

Introdução aos Testes Estatísticos

Tecnologia da Informação em Saúde – UFRN – 2018

DEB 1120 - Beatriz Stransky

DIM 0122 - Edgard Corrêa e e Fabricia Costa

IMD 0920 - Anna Giselle Rodrigues e Julio Melo

Conceito de Estatística

“Estatística é um ramo do conhecimento científico que consta de um conjunto de processos que têm, por objeto a **observação**, a **classificação formal** e a **análise dos fenômenos coletivos ou de massa** (**finalidade descritiva**) e, por fim, **investigar a possibilidade de fazer inferências indutivas válidas a partir dos dados observados** e buscar métodos capazes de permitir esta inferência (**finalidade indutiva**)”

Elza Salvatori Berquó

Divisões da Estatística

Estatística Descritiva ou Dedutiva

Como o próprio nome indica, pode ser usada para somente descrever os dados através de suas medidas mais comuns (frequências, médias, distribuições, etc.)

Estatística Inferencial, Indutiva ou Analítica

Utilizada para comparar grupos e fazer generalizações a partir dos resultados obtidos

Conceito de Variável

Resumidamente, uma variável pode ser definida como a expressão numérica de um evento qualquer. Isso significa, portanto, que a variável é o elemento-chave na análise estatística e que a estatística somente pode ser utilizada em eventos quantificáveis.

Classificação das Variáveis

Tipos de Variáveis

Qualitativas

Nominal

- Profissão
- Sexo
- Religião

Ordinal

- Escolaridade
- Estágio da doença
- Classe social

Quantitativas

Discreta

- Nº de filhos
- Nº de acessos à plataforma

Contínua

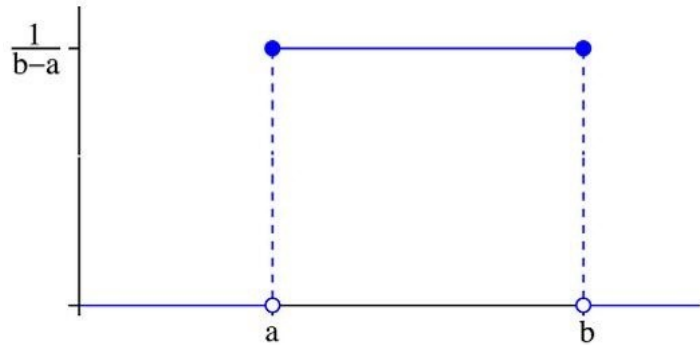
- Altura
- Peso
- Salário

Densidade e Distribuições

- **Função de densidade de probabilidade**
 - Função gerada pelo histograma das variáveis aleatórias, onde a amplitude representa a probabilidade da variável ser igual aquele valor.
 - Definido por $P(x = x_0)$ ou $f(x_0)$.
- **Distribuição de variáveis aleatórias**
 - Funções de densidade de probabilidade que são largamente conhecidas na ciência:
 - Distribuição Uniforme
 - Distribuição Binomial
 - Distribuição Normal ou Gaussiana
 - Distribuição de Poisson
 - etc

