6

Question n° 1 : (8 points)

Soit Z_1, Z_2, Z_3, Z_4, Z_5 un échantillon aléatoire de la population $Z \sim N(0,1)$.

1,5

a) (2 points) On définit $W = 4Z_1 - Z_2 + 3Z_3 - 3Z_4 + Z_5$. Calculer $P(W \leq -9)$.

$E(W) = 4E(Z_1) - E(Z_2) + 3E(Z_3) - 3E(Z_4) + E(Z_5) = 0 = \mu_W$

$V(W) = 16V(Z_1) + V(Z_2) + 9V(Z_3) + 9V(Z_4) + V(Z_5) = 36 = \sigma_W^2$

$W \sim N(\mu_W, \sigma_W^2) \rightarrow N(0, 36)$

$P(W \leq -9) = P\left(\frac{W-0}{6} \leq \frac{-9}{6}\right) = P\left(Z \leq -\frac{3}{2}\right)$

$= P(Z \leq -1{,}5)$

$= \cancel{\Phi(1{,}5) - 0{,}5}$ $1 - \Phi(1.$

$= 0{,}93319 - 0{,}5$

$= 0{,}43319$

b) (2 points) Pour vérifier si la moyenne théorique de Z est bien égale à 0, on utilise la moyenne échantillonnale $\bar{Z}$ comme estimateur du paramètre $\theta = E(Z)$. Calculer l'erreur quadratique moyenne de $\bar{Z}$.

0.5

$EQM = E((\hat{\theta} - \theta)^2)$

$= E(\hat{\theta}^2 + \theta^2 - 2\hat{\theta}\theta)$

$= E(\hat{\theta}^2) + E(\theta^2) - 2E(\hat{\theta})E(\theta)$

$= 0$

Question n° 1 : (suite)

c) **(2 points)** Déterminer la valeur de $x > 0$ telle que

2

$$P\left(-x \leq \frac{Z_1}{\sqrt{Z_2^2 + Z_3^2}} \leq x\right) = 0{,}90.$$

$W = Z_2^2 + Z_3^2$

$W \sim \chi^2_2$

$T = \dfrac{Z_1}{\sqrt{\dfrac{W}{2}}} \sim t_2$

$\Leftrightarrow 1 - P(T > x\sqrt{2}) - P(T < -x\sqrt{2}) = 0{,}90$

$\Leftrightarrow -(P(T > x\sqrt{2}) + P(T < -x\sqrt{2})) = 0{,}90$

$\Leftrightarrow 2P(T > t_{\alpha;2}) = 0{,}10$

$\Leftrightarrow P(T > t_{\alpha;2}) = 0{,}05 = \alpha$

$x\sqrt{2} = t_{0{,}05;2} = 2{,}92$

$x = \dfrac{2{,}92}{\sqrt{2}} = 2{,}064$ ✓

d) **(2 points)** Déterminer la valeur de d telle que $P(Z_1^2 + Z_2^2 + Z_3^2 \leq d) = 0{,}95$.

2

$W = Z_1^2 + Z_2^2 + Z_3^2$

$W \sim \chi^2_3$

$P(W \leq d) = 0{,}95$

$\Rightarrow 1 - P(W > d) = 0{,}95$

$\Rightarrow P(W > d) = 0{,}05 = \alpha$

$d = \chi^2_{0{,}05;3} = 7{,}81$ ✓

Question n° 2 : (7 points)

6

Dans un cours d'informatique, le professeur s'intéresse au temps X (en heures) nécessaire à la réalisation d'un devoir de programmation. Le tableau d'effectifs suivant présente un échantillon aléatoire de X obtenu dans un sondage auprès de $n = 100$ étudiants.

Classe	$[0,2)$	$[2,5)$	$[5,10]$
Effectif	25	18	57

a) (2 points) Calculer approximativement la moyenne et la médiane de cet échantillon en utilisant le point milieu des classes.

2

$$\bar{x} = \frac{1}{n}\sum y_i = \frac{1}{100}(25 \times 1 + 18 \times 3{,}5 + 7{,}5 \times 57)$$

$$= 5{,}065$$

médiane $= 7{,}5$

b) (2 points) Soit p le paramètre désignant la probabilité de compléter le devoir en moins de 5 heures. Donner une estimation ponctuelle et un intervalle de confiance de niveau 95% pour p.

