

*Ecole Hassania
des Travaux Publics*



المدرسة الحسنية للأشغال العمومية

Département de Mathématiques et Informatique

STATISTIQUE

Exercices Corrigés

Abdelhamid El Mossadeq

Professeur à l'EHTP

2006-2007

TABLE DES MATIERES

Structures Statistiques et Estimation	1
Les Procédures Usuelles des Tests d'Hypothèses :	
1. Les Fréquences	45
Les Procédures Usuelles des Tests d'Hypothèses :	
2. Les Tests du Khi-Deux	61
Les Procédures Usuelles des Tests d'Hypothèses :	
3. Moyennes et Variances	95

Structure Statistique et Estimation

Exercice 1

Déterminer et étudier les propriétés de l'estimateur du maximum de vraisemblance d'un r -échantillon pour :

1. le paramètre p d'une loi de *Bernouilli*
2. le paramètre p d'une loi *géométrique*
3. le paramètre p d'une loi *binomiale* d'ordre n
4. le paramètre α d'une loi de *Poisson*
5. le paramètre λ d'une loi *exponentielle*
6. les paramètres μ et σ^2 d'une loi *normale*
7. le paramètre θ d'une loi *uniforme* sur l'intervalle $[0, \theta]$

Solution 1

1. Soit X une variable aléatoire de Bernouilli de paramètre p .

Pour tout $x \in \{0, 1\}$, la probabilité élémentaire $p(x)$ de x est :

$$p(x) = p^x (1 - p)^{1-x}$$

de plus :

$$\begin{cases} E[X] = p \\ V[X] = p(1 - p) \end{cases}$$

- (a) **Recherche du maximum de vraisemblance :**

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $p \in [0, 1]$ et tout $(x_1, \dots, x_r) \in \{0, 1\}^r$ par :

$$\begin{aligned} \mathcal{L}(p; x_1, \dots, x_r) &= \prod_{i=1}^r p(x_i) \\ &= p^{\sum_{i=1}^r x_i} (1 - p)^{r - \sum_{i=1}^r x_i} \end{aligned}$$

d'où :

$$\ln \mathcal{L}(p; x_1, \dots, x_r) = \left(\sum_{i=1}^r x_i \right) \ln p + \left(r - \sum_{i=1}^r x_i \right) \ln (1 - p)$$

Il en résulte que :

$$\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) = \frac{\sum_{i=1}^r x_i}{p} - \frac{r - \sum_{i=1}^r x_i}{1-p}$$

d'où :

$$\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) = 0 \implies p = \frac{1}{r} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial p^2} \ln \mathcal{L}(p; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure de Bernoulli est :

$$\hat{p} = \frac{1}{r} \sum_{i=1}^r X_i$$

C'est la fréquence empirique du r -échantillon.

(b) **Etude des propriétés de \hat{p} :**

Puisque :

$$\begin{aligned} E[\hat{p}] &= E[X] \\ &= p \end{aligned}$$

et :

$$\begin{aligned} V[\hat{p}] &= \frac{V[X]}{r} \\ &= \frac{p(1-p)}{r} \end{aligned}$$

On en déduit que \hat{p} est un estimateur sans biais et convergent du paramètre p d'une loi de Bernoulli.

2. Soit X une variable aléatoire de géométrique de paramètre p .

Pour tout $x \in \mathbb{N}^*$, la probabilité élémentaire $p(x)$ de x est :

$$p(x) = p(1-p)^{x-1}$$

de plus :

$$\begin{cases} E[X] = \frac{1}{p} \\ V[X] = \frac{1-p}{p^2} \end{cases}$$

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $p \in [0, 1]$ et tout $(x_1, \dots, x_r) \in (\mathbb{N}^*)^r$ par :

$$\begin{aligned}\mathcal{L}(p; x_1, \dots, x_r) &= \prod_{i=1}^r p(x_i) \\ &= p^r (1-p)^{\sum_{i=1}^r x_i - r}\end{aligned}$$

d'où :

$$\ln \mathcal{L}(p; x_1, \dots, x_r) = r \ln p + \left(\sum_{i=1}^r x_i - r \right) \ln (1-p)$$

Il en résulte que :

$$\begin{aligned}\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) &= \frac{r}{p} - \frac{\sum_{i=1}^r x_i - r}{1-p} \\ &= \frac{r - p \sum_{i=1}^r x_i}{p(1-p)}\end{aligned}$$

d'où :

$$\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) = 0 \implies p = \frac{r}{\sum_{i=1}^r x_i}$$

et comme :

$$\frac{\partial^2}{\partial p^2} \ln \mathcal{L}(p; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure géométrique est :

$$\hat{p} = \frac{r}{\sum_{i=1}^r X_i}$$

C'est l'inverse de la moyenne empirique du r -échantillon.

3. Soit X une variable aléatoire binomiale d'ordre n et de paramètre p .
pour tout $x \in \{0, 1, \dots, n\}$, la probabilité élémentaire $p(x)$ de x est :

$$p(x) = C(n, x) p^x (1-p)^{n-x}$$

de plus :

$$\begin{cases} E[X] = np \\ V[X] = np(1-p) \end{cases}$$

(a) **Recherche du maximum de vraisemblance :**

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $p \in [0, 1]$ et tout $(x_1, \dots, x_r) \in \{0, 1, \dots, n\}^r$ par :

$$\begin{aligned} \mathcal{L}(p; x_1, \dots, x_r) &= \prod_{i=1}^r p(x_i) \\ &= \left[\prod_{i=1}^r C(n, x_i) \right] p^{\sum_{i=1}^r x_i} (1-p)^{rn - \sum_{i=1}^r x_i} \end{aligned}$$

d'où :

$$\ln \mathcal{L}(p; x_1, \dots, x_r) = \ln \prod_{i=1}^r C(n, x_i) + \sum_{i=1}^r x_i \ln p + \left(rn - \sum_{i=1}^r x_i \right) \ln(1-p)$$

Il en résulte que :

$$\begin{aligned} \frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) &= \frac{\sum_{i=1}^r x_i}{p} - \frac{rn - \sum_{i=1}^r x_i}{1-p} \\ &= \frac{\sum_{i=1}^r x_i - rnp}{p(1-p)} \end{aligned}$$

d'où :

$$\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, \dots, x_r) = 0 \implies p = \frac{1}{rn} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial p^2} \ln \mathcal{L}(p; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure de binomiale est :

$$\hat{p} = \frac{1}{rn} \sum_{i=1}^r X_i$$

(b) **Etude des propriétés de \hat{p} :**

Puisque :

$$\begin{aligned} E[\hat{p}] &= \frac{1}{n} E[X] \\ &= p \end{aligned}$$

et :

$$\begin{aligned} V[\hat{p}] &= \frac{V[X]}{rn^2} \\ &= \frac{p(1-p)}{rn} \end{aligned}$$

on en déduit que \hat{p} est un estimateur sans biais et convergent de p .4. Soit X une variable aléatoire de Poisson de paramètre α .Pour tout $x \in \mathbb{N}$, la probabilité élémentaire $p(x)$ de x est :

$$p(x) = \frac{\alpha^x}{x!} \exp -\alpha$$

de plus :

$$\begin{cases} E[X] = \alpha \\ V[X] = \alpha \end{cases}$$

(a) **Recherche du maximum de vraisemblance :**Considérons un r -échantillon de cette structure.Sa fonction de vraisemblance est définie pour tout α , $\alpha > 0$, et tout $(x_1, \dots, x_r) \in \mathbb{N}^r$ par :

$$\begin{aligned} \mathcal{L}(\alpha; x_1, \dots, x_r) &= \prod_{i=1}^r p(x_i) \\ &= \frac{\alpha^{\sum_{i=1}^r x_i}}{x_1! \dots x_r!} \exp -r\alpha \end{aligned}$$

d'où :

$$\ln \mathcal{L}(\alpha; x_1, \dots, x_r) = -\ln(x_1! \dots x_r!) + \sum_{i=1}^r x_i \ln \alpha - r\alpha$$

Il en résulte que :

$$\frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) = \frac{\sum_{i=1}^r x_i}{\alpha} - r$$

d'où :

$$\frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) = 0 \implies p = \frac{1}{r} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial \alpha^2} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure de Poisson est :

$$\hat{\alpha} = \frac{1}{r} \sum_{i=1}^r X_i$$

C'est la moyenne empirique du r -échantillon.

(b) **Etude des propriétés de $\hat{\alpha}$:**

Puisque :

$$\begin{aligned} E[\hat{\alpha}] &= E[X] \\ &= \alpha \end{aligned}$$

et :

$$\begin{aligned} V[\hat{\alpha}] &= \frac{V[X]}{r} \\ &= \frac{\alpha}{r} \end{aligned}$$

On en déduit que $\hat{\alpha}$ est un estimateur sans biais et convergent de α .

5. Soit X une variable aléatoire exponentielle de paramètre λ .

Sa densité de probabilité f est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \lambda \exp -\lambda x & \text{si } x > 0 \end{cases}$$

de plus :

$$\begin{cases} E[X] = \frac{1}{\lambda} \\ V[X] = \frac{1}{\lambda^2} \end{cases}$$

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout λ , $\lambda > 0$, et tout (x_1, \dots, x_r) dans \mathbb{R}^r , tous strictement positifs, par :

$$\begin{aligned}\mathcal{L}(\lambda; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \lambda^r \exp -\lambda \sum_{i=1}^r x_i\end{aligned}$$

d'où :

$$\ln \mathcal{L}(\lambda; x_1, \dots, x_r) = r \ln \lambda - \lambda \sum_{i=1}^r x_i$$

Il en résulte que :

$$\frac{\partial}{\partial \lambda} \ln \mathcal{L}(\lambda; x_1, \dots, x_r) = \frac{r}{\lambda} - \sum_{i=1}^r x_i$$

d'où :

$$\frac{\partial}{\partial \lambda} \ln \mathcal{L}(\lambda; x_1, \dots, x_r) = 0 \implies \lambda = \frac{r}{\sum_{i=1}^r x_i}$$

et comme :

$$\frac{\partial^2}{\partial \lambda^2} \ln \mathcal{L}(\lambda; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure exponentielle est :

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^r X_i}$$

C'est l'inverse de la moyenne empirique du r -échantillon.

6. Soit X une variable aléatoire normale de paramètres μ et σ^2 .
Sa densité de probabilité f est définie pour tout $x \in \mathbb{R}$ par :

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

de plus :

$$\begin{cases} E[X] = \mu \\ V[X] = \sigma^2 \end{cases}$$

(a) **Recherche du maximum de vraisemblance :**

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\mu \in \mathbb{R}$, tout $\sigma > 0$ et tout $(x_1, \dots, x_r) \in \mathbb{R}^r$ par :

$$\begin{aligned}\mathcal{L}(\mu, \sigma; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{(\sigma\sqrt{2\pi})^r} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^r (x_i - \mu)^2\end{aligned}$$

d'où :

$$\ln \mathcal{L}(\mu, \sigma; x_1, \dots, x_r) = -r \ln \sqrt{2\pi} - r \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^r (x_i - \mu)^2$$

Il en résulte que :

$$\begin{cases} \frac{\partial}{\partial \mu} \mathcal{L}(\mu, \sigma; x_1, \dots, x_r) = \frac{1}{\sigma^2} \sum_{i=1}^r (x_i - \mu) \\ \frac{\partial}{\partial \sigma} \mathcal{L}(\mu, \sigma; x_1, \dots, x_r) = -\frac{r}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^r (x_i - \mu)^2 \end{cases}$$

d'où :

$$\begin{cases} \frac{\partial}{\partial \mu} \mathcal{L}(\mu, \sigma; x_1, \dots, x_r) = 0 \\ \frac{\partial}{\partial \sigma} \mathcal{L}(\mu, \sigma; x_1, \dots, x_r) = 0 \end{cases} \implies \begin{cases} \mu = \frac{1}{r} \sum_{i=1}^r x_i \\ \sigma^2 = \frac{1}{r} \sum_{i=1}^r (x_i - \mu)^2 \end{cases}$$

Donc les estimateurs du maximum de vraisemblance d'un r -échantillon d'une structure normale est :

$$\begin{cases} \hat{\mu} = \frac{1}{r} \sum_{i=1}^r X_i \\ \hat{\sigma}^2 = \frac{1}{r} \sum_{i=1}^r (X_i - \hat{\mu})^2 \end{cases}$$

(b) **Etude des propriétés de $\hat{\mu}$ et $\hat{\sigma}$:**

On a :

$$\begin{aligned} E[\hat{\mu}] &= E[X] \\ &= \mu \end{aligned}$$

et :

$$\begin{aligned} E[\hat{\sigma}^2] &= \frac{r-1}{r} V[X] \\ &= \frac{r-1}{r} \sigma^2 \end{aligned}$$

On en déduit que $\hat{\mu}$ est un estimateur sans biais et convergent de μ , mais $\hat{\sigma}$ est un estimateur biaisé de σ .

7. Soit X une variable aléatoire uniforme sur l'intervalle $[0, \theta]$.

Sa densité de probabilité f est définie pour tout $x \in [0, \theta]$ par :

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0, \theta] \\ 0 & \text{si } x \notin [0, \theta] \end{cases}$$

de plus :

$$\begin{cases} E[X] = \frac{\theta}{2} \\ V[X] = \frac{\theta^2}{12} \end{cases}$$

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\theta, \theta > 0$, et tout $(x_1, \dots, x_r) \in [0, \theta]^r$:

$$\begin{aligned} \mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{\theta^r} \end{aligned}$$

La fonction :

$$\theta \longrightarrow \mathcal{L}(\theta; x_1, \dots, x_r)$$

est strictement décroissante, donc elle atteint son maximum lorsque θ est minimum.

Et comme :

$$\forall i \in \{1, \dots, r\} : \theta \geq x_i$$

donc θ est minimum lorsque :

$$\theta = \max(x_1, \dots, x_r)$$

Donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure uniforme est :

$$\hat{\theta} = \max(X_1, \dots, X_r)$$

Exercice 2

Soit X une variable aléatoire dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} \frac{1}{\theta} \exp -\frac{x}{\theta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

où θ est un paramètre réel strictement positif.

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ d'un r -échantillon de variable parente X .
2. $\hat{\theta}$ est-il un résumé exhaustif ?
3. Calculer l'espérance mathématique et la variance de $\hat{\theta}$.
Que peut-on conclure ?
4. Calculer la quantité d'information de *Fisher*.
En déduire que $\hat{\theta}$ est efficace.

Solution 2

Soit X une variable aléatoire exponentielle dont la densité de probabilité f est définie pour tout x , $x > 0$, par :

$$f(x) = \begin{cases} \frac{1}{\theta} \exp -\frac{x}{\theta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

où θ est un paramètre réel strictement positif.

On a :

$$\begin{cases} E[X] = \theta \\ V[X] = \theta^2 \end{cases}$$

1. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\theta, \theta > 0$, et tout $(x_1, \dots, x_r) \in \mathbb{R}^r$, tous strictement positifs, par :

$$\begin{aligned}\mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{\theta^r} \exp - \frac{\sum_{i=1}^r x_i}{\theta}\end{aligned}$$

d'où :

$$\ln \mathcal{L}(\theta; x_1, \dots, x_r) = -r \ln \theta - \frac{\sum_{i=1}^r x_i}{\theta}$$

Il en résulte que :

$$\frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; x_1, \dots, x_r) = -\frac{r}{\theta} + \frac{\sum_{i=1}^r x_i}{\theta^2}$$

d'où :

$$\frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; x_1, \dots, x_r) = 0 \implies \theta = \frac{1}{r} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}(\theta; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure exponentielle est :

$$\hat{\theta} = \frac{1}{r} \sum_{i=1}^r X_i$$

C'est la moyenne empirique du r -échantillon.

2. Pour tout $\theta, \theta > 0$, et tout $(x_1, \dots, x_r) \in \mathbb{R}^r$, tous strictement positifs, on a :

$$\begin{aligned}\mathcal{L}(\theta; x_1, \dots, x_r) &= \frac{1}{\theta^r} \exp - \frac{\sum_{i=1}^r x_i}{\theta} \\ &= \frac{1}{\theta^r} \exp - \frac{r \hat{\theta}(x_1, \dots, x_r)}{\theta}\end{aligned}$$

D'après le théorème de factorisation, $\hat{\theta}$ est un résumé exhaustif puisque :

$$\mathcal{L}(\theta; x_1, \dots, x_r) = g\left(\theta; \hat{\theta}(x_1, \dots, x_r)\right) h(x_1, \dots, x_r)$$

où :

$$g\left(\theta; \hat{\theta}(x_1, \dots, x_r)\right) = \frac{1}{\theta^r} \exp - \frac{r\hat{\theta}(x_1, \dots, x_r)}{\theta}$$

et :

$$h(x_1, \dots, x_r) = 1$$

3. Comme :

$$\hat{\theta} = \frac{1}{r} \sum_{i=1}^r X_i$$

alors :

$$\begin{aligned} E[\hat{\theta}] &= E[X] \\ &= \theta \end{aligned}$$

et :

$$\begin{aligned} V[\hat{\theta}] &= \frac{V[X]}{r} \\ &= \frac{\theta^2}{r} \end{aligned}$$

On en déduit que $\hat{\theta}$ est un estimateur sans biais et convergent de θ .

