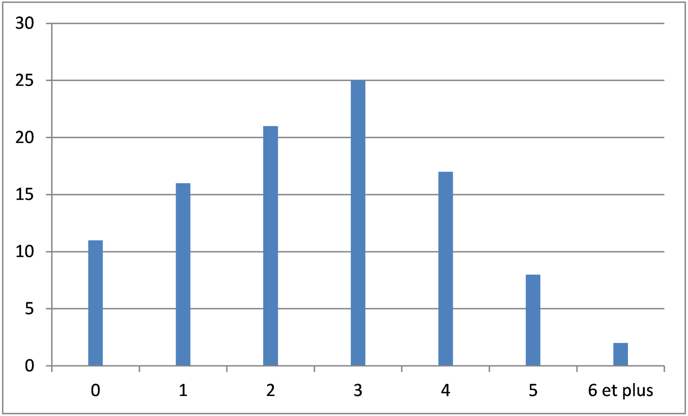
Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

X Nombre d'enfants	Effectif
0	11
1	16
2	21
3	25
4	17
5	8
6 et plus	2
Total	100

Représentation graphique

Variable discrète (Caractère quantitatif discret) (distribution des effectifs)



Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Histogramme (cas de classes de même amplitude)

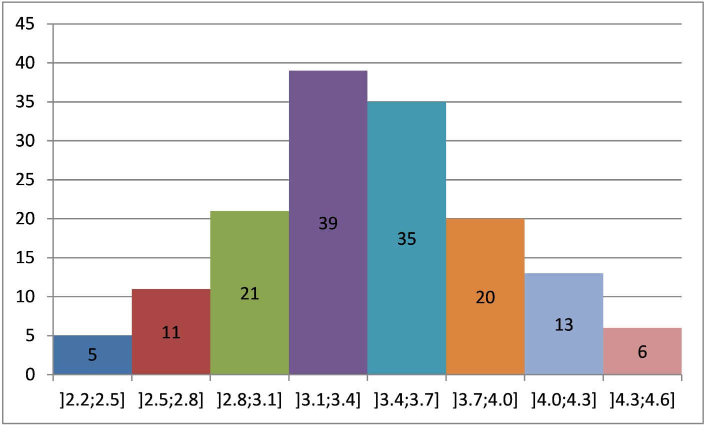
Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Histogramme (cas de classes de même amplitude)

X Nombre d'enfants	Centre	Effectif
$]2.2; 2.5]$	2.35	5
$]2.5; 2.8]$	2.65	11
$]2.8; 3.1]$	2.95	21
$]3.1; 3.4]$	3.25	39
$]3.4; 3.7]$	3.55	35
$]3.7; 4.0]$	3.85	20
$]4.0; 4.3]$	4.15	13
$]4.3; 4.6]$	4.45	6
Total	—	150

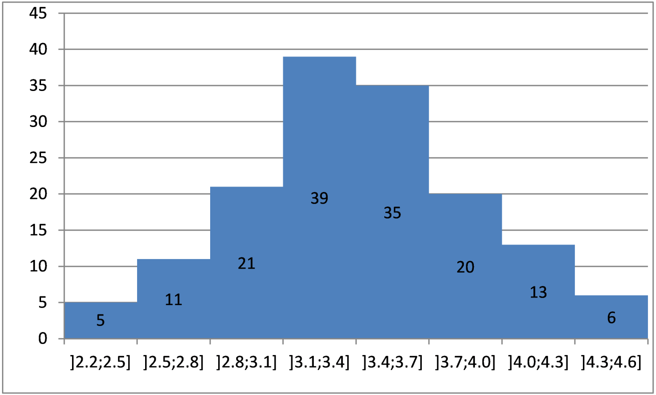
Histogramme (cas de classes de même amplitude)

Histogramme (cas de classes de même amplitude)



Histogramme (cas de classes de même amplitude)

Histogramme (cas de classes de même amplitude)



Histogramme (cas de classes d'amplitudes différentes)

Histogramme (cas de classes d'amplitudes différentes)
Si nous considérons que l'amplitude de la classe représente l'unité (longueur égale à 1) alors la surface de l'histogramme est $5 + 11 + 21 + 39 + 35 + 20 + 13 + 6 = 150$ qui est la taille de l'échantillon.

Histogramme (cas de classes d'amplitudes différentes)
Si nous considérons que l'amplitude de la classe représente l'unité (longueur égale à 1) alors la surface de l'histogramme est $5 + 11 + 21 + 39 + 35 + 20 + 13 + 6 = 150$ qui est la taille de l'échantillon.
La surface totale de l'histogramme doit être toujours égale à la taille de l'échantillon.

Histogramme (cas de classes d'amplitudes différentes)

Histogramme (cas de classes d'amplitudes différentes)
Nous regroupons les deux premières classes puis les trois dernières classes

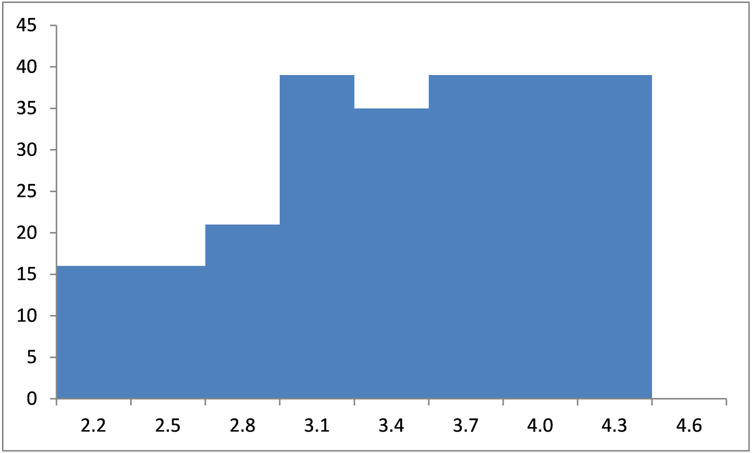
Histogramme (cas de classes d'amplitudes différentes)

