ME731 - Métodos em Análise Multivariada
– MANOVA I –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 08

Agenda I

Introdução

Introdução

- Na aula anterior vimos o TRV (um método geral para construir testes de hipóteses).
- Com ajuda do TRV contruimos testes para testar

$$H_0 : \mu = \mu_0 \quad vs. \quad H_1 : \mu \neq \mu_0$$

- Na aula de hoje aprenderemos a testar hipóteses do tipo

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

- Utilizaremos **MANOVA** (*multivariate analisys of variance*), a versão multivariada do ANOVA.

MANOVA

MANOVA

Sejam

- $\mathbf{X}_{11}, \cdots, \mathbf{X}_{1n_1} \in \mathbb{R}^p$ uma a.a da população 1,
- $\mathbf{X}_{21}, \cdots, \mathbf{X}_{2n_2} \in \mathbb{R}^p$ uma a.a da população 2,
- $\cdots$
- $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kn_k} \in \mathbb{R}^p$ uma a.a da população k,

com $N_p(\mu_g, \Sigma)$ $(g = 1, \cdots, k)$ a distribuição da g-éssima população e as a.as. das diferentes populações são independentes.

MANOVA

Sejam

- $\mathbf{X}_{11}, \cdots, \mathbf{X}_{1n_1} \in \mathbb{R}^p$ uma a.a da população 1,
- $\mathbf{X}_{21}, \cdots, \mathbf{X}_{2n_2} \in \mathbb{R}^p$ uma a.a da população 2,
- $\cdots$
- $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kn_k} \in \mathbb{R}^p$ uma a.a da população k,

com $N_p(\mu_g, \Sigma)$ $(g = 1, \cdots, k)$ a distribuição da g-éssima população e as a.as. das diferentes populações são independentes.

MANOVA é utilizado para testar

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

em que μ_i é o vetor de médias da i-éssima população.

MANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

Para encontrar a estatística de teste, utilizaremos o TRV.

MANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

Para encontrar a estatística de teste, utilizaremos o TRV.

A função de verossimilhança é dada por

$$L(\boldsymbol{\theta}) = \frac{1}{|2\pi\Sigma|^{(n_1+\cdots+n_k)/2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\mathbf{X}_{ji} - \mu_j)' \Sigma^{-1} (\mathbf{X}_{ji} - \mu_j) \right)$$

e a log-verossimilhança é dada por

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\mathbf{X}_{ji} - \mu_j)' \Sigma^{-1} (\mathbf{X}_{ji} - \mu_j) - \log(|2\pi\Sigma|^{(n_1+\cdots+n_k)/2})$$

MANOVA

ou, **equivalentemente**,

$$l(\theta) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}Tr(\Sigma^{-1}\sum_{j=1}^{k} n_j \mathbf{S}_j) - \frac{1}{2}\sum_{j=1}^{k} n_j(\bar{\mathbf{X}}_j - \mu_j)'\Sigma^{-1}(\bar{\mathbf{X}}_j - \mu_j).$$

MANOVA

ou, **equivalentemente**,

$$l(\boldsymbol{\theta}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}Tr(\Sigma^{-1}\sum_{j=1}^{k} n_j\mathbf{S}_j) - \frac{1}{2}\sum_{j=1}^{k} n_j(\bar{\mathbf{X}}_j - \mu_j)'\Sigma^{-1}(\bar{\mathbf{X}}_j - \mu_j).$$

Assim, os EMV sob o modelo restrito e o modelo irrestrito são

- $\sup_{\boldsymbol{\theta}\in\Omega_0} l(\boldsymbol{\theta}) \rightarrow \quad \hat{\mu} = \bar{\mathbf{X}} \quad e \quad \hat{\Sigma} = \mathbf{S}.$
- $\sup_{\boldsymbol{\theta}\in\Omega} l(\boldsymbol{\theta}) \rightarrow \quad \hat{\mu}_i = \bar{\mathbf{X}}_i \quad e \quad \hat{\Sigma} = \underbrace{(n_1 + \cdots + n_k)^{-1}}_{n^{-1}} \underbrace{\sum_{i=1}^{k} n_i\mathbf{S}_i}_{\mathbf{W}}.$

