Solution du TD N°4
Ajustement d'une courbe et théorie de la corrélation

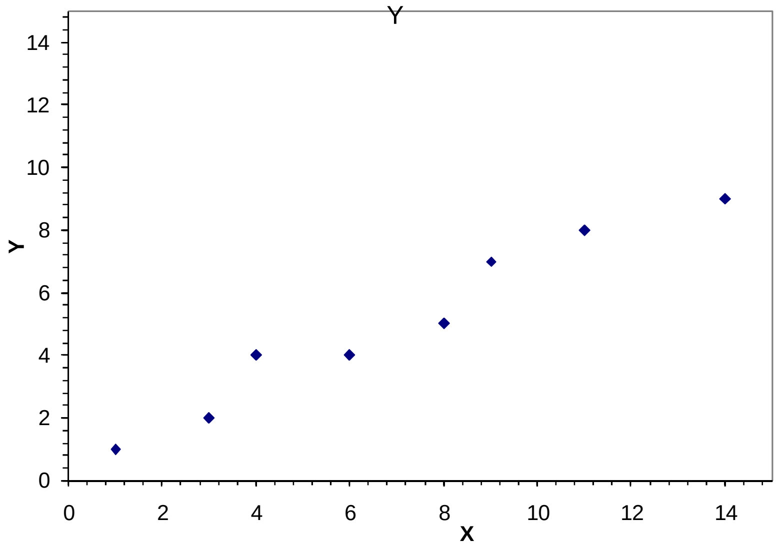
Solution de l'exercice

1) Construire le nuage de points correspondant aux données (x, y) du tableau ci-dessous :

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Solution

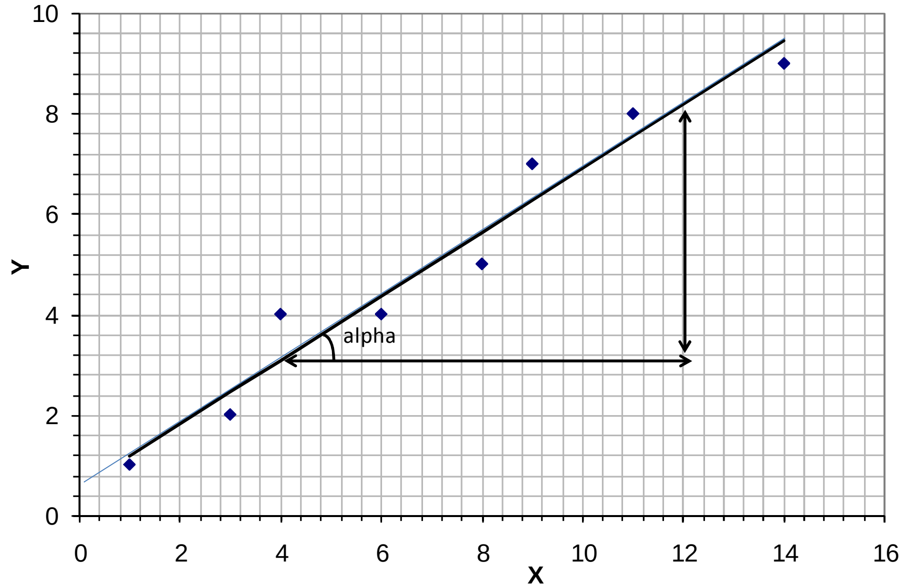
On trace les points (1, 1), (3, 2), (4, 4), (6,4), (8, 5) (9, 7) (11, 8) et (14, 9) dans un système d'axes rectangulaires comme sur la figure.



2) Trouver l'équation de la droite d'ajustement des points par la méthode graphique.

Solution

On trace une droite qui passe au milieu des points, puis on calcul les constants a et b de l'équation de la droite.



Y = ax +b

- a = ponte de la droite = tg (alpha) = dy/dx = (8.5 – 3.5) / (12 – 4) =5/8 = 0.63
- b = intersection de la droite avec y = 0.6

L'équation cherchée est alors Y = 0,63 X + 0.6.

3) Trouver l'équation de la droite d'ajustement des points par la méthode des moindres carrés.

X	Y	X^2	XY	Y^2
1	1	1	1	1
3	2	9	6	4
4	4	16	16	16
6	4	36	24	16
8	5	64	40	25
9	7	81	63	49
11	8	121	88	64
14	9	196	126	81
Σ X= 56	ΣY = 40	ΣX^2 = 524	ΣXY = 364	ΣY^2 = 256

