Cours 2 : Rappels de Statistique descriptive

A- Introduction

B- Statistique descriptive unidimensionnelle

C- Statistique descriptive bidimensionnelle

A- Introduction

A- Introduction

- Rappel : Série statistique = ensemble de mesures d'une ou plusieurs variables faites sur une population ou un échantillon d'individus.

Individu	X_1	X_2		X_j		X_p
e1	x_{11}	x_{12}		x_{1j}		x_{1p}
e2	x_{21}	x_{22}		x_{2j}		x_{2p}
ei	x_{i1}	x_{i2}		x_{ij}		x_{ip}
en	x_{n1}	x_{n2}		x_{nj}		x_{np}

A- Introduction

- ✓ Objectifs de la statistique descriptive (ou exploratoire):
  - résumer, synthétiser l'information contenue dans la série statistique, mettre en évidence ses propriétés.
  - suggérer des hypothèses relatives à la population dont est issu l'échantillon.

- ✓ Outils utilisés :
  - Tableaux (table des fréquences, de contingence, …)
  - Graphiques (box-plots, histogrammes,..)
  - indicateurs (moyenne, corrélation,..).

A- Introduction

✓ Le type d'outils utilisé dépend

- De la nature de la série (uni ou multi dimensionnelle)
- De la nature des variables (quantitatives discrètes, continues ou qualitatives).

A- Introduction

Exemple : observation de la séquence d'un brin d'ADN

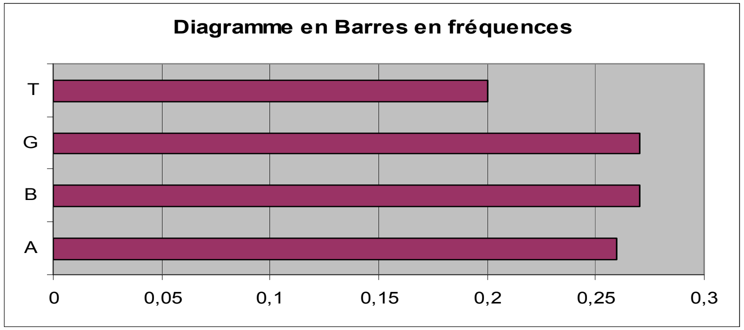
GGGAGTGTBTATTAABTBBGAA
BTBBBAGBGBTAGBTBGBGBGG
AGTGABBGAGBBTABATGAGGG
TABTGTBAATAABGBATGTTABB
AGAAGGA

Série unidimensionnelle de taille 100 de la variable qualitative « base du brin d'ADN ».

Table des fréquences:

valeurs	effectifs	frequences
A	26	0,26
C	27	0,27
G	27	0,27
T	20	0,2

Visualisation :



Indicateur: Modes=C et G

A- Introduction

Exemple : Série des âges de 50 salariés dans une entreprise

36.44460 30.63702 30.36399 56.13572 62.31707 48.87932 25.22967 45.07674

41.22021 18.45797 46.82866 57.83412 26.93824 51.17832 42.42865 25.00991

39.49332 61.49174 41.12957 48.73509 24.84856 62.86307 31.46099 18.30140

58.65384 22.66574 28.69191 43.23656 29.99305 37.23314 25.34647 56.18528

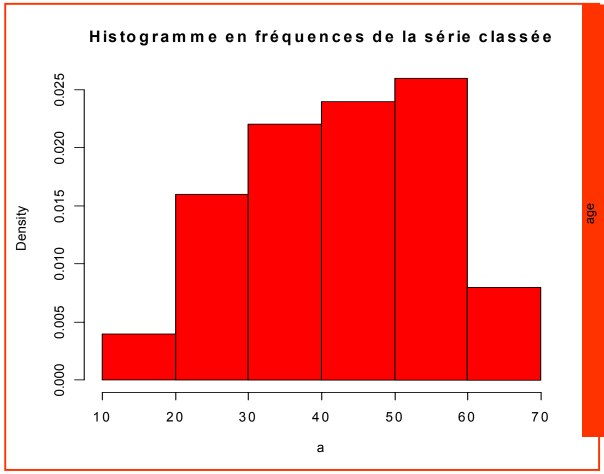
59.60421 56.78237 34.86674 55.49477 52.80441 58.90374 64.61624 57.62305

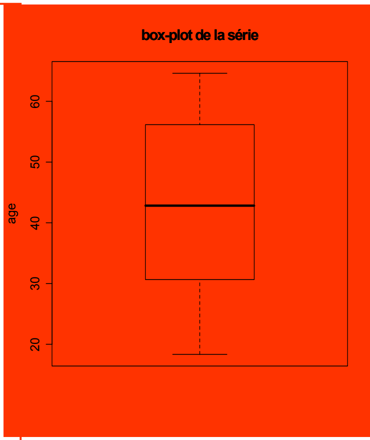
41.92750 39.26187 43.79833 33.12420 44.39254 58.30465 30.01482 56.69020

45.00456 39.18792

Série unidimensionelle de la variable quantitative continue « age ».

classes	centres	amplitudes	effectifs	frequences	eff. Cum.	freq. Cum.
(18.3,26]	22,15	7,7	7	0,14	7	0,14
(26,33.7]	29,85	7,7	8	0,16	15	0,3
(33.7,41.5]	37,6	7,7	8	0,16	23	0,46
(41.5,49.2]	45,35	7,7	10	0,2	33	0,66
(49.2,56.9]	53,05	7,7	7	0,14	40	0,8
(56.9,64.7]	60,8	7,7	10	0,2	50	1





