CENTRO UNIVERSITÁRIO SAGRADO CORAÇÃO

UNISAGRADO

CESAR OTÁVIO DA SILVA BOIANI, FELIPE MELGES RUARO, GIOVANA GIRALDELI, GIOVANI NOGUEIRA PIRES, MARCUS VINICIUS DA SILVA CAPETERUCHI

CONTEÚDO DE ESTATÍSTICA

BAURU

2025

CESAR OTÁVIO DA SILVA BOIANI, FELIPE MELGES RUARO, GIOVANA GIRALDELI, GIOVANI NOGUEIRA PIRES, MARCUS VINICIUS DA SILVA CAPETERUCHI

CONTEÚDO DE ESTATÍSTICA

Trabalho apresentado como exigência da disciplina Estatística, do curso de Ciência da Computação do Centro Universitário Sagrado Coração.

Orientadora: Prof. Rosane Maria Lima Araújo.

BAURU

2025

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE GRÁFICOS

SUMÁRIO

RESUMO

Este estudo apresenta uma análise aprofundada sobre tópicos fundamentais na área de dados e estatística, abordando conceitos e técnicas essenciais para o desenvolvimento em matemática aplicada. Discute-se os fundamentos da Estatística, incluindo população, amostra e tipos de variáveis (quantitativas e qualitativas). A organização e representação dos dados são exploradas por meio de tabelas de frequência e gráficos estatísticos. Além disso, analisam-se medidas de tendência central (média, mediana, moda), separatrizes (quartis, decis, percentis) e medidas de dispersão (amplitude, variância, desvio padrão). O estudo também inclui distribuições de probabilidade (Binomial, Poisson, Normal), intervalos de confiança, testes de hipóteses e significância. As distribuições t de Student e Qui-quadrado são abordadas, assim como a análise de correlação e regressão linear para investigar relações entre variáveis. O objetivo principal é compreender a teoria por trás de cada conceito e aplicá-los na prática, utilizando exemplos detalhados do desempenho do jogador Neymar, especialmente em relação aos seus gols, para facilitar a assimilação. Destaca-se a importância dos conceitos estatísticos na análise dos dados sobre os gols marcados por Neymar ao longo da sua carreira, contribuindo para uma tomada de decisão mais embasada nas estratégias esportivas.

Palavras-chave: Estatística; Análise de dados; Futebol; Frequência; Probabilidade; Hipóteses.

ABSTRACT

This study presents an in-depth analysis of fundamental topics in the area of data and statistics, addressing concepts and techniques essential for development in applied mathematics. The fundamentals of Statistics are discussed, including population, sample, and types of variables (quantitative and qualitative). The organization and representation of data are explored through frequency tables and statistical graphs. In addition, measures of central tendency (mean, median, mode), separatrices (quartiles, deciles, percentiles), and measures of dispersion (range, variance, standard deviation) are analyzed. The study also includes probability distributions (Binomial, Poisson, Normal), confidence intervals, hypothesis tests, and significance. Student's t-distributions and Chi-squared are covered, as well as correlation analysis and linear regression to investigate relationships between variables. The main objective is to understand the theory behind each concept and apply them in practice, using detailed examples of the player Neymar's performance, especially in relation to his goals, to facilitate assimilation. The importance of statistical concepts in the analysis of data on goals scored by Neymar throughout his career is highlighted, contributing to more informed decision-making in sports strategies.

Keywords: Statistics; Data analysis; Football; Frequency; Probability; Hypotheses.

INTRODUÇÃO

A Estatística emerge como uma disciplina científica indispensável no mundo contemporâneo, permeando desde pesquisas acadêmicas até decisões estratégicas em empresas, governos e esportes. Seu poder reside na capacidade de transformar dados brutos em conhecimento estruturado, permitindo a identificação de padrões, tendências e relações que seriam imperceptíveis a uma análise superficial. No contexto da Ciência da Computação, a estatística assume um papel ainda mais crítico, servindo como alicerce para áreas como machine learning, mineração de dados, otimização de sistemas e análise de desempenho. Este trabalho busca explorar os fundamentos da estatística, desde conceitos introdutórios até técnicas avançadas, demonstrando sua aplicação prática por meio de um estudo de caso envolvendo o desempenho do atacante brasileiro Neymar Jr.

A análise estatística começa com a definição de conceitos básicos, como população e amostra, que norteiam a coleta e interpretação de dados. Em seguida, abordamos a classificação de variáveis qualitativas e quantitativas, essenciais para determinar as técnicas adequadas de análise. A representação gráfica e tabular desses dados — por meio de histogramas, polígonos de frequência e gráficos de setores — facilita a visualização e interpretação de conjuntos complexos de informações.

À medida que avançamos, exploramos medidas de tendência central (média, moda e mediana) e medidas de dispersão (variância, desvio padrão e coeficiente de variação), que permitem resumir e comparar conjuntos de dados. Além disso, discutimos separatrizes (quartis, decis e percentis), fundamentais para entender a distribuição de valores em um dataset. Por fim, adentramos o campo das probabilidades e distribuições estatísticas, com ênfase na Distribuição Normal, amplamente utilizada em modelos preditivos e testes de hipóteses.

A escolha do futebol como estudo de caso não é aleatória. O esporte, especialmente em sua vertente profissional, é cada vez mais guiado por análises estatísticas que influenciam desde escalações táticas até estratégias de mercado. Neymar Jr., um dos atletas mais emblemáticos da atualidade, serve como exemplo ideal para ilustrar como métricas estatísticas podem quantificar desempenho, eficiência e até mesmo o impacto de um jogador em diferentes contextos.

Ao final deste trabalho, espera-se não apenas consolidar o entendimento teórico da estatística, mas também demonstrar sua versatilidade e aplicabilidade em problemas reais, reforçando sua importância tanto na academia quanto no mercado profissional.

OBJETIVO

O objetivo principal deste trabalho é apresentar e aplicar os conceitos fundamentais da estatística, destacando sua importância na análise de dados. Buscamos:

1. Compreender os conceitos teóricos de população, amostra, variáveis qualitativas e quantitativas, e suas representações por meio de tabelas e gráficos.

2. Explorar medidas estatísticas como tendência central (média, moda, mediana), separatrizes (quartis, decis, percentis) e dispersão (variância, desvio padrão).

3. Aplicar técnicas estatísticas em exemplos práticos, utilizando dados do jogador Neymar para ilustrar a análise de desempenho esportivo.

4. Introduzir conceitos avançados como distribuições de probabilidade (Binomial, Poisson, Normal) e sua relevância em problemas reais.

Através desse estudo, esperamos consolidar o conhecimento estatístico e demonstrar sua utilidade em contextos variados, especialmente na ciência da computação e no esporte.

JUSTIFICATIVA

A estatística desempenha um papel crucial na ciência de dados e na tomada de decisões, sendo essencial para interpretar fenômenos complexos e extrair insights significativos. Este trabalho justifica-se pelos seguintes motivos:

1. Relevância Acadêmica: A disciplina de Estatística é fundamental para estudantes de Ciência da Computação, pois fornece as bases necessárias para o desenvolvimento de algoritmos, análise de dados e machine learning. Dominar esses conceitos é essencial para a formação profissional na área.

2. Aplicação Prática: A utilização de exemplos reais, como os dados de desempenho do jogador Neymar, facilita a compreensão dos conceitos teóricos e demonstra como a estatística pode ser aplicada em contextos do mundo real, como o esporte.

3. Interdisciplinaridade: O trabalho conecta a estatística com outras áreas, como a computação e o esporte, mostrando sua versatilidade e importância em diferentes campos.

4. Preparação para o Mercado de Trabalho: O domínio de técnicas estatísticas é altamente valorizado no mercado de trabalho, especialmente em áreas como análise de dados, inteligência artificial e desenvolvimento de software. Este trabalho contribui para o desenvolvimento dessas habilidades.

Além disso, a escolha do tema relacionado ao futebol busca despertar o interesse dos alunos, mostrando como a estatística pode ser envolvente e aplicável até mesmo em assuntos do cotidiano, como o esporte.

ANÁLISE DOS RESULTADOS, DISCUSSÃO, TEORIAS, CURIOSIDADES E APLICAÇÕES

AULA 1: População, Amostra, Variável Quantitativa E Qualitativa

Estatística

É a ciência que oferece uma coleção de métodos para produzir e obter dados, organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conhecimento. Deste modo, a Estatística contribui para que dados gerem conhecimento e, como tal, deve ter como objetivo não só a produção de dados, como também a interpretação de dados já existentes, utilizando a combinação de gráficos, tabelas e medidas numéricas que permitam interpretar o que esses dados significam.

População

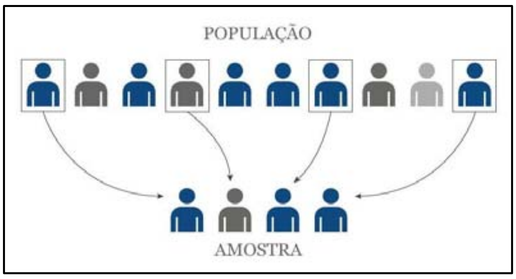
- Conjunto de elementos que possuem características em comum;
- Pode ser finita ou infinita;
- Pode ser composta por pessoas, objetos, eventos ou elementos.

Na estatística, a população é o conjunto de elementos que possuem características em comum, enquanto a amostra é um subconjunto da população.

Amostra

- Subconjunto da população que será estudado;
- Deve ser representativa da população;
- Pode ser probabilística ou não probabilística;
- Pode ser extraída por meio de técnicas de amostragem, como a casual simples, sistemática ou estratificada. A amostra é utilizada para analisar a população, permitindo obter resultados representativos sem precisar coletar dados de todos os indivíduos.

Figura 1 - Diagrama de árvore (população/amostra)



Fonte: Educa Mais Brasil, 2025.

Variáveis qualitativas

São aquelas cujos valores podem ser separados em diferentes categorias que se distinguem por alguma característica não numérica. Por exemplo: sexo (masculino e feminino), cor dos cabelos (preto, loiro, ruivo, castanho, etc). Podem ser subdivididas em:

- **Variáveis qualitativas ordinais:** quando existe uma ordem nos seus valores. Por exemplo, a variável "Grau de instrução" pode ter seus valores ordenados (fundamental, médio, superior, etc). O mesmo não ocorre com a variável "cor da pele".

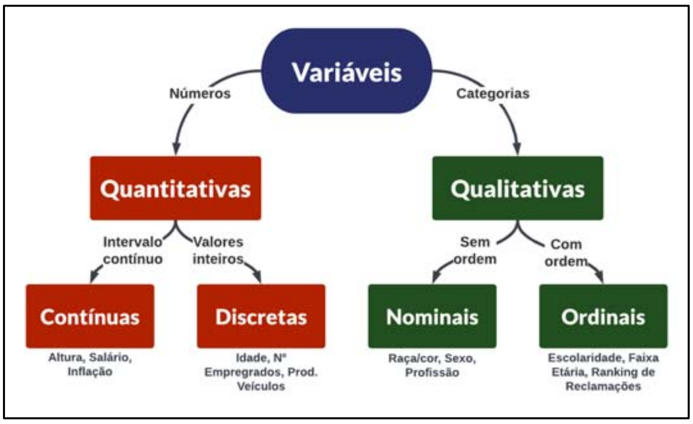
- **Variáveis qualitativas nominais:** quando uma ordem não pode ser estabelecida entre seus valores.

Variáveis quantitativas

São aquelas cujos valores são expressos em números. As variáveis quantitativas se subdividem, ainda, em:

- **Variáveis quantitativas discretas:** quando resultam de um conjunto finito (ou enumerável) de valores possíveis. Por exemplo, o número de livros de Matemática da biblioteca da escola e a quantidade de carros presentes, agora, num estacionamento, são variáveis discretas.

Figura 2 - Diagrama de árvore (variáveis)



Fonte: Medium, 2025.

- **Variáveis quantitativas contínuas:** quando resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma escala contínua, de modo que não haja lacunas ou interrupções. Por exemplo, o peso de um pacote de arroz e o comprimento de um parafuso constituem variáveis contínuas.

