QUESTION # 1 (10 points)

Le diamètre extérieur X d'une tige en acier est une variable aléatoire uniformément distribuée sur l'intervalle $[\theta\ , 3\theta]$. Afin d'estimer le paramètre réel θ, on propose les deux estimateurs ponctuels $\hat{\theta}1$ et $\hat{\theta}2$ calculés à partir d'un échantillon aléatoire $\{X_1, X_2,\ldots,X_n\}$, d'une taille n suffisamment grande, comme suit :

$$\hat{\theta}1 = \frac{\overline{X}}{2} \quad \text{et} \quad \hat{\theta}2 = \frac{2X_1 + nX_2 - X_{n-1} - X_n}{2n}$$

a) (2.5 points) Calculer pour chaque estimateur son biais et son erreur quadratique moyenne.

b) (1 point) Compte tenu des résultats obtenus en a), lequel des deux estimateurs est préférable?

Afin d'estimer le paramètre θ par un intervalle de confiance bilatéral $[L, U]$ au seuil 1-α, on utilise le même échantillon que ci haut (n toujours grand).

c) (1.5 points) Quelle est la loi de $\overline{X}$, quels sont ses paramètres ?

d) (2 points) À l'aide de la réponse en c) et en utilisant $\hat{\theta}1$ comme meilleur estimateur ponctuel de θ, calculer les deux bornes L et U de l'intervalle de confiance.

Supposons maintenant que X suit une loi normale $N(\mu,\sigma^2= 4)$. On prélève un échantillon aléatoire de taille n = 16, les calculs montrent que la moyenne et la variance de l'échantillon sont respectivement $\overline{X}$ = 10.8 et S^2 = 12.96

e) (1 point) Construire un intervalle de confiance bilatéral au seuil 95% pour la moyenne μ.

f) (1 points) Que doit être la taille n' de l'échantillon à prélever si on veut obtenir un intervalle de confiance 2 fois plus précis que celui construit en e) ?

g) (1 point) On prélève une observation supplémentaire et on constate qu'elle n'appartient pas à l'intervalle calculé en g). Cela semble-t-il normal ? Justifier !

<u>QUESTION # 2</u> (10 points)

Une entreprise de développement de films envisage de remplacer sa développeuse actuelle A par une nouvelle développeuse B. Le temps de développement d'un rouleau est important et constitue le critère sur lequel l'entreprise se base pour prendre une décision quant au changement de la développeuse. Dans tout le problème, on considère que le temps de développement suit une loi normale et que tous les tests doivent être effectués avec $\alpha = 0.05$.

Dans un premier temps, la compagnie décide d'effectuer le remplacement dès que le temps de développement moyen est supérieur à 9 min. Un échantillon aléatoire de 16 rouleaux est soumis à la développeuse A et on obtient, pour cet échantillon, un temps moyen de développement de 9.5 min et un écart type de 1.5 min.

a) (1.5 points) Devrait-on effectuer le changement de développeuse ?

b) (1 point) L'affirmation suivante au sujet du test effectué en a) est-elle **VRAIE** ou **FAUSSE** ?

« Si le temps de développement moyen est de 9 min, la probabilité que le test mène à un remplacement de développeuse est de 95% »

c) (1.5 points) Si la moyenne de développement est en réalité 9.9 min, quelle est, approximativement, la probabilité que le remplacement soit effectué ?

L'entreprise s'intéresse maintenant à la développeuse B et décide de la mettre en marche (à la place de A) si son temps moyen de développement est inférieur à celui de A. Ainsi, on soumet 16 rouleaux à B et on obtient un temps moyen de développement de 8.9 min et un écart type de 1.2 min. Afin de comparer les deux développeuses, on utilise les mêmes résultats obtenus avec le premier échantillon soumis à la développeuse A.

d) (1.5 points) Peut-on affirmer que les deux développeuses ont la même variabilité ?

e) (1.5 points) Devrait-on mettre la développeuse B en marche ?

f) Si l'entreprise décide d'effectuer le remplacement seulement si le rapport :

(temps moyen de A/ temps moyen de B) > 1,2

(i) **(1 point)** Formuler les hypothèses H_0 et H_1 du test à effectuer.
(ii) **(2 points)** Devrait-on effectuer le remplacement sous cette nouvelle condition ? (On effectue le test avec des variances sont égales)

QUESTION # 3 (10 points)

