

Centro Universitário
Christus- UNICHRISTUS

Estatística Computacional

Especialização em Ciência
de Dados e Inteligência
de Negócios (Big Data e
BI)

Prof. Dr. Manoel Ribeiro

Apresentação Pessoal



1990 – Graduação em Computação de Computação na UFC



2005 – Mestrado em Computação UECE



2018 - Doutorado em Computação UFC



1999 – Empreendedor, sócio em empresa de base tecnológica



2010 – Fundação do GGTIC-CE



1990 a 2007 – Grupo J.Macedo



2007 a 2012 – Grupo Marquise

Ementa da Disciplina

- Estatísticas Descritivas
 - Medidas de tendência central;
 - Medidas de variabilidade;
 - Medidas de posição relativa;
 - Gráficos exploratórios de dados.
- Fundamentos de probabilidade.
- Distribuições discretas de probabilidade.
- Distribuições contínuas de probabilidade.
- Teste de Hipóteses
- Correlação

Bibliografia

- BARBETTA, P. A. Estatística Aplicada às Ciências Sociais. UFSC, 6a. ed. 2006.
- WHEELAN, C. Estatística, o que é, para que serve, como funciona. Zahar, 2016.
- TAVARES, Marcelo. Estatística aplicada à administração. Sistema Universidade Aberta do Brasil, 2007.
- Tsitsiklis, J. Probabilistic Systems Analysis and Applied Probability. Disponível em: <https://www.youtube.com/watch?v=j9WZyLZCBzs>.

Bibliografia Complementar

- KAHNEMAN, D. Rápido e Devagar - Duas Formas de Pensar. Objetiva, 1a. ed. 2012.
- SILVER, N. Sinal e Ruído. Intrínseca, 1ª ed. 2013.
- TALEB, N. N. Iludido pelo Acaso – A influência oculta da sorte nos mercados e na vida. Record, 2004.
- TALEB, N. N. A lógica do cisne negro. Best Seller, 15ª ed. 2008.
- DARRELL, H. Como mentir com estatística. Intrínseca, 2016.

Resultados / Impactos da Disciplina

- Aquisição de domínio conceitual e prático
 - Sobre os conceitos básicos da estatística descritiva.
- Aprender conceitos básicos de Linguagem R
 - O mínimo necessário e suficiente para realizar os cálculos estatísticos.

Metodologia

Aulas expositivas com discussões.

Práticas em laboratório.

Leituras.

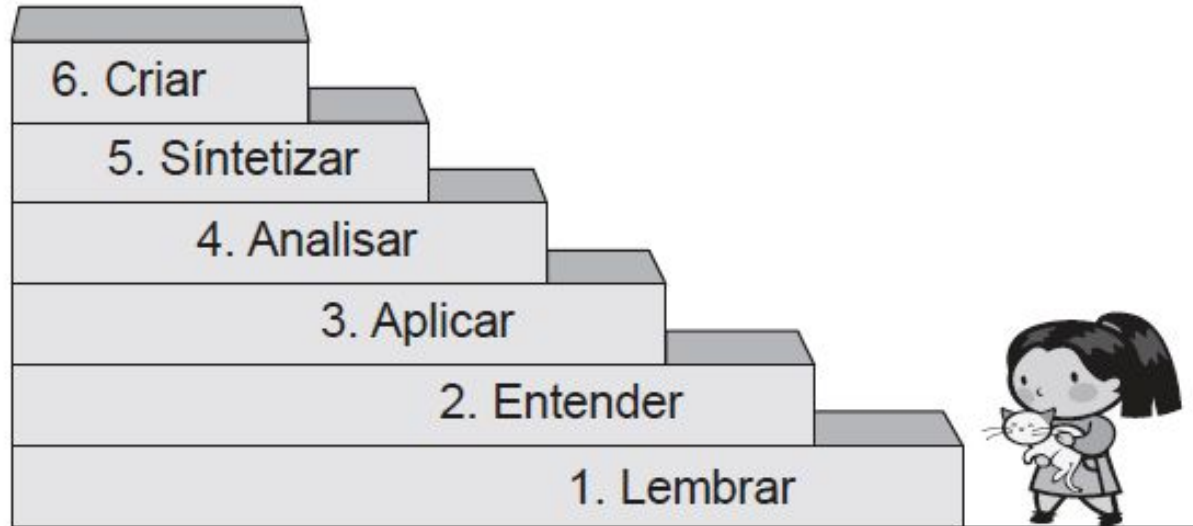
Tarefas individuais.

Avaliação (em dupla).

Método Pedagógico

● Taxonomia de Bloom

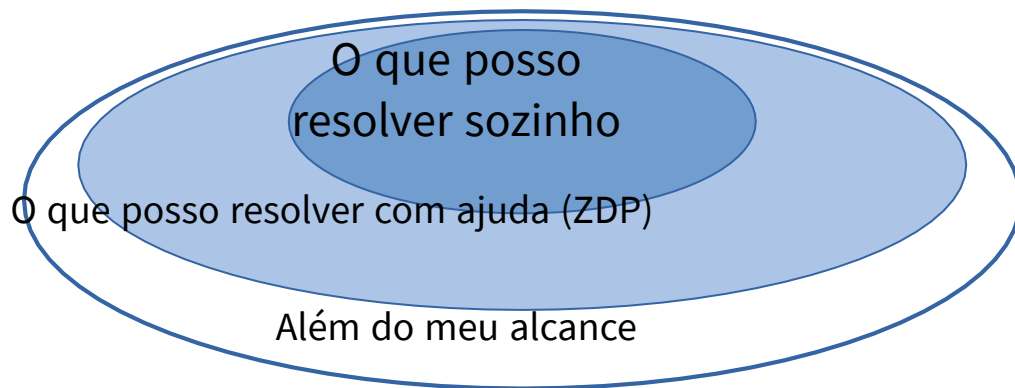
- para adquirir uma nova habilidade pertencente ao próximo nível, deve-se ter dominado e adquirido a habilidade do nível anterior.



Processo Cognitivo

● Zona de Desenvolvimento Proximal - ZDP (Vygotsky [1])

- a distância entre o nível de desenvolvimento real,
 - determinado pela capacidade de resolver tarefas de forma independente,
- e o nível de desenvolvimento potencial,
 - determinado por desempenhos possíveis, com ajuda de adultos ou de colegas mais avançados ou mais experientes.



Planejamento das aulas

- Dia 1 (sexta 18:00 às 22:00h)
 - Apresentação
 - Aula motivacional - Aplicação da Estatística na Análise de Dados?
- Dia 2 (sábado 8:00 às 18:00)
 - Fundamentos de Estatística Aplicada
 - Métodos e conceitos
 - Frequências e medidas
 - Propriedades e distribuição
 - Teste de hipóteses

Conteúdo da disciplina

- Dia 3 (sexta 18:00 às 22:00h)
 - Linguagem R
 - RStudio
- Dia 4 (sábado 8:00 às 18:00)
 - Estudos de caso com R
 - Avaliação

Repositório

<https://github.com/antoniomralmeida/EC>

Pré-requisitos da disciplina

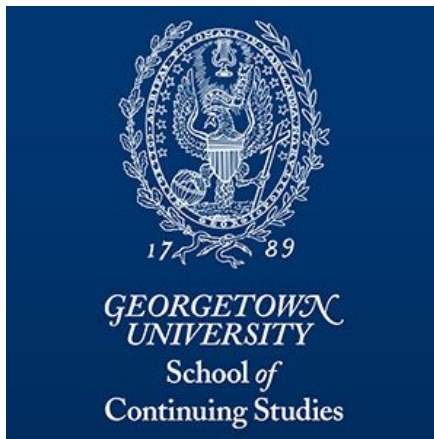
- Pré-requisitos da disciplina
 - Teoria dos Conjuntos
 - Matemática Aplicada
 - Linguagem de Programação

Certificações em Data Science

Northwestern

**SCHOOL OF
PROFESSIONAL STUDIES**

Advanced Data Science
US\$ 3.015, por disciplina, 1
ano, online

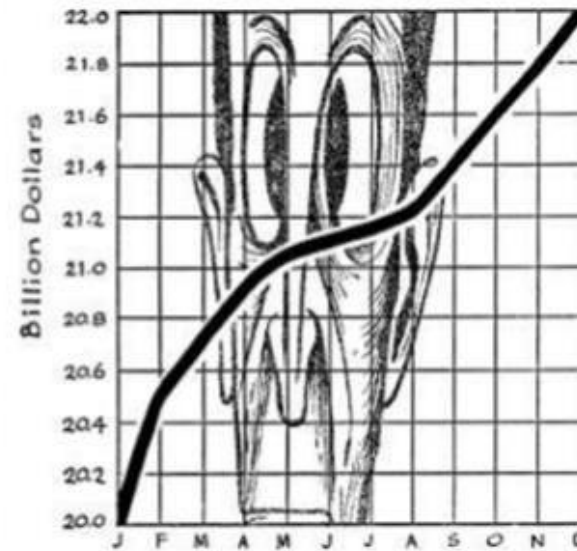
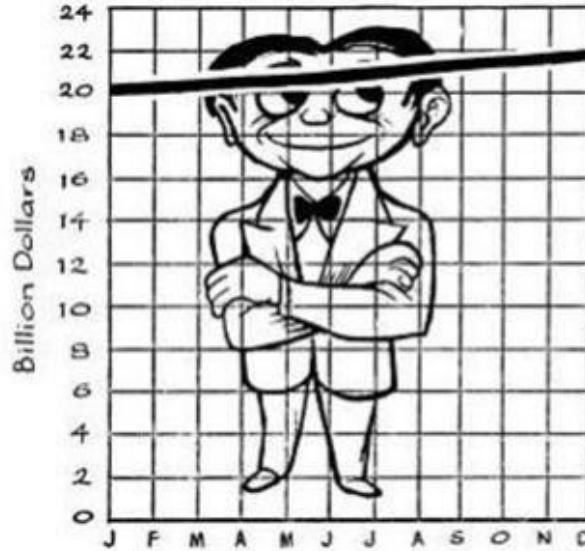


US\$ 7.496, presencial, 8
meses

Contextualização



Contextualização



“É fácil mentir com estatísticas, mas é difícil dizer a verdade sem elas”

(Andrejs Dunkels / Matemático / 1939-1998)

Por que estamos aqui ?