MANOVA

Então,

- $l(\bar{\mathbf{X}}, \mathbf{S}) = -\frac{n}{2}\log(|\mathbf{S}|) - \frac{np}{2}\log(2\pi) - \frac{np}{2}.$
- $l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\mathbf{W}) = -\frac{n}{2}\log(|n^{-1}\mathbf{W}|) - \frac{np}{2}\log(2\pi) - \frac{np}{2}.$

MANOVA

Então,

- $l(\bar{\mathbf{X}}, \mathbf{S}) = -\frac{n}{2}\log(|\mathbf{S}|) - \frac{np}{2}\log(2\pi) - \frac{np}{2}.$
- $l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\mathbf{W}) = -\frac{n}{2}\log(|n^{-1}\mathbf{W}|) - \frac{np}{2}\log(2\pi) - \frac{np}{2}.$

A estatística de teste é dada por

$$2\log(\Lambda(\mathbf{X})) = 2\left[l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\mathbf{W}) - l(\bar{\mathbf{X}}, \mathbf{S})\right] = n\log\left(\frac{|n\mathbf{S}|}{|\mathbf{W}|}\right),$$

com região de rejeição $R = \{\mathbf{x} : n\log\left(\frac{|n\mathbf{S}|}{|\mathbf{W}|}\right) > k\} \equiv \{\mathbf{x} : \frac{|\mathbf{W}|}{|n\mathbf{S}|} < c\}$

MANOVA

Quando $n \to \infty$,

$$n \log \left(\frac{|n\mathbf{S}|}{|\mathbf{W}|} \right) \to^{D} \chi^2_{p(k-1)}$$

MANOVA

Quando $n \to \infty$,

$$n \log\left(\frac{|n\mathbf{S}|}{|\mathbf{W}|}\right) \to^{D} \chi^2_{p(k-1)}$$

O que fazer se temos amostras pequenas?

- Pode-se utilizar a seguinte correção para pequenas amostras (veja Peña (2002)),

$$[(n-1)-(p+k)/2] \log\left(\frac{|n\mathbf{S}|}{|\mathbf{W}|}\right) \sim^{approx} \chi^2_{p(k-1)}$$

- Outra opção é utilizar a distribuição exata de $\dfrac{|\mathbf{W}|}{|n\mathbf{S}|}$ que será uma distribuição chamada **Lambda de Wilks** $(\Lambda_p(n-k, k-1))$

MANOVA: Exemplo

O Departamento de saúde e serviço social (DSSS) de Wisconsin reembolsa casas de repouso pelos serviços prestados. As casas de repouso podem ser classificadas como privadas, sem fins lucrativos e governamentais.

MANOVA: Exemplo

O Departamento de saúde e serviço social (DSSS) de Wisconsin reembolsa casas de repouso pelos serviços prestados. As casas de repouso podem ser classificadas como privadas, sem fins lucrativos e governamentais.

O DSSS está interessado em investigar se existe diferença, por tipo de casa de repouso, nos custos de quatro variáveis (que chamaremos X_1, X_2, X_3, X_4).

MANOVA: Exemplo

O Departamento de saúde e serviço social (DSSS) de Wisconsin reembolsa casas de repouso pelos serviços prestados. As casas de repouso podem ser classificadas como privadas, sem fins lucrativos e governamentais.

O DSSS está interessado em investigar se existe diferença, por tipo de casa de repouso, nos custos de quatro variáveis (que chamaremos X_1, X_2, X_3, X_4).

O total de observações por grupo, bem como o vetor de médias e a matriz de covariância são apresentados a seguir.

