HEG Genève
Dr S. Varone

Statistique inférentielle

Corrigé 7

Problème 1

Les données suivantes représentent le taux d'acceptation d'une initiative pour un échantillon de communes en Suisse.

62.3	44.4	49.2	63.3	47.6	60.1
37.4	55.8	57.5	58.3	56.2	54.3

On suppose que l'écart type des taux pour l'ensemble des communes suisses vaut 5, et que les données proviennent d'une distribution normale.

```
x <- c(62.3,44.4,49.2,63.3,47.6,60.1,37.4,55.8,57.5,58.3,56.2,54.3)
library(TeachingDemos); z.test(x, mu=50, stdev=5, alternative="greater")
```

a) En prenant un risque de première espèce $\alpha = 0.05$, la moyenne μ des taux de la population de communes est-elle significativement supérieure à 50% ? Commenter.

i) La valeur de la population d'intérêt est la moyenne des taux d'acceptation

ii) Les hypothèses nulle et alternative sont

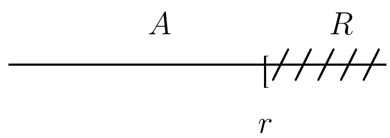
$$\begin{array}{rl} & H_0 : \mu = \mu_0 \leq 50 \\ \text{contre} & H_1 : \mu = \mu_1 > 50 \end{array}$$

iii) Le niveau de signification est $\alpha = 0.05$

iv) Test de la moyenne avec variance connue :

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma_{\bar{X}}} \sim N(0,1)$$

La région critique est déterminée par un test unilatéral à droite :



Le seuil de rejet est $r = z_\alpha = z_{0.05} = 1.645$

v) Valeur observée de la statistique :

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.87 - 50}{\frac{5}{\sqrt{12}}} = 2.68$$

vi) $2.68 > 1.645$, l'hypothèse H_0 est donc rejetée.

vii) La moyenne des taux est donc significativement supérieure à 50%

b) Soit un échantillon de taille 100 pour lequel on a obtenu la même moyenne que pour l'échantillon précédent de taille 12. Déterminer la région critique du test pour ce nouvel échantillon et comparer avec celle obtenue au point 1. Commenter.

   i) Le seuil de rejet ne dépendant pas de la taille de l'échantillon, il reste le même que précédemment, soit $r = z_\alpha = z_{0.05} = 1.645$

   ii) La valeur observée de la statistique dépend de la taille de l'échantillon. En fait, seul l'écart type de la statistique change : elle diminue. Et donc la statistique augmente sa valeur, et reste donc donc la région critique.

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.87 - 50}{\frac{5}{\sqrt{100}}} = 7.74$$

   iii) Comme $7.74 > 1.645$ l'hypothèse H_0 est rejetée.

   iv) La moyenne des taux est donc dans ce cas aussi significativement supérieure à 50%

Problème 2

En 2005, une enquête menée auprès de femmes quant à leur âge x_i à la naissance de leur premier enfant a donné, pour 51 femmes interrogées au hasard, les résultats suivants :

$$\bar{x} = 28 \text{ années}$$

$$\sum (x_i - \bar{x})^2 = 816 \text{ années}^2$$

Les âges x_i représentent les réalisations d'une variable aléatoire X distribuée selon une loi normale (μ, σ^2).

a) On sait qu'en 1990, pour la population, l'âge moyen des femmes à la naissance de leur premier enfant était de 27 ans. Peut-on affirmer, avec un degré de confiance de 95 %, que l'âge moyen a augmenté en 3 ans ? Pour cela, on testera l'hypothèse $H_0 : \mu = 27$ contre l'hypothèse $H_1 : \mu > 27$.

