

ESTATÍSTICA PARA SAÚDE COLETIVA

Aula 9

Recados

- Na próxima aula inicia-se a apresentação de seminário!

Nomes (Aluno ou Dupla)	Nome do artigo
ALEX SILVA + WILLIAM	Quedas entre idosos brasileiros residentes em áreas urbanas: ELSI-Brasil
DIEGO REIS	Resistência insulínica e sua relação com excesso de peso em adolescentes de uma capital do Nordeste Brasileiro
MARYNA LUZ + MARINA MENOTTI	Problemas emocionais e uso de medicamentos psicotrópicos: uma abordagem da desigualdade racial

Próxima semana

Nomes (Aluno ou Dupla)	Data seminário	Nome do artigo
ELZA DA SILVA	06/10/2020	Carga de trabalho de enfermagem em unidade de terapia intensiva destinada a pacientes com queimaduras
LETICIA GOMES	06/10/2020	Clima organizacional e trabalho em equipe na estratégia saúde da família
THAIS RIBEIRO	06/10/2020	Estudo quanti-qualitativo sobre amamentação exclusiva por gestantes de alto risco
CLAUDIA MARTINS	08/10/2020	Análise Espacial dos Casos de COVID-19 e leitos de terapia intensiva no estado do Ceará, Brasil
CLAUDINEIA SOARES TORRES	08/10/2020	Estimativas de impacto do Covid-19 na mortalidade de idosos institucionalizados no Brasil
PRISCILA PRATES	08/10/2020	Promoção comercial ilegal de produtos que competem com o aleitamento materno

Feedback lista 7 e 8

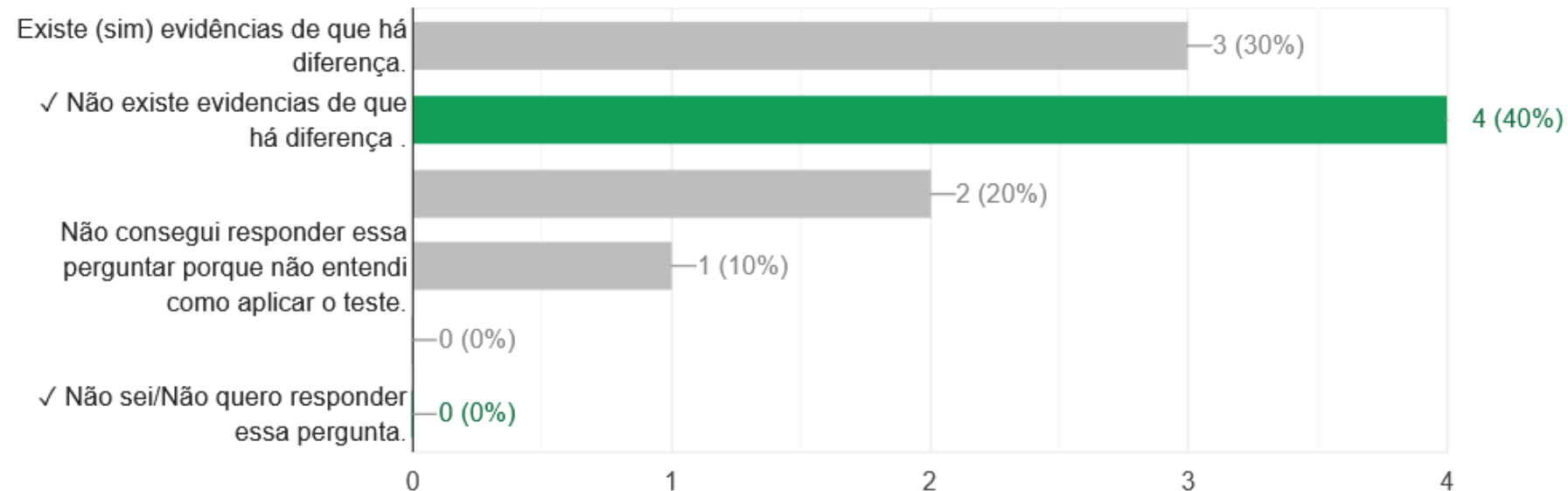
- Revisão das questões com maior erro

Feedback lista 7

Pergunta 6. Existe evidências de que há diferença na média do número de anos de prática de atividade física (Coluna) entre homens e mulheres (Coluna Sexo)? Para responder essa questão, aplique o teste de Wilcoxon-Mann-Whitney para amostras independentes. Assuma um nível de significância de 5%. Assuma que a distribuição dos dados não é normal.(Basta adaptar o nome das colunas da tabela linha 83 do script da aula)



4 / 10 respostas corretas



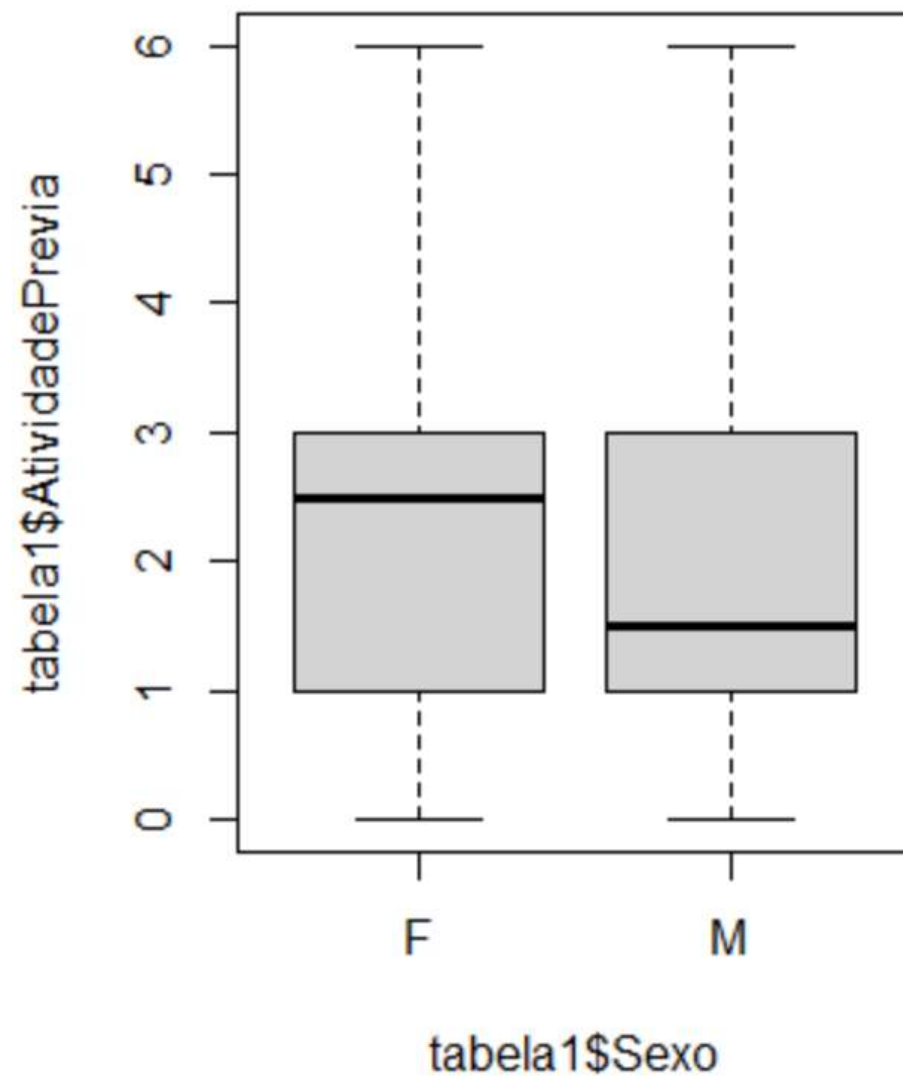
Feedback lista 7

```
> # Pergunta 6  
> boxplot(tabela1$AtividadePrevia~tabela1$Sexo)  
> wilcox.test(tabela1$AtividadePrevia~tabela1$Sexo)
```

wilcoxon rank sum test with continuity correction

```
data:  tabela1$AtividadePrevia by tabela1$Sexo  
W = 1488, p-value = 0.09546  
alternative hypothesis: true location shift is not equal to 0
```

Feedback lista 7



Feedback lista 7

```
> # Pergunta 6  
> boxplot(tabela1$AtividadePrevia~tabela1$Sexo)  
> wilcox.test(tabela1$AtividadePrevia~tabela1$Sexo)
```

wilcoxon rank sum test with continuity correction

```
data: tabela1$AtividadePrevia by tabela1$Sexo  
W = 1488, p-value = 0.09546  
alternative hypothesis: true location shift is not equal to 0
```


$0.09 > 0.05$

Feedback lista 7

 Selecione as respostas corretas:

Pergunta 6. Existe evidências de que há diferença na média do número de anos de prática de atividade física (Coluna) entre homens e mulheres (Coluna Sexo)? Para responder essa questão, aplique o teste de Wilcoxon-Mann-Whitney para amostras independentes. Assuma um nível de significância de 5%. Assuma que a distribuição dos dados não é normal.(Basta adaptar o nome das colunas da tabela linha 83 do script da aula)

1  pontos

- ☐ Existe (sim) evidências de que há diferença.
- ☒ Não existe evidencias de que há diferença . 
- ☐ Não é possível responder essa pergunta com os dados apresentados.
- ☐ Não consegui responder essa perguntar porque não entendi como aplicar o teste.
- ☐ Não consegui responder essa perguntar porque não tenho computador disponível para executar o teste.
- ☐ Não sei/Não quero responder essa pergunta.

