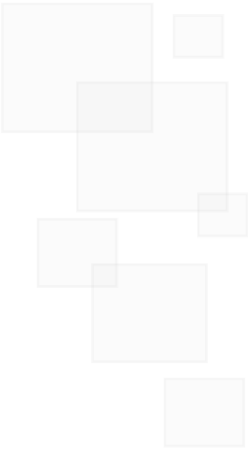


Corrélations et Régressions linéaires



Exemple d'article scientifique....



dentistry journal





Lien vers l'article:

<https://www.mdpi.com/2304-6767/11/9/205>

Article

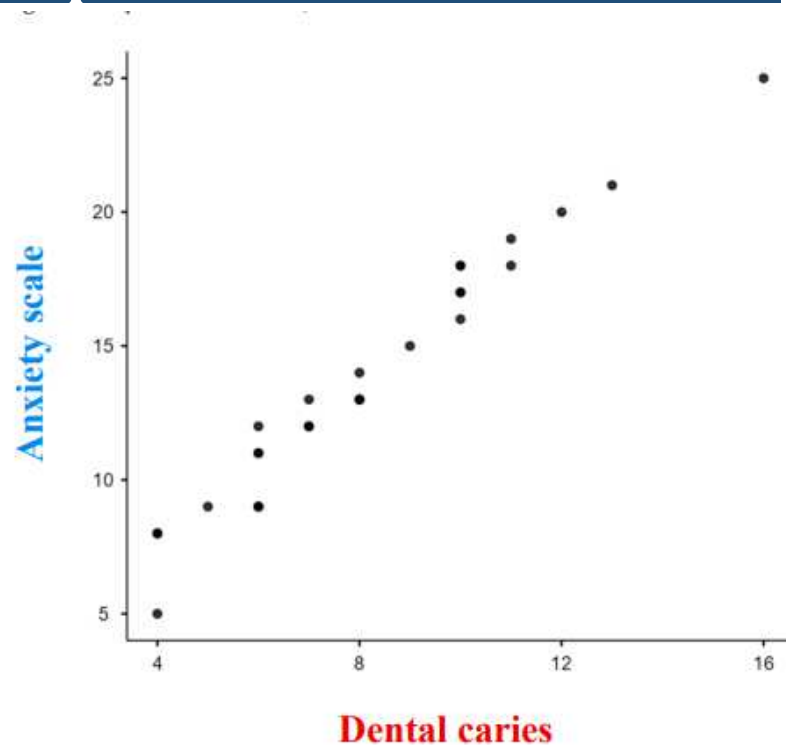
Association between Salivary Cortisol Levels, Dental Anxiety, and Dental Caries in Children: A Cross-Sectional Study

Vivek Padmanabhan ^{1,*} , Md Sofiqul Islam ¹ , Muneera Habib ¹, Zainab Abdulaziz ¹, Manjunatha Goud ², Nallan CSK Chaitanya ¹, Sheela Haridas ² and Muhammed Mustahsen Rahman ¹

Abstract:

Aim: This study aimed to investigate the **relationship between dental caries, dental anxiety, and salivary cortisol levels in children visiting pediatric dental clinics** and their implications on pediatric oral health. **Materials and Methods:** A cross-sectional study was conducted at a dental university in the UAE. **A total of 60 children, aged 4–12 years, were included. Salivary cortisol levels were measured using an Enzyme-Linked Immunosorbent Assay (ELISA) kit. Dental caries status was evaluated, and dental anxiety levels were assessed using the Modified Dental Anxiety Scale (MDAS).** Statistical analyses, including Mann-Whitney U test and Pearson's correlation coefficient, were performed to determine significant differences and associations.

Exemple d'article scientifique....



dentistry journal



Article

Association between Salivary Cortisol Levels, Dental Anxiety, and Dental Caries in Children: A Cross-Sectional Study

Vivek Padmanabhan ^{1,*}, Md Sofiqul Islam ¹, Muneera Habib ¹, Zainab Abdulaziz ¹, Manjunatha Goud ², Nallan CSK Chaitanya ¹, Sheela Haridas ² and Muhammed Mustahsen Rahman ¹

Figure 2. Correlation between dental caries and dental anxiety score. As dental caries increases; there is an increase in dental anxiety levels. There is a strong positive correlation (0.9803) which is statistically significant ($p \leq 0.0001$).

