

STATISTIQUE
EXAMEN N° 1 (sujet A)
Mme B. BOTTOLIER

<i>Conditions d'examens</i>	Documents		Autorisés
		X	Non autorisés
	Calculatrice		Non autorisée
		X	4 opérations autorisée
			Tout type autorisée

Remarques particulières

Tables statistiques autorisées.

A la fin de l'épreuve vous devez rendre:

- le questionnaire de 5 pages et vos brouillons, sous peine de perdre 2 points!!!
- les feuilles de réponses (pages 6 et 7) avec votre nom. Sur ces feuilles figure le barème de chaque question entre parenthèses (total de 40 points).

Choisissez la (les) bonne(s) réponse(s) en reportant la(les) bonne(s) lettre(s) sur la feuille des réponses ci jointe.

1 - Quand on se sert des paramètres calculés d'un échantillon, l'application de la théorie de l'échantillonnage est utilisée quand on veut:

- a. Réaliser des contrôles portant sur une caractéristique de qualité à la fin d'une production.
- b. Proposer un intervalle dans lequel devrait se trouver un certain pourcentage des moyennes des échantillons de taille n
- c. Définir si un critère de qualité suit une loi Normale
- d. Proposer un intervalle dans lequel la moyenne du critère de qualité étudié dans une population a une probabilité donnée de ne pas se trouver.

2 - Le symbole $\sigma(m)$ représente:

- a. L'écart type estimé de la variable mesurée dans la population.
- b. L'écart type de la moyenne des échantillons de taille n.
- c. L'écart type de la variable mesurée dans la population de taille N.
- d. L'écart type de la variable mesurée dans l'échantillon de taille n.

3 - Le symbole $\hat{\sigma}_x$ représente:

- a. L'écart type estimé de la variable mesurée dans la population.
- b. L'écart type de la moyenne des échantillons de taille n.
- c. L'écart type de la variable mesurée dans la population de taille N.
- d. L'écart type de la variable mesurée dans l'échantillon de taille n.

21 - Le fabricant de pâtes Panzato compare 2 chaînes de mises en paquets des pâtes. Des études antérieures avaient donné une dispersion de 20g et 30g pour les chaînes 1 et 2 respectivement. Afin de décider si les 2 chaînes étaient équivalentes, on prélève sur les 500 paquets de la première 40 paquets dont le poids moyen est égal à 400g, et, sur les 400 paquets de la deuxième on en prélève 30 et trouve un poids moyen de 410g. Sans faire les calculs donnez l'application numérique du calcul de l'écart type de la différence des moyennes nécessaire pour répondre à la question sous-entendue dans cette étude.

22 - Deux échantillons aléatoires sont tirés de façon indépendante dans deux populations normales d'où sont prélevées $n_1 = 18$ et $n_2 = 21$ unités indépendantes avec remise. On obtient respectivement pour les échantillons 1 et 2: $m_1=2$ et $m_2=1$; $SCE_1 = 3849$ et $SCE_2 = 3948$. Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet la comparaison des 2 moyennes.

23 - Deux échantillons aléatoires sont tirés de façon indépendante dans deux populations normales d'où sont prélevées $n_1 = n_2 = 20$ unités indépendantes avec remise. On obtient respectivement pour les échantillons 1 et 2: somme des valeurs = 60 et 40; $\sum x^2_{(1)} = \sum x^2_{(2)} = 280$. Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet la comparaison des 2 variances.

24 - On veut construire un intervalle d'estimation, pour un risque alpha de 1%, d'une moyenne μ de population de taille $N=200$, à partir d'un échantillon de taille 25 de variance égale à 4 et de moyenne 120. Sans faire les calculs donnez l'application numérique du calcul de la limite inférieure.

25 - On veut construire un intervalle d'estimation, pour un risque alpha de 1%, d'une moyenne μ de population de taille $N=800$, à partir d'un échantillon de taille 40 de variance égale à 4 et de moyenne 120. Sans faire les calculs donnez l'application numérique du calcul de la marge d'erreur absolue de cet intervalle.

26 - On veut construire un intervalle d'estimation, pour un risque alpha de 1%, d'une moyenne μ de population de taille $N=800$, dans laquelle la variable X étudiée suit une loi Normale, à partir d'un échantillon de taille 20 de variance égale à 4 et de moyenne 120. Quelle est la valeur du critère statistique théorique qu'il faut utiliser?

27 - On veut construire l'intervalle dans lequel doit de trouver 95% des fréquences des échantillons de taille 100 issus d'une population de taille 2500, centrée sur une proportion égale à 2% de pièces défectueuses. Sans faire les calculs donnez l'application numérique du calcul de la marge d'erreur relative de cet intervalle.

28 - Soit une variable aléatoire X dans une population de 1000 unités, de moyenne égale à 20 et d'écart type égal à 2. Sans faire les calculs donnez l'application numérique du calcul de la limite supérieure de l'intervalle dans lequel on a 98% de chance de trouver la moyenne d'un échantillon de taille 121.

29 - On veut construire un intervalle d'estimation d'une proportion p de population, de taille $N = 1800$, à partir d'un échantillon de taille 150, sur lequel 20 individus possèdent la caractéristique étudiée, pour un risque d'erreur égal à 4%. Sans faire les calculs donnez l'application numérique du calcul de la marge d'erreur absolue de l'estimation.

30 - On veut construire un intervalle d'estimation d'une moyenne μ de population, de taille $N=800$, dans laquelle la variable X étudiée suit une loi Normale, à partir d'un échantillon de taille 90 dont la moyenne = 120 et l'écart type = 10, pour un risque alpha de 3%. Sans faire les calculs donnez l'application numérique du calcul de la limite inférieure de cet intervalle.

- 4 - L'estimation ponctuelle d'une variance de population centrée sur μ se fait en utilisant:
- La loi du X^2
 - la $SCE_x(\mu) / (N - 1)$
 - $SCE_x(m) / n$
 - $SCE_x(m) / (n - 1)$
 - $n * SCE_x(m) / (n - 1)$
- où m est la moyenne d'un échantillon de taille n .

