BIOLOGIA/BIOMEDICINA

BIOESTATÍSTICA

Profª. Letícia Raposo
profleticiaraposo@gmail.com

RELEMBRANDO...

Em geral, na aplicação de um teste estatístico, devemos saber:

a) formular H0 e H1 em termos de parâmetros populacionais;

b) como obter a estatística do teste;

c) qual é a distribuição de referência para calcular o valor p;

d) quais as suposições básicas para o uso do teste escolhido.

A decisão do teste estatístico é feita pela comparação do valor p com o nível de significância preestabelecido, mas a implicação do resultado estatístico depende da aplicação em questão.

DIFERENÇA ENTRE DOIS GRUPOS

OBJETIVOS

- Observaremos estatísticas que nos dizem se duas condições (ou grupos) diferem entre si em uma ou mais variáveis. As duas condições podem ser:
  - O mesmo grupo de pessoas/elementos testado em duas condições (tradicionalmente chamadas de A e B);
  - Dois grupos diferentes de pessoas/elementos que foram submetidos à condição A ou à condição B.

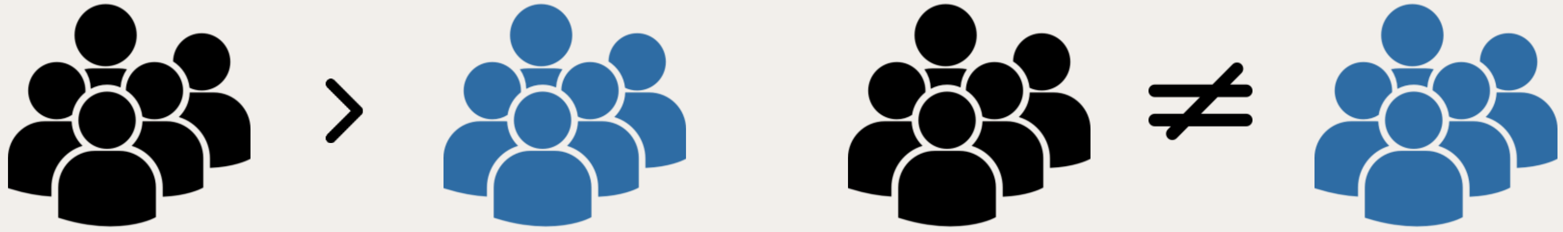
TESTES PARAMÉTRICOS X NÃO PARAMÉTRICOS

Testes paramétricos: supõem que a população de que os valores foram retirados são normalmente distribuídos – fazem suposições sobre os parâmetros da população subjacente.
São geralmente os mais sensíveis que podemos usar (mais poderosos).

Se os dados não satisfazem as suposições para os testes paramétricos, existem os não paramétricos – não têm suposições rígidas sobre as distribuições da população, mas tendem a ter uma menor probabilidade para detectar um efeito que existe na população de interesse.

INTRODUÇÃO

- Na comparação entre dois grupos ou duas condições, os pesquisadores lançam uma hipótese de que haverá uma diferença significativa entre eles.
- H0: qualquer diferença nos valores entre as condições ocorre devido ao erro amostral.



EXEMPLO 1

Pense se o delineamento é de grupos independentes ou de medidas repetidas

Yu e colaboradores (2007) estavam examinando a eficácia de um programa de treinamento de rastreamento do câncer de mama.

O grupo testado era formado por de estagiários.

Foi dada a eles uma pré-intervenção, mensurando o conhecimento e a eficácia pessoal. O grupo foi engajado em um autoestudo de materiais de treinamento. No fim do programa, eles responderam a questionários para mensurar o conhecimento e a autoeficácia.

EXEMPLO 2

Pense se o delineamento é de grupos independentes ou de medidas repetidas

Skumlien e colaboradores (2007) investigaram os benefícios da reabilitação intensiva em pacientes com doença pulmonar obstrutiva crônica (DPOC).

Eles observaram as mudanças na incapacidade funcional e na saúde em relação à reabilitação pulmonar (RP).

Quarenta pessoas com DPOC participaram da reabilitação pulmonar multidisciplinar de pacientes, o que consistiu em um treinamento de tolerância, em um treinamento de paciência, em uma sessão de educação e em sessões individuais de aconselhamento. Esse grupo foi, então, comparado a um grupo de pacientes que estavam na lista de espera (chamado de grupo-controle da lista de espera).

EXEMPLO 3

Pense se o delineamento é de grupos independentes ou de medidas repetidas

Shearer e colaboradores (2009) compararam os valores da glicose nos pontos de cuidado com valores de laboratório em pacientes críticos.

Sessenta e três desses indivíduos tinham os níveis de glicose medidos no leito por meio de um glicosímetro, utilizando sistemas de ponto de cuidados (POC). Esse método é geralmente utilizado com o medidor de glicose, em vez de enviar amostras a um laboratório.

Os pesquisadores, então, compararam os valores da glicose obtidos dos 63 pacientes por ambos os métodos.

EXEMPLO 4

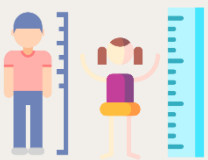
Pense se o delineamento é de grupos independentes ou de medidas repetidas

Giovannelli e colaboradores (2007) queriam observar se a fisioterapia aumentava os efeitos da toxina botulínica tipo A na redução da espasticidade em pacientes com esclerose múltipla (EM).

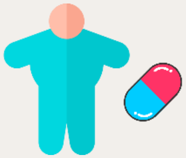
Havia 38 pacientes no estudo (um ensaio aleatório controlado), consistindo em um grupo de intervenção (que tomou uma injeção da toxina botulínica mais fisioterapia adicional) e um grupo-controle (injeção botulínica).

TESTE T

- Teste t para 1 amostra
  - A altura média das estudantes universitárias é maior que 1,65 m?



- Teste t para 2 amostras
  - A altura média das estudantes universitárias é significativamente diferente da altura média dos estudantes universitários?



- Teste t pareado
  - Se você mede o peso de estudantes universitários do sexo masculino antes e depois que cada sujeito toma uma pílula para perder peso, a perda de peso média é suficientemente significante para concluir que a pílula funciona?

DESCRIÇÃO CONCEITUAL DOS TESTES T

Paciente	Tratamento	Paciente	Controle
1	14	11	9
2	18	12	7
3	10	13	12
4	13	14	11
5	15	15	14
6	15	16	5
7	12	17	4
8	12	18	10
9	10	19	9
10	12	20	9
Média	13,1	Média	9

- Grupo 1: pacientes com esclerose múltipla em um novo tratamento;
- Grupo 2: pacientes na lista de espera (grupo-controle).
- A variável independente é "Tratamento", e a variável dependente é a medida da memória.

Há variação intragrupos e entre grupos.

DESCRIÇÃO CONCEITUAL DOS TESTES T

- Os dois grupos diferem? → não podemos simplesmente olhar as suas médias. Precisamos saber se a variância entre grupos difere daquela intragrupos.
- O teste estatístico t é calculado obtendo-se a diferença entre as duas médias e dividindo o resultado por uma medida representando a variação nos valores para os grupos. Essa medida de variância é o "erro padrão da diferença".

t = Diferença entre as médias (sinal)
÷ Variabilidade indesejada (ruído)

O "ruído" é a variabilidade para cada grupo. Quanto maior o ruído, mais baixa será a razão sinal-ruído (ou seja, o valor t será mais baixo); quanto menor o ruído, maior a razão sinal-ruído (o valor t será mais alto).

DESCRIÇÃO CONCEITUAL DOS TESTES T

- O valor t pode ser negativo ou positivo; isso dependerá da ordem.
  - Grupo 1 — grupo do tratamento: média 13,1.

    Grupo 2 — grupo-controle: média 9.

    A diferença das médias é positiva e o valor t é, portanto, positivo.
- O que acontece se codificamos o grupo-controle como grupo 1 e o grupo do tratamento como grupo 2?
  - Não importa como você codifica os grupos, um valor t negativo é tão importante e significativo quanto um valor positivo.

GENERALIZANDO PARA A POPULAÇÃO

- Qualquer pessoa que realiza uma pesquisa envolvendo a observação de diferenças entre dois grupos ou condições tenta assegurar que suas amostras do grupo (ou dos grupos) sejam representativas da população da qual foram retiradas.
- Embora possamos achar que nossas duas condições diferem significativamente, queremos ser capazes de generalizar para populações maiores — isto é, a partir de esquemas executado com intervalos de confiança.

Os pesquisadores querem ser capazes de generalizar para a população.

GENERALIZANDO PARA A POPULAÇÃO

- A média de ambos são "estimativas por pontos".
  - Se tivéssemos executado o experimento em um dia diferente, com uma nova amostra, ou até mesmo com a mesma amostra, é provável que as médias fossem diferentes.
  - Portanto, precisamos do intervalo de confiança → 1,49 — 6,71: embora na amostra a diferença média seja de 4,1, estamos 95% confiantes de que a diferença média na população esteja entre 1,49 e 6,71.
- Quanto mais estreitos os intervalos de confiança, melhor.
- Quando o intervalo de confiança inclui 0 (zero), o valor t será baixo e não será estatisticamente significativo.