Nous regroupons les deux premières classes puis les trois dernières classes

X Nombre d'enfants	Centre	Effectif	Fréquence	%	Effectif cumulé croissant	Effectif cumulé décroissant
$]2.2; 2.8]$	2.50	16	0.1067	10.67	16	150
$]2.8; 3.1]$	2.95	21	0.1400	14	37	134
$]3.1; 3.4]$	3.25	39	0.2600	26	76	113
$]3.4; 3.7]$	3.55	35	0.2333	23.33	111	74
$]3.7; 4.6]$	4.15	39	0.2600	26	150	39
Total	—	150	1	100	—	—

Histogramme (cas de classes d'amplitudes différentes)

Histogramme (cas de classes d'amplitudes différentes)



Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

Exemple 2. On reprend l'exemple sur la distribution des familles suivant le nombre d'enfants.

X Nombre d'enfants	Effectif cumulé croissant
0	11
1	27
2	48
3	73
4	90
5	98
6 et plus	100
Total	—

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Courbe des effectifs cumulés croissants

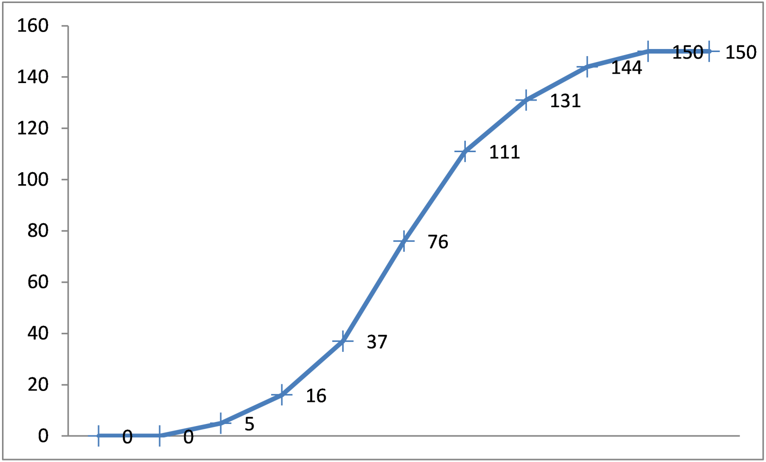
Exemple 3. On reprend l'exemple sur la distribution des nouveaux né suivant leur poids.

Courbe des effectifs cumulés croissants

X Nombre d'enfants	Centre	Effectif cumulé croissant
$]2.2; 2.5]$	2.35	5
$]2.5; 2.8]$	2.65	16
$]2.8; 3.1]$	2.95	37
$]3.1; 3.4]$	3.25	76
$]3.4; 3.7]$	3.55	111
$]3.7; 4.0]$	3.85	131
$]4.0; 4.3]$	4.15	144
$]4.3; 4.6]$	4.45	150
Total	—	—

Courbe des effectifs cumulés croissants

Courbe des effectifs cumulés croissants



Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique.

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1er ordre ou paramètres de position :

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1er ordre ou paramètres de position : **moyenne arithmétique**,

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1^{er} ordre ou paramètres de position : **moyenne arithmétique**, **mode**

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1^{er} ordre ou paramètres de position : **moyenne arithmétique**, **mode** et **médiane**.

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1er ordre ou paramètres de position : **moyenne arithmétique**, **mode** et **médiane**.
2. Les paramètres du 2ème ordre ou paramètres de dispertion :

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1^{er} ordre ou paramètres de position : **moyenne arithmétique**, **mode** et **médiane**.
2. Les paramètres du $2^{\text{ème}}$ ordre ou paramètres de dispertion : **écart-type**,

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1^{er} ordre ou paramètres de position : **moyenne arithmétique**, **mode** et **médiane**.
2. Les paramètres du $2^{ème}$ ordre ou paramètres de dispertion : **écart-type**, **coefficient de variation**

Paramètres de position et de dispertion

L'ensemble des données d'une série statistique est d'un maniement lourd et difficile. Il s'impose donc de définir un ensemble de valeurs caractéristiques (ou paramètres) qui permettent une représentation condensée de l'information contenue dans la série statistique. On distingue deux catégories de valeurs typiques :

1. Les paramètres du 1^{er} ordre ou paramètres de position : **moyenne arithmétique**, **mode** et **médiane**.
2. Les paramètres du $2^{ème}$ ordre ou paramètres de dispertion : **écart-type**, **coefficient de variation** et **étendue interquartile**.

Paramètres de position

Le Mode

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence.

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1 le mode est $Mo =$ Célibataire.

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1 le mode est $Mo =$ Célibataire.
- Pour l'exemple 2

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1 le mode est $Mo =$ Célibataire.
- Pour l'exemple 2 le mode est $Mo = 3$.

Le mode, noté *Mo*, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1 le mode est $Mo =$ Célibataire.
- Pour l'exemple 2 le mode est $Mo = 3$.
- Pour l'exemple 3

Le mode, noté Mo, d'un caractère (qualitatif ou quantitatif) est la modalité la plus observée, c'est à dire celle qui a le plus grand effectif ou la plus grande fréquence. Le mode peut ne pas exister et s'il existe, ne pas être unique.

