Análise de Variância

Biologia Quantitativa
Departamento de Zoologia – UnB
Módulo 03

11 de abril de 2023

Objetivo

- Determinar se amostras são significativamente diferentes em desenhos experimentais ou observacionais com múltiplos fatores e interações
- A variável dependente é contínua
- As variáveis independentes formam classes
- O método foi desenvolvido por Fisher e é baseado na partição das variâncias dentro e entre grupos usando mínimos quadrados

Objetivo

- Hoje em dia a ANOVA é também utilizada como ferramenta para ajuste de modelos. A proporção da variância explicada informa a qualidade do ajuste.
- É um modelo onde as variáveis independentes são discretas (categorias). Se as variáveis independentes são contínuas lidamos com regressão.

Modelos Lineares

- Modelos Lineares Gerais: a variável dependente segue uma distribuição normal
- Modelos Lineares Generalizados: a variável dependente pode ter outros tipos de distribuição: regressão logística variável dependente é categórica, regressão poisson variável dependente é contagem ou frequência.
- No modelo linearizado a variável dependente também é transformada - log, etc.

Premissas - 1

- Amostras aleatórias e independentes
- Dados distribuídos normalmente
- Variâncias homogêneas
- Fatores aditivos
-
- Como testar?
- Normalidade: Shapiro-Wilk, KS, e visualizar histogramas e box-plots
- Homoscedascidade: Teste F (2 amostras) ou Levene, Bartlett (múltiplas)

Premissas - 2

- Amostras aleatórias e independentes
- Dados distribuídos normalmente
- Variâncias homogêneas
- Fatores aditivos
-
- Como corrigir? Transformar dados **ou:**
- Normalidade – se for balanceado e as distribuições forem simétricas – tolerar.
- Homoscedascidade – modelo balanceado aceitar até 3-4x razão maior/menor var.

Modelo da Anova

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\varepsilon}_{ij}$$

The individual's outcome

The average outcome over all individuals

The average outcome over all individuals in group i

The random noise that makes individual j unique

Fonte: https://rlbarter.github.io/Practical-Statistics/2017/02/20/anova/

Método

- Estimar a variância da população dentro dos grupos, assim como a variância entre grupos
- Se todas as amostras vem da mesma população, ambas as estimativas são aproximadamente iguais
- A distribuição F dá a estatística da razão entre variâncias.
- Para amostras repetidas n1 e n2, a razão das variâncias tende a (n2-1)/(n2-3)

