

Examen de Bio - Statistique 1

Durée : 2 heures

Documents Autorisés

Dans le package "UsingR", on trouve un jeu de données appelé "babyboom" contenant des informations sur 44 bébés nés dans un hôpital de Brisbane (Etat du Queensland, Australie) sur une période de 24 heures :

- le temps exprimé en minutes (`clock.time`)
- le sexe (`gender` : `boy` ou `girl`)
- le poids de naissance exprimé en grammes (`wt`)
- le temps à partir de minuit, exprimé en minutes (`running.time`)

Question 1 : Naissance de garçons

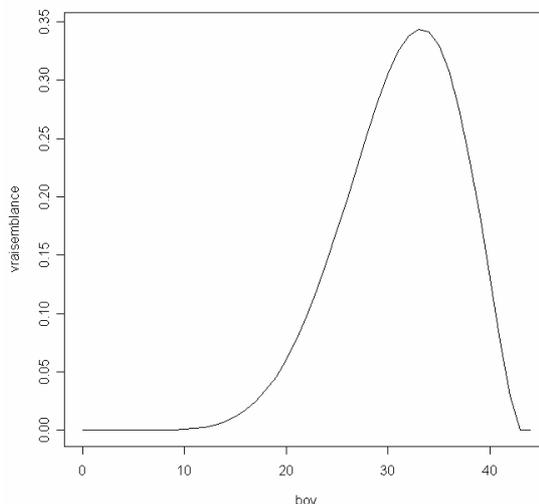
On extrait 8 bébés au hasard et on note le nombre de garçons.

```
> echal = sample(babyboom$gender, 8)
> echal
[1] boy boy boy girl boy girl boy boy
Levels: girl boy
> summary(echal)
girl boy
  2    6
```

1a) Donner la fréquence relative associée au nombre de garçons.

1b) On désire construire la valeur de la vraisemblance en fonction de l'hypothèse : nombre total de garçons dans la population. Donner la loi de probabilité associée.

1c) On construit le graphe. Que trouve-t-on en abscisse ? en ordonnée ?



1d) A l'aide des informations complémentaires suivantes, quelle est l'hypothèse la plus vraisemblable ? Donner l'estimation au maximum de vraisemblance du nombre de garçons.

```
> vraisemblance[29:39]
[1] 0.2550817 0.2814217 0.3048735 0.3240369 0.3374586 0.3437078 0.3414760
[8] 0.3297009 0.3077209 0.2754598 0.2336489
> boy[29:39]
[1] 28 29 30 31 32 33 34 35 36 37 38
```

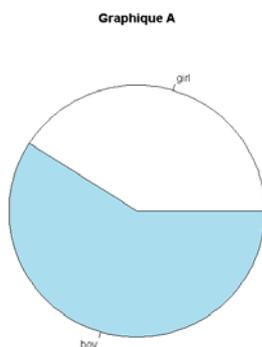
Question 2 : Comparaison générale

Le nombre de garçons dans la distribution totale est 26.

2a) Donner le pourcentage de garçons durant cette période.

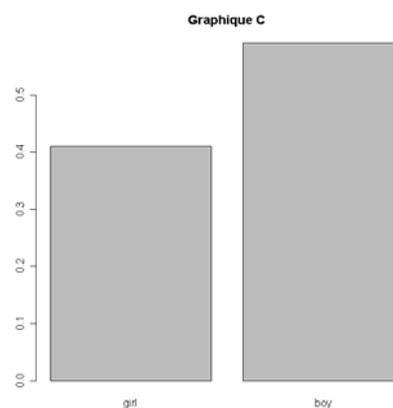
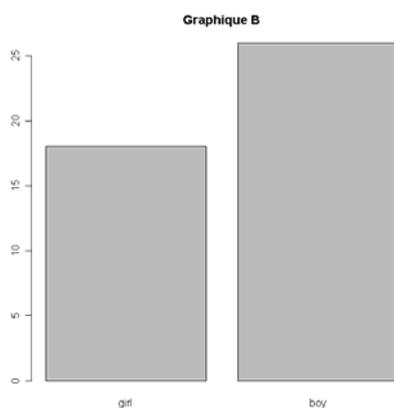
2b) On sait que la proportion de naissance masculine dans la population humaine est de 51%. Que dire de l'échantillon ?

