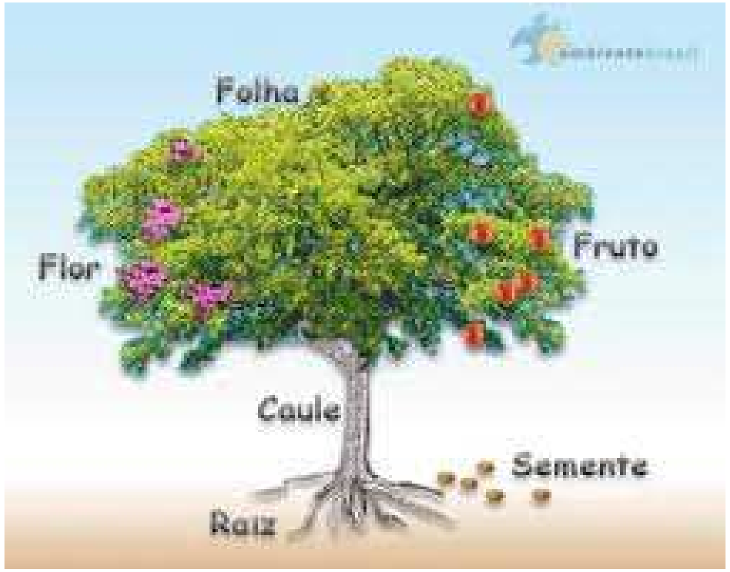
Delineamento Inteiramente Casualizado

Professora Ângela

Exemplo

- Um pesquisador pretende comparar 4 cultivares de pêssego quanto ao enraizamento de estacas.
- A área experimental disponibilizada para o experimento é um viveiro em condições controladas com capacidade para 20 parcelas.
- O pesquisador pode conseguir um máximo de 100 estacas por cultivar.
- Na literatura o número de estacas por parcela varia de 8 a 15.



Experimento Inteiramente Casualizado

- Sem a necessidade de controle local;
- Exemplo (Barbin, 2003):
  - Objetivo: Comparar 4 cultivares de pêssego quanto ao enraizamento de estacas;
  - Var. Resp.: número de estacas enraizadas por parcela;
  - Fator em potencial / Tratamentos: 4 cultivares de pêssego.
  - Fator de perturbação: Não tem, pois a área experimental é um viveiro em condições controladas.
  - Parcela: 20 estacas.
  - Área Exp.: Viveiro em condições controladas com capacidade para 20 parcelas.
  - Cada tratamento foi repetido 5 vezes;
  - Os tratamentos foram designados às parcelas por meio de sorteio.

Experimento Inteiramente Casualizado

P 1	P 5	P 9	P 13	P 17
P 2	P 6	P 10	P 14	P 18
P 3	P 7	P 11	P 15	P 19
P 4	P 8	P 12	P 16	P 20

Croqui da Área Experimental

Urna para sorteio dos tratamentos

- Cultivar 1
- Cultivar 2
- Cultivar 3
- Cultivar 4

Experimento Inteiramente Casualizado

Resultados:

Tratamentos	Repetições					Total
	1ª	2ª	3ª	4ª	5ª	
Var 1 (A)	2	2	1	1	0	6
Var 2 (B)	1	0	0	1	1	3
Var 3 (C)	12	10	14	17	11	64
Var 4 (D)	7	9	15	8	10	49
						122

Análise dos Dados

- Análise da Variância:
  - Nem sempre pode ser utilizada;
  - Só é indicada se o modelo matemático respeitar certas exigências.
- Modelo Matemático:
  - Representação simplificada da realidade;
  - Na experimentação é a representação da variável resposta levando em consideração os fatores envolvidos no experimento.

Exemplo: Variedades de Pêssego

Número de estacas, da variedade *i* (*i* = 1, ..., 4), fixadas na repetição *j* (*j* = 1, ..., 5),

$$\boxed{y_{ij}} = \boxed{m} + \boxed{t_i} + \boxed{e_{ij}}$$

Efeito do tratamento *i*, ou variedade *i*.

Média geral (número médio de estacas fixadas por P, independente de tratamento ou repetição)

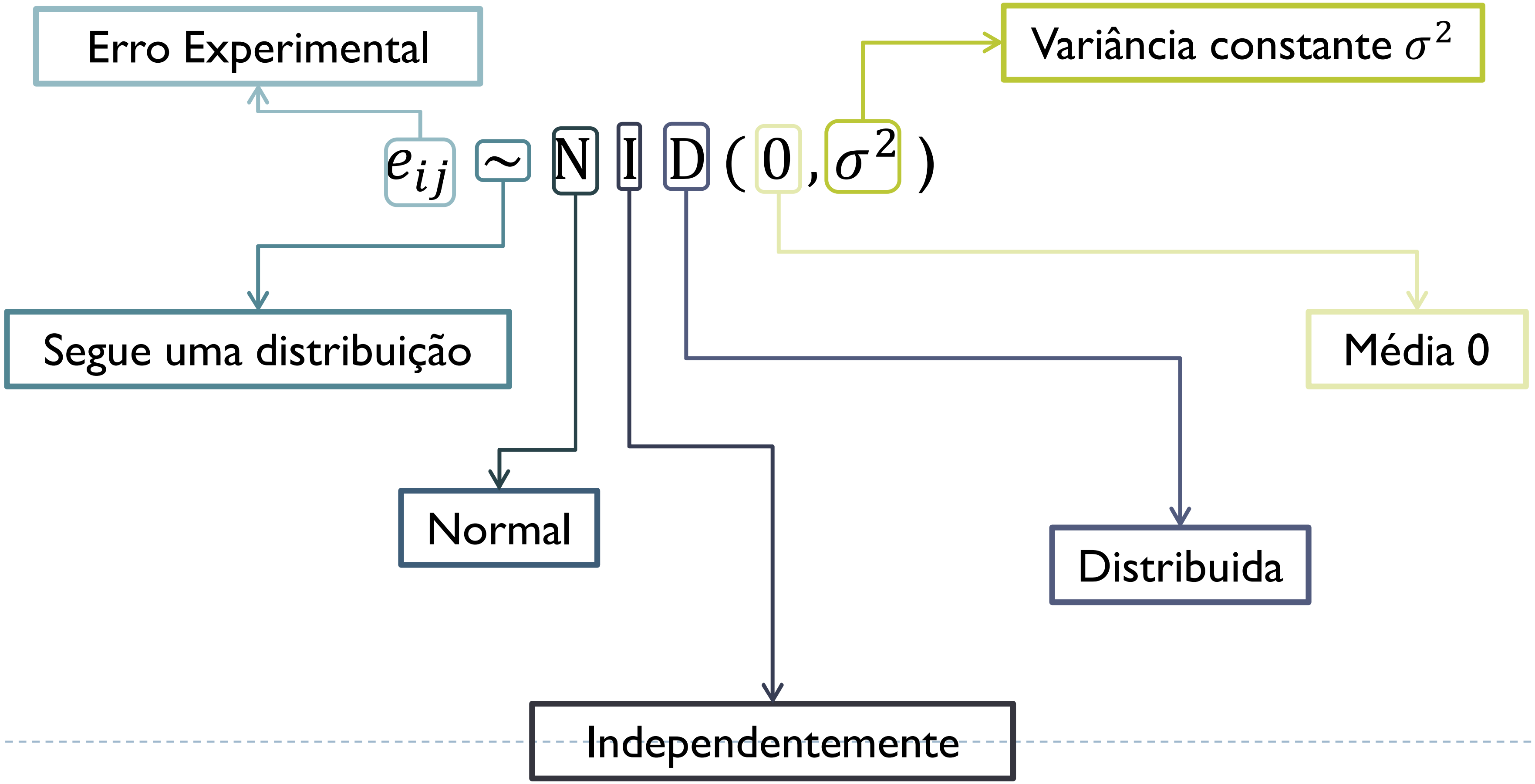
Erro aleatório ligado à variedade *i* na repetição *j*.