- L'ensemble 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a pour mode $Mo = 9$
- L'ensemble 3, 5, 8, 10, 12, 15, 16 n'a pas de mode.
- L'ensemble 2, 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 a deux modes $Mo_1 = 2$ et $Mo_2 = 9$, on dit que c'est une série bimodale.
- Pour l'exemple 1 le mode est $Mo =$ Célibataire.
- Pour l'exemple 2 le mode est $Mo = 3$.
- Pour l'exemple 3 la classe modale est $]3,1; 3,4]$ et le mode sera le centre de la classe c'est à dire $Mo = 3,250$.

La moyenne arithmétique d'une série statistique $x_1, x_2, \cdots, x_k$, d'un caractère quantitatif X, et d'effectifs respectifs $n_1, n_2, \cdots, n_k$ est donné par le nombre réel $\overline{X}$ défini par

La moyenne arithmétique d'une série statistique $x_1, x_2, \cdots, x_k$, d'un caractère quantitatif X, et d'effectifs respectifs $n_1, n_2, \cdots, n_k$ est donné par le nombre réel $\overline{X}$ défini par

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i$$

où x_i est la modalité i de la variable X dans le cas d'une variable discrète et c'est le centre de la classe i dans le cas d'une variable continue et $n = \sum_{i=1}^{k} n_i$.

Exemple : On reprend l'exemple 2

X Nombre d'enfants	Effectif	$n_i x_i$
0	11	0
1	16	16
2	21	42
3	25	75
4	17	68
5	8	40
6 et plus	2	12
Total	100	253

Exemple : On reprend l'exemple 2

X Nombre d'enfants	Effectif	$n_i x_i$
0	11	0
1	16	16
2	21	42
3	25	75
4	17	68
5	8	40
6 et plus	2	12
Total	100	253

d'où $\overline{X} = \frac{253}{100} = 2,53.$

Soit une série statistique d'une variable X ayant pour modalités (ordonnées par ordre croissant) $x_1 < x_2 < \cdots < x_n$.

Soit une série statistique d'une variable X ayant pour modalités (ordonnées par ordre croissant) $x_1 < x_2 < \cdots < x_n$. On appelle médiane de X, le nombre Me, s'il existe, qui partage la série statistique en deux parties d'égal effectif.

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

On ordonne la série par ordre croissant, on a alors

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

On ordonne la série par ordre croissant, on a alors

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

On ordonne la série par ordre croissant, on a alors

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 9 alors

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

On ordonne la série par ordre croissant, on a alors

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 9 alors

$$\underbrace{4, 5, 6, 7, 8, 9, 9, 10, 10}_{\text{9 valeurs}}, \boxed{11}, \underbrace{11, 11, 12, 12, 12, 13, 15, 15, 16}_{\text{9 valeurs}}.$$

Considérons les notes de 19 étudiants obtenues à l'examen de statistiques :

$$9, 15, 4, 8, 16, 10, 11, 10, 5, 12, 9, 10, 7, 12, 15, 11, 6, 13, 12.$$

On ordonne la série par ordre croissant, on a alors

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 9 alors

$$\underbrace{4, 5, 6, 7, 8, 9, 9, 10, 10}_{\text{9 valeurs}}, \boxed{11}, \underbrace{11, 11, 12, 12, 12, 13, 15, 15, 16}_{\text{9 valeurs}}.$$

la médiane est le nombre qui partage la série en deux parties de même efectif, la valeur qui réalise ceci est 11 donc $Me = 11$.

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 10 alors

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 10 alors

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}_{\text{10 valeurs}}, \underbrace{11, 11, 12, 12, 12, 13, 15, 15, 16}_{\text{10 valeurs}}.$$

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 10 alors

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}_{10 \text{ valeurs}}, \underbrace{11, 11, 12, 12, 12, 13, 15, 15, 16}_{10 \text{ valeurs}}.$$

la médiane est le nombre qui partage la série en deux parties de même efectif, elle se trouve entre le dernier 10 et le premier 11,

On ajoute maintenant la note 7 à la série précédente, on a alors une série de 20 notes :

$$4, 5, 6, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 15, 15, 16.$$

La moitié de la série est 10 alors

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}_{\text{10 valeurs}}, \underbrace{11, 11, 12, 12, 12, 13, 15, 15, 16}_{\text{10 valeurs}}.$$

la médiane est le nombre qui partage la série en deux parties de même efectif, elle se trouve entre le dernier 10 et le premier 11, dans ce cas on prendra la valeur moyenne de ces deux notes, alors $Me = \frac{10+11}{2} = 10,5$.

