STATISTIQUES DESCRIPTIVES

INTRODUCTION

VOCABULAIRE STATISTIQUE

Population statistique :

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

Individu (ou unités statistiques) :

Les individus sont les éléments de la population statistique étudiée.

Caractère statistique ou variable statistique :

C'est ce qui est observé ou mesuré sur les individus d'une population statistique.

VARIABLES QUANTITATIVES

Variable quantitative :

Une variable statistique est quantitative si ses valeurs sont des nombres exprimant une quantité, sur lesquels les opérations arithmétiques (somme, etc...) ont un sens.

Variable quantitative discrète:

Une variable quantitative est discrète si elle ne peut prendre que des valeurs isolées, généralement entières.

Variable quantitative continue:

Une variable quantitative est continue si ses valeurs peuvent être n'importe lesquelles d'un intervalle réel.

VARIABLES QUALITATIVES

Variable qualitative :

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.

Variable qualitative nominale :

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

Variable qualitative ordinale :

C'est une variable qualitative dont les modalités sont naturellement ordonnées

(1) UN OUTIL : L 'OPERATEUR SOMME Σ

DEFINITION:

p et q étant 2 entiers relatifs

$$\sum_{i=p}^{q} x_i = x_p + x_{p+1} + \ldots\ldots + x_q$$

REMARQUE 1: i est une variable muette

$$\sum_{i=p}^{q} x_i = \sum_{j=p}^{q} x_j = \sum_{h=p}^{q} x_h$$

REMARQUE 2:

Quand il n'y a pas d'ambiguïté sur le domaine de variation de i, celui-ci peut être omis

$$\sum_{i=1}^{n} x_i = \sum_{i} x_i = \sum x_i$$

(2) UN OUTIL : L 'OPERATEUR SOMME Σ

PROPRIETE 1: $\sum_i ka_i = k\sum_i a_i$

$$\sum_{i=p}^{q} ka_i = ka_p + ka_{p+1} + \ldots\ldots + ka_q = k\left(a_p + a_{p+1} + \ldots\ldots + a_q\right) = k\sum_{i=p}^{q} a_i$$

PROPRIETE 2: $\sum_i (a_i + b_i) = \sum_i a_i + \sum_i b_i$

$$\sum_{i=p}^{q} (a_i + b_i) = (a_p + b_p) + (a_{p+1} + b_{p+1}) + \ldots\ldots + (a_q + b_q)$$

PROPRIETE 3 : $\sum_i k(a_i + b_i) = k\sum_i a_i + k\sum_i b_i$

$$= \left(a_p + a_{p+1} + \ldots\ldots + a_q\right) + \left(b_p + b_{p+1} + \ldots\ldots + b_q\right) = \sum_{i=p}^{q} a_i + \sum_{i=p}^{q} b_i$$

PROPRIETE 4 : $\sum_{i=1}^{n} k = nk$

$$\sum_{i=1}^{n} k = \underbrace{k + k + \ldots\ldots + k}_{n} = nk$$

PROPRIETE 5: $\sum_{i=p}^{q} k = (q - p + 1)k$

TABLEAUX ET GRAPHIQUES

(1) VARIABLES QUALITATIVES NOMINALES

Noms	Couleur des yeux
M. Alberro	Vert
M. Hondarrague	Noir
Mme Claverotte	Noir
Melle Lopez	Noisette
M. Paulien	Bleu
M. Guillou	Noir
M. Lahitette	Noisette
Mme Vigouroux	Noir
Melle Maleig	Bleu
M. Duclos	Vert
M. Carricaburu	Bleu
Mme Vidal	Noir
….	….

Modalités	Effectifs	Fréquences	%
Bleu	60	0,200	20,0
Noir	160	0,533	53,3
Noisette	40	0,133	13,3
Vert	40	0,133	13,3
Total :	300	1	100

Modalités	Effectifs	Fréquences	%
modalité 1	n_1	$f_1 = n_1/n$	$f_1 \times 100$
…	…	…	
modalité i	n_i	$f_i = n_i/n$	$f_i \times 100$
…	…	…	
modalité k	n_k	$f_k = n_k/n$	$f_k \times 100$
Total :	$\sum n_i = n$	$\sum f_i = 1$	100

(2) VARIABLES QUALITATIVES NOMINALES

Modalités	Effectifs	Fréquences	%
Bleu	60	0.200	20,0
Noir	160	0,533	53,3
Noisette	40	0,133	13,3
Vert	40	0,133	13,3
Total :	300	1	100

Diagramme circulaire ou camembert

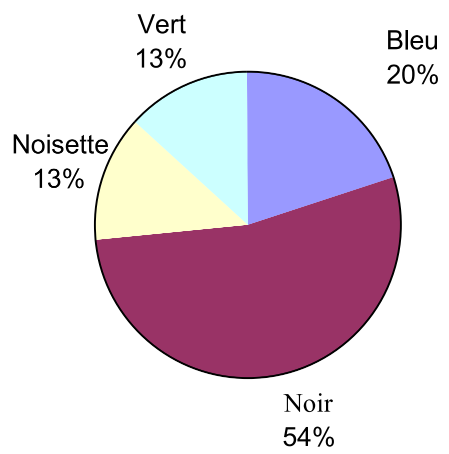
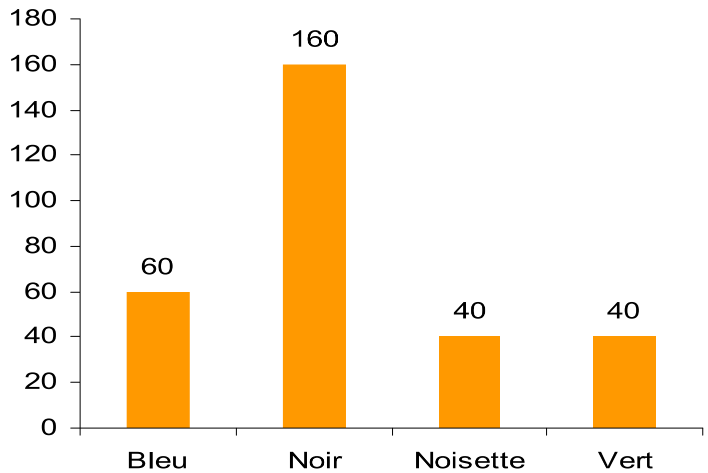


Diagramme en barres

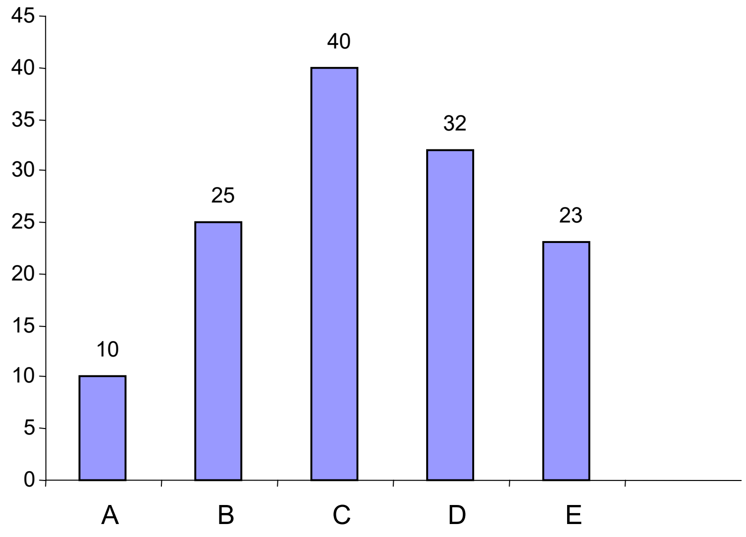


VARIABLES QUALITATIVES ORDINALES

130 personnes ont été interrogées sur leur addiction au chocolat

Les modalités sont présentées dans l'ordre

Modalités	Effectifs = Nombre de personnes
Pas du tout (A)	10
Un peu (B)	25
Beaucoup (C)	40
Passionnément (D)	32
A la folie (E)	23



(1) VARIABLES QUANTITATIVES DISCRETES

EFFECTIFS ET FREQUENCES

Clients	Nombre de produits financiers
Bredat	2
Gauguet	3
Leremboure	0
Coustere	0
Lalisou	1
Aussagne	0
Vittorello	1
Diaz	0
Etcheverry	2
Bernadet	4
Miramon	1
Jaime	3
Dartus	2
Domege	0
Train	0
Piquemal	1
Laffargue	2
……	…….

Nombre de produits financiers	Nombre de clients
0	103
1	115
2	95
3	35
4	10
5	2

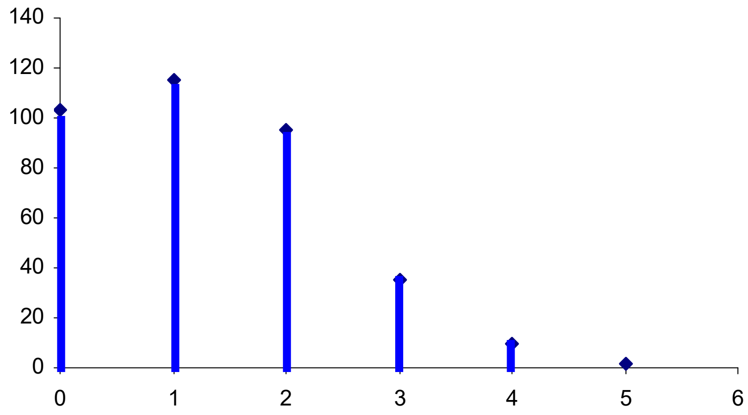
Valeurs de la variable	Effectifs	Fréquences	%
x_1	n_1	$f_1 = n_1/n$	$f_1 \times 100$
…	…	…	
x_i	n_i	$f_i = n_i/n$	$f_i \times 100$
…	…	…	
x_k	n_k	$f_k = n_k/n$	$f_k \times 100$
Total :	$\sum n_i = n$	$\sum f_i = 1$	100

(2) VARIABLES QUANTITATIVES DISCRETES

REPRESENTATION GRAPHIQUE DES EFFECTIFS ET FREQUENCES

Nbre de produits financiers x_i	Effectif n_i	Fréquence f_i
0	103	0,286
1	115	0,319
2	95	0,264
3	35	0,097
4	10	0,028
5	2	0,006

Diagramme en bâtons



(3) VARIABLES QUANTITATIVES DISCRETES

EFFECTIFS ET FREQUENCES CUMULES

Effectifs cumulés croissants:

Nombre d'individus pour lesquels la variable est **inférieure ou égale** à x_i. Résultat de l'addition, de proche en proche, des effectifs d'une distribution observée en commençant par le 1er.

Nbre produits financiers	Nombre de Clients	Effectifs cumulés croissants	Effectifs cumulés décroissants
0	103	103	360
1	115	218	257
2	95	313	142
3	35	348	47
4	10	358	12
5	2	360	2
Total :	360		

Effectifs cumulés décroissants:

Nombre d'individus pour lesquels la variable est **supérieure ou égale** à x_i. Résultat de l'addition, de proche en proche, des effectifs d'une distribution observée en commençant par le dernier.

Valeurs de la variable x_i	Effectif n_i	Effectifs cumulés croissants N_i	Effectifs cumulés décroissants N'_i
x_1	n_1	$N_1 = n_1$	$N'_1 = n_k + \ldots + n_1 = n$
x_2	n_2	$N_2 = n_1 + n_2$	$N'_2 = n_k + \ldots + n_2$
x_3	n_3	$N_3 = n_1 + n_2 + n_3$	$N'_3 = n_k + \ldots + n_3$
…	…	….	….
x_{k-1}	n_{k-1}	$N_{k-1} = n_1 + \ldots + n_{k-1}$	$N'_{k-1} = n_k + n_{k-1}$
x_k	n_k	$N_k = n_1 + \ldots + n_k = n$	$N'_k = n_k$
Total :	n		

(4) VARIABLES QUANTITATIVES DISCRETES
EFFECTIFS ET FREQUENCES CUMULES

Nombre de produits financiers x_i	Nombre de clients n_i	Effectifs cumulés croissants N_i	Effectifs cumulés décroissants N'_i	Fréquences f_i	Fréquences cumulées croissantes F_i	Fréquences cumulées décroissantes F'_i
0	103	103	360	0,2861	0,2861	1
1	115	218	257	0,3194	0,6055	0,7139
2	95	313	142	0,2639	0,8694	0,3945
3	35	348	47	0,0972	0,9666	0,1306
4	10	358	12	0,0278	0,9944	0,0334
5	2	360	2	0,0056	1	0,0056
Total :	360			1		

Il y a 313 clients possédant un nombre de produits financiers inférieur ou égal à 2

Il y a 47 clients possédant un nombre de pro. fin. supérieur ou égal à 3

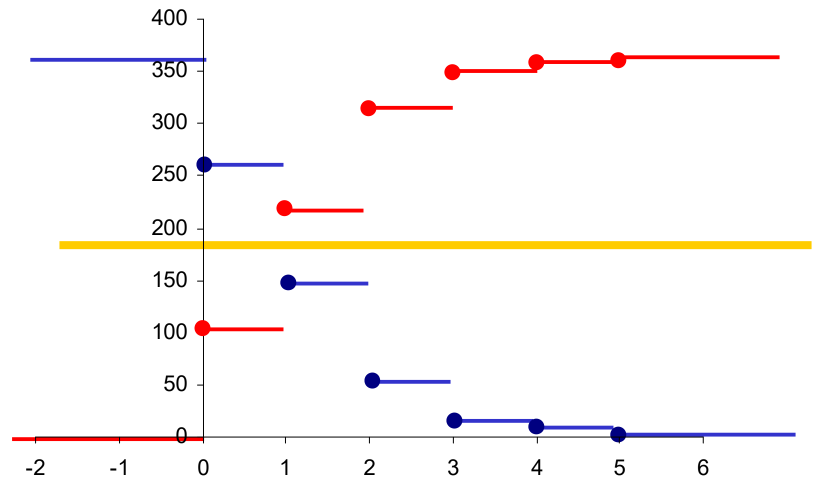
La proportion de clients possédant un nombre de pro. fin. inférieur ou égal à 4 est de 99,44%

La proportion de clients possédant un nombre de pro. fin. supérieur ou égal à 1 est de 71,39%

(5) VARIABLES QUANTITATIVES DISCRETES

COURBES CUMULATIVES

x	x_i	n_i	N_i	N(x)	N'_i	N '(x)
$-\infty$				0		360
0	0	103	**103**		**360**	
				103		257
1	1	115	**218**		**257**	
				218		142
2	2	95	**313**		**142**	
				313		47
3	3	35	**348**		**47**	
				348		12
4	4	10	**358**		**12**	
				358		2
5	5	2	**360**		**2**	
$+\infty$				360		0



On appelle courbe cumulative croissante le tracé de la fonction N (ou F pour les fréquences) qui à **tout réel x** associe N(x) = nombre d'observations inférieur ou égal à x.