Distribuição Uniforme

- Distribuição de probabilidade para variáveis aleatórias



Valor esperado: $\bar{X} = \frac{a + b}{2}$

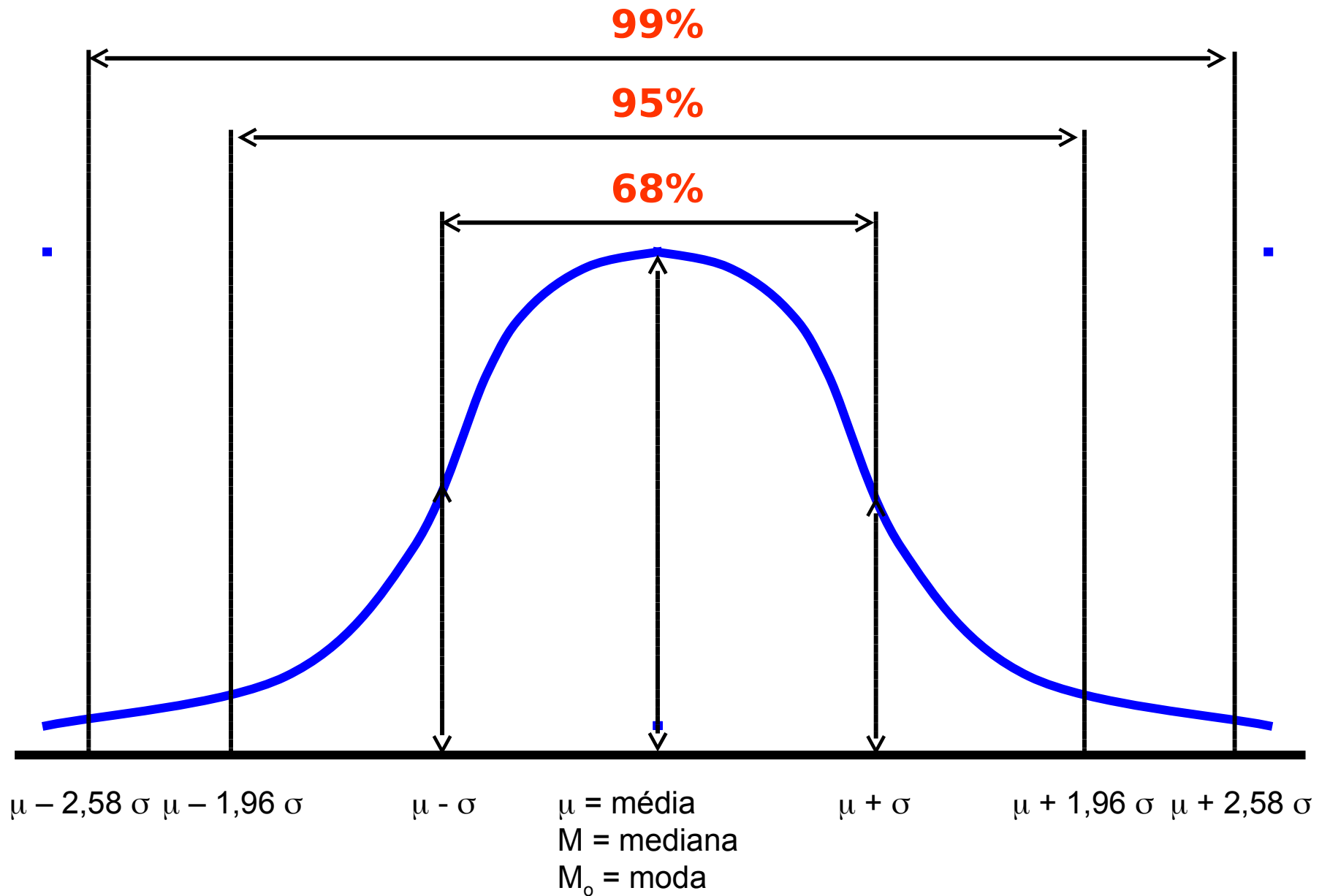
Variância: $S^2 = \frac{1}{12} (b - a)^2$

Densidade:
$$P(x = x_0) = \begin{cases} \frac{1}{b-a} & \text{se } a < x_0 < b \\ 0 & \text{caso contrário} \end{cases}$$