Pressuposições da Análise da Variância (ANOVA)

- O modelo deve ser aditivo;
- Os erros devem ter distribuição normal;
- Os erros devem ser independentes;
- Os erros devem ter a mesma variância (Homocedasticidade dos erros).

Pressuposições da Análise da Variância (ANOVA)

- É comum unir as 3 últimas exigências na seguinte expressão:



Verificação das Pressuposições

- Geralmente considera-se o modelo como aditivo por hipótese:
  - Teste de não aditividade de Tukey.
- Normalidade dos erros:
  - Teste de χ^2;
  - Teste de Lilliefors;
  - Teste de Shapiro Wilk;
  - Teste de Kolmogorov-Smirnov.
- Independência dos erros:
  - Princípio da casualização;
  - Verificação gráfica.
- Homocedasticidade das variâncias:
  - Teste de Hartley ou da razão máxima (Fmáx);
  - Teste de Bartlett;
  - Teste de Levene.

Aditividade do Modelo

- Modelos para experimentos são considerados aditivos pela sua simplicidade;
- Pode-se aplicar testes de não aditividade a modelos de mais de um fator ou que tenham um fator e algum tipo de controle local aplicado;
- Muitas vezes se o planejamento do experimento foi feito da maneira correta o modelo pode ser considerado como aditivo sem a necessidade de testes adicionais.

Análise de Resíduos

- Para aplicar um teste de normalidade dos erros, primeiro, é necessário obter as estimativas dos erros;
- Quanto mais simples o modelo matemático e o delineamento experimental, mais simples a obtenção das estimativas dos erros experimentais;
- A análise de resíduos permite a verificação da normalidade dos erros, homocedasticidade das variâncias e independência dos erros.

Como Obter os resíduos?

- Basta conhecer o modelo matemático:
- No caso do delineamento inteiramente casualizado, tem-se:
  - $y_{ij} = m + t_i + e_{ij}$, em que *i* representa o tratamento (de 1 a *I*) e *j* as repetições (de 1 a *J*)
- A estimativa da média geral ($\widehat{m}$) é dada por:
  - $\widehat{m} = \frac{G}{IJ}$, em que $G = \sum_{i,j} y_{ij}$.
- As estimativas dos efeitos de tratamento ($\hat{t}_i$) são dadas por:
  - $\hat{t}_i = \frac{T_i}{J} - \widehat{m}$, em que $T_i = \sum_j y_{ij}$.
- As estimativas dos erros são dadas por:
  - $\hat{e}_{ij} = y_{ij} - \widehat{m} - \hat{t}_i$.