 [Adicionar feedback da resposta](#)

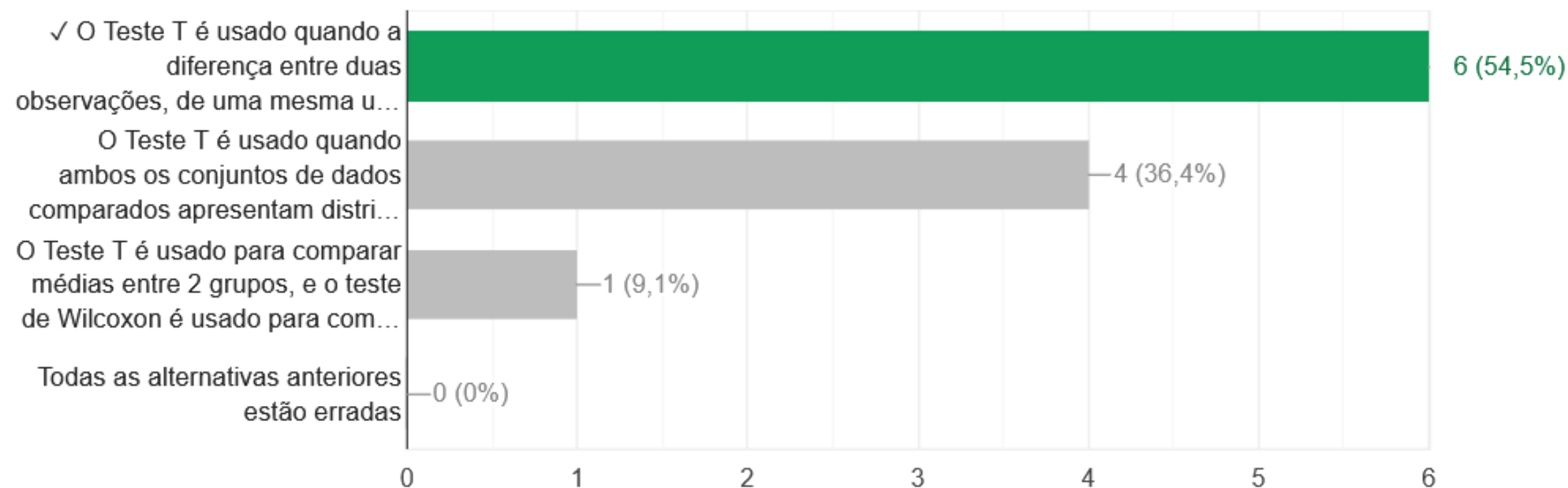
Concluido

Feedback lista 8


Pergunta 2. De forma geral, qual é principal diferença entre o Teste T (teste paramétrico) e um teste de Wilcoxon (teste não-paramétrico) - para amostras pareadas?




6 / 11 respostas corretas



Feedback lista 8

 Selecione as respostas corretas:

Pergunta 2. De forma geral, qual é principal diferença entre o Teste T (teste paramétrico) e um teste de Wilcoxon (teste não-paramétrico) - para amostras pareadas?

1  pontos



O Teste T é usado quando a diferença entre duas observações, de uma mesma unidade amostral, apresenta distribuição normal. Ao passo que o teste de Wilcoxon é usado essa diferença não apresenta distribuição normal.



O Teste T é usado quando ambos os conjuntos de dados comparados apresentam distribuição normal, e o teste de Mann-Whitney é usado quando qualquer um dos conjuntos de dados comparados apresentam distribuição diferente da normal.



O Teste T é usado para comparar médias entre 2 grupos, e o teste de Wilcoxon é usado para comparar médias entre mais de 2 grupos



Todas as alternativas anteriores estão erradas



Adicionar feedback da resposta

Concluído

ANOVA

Aula de hoje

Contínua

Categórica

Expectativa de vida ← UF de residência

Mesma ideia do teste t, com a diferença que a variável categórica pode mais de 2 grupos

