Metodologia Ecológica
Abordagens e Métodos Estatísticos em Ecologia

Análise de Variância com um modelo linear

- Soluções: ANOVA através de delineamento de regressão linear
- Matriz de delineamento (*Design matrix*)
- Interpretação dos coeficientes e resultados
- Como calcular valores esperados
- Testes *a posteriori* e problemas de múltiplos testes
- Análise de Covariancia - ANCOVA

Prof. Marcus Vinícius Vieira - Instituto de Biologia UFRJ

ANOVA como um modelo linear

$$y_i = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + e_i$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

$$[y] = [\beta] * [X] + [e]$$

Códigos para matriz de delineamento (*design matrix)*

- Codifica um fator ou variável categórica
- No exemplo, um fator com quatro níveis
- A média de um nível é o intercepto (β_0)
- Os demais são $\beta_0 + \beta_i$

Matriz de delineamento

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Códigos para matriz de delineamento (*design matrix)*

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- Quatro níveis: A, B, C, D
- Nível A = β_0 (intercepto)
- Nível B = $\beta_0 + \beta_B$ = $\beta_0 + \beta_B * \mathbf{1} + \beta_C * 0 + \beta_D * 0$
- Nível C = $\beta_0 + \beta_C$ = $\beta_0 + \beta_B * 0 + \beta_C * \mathbf{1} + \beta_D * 0$
- Nível D = $\beta_0 + \beta_D$ = $\beta_0 + \beta_B * 0 + \beta_C * 0 + \beta_D * \mathbf{1}$

TABLE 10.2 ANOVA table for one-way layout

Source	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	Expected mean square	F-ratio	P-value
Among groups	$a-1$	$\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{Y}_i-\bar{Y})^2$	$\frac{SS_{among\ groups}}{(a-1)}$	$\sigma^2 + n\sigma_A^2$	$\frac{MS_{among\ groups}}{MS_{within\ groups}}$	Tail of the F-distribution with $(a-1)$ $a(n-1)$ degrees of freedom
Within groups (residual)	$a(n-1)$	$\sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_i)^2$	$\frac{SS_{within\ groups}}{a(n-1)}$	σ^2		
Total	$an-1$	$\sum_{i=1}^{a}\sum_{i=1}^{n}(Y_{ij}-\bar{Y})^2$	$\frac{SS_{total}}{(an-1)}$	σ_Y^2		

TABLE 10.3 One-way ANOVA table for the hypothetical data in Table 10.1

Source	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	F-ratio	P-value
Among groups	2	22.17	11.08	5.11	0.033
Within groups (residual)	9	19.50	2.17		
Total	11	41.67			

A posteriori:
Diferença honestamente significante de Tukey (*Tukey honestly significant difference – HSD*):

ajusta o valor de *P* para cada teste individual de forma a manter α = 0,05

$$HSD = q\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) QM_{res}}$$

q = valor obtido na tabela de t de Student

n_i e n_j = tamanho da amostra i e j respectivamente

QM_{res} = quadrado médio dos resíduos (soma dos desvios ao quadrado div. pelos gl)

4 tratamentos: não manipulado, isopor, *Tedania ignis*, *Haliclona implexiforms*

14 réplicas por tratatamento x 3 tratamentos = 56 (tamanho amostral)

	Graus de Liberdade	Soma dos Quadrados	Quadrado Médio	F	*P*
Tratamentos	3	2,602	0,867	5,286	0,003
Resíduo	52	8,551	0,164		
Total	52	11,153			

Para a tabela anterior:

q = 3,42

$n_i = n_j$ = 14

QM_{res} = 0,164

HSD = 0,523

Um par de médias para ser considerado estatísticamente significativo deve diferir de pelo menos 0,523

TABLE 10.13 Mean differences between all possible pairs of treatments in an experimental study of effects of sponges on root growth of red mangrove

	Unmanipulated	Foam	*Tedania*	*Haliclona*
Unmanipulated	0.000			
Foam	0.383 (0.072)	0.000		
Tedania	0.536 (0.031)	0.053 (0.986)	0.000	
Haliclona	0.584 (0.002)	0.202 (0.557)	0.149 (0.766)	0.000

$$HSD = q\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)QM_{res}} \qquad q = \frac{HSD}{\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)QM_{res}}}$$

Correções de Bonferroni e testes múltiplos

O problema: nível de significancia de 5% quer dizer que, por exemplo, em 100 vezes que o teste for repetido, em 5 delas o valor de *F* (ou outra estatística) pode ser maior ou igual o observado.

Se o teste for repetido 20 vezes, em 1

Correção de Bonferroni simples

Correção de Dunn-Sidak

Problemas

Assume que os testes são independentes

Assume que todas as hipóteses nulas são verdadeiras

Não há mais um valor padrão de 0,05 e o espaço amostral é arbitrário: qual o universo de testes repetidos? Todos os artigos que o autor já publicou sobre a hipótese? Que todos os autores publicaram?

Se os testes são independentes, qual a chance de digamos em 3 testes obter P = 0,11 quando a hipótese nula é verdadeira?

Bonferroni: 0,0167   Dunn-Sidak: 0,0169

Teste de probabilidades combindas de Fisher: CP = 13,24, 6 df, P = 0,039

Comparações *a priori*

Vantagem: pode-se definir comparações que sejam ortogonais umas às outras: **contrastes ortogonais**

Sendo **ortogonais**, são independentes um do outro: soma dos quadrados entre grupos pode ser repartida

Se não forem ortogonais, a soma dos quadrados entre grupos não pode ser repartida

Maior poder estatístico (maior capacidade de detectar de rejeitar hipótese nula quando é falsa)

ANCOVA

- Modelo linear
- Objetivos:
  - Remover de *Y* a variação devido a uma covariável *X*
  - Ajustar a média de *Y* para um valor padrão da variável *X* (desta forma “fixando” o valor de X)
- Como fazer isso? Como medir a variação comum de *X* e *Y* e descontá-la de *Y*?

Modelo de ANCOVA

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

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

$$Y_{ij} = \mu + A_i + B_i\left(X_{ij} - \bar{X}_{ij}\right) + \varepsilon_{ij}$$

$$Y_{ij} = \mu + A_i + B_C\left(X_{ij} - \bar{X}_{ij}\right) + \varepsilon_{ij}$$

Interações em ANCOVA