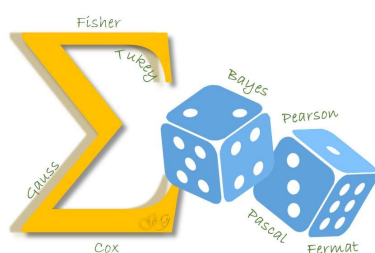




HEP 0170 - ESTATÍSTICAS DE SAÚDE

GLEICE M S CONCEIÇÃO
FSP USP



BIOESTATÍSTICA

ANÁLISE DESCRITIVA

GLEICE M S CONCEIÇÃO
FSP USP

ANÁLISE DESCRITIVA

Introdução - Tipos de Variáveis
(veja o vídeo no Moodle)

GLEICEM S CONCEIÇÃO
FSP - USP

Análise descritiva

A análise descritiva consiste em "resumir" o
conjunto de dados, por meio de

- ✓ tabelas
- ✓ gráficos
- ✓ medidas resumo numéricas

para que a informação contida nos dados seja
visualizada com facilidade.

GLEICEM S CONCEIÇÃO
FSP - USP

Análise descritiva



Essas ferramentas de visualização, se forem aplicadas adequadamente, revelam com detalhes o comportamento dos dados e as relações entre as variáveis de um modo que não seria possível apenas observando o conjunto de dados brutos.

GLEICEM S CONCEIÇÃO
FSP - USP

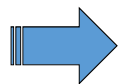
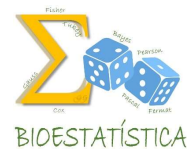
O que vamos aprender



- ✓ Construir e interpretar algumas medidas resumo numéricas e gráficos comumente utilizados em uma análise descritiva
- ✓ Entender quais medidas e gráficos são adequados a cada situação
- ✓ Tais medidas e gráficos devem possibilitar a apreensão da informação contida nos dados de forma simples e direta.

GLEICEM S CONCEIÇÃO
FSP - USP

Tipos de Variáveis

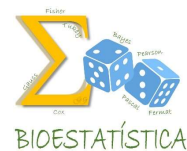


Para escolher a medida ou o gráfico mais adequado devemos levar em conta o tipo de variável que está sendo analisada.

- **variáveis qualitativas ou categóricas**
 - qualitativa nominal (sexo, tipo de patologia)
 - qualitativa ordinal (escolaridade)
- **variáveis quantitativas ou numéricas**
 - quantitativa discreta (número de filhos)
 - quantitativa contínua (peso, altura)

GLEICEM S CONCEIÇÃO
FSP - USP

O conjunto de dados "DC.pdf", disponível no Moodle, contém informações sobre a presença de doença coronariana, idade (anos), sexo, prática regular de exercícios (em horas/semana), obesidade, IMC (kg/m²) e pressão arterial sistólica (em mmHg) após teste ergométrico em 40 pacientes da Clínica TAL, na cidade de Araraquara, em 2015.



Paciente	Doença Coronariana	Idade	Sexo	Exercício	IMC	Obesidade	Pressão Sistólica
1	N	42.0	F	2	19.1	ausente	90.2
2	N	34.6	F	0	22.3	ausente	104.9
3	N	45.8	F	0	18.7	ausente	120.2
4	N	44.8	F	0	19.8	ausente	124.5
5	N	40.4	M	0	22.9	ausente	131.3
.
.
.
36	S	64.6	F	0	30.7	grau II	202.6
37	N	29.7	M	0	36.1	grau II	207.5
38	N	82.2	M	0	36.9	grau II	209.1
39	S	85.6	F	0	35.7	grau II	224.8
40	N	61.7	F	0	38.1	grau II	239.0

GLEICEM S CONCEIÇÃO
FSP - USP

Exercício 1



Utilizando o conjunto de dados "DC.pdf", classifique, quanto à natureza, cada uma das variáveis do estudo envolvendo os 40 pacientes da Clínica TAL.

Depois, confira a solução no próximo vídeo.