On appelle courbe cumulative décroissante le tracé de la fonction N' (ou F' pour les fréquences) qui a **tout réel x** associe N'(x) = nombre d'observations supérieur strictement à x.

Les courbes cumulatives N(x) et N'(x) sont symétriques par rapport à n/2 : N(x) + N'(x) = n

Les courbes cumulatives F(x) et F'(x) sont symétriques par rapport à 0,5 : F(x) + F'(x) = 1

(1) VARIABLES QUANTITATIVES CONTINUES

Variable observée: augmentation moyenne mensuelle du salaire, en €, des employés d'une multinationale au cours de l'année 2005.

18	38	10	35	0	4
4	11	27	2	41	16
2	25	43	22	26	11
34	34	1	28	5	5
21	0	2	30	1	8
9	37	22	39	11	0
36	16	6	42	42	1
8	33	31	33	4	4
9	19	15	2	21	0
12	18	....	....	....	....

Augmentation (€)	Effectif
0	257
1	318
2	255
3	307
4	308
5	159
6	140
7	84
8	72
9	55
10	22
11	13
12	9
13	7
14	8
15	21
16	6
17	2
…..	….
Total	2125

Remarque1 : la variable augmentation moyenne mensuelle peut être considérée comme continue. En arrondissant à l'euro, on l'a discrétisée.
Une augmentation de 10 € est en fait une augmentation comprise entre 9,5 € et 10,5 €.

Remarque2 : Une variable continue ne prend pas des valeurs isolées, mais des valeurs appartenant à des intervalles. C'est pourquoi, au lieu de définir des effectifs par valeurs, on définira des effectifs par intervalles, appelés **classes**.

Remarque3 : Une variable discrète comportant trop de valeurs est aussi traitée comme une variable continue.

(2) VARIABLES QUANTITATIVES CONTINUES

Augmentation (€)	Effectifs
[0 – 3[	830
[3 – 5[	615
[5 – 10[	510
[10 – 20[	92
[20 – 30[	63
[30 – 50[	15

Classes	Effectifs
$[e_1 - e_2[$	n_1
$[e_2 - e_3[$	n_2
….	….
$[e_k - e_{k+1}[$	n_k

Remarque 1: Le choix des classes et arbitraire, mais elles doivent être contigües et recouvrir l'ensemble des valeurs.

Remarque 2: Il est préférable de prendre des classes d'amplitudes égales.

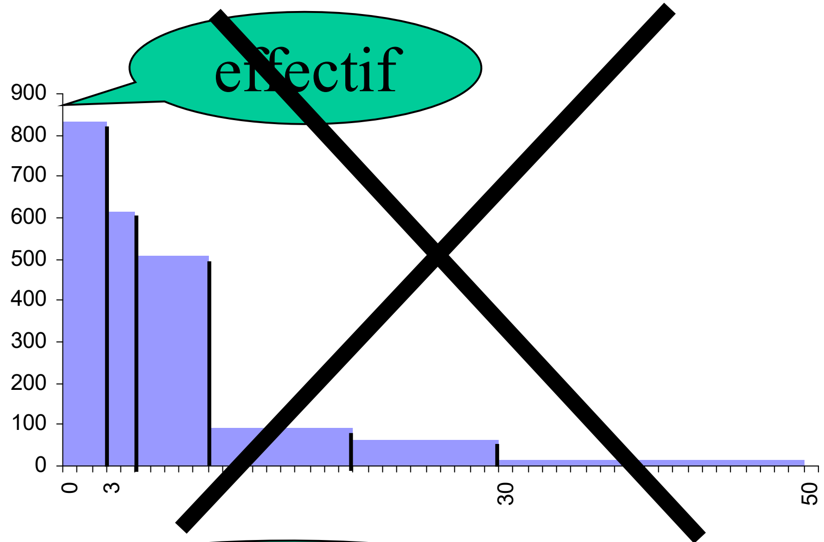
Remarque 3: Il ne faut prendre ni trop ni trop peu de classes.

Remarque 4: Le choix et le nombre de classes influent sur les représentations graphiques.

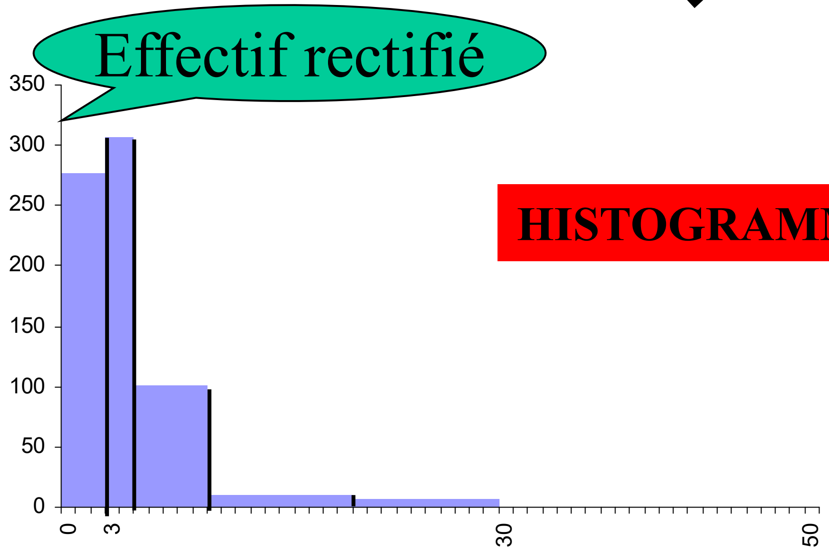
(3) VARIABLES QUANTITATIVES CONTINUES

REPRESENTATION GRAPHIQUE DES EFFECTIFS ET FREQUENCES

Classes	Effectifs
[0 – 3[	830
[3 – 5[	615
[5 – 10[	510
[10 – 20 [	92
[20 – 30[	63
[30 – 50[	15



Classes	Effectifs n_i	Amplitude a_i	Effectifs rectifiés n_i /a_i
[0 – 3[	830	3	276,7
[3 – 5[	615	2	307,5
[5 – 10[	510	5	102,0
[10 – 20 [	92	10	9,2
[20 – 30[	63	10	6,3
[30 – 50[	15	20	0,75

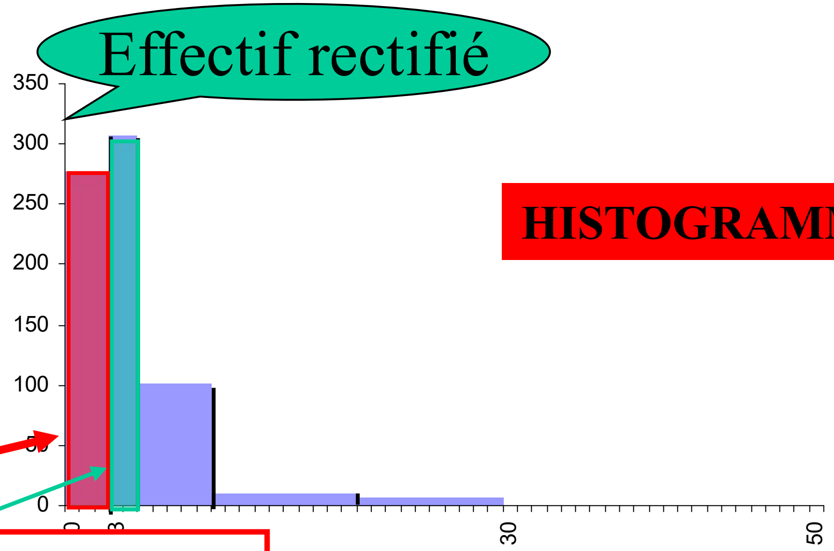


HISTOGRAMME

(4) VARIABLES QUANTITATIVES CONTINUES

REPRESENTATION GRAPHIQUE DES EFFECTIFS ET FREQUENCES

Classes	Effectifs n_i	Amplitude a_i	Effectifs rectifiés n_i /a_i
[0 – 3[	830	3	276,7
[3 – 5[	615	2	307,5
[5 – 10[	510	5	102,0
[10 – 20[	92	10	9,2
[20 – 30[	63	10	6,3
[30 – 50[	15	20	0,75



HISTOGRAMME

La surface = $a_i \times (n_i/a_i)$ est de 830 unités

La surface = $a_i \times (n_i/a_i)$ est de 615 unités

Dans un histogramme, ce sont les surfaces des rectangles (ce que l'œil voit), qui sont proportionnelles aux effectifs, et non les hauteurs de ces rectangles

Remarque: Le tracé de l'histogramme des fréquences est identique. Il suffit de porter en ordonnées la fréquence rectifiée $d_i = f_i/a_i$, appelée densité.

(5) VARIABLES QUANTITATIVES CONTINUES
EFFECTIFS ET FREQUENCES CUMULES

Variable observée: augmentation moyenne mensuelle du salaire, en €, des employés d'une multinationale au cours de l'année 2005.

Classes $[e_i - e_{i+1}[$	Effectifs n_i	Effectifs cumulés croissants N_i	Effectifs cumulés décroissants N'_i	Fréquences cumulées croissantes F_i	Fréquences cumulées décroissantes F'_i
[0 – 3[	830	830	2125	0,391	1,000
[3 - 5 [	615	1445	1295	0,680	0,609
[5 - 10 [	510	1955	680	0,920	0,320
[10 - 20 [	92	2047	170	0,963	0,080
[20 - 30 [	63	2110	78	0,993	0,037
[30 – 50[	15	2125	15	1,000	0,007
Total :	2125				

Il y a 1445 employés dont l'augmentation est strictement inférieure à 5

Il y a 170 employés dont l'augmentation est supérieure ou égale à 10

Combien y-a-t-il d'employés dont l'augmentation est inférieure à 17 ?

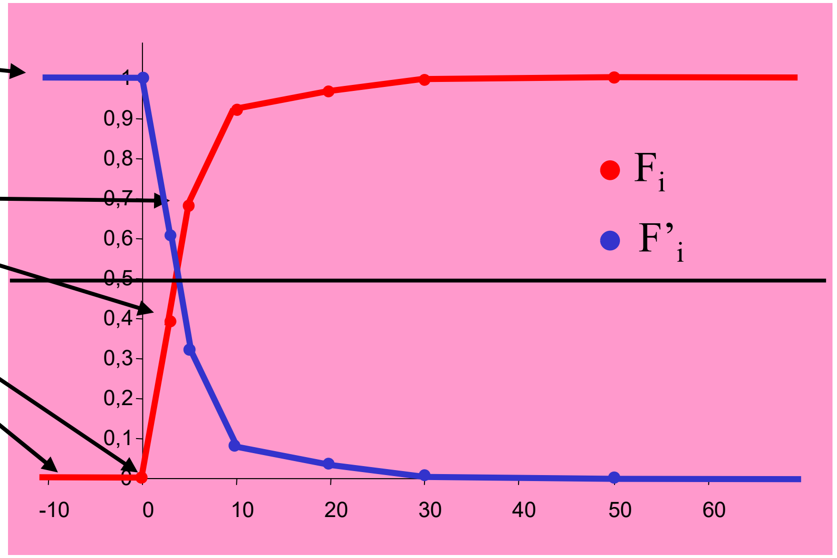
(6) VARIABLES QUANTITATIVES CONTINUES

COURBES CUMULATIVES

$[e_i - e_{i+1}[$	F_i	F'_i
[0 - 3[	0,391	1,000
[3 - 5[	0,680	0,609
[5 - 10[	0,920	0,320
[10 - 20[	0,963	0,080
[20 - 30[	0,993	0,037
[30 - 50[	1,000	0,007

x	F(x)	F'(x)
−∞	?	?
0	0	1
3	0,391	0,609
5	0,680	0,320
10	0,920	0,080
20	0,963	0,037
30	0,993	0,007
50	1	0
+∞		

ntérieur chaque e, on fait pothèse ue la tition est iforme



On appelle courbe cumulative croissante le tracé de la fonction F (N pour les effectifs) qui à tout réel x associe **F(x)** = nombre d'observations inférieur ou égal à x.

On appelle courbe cumulative décroissante le tracé de la fonction F' (N' pour les effectifs) qui a tout réel x associe F'(x) = nombre d'observations supérieur strictement à x.

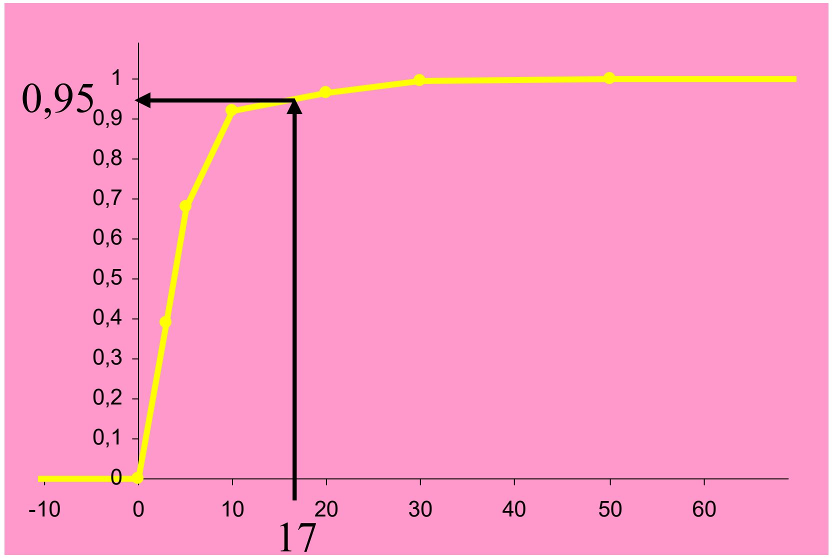
Remarque: Pour une variable continue, il est indifférent de dire « inférieur ou égal » ou « strictement inférieur ». Il en est de même pour « supérieur ou égal » ou « strictement supérieur » ... l'instrument de mesure et un mauvais choix des bornes qui pourrait conduire à ce résultat.