2

$n = 100$ $\quad \alpha = 0{,}05$

$z_{\alpha/2} = 1{,}960$ $\quad \alpha/2 = 0{,}025$

$$\hat{p} = \frac{25+18}{100} = 0{,}43$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow [0{,}332 \; ; \; 0{,}527]$$

Question n° 2 : (suite)

c) (3 points) Peut-on rejeter l'hypothèse que X suit une loi uniforme sur l'intervalle $[0, 10]$? Formuler les hypothèses testées H_0 et H_1. Utiliser un test d'hypothèse approprié au seuil $\alpha = 0{,}05$.

Test Khi deux

H_0 : X suit une loi uniforme sur l'intervalle $[0;10]$

H_1 : X ne suit pas une loi uniforme sur l'intervalle $[0,10]$

Tableau E_i	$[0, 2)$	$(2, 5)$	$[5 \quad 10]$
	~~33,33~~	~~33,33~~	~~33,33~~

$$\chi_0^2 = \frac{(25-33{,}33)^2}{33{,}33} + \frac{(18-33{,}33)^2}{33{,}33} + \frac{(57-33{,}33)^2}{33{,}33}$$

$= 25{,}94$

$\chi_{0,05;2} = 5{,}99$

$\chi_0^2 > \chi_{0,05;2}^2$

donc on rejete H_0

Question n° 3 : (4 points)

3/4

On a dressé un tableau de contingence afin de déterminer s'il existe un lien entre deux variables qualitatives X et Y. Ce tableau possède deux lignes et deux colonnes (i.e. chaque variable a 2 modalités).

a) **(1 point)** Écrire les hypothèses testées H_0 et H_1.

H_0 : X et Y sont indépendant

H_1 : X et Y sont dépendant ✓

b) **(2 points)** La statistique du test est égale à 2,4108. Doit-on rejeter H_0 au seuil $\alpha = 0{,}05$?

$v = (2-1)(2-1) = 1$

$\chi^2_{0,05;1} = 3{,}84$ $\qquad$ $\chi^2_{0,05;1} > 2{,}4102 = \chi^2_0$

On accepte H_0 ✓

c) **(1 point)** À l'aide des centiles fournis à l'annexe, déterminer une borne supérieure et une borne inférieure pour la valeur-P du test.

$1{,}32 < \text{Valeur-P} \leq 2{,}71$

0

Question n° 4 : (12 points)

Deux machines (M_1 et M_2) servent à produire un type de pièce dont le diamètre est une caractéristique importante. Le tableau ci-dessous présente les mesures de diamètre et quelques statistiques pour deux échantillons de pièces produites par les deux machines (un échantillon par machine)

Machine	Mesures de diamètre en mm							Taille	Moyenne	Écart-type
M_1	39,5	41,7	38,1	39,5	40,1			5	39,78	1,30
M_2	38,6	40,1	40,2	42,8	39,5	40,8	41,9	7	40,56	1,42

On suppose que les deux échantillons sont indépendants et que les mesures de diamètre X_1 et X_2 des pièces des deux machines suivent respectivement des lois normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$.

Important *: formuler les hypothèses H_0 et H_1 pour chaque test.*

a) (2 points) Calculer un intervalle de confiance pour la variance du diamètre des pièces de la machine M_1 au niveau de confiance 95%.

+2

$X_1 \sim N(\mu_1, \sigma_1^2)$

$S^2 = (1,30)^2$
$\bar{X} = 39,78$
$n = 5$
σ^2 inconnue
$\alpha = 0,05$

$\Rightarrow \left[\dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \quad \dfrac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}} \right]$

$[0,60 \quad 14,08]$

b) (2 points) On compte effectuer une sixième mesure de diamètre d'une nouvelle pièce produite par M_1. Calculer un intervalle de prévision pour cette mesure à un niveau de confiance 95 %; interpréter brièvement le résultat.

