Especialização em *Data Science* e Estatística Aplicada

Módulo IV - Métodos Não Paramétricos

Profa. Dra. Teresa Cristina Martins Dias

Goiânia, 2025

IME INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
FEN FACULDADE DE ENFERMAGEM
UFG UNIVERSIDADE FEDERAL DE GOIÁS

Conteúdo Programático

- Introdução aos métodos não paramétricos: apresentação e discussão do uso.
- Revisão de conceitos básicos: população e amostra, tipos de variáveis, escala de medidas e medidas descritivas, testes de hipóteses e p-valor.
- Testes de hipóteses não paramétricos baseados na distribuição binomial.
- Testes baseados na distribuição qui-quadrado.
- Teste para duas e mais amostras pareadas.
- Teste para duas e mais amostras independentes.
- Testes de dispersão.
- Aplicações em dados na área da saúde na linguagem **R**.

Aula 1 - Introdução

Para realizar uma análise estatística ou fazer inferência estatística, aplicamos técnicas, métodos ou testes estatísticos, sendo que nestes existem várias suposições para que os resultados obtidos sejam válidos, incluindo análise descritiva adequada ao tipo de variável.

Duas situações:

a) Estamos interessados em verificar se existe diferença entre médias de duas amostras, pareadas ou independentes, ou ainda se uma amostra foi retirada de uma população com determinada média

(verificar se a diferença entre as médias de duas amostras é estatisticamente significativa ou se a diferença é "casual").

=> testes t-Student para comparação:

- para duas amostras independentes:
- para duas amostras dependentes (ou pareadas)
- para uma amostra (comparar com a média da população).

Suposições:

- independência entre as observações
- normalidade (ou *n* suficientemente)
- igualdade de variâncias (no caso de amostras independentes)
- variável quantitativa.

b) Estamos interessados em comparar várias amostras (três ou mais) para verificar se as médias são diferentes

(verificar se, pelo menos, um dos tratamentos produz resultados estatisticamente diferentes dos demais)

Suposições:
- variável quantitativa
- normalidade (ou *n* suficientemente grande)
- independência entre as observações
- igualdade de variâncias (amostras independentes)

=> ANOVA: análise de variância

… e se …

- a(s) variável(is) for(em) qualitativa(s)?

- a(s) amostra(s) for(em) pequena(s)?

- a(s) amostras(s) não tem(tiverem) distribuição normal ou (aproximadamente normal)?

⇩

Métodos Não Paramétricos

- A estatística não paramétrica pode ser definida como uma coleção de métodos e testes para inferência estatística, nos quais não fazemos (muitas!) suposições sobre a (forma da) população da qual os dados foram coletados.

- Suposições
  nos casos em que existem, estas são mais brandas.

- Em que situações pode ser aplicada?
  os métodos e testes são usados em substituição aos testes paramétricos usuais em diversas situações (não normalidade dos dados, variáveis qualitativas, *etc*.).

Uso de testes não paramétricos

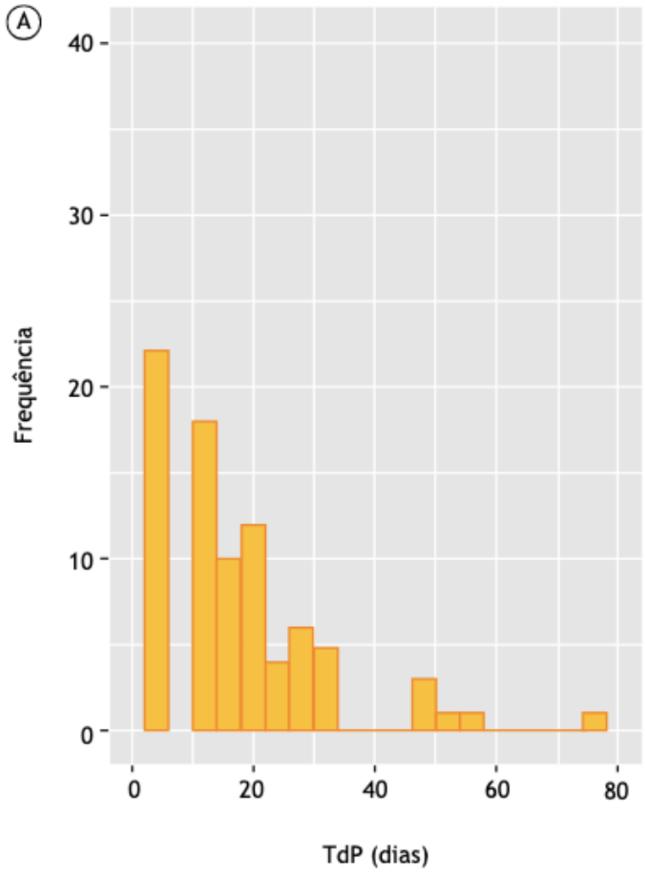
Vantagens	Desvantagens
Dispensa normalidade e, são mais eficientes do que os paramétricos no caso de não normalidade	Desperdício de informação (geralmente não são levadas em consideração a magnitude dos dados)
Pode ser usado para variáveis qualitativas e, úteis no caso de ser difícil estabelecer uma escala quantitativa de valores para os dados	Se as suposições necessárias são atendidas, muitos dos testes paramétricos são conhecidos por serem os mais poderosos. Nos estudos de comparações (para alguns testes) os não paramétricos são frequentemente quase tão poderosos quanto os paramétricos, especialmente para amostras pequenas
Pode ser aplicado para amostras pequenas	
Testes mais simples	A utilização das tabelas dos testes é mais complicada

Discussão:

A chefe de uma UTI gostaria de avaliar se pacientes obesos admitidos por exacerbação da doença pulmonar obstrutiva crônica (DPOC) têm um tempo de permanência (TdP) hospitalar mais longo do que pacientes não obesos.

Depois de selecionar **200** pacientes, ela descobriu que a distribuição do TdP é fortemente assimétrica a direita. Se ela fizesse um teste de hipótese, seria apropriado usar um teste t para comparar o TdP de pacientes obesos e não obesos com exacerbação de DPOC?

(Politi, M. T.; Ferreira, J. C.; Patino,C. M. "Testes estatísticos não paramétricos: mocinho ou bandido?. " Jornal Brasileiro de Pneumologia 47 (2021): e20210292).


A
Frequência
40
30
20
10
0
0
20
40
60
80
TdP (dias)

Por fim…

Nesta situação, como a distribuição do tempo de permanência (TdP) hospitalar é fortemente assimétrica à direita, a relação entre obesidade e TdP em pacientes hospitalizados por exacerbações de DPOC deve ser analisada com um **teste não paramétrico** (em vez de aplicar o teste t-Student).

Alguns testes não paramétricos

Uma amostra	para proporção, mediana, quantis, média (Wilcoxon)
Duas amostras dependentes	McNemar, do sinal, Wilcoxon e medidas de associação
Três ou mais amostras dependentes	Friedmam
Duas amostras independentes	Mann-Whitney
Três ou mais amostras independentes	Kruskal-Wallis

Revisão

1. População e amostra

2. Tipos de variáveis e escala de medidas

3. Testes de hipóteses e p-valor

1. População e Amostra

População Conjunto de todos os elementos (?) que contêm a(s) característica(s) de interesse.	**Amostra** Subconjunto da população.
Parâmetro Quantidade relacionada à população - usada para descrever a população. Exemplos	**Estatística** Qualquer função dos elementos da amostra. Exemplos

A amostra deve ser:

- representativa (?)
- suficiente (?) e,
- aleatória (?).

2. Tipos de variáveis (características) e escala de medidas

qualitativas: nominal ou ordinal

quantitativas: discreta ou contínua

Nos métodos não-paramétricos o primeiro critério a ser verificado é a escala de mensuração.

Com esta podemos determinar a ("quantidade" de) informação contida nos dados e quais as maneiras de resumi-los - análises estatísticas apropriadas.

↓

Escala de medidas

Níveis de mensuração:

- **nominal**: as respostas são nomes ou números (que não representam quantidades) => as operações aritméticas não fazem sentido

- **ordinal**: as respostas são nomes ou números (que não representam quantidades), como na escala nominal, porém podem ser ordenadas, segundo algum critério => as operações aritméticas não fazem sentido

- **intervalar**: as respostas são numéricas sem ponto zero natural (o zero é arbitrário) e os intervalos (diferenças) são significativos, mas as razões não são (temperatura de 0ºC não significa falta de temperatura, e 20ºC não significa o dobro de calor do que 10ºC)

- **razão**: as respostas têm as propriedades da escala intervalar, e possuem um zero verdadeiro (representa uma origem absoluta e não há valor numérico negativo – 0 significa "falta de"); tanto os intervalos como as razões são significativos => todas as operações aritméticas podem ser aplicadas.

Por fim,

- uma variável pode ser especificada de formas diferentes => importância da escala: como queremos obter as informações (quanto de informação queremos extrair no estudo!)

- a escala de razão é a de nível mais alto:

(mais informativa) **razão => intervalar => ordinal => nominal** (menos informativa)

3. Testes de hipóteses e p-valor: contexto de inferência estatística

- Hipótese estatística: declaração a respeito de um ou mais parâmetros. Esta é testada com evidências contidas na amostra:
  - se rejeitada => a evidência obtida na amostra não é suficiente para afirmarmos, com certo grau de confiança, que a hipótese é verdadeira
  - se aceita => a evidência obtida na amostra é suficiente para afirmarmos, com certo grau de confiança, que a hipótese é verdadeira.

- Teste de hipótese (t.h.): para responder uma questão estatística, tal questão é **transformada** em uma hipótese, usando parâmetros – declaração que podemos testar, via teste estatístico, cujo resultado nos leva a aceitar, ou não, a hipótese.

- Teste estatístico: uma função da amostra e do parâmetro (em questão) usado para a tomada de decisão.

(observe que no teste fazemos inferência, isto é, usamos a estatística para inferir a respeito do parâmetro e, aceitamos ou não a declaração a respeito da população)

- Como realizar um t.h.?

a) construir as hipóteses
- H_0: hipótese a ser testada
  (se desejamos determinar se a declaração a respeito da população é falsa, fazemos desta a hipótese nula)

- H_1: hipótese alternativa: complementar à H_0
  (se desejamos determinar se a declaração a respeito da população é verdadeira, fazemos desta a hipótese alternativa)

b) selecionar o teste estatístico: relacionado à quantidade populacional.

c) aplicar o teste e tomar a decisão: rejeitar ou não H_0.

d) Pontos importantes:
- região crítica (r.c.): valores no espaço amostral que nos levam a rejeitar H_0.
- nível de significância: $\alpha = P[\text{rej. } H_0 \mid H_0 \text{ verd.}]$
- nível crítico: **p-valor**: menor nível de significância no qual H_0 seria rejeitada (medida intuitiva da força dos dados contra H_0).

Observações:

- As hipóteses são especificadas em termos de parâmetros e a escolha do teste depende do parâmetro a ser testado e de como os valores amostrais foram registrados (natureza da variável).

- Usando o teste estatístico e definindo a região crítica (ou calculando o p-valor), tomamos uma decisão => não existe evidência suficiente na amostra para rejeitar/aceitar H_0.

- Não rejeitar a hipótese nula não significa dizer (ou afirmar) que a hipótese alternativa é válida com o mesmo nível de confiança.

Quais os tipos de hipóteses?

Alguns tipos de hipóteses:

Denotando por θ o parâmetro de interesse (média, variância, proporção, mediana, *etc.*) e θ_0 um valor especificado de θ:

a) H_0: $\theta = \theta_0$ *versus* H_1: $\theta \neq \theta_0$

b) H_0: $\theta \geq \theta_0$ *versus* H_1: $\theta < \theta_0$

c) H_0: $\theta \leq \theta_0$ *versus* H_1: $\theta > \theta_0$

(observe que o sinal “=“ está sempre em H_0).

Exemplos:

1. Produção de peças: uma amostra foi selecionada a fim de verificar se a máquina, que produz as pecas, está sob especificação, isto é, se esta produz, no máximo, 5% de itens que apresentam algum tipo de defeito.

H_0: a máquina está em condições de se operada
H_1: a máquina não está em condições de se operada

=> H_0: $p \leq 0{,}05$ *versus* H_1: $p > 0{,}05$.

2. A fim de verificar se existe diferença de comportamento entre crianças que passaram pela educação infantil (até cinco anos) e crianças que entraram (direto) na educação básica (a partir de cinco anos), uma escola aplicou um teste nas crianças.

H_0: não existe diferença comportamental
H_1: existe diferença comportamental.

Alguns exemplos para discussão

- Tipos de variáveis e escalas de medidas

- Testes de hipóteses

Exemplos para discutir: tipos de variáveis e algumas medidas descritivas apropriadas

- queixa de dor corporal
- grupo sanguíneo
- grau de edema
- escolaridade
- incapacidade física
- quantidade de nódulos pulmonares
- frequência cardíaca
- peso
- altura
- largura
- comprimento
- rendimento
- diâmetro abdominal

Possíveis classificações

- queixa de dor corporal => sim ou não => nominal: distribuição de frequência e moda

- grupo sanguíneo: A, B, O, AB => nominal

- grau de edema => 0: sem edema; 1: leve; 2: moderado; 3: grave; 4: muito grave => ordinal: distribuição de frequência, gráfico de barras ou de setores, moda e quantis (classes ordenadas)

- escolaridade: 1° grau, 2° grau, 3° grau => ordinal

- incapacidade física: leve, moderada, severa => ordinal

- quantidade de nódulos pulmonares, frequência cardíaca => discreta (intervalar): distribuição de frequência; média, variância e quantis

- peso, altura, largura, comprimento, rendimentos, diâmetro cervical e abdominal: contínuas (razão): distr. de frequência; média, variância, quantis e coef. de variação.

Escreva H_0 e H_1 para as questões abaixo:

1) Uma empresa quer verificar se "uma pausa para um cafezinho", melhora a produtividade de seus funcionários. Para tanto, implementou a seguinte medida:

   a) sorteou 20 funcionários (de uma seção) que tiveram direito a 15 minutos de intervalo (na parte da manhã e da tarde) e os outros 20 (de outra seção) que não tiveram. O número de peças produzidas foi registrado, por funcionário. Existe melhora na produtividade?

   b) na seção com os 40 funcionários, todos tiveram direito a 15 minutos de intervalo em uma semana e, na seguinte não tiveram. O número de peças produzidas foi registrado, por funcionário. Existe melhora na produtividade?

2) Um novo medicamento para dores musculares é testado para verificar se é melhor do que o usual. Participantes do estudo foram acompanhados e divididos aleatoriamente em dois grupos: um grupo tomou o medicamento usual e o outro tomou o novo medicamento. Os tempos até a "melhora" dos pacientes foram registrados. Existe diferença entre os medicamentos?

Duas aplicações

a) Partindo da hipótese de que os exercícios inspiratórios com exercitadores e incentivadores respiratórios poderiam impactar positivamente na voz, um estudo foi com realizado com o objetivo de avaliar o efeito imediato destes exercícios, nos dados perceptivo-auditivos, acústicos e aerodinâmicos.

- 25 mulheres sem queixas vocais, com idade entre 18 e 34 anos e média (aproximada) de 21 anos.

- a coleta de dados foi realizada nos momentos **antes e após** a realização de exercício inspiratório: gravação de vogal sustentada /a/, fala encadeada e tempos máximos fonatórios de vogais, fonemas fricativos e contagem de números.

- foi utilizada a Escala de Desvio Vocal para verificar o grau geral do desvio vocal.

(Beraldo, A. T., *et al.*, Efeito imediato de exercício inspiratório com exercitador e incentivador respiratório em mulheres sem queixas vocais. *CoDAS*. Vol. 36. No. 4. Sociedade Brasileira de Fonoaudiologia, 2024).

Para discutir: quais as são variáveis, escala de medidas e as hipóteses?

- **Variáveis de estudo**
  - dependentes (desfechos): o julgamento perceptivo-auditivo (JPA), as medidas acústicas e aerodinâmicas da emissão
  - independente: foi a realização do exercício com exercitador e incentivador respiratório.

- **Análise**
  - teste t-Student: amostras dependentes (antes e depois e, com distribuição normal)
  - teste de Wilcoxon (antes e depois e, distribuição não normal).

Resultados

- não houve diferenças entre os resultados do JPA e das medidas acústicas, nos momentos pré e pós exercício inspiratório. Quanto às medidas aerodinâmicas foi possível observar aumento significativo no valor do tempo máximo fonatório (TMF) /s/ ($p=0,008$).

Conclusão

- não houve modificação na qualidade vocal após o exercício inspiratório com incentivador e exercitador respiratório, porém foi observado aumento do TMF do fonema /s/ após a realização do exercício.

b) Um estudo foi realizado com o objetivo de analisar os efeitos da dança Sênior® nos aspectos cognitivos em pessoas idosas de 60 a 85 anos. O ensaio clínico foi não randomizado, controlado, e composto por dois grupos: (G1) pessoas idosas com declínio cognitivo, institucionalizadas e (G2) pessoas idosas sem declínio cognitivo, não institucionalizadas:

- amostra: G1: 15 pessoas com comprometimento cognitivo leve e demência leve e, G2: 32 pessoas
- a coleta de dados foi realizada em dois momentos: **Pré-intervenção e Pós-intervenção.**

(Piazentin, G. X. P. *et al.*, Efeitos da Dança Sênior® modalidade sentada nas funções cognitivas em pessoas idosas com e sem declínio cognitivo: ensaio clínico controlado, *Revista Brasileira de Geriatria e Gerontologia*, 27, 2024, e240036).

Para discutir: quais as são variáveis, escala de medidas e as hipóteses?

- Variáveis de estudo

- alguns domínios cognitivos de atenção e cálculo e linguagem foram avaliados, nos dois grupos (obtidos via exames: MEMM e BBRC)
- o estadiamento do declínio cognitivo das pessoas idosas (obtidos via escala Clinical Dementia Rating - CDR).