Les courbes cumulatives F(x) et F'(x) sont symétriques par rapport à 0,5 : F(x) + F'(x) = 1

(7) VARIABLES QUANTITATIVES CONTINUES
COURBES CUMULATIVES

Quelle est la proportion p d'employés dont l'augmentation est inférieure à 17 € ?

x	$[e_i - e_{i+1}[$	F_i	F(x)
0			0
	[0 - 3 [	0,391	
3			0,391
	[3 - 5 [	0,680	
5			0,680
	[5 - 10 [	0,920	
10			0,920
17	[10 - 20 [	0,963	p
20			0,963
	[20 - 30 [	0,993	
30			0,993
	[30 - 50 [	1	
50			1



$$\frac{17 - 10}{20 - 10} = \frac{p - 0{,}92}{0{,}963 - 0{,}920}$$

D'où $p = 0{,}92 + \dfrac{17-10}{20-10}(0{,}963 - 0{,}920) \approx 95\%$

TABLEAUX ET GRAPHIQUES

RESUME

VARIABLE QUALITATIVE

Nominale

Effectifs ou Fréquences

Diagramme en barres

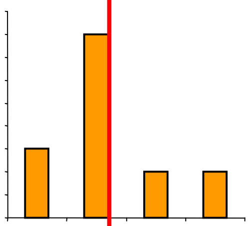
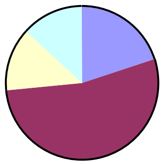
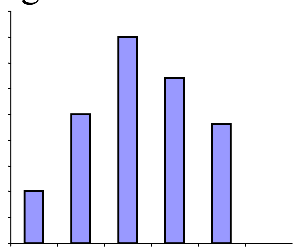


Diagramme circulaire



Ordinale

Diagramme en barres



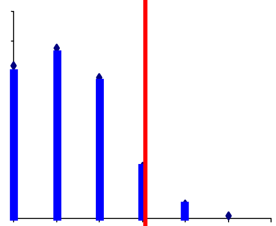
Modalités dans l 'ordre

VARIABLE QUANTITATIVE

Discrète

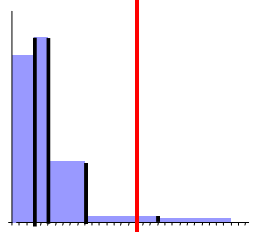
Effectifs ou Fréquences

Diagramme en bâtons

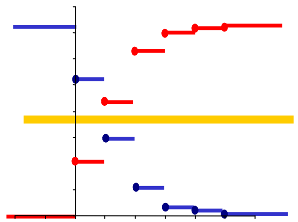


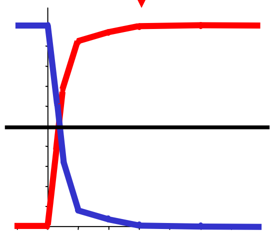
Continue

Histogramme



Courbes cumulatives des effectifs ou des fréquences





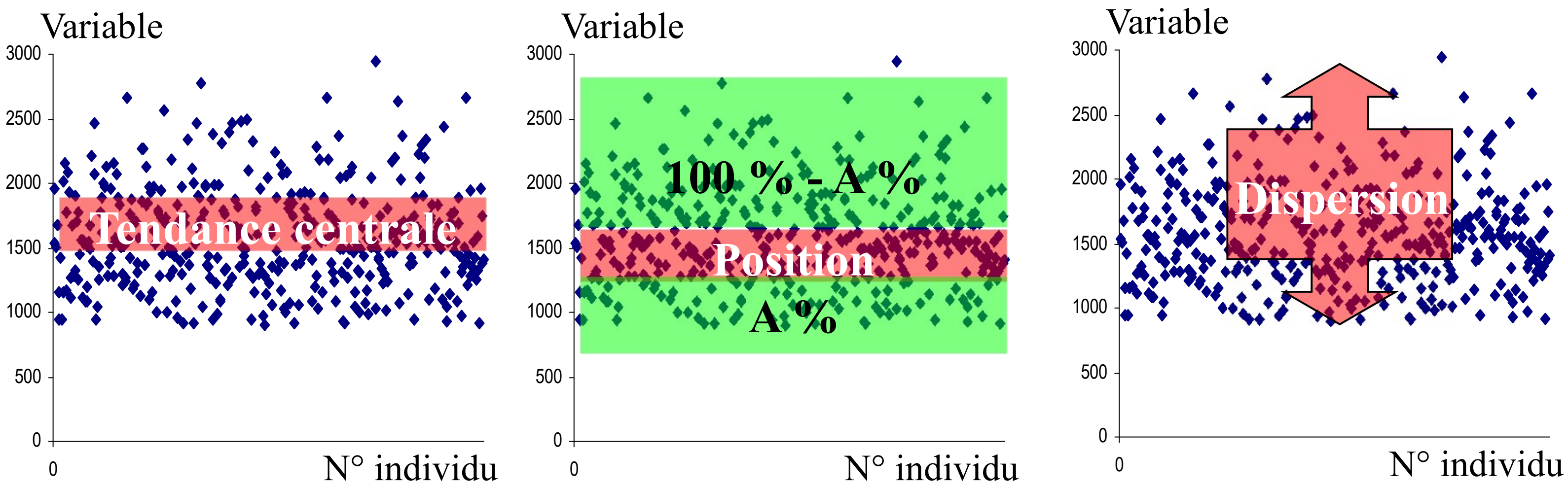
PARAMETRES STATISTIQUES

PARAMETRES STATISTIQUES

Les représentations graphiques ont permis une première synthèse visuelle de la distribution des observations

Un paramètre statistique permet de résumer par une seule quantité numérique une information contenue dans une distribution d'observations.

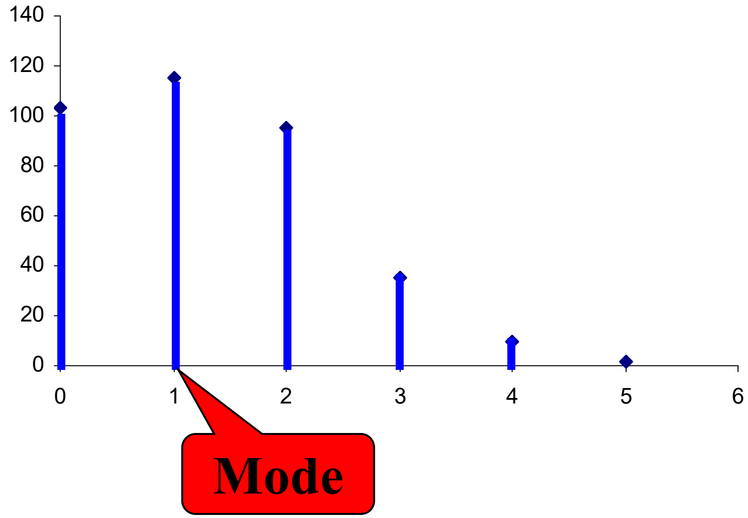
Les paramètres statistiques ne concernent que les variables **quantitatives**

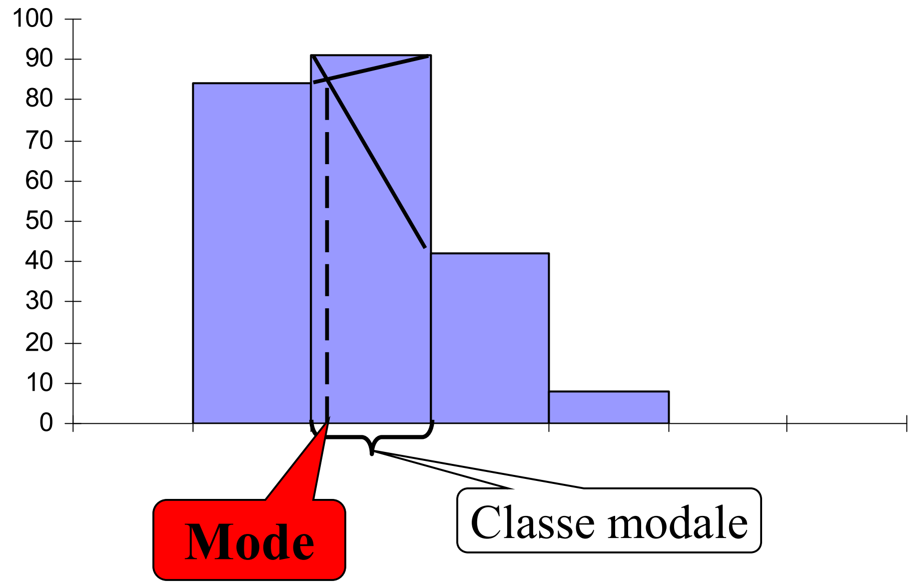


(1) PARAMETRES DE TENDANCE CENTRALE

LE MODE

Une distribution est **unimodale** si elle présente un maximum marqué, et pas d'autres maxima relatifs.
La lecture s'effectue sur le diagramme en bâtons ou l'histogramme.





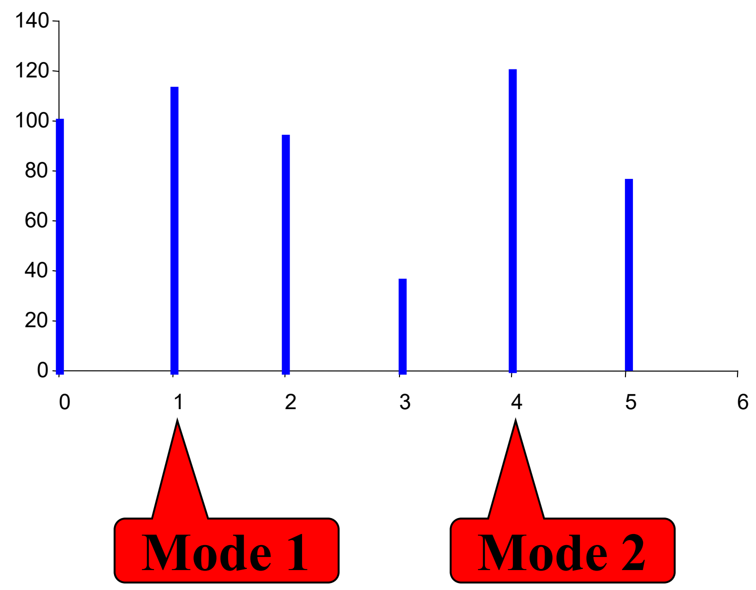
Le **mode** correspond à l'abscisse du maximum, c.à.d. la valeur la plus fréquente

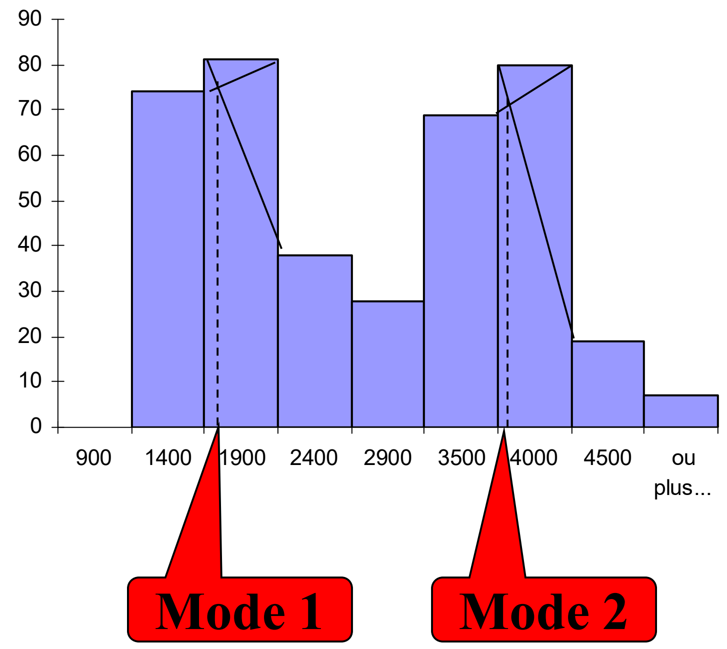
(2) PARAMETRES DE TENDANCE CENTRALE

LE MODE

Si la distribution présente 2 ou plus maxima relatifs, on dit qu'elle est **bimodale** ou **plurimodale**.

La population est composée de plusieurs sous-populations ayant des caractéristiques de tendance centrale différentes.





(3) PARAMETRES DE TENDANCE CENTRALE

LA MEDIANE

Les valeurs observées doivent être rangées par ordre croissant.

La **médiane** M est la valeur du milieu de la série d’observations, c.à.d. telle qu'il y ait autant d'observations "au-dessous" que "au-dessus".

