Capítulo 9

Análise de Variância

Texto baseado no livro:

Estatística Aplicada - Larson / Farber – Editora Pearson – 2010

Neste Capítulo, o procedimento de teste de hipóteses será utilizado para a comparar as médias de mais de duas populações.

Embora o nome não mostre o objetivo real do procedimento, a análise da variância ou ANOVA é um teste de hipóteses de médias de duas ou mais populações

É um procedimento muito útil para comparar, por exemplo:

A eficiência de diversas marcas de remédios para o tratamento de uma mesma doença, o controle de pressão alta.

O consumo em km/litro de um modelo de carro abastecido com combustíveis do mesmo tipo, porém de marcas diferentes.

A eficiência de uma lavoura tratada com diferentes fertilizantes.

O tempo de reação de uma pessoa em função de estímulo de luz de quatro cores diferentes.

Etc ...

A classificação dos testes de análise da variância é de acordo ao número de fatores de interesse ou que influem na variável dependente.

- Por exemplo, na verificação da eficiência do crescimento de uma lavoura tratada com quatro tipos de fertilizantes, cada um dos fertilizantes é um fator.

- Da mesma maneira, na comparação do consumo de carros abastecidos com o mesmo tipo de combustível, porém de três marcas diferentes, cada marca de combustível é um fator.

Por que é denominada análise da variância um procedimento que compara médias de grupos diferentes?

Porque na preparação das variabilidades *entre* e *dentro* são utilizados os quadrados dos desvios dos valores das amostras, que fazem parte da definição da variância

Em um teste ANOVA unidirecional, as seguintes condições devem ser preenchidas:

1. Cada amostra deve ser aleatoriamente selecionada de uma populaçã normal ou aproximadamente normal.

2. As amostras devem ser independentes entre si.

3. Cada população deve ter a mesma variância.

Análise de variância com um fator

Hipóteses:

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$.

H_a: Pelo menos uma das médias é diferente das outras.

A distribuição F conduzirá a decisão de aceitar o rejeitar a hipótese nula, comparando o F observado F_0 calculado com a expressão: MS_B

$$F_0 = \frac{Variância\ entre\ as\ amostras}{Variância\ dentro\ das\ amostras} = \frac{MS_B}{MSw}$$

com o F crítico F_c (obtido na tabela F) correspondente ao nível de significância α adotado.

1. A variância entre as amostras MS_B medem a diferença relacionada ao tratamento dado à cada amostra e é às vezes chamada de **média entre quadrados**.

2. A variância dentro das amostras MS_W medem as diferenças relacionadas aos lançamentos dentro da mesma amostra. Esta variância, às vezes chamada de **média dos quadrados internos**, geralmente se dá por causa do erro de amostragem.

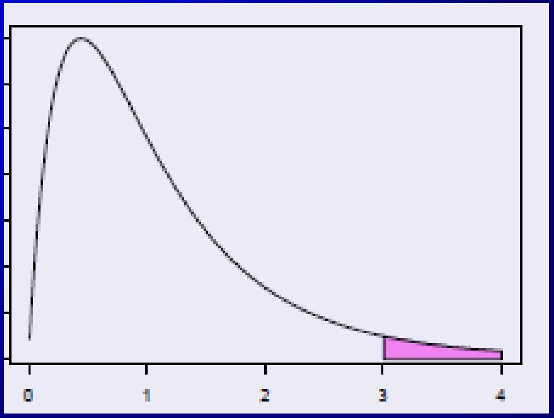
Se as condições para uma análise de variância de um fator são satisfeitas, então a distribuição amostral para o teste é aproximada pela distribuição *F de Fisher* .

