

# **ESTATÍSTICA PARA SAÚDE COLETIVA**

## **Aula 4**

Revisão última aula

# Bem vindo ao R



# Revisão última aula: objetos

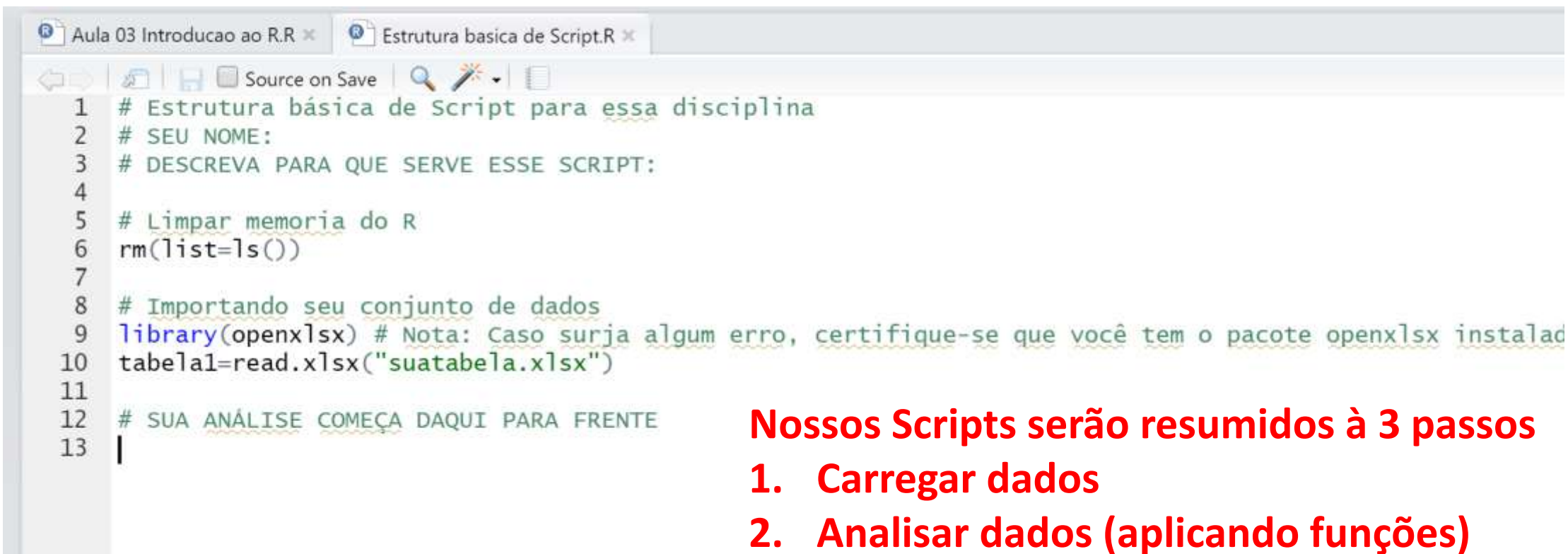


- Podem receber
  - Números
  - Textos
  - Tabelas
  - Funções

# Revisão última aula: Funções e Pacotes

- Funções são “as análises” que o R faz
- Como usar uma função?
  - função(parâmetros)
- Você pode escrever suas próprias funções, ou baixar funções escritas por outras pessoas por meio dos pacotes

# Revisão última aula: estrutura básica de script



```
Aula 03 Introducao ao R.R x Estrutura basica de Script.R x
Source on Save
1 # Estrutura básica de Script para essa disciplina
2 # SEU NOME:
3 # DESCREVA PARA QUE SERVE ESSE SCRIPT:
4
5 # Limpar memoria do R
6 rm(list=ls())
7
8 # Importando seu conjunto de dados
9 library(openxlsx) # Nota: Caso surja algum erro, certifique-se que você tem o pacote openxlsx instalado
10 tabela1=read.xlsx("suatabela.xlsx")
11
12 # SUA ANÁLISE COMEÇA DAQUI PARA FRENTE
13 |
```

**Nossos Scripts serão resumidos à 3 passos**

- 1. Carregar dados**
- 2. Analisar dados (aplicando funções)**
- 3. Interpretar resultados**

# Revisão última aula: estrutura básica de script

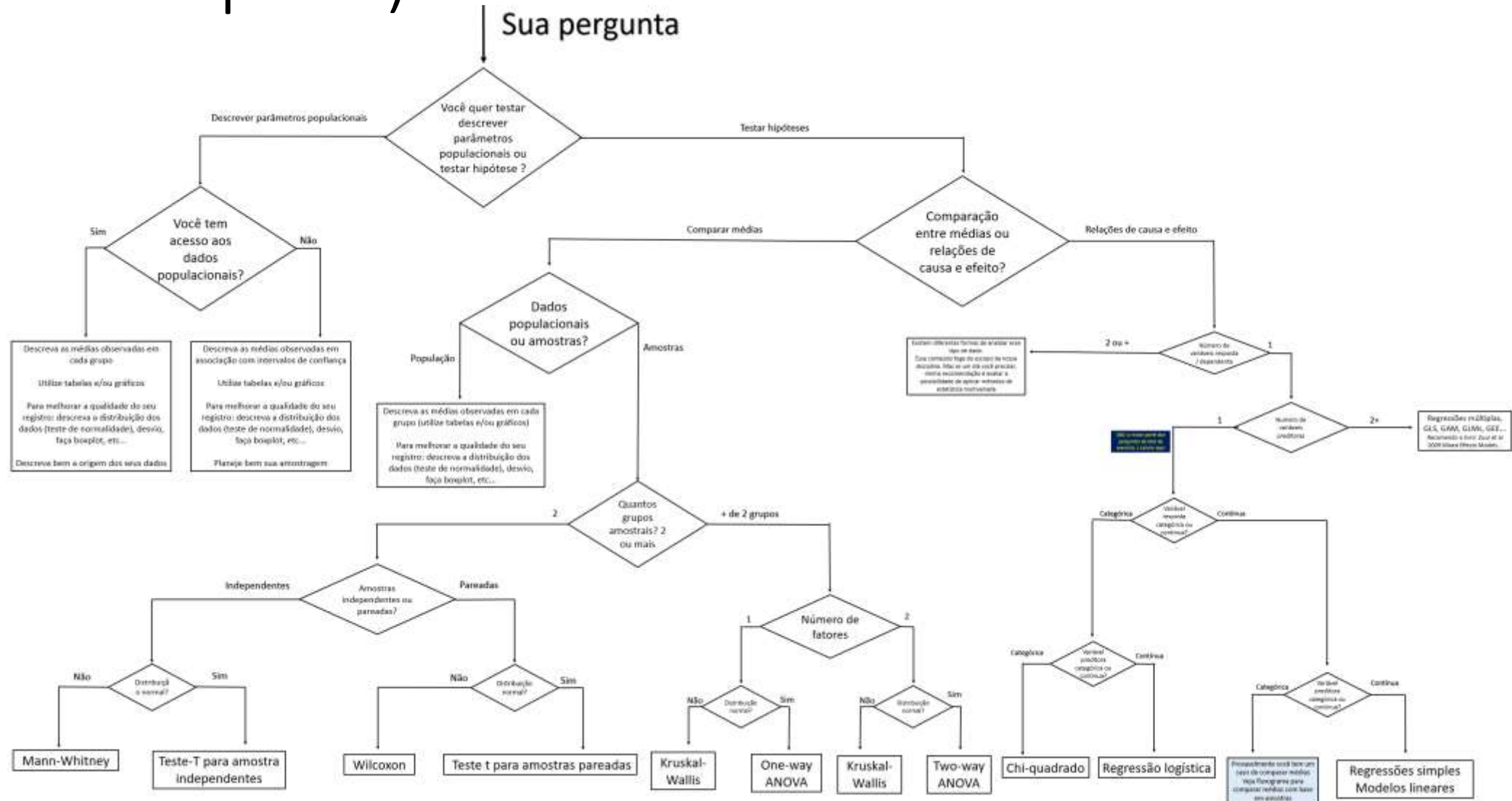
- Você não precisa ser um programador para usar o R

Aula de hoje

- Bioestatística é a nossa ferramenta para descrever padrões!



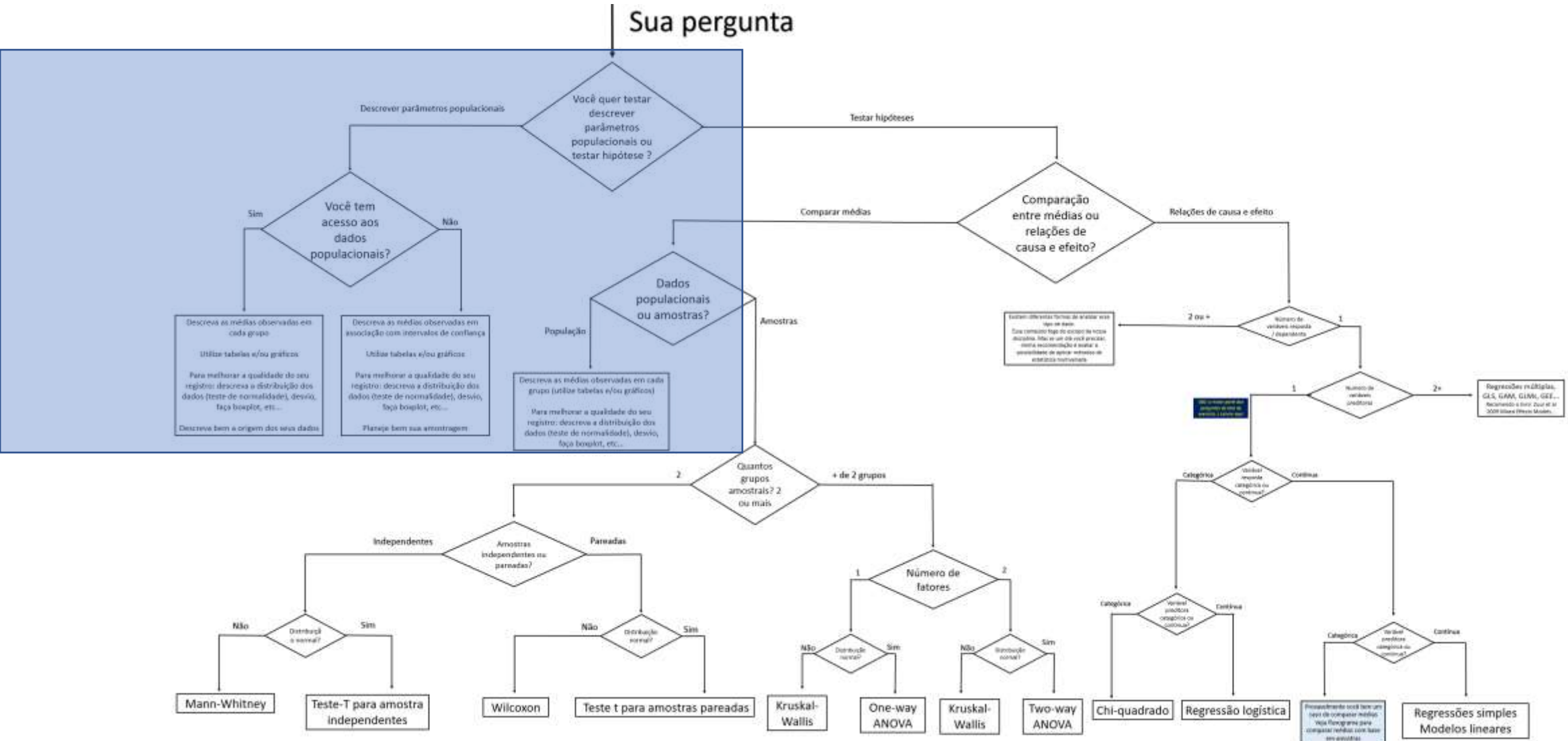
# Hoje começamos descrever padrões (resumo da disciplina)



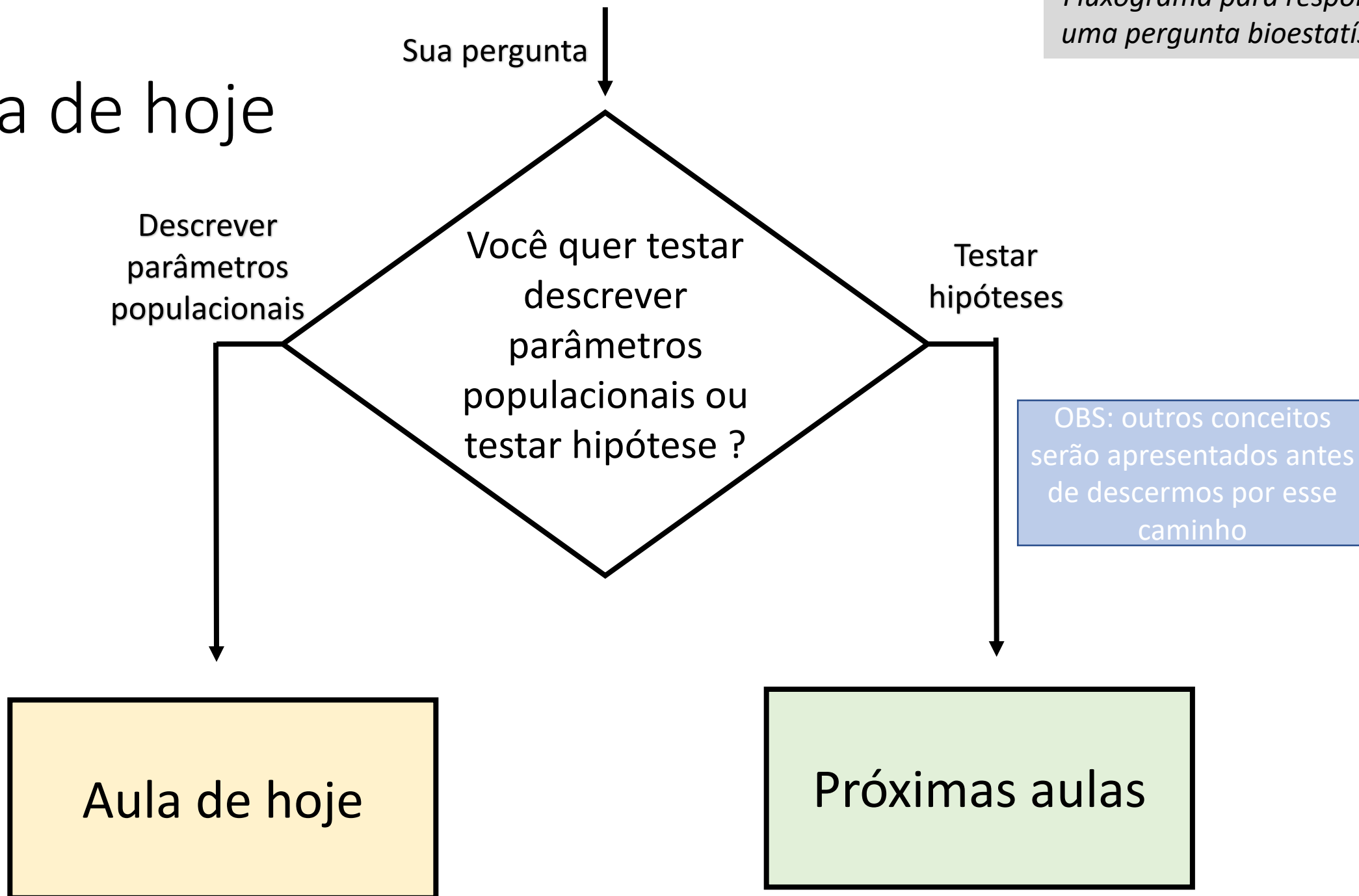
# Lembre

- Sua pergunta deve ser clara e objetiva
- Você nunca vai responder perguntas complexas sozinha(o)
- Em geral nossas perguntas vão ser relacionadas a descrição de parâmetros populacionais, comparações entre grupos ou relações de causa e efeito

# Tópicos da aula de hoje



# Aula de hoje



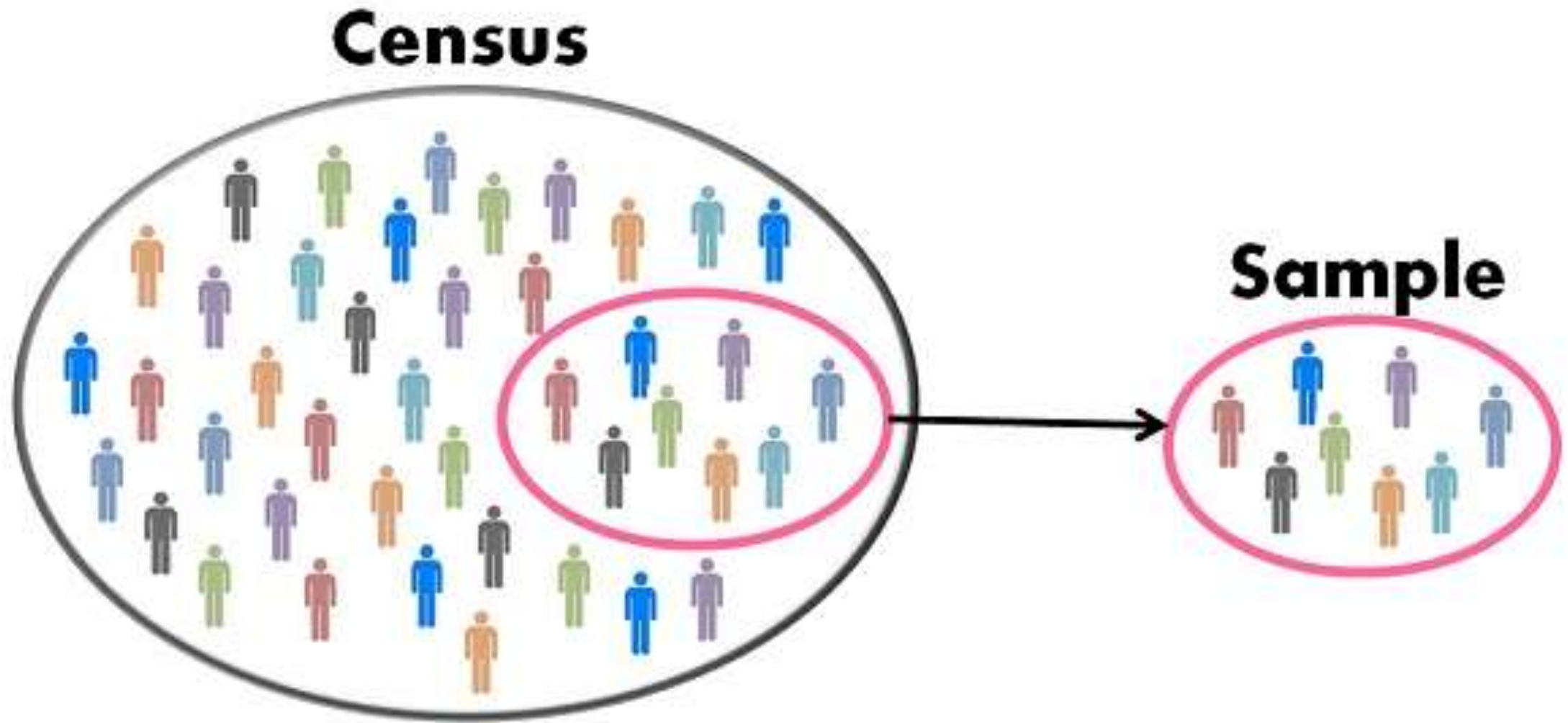
# Tema de hoje: Como resumir e descrever grandes volumes de dados.