4. Calculons la quantité d'information de Fisher, $I[X, \theta]$, concernant θ .

On a :

$$\begin{aligned} I[X, \theta] &= -E\left[\frac{\partial^2}{\partial \theta^2} \ln f(\theta, X)\right] \\ &= -E\left[\frac{\partial^2}{\partial \theta^2} \left(-\ln \theta - \frac{X}{\theta}\right)\right] \\ &= E\left[-\frac{1}{\theta^2} + \frac{2X}{\theta^3}\right] \\ &= \frac{1}{\theta^2} \end{aligned}$$

Donc la quantité d'information de Fisher, $I[X_1, \dots, X_r, \theta]$, concernant θ fournie par le r -échantillon est :

$$\begin{aligned} I[X_1, \dots, X_r, \theta] &= rI[X, \theta] \\ &= \frac{r}{\theta^2} \end{aligned}$$

Calculons l'efficacité $e[\hat{\theta}]$ de $\hat{\theta}$.

On a :

$$\begin{aligned} e[\hat{\theta}] &= \frac{1}{I[X_1, \dots, X_r, \theta] V[\hat{\theta}]} \\ &= 1 \end{aligned}$$

donc, $\hat{\theta}$ est efficace.

Exercice 3

Soit X une variable aléatoire dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{\lambda}{\theta^k} x^{k-1} \exp -\frac{x}{\theta} & \text{si } x > 0 \end{cases}$$

où θ est un paramètre réel strictement positif, k un entier naturel non nul et λ une constante réel.

1. Déterminer la constante λ .
2. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ d'un r -échantillon de variable parente X .
3. $\hat{\theta}$ est-il un résumé exhaustif ?
4. Calculer l'espérance mathématique et la variance de $\hat{\theta}$.
Que peut-on conclure ?
5. Calculer la quantité d'information de *Fisher*.
En déduire que $\hat{\theta}$ est efficace.

Solution 3

La densité de probabilité de la variable aléatoire X est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{\lambda}{\theta^k} x^{k-1} \exp -\frac{x}{\theta} & \text{si } x > 0 \end{cases}$$

Rappelons que pour tout $k \in \mathbb{N}$:

$$\int_0^{+\infty} u^k \exp -u du = k!$$

1. Ainsi :

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x) dx &= \int_0^{+\infty} \frac{\lambda}{\theta^k} x^{k-1} \exp -\frac{x}{\theta} dx \\ &= \int_0^{+\infty} \lambda u^{k-1} \exp -u du \\ &= \lambda (k-1)!\end{aligned}$$

d'où

$$\lambda = \frac{1}{(k-1)!}$$

puisque :

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

De plus :

$$\begin{aligned}E[X] &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \int_0^{+\infty} \frac{1}{(k-1)!\theta^k} x^k \exp -\frac{x}{\theta} dx \\ &= k\theta\end{aligned}$$

et :

$$\begin{aligned}E[X^2] &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\ &= \int_0^{+\infty} \frac{1}{(k-1)!\theta^k} x^{k+1} \exp -\frac{x}{\theta} dx \\ &= k(k+1)\theta^2\end{aligned}$$

d'où :

$$\begin{aligned}V[X] &= E[X^2] - E[X]^2 \\ &= k\theta^2\end{aligned}$$

2. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\theta, \theta > 0$, et tout $(x_1, \dots, x_r) \in \mathbb{R}^r$, tous strictement positifs, par :

$$\begin{aligned}
\mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\
&= \frac{1}{[(k-1)!]^r \theta^{rk}} (x_1 \dots x_r)^{k-1} \exp - \frac{\sum_{i=1}^r x_i}{\theta}
\end{aligned}$$

d'où :

$$\ln \mathcal{L}(\theta; x_1, \dots, x_r) = -r \ln(k-1)! - \ln(x_1 \dots x_r)^{k-1} - rk \ln \theta - \frac{\sum_{i=1}^r x_i}{\theta}$$

Il en résulte que :

$$\frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; x_1, \dots, x_r) = -\frac{rk}{\theta} + \frac{\sum_{i=1}^r x_i}{\theta^2}$$

d'où :

$$\frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta; x_1, \dots, x_r) = 0 \implies \theta = \frac{1}{rk} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}(\theta; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\hat{\theta} = \frac{1}{rk} \sum_{i=1}^r X_i$$

3. Pour tout θ , $\theta > 0$, et tout $(x_1, \dots, x_r) \in \mathbb{R}^r$, tous strictement positifs, on a :

$$\begin{aligned}
\mathcal{L}(\theta; x_1, \dots, x_r) &= \frac{1}{[(k-1)!]^r \theta^{rk}} (x_1 \dots x_r)^{k-1} \exp - \frac{\sum_{i=1}^r x_i}{\theta} \\
&= \frac{1}{[(k-1)!]^r \theta^{rk}} (x_1 \dots x_r)^{k-1} \exp - \frac{rk \hat{\theta}(x_1, \dots, x_r)}{\theta}
\end{aligned}$$

D'après le théorème de factorisation, $\hat{\theta}$ est un résumé exhaustif puisque :

$$\mathcal{L}(\theta; x_1, \dots, x_r) = g\left(\theta; \hat{\theta}(x_1, \dots, x_r)\right) h(x_1, \dots, x_r)$$

où :

$$g\left(\theta; \hat{\theta}(x_1, \dots, x_r)\right) = \frac{1}{\theta^{rk}} \exp - \frac{rk\hat{\theta}(x_1, \dots, x_r)}{\theta}$$

et :

$$h(x_1, \dots, x_r) = \frac{1}{[(k-1)!]^r} (x_1 \dots x_r)^{k-1}$$

4. Puisque :

$$\hat{\theta} = \frac{1}{rk} \sum_{i=1}^r X_i$$

alors :

$$E[\hat{\theta}] = \frac{1}{k} E[X] = \theta$$

et :

$$V[\hat{\theta}] = \frac{V[X]}{rk^2} = \frac{\theta^2}{rk}$$

On en déduit que $\hat{\theta}$ est un estimateur sans biais et convergent de θ .

5. Calculons la quantité d'information de *Fisher*, $I[X, \theta]$, concernant θ .

On a :

$$\begin{aligned} I[X, \theta] &= -E \left[\frac{\partial^2}{\partial \theta^2} \ln f(\theta, X) \right] \\ &= -E \left[\frac{\partial^2}{\partial \theta^2} \left(-\ln(k-1)! + (k-1) \ln X - k \ln \theta - \frac{X}{\theta} \right) \right] \\ &= E \left[-\frac{k}{\theta^2} + \frac{2X}{\theta^3} \right] \\ &= \frac{k}{\theta^2} \end{aligned}$$

Donc la quantité d'information de *Fisher*, $I[X_1, \dots, X_r, \theta]$, concernant θ fournie par le r -échantillon est :

$$\begin{aligned} I[X_1, \dots, X_r, \theta] &= rI[X, \theta] \\ &= \frac{rk}{\theta^2} \end{aligned}$$

Calculons l'efficacité $e[\hat{\theta}]$ de $\hat{\theta}$.

On a :

$$e[\hat{\theta}] = \frac{1}{I[X_1, \dots, X_r, \theta] V[\hat{\theta}]} = 1$$

donc, $\hat{\theta}$ est efficace.

Exercice 4

Soit X une variable aléatoire dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \notin [0, \theta] \\ \frac{1}{\theta} & \text{si } x \in [0, \theta] \end{cases}$$

où θ est un paramètre réel.

1. Déterminer la fonction de répartition de X .
2. Calculer la quantité d'information de Fisher.
3. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ d'un n -échantillon de variable parente X .
4. Calculer l'espérance mathématique et la variance de $\hat{\theta}$.
Que peut-on conclure ?
5. Dans le cas où $\hat{\theta}$ est biaisé, proposer un estimateur sans biais de θ .

Solution 4

1. La fonction de répartition F de X est définie pour tout $x \in \mathbb{R}$ par :

$$F(x) = \int_{-\infty}^x f(t) dt$$

d'où :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{\theta} & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

de plus :

$$\begin{cases} E[X] = \frac{\theta}{2} \\ V[X] = \frac{\theta^2}{12} \end{cases}$$

2. Puisque le domaine D_θ :

$$\begin{aligned} D_\theta &= \{x \in \mathbb{R} \mid f(x) > 0\} \\ &= [0, \theta] \end{aligned}$$

dépend de θ , donc la quantité d'information de *Fisher* n'existe pas.

3. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout θ , $\theta > 0$, et tout $(x_1, \dots, x_r) \in [0, \theta]^r$:

$$\begin{aligned}\mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{\theta^r}\end{aligned}$$

La fonction :

$$\theta \longrightarrow \mathcal{L}(\theta; x_1, \dots, x_r)$$

est strictement décroissante, donc elle atteint son maximum lorsque θ est minimum.

Et comme :

$$\forall i \in \{1, \dots, r\} : \theta \geq x_i$$

Il en résulte que θ est minimum lorsque :

$$\theta = \max(x_1, \dots, x_r)$$

Donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure uniforme est :

$$\hat{\theta} = \max(X_1, \dots, X_r)$$

4. Pour déterminer la densité de probabilité de $\hat{\theta}$, commençons d'abord par calculer sa fonction de répartition.

(a) **Fonction de répartition de $\hat{\theta}$:**

Pour tout $u \in \mathbb{R}$ on a :

$$\begin{aligned}
 F_{\hat{\theta}}(u) &= P[\hat{\theta} < u] \\
 &= P[\max(X_1, \dots, X_r) < u] \\
 &= P[X_1 < u, \dots, X_r < u] \\
 &= \prod_{k=1}^r P[X_k < u] \\
 &= [F(u)]^r \\
 &= \begin{cases} 0 & si \quad u \leq 0 \\ \left(\frac{u}{\theta}\right)^r & si \quad 0 \leq u \leq \theta \\ 1 & si \quad u \geq \theta \end{cases}
 \end{aligned}$$

(b) **Densité de probabilité de $\hat{\theta}$:**

Pour tout $u \in \mathbb{R} - \{0, \theta\}$ on a :

$$\begin{aligned} f_{\hat{\theta}}(u) &= \frac{d}{du} F_{\hat{\theta}}(u) \\ &= \begin{cases} 0 & \text{si } u \notin]0, \theta[\\ r \frac{u^{r-1}}{\theta^r} & \text{si } u \in]0, \theta[\end{cases} \end{aligned}$$

(c) **Espérance mathématique de $\hat{\theta}$:**

$$\begin{aligned} E[\hat{\theta}] &= \int_{\mathbb{R}} u f_{\hat{\theta}}(u) du \\ &= \int_0^{\theta} r \frac{u^r}{\theta^r} du \\ &= \frac{r}{r+1} \theta \end{aligned}$$

(d) **Espérance mathématique de $\hat{\theta}^2$:**

$$\begin{aligned} E[\hat{\theta}^2] &= \int_{\mathbb{R}} u^2 f_{\hat{\theta}}(u) du \\ &= \int_0^{\theta} r \frac{u^{r+1}}{\theta^r} du \\ &= \frac{r}{r+2} \theta^2 \end{aligned}$$

(e) **Variance de $\hat{\theta}$:**

$$\begin{aligned} V[\hat{\theta}] &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 \\ &= \frac{r}{(r+1)^2 (r+2)} \theta^2 \end{aligned}$$

L'estimateur $\hat{\theta}$ de θ est biaisé, mais il est asymptotiquement sans biais.

5. Considérons l'estimateur :

$$T = \frac{r+1}{r} \hat{\theta}$$

Alors :

$$\begin{aligned} E[T] &= \frac{r+1}{r} E[\hat{\theta}] \\ &= \theta \end{aligned}$$

et :

$$\begin{aligned} V[T] &= \left(\frac{r+1}{r} \right)^2 V[\hat{\theta}] \\ &= \frac{1}{r(r+2)} \theta^2 \end{aligned}$$

T est donc un estimateur sans biais et convergent de θ .

Exercice 5

Soit X une variable aléatoire dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < \theta \\ \exp(\theta - x) & \text{si } x \geq \theta \end{cases}$$

où θ est un paramètre réel.

1. Déterminer la fonction de répartition de X .
2. Calculer la quantité d'information de Fisher.
3. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ d'un r -échantillon de variable parente X .
4. Calculer l'espérance mathématique et la variance de $\hat{\theta}$.
Que peut-on conclure ?
5. Dans le cas où $\hat{\theta}$ est biaisé, proposer un estimateur sans biais de θ .

Solution 5

1. La fonction de répartition F de X est définie pour tout $x \in \mathbb{R}$ par :

$$F(x) = \int_{-\infty}^x f(t) dt$$

d'où :

$$F(x) = \begin{cases} 0 & \text{si } x \leq \theta \\ 1 - \exp(\theta - x) & \text{si } x \geq \theta \end{cases}$$

de plus :

$$\begin{cases} E[X] = \theta + 1 \\ V[X] = 1 \end{cases}$$

2. Puisque le domaine D_θ :

$$\begin{aligned} D_\theta &= \{x \in \mathbb{R} \mid f(x) > 0\} \\ &= [\theta, +\infty[\end{aligned}$$

dépend de θ , donc la quantité d'information de *Fisher* n'existe pas.

3. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\theta \in \mathbb{R}$, et tout $(x_1, \dots, x_r) \in ([\theta, +\infty[)^r$:

$$\begin{aligned} \mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \exp \sum_{i=1}^r (\theta - x_i) \end{aligned}$$

La fonction :

$$\theta \longrightarrow \mathcal{L}(\theta; x_1, \dots, x_r)$$

est strictement croissante, donc elle atteint son maximum lorsque θ est maximum.

Et comme :

$$\forall i \in \{1, \dots, r\} : \theta \leq x_i$$

Il en résulte que θ est maximum lorsque :

$$\theta = \min(x_1, \dots, x_r)$$

Donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\hat{\theta} = \min(X_1, \dots, X_r)$$

4. Pour déterminer la densité de probabilité de $\hat{\theta}$, commençons d'abord par calculer sa fonction de répartition.

(a) **Fonction de répartition de $\hat{\theta}$:**

Pour tout $v \in \mathbb{R}$ on a :

$$\begin{aligned}
 F_{\hat{\theta}}(v) &= P[\hat{\theta} < v] \\
 &= P[\min(X_1, \dots, X_r) < v] \\
 &= 1 - P[\min(X_1, \dots, X_r) \geq v] \\
 &= 1 - P[X_1 \geq v, \dots, X_r \geq v] \\
 &= 1 - \prod_{k=1}^r P[X_k \geq v] \\
 &= 1 - \prod_{k=1}^r (1 - P[X_k < v]) \\
 &= 1 - [1 - F(v)]^r \\
 &= \begin{cases} 0 & \text{si } v \leq \theta \\ 1 - \exp r(\theta - v) & \text{si } v \geq \theta \end{cases}
 \end{aligned}$$

(b) **Densité de probabilité de $\hat{\theta}$:**

Pour tout $u \in \mathbb{R} - \{\theta\}$ on a :

$$\begin{aligned}
 f_{\hat{\theta}}(v) &= \frac{d}{dv} F_{\hat{\theta}}(v) \\
 &= \begin{cases} 0 & \text{si } v < \theta \\ r \exp r(\theta - v) & \text{si } v > \theta \end{cases}
 \end{aligned}$$

(c) **Espérance mathématique de $\hat{\theta}$:**

$$\begin{aligned}
 E[\hat{\theta}] &= \int_{\mathbb{R}} v f_{\hat{\theta}}(v) dv \\
 &= \int_{\theta}^{+\infty} r v \exp r(\theta - v) dv \\
 &= \theta + \frac{1}{r}
 \end{aligned}$$

(d) **Espérance mathématique de $\hat{\theta}^2$:**

$$\begin{aligned} E \left[\hat{\theta}^2 \right] &= \int_{\mathbb{R}} v^2 f_{\hat{\theta}}(v) dv \\ &= \int_{\theta}^{+\infty} r v^2 \exp r (\theta - v) dv \\ &= \left(\theta + \frac{1}{r} \right)^2 + \frac{1}{r^2} \end{aligned}$$

(e) **Variance de $\hat{\theta}$:**

$$V \left[\hat{\theta} \right] = E \left[\hat{\theta}^2 \right] - E \left[\hat{\theta} \right]^2 = \frac{1}{r^2}$$

L'estimateur $\hat{\theta}$ de θ est biaisé, mais il est asymptotiquement sans biais.

5. Considérons l'estimateur :

$$T = \hat{\theta} - \frac{1}{r}$$

Alors :

$$E [T] = E \left[\hat{\theta} \right] - \frac{1}{r} = \theta$$

et :

$$V [T] = V \left[\hat{\theta} \right] = \frac{1}{r^2}$$

T est donc un estimateur sans biais et convergent de θ .

Exercice 6

Les éléments d'une population possèdent un caractère X qui suit une loi de *Poisson* de paramètre inconnu α .

Une suite de r expériences a fourni les valeurs k_1, \dots, k_r .

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\alpha}$ de α et étudier les propriétés de cet estimateur.
2. $\hat{\alpha}$ est-il un résumé exhaustif ?
3. On désire estimer la quantité :

$$\delta = P [X = 0]$$

Déterminer l'estimateur du maximum de vraisemblance $\hat{\delta}$ de δ .

Que remarquez-vous ?