+2

$X_1 \sim N(\mu_1, \sigma_1^2)$

I.P. $\bar{X} \pm t_{\alpha/2; n-1} \sqrt{S^2(1 + \frac{1}{n})}$

σ_1^2 inconnue

$\bar{X} = 39,78$
$\alpha = 0,05$
$S^2 = (1,30)^2$
$n = 5$

$\Rightarrow [35,82 \quad 43,73]$

la prochaine valeur de M_1 sera comprise dans I.P

Question n° 4 : (suite)

c) **(2 points)** Au seuil $\alpha = 0{,}05$, peut-on conclure que les pièces de M_1 présentent en moyenne un diamètre inférieur à 40 mm ?

1/2

$X_1 \sim N(\mu_1, \sigma_1^2)$

$H_0 : \mu_1 \leq 40$ (.015

$H_1 : \mu_1 > 40$

$T_0 = \dfrac{\bar{X} - 40}{S/\sqrt{n}} = -0{,}37$

$S = 1{,}3$

$n = 5$

$\alpha = 0{,}05$

$\bar{x} = 39{,}78$

σ^2 inconnue

$t_{0{,}05;4} = 2{,}14$

$T_0 < t_{0{,}05;4}$

on accepte H_0

conclu (-0,5)

d) **(3 points)** Au seuil $\alpha = 0{,}05$, peut-on dire que les variances des diamètres des pièces sont différentes pour les deux machines M_1 et M_2 ?

$\frac{2{,}5}{3}$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$F_0 = \dfrac{S_1^2}{S_2^2} = \dfrac{(1{,}30)^2}{(1{,}42)^2} = 0{,}838$

$\alpha = 0{,}05$

$S_1 = 1{,}30$

$S_2 = 1{,}42$

$F_{0{,}975;4,6} = \dfrac{1}{F_{0{,}025;6,4}} = \dfrac{1}{9{,}20} \simeq 0{,}10$

$F_{0{,}025;4,6} = 6{,}23$

$\begin{cases} 0{,}838 > 0{,}10 \\ 0{,}838 < 6{,}23 \end{cases} \Rightarrow$ on accepte H_0

conclu (-0,5)

Question n° 4 : (suite)

e) (3 points) Au seuil $\alpha = 0{,}05$, peut-on dire que les moyennes des diamètres des pièces sont différentes pour les deux machines M_1 et M_2 ?

$\frac{25}{3}$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$ H_0 $\mu_1 = \mu_2$

H_1 $\mu_1 \neq \mu_2$

(on utilise le test avec $\sigma_1^2 = \sigma_2^2$ vu que dans la question précédente on testen et accepten H_0)

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 1{,}88$$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{39{,}78 - 40{,}56}{\sqrt{1.88\left(\frac{1}{5}+\frac{1}{7}\right)}} = -0{,}971$$

$\alpha = 0{,}05$

$t_{0{,}025;10} = 2{,}23$

$|T_0| = 0{,}971 < 2{,}23$

$\hookrightarrow$ on accepte H_0

correct

0,5

9,5

Question n° 5 : (14 points)

Un ingénieur étudie le temps de transfert de fichiers de grande taille sur un réseau informatique. Il construit un modèle de régression linéaire simple d'équation

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

où x représente la taille d'un fichier (en Go, giga-octets) et Y, le temps de transfert (en minutes). De plus, β_0 et β_1 sont des paramètres, et ε une erreur aléatoire que l'on suppose de loi $N(0,\ \sigma^2)$. Les données recueillies pour estimer les paramètres du modèle sont fournies dans le tableau suivant.

i	1	2	3	4	5	6	7	8	9	10
x_i	50	50	100	100	150	150	200	200	250	250
y_i	7,4	7,3	14,9	15,1	21,9	22,2	29,4	29,1	36,5	37,0
i	11	12	13	14	15	16	17	18	19	20
x_i	300	300	350	350	400	400	450	450	500	500
y_i	43,8	44,3	51,7	52,8	61,5	59,2	66,9	70,7	73,6	78,4

L'ingénieur trouve que

$$\bar{x} = 275{,}00\ ; \quad S_{xx} = 412\,500\ ; \quad SS_R = 9\,656{,}2125;$$
$$\bar{y} = 41{,}185\ ; \quad S_{yy} = 9\,685{,}0255.$$

4 a) **(4 points)** Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$. Donner l'équation de la droite des moindres carrés.

