Resumo Bioestatística Cheat Sheet

by malandro123 via cheatography.com/23628/cs/5260/

Objetivos da estatística

Fazer inferências sobre a população com base em dados amostrais

Planejar o experimento e o tamanho da amostra para que tais inferências tenham uma base confiavel

Variáveis

O que é?

É uma característica da população

Tipos Variavéis

Dependentes	Efeitos são esperados
Independentes	Efeitos que queremos medir

Variáveis Qualitativas

Nominais: não existe ordenação dentre as categorias

Ordinais: existe uma ordenação entre as categorias

Variáveis Quantitativas

Discretas: geralmente são o resultado de contagens

Contínuas: características mensuráveis que assumem valores em uma escala contínua (na reta real)

Tipos de Gráficos

Colunas e Barras

Histograma

Setores

Linhas

Diagrama de Ordenadas

Medidas Descritivas - Objetivo

Sumarizar conjuntos de dados quanto a:

Centralidade -> Medidas de locação

Variabilidade -> Medidas de dispersão

Medidas de Dispersão

Variância	(s2)=Σ[(xi-x̄)²]/n-1
Desvio	(xi - x̄)
Desvio Médio	Σ[(xi - x̄)²]/n
Amplitude	maior valor-menor valor
CV	s/x̄

Medidas de Locação

Moda	Valor que mais se repete
Média	xi/n
Mediana	Valor médio quando ordenado

Intervalo de Variação

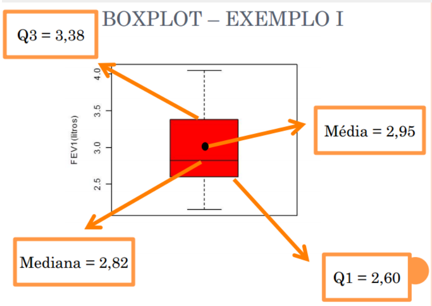
Combinação entre medidas de locação e dispersão

Lei de Chebychev

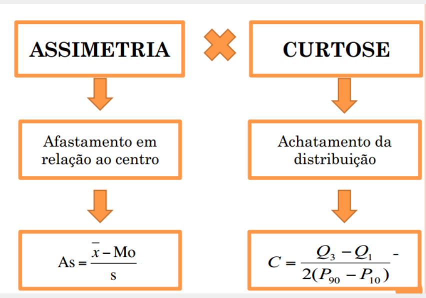
510g de Vitamina C, desvio padrão de 3g

75% (1-1/2²) possuem entre 504g e 516g

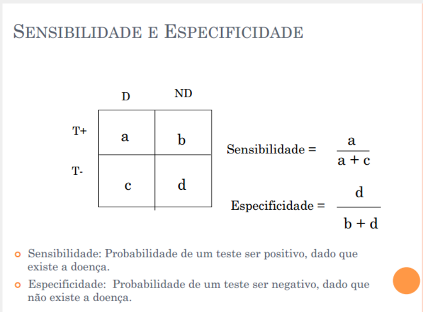
Boxplot



Assimetria x Curtose



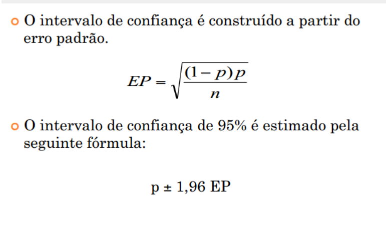
Sensibilidade e Especificidade



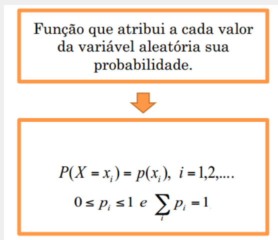
Razão de Verossimilhança Positiva e Negativa

Positiva	É uma razão entre a probabilidade de um teste ser positivo, dado que existe a doença, e a probabilidade de um teste ser positivo, dado que não existe a doença
Negativa	É uma razão entre a probabilidade de um teste ser negativo, dado que existe a doença, e a probabilidade de um teste ser negativo, dado que não existe a doença.

Cálculo do Intervalo de Confiança



Função Discreta de Probabilidade



Distribuição Normal

Mais importante distribuição de probabilidade, aplicada a inúmeros fenômenos e utilizada para o desenvolvimento teórico da Estatística.

Seja X uma variável aleatória. X terá distribuição normal se:

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

Sendo: μ = média da distribuição
σ = desvio-padrão da distribuição
π = 3,1416
e = 2,7
e $-\infty < X < +\infty$

X tem distribuição Normal com média μ e variância σ² : $X \approx N(\mu, \sigma↑2)$.

Distribuição Normal Padrão

Para calcular probabilidades $P(X \le x)$, é necessário integrar f(X) para diferentes valores de μ e σ. A solução é transformar a variável X na variável Z, dada por:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

A distribuição de probabilidades associada à variável Z denomina-se distribuição normal padrão:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}}$$

Sendo: média da distribuição = 0
variância = 1
e $-\infty < Z < +\infty$

Z tem distribuição Normal com média 0 e variância 1: $Z \approx N(0,1)$.

Exemplo

As alturas dos alunos de determinada escola são normalmente distribuídas com média 1,60 m e desvio-padrão 0,30 m. Encontre a probabilidade de um aluno medir:

a) entre 1,50 m e 1,80 m – $P(1{,}50 \le X \le 1{,}80)$;
b) mais de 1,75 m – $P(X > 1{,}75)$
c) menos de 1,48 m – $P(X < 1{,}48)$

Qual intervalo compreende 90% da população? – $P(a \le X \le b = 90\%)$

a) $P(1{,}50 \le X \le 1{,}80) = P(1{,}50 - 1{,}60/0{,}30 \le Z \le 1{,}80 - 1{,}60/0{,}30) = P(-0{,}33 \le Z \le 0{,}67)$
Pela tabela: $P = 1 - (0{,}3707 + 0{,}2514) = 0{,}3779$

b) $P(X > 1{,}75) = P(Z > 1{,}75 - 1{,}60/0{,}30) = P(Z > 0{,}50)$
Pela tabela: $P = 0{,}3085$

c) $P(X < 1{,}48) = P(Z < 1{,}48 - 1{,}60/0{,}30) = P(Z < -0{,}40)$
Pela tabela: $P = 0{,}3446$

$P(a \le X \le b = 90\%) = P(a - 1{,}60/0{,}30 \le Z \le b - 1{,}60/0{,}30 = 90\%) \rightarrow a - 1{,}60/0{,}30 = -Z↓5\%$ e $b - 1{,}60/0{,}30 = Z↓5\%$
$a - 1{,}60/0{,}30 = -1{,}64$ e $b - 1{,}60/0{,}30 = 1{,}64 \rightarrow a = 1{,}11\ m$ e $b = 2{,}09\ m$
$\rightarrow (1{,}11 \le X \le 2{,}09)$