Aula de hoje

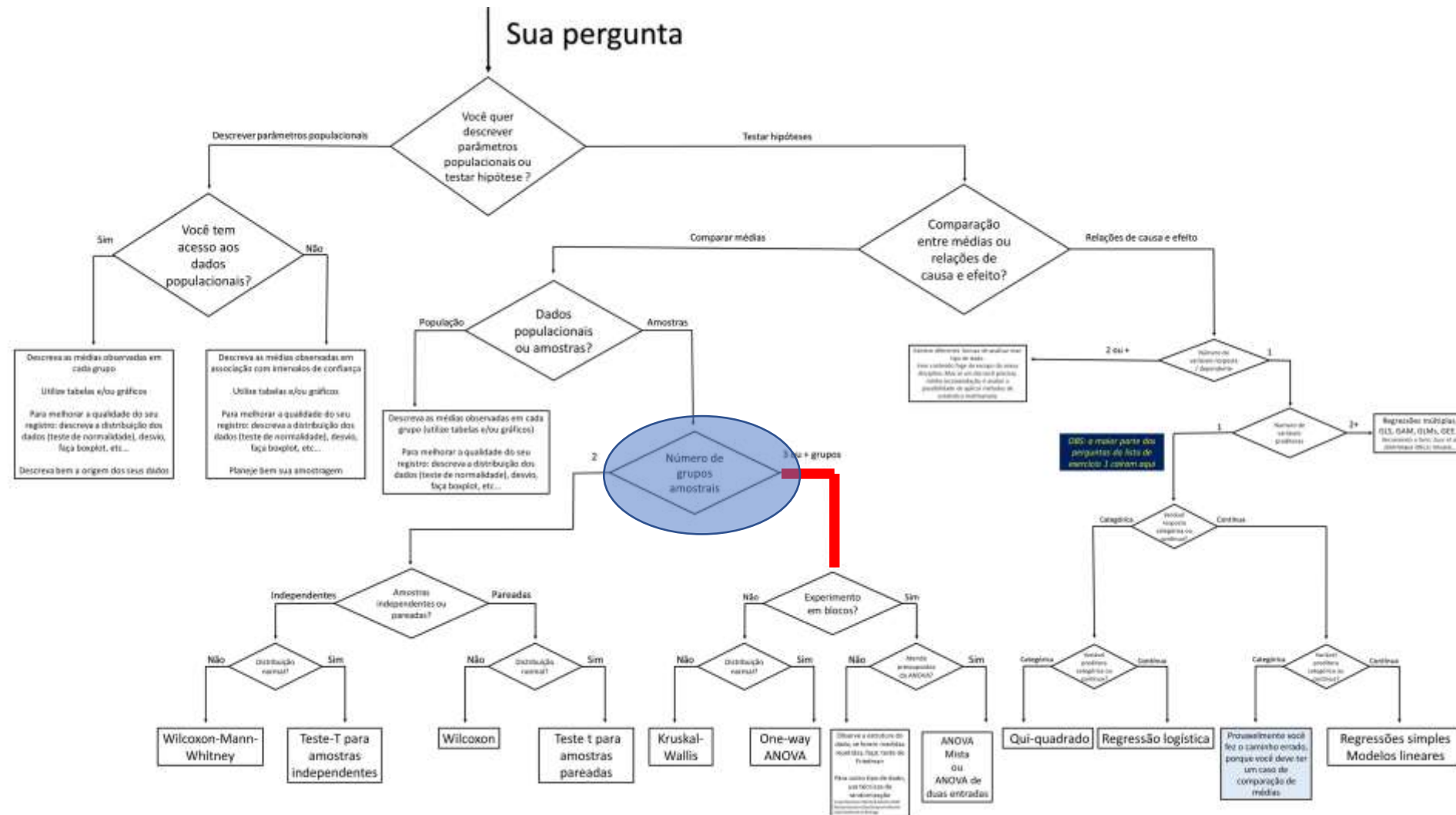
Variável resposta

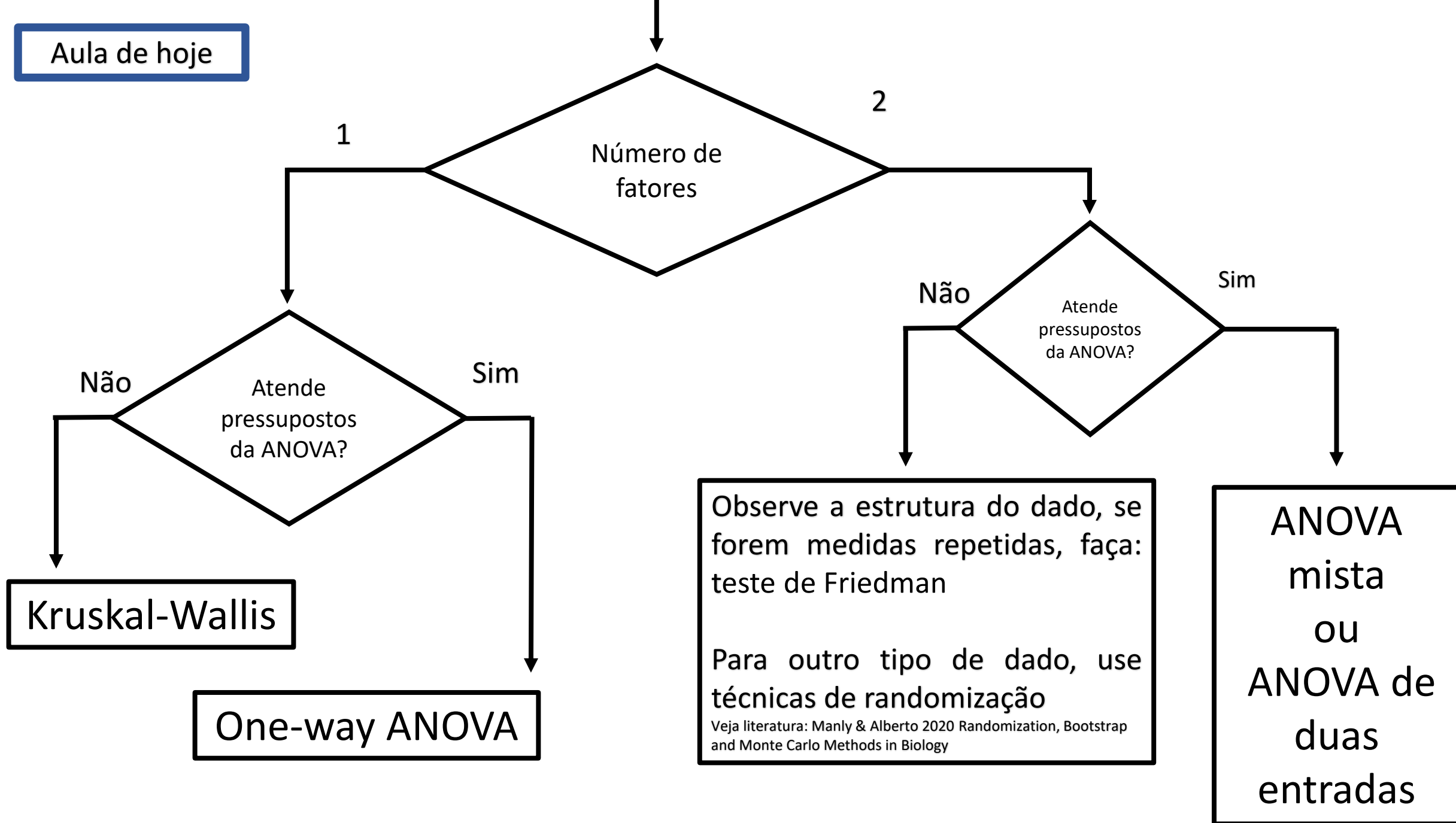
Variável preditora

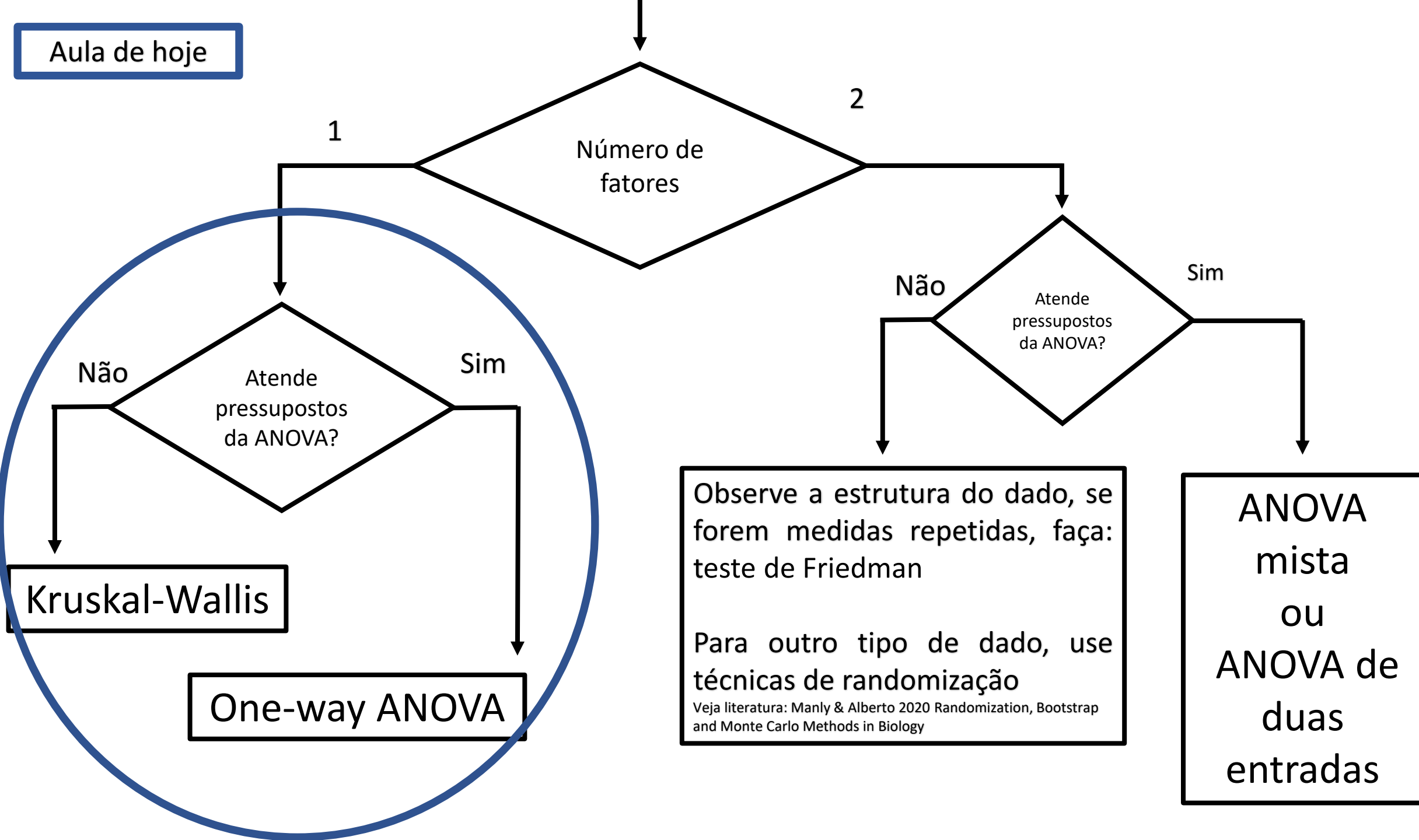
Expectativa de vida ← UF de residência

Mesma ideia do teste t, com a diferença que a variável categórica pode mais de 2 grupos

Tópicos dessa semana







1 variável preditora

Variável resposta

Variável preditora

Expectativa de vida ← UF de residência

2 variáveis preditoras

Variável resposta

Variáveis preditoras

Expectativa de vida ← UF de residência + Sexo

ANOVA simples (uma entrada)

- 1 Variável preditora: Variáveis qualitativa que divide os dados em 3 grupos ou mais
- Variável resposta: 1 Variável quantitativas contínuas

ANOVA dupla (duas entradas)

- 2 Variáveis preditoras: Variáveis qualitativas que dividem os dados em grupos
- Variável resposta: 1 Variável quantitativa contínua

Exemplos de adaptações de perguntas da Lista um que podem ser respondidas com essa metodologia

- 1) A incidências de Síndrome de Bornout é maior entre os profissionais de enfermagem, comparado a outros profissionais da área da saúde? (Alex, Rafaela Pereira)

OBS: próxima aula vamos ver uma forma alternativa de testar a mesma ideia

Exemplos de adaptações de perguntas: identificar variáveis

- Qual categoria de profissional da saúde, apresenta maior incidência de Síndrome de Bornout?
 - Incidência de Síndrome de Bornout:
 - variável resposta
 - quantitativas contínua (número de 0-100%)
 - Categoria de profissional da saúde:
 - variável preditora
 - qualitativa (Enfermeiro, Médico, Nutricionista ou Psicólogo)

Desenho experimental

- Para comparar valores de média entre 3 ou mais grupos, você deve usar alguma técnica para obter dados de cada um dos grupos que estão comparados. Exemplos:
 - a. Entrevistar aleatoriamente diferentes profissionais da saúde em diferentes hospitais, e a frequência de profissionais acometidos pela síndrome Burnout em cada hospital. Nesse estudo cada hospital seria uma unidade amostral.

Hipóteses

- H_0 : Não há diferenças entre as médias de todos os grupos comparados
- H_1 : Existe diferença de pelo menos uma das médias comparadas

Como fica uma tabela desse tipo de dado?

Nome	Profissão	Numero de horas de trabalho semanal
Lorena	Enfermeiro	30
Livia	Enfermeiro	27
Maria Luiza	Enfermeiro	40
Cecilia	Médico	40
Eloa	Médico	20
Giovanna	Médico	35
Maria Clara	Nutricionista	30
Maria Eduarda	Nutricionista	25
Mariana	Nutricionista	15
Lara	Psicólogo	40
Beatriz	Psicólogo	23
Antonella	Psicólogo	35

Identificação do
dado

Variável categórica

Variável numérica

Seja uma base desse tipo

Nome	Profissão	Numero de horas de trabalho semanal
Lorena	Enfermeiro	30
Livia	Enfermeiro	27
Maria Luiza	Enfermeiro	40
Cecilia	Médico	40
Eloa	Médico	20
Giovanna	Médico	35
Maria Clara	Nutricionista	30
Maria Eduarda	Nutricionista	25
Mariana	Nutricionista	15
Lara	Psicólogo	40
Beatriz	Psicólogo	23
Antonella	Psicólogo	35

Identificação do
dado

Variável categórica

Variável numérica

seja uma

esse tipo

o

Nome	Profissão	Numero de horas de trabalho semanal
Lorena	Enfermeiro	30
Livia		
Maria		
Cecilia		
Eloa		
Giovan		
Maria		
Maria		
Maria		
Lara		
Beatri		
Antonena	Psicólogo	35

Igual ao teste t, a diferença é que
a variável categórica tem mais de
2 grupos!!!

E se tiver mais de uma entrada?

Exemplo de 1 entrada

Nome	Profissão	Numero de horas de trabalho semanal
Lorena	Enfermeiro	30
Livia	Enfermeiro	27
Maria Luiza	Enfermeiro	40
Cecilia	Médico	40
Eloa	Médico	20
Giovanna	Médico	35
Maria Clara	Nutricionista	30
Maria Eduarda	Nutricionista	25
Mariana	Nutricionista	15
Lara	Psicólogo	40
Beatriz	Psicólogo	23
Antonella	Psicólogo	35

Exemplo de 1 entrada

Observação única

Nome	Profissão	Numero de horas trabalho semanal
Lorena	Enfermeiro	30
Livia	Enfermeiro	27
Maria Luiza	Enfermeiro	40
Cecilia	Médico	40
Eloa	Médico	20
Giovanna	Médico	35
Maria Clara	Nutricionista	30
Maria Eduarda	Nutricionista	25
Mariana	Nutricionista	15
Lara	Psicólogo	40
Beatriz	Psicólogo	23
Antonella	Psicólogo	35

Exemplo de 2 entradas

Nome	Profissão	Sexo	Dados da variável 1
Lorena	Enfermeiro	M	18.6
Livia	Enfermeiro	F	24.33
Maria Luiza	Enfermeiro	F	10.42
Cecilia	Médico	M	22.04
Eloa	Médico	F	21.2
Giovanna	Médico	M	28.3
Maria Clara	Nutricionista	F	17.95
Maria Eduarda	Nutricionista	F	17.17
Mariana	Nutricionista	M	17.57
Lara	Psicólogo	M	19.68
Beatriz	Psicólogo	M	18.6
Antonella	Psicólogo	F	24.33

Exemplo de trade

Variável categórica

Variável categórica

Dados

Nome	Profissão	Sexo	Dados da variável
Lorena	Enfermeiro	M	18.6
Livia	Enfermeiro	F	24.33
Maria Luiza	Enfermeiro	F	10.42
Cecilia	Médico	M	22.04
Eloa	Médico	F	21.2
Giovanna	Médico	M	28.3
Maria Clara	Nutricionista	F	17.95
Maria Eduarda	Nutricionista	F	17.17
Mariana	Nutricionista	M	17.57
Lara	Psicólogo	M	19.68
Beatriz	Psicólogo	M	18.6
Antonella	Psicólogo	F	24.33

Exemplo de 2 entradas

Nome	Profissão	Dados da variável 1	Dados da variável 2
Lorena	Enfermeiro	18.6	12.92
Livia	Enfermeiro	24.33	19.1
Maria Luiza	Enfermeiro	10.42	5.57
Cecilia	Médico	22.04	17.09
Eloa	Médico	21.2	16.97
Giovanna	Médico	28.3	23.12
Maria Clara	Nutricionista	17.95	13.06
Maria Eduarda	Nutricionista	17.17	11.63
Mariana	Nutricionista	17.57	12.08
Lara	Psicólogo	19.68	14.55
Beatriz	Psicólogo	18.6	12.92
Antonella	Psicólogo	24.33	19.1

Exemplo de 2 entradas

Dados 1

Dados 2

Nome	Profissão	Dados da variável 1	Dados da variável 2
Lorena	Enfermeiro	18.6	12.92
Livia	Enfermeiro	24.33	19.1
Maria Luiza	Enfermeiro	10.42	5.57
Cecilia	Médico	22.04	17.09
Eloa	Médico	21.2	16.97
Giovanna	Médico	28.3	23.12
Maria Clara	Nutricionista	17.95	13.06
Maria Eduarda	Nutricionista	17.17	11.63
Mariana	Nutricionista	17.57	12.08
Lara	Psicólogo	19.68	14.55
Beatriz	Psicólogo	18.6	12.92
Antonella	Psicólogo	24.33	19.1