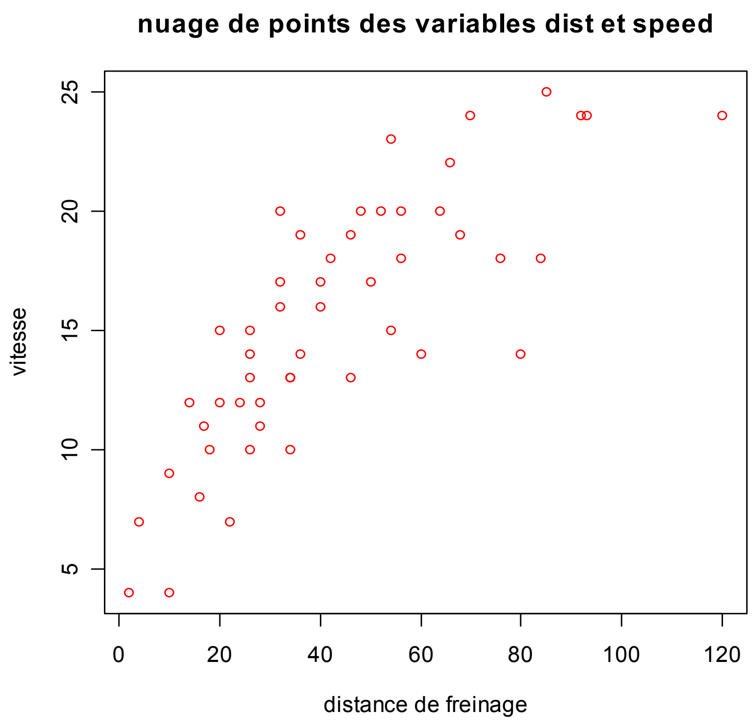
Min.	Q1	Median	Mean	Q3	Max.
18.30	30.84	42.83	42.95	56.17	64.62

A- Introduction

Ex: observation de la vitesse et de la distance de freinage de 50 voitures.

	speed	dist
1	4	2
2	4	10
3	7	4
4	7	22
5	8	16
6	9	10
7	10	18
8	10	26
9	10	34
10	11	17
11	11	28

………………



B- Statistique descriptive unidimensionnelle

1-Généralités

2- Etude d’une variable quantitative

3- Etude d’une variable qualitative

B-1 Généralités

✓On considère une variable statistique X, observée sur n individus. On dispose alors d'une série statistique unidimensionnelle $x = (x_1, \ldots, x_n)$ que l'on peut mettre sous forme d'un tableau de données :

individu	1	2	....	i	.....	n
Valeur de X	x_1	x_2	....	x_i		x_n

x_i = valeur de X pour l'individu i de la série.

✓On veut mettre en évidence les principales caractéristiques de la série.

B-1 Généralités

- effectif d'une valeur de X : nb. d'individus ayant cette valeur. $\boxed{n_i}$
- fréquence d'une valeur de X : prop. d'individus ayant cette valeur : $\boxed{f_i = \frac{n_i}{n}}$
- effectif cumulé de la i°valeur de X : nb. d'individus ayant l'une des i premières valeurs de X :

$$\boxed{N_i = \sum_{j=1}^{i} n_j = n_1 + n_2 + \ldots + n_j + \ldots n_i}$$

- fréquence cumulée d'une valeur de X : prop. des individus ayant l'une des i premières valeurs de X :

$$\boxed{F_i = \sum_{j=1}^{i} f_j = f_1 + f_2 + \ldots + f_j + \ldots f_i}$$

ménage	Nb. d'enfants X
1	3
2	2
3	5
4	3
5	6
6	3
7	5
8	5
9	1
10	5

Table des fréquences :

valeurs de X	effectif	effectif cum.	fréquence	fréq. Cum
1	1	1	0.1	0.1
2	1	2	0.1	0.2
3	3	5	0.3	0.5
5	4	9	0.4	0.9
6	1	10	0.1	1

B-2 Etude d'une variable quantitative

- Les différentes étapes de l'étude

  - ✓ Construction de la table des fréquences (par valeurs ou classes de valeurs).
  - ✓ Visualisation de la distribution des fréquences (ou des effectifs) de la série.
  - ✓ résumé des caractéristiques de la série par des indicateurs et des graphiques.

B-2.1 Etude d'une variable quantitative: Table des fréquences

Variable quantitative discrète

- ✓ classement des valeurs de x par ordre croissant
- ✓ Dénombrement des m valeurs distinctes de la série

$$\boxed{v_1 < \ldots < v_k < v_m}$$

Variable quantitative continue

- ✓ Création d'une série classée regroupement des valeurs de x en m classes (intervalles) disjointes de valeurs: $\boxed{I_k = [d_k, d_{k+1}[}$
- ✓ Définitions :
  - borne inférieure (resp.supérieure) de la classe I_k : d_k (resp. d_{k+1})
  - amplitude de I_k : $\boxed{a_k = d_{k+1} - d_k}$
  - centre de I_k : $\boxed{c_k = \frac{1}{2}(d_k + d_{k+1})}$
- ✓ NB : classement d'une série $\Rightarrow$ perte d'information; la constitution des classes est une étape délicate.