PRESSUPOSTOS DA ANÁLISE

Para realizar o teste t para amostras independentes existem 2 pressupostos:

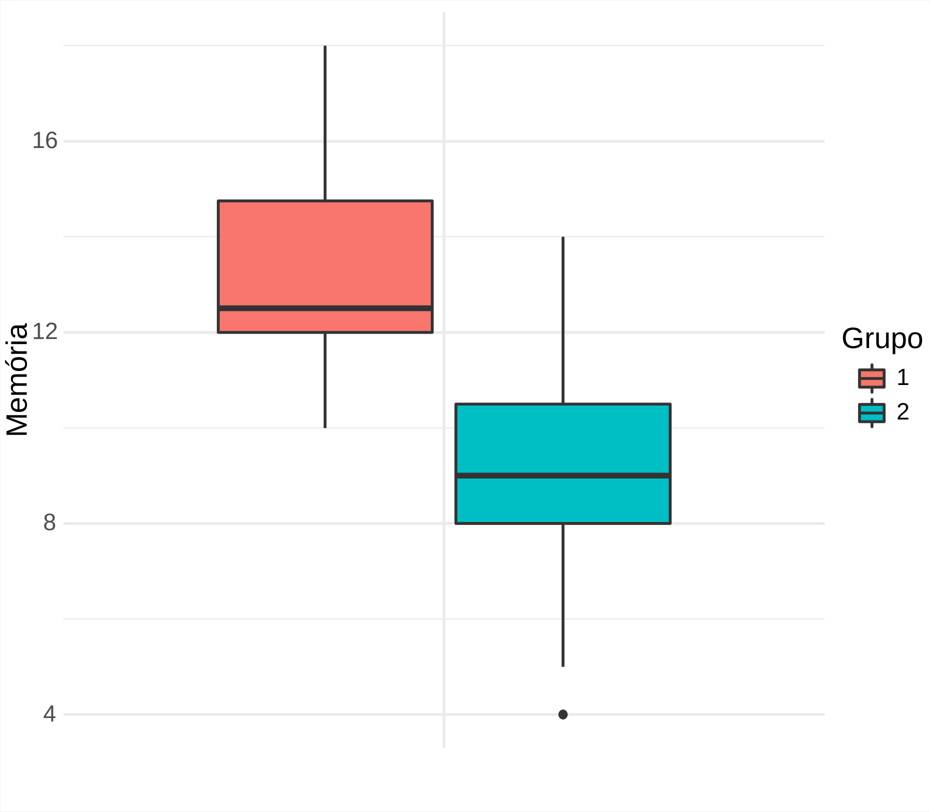
- A distribuição dos dados seja normal e as variâncias sejam homogêneas.
- O teste de Shapiro-Wilk é uma função que testa a normalidade dos dados.

H0: a amostra provém de uma população Normal
H1: a amostra não provém de uma população Normal

- O teste de Levene é uma função que testa a homogeneidade das variâncias.

H0: as variâncias são homogêneas
H1: as variâncias não são homogêneas

TESTE T PARA GRUPOS INDEPENDENTES NO R



TESTE T PARA GRUPOS INDEPENDENTES NO R

```
> # Avaliando a normalidade
>
> library(RVAideMemoire)
> byf.shapiro(memory ~ group, data = dadost)

        Shapiro-Wilk normality tests

data:  memory by group

       W p-value
1 0.9347  0.4952
2 0.9575  0.7396

>
> # Avaliando a homocedasticidade
>
> library(car)
> leveneTest(memory ~ group, data = dadost)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.0159 0.9008
      19
```

```
> t.test(memory ~ group, var.equal = T, data = dadost)

        Two Sample t-test

data:  memory by group
t = 3.4705, df = 19, p-value = 0.002561
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.627355 6.572645
sample estimates:
mean in group 1 mean in group 2
           13.1             9.0
```

EXEMPLO DA LITERATURA

Ornam e colaboradores (2008), em um ensaio aleatório controlado, perceberam que a <u>meditação diminui o estresse e encoraja o perdão entre estudantes universitários</u>.

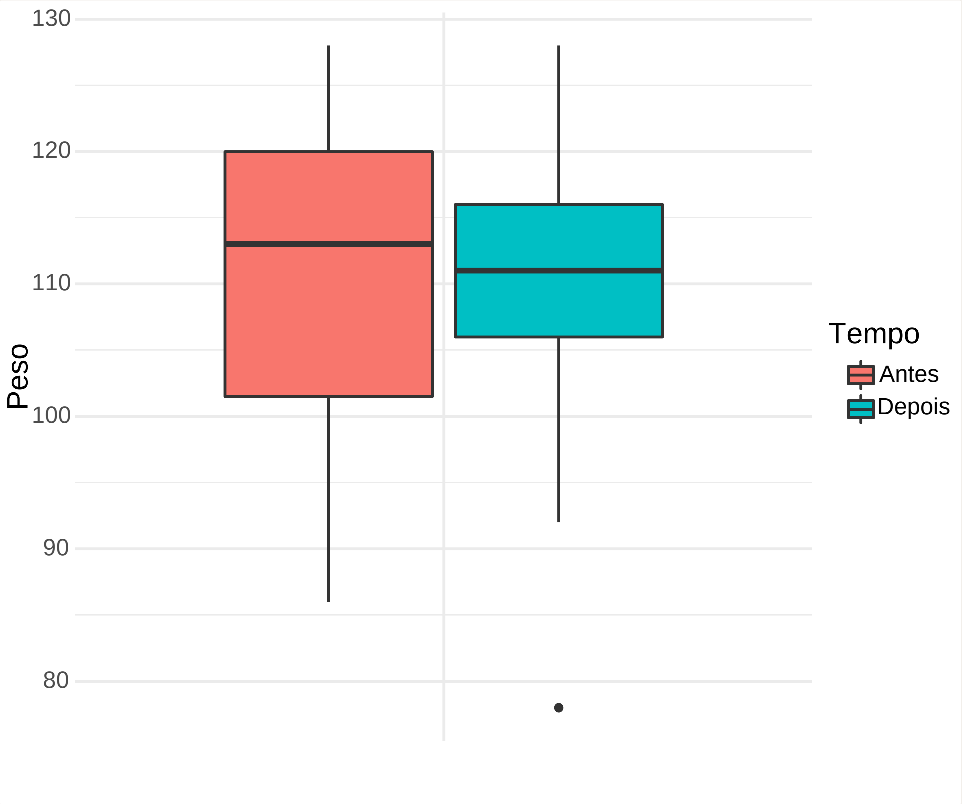
- O estudo foi organizado em dois grupos: estudantes que foram "tratados" por meio de aulas de técnicas de meditação e estudantes que aguardavam tratamento. Os autores dizem:

  Comparados aos controles, os participantes tratados (N = 29) demonstraram benefícios significativos para o estresse (p < 0,05) e perdão (p < 0,05).

TESTE T PAREADO

- Usado quando os mesmos sujeitos participam em ambas as condições.
- Compara cada participante consigo próprio, e, assim, esperamos que os dois valores estejam correlacionados. Isso reduz o erro da variância e leva a um teste mais sensível.

TESTE T PAREADO NO R



TESTE T PAREADO NO R

Para ilustrar o teste t pareado usaremos 20 participantes de uma dieta alimentar. Todos os participantes tiveram seus pesos mensurados antes e após a dieta.

```
> # Avaliando a normalidade
>
> shapiro.test(dadostpareado$peso_antes)

        Shapiro-Wilk normality test

data:  dadostpareado$peso_antes
W = 0.91958, p-value = 0.06526

> shapiro.test(dadostpareado$peso_depois)

        Shapiro-Wilk normality test

data:  dadostpareado$peso_depois
W = 0.94621, p-value = 0.2438
```

TESTE T PAREADO NO R

```
> # Teste t pareado
>
> t.test(dadostpareado$peso_antes, dadostpareado$peso_depois, alternative = "greater
", paired = T, data = dadostpareado)

        Paired t-test

data:  dadostpareado$peso_antes and dadostpareado$peso_depois
t = 0.39629, df = 22, p-value = 0.3479
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -3.333056       Inf
sample estimates:
mean of the differences
                      1
```

Embora a diferença entre os grupos esteja na direção esperada, os resultados mostraram que ela não foi estatisticamente significativa: t=0,39629 (22), valor-p = 0,3479.

EXEMPLO DA LITERATURA

Anzalone (2008) queria determinar se havia diferenças significativas na glucose do sangue amostrado no lóbulo da orelha em relação àquela retirada na ponta dos dedos.

- Cinquenta participantes forneceram amostras tanto da ponta dos dedos quanto do lóbulo da orelha. Isto é o que foi dito:

  Os resultados indicaram que o valor médio para a punção da polpa digital (M = 180, 14 , DP = 64,16) foi significativamente maior do que o resultado do lóbulo da orelha (M = 174,38, DP = 63,18, t(49) = 2,81, valor-p = 0,007). O tamanho do efeito padronizado (d) foi de 0,40 (tamanho pequeno). O intervalo de confiança de 95% da diferença média entre os dois resultados foi de 1,64 a 9,89.

TESTES NÃO PARAMÉTRICOS

- Os seguintes testes não paramétricos transformam os valores em dados ordenados.
- Desse modo, são resistentes a valores atípicos e à assimetria, o que os torna ideais para amostras pequenas e assimétricas.

MANN-WHITNEY PARA GRUPOS INDEPENDENTES

- Usa os postos (posições) em uma fórmula que calcula a estatística teste "U".
- Os escores são <u>postos atribuídos aos grupos em conjunto</u>. A partir deles, é encontrado um posto médio para cada grupo.
- Se não existem diferenças significativas entre as condições A e B, então os postos devem ser igualmente distribuídos nas duas condições.