Solution 6

1. Soit X une variable aléatoire de Poisson de paramètre α .
pour tout $x \in \mathbb{N}$, la probabilité élémentaire $p(x)$ de x est :

$$p(x) = \frac{\alpha^x}{x!} \exp -\alpha$$

de plus :

$$\begin{cases} E[X] = \alpha \\ V[X] = \alpha \end{cases}$$

(a) Recherche du maximum de vraisemblance :

Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout α , $\alpha > 0$, et tout $(k_1, \dots, k_r) \in \mathbb{N}^r$ par :

$$\begin{aligned} \mathcal{L}(\alpha; k_1, \dots, k_r) &= \prod_{i=1}^r p(k_i) \\ &= \frac{\alpha^{\sum_{i=1}^r k_i}}{k_1! \dots k_r!} \exp -r\alpha \end{aligned}$$

d'où :

$$\ln \mathcal{L}(\alpha; k_1, \dots, k_r) = -\ln(k_1! \dots k_r!) + \sum_{i=1}^r k_i \ln \alpha - r\alpha$$

Il en résulte que :

$$\frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; k_1, \dots, k_r) = \frac{\sum_{i=1}^r k_i}{\alpha} - r$$

d'où :

$$\frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; k_1, \dots, k_r) = 0 \implies p = \frac{1}{r} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial \alpha^2} \ln \mathcal{L}(\alpha; k_1, \dots, k_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure de Poisson est :

$$\hat{\alpha} = \frac{1}{r} \sum_{i=1}^r X_i$$

C'est la moyenne empirique du r -échantillon.

(b) **Etude des propriétés de $\hat{\alpha}$:**

Puisque :

$$\begin{aligned} E[\hat{\alpha}] &= E[X] \\ &= \alpha \end{aligned}$$

et :

$$\begin{aligned} V[\hat{\alpha}] &= \frac{V[X]}{r} \\ &= \frac{\alpha}{r} \end{aligned}$$

On en déduit que $\hat{\alpha}$ est un estimateur sans biais et convergent de α .

2. Pour tout α , $\alpha > 0$, et tout $(k_1, \dots, k_r) \in \mathbb{N}^r$ on a :

$$\mathcal{L}(\alpha; x_1, \dots, x_r) = \frac{\alpha^{r\hat{\alpha}(k_1, \dots, k_r)}}{x_1! \dots x_r!} \exp -r\alpha$$

D'après le théorème de factorisation, $\hat{\theta}$ est un résumé exhaustif puisque :

$$\mathcal{L}(\alpha; x_1, \dots, x_r) = g(\alpha; \hat{\alpha}(x_1, \dots, x_r)) h(x_1, \dots, x_r)$$

où :

$$g(\alpha; \hat{\alpha}(x_1, \dots, x_r)) = \alpha^{r\hat{\alpha}(k_1, \dots, k_r)} \exp -r\alpha$$

et :

$$h(x_1, \dots, x_r) = \frac{1}{x_1! \dots x_r!}$$

3. On a :

$$\begin{aligned} \delta &= P[X = 0] \\ &= \exp -\alpha \end{aligned}$$

Pour tout δ , $\delta > 0$, et tout $(k_1, \dots, k_r) \in \mathbb{N}^r$ par :

$$\begin{aligned} \mathcal{L}(\delta; k_1, \dots, k_r) &= \prod_{i=1}^r p(k_i) \\ &= \frac{(-\ln \delta)^{\sum_{i=1}^r k_i}}{k_1! \dots k_r!} \delta^r \end{aligned}$$

d'où :

$$\ln \mathcal{L}(\delta; k_1, \dots, k_r) = -\ln(k_1! \dots k_r!) + \sum_{i=1}^r k_i \ln(-\ln \delta) + r \ln \delta$$

Il en résulte que :

$$\frac{\partial}{\partial \delta} \ln \mathcal{L}(\delta; k_1, \dots, k_r) = \frac{\sum_{i=1}^r k_i}{\delta \ln \delta} + \frac{r}{\delta}$$

d'où :

$$\frac{\partial}{\partial \delta} \ln \mathcal{L}(\delta; k_1, \dots, k_r) = 0 \implies \delta = \exp - \left(\frac{1}{r} \sum_{i=1}^r k_i \right)$$

et comme :

$$\frac{\partial^2}{\partial \delta^2} \ln \mathcal{L}(\delta; k_1, \dots, k_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\begin{aligned} \hat{\delta} &= \exp - \left(\frac{1}{r} \sum_{i=1}^r X_i \right) \\ &= \exp - \hat{\alpha} \end{aligned}$$

Exercice 7

Soit α un réel appartenant à $]1, +\infty[$ et X une variable aléatoire telle que :

$$P[X = k] = \frac{1}{\alpha} \left(1 - \frac{1}{\alpha} \right)^{k-1}, \quad k \in \mathbb{N}^*$$

1. Calculer l'espérance mathématique et la variance de X .
2. Déterminer l'estimateur du maximum de vraisemblance $\hat{\alpha}$ de α d'un r -échantillon de variable parente X et étudier ses propriétés.
3. $\hat{\alpha}$ est-il un résumé exhaustif ?

Solution 7

1. On a :

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k P[X = k] \\ &= \sum_{k=1}^{\infty} \frac{k}{\alpha} \left(1 - \frac{1}{\alpha} \right)^{k-1} \\ &= \frac{1}{\alpha} \end{aligned}$$

et :

$$\begin{aligned}
 E[X(X-1)] &= \sum_{k=1}^{\infty} k(k-1) P[X=k] \\
 &= \sum_{k=1}^{\infty} \frac{k(k-1)}{\alpha} \left(1 - \frac{1}{\alpha}\right)^{k-1} \\
 &= 2\alpha^2 \left(1 - \frac{1}{\alpha}\right)
 \end{aligned}$$

d'où :

$$\begin{aligned}
 E[X^2] &= E[X(X-1)] + E[X] \\
 &= \alpha(2\alpha - 1)
 \end{aligned}$$

et :

$$\begin{aligned}
 V[X] &= E[X^2] - E[X]^2 \\
 &= \alpha(\alpha - 1)
 \end{aligned}$$

2. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\alpha \in]1, +\infty[$ et tout $(x_1, \dots, x_r) \in (\mathbb{N}^*)^r$ par :

$$\begin{aligned}
 \mathcal{L}(\alpha; x_1, \dots, x_r) &= \prod_{i=1}^r p(x_i) \\
 &= \frac{1}{\alpha^r} \left(1 - \frac{1}{\alpha}\right)^{\sum_{i=1}^r x_i - r}
 \end{aligned}$$

d'où :

$$\ln \mathcal{L}(\alpha; x_1, \dots, x_r) = -r \ln \alpha + \left(\sum_{i=1}^r x_i - r\right) \ln \left(1 - \frac{1}{\alpha}\right)$$

Il en résulte que :

$$\begin{aligned}
 \frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) &= -\frac{r}{\alpha} + \frac{\sum_{i=1}^r x_i - r}{\alpha(\alpha - 1)} \\
 &= \frac{\sum_{i=1}^r x_i - r\alpha}{\alpha(\alpha - 1)}
 \end{aligned}$$

d'où :

$$\frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) = 0 \implies \alpha = \frac{1}{r} \sum_{i=1}^r x_i$$

et comme :

$$\frac{\partial^2}{\partial \alpha^2} \ln \mathcal{L}(p; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon d'une structure géométrique est :

$$\hat{\alpha} = \frac{1}{r} \sum_{i=1}^r X_i$$

C'est la moyenne empirique du r -échantillon.

3. Puisque :

$$\hat{\alpha} = \frac{1}{r} \sum_{i=1}^r X_i$$

alors :

$$\begin{aligned} E[\hat{\alpha}] &= E[X] \\ &= \alpha \end{aligned}$$

et :

$$\begin{aligned} V[\hat{\alpha}] &= \frac{V[X]}{r} \\ &= \frac{\alpha(\alpha - 1)}{r} \end{aligned}$$

On en déduit que $\hat{\alpha}$ est un estimateur sans biais et convergent du paramètre α d'une structure géométrique.

Exercice 8

Soit X une variable aléatoire qui suit une loi de *Pareto* dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} \frac{\alpha a^\alpha}{x^{\alpha+1}} & \text{si } x \geq a \\ 0 & \text{si } x < a \end{cases}$$

où X représente le revenu par habitant, a le revenu minimum et α , $\alpha > 2$, un coefficient dépendant du type du pays où l'on se place.

1. Vérifier que f est bien une densité de probabilité.
2. Calculer l'espérance mathématique et la variance de X .
3. Calculer la fonction de répartition de X .
4. Déterminer l'estimateur du maximum de vraisemblance \hat{a} de a d'un r -échantillon issu X .
5. Dans le cas où \hat{a} est biaisé, proposer un estimateur sans biais de a .

Solution 8

1. La densité de probabilité de la loi de Pareto est définie par :

$$f(x) = \begin{cases} \frac{\alpha a^\alpha}{x^{\alpha+1}} & \text{si } x \geq a \\ 0 & \text{si } x < a \end{cases}$$

f est bien une densité de probabilité.

En effet :

$$\begin{aligned} \int_{\mathbb{R}} f(x) dx &= \int_a^{+\infty} \frac{\alpha a^\alpha}{x^{\alpha+1}} dx \\ &= 1 \end{aligned}$$

2. On a :

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x f(x) dx \\ &= \int_a^{+\infty} \frac{\alpha a^\alpha}{x^\alpha} dx \\ &= \frac{\alpha}{\alpha-1} a \end{aligned}$$

et :

$$\begin{aligned} E[X^2] &= \int_{\mathbb{R}} x^2 f(x) dx \\ &= \int_a^{+\infty} \frac{\alpha a^\alpha}{x^{\alpha-1}} dx \\ &= \frac{\alpha}{\alpha-2} a^2 \end{aligned}$$

d'où :

$$\begin{aligned} V[X] &= E[X^2] - E[X]^2 \\ &= \frac{\alpha}{(\alpha-2)(\alpha-1)^2} a^2 \end{aligned}$$

3. La fonction de répartition F de X est définie pour tout $x \in \mathbb{R}$ par :

$$\begin{aligned}
 F(x) &= \int_{-\infty}^x f(t) dt \\
 &= \begin{cases} 0 & \text{si } x \leq a \\ \int_a^x \frac{\alpha a^\alpha}{t^{\alpha+1}} dt & \text{si } x > a \end{cases} \\
 &= \begin{cases} 0 & \text{si } x \leq a \\ 1 - \frac{a^\alpha}{x^\alpha} & \text{si } x > a \end{cases}
 \end{aligned}$$

4. Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $a \in \mathbb{R}$ et tout $(x_1, \dots, x_r) \in (]a, +\infty[)^r$, par :

$$\begin{aligned}
 \mathcal{L}(a; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\
 &= \frac{\alpha^r a^{r\alpha}}{(x_1 \dots x_r)^{\alpha+1}}
 \end{aligned}$$

La fonction :

$$a \longrightarrow \mathcal{L}(a; x_1, \dots, x_r)$$

est strictement croissante, donc elle atteint son maximum lorsque a est maximum.

Et comme :

$$\forall i \in \{1, \dots, r\} : a \leq x_i$$

Il en résulte que θ est maximum lorsque :

$$a = \min(x_1, \dots, x_r)$$

Donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\hat{a} = \min(X_1, \dots, X_r)$$

5. Pour déterminer la densité de probabilité de $\hat{\theta}$, commençons d'abord par calculer sa fonction de répartition.

(a) **Fonction de répartition de \hat{a} :**Pour tout $x \in \mathbb{R}$ on a :

$$\begin{aligned}
F_{\hat{a}}(x) &= P[\hat{a} < x] \\
&= P[\min(X_1, \dots, X_r) < x] \\
&= 1 - P[\min(X_1, \dots, X_r) \geq x] \\
&= 1 - P[X_1 \geq x, \dots, X_r \geq x] \\
&= 1 - \prod_{k=1}^r P[X_k \geq x] \\
&= 1 - \prod_{k=1}^r (1 - P[X_k < x]) \\
&= 1 - [1 - F(x)]^r \\
&= \begin{cases} 0 & \text{si } x \leq a \\ 1 - \left(\frac{a^\alpha}{x^\alpha}\right)^r & \text{si } x \geq a \end{cases}
\end{aligned}$$

(b) **Densité de probabilité de $\hat{\theta}$:**Pour tout $x \in \mathbb{R} - \{a\}$ on a :

$$\begin{aligned}
f_{\hat{a}}(x) &= \frac{d}{dv} F_{\hat{a}}(x) \\
&= \begin{cases} 0 & \text{si } x < a \\ \frac{r\alpha a^{r\alpha}}{x^{r\alpha+1}} & \text{si } x > a \end{cases}
\end{aligned}$$

(c) **Espérance mathématique de \hat{a} :**

$$\begin{aligned}
E[\hat{a}] &= \int_{\mathbb{R}} v f_{\hat{a}}(v) dv \\
&= \int_a^{+\infty} \frac{r\alpha a^{r\alpha}}{v^{r\alpha}} dv \\
&= \frac{r\alpha}{r\alpha - 1} a
\end{aligned}$$

(d) **Espérance mathématique de \hat{a}^2 :**

$$\begin{aligned}
E[\hat{a}^2] &= \int_{\mathbb{R}} v^2 f_{\hat{a}}(v) dv \\
&= \int_a^{+\infty} \frac{r\alpha a^{r\alpha}}{v^{r\alpha-1}} dv \\
&= \frac{r\alpha}{r\alpha - 2} a^2
\end{aligned}$$

(e) **Variance de \hat{a} :**

$$\begin{aligned} V[\hat{a}] &= E[\hat{a}^2] - E[\hat{a}]^2 \\ &= \frac{r\alpha}{(r\alpha - 2)(r\alpha - 1)^2} a^2 \end{aligned}$$

L'estimateur \hat{a} de a est biaisé, mais il est asymptotiquement sans biais.

(f) Considérons l'estimateur :

$$T = \frac{r\alpha - 1}{r\alpha} \hat{a}$$

Alors :

$$E[T] = a$$

et :

$$V[T] = \left(\frac{r\alpha - 1}{r\alpha} \right)^2 V[\hat{a}] = \frac{1}{r\alpha(r\alpha - 2)} a^2$$

T est donc un estimateur sans biais et convergent de a .

Exercice 9

Soit X une variable aléatoire dont la densité de probabilité f est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \leq \theta \\ \frac{1}{\alpha} \exp \frac{(\theta - x)}{\alpha} & \text{si } x > \theta \end{cases}$$

où θ est un paramètre réel et α un paramètre réel strictement positif.

1. Vérifier que f est bien une densité de probabilité.
2. Calculer l'espérance mathématique et la variance de X .
3. Calculer la fonction de répartition de X .
4. On suppose θ connu et α inconnu.
 - (a) Déterminer l'estimateur du maximum de vraisemblance $\hat{\alpha}$ de α d'un r -échantillon issu X .
 - (b) Etudier les propriétés de $\hat{\alpha}$.
 - (c) Dans le cas où $\hat{\alpha}$ est biaisé, proposer un estimateur sans biais de α .
5. On suppose α connu et θ inconnu.
 - (a) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ d'un r -échantillon issu de X .
 - (b) Etudier les propriétés de $\hat{\theta}$.
 - (c) Dans le cas où $\hat{\theta}$ est biaisé, proposer un estimateur sans biais de θ .

6. On suppose que θ et α sont tous les deux inconnus.

- (a) Déterminer l'estimateur du maximum de vraisemblance $(\hat{\alpha}, \hat{\theta})$ de (α, θ) d'un r-échantillon issu de X .
- (b) Etudier les propriétés de $(\hat{\alpha}, \hat{\theta})$
- (c) Proposer un estimateur sans biais de (α, θ) .

Solution 9

1. f est bien une densité de probabilité.

