UE Probabilité et Statistique - Mathématiques

Devoir Maison n°8

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

09 May 2022

Contents

Préparation de l'environnement

R et Rstudio seront utilisés pour la rédaction de ce DM, ainsi que les packages suivants :

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

Exercice 1

Contexte

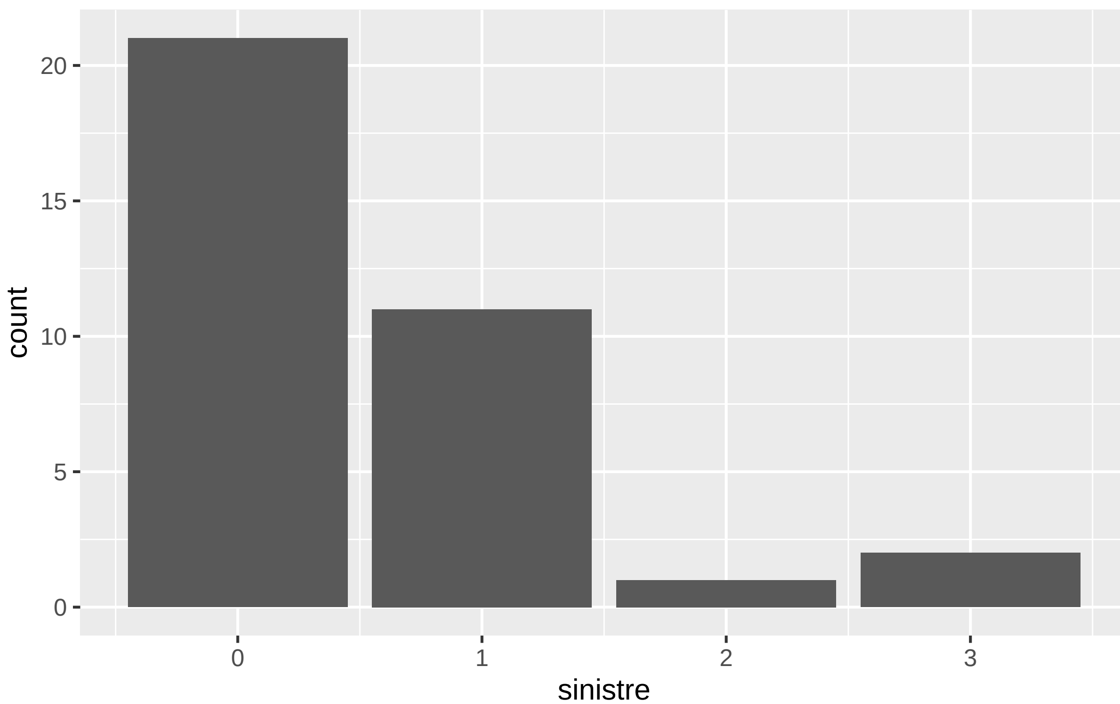
Soit X, la variable aléatoire représentant le nombre d'accidents par assuré. X est une variables aléatoire discrète. Il est admis que l'occurence des sinistres X suit une loi de Poisson pour laquelle on recherche le paramètre inconnu λ. Le modèle d'échantillonage de nos données est le suivant : $(\mathbb{N}, (\mathcal{P}_\lambda)_{\lambda>0})^n$.