Mann-Whitney U Test

Original data		Ranks				
					Control	Drug
Control	Drug	Control	Drug	count	12	11
11	34	4	21.5	rank sum	117.5	158.5
15	31	10	19.5	U	92.5	39.5
9	35	2	23			
4	29	1	17	α	0.05	
34	28	21.5	16	tails	2	
17	12	11	5.5	U	39.5	
18	18	12.5	12.5	U-crit	33	
14	30	8.5	18	sig	no	
12	14	5.5	8.5			
13	22	7	14			
26	10	15	3			
31		19.5				
17	23.90909	117.5	158.5			

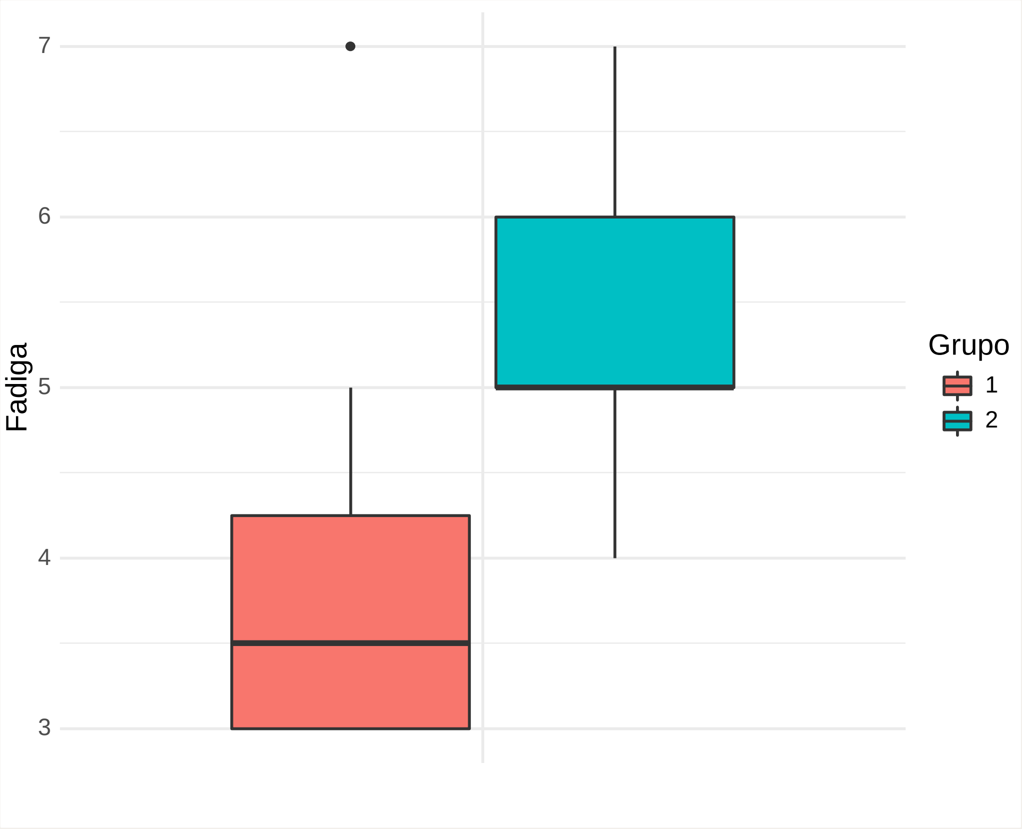
MANN-WHITNEY PARA GRUPOS INDEPENDENTES

Imagine que tenhamos uma amostra pequena de pacientes; oito tiveram um tratamento para melhorar a fadiga, e sete estão na lista de espera.

A fadiga é avaliada em uma escala de 1 (dificilmente com fadiga) até 7 (fadiga severa).

A hipótese é que a amostra tratada terá menos fadiga. Essa é uma hipótese direcional, assim poderemos usar um teste unicaudal.

MANN-WHITNEY NO R



MANN-WHITNEY NO R

```
> byf.shapiro(fatigue ~ group, data = dadosmann)

        Shapiro-Wilk normality tests

data:   fatigue by group

       W p-value
1 0.7705 0.01374 *
2 0.9366 0.60851
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>
> # Mann-Whitney
>
> wilcox.test(fatigue ~ group, alternative = "less", exact = F, data = dadosmann)

        Wilcoxon rank sum test with continuity correction

data:   fatigue by group
W = 10, p-value = 0.01904
alternative hypothesis: true location shift is less than 0
```

O teste de Mann-Whitney mostrou que o grupo tratado avaliou a si próprio como menos fatigado do que o grupo-controle ($valor - p = 0{,}01904$).

EXEMPLO DA LITERATURA

Blake e Batson (2008) tinham a seguinte pergunta de pesquisa:

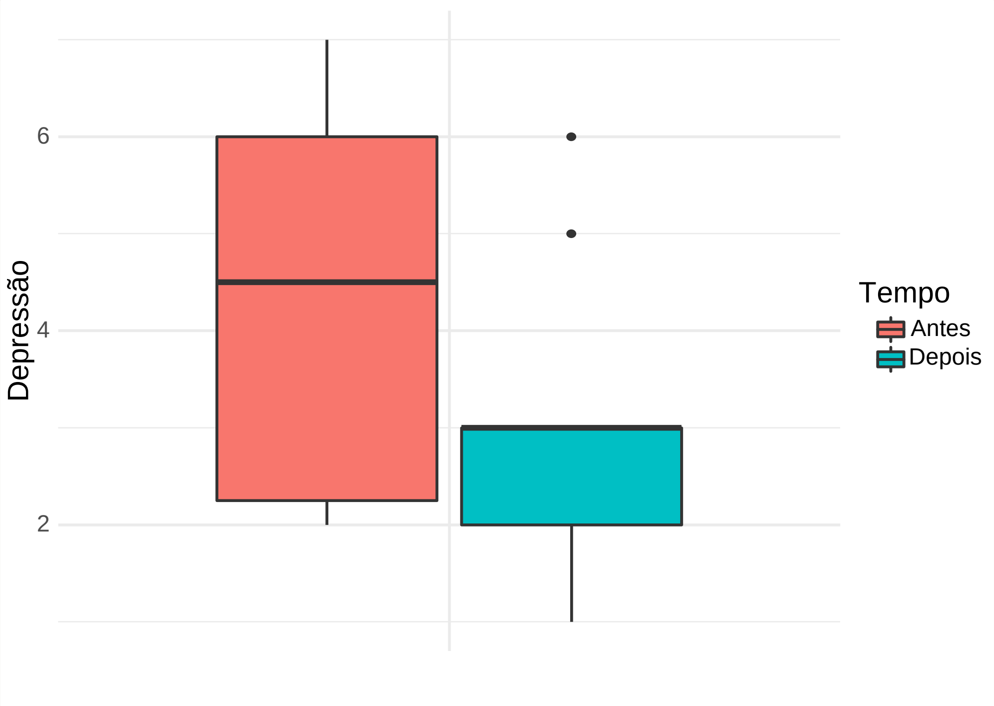
A participação em um exercício de intervenção Qigong melhora o humor, a autoestima, a flexibilidade perceptivo, a coordenação, a atividade física e o suporte social em pessoas com lesão cerebral traumática?

- Blake e Batson executaram o exercício de intervenção Tai Chi Qigong por uma hora semanal, ao longo de oito semanas, em 20 pessoas com lesão cerebral. Outras 20 pessoas participaram em uma atividade não baseada no exercício por oito semanas, agindo como controles.
- Para cada uma das variáveis medidas, os autores executaram um teste Mann-Whitney entre os grupos do exercício e de controle presentes no acompanhamento.

Os testes U de Mann-Whitney mostraram uma diferença significativa no humor avaliado pelo Questionário Geral da Saúde-12 entre os grupos do exercício e de controle. Não houve diferenças significativas entre os grupos nas medidas de Suporte Social para Hábitos de Exercícios de familiares e amigos de coordenação do Questionário de Autodescrição Física, de autoestima, de flexibilidade e de atividade física.

TESTES DOS POSTOS COM SINAIS DE WILCOXON PARA MEDIDAS REPETIDAS NO R

Vinte pacientes foram medidos em uma escala de depressão antes e depois da intervenção.



TESTES DOS POSTOS COM SINAIS DE WILCOXON PARA MEDIDAS REPETIDAS NO R

Vinte pacientes foram medidos em uma escala de depressão antes e depois da intervenção.

```
> wilcox.test(dadoswilcoxon$depression01, dadoswilcoxon$depression02, exact = F,
+             alternative = "greater", paired = T)

        Wilcoxon signed rank test with continuity correction

data:  dadoswilcoxon$depression01 and dadoswilcoxon$depression02
V = 26.5, p-value = 0.02071
alternative hypothesis: true location shift is greater than 0
```

Os resultados mostram que houve uma diferença estatisticamente significativa entre as duas condições (p = 0,02071), o que mostra que a intervenção surtiu efeito.

EXEMPLO DA LITERATURA

Taylor-Ford e colaboradores (2008) queriam determinar se o programa de redução do ruído diminuiu os níveis de som e distúrbios devidos ao som conforme percebido pelos pacientes em duas unidades hospitalares diferentes.

- Os pacientes, que completaram um questionário sobre distúrbios causados pelo som para pacientes durante as fases de pré-intervenção e pós-intervenção, forneceram os dados.
- Os pesquisadores escolheram Wilcoxon como o seu teste e alegaram o seguinte:

    Os escores possíveis variaram entre 30, indicando nenhum distúrbio de som, e 150, indicando graves distúrbios do som. A diferença entre médias e medianas da soma dos postos de Wilcoxon nos escores das duas unidades não foi significativa (p = 0,16), e a diferença entre os escores pré e pós na unidade de tratamento não foram significativos (p = 0,67).