Censo

A palavra **censo** vem do latim *census* e quer dizer "conjunto dos dados estatísticos dos habitantes de uma cidade, província, estado, nação".

Amostragem

- Amostragem é o processo de selecionar uma parte de uma população para estudar e fazer inferências sobre o todo.
- A amostragem é importante porque permite economizar recursos e tempo, pois não é necessário pesquisar toda a população.

Existem dois tipos de amostragem:

- **Amostragem probabilística**

A seleção dos elementos da amostra é aleatória, ou seja, cada elemento da população tem a mesma chance de ser escolhido.

- **Amostragem não-probabilística**

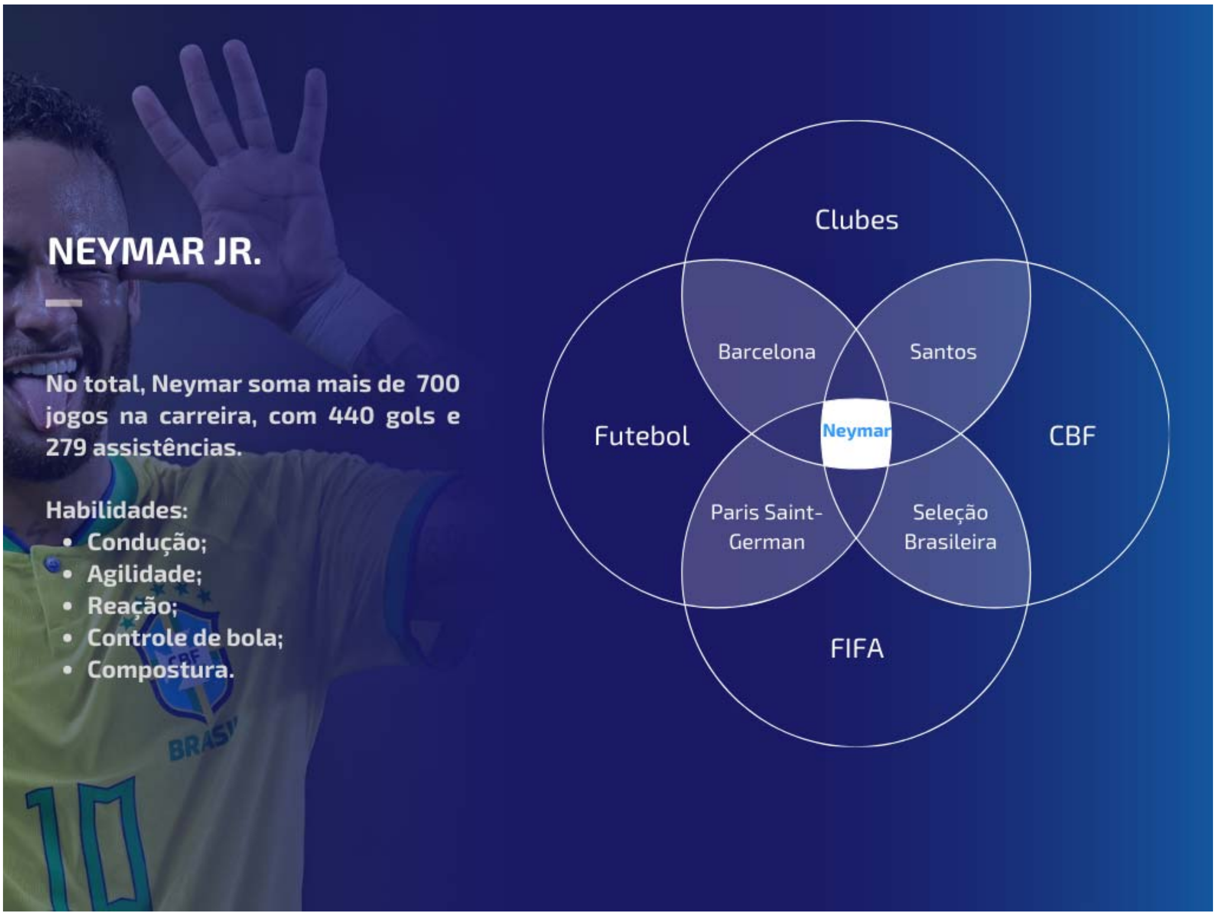
A seleção dos elementos da amostra é feita de forma deliberada, com base em critérios definidos pelo pesquisador.

EXERCÍCIO:

Construir um diagrama (conjunto/subconjunto), para registrar exemplos de população, amostra e elemento.

Resposta:

Gráfico 1 - Diagrama de Venn



CURIOSIDADES:

Biografia resumida de Neymar

Nome completo: Neymar da Silva Santos Júnior

Data de nascimento: 5 de fevereiro de 1992.

Local de nascimento: Mogi das Cruzes, São Paulo, Brasil.

Neymar é um dos jogadores de futebol mais conhecidos e talentosos do mundo. Desde jovem, mostrou grande habilidade com a bola e começou sua carreira nas categorias de base do Santos Futebol Clube, onde se destacou rapidamente.

Carreira

• **Santos (2009-2013):** Neymar começou sua carreira profissional no Santos, onde rapidamente se destacou como um dos melhores jovens talentos do Brasil. Ele ajudou o clube a conquistar a Copa do Brasil em 2010 e a Copa Libertadores em 2011.

• **Barcelona (2013-2017):** Em 2013, transferiu-se para o FC Barcelona, onde formou um trio lendário com Lionel Messi e Luis Suárez. Durante sua passagem pelo clube, ganhou diversos títulos, incluindo a UEFA Champions League em 2015.

• **Paris Saint-Germain (2017-2024):** Neymar se transferiu para o PSG em uma transferência recorde em 2017. Ele se tornou uma das estrelas mais brilhantes da equipe e conquistou vários títulos da Ligue 1. No entanto, em 2024, ele deixou o PSG após uma série de controvérsias e decisões pessoais.

• **Retorno ao Santos (2025):** Em uma reviravolta emocionante, Neymar retornou ao Santos em 2025, onde começou sua carreira. O retorno foi calorosamente recebido pelos torcedores e promete trazer nova energia ao clube.

Seleção Brasileira

Neymar tem sido um pilar na seleção brasileira desde sua estreia em 2010. Ele participou de várias Copas do Mundo e Copas América, sendo fundamental na conquista da medalha de ouro nos Jogos Olímpicos de 2016 no Rio de Janeiro. Em 2025, ele continua a ser uma figura chave na busca do Brasil por títulos internacionais.

Estilo de Jogo

Reconhecido por sua velocidade impressionante, dribles habilidosos e capacidade de finalização, Neymar é admirado por fãs e críticos. Seu estilo carismático dentro e fora do campo fez dele um ícone global.

AULA 2: Tabela de Distribuição de Frequência – TDF
Frequência

Refere-se ao número de vezes que um determinado valor ou categoria aparece em um conjunto de dados.

- **Frequência Absoluta ou Frequência Simples**

  É a contagem total de ocorrências de um determinado valor ou categoria em um conjunto de dados. Ex.: Se em uma lista de notas de alunos, a nota 7 aparece 5 vezes, a frequência absoluta da nota 7 é 5.

- **Frequência Relativa**

  É a proporção da frequência absoluta em relação ao total de observações. É calculada pela fórmula: $Fequência\ relativa = \frac{frequência\ absoluta}{total\ de\ observações}$. Ex.: Se a frequência absoluta da nota 7 é 5 e o total de notas é 20, a frequência relativa de nota 7 seria $\frac{5}{20}$ = 0,25 ou 25%.

- **Frequência Acumulada:**

  Representa a soma das frequências absolutas até um certo ponto na tabela. Ela mostra quantas observações estão abaixo ou iguais a determinado valor. Para calcular, é só somar as frequências absolutas sucessivamente.

Porcentagem:

A porcentagem é uma forma de expressar uma fração com base em 100. Para calcular a porcentagem da frequência relativa, multiplica-se por 100:

$porcentagem = frequência\ relativa \times 100.$

Grau:

Em algumas situações, especialmente em gráficos circulares (pizza), os dados podem ser representados em graus. A soma total do círculo é 360 graus. Para calcular quantos graus corresponde à frequência relativa de uma categoria, usa-se:

$grau = frequência\ relativa \times 360$. No exemplo da nota 7 com frequência relativa de 0,25, o cálculo seria: $0,25 \times 360 = 90$°.

EXERCÍCIO:

Construir e completar a Tabela de Distribuição de Frequência (TDF) para variáveis qualitativas e quantitativas.

Resposta:

Tabela 1 - Tabela variável qualitativa

GOLS NEYMAR									
i	CLUBES	GOLS (fi)	fri	%fri	°fri	Fi	Fri	%Fri	°Fri
1	Al Hilal	1	0,002	0,22%	0,79	1	0,002	0,22%	0,79
2	Barcelona	105	0,239	24%	86,04	106	0,241	24%	86,76
3	Paris Saint German	118	0,268	27%	96,48	224	0,510	51%	183,60
4	Santos	137	0,311	31%	111,96	361	0,820	82%	295,20
5	Seleção Brasileira	79	0,179	18%	64,44	440	1,000	100%	360,00
K = 5		440	1,000	100%	360,00				

Tabela 2 - Tabela variável quantitativa

GOLS DE NEYMAR POR MINUTOS										
i	MINUTOS	GOLS (fi)	fri	%fri	°fri	Fi	Fri	%Fri	°Fri	xi
1	1|-|15	55	0,125	13%	45,00	55	0,125	13%	45,00	8
2	16|-| 30	64	0,145	15%	52,20	119	0,270	27%	97,20	23
3	31|-|45	84	0,190	19%	68,40	203	0,461	46%	165,96	38
4	46|-|60	71	0,161	16%	57,96	274	0,622	62%	223,92	53
5	61|-|75	76	0,172	17%	61,92	350	0,795	80%	286,20	68
6	76|-|90	68	0,154	15%	55,44	418	0,950	95%	342,00	83
7	91|-|105	22	0,050	5%	18,00	440	1,000	100%	360,00	98
K = 7	hi = 14	440	1,000	100,00	360,00					
At = 105 - 1 = 104		Aa = 105 - 1 = 104								

$K = 1 + 3{,}322 \times \log 440 = \mathbf{9,781} \cong \mathbf{10}\ \boldsymbol{linhas}$

CURIOSIDADES:

Estatística no Futebol

1. Análise e desempenho dos jogadores: As estatísticas são essenciais para avaliar o desempenho individual dos jogadores. Algumas métricas comuns incluem:

- Gols e assistências;
- Passes completos;
- Dribles e faltas.

2. Análise Tática: As estatísticas também são usadas para analisar táticas de equipe. Por exemplo:

- Posse de bola;
- Distribuição de passes;
- Distribuição espacial (mapas de calor).

3. Avaliação de equipes: além das análises individuais, as estatísticas também permitem avaliar o desempenho geral das equipes:

- Resultados em Casa vs. Fora;
- Média de gols marcados e sofridos;
- Ranking estatístico.

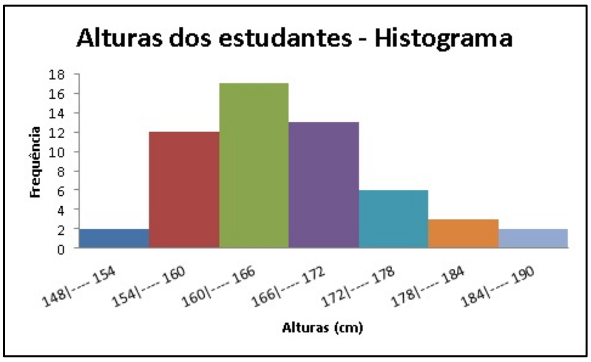
4. Modelagem Preditiva: Com o aumento da análise de dados, muitos clubes começam a usar modelagem preditiva para antecipar resultados:

- **Simulações estatísticas:** modelos que utilizam dados históricos para prever resultados futuros ajudam os clubes a fazer melhores escolhas em termos de escalação, táticas e estratégias durante as partidas.