D'une manière générale soit X est une variable discrète prenant les valeurs ordonnées $x_1, x_2, \cdots, x_p, x_{p+1,\cdots}, x_n,$

D'une manière générale soit X est une variable discrète prenant les valeurs ordonnées $x_1, x_2, \cdots, x_p, x_{p+1,\ldots}, x_n$, alors si

D'une manière générale soit X est une variable discrète prenant les valeurs ordonnées $x_1, x_2, \cdots, x_p, x_{p+1,\ldots}, x_n$, alors si

$$n = 2p \Longrightarrow Me = \frac{x_p + x_{p+1}}{2}$$

D'une manière générale soit X est une variable discrète prenant les valeurs ordonnées $x_1, x_2, \cdots, x_p, x_{p+1,\cdots}, x_n$, alors si

$$
\begin{aligned}
n &= 2p \Longrightarrow Me = \frac{x_p + x_{p+1}}{2} \\
n &= 2p + 1 \Longrightarrow Me = x_{p+1}.
\end{aligned}
$$

On considère l'exemple des nouveaux né

X Poids en Kg	Centre	Effectif	Effectif Cumulé croissant	$n_i x_i$
$]2,2;2,5]$	2,35	5	5	11,75
$]2,5;2,8]$	2,65	11	16	29,15
$]2,8;3,1]$	2,95	21	37	61,95
$]3,1;3,4]$	3,25	39	76	126,75
$]3,4;3,7]$	3,55	35	111	124,25
$]3,7;4,0]$	3,85	20	131	77
$]4,0;4,3]$	4,15	13	144	53,95
$]4,3;4,6]$	4,45	6	150	26,7
Total		150		511,5

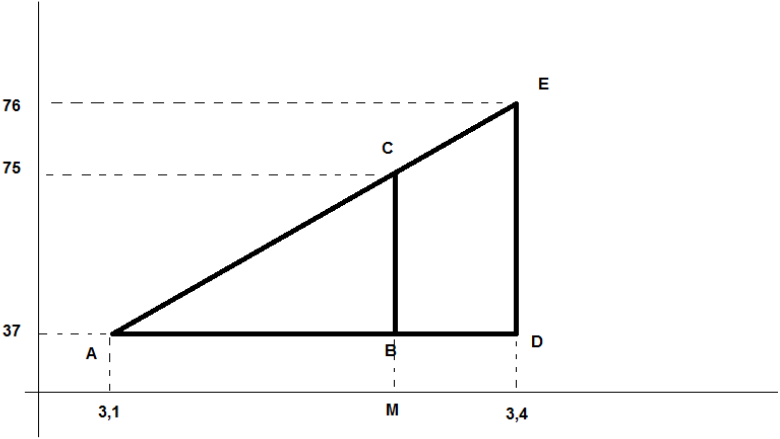
On considère l'exemple des nouveaux né

X Poids en Kg	Centre	Effectif	Effectif Cumulé croissant	$n_i x_i$
$]2,2;2,5]$	2,35	5	5	11,75
$]2,5;2,8]$	2,65	11	16	29,15
$]2,8;3,1]$	2,95	21	37	61,95
$]3,1;3,4]$	3,25	39	76	126,75
$]3,4;3,7]$	3,55	35	111	124,25
$]3,7;4,0]$	3,85	20	131	77
$]4,0;4,3]$	4,15	13	144	53,95
$]4,3;4,6]$	4,45	6	150	26,7
Total		150		511,5

On détermine d'abord la classe médiane qui correspond à la moitié des effectifs $\left(\frac{n}{2} = 75\right)$ d'où $M \in]3,1;3,4]$.

Pour déterminer M on considère la courbe des effectifs cumulés croissants mais on ne s'intéressera qu'à la portion de la classe $]3,1;3,4]$

Pour déterminer M on considère la courbe des effectifs cumulés croissants mais on ne s'intéressera qu'à la portion de la classe $]3,1;3,4]$



D'après le théorème de Thales on a

$$\frac{AB}{AD} = \frac{BC}{DE}$$

D'après le théorème de Thales on a

$$\frac{AB}{AD} = \frac{BC}{DE} \Longrightarrow \frac{M-3,1}{3,4-3,1} = \frac{75-37}{76-37}$$

D'après le théorème de Thales on a

$$\frac{AB}{AD} = \frac{BC}{DE} \Longrightarrow \frac{M-3,1}{3,4-3,1} = \frac{75-37}{76-37}$$
$$\Longrightarrow M = \frac{38}{39}0,3+3,1$$

D'après le théorème de Thales on a

$$\begin{aligned}
\frac{AB}{AD} &= \frac{BC}{DE} \Longrightarrow \frac{M-3,1}{3,4-3,1} = \frac{75-37}{76-37} \\
&\Longrightarrow M = \frac{38}{39}0,3+3,1 \\
&\Longrightarrow M \approx 3,392\ Kg
\end{aligned}$$

D'après le théorème de Thales on a

$$\begin{aligned}\frac{AB}{AD} &= \frac{BC}{DE} \Longrightarrow \frac{M-3,1}{3,4-3,1} = \frac{75-37}{76-37} \\ &\Longrightarrow M = \frac{38}{39}0,3+3,1 \\ &\Longrightarrow M \approx 3,392\ Kg\end{aligned}$$

On a aussi pour cet exemple $\overline{X} = \frac{511,5}{150} \approx 3,41\ Kg$

L'écart type d'une série statistique $(x_1, x_2, \cdots, x_k)$, d'un caractère X, et d'effectif respectif respectif $(n_1, n_2, \cdots, n_k)$ est donné par le nombre réel σ_X défini par

Paramètres de dispersion

L'écart type

L'écart type d'une série statistique $(x_1, x_2, \cdots, x_k)$, d'un caractère X, et d'effectif respectif respectif $(n_1, n_2, \cdots, n_k)$ est donné par le nombre réel σ_X défini par

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2}$$

L'écart type d'une série statistique $(x_1, x_2, \cdots, x_k)$, d'un caractère X, et d'effectif respectif respectif $(n_1, n_2, \cdots, n_k)$ est donné par le nombre réel σ_X défini par

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2}$$

où $n = \sum_{i=1}^{n} n_i$.

