Fiche TD avec le logiciel R : tdr211

Pratique des tests en A.P.A.

A.B. Dufour & T. Jombart

La fiche contient quelques exercices portant sur les tests clasiques de comparaison d'une valeur observée à une valeur théorique ; de deux valeurs observées.

Table des matières

1 Latéralisation et sport d'opposition **2**

2 Etudiants du campus de la Doua et groupes sanguins **2**

3 Rugby et recettes **2**

4 Indice de Quételet et Syndrôme de Turner [1] **4**

5 Tabac et Cancer **5**

6 Lecture et Apprentissage **6**

7 Retard de croissance et Indice de Quételet **6**

8 Dépense énergétique et poids **7**

Références **7**

1 Latéralisation et sport d'opposition

Les études statistiques portant sur la latéralisation permettent d'estimer à environ 10% de la population la proportion d'individus qui, dans nos sociétés, utilisent préférentiellement leur main gauche dans les tâches motrices et dans l'écriture en particulier. Parmi les meilleurs tennismen et escrimeurs mondiaux, on a dénombré 18 gauchers sur 64. Peut-on admettre que les gauchers sont plus nombreux dans ces sports d'opposition que dans la population totale ?
Si nous appliquons la valeur de la statistique du test à ces données :

$$z = \frac{f - p_0}{\sqrt{(\frac{p_0(1-p_0)}{n}}}$$

```
z <- (((18/64) - 0.1)/sqrt(0.1 * 0.9/64))^2
z
[1] 23.36111
2 * (1 - pnorm(z))
[1] 0
```

Il existe une procédure dans R pour comparer une proportion observée à une proportion théorique.

```
prop.test(18, 64, 0.1, correct = F, alternative = "two.sided")
        1-sample proportions test without continuity correction
data:  18 out of 64, null probability 0.1
X-squared = 23.3611, df = 1, p-value = 1.343e-06
alternative hypothesis: true p is not equal to 0.1
95 percent confidence interval:
 0.1859337 0.4013393
sample estimates:
      p
0.28125
```

La fonction donne la proportion observée dans l'échantillon (`sample estimates`).

```
18/64
[1] 0.28125
```

2 Etudiants du campus de la Doua et groupes sanguins

Les centres de transfusion sanguine ont diffusé le tableau des répartitions, en France, des principaux groupes sanguins. La proportion de O^-et O^+ est de 0.44. Une collecte de sang est organisée sur le campus de la Doua. Sur 356 étudiants, 148 sont de type O. Quelle conclusion pouvez-vous tirer de cette étude ?

3 Rugby et recettes

Le journal *l'équipe* donne les recettes (en francs) de 14 matches de rugby joués au cours de la journée du 13 janvier 1991 :

22550	22180	25000	38580	41860	39800	178525
31670	75785	17680	62595	23825	40550	144000

```
recette <- c(22550, 22180, 25000, 38580, 41860, 39800, 178525, 31670,
    75785, 17680, 62595, 23825, 40550, 144000)
```

1) Calculer la moyenne, la variance et l'écart-type des recettes de la journée du 13 janvier.
2) En 1990, la recette moyenne d'un match a été de 65 000 F. Peut-on considérer que la recette de cette journée du 13 janvier 1991 est comparable à la recette moyenne de l'année précédente ?

```
mean(recette)
[1] 54614.29
var(recette)
[1] 2345251411
sd(recette)
[1] 48427.8
```

Si nous appliquons la valeur de la statistique du test à ces données :

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\widehat{\sigma^2}}{n}}}$$

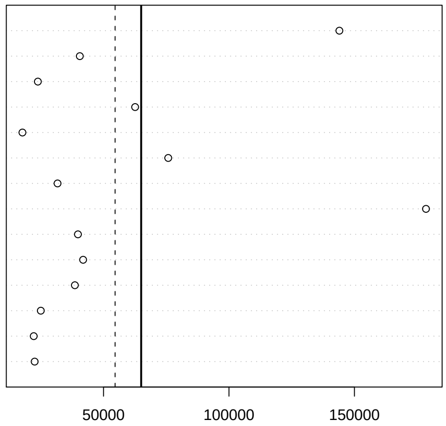
```
vt <- (mean(recette) - 65000)/sqrt(var(recette)/length(recette))
vt
[1] -0.8024273
2 * pt(vt, length(recette) - 1)
[1] 0.4367301
```

