

STATISTIQUES

Examen – Emmanuel PERINEL

CONSIGNES :

Consignes à lire attentivement

- (1) Le sujet comporte 5 pages, **4 exercices**.
- (2) Pour chaque exercice, vous trouverez les **données à traiter** dans un des 4 onglets du fichier Excel **data.xlsx**
- (3) **Barème indicatif** : 5 points par exercice.
- (4) **Rapport à rendre sous Moodle**.
 - Un fichier au format pdf, Word ou Open office.
 - Nom du fichier : indiquez vos nom et prénom, par exemple **NomPrénom.pdf**
 - La remise des fichiers est autorisée une heure après le début de l'épreuve (c'est-à-dire à partir de 15h00).
 - En cas de difficulté pour déposer le fichier sous Moodle, vous pouvez me l'envoyer par mail à l'adresse **e.perinel@unistra.fr**
 - Pour tenir compte d'éventuels problèmes de connexion, l'heure limite de remise de votre rapport est fixée à **16h30, dernier délai**.
- (5) En cas de difficulté d'ordre technique (connexion internet ou au serveur d'application, problème logiciel), prévenez-moi rapidement, soit par mail soit par SMS au 06 51 53 96 44.
- (6) Les tests d'hypothèses seront effectués **au seuil de significativité de 5%**.
- (7) Pour chaque question nécessitant la mise en œuvre du logiciel XLSTAT, il est indispensable d'incorporer à votre rédaction la **sortie produite par le logiciel** (copié-collé d'un tableau et/ou un graphique) qui vous permet de justifier votre réponse

Important ! Pour éviter de mauvaises surprises (déconnexion intempestive, plantage d'ordinateur, etc.) n'oubliez pas de procéder régulièrement à des **sauvegardes** de votre travail !

SOMMAIRE

1 Exercice 1

Exercice 1.

(D'après Shafawi A., Ebdon L., Foulkes M. and Corns W., 1999)

Dix analyses répétées de la concentration de mercure dans un échantillon de gaz commercial condensé ont conduit aux résultats suivants (en ng/mL) :

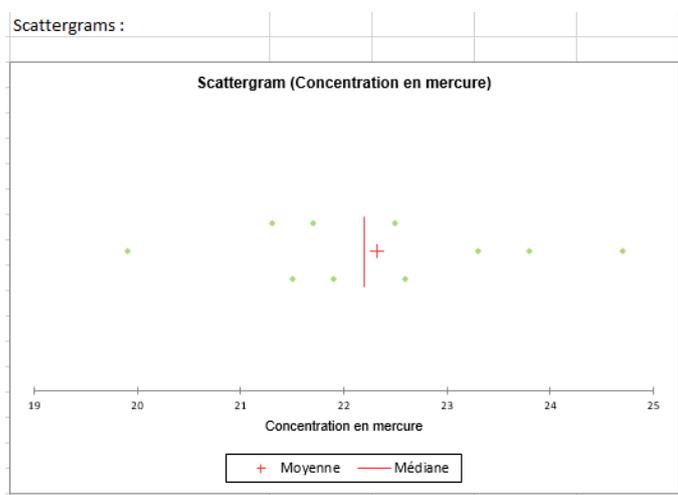
23,3	22,5	21,9	21,5	19,9	21,3	21,7	23,8	22,6	24,7
------	------	------	------	------	------	------	------	------	------

Questions

- 1) Représenter les données de cet échantillon à l'aide d'un nuage de points puis d'une boîte à moustaches. Les graphiques devront être disposés de façon horizontale, comporter un titre et des libellés d'axes explicites.
- 2) Calculer la moyenne, l'étendue des valeurs et l'écart type pour la concentration de mercure dans cet échantillon.
- 3) Cet échantillon comporte-t-il une ou plusieurs observations que l'on pourrait qualifier d'extrêmes ou remarquables ? Justifier.
- 4) Construire un intervalle de confiance de la concentration moyenne au niveau de confiance de 95%. Donner l'amplitude de cet intervalle puis interprétez celui-ci.
- 5) L'intervalle de confiance construit par XLSTAT est-il basé sur l'utilisation de la loi Normale ou la loi de Student ? Les conditions permettant le calcul de cet intervalle sont-elles vérifiées pour cet exemple ? Justifier votre réponse et procéder si besoin à la vérification des conditions d'application.

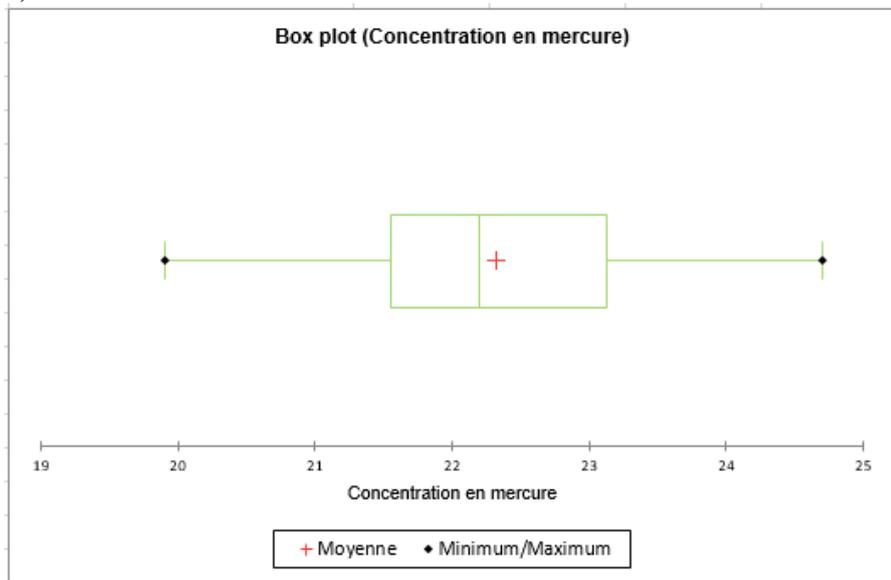
QUESTION 1

Tout d'abord, nous allons représenter les données de cet échantillon à l'aide d'un nuage de point (=scattergram) :



Ce nuage de point montre une médiane et une moyenne presque équivalente, ainsi qu'une assez faible étendue.