   i) Le paramètre d'intérêt est la moyenne $\mu = 27$

   ii)

$$\begin{aligned} & H_0 : \mu = \mu_0 = 27 \\ \text{contre} \quad & H_1 : \mu = \mu_1 > 27 \end{aligned}$$

   iii) Le niveau de signification $\alpha = 0.05$

   iv) Test de la moyenne avec variance inconnue :

$$T_0 = \frac{\overline{X} - \mu_0}{\hat{\sigma}_{\bar{X}}} \sim St_{n-1}$$

   La région critique est déterminée par un test unilatéral à droite. La valeur du seuil de rejet est

$$r = t_{0.05}^{(50)} = 1.676 \approx z_{0.05} \approx 1.645$$

v) La valeur de la statistique est

$$\sigma_{\bar{x}} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}}{\sqrt{n}} = \frac{\sqrt{816}}{\sqrt{51}\cdot\sqrt{50}} = 0.5657$$

$$t_0 = \frac{\bar{x}-\mu_0}{\hat{\sigma}_{\bar{x}}} = \frac{28-27}{0.5657} = 1.768$$

vi) Comme $t_0 = 1.768 > 1.676 = r$, l'hypothèse H_0 est rejetée.

vii) En conclusion, l'âge moyen a augmenté.

b) Selon certains chercheurs, l'écart type s de l'échantillon sous-estime l'écart type σ de la population, qu'ils évaluent à 5 ans. Refaire le test précédent en supposant que la vraie valeur de σ est égale à 5 et commenter.

iv) Test de la moyenne avec variance connue :

$$Z_0 = \frac{\overline{X}-\mu_0}{\sigma_{\bar{X}}} \sim N(0,1)$$

v) Seuil de rejet : $r = z_{0.05} = 1.645$ Valeur observée de la statistique :

$$\sigma_{\bar{x}} = \frac{\sigma_0}{\sqrt{n}} = \frac{5}{\sqrt{51}} = 0.7$$

$$z_0 = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}} = \frac{28-27}{0.7} = 1.43$$

vi) Comme $z_0 = 1.43 < 1.645 = r$, l'hypothèse H_0 n'est pas rejetée.

vii) En d'autres termes, il n'y a pas suffisamment d'évidence pour affirmer que l'âge moyen aurait augmenté, i.e. l'âge moyen n'a pas augmenté.

Problème 3

Dans une grande entreprise américaine, le salaire annuel moyen des hommes possédant entre 3 et 5 ans d'expérience est de 58'000\$. Les salaires (en milliers de dollars) d'un échantillon aléatoire de 10 femmes possédant entre 3 et 5 ans d'expérience sont les suivants :

54 57 61 51 49 56 60 52 45 66

y a-t-il suffisamment d'évidence pour attester que les femmes gagnent moins que les hommes, en prenant un risque $\alpha = 0.05$?

Moyenne $\bar{x} = 55.1$

```
x <- c(54, 57, 61, 51, 49, 56, 60, 52, 45, 66)
t0 <- qt(0.95,9)
taille <- length(x)
seuilcritique <- mean(x) + t0 * sd(x)/sqrt(taille)
t.test(x,mu=58,alternative="less")
```

a) Le paramètre d'intérêt est la moyenne des salaires, qui vaut 58 milliers de \$

b) Comme il peut être supposé que le salaire moyen des femmes soit inférieur à celui des hommes, nous pouvons utiliser un test unilatéral à gauche.

$$\begin{aligned} & H_0 : \mu = 58 \\ \text{contre} \quad & H_1 : \mu < 58 \end{aligned}$$

c) Le niveau de signification $\alpha = 0.05$

d) Le seuil critique vaut $t^9_{0.95} = -1.8331$ et donc la région critique est $[-\infty; -1.8331]$

e) La statistique associée à l'échantillon vaut $\sigma_{\bar{x}} = \sqrt{\frac{348.9}{10\times 9}} = 1.97$

$$t_0 = \frac{55.1 - 58}{1.97} = -1.47$$

f) Comme $t_0 \notin [-\infty; -1.8331]$ l'hypothèse H_0 n'est pas rejetée.

g) Il n'y a donc pas suffisamment d'évidence pour rejeter l'égalité moyenne des salaires hommes et femmes.