EXERCICES SUPPLEMENTAIRES

Sur les tests en ***Régression linéaire***

Exercice n° 1 :

Dans un certain type d'éprouvette métallique, la contrainte normale sur une éprouvette est connue pour être fonctionnellement liée à la résistance au cisaillement. Voici un ensemble de données expérimentales codées sur les deux variables :

Contrainte, x	**cisaillement, y**	**Contrainte, x**	**cisaillement, y**
26.8	26.5	24.7	26.3
25.4	27.3	28.1	22.5
28.9	24.2	26.9	21.7
23.6	27.1	27.4	21.4
27.7	23.6	22.6	25.8
23.9	25.9	25.6	24.9

a) Estimer la droite de régression $\mu Y \mid x = \alpha + \beta * x$.
b) Estimer la résistance au cisaillement pour une contrainte normale de 24,5.
c) Evaluer $\hat{\sigma}^2$.
d) Tester l'hypothèse H_0 : $\alpha = 0$ avec un risque de 1 %.
e) Construire un intervalle de confiance à 99% pour α.
f) Tester l'hypothèse H_0 : $\beta = 0$ avec un risque de 1 %
g) Construire un intervalle de confiance de 99% pour β.
h) Utiliser $\hat{\sigma}^2$ pour calculer un intervalle de confiance à 95% pour la résistance moyenne au cisaillement lorsque x = 24,5.
i) Utiliser $\hat{\sigma}^2$ pour calculer un intervalle de prédiction de 95% pour une seule valeur prédite de la résistance au cisaillement lorsque x = 24,5.

Solution :

La calculatrice a donné : A = 42.58180269 B = – 0.686077125 $\bar{x} = 25.96666667$

r = – 0.655567185 r² = 0.429768335 $\sum x_i^2 = 8134.26$

$\sigma_x^2 = 3.587222222$ $\sigma_y^2 = 3.928888889$ $S_{xy} = n\,\mathrm{Cov}(x\,;y) = -29.53333333$

$SSE = \sum y_i^2 - A\sum y_i - B\sum x_i y_i = 26.88452222$ $S_{yy} = n\,\sigma_y^2 = 47.14666667$

$S_{xx} = n * \sigma_x^2 = 43.046666667$

a) Estimer la droite de régression $\mu(Y \mid x) = \alpha + \beta * x =$ **42.582 – 0.686 x.**
b) Estimer $\hat{y}(24.5) = A + 24.5 * B \Rightarrow$ **$\hat{y}(24.5) = 25.77291312$**
c) $S_{yy} - B\,S_{xy} = 47.14666667 - (-0.686077125)(-29.53333333) = 26.88452222$
Evaluation de $\hat{\sigma}^2 = \frac{S_{yy} - B\,S_{xy}}{n-2} = \frac{SSE}{n-2} = 2.688452222 \Rightarrow$ **$\hat{\sigma} = 1.63965003$**
d) Test d'hypothèse sur l'ordonnée à l'origine α de la droite de régression de Y en X.
Testez l'hypothèse $H_0 = \{\alpha = 0\}$ contre $H_1 = \{\alpha \neq 0\}$ en prenant un risque de 1 %.
C'est un test bilatéral (égal contre inégal) ; le seuil critique $t_{(0.01\,;\,10)} = 3.169 \Rightarrow$
Si To > 3.169 ou To < – 3.169 alors il y a Rejet de H_0
Calcul de $Se(A) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = 1.63965003\sqrt{\frac{1}{12} + \frac{25.966666667^2}{43.046666667}} = 6.506635437$
Calcul de la statistique de test $To = \frac{A}{Se(A)} = \frac{42.58180269}{6.506635437} = 6.544466422 \Rightarrow$ **To ≈ 6.544**
Comme **$|To| = 6.544 > 2.262$** alors on doit rejeter H_0. **On conclut que $\alpha \neq 0$.**

e) Intervalle de confiance pour α : $IC_{0.99}(\alpha) = [A \pm t_{\left(\frac{\alpha}{2}\right)} \frac{\hat{\sigma}}{\sqrt{n\,S_{xx}}} \sqrt{\sum x_i^2}\ ]$

$IC_{0.99}(\alpha) = [42.58180269 \pm 3.169 \frac{1.63965003}{\sqrt{12 * 43.046666667}} \sqrt{8134.26}] =$

$[42.58180269 \pm 20.61921081]$ = **[21.96259189 ; 63.2010135]**

f) Test d'hypothèse sur la β de la droite de régression de Y en X.

