Análise de variância multivariada

Jhessica Letícia Kirch
Universidade de São Paulo

Simpósio de Microbiologia Agrícola
11 de abril de 2023

RELEMBRANDO

- A Análise de Variância (ANOVA) permite comparar médias de várias populações independentes.

- O teste F é utilizado para verificar diferenças entre médias de populações distintas.

EXEMPLO ANOVA

- Hipótese da ANOVA de um fator:

  - Hipótese nula: a média de todas as populações são iguais, ou seja, o tratamento (fator) não tem efeito (nenhuma variação em média entre os grupos).

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

  - Hipótese alternativa: Pelo menos uma média é diferente, isto é, existe efeito do tratamento (fator).

EXEMPLO ANOVA

- O conjunto de dados Egyptian skulls apresenta medições feitas em crânios masculinos da área de Tebas no Egito.

- Há cinco amostras de 30 crânios cada uma dos períodos:
  - Pré-dinástico primitivo (cerca de 4000 a.C.);
  - Pré-dinástico antigo (cerca de 3300 a.C.);
  - 12ª e 13ª dinastias (cerca de 1850 a.C.);
  - Ptolemaico (cerca de 200 a.C.);
  - Romano (cerca de 150 d.C.).

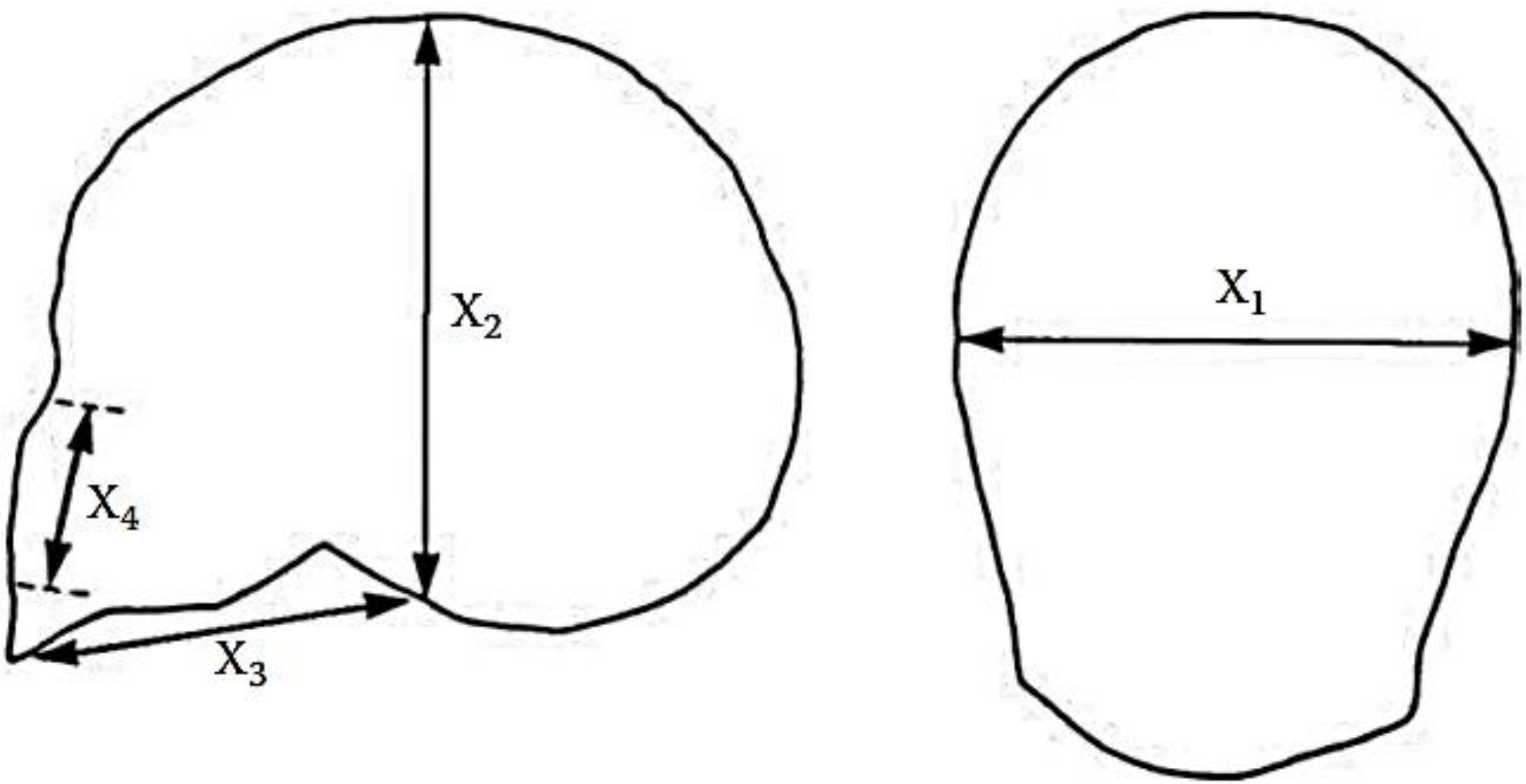
EXEMPLO ANOVA

- Quatro medidas são apresentadas para cada crânio.