Question 3 : Graphes de la répartition sexuelle des naissances



3a) Quelle instruction nous permet d'obtenir le graphique A ? Pourquoi ?

```
[1] pie(babyboom$gender)
[2] pie(summary(babyboom$gender))
```



3b) Commenter la répartition du sexe à la naissance.

3c) D'autres données ont été obtenues dans un hôpital de Cairns (Etat du Queensland, Australie). Choisir parmi les trois graphiques ci-dessus (A, B ou C) celui qui permettrait le mieux de comparer les répartitions dans les hôpitaux de Brisbane et Cairns. Expliquer votre choix.

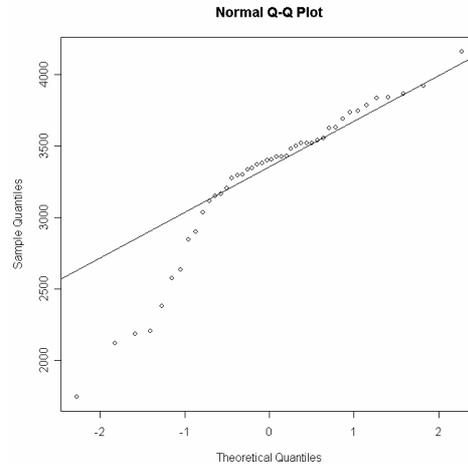
Question 3 : Ajustement à une loi de Poisson, nombre de naissances par heures

Question 4 : Temps d'attente entre deux naissances

Question 5 : Normalité du poids -A

5a) On réalise la représentation graphique ci-dessous. Commenter le résultat.

```
> qqnorm(babyboom$wt)
> qqline(babyboom$wt)
```



5b) Quelle hypothèse pourriez-vous émettre sur la normalité de la distribution des poids des 44 nouveaux nés ?

Question 6 : Normalité du poids -B

Pour tester la normalité d'une distribution, S.S. Shapiro et M.B. Wilk proposent la statistique suivante et soulignent que celle-ci « *...is very simple to compute once the table of linear coefficients is available...* ».

Prenons par exemple les poids de 4 bébés :

```
> testwt = sample(wt,4)
> testwt
[1] 3838 2184 3500 3166
```

Les valeurs sont rangées par ordre croissant :

```
> ordtestwt
[1] 2184 3166 3500 3838
```

6a) Quelle est la fonction R qui permet de classer des valeurs par ordre croissant ?

La somme des carrés des écarts est alors calculée :

```
> sum2
[1] 1527320
```

6b) Donner l'instruction qui permet de calculer `sum2` à partir de `var`.

ecrire b

où les `ai` sont les coefficients de normalité définis ci-dessus.

```
ctb = 0.6872*(ordtestwt[4]-ordtestwt[1]) +
      0.1677*(ordtestwt[3]-ordtestwt[2])
ctb
[1] 1192.641
```

6c) Calculer $W = b^2 / S^2$

Question 7 : Normalité du poids -C

7a) On teste la normalité du poids des 44 bébés par Kolmogorov-Smirnov et Shapiro-Wilks. Les résultats sont donnés ci-dessous :

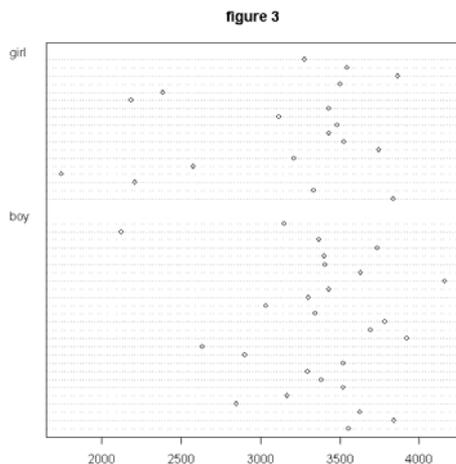
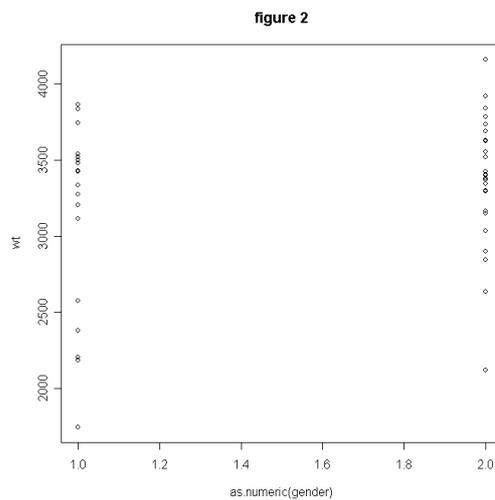
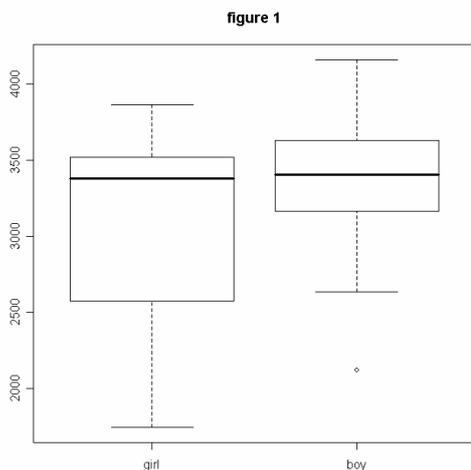
```
One-sample Kolmogorov-Smirnov test
data:  babyboom$wt
D = 0.1834, p-value = 0.1038
alternative hypothesis: two.sided
```

```
Shapiro-Wilk normality test
data:  babyboom$wt
W = 0.8987, p-value = 0.0009944
```