GLEICEM S CONCEIÇÃO
FSP - USP

Análise descritiva



1. Descrever uma variável de cada vez
2. Descrever duas variáveis conjuntamente

GLEICEM S CONCEIÇÃO
FSP - USP

1. Descrevendo / visualizando uma variável de cada vez

GLEICEM S CONCEIÇÃO
FSP - USP

Variáveis Qualitativas

Medidas resumo

- ✓ Tabela de frequência
contendo frequências e %

Gráficos

- ✓ Gráfico em barras
- ✓ Diagrama circular ou “pizza”

Variáveis Quantitativas

Medidas resumo

- ✓ Medidas de posição (média, mediana, moda)
- ✓ Medidas de dispersão
(variância, desvio-padrão, amplitude, quartis)

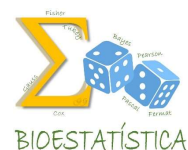
Gráficos

- ✓ Box plot
- ✓ Histograma
- ✓ Polígono de frequências
- ✓ Ogiva de frequências

Para as variáveis quantitativas
discretas que assumem poucos
valores:

- ✓ Além destas ferramentas,
podem ser utilizadas aquelas
descritas para variáveis as
qualitativas
- ✓ De fato, em algumas
situações, aquelas podem ser
mais interessantes

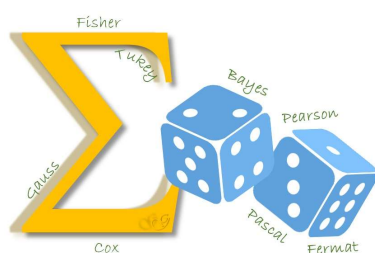
GLEICEM S CONCEIÇÃO
FSP - USP



ANÁLISE DESCRITIVA

Aprendendo a construir uma Tabela de Frequências
(veja o vídeo no Moodle)

GLEICE M S CONCEIÇÃO
FSP - USP



BIOESTATÍSTICA



ANÁLISE DESCRITIVA

GLEICE M S CONCEIÇÃO
FSP USP

Descrevendo / visualizando
uma variável de cada vez

Exercício 1 - solução

No conjunto de dados “DC.pdf”

- ✓ Doença coronariana – Qualitativa nominal
- ✓ Idade – Quantitativa contínua
- ✓ Sexo – Qualitativa nominal
- ✓ Exercício – Quantitativa discreta
- ✓ IMC – Quantitativa contínua
- ✓ Obesidade – Qualitativa ordinal
- ✓ Pressão sistólica – Quantitativa contínua

Variáveis Qualitativas



No conjunto de dados “DC.pdf”

- ✓ Doença coronariana
- ✓ Sexo
- ✓ Obesidade

GLEICEM S CONCEIÇÃO
FSP - USP

Variáveis Qualitativas



Medidas resumo

- ✓ Tabela de frequências

Gráficos

- ✓ Gráfico em barras
- ✓ Diagrama circular ou “pizza”

GLEICEM S CONCEIÇÃO
FSP - USP

Doença coronariana	Frequência n_i	Proporção p_i	Porcentagem $100 \times p_i$
Sim			
Não			
Total	40		

Doença coronariana	Frequência n_i	Proporção p_i	Porcentagem $100 \times p_i$
Sim	13		
Não	27		
Total	40		

Doença coronariana	Frequência n_i	Proporção p_i	Porcentagem $100 \times p_i$
Sim	13	0,325	
Não	27	0,675	
Total	40	1	

Distribuição dos pacientes segundo a presença de doença coronariana. Araraquara, 2015.

Doença coronariana	Frequência n_i	Proporção p_i	Porcentagem $100 \times p_i$
Sim	13	0,325	32,5%
Não	27	0,675	67,5%
Total	40	1	100%

Em um artigo científico...

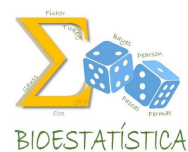


Tabela 1. Distribuição dos pacientes segundo a presença de doença. Araraquara, 2015.

Doença coronariana	n	%
Presente	13	32,5
Ausente	27	67,5
Total	40	100

GLEICEM S CONCEIÇÃO
FSP - USP

Exemplo



Tabela 1. Distribuição de crianças segundo níveis séricos de retinol. Cansação – Bahia, 1992.

Nível de retinol sérico ⁽¹⁾	N	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

⁽¹⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl
Fonte: Prado MS et al., 1995.

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Tabular



Exemplo

Tabela 1. Distribuição de crianças segundo níveis séricos de retinol. Cansação – Bahia, 1992.