- Análise

- teste de Wilcoxon: comparação entre os momentos pré-intervenção e pós-intervenção (dentro de cada grupo).
- teste de Mann-Whitney: comparação entre grupos (considerando a diferença entre os momentos pré e pós-intervenção).

- Resultados

- teste de Wilcoxon: apontou diferença estatisticamente significante em vários domínios cognitivos avaliados no MEEM e BBRC em ambos os grupos, abrangendo os domínios atenção e cálculo e linguagem no G1, e linguagem no G2 pelo MEEM.

- Conclusão

- os autores concluíram que a DS proporcionou impactos positivos nas funções cognitivas dos participantes de ambos os grupos, contribuindo para a prevenção e promoção da saúde.

Testes baseados na distribuição binomial

- teste para uma amostra:
  - binomial
- testes para duas amostras (pareadas):
  - do sinal
  - de McNemar
  - de Wilcoxon

Algumas situações:

- um medicamento é aplicado em animais com determinada doença, e verificamos se o tratamento deu resultado (cura do animal) – ou não. A empresa que fabrica o medicamento afirma que 95% dos animais são curados com a aplicação do medicamento. Como confirmar esta afirmação?

- alunos matriculados em disciplinas, ao final do período são classificados como aprovado ou reprovado. Historicamente em determinada disciplina, a proporção de aprovados é, pelo menos, maior do que a de não aprovados. Como confirmar esta afirmação?

- certa indústria de refrigerantes desenvolveu um produto. O gerente de vendas gostaria de saber se o novo produto é tão preferido quanto o tradicional. Como verificar esta "dúvida"?

- em uma plantação (planta específica) foram observadas ao acaso dez plantas e, entre as dez apenas uma estava infectada com certo fungo. Se a proporção de infestação estiver abaixo de 30%, esta é considerada controlada. Como confirmar esta situação?

O que existe em comum nestes exemplos?

- uma **variável qualitativa** com duas classes de respostas (interesse/não interesse ou sucesso/fracasso)

- informação populacional sobre a quantidade de interesse (**proporção!**)

- objetivo: verificar se a proporção de sucesso (da classe de interesse) observada na amostra indica que esta pode pertencer a uma população com determinada **probabilidade de sucesso**.

Como “contamos” os “sucessos” na amostra e temos as mesmas suposições da **distribuição binomial**, usamos esta distribuição para realizar o teste de hipótese.

Teste binomial

Considere **uma** variável categórica dicotômica observada em uma amostra com *n* observações (unidades amostrais: peças, pessoas, pacientes, *etc*.), cujo objetivo, no estudo, é fazer inferência a respeito da probabilidade de sucesso.

- **Amostra**
  - resultado de ***n*** ensaios (realizações), cujas respostas pertencem ou à classe 1 ou à classe 2 (S/F, S/N, 0/1)
  - obtemos o número de observações na classe 1 (O_1) e número de observações na classe 2 ($O_2 = n - O_1$).

- **Suposições**
  - os *n* ensaios são independentes
  - todos os ensaios têm mesma probabilidade p de pertencer à classe 1 ("sucesso").

- **Hipóteses**

a) H_0: $p = p_0$ *vs* H_1: $p \neq p_0$

b) H_0: $p \leq p_0$ *vs* H_1: $p > p_0$

c) H_0: $p \geq p_0$ *vs* H_1: $p < p_0$

- **Teste estatístico**
  - como o interesse está nos resultados da classe "sucesso" => $T = O_1$.

- **Região crítica (r.c.) e a decisão**

  dependem do tipo de hipótese (unilateral ou bilateral). Nos três casos, usamos valores da distribuição binomial, com parâmetros *n* e p_0, para determinar a r.c. e tomar uma decisão.
  Definindo um nível de significância α:

  a) (teste bilateral): a r.c. de tamanho α corresponde às duas caudas da distribuição binomial(n, p_0).
  As caudas inferior e superior têm tamanhos α_1 e α_2, respectivamente ($\alpha_1 \approx \alpha_2$ e $\alpha = \alpha_1 + \alpha_2$).
  Denotando Y uma v.a. com distribuição binomial(n e p_0), encontramos os valores t_1 e t_2 que satisfazem:

$$P[Y \leq t_1] = \alpha_1 \text{ e } P[Y > t_2] = \alpha_2 \text{ (ou } P[Y \leq t_2] = 1 - \alpha_2)$$

  e, rejeite H_0 se $T > t_2$ ou se $T \leq t_1$. Caso contrário, aceite H_0

  b) como valores grandes de T indicam que H_0 é falso,

$$P[Y > t] = \alpha \text{ (ou } P[Y \leq t] = 1 - \alpha) \Rightarrow \text{ rejeite } H_0 \text{ se } T > t$$

  c) como valores pequenos de T indicam que H_0 é falso,

$$P[Y \leq t] = \alpha \Rightarrow \text{ rejeite } H_0 \text{ se } T \leq t.$$

Observação:

- se n > 20 utilizamos a aproximação pela distribuição normal. O valor exato de p é obtido para amostras pequenas. No R, já existe esta implementação.

Aplicação: “Comportamento anoréxico e percepção corporal em universitário”

(Gonçalves, T. D., et al. "Comportamento anoréxico e percepção corporal em universitários" Jornal Brasileiro de Psiquiatria, v. 57, pg. 166-170, 2008)

Um estudo foi realizado para avaliar a percepção corporal e a prevalência de sintomas de anorexia nervosa em estudantes universitários:

- amostra: participaram do estudo: 149 estudantes de Nutrição e 78 estudantes de Educação Física.

- método: o questionário Teste de Atitudes Alimentares (EAT- 26) e Teste de Imagem Corporal foram aplicados para avaliar os sintomas de anorexia nervosa e a percepção corporal, respectivamente.

- Dados:

	Insatisfação corporal			
Sintomas de Anorexia Nervosa	Nutrição		Educação Física	
	Sim	Não	Sim	Não
Sim	20	1	8	0
Não	93	35	53	17

Considerando os resultados da tabela acima e as afirmações:

- de acordo com dados da Organização Mundial da Saúde (OMS), 4,7% dos brasileiros sofrem de compulsão alimentar, anorexia nervosa e
  transtorno alimentar, quase o dobro da média global, que é de 2,6% (22/7/24)
- insatisfação corporal atinge mais de 50% dos brasileiros,

especifique o tamanho amostral e estatística teste, discuta, construa e teste as hipóteses, para os casos abaixo:

a) os estudantes de Nutrição tem tendência a apresentarem comportamento de risco para anorexia nervosa
b) os estudantes de Educação Física têm tendência a apresentarem proporção elevada de insatisfação corporal
c) as duas turmas (juntas) têm tendência a apresentarem comportamento de risco para anorexia nervosa

d) as duas turmas (juntas) têm tendência a apresentarem proporção elevada de insatisfação corporal
e) os estudantes de Nutrição têm tendência a apresentarem proporção elevada de insatisfação corporal
f) os estudantes de Educação Física têm tendência a apresentarem comportamento de risco para anorexia nervosa.

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste binomial

Exemplos apresentados na aula 04: testes baseados na distribuiçao binomial

1. Um medicamento foi aplicado em animais com determinada doença e foram registrados: cura do animal ou não cura. A empresa que fabrica o medicamento afirma que 95% dos animais são curados com a aplicação de tal medicamento. No estudo, o medicamento foi aplicado em 15 animais e, destes, 7 foram curados. O que podemos afirmar?

Informações:

- n = 15
- "sucesso": cura do animal
- 7: quantidade de animais curados $\Rightarrow T = O_1 = 7$
- hipóteses:

$$H_0 : p \geq 0.95$$
$$H_1 : p < 0.95$$

No R, usamos o comando

binom.test(x, n, p, alternative = c("two.sided", "less", "greater"), conf.level = 0.95)

escolhendo o tipo de hipótese, mais especificamente, definindo a hipótese alternativa e, o nível de confiança.

```
teste.1 <-binom.test(7, 15, p = 0.95, alternative = c("less"), conf.level = 0.95)
teste.1
```

```
##
##  Exact binomial test
##
## data:  7 and 15
## number of successes = 7, number of trials = 15, p-value = 1.83e-07
## alternative hypothesis: true probability of success is less than 0.95
## 95 percent confidence interval:
##  0.0000000 0.7000135
## sample estimates:
## probability of success
##              0.4666667
```

Podemos dizer que não existe evidência na amostra para aceitar a hipótese H_0. Observe a proporção (sucesso) estimada $\hat{p} \approx 0,47$, considerando p-valor $< 0,001$.

2. Alunos matriculados em disciplinas, ao final do período são classificados, como aprovado ou reprovado. Historicamente em determinada disciplina, a proporção de aprovados é, no máximo, igual 60%. Na última oferta da disciplina, entre os 21 alunos, 8 foram aprovados. O que podemos dizer?

Informações:

- n = 21
- "sucesso": aprovação do aluno
- 8: quantidade de alunos aprovados $\Rightarrow T = O_1 = 8$
- hipóteses:

$$H_0 : p \leq 0.60$$

$$H_1 : p > 0.60$$

```r
teste.2 <- binom.test(8, 21, p = 0.6, alternative = c("greater"), conf.level = 0.95)
teste.2
```

```
##
##  Exact binomial test
##
## data:  8 and 21
## number of successes = 8, number of trials = 21, p-value = 0.9877
## alternative hypothesis: true probability of success is greater than 0.6
## 95 percent confidence interval:
##  0.2057499 1.0000000
## sample estimates:
## probability of success
##              0.3809524
```

Podemos dizer que não existe evidência na amostra para rejeitar a hipótese H_0. Observe a proporção estimada $\hat{p} \approx 0,38$, considerando p-valor $= 0,9877$.

3. Certa indústria de refrigerantes desenvolveu um produto. O gerente de vendas gostaria de saber se o novo produto é tão preferido quanto o tradicional. Para tanto, realizou um experimento e entre os 13 degustadores, 5 preferem o novo. O que podemos afirmar?

Informações:

- n = 13
- "sucesso": preferência pela nova bebida
- 5: número de degustadores que preferem a nova bebida
- hipóteses:

$$H_0 : p \leq 0.50$$

$$H_1 : p > 0.50$$

```
teste.3 <- binom.test(5, 13, p = 0.5, alternative = c("greater"), conf.level = 0.95)
teste.3
```

```
##
##  Exact binomial test
##
## data:  5 and 13
## number of successes = 5, number of trials = 13, p-value = 0.8666
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.1656594 1.0000000
## sample estimates:
## probability of success
##              0.3846154
```

Podemos dizer que não existe evidência na amostra para rejeitar a hipótese H_0. Observe a proporção estimada $\hat{p} \approx 0,38$ para p-valor $= 0,8666$.

4. Em uma plantação (planta específica) foram observadas ao acaso 16 plantas e, entre estas, quatro estavam infectada com certo fungo. Se a proporção de infestação estiver abaixo de 30%, esta é considerada controlada. O que podemos afirmar?

Informações:

- n = 16
- "sucesso": planta estar infectada
- 4: número de plantas infectadas
- hipóteses:

$$H_0 : p \leq 0.30$$
$$H_1 : p > 0.30$$

```
teste.4 <- binom.test(4, 16, p = 0.3, alternative = c("greater"), conf.level = 0.95)
teste.4
```

```
##
##  Exact binomial test
##
## data:  4 and 16
## number of successes = 4, number of trials = 16, p-value = 0.7541
## alternative hypothesis: true probability of success is greater than 0.3
## 95 percent confidence interval:
##  0.09025243 1.00000000
## sample estimates:
## probability of success
##                   0.25
```

Podemos dizer que não existe evidência na amostra para rejeitar a hipótese H_0. Observe a proporção estimada $\hat{p} = 0,25$, para p-valor $= 0,7541$.

Aplicação: ***Comportamento anoréxico e percepção corporal em universitário***

(Jornal Brasileiro de Psiquiatria, v. 57, p. 166-170, 2008)

Um estudo foi realizado para avaliar a percepção corporal e a prevalência de sintomas de anorexia nervosa em estudantes universitários:

- participaram do estudo: 149 estudantes de Nutrição e 78 estudantes de Educação Física.
- método: Seleção da amostra foi realizada por conveniência. O questionário teste de atitudes alimentares (EAT- 26) e teste de imagem corporal foram aplicados para avaliar os sintomas de anorexia nervosa e a percepção corporal, respectivamente.

Considerando os resultados da tabela acima e as afirmações:

- de acordo com dados da Organização Mundial da Saúde (OMS), 4,7% dos brasileiros sofrem de compulsão alimentar (22/7/24)
- insatisfação corporal atinge mais de 50% dos brasileiros,

especifique o tamanho amostral e estatística teste, discuta, construa e teste as hipóteses, para os casos abaixo:

1. os estudantes de Nutrição têm tendência de apresentarem comportamento de risco para anorexia nervosa
2. os estudantes de Educação Física tendem a ter proporção elevadas de insatisfação corporal.
3. os estudantes, das duas turmas, têm tendência de apresentarem comportamento de risco para anorexia nervosa o que podemos afirmar?

caso 1:

- hipóteses:

$$H_0 : p \leq 0.047 \;\; vs \;\; H_1 : p > 0.047$$

```r
aplicacao.1 <- binom.test(21, 149, p = 0.047, alternative = c("greater"), conf.level = 0.95)
```

caso 2:

- hipóteses:

$$H_0 : p \leq 0.50 \;\; vs \;\; H_1 : p > 0.5$$

```r
aplicacao.2 <- binom.test(61, 78, p = 0.5, alternative = c("greater"), conf.level = 0.95)
```

caso 3:

- hipóteses:

$$H_0 : p \leq 0.047 \;\; vs \;\; H_1 : p > 0.047$$

```r
aplicacao.3 <- binom.test(29, 227, p = 0.047, alternative = c("greater"), conf.level = 0.95)
```

Exercício: interpretar os resultados dos casos 1) a 3) e, realizar testes de hipóteses considerando os demais itens.

Teste para duas amostras pareadas

Em alguns estudos, o interesse está em comparar ou o efeito de dois grupos ou efeito de dois tratamentos: verificar a diferença em entre duas situações!

Alguns exemplos:

a) aplicação de herbicida (em plantas) para comparar o efeito do produto, registrando a quantidade de ervas daninhas, nas plantas sem e com herbicida. Existe diferença entre os tratamentos (**controle e tratamento** - herbicida)?

b) antes do sistema de cotas ser implantado, alunos de uma turma foram entrevistados sobre a implantação, se manifestando **a favor ou contra**. Após ser implantado, os mesmos alunos foram entrevistados e novamente foram interrogados sobre serem favoráveis ou não à implantação. **Houve mudança significativa de opinião**?

c) Vinte pessoas foram selecionadas para um "experimento": para cada uma destas foi registrado o **tempo** de reação a determinado evento ao dirigir um carro (frear, desviar, *etc*.), **antes e após** ingerir bebida alcoólica. O álcool afeta o tempo de reação das pessoas?

d) Homens e mulheres tem a mesma altura? Para verificar esta afirmação, **uma amostra da população feminina** foi observada e registradas as alturas das mulheres e **uma amostra da população masculina** foi retirada e suas alturas registradas. O que podemos concluir?

Quais as diferenças nestes exemplos?

Diferenças:

- temos dois tratamentos diferentes aplicados um em cada amostra (grupo) ou temos dois tratamentos aplicado na mesma amostra?
- qual o tipo de variável e escala de medidas?
- os grupos são independentes ou são pareados?
- ? ? ?

Características:

- caso a) duas amostras independentes (porém, com elementos parecidos)

- casos b) e c) a mesma amostra é utilizada duas vezes: o elemento é usado como seu próprio controle, submetendo-o aos dois tratamentos, em ocasiões diferentes. Este tipo é chamado de pré e pós tratamentos (antes e depois). Outro tipo de pareamento é o pareamento natural, como exemplo, gêmeos, órgãos (ouvidos, braços, pés, etc.), ou ainda quando selecionamos (ou tentamos selecionar) indivíduos muito semelhantes!

- caso d) amostras retiradas de populações independentes

Outros pontos importantes:

- caso a) variável quantitativa ou ainda, o registro pode ser uma resposta ordinal (menor/maior, pior/melhor, menos/mais)
- caso b) variável qualitativa
- casos c) e d): variável quantitativa

=> situações diferentes => testes diferentes!

Teste do sinal para dados ordinais

- as observações são emparelhadas, seguindo algum critério.

- útil nos casos em que a mensuração quantitativa não é possível de ser estabelecida, sendo possível estabelecer uma **relação** para cada membro do par. Pode ser aplicado para dados quantitativos

- **ideia do teste**: uma amostra é submetida a um tratamento e os resultados da variável são registrados. A mesma amostra é submetida a outro tratamento e os resultados da variável são registrados. Comparamos o efeito dos dois tratamentos, para cada par.

- podemos observar as seguintes situações: ou (a) o efeito aumentou; ou (b) o efeito diminuiu ou ainda, (c) o efeito permaneceu constante

- a cada par é atribuído um sinal (+) ou (-) (daí vem o nome do teste!), dependendo do resultado observado (uso apenas do sinal!)

- eficiente para pequenas amostras (ideal quando $n \leq 50$ e não há empates de postos)

- **teste binomial** no caso em que $p = 0,5$

- **teste usado para comparar evoluções avaliadas em uma escala ordinal**.

- **Dados**:
  - consiste de *n* observações de uma v.a. bivariada $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.
  - os valores registrados, para cada par, são comparados e classificados da seguinte forma:

"+" , se $X_i < Y_i$; "-" , se $X_i > Y_i$ e "0" , se $X_i = Y_i$

- **Suposições**:
  - as v.as. bivariadas (X_i, Y_i) são mutuamente independentes
  - a escala de medidas é, pelo menos, ordinal dentro de cada par
  - os pares (X_i, Y_i) são internamente consistentes: se para um par $P(+) > P(-)$ então esta probabilidade vale para todos os pares; o mesmo ocorre para $P(+) < P(-)$ e para $P(+) = P(-)$.

