MAF 261 - Estatística Experimental

Prof. Fernando de Souza Bastos

Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Campus UFV - Florestal

2018

Sumário

No Delineamento Inteiramente Casualizado (DIC) a distribuição dos tratamentos às unidades experimentais é feita inteiramente ao acaso. Os outros delineamentos experimentais, por exemplo: blocos casualizados e quadrado latino, se originam do DIC pelo uso de restrição na casualização. O DIC utiliza apenas os princípios básicos da repetição e da casualização.

Como não faz restrições na casualização, o uso do DIC pressupõe que as unidades experimentais estão sob condições homogêneas. Estas condições homogêneas geralmente são obtidas em locais com ambientes controlados tais como laboratórios, estufas e casas de vegetação.

A título de exemplo, considere um experimento instalado no DIC com I tratamentos e J repetições. A coleta de dados da pesquisa pode ser resumida, num quadro do tipo a seguir:

	Tratamentos			
Repetições	1	2	...	I
1	Y_{11}	Y_{21}	...	Y_{I1}
2	Y_{12}	Y_{22}	...	Y_{I2}
...	...	...	...	...
J	Y_{1J}	Y_{2J}	...	Y_{IJ}
Totais	T_1	T_2	...	T_I

Deste quadro pode-se retirar algumas informações de interesse:

- Número de unidades experimentais: $N = I \times J$

Deste quadro pode-se retirar algumas informações de interesse:

- Número de unidades experimentais: $N = I \times J$
- Total geral: $G = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} = \sum_{i=1}^{I} T_i = Y_{\bullet\bullet}$

Deste quadro pode-se retirar algumas informações de interesse:

- Número de unidades experimentais: $N = I \times J$
- Total geral: $G = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} = \sum_{i=1}^{I} T_i = Y_{\bullet\bullet}$
- Total para o tratamento i : $T_i = \sum_{j=1}^{J} Y_{ij} = Y_{i\bullet}$

Deste quadro pode-se retirar algumas informações de interesse:

- Número de unidades experimentais: $N = I \times J$
- Total geral: $G = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} = \sum_{i=1}^{I} T_i = Y_{\bullet\bullet}$
- Total para o tratamento i : $T_i = \sum_{j=1}^{J} Y_{ij} = Y_{i\bullet}$
- Média para o tratamento i : $\mu_i = \dfrac{T_i}{J}$

Deste quadro pode-se retirar algumas informações de interesse:

- Número de unidades experimentais: $N = I \times J$
- Total geral: $G = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} = \sum_{i=1}^{I} T_i = Y_{\bullet\bullet}$
- Total para o tratamento i : $T_i = \sum_{j=1}^{J} Y_{ij} = Y_{i\bullet}$
- Média para o tratamento i : $\mu_i = \frac{T_i}{J}$
- Média geral do experimento: $\mu = \frac{G}{IJ}$

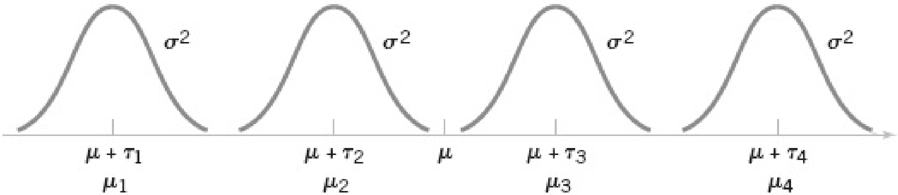
Modelo Estatístico

Existe um modelo estatístico específico para cada tipo de delineamento. O modelo estatístico identifica quais são as fontes de variação dos valores de uma variável resposta em estudo. Para os dados oriundos de um experimento instalado segundo o DIC, o seguinte modelo estatístico deve ser utilizado nas análises estatísticas:

$$Y_{ij} = \mu + t_i + e_{ij}$$

em que, Y_{ij} é o valor observado para a variável resposta obtido para o i-ésimo tratamento em sua j-ésima repetição; μ é a média de todos os valores possíveis da variável resposta; t_i é o efeito do tratamento i no valor observado Y_{ij}, ou seja, $t_i = \mu_i - \mu$; Por fim, e_{ij} é o erro experimental associado ao valor observado Y_{ij}, isto é, $e_{ij} = Y_{ij} - \mu_i$.