Soma dos Quadrados e Variância dentro/entre grupos

$$\frac{1}{a(n-1)} \sum_{i=1}^{i=a} \sum_{j=1}^{j=n} (Y_{ij} - \bar{Y}_i)^2$$

$$\frac{n}{a-1} \sum_{i=1}^{i=a} (\bar{Y}_i - \bar{\bar{Y}})^2$$

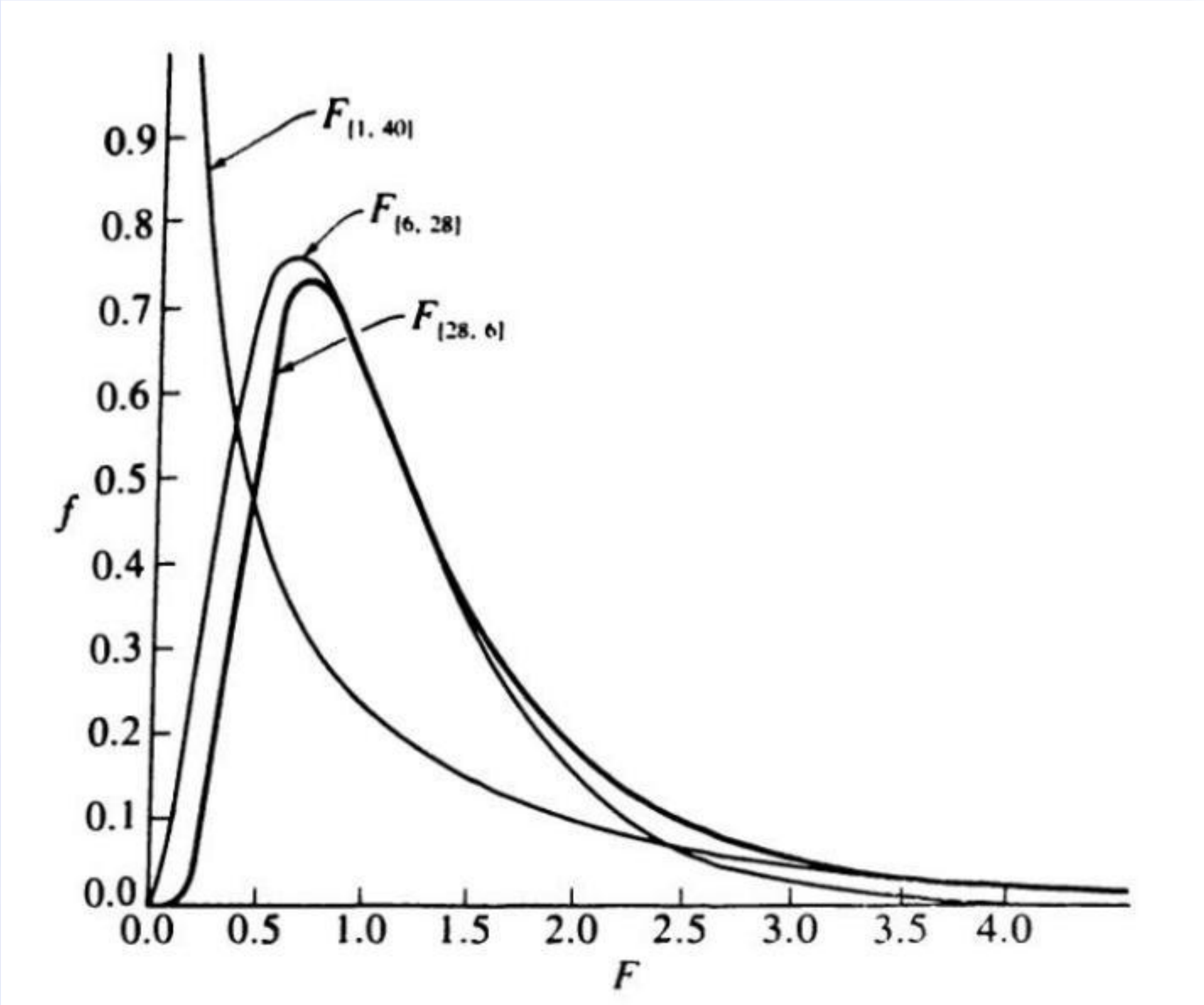
Cálculo

- Soma dos mínimos quadrados
- Média ponderada (por graus de liberdade) das variâncias dentro dos grupos
- Variância das médias dos grupos em relação à média das médias
- 2 estimativas independentes da mesma variância
- Hipótese nula: ambas as variâncias estimam a mesma variância paramétrica