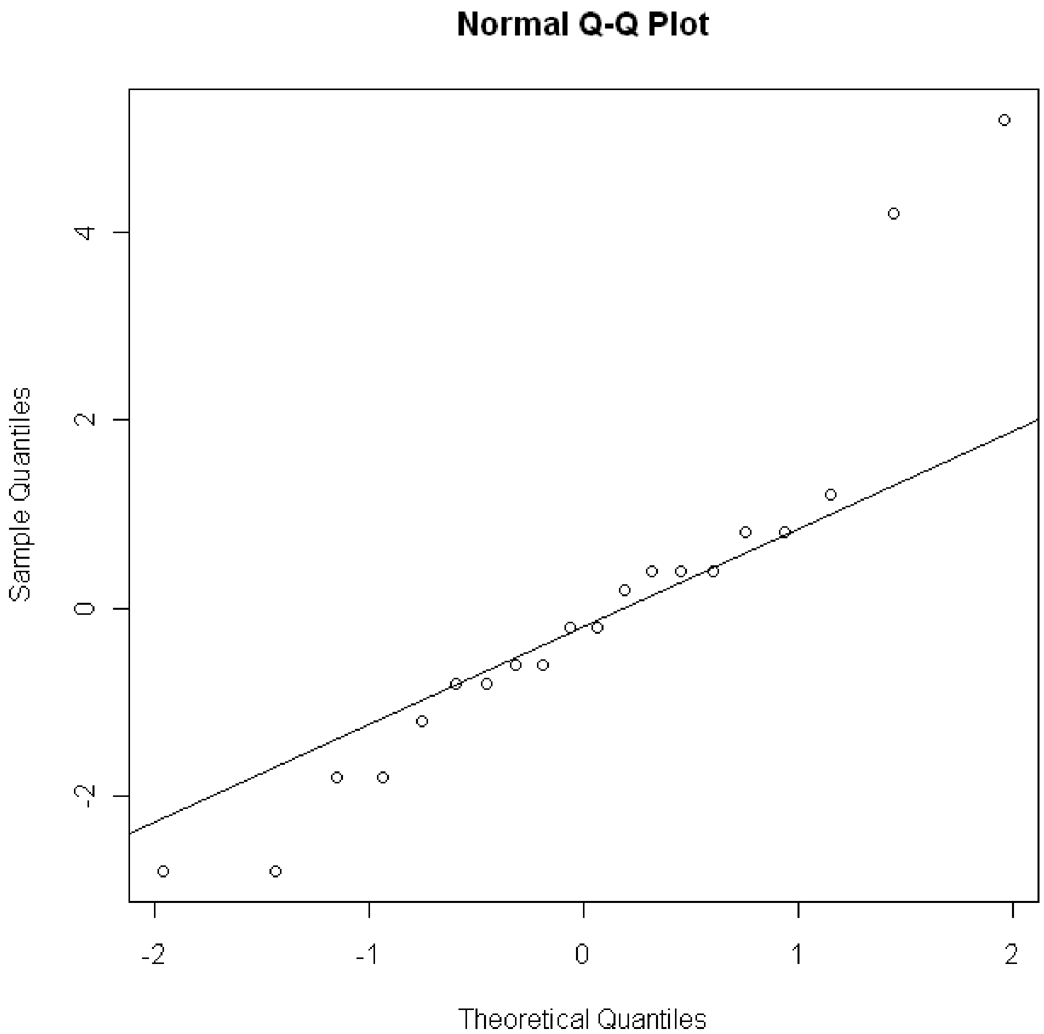
Análise de Resíduos

- Obtenção dos erros padronizados:

$$Z_i = \frac{X_i - \bar{X}}{s}$$

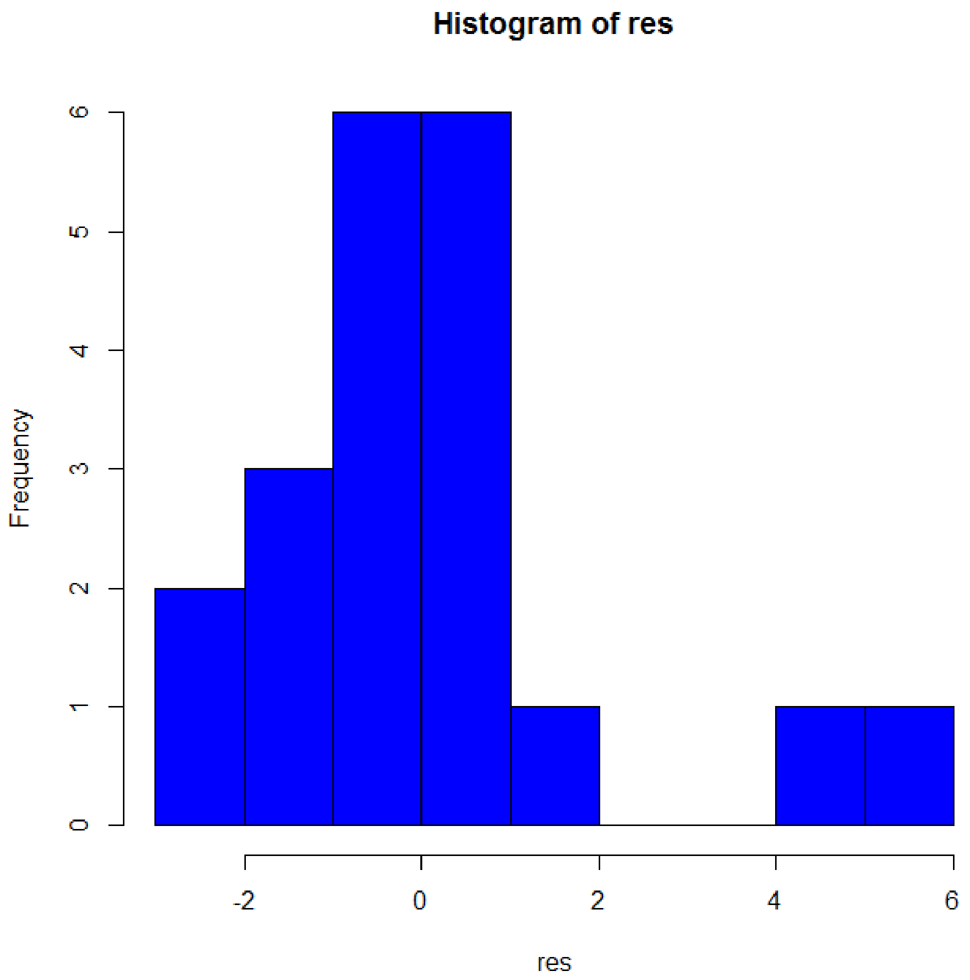
- no nosso caso:
  - $X_i = e_{ij}$;
  - $\bar{X} = 0$ e
  - s é a estimativa do desvio padrão médio dos erros:
    - $Z_i = \frac{e_{ij}}{\bar{s}} = d_{ij}$ = desvios padronizados;
    - $\bar{s} = \sqrt{\bar{s}^2}$;
    - $\bar{s}^2 = \frac{\sum e_{ij}^2}{I(J-1)} = \frac{\sum s_i^2}{I}$;

Q-Q Plot



O Q-Q Plot é um gráfico bastante útil na análise dos resíduos. Se os resíduos se posicionarem de maneira a formar uma reta (aproximada), tem-se evidência de normalidade dos mesmos, se não, tem-se evidência de falta de normalidade (como é o caso do exemplo dado).

Histograma



O histograma dos resíduos também é um bom indicador de normalidade dos dados. Se os resíduos forem normais, o seu histograma deve representar uma amostra retirada de uma distribuição normal com média zero. Novamente, o gráfico produzido com os dados do exemplo é um indicativo da falta de normalidade dos resíduos.