- Tabelas
- Figuras
- Medidas de descrição
  - Média
  - Moda
  - Mediana; quartil; mínimo e máximo
  - Variância/Desvio

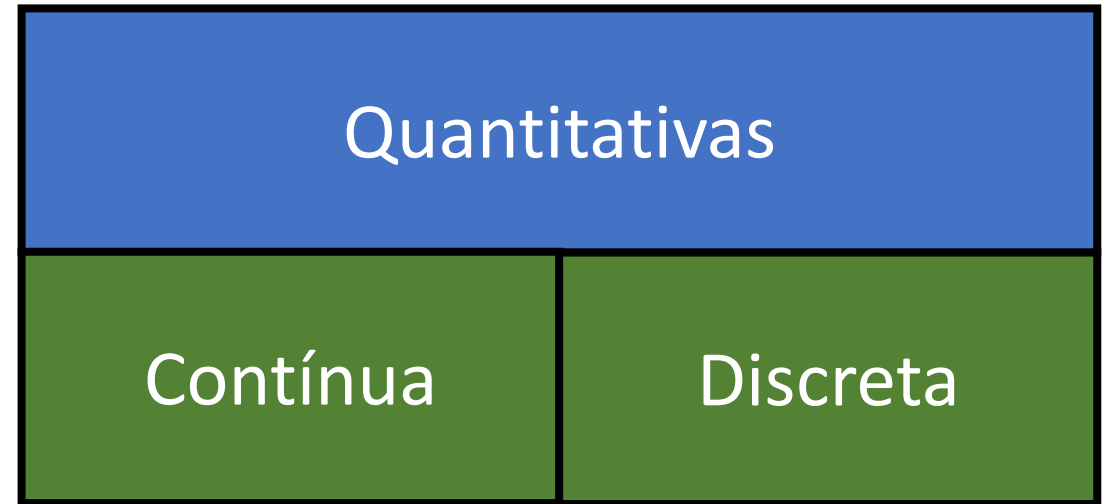
# Exemplos de perguntas que envolvem descrição de parâmetros populacionais

- Comparações a nível de população:
  - Qual estado brasileiro tem a maior taxa de mortalidade infantil?
- Descrições:
  - Qual é a altura média das mulheres da cidade de São Paulo?

# Revisão de conceitos importantes: Censo vs. Amostra



# Revisão de conceitos importantes: Tipos de variável





# Revisão de conceitos importantes: Variáveis qualitativas

- Dados mutualmente exclusivos
- Exemplos
  - Cidade de nascimento
  - Sexo
  - Toma ou não toma medicamento

# Revisão de conceitos importantes: Variáveis quantitativas

- Dados são expressos por números
- Exemplos
  - Idade
  - Altura
  - Número de filhos

# Bancos de dado

- Como você deve arquivar seus dados em uma tabela?
- Muitas vezes, transmitir para o próximo toda o banco de dados não é tão informativo quanto descrever esse banco de dados por meio de tabelas, figuras e descritores estatísticos

Identidade do bebê	Sexo	Sexo feminino	Peso (Kg)
Bebê 1	M	0	2,750
Bebê 2	F	1	2,950
Bebê 3	F	1	3,000
Bebê 4	F	1	4,000
Bebê 5	M	0	3,400
Bebê 6	M	0	3,200
Bebê 7	F	1	2,870
Bebê 8	F	1	3,500
Bebê 9	F	1	3,720
Bebê 10	M	0	3,940
...	...	...	...

Tabelas

# Regras para todos os tipos de tabela

- Devem ser introduzidas no texto próximo a primeira vez que são citadas.
- Toda tabela deve conter: título, corpo, cabeçalho e colunas indicadoras

# Recomendações

- Use apenas traços horizontais
- Cabeçalho deve ser separado do corpo do texto por um traço horizontal
- Se os dados não foram coletados por você, introduza a fonte no rodapé

# Exemplo de tabela

**TABELA 2.2**

**Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.**

<i><b>Grupo de doenças</b></i>	<i><b>Número</b></i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS.  
Nota: Suprimidos os casos com idade ou local de residência ignorados.

# Exemplo de tabela

## Cabeçalho

**TABELA 2.2**

**Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.**

<i>Grupo de doenças</i>	<i>Numero</i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS.  
Nota: Suprimidos os casos com idade ou local de residência ignorados.



# Exemplo de tabela

## Coluna indicadora

**TABELA 2.2**

**Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.**

<i><b>Grupo de doenças</b></i>	<i><b>Número</b></i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS.  
Nota: Suprimidos os casos com idade ou local de residência ignorados.

# Exemplo de tabela

# Corpo da tabela

**TABELA 2.2**

**Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.**

<i>Grupo de doenças</i>	<i>Número</i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS.  
Nota: Suprimidos os casos com idade ou local de residência ignorados.

# Exemplo de tabela

**TABELA 2.2**

**Número de internações hospitalares, de mulheres, pelo Sistema Único de Saúde (SUS). Brasil, 2005.**

<i>Grupo de doenças</i>	<i>Número</i>
Gravidez, parto e puerpério	2.640.438
Doenças do aparelho respiratório	736.012
Doenças do aparelho circulatório	612.415
Doenças do aparelho geniturinário	507.295
Doenças infecciosas e parasitárias	480.165
Doenças do aparelho digestivo	452.894
Transtornos mentais e comportamentais	105.354
Neoplasias	355.570
Causas externas	233.787
Demais causas	801.123
Total	6.925.053

Fonte: Ministério da Saúde/SE/Datasus — Sistema de Informações Hospitalares do SUS — SIH/SUS  
Nota: Suprimidos os casos com idade ou local de residência ignorados.

Fonte  
&  
Nota

Tabelas para dados qualitativos

# Tabelas para dados qualitativos

- Tabelas para dados qualitativos devem expressar dados de frequência absoluta e/ou relativa
  - *Frequência absoluta*: contagem do número de vezes que é feito o registro de uma amostra com a dada categoria
  - *Frequência relativa*: numero de vezes que o registro é feito / numero total de registros

# Tabelas para dados qualitativos

Tabela X. Opinião dos pacientes de um hospital sobre a qualidade do atendimento

Resposta	Frequência absoluta	Frequência relativa
Ruim	500	0,11 ou 11%
Regular	750	0,17 ou 17%
Bom	650	0,14 ou 14%
Muito bom	2500	0,56 ou 56%
Não sabe	100	0,02 ou 2%
Total	4500	1

# Tabelas de contingência

- Utilizadas quando você tem dados de elementos que foram classificados de acordo com duas variáveis qualitativas

# Tabelas de contingência

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	



# Tabelas de contingência      Número total de participantes da pesquisa

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	

# Tabelas de contingência

## Número de obesos

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	25%

# Tabelas de contingência

## Número de não-obesos

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	

# Tabelas de contingência

## Número de homens

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	

# Tabelas de contingência

## Número de mulheres

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	

# Tabelas de contingência

## Frequência

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	25%

# Tabelas de contingência

## Frequência

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	25%

Tabelas para dados quantitativos



# Tabelas para dados quantitativos discretos

- **Dados quantitativos discretos** também podem ser apresentados em tabelas de distribuição de frequência
  - Dados em ordem crescente
  - Contagem do numero de vezes que é feito registro de uma observação

# Tabelas de distribuição de frequência: para dados quantitativos discretos

Tabela X. Numero de faltas em uma equipe de 50 enfermeiros ao longo de 1 mês

Número de faltas	Frequências	Frequência relativa
0	32	0,64
1	8	0,16
2	4	0,08
3	3	0,06
4	2	0,04
5	1	0,02

# Tabelas para dados quantitativos contínuos

- **Dados quantitativos contínuos** também podem ser apresentados em tabelas de distribuição de frequência
  - Identifique o valor mínimo e máximo
  - Escolha um numero de classes
  - Contagem do numero de vezes que é feito registro dentro de cada classe
- Exemplo. Suponha que você tem dados de peso ao nascimento de 100 bebês
  - O peso mínimo é 1.590 e peso máximo 4.950
  - Você deseja separar os dados em 7 categorias
  - Amplitude = 3.360
  - Para 7 classes , você deveria separar as classes por 480.
  - Mas por conveniência, você pode fazer adaptações, veja a tabela a seguir

# Tabelas de distribuição de frequência: para dados quantitativos contínuos

Tabela X. Distribuição de frequência de para peso ao nascimento vivo (Kg).

Classe	Frequência absoluta	Frequência relativa
1.500  — 2.000	3	3%
2.000  — 2.500	16	16%
2.500  — 3.000	31	31%
3.000  — 3.500	34	34%
3.500  — 4.000	11	11%
4.000  — 4.500	4	4%
4.500  — 5.000	1	1%

# Tabelas descritivas para dados quantitativos contínuos

- Se você quer apresentar vários dados quantitativos, apresente os dados na ordem de:
  - Coleta
  - Importância

# Tabelas para dados quantitativos

Tabela X. Idade de mães e peso médio em 5 cidades brasileiras

Nome da cidade	Idade média das mães	Peso médio dos bebês
Cidade 1	25.2	2.750
Cidade 2	27.3	2.950
Cidade 3	26.7	3.000
Cidade 4	24.3	4.000
Cidade 5	29.5	3.400

# Tabelas para dados quantitativos

- **IMPORTANTE:** essa tabela está incompleta
  - Falta dados de medida de tendência central
  - Se você estiver trabalhando com amostras, também falta uma estimativa do erro
- Essas questões serão vistas no final dessa aula

# Gráficos



# Regras para todos os tipos de gráficos

- Devem ser introduzidas no texto próximo a primeira vez que são citadas.
- Todo deve conter: título e escala
- Escala: deve crescer da esquerda para direita e de baixo para cima

# Recomendações

- Legendas explicativas devem ser colocada preferencialmente a direita do gráfico

Gráficos para dados qualitativos

# Gráfico de barras

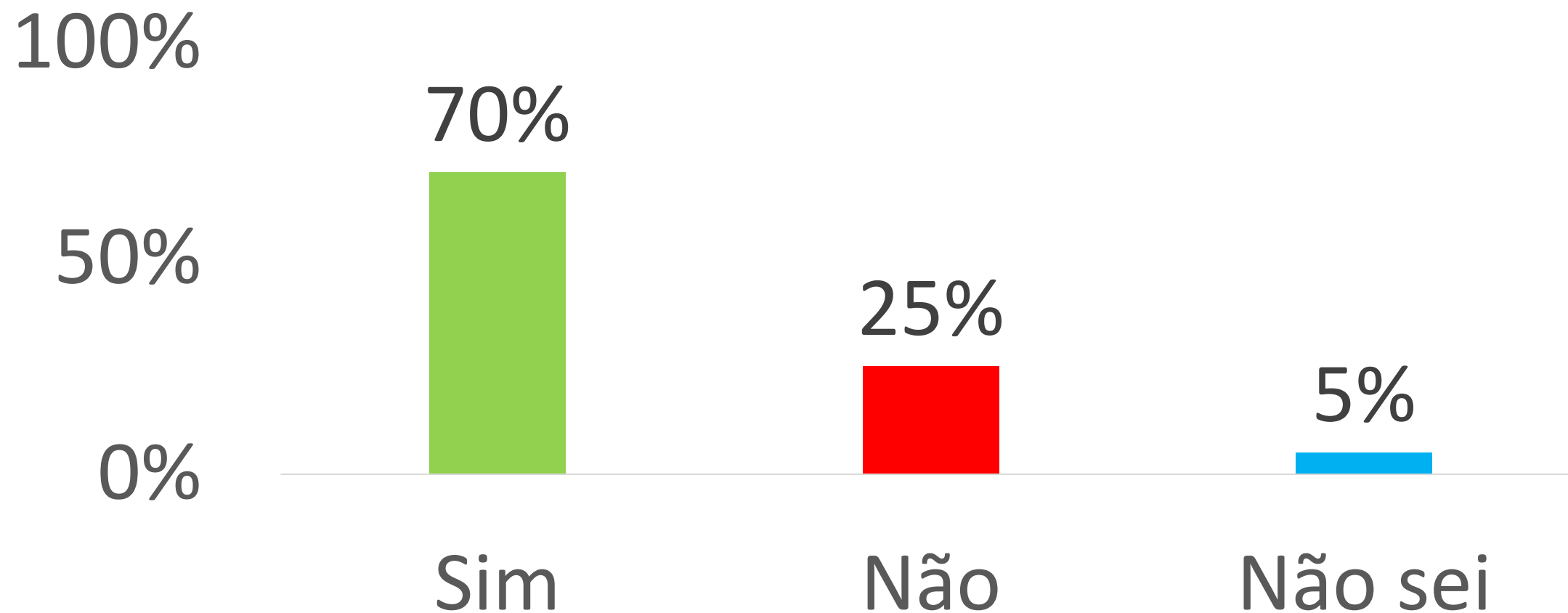
- Usado para facilitar a leitura de percentuais frequências em determinadas categorias

# Dados

Tabela X. Opinião sobre uso de mascara. Participantes foram questionados se acham que uso de mascara diminui o contagio do COVID.

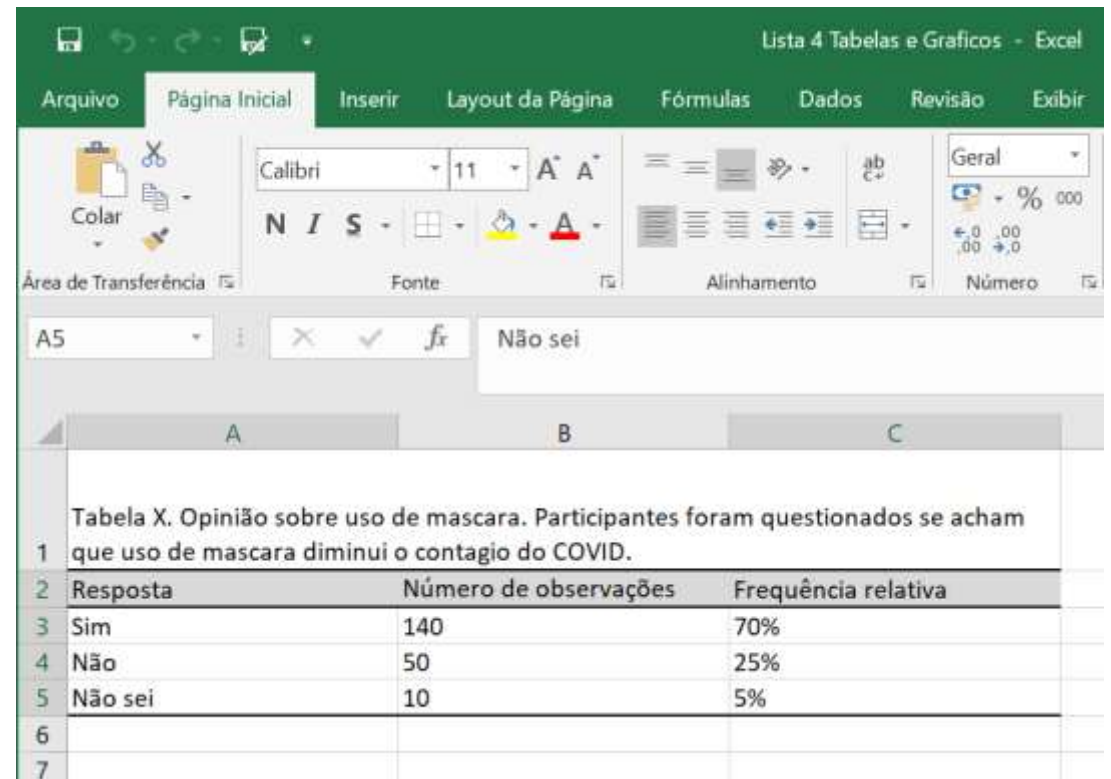
Resposta	Número de observações	Frequência relativa
Sim	140	70%
Não	50	25%
Não sei	10	5%
Total	200	100%

