

DM de statistique

Exercice 1 : Soit Y une variable aléatoire ayant pour densité f_Y définie pour tout $x \in \mathbb{R}$ par

$$f_Y(x) = \frac{1}{2(1-x)^{1/2}} \mathbf{1}_{[0,1]}(x).$$

avec $f_Y(1) = +\infty$. Soit $\theta > 0$, on considère maintenant la variable aléatoire $X = \theta Y$ et on admet que sa densité f_θ est définie pour tout $x \in \mathbb{R}$ par

$$f_\theta(x) = \frac{1}{2\theta(1-x/\theta)^{1/2}} \mathbf{1}_{[0,\theta]}(x).$$

Dans tout ce qui suit, on considère des variables aléatoires i.i.d X_1, \dots, X_n de même loi que X .

1. On admet que $\mathbb{E}[Y] = \frac{2}{3}$ et $\mathbb{V}[Y] = \frac{4}{45}$. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Par la méthode des moments, proposer un estimateur de θ .
3. Est-il consistant? Asymptotiquement normal?
4. Soit $\alpha \in (0, 1)$, donner un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ .
5. Montrer que l'estimateur du maximum de vraisemblance vaut $X_{(n)}$ où $X_{(n)} = \max_{i=1, \dots, n} X_i$. A noter que l'on a alors une vraisemblance égale à $+\infty$.
6. Montrer que pour tout $x \in [0, 1]$, $F_Y(x) = 1 - (1-x)^{1/2}$ où F_Y est la fonction de répartition de Y .
7. En déduire les fonctions de répartition de X et $X_{(n)}$.
8. En déduire que $n \left(\theta - X_{(n)} \right)^{1/2}$ converge en loi vers $E \sim \mathcal{E}(\theta^{-1/2})$
9. Soit $\alpha \in (0, 1)$, trouver $c_\alpha > 1$ tel que

$$\mathbb{P} \left[\theta \leq X_{(n)} c_\alpha \right] = 1 - \alpha.$$

10. En déduire un nouvel intervalle de confiance.
11. Quel estimateur choisiriez-vous?

Exercice 2 : On relève chez 9 patients une glycémie moyenne de 1.12 g/l et une variance de 0.01. On admettra que la glycémie des patients suit une loi normale. Ces patients sont issus d'une population dont la glycémie moyenne vaut $m_0 = 1$ g/l. Au risque de 5%, tester si cet échantillon est représentatif du reste de la population. On pourra s'aider des résultats suivants :

$$\begin{aligned} \mathbb{P}[T_8 \leq 2.306] &= 0.975 & \mathbb{P}[T_8 \leq 1.860] &= 0.95 & \mathbb{P}[T_9 \leq 2.262] &= 0.975 \\ \mathbb{P}[T_9 \leq 1.833] &= 0.95 & \mathbb{P}[N \leq 1.960] &= 0.975 & \mathbb{P}[N \leq 1.645] &= 0.95 \end{aligned}$$

où T_k suit une loi de Student à k degrés de liberté et N suit une loi normale centrée réduite.

Exercice 3 : En vue d'estimer les différences de productivité qui peuvent exister entre plusieurs types de forêts de hêtre (*Fagus sylvatica* L.) de l'Ardenne belge, on a mesuré, en différents endroits, la hauteur des arbres les plus gros [Dagnélie, 1956-1957]. La hauteur de ces arbres, qui est étroitement liée à la production en volume, peut en effet être considérée comme une mesure simple, mais fiable, du niveau de productivité des forêts. Nous n'envisageons ici que trois types de hêtraies, au sein desquels on a observé les hauteurs des arbres, respectivement en 13 endroits, en 14 endroits et en 10 endroits différents, choisis au hasard et indépendamment les uns des autres. En chaque endroit, les cinq arbres les plus gros situés au sein d'une parcelle circulaire d'environ 15 m de rayon (soit environ 7 ares) ont été mesurés, chacun deux fois, et pour chaque lieu, la moyenne des 10 observations a été calculée. Pour éviter la présentation de données trop nombreuses, nous ne considérons que ces moyennes.

On souhaite comparer la productivité des forêts 1 et 2 en utilisant la hauteur des arbres.

On trouvera dans le tableau ci-dessous les mesures obtenues.

Forêt 1				Forêt 2			
$(x_{j,1})_{1 \leq j \leq 13}$				$(x_{j,2})_{1 \leq j \leq 14}$			
23.4	24.4	24.6	24.9	22.5	22.9	23.7	24.0
25.0	26.2	26.3	26.8	24.4	24.5	25.3	26.0
26.8	26.9	27.0	27.6	26.2	26.4	26.7	26.9
27.7				27.4 28.5			
$n_1 = 13$				$n_2 = 14$			

Tab. 1: Mesures des hauteur des arbres des forêts 1 et 2.

