Université Claude Bernard Lyon 1

Tutorat Lyon Est

Unité d'Enseignement 3

BANQUE DE QCM

2013 - 2022

Corrélation - Régression

QUESTIONS - REPONSES

Question 1 :

Après les blessures d'un nombre important de ses joueurs, l'Olympique Lyonnais aimerait identifier des facteurs associés.
En 2014, une étude a été réalisée sur 102 joueurs du club à propos du lien entre la consommation X du complément alimentaire Gourcuffine© et la concentration sanguine de créatine kinase, témoin d'une atteinte musculaire.

On obtient : $\rho_{X,Y} = 0.3$; $\sigma_X{}^2 = 2.4$; $\sigma_Y{}^2 = 2.55$; $\sigma_X \sigma_Y \approx 2$

Aide aux calculs : $\frac{2.55}{2.4} \approx 1.0625$; $0.25^2 = 0.0625$

L'équation de la droite de régression est : y = b0 + b1x + e

A. $cov_{X,Y} \approx - 0.6$
B. $b_1 = 0.25$
C. On ne peut pas calculer b_0.
D. $b_0 = 0.3$
E. La droite de régression a une pente non significativement différente de 0.

Question 1 : BC

A FAUX On connaît la formule : $p_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\ \sigma_Y}$ ⇔ $cov_{X,Y} = \rho_{X,Y} * \sigma_X * \sigma_Y = 0.3 * 2 \approx 0.6$

MAIS ATTENTION AU SIGNE – (piège fréquent) dans l'énoncé

B VRAI On connaît la formule : $b_1 = \frac{\sigma_{X,Y}}{\sigma^2{}_X} = \frac{0,6}{2,4} = 0.25$

C VRAI

D FAUX Pour pouvoir calculer b_0, il faut connaître les moyennes de Y et de X puisque $b_0 = \bar{Y} - \bar{X}$ (cf diapo 29 du cours). Or on ne dispose pas de ces informations.

E FAUX On teste la pente à 0, ce qui revient à tester la significativité de la relation entre X et Y :
H0 : b1=0

t = $\frac{b_1}{\sigma_{b_1}}$ avec $\sigma_{b_1} = \sqrt{\frac{\frac{s_Y{}^2}{s_X{}^2} - b_1{}^2}{n-2}} = \sqrt{\frac{1.0625-0.0625}{100}} = \frac{\sqrt{1}}{\sqrt{100}} = 0.1$ (voir aides aux calculs)

Donc t = $\frac{b_1}{\sigma_{b_1}} = \frac{0.25}{0.1} = 2.5$

On lit dans la table de Student à n-2 ddl ⇔ à 100 ddl et pour un risque α à 5% : qui nous donne la valeur 1.9840.
$t^{n-2} > 1.984$ donc on peut rejeter l'hypothèse nulle H_0, on conclut que la pente de la droite de régression est significativement différente de 0.

Enoncé commun aux questions 2 et 3 :

On souhaite diminuer l'incidence des accidents cardio-vasculaires. On teste l'efficacité d'une nouvelle statine sur le taux sanguin de LDL. Notre échantillon comprend 36 sujets, tous issus de la population cible. Le taux sanguin moyen de LDL des sujets ayant reçus cette statine 1 fois/jour est estime à 1.6 g/L

avec un écart type estime de 0.2g/L. Dans le même temps, la concentration sanguine moyenne en troponine, témoin d'une atteinte cardiaque, est de 2 mg/L avec un écart-type de 0.1 mg/L.
Dans la population cible, la fréquence d'un taux sanguin de LDL compris entre 1.4g/L et 1.8g/L est de 85%.
Soient X, Y et F trois variables aléatoires. X correspond au taux sanguin moyen de LDL des sujets ayant reçu cette statine une fois par jour, Y a la concentration sanguine moyenne en troponine et F à la proportion de sujets ayant un taux de LDL compris entre 1.4g/L et 1.8g/L.

Données supplémentaires : E(XY) = 3.2005 ; s_p= 0.36

Aides au calcul : 0.85 * 0.15 = 0.1275 et $\sqrt{\frac{0.1275}{36}} = 0.06$

Pour un risque alpha égal à 5% et 1.96 ≈ 2

Question 2 :

On souhaite maintenant s'assurer qu'une baisse du taux sanguin de LDL est corrélée à une baisse du taux sanguin de troponine. Elle représente en effet une baisse de la prévalence des accidents cardiovasculaires.
On suppose ici que X et Y suivent approximativement une loi normale respectivement de paramètres $N(\mu_X ; \sigma_X)$ et $N(\mu_Y ; \sigma_Y)$.

A. cov(X,Y) = 0.05
B. $\rho_{X,Y} = 0.1$
C. D'après le coefficient de corrélation de Pearson, qui est paramétrique, X et Y semblent très corrélés.
D. La valeur du test du coefficient de corrélation linéaire est égale à 2.5.
E. On rejette l'hypothèse nulle avec un risque de rejeter à tort H_0 égal à 1%.

Question 2 : CD

A FAUX On connaît la formule pour calculer cov(X,Y) = $\sigma_{X,Y}$ = E(XY) – E(X) E(Y).
On a E(XY) = 3.2005
Pour trouver E(X) et E(Y), on sait que n= 36>30 ⇔ on peut affirmer que X et Y suivent approximativement une loi normale de paramètre $(\mu; \frac{\sigma}{\sqrt{n}})$ donc respectivement : $N(1.6 ; \frac{0,2}{6})$ pour X et $N(2 ; \frac{0,1}{6})$ pour Y.
/!\ Pour cet énoncé commun, il manquait de précision. L'écart-type donné ne correspondait pas à l'écart-type de la moyenne du taux sanguin de LDL, mais uniquement à l'écart-type du taux sanguin. La différence était donc entre une variable aléatoire correspondant au taux sanguin et une variable aléatoire correspond à la moyenne de ce taux. Il en était de même pour l'autre variable. On vient maintenant au calcul : on utilise le théorème centré limite. On doit donc diviser notre écart-type (du taux sanguin) par racine(n) pour avoir l'écart-type de la moyenne de ce taux. C'est dans le polycopié page 61 et 75.

Donc μ_X = E(X) et μ_Y = E(Y)⇔ E(X) = 1.6 et E(Y) = 2
On remplace donc dans la formule : cov(X,Y) = 3.2005 – 1.6*2 = 3.2005 – 3.2 = 0.0005

B FAUX

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} = \frac{0.0005}{\frac{0.2}{6} * \frac{0.1}{6}} = \frac{0.0005}{0.02} * 36 = \frac{0.05}{2} * 36 = 0.9$$

C VRAI On a un coefficient de corrélation très proche de 1 ($\rho_{X,Y}$ = 0.9) donc X et Y sont très corrélés.

D VRAI On sait que X et Y suivent tout deux une loi normale, on peut donc effectuer un test du coefficient de corrélation linéaire.

Pour ce test, la formule que l'on utilise classiquement est $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

Or il faut bien comprendre que cette formule est à l'origine $t = \frac{|r-p|}{\sqrt{\sigma^2{}_p}} = \frac{|r|}{\sqrt{\sigma^2{}_p}}$ (car p=0 sous H_0).

Ici, nous avons les valeurs pour calculer avec la formule « classique » : cela donne $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}9\sqrt{34}}{\sqrt{0{,}19}}$

mais il n'est pas évident de simplifier le calcul.

De plus, on nous donne s_p, qui nous permet d'estimer σ_p : il est donc beaucoup plus rapide d'utiliser la formule originelle.

Ainsi $t = \frac{|r|}{\sqrt{\sigma^2{}_p}} = \frac{|r|}{\sigma_p} = \frac{0{,}9}{0{,}36} = \frac{9}{3{,}6} = \frac{9}{9*0{,}4} = \frac{1}{0{,}4} = 2{,}5$

[ATTENTION ! Ici, le(a) tuteur(trice) qui a fait cet énoncé a voulu que vous calculiez la statistique de test avec le σ_p donné dans l'énoncé. Toutefois, si vous calculez cette statistique avec la formule « habituelle » vous ne tombez pas sur le même résultat… En fait ce qu'il se passe ici c'est que le (a) tuteur(tirce) a donné le **σp** pour que vous l'utilisiez et que vous calculiez la valeur de t en utilisant la formule : **t=r/σp**

Il s'agit d'un **ERRATA**, normalement les deux formules devraient donner le **MEME** résultat, ici c'est une erreur de la part du tuteur qui a voulu simplifier les calculs sans vérifier que la valeur de **σp** valait bien **σp = (1-r^2)^0,5 / (n-2)^0,5**

Toutefois, le but ici est de vous faire utiliser le **σp** car sinon les calculs sont impossibles (trop longs) de tête. C'est une erreur car il n'y a pas concordance entre les 2 résultats. On vous laisse la question car on pouvait tout à faut y répondre avec les données de l'énoncé.]

E FAUX On fait un test bilatéral : on cherche à démontrer une différence (≠ unilatéral : on présompte le sens de cette différence).
On cherche dans la table de Student, pour n-2ddl = 34ddl, la valeur de α correspondant à t = 2.5.
Elle se situe entre t = 2.4411 et t = 2.7284 donc pour α compris entre 0.02 et 0.01.
On peut donc rejeter l'hypothèse nulle avec un risque alpha (= probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie) égal à 2%, mais pas à 1%.

Question 3 :

Enfin on aimerait certifier que l'augmentation de la troponine sanguine est bien due à une augmentation de LDL sanguin.
Soit l'équation de la régression : $Y = \beta_0 + \beta_1 x + e$, estimée dans l'échantillon par $y = b_0 + b_1 x + e$

On a : $SCE_R = 0{,}0015$ et $SCE_E = 0{,}0063$

Aides au calcul : $\frac{63}{78} \approx 0{,}81$; $\frac{15}{78} \approx 0{,}19$; $\frac{15}{63} \approx 0{,}24$; $\frac{8}{150} \approx 0{,}05$

A. Le coefficient de régression $b_1 = 0{,}025$.
B. L'hypothèse nulle du coefficient de régression est $b_1 = 0$.
C. Le coefficient de détermination $\rho^2 = 0{,}80$
D. La variation de X explique 20 % de la variabilité de Y.
E. Ces résultats sont cohérents avec ceux des questions précédentes.

Question 3 : BCE

A FAUX $b_1 = \frac{\sigma_{X,Y}}{\sigma^2{}_X} = \frac{5.10^{-4}}{10^{-3}} = 5.10^{-1} = 0.5$

Car : $\sigma^2{}_X = (\frac{0,2}{6})^2 = \frac{0,04}{36} = \frac{0,01}{9} \approx \frac{0,01}{10} \approx 10^{-3}$

B VRAI Sous H_0, il n'y a pas de relation entre X et Y, donc la droite de régression n'est pas linéaire donc le coefficient directeur de la droite b_1 est égal à 0.

C VRAI On a : $SCE_T = SCE_R + SCE_E = 0,0015 + 0,0063 = 0,0078$

$\rho^2 = \frac{SCE_E}{SCE_T} = \frac{0.0063}{0,0078} = \frac{63}{78} = 0,8$

D FAUX Par définition, la variation de X explique une fraction ρ^2 de la variabilité de Y. Ainsi, dans notre exercice, la variation de X explique une fraction de 0.8 de la variabilité de Y : elle explique donc 80 % de la variabilité de Y.

E VRAI En fait, le coefficient de détermination est le coefficient de corrélation au carré. Nous pouvions donc répondre à la question C et D à partir des résultats du QCM 4B.

Ici on retrouve bien l'égalité $\rho^2 = \rho_{X,Y} * \rho_{X,Y} = 0.81 = 0.9 * 0.9$

Question 4 :

L'UCLA (Université de Californie à Los Angeles) a étudié la relation entre le nombre d'heures de méditation de type mindfulness (pleine conscience) et le nombre de gyrifications du cortex cérébral. La gyrification représente le degré de plissement du cerveau, caractéristique d'une plus ou moins grande surface corticale. Le protocole de l'étude prévoit d'utiliser un test du coefficient de corrélation de Spearman avec un risque alpha de 0.5%. Le nombre d'heure de méditation de type mindfulness et le nombre de gyrifications du cortex cérébral sont recueillis chez 11 patients. On note X le nombre d'heures de méditation et Y le nombre de gyrifications du cortex cérébral.

On donne : $\sum_i {d_i}^2 = 50$

Aide au calcul : $\frac{300}{1320} \approx \frac{300}{1200}$; $\frac{50}{1080} \approx \frac{50}{1000}$; $\sqrt{1-0.25^2} \approx 1$; $\sqrt{1-0.75^2} \approx 0.6$

A. Le coefficient de corrélation de Spearman est estimé à 0.25.
B. La statistique du test du coefficient de corrélation de Spearman vaut 3.75.
C. La statistique du test du coefficient de corrélation de Spearman est à comparer à la valeur seuil dans la table de Student au risque alpha 0.5% et à 11ddl.
D. Le coefficient de corrélation est significativement diffèrent de 0 au risque alpha 0.5%.
E. Plus le nombre d'heures de méditation est important, plus le nombre de gyrifications du cortex cérébral est important, et vice versa.

Question 4 : BDE

A FAUX *Conditions d'application :* n=11>10

$$r_s = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)} = 1 - \frac{6*50}{11(121-1)} = 1 - \frac{300}{1320} \approx 1 - \frac{300}{1200} \approx \mathbf{1} - 0.25 = 0.75$$

B VRAI

$$t_s = \frac{|r_s|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.75}{\sqrt{\frac{1-0.75^2}{9}}} = \frac{0.75}{\frac{\sqrt{1-0.75^2}}{\sqrt{9}}} = \frac{0.75}{\frac{0.6}{3}} = \frac{0.75}{0.2} = 0.75 * 5 = 3.75$$

C FAUX La statistique du test du coefficient de corrélation de Spearman est à comparer à la valeur seuil dans la table de Student au risque alpha 0.5% et **n-2ddl = 11-2 ddl = 9ddl.**

D VRAI En lisant la valeur seuil dans la table de Student à la colonne p = 0.05 et à la ligne 9ddl, on obtient t_{seuil} = 3.6897 or t_s = 3.75>t_{seuil} donc le coefficient de corrélation de Spearman est significativement différent au risque alpha de 0.5%.