Afin d'étudier l'impact du pourcentage de concentration en bois dur dans la pâte originale (X) sur la résistance du papier produit par cette pâte (Y), on effectue n tests aléatoires. À chaque test i, on obtient la valeur x_i (de la variable X), ainsi que la valeur y_i (de la variable Y) correspondante. On s'intéresse au modèle linéaire simple de la forme :

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \varepsilon \quad \text{..........(I)}$$

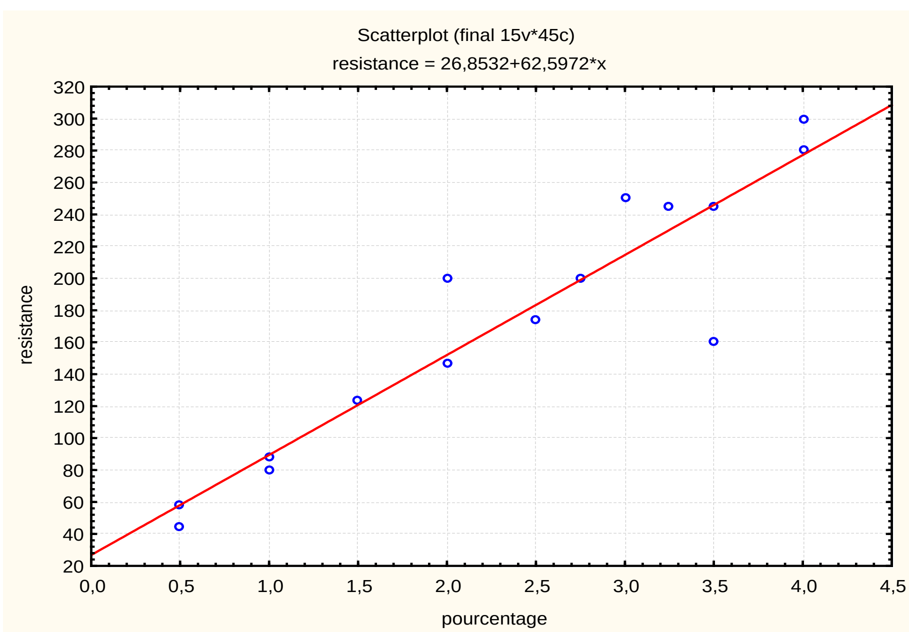
ε représente les résidus du modèle.

On utilise le logiciel ***STATISTICA*** afin de tracer le nuage de points $\{(x_i, y_i)\}$ et ajuster le modèle (I) et on obtient le graphique et le tableau ci-dessous (Graphe 3.1 et Tableau 3.1)

Tableau3.1

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(13)	p-level
Intercept			a	17,41664	c	0,147103
pourcentage	0,933685	0,099318	b	6,65861	d	e

Graphe 3.1



a) (2 points) Remplir les cases (a), (b), (c) et (d) du tableau.

b) (1.5 points) Sans calculer la valeur de la case (e) du tableau, peut-on dire, au seuil critique 0.05, que la variable X est significative ? Calculer ensuite la valeur de la case (e). Que peut-on conclure ?

c) (1.5 points) Calculer un intervalle de confiance bilatéral au seuil critique 0.05 pour le coefficient β_0. Existe-t-il un lien entre cet intervalle et p-value(β_0) ?

d) (2.5 points) En utilisant la valeur de $SS_E = 12080$, trouver les valeurs de S_{XX}, S_{XY} et S_{YY}

e) (1 point) Calculer le coefficient de corrélation entre X et Y. Interpréter brièvement.

f) (1.5 points) L'idée d'améliorer le modèle (I) en incorporant le terme X^2 semble-t-elle raisonnable ? Que doit-on vérifier avant d'entreprendre une telle démarche ?

