

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Tests

Eric Marcon

15 février 2024

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Motivations

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Savoir tester un ensemble de valeurs observées, y_i , issues d'une variable aléatoire Y , contre plusieurs hypothèses nulles :

- une valeur est un tirage vraisemblable de Y ;
- une autre distribution, éventuellement appariée, provient de Y ;
- une autre distribution provient de X qui est corrélée à Y ;

Les tests peuvent être paramétrique, c'est-à-dire s'appuyer sur des lois connues (la loi normale surtout), ou non paramétrique, et s'appuyer sur les rangs.

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contre une valeur

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte: $Y \sim \mathcal{N}(\mu, \sigma^2)$.

```
n <- 1000 ; mu <- 1 ; sigma = 10  
Y <- rnorm(n, mean = mu, sd = sigma)
```

Hypothèse nulle : $\mu = 0$.

Test de Student : `t.test(Y, mu = 0)`

Modèle linéaire : `lm(Y ~ 1)`

→ Faire varier n et σ .

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte: Y n'est pas forcément distribué normalement.

Hypothèse nulle : $\mu = 0$.

Test de Wilcoxon : `wilcox.test(Y, mu = 0)`

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contre une distribution

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Teste l'hypothèse que deux échantillons sont issus de la même distribution normale

```
Y_double <- rnorm(2 * n, mean = mu, sd = sigma)
Y_1 <- Y_double[1:n] ; Y_2 <- Y_double[(n + 1):(2 * n)]
ks.test(Y_1, Y_2)
```

```
##
## Asymptotic two-sample Kolmogorov-Smirnov
## test
##
## data: Y_1 and Y_2
## D = 0.035, p-value = 0.5727
## alternative hypothesis: two-sided
```

→ Interpréter `ks.test(Y_1, Y_2 + 2)` (faire un graphique)

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte:

- $Y \sim \mathcal{N}(\mu, \sigma^2)$
- $Y_0 \sim \mathcal{N}(\mu_0, \sigma^2)$

```
mu_0 <- 2 ;  
Y_0 <- rnorm(n, mean = mu_0, sd = sigma)
```

Hypothèse nulle : $\mu = \mu_0$.

Test de Student : `t.test(Y, Y_0, var.equal = TRUE)`

Modèle linéaire : Anova à un facteur.

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte:

- $Y \sim \mathcal{N}(\mu, \sigma^2)$
- $Y_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$

```
sigma_0 <- 20 ;  
Y_0 <- rnorm(n, mean = mu_0, sd = sigma_0)
```

Hypothèse nulle : $\mu = \mu_0$.

Test de Welch : `t.test(Y, Y_0)`

Modèle linéaire : Anova de Welch.

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte: Y ou Y_0 ne sont pas distribuées normalement.

Hypothèse nulle : $\mu = \mu_0$.

Test U de Mann-Whitney : `wilcox.test(Y, Y_0)`

Modèle linéaire : Anova à un facteur sur les rangs signés.

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contre une distribution appariée

Tests

Eric Marcon

Motivations

Contre une
valeurContre une
distributionContre une
distribution
appariée

Corrélation

Conclusion

Contexte: Y et Y_0 sont deux observations du même phénomène.

- $Y \sim \mathcal{N}(\mu, \sigma^2)$
- $Y_0 \sim \mathcal{N}(\mu_0, \sigma^2)$

```
Y_0 <- Y + rnorm(n, mean = mu_0 - mu)
```

Hypothèse nulle : $\mu = \mu_0$.

Test de Student : `t.test(Y, Y_0, paired = TRUE)`

Modèle linéaire : `lm(Y - Y_0 ~ 1)`

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte:

- Y et Y_0 sont deux observations du même phénomène.
- Y ou Y_0 ne sont pas distribuées normalement.

Hypothèse nulle : $\mu = \mu_0$.

Test de Wilcoxon apparié : `wilcox.test(Y, Y_0, paired = TRUE)`

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Corrélation

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte:

- Y et X sont liées linéairement
- $Y - Y^* \sim \mathcal{N}(0, \sigma^2)$

Hypothèse nulle : $\text{Cor}(X, Y) = 0$.

Test de corrélation : `cor.test(X, Y, method = "Pearson")`

Modèle linéaire : `lm(Y ~ 1 + X)`

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Contexte:

- Y et X sont liées non linéairement, mais la relation est monotone.

Hypothèse nulle : $\text{Cor}(X, Y) = 0$.

Test de corrélation : `cor.test(X, Y, method = "Spearman")`

Modèle linéaire : `lm(Y ~ 1 + X)`

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Simulation de tirages corrélés

```
library(MASS) # Attention à MASS::select()
# Matrice de covariance
Sigma <- matrix(c(1, 0.8, 0.8, 1), nrow = 2)
# Simulation de X et Y
XY <- mvrnorm(n, mu = c(mu, mu_0), Sigma = Sigma)
```

Test:

```
cor.test(XY[, 1], XY[, 2])
```

```
##
## Pearson's product-moment correlation
##
## data: XY[, 1] and XY[, 2]
## t = 39.941, df = 998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.7592478 0.8070751
## sample estimates:
## cor
## 0.7843242
```

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Conclusion

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion

Tous ces tests s'appuient sur le modèle linéaire, y compris les tests non paramétriques qui utilisent le modèle linéaire sur les rangs (éventuellement signés) des variables.

Référence.

Tests

Eric Marcon

Motivations

Contre une
valeur

Contre une
distribution

Contre une
distribution
appariée

Corrélation

Conclusion