MAI 2007

Exercice 1 – Test 6 / Test 5→4

En 2006, les étudiants inscrits à l'enquête inter-années de l'UFR des sciences Sociales enquêtaient auprès de la population des jeunes de 15 à 24 ans du Bas Rhin. Le but de cette enquête est de voir comment les jeunes utilisent les nouvelles technologies. 997 jeunes ont été interrogés, dont 513 filles et 484 garçons.

1) Je démontre au risque de 5 % que les garçons sont plus nombreux à avoir déjà joué aux jeux vidéo. En effet 94 % des garçons ont déjà joué à des jeux vidéo contre 80,3 % des filles. Quel test ai-je utilisé ? Argumentez votre réponse (ne faîtes pas les calculs).

J'utilise pour ce faire un test de porportion entre 2 populations : **test 6**.

P = {Garçons de 15-24 ans du Bas-Rhin en 2006}	P' = {Filles de 15-24 ans du Bas-Rhin en 2006}
X = Indicateur de joueur de jeux vidéo	X = Indicateur de joueuse de jeux vidéo
$X = \begin{cases} 1 & \text{joueur} \quad p \\ 0 & \text{sinon} \quad 1-p=q \end{cases}$	$X = \begin{cases} 1 & \text{joueuse} \quad p' \\ 0 & \text{sinon} \quad 1-p'=q' \end{cases}$
$L(X) = B(1,p)$	$L(X) = B(1,p')$
p = proportion de joueurs de jeux vidéo parmi les garçons de 15-24 ans du Bas-Rhin en 2006 → inconnu	p' = proportion de joueuses de jeux vidéo parmi les filles de 15-24 ans du Bas-Rhin en 2006 → inconnu

Hypothèse : $H_0 : p = p'$ / $\overline{H_0} : p > p'$

H_0 : Les garçons et filles sont autant à jouer aux jeux vidéo, parmi les jeunes du Bas-Rhin en 2006.

Si on rejette H_0, on accepte son hypothèse contraire à savoir que les garçons jouent *plus* aux jeux vidéo que les filles, parmi les jeunes du Bas-Rhin en 2006 (on réalise alors un test unilatéral : on rejette la différence d'un seul côté).

Résolution

*** A priori**

n = 484 garçons (issus de la population P)

$X_1, X_2, \ldots, X_{484}$ 484 V.A. de Bernouilli, indépendantes (taux de sondage < 1/10).

$F_{484} = \dfrac{\sum X_i}{484}$ estimateur de p

Proportion théorique de garçons du BR qui jouent aux jeux vidéo dans un échantillon quelconque de 484 garçons.

$L(F_{484}) = PB(484, p)$

n' = 513 filles (issues de la population P')

$X_1^{'}, X_2^{'}, \ldots, X_{513}^{'}$ 513 V.A. de Bernouilli, indépendantes (taux de sondage < 1/10).

$F_{513}^{'} = \dfrac{\sum X_i^{'}}{513}$ estimateur de p'

Proportion théorique de filles du BR qui jouent aux jeux vidéo dans un échantillon quelconque de 513 filles.

$L(F_{513}^{'}) = PB(513, p')$

Sous H_0 : $p = p' \quad \Rightarrow \quad p - p' = 0$

Sous H_0 : $p = p' = p_0$

Estimation de p_0. $p_0 = \dfrac{nf_n + n' f_{n'}}{n+n'} = \dfrac{484 \times 0,94 + 513 \times 0,803}{997} = \dfrac{455+412}{997} = 0,870$

On estime à 87 % les jeunes du Bas-Rhin en 2006 qui jouent aux jeux vidéo (si garçons et filles jouent dans les mêmes proportions).

