



# Examen de Statistique Générale

Noircir 0, 1 ou plusieurs cases avec un crayon bleu ou noir  
2 points par question ; 0pt si 1 mauvaise réponse (case  
cochée par erreur ou case non cochée)

Nom et prénom :  
.....

## Exercice 1 (questions 1 à 3) : questions de cours.

**Question 1 ♣** Quelles problématiques, parmi les suivantes, peuvent être traitées par un modèle linéaire (régression multiple, analyse de variance ou analyse de covariance) ?

- On veut prédire la densité de cloportes en fonction de l'épaisseur de la litière (en mm), du pH et de l'hygrométrie du sol
- On veut étudier l'effet de l'âge, du sexe, du niveau d'étude et du nombre d'heures passées sur les réseaux sociaux sur la pratique ou non d'un sport
- On veut étudier la teneur en eau d'un gâteau en fonction de la température et de la durée de cuisson
- On veut étudier l'effet de la consommation d'alcool, de l'âge, du tabagisme sur l'apparition d'un cancer de la gorge
- On veut prédire la catégorie socio-professionnelle d'individus en fonction de leur lieu d'habitation et de leur revenu

**Question 2 ♣** En régression,

- le meilleur modèle prédictif est celui pour lequel la probabilité critique associée au test du  $R^2$  est la plus petite
- le meilleur modèle prédictif est toujours celui qui a le plus de variables
- le meilleur modèle a au plus 3 variables explicatives
- le modèle nul (tel que  $\forall j, \beta_j = 0$ ) est un bon modèle prédictif
- le meilleur modèle prédictif est celui ayant le  $R^2$  le plus grand

**Question 3 ♣** En analyse de variance,

- le nombre de répétitions pour chaque combinaison de 2 facteurs doit être identique
- on fait l'hypothèse que la variable réponse  $Y$  suit une loi normale
- deux variables explicatives peuvent être identiques
- on fait l'hypothèse que les résidus ont la même variance
- on fait l'hypothèse que les résidus du modèle suivent une loi normale

## Exercice 3 (questions 4 et 5) : construction d'un plan d'expériences.

On construit un plan d'expérience  $2^{6-2}$  et on nomme les facteurs A, B, C, D, E et F. On confond le facteur E avec l'interaction triple ABC et le facteur F avec l'interaction triple ABD.

**Question 4 ♣** Quelles interactions sont confondues avec l'interaction AB :

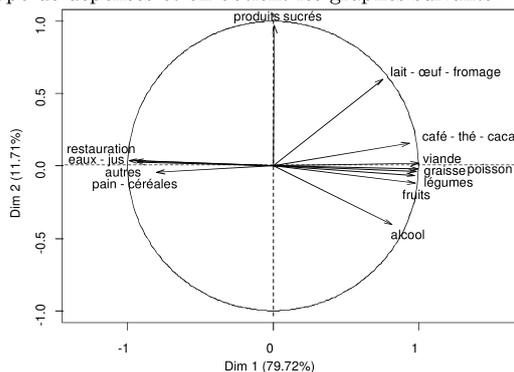
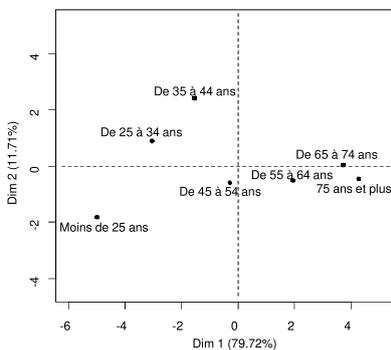
- DF
- A
- CD
- C
- ABCDEF
- CE

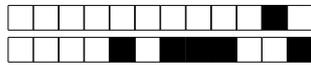
**Question 5 ♣** Les résultats de ce plan  $2^{6-2}$  peuvent être analysés par

- un modèle d'analyse de variance avec les effets principaux et les interactions AB et EF
- un modèle d'analyse de variance avec les effets principaux et au maximum 9 interactions
- un modèle d'analyse de variance avec tous les effets principaux et toutes les interactions
- un modèle d'analyse de variance à 6 facteurs à 2 modalités (avec juste les effets principaux)
- un modèle d'analyse de variance avec les effets principaux et les interactions AB et CE

## Exercice 4 (questions 6 et 7) : ACP.

On étudie la répartition des dépenses alimentaires des Français (exprimées en pourcentage de la dépense alimentaire), grâce à une ACP sur le tableau croisant classes d'âge et type de dépenses et on obtient les graphes suivants :





**Question 6 ♣** On peut affirmer que, dans la part de leur budget,

- les personnes plus âgées dépensent beaucoup au restaurant
- une part importante du budget des moins de 25 ans concerne le pain et l'eau
- les moins de 25 ans achètent beaucoup de produits sucrés
- ceux qui achètent beaucoup de fruits achètent beaucoup de viande
- ceux qui achètent beaucoup de produits sucrés achètent beaucoup de viande