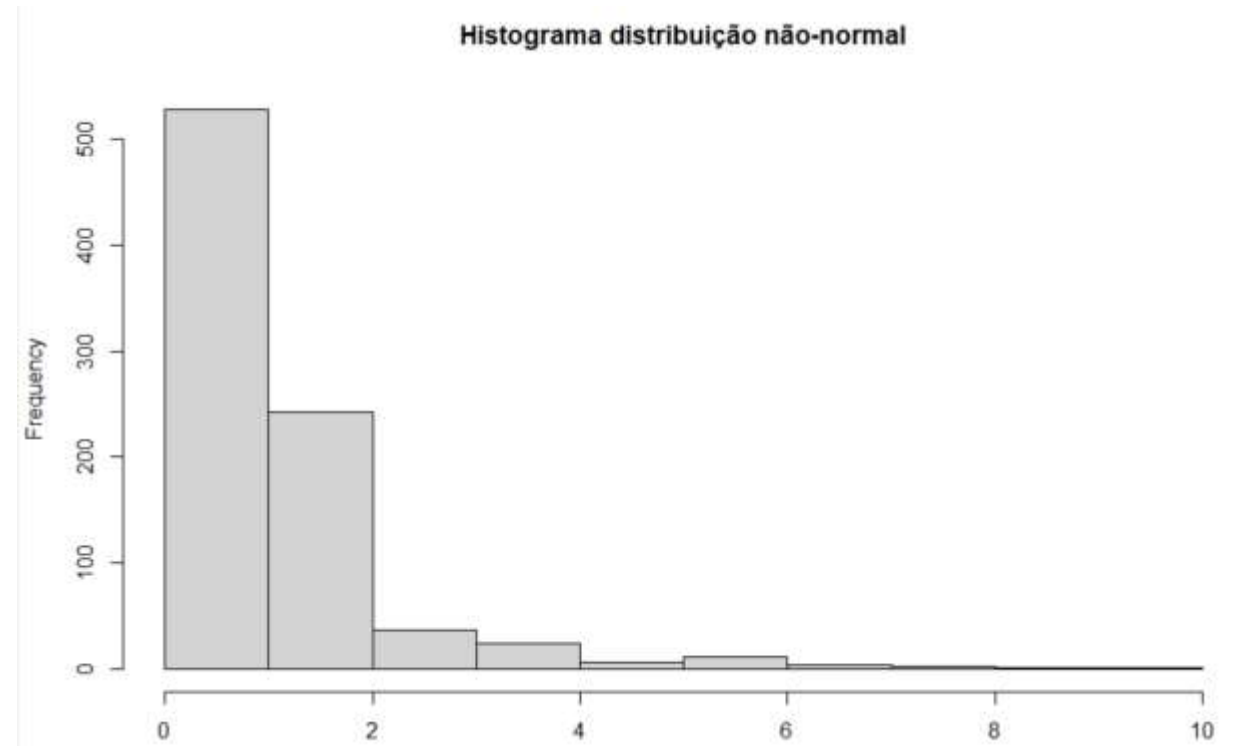
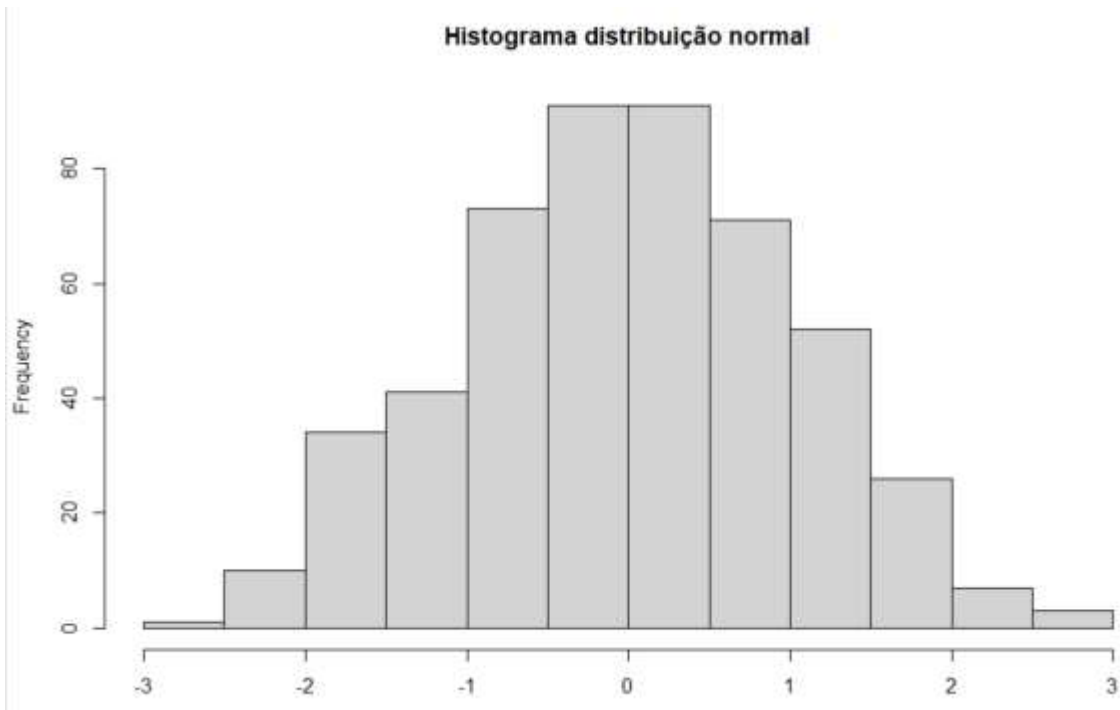
Portanto

- Se você tem 1 fator, seu teste vai avaliar apenas as diferenças na média entre as médias daquele fator
 - No exemplo, você vai saber apenas se tem ou não diferença no número de horas médio trabalhado por cada profissional
- Se você tem 2 fatores, pode ser que você esteja interessado no efeito do segundo fator ou não
 - Se você tem interesse no segundo fator: ANOVA de duas entradas
 - Se você não tem interesse no segundo fator: ANOVA mista (de medidas repetidas)

Qual diferença entre anova mista e anova duas entradas?

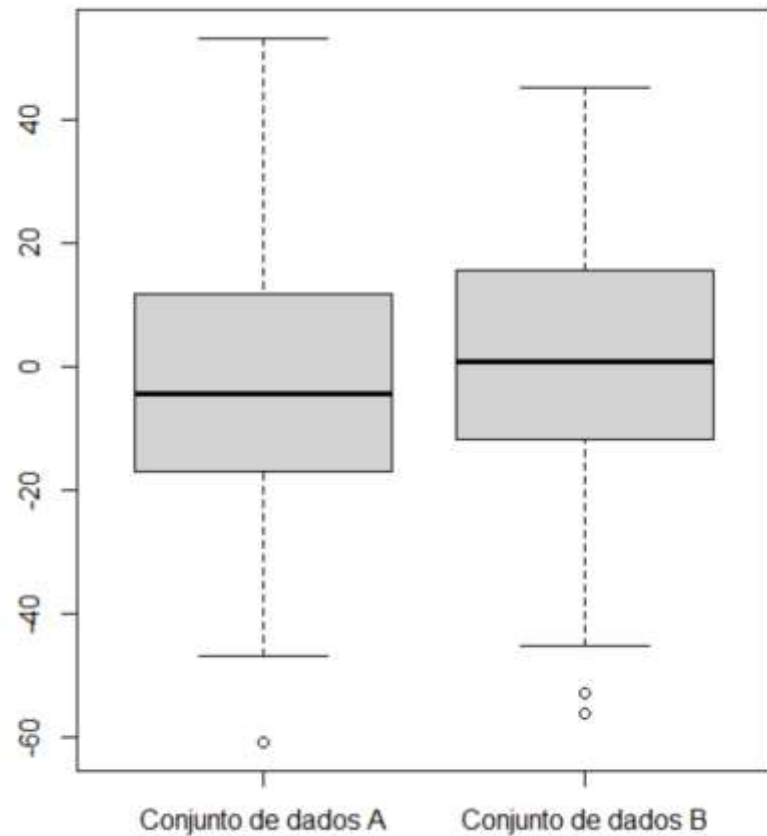
- A diferença é que na anova mista você não está interessado em saber o efeito das duas variáveis. O seu objetivo é apenas descontar o “peso” de uma mesma variável ter sido medida repetidas vezes para uma mesma unidade amostral

Pressupostos: Normalidade

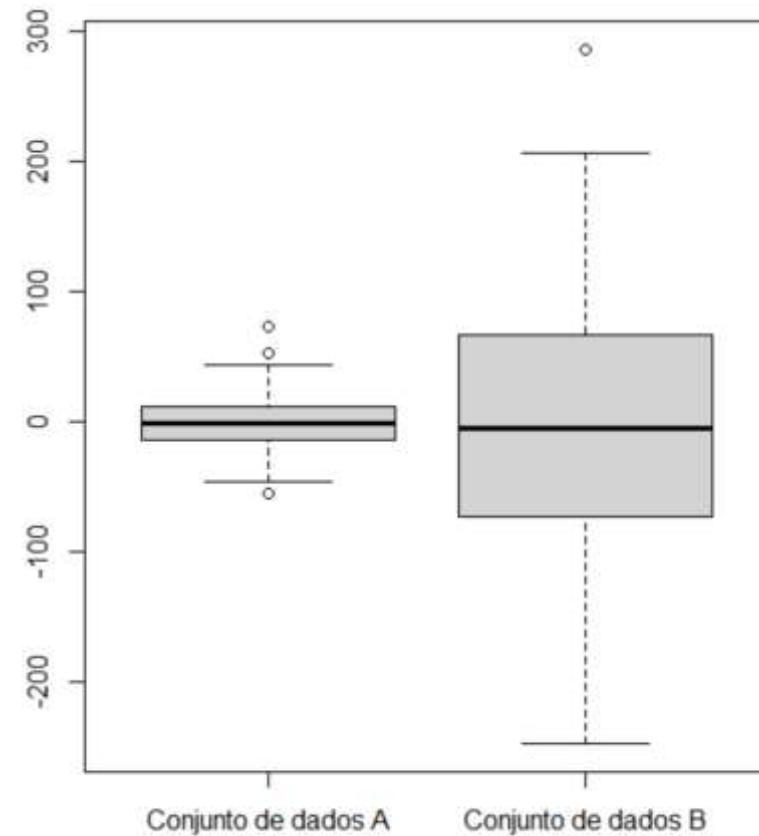


Pressupostos: Homogeneidade de variâncias

Padrão que sugere variâncias iguais



Padrão que sugere variâncias diferentes



Pressupostos: Outliers

- Confira se não tem nenhum numero muito discrepante na sua tabela.
- Pode ser que ao digitar num número, você sem querer escreveu errado. Exemplo: se a pessoa não digitar a virgula em 2,50 o dado vira 250.

