

A photograph of a modern building with a glass facade and a courtyard. The building has a white brick wall on the left and a glass facade on the right. The courtyard is filled with green plants and a paved path. The sky is blue with a few clouds.

Les tests d'hypothèses

Vincent Guillemot

Amoury Vaysse

Vendredi 18/11/2021

swiss post

OMICS

Rappels

- Hypothèse nulle \mathcal{H}_0 , c'est l'hypothèse du *statu quo*
- Hypothèse alternative \mathcal{H}_1 , c'est la situation intéressante ! (signal)
- α : risque de première espèce, rejeter \mathcal{H}_0 lorsqu'elle est vraie ("erreur de détection")
- β : risque de deuxième espèce, ne pas rejeter \mathcal{H}_0 alors que \mathcal{H}_1 est vraie ("rater un signal")
- Puissance : $1 - \beta$

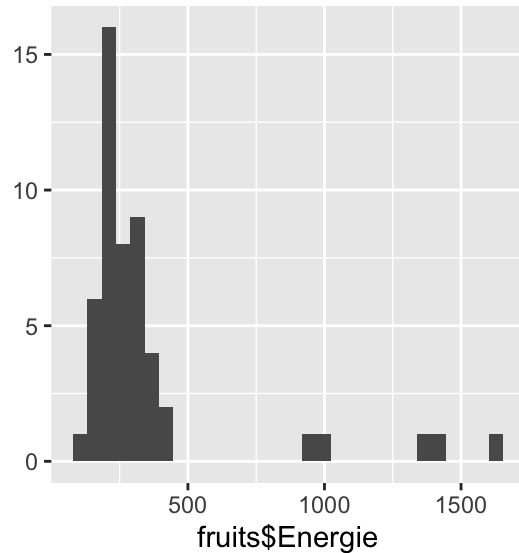
Décision / Verité	Non rejet de \mathcal{H}_0	Rejet \mathcal{H}_0
\mathcal{H}_0	Confiance	Erreur de 1ère esp.
\mathcal{H}_1	Erreur de 2ème esp.	Puissance

Expérience de Chastaing

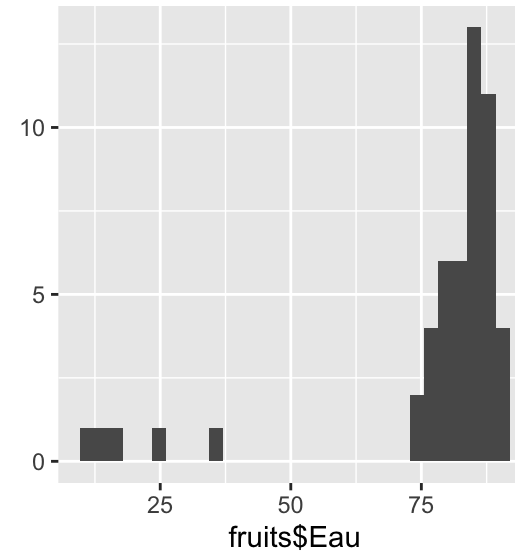
L'hypothèse nulle

Création d'un exemple

```
qplot(fruits$Energie)
```



```
qplot(fruits$Eau)
```



```
energiequal <- cut(fruits$Energie,  
  c(0, 250, 2000))
```

```
eauqual <- cut(fruits$Eau,  
  c(0, 85, 100))
```

Table de contingence

Une table de contingence, ou table de comptage, est un tableau croisé (de comptage) entre deux variables qualitatives ou plus.

```
(tab <- table(energiequal, eauqual))  
#>           eauqual  
#> energiequal  (0,85] (85,100]  
#>  (0,250]           3       23  
#>  (250,2e+03]      25        0
```

On peut aussi calculer les proportions

```
prop.table(tab)  
#>           eauqual  
#> energiequal  (0,85]  (85,100]  
#>  (0,250]      0.05882353 0.45098039  
#>  (250,2e+03] 0.49019608 0.00000000
```

Profils lignes et profils colonnes

Proportions conditionnellement aux lignes :

Proportions conditionnellement aux colonnes :

```
prop.table(tab, margin = 1)
#>          eauqual
#> energiequal (0,85] (85,100]
#>  (0,250]    0.1153846 0.8846154
#>  (250,2e+03] 1.0000000 0.0000000
```

```
prop.table(tab, margin = 2)
#>          eauqual
#> energiequal (0,85] (85,100]
#>  (0,250]    0.1071429 1.0000000
#>  (250,2e+03] 0.8928571 0.0000000
```