L'écart type d'une série statistique $(x_1, x_2, \cdots, x_k)$, d'un caractère X, et d'effectif respectif respectif $(n_1, n_2, \cdots, n_k)$ est donné par le nombre réel σ_X défini par

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2}$$

où $n = \sum_{i=1}^{n} n_i$.
x_i est la modalité i de la variable X dans le cas d'une variable discrète et c'est le centre de la classe i dans le cas d'une variable continue.
Remarque 1. On peut aussi décrire σ_X sous la forme suivante

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i x_i^2 - \overline{X}^2}.$$

L'écart type d'une série statistique $(x_1, x_2, \cdots, x_k)$, d'un caractère X, et d'effectif respectif respectif $(n_1, n_2, \cdots, n_k)$ est donné par le nombre réel σ_X défini par

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2}$$

où $n = \sum_{i=1}^{n} n_i$.
x_i est la modalité i de la variable X dans le cas d'une variable discrète et c'est le centre de la classe i dans le cas d'une variable continue.
Remarque 1. On peut aussi décrire σ_X sous la forme suivante

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^{n} n_i x_i^2 - \overline{X}^2}.$$

Remarque 2. On appelle variance de la variable X, noté $Var(X)$, le carré de l'écart type. $Var(X) = \sigma_X^2$.

Exemple : On reprend l'exemple de la distribution des nouveaux né suivant leur poids.

Exemple : On reprend l'exemple de la distribution des nouveaux né suivant leur poids.

X Poids en Kg	Centre	Effectif	Effectif Cumulé croissant	$n_i x_i$	$n_i x_i^2$
$]2,2;2,5]$	$2,35$	5	5	$11,75$	$27,6125$
$]2,5;2,8]$	$2,65$	11	16	$29,15$	$77,2475$
$]2,8;3,1]$	$2,95$	21	37	$61,95$	$182,7525$
$]3,1;3,4]$	$3,25$	39	76	$126,75$	$411,9375$
$]3,4;3,7]$	$3,55$	35	111	$124,25$	$441,0875$
$]3,7;4,0]$	$3,85$	20	131	77	$296,45$
$]4,0;4,3]$	$4,15$	13	144	$53,95$	$223,8925$
$]4,3;4,6]$	$4,45$	6	150	$26,7$	$118,815$
Total		150		$511,5$	$1779,795$

On a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{8} n_i x_i$$

On a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{511,5}{150} \approx 3,41 Kg$$

On a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{511,5}{150} \approx 3,41 Kg$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{8} n_i x_i^2 - \overline{X}^2$$

On a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{511,5}{150} \approx 3,41 Kg$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{8} n_i x_i^2 - \overline{X}^2 = \frac{1779,795}{150} - 3,41^2$$

On a

$$\begin{aligned}
\overline{X} &= \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{511,5}{150} \approx 3,41 Kg \\
\sigma_X^2 &= \frac{1}{n}\sum_{i=1}^{8} n_i x_i^2 - \overline{X}^2 = \frac{1779,795}{150} - 3,41^2 \\
\Longrightarrow & \quad \sigma_X^2 \approx 0,2372
\end{aligned}$$

On a

$$\begin{aligned}\overline{X} &= \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{511,5}{150} \approx 3,41 Kg \\ \sigma_X^2 &= \frac{1}{n}\sum_{i=1}^{8} n_i x_i^2 - \overline{X}^2 = \frac{1779,795}{150} - 3,41^2 \\ &\Longrightarrow \sigma_X^2 \approx 0,2372 \Longrightarrow \sigma_X \approx 0,4870 Kg.\end{aligned}$$

Le coefficient de variation est un paramètre de dispersion relative exprimé en pourcentage et défini par

$$CV_X = 100\frac{\sigma_X}{\overline{X}}.$$

Le coefficient de variation est un paramètre de dispersion relative exprimé en pourcentage et défini par

$$CV_X = 100\frac{\sigma_X}{\overline{X}}.$$

Pour l'exemple précédent on obtient

$$CV_X = 100\frac{0,4870}{3,41}$$

Le coefficient de variation est un paramètre de dispersion relative exprimé en pourcentage et défini par

$$CV_X = 100\frac{\sigma_X}{\overline{X}}.$$

Pour l'exemple précédent on obtient

$$\begin{aligned} CV_X &= 100\frac{0,4870}{3,41} \\ &\Longrightarrow 14,28\%. \end{aligned}$$

Les quartiles d'une série statistique sont les valeurs qui partagent la série en quatre parties de même effectif. Alors il existe trois quartiles, le premier quartile Q_1, le deuxième quartile Q_2 et le troisième quartile Q_3. Le deuxième quartile Q_2 étant la médiane M.

Les quartiles d'une série statistique sont les valeurs qui partagent la série en quatre parties de même effectif. Alors il existe trois quartiles, le premier quartile Q_1, le deuxième quartile Q_2 et le troisième quartile Q_3. Le deuxième quartile Q_2 étant la médiane M.
L'étendue interquartile est la différence entre le troisième et le premier quartile et on le note

$$IQR = Q_3 - Q_1.$$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10.$

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10$.

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10.$

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5 \Longrightarrow p = 5$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10$.

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5 \Longrightarrow p = 5$ d'où $Q_1 = \frac{x_5 + x_6}{2} = \frac{7+8}{2} = 7,5$.

