Análise de Variância – ANOVA

Introdução

Quando se usa a ANOVA?

Comparação envolvendo mais de dois tratamentos (grupos), ou situações experimentais.

O que a ANOVA faz?

Compara variação dentro de cada tratamento e entre os tratamentos.

ANOVA – Fator único

Passos a serem realizados para Análise de Variância:

1. Estabelecer as hipótese:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_m \\ H_1 : existe\ pelo\ menos\ uma\ média\ diferente \end{cases}$$

2. Estabelecer o índice de significância: α

3. Estatística: Teste F

4. Definir a região Crítica

Para localizar o $F_{crit} = F5\%;\ gl_{trat};\ gl_{erro}$

5. Conclusão do teste:

F calculado < F tabelado = H_0 não rejeitada

F calculado > F tabelado = H_0 rejeitada.

ANOVA – Fator único

Soma de Quadrados

A abordagem da análise da variância pode ser mais bem compreendida observando-se a soma dos quadrados dos desvios (ou soma de quadrados).

Quadro da ANOVA – Fator único para m grupos distintos. Cada um com n elementos

Fonte de Variação	SQ	gl	QM	F
Entre os grupos (tratamento)	SQTR	m – 1	$QM_{trat} = SQTR/gl_{trat}$	QM_{trat}/QM_{erro}
Dentro dos grupos (erro/resíduo)	SQER	m(n – 1)	$QM_{erro} = SQER/gl_{erro}$	
Total	STQ	$gl_{trat} + gl_{erro}$		

SQ = soma dos quadrados

QM = variância do quadrado médio

gl = grau de liberdade

F = F_{calc}

ANOVA – Fator único

Formulário

SQTR - Soma de quadrados de tratamentos: Desvio de cada média de grupo em relação ao média global. (Esses desvios podem ser considerados como devidos ao fato de os grupos terem recebidos tratamentos diferentes.)

$$SQTR = n\sum_{j=1}^{m}(\bar{x}_j - \bar{\bar{x}})^2$$

onde, a Média global é: $\bar{\bar{x}} = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}}{mn}$

m= número de grupos e

n=número de elementos em cada grupo.

$\bar{x}_j$ é a média de cada grupo , $j = 1, \ldots, m$

ANOVA – Fator único

Formulário

SQER - Soma de quadrados dos erros: Desvio de cada elemento em relação a média do grupo. (Podemos considerar essas diferenças como decorrentes de fatores aleatórios desconhecidos.)

$$SQER = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{x}_j)^2$$

STQ - É a soma total:

$$STQ = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{\bar{x}})^2$$

ou ainda, $STQ = SQTR + SQER$

ANOVA – Fator único

Variância Quadrática Média

- Cada estatística do tipo soma de quadrados tem uma grandeza associada, que são seus graus de liberdade;
- O número de graus de liberdade de SQTR é m – 1;
- O número de graus de liberdade de SQER é m(n – 1);
- Quando uma soma de quadrados é dividida pelos seus graus de liberdade, o seu resultado é chamado de variância quadrática média;
- A variância quadrática média dos tratamentos é dada por:

$$QM_{trat} = \frac{SQTR}{m-1}$$

- A variância quadrática média do erro é dada por:

$$QM_{erro} = \frac{SQER}{m(n-1)}$$

ANOVA – Fator único

Estatística do Teste

Tomando-se a variância quadrática média dos tratamentos dividida pela variância quadrática média dos erros, o resultado é a estatística F:

$$F_{cal} = \frac{QM_{trat}}{QM_{erro}}$$

ANOVA – Fator único

Exemplo 1

Um professor deseja testar três métodos de ensino diferentes, I, II e III. Para isso, escolhe aleatóriamente três grupos de 5 alunos, e cada grupo é ensinado por um método diferente. Dá-se então uma mesma prova a todos os alunos; os resultados (notas) constam na tabela abaixo. Determine se há diferença significativa entre os três métodos, ao nível de significância de 5%.