Description des données

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
```



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

Intervalle de confiance (IC) asymptotique pour λ

Comme la variable aléatoire X suit une loi de Poisson de paramètre λ et que notre effectif d'échantillon n est grand, on utilise la fonction pivotale asymptotique :

$$\frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\overline{X}_n}} \rightsquigarrow N(0,1)$$

Ainsi, on peut en déduire un intervalle de confiance (IC) asymptotique.

$$IC_{\underline{X}_n}^{1-\alpha \text{ asymp}}(\lambda) = \left[\overline{X}_n - Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{X}_n}{n}} ; \overline{X}_n + Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{X}_n}{n}}\right]$$

1. Calcul d'un interval de confiance asymptotique unilatéral gauche de niveau 95% pour λ (IC_g)

On obtient un IC tel que :

$$IC_{\underline{X}_n}^{1-\alpha \text{ asymp}}(\lambda) = \left[\overline{X}_n - Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{X}_n}{n}} < \lambda\right] : (1)$$

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

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

$$(1) \Leftrightarrow IC_{\underline{X}_n}^{0,95 \text{ asymp}}(\lambda) = [0,298 < \lambda]$$

2. Calcul d'un interval de confiance asymptotique unilatéral droit de niveau 95% pour λ (IC_d)

On obtient un tel que :

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

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

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

$$(2) \Leftrightarrow IC_{\underline{X}_n}^{0,95 \text{ asymp}}(\lambda) = [0,787 > \lambda]$$

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

Test d'hypothèses

On pose $H_0 : \lambda = 1$ et $H_1 : \lambda \neq 1$.

Observations

On a $\lambda \notin IC_g$ et $\lambda \notin IC_d$, autrement dit $\lambda \notin [0.298; 0.787]$.

Décision

On rejette H_0.

Conclusion

Par conséquent, on a $\lambda \neq 1$.

Commentaires

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

$$W_{\underline{X}_n}^{\alpha}(\lambda) = \left[\frac{\underline{X}_n}{\overline{\overline{X}}_n} < a_\alpha\right] \text{ ou } W_{\underline{X}_n}^{\alpha}(\lambda) = \left[\overline{X}_n > b_\alpha\right]$$
$$P_{\lambda=1}(W_{\underline{X}_n}^{-\alpha}(\lambda)) \leq \alpha \text{ et } P_{\lambda\neq 1}(W_{\underline{X}_n}^{\alpha}(\lambda)) \geq \alpha$$

Exercice 2

Contexte

Soit X, une variable aléatoire discrète représentant le nombre de pièces défectueuses par échantillon (représentée par une commande ou un lot). Les pièces peuvent être soient défectueuses, soient fonctionnelles. Notre échantillonage est donc extrait d'une épreuve de Bernouilli. La taille de l'échantillon est grande : $n > 100$ avec $n = 140$.

