ANOVA

- Para o modelo de efeitos (ANOVA):

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., n \qquad (1)$$

em que μ é a média geral, τ_i é a média ou efeito dos tratamentos e ε_{ij} componente de erro aleatório.

- As hipótese de interesse são definidas por:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0, & \text{(O efeito de tratamento é nulo)} \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

- Ou de forma equivalente:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = ... = \mu_a \\ H_1 : \exists \mu_i \neq \mu_j,\ i \neq j, \end{cases}$$

em que $\mu_i = \mu + \tau_i$.

Anova

- Só é possível realizar a análise de variância se certas condições, ou seja, certas exigências do modelo matemático forem satisfeitas:
    - Os erros devem ter distribuição normal;
    - Os erros devem ser independentes;
    - Os erros deve ter a mesma variância, ou seja, deve existir homocedasticidade.

- Pode-se representar essas condições por:

$$\varepsilon_{ij} \sim N(0, \sigma^2). \tag{2}$$

ANOVA

$$E(QM_{Res}) = E\left(\frac{SQ_{Res}}{N-a}\right) = \frac{1}{N-a}E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y_{i.}})^2\right] = \sigma^2$$

e

$$E(QM_{Trat}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

- Ou seja, $QM_{Res} = SQ_{Res}/(N-a)$ estima σ^2;
- E $QM_{Trat} = SQ_{Trat}/(a-1)$ também estima σ^2, caso não há diferença entre as médias dos tratamentos (o que implica em $\tau_i = 0$).

ANOVA

- Como os graus de liberdade de SQ_{Trat} e SQ_{Res} somam $N-1$, que é o número de graus de liberdade da SQ_T, o Teorema de Cochran estabelece que SQ_{Trat}/σ^2 e SQ_{Res}/σ^2 são variáveis aleatórias independentes.

- Portanto, se a hipótese nula de nenhuma diferença nas médias de tratamento for verdadeira, a razão

$$F_0 = \frac{SQ_{Trat}/(a-1)}{SQ_{Res}/(N-a)} = \frac{QM_{Trat}}{QM_{Res}}, \tag{3}$$

  segue distribuição F com $a-1$ e $N-a$ graus de liberdade.

- A equação (3) é a estatística do teste para a hipótese de não haver diferenças nas médias de tratamento.

ANOVA

- A partir da esperança dos quadrados médios, verifica-se que, em geral, QM_{Res} é um estimador imparcial de σ^2;

- Além disso, sob a hipótese nula ser verdadeira, QM_{Trat} também é um estimador imparcial de σ^2;

- No entanto, se a hipótese nula for falsa, o valor esperado de QM_{Trat} é maior que σ^2;

- Portanto, sob a hipótese alternativa, o valor esperado do numerador da estatística de teste (3) é maior que o valor esperado do denominador, e deve-se rejeitar H_0 para valores da estatística do teste que são muito grandes;

- Portanto, deve-se rejeitar H_0 e concluir que existem diferenças nas médias de tratamento se $F_0 > F_{\alpha,(a-1),(N-a)}$.

ANOVA

Table: Tabela de Análise de Variância

Fonte de Variação	SQ	g.l.	QM	F
Tratamentos	SQ_{Trat}	a - 1	QM_{Trat}	QM_{Trat}/QM_{Res}
Resíduo	SQ_{Res}	N - a	QM_{Res}	
Total	SQ_T	N - 1		

em que $N = an$.

Diagnóstico

- É fundamental verificar a adequabilidade do modelo antes de tirar conclusões sobre o processo inferencial.

- Verificar as características do modelo.

- Para avaliar a adequabilidade do modelo serão usados:
  - Métodos gráficos
  - Testes estatísticos

Estimação dos Parâmetros do Modelo

- Após verificar a qualidade de ajuste do modelo, o próximo passo é estimar os parâmetos do modelo;
- Ao considerar o modelo de efeitos:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, ..., a; \quad j = 1, 2, ..., n$$

  em que μ é a média geral, τ_i é a média ou efeito dos tratamentos e ε_{ij} componente de erro aleatório.
- Os parâmetros a serem estimados são: μ e τ_i;
- O método de estimação de Mínimos Quadrados será utilizado.

Estimação dos Parâmetros do Modelo

- Para encontrar os estimadores de mínimos quadrados de μ e τ_i, é necessário escrever a soma dos quadrados dos erros:

$$L = \sum_{i=1}^{a} \sum_{j=1}^{n} \varepsilon_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \mu - \tau_i)^2 \tag{4}$$

e os valores de μ e τ_i que minimizam a equação (4) são os estimadores de mínimos quadrados, $\hat{\mu}$ e $\hat{\tau}_i$.