AJUSTES PARA MÚLTIPLOS TESTES

- Muitos pesquisadores executam comparações com duas condições em mais de uma variável.
- Executar testes múltiplos dessa forma torna mais provável que um ou mais testes t sejam significativos apenas pelo erro amostral ou pelo acaso.
- É conveniente, então, “ajustar” o nível de significância para torna-lo mais estrito.
  - P. ex., se tivéssemos executado três testes t, então o ajuste seria dividir o nosso nível de significância padrão (0,05) pelo número de comparações feitas (3): $0{,}05/3 = 0{,}017$.
  - Isso significa que somente testes t com $p < 0{,}017$ seriam declarados estatisticamente significativos, embora relatemos isso como $p < 0{,}05$.
- Essa alteração do nível de significância é conhecida como correção de Bonferroni.

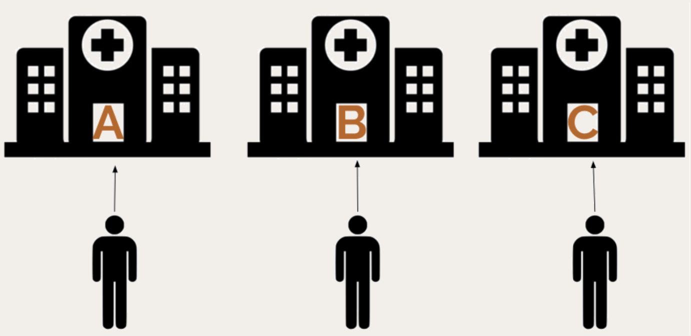
DIFERENÇAS ENTRE TRÊS OU MAIS CONDIÇÕES

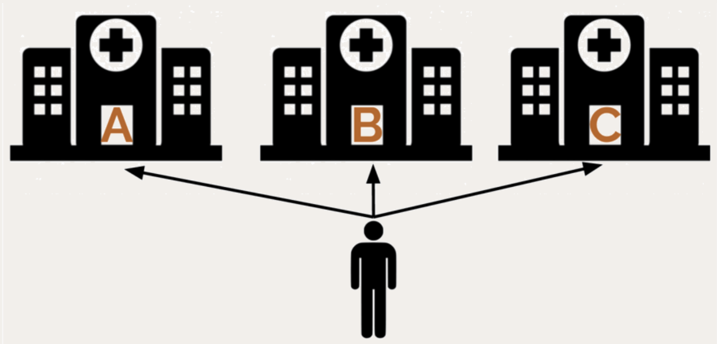
OBJETIVOS

- Serão observadas estatísticas que apontam se três ou mais condições ou grupos diferem entre si em uma ou mais variáveis.
  - Obter um entendimento conceitual da Análise de Variância (ANOVA);
  - Ser capaz de decidir quando usar a ANOVA — um teste paramétrico — ou testes não paramétricos equivalentes: Kruskal-Wallis ou Friedman;
- Aprender a executar Análises de Variância no R.

INTRODUÇÃO

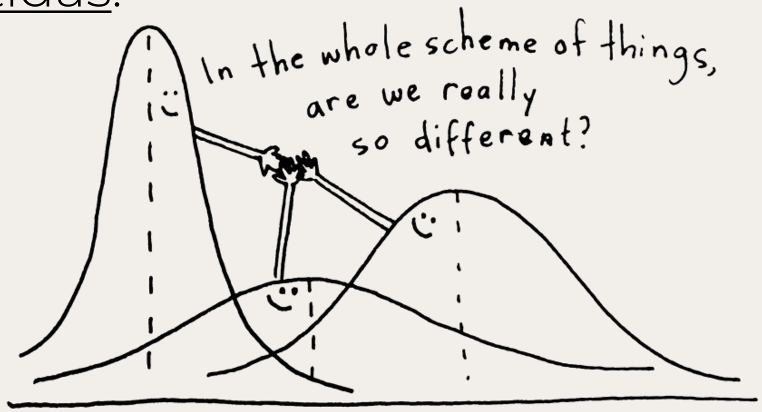
- Uma única variável com 3+ condições:
  - Hipótese experimental: existirá uma diferença significativa entre algumas ou todas as condições.
  - H0: quaisquer diferenças nos escores entre as condições devem-se ao erro amostral (acaso).
- Teste paramétrico para 3+ condições: Análise de Variância (*ANalysis Of Variance*, ANOVA).
  - Grupos independentes;
  - Medidas repetidas.





INTRODUÇÃO

- A ANOVA é recomendada quando:
  - Amostras aleatórias e independentes;
  - Populações têm distribuição normal (o teste é paramétrico);
  - Variâncias populacionais são iguais.
- Dados assimétricos → testes equivalentes não paramétricos: Kruskal-Wallis para grupos independentes e Friedman para medidas repetidas.



EXEMPLOS

Pense se o delineamento é de grupos independentes ou de medidas repetidas

- Scarpellini e colaboradores (2008) queriam determinar se havia diferenças significativas nos peptídeos citrulinados cíclicos (PCC) entre pessoas com artrite reumatoide (AR), com osteoartrite, com artrite psoriática e com outras condições de artrite.

- Button (2008) investigou a possível diferença entre fatores de estresse e níveis de saúde em diferentes níveis de enfermeiros (auxiliares, de equipe, chefe e parteiras).

EXEMPLOS

Pense se o delineamento é de grupos independentes ou de medidas repetidas

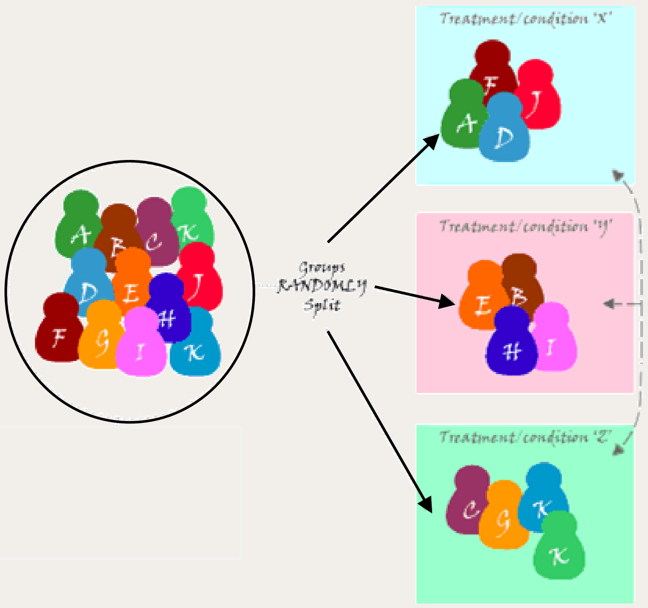
- Paterson e colaboradores (2009) queriam descobrir se a cafeína iria interromper o sono e também se duas drogas de indução ao sono iriam reverter a potencial interrupção. Doze homens saudáveis participaram em um estudo duplo-cego em que todos receberam placebo, cafeína, cafeína mais zolpidem e cafeína mais trazodone.

- Gariballa e Forster (2009) executaram um estudo sobre os efeitos do fumo no estado nutricional e na resposta aos suplementos alimentares durante uma doença grave. Como parte da pesquisa, o estado nutricional foi comparado entre fumantes, ex-fumantes e aqueles que nunca haviam fumado.

INTRODUÇÃO

A ANOVA mostra a existência de <u>quaisquer diferenças</u> significativas entre as condições. Para um delineamento de três condições, uma ANOVA estatisticamente significativa costuma apontar que:

a) a condição 1 pode ser significativamente diferente da condição 2; <u>ou</u>
b) a condição 1 pode ser significativamente diferente da condição 3; <u>ou</u>
c) a condição 2 pode ser significativamente diferente da condição 3.



DESCRIÇÃO CONCEITUAL DA ANOVA

A ANOVA investiga as <u>diferentes fontes em que surge a variação</u> nos escores.

Os dados são escores de um teste.

Grupo A	Grupo B	Grupo C
3	10	20
3	9	18
3	10	28
3	11	22
3	10	24
3	11	20
3	10	16
3	9	20
3	10	12

Qual grupo mostra a maior variabilidade?
Qual grupo mostra a menor variabilidade?

DESCRIÇÃO CONCEITUAL DA ANOVA

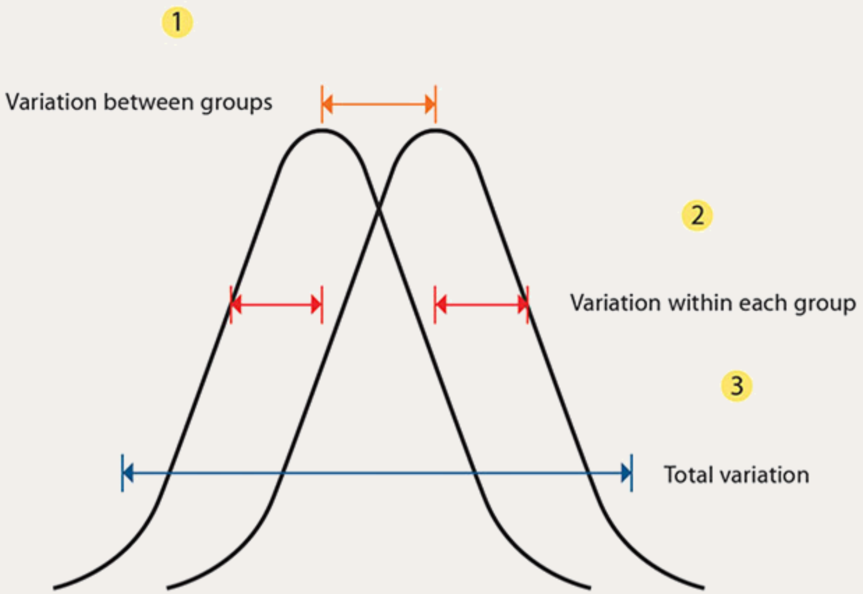
- A variação dentro dos grupos é chamada de variância intragrupos.
  - Indivíduos diferentes que nem sempre reagem da mesma forma aos tratamentos;
  - Erro devido a problemas nos métodos ou nas execuções de estudos e experimentos.
- A variância do erro está presente tanto na estimativa da variância intragrupos quanto entre grupos.