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10.$

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5 \Longrightarrow p = 5$ d'où $Q_1 = \frac{x_5 + x_6}{2} = \frac{7+8}{2} = 7, 5.$

$$\underbrace{4, 5, 6, 7, 7}, \boxed{8{,}5}, \underbrace{8, 9, 9, 10, 10},$$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10.$

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5 \Longrightarrow p = 5$ d'où $Q_1 = \frac{x_5 + x_6}{2} = \frac{7+8}{2} = 7, 5.$

$$\underbrace{4, 5, 6, 7, 7}, \boxed{8{,}5}, \underbrace{8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12},$$

Cas d'une variable discrète

Exemple : Considérons les notes de 21 étudiants obtenues à l'examen de biostatistique

$$4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15, 15, 16.$$

On a $n = 21 = 2 \times 10 + 1 \Longrightarrow p = 10$ d'où $Q_2 = M = x_{p+1} = x_{11} = 10.$

$$\underbrace{4, 5, 6, 7, 7, 8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12, 13, 13, 15, 15, 16}.$$

et $p = 10 = 2 \times 5 \Longrightarrow p = 5$ d'où $Q_1 = \frac{x_5 + x_6}{2} = \frac{7+8}{2} = 7,5.$

$$\underbrace{4, 5, 6, 7, 7}, \boxed{8{,}5}, \underbrace{8, 9, 9, 10, 10}, \boxed{10}, \underbrace{11, 11, 11, 12, 12}, \boxed{12{,}5}, \underbrace{13, 13, 15, 15, 16}.$$

Cas d'une variable continue

Exemple : On considère l'exemple des nouveaux né.

X Poids en Kg	Centre	Effectif	Effectif Cumulé croissant
$]2,2;2,5]$	2,35	5	5
$]2,5;2,8]$	2,65	11	16
$]2,8;3,1]$	2,95	21	37
$]3,1;3,4]$	3,25	39	76
$]3,4;3,7]$	3,55	35	111
$]3,7;4,0]$	3,85	20	131
$]4,0;4,3]$	4,15	13	144
$]4,3;4,6]$	4,45	6	150
Total		150	

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in \left]3,1; 3,4\right]$

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in]3,1; 3,4]$ d'où

$$Q_1 = \frac{37,5 - 37}{76 - 37} \times (3,4 - 3,1) + 3,1 = 3,104 Kg$$

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in]3,1;3,4]$ d'où

$$Q_1 = \frac{37,5-37}{76-37} \times (3,4-3,1) + 3,1 = 3,104Kg$$

et $\frac{3n}{4} = 112,5$ alors $Q_3 \in]3,7;4,0]$

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in]3,1;3,4]$ d'où

$$Q_1 = \frac{37,5 - 37}{76 - 37} \times (3,4 - 3,1) + 3,1 = 3,104Kg$$

et $\frac{3n}{4} = 112,5$ alors $Q_3 \in]3,7;4,0]$ d'où

$$Q_3 = \frac{112,5 - 111}{131 - 111} \times (3,7 - 4,0) + 3,7 = 3,723Kg$$

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in \left]3,1;3,4\right]$ d'où

$$Q_1 = \frac{37,5 - 37}{76 - 37} \times (3,4 - 3,1) + 3,1 = 3,104Kg$$

et $\frac{3n}{4} = 112,5$ alors $Q_3 \in \left]3,7;4,0\right]$ d'où

$$Q_3 = \frac{112,5 - 111}{131 - 111} \times (3,7 - 4,0) + 3,7 = 3,723Kg$$

d'où

$$IQR \quad = \quad Q_3 - Q_1 = 3,723 - 3,104$$

On a $\frac{n}{4} = 37,5$ alors $Q_1 \in]3,1;3,4]$ d'où

$$Q_1 = \frac{37,5-37}{76-37} \times (3,4-3,1) + 3,1 = 3,104Kg$$

et $\frac{3n}{4} = 112,5$ alors $Q_3 \in]3,7;4,0]$ d'où

$$Q_3 = \frac{112,5-111}{131-111} \times (3,7-4,0) + 3,7 = 3,723Kg$$

d'où

$$\begin{aligned} IQR &= Q_3 - Q_1 = 3,723 - 3,104 \\ &\Longrightarrow IQR = 0,619Kg. \end{aligned}$$

Soit $\mathcal{P}$ une population d'effectif total n, sur laquelle on étudie deux caractères quantitatif X et Y, on s'intéresse à la liaison entre ces deux variables.

Soit $\mathcal{P}$ une population d'effectif total n, sur laquelle on étudie deux caractères quantitatif X et Y, on s'intéresse à la liaison entre ces deux variables.
On commence par définir la série statistique double de $\mathcal{P}$ pour les caractères X et Y

Soit $\mathcal{P}$ une population d'effectif total n, sur laquelle on étudie deux caractères quantitatif X et Y, on s'intéresse à la liaison entre ces deux variables.
On commence par définir la série statistique double de $\mathcal{P}$ pour les caractères X et Y

$$\begin{array}{ccc} \mathcal{P} & \longrightarrow & \mathbb{R}^2 \\ e_{ij} & \longmapsto & (x_i, y_j) \end{array}$$

Soit $\mathcal{P}$ une population d'effectif total n, sur laquelle on étudie deux caractères quantitatif X et Y, on s'intéresse à la liaison entre ces deux variables.
On commence par définir la série statistique double de $\mathcal{P}$ pour les caractères X et Y

$$\begin{array}{ccc} \mathcal{P} & \longrightarrow & \mathbb{R}^2 \\ e_{ij} & \longmapsto & (x_i, y_j) \end{array}$$

Une première idée pour essayer de mettre en évidence la liaison entre X et Y consiste à tracer le nuage de points associé à la série statistique double.