E VRAI Le coefficient de corrélation de Sperman r_s est égal à 0.75 > 0 donc plus le nombre d'heures de méditation est important, plus le nombre de gyrifications du cortex cérébral est important. La réciproque est vraie parce qu'on affaire au coefficient de corrélation et non au coefficient de régression.

Question 5 :

D'après ≪ the relationship between cognitiv anxiety and sports performances on basketball ≫, Parnabas and co.

Une équipe de chercheurs de l'université de Shah Alam en Malaisie travaille sur les relations stressperformance.

Elle aimerait connaitre la relation entre le stress et les performances au basketball. Pour cela, elle cherche à savoir si l'augmentation de la concentration sanguine en cortisol, une hormone du stress, est corrélée à la baisse du nombre de points inscrits lors d'un match de basketball.

Les données sont recueillies chez 102 patients. Le coefficient de corrélation de Pearson entre les 2 variables est estimé à -0.70 (p<0.05).

On note X la concentration sanguine en cortisol en mg/L et Y le nombre de points inscrits par match. La régression linéaire de Y en fonction de X s'écrit $Y = b_0 + b_1 * X + \varepsilon$.

On donne : $\sigma_Y = \sigma_X = 1$; $b_0 = 25$; ε suit une distribution centrée sur 0

Aide au calcul : $\sqrt{0.51} \approx \sqrt{0.5}$; $\sqrt{2} \approx 1.5$

A. La statistique du test du coefficient de corrélation vaut environ – 10.5.
B. Lorsque la concentration sanguine de cortisol diminue, le nombre de points inscrit augmente.
C. Le coefficient de régression b_1= - 0.7
D. Dans ce cas, l'ordonnée à l'origine a un sens.
E. Le nombre de points inscrits par match pour une concentration sanguine de cortisol de 10mg/L est d'en moyenne 18.

Question 5 : ABCDE

A VRAI $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0,7\sqrt{100}}{\sqrt{1-0,49}} = \frac{-0,7*10}{\sqrt{0,5}} = \frac{-7}{\sqrt{\frac{1}{2}}} \approx -7*\sqrt{2} \approx -7*1.5 \approx -10.5$

B VRAI Le coefficient de corrélation de Pearson = -0.7 donc les 2 variables X et Y évoluent en sens inverse.

C VRAI On a $r_{X,Y} = b_1 * \frac{\sigma_X}{\sigma_Y} \Leftrightarrow b_1 = r_{X,Y} * \frac{\sigma_Y}{\sigma_X} = -0.7 * 1 = -0.7$

D VRAI Vrai, elle correspond au nombre de points marqués pour une concentration sanguine de cortisol nulle. (En pratique, ce n'est pas possible qu'elle soit nulle, mais vous n'êtes pas censés le savoir.)

E VRAI On nous donne la valeur de X : X = 10, on remplace dans l'équation :

$\hat{Y} = b_0 + b_1 * X = 25 + (-0.7 * 10) + 0 = 25 - 7 = 18$

Question 6 :

Une équipe de recherche travaille sur les facteurs pronostiques des cancers du sein. Une de leurs hypothèses est que l'exposition aux pesticides sur une longue durée est associée à une plus grande agressivité des cancers mesurée par la taille de la plus grosse métastase au diagnostic. Pour tester leur hypothèse ils soumettent à un échantillon de 36 souris des pesticides pendant des durées d'exposition variables.

On pose : Y : « taille de la métastase » en mm et X : « temps d'exposition » en années

On écrit l'équation de la régression de la taille de la plus grosse métastase sur le temps d'exposition Y = a + b*X

On donne Cov(X,Y)=3 $S^2x=1.5$ Sb1=0.5

A. Le coefficient de régression est significativement différent de 0 au risque alpha=5%
B. Le coefficient de corrélation n'est pas significativement différent de 0 au risque alpha=5%
C. La taille de la plus grosse métastase est plus grosse de 2mm à chaque année d'exposition supplémentaire.
D. On trouve une valeur du test de Student de t=4
E. Il y a une corrélation statistiquement significative entre la durée d'exposition aux pesticides et la taille de la métastase.

Question 6 : ACDE

A VRAI Comme vu ci-dessous (item D), le test est significatif, on peut donc conclure que le coefficient de régression est significativement différent de 0.

B FAUX Le test du coefficient de régression est significatif au risque alpha=5%, or tester le coefficient de régression revient à tester le coefficient de corrélation, ainsi le coefficient de corrélation est significativement différent de 0.

C VRAI En ce qui concerne l'item C on a la formule :

On a $b_1 = \frac{\sigma_{X,Y}}{\sigma^2{}_X} = \frac{3}{1,5} = 2$ donc la taille de la métastase la plus grosse est plus grosse de 2 mm à chaque année d'exposition. C'est ce que représente b1, la pente de la droite de régression.

D VRAI $t = \frac{b_1}{Sb_1} = \frac{2}{0,5} = 4\ t_{34\ ddl}$

Dans la table de Student pour alpha=5% et 34 ddl on trouve une valeur $t_{seuil} = 2,0322$ et $t_{calculé} > t_{seuil}$ Donc on rejette l'hypothèse nulle d'une pente nulle : la pente liant les deux variables est significativement différente de 0.

E VRAI On peut bien conclure à une liaison linéaire entre les deux variables.

Question 7 :

La société française de Cardiologie a publié une étude estimant que le coefficient de corrélation de Pearson entre la quantité de soda sucré absorbée quotidiennement et le diamètre intraluminal de l'artère carotidienne, est de 0.90. L'écart-type associé est de Sr=0.090. Cette étude a été menée sur un échantillon de 40 patients. Les variables de cette étude sont des variables aléatoires gaussiennes.

A. Le test du coefficient de corrélation de Pearson est à 38ddl.
B. Le coefficient de corrélation est significativement différent de 0 au risque alpha=1%
C. Le coefficient de détermination est d'environ 0.81.
D. La diminution du diamètre intraluminal de l'artère carotidienne est d'autant plus grande que l'absorption quotidienne de soda sucré est importante.
E. Tous les items sont vrais.

Question 7 : ABC

A VRAI Le test est un test coefficient de corrélation de Pearson via un test de student à n-2ddl soit 40-2=38 ddl

B VRAI On a comme formule : t = $\frac{r}{Sr} = \frac{0,9}{0,09} = 10$

Dans la table de student pour alpha=1% et 38 ddl on lit $t_{seuil} = 2,7116$ donc, la valeur calculée est supérieure à la valeur seuil, on rejette alors l'hypothèse nulle au risque alpha=1%.

C VRAI Le coefficient de détermination est égal au carré du coefficient de corrélation soit : $0,9^2=0,81$.

D FAUX R >0 (0,9 dans l'énoncé) donc les deux variables varient dans le même sens. Plus on consomme de soda, plus le diamètre augmente.

(*NB : ceci est faux physiologiquement parlant → le diamètre diminue normalement mais ici on est en UE4 et pas en physio ;)*)

E FAUX

Question 8 :

Le cœur est un organe noble, dont l'importance pour notre santé est capitale. Une étude a été menée sur 400 sujets masculins pour détecter un lien entre la longueur du cœur à l'âge adulte et l'espérance de vie. La longueur du cœur a été mesurée par échocardiographie et est exprimée en cm (on s'intéresse à la longueur entre l'oreillette gauche et l'apex). L'espérance de vie est exprimée en année. Au sein de cet échantillon, la longueur moyenne du cœur est de 12 cm et l'espérance de vie moyenne est de 75 ans. On observe un écart-type de 2 cm sur la longueur du cœur et un écart-type de 10 ans sur l'espérance de vie. La covariance des variables aléatoires modélisant la longueur du cœur et l'espérance de vie est égale à 2 cm.an. De plus, les deux variables aléatoires modélisant la longueur du cœur et l'espérance de vie suivent une loi normale.

Aide aux calculs : $\frac{\sqrt{398}}{\sqrt{0.99}} \approx 20 \;\; ; \;\; \frac{\sqrt{399}}{\sqrt{1}} \approx 21$

A. L'estimation du coefficient de corrélation de Pearson est égale à 0.1.
B. La statistique de test de ce coefficient de corrélation de Pearson est de 2.
C. La statistique de test du coefficient de corrélation de Pearson suit une loi de Student à 400 degrés de liberté.
D. On rejette l'hypothèse nulle d'un coefficient de corrélation de Pearson égal à 0 au risque alpha=5%.
E. On peut affirmer qu'une augmentation de la longueur du cœur de 1 cm augmente l'espérance de vie d'une demi-année.

Question 8 : ABDE

A VRAI On a $r_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \, \sigma_Y} = \frac{2}{2*10} = 0,1$

B VRAI Pour tester notre coefficient de corrélation, c'est-à-dire pour calculer la probabilité que notre coefficient soit égal à 0 mais que la différence qu'on observe sur cet échantillon soit due à l'inférence statistique, il faut calculer un paramètre que l'on va comparer à un seuil.

On a $t = \frac{|r| \times \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{|0.1| \times \sqrt{398}}{\sqrt{0.99}}$. Grace à l'aide aux calculs, on t = 2.

C FAUX On compare ensuite cette valeur dans la table de la loi de Student, en prenant 398 dans l'idéal mais +l'infini ici.

D VRAI On voit que $t_{calculée} > t_{seuil}$ pour α=5%, t $_{seuil}$ =1,96. Donc la probabilité que l'hypothèse nulle selon laquelle r=0 et que les différences observées soit dues au hasard est plus faible que 0,05. On peut donc rejeter cette hypothèse nulle et affirmer qu'il y a une corrélation entre la longueur du cœur et l'espérance de vie.

E VRAI On peut affirmer qu'une augmentation de la longueur du cœur de 1 cm augmente l'espérance de vie d'une demi-année.
Si l'estimation du coefficient de corrélation de Pearson nous donne une idée de la force de l'association (plus il est proche de 1 ou -1, plus l'association est forte), seul le calcul de b_1 va pouvoir quantifier de combien l'espérance de vie augmente en fonction de la longueur du cœur.

On a $b_1 = \frac{\sigma_{X,Y}}{\sigma^2_X} = \frac{2}{2*2} = 0{,}5.$

Donc il semblerait que qu'une augmentation de 1 cm de la longueur du cœur entraine une augmentation de l'espérance de vie d'une demi année. Maintenant, il faut encore une fois tester cette hypothèse et calculer la probabilité qu'en réalité b_1 soit égal à 0 et que la différence observée soit due au hasard. Cependant, tester le coefficient de corrélation de Pearson est équivalent à tester ce coefficient b_1 et on peut donc tout de suite affirmer qu'une augmentation de la longueur du cœur entraine significativement une augmentation de l'espérance de vie d'une demi-année.

Question 9 :

On étudie la corrélation entre la consommation de protoxyde d'azote chez 27 jeunes adultes et les résultats scolaires. L'hypothèse est qu'il existe une relation entre le nombre de cartouches consommées par semaine et les résultats au baccalauréat.

Soit X la variable aléatoire associée au nombre de cartouches consommées par semaine, et Y la variable aléatoire associée à la moyenne au baccalauréat. Les distributions des 2 variables sont supposées normales.

On obtient pour cet échantillon : $\mathrm{var(X)} = 1.125$ $\quad \mathrm{var(Y)} = 4$ $\quad \mathrm{cox(X,Y)} = -0.85$

Aide aux calculs : $\frac{0.85}{\sqrt{4.5}} \approx 0.4$ $\quad \frac{2}{\sqrt{0.84}} \approx 2.2$ $\quad \frac{0.85}{4.5} \approx 0.2$ $\quad \frac{1}{\sqrt{0.96}} \approx 1.02$

A. $\rho_{X,Y} \approx 0.4$
B. Le seuil de décision pour le test du coefficient est lu dans la table de Student à 26 degrés de liberté.
C. Au risque $\alpha = 0.01$, on peut rejeter l'hypothèse nulle de l'absence de corrélation entre les deux variables aléatoires.
D. Au risque $\alpha = 0.05$, on conclut que la consommation de protoxyde d'azote réduit la moyenne au bac.
E. La covariance permet d'évaluer le sens de variation des 2 variables aléatoires l'une par rapport à l'autre.

Question 9 : DE

A FAUX On calcule le coefficient de corrélation de Pearson $\rho_{X,Y}$, dont la formule est à savoir

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X) \times var(Y)}} = \frac{-0.85}{\sqrt{4 \times 1.125}} = \frac{-0.85}{\sqrt{4.5}} \approx -0.4$$

ATTENTION AU SIGNE!

B FAUX On lit dans la table de student à n-2 degrés de liberté soit 27-2=25 ddl.

C FAUX On doit calculer la statistique du test de corrélation à partir de l'estimation du coefficient de corrélation :

$$t_{25ddl} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0.4\sqrt{27-2}}{\sqrt{1-(-0.4)^2}} = \frac{-2}{\sqrt{0.84}} \approx -2.2$$

On lit dans la table de Student à 25 ddl.

$$2.0595 < |-2.2| < 2.4851$$

$$0.02 < P(|T| > -2.2) < 0.05$$

$P(|T| > -2.2) > 0.01$ donc on ne peut pas rejeter l'hypothèse nulle d'une absence de corrélation entre X et Y au risque $\alpha = 1\%$.

D VRAI Le test permet de rejeter H0 au risque 5% et le signe de la corrélation permet de conclure sur le sens de l'association donc si la consommation de protoxyde d'azote augmente, la moyenne au bac diminue et inversement.

E VRAI La covariance permet d'évaluer le sens de la variation entre deux variables aléatoires. Si elle est négative, les 2 variables varient en sens inverse ; si elle est positive, les 2 variables varient dans le même sens.

