
---

INITIATION AU LOGICIEL R - TP 3

## Fonctions

---

**Exercice 1 :** Le code fourni dans le fichier **function1.txt** permet de définir une fonction appelée **facto**. Cette fonction prend un nombre $n$ comme argument. Soumettre le code dans la console, en utilisant successivement les valeurs suivantes pour $n$ : -1, 0, 3.5, 4. Commenter les résultats otenus et le code proposé (on s'attachera en particulier à comprendre comment est définie une fonction sous `R`).

**Exercice 2 : Simulation de lois usuelles**

Sans utiliser les fonctions `rchisq` et `rt`,

1. Ecrire une fonction nommée **chideux** permettant de simuler des réalisations d'une loi $\chi^2_n$. Cette fonction aura pour arguments le nombre de tirages souhaités et le nombre de degrés de liberté de la loi du chi-deux. Utiliser cette fonction pour simuler 1000 observations du $\chi^2_3$. Tracer un histogramme de ces 1000 observations et superposer la densité du $\chi^2_3$.
2. Ecrire une fonction **student** permettant de simuler des réalisations suivant une loi de Student à $n$ degrés de liberté. Cette fonction aura pour arguments le nombre de tirages souhaités et le nombre de degrés de liberté de la loi de Student.
   Utiliser la fonction précédente pour simuler 1000 observations suivant la loi de Student à 7 degrés de liberté. Tracer un histogramme de ces 1000 observations et superposer la densité de la loi de Student à 7 degrés de liberté.

**Rappels**

- Si pour tout $i = 1, \dots, n$, $X_i$ i.i.d. $\sim \mathcal{N}(0,1)$, alors $\displaystyle\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$.
- Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2_n$ sont indépendantes, alors $T = \dfrac{X}{\sqrt{Y/n}} \sim \mathcal{T}_n$.

**Exercice 3 : Régression linéaire**

1. On souhaite ajuster une droite d'équation $y = ax + b$ à un nuage de $n$ points $M_i(x_i, y_i)$. Les estimateurs des moindres carrés ordinaires de $a$ et $b$ minimisent la somme

$$S = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2.$$

   On rappelle l'expression de ces estimateurs :

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

   (a) Ecrire une fonction **resume.reg** qui prend comme arguments les vecteurs `x` et `y` et retourne à l'utilisateur les estimations $\hat{a}$ et $\hat{b}$.

   (b) Ecrire une fonction **affiche.reg** qui prend comme arguments les vecteurs `x` et `y` et trace le nuage de points $(x_i, y_i)$ ainsi que la droite de régression.

(c) On reprend le jeu de données `mtcars`. Utiliser la fonction **resume.reg** pour calculer les estimations $\hat{a}$ et $\hat{b}$ dans le modèle mpg $= b + a\,\mathrm{hp} + \varepsilon$. Représenter la droite de régression avec `affiche.reg`. Vérifier les résultats obtenus en utilisant la fonction **lm**.

2. Considérons une variable quantitative à expliquer $\mathbf{y}$ et $p$ variables quantitatives explicatives $\mathbf{x}^1, \ldots, \mathbf{x}^p$. On dispose d'un échantillon

$$(x_i^1, \ldots, x_i^j, \ldots, x_i^p, y_i) \quad i = 1, \ldots, n.$$

d'observations indépendantes. Le modéle linéaire (multiple) s'écrit de la façon suivante :

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_i^j + \varepsilon_i \quad i = 1, \ldots, n$$

On suppose que les variables aléatoires $\varepsilon_i$ sont indépendantes et identiquement distribuées selon une loi normale centrée. Ce modèle admet l'écriture matricielle suivante : $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ où :

- $\mathbf{y} = (y_1, \ldots, y_n)^\top$
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$
- $\mathbf{X}$ est une matrice de taille $n \times (p+1)$ de terme général $x_i^j$ avec $x_i^0 = 1$ pour tout $i = 1, \ldots, n$.

L'estimateur MCO de $\beta$ est donné par $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.

(a) Ecrire une fonction `reg` qui prend comme arguments le tableau contenant les variables explicatives `var_exp` et le vecteur `y` et qui rend l'estimation du vecteur $\boldsymbol{\beta}$.

(b) On reprend le jeu de données `mtcars`. Utiliser la fonction `reg` pour calculer les estimations des MCO dans le modèle mpg $= \beta_0 + \beta_1 \mathrm{hp} + \beta_2 \mathrm{wt} + \varepsilon$. Vérifier les résultats obtenus en utilisant la fonction **lm**.