8 STT – 105
PROBABILITÉ ET STATISTIQUE I
EXERCICES – Test d'hypothèses

1. Voici les résumés de résultats d'observations sur les rythmes cardiaques d'individus vivant en milieu urbain (U) et en milieu rural (R) :

$$\begin{array}{lll} n_U = 200 & \overline{x}_U = 81 & s_U^2 = 147 \\ n_R = 300 & \overline{x}_R = 76 & s_R^2 = 126 \end{array}$$

a) Avec un niveau $\alpha = 0,05$ et en supposant que les populations sont normales de même variance, peut-on croire que la vie en milieu urbain est associée à un rythme cardiaque plus élevé?

Solution : On cherche à confronter les hypothèses

$$\begin{aligned} H_0 &: \mu_U = \mu_R \\ H_1 &: \mu_U > \mu_R \end{aligned}$$

où μ_U et μ_R sont respectivement les moyennes de la mesure du rythme cardiaque en milieu urbain et rural. Le test[1] est de rejeter H_0 si

$$\frac{\overline{X}_U - \overline{X}_R}{\sqrt{\frac{(n_U-1)S_U^2+(n_R-1)S_R^2}{n_U+n_R-2}}\sqrt{\frac{1}{n_U}+\frac{1}{n_R}}} > t_{n_u+n_R-2;a}$$

Le niveau est fixé à 5%. On observe

$$\begin{aligned} \frac{\overline{x}_U - \overline{x}_R}{\sqrt{\frac{(n_U-1)s_U^2+(n_R-1)s_R^2}{n_U+n_R-2}}\sqrt{\frac{1}{n_U}+\frac{1}{n_R}}} &= \frac{81-76}{\sqrt{\frac{(199)147+(299)126}{200+300-2}}\sqrt{\frac{1}{200}+\frac{1}{300}}} \\ &= 4.7247 \end{aligned}$$

et le point critique est $t_{498;0.05} = 1.6479$. On rejette H_0 au niveau 5% et on peut dire que le rythme cardiaque est plus élevé en milieu urbain.

b) À partir de quel niveau ne peut-on plus le croire?

Solution : on cherche le niveau du test si le point critique était la valeur observée, c'est-à-dire le niveau de signification, ou

$$\Pr\left(\frac{\overline{X}_U - \overline{X}_R}{\sqrt{\frac{(n_U-1)S_U^2+(n_R-1)S_R^2}{n_U+n_R-2}}\sqrt{\frac{1}{n_U}+\frac{1}{n_R}}} > 4.7247 \mid H_0\right)$$

Or on sait que sous H_0 la statistique du test suit une distribution de Student à 498 degrés de liberté (puisque le point critique est lu dans la table de Student...). Ainsi on cherche

$$\Pr(T > 4.7247)$$

pour $T \sim t_{498}$. Dans la table de Student on trouve que cette probabilité est plus petite que 0.005.

[1] C'est le test pour deux échantillon avec mêmes variances en utilisant le point critique $t_{n_U+n_R-2;\alpha}$ puisque c'est un test unilatéral.

2. Un professeur veut vérifier si une méthode d'enseignement a un effet réel sur la capacité de concentration des élèves. La mesure est un score à un test de concentration avant l'application de la méthode et une autre mesure après 6 mois. La différence de mesure (après moins avant) suit une distribution normale de variance 12. Le professeur veut utiliser un test sur π, la probabilité d'augmenter le score après la période, parce qu'il semble plus facile à faire. A-t-il raison ?

   Solution : Le test sur la binomiale est de rejeter H_0 si

$$\frac{p-\pi_0}{\sqrt{\pi_0\left(1-\pi_0\right)}}\sqrt{n}\geq z_\alpha$$

   En considérant $\pi=\pi_1$, la distribution de p est $N\left(\pi_1,\pi_1\left(1-\pi_1\right)/n\right)$ et la puissance est donnée par

$$\begin{aligned}
\Pr\left(\text{rejeter } H_0 \mid H_1\right) &= \Pr\left(\frac{p-\pi_0}{\sqrt{\pi_0\left(1-\pi_0\right)}}\sqrt{n}\geq z_\alpha \mid H_1\right) \\
&= \Pr\left(p\geq z_\alpha\frac{\sqrt{\pi_0\left(1-\pi_0\right)}}{\sqrt{n}}+\pi_0 \mid H_1\right) \\
&= \Pr\left(\frac{p-\pi_1}{\sqrt{\pi_1\left(1-\pi_1\right)}}\sqrt{n}\geq \frac{z_\alpha\frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}}+\pi_0-\pi_1}{\sqrt{\pi_1\left(1-\pi_1\right)}}\sqrt{n} \mid H_1\right) \\
&\simeq \Pr\left(Z\geq \frac{z_\alpha\frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}}+\pi_0-\pi_1}{\sqrt{\pi_1\left(1-\pi_1\right)}}\sqrt{n} \mid H_1\right) \\
&= \Pr\left(Z\geq \frac{\sqrt{n}\left(\pi_0-\pi_1\right)}{\sqrt{\pi_1\left(1-\pi_1\right)}}+z_\alpha\frac{\sqrt{\pi_0\left(1-\pi_0\right)}}{\sqrt{\pi_1\left(1-\pi_1\right)}} \mid H_1\right)
\end{aligned}$$

   Il faut déterminer les contres hypothèses. Si la différence est de a, alors en posant X la v.a. qui donne le score au test, la probabilité d'améliorer le score est de

$$\Pr\left(X\geq 0\right)=\Pr\left(Z\geq\frac{0-a}{\sqrt{12}}\right)$$

   où $Z\sim N\left(0,1\right)$.

   Le test sur la moyenne est de rejeter H_0 si

$$\frac{\overline{X}}{S}\sqrt{n}\geq z_\alpha$$

en considérant que sous $H_1 : \mu = a$, la puissance pour une valeur a est

$$\begin{aligned} \Pr\left(\text{rejeter } H_0 \mid H_1\right) &= \Pr\left(\frac{\overline{X}}{S}\sqrt{n} \geq z_\alpha \mid H_1\right) \\ &= \Pr\left(\overline{X} \geq z_\alpha \frac{S}{\sqrt{n}} \mid H_1\right) \\ &= \Pr\left(\frac{|\overline{X} - a|}{S}\sqrt{n} \geq \frac{z_\alpha \frac{S}{\sqrt{n}} - a}{S}\sqrt{n} \mid H_1\right) \\ &\simeq \Pr\left(Z \geq z_\alpha - a\frac{1}{S}\sqrt{n} \mid H_1\right) \end{aligned}$$

On peut alors comparer les différentes valeurs de la puissance pour un test à $\alpha = 5\%$ et $n = 50$

a	π_1	$1 - \beta(a)$	$1 - \beta(\pi_1)$
1	0.613 59	0.654 04	0.484 19
2	0.718 15	0.992 61	0.945 25
1.5	0.667 50	0.921 74	0.778 81