Nível de retinol sérico ⁽¹⁾	N	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

⁽¹⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl

Fonte: Prado MS et al., 1995.

Elementos essenciais

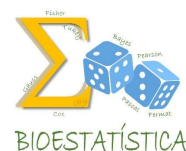
- ✓ Título autoexplicativo (o quê, quem, onde, quando), mas não repetitivo
- ✓ Cabeçalho
- ✓ Coluna indicadora
- ✓ Corpo
- ✓ Inclua as unidades de medida

Eventualmente, no rodapé

- ✓ Notas, chamadas
- ✓ Fonte

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Tabular



Exemplo

Tabela 1. Distribuição de crianças segundo níveis séricos de retinol. Cansação – Bahia, 1992.

Nível de retinol sérico ⁽¹⁾	N	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

⁽¹⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl

Fonte: Prado MS et al., 1995.

Bordas

- ✓ Para destacar a coluna indicadora, os totais e onde termina a tabela Não coloque mais bordas do que o necessário.
- ✓ Evite bordas nas laterais

No corpo da tabela

- ✓ Nenhuma casela (intersecção entre linha e coluna) deve ficar em branco
- ✓ 0 (zero) quando o valor numérico é nulo ou
- ✓ --- quando não se dispõe do dado
- ✓ Inclua totais de linhas e/ou colunas para facilitar as comparações.

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Tabular



Exemplo

Tabela 1. Distribuição de crianças segundo níveis séricos de retinol. Cansação – Bahia, 1992.

Nível de retinol sérico ⁽¹⁾	N	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

⁽¹⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl

Fonte: Prado MS et al., 1995.

Casas decimais

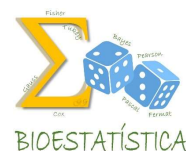
- ✓ A tabela deve ser uniforme quanto ao número de casas decimais;
- ✓ Não utilize mais casas decimais do que o necessário

Ao analisar uma tabela

- ✓ Evite ler o seu conteúdo, mas aponte os padrões e as principais tendências sugeridas pelos dados
- ✓ Resuma as similaridades e destaque as diferenças.

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Tabular



Elementos essenciais

- ✓ Título autoexplicativo (o quê, quem, onde, quando), mas não repetitivo
- ✓ Corpo
- ✓ Cabeçalho
- ✓ Coluna indicadora
- ✓ Inclua as unidades de medida

Eventualmente, no rodapé

- ✓ Fonte
- ✓ Notas, chamadas

Bordas

- ✓ Para destacar a coluna indicadora, os totais e onde termina a tabela Não coloque mais bordas do que o necessário.
- ✓ Evite bordas nas laterais

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Tabular



No corpo da tabela

- ✓ Nenhuma casela (intersecção entre linha e coluna) deve ficar em branco
 - 0 (zero) quando o valor numérico é nulo ou
 - --- quando não se dispõe do dado
- ✓ Inclua totais de linhas e/ou colunas para facilitar as comparações.

Casas decimais

- ✓ A tabela deve ser uniforme quanto ao número de casas decimais; não utilize mais casas decimais do que o necessário

Ao analisar uma tabela

- ✓ Evite ler o seu conteúdo, mas aponte os padrões e as principais tendências sugeridas pelos dados
- ✓ Resuma as similaridades e destaque as diferenças.

GLEICEM S CONCEIÇÃO
FSP - USP

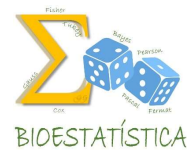
Exercício 2



- Utilizando o conjunto de dados "DC.pdf", construa uma tabela contendo a distribuição de frequências da variável "Sexo".
- Como você apresentaria as distribuições de frequências de ambas as variáveis, "Doença Coronariana" e "Sexo", em um artigo científico? Dê especial atenção aos cuidados com a apresentação tabular.
- Seria razoável apresentar a distribuição de frequências da variável "Obesidade"? Sim? Não? Por quê?

Depois, confira os resultados no Moodle.