Approche par la loi normale (toujours sous H_0)

$$\left.\begin{array}{l} np_0 = 484 \times 0,87 = 421 \\ nq_0 = 484 \times 0,13 = 63 \end{array}\right\} \geq 10$$

En vertu du Théorème Central Limite :

$$L_0(F_{484}) \cong N\left(p_0, \sqrt{\frac{p_0 q_0}{n}}\right)$$

$$\left.\begin{array}{l} n' p_0 = 513 \times 0,87 = 446 \\ n' q_0 = 513 \times 0,13 = 67 \end{array}\right\} \geq 10$$

En vertu du Théorème Central Limite :

$$L_0(F_{513}) \cong N\left(p_0, \sqrt{\frac{p_0 q_0}{n'}}\right)$$

Soit $F_n - F_{n'}$ estimateur de $p - p'$

différence théorique de proportion de joueurs de jeux vidéo entre les garçons et les filles, dans des échantillons respectifs de 484 garçons et 513 filles.

$$L_0(F_{484} - F_{513}) \cong N\left(0, \sqrt{p_0 q_0 \left(\frac{1}{n} + \frac{1}{n'}\right)}\right)$$

$$L_0\left(F_{484}-F_{513}\right)\cong N\left(0,\sqrt{0,87\times 0,13\times\left(\frac{1}{484}+\frac{1}{513}\right)}\right)$$

$$\boxed{L_0\left(F_{484}-F_{513}\right)\cong N\left(0\,;0,021\right)}$$

* **A posteriori**

$f_{484}-f_{513}=0,94-0,803=0,137$

On observe une différence de 13,7 points entre la proportion de garçons et de filles qui jouent aux jeux video, dans nos échantillons.

Test

i) $\varepsilon=0,05 \xrightarrow{\text{t.2}} t=1,645$

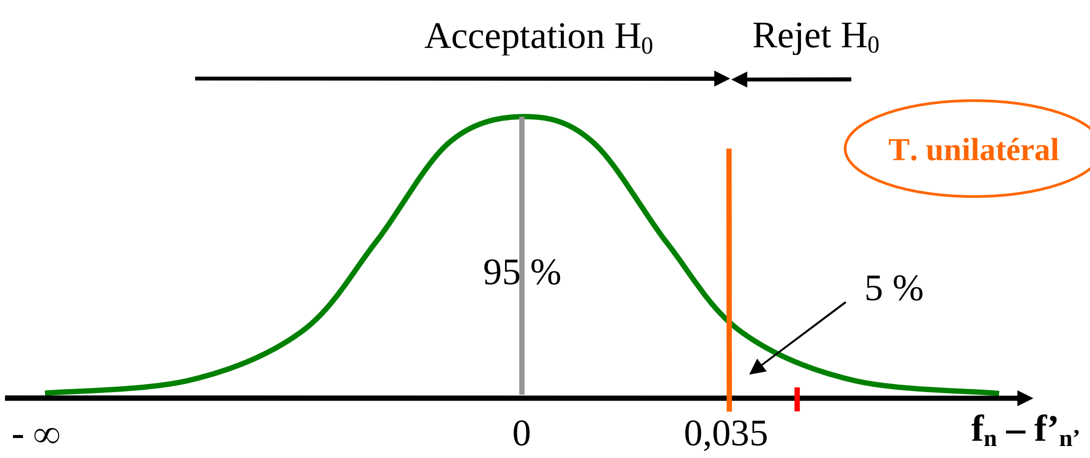
ii) $p\left(f_{484}-f_{513}\in I_\varepsilon\right)=0,95$ avec $I_\varepsilon=\left]-\infty\,;t\cdot\sqrt{p_0q_0\left(\frac{1}{n}+\frac{1}{n'}\right)}\right]$

Soit : $I_\varepsilon=\left]-\infty\,;1,645\cdot 0,021\right]=\left]-\infty\,;0,035\right]$

Si garçons et filles jouent aux jeux video dans les mêmes porportions, on peut accepter une différence allant jusqu'à 3,5 %.

iii) Décision : $\left(f_{484}-f_{513}\right)\notin I_A$ → Rejet de H_0 au seuil de 5 %.