Você (aluno) e eu (professor), por que
estamos aqui ?

Mais dados

- Maior quantidade de informação
 - Aumentando a cada dia a proporção entre informação digital e não digital



Como as robôs Alice, Sofia e Monica ajudam o TCU a caçar irregularidades em licitações



Eu sou Victor, o robô!

Estou sendo criado para trabalhar no STF. Vou ajudar ministros, assessores e estagiários. Minha especialidade serão **minutas de votos!**

Ride the Big Data Wave

90%

of the world's existing data has been created in the last

2 YEARS

By 2020

35 Zettabytes worth of data will be created

*35 Zettabytes = 35,000,000,000,000,000,000 bytes

1/3 of that data will be stored in or will have passed through the cloud



10,000

Libraries of Congress worth of data are created annually by U.S. companies



80,000 km

is the distance past the moon the current global storage capacity for digital information stored on a CD stack would stretch



Every 2 days as much information is created as from the dawn of civilization up until 2003



There are actually more pieces of digital information than stars in the universe

EVERY DAY

1 Billion



pieces of content are shared via Facebook's Open Graph

2.5 Quintillion



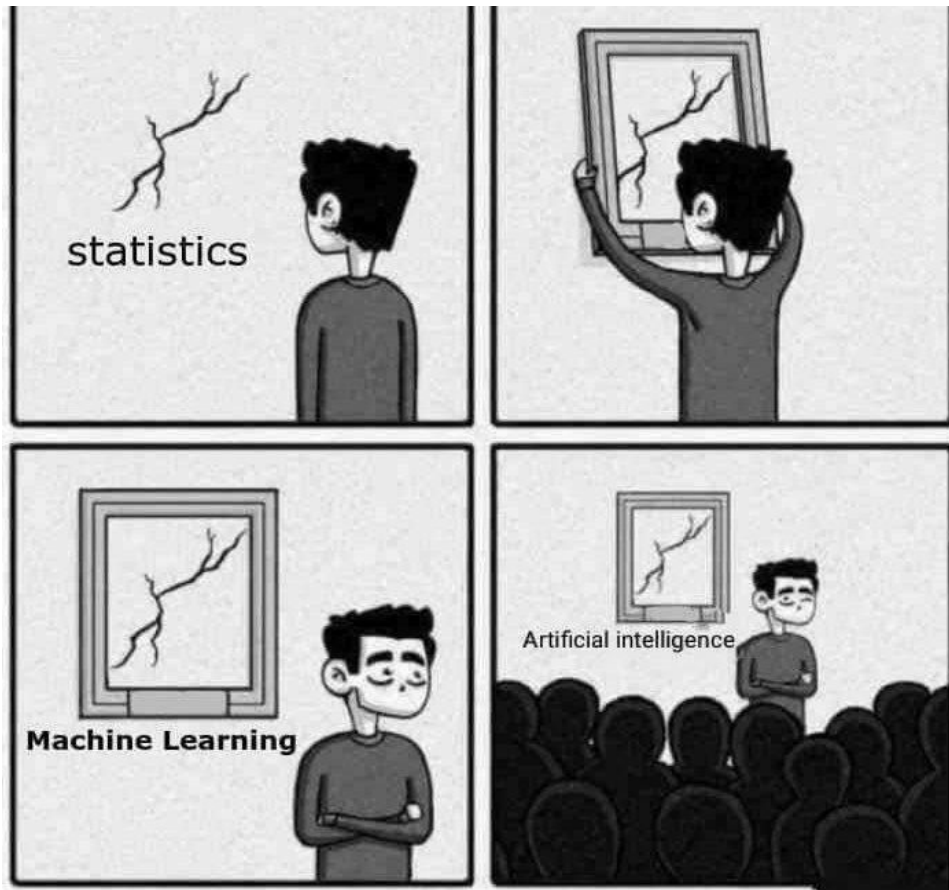
bytes of data are generated by people

8 Years



of video is uploaded to YouTube

A Estatística é a base de outras ferramentas de análise de dados



Estatística - Definições

- Estatística

- É a ciência de coletar, organizar, apresentar, analisar e interpretar dados para auxiliar na tomada de decisão efetiva.

- Análise Estatística

- é usada para manipular, resumir e investigar dados que resultem em informação útil para a tomada de decisão.

Estatística - Definições

- Estatística Descritiva

- Método de organizar, resumir e apresentar dados de uma maneira informativa

- Estatística Inferencial

- Os métodos usados para determinar alguma coisa sobre uma população baseado numa amostra

- População – Todo o conjunto de indivíduos (ou objetos de interesse) ou medidas obtidas de todos os indivíduos (ou objetos de interesse)

- Amostra – Uma porção, ou parte, de uma população de interesse

Objetivos da Pesquisa

- Os objetivos da pesquisa devem ser estabelecidos de forma clara.
 - *Se você não sabe aonde quer chegar, nenhum vento lhe é favorável (Sêneca)*
- Exemplo:
 - Conhecer o perfil de trabalho dos funcionários de determinada empresa.
 - Conhecer o tempo médio de serviço dos funcionários na Empresa;
 - Conhecer a distribuição do grau de instrução dos funcionários;
 - Verificar o interesse em programas de treinamento;
 - Avaliar o grau de satisfação dos funcionários com a Empresa;
 - Verificar se existe associação entre grau de satisfação e sua produtividade.

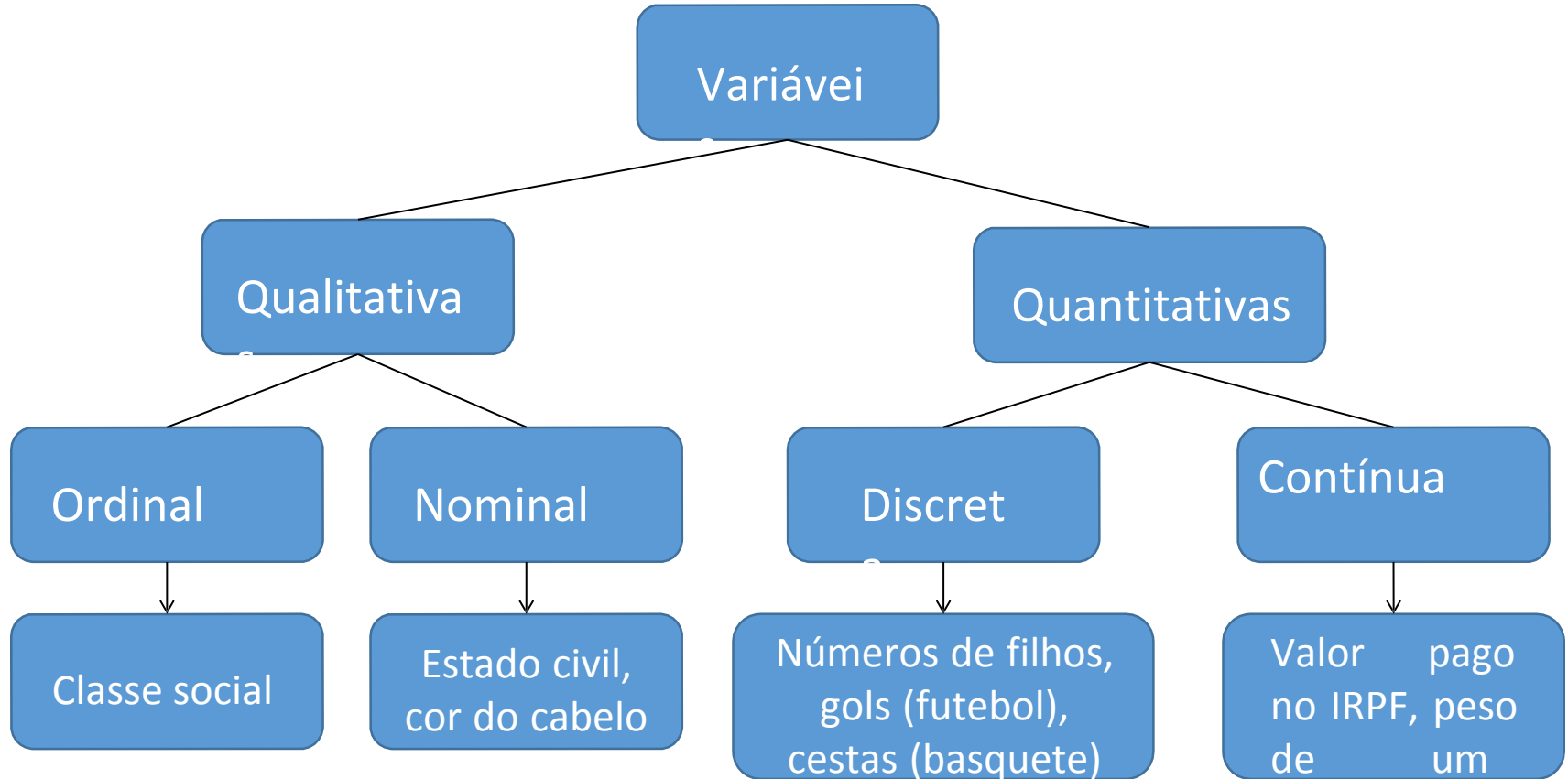
Objetivos da Pesquisa: Exemplo

- Os objetivos específicos devem fornecer uma indicação do que se precisa medir.
 - Tempo médio de serviço;
 - Grau de instrução;
 - Interesse em programas de treinamento;
 - Grau de satisfação dos funcionários com a Empresa; e
 - Produtividade.

