ANÁLISE DE VARIÂNCIA (*ANOVA*)

Prof. Anderson Rodrigo da Silva
anderson.silva@ifgoiano.edu.br

Exemplo 1 de Introdução

Medley & Clements (1998) estudaram o efeito de metais pesados, especialmente zinco, sobre a riqueza e diversidade de algas diatomáceas em rios do Colorado (EUA).

34 estações de coleta foram divididas de acordo com a concentração de metais pesados, em:

- *Background* (20 g /L), 8 estações
- *Low* (21–50 g /L), 8 estações
- *Medium* (51–200 g /L), 9 estações
- *High* (200 g /L), 9 estações

Ver detalhes no livro de Quinn & Keough (p.173)

Exemplo 1 de Introdução

A seguinte hipótese foi levantada: não há diferença entre os 4 grupos de níveis de zinco em relação a diversidade média de espécies.

	Nível de Zn			
Estação	***B***	***L***	***M***	***H***
1	2.27	1.4	1.62	1.25
2	1.7	2.18	2.19	1.15
3	2.05	1.83	2.1	0.63
4	1.98	1.88	2.06	1.04
5	2.2	2.1	2.02	1.9
6	1.53	2.38	1.94	1.88
7	0.76	2.83	1.75	0.85
8	1.89	1.66	0.8	1.43
9	--	--	0.98	1.37
Média	***1.80***	***2.03***	***1.72***	***1.28***

Exemplo 2 de Introdução

Sejam dados de produtividade (t/ha) de três variedades cana-de-açúcar (V1, V2e V3), cada uma plantada em duas parcelas ou unidades experimentais de $24m^2$.

V2 (70 t/ha)	**V3** (72 t/ha)	**V3** (70 t/ha)
V1 (120 t/ha)	**V2** (80 t/ha)	**V1** (125 t/ha)

Exemplo 2 de Introdução

Quanto da variabilidade total dos dados é devido ao fator *variedade* e quanto é devido ao “acaso”?

Análise de Variância - ANOVA

- Decomposição da variação total em causas conhecidas e desconhecidas

- Hipótese em teste: a variação devida a tratamentos é igual a variação devida ao acaso ou erro experimental ou resíduo

- Em particular, a variabilidade é medida em termos de **somas de quadrados**!

Obtenção das somas de quadrados

$$SQtotal = SQtrat + SQerro$$

Obs.: SQtrat pode ser calculada a partir dos totais de tratamentos.

Tabela de ANOVA - exemplo

FV	GL	SQ	QM	F
Tratamentos	2			
Resíduo	3			--
Total	5		--	--

Teste de Hipóteses: F

Passos:

1) Hipóteses
2) Nível de significância
3) Região crítica
4) Estatística do teste
5) Conclusão

Regra Decisória para o Teste de Hipóteses

$$\text{Se} \begin{cases} \mathrm{F_{Calc}} \geq \mathrm{F_{Tab}},\ \text{rejeita} - \text{se H}_0 \text{ ao nivel } \alpha \\ \mathrm{F_{Calc}} < \mathrm{F_{Tab}},\ \text{não se rejeita H}_0 \text{ ao nivel } \alpha \end{cases}$$

Coeficiente de variação experimental (CV%)

$$CV\% = 100\frac{\sqrt{QMres}}{\bar{y}}$$

- Precisão experimental *e* fontes de erro experimental
  - Material experimental (vegetal ou animal)
  - Unidades experimentais
  - Erros grosseiros
- Classificação de Pimentel-Gomes*
  - Baixo: < 10%
  - Médio: 10 – 20%
  - Alto: 20 – 30%
  - Muito alto: >30%

Modelo Estatístico de Análise de Variância

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Em que:

Y_{ij} : valor observado da variável Y na j – ésima repetição do i – ésimo tratamento

μ : média geral do experimento

τ_i : efeito do i – ésimo tratamento

ε_{ij} : erro ou efeito da variável residual associado a observação Y_{ij}

Modelo Estatístico de Análise de Variância

Índices ou indexadores:

$$\text{de tratamento}: \quad i = 1, 2, \ldots, I$$

$$\text{de repetição}: \quad j = 1, 2, \ldots, J$$

Obs.: Num DIC balanceado, o número total (N) de observações pode ser obtido por: N=IJ

Modelo Estatístico de Análise de Variância

Estimadores dos parâmetros

- Média geral $\hat{\mu} = \bar{y}$

- Efeito do i-ésimo tratamento $\hat{\tau}_i = \bar{y}_i - \bar{y}$

Propriedades:

$$\sum_{i=1}^{I} \hat{\tau}_i = 0 \qquad \hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_i \qquad \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{\varepsilon}_{ij} = 0$$

Flash Exercise

- Calcule os erros ou resíduos associados a cada parcela do exemplo
- Estime os efeitos de tratamento (variedade) em cada parcela
- Mostre que a soma destes é nula

Exigências ou pressuposições do modelo:

- Independência dos erros ou resíduos

(Teste de Durbin-Watson)

- Normalidade dos erros

(Teste de Shapiro-Wilk, Lilliefors...)

- Homocedasticidade ou homogeneidade de variâncias de tratamentos

(Teste de Bartlett, do F Máximo de Hartley, de Cochran, etc.)

Teste de Cochran

H_0: As variâncias dos I tratamentos são iguais
H_1: Ao menos uma das variâncias difere de outra(s)

$$C = \frac{maior\, S_i^2}{\sum_{i=1}^{I} S_i^2}$$

Se $C \geq C_{Tab}$, rejeita-se H_0.

Tabela de Cochran

O que fazer se os dados não atenderem às exigências do modelo de ANOVA?

Transformação Box-Cox

$$y^* = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

Análise não paramétrica: teste de Kruskal-Wallis

EXERCÍCIOS

#1 Teste a hipótese apresentada no Exemplo 1

#2 O resultado das vendas efetuadas por 3 vendedores de uma indústria de inseticidas durante certo período é dado a seguir:

	Vendedor		
	A	***B***	***C***
	29	27	30
	27	27	30
	31	30	31
	29	28	27
	32		29
	30		
Totais	*178*	*112*	*147*

a) Faça a ANOVA e avalie a hipótese de que os vendedores vendem em média a mesma quantidade de inseticida.

b) Calcule e interprete o coef. de variação experimental

Leitura recomendada

- Capítulo 8 do livro Quinn & Keough
  - Seções: 8.1 (p. 173 - 175), 8.1.3, 8.1.4, 8.3, 8.4, 8.6.2