- 5 - La distribution des moyennes des échantillons de taille n suit une loi de Student si:
- $n = 20$ et si la loi de la variable aléatoire étudiée (X) dans la population suit une loi uniforme
 - $n = 20$ et si la loi de la variable aléatoire étudiée (X) dans la population suit une loi normale de variance connue
 - $n = 20$ et si la loi de la variable aléatoire étudiée (X) dans la population suit une loi normale de variance inconnue
 - $n = 40$ et si la loi de la variable aléatoire étudiée (X) dans la population suit une loi normale de variance inconnue
 - $n = 40$ et si la loi de la variable aléatoire étudiée (X) dans la population suit une loi normale de variance connue

- 6 - Un tirage est exhaustif si:
- La taille de l'échantillon est faible devant celle de la population
 - Le taux de sondage est inférieur à 10%
 - Le taux de sondage est supérieur à 10%
 - La taille de l'échantillon n'est pas négligeable devant celle de la population

- 7 - Pour que la marge d'erreur associée à une estimation de moyenne soit plus grande il faut que (pour un même écart type donné):
- La racine carrée de la taille de l'échantillon augmente
 - Le niveau de confiance diminue
 - La taille de l'échantillon diminue
 - Le tirage se fasse sans remise
 - La valeur de α diminue
 - Le risque de 1^{ère} espèce augmente

Portez vos réponses aux questions suivantes sur les pages 7 et 8.

8 - On veut construire un intervalle d'estimation d'une proportion p de population, de taille $N = 7000$, à partir d'un échantillon de taille 300, sur lequel 40 individus possèdent la caractéristique étudiée, pour un risque d'erreur égal à 0.05.
Sans faire les calculs donnez l'application numérique du calcul de la marge d'erreur relative de l'estimation.

9 - On veut construire un intervalle d'estimation d'une moyenne μ de population, de taille $N = 80$, dans laquelle la variable X étudiée suit une loi Normale, à partir d'un échantillon de taille 18, de moyenne = 2.1 et de $SCE = 0.05$ pour un risque d'erreur égal à 0.05.
Sans faire les calculs donnez l'application numérique du calcul de la limite inférieure de l'intervalle.

10 - Soit une variable aléatoire X dans une population de $N = 1000$, telle que $L(X) = N(1200; 100)$.
Sans faire les calculs donnez l'application numérique du calcul de la limite inférieure de l'intervalle dans lequel se trouvent 99% des moyennes des échantillons de taille 32.

11 - On veut construire l'intervalle dans lequel doit de trouver 99% des fréquences des échantillons de taille 900 issus d'une population de taille 8000, centrée sur une proportion égale à 1% de pièces défectueuses.

Sans faire les calculs donnez l'application numérique du calcul de la limite inférieure de cet intervalle.

12 - Le fabricant de pâtes Kilus cru compare 2 chaînes de mises en paquets des pâtes, le poids des paquets étant distribué normalement. Afin de décider si les 2 chaînes étaient équivalentes, on prélève sur les 100 paquets de la première 18 paquets dont le poids moyen est égal à 90g avec une $SCE = 900$, et, sur les 100 paquets de la deuxième on en prélève 25 et trouve un poids moyen de 93.5g et une $SCE = 920$.

Donnez (au 1/100ème), pour un risque égal à 5%, le critère statistique théorique qui permet de répondre à la question sous-entendue dans cette étude

13 - Le fabricant de pâtes Kilus cru compare 2 chaînes de mises en paquets des pâtes. Afin de décider si une des chaînes sous dose les paquets, on prélève sur les 100 paquets de la première 36 paquets dont le poids moyen est égal à 90g avec une $SCE = 900$, et, sur les 100 paquets de la deuxième on en prélève 40 et trouve un poids moyen de 92.3g et une $SCE = 950$.

Sans faire les calculs donnez l'application numérique du calcul de l'écart type de la différence des moyennes.

14 - On effectue un sondage sur un échantillon de 2600 individus issus d'une population de 13000 individus, 350 possèdent une caractéristique donnée alors que dans un autre sondage sur un échantillon de 3200 individus issus de la même population, 460 possèdent cette caractéristique. On se demande si le 2ème sondage donne un résultat supérieur.

Donnez (au 1/100ème) le critère statistique théorique au risque de 5%, qui permet de répondre à la question posée.

15 - Une entreprise enregistre habituellement 4% de ses produits qui présentent un défaut mineur. Craignant une augmentation de ce taux elle prélève 900 unités sur les 8000 fabriquées et observe un taux de défaut égal à 5%.

Donnez (au 1/100ème), pour un risque égal à 0.03, le critère statistique théorique qui permet de répondre à la question sous-entendue dans cette étude

16 - Un fabricant de céréales affiche sur ses paquets une teneur en vitamine B1 = 1.125mg avec un écart type égal à 0.05 mg. Sur 400 paquets fabriqués un contrôleur prélève 18 paquets et trouve une teneur moyenne égale à 1.105 mg.

Donnez (au 1/100ème), pour un risque égal à 0.001, le critère statistique théorique qui permet de répondre à la question " le poids affiché sur chaque paquet est-il vrai?"

17 - Une entreprise emballe des tranches de saucisson (par 5) pour proposer des paquets de 64g. Le poids des tranches suit une loi Normale d'écart type égal à 5. Sur une fabrication de 400 paquets on en contrôle 36 et le contrôleur trouve un poids moyen du paquet égal à 61g.

Donnez (au 1/100ème), pour un risque de 0.1%, le critère statistique théorique qui permet de répondre à la question " les tranches sont-elles découpées trop finement?"

18 - Le fabricant de pâtes Kiluscu compare 2 chaînes de mises en paquets des pâtes, le poids des paquets étant distribué normalement. Afin de décider si les 2 chaînes étaient équivalentes, on prélève sur les 400 paquets de la première 18 paquets dont le poids moyen est égal à 90g avec une SCE = 900, et, sur les 400 paquets de la deuxième on en prélève 25 et trouve un poids moyen de 94g et une SCE = 920.

Pour un risque égal à 5%, sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question sous-entendue dans cette étude.

19 - Un fabricant de céréales affiche sur ses paquets une teneur lipides = 2.25g (pour 30g de produit). Sur 400 paquets fabriqués un contrôleur prélève 36 paquets et trouve une teneur moyenne égale à 2.21g et une SCE = 0.5.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question " le poids affiché sur chaque paquet est-il vrai?"