Nombre impair d’observations

$$\underbrace{3 \quad 4 \quad 4 \quad 5}_{4\ \text{valeurs}} \quad \underset{\underset{\mathbf{M}}{\uparrow}}{6} \quad \underbrace{8 \quad 8 \quad 9 \quad 10}_{4\ \text{valeurs}}$$

Nombre pair d’observations

$$\underbrace{3 \quad 4 \quad 4 \quad 5}_{4\ \text{valeurs}}\!\!\!\!\underset{\uparrow}{\underbrace{\quad\;\;\;\;\;}}\!\!\!\!\underbrace{6 \quad 8 \quad 8 \quad 9}_{4\ \text{valeurs}}$$

Intervalle médian

M = milieu = 5,5

(4) PARAMETRES DE TENDANCE CENTRALE

LA MEDIANE à partir d'une distribution discrète

x_i	n_i	F_i	F(x)
			0
0	103	0,286	0,286
M 1	115	0,606	0,606
2	95	0,869	0,869
3	35	0,967	0,967
4	10	0,994	0,994
5	2	1	1

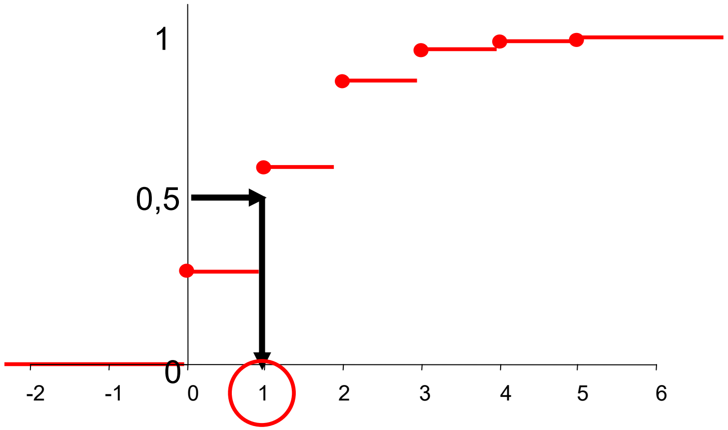
0,5

Intervalle médian

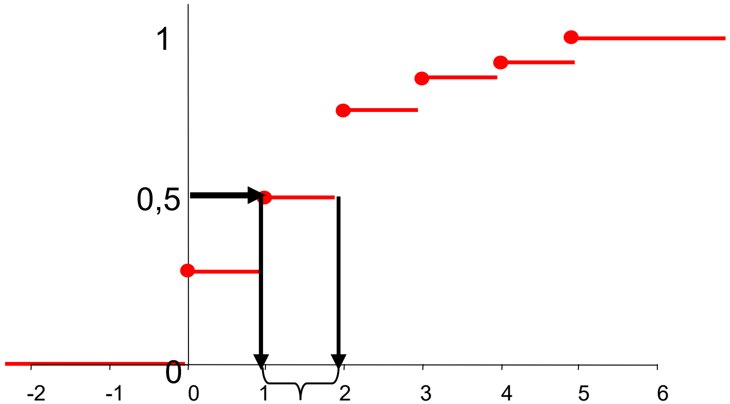
M = milieu = 1,5

x_i	n_i	F_i	F(x)
			0
0	103	0,286	0,286
1	77	0,500	0,500
2	95	0,764	0,764
3	35	0,861	0,861
4	10	0,889	0,889
5	40	1	1

0,5



M



Intervalle médian

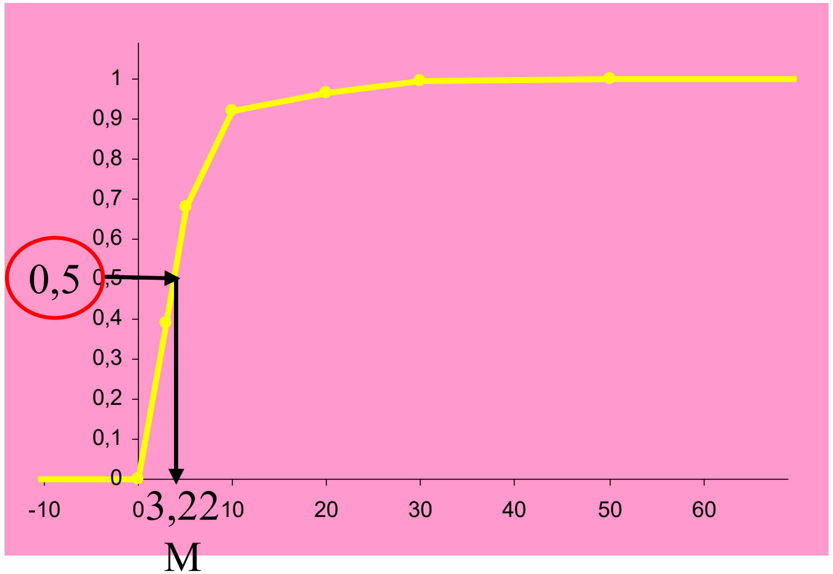
M = milieu = 1,5

(5) PARAMETRES DE TENDANCE CENTRALE

LA MEDIANE à partir d'une distribution continue

x	$[e_i - e_{i+1}[$	F_i	F(x)
0			0
	[0 - 3 [	0,391	
3			0,391
M			**0,5**
	[3 - 5 [	0,680	
5			0,680
	[5 - 10 [	0,920	
10			0,920
	[10 - 20 [	0,963	
20			0,963
	[20 - 30 [	0,993	
30			0,993
	[30 - 50 [	1	
50			1

$$\frac{M - 3}{5 - 3} = \frac{0{,}5-0{,}391}{0{,}680-0{,}391}$$



$$\text{D'où } M = 3 + \frac{0{,}5 - 0{,}391}{0{,}680 - 0{,}391}(5-3) \approx 3{,}22$$

(6) PARAMETRES DE TENDANCE CENTRALE

LA MOYENNE ARITHMETIQUE

La moyenne arithmétique est notée $\overline{x}$

Série brute $x_1, x_2, \ldots, x_n$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Série groupée

Valeurs de la variable	Effectifs	Fréquences
x_1	n_1	$f_1 = n_1/n$
…	…	…
x_i	n_i	$f_i = n_i/n$
…	…	…
x_k	n_k	$f_k = n_k/n$

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

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

(7) PARAMETRES DE TENDANCE CENTRALE

LA MOYENNE ARITHMETIQUE

Série classée

Classes	Effectifs	Fréquences	Centres de classe
$[e_1 - e_2[$	n_1	f_1	$x_1 = (e_1 + e_2)/2$
$[e_2 - e_3[$	n_2	f_2	$x_2 = (e_2 + e_3)/2$
….	….	….	….
$[e_k - e_{k+1}[$	n_k	f_k	$x_k = (e_k + e_{k+1})/2$

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

(8) PARAMETRES DE TENDANCE CENTRALE
LA MOYENNE ARITHMETIQUE

Comment faire la moyenne de plusieurs populations ?

Population P_1
Effectif n_1
Moyenne $\overline{x}_1$

Population P_2
Effectif n_2
Moyenne $\overline{x}_2$

Population $P = P_1 \cup P_2$
Effectif $n = n_1 + n_2$
Moyenne $\overline{\overline{x}}$?

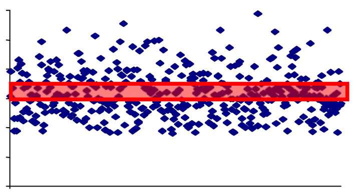
$$\overline{\overline{x}} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n} = \sum_{i=1}^{k} \frac{n_i\overline{x}_i}{n}$$

Moyenne globale = moyenne des moyennes

(9) PARAMETRES DE TENDANCE CENTRALE

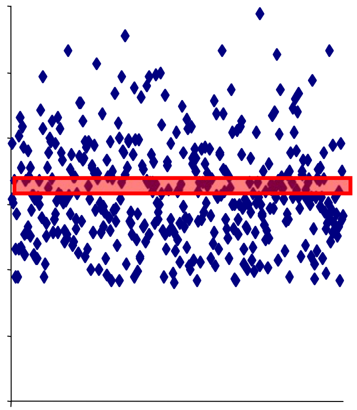
PROPRIETES GENERALES

x



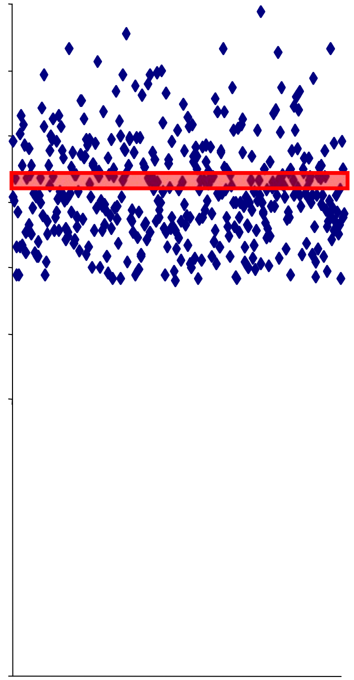
P (x) = moyenne, médiane, mode

y = a x



P (y) = a P (x)

z = a x + b



P (z) = a P (x) + b

(10) PARAMETRES DE TENDANCE CENTRALE
MOYENNES GEOMETRIQUE ET HARMONIQUE

Moyenne géométrique

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \ldots\ldots x_k^{n_k}}$$

Utilisée dans le cas de phénomènes multiplicatifs (taux de croissance moyen)

Moyenne harmonique

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

Utilisée dans le cas où l'on combine 2 variables sous forme de rapport (pièces/heure, km/litre,…)

(1) PARAMETRES DE POSITION

LES FRACTILES OU QUANTILES

On appelle fractiles ou quantiles d'ordre k les (k-1) valeurs qui divisent les observations en k parties d'effectifs égaux.

1 médiane M qui divise les observations en 2 parties égales

3 quartiles Q_1, Q_2, Q_3 qui divisent les observations en 4 parties égales

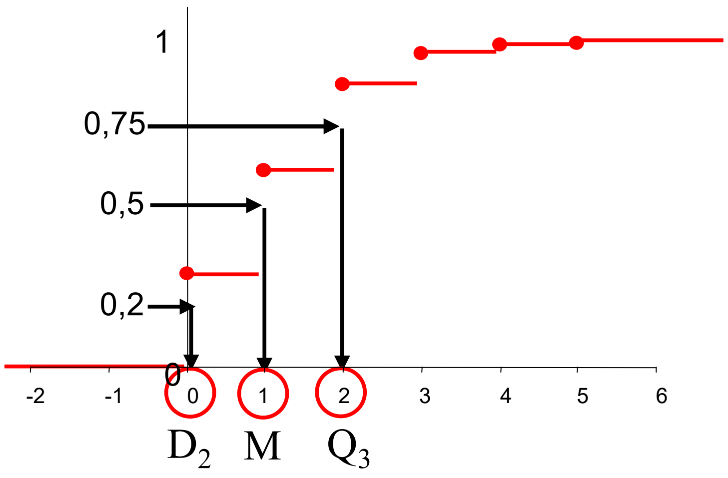
9 déciles D_1, D_2, …, D_9 qui divisent les observations en 10 parties égales

99 centiles C_1, C_2, …, C_{99} qui divisent les observations en 100 parties égales

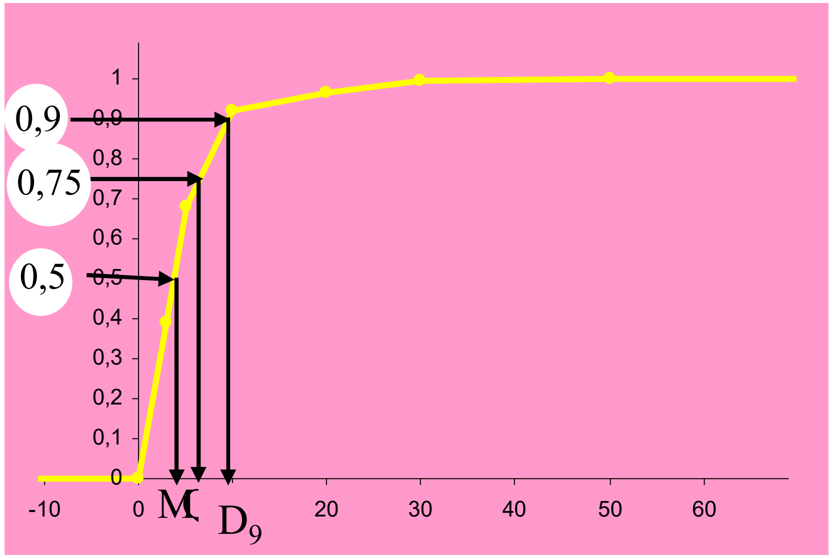
(2) PARAMETRES DE POSITION
LES FRACTILES OU QUANTILES

Quartiles, déciles, centiles s'obtiennent de la même façon que la médiane.

Variable discrète



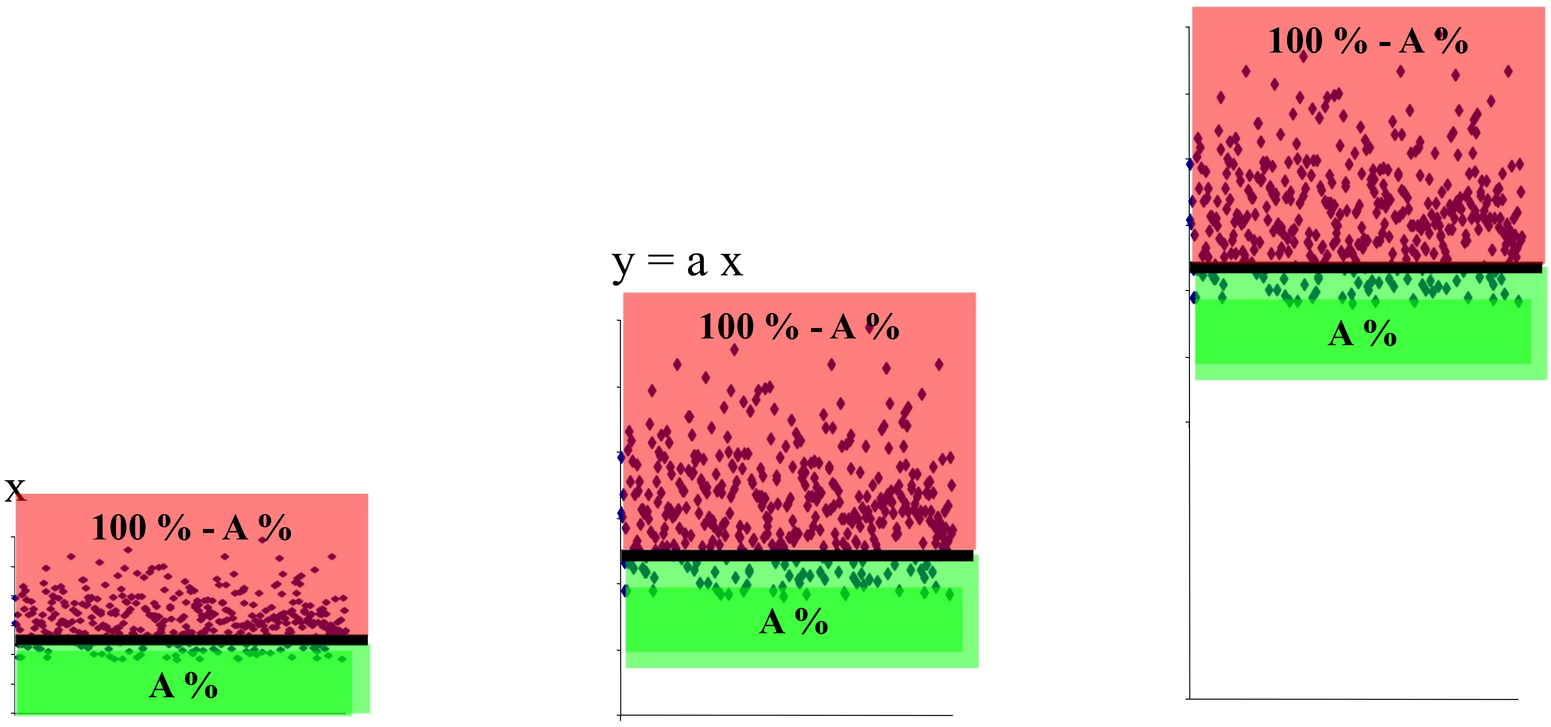
Variable continue



(3) PARAMETRES DE POSITION
PROPRIETES GENERALES

$z = a\,x + b$

$y = a\,x$



$Q(x)$ = quantile

$Q(y) = a\,Q(x)$

$Q(z) = a\,Q(x) + b$

(1) PARAMETRES DE DISPERSION

Etendue : $R = x_{max} - x_{min}$

Intervalle interquartile : $IQ = Q_3 - Q_1$

Variance :

Série brute :

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Série groupée ou classée :

$$V = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{k} f_i (x_i - \overline{x})^2$$

$$V = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 \quad = \text{Moyenne des carrés - Carré de la moyenne}$$

Ecart-type : $\sigma = \sqrt{V}$

(2) PARAMETRES DE DISPERSION

Comment faire la variance de plusieurs populations ?