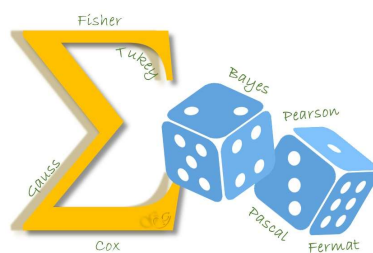
GLEICEM S CONCEIÇÃO
FSP - USP



ANÁLISE DESCRITIVA

Aprendendo a construir um
Gráfico em Barras

GLEICE M S CONCEIÇÃO
FSP - USP



BIOESTATÍSTICA



ANÁLISE DESCRITIVA

GLEICE M S CONCEIÇÃO
FSP USP

Variáveis Qualitativas



No conjunto de dados “DC.pdf”

- ✓ Doença coronariana
- ✓ Sexo
- ✓ Obesidade

GLEICEM S CONCEIÇÃO
FSP - USP

Variáveis Qualitativas



Medidas resumo

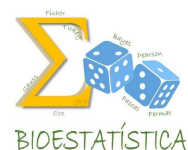
- ✓ Tabela de frequência

Gráficos

- ✓ Gráfico em barras
- ✓ Diagrama circular ou “pizza”

GLEICEM S CONCEIÇÃO
FSP - USP

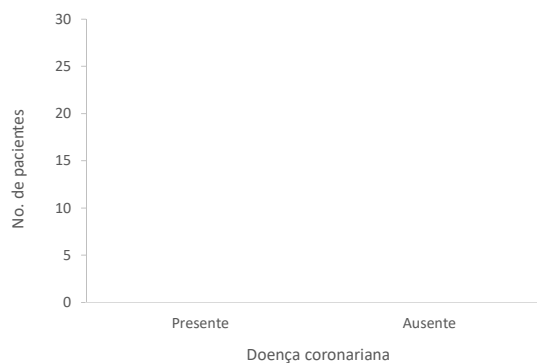
Distribuição dos pacientes segundo a presença de doença coronariana. Araraquara, 2015.



Doença coronariana	Frequência n_i	Proporção p_i	Porcentagem $100 \times p_i$
Sim	13	0,325	32,5%
Não	27	0,675	67,5%
Total	40	1	100%

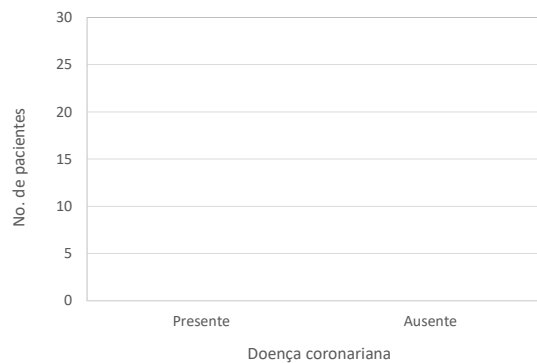
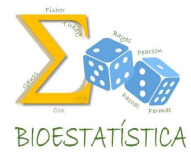
GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras



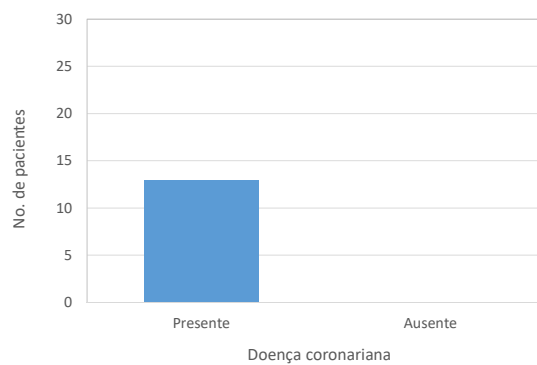
GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras



GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras



GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras

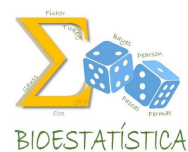
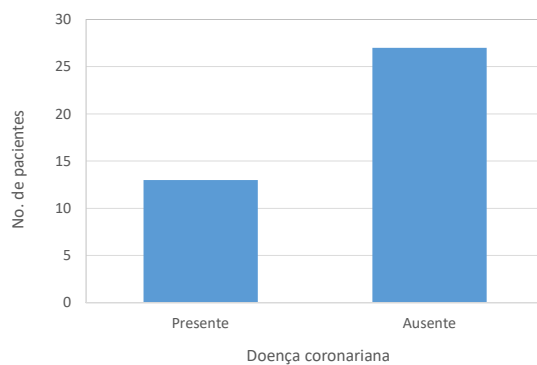


Gráfico 1. Distribuição dos pacientes segundo a presença de doença coronariana . Araraquara, 2015.

