

LEXIQUE :

Choisissez le terme à expliquer...

Amplitude d'une classe
 Bimodale
 Bivariée
 Caractère qualitatif
 Caractère quantitatif
 Caractère statistique

Amplitude d'une classe (ou d'un intervalle) :

C'est la longueur de l'intervalle.

L'amplitude de la classe $[e_i \quad e_{i+1} [$ est $e_{i+1} - e_i$.

L'amplitude de la classe $[55 \quad 68 [$ est $68 - 55 = 13$ (unités de mesure)

**Caractère qualitatif :**

Une variable statistique est qualitative si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage sur lequel les opérations arithmétiques telles que moyenne, somme, ... , n'ont pas de sens.

- Exemples :
- Sexe de la personne interrogée, situation familiale, numéro de son département de naissance, ...
 - Etat du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)

**Caractère quantitatif :**

Une variable statistique est quantitative si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

- Exemples :
- Taille, poids, salaire
 - Rendement
 - Note à un examen
 - PNB / habitant, espérance de vie, nombre d'habitants d'un ensemble de pays



Caractère statistique (ou variables statistiques) :

C'est ce qui est observé ou mesuré sur les individus d'une population statistique.
Il peut s'agir d'une variable qualitative ou quantitative.

- Exemples :
- Taille, poids, salaire, sexe, profession d'un groupe donné d'individus
 - Rendement d'un ensemble de parcelles cultivées
 - Température maximale et minimale, pluviométrie, ensoleillement, mesurés à un endroit donné tous les jours.

**Centiles :**

Les centiles C_1, C_2, \dots, C_{99} divisent une série statistique en 100 parties d'effectifs égaux.

Ce sont les abscisses respectives des points d'ordonnée 0.01 ; 0.02 ; ... ; 0.99 sur la courbe cumulative croissante.

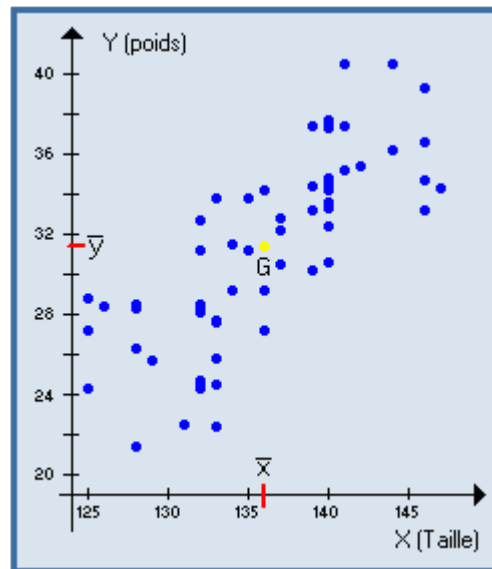
Par exemple le centile C_{98} est une valeur dépassée par 2 % des observations ; les centiles n'ont de sens que si on dispose d'un grand nombre (plusieurs centaines) d'observations.

**Centre de gravité :**

Le centre de gravité d'un nuage de points M_i de coordonnées (x_i, y_i) est le point G de coordonnées (\bar{x}, \bar{y}) ; c'est le point moyen du nuage.

\bar{x} est la moyenne des x_i

et \bar{y} la moyenne des y_i



Classe modale :

C'est la classe correspondant au maximum de l'histogramme, dans le cas d'une distribution continue unimodale.



Classes :

Intervalles de valeurs d'une variable continue, l'ensemble des classes formant une partition de l'ensemble des valeurs possibles de la variable. Par exemple, si tous les salaires des employés d'une entreprise se situent entre 750 et moins de 3 000 €, on peut construire (par exemple) les classes :

$[750 - 900 [$, $[900 - 1 500 [$, $[1 500 - 2 250 [$, $[2 250 - 3 000 [$

Chaque valeur observée de la variable doit appartenir à une classe et une seule.



Coefficient de corrélation (linéaire) :

Le coefficient de corrélation entre deux variables statistiques X et Y sur les mêmes individus est le nombre :

$$r(X, Y) = \frac{\text{cov}(X, Y)}{s_X s_Y}$$

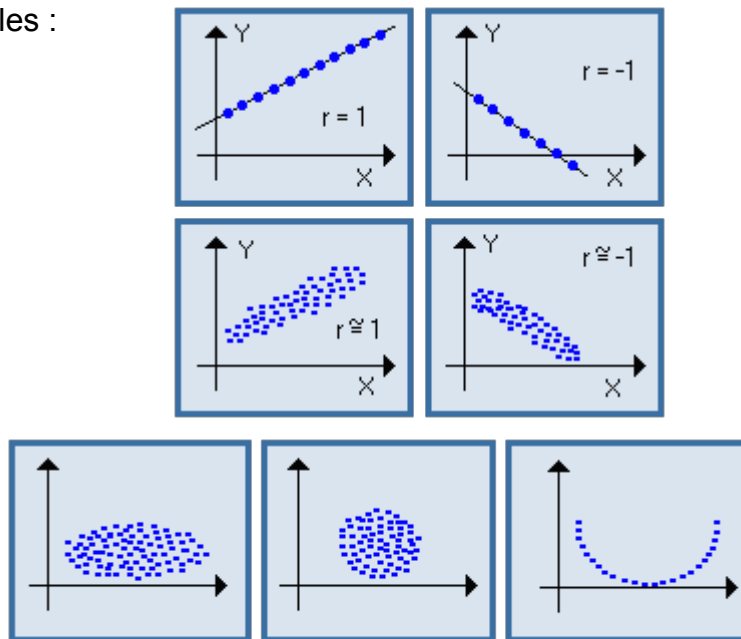
où $\text{cov}(X, Y)$ est la covariance entre X et Y,

et $s_X s_Y$ les écarts-types de X et Y.