Population P_1
Effectif n_1
Moyenne $\overline{x}_1$
Variance V_1

Population P_2
Effectif n_2
Moyenne $\overline{x}_2$
Variance V_2

Population $P = P_1 \cup P_2$
Effectif $n = n_1 + n_2$
Moyenne $\overline{\overline{x}}$
Variance V ?

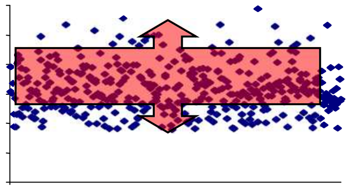
$$V = \frac{1}{n}\sum_{i=1}^{k} n_i V_i + \frac{1}{n}\sum_{i=1}^{k} n_i \left(\overline{x}_i - \overline{\overline{x}}\right)^2$$

Variance globale = Moyenne des variances + Variance des moyennes

(3) PARAMETRES DE DISPERSION

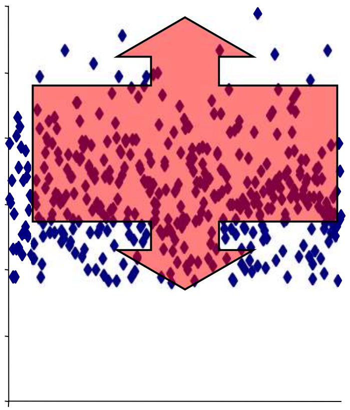
PROPRIETES GENERALES

x



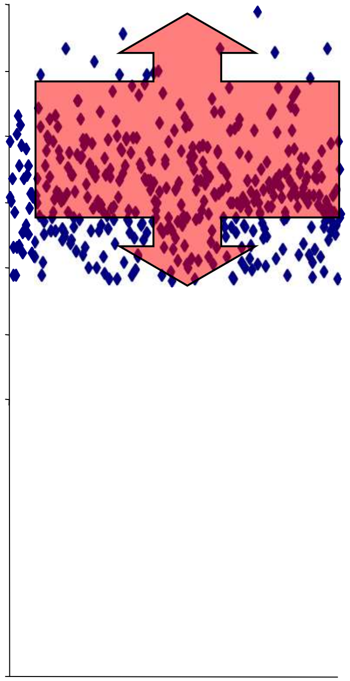
P (x) = étendue, écart-type,
intervalle interquartile

y = a x



P (y) = a P (x)

z = a x + b



P (z) = a P (x)

PROPRIETES IMPORTANTES
DE LA MOYENNE ET DE LA VARIANCE

Comment se comportent la moyenne et la variance lorsqu'on fait subir un changement de variable aux observations?

$$x_i \longrightarrow y_i = a\,x_i + b$$

$$\overline{y} = a\,\overline{x} + b \qquad V(y) = a^2\,V(x) \qquad \sigma(y) = |a|\,\sigma(x)$$

Comment se comportent la moyenne et la variance de la somme de deux séries d'observations?

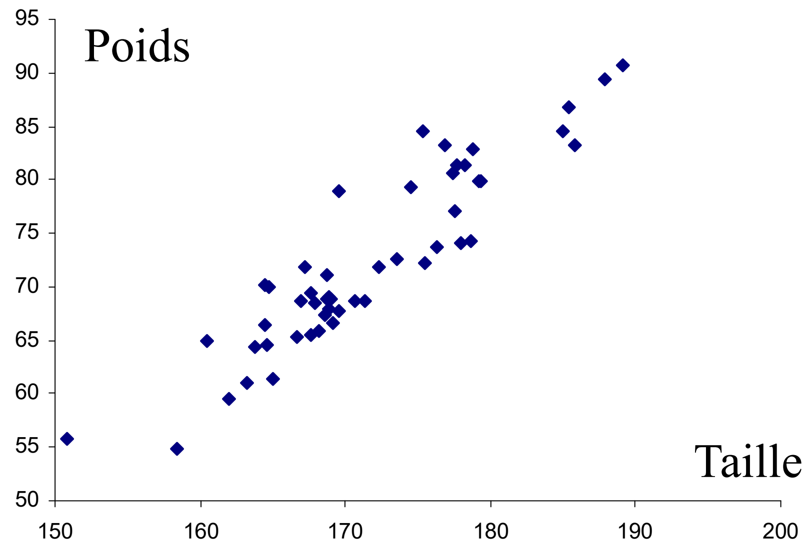
$$\begin{matrix} x_i \\ y_i \end{matrix} \longrightarrow z_i = x_i + y_i$$

$$\overline{z} = \overline{x} + \overline{y} \qquad V(z) \neq V(x) + V(y)$$

ETUDE DE 2 VARIABLES QUANTITATIVES

(1) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

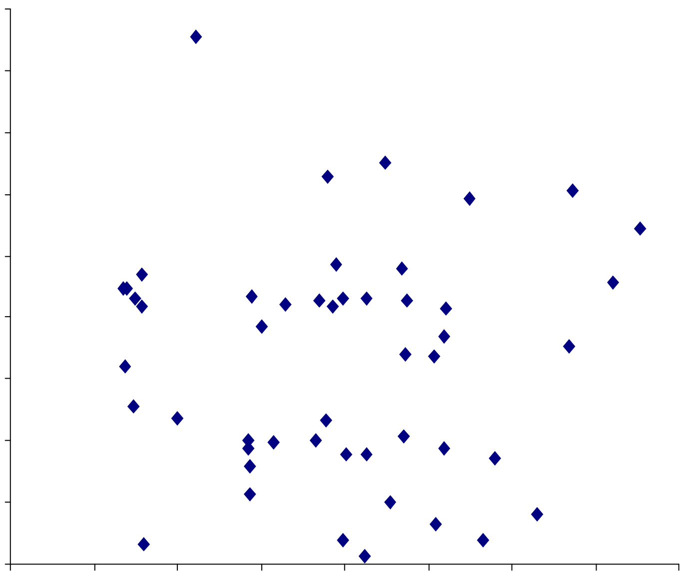
Nom	Taille x_i (cm)	Poids y_i (kg)
Pierre	175	73
Arantxa	168	56
…..	…..	…..
Martin	185	87



La connaissance de la taille x apporte une certaine information sur le poids y

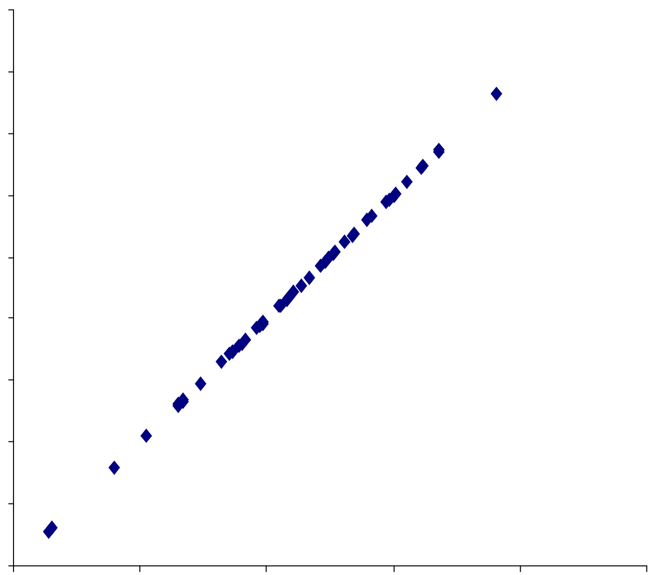
Il existe une relation de dépendance entre x et y

(2) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES



La connaissance de x n'apporte aucune certaine information sur y

x et y sont indépendantes



La connaissance de x permet de connaître exactement la valeur de y

Il existe une relation fonctionnelle entre x et y

(3) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

Covariance : $$\mathrm{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Propriétés :

$\mathrm{Cov}(x,y) > 0 \iff$ x et y varient dans le même sens

$\mathrm{Cov}(x,y) < 0 \iff$ x et y varient en sens contraire

$$\mathrm{Cov}(x,y) = \mathrm{Cov}(y,x)$$

$$\mathrm{Cov}(x,x) = V(x)$$

$$\mathrm{Cov}(a\,x + b\,y\,,\,z) = a\,\mathrm{Cov}(x,z) + b\,\mathrm{Cov}(y,z)$$

(4) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

Corrélation linéaire: $\rho = \dfrac{\text{cov}(x,y)}{\sigma(x)\,\sigma(y)}$

Propriétés :

$-1 \leq \rho \leq 1$

$$y = a\,x + b \;\Leftrightarrow \begin{cases} \rho = 1 & \text{si} \quad a > 0 \\ \rho = -1 & \text{si} \quad a < 0 \end{cases}$$

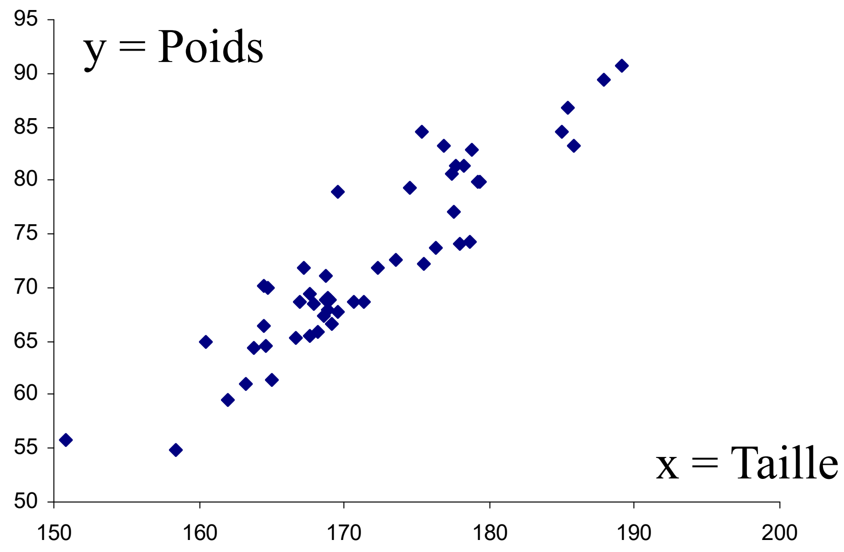
$|\rho| = 1 \;\Leftrightarrow\;$ Il existe une relation fonctionnelle entre x et y

$\rho = 0 \;\Leftrightarrow\;$ x et y sont indépendantes

$0 < |\rho| < 1 \;\Leftrightarrow\;$ Il existe une dépendance linéaire d'autant plus forte que $|\rho|$ est grand

Ne pas confondre causalité et corrélation

(1) AJUSTEMENT LINEAIRE

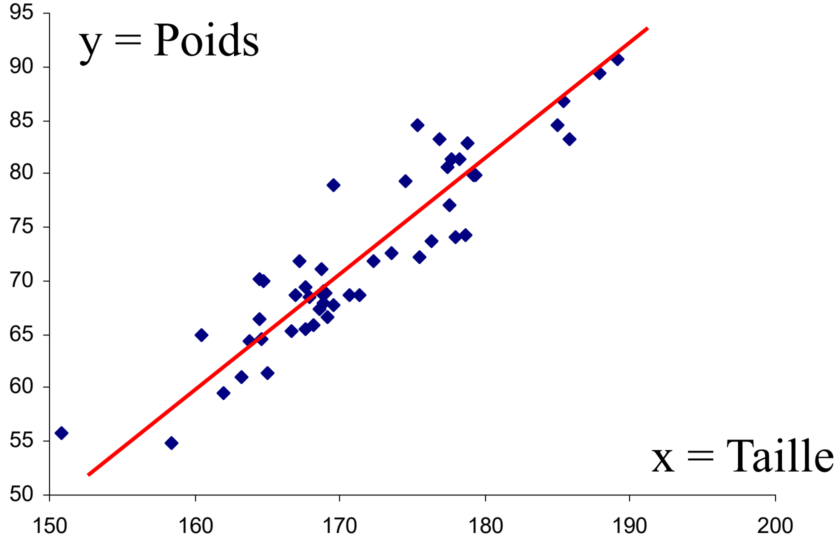


Est-il possible de trouver une fonction numérique f telle que y = f (x) ?

Si une telle fonction existe, on dit que f est un modèle du phénomène étudié.

x est la variable explicative.
y est la variable expliquée.