- $X1$ = largura máxima;
- $X2$ = altura basibregmática;
- $X3$ = comprimento basialveolar;
- $X4$ = altura nasal.

EXEMPLO ANOVA

Figura 1: Quatro medições feitas em crânios masculinos egípcios.



EXEMPLO ANOVA

- Hipótese a ser testada
  - Hipótese nula: a média da largura máxima dos crânios de todos os períodos são iguais, ou seja, o período não tem efeito sobre a largura máxima do crânio
  - Hipótese alternativa: Pelo menos uma média é diferente, isto é, existe efeito de período.

EXEMPLO ANOVA

Tabela 1. Quadro da ANOVA de um fator.

Causas de variação	Graus de liberdade	Soma de Quadrados	Quadrados médios	F
Fatores	$k-1$	SQ_{Trat}	QM_{Trat}	$F = \frac{QM_{Trat}}{QM_{Res}}$
Resíduo	$k(r-1)$	SQ_{Res}	QM_{Res}	
Total	$\boldsymbol{kr-1}$	$\boldsymbol{SQ_{Total}}$		

- $SQ_{Total} = \sum y^2 - C$
- $SQ_{Res} = SQ_{Total} - SQ_{Trat}$
- $SQ_{Trat} = \frac{\sum T^2}{r} - C$
- $C = \frac{(\sum y)^2}{n}$

EXEMPLO ANOVA

- Os QM's são obtidos dividindo as somas de quadrados pelos respectivos graus de liberdade.

- Para testar as hipóteses utiliza-se a estatística F com $(k - 1)$ graus de liberdade no numerador e $k(r - 1)$ graus de liberdade no denominador.

- Se $F_{calc} > F_{tab}$, rejeita-se H_0. Neste caso dizemos que existem diferenças estatisticamente significativas entre as médias.

EXEMPLO ANOVA

- Um procedimento de teste equivalente usa a probabilidade de significância (p-valor).

- O p-valor representa a probabilidade de ser obtida uma observação da $F_{tab} \geq F_{cal}$.

- Se o p-valor for menor que α, rejeitamos H_0.

- Exemplo no R.

MANOVA

- Quando há mais de uma variável resposta, o método geral de análise é a MANOVA, análise de variância multivariada, em que vetores de médias são comparados

- Ou seja: as variáveis resposta são analisadas simultaneamente

- Leva-se em consideração as relações existentes entre as variáveis.

MANOVA

Tabela 2. Quadro da MANOVA de um fator.

Fonte de Variação	SQPC	gl
Dentro de Amostras/grupos/tratamentos	W	$m-1$
Entre amostras ou Resíduo	B	$n-m$
Total	$\boldsymbol{T}$	$\boldsymbol{n-1}$

MANOVA

- A hipótese a ser verificada é:

$$H_0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_I$$

- Ou seja, a hipótese é que não há diferenças entre os verdadeiros vetores de médias de tratamentos.

- A hipótese alternativa é de que ao menos um dos vetores de médias é diferente dos demais.

ESTATÍSTICAS DE TESTES

Tabela 3. Estatísticas de testes usadas para comparar vetores de médias amostrais com testes F aproximados na MANOVA

Teste	Estatística	F	gl_1	gl_2	Comentários
Lambda de Wilks	Λ	$\frac{(1-\Lambda^{1/t})}{\Lambda^{1/t}}(\frac{gl_2}{gl_1})$	$p(m-1)$	$wt-\left(\frac{gl_1}{2}\right)+1$	$w = n - 1\{(p+m)/2\}$ $t = [(gl_1{}^2 - 4)/\{p^2 + m - 1)2 - 5\}]1/2$ Se $gl_1 = 2$, faça $t = 1$
Maior raiz de Roy	λ_1	$\left(\frac{gl_2}{gl_1}\right)\lambda_1$	D	$n-m-d-1$	O nível de significância obtido é um limite inferior $d = \max(p, m-1)$
Traço de Pillai	$V = \sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i}$	$\frac{(n-m-p+s)V}{d(s-V)}$	sd	$s(n-m-p+s)$	$s = \min(p, m-1)$ $=$ n° de λ_i ‘s posit.
Traço de Lawley-Hotelling	$U = \sum_{i=1}^{p} \lambda_i$	$\frac{gl_2 U}{s\, gl_1}$	$s(2A+s+1)$	$2(sB+1)$	$A = (\|m-p-1)/2$ $B = (n-m-p-1)/2$

Nota: Assume-se que há p variáveis em m amostras, com a j-ésima de tamanho n_j e $n = \Sigma n_j$

MANOVA

- No caso multivariado, quatro estatísticas de teste são comumente utilizadas:
    - Lambda de Wilks
    - Maior raiz de Roy
    - Traço de Pillai
    - Traço de Hotelling-Lawley
- É esperado que elas apresentem valores-p semelhantes, de modo que não há real necessidade de se escolher uma delas.

PRESSUPOSIÇÕES

- Normalidade multivariada;

- Igualdade de matrizes de variância e covariância.

MANOVA

- Exemplo no R