Ajustement linéaire.

Éléments de correction du partiel mai 2024

Exercice 1

Une exploitation extrait un minerai rare dans des gisements depuis l'année 1968. Le tableau suivant indique la quantité extraite q_i en tonnes durant l'année désignée par son rang t_i :

Année	1968	1973	1995	1998	2003	2008	2013	2018	2023
t : Rang de l'année	0	1	2	3	4	5	6	7	8
q : Quantité extraite (en tonnes)	$15,7$	$13,3$	11	$9,3$	$7,8$	$7,1$	$6,1$	$5,2$	$4,3$

(On gardera 3 décimales dans les calculs de cet exercice)

1. En première approximation, on envisage de représenter q en fonction de t. Calculer le coefficient de corrélation linéaire entre q et t. Un ajustement linéaire est-il envisageable ? Expliquer.

   On a $r(q,t) = \frac{cov(q,t)}{\sigma(q)\sigma(t)}$. On calcule les quantités nécessaires et on obtient :

   - $\bar{q} \approx 8,867,\ \ var(q) \approx 13,033\ donc\ \sigma(q) \approx 3,61$
   - $\bar{t} = 4,\ \ var(t) \approx 6,667\ donc\ \sigma(t) \approx 2,582$
   - $\sum q_i t_i = 237,3\ puis\ cov(q,t) = -9,1$

   On obtient ainsi $\boxed{r(q,t) \approx -0,976}$, qui est assez proche de 1 en valeur absolue, donc un ajustement linéaire est envisageable.

2. Déterminez l'équation de l'ajustement linéaire de q en fonction de t.

   Par les formules du cours, on a $a = \frac{cov(q,t)}{var(t)} \approx -1,365\ et\ b = \bar{q} - a\bar{t} \approx 14,327$. Ainsi $\boxed{q(t) = -1,365t + 14,327}$.

3. En considérant cet ajustement affine, quelle quantité de minerai, au dixième de tonne près, l'exploitation peut-elle prévoir d'extraire durant l'année 2028 ?

   L'année 2028 correspond au rang 9. Ainsi, à l'aide de l'ajustement précédent, on aurait $\boxed{q(9) = -1,365 \times 9 + 14,327 \approx 2,042 \ \text{ tonnes}}$.

4. On souhaite maintenant savoir si il est raisonnable de penser que q et t puissent être liés par une relation du type $q \approx ae^{bt}$. Quelles variables (que l'on notera y et x) doit on considèrer pour répondre à cette question par un ajustement linéaire ?

   La relation $q = ae^{bt}$ se re écrit $\ln(q) = \ln(a) + bt$, qui est bien une relation affine du type $y = Ax + B$ en posant : $\boxed{y := \ln(q) \text{ et } x := t}$

5. Calculer le coefficient de correlation des ces 2 nouvelles variables, puis effectuer cet ajustement (ie : déterminez a et b.)

   On ajoute une ligne au tableau et on y calcule les $\ln(q)$. On calcule ensuite les quantités nécessaires à l'obtention du coeff de corrélation. On a :

- $\bar{y} = \overline{\ln(q)} \approx 2,1$ $var(y) \approx 0,166$ $donc$ $\sigma(y) \approx 0,407$
- $\bar{x} = 4,$ $var(x) \approx 6,667$ $donc$ $\sigma(x) \approx 2,582$
- $\sum x_i y_i = 66,150$ $puis$ $cov(\ln(q), t) = cov(y, x) = -1,05$

On obtient ainsi $r(x, y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)} = \boxed{r\Big(\ln(q), t\Big) \approx -0,998}$, qui est davantage proche de 1 en valeur absolue que le $r(q, t)$. Cet ajustement est donc de meilleure qualité.
Numériquement, on obtient : $\ln(q) = -0,157t + 2,793$. Soit $q(t) = e^{-0,157t+2,793}$ qui peut se mettre sous la forme : $\boxed{q(t) \approx 15,328\ e^{-0,157t}}$

6. Avec ce modèle, quel serait la prévision de la quantité de minerai extrait, au dixième de tonne près, pour l'année 2028 ? Comparez avec votre résultat de la question 3.