A Curva Normal ou Gaussiana

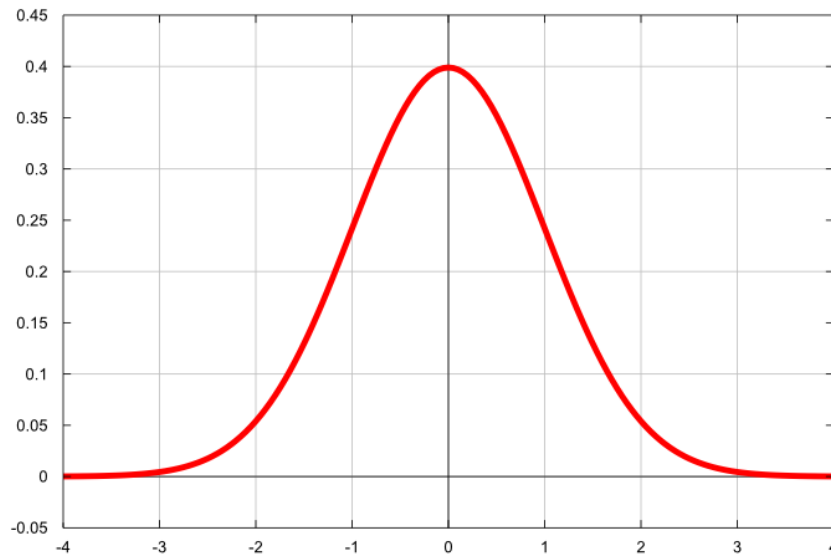


Johann Carl Friedrich Gauss



Distribuição Normal ou Gaussiana

- Distribuição de probabilidade para variáveis aleatórias



Valor esperado: $\bar{X} = \mu$

Variância: $S^2 = \sigma^2$

Densidade: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu-x)^2}{2\sigma^2}}$

“O conhecimento, pelo investigador, dos tipos de variáveis e dos testes estatísticos apropriados são análogos ao conhecimento do pintor a respeito dos tipos de pinturas (óleo, têmpera, aquarela etc.) e dos pincéis apropriados e as técnicas a serem usadas.”

Jekel et al., 2002

Princípio para aplicação de Testes

A aplicação adequada de um teste estatístico depende de alguns fatores:

- **O tipo de pergunta que se deseja responder**
- **A classificação da variável dependente e da independente**



Princípio para aplicação de Testes

Desse modo, são possíveis várias combinações em função destes fatores:

Variável Dependente	Variável Independente	Pergunta
Quantitativa Contínua/ Discreta	Quantitativa Contínua/ Discreta	As variáveis se correlacionam entre si? Com que magnitude?
Categórica	Categórica	
Quantitativa Contínua/ Discreta	Categórica	As médias (ou medianas) da variável dependente diferem entre as categorias da variável independente?

Pergunta da Pesquisa: As médias da variável dependente diferem entre os grupos estudados?

1ª Pergunta: Os dados apresentam distribuição normal?

Sim - Testes Paramétricos

Não - Testes Não-Paramétricos

2ª Pergunta: Com quantos grupos ou com quantas categorias da variável estou trabalhando?

2 grupos

Mais de 2 grupos

2 grupos

Mais de 2 grupos

3ª Pergunta: Existe vinculação entre as categorias da variável?

Sim

Não

Sim

Não

Sim

Não

Sim

Não

4ª Pergunta: Qual teste devo usar?

Teste "t" Pareado

Teste "t"

Análise Variância

Análise Variância

Wilcoxon

Mann-Whitney

Friedman

Kruskal-Wallis

Pergunta da Pesquisa: **As variáveis dependente e independente se correlacionam entre si? Qual a magnitude?**

1ª Pergunta: Qual o tipo de variável?

Quantitativa

Categórica

2ª Pergunta: Com quantos grupos ou com quantas categorias da variável estou trabalhando?

2 grupos

Mais de 2 grupos

2 grupos

Mais de 2 grupos

3ª Pergunta: Os dados apresentam distribuição normal?

Sim

Não

Sim

Não

Não

Não

4ª Pergunta: Qual teste devo usar?

**Correlação
Pearson**

**Correlação
Spearman**

**Regressão
Múltipla**

**Regressão
Múltipla**

**Qui-
Quadrado**

**Exato de
Fisher**

**Regressão
Logística**

$n > 20$

$n \leq 20$

ALGUNS EXEMPLOS....



O Teste “t” de Student

É o teste estatístico indicado para variáveis com distribuição normal, estabelecendo a comparação entre duas médias.