# Gráfico de barras



# Como fazer esse gráfico no Excel?

## 1. Digite sua tabela

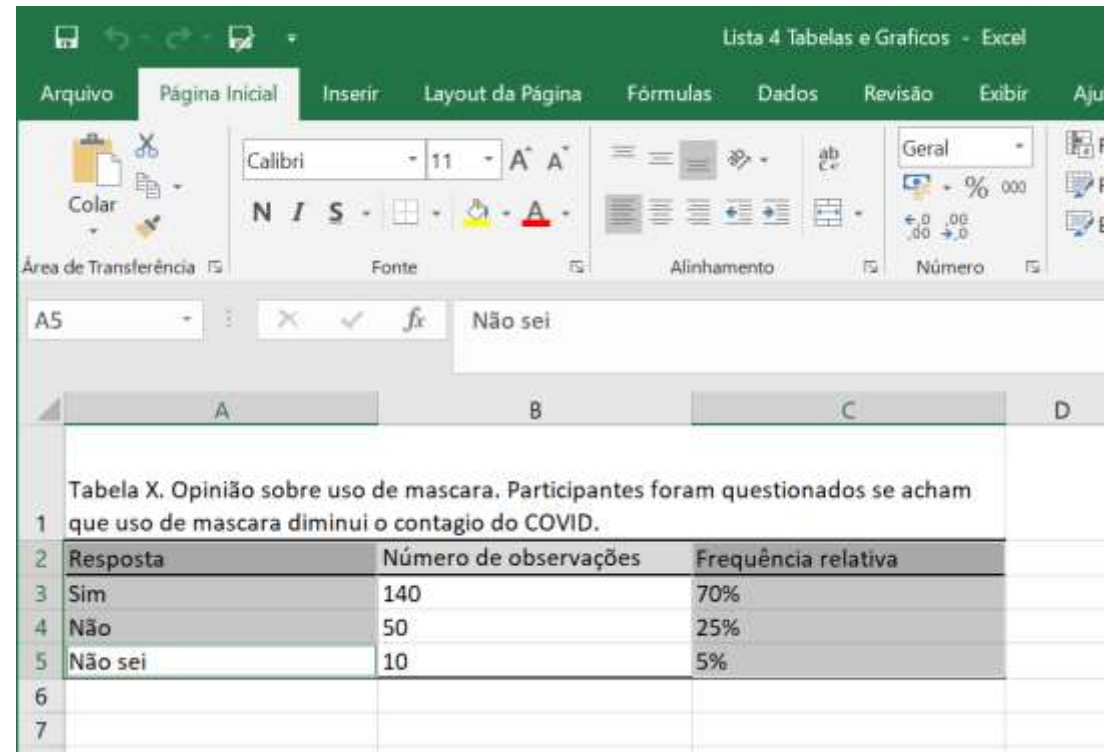


The screenshot shows the Microsoft Excel interface with the 'Página Inicial' (Home) tab selected. The ribbon includes options for 'Arquivo', 'Página Inicial', 'Inserir', 'Layout da Página', 'Fórmulas', 'Dados', 'Revisão', and 'Exibir'. The 'Fonte' (Font) group shows 'Calibri' font and size '11'. The 'Alinhamento' (Alignment) group shows text alignment options. The 'Número' (Number) group shows the 'Geral' (General) format. The active cell is A5, containing the text 'Não sei'. Below the ribbon, a table is displayed with the following data:

	A	B	C
1	Tabela X. Opinião sobre uso de mascara. Participantes foram questionados se acham que uso de mascara diminui o contágio do COVID.		
2	Resposta	Número de observações	Frequência relativa
3	Sim	140	70%
4	Não	50	25%
5	Não sei	10	5%
6			
7			

# Como fazer esse gráfico no Excel?

2. Selecione a coluna respostas +  
colunas frequência



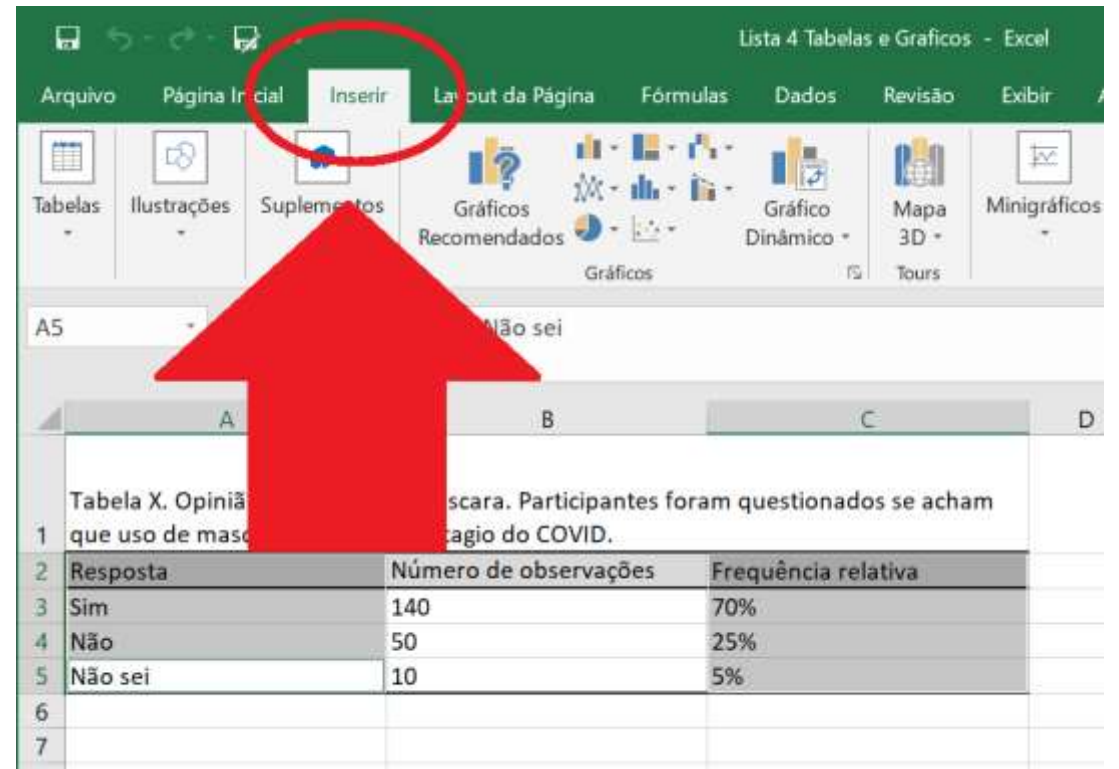
The screenshot shows the Microsoft Excel interface. The title bar reads 'Lista 4 Tabelas e Graficos - Excel'. The ribbon is set to 'Página Inicial' (Home). The 'Fonte' (Font) group shows 'Calibri' font and size '11'. The 'Alinhamento' (Alignment) group shows text aligned to the right. The 'Número' (Number) group shows the 'Geral' (General) format. The active cell is A5, containing the text 'Não sei'. Below the ribbon, a table is displayed with the following data:

	A	B	C	D
1	Tabela X. Opinião sobre uso de mascara. Participantes foram questionados se acham que uso de mascara diminui o contágio do COVID.			
2	Resposta	Número de observações	Frequência relativa	
3	Sim	140	70%	
4	Não	50	25%	
5	Não sei	10	5%	
6				
7				



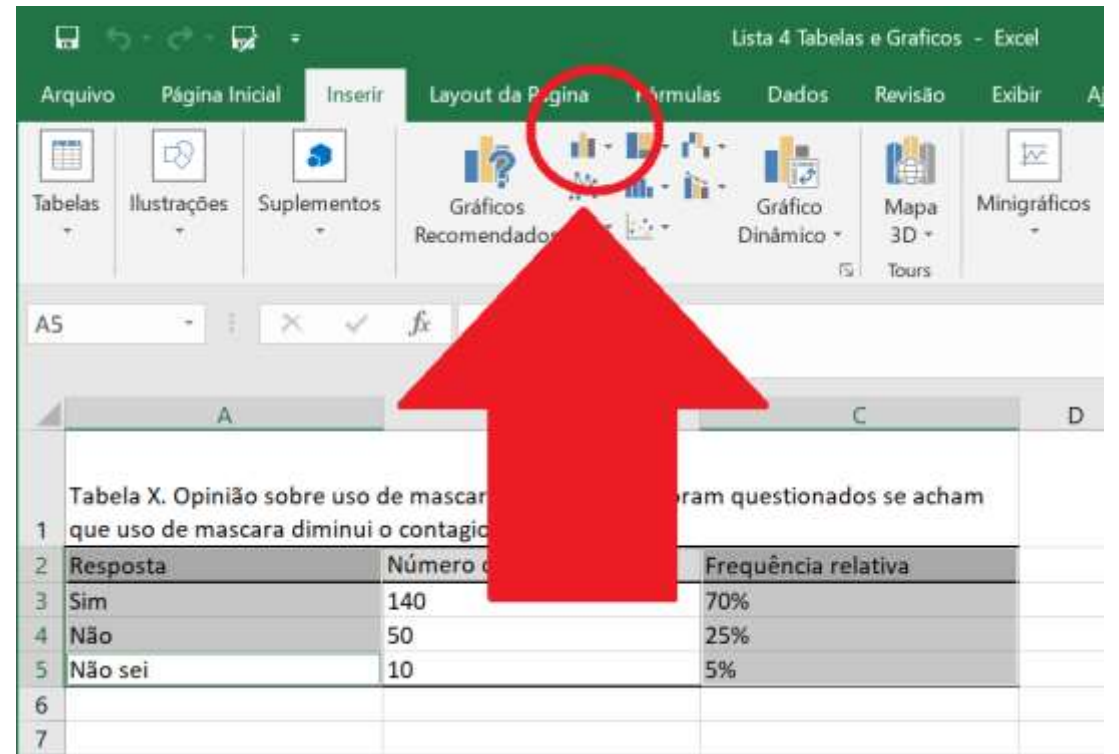
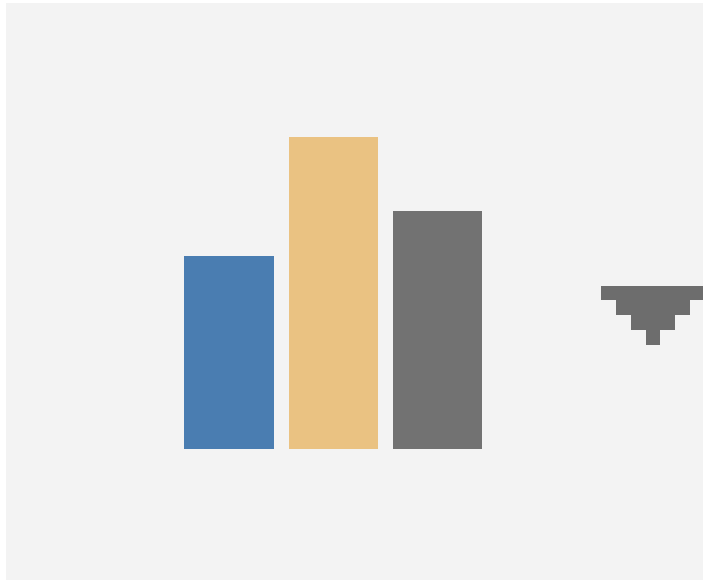
# Como fazer esse gráfico no Excel?

## 3. Clique na aba Inserir



# Como fazer esse gráfico no Excel?

4. Clique no símbolo de gráfico de barras



Lista 4 Tabelas e Gráficos - Excel

Arquivo Página Inicial Inserir Layout da Página Formulas Dados Revisão Exibir

Tabelas Ilustrações Suplementos Gráficos Recomendados Gráfico Dinâmico Mapa 3D Minigráficos

A5

Resposta	Número de	Frequência relativa
Sim	140	70%
Não	50	25%
Não sei	10	5%

# Gráfico em barras também pode aparecer na horizontal

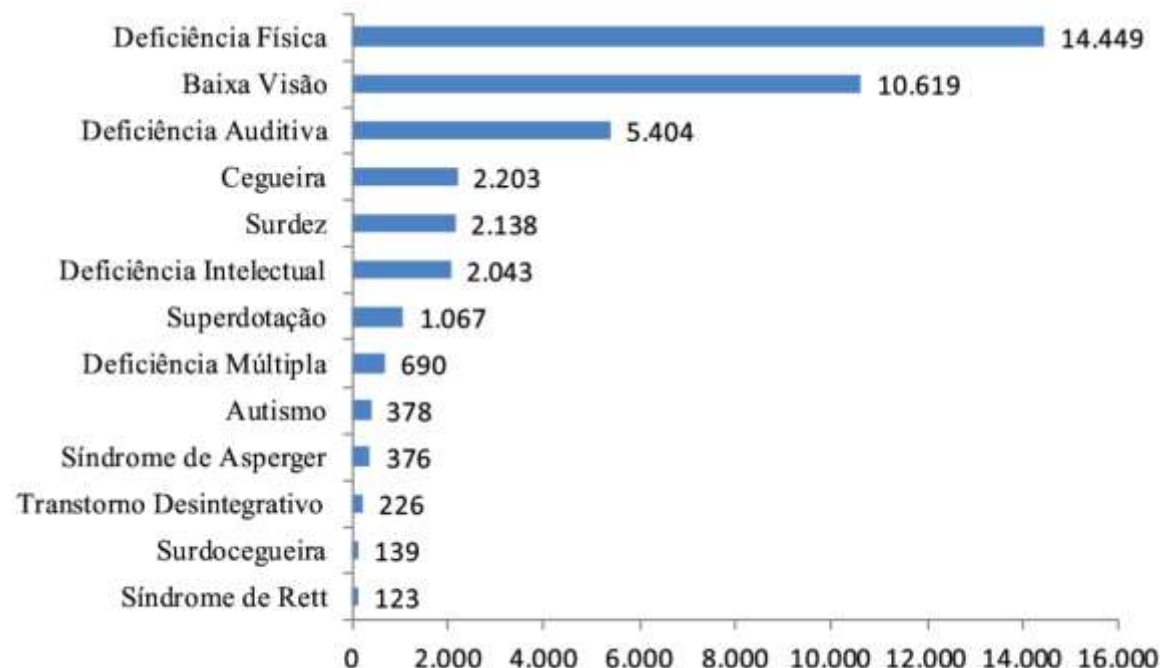


GRÁFICO 4

TOTAL DE MATRÍCULAS DE GRADUAÇÃO CONFORME O TIPO DE DEFICIÊNCIA,  
TRANSTORNO GLOBAL DO DESENVOLVIMENTO OU ALTAS HABILIDADES/SUPERDOTAÇÃO  
DECLARADOS – BRASIL – 2017

Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior.

Nota: Um mesmo vínculo discente pode ter mais de um tipo de deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação.

Gráfico obtido em:

[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/resumo\\_tecnico/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2017.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf)

# Dados

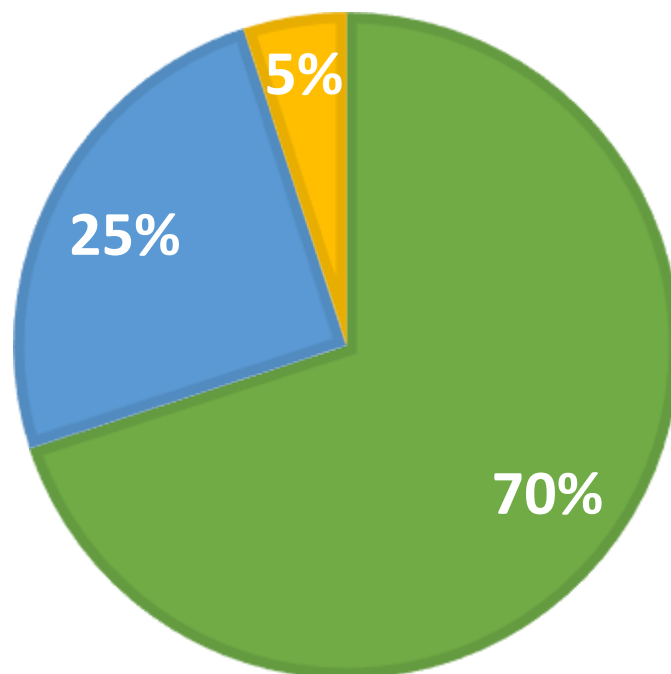
Tabela X. Opinião sobre uso de mascara. Participantes foram questionados se acham que uso de mascara diminui o contagio do COVID.