On peut dire, avec un risque de 5 %, que les garçons sont plus nombreux à jouer aux jeux vidéo que les filles, parmi les jeunes du Bas-Rhin en 2006.



La différence de proportion entre les filles et les garçons est trop importante pour être due aux aléas de l'échantillon.

2) On leur a demandé combien d'heures, ils avaient joué par semaine dans les 4 dernières semaines. Sur les 367 garçons qui avaient joué aux jeux vidéo les 4 dernières semaines, on observe un nombre moyen d'heures de jeux de 10,9 heures par semaine avec un écart-type observé de 37,3 heures et sur les 176 filles, 5,7 heures de jeux en moyenne avec un écart-type de 14,9 heures.

Démontrez au risque de 5 % que les garçons jouent plus longtemps aux jeux vidéo que les filles. Commentez.

P = {Garçons de 15-24 ans du Bas-Rhin en 2006 qui ont joué aux jeux vidéo les 4 dernières semaines}	P' = {Filles de 15-24 ans du Bas-Rhin en 2006 qui ont joué aux jeux vidéo les 4 dernières semaines}
X = Nombre d'heures à jouer par semaine.	X = Nombre d'heures à jouer par semaine.
$L(X) = L(\mu, \sigma)$	$L(X') = L(\mu', \sigma')$
μ = nombre moyen d'heures jouées par les garçons de 15-24 ans du Bas-Rhin en 2006 → inconnu	μ' = nombre moyen d'heures jouées par les filles de 15-24 ans du Bas-Rhin en 2006 → inconnu

Hypothèse : $H_0 : \mu = \mu'$ / $\overline{H_0} : \mu > \mu'$

H_0 : Garçons et filles du Bas-Rhin en 2006 jouent autant d'heures par semaine aux jeux vidéo. Si on rejette l'hypothèse nulle, on accepte que les garçons jouent plus aux jeux vidéo que les filles.

Résolution

* **A priori**

n = 367 garçons	n' = 176 filles
$X_1, X_2, \ldots, X_{367}$ 367 V.A. issues de X, indépendantes (taux de sondage < 1/10).	$X_1', X_2', \ldots, X_{176}'$ 176 V.A. issues de X', indépendantes (taux de sondage < 1/10).
$\overline{X} = \frac{\sum X_i}{367}$ estimateur de μ	$\overline{X'} = \frac{\sum X_i'}{176}$ estimateur de μ'
Nombre moyen théorique d'heures jouées par les garçons aux jeux vidéo par semaine, dans un échantillon de 367 garçons.	Nombre moyen théorique d'heures jouées par les filles aux jeux vidéo par semaine, dans un échantillon de 176 filles.

$L\left(\overline{X}\right) = ?$ | $L\left(\overline{X'}\right) = ?$

On ne connaît pas les lois respectives des estimateurs de μ et μ’ car σ et σ’ sont inconnus. Il faudrait alors les estimer et passer par des lois de Student (soit un $\boxed{\textbf{test 5}}$) ; or, comme n et n’ sont grands (respectivement 367 garçons et 176 filles), on peut admettre que $\frac{\sigma}{\sqrt{n}} \approx \frac{s_c}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ et $\frac{\sigma'}{\sqrt{n}} \approx \frac{s_c'}{\sqrt{n}} \approx \frac{s'}{\sqrt{n}}$; de même qu’on peut approcher la loi de Student par la loi Normale [On passe alors dans une résolution en $\boxed{\textbf{test 4}}$]. Alors, en vertu du Théôrème Central Limite :

$L\left(\overline{X}\right) \cong N\left(\mu, \frac{s}{\sqrt{n}}\right)$

Avec $s = 37{,}3\ h$ (écart-type observé)