O erro experimental ocorre em todos os experimentos, porque não é possível controlar o efeito de fontes de variações que ocorrem de forma aleatória e desconhecida. Este erro é o responsável pela variação observada entre as observações obtidas nas repetições para cada tratamento.



Análise de Variância é uma técnica de análise estatística que permite decompor a variação total, ou seja, a variação existente entre todas as observações, na variação devido à diferença entre os efeitos dos tratamentos e na variação devido ao acaso, que também é denominada de erro experimental ou resíduo.

No entanto, para que esta técnica seja empregada é necessário que sejam satisfeitas as seguintes pressuposições:

1. os efeitos do modelo estatístico devem ser aditivos;

No entanto, para que esta técnica seja empregada é necessário que sejam satisfeitas as seguintes pressuposições:

1. os efeitos do modelo estatístico devem ser aditivos;
2. os erros experimentais devem ser normalmente distribuídos, independentes, com média zero e com variância comum.

Partindo do modelo estatístico, pode-se decompor a variação entre os valores observados nas diferentes causas de variabilidade, como demonstrado a seguir.

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Fazendo $t_i = \mu_i - \mu$ e $e_{ij} = Y_{ij} - \mu_i$, tem-se:

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Fazendo $t_i = \mu_i - \mu$ e $e_{ij} = Y_{ij} - \mu_i$, tem-se:

$$Y_{ij} - \mu = (\mu_i - \mu) + (Y_{ij} - \mu_i),$$

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Fazendo $t_i = \mu_i - \mu$ e $e_{ij} = Y_{ij} - \mu_i$, tem-se:

$$Y_{ij} - \mu = (\mu_i - \mu) + (Y_{ij} - \mu_i),$$

Substituindo μ, μ_i e e_{ij} por seus estimadores tem-se:

Considere o modelo estatístico para um experimento instalado segundo o DIC:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Fazendo $t_i = \mu_i - \mu$ e $e_{ij} = Y_{ij} - \mu_i$, tem-se:

$$Y_{ij} - \mu = (\mu_i - \mu) + (Y_{ij} - \mu_i),$$

Substituindo μ, μ_i e e_{ij} por seus estimadores tem-se:

$$Y_{ij} - \hat{\mu} = (\hat{\mu}_i - \hat{\mu}) + (Y_{ij} - \hat{\mu}_i),$$

elevando ambos os membros ao quadrado e aplicando o somatório, temos:

$$\sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} [(\hat{\mu}_i - \hat{\mu}) + (Y_{ij} - \hat{\mu}_i)]^2$$

$$\Rightarrow \sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu}_i)^2 + \sum_{i=i,j=1}^{I,J} \text{duplos produtos}$$

É fácil mostrar que $\sum_{i=i,j=1}^{I,J} \text{duplos produtos} = 0$. Logo, podemos escrever, simplificadamente:

$$SQTotal = SQTrat + SQRes$$

Por meio destas fórmulas, pode-se obter os valores para as respectivas somas de quadrados. No entanto, essas fórmulas demandam muitos cálculos. Fórmulas de mais fácil aplicação podem ser obtidas, conforme é mostrado a seguir. Inicialmente trabalharemos com a fórmula da SQTotal. Tem-se:

$$SQTotal = \sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2$$

Por meio destas fórmulas, pode-se obter os valores para as respectivas somas de quadrados. No entanto, essas fórmulas demandam muitos cálculos. Fórmulas de mais fácil aplicação podem ser obtidas, conforme é mostrado a seguir. Inicialmente trabalharemos com a fórmula da SQTotal. Tem-se:

