Stat : TD de révision

1/e

180	60	Amé
	60	Euro
	60	Jap

MPG = miles parcourus par gallon

(1) (a) Soit X_i la va $= \begin{cases} 0 \text{ si } mpg < 20 \\ 1 \text{ si } mpg > 20 \end{cases}$

X_i

X_i suit une loi de Bernoulli de paramètre p inconnu.

On estime p par $\hat{p} = \hat{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (n = 60)$

$= \frac{23}{60} = 0,38.$

(b) donner la loi de $\hat{X} = \hat{p} \quad \mathcal{L}(\hat{X}) = \mathcal{L}(N)$

$\hat{X} = \frac{\sum x_i}{n} = \frac{X}{n}$ où $X = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, p)$

$n\bar{X} \sim \mathcal{B}(n, p)$

Approximation

On a $n = 60 > 30$

$np = n\hat{p} = 60 \times \frac{23}{60} = 23 > 15$

$n\hat{p}(1 - \hat{p}) = 14,18 \overset{?}{>} 15$

Mais $\hat{p} = 0,38$ est assez grande $\Rightarrow$ on peut approximer par une loi normale.

$\mathcal{L}(n\bar{X}) \equiv N(n\hat{p}; n\hat{p}(1-\hat{p}))$

(c) $mpg = X \to \mathcal{N}(\mu, \sigma^2)$

où μ et σ^2 inconnus.

Estimation

i) la moyenne μ

$$\mu \to \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad (n = 180)$$

$$= 26.36$$

(d'après le tableau 1)

ii) la variance σ^2

$$\sigma^2 \to S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

Estimateur non biaisé

$S_1^2 = \frac{1}{n} \sum^{n} (X_i - \hat{\mu})^2$ est biaisé

$$S^2 = \frac{n}{n-1} \cdot S_1^2$$

où $S_1^2 = Var(mpg) = 52,06$

$$S^2 = \frac{180}{179} \times 52,06 = 52,99$$

iii) Estimation par intervalle

IDC (μ)

$$P(\mu \in [\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}]) = 1 - \alpha$$

où $z_{\frac{\alpha}{2}}$ est le quantile d'ordre $\frac{\alpha}{2}$ de $\mathcal{N}(0,1)$

$\alpha = 5\%$ $\quad P(N(0,1) > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} = 0,025$

$\Rightarrow z_{\frac{\alpha}{2}} = 1,96$

D' où

$IdC(\mu) = \left[26{,}36 - 1{,}96 \times \frac{7{,}28}{\sqrt{180}} ; 26{,}36 + 1{,}96 \times \frac{7{,}28}{\sqrt{180}}\right]$

$= [25{,}3 ; 27{,}42]$

v) la taille minimale n_{min} pour calculer $IdC(\mu)$ avec un risque $\alpha = 0{,}2\%$ et une marge d'erreur $R_0 = 1$

$n_{min} = \left[\frac{z_{\frac{\alpha}{2}} S}{R_0}\right]^2$ où $z_{\frac{\alpha}{2}} = z_{0{,}001}$

la table donne $z_{0{,}001} = 3{,}0902$

$n_{min} = \left[\frac{3{,}0902 \times 7{,}28}{1}\right] = 507$

c) $IcD(\sigma^2)$

$P\left(\sigma^2 \in \left[\frac{nS^2}{b} ; \frac{nS^2}{a}\right]\right) = 1 - \alpha$

où $a = \chi^2_{n-1, 1-\frac{\alpha}{2}}$

$b = \chi^2_{n-1, \frac{\alpha}{2}}$

$n = 180$, $\nu = n - 1 = 179$ $(\alpha = 5\%)$

On trouve $\chi^2_{179, 0.95}$ $\quad (P(\chi^2_{179} > a) = 0{,}95$

$P(\chi^2_{179} > b) = 0{,}05$

$\begin{cases} a = 149{,}05 \\ b = 211{,}22 \end{cases}$

$\Rightarrow IdC(\sigma^2) = \left[\frac{180 \times 52.99}{211.22} ; \frac{180 \times 5,99}{149.05}\right]$

d) la longueur d'intervalle est

$\left(\frac{1}{a} - \frac{1}{b}\right) nS^2 = R_0$

$\Rightarrow n = \frac{R_0}{S^2} \times \frac{1}{\frac{1}{a} - \frac{1}{b}}$

où $R_0 = 63.99 - 45.17 = 18.82$

D'où $n = \frac{18,82}{52,106} \times \frac{1}{\frac{1}{149.05} - \frac{1}{211.22}}$

$= 182,9 \simeq 183 \simeq 180.$

(c) $mpg = X \sim \mathcal{N}(\cancel{\mu}, \cancel{\sigma}^2)$

où μ et σ^2 inconnus.