Commenter.

7b) Dans leur article publié en 1965, S.S. Shapiro et M.B. Wilk démontrent que leur test est plus puissant que celui de Kolmogorow et Smirnov.
Qu'est-ce que la puissance d'un test ?

Question 7 : Interprétation de graphiques du poids en fonction du sexe



7a) Mettre un titre pour chacune des trois figures.

7b) Commenter.

Question 8 : Normalité du poids au sein de chaque groupe : garçon, fille

```
> summary(wt[gender=="boy"])
```

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      2121   3198   3404   3375   3629   4162
> sd(wt[gender=="boy"])
[1] 428.0461
> summary(wt[gender=="girl"])
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1745   2711   3381   3132   3517   3866
> sd(wt[gender=="girl"])
[1] 631.5825

```

8a) A l'aide des tests ci-dessous, peut-on admettre la normalité du poids chez les garçons ? chez les filles ?

```

> shapiro.test(wt[gender=="boy"])
      Shapiro-Wilk normality test
data:  wt[gender == "boy"]
W = 0.9475, p-value = 0.2022

> shapiro.test(wt[gender=="girl"])
      Shapiro-Wilk normality test
data:  wt[gender == "girl"]
W = 0.8703, p-value = 0.01798

```

Question 8 : Comparaison des deux variances.

8a) Soit une loi du Chi-Deux à 26 degrés de liberté et soit une loi du Chi-Deux à 17 degrés de liberté. Donner la loi du rapport de deux Chi-Deux.

8b) On donne les résultats au test de comparaison de deux variances selon Fisher. Commenter.

```

var.test(wt[gender=="boy"],wt[gender=="girl"])

      F test to compare two variances

data:  wt[gender == "boy"] and wt[gender == "girl"]
F = 0.4593, num df = 25, denom df = 17, p-value = 0.07526

```

Question 9 : Comparaison des moyennes.

A l'aide de toutes les informations à votre disposition, répondre à la question : "Existe-t-il une différence, en moyenne, entre le poids des garçons nouveaux nés et celui des filles nouveaux nés ?"

```

> t.test(wt[gender=="boy"],wt[gender=="girl"],var.equal=T)

      Two Sample t-test

data:  wt[gender == "boy"] and wt[gender == "girl"]
t = 1.5229, df = 42, p-value = 0.1353
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -78.9779 564.7044
sample estimates:
mean of x mean of y
 3375.308  3132.444

> wilcox.test(wt[gender=="boy"],wt[gender=="girl"])

```

Wilcoxon rank sum test with continuity correction

```
data: wt[gender == "boy"] and wt[gender == "girl"]  
W = 273.5, p-value = 0.3519  
alternative hypothesis: true mu is not equal to 0
```

Question 10 : Lien entre les statistiques de Wilcoxon et de Kruskal-Wallis.

10a) Centrer et réduire la statistique W de Wilcoxon.

10b) Elever le résultat au carré et comparer au résultat ci-dessous de Kruskal-Wallis.
> `kruskal.test(wt,gender)`

Kruskal-Wallis rank sum test

```
data: wt and gender  
Kruskal-Wallis chi-squared = 0.8891, df = 1, p-value = 0.3457
```

10c) Conclure.