AULA 3: Gráficos Estatísticos
Histograma

- É uma representação gráfica que mostra a distribuição de um conjunto de dados quantitativos. Ele é construído a partir de intervalos (ou classes) em que os dados são agrupados. O eixo *x* representa os intervalos, enquanto o eixo *y* representa a frequência absoluta de cada intervalo.
- Cada barra do histograma é proporcional à frequência dos dados naquele intervalo. É uma boa maneira de visualizar a densidade e a distribuição dos dados.

Figura 3 - Exemplo histograma

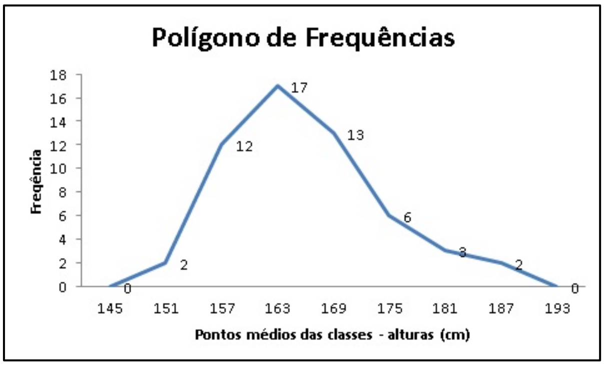


Fonte: Blog de Estatística do Professor Alexandre, 2025.

Polígono de Frequência

- É uma linha que conecta os pontos médios do topo das barras de um histograma. Para construí-lo, primeiro traça um histograma e, em seguida, usa os pontos médios das barras como coordenadas para criar uma linha.
- É ideal para ajudar a visualizar as tendências dos dados e facilita a comparação entre diferentes conjuntos de dados.

Figura 4 - Exemplo polígono de frequência

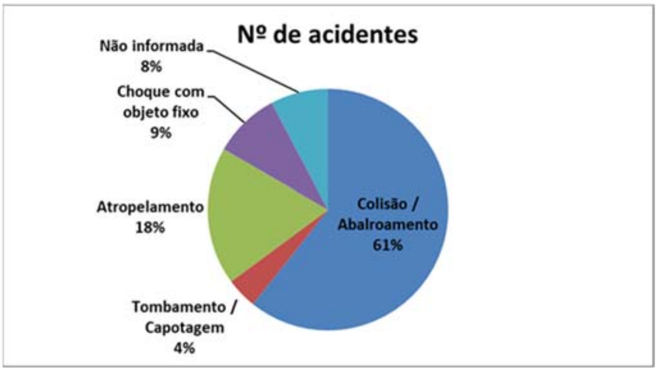


Fonte: Ajudando na Matemática, 2025.

Gráficos

- **Setores:** É utilizado para mostrar a proporção relativa das partes em relação ao todo. Cada setor representa uma categoria e o tamanho do setor é proporcional à quantidade ou porcentagem dessa categoria em relação ao total. É ideal para visualizar dados qualitativos ou categorias.

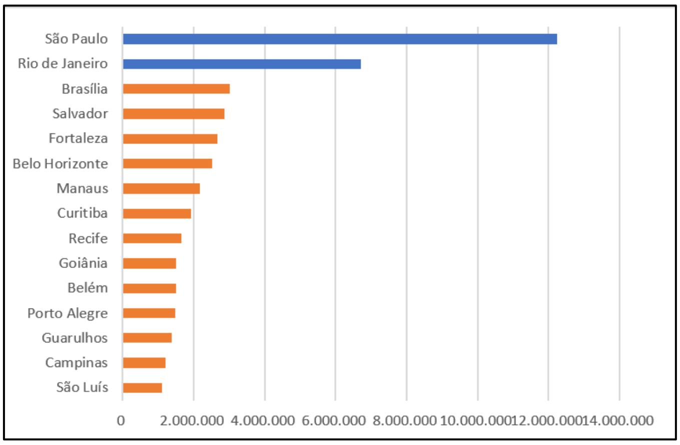
Figura 5 - Exemplo gráfico de setores



Fonte: Professor Guru, 2025.

- **Barras:** Utilizado para comparar diferentes categorias. As barras podem ser verticais ou horizontais e representam a frequência ou a quantidade de cada categoria.

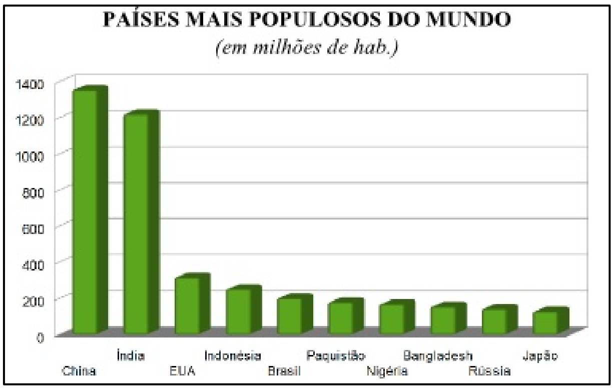
Figura 6 - Exemplo gráfico de barras



Fonte: Hashtag Treinamentos, 2025.

- **Colunas:** Semelhante ao gráfico de barras, mas com as barras dispostas verticalmente. Frequentemente utilizado para mostrar variações ao longo do tempo.

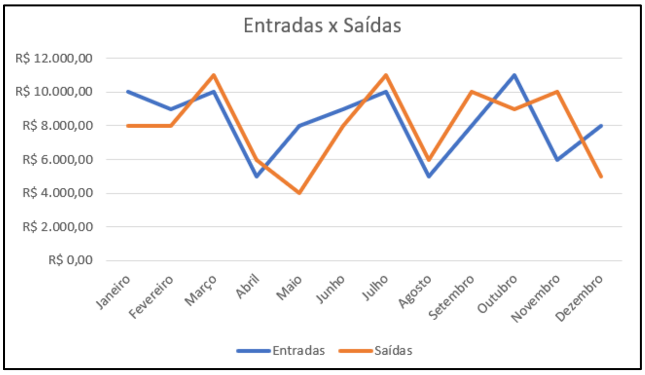
Figura 7 - Exemplo gráfico de colunas



Fonte: Mundo Educação, 2025.

- **Linhas:** Ideal para mostrar tendências ao longo do tempo, esse gráfico conecta pontos de dados com linhas. Muito utilizado em análises financeiras e científicas.

Figura 8 - Exemplo gráfico de linhas



Fonte: Excel Easy, 2025.

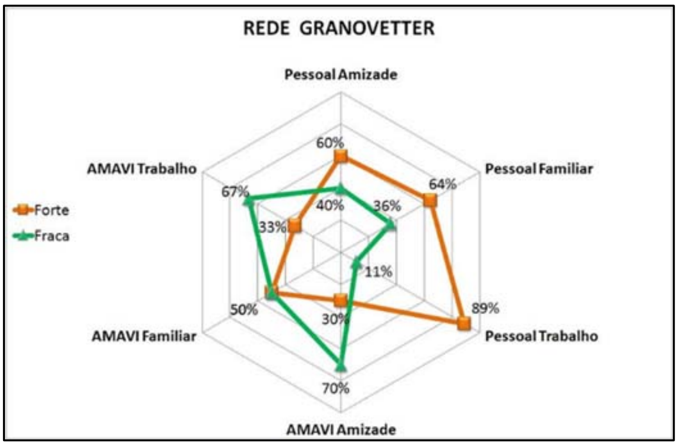
- **Áreas:** É semelhante ao gráfico de linhas, mas a área abaixo da linha é preenchida com uma cor. É utilizado para mostrar a magnitude das mudanças ao longo do tempo e pode representar partes do todo.

Figura 9 - Exemplo gráfico de áreas

Fonte: Ninja do Excel, 2025.

- **Redes:** É usado para representar relações entre diferentes entidades. Comum em análises sociais e biológicas, onde se deseja visualizar conexões e interações entre elementos.

Figura 10 - Exemplo gráfico de redes



Fonte: ResearchGate, 2025.

EXERCÍCIO:

Gerar os gráficos das variáveis qualitativas e quantitativas.

Resposta:

Gráficos qualitativos:

Gráfico 2 - Gols de Neymar por clube (fi)

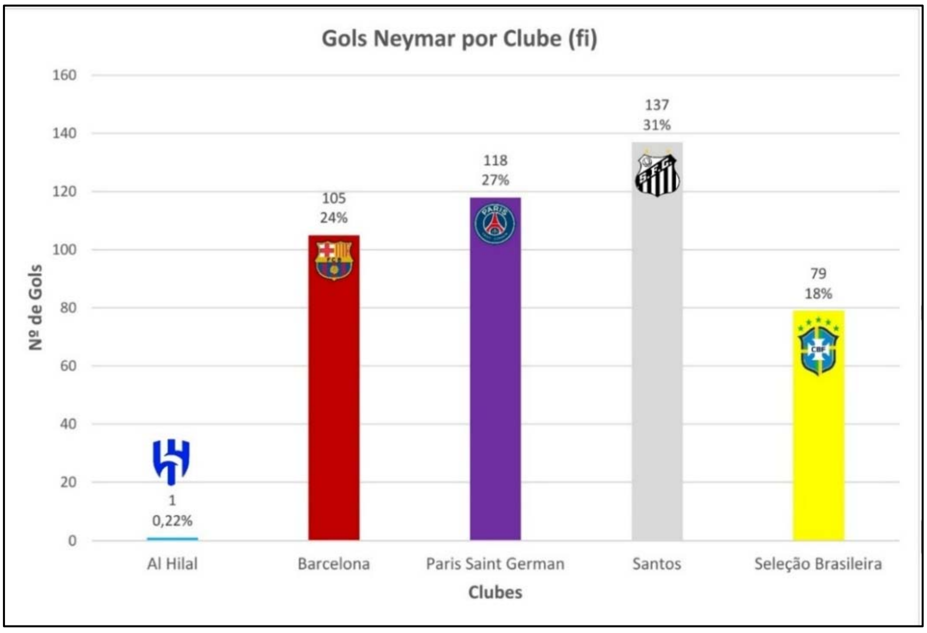


Gráfico 3 - Gols de Neymar por clube (FI)

