



www.estudar.com.vc

Estatística 1

Resumo Teórico





Conceitos do Curso

1. Tipos de Variáveis e Representações Gráficas

- a. Tipos de Variáveis
- b. Distribuição de Frequências
- c. Histograma

2. Estatística Descritiva – Medidas Estatísticas

- a. Medidas de Posição (média, mediana, percentil e quartil)
- b. Medidas de Dispersão (variância, desvio padrão, coeficiente de variação)

3. Estatística Descritiva – Outliers e Padronização

- a. Regra Empírica
- b. Padronização
- c. Como identificar outliers a partir da distância em relação a média
- d. Boxplot

4. Análise Bidimensional

- a. Quali x Quali
- b. Quali x Quanti
- c. Quanti x Quanti



5. Probabilidade

- a. Variáveis Aleatórias Discretas
- b. Variáveis Aleatórias Contínuas

6. Distribuição Normal

- a. Tabela Z
- b. Cálculo da probabilidade normal para valores acima da média
- c. Cálculo da probabilidade normal para valores abaixo da média
- d. Cálculo da probabilidade normal para intervalos de valores
- e. Cálculo de valores a partir da probabilidade normal

7. Intervalos de Confiança

- a. Parâmetro, estimador e estimativa
- b. Teorema do Limite Central
- c. Margem de Erro
- d. Intervalo de Confiança para a Média Populacional com Desvio Padrão Conhecido
- e. Intervalo de Confiança para a Média Populacional com Desvio Padrão Desconhecido
- f. Tabela T-student
- g. Intervalo de Confiança para Proporção Populacional



8. Amostragem

- a. Como determinar o tamanho da amostra
- b. Métodos de Amostragem (probabilísticos x não probabilísticos)

9. Funções do Excel

- a. Análise Bidimensional
- b. Distribuição Normal
- c. Intervalos de Confiança



1. Tipos de Variáveis e Representações Gráficas

a. Tipos de Variáveis

Qualitativas: resultados são características da variável, não podem ser contados. Podem ser nominais (quando não podem ser ordenadas) ou ordinais (quando podem ser ordenadas). Exemplos: estado civil, gênero, CPF, grau de satisfação.

Quantitativas: resultados são números e podem ser contados. Podem ser discretas (números inteiros) ou contínuas (existem valores entre números consecutivos, números “quebrados”). Exemplos: altura, nota da prova, salário.

b. Distribuição de Frequências

Tabela de distribuição de frequências possui 3 colunas:

1. Valores possíveis da variável
2. Frequência Absoluta (quantas vezes cada valor apareceu)
3. Frequência Relativa (frequência absoluta de cada resultado/tamanho da amostra)

Para variáveis quantitativas contínuas: divisão dos valores em faixas

- Número de Faixas: \sqrt{n}
- Amplitude dos Dados: Valor Máximo – Valor Mínimo
- Amplitude das Faixas: $\frac{\text{Amplitude dos Dados}}{\text{Número de Faixas}}$
- As faixas aparecerão na primeira coluna da Tabela de Distribuição de Frequências



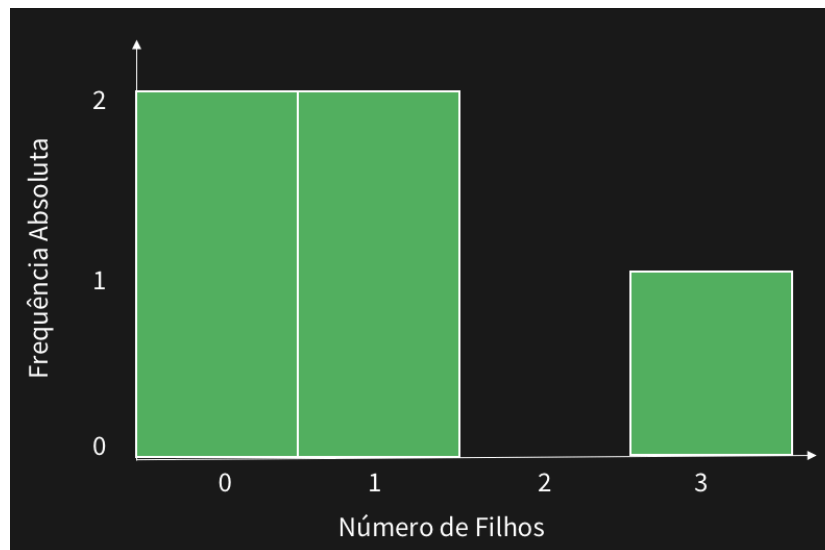
c. Histograma

Representação gráfica de variáveis quantitativas.

