STATISTIQUE DESCRIPTIVE DOUBLE

Introduction :

On s'intéresse à deux variables $\boldsymbol{X}$et $\boldsymbol{Y}$ qui sont mesurées sur les $\boldsymbol{n}$ unités d'observation. Pour chaque unité, on obtient deux mesures. La série statistique est alors une suite $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$ de $\boldsymbol{n}$ couples des valeurs prises par les deux variables sur chaque individu. Chacune des deux variables peut être, soit quantitative, soit qualitative.

1/ Tableaux de données et de contingence :

X est une variable pouvant prendre k modalités et Y est une variable pouvant prendre l modalités. La série $\left(x_i, y_i\right)_{i=1,\dots,n}$ des observations est présentée dans un **tableau de données** :

x_i	x_1	x_2	…	x_n
y_i	y_1	y_2	…	y_n

On construit le **tableau de contingence** qui représente la distribution d'effectif du couple de variable (X, Y) pour $\boldsymbol{n}$ suffisamment grand :

X \ Y	y_1	y_2	y_j	y_l
x_1	n_{11}	n_{12}	n_{1j}	n_{1l}
x_2	n_{21}	n_{22}	n_{2j}	n_{2l}
x_i	n_{i1}	n_{i2}	n_{ij}	n_{il}
x_k	n_{k1}	n_{k2}	n_{kj}	n_{kl}

$\boldsymbol{n_{ij}}$: l'effectif de la cellule $\left(x_i, y_j\right)$ est le nombre d'individus présentant simultanément les modalités x_i de X et y_j de Y.

$\boldsymbol{f_{ij}} = \frac{n_{ij}}{n}$: la fréquence de la cellule$\left(x_i, y_j\right)$.

Exemple 1 : On mesure le poids Y et la taille X de 20 individus. Les observations sont données dans le premier tableau (à gauche), et après répartition en 5 classes d'égales amplitudes pour chacune des deux variables, nous obtenons le **tableau de contingence** ci-dessous (à droite).

y_i	x_i	y_i	x_i
60	155	75	180
61	162	76	175
64	157	78	173
67	170	80	175
68	164	85	179
69	162	90	175
70	169	96	180
70	170	96	185
72	178	98	189
73	173	101	187

X \ Y	[60,69[	[69,78[	[78,87[	[87,96[	[96,105[
[155,162[	2				
[162,169[	2	1			
[169,176[	1	4	2	1	
[176,183[		2	1		1
[183,190[					3

Remarque : Si $\boldsymbol{n}$ est petit, nous n'avons pas besoin de construire le tableau de contingence, nous effectuons les calculs sur le tableau de données.

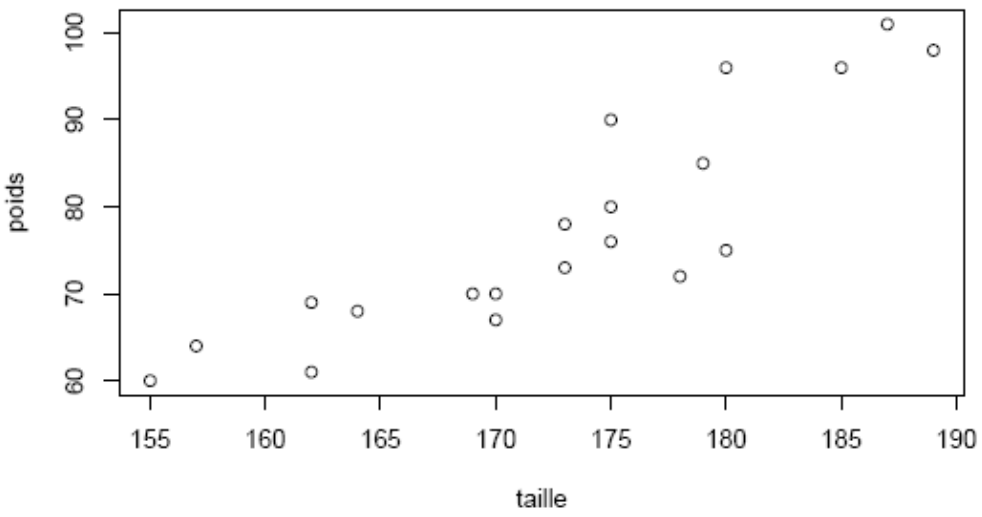
Exemple 2 : Un responsable logistique a effectué 8 observations mesurant le temps de préparation d'une commande en minutes Y et le nombre de colis à préparer X. Ces 8 mesures sont consignées dans le tableau suivant.