Question 10 :

Un chercheur en diabétologie étudie la dépendance entre la moyenne de la glycémie sur 2 mois, et le taux d'hémoglobine glyquée par rapport à l'hémoglobine totale. On appelle G la variable aléatoire de la glycémie moyenne, et H la variable aléatoire de l'hémoglobine glyquée. Il obtient après son étude réalisée sur un échantillon de 1000 personnes, une équation de régression : $h_i = 2g_i + 1 + e_i$.

A. Cov(G,H) = 2 var(H)
B. ei tend vers 0.
C. p² = somme $(h_i - \bar{h})^2/(\hat{h}_i - \bar{h})^2$ est la part de la variance de H expliquée par G.
D. La droite d'équation $y_i = 2x_i + 1$ est appelée droite de régression.
E. Si on étudie sur un échantillon de personnes les relations entre H et G, on obtiendra des valeurs qui ne correspondront pas parfaitement au modèle.

Question 10 : BDE

A FAUX Par définition, on sait que si y_i= b0 + b1 x_i + ei, alors b1 = cov (X, Y)/Var(X)
Donc adaptée aux notations et données de notre énoncé, ça donne : 2 = cov(G, H) / Var(G). Donc on a cov(G, H) = 2 Var(G).

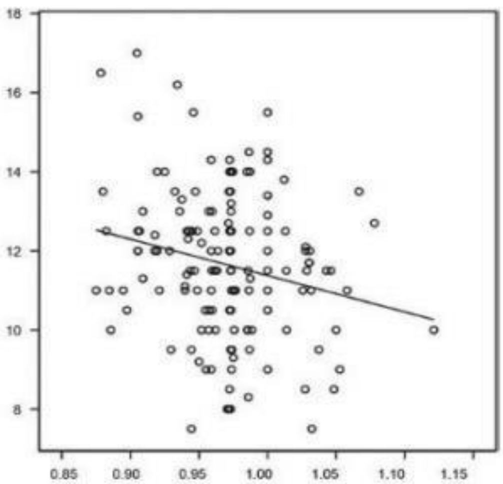
B VRAI En effet, dans le modèle de régression, $e_i$représente l'erreur potentielle inhérente au modèle. Et pour avoir un modèle le plus juste et précis possible on essaie de faire tendre cette erreur vers 0.

C FAUX C'est l'inverse ! Bien retenir que $\widehat{hi}$ est l'estimation de hi par le modèle de régression. Ici la partie expliquée du modèle est au dénominateur, on a donc le pourcentage de "non-expliqué".

D VRAI La définition de la droite de régression est en effet d'approximer la relation entre deux variables par le modèle, on peut donc reprendre l'équation de régression, en enlevant l'erreur car on est dans le cas théorique (donc pas d'erreur).

E VRAI Le principe d'un modèle de dépendance entre deux variables aléatoires est malheureusement de contenir des erreurs dans la vie réelle, car chaque personne est unique et il peut exister des variations inter- et intra-individus importantes. Il existe donc une multitude de couples de valeurs qui ne correspondent pas au modèle, mais la moyenne du lien entre les deux variables restent approximable par la droite.

Question 11 :



Un zoologiste passionné de chiens cherche à voir s'il existe un lien entre la taille de la queue et le rapport de la taille des pattes postérieures sur la taille des pattes antérieures (leg ratio). Pour cela il réalise une étude observationnelle sur 144 braques.

Soit X la variable aléatoire représentant le leg ratio. Soit Y la variable aléatoire représentant la taille de la queue du chien.

Il estime la droite de régression suivante dont l'équation est y = -9x + 20.6

On a $s_x = 0.04$ et $s_y = 1.6$ cm

Aides aux calculs :

$\frac{0.225\sqrt{144-2}}{\sqrt{1-0.225^2}} \approx 2.7$ $\quad \frac{9\times4}{16} \approx 2.25$ $\quad \frac{9\times16}{4} = 36$

A. Le coefficient de corrélation entre la leg ratio et la taille de la queue est estimé à 0.225.
B. Le coefficient de régression de la taille de la queue en fonction du leg ratio est estimé à -9.
C. Le coefficient de corrélation est significativement différent de 0 au risque α = 5%.
D. $covar_{X,Y} < 0$.
E. Pour x = 1.2 on a $\hat{y}$ = 9.8 cm.

Question 11 : BCD

A FAUX On a $r = b_1 \frac{s_X}{s_Y}$ = - 9 x $\frac{0,04}{1,6}$ = - 0,225

B VRAI Le coefficient de régression est $b_1 = -9$

C VRAI On calcule la statistique du test $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = t = \frac{-0,225\sqrt{144-2}}{\sqrt{1-0,225^2}} \approx$ - 2.75

On lit dans la table de Student à 144-2=142 ddl et on trouve $2.576 \leq |t| \leq 2.8072$ et on trouve $0.005 \leq P(|T| \geq t) \leq 0.01$

On peut dire que le coefficient de corrélation est significativement différent de 0 au risque alpha = 5%.

D VRAI Le coefficient de corrélation r est négatif donc la covariance l'est forcément aussi.

E FAUX Les valeurs observées ne vont pas au-delà de 1.15 (cf graphique), on ne peut donc pas donner d'estimation de y quand x > 1.15 car on ne sait pas si le modèle reste valide en dehors de la zone des valeurs observées.

Question 12 :

Les flavonoïdes, de puissants antioxydants que l'on trouve en grande quantité dans de nombreux aliments dérivés des plantes et notamment dans les fèves de cacao, ont montré qu'ils étaient susceptibles d'avoir un impact positif sur les fonctions cognitives. Un chercheur s'est donc demandé s'il pouvait exister une certaine corrélation entre la consommation de cacao dans un pays et les capacités mentales de ses habitants.

Pour cela, on pose X, la variable aléatoire gaussienne représentant la consommation de chocolat en kg/an/personne et Y, la variable aléatoire gaussienne symbolisant le nombre de lauréats au Prix Nobel pour 10 millions d'habitants. Cette étude a été réalisée sur 38 pays.

Données : $\sqrt{38} \approx 6.16$; $\sqrt{37} \approx 6.08$; $r_{X,Y} = 0.8$

A. Le test du coefficient de corrélation de Pearson, non paramétrique, est à 37 ddl.
B. Pour un risque α=5%, notre valeur t_{seuil} est 2.9905.
C. Le coefficient de corrélation de Pearson est significativement différent de 0 au risque α=1%.
D. Le coefficient de détermination est de 0.6.
E. $cov_{X,Y}$ est positive. Il semblerait que plus nous mangeons de chocolat, plus nous avons de chance d'être lauréat.

Question 12 : CE

A FAUX Le test du coefficient de corrélation de Pearson est un test paramétrique effectué via un test de Student à n-2 ddl soit 36 ddl ici.

B FAUX On a X et Y des variables aléatoires suivant chacune une loi Normale. On pose l'hypothèse nulle H_0 : $\rho = 0$, ρ étant le coefficient de corrélation de Pearson et H_1 : l'hypothèse alternative : $\rho \neq 0$. Dans l'énoncé, on nous donne r=0.8, l'estimateur du coefficient de corrélation de Pearson calculé à partir de notre étude. On cherche à savoir via ce test si notre coefficient de corrélation est significativement différent de 0, et donc si l'on pourrait conclure à une corrélation entre la consommation de chocolat et le nombre de lauréats au Prix Nobel.

On effectue donc un test sous H_0, avec la formule : $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

Puis nous comparerons notre valeur test t calculée à un $t_{seuil\ n\text{-}2ddl,\alpha}$ lu dans la table de Student. Par exemple ici, pour α=5% et n-2 ddl = 36 ddl : $t_{seuil\ 36\ ddl,0.05} = 2{,}0281$.

C VRAI Nous allons à maintenant calculer notre valeur t :

$$t = \frac{0{,}8\sqrt{38-2}}{\sqrt{1-0{,}8^2}} = \frac{0{,}8*6}{\sqrt{0{,}36}} = \frac{0{,}8*6}{0{,}6} = 8$$

Pour α=1%, $t_{seuil\ 36\ ddl,0.01}$ = 2,7195. $|t| > t_{seuil\ 36ddl,0.01}$ car 8 ≥ 2,7195, donc nous rejetons H_0 et nous pouvons conclure que le coefficient de corrélation de Pearson est significativement différent de 0 au risque α=1%.

D FAUX Le coefficient de détermination est égal au carré du coefficient de corrélation soit : $0{,}8^2 = 0{,}64$.

E VRAI Le signe de $p_{X,Y}$ est le signe de $cov_{X,Y}$ (ou $\sigma_{X,Y}$), donc $cov_{X,Y}$ est bien positive. Il semblerait effectivement que plus nous mangeons de chocolat, plus nous avons de lauréat dans le pays.

Question 13 :

Les bienfaits du chocolat, le retour. La publication sur la corrélation observée entre la quantité de chocolat consommée dans un pays et le nombre de ses lauréats du Nobel ayant eu un grand retentissement dans la presse généraliste, deux autres chercheurs ont tenté de découvrir quelles surprises cette gourmandise nous réservait-elle encore. A l'issue de leurs travaux, ils ont pu trouver un coefficient de corrélation r=0,52 (p<0,02) entre la quantité de chocolat consommée en kg/an/habitant dans un pays (notée CC) et le nombre de tueurs en série originaires de ce pays par million d'habitants depuis 1900 (noté SK). Une droite de régression linéaire a été construite, d'équation : SK=0,5CC+1.

A. La statistique de test du coefficient de régression est à comparer à une valeur seuil lue dans la table de distribution de la loi de Student au risque α et à n-2 degrés de liberté.
B. Le coefficient de régression est significativement différent de 0 au risque α=0,05.
C. Lorsque CC augmente d'une unité, SK augmente de 1 pour un million d'habitants.
D. Plus la quantité de chocolat consommée augmente, plus le nombre de tueurs en série diminue au sein d'un pays grâce aux bienfaits du chocolat.
E. Il manque des données dans l'énoncé pour répondre à l'item B.

Question 13 : AB

A VRAI

B VRAI Tester le coefficient de régression revient à tester le coefficient de corrélation. D'après l'énoncé, on nous donne un coefficient de corrélation r=0,52 avec un niveau de significativité petit p inférieur à 0,02. L'hypothèse nulle H_0 étant : $\rho = 0$ et l'hypothèse alternative H_1 étant : $\rho \neq 0$. On rejette donc l'hypothèse nulle au risque α=0,05 car p< α. Le coefficient de corrélation étant significativement différent de 0, on peut donc conclure que le coefficient de régression est significativement différent de 0 au risque α=0,05.

C FAUX La droite de régression linéaire donnée est de la forme $Y=b_1X+b_0$, avec Y=SK, b_1=0,5, X=CC et b_0=1. D'après notre cours on sait que b1 correspond à la variation moyenne de Y par unité d'augmentation de X. On cherche donc b_1 dans cet item, il est donc de 0,5. On peut donc en conclure que lorsque la consommation en chocolat (CC) augmente de 1 unité, le nombre de tueurs en séries (SK) augmente de 0,5 pour 1 million d'habitants.

D FAUX c'est l'inverse car le coefficient de corrélation est positif. Donc plus la quantité de chocolat consommée augmente, plus le nombre de tueurs en série augmente au sein d'un pays.

E FAUX cf. correction item B.

Question 14 :

Deux amis tombent sur des articles de corrélations farfelues. Ayant bien appris leur cours sur la Corrélation-Régression de PACES, ils décident de vérifier l'une d'entre elle avec un test de Spearman : « Quel est le rapport entre le score du vainqueur à l'Eurovision et nombre de décès par la foudre en Allemagne ? » Pour faciliter leur travail, ils mettent au point le tableau suivant :

	Score à l'Eurovision	r_i	Nb de décès par la foudre	s_i	d_i	d_i^2
2008	272		1			
2009	387		5			
2010	246		0			
2011	221		1			
2012	372		6			
2013	281		1			
					$\sum_i d_i^2$	

On admettra que $\frac{8}{7} \approx 1$, et que le test effectué est bilatéral avec un risque α=0,05.

A. Le test de Spearman est paramétrique, c'est pourquoi il n'y a pas d'hypothèse sur la distribution des deux variables étudiées ici.
B. On attribue des rangs par ordre croissant des valeurs des deux variables.
C. L'hypothèse nulle testée ici est ρ_S=0.
D. Le coefficient de corrélation de Spearman calculé vaut environ 0,90.
E. D'après la table ci-dessous, on rejette l'hypothèse nulle et on admet l'existence d'une corrélation entre le score du gagnant de l'Eurovision et le nombre de décès par la foudre.

Valeur seuil critique du coefficient de corrélation de Spearman suivant α et n.

Table H. Critical Values of r_s (Spearman Rank-Order Correlation Coefficient)

	Level of Significance for One-Tailed Test			
	.05	.025	.01	.005
	Level of Significance for Two-Tailed Test			
df	.10	.05	.02	.01
5	.900	1.000	1.000	—
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.683	.783	.833

Two-tailed test = test bilatéral

Df = degré de liberté

Question 14 : BC

Le test de Spearman est **non paramétrique** ! Dans un test non paramétrique le calcul ne porte pas sur les valeurs numériques des mesures issues des échantillons représentatifs des populations, mais sur leurs rangs attribués suite au classement des valeurs par ordre croissant.

Donc item A FAUX et item B VRAI.

Le test est réalisé sous l'hypothèse nulle est ρ_S=0 (comme dans le test de Pearson). Donc item C VRAI.