MANOVA: Exemplo

```r
n_1 <- 271
n_2 <- 138
n_3 <- 107
xbarra_1 <- matrix(c(2.066, 0.480, 0.082, 0.360), ncol = 1)
xbarra_2 <- matrix(c(2.167, 0.596, 0.124, 0.418), ncol = 1)
xbarra_3 <- matrix(c(2.273, 0.521, 0.125, 0.383), ncol = 1)
```

MANOVA: Exemplo

Com matrizes de covariância amostrais

```
cov_1 <- matrix(c(0.291, -0.001, 0.002, 0.010,
                  -0.001, 0.011, 0.000, 0.003,
                   0.002, 0.000, 0.001, 0.000,
                  0.010, 0.003, 0.000, 0.010), ncol = 4)
cov_2 <- matrix(c(0.561, 0.011, 0.001, 0.037,
                  0.011, 0.025, 0.004, 0.007,
                  0.001, 0.004, 0.005, 0.002,
                  0.037, 0.007, 0.002, 0.019), ncol = 4)
cov_3 <- matrix(c(0.261, 0.030, 0.003, 0.018,
                  0.030, 0.017, 0.000, 0.006,
                  0.003, 0.000, 0.004, 0.001,
                  0.018, 0.006, 0.001, 0.013), ncol = 4)
```

MANOVA: Exemplo

```
cov_1
```

```
##            [,1]   [,2]  [,3]  [,4]
## [1,]  0.291 -0.001 0.002 0.010
## [2,] -0.001  0.011 0.000 0.003
## [3,]  0.002  0.000 0.001 0.000
## [4,]  0.010  0.003 0.000 0.010
```

```
cov_2
```

```
##          [,1]  [,2]  [,3]  [,4]
## [1,] 0.561 0.011 0.001 0.037
## [2,] 0.011 0.025 0.004 0.007
## [3,] 0.001 0.004 0.005 0.002
## [4,] 0.037 0.007 0.002 0.019
```

MANOVA: Exemplo

```
cov_3
```

```
##        [,1]  [,2]  [,3]  [,4]
## [1,] 0.261 0.030 0.003 0.018
## [2,] 0.030 0.017 0.000 0.006
## [3,] 0.003 0.000 0.004 0.001
## [4,] 0.018 0.006 0.001 0.013
```

MANOVA: Exemplo

```
cov_3
```

```
##        [,1]  [,2]  [,3]  [,4]
## [1,] 0.261 0.030 0.003 0.018
## [2,] 0.030 0.017 0.000 0.006
## [3,] 0.003 0.000 0.004 0.001
## [4,] 0.018 0.006 0.001 0.013
```

Como proceder?

MANOVA: Exemplo

```
cov_3
```

```
##        [,1]  [,2]  [,3]  [,4]
## [1,] 0.261 0.030 0.003 0.018
## [2,] 0.030 0.017 0.000 0.006
## [3,] 0.003 0.000 0.004 0.001
## [4,] 0.018 0.006 0.001 0.013
```

Como proceder?

1. Quais as hipóteses a serem testadas?
2. Qual a estatística de teste?
3. Qual a decisão?

MANOVA: Exemplo

```
cov_3
```

```
##        [,1]  [,2]  [,3]  [,4]
## [1,] 0.261 0.030 0.003 0.018
## [2,] 0.030 0.017 0.000 0.006
## [3,] 0.003 0.000 0.004 0.001
## [4,] 0.018 0.006 0.001 0.013
```

Como proceder?

1. Quais as hipóteses a serem testadas?
2. Qual a estatística de teste?
3. Qual a decisão?

Com fins ilustrativos assumiremos normalidade e igualdade de variâncias.

MANOVA: Exemplo

1. Definimos as hipóteses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

MANOVA: Exemplo

1. Definimos as hipóteses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

2. Estatística de teste

```
n <- n_1 + n_2 + n_3
```

como $n = 516$ (grande), podemos utilizar

$$n \log \left(\frac{|n\mathbf{S}|}{|\mathbf{W}|} \right) \sim^{approx} \chi^2_{4(3-1)}$$

MANOVA: Exemplo

3. Decisão

Para tomar uma decisão, precisamos primeiro **W** e **S**

```
W <- (n_1 - 1)*cov_1 + (n_2 - 1)*cov_2 + (n_3 - 1)*cov_3
```