$$SQTotal = \sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2$$

Desenvolvendo o quadrado perfeito,

$$\sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} (Y_{ij}^2 - 2\hat{\mu}Y_{ij} + \hat{\mu}^2)$$

aplicando-se as propriedades de somatório, temos:

$$\sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} Y_{ij}^2 - 2\hat{\mu} \sum_{i=i,j=1}^{I,J} Y_{ij} + \sum_{i=i,j=1}^{I,J} \hat{\mu}^2 \quad (1)$$

$$= \sum_{i=i,j=1}^{I,J} Y_{ij}^2 - 2\hat{\mu} \sum_{i=i,j=1}^{I,J} Y_{ij} + IJ\hat{\mu}^2 \quad (2)$$

Sabemos que $\hat{\mu} = \dfrac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ}$, assim:

$$\sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} Y_{ij}^2 - 2\left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ}\right)\sum_{i=i,j=1}^{I,J} Y_{ij} + IJ\left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ}\right)^2$$

simplificando tem-se,

$$\sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} Y_{ij}^2 - 2\frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ} + \frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

Finalmente, temos:

$$SQTotal = \sum_{i=i,j=1}^{I,J} (Y_{ij} - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} Y_{ij}^2 - \frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

que é a fórmula mais prática para se calcular a SQTotal.

Para a SQTratamentos tem-se:

$$SQTrat = \sum_{i=1,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2$$

Para a SQTratamentos tem-se:

$$SQTrat = \sum_{i=1,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2$$

desenvolvendo o quadrado perfeito,

$$\sum_{i=1,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = \sum_{i=1,j=1}^{I,J} (\hat{\mu}_i^2 - 2\hat{\mu}\hat{\mu}_i + \hat{\mu}^2)$$

aplicando-se as propriedades de somatório, temos:

$$\sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = \sum_{i=i,j=1}^{I,J} \hat{\mu}_i^2 - 2\hat{\mu} \sum_{i=i,j=1}^{I,J} \hat{\mu}_i + \sum_{i=i,j=1}^{I,J} \hat{\mu}^2 \tag{3}$$

$$= J \sum_{i=i}^{I} \hat{\mu}_i^2 - 2\hat{\mu} J \sum_{i=i}^{I} \hat{\mu}_i + IJ\hat{\mu}^2 \tag{4}$$

A média geral e a média para tratamentos podem ser escritas respectivamente como:

$$\hat{\mu} = \frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ} \text{ e } \hat{\mu}_i = \frac{T_i}{J}$$

Substituindo na expressão anterior, tem-se:

$$\sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = J \sum_{i=i}^{I} \frac{T_i^2}{J^2} - 2 \left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ} \right) J \sum_{i=i}^{I} \frac{T_i}{J} + IJ \left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ} \right)^2$$

Sabe-se que $T_i = \sum_{j=1}^{J} Y_{ij}$, então:

$$\sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = J \sum_{i=i}^{I} \frac{T_i^2}{J^2} - 2 \left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ} \right) J \left(\frac{\sum_{j=1}^{J} Y_{ij}}{J} \right) + IJ \left(\frac{\sum_{i=i,j=1}^{I,J} Y_{ij}}{IJ} \right)^2$$

simplificando, tem-se:

$$\sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = \sum_{i=i}^{I} \frac{T_i^2}{J} - 2\frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ} + \frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

finalmente tem-se:

$$SQTrat = \sum_{i=i,j=1}^{I,J} (\hat{\mu}_i - \hat{\mu})^2 = \sum_{i=i}^{I} \frac{T_i^2}{J} - \frac{\left(\sum_{i=i,j=1}^{I,J} Y_{ij}\right)^2}{IJ}$$

A fórmula anterior é utilizada quando o número de repetições é igual para todos os tratamentos. No caso em que o número de repetições varia de acordo com o tratamento a fórmula apropriada é:

$$SQTrat = \sum_{i=i}^{I} \frac{T_i^2}{r_i} - \frac{\left(\sum_{i=i,j=1}^{I,r_i} Y_{ij}\right)^2}{N}$$

em que, N é o número de unidades experimentais $= \sum_{i=1}^{I} r_i$, e r_i é número de unidades experimentais do tratamento i.