Ensuite, nous allons représenter les données de cet échantillon à l'aide d'une boîte à moustache (box-plot) :



La boîte à moustache nous permet de nous représenter les écarts interquartiles de manière plus visuelle. Ainsi, l'écart entre la médiane et Q3 semble plus élevée qu'entre Q1-médiane. Il semble également ne pas avoir des valeurs extrêmes.

QUESTION 2

Ensuite, j'ai utilisé la fonction des statistiques descriptives pour calculer la moyenne, l'étendue (max-min) et l'écart-type :

Statistiques descriptives (Données quantitatives) :	
Statistique	Concentration en mercure
Nb. d'observations	10
Minimum	19,900
Maximum	24,700
1er Quartile	21,550
Médiane	22,200
3ème Quartile	23,125
Moyenne	22,320
Variance (n-1)	1,895
Ecart-type (n-1)	1,377

Nous observons ainsi une étendue de **4.8**, une moyenne de **22.320** ainsi qu'un écart-type à la moyenne de **1.377**

QUESTION 3

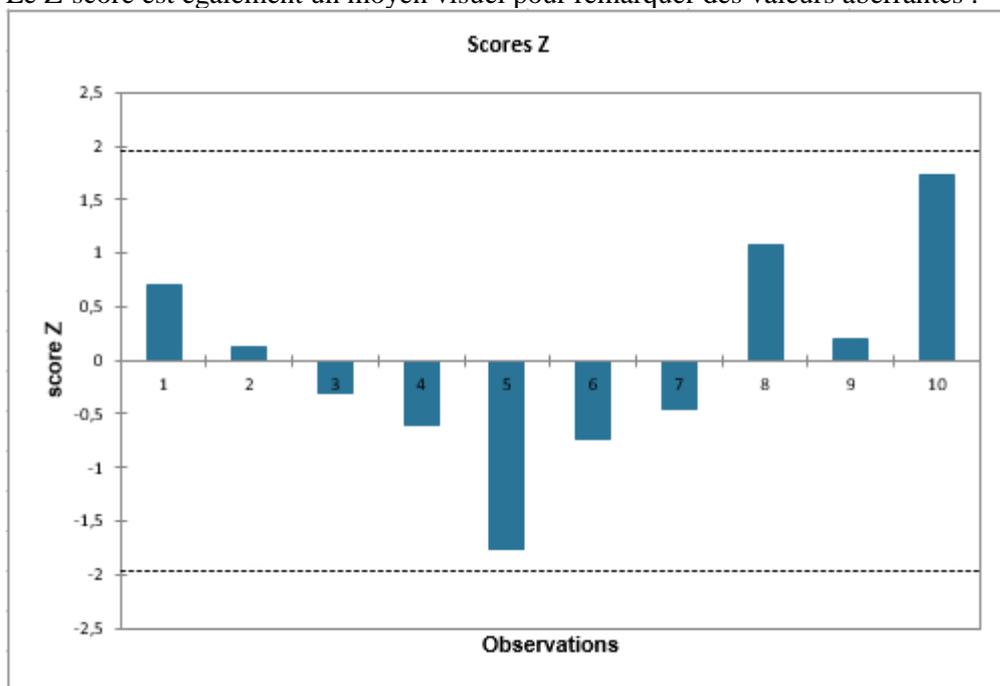
Pour connaître le nombre de valeurs extrêmes, j'ai utilisé un test de Grubbs. En règle générale, on se réfère à une table mais XCEL Stat sait produire la méthode seul.

Test de Grubbs pour les valeurs extrêmes / Test bilatéral :	
G (Valeur observée)	1,758
G (Valeur critique)	2,290
p-value (bilatérale)	0,571
alpha	0,05

Interprétation du test :
 H0 : Il n'y a pas de valeur extrême dans les données
 Ha : Le minimum ou le maximum est une valeur extrême
 Etant donné que la p-value calculée est supérieure au niveau de signification seuil $\alpha=0,05$, on ne peut pas rejeter l'hypothèse nulle H0.

D'après les résultats obtenus pour ce test, nous observons que la p-value est égale à 57%, ce qui signifie que l'on ne peut pas rejeter H0, ainsi il n'y a pas de valeurs aberrantes.

Le Z-score est également un moyen visuel pour remarquer des valeurs aberrantes :

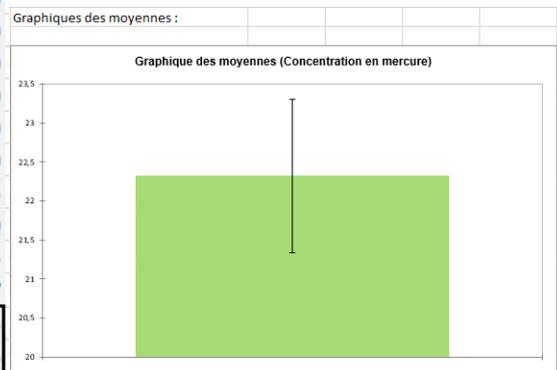


QUESTION 4

J'ai utilisé la fonction « Borne inf de la moyenne (95%) » et « Borne sup de la moyenne (95%) » afin de pouvoir calculer l'intervalle de confiance de la concentration moyenne :

Statistiques descriptives (Données quantitatives) :

Statistique	Concentration en mercure
Nb. d'observations	10
Minimum	19,900
Maximum	24,700
Amplitude	4,800
1er Quartile	21,550
Médiane	22,200
3ème Quartile	23,125
Moyenne	22,320
Variance (n-1)	1,895
Ecart-type (n-1)	1,377
Borne inf. de la moyenne (95%)	21,335
Borne sup. de la moyenne (95%)	23,305



Ces valeurs indiquent que nous pouvons être sûrs à 95% que la moyenne se situe entre 21.335 et 23.305. L'intervalle étant assez restreintes, nous pouvons considéré cette information utile.