Pour répondre aux items suivants, on va donc calculer une estimation du coefficient de corrélation de Spearman d'après nos données avec la formule : $r_s = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}$

avec r_i le rang des valeurs du score du gagnant de l'Eurovision, s_i le rang des valeurs du nombre de décès par la foudre en Allemagne et $d_i=r_i-s_i$

Pour s'aider on va donc compléter le tableau donné dans l'énoncé :

	Score à l'Eurovision	r_i	Nb de décès par la foudre	s_i	$d_i= r_i-s_i$	d_i^2
2008	272	3	1	3	0	0
2009	387	6	5	5	1	1
2010	246	2	0	1	1	1
2011	221	1	1	3	-2	4
2012	372	5	6	6	1	1
2013	281	4	1	3	1	1
					$\sum_i d_i^2$	8

On attribue donc un rang 1 au plus petit score du gagnant de l'Eurovision, (soit le gagnant de 2011) et ainsi de suite par ordre de score croissant. On fait de même pour les rangs des valeurs du nombre de décès par foudre avec 1 pour le plus petit nombre de décès. Lorsqu'on se retrouve face à des ex-aequo, on prend le rang moyen : c'est-à-dire que lorsqu'on a trois années (2008, 2010 et 2013) où il y a le même nombre de décès, on fait la moyenne des rangs qu'ils auraient dû occuper (rang 2, 3, 4) donc (2+3+4)/3=3. Ils prennent donc tous les trois le rang n°3.

On remplit ensuite d_i, d_i2 et on fait la somme des d_i2. On trouve donc un résultat de 8. On peut donc ensuite compléter notre formule :

$$r_s = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)} = 1 - \frac{6\times 8}{6\times(6^2-1)} = 1 - \frac{8}{36-1} = 1 - \frac{8}{7\times 5} \approx 1 - \frac{1}{5} \approx 1 - 0{,}2 \approx 0{,}8$$

Donc item D FAUX.

On va ensuite comparer notre r_S à la valeur critique dans la table donnée avec un α=0,05 et à n-2 ddl soit 6-2=4 (car 6 années), sachant qu'on est dans un test bilatéral.

La table ne nous donnant des valeurs critiques qu'à partir de 5 degrés de liberté, nous ne pouvons pas conclure sur la corrélation entre le score du gagnant de l'Eurovision et le nombre de décès par la foudre. Donc item E FAUX.

Nota bene : on remarque néanmoins que pour 5 ddl, un test bilatéral et un risque α=0,05, la valeur seuil donnée par la table est un coefficient de corrélation de 1,000. De plus, plus le nombre de ddl diminue, plus la valeur seuil du coefficient de corrélation augmente. Notre coefficient de corrélation de

Spearman calculé étant de 0,8 environ, il est de toute façon inférieur à 1. Nous ne pouvons donc pas rejeter l'hypothèse nulle et conclure sur une corrélation significative.

Question 15 :

On s'intéresse à l'impact des céréales secrètes du petit déjeuner d'Usain Bolt sur ses performances sportives à l'entraînement du matin. Bien que le mécanisme physiologique de métabolisation de ces céréales n'ait pas encore été totalement découvert, la quantité de céréales mangée QC (en grammes) et le temps T(en secondes) d'Usain Bolt au 100m quelques heures après sont liés linéairement. La relation a été estimée sur 110 entraînements : T = -0,001QC + 10.

On fait l'hypothèse que QC et T sont distribuées normalement avec des écart-types respectifs de 50 et 0,1.

Données : 36 x 3 = 108

A. Le coefficient de corrélation entre QC et T est de 0,5.
B. La covariance de QC et T est égale à $5b_1$.
C. La statistique de test du coefficient de régression vaut -3.
D. Dans cet exercice, l'ordonnée à l'origine à un sens.
E. On peut estimer grâce à la droite de régression que lors de son record du monde en 9s58 aux Championnats du monde de 2009, Usain Bolt aurait mangé environ 420 grammes de céréales au petit déjeuner.

Question 15 : DE

A FAUX L'énoncé nous donne l'équation de la droite de régression : T = -0,001×QC + 10, qui correspond dans le cours à l'équation Y=b1*X+b0. On peut donc dire que le coefficient de régression b1 est estimé à -0,001.

Grâce à la formule $\rho_{QC,T} = b_1 \frac{\sigma_{QC}}{\sigma_T}$, on peut dire que $\rho_{QC,T} = -0{,}001 \times \frac{50}{0{,}1} = -0{,}5$

Attention au signe !

B FAUX On a $b_1 = \frac{cov_{QC,T}}{\sigma^2{}_{QC}}$ donc $cov_{QC,T} = b_1 \text{x } 50^2 = b_1 \text{x } 2500$

C FAUX

$$t = \frac{b_1}{\sqrt{\frac{\frac{s_T^2}{s_{QC}^2} - b_1^2}{n-2}}} = \frac{b_1 \times \sqrt{n-2}}{\sqrt{\frac{s_T^2}{s_{QC}^2} - b_1^2}} = \frac{-10^{-3} \times \sqrt{110-2}}{\sqrt{\frac{0{,}1^2}{50^2} - (10^{-3})^2}} = \frac{-10^{-3} \times \sqrt{108}}{\sqrt{\frac{10^{-2}}{\frac{1}{4} \times 10^4} - 10^{-6}}} = \frac{-10^{-3} \times \sqrt{36 \times 3}}{\sqrt{4 \times 10^{-6} - 10^{-6}}}$$

$$\frac{-10^{-3} \times \sqrt{36} \times \sqrt{3}}{\sqrt{3} \times \sqrt{10^{-6}}} = -6$$

D VRAI L'ordonnée à l'origine correspond ici au temps que fait Usain Bolt à l'entraînement s'il n'a pas mangé ses céréales secrètes au petit déjeuner.

E VRAI Si l'on prend la droite de régression linéaire : T = -0,001QC + 10 et que l'on remplace T par 9,58, on obtient : 9,58 = -0,001*QC + 10

Donc QC * (-0,001) = 9,58 – 10

Donc QC = (-0,42)/(-0,001) = 420 grammes de céréales.

Question 16 :

Chez 27 enfants qui ont une mauvaise moyenne générale (noté Y) : moins de 10/20, on étudie l'existence d'un manque de travail personnel (noté X), c'est-à-dire un nombre d'heures consacrées aux études par semaine inférieur à 7. On suppose que les variables X et Y sont normalement distribuées. Le coefficient de corrélation de Pearson est estimé à 0.6, et la grandeur test vaut t=3,75. L'équation mettant en relation X et Y est de la forme : $Y = b_0 + b_1X$

On donne :

- $m_x=70$ $s_x=20$
- $m_y=20$ $s_y=10$
- $s_{x,y} = 120$

A. Le coefficient de régression vaut 1,2.
B. La variabilité du manque de travail personnel explique 60% de la variabilité des mauvais résultats scolaire.
C. On effectue un test à 25ddl.
D. b_1 est significativement différent de 0 au risque $\alpha=0{,}05$.
E. En moyenne, le risque d'avoir de mauvais résultats scolaires augmente quand le manque de travail personnel augmente.

Question 16 : CDE

A FAUX Le coefficient de corrélation est b_1, $b_1 = \frac{\sigma_{X,Y}}{\sigma^2_X} = \frac{120}{20^2} = 0.3$

B FAUX La variance de X explique ρ^2 de la variance de Y. ρ correspond au coefficient de corrélation de Pearson => 0.6 d'après l'énoncé. Donc, la variance du manque de travail explique 0.6^2 soit 36% de la variabilité des mauvais résultats scolaires.

C VRAI C'est du cours, on regarde bien à n-2 ddl.

D VRAI Il faut comparer la valeur de notre test statistique (soit t= 3.75) à la valeur seuil du test à 25ddl pour p=0.05. On trouve $t_{seuil}= 2.0595$

$t > t_{seuil}$ => donc $p< 0.05$ il y a donc RH0. Or H0 est l'hypothèse nulle selon laquelle $b_1=0$ => puisqu'on rejette cette hypothèse, b_1 est significativement différent de 0.

E VRAI $\rho> 0$, donc X et Y varient dans le même sens.

Question 17 :

Une étude a été menée sur 51 personnes grands amateurs de chocolat. Cette étude souhaite étudier la relation entre la quantité de chocolat mangé en gramme (noté X) et la glycémie exprimée en mmol/L (noté Y). On suppose que X et Y sont normalement distribuées. L'ordonnée à l'origine de l'équation mettant en relation X et Y vaut 2. On prendra $\sqrt{7}=2.5$

On donne :

$m_X = 3$

$s_{X,Y} = 4.5$

$s_X = 1.5$

$E(Y^2)-E(Y)^2 = 16$

A. La moyenne estimée de Y vaut 7.
B. En moyenne, le taux de glycémie augmente lorsque la quantité de chocolat mangé par jour augmente.

C. La valeur du test du coefficient de régression vaut 8.4
D. B1 est significativement différent de 0 au risque α=0.03
E. Sous l'hypothèse nulle, la probabilité d'observer un coefficient de régression au moins aussi grande que 8.4 est inférieur à 0.01%

Question 17 : BCD

A FAUX L'équation mettant en relation Y et X est de la forme : Y= b1X + b0. L'ordonnée à l'origine est la valeur que prend Y lorsque X=0, on a donc b0 = 2. On cherche b1 afin de trouver m_y via m_x dans notre équation : $b_1 = \frac{\sigma_{X,Y}}{\sigma^2_X} = \frac{4.5}{1.5^2} = 2$

On a donc maintenant : $m_y = 2*m_x + 2 \Rightarrow m_y = 8$ (=> valeur que prend Y quand X est centrée sur sa moyenne)

B VRAI Rappel : Si ρ < 0 => X et Y varient en sens opposé
Si ρ > 0 => X et Y varient dans le même sens
Ici $b_1>0$ donc ρ>0 ainsi X et Y varient donc dans le même sens

C VRAI

$$t = \frac{b1-0}{\sqrt{\frac{sy^2}{sx^2}-b1^2}} * \sqrt{n-2} = \frac{2-0}{\sqrt{\frac{4^2}{1.5^2}-(2)^2}} * \sqrt{51-2} = \frac{2}{\sqrt{\frac{16}{2.25}-4}} * 7 = \frac{2}{\sqrt{\frac{16-9}{2.25}}} * 7 = \frac{14}{\sqrt{\frac{7}{2.25}}} = \frac{14}{\sqrt{7}} * 1.5 = \frac{14*1.5}{2.5}$$

$$t = 14 * \frac{3}{5} = 14*0.6 = 8.4$$

D VRAI On regarde dans la table de Student, pour 49 ddl, notre valeur t se trouve pour un p < 0,001 donc p< α. On rejette l'hypothèse nulle selon laquelle b1 = 0.

E FAUX Il s'agit de la valeur du petit p trouvé à l'item D, p < 0.001 soit 0.1%

Question 18 :

Des étudiants en 2ème année de médecine souhaitent savoir s'il y a un lien entre la quantité de travail fournie exprimée en heures de révision en PACES par les étudiants (notée X) et le niveau d'altération de la vue des étudiants exprimé en unité visuelle (noté Y).
On donne : $\mu_Y = 8$, $\sigma_x = 2$, $\mu_X = 3$; cov(X,Y) = 10

Quelle est l'équation de la droite de régression modélisant la situation ?

A. $Y = 5x - 7$
B. $Y = 2,5x + 0,5$
C. $Y = 2,5 - 0,5$
D. $Y = 5x + 7$
E. Il manque une donnée pour répondre à l'énoncé.

Question 18 : B

$b_1 = \frac{\text{cov (X,Y)}}{(\sigma_x)^2} = \frac{10}{4} = 2,5$

$Y = b_1X + b_0$

D'où, $b_0 = Y - b_1X \Rightarrow b_0 = 8 - 2,5 \times 3 \Rightarrow b_0 = 0,5$

Donc $Y = 2,5x + 0,5$

Enoncé commun aux questions 19 et 20 :

L'anémie de Biermer se caractérise par une malabsorption de la vitamine B12, et donc une carence de cette dernière. Cette vitamine a un rôle essentiel dans la biosynthèse de l'ADN et dans la production des globules rouges.
On s'intéresse à deux variables aléatoires suivant toutes deux une loi normale :
X correspond à la concentration de vitamine B12 dans le sang et Y correspondant au taux de globules rouges.
Pour cela, on étudie un échantillon de 66 personnes.
On a : coefficient de corrélation : $\rho_{X,Y}$ = -0,6 σ_X = 4 σ_Y = 5

Question 19 :

A. Le test du coefficient de corrélation est à 66 ddl.
B. La covariance $\sigma_{X,Y}$ = -12
C. Y explique 36% de la variabilité de X.
D. Pour α=1%, le coefficient de corrélation est significativement différent de 0, X et Y sont dépendants.
E. La statistique du test paramétrique vaut -6.

Question 19 : BDE

A FAUX Le test est à n-2 ddl, or ici n=66, ainsi le test est à 64 ddl.

B VRAI Pour calculer la covariance, on utilise la formule : $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\,\sigma_Y}$ avec $\rho_{X,Y}$ = -0,6, σ_X = 4 et σ_Y = 5.
On a donc : $\sigma_{X,Y} = \rho_{X,Y}.\sigma_X.\sigma_Y$ = -0,6 x 4 x 5 = -12.

C FAUX C'est l'inverse : X explique $\rho_{X,Y}{}^2$ de la variabilité de Y, avec $\rho_{X,Y}{}^2$ = $(-0,6)^2$ = 0,36
Ainsi c'est X qui explique 36% de la variabilité de Y.
X est la variable explicative, indépendante de Y et c'est Y qui dépend de X et donc s'explique selon X.

D VRAI Pour calculer la statistique de test, on peut utiliser la formule suivante : $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, avec r l'estimateur de $\rho_{X,Y}$.

$$t = \frac{-0{,}6\sqrt{64}}{\sqrt{1-0{,}6^2}} = \frac{-0{,}6 \text{ x } 8}{\sqrt{1-0{,}36}} = \frac{-0{,}6 \text{ x } 8}{\sqrt{0{,}64}} = \frac{-0{,}6 \text{ x } 8}{0{,}8} = \frac{-0{,}6}{0{,}1} = -6$$

Attention à ne pas oublier que t est en valeur absolue pour le comparer au t_{seuil} pour un test bilatéral.
On cherche la valeur seuil dans notre table pour α=0,01 et pour 64 ddl, ainsi on trouve $t_{seuil\ 64ddl}$ = 2,6603, ainsi $|t| > t_{seuil\ 64ddl}$, on rejette H_0. Donc le coefficient de corrélation est significativement différent de 0.

