

Le 12 juin 2014

**EXAMEN US : STATISTIQUES APPLIQUÉES Code : 413Q**

Sujet de France LAPLUME et Colette VUILLET

Durée 2 h 30

**PARTIE I. Régression linéaire (Fiche n°1)**

Pour contrôler l'influence du fertilisant x, sur la biomasse aérienne Y d'un blé, on a réalisé une statistique résumée dans la Fiche n°1.

Y est la variable à expliquer, x est la variable explicative.

On considère que le lien entre x et Y est linéaire si une droite d'équation  $y = a + bx$  s'ajuste au nuage de points.

- 1) À l'aide de la Fiche n°1, conclure quant à ce lien.
- 2) Si  $x = 120$ , quelle valeur de Y proposez-vous ?

**PARTIE II. Analyse de la variance à un facteur (Fiches n°2 et n°3)**

À l'aide des Fiches n°2 et n°3, étudiez la quantité de matière azotée d'un fromage en fonction de la recette utilisée (type de fromage), quel modèle vous paraît le mieux adapté ?

**PARTIE III. Analyse en composantes principales (Document 1)**

Le Document 1 donne les résultats d'une ACP normée concernant l'analyse d'une expérimentation menée sur des melons. Pour chacun de ces melons, les variables quantitatives suivantes ont été étudiées :

- le poids,
- le diamètre équatorial : « D-equatorial »,
- le diamètre haut : « D-haut »,
- le diamètre du calice : « D-calice »,
- la fermeté : « fermete »,
- l'indice de réfraction, mesuré par infra rouge, donnant une bonne mesure du goût sucré du melon, donc de sa qualité gustative ; il est noté « refraction ».

- 1) Valait-il mieux faire une ACP brute ou normée ? Les deux sont-elles possibles ici ?
- 2) Quel pourcentage d'inertie est obtenu en projection dans le plan 1-2 ?
- 3) Interpréter les axes 1 et 2.
- 4) Expliquer la position des vecteurs « fermete » et « poids ». Quelle variable influence le plus l'indice de réfraction ?

## PARTIE IV : Tests d'hypothèses

*Les 3 exercices de cette partie sont indépendants.  
On s'aidera avec profit des pages 4 et 5 du résumé.*

- 1)** Un jury de 40 consommateurs réguliers de pain est réuni en vue d'un test d'analyse sensorielle. Chaque personne reçoit trois morceaux de pain courant français apparemment identiques. En fait, l'un des morceaux est de type I (il contient moins de sel) et les deux autres sont de type II. Après avoir goûté les trois morceaux de pain, chaque personne doit noter quel est le morceau qui lui semble contenir le moins de sel. Chaque personne doit apporter une réponse même si elle est indécise.

À partir de combien de bonnes réponses peut-on décider que les perceptions salées des 2 types de pain sont significativement différentes (risque  $\alpha = 0,05$ ) ?

Dans ce test d'hypothèse, on expliquera pourquoi les hypothèses sont :

$$H_0 : p = 1/3 \quad \text{et} \quad H_1 : p > 1/3$$

- 2)** Une remplissage de paquets de farine est considérée conforme si la moyenne des poids des paquets est de 250 g avec un écart type inférieur à 1,7 g. On contrôle le processus en pesant 5 paquets pour lesquels on trouve une moyenne de 249 g et un écart type de 2 g. On suppose que le poids des paquets produits suit une loi normale dont la moyenne et l'écart type sont tous les deux inconnus.

Effectuer un test d'hypothèse sur la moyenne puis sur l'écart type pour décider si le processus est satisfaisant (risque  $\alpha = 0,05$  pour les deux tests).

- 3)** On suppose que les rendements des variétés de blés suivent des lois normales.

Les rendements de deux variétés de blé A et B ont été relevés dans 10 parcelles différentes. On a obtenu les résultats suivants en q/ha :

Blé A	75	62	86	79	75	68	77	81	72	68
Blé B	77	64	82	81	78	74	75	86	76	74

Calculer la moyenne et l'écart type estimé de ces 2 échantillons à  $10^{-2}$  près.

Montrer par un test d'hypothèse que les 2 écarts types ne sont pas significativement différents (risque  $\alpha = 0,05$ ) et calculer une estimation de cet écart type commun.

On utilisera ensuite cette estimation pour tester si les moyennes des rendements sont significativement différentes (risque  $\alpha = 0,05$ ).

**FICHE n°1****Numéro du candidat :**

Population : des blés

Variable à expliquer : Y = Biomasse aérienne du blé

Variable explicative : x = Dose de fertilisant

**Valeurs observées de Y et x**

n° obs	1	2	3	4	5	6	7	8	9	10
x <sub>i</sub>	0	0	0	160	160	160	240	240	240	240
Y <sub>i</sub>	848	1070	963	1153	1080	1334	1276	1319	1400	1200

**Équation de la droite de régression :  $\hat{Y} =$** **Résidus**

	1	2	3	4	5	6	7	8	9	10
$\hat{Y}_i$	961	961	961					1300	1300	1300
$\hat{E}_i = Y_i - \hat{Y}_i$	-113	109	-2							-100

**Nuage de points****Graphe des résidus****Table d'analyse de la variance**

Source de variabilité	Somme des carrés	ddl	Carré moyen	Fischer
Modèle	SCM = 198925			
Résidus	SCR =			
Total	SCT = 278730			

$R^2 =$

$\hat{\sigma} =$

Répondre aux questions au dos de la feuille

**FICHE n°2****Numéro du candidat :**

Population : Fromages

Variable à expliquer : Y = Matière azotée en g pour 100 g

Facteur A : type de fromage, A1, A2, A3

Y	A1				A2				A3			
	36.7	32.1	26.8	32.3	46.0	56.0	45.8	35	17.4	28.6	25.0	26.4
Y <sub>ik</sub>												