MANOVA: Exemplo

3. Decisão

Para tomar uma decisão, precisamos primeiro **W** e **S**

```
W <- (n_1 - 1)*cov_1 + (n_2 - 1)*cov_2 + (n_3 - 1)*cov_3
```

$$n\mathbf{S} = \sum_{j=1}^{3}\sum_{i=1}^{n_j} \underbrace{(\mathbf{X}_{ji} - \bar{\mathbf{X}})(\mathbf{X}_{ji} - \bar{\mathbf{X}})'}_{(\mathbf{X}_{ji}-\bar{\mathbf{X}}_j+\bar{\mathbf{X}}_j-\bar{\mathbf{X}})(\mathbf{X}_{ji}-\bar{\mathbf{X}}_j+\bar{\mathbf{X}}_j-\bar{\mathbf{X}})'} = \sum_{j=1}^{3}[n_j\mathbf{S}_j + n_j(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)']$$

MANOVA: Exemplo

3. Decisão

Para tomar uma decisão, precisamos primeiro **W** e **S**

```
W <- (n_1 - 1)*cov_1 + (n_2 - 1)*cov_2 + (n_3 - 1)*cov_3
```

$$n\mathbf{S} = \sum_{j=1}^{3}\sum_{i=1}^{n_j} \underbrace{(\mathbf{X}_{ji} - \bar{\mathbf{X}})(\mathbf{X}_{ji} - \bar{\mathbf{X}})'}_{(\mathbf{X}_{ji} - \bar{\mathbf{X}}_j + \bar{\mathbf{X}}_j - \bar{\mathbf{X}})(\mathbf{X}_{ji} - \bar{\mathbf{X}}_j + \bar{\mathbf{X}}_j - \bar{\mathbf{X}})'} = \sum_{j=1}^{3}[n_j\mathbf{S}_j + n_j(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)']$$

$$n\mathbf{S} = \mathbf{W} + \underbrace{\sum_{j=1}^{3} n_j(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)(\bar{\mathbf{X}} - \bar{\mathbf{X}}_j)'}_{\mathbf{B}}$$

MANOVA: Exemplo

```
xbarra <- (n_1*xbarra_1 + n_2*xbarra_2 + n_3*xbarra_3)/n
B <- n_1*(xbarra - xbarra_1) %*% t(xbarra - xbarra_1) +
     n_2*(xbarra - xbarra_2) %*% t(xbarra - xbarra_2) +
     n_3*(xbarra - xbarra_3) %*% t(xbarra - xbarra_3)
nS <- B + W
```

MANOVA: Exemplo

```
xbarra <- (n_1*xbarra_1 + n_2*xbarra_2 + n_3*xbarra_3)/n
B <- n_1*(xbarra - xbarra_1) %*% t(xbarra - xbarra_1) +
     n_2*(xbarra - xbarra_2) %*% t(xbarra - xbarra_2) +
     n_3*(xbarra - xbarra_3) %*% t(xbarra - xbarra_3)
nS <- B + W
```

Precisamos calcular $n \log \left(\frac{|n\mathbf{S}|}{|\mathbf{W}|} \right)$

```
est_teste <- n*log(det(nS)/det(W)) # 139.7401
ifelse(est_teste > qchisq(1 - 0.05, 8),
       "Rejeito H0",
       "Não rejeito H0")
```

```
## [1] "Rejeito H0"
```

MANOVA: Exemplo

O *data set* disponível aqui contem 20 observações nas quais foram medidas 5 variáveis:

- $V1$: taxa de extrusão (baixa, alta)
- $V2$: quantidade de um aditivo (baixa, alta)
- $V3$: resistência (ao rasgo)
- $V4$: brilho
- $V5$: opacidade

MANOVA: Exemplo

O *data set* disponível aqui contem 20 observações nas quais foram medidas 5 variáveis:

- $V1$: taxa de extrusão (baixa, alta)
- $V2$: quantidade de um aditivo (baixa, alta)
- $V3$: resistência (ao rasgo)
- $V4$: brilho
- $V5$: opacidade

Assumindo normalidade e variâncias iguais, teste se o vetor de médias de reistência, brilho e opacidade é o mesmo para as difentes taxas de extrusão ($V1$).