E VRAI Le test du coefficient de Pearson est bien un test paramétrique.

Question 20 :

La valeur de la concentration moyenne de vitamines B12 dans le sang vaut 3 ng/cl, et le taux moyen de globules rouges vaut 5 T/L.
On sait que Y = b_1X + b_0

A. Le coefficient de régression est égal à b_1 = -0,75.
B. Le taux de globules rouges diminue en moyenne de 0,75 T/L par augmentation de la concentration de vitamines B12 de 1 ng/cl.
C. La concentration de vitamine B12 dépend du taux de globules rouges.

D. L'écart-type de b_1 vaut – 0,125.
E. La valeur moyenne de Y quand X est centré sur sa moyenne est égale à b_0 = 2,75.

Question 20 : AB

A VRAI On peut utiliser deux formules du cours : $b_1 = \frac{\sigma_{X,Y}}{\sigma^2{}_X}$ ou $b_1 = p_{X,Y}\frac{\sigma_Y}{\sigma_X}$

On obtient le même résultat avec chaque formule : $b_1 = \frac{-12}{4^2} = \frac{-12}{16} = -0{,}75$

Ou $b_1 = -0{,}6 \text{ x} \frac{5}{4} = -0{,}6 \text{ x } 1{,}25 = -0{,}75$

B VRAI b_1 correspond à l'augmentation de Y par unité d'augmentation de X. Or b_1 est négatif ainsi quand la variable aléatoire X augmente, Y diminue.

C FAUX On exprime Y en fonction de X, ainsi X ne dépend pas de Y.
C'est seulement Y qui dépend de X si le test du coefficient de Pearson est significatif.

D FAUX Comme tester b_1 équivaut à tester $p_{X,Y}$, et que l'on connait la statistique de test t et b_1, on peut utiliser la formule : t = $\frac{b_1}{\sigma_{b_1}}$. Donc $\sigma_{b_1} = \frac{b_1}{t}$.

On a b_1 = -0,75 et t = -6

Attention, ici on ne prend pas la valeur absolue de t. En effet, il faut absolument conserver le signe de la statistique de test pour être cohérent avec le signe du coefficient de régression. La formule de la statistique de test en valeur absolue est utile pour les tests bilatéraux afin de comparer la valeur de la statistique de test à la valeur seuil de la table.

Donc $\sigma_{b_1} = \frac{b_1}{t} = \frac{-0{,}75}{-6}$ = 0,125

E FAUX On utilise l'équation $\bar{Y} = b_1\bar{X} + b_0$, avec b_1 = -0,75, la valeur moyenne de X = 3 et la valeur moyenne de Y = 5.

On a donc $b_0 = \bar{Y} - b_1\bar{X}$ = 5 – (-0,75) x 3 = 5 + 0,75 x 3 = 5 + 2,25 = 7,25.

Question 21 :

Une étude porte sur le lien potentiel entre la consommation de fruits par semaine et l'apaisement du rythme cardiaque en période de stress. En effet, les fruits sont des sources de minéraux et de vitamines supposés soulager le stress. Cette étude recense 38 étudiants en PACES une heure avant le concours blanc du Tut.

Soit X la variable aléatoire qui correspond au poids de fruits consommés par semaine en hg, et Y correspondant à la diminution de la fréquence cardiaque en période de stress.

Ces deux variables aléatoires suivent une loi normale.

On obtient : σ_Y = 2,5 $\sigma_{X,Y}$ = 4 Y = X + 1,3

A. σ_X = 4.
B. Le coefficient de corrélation vaut 0,8.
C. X explique 80% de la variabilité de Y.
D. La statistique de test calculée à 36 ddl, vaut 8.
E. Pour α=5%, on rejette l'hypothèse nulle, on conclut que les fruits participent à soulager le stress par la diminution de la fréquence cardiaque.

Question 21 : BDE

A FAUX On utilise la formule : $b_1 = \frac{\sigma_{X,Y}}{\sigma^2{}_X}$ avec b_1= 1 (que l'on trouve dans la formule Y = b_1X + b_0) et $\sigma_{X,Y}$ = 4.

Ainsi on a : $\sigma^2_X = \frac{\sigma_{X,Y}}{b_1} = \frac{4}{1} = 4$, donc $\sigma_X = 2$.

B VRAI Le coefficient de corrélation est $p_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\,\sigma_Y} = \frac{4}{2 \times 2.5} = \frac{4}{5} = 0{,}8$.

C FAUX On sait que X explique $p^2_{X,Y}$ de la variabilité de Y, avec $p^2_{X,Y}$ le coefficient de détermination. Ainsi X explique 64% de la variabilité de Y.

D VRAI Le test est à n-2 ddl, or ici n=38, ainsi notre test est à 36 ddl.

Pour calculer la statistique de test, on peut utiliser la formule suivante : $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, avec r l'estimateur de $p_{X,Y}$.

$$t = \frac{0{,}8\sqrt{36}}{\sqrt{1-0{,}8^2}} = \frac{0{,}8 \times 6}{\sqrt{1-0{,}64}} = \frac{0{,}8 \times 6}{\sqrt{0{,}36}} = \frac{0{,}8 \times 6}{0{,}6} = \frac{0{,}8}{0{,}1} = 8$$

E VRAI On cherche la valeur seuil dans notre table pour α=0,05 et pour 36 ddl, ainsi on trouve $t_{seuil\ 36ddl} = 2{,}0281$, ainsi $|t| > t_{seuil\ 36ddl}$, donc on rejette H_0. On conclut à une dépendance entre la consommation de fruits et la diminution du rythme cardiaque en période de stress.

Question 22 :

On s'intéresse aux capacités du repas de Noël à agir sur l'apparition de poignées d'amour chez les français.
Pour cela, on étudie l'influence du temps passé à table (en heures) sur le nombre de kilogrammes pris (en kg). Ces deux variables aléatoires, notées X et Y, sont supposées normales.
On recense 38 sujets que l'on pèsera un jour avant le 25 décembre ainsi que un jour après.
On a : $\sigma_X = 1$ $\sigma_Y = 4$ $\sigma_{X,Y} = 1{,}6$

Aide au calcul : $0{,}84 \approx 0{,}81$

A. Le coefficient de corrélation vaut 0,1.
B. Dans cette étude, l'estimation du coefficient de régression est égale à l'estimation de la covariance.
C. Pour un risque alpha de 1%, à 36 ddl, on ne rejette pas l'hypothèse nulle.
D. Pour un risque alpha de 5%, on rejette l'hypothèse nulle.
E. X explique 40% de la variabilité de Y.

Question 22 : **BCD**

A FAUX Le coefficient de corrélation est $p_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\,\sigma_Y} = \frac{1{,}6}{1 \times 4} = 0{,}4$.

B VRAI Le coefficient de régression est $b_1 = \frac{\sigma_{X,Y}}{\sigma^2_X} = \frac{1{,}6}{1^2} = 1{,}6 = \sigma_{X,Y}$

C VRAI Le test est à n-2 ddl, or ici n=38, ainsi notre test est à 36 ddl.

Pour calculer la statistique de test, on peut utiliser la formule suivante : $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, avec r l'estimateur de $p_{X,Y}$.

$$t = \frac{0{,}4\sqrt{36}}{\sqrt{1-0{,}4^2}} = \frac{0{,}4 \times 6}{\sqrt{1-0{,}16}} = \frac{0{,}4 \times 6}{\sqrt{0{,}84}} = \frac{0{,}4 \times 6}{\sqrt{0{,}81}} = \frac{0{,}4 \times 6}{0{,}9} = \frac{4 \times 2}{3} = \frac{8}{3} = 2{,}66$$

On cherche la valeur seuil dans notre table pour α=0,01 et pour 36 ddl, ainsi on trouve $t_{seuil\ 36ddl} = 2{,}7195$, ainsi $|t| < t_{seuil\ 36ddl}$, donc on ne peut pas rejeter H_0.

D VRAI On cherche la valeur seuil dans notre table pour α=0,05 et pour 36 ddl, ainsi on trouve $t_{seuil\ 36ddl} = 2{,}0281$, ainsi $|t| > t_{seuil\ 36ddl}$, donc on rejette H_0.

E FAUX On sait que X explique $p^2_{X,Y}$ de la variabilité de Y, avec $p^2_{X,Y}$ le coefficient de détermination. Ainsi X explique 16% de la variabilité de Y.

Question 23 - C'est fort en chocolat :

Des études montrent qu'un des constituants du chocolat, l'anandamine, est responsable de la libération d'endorphine dans le cerveau. L'endorphine est l'hormone du plaisir : elle active le circuit de récompense de notre cerveau.
On souhaite étudier le lien entre la consommation journalière de chocolat et la bonne humeur considérée comme un reflet fiable de la sensation de plaisir. On note X le nombre de carreaux de chocolat mangés par jour et Y le nombre d'heures où l'on se sent de bonne humeur (sensation de plaisir) par jour.
Le protocole de l'étude prévoit l'utilisation du test de Spearman bilatéral chez 16 patients.
On donne : $\sum_i {d_i}^2 = 136$
Aides aux calculs : $\sqrt{14} \approx 3{,}5$; $6 * 136 = 816$; $16 * 255 = 4080$

A. On pose H1 l'hypothèse « Il existe une corrélation entre la consommation journalière de chocolat et la bonne humeur ».
B. Le coefficient de corrélation non paramétrique de Spearman vaut 0,8.
C. La statistique du test vaut environ 4,67.
D. Au risque alpha = 0,1%, on ne rejette pas l'hypothèse nulle H0.
E. Le nombre de carreaux de chocolat consommés par jour explique 80 % de la sensation de plaisir quotidien.

Question 23 : ABC

A VRAI L'hypothèse nulle H0 est l'hypothèse d'absence de corrélation / de lien entre deux variables donc ici H0 correspond à l'absence de corrélation entre la consommation journalière de chocolat et la sensation de plaisir. Donc l'hypothèse H1 est la présence d'une corrélation entre les 2 variables.

B VRAI Attention ! Le coefficient de corrélation de Spearman est bien NON paramétrique
On applique ensuite la formule :

$$r_s = 1 - \frac{6 * \sum_i {d_i}^2}{n(n^2-1)} = 1 - \frac{6 * 136}{16 * (256-1)} = 1 - \frac{6 * 136}{16 * 255} = 1 - \frac{816}{4080} = 1 - 0{,}2 = 0{,}8$$

C VRAI On applique ensuite la formule :

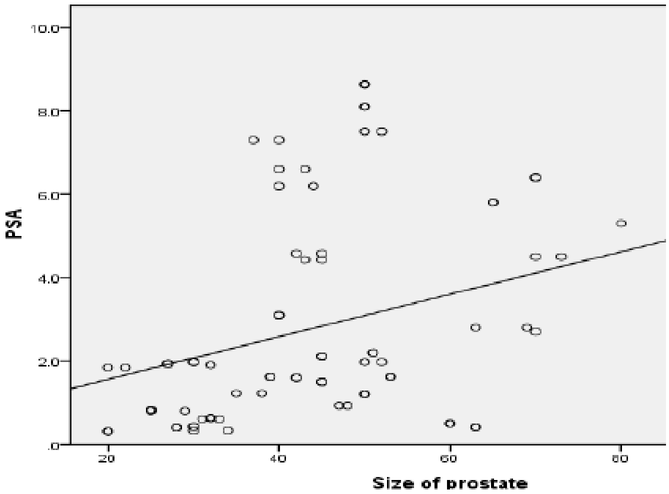
$$T = \frac{r_s * \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}8 * \sqrt{14}}{\sqrt{1-0{,}64}} = \frac{0{,}8 * 3{,}5}{\sqrt{0{,}36}} = \frac{2{,}8}{0{,}6} \approx 4{,}67$$

D FAUX On cherche dans la table de Student pour n - 2 ddl soit 16 – 2 = 14 ddl et α = 0,001
Dans la table on trouve $T_{seuil_{14\,ddl}} = 4{,}1405$ donc $|T| > T_{seuil\ 14\ ddl}$, donc on rejette H0 au risque alpha = 0,1%.

E FAUX Le nombre de carreaux de chocolat consommés par jour (X) explique bien la sensation de plaisir quotidien (Y) mais à hauteur de 64% (${r_s}^2 = 0{,}8 * 0{,}8 = 0{,}64$).

Question 24 :

Le cancer de la prostate est une des premières causes de cancer chez l'homme. La prostate est une glande masculine qui fabrique du PSA, une enzyme utilisée comme marqueur des cancers de la prostate.
On s'intéresse à la relation entre le taux de PSA dans le sang et la taille de la prostate.
X est la variable aléatoire qui modélise la taille de la prostate (en UI) et Y est la variable aléatoire qui modélise le taux de PSA dans le sang (en UI).
On donne : $r_{X,Y} = 0{,}4$; $\sigma_X = 5$; $\sigma_Y = 1$



A. Le coefficient de corrélation vaut 0,08.
B. L'ordonnée à l'origine est b_0 est positive et inférieure à 2 UI.
C. b_1 vaut 0,08.
D. Le taux de PSA dans le sang (Y) augmente en moyenne de 0,4 UI par augmentation de la taille de la prostate (X) de 1 unité.
E. La covariance est négative.

Question 24 : BC

ATTENTION, cet énoncé n'est pas rigoureusement exact car la pente de la droite représentée sur le graphique (coefficient directeur) ne correspond pas à celle calculée. MAIS l'exercice est faisable avec les données de l'énoncé. (rassurez-vous, il n'y aura jamais ce genre d'erreur au concours).