- **Hipóteses**: (observando que, para X < Y => "+"):

  a) H_0: $E(X_i) = E(Y_i)$ vs H_1: $E(X_i) \neq E(Y_i)$ (mesmo parâmetro de locação!)

  b) H_0: $E(X_i) \geq E(Y_i)$ vs H_1: $E(X_i) < E(Y_i)$ (os valores de X tendem a ser maiores do que de Y => H_1: "+" é mais provável do que um "-")

  c) H_0: $E(X_i) \leq E(Y_i)$ vs H_1: $E(X_i) > E(Y_i)$, para todo i (os valores de X tendem a ser menores do que de Y).

- **Teste estatístico**:
  número de pares em que $X_i < Y_i$ => T = número de "+" .

- **Regra de decisão**:
  considerando apenas os ***n'*** pares em que $X_i \neq Y_i$, um nível α de significância e as hipóteses temos (usando a distribuição binomial com parâmetros n' e p = 0,5):

  a) rejeite H_0 se $T \leq t$ ou $T \geq n' - t$

  b) rejeite H_0 se $T \geq t$ (valores grande de T indicam H_1)

  c) rejeite H_0 se $T \leq t$ (valores pequenos de T indicam H_1).

Aplicação: **“Tower Building”** (Johnson e Courtney, Child Development, Vol. 3)

Usando blocos, os autores avaliaram os tempos (em segundos) que crianças pequenas levaram para construir uma torre (a mais alta possível). A fim de fornecer alguma medida de desenvolvimento mental, eles registraram os tempos para uma atividade inicial de construção de torres e, um mês depois, observando as mesmas crianças, registraram os tempos para uma segunda experiência de construção de torres:

- qual é a população de interesse?
- qual é a questão inerente de interesse aqui?
- quais as são amostras?
- qual é a variável aleatória avaliada?
- qual é o parâmetro populacional de interesse?
- qual é a estatística amostral correspondente a esse parâmetro?
- qual é uma declaração ou afirmação populacional apropriada?
- quais são as hipóteses estatísticas correspondentes?
- qual é a estatística teste usada para testar essas hipóteses?
- qual a decisão tomada e a conclusão, com base nos dados disponíveis?

- **definição de sucesso**: X > Y (evolução do experimento)

- **dados**: cada uma das crianças construiu duas torres (uma em cada momento) e, os tempos, em segundos, das construções foram registrados:

Crianças	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1ª tent.	30	19	19	23	29	178	42	20	12	39	14	81	17	31	52
2ª tent.	30	6	14	8	14	52	14	22	17	8	11	30	14	17	15
sinais															

- **hipóteses**:
  H_0:
  H_1:

  ou ainda, H_0: $E(X_i)$?? $E(Y_i)$ *vs* H_1: $E(X_i)$?? $E(Y_i)$

- T = ??

O que podemos concluir? (resolvemos no R!)

Importante:

- Nos casos em que a variável (bivariada) é quantitativa (conhecemos a hierarquia e o tamanho da diferença entre os valores - (X,Y)):

  - as amostras não tem distribuição normal, são pequenas e pareadas => teste de Wilcoxon associado a postos

  - as amostras não são pareadas e são pequenas => teste de Mann-Whitney

  - se as amostras têm distribuição normal => teste t-Student.

Teste de McNemar

- variação do teste do sinal, sendo que as observações pertencem a uma **escala nominal** com duas categorias (0/1, N/S, etc.). As frequências das classes são conhecidas

- como verificar se existe diferença entre as duas categorias (**probabilidades** diferentes)?

- **ideia do teste**: uma amostra é submetida a um tratamento e o resultado é registrado (0/1). Em um segundo momento, essa mesma amostra é submetida a outro tratamento, e novamente os valores da variável são registrados (0/1).
  Geralmente, o tratamento é o mesmo porém, aplicado na amostra em dois momentos diferentes. O tratamento é aplicado a primeira vez, e após alguma intervenção, o mesmo tratamento é aplicado novamente

- o interesse é verificar se as **mudanças** (de respostas de uma categoria para a outra) **são significativas.**

- **Dados**:
  consistem de n observações independentes de uma v.a. bivariada (X_i,Y_i), $i = 1,..., n$, sendo classificados em uma entre
  duas categorias:

	Classificação dos Y's	
Classificação dos X's	0	1
0	a = número de pares (0,0)	b = número de pares (0,1)
1	c = número de pares (1,0)	d = número de pares (1,1)

- **Suposições**:
  - os pares (X_i, Y_i) são mutuamente independentes, e X representa o tratamento 1 ou "antes" e Y representa o tratamento 2 ou "depois" e Y ;
  - a escala de dados é nominal com duas categorias, para todo par (X_i, Y_i), $i = 1,..., n$;
  - a diferença $P(X_i = 0,Y_i = 1) - P(X_i = 1,Y_i = 0)$ é negativa, ou igual a zero ou é positiva, para todo i.

- **Hipóteses**:
  H_0: $P(X_i = 0,Y_i = 1) = P(X_i = 1,Y_i = 0)$, para todo i.
  H_1: $P(X_i = 0,Y_i = 1) \neq P(X_i = 1,Y_i = 0)$, para todo i.
  (H_0: declaramos que as mudanças que ocorreram nos dois sentidos não são significativas).

- **Teste estatístico:**

temos dois casos:

a) se (b+c) > 20 => $T_1 = (b-c)^2/(b+c)$

b) se (b+c) ≤ 20 => $T_2 = b$

- **Regra de decisão**:

a) rejeite H_0 se $T_1 \geq w_{1-\alpha}$, obtida da distribuição qui-quadrado com 1 g.l.
b) rejeite H_0 se $T_2 \geq$ se $T_2 \leq t$ ou se $T_2 \geq n - t$, em que t é um valor obtido da distribuição binomial com parâmetros $n = b + c$ e, p = 0,5.

Aplicação: **Eficácia de ação educativa com reedducandas de cadeia pública de Mato Grosso sobre o vírus HPV**

(Corsino, P. K. D. *et al. Saúde e Pesquisa*, v. 11, n.1, pg 115-126, 2018)

- objetivo: analisar o impacto de ação educativa, sobre o vírus HPV, realizada com reeducandas de Cadeia Pública Feminina de Mato Grosso.

- estudo do tipo “antes” e depois”

- amostra: 37 mulheres (10/2016), em que

antes: explicação e sobre o tema da pesquisa e aplicação do questionário.

intervenção: duração de 30 minutos: atividade expositiva, utilizando cartazes com imagens relacionado ao HPV, modo de transmissão, diagnóstico, tratamento e prevenção, sendo possibilitado o moment reflexivo com as discussões e respostas aos questionamentos feitos pelas reeducandas.

após: após sete dias aplicou-se o mesmo questionário que precedeu a ação educativa.

Locais onde adquiriu conhecimento sobre HPV antes e após ação educativa realizada na Cadeia Pública Feminina.

Casa				Cadeia			
		Pós				Pós	
		S	N			S	N
Pré	S	2	5	Pré	S	2	1
	N	4	26		N	22	12

O que podemos concluir?

A realização da ação educativa contribuiu significativamente com a informação dessas mulheres sobre o HPV?

Vamos ao R!

Teste de Wilcoxon
(associado a postos)

Exemplo: “Tower Building”

É possível comparar, agora, os valores como quantidades e não apenas usando o sinal das diferenças

- teste aplicado para verificar se determinada amostra veio de uma população com **mediana** especificada (variável quantitativa)

- pode ser usado em situações nas quais as observações são **pareadas** (do tipo antes - depois), para verificar se as amostras têm mesma mediana

- as diferenças (entre observações do par e em módulo) são ordenadas e, **postos** são atribuídos a estas diferenças.

Observações:

a) **postos**: ordenação dos dados (do menor para o maior), atribuindo a "posição"

Exemplo

D= X-Y	-5	5	5	7	10	-13	13	15
|D|	5	5	5	7	10	13	13	15
Postos	2	2	2	4	5	6,5	6,5	8

- diferenças 5 e 13 **???**

b) **teste:**

- do sinal: aplicado nos caos em que apenas os sinais das diferenças são conhecidos
- de Wilcoxon: aplicado nos casos em que as diferenças, entre os valores dos tratamentos, são possíveis de serem obtidas (quantificar!).

- **Dados**:
  - consistem de *n* observações independentes de uma v.a. bivariada (Xi,Yi), i = 1,..., *n*.
  - defina as diferenças em módulo: $|D_i| = |Y_i - X_i|$, $i=1,\cdots,n$.
  - seja $\boldsymbol{n'}$ o número de pares em que $X_i \neq Y_i$ (o teste não considera pares nos quais $X_i = Y_i$).
  - atribua postos, denotado por **Ri,** para as n' diferenças em módulo.

- **Suposições**:
  - Di's são mutuamente independentes;
  - Di's são v.as. contínuas com distribuição simétrica e tem a mesma mediana
  - a escala de medidas para os Di's é pelo menos intervalar.

- **Hipóteses**:

  a) $H_0: d_{.5} = 0$ vs $H_1: d_{.5} \neq 0$

  b) $H_0: d.5 \leq 0$ vs $H_1: d.5 > 0$

  c) $H_0: d.5 \geq 0$ vs $H_1: d.5 < 0$

- **Teste estatístico:**

  seja Ri o posto associado para cada par (se Xi ≠ Yi) definido por:

  se Di < 0 => Ri é o **negativo do posto** associado ao par (Xi, Yi)

  se Di > 0 => Ri é o **posto** associado ao par (Xi, Yi).

  O teste é dado por:

$$T = \frac{\sum_{i=1}^{n'} R_i}{\sqrt{\sum_{i=1}^{n'} R_i^2}}.$$

  Se não houver igualdade:

$$T^+ = \sum_{i=1}^{n'} R_i.$$

- **Regra de decisão:**

a) rejeite H_0, ao nível α de significância se $T > w_{1-\alpha/2}$ ou se $T < w_{\alpha/2}$

b) rejeite H_0 ao nível α de significância se $T < w_\alpha$

c) reject H_0 se $T > w_{1-\alpha}$

(w: tabela de quantis do teste de Wilcoxon)

Exemplo: construção das torres por 15 crianças

Crianças	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1ª tent.	30	19	19	23	29	178	42	20	12	39	14	81	17	31	52
2ª tent.	30	6	14	8	14	52	14	22	17	8	11	30	14	17	15
diferenças															
postos															

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Aplicações apresentadas na aula 06: testes para duas amostras pareadas:

Teste do sinal

1. *Tower Building: construção de torres altas*
   Usando blocos, as crianças devem construir a torre mais alta que conseguirem, em dois momentos diferentes (após a primeira, a segunda tentativa é realizada um mês depois). A criança "aprende" com a primeira tentativa?

- dados: cada uma das crianças construiu duas torres (em dois momentos diferentes) e, os tempos de construção foram anotados. Valores registrados, diferenças, sinais e soma dos sinais positivos são

```
x <- c(30,19,19,23,29,178,42,20,12,39,14,81,17,31,52)
y <- c(30,6,14,8,14,52,14,22,17,8,11,30,14,17,15)

dif <- x - y
dif
```

```
##  [1]   0  13   5  15  15 126  28  -2  -5  31   3  51   3  14  37
```

```
dif <- dif[-1]      # retira a 1a. diferenca = 0
dif
```

```
##  [1]  13   5  15  15 126  28  -2  -5  31   3  51   3  14  37
```

```
ifelse(dif > 0, yes = "+", no = "-")  # atribuindo postos
```

```
##  [1] "+" "+" "+" "+" "+" "+" "-" "-" "+" "+" "+" "+" "+" "+"
```

```
dif.0.1 <- ifelse(dif > 0, yes = 1, no = 0)  # atribuindo 0's e  1's
dif.0.1
```

```
##  [1] 1 1 1 1 1 1 0 0 1 1 1 1 1 1
```

```
estatistica <- sum(dif.0.1); estatistica
```

```
## [1] 12
```

```
length(dif.0.1)
```

```
## [1] 14
```

Usando o comando do R, temos:

```
teste.1 <- binom.test(12, 14, p = 0.5, alternative = c("greater"), conf.level = 0.95)
teste.1
```

```
##
```

```
##  Exact binomial test
##
## data:  12 and 14
## number of successes = 12, number of trials = 14, p-value = 0.00647
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.6146103 1.0000000
## sample estimates:
## probability of success
##              0.8571429
```

Podemos dizer que, baseado nos valores amostrais, existe evidência para rejeitar H_0, ou seja, existe diferença entre os tempos da primeira e segunda tentativas (p-valor $= 0,00647$).

Teste de McNemar

2. *Eficácia de ação educativa com reeducandas de cadeia pública de Mato Grosso sobre o vírus HPV* O estudo foi realizado na Cadeia Feminina de Mato Grosso, com o objetivo de analisar o impacto de ação educativa, sobre o vírus HPV.

- estudo do tipo "antes" e depois" (a administração do tratamento experimental)
- considerando "casa", que podemos concluir?
- dados: 37 mulheres

```
casa <- matrix(c(2,4,5,26), nrow = 2,
            dimnames = list("Antes" = c("Sim","Nao" ),
                            "Depois" = c("Sim", "Nao")))
casa
```

```
##      Depois
## Antes Sim Nao
##   Sim   2   5
##   Nao   4  26
```

```
teste.1 <-mcnemar.test(casa)
teste.1
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  casa
## McNemar's chi-squared = 0, df = 1, p-value = 1
```

Podemos dizer, com base na amostra, que existe evidência para rejeitar H_0, considerando p-valor $< 0,001$, ou seja, podemos dizer que existe mudança significativa de opinião após a intervenção (vídeos, explicações, *etc.*).

Exercícios:

1. Considerando "cadeia", que podemos concluir?

```
cadeia <- matrix(c(2,22,1,12), nrow = 2,
            dimnames = list("Antes" = c("Sim","Nao" ),
                            "Depois" = c("Sim", "Nao")))
cadeia
```

```
##      Depois
## Antes Sim Nao
##   Sim   2   1
##   Nao  22  12
```

```
teste.2 <- mcnemar.test(cadeia)
```

2. Sessenta pacientes do sexo masculino que estavam sendo tratados para impotência sexual com medicação injetável, foram submetidos a um tratamento com medicação oral. Dentre os 40 pacientes que obtiveram sucesso com a medicação oral, 25 também já tinham apresentado melhora com a injeção local. Oito pacientes não obtiveram sucesso com nenhum tratamento. Construa uma tabela que mostre todos os resultados obtidos. Há diferença entre os medicamentos?

Teste de Wilcoxon associado a postos

3. *Tower Building: construção de torres altas*
   (voltando à aplicação da construção das torres)

```
x <- c(30,19,19,23,29,178,42,20,12,39,14,81,17,31,52)
y <- c(30,6,14,8,14,52,14,22,17,8,11,30,14,17,15)
```

Observando igualdade nas respostas, retiramos primeiro par de tempos:

```
x <- x[-1]
y <- y[-1]
```

e aplicamos o teste de Wilcoxon associado a postos:

```
teste.3 <- wilcox.test(x, y, alternative = c("two.sided"),
                       paired = TRUE, conf.level = 0.95)
```

```
## Warning in wilcox.test.default(x, y, alternative = c("two.sided"), paired =
## TRUE, : cannot compute exact p-value with ties
```

```
teste.3
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x and y
## V = 99.5, p-value = 0.003486
## alternative hypothesis: true location shift is not equal to 0
```

Podemos dizer, baseados nas informações da amostra, que existe evidência para rejeitar H_0, considerando p-valor $= 0,003486$.

Exemplo

Um nutricionista afirma que uma dieta de emagrecimento consegue produzir bons resultados em seis meses. Para verificar se a dieta é eficiente, este nutricionista selecionou, aleatoriamente, dez pacientes e pediu a estes pacientes que seguissem a dieta, por seis meses. O nutricionista pesou os pacientes em dois momentos: antes de começar a dieta e ao final de seis meses. Observando os pesos destes pacientes, o que voce conclui?

```
x <- c(82, 65, 69, 76, 73, 56, 43, 58, 61, 70)
y <- c(63, 73, 68, 70, 75, 53, 45, 50, 58, 65)
```

Teste para três ou mais amostras pareadas

Exemplos:

a) Dez juízes de vinhos avaliam algumas características de determinado tipo de vinho, produzido por quatro vinícolas. Os critérios são: intensidade, limpidez, álcool, transparência, e ao final, atribuem uma nota de 0 a 10. Alguma vinícola produz vinho "melhor" (nota mais alta) do que as demais?

b) Quinze soldadores usam, cada um, três tochas de soldagem, e as soldas resultantes são avaliadas quanto à qualidade. Pelo menos uma das tochas produz soldas consistentemente melhores ou piores do que as demais?

Duas aplicações:

1) Em um estudo sobre hipnose, oito pessoas receberam uma tensão elétrica na superfície da pele (em milivolts), em quatro situações emocionais distintas: medo, alegria, tristeza e calma. Podemos afirmar que existe diferença na tensão entre os diferentes estados emocionais?

Pacientes	Medo	Alegria	Tristeza	Calma
1	23,1	22,7	22,5	22,6
2	57,6	53,2	53,7	53,1
3	10,5	9,7	10,8	8,3
4	23,6	19,6	21,1	21,6
5	11,9	13,8	13,7	13,3
6	54,6	47,1	39,2	37,0
7	21,0	13,6	13,7	14,8
8	20,3	23,6	16,3	14,8

(Lehmann, E. L. and D’abrera, H. J. M. Nonparametrics : statistical methods based on ranks, 2006)

2) Em um estudo sobre variedades de cana-de-açúcar, foram registrados a produção (em toneladas por hectare), de seis canteiros para as cinco variedades da cana. O que podemos concluir sobre as cinco variedades?