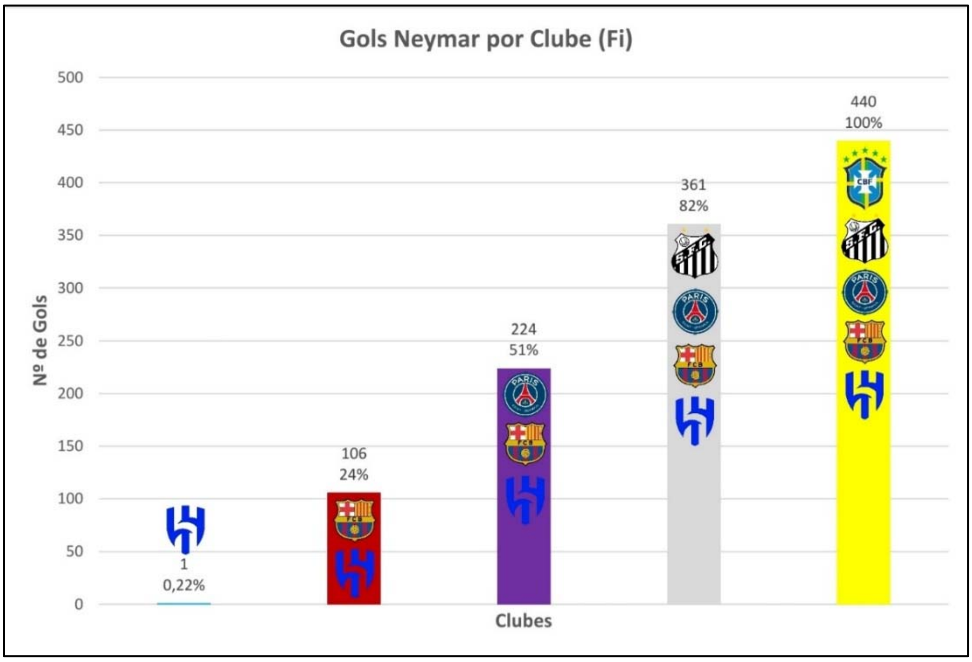
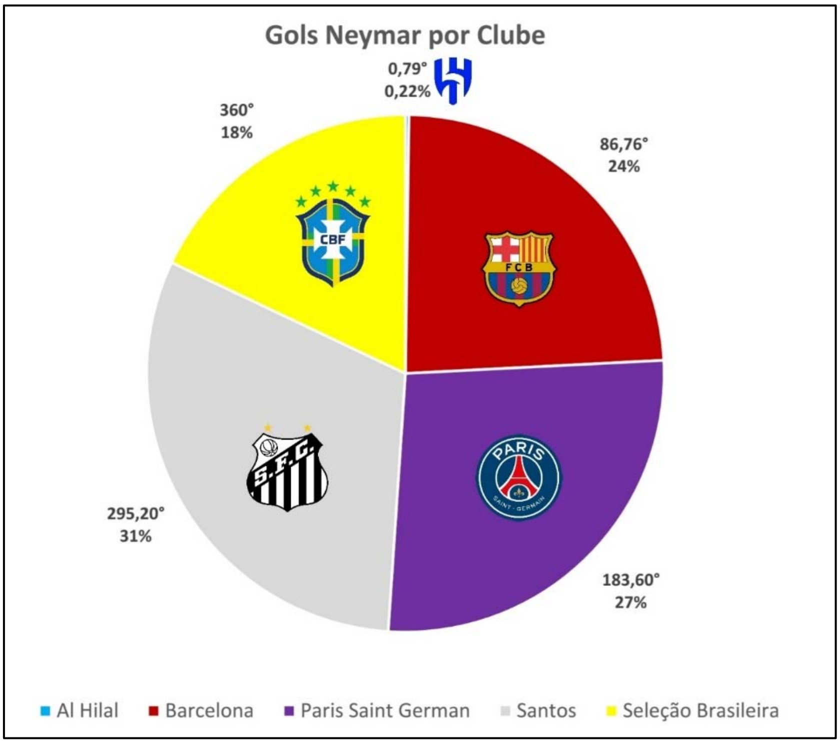


Gráfico 4 – Setores: Gols de Neymar por clube (FI)



Gráficos quantitativos:

Gráfico 5 - Histograma + polígono fi(xi): Gols por minutos

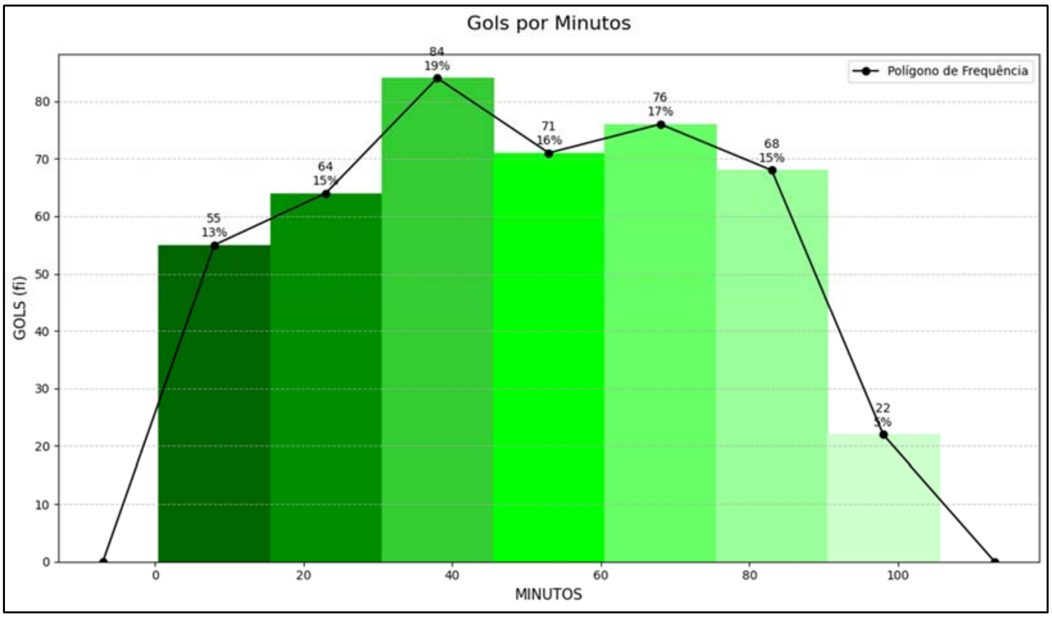


Gráfico 6 - Histograma + polígono Fi(Li): Gols por minutos

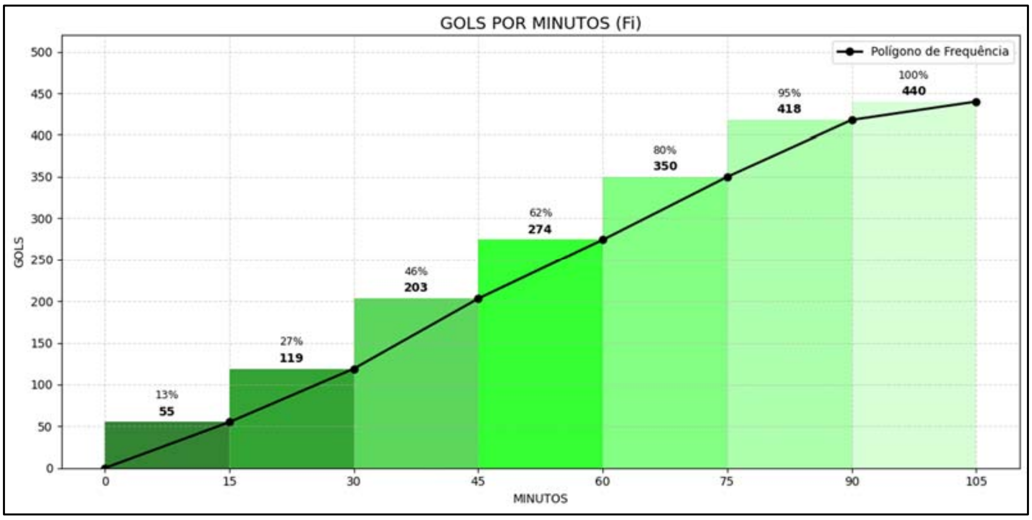
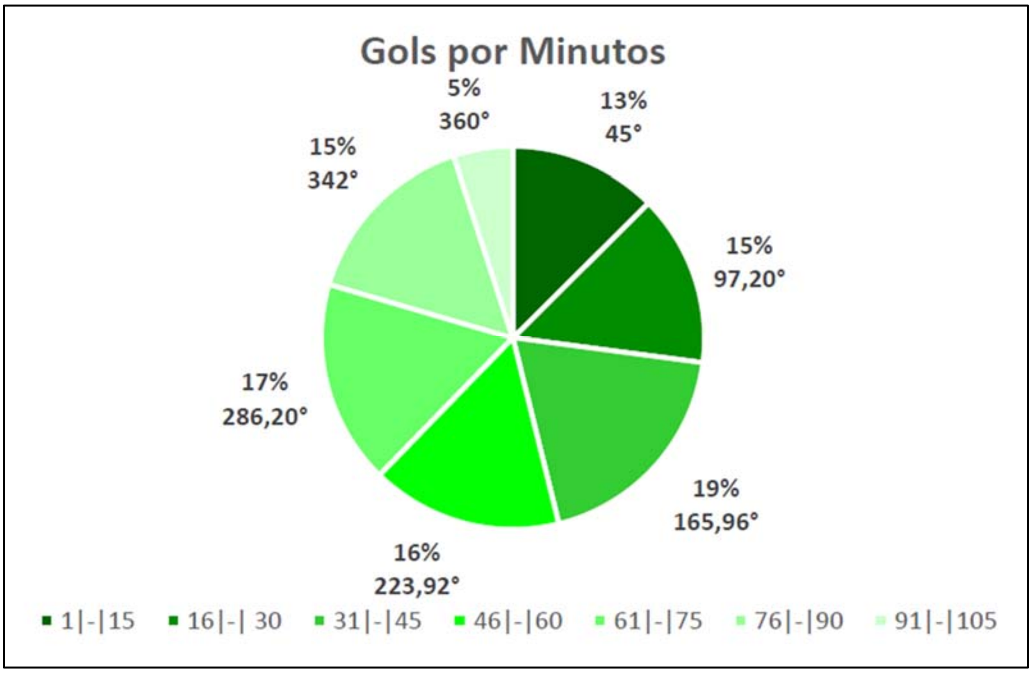


Gráfico 7 - Setores: Gols por minutos (FI)



CURIOSIDADES:

Figura 11 - Infográfico Neymar Seleção Brasileira (2014)

Fonte: Pinterest, 2025.

Figura 12 - Infográfico Neymar (2018)

Fonte: Pinterest, 2025.

Saiba mais em: https://www.transfermarkt.com.br/neymar/profil/spieler/68290

AULA 4: Medidas de Tendência Central
Medidas de Tendência Central (MTC)

Média

- É a soma de todos os valores de um conjunto dividida pelo número total de valores. É uma medida comum, mas pode ser afetada por valores extremos (outliers).
- Fórmula: $Média = \frac{\sum(fi \; x \; xi)}{\sum fi}$

Moda

- É o valor que aparece com mais frequência em um conjunto de dados. Um conjunto pode ter uma única moda (unimodal), mais de uma moda (multimodal) ou nenhuma moda (se todos os valores forem únicos).

Mediana

- É o valor que separa o conjunto de dados em duas partes iguais quando os dados estão ordenados. Se o número de observações for ímpar, a mediana é o valor do meio. Se for par, a mediana é a média dos dois valores centrais.
- Passo-a-passo:
  1. Ordenar os dados
  2. Se *n* (número total de valores) for ímpar, a mediana é o valor na posição $\frac{n+1}{2}$
  3. Se *n* for par, a mediana é a média dos valores nas posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

EXERCÍCIOS:

1- Construir a tabela de distribuição de frequência e fazer o histograma e o polígono fi.

Tabela 3 – TDF para medidas de tendência central

GOLS DE NEYMAR POR MINUTOS					
i	MINUTOS	GOLS (fi)	Fi	xi	fixi
1	1|-|15	55	55	8	440
2	16|-| 30	64	119	23	1472
3	31|-|45	84	203	38	3192
4	46|-|60	71	274	53	3763
5	61|-|75	76	350	68	5168
6	76|-|90	68	418	83	5644
7	91|-|105	22	440	98	2156
K = 7	hi = 14	440			21835

2- Calcular as medidas centrais: média (precisa), moda (imediata) e mediana (central).

$$Média = Me = \frac{21835}{440} = \mathbf{49{,}625\ minutos}$$

$$Moda\ (rápida) = Mo = \frac{(31 + 45)}{2} = \mathbf{38\ minutos}$$

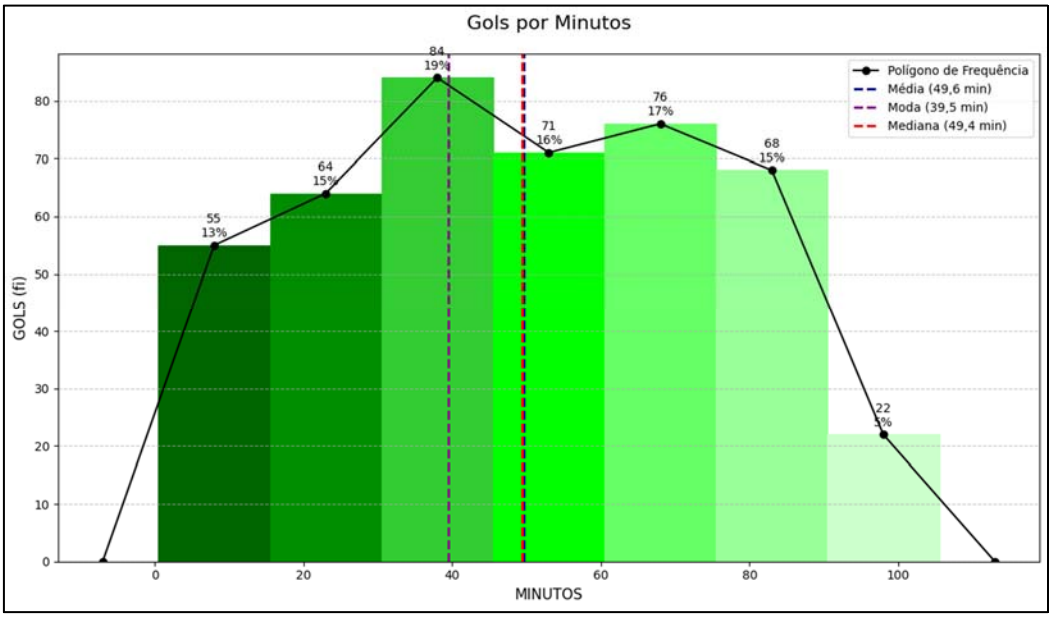
$$Moda\ (czuber) = Mo = 31 + \left(\frac{20}{33} \times 14\right) = \mathbf{39{,}485\ minutos}$$

$$Mediana = Md = 46 + \left[\frac{(220 - 203)}{71} \times 14\right] = \mathbf{49{,}352\ minutos}$$

$$Mo < Md < Me \rightarrow Tendência: Dir/+$$

3- Dar posição relativa Me, Mo e Md no histograma e polígono frequência absoluta (fi).

