

Análise de Variância

Biologia Quantitativa
Departamento de Zoologia – UnB

23 de fevereiro de 2021

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.

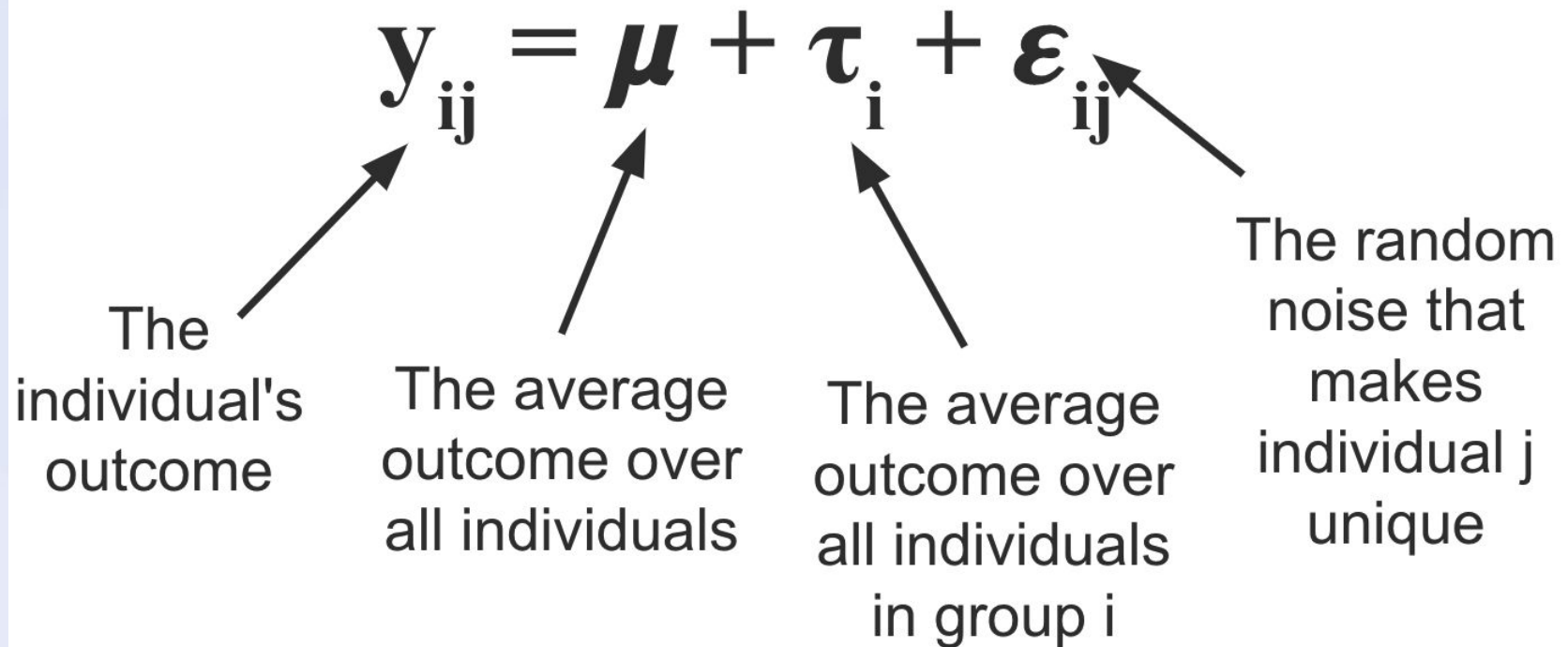
Premissas - 1

- Amostras aleatórias e independentes
- Dados distribuídos normalmente
- Variâncias homogêneas
- Fatores aditivos
-
- Como testar?
- Normalidade: Shapiro-Wilk, KS, e visualizar histogramas e box-plots
- Homoscedasticidade: 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.
- Homoscedasticidade – modelo balanceado aceitar até 3-4x razão maior/menor var.