Eixo X: Valores da variável

Eixo Y: Frequência (relativa, absoluta ou densidade)

Exemplo:



2. Estatística Descritiva – Medidas Estatísticas

a. Medidas de Posição

Média:

$$\bar{x} = \frac{\sum x_i * n_i}{n}$$

Mediana:

Divide os dados ordenados em 2 partes iguais

Para amostras com n ímpar:

$$\frac{x_{n+1}}{2}$$



Para amostras com n par:

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Percentil:

Divide os dados ordenados em 100 partes iguais

- Multiplicar o tamanho da amostra pela porcentagem Y do percentil desejado
- Se $n * Y$ for inteiro:

$$P(Y) = \frac{x_{(n*Y)} + x_{(n*Y)+1}}{2}$$

- Se $n * Y$ não for inteiro:

$P(Y)$ será o termo correspondente ao próximo número inteiro

Quartil:

Divide os dados ordenados em 4 partes iguais

- $Q1 = P(25)$
- $Q2 = P(50) = \text{Mediana}$
- $Q3 = P(75)$
- $Q4 = P(100) = \text{Valor Máximo}$

b. Medidas de Dispersão

Variância:

- Amostral

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$



- Populacional

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

Desvio Padrão:

- Amostral

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

- Populacional

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

Coefficiente de Variação:

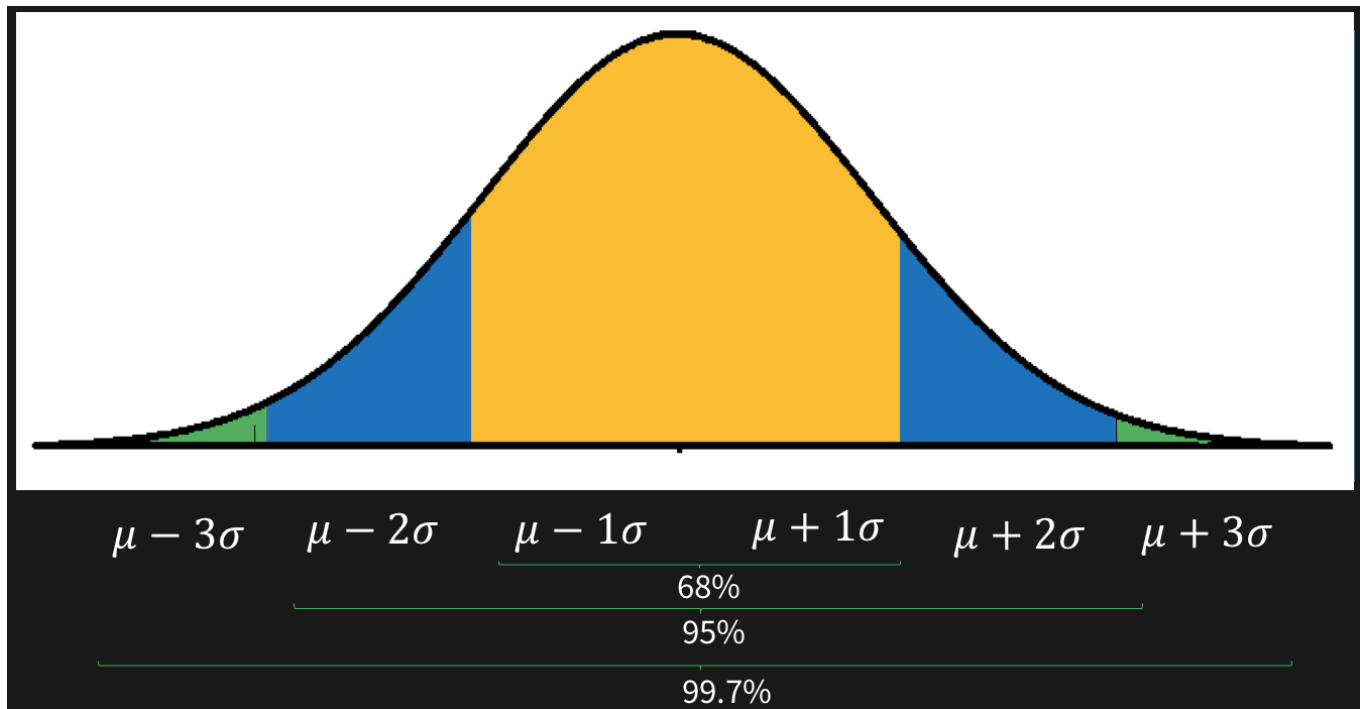
$$CV = \frac{\text{desvio padrão}}{\text{média}} * 100\%$$

