Planejamento de Experimentos
Suposições do Modelo e Comparações Múltiplas

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG

Exemplo típico: Resistência de uma nova fibra sintética para camisas de homens. (Montgomery, 1997, p.63)

- Um engenheiro de produto está interessado em investigar a resistência de uma nova fibra sintética.
- A resistência é afetada pela porcentagem de algodão usada na produção do tecido.
- O engenheiro decide testar o novo tecido com diferentes porcentagens de algodão: 15, 20, 25, 30 e 35 %.
- A resposta é a resistência da fibra medida em libra/polegada^2.
- O experimento usou 5 réplicas para cada nível do fator.

Exemplo típico: Resistência de uma nova fibra sintética para camisas de homens (Montgomery, 1997, p.63)

- Como realizar a aleatorização se temos somente um processo para produzir a fibra?
- Necessário aleatorizar a ordem/sequência de produção/medida da fibra.
- A aleatorização evita possíveis efeitos de variáveis de perturbação. Por exemplo: aquecimento, desgaste, aprendizado, etc.

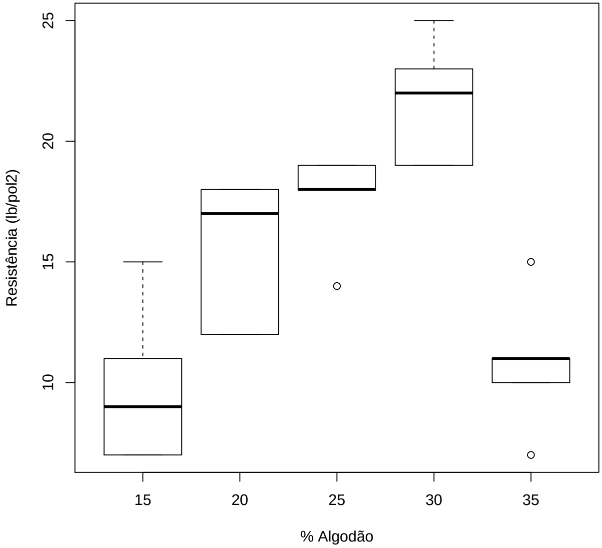
Exemplo: Resistência da fibra sintética

Tabela: Resistência da fibra sintética.(em libra/polegada^2)

Réplicas	**Porcentagem de Algodão**				
	15	20	25	30	35
1	7	12	14	19	7
2	7	17	18	25	10
3	15	12	18	22	11
4	11	18	19	19	15
5	9	18	19	23	11
	7	12	14	19	7
Média	9,8	15,4	17,6	21,6	10,8
DP	3,3	3,1	2,1	2,6	2,9

Média e desvio-padrão para o conjunto de dados são: 15,0 e 5,2.

Boxplot



Fontes de variação

- De forma análoga, temos a seguinte tabela de análise de variância.

Tabela: Tabela da Análise de Variância - ANOVA.

Fontes de variação	SQ	GL	QM	*F*	*p-valor*
Entre os grupos	475,8	4	$118,94$	14,765	< 0.001
Dentro dos grupos	161.2	20	$8,06$	-	-
Total	637,0	24	-	-	-

Modelo Estatístico para Efeitos Fixos

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

em que Y_{ij} é o valor da resposta da j-ésima observação ($j = 1, \ldots, n$) no i-ésimo nível do fator ($i = 1, \ldots, a$); μ: efeito geral da média: τ_i: efeito do i-ésimo tratamento.

- $Var(\epsilon_{ij}) = \sigma^2$
- $Y_{ij} \sim N(\mu + \tau_i, \sigma^2)$

Condições para o uso da ANOVA

- Para que os resultados da Análise de Variância sejam válidos, é necessário que:
  - Os desvios padrões das distribuições das respostas dos grupos serem iguais (HOMOCEDASTICIDADE);
  - e, a distribuição das respostas de cada grupo deve ser normal (NORMALIDADE).
  - Independência das observações.
- A ANOVA é razoavelmente robusta a afastamentos da normalidade e da homocedasticidade, especialmente se os tamanhos amostrais forem grandes.

Como verificar as suposições da ANOVA?

1. Uma ferramente útil para esta tarefa são os resíduos do ajuste da ANOVA.

2. Os resíduos são definidos da seguinte forma:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu} - \hat{\tau}_i = Y_{ij} - \bar{Y}_{i.}$$

em que $\hat{\mu}$ e $\hat{\tau}_i$ são os valores estimados pelos dados.