QUESTION 5

La loi utilisé par XCEL Stat est la loi de Student, car la variance σ^2 est inconnue. Ainsi :

- la variance σ est remplacé par celle de l'échantillon s,
- la loi normale normalement utilisé, est remplacé par la loi de Student
- Attention : la variable doit cependant toujours être distribuée selon une loi Normale.

Pour vérifier la condition de la loi Normale, nous allons procéder à un Q-Q Plot et à un Shapiro-Wilk. Pour ce faire, nous considérons H_0 : « la variable dont provient l'échantillon suit une Loi Normale », et l'hypothèse H_a étant son alternative.

Test de Shapiro-Wilk (Concentration en mercure) :

W	0,984
p-value (bila	0,983
alpha	0,050

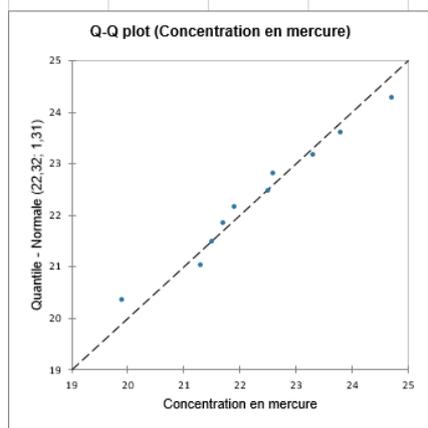
Interprétation du test :

H0 : La variable dont provient l'échantillon suit une loi Normale.

Ha : La variable dont provient l'échantillon ne suit pas une loi Normale.

Etant donné que la p-value calculée est supérieure au niveau de signification seuil $\alpha=0,05$, on ne peut pas rejeter l'hypothèse nulle H0.

Graphiques Q-Q (loi normale) :



La p-value vaux 98%, cela signifie que H0 est acceptable et donc que la variable suit une loi Normale. Le graphique Q-Q plot montre également des valeurs proches de la tendance. Cela conforte notre test de Shapiro-Wilk induisant une variable suivant une Loi Normale.

2 Exercice 2

Exercice 2.

Dans l'industrie nucléaire, des registres indiquent la quantité de plutonium reçue, transportée et utilisée. Chaque envoi de granulés de plutonium reçu est analysé soigneusement pour vérifier la pureté (en %) déclarée par le fournisseur. Un envoi spécifique analysé fournit les quatre mesures suivantes :

99,93% 99,87% 99,91% 99,86%

La pureté déclarée par le fournisseur est de 99,95%. Notre objectif est de savoir si cet envoi peut être considéré comme acceptable. Nous admettrons comme raisonnable l'hypothèse selon laquelle le pourcentage de plutonium est distribué selon une loi normale.

Questions

- 1) Effectuer sous XLSTAT le test d'hypothèses permettant de répondre à la question posée. Dans votre rédaction :
 - Rappeler les conditions d'application du test choisi.
 - Préciser les deux hypothèses testées.
 - Énoncer clairement la décision prise à l'issue de ce test en la justifiant.
- 2) En quoi un *Graphique des moyennes* (Menu *Statistique descriptives*) associé à des barres d'erreurs judicieusement choisies permet également de répondre à la question posée ? Produire le graphe en question à l'aide de XLSTAT et justifier précisément votre réponse.

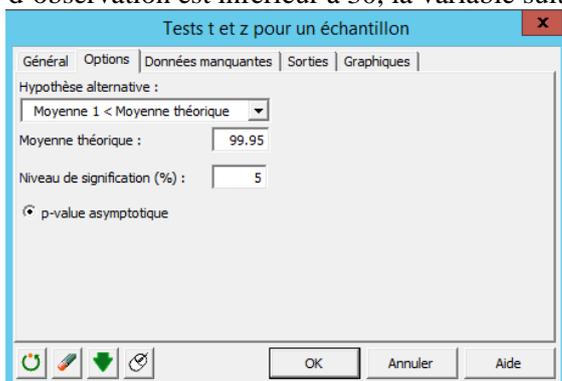
QUESTION 1

Tout d'abord, nous savons que les 4 valeurs présentées proviennent de la répétabilité d'une mesure de pureté. Ensuite, nous n'avons pas de variance dans l'énoncé, par contre nous savons que le pourcentage de plutonium est distribué selon une Loi Normale. Enfin, étant donné que nous voulons savoir si cet envoi est considéré acceptable vis-à-vis de la pureté de 99.95% déclarée par le fournisseur, l'utilisation d'un test unilatéral à gauche semble adéquate.

Les hypothèses testées :

- H_0 : « La moyenne de la pureté est de 99.95% »
- H_1 : « La moyenne de la pureté est inférieure à 99.95% »

Le choix du test à réaliser est un test de Student, car nous n'avons pas de variance et bien que le nombre d'observation est inférieur à 30, la variable suit une loi Normale. Donc un test t est adéquat.



