TD de statistiques

Partie 1 : Statistique descriptive

licence de sciences de la vie — deuxième année

Téofil Adamski

janvier 2024

Introduction

7 séances de TD

Me contacter :
`teofil.adamski@univ-smb.fr`
Bureau 20, LAMA

Exercice 1.1

La capacité vitale est le volume d'air maximum pouvant être mobilisé en une seule inspiration. On l'a mesurée sur un échantillon de 18 hommes d'âge compris entre 19 et 23 ans et de taille comprise entre 169 et 173 cm.

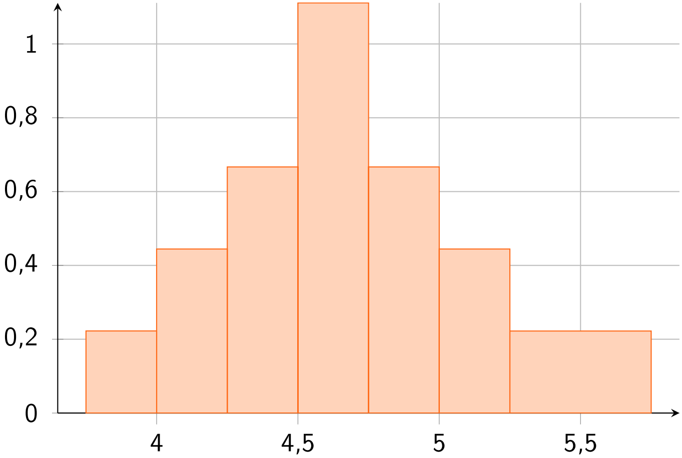
capacité (litres)	effectifs
de 3,75 à moins de 4	1
de 4 à moins de 4,25	2
de 4,25 à moins de 4,5	3
de 4,5 à moins de 4,75	5
de 4,75 à moins de 5	3
de 5 à moins de 5,25	2
de 5,25 à moins de 5,75	2

1. Calculer les fréquences, les pourcentages, les amplitudes et les valeurs centrales de chaque classe.
2. Tracer l'histogramme des fréquences et tracer le polygone des fréquences cumulées.

Exercice 1.1 — correction (question 1)

capacité	effectifs	fréquence	pourcentage	amplitude	densité	valeur centrale
[3,75 ; 4[	1	0,0556	5,56	0,25	0,2224	3,875
[4 ; 4,25[	2	0,1111	11,11	0,25	0,4444	4,125
[4,25 ; 4,5[	3	0,1667	16,67	0,25	0,6668	4,375
[4,5 ; 4,75[	5	0,2778	27,78	0,25	1,1112	4,625
[4,75 ; 5[	3	0,1667	16,67	0,25	0,6688	4,875
[5 ; 5,25[	2	0,1111	11,11	0,25	0,4444	5,125
[5,25 ; 5,75[	2	0,1111	11,11	0,50	0,2222	5,500

Exercice 1.1 — correction (question 1)

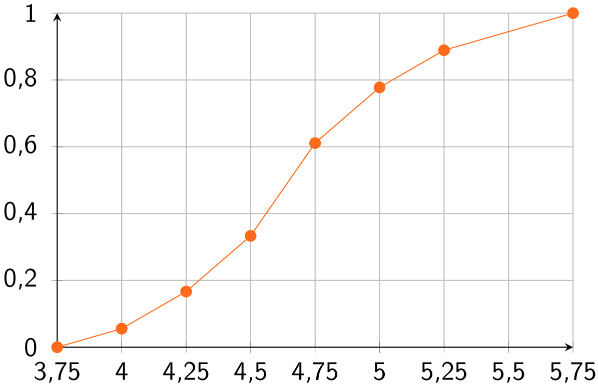


Exercice 1.1 — correction (question 2)

On calcule le tableau des fréquences cumulées.

