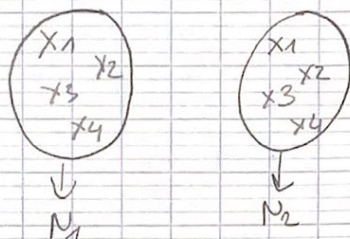


TDS - Statistiques

Exercice 1



1. Les tirages sont indépendants, X_1 et X_2 indépendants ← une loi quelconques
2. La loi de X_1 et X_2

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

$$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

$$\sigma_1 \text{ et } \sigma_2 = \text{connus (1)}$$

$$\text{1. connus } \sigma_1 = \sigma_2 = \sigma \text{ (mais écarts sont égaux)}$$

$$\text{2. un connu (3)}$$
3. Ecart-types?

Cas 1

$X_1, X_2 \sim \mathcal{N}(\mu, \sigma)$ et on connaît σ

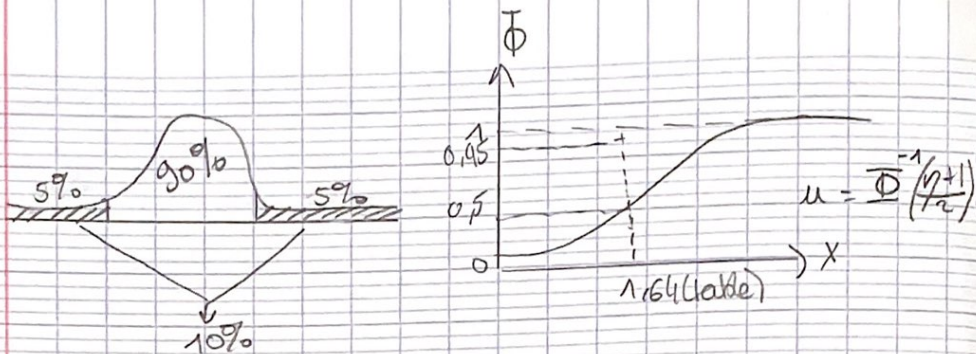
$$E[\bar{X}_1 - \bar{X}_2] = E[\bar{X}_1] - E[\bar{X}_2] = \mu_1 - \mu_2$$

$$\text{Var}[\bar{X}_1 - \bar{X}_2] = \text{Var}[\bar{X}_1] + (-1)^2 \text{Var}[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

$$P\left(-u \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq u\right) \approx \eta \quad \begin{array}{l} \text{rela} \\ \eta = 0.9 \\ \text{↳ niveau de confiance} \end{array}$$



$$\begin{aligned}\Phi(u) - \Phi(-u) &= 0.9 \Leftrightarrow 2\Phi(u) - 1 = \eta \\ \Phi(u) &= \frac{\eta+1}{2} = \frac{0.9+1}{2} \\ &= 0.95\end{aligned}$$

$$(\bar{X}_1 - \bar{X}_2) - u \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + u \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

↳ Ici, on peut pas calculer l'intervalle comme on connaît pas σ (hypothèse fautive)

Cas 2 $\sigma_1 = \sigma_2 = \sigma$ inconnus

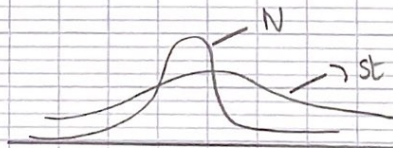
$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$$

$$\text{Var}[\bar{X}_1 - \bar{X}_2] = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Il faut estimer σ^2 :

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad S_1^2, S_2^2 \text{ étant les variances empiriques}$$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{St}(n_1 + n_2 - 2)$$



$$P(-t \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t) \quad \eta$$

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$*(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq N_1 - N_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$t = F^{-1}\left(\frac{\eta+1}{2}\right)$$

$$n_1 = 10, \bar{X}_1 = 1207,5, \bar{X}_2 = 716,62$$

$$n_2 = 8, S_1 = 488,06, S_2 = 232,06, S^2 = \frac{1}{n_1+n_2-2} \sum (X_i - \bar{X})^2$$

$$= \frac{1}{n_1+n_2-2} \sum X_i^2 - \bar{X}^2$$

$$S_p^2 = \frac{9 \times 488,06^2 + 7 \times 232,06^2}{10+8-2} = 157500,4$$

$$S_p = 396,93$$

$$\frac{\eta+1}{2} = \frac{0,9+1}{2} = 0,95$$

↳ dans la table de $t(k)$

$$t = F(0,95) = 1,7459$$

$$*(1207,5 - 716,62) - 1,75 \times 396,93 \sqrt{\frac{1}{10} + \frac{1}{8}} \leq N_1 - N_2$$

$$\leq (1207,05 - 716,62) + 1,75 \times 396,93 \sqrt{\frac{1}{10} + \frac{1}{8}}$$

$$171,72 \leq N_1 - N_2 \leq 810,03$$

↳ hypothèse $\sigma_1 = \sigma_2 = \sigma$ est peu réaliste

Cas 3 σ_1^2, σ_2^2 inconnus et $\sigma_1^2 \neq \sigma_2^2$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (N_1 - N_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim St(r)$$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$r = \frac{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}{\frac{\sigma_1^4}{(n_1-1)^2} + \frac{\sigma_2^4}{(n_2-1)^2}} \approx 17 \rightarrow k = 17, 4$$