Il existe une procédure dans R pour comparer une proportion observée à une proportion théorique.

```
t.test(recette, mu = 65000)
        One Sample t-test
data:  recette
t = -0.8024, df = 13, p-value = 0.4367
alternative hypothesis: true mean is not equal to 65000
95 percent confidence interval:
 26652.91 82575.66
sample estimates:
mean of x
 54614.29
```

N'oubliez pas que les graphiques contribuent toujours à bien comprendre les données.

```
dotchart(recette)
abline(v = mean(recette), lty = 2)
abline(v = 65000, lwd = 2)
```

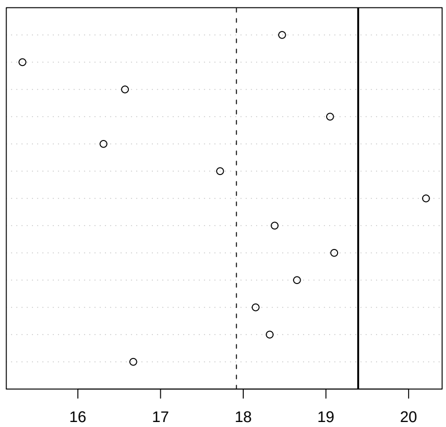


4 Indice de Quételet et Syndrôme de Turner [1]

On a mesuré l'indice de Quételet chez 13 jeunes filles âgées de 14 ans, atteintes du syndrôme de Turner. On a :

```
quetelet <- c(16.67, 18.32, 18.15, 18.65, 19.1, 18.38, 20.21, 17.72,
    16.31, 19.05, 16.57, 15.33, 18.47)
```

1) Calculer la moyenne, la variance et l'écart-type.
2) Sachant que l'indice moyen de référence pour des jeunes filles du même âge est de 19.39, peut-on conclure que l'indice de Quételet moyen de l'échantillon est identique à l'indice moyen de référence ?
3) Quelle conclusion en tirez-vous ?



5 Tabac et Cancer

On a suivi, sur une période de 20 ans, deux cohortes : 200 sujets fumeurs et 200 sujets non fumeurs. On a noté le nombre d'apparition de cancer dans chacune des cohortes : 40 chez les fumeurs ; 20 chez les non fumeurs. La différence d'apparition de cancer dans les deux cohortes est-elle significative ?

```
tabac <- matrix(c(40, 160, 20, 180), 2)
colnames(tabac) = c("coh1", "coh2")
rownames(tabac) = c("fumeur", "non_fumeur")
tabac
           coh1 coh2
fumeur       40   20
non_fumeur  160  180
```

Si nous appliquons la valeur de la statistique du test à ces données :

$$z = \frac{f_1 - f_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

avec $\hat{p} = \frac{k_1+k_2}{n_1+n_2}$ et $\hat{q} = 1 - \hat{p}$.

```
f1 <- 40/200
f1
[1] 0.2
f2 <- 20/200
f2
[1] 0.1
pest <- (40 + 20)/(200 + 200)
pest
[1] 0.15
z <- (f1 - f2)/sqrt(pest * (1 - pest) * ((1/200) + (1/200)))
z
[1] 2.80056
2 * (1 - pnorm(z))
[1] 0.0051014
```

La procédure dans R pour comparer deux proportions observées est :

```
prop.test(c(40, 20), c(200, 200), correct = F)
        2-sample test for equality of proportions without continuity correction
data:  c(40, 20) out of c(200, 200)
X-squared = 7.8431, df = 1, p-value = 0.005101
alternative hypothesis: two.sided
95 percent confidence interval:
 0.03070481 0.16929519
sample estimates:
prop 1 prop 2
   0.2    0.1
```