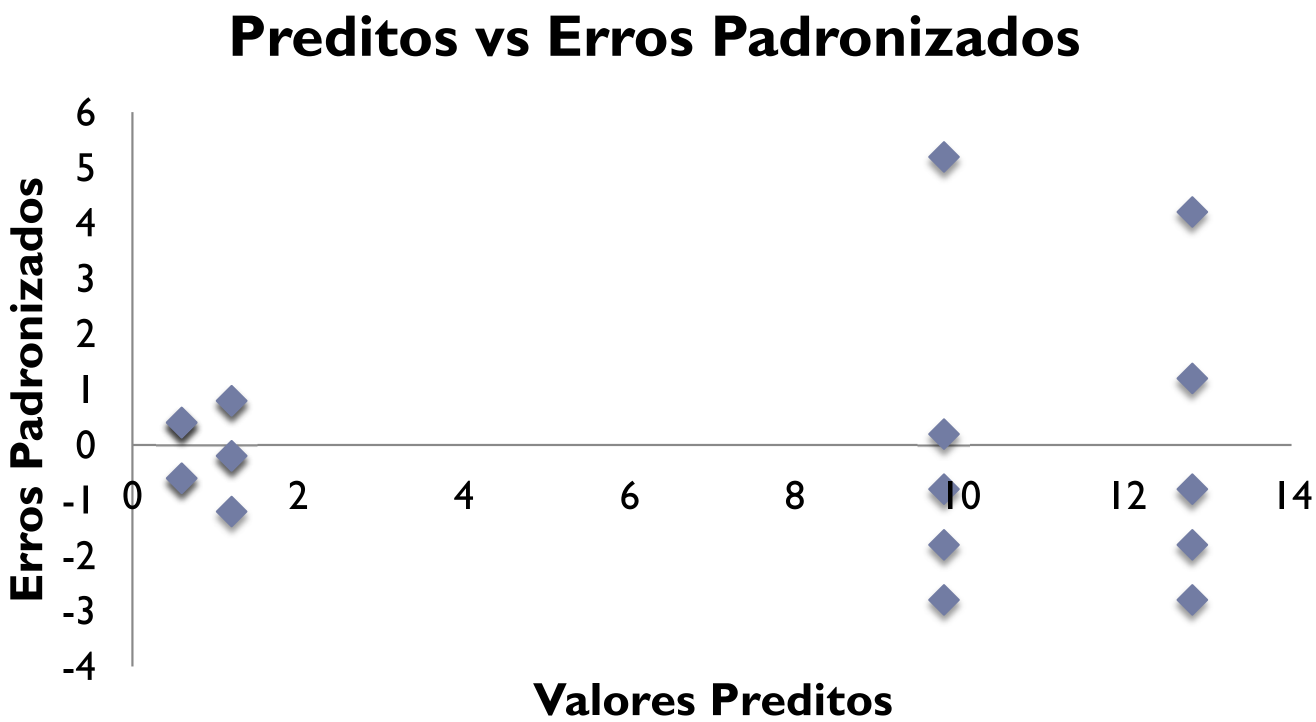
Testes de Normalidade

- Teste de Shapiro Wilk:
  - Teste mais poderoso e que melhor se adequa à maioria dos casos, porém, perde poder quando utilizado em amostras pequenas
- Teste de Kolmogorov-Smirnov:
  - Pode ser utilizado em amostras relativamente pequenas, porém, supõe-se que os parâmetros μ e σ^2 são conhecidos e pode ser utilizado, apenas, para dados contínuos.
- Teste de Lilliefors:
  - Uma adaptação do teste de Kolmogorov-Smirnov, que considera os parâmetros μ e σ^2 como desconhecidos e utiliza suas estimativas.
  - p-value = 0.05114

Testes de Homocedasticidade

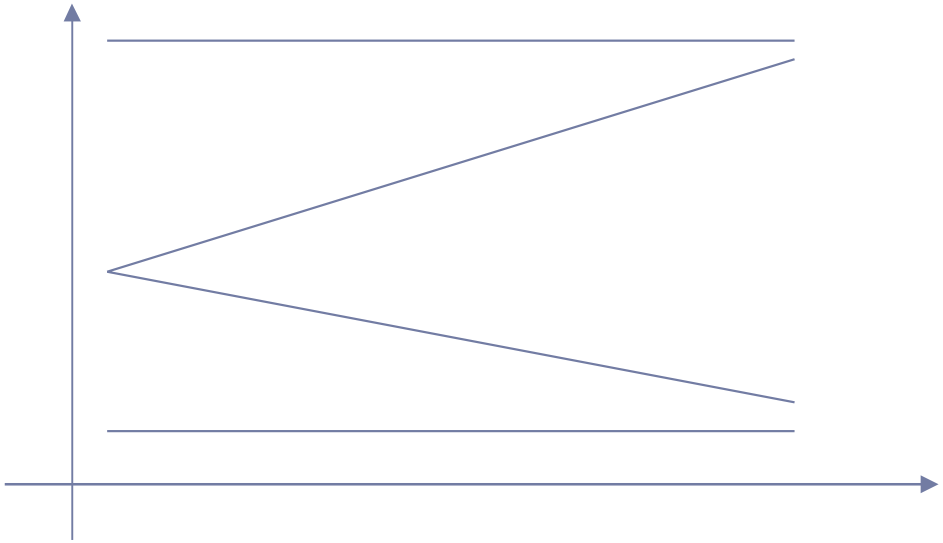
- Teste de Hartley (F máximo)
  - Indicado apenas quando o experimento for balanceado.
- Teste de Bartlett:
  - Deve ser utilizado, apenas, quando a suposição de normalidade for respeitada.
  - Pode ser utilizado em experimentos desbalanceados.
- Teste de Levene:
  - Pode ser utilizado quando a pressuposição de normalidade não for respeitada.

