

Le 9 juin 2011

EXAMEN UE : STATISTIQUES APPLIQUÉES Code : BCA016

Sujet de France LAPLUME et Colette VUILLET

Durée 2 h 30

PARTIE I. Régression linéaire (Fiche n°1)

Pour vérifier l'étalonnage d'un spectrophotomètre proche infrarouge, on a mesuré la teneur en protéines d'un blé par la méthode de référence (x), puis avec le spectrophotomètre (Y).

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

On considère que l'appareil est bien réglé si la droite d'équation $y = 1x + 0$ s'ajuste au nuage de points.

- 1) À l'aide de la Fiche n°1, conclure quant à au réglage de l'appareil.
- 2) Si $y = 11,2$, quelle valeur de x proposez-vous ? Quel est l'écart type de cette mesure ?

PARTIE II. Analyse de la variance à un facteur (Fiche n°2)

On reprend l'exemple de la pollution des poissons, mais cette fois, l'âge A est considéré comme un facteur à trois modalités : A1 jeune, A2 mûre, A3 âgé.

La variable à expliquer Y est le logarithme du taux de DDT.

À l'aide de la Fiche n°2, conclure à propos de l'influence de l'âge sur le taux de pollution par le DDT.

PARTIE III. Analyse en composante principale (Fiche n°3)

Le document donne les résultats d'une ACP standardisée (ou normée) sur l'analyse nutritionnelle de 98 produits proposés par la chaîne de restauration Quick : hamburgers, clubs, salades, finger food, frites, boissons, sauces, desserts, et menus enfants.

Pour chacun de ces produits, les variables quantitatives suivantes ont été étudiées :

- | | |
|-------------------|----------------------|
| - Quantités (g) | - Lipides (g) |
| - Calories (kcal) | - Protéines (g) |
| - Acides gras (g) | - Sucres rapides (g) |
| - Fibres (g) | - Sel (g) |
| - Glucides (g) | |

- 1) Valait-il mieux faire une ACP brute ou standardisée ?
- 2) Quel pourcentage d'inertie est obtenu en projection dans le plan 1-2 sachant que :
 $\lambda_1 = 4,2$ et $\lambda_2 = 1,9$?
- 3) Interpréter les 4 zones (clusters) de l'ACP.

PARTIE IV. Détection de valeurs aberrantes (Fiches n°4, 5 et Annexe)

12 laboratoires ont participé à un *ring test* relatif à la détermination du temps de développement d'une farine par une nouvelle méthode.

Chaque laboratoire a reçu 3 échantillons de farine notés A, B, C. Les laboratoires ont réalisé 2 mesures du temps de développement sur chacun des 3 types de farine et calculé la moyenne \bar{y} et la variance estimée s^2 de ces 2 mesures. On rappelle que :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Les questions 2. et 3. sont indépendantes.

1. Compléter les cases grisées.

2. Détection des **variances d'échantillon** douteuses ou aberrantes par le test de Cochran

Pour chacune des 3 farines, déterminer s_{\max}^2 la plus grande variance estimée obtenue par les laboratoires, puis calculer le nombre g défini par la formule :

$$g = \frac{s_{\max}^2}{\sum s_i^2}$$

Comparer le nombre g obtenu aux valeurs critiques à 5 % et à 1 % (C5% et C1%) données par la table du test de Cochran (Annexe) et conclure en utilisant la règle de décision suivante :

- si g est inférieur aux 2 valeurs critiques C5% et C1%, on décide qu'aucune variance n'est aberrante,
- si g est compris entre ces 2 valeurs critiques, l'échantillon de variance maximum est déclaré douteux ; il est éliminé et on recommence le test avec les observations restantes,
- si g est supérieur aux 2 valeurs critiques, l'échantillon de variance maximum est déclaré aberrant ; il est éliminé et on recommence le test avec les observations restantes.

3. Détection des **moyennes d'échantillon** douteuses ou aberrantes par le test de Dixon

Pour chacune des 3 farines, ordonner les moyennes obtenues par tous les laboratoires.

Calculer ensuite (en détaillant les calculs) la valeur de r_1 puis celle de r_n selon les formules données par la table du test de Dixon (Annexe).

