Examen de Statistique Inférentielle

Durée : 2h00.
La qualité de la rédaction sera prise en compte dans la notation. Les documents, les téléphones mobiles et les calculatrices sont interdits.

On rappelle que la densité d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ est

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Exercice 1.

On considère un échantillon $X_1, \ldots, X_n$ où $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et on souhaite tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu = \mu_1$.
(On suppose que $\mu_0 < \mu_1$ et que la valeur de σ^2 est connue).

1. Montrer que la région critique du test du rapport de vraisemblance est de la forme :
   On rejette H_0 avec un risque de première espèce α si $\overline{X}_n > k_\alpha$.
2. Ecrire le code (langage $\mathbb{R}$) permettant de calculer la valeur de k_α.
3. Calculer la puissance de ce test. Pourrait-on trouver un test plus puissant ?
4. Montrer qu'il s'agit d'un test convergent.

Exercice 2.

Les données proviennent d'une étude sur un échantillon de 62 espèces de mammifères terrestres pour lesquelles on a calculé le poids moyen du cerveau (variable `brain`, en grammes) ainsi que le poids moyen du corps (variable `body` en kg).

1. Que remarquez vous d'inhabituel sur ces statistiques descriptives ?

```
summary(mammals)
      body               brain
 Min.   :   0.005   Min.   :   0.14
 1st Qu.:   0.600   1st Qu.:   4.25
 Median :   3.342   Median :  17.25
 Mean   : 198.790   Mean   : 283.13
 3rd Qu.:  48.203   3rd Qu.: 166.00
 Max.   :6654.000   Max.   :5712.00
```

2. On souhaite établir une relation (statistique) entre le poids du cerveau et le poids du corps. On construit pour cela les nouvelles variables $Y = \log(brain)$ et $X = \log(body)$.
   Expliquer le choix de modéliser Y en fonction de X plutôt qu'avec les variables de départ `brain` en fonction de `body`.
3. On trace dans (voir Figure 1) le nuage de points de Y en fonction de X. Le choix précédent vous semble-t-il justifié ?
4. On considère le modèle linéaire suivant

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1 = 1, \ldots, 62$$

   où les $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ et sont indépendants.
   Rappeler l'écriture de la vraisemblance du modèle et écrire les équations du score (on raisonnera comme dans le cours, conditionnellement aux valeurs de X).

5. On obtient :

```
lm(formula = log(brain) ~ log(body), data = mammals)


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.13479    0.09604   22.23   <2e-16 ***
log(body)    0.75169    0.02846   26.41   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6943 on 60 degrees of freedom
Multiple R-squared:  0.9208,Adjusted R-squared:  0.9195
F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16
```

   Expliquer et interpréter les 3 points suivants

   - `F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16`
   - `Multiple R-squared:  0.9208`
   - `Residual standard error: 0.6943 on 60 degrees of freedom`

6. Un mammifère pèse 1 kg. Quelle valeur de Y ajustez vous par ce modèle ?
7. Peut-on supposer qu'il y a une relation de proportionnalité en Y et X, soit $\beta_0 = 0$?
8. Construire un intervalle de confiance 0.95 pour β_1.
   (on supposera pour simplifier les calculs que le quantile de la loi correspondante vaut 2).
   Quelle instruction du langage $\mathbb{R}$ permet d'obtenir la vraie valeur de ce quantile ?

Exercice 3.

On s'intéresse au modèle linéaire suivant :

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \ldots, n$$

où les $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ et sont indépendants. Les x_i ne sont pas aléatoires et on supposera que $x_i > 0$, pour $i = 1, \ldots, n$ et que $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu_x > 0$.

1. Construire les estimateurs du maximum de vraisemblance de β et de σ^2. Montrer que l'estimateur de β (noté $\hat{\beta}$) est sans biais, calculer sa variance et en déduire qu'il est convergent.
2. En remarquant que $\mathbb{E}(\epsilon) = 0$, construire l'estimateur de β obtenu par la méthode des moments.
3. Construire un intervalle de confiance asymptotique $1 - \alpha$ de β.
4. Comment tester l'hypothèse $H_0 : \beta = 1$ contre $H_1 : \beta < 1$.
5. On s'intéresse au paramètre $\theta = \exp \beta$. Proposer un estimateur de θ, montrer qu'il est convergent et donner sa loi asymptotique.

Exercice 4.

Expliquer le principe du test du khi-2 d'ajustement à une loi donnée : hypothèses testées, mise en forme des données et choix de la statistique de test, construction de la région critique et règle de décision.

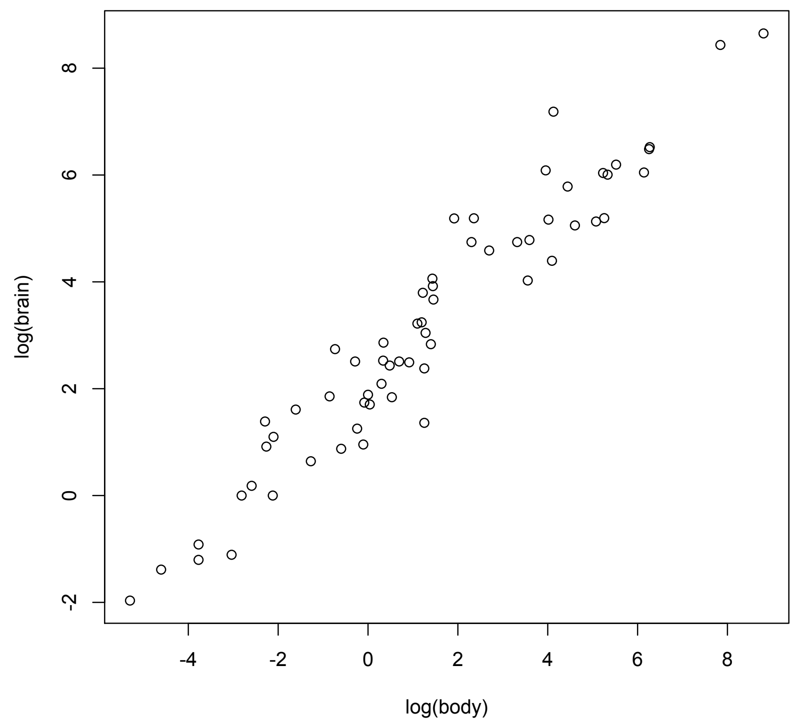

FIGURE 1 – Poids du cerveau (log) en fonction du poids du corps (log).