Por fim, A Soma de Quadrados do Resíduo ($SQRes$) é obtida por diferença,

$$SQRes = SQTotal - SQTrat$$

Análise de Variância

O valor esperado da soma dos quadrados dos tratamentos é:

$$E(SQTrat) = (I-1)\sigma^2 + n\sum_{i=1}^{I} t_i^2$$

e o valor esperado da soma de quadrados do resíduo é:

$$E(SQRes) = I(n-1)\sigma^2$$

Existe também uma divisão do número de graus de liberdade que corresponde à identidade das somas de quadrados. Ou seja, há $IJ = n$ observações; assim, SQTot tem $IJ - 1$ graus de liberdade. Existem I níveis do fator; logo, SQTrat tem $I - 1$ graus de liberdade. Finalmente, dentro de qualquer tratamento, existem replicatas (ou réplicas) fornecendo $J - 1$ graus de liberdade, com os quais se estima o erro experimental. Já que há I tratamentos, temos $I(J - 1)$ graus de liberdade para o erro. Consequentemente, a divisão dos graus de liberdade é:

$$IJ - 1 = I - 1 + I(J - 1)$$

A razão

$$QM_{Trat} = \frac{SQTrat}{I-1}$$

é chamada de média quadrática dos tratamentos ou quadrado médio dos tratamentos. Assim, se a hipótese nula $H_0 : t_1 = t_2 = \cdots = t_I = 0$ for verdadeira, $QMTrat$ será um estimador não tendencioso de σ^2 porque $\sum_{i=1}^{I} t_i = 0$. Entretanto, se H_1 for verdadeira, $QMTrat$ estimará σ^2 mais um termo positivo que incorpora a variação em razão da diferença sistemática nas médias dos tratamentos.

Notem que a média quadrática do erro ou quadrado médio dos resíduos, dada por

$$QMRes = \frac{SQRes}{I(J-1)}$$

é um estimador não tendencioso de σ^2, independente de se H_0 é ou não verdadeira. Podemos mostrar também que QMTrat e QMRes são independentes. Consequentemente, podemos mostrar que, se a hipótese nula H_0 for verdadeira, a razão

$$F_{cal} = \frac{QMTrat}{QMRes}$$

Tem distribuição F com $I-1$ e $I(J-1)$ graus de liberdade.

O quadro da análise de variância, geralmente denotada por ANOVA (ANalysis Of VAriance) para a análise de um experimento instalado segundo o DIC, com igual número de repetições para todos os tratamentos é do seguinte tipo:

FV	GL	SQ	QM	F	$F_{tab;\alpha}$
Tratamentos	$(I-1)$	SQTrat	$\dfrac{SQTrat}{I-1}$	$\dfrac{QMTrat}{QMRes}$	$[(I-1); I(J-1)]$
Resíduo	$I(J-1)$	SQRes	$\dfrac{SQRes}{I(J-1)}$		
Total	IJ-1	SQTotal			

A partir das *SQTrat* e *SQRes*, obtém-se os respectivos quadrados médios, por meio do quociente entre a soma de quadrados com o respectivo número de graus de liberdade.

A partir das *SQTrat* e *SQRes*, obtém-se os respectivos quadrados médios, por meio do quociente entre a soma de quadrados com o respectivo número de graus de liberdade.

Para se concluir se existe diferença entre tratamentos, calcula-se o valor de F, que é obtido pelo quociente do *QMTrat* com o *QMRes*. Este valor de F calculado deve ser comparado com o valor de F tabelado, o qual é obtido na tabela de distribuição da variável aleatória F, de acordo com o nível de significância do teste, graus de liberdade para tratamentos e graus de liberdade para resíduo.