O que espera-se é que a variância intragrupos e a variância devido a erros experimentais sejam pequenas.

DESCRIÇÃO CONCEITUAL DA ANOVA

- A variação entre os grupos é chamada de variação entre grupos.

- Quando o experimento ou estudo são executados, espera-se que os grupos divirjam em virtude da intervenção ou de suas diferenças (e não apenas devido ao erro de medida, ao erro amostral ou a algo que os pesquisadores fizeram de errado).

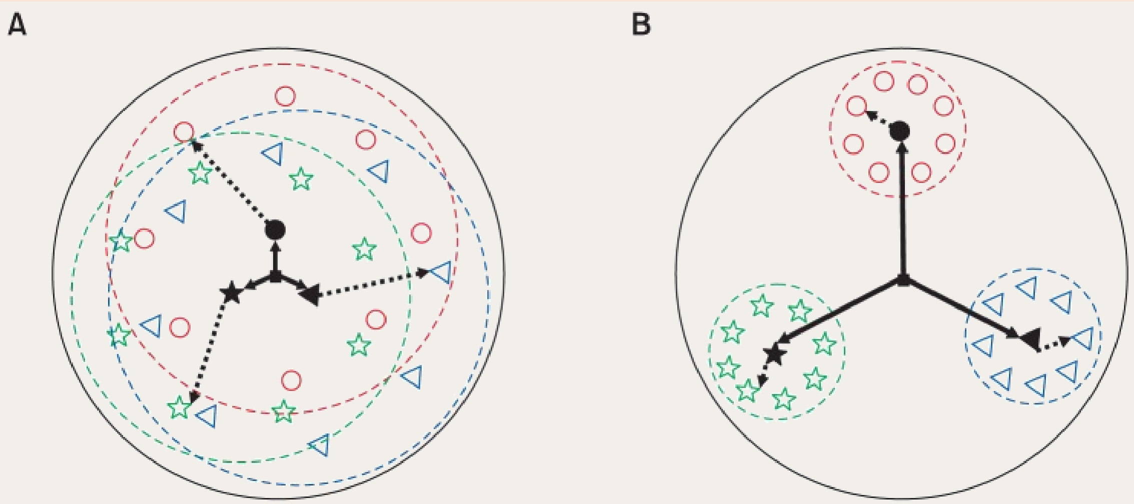


DESCRIÇÃO CONCEITUAL DA ANOVA

Ao executar-se uma ANOVA entre grupos, deve-se dividir toda a variância do conjunto de dados em variância intragrupos e variância entre grupos:

a) variância entre grupos (variância EG) — devido aos efeitos do tratamento, às diferenças individuais e ao erro experimental.

b) variância intragrupos (variância IG) — devido a diferenças individuais e erro experimental.



ANOVA DE UM FATOR

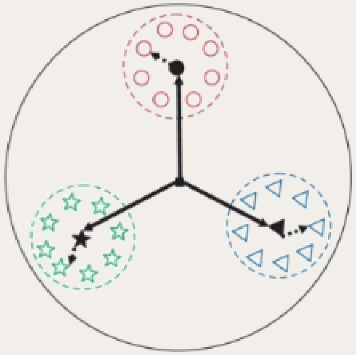
- A ANOVA de um fator → uma variável independente.
  - As variáveis independentes são chamadas de "fatores".
- A ANOVA de um fator pode ter três ou mais níveis do fator.

- P. ex., na testagem de três grupos diferentes de doença, o fator poderia ser chamado de "grupo da doença", e os três níveis poderiam ser "esclerose múltipla", "síndrome do intestino irritável" e "síndrome da fadiga crônica".

- P. ex., na testagem de quatro diferentes doses de medicações em pessoas com esclerose múltipla, o fator seria "níveis das medicações", e haveria quatro níveis, como, por exemplo, placebo, 5 mg, 10 mg e 15 mg.

ANOVA DE UM FATOR

- A variância total mede a dispersão de cada observação em torno da média geral de todas as observações, ignorando quaisquer grupos diferentes ou condições especificadas no delineamento.
- ANOVA de um fator: divide a variância total em variância EG e variância IG.
  - A estimativa da variância IG inclui diferenças individuais e erros de mensuração, enquanto a variância EG inclui também os efeitos do tratamento.
  - Portanto, se existe um efeito da variável independente, a estimativa da EG será maior do que a da IG.



ANOVA DE UM FATOR

- Se a variância EG for relativamente maior do que a variância IG, o valor F será maior; se a variância IG for maior, então o valor F será menor.
- O tamanho relativo da variância EG em relação à variância IG é expresso como uma razão, chamada de razão F.

$$Razão\ F^2 = \frac{estimativa\ da\ variância\ EG}{estimativa\ da\ variância\ IG}$$

EXEMPLO

- Suponha um curso preparatório para o ENEM que tenha em seu corpo docente três professores de matemática, que são responsáveis por diferentes turmas de alunos.
- A direção da escola suspeita que a variação do desempenho dos alunos nas provas de matemática do ENEM pode ser explicada pelo trabalho desenvolvido pelos seus professores.
- Sendo assim, a direção resolveu verificar as notas na prova de matemática dos alunos de cada professor e calculou a média das notas de cada turma.

http://www.abgconsultoria.com.br/blog/como-interpretar-uma-analise-de-variancia-anova/

EXEMPLO

- Mas será que essa informação é suficiente para afirmar que o desempenho dos alunos de cada professor é realmente diferente? E se um dos professores tiver em sua turma um aluno que não se preparou e errou quase todas as questões? Esse aluno não seria responsável por ter diminuído a média do grupo de alunos desse professor?
- Para verificar então se realmente o desempenho dos alunos variou de acordo com o professor, se faz necessário a utilização de teste estatístico, que além de considerar a média das notas, leva também em conta a variação das notas dentro de cada turma.

H0: Não existe diferença entre o desempenho das notas dos alunos de cada professor.
H1: Há pelo menos um professor com alunos com desempenho diferente.

EXEMPLO

Caso os três grupos de alunos apresentem mesma variabilidade e mesma média de desempenho, suas distribuições tendem a se sobrepor.

Quando os grupos apresentam mesma variabilidade interna e médias diferentes, as distribuições se distanciam quanto mais as médias se diferenciam.

EXEMPLO

Os grupos de alunos de cada professor podem ser vistos como três níveis de um mesmo fator, sendo que o objetivo é saber se o fator professor exerce alguma influência na variação do desempenho das notas de matemática.

Variação entre os professores de cada turma

Razão entre a soma de quadrados e os graus de liberdade

Análise de Variância das Notas dos Alunos por Turma

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	Estatística F	Valor P
Tratamentos	2	56.806	28.403	5,25	0,010
Resíduos	33	178.783	5.418		
Total	35	235.589	–		

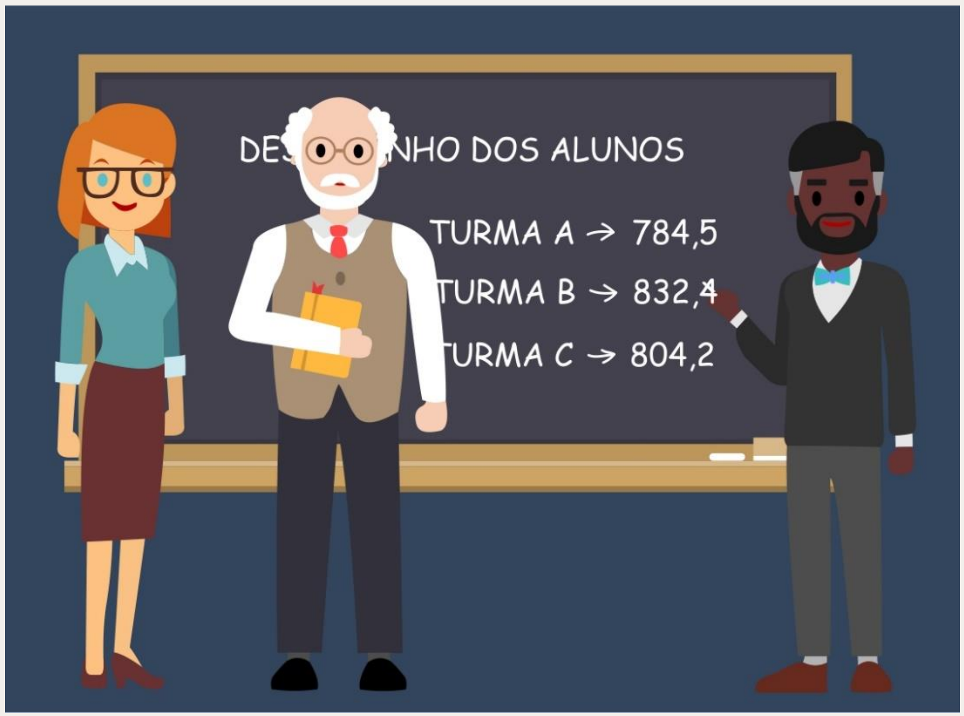
Variação dentro de cada turma

Variação total nos dados

TESTES DE MÚLTIPLAS COMPARAÇÕES

- O valor F geral não diz onde estão as diferenças significativas. Mas também não recomenda-se realizar cada comparação pareada, pois irá aumentar o erro do Tipo 1.
- Para saber quem se diferencia: teste *post hoc* - executa todas as comparações possíveis com o conjunto de médias.
  - Por exemplo, ao ter-se um fator com três condições (grupo I, grupo 2 e grupo 3), as comparações seriam: I versus 2; 1 versus 3; e 2 versus 3.
  - A probabilidade de se cometer um erro do Tipo I é controlada ao longo do conjunto de comparações.