Método I	75	62	71	58	73
Método II	81	85	68	92	90
Método III	73	79	60	75	81

ANOVA – Fator duplo sem repetição

Nesses modelos, os elementos são classificados segundo dois critérios, ou seja, existem duas características de interesse que deverão ser testadas simultaneamente, constituindo duas classificações cruzadas.

Testes a serem realizados

$$I: \begin{cases} H_0\ linha:\ \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_n \\ H_1\ linha: existe\ pelo\ menos\ uma\ diferente \end{cases}$$

$$II: \begin{cases} H_0\ coluna:\ \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_m \\ H_1\ coluna: existe\ pelo\ menos\ uma\ diferente \end{cases}$$

ANOVA – Fator duplo sem repetição

Quadro da ANOVA – Fator duplo sem repetição

Fonte de Variação	SQ	gl	QM	F
Entre colunas	SQC	$m-1$	$QMC = \frac{SQC}{m-1}$	QMC/QMER
Entre linhas	SQL	$n-1$	$QML = \frac{SQL}{n-1}$	QML/QMER
Dentro dos grupos (erro/resíduo)	SQER	$(m-1)(n-1)$	$QMER = SQER/(m-1)(n-1)$	
Total	STQ	$gl_c + gl_l + gl_{er}$		

SQ = soma dos quadrados

QM = variância do quadrado médio

gl = grau de liberdade

F = F_{calc}

ANOVA – Fator duplo sem repetição

Formulários

Para encontrar os F críticos:

$F_{crit/linha} = F_{\alpha\,;\,gl_L;\,gl_{erro}}$ $\qquad$ $F_{crit/coluna} = F_{\alpha\,;\,gl_C;\,gl_{erro}}$

STQ - Soma total de quadrados:

$$STQ = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{\bar{x}})^2$$

ou ainda, $STQ = SQlin + SQcol + SQerro$

ANOVA – Fator duplo sem repetição

Formulários

SQcol - soma de quadrados entre colunas:

$$SQcol = n\sum_{j=1}^{m}(\bar{x}_{cj} - \bar{\bar{x}})^2$$

$\bar{x}_{cj}$ representa a média de todos os elementos da coluna j.

- SQcol tem (m-1) graus de liberdade

SQlin - Soma de quadrados entre linhas:

$$SQlin = m\sum_{i=1}^{n}(\bar{x}_{li} - \bar{\bar{x}})^2$$

$\bar{x}_{li}$ representa a média de todos os elementos da linha i.

- SQlin tem (n-1) graus de liberdade.

ANOVA – Fator duplo sem repetição

Formulários

SQerro= Soma dos quadrados dos erros:

$$SQER = \sum_{i=1}^{n}\sum_{j=1}^{m}[x_{ij} - \bar{\bar{x}} - (\bar{x}_{cj} - \bar{\bar{x}}) - (\bar{x}_{li} - \bar{\bar{x}})]^2$$

ou ainda, $SQER = STQ - SQcol - SQcol$

- SQER tem (m-1)(n-1) graus de liberdade.

ANOVA – Fator duplo sem repetição

$$QMC = \frac{SQcol}{m-1}$$

$$QML = \frac{SQlin}{n-1}$$

$$QMER = \frac{SQER}{(m-1)(n-1)}$$

ANOVA – Fator duplo sem repetição

F_{col} - Estatística F para colunas:

$$F_{col} = \frac{QMC}{QMER}$$

F_{lin} - Estatística F para linhas:

$$F_{lin} = \frac{QML}{QMER}$$

ANOVA – Fator duplo sem repetição

Exemplo 2

A tabela abaixo apresenta o número de peças defeituosas produzidas por quatro operários trabalhando em três máquinas diferentes. Faça uma análise de variância considerando os dois fatores (máquinas e operários) utilizando 5% de significância.

Máquinas	Operários				Total
	B_1	B_2	B_3	B_4	
A_1	35	38	41	32	146
A_2	31	40	38	31	140
A_3	36	35	43	25	139
Total	102	113	122	88	425