On souhaite savoir si le client acceptera cet echantillon, et pour cela il faut qu'il contienne au moins 120 composants fonctionnels.

Description du modèle de données

Notre échantillonnage aléatoire est simple tel que : $X_n = (X_i)_{1\leq i\leq n}$ avec le modèle suivant : $(0,1(B(1,p)_p)_{p\in[0,1]})$ sachant $p = 10\% = 0,1$.

On va comparer ce paramètre avec une estimation ponctuelle, afin de savoir si l'affirmation du fabricant est vraie. Pour cela nous allons réaliser deux tests : le premier du point de vue du fabricant (TH_1), le second du point de vue du client (TH_2).

Observation ponctuelle

Dans notre observation ponctuelle (lot), on a observé que la proportion de pièces défectueuses est de 0.3 (`= 12/40`).

Point de vue du fabricant (TH_1)

Test d'hypothèses

$$H_0 : p_0 \leq 0,1 \text{ contre } H_1 : p_0 > 0,1$$

$$W_{X_n}^{\alpha \text{ asymp}}(p) = \left\{ \frac{X_n}{\overline{X}_n} > p + \frac{Z_{1-\alpha}}{\sqrt{n}} \times \sqrt{p(1-p)} \right\} : (3)$$

$$(3) \Leftrightarrow W_{140}^{0,05 \text{ asymp}}(p) = \left\{ \frac{X_n}{\overline{X}_n} > 0,1 + \frac{1,96}{\sqrt{140}} \times \sqrt{0,1(1-0,1)} \right\}$$

$$(3) \Leftrightarrow W_{140}^{0,05 \text{ asymp}}(p) = \left\{ \frac{X_n}{\overline{X}_n} > 0,1496 \right\}$$

Observations

$$\frac{X_n}{\overline{X}_n} = 0,0857 \notin W_{X_n}^{\alpha \text{ asymp}}(p)$$

Décision

On accepte donc H_0 de TH_1.

Conlcusion

Jusqu'à preuve du contraire, le client acceptera le lot.

Point de vue du client (TH_2)

Test d'hypothèses

$$H_0 : p_0 \geq 0,1 \text{ contre } H_1 : p_0 < 0,1$$

$$W_{X_n}^{\alpha \text{ asymp}}(p) = \left\{ \frac{\underline{X}_n}{\overline{X}_n} > p - \frac{Z_{1-\alpha}}{\sqrt{n}} \times \sqrt{p(1-p)} \right\} : (3)$$

$$(3) \Leftrightarrow W_{140}^{0,05 \text{ asymp}}(p) = \left\{ \frac{\underline{X}_n}{\overline{X}_n} > 0,1 - \frac{1,96}{\sqrt{140}} \times \sqrt{0,1(1-0,1)} \right\}$$

$$(3) \Leftrightarrow W_{140}^{0,05 \text{ asymp}}(p) = \left\{ \frac{\underline{X}_n}{\overline{X}_n} > 0,0503 \right\}$$

Observations

$$\frac{\underline{X}_n}{\overline{X}_n} = 0,0857 \notin W_{X_n}^{\alpha \text{ asymp}}(p)$$

Décision

On accepte donc H_0 TH_2.

Conlcusion

Jusqu'à preuve du contraire, le client n'acceptera pas le lot.

Commentaires

L'acceptation du lot a été évaluée depuis les points de vues des deux parties.

On se place du point de vue du client, il ne faudrait donc pas tester TH_2 mais seulement TH_1. En effet, en se placant du côté du client, on souhaite seulement évaluer ce que craint le client de rejeter à tord.