B-2.1 Etude d'une variable quantitative: Table des fréquences

Valeur de X	Effectif	Eff. cumulé	Fréquence	Fréq. cumulée
v_1	n_1	$N_1 = n_1$	f_1	$F_1 = f_1$
v_2	n_2	N_2		F_2
....	………	………………	……… …….	
v_i	n_i	$N_i = \sum_{j=1}^{i} n_j$	$f_i = \frac{n_i}{n}$	$F_i = \sum_{j=1}^{i} f_j$
……………… ……	………	………………		
v_k	n_k	$N_k = n$	f_k	$F_k = 1$

classe	amplitude	centre	effectif	Eff. cumulé	fréquence	Fréq. cumulée
$[d_1, d_2[$	a_1	c_1	n_1	$N_1 = n_1$	f_1	$F_1 = f_1$
$[d_2, d_3[$	a_2	c_2	n_2	N_2	f_2	F_2
$[d_i, d_{i+1}[$	a_i	c_i	n_3	N_i	f_i	F_i
$[d_k, d_{k+1}[$	a_k	c_k	n_k	$N_k = n$	f_k	$F_k = 1$

B-2.1 Etude d'une variable quantitative: Table des fréquences

ménage	Nb. d'enfants X
1	3
2	2
3	5
4	3
5	6
6	3
7	5
8	5
9	1
10	5

- ✓ Classement : 1,2,3,3,3,5,5,5,5,6
- ✓ Modalités : 1,2,3,5,6
- ✓ Table des fréquences :

valeurs de X	effectif	effectif cum.	fréquence	fréq. Cum
1	1	1	0.1	0.1
2	1	2	0.1	0.2
3	3	5	0.3	0.5
5	4	9	0.4	0.9
6	1	10	0.1	1

menage	1	2	3	4	5	6	7	8	9	10
superficie	8	8,5	10	12,5	11	13	20	25	33	15

- ✓ Nombre de classes par la règle de Sturges : k~5, amplitude des classes égales à E/k =33-8/5=5.
- ✓ Classes : [8,13[,[13,18[, [18,23[, [23,28[,[28,33].
- ✓ Table des fréquences :

classes	centres	eff.	eff.cum	freq.	freq.cum.
[8,13[	10,5	5	5	0,5	0,5
[13,18[	15,5	2	7	0,2	0,7
[18,23[	20,5	1	8	0,1	0,8
[23,28[	15,5	1	9	0,1	0,9
[28,33]	30,5	1	10	0,1	1

B-2.1 Etude d'une variable quantitative: Table des fréquences

Info

Règle de constitution des classes

- Le nombre de classes ne devrait être ni inférieur à 5, ni supérieur à 20 (il varie généralement entre 6 et 12). Ce choix est fonction du nombre d'observations et de leur dispersion. En pratique, on peut utiliser la *formule de Sturges : l*e nombre k indiqué de classes pour une série de n observations est donné approximativement par :

$$k = 1 + 3,322 \log_{10} n$$

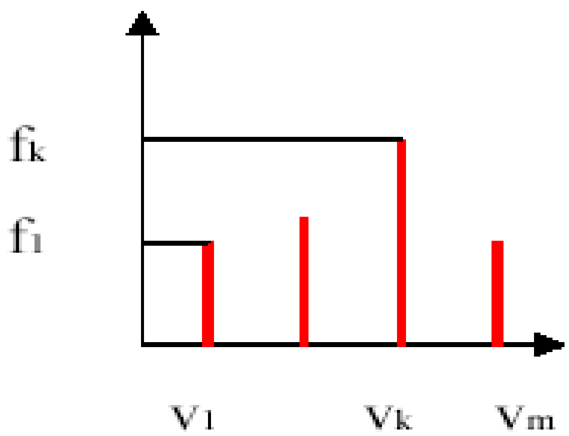
Cependant, le choix définitif du nombre de classes sera dicté par un souci de clarté.

- Il s'agit ensuite de choisir l'amplitude des classes. On les choisit généralement égales, d'amplitude approximativement égale à a=E/k où $E = x_{\max} - x_{\min}$ est l'étendue de la série.

B-2.2 Etude d'une variable quantitative: Visualisation

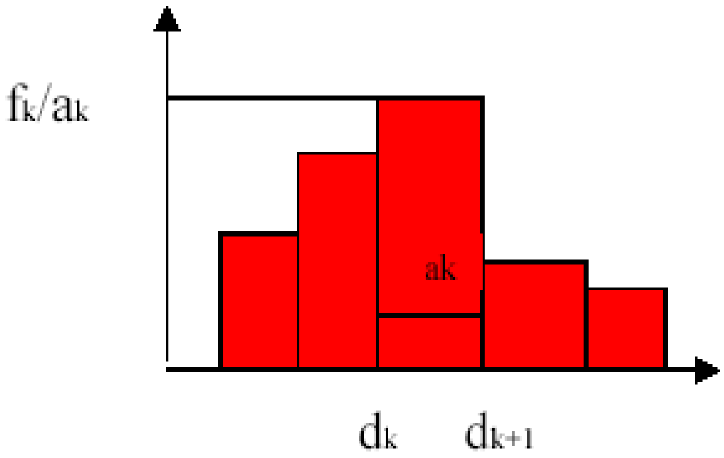
Variable quantitative discrète

- *Diagramme en bâtons* : valeurs de X en abscisse, bâton de longueur égale à la fréquence (ou à l'effectif) de ces valeurs en ordonnée.