A FAUX Le coefficient de corrélation est $r_{X,Y}$, il est donné et vaut 0,4. Le coefficient de **régression** est b_1 !

B VRAI On lit sur le graphique l'endroit où la courbe de régression croise l'axe des ordonnées : c'est l'ordonnée à l'origine (b_0) qui vaut environ 1,5

C VRAI Calculons b_1 avec la formule : $r_{X,Y} = b_1 * \frac{\sigma_X}{\sigma_Y}$
donc $b_1 = r_{X,Y} * \frac{\sigma_Y}{\sigma_X} = 0{,}4 * \frac{1}{5} = 0{,}4 * 0{,}2 = 0{,}08$

D FAUX Y est la variable expliquée et X la variable explicative, donc X explique bien Y MAIS b_1 vaut 0,08 donc le taux de PSA dans le sang augmente en moyenne de 0,08 UI par augmentation de la taille de la prostate de 1 unité.

E FAUX La covariance et le coefficient de corrélation $r_{X,Y}$ varient dans le même sens. Or, $r_{X,Y} = 0{,}4 > 0$ donc la covariance est positive.

Question 25 – Les produits laitiers sont nos amis pour la vie :

Le calcium, retrouvé en grande quantité dans les produits laitiers, est essentiel à la croissance car il participe au métabolisme osseux. La vitamine D favorise l'absorption du calcium dans l'intestin grêle. Soit X la variable aléatoire qui modélise la concentration de vitamine D dans les cellules intestinales et Y la variable aléatoire modélisant le taux d'absorption du calcium au niveau des cellules intestinales. X etY suivent chacune une loi normale. On étudie un échantillon de 82 personnes.

$$cov_{Y,X} = 18 \ ; \ \sigma_X = 5 \ ; \ \sigma_Y = 4$$

Aides aux calculs : $\sqrt{19} \approx \sqrt{20} \approx 4{,}5$ $\qquad \sqrt{82} \approx \sqrt{80} \approx 9$

A. Le coefficient de corrélation $p_{X,Y}$ vaut 18.
B. Pour α=5%, on rejette l'hypothèse nulle, on conclut que le taux d'absorption du calcium et la concentration de vitamine D sont liés.
C. Pour α=0,1%, on rejette l'hypothèse nulle, on conclut que le taux d'absorption du calcium et la concentration de vitamine D sont liés.
D. Le coefficient de régression b1 vaut 1,125.
E. Le taux de calcium absorbé dans l'intestin grêle explique 81 % de la concentration en vitamine D dans les cellules intestinales.

Question 25 : BC

A FAUX Le coefficient de corrélation c'est $p_{X,Y}$

$$p_{X,Y} = \frac{cov_{Y,X}}{\sigma_X * \sigma_Y} = \frac{cov_{X,Y}}{\sigma_X * \sigma_Y} = \frac{18}{20} = 0{,}9$$

NB j'ai mis cet item pour que vous appliquiez le fait que $cov_{Y,X} = cov_{X,Y}$. D'autre part, si $cov_{X,Y}$ et $p_{X,Y}$ varient dans le même sens, ils ne sont PAS du tout égaux !!!

B VRAI On calcule la statistique de test t.

$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$ avec $r = \frac{cov_{Y,X}}{\sqrt{\sigma_X{}^2 * \sigma_Y{}^2}} = p_{X,Y} = 0{,}9$ et n=83

Donc $t = \frac{|0{,}9|\sqrt{83-2}}{\sqrt{1-0{,}9^2}} = \frac{0{,}9*9}{\sqrt{1-0{,}81}} = \frac{0{,}9*9}{\sqrt{0{,}19}} = 0{,}9 * 9 * \frac{\sqrt{100}}{\sqrt{19}} = \frac{9}{10} * 9 * \frac{10}{\frac{9}{2}} = 9 * 9 * \frac{2}{9} = 18$

On cherche ensuite dans la table de Student pour n-2 ddl soit 82-2=80ddl et $\alpha = 0{,}05$

Dans la table on trouve $t_{\text{seuil 80ddl}}$= 1,9901 donc $|t| > t_{\text{seuil 80ddl}}$, donc on rejette H0 et on conclut que le taux d'absorption du calcium et la concentration de vitamine D sont liés avec une confiance de 95%

C VRAI Idem que question C : on cherche ensuite dans la table de Student pour n-2 ddl soit 83-2=81 ddl et $\alpha = 0{,}001$

Dans la table on trouve $t_{\text{seuil 81 ddl}}$= 3,4163 donc $|t| > t_{\text{seuil 81 ddl}}$, donc on rejette H0 et on conclut que le taux d'absorption du calcium et la concentration de vitamine D sont liés avec une confiance de 99,9%

D FAUX J'ai mis cet item pour ceux qui ont inversé σ_X et σ_Y dans la formule (cf calcul ci-dessous).

$r_{X,Y} = b_1 * \frac{\sigma_X}{\sigma_Y}$ donc $b_1 = r_{X,Y} * \frac{\sigma_Y}{\sigma_X} = 0{,}9 * \frac{4}{5} = 0{,}9 * 0{,}8 = 0{,}72$

Attention, la prof me l'a fait remarquer, le coefficient de régression b1 peut prendre toutes les valeurs dans R ($-\infty < \text{b1} < +\infty$) contrairement au coefficient de corrélation $r_{X,Y}$.

E FAUX Il est vrai que ${p_{X,Y}}^2 = 0{,}81$ MAIS c'est X qui explique Y car Y dépend de X (l'absorption de Ca2+ (Y) dépend de la concentration en vitamine D (X)) donc c'est la concentration en vitamine D dans les cellules intestinales (X) qui explique 81% du taux d'absorption de Ca 2+ (Y) par les cellules intestinales.

Question 26 – On arrête avec la bouffe, maintenant c'est la boisson :

Une étude réalisée par vos tuteurs s'intéresse au lien entre la consommation de boisson caféinée avant une épreuve et le stress durant l'épreuve. Pour approcher le stress, on s'intéresse à la fréquence cardiaque de l'étudiant en PASS. On pose X la variable aléatoire correspondant à la quantité de boisson caféinée consommée, en décilitres, et Y la variable aléatoire quantifiant la fréquence cardiaque. On considère pour l'exercice que X et Y suivent toutes deux une loi normale.
On étudie un échantillon de 51 PASS.
Données : Y= b_0 + 6,4X σ_{XY}=6,4 σ_Y=8 $\bar{Y} = 70$ $\bar{X} = 2$

A. Le coefficient de corrélation de Pearson vaut 0,8.
B. On lit la statistique de test dans la table de la loi de Student à 50 ddl.
C. Pour α=5%, on rejette l'hypothèse nulle, on conclut que la consommation de boisson caféinée participe à l'augmentation de la fréquence cardiaque.
D. b_0 vaut 57,2.
E. Y explique 64% de la variabilité de X.

Question 26 – On arrête avec la bouffe, maintenant c'est la boisson : ACD

A VRAI On connait la formule du coefficient de corrélation de Pearson : $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$. On a déjà σ_{XY} dans l'énoncé, de même que σ_Y. Il nous manque néanmoins σ_X.
Pour le calculer, on va utiliser la valeur de b_1 qu'on trouve dans l'équation de la droite de régression, ici b_1 vaut 6,4.
On sait que : $b_1 = \frac{\sigma_{XY}}{\sigma_X{}^2}$. On va donc pouvoir trouver la valeur de σ_X :

$$b_1 = \frac{\sigma_{XY}}{\sigma_X^2} \ donc \ \sigma_X^2 = \frac{\sigma_{XY}}{b_1} = \frac{6,4}{6,4} = 1$$
$$ainsi, \sigma_X = \sqrt{1} = 1$$

Maintenant qu'on a trouvé ça, on peut calculer le coefficient de corrélation de Pearson :

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{6,4}{1 \times 8} = 0,8$$

B FAUX On se place à n-2 ddl soit 51-2 = 49 ddl (avec n l'effectif total de l'échantillon).
C VRAI On se trouve dans le cas où X et Y suivent une loi normale : on va donc utiliser le test paramétrique de Pearson pour calculer la valeur de la statistique du test t :

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,8 \times \sqrt{49}}{\sqrt{1-0,8^2}} = \frac{0,8 \times 7}{\sqrt{1-0,64}} = \frac{5,6}{\sqrt{0,36}} = \frac{5,6}{0,6} = \frac{56}{6} = \frac{54}{6} + \frac{2}{6} = 9 + \frac{1}{3} \approx 9,33$$

On se place maintenant dans la table de la loi de Student à 49 ddl et à alpha=5% pour lire notre valeur seuil :

ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	636,6192
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	31,5991
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	12,9240
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	8,6103
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	6,8688
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,9588
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	5,4079
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	5,0413
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,7809
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,5869
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,4370
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	4,3178
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	4,2208
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	4,1405
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	3,2860	4,0728
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	4,0150
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,9651
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,9216
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,8834
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,8495
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,8193
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,7921
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,7676
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,7066
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,6896
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,6739
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,6594
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,6460
31	0,1267	0,2555	0,3889	0,5298	0,6825	0,8534	1,0541	1,3095	1,6955	2,0395	2,4528	2,7440	3,0221	3,6335
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,6218
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448	2,7333	3,0082	3,6109
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,6007
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238	2,9960	3,5911
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,5821
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314	2,7154	2,9852	3,5737
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,5657
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258	2,7079	2,9756	3,5581
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,5510
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208	2,7012	2,9670	3,5442
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,5377
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163	2,6951	2,9592	3,5316
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,5258
45	0,1264	0,2549	0,3878	0,5281	0,6800	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896	2,9521	3,5203
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,5150
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083	2,6846	2,9456	3,5099
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,5051
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	2,9397	3,5004
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	[illegible]	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,4602
70	0,1261	0,2543	0,3869	0,5268	0,6780	0,8468	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,4350

On trouve que : $t_{49ddl,5\%} = 2{,}0096$

On constate alors :

$$|t| \geq t_{seuil\ 49ddl,5\%}$$

Donc : On rejette l'hypothèse nulle au risque α=5%. On peut dire que la consommation de boisson caféinée participe à augmenter la fréquence cardiaque. On sait qu'elle fait augmenter la fréquence cardiaque, et pas diminuer, car le coefficient de corrélation de Pearson est positif.

D VRAI On peut trouver b_0 grâce aux moyennes de X et Y données par l'énoncé :

$$b_0 = \bar{Y} - b_1 \times \bar{X} = 70 - 6{,}4 \times 2 = 70 - 12{,}8 = 57{,}2$$

E FAUX C'est X qui explique $(\rho_{XY})^2$ de la variabilité de Y.

Avec $(\rho_{XY})^2$ le coefficient de détermination : $(\rho_{XY})^2 = 0{,}8^2 = 0{,}64$

Ici, la consommation de boisson caféinée explique 64% de la variabilité de la fréquence cardiaque.

Question 27 – Boo !!!! :

À l'occasion d'Halloween, les tuteurs de Biostats ont décidé de parcourir 100 films d'horreur, des plus vieux aux plus récents, en 2 semaines. Ils ont remarqué au passage qu'il y a de plus en plus de jumpscares dans les derniers films. Un jumpscare est un changement brutal intégré dans un film pour surprendre le spectateur ou utilisateur.

Curieux, ils souhaitent connaître la relation entre le nombre de jumpscares perçus dans *le dernier des films d'horreur* et le taux d'adrénaline secrété par l'organisme, connue pour être l'hormone de la peur. L'étude est menée dans un échantillon de 102 étudiants. En moyenne, le nombre de jumpscares perçus est estimé à 50 avec un écart-type de 20 et la concentration sanguine moyenne en adrénaline est estimée à 200 ng/L avec une variance de 2500 ng/L.

On note X le nombre de jumpscares perçus dans *le dernier film* et Y la concentration sanguine en adrénaline en ng/L. Ces deux variables aléatoires sont considérées suivre une loi normale.

Données : $\frac{2{,}5}{\sqrt{0{,}9375}} \approx 2{,}58$; $\frac{2{,}5}{\sqrt{0{,}75}} \approx 2{,}89$; E(XY) = 10250

A. Le coefficient de corrélation de Pearson est égal à ¼.
B. Pour α = 0.01, le coefficient de corrélation est significativement différent de 0.
C. Pour α = 0.5%, on peut conclure à une dépendance entre le nombre de jumpscares dans *ce film et* le taux d'adrénaline sécrété par l'organisme.
D. Un test non paramétrique de Spearman ne nécessite pas d'hypothèse concernant la distribution de X et Y pour être appliqué.
E. Quand le nombre de jumpscares dans *le film* augmente, la concentration sanguine en adrénaline diminue.

Question 27 – Boo !!!! : AD

On commence par noter toutes les valeurs données par l'énoncé :

$m_X = 50$

$m_Y = 200$

$s_X = 20$

$s_Y = \sqrt{s_Y{}^2} = \sqrt{2500} = 50$

A VRAI La formule du coefficient de corrélation de Pearson est :

$$r_{X,Y} = \frac{s_{X,Y}}{s_X s_Y}$$

Avec : $s_{X,Y}$ la covariance.

$$s_{X,Y} = E(XY) - E(X)E(Y)$$
$$s_{X,Y} = 10250 - 50.200$$
$$s_{X,Y} = 10250 - 10000$$
$$s_{X,Y} = 250$$

Donc :

$$r_{X,Y} = \frac{s_{X,Y}}{s_X s_Y}$$
$$r_{X,Y} = \frac{250}{20.50} = \frac{250}{1000} = \frac{1}{4} = 0.25$$

B FAUX Ici, on est en présence de deux variables aléatoires suivant une loi normale, on peut alors utiliser le test paramétrique de Pearson.

