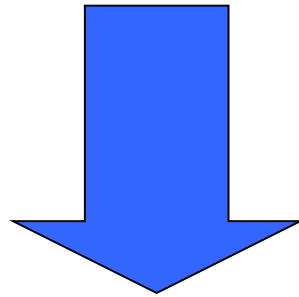


TESTES DE SIGNIFICÂNCIA

Conforme já discutimos, qualquer estatística obtida a partir de uma amostra ou de um experimento, provavelmente não terá o mesmo valor em outro experimento.



Por mais que sejam utilizados **os mesmos tratamentos**, será impossível obter condições ambientais e de manejo, idênticas ao que foi realizado no experimento anterior, pois há variações impossíveis de serem controladas.

A estatística F calculada através da análise de variância é um tipo de teste de significância ou teste de hipóteses muito utilizado.

VAMOS ENTENDER UM POQUINHO MAIS SOBRE O TESTE F....

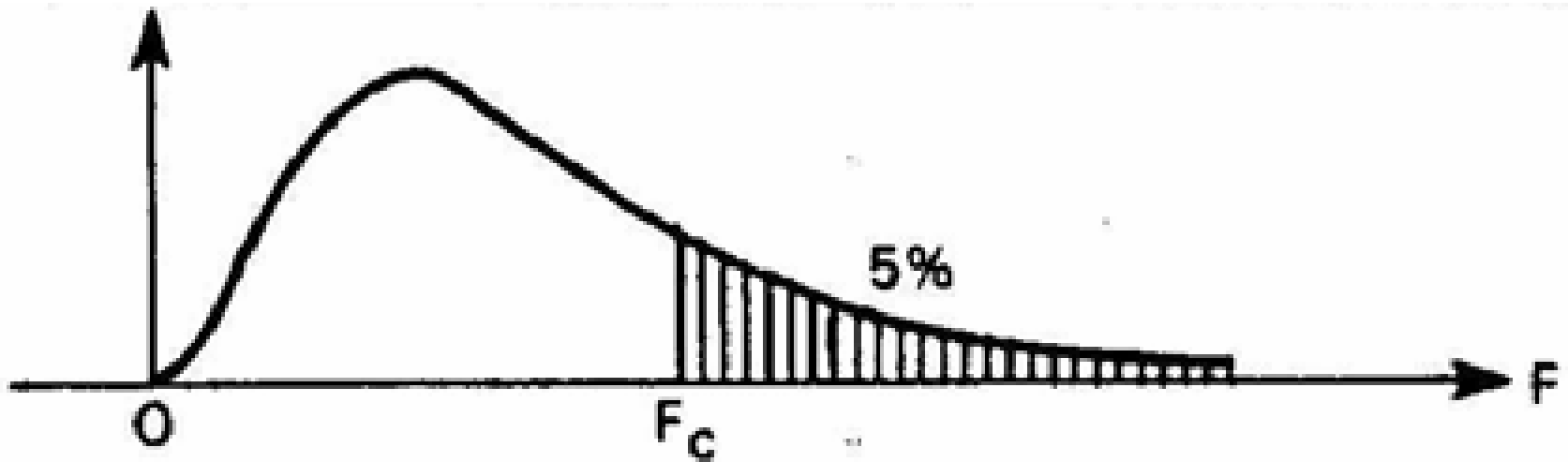
Quando estivermos testando “n” tratamentos....e existir diferença entre eles, ou seja existir variabilidade o valor F calculado no teste será maior que ZERO.

A estatística F é calculada através da razão $QM_{Tratamento}/QM_{Resíduo}$.

Quanto maior for a estatística F , menor a chance desse valor acontecer sob a hipótese de um modelo de efeito nulo (igualdade de tratamentos).

Ou seja, o valor de **F deve ser grande** (maior que o F_c tabelado, em nível de significância) para rejeitar H_0 e concluir pela diferença significativa entre algum contraste de médias de tratamentos.

Quando isso ocorre ($F \geq F_c$) dizemos que o valor da estatística “caiu” na região crítica ou de rejeição da hipótese de nulo efeito do modelo. Na Figura a seguir podemos observar a região crítica para a distribuição F de Snedecor, para nível de significância ($\alpha = 0,05$) ou 5%. A região crítica é formada por valores maiores que F_c .