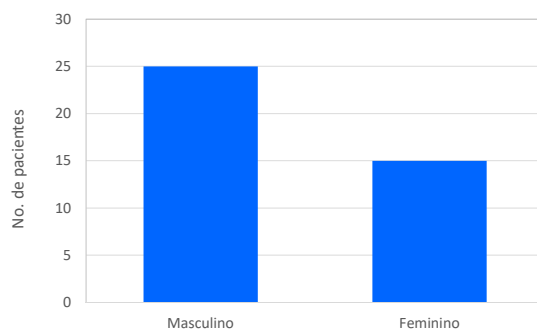


GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras



Gráfico 2. Distribuição dos pacientes segundo o sexo. Araraquara, 2015.



GLEICEM S CONCEIÇÃO
FSP - USP

Gráfico em barras



Gráfico 2. Distribuição dos pacientes segundo o sexo.
Araraquara, 2015.



GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Gráfica



Elementos essenciais

- ✓ Título do gráfico (o quê , como, onde, quando)
- ✓ Escala e numeração dos eixos
- ✓ Títulos dos eixos, rótulos das categorias
- ✓ Unidade de medida da variável

Eventualmente, no rodapé

- ✓ Notas, chamadas
- ✓ Fonte

Além disso

- ✓ Não utilizar mais casas decimais do que o necessário
- ✓ Não utilizar mais cores do que o necessário

GLEICEM S CONCEIÇÃO
FSP - USP

Apresentação Gráfica

Ao analisar um gráfico

- ✓ Evite ler o seu conteúdo, mas aponte os padrões e as principais tendências sugeridas pelos dados
- ✓ Resuma as similaridades e destaque as diferenças



GLEICEM S CONCEIÇÃO
FSP - USP

Exercício 3

Segundo o Ministério da Saúde, durante o mês de março de 2015, foram registradas 377 internações por dengue em hospitais do Sistema Único de Saúde (SUS) na região Norte, 1783 no Nordeste, 3600 no Sudeste, 371 no Sul e 1349 no Centro-Oeste.

- Construa uma tabela e um gráfico que permita visualizar essas informações adequadamente. Não se esqueça dos cuidados com a apresentação.
- Interprete a tabela e o gráfico que você construiu.

Depois, confira os resultados no Moodle.



GLEICEM S CONCEIÇÃO
FSP - USP

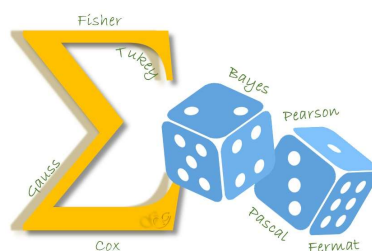


ANÁLISE DESCRITIVA

Apresentando o Diagrama Circular

GLEICE M S CONCEIÇÃO
FSP - USP

USP



BIOESTATÍSTICA



ANÁLISE DESCRITIVA

GLEICE M S CONCEIÇÃO
FSP USP

Variáveis Qualitativas

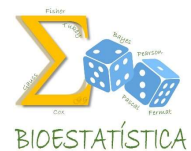


No conjunto de dados “DC.pdf”

- ✓ Doença coronariana
- ✓ Sexo
- ✓ Obesidade

GLEICEM S CONCEIÇÃO
FSP - USP

Variáveis Qualitativas



Medidas resumo

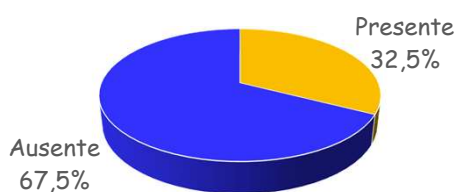
- ✓ Tabela de frequência

Gráficos

- ✓ Gráfico em barras
- ✓ Diagrama circular ou “pizza”