capacité	fréquence	fréquence cumulée
[3,75 ; 4[	0,0556	0,0556
[4 ; 4,25[	0,1111	0,1667
[4,25 ; 4,5[	0,1667	0,3333
[4,5 ; 4,75[	0,2778	0,6111
[4,75 ; 5[	0,1667	0,7778
[5 ; 5,25[	0,1111	0,8889
[5,25 ; 5,75[	0,1111	1,0000

Exercice 1.1 — correction (question 2)



Exercice 1.2

Le tableau suivant donne la répartition des résultats de 45 mesures de contrôle de la teneur en sulfures, exprimée en g/L, d'une amine régénérée par chauffage. Cette amine est utilisée pour la purification d'un gaz naturel contenant de l'hydrogène sulfuré.

teneur en H_2S	pourcentage
$[0{,}70\,;0{,}75[$	6,67
$[0{,}75\,;0{,}80[$	8,89
$[0{,}80\,;0{,}85[$	13,33
$[0{,}85\,;0{,}90[$	15,56
$[0{,}90\,;0{,}95[$	20,00
$[0{,}95\,;1{,}00[$	13,33
$[1{,}00\,;1{,}05[$	11,11
$[1{,}05\,;1{,}10[$	6,67
$[1{,}10\,;1{,}15[$	4,44

Déterminer la classe modale, le mode, la médiane et les quartiles.

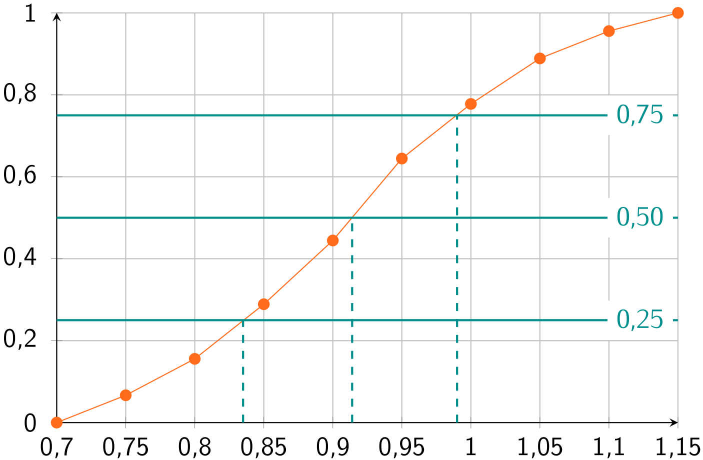
Exercice 1.2 — correction

La classe modale est l'intervalle $[0{,}90\,;0{,}95[$ et elle correspond au mode 0,925.

Pour la médiane et les quartiles, on calcule les fréquences cumulées.

teneur en H_2S	pourcentage	pourcentage cumulée
$[0{,}70\,;0{,}75[$	6,67	6,67
$[0{,}75\,;0{,}80[$	8,89	15,56
$[0{,}80\,;0{,}85[$	13,33	28,89
$[0{,}85\,;0{,}90[$	15,56	44,45
$[0{,}90\,;0{,}95[$	20,00	64,45
$[0{,}95\,;1{,}00[$	13,33	77,78
$[1{,}00\,;1{,}05[$	11,11	88,89
$[1{,}05\,;1{,}10[$	6,67	95,56
$[1{,}10\,;1{,}15[$	4,44	100,00

Exercice 1.2 — correction



Médiane. On cherche un réel $\alpha \in [0\,;1]$ tel que

$$44{,}45(1-\alpha) + \alpha 64{,}45 = 50{,}00.$$

Après calcul, on trouve

$$\alpha = \frac{50{,}00 - 44{,}45}{64{,}45 - 44{,}45} = 0{,}2775.$$

Ainsi la médiane est égale à

$$0{,}90(1-\alpha) + 0{,}95\alpha = 0{,}914.$$

Exercice 1.2 — correction

Premier quartile. On cherche un réel $\alpha_1 \in [0\,;1]$ tel que

$$15{,}56(1-\alpha_1)+\alpha_1 28{,}89 = 25{,}00.$$

Après calcul, on trouve

$$\alpha_1 = \frac{25{,}00-15{,}56}{28{,}89-15{,}56} = 0{,}7082.$$

Ainsi le premier quartile est égal à

$$0{,}80(1-\alpha_1)+0{,}85\alpha_1 = 0{,}835.$$

Deuxième quartile. C'est la médiane.