Variáveis

- São características que podem ser observadas
- Quando coletar variáveis por meio de perguntas
 - Há quanto o Sr.(a) trabalha nessa empresa?
 - Qual seu estado civil?
- Elaborar perguntas que aceitam respostas precisas
 - Há quanto o Sr.(a) trabalha nessa _____ anos
empresa? completos
 - Qual seu estado civil? () solteiro () casado () viúvo ...
- Podem ser quantitativas ou qualitativas
(categóricas)

Variáveis



Coleta de dados: recomendações

- Identificar as características de interesse
- Revisão bibliográfica para
 - Verificar como mensurar adequadamente as características
 - Consultar pesquisas de referência (Ex.: IBGE)
- Definir como mensurar as variáveis
 - Unidades de medida (Km, Kg,... etc) ou categorias
- Elaborar uma ou mais perguntas para característica
 - Exemplo: Grau de satisfação com o trabalho
 - Satisfação com o salário, com a segurança do emprego, com a autonomia de trabalho, etc.

Coleta de dados: recomendações

- Usar uma linguagem suficientemente clara
 - Compreensível a todos os elementos da população
- Verificar se a pergunta induz uma resposta
 - 64 a 70% dos americanos responderam que concordam com a pena de morte por cometer assassinato
 - Quando puderam escolher, 48% preferiram a prisão perpétua [2]
- Verificar se a resposta é óbvia
 - Dependendo de como se pergunta sobre a satisfação com o salário, a resposta sempre será não.
- Quanto mais longo o questionário
 - menor a qualidade e confiabilidade das respostas

[2] - Fonte: Lydia Saad, "Americans hold firm support for death penalty", Gallup.com, 17 nov 2008.

População

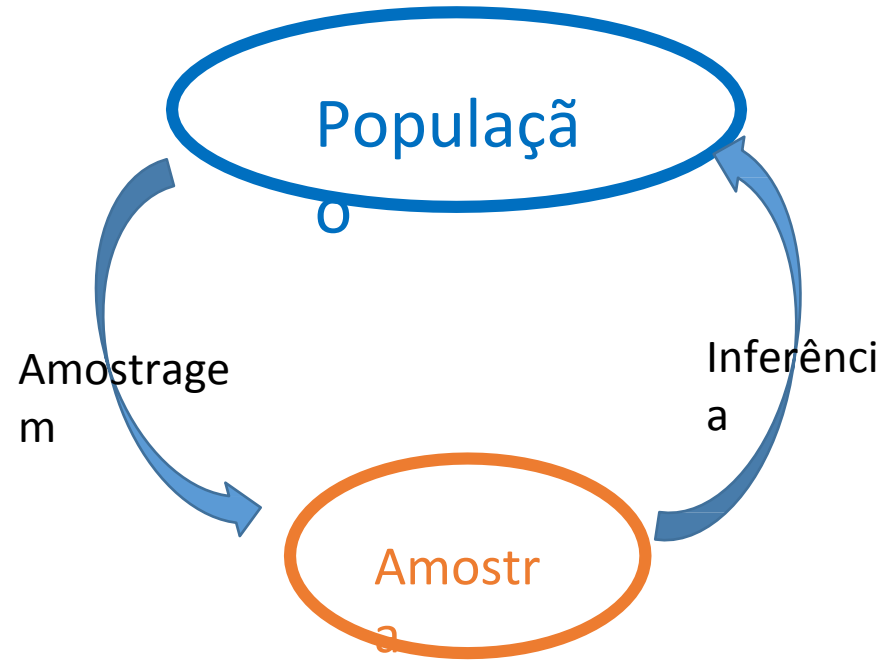
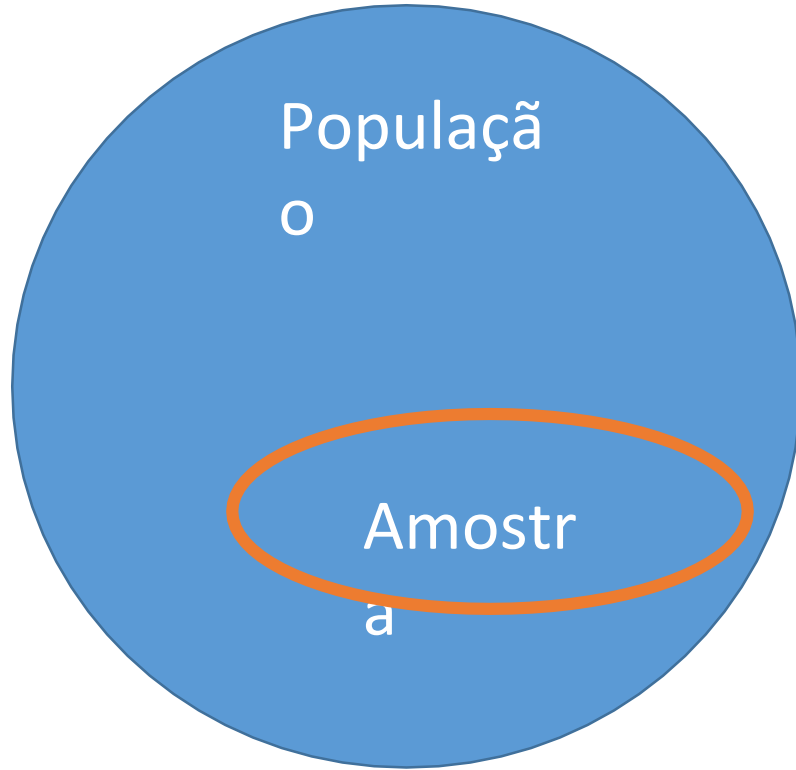
- População Alvo

- Conjunto de elementos que se quer abranger no estudo.
 - Exemplo: O conjunto de todos os indivíduos de uma Empresa, num determinado tempo.

- População Acessível (ou simplesmente População)

- Conjunto de elementos (indivíduos) observáveis
 - Exemplo: funcionários que não estão de férias nem licença
 - Veja que a variável tempo é relevante.

Amostragem



Amostragem

- Por que amostrar ?
 - Viabilizar o custo.
 - Entrevistar 1000 pessoas para fazer uma pesquisa eleitoral quinzenal com margem de erro de 5%.
 - Não consumir todo o estoque (experimentar uma sopa)
- Uma amostra deve ter as mesmas características da população subjacente (que está representando)
- Amostragem pode ser:
 - Com reposição: Um membro poderá ser escolhido mais de uma vez
 - Retirar bolas de uma urna (devolvendo-as)
 - Sem reposição: Um membro poderá ser escolhido apenas uma vez
 - Loteria, sorteio, bingo
- Útil para elaborar estimativas

Plano de Amostragem

- Deve conter a definição de:
 - Objetivos da pesquisa
 - População
 - Parâmetros (características da população). Ex: médias, proporção
 - Unidade de amostragem
 - Indivíduos, grupos, famílias, domicílios, países
 - Forma de seleção dos elementos
 - Aleatória simples,
 - Amostragem sistemática (coleta de dados ecológicos)
 - Amostragem estratificada (estratificação por características. Ex: renda
 - Amostragem de conglomerados (Ex.: bairros de uma cidade)

Unidade de análise

- A globalização está aumentando a desigualdade de renda?
 - De 1980 a 2000,
 - Países ricos ficaram mais ricos (Unidade de análise: países)
- Nos importamos com países pobres ou com pessoas pobres ?
 - E se a unidade de análise fosse pessoas ?
 - A china e a Índia detêm uma grande proporção de pessoas pobres
 - Eles eram relativamente pobres em 1980, e cresceram rapidamente
 - As Ilhas Maurício deveriam ter o mesmo peso que a china na análise?
- Segundo a revista The Economist
 - Se você considerar pessoas, e não países, a desigualdade está caindo.

Amostragem: outros conceitos

- Estimativa
 - Valor de um parâmetro referente a uma amostra.
- Erro amostral
 - Diferença entre a estimativa (amostra) e
 - o valor real (da população)
- Erro amostral tolerável
 - O quanto se admite errar.
 - É um requisito de projeto.
 - Exemplo: margem de erro de 2% numa pesquisa eleitoral.

Tamanho mínimo da amostra (método genérico)

- N : tamanho da população
- n : tamanho da amostra
- n_0 : uma primeira aproximação para o tamanho da amostra
- E_0 : erro amostral tolerável
- 1ª Aproximação (sem o tamanho da população) $n_0 = \frac{1}{E^2}$
- Refinamento (com o tamanho da população) $n = \frac{N \cdot n_0}{N + n_0}$

Exercício 1

- Nas eleições de 2019 tivemos 147,3 milhões de eleitores aptos a votar. Considerando um erro máximo tolerável de 2% qual deveria ser uma amostra mínima de uma pesquisa eleitoral para ser confiável?

Fontes de erro de amostragem

- População acessível diferente da população alvo
 - Consultar domicílios numa pesquisa eleitoral
 - E os eleitores com domicílio eleitoral em outro município ?
- Falta de respostas
 - Indivíduos da amostra podem se recusar a responder ou não serem encontrados
 - Empreender esforço encontrar e convencer os participantes.
- Respostas incorretas
 - Indivíduos podem exagerar a renda ou o número de vezes que fazem sexo por mês, não admitir que não votam
 - Pode-se evitar esse viés, perguntando se ele votou nas últimas eleições

A organização social da sexualidade: práticas sexuais nos EUA

- Conhecer quem faz o que, com quem e com que frequência
 - Para entender como o comportamento sexual nos EUA afeta a disseminação do HIV/AIDS
- Metodologia: entrevistas de 90 minutos
- Resultados:
 - Pessoas geralmente fazem sexo com outras parecidas com elas
 - 80% teve 1 ou nenhum parceiro sexual no ano anterior
 - 25% dos homens e 10% das mulheres relataram sexo extraconjugal
 - 5% dos homens e 4% das mulheres relataram atividade homossexual
- Crítica[3]: a pesquisa pressupõe uma amostragem representativa e que deram respostas acuradas

[3] - Presser, S. (1995). Sex, Samples, and Response Errors. *Contemporary Sociology*, 24(4), 296-298.