$H_0 = \{\beta = 0\}$; contre $H_1 = \{\beta \neq 0\}$, avec un risque de 1 %

C'est un test bilatéral (égal contre inégal) et le seuil critique $t_{(0.01\ ;\ 10)} = 3.169$

$Se(B) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}} = \frac{\sqrt{\frac{SSE}{n-2}}}{\sqrt{S_{xx}}} = \frac{\mathbf{1.63965003}}{\sqrt{43.046666667}} = 0.249908712$ $\qquad B = -\,0.686077125$

Calcul de To $= \frac{B-\beta}{Se_{(B)}} = \frac{-\,0.686077125 - 0}{0.249908712} = -\,2.745310946 \Rightarrow$ **To ≈ – 2.7453**

Décision : comme **|To| = 2.7453 < 3.169** on accepte H_0 et on conclut que β = 0

g) Intervalle de confiance pour β : $IC_{0.99}(\beta) = [B \pm t_{\left(\frac{\alpha}{2}\right)} \frac{\hat{\sigma}}{\sqrt{S_{xx}}}]$.

$IC_{0.99}(\alpha) = [-\,0.686077125 \pm 3.169 \frac{1.63965003}{\sqrt{43.046666667}}] =$

$[-\,0.686077125 \pm 0.791960711]$ = **[–1.478037837 ; 0.105883585]**

h) Calcul de l'intervalle de confiance à 95% pour le cisaillement moyen quand x = 24.5

$t_{(0.05\ ;\ 10)} = t_{\left(\frac{\alpha}{2}\right)} = 2.228$ $\qquad \hat{y}(24.5) = 25.77291312$ $\qquad n = 12$

$E(Y|_{x}) = \hat{y}(x) \pm t_{\left(\frac{\alpha}{2}\right)} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = \hat{y}(24.5) \pm 1.312368746$

$E(Y|_{x=24.5}) = \hat{y}(24.5) \pm 2.228 * 1.63965003 \sqrt{\frac{1}{n} + \frac{(24.5-25.96666667)^2}{47.04666667}} =$

$E(Y|_{x=24.5}) = 25.77291312 \pm 1.312368746 =$ **[24.46054437 ; 27.08528186]**

i) Calcul d'un intervalle de prédiction de 95% pour une seule valeur prédite du cisaillement lorsque la contrainte x = 24.5

$E(Y|_{x}) = \hat{y}(x) \pm t_{\left(\frac{\alpha}{2}\right)} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = \hat{y}(24.5) \pm 3.881719404$

$E(Y|_{x=24.5}) = \hat{y}(24.5) \pm 2.228 * 1.63965003 \sqrt{1 + \frac{1}{n} + \frac{(24.5-25.96666667)^2}{47.04666667}} =$

$E(Y|_{x=24.5}) = 25.77291312 \pm 3.881719404 =$ **[21.89119371 ; 29.65463252]**

Solution de vérification avec EXCEL :

Rapport détaillé

Statistiques de la régression	
Coefficient de corrélation : r	0,65556719
Coefficient de détermination : r²	0,42976834
Erreur-type $\hat{\sigma}$	1,63965003
Observations n	12

Analyse de la variance pour la régression

	Degré de liberté	*Somme des carrés*	*Moyenne des carrés*	*To = Fo*	*P-Valeur*
Régression	1	$B * S_{xy}$ = 20,2621444	MSR = 20,2621444	7,5367322	0,020643715
Résidus	10	SSE = SST – $B*S_{xy}$ = 26,8845222	MSE = 2,68845222		
Total	11	SST = 47,1466667			