- **Existe diferença significativa entre as médias da altura dos homens e das mulheres de Natal?**
- MÉDIA da altura de “n” Mulheres
- MÉDIA da altura de “n” Homens

A Análise de Variância - ANOVA

É o teste estatístico indicado para variáveis com distribuição normal, estabelecendo a comparação entre três ou mais médias.

As médias de equilíbrio de 03 grupos pós experimento fisioterapêutico são diferentes???



Grupo 01 – Controle

Grupo 02 – Tratamento convencional

Grupo 03 – Tratamento por RV

A Análise de Variância - ANOVA

O Pós-Teste de Tukey-Kramer

O teste ANOVA nos informa somente se há diferença entre os grupos. Para saber onde residem as diferenças, usa-se o Pós-Teste de Tukey

As Tabelas de Contingência

O Teste Qui Quadrado

- **Testa a associação entre duas variáveis categóricas**

Quem fuma mais homens ou mulheres?

TABELA 02 x 02

Grupo 01 – Homens que fumam

Grupo 02 – Homens que não fumam

Grupo 03 – Mulheres que fumam

Grupo 04 – Mulheres que não fumam

O Coeficiente de Correlação

É o número que expressa em que medida duas variáveis se correlacionam. É dado pelo **Coeficiente de Correlação de Pearson** e varia de **-1 a +1**,

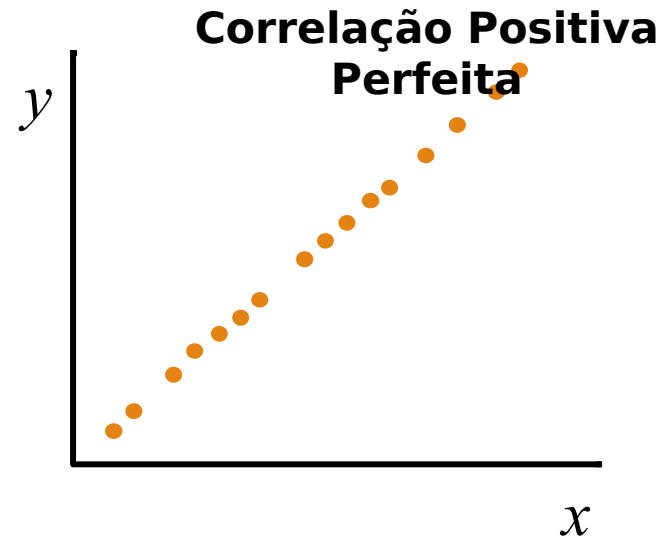
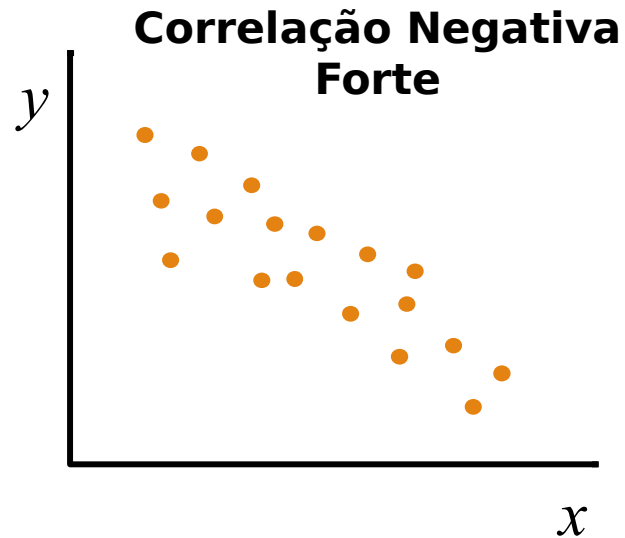
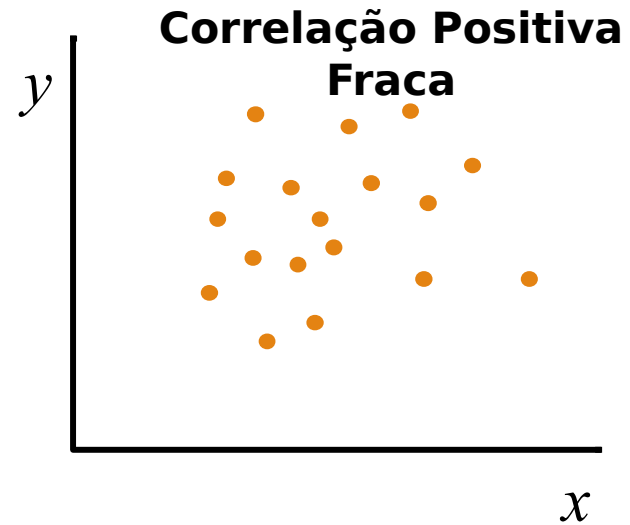
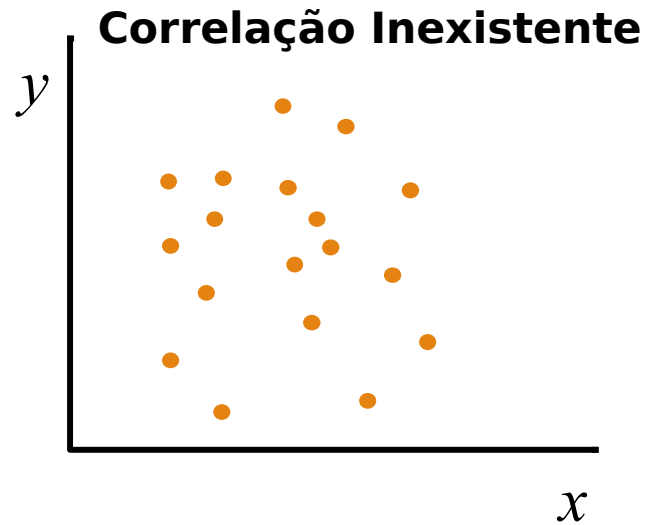
$r = 1$ correlação perfeita positiva entre as duas variáveis.

