Delineamento Inteiramente Casualizado

Rodrigo R. Pescim

Universidade Estadual de Londrina

26 de abril de 2020

- Suponha que temos a tratamentos ou diferentes níveis de um único fator que se queira comparar. A resposta observada de cada um dos a tratamentos é uma variável aleatória. Os dados seriam da forma:

Tratamentos	Observações				Totais	Médias
1	y_{11}	y_{12}	$\cdots$	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

em que y_{ij} representa a b-ésima observação do a-ésimo tratamento;

$$y_{i\cdot} = \sum_{j=1}^{b} y_{ij} \qquad \bar{y}_{i\cdot} = \sum_{j=1}^{b} \frac{y_{ij}}{b} \qquad e \qquad \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{ab} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij}}{ab}.$$

- Por levar em consideração apenas os princípios da **repetição** e da **casualização**, são considerados os mais simples delineamentos experimentais.

- São instalados em situação de **homogeneidade**, por isso, são muito usados em laboratórios, casas de vegetação, etc.

- O modelo estatístico para o delineamento inteiramente casualizado é:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{1}$$

- O objetivo é testar hipóteses apropriadas sobre os efeitos dos tratamentos e estimá-los.

Vantagens

i) Pode-se utilizar qualquer número de tratamentos e repetições, sendo que o número de repetições pode variar de um tratamento para outro (ensaio desbalanceado) sem dificultar as análises.

ii) O número de repetições depende apenas do número de unidades experimentais disponíveis;

iii) Apresenta maior número de graus de liberdade associado ao erro em relação a outros delineamentos.

Desvantagens

1. Exige homogeneidade total das condições experimentais;

2. Pode-se obter uma **estimativa da variância devido ao erro experimental bastante alta**, quando não utilizado corretamente, pois, uma vez que não se considera o princípio do controle local,**todas as variações exceto as devidas aos tratamentos**, são consideradas como **variação ao acaso**.

Quando se instala um experimento no delineamento inteiramente casualizado, o objetivo é, em geral, **verificar se existe diferença significativa entre pelo menos duas médias de tratamentos.** As hipóteses testadas são:

$$H_0 \quad : \quad \mu_1 = \mu_2 = \cdots = \mu_a$$

$$H_1 \quad : \quad \mu_i \neq \mu_j \qquad \text{Pelo menos duas médias de trat. diferem entre si}$$

Uma forma equivalente de escrever as hipóteses anteriores é em termos dos efeitos dos tratamentos τ_i, que é:

$$H_0 \quad : \quad \tau_1 = \tau_2 = \cdots = \tau_a = 0$$

$$H_1 \quad : \quad \tau_i \neq 0 \qquad \text{Pelo menos um tratamento}$$

A análise de variância para testar essas hipóteses **só é válida** se forem satisfeitas as seguintes condições:

1. **aditividade:** os efeitos devem se somar (não há interação);
2. **independência:** os erros (ϵ_{ij}) devem ser independentes;
3. **normalidade:** os erros (ϵ_{ij}) devem possuir uma distribuição normal;
4. **homocedasticidade ou homogeneidade de variâncias:** os erros (ϵ_{ij}) devem possuir uma variância comum σ^2;

- Para a **verificação da normalidade dos erros**, em geral, utilizam-se os testes de normalidade, tais como Lilliefors e Shapiro-Wilks

- A **homogeneidade das variâncias** pode ser verificada por meio do teste de Bartlett, teste do F máximo e teste de Levene.

Para verificarmos se a hipótese nula (H_0) é aceita ou não, completa-se o seguinte **Quadro da Análise de Variância:**

Tabela 1: Quadro da Análise de Variância.

Causas de Variação	g.l.	S.Q.	Q.M.	F_{calc}	F_{tab}
Tratamentos	a - 1	SQTrat	$\frac{SQTrat}{a-1}$	$\frac{QMTrat}{QMRes}$	$F_{\alpha;a-1,a(b-1)}$
Resíduo	a(b-1)	SQRes	$\frac{SQRes}{a(b-1)}$		
Total	ab - 1	SQTotal			

Fazendo-se a usual suposição de normalidade, a estatística apropriada para $H_0 : \tau_i = 0$ é

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

Se $F_{cal} > F_{\alpha;(a-1),a(b-1)}$, rejeita-se H_0.