En effet :

$$\begin{aligned} \int_{\mathbb{R}} f(x) dx &= \int_{\theta}^{+\infty} \frac{1}{\alpha} \exp \frac{(\theta - x)}{\alpha} dx \\ &= \int_0^{+\infty} \exp -tdt \\ &= 1 \end{aligned}$$

2. On a :

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x f(x) dx \\ &= \int_{\theta}^{+\infty} \frac{x}{\alpha} \exp \frac{(\theta - x)}{\alpha} dx \\ &= \int_0^{+\infty} (\alpha t + \theta) \exp -tdt \\ &= \alpha + \theta \end{aligned}$$

et :

$$\begin{aligned} E[X^2] &= \int_{\mathbb{R}} x^2 f(x) dx \\ &= \int_{\theta}^{+\infty} \frac{x^2}{\alpha} \exp \frac{(\theta - x)}{\alpha} dx \\ &= \int_0^{+\infty} (\alpha t + \theta)^2 \exp -tdt \\ &= 2\alpha^2 + 2\alpha\theta + \theta^2 \\ &= (\alpha + \theta)^2 + \alpha^2 \end{aligned}$$

d'où :

$$\begin{aligned} V[X] &= E[X^2] - E[X]^2 \\ &= \alpha^2 \end{aligned}$$

3. La fonction de répartition F de X est définie pour tout $x \in \mathbb{R}$ par :

$$\begin{aligned}
 F(x) &= \int_{-\infty}^x f(t) dt \\
 &= \begin{cases} 0 & \text{si } x \leq \theta \\ \int_{\theta}^x \frac{1}{\alpha} \exp \frac{(\theta - t)}{\alpha} dt & \text{si } x \geq \theta \end{cases} \\
 &= \begin{cases} 0 & \text{si } x \leq \theta \\ 1 - \exp \frac{(\theta - x)}{\alpha} & \text{si } x \geq \theta \end{cases}
 \end{aligned}$$

4. On suppose θ connu et α inconnu.

(a) Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout α , $\alpha > 0$, $\theta \in \mathbb{R}$ et tout $(x_1, \dots, x_r) \in]\theta, +\infty[)^r$ par :

$$\begin{aligned}
 \mathcal{L}(\alpha; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\
 &= \frac{1}{\alpha^r} \exp \sum_{i=1}^r \frac{(\theta - x_i)}{\alpha}
 \end{aligned}$$

d'où :

$$\ln \mathcal{L}(\alpha; x_1, \dots, x_r) = -r \ln \alpha + \frac{1}{\alpha} \sum_{i=1}^r (\theta - x_i)$$

Il en résulte que :

$$\begin{aligned}
 \frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) &= -\frac{r}{\alpha} - \frac{1}{\alpha^2} \sum_{i=1}^r (\theta - x_i) \\
 &= \frac{1}{\alpha} \left[r - \frac{1}{\alpha} \sum_{i=1}^r (\theta - x_i) \right]
 \end{aligned}$$

d'où :

$$\begin{aligned}
 \frac{\partial}{\partial \alpha} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) = 0 &\implies \alpha = \frac{1}{r} \sum_{i=1}^r (x_i - \theta) \\
 &\implies \alpha = \left[\frac{1}{r} \sum_{i=1}^r x_i \right] - \theta
 \end{aligned}$$

et comme :

$$\frac{\partial^2}{\partial \alpha^2} \ln \mathcal{L}(\alpha; x_1, \dots, x_r) < 0$$

donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\hat{\alpha} = \left[\frac{1}{r} \sum_{i=1}^r X_i \right] - \theta$$

(b) On a :

$$\begin{aligned} E[\hat{\alpha}] &= E \left[\left(\frac{1}{r} \sum_{i=1}^r X_i \right) - \theta \right] \\ &= \alpha \end{aligned}$$

et :

$$\begin{aligned} V[\hat{\alpha}] &= V \left[\left(\frac{1}{r} \sum_{i=1}^r X_i \right) - \theta \right] \\ &= \frac{V[X]}{r} \\ &= \frac{\alpha^2}{r} \end{aligned}$$

5. On suppose α connu et θ inconnu.

(a) Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout α , $\alpha > 0$, $\theta \in \mathbb{R}$ et tout $(x_1, \dots, x_r) \in (]\theta, +\infty[)^r$, tous strictement positifs, par :

$$\begin{aligned} \mathcal{L}(\theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{\alpha^r} \exp \sum_{i=1}^r \frac{(\theta - x_i)}{\alpha} \end{aligned}$$

La fonction :

$$\theta \longrightarrow \mathcal{L}(\theta; x_1, \dots, x_r)$$

est strictement croissante, donc elle atteint son maximum lorsque θ est maximum.

Et comme :

$$\forall i \in \{1, \dots, r\} : \theta \leq x_i$$

Il en résulte que θ est maximum lorsque :

$$\theta = \min(x_1, \dots, x_r)$$

Donc l'estimateur du maximum de vraisemblance d'un r -échantillon de cette structure est :

$$\hat{\theta} = \min(X_1, \dots, X_r)$$

(b) Pour déterminer la densité de probabilité de $\hat{\theta}$, commençons d'abord par calculer sa fonction de répartition.

(i) **Fonction de répartition de $\hat{\theta}$:**

Pour tout $v \in \mathbb{R}$ on a :

$$\begin{aligned} F_{\hat{\theta}}(v) &= P[\hat{\theta} < v] \\ &= P[\min(X_1, \dots, X_r) < v] \\ &= 1 - P[\min(X_1, \dots, X_r) \geq v] \\ &= 1 - P[X_1 \geq v, \dots, X_r \geq v] \\ &= 1 - \prod_{k=1}^r P[X_k \geq v] \\ &= 1 - \prod_{k=1}^r (1 - P[X_k < v]) \\ &= 1 - [1 - F(v)]^r \\ &= \begin{cases} 0 & \text{si } v \leq \theta \\ 1 - \exp r \left(\frac{\theta - v}{\alpha} \right) & \text{si } v \geq \theta \end{cases} \end{aligned}$$

(ii) **Densité de probabilité de $\hat{\theta}$:**

Pour tout $v \in \mathbb{R} - \{\theta\}$ on a :

$$\begin{aligned} f_{\hat{\theta}}(v) &= \frac{d}{dv} F_{\hat{\theta}}(v) \\ &= \begin{cases} 0 & \text{si } v < \theta \\ \frac{r}{\alpha} \exp r \left(\frac{\theta - v}{\alpha} \right) & \text{si } v > \theta \end{cases} \end{aligned}$$

(iii) **Espérance mathématique de $\hat{\theta}$:**

$$\begin{aligned}
 E[\hat{\theta}] &= \int_{\mathbb{R}} v f_{\hat{\theta}}(v) dv \\
 &= \int_{\theta}^{+\infty} \frac{r}{\alpha} v \exp r \left(\frac{\theta - v}{\alpha} \right) dv \\
 &= \int_0^{+\infty} \left(\frac{\alpha}{r} t + \theta \right) \exp -t dt \\
 &= \frac{\alpha}{r} + \theta
 \end{aligned}$$

(iv) **Espérance mathématique de $\hat{\theta}^2$:**

$$\begin{aligned}
 E[\hat{\theta}^2] &= \int_{\mathbb{R}} v^2 f_{\hat{\theta}}(v) dv \\
 &= \int_{\theta}^{+\infty} \frac{r}{\alpha} v^2 \exp r (\theta - v) dv \\
 &= \left(\frac{\alpha}{r} + \theta \right)^2 + \left(\frac{\alpha}{r} \right)^2
 \end{aligned}$$

(v) **Variance de $\hat{\theta}$:**

$$\begin{aligned}
 V[\hat{\theta}] &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 \\
 &= \left(\frac{\alpha}{r} \right)^2
 \end{aligned}$$

L'estimateur $\hat{\theta}$ de θ est biaisé, mais il est asymptotiquement sans biais.

(c) Considérons l'estimateur :

$$T = \hat{\theta} - \frac{\alpha}{r}$$

Alors :

$$\begin{aligned}
 E[T] &= E[\hat{\theta}] - \frac{\alpha}{r} \\
 &= \theta
 \end{aligned}$$

et :

$$\begin{aligned}
 V[T] &= V[\hat{\theta}] \\
 &= \frac{\alpha}{r^2}
 \end{aligned}$$

T est donc un estimateur sans biais et convergent de θ .

6. On suppose que θ et α sont tous les deux inconnus.

(a) Considérons un r -échantillon de cette structure.

Sa fonction de vraisemblance est définie pour tout $\alpha, \alpha > 0, \theta \in \mathbb{R}$ et tout $(x_1, \dots, x_r) \in (]\theta, +\infty[)^r$, tous strictement positifs, par :

$$\begin{aligned}\mathcal{L}(\alpha, \theta; x_1, \dots, x_r) &= \prod_{i=1}^r f(x_i) \\ &= \frac{1}{\alpha^r} \exp \sum_{i=1}^r \frac{(\theta - x_i)}{\alpha}\end{aligned}$$

Compte tenu des questions précédentes, la fonction :

$$(\alpha, \theta) \longmapsto \mathcal{L}(\alpha, \theta; x_1, \dots, x_r)$$

atteint son maximum pour :

$$\begin{cases} \theta = \min(x_1, \dots, x_r) \\ \alpha = \left[\frac{1}{r} \sum_{i=1}^r x_i \right] - \theta \end{cases}$$

d'où, les estimateurs du maximum de vraisemblance $(\hat{\alpha}, \hat{\theta})$ de (α, θ) sont donnés par :

$$\begin{cases} \hat{\theta} = \min(X_1, \dots, X_r) \\ \hat{\alpha} = \left[\frac{1}{r} \sum_{i=1}^r X_i \right] - \hat{\theta} \end{cases}$$

(b) On a :

$$E[\hat{\theta}] = \frac{\alpha}{r} + \theta$$

et :

$$\begin{aligned}E[\hat{\alpha}] &= E[X] - E[\hat{\theta}] \\ &= \frac{r-1}{r} \alpha\end{aligned}$$

Donc les estimateurs $\hat{\alpha}$ et $\hat{\theta}$ sont biaisés.

(c) Considérons les estimateurs T et S de α et θ respectivement définis par :

$$\begin{cases} T = \frac{r}{r-1} \hat{\alpha} \\ S = \hat{\theta} - \frac{1}{r-1} \hat{\alpha} \end{cases}$$

alors :

$$\begin{cases} E[T] = \alpha \\ E[S] = \theta \end{cases}$$

Donc T et S sont des estimateurs sans biais de α et θ respectivement.

Exercice 10

Soient X et Y deux variables aléatoires indépendantes, la première prenant les valeurs 1 et 0 avec les probabilités respectives α et $1 - \alpha$, et la deuxième prenant les valeurs 1 et 0 avec les probabilités respectives P et $1 - P$. On suppose α inconnue et P connue, $P > 0.5$.

On définit la variable aléatoire Z par :

$$\begin{cases} Z = 1 & \text{si } X = Y \\ Z = 0 & \text{si } X \neq Y \end{cases}$$

On considère un n -échantillon $((X_1, Y_1), \dots, (X_n, Y_n))$ de (X, Y) et on définit Z_i , $1 \leq i \leq n$, à partir de X_i et Y_i comme on a défini Z à partir de X et Y .

1. Montrer que (Z_1, \dots, Z_n) est un n -échantillon de Z .
2. Etudier les propriétés de l'estimateur :

$$T = \frac{1}{n} (Z_1 + \dots + Z_n)$$

3. Proposer alors un estimateur sans biais S de α .
4. Etudier la variance de S en fonction de P .
5. Indiquer un intervalle de confiance pour α lorsque n est grand, en supposant qu'on dispose d'une observation p de la variable :

$$T = \frac{1}{n} (Z_1 + \dots + Z_n)$$

6. Voyez-vous une application de ce qui précède dans le domaine des sondages d'opinion ?

Solution 10

On a :

$$\begin{cases} P[X = 0] = 1 - \alpha \\ P[X = 1] = \alpha \end{cases}$$

et :

$$\begin{cases} P[Y = 0] = 1 - P \\ P[Y = 1] = P \end{cases}$$

X et Y deux variables aléatoires de Bernoulli de paramètres α et P respectivement. Déterminons la loi de probabilité de Z :

$$\begin{aligned} P[Z = 0] &= P[X \neq Y] \\ &= P[\{(X, Y) = (0, 1)\} \oplus \{(X, Y) = (1, 0)\}] \\ &= P[X = 0] P[Y = 1] + P[X = 1] P[Y = 0] \\ &= (1 - \alpha) P + \alpha (1 - P) \end{aligned}$$

et :

$$\begin{aligned} P[Z = 1] &= P[X = Y] \\ &= P[\{(X, Y) = (0, 0)\} \oplus \{(X, Y) = (1, 1)\}] \\ &= P[X = 0] P[Y = 0] + P[X = 1] P[Y = 1] \\ &= (1 - \alpha) (1 - P) + \alpha P \end{aligned}$$

Z est donc une variable aléatoire de Bernoulli de paramètre :

$$\theta = (1 - \alpha) (1 - P) + \alpha P$$

de plus :

$$\begin{aligned} E[Z] &= \theta \\ V[Z] &= \theta (1 - \theta) \end{aligned}$$

1. Puisque $(X_1, Y_1), \dots, (X_n, Y_n)$ sont indépendants et suivent la même loi que (X, Y) , on en déduit que (Z_1, \dots, Z_n) sont indépendants et suivent la même loi que Z , donc c'est un n -échantillon de Z .
2. Soit l'estimateur :

$$T = \frac{1}{n} (Z_1 + \dots + Z_n)$$

On a :

$$\begin{aligned} E[T] &= E[Z] \\ &= (1 - \alpha) (1 - P) + \alpha P \end{aligned}$$

et :

$$\begin{aligned} V[T] &= \frac{1}{n} V[Z] \\ &= \frac{1}{n} [(1-\alpha)(1-P) + \alpha P] [(1-\alpha)P + \alpha(1-P)] \end{aligned}$$

3. T est donc un estimateur biaisé de α sauf lorsque :

$$\alpha = \frac{1}{2}$$

ou :

$$P = 1$$

(a) Si :

$$\alpha = \frac{1}{2} \text{ ou } P = 1$$

alors il suffit de prendre :

$$S = T$$

(b) Si :

$$\alpha \neq \frac{1}{2} \text{ et } P \neq 1$$

alors il suffit de prendre :

$$S = \frac{1}{2P-1} [T - (1-P)]$$

4. On a :

$$\begin{aligned} V[S] &= \frac{1}{(2P-1)^2} V[T] \\ &= \frac{1}{n(2P-1)^2} [(1-\alpha)(1-P) + \alpha P] [(1-\alpha)P + \alpha(1-P)] \end{aligned}$$

Tests d'Hypothèses

Les Fréquences

Exercice 1

A la veille d'une consultation électorale, on a interrogé cent électeurs constituant un échantillon au hasard. Soixante ont déclaré avoir l'intention de voter pour le candidat C .

En quelles limites, au moment du sondage, la proportion du corps électoral favorable à C se situe-t-elle ?

Solution 1

Construisons l'intervalle de confiance correspondant à la fréquence $f = 0.6$ du corps électoral favorable à C observée sur un échantillon de taille $n = 100$.

Au seuil α , cet intervalle est défini par :

$$\left[f - t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on obtient alors l'intervalle :

$$[.504, .696]$$

A 95%, le candidat C serait élu.

Exercice 2

On sait que le taux de mortalité d'une certaine maladie est de 30%.

Sur 200 malades testés, combien peut-on envisager de décès ?

Solution 2

Construisons d'abord l'intervalle de pari, pour un échantillon de taille $n = 200$, correspondant à la probabilité de décès $p = 0.3$.

Au seuil α , cet intervalle est défini par :

$$\left[p - t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on obtient alors l'intervalle :

$$[.24, .36]$$

Il en résulte que sur les 200 malades, le nombre de décès à envisager serait compris, à 95%, entre 48 et 72 décès.

Exercice 3

Dans une pré-enquête, on sélectionne, par tirage au sort cent dossiers. Quinze d'entre eux sont incomplets. Combien de dossiers incomplets trouvera-t-on sur dix milles dossiers ?

Solution 3

Construisons l'intervalle de confiance correspondant à la fréquence $f = 0.15$ de dossiers incomplets observée sur un échantillon de taille $n = 100$.

Au seuil α , cet intervalle est défini par :

$$\left[f - t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on obtient alors l'intervalle :

$$[.08, .22]$$

Il en résulte que sur les 10000 dossiers, le nombre de dossiers incomplets serait compris, à 95%, entre 800 et 2200 dossiers.

Exercice 4

Dans une maternité, on fait le point de la proportion de filles toutes les cent naissances.

Comment peut varier cette proportion d'une fois à l'autre si l'on admet qu'il nait en moyenne 51% de filles ?

Solution 4

Construisons l'intervalle de pari, pour un échantillon de taille $n = 100$, correspondant à la probabilité d'obtenir une fille $p = 0.51$.

Au seuil α , cet intervalle est défini par :

$$\left[p - t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on obtient alors l'intervalle :

$$[.41, .61]$$

Il en résulte, qu'à 95%, la proportion de filles varie d'une fois à l'autre, entre 41% et 61%.

Exercice 5

Une machine à former des pilules fonctionne de façon satisfaisante si la proportion de pilules non réussies est de 1 pour 1000.

Sur un échantillon de 10000 pilules, on a trouvé 15 pilules défectueuses.

Que faut-il conclure ?

Solution 5

Ici on :

$$\begin{cases} n = 10^4 \\ f = 15 \times 10^{-4} \\ p = 10^{-3} \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la machine est bien réglée"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \\ &= 1.58 \end{aligned}$$

on accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 5\%$, c'est à dire, qu'au seuil $\alpha = 5\%$, *la machine fonctionne de façon satisfaisante.*

Exercice 6

Sur un échantillon de 600 sujets atteints du cancer des poumons, on a trouvé 550 fumeurs.

Que peut-on dire du pourcentage de fumeurs parmi les cancéreux ?

Solution 6

Construisons l'intervalle de confiance correspondant à la fréquence $f = \frac{11}{12}$ des cancéreux parmi les fumeurs observée sur un échantillon de taille $n = 600$.

Au seuil α , cet intervalle est défini par :

$$\left[f - t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on obtient alors l'intervalle :

$$[.9, .94]$$

Il en résulte que parmi, les fumeurs, la proportion des atteints par le cancer des poumons est comprise, à 95%, entre 90% et 94%.

Exercice 7

Avant de procéder au lancement d'un produit, une entreprise a fait procéder à une enquête portant sur deux régions géographiques A et B .

Sur 1800 réponses provenant de la région A , 630 se déclarent intéressées par le produit.

En provenance de B , 150 réponses sur 600 se déclarent favorables.

Tester, au seuil de 5%, l'hypothèse de l'identité des opinions des régions A et B quant au produit considéré.