20 - Une entreprise emballe 16 minis bouchés à garnir pour proposer des paquets de 90 g. Le poids du mini bouché suit une loi Normale. Sur une fabrication de 400 paquets on en contrôle 36 et le contrôleur trouve un poids moyen du paquet égal à 92g avec une SCE égale à 900.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question " le poids de chaque paquet est-il aux normes?"

21 - Afin de vérifier si l'effet d'une substance modifie de 12.6 points le score de mémorisation de 200 individus, on fait subir à 36 d'entre eux un test de mémoire avant et après la prise de la substance et on obtient: moyenne des différences = 12 points et SCE des différences = 120.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique pour un risque de 5%, qui permet de répondre à la question de cette étude

22 - Afin de vérifier si l'effet d'une substance dégrade le temps de mémorisation de 50 individus, on fait subir à 9 d'entre eux un test de mémoire avant et après la prise de la substance et on obtient:

Avant	52	53	58	59	51	54	52	57	56
Après	51	54	58	58	52	55	53	58	58

Calculez au 1/1000^{ème} le critère statistique qui permet de répondre à la question posée dans cette étude.

23 - Un fabricant de céréales affiche sur ses paquets une teneur en vitamine B1 = 1.125mg avec un écart type égal à 0.05 mg. Sur 400 paquets fabriqués un contrôleur prélève 36 paquets et trouve une teneur moyenne égale à 1.105 mg.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question " le poids affiché sur chaque paquet est-il vrai?"

24 - Un fabricant de céréales affiche sur ses paquets une teneur en vitamine B1 = 1.125mg avec un écart type égal à 0.05 mg. Sur 100 paquets fabriqués un contrôleur prélève 18 paquets et trouve une teneur moyenne égale à 1.105 mg.

Donnez (au 1/100ème), pour un risque égal à 0.05, le critère statistique théorique qui permet de répondre à la question " le poids affiché sur chaque paquet est-il vrai?"

25 - Un fabricant de céréales affiche sur ses paquets une teneur en vitamine B1 = 1.125mg avec un écart type égal à 0.05 mg. Sur 100 paquets fabriqués un contrôleur prélève 36 paquets et trouve une teneur moyenne égale à 1.105 mg.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question " le poids réel de vitamine B1 dans chaque paquet est-il plus faible?"

26 - Le fabricant de pâtes Kilus cru compare 2 chaînes de mises en paquets des pâtes, le poids des paquets étant distribué normalement. Afin de décider si les 2 chaînes étaient équivalentes, on prélève sur les 100 paquets de la première 18 paquets dont le poids moyen est égal à 90g avec une $SCE = 900$, et, sur les 100 paquets de la deuxième on en prélève 25 et trouve un poids moyen de 93.5g et une $SCE = 920$.

Sans faire les calculs donnez l'application numérique du calcul du critère statistique qui permet de répondre à la question sous-entendue dans cette étude.

27 - Le fabricant de pâtes Kilus cru compare 2 chaînes de mises en paquets des pâtes. Des études antérieures avaient donné une dispersion de 20g et 30g pour les chaînes 1 et 2 respectivement et une distribution des poids normale. Afin de décider si une des chaînes surdose les paquets, on prélève sur les 400 paquets de la première 16 paquets dont le poids moyen est égal à 493g, et, sur les 400 paquets de la deuxième on en prélève 25 et trouve un poids moyen de 510g.

Sans faire les calculs donnez l'application numérique du calcul de l'écart type de la différence des moyennes.

28 - Un fabricant de céréales affiche sur ses paquets une teneur lipides = 2.25g (pour 30g de produit). Sur 100 paquets fabriqués un contrôleur prélève 36 paquets et trouve une teneur moyenne égale à 2.21g et une $SCE = 0.5$.

Donnez (au 1/100ème), pour un risque égal à 0.01, le critère statistique théorique qui permet de répondre à la question " le poids réel de lipides dans chaque paquet est-il plus faible?"

29 - Dans une fabrication une doseuse est réglée sur 125g, avec un écart type de 1g, le fabricant craint un dérèglement de la machine entraînant un surdosage égal à 70% de l'écart type. Sur 500 produits fabriqués il décide de contrôler 30 pièces et accepte de détecter à tort un dérèglement dans 5% des contrôles.

Sans faire les calculs donnez l'application numérique du calcul de la valeur critique de ce plan d'échantillonnage.

30 - Dans une fabrication on observe un problème sur 50% des pièces, le fabricant améliore le réglage de la machine et suppose avoir ramené ce taux à 30%. Sur 1000 produits fabriqués il décide de contrôler 60 pièces et accepte de détecter à tort une amélioration dans 5% des contrôles.

Sans faire les calculs donnez l'application numérique du calcul de la valeur critique de ce plan d'échantillonnage.

Statistique
EXAMEN N°1 – 15 Juin 2009
Mme B. BOTTOLIER

Conditions d'examens

Documents : **Poly cours complétés partie 6 autorisés + 1
feuille A4 manuscrite pour la partie 7
TD Non autorisés**
Tables statistiques autorisées
Calculatrice : **Tout type autorisée**

Remarques importantes :

A la fin de l'épreuve vous devez rendre le sujet de 7 pages, complété et vos brouillons (sous peine de perdre 2 points). Le sujet est noté sur 20 points.

Sauf indication contraire, vous proposerez vos résultats avec 2 chiffres significatifs.

NOM : **PRENOM :**

Application 1 (10 pts): (Inspiré de : Modélisation et optimisation par la méthode des plans d'expériences de la décoloration d'un sirop de sucre utilisé dans la formulation de boissons gazeuses. S. Belouafa et al, Sciences des aliments, 26 (2006) 29-36)

L'industrie des boissons gazeuses emploie des sirops de sucre qui doivent répondre à des normes qui touchent entre autres : la couleur. Or l'utilisation de sucre local pose un problème de coloration que l'on va essayer de résoudre en décolorant le sirop de sucre par adsorption sur charbon actif. Pour cela on va étudier les effets de la variation de paramètres :

- U1 = masse d'adsorbant (charbon actif) en grammes,
- U2 : la durée du contact adsorbât-adsorbant en minutes,
- U3 : le rapport des masses des farines constituant le gâteau de filtration,

La réponse Y étant le coefficient d'extinction spécifique noté ES qui est à minimiser ($ES < 20$). Le dispositif est dans l'annexe 1.