Visualização Gráfica

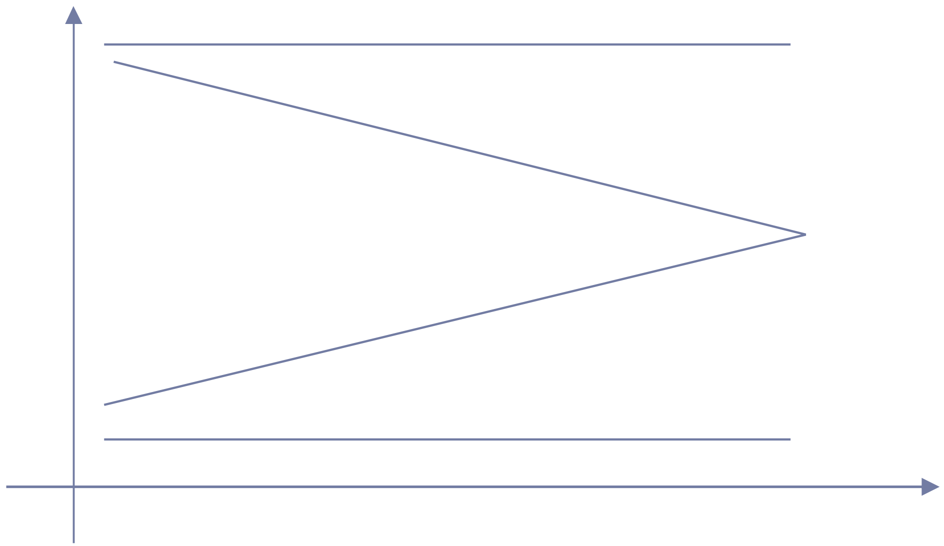


Indicativo de uma possível heterogeneidade das variâncias

Alguns padrões

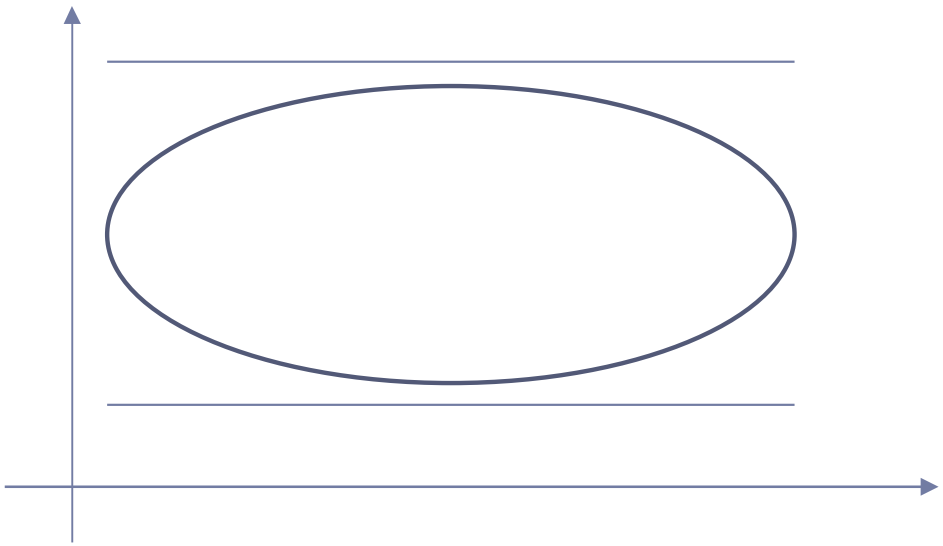


Variância aumenta conforme aumenta o valor da média de tratamento

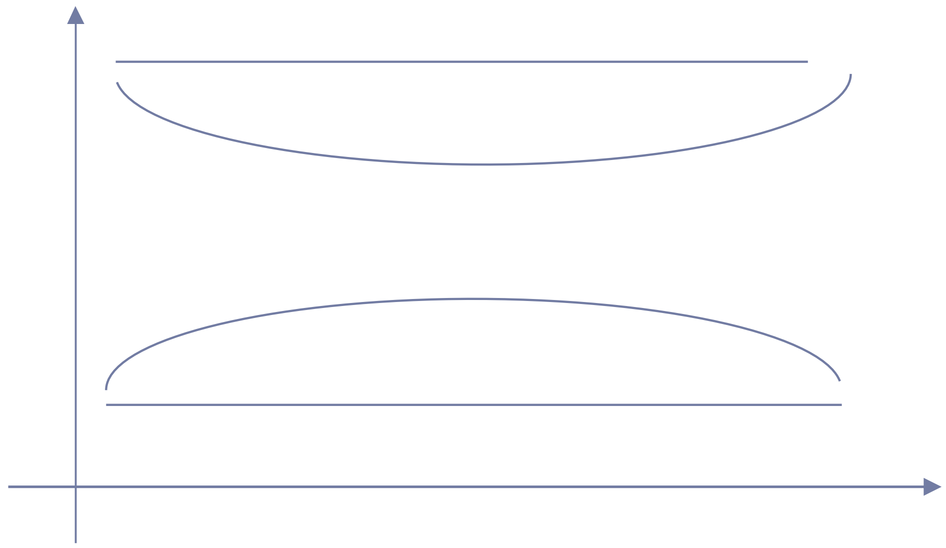


Variância diminui conforme aumenta o valor da média de tratamento

Alguns padrões

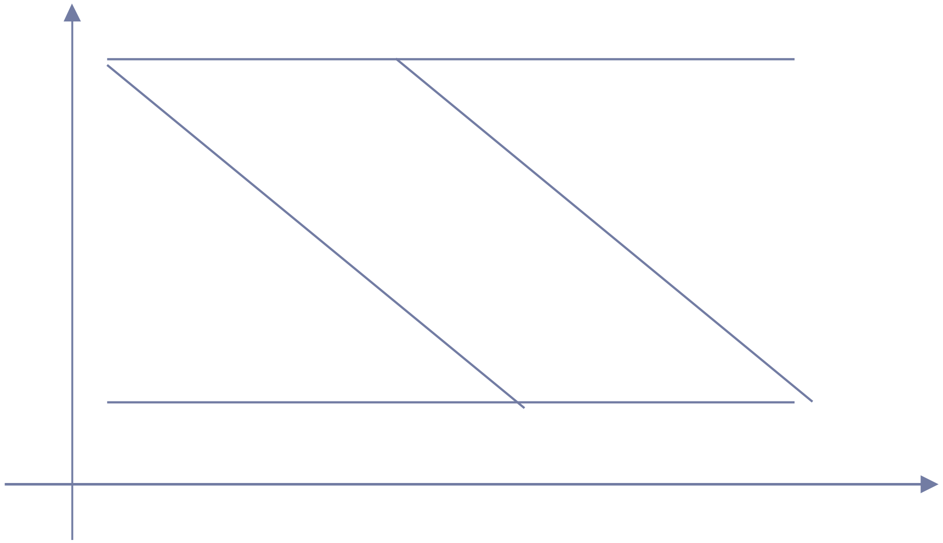


Variância aumenta conforme média de tratamento se aproxima da média geral

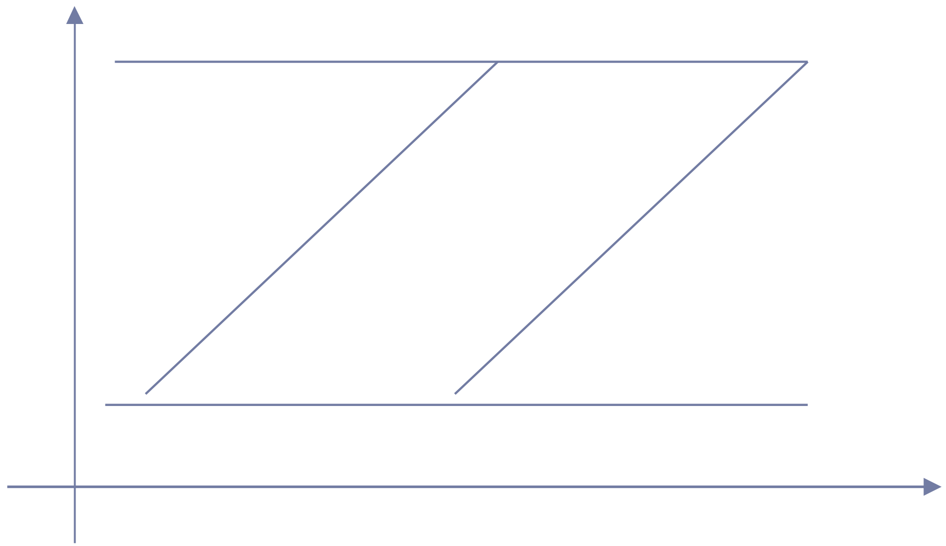


Variância aumenta conforme média de tratamento se afasta da média geral

Alguns padrões



Correlação negativa entre os erros e os valores preditos – indicativo de dependência dos erros