↳ difference - variance
- paramètre de la loi $st(t)$

$$(1207,05 - 716,62) - 1,74 \sqrt{\frac{214383}{9} + \frac{47122}{7}} \leq \mu_1 - \mu_2$$

$$< (1207,05 - 716,62) + 1,74 \sqrt{\frac{214383}{9} + \frac{47122}{7}}$$

$$186,74 \leq \mu_1 - \mu_2 \leq 795,01$$

Exercice 2

$$D \sim cr(m=2, \sigma^2)$$

$$S'^2 = \frac{1}{n} \sum_{i=1}^n (d_i - m)^2 = 0,13$$

n = taille de l'échantillon = 30

On va déterminer une IC unilatérale de type $\sigma^2 > A$, avec un niveau de confiance de 0,95 %

D'après le cours $\sigma^2 > \frac{n \hat{\sigma}^2}{R_1}$

D'après la table de $\chi^2(30)$ on a pour 95% $R = 43,77$.

$$\text{donc } \sigma^2 > \frac{30 \times 0,13}{43,77} = 0,089$$

Exercice 3

1) Soit X_i une VA Bernoulli

$$X = \begin{cases} 1 & \text{fumeur} \\ 0 & \text{non} \end{cases}$$

$F_n = \sum \frac{X_i}{n}$ la proportion d'échantillon

On sait la loi normale
 $f_n \sim \text{cr}(p, \sqrt{\frac{pq}{n}})$

$$f_n \sim \text{cr}(0,54; \sqrt{\frac{0,54 \times 0,46}{100}})$$

$$f_n \sim \text{cr}(0,54; 0,05)$$

$$P(f_n > 65\%) \Rightarrow P\left(\frac{f_n - 0,54}{0,05} > \frac{0,65 - 0,54}{0,05}\right)$$

$$\begin{aligned} \Rightarrow P(Z > 2,2) &= 1 - P(Z \leq 2,2) \\ &= 1 - 0,9861 \\ &= 0,0139 \end{aligned}$$

2.1 Soit X le nombre de cigarettes

on a $m = 16$ et $\sigma = 4,2$

On cherche $\bar{X}$ la consommation moyenne de cigarette dans l'échantil
 $n = 80$

$$\text{On sait } \bar{X} \sim \text{cr}(m, \frac{\sigma}{\sqrt{n}})$$

$$\bar{X} \sim \text{cr}(16, \frac{4,2}{\sqrt{80}}) \Rightarrow \bar{X} \sim \text{cr}(16; 0,47)$$

$$2.2 P(\bar{X} < 15) \Rightarrow P\left(\frac{\bar{X} - 16}{0,47} < \frac{15 - 16}{0,47}\right)$$

$$\begin{aligned} \Rightarrow P(Z < -2,13) &= 1 - P(Z < 2,13) \\ &= 1 - 0,9834 \\ &= 0,0166 \end{aligned}$$

3.1 Echantillon $n = 200$ 90 fumeurs.

* Non biaisée: $E[f_n] = p$
 Convergence: $\lim \text{Var}(f_n) = \lim \frac{p(1-p)}{n} = 0$

X la VA représentant les habitants
 $X \sim B(1, p)$ avec 1 si joueur
 0 sinon

*

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = f_n \Rightarrow \hat{p} = \frac{90}{200}$$

$$\hat{p} = 0,45$$

3-2 Intervalle de confiance pour p avec $L=1$

$X \sim B(1, p)$

$$E(\hat{p}) = p \quad \text{Var}(X) = p(1-p)$$

$$IC = f_n - u \sqrt{\frac{f_n(1-f_n)}{n}} \leq p \leq f_n + u \sqrt{\frac{f_n(1-f_n)}{n}}$$

avec $f_n = 0,45$

$$u = \Phi^{-1}\left(\frac{2+1}{2}\right) = \Phi^{-1}\left(\frac{0,99+1}{2}\right) = \Phi^{-1}(0,995) = 2,58$$

$$IC = 0,45 - 2,58 \sqrt{\frac{0,45(1-0,45)}{200}} \leq p \leq 0,45 + 2,58 \sqrt{\frac{0,45(1-0,45)}{200}}$$

$$\Rightarrow 0,36 \leq p \leq 0,54$$

4) Soit $X \sim cr(14, \frac{4,2}{\sqrt{100}})$ donc $X \sim cr(14, 0,42)$

σ est connu.

$$IC = \bar{X} - u \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + u \frac{\sigma}{\sqrt{n}}$$

$$\text{avec } u = \Phi^{-1}\left(\frac{2+1}{2}\right) = \Phi^{-1}(0,995) = 2,58$$

$$IC = 14 - 2,58 \times \frac{4,2}{\sqrt{100}} \leq m \leq 14 + 2,58 \times \frac{4,2}{\sqrt{100}}$$

$$IC = 12,9 \leq m \leq 15,08$$

Avant 16

Comme 16 est nettement au dessus de l'intervalle de confiance la campagne est efficace