MANOVA: Exemplo

```
library(dplyr)
dados = read.table("https://raw.githubusercontent.com/ctruciosm/.
glimpse(dados)
```

```
## Rows: 20
## Columns: 5
## $ V1 <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
## $ V2 <int> 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1,
## $ V3 <dbl> 6.5, 6.2, 5.8, 6.5, 6.5, 6.9, 7.2, 6.9, 6.1, 6.3, 6
## $ V4 <dbl> 9.5, 9.9, 9.6, 9.6, 9.2, 9.1, 10.0, 9.9, 9.5, 9.4,
## $ V5 <dbl> 4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0, 3.9, 1.9, 5.7, 2
```

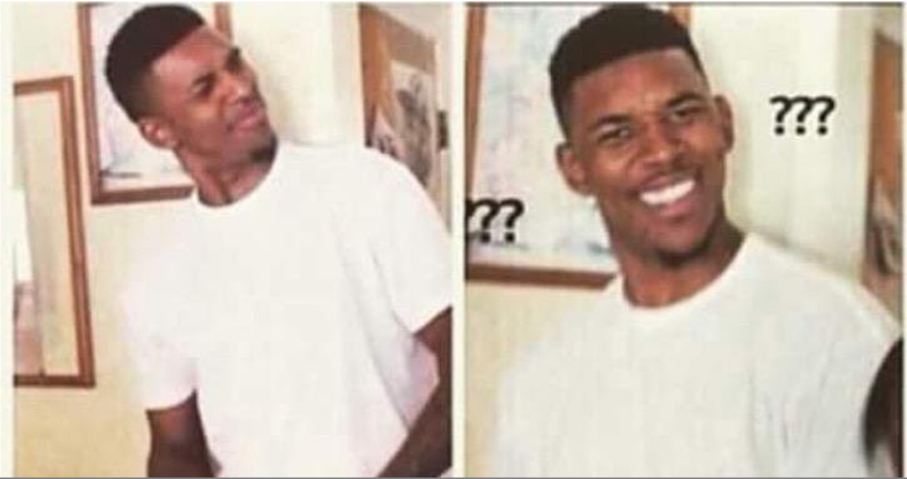
MANOVA: Exemplo

```
manova_test <- manova(cbind(V3, V4, V5) ~ V1, data = dados)
summary(manova_test, test = 'Wilks')
```

```
##             Df   Wilks approx F num Df den Df   Pr(>F)
## V1           1 0.41362    7.561      3     16 0.002273 **
## Residuals 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MANOVA: Exemplo

Estatística F? mas não era distribuição χ^2?



MANOVA:

Distribuição Λ de Wilks

Sejam $\mathbf{A} \sim W_p(\mathbf{I}, m)$ e $\mathbf{B} \sim W_p(\mathbf{I}, n)$ independentes com $m, n \geq p$. Então,

$$\frac{|\mathbf{A}|}{|\mathbf{A}+\mathbf{B}|} \sim \Lambda(p, m, n) \equiv \Lambda_p(m, n)$$

($\Lambda(p, m, n)$ denota uma distribuição Λ de Wilks com parâmetros, p, m e n).

Para alguns valores de p, m e n podemos transformar Λ em uma distribuição F. Para outros valores, temos uma aproximação.

MANOVA

$$\mathbf{A} \sim W_p(\mathbf{I}, m), \mathbf{B} \sim W_p(\mathbf{I}, n) \to |\mathbf{A}|/|\mathbf{A}+\mathbf{B}| \sim \Lambda_p(m, n)$$

Parâmetros	Distribuição
$p \geq 1, \; n = 1$	$\frac{1-\Lambda}{\Lambda} \times \frac{m-p+1}{p} \sim F_{(p,m-p+1)}$
$p \geq 1, \; n = 2$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \times \frac{m-p+1}{p} \sim F_{(2p,2(m-p+1)}$
$p = 1, \; n \geq 1$	$\frac{1-\Lambda}{\Lambda} \times \frac{m}{n} \sim F_{n,m}$
$p = 2, \; n \geq 1$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \times \frac{m-1}{n} \sim F_{2n,2(m-1)}$

MANOVA

Para outros valores de p e n podemos utilizar a seguinte aproximação.

$$\frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \times \frac{rt-(pn-2)/2}{pn} \sim^{approx} F_{pn, rt-(pn-2)/2}$$

- $r = m + n - (p + n + 1)/2$
- $t = \sqrt{\frac{p^2n^2-4}{p^2+n^2-5}}$ (desde que $p^2 + n^2 - 5 > 0$, caso contrario $t = 1$).

MANOVA: Exemplo

Voltando ao nosso exemplo...

$$\Lambda = \frac{|\mathbf{W}|}{|n\mathbf{S}|} \sim \Lambda_{\underbrace{p}_{3}}(\underbrace{n-k}_{20-2}, \underbrace{k-1}_{2-1})$$

MANOVA: Exemplo

Voltando ao nosso exemplo...

$$\Lambda = \frac{|\mathbf{W}|}{|n\mathbf{S}|} \sim \Lambda_{\underbrace{p}_{3}}(\underbrace{n-k}_{20-2}, \underbrace{k-1}_{2-1})$$

Então,

$$\frac{1-\Lambda}{\Lambda} \times \frac{n-k-p+1}{p} \sim F_{(p, \underbrace{n-k-p+1}_{16})}$$

MANOVA: Exemplo