Verificando as suposições da ANOVA

1. HOMOCEDASTICIDADE
   - Teste Bartlett ou Levene ($\sigma_1^2 = \ldots = \sigma_a^2$).

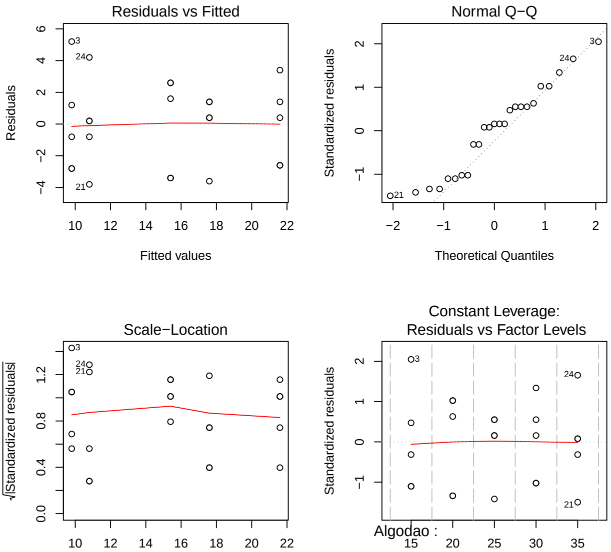
- Gráfico de resíduo (e_{ij}) vs ajustados ($\overline{Y}_{i.}$) (não deve exibir tendências sob homocedasticidade) ou o gráfico $\overline{Y}_{i.}$ vs $|e_{ij}|$.

2. NORMALIDADE
   - Teste Shapiro-Wilks, etc.
   - Gráfico de Probabilidade Normal dos resíduos.

Exemplo: Resistência de Tecido

- Bartlett: valor-p = 0,9198 / Levene: valor-p=0,8626
  (Validando a homocedasticidade)
- Shapiro-Wilks: valor-p = 0,182
  (Validando a normalidade).
- As suposições, desta forma, foram consideradas satisfeitas.

Exemplo: Resistência de Tecido - Gráfico de Resíduos



O que fazer se as suposições não valerem?

1. Testes Não-Paramétrico: Kruskal-Wallis, permutação, etc.
2. Transformação na Resposta.
3. Etc e etc....

Procedimentos de comparações múltiplas

- Um valor de F significativo indica a existência de pelo menos uma diferença entre os grupos estudados.
- A identificação de diferenças particulares entre médias, tomando-as duas a duas, deve ser realizada por um dos vários métodos de *Comparações Múltiplas entre Médias* existentes na literatura.
- Estes testes são semelhantes ao test t, com a diferença de que controlam o nível de significância ao levar em consideração o número de comparações a serem realizadas.

Tipo de Erro

- Família: Uma família de testes é um conjunto de testes baseado em alguma medida de erro global.
- Taxa de erro da família: (**familywise error rate**): é a probabilidade de rejeitarmos incorretamente ao menos umas das hipóteses nulas que compõem a família.
- Taxa de erro por comparação (per-comparison error rate): é a probabilidade de rejeitarmos incorretamente cada uma das hipóteses nulas que compõem a família.

Probabilidade de Erro

- Na presença de k tratamentos (ou níveis), temos $\begin{pmatrix} a \\ 2 \end{pmatrix} = \frac{a(a-1)}{2}$ comparações de pares de médias.

- A probabilidade de cometermos pelo menos um erro do tipo I é muito maior que α (note que estamos realizando vários testes simultaneamente).

- É difícil calcular de maneira exata a probabilidade global de erro (do tipo I) envolvendo comparações múltiplas.

- Uma cota superior para a probabilidade global de erro é dada por

$$P(\text{m erros tipo I} \geq 1) = 1 - (1 - \alpha)^m.$$

- OBS: a cota superior apresentada acima é obtida assumindo que as comparações entre os pares de médias são independentes (o que de fato não ocorre na prática).

Correção de Bonferroni: simples e ineficiente.

- O procedimento de Bonferroni consiste em corrigir o valor do nível de significância α, calculando-se

$$\alpha^* = \frac{\alpha}{m}$$

em que α é o nível de significância global e m é o número de comparações a serem realizadas ($m = \binom{a}{2}$, para a grupos).