- Isto é, se uma aumenta, a outra também aumenta.

$r = -1$ correlação negativa perfeita entre as duas variáveis

- Isto é, se uma aumenta, a outra sempre diminui.

O Coeficiente de Correlação



*A escolha do teste estatístico – um tutorial em forma de apresentação em PowerPoint**

David Normando**, Leo Tjäderhane***, Cátia Cardoso Abdo Quintão****

Dental Press J. Orthod. v. 15, no. 1, p. 101-106, Jan./Feb. 2010

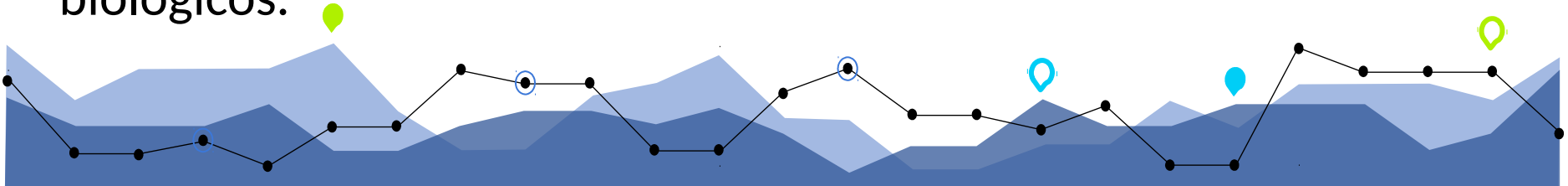
Apresenta um grupo de ícones (links) que levam o usuário a prosseguir de acordo com as suas necessidades.

Existem 5 opções:

- 1) Quero examinar se os meus dados apresentam uma distribuição Normal.
- 2) Desejo comparar grupos (procurando por diferença entre amostras).
- 3) Quero fazer correlação ou predição (regressão).
- 4) Quero checar a replicabilidade de dados (análise do erro sistemático e casual).
- 5) Desejo escolher o grafico apropriado para os meus dados.