Le lien entre les deux est donné par $z^2 = \chi^2$.

```
z^2
[1] 7.843137
```

6 Lecture et Apprentissage

Deux institutrices spécialisées désirent comparer deux méthodes d'apprentissage de lecture différentes M1 et M2. Elles choisissent chacune une méthode qu'elles appliquent dans leurs classes respectives. Pour la méthode M1, sur 25 enfants, l'institutrice compte 60 % d'enfants ayant bien appris à lire ; pour la méthode M2, sur 20 enfants, on en a compté 50%. Que peut-on conclure sur l'efficacité de ces deux méthodes ?

7 Retard de croissance et Indice de Quételet

Précédemment, nous avons comparé la moyenne de l'indice de Quételet chez 13 jeunes filles atteintes du syndrôme de Turner.

```
turner <- quetelet
```

Nous possédons également l'indice de Quételet chez 8 jeunes filles atteintes d'un déficit en hormone de croissance.

```
deficit <- c(16.79, 19.22, 23, 20.52, 16.06, 20, 16.61, 17.81)
deficit
[1] 16.79 19.22 23.00 20.52 16.06 20.00 16.61 17.81
```

Existe-t-il une différence, en moyenne, de l'indice de Quételet entre ces deux pathologies de croissance ?

Si nous appliquons la valeur de la statistique du test à ces données :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

avec

$$\widehat{\sigma^2} = \frac{(n_1 - 1)\widehat{\sigma_1^2} + (n_2 - 1)\widehat{\sigma_2^2}}{n_1 + n_2 - 2}$$

```
valt <- (mean(turner) - mean(deficit))/sqrt((12 * var(turner) +
    7 * var(deficit))/(13 + 8 - 2) * (1/13 + 1/8))
valt
[1] -1.031742
2 * pt(valt, 13 + 8 - 2)
[1] 0.3151481
```

```
t.test(turner, deficit, var.eq = T)
        Two Sample t-test
data:  turner and deficit
t = -1.0317, df = 19, p-value = 0.3151
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.524538  0.857423
sample estimates:
mean of x mean of y
 17.91769  18.75125
```

Le dernier argument porte sur l'égalité des variances $F = \frac{\widehat{\sigma_1^2}}{\widehat{\sigma_2^2}}$.

```
var.test(turner, deficit)
        F test to compare two variances
data:  turner and deficit
F = 0.3232, num df = 12, denom df = 7, p-value = 0.08294
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.06927673 1.16574614
sample estimates:
ratio of variances
         0.3232334
```

8 Dépense énergétique et poids

Les données proviennent du package `ISwR` et contiennent les dépenses énergétiques sur une période de 24 heures (en MJ) en fonction du statut pondéral obèse (`obese`) ou mince (`lean`) de 22 femmes. Peut-on mettre en évidence une différence de dépense énergétique selon le statut pondéral ?

```
depenergie <- c(9.21, 7.53, 7.48, 8.08, 8.09, 10.15, 8.4, 10.88,
    6.13, 7.9, 11.51, 12.79, 7.05, 11.85, 9.97, 7.48, 8.79, 9.69,
    9.68, 7.58, 9.19, 8.11)
statpond <- c("obese", "lean", "lean", "lean", "lean", "lean", "lean",
    "lean", "lean", "lean", "obese", "obese", "lean", "obese", "obese",
    "lean", "obese", "obese", "obese", "lean", "obese", "lean")
statpond <- factor(statpond)
```

Références

[1] A.B. Dufour, M. Pagès, and M. Sempé. L'analyse de la variance utilisée pour étudier l'évolution de l'indice de corpulence avec l'âge selon les différentes pathologies de croissance. *Cahiers d'Anthropologie et Biométrie Humaine*, 14(1-2) :79–89, 1996.