Exemplo de Amostragem: Pesquisa por telefone

- Escolher aleatoriamente um conjunto de troncos de telefonia fixa
 - Tronco: código de área + 3 primeiros dígitos do número
- A probabilidade de um tronco ser selecionado é
 - proporcional a sua participação no total de números telefônicos
 - Obtêm-se uma distribuição geográfica representativa da população
 - Assim, cada região do país está representada proporcionalmente
 - À sua participação em todos os números telefônicos
 - É um proxy (aproximação) da proporção do número de indivíduos da população

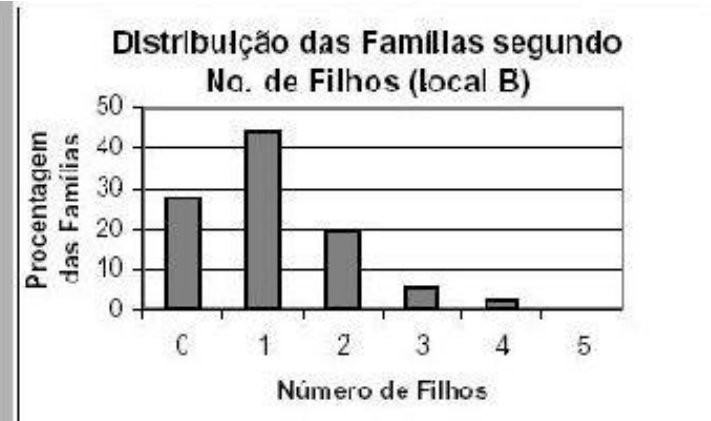
Exemplo de Amostr.: Pesquisa por telefone

- Os outros dígitos são escolhidos aleatoriamente
- Ao ligar na casa,
 - Um adulto é designado como respondente por um procedimento aleatório
 - Por exemplo, o adulto mais jovem presente na casa
- Ligar várias vezes durante o dia e à noite
 - Para conseguir alcançar a maior quantidade de respondentes
 - Importante para evitar um viés (pessoas que sempre estão em casa)
- A taxa de resposta é um indicador de validade da pesquisa
 - Baixa taxa indica um possível viés de amostragem
- A telefonia celular impôs novos desafios a pesquisa por telefone

Distribuição de Frequências

- Compreende a organização dos dados de acordo com as ocorrências dos diferentes resultados observados

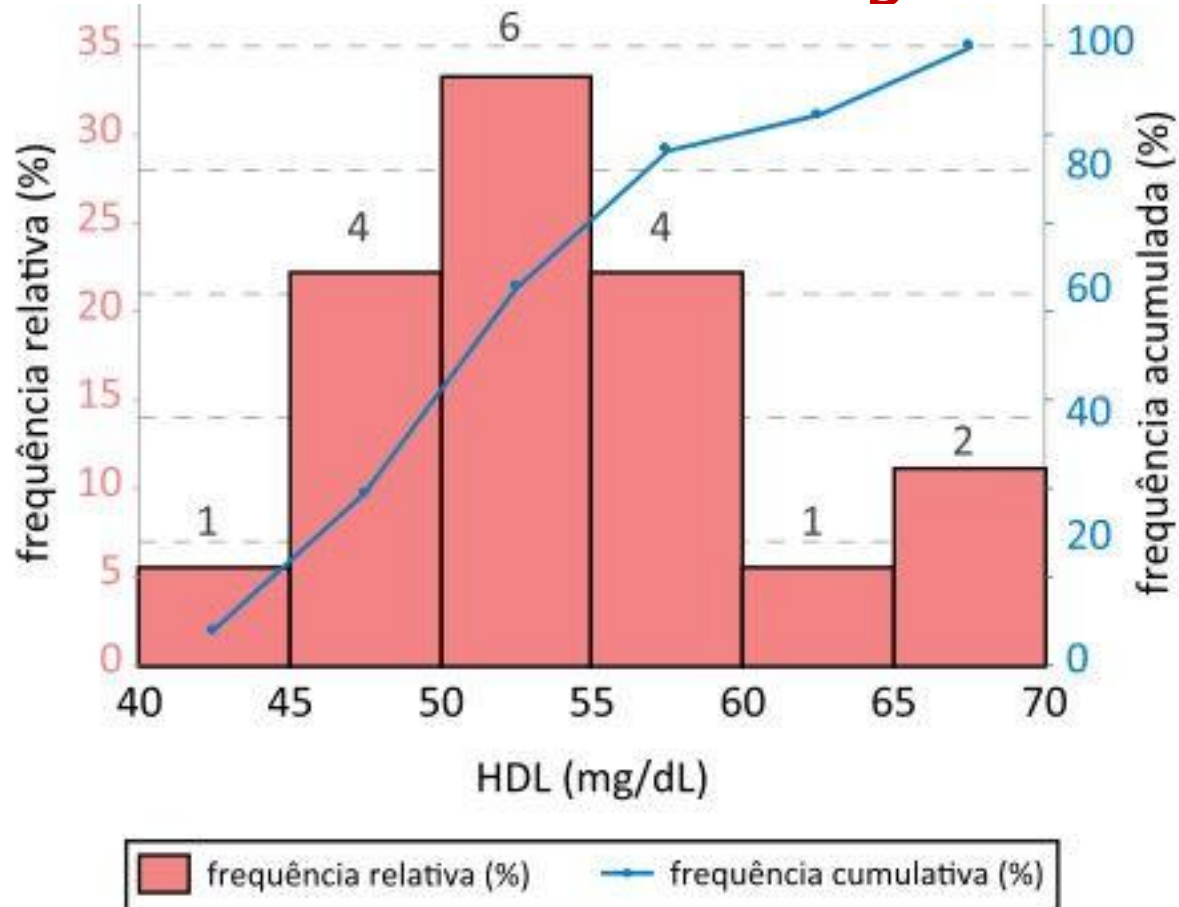
Número da classe	Salário do mês – R\$	Número de empregados
1	1 000 → 2 000	20
2	2 000 → 3 000	18
3	3 000 → 4 000	9
4	4 000 → 5 000	3



Distribuição de frequências (Variável

coleta aleatória		dados ordenados		intervalo	frequência	classe
pacientes	HDL (mg/dL)	pacientes	HDL (mg/dL)			
1	55	7	44	HDL < 45	1	1
2	57	8	45	45 ≤ HDL < 50	4	2
3	53	16	46			
4	49	14	47			
5	54	4	49			
6	52	9	50	50 ≤ HDL < 55	6	3
7	44	10	52			
8	45	6	52			
9	50	13	53			
10	52	3	53			
11	55	5	54	55 ≤ HDL < 60	4	4
12	67	1	55			
13	53	11	55			
14	47	2	57			
15	65	18	59	60 ≤ HDL < 65	1	5
16	46	17	64			
17	64	15	65	65 ≤ HDL	2	6
18	59	12	67			

Histograma



Método para variáveis discretas ou qualitativas

- O número de classes será igual ao conjunto de valores possíveis
- Obtém-se a FA pela contagem simples das ocorrências de cada valor
- Obtém-se a FR dividido-de a FA/n onde n é o número de observações

GÊNERO	FA	FR
Masculino	20	0,40
Feminino	30	0,60
Total	50	1,00

NÚMERO DE RECLAMAÇÕES POR DIA	NÚMERO DE DIAS (FA)	FREQ. RELATIVA
0	9	0.3
1	5	0.17
2	7	0.23
3	5	0.17
4	2	0.07
5	2	0.07
Total	30	1

Método para variáveis quantitativas contínuas

- Calcular o número de classes k

$$k = \sqrt{n}, \text{ para } n \leq 100$$

- Calc $k = 5 \log n, \text{ para } n > 100$ corresponde à diferença entre o maior valor e o menor valor observados.

- Calcular a amplitude do intervalo de classe (c)

- As classes são criadas iniciando com:

- Limite inferior = menor valor – $c/2$ (fechado)
- Limite superior = inferior + c (aberto)
- Ponto médio = (limite inferior + limite superior)/2

$$c = \frac{A}{k - 1}$$

Exercício 2

- Um levantamento revelou que em uma empresa com 100 funcionários os salários variam de R\$ 1000 a R\$15000. Calcule as classes do histograma que melhor representa este dados

Medidas de tendência central

(Introdução ao R)

Médias

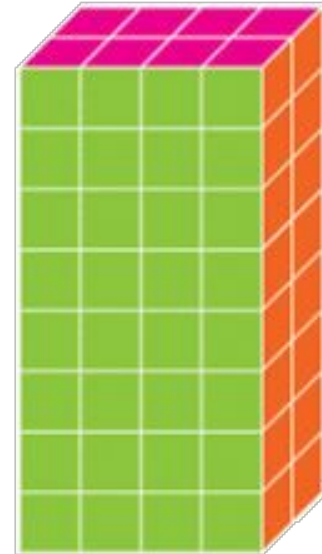
- Média aritmética

$$\mu = \frac{\sigma_1^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

- Média

geométrica

$$\left(\prod_{i=1}^n a_i \right)^{1/n} = \sqrt[n]{a_1 a_2 \cdots a_n}.$$



Mediana

- É o valor que divide uma distribuição ao meio.
 - Metade dos valores (os menores) fica de um lado e a metade (os maiores) fica de outro.
- Procedimento de cálculo
 - Ordena-se os valores, e escolhe-se o valor do centro
 - Qual é a mediana de: 1 1 1 4 20 680 2300
 - Com uma quantidade par de números, calcula-se a média dos dois números centrais
 - Qual é a mediana de: 1 1 1 4 20 680
- Consegue filtrar valores extremos (outliers)

Média vs Mediana: Exemplo fictício

- Imagine você num bar com mais 8 clientes presentes
 - Considere que a renda anual de cada um dos clientes seja esta:
 - 15 15 16 18 20 20 21 21 84
 - Média = 25.5
 - Mediana = 20
- O Bill Gates entra no bar (renda anual de 10 milhões)
 - 15 15 16 18 20 20 21 21 24 10×10^6
 - Média ~ 1.1 milhão
 - Mediana = 20
- Transmitir a mensagem sobre o ambiente,
 - dizer que no bar onde você toma cerveja a renda média anual dos frequentadores é
 - um pouco mais de 1 milhão ?