1°) (1 pt) Ecrivez le modèle préconisé.

2°) (1.5 pts) Dans le tableau ci-dessous, calculer les coefficients du modèle au $1/1000^{\text{ème}}$. Dans l'annexe 1 on donne la matrice de dispersion et la matrice $X'Y$.

j	0		2	
Coefficients b_j	3.265		2,338	

Interprétation de b_0 :

Interprétation de b_1 :

3°) (1 pt) Compléter le tableau ci-dessous :

N° essai	Y	Y^	e
1	20	23,234	-3,234
2	18	17,145	0,855
3	55	51,286	3,714
4	45	45,198	-0,198
5	20	22,019	-2,019
6	17	15,931	1,069
7	50	50,072	-0,072
8	46	43,983	2,017
9	40	38,784	1,216
10	40	37,262	2,738

N° essai	Y	Y^	e
11	11	9,998	1,002
12	53	57,219	-4,219
13	37	34,650	2,350
14	35
15	32
16	31	-2,609
17	25	-8,609
18	37	3,391
19	39	5,391
20	30	-3,609

Somme	681
SC	26203,000	25992,371

4°) (4 pts) Sous forme de tableau, Valider le modèle. Hypothèses, critères statistiques, conclusions. Résultats au 1/1000^{ème}. On donne :

Variation					p-value
De la régression					0,00000

H₀:

H₁:

Conclusion :

5°) (2.5 pts) Tester la significativité des effets des facteurs, hypothèses et conclusion.

Application 2 (10 pts): Dans un souci de respect de confidentialité le contexte de cette étude n'est pas détaillé et les résultats numériques ne sont pas fidèles.

Dans le cadre d'une certification AOC, une expérimentation est menée dans le but d'étudier l'influence de la mise en place d'un parcours grillagé pour inciter les poules à s'éloigner de leur bâtiment d'élevage. La réponse mesurée sera donc la distance (en m) parcourue par des poules. On dispose de 2 bâtiments ($i = 1, 2$) (bâtiment 1 sans grillage, bâtiment 2 avec grillage) et les mesures seront faites à 3 moments différents de la journée ($j = 1, 2, 3$) (H1, H2 et H3) et sur $r = 11$ poules différentes à chaque fois.

Vous tolérerez tout au long de cette étude un **risque d'erreur maximal de 10%**, les données sont dans l'annexe 2. **Rappel !!! SC = somme des carrés**

1°) (1pt)

- Combien y a-t-il de facteur(s) étudié(s) ?.....
- Combien y a-t-il d'unités statistique ?
- Combien y a-t-il de traitements ?.....
- Ecrivez le modèle associé :

2°) (1 pt) On supposera pour la suite que la distribution des résidus suit une loi normale et que les résidus sont indépendants entre eux. Compléter le tableau de la liste des résidus suivant :

e_{1jr}	14h30	17h45	11h30
Bâtiment 1 $\sum e^2_{1jr} = 292$	*****	*****	*****
	*****	*****	*****
	0	-6	-1
	4	3	-3
	-2	0	4
	2	2	0
	6	-1	2
	-6	-4	-1
	-4	-3	-3
	-1	1	0
	-3	3	-1
SC	132	110	50

e_{2jr}	14h30	17h45	11h30
Bâtiment 2 $\sum e^2_{2jr} = 508$	3	-7	-2
	2	9	2
	1	-7	10
	2	-2	1
	5	1	-1
	-4	5	3
	-4	4	2
	-4	0	-4
	-2	-1	-3
	0	-2	-5
	1	0	-3
SC	96	230	182

3°) (2 pts)

a) La vérification de l'hypothèse d'homoscédasticité entre les répétitions a donné un critère calculé = 6,55. Quelle est votre conclusion en encadrant la p-value ?

H0 :

H1 :

Conclusion :

b) Calculer le critère statistique pour vérifier l'homoscédasticité entre les bâtiments et concluez (on sait que la p-value du test est égale à 0,30). Précisez le (les) ddl.

H0 :

H1 :

Conclusion :

c) Pour vérifier l'homoscédasticité entre les heures de sortie des poules on a réalisé un test du X^2 , compléter le tableau suivant :

X^2 calculé =	1,11
ddl =	
X^2 théorique =	

H0 :

H1 :

Conclusion :

4°) (4 pts) Sous forme de tableau, réaliser l'analyse de la variance (utiliser les sommes données dans l'annexe 2). Hypothèses, critères statistiques, conclusions (tables en annexe 2). Résultats au 1/100^{ème}.

!!! en cas de SCE négative (problème d'arrondi) prenez la valeur 0 !!!

Hypothèses :

Conclusions :

5°) (2 pts) Proposer des groupements des moyennes des 2 facteurs pour un risque de 10% en utilisant la méthode de la ppds. Conclusion de l'étude.

Conclusions :

Annexe 1

N° essai	U1	U2	U3	Y
1	2	12	2, 3	20
2	4	12	2, 3	18
3	2	24	2, 3	55
4	4	24	2, 3	45
5	2	12	3, 7	20
6	4	12	3, 7	17
7	2	24	3, 7	50
8	4	24	3, 7	46
9	1, 3	18	3	40
10	1, 8	18	3	40

N° essai	U1	U2	U3	Y
11	3	7, 9	3	11
12	3	28, 1	3	53
13	3	18	1, 8	37
14	3	18	4, 2	35
15	3	18	3	32
16	3	18	3	31
17	3	18	3	25
18	3	18	3	37
19	3	18	3	39
20	3	18	3	30

inv			
2,716453	-0,239725	-0,036584	-0,441176
-0,239725	0,083967	0,000000	0,000000
-0,036584	0,000000	0,002032	0,000000
-0,441176	0,000000	0,000000	0,147059