Droite de la régression Linéaire

	Coefficients	*Erreur-type*	*Stat t_0*	*P-valeur*	*inférieure IC = 99 %*	*supérieure IC = 99 %*
A =	42,5818027	6,50653544	6,54446642	6,5182E-05	21,9608177	63,2027877
B =	– 0,68607713	0,24990871	– 2,74531095	0,02064371	– 1,47810598	0,10595173

Exercice n° 2 :

Une étude a été réalisée sur une quantité de sucre converti par un certain procédé à différentes températures. Les données ont été codées et enregistrées comme suit :

Température, *x*	**sucre converti, *y***	**Température, *x***	**sucre converti, *y***
1.0	8.1	1.6	8.6
1.1	7.8	1.7	10.2
1.2	8.5	1.8	9.3
1.3	9.8	1.9	9.2
1.4	9.5	2.0	10.5
1.5	8.9		

a) Estimer la droite de régression linéaire de Y en X.
b) Estimer la quantité moyenne de sucre converti produite lorsque la température codée x est de 1,75.
c) évaluer $\widehat{\sigma}$
d) Tester l'hypothèse $H_0 : \alpha = 0$ avec un risque de 5 %.
e) Construire un intervalle de confiance à 95 % pour α.
f) Tester l'hypothèse $H_0 : \beta = 0$ avec un risque de 5 %.
g) Construire un intervalle de confiance à 95 % pour β.
h) Construire un intervalle de confiance à 95 % pour la quantité de sucre converti correspondant à x = 1,6 pouces.
i) Trouver un intervalle de prédiction à 0.95 de contenir non pas un paramètre mais une future valeur prédite y_0 de la variable aléatoire Y_0 lorsque x = 1,6 pouces.

Solution :

La calculatrice a donné : A = 6.413636364 B = 1.809090909 r = 0.707026443
$r^2 = 0.499886392$ $\bar{x} = 1.5$ $\sigma_x^2 = 0.1$ $\sum x_i^2 = 25.85$ $\sigma_y^2 = 0.654710743$
$S_{xx} = n * \sigma_x^2 = 1.1$ $S_{yy} = n\, \sigma_y^2 = 7.201818182$ $S_{xy} = n\, \mathrm{Cov}(x\,;\,y) = 1.99$
$SSE = \sum y_i^2 - A \sum y_i - B \sum x_i\, y_i = 3.601727273$

a) Estimer la droite de régression $\mu(Y \mid x) = \alpha + \beta * x =$ **6.414 + 1.809 x.**
b) Estimer $\hat{y}\,(1.75) = A + 1.75 * B \Rightarrow$ **$\hat{y}$ (1.75) = 9.579545455**
c) $S_{yy} - B\, S_{xy} = 7.201818182 - (1.809090909)(1.99) = 3.601727273$

Evaluation de $\hat{\sigma}^2 = \frac{S_{yy} - B\ S_{xy}}{n-2} = \frac{SSE}{n-2} = 0.400191919 \Rightarrow \boldsymbol{\hat{\sigma} = 0.632607239}$

d) Testez l'hypothèse $H_0 = \{\alpha = 0\}$ contre $H_1 = \{\alpha \neq 0\}$ en prenant un risque de 5 %.

C'est un test bilatéral (égal contre inégal) ; le seuil critique $t_{(0.05\,;\,9)} = 2.262 \Rightarrow$

Si To > 2.262 ou To < – 2.262 alors il y a Rejet de H_0

Calcul de Se(A) = $\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = 0.632607239\sqrt{\frac{1}{11} + \frac{1.5^2}{1.1}} = 0.924638017$

Calcul de la statistique de test To = $\frac{A}{Se(A)} = \frac{6.413636364}{0.924638017} = 6.936375364 \Rightarrow$ **To ≈ 6.936**

Comme | **To** | **= 6.936 > 2.262** alors on doit rejeter H_0. **On conclut donc que α ≠ 0.**

e) Intervalle de confiance pour α : $IC_{0.95}(\alpha) = [A \pm t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n\, S_{xx}}} \sqrt{\sum x_i^2}\ ]$

$IC_{0.95}(\alpha) = [6.413636364 \pm 2.262\ \frac{0.632607239}{\sqrt{11 * 1.1}} \sqrt{25.85}] =$