Ce coefficient est toujours compris entre -1 et + 1.

S'il est proche de + 1 ou - 1, X et Y sont bien corrélées, c'est-à-dire qu'elles sont liées entre elles par une relation presque affine ; le nuage de points est presque aligné le long d'une droite (croissante si $r = + 1$, décroissante si $r = - 1$). S'il n'y a aucun lien entre X et Y, ce coefficient est nul, ou presque nul.

Exemples :



Exemples de cas où r est proche de 0

(Voir plus de détails dans le module "Ajustement Linéaire")



Coefficient de Spearman (ou coefficient de corrélation des rangs) :

C'est, dans le cas de deux variables ordinales X et Y mesurées sur les mêmes individus, le coefficient de corrélation entre le rang des individus pour X et le rang des individus pour Y.



Courbe cumulative croissante (ou fonction de répartition) :

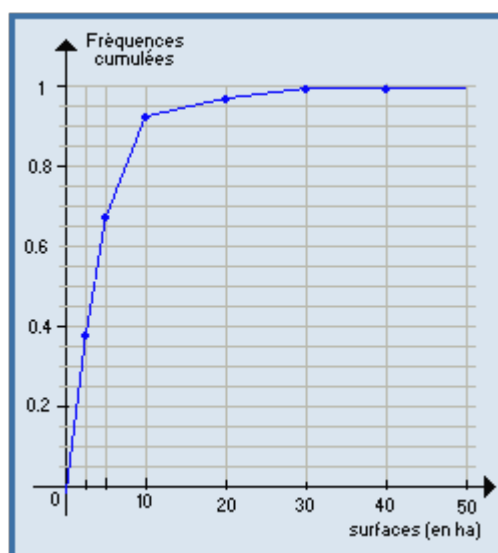
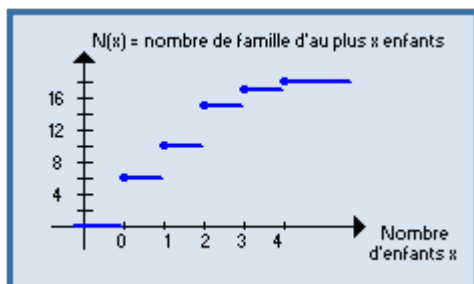
C'est le tracé de la fonction N qui à tout x associe $N(x) =$ nombre d'observations $\leq x$. Il s'obtient au moyen des effectifs cumulés croissants.

Dans le cas discret on a une fonction en escalier, dans le cas continu une fonction continue, affine par morceaux.

Si on raisonne en fréquences (au lieu d'effectifs), on a le tracé de la fonction de répartition.

$F(x) =$ proportion d'observations $\leq x$

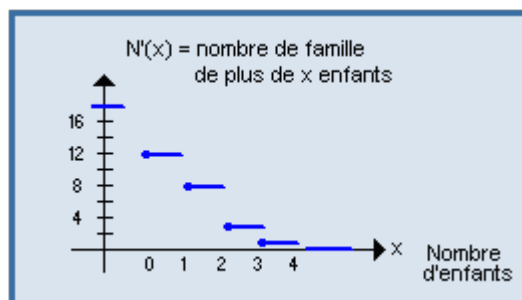
Exemples :

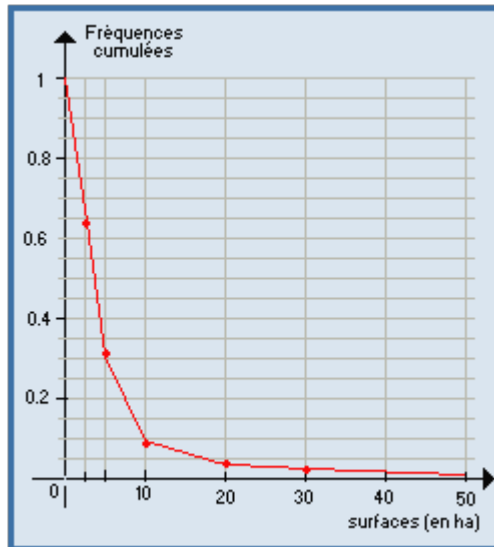


Courbe cumulative décroissante :

C'est le tracé de la fonction N' qui à tout x associe $N'(x) =$ nombre d'observations $> x$. Il s'obtient au moyen des effectifs cumulés décroissants.

Exemples :



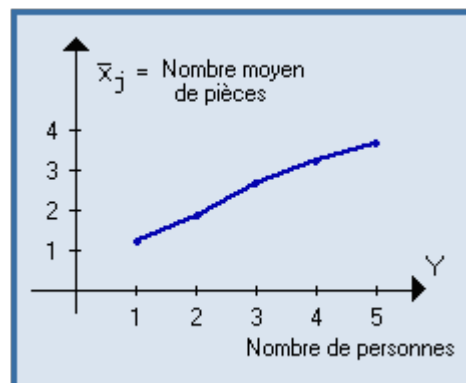
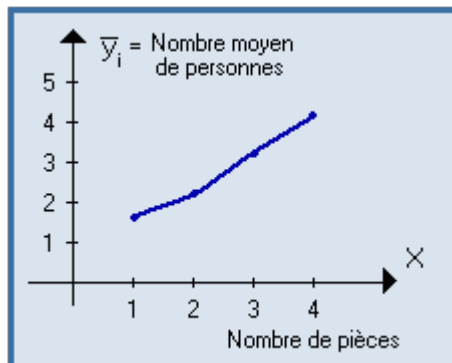


Courbe de régression :

Si X et Y sont 2 variables quantitatives, la courbe de régression de Y en X est la courbe représentant les moyennes conditionnelles de Y, à X fixé.