Pós-teste!

- Ao aplicar a ANOVA você descobre se existe ou não diferença entre ao menos 1 dos grupos comparados, mas como saber qual grupo é diferente de qual?

Teste de Tukey

- Faz comparações entre todos os pares
- Teste de Tukey HSD (Teste de Tukey da Diferença Honestamente Significativa)

ANOVA vs. Kruskal-Wallis

- ANOVA:
 - todas as populações em confronto são normalmente distribuídas
 - todas as populações em confronto apresentam variâncias iguais
- Kruskal-Wallis:
 - não coloca nenhuma restrição sobre a comparação

Prática – 1 Existe diferença no número de plaquetas em relação ao estado?

- Tabela “Dados Fleury.xlsx”
- Variável preditora: UF
- Variável resposta: Plaquetas
- Tipo de dado: distribuição normal
- ANOVA de uma entrada

Interpretação dos resultados

Resultado: P maior que 0.05
Conclusão: H_0 é verdadeira

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tabela1\$UF	2	3.770e+09	1.885e+09	0.69	0.502
Residuals	578	1.579e+12	2.732e+09		

Prática – 2 Existe diferença no número de plaquetas em relação a categoria de idade?

- Tabela “Dados Fleury.xlsx”
- Variável preditora: Categoricalidade
- Variável resposta: Plaquetas
- Tipo de dado: distribuição normal
- ANOVA de uma entrada

Interpretação dos resultados

Resultado: P menor que 0.05
Conclusão: H_1 é verdadeira

```
> summary(modelo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tabela1\$CategoriaIdade	4	3.127e+10	7.817e+09	2.902	0.0214 *
Residuals	576	1.551e+12	2.693e+09		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> |
```

Pós teste

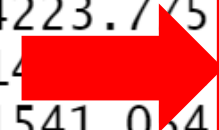
```
> TukeyHSD(modelo)
```

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

```
Fit: aov(formula = tabela1$Plaquetas ~ tabela1$CategoriaIdade)
```

```
$`tabela1$CategoriaIdade`
```

	diff	lwr	upr	p adj
De 30 a 44 anos-De 15 a 29 anos	767.7645	-19127.50	20663.027	0.9999719
De 45 a 59 anos-De 15 a 29 anos	-8945.9727	-30136.44	12244.494	0.7767881
De 60 a 74 anos-De 15 a 29 anos	-20321.0685	-45627.93	4985.797	0.1821229
Mais de 75 anos-De 15 a 29 anos	-26274.1935	-86992.81	34444.427	0.7605407
De 45 a 59 anos-De 30 a 44 anos	-9713.7372	-23651.25	4223.775	0.3146047
De 60 a 74 anos-De 30 a 44 anos	-21088.8330	-40726.95	-14	0.0282024
Mais de 75 anos-De 30 a 44 anos	-27041.9580	-85624.97	31541.054	0.7139549
De 60 a 74 anos-De 45 a 59 anos	-11375.0959	-32324.33	9574.134	0.5721227
Mais de 75 anos-De 45 a 59 anos	-17328.2209	-76363.67	41707.224	0.9296039
Mais de 75 anos-De 60 a 74 anos	-5953.1250	-66587.98	54681.727	0.9988549



Prática – 3 Existe diferença na testosterona em relação a categoria de idade?

- Tabela “Dados Fleury.xlsx”
- Variável preditora: Categoricalidade
- Variável resposta: Testosterona
- Tipo de dado: distribuição não-normal
- Kruskal-Wallis

Interpretação dos resultados

Resultado: P menor que 0.05
Conclusão: H_1 é verdadeira

Kruskal-Wallis rank sum test

data: Testosterona by CategoriaIdade
Kruskal-Wallis chi-squared = 13.336, df = 4, p-value = 0.009746

Pós teste

```
" A CRISTO 10 A 5
  .y.      group1      group2      n1      n2 statistic      p      p.adj      p.adj.signif
* <chr>    <chr>      <chr>      <int> <int>      <dbl>      <dbl>      <dbl>      <chr>
1 Testoste~ De 15 a 29~ De 30 a 44~    20    93    -1.60    0.109    1          ns
2 Testoste~ De 15 a 29~ De 45 a 59~    20    75    -2.58    0.00979  0.0979    ns
3 Testoste~ De 15 a 29~ De 60 a 74~    20    36    -2.57    0.0101  0.101     ns
4 Testoste~ De 15 a 29~ Mais de 75~    20     5    -2.78    0.00537  0.0537    ns
5 Testoste~ De 30 a 44~ De 45 a 59~    93    75    -1.64    0.100    1          ns
6 Testoste~ De 30 a 44~ De 60 a 74~    93    36    -1.64    0.100    1          ns
7 Testoste~ De 30 a 44~ Mais de 75~    93     5    -2.17    0.0299  0.299     ns
8 Testoste~ De 45 a 59~ De 60 a 74~    75    36    -0.335   0.738    1          ns
9 Testoste~ De 45 a 59~ Mais de 75~    75     5    -1.61    0.108    1          ns
10 Testoste~ De 60 a 74~ Mais de 75~    36     5    -1.41    0.158    1          ns
> |
```

Prática – 4 Existe diferença na glicose em relação a categoria de idade e registro de COVID?

- Tabela “Dados Fleury.xlsx”
- Variável preditora: Categoricalidade + Covid19PCR
- Variável resposta: Glicose
- Tipo de dado: distribuição normal
- ANOVA de duas entradas

Interpretação dos resultados

Resultado: P menor que 0.05
Conclusão: H_1 é verdadeira

```
> summary(modelo)
```

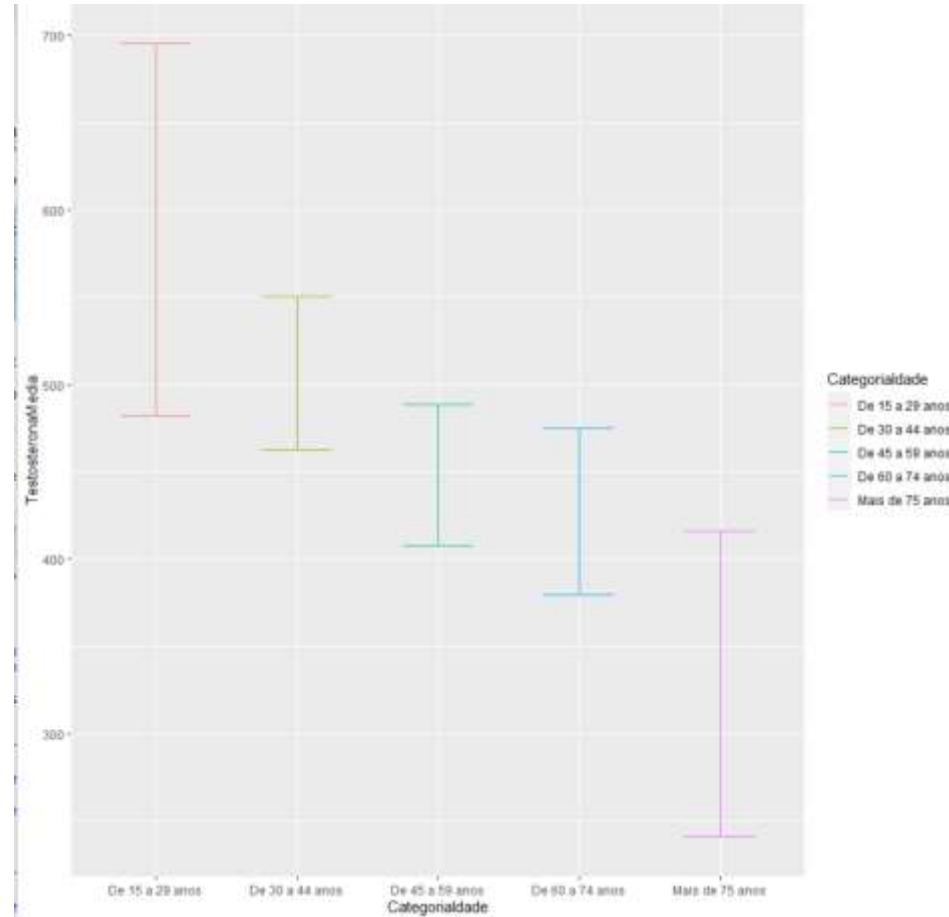
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tabela1\$CategoriaIdade	4	8630	2157.5	17.756	9.37e-14 ***
tabela1\$Covid19PCR	1	45	45.2	0.372	0.542
Residuals	575	69870	121.5		

