Métodos Estatísticos Avançados em Epidemiologia

Análise de Variância - ANOVA

Referência: Cap. 12 - Pagano e Gauvreau (2004) - p.254

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG

- **Pergunta Clínica Típica**: comparar **dois ou mais grupos** experimentais com relação a um desfecho quantitativo.
- **Exemplos:**
    - Comparar três drogas para o tratamento de fibrose cística em que o desfecho é uma medida quantitativa de capacidade pulmonar.
    - Comparar a **idade** de pacientes entre três grupos de risco (baixo, médio, alto).
- Comparação de dois grupos: testes t ou Mann-Whitney.

Exemplo

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Prednisone, Azathioprine, and N-Acetylcysteine for Pulmonary Fibrosis

The Idiopathic Pulmonary Fibrosis Clinical Research Network*

ABSTRACT

BACKGROUND

A combination of prednisone, azathioprine, and N-acetylcysteine (NAC) has been widely used as a treatment for idiopathic pulmonary fibrosis. The safety and efficacy of this three-drug regimen is unknown.

METHODS

In this randomized, double-blind, placebo-controlled trial, we assigned patients with idiopathic pulmonary fibrosis who had mild-to-moderate lung-function impairment to one of three groups — receiving a combination of prednisone, azathioprine, and NAC (combination therapy), NAC alone, or placebo — in a 1:1:1 ratio. The primary outcome was the change in longitudinal measurements of forced vital capacity during a 60-week treatment period.

N ENGL J MED 366;21 NEJM.ORG MAY 24, 2012

STATISTICAL ANALYSIS

The trial was designed with a two-step procedure to control the experiment-wise error rate at the 0.05 level. The first step was based on an overall test of 2 degrees of freedom. If any difference between study groups was statistically significant at the 0.05 level, then each of the three pairwise comparisons would be tested at the 0.05 level.

Introdução

- A primeira vista, pode parecer correto realizar vários testes *t* entre os grupos, comparando-os *dois a dois*.
- No caso da comparação de três grupos (grupo A, grupo B e grupo C), temos **três testes** *t* de comparação entre médias: μ_A *vs* μ_B, μ_A *vs* μ_C e μ_B *vs* μ_C.
- Na comparação de quatro grupos, temos **seis testes** *t* de comparação entre médias.
- Se o número de grupos é igual a 10, precisaríamos de **45 testes** *t* dois a dois.

- **Observação 1:** O número de testes aumenta conforme o número de grupos aumenta. Para k grupos temos $\binom{k}{2}$ comparações.
- **Observação 2:** Tal procedimento (a realização de todas as comparações dois a dois) é estatisticamente incorreto.
  - O teste t foi proposto para, **em um mesmo experimento**, comparar-se uma média A com apenas outra, B, com probabilidade fixa do erro tipo I (α=0,05).
  - Se forem feitas mais de uma comparação envolvendo a média A, a probabilidade do erro tipo I (ou erro experimental conjunto) **passa a ser maior do que** α.

Introdução

- O procedimento mais indicado para se evitar esse aumento no nível global de significância do experimento consiste em utilizar a técnica da **Análise de Variância** (ANOVA) que consiste nos seguintes passos:
  - PASSO 1 - Comparar todas as médias em um único teste. O objetivo inicial é identificar a existência de **ao menos uma diferença** entre grupos.
  - PASSO 2 - Caso o resultado anterior for significativo, aplica-se um ou mais métodos de **comparações múltiplas**. O objetivo é identificar quais as médias são diferentes, controlando o nível global de significância.

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 $H_0 : \mu_1 = \mu_2 = \mu_3$ contra a alternativa de que pelo menos duas médias populacionais são diferentes.

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

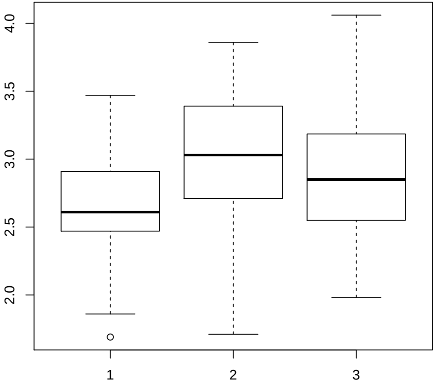


Figura: 1- John Hopkins, 2- Rancho Los Amigos e 3- St. Louis.