O teste estatístico é: $$F_0 = \frac{MS_B}{MS_W}$$

Os graus de liberdade para o teste *F* são:

$$gl_N = k - 1 \quad \text{e} \quad gl_D = N - k.$$

Onde k é o número de amostras e N é a soma dos tamanhos das amostras.



$$\boxed{\text{Rejeita-se } H_0\,,\ \ se \quad F_0 > F_c}$$

Teste estatístico para um teste ANOVA de um fator

Em palavras	*Em símbolos*
1. Encontre a média e a variância de cada amostra.	$\bar{x} = \dfrac{\sum x}{n}$ e $s^2 = \dfrac{\sum (x - \bar{x})^2}{n - 1}$
2. Encontre a média de todas as entradas em todas as amostras (a grande média).	$\bar{\bar{x}} = \dfrac{\sum x}{N}$
3. Encontre a soma dos quadrados entre as amostras.	$SS_B = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$
4. Encontre a soma dos quadrados dentro das amostras.	$SS_W = \sum (n_i - 1) s_i^2$

Em palavras	*Em símbolos*
5. Encontre a variância entre as amostras.	$MS_B = \frac{SS_B}{k-1}$
6. Encontre a variância dentro das amostras.	$MS_W = \frac{SS_W}{N-k}$
7. Encontre o teste estatístico.	$F_0 = \frac{MS_B}{MS_W}$

Análise simples do teste de variância de um fator ANOVA

Em palavras	*Em símbolos*
1. Identifique a afirmação. Expresse as hipóteses nula e alternativa.	Expresse H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Identifique os graus de liberdade.	$gl_N = k - 1$
4. Determine o valor crítico.	$gl_D = N - k$

Em palavras	*Em símbolos*
5. Determine a região de rejeição.	
6. Calcule o teste estatístico.	$F_0 = \dfrac{MS_B}{MS_W}$
7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula.	Se F est na região de rejeição, rejeite H_0 $(F_o > F_c)$
8. Interprete a decisão no contexto da afirmação original.	Se não, falhe em rejeitar H_0.

Tabela de resumo ANOVA

Uma tabela é um modo conveniente de resumir os resultados de um teste com um fator ANOVA.

Variação	Soma dos quadrados	Graus de liberdade	Médias quadradas	*F*
Entre	SS_B	gl_N	$MS_B = \frac{SS_B}{d.f._N}$	$\frac{MS_B}{MS_W}$
Interna	SS_W	$gl._D$	$MS_W = \frac{SS_W}{d.f._D}$	

Exemplo: teste ANOVA simples

Um médico pesquisador quer determinar se há uma diferença na média de tempo que três tipos de analgésicos levam para aliviar a dor de cabeça. Várias pessoas que sofrem com dores de cabeça são selecionadas aleatoriamente e tomam um dos três medicamentos. Cada pessoa diz o tempo (em minutos) que o medicamento começou a fazer efeito. Os resultados são mostrados no próximo slide. Com $\alpha = 0{,}01$, você pode concluir que as médias de tempo são diferentes? Suponha que cada população de tempo de alívio seja normalmente distribuída e que a população de variâncias seja igual.

Medicação 1	Medicação 2	Medicação 3
12	16	14
15	14	17
17	21	20
12	15	15
	19	
$\bar{x}_1 = \frac{56}{4} = 14$	$\bar{x}_2 = \frac{85}{5} = 17$	$\bar{x}_3 = \frac{66}{4} = 16.5$
$s_1^2 = 6$	$s_2^2 = 8.5$	$s_3^2 = 7$

Solução:

$k = 3$ (3 amostras)

$N = n_1 + n_2 + n_3 = 4 + 5 + 4 = 13$ (soma dos tamanhos das amostras)

Para encontrar o teste estatístico, o seguinte deve ser calculado:

$$\bar{\bar{x}} = \frac{\sum x}{N} = \frac{56 + 85 + 66}{13} \approx 15.92$$

$$MS_B = \frac{SS_B}{\mathrm{d.f.}_\mathrm{N}} = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k - 1}$$

$$= \frac{4(14 - 15.92)^2 + 5(17 - 15.92)^2 + 4(16.5 - 15.92)^2}{3 - 1}$$

$$\approx \frac{21.92}{2} = 10.96$$

Para encontrar o teste estatístico, o seguinte deve ser calculado:

$$MS_W = \frac{SS_W}{\text{d.f.}_\text{D}} = \frac{\sum(n_i - 1)s_i^2}{N - k}$$

$$= \frac{(4-1)(6) + \ (5-1)(8.5) + (4-1)(7)}{13-3}$$

$$= \frac{73}{10} = 7.3$$

$$F_0 = \frac{MS_B}{MS_W} = \frac{10.96}{7.3} \approx 1.50$$

Solução: teste ANOVA simples

H_0: $\mu_1 = \mu_2 = \mu_3$

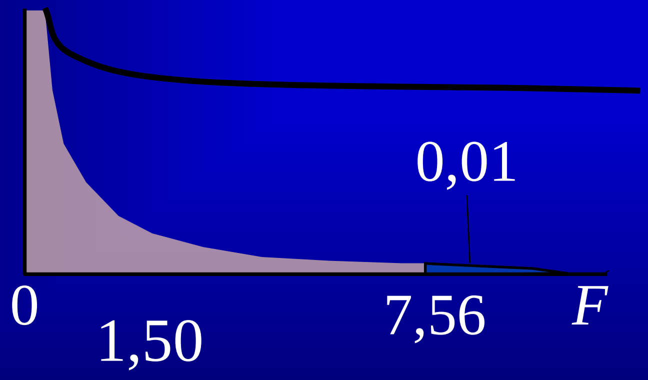
H_a: **Pelo menos uma média é diferente**

α = **0,01**

gl$_N$**=** **3 – 1 = 2**

gl$_D$**=** **13 – 3 = 10**

Região de rejeição:



Teste estatístico:

$$F = \frac{MS_B}{MS_W} \approx 1.50$$

Decisão:

Falhar em rejeitar H_0

Não há evidência suficiente no nível de significância 1% para concluir que há uma diferença no tempo médio que os três analgésicos levam para aliviar as dores de cabeça.

Anova: fator único

RESUMO

Grupo	*Contagem*	*Soma*	*Média*	*Variância*
Coluna 1	4	56	14	6
Coluna 2	5	85	17	8,5
Coluna 3	4	66	16,5	7

ANOVA

Fonte da variação	*SQ*	*gl*	*MQ*	*F*	*valor-P*	*F crítico*
Entre grupos	21,92308	2	10,96154	1,501581	0,269002	7,559432
Dentro dos grupos	73	10	7,3			
Total	94,92308	12				

Resultados do EXCEL

Exemplo: usando a tecnologia para realizar um teste de um fator ANOVA

Três empresas de aviação oferecem vôos entre duas cidades. Vários tempos de vôo selecionados aleatoriamente (em minutos) entre as cidades para cada empresa podem ser observados no próximo slide. Suponha que as populações de tempo de vôo sejam normalmente distribuídas, as amostras sejam independentes e as variâncias populacionais sejam iguais. Com $\alpha = 0{,}01$, você pode concluir que há uma diferença nas médias de tempo dos vôos?

Empresa 1	Empresa 2	Empresa 3
122	119	120
135	133	158
126	143	155
131	149	126
125	114	147
116	124	164
120	126	134
108	131	151
142	140	131
113	136	141

Solução: usando o excel para realizar um teste de um fator ANOVA

H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: **Pelo menos uma média é diferente**

Anova: fator único						
RESUMO						
Grupo	*Contagem*	*Soma*	*Média*	*Variância*		
Coluna 1	10	1238	123,8	106,6222		
Coluna 2	10	1315	131,5	120,2778		
Coluna 3	10	1427	142,7	215,1222		
ANOVA						
Fonte da variação	*SQ*	*gl*	*MQ*	*F*	*valor-P*	*F crítico*
Entre grupos	1806,467	2	903,2333	6,130235	0,006383	5,488118
Dentro dos grupos	3978,2	27	147,3407			
Total	5784,667	29				

Decisão: Rejeitar H_0. Há evidência suficiente para apoiar a afirmação. Você pode concluir que há uma diferença nas médias dos tempos de voo.