Resposta	Número de observações	Frequência relativa
Sim	140	70%
Não	50	25%
Não sei	10	5%
Total	200	100%

# Gráfico de setores (pizza)

## FREQUÊNCIAS

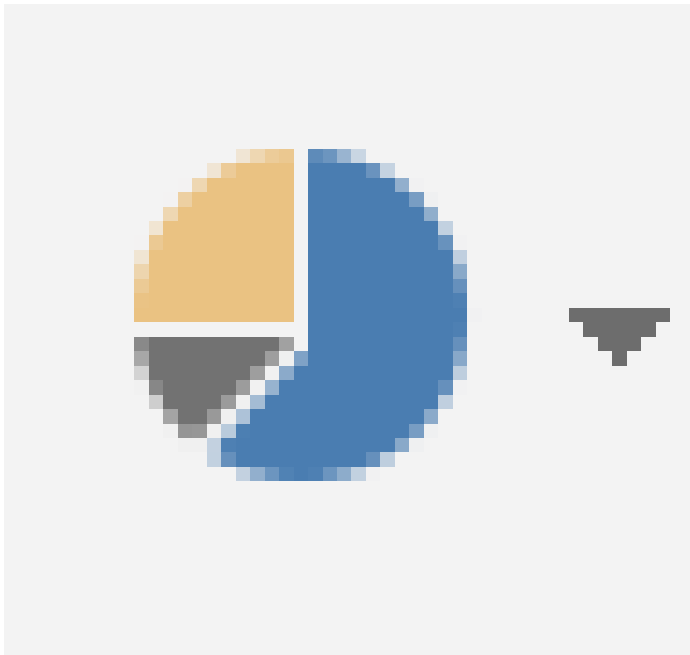
■ Sim ■ Não ■ Não sei



# Como fazer esse gráfico no Excel?

Tudo igual nos passos 1 até 3

4. Clique no símbolo de gráfico de barras



Lista 4 Tabelas e Graficos - Excel

Arquivo Página Inicial **Inserir** Layout da Página Fórmulas Dados Revisão Exibir

Tabelas Ilustrações Suplementos Gráficos Recomendados Gráfico Dinâmico Mapa 3D Minigráficos

A5 Não

Tabela X. Opinião sobre uso de máscara por questionados se acham que uso de máscara diminui o contágio

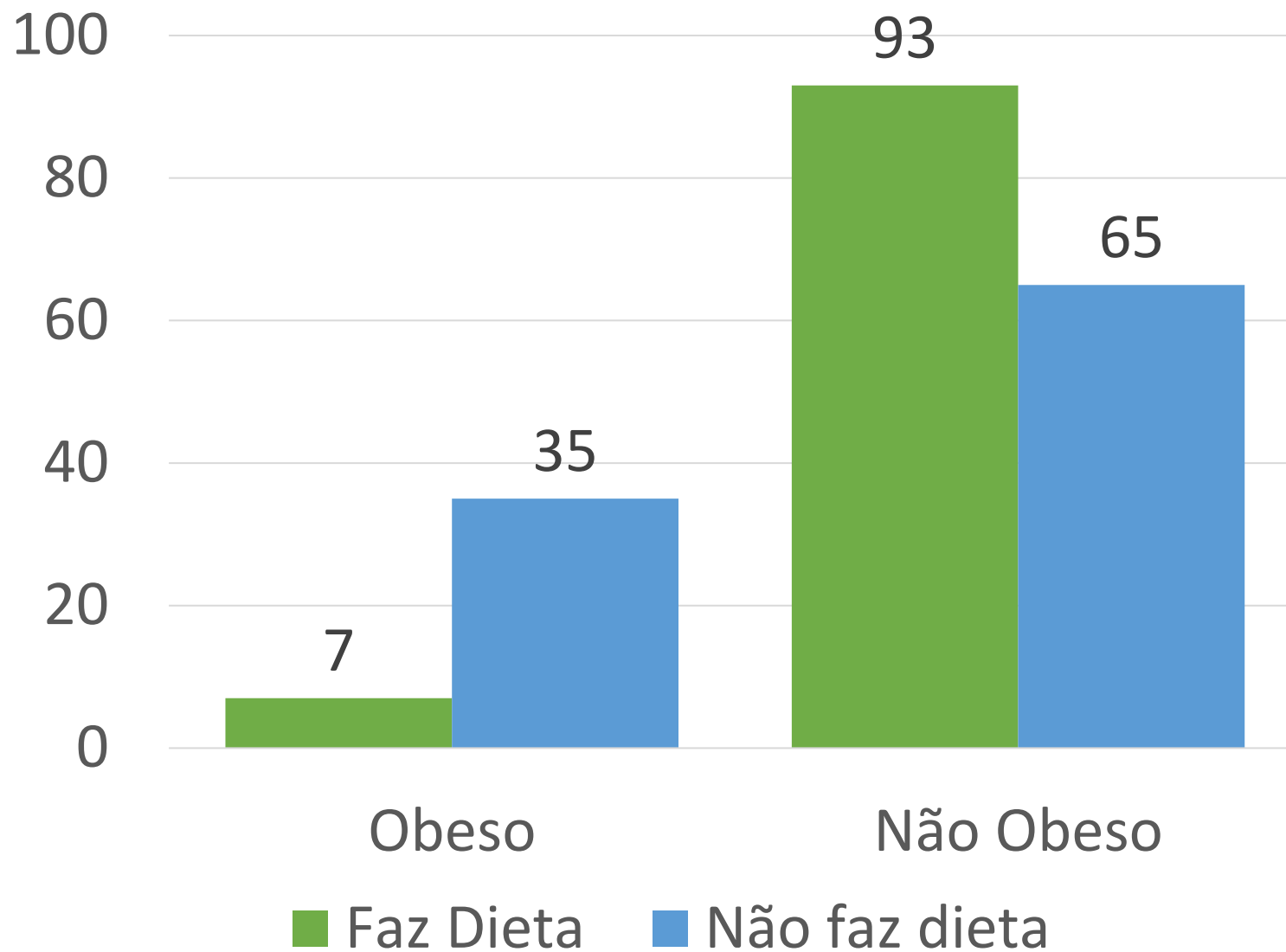
Resposta	Número de	Frequência relativa
Sim	140	70%
Não	50	25%
Não sei	10	5%

# Dados

Tabela X. Obesidade registrada em função do sexo

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Faz dieta	7	93	100	7%
Não faz dieta	35	65	100	35%
Total	42	158	200	25%

# Gráfico de barra

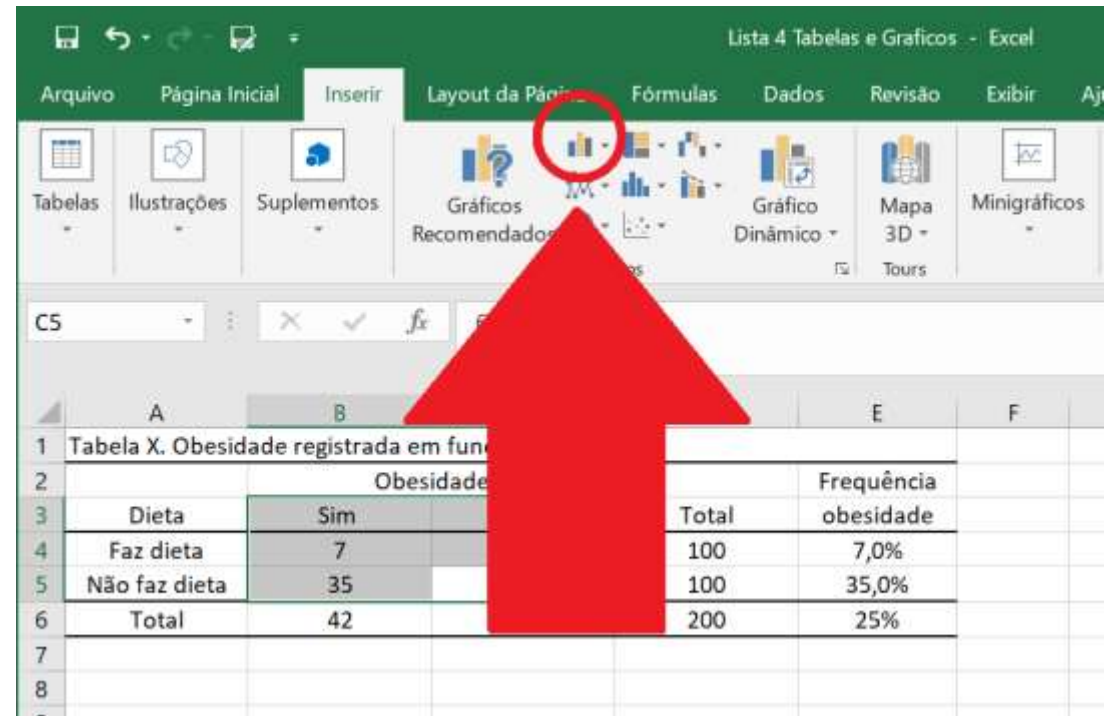
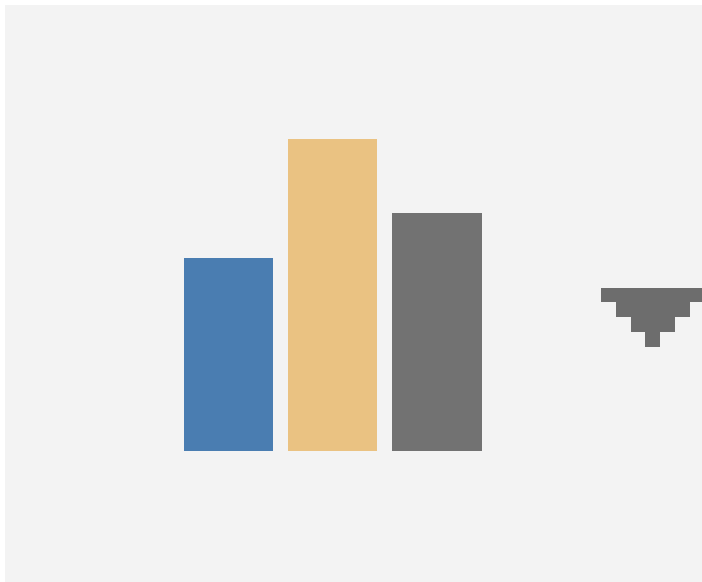




# Como fazer esse gráfico no Excel?

Tudo igual nos passos 1 até 3

4. Clique no símbolo de gráfico de barras



Gráficos para dados quantitativos

# Gráficos para dados quantitativos discretos

- Diagrama de linhas

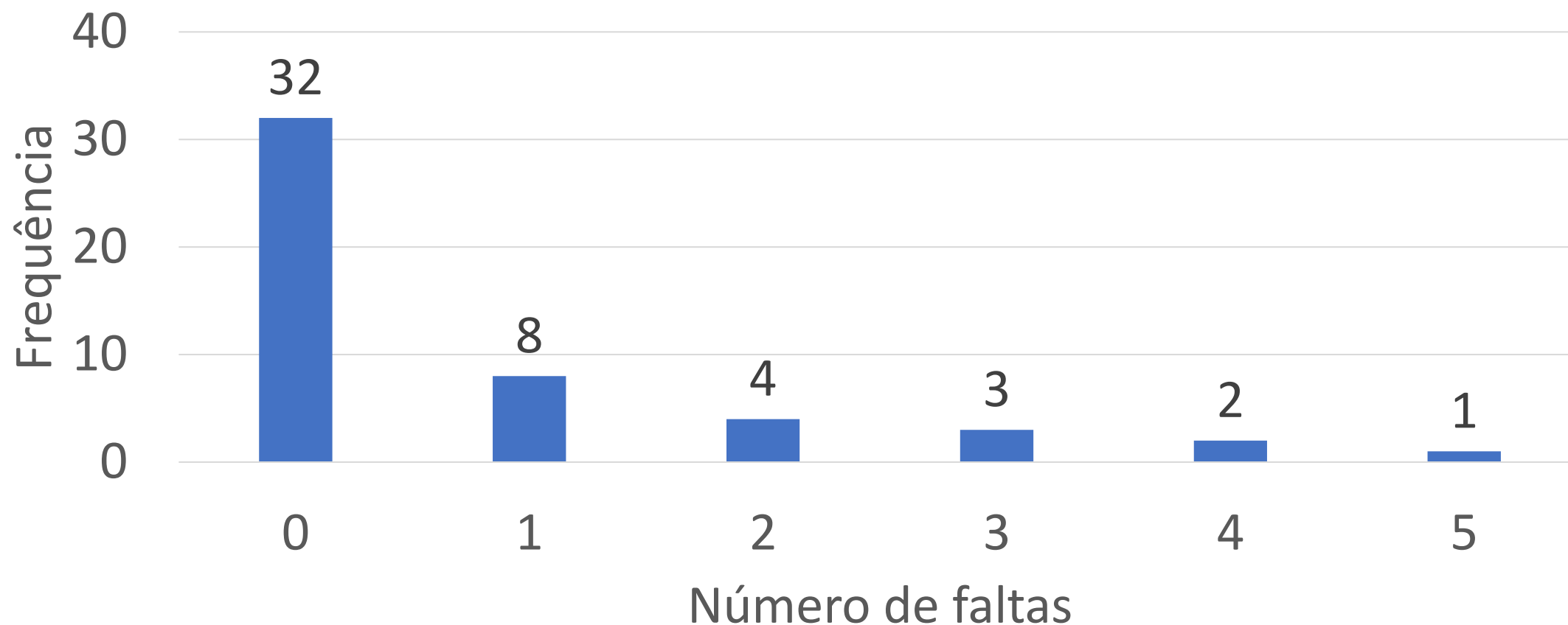
# Tabelas de distribuição de frequência: para dados quantitativos discretos

Tabela X. Numero de faltas em uma equipe de 50 enfermeiros ao longo de 1 mês

Número de faltas	Frequências	Frequência relativa
0	32	0,64
1	8	0,16
2	4	0,08
3	3	0,06
4	2	0,04
5	1	0,02

# Gráficos para dados quantitativos discretos:

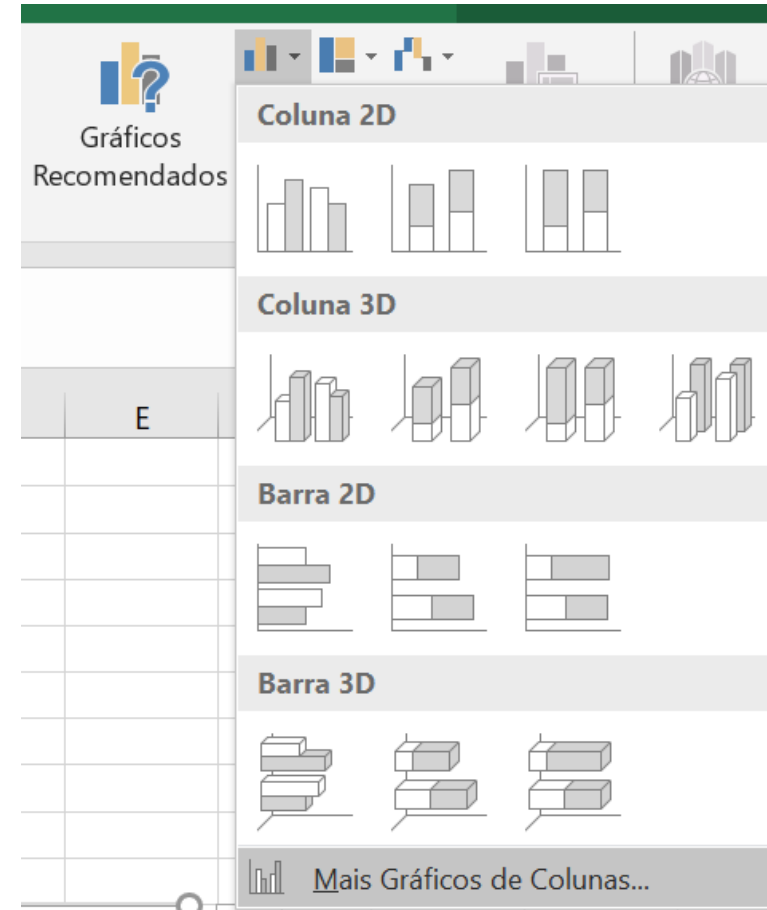
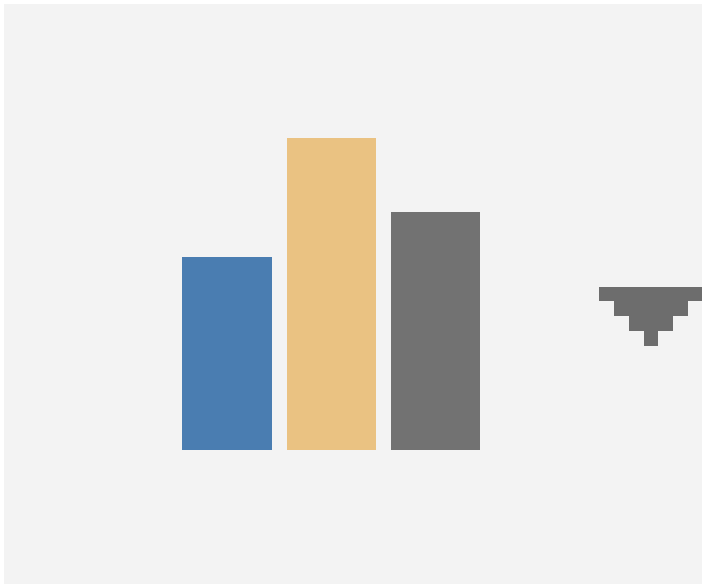
## Diagrama de linhas