$SS_R = \hat{\beta}_1^2 S_{xx}$

$\hat{\beta}_1 = \sqrt{\frac{SS_R}{S_{xx}}} = 0{,}153$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\hat{\beta}_0 = -0{,}89$

$E(\varepsilon) = 0$

$\hat{Y} = -0{,}89 + 0{,}153x$

Question n° 5 : (suite)

b) **(3 points)** Compléter le tableau d'analyse de la variance ci-dessous. Quelle conclusion peut-on en tirer au seuil $\alpha = 0{,}05$?

2

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	F_0
Régression	9 656,2125 SSR	1	9656,2125	6035,132
Erreur	28,813	18	1,6	
Totale	9 685,0255 S_{YY}	19		

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0{,}05$

$F_{0{,}05;1,18} = \cancel{5{,}98}\ 4{,}11 < F_0$

On rejette H_0 régr. signif.

c) **(2 points)** En négligeant des délais minimaux indépendants de la taille des fichiers, l'ingénieur s'attendrait à ce que $\beta_0 = 0$. Peut-on rejeter cette hypothèse ? Justifier à l'aide d'un test approprié au seuil $\alpha = 0{,}05$.

2

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$$T_0 = \frac{\hat{\beta}_0 - 0}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)}} = -1{,}456$$

$t_{\alpha/2, n-2} = t_{0{,}025;18} = 2{,}10$

$|T_0| = 1{,}456 < 2{,}10$

On accepte H_0

Question n° 5 : (suite)

d) **(2 points)** Quelle valeur ponctuelle de temps de transfert l'ingénieur peut-il prévoir pour un fichier de 5 giga-octets ? Interpréter ce résultat.

$n = 5$

0,5 $\hat{Y} = -0,89 + 0,153n = -0,125$

un transfert de donnée instantané

e) **(1 point)** Calculer le coefficient de détermination R^2 et interpréter ce résultat.

1 $R^2 = \frac{SS_R}{S_{YY}} = 0,99$

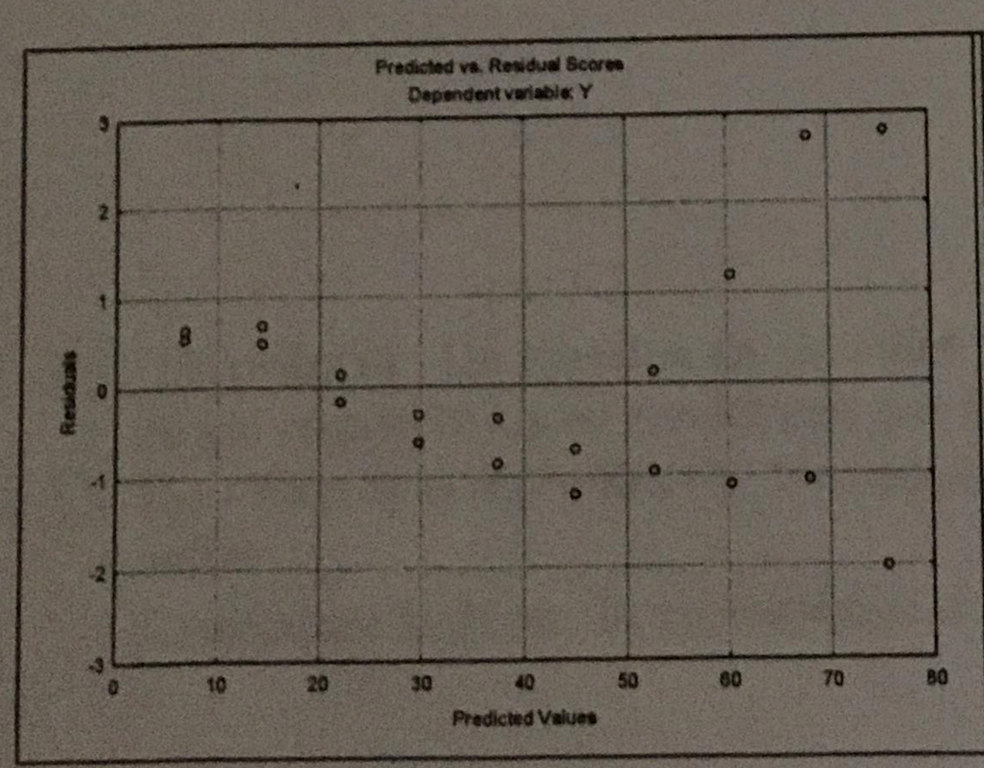
le modèle de régression est adéquat

f) **(2 points)** Les graphiques de la page suivante ont été produits avec Statistica. Que peut-on dire sur le modèle de régression en analysant les résidus ?