Troisième quartile. Il vaut 0,990.

Exercice 1.3

On a compté le nombre de segments d'un échantillon de 112 annélides polychètes (*polydora ciliata*). Le tableau suivant en indique la répartition.

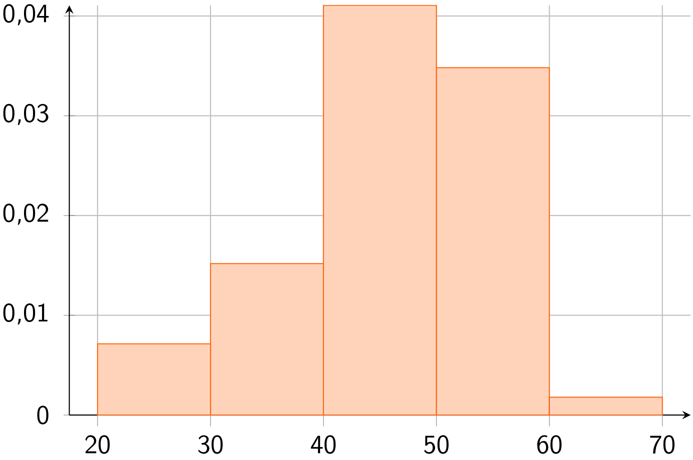
nombre de segments	fréquence
de 20 à moins de 30	0,0714
de 30 à moins de 40	0,1518
de 40 à moins de 50	0,4107
de 50 à moins de 60	0,3482
de 60 à moins de 70	0,0179

1. Calculer les effectifs, les pourcentages, les amplitudes et les valeurs centrales de chaque classe.
2. Tracer l'histogramme des fréquences et tracer le polygone des fréquences cumulées.
3. Déterminer la classe modale, le mode, la médiane et les quartiles.
4. Calculer la moyenne, la variance, l'écart-type et l'intervalle moyen.

Exercice 1.3 — correction (question 1)

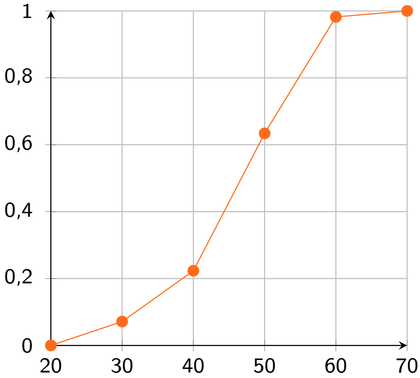
nombre de segments	effectif	fréquence	pourcentage	amplitude	valeur centrale
$[20\,;30[$	8	0,0714	7,14	10	25
$[30\,;40[$	17	0,1518	15,18	10	35
$[40\,;50[$	46	0,4107	41,07	10	45
$[50\,;60[$	39	0,3482	34,82	10	55
$[60\,;70[$	2	0,0179	1,79	10	65

Exercice 1.2 — correction (question 1)



Exercice 1.3 — correction (question 2)

nombre de segments	fréquence	fréquence cumulée
[20 ; 30[	0,0714	0,0714
[30 ; 40[	0,1518	0,2232
[40 ; 50[	0,4107	0,6339
[50 ; 60[	0,3482	0,9821
[60 ; 70[	0,0179	1,0000



Exercice 1.3 — correction (question 3)

La classe modale est l'intervalle $[40\,;50[$ et le mode vaut 45.

Premier quartile. On cherche un réel $\alpha_1 \in [0\,;1]$ tel que

$$0{,}2232(1-\alpha_1) + \alpha_1 0{,}6339 = 0{,}2500.$$

Après calcul, on trouve

$$\alpha_1 = \frac{0{,}2500 - 0{,}2232}{0{,}6339 - 0{,}2232} = 0{,}0652.$$

Ainsi la médiane est égale à

$$40(1-\alpha_1) + 50\alpha_1 = 40{,}65.$$

Deuxième quartile = médiane. Il vaut 46,74.