(2) AJUSTEMENT LINEAIRE



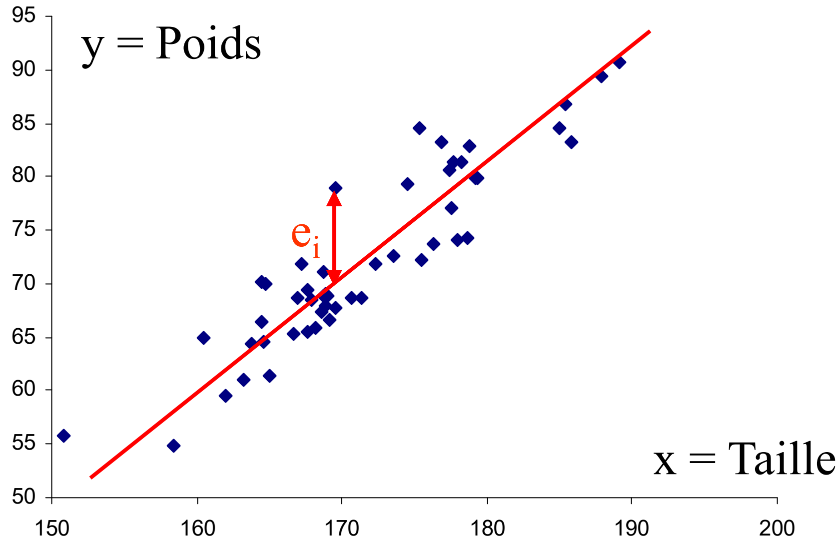
On désire trouver la droite qui passe « au mieux » à l'intérieur du nuage de points

(3) AJUSTEMENT LINEAIRE

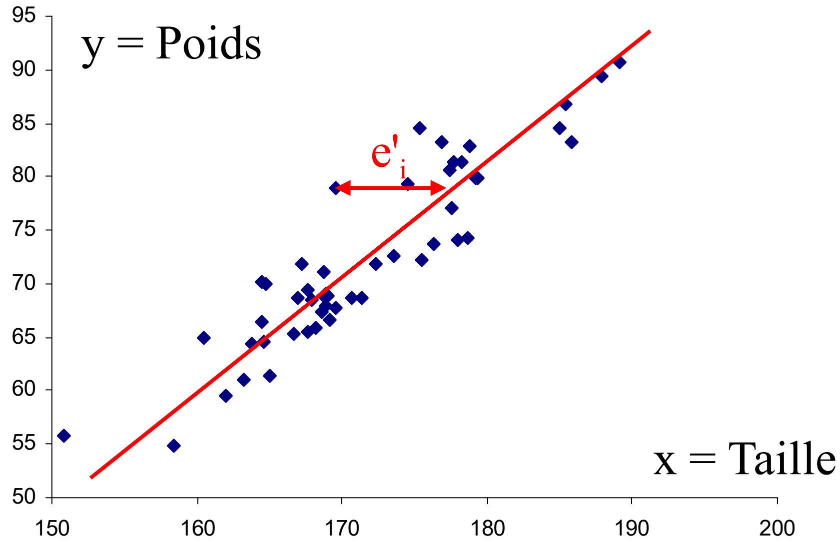
« au mieux »

Minimiser $S = \sum_{i=1}^{n} e_i^2$

Minimiser $S' = \sum_{i=1}^{n} e'^2_i$



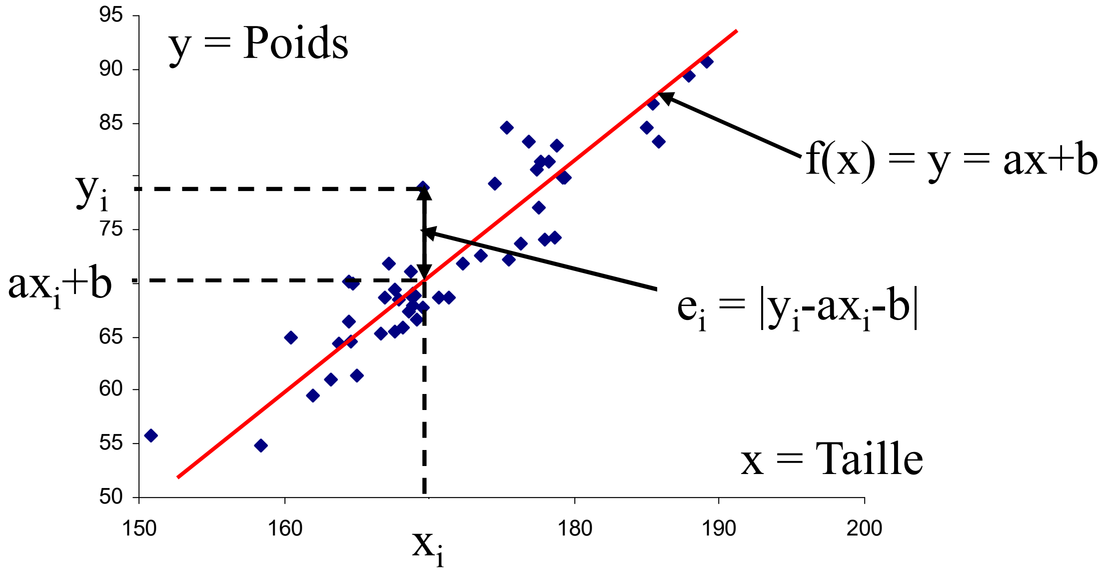
Droite de régression de y en x



Droite de régression de x en y

(4) AJUSTEMENT LINEAIRE
REGRESSION LINEAIRE DE Y EN X

Droite de régression linéaire de y en x
$y = f(x) = ax + b$



La droite de régression linéaire de y en x, notée $D_{y/x}$, minimise $S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{Cov(x,y)}{V(x)}$$

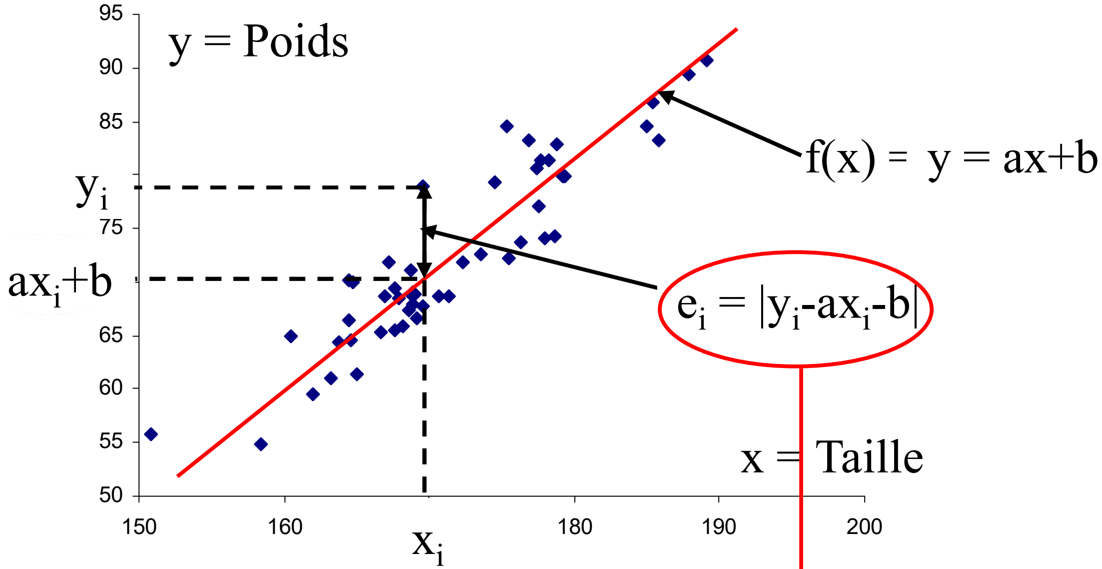
$$b = \bar{y} - a\bar{x}$$

$D_{y/x}$ passe par le point moyen $(\bar{x}, \bar{y})$

(5) AJUSTEMENT LINEAIRE

REGRESSION LINEAIRE DE Y EN X

Droite de régression linéaire de y en x
$y = f(x) = ax + b$



$y = a\,x + b$ définit un modèle affine

$\hat{y}_i = a\,x_i + b$ = valeur de y_i prévue par le modèle

$r_i = y_i - \hat{y}_i$ = résidu de la ième observation

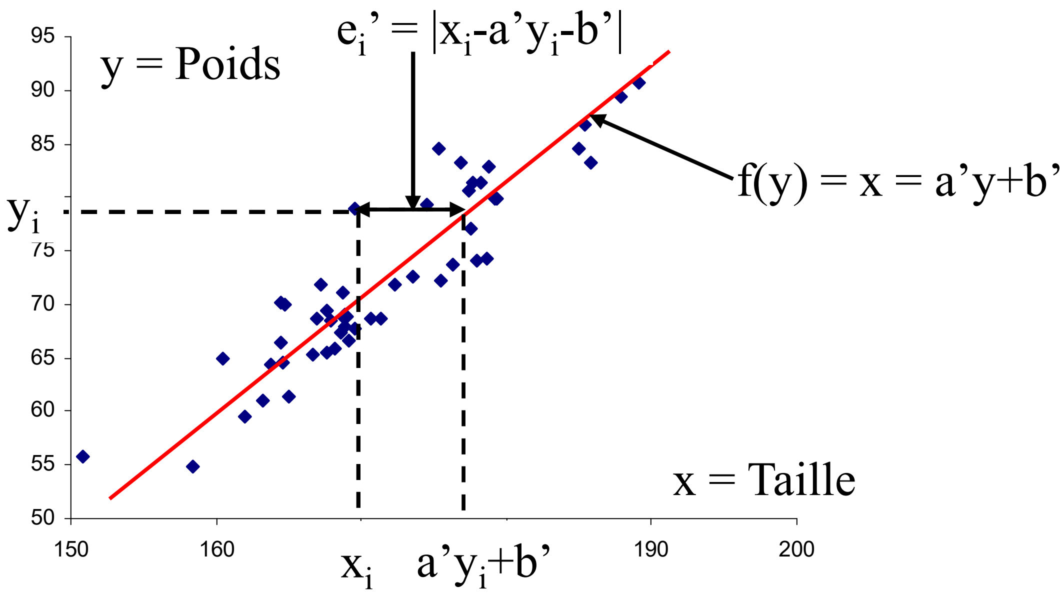
$e_i = |r_i| = |y_i - a\,x_i - b|$ = erreur due au modèle

ETUDE DE 2 VARIABLES QUANTITATIVES

(6) AJUSTEMENT LINEAIRE

REGRESSION LINEAIRE DE X EN Y

Droite de régression linéaire de x en y
$x = f(y) = a'y + b'$



La droite de régression linéaire de x en y, notée $D_{x/y}$, minimise $S' = \sum_{i=1}^{n} e_i'^2 = \sum_{i=1}^{n} (x_i - a'y_i - b')^2$

$$a' = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{Cov(x,y)}{V(y)}$$

$$b' = \bar{x} - a'\,\bar{y}$$

$D_{x/y}$ passe par le point moyen $(\bar{x}, \bar{y})$

LIENS ENTRE CORRELATION ET DROITES DE REGRESSION

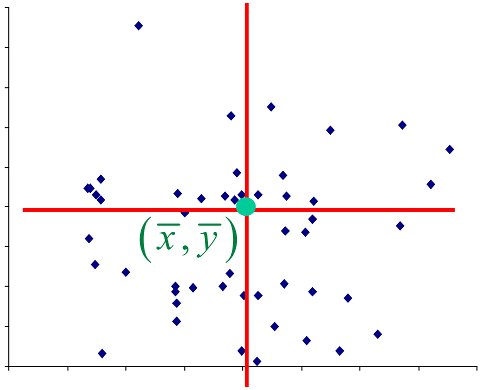
$D_{y/x}$: $y = ax + b$ $\quad a = \dfrac{Cov(x,y)}{V(x)}$ $\quad b = \overline{y} - a\overline{x}$

$D_{x/y}$: $x = a'y + b'$ $\quad a' = \dfrac{Cov(x,y)}{V(y)}$ $\quad b' = \overline{x} - a'\,\overline{y}$

$\Leftrightarrow y = \dfrac{1}{a'}x - \dfrac{b'}{a'}$

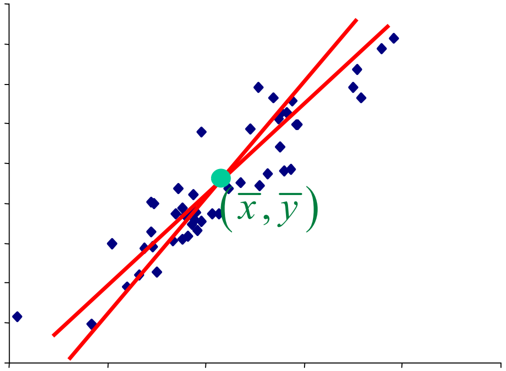
$\rho^2 = a\,a'$

$\rho = a\dfrac{\sigma(x)}{\sigma(y)} = a'\dfrac{\sigma(y)}{\sigma(x)}$



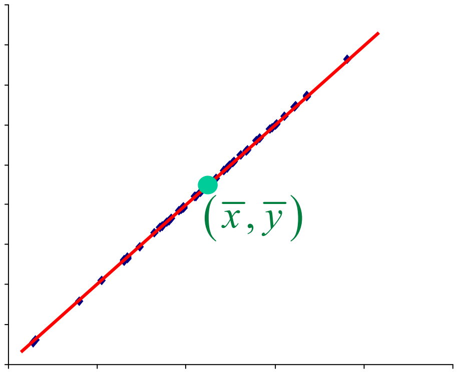
$\rho^2 = a\,a' = 0$

Indépendance linéaire



$0 < \rho^2 = a\,a' < 1$

Le degré de dépendance linéaire se mesure à la proximité des droites de régression



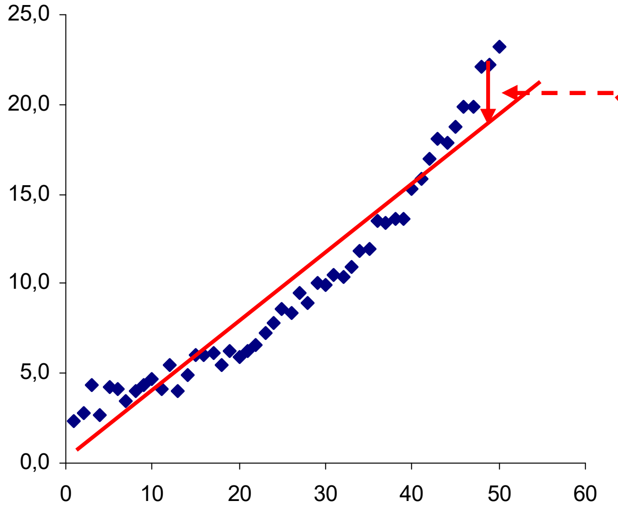
$\rho^2 = a\,a' = 1$

Liaison fonctionnelle linéaire

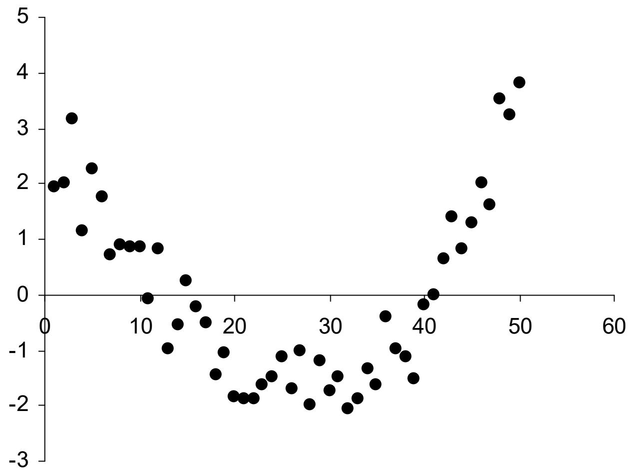
ETUDE DE 2 VARIABLES QUANTITATIVES

(1) AJUSTEMENT A UNE FONCTION EXPONENTIELLE

x_i	y_i
2,8	0,8
4,3	1,2
2,7	1,5
4,2	1,9
4,1	2,3
….	….
4,0	3,1



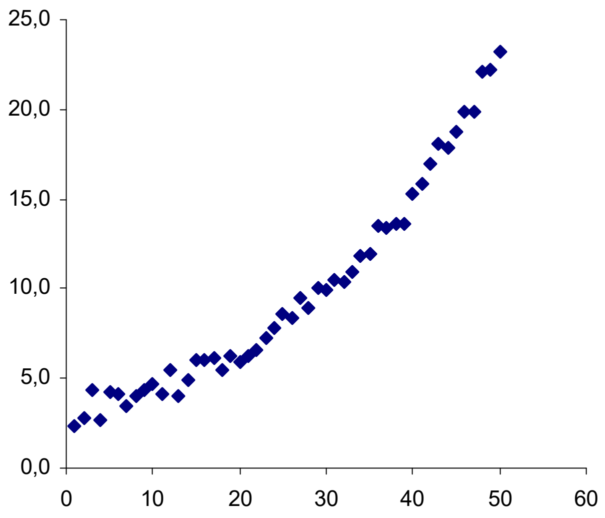
droite de régression linéaire
de y en x



Analyse des résidus

Les résidus devraient se répartir au hasard autour de l'axe des abscisses:
le modèle affine ne convient pas