X'Y
681,000
1908,000
13408,200
2037,100

Si besoin vous utiliserez la matrice X'Y pour identifier l'ordre d'introduction des facteurs

Annexe 2

xijr	14h30	17h45	11h30
Bâtiment 1 x1jr = 616	20	20	22
	18	25	19
	17	14	18
	21	23	16
	15	20	23
	19	22	19
	23	19	21
	11	16	18
	13	17	16
	16	21	19
	14	23	18
Somme	187	220	209

xijr	14h30	17h45	11h30
Bâtiment 2 x2jr = 715	23	16	20
	22	32	24
	21	16	32
	22	21	23
	25	24	21
	16	28	25
	16	27	24
	16	23	18
	18	22	19
	20	21	17
	21	23	19
Somme	220	253	242

xijr	14h30	17h45	11h30
Somme	407	473	451
SC	7807	10559	9527

Extraits des tables de la loi F . Valeur de $F_p(v1;v2)$

p = 0.9						p = 0.95					
v1	1	2	3	4	5	1	2	3	4	5	
v2											
1	39,86	49,50	53,59	55,83	57,24	161,4	199,5	215,7	224,6	230,2	
2	8,53	9,00	9,16	9,24	9,29	18,51	19,00	19,16	19,25	19,30	
3	5,54	5,46	5,39	5,34	5,31	10,13	9,55	9,28	9,12	9,01	
4	4,54	4,32	4,19	4,11	4,05	7,71	6,94	6,59	6,39	6,26	
5	4,06	3,78	3,62	3,52	3,45	6,61	5,79	5,41	5,19	5,05	
10	3,29	2,92	2,73	2,61	2,52	4,96	4,10	3,71	3,48	3,33	
11	3,23	2,86	2,66	2,54	2,45	4,84	3,98	3,59	3,36	3,20	
15	3,07	2,70	2,49	2,36	2,27	4,54	3,68	3,29	3,06	2,90	
16	3,05	2,67	2,46	2,33	2,24	4,49	3,63	3,24	3,01	2,85	
17	3,03	2,64	2,44	2,31	2,22	4,45	3,59	3,20	2,96	2,81	
18	3,01	2,62	2,42	2,29	2,20	4,41	3,55	3,16	2,93	2,77	
20	2,97	2,59	2,38	2,25	2,16	4,35	3,49	3,10	2,87	2,71	
21	2,96	2,57	2,36	2,23	2,14	4,32	3,47	3,07	2,84	2,68	
26	2,91	2,52	2,31	2,17	2,08	4,23	3,37	2,98	2,74	2,59	
27	2,90	2,51	2,30	2,17	2,07	4,21	3,35	2,96	2,73	2,57	
28	2,89	2,50	2,29	2,16	2,06	4,20	3,34	2,95	2,71	2,56	
29	2,89	2,50	2,28	2,15	2,06	4,18	3,33	2,93	2,70	2,55	
30	2,88	2,49	2,28	2,14	2,05	4,17	3,32	2,92	2,69	2,53	
31	2,87	2,48	2,27	2,14	2,04	4,16	3,30	2,91	2,68	2,52	
32	2,87	2,48	2,26	2,13	2,04	4,15	3,29	2,90	2,67	2,51	
35	2,85	2,46	2,25	2,11	2,02	4,12	3,27	2,87	2,64	2,49	
50	2,81	2,41	2,20	2,06	1,97	4,03	3,18	2,79	2,56	2,40	
60	2,79	2,39	2,18	2,04	1,95	4,00	3,15	2,76	2,53	2,37	

STATISTIQUE
Examen N°2 – 13 Juin 2008
Mme B. BOTTOLLIER

Conditions d'examens

Documents autorisés : **2 photocopiés complétés : partie 6 (analyse de la variance) + connaissances théoriques ; feuille format A4 manuscrite non photocopiée ; tables statistique**
TD Non autorisés

Calculatrice : **Tout type autorisée**

Remarques importantes :

A la fin de l'épreuve vous devez rendre le sujet de 5 pages complété et vos brouillons (sous peine de perdre 2 points). Le sujet est noté sur 20 points.

Sauf indication contraire, vous proposerez vos résultats avec 2 chiffres significatifs.

NOM : **PRENOM :**

APPLICATION 1 : (10 PTS)

Une expérimentation en randomisation a été menée pour mesurer les effets de 3 variétés d'une céréale (A, B, C) semées en 4 densités (d1 à d4) sur le rendement d'une culture. La réponse X mesurée en unité de poids par parcelle (annexe 1) est à maximiser.

1°) (1 pt) Donner l'équation modèle associé. Définissez chaque terme. Combien y a-t-il de traitements ?

2°) (1 pt) Interpréter l'analyse des résidus donnée dans l'annexe 2

3°) (2pts)

a) La vérification de l'hypothèse d'homoscédasticité entre les densités a donné un critère calculé = 1.85. Quelle est votre conclusion (encadrer le risque)?

b) Vérifier l'homoscédasticité entre les variétés. On donne le tableau ci-dessous à compléter.

Ecart-types	résiduels	estimé	facteur	Variété
A	B			C
0.267	0.118			

c) Quelle autre hypothèse portant sur les résidus doit être vérifiée ? On admettra pour la suite de l'étude que cette hypothèse est valide.

4°) (3 pts) Sous forme de tableau, réaliser l'analyse de la variance pour un risque de 5% (utiliser les sommes données dans l'annexe 3). Hypothèses, critères statistiques, conclusions. Résultats au 1/100^{ème}.

5°) (3 pts) Sous forme d'un tableau, grouper les moyennes du facteur variété pour un risque de 5% en utilisant la méthode de la ppas.

APPLICATION 2 : (10 PTS)

Un expérimentateur désirant minimiser les pertes d'un procédé a décidé d'étudier l'influence de 4 facteurs U_1 à U_4 . La réponse mesurée Y représente une perte financière sur 1 jour de fabrication, on admettra qu'elle suit une loi normale. Chaque essai est effectué à partir d'une même masse de matières premières. La matrice d'expérience optimale utilisée pour conduire l'expérimentation, ainsi que les résultats observés figurent dans l'annexe 4.

