Exercice 2 :

Un biologiste veut étudier la corrélation entre le nombre de graines dans un petits pois (**noté par X**) et la longueur de la gousse mesurée en mm (**noté par Y**). Les résultats sont donnés dans le tableau de contingence suivant :

X \ Y	[40,50[	[50,60[	[60,70[	[70,80[
2	6	2	1	
4	2	8	3	1
6		1	7	2
8			1	4

1- Les variables X et Y sont-elles indépendantes ?
2- Calculer la covariance entre les variables X et Y.
3- Déduire le coefficient de corrélation linéaire $r(X,Y)$. Interpréter.

4- Donner l'équation des deux droites de régression. Tracer les deux droites dan repère orthonormé. Que peut-conclure ?
5- Peut-on avoir la taille d'une gousse qui contient 10 graines ?

Exo2:

X : le nombre de graines ; Y : la longueur (mm).

X \ Y	[40,50[45	[50,60[55	[60,70[65	[70,80[75	[80,90[85	$n_{i.}$	$n_{i.} x_i$	$n_{i.} x_i^2$
2	6_{540}	2_{220}	1_{130}			9	18	36
4	2_{360}	8_{1760}	3_{780}	1_{300}		14	56	224
6		1_{330}	7_{2730}	2_{900}		10	60	360
8			1_{520}	4_{2400}	2_{1360}	7	56	448
$n_{.j}$	8	11	12	7	2	40	190	1068
$n_{.j} y_j$	360	605	780	525	170	2440	12330	
$n_{.j} y_j^2$	16200	33275	50700	39375	14450	154 00		

1. Les variables X et Y sont-elles indépendantes :

$f_{ij} = f_{i.} \times f_{.j}$

$f_{11} = \frac{6}{40} \stackrel{?}{=} f_{1.} \times f_{.1}$

$\frac{6}{40} \neq \frac{9}{40} \times \frac{8}{40}$ ⇒ X et Y ne sont pas indépendantes.

⇒ il existe un lien entre les variables X et Y.

2. Calculer la covariance entre les variables X et Y.

$$Cov(X,Y) = \frac{1}{n} \sum_i \sum_j n_{ij} (x_i - \overline{x})(y_j - \overline{y})$$

$$= \frac{1}{n} \sum_i \sum_j n_{ij} x_i y_j - \overline{x}.\overline{y}$$

$$\begin{cases} \overline{x} = \frac{1}{n} \sum_i n_{i.} x_i = \frac{190}{40} = 4,75 \\ \overline{y} = \frac{1}{n} \sum_j n_{.j} y_j = \frac{2440}{40} = 61 \text{ mm} \end{cases}$$

$$Cov(X;Y) = \frac{1}{40} . 12330 - (4,75)(61)$$

$$Cov(X;Y) = 18,5$$

3. Déduire le coefficient de corrélation

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_x . \sigma_y} = \frac{18,5}{(2,13)(11,36)} \Rightarrow r(X;Y) = 0,7$$

il faut calculer :

$$V(X) = \frac{1}{n} \sum_i n_{i.} x_i^2 - \overline{x}^2$$

$$= \frac{1}{40} (1068) - (4,75)^2 \Rightarrow V(X) = 4,14 \ , \ \sigma_x = 2,1$$

$$V(Y) = \frac{1}{n} \sum_j n_{.j} y_j^2 - \overline{y}^2$$

$$= \frac{1}{40} (154\,000) - (61)^2 \Rightarrow V(Y) = 129 \ , \ \sigma_y = 11,36$$

$r(X;Y) = 0,76 \Rightarrow r^2(X,Y) = 0,57 \geqslant 0,7$ (forte corrélation).

$\begin{cases} \underline{\text{Table :}} \\ n = 40 \Rightarrow r^2 = 0,7 \text{ (C. Théorique)} \\ \text{décision : si } \underset{\text{calculé}}{r^2(X,Y)} > 0,7 \text{ ; forte corrélation} \\ \qquad\qquad\; r^2(X;Y) < 0,7 \text{ ; faible corrélation.} \end{cases}$

$\overline{X} = 4,15$; $V(X) = 4,14$; $\sigma_x = 2,03$

$\overline{Y} = 67$; $V(Y) = 129$; $\sigma_y = 11,36$

$Cov(X;Y) = 18,5$

4. Donner l'équation des deux droites de régression :

. $(D) : Y = aX + b$ $\begin{cases} a = \dfrac{Cov(X;Y)}{V(X)} = 4,47 \\ b = \overline{Y} - a\overline{X} = 39,7 \end{cases}$