Test t pour un échantillon / Test unilatéral à gauche :

Intervalle de confiance à 95% autour de la moyenne :
] -Inf ; 99,931]

Différence	-0,058
t (Valeur obs)	-3,481
t (Valeur crit)	-2,353
DDL	3
p-value (unil)	0,020
alpha	0,050

Interprétation du test :

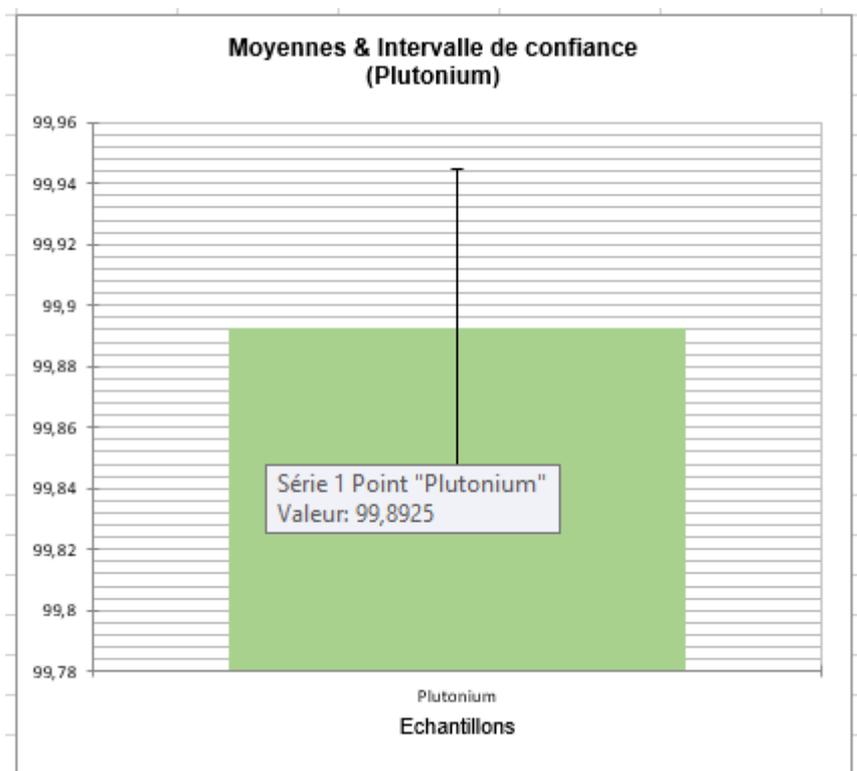
H0 : La moyenne est égale à 99,95.

Ha : La moyenne est inférieure à 99,95.

Etant donné que la p-value calculée est inférieure au niveau de signification $\alpha=0,05$, on doit rejeter l'hypothèse nulle H0, et retenir l'hypothèse alternative Ha.

D'après les résultats obtenus, nous pouvons conclure avec l'intervalle de confiance : 99.95% appartenant pas à l'intervalle de confiance à 95% autour de la moyenne, l'hypothèse H0 est rejetée et l'hypothèse alternative est acceptée. La p-value est de 2%, ce qui est inférieur à notre alpha de 5%. Cela indique que nous avons 2% de probabilité d'avoir une pureté de 99.95%. Nous prenons donc peu de risque à rejeter H0.

QUESTION 2



Comme expliqué précédemment, la p-value et l'intervalle de confiance sont deux moyens pour prendre une décision et conclure sur le test de Student utilisé comme test de conformité d'une moyenne. Ainsi, nous observons que 99.95% est supérieur à la limite de l'intervalle de confiance de 99.931%, donc on rejette H0. Le fournisseur ne nous a pas envoyé un envoi acceptable.

3 Exercice 3

Exercice 3.

Les résultats de la concentration de la norépinéphrine en $\mu\text{mol/g}$ de créatinine, dans l'urine de jeunes patients est représentée dans le tableau suivant :

Homme	0,48	0,36	0,20	0,55	0,45	0,46	0,47	0,23
Femme	0,35	0,37	0,27	0,29				

Nous supposons que les quantités de norépinéphrine mesurées suivent une loi normale pour chacune des deux populations.

L'objectif de cette analyse est de déterminer s'il existe une différence significative entre les concentrations de norépinéphrine moyennes des patients des deux sexes.

Questions

- 1) Représenter les données recueillies à l'aide d'un graphique approprié.
- 2) Calculer la moyenne et l'écart type pour les deux échantillons comparés.
- 3) Répondre au problème posé à l'aide d'un test d'hypothèses. Dans votre rédaction :
 - Rappeler les conditions d'application du test choisi.
 - Préciser les deux hypothèses testées pour chacun des tests réalisés par le logiciel.
 - Énoncer clairement la décision prise à l'issue de chacun des tests en la justifiant.
- 4) Quel(s) autre(s) test(s) auriez-vous pu employer pour réponse au problème posé ?