Troisième quartile. Il vaut 53,33.

Exercice 1.3 — correction (question 4)

Moyenne. Elle vaut

$$\begin{aligned}\overline{x} &= 25 \times 0{,}0714 + 35 \times 0{,}1518 + 45 \times 0{,}4107 + 55 \times 0{,}3482 + 65 \times 0{,}0179 \\ &= 45{,}89.\end{aligned}$$

Variance et écart-type. La variance vaut

$$\begin{aligned}v_x &= \overline{x^2} - \overline{x}^2 \\ &= 25^2 \times 0{,}0714 + 35^2 \times 0{,}1518 + 45^2 \times 0{,}4107 + 55^2 \times 0{,}3482 + 65^2 \times 0{,}0179 - 45{,}89^2 \\ &= 84{,}92.\end{aligned}$$

Ainsi l'écart-type vaut $\sigma_x = \sqrt{v_x} = 9{,}22$.

Intervalle moyen. Il vaut donc

$$[\overline{x} - \sigma_x\,;\overline{x} + \sigma_x] = [36{,}67\,;55{,}11].$$

Exercice 1.4

On a prélevé 54 colonies d'alcyon jaune (*alcyonium digitatum*) et on a mesuré leurs tailles. Les résultats sont présentés dans le tableau suivant.

taille (mm)	fréquence
inférieure à 22	0,2037
de 22 à moins de 44	0,1111
de 44 à moins de 66	0,4630
de 66 à moins de 88	0,1481
de 88 à moins de 110	0,0741

Calculer la moyenne, la variance, l'écart-type et l'intervalle moyen.

Exercice 1.4 — correction

Moyenne. Elle vaut

$$\begin{aligned}\overline{x} &= 11 \times 0{,}2037 + 33 \times 0{,}1111 + 55 \times 0{,}4630 + 77 \times 0{,}1481 + 99 \times 0{,}0741 \\ &= 50{,}11.\end{aligned}$$

Variance et écart-type. La variance vaut

$$\begin{aligned}v_x &= 11^2 \times 0{,}2037 + 33^2 \times 0{,}1111 + 55^2 \times 0{,}4630 + 77^2 \times 0{,}1481 + 99^2 \times 0{,}0741 - \overline{x}^2 \\ &= 639{,}4.\end{aligned}$$

Ainsi l'écart-type vaut $\sigma_x = \sqrt{v_x} = 25{,}29$.

Intervalle moyen. Il vaut donc

$$[\overline{x} - \sigma_x\,;\overline{x} + \sigma_x] = [24{,}83\,;74{,}41].$$

Exercice 1.5

Dans une observation sur la croissance de plants d'eucalyptus, on a relevé les hauteurs de jeunes eucalyptus. Les résultats, regroupés par classes, sont dans le tableau suivant.

hauteur	effectif
$[30\,;80[$	10
$[80\,;100[$	9
$[100\,;110[$	13
$[110\,;120[$	14
$[120\,;130[$	19

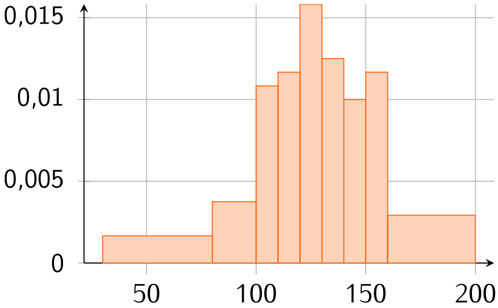
hauteur	effectif
$[130\,;140[$	15
$[140\,;150[$	12
$[150\,;160[$	14
$[160\,;200[$	14