Canteiros	Variedade 1	Variedade 2	Variedade 3	Variedade 4	Variedade 5
1	110,6	116,7	128,7	140,3	143,4
2	119,5	128,4	140,2	150,0	153,8
3	120,1	131,5	130,3	150,9	151,5
4	105,3	114,8	138,7	144,7	144,1
5	130,8	146,8	146,0	153,9	154,6
6	138,1	155,5	149,8	156,0	159,3

Teste de Friedman

- extensão do teste do sinal

- comparar mais de duas amostras pareadas

- útil para estudos de medidas repetidas

- verificar se as amostras provêm da mesma população

- o mesmo elemento amostral é observado várias vezes (uma vez em cada um dos tratamentos) => **detectar possíveis diferenças entre os diferentes tratamentos**

- os dados numéricos não são usados diretamente, mas sim os **postos** ocupados pelos valores

- A ordenação é realizada para cada bloco separadamente.

- **Amostra**
  - considere v.as. independentes k-variadas (Xi1, Xi2,...,Xik) (i = 1,…, n) (i representa o elemento amostral e k e o número de tratamentos)
  - atribua postos dentro de cada bloco (elemento amostral). Em caso de valores iguais, atribua a média dos postos.

- **Suposições**:
  - os dados consistem de blocos (u.a.’s) independentes, para os k tratamentos
  - as observações são independentes dentro e entre as amostras
  - a **variável de estudo é contínua**
  - a **escala de medidas é, pelo menos, ordinal**, dentro de cada bloco
  - não existe interação entre blocos e tratamentos.

- **Hipóteses**:

H_0: a ordenação das v.as. dentro dos blocos é igualmente provável
(ou ainda os **tratamentos têm efeitos idênticos**)

vs

H_1 : a ordenação das v.as. dentro dos blocos não é igualmente provável para, pelo menos, um par de tratamentos.
(**pelo menos um dos tratamentos tende a produzir valores diferentes dos demais**).

- **Teste estatístico**:

$$T = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

em que

$$R_j = \sum_{i=1}^{n} R(X_{ij}), \; j = 1, \cdots, k.$$

- **Regra de Decisão**:
  - rejeite H_0, ao nível α de significância, se T exceder o quantil (1−α) da distribuição qui-quadrado com (k-1) g.l.

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste de Friedman

Aplicações:

1. *Situações emocionais*

Em um estudo sobre hipnose, oito pessoas receberam uma tensão elétrica na superfície da pele (em milivolts), em quatro situações emocionais distintas: medo, alegria, tristeza e calma. Podemos afirmar que existe diferença na tensão entre os diferentes estados emocionais?

- dados: cada oito pessoas receberam o choque e as tensoes, para cada emoção, foram registrados.
- análise descritiva:

```
sensacoes <- matrix(c(23.1, 57.6, 10.5, 23.6, 11.9, 54.6, 21.0, 20.3,
              22.7, 53.2,  9.7, 19.6, 13.8, 47.1, 13.6, 23.6,
              22.5, 53.7, 10.8, 21.1, 13.7, 39.2, 13.7, 16.3,
              22.6, 53.1,  8.3, 21.6, 13.3, 37.0, 14.8, 14.8),
              ncol = 4,byrow=FALSE, dimnames = list(1:8,
              c("medo","alegria","tristeza", "calma")))

sensacoes
```

```
##   medo alegria tristeza calma
## 1 23.1    22.7     22.5  22.6
## 2 57.6    53.2     53.7  53.1
## 3 10.5     9.7     10.8   8.3
## 4 23.6    19.6     21.1  21.6
## 5 11.9    13.8     13.7  13.3
## 6 54.6    47.1     39.2  37.0
## 7 21.0    13.6     13.7  14.8
## 8 20.3    23.6     16.3  14.8
```

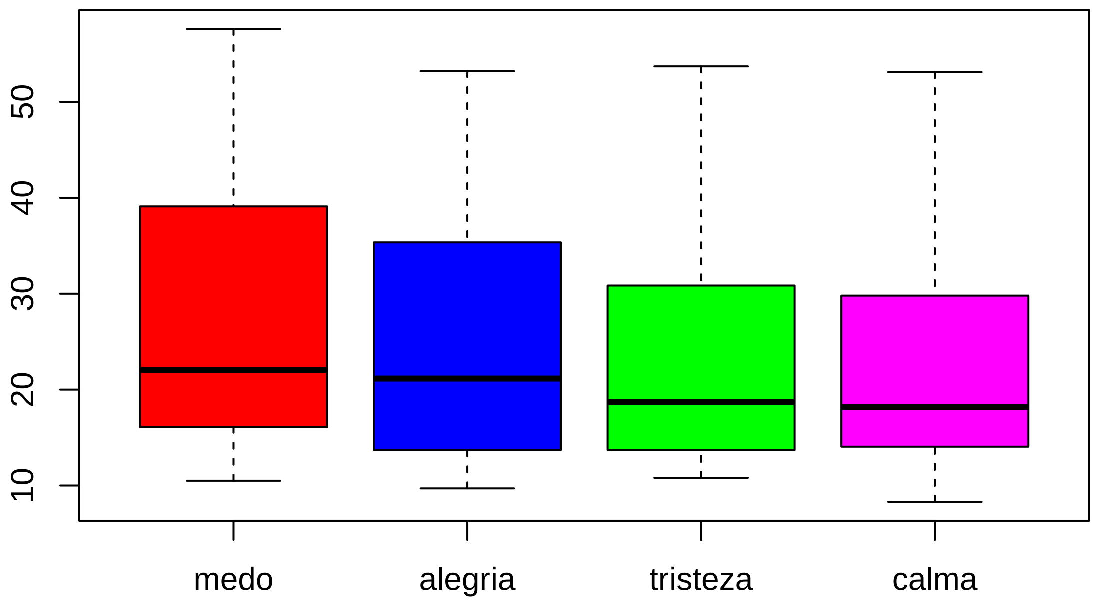
```r
summary(sensacoes)
```

```
##       medo          alegria         tristeza          calma      
##  Min.   :10.50   Min.   : 9.70   Min.   :10.80   Min.   : 8.30  
##  1st Qu.:18.20   1st Qu.:13.75   1st Qu.:13.70   1st Qu.:14.43  
##  Median :22.05   Median :21.15   Median :18.70   Median :18.20  
##  Mean   :27.82   Mean   :25.41   Mean   :23.88   Mean   :23.19  
##  3rd Qu.:31.35   3rd Qu.:29.48   3rd Qu.:26.68   3rd Qu.:26.20  
##  Max.   :57.60   Max.   :53.20   Max.   :53.70   Max.   :53.10  
```

```r
apply(sensacoes, 2, sd)
```

```
##     medo  alegria tristeza    calma 
## 18.12888 16.07211 14.97319 14.87332 
```

```r
boxplot(sensacoes, outline = TRUE, col = c("red","blue","green", "magenta"))
```



- aplicação do teste de Friedman:

```r
friedman.test(sensacoes)
```

```
## 
##  Friedman rank sum test
## 
## data:  sensacoes
## Friedman chi-squared = 6.45, df = 3, p-value = 0.09166
```

O que podemos concluir?

2. *Plantação de cana-de-acúcar*

Em um estudo sobre variedades de cana-de-açúcar, foram registrados a produção (em toneladas por hectares), dos seis canteiros para as cinco variedades da cana. O que podemos concluir sobre as cinco variedades de cana?

- dados: nos seis canteiro foram plantadas as cinco variedades de cana e, a produção, para cada canteiro foi registrada.
- análise descritiva:

```r
producao <- matrix(c(110.6, 119.5, 120.1, 105.3, 130.8, 138.1,
                     116.7, 128.4, 131.5, 114.8, 146.8, 155.5,
                     128.7, 140.2, 130.3, 138.7, 146.0, 149.8,
                     140.3, 150.0, 150.9, 144.7, 153.9, 156.0,
                     143.4, 153.8, 151.5, 144.1, 154.6, 159.3),
                     ncol = 5, byrow=FALSE, dimnames = list(1:6,
                     c("V.1", "V.2", "V.3", "V.4", "V.5")))

producao
```

```
##     V.1   V.2   V.3   V.4   V.5
## 1 110.6 116.7 128.7 140.3 143.4
## 2 119.5 128.4 140.2 150.0 153.8
## 3 120.1 131.5 130.3 150.9 151.5
## 4 105.3 114.8 138.7 144.7 144.1
## 5 130.8 146.8 146.0 153.9 154.6
## 6 138.1 155.5 149.8 156.0 159.3
```

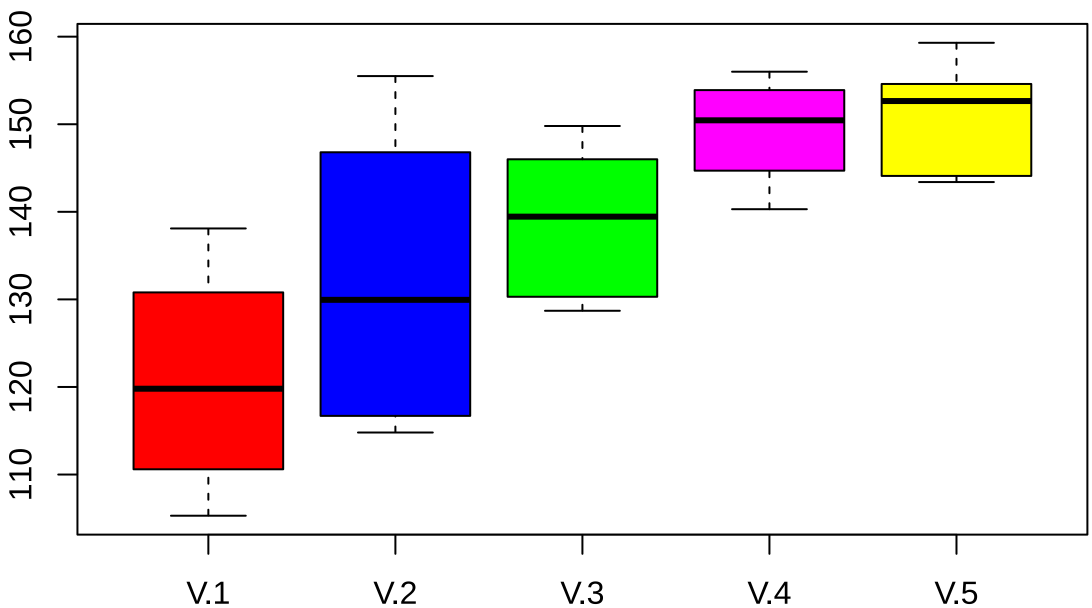
```r
summary(producao)
```

```
##       V.1              V.2              V.3              V.4
##  Min.   :105.3   Min.   :114.8   Min.   :128.7   Min.   :140.3
##  1st Qu.:112.8   1st Qu.:119.6   1st Qu.:132.4   1st Qu.:146.0
##  Median :119.8   Median :129.9   Median :139.4   Median :150.4
##  Mean   :120.7   Mean   :132.3   Mean   :138.9   Mean   :149.3
##  3rd Qu.:128.1   3rd Qu.:143.0   3rd Qu.:144.6   3rd Qu.:153.2
##  Max.   :138.1   Max.   :155.5   Max.   :149.8   Max.   :156.0
##       V.5
##  Min.   :143.4
##  1st Qu.:145.9
##  Median :152.7
##  Mean   :151.1
##  3rd Qu.:154.4
##  Max.   :159.3
```

```r
apply(producao, 2, sd)
```

```
##       V.1       V.2       V.3       V.4       V.5
## 12.212562 16.210789  8.353861  5.852521  6.248653
```

```r
boxplot(producao, outline = TRUE, col = c("red","blue","green", "magenta", "yellow"))
```



- aplicação do teste de Friedman:

```r
friedman.test(producao)
```

```
##
##  Friedman rank sum test
##
## data:  producao
## Friedman chi-squared = 22.133, df = 4, p-value = 0.0001885
```

O que podemos concluir?

Teste de comparações múltiplas para amostras independentes

Aplicação

Em um estudo sobre variedades de cana-de-açúcar foram registrados as produções, em toneladas por hectare, de seis canteiros para a= cinco variedades da cana. O que podemos concluir sobre as cinco variedades?

Canteiros	Variedade 1	Variedade 2	Variedade 3	Variedade 4	Variedade 5
1	110,6	116,7	128,7	140,3	143,4
2	119,5	128,4	140,2	150,0	153,8
3	120,1	131,5	130,3	150,9	151,5
4	105,3	114,8	138,7	144,7	144,1
5	130,8	146,8	146,0	153,9	154,6
6	138,1	155,5	149,8	156,0	159,3

Lembrando…

Teste de Friedman

- comparar mais do que duas amostras pareadas

- **detectar possíveis diferenças entre os diferentes tratamentos**

- **hipóteses:**

H_0: a ordenação das v.as. dentro dos blocos é igualmente provável

vs

H_1: a ordenação das v.as. dentro dos blocos não é igualmente provável para, pelo menos, um par de tratamentos.

- **Amostra**
  - considere v.as. independentes *k*-variadas (X_{i1}, X_{i2},…, X_{ik}) (i = 1,…, n) (i representa o elemento amostral e k e o número de tratamentos)
  - atribua postos dentro de cada bloco (elemento amostral). Em caso de valores iguais, atribua a média dos postos.

- **Suposições**:
  - os dados consistem de blocos (u.a.'s) independentes, para os *k* tratamentos
  - as observações são independentes dentro e entre as amostras
  - a **variável de estudo é contínua**
  - a **escala de medidas é, pelo menos, ordinal**, dentro de cada bloco
  - não existe interação entre blocos e tratamentos.

- **Estatística do teste:**

$$T = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

em que $R_j = \sum_{i=1}^{n} R(X_{ij}), \; j = 1, \cdots, k.$

Rejeite H_0, ao nível α de significância, se T exceder o quantil (1−α) da distribuição qui-quadrado com (k-1) g.l.

Se no teste de Friedman, a hipótese nula for rejeitada (diferenças serem detectadas), quais tratamentos produzem valores diferentes dos demais?

Testes de comparações múltiplas

- Para todos os pares de tratamentos, fazemos comparações

- **Hipóteses**:

  H_0: tratamentos *i* e *j* **não** produzem resultados diferentes
  vs
  H_1 :tratamentos *i* e *j* produzem resultados diferentes

- **Estatística:**

$$|R_i - R_j| > t_{1-\alpha/2}\left[\frac{2b(A_1 - B_1)}{(b-1)(k-1)}\right]^{1/2}$$

em que: $t_{1-\alpha/2}$ a distribuição t-Student com (*b-1*)(*k-1*) g.l. e,

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k}(R(X_{ij}))^2 \qquad \text{e} \qquad B_1 = \frac{1}{b}\sum_{j=1}^{k} R_j^2$$

- **Regra de Decisão:**
  - rejeite H_0 se a desigualdade é satisfeita.

- **Observações:**
  - em caso de igualdades: A_1 é modificada.

- Existem outros testes para mais do que duas amostra pareadas:

  - **Teste Q de Cochran**:
    - é extensão do teste de McNemar para mais do que duas amostras relacionadas
    - neste teste, não são “avaliadas” a extensão da mudança, apenas se houve mudança.

- Várias autores descrevem o procedimento de comparação para o teste de Friedman como sendo uma **análise de variância dois fatores** ("two-way")

- Nos testes de comparações múltiplas, postos são usados para detectar diferenças

- Testes mais usados (Pós Friedman):
  - Testes de Nemeny e de Conover: compara todos os pares de tratamentos usando a soma dos postos, ajustando o p-valor

- Corrija o p-valor (p_{novo}) e compare-o com nível de significância (α):
  - rejeite a hipótese H_0 se $p_{novo} < \alpha$ (ou seja, os tratamentos produzem resultados, estatisticamente diferentes)
  - a correção do p-valor reduz a chance de falsos positivos => teste mais conservador.

Vamos para o R!

Duas aplicações

1. Revisão da medicação em idosos institucionalizados: aplicação dos critérios STOPP e START

(Periquito, Catarina Metelo de Nápoles, *et al.* Revisão da medicação em idosos institucionalizados: aplicação dos critérios STOPP e START. *Revista Portuguesa de Farmacoterapia* 6 (2014): 211-7)

Objetivo: caracterizar e quantificar a ocorrência de medicamentos potencialmente inadequados (MPI) e medicamentos potencialmente omissos (MPO) em um grupo de idosos institucionalizados através da revisão da medicação.

Amostra: os doentes com idade ≥ 65 anos que utilizam cinco ou mais medicamentos

Análise estatística: testes de Mann-Whitney e de Kruskal-Wallis.

2. Investigação sobre a capacidade de sustentar a atenção no Teste do Relógio de Mackworth, segundo a idade e sexo

(Giambra, Leonard M., and Reginald E. Quilter. Sex differences in sustained attention across the adult life span. Journal of applied psychology 74.1 (1989): 91.)

Objetivo: investigar a diferença entre idades na capacidade de sustentar a atenção quando as pessoas foram submetidas ao Teste do Relógio de Mackworth (relógio de metal com um mostrador branco simples e um ponteiro preto que se move ao redor do mostrador em 100 passos discretos de 36º cada). Durante o período de teste, o ponteiro fez 23 saltos duplos, definidos como mover-se duas vezes a distância normal ou 7,2º no mesmo período de tempo, em intervalos aleatórios irregulares. **As pessoas foram informadas de que os saltos duplos ocorreriam e estas sinalizavam seu reconhecimento da ocorrência pressionando um botão. As pontuações foram o número de reconhecimentos corretos dos saltos duplos.**

Amostra: 10 homens, de18 a 29 anos e 10 homens de 50 a 59 anos.

O número mediano de pontuações corretas é maior para homens jovens do que para homens mais velhos?

- O que são estes testes citados na 1ª aplicação: Mann-Whitney? Kruskall-Wallis?

