ME731 - Métodos em Análise Multivariada
– MANOVA II –

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

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

Aula 09

Agenda I

Introdução

Introdução

- One-Way e Two-Way MANOVA foram introduzidos.
- Foram derivadas as estatísticas de teste e foram também discutidos os resultados obtidos no software R.
- Na aula de hoje:
  - aprenderemos a testar igualdade das matrizes de covariância,
  - aprenderemos algumas ferramentas que nos ajudarão na interpretação,
  - aprenderemos a lidar quando algumas das suposições do modelo não são verificadas.

Testando a igualdade das matrizes de covariância

Testando a igualdade das matrizes de covariância

Sejam

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

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

Queremos testar:

$$H_0 : \Sigma_1 = \cdots = \Sigma_k \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

Testando a igualdade das matrizes de covariância

$$H_0 : \Sigma_1 = \cdots = \Sigma_k \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

TRV:

- $\sup_{\theta \in \Omega_0} l(\theta) \rightarrow \quad \hat{\mu}_i = \bar{\mathbf{X}}_i \quad e \quad \hat{\Sigma} = n^{-1} \sum_{i=1}^{k} n_i \mathbf{S}_i$
- $\sup_{\theta \in \Omega} l(\theta) \rightarrow \quad \hat{\mu}_i = \bar{\mathbf{X}}_i \quad e \quad \hat{\Sigma}_i = \mathbf{S}_i$

Testando a igualdade das matrizes de covariância

$$H_0 : \Sigma_1 = \cdots = \Sigma_k \quad vs. \quad H_1 : H_0 \text{ não é verdade}$$

TRV:

- $\sup_{\theta \in \Omega_0} l(\theta) \rightarrow \quad \hat{\mu}_i = \bar{\mathbf{X}}_i \quad e \quad \hat{\Sigma} = n^{-1} \sum_{i=1}^{k} n_i \mathbf{S}_i$
- $\sup_{\theta \in \Omega} l(\theta) \rightarrow \quad \hat{\mu}_i = \bar{\mathbf{X}}_i \quad e \quad \hat{\Sigma}_i = \mathbf{S}_i$

Agora vamos obter $l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1} \sum_{i=1}^{k} n_i \mathbf{S}_i)$ e $l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, \mathbf{S}_1, \cdots, \mathbf{S}_k)$

Testando a igualdade das matrizes de covariância

Note que podemos escrever a verrosimilhança como

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{n_1/2}\ldots|\Sigma_k|^{n_k/2}} e^{-\frac{1}{2}\sum_{i=1}^{k} Tr(\Sigma_i^{-1}\sum_{j=1}^{n_i}(X_{ij}-\mu_j)(X_{ij}-\mu_j)')}$$

Testando a igualdade das matrizes de covariância

Note que podemos escrever a verrosimilhança como

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{n_1/2}\dots|\Sigma_k|^{n_k/2}} e^{-\frac{1}{2}\sum_{i=1}^{k} Tr(\Sigma_i^{-1}\sum_{j=1}^{n_i}(X_{ij}-\mu_j)(X_{ij}-\mu_j)')}$$

Então,

- $L(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\sum_{i=1}^{k} n_i\mathbf{S}_i) = \dfrac{1}{(2\pi)^{n/2}|n^{-1}\sum_{i=1}^{k} n_i\mathbf{S}_i|^{n/2}} e^{-\frac{n}{2}}$
- $L(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, \mathbf{S}_1, \cdots, \mathbf{S}_k) = \dfrac{1}{(2\pi)^{n/2}|n_1^{-1}\mathbf{S}_1|^{n_1/2}\dots|n_k^{-1}\mathbf{S}_k|^{n_k/2}} e^{-\frac{n}{2}}$

Testando a igualdade das matrizes de covariância

Quando $n_1, \cdots, n_k \to \infty$,

$$\underbrace{-2[l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, \mathbf{S}_1, \cdots, \mathbf{S}_k) - l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\sum_{i=1}^{k} n_i\mathbf{S}_i)]}_{n \log|\mathbf{S}| - \sum_{i=1}^{k} n_i \log|\mathbf{S}_i| = \sum_{i=1}^{k} n_i \log(|\mathbf{S}_i^{-1}\mathbf{S}|)} \sim \chi^2_{(k-1)p(p+1)/2},$$

com $\mathbf{S} = n^{-1}\sum_{i=1}^{k} n_i\mathbf{S}_i$.

