



1^{ère} SESSION DU 1^{er} SEMESTRE 2007-2008

Mention / Parcours / Spécialité : **Master 1**

Enseignant : Nils Berglund

Intitulé de l'épreuve : **Analyse de données**

Durée de l'épreuve : 2 heures

Documents autorisés : Résumé manuscrit de 4 pages (4 côtés de feuille)

Matériels autorisés : Calculatrice (non programmable)

page 1 / .4

SUJET

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

Problème 1 [5 points]

En Parcimonie Orientale, une entreprise rémunère ses 70 employés à la semaine, en fonction du type de travail effectué et de leur ancienneté. On dispose des informations suivantes :

- salaire de 50 à 59.99 crédits : 8 employés;
- salaire de 60 à 69.99 crédits : 10 employés;
- salaire de 70 à 79.99 crédits : 16 employés;
- salaire de 80 à 89.99 crédits : 15 employés;
- salaire de 90 à 99.99 crédits : 10 employés;
- salaire de 100 à 119.99 crédits : 8 employés;
- salaire de 120 à 179.99 crédits : 3 employés.

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 des valeurs approchées de la moyenne, de la variance et de l'écart-type du salaire hebdomadaire.

Problème 2 [5 points]

On lance 2 dés équilibrés. Calculer les probabilités que

1. la somme des points obtenus soit égale à 8;
2. la somme des points obtenus soit supérieure ou égale à 10;
3. la somme des points obtenus soit égale à 8, sachant que le premier dé est tombé sur 4;
4. le premier dé soit tombé sur 4, sachant que la somme des points obtenus est égale à 8;
5. le premier dé soit tombé sur 4, sachant que la somme des points obtenus est supérieure ou égale à 10.

Problème 3 [5 points]

Une urne contient 5 billets de 5 dollars, 4 billets de 10 dollars, et 1 billet de 100 dollars. Ces billets sont indiscernables au toucher.

Un joueur tire un premier billet de l'urne. Si ce n'est pas le billet de 100 dollars, alors il tire un second billet de l'urne.

S'il a tiré le billet de 100 dollars, alors le joueur peut garder ce billet uniquement. Sinon, il peut garder tous les billets tirés.

On considère deux variables aléatoires:

- X égale à la somme gagnée (la valeur totale des billets gardés);
- Y égale au nombre total de billets tirés.

Déterminer

- la loi conjointe de X et Y ,
- les lois (marginales) de X et Y , leur espérance et leur variance,
- la covariance de X et Y ,
- la variance de $X + Y$,
- le coefficient de corrélation de X et Y .

Problème 4 [5 points]

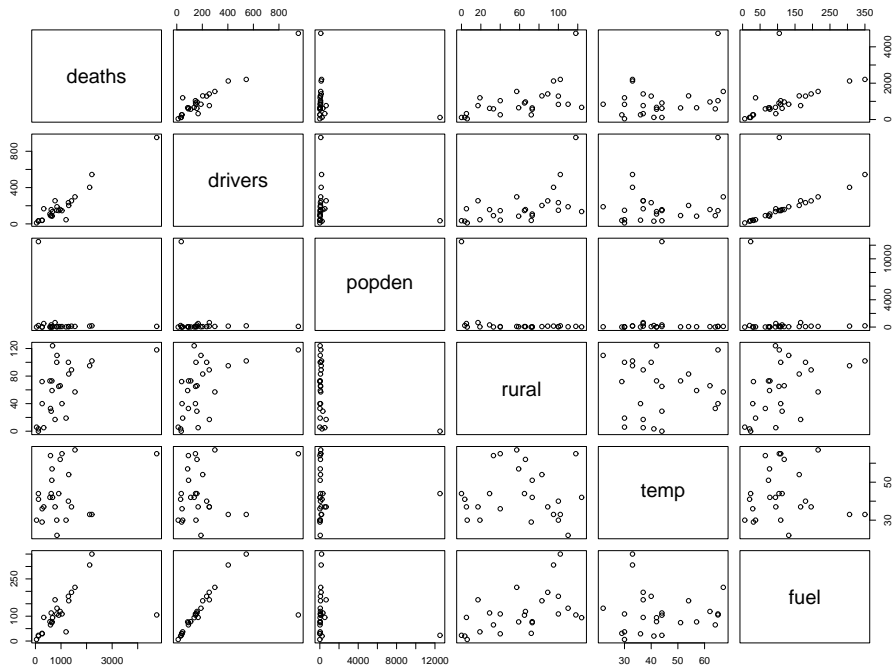
L'échantillon statistique `road` contient les données suivantes pour 26 états américains:

| | |
|----------------------|---|
| <code>state</code> | nom de l'état |
| <code>deaths</code> | nombre annuel de décès dans des accidents de la route |
| <code>drivers</code> | nombre de conducteurs (en dizaines de milliers) |
| <code>popden</code> | densité de population (habitants par mile carré) |
| <code>rural</code> | longueur totale des routes rurales (en milliers de miles) |
| <code>temp</code> | moyenne de la température diurne maximale en janvier (° F) |
| <code>fuel</code> | consommation annuelle de carburant (en 10'000'000 gallons US) |