CORRELATIONS et REGRESSIONS

Formulation du problème

Etude de la **relation** ou **lien** entre **2 variables QUANTITATIVES (X et Y)**:

Exemple: La taille du cerveau et le niveau d'intelligence?

Taille du cerveau (X) avec des valeurs: x_1, x_2, \dots, x_n

Niveau d'intelligence (Y): y_1, y_2, \dots, y_n .

0. L' évaluation graphique de la relation :

- La diagramme de dispersion/nuage des points/scatter plot

1. L' existence d' une relation entre les variables X et Y

- A l' aide d'un test statistique sur le coefficient de corrélation

2. L' intensité / la force/l'importance/le degré de cette relation

- Le coefficient de corrélation de Pearson ou Spearman
- Le coefficient de détermination

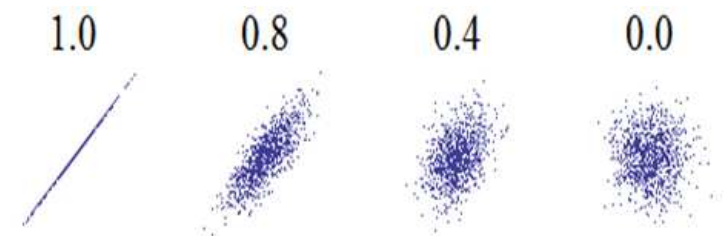
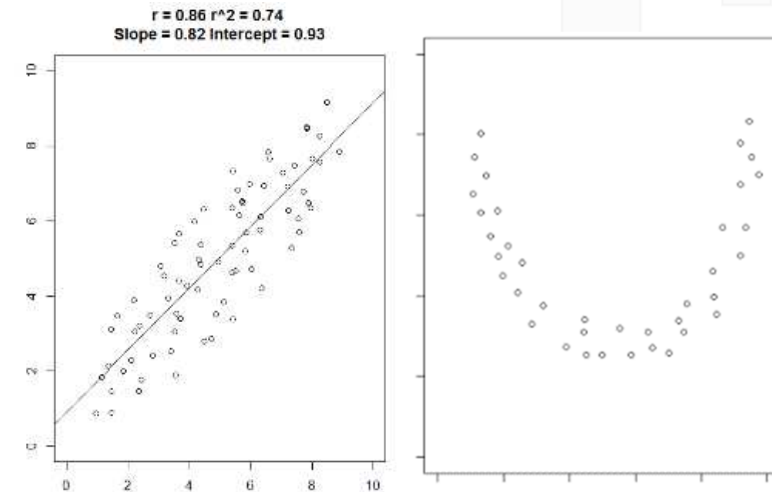
3. Prédiction : prédire les valeurs d'une variable sachant les valeurs de l'autre

- La régression: déterminer une fonction (fonction de régression) telle que $Y=f(X)$?

Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

Evaluer la linéarité, et l'importance de la corrélation:

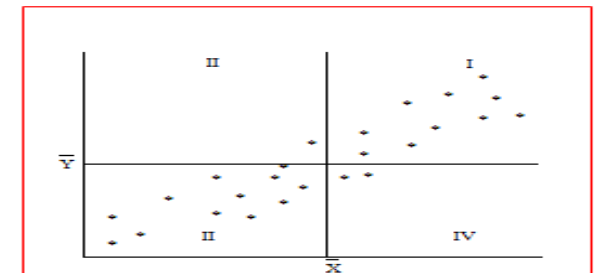
- Si les points semblent suggérer un droite – la relation est plus probable une relation linéaire
- Si les points semblent suggérer des tendances qui ne sont pas linéaires, la relation est peut être non linéaire (exponentielle, quadratique)
- Si la relation est plus probable linéaire, on peut évaluer d'une manière subjective la corrélation linéaire.
- Plus les points se rapprochent d'une droite de tendance, plus la corrélation est forte, plus les points sont distants, plus la corrélation est faible