Individu i	x_i	y_i	x_i^2	y_i^2
1	7	38	49	1444
2	9	42	81	1764
3	11	53	121	2809
4	13	86	169	7396
5	14	104	196	10816
6	16	144	256	20736
7	18	201	324	40401
8	20	292	400	85264
Σ	**108**	**960**	**1596**	**170630**

2/ Représentation graphique (Nuage de points) :

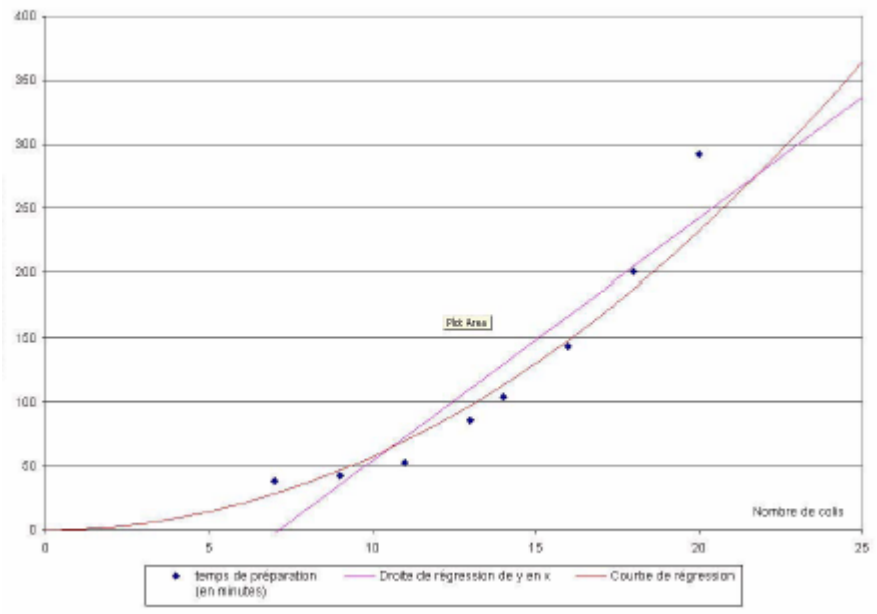
Chaque couple est composé de deux valeurs numériques si les deux caractères sont quantitatifs. Un couple de nombres (entiers ou réels) peut toujours être représenté comme un point dans un plan de coordonnées(x_i, y_i).

Exemple 1 :



Le nuage de points

Exemple 2 :



3/ Distributions :

a/ Distributions marginales :

Y / X	y_1	y_2	…	y_j	…	y_l	Distribution marginale de X
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1l}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2l}	$n_{2\bullet}$
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{il}	$n_{i\bullet}$
x_k	n_{k1}	n_{k2}	…	n_{kj}	…	n_{kl}	$n_{k\bullet}$
Distribution marginale de Y	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet l}$	$n_{\bullet\bullet} = n$

$\boldsymbol{n_{i\bullet}}$: effectif des individus qui présentent la modalité x_i de X. $\boldsymbol{n_{i\bullet}} = \sum_{j=1}^{l} n_{ij}$.

Les effectifs $n_{i\bullet}$ de la **dernière colonne** du tableau de contingence définissent la **distribution marginale de $\boldsymbol{X}$**, alors la fréquence marginale de la modalité x_i est : $f_{i\bullet} = \frac{n_{i\bullet}}{n}$.

$\boldsymbol{n_{\bullet j}}$: effectif des individus qui présentent la modalité y_j de Y. $\boldsymbol{n_{\bullet j}} = \sum_{i=1}^{k} n_{ij}$.

De même, on définit la **distribution marginale de $\boldsymbol{Y}$** par la **dernière ligne** et la fréquence marginale est : $f_{\bullet j} = \frac{n_{\bullet j}}{n}$.

