Two-Way ANOVA

Prof. Lorí Viali, Dr.
http://www.ufrgs.br/~viali/
viali@mat.ufrgs.br

A análise de variância de uma classificação (One-Way ANOVA) verifica se as médias de "k" amostras independentes (tratamentos) diferem entre si.

Um segundo tipo de análise de variância, denominado de ANOVA de Dupla Classificação (Two-Way ANOVA) testa se existe diferença entre duas variáveis categóricas.

Neste caso, a segunda variável categórica é denominada de “Bloco”. Assim é possível testar se existe diferença simultânea entre os tratamentos (médias das amostras independentes) e se simultaneamente estas diferenças podem ser debitadas a segunda variável ou blocos.

Esta ANOVA é de blocos completos, isto é, cada bloco inclui todos os tratamentos e sem repetição, isto é, cada bloco apresenta apenas uma parcela com cada tratamento. Este desenho pode incluir combinações mais complexas, como blocos incompletos ou repetição dos tratamentos.

Seja Y_{ij} a variável dependente, neste caso o índice "i" indica o tramento e o índice "j" o bloco. Por exemplo, a variável Y_{ij} pode representar a "Renda de Pessoas" pertencentes a "k" categorias profissionais (tratamentos) em "l" empresas diferentes (blocos).

Vamos admitir o seguinte modelo linear:

$$Y_{ij} = \mu + \alpha_i + \beta_j + u_{ij}, \text{ com:}$$

$$\sum_{i=1}^{k} \alpha_i = 0 \qquad e \qquad \sum_{j=1}^{l} \beta_i = 0$$

Neste modelo μ *é a média geral,* α_i *representa o efeito dos tratamentos e* β_j *o efeito dos blocos.*

As hipóteses feitas sobre u_{ij} para o modelo anterior continuam válidas aqui. Isto é, os termos erro são variáveis aleatórias independentes com distribuição normal de média "zero" e desvio padrão "σ^2".

Sejam m, a_i e b_j as estimativas de μ, dos α_i e dos β_j respectivamente. Então o modelo amostral será:

$$Y_{ij} = m + a_i + b_j + e_{ij}$$

Dados os valores observados Y_{ij}, as estimativas dos parâmetros μ, α_i e β_j podem ser determinadas através do Método dos Mínimos Quadrados. Para isto deve-se minimizar a soma dos quadrados dos resíduos, isto é:

$$Q = S.Q.R = \sum_{i}^{k} \sum_{j}^{l} E_{ij}^{2} = \sum_{i}^{k} \sum_{j}^{l} (Y_{ij} - m - a_i - b_j)^2$$

Derivando e igualando a zero, tem-se:

$$\frac{\partial Q}{\partial m_i} = 2 \sum_{i=1}^{k} \sum_{j=1}^{l} (Y_{ij} - m - a_i - b_j)(-1) = 0$$

$$\frac{\partial Q}{\partial a_i} = 2\sum_{j=1}^{l}(Y_{ij} - m - a_i - b_j)(-1) = 0$$

$$i = 1,2,\ldots,k$$

e

$$\frac{\partial Q}{\partial b_j} = 2\sum_{i=1}^{k}(Y_{ij} - m - a_i - b_j)(-1) = 0$$

$$j = 1,2,\ldots,l$$

Fazendo n = k.l,

$$\mathcal{A}_i = \sum_{j=1}^{l} Y_{ij} \quad (i = 1, 2, \ldots, k)$$

$$\mathcal{B}_j = \sum_{i=1}^{k} Y_{ij} \quad (j = 1, 2, \ldots, l)$$

e

$$\mathcal{G} = \sum_{i=1}^{k} \sum_{j=1}^{l} Y_{ij} = \sum_{i=1}^{k} \mathcal{A}_i = \sum_{j=1}^{l} \mathcal{B}_j$$

Obtém-se, o seguinte sistema de equações:

$$G = nm + l\sum_{i=1}^{k} a_i + k\sum_{j=1}^{l} b_j$$

$$A_i = lm + l\,a_i + \sum_{j=1}^{l} b_j \quad (i = 1, 2, ..., k)$$

$$B_j = km + \sum_{i=1}^{k} a_i + k\,b_j \quad (j = 1, 2, ..., l)$$

Este sistema tem $l + k + 1$ equações, com o mesmo número de incógnitas. Destas equações apenas $k + l - 1$ são Linearmente Independentes. Para resolver este sistema utilizam-se as restrições sobre a_i e b_j.

$$\sum_{i=1}^{k} \alpha_i = 0 \qquad e \qquad \sum_{j=1}^{\ell} \beta_j = 0$$

Tem-se, então: $m = G/n$

$$a_i = (A_i/\ell) - (G/n) = (A_i/\ell) - m$$

$$b_j = (B_j/k) - (G/n) = (B_j/k) - m$$

Substituindo, estes resultados na expressão dos Mínimos Quadrados, tem-se:

$$S.Q.R = \sum_{i}^{k} \sum_{j}^{l} \left(Y_{ij} - \frac{A_i}{l} - \frac{B_j}{k} + \frac{G}{n} \right)^2$$