$L\left(\overline{X'}\right) \cong N\left(\mu', \frac{s'}{\sqrt{n'}}\right)$

Avec $s' = 14{,}9\ h$

Sous H_0 : $\mu = \mu' \Rightarrow \mu - \mu' = 0$

Soit $\overline{X} - \overline{X'}$ estimateur de $\mu - \mu'$

$L_0\left(\overline{X} - \overline{X'}\right) \cong N\left(0, \sqrt{\frac{s^2}{n} + \frac{s'^2}{n'}}\right)$ soit : $L_0\left(\overline{X} - \overline{X'}\right) \cong N\left(0, \sqrt{\frac{37{,}3^2}{367} + \frac{14{,}9^2}{176}}\right)$

$L_0\left(\overline{X} - \overline{X'}\right) \cong N\left(0\,; 2{,}25\right)$ Si garçons et filles jouent autant, la différence d’heures entre garçons et filles suit une loi normale d’espérance 0 h et d’écart-type 2,25 h.

*** A posteriori**

$\overline{x} - \overline{x'} = 10{,}9 - 5{,}7 = 5{,}2$ h

On a observé une différence de 5,2 h entre les temps moyens de jeu des garçons et des filles dans notre enquête. On se pose la question de savoir si cette différence est significative.

Test

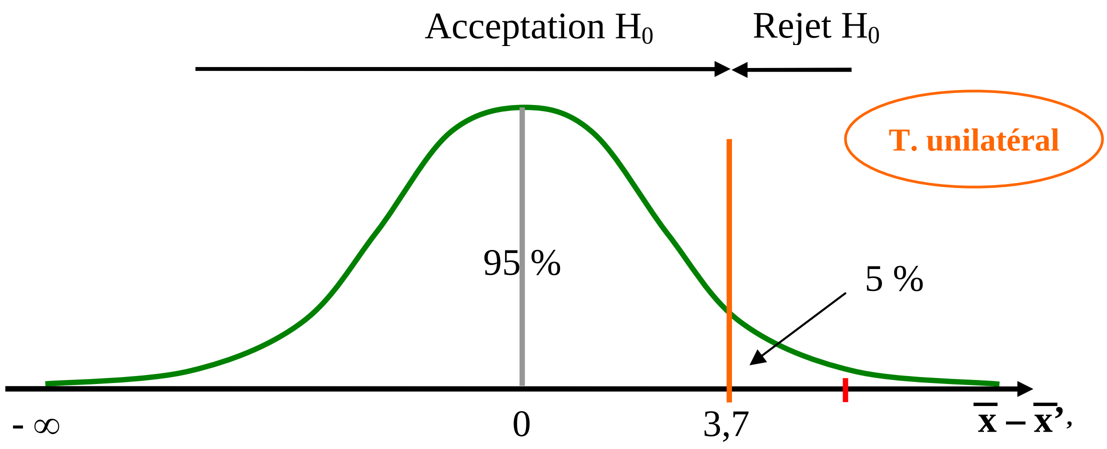
i) $\varepsilon = 0{,}05 \xrightarrow{\text{t.2}} t = 1{,}645$

ii) $p\left(\overline{x} - \overline{x'} \in I_\varepsilon\right) = 0{,}95$ avec $I_\varepsilon = \left]-\infty\,; t \cdot \sqrt{\frac{s^2}{n} + \frac{s'^2}{n'}}\right]$

Soit : $I_\varepsilon = \left]-\infty\,; 1{,}645 \cdot 2{,}25\right] = \left]-\infty\,; 3{,}70\right]$

iii) Décision : $\left(\overline{x} - \overline{x'}\right) \notin I_A$ → Rejet de H_0 au seuil de 5 %.

On peut dire, avec un risque de 5 %, que les garçons jouent plus longtemps aux jeux vidéo que les filles, dans le Bas-Rhin en 2006.



→ Raisons ? Interprétations ?

Les geeks sont des garçons : ce qui pousse la moyenne d'heures jouées (peu de garçons qui jouent beaucoup ; l'écart-type est largement plus important chez les garçons que chez les filles, avec une différence de plus de 12 h, bien que les garçons soient plus nombreux). On observe donc une disparité entre les filles et les garçons, mais sachant que le comportement des joueurs n'est pas homogène.