1ère SESSION DU 1er SEMESTRE 2008-2009

Mention / Spécialité / Option : MO – MPME

Année : 1

Intitulé de l'épreuve : Analyse de données

Durée de l'épreuve : 2 heures

Documents autorisés : Résumé manuscrit de 4 pages A4

Matériels autorisés : Calculatrice non programmable

SUJET

Il sera tenu compte de la qualité de la rédaction.
Les points sont donnés à titre indicatif.

Problème 1 [5 points]

Lors de la fabrication d'un lot de fromages de chèvre, on a relevé la masse des fromages fabriqués :

- poids entre 70 et 85 g : 8 fromages;
- poids entre 85 et 90 g : 9 fromages;
- poids entre 90 et 95 g : 14 fromages;
- poids entre 95 et 100 g : 18 fromages;
- poids entre 100 et 105 g : 25 fromages;
- poids entre 105 et 110 g : 16 fromages;
- poids entre 110 et 120 g : 10 fromages;

1. Représenter les données sous forme de tableau.
2. Représenter les données sous forme d'histogramme.
3. Tracer sur un graphique séparé la courbe des fréquences cumulées.
4. Déterminer la médiane de la distribution.
5. Calculer la moyenne, de la variance et de l'écart-type du poids.

Problème 2 [5 points]

Une grave maladie affecte le cheptel bovin d'un certain pays. On estime que 7% des bovins sont atteints. On vient de mettre au point un test pour diagnostiquer la maladie, et on a établi que

- quand un animal est malade, le test est positif dans 87% des cas;
- quand un animal n'est pas malade, le test est négatif dans 98% des cas;

On note M l'événement "être malade" et P l'événement "avoir un test positif".

1. Calculer la probabilité des trois événements "M et P", "non M et non P", et "M et non P".
2. Calculer la probabilité que le test soit négatif.
3. Calculer la probabilité pour qu'un animal ayant un test négatif soit malade (avec une précision de 0.1 %).
4. Les événements M et P sont-ils indépendants?
5. Si l'on effectue un test d'indépendance sur un échantillon de 1000 bovins, distribués selon les probabilités ci-dessus, à l'aide du logiciel `R`, va-t-il retourner une valeur de `p-value` grande ou petite?

Problème 3 [5 points]

On lance deux dés équilibrés, de manière indépendante.
On considère deux variables aléatoires:

- X est égale au plus grand des deux nombres de points indiqués par les dés;
- Y est égale au nombre de dés indiquant un multiple de 3.

Déterminer

1. la loi conjointe de X et Y,
2. les lois (marginales) de X et Y, leur espérance et leur variance,
3. la covariance de X et Y,
4. la variance de $X + Y$,
5. le coefficient de corrélation de X et Y.

Problème 4 [5 points]

Le crâne d'un animal préhistorique appartenant à la famille des canidés a été découvert il y a quelques années, dans la région de Jussac (Auvergne). L'une des questions que se posaient les paléontologues était de savoir si cet animal se rapprochait plus d'un chien ou d'un loup.

On a mesuré 6 grandeurs caractéristiques sur des crânes de chiens de même taille que celle de l'animal inconnu, et sur des crânes de loups. Ces variables sont

`LCB`	longueur condylo-basale
`LMS`	longueur de la mâchoire supérieure
`LBM`	largeur bi-maxilaire
`LP`	longueur de la carnassière supérieure
`LM`	longueur de la première molaire supérieure
`LAM`	largeur de la première molaire supérieure

On a créé dans le logiciel `R` un tableau `Jussac` contenant les données. Ici la lettre C désigne les mesures effectuées sur des chiens, la lettre L celles effectuées sur des loups, et l'étoile celles effectuées sur l'animal inconnu.