La courbe de régression de X en Y représente les moyennes conditionnelles de X, à Y fixé.

Exemples :



Covariance :

On appelle covariance de deux variables statistiques X et Y sur les mêmes n individus le nombre :

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}$$

"Moyenne des produits moins le produit des moyennes"

Ce nombre est positif si X et Y ont tendance à varier dans le même sens, et négatif si elles ont tendance à varier en sens contraire.

Si les données sont groupées en (x_i, y_i) d'effectifs n_i ,

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^K n_i (x_i - \bar{x})(y_i - \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^K n_i x_i y_i \right) - \bar{x} \bar{y}$$

**Déciles :**

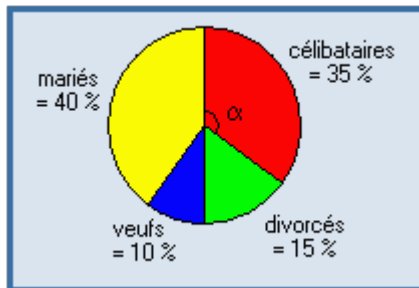
Les déciles D_1, D_2, \dots, D_9 divisent une série statistique en 10 parties d'effectifs égaux.

Ce sont les abscisses respectives des points d'ordonnée 0.1 ; 0.2 ; ... ; 0.9 sur la courbe cumulative croissante.

**Diagramme circulaire (ou à secteurs circulaires, ou en camembert) :**

Diagramme permettant de représenter la distribution d'une variable qualitative : les modalités sont représentées par des portions de disque proportionnelles à leur effectif, ou à leur fréquence.

Exemple :



l'angle a est proportionnel à l'effectif, ou à la fréquence par exemple pour représenter 15 % :

$$a = 0.15 \cdot 360 = 54^\circ$$

**Diagramme en barres (ou en tuyaux d'orgue) :**

Diagramme représentant la distribution d'une variable qualitative : les modalités sont placées en abscisse, formant des bases de rectangles égales et équidistantes, et les effectifs (ou fréquences) en ordonnée, suivant une échelle arithmétique.

Les surfaces des rectangles obtenus sont proportionnelles aux effectifs (ou aux fréquences).

Exemple :

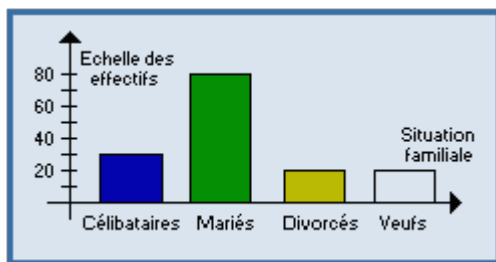


Diagramme en bâtons :

Diagramme représentant la distribution d'une variable quantitative discrète : les valeurs sont placées en abscisse, les effectifs (ou fréquences) en ordonnée, au moyen de segments verticaux.

Exemple :

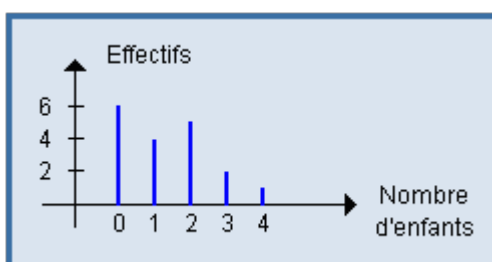
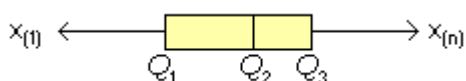
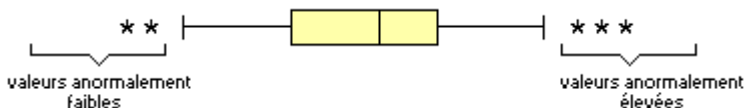


Diagramme en boîte (ou boîte à moustaches) :

Il s'agit d'un diagramme permettant de positionner les quartiles Q_1 , Q_2 , Q_3 , au moyen de rectangles de largeur arbitraire, prolongés par des "moustaches" de part et d'autre, de longueur au plus égale à une fois et demie $Q_3 - Q_1$



Si la plus petite ou la plus grande valeur observée se trouvent à l'intérieur, on raccourcit les moustaches correspondantes ; si elles se trouvent à l'extérieur, on positionne à part les valeurs "aberrantes" qui dépassent des moustaches :



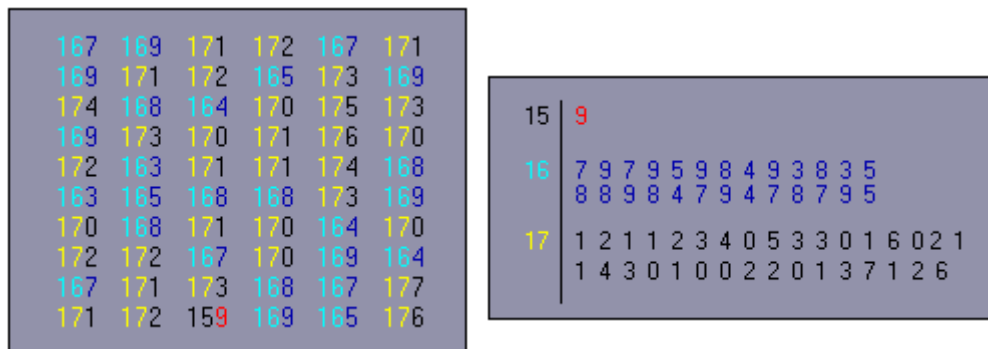
Ces diagrammes sont surtout utiles pour comparer rapidement l'allure générale de plusieurs distributions.