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_l	Effectif marginanl
x_1	n_{11}	n_{12}		n_{1j}		n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{il}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k\bullet}$
Effectif marginal	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet l}$	n

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_l	Effectif marginanl
x_1	n_{11}	n_{12}		n_{1j}		n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{il}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k\bullet}$
Effectif marginal	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet l}$	n

n_{ij} est l'effectif partiel du couple (x_i, y_j) ,

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_l	Effectif marginanl
x_1	n_{11}	n_{12}		n_{1j}		n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{il}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k\bullet}$
Effectif marginal	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet l}$	n

n_{ij} est l'effectif partiel du couple (x_i, y_j) , $n_{i\bullet}$ est l'effectif marginal de x_i et

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_l	Effectif marginanl
x_1	n_{11}	n_{12}		n_{1j}		n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{il}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k\bullet}$
Effectif marginal	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet l}$	n

n_{ij} est l'effectif partiel du couple (x_i, y_j), $n_{i\bullet}$ est l'effectif marginal de x_i et $n_{\bullet j}$ est l'effectif marginal de y_j

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_l	Effectif marginanl
x_1	n_{11}	n_{12}		n_{1j}		n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{il}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k\bullet}$
Effectif marginal	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet l}$	n

n_{ij} est l'effectif partiel du couple (x_i, y_j) , $n_{i\bullet}$ est l'effectif marginal de x_i et $n_{\bullet j}$ est l'effectif marginal de y_j

$n_{i\bullet} = \sum_{j=1}^{l} n_{ij}$ et $n_{\bullet j} = \sum_{i=1}^{k} n_{ij}$

Definition

Le couple (X, Y) est statistiquement indépendant si on a
$\forall i = 1, \cdots, k; j = 1, \cdots, l$

$$f_{ij} = \frac{n_{ij}}{n} = f_{i\bullet} \times f_{\bullet j} = \frac{n_{i\bullet}}{n} \times \frac{n_{\bullet j}}{n}$$

Séries statistiques doubles

Definition

Le couple (X, Y) est statistiquement indépendant si on a $\forall i = 1, \cdots, k; j = 1, \cdots, l$

$$f_{ij} = \frac{n_{ij}}{n} = f_{i\bullet} \times f_{\bullet j} = \frac{n_{i\bullet}}{n} \times \frac{n_{\bullet j}}{n}$$

Definition

On appelle covariance des variables X et Y et on note $Cov\,(X, Y)$, le nombre

$$\begin{aligned} Cov\,(X, Y) &= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(x_i - \overline{X}\right) \left(y_j - \overline{Y}\right) \\ &= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} x_i y_j - \overline{X}\,\overline{Y}. \end{aligned}$$

Remarque

1. *Les moyennes marginales* $\overline{X}$ *et* $\overline{Y}$ *sont données par*

Remarque

1. *Les moyennes marginales $\overline{X}$ et $\overline{Y}$ sont données par*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}x_i \text{ et } \overline{Y} = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}y_j.$$

Remarque

1. *Les moyennes marginales* $\overline{X}$ *et* $\overline{Y}$ *sont données par*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}x_i \text{ et } \overline{Y} = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}y_j.$$

2. *Si les variables* X *et* Y *sont statistiquement indépendantes alors* $Cov\left(X, Y\right) = 0.$

Séries statistiques doubles

Remarque

1. *Les moyennes marginales* $\overline{X}$ *et* $\overline{Y}$ *sont données par*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}x_i \ \text{et}\ \overline{Y} = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}y_j.$$

2. *Si les variables* X *et* Y *sont statistiquement indépendantes alors* $Cov\,(X,Y) = 0$. *Mais la réciproque n'est pas vraie.*

Remarque

1. *Les moyennes marginales* $\overline{X}$ *et* $\overline{Y}$ *sont données par*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}x_i \text{ et } \overline{Y} = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}y_j.$$

2. *Si les variables* X *et* Y *sont statistiquement indépendantes alors* $Cov\,(X, Y) = 0$. *Mais la réciproque n'est pas vraie.*

Definition

Si $Cov\,(X, Y) = 0$ on dit que les variables X et Y sont non corrélées.

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho\left(X, Y\right) = \frac{Cov\left(X, Y\right)}{\sigma_X \sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho\left(X, Y\right) = \frac{Cov\left(X, Y\right)}{\sigma_X \sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.
2. On a $-1 \leq \rho\left(X, Y\right) \leq 1$

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.
2. On a $-1 \leq \rho(X, Y) \leq 1$
3. Si $\rho(X, Y) = 0$, les variables X et Y sont non corrélées.

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho\left(X,Y\right)=\frac{Cov\left(X,Y\right)}{\sigma_X\sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.
2. On a $-1\leq\rho\left(X,Y\right)\leq 1$
3. Si $\rho\left(X,Y\right)=0$, les variables X et Y sont non corrélées.
4. Soient X et Y deux variables statistiques liées par la relation $Y=aX+b$

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho\left(X, Y\right) = \frac{Cov\left(X, Y\right)}{\sigma_X \sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.
2. On a $-1 \leq \rho\left(X, Y\right) \leq 1$
3. Si $\rho\left(X, Y\right) = 0$, les variables X et Y sont non corrélées.
4. Soient X et Y deux variables statistiques liées par la relation $Y = aX + b$
   On a $Cov\left(X, Y\right) = a\sigma_X^2$ et comme $\sigma_Y^2 = a^2\sigma_X^2$ alors