Comparer des proportions

Avec la fonction `prop.test` :

```
prop.test(table(energiequal, eauqual))
#>
#> 2-sample test for equality of proportions with
#> continuity correction
#>
#> data:  table(energiequal, eauqual)
#> X-squared = 36.788, df = 1, p-value = 1.317e-09
#> alternative hypothesis: two.sided
#> 95 percent confidence interval:
#> -1.0000000 -0.7225806
#> sample estimates:
#>      prop 1      prop 2
#> 0.1153846 1.0000000
```

Attention, le test des proportions a besoin de données de comptage, pour lui :

$$\frac{2}{4} \neq \frac{50}{100}$$

La fonction `prop.test`

- Accepte des tables de contingences,
- Ou bien deux vecteurs : `x` pour les “succès”, `n` pour le nombre total,
- Eventuellement un vecteur de proportions de référence `p`

Un des exemples de la fonction (cf. `?prop.test`) :

```
smokers <- c( 83, 90, 129, 70 )
patients <- c( 86, 93, 136, 82 )
prop.test(smokers, patients)
#>
#> 4-sample test for equality of proportions without
#> continuity correction
#>
#> data: smokers out of patients
#> X-squared = 12.6, df = 3, p-value = 0.005585
#> alternative hypothesis: two.sided
#> sample estimates:
#>      prop 1      prop 2      prop 3      prop 4
#> 0.9651163 0.9677419 0.9485294 0.8536585
```

Test du “khi-deux”

Avec la fonction `chisq.test` :

```
chisq.test(energiequal, eauqual)
#>
#>  Pearson's Chi-squared test with Yates' continuity
#>  correction
#>
#> data:  energiequal and eauqual
#> X-squared = 36.788, df = 1, p-value = 1.317e-09
```

La fonction `chisq.test`

- Accepte deux variables qualitatives,
- Ou une table de contingence

Un des exemples de la fonction (cf. `?chisq.test`):

```
M <- as.table(rbind(c(762, 327, 468), c(484, 239, 477)))
dimnames(M) <- list(gender = c("F", "M"),
                    party = c("Democrat", "Independent", "Republican"))
(Xsq <- chisq.test(M)) # Prints test summary
#>
#> Pearson's Chi-squared test
#>
#> data:  M
#> X-squared = 30.07, df = 2, p-value = 2.954e-07
Xsq$expected # expected counts under the null
#>      party
#> gender Democrat Independent Republican
#>      F 703.6714      319.6453      533.6834
#>      M 542.3286      246.3547      411.3166
```

La statistique du χ^2

Elle compare les fréquences observées aux fréquences attendues. Les fréquences attendues sont calculées à partir des fréquences marginales sous hypothèse d'indépendance.

$$\chi^2 = \sum \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n} \right)^2}{\frac{n_{i.} n_{.j}}{n}},$$

avec n_{ij} l'effectif observé, $n_{i.}$ l'effectif marginal ligne, $n_{.j}$ l'effectif marginal colonne et n l'effectif total.

Rappel : quand A et B sont indépendants, $P(A \cap B) = P(A)P(B)$.

Test exact de Fisher

Avec la fonction `fisher.test` :

```
fisher.test(energiequal, eauqual)
#>
#> Fisher's Exact Test for Count Data
#>
#> data:  energiequal and eauqual
#> p-value = 1.474e-11
#> alternative hypothesis: true odds ratio is not equal to 1
#> 95 percent confidence interval:
#>  0.00000000 0.03850145
#> sample estimates:
#> odds ratio
#>          0
```

Comparer des moyennes

Avec la fonction `t.test` :

```
t.test(fruits$VitamineC ~ eauqual)
#>
#> Welch Two Sample t-test
#>
#> data:  fruits$VitamineC by eauqual
#> t = -1.6272, df = 37.768, p-value = 0.112
#> alternative hypothesis: true difference in means between group (0,85] and group (85,100]
#> 95 percent confidence interval:
#> -21.202176  2.308077
#> sample estimates:
#> mean in group (0,85] mean in group (85,100]
#>          10.82643          20.27348
```

Les formules

Les formules permettent à l'utilisateur de décrire un modèle :

$$Y = X_1 + X_2 + X_3 + X_2 * X_3 + X_3 * X_4$$

deviendra

$$y \sim x1 + x2 * x3 + x3:x4$$

Repérez le tilde sur votre clavier, il est très important en R !

Comparer des moyennes sur données appariées

Equivalent non-paramétrique

ANOVA

ANOVA non paramétrique

Modèles linéaires