Modelo da Anova



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 n_1 e n_2 , a razão das variâncias tende a $(n_2-1)/(n_2-3)$

Soma dos Quadrados e Variância dentro/entre grupos

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

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

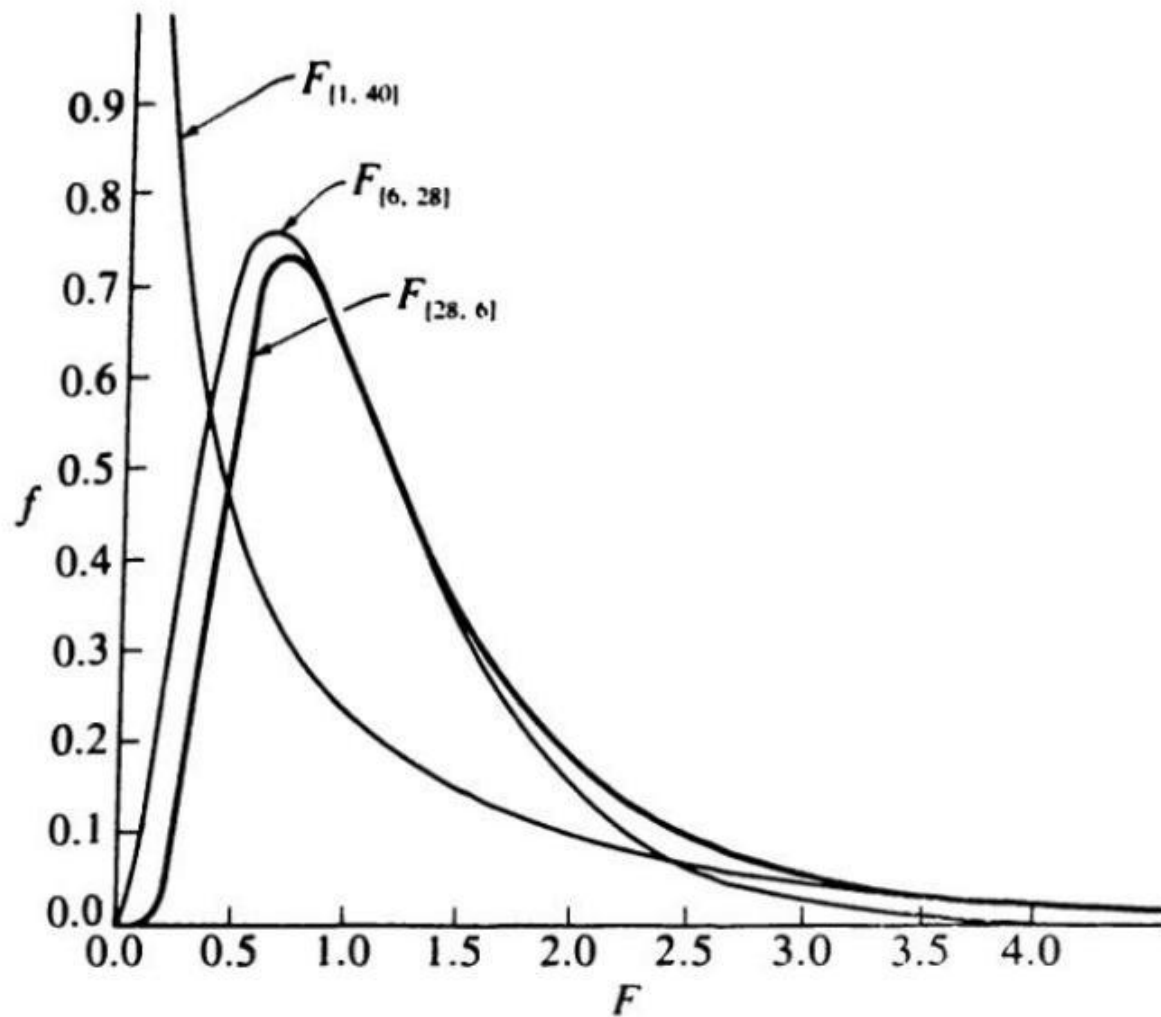
Cálculo

- 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