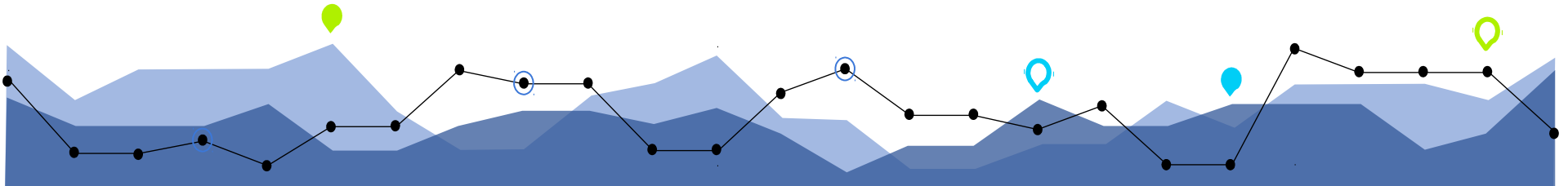
Bioestatística

- A bioestatística é **caracterizada pelos dados** e não pela ciência.
- Bases de dados biológicos (que na maioria das vezes são numéricas).
- **Médias, medianas, desvio padrão e variância** passam a ter significado biológico ao invés de puramente numérico.
- Estudos passam a requerer **maior espaço amostral** devido à grande diversidade e aleatoriedade dos processos biológicos.



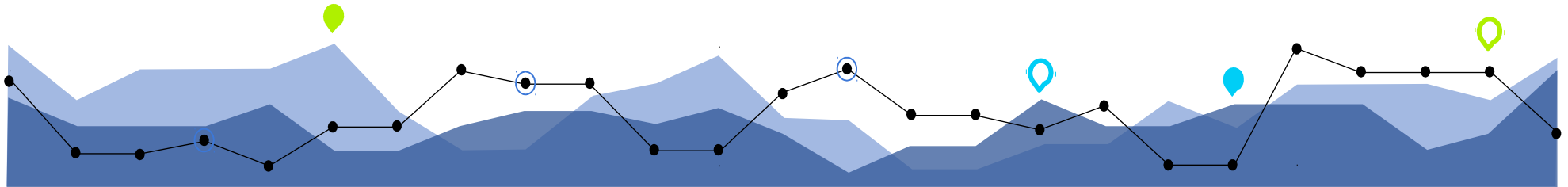
Bioestatística

- **Base de dados Zika Virus – Epidemic^[1]**
- Reporta alguns casos reportados de Zika Virus a partir de 2015.
- Base contem informações como local, quantidade de casos reportados, período da informação dos dados (*report*) e outros.
 - `report_date`
 - `location/location_type`
 - `data_field`
 - `value, unit`



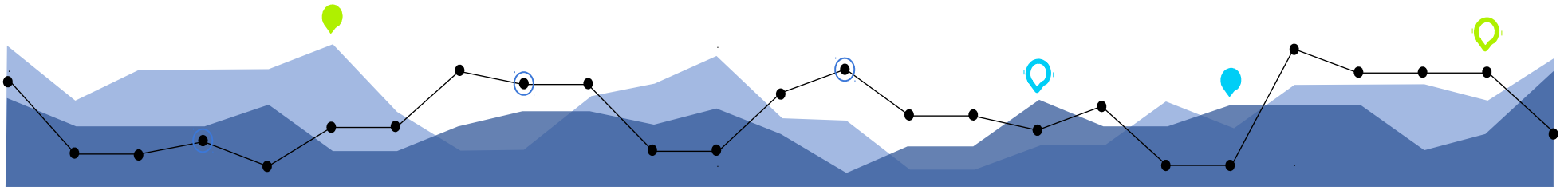
Bioestatística - Exercícios

1. Mostre, através de histogramas, os países com mais reports na base de dados.
2. Construa 3 histogramas que mostrem, por mês, incidências de Zika nos países onde houve maior incidência no ano de 2016.
3. Mostre, através de Box Plot, os números de casos por região no Brasil. (Use os estados como sendo as amostras, para cada boxplot).
4. Crie um histograma que mostre a incidência total de casos no Brasil por estado e por mês.
 - Através desse histograma construa a função de densidade de probabilidade que mostra a distribuição de casos de zika por estado e por mês.