TESTE DE TUKEY



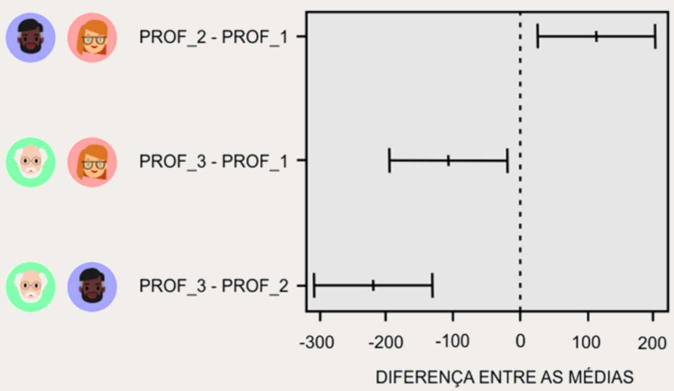
http://www.abgconsultoria.com.br/blog/comparacoes-multiplas-teste-de-tukey/

TESTE DE TUKEY

O módulo da diferença da média entre os pares de professores foi maior que o valor da D.M.S. obtido. Isso nos leva a concluir que o desempenho médio dos alunos dos professores (1 e 2), (1 e 3) e (2 e 3) são significativamente diferentes.

O valor 0 (zero) não está contido nos intervalos de confiança.

Diferença Mínima Significativa	Professores	Diferença	I.C. – 95%	Valor P
88,18	PROF_2 – PROF_1	113,75	[25,57 ; 201,93]	0,009
	PROF_3 – PROF_1	-107,08	[-195,27 ; -18,90]	0,014
	PROF_3 – PROF_2	-220,83	[-309,02 ; -132,65]	0,001



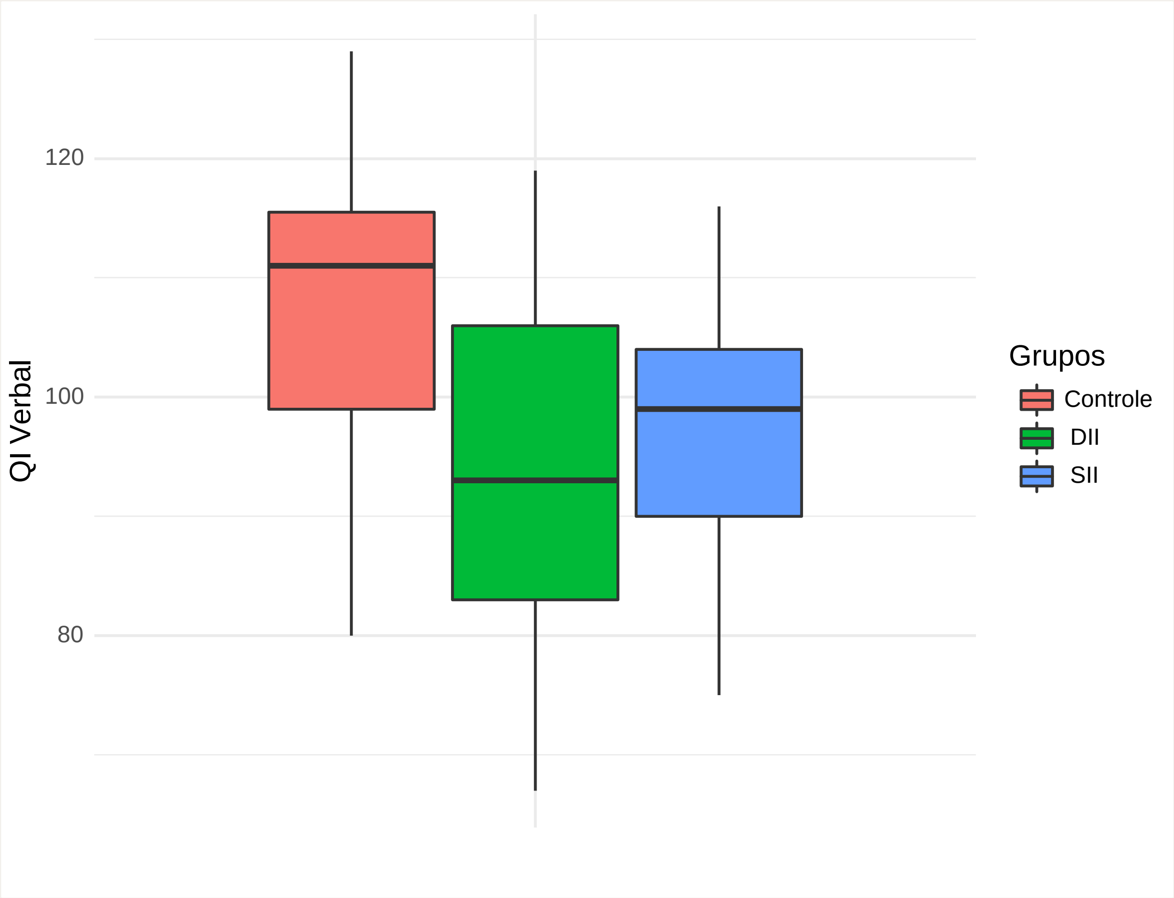
Notamos que todos eles são menores que o nível de significância adotado (valor-p < 0,05). Dessa maneira, chegamos a mesma conclusão baseada na D.M.S e nos intervalos de confiança.

ANOVA DE UM FATOR NO R

- Neste exemplo, busca-se descobrir se os grupos (doença inflamatória intestinal – DII; síndrome do intestino irritável – SII; e controle saudável – CS) diferem um dos outros.
- Previmos que as pessoas com <u>DII e SII teriam escores do QI Verbal mais baixos do que os controles saudáveis</u>, isso porque a literatura prévia mostrou que outros grupos de doentes tiveram QI verbais mais baixos do que os controles saudáveis.

H0: Não existe diferença dos escores do QI verbal entre as doenças.
H1: Há pelo menos uma doença com QI verbal diferente.

ANOVA DE UM FATOR NO R



ANOVA DE UM FATOR NO R

```
> # Avaliando a normalidade
>
> library(RVAideMemoire)
> byf.shapiro(viq ~ groups, data = dadosanova)

        Shapiro-Wilk normality tests

data:  viq by groups

              W p-value
controls 0.9726  0.6121
ibd      0.9710  0.5867
ibs      0.9625  0.3782

>
> # Avaliando a homocedasticidade
>
> library(car)
> leveneTest(viq ~ groups, data = dadosanova)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.8307 0.4393
      85
```

ANOVA DE UM FATOR NO R

```
> anova <- aov(viq ~ groups, data = dadosanova)
> anova
Call:
   aov(formula = viq ~ groups, data = dadosanova)

Terms:
                  groups Residuals
Sum of Squares   3365.397 12543.467
Deg. of Freedom         2        85

Residual standard error: 12.14785
Estimated effects may be unbalanced
> summary(anova)
            Df Sum Sq Mean Sq F value   Pr(>F)    
groups       2   3365  1682.7    11.4 4.1e-05 ***
Residuals   85  12543   147.6                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Diferença estatisticamente significante entre alguns grupos ou todos os grupos

ANOVA DE UM FATOR NO R

Teste *post hoc*

```
> tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = viq ~ groups, data = dadosanova)