Comparer r_1 et r_n aux valeurs critiques à 5 % et à 1 % (D5% et D1%) données par cette même table et conclure en utilisant la règle de décision suivante :

- si r_1 et r_n sont inférieurs aux 2 valeurs critiques D5% et D1%, on décide qu'aucune moyenne n'est aberrante,
- si r_1 ou r_n est compris entre ces 2 valeurs critiques, l'échantillon de moyenne minimum ou maximum est déclaré douteux ; il est éliminé et on recommence le test avec les observations restantes,
- si r_1 ou r_n est supérieur aux 2 valeurs critiques, l'échantillon de moyenne minimum ou maximum est déclaré aberrant ; il est éliminé et on recommence le test avec les observations restantes.

FICHE n°1**Numéro du candidat :**

Population : blés

Variable à expliquer : Y = taux de protéines par infrarouge

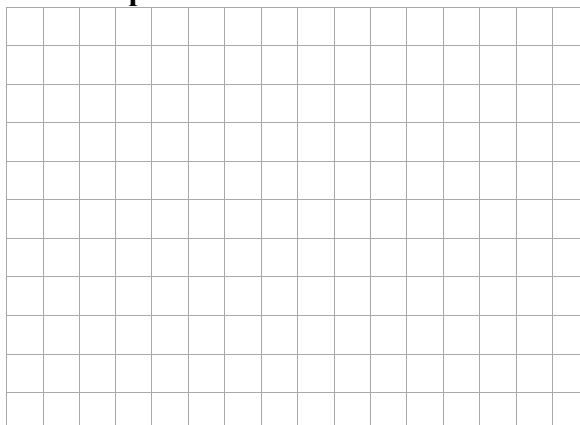
Variable explicative : x = taux de protéines par la méthode de référence

Valeurs observées de Y et x

n° obs	1	2	3	4	5	6	7	8	9	10
x_i	9,1	12,4	12,5	10,1	10,6	11,5	10,8	11,3	11,6	12,9
Y_i	8,6	11,1	11,6	10,9	10,0	11,6	10,4	11,6	12,9	13

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

	1	2	3	4	5	6	7	8	9	10
\hat{Y}_i	9,20	12,18	12,27	10,10	10,56	11,36				12,63
$\hat{E}_i = Y_i - \hat{Y}_i$										