3. Estatística Descritiva – Outliers e Padronização

a. Regra Empírica



Para variáveis com distribuição simétrica:



b. Padronização

$$Z = \frac{x - \mu}{\sigma}$$

c. Como identificar outliers a partir da distância em relação a média

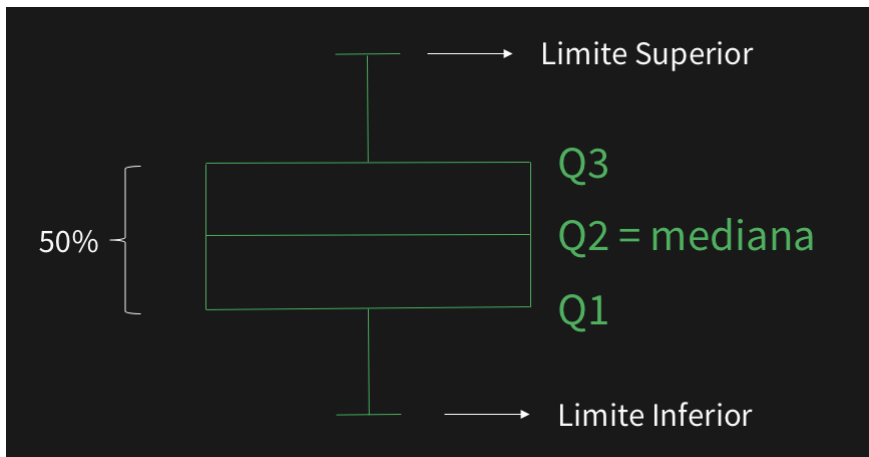
Outliers:

- Valores com $Z > 3$
- Valores com $Z < -3$

d. Boxplot

Desenho esquemático.

Permite identificar outliers de variáveis que não possuem distribuição simétrica.



- $Q3 - Q1 = IQ$
- Limite Superior = $Q3 + 1.5IQ$
- Limite Inferior = $Q1 - 1.5IQ$

Outliers:

- Valores que não se encontrarem dentro do intervalo estabelecido pelos limites: $X > \text{Limite Superior}$ ou $X < \text{Limite Inferior}$

4. Análise Bidimensional

a. Quali x Quali

Tabelas de Frequência

- Análise Vertical: total por coluna
- Análise Horizontal: total por linha

b. Quali x Quanti

Separação em grupos.

Há sinais de associação forte entre as variáveis quando:

- As médias entre os grupos são bastante diferentes
- O desvio padrão de cada grupo é menor que o desvio padrão sem dividir por grupos
- Boxplots dos grupos possuem formatos muito diferentes