$[6.413636364 \pm 2.091531195]$ = **[4.322105169 ; 8.505167559]**

f) Tester l'hypothèse $H_0 = \{\beta = 0\}$; contre $H_1 = \{\beta \neq 0\}$, avec un risque de 5 %

C'est un test bilatéral (égal contre inégal) et le seuil critique $t_{(0.05\,;\,9)} = 2.262$

$Se(B) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}} = \frac{\sqrt{\frac{SSE}{n-2}}}{\sqrt{S_{xx}}} = \frac{\mathbf{0.632607239}}{\sqrt{1.1}} = 0.603167336$ B = 1.809090909

Calcul de To = $\frac{B - \beta}{Se_{(B)}} = \frac{1.809090909 - 0}{0.603167336} = 2.999318433 \Rightarrow$ **To ≈ 2.999**

Décision : comme | **To** | **≈ 2.999 > 2.262** on rejette H_0 et on conclut que β ≠ 0

g) Intervalle de confiance pour β : $IC_{0.95}(\beta) = [B \pm t_{\left(\frac{\alpha}{2}\right)} \frac{\hat{\sigma}}{\sqrt{S_{xx}}}]$.

$IC_{0.99}(\alpha) = [1.809090909 \pm 2.262\ \frac{0.632607239}{\sqrt{1.1}}] =$

$[1.809090909 \pm 1.364364514]$ = **[0.444726394 ; 3.173455423]**

h) Calcul de l'intervalle de confiance à 95% pour la quantité de sucre converti correspondant à x = 1.6

$t_{(0.05\,;\,9)} = t_{\left(\frac{\alpha}{2}\right)} = 2.262$ $\hat{y}(1.6) = 9.308181818$ n = 11

$E(Y|_{x}) = \hat{y}(x) \pm t_{\left(\frac{\alpha}{2}\right)}\, \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} = \hat{y}(1.6) \pm 0.452508517$

$E(Y|_{x = 1.6}) = \hat{y}(1.6) \pm 2.262 * 0.632607239 \sqrt{\frac{1}{11} + \frac{(1.6 - 1.5)^2}{1.1}} =$

$E(Y|_{x = 1.6}) = 9.308181818 \pm 0.452508517$ = **[8.855673301 ; 9.760690335]**

i) Calcul de l'intervalle de prédiction à 0.95 qui contient non pas un paramètre de la population mais une future valeur prédite y_0 de la variable aléatoire Y_0 lorsque x = 1.6

$t_{(0.05\,;\,9)} = t_{\frac{\alpha}{2}} = 2.262$ $y_0 = \hat{y}(1.6) = 9.308181818$ n = 11

$E(Y|_{x}) = \hat{y}(x) \pm t_{\frac{\alpha}{2}}\, \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} = \hat{y}(1.6) \pm 1.500800966$

$E(Y|_{x = 1.6}) = \hat{y}(1.6) \pm 2.262 * 0.632607239 \sqrt{1 + \frac{1}{11} + \frac{(1.6 - 1.5)^2}{1.1}} =$

$E(Y|_{x = 1.6}) = 9.308181818 \pm 1.500800966$ = **[7.807380853 ; 10.80898278]**

Solution de vérification avec EXCEL :

Rapport détaillé

Statistiques de la régression		
Coefficient de corrélation :	r =	0,70702644
Coefficient de détermination :	r^2 =	0,49988639
Erreur-type	$\widehat{\sigma}$	0,63260724
Observations	n =	11

Test sur la signification de la régression avec l'analyse de la variance

	DDL	*Somme des carrés*	*Moyenne des carrés*	*F*	*P-Valeur*
Régression	1	$B * S_{xy}$ = 3,60009091	MSR = 3,60009091	8,99591105	0,0149729
Résidus	9	SSE = 3,60172727	MSE = 0,40019192		
Total	10	SST = 7,20181818			

Test sur les coefficients de la droite de régression linéaire

	Coefficients	*Erreur-type*	*Stat t_0*	*P-valeur*	*inférieure IC = 95%*	*supérieure IC = 95%*
A =	6,413636364	0,92463802	6,936375364	6,7865E-05	4,32195985	8,50531288
B =	1,809090909	0,60316734	2,999318433	0,0149729	0,4446316	3,17355022

Exercice n° 3 :

L'examen final a donné les notes des 20 étudiants sélectionnés au hasard qui suivent un cours de statistiques et un cours de recherche opérationnelle. Supposons que les notes finales soient distribuées normalement conjointement.