Solution 7

Ici on :

$$\begin{cases} n_A = 1800 & \text{et} & f_A = \frac{7}{20} \\ n_B = 600 & \text{et} & f_B = \frac{1}{4} \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"les opinions des régions A et B sont identiques"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f_A - f_B}{\sqrt{\frac{f_A(1-f_A)}{n_A} + \frac{f_B(1-f_B)}{n_B}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{f_A - f_B}{\sqrt{\frac{f_A(1-f_A)}{n_A} + \frac{f_B(1-f_B)}{n_B}}} \\ &= 4.77 \end{aligned}$$

on rejette donc l'hypothèse nulle H_0 à 95% (et même à 99.98%), c'est à dire, les deux régions A et B ont des opinions différentes.

Exercice 8

Dans un groupe de 200 malades atteints du cancer du col de l'utérus, un traitement par application locale du radium a donné 50 guérisons.

Un autre groupe de 150 sujets atteints de la même maladie a été traité par chirurgie, on a trouvé 50 guérisons.

Que peut-on conclure ?

Solution 8

Ici on :

$$\begin{cases} n_1 = 200 & , & f_1 = \frac{1}{4} \\ n_2 = 150 & , & f_2 = \frac{1}{3} \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"les deux traitements sont équivalents"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} \\ &= -1.69 \end{aligned}$$

on accepte donc l'hypothèse nulle H_0 au seuil 5%, c'est à dire, *les deux méthodes sont équivalentes*.

Exercice 9

Aux guichets d'une gare parisienne, sur les 350 billets vendus vendredi après-midi, 95 étaient des billets de 1^{ère} classe. Sur les 250 billets vendus la matinée du lundi suivant, 55 étaient de 1^{ère} classe.

Peut-on considérer qu'il y a une différence entre les proportions de vente de parcours en 1^{ère} classe pour les fins et débuts de semaines ?

Solution 9

Ici on :

$$\begin{cases} n_1 = 350 & , & f_1 = \frac{19}{70} \\ n_2 = 250 & , & f_2 = \frac{11}{50} \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \quad \text{''les taux de billets de 1^{ère} classe vendus en fin et début de semaines sont identiques''}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}}$$

peut être considérée comme une réalisation d'une variable normale centrée réduite. Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$t = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} = 1.45$$

on accepte donc l'hypothèse nulle H_0 au seuil 5%, c'est à dire, les *taux de billets de parcours en 1^{ère} classe vendus en fins et débuts de semaines sont identiques* et qu'on peut estimer par :

$$\begin{aligned} f &= \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} \\ &= 0.25 \end{aligned}$$

Exercice 10

On a lancé cent fois une pièce de monnaie et l'on a obtenu soixante fois "pile" et quarante fois "face".

Tester au seuil de 5%, puis 1%, l'hypothèse de la loyauté de la pièce.

Solution 10

Ici on :

$$\begin{cases} n = 100 \\ f = 0.6 \end{cases}$$

où f est la fréquence de "pile".

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la pièce est loyale"}$$

Sous cette hypothèse, on a :

$$p = 0.5$$

et la quantité :

$$t = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

on a :

$$\begin{aligned} t &= \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \\ &= 2 \end{aligned}$$

(1) Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on rejette donc l'hypothèse nulle H_0 à 95%, c'est à dire, qu'à 95%, *la pièce est truquée*.

(2) Pour $\alpha = 1\%$, on a :

$$2.57 < t_{.995} < 2.58$$

on accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 1\%$, c'est à dire, qu'au seuil $\alpha = 1\%$, *la pièce est normale*.

Exercice 11

Un échantillon de taille n a donné lieu au calcul d'une fréquence observée f correspondant à l'intervalle de confiance $[.22 - .34]$ au seuil $\alpha = 5\%$.

1. Calculer n .
2. Par rapport à la proportion $p = 0.3$, l'écart est-il significatif au seuil $\alpha = 5\%$?
3. Déterminer l'intervalle de confiance de $|f - p|$ au seuil $\alpha = 5\%$.

Solution 11

1. Au seuil α , l'intervalle de confiance correspondant à une fréquence f observée sur un échantillon de taille n est défini par :

$$\left[f - t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right]$$

On en déduit :

$$\begin{cases} f = \frac{0.22 + 0.34}{2} \\ n = t_{1-\alpha/2}^2 \frac{f(1-f)}{(f - 0.22)^2} \end{cases}$$

Pour $\alpha = 5\%$, on a :

$$t_{0.975} = 1.96$$

on obtient alors :

$$\begin{cases} f = .28 \\ n = 215 \end{cases}$$

2. Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"l'écart n'est pas singificatif"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

On a :

$$\begin{aligned} t &= \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \\ &= -0.64 \end{aligned}$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

on accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 5\%$.

3. Au seuil α :

$$\frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \in [-t_{1-\alpha/2}, t_{1-\alpha/2}]$$

donc, au seuil α :

$$|f - p| \in \left[0, t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

d'où :

$$|f - p| \in [0, 0.06]$$

Exercice 12

L'étude du taux de défauts afférentes aux caractéristiques de traitements thermiques d'une même pièce, traitée par deux fours différents, a donné lieu aux résultats suivants :

* Pour le premier four, 20 pièces défectueuses sur un échantillon de 200 pièces traitées.

* Pour le second four, 120 pièces défectueuses sur un échantillon de 800 pièces traitées.

Que peut-on conclure ?

Solution 12

Ici on :

$$\begin{cases} n_1 = 200 & , & f_1 = 0.10 \\ n_2 = 800 & , & f_2 = 0.15 \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"les deux traitements thermiques sont équivalents"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} \\ &= -2.03 \end{aligned}$$

on rejette donc l'hypothèse nulle H_0 à 95%, c'est à dire, *les deux traitements ne sont pas équivalents*.

Exercice 13

Un questionnaire auquel on ne peut répondre que par "oui" ou par "non", a été rempli par un échantillon de taille n .

L'intervalle de confiance de la fréquence observée f des réponses "oui" est $(0.35 - 0.43)$ au seuil $\alpha = 5\%$.

1. Quelle est la taille n de l'échantillon.
2. Par rapport à la proportion $p = 0.4$, l'écart est-il significatif au seuil $\alpha = 5\%$?
3. Déterminer l'intervalle de confiance de $|f - p|$ au seuil $\alpha = 5\%$.

Solution 13

1. Au seuil α , l'intervalle de confiance correspondant à une fréquence f observée sur un échantillon de taille n est défini par :

$$\left[f - t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right]$$

On en déduit :

$$\begin{cases} f = \frac{0.35 + 0.43}{2} \\ n = t_{1-\alpha/2}^2 \frac{f(1-f)}{(f - 0.35)^2} \end{cases}$$

Pour $\alpha = 5\%$, on a :

$$t_{0.975} = 1.96$$

on obtient alors :

$$\begin{cases} f = 0.39 \\ n = 571 \end{cases}$$

2. Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"l'écart n'est pas singificatif"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

On a :

$$\begin{aligned} t &= \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \\ &= -0.49 \end{aligned}$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

On accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 5\%$.

3. Au seuil α :

$$\frac{f - p}{\sqrt{\frac{p(1-p)}{n}}} \in [-t_{1-\alpha/2}, t_{1-\alpha/2}]$$

donc, au seuil α :

$$|f - p| \in \left[0, t_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

d'où :

$$|f - p| \in [0, 0.04]$$

Exercice 14

Parmi 470 sujets exposés à une infection, 370 n'ayant pas été immunisés.

Parmi ces derniers, 140 contractent la maladie ainsi que 25 sujets immunisés.

Le traitement donne-t-il une protection significative ?

Solution 14

Soient f_1 la fréquence de contracter la maladie pour un sujet non immunisé et f_2 la fréquence de contracter la maladie pour un sujet immunisé.

Ici on :

$$\begin{cases} n_1 = 370 & \text{et} & f_1 = \frac{14}{37} \\ n_2 = 100 & \text{et} & f_2 = \frac{1}{4} \end{cases}$$

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"le traitements n'est pas efficace"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} \\ &= 2.56 \end{aligned}$$

On rejette donc l'hypothèse nulle H_0 à 95%, c'est à dire, *le traitement donne une protection significative.*

Les Tests du Khi-deux

Exercice 1

Avant de procéder au lancement d'un produit, une entreprise a fait procéder à une enquête portant sur deux régions géographiques A et B .

Sur 1800 réponses provenant de la région A , 630 se déclarent intéressées par le produit.

En provenance de B , 150 réponses sur 600 se déclarent favorables.

Tester, au seuil de 5%, l'hypothèse de l'identité des opinions des régions A et B quant au produit considéré.

Solution 1

La répartition observée est :

Tableau des effectifs observés

<i>Région \ Opinion</i>	<i>favorable</i>	<i>non favorable</i>	<i>Total</i>
<i>Région A</i>	630	1170	1800
<i>Région B</i>	150	450	600
<i>Total</i>	780	1620	2400

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"les régions } A \text{ et } B \text{ ont la même opinion"}$$

Calculons, sous cette hypothèse, la répartition théorique :

Tableau des effectifs théoriques

<i>Région \ Opinion</i>	<i>favorable</i>	<i>non favorable</i>	<i>Total</i>
<i>Région A</i>	585	1215	1800
<i>Région B</i>	195	405	600
<i>Total</i>	780	1620	2400

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1) (2 - 1) = 1$$

degré de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{1;.95}^2 = 3.84$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 20.51 \end{aligned}$$

On rejette alors H_0 à 95% (et même à 99.5%), c'est à dire, *les deux régions ont des opinions différentes quant au produit considéré.*

Exercice 2

Dans un groupe de 200 malades atteints du cancer du col de l'utérus, un traitement par application locale du radium a donné 50 guérisons.

Un autre groupe de 150 sujets atteints de la même maladie a été traité par chirurgie, on a trouvé 54 guérisons.

Que peut-on conclure ?

Solution 2

La répartition observée est :

Tableau des effectifs observés

<i>Traitement \ Résultat</i>	<i>guéri</i>	<i>non guéri</i>	<i>Total</i>
<i>radium</i>	50	150	200
<i>chirurgie</i>	54	96	150
<i>Total</i>	104	246	350

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"les deux traitements sont équivalents"}$$

Calculons, sous cette hypothèse, la répartition théorique :

Tableau des effectifs théoriques

<i>Traitement \ Résultat</i>	<i>guéri</i>	<i>non guéri</i>	<i>Total</i>
<i>radium</i>	59.4	140.6	200
<i>chirurgie</i>	44.6	105.4	150
<i>Total</i>	104	246	350

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(2 - 1) = 1$$

degré de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{1;95}^2 = 3.84$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 4.94 \end{aligned}$$

On rejette alors H_0 à 95% , c'est à dire, *les deux traitements ne sont pas équivalents.*

Exercice 3

Aux guichets d'une gare parisienne, sur les 350 billets vendus vendredi après-midi, 95 étaient des billets de 1^{ère} classe. Sur les 250 billets vendus la matinée du lundi suivant, 55 étaient de 1^{ère} classe.

Peut-on considérer qu'il y a une différence entre les proportions de vente de parcours en 1^{ère} classe pour les fins et débuts de semaines ?

Solution 3

La répartition observée est :

Tableau des effectifs observées

<i>jour \ Classe</i>	<i>1^{ère} classe</i>	<i>2^{ème} classe</i>	<i>Total</i>
<i>Vendredi A.M</i>	95	255	350
<i>Lundi matin</i>	55	195	250
<i>Total</i>	150	450	600

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \quad \text{''les taux de billets de parcours en 1^{ère} classe vendus en fin et début de semaines sont identiques''}$$

Calculons, sous cette hypothèse, la répartition théorique :

Tableau des effectifs théoriques

<i>Jour \ Classe</i>	<i>1^{ère} classe</i>	<i>2^{ème} classe</i>	<i>Total</i>
<i>Vendredi A.M</i>	87.5	262.5	350
<i>Lundi matin</i>	62.5	187.5	250
<i>Total</i>	150	450	600

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(2 - 1) = 1$$

degré de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{1;95}^2 = 3.84$$

Et comme :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 2.06\end{aligned}$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *les taux de billets de parours en 1^{ère} classe vendus en fins et débuts de semaines sont identiques.*

Exercice 4

On a lancé cent fois une pièce de monnaie et l'on a obtenu soixante fois "pile" et quarante fois "face".

Tester au seuil de 5% puis 1%, l'hypothèse de la loyauté de la pièce.

Solution 4

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la pièce est loyale"}$$

Sous cette hypothèse, on a :

$$p = 0.5$$

d'où les répartitions :

<i>Côté</i>	<i>Répartition Observée</i>	<i>Répartition Théorique</i>
<i>pile</i>	60	50
<i>face</i>	40	50
<i>Total</i>	100	100

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$2 - 1 = 1$$

degré de liberté.

On a :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^2 \frac{(o_i - t_i)^2}{t_i} \\ &= 4\end{aligned}$$

(1) Pour $\alpha = 5\%$, on a :

$$\chi_{1;95}^2 = 3.84$$

On rejette donc l'hypothèse nulle H_0 à 95%, c'est à dire, qu'à 95%, *la pièce est truquée*.

(2) Pour $\alpha = 1\%$, on a :

$$\chi_{1;99}^2 = 6.63$$

On accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 1\%$, c'est à dire, qu'au seuil $\alpha = 1\%$, *la pièce est normale*.

Exercice 5

On veut savoir si la réussite (R) d'un traitement est indépendantes du niveaux de la tension artérielle du malade (T).

On dispose pour cela de 250 observations réparties comme suit :

$T \setminus R$	<i>échec</i>	<i>succès</i>
<i>basse</i>	21	104
<i>élevée</i>	29	96

Que peut-on conclure ?

Solution 5

La répartition observée est :

Tableau des effectifs observées

$T \setminus R$	<i>Echec</i>	<i>Succès</i>	<i>Total</i>
<i>Basse</i>	21	104	125
<i>Elevée</i>	29	96	125
<i>Total</i>	50	200	250

Testons, au seuil α , l'hypothèse nulle :

H_0 : "la réussite du traitement est indépendante du niveau de la tension artérielle"

Calculons, sous cette hypothèse, la répartition théorique, le tableau de cette répartition est donné ci-après.

Tableau des effectifs théoriques

$T \setminus R$	<i>Echec</i>	<i>Succès</i>	<i>Total</i>
<i>Basse</i>	25	100	125
<i>Elevée</i>	25	100	125
<i>Total</i>	50	200	250

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(2 - 1) = 1$$

degré de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{1;.95}^2 = 3.84$$

Et comme :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 1.6\end{aligned}$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *la réussite du traitement est indépendante du niveau de la tension artérielle.*

Exercice 6

On veut savoir s'il y a une liason entre la localisation (L) du cancer du poumon (périphérique , non périphérique) et le côté (C) de la lésion (poumon gauche , poumon droit).

L'étude a porté sur 1054 malades :

$L \setminus C$	<i>gauche</i>	<i>droit</i>
<i>périphérique</i>	26	62
<i>non périphérique</i>	416	550

Que peut-on conclure ?

Solution 6

La répartition observée est :

Tableau des effectifs observées

$L \setminus C$	<i>gauche</i>	<i>droit</i>	<i>Total</i>
<i>périphérique</i>	26	62	88
<i>non périphérique</i>	416	550	966
<i>Total</i>	442	612	1054

Testons, au seuil α , l'hypothèse nulle :

H_0 : "la localisation du cancer est indépendante du côté de la lésion"

Calculons, sous cette hypothèse, la répartition théorique.

Le tableau de cette répartition est donnée ci-après.

Tableau des effectifs théoriques

$L \setminus C$	<i>gauche</i>	<i>droit</i>	<i>Total</i>
<i>périphérique</i>	36.9	51.1	88
<i>nonpériphérique</i>	405.1	560.9	966
<i>Total</i>	442	612	1054

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(2 - 1) = 1$$

degré de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{1;95}^2 = 3.84$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 6.05 \end{aligned}$$

On rejette alors H_0 à 95% (même à 97.5%), c'est à dire, *la localisation du cancer dépend du côté de la lésion.*

Exercice 7

De nombreuses observations cliniques ont montré que jusque là :

- 30% des malades atteints de M ont une survie inférieure à un an
- 50% ont une survie entre un an et deux ans
- 10% ont une survie entre deux ans et cinq ans
- 10% ont une survie supérieure à cinq ans.

On applique un nouveau traitement à 80 malades atteint de la maladie M et on constate :

- 12 ont une survie inférieure à un an
- 56 ont une survie entre un an et deux ans
- 8 ont une survie entre deux ans et cinq ans
- 4 ont une survie supérieure à cinq ans.

Que peut-on conclure ?

Solution 7

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"le nouveau traitement n'est pas actif contre la maladie } M\text{"}$$

Sous cette hypothèse, on a les répartitions :

<i>Survie</i>	<i>Répartition Observée</i>	<i>Répartition Théorique</i>
<i>survie ≤ 1 an</i>	12	24
<i>1 an $<$ survie ≤ 2 ans</i>	56	40
<i>2 an $<$ survie ≤ 5 ans</i>	8	8
<i>survie > 5 ans</i>	4	8
<i>Total</i>	80	80

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^4 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$4 - 1 = 3$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{3;95}^2 = 7.81$$

Et comme :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^2 \frac{(o_i - t_i)^2}{t_i} \\ &= 14.4\end{aligned}$$

on rejette donc l'hypothèse nulle H_0 à 95% (même à 99.5%), c'est à dire, qu'à 99.5%, *le nouveau traitement est actif contre la maladie M.*

Exercice 8

On suppose pouvoir classer les malades atteints d'une maladie M en trois catégories cliniques : A , B , C .