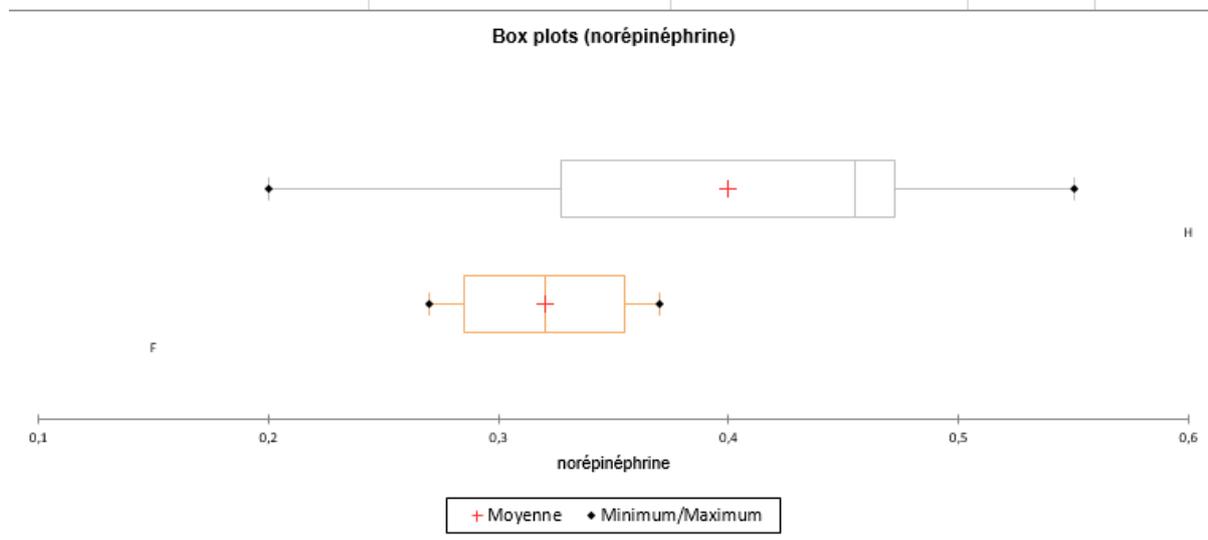
QUESTION 1

Tout d'abord, comme à l'accoutumé, quelques statistiques descriptives :

Statistiques descriptives (Données quantitatives) :			
Statistique	norépinéphrine Sexe-F		norépinéphrine Sexe-H
Nb. d'observations	4		8
Minimum	0,270		0,200
Maximum	0,370		0,550
Amplitude	0,100		0,350
1er Quartile	0,285		0,328
Médiane	0,320		0,455
3ème Quartile	0,355		0,473
Moyenne	0,320		0,400
Variance (n-1)	0,002		0,016
Ecart-type (n-1)	0,048		0,126
Borne inf. de la moyenne (95%)	0,244		0,295
Borne sup. de la moyenne (95%)	0,396		0,505

Ensuite, un box-plot permettrait de visualiser les données et de faciliter la comparaison entre les résultats de la concentration en norépinéphrine moyenne des patients des deux sexes :

Box plots :



Le box-plot supérieur, correspond aux résultats obtenus pour les Hommes, tandis que celui du bas correspond à ceux des femmes. Cette représentation graphique montre une symétrie des résultats plus probant chez les femmes que chez les hommes. De plus, il y a une moyenne plus éloignée de la médiane chez les hommes que chez les femmes. Nous observons également une étendue, une moyenne et une médiane plus grande pour les hommes. Cela semble tout à fait envisageable, puisque nous avons uniquement 4 valeurs pour les femmes, ce qui est 2 fois moins d'observations que chez les hommes.

QUESTION 2

D'après les statistiques descriptives élaborées précédemment, Excel Stat calcule une moyenne de $0.320 \mu\text{mol/g}$ de norépinéphrine chez les femmes et $0.400 \mu\text{mol/g}$ de norépinéphrine chez les hommes.

Également, nous pouvons observer un écart-type ($n-1$) de 0.048 chez les femmes et 0.126 chez les hommes.

QUESTION 3

Tout d'abord, nous savons que les observations présentées proviennent d'individu de sexes féminin et masculin, ce qui correspond à un échantillon indépendant, car aucun individu appartient au groupe F et H en même temps. Ensuite, nous n'avons pas de variance dans l'énoncé, par contre nous savons que les quantités de norépinéphrine mesurées suivent une Loi Normale. Enfin, étant donné que nous voulons savoir s'il existe une différence significative entre les concentrations de norépinéphrine moyennes des patients des deux sexes, l'utilisation d'un test bilatéral semble adéquate.

Les hypothèses testées sont :

- H_0 : « Les moyennes de concentration en norépinephrine des hommes et des femmes sont équivalentes »
- H_1 : « Les moyennes de concentration en norépinephrine des hommes et des femmes ne sont pas équivalentes »

Le choix du test à réaliser est un test de Student, car nous n'avons pas de variance et bien que le nombre d'observation soit inférieur à 30, la variable suit une loi Normale. Donc un test t, utilisé en tant que test de comparaison de moyenne pour deux échantillons semble adéquat.

Afin de traiter les données, je commence par vérifier les résultats du test de Fischer, qui va me permettre de comparer les 2 variances. En effet, la variance étant inconnue, nous devons nous baser sur les estimations corrigées s_1^2 et s_2^2 , et vérifier leurs égalités, sinon nous devons procéder à un test non paramétrique.

Test F de Fisher / Test bilatéral :	
Intervalle de confiance à 95% autour du rapport des variances : [0,024; 2,102]	
Rapport	0,144
F (Valeur obs)	0,144
F (Valeur c)	5,890
DDL1	3
DDL2	7
p-value (bila)	0,139
alpha	0,050
Interprétation du test :	
H0 : Le rapport entre les variances est égal à 1.	
Ha : Le rapport entre les variances est différent de 1.	
Etant donné que la p-value calculée est supérieure au niveau de signification seuil alpha=0,05, on ne peut pas rejeter l'hypothèse nulle H0.	

D'après les résultats de Fischer, le rapport des variances à 13% de chance d'être égale à 1, ce qui est supérieur à 5%. Donc, l'hypothèse H_{0F} : « le rapport des variances est égale à 1 » est acceptable et ne peut être rejetée.