Tests.

iv) Comparaison de deux produits

Test bilatéral, sur les moyennes des populations normales. Variances inconnues et inégales.

Etapes :

1) $H_0 : \mu_X - \mu_Y = 0,$ $\quad W = X - Y$

$H_1 : \mu_X \neq \mu_Y$

où X = mpg voiture ER

Y = ——— JP

2) $\alpha = 0,05$, $n_x = n_y = n = 60.$

$V_\alpha = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma_W}$

$\sigma_W = \frac{S_X^2}{n_x} + \frac{S_Y^2}{n_y}$. $\sigma_W^2 = Var(X-Y)$

On montre que $V_\alpha \sim t_{\frac{\alpha}{2}, \nu}$

Statistique
TD de révision
①/2

où $\nu = \dfrac{\left(\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}\right)^2}{\left(\dfrac{S_x^2}{n_x}\right)^2 + \left(\dfrac{S_y^2}{n_y}\right)^2} - 2 = 117.24 \simeq 117$

On $\nu \simeq 117 < 120 - 2 = 118$

4) Calcul de la valeur du critère

$c = t_{\frac{\alpha}{2}, \nu} \times \sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}$ où $t_{\frac{\alpha}{2}, \nu} = t_{0,025, 117}$ à déterminer dans la table de Student avec

$\begin{cases} \alpha = 0,025 \\ \nu = 117 \end{cases}$

$\longrightarrow t_{\frac{\alpha}{2}, \nu} = 1,984$.

d'où $C = 1,984 \times 1,017 = 2,017$

5) Décision : Si $\bar{X} - \bar{Y} \in [-C, C]$ alors on accepte H_0

or $\bar{Y} - \bar{X} = 32.10 - 27.85 = 4,35 > c = 2,017$

On rejette l'hypothèse H_0.

H_D) $\left(\dfrac{\sigma_x}{\sigma_y}\right)^2 = \rho_1^2 = 2$ // ou $\left(\dfrac{\sigma_x}{\sigma_y}\right)^2 = \dfrac{1}{S_1^2}$

$\alpha = 0,1$ et $\beta = 0,1$

On accepte l'équivalentre soit jop et ano du point de vue variance

A) ANOVA

Facteur : pays de provenance
2 modalités

$1 \leq i \leq p = 2$

$1 \leq k \leq n_i$ | $n_1 = n_2$
$n_1 = n_2 = 60$

sommes carrés $SSA =$ $\qquad\qquad$ $\nu_A = p - 1 = 1$

SS_R $\qquad\qquad$ $\nu_R = n_T - p$

SS_T $\qquad\qquad$ $n_T = \sum_{i=1} n_i = 120$

$= n_0 - 2 = 118.$

Le test

$SS_A = \sum_{i=1}^{2} n_i (\bar{x}_i - \bar{x}_T)^2$

$= n_1 (\bar{x}_1 - \bar{x}_T) + n_2 (\bar{x}_2 - \bar{x}_T)^2$

$= 60 [(27,85 - 30,03)^2 + (32,20 - 30,03)^2]$

D'où $SS_R = 4293.6$

Ainsi $F_A = 17,98$ à comparer avec la valeur F_{table} lue dans la table $F_{\alpha, \nu_A, \nu_R} = 3,93$,

Donc $F_A = F_{calculé} > F_{table}$: il y a une diff de mpg selon

Sur la base des observations, la région a de l'~~influence~~

provenance.

L'analyse de la variance (rejet) → diff significative

$Y = a_1 X_1 + b_2 X_2$

TP. principe

avec une var exp → le meilleur est le déplacement

avec deux var explicatives, la meilleure regr. est obtenue avec déplacement et accélération.

Ensuite on n'a pas vraiment d'amélioration si on ajoute des variables.

On choisit 2 comp → 90% de l'information.

Si on veut introduire les années on fait 70,1 ; 80,1 ; ...

Une var proche du cercle est bien representée