On se demande si ces trois catégories diffèrent par leurs survies à un an.

Les effectifs observés sont les suivants :

<i>Survie \ Catégorie</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>survie à un an</i>	5	20	45
<i>décès avant un an</i>	15	50	145

Que peut-on conclure ?

Solution 8

La répartition observée est :

Tableau des effectifs observées

<i>Survie \ Catégorie</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>Total</i>
<i>Survie à un an</i>	5	20	45	70
<i>Décès avant un an</i>	15	50	145	210
<i>Total</i>	20	70	190	280

Testons, au seuil α , l'hypothèse nulle :

H_0 : "la survie à un an est indépendante de la catégorie clinique"

Calculons, sous cette hypothèse, la répartition théorique.

Tableau des effectifs théoriques

<i>Survie</i> \ <i>Catégorie</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>Total</i>
<i>Survie à un an</i>	5	17.5	47.5	70
<i>Décès avant un an</i>	15	52.5	142.5	210
<i>Total</i>	20	70	190	280

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(3 - 1) = 2$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{2;.95}^2 = 5.99$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= .65 \end{aligned}$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *la survie à un an est indépendante de la catégorie clinique.*

Exercice 9

75 enfants sont vus en consultation pour un asthme. On relève chez eux les deux symptômes suivants :

- * Intensité de la maladie asmathique : légère , moyenne , forte
- * Existence ou absence d'un eczéma au moment de l'observation ou dans le passé.

On peut classer les enfants selon la répartition suivante :

$E \setminus A$	<i>fort</i>	<i>moyen</i>	<i>léger</i>
<i>présent</i>	8	2	2
<i>passé</i>	11	11	3
<i>jamaïs</i>	6	18	14

Au vu de ces résultats, existe-t-il une association entre l'intensité de l'asthme et l'existence d'un eczéma ?

Solution 9

Le tableau de la répartition observée est donnée ci-après:

Tableau des effectifs observées

<i>Eczéma \ Asthme</i>	<i>fort</i>	<i>moyen</i>	<i>léger</i>	<i>Total</i>
<i>présent</i>	8	2	2	12
<i>passé</i>	11	11	3	25
<i>jamaïs</i>	6	18	14	38
<i>Total</i>	25	31	19	75

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{ "l'intensité de l'asthme est indépendante de l'existence d'un eczéma" }$$

Calculons, sous cette hypothèse, la répartition théorique.

Le tableau de cette répartition est donnée ci-après.

Tableau des effectifs théoriques

<i>Eczéma\Asthme</i>	<i>fort</i>	<i>moyen</i>	<i>léger</i>	<i>Total</i>
<i>présent</i>	4	4.96	3.04	12
<i>passé</i>	8.33	10.33	6.34	25
<i>jamais</i>	12.67	15.71	9.62	38
<i>Total</i>	25	31	19	75

Les effectifs théoriques sur la première ligne sont strictement inférieurs à cinq, ce qui empêche l'application d'un test du Khi-deux. On peut remédier à cet état en opérant le groupement "logique" des classes "*présent*" et "*passé*".

Les nouveaux tableaux des effectifs observés et théoriques, obtenus après regroupement de ces deux classes sont donnés ci-après.

Tableau des effectifs observées

<i>Eczéma\Asthme</i>	<i>fort</i>	<i>moyen</i>	<i>léger</i>	<i>Total</i>
<i>présent ou passé</i>	19	13	5	37
<i>jamais</i>	6	18	14	38
<i>Total</i>	25	31	19	75

Tableau des effectifs théoriques

<i>Eczéma\Asthme</i>	<i>fort</i>	<i>moyen</i>	<i>léger</i>	<i>Total</i>
<i>présent ou passé</i>	12.33	15.29	9.38	37
<i>jamais</i>	12.67	15.71	9.62	38
<i>Total</i>	25	31	19	75

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(2 - 1)(3 - 1) = 2$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{2;.95}^2 = 5.99$$

Et comme :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 11.84\end{aligned}$$

On rejette alors H_0 à 95% (même à 99.5%), c'est à dire, *l'intensité de l'asthme dépend de l'existence d'un eczéma.*

Exercice 10

Une étude statistique relative aux résultats d'admission du concours d'une grande école fait ressortir la répartition des admis selon la profession des parents lorsque celle-ci est connue.

1. La profession des parents a-t-elle une influence sur l'accès à cette école ?
2. Cette conclusion persiste-t-elle lorsqu'on tient compte pour compléter la statistique précédente de 961 candidats dont l'origine socio-professionnelle est inconnue et qui ont obtenus 43 succès ?

<i>Profession des Parents</i>	<i>Candidats</i>	<i>Admis</i>
<i>Fontionnaires et Assimilés</i>	2224	180
<i>Commerce et Industrie</i>	998	89
<i>Professions Libérales</i>	575	48
<i>Propriétaires Rentiers</i>	423	37
<i>Propriétaires Agricoles</i>	287	13
<i>Artisans</i>	210	18
<i>Banques et Assurances</i>	209	17

Solution 10

1. La répartition observée est :

<i>Profession des Parents</i>	<i>Candidats</i>	<i>Admis</i>	<i>Non admis</i>
<i>Fontionnaires et Assimilés</i>	2224	180	2044
<i>Commerce et Industrie</i>	998	89	899
<i>Professions Libérales</i>	575	48	527
<i>Propriétaires Rentiers</i>	423	37	386
<i>Propriétaires Agricoles</i>	287	13	274
<i>Artisans</i>	210	18	192
<i>Banques et Assurances</i>	209	17	192
<i>Total</i>	4916	402	4514

Testons, au seuil α , l'hypothèse nulle :

H_0 : "l'accès à l'Ecole est indépendant de la profession des parents"

Calculons, sous cette hypothèse, la répartition théorique :

<i>Profession des Parents</i>	<i>Candidats</i>	<i>Admis</i>	<i>Non admis</i>
<i>Fontionnaires et Assimilés</i>	2224	181.9	2042.1
<i>Commerce et Industrie</i>	998	80.8	907.2
<i>Professions Libérales</i>	575	47	528
<i>Propriétaires Rentiers</i>	423	34.6	388.4
<i>Propriétaires Agricoles</i>	287	23.5	263.5
<i>Artisans</i>	210	17.2	192.8
<i>Banques et Assurances</i>	209	17.1	191.9
<i>Total</i>	4916	402	4514

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^7 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(7 - 1)(2 - 1) = 6$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{6;.95}^2 = 12.6$$

Et comme :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} = 6.28$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *l'accès à l'Ecole est indépendant de la profession des parents.*

2. Si l'on tient compte des 961 candidats dont l'origine socio-professionnelle est inconnue et qui ont obtenus 43 succès, la répartition observée et la répartition théorique, sous la même hypothèse nulle, deviennent comme consognés ci-après.

Tableau des effectifs observées

<i>Profession des Parents</i>	<i>Candidats</i>	<i>Admis</i>	<i>Non admis</i>
<i>Fontionnaires et Assimilés</i>	2224	180	2044
<i>Commerce et Industrie</i>	998	89	899
<i>Professions Libérales</i>	575	48	527
<i>Propriétaires Rentiers</i>	423	37	386
<i>Propriétaires Agricoles</i>	287	13	274
<i>Artisans</i>	210	18	192
<i>Banques et Assurances</i>	209	17	192
<i>Autres</i>	961	43	918
<i>Total</i>	5877	445	5432

Tableau des effectifs théoriques

<i>Profession des Parents</i>	<i>Candidats</i>	<i>Admis</i>	<i>Non admis</i>
<i>Fontionnaires et Assimilés</i>	2224	168.4	2055.6
<i>Commerce et Industrie</i>	998	74.8	913.2
<i>Professions Libérales</i>	575	43.5	531.5
<i>Propriétaires Rentiers</i>	423	32	391
<i>Propriétaires Agricoles</i>	287	21.7	265.3
<i>Artisans</i>	210	15.9	194.1
<i>Banques et Assurances</i>	209	15.8	193.2
<i>Autres</i>	961	72.8	888.2
<i>Total</i>	5877	445	5432

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^8 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(8 - 1)(2 - 1) = 7$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{7;.95}^2 = 14.1$$

Et comme :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} = 22.5$$

On rejette alors H_0 à 95% (même à 99.5%) , c'est à dire, *l'accès à l'Ecole est indépendant de la profession des parents.*

Exercice 11

Sur un échantillon de 84 prématurés, on cherche s'il existe une liaison entre la survenue d'une hypoglycémie et la survenue d'un ictère :

- sur 43 enfants n'ayant pas d'ictère, 23 sont hypoglycémiques
- sur 20 enfants ayant un ictère modéré, 6 sont hypoglycémiques
- sur 21 enfants ayant un ictère intense, 4 sont hypoglycémiques

Que peut-on conclure ?

Solution 11

La répartition observée est donnée dans le tableau :

Tableau des effectifs observés

<i>Ictère \ Hypoglycémie</i>	<i>hypoglycémique</i>	<i>non hypoglycémique</i>	<i>Total</i>
<i>pas d'ictère</i>	23	20	43
<i>ictère modéré</i>	6	14	20
<i>ictère intense</i>	4	17	21
<i>Total</i>	33	51	84

Testons, au seuil α , l'hypothèse nulle :

H_0 : "la survenue d'une hypoglycémie est indépendante de la survenue d'un ictère"

Calculons, sous cette hypothèse, la répartition théorique :

Tableau des effectifs théoriques

<i>Ictère \ Hypoglycémie</i>	<i>hypoglycémique</i>	<i>non hypoglycémique</i>	<i>Total</i>
<i>pas d'ictère</i>	16.89	26.11	43
<i>ictère modéré</i>	7.86	12.14	20
<i>ictère intense</i>	8.25	12.75	21
<i>Total</i>	33	51	84

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(3 - 1)(2 - 1) = 2$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{2;.95}^2 = 5.99$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^3 \sum_{j=1}^2 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 7.97 \end{aligned}$$

On rejette alors H_0 à 95% (même à 97.5%), c'est à dire, *la survenue d'une hypoglycémie dépend de la survenue d'un ictère.*

Exercice 12

Un médicament essayé sur 42 patients est contrôlé quant aux effets secondaires qu'il peut avoir sur le poids des malades.

On peut considérer que :

- quinze d'entre eux ont maigri
- dix sept n'ont pas changé de poids
- dix ont grossi

En supposant que la maladie est sans effet sur les variations de poids, le médicament a-t-il un effet significatif sur le poids ?

Solution 12

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"le traitement est sans effet sur les variations du poids"}$$

Si le traitement est sans effet sur les variations du poids, alors ces variations sont dûes seulement au hasard.

La loi de probabilité est donc la loi uniforme, c'est à dire la probabilité de chaque classe est la même et est égale à $\frac{1}{3}$.

D'où les répartitions :

<i>Variations</i>	<i>Répartition Observée</i>	<i>Répartition Théorique</i>
<i>ont maigri</i>	15	14
<i>n'ont pas changé</i>	17	14
<i>ont grossi</i>	10	14
<i>Total</i>	42	42

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^3 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$3 - 1 = 2$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{2, .95}^2 = 5.99$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \frac{(o_i - t_i)^2}{t_i} \\ &= 1.86 \end{aligned}$$

on accepte donc l'hypothèse nulle H_0 au seuil $\alpha = 5\%$, c'est à dire, *le traitement est sans effet sur les variations du poids*.

Exercice 13

Pour étudier la densité de poussières dans un gaz, on a procédé à une série d'observations de petits échantillons de gaz au moyen d'un microscope.

On a ainsi effectué 143 observations et les résultats sont les suivants :

<i>Nombre de particules en suspension</i>	<i>Nombre d'échantillons de gaz</i>
0	34
1	46
2	38
3	19
4	4
5	2
> 5	0

Peut-on admettre, au seuil $\alpha = 5\%$, que le nombre de particules en suspension est une variable de *Poisson* ?

Solution 13

Testons, au seuil α , l'hypothèse nulle :

H_0 : "le nombre de particules en suspension est une variable de *Poisson*"

Calculons une estimation ponctuelle du paramètre α de cette loi :

$$P[X = k] = \frac{\alpha^k}{k!} \exp -\alpha$$

où X est la variable aléatoire représentant le nombre de particules en suspension. On sait que :

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n X_i$$

est un estimateur sans biais et convergent de α .

Une estimation ponctuelle $\tilde{\alpha}$ de α est donnée par :

$$\begin{aligned} \tilde{\alpha} &= \frac{1}{143} \sum_{i=0}^5 i n_i \\ &= 1.4336 \end{aligned}$$

D'où les répartitions :

<i>Particules en suspension</i>	<i>Répartition observée</i>	<i>Répartition théorique</i>
0	34	34.1
1	46	48.9
2	38	35.0
3	19	16.7
4	4	06.0
5	2	01.7
> 5	0	00.6
<i>Total</i>	143	143

L'effectif théorique t_k , $k \geq 0$, représentant le nombre particules en suspension k est donné par :

$$t_k = nP[X = k]$$

On constate que le tableau contient des effectifs théoriques strictement inférieurs à 5, ce qui empêche l'utilisation d'un test du khi-deux.

On peut remédier à cet état en opérant le groupement "logique" des classes "4 et plus". Les tableaux des effectifs observés et théoriques deviennent comme consignés ci-après.

<i>Particules en suspension</i>	<i>Répartition observée</i>	<i>Répartition théorique</i>
0	34	34.1
1	46	48.9
2	38	35.0
3	19	16.7
≥ 4	4	08.3
<i>Total</i>	143	143

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=0}^4 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$5 - 1 - 1 = 3$$

degrés de liberté. puisque pour calculer les effectifs théoriques, nous avons utilisé l'estimation, et non la valeur réel, du paramètre α de la loi de Poisson.

Pour $\alpha = 5\%$, on a :

$$\chi_{3;95}^2 = 7.81$$

Et comme :

$$\chi^2 = 2.97$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *le nombre de particules en suspension peut être ajusté par une loi de Poisson dont le paramètre α est estimé* par :

$$\tilde{\alpha} = 1.4336$$

Exercice 14

Le tableau ci-après concerne le nombre annuel de cyclones tropicaux ayant atteint la côte orientale des Etats-Unis entre 1887 et 1956 :

<i>Nombre annuel de cyclones</i>	<i>Nombre d'années</i>
0	1
1	6
2	10
3	16
4	19
5	5
6	8
7	3
8	1
9	1
> 9	0

Peut-on admettre, au seuil $\alpha = 5\%$, que ce nombre annuel de cyclones est une variable de *Poisson* ?

Solution 14

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"le nombre annuel de cyclones est une variable de Poisson"}$$

Calculons une estimation ponctuelle du paramètre α de cette loi :

$$P[X = k] = \frac{\alpha^k}{k!} \exp -\alpha$$

où X est la variable aléatoire représentant le nombre annuel de cyclones.

On sait que :

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n X_i$$

est un estimateur sans biais et convergent de α .

Une estimation ponctuelle $\tilde{\alpha}$ de α est donnée par :

$$\tilde{\alpha} = \frac{1}{70} \sum_{i=0}^9 in_i = 3.7286$$

L'effectif théorique t_k , $k \geq 0$, représentant le nombre d'années à k cyclones est donné par :

$$t_k = nP[X = k]$$

D'où les répartitions :

<i>Nombre annuel de cyclones</i>	<i>Effectifs observés</i>	<i>Effectifs théoriques</i>
0	1	1.68
1	6	6.27
2	10	11.69
3	16	14.53
4	19	13.54
5	5	10.1
6	8	6.28
7	3	3.34
8	1	1.56
9	1	0.65
> 9	0	0.36
<i>Total</i>	70	70

On constate que le tableau contient des effectifs théoriques strictement inférieurs à 5, ce qui empêche l'utilisation d'un test du khi-deux.

On peut remédier à cet état en opérant le groupement "logique" :

- * des classes "0" et "1" d'une part,
- * et des classes "7 et plus" d'autre part.

Les tableaux des effectifs observés et théoriques deviennent :

<i>Nombre annuel de cyclones</i>	<i>Effectifs observés</i>	<i>Effectifs théoriques</i>
0 ou 1	7	7.95
2	10	11.69
3	16	14.53
4	19	13.54
5	5	10.10
6	8	6.28
≥ 7	5	5.91
<i>Total</i>	70	70

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^7 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$7 - 1 - 1 = 5$$

degrés de liberté. puisque pour calculer les effectifs théoriques, nous avons utilisé l'estimation, et non la valeur réel, du paramètre α de la loi de Poisson.

Pour $\alpha = 5\%$, on a :

$$\chi_{5;.95}^2 = 5.8948$$

Et comme :

$$\chi^2 = 5.81$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *le nombre annuel de cyclones peut être ajusté par une loi de Poisson dont le paramètre α est estimé par :*

$$\tilde{\alpha} = 3.7286$$

Exercice 15

Le tableau suivant indique le résultat de l'examen de 124 sujets, classés d'après la couleur de leurs yeux (Y) et la couleur de leurs cheveux (C) :

$Y \setminus C$	<i>Blonds</i>	<i>Bruns</i>	<i>Noirs</i>	<i>Roux</i>
<i>Bleus</i>	25	9	3	7
<i>Gris ou Verts</i>	13	17	10	7
<i>Marrons</i>	7	13	8	5

Existe-t-il une liaison entre ces deux caractères ?