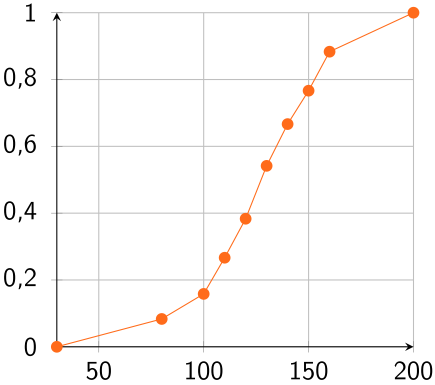
1. Calculer les fréquences, les pourcentages, les amplitudes et les valeurs centrales de chaque classe.
2. Tracer l'histogramme des fréquences et tracer le polygone des fréquences cumulées.
3. Déterminer la classe modale, le mode, la médiane et les quartiles.
4. Calculer la moyenne, la variance, l'écart-type et l'intervalle moyen.

Exercice 1.5 — correction (question 1)

hauteur	effectif	fréquence	fréq. cumulée	pourcentage	amplitude	densité	valeur centrale
[30 ; 80[	10	0,0833	0,0833	8,33	50	0,00166	55
[80 ; 100[	9	0,0750	0,1583	7,50	20	0,00375	90
[100 ; 110[	13	0,1083	0,2666	10,83	10	0,01083	105
[110 ; 120[	14	0,1167	0,3833	11,67	10	0,01167	115
[120 ; 130[	19	0,1583	0,5416	15,83	10	0,01583	125
[130 ; 140[	15	0,1250	0,6666	12,50	10	0,01250	135
[140 ; 150[	12	0,1000	0,7666	10,00	10	0,01000	145
[150 ; 160[	14	0,1167	0,8833	11,67	10	0,01167	155
[160 ; 200[	14	0,1167	1,0000	11,67	40	0,00293	180

Exercice 1.5 — correction (question 2)





Exercice 1.5 — correction (questions 3)

Médiane. En regardant le tableau, la fréquence centrale, c'est-à-dire 0,5000, se trouve entre les fréquences cumulées 0,3833 et 0,5416 et on doit donc considérer l'intervalle $[120\,;130[$. On cherche donc un réel $\alpha \in [0\,;1]$ tel que

$$0{,}3833(1-\alpha) + 0{,}5416\alpha = 0{,}5000.$$

En résolvant cette dernière équation d'inconnue α, on trouve

$$\alpha = \frac{0{,}5000 - 0{,}3833}{0{,}5416 - 0{,}3833} = 0{,}737.$$

La médiane vaut donc

$$120(1-\alpha) + 130\alpha = 127{,}37.$$

Exercice 1.5 — correction (questions 3)

Premier quartile. En regardant le tableau, la fréquence 0,2500 se trouve entre les fréquences cumulées 0,1583 et 0,2666 et on doit donc considérer l'intervalle $[100\,;110[$. On cherche donc un réel $\alpha_1 \in [0\,;1]$ tel que

$$0{,}1583(1-\alpha_1) + 0{,}2666\alpha_1 = 0{,}2500.$$

En résolvant cette dernière équation d'inconnue α_1, on trouve

$$\alpha_1 = \frac{0{,}2500 - 0{,}1583}{0{,}2666 - 0{,}1583} = 0{,}847.$$

Le premier quartile vaut donc

$$100(1-\alpha_1) + 110\alpha_1 = 108{,}47.$$

Exercice 1.5 — correction (questions 3)

Troisième quartile. En regardant le tableau, la fréquence 0,7500 se trouve entre les fréquences cumulées 0,6666 et 0,7666 et on doit donc considérer l'intervalle $[140\,;150[$. On cherche donc un réel $\alpha_3 \in [0\,;1]$ tel que

$$0{,}6666(1-\alpha_3) + 0{,}7666\alpha_3 = 0{,}7500.$$

En résolvant cette dernière équation d'inconnue α_3, on trouve

$$\alpha_3 = \frac{0{,}7500 - 0{,}6666}{0{,}7666 - 0{,}6666} = 0{,}834.$$

Le premier quartile vaut donc

$$140(1-\alpha_3) + 150\alpha_3 = 148{,}34.$$

Exercice 1.5 — correction (question 4)

