DEVOIR SURVEILLÉ 2

Calculatrice autorisée

Lundi 18 novembre 2024

EXERCICE 1 (10 POINTS)

On considère la série statistique $(x; y)$ à deux variables suivante.

x_i	2	5	9	0	4
y_i	130	160	160	100	140

1. Donner l'effectif total n associé à cette série.

2. Calculer, **en détaillant**, les moyennes $\overline{x}$ et $\overline{y}$.

3. Calculer, **en détaillant**, les variances $\text{Var}(x)$ et $\text{Var}(y)$.

4. Calculer, **en détaillant**, la covariance $\text{Cov}(x; y)$.

5. Déterminer l'équation approchée (à 10^{-2}) de la droite de régression par les moindres carrés.

Détailler le calcul des coefficients.

6. L'ajustement affine par les moindres carrés est-il justifié ? Pourquoi ?

CORRECTION

1. $n = 5$

2.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2+5+9+0+4}{5} = 4$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{130+160+160+100+140}{5} = 138$$

3.

$$\begin{aligned}\text{Var}(x) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \\ &= \frac{(2-4)^2+(5-4)^2+(9-4)^2+(0-4)^2+(4-4)^2}{5} \\ &= 9{,}2\end{aligned}$$

$$\begin{aligned}\text{Var}(y) &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2 \\ &= \frac{(130-138)^2+(160-138)^2+(160-138)^2+(100-138)^2+(140-138)^2}{5} \\ &= 496\end{aligned}$$

4.

$$\begin{aligned}\text{Cov}(x;y) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \\ &= \frac{(2-4)(130-138)+(5-4)(160-138)+(9-4)(160-138)+(0-4)(100-138)+(4-4)(140-138)}{5} \\ &= 60\end{aligned}$$

5. L'ajustement affine par les moindres carrés admet pour équation : $y = ax + b$ avec :

- $a = \dfrac{\text{Cov}(x; y)}{\text{Var}(x)} = \dfrac{60}{9.2} \simeq 6{,}52$
- $b = \overline{y} - a\overline{x} = 138 - \dfrac{60}{9.2} \times 4 \simeq 111{,}91$.

Ainsi, on a : $y \simeq 6{,}52x + 111{,}91$

6. $r = \dfrac{\text{Cov}(x; y)}{\sigma(x)\sigma(y)} = \dfrac{60}{\sqrt{9.2}\sqrt{496}} \simeq 0{,}89$ est très proche de 1 donc il y a forte corrélation linéaire, l'ajustement affine est justifié.

EXERCICE 2 (10 POINTS)

Une entreprise a conçu un logiciel de gestion d'hôtels (gestion des réservations, planning du personnel, gestion des fournitures, etc.). Pour savoir à quel prix elle peut vendre son logiciel, l'entreprise a effectué une enquête auprès de 100 hôtels susceptibles de l'acheter.

Elle a résumé les résultats de l'enquête dans le tableau suivant, où p est la variable qui prend pour valeurs les prix de vente p_i (en €) et y la variable qui prend pour valeurs les nombres y_i d'hôtels qui acceptent d'acheter le logiciel au prix p_i.

p_i	600	650	700	750	800	850	900	990
y_i	76	70	65	61	55	49	45	35

1. Tracer, dans un repère orthogonal, le nuage de points de cette série statistique à deux variables p et y. On prendra comme unité 2 carreaux pour 100 € sur l'axe des abscisses et 2 carreaux pour 10 hôtels sur l'axe des ordonnées.

2. Déterminer les coordonnées du point moyen G et le placer sur le graphique.

3. **a.** Tracer un ajustement affine **au jugé**.

   b. En utilisant cet ajustement affine, déterminer le nombre d'hôtels prêts à payer le logiciel 680 €.

4. Afin d'être plus précis, l'entreprise veut déterminer, par la méthode des moindres carrés, l'équation de la droite de régression de y en p de ce nuage de points.

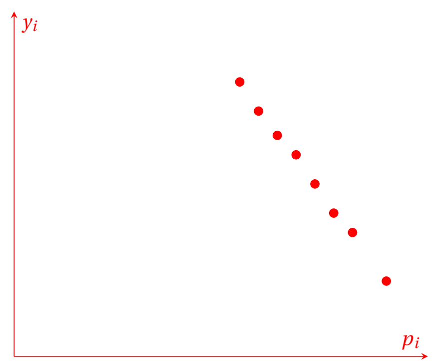
   a. Déterminer cette équation (approchée à 10^{-2}). On pourra utiliser la calculatrice.

   b. En utilisant cet ajustement affine, combien d'hôtels seraient prêts à payer le logiciel 680 € ?
   Comparer le résultat avec la question **3.b.**.

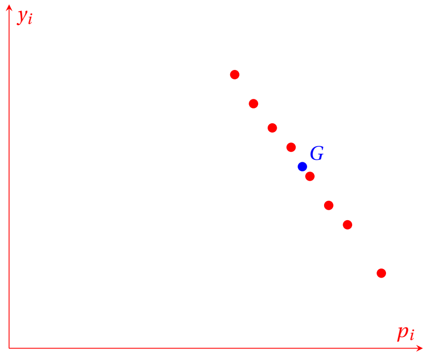
   c. Calculer le coefficient de corrélation linéaire entre y et p.
   Peut-on dire que les deux variables sont corrélées ? Avons-nous une relation de cause à effet entre y et p ?

CORRECTION

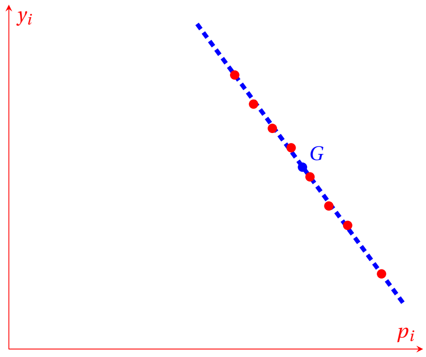
1.



2. G a pour coordonnées $(\overline{p}; \overline{y}) = (780; 57)$.



3. **a.** On peut tracer plusieurs ajustement affine, nous allons choisir ici la droite des points extrêmes : celle passant par les deux points aux abscisses extrémales.



b. Avec cet ajustement, on lit **graphiquement** que l'image de 680 est environ 68.

4. **a.** Par les moindres carrés, la droite de régression est : $y = -0{,}10p + 138{,}17$.

b. Via cet ajustement, pour $p = 680$ alors $y \simeq 67{,}41$ ce qui est très proche de notre résultat en **3.b.**.

c. $r = \dfrac{\text{Cov}(x; y)}{\sigma(x)\sigma(y)} \simeq -0{,}99$ est très proche de -1 donc il y a forte corrélation linéaire entre y et p. Pour autant, rien dans le contexte ne nous permet d'affirmer qu'il y a causalité entre y et p même si c'est probable.