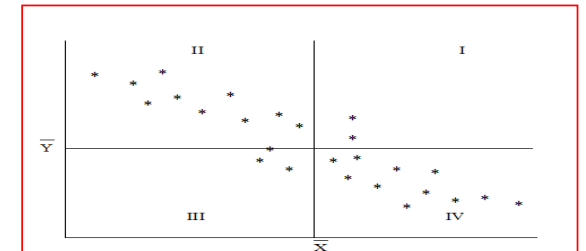
Evaluation graphique : diagramme de dispersion (nuage des points/scatter)

But : évaluation visuelle de la relation entre deux variables quantitatives:
l'utilisation des cadrans pour identifier **la tendance/sens/direction**
(lien de proportionnalité directe ou inverse)

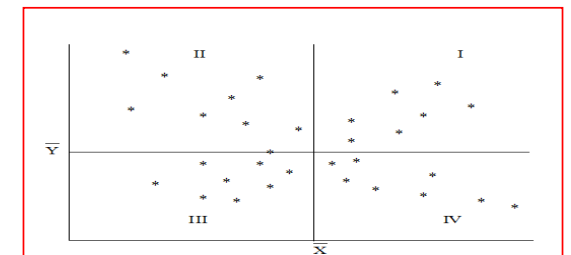
i) presque toutes les points sont dans les cadrans I et III ☐
tendance croissante/ pente ascendante/ pente positive/
lien de proportionnalité directe



ii) presque toutes les points sont dans les cadrans II et IV ☐
tendance décroissante / pente descendante/ pente
négative/ lien de proportionnalité inverse



iii) les points sont distribués uniformément dans tous les cadrans ☐ aucune tendance



Le coefficient de corrélation linéaire de Pearson

But: évaluer la corrélation des deux variables **quantitatives** du point de vue de la **direction** de l'association et l' **importance** de l' association

Le coefficient de corrélation Pearson:

- est une mesure de la force de **la corrélation (association) linéaire** des deux variables quantitatives
- montre le degré de rapprochement des points a une droite qui passe entre les points.

Condition d' application:

- Les paires des observations indépendantes dans l'échantillon
- Variables quantitatives
- Les deux variables soient normalement distribuées
(les variables suivent une distribution normale bi variée)
- La relation entre les deux variables est linéaire simple (pas quadratique, exponentielle)

Le coefficient de corrélation linéaire de Pearson

Le calcul du coefficient de corrélation linéaire Pearson:

Covariance d'échantillonnage $COV(X,Y)$:

$$COV(X,Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$$

Coefficient de corrélation linéaire Pearson

$$r = \frac{COV(X,Y)}{S_X \cdot S_Y}$$

Coefficient de détermination $d = r^2$

x_i, y_i = les valeurs des deux séries des données. \bar{X} et \bar{Y} sont les moyennes des deux séries. n = nombre des observations. S_x et S_y sont les déviations standard d'échantillonnage, r = coefficient de corrélation Pearson, d – coefficient de détermination

Corrélation linéaire de Pearson: Interprétations

Interprétation de la direction/sens/tendance:

$r > 0$ tendance croissante/ pente ascendante/ lien de proportionnalité directe/ corrélation positive

$r < 0$ tendance décroissante/ pente descendante/ lien de proportionnalité inverse/ corrélation négative

$r \simeq 0 \Rightarrow$ aucune tendance **linéaire**

- plus **r** est grand (en valeur absolue) plus la relation est forte pendant que plus **r** est proche du 0, plus la relation est faible

Corrélation linéaire de Pearson: interprétations

Interprétation de l'intensité/degré/importance de la corrélation linéaire avec les règles empiriques de Colton [*Colton T. Statistics in Medicine. Little Brown and Company, New York, NY 1974*]

Si r se situe dans l'intervalle $]-0.25; 0.25[$

=> une corrélation **négligeable** ou **aucune** corrélation linéaire entre les variables
 $[0.25 \text{ et } 0.50[$ ou $[-0.25 \text{ et } -0.50[$

=> un degré de corrélation **faible/acceptable**
 $[0.50 \text{ et } 0.75[$ ou $[-0.50 \text{ et } -0.75[$

=> un degré de corrélation **modérée à bonne**
 $[0.75 \text{ et } 1]$ ou $[-0.75 \text{ et } -1]$ => une **très bonne à excellente** corrélation

Il y a autre divisions possibles aussi.

!!! Ces règles doit être utilisée avec soins.