ANOVA - Fontes de variação

- A ANOVA é baseada em estimativas de dispersão/variância.
- Neste caso, existem duas diferentes fontes de variação:
  - Variação natural ou intra-grupo: valores indivíduos em torno das médias populacionais (desvio-padrão intra-grupo);
  - Variação entre-grupos: médias populacionais em torno da média global (desvio-padrão inter-grupos).
- Se a variabilidade intra populações é menor que a variabilidade entre grupos, sugere que as médias populacionais são de fato diferentes.

Desvio-padrão Intra-Grupo

- Objetivo: testar a seguinte hipótese:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

para as médias de k populações.

- Estimativa do desvio-padrão Intra-Grupo (s_D):

$$s_D^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k},$$

em que, s_j e n_j são o desvio-padrão e tamanho de amostra do j-ésimo grupo e n é o tamanho amostra total.

- Esta quantidade é, simplesmente, a média ponderada das k variâncias amostrais. O subscrito D se refere a variabilidade "dentro de grupos".

Desvio-padrão Entre-Grupos

- Necessitamos de uma estimativa da variação das médias em torno da média global.
- Estimativa do desvio-padrão Entre-Grupos (s_E):

$$s_E^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k-1},$$

  em que $\bar{x}_j$ é a média amostral do j-ésimo grupo e $\bar{x}$ é a média global das n observações

- Se a hipótese nula é verdadeira, esta quantidade também estima a variância intra-grupo, s_D^2.

Testa F

- A pergunta clínica pode ser traduzida pela seguinte questão: as médias amostrais variam em torno da média global mais do que as observações individuais variam em torno das médias amostrais?

- Em caso positivo, isto inidica que existe alguma diferença entre as médias populacionais.

- Precisamos de um estatístca para avaliar o tamanho desta diferença. A estatística F é usada para este propósito:

$$F = \frac{s_E^2}{s_D^2}.$$

Teste F

- Sob a hipótese nula, que as médias são iguais, tanto s_E^2 quanto s_D^2 estimam a variância comum σ^2, e F deve estar próximo de 1.
- Se existe uma diferença entre as populações, então a variância entre os grupos é maior que a variância dentro dos grupos, e F é maior que 1.
- Sob H_0, a razão F tem uma distribuição F com $k-1$ e $n-k$ graus de liberdade.

Distribuição F

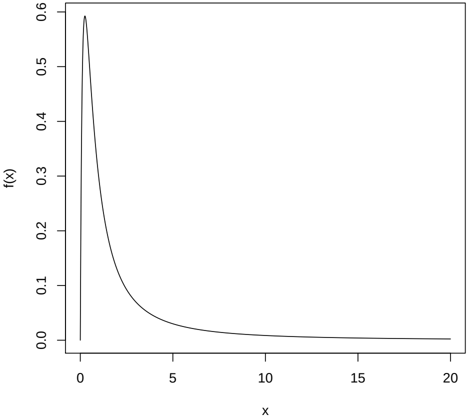
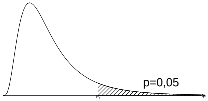


Figura: Distribuição F com 4 e 2 graus de liberdade.