# Como fazer esse gráfico no Excel?

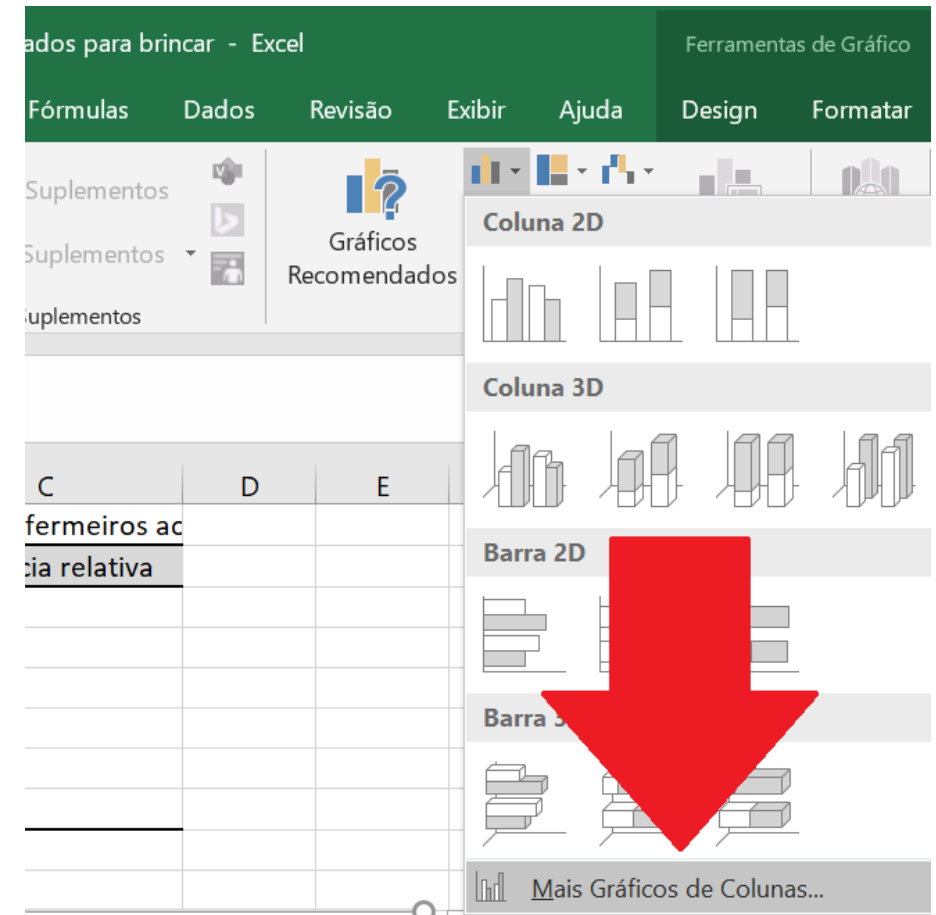
Tudo igual nos passos 1 até 3

4. Clique na setinha para baixo  
símbolo de gráfico de barras



# Como fazer esse gráfico no Excel?

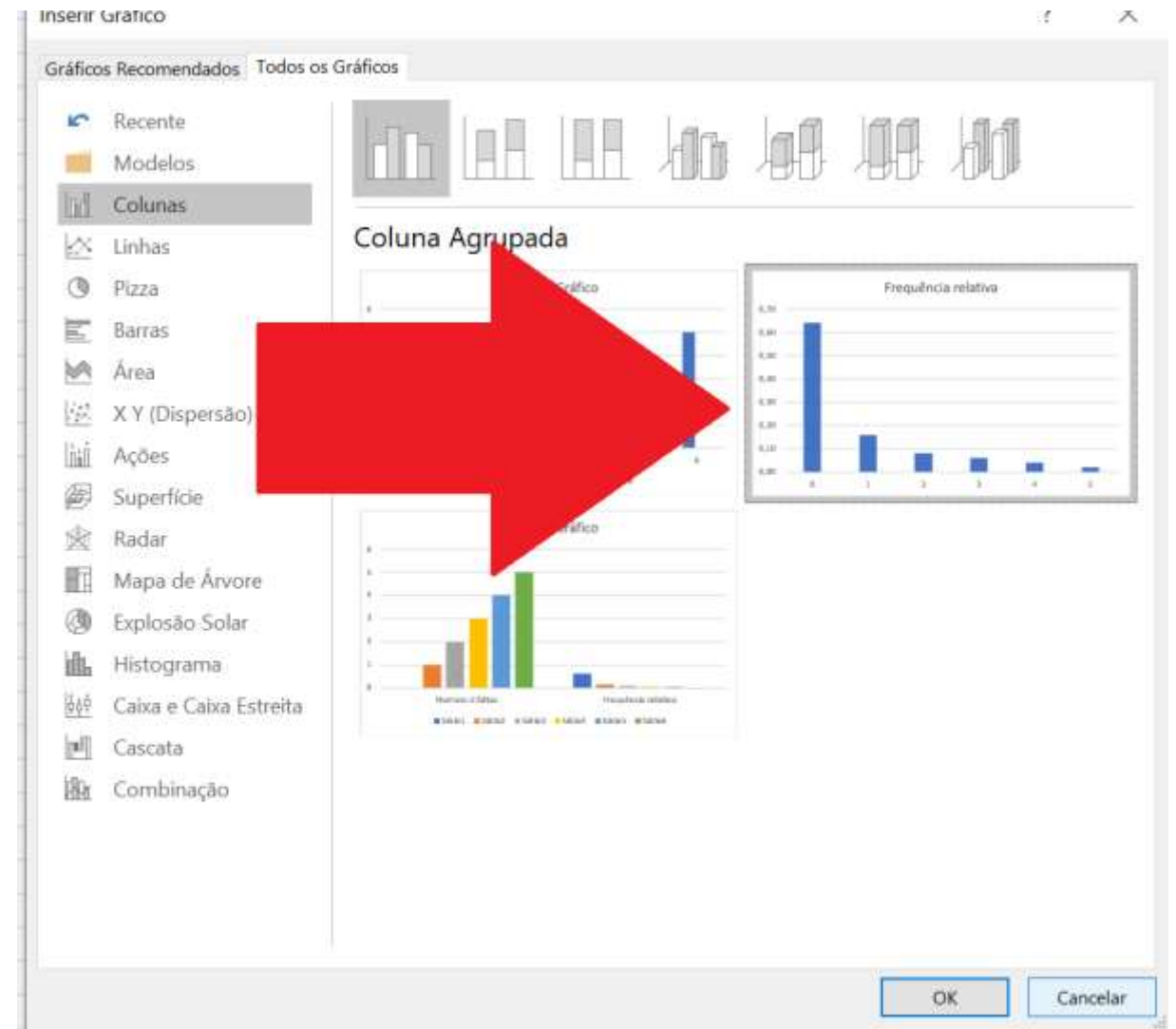
5. Selecione “Mais gráficos de colunas”



Mais Gráficos de Colunas...

# Como fazer esse gráfico no Excel?

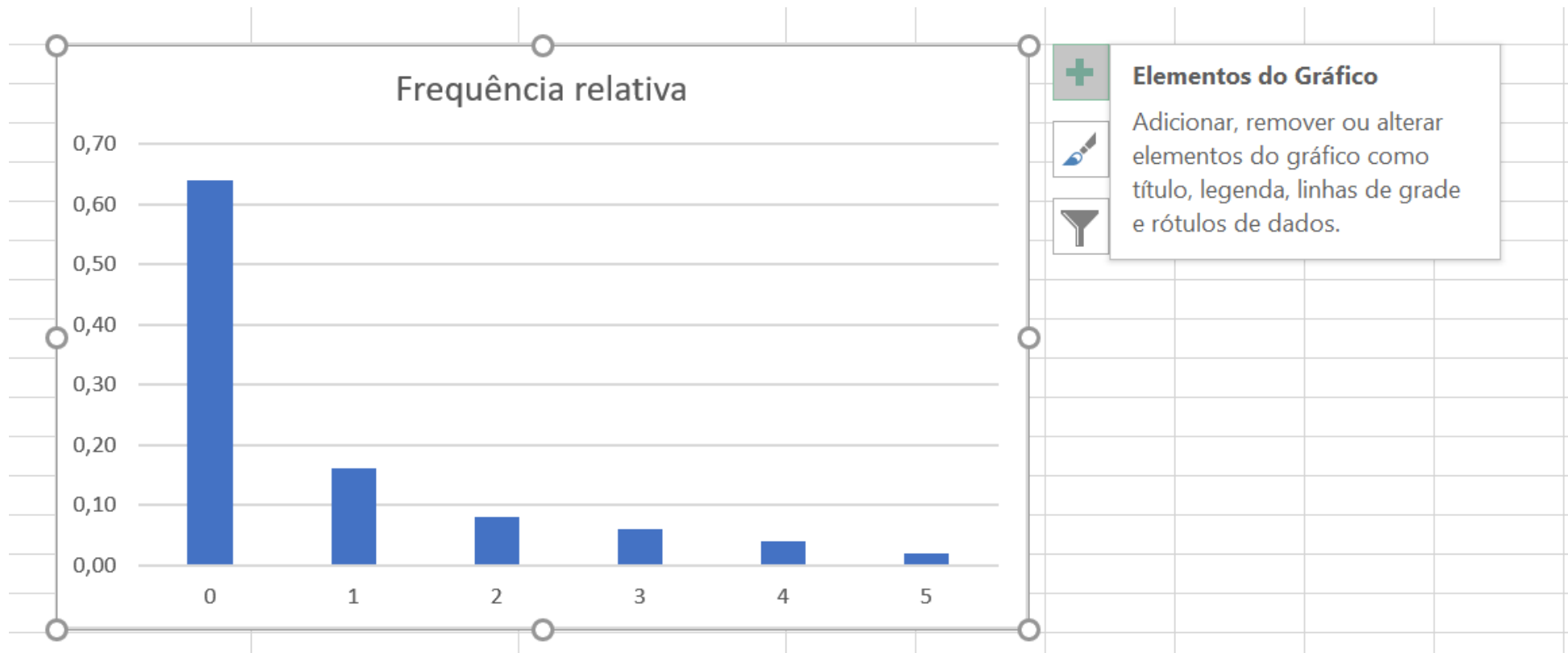
## 6. Selecione o gráfico desejado





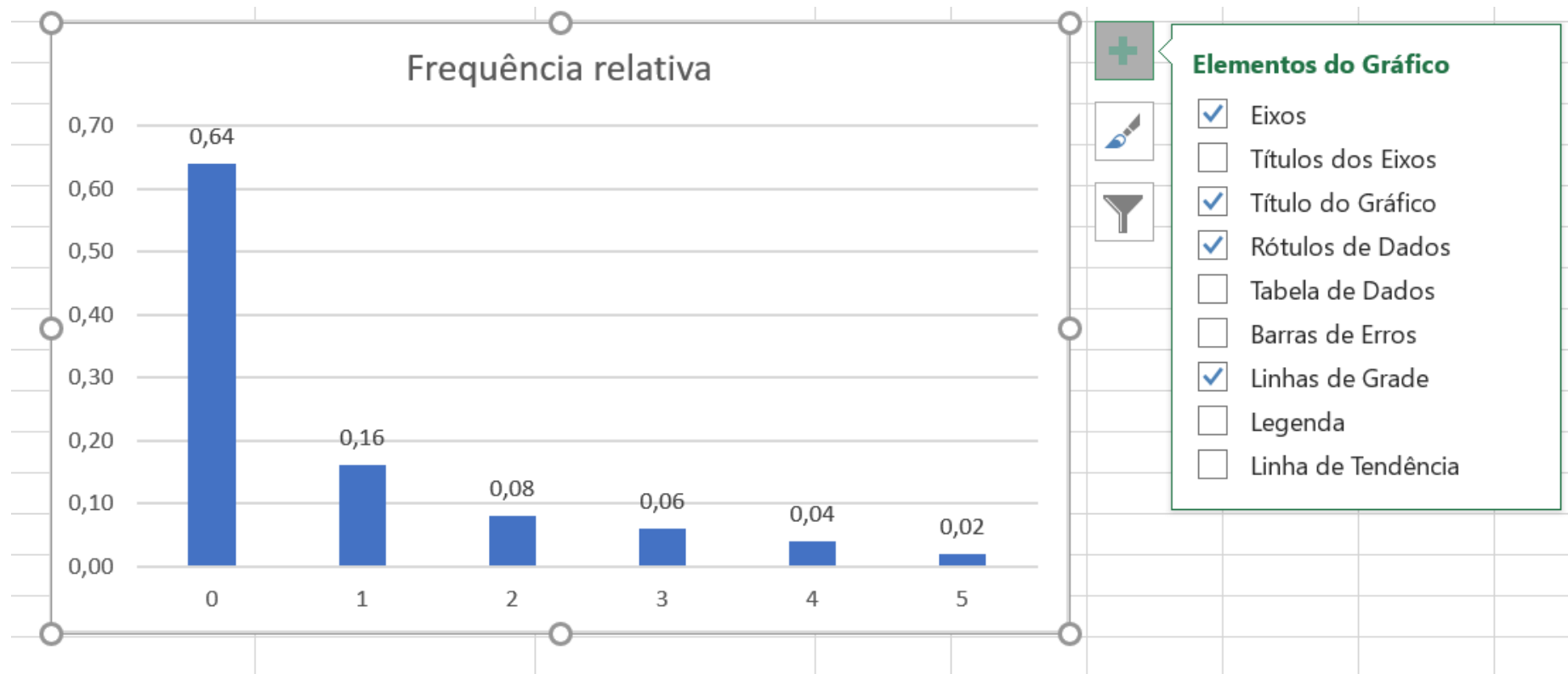
# Como fazer esse gráfico no Excel?

- 7. Para personalizar o gráfico clique no símbolo +



# Como fazer esse gráfico no Excel?

- 8. Selecione as características desejadas



# Gráficos para dados quantitativos contínuos

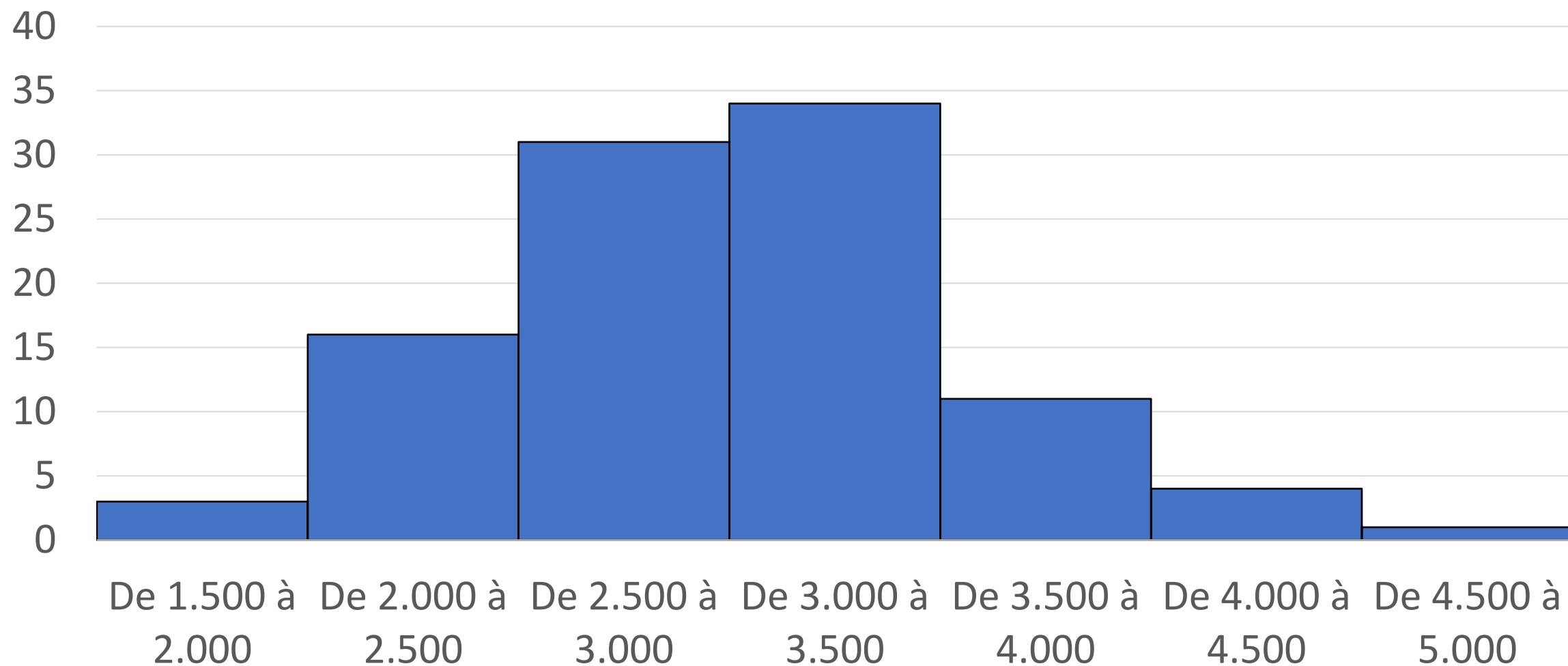
- Histogramas

# Dados

Tabela X. Distribuição de frequência de para peso ao nascimento vivo, em quilogramas

Classe	Frequencias	Frequência relativa
1.500  — 2.000	3	3%
2.000  — 2.500	16	16%
2.500  — 3.000	31	31%
3.000  — 3.500	34	34%
3.500  — 4.000	11	11%
4.000  — 4.500	4	4%
4.500  — 5.000	1	1%

# Histograma



# Como fazer esse gráfico no Excel?

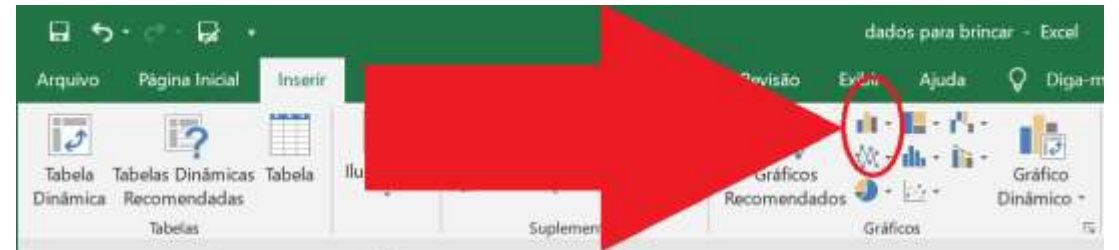
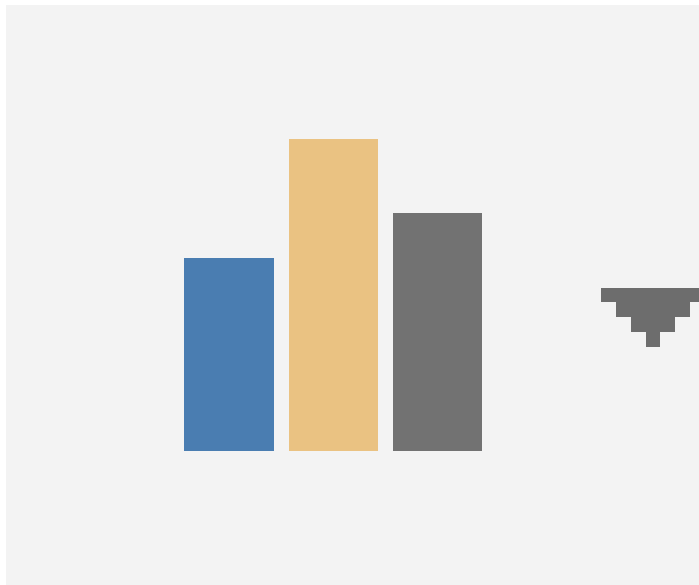
- 2 opções
  - Gráfico de barras
  - Histograma propriamente dito

# Opção 1

- Gráfico de barras

# Como fazer esse gráfico no Excel?

- Fazer gráfico de barras como anteriormente
- Ou seja, repita até o passo 4. do gráfico de barras



dados para brincar - Excel

Arquivo Página Inicial Inserir Revisão Referências Ajuda Diga-nos o que achou

Tabela Tabelas Dinâmicas Tabela Ilustrações Gráficos Gráfico Dinâmico

Tabelas Recomendadas Gráficos Recomendados Gráficos

B5 5

A B C D E F

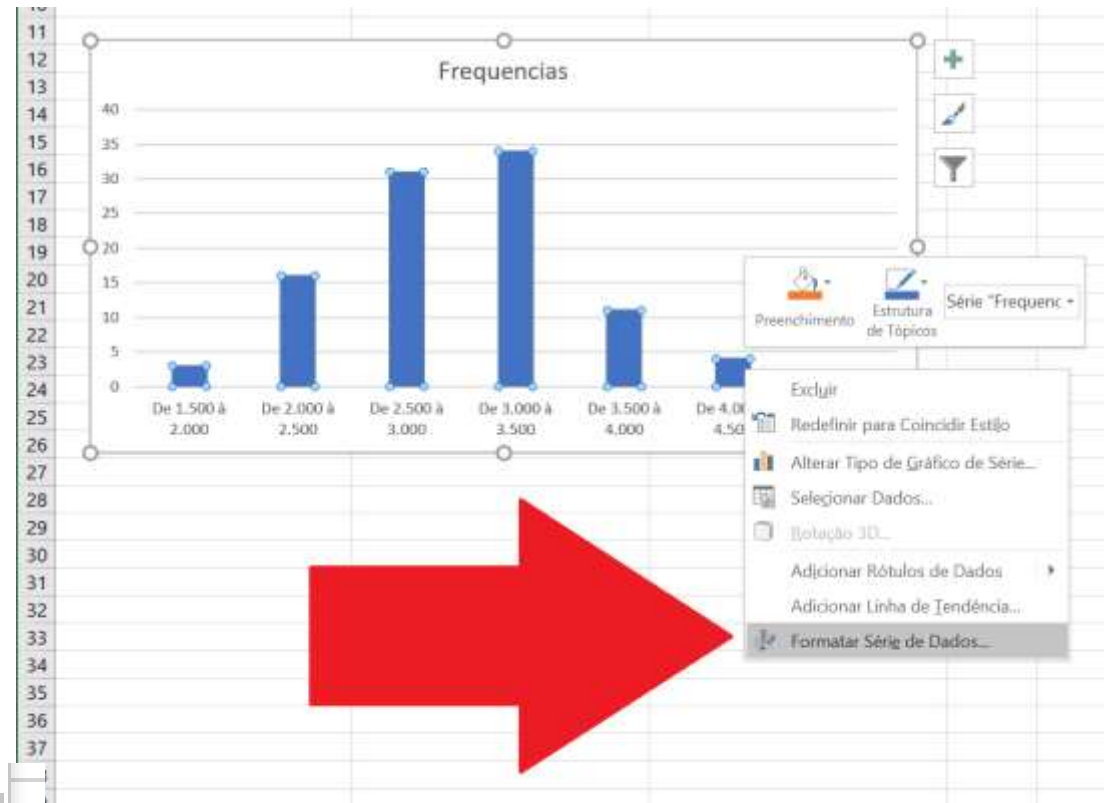
Tabela X. Opinião sobre uso de máscara. Participantes foram questionados se acham que uso de máscara diminui o contágio do COVID.