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

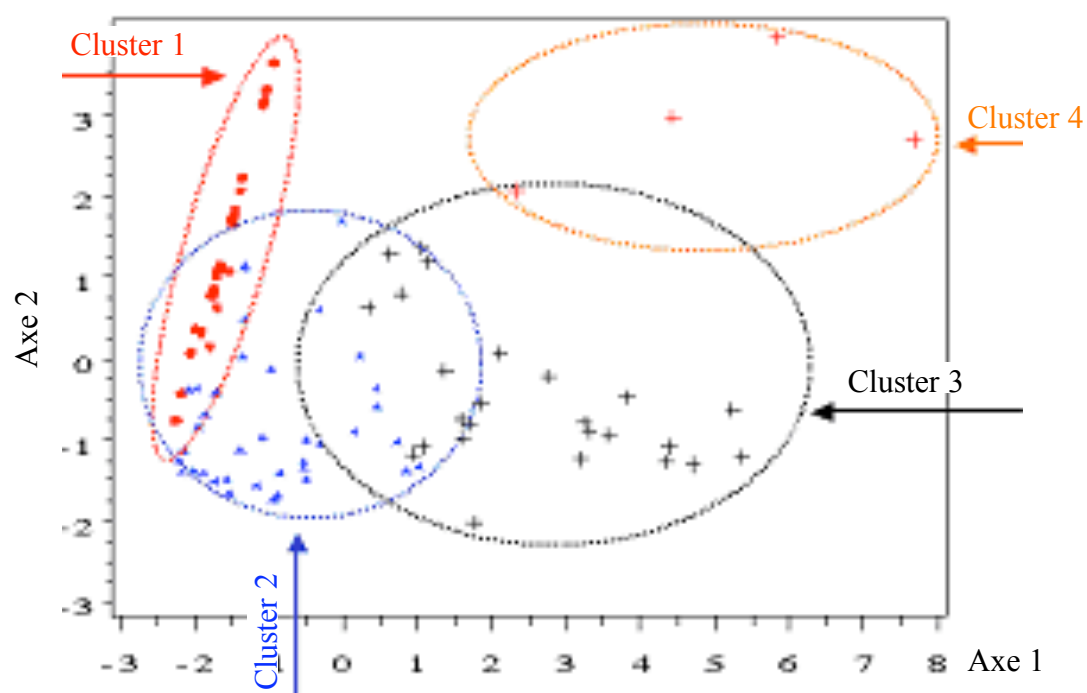
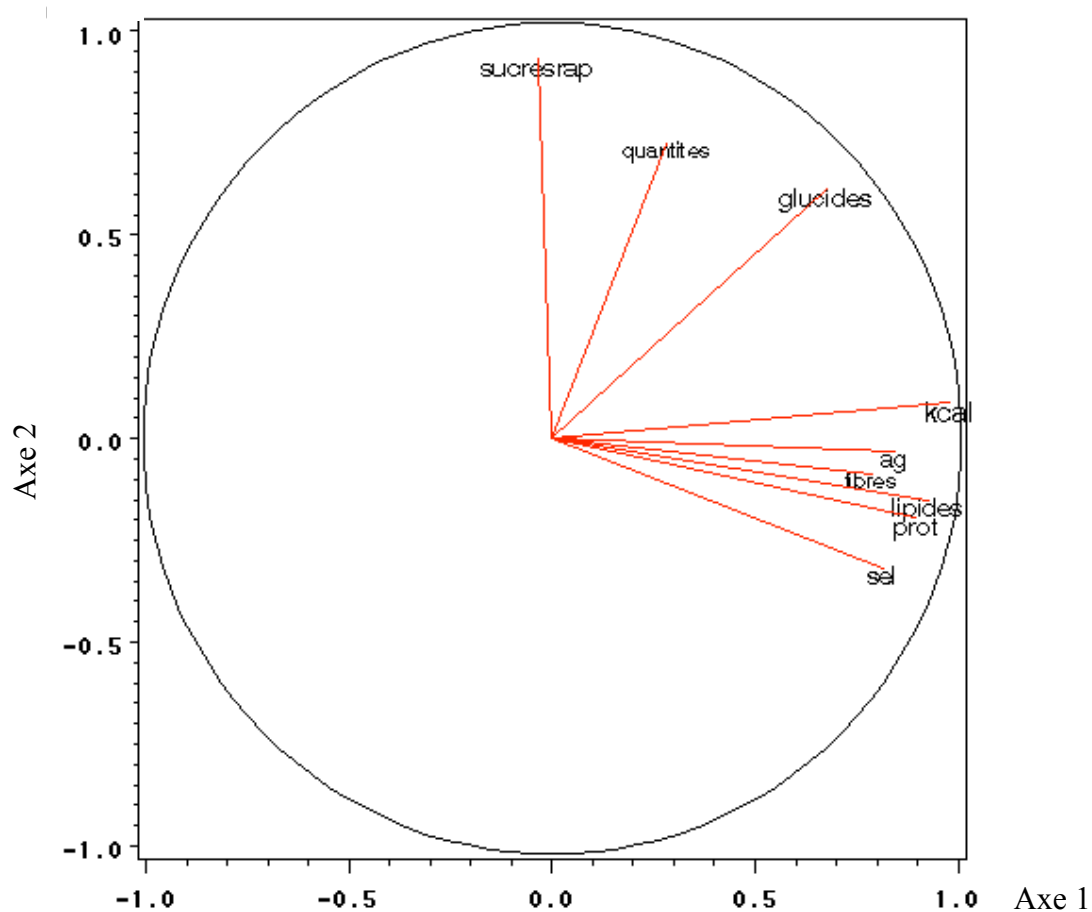
Source de variabilité	Somme des carrés	ddl	Carré moyen	Fischer
Modèle	SCM =			
Résidus	SCR = 5,47			
Total	SCT = 15,54			

 $R^2 =$ $\hat{\sigma} =$ **Répondre aux questions au dos de la feuille**

FICHE n°3

Numéro du candidat :

Résultats ACP normée des données Quick
Cercle des corrélations



FICHE n°4

Numéro du candidat :

