Fiche de révision 1.

Exercice 1.

Une société financière a étudié la relation entre le **montant investi** par des clients (en milliers d'euros), noté X, et le **revenu annuel généré** par ces placements (en milliers d'euros), noté Y. Les observations recueillies pour six clients sont :

$$X = (20, 25, 30, 35, 40, 50) \quad \text{et} \quad Y = (4.2, 5.1, 5.9, 6.8, 7.4, 9.1).$$

On souhaite déterminer si une relation linéaire simple existe entre le montant investi et le revenu généré. À partir de ces données, on vous demande de réaliser une étude statistique basée sur la régression linéaire.

1. Déterminer l'équation de la droite de régression de Y sur X à l'aide de la méthode des moindres carrés.
2. Calculer le coefficient de corrélation linéaire r entre X et Y.
3. En déduire le coefficient de détermination R^2.
4. Interpréter les résultats dans ce contexte financier.

Exercice 2.

Une compagnie d'assurance a analysé le nombre de **sinistres mensuels** déclarés par un portefeuille de clients. Les données observées sur 200 mois sont les suivantes (nombre de sinistres par mois) :

Sinistres	0	1	2	3	4	5	6
Effectif	78	22	10	15	5	12	58

L'objectif est d'étudier la distribution du nombre de sinistres.

1. Déterminer la médiane.
2. Calculer les quartiles puis le mode.
3. Interpréter les résultats.

Exercice 3.

Un service comptable a regroupé les **montants des factures mensuelles** (en milliers d'euros) dans le tableau suivant, représentant la distribution des 50 factures reçues :

Classe (k€)	[2 ; 4[	[4 ; 6[	[6 ; 8[	[8 ; 12[	[12 ; 20]
Effectif	8	12	15	10	5

On souhaite estimer le niveau central des dépenses mensuelles d'approvisionnement.

1. Déterminer la médiane.
2. Calculer les quartiles puis le mode.
3. Interpréter les résultats.

Correction Fiche révision 1.

Exercice 1

Données

$$X = (20, 25, 30, 35, 40, 50), \quad Y = (4.2, 5.1, 5.9, 6.8, 7.4, 9.1)$$

Nombre d'observations : $n = 6$**.**

1. Droite de régression de Y sur X

Méthode des moindres carrés : $\hat{Y} = aX + b$
Formules :

$$a = \frac{\overline{x_i y_i} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2}, \quad b = \bar{y} - a\bar{x}$$

Calculs intermédiaires

- Moyennes :

$$\bar{x} = \frac{20 + 25 + 30 + 35 + 40 + 50}{6} = \frac{200}{6} \approx 33.333$$
$$\bar{y} = \frac{4.2 + 5.1 + 5.9 + 6.8 + 7.4 + 9.1}{6} = \frac{38.5}{6} \approx 6.4167$$

- Tableau des écarts :

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
20	**4.2**	**-13.333**	**-2.2167**	**29.5556**
25	**5.1**	**-8.333**	**-1.3167**	**10.9722**
30	**5.9**	**-3.333**	**-0.5167**	**1.7222**
35	**6.8**	**1.667**	**0.3833**	**0.6389**
40	**7.4**	**6.667**	**0.9833**	**6.5556**
50	**9.1**	**16.667**	**2.6833**	**44.7222**
Σ				**94.1667**

$$\sum (x_i - \bar{x})^2 = (-13.333)^2 + (-8.333)^2 + (-3.333)^2 + (1.667)^2 + (6.667)^2 + (16.667)^2$$
$$= 177.778 + 69.444 + 11.111 + 2.778 + 44.444 + 277.778 = 583.333$$

Calcul de a et b

$$a = \frac{94.1667}{583.333} \approx 0.1614$$
$$b = \bar{y} - a\bar{x} = 6.4167 - 0.1614 \times 33.333 \approx 6.4167 - 5.3800 = 1.0367$$

Équation de la droite

$$\boxed{\hat{Y} = 0.1614X + 1.0367}$$

2. Coefficient de corrélation linéaire r

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Calcul de $\sum (y_i - \bar{y})^2$:

$$(-2.2167)^2 + (-1.3167)^2 + (-0.5167)^2 + (0.3833)^2 + (0.9833)^2 + (2.6833)^2$$

$$= 4.913 + 1.733 + 0.267 + 0.147 + 0.967 + 7.200 = 15.227$$

$$r = \frac{94.1667}{\sqrt{583.333 \times 15.227}} = \frac{94.1667}{\sqrt{8877.78}} \approx \frac{94.1667}{94.226} \approx 0.9994$$

$$\boxed{r \approx 0.9994}$$

3. Coefficient de détermination R^2

$$R^2 = r^2 = (0.9994)^2 \approx 0.9988$$

$$\boxed{R^2 \approx 0.9988}$$

4. Interprétation

Interprétation de a et b

- **Pente $a \approx 0.1614$: Pour chaque augmentation de 1 millier d'euros investi, le revenu annuel augmente en moyenne de 0.1614 millier d'euros (soit 161.4 €).**
- **Ordonnée à l'origine $b \approx 1.0367$: Si aucun investissement n'est réalisé ($X = 0$), le revenu annuel serait d'environ 1.0367 milliers d'euros. Cette valeur peut s'interpréter comme un revenu de base ou un effet fixe.**

Interprétation de r

Le coefficient de corrélation est très proche de 1, ce qui indique une relation linéaire positive très forte entre le montant investi et le revenu généré.