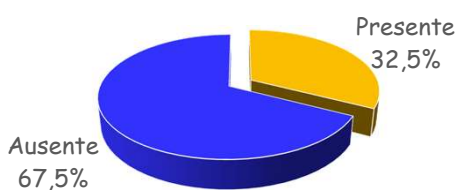
GLEICEM S CONCEIÇÃO
FSP - USP

Doença Coronariana



GLEICEM S CONCEIÇÃO
FSP - USP

Doença Coronariana



GLEICE M S CONCEIÇÃO
FSP - USP

Diagrama circular

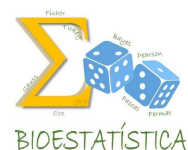
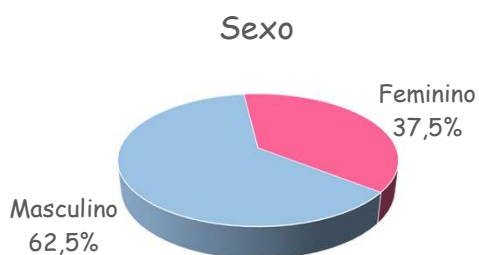
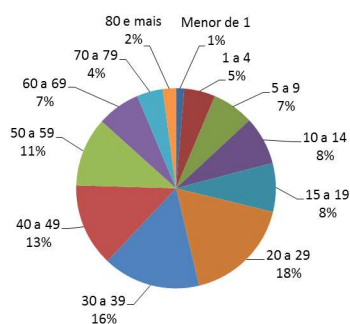
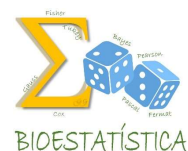


Gráfico 4. Distribuição percentual dos pacientes segundo sexo. Araraquara, 2015.



GLEICEM S CONCEIÇÃO
FSP - USP

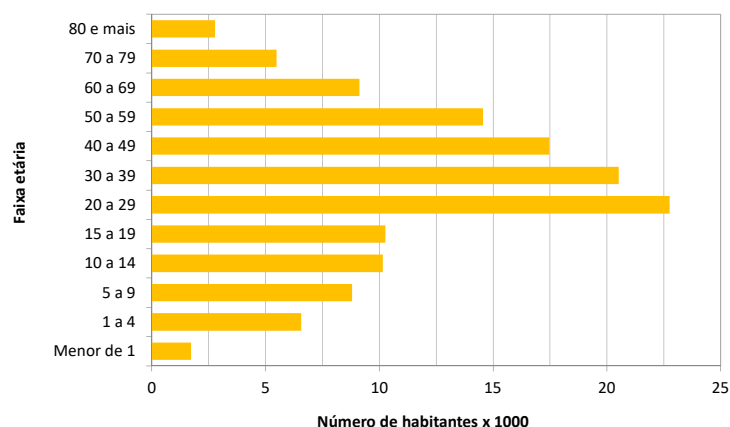
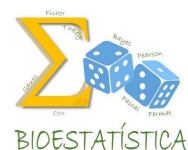
Distribuição percentual de habitantes segundo faixa etária.
Município de Botucatu, 2012.



Difícil de apreender a informação!

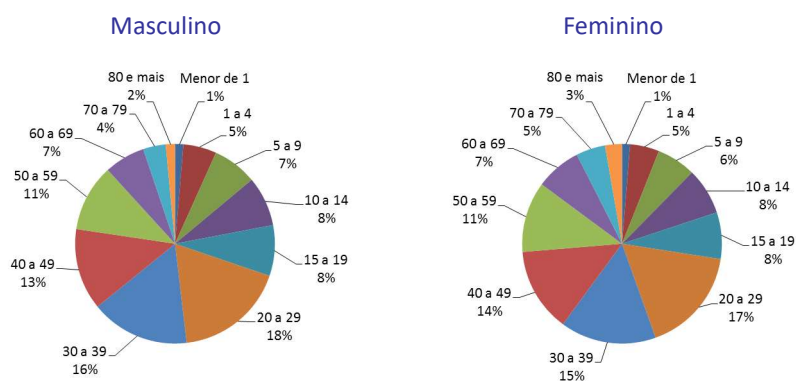
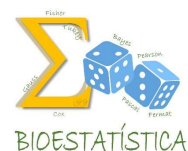
GLEICEM S CONCEIÇÃO
FSP - USP

Distribuição percentual de habitantes segundo faixa etária.
Município de Botucatu, 2012.