$groups
                    diff        lwr       upr     p adj
ibd-controls -14.659770 -22.206115 -7.113426 0.0000379
ibs-controls -10.383908 -17.930252 -2.837564 0.0042248
ibs-ibd        4.275862  -3.334166 11.885890 0.3770103
```

Os controles tinham um QI significantemente mais altos do que ambos os grupos DII e SII.

ANOVA DE UM FATOR NO R

EXEMPLO DA LITERATURA

- Button (2008) queria investigar possíveis diferenças nos fatores de estresse e níveis de saúde entre diferentes graduações de enfermeiros. Havia 212 enfermeiros no estudo.
- Eles executaram uma ANOVA de um fator para determinar as diferenças no tempo de estresse entre as graduações do trabalho.

Havia uma diferença significativa nos níveis de tempo do estresse, com as irmãs enfermeiras relatando um escore mais alto do que os enfermeiros auxiliares e as parteiras [$F(4, 207) = 6,72$, $p < 0,005$]. Em adição às irmãs enfermeiras, os enfermeiros auxiliares também relataram tempo do estresse mais baixo do que os enfermeiros chefes.

ANOVA DE MEDIDAS REPETIDAS

- Todos tomam parte em todas as condições.
- Os participantes agem como os seus próprios controles.
  - Isso significa que a variância entre condições não pode incluir diferenças individuais.
  - Um participante com características peculiares aparece em todas as condições e irá afetar todas as suas médias da mesma maneira.
- A variância devido a diferenças individuais no conjunto de dados pode ser estimada e removida da equação:

$$F = \frac{variância\ EG}{variância\ IG\ (com\ diferenças\ individuais\ removidas)}$$

ANOVA DE MEDIDAS REPETIDAS

- São baseadas na mesma suposição de normalidade dos dados.
- Diferença: ANOVA de medidas repetidas presume que as variâncias das diferenças entre todos os pares de medidas devem ser similares → esfericidade.
- Comprova-se com a prova de esfericidade de Mauchly.
  - Caso o resultado seja significativo, ele indica, então, a violação da esfericidade.
  - Se os dados violam a condição de esfericidade, há algumas correções que ser aplicadas para conseguir um valor de F adequado ajustando os graus de liberdade:
    - Correção de Greenhouse-Geisser.
    - Correção de Huynh-Feldt.

TESTE DE KRUSKAL-WALLIS

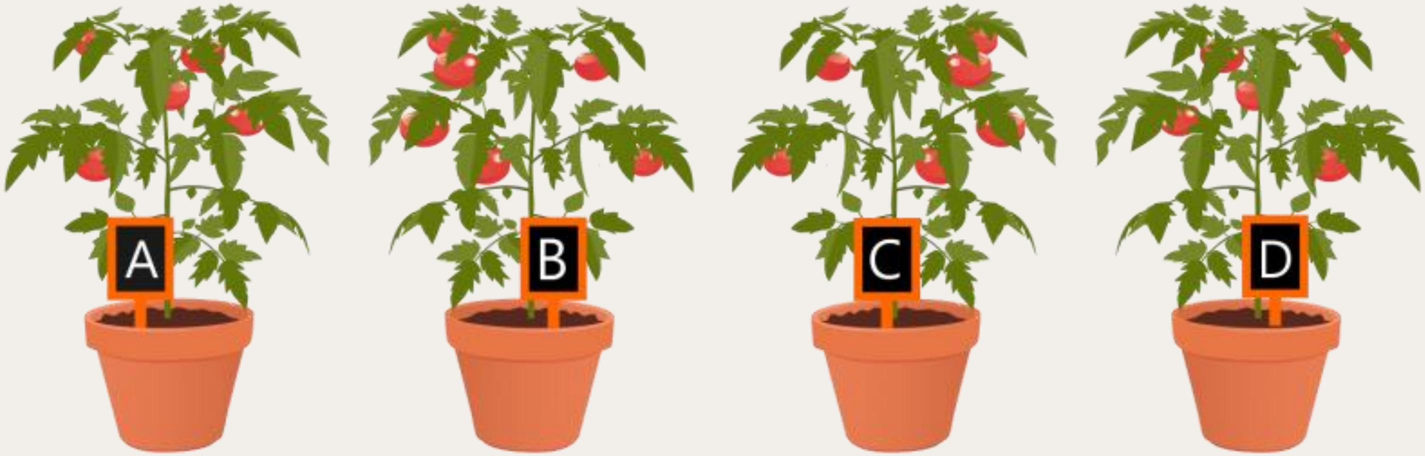
- Nas ciências da saúde/biológicas têm-se geralmente amostras pequenas e distribuições assimétricas. Em caso de dados significativamente assimétricos, use, então, os testes não paramétricos.
- O teste de Kruskal-Wallis é o equivalente não paramétrico da ANOVA paramétrica entre participantes.
  - Neste caso, devemos considerar apenas a suposição de que as observações sejam independentes e que as variáveis sejam do tipo contínuas ou ordinais.
  - O teste de Kruskal-Wallis é usado para três ou mais grupos, e é baseado na ordem dos escores, procurando por uma diferença significativa entre as ordens médias dos grupos.

TESTE DE KRUSKAL-WALLIS

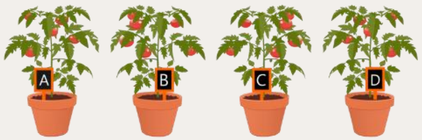
- A fórmula para o teste consiste em ordenar os escores de todas as condições. As ordens de cada grupo são, então, somadas, obtendo-se uma média das ordens de cada grupo. As medidas são, então, comparadas por meio do teste qui-quadrado.
- H0: os grupos têm ordens similares (uma vez que, se realmente não houver diferenças, as ordens devem ser distribuídas aleatoriamente nos diferentes grupos);
- H1: há uma diferença nas ordens dos grupos.

EXEMPLO

- Vamos supor um experimento que foi elaborado com o objetivo de verificar o comportamento do peso de tomates italianos cultivados em solos enriquecidos com diferentes tipos de nutrientes.
- Quatro solos foram enriquecidos com diferentes nutrientes e nomeados como “A”, “B”, “C” e “D”. Em seguida foram cultivados tomates italianos e, após o período de crescimento, foram colhidos e pesados seis tomates de cada solo.



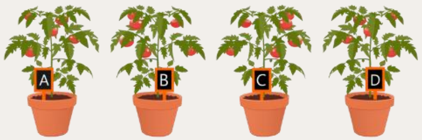
EXEMPLO

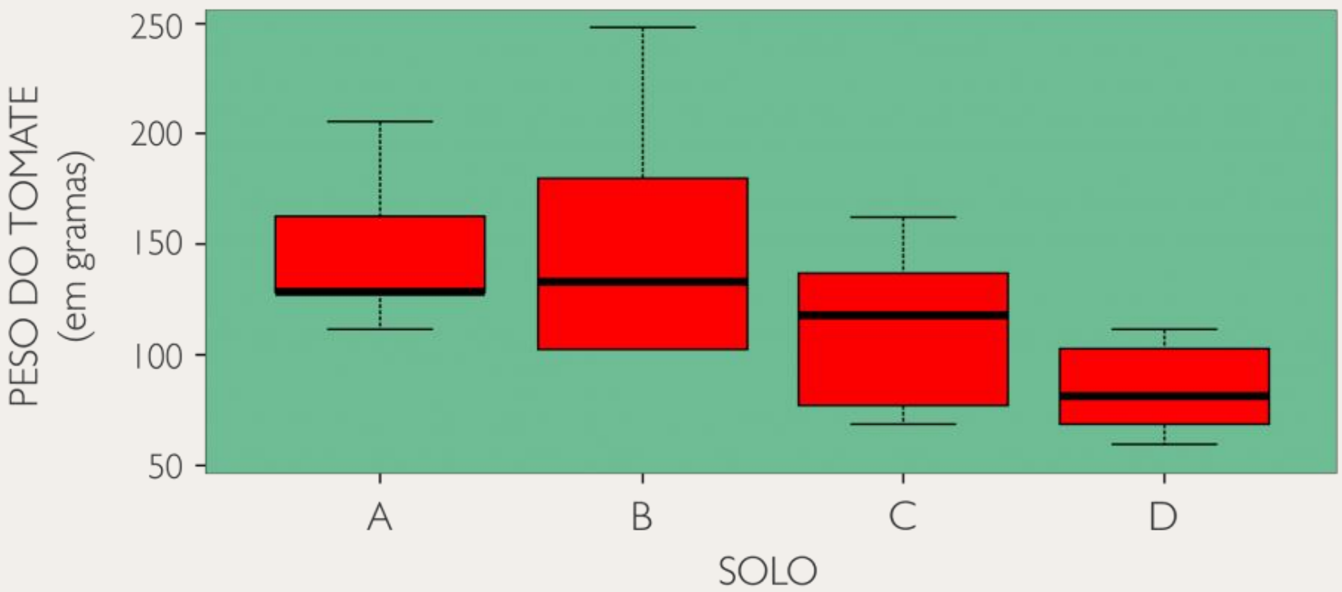


Grupo	Peso dos tomates italianos colhidos (em gramas)					
Solo “A”	128,5	162,8	111,4	128,5	205,7	128,5
Solo “B”	128,5	102,8	102,8	179,9	248,5	137,1
Solo “C”	162,8	137,1	68,6	98,5	77,1	137,1
Solo “D”	85,7	102,8	111,4	77,1	60	68,6

H0: Não existe diferença entre o peso do tomate italiano cultivado nos quatro solos.
H1: Há pelo menos um solo com o peso do tomate italiano cultivado diferente.

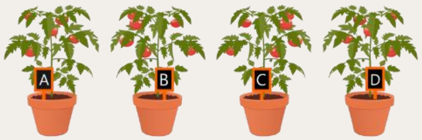
EXEMPLO





Essa aparente diferença do pesos dos tomates cultivados nos diferentes solos é significativa?

EXEMPLO



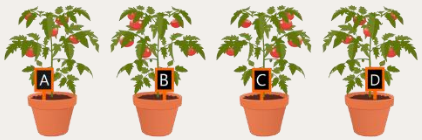

```
> kruskal.test(peso ~ cultivo, data = dadoskruskal)

        Kruskal-Wallis rank sum test

data:  peso by cultivo
Kruskal-Wallis chi-squared = 9.0028, df = 3, p-value = 0.02925
```

Mas como saber em quais solos o peso dos tomates se diferem?

Ao utilizar o teste de Kruskal-Wallis, o teste de comparações múltiplas adequado é o de Nemenyi.

O teste consiste em fazer comparações em pares com o intuito de verificar qual dos fatores que diferem entre si. No entanto, o teste de Nemenyi é muito conservador e pode não encontrar diferença significativa entre os pares testados.

EXEMPLO




```
> library(PMCMR)
> posthoc.kruskal.nemenyi.test(peso ~ cultivo, data = dadoskruskal)