Correlação positiva entre os erros e os valores preditos – indicativo de dependência dos erros

Verificação das Pressuposições

- O experimento respeita 2 pressuposições: Aditividade do modelo (por hipótese); Independência dos erros (aleatorização).
- Porém duas pressuposições parecem não serem totalmente respeitadas: Normalidade dos erros (Lilliefors); Homocedasticidade das variâncias (Hartley).
- Os problemas de falta de Homocedasticidade e/ou Normalidade podem ser resolvidos (algumas vezes) com a transformação dos dados.
- Qual transformação usar?

Transformação de Dados

- Mais Comuns:
  - $\sqrt{x+k}$, com k sendo uma constante positiva, para dados de contagem;
  - $arc\ sen\sqrt{p/100}$, para dados de percentagem, geralmente para $0 \leq p \leq 30\%$ ou $70 \leq p \leq 100\%$;
  - $\log(x+k)$, quando hà proporcionalidade entre médias e desvios padrões.

$\hat{b}$	λ	Transformação Box Cox
0	1	Nenhuma
1	1/2	$\sqrt{x}$
2	0	$\log(x)$
3	-1/2	$1/\sqrt{x}$
4	-1	$1/x$

Transformação Box Cox

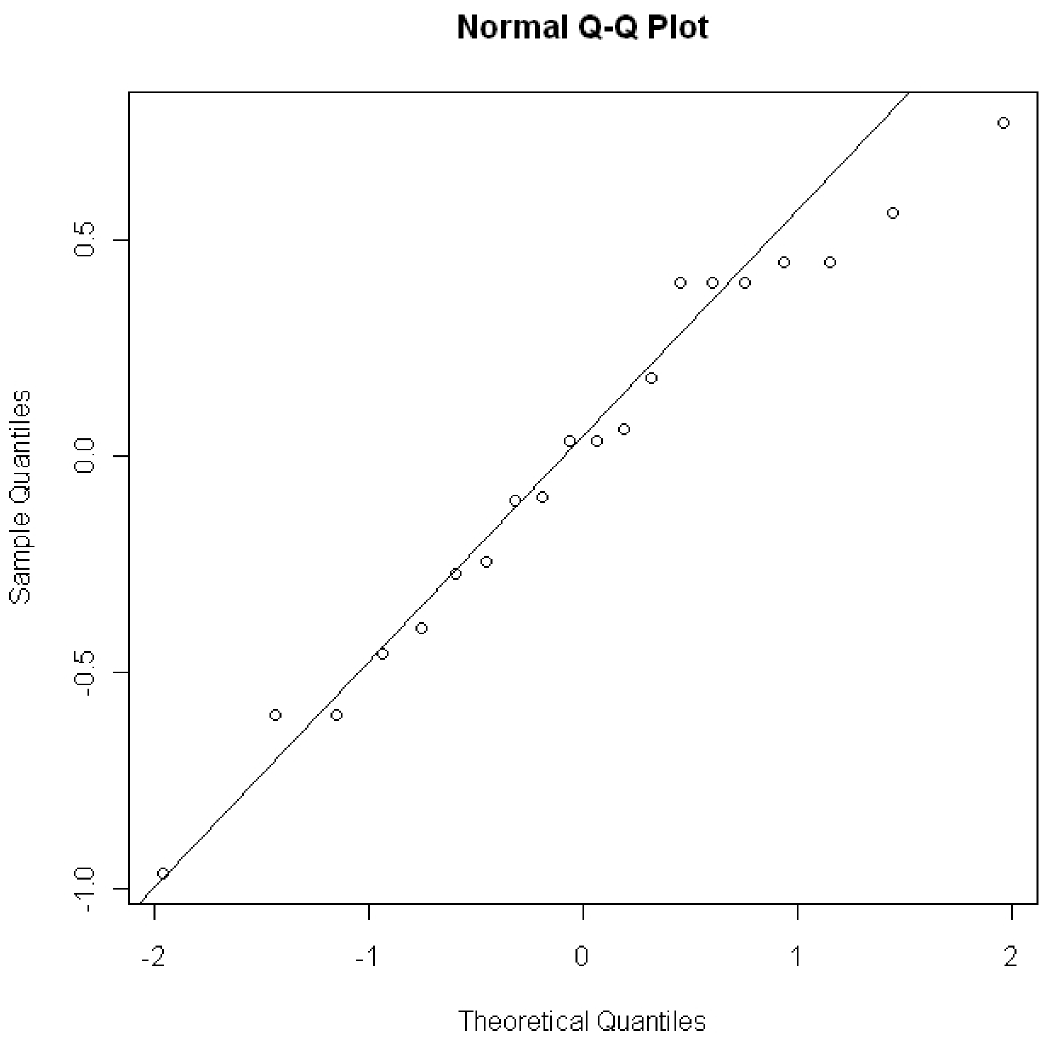
- Consiste em examinar possíveis relações entre os logarítimos das variâncias dos tratamentos com os logarítimos das médias dos mesmo utilizando uma regressão linear simples:
  - $\log({s_i}^2) = b \cdot \log(\widehat{m}_i) + c$;
- Deve-se estimar b, o coeficiente de inclinação da reta, e encontrar λ, em que:
  - $\lambda = 1 - {}^{\hat{b}}/_{2}$
- OBS:

$$\hat{b} = \frac{\sum_{i=1}^{I} log\left(s_i^2\right) log(\widehat{m}_i) - I\left(\sum_{i=1}^{I} \frac{log\left(s_i^2\right)}{I}\right)\left(\sum_{i=1}^{I} \frac{log(\widehat{m}_i)}{I}\right)}{\sum_{i=1}^{I}[log(\widehat{m}_i)]^2 - I\left(\sum_{i=1}^{I} \frac{log(\widehat{m}_i)}{I}\right)^2}$$