QUESTION # 4 (10 points)

Un ingénieur mécanicien analyse le fini de surface des pièces métalliques usinées (F) sur un tour et sa relation avec la vitesse de rotation (V : *tour/min*) ainsi que la profondeur de la coupe (P : *mm*). Il existe 4 types d'outils de coupe (*Type* : A, B, C et D). À partir d'un tableau de données d'une taille n = 40, On cherche à établir une équation de la forme :

$$F = \hat{\beta}_0 + \hat{\beta}_1 V + \hat{\beta}_2 P + \varepsilon \text{ ....(II)}$$

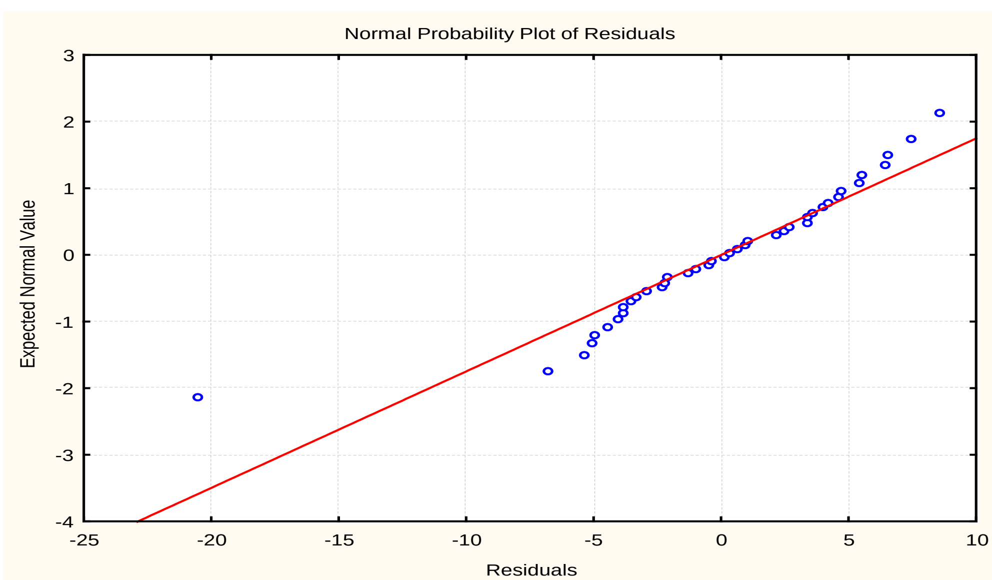
ε représente les résidus du modèle.

On utilise le logiciel ***STATISTICA*** afin d'ajuster le modèle (II) et on obtient le tableau et le graphique ci-dessous (Tableau 4.1 et Graphe 4.1) :

Tableau 4.1

	Analysis of Variance; DV: Fini (final)				
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	6577,530	b	3288,765	e	f
Residual	a	c	d		
Total	7645,900				

Graphe 4.1



a) (2.5 points) Calculer les valeurs des cases (a), (b), (c), (d) et (e)

b) (1 point) Effectuer un test de signification globale au seuil 95%. Conclure !

c) (1 point) Quelle conclusion peut-on tirer du graphe ci-dessus concernant les résidus ?

QUESTION # 4 (suite)

d) (1.5 points) Soit le triplet (Vi , Pi, Fi) relatif a la $i^{\text{ème}}$ observation. Calculer la probabilité qu'il y ait un écart supérieur à 0.5 entre la valeur Fi observée et la valeur $\hat{F}_i$ prédite par le modèle.

e) (1 point) Proposer une solution qui permettrait d'incorporer la variable *Type* dans le modèle (II)

Afin d'évaluer l'influence de la variable Type (type d'outil) sur la variable F (fini), on fait appel au module ANOVA qui nous génère un tableau (voir ci-dessous, Tableau 4.2) des intervalles de confiance pour la différence des moyennes du fini suivant les différents types d'outils.

Tableau 4.2

	LSD test; variable Fini (final) Simultaneous confidence intervals				
Cell No.	versus Cell No.	Mean Differ.	Standard Error	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
1	2	-0,2500	3,733553	-7,8220	7,3220
	3	-25,2778	4,006426	-33,4032	-17,1524
	4	-18,6111	4,006426	-26,7365	-10,4857
2	3	-25,0278	3,845027	-32,8259	-17,2297
	4	-18,3611	3,845027	-26,1592	-10,5630
3	4	6,6667	4,110506	-1,6698	15,0032

Ainsi, on peut lire par exemple : $\mu_A - \mu_B \in [-7{,}822;\, 7{,}322]$

f) (2 points) D'après le tableau 4.2, quelles sont les paires de types d'outils qui sont significativement différentes au seuil 95% ? Peut-on ainsi dire que la variable *Type* a un impact sur la variable F ?

g) (1 point) Donner la forme de la matrice qu'on aurait pu obtenir en utilisant un test de Tukey (utiliser 1 si la différence est significative et 0 sinon).