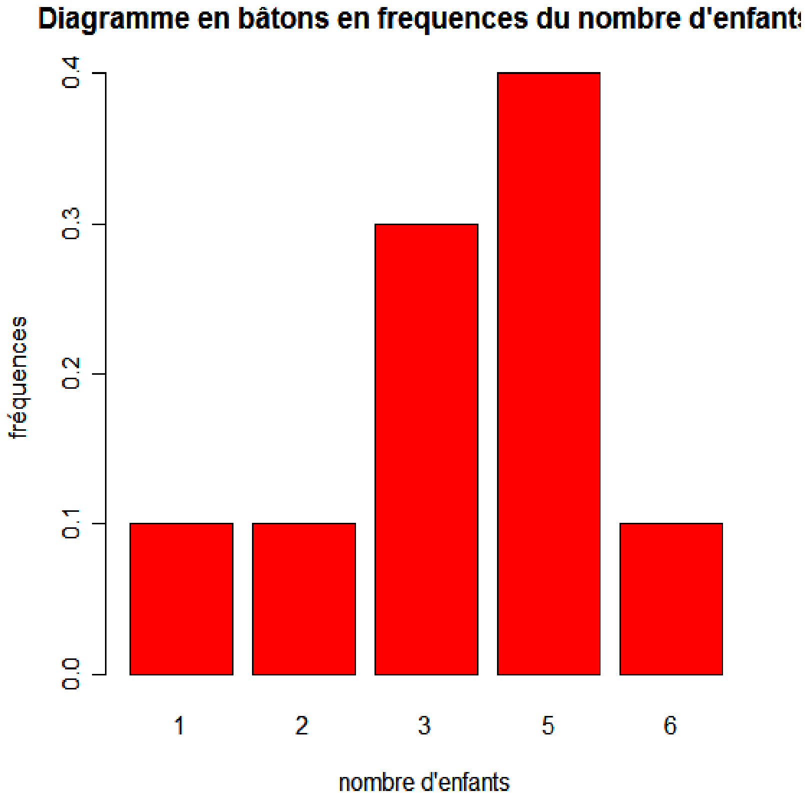


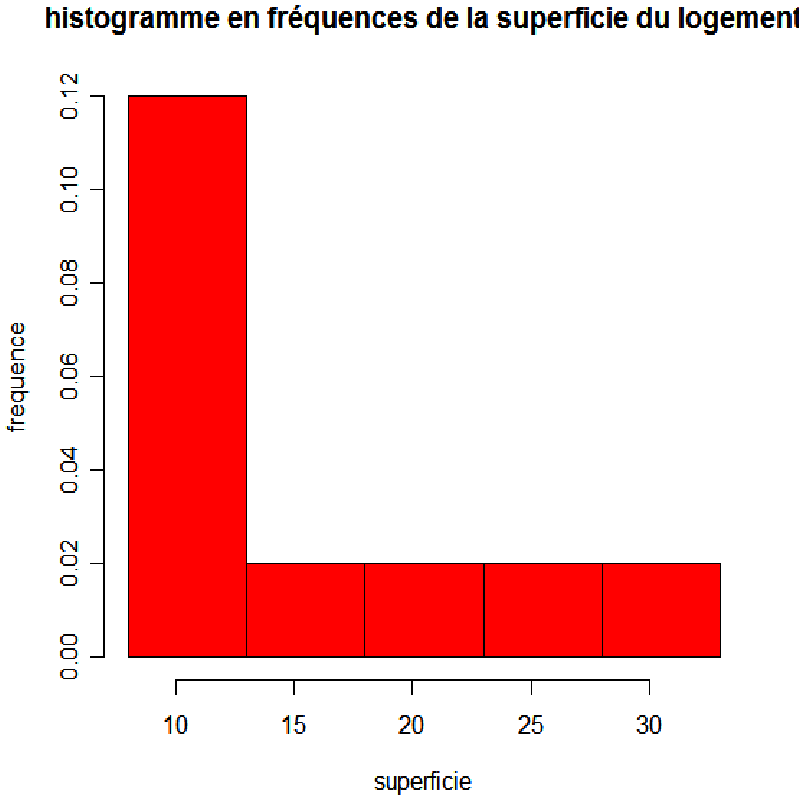
Variable quantitative continue

- *histogramme* : rectangles juxtaposés de base égale à a_k et de hauteur proportionnelle à la fréquence (ou effectif). Généralement, on prend comme hauteur f_k / a_k (l'aire de l'histogramme *est égale à 1*).



B-2.2 Etude d'une variable quantitative: visualisation





B-2.2 Etude d’une variable quantitative: Visualisation

Remarques :

- ✓ La distribution des fréquences d'une série statistique de la variable X, considérée comme un échantillon prélevé sur une population est une approximation de la distribution de probabilité de cette variable sur la population. C’est pourquoi il est préférable de tracer le diagramme en bâtons ou l’histogramme en fréquences plutôt que celui en effectifs

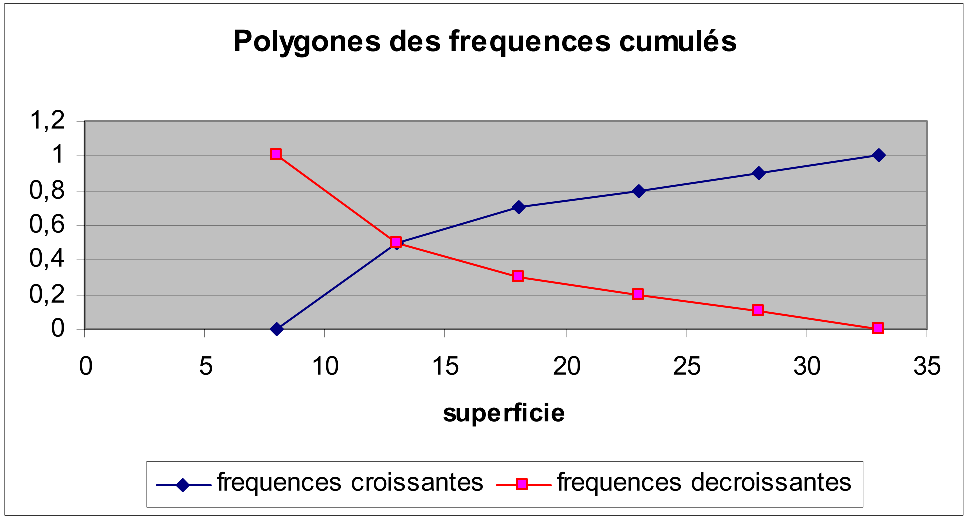
- ✓ La visualisation d’une série en fréquence permet la comparaison de plusieurs échantillons de tailles différentes.