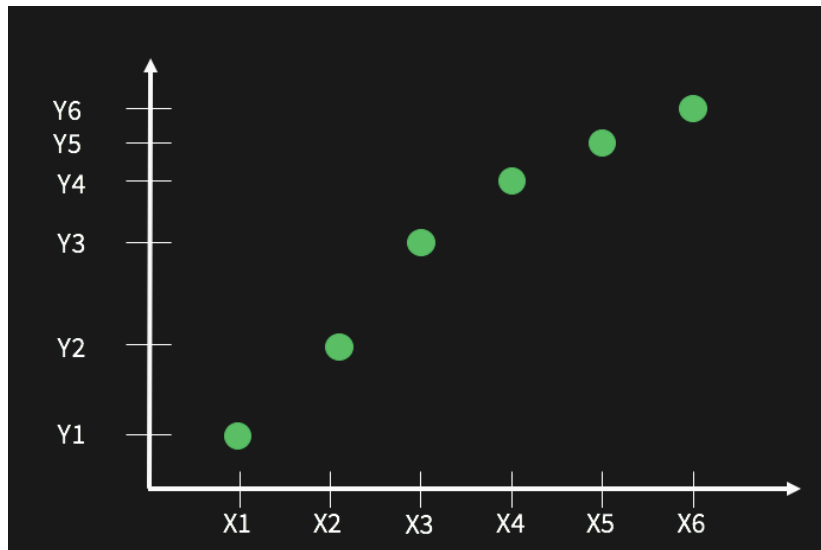
c. Quanti x Quanti

Gráfico de Dispersão



Eixo X: Valores da variável X

Eixo Y: Valores da variável Y



Associação Linear

- Positiva: X e Y caminham em direções iguais (+/+ ou -/-)
- Negativa: X e Y caminham em direções opostas (+/- ou -/+)
- Nula: não é possível identificar uma tendência entre os valores de X e Y

Covariância

$$Cov(X, Y) = \frac{\sum(x - \bar{x}) * (y - \bar{y})}{n}$$

Se $Cov(X, Y) > 0$, associação positiva

Se $Cov(X, Y) < 0$, associação negativa



Coefficiente de Correlação

$$r_{xy} = \frac{Cov(X, Y)}{\sigma_x * \sigma_y}$$

Quanto mais próximo de 1, mais forte e positiva a associação
Quanto mais próximo de -1, mais forte e negativa a associação

Coefficiente de Determinação

$$R^2 = (r_{xy})^2$$

Porcentagem da variabilidade de Y que é explicada pela variável X

Equação de Reta da Linha de Tendência

$$y = ax + b$$

Usada para prever valores estimados de Y a partir de valores de X

5. Probabilidade

a. Variáveis Aleatórias Discretas

Função de Probabilidade: associa uma probabilidade para cada valor da variável

Propriedades da Função de Probabilidade:

$$\sum f(x) = 1$$

$$f(x) \geq 0$$



Esperança:

$$E(x) = \mu = \sum x * f(x)$$

Variância:

$$Var(x) = \sum (x - \mu)^2 * f(x)$$

Distribuições de Probabilidade para Variáveis Aleatórias Discretas mais importantes:

- Ensaio de Bernoulli
- Distribuição Binomial
- Distribuição de Poisson

b.Variáveis Aleatórias Contínuas

Função Densidade de Probabilidade:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Probabilidade da variável X se encontrar entre os pontos a e b:

$$\int_a^b f(x)dx$$



Esperança:

$$E(x) = \mu = \int xf(x)dx$$

Variância:

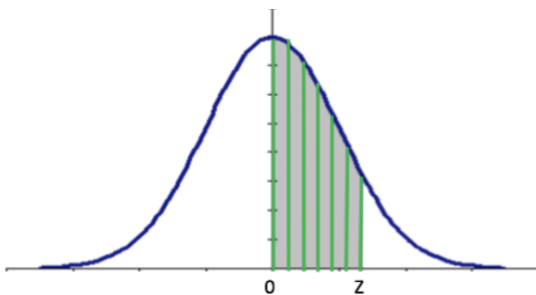
$$\int (x - \mu)^2 * f(x)dx$$

Distribuições de Probabilidade para Variáveis Aleatórias Contínuas mais importantes:

- Distribuição Uniforme
- Distribuição Normal
- Distribuição Exponencial

6. Distribuição Normal

a. Tabela Z



Z	0,00	0,01
0,0	0,0000	0,0040
0,1	0,0398	0,0438
0,2	0,0793	0,0832

- Primeira Coluna: 1ª casa decimal de Z
- Primeira Linha: 2ª casa decimal de Z
- Corpo da Tabela: área embaixo da curva normal padronizada de 0 a Z



b. Cálculo da probabilidade normal para valores acima da média

- Encontrar valor Z (> 0)
- Desenhar área correspondente a probabilidade desejada
- Buscar valor na tabela Z

c. Cálculo da probabilidade normal para valores abaixo da média

- Encontrar valor Z (< 0)
- Desenhar área correspondente a probabilidade desejada
- Desenhar área simétrica a área desejada
- Buscar valor na tabela Z

d. Cálculo da probabilidade normal para intervalos de valores

- Encontrar valor Z para os limites do intervalo
- Desenhar área correspondente a probabilidade desejada
- Buscar valores na tabela Z

e. Cálculo de valores a partir da probabilidade normal

- Inverso
- Desenhar área correspondente a probabilidade dada
- Buscar valor Z para a probabilidade na tabela Z
- Converter valor Z para a variável estudada