Stat y	86	75	69	75	90	94	83	86	71	65
OR x	80	81	75	81	92	95	80	81	76	72
Stat y	84	71	62	90	83	75	71	76	84	97
OR x	85	72	65	93	81	70	73	72	80	98

a) Trouvez la droite de régression liant la note finale des statistiques à la note finale de RO.
b) Estimer le coefficient de corrélation.
c) Testez la signification de la régression en utilisant un risque de 5%.
d) Testez l'hypothèse que $\rho = 0$ en utilisant un risque de 5%
e) Testez l'hypothèse que $\rho = 0,5$ en utilisant un risque de 5%
f) Construire un intervalle de confiance de 95% pour le coefficient de corrélation.

Solution :

La calculatrice a donné : $A = -0.028041126$ $B = 0.99098678$ $r = 0.903339735$

$r^2 = 0.816022678$ $\bar{x} = 80.1$ $\sigma_x^2 = 74.89$ $\sum x_i^2 = 129\,818$

$S_{xx} = n * \sigma_x^2 = 1497.8$ $S_{yy} = n\, \sigma_y^2 = 1802.55$ $S_{xy} = n\, Cov(x\, ;\, y) = 1484.3$

$SSE = \sum y_i^2 - A \sum y_i - B \sum x_i\, y_i = 331.6283215$ $\sigma_y^2 = 90.1275$

a) L'estimation de la droite de régression $\mu(Y \mid x) = \alpha + \beta * x =$ **$- 0.028 + 0.991\, x$.**
b) L'estimation du coefficient de corrélation est : **$r = 0.903339735$**

c) Test de la signification de la régression

Calcul de SST = $S_{yy} = n\,\sigma_y^2$ = 1802.55

Calcul de SSR = B S_{xy} = 0.99098678 ∗ 1484.3 = 1470.921678 ⇒ MSR = SSR/1

Calcul de SSE = SST – SSR = 331.6283224 ⇒ MSE = $\frac{\text{SSE}}{\text{n}-2}$ = 18.42379564

Calcul de la statistique de test To = Fo = $\frac{\text{MSR}}{\text{MSE}} = \frac{1470.921678}{18.42379564}$ = 79.83814556

Le seuil critique $f_{(\alpha\,;\,\nu_1\,;\,\nu_2)} = f_{(0.05\,;\,1\,;\,18)}$ = 4.4139

Test de la signification de la régression avec EXCEL (pour vérifier les calculs)

	DDL	*Somme des carrés*	*Moyenne des carrés*	*To = Fo*	*F critique*
Régression	1	B ∗ S_{xy} = 1470,92168	MSR = 1470,92168	79,83814556	4,4139
Résidus	18	SSE = 331,6283224	MSE = 18,42379564		
Total	19	SST = 1802,55			

$H_0 = \{\beta = 0\}$ contre $H_1 = \{\beta \neq 0\}$ avec risque de 5 %. C'est un test bilatéral

Le seuil critique $t_{(0.05\,;\,18)}$ = 2.101 si ∣To∣ > 2.101 on rejette H_0.

Evaluation de $\hat{\sigma}^2 = \frac{S_{yy} - B\ S_{xy}}{n-2} = \frac{\text{SSE}}{n-2}$ = 18.42379564 ⇒ $\boldsymbol{\hat{\sigma}}$ **= 4.292294915**

Evaluation de Se(B) = $\frac{\hat{\sigma}}{\sqrt{S_{xx}}} = \frac{4.292294915}{\sqrt{1497.8}}$ = 0.11090794

Calcul de To = $\frac{B}{\sqrt{\frac{\hat{\sigma}^2}{Sxx}}} = \frac{0.99098678}{0.11090794}$ = 8.935219386 ⇒ **On rejette H_0.**

d) Test de l'hypothèse que ρ = 0.