Diagramme tige-feuilles :

Diagramme permettant simultanément de dépouiller les données d'une série statistique et de faire une représentation graphique.

Chaque observation individuelle est représentée par sa tige (premiers chiffres, communs à plusieurs valeurs observées), et sa feuille (derniers chiffres).



Dispersion :

Un paramètre statistique est dit de dispersion s'il s'agit d'un nombre clé résumant la plus ou moins grande disparité des observations, leur plus ou moins grande variabilité de part et d'autre de la tendance centrale : étendue, écart-type sont des paramètres de dispersion.



Distribution :

Ensemble des valeurs, modalités ou classes d'une variable statistique, et des effectifs ou fréquences associées :

Par exemple :

pour une variable qualitative :

Modalités	Effectifs
A	n_1
B	n_2
...	...
L	n_K

pour une variable continue :

Classes	Effectifs	Fréquences
[10 - 12 [n_1	f_1
[12 - 14 [n_2	f_2
[14 - 16 [n_3	f_3

[16 - 20 [n_4	f_4
Total	n	1



Distribution conditionnelle :

La distribution conditionnelle d'une variable Y, pour X fixé, (X égal à x_i , modalité ou valeur, ou X appartenant à une classe donnée) est la distribution statistique des valeurs de Y, en se limitant aux individus pour lesquels X est égal à x_i (ou appartient à une classe donnée).

Exemple : 1) Distributions conditionnelles de l'âge, pour la catégorie X fixée :

Catégorie	Age (années)
A	32 ; 35 ; 40 ; 42 ; 43 ; 43 ; 49 ; 50 ; 55 ; 58
B	22 ; 26 ; 27 ; 27 ; 29 ; 30 ; 31 ; 31 ; 33 ; 34 ; 36 ; 36 ; 38 ; 39 ; 39 ; 42 ; 44 ; 46 ; 51 ; 53
C	20 ; 20 ; 21 ; 22 ; 23 ; 24 ; 24 ; 24 ; 26 ; 27 ; 28 ; 28 ; 28 ; 29 ; 29 ; 30 ; 32 ; 33 ; 33 ; 35 ; 38 ; 41 ; 43 ; 45 ; 45

Exemple : 2) Distributions conditionnelles de la catégorie, par tranche d'âge :

Pour les 20 - 30 ans : B B C B C C C C C C C B C B C C C C C C	
30 - 40 ans : A C C B B B B B B B B B C B B A C C C	
40 - 50 ans : B A C A A A B A B C C C	
50 - 60 ans : A A A B B	



Distribution marginale :

Distribution d'une variable statistique, obtenue dans la marge d'un tableau de contingence, en ajoutant les effectifs ligne par ligne, ou colonne par colonne.



Données brutes (Données statistiques brutes) :

Ensemble de mesures ou d'observations concernant l'état ou l'évolution d'un phénomène. Ce sont les valeurs prises par une ou plusieurs variables sur un certain

nombre d'individus.



Données statistiques :

Séries de mesures, d'observations, d'une ou plusieurs variables statistiques sur un ensemble d'individus.



Ecart interquartile :

C'est la différence entre le 3ème et le 1er quartile, amplitude de l'intervalle interquartile :

$$Q_3 - Q_1$$



Ecart-type :

C'est la racine carrée de la variance :

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{ou} \quad \sqrt{\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2} \quad \text{pour des données groupées.}$$

Une distribution aura un écart-type d'autant plus faible (proche de 0) qu'elle sera ramassée autour de la moyenne, avec des valeurs très peu différentes les unes des autres.



Effectif :

Nombre d'individus pour lesquels une variable statistique a pris une valeur donnée. Si, sur 150 familles, 50 ont 2 enfants, on dira que l'effectif n_i correspondant à la valeur $x_i = 2$ de la variable "nombre d'enfants", est 50.



Effectifs cumulés :

Résultat de l'addition, de proche en proche, des effectifs d'une distribution observée, soit en commençant par le 1er :

$$N_1 = n_1, N_2 = n_1 + n_2, \dots, N_i = n_1 + n_2 + \dots + n_i$$

(effectifs cumulés croissants),

soit en commençant par le dernier :

$$N'_K = n_K, N'_{K-1} = n_K + n_{K-1}, \dots, N'_i = n_K + n_{K-1} + \dots + n_i$$

(effectifs cumulés décroissants).

Exemples :

Nombre d'appels	Nombre de jours	Effectifs cumulés croissants	Effectifs cumulés décroissants
0	2	2	96
1	14	16	94
2	23	39	80
3	24	63	57
4	18	81	33
5	9	90	15
6	6	96	6
Total :	96		



Effectif total :

C'est le nombre d'observations, d'une série statistique brute, nombre d'individus de la population étudiée.

Il est égal à la somme des effectifs associés aux différentes modalités, valeurs ou classes :

$$n = \sum_{i=1}^K n_i$$



Etendue :

Différence entre la plus grande et la plus petite des observations d'une série statistique. C'est un paramètre de dispersion.