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.

Estimation du paramètre θ

Pour estimer le paramètre θ, on fait $n(n \geq 1)$ essais indépendants avec des doses de l'agent A notées $X_1, ... X_n$. De Y, on extrait donc un échantillon aléatoire bernoullien $Y_n = (Y_1 ... Y_n)$.

Modèle statistique associé à Y_n

On pose le modèle statistique suivant à deux paramètres :

$$(\mathbb{R}^n, \otimes_{i=1}^n N(\theta x^2, a^2)_{(\theta,a)\in\mathbb{R}^2\times]0;+\infty[})$$

Avec pour vraissemblance :

$$L(Y_{\underline{X}_n}, \theta, a^2) = (2\pi a^2)^{\frac{-n}{2}} \times e^{\frac{\sum_{i=1}^n (Y_i - \theta_i^2)^2}{2a^2}}$$

et

$$logL(Y_{\underline{X}_n}, \theta, a^2)) = log(2\pi)^{\frac{-n}{2}} - \frac{n}{2} \ln x - \frac{\sum_{i=1}^n (Y_i - \theta_i^2)^2}{2a^2} = g(\theta, a^2)$$

Régularité du modèle Y_n

On pose Z une variable variable aléatoire normale centrée réduite, tel qu'en utilisant les notations précédentes : $Z = \frac{Y - \theta x^2}{a}$.

Or on sait selon le théorème 1 du chapitre sur la régression linéaire que le modèle : $(\mathbb{R}^n, \otimes_{i=1}^n N(\theta_0 + \theta_1 x^2, a^2)_{(\theta_0,\theta_1,a)\in\mathbb{R}^3\times]0;+\infty[})$: (3) Et sachant $\theta_0 = 0$ et $x' = x^2$ on a : (3) $\Rightarrow (\mathbb{R}^n, \otimes_{i=1}^n N(\theta x^2, a^2)_{(\theta,a)\in\mathbb{R}^2\times]0;+\infty[})$

Ainsi le modèle de Y_n, cité plus haut, est régulier.

Calcul de la borne de Cramer-Rao pour θ

Pour θ on a la borne de Cramer-Rao suivante :

$$I_{\underline{X}_n}(\theta)^{-1} = \frac{a^2}{nS^2_{\underline{X}_n}}$$

Calcul de l'EMV $\hat{\theta}_n$

On a :

$$EMV\hat{\theta}_n = \frac{\sum_{i=1}^n (x_i^2 - \overline{x_n^2})(Y_n - Y_i)}{nS^2_{X_n}}$$

??? Montrer que c'est une EMM e θ???

Démonstration de $\hat{\theta}_n$ en tant qu'estimateur efficace de θ

??? Need help here ???

Identification de la loi de $\hat{\theta}_n$

On a $\hat{\theta}_n$ tel que :

$$\hat{\theta}_n \rightsquigarrow N(\theta_1, \frac{a^2}{nS^2_{\underline{X}_n}})$$

Calcul d'un intervalle de confiance (IC) de niveau $1-\alpha$ pour θ

Soit l'intervalle de confiance (IC) de $\hat{\theta}_n$ tel que :

$$IC^{1-\alpha}_{Y_{\underline{X}_n}}(\theta) = \left[\theta - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{a^2}{nS^2_{\underline{X}_n}}} ; \theta + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{a^2}{nS^2_{\underline{X}_n}}}\right]$$

TODO : Mettre à jour les intervalles avec ';' et non ','

Estimation ponctuelle du paramètre θ

Description des données

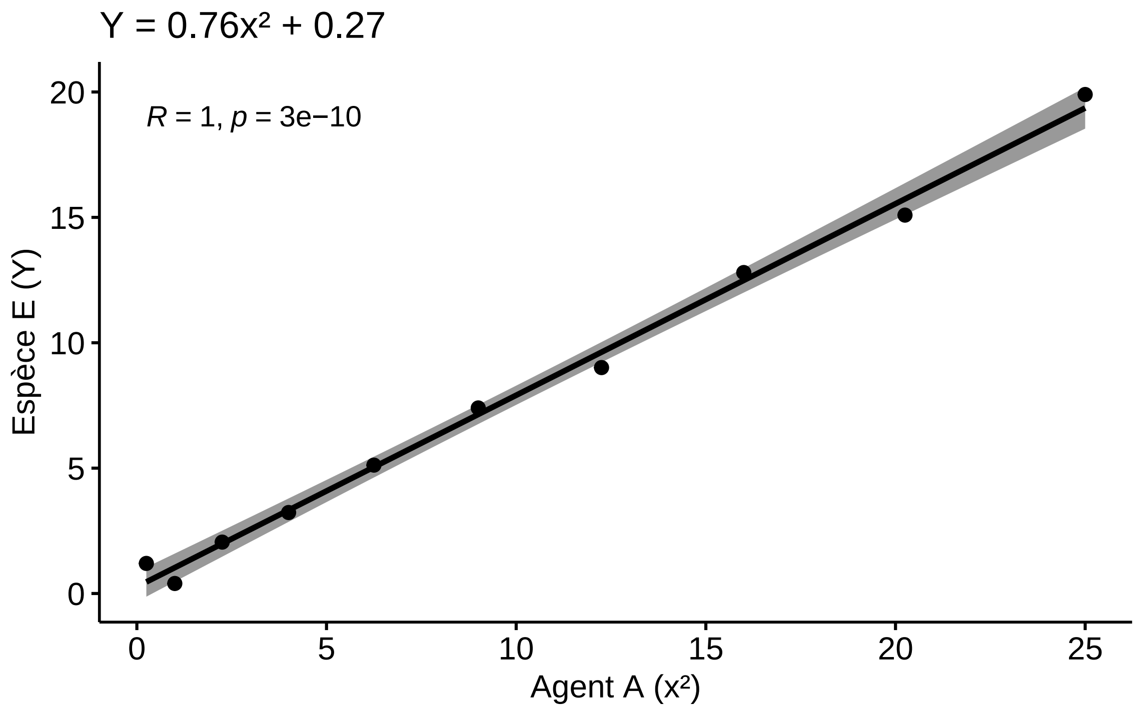
L'effectif des données est de 10.

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
```