Moyenne. Elle vaut

$$\begin{aligned}\overline{x} &= \frac{1}{120}(10 \times 55 + 9 \times 90 + 13 \times 105 + 14 \times 115 + 19 \times 125 + 15 \times 135 \\ &\qquad + 12 \times 145 + 14 \times 155 + 14 \times 180) \\ &= 148{,}34.\end{aligned}$$

Variance et écart-type. La variance vaut alors

$$\begin{aligned}v_x &= \frac{1}{120}(10 \times 55^2 + 9 \times 90^2 + 13 \times 105^2 + 14 \times 115^2 + 19 \times 125^2 + 15 \times 135^2 \\ &\qquad + 12 \times 145^2 + 14 \times 155^2 + 14 \times 180^2) - 148{,}34^2 \\ &= 1063{,}73.\end{aligned}$$

L'écart-type vaut donc $\sigma_x = \sqrt{1063{,}73} = 32{,}61$.

Intervalle moyen. Il vaut ainsi

$$[\overline{x} - \sigma_x\,;\overline{x} + \sigma_x] = [93{,}77\,;158{,}99].$$

Exercice 1.6

Pour une certaine mouche (*lucilia cuprina*), le biologiste L. G. Weber a étudié la relation entre le poids P de la pupe et le nombre N d'ovarioles pour des femelles. Les résultats sont indiqués dans le tableau suivant.

P (mg)	N
6,30	62
7,91	76
8,20	81
8,25	80
8,65	83
9,48	105
16,08	142

P (mg)	N
18,15	152
21,70	184
23,12	186
24,88	182
26,20	230
28,82	193
32,00	260

1. Calculer la covariance des deux variables et leur coefficient de corrélation linéaire.
2. Représenter graphiquement ces deux variables par un nuage de points. Déterminer et tracer les deux droites d'ajustement linéaire d'une variable par rapport à l'autre.

Exercice 1.6 — correction (question 1)

On doit calculer la covariance

$$c_{p,n} = \frac{1}{14}\sum_{k=1}^{14} p_k n_k - \overline{p} \times \overline{n}$$

et le coefficient de corrélation linéaire

$$r_{p,n} = \frac{c_{p,n}}{\sigma_p \sigma_n}.$$

Exercice 1.6 — correction (question 1)

	p_k	n_k	p_k^2	$p_k n_k$	n_k^2
1	6,30	62	39,690 0	390,60	3 844
2	7,91	76	62,586 1	601,16	5 776
3	8,20	81	67,240 0	664,20	6 561
4	8,25	80	68,062 5	660,00	6 400
5	8,65	83	74,822 5	717,95	6 889
6	9,48	105	89,870 4	995,40	11 025
7	16,08	142	258,566 4	2 283,36	20 164
8	18,15	152	329,422 5	2 758,80	23 104
9	21,70	184	470,890 0	3 992,80	33 856
10	23,12	186	534,534 4	4 300,32	34 596
11	24,88	182	619,014 4	4 528,16	33 124
12	26,20	230	686,440 0	6 026,00	52 900
13	28,82	193	830,592 4	5 562,26	37 249
14	32,00	260	1024,000 0	8 320,00	67 600
Total	239,74	2 016	5 155,713 6	41 801,01	343 088

Exercice 1.6 — correction (question 1)

La covariance des variables P et N vaut

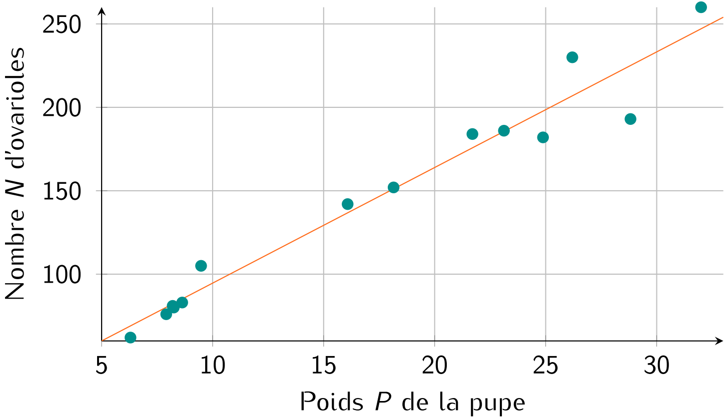
$$c_{p,n} = \frac{1}{14}\sum_{k=1}^{14} p_k n_k - \overline{p} \times \overline{n}$$
$$= \frac{41\ 801{,}01}{14} - \frac{239{,}74}{14} \times \frac{2016}{14} = 519{,}89.$$