B-2.2 Etude d'une variable quantitative: Visualisation

✓ Variable quantitative continue : les polygones des fréquences cumulées

- Objectif : Outils utiles pour répondre à des questions du type : quelle est la proportion (ou le nombre) de ménages ayant un logement de moins de 20 m% ? entre 40 et 60 m² ? Quelle est la valeur de la médiane (des quantiles) de la distribution ?

8	0	1
13	0,5	0,5
18	0,7	0,3
23	0,8	0,2
28	0,9	0,1
33	1	0



B-2.2 Etude d'une variable quantitative: Visualisation

- Méthode :
  - Faire un tableau :

bornes	d1	d2		di		dn	dn+1
<à	$p_1 = 0$	$p_2 = f_1$	$p_3 = F_2$	$p_i = F_{i-1}$		$p_n = F_{n-1}$	$p_{n+1} = F_n$
≥à	$q_1 = 1$	$q_2 = 1 - F_{i-1}$		$q_i = 1 - F_{i-1}$		$q_n = 1 - F_{n-1}$	0

  - Le polygone en fréquences croissantes (resp. décroissantes) est obtenu en traçant les points de coordonnées (d_k, p_k) (resp. (d_k, q_k)) et en interpolant linéairement entre ces points.

B-2.3 Etude d'une variable quantitative: Indicateurs

- ✓ Objectif : caractériser la distribution de la série à l'aide de nombres et éventuellement de graphiques résumant de façon suffisamment complète l'ensemble ses valeurs. Ces indicateurs faciliteront la comparaison d'échantillons.

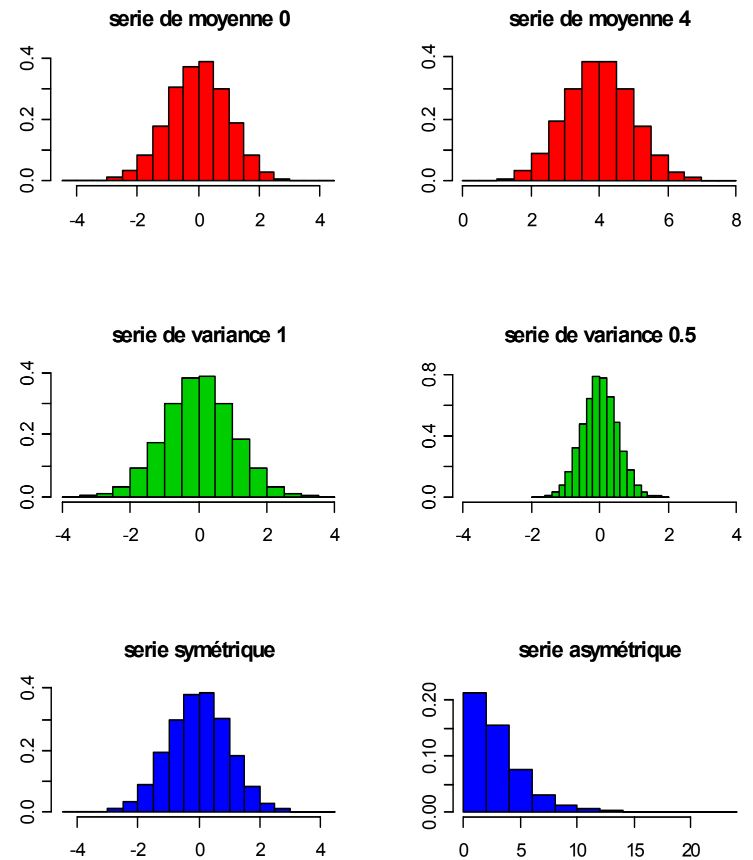
- ✓ 3 types d'indicateurs :
  - Indicateurs de tendance centrale
  - Indicateurs de dispersion
  - Indicateurs de forme

B-2.3 Etude d'une variable quantitative: Indicateurs

- Indicateurs *de tendance centrale* : fournissent *l'ordre de grandeur* des valeurs de la série et la position où se rassemblent ces valeurs.

- Indicateurs de *dispersion* : quantifient les *fluctuations* des valeurs autour de la valeur centrale. Permettent d'apprécier l'étalement des valeurs de la série (les unes par rapport aux autres ou à la valeur centrale).

- Indicateurs *de forme* : donnent une idée de la *symétrie* et de l'*aplatissement* d'une distribution. Leur usage est moins fréquent.



B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

- **La moyenne arithmétique**

  - ✓ Définition

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_i + \ldots + x_n}{n}}$$

- Sur une série discrète : $\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i v_i}$

- Sur série continue classée : $\boxed{\bar{x} \approx \frac{1}{n}\sum_{i=1}^{k} n_i c_i}$ (perte d'information)

B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

- Propriétés $$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

  La moyenne de la série $(ax_1 + b, \ldots, ax_n + b)$ est $a\bar{x} + b$

  Lorsque la distribution des fréquences est symétrique par rapport à la droite x=a, la moyenne vaut a.

- Limites

  Indicateur très affecté par les valeurs extrêmes (attention aux points aberrants).

B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

- **La médiane**

  - Définition : c'est la valeur *observée ou possible* de la série ordonnée en ordre croissant ou décroissant, qui partage cette série en deux sous-séries, chacune comprenant le même nombre d'observations.

    - si n impair $$\boxed{Me = x_{(n+1)/2}}$$
    - si n pair $$\boxed{Me = \frac{x_{(n/2)} + x_{(n/2)+1}}{2}}$$

NB : Si la variable est discrète et n pair, il se peut qu'il n'y ait pas de valeur médiane car Me doit correspondre à une valeur possible de la série.