Interprétation de R^2

Le coefficient de détermination indique que 99.88% de la variabilité du revenu annuel est expliquée par le montant investi. Le modèle linéaire est donc très bien adapté aux données.

Exercice 2.

Données

Nombre total de mois : $N = 78 + 22 + 10 + 15 + 5 + 12 + 58 = 200$

Sinistres	0	1	2	3	4	5	6
Effectif	78	22	10	15	5	12	58
E.C.C	78	100	110	125	130	142	200

1. Médiane

Rang médian : $\frac{N+1}{2} = \frac{201}{2} = 100.5 \rightarrow$ **moyenne des valeurs aux rangs 100 et 101.**
Le rang 100 correspond à 1 sinistre, le rang 101 correspond à 2 sinistres.

$$\text{Médiane} = \frac{1+2}{2} = 1.5$$

$$\boxed{Me = 1.5 \text{ sinistres}}$$

2. Quartiles et mode

Quartile 1 Q_1

Rang : $\frac{n+1}{4} = \frac{201}{4} = 50.25 \rightarrow$ **valeur au rang 51.**
Le rang 51 se trouve dans la première classe (0 sinistre).

$$\boxed{Q_1 = 0}$$

Quartile 3 Q_3

Rang : $\frac{3(n+1)}{4} = \frac{603}{4} = 150.75 \rightarrow$ **valeur au rang 151.**
Le rang 151 se trouve dans la classe "6 sinistres".

$$\boxed{Q_3 = 6}$$

Mode

Classe ayant l'effectif le plus élevé : 78 (pour 0 sinistre).

$$\boxed{\text{Mode} = 0}$$

3. Interprétation

- **Médiane (1.5) : Dans 50% des mois, le nombre de sinistres est inférieur ou égal à 1.5.**
- **Quartiles :**
  - $Q_1 = 0$ **: 25% des mois ont 0 sinistre.**
  - $Q_3 = 6$ **: 75% des mois ont moins de 6 sinistres, mais la distribution est très asymétrique.**
- **Mode (0) : Le nombre de sinistres le plus fréquent est 0.**

Exercice 3.

Données

Classe (k€)	[2 ; 4[	[4 ; 6[	[6 ; 8[	[8 ; 12[	[12 ; 20[
Effectif	8	12	15	10	5
n_i^c	8	20	35	45	50

Total : $n = 50$

1. Médiane

Rang médian : $\frac{n}{2} = 25$ **(pour les données groupées).**
La médiane est dans la classe [6;8[.
Formule pour données groupées :

$$Me = b_i + \frac{\frac{n}{2} - n_{i-1}^c}{n_i^c - n_{i-}^c} \times a$$

Avec :

- $b_i = 6$ **(borne inférieure)**
- $\frac{n}{2} = 25$
- $n_{i-1}^c = 20$ **(fréquence cumulée avant la classe)**
- $n_i^c - n_{i-1}^c = 15$
- $a = 2$ **(amplitude)**

$$Me = 6 + \frac{25 - 20}{15} \times 2 = 6 + \frac{5}{15} \times 2 = 6 + 0.6667 = 6.6667$$

$$\boxed{Me \approx 6.67 \textbf{ k€}}$$

2. Quartiles et mode

Quartile 1 Q_1

Rang : $\frac{n}{4} = 12.5$
Classe contenant Q_1 **: [4;6[(fréquence cumulée : 20)**

$$Q_1 = 4 + \frac{12.5 - 8}{12} \times 2 = 4 + \frac{4.5}{12} \times 2 = 4 + 0.75 = 4.75$$

$$\boxed{Q_1 = 4.75 \textbf{ k€}}$$

Quartile 3 Q_3

Rang : $\frac{3n}{4} = 37.5$
Classe contenant Q_3 **: [8;12[(fréquence cumulée : 35, amplitude** $h = 4$**)**

$$Q_3 = 8 + \frac{37.5 - 35}{10} \times 4 = 8 + \frac{2.5}{10} \times 4 = 8 + 1 = 9$$

$$\boxed{Q_3 = 9 \textbf{ k€}}$$

Mode

Classe modale : celle avec l'effectif le plus élevé → [6;8[(effectif=15).
Formule de Czuber :

$$M_0 = L_1 + \frac{d_1}{d_1 + d_2} \times (L_2 - L_1)$$

Avec :

- $L_1 = 6$
- $d_1 = 15 - 12 = 3$
- $d_2 = 15 - 10 = 5$
- $L_2 - L_1 = 2$

$$M_0 = 6 + \frac{3}{3+5} \times 2 = 6 + \frac{3}{8} \times 2 = 6 + 0.75 = 6.75$$

$$\boxed{\textbf{Mode} \approx 6.75 \textbf{ k€}}$$

3. Interprétation

- **Médiane (6.67 k€) : 50% des factures sont inférieures ou égales à 6 670 €.**
- **Quartiles :**
  - $Q_1 = 4.75$ **k€ : 25% des factures sont inférieures à 4 750 €.**
  - $Q_3 = 9$ **k€ : 75% des factures sont inférieures à 9 000 €.**
  - **Écart interquartile :** $Q_3 - Q_1 = 9 - 4.75 = 4.25$ **k€, indiquant une dispersion modérée.**
- **Mode (6.75 k€) : Le montant le plus fréquent se situe autour de 6 750 €.**
- **Analyse globale : La distribution est centrée autour de 6 500-7 000 €, avec une certaine asymétrie vers la droite (quelques grosses factures jusqu'à 20 k€).**