Testando a igualdade das matrizes de covariância

Quando $n_1, \cdots, n_k \to \infty$,

$$\underbrace{-2[l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, \mathbf{S}_1, \cdots, \mathbf{S}_k) - l(\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_k, n^{-1}\sum_{i=1}^{k} n_i \mathbf{S}_i)]}_{n \log|\mathbf{S}| - \sum_{i=1}^{k} n_i \log|\mathbf{S}_i| = \sum_{i=1}^{k} n_i \log(|\mathbf{S}_i^{-1}\mathbf{S}|)} \sim \chi^2_{(k-1)p(p+1)/2},$$

com $\mathbf{S} = n^{-1}\sum_{i=1}^{k} n_i \mathbf{S}_i$.

Assim, rejeitamos H_0 se
$R = \{\mathbf{x} : \sum_{i=1}^{k} n_i \log(|\mathbf{S}_i^{-1}\mathbf{S}|) > \chi^2_{1-\alpha,(k-1)p(p+1)/2}\}$

Testando a igualdade das matrizes de covariância

Para os casos em que os n_is não são muito grandes, Box (1949) propôs a seguinte estatística de teste,

$$M = \gamma \sum (n_i - 1) \log |\mathbf{S}_{ui}^{-1} \mathbf{S}_u| \sim \chi^2_{(k-1)p(p+1)/2},$$

em que $\mathbf{S}_u = \frac{n}{n-k}\mathbf{S}$, $\mathbf{S}_{ui} = \frac{n_i}{n_i - 1}\mathbf{S}_i$ e

$$\gamma = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \times \left(\sum_{i=1}^{k} \frac{1}{n_i - k} - \frac{1}{n-k} \right).$$

Box costuma funcionar bem quando $n_1, \cdots, n_k > 20$ e se $k, p \leq 5$. Em situações quando isso não acontece, Box fornece uma aproximação à distribuição F.

Testando a igualdade das matrizes de covariância: Exemplo

O *data set* `penguins` do pacote `palmerpenguins` contém informação sobre medias de três tipos de penguins. Queremos testar se

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$$

Com fins ilustrativos, assumiremos normalidade.

Testando a igualdade das matrizes de covariância: Exemplo

```
library(palmerpenguins)
library(dplyr)
glimpse(penguins)
```

```
## Rows: 344
## Columns: 8
## $ species           <fct> Adelie, Adelie, Adelie, Adelie, Adel
## $ island            <fct> Torgersen, Torgersen, Torgersen, Tor
## $ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38
## $ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 19
## $ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 36
## $ sex               <fct> male, female, female, NA, female, ma
## $ year              <int> 2007, 2007, 2007, 2007, 2007, 2007,
```

Testando a igualdade das matrizes de covariância: Exemplo

```
dados <- penguins %>%
  dplyr::select(species, ends_with("_mm")) %>%
  na.omit()
biotools::boxM(dados[,-1], dados$species)
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  dados[, -1]
## Chi-Sq (approx.) = 59.682, df = 12, p-value = 2.58e-08
```

MANOVA

For practitioners:

- O teste M de Box é bastante sensível à falta de normalidade (muitas vezes rejeitamos H_0 mas devido à falta de normalidade).
- Para amostras grandes, MANOVA é pouco afetado pela falta de normalidade.
- Com iguais tamanhos de amostra ($n_1 = \cdots = n_k$), diferenças nas matrizes de covariância tem pouca influência no MANOVA

Quando as suposições falham

Quando as suposições falham

Os testes desenvolvidos até agoram assumem:

1. Normalidade Multivariada
2. Igualdade das matrizes de covariância

O que fazer se isto não é verificado?

Quando as suposições falham

Os testes desenvolvidos até agoram assumem:

1. Normalidade Multivariada
2. Igualdade das matrizes de covariância

O que fazer se isto não é verificado?

1. Utilizar transformações,
2. Se conhecemos a distribuição, podemos desenvolver TRV,
3. Teste de Kruskal-Wallis multivariado,
4. Teste de Scheirer–Ray–Hare,
5. Teste de Permutação,
6. Teste Bootstrap.

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.