

Remarque : En 2020-2021, les examens ont eu lieu en ligne. Ce document retranscrit le corrigé correspondant à un exemple d'énoncé du contrôle commun du 11 mars.

Exercice 1 : Niveau d'empathie

On étudie le niveau d'empathie d'un échantillon d'adultes Bourguignons. On désigne par X leur note à un test qui détermine leur niveau d'empathie (les niveaux élevés indiquent des individus très empathiques). Ces notes sont résumées ci-dessous :

Empathie X	[26 ; 64[[64 ; 83[[83 ; 102[[102 ; 121[[121 ; 140[[140 ; 178[
Effectif	8	43	81	85	27	6

1. Combien de personnes y a-t'il dans cet échantillon ?

Il y en a $8 + 43 + 81 + 85 + 27 + 6 = 250$,

2. Au sein de cet échantillon, combien vaut la proportion $\mathbb{P}_r[64 \leq X < 102]$?

En ajoutant les effectifs des classes $[64 ; 83[$ et $[83 ; 102[$ on compte $43 + 81 = 124$ personnes, donc cette proportion vaut $\frac{124}{250} = 0,496$.

3. Calculez $m(X)$. Dans cette question, on vous demande d'indiquer les calculs effectués.

$$m(X) = \frac{\sum c_i n_i}{n} = \frac{45 \times 8 + 73,5 \times 43 + 92,5 \times 81 + \dots + 159 \times 6}{250} = \frac{24968}{250} \simeq 99,87$$

4. Combien vaut l'écart type $s(X)$?

$$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{45^2 \times 8 + 73,5^2 \times 43 + 92,5^2 \times 81 + \dots + 159^2 \times 6}{250} = \frac{2609797}{250}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{2609797}{250} - \left(\frac{24968}{250}\right)^2 \simeq 464,77$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 21,56$$

5. Calculez les fréquences et les fréquences cumulées.

Empathie X	[26 ; 64[[64 ; 83[[83 ; 102[[102 ; 121[[121 ; 140[[140 ; 178[
Effectif	8	43	81	85	27	6
fréquence	0,032	0,172	0,324	0,340	0,108	0,024
fréquence cumulée	0,032	0,204	0,528	0,868	0,976	1,000

6. Calculer le premier quartile. Dans cette question, on vous demande d'indiquer les calculs effectués

Classe du premier quartile : $[83 ; 102[$

$$Q_1 \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)}(0,25 - F_X(a_i)) \simeq 83 + \frac{102 - 83}{0,528 - 0,204}(0,25 - 0,204) \simeq 85,7$$

7. Calculer le troisième quartile.

Classe du troisième quartile : $[102 ; 121[$

$$Q_3 \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)}(0,75 - F_X(a_i)) \simeq 102 + \frac{121 - 102}{0,868 - 0,528}(0,75 - 0,528) \simeq 114,41$$

Exercice 2 : Taux de CO_2 et cognition

On étudie le lien entre le taux de CO_2 dans l'air que nous respirons et nos capacités cognitives.

On choisit pour cela un groupe de 11 volontaires, que l'on soumet pendant une journée à un environnement intérieur où l'on a fixé le taux CO_2 . Ce taux de CO_2 est exprimé en ppm ("parties par million") et désigné par la lettre X . À l'issue de cette journée, on leur fait passer un test cognitif, qui attribue à chacun d'entre eux une note Y . On obtient les données ci-dessous :

Taux X de CO_2 (en ppm)	700	700	700	950	950	950	1200	1200	1200	1450	1450
Note Y au test cognitif	112	81	107	75	92	83	78	73	79	59	69

1. Calculer le coefficient de corrélation des rangs (de "spearman"). Que peut-on conclure ?

sujet	1	2	3	4	5	6	7	8	9	10	11
X	700	700	700	950	950	950	1200	1200	1200	1450	1450
Y	112	81	107	75	92	83	78	73	79	59	69
rang X'	2	2	2	5	5	5	8	8	8	10,5	10,5
rang Y'	11	7	10	4	9	8	5	3	6	1,0	2
$(X' - Y')^2$	81	25	64	1,0	16	9	9	25	4	90,25	72,25

le coefficient de corrélation des rangs de Spearman est donc $1 - \left(6 \times \frac{81+25+64+1,0+16+\dots+72,25}{11(11^2-1)} \right) \simeq -0,802$

Cela montre un fort lien entre les deux variables, où Y diminue quand X augmente.

