UNIVERSITÉ NATIONALE DES SCIENCES, TECHNOLOGIES, INGÉNIERIE ET MATHÉMATIQUES (UNSTIM)

ÉCOLE NATIONALE SUPÉRIEURE DE GÉNIE MATHÉMATIQUE ET MODÉLISATION (ENSGMM)

****

Travaux Pratique de STATISTIQUE INFERENTIELLE

Unité d'enseignement : STATISTIQUE INFERENTIELLE

Membres :

- AKAKPO Expéra
- HOUNSOUNOU Marius
- SOGBEDJI Prosper

Sous la supervision :
Dr (MA) Nicodème ATCHADE

2024-2025

Table des matières

1 A3.

Soient

$$X_{1i} \sim \mathcal{N}\left(\mu_1, \sigma^2\right)$$
$$X_{2j} \sim \mathcal{N}\left(\mu_2, \sigma^2\right)$$

(i) Trouvons un estimateur sans biais $\hat{\sigma}^2$ de la variance commune σ^2

Méthode du maximum de vraissemblance

Pour la variable aléatoire suivant la loi $\mathcal{N}(\mu_1, \sigma^2)$, considérons s_1^2 unbiased l'un estimateur sans biais de σ^2.

$$L(x_{11}, x_{12}, \ldots, x_{1n}) = \prod_{i=1}^{n} f(x_i) \tag{1}$$
$$= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\frac{1}{2}\left(\frac{x_{1i}-\mu_1}{\sigma}\right)^2} \tag{2}$$

alors,

$$\ln L(x_{1i}, \sigma^2) = -n\ln\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_{1i}-\mu_1}{\sigma}\right)^2 \tag{3}$$
$$\frac{d\ln L(x_{1i}, \sigma^2)}{d\sigma} = -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{n}(x_{1i}-\mu_1)^2 \tag{4}$$

$\frac{d\ln L(x_{1i},\sigma)}{d\sigma} = 0, \quad \Leftrightarrow \quad -\frac{n}{\hat{\sigma}_{\text{MLE}}} + \frac{1}{\hat{\sigma}^3_{\text{MLE}}}\sum_{i=1}^{n}(x_i-\mu_1)^2 = 0$ donc.

$$\boxed{\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(X_{1i}-\bar{X}_n)^2} \tag{5}$$

De plus $\frac{d2\ln L(x_{1i},\sigma)}{d\sigma^2} = \frac{n}{\hat{\sigma}^2} - \frac{3}{\hat{\sigma}^4}\sum(x_{1i}-\mu_1)^2 < 0$
$\hat{\sigma}^2_{\text{MLE}}$ est biaisé.

$$E(\hat{\sigma}^2_{\text{MLE}}) = \frac{1}{n}\sum E[(X_{1i}-\bar{X}_n)^2] \tag{6}$$
$$= E\left[\frac{1}{n}\sum X_{1i}^2 - \bar{X}_n^2\right] \tag{7}$$
$$= E(X_{1i}^2) - E(\bar{X}_n^2) \tag{8}$$

$$\text{or,}\quad E(X_{1i}^2) = \text{Var}(X_{1i}) + E(X_{1i})^2 \tag{9}$$
$$= \sigma^2 + \mu_1^2 \tag{10}$$

$$E(\bar{X}_n^2) = \text{Var}(\bar{X}_n) + E(\bar{X}_n)^2 \tag{11}$$
$$= \frac{\sigma^2}{n} + \mu_1^2 \quad \text{thus,} \quad E(\hat{\sigma}^2_{\text{MLE}}) = \frac{n-1}{n}\sigma^2 \tag{12}$$

Pour un estimateur sans biais de σ^2, nous avons $E(s_1^2) = \sigma^2$

$$E(s_1^2) = \sigma^2 \Leftrightarrow E\left(\frac{n-1}{n}s_1^2\right) = \frac{n-1}{n}\sigma^2 \tag{13}$$
$$\Leftrightarrow E\left(\frac{n-1}{n}s_1^2\right) = E\left(\hat{\sigma}^2_{\text{MLE}}\right) \tag{14}$$
$$s_1^2 = \frac{n}{n-1}\hat{\sigma}^2_{\text{MLE}} \tag{15}$$

So

$$\boxed{s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{1i} - \bar{X}_n\right)^2} \tag{16}$$

Pour la variable aléatoire suivant la loi $\mathcal{N}(\mu_2, \sigma^2)$, considérons s_2^2 unbiased l'un estimateur sans biais de σ^2

De façon analogique,

$$\boxed{s_2^2 = \frac{1}{n-1}\sum_{j=1}^{m}\left(X_{2j} - \bar{X}_m\right)^2} \tag{17}$$

