UE Probabilité et Statistique - Mathématiques

Devoir Maison n°8

Vincent Escoffier, Adrien Jallais, Théo Martel, Louis Muzellec.

04 mai 2022

Contents

Préparation de l'environnement

```
library(readr)
library(dplyr)
library(ggplot2)
library(DescTools)
```

Exercice 1

Contexte

Soit X, la variable aléatoire représentant le nombre d'accidents par assuré. Les mesures de l'échantillon sont entières. Il est admis que l'occurence des sinistres X suit une loi de Poisson pour laquelle on recherche le paramètre inconnu λ.

Description des données

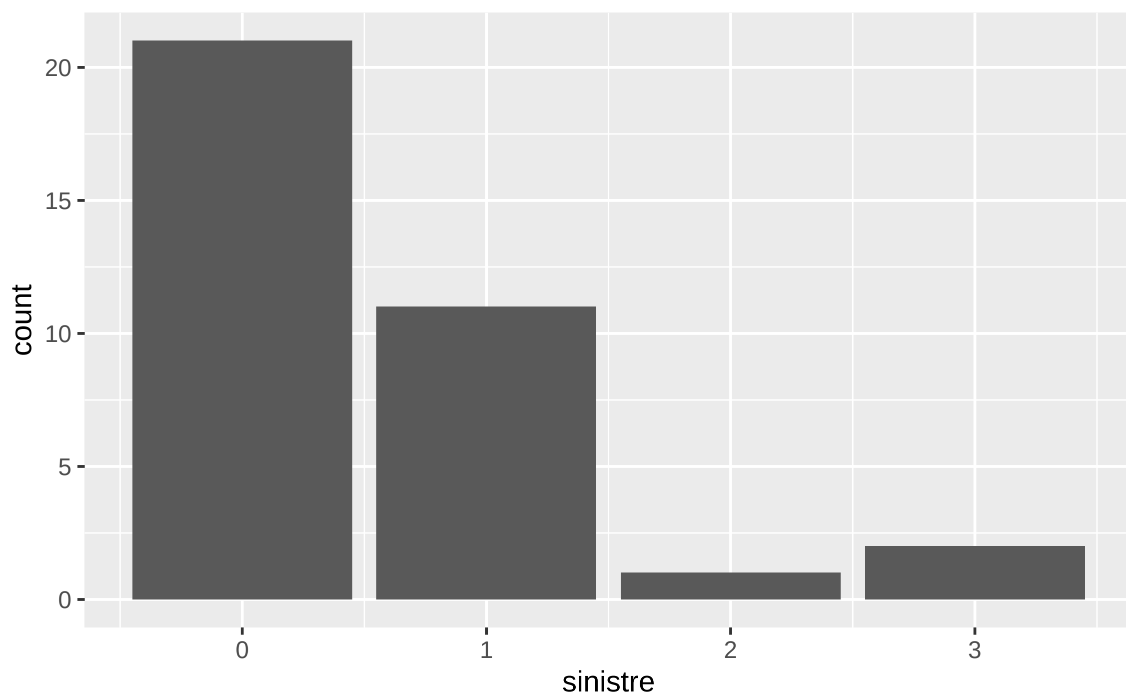
On importe les données avec la commande suivante :

```
d1 <- read_delim("../data/donnees.exo1.csv", delim = ";", escape_double = FALSE,
    trim_ws = TRUE, show_col_types = FALSE)
```

L'effectif de l'échantillon est de 35. Celui-ci peut donc être considéré de grande taille.