Teste Estatístico

- Testes usados para determinar se certos dados descrevem ou não determinada característica:
 - Hipótese H_0 vs H_1
 - Probabilidade de que, em uma determinada população, a característica testada seja observada na maioria dos subconjuntos amostrais possíveis.
 - Na maioria dos casos, a distribuição '*pai*', de onde as amostras vem, é considerada com sendo normal, o que é uma suposição válida pelo teorema do limite central ^[1].



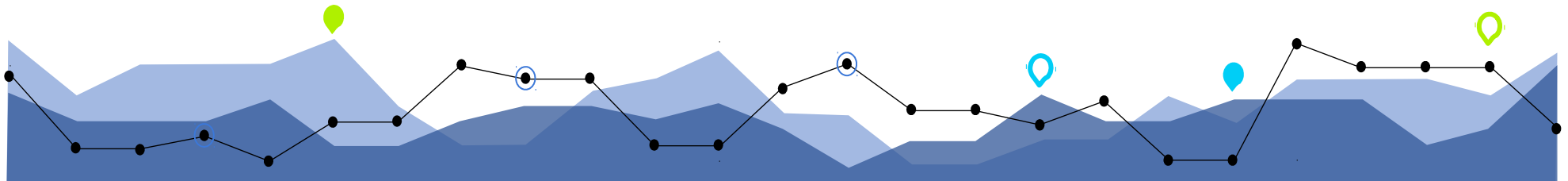
Teste Estatístico

- **Teste t-student**

- Usa uma distribuição bem conhecida pra testar variáveis dentro de um determinado intervalo.
- Define um valor de p que indica a probabilidade de erro dentro do grupo amostral escolhido.

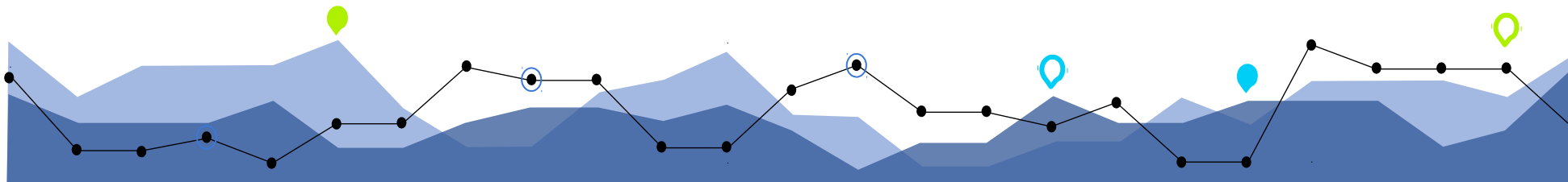
$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$


Equação para determinação de t , em um teste de médias



Teste Estatístico - Exercício

1. Calcule a média do aparecimento de Zika no Brasil no ano de 2016, em cada estado brasileiro, reportado na base de dados do Zika Virus. Calcule, através desses valores, a média de casos por 100 mil habitantes em cada estado.
2. Verifique, através de teste estatístico, se os valores calculados condizem com a média *real* proposta pelo ministério da saúde^[1] para cada região do Brasil (em casos/100 mil habitantes):
 - Centro Oeste: 113,4;
 - Nordeste: 53,5;
 - Sudeste: 41,4;
 - Norte: 36,0;
 - Sul: 6,1.



A vibrant sunset scene over a vast blue ocean. The sun is a bright, glowing orb positioned just above the horizon, partially obscured by a layer of clouds. Its light creates a brilliant, shimmering reflection that stretches down the center of the water towards the viewer. The sky is a deep, clear blue, with scattered, wispy clouds that catch the low light of the setting sun, appearing in shades of orange and yellow. The water in the foreground is a rich, textured blue with small, rhythmic waves.

Está estressado?..... Vá pescar!!!