Média vs Mediana: Exemplo real

- A administração de George W. Bush anunciou um plano de cortes de impostos benéficos para a “maioria das famílias” americanas
 - O argumento a favor do plano era
 - 92 milhões de americanos receberiam uma redução tributária média de U\$1083
- Seria essa a melhor descrição dos benefícios ?
 - O corte mediano era menos de U\$100
 - Indivíduos extremamente ricos eram elegíveis para cortes muito grandes
 - Distorcendo a média

Moda

- É o valor mais frequente de uma distribuição de frequência
- Útil como tendência central para variáveis qualitativas
 - Sim, Sim, Sim, Sim, Não, Não, Não sei, Não sei

Média vs Mediana

- A mediana consegue filtrar valores extremos (outliers)
- Uma boa análise estatística
 - Apresenta as duas métricas
- Qual é a mais apropriada depende de se os valores extremos são outliers
 - Ou são parte da mensagem que você quer transmitir

Desvio padrão e Variância

- São medidas de dispersão (espalhamento)
 - Em relação ao valor médio
- São medidas quantitativas para expressar
 - o quanto os elementos distam da média
- Exemplo:
 - Peso médio dos passageiros de um avião que carrega competidores de uma maratona;
 - Peso médio de passageiros de um voo comercial comum.
 - Crianças, jovens, adultos
 - O peso pode ser parecido, mas a dispersão dos pesos em relação a média será parecida ?

Variância: Média do Desvio

Desvio
quadrático

Grupo 1	Altura ($\mu = 175$ cm)	Média = Valor absoluto de $(x_n - \mu)^*$	$(x_n - \mu)^2$
Nick	185	10	100
Elana	165	10	100
Dinah	170	5	25
Rebecca	173	2	4
Ben	183	8	64
Charu	175	0	0
		Total = 35	Total = 293
			Varlância = $293/6 = 48,83$
			Desvio padrão = $\sqrt{48,83} = 6,988 = 7$

Desvio padrão

- Variância

$$v = \frac{\sum (X_i - \mu)^2}{n - 1}$$

- Desvio padrão

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{n - 1}}$$

- Ou

$$\sigma = \sqrt{v} \quad \text{ou} \quad \sigma^2 = v$$

- Obs.: quando se trata de toda a população, alguns autores

- Usam N no lugar de n-1 da fórmula.

Os dois grupos têm a mesma média

Grupo 2	Altura ($\mu = 175$ cm)	Média = Valor absoluto de $(x_n - \mu)^*$	$(x_n - \mu)^2$
Sahar	163	12	144
Maggie	170,5	4,5	20,25
Faisal	174	1	1
Ted	175	0	0
Jeff	180,5	5,5	30,25
Narciso	187	12	144
		Total = 35	Total = 339,5
		Variancia = $339,5/6 = 56,583$	
		Desvio padrão = $\sqrt{56,583} = 7,522 = 7,5$	

Intepretação do desvio padrão

- Numa distribuição Normal (curva caracterizada por μ e σ)
 - 68,2% das medições estão dentro de 1 σ da média
 - 95,4% estão dentro de 2 σ
 - 99,7% estão dentro de 3 σ
- Útil quando você desconhece
 - Os valores envolvidos (contexto)



Quartis

- Divide uma distribuição em 4 partes iguais
 - Cada parte tem $\frac{1}{4}$ da amostra (ou da população)
- Como calcular os quartis
- $Q_{1/4}$ = arredondar $0.25 \cdot (N+1)$
- $Q_{2/4}$
 - Se N for par:
 - $Q_{2/4}$ = média dos itens na posição $(N/2)$ e $(N/2)+1$
 - Se N for ímpar:
 - $Q_{2/4}$ = o item na posição $(N+1)/2$
- $Q_{3/4}$ = arredondar $0.75 \cdot (N+1)$

Diagrama de Caixa (boxplot)

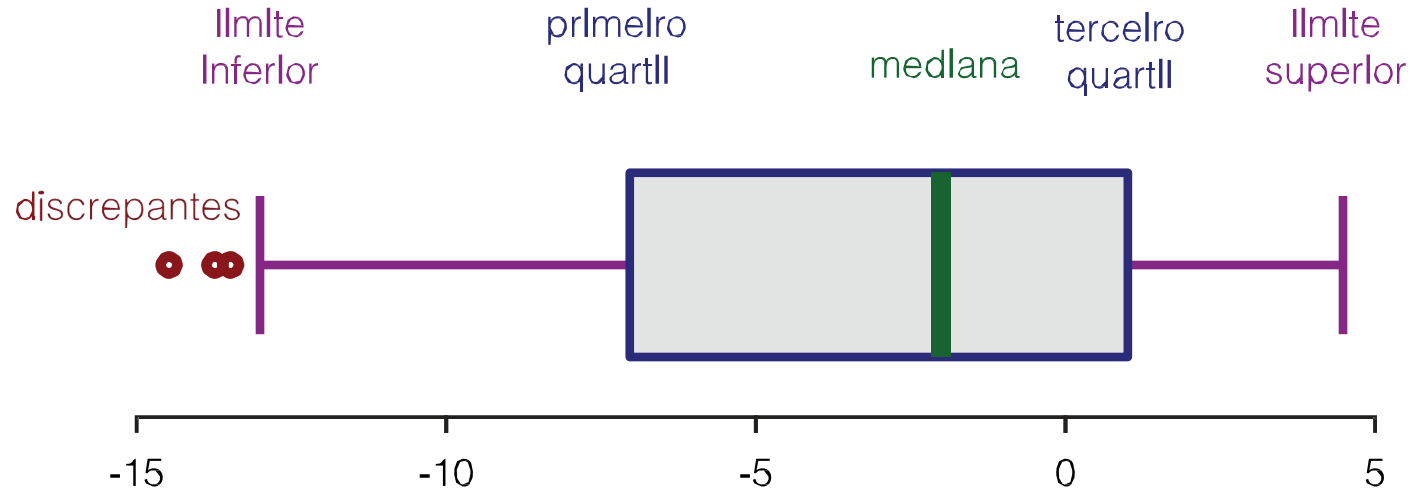
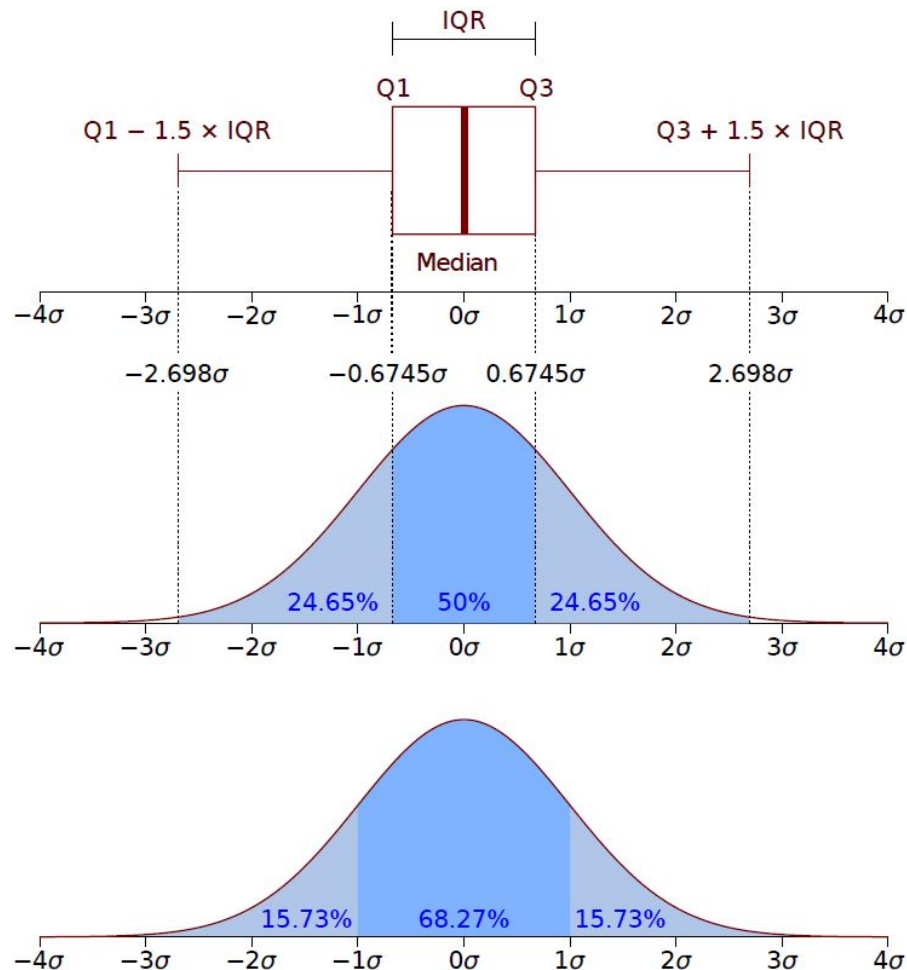


