Université Claude Bernard - Lyon 1

L3 - MIV **Année 2006-2007**
Bio-Statistique 1 **Session 1**

Mardi 5 juin 2007
Durée : 2 heures

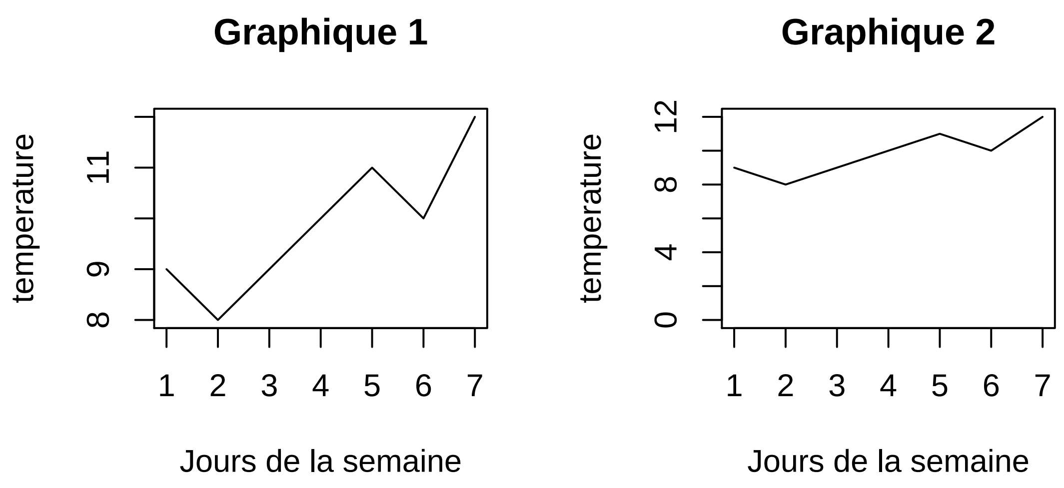
A.B. Dufour

Tous les documents sont autorisés.

1 Graphiques et Origines

Dans un des chapitres de son livre *Statistiques. Méfiez-vous !* (Ellipses, 2007), Nicolas Gauvrit mène une réflexion sur les échelles dans les représentations graphiques. En voici deux exemples.

1) On représente la température en degrés celsius des jours d'une semaine de février.

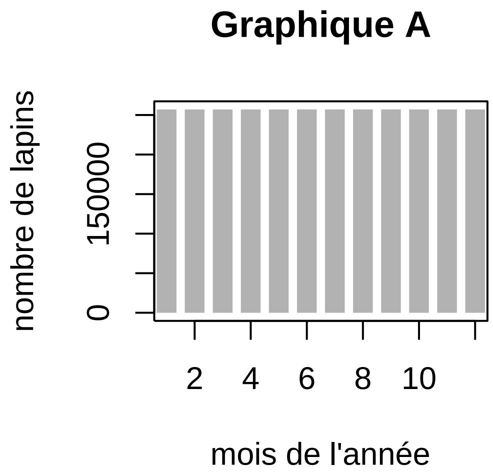


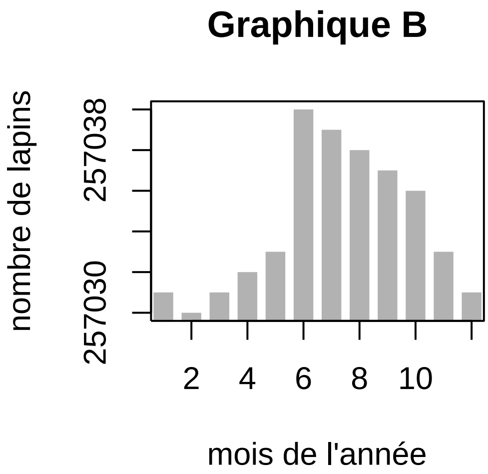
a) Commentez le graphique 1.

b) Commentez le graphique 2.

c) De ces deux graphiques, expliquez celui qui vous parait le mieux refléter les données.

2) On représente l'évolution du nombre de lapins au cours d'une année sur le territoire Australien.





a) Commentez le graphique A.
b) Commentez le graphique B.
c) De ces deux graphiques, expliquez celui qui vous parait le mieux refléter les données.

2 Poissons d'élevage et type de nourriture

L'accroissement de la population Vietnamienne nécessite des changements dans les techniques d'agriculture et d'élevage afin de pourvoir à l'alimentation de tous. Une expérience a été menée dans une ferme piscicole. 60 poissons du même âge ont été pesés, puis répartis en 4 groupes de 15. Chaque groupe a reçu un régime alimentaire différent :

- nourriture classique : temoins
- nourriture à base de 20% de carcasse bovine : MBM20
- nourriture à base de 40% de carcasse bovine : MBM40
- nourriture à base de 80% de carcasse bovine : MBM80.