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

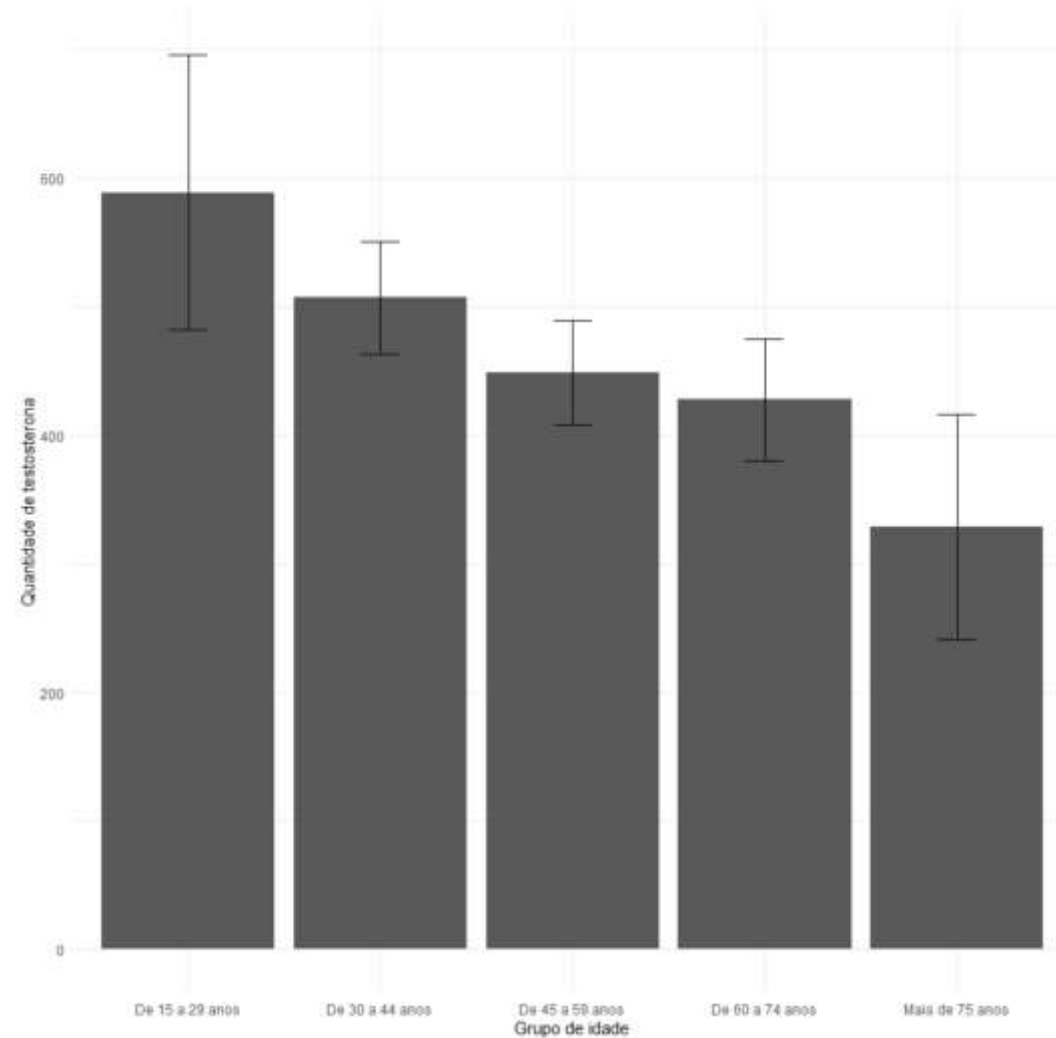
OBS: $9.37e-14 = 9.37 \times 10^{-14} = 0.0000000000000000937$

Gráfico de barras

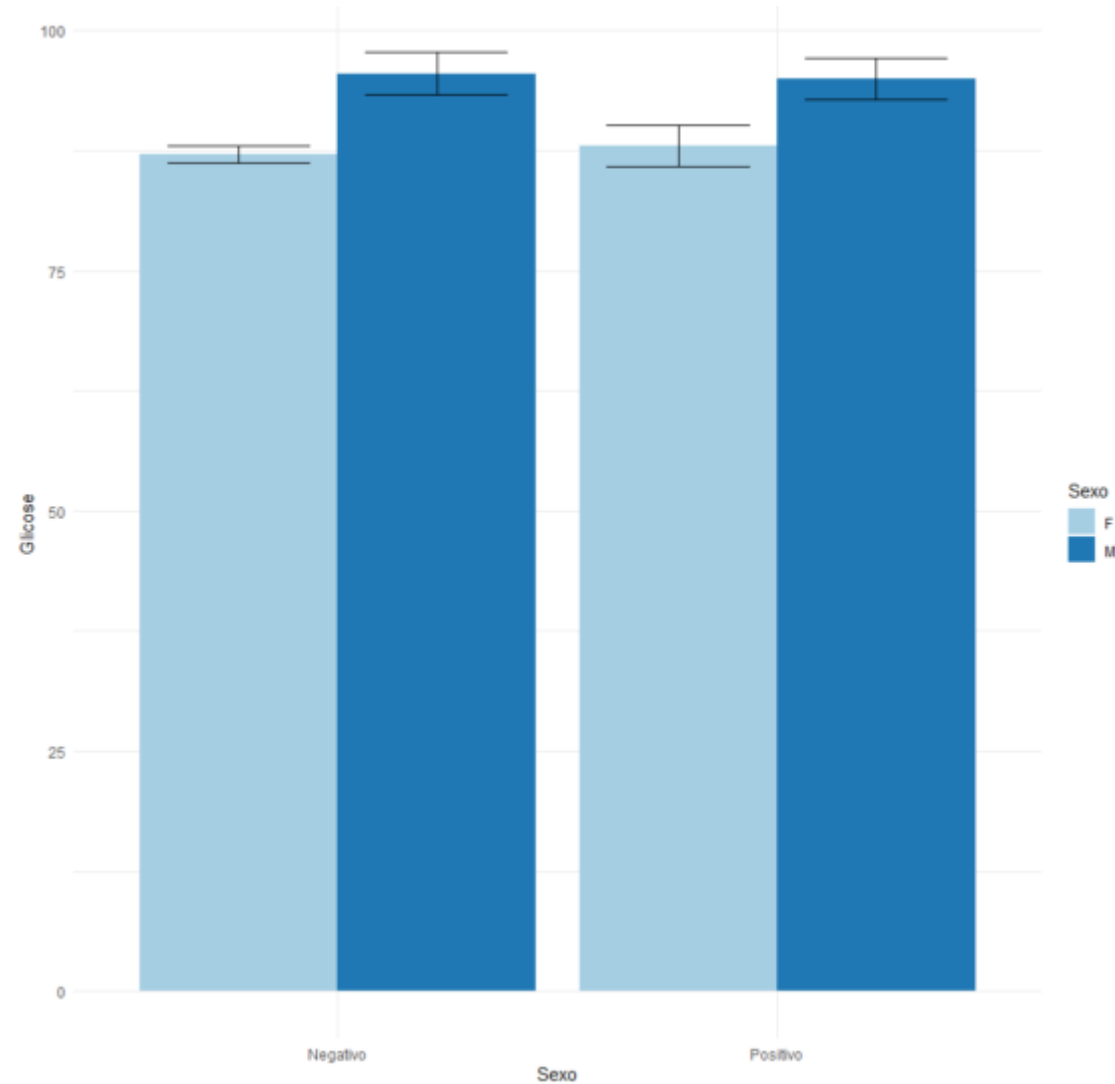
Gráficos presentes no material extra do R



Gráficos presentes no material extra do R



Gráficos presentes no material extra do R



Gráficos presentes no material extra do R

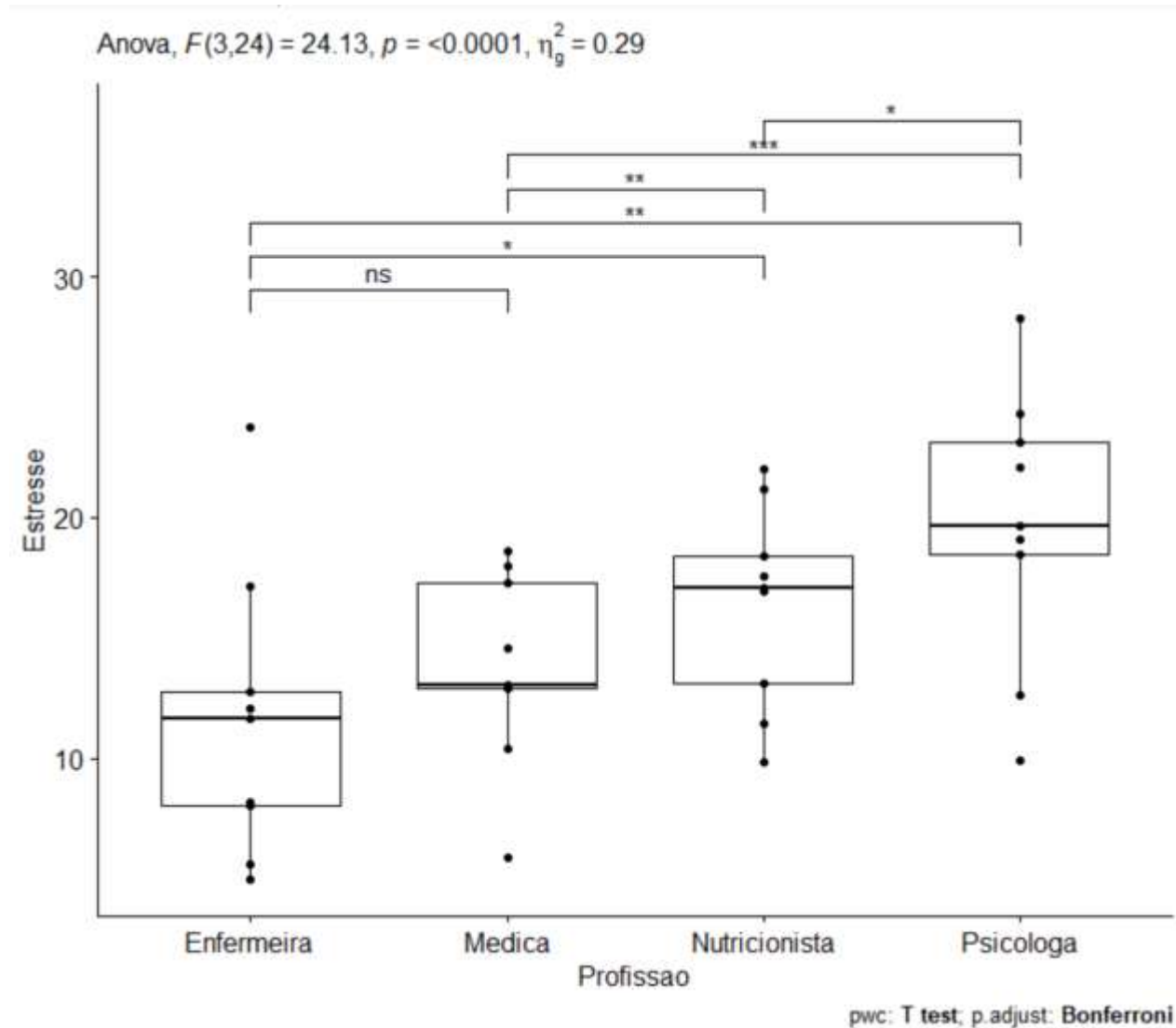


Gráfico de barras (Excel)