Ex : dans la série du nombre d'enfants : 1,2,3,3,3,5,5,5,5,6, Me=4.

dans la série de la superficie : 8,8.5,10,11,12.5,13,15,20,25,33, Me=12,75.

B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

- ✓ Limites : La médiane est plus robuste que la moyenne (pas influencée par les valeurs extrêmes) mais elle est influencée par le nombre d'observations.

Remarque : La médiane correspond à la valeur telle que la fréquence cumulée est égale à ½.

B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

- ✓ CP d'une série continue classée: Approximation de Me à partir de la table des fréquences par interpolation linéaire.

- Repérage de la classe médiane = première classe contenant au moins 50% des effectifs cumulés

$$I_j = [d_j, d_{j+1}[$$

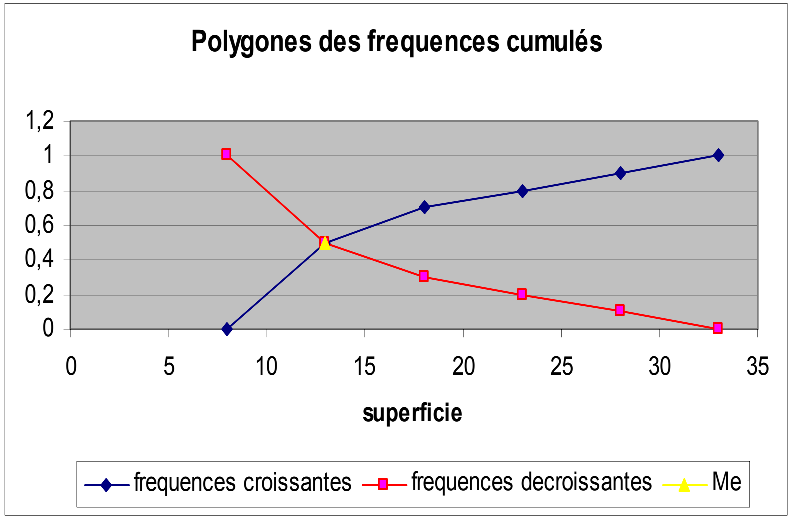
- Interpolation linéaire

$$\boxed{Me \approx d_j + \frac{\frac{n}{2} - N_{j-1}}{n_j} \times a_j}$$

Ex : Série superficie :

Par la définition : Me=12. 75

Par interpolation : Me~13



B-2.3 Etude d’une variable quantitative: Indicateurs de tendance centrale

- **Le mode**

  - ✓ Définition : c’est la valeur qui a été observée le plus grand nombre de fois.

NB : Dans le cas d'une variable continue en classes, ce critère est peu objectif. On parlera plutôt de classe modale : classe ayant la fréquence la plus élevée. Le mode n’est pas unique.

Ex : série nombre d’enfants : mode=5; série superficie : intervalle modal= [8,13[.

B-2.3 Etude d'une variable quantitative: Indicateurs de tendance centrale

Info

Il existe d'autres moyennes moins utilisées car elles ne disposent pas des propriétés algébriques valables pour la moyenne arithmétique :

La moyenne **géomètrique** : $g = \sqrt[n]{\prod f_i x_i}$

La moyenne **harmonique** : $\frac{1}{h} = \frac{1}{n} \sum \frac{f_i}{x_i}$

La moyenne **quadratique** : $q = \sqrt[n]{f_i x_i^2}$

On a la relation suivante entre ces moyennes : $h < g < \bar{x} < q$.

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

➢ **La variance et l'écart-type de la série**

✓ Définition : La variance est la somme pondérée des carrés des écarts des valeurs de la série à la moyenne.

- Variance de la série $$\boxed{s_x^2 = s^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

- Variance d'échantillonnage $$\boxed{s_x^{*2} = s^{*2}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

L'écart type est la racine carrée de la variance $\boxed{s_x = \sqrt{s_x^2}}$ $\boxed{s_x^* = \sqrt{s_x^{*2}}}$

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

Lorsque la série est un échantillon issu d'une population et que l'on s'intéresse aux caractéristiques de cette population via l'échantillon (inférence), on utilise plutôt s_n^{*2} qui est un meilleur estimateur de la variance théorique de la population. Dès lors que la taille n de la série est assez grande, ces deux quantités sont pratiquement égales.

Info

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

✓ Propriétés $(s_n^2 \text{ ou } s_n^{*2})$

- La variance (ou écart-type) est toujours positive ou nulle $\boxed{s^2_x \geq 0}$ $\boxed{s_x \geq 0}$
- La variance est une forme quadratique $\boxed{s_{ax+b}^2 = a^2 s_x^2}$ $\boxed{s_{ax+b} = |a| s_x}$
- Théorème de Koenig $\boxed{s_x^2 = \dfrac{n-1}{n} s_x^{*2} = \overline{x^2} - \overline{x}^2}$

Une série peu dispersée (ayant des valeurs regroupées autour de la valeur moyenne) aura un écart-type plutôt faible.

Remarque : Pour une distribution symétrique, pratiquement toutes les observations sont situées entre x-3s et x+3s.