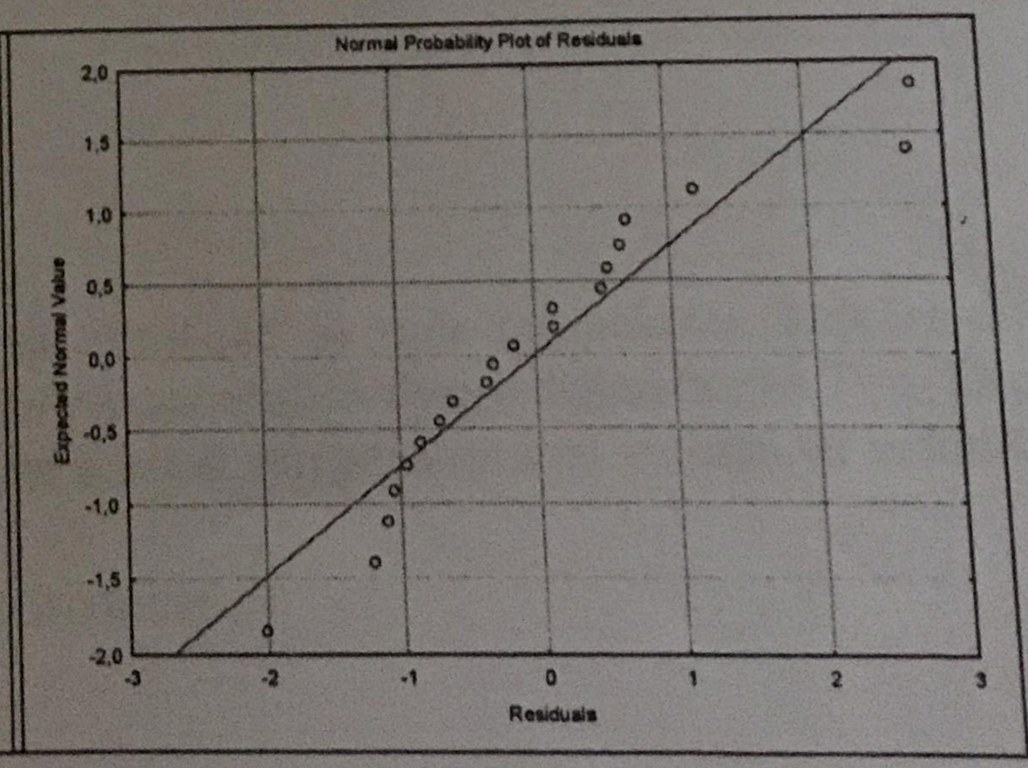
0

le modèle de regression est adéquat

Question n° 5 : (suite)



1) Résidus en fonction des valeurs prédites.



2) Graphe de probabilité normale des résidus.

Question n° 6 : (5 points)

0

Cette question est constituée de deux parties indépendantes l'une de l'autre.

a) **(2 points)** Des clients se présentent à un guichet automatique en formant une file d'attente $M/M/1$ avec $\lambda = 27$ clients par heure. On estime que 90% des clients sont présents plus de 3 minutes au total pour l'attente et le service.

0

Déterminer le nombre moyen de clients dans le système à l'équilibre, $\bar{N}$.

b) **(3 points)** Un système en parallèle est constitué de trois composants dont les durées de vie respectives sont modélisées par les variables aléatoires indépendantes $T_1 \sim \text{Exp}(2)$, $T_2 \sim \text{Exp}(5)$ et $T_2 \sim \text{Exp}(8)$. On suppose que ces composants sont utilisés en redondance active.

0

Déterminer la durée de vie moyenne du système.