Ensuite, nous pouvons passer aux résultats du test de comparaison pour deux échantillons indépendants par un test de Student t :

Test t pour deux échantillons indépendants / Test bilatéral :	
Intervalle de confiance à 95% autour de la différence des moyennes : [-0,228; 0,068]	
Différence	-0,080
t (Valeur obs)	-1,207
t (Valeur c)	2,228
DDL	10
p-value (bila)	0,255
alpha	0,050
Interprétation du test :	
H0 : La différence entre les moyennes est égale à 0.	
Ha : La différence entre les moyennes est différente de 0.	
Etant donné que la p-value calculée est supérieure au niveau de signification seuil alpha=0,05, on ne peut pas rejeter l'hypothèse nulle H0.	

Nous allons prendre une décision selon la p-value obtenu à ce test : l'hypothèse H_0 est acceptable, car nous avons une p-value de 25%, ce qui est supérieur à notre alpha de 5%. Cela signifie que la moyenne de concentration en norépinephrine ne semble pas significativement différente entre les Hommes et les Femmes.

QUESTION 4

Pour répondre à ce problème, j'aurais également pu utiliser un test ANOVA à un facteur et deux modalités (F et H), car nous pouvons l'utiliser pour comparer un nombre de moyenne supérieur ou égale à deux. Ce qui était le cas. Les variables auraient été considérées comme fixe, car on ne compare pas des hommes aléatoirement, mais des jeunes patients (donc pas de vieux).

4 Exercice 4

Exercice 4.

Hyptis suaveolens est une plante d'environ 2 mètres de haut à fleurs de couleur pourpre ou blanche, native du continent américain. De ses feuilles et ses fleurs est extraite une huile essentielle utilisée dans la parfumerie, le soin ou encore dans la lutte contre les insectes.

Une étude a été menée afin de comparer la composition chimique de différentes huiles élaborées à partir de plantes provenant de quatre lieux géographiques du Salvador. Les données recueillies sont disponibles dans le fichier `data.xlsx`. Parmi les composés mesurés, nous allons nous intéresser à deux d'entre eux : β -pinène et sabinène.

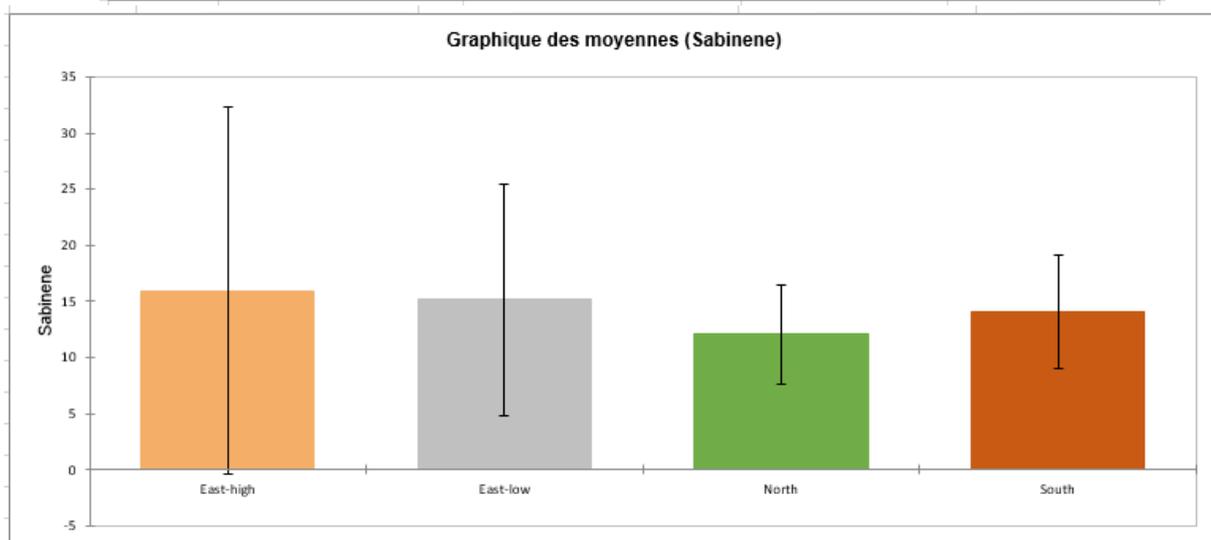
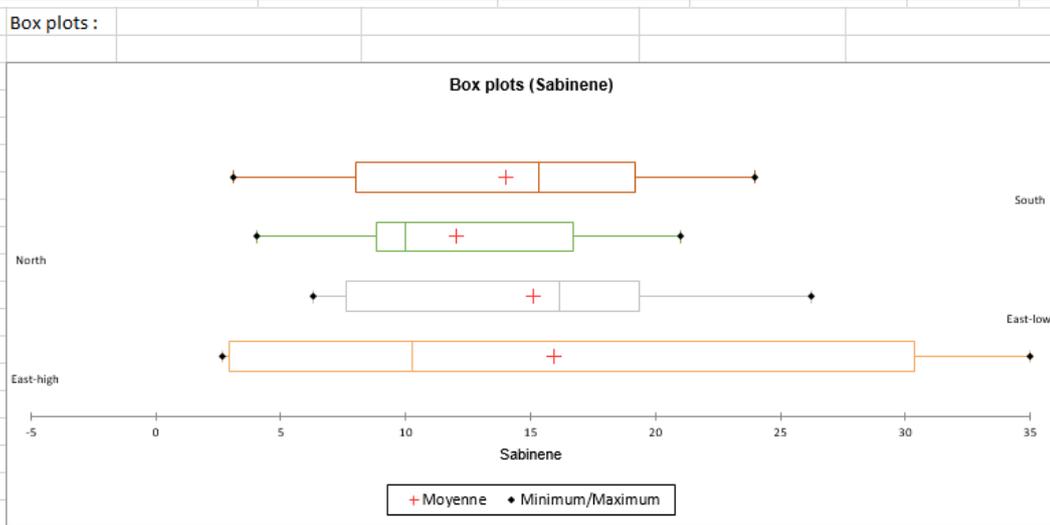
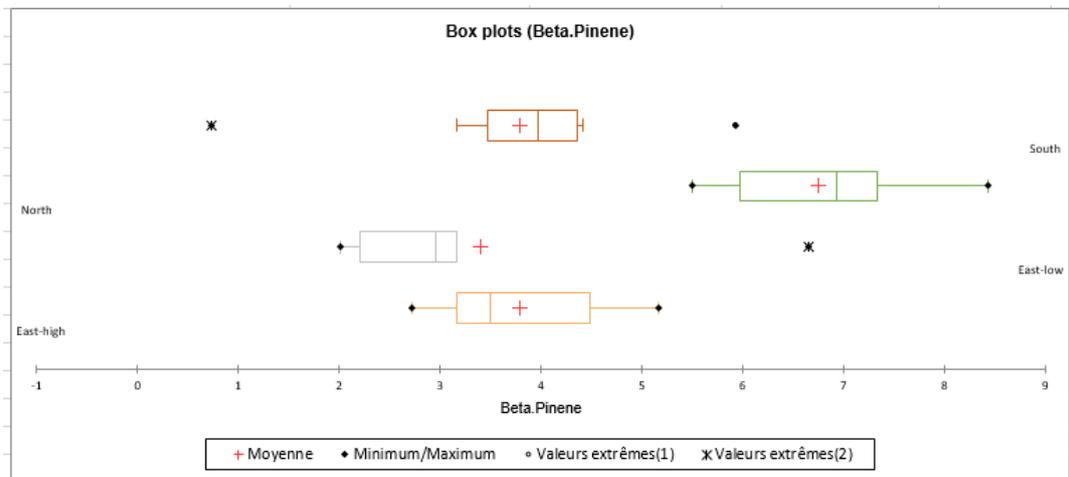
Objectif. Il s'agit de déterminer, pour chacun de ces deux composés, si la teneur moyenne diffère selon le lieu géographique. Et dans le cas où des différences apparaîtraient, de préciser entre quels lieux géographiques se situent ces différences.