Gráfico 8 - Histograma + polígono + MTC



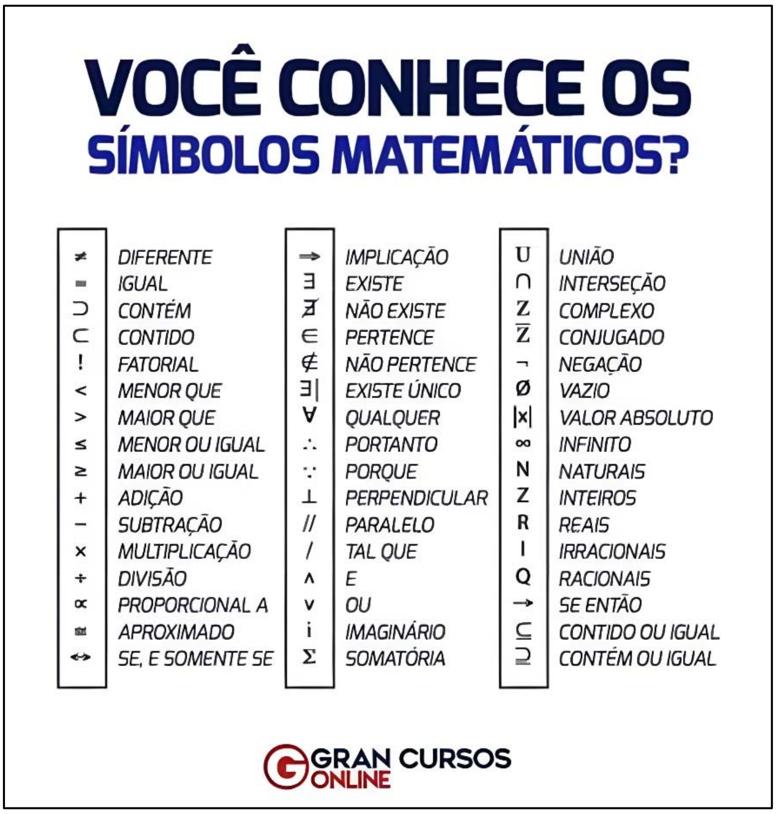
CURIOSIDADES:

Figura 13 - Alfabeto Grego

Αα	Ββ	Γγ	Δδ	Εε	Ζζ	Ηη	Θθ
Alpha	Beta	Gamma	Delta	Epsilon	Zeta	Eta	Theta
Ιι	Κκ	Λλ	Μμ	Νν	Ξξ	Οο	Ππ
Iota	Kappa	Lambda	Mu	Nu	Xi	Omicron	Pi
Ρρ	Σσς	Ττ	Υυ	Φφ	Χχ	Ψψ	Ωω
Rho	Sigma	Tau	Upsilon	Phi	Chi	Psi	Omega

Fonte: Mundo Educação, 2025.

Figura 14 - Simbologia Matemática



Fonte: Gran Cursos Online, 2025.

AULA 5: Separatrizes: Quartil, Quintil, Decil e Percentil

Separatrizes: são elementos que separam a sequência de valores em partes iguais, permitindo uma análise mais fácil da distribuição dos dados.

Quartis

Dividem os dados em quatro partes iguais.

- Q1 (primeiro quartil): 25% dos dados estão abaixo dele.
- Q2 (segundo quartil ou mediana): 50% dos dados estão abaixo dele.
- Q3 (terceiro quartil): 75% dos dados estão abaixo dele.

Quintis

Dividem os dados em cinco partes iguais.

- Cada quintil representa 20% dos dados. O primeiro quintil (K1) é o ponto que separa os 20% inferiores do restante.

Decis

Dividem os dados em dez partes iguais.

- Cada decil representa 10% dos dados. O primeiro decil (D1) é o ponto que separa os 10% inferiores.

Percentis

Dividem os dados em cem partes iguais.

- Cada percentil representa 1% dos dados. O percentil 25, por exemplo, é o valor abaixo do qual 25% dos dados se encontram.

Correspondência:

Q1 = P25	K1 = D2 = P20	D1 = P10
Q2 = P50 = Md	K2 = D4 = P40	D5 = Q2 = P50 = Md
Q3 = P75	K3 = D6 = P60	D9 = P90

EXERCÍCIOS:

1- Construir a tabela de distribuição de frequência e representar o polígono Fi;

Resposta:

Tabela 4 - TDF para separatrizes

GOLS DE NEYMAR POR MINUTOS			
i	MINUTOS	GOLS (fi)	Fi
1	1|-|15	55	55
2	16|-| 30	64	119
3	31|-|45	84	203
4	46|-|60	71	274
5	61|-|75	76	350
6	76|-|90	68	418
7	91|-|105	22	440
K = 7	hi = 14	440	

2- Determinar as seguintes separatrizes: Quartil (Q2), Quintil (K3), Decil (D9) e Percentil (P30);

Resposta:

$$\boldsymbol{Q2} = \frac{2\sum fi}{4} = \frac{2(440)}{4} = \frac{880}{4} = 220$$

$$\boldsymbol{Q2} = 46 + \left[\frac{220-203}{71} \times 14\right] = 46 + 3{,}352 = \boldsymbol{49,352\ minutos = 50\%}$$

$$\boldsymbol{K3} = \frac{3\sum fi}{5} = \frac{3(440)}{5} = \frac{1320}{5} = 264$$

$$\boldsymbol{K3} = 46 + \left[\frac{264-203}{71} \times 14\right] = 46 + 12{,}028 = \boldsymbol{58,028\ minutos = 60\%}$$

$$\boldsymbol{D9} = \frac{9\sum fi}{10} = \frac{9(440)}{10} = \frac{3960}{10} = 396$$

$$\boldsymbol{D9} = 76 + \left[\frac{396-350}{68} \times 14\right] = 76 + 9{,}470 = \boldsymbol{85,470\ minutos = 90\%}$$

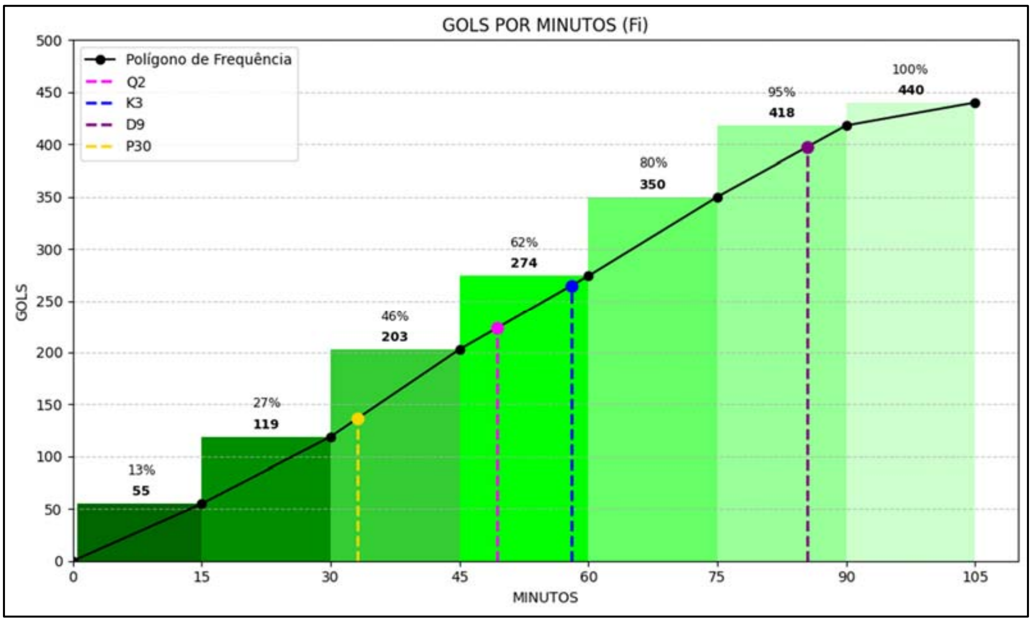
$$\boldsymbol{P30} = \frac{30\sum fi}{100} = \frac{30(440)}{100} = \frac{13200}{100} = 132$$

$$\boldsymbol{P30} = 31 + \left[\frac{132-119}{84} \times 14\right] = 31 + 2{,}166 = \boldsymbol{33,166\ minutos = 30\%}$$

3- Posicionar os resultados Q2, K3, D9, P90 no polígono frequência acumulada (Fi).

Resposta:

Gráfico 9 – Histograma + polígono + separatrizes



CURIOSIDADES:

Unidades de Medida

Unidades de medida são padrões que permitem quantificar propriedades como comprimento, massa, volume, tempo e capacidade.

Algumas unidades de medida são:

Quilômetro (km), Hectômetro (hm), Decâmetro (dam), Metro (m), Decímetro (dm), Centímetro (cm), Milímetro (mm), Quilograma (kg), Segundo (s), Litro (L).

Regras de arredondamento

As regras de arredondamento dependem do algarismo imediatamente à direita dos que serão mantidos.

Para o arredondamento, a referência é o algarismo 5.

- **Algarismo à direita menor que 5**

  Os algarismos mantidos são conservados sem alteração.

Exemplo: Arredondar **5,1328** com duas casas decimais.

Os algarismos na primeira e segunda casa decimal são o 1 e o 3. Logo após, o 2 está à direita. Como 2 é menor que 5, não se altera o 3. Assim, o 1 e o 3 são mantidos.

O arredondamento com duas casas decimais é: **5,13**.

- **Algarismo à direita maior que 5**

Acrescenta-se uma unidade ao último algarismo mantido.

Exemplo: Arredondar **1,236** com duas casas decimais.

As primeiras duas casas decimais são ocupadas pelos algarismos 2 e o 3. À direita do 3 está o 6. Verifica-se que 6 é maior que 5. Desta forma, é acrescentado uma unidade ao último algarismo que ficará o 3.

Como 3 + 1 = 4, substitui-se o 3 pelo 4, assim: **1,24**.

- **Algarismo à direita igual a 5**

Quando o algarismo à direita dos que serão mantidos é cinco, há três possibilidades:

Se houver pelo menos um algarismo diferente de zero, depois do 5, acrescenta-se uma unidade à última casa decimal mantida.

Caso 1

Exemplo: Arredondar **1,35000001** com uma casa depois da vírgula.

Com uma casa decimal, inicialmente o número fica: 1,3.

Verifica-se que à direita do 3 está o 5. Também, que há pelo menos um algarismo diferente de 0, no caso, o 1.