Diagrama de Caixa

- Distribuição Normal
- $N(0, 1\sigma^2)$

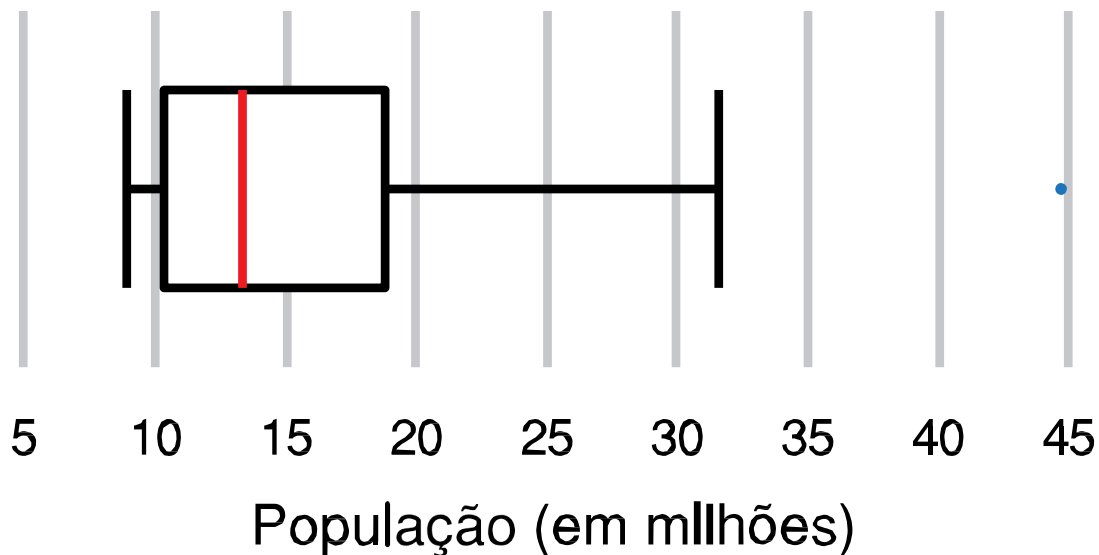
amplitude inter-quartis, $IQR = Q3 - Q1$



Exercício 3

- Represente através do diagrama de caixa (boxplot) a seguinte situação
- . O dono de um restaurante queria conhecer o perfil de deslocamento dos seus clientes e fez a pesquisa sobre a distância percorrida por cada um:
- . 14,6,3,2,4,15,11,8,1,7,2,1,2,4,10,22,20

População dos estados brasileiros segundo o IBGE



Fonte 1:

ftp://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2016/estimativa_dou_2016_20160913.pdf

Distribuição de renda de duas localidades

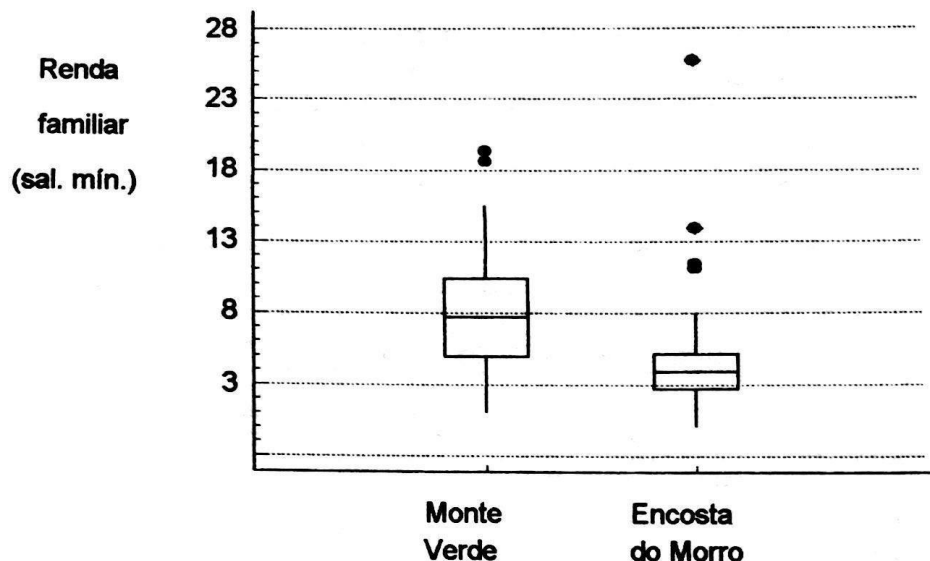
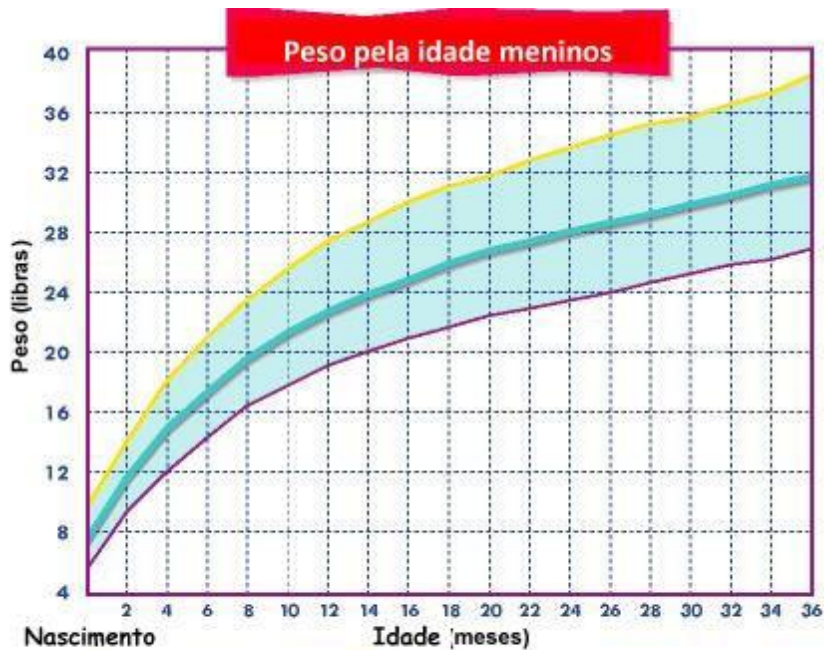


FIG. 6.10 Representação das distribuições de renda do Exemplo 6.4 em *diagramas em caixas*.

Percentil

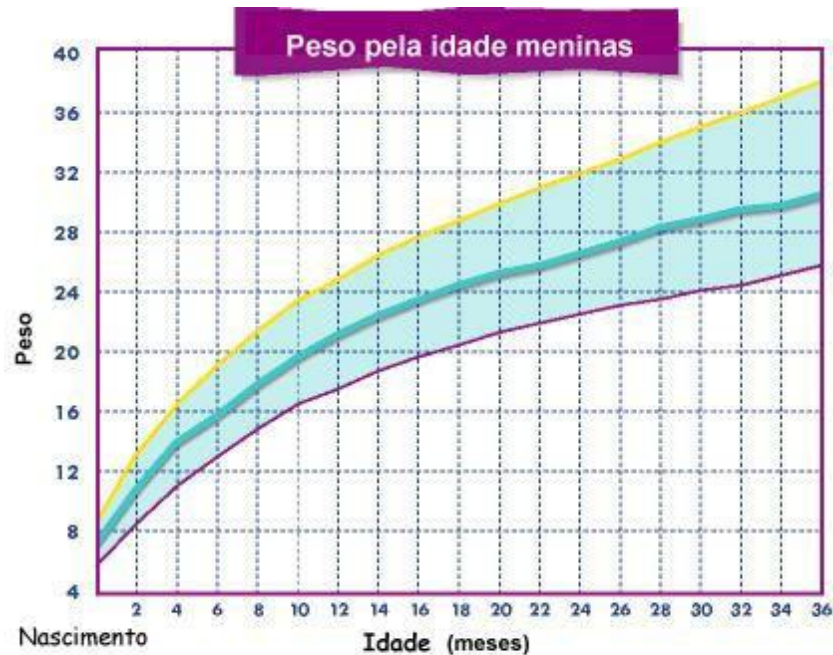
- Divide-se a distribuição em 100 partes



Percentil 95%

Percentil 50%

Percentil 5%



Percentil 95%

Percentil 50%

Percentil 5%

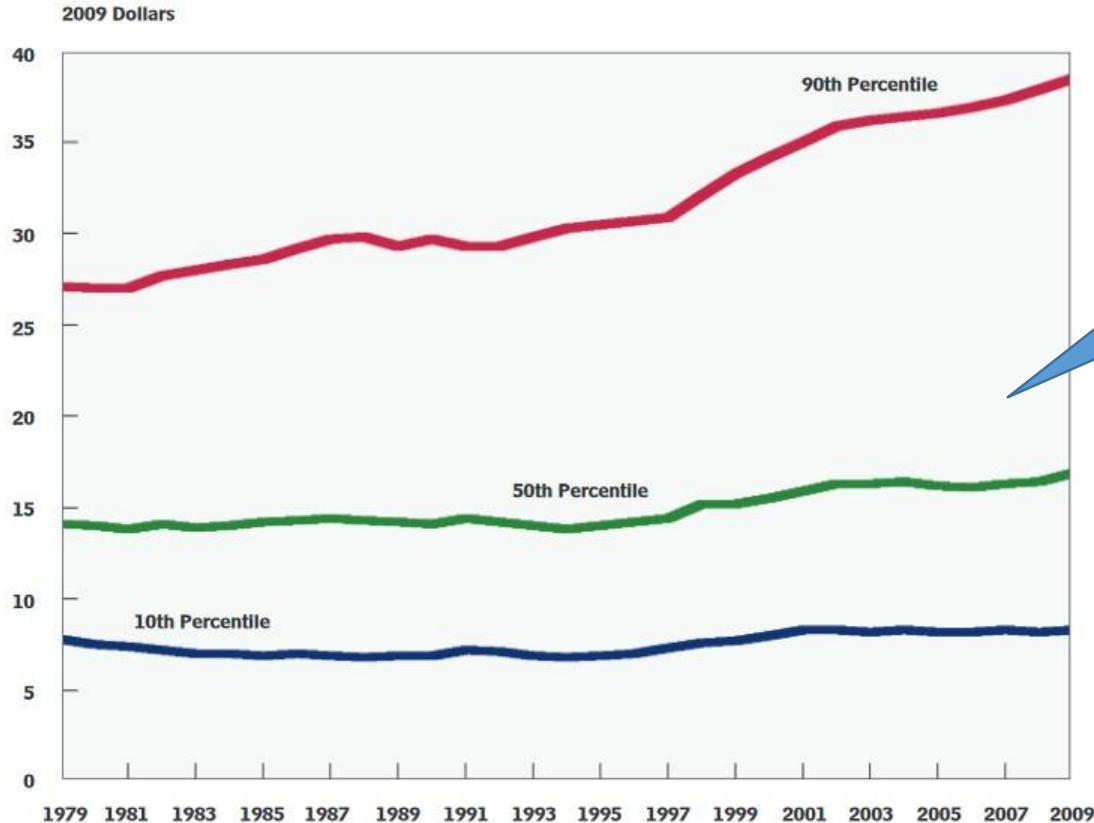


Examinar a saúde econômica da classe média americana

- Segundo Jeff Grogger
 - PhD em Economia e professor de Política Pública na Univ. de Chicago
- E Alan Krugger
 - Chefe do conselho de assessores econômicos do presidente Obama
- Duas boas medidas para avaliar a saúde econômica da classe média:
 - As mudanças no salário mediano (corrigido pela inflação) durante as últimas décadas; e
 - As mudanças nos salários no 25º e 75º percentis
 - Esses valores podem ser interpretados como os limites inferior e superior da classe média
- Renda é diferente de salário. Qual delas é mais apropriada ?