**Moyennes des valeurs observées**

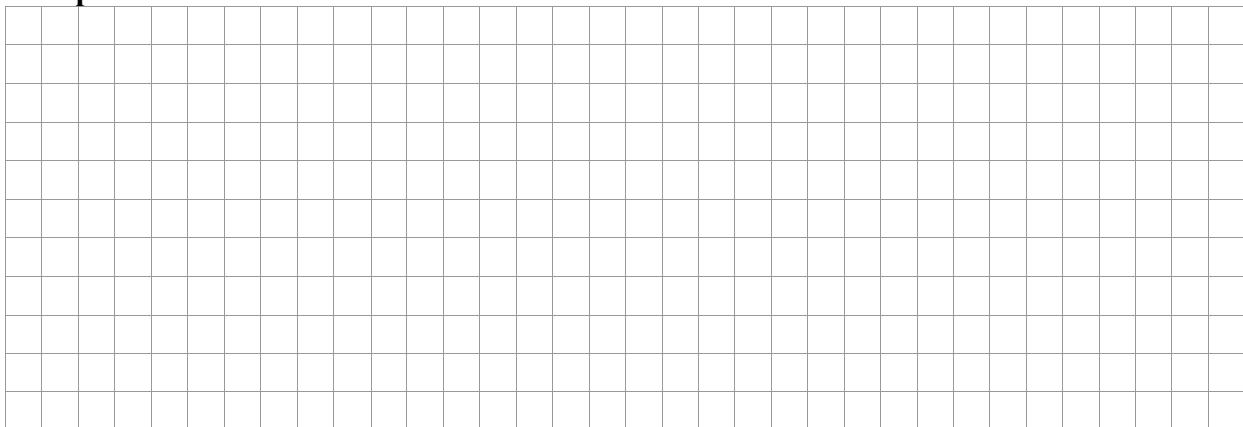
	A1	A2	A3	
	Y <sub>1.</sub> =	Y <sub>2.</sub> =	Y <sub>3.</sub> =	Y <sub>.</sub> =

**Modèle : Y<sub>1k</sub> =****Estimations :****Moyennes prédites**

	A1	A2	A3	
	ŷ <sub>1</sub> =	ŷ <sub>2</sub> =	ŷ <sub>3</sub> =	ŷ =

**Résidus**

	A1				A2				A3			
					-1.7	10.3	0.1	10.7	-6.9	4.3	1.7	2.1

**Graphe des résidus :****Que concluez-vous avec le graphe des résidus ?**

**FICHE n°3****Numéro du candidat :**

Population : Fromages

Variable à expliquer : ln(Matière azotée)

Facteur A : Type de fromage, A1, A2, A3

Y	A1				A2				A3			
	3.6	3.47	3.29	3.47	3.82	3.55	2.86	3.35	2.85	3.35	3.22	3.27
Y <sub>ik</sub>												

**Moyennes des valeurs observées**

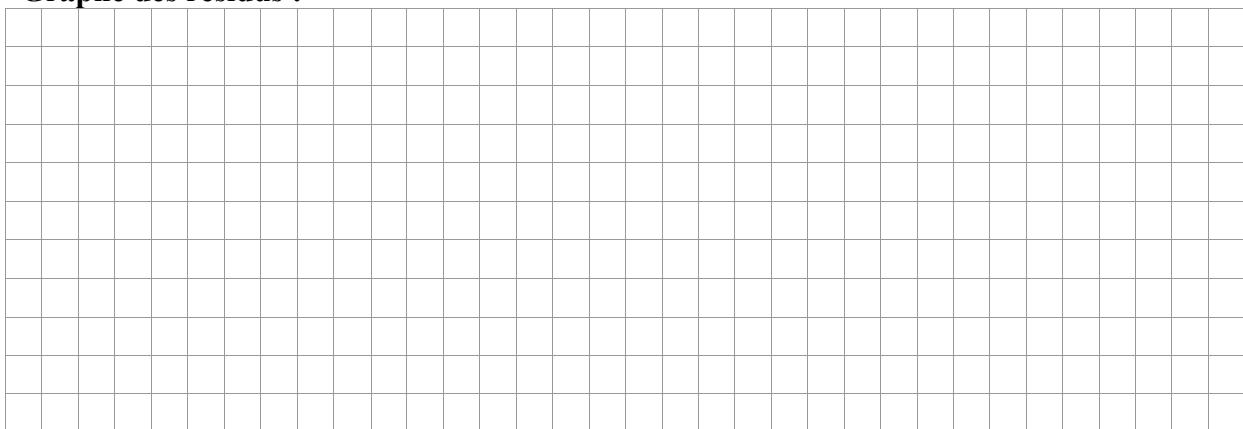
	A1	A2	A3	
Y <sub>1.</sub> =				Y <sub>..</sub> =

**Modèle :** Y<sub>1k</sub> =**Estimations :****Moyennes prédites**

	A1	A2	A3	
ŷ <sub>1</sub> =				ŷ =

**Résidus**

	A1			A2			A3					
	1.4	0.01		0.01	0.42	0.15				0.18	0.05	0.1

**Graphe des résidus :****Table d'analyse de la variance**

Source	Somme des carrés	ddl	Carré moyen	Fischer
Modèle	SCM =			
Résidus	SCR = 0.307			
Total	SCT = 1.11			

$$\frac{SCM}{SCT} = \hat{\sigma}^2 =$$

# DOCUMENT 1

## Principal Components Analysis

Component              Percent of Cumulative

Number	Eigenvalue	Variance	Percentage
1	3,22958	53,826	53,826
2	1,34564	22,427	76,254
3	0,696353	11,606	87,860