As hipóteses para o teste F da análise de variância para tratamentos são as seguintes:

$H_0 : \mu_1 = \cdots = \mu_I = \mu$, o que equivale a dizer que todos os possíveis contrastes entre as médias dos tratamentos, são estatisticamente nulos, ao nível de probabilidade que foi executado o teste.

As hipóteses para o teste F da análise de variância para tratamentos são as seguintes:

$H_0 : \mu_1 = \cdots = \mu_I = \mu$, o que equivale a dizer que todos os possíveis contrastes entre as médias dos tratamentos, são estatisticamente nulos, ao nível de probabilidade que foi executado o teste.

$H_a : \text{não } H_0$, o que equivale a dizer que existe pelo menos um contraste entre as médias dos tratamentos, estatisticamente diferentes de zero, ao nível de probabilidade que foi realizado o teste.

A regra decisória para o teste F é a seguinte:

- Se o valor do F calculado for maior ou igual ao valor do F tabelado, então rejeita-se H_0 e conclui-se que os tratamentos tem efeito diferenciado ao nível de significância em que foi realizado o teste;

A regra decisória para o teste F é a seguinte:

- Se o valor do F calculado for maior ou igual ao valor do F tabelado, então rejeita-se H_0 e conclui-se que os tratamentos tem efeito diferenciado ao nível de significância em que foi realizado o teste;
- Se o valor de F calculado for menor que o valor do F tabelado, então não rejeita-se H_0 e conclui-se que os tratamentos têm efeitos iguais ao nível de significância em que foi realizado o teste.

Coeficiente de Variação

O coeficiente de variação é calculado da seguinte maneira:

$$CV = \frac{\sqrt{QMRes}}{\hat{\mu}}.100$$

O CV é utilizado para avaliação da precisão de experimentos. Quanto menor o CV mais preciso tende a ser o experimento.

Vantagens e Desvantagens do delineamento inteiramente casualizado

Vantagens:

- Não existem exigências quanto ao número de tratamentos e repetições;

Vantagens e Desvantagens do delineamento inteiramente casualizado

Vantagens:

- Não existem exigências quanto ao número de tratamentos e repetições;
- É o delineamento com maior valor para os graus de liberdade do resíduo.

Vantagens e Desvantagens do delineamento inteiramente casualizado

Vantagens:

- Não existem exigências quanto ao número de tratamentos e repetições;
- É o delineamento com maior valor para os graus de liberdade do resíduo.

Desvantagens:

- Não é fácil conseguir e manter total homogeneidade das condições;

Vantagens e Desvantagens do delineamento inteiramente casualizado

Vantagens:

- Não existem exigências quanto ao número de tratamentos e repetições;
- É o delineamento com maior valor para os graus de liberdade do resíduo.

Desvantagens:

- Não é fácil conseguir e manter total homogeneidade das condições;
- todas as variações exceto a devida a tratamentos, são consideradas aleatórias. Isto pode acarretar uma estimativa muito alta para o erro experimental.

Para comparar a produtividade de quatro variedades de milho, um agrônomo tomou vinte parcelas similares e distribuiu, inteiramente ao acaso, cada uma das 4 variedades em 5 parcelas experimentais. A partir dos dados experimentais fornecidos, é possível concluir que existe diferença significativa entre as variedades com relação a produtividade, utilizando o nível de significância de 5%?

	Variedades			
	A	B	C	D
	25	31	22	33
	26	25	26	29
	20	28	28	31
	23	27	25	34
	21	24	29	28
Totais	115	135	130	155
Médias	23	27	26	31

Referências Bibliográficas I

A. J. d. A. Calegare. *Introdução ao Delineamento de Experimentos.* Edgard Blucher, São Paulo, 2 edition, 2009.

A. P. Carneiro, J. I. R. Júnior, e N. T. Santos. Apostila de Estatística Experimental - UFV, 2018. Apostila gentilmente cedida pelos autores para o curso de Estatística Experimental do Campus UFV - Florestal.