

TP 2 : Estimations et Tests d'Hypothèses

Exercice 1 : Finir le TP1. Rendre un rapport écrit sur la partie 3 avant le prochain TP.

Exercice 2 : Etude du niveau sonore autour d'un aéroport (Rapport écrit à rendre avant le prochain TP)

Dans cet exercice, on s'intéresse au niveau de bruit sur un aéroport généré par les décollages et atterrissages des avions ainsi qu'aux nuisances que cela entraîne chez les riverains. On mesure, grâce à une sonde, le niveau de bruit en décibel sur un site S voisin de l'aéroport. Les résultats des mesures, effectuées sur un mois, sont donnés dans le fichier `aeroport-1.txt` disponible à l'adresse <http://epoc.isima.fr/~diarrassouba/enseignements.htm>.

Les riverains du site S se plaignent de l'intensité du bruit engendré par l'activité de l'aéroport et affirme que le niveau sonore est supérieur à 80dB, qui est la limite légale autorisée. Il disent aussi que cela a une incidence négative sur le sommeil chez la majorité d'entre eux. La direction de l'aéroport, quant à elle, affirme que le niveau moyen de bruit à proximité du site S est de 78.35dB. Un expert (c'est-à-dire vous) est dépêché sur les lieux pour trancher entre les deux parties. On modélise le niveau de bruit par une variable aléatoire X suivant une loi dont on estimera la forme, la moyenne μ et l'écart-type σ .

Q1. Tests et estimations du niveau moyen de bruit

1. Téléchargez le fichier `aeroport-1.txt` et lisez les données qu'ils contiennent (fonction `scan()`, précisez `comment.char="#"` pour éviter les commentaires dans le fichier). Affichez le vecteur résultat et déterminez le nombre n de mesures effectuées.
2. A partir de ces mesures déterminez les niveaux de bruit minimum et maximum enregistrés sur le site et dites combien de mesures sont ≥ 80 dB (utilisez les expressions d'extraction de sous-tableaux `x[x >= 80]`).
3. Tracez l'histogramme de la répartition des bruits (fonction `hist()`). Vous préciserez comme vecteur de classes le vecteur `(60,65,70,75,80,85,90,95)`.
4. De quelle distribution semble s'approcher la forme de l'histogramme ?
5. On souhaite connaître le niveau moyen de bruit et savoir qui des riverains ou de l'aéroport a raison. On va donc tester, avec un risque d'erreur $\alpha = 5\%$, les hypothèses nulle $H_0 : \mu = 80$ et alternative $H_1 : \mu = 78.35$. Déterminez la forme de la région critique.
6. Calculez la moyenne et l'écart-type empirique de l'échantillon (fonctions `mean()` et `sd()`).
7. Dites pourquoi peut-on considérer que la variable $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ suit une loi normale centrée réduite, même si on ne connaît pas la distribution de la variable X .
8. Calculez $z_{\frac{\alpha}{2}}$, le quantile d'ordre $\frac{\alpha}{2}$ d'une loi normale centrée réduite. Utilisez la fonction `qnorm()` en précisant la valeur $\frac{\alpha}{2}$ comme argument.
9. Calculez la valeur $z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ (s est l'estimation de l'écart-type calculé avec la fonction `sd()`). Situez la valeur de z par rapport à la région critique. Quel est le résultat du test.
10. Déterminez un intervalle de confiance pour μ avec un niveau de confiance de 95%.

11. Que peut-on en déduire quant à l'affirmation des riverains ?
12. **Utilisation de la fonction "t.test()"** : La fonction `t.test(x,conf.level)` peut être utilisée pour estimer une moyenne par intervalle de confiance pour des échantillons de grandes tailles ($n \geq 30$) ou de petites tailles ($n < 30$). L'argument `x` est le vecteur de données sur lequel est réalisée l'estimation. L'argument `conf.level`, comme son nom l'indique, permet de spécifier le niveau de confiance de l'estimation. `t.test()` renvoie la valeur de la moyenne empirique de l'échantillon `x` ainsi que les bornes de l'intervalle de confiance.

Tapez les commandes suivantes et comparez les résultats retournés avec ceux que vous avez obtenus à la question 10.

```
t.test(x,conf.level=0.95) ; //Intervalle de confiance à 95%
t.test(x,conf.level=0.80) ; //Intervalle de confiance à 80%
t.test(x,conf.level=0.10) ; //Intervalle de confiance à 10%
t.test(x,conf.level=0.01) ; //Intervalle de confiance à 1%
```

13. Que constatez-vous en ce qui concerne la largeur de l'intervalle de confiance lorsque le niveau de confiance diminue ? Donnez une explication.
14. Quelle semble être la vraie valeur de μ ?

Q2. Estimation de la variance de X par intervalle de confiance

On va maintenant déterminer un intervalle de confiance pour la variance de X à 95% (risque d'erreur $\alpha = 5\%$).

1. Déterminez les quantiles d'ordre $1 - \frac{\alpha}{2}$ et $\frac{\alpha}{2}$ de degré de liberté $n - 1$ pour la loi du χ^2 (utilisez la fonction `qchisq()` : consultez l'aide en ligne pour connaître les paramètres).
2. Calculez les bornes de l'intervalle de confiance de σ^2 . Que constatez-vous quant à la largeur de cet intervalle ?

Dans la suite, on supposera que la variable X suit une loi normale de moyenne $\mu = 78.35$ et de variance $\sigma^2 = 22$.

Q3. Estimation du niveau de bruit durant la période de 22h-6h

Selon certaines études, le sommeil d'un riverain est perturbé, dans 90% des cas, lorsque celui-ci est soumis pendant sommeil à plus de 8 bruits d'au moins 80dB. Pour savoir si le sommeil des riverains du site S est perturbé ou non, il nous faut donc savoir combien d'avions peuvent générer un bruit supérieur à 80dB sur la période 22h-6h. La connaissance de la distribution de X ainsi que sa moyenne μ et de sa variance σ^2 nous permet de déterminer la proportion théorique d'avions qui engendrent un niveau de bruit ≥ 80 dB, et ce sur un effectif donné. Le fichier `aeroport-2.txt` décrit le trafic moyen sur l'aéroport à chaque heure de 0h jusqu'à 24h. La première valeur correspond au trafic entre 0h et 1h, la deuxième correspond au trafic entre 1h et 2h, etc. Ces données nous serviront à estimer le nombre moyen d'avions qui décollent ou atterrissent sur l'aéroport entre 22h et 6h du matin.

1. Téléchargez le fichier `aeroport-2.txt` et lisez les données contenues dans le fichier. Tracez avec la fonction `plot()` le graphique associé (précisez comme paramètre `type="h"`, `col="red"` et `lwd = 5` : consultez l'aide en ligne pour connaître les paramètres).
2. Que constatez-vous en ce qui concerne la répartition des vols. A quelles heures observe-t-on les pics d'activités ?

3. Calculez le nombre n d'avions qui ont atterri ou décollé entre 22h et 6h.
4. Calculez avec la fonction `pnorm()` la probabilité $p_{80} = P(X \geq 80)$, proportion d'avions qui génèrent un bruit ≥ 80 dB. Attention au sens de l'inégalité et à la valeur renvoyée par `pnorm()`.
5. A partir de n et p_{80} , calculez le nombre théorique d'avions qui circulent sur l'aéroport entre 22h et 6h et qui génère un bruit ≥ 80 dB.
6. Concluez quant à l'inquiétude des riverains sur l'impact qu'a les activités de l'aéroport sur leur sommeil.