Por definição, a soma dos quadrados total, a soma dos quadrados dos tratamentos e a soma dos quadrados dos blocos são dadas por:

$$S.Q.Total = \sum_{i}^{k} \sum_{j}^{l} (Y_{ij} - m)^2$$

$$S.Q.Trat. = l \sum_{i}^{k} a_i^2 = l \sum_{i}^{k} \left(\frac{A_i}{l} - m\right)^2$$

$$S.Q.Blocos = k \sum_{j}^{l} b_j^2 = k \sum_{i}^{k} \left(\frac{B_j}{k} - m\right)^2$$

Deve-se notar que todas as somas de quadrados são somas de $n = kl$ *parcelas, onde cada parcela é um quadrado. Assim para obter a soma de quadrados total, somamos os quadrados dos desvios dos* $n = kl$ *valores observados em relação as respectivas estimativas, dadas por* $m + a_i + b_j$.

Para obter a soma de quadrados de tratamentos multiplicamos por "l"" as somas dos quadrados das "k" diferenças de médias estimadas de tratamentos, dadas por A_i/l, em relação a média "m", ou seja, multiplicamos por "l " a soma dos quadrados das estimativas dos "k" efeitos de tratamentos.

Finalmente, para se obter a soma dos quadrados de blocos multiplicamos por “k” as somas dos quadrados das “l” diferenças de médias estimadas de blocos, dadas por B_j/k, em relação a média “m”, ou seja, multiplicamos por “k” as somas dos quadrados das estimativas dos “n” efeitos de blocos.

Pode-se mostrar através de manipulação algébrica que:

$$S.Q.Total = \sum_{i=1}^{k} \sum_{j=1}^{l} (Y_{ij} - m)^2 =$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{l} Y_{ij}^2 - \frac{G^2}{n}$$

e:

$$S.Q.Trat. = r\sum_{i=1}^{k} a_i^2 = r\sum_{i=1}^{k}\left(\frac{A_i}{r} - m\right)^2 =$$

$$= \frac{1}{r}\sum_{i=1}^{k} {A_i}^2 - \frac{G^2}{n}$$

E também:

$$S.Q.Blocos = k\sum_{j=1}^{l} b_j^2 = k\sum_{j=1}^{l}\left(\frac{B_j}{k} - m\right)^2 =$$

$$= \frac{1}{k}\sum_{j=1}^{l} B_j^{\,2} - \frac{G^2}{n}$$

Finalmente:

$$S.Q.R = \sum_{i}^{k} \sum_{j}^{l} \left(Y_{ij} - \frac{A_i}{l} - \frac{B_j}{k} + \frac{G}{n} \right)^2 =$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{l} Y_{ij}^{2} - \frac{1}{l} \sum_{i=1}^{k} A_i^{2} - \frac{1}{k} \sum_{j=1}^{l} B_j^{2} + \frac{G^2}{n}$$

Resumindo, tem-se então:

$$S.Q.R = S.Q.Total - S.Q.Trat. - S.Q.Blocos$$

Dados

Bloco	Tratamentos			Total do Bloco
	1	2	3	
1	12	17	21	50
2	14	19	23	56
3	15	18	19	52
4	18	19	18	55
5	16	21	22	59
6	13	20	19	52
Total Trat.	88	114	122	324

Tem-se:

$$k = 3; \;\; l = 6; \;\; n = k.l = 3.6 = 18$$

$$G = 324 \quad G^2/n = 5832$$

$$\Sigma A_i = \Sigma B_i = \Sigma Y_{ij} = 324$$

$$\Sigma A_i^2 = 35624; \Sigma B_i^2 = 17550$$

Então:

$$S.Q.Total = \sum_{i=1}^{k} \sum_{j=1}^{l} Y_{ij}^{2} - \frac{G^2}{n} = 158$$

$$S.Q.Trat. = \frac{1}{l} \sum_{i=1}^{k} A_i^{2} - \frac{G^2}{n} = 105{,}33$$

$$S.Q.Blocos = \frac{1}{k} \sum_{j=1}^{l} B_j^{2} - \frac{G^2}{n} = 18$$

Portanto:

$$S.Q.R = 158 - 105{,}33 - 18 = 34{,}67$$

O modelo para a análise de variância de dupla classificação é dado por: $Y_{ij} = \mu + \alpha_i + \beta_j + u_{ij}$,

Considerando μ, α_i e β_j como valores fixos e lembrando que u_{ij} são variáveis aleatórias independentes com média zero e variância σ^2, vem:

$$E(Y_{ij}^{2}) = \mu^2 + \alpha_i^2 + \beta_j^2 + \sigma^2$$

Então:

$$E\left(\sum_{i=1}^{k}\sum_{j=1}^{l} Y_{ij}^{2}\right) = n\mu^2 + l\sum_{i=1}^{k}\alpha_i^2 + k\sum_{j=1}^{l}\beta_j^2 + n\sigma^2$$

Como: $$A_i = \sum_{j=1}^{l} Y_{ij}$$

De acordo com o modelo dado e as suas restrições, tem-se: $$A_i = l(\mu + \alpha_i) + \sum_{j=1}^{l} U_{ij}$$