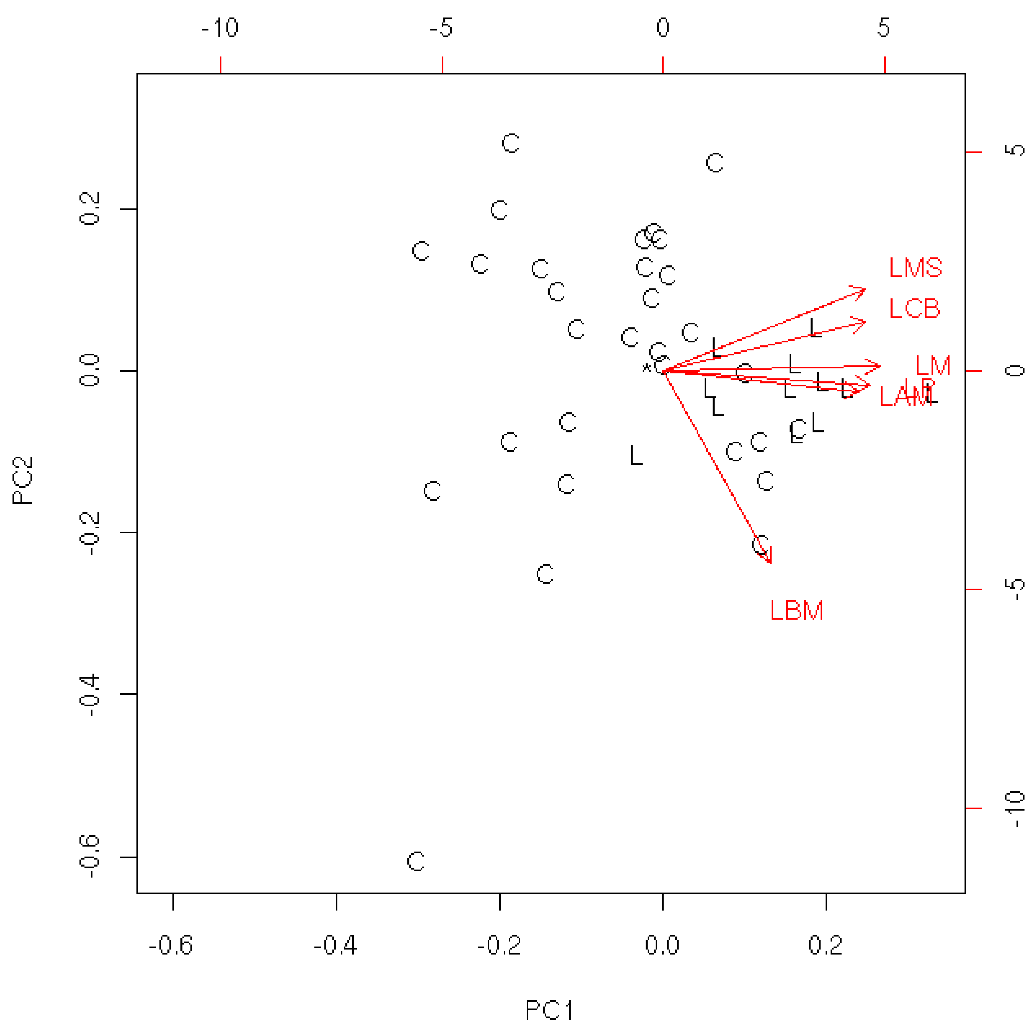
```
> Jussac
  LCB LMS LBM   LP   LM  LAM
C 129  64  95 17.5 11.2 13.8
C 154  74  76 20.0 14.2 16.5
C 170  87  71 17.9 12.3 15.9
C 188  94  73 19.5 13.3 14.8
C 161  81  55 17.1 12.1 13.0
C 164  90  58 17.5 12.7 14.7
C 203 109  65 20.7 14.0 16.8
C 178  97  57 17.3 12.8 14.3
C 212 114  65 20.5 14.3 15.5
C 221 123  62 21.2 15.2 17.0
C 183  97  52 19.3 12.9 13.5
C 212 112  65 19.7 14.2 16.0
C 220 117  70 19.8 14.3 15.6
C 216 113  72 20.5 14.4 17.7
C 216 112  75 19.6 14.0 16.4
C 205 110  68 20.8 14.1 16.4
C 228 122  78 22.5 14.2 17.8
C 218 112  65 20.3 13.9 17.0
C 190  93  78 19.7 13.2 14.0
C 212 111  73 20.5 13.7 16.6
C 201 105  70 19.8 14.3 15.9
C 196 106  67 18.5 12.6 14.2
C 158  71  71 16.7 12.5 13.3
C 255 126  86 21.4 15.0 18.0
C 234 113  83 21.3 14.8 17.0
C 205 105  70 19.0 12.4 14.9
C 186  97  62 19.0 13.2 14.2
C 241 119  87 21.0 14.7 18.3
C 220 111  88 22.5 15.4 18.0
C 242 120  85 19.9 15.3 17.6
L 199 105  73 23.4 15.0 19.1
L 227 117  77 25.0 15.3 18.6
L 228 122  82 24.7 15.0 18.5
L 232 123  83 25.3 16.8 15.5
L 231 121  78 23.5 16.5 19.6
L 215 118  74 25.7 15.7 19.0
L 184 100  69 23.3 15.8 19.7
L 175  94  73 22.2 14.8 17.0
L 239 124  77 25.0 16.8 27.0
L 203 109  70 23.3 15.0 18.7
L 226 118  72 26.0 16.0 19.4
L 226 119  77 26.5 16.8 19.3
* 210 103  72 20.5 14.0 16.7
```

Interpréter les résultats suivants fournis par R. En particulier, quelle réponse donneriez-vous à la question des paléontologues?

```
> cor(Jussac)

          LCB       LMS       LBM        LP        LM       LAM
LCB 1.0000000 0.9587410 0.3481835 0.6129486 0.7179356 0.5872510
LMS 0.9587410 1.0000000 0.2003331 0.6610016 0.7359561 0.5946533
LBM 0.3481835 0.2003331 1.0000000 0.3699619 0.3502798 0.3547771
LP  0.6129486 0.6610016 0.3699619 1.0000000 0.8935121 0.7626431
LM  0.7179356 0.7359561 0.3502798 0.8935121 1.0000000 0.7892164
LAM 0.5872510 0.5946533 0.3547771 0.7626431 0.7892164 1.0000000

> acp<-prcomp(Jussac, scale=TRUE)
> acp
Standard deviations:
[1] 2.0248394 0.9394336 0.7992505 0.5093948 0.3119382 0.1479871
> biplot(acp)
```



La commande suivante représente les composantes principales 1 et 3.

```
> biplot(acp, choices=c(1,3))
```