Teste de Mann-Whitney

- Equivalente ao teste paramétrico **t-Student** (não-pareado - duas amostras independentes)
  - diferença: no teste não paramétrico são usados postos
  - vantagem: as suposições de normalidade e homogeneidade de variâncias não são necessárias
  - se existirem pontos discrepantes nas amostras, estes perdem a "influência", pois neste teste, os postos são considerados (a informação de valor grande está na posição).

- Suposição: variável sob estudo deve ser, pelo menos, ordinal.

- As amostras não precisam ser do mesmo tamanho.

- Podemos ter duas situações:
  - as amostras foram retiradas de duas populações diferentes
  - u.a.'s da mesma população foram alocados, aleatoriamente, em (dois) tratamentos diferentes.

- As duas amostras vieram da mesma população? (com a mesma média/mediana?)

- Ideia do teste:

  - **combine** as duas amostras X e Y e ordene os valores de forma crescente
  - **atribua** postos para todas as observações
  - considere apenas os **postos da amostra X**
  - **a estatística teste é baseada na soma dos postos da amostra X: quanto menor for o valor da estatística, maior a evidência de que as populações são diferentes.**

Teste de Mann-Whitney (Wilcoxon - Mann - Whitney)

- **Amostra:**
  - considere duas amostras (X_1, X_2, … , X_n) e (Y_1, Y_2, … , Y_m), retiradas de populações diferentes
  - atribua postos para as (n + m) observações (amostra combinada)
  - denote:

    $R(X_i)$ o posto associado à observação X_i, $i = 1, \dots, n$ e,

    $R(Y_j)$ o posto associado à observação Y_j , $j = 1, \dots, m$. Denote $N = n + m$.
  - se existirem observações iguais, atribua a média dos postos.

- **Suposições**:
  - as duas amostras foram retiradas aleatoriamente
  - as amostras são independentes entre si
  - a escala de medidas é, pelo menos ,ordinal.

- **Hipóteses**:

  a) H_0: E(X) = E(Y) *vs* H_1: E(X) = E(Y)

  b) H_0: E(X) ≤ E(Y) *vs* H_1: E(X) > E(Y)

  c) H_0: E(X) ≥ E(Y) *vs* H_1: E(X) < E(Y)

(se existe diferença entre as amostras => diferença em locação!)

- **Teste estatístico**:

$$T = R_X - \frac{n(n+1)}{2}$$

- **Região crítica e decisão**:
  - esta depende da hipótese a ser testada.
  - existe uma tabela para o teste: w_p para p = 0,001; 0,01; 0,1; 0,005; 0,05 e 0,025.

a) rejeite H_0 se $T < w_{\alpha/2}$ ou se $T > w_{1-\alpha/2}$

b) rejeite H_0 se $T > w_{1-\alpha}$ ou se $T' < w_\alpha$

c) rejeite H_0 se $T < w_\alpha$.

- **Observação**: se:
  - *n* e *m* < 20: distribuição teórica e exata para a estatística (quantis w são obtidos na tabela)
  - *n* e *m* ≥ 20: aproximação pela distribuição normal

- **Observação:**

  - existem outras definições para os testes de soma de postos de Wilcoxon-Mann-Whitney
  - por exemplo,

$$U_X = nm + \frac{n(n+1)}{2} - R_X$$

e

$$U_Y = nm + \frac{m(m+1)}{2} - R_Y$$

O teste é dado por:

$$\boldsymbol{T = min(U_X, U_Y)}$$

Observação:

- este teste pode ser modificado para **testar variabilidade das amostras**
- por exemplo, teste unilateral:

H_0: Var(X) ≤ Var(Y)

H_1: Var(X) > Var(Y)

As amostras são combinadas e os postos são atribuídos da seguinte maneira:

- atribua o valor 1 para a menor observação, 2 para a maior observação, 3 para a segunda maior observação 4 para a segunda menor observação, 5 para a terceira menor observação, 6 para a terceira maior observação e, assim sucessivamente.
- se H_0 é falso, os valores de X tendem a estar nas caudas da amostra combinada e portanto receberá os menores postos.

Aplicação 3:

Investigação sobre da capacidade de sustentar a atenção no Teste do Relógio de Mackworth, segundo a idade e sexo

Dados:

Jovens: 11; 13; 15; 15; 17; 19; 20; 21; 21; 22

Adultos: 8; 9; 10; 11; 12; 13; 5; 17; 19; 23

O número mediano de pontuações corretas é maior para homens jovens do que para homens mais velhos?

Hipóteses: $H_0: (X) \leq E(Y)$ *vs* $H_1: E(X) > E(Y)$

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Testes de comparações múltiplas: Friedman

Aplicações apresentadas:

1. *Situações emocionais*

Em um estudo sobre hipnose, oito pessoas receberam uma tensão elétrica na superfície da pele (em milivolts), em quatro situações emocionais distintas: medo, alegria, tristeza e calma. Podemos afirmar que existe diferença na tensão entre os diferentes estados emocionais?

```r
respostas <- matrix(c(23.1, 57.6, 10.5, 23.6, 11.9, 54.6, 21.0, 20.3,
                      22.7, 53.2,  9.7, 19.6, 13.8, 47.1, 13.6, 23.6,
                      22.5, 53.7, 10.8, 21.1, 13.7, 39.2, 13.7, 16.3,
                      22.6, 53.1,  8.3, 21.6, 13.3, 37.0, 14.8, 14.8),
                      ncol = 4, byrow = FALSE)

grupos <- matrix(c(rep("medo", 8), rep("alegria", 8), rep("tristeza", 8),
                   rep("calma", 8)), ncol = 4, byrow = FALSE)

blocos <- seq(1:8)

friedman.test(respostas, grupos, blocos )
```

```
##
##  Friedman rank sum test
##
## data:  respostas
## Friedman chi-squared = 6.45, df = 3, p-value = 0.09166
```

Pelo teste de Friedman, não existe evidências na amostra para rejeitar a hipótese de que as tensões são “iguais” para as diferentes emoções (p-valor $= 0,09166$).

Como parte da análise descritiva dos dados, vamos construir um gráfico de pontos, com informações da média e desvio-padrão.

```
dp <- apply(respostas, 2, sd)
dp
```

```
## [1] 18.12888 16.07211 14.97319 14.87332
```

```
media <- apply(respostas, 2, mean)
media
```

```
## [1] 27.8250 25.4125 23.8750 23.1875
```

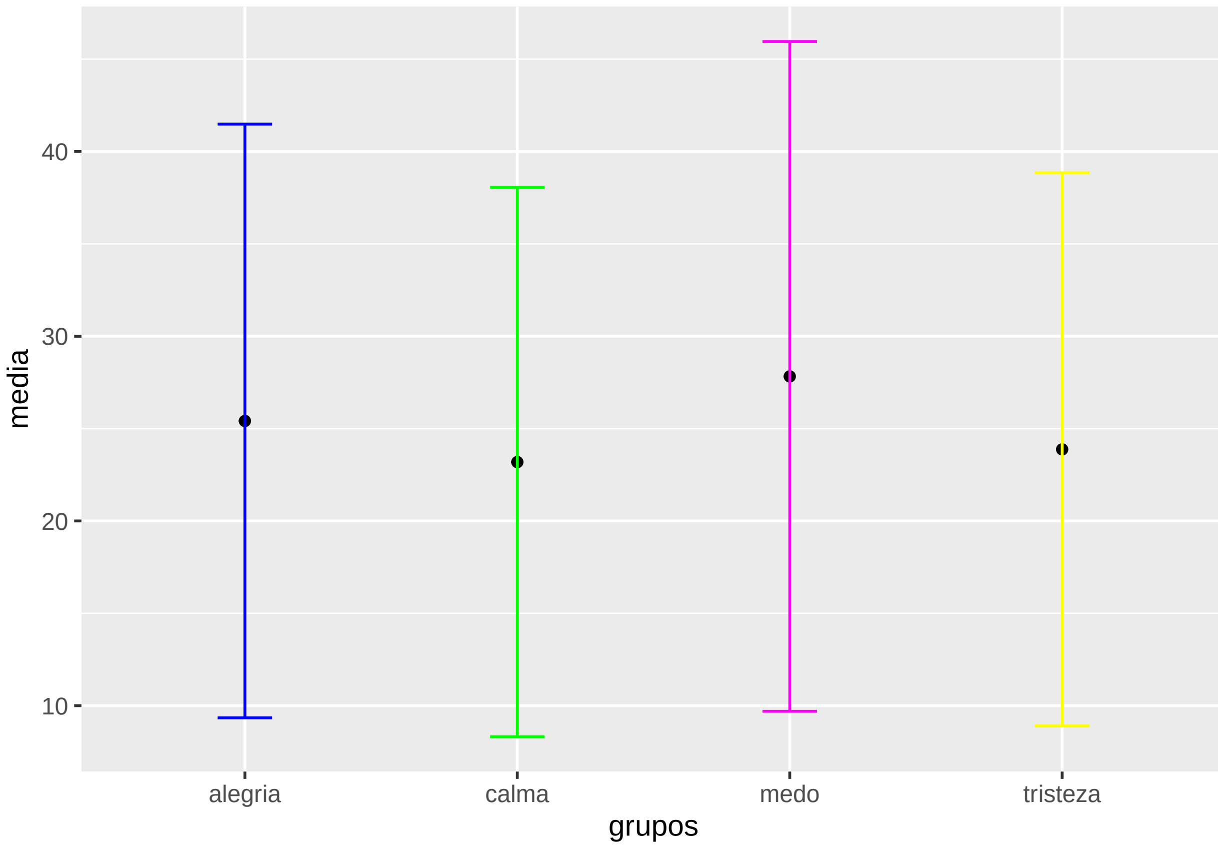
```
df <- data.frame(Mean=media, sd=dp,
                 grupos=as.factor(c("medo", "alegria", "tristeza", "calma")))
df
```

```
##      Mean       sd   grupos
## 1 27.8250 18.12888     medo
## 2 25.4125 16.07211  alegria
## 3 23.8750 14.97319 tristeza
## 4 23.1875 14.87332    calma
```

Construindo o gráfico:

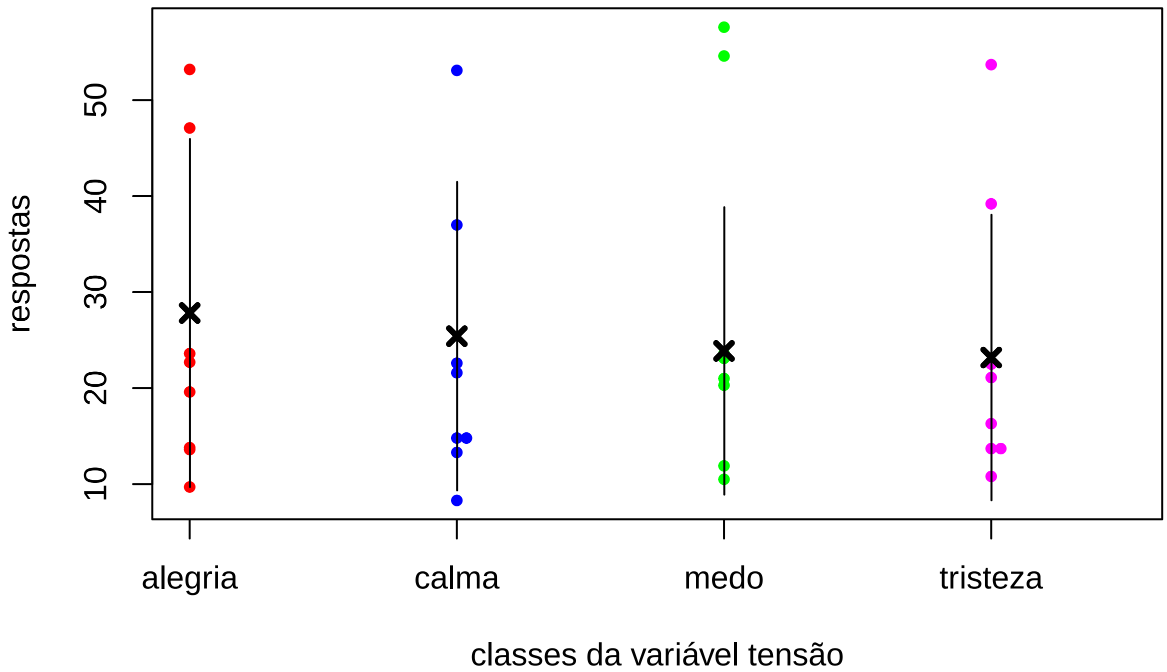
```
grafico.m.dp <- ggplot(df, aes(x=grupos, y=media)) +
                       geom_point()+
                       geom_errorbar(aes(ymin = media-dp, ymax = media+dp), width=.2,
                       position=position_dodge(0.05),
                       colour= c("magenta", "blue", "yellow", "green") )

grafico.m.dp
```




```
stripchart(respostas ~ grupos, col = c("red","blue","green", "magenta"),
           pch = c(20,20,20,20), method = "stack",
           xlab = "classes da variável tensão", vertical = TRUE)
points(media, col = "black", pch = 4, cex = 1.1, bg = 2, lwd = 3)

lines(c(1.001,1.001), c(media[1] - dp[1], media[1] + dp[1]),lty=1)
lines(c(2.001,2.001), c(media[2] - dp[2], media[2] + dp[2]),lty=1)
lines(c(3.001,3.001), c(media[3] - dp[3], media[3] + dp[3]),lty=1)
lines(c(4.001,4.001), c(media[4] - dp[4], media[4] + dp[4]),lty=1)
```



Apenas como um exercício, vamos aplicar o teste de comparações múltiplas (Conover, 1999), pois não rejeitamos a hipótese de que tensões são "iguais" para as diferentes emoções.

```
sensacoes <- matrix(c(23.1, 57.6, 10.5, 23.6, 11.9, 54.6, 21.0, 20.3,
                      22.7, 53.2,  9.7, 19.6, 13.8, 47.1, 13.6, 23.6,
                      22.5, 53.7, 10.8, 21.1, 13.7, 39.2, 13.7, 16.3,
                      22.6, 53.1,  8.3, 21.6, 13.3, 37.0, 14.8, 14.8),
                    ncol = 4,byrow=FALSE, dimnames = list(1:8,
                                                          c("medo", "alegria", "tristeza", "calma")))

frdAllPairsConoverTest(sensacoes, groups = grupos,
                       blocks = blocos, p.adjust = "bonferroni")
```

```
##
##  Pairwise comparisons using Conover's all-pairs test for a two-way balanced complete block design
## data: y
##          medo  alegria tristeza
## alegria  0.918 -       -
## tristeza 0.630 1.000   -
## calma    0.071 1.000   1.000
##
## P value adjustment method: bonferroni
```

Observe que, pelo teste de comparações múltiplas, não podemos dizer que as diferentes tensões produzem resultados diferentes, sob as quatro sensações.

2. Exercício: *Plantação de cana-de-acúcar*

Em um estudo sobre variedades de cana-de-açúcar, foram registrados as produções (em toneladas por hectare) de seis canteiros, para cinco variedades de cana.

- dados: nos seis canteiro foram plantadas as cinco variedades de cana e, a produção, para cada canteiro foi registrada.
- análise descritiva:

```
producao <- matrix(c(110.6, 119.5, 120.1, 105.3, 130.8, 138.1,
                     116.7, 128.4, 131.5, 114.8, 146.8, 155.5,
                     128.7, 140.2, 130.3, 138.7, 146.0, 149.8,
                     140.3, 150.0, 150.9, 144.7, 153.9, 156.0,
                     143.4, 153.8, 151.5, 144.1, 154.6, 159.3),
                     ncol = 5, byrow=FALSE, dimnames = list(1:6,
                     c("V.1", "V.2", "V.3", "V.4", "V.5")))

blocos <- seq(1:6)
```

O que podemos dizer a respeito da produção das cinco variedades de cana?

Um Exemplo

(Sheskin, D. Handbook of Parametric and Nonparametric Statistical Procedures.(2003).)

Um psicólogo conduz um estudo para determinar se o ruído pode ou não inibir o aprendizado.

Cada um das 15 pessoas é, aleatoriamente, designada para um dos três grupos.
Cada pessoa tem 20 minutos para memorizar uma lista de 10 sílabas sem sentido, as quais serão usadas no teste no dia seguinte.

Grupo1 - condição sem ruído: cinco pessoas, designados para este grupo, estudam a lista de sílabas sem sentido enquanto estão em uma sala silenciosa.

Grupo 2 - condição de ruído moderado: cinco pessoas, designadas para este grupo, estudam a lista de sílabas sem sentido enquanto ouvem música clássica.

Grupo 3 - condição de ruído extremo: cinco pessoas, designados para este grupo, estudam a lista de sílabas sem sentido enquanto ouvem música rock.

Os dados indicam que o ruído influenciou o desempenho dos sujeitos?

Duas aplicações

Efeito da infraestrutura sustentável e da prestação de serviços no turismo sustentável: aplicação do teste de Kruskal Wallis (não paramétrico)

(Wani, Gowhar Ahmad, and V. Nagaraj. "Effect of Sustainable Infrastructure and Service Delivery on Sustainable Tourism: Application of Kruskal Wallis Test (Non-parametric)." *International Journal of Sustainable Transportation Technology* 5.2 (2022): 38-50.)

Objetivo: examinar o efeito da infraestrutura sustentável e da prestação de serviços no turismo sustentável no Vale da Caxemira por meio dos cinco principais destinos, abrangendo as partes interessadas essenciais. Os autores apresentam resumos da investigação empírica do efeito da infraestrutura de transporte sustentável e da prestação de serviços no turismo sustentável no Vale da Caxemira.

Amostra: principais destinos importantes do Vale da Caxemira (Srinagar, Pahalgam, Gulmarg, Kokernag e Yusmarg).