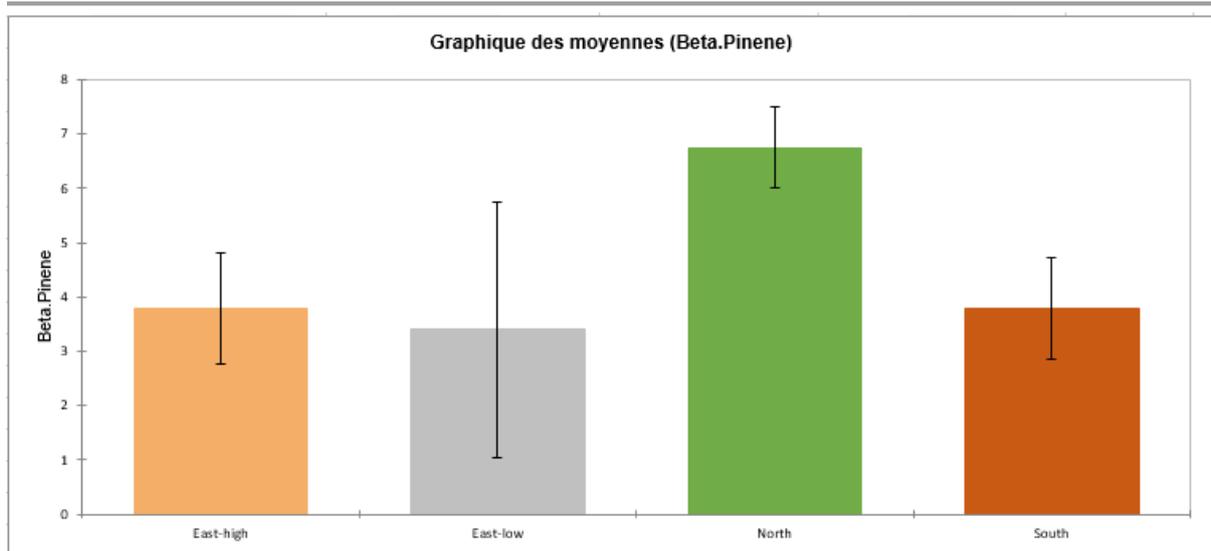
Dans votre rédaction, vous ferez apparaître comme d'habitude les points suivants :

- Statistiques descriptives et représentation graphique des données étudiées.
- Conditions d'applications du ou des tests effectués.
- Les hypothèses testées.
- La décision prise à l'issue de chacun des tests en la justifiant et en exprimant concrètement votre conclusion en regard du problème posé.

Tout d'abord, quelques statistiques descriptives :

Statistiques descriptives (Données quantitatives) :									
Statistique	Sabinene Lieu géographique-East-high	Sabinene Lieu géographique-East-low	Sabinene Lieu géographique-North	Sabinene Lieu géographique-South	Beta.Pinene Lieu géographique-East-high	Beta.Pinene Lieu géographique-East-low	Beta.Pinene Lieu géographique-North	Beta.Pinene Lieu géographique-South	
Nb. d'observ	6	5	9	10	6	5	9	10	
Minimum	2,670	6,300	4,040	3,090	2,720	2,010	5,500	0,740	
Maximum	35,000	26,220	20,980	23,980	5,160	6,660	8,430	5,930	
Amplitude	32,330	19,920	16,940	20,890	2,440	4,650	2,930	5,190	
1er Quartile	2,938	7,610	8,830	8,005	3,163	2,200	5,970	3,475	
Médiane	10,280	16,130	9,960	15,335	3,495	2,950	6,930	3,965	
3ème Quartile	30,358	19,370	16,730	19,185	4,488	3,160	7,330	4,358	
Moyenne	15,962	15,126	12,046	14,032	3,790	3,396	6,753	3,784	
Variance (n-1)	244,836	69,121	32,772	50,046	0,929	3,565	0,924	1,727	
Ecart-type (n-1)	15,647	8,314	5,725	7,074	0,964	1,888	0,961	1,314	
Borne inf. de confiance	-0,459	4,803	7,645	8,971	2,779	1,052	6,014	2,844	
Borne sup. de confiance	32,382	25,449	16,446	19,093	4,801	5,740	7,492	4,724	





Etant donnée que nous voulons savoir si la composition chimique des huiles change en fonction du lieu géographique. Nous allons réaliser une ANOVA à un facteur (lieu) à quatre modalités.