GLEICEM S CONCEIÇÃO
FSP - USP

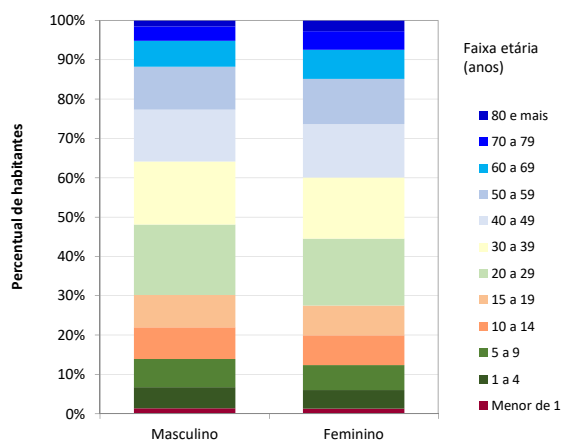
Distribuição percentual de habitantes segundo sexo e faixa etária.
Município de Botucatu, 2012.



Difícil fazer comparações!

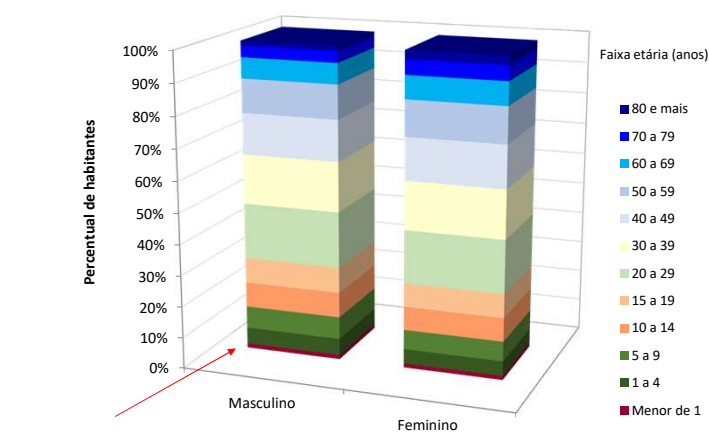
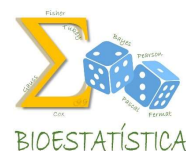
GLEICEM S CONCEIÇÃO
FSP - USP

Distribuição percentual de habitantes segundo sexo e faixa etária.
Município de Botucatu, 2012.



GLEICEM S CONCEIÇÃO
FSP - USP

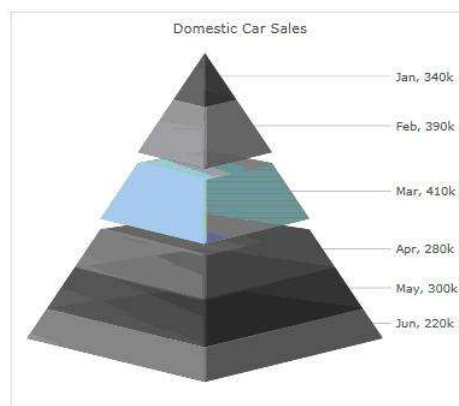
Distribuição percentual de habitantes segundo sexo e faixa etária.
Município de Botucatu, 2012.



Cuidado com distorções

GLEICEM S CONCEIÇÃO
FSP - USP

Cuidado!



GLEICEM S CONCEIÇÃO
FSP - USP



GLEICEM S CONCEIÇÃO
FSP - USP



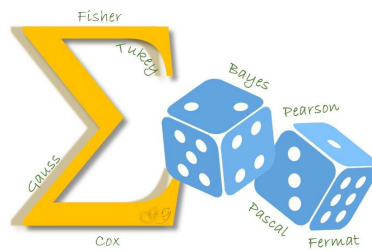
GLEICEM S CONCEIÇÃO
FSP - USP

ANÁLISE DESCRITIVA

Aprendendo a calcular

Medidas de Posição e Dispersão

GLEICEM S CONCEIÇÃO
FSP - USP



BIOESTATÍSTICA

ANÁLISE DESCRITIVA

GLEICE M S CONCEIÇÃO
FSP USP

Variáveis Quantitativas



BIOESTATÍSTICA

Medidas resumo

- ✓ Medidas de posição (média, mediana, moda)
- ✓ Medidas de dispersão (variância, desvio-padrão, amplitude, quartis)