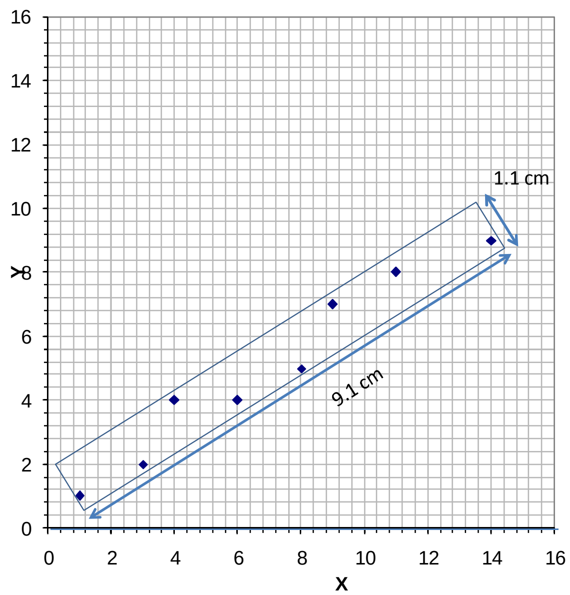
$$a = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{N\sum X^2 - \left(\sum X\right)^2} \qquad \frac{(8)(364)-(56)(40)}{(8)(524)-(56)^2} = \frac{7}{11} = 0.636$$

$$b = \frac{\left(\sum Y\right)\left(\sum X^2\right) - \left(\sum X\right)\left(\sum XY\right)}{N\sum X^2 - \left(\sum X\right)^2} \qquad \frac{(40)(524)-(56)(364)}{(8)(524)-(56)^2} = \frac{6}{11} = 0.545$$

Alors comme Y = a X + b, il vient Y = 0,636X + 0.545.

4) Calculer le coefficient de corrélation de la droite de régression de Y en fonction de X par la méthode graphique

Au début, il faut tracer un graphe y = f(x) avec la même échelle, puis tracer un rectangle qui contenait tous les points et le plus petit possible. Lorsqu'il est tracé, il suffit de prendre une règle graduée et de mesurer les segments.



$$r \approx \pm 1 - \frac{l}{L} \approx \pm 1 - \frac{1.1}{9.1} \approx 0.88$$

5) Calculer le coefficient de corrélation de la droite de régression de Y en fonction de X par la méthode statistique

X	Y	X^2	XY	Y^2
1	1	1	1	1
3	2	9	6	4
4	4	16	16	16
6	4	36	24	16
8	5	64	40	25
9	7	81	63	49
11	8	121	88	64
14	9	196	126	81
Σ X= 56	ΣY = 40	ΣX^2 = 524	ΣXY = 364	ΣY^2 = 256

$$r = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

r = 0.977 = ≈ 0.98

6) Comparer entre les deux coefficients de corrélation.

Les deux coefficients de corrélations sont proche 0.88 et 0.98. La méthode graphique présente une sous estimation, elle est souvent utilisée pour un calcul rapide à titre d'indication. Pour les calculs précis, il faut utiliser la méthode statistique de régression.

7) Est-ce que ce coefficient de corrélation est significatif ?

Il existe plusieurs méthodes pour déterminer la signification du coefficient de corrélation :

- La formule de Mangin (1975) : $r^2 = \frac{2}{\sqrt{n-3}}$ n : nombre des points

  $r^2 = \frac{2}{\sqrt{8-3}}$ = 2/2.24= 0.89 r = 0,94

- Formule 2 de Student : $r = \frac{t}{\sqrt{(n-2)+t^2}}$

  n= nombre des points t= le fractile d'ordre α/2 de la loi de Student à n-2 degrés de liberté (table de student)

  On se fixe un risque acceptable de 5%, c.-à-d. 95% vrai

$$r = \frac{t(8-2)}{\sqrt{(8-2)+t(8-2)^2}} == \frac{t(6)}{\sqrt{(6)+t(6)^2}} == \frac{2.447}{\sqrt{6+(2.447)^2}} = 0.71$$

Dans les deux cas le coefficient de corrélation de la droite (r=0.98) est supérieur aux seuils obtenus par les deux formule (0.94 et 0.71). Cela indique que la droite est représentative des points.

8) Vérifier la formule suivante du coefficient de corrélation (r) : $r = \dfrac{\sum x_r y_r}{\sqrt{\left(\sum x_r^2\right)\left(\sum y_r^2\right)}}$

Où : $x_r = x - x_{moy}$ et $y_r = y - y_{moy}$

X	*Y*	*X²*	*XY*	*Y²*	*X moyenne*	*Xr*	*Yr*	*Xr*Yr*	*(Xr)^2*	*(Yr)^2*
1	1	1	1	1	7	-6	-4	24	36	16
3	2	9	6	4		-4	-3	12	16	9
4	4	16	16	16	Y moyenne	-3	-1	3	9	1
6	4	36	24	16	5	-1	-1	1	1	1
8	5	64	40	25		1	0	0	1	0
9	7	81	63	49		2	2	4	4	4
11	8	121	88	64		4	3	12	16	9
14	9	196	126	81		7	4	28	49	16
$\Sigma X = 56$	$\Sigma Y = 40$	$\Sigma X^2 = 524$	$\Sigma XY = 364$	$\Sigma Y^2 = 256$	Somme	0	0	84	132	56

$$r = \frac{\sum x_r y_r}{\sqrt{\left(\sum x_r^2\right)\left(\sum y_r^2\right)}}$$