Examinar a saúde econômica da classe

média



7

Compare o desempenho do
50º percentil com o
desempenho do

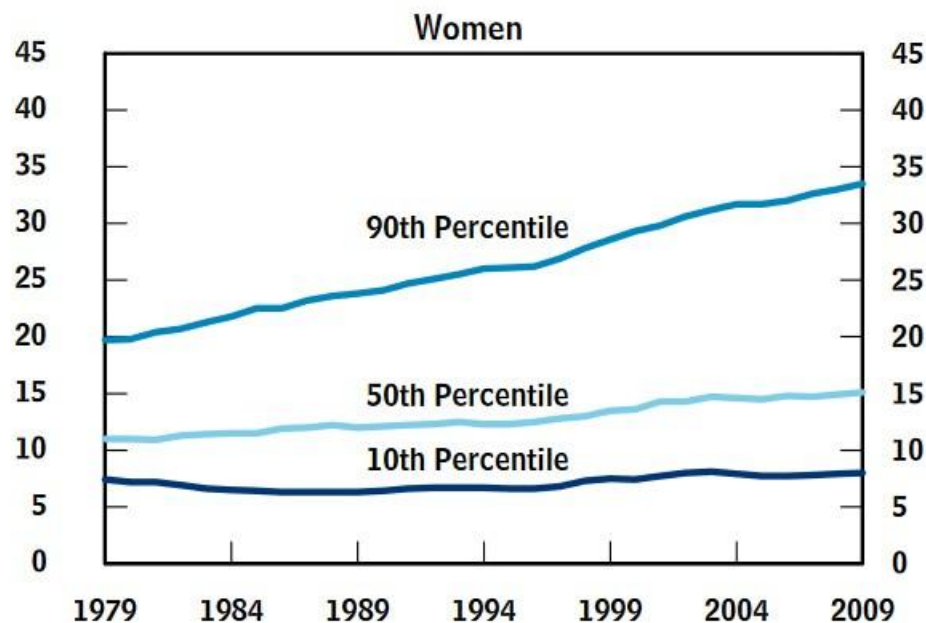
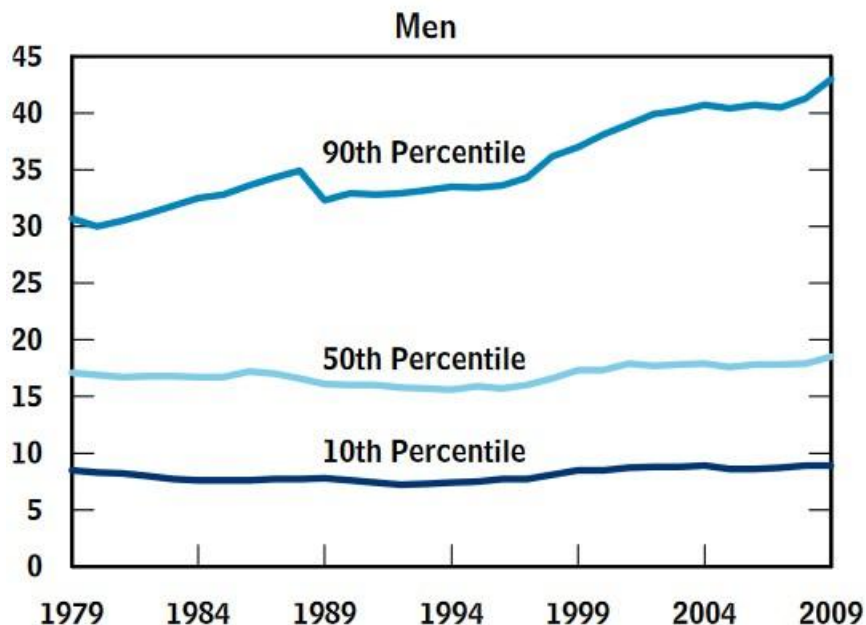
Fonte: <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/120xx/doc12051/02-16-wagedispersion.pdf>

Hourly Wages at Selected Percentiles for Workers Ages 16 to 64

Examinar a saúde econômica da classe

Hourly Wages at Selected Percentiles for Men and Women Ages 16 to 64

(2009 dollars)



Source: Congressional Budget Office based on monthly data from Census Bureau, Current Population Survey, Outgoing Rotation Groups, 1979 to 2009.

Medidas de Curtose

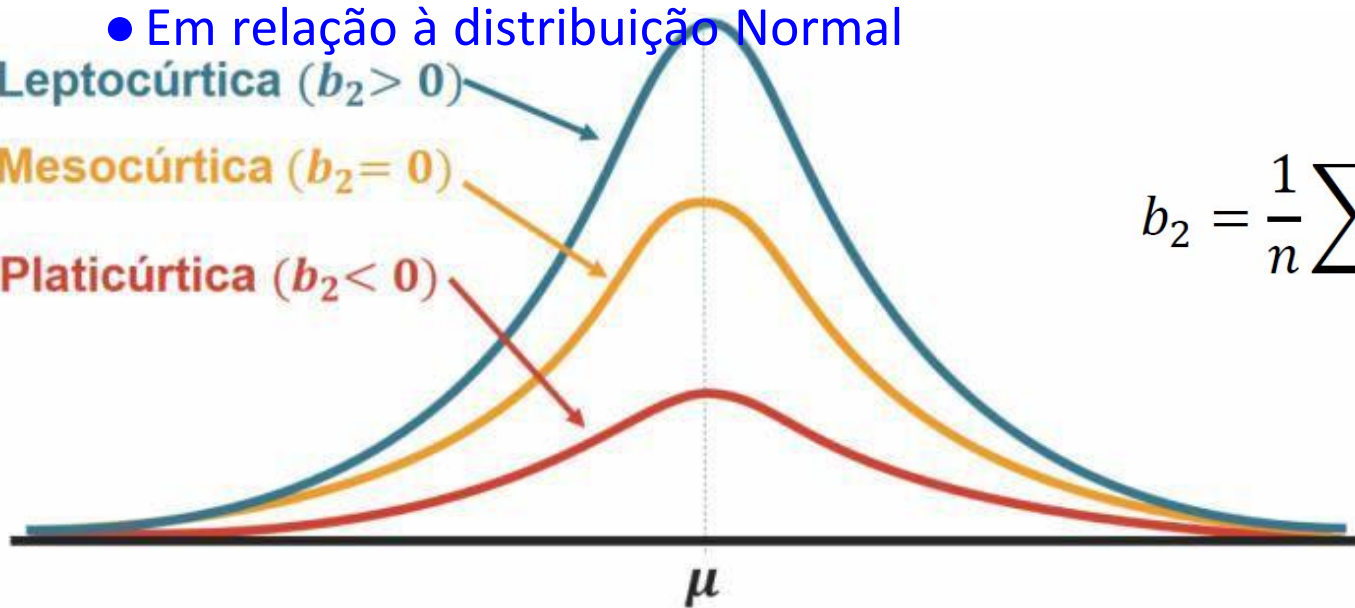
- É uma medida do achatamento de uma distribuição

- Em relação à distribuição Normal

Leptocúrtica ($b_2 > 0$)

Mesocúrtica ($b_2 = 0$)

Platicúrtica ($b_2 < 0$)