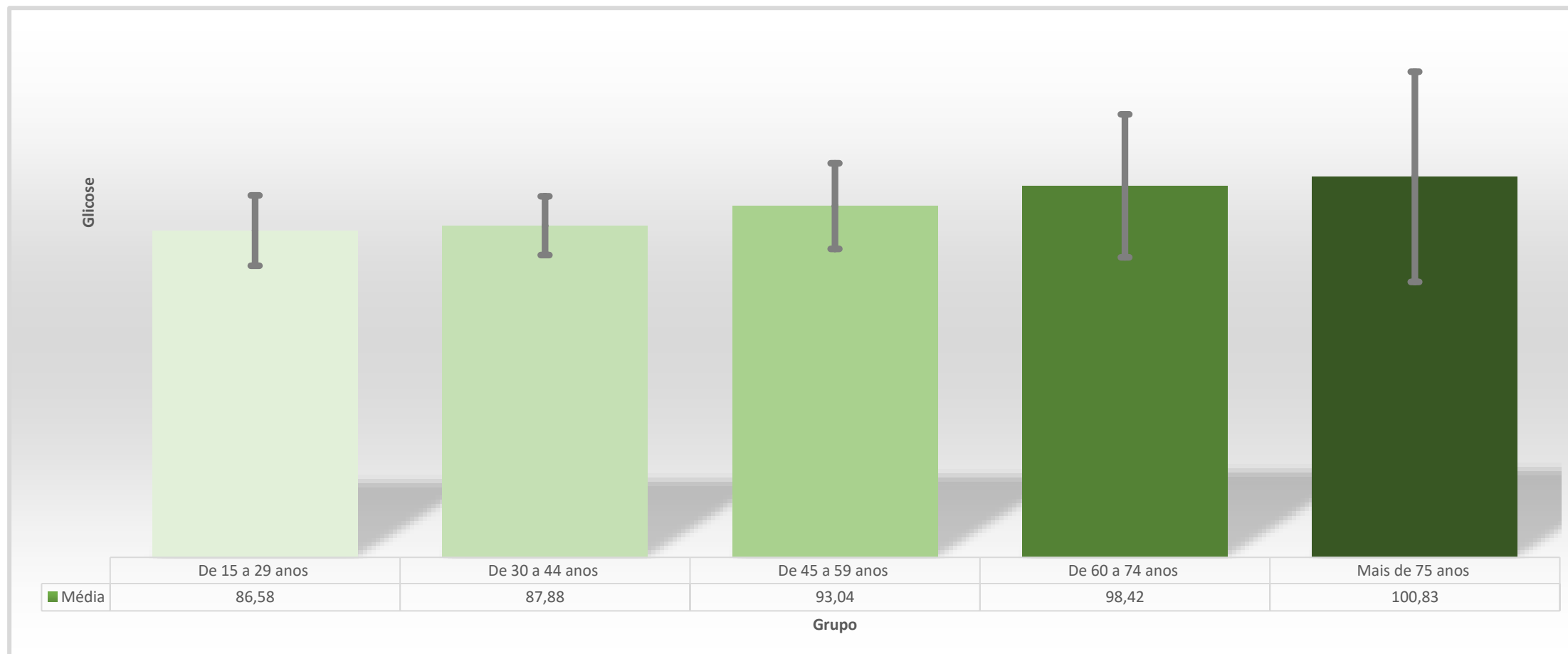
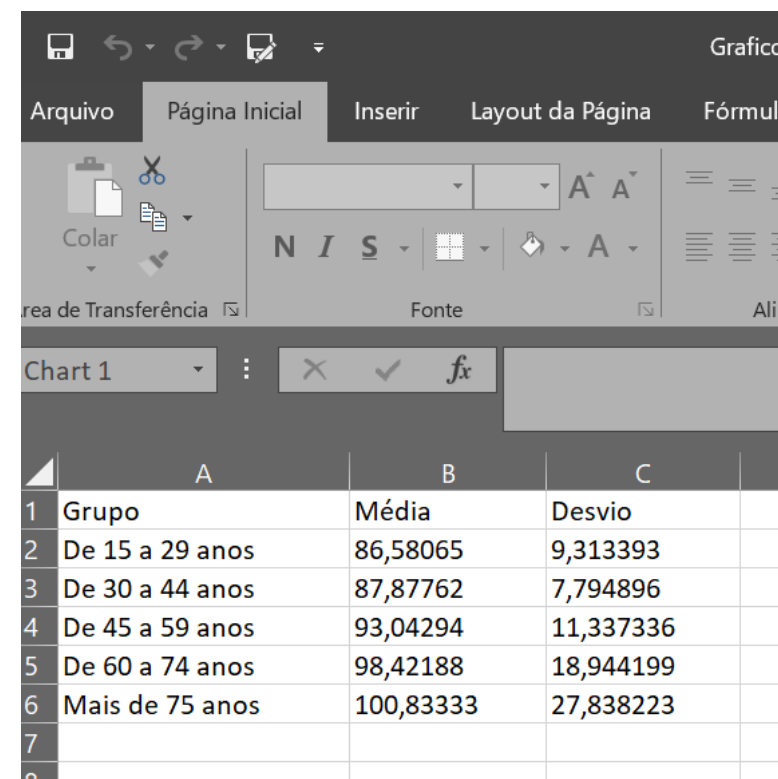


Gráfico de barras (Excel)

Utilize o R para conhecer a média de desvio

```
> with(tabela1, tapply(Glicose, CategoriaIdade,
  mean)) # Media
De 15 a 29 anos De 30 a 44 anos
      86.58065      87.87762
De 45 a 59 anos De 60 a 74 anos
      93.04294      98.42188
Mais de 75 anos
      100.83333
> with(tabela1, tapply(Glicose, CategoriaIdade,
  sd)) # Desvio
De 15 a 29 anos De 30 a 44 anos
      9.313393      7.794896
De 45 a 59 anos De 60 a 74 anos
      11.337336      18.944199
Mais de 75 anos
      27.838223
> |
```

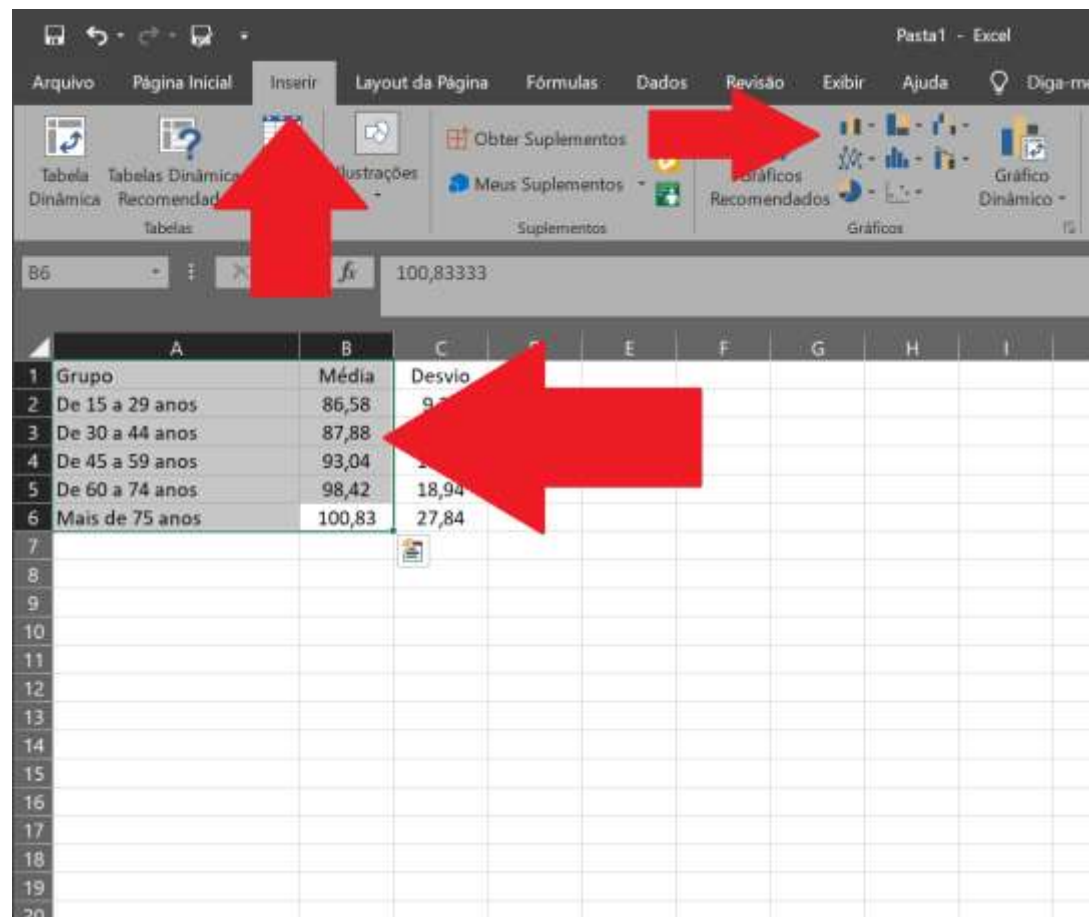
Transcreve esses dados para o Excel



The screenshot shows the Microsoft Excel interface with the 'Página Inicial' (Home) tab selected. The ribbon includes options for 'Arquivo', 'Página Inicial', 'Inserir', 'Layout da Página', and 'Fórmula'. The 'Fonte' (Font) section is visible, showing font face, size, bold, italic, underline, and color options. Below the ribbon, a chart titled 'Chart 1' is partially visible. The main data table is displayed in the worksheet area, with columns A, B, and C. The data is as follows:

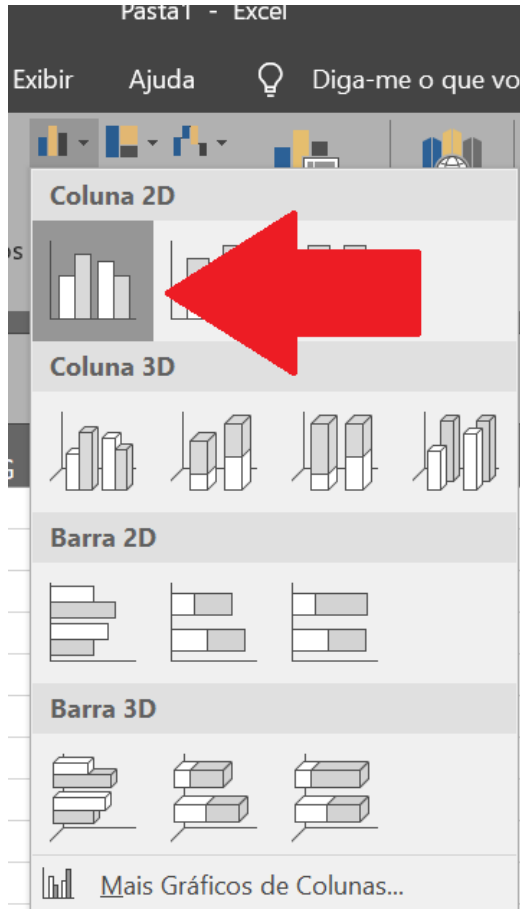
	A	B	C
1	Grupo	Média	Desvio
2	De 15 a 29 anos	86,58065	9,313393
3	De 30 a 44 anos	87,87762	7,794896
4	De 45 a 59 anos	93,04294	11,337336
5	De 60 a 74 anos	98,42188	18,944199
6	Mais de 75 anos	100,83333	27,838223
7			
8			

Gráfico de barras (Excel)