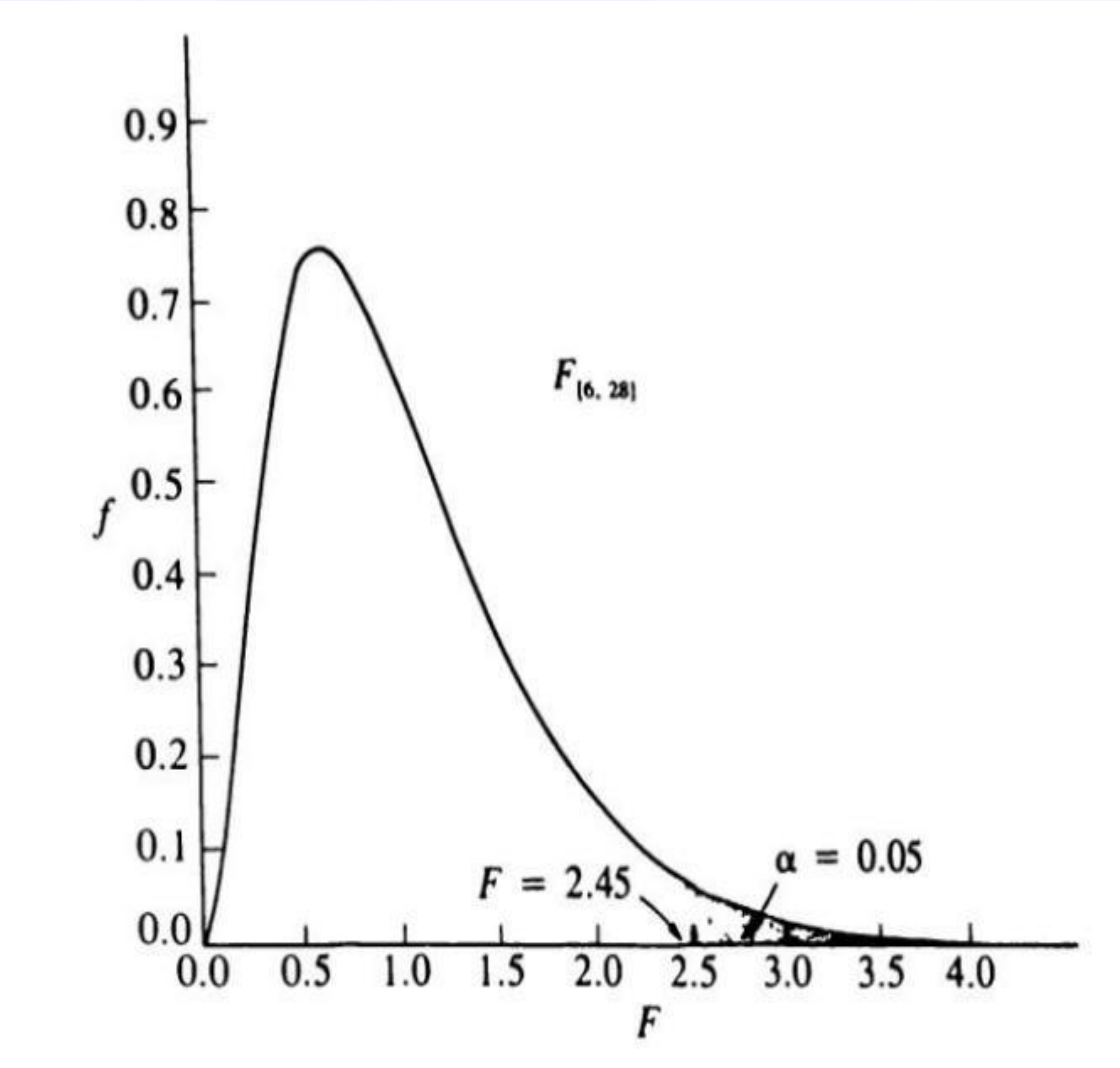
Método

- Quando n2 é grande F tende a 1
- A razão F depende de dois graus de liberdade
- A distribuição F pode ser obtida por amostras repetidas da mesma distribuição normal, ou por amostras de distribuições normais distintas com a mesma variância
- Refs: Sokal & Rohlf cap. 8