   Avec ce nouveau modèle, on a $\boxed{q(9) \approx 3,716 \text{ tonnes}}$

Exercice 2 (Les questions sont indépendantes)

1. Soit X et Y deux séries ayant chacune un effectif de 2, avec $x_1 \neq x_2$ et $y_1 \neq y_2$. Sans aucun calcul, que dire du coefficient de correlation ?

   2 points distincts étant alignés, $\boxed{\text{le coeff de corrélation vaudra donc 1 ou -1.}}$
2. Lors d'un ajustement linéaire de la variable y en fonction de la variable u, un étudiant a obtenu : $y = -3,2u + 5,1$ et un coefficient de corrélation $r(y, u) = 0,992$. Que pensez vous de son calcul ? Justifiez.

   Le coeff de corrélation et le coeff directeur de l'ajustement étant du même signe, on en déduit que $\boxed{\text{l'étudiant a commis une erreur.}}$
3. On suppose que la quantité demandée y (en tonnes) d'un certain article dépend du prix au kilos p à travers une relation du type $\frac{100}{y} = ap + b$. Peut on effectuer directement un ajustement linéaire entre y et p ? Si non, que proposeriez vous afin d'obtenir de réaliser l'ajustement proposé ?

   La relation proposée n'étant pas linéaire en les varibales y et p, on ne peut pas effectuer directement une ajustement linéaire. Cependant, à l'aide d'un $\boxed{\text{changement d'un variable}}$, on peut s'y ramener. En posant par exemple, $\boxed{z = \frac{100}{y} \text{ et en gardant la variable } p\text{, on obtient } z = ap + b.}$ On pourra donc effectuer un ajustement entre z et p.
4. Le responsable d'une chaîne de magasins de bricolage pense qu'il y a une relation entre le nombre de personnes qui s'installent dans la région et le chiffre d'affaire des magasins. Il a noté, pour chacune des 6 dernières années, la valeur n_k du nombre de personnes (en dizaines) ayant déménagé pour s'installer dans la région pendant l'année et c_k le chiffre d'affaire cumulé (en milliers d'euros) de l'ensemble des magasins de la chaîne lors de l'année. Le tableau suivant reproduit ces valeurs :

Année	2018	2019	2020	2021	2022	2023
n_k	$6,4$	$7,8$	$3,6$	$4,9$	$2,6$	$3,7$
c_k	32	$36,1$	$23,6$	26	$18,5$	$22,2$

Y a t'il une dépendance entre ces 2 deux variables selon vous ? De quelle nature ? Justifier votre réponse en calculant une certaine quantité et en la commentant.

La comparaison entre $\boxed{\text{la valeur absolue du coeff de corrélation } r(n_k, c_k) \text{ et } 1,}$ est le critère qui permettra de répondre à la question sur une dépendance linéaire. Après calculs, on obtient $r(n_k, c_k) \approx 0,994$.

5. Deux étudiants s'entrainent sur les ajustements de deux variables z et t. Le premier a réalisé un ajustement de z en fonction de t, il a obtenu $z = 2,44t + 1,5$, ainsi qu'un coefficient de correlation $r(z,t) = 0,96$. L'autre étudiant a réalisé l'ajustement de t en fonction de z. il a obtenu $t = 0,47z + 1,2$. L'un des deux s'est il trompé ?

   On calcule $\boxed{\text{le produit des coeff directeurs } aa' \text{ que l'on compare à } r^2.}$ On a : $aa' = 2,44 \times 0,47 = 1,1468 \neq 0,96^2$. Donc, un des étudiants s'est trompé !

6. Soit (y_t) une série chronologique de période 4. Précisez comment on calcule la tendance générale (g_t).

   La tendance générale se calcule à l'aide des $\boxed{\text{moyennes mobiles d'ordre 4.}}$