- Por exemplo, para o caso de $a = 5$ populações, o total de testes é $m = 10$. Se definimos o nível de significância global em 5%, devemos utilizar

$$\alpha^* = \frac{0,05}{10} = 0,005$$

para cada teste individual.

- Para realizar um teste da hipótese nula

$$H_0 : \tau_i = \tau_j$$

devemos calcular

$$t_{ij} = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{s_E^2[(1/n) + (1/n)]}}.$$

- Este é um teste t para duas médias. No entanto, utilizamos a informação das a amostras, s_E^2. Sob a hipótese nula, t_{ij} tem uma distribuição t com $N - a$ graus de liberdade.

- Cada método fornece um valor de referência que deve ser comparado às diferenças de médias amostrais.
- De forma equivalente, eles fornecem um intervalo de confiança para a diferença de médias.
- Um procedimento usual consiste em: (1) ordenar as médias amostrais; (2) compará-las utilizando um método de comparação múltipla.

- O método de comparações múltiplas de Tukey é bastante popular por ser um dos mais antigos e razoavelmente eficiente.

- Neste teste, duas média amostrais são comparadas usando

$$\frac{S_{T(1-\alpha);a,N-a}}{\sqrt{2}}\sqrt{s_E^2\left(\frac{1}{n}+\frac{1}{n}\right)}$$

em que $S_{T(1-\alpha);a,N-a}$ é o quantil de probabilidade $(1-\alpha)$ da distribuição Studentizada com a e $N-a$ graus de liberdade.

- A hipótese $H_0 : \tau_i = \tau_j$ é rejeitada se

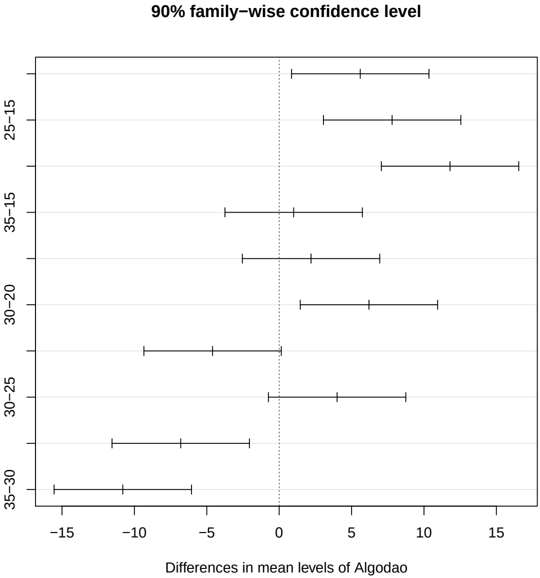
$$|\bar{y}_{i.} - \bar{y}_{j.}| \geq \frac{S_{T(1-\alpha);a,N-a}}{\sqrt{2}} \sqrt{s_E^2 \left(\frac{1}{n} + \frac{1}{n} \right)}$$

- Neste teste a hipótese nula $H_0 : \tau_i = \tau_j$ é rejeitada se

$$|\bar{y}_{i.} - \bar{y}_{j.}| \geq \sqrt{(a-1)F_{(1-\alpha);a-1,N-a}}\sqrt{s_E^2\left(\frac{1}{n}+\frac{1}{n}\right)}$$

em que, $F_{(1-\alpha)}$ é o quantil de probabilidade $(1-\alpha)$ da distribuição $F_{a-1,N-a}$.

Comparações Múltiplas
Forma alternativa de apresentar o Método de Tukey



Outro Exemplo: Volume expiratório forçado (FEV) (Pagano e Gauvreau, 2004, p.256)

- Deseja-se comparar o *volume expiratório forçado* de pacientes com doença coronária oriundos de três centros médicos diferentes (21 pacientes da Johns Hopkins University School of Medicine, 16 pacientes do Rancho Los Amigos Medical Center e 23 pacientes da St. Louis University School of Medicine).
- Ou seja, deseja-se testar a hipótese $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ contra a alternativa de que pelo menos duas médias populacionais são diferentes.
- É possível aleatorizar?