Desta forma, acrescenta-se uma unidade ao 3. Como 3 + 1 = 4, o número fica: **1,4**.

Caso 2

Se todos os algarismos após o 5 são zeros e o número mantido é ímpar, soma-se uma unidade à última casa decimal mantida e descartam-se os posteriores.

Exemplo: Arredondar **4,2750** com duas casas decimais.

Com duas casas decimais, o número é: 4,27.

Verifica-se que à direita do 7 está o 5. Também, que não há algarismos diferentes de 0 após o 5. Por fim, que 7 é ímpar.

Com estas condições, é somado uma unidade ao último algarismo que fica, no caso, o 7. Como 7 + 1 = 8, o número 4,2750 arredondado com duas casas decimais é: **4,28.**

Caso 3

Se todos os algarismos após o 5 são zeros e o número mantido é par, não há mudança no número mantido e este permanece o mesmo. Os posteriores são descartados.

Exemplo: Arredondar **16,825000** com duas casas decimais.

Com duas casas decimais, o número fica: 16,82.

Verifica-se que à direita do 2 está o algarismo 5 e que, após, não há algarismos diferentes de 0.

Como o último algarismo que fica é o 2, que é par, ele é mantido sem alteração.

Arredondado com duas casas decimais, o número 16,82500000 fica: **16,82**.

AULA 6: Medidas de Dispersão ou Variabilidade

Medidas de Dispersão ou Variabilidade

- São medidas fundamentais para entender a dispersão dos dados em um conjunto.

Desvio Padrão:

- É a raiz quadrada da variância e fornece uma medida de dispersão dos dados que está na mesma unidade dos dados originais.
- Fórmula:

$$\boldsymbol{DP\ Normal} = S_N = \sqrt{\frac{\sum fi(xi)^2}{\sum fi} - \left(\frac{\sum fixi}{\sum fi}\right)^2}$$

$$\boldsymbol{DP\ Breve} = S_B = hi \times \sqrt{\frac{\sum fi(yi)^2}{\sum fi} - \left(\frac{\sum fiyi}{\sum fi}\right)^2}$$

Variância:

- É uma medida que indica o quanto os dados de um conjunto se dispersam em relação à medida.
- Fornece a ideia de quão longe os valores estão da média.
- Fórmula: $S^2 = S_N \times S_B$

Coeficiente de Variação

- É uma medida relativa de dispersão e é calculado como a razão do desvio padrão pela média, expressa em porcentagem
- Permite comparar a variabilidade entre diferentes conjuntos de dados com médias diferentes.
- Fórmula: $CV = \left(\frac{S}{Me}\right) x\ 100(\%)$
- Onde: S é $o\ desvio\ padrão.$

  Me é $a\ média.$
- Um CV mais alto indica maior variabilidade em relação à média, enquanto um CV mais baixo sugere menor variabilidade.

Tipo de dispersão:

Cv < 5% (baixa)

5% < Cv < 30% (moderada)

30% < Cv < 50% (alta)

Cv > 50% (muito alta)

EXERCÍCIOS:

1- Construir a tabela de distribuição de frequência;

Resposta:

Tabela 5 - TDF para medidas de dispersão

GOLS DE NEYMAR POR MINUTOS										
i	MINUTOS	GOLS (fi)	xi	fi xi	xi²	fi(xi)²	yi	fi yi	yi²	fi(yi)²
1	1|-|15	55	8	440	64	3520	-2	-110	4	220
2	16|-| 30	64	23	1472	529	33856	-1	-64	1	64
3	31|-|45	84	38	3192	1444	121296	0	0	0	0
4	46|-|60	71	53	3763	2809	199439	1	71	1	71
5	61|-|75	76	68	5168	4624	351424	2	152	4	304
6	76|-|90	68	83	5644	6889	468452	3	204	9	612
7	91|-|105	22	98	2156	9604	211288	4	88	16	352
K = 7	hi = 14	440		21835		1389275		341		1623

2- Obter o Desvio Padrão (S) (processo normal e breve), Variância (S²) e Coeficiente de variação (Cv);

Resposta:

$$\boldsymbol{DP\ Normal} = S_N = \sqrt{\frac{1389275}{440} - \left(\frac{21835}{440}\right)^2} = \sqrt{3157{,}443 - 2462{,}640} = \mathbf{26{,}359}\ \boldsymbol{minutos}$$

$$\boldsymbol{DP\ Breve} = S_B = 14 \times \sqrt{\frac{1623}{440} - \left(\frac{341}{440}\right)^2} = 14 \times \sqrt{3{,}688 - 0{,}600} = \mathbf{24{,}601}\ \boldsymbol{minutos}$$

$$\boldsymbol{S^2} = 26{,}359^2 = 694{,}796 \cong \mathbf{695}\ \boldsymbol{minutos}$$

$$\boldsymbol{Cv} = \frac{26{,}359}{49{,}625} \times 100(\%) = \mathbf{53{,}11}\%$$

3- Aplicar a regra empírica e classificar o coeficiente de variação quanto ao tipo de dispersão.

Resposta:

Coeficiente de Variação = 53,11%. Dispersão **muito alta**.

CURIOSIDADES:

As medidas de dispersão podem ser aplicadas em diversas situações cotidianas, como no contexto dos investimentos, por exemplo. Para ilustrar essa aplicação, utilizaremos o desvio padrão.

Consideremos dois tipos de aplicações que apresentam, em média, o mesmo rendimento. A diferença entre elas reside no fato de que uma apresenta resultados consistentes, enquanto a outra exibe variações significativas - oscilando constantemente. Mesmo com a mesma média, a aplicação que apresenta maior variabilidade é considerada mais arriscada. Isso ocorre porque o desvio padrão indica o grau em que os valores se afastam da média. Assim, investidores tendem a analisar não apenas o potencial de rendimento do capital, mas também a possibilidade de oscilações excessivas. Dessa forma, é possível optar por alternativas mais seguras, com menores surpresas ao longo do percurso.

Em síntese, neste exemplo, o desvio padrão é utilizado para mensurar a variação dos valores em relação à média. Quanto maior for esse índice, maior será o risco e a instabilidade dos resultados.

AULA 7: Coeficientes Estatísticos: Assimetria e Curtose

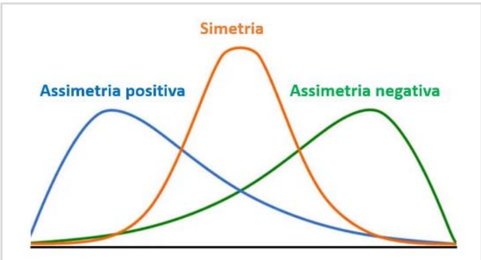
Coeficiente de Assimetria e Coeficiente de Curtose

São medidas estatísticas importantes para descrever a forma de uma distribuição de dados.

Coeficiente de Assimetria

- Mede a simetria da distribuição dos dados.
- Indica se os dados estão distribuídos de maneira simétrica em torno da média ou se há uma inclinação em uma direção específica.
- **Assimetria positiva:** se o coeficiente for maior que 0, isso indica que a cauda direita da distribuição é mais longa ou mais pesada, ou seja, há valores extremos maiores.
- **Assimetria negativa:** se o coeficiente for menor que 0, isso indica que a cauda esquerda da distribuição é mais longa ou mais pesada, ou seja, há valores extremos menores.
- **Simetria:** um coeficiente igual a 0 sugere que a distribuição é simétrica.

Figura 15 - Exemplo de Assimetrias



Fonte: Sonia Vieira, 2025.

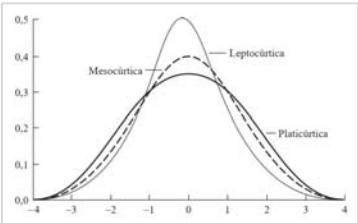
- Fórmula tipo de assimetria: $Ta = (Me - Mo)\ ou\ Ta = \frac{(Me-Mo)}{S}$
- Fórmula coeficiente de assimetria: $Ca = \frac{3\times(Me-Md)}{S}$

Coeficiente de Curtose:

- Mede a "altura" e a "largura" da distribuição de dados.
- Indica se os dados têm caudas pesadas (muitos valores extremos) ou leves (poucos valores extremos) em comparação com uma distribuição normal.
- **Curtose alta (leptocúrtica):** se o coeficiente for maior que 0,263, isso indica que a distribuição tem caudas pesadas e um pico mais alto do que uma distribuição normal. Isso significa que há mais valores extremos.

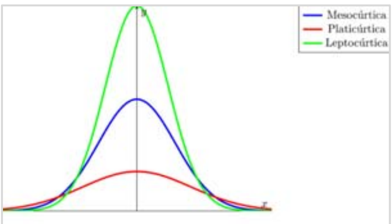
- **Curtose baixa (platicúrtica):** se o coeficiente for menor que 0,263, isso indica que a distribuição tem caudas leves e um pico mais baixo do que uma normal. Isso significa que há menos valores extremos.
- **Curtose normal (nula/mesocúrtica):** um coeficiente igual a 0,263 sugere uma distribuição normal.

Figura 16 - Exemplo de curtose 1



Fonte: Blog Prof Fernanda Maciel, 2025.

Figura 17 - Exemplo de curtose 2



Fonte: Estatística para Concurso, 2025.

- Fórmula: $Cc = \frac{(Q3-Q1)}{2\times(D9-D1)}$

Relação entre Assimetria e Curtose

- A **assimetria** está relacionada à direção da distribuição, enquanto a **curtose** está relacionada à forma e ao peso das caudas.
- Ambas as medidas são úteis para entender melhor a natureza dos dados e podem influenciar técnicas estatísticas e inferências.

EXERCÍCIOS:

1- Construir a tabela de distribuição de frequência;

Resposta:

Tabela 6 - TDF para Coeficientes estatísticos

GOLS DE NEYMAR POR MINUTOS										
i	MINUTOS	GOLS (fi)	xi	fi xi	xi²	fi(xi)²	yi	fi yi	yi²	fi(yi)²
1	1|-|15	55	8	440	64	3520	-2	-110	4	220
2	16|-| 30	64	23	1472	529	33856	-1	-64	1	64
3	31|-|45	84	38	3192	1444	121296	0	0	0	0
4	46|-|60	71	53	3763	2809	199439	1	71	1	71
5	61|-|75	76	68	5168	4624	351424	2	152	4	304
6	76|-|90	68	83	5644	6889	468452	3	204	9	612
7	91|-|105	22	98	2156	9604	211288	4	88	16	352
K = 7	**hi = 14**	**440**		**21835**		**1389275**		**341**		**1623**

2- Calcular o tipo e coeficiente de assimetria (Ta, Ca) e curtose (Cc);

Reposta:

$$\boldsymbol{Ta} = (49{,}625 - 39{,}485) = \mathbf{10,14}\ ou\ \boldsymbol{Ta} = \frac{(Me - Mo)}{S} = \frac{10{,}14}{26{,}359} = \mathbf{0,384}$$

$$\boldsymbol{Ca} = \frac{3 \times (Me - Md)}{S} = \frac{3 \times (49{,}625 - 49{,}352)}{26{,}359} = \mathbf{0,0310}$$

$$\boldsymbol{Cc} = \frac{(Q3 - Q1)}{2 \times (D9 - D1)} = \frac{(71{,}315 - 28{,}031)}{2 \times (85{,}470 - 12{,}20)} = \frac{43{,}284}{146{,}54} = \mathbf{0,295}$$

$$Q3 = \frac{3\sum fi}{4} = \frac{3(440)}{4} = 330$$

$$Q3 = 61 + \left[\frac{330 - 274}{76} \times 14\right] = \mathbf{71,315}\ \boldsymbol{minutos}$$

$$Q1 = \frac{\sum fi}{4} = \frac{440}{4} = 110$$

$$Q1 = 16 + \left[\frac{110 - 55}{64} \times 14\right] = \mathbf{28,031}\ \boldsymbol{minutos}$$

$$D9 = \frac{9\sum fi}{10} = \frac{9(440)}{10} = \frac{3960}{10} = 396$$

$$D9 = 76 + \left[\frac{396 - 350}{68} \times 14\right] = 76 + 9{,}470 = \mathbf{85,470}\ \boldsymbol{minutos} = \mathbf{90\%}$$

$$D1 = \frac{\sum fi}{10} = \frac{440}{10} = 44$$

$$D1 = 1 + \left[\frac{44 - 0}{55} \times 14\right] = \mathbf{12,20\ minutos}$$

3- Aplicar a regra empírica de dispersão e classificar quanto à assimetria e curtose (grau achatamento).