Elle sont pour donner une idée, mais pour chaque problème, l'intensité de la relation est relative au domaine de recherche. Pour certain situations les valeurs en dessous de 0,8 peut être faibles.

Test de significativité statistique pour le coefficient de Pearson

- **But:** tester si entre deux variables **quantitatives** il y a une corrélation **linéaire** statistiquement significative
- On utilise un test statistique sur le coefficient de corrélation pour voir s' il est différent de 0 (absence de la corrélation H_0)
- On peut toujours calculer un coefficient de corrélation mais il faut vérifier certaines conditions si on veut tester s' il est statistiquement significatif.
- **Condition d' application:**
 - les paires des observations indépendantes dans l'échantillon
 - variables quantitatives
 - les deux variables soient normalement distribuées
 - *(les variables suit un distribution normale bi variée)*
 - sans valeurs aberrantes (très éloignées du nuage des points)

Test de significativité statistique pour le coefficient de Pearson

- Etape 1.** $H_0 : r=0$; (il **n'y a pas** de **corrélation linéaire statistiquement significative** entre les variables X et Y **sur la population**)
 $H_1 : r \neq 0$ (il **y a une** **corrélation linéaire statistiquement significative** entre les variables X et Y **sur la population**)

Etape 2 . Statistique du test:
$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

suit la loi de « t » avec « n-2 » degrés de liberté; n=taille de l' échantillon

Etape 3. $\alpha = 0,05$

Etape 4. Région du rejet $RR = (-\infty, -t_{n-2, \alpha/2}] \cup [t_{n-2, \alpha/2}, \infty)$

Etape 5. Calcul du t

Etape 6. Si t est contenu dans la region du rejet RR on rejette H_0 , alors on prends H_1
Si $p < 0,05$ on rejette H_0 , alors on prends H_1

Calcul du coef. de corrélation Pearson

- Les valeurs du HDL Cholestérol (mmol/l) et les poids (kg) pour 10 malades sont:
Les deux variables sont quantitatives, paires des observations indépendantes, normale distribuées. La relation est linéaire (le graphique Scatter)
- On a les moyennes de cholestérol et de poids $\bar{x} = 5.24, \bar{y} = 74.9$ et la variance de cholestérol = 1,15, la variance de poids = 214,89 et par conséquence les écarts types $s_x = 1,07, s_y = 14,65$.
- La covariance sur la population est $\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$
- $\text{COV}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) \times (Y_i - \bar{Y})$
- $\text{COV}(x, y) = (1/9) \times [(6.8-5.24) \times (90-74.9) + (5.3-5.24) \times (75-74.9) + \dots]$

$$r = \frac{\text{COV}(x, y)}{s_x \times s_y} = \frac{14.324}{1.07 \times 14.65} = 0.91$$

Interprétation :

$r > 0 \Rightarrow$ tendance croissante / relation (directe) proportionnelle, $r > 0,75 \Rightarrow$ intensité de corrélation très bonne (règles du Colton)

HDL	6.8	5.3	4.3	5.0	7.1	5.5	3.8	4.6	4.0	6.0
Poids	90	75	70	73	110	67	60	65	59	80

Test statistique de signifiante pour le coefficient de corrélation linéaire du Pearson

Exemple

Le coefficient de corrélation pour la relation entre les triglycérides et le poids pour 50 sujets est 0,72, et la valeur du p associée est 0,001. Les paires des observations sont indépendantes, les données sont normalement distribuées, la relation est linéaire

Interprétation de la valeur du p associée au coefficient de corrélation:

=> $p < 0,05$ il y a une **corrélation statistiquement significative** entre les **triglycérides** et le **poids**

La **corrélation** entre les **triglycérides** et le **poids** est **statistiquement significative**

Interprétation de la direction et de l'intensité de la corrélation: dans l'échantillon d'étude, la relation est (directe) proportionnelle ($r = 0,72 > 0$), et l'intensité de la corrélation est modérée à bonne ($r = 0,72$ – est compris dans l'intervalle $[0,50; 0,75]$)

Test statistique de signifiante pour le coefficient de corrélation linéaire du Pearson