Exemple 1 (distributions marginales) :

On peut représenter les distributions marginales de chaque variable dans un tableau ou les deux à la fois dans le tableau de contingence.

Classes	X_i	$n_{i\bullet}$
[155,162[	158.5	2
[162,169[	165.5	3
[169,176[	172.5	8
[176,183[	179.5	4
[183,190[	186.5	3
Somme		**20**

Distribution marginale de X

Classes	Y_j	$n_{\bullet j}$
[60,69[	64.5	5
[69,78[	73.5	7
[78,87[	82.5	3
[87,96[	91.5	1
[96,105[	100.5	4
Somme		**20**

Distribution marginale de Y

X \ Y		[60,69[	[69,78[	[78,87[	[87,96[	[96,105[	
	x_i \ y_j	64.5	73.5	82.5	91.5	100.5	$n_{i\bullet}$
[155,162[	158.5	2					2
[162,169[	165.5	2	1				3
[169,176[	172.5	1	4	2	1		8
[176,183[	179.5		2	1		1	4
[183,190[	186.5					3	3
	$n_{\bullet j}$	5	7	3	1	4	**20**

b/ Distributions conditionnelles :

La $j^{ème}$ colonne du tableau statistique décrit la sous population des individus possédant la modalité y_j suivant le caractère X. La fréquence conditionnelle de la modalité x_i sachant y_j (ou liée à y_j) est :

$f_{i/j} = \frac{f_{ij}}{f_{\bullet j}} = \frac{n_{ij}}{n_{\bullet j}}$ $\forall\, 1 \leq i \leq k$ pour j fixé.

De même, la distribution conditionnelle sachant x_i :

$f_{j/i} = \frac{f_{ij}}{f_{i\bullet}} = \frac{n_{ij}}{n_{i\bullet}}$ $\forall\, 1 \leq j \leq l$ pour i fixé.

Remarque :

- $f_{ij} = f_{i\bullet} f_{j/i} = f_{\bullet j} f_{i/j}$
- $\sum_{i=1}^{k} f_{i/j} = 1$ et $\sum_{j=1}^{l} f_{j/i} = 1$

Exemple 1 :

- Distribution de X conditionnée par $Y \in [\,69\,,78\,[\,$.

Classes	[162,169[	[169,176[	[176,183[	Σ
x_i	165.5	172.5	179.5	/
$f_{X=x_i/Y\in[69,78[}$	1/7	4/7	2/7	1

- Distribution de Y conditionnée par $X \in [\,169\,,176\,[\,$.

Classes	[60,69[	[69,78[	[78,87[	[87,96[	Σ
y_j	64.5	73.5	82.5	91.5	/
$f_{Y=y_j/X\in[169,176[}$	1/8	4/8	2/8	1/8	1

X \ Y		[60,69[	[69,78[	[78,87[	[87,96[	[96,105[	
	x_i \ y_j	64.5	73.5	82.5	91.5	100.5	$n_{i\bullet}$
[155,162[	158.5	2					2
[162,169[	165.5	2	1				3
[169,176[	172.5	1	4	2	1		8
[176,183[	179.5		2	1		1	4
[183,190[	186.5					3	3
	$n_{\bullet j}$	5	7	3	1	4	**20**

4/ Indépendance de deux variables :

Les deux variables X et Y sont dites indépendantes si les variations de l'un des caractères n'entrainent pas de variations pour l'autre caractère. On posera alors la définition suivante :

Définition :
Les séries statistiques $(x_i, n_{i\bullet})$; $1 \le i \le k$ et $(y_j, n_{\bullet j})$; $1 \le j \le l$ sont dites indépendantes si l'on a :

$$f_{ij} = f_{i\bullet} \times f_{\bullet j} \; ; \; \forall\, 1 \le i \le k \text{ et }, \forall\, 1 \le j \le l.$$

Remarque : En pratique, pour montrer que deux variables **ne** sont **pas indépendantes**, il suffit de trouver un i_0 et un j_0 tels que

$f_{i_0 j_0} \neq f_{i_0 \bullet} \times f_{\bullet j_0}$, ce qui donne $n_{i_0 j_0} \times n \neq n_{i_0 \bullet} \times n_{\bullet j_0}$

Exemple 1 : (Poids et taille de 20 individus)

Pour $i=1$ et j$=1$ on a

2x20=40 et 2x5=10

Alors le poids et la taille ne sont pas indépendants.