Resposta	Frequências	Frequência relativa
Sim	70	0,70
Não	25	0,25
Não sei	5	0,05



# Como fazer esse gráfico no Excel?

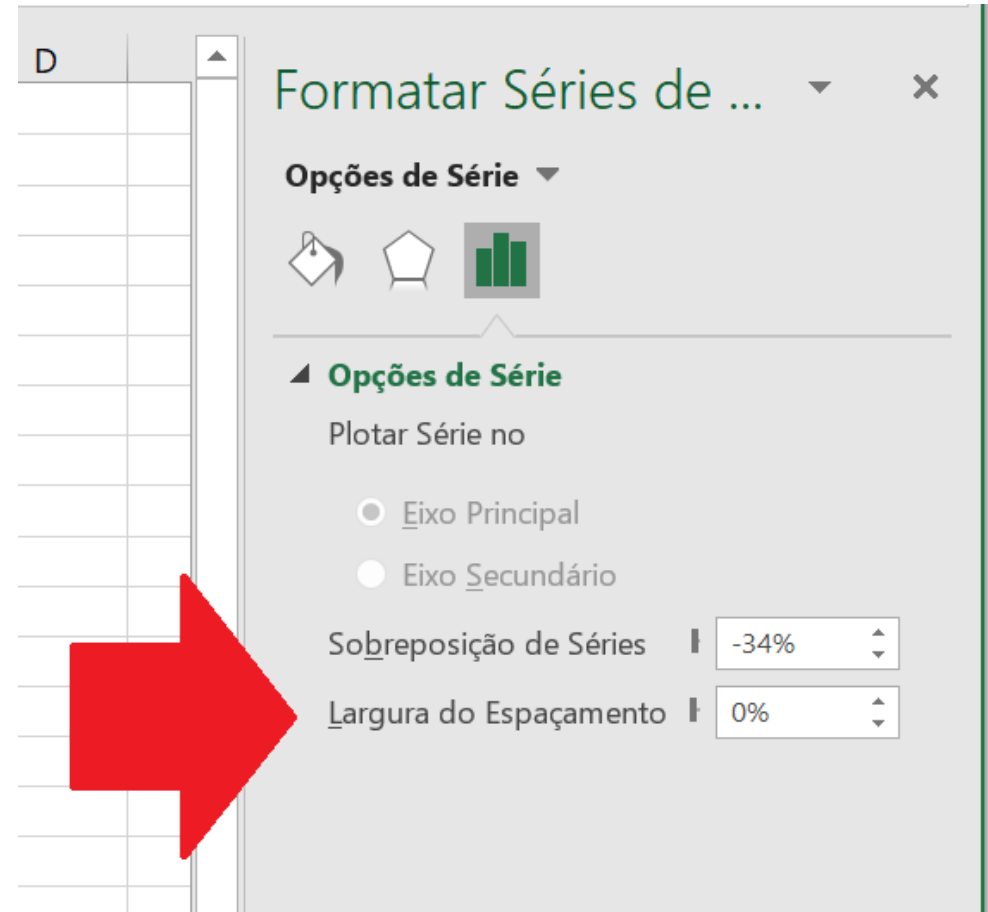
- 5. Clica com o botão direito em cima da uma das barras, em seguida clique em
- “Formatar série de dados”



Formatar Série de Dados...

# Como fazer esse gráfico no Excel?

- 6. Troque a largura do espaçamento para 0



# Opção 2

- Histograma propriamente dito

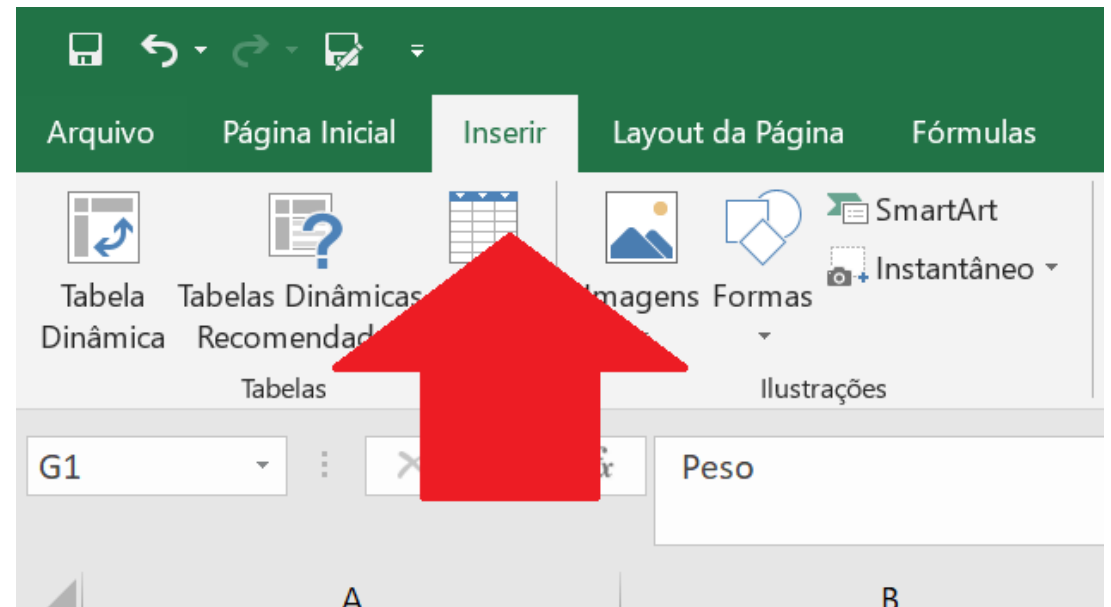
# Como fazer esse gráfico no Excel?

1. Selecione sua coluna com dados bruto

F	G	H
	Peso	
	1800	
	1850	
	1900	
	2110	
	2120	
	2130	
	2140	
	2150	
	2160	
	2170	
	2180	
	2190	
	2200	
	2210	
	2220	
	2230	

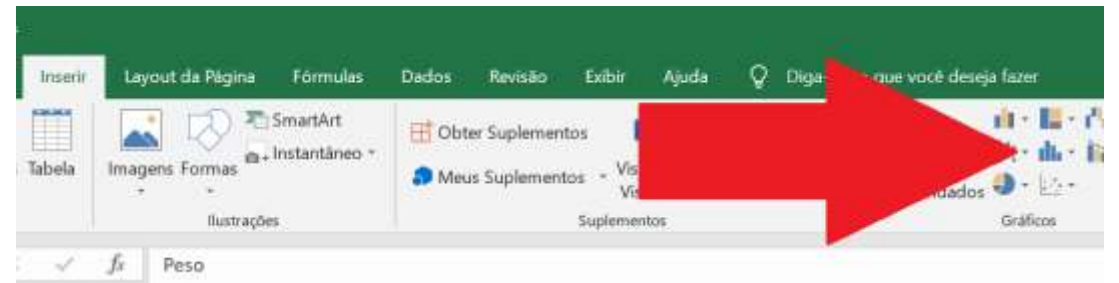
# Como fazer esse gráfico no Excel?

2. Vá na aba inserir



# Como fazer esse gráfico no Excel?

3. Selecione a opção de criar histograma



# Medidas de descrição

Média

Mediana

Quartil

Moda

# Medidas de tendência central

- Você consegue mostrar características gerais dos seus dados por meio de figuras e tabelas
- Mas na prática, medidas numéricas são mais úteis.
  - São números exatos (passível de comparar com outros estudos)
  - Podem ser escritas e faladas em poucas palavras



# Média

- Média é o parâmetro mais simples e popular

- $$\frac{\textit{Soma de todos os dados}}{\textit{Número de observações}}$$

- Influenciada por valores extremos

# Média

- Média é como o centro de gravidade de um conjunto de dados



**FIGURA 4.1** Distribuição de dados de circunferência abdominal, em centímetros, sobre um eixo e a respectiva média.

# Média

- Média é cor



FIGURA 4.1 D  
a respectiva m



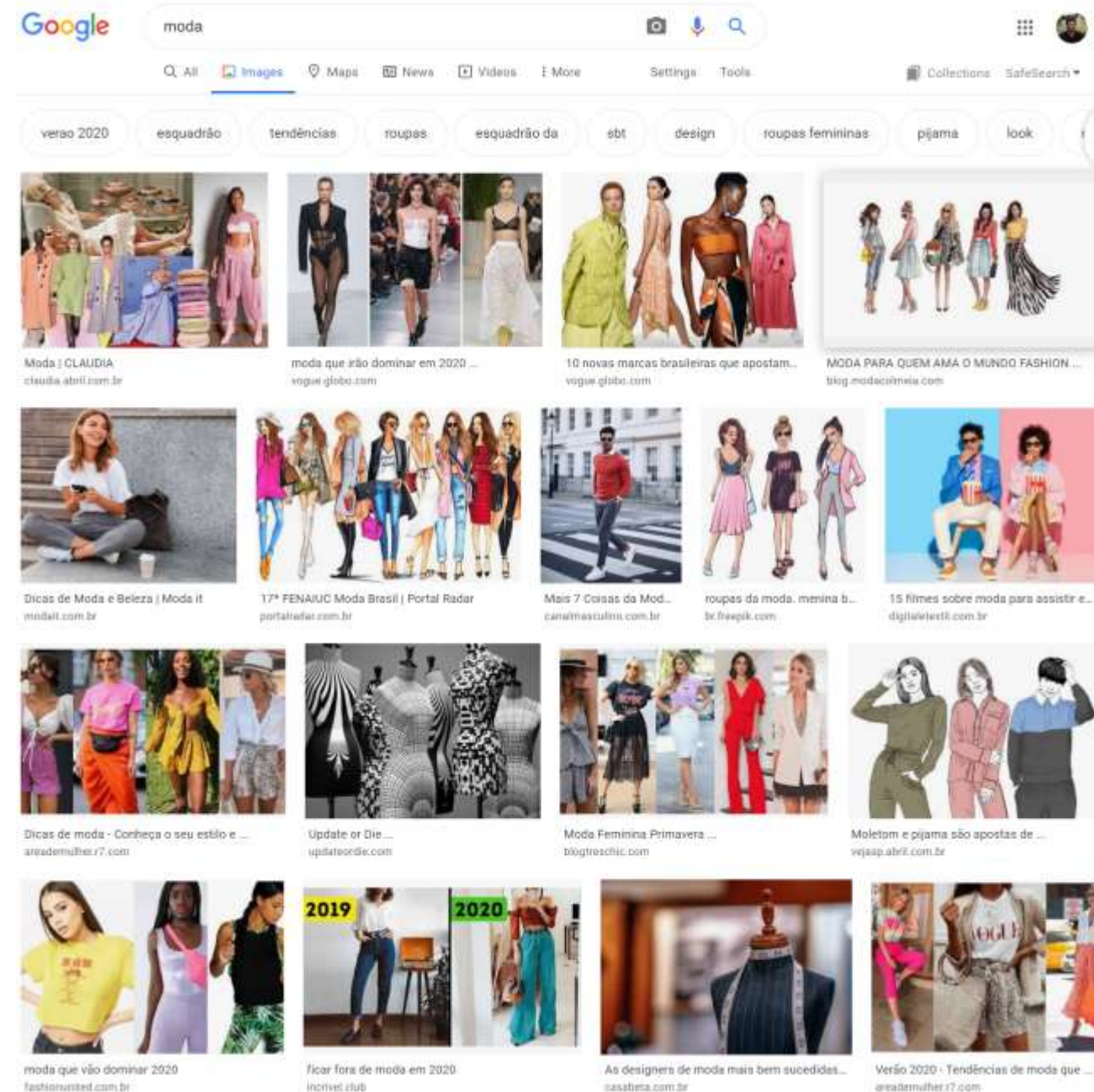
dos

# Moda

- Valor com maior frequência
  - Exemplo, moda de 0,1,3,3,3,4 = 3
- Conjuntos de dados que nenhum número se repete, não tem moda
  - Exemplo, moda de 0,1,2,3,4 = não existe
- Conjuntos de dados que dois ou mais número se repete um mesmo número de vezes, tem mais de uma moda
  - Exemplo, moda de 0,1,1,2,3,3,4 = 1 e 3

# Moda

- Para lembrar: moda é aquilo que está sendo mais usado!



# Mediana

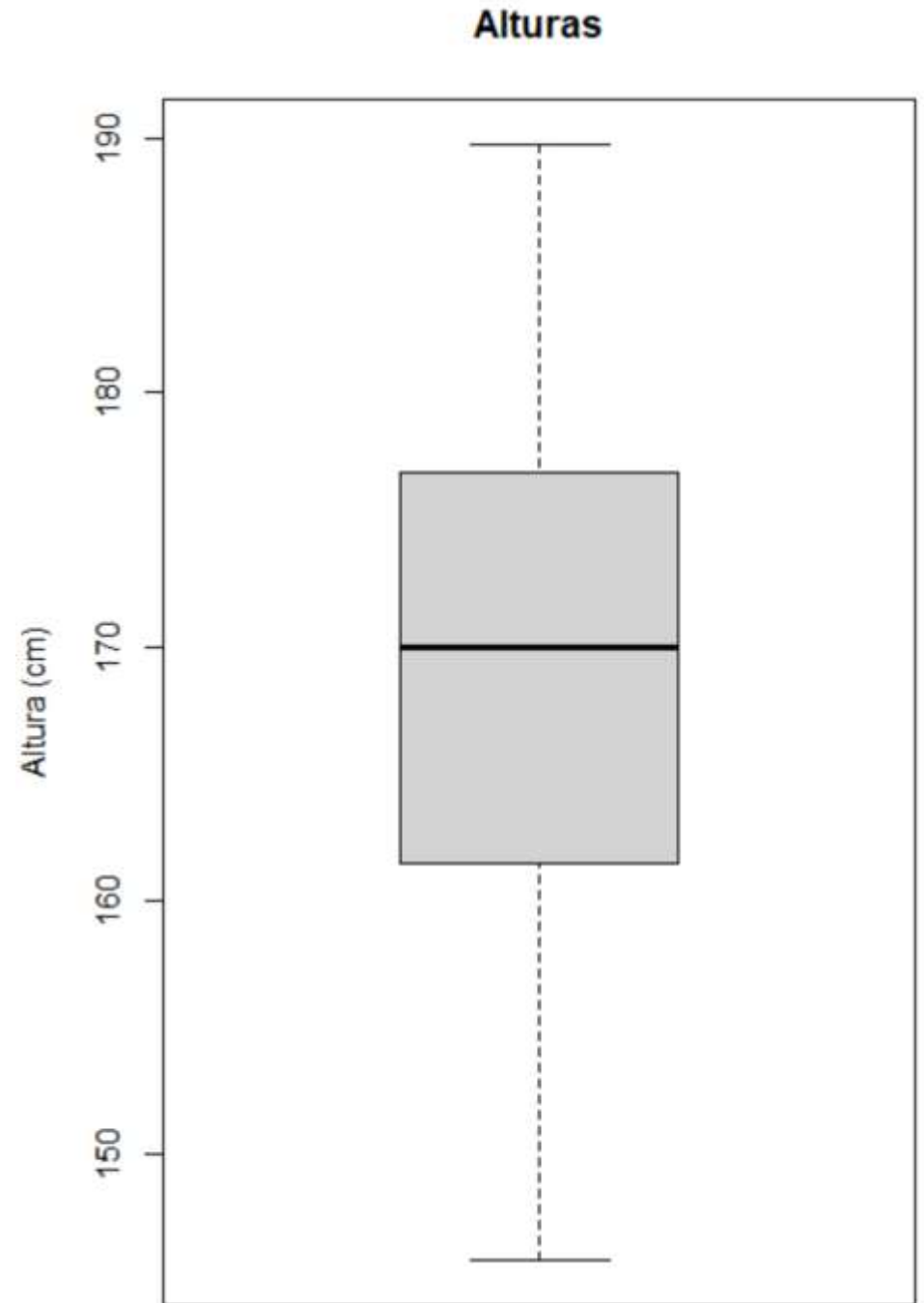
- Valor que ocupa a posição central do conjunto dos dados
- Exemplos
  - mediana de  $\{5,6,7,8,9\} = 7$
  - mediana de  $\{2,4,6,8\} = 5$