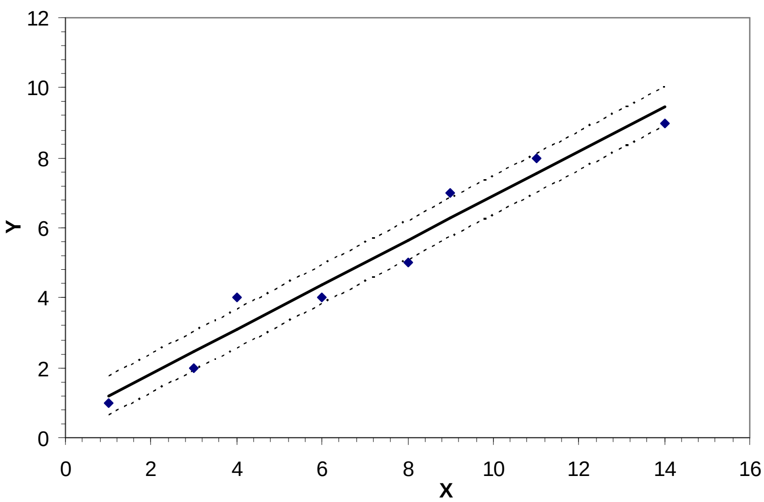
r = 0.977

9) Calculer l'écart-type $S_{Y.X}$.

$$S_{Y.X} = \sqrt{\frac{\sum (Y - Y_{est})^2}{N}}$$

X	*Y*	*Y est*	*(Y - Y est)^2*	*Sy.x*
1	1	1.18	0.03	0.56
3	2	2.45	0.21	
4	4	3.09	0.83	
6	4	4.36	0.13	
8	5	5.63	0.40	
9	7	6.27	0.53	
11	8	7.54	0.21	
14	9	9.45	0.20	

10) Construire deux droites parallèles à la droite de régression situent à une distance $S_{Y.X}$ suivant l'axe de Y.



11) Déterminer la pourcentage des points du nuage compris à l'intérieur de la bande comprise entre les deux droites.
Le pourcentage = 5 / 8 = 0.625 = 62.5 %

12) Si l'on considère que la variable expliquée est X, l'équation de la droite des moindres carrés s'écrit X = a Y + b et les équations normales sont données par :

$$a = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{N\sum Y^2 - \left(\sum Y\right)^2} \qquad \frac{(8)(364)-(56)(40)}{(8)(256)-(40)^2} = \frac{3}{2} = 1.5$$

$$b = \frac{\left(\sum X\right)\left(\sum Y^2\right) - \left(\sum Y\right)\left(\sum XY\right)}{N\sum Y^2 - \left(\sum Y\right)^2} \qquad \frac{(56)(256)-(40)(364)}{(8)(256)-(40)^2} = -\frac{1}{2} = -0.5$$

Ainsi l'équation recherchée de la droite des moindres carrés est X = 1,50 Y – 0.5.
Notons qu'en résolvant l'équation en Y on obtient Y = 0.5/1.5 + 1/1.5 X Y = 1/3 + 2/3 X, soit Y = 0,333 + 0,667X, qui n'est pas la même que la droite obtenue dans le précédent problème à la question e.

13) Calculer le coefficient de corrélation linéaire des variables X et Y.

On peut présenter le problème selon le modèle du tableau suivant : (*attention : petit x est l'équivalent de* x_r *et petit y est l'équivalent de* y_r)

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	xy	y^2
1	1	−6	−4	36	24	16
3	2	−4	−3	16	12	9
4	4	−3	−1	9	3	1
6	4	−1	−1	1	1	1
8	5	1	0	1	0	0
9	7	2	2	4	4	4
11	8	4	3	16	12	9
14	9	7	4	49	28	16
$\Sigma X = 56$ $\bar{X} = 56/8 = 7$	$\Sigma Y = 40$ $\bar{Y} = 40/8 = 5$			$\Sigma x^2 = 132$	$\Sigma xy = 84$	$\Sigma y^2 = 56$

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{84}{\sqrt{(132)(56)}} = 0{,}977$$

Ceci montre qu'il y a une très forte corrélation linéaire entre les deux variables.

Annexe 1 : Table de la loi de Student

α *(bilatéral)*	50 %	60 %	70 %	80 %	90 %	95 %	98 %	99 %	99,5 %	99,8 %	99,9 %
$1-\gamma$ *(unilatéral)*	75 %	80 %	85 %	90 %	95 %	97,5 %	99 %	99,5 %	99,75 %	99,9 %	99,95 %
k											
1	1,000	1,376	1,963	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,767
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
50	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
80	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639	2,887	3,195	3,416
100	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626	2,871	3,174	3,390
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	2,860	3,160	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291