```
p <- 3; k <- 2
n_1 <- sum(dados$V1 == 0)
n_2 <- sum(dados$V1 == 1)
n <- n_1 + n_2
S_1 <- dados %>% filter(V1 == 0) %>% select(V3, V4, V5) %>% cov()
S_2 <- dados %>% filter(V1 == 1) %>% select(V3, V4, V5) %>% cov()
S <- dados %>% select(V3, V4, V5) %>% cov()
S <- (n - 1)/n * S
W <- (n_1 - 1)*S_1 + (n_2 - 1)*S_2
Lambda <- det(W)/det(n*S)
Lambda
```

```
## [1] 0.4136192
```

MANOVA: Exemplo

$$\frac{1-\Lambda}{\Lambda} \times \frac{n-k-p+1}{p} \sim F_{(p,\underbrace{n-k-p+1}_{16})}$$

```
F0 <- (1 - Lambda)/Lambda * (n - k - p + 1)/p
F0
```

```
## [1] 7.560974
```

```
qf(1 - 0.05, p, n - k - p + 1)
```

```
## [1] 3.238872
```

```
1 - pf(F0, p, n - k - p + 1) #pvalor
```

```
## [1] 0.002273044
```

MANOVA

Outras estatística utilizadas para testar

$$H_0 : \mu_1 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade, são:}$$

- Traço Lawley-Hotelling $= Tr[\mathbf{BW}^{-1}]$.
- Traço de Pillai $= Tr[\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}]$.
- Roy $=$ maior autovalor de $\mathbf{W}(\mathbf{B}+\mathbf{W})^{-1}$

MANOVA

Outras estatística utilizadas para testar

$$H_0 : \mu_1 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade, são:}$$

- Traço Lawley-Hotelling $= Tr[\mathbf{BW}^{-1}]$.
- Traço de Pillai $= Tr[\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}]$.
- Roy $=$ maior autovalor de $\mathbf{W}(\mathbf{B}+\mathbf{W})^{-1}$