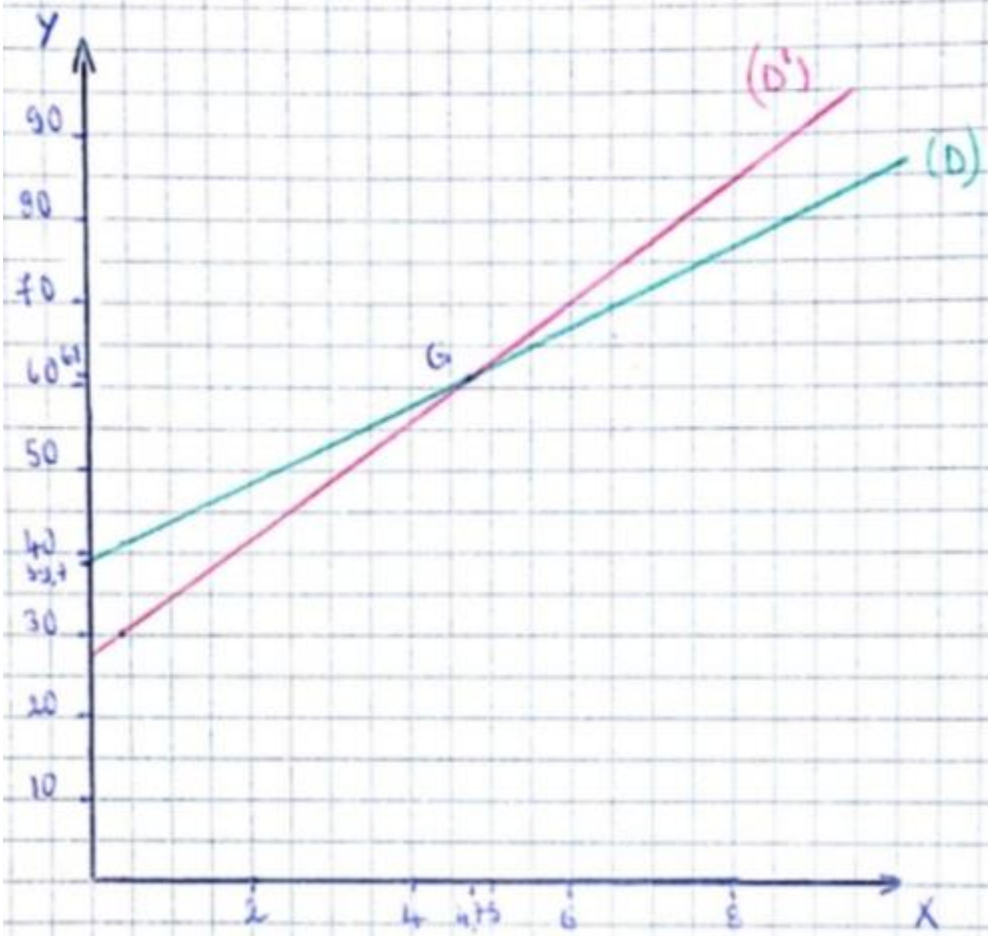
$(D) : Y = 4,47X + 39,7$

. $(D') : X = \alpha Y + \beta$ $\begin{cases} \alpha = \dfrac{Cov(X;Y)}{V(Y)} = 0,14 \\ \beta = \overline{X} - \alpha\overline{Y} = -3,8 \end{cases}$

$(D') : X = 0,14Y - 3,8$

- Les 2 droites passent par le point $G(\overline{X}, \overline{Y})$

4. Tracer les 2 droites dans repère orthonormé :



- On conclue que les 2 droites sont très proches <u>car</u> on a une forte corrélation

5. Quel est la taille d'une gousse qui contient 10 graines ?

Si $X = 10 \Rightarrow Y = 100$ mm

- On pose : $y' = \dfrac{y - 65}{10} = \dfrac{1}{10} y - 6{,}5$

. $\overline{y'} = \dfrac{1}{10} \overline{y} - 6{,}5 = \dfrac{-16}{40} = -0{,}4 \Rightarrow \overline{y} = 61$

- $V(y') = \dfrac{1}{100} V(y) = \dfrac{58}{40} - (-0{,}4) = 1{,}29 \Rightarrow V(y) = 1\ 9$

. $Cov(x;y') = \dfrac{1}{10} Cov(x.y') = \dfrac{-2}{40} - (4{,}75)(-0{,}4) = 1{,}85$

$\Rightarrow Cov(x.y) = 18{,}5$

Changement de variable

TD 1

Exercice 1

Dans une expérimentation animale, un nombre de 16 rats adultes de poids moyen 380 ± 15 g sont soumis à deux régimes alimentaire ,un régime A (riche en sucre) et un régime B (riche en matière grasse).

Les résultats de gain des poids (g) pour chaque groupe sont les suivants :

Groupe A	45	50	30	20	40	20	45
Groupe B	30	10	30	25	14	20	35

1-Calculer la moyenne et l'écart-type.
2-Décrire vos résultats dans un tableau en présentant : la moyenne, écart-type, la valeur minimale et la valeur maximale pour chaque type de régime.
3-Présenter vos résultats sous forme d'un histogramme (Moyenne ± écart-type).
4-Quelle est la série des données qui a un écart-type le plus large ? Expliquer.