Les indicateurs et le graphe suivant résument la dispersion de nos données :

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.5429  1.0000  3.0000
```



On a également X :

- $X2$: La valeur critique du khi-carré
- N : Le nombre d'événements observés
- α : Le niveau de signification

A partir de $\overline{X} = 35$, on a $n\overline{X_n} = 19$.

Le modèle de nos données est le suivant : ???.

Calcul d'un interval de confiance IC unilatéral de niveau 95% pour λ

Comme ???, on peut approximer la loi de Poisson par une loi Normale centrée réduite $N(0,1)$ par :

$$\frac{\sqrt{n}(Xn-\lambda)}{\sqrt{Xn}} \to N(0,1)$$

Avec cette approximation, on peut en déduire un intervalle de confiance (IC) asymptotique (car $n = 35$ < 100).

IC_g asymptotique gauche

On obtient un IC tel que :

$$IC^{1-\alpha asymp}_{\underline{X_n}}(\lambda) = \left[\overline{X_n} - Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{X_n}}{n}} < \lambda\right] : (1)$$

$$(1) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = \left[0,543 - Z_{1-\frac{0,05}{2}} \times \sqrt{\frac{0,543}{35}} < \lambda\right]$$

$$(1) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = \left[0,543 - 1,96 \times \sqrt{\frac{0,543}{35}} < \lambda\right]$$

$$(1) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = [0,298 < \lambda]$$

IC_d asymptotique droit

On obtient un tel que :

$$IC^{1+\alpha asymp}_{\underline{X_n}}(\lambda) = \left[\overline{X_n} + Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{X_n}}{n}} > \lambda\right] : (2)$$

$$(2) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = \left[0,543 + Z_{1-\frac{0,05}{2}} \times \sqrt{\frac{0,543}{35}} > \lambda\right]$$

$$(2) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = \left[0,543 + 1,96 \times \sqrt{\frac{0,543}{35}} > \lambda\right]$$

$$(2) \Leftrightarrow IC^{0,95 asymp}_{\underline{X_n}}(\lambda) = [0,787 > \lambda]$$

Peut-on accepter l'hypothèse que $\lambda = 1$?

On pose $H_0 : \lambda = 1$ et $ H_1 : 1$. On ne peut accepter H_0, car $\lambda \notin IC_g$ et $\lambda \notin IC_d$, autrement dit $\lambda \notin [0.298; 0.787]$

On peut vérifer cette négation, avec l'estimateur $W^{-\alpha}_{\underline{X_n}}(\lambda)$ de la manière suivante :

$$W^{\alpha}_{\underline{X_n}}(\lambda) = \left[\frac{X_n}{\overline{\overline{\overline{X_n}}}} < a_\alpha\right] \text{ ou } W^{\alpha}_{\underline{X_n}}(\lambda) = [\overline{X_n} > b_\alpha]$$

$$P_{\lambda=1}(W^{-\alpha}_{\underline{X_n}}(\lambda)) \leq \alpha \text{ et } P_{\lambda\neq 1}(W^{\alpha}_{\underline{X_n}}(\lambda)) \geq \alpha$$

Exercice 2

Description du modèle de données

Exercice 3

Contexte

Nous introduisons les variables x, A, S, E, Y, θ, et Z comme suit : Une expérience chimique consiste à ajouter une dose x d'un agent A dans une solution S. Après réaction, on mesure la quantité d'une espèce E. Pour x donné, on suppose qu'il est pertinent de représenter cette mesure par une variable aléatoire $Y = \theta x^2 + aZ$, a connu, θ un paramètre réel inconnu et Z une variable aléatoire normale centrée réduite.

Description des données

On importe les données avec les commandes suivantes :

```
d3 <- read_delim("../data/donnees.exo3.csv", ";", col_names = TRUE, escape_double = FALSE,
    show_col_types = FALSE, locale = locale(decimal_mark = ","))
d3$x <- as.numeric(d3$x)
d3$y <- as.numeric(d3$y)
d3 <- select(d3, x, y)
```

L'effectif des données est de 10.

La table suivante résume la dispersion de nos données :

```
##        x               y         
##  Min.   :0.500   Min.   : 0.400  
##  1st Qu.:1.625   1st Qu.: 2.345  
##  Median :2.750   Median : 6.260  
##  Mean   :2.750   Mean   : 7.620  
##  3rd Qu.:3.875   3rd Qu.:11.852  
##  Max.   :5.000   Max.   :19.900  
```