Le graphique, ci-dessous illustre le modèle aléatoire suivant $Y = \theta x^2 + aZ$ soulignant la relation linéaire entre Y et x^2, tel que $Y = ax^2 + b$ avec \$a=\$0.76 et \$b=\$0.27.

??? supprimer les estimations a et b ci dessus ???



Calcul d'une estimation ponctuelle de θ

Il a pu être défini que :

??? Pas certain : de Yxi et Yxn ???

$$\hat{\theta}_{1n} = \frac{\sum_{i=1}^{n}(x_i^2 - \overline{x_n^2})(Y_{x_i} - \overline{Y_{\underline{X}_n}})}{nS^2_{\underline{X}_n}}$$

Ainsi à partir de l'échantillon, comme on a :

$$nS^2_{\underline{X}_n} = \sum_{i=1}^{n}(x^2 - \overline{x_n^2}) = 656,906$$

et

$$nS^2_{\underline{X}_n} = \sum_{i=1}^{n}(x^2 - \overline{x_n^2})(Y_{x_i} - \overline{Y_{\underline{X}_n}}) = 501,653$$

On peut en déduire $\hat{\theta}_{1n}$, tel que :

$$\hat{\theta}_{1n} = \theta = 0,7637$$

Par ailleurs, on obtient un écart résiduel non nul, en effet, on a :

$$\sum_{i=1}^{n}\tilde{e}_i = 2,7$$

Calcul d'un intervalle de confiance (IC) de niveau 95% pour θ

$$\left[\theta - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{a^2}{nS^2_{\underline{X}_n}}};\theta + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{a^2}{nS^2_{\underline{X}_n}}}\right] : (3)$$

Or on support que l'écart-type (a) est tel que $a = 0,5$, ainsi :

$$(3) \Leftrightarrow \left[0,7637 - 1,96 \times \sqrt{\frac{0,5^2}{656,906}};0,7637 + 1,96 \times \sqrt{\frac{0,5^2}{656,906}}\right]$$

$$(3) \Leftrightarrow [0,7254;0,8019]$$

Evaluation d'une évolution significative de θ

On ne souhaite ici évaluer une croissance ou une décroissance, de θ ; mais seulement savoir s'il y a une différence.

Ainsi, on souhaite savoir si l'on peut accepter au seuil de 5% l'hypothèse $H_1 : \theta = \theta_{pass} \Leftrightarrow H_0 : \theta = 0,9$ contre $H_1 : \theta \neq 0,9$.

Test d'hypothèses

$$W^{\alpha}_{Y_{\underline{X}_n}}(\theta) = \left\{\mid \hat{\theta}_n - 0,9 \mid > Z_{1-\frac{\alpha}{2}}\sqrt{\frac{a^2}{nS^2_{\underline{X}_n}}}\right\} : (4)$$

$$(4) \Leftrightarrow W^{\alpha}_{Y_{\underline{X}_n}}(\theta) = \left\{\mid \hat{\theta}_n - 0,9 \mid > 1,96 \times \sqrt{\frac{0,5^2}{656,906}}\right\}$$

$$(4) \Leftrightarrow W^{\alpha}_{Y_{\underline{X}_n}}(\theta) = \left\{\mid \hat{\theta}_n - 0,9 \mid > 0,038\right\}$$

Observations On observe :

$$\mid \hat{\theta}_n - 0,9 \mid = \mid 0,7637 - 0,9 \mid = \mid -0.1363 \mid = 0.1363 : (5)$$
$$(5) \Leftrightarrow \mid \hat{\theta}_n - 0,9 \mid \in W^{\alpha}_{Y_{\underline{X}_n}}$$

Décision On accepte H_1, au seuil de 5%.

Conlcusion On en conclue que la valeur du paramètre θ a évolu significativement par rapport aux mesures précédentes.