Análise: aplicação doo teste Kruskal-Wallis para avaliar a variação no turismo sustentável com base no desempenho da infraestrutura sustentável e da prestação de serviços nos destinos. **O turismo sustentável difere pelo desempenho da infraestrutura de transporte, outros elementos e qualidade do serviço dos destinos.**

Angiogênese na Neoplasia Escamosa do Colo Uterino:

Comparação entre dois marcadores de Células Endoteliais

(Calux, Nilciza Maria de Carvalho Tavares, et al. "Angiogênese na neoplasia escamosa do colo uterino: comparação entre dois marcadores de células endoteliais." *Revista Brasileira de Ginecologia e Obstetrícia,* 23 (2001): 313-319.)

Objetivo: comparar a acurácia de dois marcadores de células endoteliais anti-CD34 e anti-fator VIII em neoplasia cervical uterina, em lesões intra-epiteliais e no colo normal.

Amostra: 58 pacientes atendidas no Departamento de Ginecologia da Universidade Federal de São Paulo - Escola Paulista de Medicina, no período de maio de 1989 a abril de 1997.

- Grupo A: mulheres com diagnóstico anatomopatológico de neoplasia escamosa invasiva
- Grupo B: neoplasia intra-epitelial de alto grau
- Grupo C: neoplasia intra-epitelial de baixo grau
- Grupo D: mulheres sem qualquer processo neoplásico.

Análise: testes de Wilcoxon e de Kruskal-Wallis.

O desempenho dos marcadores nos quatro grupos é diferente?

Observações:

- as amostras são independentes

- teste usado para comparar mais de duas amostras independentes

- é uma extensão do teste de Mann-Whitney-Wilcoxon

- equivale à análise de variância, sem as suposições de normalidade e homogeneidade

- aqui também são usados postos (escala de medidas pelo menos ordinal)

- usado testar hipóteses de que diferentes amostras foram retiradas da mesma população ou de populações idênticas com a mesma mediana.

- para aplicar, são necessárias pelo menos três amostras independentes e, cada amostra com, no mínimo, cinco observações.

Teste de Kruskal-Wallis

- **Amostras:**
  - k amostras
  - as amostras tem tamanhos, possivelmente, diferentes: $n_1, n_2, \ldots, n_k$
  - atribuímos postos para todas as observações (independente da amostra a que pertence)

- **Suposições**:
  - todas as amostras são selecionadas aleatoriamente independentes, de suas respectivas populações
  - escala de medidas é, pelo menos, ordinal

- **Hipóteses:**

H_0: todas as k funções distribuições são idênticas

vs

H_1: pelo menos uma tende a produzir valores maiores dos que, pelo menos, uma das populações.

- **Teste estatístico**:

$$T = \frac{12}{N(N+1)} \left(\sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1)$$

em que: $R_j = \sum_{i=1}^{n_j} R(X_{ij}), \quad j = 1, \ldots, k$ e $N = \sum_{j=1}^{k} n_j$

- **Regra de Decisão**:
  - rejeite H_0, ao nível α de significância, se T exceder o quantil (1−α) da distribuição qui-quadrado com (*k*-1) g.l.

ANOVA *vs* Kruskal-Wallis

- ANOVA
  - as variáveis devem ser quantitativas e, devemos verificar as suposições de normalidade, variância constante e independência dos resíduos
  - teste para médias (e a variação entre os grupos).

- Kruskal-Wallis
  - suposição de que as observações são independentes e que a escala de medidas seja, pelo menos, ordinal
  - teste para as diferenças de médias de ordens (postos), que não são (necessariamente) iguais às medianas dos grupos.

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste de Kruskal-Wallis

1. *Teste do psicólogo: ruídos e aprendizado*

Um psicólogo conduz um estudo para determinar se o ruído pode ou não inibir o aprendizado. Cada um das 15 pessoas é, aleatoriamente, designada para um dos três grupos. Cada pessoa tem 20 minutos para memorizar uma lista de 10 sílabas sem sentido, as quais serão usadas no teste no dia seguinte.

- Grupo1 - condição sem ruído: cinco pessoas estudam a lista de sílabas enquanto estão em uma sala silenciosa
- Grupo 2 - condição de ruído moderado: cinco pessoas estudam a lista de sílabas enquanto ouvem música clássica
- Grupo 3 - condição de ruído extremo: cinco pessoas estudam a lista de sílabas enquanto ouvem música rock.

O número de sílabas sem sentido lembradas corretamente pelas 15 pessoas é:

Grupo 1: (8, 10, 9, 10, 9)

Grupo 2: (7, 8, 5, 8, 5)

Grupo 3: (4, 8, 7, 5, 7)

Podemos afirmar que o ruído influenciou o desempenho das pessoas?

Vamos analisar os dados!

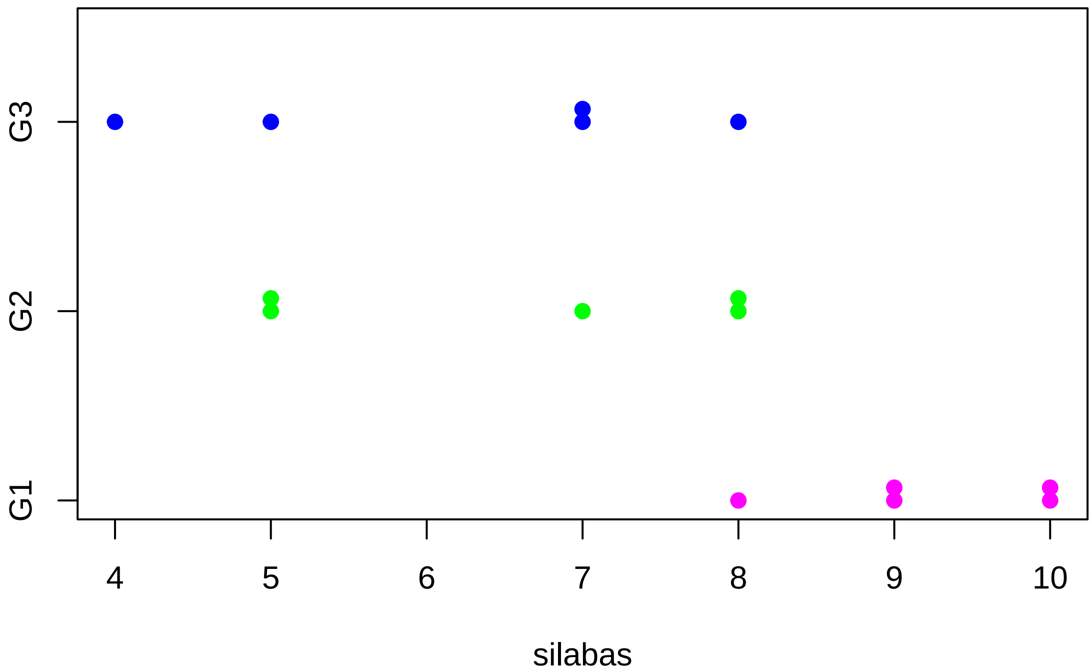
```r
silabas <- matrix(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7),
                    ncol = 3, byrow=FALSE)

grupos <- matrix(c(rep("G1", 5), rep("G2", 5), rep("G3", 5)), ncol = 3,
                 byrow = FALSE)

df <- data.frame(silabas, grupos)
```

Pelo gráfico de pontos, podemos observar o comportamento do número de sílabas lembradas corretamente.

```r
stripchart(silabas ~ grupos, col = c("magenta","green", "blue"),
           pch = c(19, 19,19), method = "stack", xlab = "silabas")
```



Para aplicação do teste de Kruskal-Wallis, fazemos:

```
G1 <- c(8, 10, 9, 10, 9)
G2 <- c(7, 8, 5, 8, 5)
G3 <- c(4, 8, 7, 5, 7)
kruskal.test(list(G1, G2, G3))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  list(G1, G2, G3)
## Kruskal-Wallis chi-squared = 8.7474, df = 2, p-value = 0.0126
```

Podemos dizer, que não existe evidências na amostra para aceitar H_0 (p-valor $= 0,0126$).

Vamos conferir a estatística teste!

- obtendo os postos para a amostra combinada

```
silabas <- matrix(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7),
                    ncol = 1, byrow=FALSE)

grupos <- matrix(c(rep("G1", 5), rep("G2", 5), rep("G3", 5)), ncol = 1,
                 byrow = FALSE)

df <- data.frame(silabas, grupos, rank(silabas))
```

```r
df
```

```
##    silabas grupos rank.silabas.
## 1        8     G1           9.5
## 2       10     G1          14.5
## 3        9     G1          12.5
## 4       10     G1          14.5
## 5        9     G1          12.5
## 6        7     G2           6.0
## 7        8     G2           9.5
## 8        5     G2           3.0
## 9        8     G2           9.5
## 10       5     G2           3.0
## 11       4     G3           1.0
## 12       8     G3           9.5
## 13       7     G3           6.0
## 14       5     G3           3.0
## 15       7     G3           6.0
```

```r
attach(df)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     grupos, silabas
```

- ordenando os postos

```r
df%>%arrange(silabas)
```

```
##    silabas grupos rank.silabas.
## 1        4     G3           1.0
## 2        5     G2           3.0
## 3        5     G2           3.0
## 4        5     G3           3.0
## 5        7     G2           6.0
## 6        7     G3           6.0
## 7        7     G3           6.0
## 8        8     G1           9.5
## 9        8     G2           9.5
## 10       8     G2           9.5
## 11       8     G3           9.5
## 12       9     G1          12.5
## 13       9     G1          12.5
## 14      10     G1          14.5
## 15      10     G1          14.5
```

- extraindo postos para os três grupos

```
posto.G1 <-df[df$grupos == "G1",]
posto.G1
```

```
##   silabas grupos rank.silabas.
## 1       8     G1           9.5
## 2      10     G1          14.5
## 3       9     G1          12.5
## 4      10     G1          14.5
## 5       9     G1          12.5
```

```
R.1 <- sum(posto.G1$rank.silabas)
R.1
```

```
## [1] 63.5
```

```
posto.G2 <-df[df$grupos == "G2",]
posto.G2
```

```
##    silabas grupos rank.silabas.
## 6        7     G2           6.0
## 7        8     G2           9.5
## 8        5     G2           3.0
## 9        8     G2           9.5
## 10       5     G2           3.0
```

```
R.2 <- sum(posto.G2$rank.silabas)
R.2
```

```
## [1] 31
```

```
posto.G3 <-df[df$grupos == "G3",]
posto.G3
```

```
##    silabas grupos rank.silabas.
## 11       4     G3           1.0
## 12       8     G3           9.5
## 13       7     G3           6.0
## 14       5     G3           3.0
## 15       7     G3           6.0
```

```
R.3 <- sum(posto.G3$rank.silabas)
R.3
```

```
## [1] 25.5
```

- calculando a estatística teste

```
T <- (12/(15*16)) * ((R.1^2)/5 + (R.2^2)/5 + (R.3^2)/5) - 3*16
T
```

```
## [1] 8.435
```

Observe que T é um valor muito próximo da estatística do teste Kruskal-Wallis obtida acima.

Exercício: ***Marcadores de células endoteliais anti-CD34 e anti-fator VIII***

Objetivo: comparar a acurácia do marcador de células endoteliais anti-CD34 e anti-fator VIII em neoplasia cervical uterina, em lesões intra-epiteliais e no colo normal.

Amostra:

- Grupo A: mulheres com diagnóstico anatomopatológico de neoplasia escamosa invasiva
- Grupo B: neoplasia intra-epitelial de alto grau
- Grupo C: neoplasia intra-epitelial de baixo grau
- Grupo D: mulheres sem qualquer processo neoplásico

Os valores correspondentes ao número de vasos avaliados nos carcinomas invasivos de colo uterino (A), lesões intra-epiteliais de alto grau (B) e de baixo grau (C) e colos uterinos normais (D) utilizando a técnica imuno-histoquímica para identificação de estruturas vasculares com anti-CD34, são:

Grupo A: (161, 160, 128, 168, 131, 107, 219, 147, 175, 119, 190, 203, 107, 153, 128, 209, 142, 124)

Grupo B: (128, 131, 125, 141, 157, 132, 143, 112, 131, 128, 139, 135, 127, 121, 168)

Grupo C: (121, 98, 81, 128, 91, 117, 136, 95, 105, 128, 90, 151, 129, 95, 112)

Grupo D: (109, 65, 97, 96, 110, 67, 106, 102, 80, 96)

O que podemos concluir a respeito do número de vasos avaliados nos carcinomas nos diferentes grupos?

Vamos analisar os dados!

- leitura dos valores:

```
GA <- c(161, 160, 128, 168, 131, 107, 219, 147, 175, 119, 190, 203, 107, 153, 128,
        209, 142, 124)

GB <- c(128, 131, 125, 141, 157, 132, 143, 112, 131, 128, 139, 135, 127, 121, 168)

GC <- c(121, 98, 81, 128, 91, 117, 136, 95, 105, 128, 90, 151, 129, 95, 112)

GD <- c(109, 65, 97, 96, 110, 67, 106, 102, 80, 96)
```

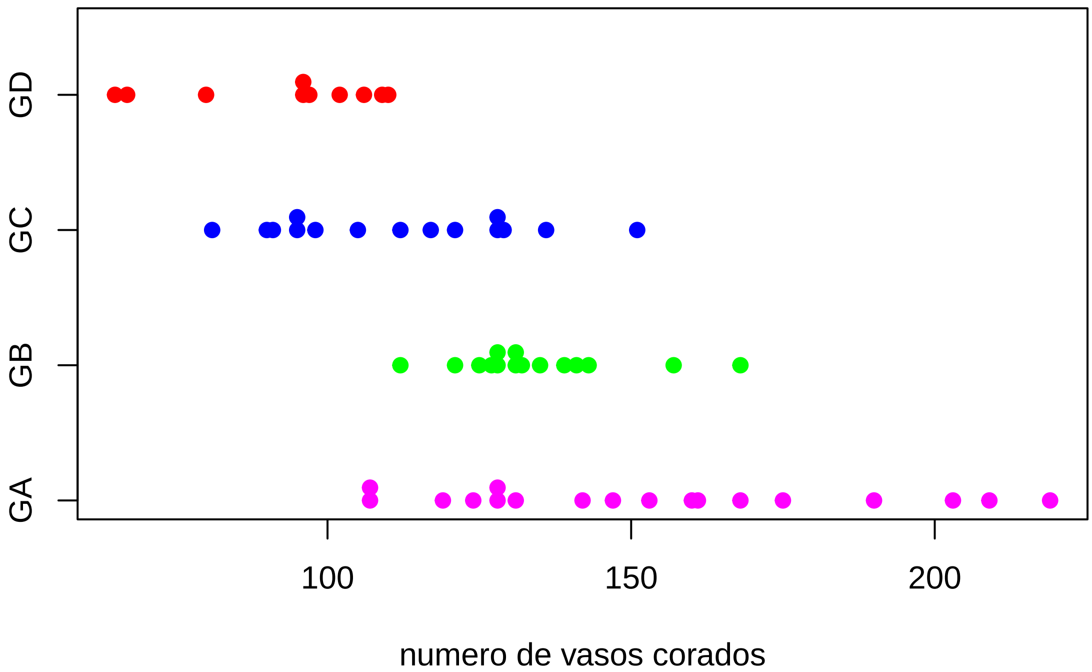
- construção da matriz de dados:

```
vasos.corados <- matrix(c(GA, GB, GC, GD),ncol = 1, byrow = FALSE)

grupos <- matrix(c(rep("GA", length(GA)), rep("GB", length(GB)),
                   rep("GC", length(GC)), rep("GD", length(GD))),
                   ncol = 1, byrow = FALSE)
```

- contrução do gráfico de pontos:

```
stripchart(vasos.corados ~ grupos, col = c("magenta","green", "blue",
           "red"),pch = c(19, 19, 19, 19), method = "stack",
            xlab = "numero de vasos corados")
```




```r
kruskal.test(list(GA, GB, GC, GD))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  list(GA, GB, GC, GD)
## Kruskal-Wallis chi-squared = 29.628, df = 3, p-value = 1.653e-06
```

Podemos dizer, para estes resultados, que os grupos apresentam comportamento diferentes $(= 1,653e-06)$.