        Pairwise comparisons using Tukey and Kramer (Nemenyi) test
                   with Tukey-Dist approximation for independent samples

data:  peso by cultivo

  A     B     C
B 1.000 -     -
C 0.663 0.713 -
D 0.044 0.055 0.456

P value adjustment method: none
```

TESTE DE FRIEDMAN

- O teste de Friedman é o equivalente não paramétrico à ANOVA de medidas repetidas e é uma extensão do teste de Wilcoxon para duas condições.
- Usado para três ou mais condições → pressupõe que a variável em análise seja medida em escala ordinal ou numérica.
- Para cada participante, as variáveis são ordenadas, e a soma dos postos dos participantes é calculada.
  - Funciona da mesma forma que o teste paramétrico da ANOVA de medidas repetidas, mas usando os postos dos escores dos participantes em vez dos próprios escores.
  - H0: postos das diferentes condições são similares;
  - H1: há diferenças nos postos entre cada grupo.

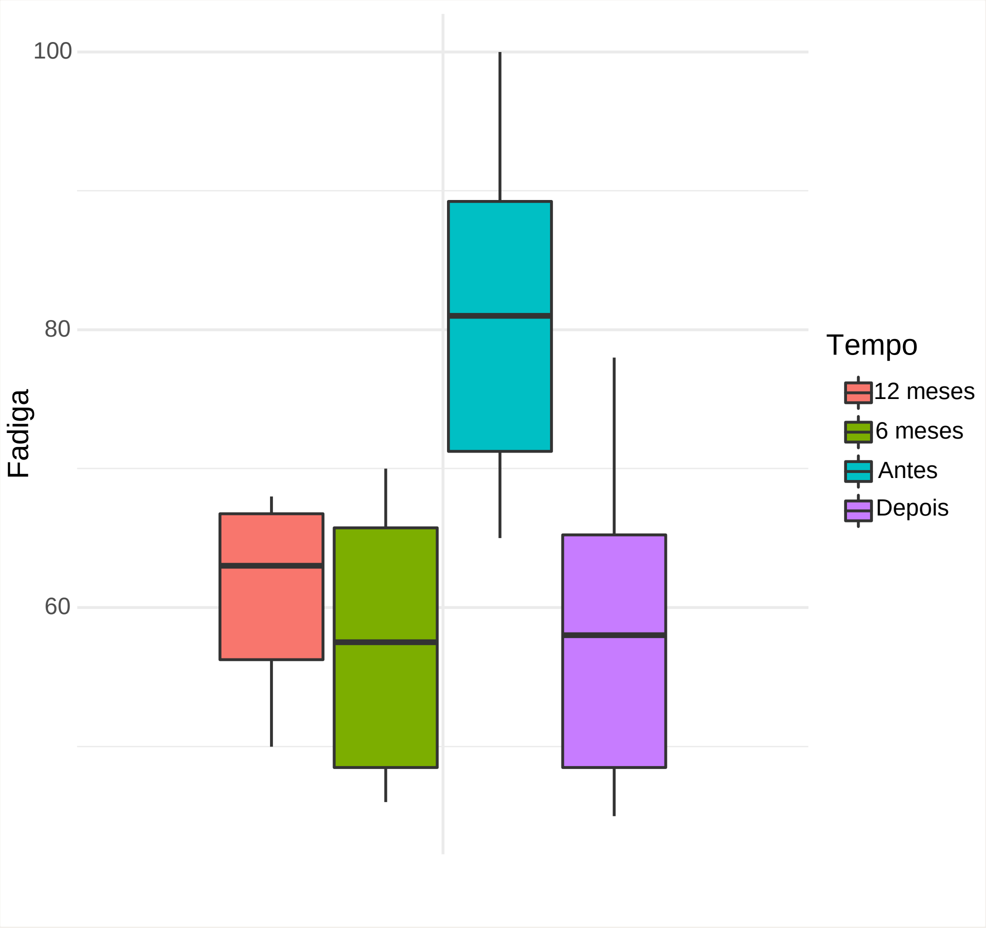
TESTE DE FRIEDMAN NO R

- Seis participantes com síndrome da fadiga crônica relataram quão exaustos se sentiam, respondendo um questionário de 20 itens. O valor total dos escores podia variar de 0 a 100.
- Foi feita uma intervenção, o que incluiu conselhos sobre diminuir o ritmo e sobre como relaxar e praticar meditação. O questionário foi aplicado novamente após a intervenção e depois de 6 e 12 meses

	Fatiguebefore	Fatigueafter	Fatigue6	Fatigue12
1	100	60	65	66
2	70	67	66	67
3	65	46	50	55
4	75	56	46	60
5	90	78	70	68
6	87	45	48	50

Fatigue before	Fatigue after	Fatigue 6	Fatigue 12
4	1	2	3
4	2,5	1	2,5
4	1	2	3
4	2	1	3
4	3	2	1
4	1	2	3
24	10,5	10	15,5

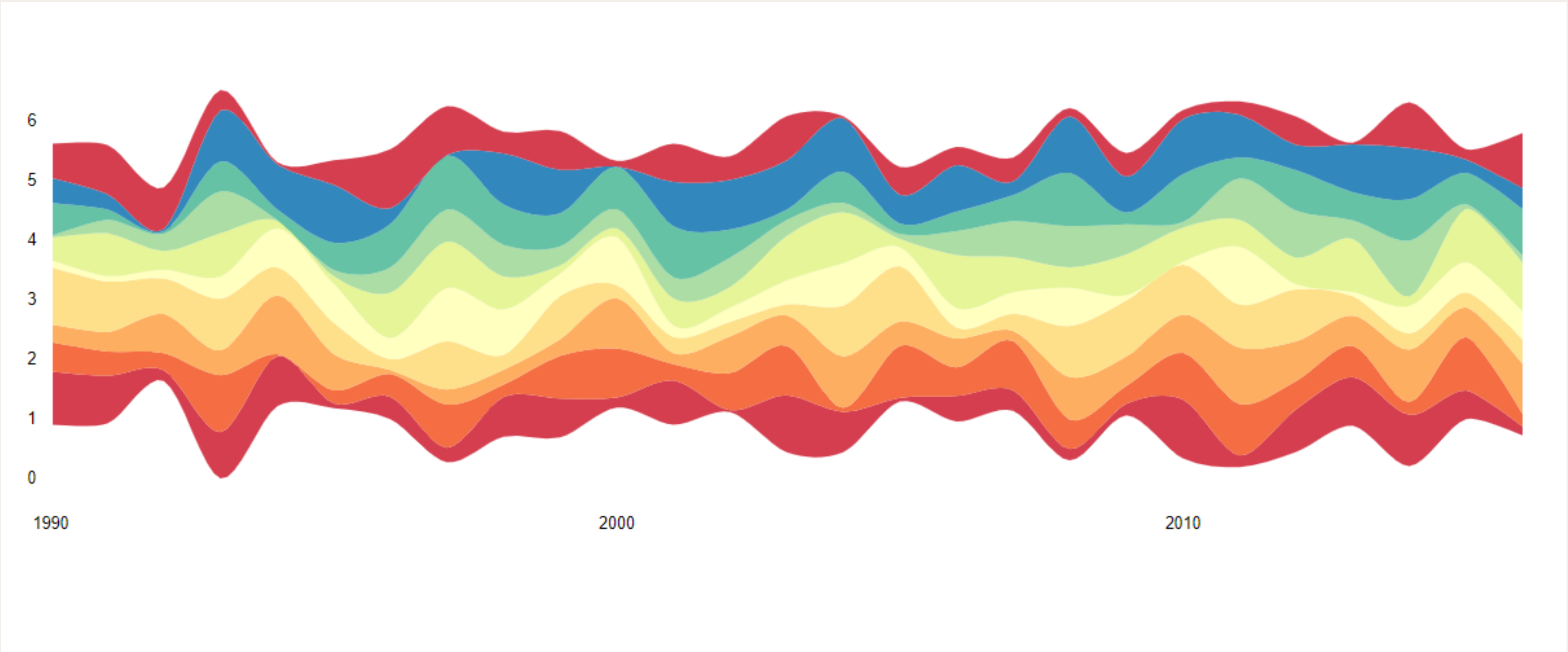
TESTE DE FRIEDMAN NO R



TESTE DE FRIEDMAN NO R

```
> friedman.test(as.matrix(dadosfriedman))

        Friedman rank sum test

data:  as.matrix(dadosfriedman)
Friedman chi-squared = 12.864, df = 3, p-value = 0.004939

> posthoc.friedman.nemenyi.test(as.matrix(dadosfriedman))

        Pairwise comparisons using Nemenyi multiple comparison test
             with q approximation for unreplicated blocked data

data:  as.matrix(dadosfriedman)

             Fatiguebefore Fatigueafter Fatigue6
Fatigueafter 0.0135        -            -
Fatigue6     0.0095        0.9995       -
Fatigue12    0.2277        0.6784       0.6078

P value adjustment method: none
```

ARTE DO DIA FEITA EM R



https://www.r-graph-gallery.com/154-basic-interactive-streamgraph-2.html

REFERÊNCIAS BIBLIOGRÁFICAS

- BARBETTA, Pedro Alberto. Estatística aplicada às ciências sociais. Ed. UFSC, 2008.

- DANCEY, Christine P.; REIDY, John G.; ROWE, Richard. Estatística Sem Matemática para as Ciências da Saúde. Penso Editora, 2017.

- MAGNUSSON, Willian E. Estatística [sem] matemática: a ligação entre as questões e a análise. Planta, 2003.