$H_0 = \{\rho = 0\}$; contre $H_1 = \{\rho \neq 0\}$, avec un risque de 5 %

C'est un test bilatéral (égal contre inégal) et le seuil critique $t_{(0.05\,;\,18)}$ = 2.101

Calcul de To = $\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.903339735}{\sqrt{\frac{1-0.903339735^2}{20-2}}}$ = 8.935219392 ⇒ **To ≈ 8.935**

Décision : comme **∣To∣ = 8.935 > 2.101** on rejette H_0 et on conclut que ρ ≠ 0

e) Test de l'hypothèse que ρ = 0.

$H_0 = \{\rho = 0.5\}$; contre $H_1 = \{\rho \neq 0.5\}$, avec un risque de 5 % et n = 20

La procédure d'un test d'hypothèse de ce genre est quelque peu complexe.

Mais il s'avère que pour des grands échantillons (n ≥ 20 ou 25), la statistique Z = arctanh R = ½ Ln$\frac{1+R}{1-R}$) suit sensiblement une loi normale de moyenne et de variance :

$$\mu_Z = \text{arctanh } R = \tfrac{1}{2}\,\text{Ln}\,\frac{1+R}{1-R} \quad ; \quad \sigma_Z^2 = \frac{1}{n-3} \quad \text{respectivement.}$$

Par conséquent pour tester $H_0 = \{\rho = 0.5\}$; contre $H_1 = \{\rho \neq 0.5\}$, on peut utiliser la statistique de test To = Z_0 = (arctanh R – arctanh ρ_0)$\sqrt{n-3}$; le seuil critique étant $z_{(\frac{\alpha}{2})}$ ⇒

Si la statistique de test ∣To∣ = ∣Z_0∣ > $z_{(\frac{\alpha}{2})}$ on rejette H_0.

Calcul de To = [½ Ln $\frac{1+R}{1-R}$ – ½ Ln $\frac{1+R}{1-R}$] $\sqrt{n-3}$ = [½ Ln $\frac{1+0.903339735}{1-0.903339735}$ – ½ Ln $\frac{1+0.5}{1-0.5}$] $\sqrt{20-3}$ =

[1.490081489 – 0.549306144] * 4.123105626 = 3.878916114 ≈ **3.879**

On a la statistique de test ∣ To ∣ = ∣ Z_0 ∣ = 3.88 > $z_{\left(\frac{\alpha}{2}\right)}$ = 1.96 ⇒ **on rejette H_0.**

f) Construction d'un intervalle de confiance de 95% pour le coefficient de corrélation.

$IC_{0.95}$ (ρ) = tanh (arctanh r – $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$) < ρ < tanh (arctanh r + $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$) avec : Tanh u = $\frac{e^{u} - e^{-u}}{e^{u} + e^{-u}}$

Arctanh r = ½ Ln $\frac{1+r}{1-r}$ = ½ Ln $\frac{1+0.903339735}{1-0.903339735}$ = 1.490081489 et $\frac{t_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$ = 0.409885206

arctanh r – $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$ = 1.080196282   arctanh r + $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$ = 1.899966695

$IC_{0.95}$ (ρ) = tanh (arctanh ½ Ln $\frac{1+r}{1-r}$ – $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$) < ρ < tanh (arctanh ½ Ln $\frac{1+r}{1-r}$ + $\frac{z_{\left(\frac{\alpha}{2}\right)}}{\sqrt{n-3}}$)

$IC_{0.95}$ (ρ) = tanh (1.080196282) < ρ < tanh (1.899966695) ⇒

tanh (1.080196282) = $\frac{e^{1.080196282} - e^{-1.080196282}}{e^{1.080196282} + e^{-1.080196282}} = \frac{2.945257595 - 0.339528875}{2.945257595 + 0.339528875}$ =