Temps de développement (min)	Farine A				Farine B				Farine C			
	mesure 1	mesure 2	moyenne	variance estimée	mesure 1	mesure 2	moyenne	variance estimée	mesure 1	mesure 2	moyenne	variance estimée
Labo 1	3,8	4,5	4,15	0,245	1,7	1,6	1,65	0,005	1,8	1,7	1,75	0,005
Labo 2	5,1	4,8	4,95	0,045	2,3	1,8	2,05	0,125	1,8	2,1	1,95	0,045
Labo 3	3,7	4,0			1,8	1,6	1,70	0,020	1,9	1,7	1,80	0,020
Labo 4	3,8	3,9	3,85	0,005	1,6	1,8	1,70	0,020	1,8	2,0	1,90	0,020
Labo 5	4,4	4,6	4,50	0,020	1,6	1,7	1,65	0,005	1,6	1,9	1,75	0,045
Labo 6	4,9	4,9	4,90	0,000	1,7	1,7	1,70	0,000	1,6	1,7	1,65	0,005
Labo 7	3,5	3,7	3,60	0,020	1,6	1,7	1,65	0,005	1,0	1,0	1,00	0,000
Labo 8	3,8	4,4	4,10	0,180	1,7	1,8	1,75	0,005	2,0	1,9	1,95	0,005
Labo 9	4,8	4,7	4,75	0,005	1,8	1,8	1,80	0,000	1,7	1,8	1,75	0,005
Labo 10	5,0	5,4	5,20	0,080	1,7	1,6	1,65	0,005	1,8	1,7	1,75	0,005
Labo 11					1,7	1,0	1,35	0,245	1,8		2,10	
Labo 12					1,7		1,70		1,5	1,6	1,55	0,005
Test de Cochran 1	$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$	
	g =				g =				g =			
	C5% =		C1% =		C5% =		C1% =		C5% =		C1% =	
	Comparaison et Conclusion				Comparaison et Conclusion				Comparaison et Conclusion			
Test de Cochran 2 (éventuellement)	$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$	
	g =				g =				g =			
	C5% =		C1% =		C5% =		C1% =		C5% =		C1% =	
	Comparaison et Conclusion				Comparaison et Conclusion				Comparaison et Conclusion			
Test de Cochran 3 (éventuellement)	$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$		$s_{\max}^2 =$		$\sum s_i^2 =$	
	g =				g =				g =			
	C5% =		C1% =		C5% =		C1% =		C5% =		C1% =	
	Comparaison et Conclusion				Comparaison et Conclusion				Comparaison et Conclusion			

Récapitulation des échantillons éliminés pour l'étude de la répétabilité :

FICHE n°5

Numéro du candidat :

Temps de développement (min)	Farine A				Farine B				Farine C			
	mesure 1	mesure 2	moyenne	variance estimée	mesure 1	mesure 2	moyenne	variance estimée	mesure 1	mesure 2	moyenne	variance estimée
Labo 1	3,8	4,5	4,15	0,245	1,7	1,6	1,65	0,005	1,8	1,7	1,75	0,005
Labo 2	5,1	4,8	4,95	0,045	2,3	1,8	2,05	0,125	1,8	2,1	1,95	0,045
Labo 3	3,7	4,0			1,8	1,6	1,70	0,020	1,9	1,7	1,80	0,020
Labo 4	3,8	3,9	3,85	0,005	1,6	1,8	1,70	0,020	1,8	2,0	1,90	0,020
Labo 5	4,4	4,6	4,50	0,020	1,6	1,7	1,65	0,005	1,6	1,9	1,75	0,045
Labo 6	4,9	4,9	4,90	0,000	1,7	1,7	1,70	0,000	1,6	1,7	1,65	0,005
Labo 7	3,5	3,7	3,60	0,020	1,6	1,7	1,65	0,005	1,0	1,0	1,00	0,000
Labo 8	3,8	4,4	4,10	0,180	1,7	1,8	1,75	0,005	2,0	1,9	1,95	0,005
Labo 9	4,8	4,7	4,75	0,005	1,8	1,8	1,80	0,000	1,7	1,8	1,75	0,005
Labo 10	5,0	5,4	5,20	0,080	1,7	1,6	1,65	0,005	1,8	1,7	1,75	0,005
Labo 11					1,7	1,0	1,35	0,245	1,8		2,10	
Labo 12					1,7		1,70		1,5	1,6	1,55	0,005
Test de Dixon 1	r ₁ =				r ₁ =				r ₁ =			
	r _n =				r _n =				r _n =			
	D5% =		D1% =		D5% =		D1% =		D5% =		D1% =	
	Comparaison et Conclusion				Comparaison et Conclusion				Comparaison et Conclusion			
Test de Dixon 2 (éventuellement)	r ₁ =				r ₁ =				r ₁ =			
	r _n =				r _n =				r _n =			
	D5%		D1%		D5% =		D1% =		D5% =		D1% =	
	Comparaison et Conclusion				Comparaison et Conclusion				Comparaison et Conclusion			