Plus on introduit de la carcasse bovine dans l'alimentation, plus le coût financier diminue. Mais qu'en est-il de la prise de poids de ces poissons ?

Remarque. On a démontré à l'aide d'un test de Shapiro que la variable poids était gaussienne dans chacun des groupes.

Partie A.

1) Pourquoi est-il important de s'assurer que les groupes sont identiques en poids au début de l'étude ?

2) On donne les résultats aux tests de comparaison de variance et de comparaison de moyenne.

```
        Bartlett test of homogeneity of variances
data:  w0 by reference
Bartlett's K-squared = 2.1943, df = 3, p-value = 0.5331
Analysis of Variance Table
Response: w0
          Df  Sum Sq Mean Sq F value Pr(>F)
reference  3  174.62   58.21  1.7792 0.1616
Residuals 56 1832.02   32.71
```

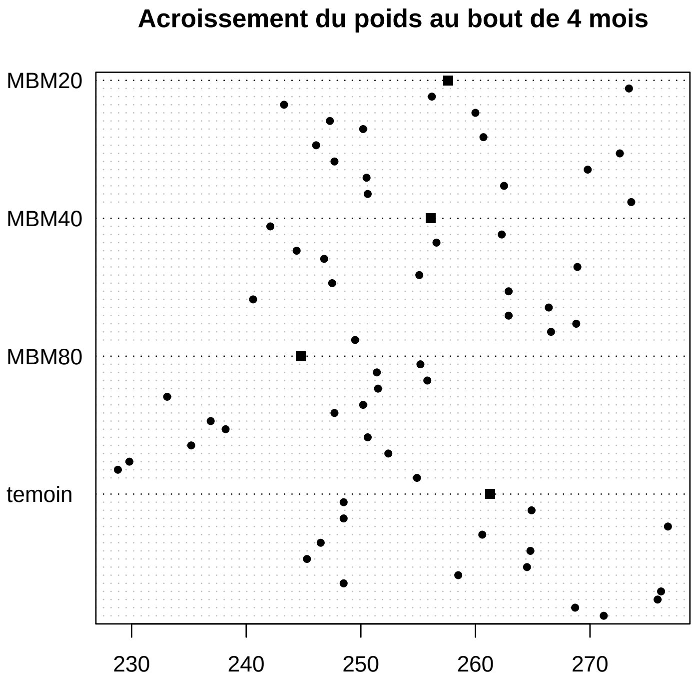
Conclure.

3) Disposez-vous de toutes les informations nécessaires à la bonne rédaction d'un document scientifique ? Argumentez votre réponse.

Partie B.

On note, quatre mois après, l'accroissement du poids de ces poissons.

1) Pourquoi choisit-on l'accroissement plus que le poids final ?
2) Commentez le graphe de Cleveland ci-dessous (les carrés désignant les moyennes par groupe).



3) Des tests de Shapiro ont montré que certaines distributions n'étaient pas gaussiennes. On choisit donc de réaliser simplement un test non paramétrique. Quel est-il ?
4) On obtient une statistique du test de 11.1638 et une probabilité critique de 0.0109. Donnez une conclusion à cette étude.

3 Autour de la loi de Poisson

1) Soit X une loi de Poisson de paramètre λ.

On pose $\lambda = 2$ et on donne les informations suivantes :

```
dpois(0:10, 2)
[1] 1.353353e-01 2.706706e-01 2.706706e-01 1.804470e-01 9.022352e-02 3.608941e-02
[7] 1.202980e-02 3.437087e-03 8.592716e-04 1.909493e-04 3.818985e-05
ppois(0:10, 2)
[1] 0.1353353 0.4060058 0.6766764 0.8571235 0.9473470 0.9834364 0.9954662 0.9989033
[9] 0.9997626 0.9999535 0.9999917
```