$\frac{2.60572872}{3.284786471}$ = 0.793271874

tanh (1.899966695) = $\frac{e^{1.899966695} - e^{-1.899966695}}{e^{1.899966695} + e^{-1.899966695}} = \frac{6.685671772 - 0.1495736}{6.685671772 + 0.1495736} = \frac{6.536098172}{6.835245373}$

= 0.956234606 ⇒ **[0.793271874 < ρ < 0.956234606]**

Les calculs ont donné r = 0.903339735

Rapport Excel détaillé

Statistiques de la régression	
Coefficient de corrélation = r	0,90333974
Coefficient de détermination = r²	0,81602268
Erreur-type = $\hat{\sigma}$	4,29229492
Observations = n	20

<u>Test sur les coefficients de la droite de régression linéaire</u>

	Coefficients	*Erreur-type*	*Stat t_0*	*P-valeur*	*inférieure = IC 95%*	*supérieure = IC 95%*
A =	– 0,02804113	8,93542266	– 0,0031382	0,9975306	– 18,8006675	18,7445853
B =	0,99098678	0,11090794	8,93521939	4,9037E-08	0,75797784	1,22399572

Exercice n° 4 :

Les valeurs x et y des variables aléatoires X et Y sont données dans le tableau ci-dessous :

x	23.1	32.8	31.8	32	30.4	24	39.5	24.2
y	10.5	16.7	18.2	17	16.3	10.5	23.1	12.4
52.5	37.9	30.5	25.1	12.4	35.1	31.5	21.1	27.6
24.9	22.8	14.1	12.9	8.8	17.4	14.9	10.5	16.1

a) Estimer la corrélation entre Y et X.
b) Testez l'hypothèse que $\rho = 0$ en utilisant un risque de 5%.
c) Ajuster un modèle de régression linéaire pour cet échantillon et tester la signification de la régression avec un risque de 5%.
d) Quelles conclusions pouvez-vous tirer ? Comment le test de signification de la régression est-il lié au test sur ρ dans la partie (b)?
e) Testez l'hypothèse $H_0 : \alpha = 0$ contre $H_1 : \alpha \neq 0$ et tirez des conclusions. Utilisez un risque de 5%.
f) Analyser les résidus et commenter l'adéquation du modèle.

Solution :

La calculatrice a donné : A = 0.725380435 B = 0.498081197 r = 0.933202938

$r^2 = 0.870867723$ $\bar{x} = 30.08823529$ $\sigma_x^2 = 72.85397924$ $\bar{y} = 15.71176471$

$\sigma_y^2 = 20.75397924$ $\sum x_i^2 = 16\,628.65$ $S_{xx} = n * \sigma_x^2 = 1238.517647$

$S_{yy} = n\,\sigma_y^2 = 352.8176471$ $S_{xy} = n\,\mathrm{Cov}(x\,;\,y) = 616.8823529$

$SSE = \sum y_i^2 - A \sum y_i - B \sum x_i\, y_i = 45.5601459$

a) L'estimation du coefficient de corrélation est : **r = 0.933202938**

b) Test de l'hypothèse que $\rho = 0$.

$H_0 = \{\rho = 0\}$; contre $H_1 = \{\rho \neq 0\}$, avec un risque de 5 %

C'est un test bilatéral (égal contre inégal) et le seuil critique $t_{(0.025\,;\,15)} = 2.131$

Calcul de $To = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \dfrac{0.933202938}{\sqrt{\frac{1-0.933202938^2}{17-2}}} = 10.05783076 \Rightarrow \mathbf{To \approx 10.058}$

Décision : comme $|\mathbf{To}| = \mathbf{10.058 > 2.131}$ on rejette H_0 et on conclut que $\rho \neq 0$

c) L'estimation de la droite de régression $\mu(Y \mid x) = \alpha + \beta * x =$ **0.72538 + 0.49808 x.**

Test de la signification de la régression avec un risque de 5 %.