Alternativas à Transformação

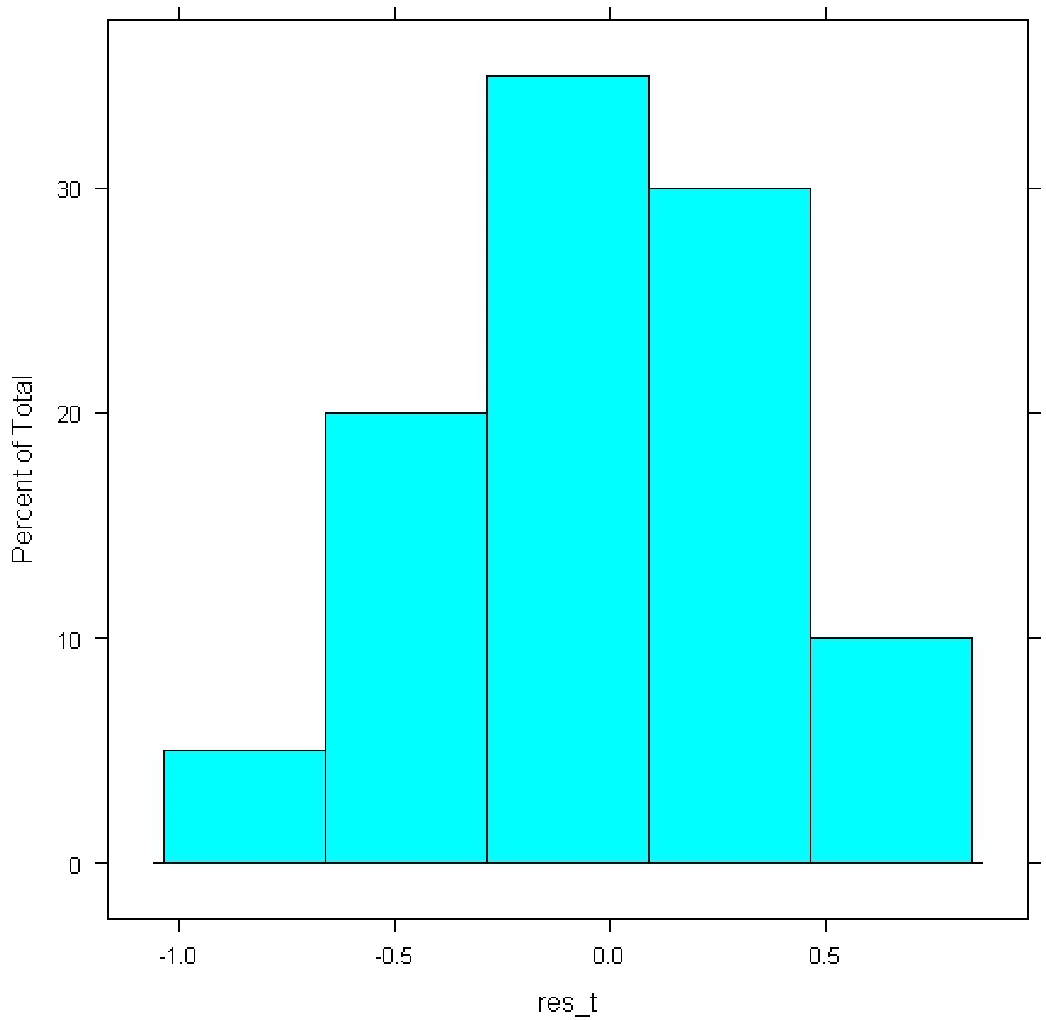
- Existem casos em que não é possível encontrar uma transformação que resolva todos os problemas e permita a utilização da técnica da ANOVA.
- Nesses casos, recomenda-se:
- Análises não paramétricas; ou
- Modelos Lineares Generalizados.

Q-Q Plot dos Resíduos – pós transformação dos dados



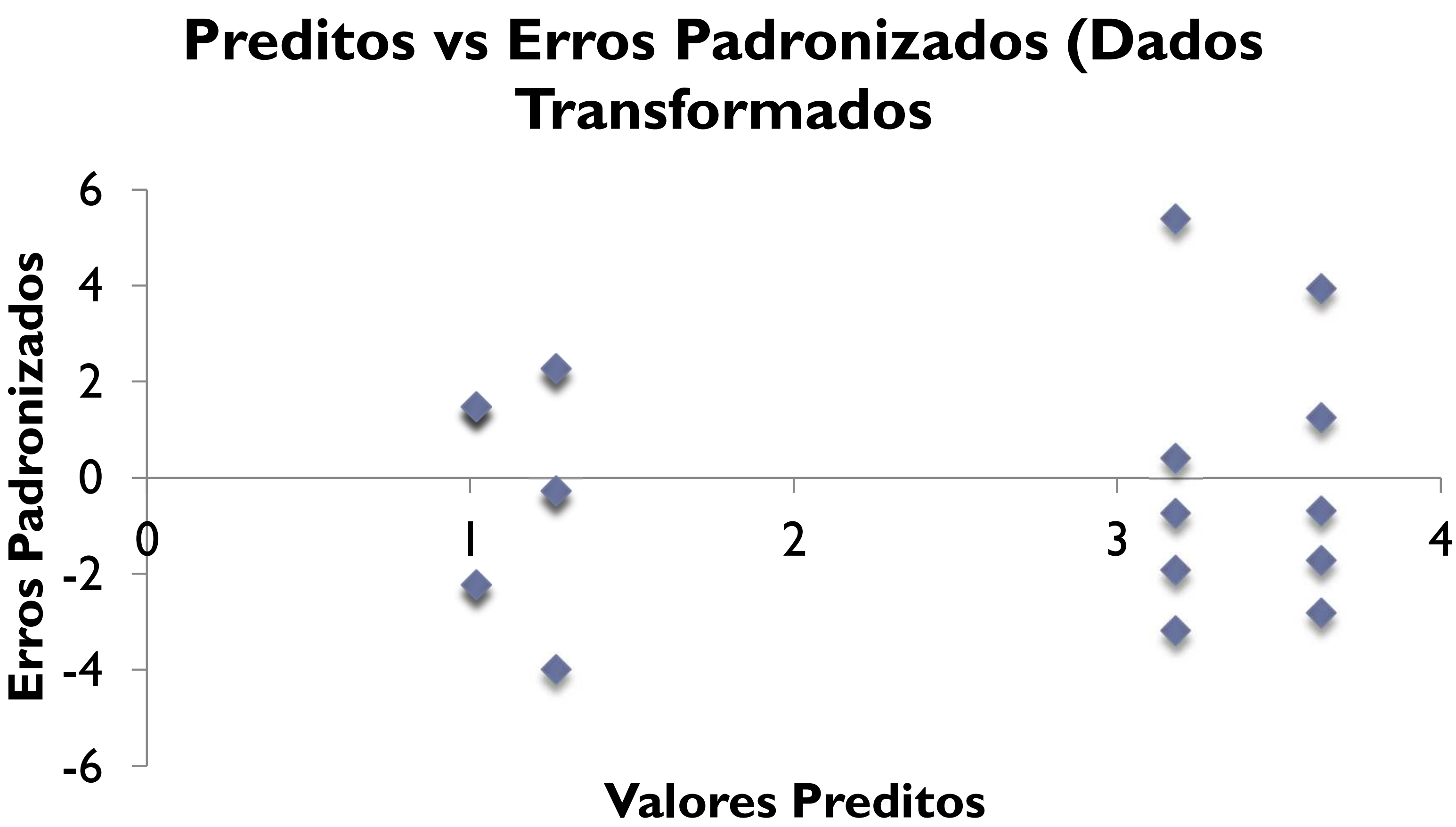
Pode-se notar que os resíduos estão bem mais aproximados da reta após a transformação dos dados, o que indica uma alta possibilidade de não se rejeitar a hipótese da normalidade dos resíduos após a transformação dos dados .