Les conditions :

- Les observations sont effectivement indépendantes
- La normalité des résidus sera observé durant l'ANOVA
- L'homogénéité des variances sera observé durant l'ANOVA

Les hypothèses testées :

- H_0 : « les moyennes de sabinene est équivalente selon les 4 lieux »
- H_a : « au moins deux moyennes de sabinene different selon les 4 lieux »

XLSTAT 2021.2.2.1143 - ANOVA - Début : 24/06/2021 à 16:26:39 / Fin : 24/06/2021 à 16:26:40
 Y / Variables dépendantes : Classeur = Classeur1 / Feuille = Feuil41 / Plage = Feuil41!\$A:\$A / 30 lignes et 1 cc
 X / Qualitatives : Classeur = Classeur1 / Feuille = Feuil41 / Plage = Feuil41!\$H:\$H / 30 lignes et 1 colonne
 Contraintes : Somme(ai)=0
 Intervalle de confiance (%) : 95
 Tolérance : 0,0001
 Utiliser les moyennes estimées : Oui



Statistiques descriptives (Données quantitatives) ▼

Statistiques descriptives (Données quantitatives) :

Variable	Observations	Obs. avec données	Obs. sans données	Minimum	Maximum	Moyenne	Ecart-type
Sabinene	30	0	30	2,670	35,000	14,004	8,861

Statistiques descriptives (Données qualitatives) :

Variable	Modalités	Comptages	Effectifs	%
Lieu géographique	East-high	6	6	20,000
	East-low	5	5	16,667
	North	9	9	30,000
	South	10	10	33,333

Matrice de corrélation :					
	Lieu géographique	Lieu géographique	Lieu géographique	Lieu géographique	Sabinene
Lieu géographique	1	0,631	0,576	-0,853	0,061
Lieu géographique	0,631	1	0,600	-0,857	0,030
Lieu géographique	0,576	0,600	1	-0,860	-0,086
Lieu géographique	-0,853	-0,857	-0,860	1	0,002
Sabinene	0,061	0,030	-0,086	0,002	1
Régression de la variable Sabinene :					
Coefficients d'ajustement (Sabinene) :					
Observation	30				
Somme des	30				
DDL	26				
R ²	0,028				
R ² ajusté	-0,084				
MCE	85,125				
RMCE	9,226				
MAPE	101,869				
DW	1,412				
Cp	4,000				
AIC	137,031				
SBC	142,635				
PC	1,271				

Analyse Type I Sum of Squares (Sabinene) :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Lieu géograp	3,000	63,817	21,272	0,250	0,861

Analyse Type II Sum of Squares (Sabinene) :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Lieu géograp	3,000	63,817	21,272	0,250	0,861

Analyse Type III Sum of Squares (Sabinene) :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Lieu géograp	3,000	63,817	21,272	0,250	0,861

Paramètres du modèle (Sabinene) :

Source	Valeur	Erreur standard	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
Constante	14,291	1,753	8,151	<0,0001	10,687	17,895
Lieu géograp	1,670	3,189	0,524	0,605	-4,884	8,225
Lieu géograp	0,835	3,404	0,245	0,808	-6,162	7,831
Lieu géograp	-2,246	2,793	-0,804	0,429	-7,988	3,496
Lieu géograp	-0,259	2,707	-0,096	0,924	-5,825	5,306

Tester les hypothèses :

Test de l'hypothèse de normalité des résidus (Sabinene) :

W	0,961
p-value (bila)	0,322
alpha	0,050

Interprétation du test :

H0 : Les résidus suivent une loi Normale.

Ha : Les résidus ne suivent pas une loi Normale.

Etant donné que la p-value calculée est supérieure au niveau de signification seuil $\alpha=0,05$, on ne peut pas rejeter l'hypothèse nulle H0.

Test d'homoscédasticité des résidus (Sabinene) - Lieu géographique :

F	5,213
DDL1	3
DDL2	26
p-value (bila)	0,006
alpha	0,050

Interprétation du test :

H0 : Les résidus sont homoscédastiques

Ha : Les résidus sont hétéroscédastiques

Etant donné que la p-value calculée est inférieure au niveau de signification $\alpha=0,05$, on doit rejeter l'hypothèse nulle H0, et retenir l'hypothèse alternative Ha.

Interprétation (Sabinene) :

Etant donné le R^2 , 3% de la variabilité de la variable dépendante Sabinene est expliquée par la variable explicative.

Etant donnée la p-value associée à la statistique F calculée dans le tableau d'analyse de la variance, et étant donné le niveau de signification de 5% choisi, les variables explicatives n'apporte pas d'information au modèle comparée à ce qu'apporterait la seule moyenne de la variable dépendante. Le fait que les variables n'apportent pas de l'information significative au modèle peut être interprété de différentes façons : soit les variables ne contribuent pas à l'explication du modèle, soit des covariables qui combinées aux variables déjà présentes sont manquantes, soit la forme du modèle n'est pas la bonne, soit les données sont entachées d'erreurs.