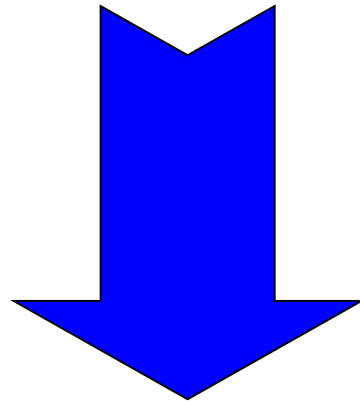


No caso da análise de variância, rejeitar H_0 significa acreditar que o modelo nulo não é correto. **Ou seja, existe diferença entre os tratamentos**

Ainda existe a possibilidade que o modelo nulo seja verdadeiro. Essa possibilidade é no máximo igual ao tamanho da região crítica.

Na agricultura, é usual trabalhar com $\alpha = 5\%$, (ou seja, desde que a hipótese nula seja rejeitada, a chance de erro será de no máximo 5%).

Como vimos na análise de variância, a estatística F é usada para testar se o modelo responde ou não por porção significativa da variação existente.



Caso seja detectada a diferença entre os tratamentos, a significância do valor F implica que nem todas médias são iguais.

Mas então vêm as perguntas:

Qual é ou são a(s) diferentes(s)?

Qual é o melhor tratamento?

E o pior?

Há várias técnicas para responder essas perguntas. Pode-se utilizar intervalos de confiança para cada média ou utilizar os PPCM (procedimentos para comparações múltiplas), que serão discutidos a seguir.

O intervalo de confiança [IC] é :

$$[\text{media} - t \cdot \text{epm} ; \text{media} + t \cdot \text{epm}]$$

é um valor tabelado  

é o erro padrão das médias.

Vamos ilustrar um caso simples, comparando
2 tratamentos.

Supondo :

Média A (M_a) = 20

Média B (M_b) = 24

$$QM_{\text{Resíduo}} = 2$$

$$r = n^{\circ} \text{ de repetições} = 4$$

Para IC de 95%: $t = 2,44$, resultando: $t \cdot \text{epm} = 1,73$

$$GL_{\text{resíduo}} = 6$$

$$\text{epm} = \sqrt{QM_{\text{Resíduo}}} / \sqrt{r} = 0,71$$

Como os IC de cada média não se interceptam, conclui-se que as médias são diferentes.

COM MUITAS MÉDIAS ESSE PROCEDIMENTO É POUCO PRÁTICO, DAÍ USAR PPCM.

PPCM (procedimentos para comparações múltiplas)

A regra é aparentemente simples.

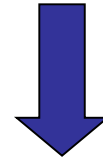
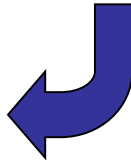
Consiste em estabelecer uma quantidade mínima significativa, denominada **DMS** (diferença mínima significativa).

Se a diferença entre duas médias (em módulo) superar a DMS, a conclusão é que as duas médias diferem entre si.

No geral usa-se

$$\text{DMS} = q \cdot \text{epm}$$

é um valor tabelado



é o erro padrão das médias.

Vamos ilustrar um caso simples, comparando
2 tratamentos.

Supondo:

Média A (M_a) = 20

Média B (M_b) = 24 (diferença observada entre as médias é igual a 4)

$$QM_{\text{Resíduo}} = 2$$

$$r = n^{\circ} \text{ de repetições} = 4$$

$$GL_{\text{resíduo}} = 6$$

$$epm = \sqrt{QM_{\text{Resíduo}}} / \sqrt{r} = 0,71$$

$q_2 = 3,46$ (valor tabelado para 2 médias e 6 GL Resíduo)

Então, a DMS mínima, também conhecida como LSD (least difference significance) é:

$$\text{LSD} = q_2 \cdot \text{epm} = 3,46 \cdot 0,71 = 2,76$$

Portanto como a diferença observada (4) é maior que a LSD (2,46), as médias são estatisticamente diferentes.

A LSD (menor DMS) é exata para experimentos de comparação entre duas médias. Para comparação de mais médias é uma medida “frouxa”, pois podem ser detectadas comparações significativas, que de fato não são. As comparações usando LSD é o chamado teste t ou teste de Student.

Agora, um exemplo com t tratamentos.