5/ Paramètres marginaux :

a/ Moyenne :

Moyenne de X : $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet} x_i = \sum_{i=1}^{k} f_{i\bullet} x_i$

Moyenne de Y : $\bar{y} = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j} y_j = \sum_{j=1}^{l} f_{\bullet j} y_j$

b/ Variance :

Variance de X : $\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}\,(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i\bullet}\, x_i^2 - \bar{x}^2$

$$\sigma_X^2 = \sum_{i=1}^{k} f_{i\bullet}\,(x_i - \bar{x})^2 = \sum_{i=1}^{k} f_{i\bullet}\, x_i^2 - \bar{x}^2$$

Variance de Y : $\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}\,(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=1}^{l} n_{\bullet j}\, y_j^2 - \bar{y}^2$

$$\sigma_Y^2 = \sum_{j=1}^{l} f_{\bullet j}\,(y_j - \bar{y})^2 = \sum_{j=1}^{l} f_{\bullet j}\, y_j^2 - \bar{y}^2$$

Exemple 1 : Calcul des moyennes et variances marginales de X et de Y.

X \ Y		[60,69[	[69,78[	[78,87[	[87,96[	[96,105[			
	x_i \ y_j	64.5	73.5	82.5	91.5	100.5	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
[155,162[	158.5	2					2	317	
[162,169[	165.5	2	1				3	496.5	
[169,176[	172.5	1	4	2	1		8	1380	
[176,183[	179.5		2	1		1	4	718	
[183,190[	186.5					3	3	559.5	
	$n_{\bullet j}$	5	7	3	1	4	**20**	**3471**	**603693**
	$n_{\bullet j}y_j$	322.5	514.5	247.5	91.5	402	**1578**		
	$n_{\bullet j}y_j^2$						**127809**		

$$\bar{x} = \frac{3471}{20} = 173.55 \qquad \bar{y} = \frac{1578}{20} = 78.9$$

$$\sigma_X^2 = 65.0475 \qquad \sigma_Y^2 = 165.24$$

Classes	X_i	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
[155,162[	158.5	2	317	
[162,169[	165.5	3	496.5	
[169,176[	172.5	8	1380	
[176,183[	179.5	4	718	
[183,190[	186.5	3	559.5	
Somme		**20**	**3471**	**603693**

Distribution marginale de X

Classes	Y_j	$n_{\bullet j}$	$n_{\bullet j}y_j$	
[60,69[	64.5	5	322.5	
[69,78[	73.5	7	514.5	
[78,87[	82.5	3	247.5	
[87,96[	91.5	1	91.5	
[96,105[	100.5	4	402	
Somme		**20**	**1578**	**127809**

Distribution marginale de Y

Exemple 2 :

Individu i	x_i	y_i	${x_i}^2$	${y_i}^2$
1	7	38	49	1444
2	9	42	81	1764
3	11	53	121	2809
4	13	86	169	7396
5	14	104	196	10816
6	16	144	256	20736
7	18	201	324	40401
8	20	292	400	85264
Σ	**108**	**960**	**1596**	**170 630**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{8} x_i = 13.5 \quad \text{et} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{8} y_i = 120.$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{8} x_i^2 - \bar{x}^2 = 199.5 - (13.5)^2 = 17.25$$

$$\sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{8} y_i^2 - \bar{y}^2 = \frac{170630}{8} - (120)^2 = 6928.75$$

c/ Moyennes et variances conditionnelles :

- Moyenne de $X/Y = y$: $\bar{x}_{/Y=y} = \sum_{i=1}^{k} f_{i/j} x_i$
- Moyenne de $Y/X = x$: $\bar{y}_{/X=x} = \sum_{j=1}^{l} f_{j/i} y_j$
- Variance de $X/Y = y$:
  $\sigma^2_{X/Y=y} = \sum_{i=1}^{k} f_{i/j} \left(x_i - \bar{x}_{/Y=y}\right)^2 = \sum_{i=1}^{k} f_{i/j}\, x_i^2 - (\bar{x}_{/Y=y})^2$
- Variance de $Y/X = x$:
  $\sigma^2_{Y/X=x} = \sum_{j=1}^{l} f_{j/i} \left(y_j - \bar{y}_{/X=x}\right)^2 = \sum_{j=1}^{l} f_{j/i}\, y_j^2 - (\bar{y}_{/X=x})^2$

Remarque : Si X et Y sont indépendantes alors $\bar{x} = \bar{x}_{/Y=y}$ et $\bar{y} = \bar{y}_{/X=x}$

Exemple 1 : Calculer moyenne et variance de $X/y \in [69,78[$.