Vamos conferir a estatística teste para esta aplicação

- construção do *dataframe* e obtenção dos postos:

```r
df <- data.frame(vasos.corados, grupos, rank(vasos.corados))
head(df, n = 6L)
```

```
##   vasos.corados grupos rank.vasos.corados.
## 1           161     GA                51.0
## 2           160     GA                50.0
## 3           128     GA                31.5
## 4           168     GA                52.5
## 5           131     GA                37.0
## 6           107     GA                16.5
```

```r
tail(df, n = 6L)
```

```
##    vasos.corados grupos rank.vasos.corados.
## 53           110     GD                19.0
## 54            67     GD                 2.0
## 55           106     GD                15.0
## 56           102     GD                13.0
## 57            80     GD                 3.0
## 58            96     GD                 9.5
```

```r
df%>%arrange(vasos.corados)
```

```
##    vasos.corados grupos rank.vasos.corados.
## 1             65     GD                 1.0
## 2             67     GD                 2.0
## 3             80     GD                 3.0
## 4             81     GC                 4.0
## 5             90     GC                 5.0
## 6             91     GC                 6.0
## 7             95     GC                 7.5
## 8             95     GC                 7.5
## 9             96     GD                 9.5
## 10            96     GD                 9.5
## 11            97     GD                11.0
## 12            98     GC                12.0
## 13           102     GD                13.0
## 14           105     GC                14.0
## 15           106     GD                15.0
## 16           107     GA                16.5
## 17           107     GA                16.5
## 18           109     GD                18.0
## 19           110     GD                19.0
## 20           112     GB                20.5
## 21           112     GC                20.5
## 22           117     GC                22.0
## 23           119     GA                23.0
## 24           121     GB                24.5
## 25           121     GC                24.5
## 26           124     GA                26.0
## 27           125     GB                27.0
## 28           127     GB                28.0
## 29           128     GA                31.5
## 30           128     GA                31.5
## 31           128     GB                31.5
## 32           128     GB                31.5
## 33           128     GC                31.5
## 34           128     GC                31.5
## 35           129     GC                35.0
## 36           131     GA                37.0
## 37           131     GB                37.0
## 38           131     GB                37.0
## 39           132     GB                39.0
## 40           135     GB                40.0
## 41           136     GC                41.0
## 42           139     GB                42.0
## 43           141     GB                43.0
## 44           142     GA                44.0
## 45           143     GB                45.0
## 46           147     GA                46.0
## 47           151     GC                47.0
## 48           153     GA                48.0
## 49           157     GB                49.0
## 50           160     GA                50.0
## 51           161     GA                51.0
```

```
## 52            168     GA                  52.5
## 53            168     GB                  52.5
## 54            175     GA                  54.0
## 55            190     GA                  55.0
## 56            203     GA                  56.0
## 57            209     GA                  57.0
## 58            219     GA                  58.0
```

```
posto.GA <-df[df$grupos == "GA",]
#posto.GA
R.A <- sum(posto.GA$rank.vasos.corados)
R.A
```

```
## [1] 753.5
```

```
posto.GB <-df[df$grupos == "GB",]
#posto.GB
R.B <- sum(posto.GB$rank.vasos.corados)
R.B
```

```
## [1] 547.5
```

```
posto.GC <-df[df$grupos == "GC",]
#posto.GC
R.C <- sum(posto.GC$rank.vasos.corados)
R.C
```

```
## [1] 309
```

```
posto.GD <-df[df$grupos == "GD",]
#posto.GD
R.D <- sum(posto.GD$rank.vasos.corados)
R.D
```

```
## [1] 101
```

- cálculo da estatística do teste:

```
N = sum(length(GA), (length(GB)), length(GC), length(GD))
N
```

```
## [1] 58
```

```
T <- (12/(N*(N+1))) * ((R.A^2)/length(GA) + (R.B^2)/length(GB) +
                  (R.C^2)/length(GC) + (R.D^2)/length(GD)) - 3*(N + 1)
T
```

```
## [1] 29.58655
```

Observe o valor da estatística obtida no teste.

… Mas, e agora?

Aplicação

Angiogênese na Neoplasia Escamosa do Colo Uterino:

Comparação entre dois marcadores de Células Endoteliais

(Calux, Nilciza Maria de Carvalho Tavares, et al. "Angiogênese na neoplasia escamosa do colo uterino: comparação entre dois marcadores de células endoteliais." *Revista Brasileira de Ginecologia e Obstetrícia,* 23 (2001): 313-319.)

Objetivo: comparar a acurácia de dois marcadores de células endoteliais anti-CD34 e anti-fator VIII em neoplasia cervical uterina, em lesões intra-epiteliais e no colo normal.

Amostra: 58 pacientes atendidas no Departamento de Ginecologia da Universidade Federal de São Paulo - Escola Paulista de Medicina, no período de maio de 1989 a abril de 1997.

- Grupo A: mulheres com diagnóstico anatomopatológico de neoplasia escamosa invasiva
- Grupo B: neoplasia intra-epitelial de alto grau
- Grupo C: neoplasia intra-epitelial de baixo grau
- Grupo D: mulheres sem qualquer processo neoplásico.

Aplicando o teste de Kruskal-Wallis (p-valor = 1,653e-06), podemos dizer que o desempenho dos marcadores nos quatro grupos é diferente.

Qual(is) grupo(s) tem comportamento diferente dos demais?

Testes de comparações múltiplas

Teste de comparações múltiplas para amostras independentes

- comparar mais do que duas amostras independentes

- **detectar possíveis diferenças entre os diferentes tratamentos**

- para todos os **pares de tratamentos**, fazemos **comparações.**

- **Hipóteses**:

H_0: tratamentos *i* e *j* **não** produzem resultados diferentes

vs

H_1 :tratamentos *i* e *j* produzem resultados diferentes

- **Estatística:**

$$|\bar{R}_i - \bar{R}_j| > \sqrt{\chi^2_{(\alpha,k-1)} \frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

em que: $\chi^2_{(\alpha,k-1)}$ é o quantil da distribuição Qui-quadrado com com (*k-1*) g.l.

- **Regra de Decisão:**

- rejeite H_0 se a desigualdade é satisfeita.

- **Observações:**

  - teste de Dunn: a aproximação neste teste é calculada como a diferença nas pontuações médias de classificação dividida pela estimativa de variância combinada de classificação para dois grupos:

$$|\bar{R}_i - \bar{R}_j| \geq z_{1-\alpha/2^*} \sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- outro teste: Nemenyi

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste comparações múltiplas: Kruskal-Wallis

1. *Teste do psicólogo: ruídos e aprendizado*

Um psicólogo conduz um estudo para determinar se o ruído pode ou não inibir o aprendizado. Cada um das 15 pessoas é, aleatoriamente, designada para um dos três grupos. Cada pessoa tem 20 minutos para memorizar uma lista de 10 sílabas sem sentido, as quais serão usadas no teste no dia seguinte.

- Grupo1 - condição sem ruído: cinco pessoas estudam a lista de sílabas enquanto estão em uma sala silenciosa
- Grupo 2 - condição de ruído moderado: cinco pessoas estudam a lista de sílabas enquanto ouvem música clássica
- Grupo 3 - condição de ruído extremo: cinco pessoas estudam a lista de sílabas enquanto ouvem música rock.

O número de sílabas sem sentido lembradas corretamente pelas 15 pessoas é:

Grupo 1: (8, 10, 9, 10, 9)

Grupo 2: (7, 8, 5, 8, 5)

Grupo 3: (4, 8, 7, 5, 7)

Os dados indicam que o ruído influenciou o desempenho dos sujeitos?

Vamos analisar os dados!

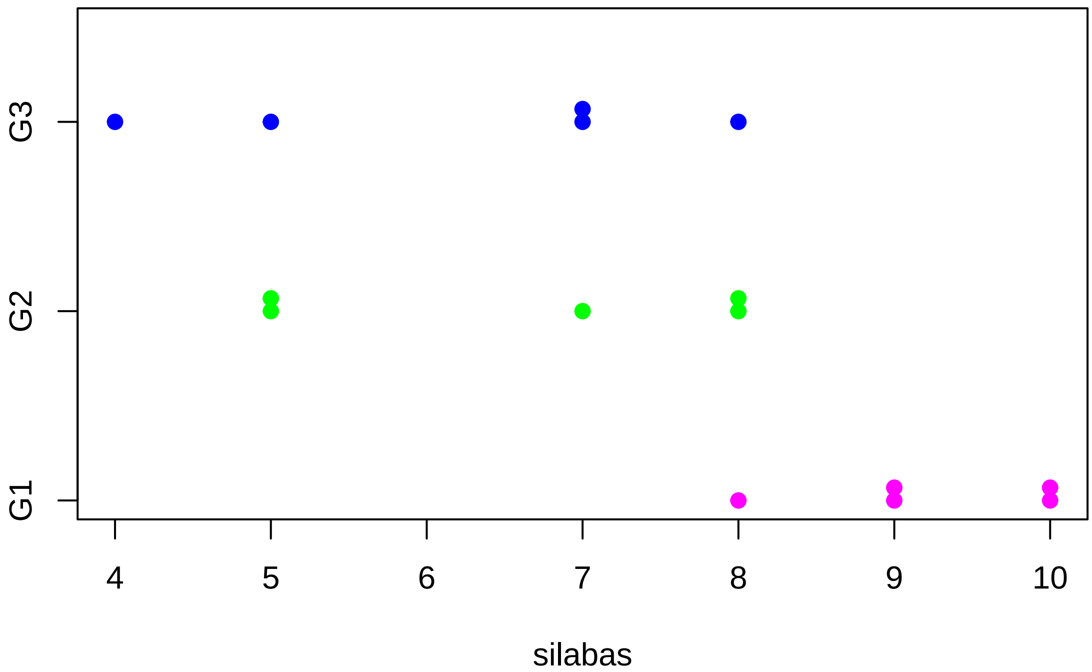
```
silabas <- matrix(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7),
                    ncol = 3, byrow=FALSE)

grupos <- matrix(c(rep("G1", 5), rep("G2", 5), rep("G3", 5)), ncol = 3,
                 byrow = FALSE)

df <- data.frame(silabas, grupos)
```

Pelo gráfico de pontos, podemos observar o comportamento do número de sílabas lembradas corretamente.

```
stripchart(silabas ~ grupos, col = c("magenta","green", "blue"),
           pch = c(19, 19,19), method = "stack", xlab = "silabas")
```



Para aplicação do teste de Kruskal-Wallis, fazemos:

```
silabas <- c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7)

grupos <- c(rep("G1", 5), rep("G2", 5), rep("G3", 5))

kruskal.test(silabas~grupos)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  silabas by grupos
## Kruskal-Wallis chi-squared = 8.7474, df = 2, p-value = 0.0126
```

A prtir do teste de Kruskal-Wallis, aplicamos o teste de comparações múltiplas.

```
kwAllPairsConoverTest(silabas~factor(grupos))
```

```
## Warning in kwAllPairsConoverTest.default(c(8, 10, 9, 10, 9, 7, 8, 5, 8, : Ties
## are present. Quantiles were corrected for ties.

##
##  Pairwise comparisons using Conover's all-pairs test

## data: silabas by factor(grupos)

##    G1     G2
## G2 0.0106 -
## G3 0.0036 0.8235

##
## P value adjustment method: single-step
```

Observando os p-valores das comparações duas a duas, pelo teste de comparçoes múultiplas, podemos dizer que existe diferença significativa entre os grupos de condição sem ruído e moderado (G1 e G2) e, entre os grupos de condição sem ruído e ruído extremo (G1 e G4).

Outro teste bastante utilzado é o teste de Dunn:

```
kwAllPairsDunnTest(silabas~factor(grupos))
```

```
## Warning in kwAllPairsDunnTest.default(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, : Ties
## are present. z-quantiles were corrected for ties.

##
##  Pairwise comparisons using Dunn's all-pairs test

## data: silabas by factor(grupos)

##    G1    G2
## G2 0.039 -
## G3 0.019 0.692

##
## P value adjustment method: holm

## alternative hypothesis: two.sided
```

no qual obtemos o memso resultado.

Exercício: ***Marcadores de células endoteliais anti-CD34 e anti-fator VIII***

Objetivo: comparar a acurácia do marcador de células endoteliais anti-CD34 e anti-fator VIII em neoplasia cervical uterina, em lesões intra-epiteliais e no colo normal.

Amostra:

- Grupo A: mulheres com diagnóstico anatomopatológico de neoplasia escamosa invasiva
- Grupo B: neoplasia intra-epitelial de alto grau
- Grupo C: neoplasia intra-epitelial de baixo grau
- Grupo D: mulheres sem qualquer processo neoplásico

Os valores correspondentes ao número de vasos avaliados nos carcinomas invasivos de colo uterino (A), lesões intra-epiteliais de alto grau (B) e de baixo grau (C) e colos uterinos normais (D) utilizando a técnica imuno-histoquímica para identificação de estruturas vasculares com anti-CD34, são:

Grupo A: (161,160,128,168,131,107,219,147,175,119,190,203,107,153,128, 209,142,124)

Grupo B: (128,131,125,141,157,132,143,112,131,128,139,135,127,121,168)

Grupo C: (121,98,81,128,91,117,136,95,105,128,90,151,129,95,112)

Grupo D: (109,65,97,96,110,67,106,102,80,96)

O que podemos concluir a respeito do número de vasos avaliados nos carcinomas nos diferentes grupos? Se existe pelo menos um grupo diferente dos demais, qual(ais) par(es) produzem resultados diferentes?

Testes para igualdade de variâncias

ou

Testes de escala

- Dois termos bastante usados na Estatística: **homogeneidade e heterogeneidade**
- Estes são usados para descrever propriedades de conjuntos de dados
- Por exemplo: podem ser usados para nos referirmos à média ou variância
- Alguns testes (paramétricos) tem suposição de homogeneidade de variâncias. Como validar esta suposição?
- A heterogeneidade pode invalidar resultados obtidos em testes ou causar erros na modelagem estatística de dados, quando da suposição de homogeneidade.

• No teste de hipótese de homogeneidade testamos se as populações têm variâncias iguais.

• Testes:

- paramétricos => **F de Fisher** para comparar variâncias de duas populações (independentes), com a suposição de normalidade dos dados.
  Este teste é bastante sensível à violação da suposição de normalidade.

- não paramétricos => de Levene (alternativa ao teste de Bartlett), baseado na soma de postos, de Mood, de Ansari-Bradley e, Fligner-Killeen, entre outros.

Testes baseados em postos

- É um teste não paramétrico usado para comparar a dispersão (variância ou escala) de **duas** amostras independentes.

- É uma alternativa ao teste de **F de Fisher**, que assume normalidade das distribuições.

- **Amostras:**
  Considere duas variáveis aleatórias X e Y, cujas amostras foram retiradas destas populações, possivelmente de tamanhos diferentes.

- **Suposições**:
  - as amostras são independentes
  - as amostras tem distribuições simétricas.

- **Hipóteses**:
  - H_0: as duas populações têm a mesma dispersão (variância/escala).
  - H_1: as dispersões das populações são diferentes.

- **Estatística**:

$$T = \sum_{i=1}^{n} (R(U_{ij}))^2$$

- em que $U_i = |X_i - \mu_1|$

- μ_1 é a média da população de X; se for desconhecida, use a média amostral.

- **Regra de decisão:**
  - Rejeite H_0 se $T > w(\alpha/2)$ ou $T < w(1-\alpha/2)$, em que w é o quantil da distribuição dos quadrados dos postos (tabelado).

- **Observações:**

  - as hipóteses podem ser unilaterais

  - se existem empates, a distribuição de T, sob H_0, não é mais exata

  - é possível utilizar a aproximação para grandes amostras

  - para três ou mais populações, esse é modificado para testar igualdade de variâncias e a comparação ocorre através da soma total dos postos de cada amostra.

Teste de Ansari-Bradley

- É um teste baseado no grau de dispersão dos postos de duas amostras combinadas.

- **Ideia:** combine as duas amostras e atribua postos,
  de tal forma que a soma de quadrados dos desvios de todos os postos é zero.

- **Hipóteses**:
  - H_0: as duas populações têm a mesma dispersão (variância/escala).
  - H_1: as dispersões das populações são diferentes
    (Hipóteses unilaterais também podem ser testadas)

- **Amostras:**
  - considere duas variáveis aleatórias X e Y, cujas amostras foram retiradas destas populações, possivelmente de tamanhos diferentes.
  - $(x_1, x_2, \ldots, x_n)$ e $(y_1, y_2, \ldots, y_m)$, amostras.

Os postos são atribuídos, para as amostras combinadas, da seguinte maneira:

- posto 1 para as menor e maior observações,
- posto 2 para as segundas menor e maior observações,
- posto 3 para as terceiras menor e maior observações
- assim sucessivamente.

- **Estatística do teste**: (soma dos postos da amostra X)
  - se H_0 é falso, os valores de X tendem a estar nas caudas da amostra combinada e portanto receberá os menores postos

- **Regra de decisão:**
  - Rejeite H_0 se $T \geq w$, tal que $P(T \geq w) = \alpha$ (w é obtido da tabela de Ansari-Bradley).

Teste de Siegel-Tuckey

- Teste “concorrente” do teste F de Fisher, porém, sem a suposição de normalidade dos dados

- **Ideia:** verificar se **duas** amostras provêm de duas populações diferentes com variabilidades diferentes.

- **Suposições:**
  - as duas amostras foram selecionadas aleatoriamente e são independentes
  - as duas populações (das quais as amostras foram retiradas) têm medianas conhecidas.

- **Amostras:**
  - considere duas variáveis aleatórias X e Y, cujas amostras, $(x_1, x_2, \ldots, x_n)$ e $(y_1, y_2, \ldots, y_m)$, foram retiradas destas populações, possivelmente de tamanhos diferentes

  - encontre a mediana de cada uma das amostras: M_x e M_y => encontre o maior valor entre as duas medianas e subtraia o menor valor do maior valor

  - da amostra com maior mediana, subtraia o valor da diferença

  - combine as duas amostras (em que uma delas foi alterada) e atribua postos da seguinte forma:
    - atribua o posto 1 para a menor observação e o posto 2 para maior observação
    - atribua o posto 3 para segunda maior observação e 4 para a segunda menor observação
    - e assim sucessivamente.

- **Estatística do teste:**

$$T = R_X + \frac{n(n+1)}{2}$$

- **Regra de decisão:**
  - Rejeite H_0 se $T \geq w$, tal que $P(T \geq w) = \alpha$ (w é obtido da tabela de Wicoxon).

- **Aplicação:**

Métodos para verificar a quantidade de ferro sérico em solução aquosa

(Jung, D. H., and A. C. Parekh. A semi-micromethod for the determination of serum iron and iron-binding capacity without deproteinization. American Journal of Clinical Pathology, Vol. 54, no. 6 p. 813-817,1970)

Objetivo: comparar diferentes métodos que medem diretamente a quantidade de ferro sérico em uma solução. O objetivo principal é verificar se o novo método não perdeu acurácia em comparação ao método proposto por Ramsay. Vinte análises foram realizadas para cada método usando uma solução que continha 105μg de ferro sérico por 10ml.

Existe diferença entre as variabilidade dos dois métodos?

Vamos para o R!

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste de comparação de variâncias

1. *Quantidade de ferro sérico em solução aquosa*

Existe diferença entre os métodos?