Definition

On appelle coefficient de corrélation linéaire des variables X et Y le nombre

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

1. Le coefficient de corrélation linéaire est invariant par changement d'origine et d'unité de mesure.
2. On a $-1 \leq \rho(X, Y) \leq 1$
3. Si $\rho(X, Y) = 0$, les variables X et Y sont non corrélées.
4. Soient X et Y deux variables statistiques liées par la relation $Y = aX + b$
   On a $Cov(X, Y) = a\sigma_X^2$ et comme $\sigma_Y^2 = a^2\sigma_X^2$ alors
   $$\rho(X, Y) = 1 \text{ si } a > 0 \text{ et } \rho(X, Y) = -1 \text{ si } a < 0.$$

Quand $\rho^2(X, Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - ax_i - b\right)^2 .$$

Quand $\rho^2(X, Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - ax_i - b\right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

Quand $\rho^2(X, Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - ax_i - b\right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

$$a = \frac{Cov(X, Y)}{\sigma_X^2}$$

Quand $\rho^2(X, Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - ax_i - b\right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

$$a = \frac{Cov(X, Y)}{\sigma_X^2} \text{ et } b = \overline{Y} - \frac{Cov(X, Y)}{\sigma_X^2}\overline{X}.$$

Quand $\rho^2(X, Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - ax_i - b \right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

$$a = \frac{Cov(X, Y)}{\sigma_X^2} \text{ et } b = \overline{Y} - \frac{Cov(X, Y)}{\sigma_X^2}\overline{X}.$$

Quand $\rho^2(X,Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - a x_i - b \right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

$$a = \frac{Cov(X,Y)}{\sigma_X^2} \text{ et } b = \overline{Y} - \frac{Cov(X,Y)}{\sigma_X^2}\overline{X}.$$

Quand $\rho^2(X,Y) > 0,9$ il existe une relation linéaire entre X et Y de la forme $Y = aX + b$ qu'on appelle droite de régression de Y en X et qui rend minimum la somme

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} \left(y_j - a x_i - b\right)^2 .$$

Theorem

La droite de régression de Y en X est la droite de la forme $Y = aX + b$ avec

$$a = \frac{Cov(X,Y)}{\sigma_X^2} \text{ et } b = \overline{Y} - \frac{Cov(X,Y)}{\sigma_X^2}\overline{X}.$$

Example

Déterminer l'équation de la droite de régression de Y en X dans le cas du tableau de contingence suivant

X \ Y	5	7	9	11	13	$n_{i\bullet}$
1				1	4	5
2			2	7	1	10
4			9	1		10
6	2	8	6	1		17
9	5	2	1			8
$n_{\bullet j}$	7	10	18	10	5	50

Example

Déterminer l'équation de la droite de régression de Y en X dans le cas du tableau de contingence suivant

$X \backslash Y$	5	7	9	11	13	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
1				1	4	5	5	5
2			2	7	1	10	20	40
4			9	1		10	40	160
6	2	8	6	1		17	102	612
9	5	2	1			8	72	648
$n_{\bullet j}$	7	10	18	10	5	50	$\boxed{239}$	$\boxed{1465}$
$n_{\bullet j}y_j$	35	70	162	110	65	$\boxed{442}$		
$n_{\bullet j}y_j^2$	175	490	1458	1210	845	$\boxed{4178}$		
$n_{ij}x_iy_j$	63	216	368	786	432	$\boxed{1865}$		

Séries statistiques doubles

$$\overline{X} = \frac{239}{50} = 4,78$$

Séries statistiques doubles

$$\overline{X} = \frac{239}{50} = 4,78$$
$$\overline{Y} = \frac{442}{50} = 8,84$$

Séries statistiques doubles

$$
\begin{aligned}
\overline{X} &= \frac{239}{50} = 4,78 \\
\overline{Y} &= \frac{442}{50} = 8,84 \\
\sigma_X &= \sqrt{\frac{1465}{50} - 4,78^2} = 2,54
\end{aligned}
$$

$$
\begin{aligned}
\overline{X} &= \frac{239}{50} = 4,78 \\
\overline{Y} &= \frac{442}{50} = 8,84 \\
\sigma_X &= \sqrt{\frac{1465}{50} - 4,78^2} = 2,54 \\
\sigma_Y &= \sqrt{\frac{4178}{50} - 8,84^2} \approx 2,3269
\end{aligned}
$$

Séries statistiques doubles

$$\begin{aligned}
\overline{X} &= \frac{239}{50} = 4,78 \\
\overline{Y} &= \frac{442}{50} = 8,84 \\
\sigma_X &= \sqrt{\frac{1465}{50} - 4,78^2} = 2,54 \\
\sigma_Y &= \sqrt{\frac{4178}{50} - 8,84^2} \approx 2,3269 \\
Cov\left(X, Y\right) &= \frac{1865}{50} - 4,78 \times 8,84 = -4,9552
\end{aligned}$$

$$
\begin{aligned}
\overline{X} &= \frac{239}{50} = 4,78 \\
\overline{Y} &= \frac{442}{50} = 8,84 \\
\sigma_X &= \sqrt{\frac{1465}{50} - 4,78^2} = 2,54 \\
\sigma_Y &= \sqrt{\frac{4178}{50} - 8,84^2} \approx 2,3269 \\
Cov\left(X,Y\right) &= \frac{1865}{50} - 4,78 \times 8,84 = -4,9552 \\
\rho\left(X,Y\right) &= \frac{-4,9552}{2,54 \times 2,3269} \approx -0,8384.
\end{aligned}
$$