- Os valores apropriados seriam as soluções para as $a + 1$ equações simultâneas:

$$\frac{\partial L}{\partial \mu}|\hat{\mu}, \hat{\tau}_i = 0$$

e

$$\frac{\partial L}{\partial \tau_i}|\hat{\mu}, \hat{\tau}_i = 0 \quad i = 1, 2, ..., a.$$

Estimação dos Parâmetros do Modelo

- Derivando a equação (4) em relação a μ e τ_i e igualando a zero, tem-se que:

$$-2\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0$$

e

$$-2\sum_{j=1}^{n}(y_{ij} + \hat{\mu} - \hat{\tau}_i) = 0 \quad i = 1, 2, ..., a.$$

- Ao aplicar a restrição: $\sum_{i=1}^{a} \hat{\tau}_i = 0$, tem-se a seguinte solução para o sistema de equações normais:

$$\hat{\mu} = \bar{y}_{..}$$

e

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

Estimação dos Parâmetros do Modelo

- Um intervalo de confiança para a média do i-ésimo tratamento pode ser facilmente determinada;
- A média do i-ésimo tratamento é

$$\mu_i = \mu + \tau_i$$

- Então, um estimador pontual para μ_i é dado por:

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$$

- Ao assumir que a distribuição dos erro é normal, a média de cada tratamento, $\bar{y}_{i.}$ tem distribuição $N(\mu_i, \sigma^2/n)$;
- Se σ^2 fosse conhecido, a distribuição normal seria usada para construir o intervalo de confiança;

Estimação dos Parâmetros do Modelo

- Ao usar o QM_{Res} como estimador de σ^2, deve-se usar a distribuição t-Student para obter o intervalo de confiança;
- Sendo assim, o intervalo $100(1-\alpha)\%$ para a i-ésima média de tratamento, μ_i, é definido por:

$$\bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{QM_{Res}}{n}} \leq \mu_i \leq \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{QM_{Res}}{n}} \quad (5)$$

- Quando o pesquisador quer construir o intervalo de confiança para várias médias de tratamento, deve-se fazer uma correção no nível de confiança usando o método de Bonferroni;
- Nesse caso deve-se usar $\alpha/(2r)$ na equação (5), em que r é a quantidade de intervalos de confiança simultâneos.

Coeficiente de Determinação

- A medida R^2 é chamada de **Coeficiente de Determinação** e representa a proporção da variação total explicada pelo modelo ANOVA e é definida por:.

$$R^2 = \frac{SQ_{Modelo}}{SQ_{Total}}$$

- Desde de que $0 \leq SQ_{Modelo} \leq SQ_{Total}$, segue que

$$0 \leq R^2 \leq 1$$

- Valores grandes de R^2 indicam que a variação total é mais reduzida/explicada pelo efeito de tratamento.

Comparações de médias

- Ao rejeitar a hipótese nula: $H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$ ou $H_0 : \mu_1 = \mu_2 = ... = \mu_a$

- A pergunta a ser feita é: Qual(is) efeito(s) de tratamento é(são) diferente(s) de zero?

CONTRASTES

Dada uma função linear da forma

$$y = f(x) = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

e se

$$\sum_{i=1}^{n} = a_1 + a_2 + \ldots + a_n = 0$$

Diz-se então, que y é um contraste na variável x. Se x for uma média, tem-se um contraste de médias.

Contrastes

Exemplos:

$$\hat{y}_1 = \hat{\mu}_1 - \hat{\mu}_2$$

$$\hat{y}_2 = (\hat{\mu}_1 + \hat{\mu}_2) - (\hat{\mu}_3 + \hat{\mu}_4)$$

$\hat{y}_1$ e $\hat{y}_2$ são estimativas de contrastes

Em geral, um contraste é uma combinação linear de parâmetros na forma

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

em que $\sum_{i=1}^{a} c_i = c_1 + c_2 + \ldots + c_a = 0$

Variância de um Contrastes

A variância de um contraste é estimada por

$$Var(\hat{y}) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2$$

em que n é o número de repetições e σ^2 é estimada para QM_{Res}.

Contrastes Ortogonais

- Dois contrastes são ditos ortogonais se a variação de um contraste é independente da variação do outro, ou seja, $Cov(\hat{y_1}, \hat{y_2}) = 0$.

- Condição 1: Três ou mais contrastes são ortogonais entre si se eles forem ortogonais dois a dois.

- Condição 2: em um experimento, existem $a - 1$ contrastes possíveis entre os a tratamentos

Contrastes Ortogonais

- Dois contrastes são ortogonais entre si quando a soma algébrica dos produtos dos coeficientes das médias correspondentes é nula.

$$C_1 = c_1\mu_1 + c_2\mu_2 + \ldots + c_a\mu_a \quad \sum_{i=1}^{a} c_i = 0$$

$$C_2 = d_1\mu_1 + d_2\mu_2 + \ldots + d_a\mu_a \quad \sum_{i=1}^{a} d_i = 0$$

- C_1 e C_2 são ortogonais entre si se:

$$\sum_{i=1}^{a} c_i d_i = 0$$

Contrastes Ortogonais

Sejam μ_1, μ_2 e μ_3 médias de três tratamentos. Verifique que os contrastes

$$C_1 = \mu_1 - \mu_2$$

$$C_2 = \mu_1 + \mu_2 - 2\mu_3$$

são ortogonais entre si.

- Três ou mais contrastes são ortogonais entre si ou mutuamente ortogonais se eles forem ortogonais dois a dois.

