Tests Statistiques Parametriques en R (3)

October 17, 2018

Greta Laage, Luc Adjengue

Contents

1 Rappel de quelques définitions

Erreur de type I: lorsque H_0 est rejetée alors qu'elle est vraie. Erreur de type II: lorsque H_0 est acceptée alors qu'elle est fausse.

α correspond à la probabilité de faire l'erreur de type I tandis que β représente la probabilité de faire l'erreur de type II. On appelle puissance du test la valeur $1-\beta$, c'est la force avec laquelle on rejette H_0 lorsque H_0 est fausse.

Les tests statistiques paramétriques sont utilisés pour tester la valeur d'un paramètre ou l'égalité des paramètres de deux distributions.

1.1 Démarche générale d'un test d'hypothèse

1. Déterminer les hypothèses nulle et alternative et fixer le seuil α du test
2. Etablir la règle de rejet de H_0 à partir de la distribution de la statistique du test et calculer la statistique du test.
3. Décider si H_0 doit être rejetée ou non en fonction de la règle de rejet établie et de la valeur trouvée à l'étape précédente.
4. Enoncer la conclusion du test.

2 Tests d'hypothèses à partir d'un seul échantillon

2.1 Moyenne d'une distribution normale de variance connue

On suppose $X \sim \mathcal{N}(\mu, \sigma)$ avec μ inconnue et σ connue.

Les hypothèses sont

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

La statistique du test est $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ et suit une loi normale si H_0 est vraie.

Exemple 10.26 du livre avec les hypothèses :

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

In [7]:
```
# Données de l'exemple 10.6 du livre

n = 25
mu0 = 40
alpha = 0.05
xbar = 41.25
sigma = 2

# valeur avec laquelle comparer la statistique du test
qnorm(1-0.025)
qnorm(0.025)
```

```
# statistique du test
z = (xbar - mu0)/(sigma/sqrt(n))
z
```

1.95996398454005
-1.95996398454005
3.125
Si $-1.96 \leq z_0 \leq 1.96$, alors H_0 est rejetée. C'est le cas ici.
Avec la même hypothèse nulle mais l'hypothèse $H_1 : \mu > \mu_0$, alors on rejette H_0 si$Z_0 > z_\alpha$

```
In [9]: z_alpha = qnorm(1-0.05)
        z_alpha
```

-1.64485362695147
Avec la même hypothèse nulle mais l'hypothèse $H_1 : \mu < \mu_0$, alors on rejette H_0 si $Z_0 < z_\alpha$

2.2 Moyenne d'une distribution normale de variance inconnue

Cette procédure de test est souvent appelée "test t" car elle fait appel à la loi de Student.
On suppose $X \sim \mathcal{N}(\mu, \sigma)$ avec μ et σ inconnues.
Les hypothèses sont

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

La statistique du test est $T_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ et suit une loi t avec $n-1$ degrés de liberté si H_0 est vraie.
La fonction R à utiliser pour faire le test-t de student est **t.test()**. On peut choisir de faire un test bilatéral (par défaut sur R) ou non, il faut le préciser avec le paramètre **alternative** :

```
* alternative = two.sided pour faire un test-t bilatéral
* alternative = less pour faire un test-t unilatéral inférieur
* alternative = greater pour faire un test-t unilatéral supérieur
```

On utilise l'ensemble de données **sleep** de R qui contient les heures de sommeil additionnel de patients ainsi que leur groupe (correspondant à un certain traitement) et un identifiant.

```
In [12]: head(sleep)
```

extra	group	ID
0.7	1	1
-1.6	1	2
-0.2	1	3
-1.2	1	4
-0.1	1	5
3.4	1	6

```
In [17]: #On teste si la moyenne de la variable *extra* suit une loi normale de moyenne 1
         res = t.test(sleep$extra, mu=1)
         print('Affichage du résultat du test')
         res
         print('Affichage de la p-value uniquement')
         res$p.value # Affichage de la p-value
         print('Affichage du degré de liberté')
         res$parameter # Affichage du dégré de liberté
         print('Affichage de la statistique t')
         res$statistic # Affichage  de la statistique t
```

```
[1] "Affichage du résultat du test"
```

```
One Sample t-test

data:  sleep$extra
t = 1.1968, df = 19, p-value = 0.2461
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 0.5955845 2.4844155
sample estimates:
mean of x
     1.54
```

```
[1] "Affichage de la p-value uniquement"
```

0.246122418040227

```
[1] "Affichage du degré de liberté"
```

df: 19

```
[1] "Affichage de la statistique t"
```

t: 1.19675395938043

```
In [18]: # test unilatéral
         t.test(sleep$extra, mu=1, alternative = "less")
```

```
One Sample t-test
```

```
data:  sleep$extra
t = 1.1968, df = 19, p-value = 0.8769
alternative hypothesis: true mean is less than 1
95 percent confidence interval:
    -Inf 2.32022
sample estimates:
mean of x
     1.54
```

3 Tests d'hypothèses à partir de deux échantillons

3.1 Moyennes de deux distributions normales de variance inconnues

On utilies la fonction **t.test()** et on doit préciser plusieurs paramètres :

- les deux échantillons
- si le test est bilatéral (**alternative = "two.sided"**) ou unilatéral (**alternative = "greater"** ou **alternative = "less"**)
- si les données sont appariées (**paired = True**) ou non (**paired = False**). Par défaut R considère que les données ne sont pas appariées
- la différence entre les moyennes des deux échantillons (**mu**). Par défaut R considère que c'est 0
- le seuil considéré (**conf.level**)
- si les variances sont inégales ou non (**var.equal = FALSE**).Par défaut la fonction **t.test()** de R suppose que les variances sont inégales.

On prend le poids de 10 hommes et 10 femmes pour faire le test. Les données sont importées depuis le fichier *poids.csv*. On veut tester si le poids moyen des hommes est significativement différent de celui des femmes.

```
In [10]: d <- read.csv("poids.csv", header = TRUE, sep = ";",dec = ".")
```

```
In [11]: # avec des données non appariées
         test1 <- t.test(Poids ~ Group, data=d,alternative = "two.sided",paired = FALSE)
         test1
```

```
Welch Two Sample t-test

data:  Poids by Group
t = -4.1493, df = 17.448, p-value = 0.0006388
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -37.76246 -12.33754
sample estimates:
mean in group Femme mean in group Homme
              55.53               80.58
```

t est la statistique de student, **df** est le degré de liberté , **p-value** est le degré de significativité du test. Comme elle est très inférieure à 5%, on peut conclure que les poids moyens sont significativement différents. Le test affiche également l'intervalle de confiance de la différence des moyennes à 95% et la valeur moyenne de chacun des groupes.

3.2 Egalité de deux variances

On se place dans le cas où les variables aléatoires suivent une loi normale avec des paramètres respectifs et on veut comparer la variance des deux lois.

On utilise la fonction **var.test()** de R.

On reprend les données de poids ci-dessus. On précise également si le test est bilatéral ou non.

```
In [13]: var.test(Poids ~ Group, data=d,alternative = "two.sided" )
```

```
F test to compare two variances

data:  Poids by Group
F = 1.4327, num df = 9, denom df = 9, p-value = 0.6008
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3558597 5.7679982
sample estimates:
ratio of variances
          1.432689
```

Références: www.sthda.com