```
summary(manova_test, test = 'Hotelling-Lawley')
summary(manova_test, test = 'Roy')
summary(manova_test, test = 'Pillai')
```

Todas elas podem ser aproximadas por uma distribuição F.

Two-Way MANOVA

Two-Way MANOVA

- Versão multivariada do Two-Way ANOVA
- One-Way MANOVA,

$$H_0 : \mu_1 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

Alternativamente, podemos escrever $\mathbf{X}_{ji} = \underbrace{\mu + \tau_j}_{\mu_j} + \underbrace{\varepsilon_{ji}}_{iid \sim N_p(0,\Sigma)}$ e testar

$$H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

Two-Way MANOVA

- Versão multivariada do Two-Way ANOVA
- One-Way MANOVA,

$$H_0 : \mu_1 = \cdots = \mu_k \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

Alternativamente, podemos escrever $\mathbf{X}_{ji} = \underbrace{\mu + \tau_j}_{\mu_j} + \underbrace{\varepsilon_{ji}}_{iid \sim N_p(0,\Sigma)}$ e testar

$$H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade.}$$

$$(\text{com } \sum_{j=1}^{k} n_j \tau_j = 0)$$

Two-Way MANOVA

- Two-Way MANOVA, $\mathbf{X}_{ijr} = \underbrace{\mu + \tau_i + \beta_j + \gamma_{ij}}_{\mu_j} + \underbrace{\varepsilon_{ijr}}_{N_p(0,\Sigma)}$. Podemos então definir as hipóteses,

$$H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

$$H_0 : \beta_1 = \cdots = \beta_g = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

$$H_0 : \gamma_{11} = \cdots = \gamma_{kg} = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

Two-Way MANOVA

- Two-Way MANOVA, $\mathbf{X}_{ijr} = \underbrace{\mu + \tau_i + \beta_j + \gamma_{ij}}_{\mu_j} + \underbrace{\varepsilon_{ijr}}_{N_p(0,\Sigma)}$. Podemos então definir as hipóteses,

$$H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

$$H_0 : \beta_1 = \cdots = \beta_g = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

$$H_0 : \gamma_{11} = \cdots = \gamma_{kg} = 0 \quad vs. \quad H_1 : H_0 \text{ não é verdade},$$

Two-Way MANOVA nos ajudará a verificar que existe diferença nas medias pelo fator 1 (τ com k grupos), pelo fator 2 (β com g grupos) ou pela interação de ambos (γ com kg grupos).

Two-Way MANOVA

Com as três hipóteses definidas, podemos obter estatísticas de teste para cada uma delas através do mesmo procedimento utilizado para o **One-Way MANOVA**

Two-Way MANOVA

Com as três hipóteses definidas, podemos obter estatísticas de teste para cada uma delas através do mesmo procedimento utilizado para o **One-Way MANOVA**

Para testar $H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0$ não é verdade, utilizamos

$$\Lambda_\tau = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_\tau|} \sim \Lambda_p(n - kg, k - 1)$$

Two-Way MANOVA

Com as três hipóteses definidas, podemos obter estatísticas de teste para cada uma delas através do mesmo procedimento utilizado para o **One-Way MANOVA**

Para testar $H_0 : \tau_1 = \cdots = \tau_k = 0 \quad vs. \quad H_1 : H_0$ não é verdade, utilizamos

$$\Lambda_\tau = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_\tau|} \sim \Lambda_p(n - kg, k - 1)$$

Para testar $H_0 : \beta_1 = \cdots = \beta_g = 0 \quad vs. \quad H_1 : H_0$ não é verdade, utilizamos

$$\Lambda_\beta = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_\beta|} \sim \Lambda_p(n - kg, g - 1)$$