# Quartil

- Quartis (Q1, Q2 e Q3): São valores do conjunto de observações ordenado em ordem crescente, que dividem a distribuição em quatro partes iguais.
  - Q1: separa o primeiro  $\frac{1}{4}$  (25 %) das observações
  - Q2: mediana
  - Q3: separa o último  $\frac{1}{4}$  (25 %) das observações

# Boxplot

- boxplot = diagrama de caixa, diagrama de extremos e quartis
- Apresenta a variação de dados observados numérica por meio de quartis

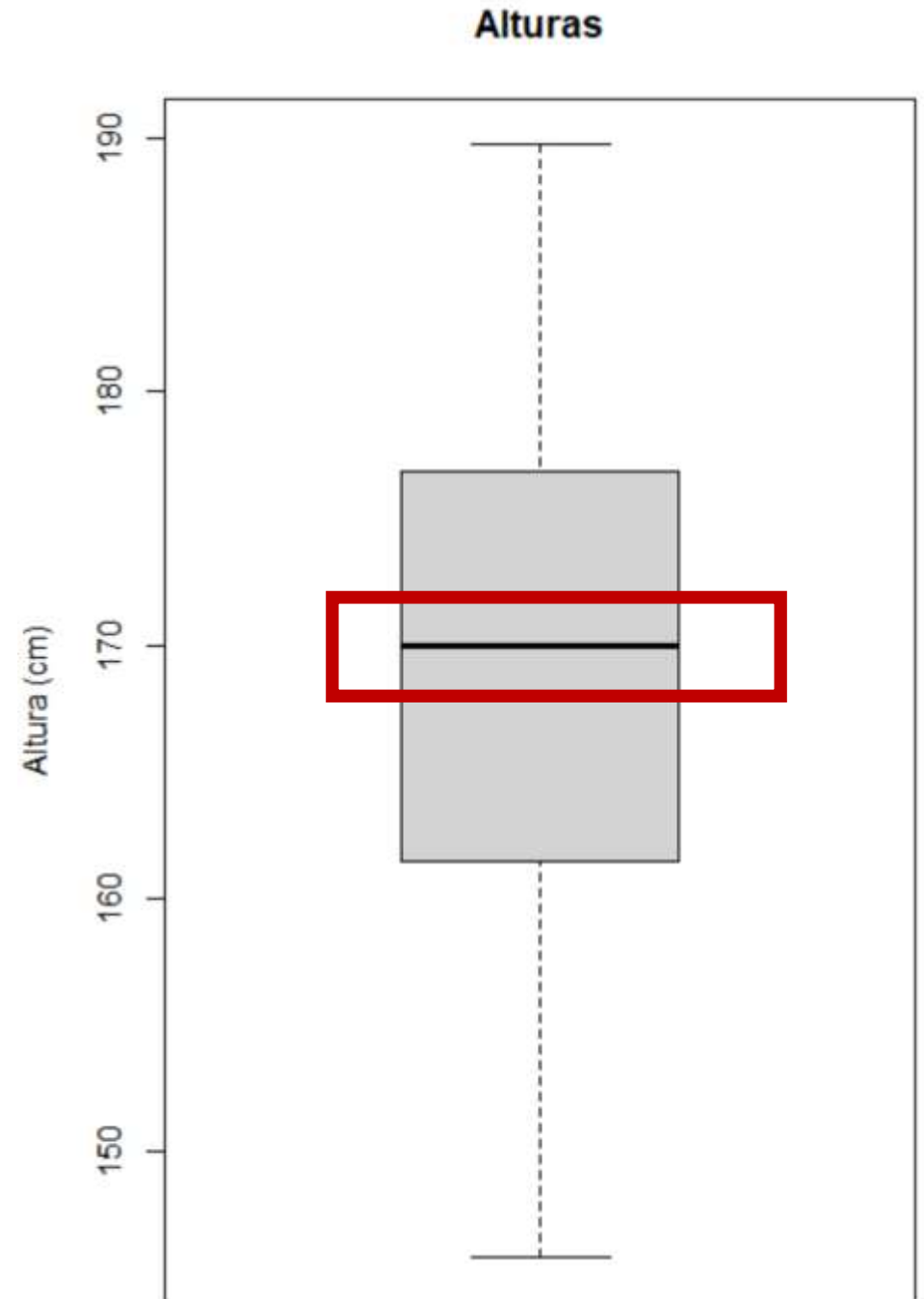




# Boxplot

- boxplot = diagrama de caixa, diagrama de extremos e quartis
- Apresenta a variação de dados observados numérica por meio de quartis

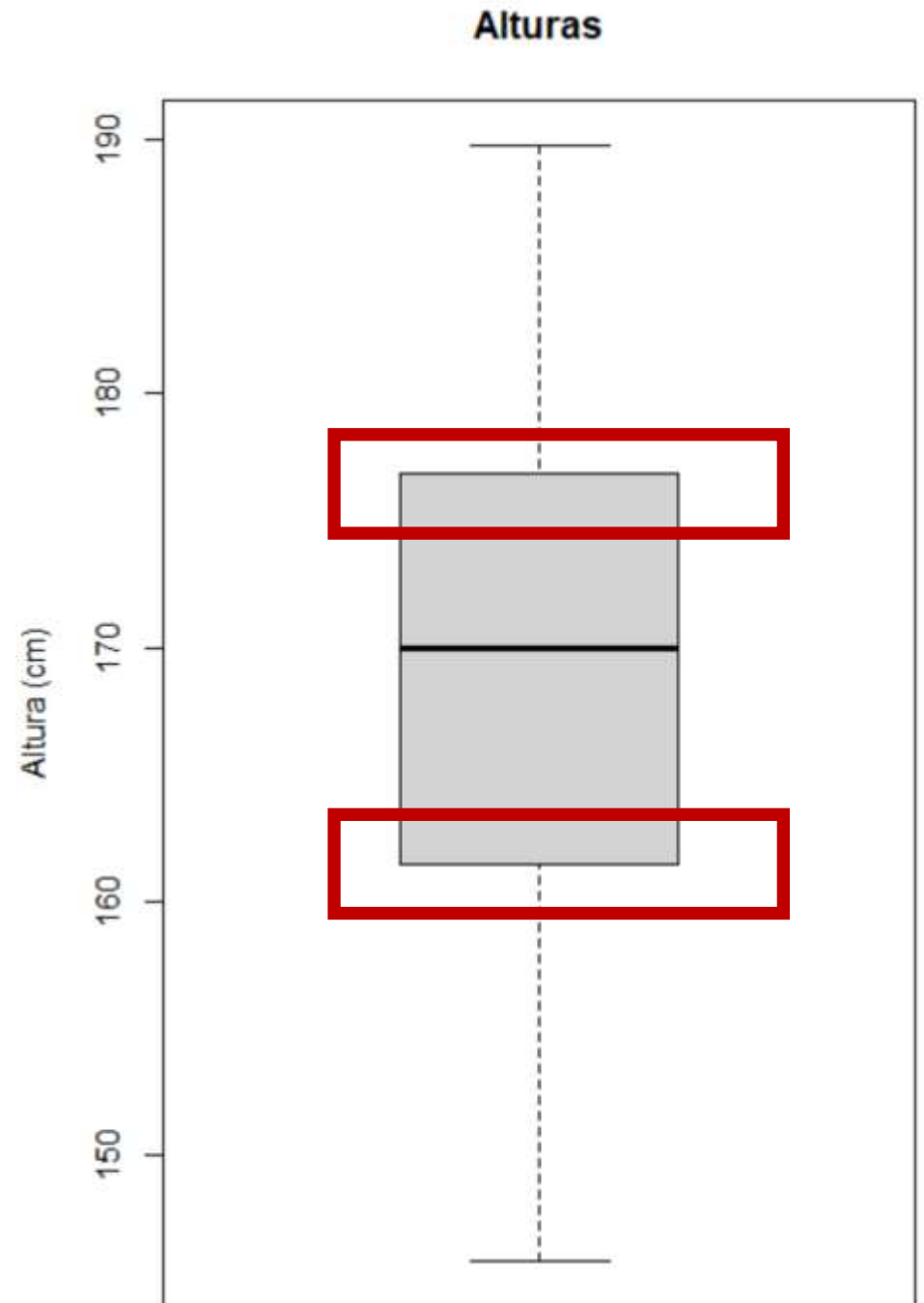
Mediana



# Boxplot

- boxplot = diagrama de caixa, diagrama de extremos e quartis
- Apresenta a variação de dados observados numérica por meio de quartis

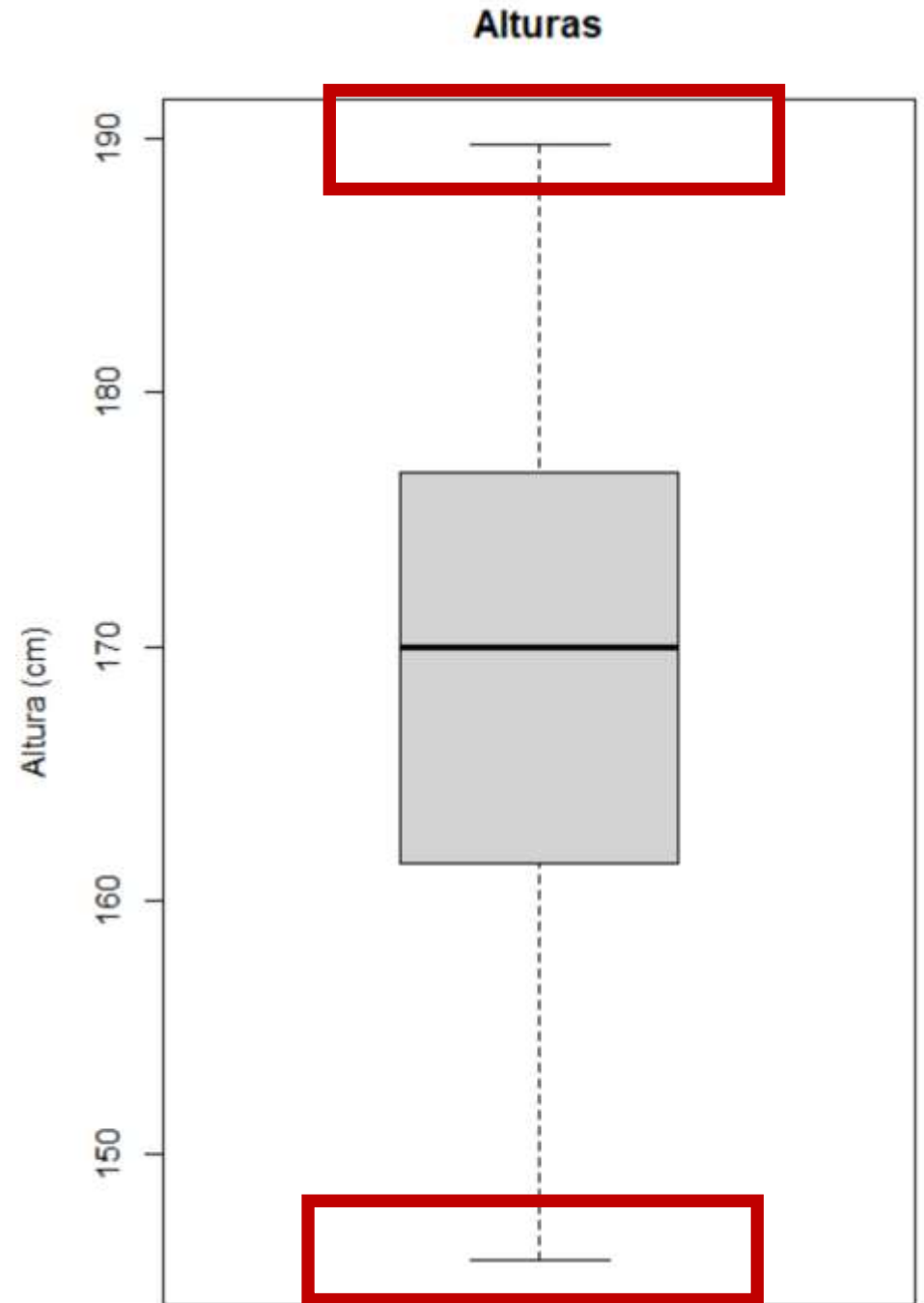
1º e 3º Quartil



# Boxplot

- boxplot = diagrama de caixa, diagrama de extremos e quartis
- Apresenta a variação de dados observados numérica por meio de quartis

Mínimo e máximo



# Variância/Desvio

- Medida de dispersão
- Diz o quanto que os dados estão espalhados em relação a média
- Exemplo, os dois conjuntos de dados a seguir tem a mesma média, mas o primeiro tem maior desvio.
  - Conjunto 1: 10, 20, 30, 40, 50
  - Conjunto 2: 28, 29, 30, 31, 32
- Essa medida é muito importante, mas vou deixar para entrar em detalhes quando formos sobre distribuições de frequência

# Exemplo prático da apresentação desse conceitos

TABELA 20

MEDIDAS ESTATÍSTICAS DESCRITIVAS PARA AS IDADES REFERENTES ÀS MATRÍCULAS NOS CURSOS DE GRADUAÇÃO, SEGUNDO A MODALIDADE DE ENSINO – BRASIL – 2017

MODALIDADE DE ENSINO	IDADE <sup>1</sup> REFERENTE À MATRÍCULA						FREQUÊNCIA MODAL <sup>2</sup>
	1º QUARTIL	MEDIANA	3º QUARTIL	MÉDIA	DESVIO-PADRÃO	MODA	
Presencial	21	23	28	25,6	7,4	21	692.549
a Distância	25	31	38	32,3	9,2	29	73.086

Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior.

<sup>1</sup> Idade consiste no cálculo produzido a partir dos dados cadastrais de alunos e docentes relativos a dia, mês e ano de nascimento, na data de referência do censo: 31 de dezembro do ano do referido censo (Brasil. Inep, 2012).

<sup>2</sup> Frequência modal corresponde ao número de observações dessa medida estatística descritiva, a qual identifica o atributo com maior frequência na distribuição do aspecto selecionado.

Gráfico obtido em:

[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/resumo\\_tecnico/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2017.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf)

# Exemplo prático da apresentação desse conceitos

50% dos matriculados tem menos de 23 anos

50% dos matriculados tem mais que 23 anos

RENTES  
DALIDADE

DE ENSINO – BRASIL – 2017

MODALIDADE DE ENSINO	IDADE <sup>1</sup> REFERENTE À MATRÍCULA						FREQUÊNCIA MODAL <sup>2</sup>
	1º QUARTIL	MÉDIA	3º QUARTIL	MÉDIA	DESVIO-PADRÃO	MODA	
Presencial	21	23	28	25,6	7,4	21	692.549
a Distância	25	31	38	32,3	9,2	29	73.086

Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior.

<sup>1</sup> Idade consiste no cálculo produzido a partir dos dados cadastrais de alunos e docentes relativos a dia, mês e ano de nascimento, na data de referência do censo: 31 de dezembro do ano do referido censo (Brasil. Inep, 2012).

<sup>2</sup> Frequência modal corresponde ao número de observações dessa medida estatística descritiva, a qual identifica o atributo com maior frequência na distribuição do aspecto selecionado.

Gráfico obtido em:

[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/resumo\\_tecnico/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2017.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf)

# Exemplo prático da apresentação desse conceitos

TABELA 20

MEDIDAS ESTATÍSTICAS DESCRITIVAS PARA AS IDADES REFERENTES ÀS MATRÍCULAS DE GRADUAÇÃO, SEGUNDO A MODALIDADE DE ENSINO – BRASIL – 2017

Distância de 2

MODALIDADE DE ENSINO	IDADE <sup>1</sup> REFERENTE À MATRÍCULA						FREQUÊNCIA MODAL <sup>2</sup>
	1º QUANTIL	MEDIANA	3º QUANTIL	MÉDIA	DESVIO-PADRÃO	MODA	
Presencial	21	23	28	25,6	7,4	21	692.549
a Distância	25	31	38	32,3	9,2	29	73.086

Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior.

<sup>1</sup> Idade consiste no cálculo produzido a partir dos dados cadastrais de alunos e docentes relativos a dia, mês e ano de nascimento, na data de referência do censo: 31 de dezembro do ano do referido censo (Brasil. Inep, 2012).

<sup>2</sup> Frequência modal corresponde ao número de observações dessa medida estatística descritiva, a qual identifica o atributo com maior frequência na distribuição do aspecto selecionado.

Gráfico obtido em:

[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/resumo\\_tecnico/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2017.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf)



# Exemplo prático da apresentação desse conceitos

TABELA 20

MEDIDAS ESTATÍSTICAS DESCRITIVAS PARA AS IDADES REFERENTES ÀS MATRÍCULAS **Distância de 5** AÇÃO, SEGUNDO A MODALIDADE DE ENSINO – BRASIL – 2017

MODALIDADE DE ENSINO	IDADE <sup>1</sup> REFERENTE À MATRÍCULA						FREQUÊNCIA MODAL <sup>2</sup>
	1º QUARTIL	MEDIANA	2º QUARTIL	MÉDIA	DESVIO-PADRÃO	MODA	
Presencial	21	23	28	25,6	7,4	21	692.549
a Distância	25	31	38	32,3	9,2	29	73.086

Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior.

<sup>1</sup> Idade consiste no cálculo produzido a partir dos dados cadastrais de alunos e docentes relativos a dia, mês e ano de nascimento, na data de referência do censo: 31 de dezembro do ano do referido censo (Brasil. Inep, 2012).