Solution 15

La répartition observée est :

$Y \setminus C$	<i>Blonds</i>	<i>Bruns</i>	<i>Noirs</i>	<i>Roux</i>	<i>Total</i>
<i>Bleus</i>	25	9	3	7	44
<i>Gris ou Verts</i>	13	17	10	7	47
<i>Marrons</i>	7	13	8	5	33
<i>Total</i>	45	39	21	19	124

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{ "les couleurs des yeux et des cheveux sont indépendantes" }$$

Calculons, sous cette hypothèse, la répartition théorique :

$Y \setminus C$	<i>Blonds</i>	<i>Bruns</i>	<i>Noirs</i>	<i>Roux</i>	<i>Total</i>
<i>Bleus</i>	16	13.8	7.4	6.8	44
<i>Gris ou Verts</i>	17	14.8	8	7.2	47
<i>Marrons</i>	12	10.4	5.6	5	33
<i>Total</i>	45	39	21	19	124

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=1}^3 \sum_{j=1}^4 \frac{(o_{ij} - t_{ij})^2}{t_{ij}}$$

est une réalisation d'une variable du Khi-deux à :

$$(3 - 1)(4 - 1) = 6$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{6;95}^2 = 12.6$$

Et comme :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^3 \frac{(o_{ij} - t_{ij})^2}{t_{ij}} \\ &= 15 \end{aligned}$$

On rejette alors H_0 à 95% (même à 97.5%), c'est à dire, *les couleurs des yeux et des cheveux ne sont pas indépendantes.*

Exercice 16

On considère les familles de quatre enfants.

Sur un échantillon de cent familles à quatre enfants, la répartition suivante a été observée :

<i>Nombre de filles</i>	<i>Nombre de familles</i>
0	7
1	20
2	41
3	22
4	10

Peut-on considérer que la probabilité qu'un enfant soit une fille est $\frac{1}{2}$?

Solution 16

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la probabilité d'avoir une fille est } \frac{1}{2} \text{"}$$

Sous l'hypothèse nulle H_0 , la variable aléatoire X égale au nombre de filles parmi les quatre enfants suit une loi binomiale d'ordre 4 et de paramètre $\frac{1}{2} : \mathcal{B}\left(4, \frac{1}{2}\right)$.

Ainsi, pour tout k , $0 \leq k \leq 4$, la probabilité p_k d'avoir k filles parmi les quatre enfants est :

$$p_k = C(4, k) \left(\frac{1}{2}\right)^4$$

L'effectif théorique t_k , $0 \leq k \leq 4$, représentant le nombre de familles ayant k filles parmi les quatre enfants est donné par :

$$t_k = np_k$$

D'où les répartitions :

<i>Nombre de filles</i>	<i>Répartition observée</i>	<i>Répartition théorique</i>
0	7	6.25
1	20	25
2	41	37.5
3	22	25
4	10	6.25
<i>Total</i>	100	100

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=0}^4 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$5 - 1 = 4$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi_{4;95}^2 = 9.49$$

Et comme :

$$\chi^2 = 4.03$$

On accepte alors H_0 au seuil $\alpha = 5\%$: la probabilité d'avoir une fille est $\frac{1}{2}$.

Exercice 17

On distribue un jeu de quarante cartes à quatre joueurs : A , B , C , D ; chacun recevant dix cartes

Un statisticien a élaboré un programme de distribution de donnes par ordinateur. Pour un ensemble de deux cents donnes, obtenues à partir de ce programme, il observe le nombre de donnes où le joueur A reçoit k as, $0 \leq k \leq 4$. Les résultats sont les suivants :

Nombre d'as	Nombre de donnes
0	64
1	74
2	52
3	8
4	2

Le programme du statisticien est-il fiable ?

Solution 17

Testons, au seuil α , l'hypothèse nulle :

$$H_0 : \text{"le programme du statisticien est fiable"}$$

Sous l'hypothèse nulle H_0 , la variable aléatoire X égale au nombre d'as du joueur A suit une loi hypergéométrique.

Ainsi, pour tout k , $0 \leq k \leq 4$, la probabilité p_k pour que le joueur A ait k as est :

$$p_k = \frac{C(4, k) C(36, 10 - k)}{C(40, 10)}$$

L'effectif théorique t_k , $0 \leq k \leq 4$, représentant le nombre de donnes à k as, du joueur A , est donné par :

$$t_k = np_k$$

D'où les répartitions :

<i>Nombre d'as</i>	<i>Répartition observée</i>	<i>Répartition théorique</i>
0	64	59.97
1	74	88.85
2	52	42.84
3	8	7.88
4	2	0.46
<i>Total</i>	200	200

On constate que le tableau contient des effectifs théoriques strictement inférieurs à 5, ce qui empêche l'utilisation d'un test du khi-deux.

On peut remédier à cet état en opérant le groupement "logique" des classes "3 et 4". Le tableau des effectifs observés et théoriques deviennent :

<i>Nombre d'as</i>	<i>Répartition observée</i>	<i>Répartition théorique</i>
0	64	59.97
1	74	88.85
2	52	42.84
3 ou 4	10	8.34
<i>Total</i>	200	200

Sous l'hypothèse nulle H_0 , la quantité :

$$\chi^2 = \sum_{i=0}^3 \frac{(o_i - t_i)^2}{t_i}$$

est une réalisation d'une variable du Khi-deux à :

$$4 - 1 = 3$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$\chi^2_{3,95} = 7.81$$

Et comme :

$$\chi^2 = 5.0418$$

On accepte alors H_0 au seuil $\alpha = 5\%$, c'est à dire, *le programme du statisticien est fiable.*

Tests d'Hypothèses

Moyennes et Variances

Exercice 1

Une série de cent mesures a donné comme résultat :

$$\left\{ \begin{array}{l} \sum_{i=1}^{100} x_i = 5200 \\ \sum_{i=1}^{100} \left[x_i - \frac{1}{100} \sum_{j=1}^{100} x_j \right]^2 = 396 \end{array} \right.$$

1. Estimer la moyenne et la variance.
2. Quel est, à 95%, l'intervalle de confiance de la moyenne ?
3. En supposant la variable mesurée gaussienne, déterminer, à 95%, l'intervalle de confiance de la variance.

Solution 1

1. Soit m l'estimation de la moyenne et s^2 celle de la variance.

On a :

$$\begin{aligned} m &= \frac{1}{100} \sum_{i=1}^{100} x_i \\ &= 52 \end{aligned}$$

et :

$$\begin{aligned} s^2 &= \frac{1}{99} \sum_{i=1}^{100} (x_i - m)^2 \\ &= 4 \end{aligned}$$

2. Au seuil α , l'intervalle de confiance de la moyenne est défini par :

$$\left[m - t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, m + t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

d'où l'intervalle de confiance à 95% :

$$[51.608, 52.392]$$

3. Au seuil α , l'intervalle de confiance de la variance est défini par :

$$\left[\frac{(n-1)}{\chi_{n-1;1-\alpha/2}^2} s^2, \frac{(n-1)}{\chi_{n-1;\alpha/2}^2} s^2 \right]$$

Pour $\alpha = 5\%$:

$$\begin{cases} \chi_{99;.025}^2 \simeq \chi_{100;.025}^2 = 74.2 \\ \chi_{99;.975}^2 \simeq \chi_{100;.975}^2 = 129.6 \end{cases}$$

d'où l'intervalle de confiance de l'écart-type à 95% :

$$[3.06, 5.34]$$

Exercice 2

La force de rupture d'un certain type de câble peut être assimilée à une variable aléatoire normale.

Des essais portant sur dix câbles ont donné une variance empirique s^2 de 1560 N². Construire un intervalle de confiance, à 95%, de l'écart-type de cette force de rupture.

Solution 2

Au seuil α , l'intervalle de confiance de l'écart-type est défini par :

$$\left[\sqrt{\frac{(n-1)}{\chi_{n-1;1-\alpha/2}^2}} s, \sqrt{\frac{(n-1)}{\chi_{n-1;\alpha/2}^2}} s \right]$$

Pour $\alpha = 5\%$:

$$\begin{cases} \chi_{9;.025}^2 = 2.7 \\ \chi_{9;.975}^2 = 19 \end{cases}$$

d'où l'intervalle de confiance de l'écart-type à 95% :

$$[27.18 \text{ N}, 72.11 \text{ N}]$$

Exercice 3

Une enquête statistique effectuée sur cent sujets permet de définir, à 95%, l'intervalle de confiance de la moyenne :

$$[49.6 - 50.4]$$

Dans quelles conditions aurait-il été possible que le résultat fût à 95% :

$$[49.8 - 50.2]$$

Solution 3

Il s'agit de déterminer la taille n' de l'échantillon à prélever pour que l'intervalle de confiance de la moyenne à 95% soit :

$$[49.8, 50.2]$$

sachant que pour un échantillon de taille $n = 100$, cet intervalle est :

$$[49.6, 50.4]$$

Puisque :

$$\left[m - t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, m + t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = [49.6, 50.4]$$

on en déduit :

$$\begin{aligned} m &= \frac{49.6 + 50.4}{2} \\ &= 50 \end{aligned}$$

et :

$$\begin{aligned} \sigma &= \frac{\sqrt{n} (50.4 - 49.6)}{2t_{1-\alpha/2}} \\ &\simeq 2.04 \end{aligned}$$

L'égalité :

$$50.2 = m + t_{1-\alpha/2} \frac{\sigma}{\sqrt{n'}}$$

implique :

$$\begin{aligned} n' &= \left(\frac{\sigma t_{1-\alpha/2}}{50.2 - m} \right)^2 \\ &= 400 \end{aligned}$$

Exercice 4

Pour déterminer le point de fusion moyen μ d'un certain alliage, on a procédé à neuf observations qui ont données une moyenne $m = 1040^\circ\text{C}$ et un écart-type $s = 16^\circ\text{C}$. Construire un intervalle de confiance de la moyenne μ à 95%.

Solution 4

Ici on a :

$$\begin{aligned} n &= 9 \\ m &= 1040^\circ\text{C} \\ s &= 16^\circ\text{C} \end{aligned}$$

Au seuil α , l'intervalle de confiance d'une telle moyenne est défini par :

$$\left[m - t_{n-1;1-\alpha/2} \frac{s}{\sqrt{n}}, m + t_{n-1;1-\alpha/2} \frac{s}{\sqrt{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{8;975} = 2.31$$

d'où l'intervalle de confiance à 95% :

$$[1027.68^\circ\text{C}, 1052.32^\circ\text{C}]$$

Exercice 5

La taille de 1200 conscrits du bureau de recrutement X a pour moyenne $\bar{X} = 172$ cm et pour écart-type $s_X = 6$ cm.

Les mêmes mesures effectuées sur les 250 conscrits du bureau de recrutement Y ont donné pour moyenne $\bar{Y} = 170$ cm et pour écart-type $s_Y = 5$ cm.

Que peut-on conclure ?

Solution 5

Testons au seuil α l'hypothèse nulle :

$$H_0 : \text{''les conscrits des bureaux de recrutement } X \text{ et } Y \text{ ont la même taille''}$$

Sous l'hypothèse nulle H_0 , la quantité :

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{975} = 1.96$$

Et comme :

$$\begin{aligned} t &= \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \\ &= 5.547 \end{aligned}$$

On rejette alors l'hypothèse nulle H_0 à 95% (même à 99%), c'est à dire, *les conscrits des bureaux de recrutement X et Y ont des tailles moyennes différentes.*

Exercice 6

On se propose de comparer le poids à la naissance chez une série de primipares (série 1) et une série de multipares (série 2) :

$$\text{Série 1 : } n_1 = 95 \quad m_1 = 3197 \text{ g} \quad s_1^2 = 210100 \text{ g}^2$$

$$\text{Série 2 : } n_2 = 105 \quad m_2 = 3410 \text{ g} \quad s_2^2 = 255400 \text{ g}^2$$

Que peut-on conclure ?

Solution 6

Testons au seuil α l'hypothèse nulle :

H_0 : "les primipares et les multipares ont le même poids moyen à la naissance"

Sous l'hypothèse nulle H_0 , la quantité :

$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite. Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

Et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= -3.1256 \end{aligned}$$

On rejette alors l'hypothèse nulle H_0 , à 95% (même à 99%), c'est à dire, *les primipares et les multipares n'ont pas le même poids moyen à la naissance*

Exercice 7

Chez cent sujet normaux, on dose l'acide urique, les résultats sont :

$$\begin{cases} m_1 &= 53.3 \text{ mg/l} \\ s_1 &= 9.1 \text{ mg/l} \end{cases}$$

Chez cent sujet atteints de la maladie de goutte, le même dosage de l'acide urique fournit les résultats :

$$\begin{cases} m_2 &= 78.6 \text{ mg/l} \\ s_2 &= 13.1 \text{ mg/l} \end{cases}$$

Que peut-on conclure ?

Solution 7

Testons au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la maladie de goutte n'a pas d'influence sur la dose de l'acide urique"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

peut être considérée comme une réalisation d'une variable aléatoire normale centrée réduite.

Pour $\alpha = 5\%$, on a :

$$t_{.975} = 1.96$$

et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= -15.862 \end{aligned}$$

On rejette l'hypothèse nulle H_0 à 95% (même à 99.99%), c'est à dire, *la maladie de goutte a une influence sur la dose de l'acide urique.*

Exercice 8

On admet que la valeur moyenne de la glycémie du sujet normal est 1 g/l.

Sur 17 sujets, on a trouvé une moyenne de .965 g/l et un écart-type estimé de .108 g/l.

Cette valeur peut-elle être considérée comme différente du taux normal ?

Solution 8

Testons au seuil α , l'hypothèse nulle :

$$H_0 : \text{"la valeur est normale"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m - \mu}{\frac{s}{\sqrt{n}}}$$

est une réalisation de la variable aléatoire T_{n-1} de Student à :

$$n - 1 = 16$$

degrés de liberté.

Pour $\alpha = 5\%$, on a :

$$t_{16; 975} = 2.12$$

et comme :

$$\begin{aligned} t &= \frac{m - \mu}{\frac{s}{\sqrt{n}}} \\ &= -1.3362 \end{aligned}$$

on accepte l'hypothèse nulle H_0 au seuil $\alpha = 5\%$, c'est à dire, *la valeur est normale*.

Exercice 9

Dans un échantillon de 17 prématurés, la moyenne du *Na*-plasmatique est :

$$\begin{cases} m_1 &= 133 \\ s_1^2 &= 81.2 \end{cases}$$

Soit un autre échantillon de 25 dysmaturés, dans lequel la moyenne du *Na*-plasmatique est :

$$\begin{cases} m_2 &= 136 \\ s_2^2 &= 56.57 \end{cases}$$

Que peut-on conclure ?

Solution 9

Testons d'abord, au seuil $\alpha = 10\%$, l'hypothèse nulle d'égalité des variances du *Na*-plasmatique chez les prématurés et les dysmaturés.

Sous cette hypothèse, la quantité :

$$f = \frac{s_1^2}{s_2^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_1 - 1, n_2 - 1) = (16, 24)$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$F_{16,24;.95} = 2.09$$

Et comme :

$$\begin{aligned} f &= \frac{s_1^2}{s_2^2} \\ &= 1.4354 \end{aligned}$$

on accepte donc l'hypothèse d'égalité des variances des deux populations.

Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= 66.42 \end{aligned}$$

et testons l'hypothèse nulle :

H_0 : "les prématurés et les dysmaturés ont la même moyenne du Na-plasmatic"

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de Student à :

$$n_1 + n_2 - 2 = 40$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$t_{40;.95} = 1.68$$

Et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= -1.17 \end{aligned}$$

On accepte l'hypothèse nulle H_0 au seuil $\alpha = 10\%$, c'est à dire, *les prématurés et les dysmaturés ont la même moyenne du Na-plasmatic* estimée par :

$$\begin{aligned} m &= \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2} \\ &= 134.79 \end{aligned}$$

Exercice 10

Lorsqu'une machine est bien réglée, elle produit des pièces dont le diamètre D est une variable gaussienne de moyenne 25 mm.

Deux heures après le réglage de la machine, on a prélevé au hasard neuf pièces.

Leurs diamètres ont pour mesure en mm :

$$22 \quad 23 \quad 21 \quad 25 \quad 24 \quad 23 \quad 22 \quad 26 \quad 21$$

Que peut-on conclure quant à la qualité du réglage après deux heures de fonctionnement de la machine ?

Solution 10

Calculons d'abord les estimations m et s^2 de la moyenne et de la variance sur cet échantillon de taille $n = 9$.

On a :

$$\begin{aligned} m &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= 23 \text{ mm} \end{aligned}$$

et :

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \\ &= 3 \text{ mm}^2 \end{aligned}$$

Testons l'hypothèse nulle :

$$H_0 : \text{"la machine est bien réglée"}$$

Sous l'hypothèse nulle H_0 , la quantité :

$$t = \frac{m - \mu}{\frac{s}{\sqrt{n}}}$$

est une réalisation d'une variable aléatoire de Student à :

$$n - 1 = 8$$

degrés de liberté : T_8 .

Pour $\alpha = 5\%$, on a :

$$t_{8;975} = 2.31$$

et comme :

$$\begin{aligned} t &= \frac{m - \mu}{\frac{s}{\sqrt{n}}} \\ &= -3.4641 \end{aligned}$$

On rejette l'hypothèse nulle H_0 à 95% (même à 99%), c'est à dire, *le réglage de la machine est rompu*.