Segue, então:

$$A_i^2 = l^2(\mu+\alpha_i)^2 + 2l(\mu+\alpha_i)\sum_{j=1}^{l} U_{ij} + \left(\sum_{j=1}^{l} U_{ij}\right)^2$$

Mas os termos erros são variáveis com média zero, assim: $E(A_i^2) = \ell^2(\mu+\alpha_i)^2 + \ell\sigma^2$

De forma semelhante, tem-se:

$$E(B_j^2) = k^2(\mu+\beta_j)^2 + k\sigma^2$$

E, também: $$E(G^2) = n^2\mu^2 + n\sigma^2$$

Então:

$$E(S.Q.Total) = l\sum_{i=1}^{k}\alpha_i^2 + k\sum_{j=1}^{l}\beta_j^2 + (n-1)\sigma^2$$

Para os tratamentos, tem-se:

$$E(S.Q.Trat.) = \ell \sum_{i=1}^{k} (\mu + \alpha_i)^2 + k\sigma^2 + n\mu^2 - \sigma^2$$

Lembrando que $\sum_{i=1}^{k} \alpha_i = 0$ *tem-se:*

$$E(S.Q.Trat.) = \ell \sum_{i=1}^{k} \alpha_i^2 + (k-1)\sigma^2$$

Para os blocos, obtemos:

$$E(S.Q.Blocos) = k\sum_{j=1}^{l}\beta_j^2 + (l-1)\sigma^2$$

Mas:

$$S.Q.R = S.Q.Total - S.Q.Trat. - S.Q.Blocos$$

Então:

$$E(S.Q.R.) = E(S.Q.Total) - E(S.Q.Trat.) - E(S.Q.Blocos)$$

Substituindo, segue:

$$E(S.Q.Res.) = (n - k - l + 1)\sigma^2 =$$
$$= (k - 1)(l - 1)\sigma^2$$

Como os U_{ij} são variáveis aleatórias independentes de média "zero" e variância constante e igual a σ^2, tem-se:

S.Q.Res.)/σ^2 apresenta uma distribuição Qui-Quadrado com $(k -1)(l - 1)$ graus de liberdade;

Supondo $\mathcal{H}_0$ verdadeira, isto é, que as médias dos tratamentos são iguais, isto é, os tratamentos não têm efeitos, ou ainda: $\mathcal{H}_0$: $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$, tem-se que: (S.Q.Trat.)/ σ^2 tem uma distribuição Qui-Quadrado com "k - 1" graus de liberade.

Supondo que as médias dos blocos são todas iguais entre si, ou que o efeito dos blocos é nulo, ou ainda que: $\mathcal{H}_0$: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$, *tem-se:* $(S.Q.Blocos)/\sigma^2$ *tem uma distribuição Qui-Quadrado com "l - 1" graus de liberdade.*

Tomando agora os quadrados médios, isto é, a soma dos quadrados divididos pelos respectivos graus de liberdade, pode-se obter a expectância dos quadrados médios. A tabela seguinte resume este e outros resultados relevantes obtidos.

O modelo para a análise de variância de dupla classificação é dado por: $Y_{ij} = \mu + \alpha_i + \beta_j + u_{ij}$,

Considerando μ, α_i *e* β_j *como valores fixos e lembrando que* U_{ij} *são VA independentes com média zero e variância* σ^2, *tem-se:*

Objetivos

A Análise de variância (ANOVA) É utilizada para mostrar os efeitos principais de variáveis categóricas independentes (denominadas de fatores) sobre uma variável quantitativa dependente.

Causa da Variação	Graus de Liberdade	Soma dos Quadrados	Esp. do Quad. Médio
Tratamentos	$k - 1$	$\frac{1}{n}\sum_{i=1}^{k} A_i^2 - \frac{G^2}{n}$	$l\frac{\sum_{i=1}^{k}\alpha_i^2}{k-1} + \sigma^2$
Blocos	$l - 1$	$\frac{1}{k}\sum_{j=1}^{l} B_j^2 - \frac{G^2}{n}$	$k\frac{\sum_{j=1}^{l}\beta_j^2}{l-1} + \sigma^2$
Resíduo	$(k-1)(l-1)$		σ^2
Total	$kn - 1$	$\sum_{i=1}^{k}\sum_{j=1}^{l} Y_{ij}^2 - \frac{G^2}{n}$	

Teste

Conforme já visto, na análise de variância com uma classificação, o quociente:

$F = (Q.M.Trat.)/(Q.M.Res.)$

pode ser utilizado para testar a hipótese:

$H_0: \alpha_1 = \alpha_2 = ... = \alpha_k = 0,$

Este resultado apresenta uma distribuição F com "k - 1" e "(k - 1)(l - 1)" graus de liberdade.

Teste

De forma semelhante o quociente:

$\mathcal{F}$ = (Q.M.Blocos.)/(Q.M.Res.)

pode ser utilizado para testar a hipótese:

$\mathcal{H}_0$: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$,

Este resultado apresenta uma distribuição $\mathcal{F}$ com "l - 1" e "(k - 1)(l - 1)" graus de liberdade.