Donc un estimateur sans biais $\hat{\sigma}^2$ de la variance σ^2, est :

$$\boxed{\hat{\sigma}^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}} \tag{18}$$

Distribution d'une version mise à l'échelle $\hat{\sigma}^2$

$$\boxed{(n+m-2)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n+m-2}} \tag{19}$$

(ii) Statistique de test et loi d'échantillonnage sous H_0

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 0$$

Une statistique appropriée (test de Student bilatéral, variances égales) est : $\bar{X}_n - \bar{X}_m \sim \mathcal{N}\left(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$ et

$$T = \frac{\bar{X}_n - \bar{X}_m}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \tag{20}$$

(iii) An approximate rejection region for the test is :

Pour un test bilatéral au niveau α la région de rejet est

$$\boxed{W = \{|T| > t_{n+m-2,\frac{\alpha}{2}}\}} \tag{21}$$

où α est le risque d'erreur et $t_{n+m-2,\frac{\alpha}{2}}$ est le quantile α de la loi de student de dégré de liberté $n+m-2$

(iv) Application numérique $n = m = 10, \quad \bar{x}_1 = 46.0 \quad \bar{x}_2 = 48.1 \quad s_1^2 = 2.04^2 \quad s_2^2 = 4.92^2$

$$\hat{\sigma}^2 = \frac{9[2.04^2 + 4.92^2]}{18} \tag{22}$$

$$\hat{\sigma}^2 = 3.924 \tag{23}$$

$$T = \frac{46.0 - 48.1}{\sqrt{3.924 \times \frac{2}{5}}} \tag{24}$$

$$= -2.32 \tag{25}$$

$$\boxed{t_{\text{obs}} = -2.32}$$

- Pour $\alpha = 0.05$
  $t_{18,0.025} = 2.1009, \quad |T| > t_{18,0.025}.$ donc on rejette H_0 au niveau $\alpha = 0.05$.
- Pour $\alpha = 0.01$
  $t_{18,0.005} = 2.8784, \quad |T| < t_{18,0.005}.$ donc on ne rejette pas H_0 au niveau $\alpha = 0.05$.

Avec les données fournies, il y a une différence significative entre μ_1 et μ_2 au niveau $\alpha = 0.05$, mais pas au niveau strict $\alpha = 0.01$.

2 A4.

Echantillon indépendant de taille $n = 1000$; proportion observée soutenant Labour $p = 0.30$. On suppose les observations Bernoulli i.i.d.

(i) Formule générale d'un intervalle de confiance approximatif à $100(1-\alpha)\%$ pour p

Pour un échantillon de taille n et proportion empirique $\hat{p}$, un intervalle de confiance approché (par la CLT) à niveau $100(1-\alpha)\%$ est

$$\boxed{IC = \hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}},$$

où $z_{1-\alpha/2}$ est le quantile $1-\alpha/2$ de la loi normale standard

Commentaire distributionnel : cette construction repose sur l'approximation par la loi normale de la variable $\hat{p}$ (par le théorème central limite) : $\hat{p} \approx N(p, p(1-p)/n)$ pour n grand. L'approximation est raisonnable si np et $n(1-p)$ sont suffisamment grands (classiquement ≥ 5 ou mieux ≥ 10).

(ii) Intervalle de confiance à 95% pour la proportion

Nous avons

$$\boxed{IC = p \pm \mu_{1-\frac{\alpha}{2}}\sqrt{\frac{p\,(1-p)}{n}}}$$

avec $p = 0.3$ et $1 - \frac{\alpha}{2} = 0.975$. $\mu_{1-\frac{\alpha}{2}} = 1.96$, alors $IC = 0.3 \pm 0.0284$.

$$\boxed{IC = [0.2716; 0.3284]}$$

(iii) Probabilité qu'au moins 150 personnes soutiennent Labour dans un nouvel échantillon

Nous avons $X \backsim \mathcal{N}\left(np, np(1-p)\right)$

$$\begin{aligned}
\mathbb{P}\left(X \geq 150\right) &= \mathbb{P}\left(Z \geq \frac{150-np}{\sqrt{np(1-p)}}\right) \text{ avec } Z \backsim \mathcal{N}\left(0,1\right) \\
&= \mathbb{P}\left(Z \geq \frac{150-np}{\sqrt{np(1-p)}}\right) \\
&= \mathbb{P}\left(Z \geq 0.99\right) \\
&= 1 - \phi\left(0.99\right), \quad \phi - \text{ loi normal standard} \\
\mathbb{P}\left(X \geq 150\right) &= 0.1611
\end{aligned}$$