Supondo 5 tratamentos temos:

$$QM_{\text{Resíduo}} = 4$$

$$r = \text{n}^{\circ} \text{ de repetições} = 4$$

$$GL_{\text{resíduo}} = 15$$

$$epm = \sqrt{QM_{\text{Resíduo}}} / \sqrt{r} = 1$$

Na Tabela a seguir temos apresentadas as médias dos 5 tratamentos e o ranking de classificação (ordenamento) das médias dos tratamentos:

Médias	Valores	Ranking das Médias
Ma	1	5 ^o
Mb	7	4 ^o
Mc	12	1 ^o
Md	8	3 ^o
Me	9	2 ^o

Para esse caso poderíamos ainda usar a LSD , ou teste t.

Mas, é comum usar a DMS máxima, também conhecida como HSD (honestly significance difference) que é o teste Tukey, calculado com:

$q_5 = 4,37$ (valor tabelado- t médias, no caso 5 e 15 GL Resíduo)

$$HSD = q_5 \cdot e_{pm} = 4,37 \cdot 1 = 4,37$$

$$M_c - M_a = 12 - 1 = 11, \text{ assim } 11 > 4,37$$

Portanto como a diferença entre a maior e a menor média (amplitude) é maior que a HSD, podemos concluir que existe diferença entre essas duas médias.

A HSD (maior DMS) é exata para comparar diferenças entre a maior e a menor média de um grupo de tratamentos. Para outras comparações é uma medida conservadora, pois pode não detectar diferenças significativas que de fato existem. Na prática, é muito eficiente para eliminar as piores médias.

Usualmente apresentamos os resultados do teste de Tukey em Tabelas, onde as médias são ordenadas de forma decrescente, de forma que **letras iguais indicam médias não significativamente diferentes** e **letras diferentes indicam diferenças significativas**, ao nível α de significância estabelecido na escolha do valor tabelado q_t .

A seguir, um exemplo da **aplicação do teste Tukey** em um experimento de competição de progênie de *Eucalyptus saligna*, instalado em DIC, com 4 repetições.

Vamos supor a seguinte análise de variância para os dados deste experimento

FV	GL	SQ	QM	F
Progênie (tratamentos)	9	160,98	17,89	6,91**
Resíduo	30	77,56	2,59	
Total	39	238,54		

Como:

$$s = \sqrt{QM_{\text{Resíduo}}} = \sqrt{2,59} = 1,61$$

$$q(t; GL_{\text{Resíduo}}) = q(10;30) = 4,82$$

$$\text{epm} = \sqrt{QM_{\text{Resíduo}}} / \sqrt{r} = 1,61 / \sqrt{4} = 0,805$$

Pelo teste de Tukey temos:

$$\text{HSD} = q_{10} \cdot \text{epm} = 4,82 \cdot 0,805 = 3,88$$

$$\text{Amplitude} = (\text{maior média}) - (\text{menor média}) = \\ 14,55 - 8,80 = 5,75$$

Como a amplitude observada ($= 5,75$) $>$ HSD ($=3,88$), podemos concluir que existe diferença significativa para esse contraste de médias (maior média – menor média).

Agora podemos organizar em ordem decrescente as médias dos tratamentos, para maior facilidade de comparação, e usar a mesma HSD.

Os resultados podem ser resumidos pelas letras colocadas ao lado das médias. Letras iguais indicam diferenças não significativas, ao nível de significância dado pela HSD, e letras diferentes indicam diferenças significativas a esse nível de probabilidade.

Progênies	Médias		
P1	14,55	A	
P2	14,30	A	
P3	14,22	A	
P4	14,12	A	
P5	13,00	A	
P6	12,42	A	B
P7	12,05	A	B
P8	11,67	A	B
P9	9,00		B
P10	8,80		B

CONCLUSÕES DO TESTE DE TUKEY APLICADO A ESSE EXPERIMENTO

No caso, entre as progênies de 1 a 5 não existem diferenças significativas para DAP.

As progênies de 1 a 8 possuem médias de DAP significativamente maiores que as progênies 9 e 10.

As progênies 6, 7 e 8 formam um grupo intermediário.

Além dos testes t (LSD) e Tukey (HSD), dezenas de outros PPCM, também são utilizados para tal fim. Entre eles:

Scott-knott, Duncan, Dunnett, SNK etc

Cada um deles apresenta vantagens e desvantagens, conforme pode ser estudado em literatura sobre o assunto. Há livros e centenas de artigos sobre isso.

Do ponto de vista do melhoramento de plantas, provavelmente o mais utilizado no mundo é o LSD, eventualmente com correção FDR (false discovery rate, Benjamini and Hochberg (1995)).