

## Modèle Linéaire Général Examen

Durée : 2h.

Les slides des cours et TD et les notes personnelles manuscrites sont autorisées.

Calculatrice autorisée.

### Exercice 1

Dans le but d'étudier l'impact des conditions de stockage de bananes sur leurs mûrissements, on a organisé l'expérience suivante : 48 bananes, manipulées par trois personnes différentes, ont été stockées selon des conditions déterminées d'éclairage et de positions. Le mûrissement des bananes a été quantifié par le % de taches noires observé au bout d'un temps donné (variable PCTNOIR). Le listing suivant donne les résultats du traitement statistique des résultats de l'expérience.

Murissement des bananes selon leurs conditions de stockage

Class	Levels	Values		
		I	II	III
Experimentateur	3			
Eclairage	2	1	2	
Pos. stockage	2	1	2	

Number of observations in data set = 48

Dependent variable: PCTNOIR : pourcentage de taches noires

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	x	1514.04166667	x	x	0.1874
Error	x	8061.87500000			
Corrected Total	x	9575.91666667			

R-Square	Coeff Var	Root MSE	y Mean
x	34.89086	13.85458064	39.70833333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Experimentateur	2	1255.79166667	627.89583333	3.27	0.0478
Eclairage	1	80.08333333	80.08333333	0.42	0.5218
Pos. stockage	1	154.08333333	154.08333333	0.80	0.3754 (**)
Eclairage*Pos. stockage	1	24.08333333	24.08333333	0.13	0.725

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Experimentateur	2	1255.79166667	627.89583333	3.27	0.0478
Eclairage	1	80.08333333	80.08333333	0.42	0.5218
Pos. stockage	1	154.08333333	154.08333333	0.80	0.3754
Eclairage*Po. stockage	1	24.08333333	24.08333333	0.13	0.7250 (*)

1. Quelle a été l'approche statistique utilisée ? Ecrire **de façon claire, précise et complète** le modèle statistique correspondant au listing fourni.
2. (a) Donner les valeurs numériques de *DF Model*, *DF Error*, *DF Corrected Total* et *R-square*.

- (b) Que représente la valeur 0.1874 ? Quelle(s) conclusion(s) tirez-vous de l'interprétation de cette valeur ?
3. Quelle est l'estimation de l'écart-type de la variable PCTNOIR ?
4. (a) Quelles hypothèses la ligne repérée par (\*) permet-elle de tester ? Quels modèles sont concernés ? Quelle est la conclusion de ce test ?
- (b) Mêmes questions pour la ligne (\*\*).
5. Les résultats  $Pr > F = 0.1874$  et  $Pr > F = 0.0478$  vous semblent-ils en contradiction ? Pourquoi ? Si vous pensez qu'il y a contradiction, comment pourriez-vous l'expliquer ?
6. Que représente *R-square* ? Quel(s) commentaire(s) vous inspire la valeur ici obtenue ?
7. (a) Compte tenu des résultats du listing, quelle démarche générale proposeriez-vous pour poursuivre l'interprétation statistique des données (en restant dans le cadre du modèle linéaire) ?
- (b) Compte tenu de la nature de la variable réponse, l'approche utilisée ici vous semble-t-elle *a priori* bien adaptée et pourquoi ?

**Exercice 2 :**

On s'intéresse aux poids d'oisillons dont les mères ont été nourries selon différents régimes caractérisés par les valeurs prises par deux variables quantitatives  $R$  et  $S$ . Ces régimes ont été donnés à cinq moments différents de la journée, moments représentés par une variable  $T$  à cinq modalités (T1, T2, T3, T4, T5). On cherche à expliquer le poids des oisillons au travers des co-variables disponibles et pour cela une approche par modèle linéaire est utilisée. Les sorties obtenues sont listées ci-dessous :

The GLM Procedure  
Informations sur le niveau de classe

Classe	Niveaux	Valeurs				
		T1	T2	T3	T4	T5
T	5					
Number of Observations Read				42		
Number of Observations Used				42		

The GLM Procedure

Dependent Variable: poids

Source	DDL	Somme des carres	Moyenne quadratique	Valeur F	Pr > F
Model	14	562.9964855	40.2140347	9.14	<.0001
Error	27	118.8092288	4.4003418		
Corrected Total	41	681.8057143			

Source	DDL	Coef de		Moyenne		Pr > F
		R-carre	Var	Racine MSE	poids Moyen	
		0.825743	9.461272	2.097699	22.17143	
Source	DDL	Type I SS	Moyenne quadratique	Valeur F	Pr > F	
T	x	184.3020833	x	10.47	<.0001	
R*T	x	349.2068690	x	15.87	<.0001	
S*T	x	29.4875331	x	1.34	0.2776	

Source	DDL	Type III SS	Moyenne quadratique	Valeur F	Pr > F
T	x	32.79967981	x	1.86	0.1459
R*T	x	18.97948351	x	0.86	0.5187
S*T	x	29.48753313	x	1.34	0.2776

1. Quelle approche de modélisation linéaire est considérée ici ?
2. Donner la forme générale de la matrice de design du modèle considéré.
3. Retrouver les valeurs des DDL remplacées par des x dans les tableaux précédents.
4. En supposant les postulats du modèle linéaire vérifiés, donner une interprétation du listing obtenu ;
5. Vous semble-t-il qu'il y ait des contradictions dans les valeurs obtenues pour les p-values ?? si oui, comment pourraient-elles s'expliquer ?
6. Quelle(s) approche(s) proposeriez vous pour continuer l'analyse statistique ?

---

### Exercice 3 :

On considère une expérience dans laquelle on voudrait expliquer une variable réponse  $Y$  en tenant compte des co-variables suivantes :

- $X$  : poids en kg de l'individu ; 3 catégories de poids sont considérées :  $\leq 60$  ,  $]60, 85]$  ;  $> 85$  ;
  - $Z$  : sexe de l'individu, de modalités  $H$  et  $F$  ;
  - $A$  : âge de l'individu (en années) ;
  - $H$  : catégorie zone d'habitation de l'individu ; 3 catégories sont considérées : 1 : urbain, 2 : semi-urbain, 3 : campagne.
1. Proposez un modèle linéaire **sans contrainte sur les paramètres** permettant d'expliquer, sur la base d'un échantillon de taille  $n$ , la variable réponse  $Y$  en tenant compte de ces 4 variables et d'une possible interaction entre  $X$  et  $H$ .
  2. Proposez ensuite une approche permettant alors de tester si l'interaction entre  $X$  et  $H$  est significative.
  3. Le recueil de données laisse apparaître un relevé de la variable  $A$  qui est différent de celui initialement prévu : 3 catégories d'âge ont en fait été considérées ( $A \leq 30$ ,  $30 < A \leq 60$ ,  $A > 60$ ) ; donner le modèle d'ANOVA qui pourrait alors être proposé pour expliquer la variable  $Y$ , toujours avec prise en compte d'une éventuelle interaction entre  $X$  et  $H$  (*donner l'écriture du modèle à considérer*).