Exemple

Le coefficient de corrélation pour la relation entre les triglycérides et la longueur du cheveux pour 24 sujets est **0,39**, et la valeur du **p associée est 0,35**. Les paires des observations sont indépendantes, les données sont normalement distribuées, la relation est linéaire

Interprétation de la valeur du p associée au coefficient de corrélation:

=> $p > 0,05$ on ne peut pas dire qu'il **ya** une **CORRÉLATION LINÉAIRE statistiquement significative** entre les triglycérides et la longueur du cheveux

Interprétation de la direction et de l'intensité de la corrélation: l'interprétation n'a aucune utilité parce que la relation n'a pas atteint la signification statistique – donc le résultat peut être beaucoup influencée par la chance. Pour le moment on sait rien. La relation semble une relation de proportionnalité directe ($r=0,35 > 0$), et l'intensité de la corrélation est faible/acceptable ($r=0,35$ – est dans $[0,25 \text{ et } 0,50]$) – si la relation était statistiquement significative.

Coefficient de corrélation **de Spearman**

But: évaluer la corrélation entre des deux variables du point de vue de la **direction** et l' **importance**

Condition d' application:

- Les paires des observations indépendantes dans l'échantillon
- Les deux variables sont ordinales/quantitatives (non normalement distribuées)

Utilité: évaluer la relation entre

- *deux variables ordinales*
- *une ordinale et une quantitative*
- *deux variables quantitatives* qui ne sont pas normalement distribuées

Coefficient de corrélation de Spearman (r_s)

Interprétation:

$r_s > 0 \Rightarrow$ une corrélation positive

$r_s < 0 \Rightarrow$ une corrélation négative

$r_s \simeq 0 \Rightarrow$ aucune corrélation

- plus r_s est grand (en valeur absolue) plus la relation est forte
- plus r_s est proche du 0, plus la corrélation est faible

Revision des coefficients de correlations

Type des variables	Nature des données	Coefficient de corrélations	Formule du coefficient
quantitative	normale distribuées	Pearson (r)	$COV(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$ $r = \frac{COV(X, Y)}{S_X \cdot S_Y}$
quantitative	pas normalement distribuées	Spearman (ρ - rho)	$\rho = r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ ou } d_i = R_{x_i} - R_{y_i}$
qualitative ordinales	-	Spearman (ρ - rho)	$\rho = r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ ou } d_i = R_{x_i} - R_{y_i}$

X_i, Y_i – sont les valeurs des deux séries des données. \bar{X} et \bar{Y} sont les moyennes des deux séries. R_{x_i} et R_{y_i} sont les rangs des valeurs X_i et Y_i après leurs rangement dans ordre croissante. n – nombre des observations. S_x et S_y sont les déviations standard d'échantillonnage. $COV(X, Y)$ = la covariance

Analyse de Régression

But: -méthode pour étudier les relations entre 2 ou plus variables

Type de régression: Selon

- **la linéarité de la fonction**
 - régression linéaire
 - régression non linéaire
- **le nombre de variables dépendantes:**
 - régression univariée (une variable dépendante)
 - régression multivariée (≥ 2 variables dépendantes)
- **le nombre de variables indépendantes:**
 - régression simple (une variable indépendante)
 - régression multiple (≥ 2 variables indépendantes)

Régression linéaire simple

But étudier la **dépendance linéaire** entre deux variables quantitatives

Objectifs:

- prédire les valeurs d' une variable dépendante en fonction des valeurs de l' autre (considéré comme indépendante)
- évaluer l'existence, la direction et l' importance de la dépendance entre deux variables

Conditions d'application:

- observations indépendants
- variables quantitatives
- existence d'une relation linéaire entre les deux variables
- erreurs normalement distribués (résidus)
- homoscédasticité - variance constante des erreurs (résidus)

Régression linéaire simple - calcul

$Y = a_0 + a_1 X + \varepsilon$ ou $a_0, a_1 =$ coefficients de la régression

- Méthode de calcul des coefficients a_0, a_1 : méthode des moindres carrés
- critère de la méthode: minimisé la somme de tous les carrés de la distance du chaque point par rapport à la droite:

$$\min \sum_{i=1}^n (Y_i^{predite} - Y_i)^2 = a_0 + a_1 X_i$$

-on obtient:

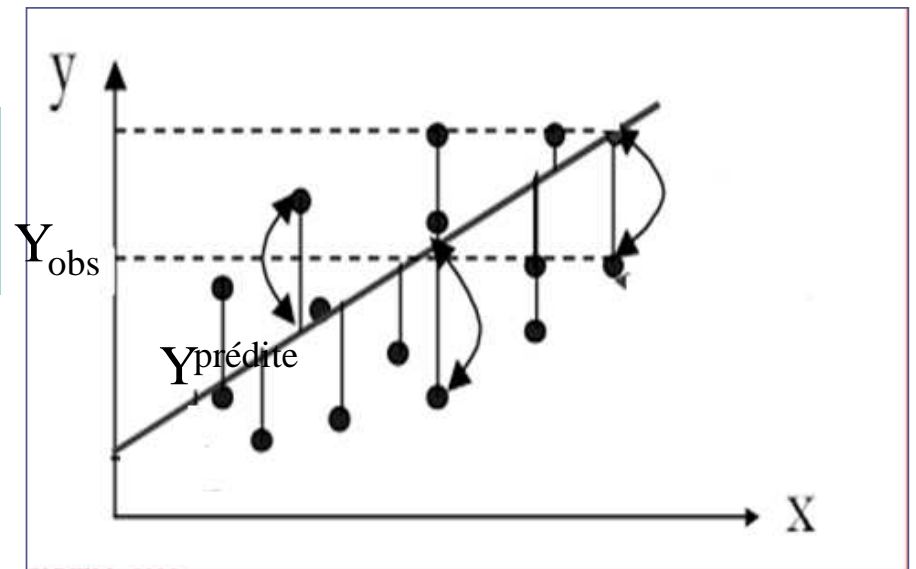
$$a_1 = \frac{COV(X, Y)}{S_X^2}$$

$$a_0 = \bar{Y} - a_1 \bar{X}$$

covariance

$Y_{obs} =$
Valeur
observée
Y

Moyenne de la variable Y, moyenne du X



Régression linéaire simple: interprétations

- La droite de régression $Y(X)$: $Y = a_0 + a_1 X$

a_0 = est l'ordonnée à l'origine

a_1 = la pente de la droite de régression.

Interprétation de a_1 : pour chaque changement d'une unité de mesure de la variable X , la variable Y va être modifiée avec a_1

- **Si $a_1 > 0$** tendance croissante/ pente ascendante/ pente positive/ lien direct proportionnel
- **Si $a_1 < 0$** tendance décroissante/ pente descendante/ pente négative/ lien inversement proportionnel
- **Si $a_1 \simeq 0$** \Rightarrow aucune tendance
- plus a_1 est grand (en valeur absolue) plus la relation de dépendance linéaire est forte
- plus a_1 est proche du 0, plus la relation de dépendance est faible

Le coefficient de détermination

Variance totale de la variable dépendante =

- Variance donné par la droite (variance de régression) + variance NON donnée par la droite (variance résiduelle)

$$r^2 = \frac{\text{Variance}_{\text{régression}}}{\text{Variance}_{\text{totale}}}$$

- Nommée : **coefficient de détermination**
- Notation: **d = r²** (le carrée du coefficient de corrélation Pearson)
- nombre de 0 à 1. On le transforme dans pourcentage pour l'interpréter (0 → 100%)

Interprétation: pourcentage de la variance (variation) de la variable dépendante expliqué par sa relation linéaire avec la variable indépendante

- plus d est grand plus la relation/association/liens est forte
- plus d est proche du 0, plus la relation est faible

Le coefficient de détermination

Exemple :

On a fait une régression entre $X = \text{Poids (kg)}$ et $Y = \text{HDL Cholestérol (mg/dl)}$.

La relation entre les deux est linéaire, les variables sont normale distribuées, les observations sont indépendantes, les erreurs sont normalement distribués, et homoscédastiques.

On a trouve le coefficient de détermination = 0,82.

Interprétation du coefficient de détermination:

$d = 0,82 \rightarrow$ **82% de la variance (variation) du cholestérol peut être explique par la relation linéaire avec le Poids.** Le reste de 18% est due a des autres facteurs.

“

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.