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

Lorsqu'on fait de l'inférence, un faible écart-type de l'échantillon permettra d'indiquer avec une plus grande précision entre quelles valeurs peuvent varier les caractéristiques de la distribution de la variable étudiée sur la population.

info

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

✓ Calcul pratique de la variance (ou de l'écart-type):

- Par la définition
- Par la formule de Koenig
- A partir de la table des fréquences
  - Pour une série discrete
    $$\boxed{s_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (v_i - \overline{x})^2}$$
  - Pour une serie en classes
    $$\boxed{s_x^2 \approx \frac{1}{n}\sum_{i=1}^{k} n_i (c_i - \overline{x})^2}$$

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

- **Une mesure de la dispersion relative : le coefficient de variation**

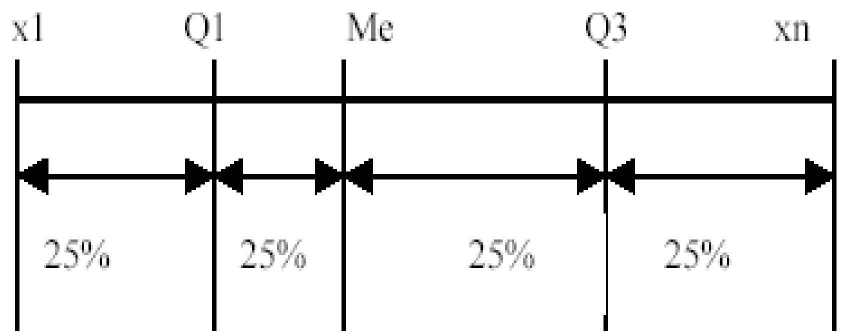
$$\boxed{CV = \frac{s_X}{\overline{x}}.100}$$

Le CV permet d'apprécier la représentativité de la moyenne par rapport à l'ensemble des observations. Il donne une bonne idée du degré d'homogénéité d'une série. Il faut qu'il soit le plus faible possible (<15% en pratique).

B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

- **Les quantiles**

  - ✓ Définition : ils correspondent à des valeurs de la variable statistique qui partagent la série ordonnée en l parties égales. Si l=4, les quantiles sont appelés quartiles. Il y a 3 quartiles, appelés Q1,Q2=Me et Q3 :



$\boxed{I = Q3 - Q1}$ est appelé *l'intervalle interquartile* et comporte 50\% de la série. C'est un indicateur de dispersion de la série.

B-2.3 Etude d’une variable quantitative: Indicateurs de dispersion

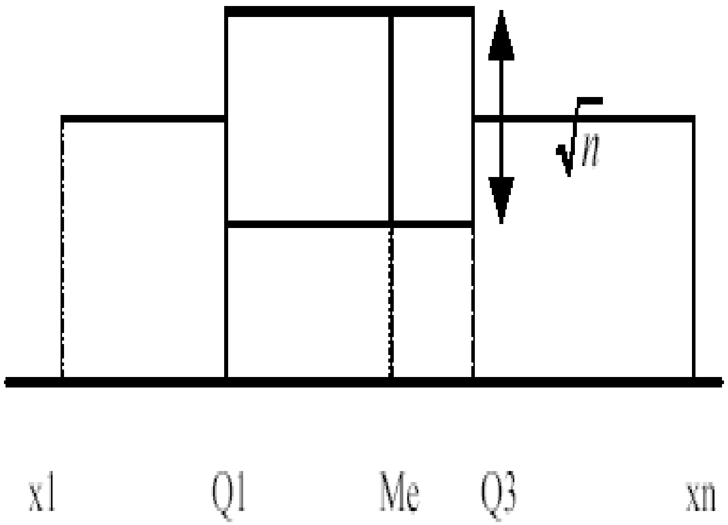
- **La boite à moustaches (box-and-Wiskers plot)**

- Résume la série à partir de ses valeurs extrêmes, ses quartiles et sa médiane.

- Permet une comparaison visuelle immédiate de plusieurs séries.

- Construction :

- Sur un axe horizontal, on place les valeurs extrêmes et les quartiles.
- on trace un rectangle de longueur l'interquartile et la largeur proportionnelle à la racine carrée de la taille de la série.
- on partage le rectangle par un segment vertical au niveau de la médiane.

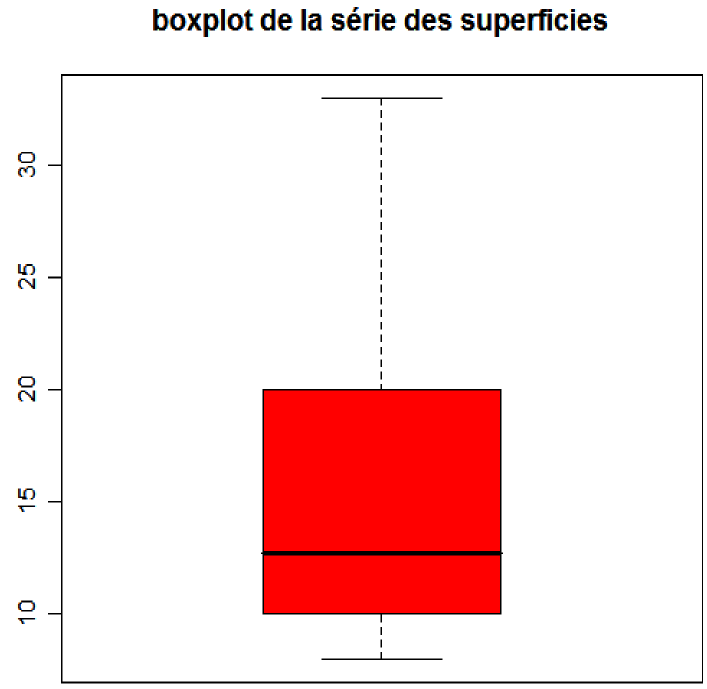


B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

Série des superficies :

8,8.5,10,11,12.5,13,15,20,25,33

Min.	Q1	Me	Mean	Q3	Max.
8.00	**10.25**	**12.75**	**15.60**	**18.75**	**33.00**



B-2.3 Etude d'une variable quantitative: Indicateurs de dispersion

✓ Autres indicateurs :

Info

– L'étendue $E = x_{\max} - x_{\min}$

– L'écart arithmétique moyen $e = \dfrac{1}{n}\sum \left| x_i - \overline{x} \right|$

B-2.4 Etude d'une variable quantitative: Indicateurs de forme

- **Symétrie**

  - Définition : Une série a une distribution symétrique si ses valeurs sont également dispersées de part et d'autre de la valeur centrale, c'est-à-dire si le graphe de la distribution - histogramme ou diagramme en bâton en fréquences - admet une axe de symétrie.