Classes	[162,169[	[169,176[	[176,183[	Σ
x_i	165.5	172.5	179.5	
$f_{X=x_i/Y\in[69,78[}$	1/7	4/7	2/7	1
$x_i f_{X=x_i/Y\in[69,78[}$	165.5/7	690/7	359/7	173.5
${x_i}^2 f_{X=x_i/Y\in[69,78[}$	27390.25/7	119025/7	64440.5/7	30122.25

$\bar{x}_{/Y\in[69,78[} = 173.50$ et $\sigma^2_{X/Y\in[69,78[} = 30122.25 - (173.5)^2 = 20$

$\bar{x} = 173.55 \neq 173.50 = \bar{x}_{/Y\in[69,78[}$ alors X et Y ne sont pas indépendantes.

6/ Covariance et coefficient de corrélation :

a/ La covariance : Elle est notée par $cov(X,Y)$ ou S_{XY} et elle est définie par :

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

Et qui peut s'écrire comme suit :

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_i y_j - \bar{x}\bar{y} = \sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}x_i y_j - \bar{x}\bar{y}$$

- **Remarque** : La covariance peut prendre des valeurs positives, négatives ou nulles, et quand $x_i = y_i$, pour tout $i = 1,\dots,n$, la covariance est égale à la variance.

- **Propriétés :**

1. la covariance est symétrique : $cov(X,Y) = cov(Y,X)$.
2. $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$.
3. $cov(X,Y)^2 \leq var(X)var(Y)$.

Exemple 1 : Poids et taille de 20 individus. Calcul de la covariance.

X \ Y	[60,69[64.5	[69,78[73.5	[78,87[82.5	[87,96[91.5	[96,105[100.5	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$	$\sum_{j=1}^{5} n_{ij}Y_j$	$\sum_{j=1}^{5} n_{ij}X_iY_j$
[155,162[158.5	2					2	317		129	20446.5
[162,169[165.5	2	1				3	496.5		202.5	33513.75
[169,176[172.5	1	4	2	1		8	1380		615	106087.5
[176,183[179.5		2	1		1	4	718		330	59235
[183,190[186.5					3	3	559.5		301.5	56229.75
$n_{\bullet j}$	5	7	3	1	4	**20**	**3471**	**603693**		**275 512.5**
$n_{\bullet j}y_j$	322.5	514.5	247.5	91.5	402	**1578**				
$n_{\bullet j}y_j^2$						**127809**				
$\sum_{i=1}^{5} n_{ij}X_i$										
$\sum_{i=1}^{5} n_{ij}X_iY_j$										

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{5}\sum_{j=1}^{5} n_{ij}X_iY_j - \bar{x}\bar{y} = \frac{1}{20}(275\,512.5) - 173.55 \times 78.9 = \mathbf{82.53}$$

b/ Le coefficient de corrélation linéaire :

C'est un indice qui mesure le degré de liaison entre X et Y. Il est noté par $corr(X,Y)$ ou ρ_{XY} , et il est défini par :

$$\rho_{XY} = \frac{S_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{i=1}^{k} f_{i\bullet}\,(x_i - \bar{x})^2}\sqrt{\sum_{j=1}^{l} f_{\bullet j}\,(y_j - \bar{y})^2}}$$

Propriétés :

1. Symétrie : $corr(X,Y) = corr(Y,X)$,

2. $\rho_{XY} \in [-1,1]$

3. $|\rho_{XY}| = 1$ ssi il existe une liaison **linéaire** entre X et Y ($\exists$a, b, c $\in$ IR / $aX + bY + c = 0$),

4. Si X et Y indépendantes alors $corr(X,Y) = 0$.

Exemple 1 : Poids et taille de 20 individus. Calcul de la covariance.