Une analyse avec le logiciel R livre les résultats suivants :

```
> road
      deaths drivers  popden rural temp  fuel
Alabama   968    158    64.0  66.0   62 119.0
Alaska     43     11     0.4   5.9   30   6.2
Arizona   588     91    12.0  33.0   64  65.0
Arkansas  640     92    34.0  73.0   51  74.0
Calif  4743   952   100.0 118.0   65 105.0
Colo     566   109    17.0  73.0   42  78.0
Conn     325   167   518.0   5.1   37  95.0
Dela     118    30   226.0   3.4   41  20.0
DC        115    35 12524.0   0.0   44  23.0
Florida  1545   298    91.0  57.0   67 216.0
Georgia  1302   203    68.0  83.0   54 162.0
Idaho     262    41     8.1  40.0   36  29.0
Ill       2207  544   180.0 102.0   33 350.0
Ind       1410  254   129.0  89.0   37 196.0
Iowa      833   150    49.0 100.0   30 109.0
Kansas    669   136    27.0 124.0   42  94.0
Kent      911   147    76.0  65.0   44 104.0
Louis    1037   146    72.0  40.0   65 109.0
Maine    1196    46    31.0  19.0   30  37.0
Maryl     616   157   314.0  29.0   44 113.0
Mass      766   255   655.0  17.0   37 166.0
Mich     2120  403   137.0  95.0   33 306.0
Minn      841   189    43.0 110.0   22 132.0
Miss      648    85    46.0  59.0   57  77.0
Mo       1289  234    63.0 100.0   40 180.0
Mont      259    38     4.6  72.0   29  31.0
```

```
> plot(road)
```



```
> cor(road)
```

| | deaths | drivers | popden | rural | temp | fuel |
|---------|----------|----------|------------|-----------|------------|-----------|
| deaths | 1.00000 | 0.95547 | -0.1924320 | 0.562871 | 0.3033855 | 0.520973 |
| drivers | 0.95547 | 1.00000 | -0.1511586 | 0.532013 | 0.2383378 | 0.589521 |
| popden | -0.19243 | -0.15115 | 1.0000000 | -0.347037 | -0.0030833 | -0.209277 |
| rural | 0.56287 | 0.53201 | -0.3470370 | 1.000000 | -0.0197696 | 0.510656 |
| temp | 0.30338 | 0.23833 | -0.0030833 | -0.019769 | 1.0000000 | -0.011212 |
| fuel | 0.52097 | 0.58952 | -0.2092773 | 0.510656 | -0.0112125 | 1.000000 |

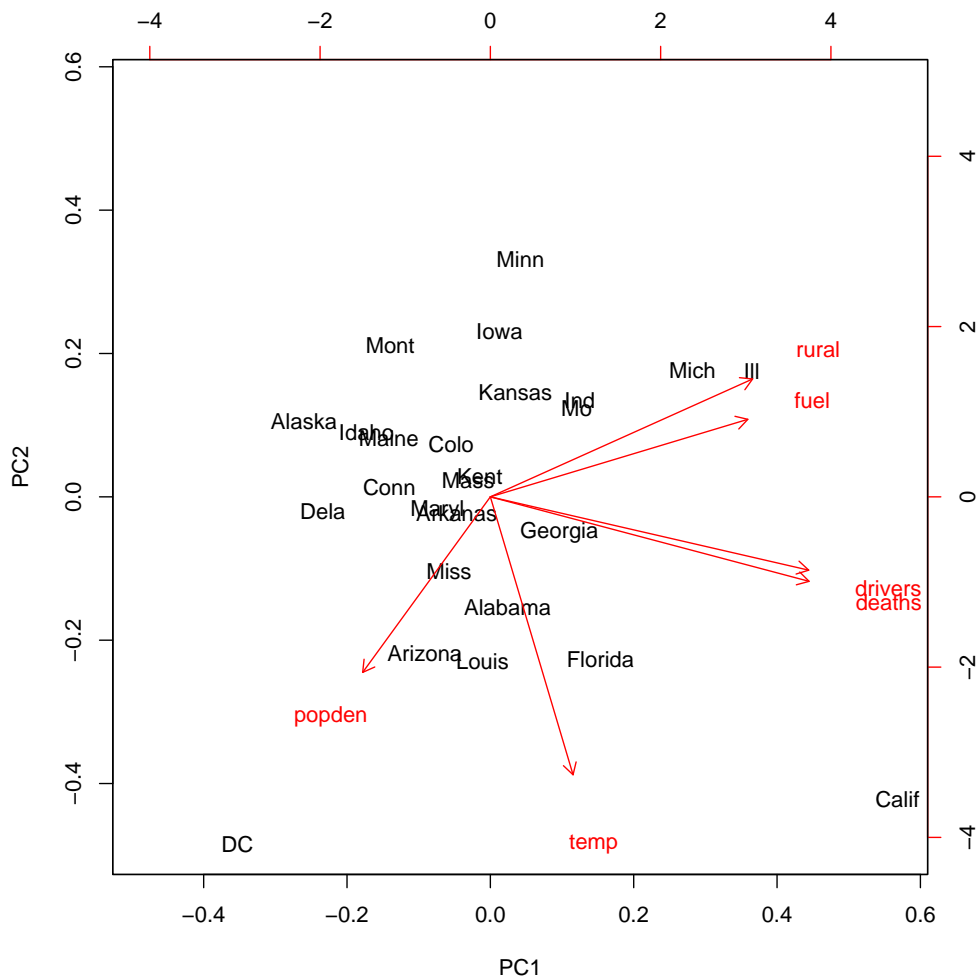
```
> acp<-prcomp(road, scale=TRUE)
```

```
> acp
```

Standard deviations:

```
[1] 1.7303513 1.0792067 0.9417306 0.6958193 0.6589479 0.1896418
```

```
> biplot(acp)
```



Interpréter ces résultats.