Fonction de répartition :

C'est le tracé de la fonction N qui à tout x associe $N(x) =$ nombre d'observations $\leq x$. Il s'obtient au moyen des effectifs cumulés croissants.

Dans le cas discret on a une fonction en escalier, dans le cas continu une fonction continue, affine par morceaux.

Si on raisonne en fréquences (au lieu d'effectifs), on a le tracé de la fonction de répartition.

$F(x) =$ proportion d'observations $\leq x$

**Fractiles :**

On appelle fractiles des valeurs F_1, F_2, \dots, F_{k-1} divisant une série en k parties d'effectifs égaux.

On a autant de valeurs $\leq F_1$ que de valeurs comprises entre F_1 et F_2 , ou entre F_2 et F_3 , etc.

Pour $k = 4$, ce sont les 3 quartiles Q_1, Q_2, Q_3 (Q_2 étant la médiane).

**Fréquence (ou fréquence relative) :**

C'est la proportion (ou le pourcentage) d'individus pour lesquels une variable statistique a pris une valeur donnée. Si, sur 150 familles, 50 ont 2 enfants, on dira que la fréquence f_i correspondant à la valeur $x_i = 2$ de la variable "nombre d'enfants", est :

$$f_i = \frac{n_i}{n} = \frac{50}{150} = 0.33 \text{ soit } 1/3 \text{ ou } 33.33 \%$$

**Fréquences cumulées :**

Résultat de l'addition, de proche en proche, des fréquences d'une distribution observée, soit en commençant par le 1er :

$$F_1 = f_1, F_2 = f_1 + f_2, \dots, F_i = f_1 + f_2 + \dots + f_i$$

(fréquences cumulées croissantes),

soit en commençant par le dernier :

$$F'_K = f_K, F'_{K-1} = f_K + f_{K-1}, \dots, F'_i = f_K + f_{K-1} + \dots + f_i$$

(fréquences cumulées décroissantes).

Exemples :

Nombre d'appels	Fréquences en %	Fréquences cumulées croissantes	Fréquences cumulées décroissantes
0	2.08	2.08	100
1	14.58	16.66	97.92
2	23.96	40.62	83.34
3	25.00	65.62	59.38
4	18.75	84.37	34.38
5	9.38	93.75	15.63

6	6.25	100	6.25
---	------	-----	------

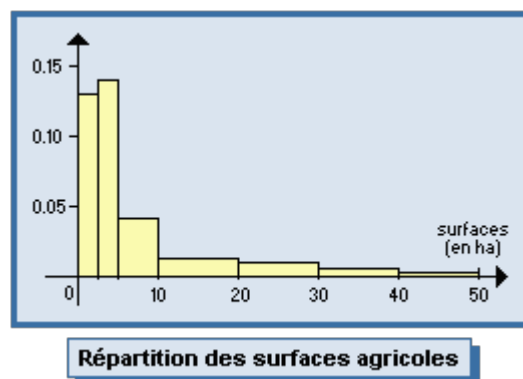
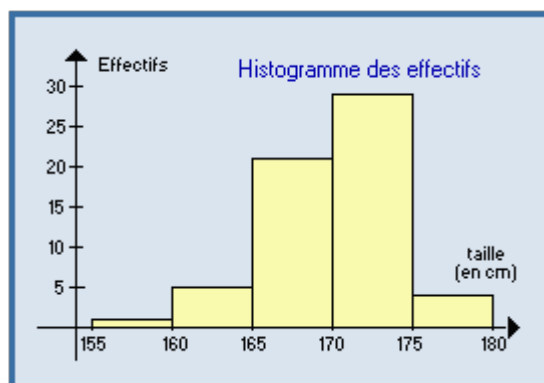


Histogramme :

Graphique permettant de représenter une distribution continue regroupée en classes : rectangles juxtaposés dont les bases sont les classes, et les surfaces sont proportionnelles aux effectifs (ou fréquences) associés.

Si les classes sont de même amplitude a_i , on place en ordonnée les effectifs n_i (ou les fréquences f_i). Si les amplitudes a_i sont différentes, on place n_i / a_i (ou f_i / a_i).

Exemples :



Indépendance :

2 variables statistiques X et Y sont dites indépendantes si la distribution conditionnelle de Y, pour tout x, est constante (c'est-à-dire ne dépend pas de x). Cela signifie que les lignes du tableau de contingence sont proportionnelles, ou de façon équivalente que les colonnes du tableau de contingence sont proportionnelles, et donc que la distribution conditionnelle de X, pour tout y, est constante.

Exemples :

Y

	Célibataires	Mariés	Veufs	Divorcés	Total
X homme	20	10	2	4	36
femme	30	15	3	6	54
Total	50	25	5	10	90

"sexe" et "situation matrimoniale" sont ici deux variables indépendantes.



Individu (ou unités statistiques) :

Les individus sont les éléments de la population statistique étudiée. Pour chaque individu, on dispose d'une ou plusieurs observations.

- Exemples :
- chacune des personnes interrogées pour une enquête
 - chaque parcelle cultivée en vue d'étudier le rendement
 - chaque pays pour lequel on étudie des données socio-économiques, ..
 - chaque jour de l'année pour lequel on dispose de données météorologiques, ...