Leur coefficient de corrélation vaut

$$r_{p,n} = \frac{c_{p,n}}{\sigma_p \sigma_n}$$
$$= \frac{519{,}89}{\sqrt{\frac{1}{14}\sum_{k=1}^{14} p_k^2 - (\frac{1}{14}\sum_{k=1}^{14} p_k)^2} \times \sqrt{\frac{1}{14}\sum_{k=1}^{14} n_k^2 - (\frac{1}{14}\sum_{k=1}^{14} n_k)^2}} = 0{,}978.$$

La droite d'ajustement linéaire est $n = ap + b$ avec

$$a = r_{p,n}\frac{\sigma_n}{\sigma_p} = 6{,}930 \qquad \text{et} \qquad b = \overline{n} - a\overline{p} = 25{,}335.$$



Exercice 1.7

Lors d'une mission en mer, deux méthodes sont employées pour estimer la concentration de chlorophylle *a* dans l'eau. La première (S) plus grossière et plus ancienne, mais servant toujours de référence, est un dosage par spectrophotométrie. La seconde (H), plus précise mais encore expérimentale, est un dosage par HPLC. On obtient sur le même groupe d'échantillons les résultats suivants en µg/L.

N° de l'essai	1	2	3	4	5	6	7	8
Méthode S	5,7	8,2	6,9	6,0	3,8	3,9	4,3	2,7
Méthode H	4,0	5,8	4,9	4,8	3,6	3,5	2,9	1,2

Soient X la concentration déterminée par la méthode S et Y la concentration déterminée par la méthode H.

1. Calculer la covariance des deux variables et leur coefficient de corrélation linéaire.
2. Représenter graphiquement les deux variables par un nuage de point. Déterminer et tracer la droite d'ajustement linéaire de la variable Y par rapport à la variable X.

Exercice 1.7 — correction (question 1)

	x_k	y_k	x_k^2	$x_k y_k$	y_k^2
1	5,7	4,0	32,49	22,80	16,00
2	8,2	5,8	67,24	47,56	33,64
3	6,9	4,9	47,61	33,81	24,01
4	6,0	4,8	36,00	28,80	23,04
5	3,8	3,6	14,44	13,68	12,96
6	3,9	3,5	15,21	13,65	12,25
7	4,3	2,9	18,49	12,47	8,41
8	2,7	1,2	7,29	3,24	1,44
Total	41,5	30,7	238,77	176,01	131,75

Exercice 1.7 — correction (question 1)

La covariance des variables X et Y vaut

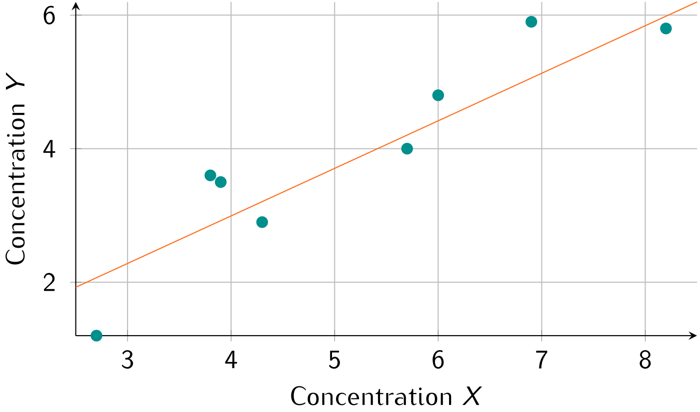
$$
\begin{aligned}
c_{x,y} &= \frac{1}{8}\sum_{k=1}^{8} x_k y_k - \overline{x} \times \overline{y} \\
&= \frac{176{,}01}{8} + \frac{41{,}5}{8} \times \frac{30{,}07}{8} = 2{,}09.
\end{aligned}
$$