Somas de Quadrados

- Soma de quadrados total

$$SQ_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - C$$

- Soma de quadrados de tratamentos

$$SQ_{trat} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - C$$

- Soma de quadrados do Resíduo

$$SQ_{res} = SQ_{total} - SQ_{trat}$$

em que, $C = \frac{(\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij})^2}{ab}$

Exemplo 1

Seja y_{ij} o valor da resistência do concreto à compressão (psi) do j-ésimo corpo de prova que recebeu a i-ésima técnica de mistura (tratamento). Os valores das resistências podem ser resumidos na forma da Tabela:

Tabela 2: Valores da resistência à compressão (psi)

	Técnica A	Técnica B	Técnica C	Técnica D	
	3129	3200	2800	2600	
	3000	3300	2900	2700	
	2865	2975	2985	2600	
	2890	3150	3050	2765	Total
$y_{i\cdot}$	11884	12625	11735	10665	46909
$\sum_{j=1}^{b} y_{ij}^2$	35350966	39903125	34462725	28455225	138172041

Ao nível de 5% de significância, conclua se as técnicas de mistura afetam a resistência do concreto.

Teste de Comparações Múltiplas: Teste de Tukey

- O teste de Tukey serve para testar qualquer diferença entre duas médias
- O teste é exato quando as duas médias de tratamento têm o mesmo número de repetições
- Por ser um teste rigoroso, geralmente, o teste de Tukey é aplicado apenas no nível de 5% de significância

Aplicação do teste de Tukey

a) Calcular a expressão para o caso das médias serem obtidas com o mesmo nº de repetições:

$$\Delta = q\sqrt{\frac{QMRes}{r}}$$

em que

- $\hookrightarrow$ $\Delta = d.m.s. =$ diferença mínima significativa,
- $\hookrightarrow$ q é a amplitude total estudentizada obtidas em tabelas em função do **número de tratamentos** e do **número de graus de liberdade do resíduo**,
- $\hookrightarrow$ r é o número de repetições dos tratamentos

Aplicação do teste de Tukey

- No caso das médias serem obtidas com nº diferentes repetições, tem-se

$$\Delta = q\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right) QMRes},$$

para $i \neq j$.

b) Calcular as estimativas de duas médias dos tratamentos $- \hat{Y}$

c) $\hookrightarrow$ Se $|\hat{Y}| \geq \Delta$, o teste é significativo, o que indica que as duas médias diferem

$\hookrightarrow$ Se $|\hat{Y}| < \Delta$, o teste não é significativo, o que indica que as duas médias não diferem

Exemplo 2

Um engenheiro está interessado no efeito, na condutividade do tubo, de cinco tipos diferentes de recobrimento de tubos de raios catódicos em uma tela de um sistema de telecomunicações. Os seguintes dados de condutividade são obtidos.

Tabela 3: Valores de condutividade

Tipo 1	Tipo 2	Tipo 3	Tipo 4	Tipo 5
143	152	134	129	147
141	149	133	127	148
150	137	132	132	144
146	143	127	129	142

Ao nível de 5% de significância, é possível afirmar que existe diferença nos tipos de recobrimento de tubos de raios catódicos ?

- O quadro da ANOVA para o Exemplo 1 é dado por:

```
> resist.aov<-aov(resist ~ trat , resist.dat)
> anova(resist.aov)
Analysis of Variance Table

Response: resist
          Df  Sum Sq   Mean Sq  F value    Pr(>F)
trat       3  489740   163247   12.728    0.0004887 ***
Residuals 12  153908   12826
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
```

$F_{tab} = 3.490295$
$p_{valor} = 0.0004887$

- Teste de normalidade de erros:

```
> shapiro.test(res)
Shapiro-Wilk normality test
data:  res
W = 0.9705, p-value = 0.846
```

- Teste de homogeneidade de variâncias:

```
> bartlett.test(resist~trat)
Bartlett test of homogeneity of variances
data:  resist by trat
Bartlett's K-squared = 0.7116, df = 3, p-value = 0.8705
```