Two-Way MANOVA

Para testar $H_0 : \gamma_{11} = \cdots = \gamma_{kg} = 0 \quad vs. \quad H_1 : H_0$ não é verdade, utilizamos

$$\Lambda_\gamma = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_\gamma|} \sim \Lambda_p(n - kg, (k-1)(g-1)), \quad \text{em que:}$$

Two-Way MANOVA

Para testar $H_0 : \gamma_{11} = \cdots = \gamma_{kg} = 0 \quad vs. \quad H_1 : H_0$ não é verdade, utilizamos

$$\Lambda_\gamma = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_\gamma|} \sim \Lambda_p(n - kg, (k-1)(g-1)), \quad \text{em que:}$$

- $\mathbf{S}_e = \sum_{i=1}^{k} \sum_{j=1}^{g} \sum_{k=1}^{n_{ij}} (\mathbf{X}_{ijr} - \bar{\mathbf{X}}_{ij.})(\mathbf{X}_{ijr} - \bar{\mathbf{X}}_{ij.})'$,
- $\mathbf{S}_\gamma = \sum_{i=1}^{k} \sum_{j=1}^{g} n_{ij}(\bar{\mathbf{X}}_{ij.} - \bar{\mathbf{X}}_{i..} - \bar{\mathbf{X}}_{.j.} + \bar{\mathbf{X}}_{...})(\bar{\mathbf{X}}_{ij.} - \bar{\mathbf{X}}_{i..} - \bar{\mathbf{X}}_{.j.} + \bar{\mathbf{X}}_{...})'$,
- $\mathbf{S}_\tau = \sum_{i=1}^{k} n_{i.}(\bar{\mathbf{X}}_{i..} - \bar{\mathbf{X}}_{...})(\bar{\mathbf{X}}_{i..} - \bar{\mathbf{X}}_{...})' \quad e \quad \mathbf{S}_\beta = \sum_{j=1}^{g} n_{.j}(\bar{\mathbf{X}}_{.j.} - \bar{\mathbf{X}}_{...})(\bar{\mathbf{X}}_{.j.} - \bar{\mathbf{X}}_{...})'$.

Two-Way MANOVA

$$n = \sum_{i=1}^{k} \sum_{j=1}^{g} n_{ij},$$

com n_{ij} sendo o número de elementos que pertencem ao grupo i do fator 1 e ao grupo j do fator 2.

Two-Way MANOVA: Exemplo

O *data set* disponível aqui contem 20 observações nas quais foram medidas 5 variáveis:

- $V1$: taxa de extrusão (baixa, alta)
- $V2$: quantidade de um aditivo (baixa, alta)
- $V3$: resistência (ao rasgo)
- $V4$: brilho
- $V5$: opacidade

Two-Way MANOVA: Exemplo

O *data set* disponível aqui contem 20 observações nas quais foram medidas 5 variáveis:

- $V1$: taxa de extrusão (baixa, alta)
- $V2$: quantidade de um aditivo (baixa, alta)
- $V3$: resistência (ao rasgo)
- $V4$: brilho
- $V5$: opacidade

Assumindo normalidade e variâncias iguais, teste se existe diferença no vetor de médias ($V3$, $V4$, $V5$), entre as taxas de extrusão e quantidade de aditivo.

Two-Way MANOVA: Exemplo

- Fator 1: $V1$ taxa de extrusão (baixa, alta).
- Fator 2: $V2$: quantidade de um aditivo (baixa, alta)

Two-Way MANOVA: Exemplo

- Fator 1: $V1$ taxa de extrusão (baixa, alta).
- Fator 2: $V2$: quantidade de um aditivo (baixa, alta)

Queremos testar

- $H_0 : \tau_1 = \tau_2 = 0$
- $H_0 : \beta_1 = \beta_2 = 0$
- $H_0 : \gamma_{11} = \cdots = \gamma_{22} = 0$

Two-Way MANOVA: Exemplo

```
manova_two_way <- manova(cbind(V3, V4, V5) ~ V1*V2,
                         data = dados)
summary(manova_two_way, test = 'Wilks')
```

```
##           Df   Wilks approx F num Df den Df   Pr(>F)
## V1         1 0.38186   7.5543      3     14 0.003034 **
## V2         1 0.52303   4.2556      3     14 0.024745 *
## V1:V2      1 0.77711   1.3385      3     14 0.301782
## Residuals 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Editon. Springer Nature. Capítulo 7.
- Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis. Sixth Edition. Pearson Prentice Hall. Capítulo 6.
- Mardia, K. V., Kent, J. T., & Bibby, J, M. (1979). Multivariate Analysis. Academic Press. Capítulo 12.