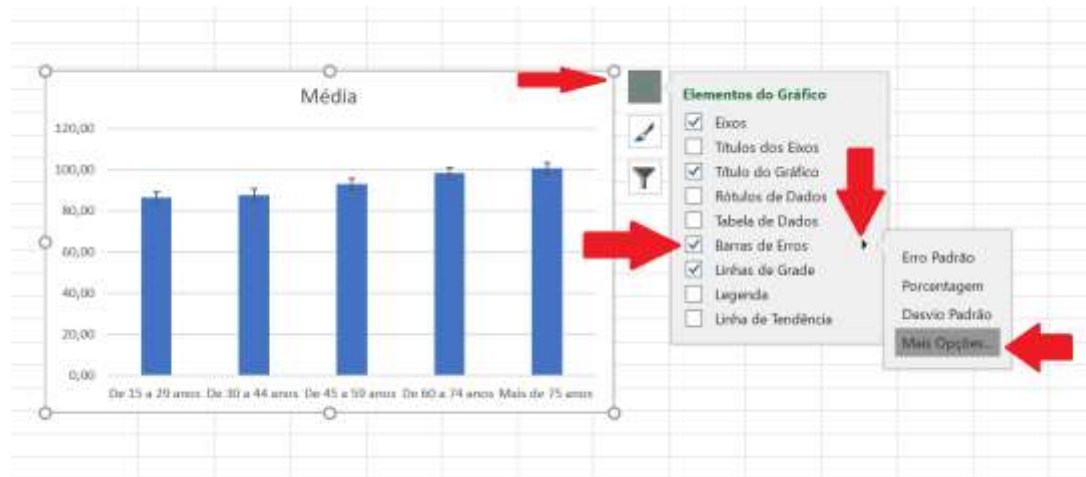
- Selecione os grupos e valores de média
- Clique em inserir
- Selecione a opção gráfico de barras

Gráfico de barras (Excel)



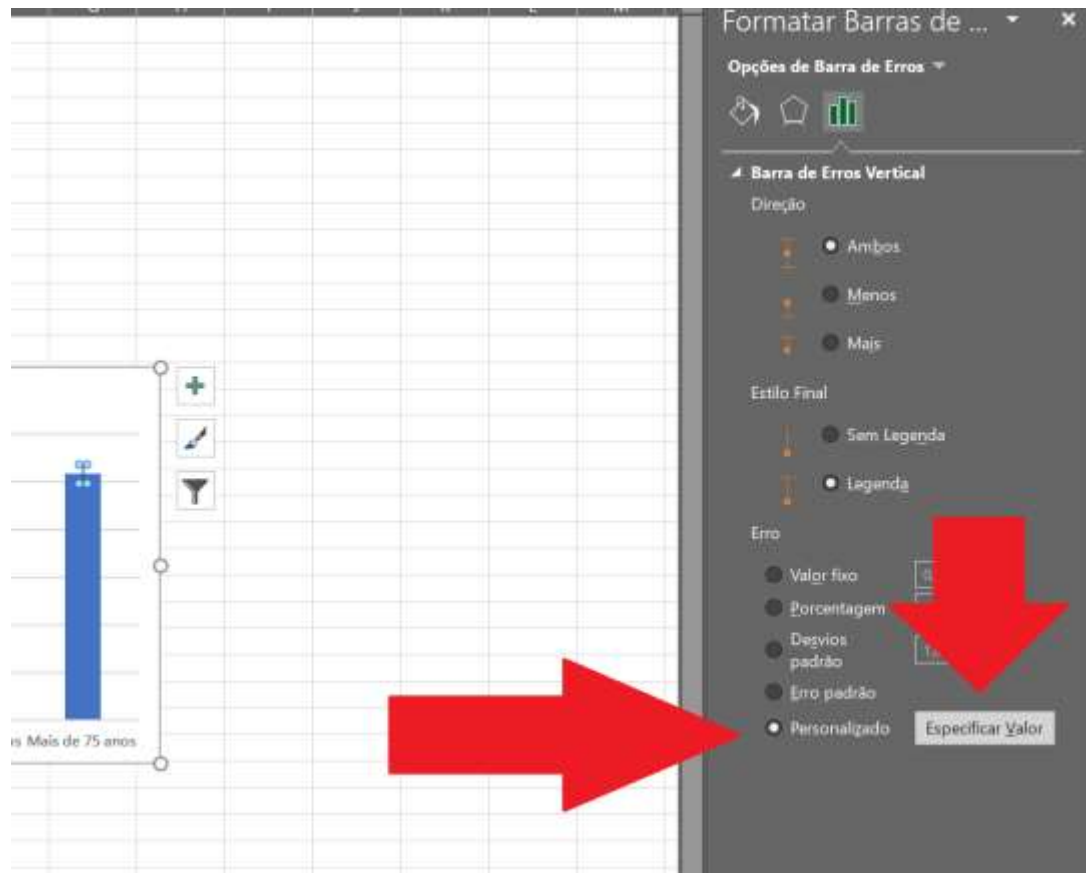
- Selecione os grupos e valores de média
- Clique em inserir
- Selecione a opção gráfico de barras

Gráfico de barras (Excel)



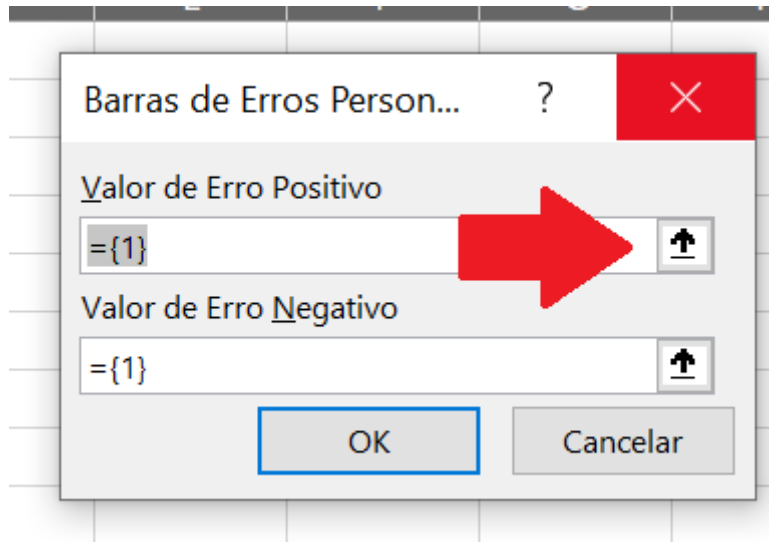
- Clique no botão “+”
- Marque a caixa Barra de erros
- Clique o símbolo >
- Clique em mais opções

Gráfico de barras (Excel)



- Clique em Personalização
- Clique em especificar valor

Gráfico de barras (Excel)



- Clique na seta apontando para cima
- Selecione a coluna de desvio
- Clique na seta apontando para baixo
- Repita o procedimento para o erro negativo

Gráfico de barras (Excel)

C	D	E	F	G
Desvio				
9,31				
7,79				
11,34				
18,94				
27,84				

Barras de Erros Personalizadas ? X

=Planilha1!\$C\$2:\$C\$6

- Clique na seta apontando para cima
- Selecione a coluna de desvio
- Clique na seta apontando para baixo
- Repita o pr

Gráfico de barras (Excel)

C	D	E	F	G
Desvio				
9,31				
7,79				
11,34				
18,94				
27,84				

Barras de Erros Personalizadas ? X

Valor de Erro Positivo
=Planilha1!\$C\$2:\$C\$6

Valor de Erro Negativo
={1}

OK Cancelar

- Clique na seta apontando para cima
- Selecione a coluna de desvio
- Clique na seta apontando para baixo
- Repita o procedimento para o erro negativo

Gráfico de barras (Excel)

C	D	E	F	G
Desvio				
9,31				
7,79				
11,34				
18,94				
27,84				

Barras de Erros Personalizadas ? X

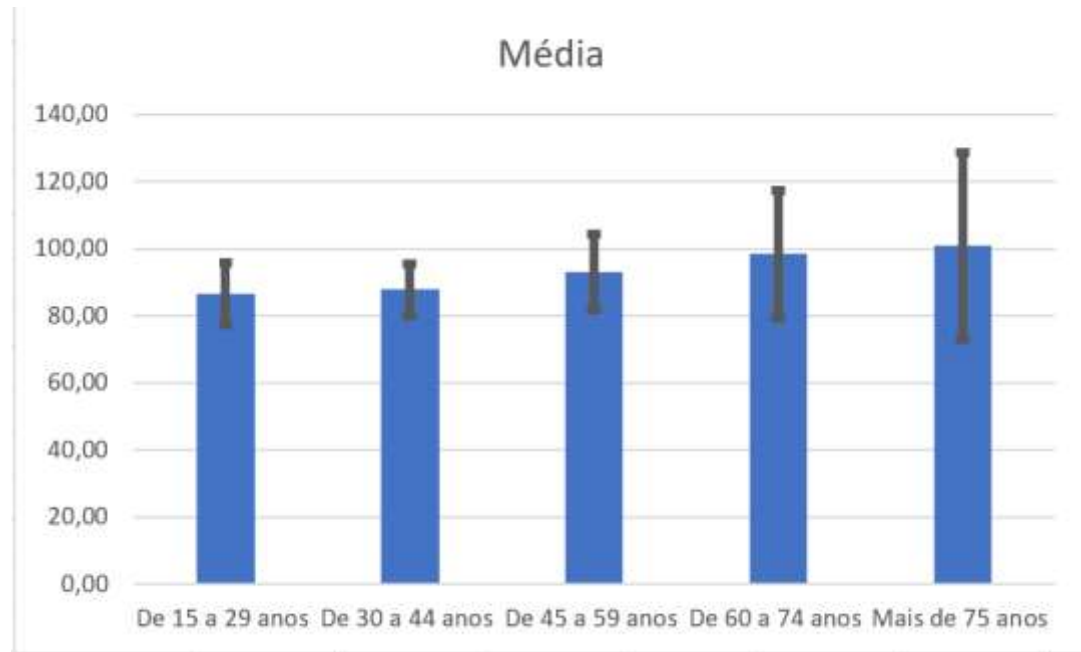
Valor de Erro Positivo
=Planilha1!\$C\$2:\$C\$6

Valor de Erro Negativo
=Planilha1!\$C\$2:\$C\$6

OK Cancelar

- Clique na seta apontando para cima
- Selecione a coluna de desvio
- Clique na seta apontando para baixo
- Repita o procedimento para o erro negativo

Gráfico de barras (Excel)



- No final deve estar algo mais ou menos assim

Gráfico de barras (Excel)

- Clique em Design para personalizar seu gráfico



Gráfico de barras (Excel)