2. Calculer la moyenne et l'écart type de X . On demande dans cette question d'indiquer les calculs effectués

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{700+700+700+\dots+1450}{11} = \frac{11450}{11} \simeq 1\,040,91 \text{ ppm}$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{700^2+700^2+700^2+\dots+1450^2}{11} = \frac{12\,702\,500}{11}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{12\,702\,500}{11} - \left(\frac{11450}{11} \right)^2 \simeq 71\,280,992$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 266,99 \text{ ppm}$$

3. Calculer le coefficient de corrélation linéaire des variables X et Y . Que peut-on conclure ?

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{112+81+107+\dots+69}{11} = \frac{908}{11} \simeq 82,55$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{112^2+81^2+107^2+\dots+69^2}{11} = \frac{77\,428}{11}$$

$$Var(Y) = m(Y^2) - m(Y)^2 = \frac{77\,428}{11} - \left(\frac{908}{11} \right)^2 \simeq 225,16$$

$$\text{Écart-type : } s(Y) = \sqrt{Var(Y)} \simeq 15,01$$

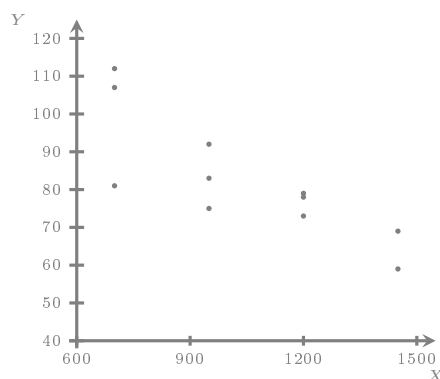
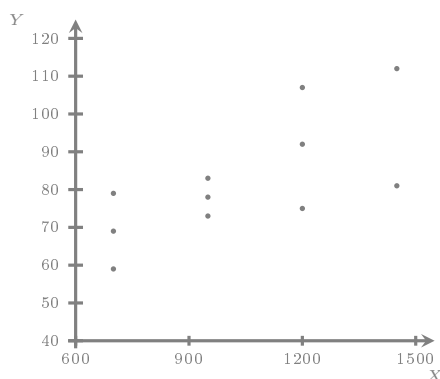
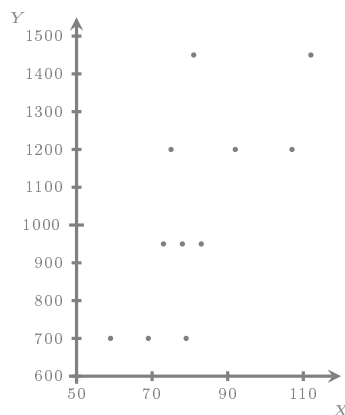
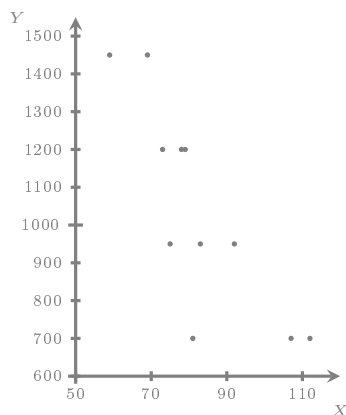
$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{700 \times 112 + 700 \times 81 + \dots + 1450 \times 69}{11} = \frac{909\,100}{11} \simeq 82\,645,455$$

$$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{909\,100}{11} - \frac{11450}{11} \frac{908}{11} \simeq -3\,276,86$$

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-3\,276,86}{\sqrt{71\,280,992 \times 225,157}} \simeq -0,818$$

Cela montre un fort lien linéaire entre les deux variables, où Y diminue quand X augmente.

4. Lequel des quatre graphiques ci-dessous est le nuage de points de ces données ? *On ne demande aucune justification : indiquez juste si c'est le premier, le deuxième, le troisième ou le quatrième*



C'est le quatrième.

5. Estimer quelle serait la note (à ce test cognitif) d'un·e volontaire soumis·e à environnement où le taux de CO_2 serait de 1117 ppm. *On vous demande de préciser quelle droite de régression vous utilisez avant de conclure.*

on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{-3\,276,86}{71\,280,992} \simeq -0,046$ et $b = m(Y) - a m(X) \simeq 82,545 - (-0,046) \times 1\,040,909 \simeq 130,427$