Inégalité de (Bienaymé)-Tchébichev :

Pour toute population de moyenne \bar{x} et d'écart-type s , la proportion de valeurs de l'intervalle $[\bar{x} - k s ; \bar{x} + k s]$ est d'au moins égale à $1 - \frac{1}{k^2}$, pour tout $k \geq 1$.

Par exemple, 75 % des valeurs au moins appartiennent à : $[\bar{x} - 2 s ; \bar{x} + 2 s]$, c'est-à-dire s'écartent de moins de 2 écart-types de la moyenne.



Intervalle interquartile :

C'est l'intervalle entre le 1er et le 3ème quartile : $[Q_1 ; Q_3]$.

Il contient 50 % des observations ; 25 % sont inférieures et 25 % sont supérieures.



Intervalle médian :

C'est l'intervalle du milieu d'une série statistique comprenant un nombre pair d'observations :

$$x_{(1)} \leq x_{(2)} \leq \dots \leq \underbrace{x_{(p)} \leq x_{(p+1)}}_{\substack{\text{intervalle} \\ \text{médian}}} \leq \dots \leq x_{(2p)}$$



Médiane :

La médiane M d'une série statistique rangée par ordre croissant

$x_{(1)} < x_{(2)} < \dots < x_{(n)}$ est la valeur "du milieu", soit $x_{(p+1)}$ si n est impair et vaut $2p + 1$,
ou

$$\frac{x_{(p)} + x_{(p+1)}}{2} \text{ si n est pair et vaut } 2p$$

M est l'abscisse du point d'intersection des courbes cumulatives, d'ordonnée $n/2$ en effectifs ou 0.5 en fréquences.



Modalité :

Les modalités d'une variable qualitative sont les différentes valeurs que peut prendre celle-ci.

Par exemple les modalités de la variable "situation familiale" sont : célibataire, marié, veuf, divorcé.

Les modalités de la variable "sexe" sont : féminin, masculin (pouvant être codées par exemple 0 et 1).



Moyenne arithmétique :

C'est le quotient \bar{x} de la somme d'une série d'observations par leur nombre.

$$\text{Pour une série brute } x_1, x_2, \dots, x_n, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Pour une série groupée } (x_i, n_i), i = 1, \dots, K, \bar{x} = \frac{1}{n} \sum_{i=1}^K n_i x_i$$



Moyenne conditionnelle :

Les moyennes conditionnelles sont les moyennes des distributions conditionnelles : valeurs moyennes de Y, pour X fixé ou valeurs moyennes de X, pour Y fixé.

**Moyenne pondérée :**

La moyenne des nombres x_1, x_2, \dots, x_n , pondérée par les poids p_1, p_2, \dots, p_n (nombres positifs de somme 1) est égale à :

$$\bar{X} = \sum_{i=1}^n p_i x_i$$

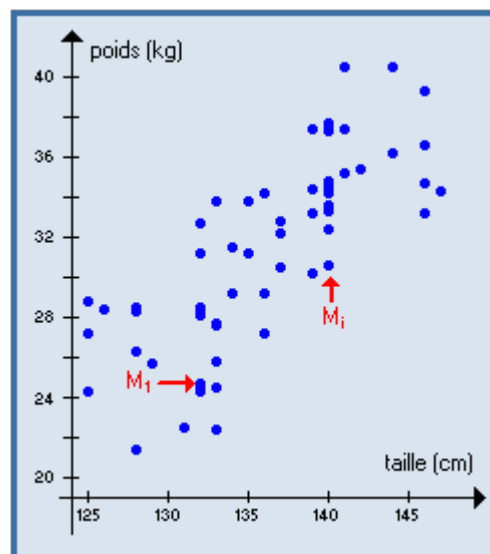
Dans le calcul de cette moyenne, les valeurs ayant un poids important comptent davantage que celles ayant un poids faible.

**Nuage de points :**

Ensemble de points isolés représentés dans un graphique cartésien : points M_1, M_2, \dots, M_n de coordonnées $(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)$.

Exemples :

taille et poids de 60 enfants

**Paramètres statistiques :**

Ce sont quelques nombres permettant de résumer numériquement les traits principaux d'une distribution statistique.

Par exemple : la moyenne, l'écart-type, l'étendue sont des paramètres statistiques.



Population statistique :

Une population statistique est l'ensemble sur lequel on effectue des observations.

- Exemples :
- ensemble de personnes interrogées pour une enquête
 - ensemble de parcelles cultivées sur lesquelles on mesure un rendement
 - ensemble de pays pour lesquels on dispose de données géographiques ou économiques, ...

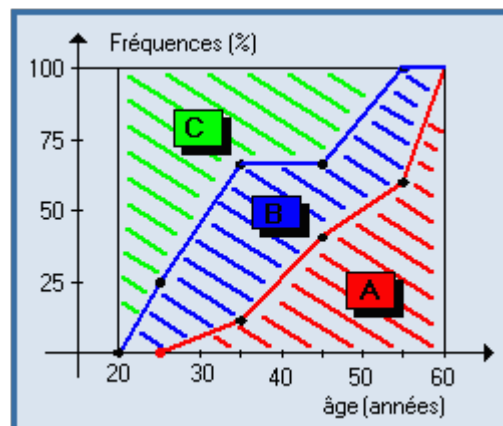
**Position :**

Un paramètre statistique est dit de position s'il s'agit d'un nombre clé permettant de préciser où se répartit une certaine fraction des observations ainsi les quartiles permettent de situer le 1/4 inférieur, la moitié, le 1/4 supérieur des observations.

**Profils :**

Ce sont les distributions conditionnelles, écrites en fréquences et non en effectifs.