<sup>2</sup> Frequência modal corresponde ao número de observações dessa medida estatística descritiva, a qual identifica o atributo com maior frequência na distribuição do aspecto selecionado.

Gráfico obtido em:

[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/resumo\\_tecnico/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2017.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/resumo_tecnico/resumo_tecnico_censo_da_educacao_superior_2017.pdf)



Populações vs. Amostras

# Margens de erro associada a estimativa de parâmetros

- Quando você trabalha com dados de censo, você tem certeza dos parâmetros que você está apresentando
- Quando você trabalha com amostras, você não tem certeza dos parâmetros



A diagram illustrating the relationship between a sample and a population. At the bottom left, the top of a person's head with dark curly hair is visible, looking upwards. Above the head is a red oval containing the word 'amostra'. To the right of the oval are three blue cloud-like shapes of increasing size, leading to a large blue cloud at the top right containing the word 'POPULAÇÃO'.

POPULAÇÃO

amostra

# Portanto

- Para descrever dados de uma população ou amostra você vai usar os mesmos parâmetros
- O que muda é que você precisa passar junto com a medida de descrição uma estimativa do erro do seu trabalho

Como saber o erro associado aos meus dados?

Estimar o erro antes de coletar os dados:

Calculo de tamanho da amostra

# Como calcular o tamanho da amostra?

<div>Revista de Odontologia da Universidade Cidade de São Paulo 2008 maio-ago; 20(2): 186-194</div> <div>AMOSTRAGEM EM PESQUISA CLÍNICA: TAMANHO DA AMOSTRA SAMPLING IN CLINICAL RESEARCH: SAMPLE'S SIZE</div>		
<div>Determinação do tamanho da amostra experimentais na área de s</div> <div>Gastão Rúbio de Sá Weyne*</div>	<div>Indian J Ophthalmol. 2010 Nov-Dec; 58(6): 517–518. doi: <a href="#">10.4103/0301-4738.71692</a></div> <div>Principles of sample size calculation</div> <div>Nithya J Gogtay</div>	<div>Int J Ayurveda Res. 2010 Jan-Mar; 1(1): 55–57. doi: <a href="#">10.4103/0974-7788.59946</a></div> <div>Sample size calculation</div> <div>Prashant Kadam and Supriya Bhalerao<sup>1</sup></div> <div>PMCID: PMC28 PMID: 205</div>
<div>ção de Bioestatística</div> <div>Cálculo de tamanho de amostra: proporções</div> <div>Sample size calculation: proportions</div> <div>Marilyn Agranonik<sup>1</sup>, Vânia Naomi Hirakata<sup>2</sup></div>	<div></div> <div>Medicina Farmacêutica</div> <div>QUAL O TAMANHO DA AMOSTRA IDEAL PARA SE REALIZAR UM ENSAIO CLÍNICO?</div> <div>Diferença entre médias</div> <div><math display="block">n = \frac{(\alpha + \beta)}{d^2}</math></div>	<div>STATISTICS IN MEDICINE Statist. Med. 17, 1623–1634 (1998)</div> <div>A SIMPLE METHOD OF SAMPLE SIZE CALCULATION LINEAR AND LOGISTIC REGRESSION</div>
<div>ARTIGO DE REVISÃO</div> <div>QUESTÕES EM BIOESTATÍSTICA: O TAMANHO DA AMOSTRA STATISTICAL QUESTIONS: THE SAMPLE SIZE</div> <div>des Alberto Mourão Júnior*</div> <div>RESUMO</div> <div>O presente artigo discute de forma crítica a questão do cálculo do tamanho da amostra e sugere estratégias práticas para a abordagem do problema.</div> <div>PALAVRAS-CHAVE</div> <div>Bioestatística. Amostra. População. Teorema central do limite.</div>	<div>MEDICAL STATISTICS</div> <div>Practical Issues in Calculating the Sample Size</div> <div>L. Naing<sup>1,2*</sup>, T. Winn<sup>2</sup>, B.N. Rusli<sup>1,2</sup></div> <div><sup>1</sup>Department of Community Dentistry, School of Dental Sciences, Faculty of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150 Teluk Anson, Penang, Malaysia <sup>2</sup>Corresponding author: <a href="mailto:naing@kck.usm.my">naing@kck.usm.my</a></div>	<div>Theory and practice</div> <div>Sample size calculation in epidemiological studies</div> <div>V. Kasiulevičius<sup>1</sup>, V. Šapoka<sup>1</sup>, R. Filipavičiūtė<sup>2</sup></div> <div><sup>1</sup> Vilnius University <sup>2</sup> Institute of Experimental and Clinical Medicine at Vilnius University</div> <div>Summary</div>

# Qual método escolher?

- Existe muita discussão, e diferentes áreas podem usar métodos diferentes.
- Na prática, o tamanho é frequentemente limitado pelo custo em coletar cada unidade amostrada (Vieira 2011).
- Em muitas áreas o tamanho amostral acaba sendo o máximo que o pesquisador conseguir coletar. Dado o tempo disponível e orçamento.



# Fórmula geral para estudos clínicos e experimentais

**Quadro 1.** Fórmulas para cálculo do tamanho de amostras para descrição de variáveis quantitativas e qualitativas em uma população.

	Variável quantitativa	Variável qualitativa
População infinita	$n = \left( \frac{Z_{\alpha/2} \cdot \delta}{E} \right)^2$	$n = \left( \frac{Z_{\alpha/2} \cdot \sqrt{p \cdot q}}{E} \right)^2$
População finita (<10000)	$n = \frac{N \cdot \delta^2 \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot (E)^2 + \delta^2 \cdot (Z_{\alpha/2})^2}$	$n = \frac{N \cdot p \cdot q \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot (E)^2 + p \cdot q \cdot (Z_{\alpha/2})^2}$

n – tamanho da amostra;  $Z_{\alpha/2}$  – valor crítico para o grau de confiança desejado, usualmente: 1,96 (95%);  $\delta$  – desvio padrão populacional da variável; E – erro padrão, usualmente:  $\pm 5\%$  da proporção dos casos (precisão absoluta), ou  $\pm 5\%$  da média ( $1,05 \times \text{média}$ ); N – tamanho da população (finita); p – proporção de resultados favoráveis da variável na população; q – proporção de resultados desfavoráveis na população ( $q=1-p$ ).

Referência:

Miot, H.A., 2011. Tamanho da amostra em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, 10(4), pp.275-278.

# Ferramentas calcular tamanho amostral

- Além dos artigos publicados, existem diversas ferramentas on-line

Fórmulas para cálculo do tamanho de amostras para descrição de variáveis quantitativas e qualitativas em uma população	
Passo 1	
Qual tamanho da população?	630
Passo 2	
Defina um nível de confiança	95,00%
Para variáveis quantitativas	
Desvio padrão	15
Erro tolerável (região de incerteza ao redor da média)	2
Resposta	
Tamanho de amostra para população maior de 10 mil e variável quantitativa	217
Tamanho de amostra para população menor de 10 mil e variável quantitativa	162
Para variáveis qualitativas	
Expectativa prévia do a frequência do parâmetro estudado	69,30%
Margem de erro tolerável ao redor da sua estimativa	5,00%
Resposta	
Tamanho de amostra para população maior de 10 mil e variável quantitativa	327
Tamanho de amostra para população menor de 10 mil e variável quantitativa	216
Referência: Miot, H.A., 2011. Tamanho da amostra em estudos clínicos e experimentais. Jornal Vascular Brasileiro, 10(4), pp.275-278.	



tamanho amostral calculadora

Todas

Imagens

Vídeos

Shopping

Aproximadamente 115.000 resultados (0,42 segundos)

# Quais parâmetros estão associados ao cálculo?

- Tamanho da amostra: número total de indivíduos que você amostrou para seu estudo.
- Tamanho da população: número total de indivíduos que pertencem a dada população.
- Nível de confiança (%): probabilidade de que sua amostra represente com precisão o parâmetro real.
  - Ou seja, se seu grau de confiança é de 95%, isso quer dizer que seu estudo tem 95% de chance de apresentar um resultado correto
- Margem de erro (%): amplitude do erro associado a estimativa

# O que você precisa saber

	Quanto maior	Quanto menor
Tamanho da população	<b>Menor precisão</b>	<b>Maior precisão</b>
Tamanho da Amostra	<b>Maior precisão</b>	<b>Menor precisão</b>
Nível de confiança	<b>Maior precisão</b>	<b>Menor precisão</b>
Margem de erro	<b>Menor precisão</b>	<b>Maior precisão</b>

# Exemplo clássico

- Em uma pesquisa de opinião de votos para eleição, é dito que
  - Candidato A tem 50% de intenções de voto, com margem de erro de 5%, e grau de confiança de 95%.
  - Isso quer dizer que existe 95% de chance de o candidato A ter de 45-55% dos votos.
- Note que em geral a TV só fala da margem de erro.

Caminho inverso....

# Erro padrão

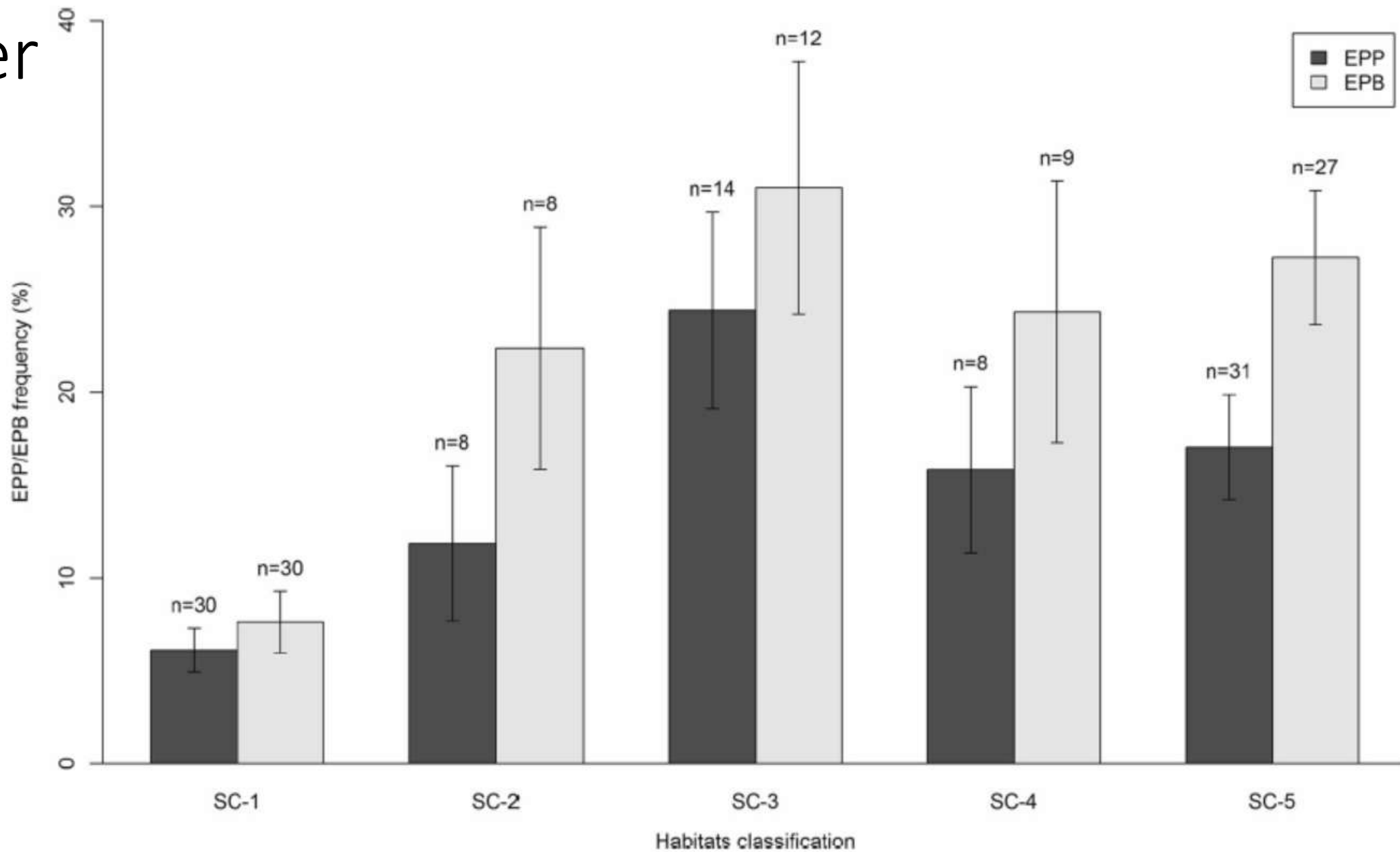
- Imagine que você não estimou um tamanho de amostra, mas foi para a rua e coletou dados de acordo como que dava para ser feito.
- Como estimar uma margem ao redor do seu dado? De forma a você saber a zona segura da estimativa do seu parâmetro?

# Intervalo de confiança

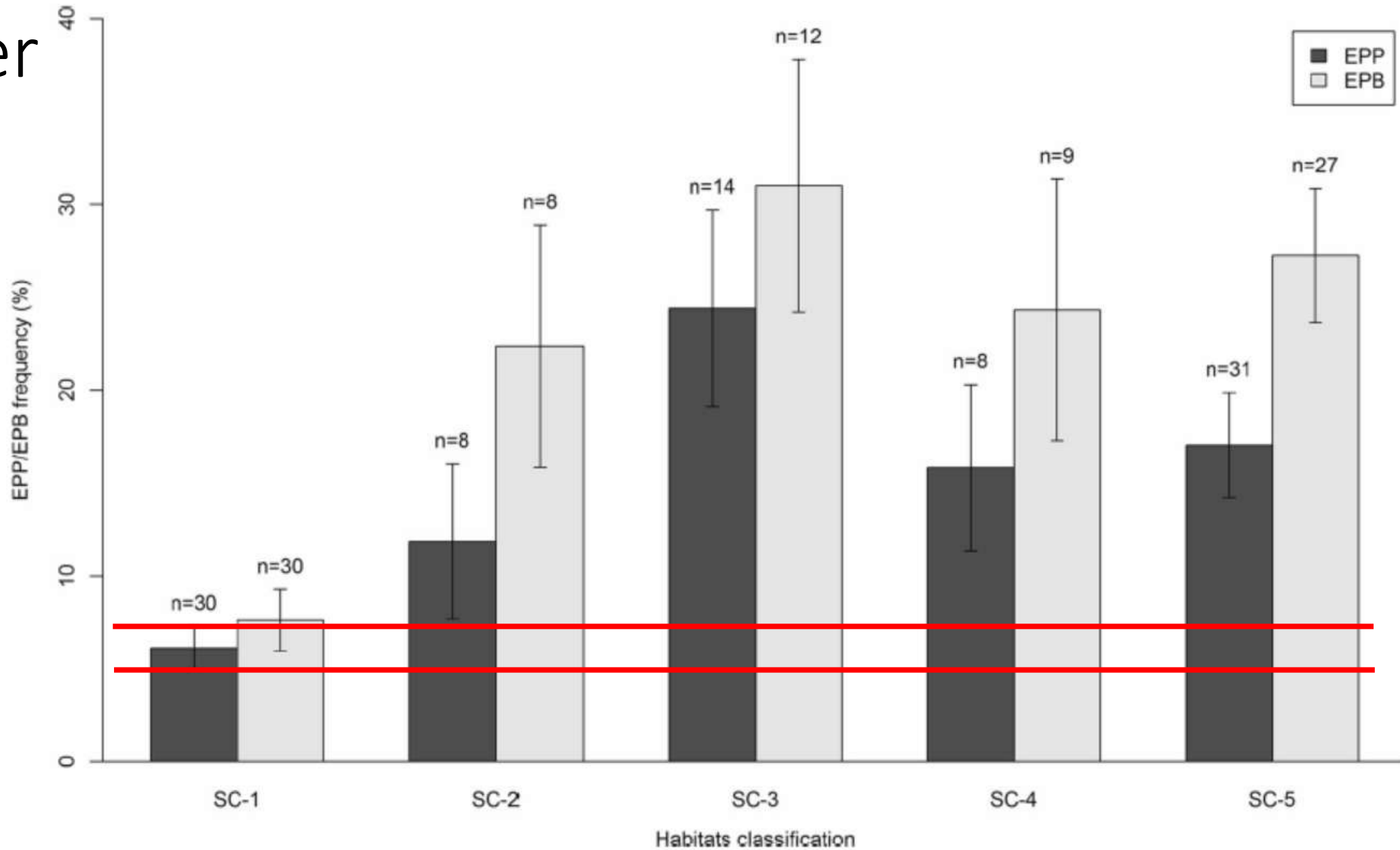
- Intervalo de confiança = intervalo ao redor média observada, onde há 95% de chance de a média populacional estar contida
- Estimou-se que a altura média de mulheres na cidade de São Paulo é de 1,60, com intervalo de confiança de 95% igual à  $\pm 10$ 
  - Isso quer dizer que temos 95% de certeza que a altura real é algum valor entre 1,50-1,70
- OBS: nos exemplos eu usei 95% de confiança, mas você pode escolher adotar outro valor



Spoiler



Spoiler



# Resumo de hoje

- Como resumir grandes conjuntos de dados em tabelas e gráficos
- Medidas de descrição
  - Média = média aritmética das observações
  - Mediana = Valor que ocupa a posição central do conjunto dos dados
  - Moda = Valor que aparece no seu conjunto de dados com mais vezes
  - Variância/Desvio = medidas associadas ao quanto os dados variam
- Descritores de erro associados a uma estimativa
  - Nível de confiança: probabilidade de que sua amostra represente com precisão o parâmetro real
  - Margem de erro: amplitude do erro associado a sua estimativa

Duvidas?

# Lista de exercício 4

- Veja o script “Aula 04 Descritores.R” para ver como calcular os parâmetros descritos no R
- Ficarei conectado no chat para tirar duvidas até às 23:00
- Links para responder a lista: <https://forms.gle/xxEF8xdjwzVPcSqE9>
  - Essa vai ser a lista mais longa, porém a mais fácil dessa disciplina