Comparações de médias

Alguns teste podem ser utilizados para tal objetivo. São eles:

- Teste de Fisher;
- Teste de Tukey;
- Teste de Duncan;
- Teste de Scheffé.

Teste de Tukey

- Suponha que, após rejeitar a hipótese nula de igualdade de tratamento na ANOVA, deseja-se testar todas as comparações de médias:

$$H_0 : \mu_i = \mu_j$$
$$H_1 : \mu_i \neq \mu_j$$

  para todo $i \neq j$.

- O procedimento de Tukey controla o erro experimental no nível selecionado. Este é um excelente procedimento de espionagem de dados quando o interesse está em pares de médias (não permite comparar grupos entre si);

Teste de Tukey

- O teste de Tukey considera a seguinte estatística de intervalo estudentizado:
$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{QM_{Res}/n}},$$
em que $\bar{y}_{max}$ e $\bar{y}_{min}$ são a maior e menor média, respectivamente, para o grupo de p amostras de médias.
- O valor de q deve ser comparado com valores de $q(p, f)$, em que:
- $q()$ é o percentil da estatística de intervalo estudentizado (valor tabelado);
- e f é o número de graus de liberdade associados ao QM_{Res};

Teste de Tukey

- Para amostras de mesmo tamanho, o teste de Tukey declara duas médias significativamente diferentes se o valor absoluto de suas diferenças amostrais excedem:

$$\Delta_\alpha = q_\alpha(a, f)\sqrt{\frac{QM_{Res}}{n}}. \tag{6}$$

  Δ Diferença Mínima Significativa (DMS)

- De forma equivalente, pode-se construir um conjunto de $100(1-\alpha)\%$ intervalos de confiança para todos os pares de médias, ou seja:

$$\bar{y_i}.-\bar{y_j}.-q_\alpha(a, f)\sqrt{\frac{QM_{Res}}{n}} \leq \mu_i-\mu_j \leq \bar{y_i}.-\bar{y_j}.+q_\alpha(a, f)\sqrt{\frac{QM_{Res}}{n}}$$

Exemplo - Teste de Tukey

Uma engenheira está interessada em investigar a relação entre a configuração de potência de rádio frequência (RF) e a taxa de gravação para esta ferramenta. O objetivo de um experimento como este é modelar a relação entre a taxa de gravação e a potência de RF e especificar a configuração de potência que dará uma taxa de gravação desejada.

Exemplo - Teste de Tukey

Ela está interessada em um determinado gás (C_2F_6) e uma abertura de 0,80cm para testar quatro níveis de potência de RF: 160, 180, 200 e 220 W. Ela decidiu testar cinco placas em cada nível de potência de RF.

Suponha que a engenheira execute o experimento de forma aleatória.

Exemplo - Teste de Tukey

As observações que ela obteve sobre a taxa de gravação são mostradas na Tabela 1.

Tabela 1 : Dados de taxa de gravação (em A/min) do experimento de gravação com plasma

	Observações					
Potência	1	2	3	4	5	Total
160	575	542	530	539	570	2756
180	565	593	590	579	610	2937
200	600	651	610	637	629	3127
220	725	700	715	685	710	3535

Exemplo - Teste de Tukey

- Considere o experimento da gravação de plasma. Como a hipótese nula foi rejeitada, sabemos que algumas configurações de energia produzem taxas de gravação diferentes das outras, mas quais realmente causam essa diferença?
- Podemos suspeitar no início do experimento que $200W$ e $220W$ produzem o mesma taxa de gravação, o que implica que gostaríamos de testar a hipótese:

$$H_0 : \mu_3 = \mu_4$$
$$H_1 : \mu_3 \neq \mu_4$$

ou equivalentemente:

$$H_0 : \mu_3 - \mu_4 = 0$$
$$H_1 : \mu_3 - \mu_4 \neq 0$$

Exemplo - Teste de Tukey

No exemplo temos:

>summary(anovap)

	Df	Sum Sq	Mean Sq	F value	Pr(¿F)
potencia	3	66871	22290	66.8	2.88e-09 ***
Residuals	16	5339	334		

>mediatrat
160 180 200 220
551.2 587.4 625.4 707.0

$\Delta(5\%) = 4,05\sqrt{\frac{334}{5}} = 33,10$
Se o contraste for maior do que Δ então as médias diferem ao nível α de significância.

Exemplo - Teste de Tukey

- Como o teste de Tukey é de certa forma independente do teste F, é possível que,mesmo sendo significativo o valor de $F_{calculado}$, não se encontre diferenças significativas entre contrastes de médias.
- Utilizaremos o método das letras para exemplificar o uso do teste.
- Inicialmente ordena-se as médias de forma crescente ou decrescente para facilitar as comparações. Coloca-se uma letra do alfabeto na primeira média e em seguida compara-se a diferença¸com as médias seguintes. Se a diferença¸for superior ao valor de Δ a diferença entre duas médias será considerada significativa. Médias sem a mesma letra diferem

Exemplo - Teste de Tukey

160	551,2	a
180	587,4	b
200	625,4	c
220	707,0	d