



1^{ère} SESSION DU 1^{er} SEMESTRE 2009-2010

Mention / Spécialité / Option : M1 SIME/MPME

Année : 2009/2010

Enseignant : N. Berglund / L. Delsol

Intitulé de l'épreuve : Analyse de données

Durée de l'épreuve : 2 heures

Documents autorisés : Résumé manuscrit 4 pages A4

Matériels autorisés : Calculatrice non programmable

page 1 / 5

SUJET

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

Problème 1 [5 points]

La hauteur de peupliers de 10 ans cultivés en peupleraie a été mesurée à partir d'une plantation de 1000 individus. Le tableau suivant présente l'effectif d'arbres en fonction des classes de hauteur (en mètres).

Classe	[14;18[[18;20[[20;22[[22;24[[24;26[[26;30[
Effectif	144	152	200	194	158	152

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, la variance et de l'écart-type de la hauteur des peupliers.

Problème 2 [5 points]

Un crime a été commis dans un village de 1000 habitants. L'habitant X est soupçonné d'avoir commis le crime. Des traces d'ADN retrouvées sur les lieux du crime coïncident avec celui de X.

On suppose que le test d'ADN est fiable à 99%, c'est-à-dire

- la probabilité qu'une personne non coupable présente un ADN coïncidant avec celui trouvé sur le lieu du crime est de 1%;
- la probabilité que le test révèle un ADN non coïncidant pour la personne coupable est de 1%.

En supposant qu'il y a exactement 1 coupable parmi les 1000 habitants, déterminer la probabilité que X soit le coupable (sachant que le test d'ADN s'est révélé positif pour lui).

Problème 3 [5 points]

On lance deux pièces équilibrées, de manière indépendante.

- Si la première pièce tombe sur Pile, alors on gagne 10 Euros (quel que soit le côté sur lequel tombe la deuxième pièce).
- Si la première pièce tombe sur Face et la deuxième tombe sur Pile, alors on gagne 5 Euros.
- Si les deux pièces tombent sur Face, alors on ne gagne rien.

On considère deux variables aléatoires:

- X est égale à la somme gagnée (0, 5 ou 10 Euros);
- Y est égale au nombre de pièces tombées sur Face (0, 1 ou 2).

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]

On cherche à étudier les différentes méthodes de production d'électricité dans différents pays du monde. On restreint notre étude à des pays qui produisent de l'énergie nucléaire. Cette étude a pour objectif de mettre en évidence les pays qui basent principalement leur production d'électricité sur des types particuliers d'énergie. On pourra par la suite envisager d'aider au développement de nouveaux modes de production moins polluants ou plus sécuritaires mais aussi mettre en avant les modèles de production basés sur de énergies renouvelables.

Pour chacun de ces pays, la production d'électricité est décrite au travers de 4 variables:

Hydraulique	proportion d'énergie hydraulique
Nucléaire	proportion d'énergie nucléaire
Fossile	proportion d'énergie fossile
Autres.Renouv	proportion d'autres types d'énergies renouvelables (géothermie, biomasse, marine, solaire, éolienne).

On dispose également d'une cinquième variable IDH correspondant à l'indice de développement humain.

On a créé dans le logiciel R un tableau `tab3` contenant les données.

```
> tab3
      Hydraulique Nucleaire Fossile Autres.Renouv IDH
Canada          57.9      14.5   25.7          2.0 0.966
USA              6.3      19.1   71.7          2.4 0.956
Mexique         10.6       3.8   81.7          4.0 0.854
Bresil          83.9       2.8    9.3          4.0 0.813
Allemagne       4.3      21.9   62.3         10.8 0.947
Belgique        1.9      54.4   40.2          2.7 0.953
Espagne         9.9      17.8   61.7         10.0 0.955
Finlande       17.4      28.8   40.6         12.9 0.959
France         11.2      77.0    9.9          1.5 0.961
Pays-Bas        0.1       4.3   86.9          7.2 0.964
Royaume-Uni     2.3      15.6   76.8          4.1 0.947
Suede          44.7      44.7    3.9          6.1 0.963
Suisse         53.9      41.0    1.5          1.7 0.960
Tchequie        2.9      29.6   66.5          1.0 0.903
Hongrie         0.5      36.7   54.3          8.1 0.879
Russie         18.7      15.9   65.0          0.0 0.817
Ukraine         6.4      48.1   45.5          0.1 0.796
Afrique.du.Sud  2.3       4.3   93.3          0.1 0.683
Chine          14.2       2.0   83.5          0.3 0.772
Japon           7.5      24.2   65.6          2.2 0.960
Inde           15.2       2.6   80.4          1.8 0.612
Nouvelle-Zelande 54.4       0.0   34.2         11.5 0.950
```

Interpréter les résultats suivants fournis par R.

```
> cor(tab3)
      Hydraulique  Nucleaire  Fossile  Autres.Renov  IDH
Hydraulique  1.0000000 -0.1679988 -0.7093665  0.06245836  0.07670006
Nucleaire   -0.16799877  1.0000000 -0.5608144  -0.12107544  0.39431619
Fossile     -0.70936651 -0.5608144  1.0000000  -0.10946411 -0.41674349
Autres.Renov 0.06245836 -0.1210754 -0.1094641  1.00000000  0.44989479
IDH         0.07670006  0.3943162 -0.4167435  0.44989479  1.00000000
> plot(data.frame(tab3))
> acp<-prcomp(tab3,scale=T)
> summary(acp)
Importance of components:
              PC1  PC2  PC3  PC4  PC5
Standard deviation  1.463 1.131 1.103 0.6037 0.01097
Proportion of Variance 0.428 0.256 0.243 0.0729 0.00002
Cumulative Proportion 0.428 0.684 0.927 1.0000 1.00000
> biplot(acp)
> biplot(acp,choices=c(1,3))
```