Exercice 4 :

Le taux du cortisol normal à 8h est de 100-250 µg/l. Le taux du cortisol dosé chez des étudiants au cours de la période des examens est le suivant :

102	280	150	300	280
130	230	180	205	240

- Calculer la médiane.
- Calculer les quartiles (Q1, Q2, Q3 et Q4).
- Calculer écart-interquartile.
- Représenter les valeurs par une boite de moustache.
- Identifier les valeurs extrêmes. Justifier

Exercice 1

Dans une expérimentation animale, un nombre de 16 rats adultes de poids moyen 380 ± 15 g sont soumis à deux régimes alimentaire ,un régime A (riche en sucre) et un régime B (riche en matière grasse).

Les résultats de gain des poids (g) pour chaque groupe sont les suivants :

Groupe A	45	50	30	20	40	20	45
Groupe B	30	10	30	25	14	20	35

1-MoyA= $\frac{45+50+30+20+40+20+45}{7}$ = 35.71

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2.$$

Var A= $\frac{(45- 35.71) =(50-35.71) (30-35.71) +(20-35.71)+(40-35.71)+ (20-35.71)+(45-35.71)}{7}$ = 131.61

Ecart-type A= $\sqrt{varA}$ = 11.47

- MoyB= $\frac{30+10+30+25+14+20+35}{7}$ = 23.43

Var B= $\frac{(30- 23.43) =(10-23.43) (30-23.43) +(25-23.43)+(14-23.43)+ (20-23.43)+(35-23.43)}{7}$ = 71.96

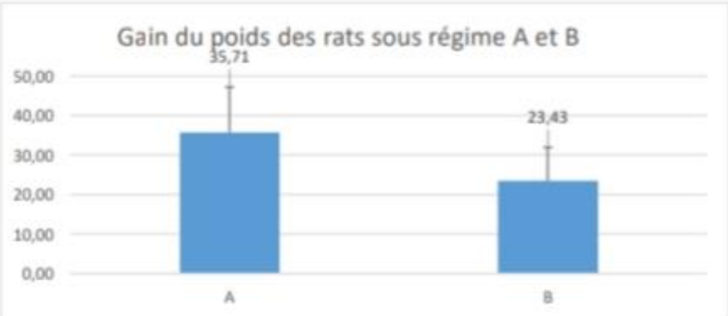
Ecart-type B =$\sqrt{varB}$ = 8.48

2- Tableau.... : gain de poids des rats ayant un régime riche en sucre (A) et un régime riche en matière

grasse (B)

	Moyenne ± Ecart-type	Maximum	Minimum
Groupe A	131.61 ± 11.47	50	20
Groupe B	23.43 ± 8.48	10	35

3- Histogramme (Moyenne ± écart-type).



4- Ecart-type du groupe A > Ecart-type du groupe B, cela indique que les variables du groupe A ne sont dispersé loin de la moyenne (série hétérogène) que celles du groupe B.

Ex 2 :

Exercice 4 :

102, 130, 150, 180, 205, 230, 240, 280, 280, 300

- Médiane= 205+230 = 217.5

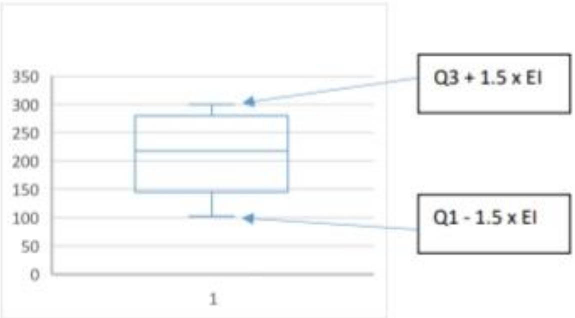
4

2

- Q1= 10/4= 2.5 , 3ème valeur 150 . Q2= Médiane= 217.5 et non $\frac{2x\ 10}{4}$= 5ème valeur :205

Q3= $\frac{3x\ 10}{4}$= 7.5, 8ème valeur : 280 Q4= $\frac{4\ x\ 10}{4}$= 10ème valeur : 300

- Ecart-interquartile= EI = $Q_3 - Q_1$= 280-150= 130.



Les valeurs aberrantes sont < Q1 - 1.5 x EI ou > Q3 + 1.5 x EI

-Voir la valeur 102 si elle est aberrante ? 150-(1.5x130) = 45.

La valeur 102> 45 donc 102 est une valeur non aberrante.

-Voir 300 si elle est aberrante ? 270 +(1.5x130) = 465

La valeur 300 < 465 donc 300 est une valeur non aberrante.