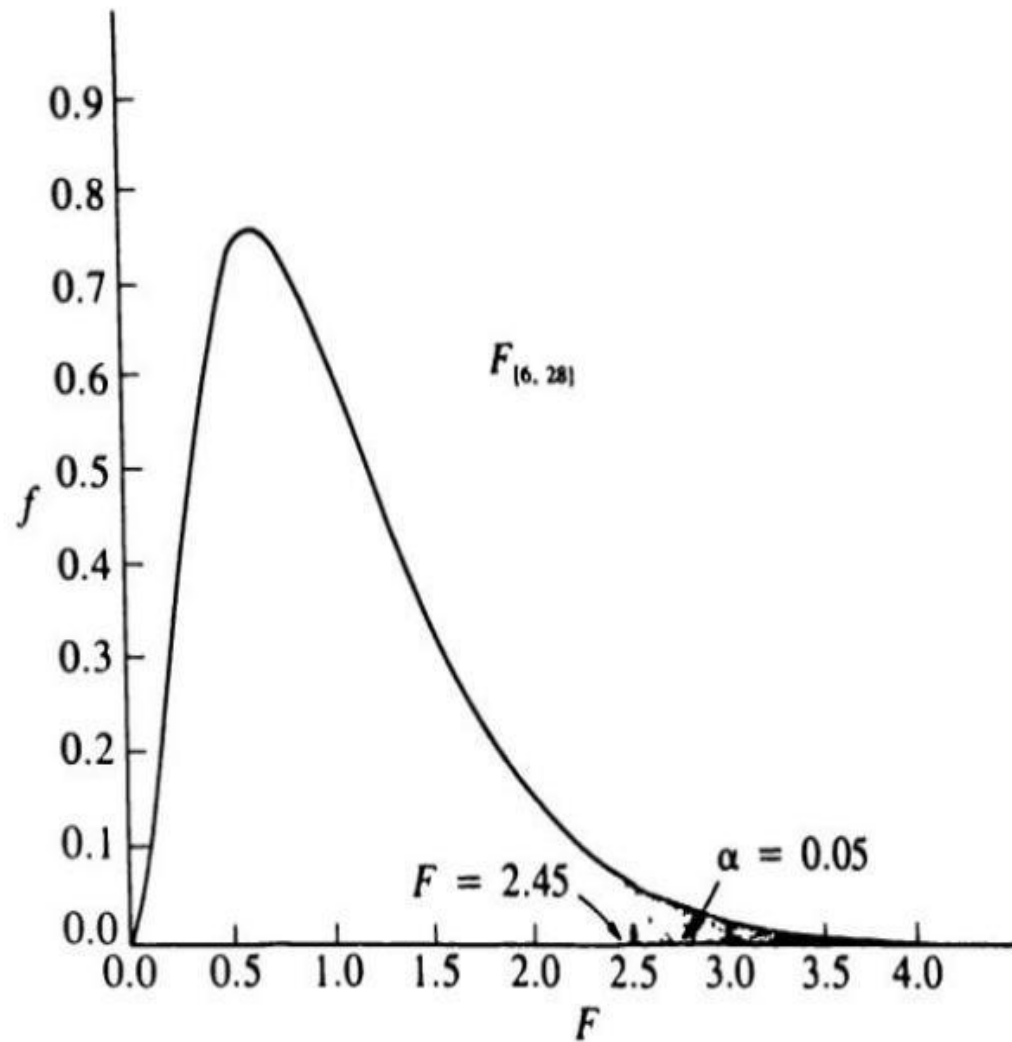
Método

- Quando n_2 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 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

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

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 $F = \text{valor de } t$

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