a) $P(X = 3) =$

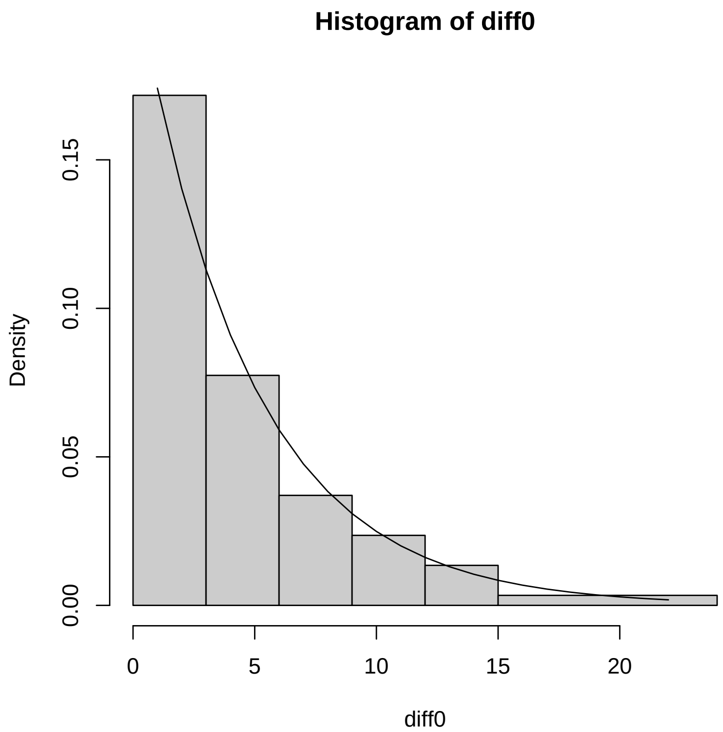
b) $P(X < 2) =$

c) $P(X > 10) =$

2) Soit Y une loi de Poisson de paramètre μ. X et Y sont indépendantes.
   a) Que vaut $P(X = x \cup Y = y)$?
   b) Donnez l'espérance et la variance de la loi de $Z = X + Y$.
   c) Le paramètre de Y est $\mu = 1$. Que vaut la variance de Z ?

4 Tremblements de terre

Dans une zone fortement sismique, on enregistre les tremblements de terre. L'histogramme ci-dessous montre la répartition des temps d'attente entre deux tremblements de terre ainsi que la distribution théorique généralement associée à ces phénomènes (facilement identifiable).



1) De quelle loi s'agit-il ?
2) Donnez, par la méthode du maximum de vraisemblance, une estimation du paramètre α de cette loi théorique T (écrire la démonstration).
3) On donne le résumé statistique suivant. Donnez l'expression numérique de α.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   3.000   4.626   6.500  22.000
```

4) On réalise un test d'ajustement du temps d'attente entre deux tremblements de terre à cette loi T :

```
        Chi-squared test for given probabilities
data:  hist1$counts
X-squared = 0.9663, df = 5, p-value = 0.9652
```

ainsi que les probabilités suivantes :

```
 1 - pchisq(reschisq$statistic, 1:7)
[1] 0.3255966 0.6168274 0.8093978 0.9148574 0.9652354 0.9868564 0.9953535
```

En vous aidant des informations ci-dessus, que pouvez-vous conclure de cet ajustement ?

5 Corrélations et Jury

Deux enseignants mc1 et mc2 constituent un jury et notent dix étudiants sur une prestation orale.

```
mc1 <- c(11, 11.5, 16, 20, 15, 15.5, 12.5, 12, 19, 13)
mc2 <- c(12.5, 12, 19, 18, 16.5, 13.5, 13, 10, 17, 16)
```

1) On regarde la corrélation linéaire simple entre les notations des deux enseignants.
   a) Quelle formulation littérale donneriez-vous à l'hypothèse testée dans ce cas-là ?
   b) A l'aide des informations ci-dessous, que peut-on dire de cette corrélation ?

```
cor.test(mc1, mc2)
        Pearson's product-moment correlation
data:  mc1 and mc2
t = 3.3976, df = 8, p-value = 0.009393
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2691753 0.9422308
sample estimates:
      cor
0.7685482
```

2) On s'intéresse maintenant aux classements des étudiants les uns par rapport aux autres.

```
rank(mc1)
[1]  1  2  8 10  6  7  4  3  9  5
rank(mc2)
[1]  3  2 10  9  7  5  4  1  8  6
```

   Soit la variable aléatoire $G = \sum_{i=1}^{n} iX_i$. i représente le classement des étudiants par un enseignant. Ce dernier est fixé. X_i représente le classement du second enseignant. C'est une variable de rang prenant également ces valeurs de 1 à n.
   a) Calculez son espérance et sa variance.
   b) Que peut nous apprendre le théorème central limite sur la variable G ?
3) On réalise le test de corrélation de rangs approprié.
   a) De quel test s'agit-il ?
   b) Quelle formulation littérale donneriez-vous à l'hypothèse testée dans ce cas ?
   c) La statistique du test est 20 et la probabilité critique associée 0.0017. Que peut-on dire de cette corrélation de rang ?