On cherche la valeur de la statistique de test :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \rightarrow t_{n-2ddl,\alpha}$$
$$t = \frac{0,25\sqrt{102-2}}{\sqrt{1-0,25^2}}$$
$$t = \frac{0,25\sqrt{100}}{\sqrt{1-\left(\frac{1}{4}\right)^2}} = \frac{0,25\sqrt{100}}{\sqrt{1-\frac{1}{16}}} = \frac{0,25\sqrt{100}}{\sqrt{1-0,0625}} = \frac{0,25.10}{\sqrt{0,9375}} = \frac{2.5}{\sqrt{0,9375}} = 2,58$$

On regarde dans la table de Student, pour α = 0.01 à 102 – 2 = 100 ddl

α / ddl	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	636,6192
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	31,5991
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	12,9240
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	8,6103
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	6,8688
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,9588
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	5,4079
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	5,0413
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,7809
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,5869
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,4370
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	4,3178
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	4,2208
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	4,1405
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	3,2860	4,0728
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	4,0150
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,9651
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,9216
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,8834
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,8495
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,8193
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,7921
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,7676
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,7066
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,6896
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,6739
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,6594
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,6460
31	0,1267	0,2555	0,3889	0,5298	0,6825	0,8534	1,0541	1,3095	1,6955	2,0395	2,4528	2,7440	3,0221	3,6335
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,6218
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448	2,7333	3,0082	3,6109
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,6007
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238	2,9960	3,5911
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,5821
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314	2,7154	2,9852	3,5737
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,5657
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258	2,7079	2,9756	3,5581
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,5510
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208	2,7012	2,9670	3,5442
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,5377
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163	2,6951	2,9592	3,5316
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,5258
45	0,1264	0,2549	0,3878	0,5281	0,6800	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896	2,9521	3,5203
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,5150
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083	2,6846	2,9456	3,5099
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,5051
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	2,9397	3,5004
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,4602
70	0,1261	0,2543	0,3869	0,5268	0,6780	0,8468	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,4350
80	0,1261	0,2542	0,3867	0,5265	0,6776	0,8461	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,4163
90	0,1260	0,2541	0,3866	0,5263	0,6772	0,8456	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,4019
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,3905
110	0,1260	0,2540	0,3863	0,5259	0,6767	0,8449	1,0413	1,2893	1,6588	1,9818	2,3607	2,6213	2,8648	3,3812
120	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,3735
130	0,1259	0,2539	0,3862	0,5257	0,6764	0,8444	1,0406	1,2881	1,6567	1,9784	2,3554	2,6142	2,8557	3,3669
140	0,1259	0,2538	0,3861	0,5256	0,6762	0,8442	1,0403	1,2876	1,6558	1,9771	2,3533	2,6114	2,8522	3,3614
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0365	1,2816	1,6449	1,9600	2,3265	2,5760	2,8072	3,2909

Ainsi, $t_{100ddl,1\%} = 2,6259$

$$|t| < t_{seuil\ 100ddl,1\%}$$

Donc : On ne rejette pas l'hypothèse nulle au risque α=0.01, le coefficient de corrélation n'est pas significativement différent de 0.

C FAUX De même, on vérifie notre valeur seuil pour α = 0.005

ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	636,6192
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	31,5991
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	12,9240
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	8,6103
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	6,8688
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,9588
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	5,4079
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	5,0413
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,7809
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,5869
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,4370
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	4,3178
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	4,2208
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	4,1405
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	3,2860	4,0728
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	4,0150
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,9651
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,9216
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,8834
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,8495
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,8193
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,7921
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,7676
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,7066
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,6896
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,6739
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,6594
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,6460
31	0,1267	0,2555	0,3889	0,5298	0,6825	0,8534	1,0541	1,3095	1,6955	2,0395	2,4528	2,7440	3,0221	3,6335
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,6218
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448	2,7333	3,0082	3,6109
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,6007
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238	2,9960	3,5911
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,5821
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314	2,7154	2,9852	3,5737
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,5657
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258	2,7079	2,9756	3,5581
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,5510
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208	2,7012	2,9670	3,5442
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,5377
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163	2,6951	2,9592	3,5316
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,5258
45	0,1264	0,2549	0,3878	0,5281	0,6800	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896	2,9521	3,5203
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,5150
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083	2,6846	2,9456	3,5099
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,5051
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	2,9397	3,5004
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,4602
70	0,1261	0,2543	0,3869	0,5268	0,6780	0,8468	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,4350
80	0,1261	0,2542	0,3867	0,5265	0,6776	0,8461	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,4163
90	0,1260	0,2541	0,3866	0,5263	0,6772	0,8456	1,0424	1,2910	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3,4019
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,3905
110	0,1260	0,2540	0,3863	0,5259	0,6767	0,8449	1,0413	1,2893	1,6588	1,9818	2,3607	2,6213	[illegible]	3,3812
120	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,3735
130	0,1259	0,2539	0,3862	0,5257	0,6764	0,8444	1,0406	1,2881	1,6567	1,9784	2,3554	2,6142	2,8557	3,3669
140	0,1259	0,2538	0,3861	0,5256	0,6762	0,8442	1,0403	1,2876	1,6558	1,9771	2,3533	2,6114	2,8522	3,3614
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0365	1,2816	1,6449	1,9600	2,3265	2,5760	2,8072	3,2909

Ici, on trouve $t_{100ddl,0.5\%} = 2{,}8707$.

$$|t| \leq t_{seuil\ 100ddl,1\%}$$

On ne rejette pas l'hypothèse nulle au risque α=0.005. On ne peut pas conclure à une dépendance entre le nombre de jumpscares dans *le film d'horreur* et le taux d'adrénaline sécrété par l'organisme. En réalité, étant donné qu'on n'a déjà pas pu rejeter l'hypothèse nulle avec α=0.01, on sait d'avance qu'on ne pourra pas non plus le rejeter avec un risque inférieur, vous n'étiez donc pas obligé de vérifier dans la table pour α=0.005.

D VRAI Le test non paramétrique ne comprend pas d'hypothèses sur la distribution de $(\boldsymbol{X}, \boldsymbol{Y})$. Les valeurs de X sont rangées dans l'ordre croissant avec des rangs ri de 1 à n. C'est le même fonctionnement pour Y, avec des rangs si de 1 à n. On utilise ensuite la formule :

$$r_S = 1 - \frac{6\Sigma_i d_i^2}{n(n^2-1)}$$

Il s'agit d'une phrase du cours.

E FAUX Le signe de notre coefficient de corrélation de Pearson détermine le sens de l'évolution de nos variables X et Y.

Si $r_{X,Y} > 0$: X et Y évoluent dans le même sens.
Si $r_{X,Y} < 0$: X et Y évoluent dans le sens opposé.

Dans cet exercice, $r_{X,Y} = 0{,}25$. On peut dire que X et Y varient dans le même sens. Quand le nombre de jumpscares dans *le film d'horreur* augmente, la concentration sanguine en adrénaline augmente.

Question 28 – 20 cm = Pas si fou :

De récentes études ont montré qu'il pourrait exister une relation entre le nombre de cadeaux sous le sapin et le diamètre de la cheminée des foyers. Cette relation a été étudiée par un groupe de lutins à l'aide d'une régression linéaire, sur un échantillon de n = 66 foyers. Le nombre de cadeaux sous le sapin, noté Y, suit une loi normale de paramètres (17 ; 4). Le diamètre de la cheminée, noté X, suit une loi normale de paramètres (30 ; 5). On note Y = $\beta_0 + \beta_1$.X et on a σ_{XY} = 12,8.

Aides au calcul : $0{,}64^2 = 0{,}4$ $0{,}6 \approx 0{,}64$

A. Le coefficient de corrélation de Pearson vaut -0,64.
B. La statistique du test du coefficient de corrélation est 3,5.
C. Pour α = 0,1 %, on ne peut pas rejeter l'hypothèse nulle d'une absence de corrélation des deux variables aléatoires.
D. L'hypothèse nulle du test du coefficient de régression est β_1=0.
E. Lorsque le diamètre de la cheminée augmente d'une unité, le nombre de cadeaux sous les sapins augmente en moyenne de 0,512 unité.

Question 28 – 20 cm = Pas si fou : DE

A FAUX Le calcul du coefficient de corrélation de Pearson est le suivant :

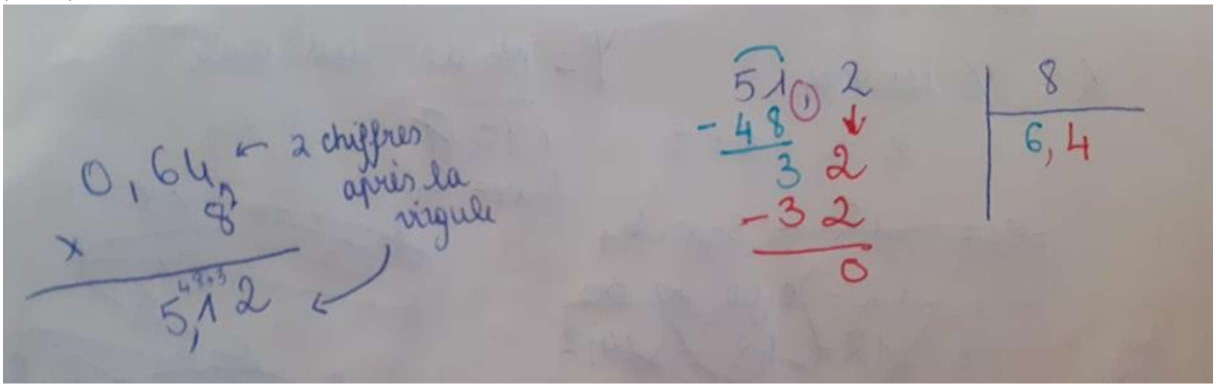
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{12{,}8}{5 \times 4} = \frac{128 \times 10^{-1}}{2 \times 10} = \frac{128}{2} \times \frac{1}{100} = \frac{64}{100} = 0{,}64$$

Il faut bien faire attention au signe, c'est une étourderie facile 😜.

B FAUX

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}64 \times \sqrt{66-2}}{\sqrt{1-0{,}64^2}} = \frac{0{,}64 \times \sqrt{64}}{\sqrt{1-0{,}4}} = \frac{0{,}64 \times 8}{\sqrt{0{,}6}} = \frac{5{,}12}{0{,}8} = \frac{51{,}2}{8} = 6{,}4$$

Ici, il y avait des aides au calcul pour le carré de 0,64, et il fallait aussi voir pour la racine de 0,6 l'approximation 0,6 ≈ 0,64. Ensuite, je vous montre comment poser 0,64 × 8 (même si je pense que vous vous rappelez comment on fait) et 51,2/8 : (il y a une explication plus détaillée de la méthode pour poser une division à l'item E)



Autre méthode pour le calcul (vous gagnerez du temps !) :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}64 \times \sqrt{66-2}}{\sqrt{1-0{,}64^2}} = \frac{0{,}64 \times \sqrt{64}}{\sqrt{1-0{,}4}} = \frac{0{,}64 \times 8}{\sqrt{0{,}6}} = \frac{0{,}64 \times 8}{0{,}8} = \frac{0{,}64 \times 8 \times 100}{0{,}8 \times 100}$$
$$= \frac{64 \times \cancel{8}}{10 \times \cancel{8}} = 6{,}4$$

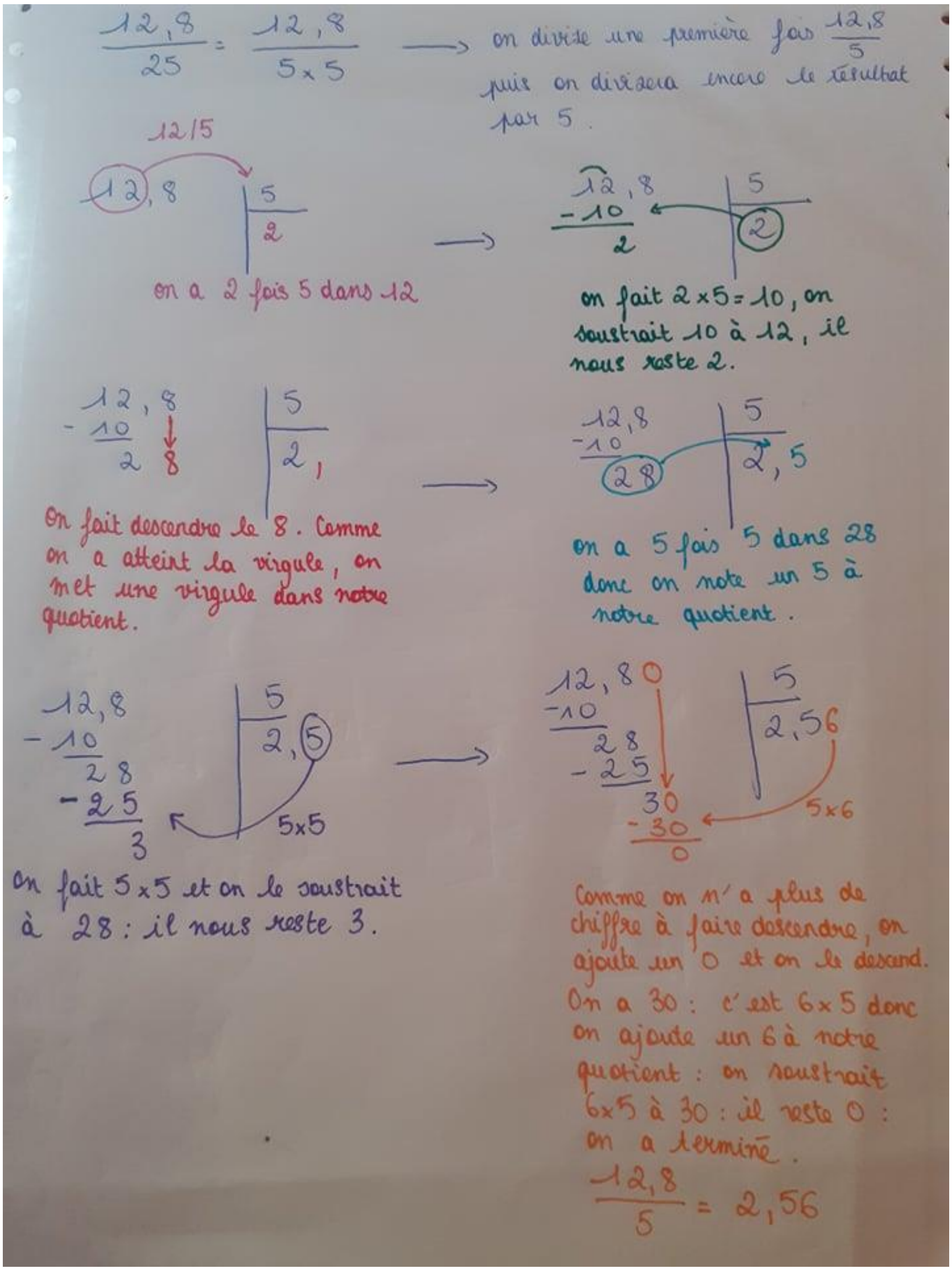
C FAUX On se place dans la table de Student à 64 ddl (comme on a 60 et 70 dans la table, on va faire un encadrement :

ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,6218
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448	2,7333	3,0082	3,6109
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,6007
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238	2,9960	3,5911
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,5821
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314	2,7154	2,9852	3,5737
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,5657
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258	2,7079	2,9756	3,5581
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,5510
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208	2,7012	2,9670	3,5442
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,5377
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163	2,6951	2,9592	3,5316
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,5258
45	0,1264	0,2549	0,3878	0,5281	0,6000	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896	2,9521	3,5203
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,5150
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083	2,6846	2,9456	3,5099
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,5051
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	2,9397	3,5004
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	$\boxed{3{,}4602}$
70	0,1261	0,2543	0,3869	0,5268	0,6780	0,8468	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	$\boxed{3{,}4350}$
80	0,1261	0,2542	0,3867	0,5265	0,6776	0,8461	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	[illegible]
90	0,1260	0,2541	0,3866	0,5263	0,6772	0,8456	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,4019
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,3905
110	0,1260	0,2540	0,3863	0,5259	0,6767	0,8449	1,0413	1,2893	1,6588	1,9818	2,3607	2,6213	2,8648	3,3812
120	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,3735
130	0,1259	0,2539	0,3862	0,5257	0,6764	0,8444	1,0406	1,2881	1,6567	1,9784	2,3554	2,6142	2,8557	3,3669
140	0,1259	0,2538	0,3861	0,5256	0,6762	0,8442	1,0403	1,2876	1,6558	1,9771	2,3533	2,6114	2,8522	3,3614
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0365	1,2816	1,6449	1,9600	2,3265	2,5760	2,8072	3,2909

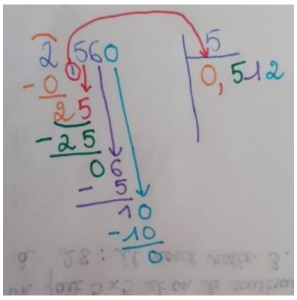
Notre valeur est seuil est donc comprise entre 3,43 et 3,46. Ainsi, avec la valeur du test trouvée à l'item D on note : 6,4 > 3,46 > 3,43, nous pouvons donc rejeter l'hypothèse nulle.
La corrélation entre X et Y est statistiquement significative, on en conclut que le nombre de cadeaux de Noël est directement expliqué par le diamètre de la cheminée.
D VRAI L'hypothèse nulle du test est que Y ne dépend pas de X : ainsi, cela revient à $\beta_1=0$.
E VRAI Cette définition est celle de β_1 : lorsque X augmente d'une unité, Y augmente de β_1. C'est une traduction de l'équation de régression $Y = \beta_0 + \beta_1.X$.
Ainsi, nous devons calculer β_1 :

$$\beta_1 = \frac{\sigma_{XY}}{\sigma^2{}_X} = \frac{12{,}8}{5^2} = \frac{12{,}8}{25} = \frac{12{,}8}{5 \times 5} = \frac{\frac{12{,}8}{5}}{5} = \frac{2{,}56}{5} = 0{,}512$$

J'ai bien conscience que ce n'est pas un calcul très simple à réaliser sans calculatrice, et pas forcément très facile à simplifier non plus mais comme cela pourrait arriver en épreuve d'être bloqué sur un calcul car on ne voit pas la simplification possible (mais les professeurs mettent toujours des calculs réalisables), je voulais faire un petit point sur comment poser une division :



On a donc divisé une première fois par 5, on le fait une seconde fois (j'ai moins détaillé cette fois) :



J'espère que ça vous parait clair ! C'est peut-être un détail mais ça peut vraiment être très utile lorsque vous ne savez pas comment aborder un calcul !

Question 29 – CHA-T-O :

Nos P1 n'étant pas tous aussi calmes, certains ont du mal à juste se laisser bronzer. Un concours de châteaux de sable s'organise alors sur la plage. La quantité de sable (exprimée en pourcentage de la quantité utilisée initialement pour faire le château) est estimée au cours du temps et modélisée grâce à une régression linéaire. Ainsi, on estime donc la capacité du château à persister dans le temps. Les conditions de régression sont validées et permettent d'écrire l'équation de la quantité de sable en fonction du temps (en heures) :

$$\text{Sable} = 100 - 0{,}2 * \text{temps}$$

On donne $r^2=0{,}77$ et $p=0{,}03$.

A. L'hypothèse nulle du test est que la persistance du château sur la plage ne dépend pas du temps.
B. La covariance du test est positive.
C. La tendance linéaire entre quantité de sable et temps est significative au risque alpha=5%.
D. Avec ce modèle, le temps explique 77% de la variabilité de la quantité de sable du château.
E. Avec ce modèle, en 24h, le sable du château diminue presque de moitié.

Question 29 – CHA-T-O : ACD

A VRAI C'est ça ! On cherche par notre test à montrer qu'il existe un **lien** entre les deux variables. L'hypothèse nulle est donc qu'il n'y a pas de lien (sur un graphique, on ne pourrait probablement pas tracer une droite passant par tous les points).

B FAUX Explications détaillées. On voit dans notre équation que b_1 est négatif. Or, la covariance est de signe identique à b_1. Ainsi, la covariance est négative.

C VRAI Oui ! On a $p = 0{,}03 < 0{,}05$. On peut donc en déduire que le test est significatif, il y a un lien entre quantité de sable et temps.

D VRAI En effet, X explique r^2 de la variabilité de Y.

E FAUX On cherche à savoir combien de sable il reste après 24h :

$$Sable = 100 - 0{,}2 * 24 = 100 - 4{,}8 = 95{,}2$$

Il reste 95,2% du château en 24h, on aura diminué de 4,8% seulement.
Je ne pouvais pas vous faire un exo qui utilise toutes les formules de ce chapitre car vous n'auriez pas eu le temps de répondre aux questions, mais pensez à bien les réviser, ce sont des points faciles :

Fiche synthèse : Corrélation Régression

COEFFICIENT DE CORRELATION DE PEARSON

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

$1 \leq \rho_{X,Y} \leq 1$
$\rho_{X,Y} = 0 \rightarrow$ X et Y sont indépendantes.
Plus il est proche de 1 ou -1, plus X et Y sont dépendantes.
X explique $(\rho_{X,Y})^2$ de la variabilité de Y.

TEST DE PEARSON (PARAMETRIQUE)

Conditions :

- X et Y suivent une loi normale
- Ou Y suit une loi normale et X et Y ont une relation linéaire.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \rightarrow t_{n-2ddl,\alpha}$$

$|t| \geq t_{seuil\ n-2ddl,\alpha} \rightarrow$ nous rejetons H_0
$|t| < t_{seuil\ n-2ddl,\alpha} \rightarrow$ nous ne pouvons pas rejeter H_0

TEST DE SPEARMAN (NON PARAMETRIQUE)

Pas de conditions, valeurs rangées dans ordre croissant.
$d_i = r_i - s_i$, avec **r_i** : rangs des valeurs de X et **s_i** : rangs des valeurs de Y.
n : nombre de rangs.

$$r_S = 1 - \frac{6\Sigma_i d_i^2}{n(n^2-1)}$$

Si n>10, on peut calculer la valeur test avec le test de Pearson.
Si n≤10, on consulte la table de Spearman.

REGRESSION

Y en fonction de X :
Y est la variable à expliquer/dépendante de X
X est la variable explicative/indépendante de Y

Droite de régression :
$$Y = b_0 + b_1.X + \epsilon$$

b_0 : l'ordonnée à l'origine, valeur moyenne de Y quand X est centré sur sa moyenne.
$b_0 = \bar{Y} - b_1.\bar{X}$

b_1 : coefficient directeur, variation moyenne de Y par unité d'augmentation de X.

Coefficient de régression :
$$b_1 = \frac{\sigma_{X,Y}}{\sigma^2_X}$$
$$\rho_{X,Y} = b_1 \frac{\sigma_X}{\sigma_Y}$$

ANALYSE DE VARIANCE : ANOVA

Test:

$$F = \frac{S_M}{S_R} = \frac{SCE_E}{SCE_R} \sim F_{(1,N-2)ddl}$$

Question 30 – Encore une histoire de centimètres :

La longueur de la ceinture du Père Noël Y en centimètres varie selon les années en suivant une loi normale N(100 ; 20). Elle dépend de la quantité de cookies X en grammes, laissées par les enfants dans un foyer, qui suit une loi normale N(160 ; 80). On essaie de quantifier la relation entre ces deux variables sur un échantillon de 402 foyers.
On écrira Y = b_0 + b_1.X.
Données : r = 0,8

A. Le test du coefficient de corrélation de Pearson est à 400 ddl.
B. La valeur test obtenue est environ 26,67.
C. Le coefficient de corrélation de Pearson est significativement différent de 0 au risque alpha = 0,1%.
D. b_1 vaut 3,2.
E. b_0 vaut 78.

Question 30 – Encore une histoire de centimètres : ABC

A VRAI L'effectif étant de 402, on est à n-2 ddl soit 400 ddl.
B VRAI On la calcule grâce à r qui est donné par l'énoncé :

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}8 \times \sqrt{402-2}}{\sqrt{1-0{,}8^2}} = \frac{0{,}8 \times \sqrt{400}}{\sqrt{1-0{,}64}} = \frac{0{,}8 \times 20}{\sqrt{0{,}36}} = \frac{\cancel{0{,}}8 \times 20}{\cancel{0{,}}6} = \frac{160}{6} = \frac{120}{6} + \frac{40}{6}$$
$$= 20 + \frac{36}{6} + \frac{4}{6} = 20 + 6 + \frac{2}{3} \approx 26{,}67$$

C VRAI On est à 400 ddl et à alpha = 0,001 mais 400 n'est pas dans la table de Student du formulaire, on va donc faire un encadrement :

ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,005	0,001
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	636,6192
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	31,5991
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	12,9240
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	8,6103
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	6,8688
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,9588
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	5,4079
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	5,0413
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,7809
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,5869
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,4370
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	4,3178
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	4,2208
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	4,1405
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	3,2860	4,0728
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	4,0150
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,9651
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,9216
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,8834
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,8495
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,8193
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,7921
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,7676
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,7454
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,7251
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,7066
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,6896
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,6739
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,6594
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,6460
31	0,1267	0,2555	0,3889	0,5298	0,6825	0,8534	1,0541	1,3095	1,6955	2,0395	2,4528	2,7440	3,0221	3,6335
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,6218
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448	2,7333	3,0082	3,6109
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,6007
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238	2,9960	3,5911
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,5821
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314	2,7154	2,9852	3,5737
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,5657
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258	2,7079	2,9756	3,5581
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,5510
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208	2,7012	2,9670	3,5442
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,5377
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163	2,6951	2,9592	3,5316
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,5258
45	0,1264	0,2549	0,3878	0,5281	0,6800	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896	2,9521	3,5203
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,5150
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083	2,6846	2,9456	3,5099
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,5051
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049	2,6800	2,9397	3,5004
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,4960
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,4602
70	0,1261	0,2543	0,3869	0,5268	0,6780	0,8468	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,4350
80	0,1261	0,2542	0,3867	0,5265	0,6776	0,8461	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,4163
90	0,1260	0,2541	0,3866	0,5263	0,6772	0,8456	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,4019
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,3905
110	0,1260	0,2540	0,3863	0,5259	0,6767	0,8449	1,0413	1,2893	1,6588	1,9818	2,3607	2,6213	2,8648	3,3812
120	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,3735
130	0,1259	0,2539	0,3862	0,5257	0,6764	0,8444	1,0406	1,2881	1,6567	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
140	0,1259	0,2538	0,3861	0,5256	0,6762	0,8442	1,0403	1,2876	1,6558	1,9771	2,3533	2,6114	2,8522	3,3614
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0365	1,2816	1,6449	1,9600	2,3265	2,5760	2,8072	3,2909

La valeur seuil est donc comprise entre 3,2909 et 3,3614. Avec une valeur test de 26,67, nous sommes largement au-dessus, nous pouvons donc rejet H0 : le coefficient de corrélation de Pearson est significativement différent de 0 au risque alpha = 0,1 %.

D FAUX Pour calculer b_1, on a cette formule :

$$r_{X,Y} = b_1 \frac{\sigma_X}{\sigma_Y} \ soit \ b_1 = r_{X,Y} \frac{\sigma_Y}{\sigma_X} = 0{,}8 \times \frac{20}{80} = 8 \times 10^{-1} \times \frac{2 \times 10}{8 \times 10} = 2 \times 10^{-1} = 0{,}2$$

E FAUX Pour calculer b_0, on utilise la formule :

$$b_0 = \bar{Y} - b_1 \times \bar{X} \ soit\ ici \ b_0 = 100 - 0{,}2 \times 160 = 100 - 2 \times 10^{-1} \times 16 \times 10 = 100 - 2 \times 16 = 100 - 32 = 68$$