Distribuição F

Distribuição de Snedecor a 5% (p=0,05)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,43	19,44	19,45	19,46	19,47	19,48	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,70	8,69	8,67	8,66	8,62	8,59	8,57	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,86	5,84	5,82	5,80	5,75	5,72	5,69	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,62	4,60	4,58	4,56	4,50	4,46	4,43	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,94	3,92	3,90	3,87	3,81	3,77	3,74	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,51	3,49	3,47	3,44	3,38	3,34	3,30	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,22	3,20	3,17	3,15	3,08	3,04	3,01	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	3,01	2,99	2,96	2,94	2,86	2,83	2,79	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,85	2,83	2,80	2,77	2,70	2,66	2,62	2,58
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,72	2,70	2,67	2,65	2,57	2,53	2,49	2,45
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,62	2,60	2,57	2,54	2,47	2,43	2,38	2,34
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,53	2,51	2,48	2,46	2,38	2,34	2,30	2,25
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,46	2,44	2,41	2,39	2,31	2,27	2,22	2,18
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,40	2,38	2,35	2,33	2,25	2,20	2,16	2,11
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,35	2,33	2,30	2,28	2,19	2,15	2,11	2,06
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,31	2,29	2,26	2,23	2,15	2,10	2,06	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,27	2,25	2,22	2,19	2,11	2,06	2,02	1,97
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,23	2,21	2,18	2,16	2,07	2,03	1,98	1,93
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,20	2,18	2,15	2,12	2,04	1,99	1,95	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,18	2,16	2,12	2,10	2,01	1,96	1,92	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,15	2,13	2,10	2,07	1,98	1,94	1,89	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,13	2,11	2,08	2,05	1,96	1,91	1,86	1,81
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,11	2,09	2,05	2,03	1,94	1,89	1,84	1,79
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,09	2,07	2,04	2,01	1,92	1,87	1,82	1,77
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,07	2,05	2,02	1,99	1,90	1,85	1,80	1,75
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,06	2,04	2,00	1,97	1,88	1,84	1,79	1,73
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,04	2,02	1,99	1,96	1,87	1,82	1,77	1,71
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,03	2,01	1,97	1,94	1,85	1,81	1,75	1,70
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	2,01	1,99	1,96	1,93	1,84	1,79	1,74	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,92	1,90	1,87	1,84	1,74	1,69	1,64	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,84	1,82	1,78	1,75	1,65	1,59	1,53	1,47
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,78	1,75	1,73	1,69	1,66	1,55	1,50	1,43	1,35

Tabela 5: Quantis da Distribuição para probabilidade . Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.

Fontes de variação - Tabela ANOVA

- **Observação:** Podemos organizar o teste F na seguinte tabela, chamada de tabela de Análise de Variância:

Tabela: Tabela da Análise de Variância (ANOVA).

Fontes de variação	SQ	GL	QM	F
Entre os grupos	SQ_E	$k-1$	$QM_E = SQ_E\ /k-1$	$QM_E\ /\ QM_D$
Dentro dos grupos	SQ_D	$n-k$	$QM_D = SQ_D\ /n-k$	-
Total	SQ Total	$n-1$	-	-

em que:

- SQ_E é a "soma de quadrados" entre os grupos e é o numerador de s_E^2, ou seja, $SQ_E = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2$.
- SQ_D é a "soma de quadrados" dentro dos grupos e é o numerador de s_D^2, ou seja, $SQ_D = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2$.
- Note que $QM_E = s_E^2$ e $QM_D = s_D^2$.

Exemplo FEV - Fontes de variação

- Retomando ao exemplo, estamos interessados em testar se a média da variável FEV é igual para os pacientes dos três diferentes centros médicos. Ou seja,

$$H_0 : \mu_1 = \mu_2 = \mu_3.$$

- Para começar, calculamos a estimativa da variância dentro dos grupos

$$\begin{aligned} s_D^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_k^2}{n_1+n_2+n_3-3} \\ &= \frac{(21-1)(0,496)^2 + (16-1)(0,523)^2 + (23-1)(0,498)^2}{21+16+23-3} \\ &= 0,254 \quad \text{litros}^2. \end{aligned}$$

Fontes de variação

- Temos que a média global (soma de todas observações dividida pelo tamanho da amostra) é

$$\begin{aligned}
\bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} \\
&= \frac{21(2,63) + 16(3,03) + 23(2,88)}{21 + 16 + 23} \\
&= 2,83 \quad \text{litros},
\end{aligned}$$

e assim a estimativa da variância entre os grupos é

$$\begin{aligned}
s_E^2 &= \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2}{3 - 1} \\
&= \frac{21(2,63 - 2,83)^2 + 16(3,03 - 2,83)^2 + 23(2,88 - 2,83)^2}{3 - 1} \\
&= 0,791 \quad \text{litros}^2.
\end{aligned}$$

Fontes de variação

- Desta forma, a estatística de teste F é

$$\begin{aligned} F &= \frac{s_E^2}{s_D^2} \\ &= \frac{0,791}{0,254} \\ &= 3,115. \end{aligned}$$

- Comparar com a distribuição F com $k-1=3-1=2$ e $n-k=60-3=57$ graus de liberdade, o $valor-p=0,052$.