$$\rho_{XY} = \frac{S_{XY}}{\sigma_X \sigma_Y} = \frac{82.53}{8.06 \times 12.85} = 0.797$$

Exemple 2 :

$$\rho_{XY} = \frac{S_{XY}}{\sigma_X \sigma_Y} = \frac{325.375}{4.1533 \times 83.24} = 0.941$$

Avec

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{8} x_i y_i - \bar{x}\bar{y} = \frac{15\,563}{8} - (13.5)(120) = 325.375$$

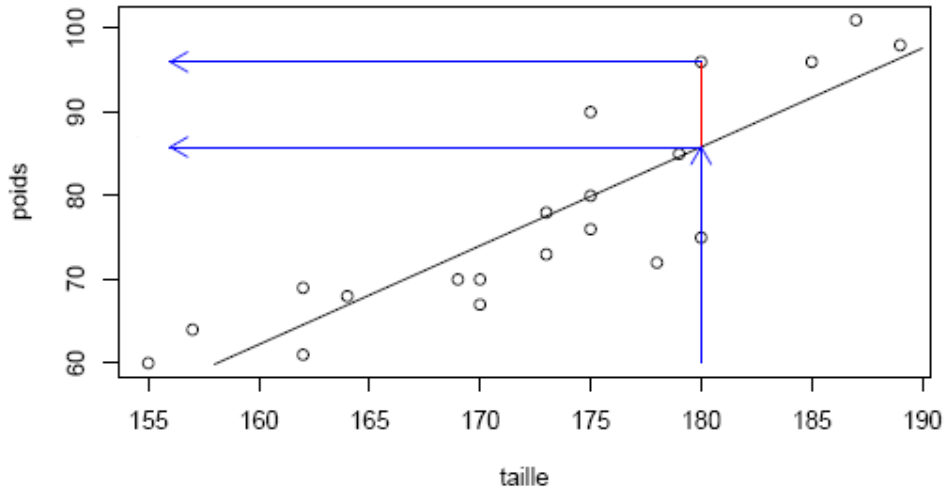
Indi	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	7	38	49	1444	
2	9	42	81	1764	
3	11	53	121	2809	
4	13	86	169	7396	
5	14	104	196	10816	
6	16	144	256	20736	
7	18	201	324	40401	
8	20	292	400	85264	
Σ	**108**	**960**	**1596**	**170 630**	**15 563**

Régression

1/ Ajustement linéaire:

a/ Droite de régression de y en x : $D_y(x)$

Le coefficient de corrélation mesure la dépendance linéaire des variables. Si cette dépendance est bonne, on peut exprimer la variable Y comme fonction linéaire de X. La méthode des moindres carrés consiste à chercher une droite telle que la somme de ses **distances** aux différents points représentant les données soit minimale. **La distance** choisie est **le carré** de la différence des ordonnées entre chaque point et le point de la droite ayant même abscisse. Cette droite a pour équation : $\hat{y} = ax + b$ ($\hat{y}$ estimé n'est pas y observé). Il reste donc à déterminer les valeurs des paramètres a et b, qui désignent respectivement la pente et l'ordonnée à l'origine de la droite d'ajustement.



La différence des ordonnées entre un point (x_i, y_i) et le point de la droite ayant même abscisse est : $y_i - \hat{y}_i = y_i - (ax_i + b)$ et la somme des carrés de ces différences doit être minimum : $D = \sum_{i=1}^{n}(y_i - ax_i - b)^2$ minimum.

La solution est donnée par $\frac{\partial D}{\partial a} = 0$ et $\frac{\partial D}{\partial b} = 0$

On trouve : $\hat{a} = \frac{S_{XY}}{{\sigma_X}^2}$ et $\hat{b} = \bar{y} - \hat{a}\,\bar{x}$

La forme du coefficient b permet de constater que la droite d'ajustement passe par le point moyen (centre de gravité) de coordonnées $\bar{x}$ et $\bar{y}$.

Son équation est : $\hat{y} = \hat{a}x + \hat{b}$.