Leur coefficient de corrélation vaut

$$
\begin{aligned}
r_{x,y} &= \frac{c_{x,y}}{\sigma_x \sigma_y} \\
&= \frac{2{,}09}{\sqrt{\frac{1}{8}\sum_{k=1}^{8} x_k^2 - (\frac{1}{8}\sum_{k=1}^{8} x_k)^2} \times \sqrt{\frac{1}{8}\sum_{k=1}^{8} y_k^2 - (\frac{1}{8}\sum_{k=1}^{8} y_k)^2}} = 0{,}924.
\end{aligned}
$$

La droite d'ajustement linéaire est $y = ax + b$ avec

$$a = r_{x,y}\frac{\sigma_y}{\sigma_x} = 0{,}712 \qquad \text{et} \qquad b = \overline{y} - a\overline{x} = 0{,}145.$$



Exercice 1.8

Pour effectuer des mesures électriques en biologie ou en médecine, on utilise souvent des électrodes en contact avec un tissu biologique. Dans certains cas, le système électrodes–tissu se comporte comme une capacité et une résistance en parallèle. Les valeurs de la capacité C et de la résistance R dépendent de la fréquence F du courant traversant les électrodes. L'étalonnage d'une sonde formée d'une paire d'électrodes parcourues par un courant d'intensité fixe a donné les résultats suivants.

X	4,91	5,64	6,64	7,64	8,23	8,97	9,97	10,97	11,55	12,29	13,29
Y	10,20	9,98	9,43	9,08	8,81	8,45	7,91	7,32	7,02	6,57	5,86

Dans ce tableau, on a posé

$$X := \log_2 F \qquad \text{et} \qquad Y := \log_2 R$$

avec F en hertz et R en ohms.

1. Calculer la covariance des variables X et Y et leur coefficient de corrélation linéaire.
2. Représenter graphiquement les deux variables par un nuage de points. Déterminer et tracer la droite d'ajustement linéaire de Y par rapport à X.

Exercice 1.8 — correction (question 1)

	x_k	y_k	x_k^2	$x_k y_k$	y_k^2
1	4,91	10,20	24,1081	50,0820	104,0400
2	5,64	9,98	31,8096	56,2872	99,6004
3	6,64	9,43	44,0896	62,6152	88,9249
4	7,64	9,08	58,3896	69,3712	82,4464
5	8,23	8,81	67,7329	72,5063	77,6161
6	8,97	8,45	80,4609	75,7965	71,4025
7	9,97	7,91	99,4009	78,8627	62,5681
8	10,97	7,32	120,3409	80,3004	53,5824
9	11,55	7,02	133,4025	81,0810	49,2804
10	12,29	6,57	151,0441	80,7453	43,1649
11	13,29	5,86	176,6241	77,8794	34,3396
Total	100,1	90,63	987,3832	785,5272	766,9657

Exercice 1.8 — correction (question 1)

La covariance des variables X et Y vaut

$$c_{x,y} = \frac{785{,}5272}{11} - \frac{100{,}1}{11} \times \frac{90{,}63}{11} = -3{,}56.$$

Leur coefficient de corrélation linéaire vaut

$$r_{x,y} = \frac{-3{,}56}{\sqrt{\frac{987{,}3831}{11} - \left(\frac{100{,}1}{11}\right)^2} \times \sqrt{\frac{766{,}9657}{11} - \left(\frac{90{,}63}{11}\right)^2}} = -0{,}996.$$

Exercice 1.8 — correction (question 2)

La droite d'ajustement linéaire est $y = ax + b$ avec

$$a = -0{,}513 \qquad \text{et} \qquad b = 12{,}90.$$