— John Tukey, *Exploratory Data Analysis*

”

**Merci de votre
attention!**



Exemples des questions pour l'examen

E1. Une étude a évalué la corrélation entre le poids (kg) et la circonférence abdominale (cm), et ils ont trouvé la valeur du coefficient égale a **0,5**, avec la valeur du P du test associée de **0,04**. Les conditions d'application du coefficient sont satisfaites. Les deux variables sont normales distribuées.

Parmi les affirmations suivantes, indiquez lesquelles sont correctes:

- A. Pour évaluer la corrélation entre les deux variable, on utilise le coefficient de corrélation de Pearson
- B. Pour évaluer la corrélation entre les deux variable, on utilise le coefficient de corrélation de Spearman
- C. Il y a une corrélation linéaire statistiquement significative entre le poids (kg) et la circonférence abdominale (cm) car $p < 0,04$
- D. Il y a une corrélation linéaire statistiquement significative entre le poids (kg) et la circonférence abdominale (cm) car la valeur du coefficient était égale a **0,5**
- E. Dans l'échantillon d'étude Il y a une corrélation linéaire positive entre le poids (kg) et la circonférence abdominale (cm) car la valeur du coefficient était égale a **0,5**

R1: A, C, E

Exemples des questions pour l'examen

E2. L'équation de régression obtenu après l'analyse des données collectées dans un étude a été ($Y=a_1*X + a_0$): **la pression artérielle diastolique PAD (mmHg) = 14,2 * la quantité de sel mange par jour (g) + 61,5.**

Parmi les affirmations suivantes, indiquez lesquelles sont correctes:

- A. la pente de la régression est égale a 14,2
- B. la pente de la régression est égale a 61,5
- C. une augmentation par 1 gramme/jour de la quantité de sel mangé va augmenter la PAD par 14,2 mmHg
- D. une augmentation par 1 gramme/jour de la quantité de sel mangé va diminuer la PAD par 14,2 mmHg
- E. L'équation de régression donnée est le modelé d'une régression linéaire simple

R2: A, C, E

Exemples des questions pour l'examen

E3. L'équation de régression de la longueur du corps de la mandibule (mm) = la hauteur du branche (mm) * 0,8 + 23,2. Le coefficient de détermination = 0,49.

Parmi les affirmations suivantes, indiquez lesquelles sont correctes:

- A. le coefficient de corrélation linéaire est positif
- B. pour une augmentation de la hauteur du branche avec 1mm, la longueur du corps de la mandibule va augmenter avec 0,8
- C. la relation entre les deux variables est inversement proportionnelle
- D. la corrélation linéaire est acceptable
- E. 49 % de la variation de la longueur du corps de la mandibule peut être expliquée par sa relation linéaire avec la hauteur de la branche

R3 : A, B, E

Exemples des questions pour l'examen

E4. L'équation de régression de la tension artérielle systolique (mmHg) = la épaisseur de la paroi ventriculaire (mm) * 2.8 + 84.5. Le coefficient de corrélation Pearson = 0,63. Lesquelles des réponses suivantes sont correctes :

- A. la covariance est négative
- B. pour chaque millimètre en plus de épaisseur de la paroi ventriculaire, la tension artérielle systolique augmente (en moyenne) avec 2,8 mmHg
- C. la relation entre les deux variables est inversement proportionnelle.
- D. la corrélation lineaire est moyenne vers bonne
- E. 84,5 mmHg est l'ordonnée à l'origine

R4: B, D, E

Exemples des questions pour l'examen

E5. L'équation de régression de la tension artérielle systolique (mmHg) = la épaisseur de la paroi ventriculaire (mm) * 2,7 + 79,2. Le coefficient de détermination = 0,36. Lesquelles des réponses suivantes sont correctes :

- A. le coefficient de corrélation est < 0
- B. pour chaque millimètre en plus de épaisseur de la paroi ventriculaire, la tension artérielle systolique augmente (en moyenne) avec 2,7 mmHg
- C. Entre les deux variables il y a une relation de proportionnalité directe
- D. la corrélation est modérée à bonne
- E. 36 % de la variation de la tension artérielle systolique peut être expliquée par sa relation linéaire avec l' épaisseur de la paroi ventriculaire

R5: B, C, D ,E

Exemples des questions pour l'examen

E6. Le poids et le taux du cholestérol des 60 malades ont été observées. Les suivantes statistiques ont été calculées. Pour le poids: coefficient d'asymétrie=0.28, coefficient d'aplatissement=0,7, coefficient de variation = 0,40. Pour le cholestérol: coefficient d'asymétrie=2,2, coefficient d'aplatissement=0,5.