On peut les représenter graphiquement par :

**Quartiles :**

Les quartiles Q_1 , Q_2 , Q_3 divisent une série statistique en 4 parties d'effectifs égaux : 25 % des valeurs sont $\leq Q_1$, 25 % comprises entre Q_1 et Q_2 , 25 % entre Q_2 et Q_3 , et 25 % supérieures à Q_3 .

Q_1 , Q_2 , Q_3 sont respectivement l'abscisse des points d'ordonnées 0.25 ; 0.5 ; 0.75 sur la

courbe cumulative croissante. Q_2 est égal à la médiane.



Rang :

Si X est une variable ordinale mesurée sur n individus, le rang de l'individu i pour X est le numéro d'ordre de i , si on range toutes les valeurs x_i par ordre croissant.

Exemple : si les x_i obtenus sont : O R D R E ; le rang de l'individu n° 3, pour l'ordre alphabétique, est 1 ; le rang de l'individu n° 5 est 2, etc ...



Rapport de corrélation :

C'est

$$e^2_{Y/X} = \frac{\text{variance expliquée}}{\text{variance de } Y}$$

coefficient compris entre 0 et 1 mesurant la part plus ou moins grande de la variabilité d'une variable Y qui peut être expliquée par les variations d'une autre variable X , qualitative, discrète, ou continue découpée en classes.



Série statistique (ou distribution observée) :

Ensemble des modalités, valeurs, ou classes d'une variable, avec les effectifs observés correspondants.

Exemples :

Modalités	Effectifs
Célibataire	30
Marié	80
Divorcé	20
Veuf	20

Nombre d'enfants x_i	Effectif n_i
0	6
1	4
2	5
3	2
4	1

Classes de tailles (en cm)	Effectifs
-----------------------------	-----------

moins de 160	1
[160 - 165 [5
[165 - 170 [21
[170 - 175 [29
175 et plus	4



Statistique descriptive univariée :

La Statistique descriptive univariée consiste en la description de chacun des caractères statistiques, un par un, et non des liens éventuels existant entre eux.



Statistique descriptive bivariée :

La Statistique descriptive bivariée consiste en la description de deux variables mesurées simultanément sur les mêmes individus. Elle permet de mettre en évidence le type de lien existant éventuellement entre ces variables.



Statistique inférentielle :

Partie de la statistique qui, contrairement à la statistique descriptive, ne se contente pas de décrire des observations, mais extrapole les constatations faites à un ensemble plus vaste, permet de tester des hypothèses sur cet ensemble, et de prendre des décisions le concernant. [Voir les modules "Echantillonnage-Estimation" et " Tests"]



Tableau de contingence :

Tableau résultant du tri croisé de deux variables.

		Y			
		Célibataires	Mariés	Veufs	Divorcés
X	homme	30	20	3	7
	femme	40	25	5	10



Tendance centrale :

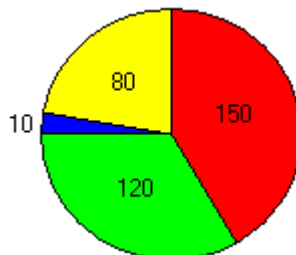
Un paramètre statistique est dit de tendance centrale s'il s'agit d'un nombre clé autour duquel les observations sont réparties : mode, médiane, moyenne sont des paramètres de tendance centrale.



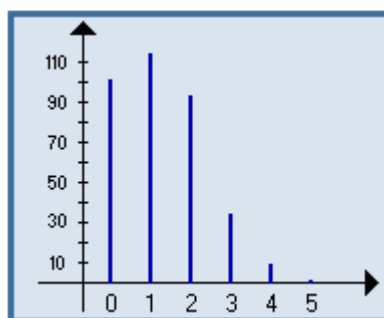
Tri à plat d'une série statistique brute :

C'est l'inventaire des modalités ou valeurs rencontrées dans la série, avec les effectifs correspondants.

Situation familiale	Nombre de personnes dans cette situation
célibataire	150
marié	120
divorcé	10
veuf	80



Nombre d'enfants	Nombre de personnes ayant ce nombre d'enfants
0	103
1	115
2	95
3	35
4	10
5	2



Tri croisé :

A partir de 2 variables X et Y mesurées sur les mêmes individus, décompte des effectifs correspondant à chaque couple (x_i, y_j) :

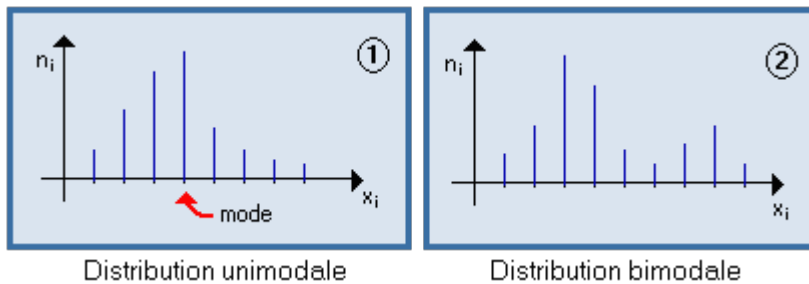
nombre d'individus pour lesquels $X = x_i$ et $Y = y_j$



Unimodale :

Une distribution est unimodale si elle présente un maximum marqué, correspondant à une valeur appelée **mode**.

S'il y a plusieurs maxima relatifs, la distribution est **plurimodale** (**bimodale** dans le cas 2)



Unité statistique (ou individu(s)) :

Les individus sont les éléments de la population statistique étudiée.