Exercice 11

Si l'écart-type de la durée de vie d'un modèle de lampe électrique est estimé à cent heures, quelle doit être la taille de l'échantillon à prélever pour que l'erreur sur l'estimation de la durée de vie moyenne n'exède pas vingt heures et ce avec une probabilité de 95% puis 99% ?

Solution 11

L'erreur sur l'estimation de la moyenne est donnée par :

$$t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

(1) Pour $\alpha = 5\%$, on a :

$$t_{1-\alpha/2} = 1.96$$

d'où :

$$t_{1-\alpha/2} \frac{s}{\sqrt{n}} \leq 20 \implies n \geq 97$$

(2) Pour $\alpha = 1\%$, on a :

$$t_{1-\alpha/2} = 2.57$$

d'où :

$$t_{1-\alpha/2} \frac{s}{\sqrt{n}} \leq 20 \implies n \geq 166$$

Exercice 12

Une machine fabrique des rondelles dont le diamètre D est une variable gaussienne. On prélève au hasard un échantillon de huit rondelles.

Leurs diamètres ont pour mesure en mm :

$$20.1 \quad 19.9 \quad 19.7 \quad 20.2 \quad 20.1 \quad 23.1 \quad 22.6 \quad 19.8$$

Construire à 95% puis 99% les intervalles de confiance de la moyenne et de la variance.

Solution 12

Calculons d'abord les estimations m et s^2 de la moyenne et de la variance sur cet échantillon de taille $n = 8$.

On a :

$$\begin{aligned} m &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= 20.6875 \text{ mm} \end{aligned}$$

et

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \\ &= 1.827 \text{ mm}^2 \end{aligned}$$

1. L'intervalle de confiance de la moyenne à $1 - \alpha$ est :

$$\left[m - t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}}, m + t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}} \right]$$

(a) Pour $\alpha = 5\%$, on a :

$$t_{7; 975} = 2.36$$

d'où l'intervalle :

$$[19.163, 22.212]$$

(b) Pour $\alpha = 1\%$, on a :

$$t_{7; 995} = 3.5$$

d'où l'intervalle :

$$[18.427, 22.948]$$

2. L'intervalle de confiance de la variance à $1 - \alpha$ est :

$$\left[\frac{(n-1)s^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1;\alpha/2}^2} \right]$$

(a) Pour $\alpha = 5\%$, on a :

$$\begin{cases} \chi_{7;.025}^2 = 1.69 \\ \chi_{7;.975}^2 = 16 \end{cases}$$

d'où l'intervalle :

$$[.79931, 7.5675]$$

(b) Pour $\alpha = 1\%$, on a :

$$\begin{cases} \chi_{7;.005}^2 = .989 \\ \chi_{7;.995}^2 = 20.3 \end{cases}$$

d'où l'intervalle :

$$[.63, 12.931]$$

Exercice 13

On effectue un dosage par deux méthodes différentes A et B .

On obtient les résultats suivants :

Méthode A .6 .65 .7 .7 .7 .7 .75 .8 .8

Méthode B .6 .6 .65 .65 .7 .6 .75 .8 .8

Peut-on considérer que les deux méthodes sont équivalentes ?

Solution 13

Calculons les estimations (m_1, s_1^2) de (μ_1, σ_1^2) et (m_2, s_2^2) de (μ_2, σ_2^2) :

$$\begin{cases} m_1 = \frac{1}{9} \sum_{i=1}^9 x_{1i} = .71 \\ s_1^2 = \frac{1}{8} \sum_{i=1}^9 (x_{1i} - m_1)^2 = .004 \end{cases}$$

et :

$$\begin{cases} m_2 = \frac{1}{9} \sum_{i=1}^9 x_{2i} = .68 \\ s_2^2 = \frac{1}{8} \sum_{i=1}^9 (x_{2i} - m_2)^2 = .007 \end{cases}$$

Testons d'abord, au seuil $\alpha = 10\%$, l'hypothèse nulle d'égalité des variances des deux méthodes de dosage.

Sous cette hypothèse, la quantité :

$$f = \frac{s_2^2}{s_1^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_2 - 1, n_1 - 1) = (8, 8)$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$F_{8,8;.95} = 3.44$$

et comme :

$$\begin{aligned} f &= \frac{s_2^2}{s_1^2} \\ &= 1.75 \end{aligned}$$

On accepte donc l'hypothèse d'égalité des variances des deux populations.

Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= 0.0055 \end{aligned}$$

et testons l'hypothèse nulle :

$$H_0 : \text{"les deux méthodes de dosage sont équivalentes."}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de Student à :

$$n_1 + n_2 - 2 = 16$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$t_{16;95} = 1.75$$

et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= 0.86 \end{aligned}$$

on accepte l'hypothèse nulle H_0 au seuil $\alpha = 10\%$, c'est à dire, *les deux méthodes de dosage sont équivalentes.*

Exercice 14

Dans deux types de forêts, on a mesuré les hauteurs de treize et quatorze peuplements choisis au hasard et indépendamment dans le but de vérifier si les hauteurs de ces deux types d'arbres sont ou ne sont pas égales.

Les résultats sont les suivants :

Type 1 : 22.5 22.9 23.7 24.0 24.4 24.5 26.0

26.2 26.4 26.7 27.4 28.6 28.7

Type 2 : 23.4 24.4 24.6 24.9 25.0 26.2 26.3

26.8 26.8 26.9 27.0 27.6 27.7 27.8

On admet que les hauteurs de ces deux types d'arbres sont des variables gaussiennes $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$.

Que peut-on conclure ?

Solution 14

Calculons les estimations (m_1, s_1^2) de (μ_1, σ_1^2) et (m_2, s_2^2) de (μ_2, σ_2^2) :

$$\left\{ \begin{aligned} m_1 &= \frac{1}{13} \sum_{i=1}^{13} x_{1i} = 25.538 \\ s_1^2 &= \frac{1}{12} \sum_{i=1}^{13} (x_{1i} - m_1)^2 = 4.1576 \end{aligned} \right.$$

et :

$$\begin{cases} m_2 = \frac{1}{14} \sum_{i=1}^{14} x_{2i} = 26.1 \\ s_2^2 = \frac{1}{13} \sum_{i=1}^{14} (x_{2i} - m_2)^2 = 1.9431 \end{cases}$$

Testons d'abord, au seuil $\alpha = 10\%$, l'hypothèse nulle d'égalité des variances des hauteurs des deux types d'arbres.

Sous cette hypothèse, la quantité :

$$f = \frac{s_1^2}{s_2^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_1 - 1, n_2 - 1) = (12, 13)$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$F_{12,13;95} = 2.6$$

et comme :

$$\begin{aligned} f &= \frac{s_1^2}{s_2^2} \\ &= 2.1398 \end{aligned}$$

on accepte donc l'hypothèse d'égalité des variances des hauteurs des deux types d'arbres.

Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= 3.0062 \end{aligned}$$

et testons l'hypothèse nulle :

$$H_0 : \text{''les deux types d'arbres ont la même hauteur''}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de Student à :

$$n_1 + n_2 - 2 = 25$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$t_{25,95} = 1.71$$

et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= -0.84155 \end{aligned}$$

on accepte l'hypothèse nulle H_0 au seuil $\alpha = 10\%$, c'est à dire, *les deux types d'arbres ont la même hauteur moyenne estimée par :*

$$\begin{aligned} m &= \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2} \\ &= 25.829 \end{aligned}$$

Exercice 15

On considère deux variétés de maïs M_1 et M_2 dont les rendements sont des variables aléatoires gaussiennes $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$.

Afin de comparer les rendements de ces deux variétés de maïs, on a choisi de cultiver dans neuf stations différentes des parcelles voisines encadrées de l'une ou l'autre des deux variétés. On a observé les rendements suivants :

Station	1	2	3	4	5	6	7	8	9
Variété 1	39.6	32.4	33.1	27	36	32	25.9	32.4	33.2
Variété 2	39.2	33.1	32.4	25.2	33.1	29.5	24.1	29.2	34.1

Que peut-on conclure ?

Solution 15

Calculons les estimations (m_1, s_1^2) de (μ_1, σ_1^2) et (m_2, s_2^2) de (μ_2, σ_2^2) :

$$\begin{cases} m_1 = \frac{1}{13} \sum_{i=1}^{13} x_{1i} = 32.4 \\ s_1^2 = \frac{1}{12} \sum_{i=1}^{13} (x_{1i} - m_1)^2 = 17.188 \end{cases}$$

et :

$$\begin{cases} m_2 = \frac{1}{14} \sum_{i=1}^{14} x_{2i} = 31.1 \\ s_2^2 = \frac{1}{13} \sum_{i=1}^{14} (x_{2i} - m_2)^2 = 21.785 \end{cases}$$

Testons d'abord, au seuil $\alpha = 10\%$, l'hypothèse nulle d'égalité des variances des rendements des deux variétés de maïs.

Sous cette hypothèse, la quantité :

$$f = \frac{s_2^2}{s_1^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_2 - 1, n_1 - 1) = (8, 8)$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$F_{8,8;95} = 3.44$$

et comme :

$$\begin{aligned} f &= \frac{s_2^2}{s_1^2} \\ &= 1.2675 \end{aligned}$$

On accepte donc l'hypothèse d'égalité des variances des hauteurs des deux types d'arbres.

Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= \frac{s_1^2 + s_2^2}{2} \\ &= 19.4865 \end{aligned}$$

et testons l'hypothèse nulle :

$$H_0 : \text{''les deux variétés de maïs ont le même rendement''}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de Student à :

$$n_1 + n_2 - 2 = 16$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$t_{16;95} = 1.75$$

et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= .42892 \end{aligned}$$

on accepte l'hypothèse nulle H_0 au seuil $\alpha = 10\%$, c'est à dire, *les deux variétés de maïs ont le même rendement moyen estimé par :*

$$\begin{aligned} m &= \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2} \\ &= 31.75 \end{aligned}$$

Exercice 16

Le relevé des températures journalières minimales de deux stations S_1 et S_2 , au cours de neuf journées consécutives a fourni les valeurs suivantes en $^{\circ}\text{C}$:

Station 1	12	8	9	10	11	13	10	7	10
Station 2	7	11	10	6	8	11	12	9	7

On admet que la distribution des températures journalières minimales des deux stations S_1 et S_2 sont des variables gaussiennes $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$.

1. Déterminer les estimations des moyennes et des variances des températures journalières minimales des deux stations S_1 et S_2 .
2. Construire, au seuil $\alpha = 5\%$, les intervalles de confiance de ces estimations.

3. Peut-on admettre, au seuil $\alpha = 10\%$, l'hypothèse selon laquelle les températures journalières minimales moyennes des deux stations S_1 et S_2 sont identiques ?

Solution 16

1. Calculons les estimations (m_1, s_1^2) de (μ_1, σ_1^2) et (m_2, s_2^2) de (μ_2, σ_2^2) .

On a :

$$\begin{cases} m_1 = \frac{1}{9} \sum_{i=1}^{11} x_{1i} = 10^\circ\text{C} \\ s_1^2 = \frac{1}{8} \sum_{i=1}^{11} (x_{1i} - m_1)^2 = 3.5 \end{cases}$$

et :

$$\begin{cases} m_2 = \frac{1}{9} \sum_{i=1}^{10} x_{2i} = 9^\circ\text{C} \\ s_2^2 = \frac{1}{8} \sum_{i=1}^{10} (x_{2i} - m_2)^2 = 4.5 \end{cases}$$

- (a) L'intervalle de confiance de μ_1 à $1 - \alpha$ est défini par :

$$\left[m_1 - t_{n-1; 1-\alpha/2} \frac{s_1}{\sqrt{n}}, m_1 + t_{n-1; 1-\alpha/2} \frac{s_1}{\sqrt{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{8; 975} = 2.31$$

d'où l'intervalle :

$$[8.56^\circ\text{C}, 11.44^\circ\text{C}]$$

- (b) L'intervalle de confiance de σ_1^2 à $1 - \alpha$ est défini par :

$$\left[\frac{(n-1) s_1^2}{\chi_{n-1; 1-\alpha/2}^2}, \frac{(n-1) s_1^2}{\chi_{n-1; \alpha/2}^2} \right]$$

Pour $\alpha = 5\%$, on a :

$$\begin{cases} \chi_{8; 025}^2 = 2.18 \\ \chi_{8; 975}^2 = 17.5 \end{cases}$$

d'où l'intervalle :

$$[1.6, 12.8]$$

(c) L'intervalle de confiance de μ_2 à $1 - \alpha$ est défini par :

$$\left[m_2 - t_{n-1;1-\alpha/2} \frac{s_2}{\sqrt{n}}, m_2 + t_{n-1;1-\alpha/2} \frac{s_2}{\sqrt{n}} \right]$$

Pour $\alpha = 5\%$, on a :

$$t_{8;975} = 2.31$$

d'où l'intervalle :

$$[7.37^\circ\text{C}, 10.63^\circ\text{C}]$$

(d) L'intervalle de confiance de σ_2^2 à $1 - \alpha$ est défini par :

$$\left[\frac{(n-1)s_2^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s_2^2}{\chi_{n-1;\alpha/2}^2} \right]$$

Pour $\alpha = 5\%$, on a :

$$\begin{cases} \chi_{8;025}^2 = 2.18 \\ \chi_{8;975}^2 = 17.5 \end{cases}$$

d'où l'intervalle :

$$[2.06, 16.51]$$

2. Testons d'abord, au seuil $\alpha = 10\%$, l'hypothèse nulle d'égalité des variances des températures journalières minimales des deux stations S_1 et S_2 .

Sous cette hypothèse, la quantité :

$$f = \frac{s_2^2}{s_1^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_2 - 1, n_1 - 1) = (8, 8)$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$F_{8,8;95} = 3.44$$

et comme :

$$f = \frac{s_2^2}{s_1^2} = 1.29$$

On accepte donc l'hypothèse d'égalité des variances.

Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= \frac{s_1^2 + s_2^2}{2} \\ &= 4 \end{aligned}$$

et testons l'hypothèse nulle :

H_0 : "les températures journalières minimales moyennes des deux stations S_1 et S_2 sont identiques"

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de Student à :

$$n_1 + n_2 - 2 = 16$$

degrés de liberté.

Pour $\alpha = 10\%$, on a :

$$t_{16; .95} = 1.75$$

et comme :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.0607$$

On accepte l'hypothèse nulle H_0 au seuil $\alpha = 10\%$, c'est à dire, les températures journalières minimales moyennes des deux stations S_1 et S_2 sont identiques.

Cette température moyenne peut être estimée par :

$$\begin{aligned} m &= \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2} \\ &= \frac{m_1 + m_2}{2} \\ &= 9.5 \end{aligned}$$

Exercice 17

On étudie l'effet d'une substance sur la croissance d'une tumeur greffée.

Les résultats sont consignés sur le tableau ci-dessous donnant la surface de la tumeur au 20^{ème} jour après sa greffe :

<i>Surface</i>	5.5	6	6.5	7	7.5	8
<i>Témoins</i>	1	2	3	8	4	3
<i>Traités</i>	4	4	8	3	1	1

Le traitement a-t-il un effet significatif sur la surface tumorale ?

On suppose que la surface tumorale est distribuée selon des lois normales $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$ chez les témoins et les traités respectivement.

Solution 17

Calculons les estimations (m_1, s_1^2) de (μ_1, σ_1^2) et (m_2, s_2^2) de (μ_2, σ_2^2) .

On a :

$$\begin{cases} m_1 = \frac{1}{21} \sum_{i=1}^6 n_{1i} x_i = 7 \\ s_1^2 = \frac{1}{20} \sum_{i=1}^6 n_{1i} (x_i - m_1)^2 = .45 \end{cases}$$

et :

$$\begin{cases} m_2 = \frac{1}{21} \sum_{i=1}^6 n_{2i} x_i = 6.4048 \\ s_2^2 = \frac{1}{20} \sum_{i=1}^6 n_{2i} (x_i - m_2)^2 = .87972 \end{cases}$$

Testons d'abord, au seuil $\alpha = 2\%$, l'hypothèse nulle d'égalité des variances des surfaces tumorales chez les populations des témoins et des traités.

Sous cette hypothèse, la quantité :

$$f = \frac{s_2^2}{s_1^2}$$

est une réalisation d'une variable aléatoire de Fisher à :

$$(n_2 - 1, n_1 - 1) = (20, 20)$$

degrés de liberté.

Pour $\alpha = 2\%$, on a :

$$F_{20,20;.99} = 2.94$$

et comme :

$$\begin{aligned} f &= \frac{s_2^2}{s_1^2} \\ &= 1.9549 \end{aligned}$$

on accepte donc l'hypothèse d'égalité des variances des deux populations. Calculons maintenant l'estimation commune s^2 de cette variance :

$$\begin{aligned} s^2 &= \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \\ &= .66486 \end{aligned}$$

et testons l'hypothèse nulle :

$$H_0 : \text{"le traitement est sans effet sur la croissance de la surface tumorale"}$$

Sous cette hypothèse, la quantité :

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

est une réalisation de la variable aléatoire de *Student* à :

$$n_1 + n_2 - 2 = 40$$

degrés de liberté.

Pour $\alpha = 2\%$, on a :

$$t_{40;.99} = 2.42$$

et comme :

$$\begin{aligned} t &= \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= 2.831 \end{aligned}$$

on rejette l'hypothèse nulle H_0 à 98%, c'est à dire, *le traitement a une influence sur la croissance de la surface tumorale.*