(2) AJUSTEMENT A UNE FONCTION EXPONENTIELLE



Modèle exponentiel

$y = e^x$ exponentielle de base e

$y = a^x$ exponentielle de base a

$y = b\,a^x$ Forme exponentielle générale

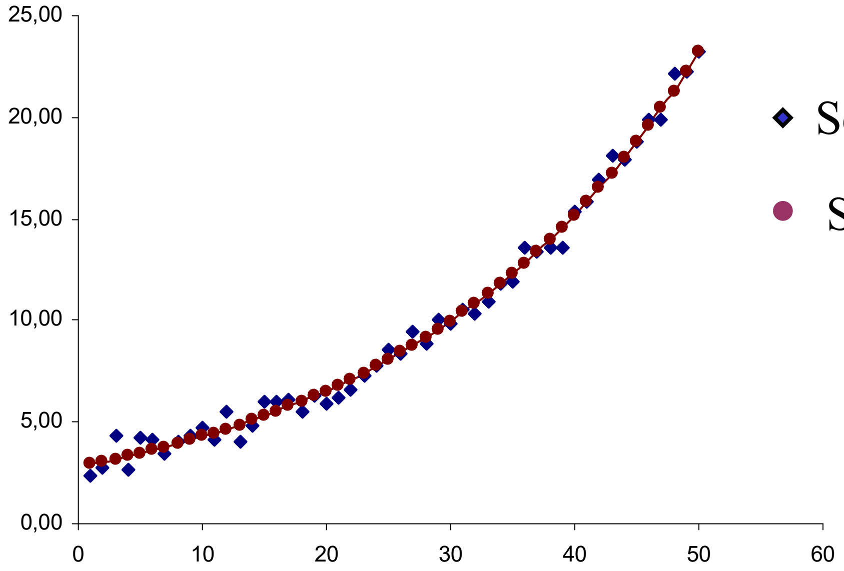
Changement de variable

$\ln y = \ln b + x \ln a$

$Y = A\,X + B$ avec $Y = \ln y$, $X = x$, $A = \ln a$, $B = \ln b$

L'ajustement affine de Y en fonction de X donne A et B,
d 'où $a = e^A$, $b = e^B$, et le modèle $y = b\,a^x$

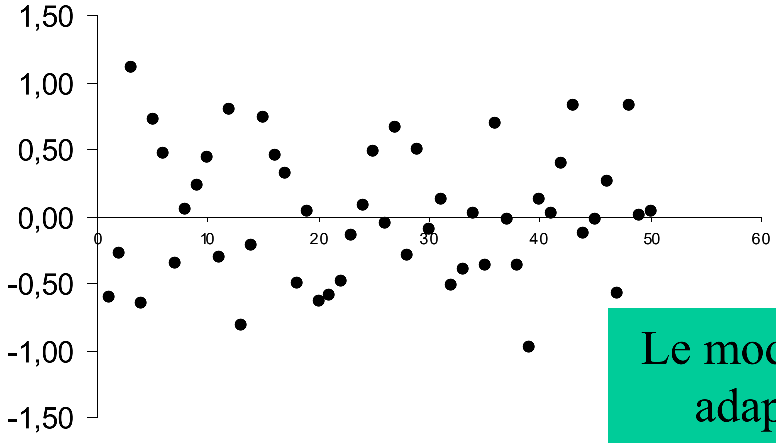
(3) AJUSTEMENT A UNE FONCTION EXPONENTIELLE



◆ Série initiale (x_i, y_i)

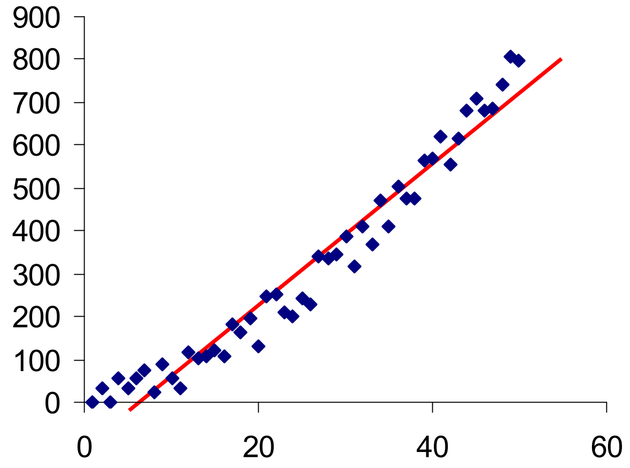
● Série prévue par le modèle $(x_i, \hat{y}_i)$

Analyse des résidus

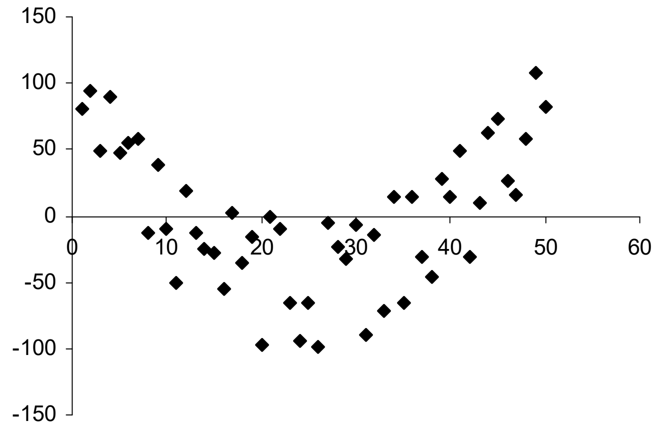


Le modèle exponentiel est mieux adapté que le modèle affine

(1) AJUSTEMENT A UNE FONCTION PUISSANCE



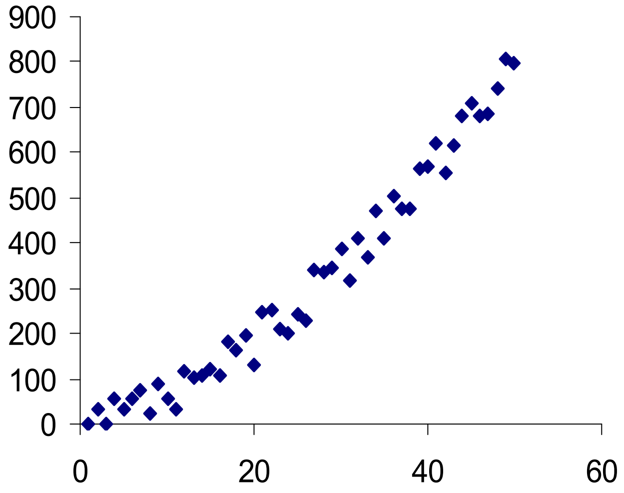
Droite de régression linéaire de y en x



Analyse des résidus

Le modèle affine ne convient pas

(2) AJUSTEMENT A UNE FONCTION PUISSANCE



Modèle puissance $y = b\,x^a$

Changement de variable

$$\ln y = \ln b + a \ln x$$

$$Y = A\,X + B$$

avec

$Y = \ln y$

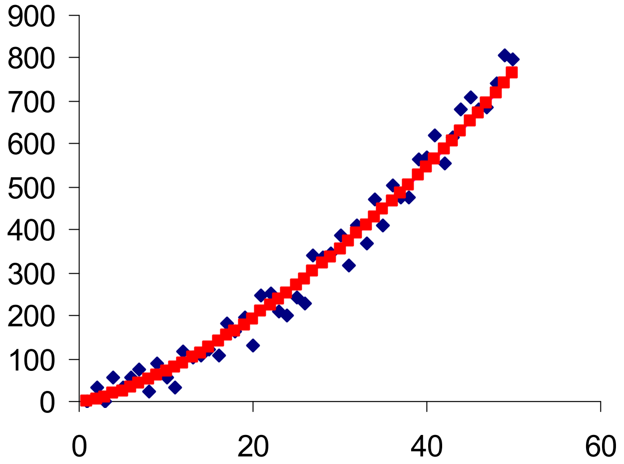
$X = \ln x$

$A = a$

$B = \ln b$

L'ajustement affine de Y en fonction de X donne A et B,
d 'où $a = A$, $b = e^B$, et le modèle $y = b\,x^a$

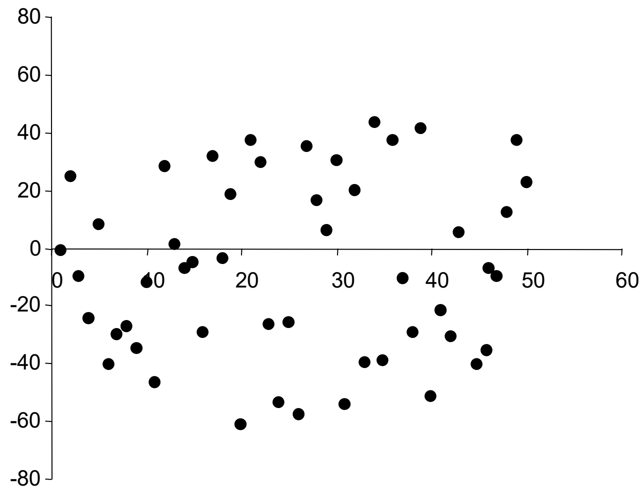
(3) AJUSTEMENT A UNE FONCTION PUISSANCE



◆ Série initiale (x_i,y_i)

● Série prévue par le modèle $\left(x_i,\hat{y}_i\right)$

Analyse des résidus



Le modèle puissance est mieux adapté que le modèle affine

QUALITE D'UN AJUSTEMENT

On montre que $$\sum(y_i-\bar{y})^2 = \sum(\hat{y}_i-\bar{y})^2 + \sum(y_i-\hat{y}_i)^2$$

$$SCT = SCM + SCR \quad \Leftrightarrow \quad 1 = \frac{SCM}{SCT} + \frac{SCR}{SCT}$$

Somme des carrés des écarts à la moyenne = Somme des carrés des écarts du modèle + Somme des carrés des résidus

L'ajustement est d'autant meilleur que SCR est proche de 0, c.à.d. que SCR/SCT est proche de 0 ou SCM/SCT est proche de 1.

$$R = \frac{SCM}{SCT} = \text{Coefficient de détermination} = \rho^2 = (\text{coef. de corrélation})^2$$

$$= \text{proportion de la variation totale due à l'ajustement}$$

$$0 \leq R \leq 1$$

LES INDICES

INDICES ELEMENTAIRES

Un indice est le rapport d'une variable mesurée à deux instants différents.
Un indice est représentatif d'une évolution

y_1 = valeur de la variable y à la date t_1

y_0 = valeur de la variable y à la date t_0

$$i_{1/0} = \frac{y_1}{y_0}$$

Indice élémentaire de la variable y à la date t_1 par rapport à la date de référence t_0

$$I_{1/0} = i_{1/0} \times 100$$

Indice élémentaire de la variable y à la date t_1 par rapport à la date de référence t_0, base 100.

<u>Propriétés</u>

$i_{n/n} = 1$	Identité
$i_{2/1} \times i_{1/2} = 1$	Réversibilité
$i_{3/1} = i_{3/2} \times i_{2/1}$	Circularité

INDICES ET TAUX DE VARIATION

$$r_{1/0} = \frac{y_1 - y_0}{y_0}$$

Taux de variation ou taux de croissance de la variable y entre la date t_0 et la date t_1

$$r_{1/0} = \frac{y_1}{y_0} - 1 = i_{1/0} - 1$$

$r = i - 1 \qquad i = 1 + r$

$$y_1 = (1 + r_{1/0})y_0 \quad \Leftrightarrow \quad y_1 = i_{1/0}\, y_0$$

$i = 1 + r =$ coefficient multiplicateur

$r = 0$	$\Leftrightarrow$	$i = 1$	Pas d'évolution
$r > 0$	$\Leftrightarrow$	$i > 1$	Croissance
$-100\% = -1 < r < 0$	$\Leftrightarrow$	$0 < i < 1$	Décroissance

INDICES ET TAUX DE VARIATION MOYENS

$y_0, y_1, \ldots, y_n$ les valeurs prises par une variable aux dates $t_0, t_1, \ldots, t_n$

$r_1, r_2, \ldots, r_n$ les taux de croissance sur chacune des périodes

$$y_n = (1+r_n) \times y_{n-1} = (1+r_n) \times (1+r_{n-1}) \times y_{n-2} = (1+r_n) \times \ldots\ldots \times (1+r_2) \times (1+r_1) \times y_0$$

r_G le taux de croissance entre t_0 et t_n

$$y_n = (1+r_G) \times y_0$$

r le taux de croissance moyen

$$y_n = (1+r) \times y_{n-1} = (1+r)^2 \times y_{n-2} = \ldots = (1+r)^n \times y_0$$

$$(1+r_G) = (1+r)^n = (1+r_n) \times \ldots\ldots \times (1+r_2) \times (1+r_1)$$

$r_1, r_2, \ldots, r_k$ indices élémentaires sur des périodes de $n_1, n_2, \ldots, n_k$ unités (jour, mois, année…)

$$(1+r_G) = (1+r)^n = (1+r_1)^{n_1} \times (1+r_2)^{n_2} \times \ldots\ldots \times (1+r_k)^{n_k}$$

