Université Claude Bernard - Lyon 1

L3 - MIV — **Année 2005-2006**

Bio-Statistique 1 — **Session 1**

Vendredi 2 juin 2006
Durée : 2 heures

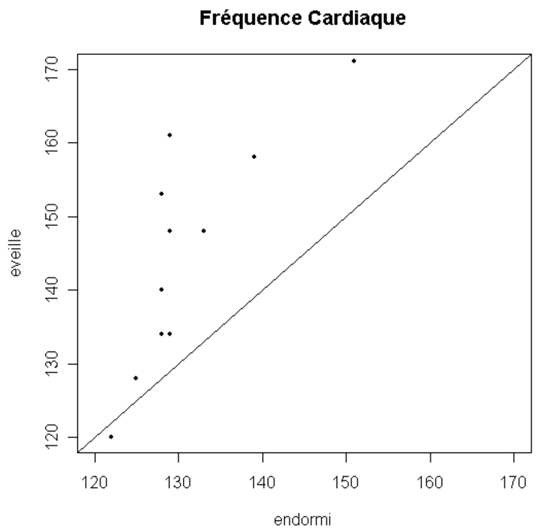
A.B. Dufour

Tous les documents sont autorisés.

1 Bébés et Fréquences cardiaques

On a mesuré la fréquence cardiaque (nombre de pulsations par minute) de 11 bébés âgés de 2 à 8 semaines à l'état endormi et à l'état éveillé.

```
endormi <- c(133, 128, 129, 139, 122, 125, 129, 129, 151, 128, 128)
eveille <- c(148, 134, 134, 158, 120, 128, 161, 148, 171, 153, 140)
```



1) A quelle hypothèse correspond la droite représentée sur ce graphique ?
2) Commenter la représentation graphique.
3) Quel test faudrait-il proposer pour répondre à l'hypothèse suivante "En moyenne, les fréquences cardiaques sont les mêmes que les bébés soient endormis ou éveillés" ?
4) Pensez-vous que la réalisation de ce test soit nécessaire ? Pourquoi ?

2 Ailes des Papillons

Un entomologiste (Brower, 1959) a mesuré la longueur de l'aile antérieure droite (`aile`, en mm) de 10 specimen mâles de papillon de l'espèce *Papilio glaucus* de chacune des régions suivantes : l'Alaska, la Colombie Britannique (`ColomBrit`), le Dakota du Sud et l'Illinois.

1) Donner l'instruction de R qui a permis d'obtenir les moyennes par groupe.

```
Alaska Colombie_Britannique           Dakota_Sud           Illinois
  41.5                 43.0                 44.4               48.7
```

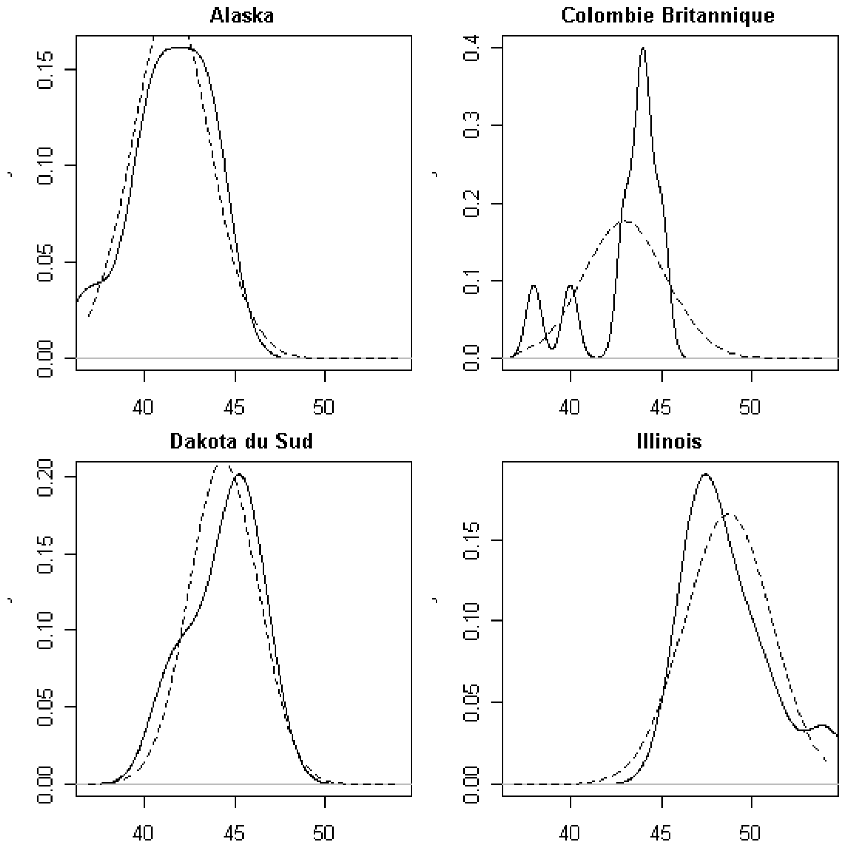
2) On réalise une analyse de la variance à un facteur.

```
Analysis of Variance Table

Response: aile
          Df  Sum Sq Mean Sq F value    Pr(>F)
region     3 288.600  96.200  20.019 8.351e-08
Residuals 36 173.000   4.806
```