Pour chaque individu, on dispose d'une ou plusieurs observations.

- Exemples :
- chacune des personnes interrogées pour une enquête
 - chaque parcelle cultivée en vue d'étudier le rendement
 - chaque pays pour lequel on étudie des données socio-économiques, ...
 - chaque jour de l'année pour lequel on dispose de données météorologiques, ...



Variable continue :

C'est une variable quantitative pouvant prendre par nature une infinité de valeurs, généralement tout un intervalle réel.

- Exemples :
- tailles, poids, salaires, surfaces cultivées, températures, ...



Variable dichotomique :

C'est une variable qualitative qui ne peut prendre que 2 modalités : OUI ou NON ; masculin ou féminin ; bon ou mauvais , etc....



Variable discrète :

C'est une variable quantitative pouvant prendre par nature un nombre fini (ou dénombrable) de valeurs.

- Exemples :
- nombre d'enfants par famille

- nombre de pièces d'un appartement
- nombre de pièces défectueuses dans un lot de pièces mécaniques ...



Variable qualitative (ou caractère qualitatif) :

Une variable statistique est qualitative si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage sur lequel les opérations arithmétiques telles que moyenne, somme, ... , n'ont pas de sens.

- Exemples :
- sexe de la personne interrogée, situation familiale, numéro de son département de naissance, ...
 - état du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)



Variable qualitative nominale :

C'est une variable qualitative dont les modalités ne sont pas ordonnées.

- Exemples :
- la variable sexe peut être notée M F, 0 1 , ou 1 0
 - la variable CSP : on ne peut pas classer les catégories socio- professionnelles selon un ordre préétabli.



Variable qualitative ordinale :

C'est une variable qualitative dont les modalités sont naturellement ordonnées selon un ordre total : on peut dire que selon un certain sens la modalité A est moins forte que la B, qui est moins forte que la C, etc...

- Exemples :
- tailles de vêtement 0 1 2 3 ... mais la taille 2 ne signifie pas que le vêtement est 2 fois plus grand que celui de la taille 1 ! Il ne s'agit pas d'une variable quantitative discrète.



Variable quantitative (ou caractère quantitatif) :

Une variable statistique est quantitative si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

- Exemples :
- taille, poids, salaire
 - rendement
 - note à un examen
 - PNB / habitant, espérance de vie, nombre d'habitants d'un ensemble de pays

**Variable statistique (ou caractère statistique) :**

C'est ce qui est observé ou mesuré sur les individus d'une population statistique. Il peut s'agir d'une variable qualitative ou quantitative.

- Exemples :
- taille, poids, salaire, sexe, profession d'un groupe donné d'individus
 - rendement d'un ensemble de parcelles cultivées
 - température maximale et minimale, pluviométrie, ensoleillement, mesurés à un endroit donné tous les jours.

**Variance :**

C'est la moyenne des carrés des écarts à la moyenne :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ pour des données isolées, et}$$

$$\frac{1}{n} \sum_{i=1}^K (n_i) (x_i - \bar{x})^2 \text{ pour des données groupées.}$$

(x_i est le centre de classe dans le cas de données regroupées en classes).

On peut aussi calculer la variance par :

$$\frac{1}{n} \sum_{i=1}^n (x_i^2) - \bar{x}^2 \text{ ou } \frac{1}{n} \left(\sum_{i=1}^K n_i x_i^2 \right) - \bar{x}^2$$

Par exemple : la variance de la série : 8 9 10 10 12 14 14 16, qui a pour moyenne $\bar{x} = 11.625$ est :

$$s^2 = \frac{1}{8} [(8 - 11.625)^2 + (9 - 11.625)^2 + 2 (10 - 11.625)^2 + (12 - 11.625)^2 + 2 (14 - 11.625)^2 + (16 - 11.625)^2]$$

$$= \frac{1}{8} (8^2 + 9^2 + 2 \cdot 10^2 + 12^2 + 2 \cdot 14^2 + 16^2) - (11.625)^2 = 6.9844$$

La racine carrée de la variance est l'écart-type, qui s'exprime dans la même unité que les x_i et mesure la plus ou moins grande dispersion des valeurs de part et d'autre de la moyenne.



Variance expliquée :

C'est la variance des moyennes des distributions conditionnelles : si Y est quantitative, et si X subdivise l'ensemble des individus en K classes d'effectifs n_1, n_2, \dots, n_K telles que la moyenne de Y sur chaque classe est : $\bar{y}_1, \dots, \bar{y}_K$,

la variance de Y expliquée par X est :

$$\frac{1}{n} \sum_{i=1}^K (n_i \bar{y}_i^2) - \bar{y}^2$$



Variance résiduelle :

C'est la moyenne des variances des distributions conditionnelles, pondérées par les effectifs. Si Y est quantitative, et si X subdivise l'ensemble des individus en K classes d'effectifs n_1, n_2, \dots, n_K telles que la moyenne de Y sur chaque classe est :

$$\bar{y}_1, \dots, \bar{y}_K,$$

avec les variances $s_1^2, s_2^2, \dots, s_K^2$, la variance de Y se décompose en :

$$s_Y^2 = \left[\frac{1}{n} \sum_{i=1}^K (n_i \bar{y}_i^2) - \bar{y}^2 \right] + \left[\frac{1}{n} \sum_{i=1}^K n_i s_i^2 \right]$$

Le premier terme de la somme est la variance de Y expliquée par X, le second la variance résiduelle.