INDICES USUELS

Indice élémentaire des prix

$$i(P)_{1/0} = \frac{P_1}{P_0}$$

Indice élémentaire des quantités
(ou des volumes)

$$i(Q)_{1/0} = \frac{Q_1}{Q_0}$$

Indice élémentaire de valeur
(ou de dépense)

$$i(V)_{1/0} = \frac{V_1}{V_0} = \frac{P_1 Q_1}{P_0 Q_0} = i(P)_{1/0}\, i(Q)_{1/0}$$

INDICES SYNTHETIQUES

Un indice synthétique mesure l'évolution simultanée de plusieurs produits

Un indice synthétique est une moyenne pondérée des indices élémentaires des différents produits

Coefficient de pondération (ou budgétaire) du produit j à la date t_n

$$\alpha_{j,n} = \frac{V_{j,n}}{\sum_{j=1}^{n} V_{j,n}} = \frac{P_{j,n} Q_{j,n}}{\sum_{j=1}^{n} P_{j,n} Q_{j,n}}$$

Remarque : $\sum_{j=1}^{n} \alpha_{j,n} = 1$

(1) INDICES SYNTHETIQUES DE LASPEYRES

Indice de Laspeyres des prix

$L\,(P)_{1/0} =$ Moyenne arithmétique des indices élémentaires des prix, base 100, pondérés par des coefficients de pondération relatifs à la date de référence t_0

$$L\,(P)_{1/0} = \sum_{j=1}^{n} \alpha_{j,0}\; I\left(P_j\right)_{1/0}$$

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,0}}{\sum_{j=1}^{n} P_{j,0} Q_{j,0}} \times 100$$

1 seul indice sur 4 doit être modifié

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,0}}{\sum_{j=1}^{n} P_{j,0} Q_{j,0}}$$

$$= \frac{\text{Dépense de la date courante avec les quantités de référence}}{\text{Dépense de la date de référence}} \times 100$$

(2) INDICES SYNTHETIQUES DE LASPEYRES

Indice de Laspeyres des quantités

$L(Q)_{1/0} =$ Moyenne arithmétique des indices élémentaires des quantités, base 100, pondérés par des coefficients de pondération relatifs à la date de référence t_0

$$L(Q)_{1/0} = \sum_{j=1}^{n} \alpha_{j,0}\ I(Q_j)_{1/0}$$

$$= \frac{\sum_{j=1}^{n} P_{j,0} Q_{j,1}}{\sum_{j=1}^{n} P_{j,0} Q_{j,0}} \times 100$$

1 seul indice sur 4 doit être modifié

$$= \frac{\sum_{j=1}^{n} P_{j,0} Q_{j,1}}{\sum_{j=1}^{n} P_{j,0} Q_{j,0}}$$

$$= \frac{\text{Dépense de la date courante avec les prix de référence}}{\text{Dépense de la date de référence}} \times 100$$

(1) INDICES SYNTHETIQUES DE PAASCHE

Indice de Paasche des prix

$P(P)_{1/0} =$ Moyenne harmonique des indices élémentaires des prix, base 100, pondérés par des coefficients de pondération relatifs à la date courante t_1

$$P(P)_{1/0} = \frac{1}{\sum_{j=1}^{n} \frac{\alpha_{j,1}}{I(P_j)_{1/0}}}$$

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,1}}{\sum_{j=1}^{n} P_{j,0} Q_{j,1}} \times 100$$

1 seul indice sur 4 doit être modifié

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,1}}{\sum_{j=1}^{n} P_{j,0} Q_{j,1}}$$

$$= \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence avec les quantités courantes}} \times 100$$

(2) INDICES SYNTHETIQUES DE PAASCHE

Indice de Paasche des quantités

$P\,(Q)_{1/0}$ = Moyenne harmonique des indices élémentaires des quantités, base 100, pondérés par des coefficients de pondération relatifs à la date courante t_1

$$P\,(Q)_{1/0} = \frac{1}{\sum_{j=1}^{n} \frac{\alpha_{j,1}}{I(Q_j)_{1/0}}}$$

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,1}}{\sum_{j=1}^{n} P_{j,1} Q_{j,0}} \times 100$$

1 seul indice sur 4 doit être modifié

$$= \frac{\sum_{j=1}^{n} P_{j,1} Q_{j,1}}{\sum_{j=1}^{n} P_{j,1} Q_{j,0}}$$

$$= \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence avec les prix courants}} \times 100$$

SERIES CHRONOLOGIQUES

SERIES CHRONOLOGIQUES

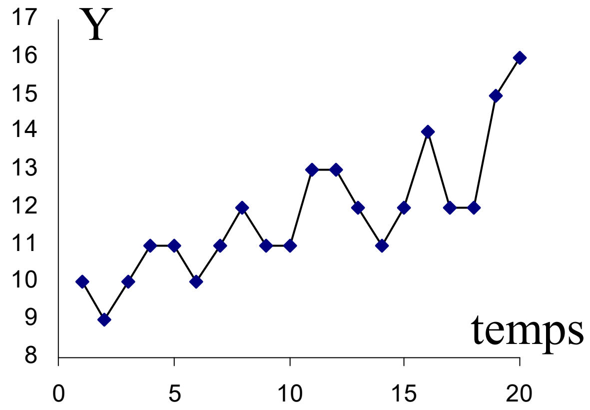
LES DONNEES

Date	Y
T1 2001	10
T2 2001	9
T3 2001	10
T4 2001	11
T1 2002	11
T2 2002	10
T3 2002	11
T4 2002	12
T1 2003	11
T2 2003	11
T3 2003	13
T4 2003	13
T1 2004	12
T2 2004	11
T3 2004	12
T4 2004	14
T1 2005	12
T2 2005	12
T3 2005	15
T4 2005	16

Y = prix d'un bien en fonction du temps

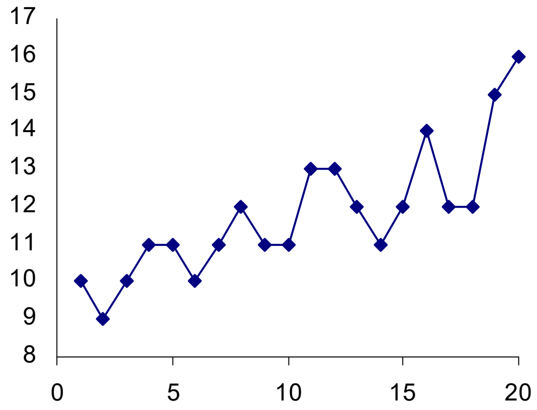
	2001	2002	2003	2004	2005
1er trimestre	10	11	11	12	12
2^{e} trimestre	9	10	11	11	12
3^{e} trimestre	10	11	13	12	15
4^{e} trimestre	11	12	13	14	16

Y = série initiale

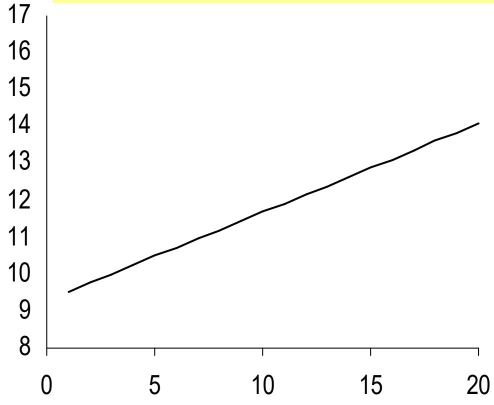


LES COMPOSANTES

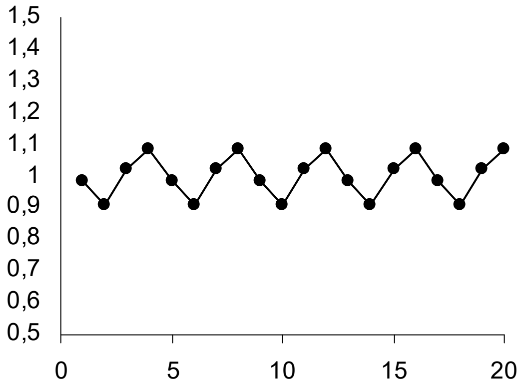
Y = série initiale



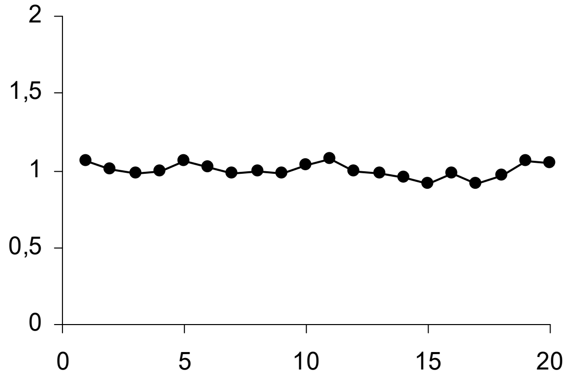
Tendance ou Trend
T



Composante Saisonnière
S

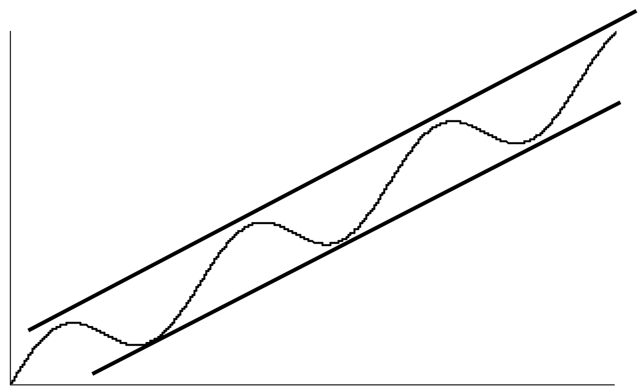


Composante Aléatoire
A



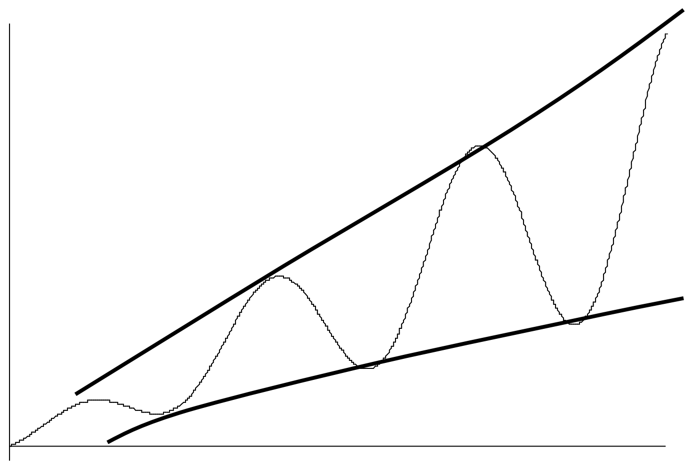
MODELES DE DECOMPOSITION

Modèle additif



$Y = T + S + A$

Modèle multiplicatif



$Y = T \cdot S \cdot A$

(1) DETERMINATION DE LA TENDANCE
REGRESSION LINEAIRE

Il s'agit de faire un lissage du nuage des points par une fonction connue.

Lorsque le nuage est linéaire on utilise la droite de régression de y en fonction du temps

T = tendance

Avantages:

Expression analytique

Inconvénients:

Un nuage ne se présente pas toujours sous une forme analytique simple

Le calcul de la tendance peut être affecté par des valeurs extrêmes ou par les valeurs de début et de fin de série.

(2) DETERMINATION DE LA TENDANCE
MOYENNES MOBILES

Moyennes mobiles
d'ordre impair

t	Y
1	y_1
2	y_2
3	y_3
4	y_4
…..	…..
n	y_n

t	mm(3)
2	$(y_1+y$ [illegible] $y_3)/3$
3	[illegible] $)/3$
	-

Moy. Mobiles
d'ordre 3

Moyennes mobiles
d'ordre pair.
On utilise une observation supplémentaire

t	Y
1	y_1
2	y_2
3	y_3
4	y_4
…..	…..
n	y_n

t	mm(2)
	-
2	$(y_1/2+y_2+y_3/2)/2$
3	$(y_2/2+y_3+y_4/2)/2$
	-

Moy. Mobiles
d'ordre 2

(3) DETERMINATION DE LA TENDANCE
MOYENNES MOBILES

Choix de l'ordre des moyennes mobiles : égal au nombre de saisons

Avantages du lissage par moyennes mobiles :

Permet de se faire une idée de la tendance lorsque le nuage ne présente pas une tendance algébrique claire

Inconvénients:

La tendance est estimée sur une partie de la période étudiée et non sur la totalité

Ne donne pas une expression analytique de la tendance en fonction du temps

Approximation pas très bonne lorsqu'il y a de fortes courbures

Sensible aux valeurs extrêmes

DETERMINATION DES COMPOSANTES SAISONNIERES

Modèle multiplicatif $Y = T.S.A$	Modèle additif $Y = T+S+A$
Rapports $Y/T = S.A$	Différences $Y-T = S+A$

Coefficients saisonniers bruts S'_j

S'_j = Moyenne des rapports de la saison j	S'_j = Moyenne des différences de la saison j

Coefficients saisonniers S_j

$S_j = S'_j / \overline{S}'$	$S_j = S'_j - \overline{S}'$

Rque: cette transformation permet de respecter le principe de conservation des aires

$\overline{S} = 1$	$\overline{S} = 0$

DETERMINATION DE LA COMPOSANTE ALEATOIRE

Modèle multiplicatif $Y = T.S.A$	Modèle additif $Y = T+S+A$
$A = \dfrac{Y}{T.S}$	$A = Y - T - S$

La composante aléatoire, ou résidu, permet d'analyser la qualité du modèle de décomposition

DESAISONNALISATION

Y_{CVS} = série désaisonnalisée ou Corrigée des Variations Saisonnières, exprime ce qu'aurait été l'évolution du phénomène sans effet saisonnier.

Modèle multiplicatif $Y = T.S.A$

$$Y_{CVS} = \frac{Y}{S}$$

Modèle additif $Y = T+S+A$

$$Y_{CVS} = Y - S$$

PREVISION

Lissage obtenu par

- Régression linéaire de Y sur le temps t

T = droite de régression $D_{Y/t}$

- Moyennes mobiles (Moyennes mobiles = T provisoire)

Régression linéaire de Y_{CVS} sur le temps t

T = droite de régression $D_{Y_{CVS}/t}$

Prévision à la date future t, correspondant à la saison j:

Modèle multiplicatif $Y = T.S.A$	Modèle additif $Y = T+S+A$
$\hat{Y}(t) = T(t) \times S_j$	$\hat{Y}(t) = T(t) + S_j$