On notera $x_{1,1}, x_{2,1}, \dots, x_{13,1}$ les mesures obtenues pour la forêt 1 et $x_{1,2}, x_{2,2}, \dots, x_{14,2}$ celles de la forêt 2. On supposera que les données $x_{1,1}, x_{2,1}, \dots, x_{13,1}$ sont des réalisations indépendantes d'une variable aléatoire d'espérance μ_1 et de variance σ_1^2 et que $x_{1,2}, x_{2,2}, \dots, x_{14,2}$ sont des réalisations indépendantes d'une variable aléatoire d'espérance μ_2 et de variance σ_2^2 .

1. On trouvera dans la figure ci-dessous les boîtes à moustaches des deux séries de données ainsi que les histogrammes en fréquences des 2 séries.

Quelles informations peut-on tirer de ces deux graphiques ?

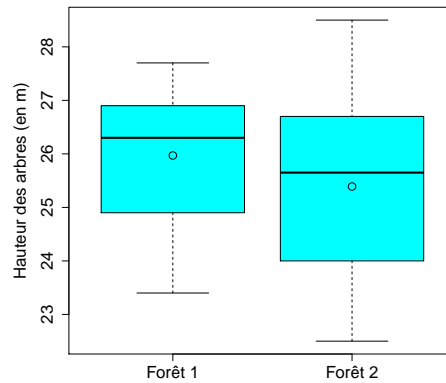


Fig. 1: Boîtes à moustaches.

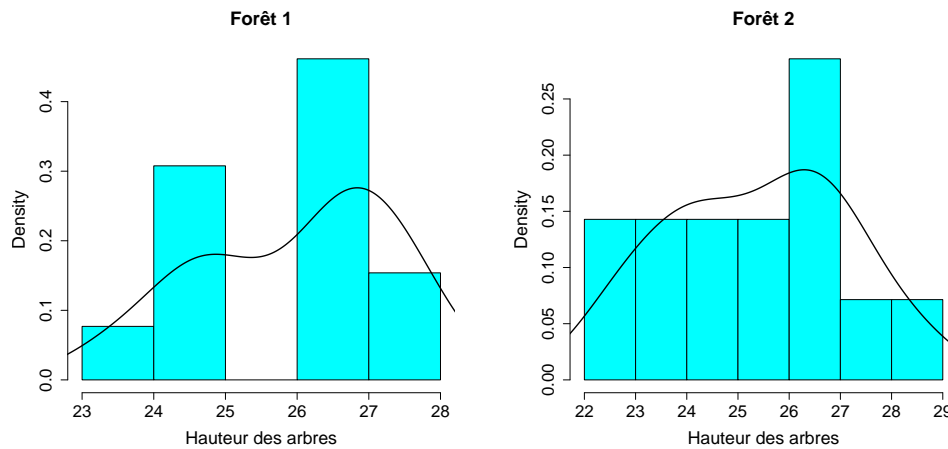


Fig. 2: Histogrammes en fréquences et densités.

2. On effectue le test de Shapiro-Wilk sur chacune des séries de mesures et on obtient les p -values respectives 0,27 et 0,91. Que pouvez-vous en conclure?
3. La mise en oeuvre, avec le logiciel R et la fonction `var.test`, du test de comparaison de variances de Fisher conduit au résultat suivant :

F test to compare two variances

data: Hauteur by Foret

F = 0.58696, num df = 12, denom df = 13, p-value = 0.3647

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1861495 1.9013248

sample estimates:

```
ratio of variances
      0.5869621
```

Tester si σ_1^2 est égale à σ_2^2 au risque de 5% et au risque de 1%.

4. On souhaite comparer la productivité des deux forêts. Quel test doit-on mettre en oeuvre pour cela. Rappeler le cadre théorique nécessaire à la mise en oeuvre de ce test.

Compte-tenu des résultats des questions précédentes, peut-on dire que les conditions d'application du test sont satisfaites? On justifiera la réponse.

5. Tester aux risques 5% et 1% l'hypothèse selon laquelle la productivité des deux forêts est la même. On précisera les hypothèses testées, la statistique de test et sa loi sous H_0 , les 2 zones de rejet et on conclura. Pour vous aider, on fournit une partie du résultat donné par la fonction `t.test` du logiciel R :

```
Two Sample t-test
```

```
data:  c1 and c2
```

```
t = 0.96377, df = 24.168, p-value = 0.3447
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.6656157  1.8326487
```

```
sample estimates:
```

```
mean of x mean of y
```

```
25.96923  25.38571
```