Commenter.

3) On a utilisé la fonction `density` pour donner une représentation graphique des longueurs des ailes de papillon en fonction de chacune des quatre régions. Commenter ces graphiques du point de vue de la normalité.



3 Loi Binomiale

1) Dans une loi binomiale où le paramètre $p < \frac{1}{2}$, la distribution est-elle asymétrique à gauche ? à droite ?

2) Soit X une loi binomiale de paramètres $n_1 = 12$ et $p_1 = 0.14$. A l'aide de R, on donne les informations suivantes :

```
dbinom(0:12, 12, 0.14)
```

```
 [1] 1.636746e-01 3.197365e-01 2.862757e-01 1.553434e-01 5.689904e-02 1.482022e-02
 [7] 2.814692e-03 3.927477e-04 3.995980e-05 2.891148e-06 1.411956e-07 4.179151e-09
[13] 5.669391e-11
```

```
pbinom(0:12, 12, 0.14)
```

```
 [1] 0.1636746 0.4834112 0.7696869 0.9250303 0.9819293 0.9967496 0.9995643 0.9999570
 [9] 0.9999970 0.9999999 1.0000000 1.0000000 1.0000000
```

Donner les probabilités :

a) $P(X = 2) =$

b) $P(X = 1) + P(X = 2) =$

c) $P(X < 5) =$

d) $P(X > 8) =$

e) $P(3 < X < 5) =$

3) Soit Y une autre loi binomiale de paramètres $n_2 = 15$ et $p_2 = 0.14$. X et Y sont indépendantes. Quelle est la loi suivie par $X + Y$ (expliciter votre réponse) ?

4 Enquête Lyon 1

Une enquête a été réalisée au mois de mars 2006 au près de l'ensemble des étudiants de l'université Lyon1.

1) Une première étude porte sur la variable `sexe`. 1259 femmes et 1105 hommes ont répondu à l'enquête. On sait par ailleurs que la proportion d'étudiantes à Lyon 1 est de 0.4846 (population).

   a) Que pouvez-vous dire de la proportion de femmes dans l'échantillon ?

```
prop.test(1259, 2364, 0.4846, correct = F)

        1-sample proportions test without continuity correction

data:  1259 out of 2364, null probability 0.4846
X-squared = 21.7818, df = 1, p-value = 3.055e-06
alternative hypothesis: true p is not equal to 0.4846
95 percent confidence interval:
 0.5124226 0.5526155
sample estimates:
        p
0.5325719
```

   b) Dans les résultats donnés par le test, apparaît l'intervalle de confiance de la proportion de femmes dans la population calculée à partir de l'échantillon. La vraie valeur se situe-t-elle dans l'intervalle ? Quel en est le sens statistique ?