**Question 7 ♣** À partir des graphes d'ACP, on peut affirmer que

- les moins de 25 ans dépensent peu car la coordonnée est faible sur l'axe 1
- le 1er axe résume 79 % de l'information du tableau de données
- les 45-54 ans sont les plus représentatifs de la répartition des dépenses alimentaires d'un français moyen
- les 55 ans et plus ont à peu près le même profil de dépenses
- les classes d'âge correspondent aux individus dans le tableau de données

**Exercice 5 (questions 8 à 10) : interprétation de résultats de modélisation.**

On s'intéresse à l'évolution morphologique des insectes et on veut comparer le poids sec d'une même espèce de mouche, en altitude puis au niveau du littoral. Pour cela on a recueilli le poids sec de 320 mouches avec 80 mouches mâles en altitude, 80 mouches mâles au niveau littoral, 80 femelle en altitude et 80 au niveau littoral. On les a pesé et mesuré leur longueur et largeur.

On construit un modèle avec comme variable réponse le poids sec et comme variables explicatives toutes les autres variables. On met également toutes les interactions possibles entre les variables qualitatives et les variables quantitatives. Les résultats de la régression sans sélection, puis ceux avec sélection AIC sont fournis (colonne de gauche), et les résultats des tests  $F$  et des tests  $T$  du modèle avec sélection sont fournis (colonne de droite).

```
> LinearModel(Poids.Sec ~ sexe+habitat+Longueur+Largeur+sexe:habitat+sexe:Longueur+habitat:Longueur+sexe:Largeur+habitat:Largeur, data=mouche, selection = "aic")
```

Results for the complete model:

=====

```
Call: LinearModel(formula = Poids.Sec ~ sexe + habitat +
Longueur +
Largeur + sexe:habitat + sexe:Longueur + habitat:Longueur +
sexe:Largeur + habitat:Largeur, data = mouche, selection =
"aic")
```

	\$Ftest	SS	df	MS	F value	Pr(>F)
sexe	2.040	1	2.040	6.724	0.0099550	
habitat	45.060	1	45.060	148.500	< 2.2e-16	
Longueur	4.585	1	4.585	15.110	0.0001237	
Largeur	2.319	1	2.319	7.640	0.0060430	
Residuals	95.590	315	0.304			

Residual standard error: 0.5543 on 310 degrees of freedom

Multiple R-squared: 0.5722

F-statistic: 46.08 on 9 and 310 DF, p-value: 4.151e-52

AIC = -367.8 BIC = -330.1

\$Ttest

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.7400	0.2980	-9.17	6.35e-18
sexe - Femelle	0.0806	0.0311	2.59	9.96e-03
sexe - Male	-0.0806	0.0311	-2.59	9.96e-03
habitat - Altitude	-0.4190	0.0344	-12.20	3.02e-28
habitat - Littoral	0.4190	0.0344	12.20	3.02e-28
Longueur	1.7300	0.4450	3.89	1.24e-04
Largeur	1.5500	0.5610	2.76	6.04e-03

Results for the model selected by AIC criterion:

=====

```
Call: LinearModel(formula = Poids.Sec ~ sexe + habitat +
Longueur +
Largeur, data = mouche, selection = "aic")
```

Residual standard error: 0.5509 on 315 degrees of freedom

Multiple R-squared: 0.5707

F-statistic: 104.7 on 4 and 315 DF, p-value: 1.335e-56

AIC = -376.6 BIC = -357.8

**Question 8 ♣** A partir des résultats du modèle complet et du modèle sélectionné, on peut dire que le modèle sélectionné :

- explique 57.07 % de la variabilité des poids
- ne contient aucune interaction
- contient des variables quantitatives et qualitatives
- a un  $R^2$  de 0.5707
- ayant un plus petit  $R^2$ , on garde le modèle complet

**Question 9 ♣** D'après les résultats, on peut dire :

- Il n'y a pas d'effet de la longueur sur le poids
- L'effet de l'habitat sur le poids est le même quel que soit le sexe
- Il n'y a pas d'effet du sexe sur le poids
- Il y a un effet de la largeur sur le poids
- Il y a un effet de l'habitat sur le poids
- L'effet de la longueur sur le poids est le même quel que soit le sexe

**Question 10 ♣** D'après les résultats, on peut dire qu'en moyenne,

- les mâles sont plus lourdes que les femelles
- on trouve plus de mâles en altitude
- les mouches plus larges sont plus lourdes
- l'écart de poids entre mâles et femelles est le même en altitude et au niveau du littoral
- les mouches sont plus légères en altitude