- Rejeitamos a hipótese nula ao nível de 10% de significância, mas não se rejeita ao nível de 5% de significância.

- Possivelmente haja alguma diferença entre as médias dos valores do FEV entre estas três populações.

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	1,58	2	$0,791$	3,115	0,052
Dentro dos grupos	14,48	57	$0,254$	-	-
Total	16,02	59	-	-	-

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.

- 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 ($m = \binom{k}{2}$, para k grupos).

- Por exemplo, para o caso de $k = 3$ populações, o total de testes é $m = 3$. Se definimos o nível de significância global em 10%, devemos utilizar

$$\alpha^* = \frac{0,10}{3} = 0,033$$

para cada teste individual.

- Para realizar um teste da hipótese nula

$$H_0 : \mu_i = \mu_j$$

devemos calcular

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{s_D^2[(1/n_i) + (1/n_j)]}}.$$

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

- Retornando ao exemplo do FEV, comparando as populações 1 e 2, os pacientes da Johns Hopkins e aqueles do Rancho Los Amigos.

$$
\begin{aligned}
t_{12} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_D^2[(1/n_1) + (1/n_2)]}} \\
&= \frac{2,63 - 3,03}{\sqrt{0,254[(1/21) + (1/16)]}} \\
&= -2,43.
\end{aligned}
$$

- Para um teste t com $n - k = 60 - 3 = 57$ graus de liberdade, o valor-p$= 0,018$.
- Usando um nível de 3,3%, concluímos que as médias do FEV dos pacientes da Johns Hopkins e do Rancho Los Amigos são diferentes.

- Comparando as populações 1 e 3 (os pacientes da Johns Hopkins e aqueles da St. Louis), temos

$$\begin{aligned} t_{13} &= \frac{\bar{x}_1 - \bar{x}_3}{\sqrt{s_D^2[(1/n_1) + (1/n_3)]}} \\ &= \frac{2,63 - 2,88}{\sqrt{0,254[(1/21) + (1/23)]}} \\ &= -1,64. \end{aligned}$$

- Como o valor-p $> 0,10$, não temos evidências suficientes para concluir pela diferença de médias.

- Comparando os pacientes do Rancho Los Amigos e aqueles da St. Louis (populações 2 e 3), temos

$$\begin{aligned} t_{23} &= \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{s_D^2[(1/n_2) + (1/n_3)]}} \\ &= \frac{3,03 - 2,88}{\sqrt{0,254[(1/16) + (1/23)]}} \\ &= 0,91. \end{aligned}$$

- Novamente o valor-p $> 0,10$, e não temos evidências suficientes para concluir pela diferença de médias.

- 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

$$\frac{S_{T(1-\alpha);k,n-k}}{\sqrt{2}}\sqrt{s_D^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

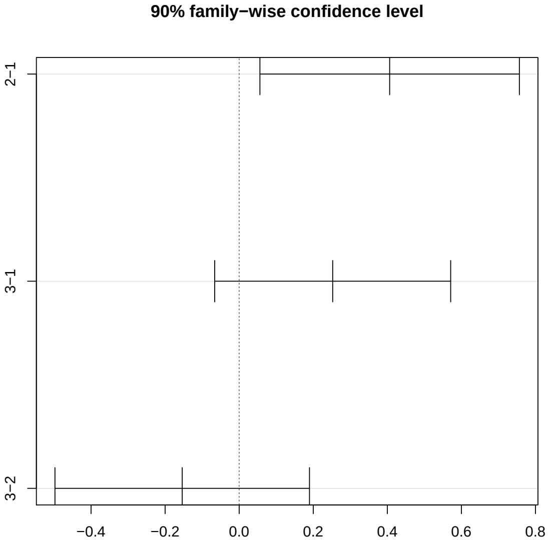
em que $S_{T(1-\alpha);k,n-k}$ é o quantil de probabilidade $(1-\alpha)$ da distribuição Studentizada com k e $n-k$ graus de liberdade.