Histograma dos Resíduos – pós transformação dos dados



Assim como o Q-Q Plot, o histograma dos resíduos encontrados após a transformação dos dados, também indica uma possível normalidade dos mesmos.

Gráfico – Dados Transformados



Pressuposições – Dados Transformados

- Após a transformação dos dados passamos a respeitar todas as pressuposições do modelo matemático;
- Se torna possível a utilização do método da ANOVA na análise dos resultados do experimento.

Análise da Variância

- Como fazer a Análise da Variância?
- Primeiro passo: Definir o esquema da ANOVA

Causa da Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
Tratamentos	I-1	SQTrat	QMTrat	QMTrat/QMRes
Resíduos	I(J-1)	SQRes	QMRes	
Total	IJ-1	SQTotal		

Somas de Quadrados

- As Somas de Quadrados (SQ) são obtidas pelas seguintes expressões:
  - $\text{SQTotal} = \sum_{i,j} {y_{ij}}^2 - \frac{\left(\sum_{i,j} y_{ij}\right)^2}{IJ}$;
    - em que $\frac{\left(\sum_{i,j} y_{ij}\right)^2}{IJ} = C$ é denominado correção
  - $\text{SQTrat} = \frac{1}{J}\sum_i {y_{i.}}^2 - C = \frac{1}{J}\sum_i {T_i}^2 - C$;
  - $\text{SQRes} = \text{SQTotal} - \text{SQTrat}.$

Quadrados Médios e Teste F

- Os Quadrados Médios (QM) são obtidos pelas seguintes expressões:
- $\text{QMTrat} = \frac{\text{SQTrat}}{\text{gl(Trat)}} = \frac{\text{SQTrat}}{I-1}$;
- $\text{QMRes} = \frac{\text{SQRes}}{\text{gl(Res)}} = \frac{\text{SQRes}}{I(J-1)}$.
- O teste F considera a hipótese de nulidade:
  - os efeitos dos tratamentos não diferem entre si ($t_r = t_s, \forall\, r \text{ e } s$);
- e a alternativa:
  - ao menos dois tratamentos diferem entre si ($t_r \neq t_s$, para ao menos um $r \neq s$).
  - A estatística do teste e: $F = \frac{\text{QMTrat}}{\text{QMRes}}$.

Qual o Significado do Teste F

- Considere um delineamento inteiramente casualizado fixo ($\overline{m}$ e t_i fixos e e_{ij} aleatório), tem-se:
- $\mathrm{E(QMRes)} = \sigma^2 \rightarrow \mathrm{QMRes} = \hat{\sigma}^2$;
- $\mathrm{E(QMTrat)} = \sigma^2 + J\Phi_\tau \rightarrow \mathrm{QMTrat} = \hat{\sigma}^2 + j\widehat{\Phi}_\tau$, com $\widehat{\Phi}_\tau = \frac{\sum_i \hat{t}_i{}^2}{I-1}$;
- Logo: $F = \frac{\widehat{\sigma}^2 + \frac{J}{I-1}\sum_i \hat{t}_i{}^2}{\widehat{\sigma}^2} = 1 + J\,\frac{\widehat{\Phi}_\tau}{\widehat{\sigma}^2}$.
- Ou seja, F se afasta de 1 a medida que $\sum_i \hat{t}_i{}^2$ aumenta; como esse valor mede a variação dos efeitos de tratamento, tem-se que valores baixos indicam igualdade entre os tratamentos e valores altos diferenças entre os tratamentos.

Teste F

- Para saber o quanto o teste F deve se afastar de 1 para ser considerado significativo basta checar a tabela do teste, com n_1 graus de liberdade de tratamentos $(I - 1)$ e n_2 graus de liberdade dos resíduos $I(J - 1)$.
- No nosso exemplo:
  - $F_{obs} = 62{,}87^{**}$;
  - $F_{tab}\begin{cases} n_1 = 3 \\ n_2 = 16 \end{cases} \rightarrow \begin{cases} F_{tab} = 3{,}24\ (5\%) \\ F_{tab} = 5{,}29\ (1\%) \end{cases}$;
  - Como o valor observado é maior do que o tabelado conclui-se que ao menos duas variedades se diferenciam quanto ao estacamento de raizes. Costuma-se indicar que o valor é significativo a um nível de 5% com * e altamente significativo (1%) com **.

Coeficiente de Variação

- Toda a Análise da Variância deve ser seguida de seu Coeficiente de Variação:
- $CV = \frac{s}{\bar{x}} 100\%$
- O CV fornece um indicativo da maneira como o experimento foi conduzido;
- Quanto menor for o CV, maior é a probabilidade de que o experimento tenha sido bem instalado e conduzido;
- Em caso de ANOVAs feitas com dados transformados, deve-se utilizar o CV relativo aos dados originais.