2) Une deuxième étude porte sur la variable `nationalite`. 179 étrangers et 2185 français ont répondu à l'enquête. On sait par ailleurs que la proportion d'étudiants étrangers à Lyon 1 est de 0.1197 (population). Que pouvez-vous dire de la proportion d'étrangers dans l'échantillon ?

```
prop.test(179, 2364, 0.1197, correct = F)

        1-sample proportions test without continuity correction

data:  179 out of 2364, null probability 0.1197
X-squared = 43.3961, df = 1, p-value = 4.471e-11
alternative hypothesis: true p is not equal to 0.1197
95 percent confidence interval:
 0.06572967 0.08708523
sample estimates:
         p
0.07571912
```

3) On décide de croiser les deux variables. La table de contingence ainsi que le résultat au test du Chi-Deux sont :

```
tns

      etranger français
femme       77     1182
homme      102     1003
```

```
chisq.test(tns)
```

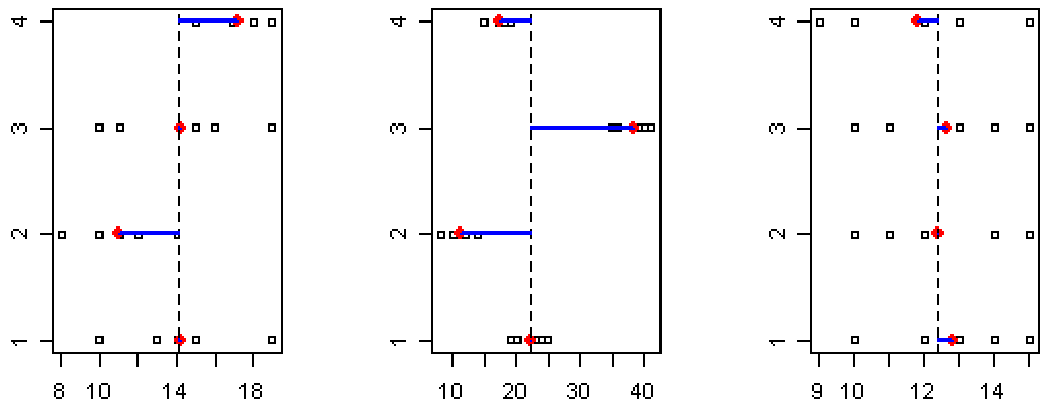
```
        Pearson's Chi-squared test with Yates' continuity correction

data:  tns
X-squared = 7.7192, df = 1, p-value = 0.005464
```

Que pouvez-vous dire de la relation entre le sexe et la nationalité dans cet échantillon d'étudiants de Lyon 1 ?

5 Graphiques

On donne les trois représentations graphiques ci-dessous et les valeurs de η^2 suivantes : 0.0374, 0.4320, 0.9610.
Replacer chaque valeur de η^2 avec le graphique correspondant.



6 Basket et réussite

Un basketteur réalise 41 lancers. On s'intéresse à la variable X : attente de réussite d'un panier. X est une loi géométrique prenant ses valeurs dans $\mathbb{N}*$.
Pour tout $x_i \geq 1$, on a $P(X = x_i) = p(1-p)^{x_i - 1}$ où p est la probabilité de réussir un panier au x_i*ième* lancer. $E(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$.

1) On souhaite estimer la valeur de p par la méthode du maximum de vraisemblance.
   a) Ecrire la fonction du maximum de vraisemblance $L(x_1, x_2, \ldots, x_n, p) =$ puis le *Log* de la vraisemblance.
   b) Obtenir la valeur estimée de p.
2) Les résultats obtenus par notre basketteur sont les suivants :

```
lancer <- rep(c(1, 2, 3, 4, 5), c(18, 3, 4, 0, 1))
table(lancer)
```

```
lancer
 1  2  3  5
18  3  4  1
```

18 fois le ballon est rentré au premier lancer ; 3 fois le ballon est rentré au deuxième lancer, ...

   a) Donner en fraction la moyenne des lancers réussis.
   b) Les probabilités théoriques de la distribution géométrique sont données dans le vecteur `proba`. Comment a-t-on obtenu la dernière probabilité ?

```
[1] 0.63414634 0.23200476 0.08487979 0.03105358 0.01791553
```

   c) On réalise un test d'ajustement du Chi-Deux de la distribution des observations.

```
reschi = chisq.test(c(18, 3, 4, 0, 1), p = proba)
reschi$statistic
```

```
X-squared
 4.539799
```

A l'aide des informations sur le chi-deux ci-dessous, peut-on accepter l'hypothèse que la distribution d'attente de réussite d'un lancer soit distribué selon une loi géométrique ?

```
ddl = 1:6
1 - pchisq(reschi$statistic, ddl)
```

```
[1] 0.03311549 0.10332258 0.20876769 0.33785443 0.47457623 0.60403629
```