1°) (4 pts) Ecrivez le modèle en variables codées (X_j) avec la valeur numérique des coefficients ($1/1000^{\text{ème}}$) puis transformez ces coefficients pour écrire le modèle en variables naturelles (U_j).

2°) (3 pts) Sous forme de tableau, réaliser l'analyse de la variance pour un risque de 5%. Hypothèses, critères statistiques, conclusions. Résultats au $1/1000^{\text{ème}}$. On donne :
 $SCE \hat{y} = 1406.500$

3°) (3 pts) Sous forme de tableau, tester les coefficients du modèle en variables codées pour un risque de 5%, hypothèses et conclusions.

Proposer les niveaux des facteurs qui permettent de minimiser les pertes Y:

U1	
U2	
U3	
U4	

Annexe 1

Variété	Densité	X	Variété	Densité	X	Variété	Densité	X
A	d1	3,8	B	d1	4,3	C	d1	4,1
A	d2	4,5	B	d2	5	C	d2	4,3
A	d3	3,8	B	d3	4,5	C	d3	4,7
A	d4	3,9	B	d4	5	C	d4	4,9

Annexe 2

INDICES DE NORMALITE (coefficients de K.PEARSON)		
SYMETRIE	BETA 1 = 0.1	PROBA : 0.59
APLATISSEMENT	BETA 2 = 2.09	PROBA : 0.46

Annexe 3

Somme des 12 valeurs		52.8	
Somme des carrés des 12 valeurs		234.480	
Sommes Facteur Variété			
1 (A)	2 (B)	3 (C)	
16	18.8	18	
Sommes Facteur densité			
1 (d1)	2 (d2)	3 (d3)	4 (d4)
12.2	13.8	13	13.8

Annexe 4

	Facteur	$X_j = (U_j - U_{0j}) / \Delta U_j$	U_j	U1	U2	U3	U4	Y
U1	Température de cuisson (°C) $\Delta U1 = 30 \quad U_{01} = 230$	-1	200	200	20	60	80	36
		1	260	260	20	60	160	40
U2	Durée de cuisson (min) $\Delta U2 = 5 \quad U_{02} = 25$	-1	20	200	30	60	160	25
		1	30	260	30	60	80	62
U3	Taux d'humidité à la cuisson (%) $\Delta U3 = 15 \quad U_{03} = 75$	-1	60	200	20	90	160	20
		1	90	260	20	90	80	57
U4	Eufs (g) $\Delta U4 = 40 \quad U_{04} = 120$	-1	80	200	30	90	80	37
		1	160	260	30	90	160	32

STATISTIQUE
Examen du 15 Juin 2007
Mme B. BOTTOLIER

Conditions d'examens

Documents : **Tables statistiques autorisées**

Calculatrice : **Non programmable autorisée**

Remarques importantes :

- Traiter chaque application sur **2 feuilles doubles séparées**,
- Présenter vos résultats sous forme de tableaux dès que cela est possible.
- Vous devez obligatoirement définir les modèles et l'interprétation de chacun des termes.
- Poser toutes les hypothèses testées et interpréter les résultats de chaque test pour un risque de 5%.
- Dans chaque application on supposera vérifiées les hypothèses de normalité.
- Vous proposerez vos résultats avec 2 chiffres significatifs. Tous les calculs intermédiaires ne sont pas nécessaires, faites le bon choix !
- Le sujet est noté sur 40 points.

Rappel :

$$\chi^2_{\text{calculé}} = \left(\frac{2.3026}{C} \right) * \left[v * \log_{10} \hat{\sigma}^2 - \sum_{j=1}^{j=k} (v_j * \log_{10} \hat{\sigma}_j^2) \right]$$

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{j=1}^{j=k} \frac{1}{v_j} - \frac{1}{v} \right]$$

$$v_j = n_j - 1 \qquad v = \sum_{j=1}^{j=k} v_j$$

$$\hat{\sigma}_j^2 = \frac{SCE_{x_j}(\bar{x}_j)}{n_j - 1} \qquad \hat{\sigma}^2 = \frac{1}{v} \sum_{j=1}^{j=k} v_j \hat{\sigma}_j^2$$

Application 1

On possède les rendements en foin de pois fourragers (en 1/100 de livres) de 3 variétés mesurées selon 2 écartements (en pouces). On désire définir l'écartement optimal pour chaque variété pour maximiser les rendements. En utilisant, lorsque cela se présente, les méthodes de votre choix, que proposeriez-vous ? Les résultats expérimentaux sont dans le tableau ci-dessous.

	Variété 1		Variété 2		Variété 3	
	4 pouces	12 pouces	4 pouces	12 pouces	4 pouces	12 pouces
	42	66	65	51	51	73
	43	57	69	52	62	77
	45	51	65	47	61	77
	54	50	61	54	54	65
somme	184	224	260	204	228	292
SC	8554	12706	16932	10430	13082	21412

NOMBRE DE MOYENNES	2	3	4	5	6
VALEURS DES PPAS (5%)	7,77	9,43	10,44	11,18	11,75

Application 2

Titre de l'étude : Gélification de denrées alimentaires

L'utilisation de denrées alimentaires re-texturées implique une stabilité dans le temps des caractéristiques des gels, notamment la rétention d'eau. Il est donc souhaitable de connaître les facteurs susceptibles de diminuer la perte en eau (synérèse) de ces gels après 24h de stockage, afin d'obtenir des produits commerciaux d'aspect appétissant. Les auteurs de l'expérimentation émettent l'hypothèse que l'absence ou la présence des composants : fibre d'avoine, amidon, Guar et Xanthane pourraient jouer un rôle sur le(s) critère(s) de qualité du gel de poivron vert.

Ci joint la liste des facteurs de l'étude ainsi que leurs niveaux en variables naturelles et codées. Ces paramètres sont indépendants et donc l'utilisation des variables codées est possible et va permettre de simplifier les calculs.

	-1	0	+1
U1 = Avoine	0	0.05%	0.1%
U2 = Amidon	0	1.5%	3%
U3 = Guar	0	0.25%	0.5%
U4 = Xanthane	0	0.25%	0.5%

Les auteurs préconisent dans cette étude un modèle linéaire multiple du 1^{er} degré sans interaction, ils ont donc réalisé 12 essais selon des critères d'optimalité, le dispositif expérimental en variables naturelles est détaillé ci après. La colonne Y correspond à la mesure de la synérèse.