Calcul de $SST = S_{yy} = n\,\sigma_y^2 = 352.8176471$

Calcul de $SSR = B\,S_{xy} = 0.498081197 * 616.8823529 = 307.2575012 \Rightarrow MSR = SSR/1$

Calcul de $SSE = SST - SSR = 45.56014595 \Rightarrow MSE = \frac{SSE}{n-2} = 3.037343063$

Calcul de la statistique de test $To = Fo = \frac{MSR}{MSE} = \frac{307.2575012}{3.037343063} = 101.1599595$

Le seuil critique $f_{(\alpha\,;\,\nu_1\,;\,\nu_2)} = f_{(0.05\,;\,1\,;\,15)} = 4.5431$

Test de la signification de la régression avec EXCEL (*pour vérifier les calculs*)

	DDL	*Somme des carrés*	*Moyenne des carrés*	*To = Fo*	*F critique*
Régression	1	$B * S_{xy}$ = 307.2575012	MSR = 307.2575012	101.1599596	4,5431
Résidus	15	SSE = 45.56014595	MSE = 3.037343063		
Total	16	SST = 352.8176471			

d) Les conclusions tirées indiquent que le coefficient de corrélation linéaire est non nul et que le test de la signification de la régression est très significatif ce qui valide le premier test sur la corrélation.

Le test de signification de la régression de la question (c) est lié au test sur ρ dans la question (b) par le fait qu'il y a une relation de carré entre les deux statistiques de test observé : la statistique de test observé To au carré de la question (b) = To^2 = $(10.05783076)^2$ = la statistique de test observé de la question (c) To = F = 101.1599595

e) Testez l'hypothèse $H_0 = \{\alpha = 0\}$ contre $H_1 = \{\alpha \neq 0\}$ en prenant un risque de 5 %.

C'est un test bilatéral (égal contre inégal) ; le seuil critique $t_{(0.05\,;\,15)} = 2.131 \Rightarrow$

Si To > 2.131 ou To < – 2.131 alors il y a Rejet de H_0

Calcul de Se(A) = $\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}} = 1.742797482\sqrt{\frac{1}{17}+\frac{30.08823529^2}{1238.517647}} = 1.548816119$

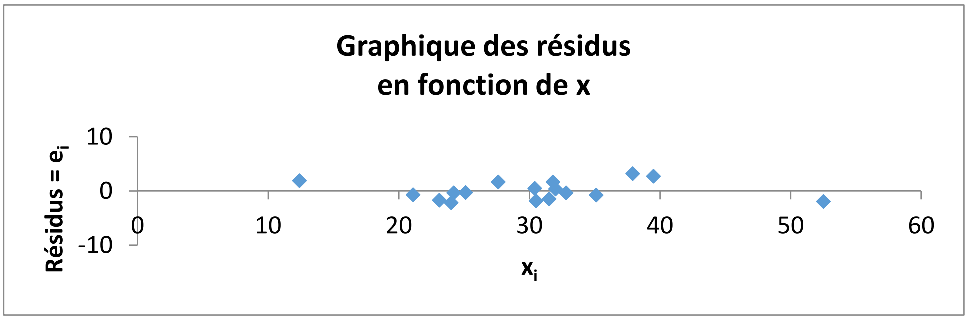
Calcul de la statistique de test To = $\frac{A}{Se(A)} = \frac{0.725380435}{1.548816119} = 0.468345097 \Rightarrow$ **To ≈ 0.468**

Comme $|$**To**$|$ = **0.468 < 2.131** alors on doit accepter H_0. **On conclut donc que α = 0**

Test sur les coefficients de la droite de régression linéaire

	Coefficients	*Erreur-type*	*Stat t_0*	*P-Valeur*	*Inférieure IC = 95 %*	*supérieure IC = 95 %*
A =	0,725380436	1,548816118	0,4683451	0,64627111	– 2,57584298	4,02660385
B =	0,498081198	0,049521732	10,0578308	4,6316E-08	0,39252812	0,60363427

f) Analyse des résidus et commentaire sur l'adéquation du modèle.



Il y a une parfaite adéquation du modèle choisi (**y = 0.72538 + 0.49808 x**) car les résidus sont répartis d'une manière aléatoire dans le graphique et $|$**e_i**$|$ < **2** sauf les observations 2 et 10.

6	12,6793292	– 2,17932918	10	19,6026578	3,19734217