7. Intervalos de Confiança



a. Parâmetro, estimador e estimativa

$$\mu \leftarrow \bar{x} = \frac{\sum x_i * n_i}{n} = 10.5$$

Parâmetro Estimador Estimativa

b. Teorema do Limite Central

Quanto maior a amostra, mais a distribuição tende à distribuição normal.

- $n > 40$ costuma ser suficiente

c. Margem de Erro

$$Z * \frac{\sigma}{\sqrt{n}}$$

d. Intervalo de Confiança para a Média Populacional com Desvio Padrão Conhecido

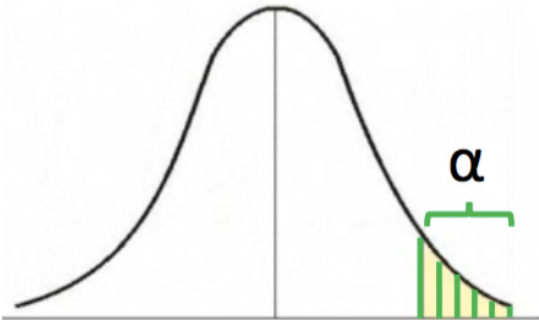
$$\bar{X} - Z * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z * \frac{\sigma}{\sqrt{n}}$$

e. Intervalo de Confiança para a Média Populacional com Desvio Padrão Desconhecido

$$\bar{X} - t_{n-1} * \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} * \frac{s}{\sqrt{n}}$$



f. Tabela T



Graus de Liberdade	25%
1	1,000
2	0,816

- Graus de Liberdade: $n - 1$
- α : área no gráfico
- Corpo da Tabela: valor t_{n-1}

g. Intervalo de Confiança para Proporção Populacional

$$\bar{p} \pm Z_{\frac{\alpha}{2}} * \sqrt{\frac{\bar{p} * (1 - \bar{p})}{n}}$$

8. Amostragem

a. Como determinar o tamanho da amostra

Para a Média Populacional:

$$n = \left(z * \frac{\sigma}{E} \right)^2$$



Para a Proporção Populacional:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 * p * (1 - p)}{E^2}$$

b. Métodos de Amostragem

Métodos Probabilísticos: permitem coletar amostras probabilísticas

- Amostra Probabilística: todos os elementos da população devem ter uma probabilidade positiva e conhecida de serem selecionados

Exemplos:

- Amostragem Aleatória Simples
- Amostragem Aleatória Estratificada
- Amostragem Sistemática

Métodos Não Probabilísticos: não permitem coletar amostras probabilísticas; seus resultados não possibilitam realizar inferências estatísticas válidas

Exemplos:

- Amostragem por Conveniência
- Amostragem Voluntária
- Amostragem por Julgamento

9. Funções do Excel



a. Análise Bidimensional

Covariância:

- Amostral: =COVARIÂNCIA.S(VALORES DE X; VALORES DE Y)
- Populacional: =COVARIÂNCIA.P(VALORES DE X; VALORES DE Y)

Coefficiente de Correlação:

=CORREL(VALORES DE X; VALORES DE Y)

b. Distribuição Normal

Probabilidade Normal a partir de valores:

- $P(X < Y)$: =DIST.NORM.N(Y; MÉDIA; DESVIO PADRÃO; VERDADEIRO)
- $P(X > Y)$: =1-DIST.NORM.N(Y; MÉDIA; DESVIO PADRÃO; VERDADEIRO)
- $P(A < X < B)$: =DIST.NORM.N(B; MÉDIA; DESVIO PADRÃO; VERDADEIRO)-DIST.NORM.N(A; MÉDIA; DESVIO PADRÃO; VERDADEIRO)

Valores a partir da probabilidade normal:

- =INV.NORM.N(PROBABILIDADE; MÉDIA; DESVIO PADRÃO)

c. Intervalos de Confiança

Valor t_{n-1} :

- =INVT(PROBABILIDADE; GRAUS DE LIBERDADE)