- A hipótese $H_0 : \mu_i = \mu_j$ é rejeitada se

$$|\bar{x}_i - \bar{x}_j| \geq \frac{S_{T(1-\alpha);k,n-k}}{\sqrt{2}} \sqrt{s_D^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

- Comparando as populações 1 e 2, os pacientes da John Hopkins e do Rancho Los Amigos, o valor-p obtido foi igual a 0,047.
- Concluímos, ao nível 10%, que as médias do FEV dos pacientes da Johns Hopkins e do Rancho Los Amigos são diferentes.
- Comparando as populações 1 e 3 (pacientes da Jonh Hopkins e da St. Louis), o valor-p obtido foi igual a 0,229.
- Comparando os pacientes do Rancho Los Amigos e da St. Louis (populações 2 e 3), o valor-p obtido foi igual a 0,619.
- Desta forma, concluímos que as médias do FEV dos pacientes da Jonh Hopkins e da St. Louis e de St. Louis e do Rancho Los Amigos não são diferentes.

Comparações Múltiplas
Forma alternativa de apresentar o Método de Tukey



- Neste teste a hipótese nula $H_0 : \mu_i = \mu_j$ é rejeitada se

$$|\bar{x}_i - \bar{x}_j| \geq \sqrt{(k-1)F_{(1-\alpha);k-1,n-k}}\sqrt{s_D^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

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

Condições para o uso da ANOVA

- Para que os resultados da Análise de Variância sejam válidos, é necessário que:
  - Os desvios padrões das distribuições da resposta (desfecho) dos grupos tem que ser iguais (HOMOCEDASTICIDADE);
  - e, a distribuição das respostas (desfechos) de cada grupo deve ser normal (NORMALIDADE).
- 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. Modelo associado com a ANOVA:

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

   em que Y_{ij} é o valor do desfecho da j-ésima observação no i-ésimo tratamento/grupo; μ: efeito geral da média: τ_i: efeito do i-ésimo tratamento.

3. Os resíduos são definidos da seguinte forma:

$$\hat{\epsilon}_{ij} = Y_{ij} - \hat{\mu} - \hat{\tau}_i$$

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

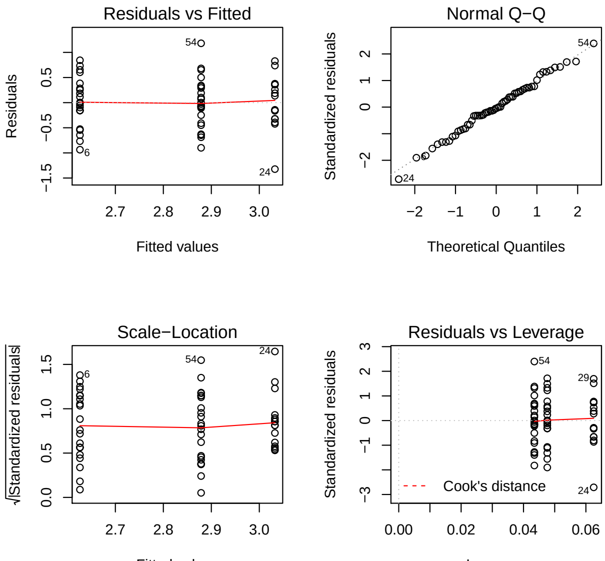
Verificando as suposições da ANOVA

1. HOMOCEDASTICIDADE
   - Teste Bartlett ou Levene ($\sigma_1^2 = \ldots = \sigma_k^2$).
   - Gráfico de resíduos vs ajustados (não deve exibir tendências sob homocedasticidade).
2. NORMALIDADE
   - Teste Shapiro-Wilks.
   - Gráfico de Probabilidade Normal dos resíduos.

Exemplo: FEV - Testes

- Considerando o exemplo do FEV, foram realizados os seguintes testes.
- Bartlett: valor-p = 0,971 / Levene: valor-p=0,959
  (Validando a homocedasticidade)
- Shapiro-Wilks: valor-p = 0,980
  (Validando a normalidade).
- As suposições, desta forma, foram consideradas satisfeitas.

Exemplo: FEV - Resíduos - Resultado do R



O que fazer se as suposições não valerem?

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

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.

Exemplo: Teste Kruskal-Wallis

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:

$$\bar{x}_1 - \bar{x}_2 \pm t_{57;1-(0,1/2*3)} \sqrt{s_D^2[(1/n_1) + (1/n_2)]} =$$

$$2,63 - 3,03 \pm 2,18\sqrt{0,25[(1/21) + (1/16)]} = (-0,77; -0,04)$$

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.