Distribuição F



Area de Rejeição Distrib F



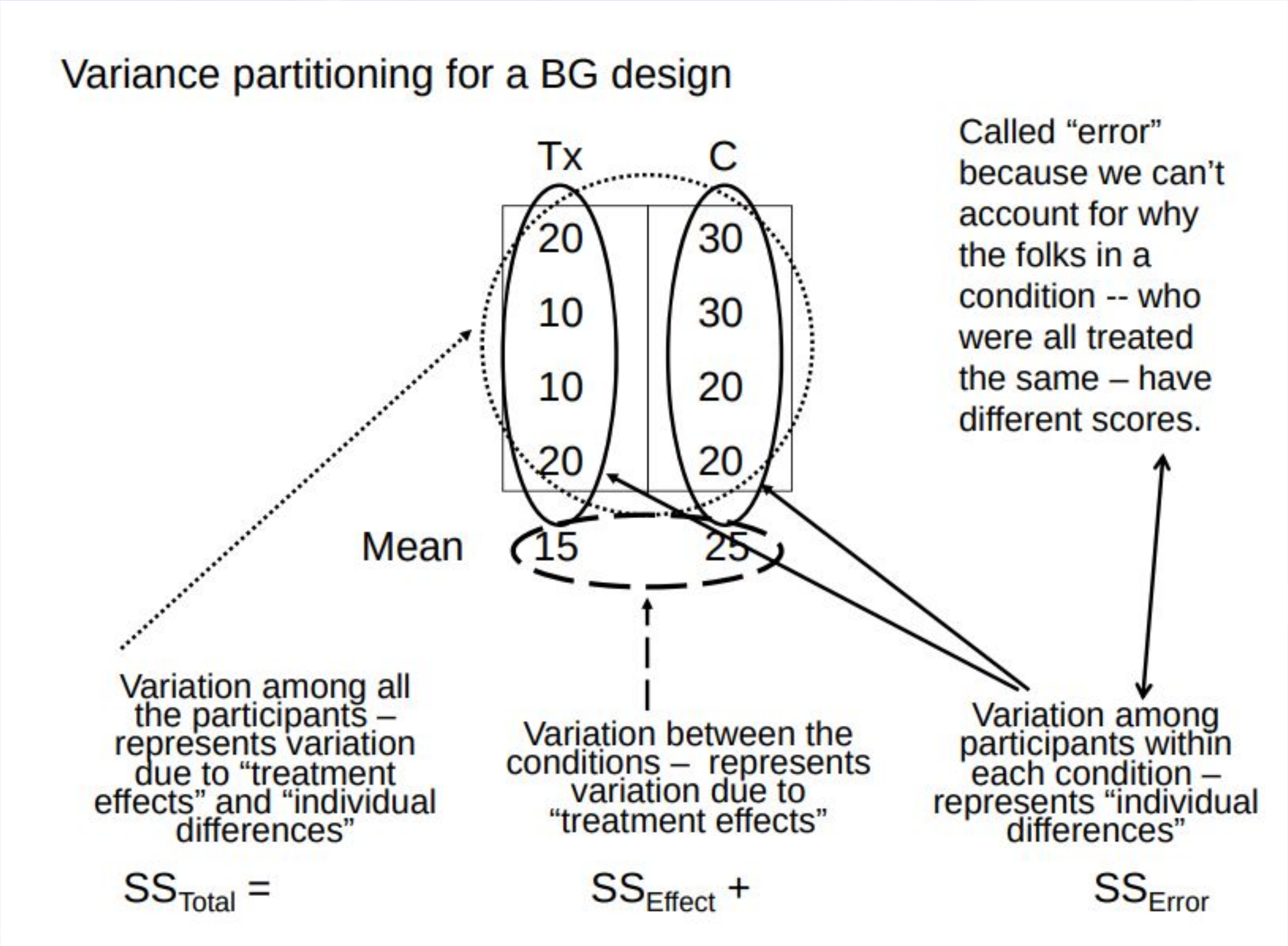
Anova com 1 fator – 1-way

-
- Calculam-se 3 variâncias no total:
- Dentro de grupos
- Entre grupos
- Total (dentro+entre grupos)

Exemplo
1 fator

	X Labels	A	B	C	D
	Animal No.	Control	Stim alone	+ Placebo	+ Antagonist
	X	Y	Y	Y	Y
1	W-1078	90	117	120	98
2	W-1079	88	116	118	90
3	W-1080	96	116	109	105
4	N-562	94	131	127	97
5	W-1129	91	123	127	99
6	N-588	97	132	129	108

Exemplo 1 fator



Anova com 2 fatores – 2-way

- Calculam-se 3 variâncias entre grupos:
- Entre linhas vs a média geral
- Entre colunas vs a média geral
- Interações
- Sokal Cap 11

Exemplo
2 fatores

	Machine 1	*Machine 2*	*Machine 3*
Operator A	16, 13, 19	9, 15, 11	22, 25, 17
Operator B	18, 17, 21	15, 13, 12	14, 16, 12
Operator C	14, 16, 13	7, 12, 9	11, 14, 12
Operator D	13, 14, 16	3, 1, 9	13, 17, 14

Exemplo
2 fatores

A General Two-Way ANOVA Table

Source of Variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS	Value of F Ratio
Between samples (due to factor A) Differences between means $\bar{X}_{i.}$ and $\bar{\bar{X}}$	$SS_A = b\sum_{i=1}^{a}\left(\bar{X}_{i.} - \bar{\bar{X}}\right)^2$	$(a-1)$	$MS_A = \frac{SS_A}{(a-1)}$	$F = \frac{MS_A}{MS_E}$
Between samples (due to factor B) Differences between means $\bar{X}_{.j}$ and $\bar{\bar{X}}$	$SS_B = a\sum_{j=1}^{b}\left(\bar{X}_{.j} - \bar{\bar{X}}\right)^2$	$(b-1)$	$MS_B = \frac{SS_B}{(b-1)}$	$F = \frac{MS_B}{MS_E}$
Within samples (due to chance errors) Differences between individual observations and fitted values.	$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{\bar{X}}\right)^2$	$(a-1)$ $\times(b-1)$	$MS_E = \frac{SS_E}{(a-1)(b-1)}$	
Totals	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \bar{\bar{X}}\right)^2$	$(ab-1)$		

Exemplo
2 fatores

Hence the two-way ANOVA table for the example under consideration is

Source of Variation	Sum of Squares SS	Degrees of Freedom	Mean Square MS	Value of F Ratio
Between samples (*due to factor A*) Differences between means $\bar{X}_{i.}$ and $\bar{\bar{X}}$	24	4	$\frac{24}{4} = 6$	$F = \frac{6}{13} = 0.46$
Between samples (*due to factor B*) Differences between means $\bar{X}_{.j}$ and $\bar{\bar{X}}$ height	150	3	$\frac{150}{3} = 50$	$F = \frac{50}{13} = 3.85$
Within samples (*due to chance errors*) Differences between individual observations and fitted values.	156	12	$\frac{156}{12} = 13$	
TOTALS	330	19		

Tipos de Anova

- Modelo 1: efeitos pré-definidos
- Modelo 2: efeitos aleatórios ou não conhecidos
- Ambos os modelos: soma dos efeitos = 0 em relação à média global
- Objetivo: estimar efeitos e particionar a soma dos quadrados

Hipóteses

- Hipótese nula: variância entre grupos e dentro de grupos são estimativas da mesma distribuição
- H1: variância entre grupos é maior do que a variância dentro dos grupos
- Relação com teste de t – caso de distribuição com 1 gl no numerador
- Raiz quadrada de F = valor de t

Aplicações

- Análises com múltiplos grupos e múltiplos fatores
- Modelos lineares em combinação com outros métodos como regressão (análise de covariância)
- Quadro lógico para partição de variâncias e estimativas de erros