Exemple 2 :

$\hat{a} = \frac{S_{XY}}{{\sigma_X}^2} = \frac{325.375}{17.25} = 18.86 \qquad \hat{b} = \bar{y} - \hat{a}\,\bar{x} = 120 - (18.86)(13.5) = -134.64$

Alors la droite de régression est : $\hat{y} = 18.86\,x - 134.64$.

b/ Droite de régression de x en y : $D_x(y)$

Le calcul précédent fait jouer un rôle dissymétrique aux variables X et Y (on inverse les rôles des deux variables). On définit une droite d'estimation de x en y d'équation :

$$\hat{x} = \hat{a}'y + \hat{b}'$$

Avec $\hat{a}' = \frac{S_{XY}}{{\sigma_Y}^2}$ et $\hat{b}' = \bar{x} - \hat{a}'\bar{y}$

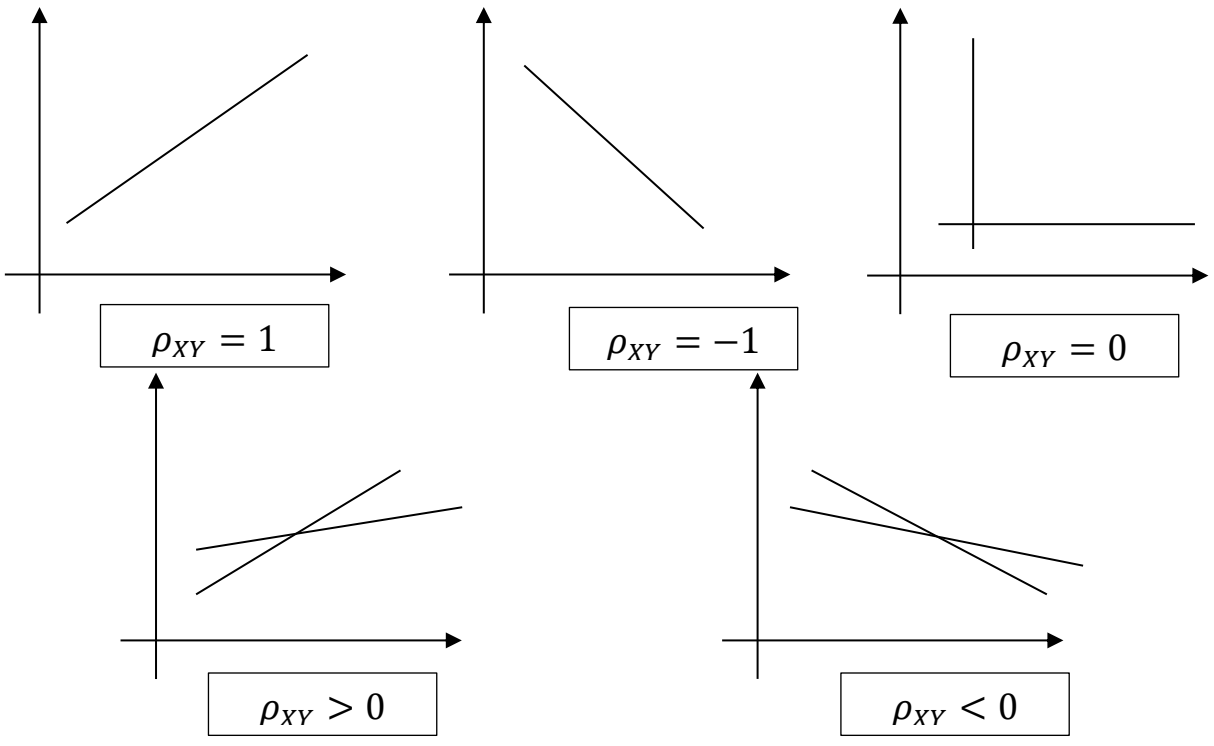
Exemple 2 :

$\hat{a}' = \frac{S_{XY}}{{\sigma_Y}^2} = \frac{325.375}{6928.75} = 0.047 \qquad \hat{b}' = \bar{x} - \hat{a}'\,\bar{y} = 13.5 - (0.047)(120) = 7.86$

Alors la droite de régression est : $\hat{x} = 0.047\,y + 7.86$.