Reposta:

$Ta \rightarrow assimetria\ Dir/+$

$Ca \rightarrow assimetria\ Dir/+$

$Cc \rightarrow alta\ (leptocúrtica)$

CURIOSIDADES:

De certa forma, os coeficientes estatísticos estão extremamente presentes na contemporaneidade, especialmente quando abordamos a tecnologia como um campo de comparação. Os coeficientes estatísticos são de suma importância na área tecnológica, sobretudo ao lidarmos com grandes volumes de dados.

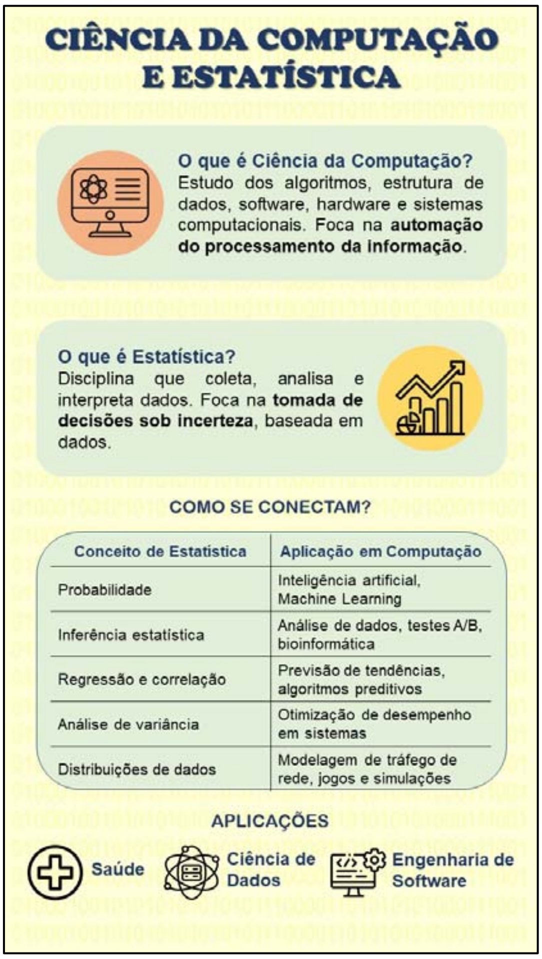
Por exemplo, o coeficiente de correlação pode ser utilizado para entender se existe uma relação entre o comportamento dos usuários em um aplicativo e o tempo que eles dedicam a ele. Em áreas como aprendizado de máquina (machine learning), esses coeficientes auxiliam na identificação das variáveis mais relevantes para o desempenho de um modelo, como na previsão da probabilidade de um cliente cancelar um serviço.

Em última análise, esses coeficientes constituem ferramentas poderosas que ajudam desenvolvedores e analistas a tomar decisões mais informadas e precisas, desde o desenvolvimento de sistemas até a criação de algoritmos de recomendação, como os utilizados no Netflix ou Spotify.

Concluindo, neste exemplo, empregamos o coeficiente de correlação, que funciona como um termômetro para medir a relação entre dois comparativos. No caso do aplicativo, ele indica se quanto mais tempo o usuário passa no app, maior é a probabilidade de interação ou realização de compras. Isso auxilia os desenvolvedores a compreender padrões e aprimorar a experiência do usuário.

AULA 8: Infográficos e Acrósticos

Figura 18 - Infográfico 1



Possibilidades em jogo,
Riscos medidos com lógica,
Ocorrência sob incerteza,
Base para decisões racionais.
Alvo da teoria estatística,
Buscando prever o improvável,
Interpretando eventos futuros,
Lançando dados e hipóteses,
Influencia escolhas reais,
Desenha cenários potenciais,
Análise que guia o acaso,
Dados transformados em chances,
Estratégia em cada número,
Sabedoria entre o aleatório.

Informação clara e visual,
Navegando entre dados e ideias,
Facilitando a compreensão,
Organiza o que parecia confuso.
Gráficos, textos e ícones,
Revelam padrões escondidos,
Apresentando com estética,
Fundindo ciência e arte,
Ilustra com precisão,
Conecta dados ao leitor,
Oferecendo valor imediato,
Simplifica sem perder conteúdo.

Fenômenos que se repetem,
Rítmicos como batimentos,
Expressos em números claros,
Quantificando repetições,
Unindo tempo e estatística,
Ênfase no padrão observado,
Notas que compõem o ruído,
Ciclos em dados contínuos,
Indicando comportamentos,
Aplicada em som, ondas e dados.

CURIOSIDADES:

Lei dos Grandes Números

A Lei dos Grandes Números é um princípio fundamental na teoria da probabilidade que descreve o resultado de realizar a mesma experiência repetidamente. Ela afirma que, à medida que o número de experimentos aumenta, a média dos resultados obtidos tende a se aproximar da média esperada (ou valor esperado) da distribuição de probabilidade da variável aleatória envolvida.

Tipos de Lei dos Grandes Números

Existem duas versões principais da Lei dos Grandes Números:

1. Lei Fraca dos Grandes Números: esta versão afirma que, para qualquer número real positivo $\in$, a probabilidade de que a média das amostras se desvie da média esperada por mais de $\in$ tende a zero à medida que o número de amostras n aumenta. Em outras palavras, conforme coletamos mais dados, a média amostral se aproxima da média populacional.

2. Lei Forte dos Grandes Números: esta versão é mais rigorosa e afirma que a média das amostras converge quase seguramente (ou seja, com probabilidade 1) para a média esperada à medida que o número de amostras vai para o infinito. Isso significa que, em um número suficientemente grande de repetições, a média será praticamente igual à média esperada.

AULA 9: Probabilidades: Distribuição Binomial e Poisson
Binômio

O Teorema do Binômio, ou Binômio de Newton, é uma ferramenta poderosa para expandir expressões da forma $(a+b)^n$.

O desenvolvimento do termo geral do binômio de Newton é dado pela fórmula:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Onde $\binom{n}{k}$ é o coeficiente binomial, que pode ser calculado como:

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ Essa fórmula permite expandir potências de binômios de forma sistemática.

Poisson

A distribuição de Poisson é uma distribuição de probabilidade discreta que expressa a probabilidade de um determinado número de eventos ocorrer em um intervalo fixo de tempo ou espaço, dado que esses eventos ocorrem com uma taxa média conhecida e independentemente do tempo desde o último evento.

Fórmula:

$$P(X=k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$$

Fatorial

O fatorial de um número inteiro não negativo *n*, denotado por *n!*, é o produto de todos os inteiros positivos até *n*. A definição é:

$n! = n \times (n-1) \times (n-2) \times \ldots \times 1$

Propriedades do fatorial:

1. Recursividade: a relação recursiva é dada por: $n! = n \times (n-1)!$ – Isso nos ajuda a calcular fatoriais maiores a partir dos menores.
2. Crescimento: o fatorial cresce muito rapidamente à medida que n aumenta. Esses conceitos são fundamentais em combinatória, probabilidade e análise matemática.

Por que se considera que 0! = 1?

- Essa definição é útil em várias áreas da matemática, especialmente em combinações e permutações.
- A razão por trás disso é que existe exatamente uma maneira de organizar zero elementos: não fazendo nada. Além disso, a fórmula que define o fatorial de um número inteiro positivo $n!$ pode ser expressa recursivamente como $n! = n \times (n-1)!$
- Se aplicarmos isso para n = 1:

  $1! = 1 \times 0!$

  Para que essa equação seja verdadeira, precisamos que 0! = 1.

Potência: por que todo número elevado a zero é igual a um?

- Todo número elevado a zero é igual a um porque, ao considerar a definição de potências, temos:

$$a^m \div a^m = a^{m-m} = a^0$$

- Por outro lado, sabemos que qualquer número dividido por ele mesmo (exceto zero) é igual a um:

$$a^m \div a^m = 1$$

  Portanto concluímos que $a^0 = 1$. Essa convenção se aplica para qualquer número real a, exceto quando $a = 0$, pois 0^0 é indeterminado em muitos contextos.

Constante de Euler (e)

A constante de Euler é aproximadamente igual a 2,718281... e é uma base importante em matemática, especialmente em cálculo e análise. Ela surge naturalmente em várias situações, como no cálculo de juros compostos e na solução de equações diferenciais. Uma das definições mais comuns de e é através da série infinita:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots,$$

EXERCÍCIOS:

1- Em uma rede de computadores, sabe-se que 15% dos servidores estão suscetíveis a falhas críticas durante uma atualização de software. Se selecionarmos 20 servidores ao acaso, qual a probabilidade de que pelo menos 5 deles apresentem falhas críticas?

Dados:

Seja *X* o número de servidores com falha.

Sabemos que *X* ~ *Binomial* (n = 20, p = 0,15).

Calcule: $P(X \geq 5)$

Resposta:

$P(X = 0)$

$$\binom{20}{0} = \frac{20!}{(20-0)! \times 0!} = 1$$

$$\binom{20}{0} \times (0{,}15)^0 \times (0{,}85)^{20} = 1 \times 1 \times (0{,}85)^{20} \cong 0{,}0387$$

$P(X = 1)$

$$\binom{20}{1} = \frac{20!}{(20-1)! \times 1!} = 20$$

$$\binom{20}{1} \times (0{,}15)^1 \times (0{,}85)^{20-1} = 20 \times 0{,}15 \times 0{,}0455 \cong 0{,}1365$$

$P(X = 2)$

$$\binom{20}{2} = \frac{20!}{(20-2)! \times 2!} = 190$$

$$\binom{20}{2} \times (0{,}15)^2 \times (0{,}85)^{20-2} = 190 \times 0{,}0225 \times 0{,}0536 \cong 0{,}2291$$

$P(X = 3)$

$$\binom{20}{3} = \frac{20!}{(20-3)! \times 3!} = 1140$$

$$\binom{20}{3} \times (0{,}15)^3 \times (0{,}85)^{20-3} = 1140 \times 0{,}003375 \times 0{,}0631 \cong 0{,}2427$$

$P(X = 4)$

$$\binom{20}{4} = \frac{20!}{(20-4)! \times 4!} = 4845$$

$\binom{20}{4} \times (0{,}15)^4 \times (0{,}85)^{20-4} = 4845 \times 0{,}0005 \times 0{,}0742 \cong 0{,}1797$

$P(X \leq 4) =$

$[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$

$0{,}0387 + 0{,}1365 + 0{,}2291 + 0{,}2427 + 0{,}1797 = 0{,}8267$

$\boldsymbol{P(X \geq 5)} =$

$1 - [P(X \leq 4)] = 1 - 0{,}8267 = \mathbf{0{,}1733 = 17{,}33}\%$

Conclusão: A probabilidade de que pelo menos 5 servidores apresentem falhas é de 17,33%.

2- Em um sistema de transmissão de pacotes, a taxa média de pacotes com erro é de 0,3 por segundo. Qual é a probabilidade de que, em um segundo, sejam observados:

Seja *X* ~ *Poisson* (λ = 0,3).

a) Exatamente dois pacotes com erro?

Resposta: A probabilidade de que sejam observados 2 pacotes com erros em 1 segundo é de 3,33%.

$$P(X = 2) = \frac{0{,}3^2 \times 0{,}741}{2!} = \frac{0{,}09 \times 0{,}741}{2} = \mathbf{0{,}0333 = 3{,}33}\%$$

b) Exatamente um pacote com erro?

Resposta: A probabilidade de que seja observado apenas 1 pacote com erro em 1 segundo é de 22,23%.

$$P(X = 1) = \frac{0{,}3^1 \times 0{,}741}{1!} = \frac{0{,}3 \times 0{,}741}{1} = \mathbf{0{,}2223 = 22{,}23}\%$$

c) Nenhum pacote com erro?

