

ANOVA

Francisco Caramelo - fcaramelo@fmed.uc.pt

A análise de variância (ANOVA) é um teste de grande importância em estatística. Aplica-se a situações em que existem mais de dois grupos e realiza-se sobre variáveis quantitativas que são independentes, seguem a distribuição normal e são homogêneas relativamente à variância. Apesar destes pressupostos a ANOVA é um teste robusto a violações de normalidade e de homocedasticidade desde que os grupos tenham tamanhos semelhantes.

Continuaremos a usar os dados *iris* para demonstrar e testar o análise de variância (ANOVA).

```
library(datasets)
head(iris,10)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa

O objectivo é determinar se existem diferenças estatisticamente significativas na largura das sépalas entre os três grupos *setosa*, *versicolor* e *virginica*.

Normalidade

Vamos começar por verificar se o pressuposto da normalidade é verificado para em todos os grupos.

```
sum(iris$Species == 'setosa')
```

```
## [1] 50
```

```
sum(iris$Species == 'versicolor')
```

```
## [1] 50
```

```
sum(iris$Species == 'virginica')
```

```
## [1] 50
```

Tendo em conta o ponto de corte igual a 25, iremos proceder a um teste de Shapiro-Wilk para testar a normalidade.

```

options(warn=-1)
setosa_data = iris[iris$Species == 'setosa',]$Sepal.Width
shapiro.test(setosa_data)

##
##  Shapiro-Wilk normality test
##
## data:  setosa_data
## W = 0.9717, p-value = 0.2715

versicolor_data = iris[iris$Species == 'versicolor',]$Sepal.Width
shapiro.test(versicolor_data)

##
##  Shapiro-Wilk normality test
##
## data:  versicolor_data
## W = 0.9741, p-value = 0.338

virginica_data = iris[iris$Species == 'virginica',]$Sepal.Width
shapiro.test(virginica_data)

##
##  Shapiro-Wilk normality test
##
## data:  virginica_data
## W = 0.9674, p-value = 0.1809

```

Teste ANOVA

Uma vez que o pressuposto de normalidade é seguido e o número de elementos na amostra é igual podemos proceder ao teste ANOVA.

```

fit <- aov(Sepal.Width ~ Species, data=iris)
summary(fit)

##              Df Sum Sq Mean Sq F value Pr(>F)
## Species      2  11.35   5.672   49.16 <2e-16 ***
## Residuals   147  16.96   0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Uma vez que o valor de p é inferior a α , rejeita-se a hipótese nula reconhecendo-se que existem diferenças estatisticamente significativas entre pelo menos dois grupos. Para determinar entre que grupos existem diferenças, vamos executar testes posthoc (testes Tukey) de comparações múltiplas.

```
TukeyHSD(fit)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Sepal.Width ~ Species, data = iris)
##
## $Species
##              diff          lwr          upr      p adj
## versicolor-setosa -0.658 -0.81885528 -0.4971447 0.0000000
## virginica-setosa   -0.454 -0.61485528 -0.2931447 0.0000000
## virginica-versicolor 0.204  0.04314472  0.3648553 0.0087802
```

Exercício

1. Realizar o teste de shapiro-Wilk sobre os três grupos presentes na base de dados *iris* para a variável largura da pétala.
2. Interprete os resultados obtidos anteriormente.
3. Independentemente dos resultados de normalidade execute o teste ANOVA sobre a mesma variável factorizada nos mesmos grupos e interprete os resultados.
4. Caso o teste ANOVA seja significativo realize testes para comparações multiplas e interprete os resultados.