Remarques:

- Les deux droites d'estimation sont différentes, mais on ne peut dire laquelle qui représente un meilleur ajustement.
- Le coefficient de corrélation linéaire est égal au produit des pentes : ${\rho_{XY}}^2 = \hat{a}\,\hat{a}'$.
- Pour s'assurer que l'ajustement est valide, on calcule le coefficient de corrélation, et s'il est voisin en valeur absolue de 1, l'ajustement est valide, $(0{,}7 < |\rho_{XY}| < 1)$.
- Si $|\rho_{XY}| = 1$ alors les points sont alignés.
- Si $\rho_{XY} = 0$ alors X et Y sont non corrélées.
- Si $\rho_{XY} > 0$ alors X et Y croient dans le même sens.
- Si $\rho_{XY} < 0$ alors X et Y croient dans le sens différent.



c/ Prévision :

- La droite de régression de y en x $D_y(x)$ permet de prédire une valeur y pour une valeur x_0 donnée : $\hat{y}_0 = \hat{a}\, x_0 + \hat{b}$.
- La droite de régression de x en y $D_x(y)$ permet de prédire une valeur x pour une valeur y_0 donnée : $\hat{x}_0 = \hat{a}'\, y_0 + \hat{b}'$.

2/ Ajustement non linéaire :

Dans certains cas, l'ajustement à une fonction linéaire n'est pas adéquat : un ajustement des données à une fonction non linéaire doit être envisagé. Les deux cas que nous considérons sont ceux où on peut se ramener par simple transformation à un ajustement affine.

a/ Ajustement à une fonction puissance :

Supposons que les variables statistiques X et Y sont liées par une relation de la forme : $Y = b\, X^a$.
Dans ce cas, cette équation peut être transformée en prenant le logarithme : $\ln Y = a \ln X + \ln b$.
En effectuant les changements de variables suivants : $V = \ln Y$, $U = \ln X$, $B = \ln b$, nous nous ramenons au cas étudié $V = a\, U + B$.

b/ Ajustement à une fonction exponentielle :

Supposons que les variables statistiques X et Y sont liées par une relation de la forme : $Y = b\, e^{aX}$.
Dans ce cas, cette équation peut être transformée en passant aux logarithmes : $\ln Y = a\, X + \ln b$.
En effectuant les changements de variables suivants : $V = \ln Y$, $B = \ln b$, nous nous ramenons au cas étudié $V = a\, X + B$.

Exemple 2 :

En examinant le nuage de points, nous nous proposons d'effectuer un ajustement linéaire ainsi qu'un ajustement à une fonction puissance pour déterminer lequel des deux ajustements est le mieux adapté à la situation.

Nous avons obtenu comme droite de régression :

$$y = 18.86\ x - 134.64.$$

Pour la courbe d'ajustement à une fonction puissance, en effectuant les changements de variables, nous aurons comme droite de régression :

$V = 2.018\, U - 0.596$ et comme $B = \ln b$ alors $b = e^B = e^{-0.596} = 0.551$

Car $\hat{a} = \frac{S_{UV}}{\sigma_U{}^2} = \frac{0.2229}{0.1104} = 2.018$ et $\hat{B} = \bar{v} - \hat{a}\,\bar{u} = 4.5493 - (2.018)(2.5505) = -0.596$

La fonction d'ajustement à une fonction puissance s'écrit donc : $y = 0.551\ x^{2.018}$.

Si nous calculons le coefficient de corrélation pour les deux fonctions d'ajustement, nous obtenons pour l'ajustement linéaire $\rho_{XY} = 0.941$, tandis que l'ajustement "puissance" donne $\rho_{UV} = 0.966$. Ce dernier est donc un meilleur ajustement que l'ajustement linéaire.

Indi	x_i	y_i	$u_i = lnx_i$	$v_i = lny_i$	${x_i}^2$	${y_i}^2$	${u_i}^2$	${v_i}^2$	$x_i y_i$	$u_i v_i$
1	7	38	1.9459	3.6376	49	1444			266	7.0784
2	9	42	2.1972		81	1764			378	
3	11	53	2.3979		121	2809				
4	13	86	2.5649		169	7396				
5	14	104			196	10816				
6	16	144			256	20736				
7	18	201			324	40401				
8	20	292	2.9957		400	85264				
Σ	**108**	**960**	**20.404**	**36.3944**	**1596**	**170 630**	**52.924**	**169.4264**	**15 563**	**94.6072**