Récapitulation des échantillons éliminés pour l'étude de la reproductibilité :

ANNEXE

Table du test de Cochran

Nombre d'échantillons	Nombre de résultats par échantillon									
	2		3		4		5		6	
	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %
2	-	-	0,975	0,995	0,939	0,979	0,906	0,959	0,877	0,937
3	0,967	0,993	0,871	0,942	0,798	0,883	0,746	0,834	0,707	0,793
4	0,906	0,968	0,768	0,864	0,684	0,781	0,629	0,721	0,590	0,676
5	0,841	0,928	0,684	0,788	0,598	0,696	0,544	0,633	0,506	0,588
6	0,781	0,883	0,616	0,722	0,532	0,626	0,480	0,564	0,445	0,520
7	0,727	0,838	0,561	0,664	0,480	0,568	0,431	0,508	0,397	0,466
8	0,680	0,794	0,516	0,615	0,438	0,521	0,391	0,463	0,360	0,423
9	0,638	0,754	0,478	0,573	0,403	0,481	0,358	0,425	0,329	0,387
10	0,602	0,718	0,445	0,536	0,373	0,447	0,331	0,393	0,303	0,357
11	0,570	0,684	0,417	0,504	0,348	0,418	0,308	0,366	0,281	0,332
12	0,541	0,653	0,392	0,475	0,326	0,392	0,288	0,343	0,262	0,310
13	0,515	0,624	0,371	0,450	0,307	0,369	0,271	0,322	0,243	0,291
14	0,492	0,599	0,352	0,427	0,291	0,349	0,255	0,304	0,232	0,274
15	0,471	0,575	0,335	0,407	0,276	0,332	0,242	0,288	0,220	0,259
16	0,452	0,553	0,319	0,388	0,262	0,316	0,230	0,274	0,208	0,246
17	0,434	0,532	0,305	0,372	0,250	0,301	0,219	0,261	0,198	0,234
18	0,418	0,514	0,293	0,356	0,240	0,288	0,209	0,249	0,189	0,223
19	0,403	0,496	0,281	0,343	0,230	0,276	0,200	0,238	0,181	0,214
20	0,389	0,480	0,270	0,330	0,220	0,265	0,192	0,229	0,174	0,205

Table du test de Dixon

Les n résultats sont classés dans l'ordre croissant : $y_1 \leq y_2 \leq y_3 \dots \dots \leq y_{n-1} \leq y_n$

$$\text{Pour } r_1 = \frac{y_2 - y_1}{y_n - y_1} \text{ ou } r_n = \frac{y_n - y_{n-1}}{y_n - y_1}$$

n	5 %	1 %
3	0,941	0,988
4	0,765	0,889
5	0,642	0,780
6	0,560	0,698
7	0,507	0,637

$$\text{Pour } r_1 = \frac{y_2 - y_1}{y_{n-1} - y_1} \text{ ou } r_n = \frac{y_n - y_{n-1}}{y_n - y_2}$$

n	5 %	1 %
8	0,554	0,683
9	0,512	0,635
10	0,477	0,597

$$\text{Pour } r_1 = \frac{y_3 - y_1}{y_{n-1} - y_1} \text{ ou } r_n = \frac{y_n - y_{n-2}}{y_n - y_2}$$

n	5 %	1 %
11	0,576	0,679
12	0,546	0,642
13	0,521	0,615

$$\text{Pour } r_1 = \frac{y_3 - y_1}{y_{n-2} - y_1} \text{ ou } r_n = \frac{y_n - y_{n-2}}{y_n - y_3}$$

n	5 %	1 %
14	0,546	0,641
15	0,525	0,616
16	0,507	0,595
17	0,490	0,577
18	0,475	0,561
19	0,462	0,547
20	0,450	0,535