Gráficos

- ✓ Box plot
- ✓ Histograma
- ✓ Polígono de frequências
- ✓ Ogiva de frequências

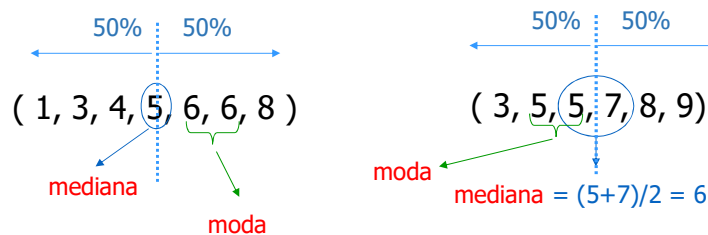
Para as variáveis quantitativas discretas que assumem poucos valores:

- ✓ Além destas ferramentas, podem ser utilizadas aquelas descritas para variáveis as qualitativas
- ✓ De fato, em algumas situações, aquelas podem ser mais interessantes

GLEICE M S CONCEIÇÃO
FSP - USP

BIOESTADÍSTICA

Moda: é a realização mais frequente do conjunto de valores observados




GLEICEM S CONCEIÇÃO
FSP - USP

mediana →	5.0	5.0
média →	5.2	7.0



- a média é afetada por valores extremos
- a mediana é “robusta”

GLEICEM S CONCEIÇÃO
FSP - USP



- GLEICEM S CONCEIÇÃO
FSP - USP

As empresas A e B têm programas de estágios e oferecem uma remuneração por 20 horas semanais com as seguintes características (em salários mínimos):

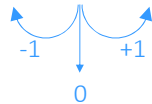
	A	B
Média	2.5	2.0
Mediana	1.7	1.9
Moda	1.5	1.9

GLEICEM S CONCEIÇÃO
FSP - USP

Medidas de dispersão



$$A = (4, 5, 6)$$

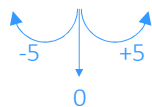


$$\text{média}_A = \bar{X}_A = 5$$

$$\text{variância}_A = S_A^2 = \frac{(-1)^2 + 0^2 + 1^2}{2} = \frac{2}{2} = 1$$

$$\text{desvio padrão}_A = \sqrt{S_A^2} = S_A = 1$$

$$B = (0, 5, 10)$$



$$\text{média}_B = \bar{X}_B = 5$$

$$\text{variância}_B = S_B^2 = \frac{(-5)^2 + 0^2 + 5^2}{2} = \frac{50}{2} = 25$$

$$\text{desvio padrão}_B = \sqrt{S_B^2} = S_B = 5$$

GLEICEM S CONCEIÇÃO
FSP - USP

Medidas de dispersão



Definição formal

- ✓ X – variável aleatória de interesse
- ✓ Seja x_1, x_2, \dots, x_n uma amostra da variável X .
- ✓ A **variância amostral** será dada por:

$$S_A^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$



Variância:

- Medida da dispersão dos dados em torno da média
- É a média do quadrado das distâncias em relação à média

GLEICEM S CONCEIÇÃO
FSP - USP

Medidas de dispersão



Definição formal

- ✓ X – variável aleatória de interesse
- ✓ Seja x_1, x_2, \dots, x_n uma amostra da variável X .
- ✓ O **desvio padrão amostral** será dado por:

$$S_A = \sqrt{S_A^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}}$$



Desvio padrão:

- Medida da dispersão dos dados em torno da média, expressa na mesma unidade dos valores do conjunto de dados
- É distância média das observações em relação à média

GLEICEM S CONCEIÇÃO
FSP - USP

Exercício 5



Utilizando o banco de dados DC.PDF:

- Considere apenas os pacientes com obesidade ausente e calcule medidas descritivas (média, mediana, desvio padrão, valores mínimo, máximo e número de observações) para a pressão sistólica. Interprete as medidas que você calculou.
- Faça o mesmo para as demais categorias de obesidade.
- Construa uma única tabela com todas essas informações. Siga as normas de apresentação tabular.

Depois, confira os resultados no Moodle.

GLEICEM S CONCEIÇÃO
FSP - USP

Exercício 6



Utilizando o banco de dados DC.PDF:

a) Construa uma tabela contendo a média e o desvio padrão para uma das variáveis abaixo, segundo o grau de obesidade.

- ✓ Pressão sistólica
- ✓ Idade
- ✓ Prática regular de exercícios

b) Interprete.

Depois, confira os resultados no Moodle.

GLEICEM S CONCEIÇÃO
FSP - USP