Essais	U1	U2	U3	U4
1	0	0	0	0
2	0,1	0	0	0,5
3	0	3	0	0,5
4	0,1	3	0	0
5	0	0	0,5	0,5
6	0,1	0	0,5	0
7	0	3	0,5	0
8	0,1	3	0,5	0,5
9	0,05	1,5	0,25	0,25
10	0,05	1,5	0,25	0,25
11	0,05	1,5	0,25	0,25

Y	Y^	e
38		
32		
22		
36	33,273	2,727
24	21,273	2,727
28	30,273	-2,273
20	21,273	-1,273
22	19,273	2,727
24	27,273	-3,273
26	27,273	-1,273
28	27,273	0,727

somme
SC

300
8528

300
8471,818

Ces essais ont été réalisés pour définir l'ordre d'importance des paramètres et aussi pour proposer une combinaison des facteurs qui permette de minimiser la synérèse. En analysant ces résultats selon un modèle complet, quelles seraient vos conclusions ?

L'analyse de cette application nécessite de passer par le calcul des coefficients, de leur intervalle d'estimation et leur test ainsi que la validation du modèle et le calcul du coefficient de détermination multiple.

STATISTIQUE
EXAMEN N° 2
Mme B. BOTTOLIER

Conditions d'examens	Documents	X	Autorisés
			Non autorisés
	Calculatrice		Non autorisée
			4 opérations autorisée
		X	Tout type autorisé

Remarques particulières

Tables statistiques autorisées.

A la fin de l'épreuve vous devez rendre les 7 feuilles du sujet complétées et vos brouillons,
Sur le sujet figure le barème de chaque partie entre parenthèses (total de 40 points).

Les 2 applications sont indépendantes.

NOM : **Prénom :** **Note/20 :**

APPLICATION 1

La mise en œuvre d'un outil de raisonnement de la fertilisation azotée du Tournesol sur plusieurs types de parcelles d'une région a amené les agriculteurs à conduire 30 parcelles de 2 manières différentes: - fertilisation au semis (SEM) - fertilisation en végétation (VEG)
Les 2 types de conduites de fertilisation azotée seront analysés au niveau des rendements.

Les 30 parcelles sont regroupées de la manière suivante en fonction de leurs réserves:
- sol superficiel - sol moyen - sol profond

On cherchera à déterminer s'il existe un effet parcelle et/ou type de conduite.

On suppose que la variable mesurée suit une loi Normale.

Les premiers résultats ont donné :

Somme des 30 valeurs = 902,30
Somme des carrés des 30 valeurs = 27876,090

Moyennes selon le type de sol

superficiel	moyen	profond
24,88	29,83	35,52

Moyennes selon la pratique culturale

SEM	VEG
29,55	30,61

Moyennes selon les répétitions

	superficiel	moyen	profond
SEM	24,46	29,18	35
VEG	25,3	30,48	36,04

Partie A (2 points): Identification de l'étude

Identifiez le(s) facteur(s) étudié(s) et la variable mesurée :

Identifier le modèle mathématique associé à cette étude expérimentale :

Partie B (6 points): Analyse des résidus

On donne un extrait des résultats en ce qui concerne tous les sols classés « profond » :

type de sol	pratique culturale	REPET	N° parcelle	Rendement	résidus
profond	SEM	r1	402	30,7	
profond	SEM	r2	201	34,6	
profond	SEM	r3	203	35,8	0,80
profond	SEM	r4	306	37,8	2,80
profond	SEM	r5	401	36,1	1,10
profond	VEG	r1	405	32,6	
profond	VEG	r2	503	36,9	
profond	VEG	r3	206	37,2	1,16
profond	VEG	r4	301	37,4	1,36
profond	VEG	r5	102	36,1	0,06

Complétez la liste des résidus (au $1/100^{\text{ème}}$) et tester si toutes les hypothèses d'invariance des variances résiduelles sont vérifiées pour un risque de 1%. On donne :

1°) Pour le test sur les variances résiduelles des répétitions on a obtenu $KH12 = 2.85$

Critère statistique théorique :

Conclusion :

2°) Les écarts types résiduels selon la pratique culturale

SEM	VEG
2,57	2,23

H0:

H1 :

Critères statistiques calculé et théorique :

Conclusion:

3°) Les écarts types résiduels selon le type de sol (à compléter) :

superficiel	moyen	profond
2,17	2,90	

H0 :

H1 ..

Critères statistiques :

Conclusions:

Partie C (5 points) : Analyse de la variance

Tester l'influence du (des) facteur(s) étudié(s) sur la réponse sous forme de tableau pour un risque de 5%. Calculs aux 1/100^{ème}.

Hypothèses :

Analyse statistique :

Conclusion(s) :

Partie D (5 points) (calculs au 1/100^{ème})

1°) Réaliser le groupement des classes de sol selon les rendements moyens calculés sous forme de tableaux. Préciser vos hypothèses, la significativité de vos réponses et justifier vos conclusions.

2°) Donner un intervalle d'estimation pour un risque de 0.1% du rendement moyen du sol profond.

APPLICATION 2

On cherche à déterminer de quelle manière la teneur en huile des graines de Tournesol varie en fonction du rendement de la parcelle, On dispose de 30 parcelles classées en sol moyen, les résultats ont donné:

	Rendement	Teneur en huile
Somme	918,500	1353,300
Somme des carrés	28455,670	61082,430
Somme des produits	41510,510	

Partie A (6 points): Les coefficients de la droite de régression

1°) Compléter le tableau suivant à 10^{-3} près :

	Coefficients	Ecart-type	Critère Statistique calculé	significativité	Lim inférieure Risque =10%	Lim supérieure Risque =10%
Constante						
Effet du régresseur						

Justifiez la significativité :

α	
0.05	
0.01	
0.001	

Conclusion sur la significativité :

2°) Donnez l'équation de la droite d'ajustement, précisez quelle variable joue le rôle du régresseur:

Régresseur :

Equation :

3°) Donnez les interprétations concrètes de :