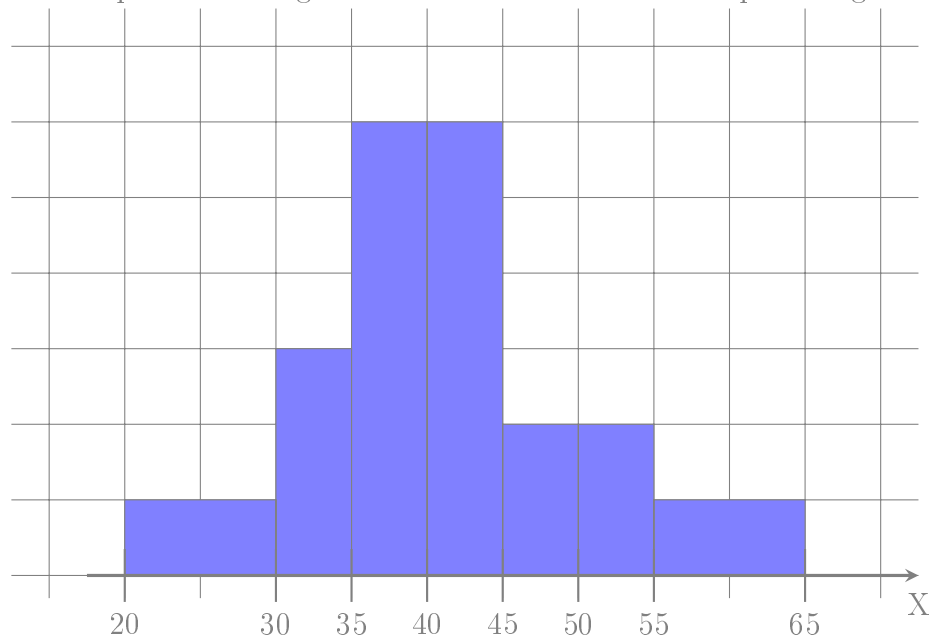
D'où l'équation de la droite $D_{Y|X} : Y = -0,046 X + 130,427$

Donc pour $x = 1117$, on s'attend à $y = -0,046 \times 1\,117 + 130,427 \simeq 79,045$.

Exercice 3 : Pollution sonore et santé

On étudie le niveau de pollution sonore à Dijon (il s'agit par exemple de bruits de circulation que l'on entend depuis son domicile au d'autres bruits récurrents). Dans cet exercice on désigne par X le niveau de pollution sonore que des individus subissent quotidiennement, qui est exprimé en décibels (dB). Ce niveau de bruit est mesuré à l'aide d'un decibelmètre puis une moyenne des mesures effectuée au cours de la journée (en accordant par exemple un coefficient plus important au bruit nocturne) pour aboutir à une valeur X .

1. On a mesuré ce niveau de pollution sonore au sein d'un échantillon d'habitants de l'agglomération dijonnaise, et réalisé l'histogramme ci-dessous (où ils se trouve que les rectangles tombent exactement sur le quadrillage de la feuille).



- (a) Indiquez quelle est la variable statistique et quelle est sa nature.
La variable statistique est la pollution sonore X . Elle est quantitative continue.
- (b) À partir de cet histogramme, déterminer les fréquences des différents intervalles.

Pour chaque intervalle, on obtient la fréquence en divisant la surface du rectangle correspondant par la surface totale. On obtient :

Pollution sonore X	$[20; 30[$	$[30; 35[$	$[35; 40[$	$[40; 45[$	$[45; 50[$	$[50; 55[$	$[55; 65[$
fréquence	0,087	0,130	0,261	0,261	0,087	0,087	0,087

- (c) En déduire quel est environ le niveau médian de pollution sonore au sein de cet échantillon.

On calcule les fréquences cumulées puis on déduit la médiane :

Pollution sonore X	[20 ; 30[[30 ; 35[[35 ; 40[[40 ; 45[[45 ; 50[[50 ; 55[[55 ; 65[
Fréquences cumulées	0,087	0,217	0,478	0,739	0,826	0,913	1,000

Classe de la médiane : [40 ; 45[

$$\text{Méd} \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 40 + \frac{45 - 40}{0,739 - 0,478} (0,5 - 0,478) \simeq 40,42$$

2. Un médecin Dijonnais s'intéresse au lien entre la santé de ses patients et le niveau de pollution sonore auquel ils sont exposés. Il choisit donc de considérer 14 patients qui sont atteints de maladies chroniques. Il constate qu'ils subissent les niveaux de pollution sonore ci dessous :

41,4 dB, 47,4 dB, 39,3 dB, 39,1 dB, 41,4 dB, 44,7 dB, 53,9 dB, 46,3 dB, 32,9 dB, 36,8 dB, 51,4 dB, 46,3 dB, 35,8 dB et 43,7 dB.

Déterminer la moyenne, l'écart type et la médiane de X au sein de cet échantillon de patients souffrant de maladies chroniques.

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{41,4 + 47,4 + 39,3 + \dots + 43,7}{14} = \frac{600,4}{14} \simeq 42,89$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{41,4^2 + 47,4^2 + 39,3^2 + \dots + 43,7^2}{14} = \frac{26\,208,6}{14}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{26\,208,6}{14} - \left(\frac{600,4}{14}\right)^2 \simeq 32,86$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 5,73$$

La médiane est la valeur numéro $\frac{14+1}{2} = 7,5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 7 et 8. C'est donc le milieu entre 41,4 et 43,7 c'est-à-dire $\frac{41,4 + 43,7}{2} = 42,55$.