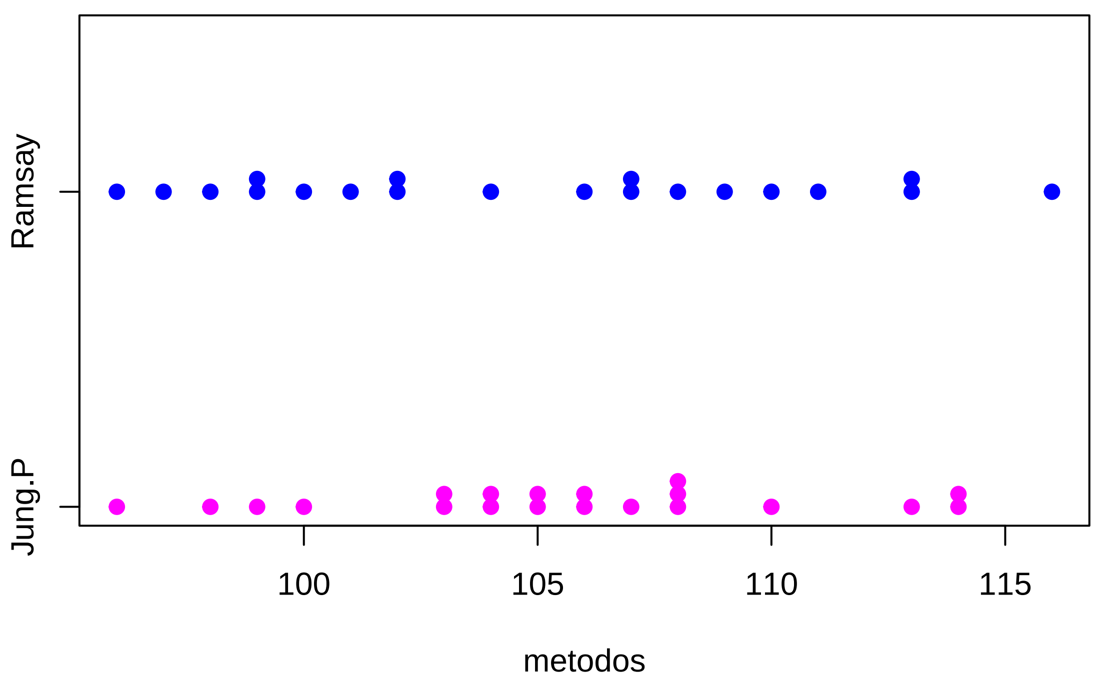
- Vamos verificar graficamente o comportamento dos dados:

```
Ramsay <- c(111, 107, 100, 99, 102, 106, 109, 108, 104, 99,
            101, 96, 97, 102, 107, 113, 116, 113, 110, 98)
Jung.P <- c(107, 108, 106, 98, 105, 103, 110, 105, 104,
            100, 96, 108, 103, 104, 114, 114, 113, 108, 106, 99)

metodos <- matrix(c(Ramsay, Jung.P), ncol =1, byrow = F)


grupos <- matrix(c(rep("Ramsay", length(Ramsay)), rep("Jung.P", length(Jung.P))),
                  ncol = 1, byrow = FALSE)


stripchart(metodos ~ grupos, col = c("magenta", "blue"),pch = c(19, 19),
          method = "stack",  xlab = "metodos")
```



- Aplicando os testes de Siegel-Tickey, Ansari-Bradley (ambos aplicados somente para duas amostras) e Levene (pode ser aplicado para mais de duas amostras), obtemos os mesmos resultados: podemos dizer que não existe evidências para rejeitar a hipótese de que as variâncias são iguais.

```
Ramsay <- c(111, 107, 100, 99, 102, 106, 109, 108, 104, 99,
            101, 96, 97, 102, 107, 113, 116, 113, 110, 98)
Jung.P <- c(107, 108, 106, 98, 105, 103, 110, 105, 104,
            100, 96, 108, 103, 104, 114, 114, 113, 108, 106, 99)

SiegelTukeyTest(Ramsay, Jung.P)
```

```
## Warning in wilcox.test.default(ranks0, ranks1, alternative = alternative, :
## cannot compute exact p-value with ties
```

```
##
##  Siegel-Tukey-test for equal variability
##
## data:  Ramsay and Jung.P
## ST = 360.83, p-value = 0.1837
## alternative hypothesis: true ratio of scales is not equal to 1
```

```
ansari.test(Ramsay, Jung.P, exact = FALSE)
```

```
##
##  Ansari-Bradley test
##
## data:  Ramsay and Jung.P
## AB = 185.5, p-value = 0.1815
## alternative hypothesis: true ratio of scales is not equal to 1
```

```
LeveneTest( c(Ramsay, Jung.P),
            factor(c(rep("Ramsay",length(Ramsay)), rep("Jung.P",length(Jung.P)))))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  1.7865 0.1893
##       38
```

2. Quinze amigos decidiram investir, cada um, 1.000 (unidades monetárias - u.m.). Um grupo (G1) investiu em CDBs, outro grupo (G2) investiu em Fundos DI e, o outro grupo (G3) investiu em Fundos de renda fixa. Os valores de retorno (u.m) foram:

G1: (146, 180, 192, 185, 153)

G2: (176, 110, 212, 108, 196)

G3: (-540, 1052, 642, 281, 67)

O que voce conclui a respeito da variância nos grupos?

(adaptado de Conover, W. J. Practical nonparametric statistics. John Wiley & Sons, 1999)

```
G1 <- c(146, 180, 192, 185, 153)
G2 <- c(176, 110, 212, 108, 196)
G3 <- c(-540, 1052, 642, 281, 67)
```

```r
LeveneTest( c(G1, G2, G3),
            factor(c(rep("G1",length(G1)), rep("G2", length(G2)),
                     rep("G3", length(G3)))))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  2  6.5017 0.01222 *
##       12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observando os resultados do teste, o que podemos afirmar?

3. Em um experimento, os resultados do comprimento da parte aérea (cm), após 60 dias da semeadura de dois híbridos de milho:

Híbrido 1: (167, 190, 217, 220, 222, 258, 262, 287)

Híbrido 2: (116, 132, 148, 162, 188, 210, 215, 252)

Para estas amostras, podemos dizer que as variâncias são diferentes?

(Cargnelutti Filho, A., *et al.* Testes não paramétricos para pesquisas agrícolas, Santa Maria, UFSM, 2001)

Existe…

- **associação** entre temperatura e umidade do ar?
- **correlação** entre índices de glicemia e colesterol alto?
- **associação** entre nota no ENEM e tipo de escola (pública/particular)?
- **associação** entre atividade física e a ocorrência de diabetes melittus em homens?
- **correlação** entre área construída e preço de imóveis?
- **correlação** entre renda e consumo?

- Estamos interessados em verificar se existe “influência” - associação/correlação - de uma variável sob outra variável.

- Se existe, qual o grau da dependência?

  - que tipo de variáveis serão analisadas (qual a escala de medidas)?
  - quantas variáveis serão analisadas?
  - os resultados “valem” em qualquer contexto?
  - as variáveis são associadas ou relacionadas: mesmo sentido?

E agora? O que usar?

As medidas de associação e correlação são usadas para avaliar relações entre variáveis,
com objetivos e interpretações diferentes.

Definições

- **Correlação**
  - formalmente: correlação significa relação mútua entre dois termos ou estabelecer relação ou correlação entre os termos
  - interesse: avaliar a "força" ou 'grau" da dependência linear entre duas **variáveis quantitativas**
  - não indica **causalidade**, apenas a intensidade e direção da relação
  - coeficientes mais utilizados:
    - **Pearson**: relação linear entre duas variáveis quantitativas; usa os valores registrados
    - **Spearman**: mede a intensidade da relação entre variáveis ordinais; usa os postos
    - **Kendall**: mede a intensidade da relação entre duas variáveis ordinais, pareadas: concordância dos pares.

- **Associação**
  - "força" e "direção" da relação entre **variáveis**, geralmente, **categóricas**
  - coeficientes mais usados:
    - **razão de chances** ou ***Odds Ratio***: usada para verificar a chance de um evento ocorrer em um grupo em relação a outro
    - coeficientes de contingência : usado para verificar a associação entre variáveis categóricas em tabelas de contingência.

- Também existem suposições para o uso destas medidas.

Duas variáveis qualitativas

- Os dados representam contagens das classes de respostas de duas **variáveis qualitativas**, geralmente expressos no formato de tabelas de contingência.

- Objetivo: testar a **associação** entre duas variáveis qualitativas.

Grupos	Variável		Total
	Não	Sim	
Grupo 1	n_{11}	n_{12}	$n_{1.}$
Grupo 2	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

- Aplicado nas situações em que os tamanhos das duas amostras são pequenos.

- Alternativa ao teste de Qui-quadrado: situações em que as frequências observadas/esperadas são menores do que 5.

- Alguns coeficientes e testes:

  - coeficiente de Pearson e de Cramér
  - teste Qui-quadrado de Pearson, exato de Fisher.

Teste exato de Fisher

- **Amostra:**
  - duas variáveis qualitativas e duas classes de resposta (em cada variável)
  - cada observação é classificada em uma única célula
  - os totais das linhas e colunas são fixos

- **Hipóteses:**
  - H_0: não existe diferença entre as proporções observadas nos dois grupos (os grupos são independentes; não existe associação entre os dois grupos)
  - H_1: as proporções são diferentes nos dois grupos.

  (também possível testar hipóteses unilaterais)

- **Estatística:**

$$T = n_{11}$$

- **Regra de decisão:**

  - Encontre o p-valor usando a distribuição da estatística T (distribuição hipergeométrica).
  - Usual: encontrar o p-valor:
    - some todas as probabilidades $p(k) = P(X = k)$ para todos os k tais que $p(k) \leq p(x_0)$, sendo x_0 o valor observado na tabela, isto é, o p-valor é tal que $p = P(p(k) \leq p(x_0))$.

- Para variáveis com mais do que duas classes de respostas:
  - obtemos a estatística de Pearson:

$$\chi^2 = \sum_{i=1}^{l}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

em que *c* é o numero de colunas, *l* é o numero de linhas e,

$$E_{ij} = \frac{(\text{total de observações na linha } i) \times (\text{total de observações na coluna } j)}{\text{total de observações}}$$

- Alguns coeficientes baseados na estatística de Pearson:

- **Coeficiente de Cramér:**

$$T_1 = \sqrt{\frac{\chi^2}{n \times (q-1)}},$$

  - q é o menor valor entre o número de linhas e o número de colunas
  - $0 < T_1 < 1$.
  - apropriado para tabelas maiores que 2 × 2.

- **Coeficiente de McNemar e Siegel:**

$$T_2 = \sqrt{\frac{\chi^2}{n + \chi^2}},$$

  - $0 < T_2 < 1$.

- Para ambas as medidas: 0 => não associação e, 1 => forte associação entre as duas variáveis.

Variáveis quantitativas

Correlação linear: coeficiente de Pearson

- Considere um par de variáveis contínuas (X, Y) e uma amostra de observações (pareadas) - independentes - destas variáveis: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

- Suposição: os valores observados têm distribuição normal (principalmente se $n < 40$).

- Usando um diagrama de dispersão é possível verificar a existência de um relação linear entre as variáveis

- **Coeficiente**:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

- **Observações 1:**

- assume valores no intervalo [-1,1]

- o sinal indica direção - positiva ou negativa - da relação:
  - se “+” => relação positiva entre as duas variáveis
  - se “-” => relação negativa entre as duas variáveis

- o valor sugere a força da relação entre as variáveis

- o **coeficiente não diferencia entre variáveis independentes e variáveis dependentes** (correlação entre X e Y é igual à correlação entre Y e X)

- o valor da correlação não muda se a unidade de mensuração das variáveis for alterada

- o coeficiente é **adimensional** (não tem unidade física)

- o coeficiente de **correlação é fortemente afetado** por valores discrepantes.

- **Observações 2:**

Esta medida:

Deve ser usada para:

- explorar relações: "ajudar" a determinar a força e a direção de uma relação linear entre variáveis, através de uma medida de coeficiente numérico
- auxiliar na seleção de variáveis em modelos preditos.

Não deve ser usada:

- para entender a causalidade: o coeficiente da análise de correlação não deve ser interpretado como uma evidência de **causalidade** entre as variáveis
- para descrever relações não-lineares
- no caso de existirem valores extremos no conjunto de dados.

- **Correlação espúria**

  - é uma relação entre variáveis que parecem estar correlacionadas, mas na verdade não têm relação de causa e efeito.

  - essa relação pode surgir devido a fatores externos ou **coincidências** estatísticas, levando a conclusões equivocadas, se não forem analisadas corretamente.

  - alguns exemplo: *site* ***Spurious correlations*****:**

  a) correlação alta (66%) entre o número de pessoas afogadas por cair em uma piscina com o número de filmes em que Nicolas Cage aparece => quanto maior o número de filmes em que Nicolas Cage aparece, maior a chance de pessoas de se afogarem?

  b) relação entre a taxa de divórcio nos EUA e o consumo *per capita* de margarina => consumir mais margarina implica em aumento de problemas conjugais?

Correlação ρ de Spearman

- Este coeficiente mede a intensidade da relação, entre variáveis **ordinais**.

- Atribuição de postos

- Este coeficiente não é sensível a assimetrias na distribuição, nem a presença de valores discrepantes. Por isso, não é necessário que os dados sejam provenientes de duas populações normais.

- Pode ser aplicado em variáveis quantitativas - como alternativa ao coeficiente de Pearson (pela suposição de normalidade dos dados)

- Nos casos em que os dados não formam uma nuvem “bem comportada” (observar no diagrama de dispersão), com alguns pontos muito afastados dos restantes, ou nos casos em que parece existir uma relação crescente ou decrescente em formato de curva, o coeficiente de Spearman é mais apropriado.

- Considere um par de variáveis, pelo menos ordinais, (X, Y) e uma amostra de observações pareadas - independentes - destas variáveis: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

- Atribua postos separadamente para as observações de X e para as observações de Y. Se ocorrerem empates, atribua o posto médio.

- **Coeficiente**:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left(R(X_i) - R(Y_i)\right)^2}{n(n^2-1)}$$

- Assume valores no intervalo entre [-1, 1]:
  - quanto mais próximo de 1 => mais forte e a correlação é positiva;
  - se for igual a (ou próximo de) zero => não existe correlação monotônica (as variáveis não têm relação ordenada).

- Como o coeficiente é obtido usando os postos, a escala das variáveis não afeta o valor da estatística

- A ideia é a mesma do coeficiente de Pearson, porém usando postos.

Coeficente τ de Kendall

- Medida que avalia a força e a direção da associação entre duas variáveis ordinais.

- Útil em situações nas quais os dados não seguem uma distribuição normal, o que a torna uma alternativa robusta ao coeficiente de correlação de Pearson.

- Usado para comparar pares de observações: para todos os pares (x_i, y_i), (x_j, y_j), verifica-se se estes estão em concordância ou discordância. Se são:

  - concordantes: ambos os valores de um par estão na mesma ordem com outro par

  - discordantes: a ordem é invertida.

- O coeficiente varia no intervalo [1, 1].

- Considere duas variáveis (pareadas), pelo menos ordinais, (X, Y) e uma amostra de observações - independentes - destas variáveis: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

- Dada todas as possíveis combinações de 2 pares, $\binom{n}{2}$, temos
  - N_C: o número de pares concordantes
  - N_D: o número de pares discordantes

- **Coeficiente**:

$$\tau = \frac{N_C - N_D}{\frac{n(n-1)}{2}}$$

  - se todos os pares forem concordantes => $\tau = 1$

  - se todos os pares forem discordantes => $\tau = -1$

- **Observações 3:**

- **Coeficiente de correlação de Pearson não** deve ser usado para testar **independência** entre duas variáveis.
  - se for próximo de zero => ausência de correlação linear, mas as variáveis ainda podem ter relação não linear.

- **Independência estatística** significa que o conhecimento de uma variável não fornece informação sobre a outra.

- **Testes adequado para independência**
  - para testar independência entre duas variáveis:
    - variáveis qualitativas => **teste Qui-quadrado de independência**
    - variáveis quantitativas => **teste de correlação de Spearman e de Kendall**.

- Com a aplicação dessas medidas (de Pearson, de Spearman e de Kendall) estimamos a influência total de uma variável aleatória sobre a outra, incluindo a influência indireta "sentida" porque a segunda variável aleatória está correlacionada não apenas com a primeira, mas talvez com uma terceira variável aleatória, que por sua vez está correlacionada com a primeira variável aleatória e, portanto, atua como portadora de influência indireta entre a primeira e a segunda variáveis aleatórias.

- Às vezes o interesse está obter uma medida da correlação entre duas variáveis, sob a condição de que a influência indireta devido às outras variáveis seja de alguma forma eliminada.

- Considerando três ou mais variáveis, os coeficientes de Pearson e de Kendall foram estendidos.

Coeficientes parciais de Pearson e de Spearman

- extensão dos coeficientes de correlação (observando a escala de medidas e suposições)

- mede a associação entre duas variáveis na situação em que a influência de uma variável sob as outras foi eliminada

- utilizado quando se deseja avaliar a relação entre duas variáveis, eliminando o efeito de uma terceira

• Caso particular: considere três variáveis, pelo menos ordinais, (X, Y, Z) e uma amostra de observações, independentes, destas variáveis: (x_1, y_1, z_1), (x_2, y_2, z_2),…, (x_n, y_n, z_n).

- **Coeficientes:**

$$\tau_{XY.Z} = \frac{\tau_{XY} - \tau_{XZ}\tau_{YZ}}{\sqrt{(1-\tau^2_{XZ})(1-\tau^2_{YZ})}} \quad \textbf{e} \quad r_{XY.Z} = \frac{r_{XY} - \tau_{XZ}r_{YZ}}{\sqrt{(1-r^2_{XZ})(1-r^2_{YZ})}}$$

- Se $\tau_{XY.Z}$ (ou $r_{XY.Z}$) for:

- próximo de 1 => X e Y estão fortemente correlacionados, eliminado o efeito de Z

- próximo de 0 => X e Y não são correlacionados, eliminado o efeito de Z

- negativo => a relação entre X e Y é inversa, eliminado o efeito de Z.

Vamos ao R!