Exemplo: Volume expiratório forçado (FEV)

Tabela: Volume expiratório forçado em 1 segundo para pacientes com doença coronária de três diferentes centros médicos.(em litros)

Johns Hopkins		Rancho Los Amigos		St. Louis	
3,23	2,57	3,22	2,61	2,79	3,17
3,47	2,08	2,88	3,39	3,22	2,23
1,86	2,47	1,71	3,17	2,25	2,19
2,47	2,47	2,89		2,98	4,06
3,01	2,74	3,77		2,47	1,98
1,69	2,88	3,29		2,77	2,81
2,10	2,63	3,39		2,95	2,85
2,81	2,53	3,86		3,56	2,43
3,28		2,64		2,88	3,20
3,36		2,71		2,63	3,53
2,61		2,71		3,38	
2,91		3,41		3,07	
1,98		2,87		2,81	
$n_1 = 21$		$n_2 = 16$		$n_3 = 23$	
$\bar{x}_1 = 2,63$ litros		$\bar{x}_2 = 3,03$ litros		$\bar{x}_3 = 2,88$ litros	
$s_1 = 0,496$ litros		$s_2 = 0,523$ litros		$s_3 = 0,498$ litros	

Exemplo: FEV - Análise Descritiva - Box-plots

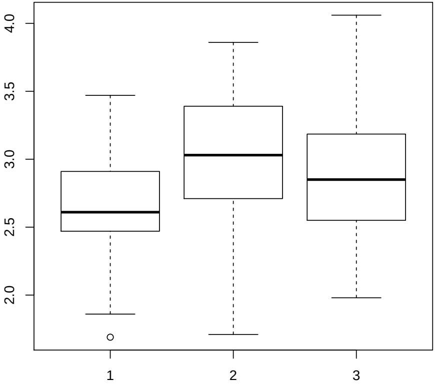


Figura: 1- John Hopkins, 2- Rancho Los Amigos e 3- St. Louis.

Teste NP - Kruskal-Wallis

- O teste de Kruskal-Wallis compara medianas e não médias.
- Todo teste não-paramétrico, basicamente, não tem suposições. O teste é baseado na ordenação dos dados amostrais.
- O teste de Kruskal-Wallis é menos eficiente que o teste-F.
- O primeiro passo de um teste não-paramétrico é ordenar todas as observações como se elas fossem de uma única amostra.

Amostra I		Amostra II		Amostra III	
Posto	Valor	Posto	Valor	Posto	Valor
2	3	1	2	10,5	9
3,5	5	3,5	5	14	12
5,5	7	5,5	7	16,5	16
7,5	8	7,5	8	18	18
10,5	9	10,5	9	19	19
10,5	9	15	15	20	20
13	11	16,5	16	21	24
T_1(soma)=52,5		T_2(soma)=59,5		T_3(soma)=119	

Exemplo: Teste Kruskal-Wallis

Estatística de Kruskal-Wallis é:

$$H = \frac{12(n_1(\overline{T}_1 - \overline{T})^2 + \ldots n_k(\overline{T}_k - \overline{T})^2)}{n(n+1)}$$

Sob a hipótese de não haver diferença entre os grupos, H tem uma distribuição qui-quadrado com $k-1$ graus de liberdade.

No exemplo, temos que: $\overline{T}_1 = 52,5/7 = 7,5$, $\overline{T}_2 = 59,5/7 = 8,5$, $\overline{T}_3 = 119/7 = 17$ e $\overline{T} = \frac{7.5+8.5+17}{3} = 11$

$$H = \frac{12(7(7,5-11)^2 + 7(8,5-11)^2 + 7(17-11)^2}{2122))} = 9,91$$

valor-p = 0,0067, obtido a partir da qui-quadrado com 2 gl.
Exemplo: VEF $\longrightarrow$ valor-p= 0,0498.

Conclusão Final: Exemplo VEF

- A média do FEV dos pacientes da Johns Hopkins é significativa menor que a daqueles do Rancho Los Amigos. Nenhuma outra diferença foi detectada.

- Intervalo de 90% de Confiança para a diferença média foi:

$$\bar{y}_{1.} - \bar{y}_{2.} \pm t_{57;1-(0,1/2*3)}\sqrt{s_E^2[(1/n_1) + (1/n_2)]} =$$
$$2,63 - 3,03 \pm 2,18\sqrt{0,25[(1/21) + (1/16)]} = (-0,77; -0,04)$$
$$(-0,68; 0,12).$$

Ou seja, o FEV médio dos pacientes do centro médico de Rancho Los Amigos é cerca de 0,4 l (IC; 90%, 0,04;0,77) maior que o FEV médio daqueles da Johns Hopkins.