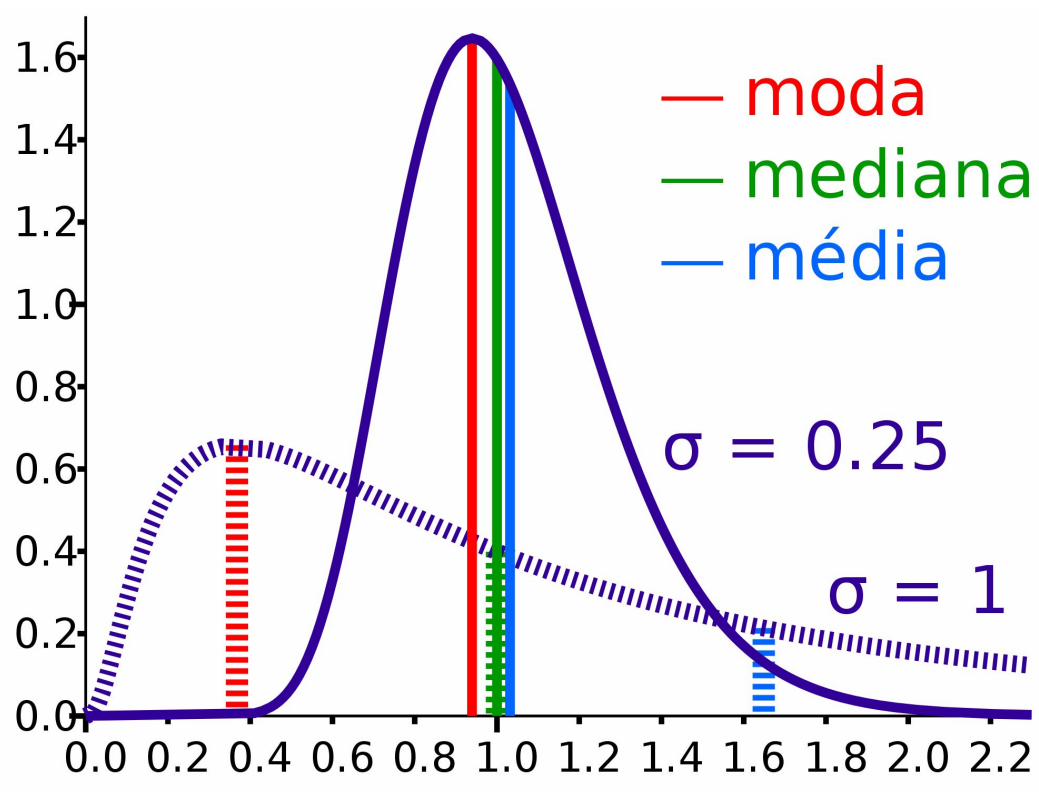


$$b_2 = \frac{1}{n} \sum \left[\frac{x_i - \mu}{\sigma} \right]^4 - 3$$

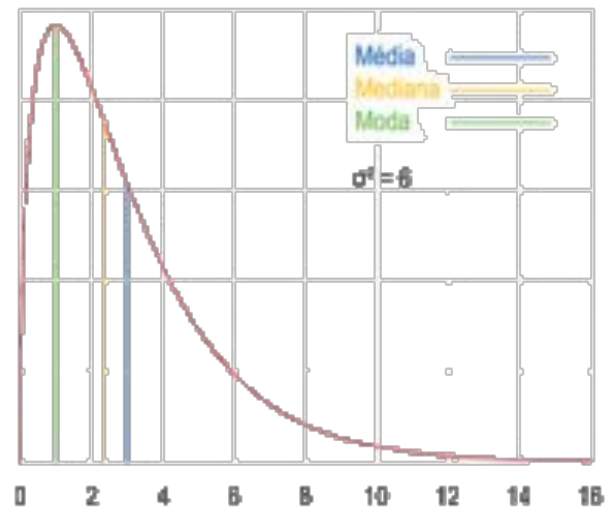
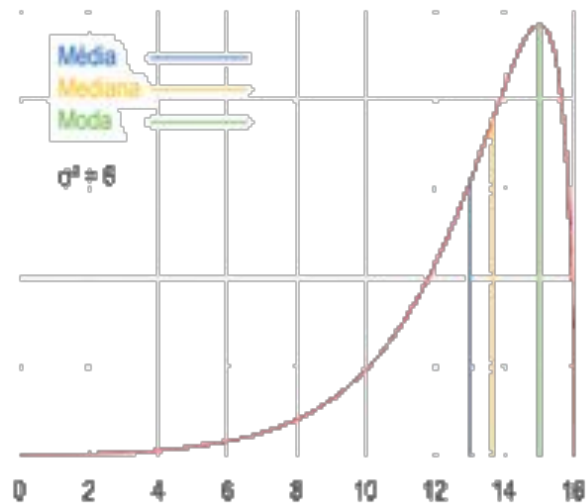
Distribuições Assimétricas

- Uma distribuição simétrica
 - Tem uma curva de frequência unimodal; e
 - Duas caudas simétricas em relação a uma linha vertical central
 - Nesta linha central estão a moda, média e mediana
- Numa distribuição assimétrica
 - Esses parâmetros não são coincidentes
 - A média sempre estará do lado da cauda mais longa
 - As caudas não são simétricas

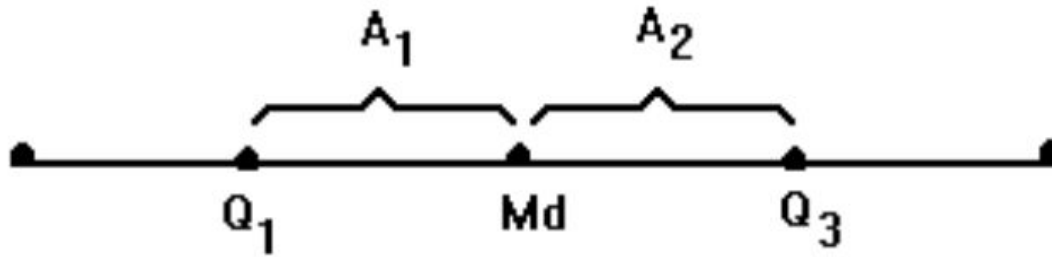
Média, Mediana e Moda de distribuições assimétricas



Média, Mediana e Moda de distribuições assimétricas

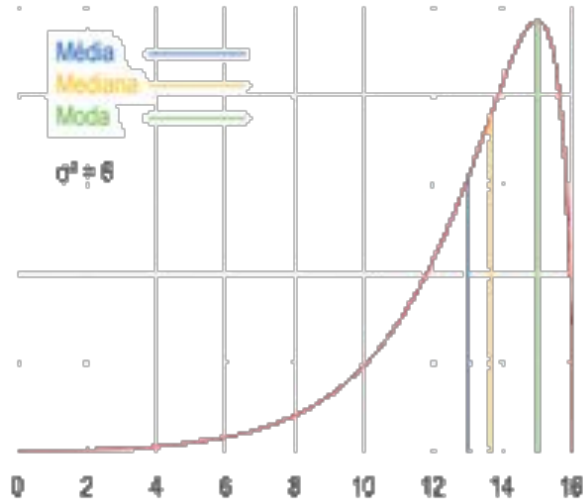


Medidas de Assimetria

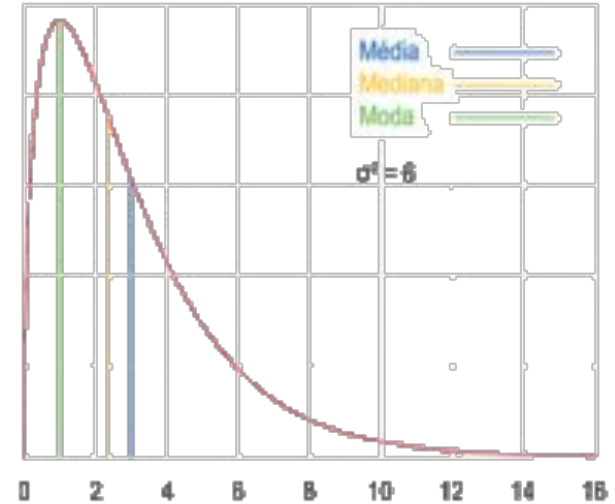


$$AS_Q = \frac{A_2 - A_1}{A_2 + A_1}$$

Média, Mediana e Moda de distribuições assimétricas

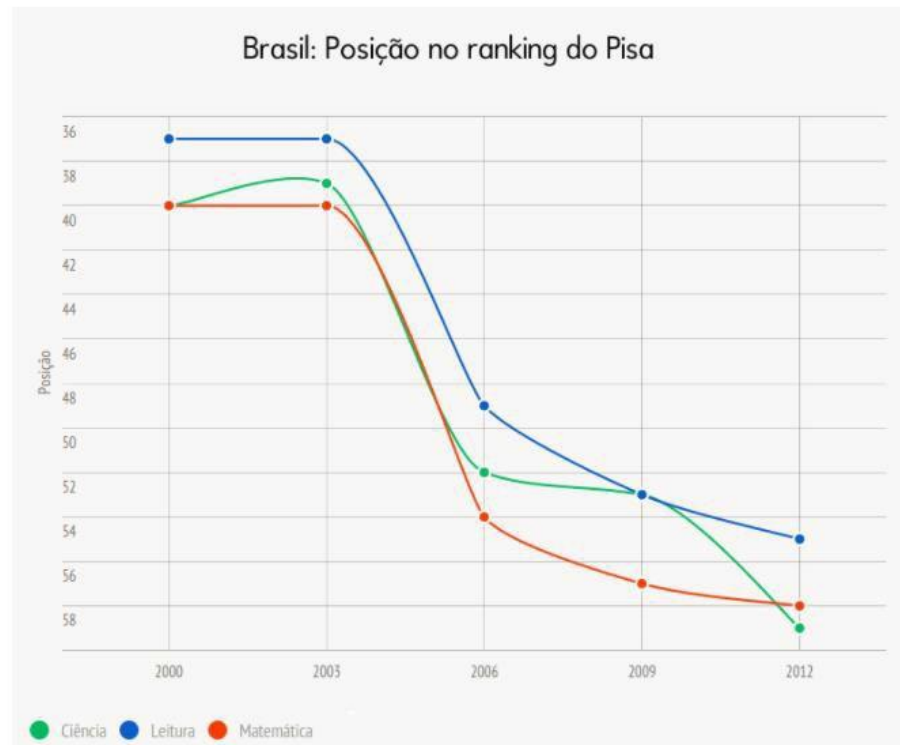
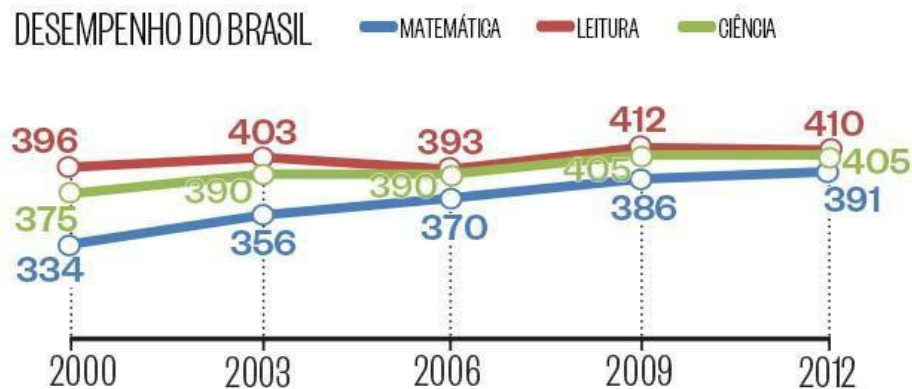


$$AS_Q < 0$$



$$AS_Q > 0$$

Atividade Prática: PISA e IDEB



Fim