Le coefficient de corrélation Pearson entre le poids et le cholestérol est: 0,30.

La valeur p pour le coefficient de corrélation est <0,01.

Lesquelles des réponses suivantes sont correctes :

- A. le poids est relativement hétérogène
- B. on doit calculer le coefficient de corrélation Spearman
- C. le coefficient de détermination est 0,09
- D. la distribution du cholestérol a une queue a la gauche
- E. 40 % de la variance (variation) du poids peut être expliquée par la relation linéaire avec le cholestérol

R6: B, c

Exemples des questions pour l'examen

E7. Lesquelles des réponses suivantes sont correctes en ce qui concerne la relation entre deux variables montrée dans le graphique à cote:

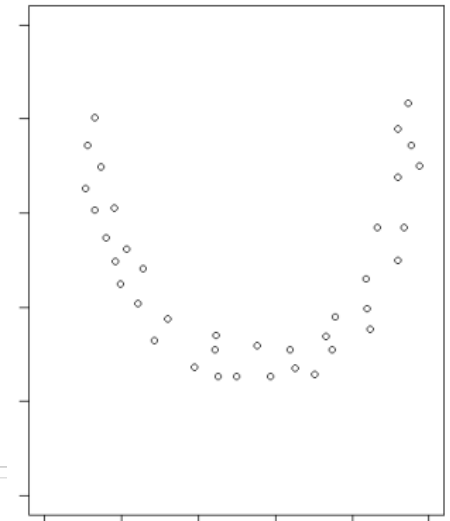
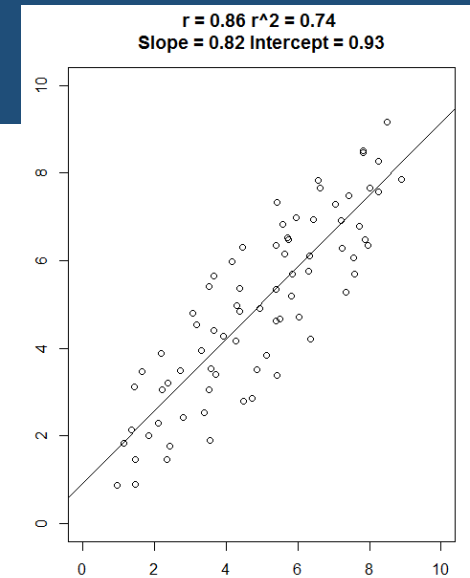
- A. la relation semble être linéaire
- B. le coefficient estimé était le coefficient de corrélation de Pearson
- C. entre les variables il y a une relation de proportionnalité directe
- D. entre les variables il y a une relation de proportionnalité inverse
- E. le pente de la droite de regression est négative

R7: A, B, C

E8. Lesquelles des réponses suivantes sont correctes en ce qui concerne la relation entre deux variables montrée dans le graphique à cote:

- A. la relation semble être non linéaire
- B. le coefficient de corrélation de Pearson ne sera utile pour quantifier la puissance de la relation
- C. entre les variables il y a une relation de proportionnalité directe
- D. entre les variables il y a une relation de proportionnalité inverse
- E. la pente de la droite est négative

R8: A, B



Exemples des questions pour l'examen

E9. Lesquelles des réponses suivantes sont correctes en ce qui concerne la relation entre deux variables montrée dans le graphique a cote:

- A. la relation entre les deux variables est plus forte dans le graphique B que pour le graphique A
- B. le coefficient de corrélation de Pearson pour le graphique C est égale a -1
- C. entre les variables il n'y a pas de relation de proportionnalité
- D. entre les variables il y a une relation de proportionnalité inverse
- E. le coefficient de corrélation Pearson pour le graphique B est plus grand en valeur absolue que pour le graphique A

R9: A, B, D, E

A.



B.



C.



“

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.

— John Tukey, *Exploratory Data Analysis*

”

**Merci de votre
attention!**