	Coefficients	Intervalles d'estimation
Constante		
Pente		

Partie B (5 points): Validité du modèle

Testez la validité de la régression linéaire en complétant (calculs au 1/1000^{ème}) :

On donne la somme des carrés des résidus = 17.341

H0 :

H1 :

	<i>ddl</i>	<i>SCE</i>	<i>CM</i>	<i>Critère statistique calculé</i>	<i>alpha</i>
Régression					1,0662E-05
Résidus					
Total					

Conclusion :

Partie C (5 points): Qualité de la régression

1°) Calculez et interprétez le coefficient de détermination

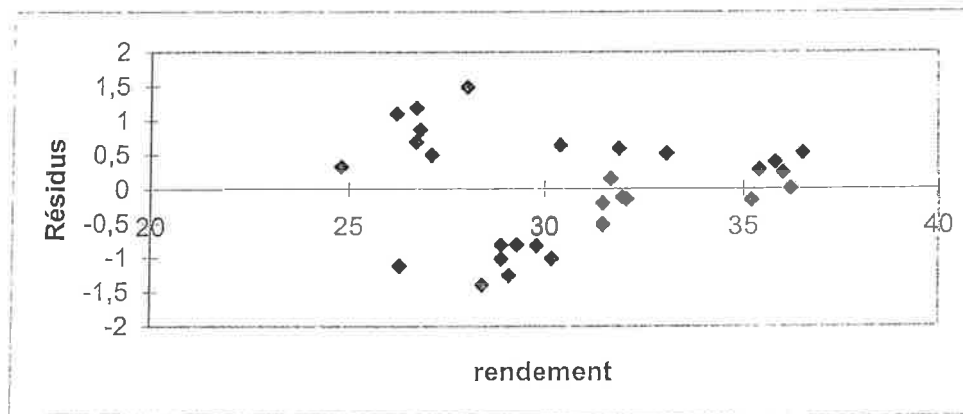
2°) Proposez un encadrement de r^2 pour un risque de 5% et concluez (au 1/1000^{ème}) :

3°) Compléter les données suivantes relatives à la parcelle N°1 :

parcelle	X	Y observé	Y estimé	Résidu	Résidu standardisé
1	26,7	44,9			

Explicitiez le rôle des résidus standardisés :

4°) Interprétez ce graphe :



Partie D (6 points) : Quelques questions

Supposez que dans cette étude on ait ajouté 2 autres régresseurs.

1°) Quelle aurait été la méthode de calcul des coefficients? précisez les valeurs numériques des dimensions mises en jeu.

2°) Quelle aurait été la méthode de calcul des écarts types des coefficients?

3°) Quel aurait été la valeur du ddl des tests sur les coefficients ?

Conditions d'examen :

Documents	(X) NON AUTORISES
Calculatrice	Autorisée
	4 opérations autorisées
	(X) NON AUTORISEE

Durée : 2 heures

Comptage des candidatures à l'ISARA-Lyon
Admissions parallèles

Objectifs :

L'entrée à l'ISARA-Lyon peut se faire en 1^{ère} année, après le passage du baccalauréat mais également en 2^e, 3^e ou 4^e année (procédure dite des **admissions parallèles**).
La Conférence des Grandes Ecoles (CGE) s'intéresse à ce type d'entrée dans les écoles d'ingénieurs et demande à chacune d'entre elles de lui fournir, chaque année, des informations sur les candidats entrés par ce système d'admission.

Les besoins décrits ci-dessous correspondent à une partie de la demande effectuée par cette commission et l'exemple fourni est un jeu d'essai et non des données réelles.

Données en entrée

- n données :
- numéro INE du candidat (11 caractères)
 - département d'origine
 - Année de candidature (2, 3 ou 4)
 - Boursier sur critères sociaux (O ou N)
 - Age

Pour arrêter l'entrée des données, on entre un n° INE de candidat égal à 0.

Exemple :

1234567890A
69
2
O
22
0987654321B
74
2
O
23
2345678901C
78
3
N
23
4592111315G
92
4
O
23
0

- comptage du nombre d'étudiants entrés (total et par année d'entrée)
- comptage du nombre de boursiers issus de la Région Rhône-Alpes (départements 01, 07, 26, 38, 42, 69, 73, 74) et hors Région Rhône-Alpes et pourcentage du nombre total d'étudiants boursiers entrés.
- comptage du nombre de non-boursiers issus de la Région Rhône-Alpes (départements 01, 07, 26, 38, 42, 69, 73, 74) et hors Région Rhône-Alpes et pourcentage du nombre total d'étudiants non boursiers entrés
- âge moyen des étudiants (sur le nombre total d'étudiants entrés)

Sortie à obtenir : (contenant les données et résultats obtenus avec l'exemple ci-dessus)

Analyse des admissions parallèles			
ISARA-Lyon			
1 ^{ère} partie			
Nombre total d'étudiants entrés en septembre :			
- entrés en 2 ^{ème} année :	4		
- entrés en 3 ^{ème} année :	2		
- entrés en 4 ^{ème} année :	1		
Etudiants boursiers			
Rhône-Alpes	nombre	2	
	%	66,66%	
Hors Rhône-Alpes	nombre	1	
	%	33,33%	
Etudiants non boursiers			
Rhône-Alpes	nombre	0	
	%	0%	
Hors Rhône-Alpes	nombre	1	
	%	100%	
Age moyen des étudiants		22,75	

Travail à réaliser :

Etablir **le pseudo-code détaillé** du programme Visual BASIC qui réalise le traitement décrit ci-dessus.

CONSIGNES PARTICULIERES :

Respecter les règles d'écriture des pseudo-codes (cf. p 4)
Il est possible de rédiger l'exercice au crayon à papier
Pour la sortie des résultats, il suffit d'indiquer les variables concernées dans les ordres de sortie; il n'y a pas dans le pseudo-code de référence à la présentation de la sortie.

BAREME

Entrée des données	/2
Boucle de traitement et valeur d'arrêt	/3
Comptages (totaux, boursiers, non boursiers)	/4
Calculs (pourcentages, âge moyen)	/4
Edition des résultats	/3
Structure générale du pseudo-code, respects des consignes (mot-clé, variables : cf p4), absence de fautes graves	/4