Dans une distribution parfaitement symétrique,

$$\boxed{Me = \overline{x} = Mode}$$

B-2.4 Etude d'une variable quantitative: Indicateurs de forme

- ✓ Coefficient d'asymétrie de Pearson

$$\boxed{\delta = \frac{\overline{x} - Me}{s_x}}$$

- ✓ Coefficient de Yule

$$\boxed{q = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1}}$$

On a $-1 \leq \delta \leq 1$

$\delta = 0 \Rightarrow$ symétrie parfaite

$\delta < 0 \Rightarrow$ série étalée à gauche

$\delta > 0 \Rightarrow$ Série étalée à droite

$q = 0 \Rightarrow$ symétrie parfaite

$q < 0 \Rightarrow$ série étalée à gauche

$q > 0 \Rightarrow$ série étalée à droite

B-2.4 Etude d’une variable quantitative: Indicateurs de forme

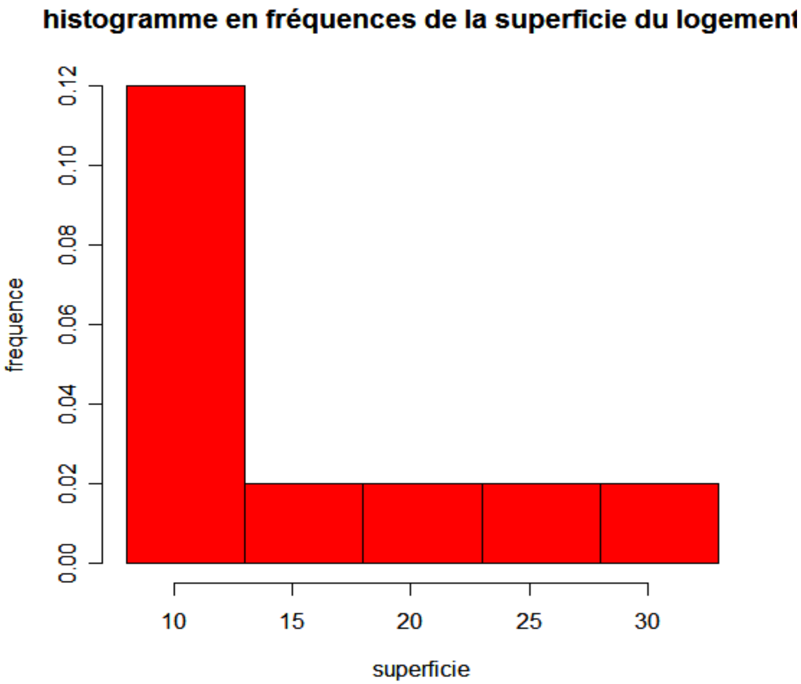
Ex : Série des superficies :

8,8.5,10,11,12.5,13,15,20,25,33

Min.	Q1	Me	Mean	Q3	Max.
8.00	**10.25**	**12.75**	**15.60**	**18.75**	**33.00**

S= 8.082216
d=1.057878
Q=0.4117647

- Série étalée à droite



B-2.4 Etude d'une variable quantitative: Indicateurs de forme

- **Applatissement**

Une distribution est plus ou moins aplatie selon que les fréquences des valeurs voisines des valeurs centrales diffèrent peu ou beaucoup les une par rapport aux autres.

- ✓ coefficient d'aplatissement de Fisher :

$$\boxed{a = \frac{m_4}{s_x^4}} \qquad m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4$$

- a=3 pour une distribution qui suit une loi normale centrée réduite.
- Si a>3, la concentration des valeurs de la série autour de la moyenne est forte : la distribution n'est pas aplatie
- Si a<3, la concentration des valeurs autour de la moyenne est faible : la distribution est aplatie

B-3 Etude d'une variable qualitative

- Table des fréquences :

- Lorsque la variable est ordinale, elle est construite de manière analogue à celle d'une variable quantitative discrète

- Lorsque la variable est nominale, n'y figurent pas les effectifs et fréquences cumulées.

B-3 Etude d'une variable qualitative

✓ Construction

- Dénombrement des modalités différentes de la série $m_1, \ldots, m_i, \ldots m_k$
- Table de la distribution des fréquences :

modalités de X	Effectif	Fréquence
m_1	n_1	f_1
m_2	n_2	..
....	....	....
.. m_i	n_i	$f_i = \frac{n_i}{n}$
....	....	
m_k	n_k	f_k

B-3 Etude d'une variable qualitative

- Visualisation : diagramme en barres (analogue au diagramme en bâtons) ou représentation en secteurs (camembert), représentant la répartition en effectif ou en fréquences des individus dans les différentes modalités de la série.

- Indicateurs : Il n'existe pas, à part le mode de caractéristiques communément adaptées pour décrire une variable qualitative.

B-3 Etude d'une variable qualitative

Exemple : observation de la séquence d'un brin d'ADN

GGGAGTGTBTATTAABTBBGAA
BTBBBAGBGBTAGBTBGBGBGG
AGTGABBGAGBBTABATGAGGG
TABTGTBAATAABGBATGTTABB
AGAAGGA

valeurs	effectifs	frequences
A	26	0,26
C	27	0,27
G	27	0,27
T	20	0,2