Resposta: A probabilidade de sejam observados nenhum pacote com erro em 1 segundo é de 74,10%.

$$P(X=0)=\frac{0{,}3^0 \times 0{,}741}{0!}=\frac{1 \times 0{,}741}{1}=\mathbf{0{,}741}=\mathbf{74{,}10}\%$$

CURIOSIDADES:

Figura 19 - Distribuições

Distribuições	Função de probabilidade
BERNOULLI	$p(k)=p^k\,(1-p)^{1-k}$
BINOMIAL	$p(k)=\frac{n!}{k!\,(n-k)!}p^k\,(1-p)^{n-k}$
POISSON	$p(k)=\frac{e^{-\lambda}\,\lambda^k}{k!}$
GEOMETRICA	$p(k)=pq^{k-1}$
HIPERGEOMETRICA	$p(k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$
BINOMIAL NEGATIVA	$p(k)=\binom{k-1}{r-1}\,p^r\,(1-p)^{k-r}$

Fonte: Geokrigagem, 2025

Teorema de Bayes

O Teorema de Bayes é uma das ferramentas mais importantes na estatística e na teoria da probabilidade. Ele fornece um método para atualizar a probabilidade de uma hipótese à medida que novas evidências se tornam disponíveis.

1. Definição do Teorema de Bayes

É expresso matematicamente da seguinte forma:

$$P(A|B)=\frac{P(B|A)\times P(A)}{P(B)}$$

Sendo:

A, B → eventos

P(A|B) → probabilidade de A se B for verdade

P(B|A) → probabilidade de B se A for verdade

P(A), P(B) → probabilidades independentes de A e B

2. Aplicações do Teorema de Bayes

O Teorema de Bayes tem várias aplicações em diferentes campos:

- **Medicina:** Diagnóstico médico e interpretação de testes.

- **Machine Learning:** Algoritmos como Naive Bayes, usados para classificação e previsão.

- **Ciência Forense:** Avaliação da evidência em investigações criminais.

- **Economia e Finanças:** Modelagem de incertezas em mercados financeiros.

3. Vantagens e Desvantagens

Vantagens:

- Permite incorporar novas evidências; Flexível e aplicável em diversas áreas.

Desvantagens:

- A seleção das probabilidades anteriores pode ser subjetiva; Pode ser computacionalmente intensivo para modelos complexos.

AULA 10: Distribuição Normal: Tabela da Variável "Z"
Distribuição Normal Padrão

A Distribuição Normal Padrão é uma distribuição de probabilidade contínua com média (μ) igual a 0 e desvio padrão (σ) igual a 1. É representada por N(0,1) e possui uma curva simétrica em forma de sino.

Essa distribuição é fundamental em estatística, pois muitos fenômenos naturais e sociais seguem esse padrão. A área total sob a curva é igual a 1, representando 100% da probabilidade.

Tabela de Distribuição Normal Padronizada

A Tabela de Distribuição Normal Padronizada, também conhecida como Tabela Z, fornece as áreas acumuladas sob a curva normal padrão para diferentes valores de z (escores Z). Essas tabelas são essenciais para calcular probabilidades associadas a valores específicos em distribuições normais.

Por exemplo, para encontrar a probabilidade de um valor ser menor que z = 1,26, consultamos a tabela, que nos dá a área correspondente.

Distribuição Normal Padrão Acumulada e Reduzida

A **Distribuição Normal Padrão Acumulada** refere-se à função de distribuição acumulada (CDF) da normal padrão, que calcula a probabilidade de uma variável aleatória ser menor ou igual a um valor específico. Matematicamente, é expressa como:

$$F(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Já a **Distribuição Normal Padrão Reduzida** envolve a padronização de uma variável aleatória X, transformando-a em Z utilizando a fórmula:

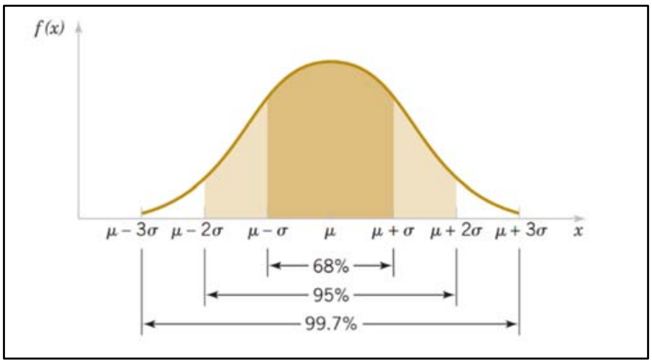
$$Z = \frac{X - \mu}{\sigma}$$

Essa transformação permite comparar diferentes distribuições normais, convertendo-as para a distribuição normal padrão.

Área sob a Curva Normal Padrão

A Área sob a Curva Normal Padrão representa a probabilidade acumulada até um determinado valor de z. Por exemplo, a área até z = 1,26 é 0,8962, indicando que há 89,62% de chance de uma variável aleatória da normal padrão ser menor que 1,26.

Figura 20 - Exemplo de Área sob a Curva Normal Padrão



Fonte: Moodle UFSC, 2025.

EXERCÍCIOS:

Considerando:

Média = 49,625 minutos

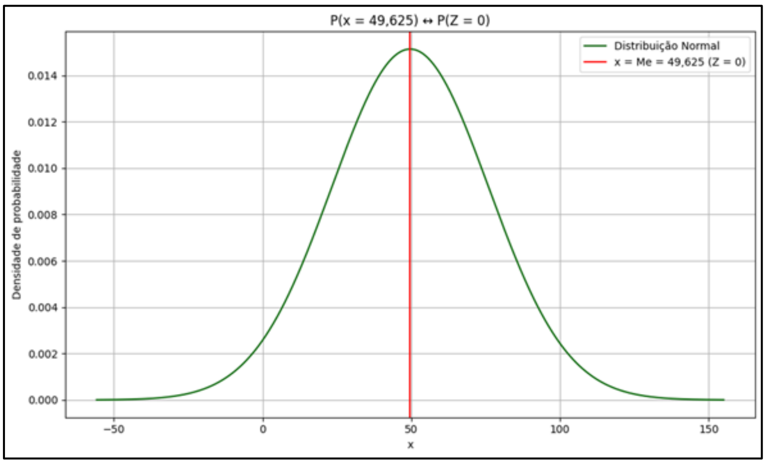
Desvio Padrão = 26,359 minutos

1- $P(x = Me) \rightarrow P(x = 49{,}625) \leftrightarrow \boldsymbol{P(Z = 0)}$

$$Z = \frac{X - \mu}{S} = \frac{49{,}625 - 49{,}625}{26{,}359} = 0$$

$Z = 0 = \mathbf{0{,}00}$

Gráfico 10 - Probabilidade para Z = 0



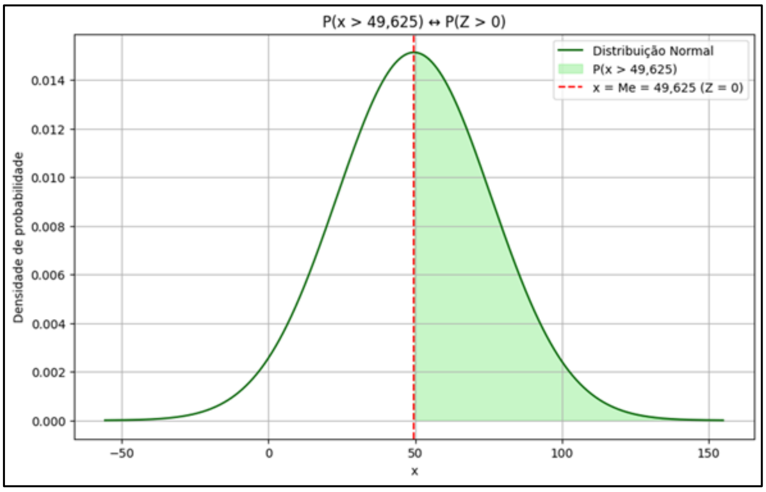
2- $P(x > Me) \rightarrow P(x > 49{,}625) \leftrightarrow \boldsymbol{P(Z > 0)}$

$$Z = \frac{X - \mu}{S} = \frac{49{,}625 - 49{,}625}{26{,}359} = 0$$

$Z = 0 = 0{,}00$

$P(Z > 0) = 0{,}5 - 0{,}00 = \mathbf{0{,}5} = \mathbf{50}\%$

Gráfico 11 - Probabilidade para Z > 0



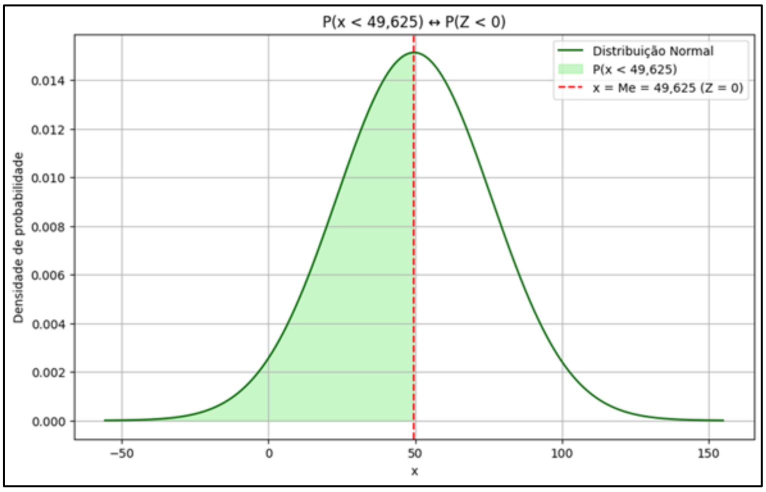
3- $P(x < Me) \rightarrow P(x < 49{,}625) \leftrightarrow \boldsymbol{P(Z < 0)}$

$$Z = \frac{X - \mu}{S} = \frac{49{,}625 - 49{,}625}{26{,}359} = 0$$

$Z = 0 = 0{,}00$

$P(Z < 0) = 0{,}5 + 0{,}00 = \mathbf{0{,}5} = \mathbf{50}\%$

Gráfico 12 - Probabilidade para Z < 0



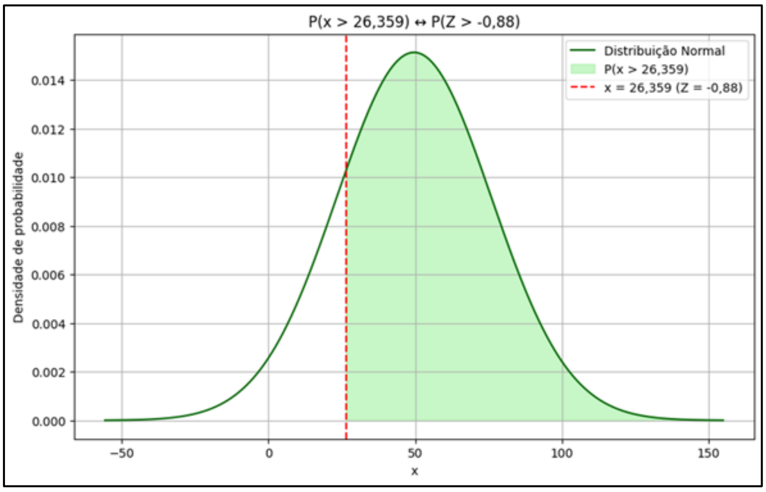
4- $P(x > S) \rightarrow P(x > 26{,}359) \leftrightarrow \boldsymbol{P(Z > -0{,}88)}$

$$Z = \frac{X - \mu}{S} = \frac{26{,}359 - 49{,}625}{26{,}359} = -0{,}88$$

$$Z = -0{,}88 = 0{,}3106$$

$$P(Z > -0{,}88) = 0{,}5 + 0{,}3106 = \mathbf{81{,}06}\%$$

Gráfico 13 - Probabilidade para Z > -0,88



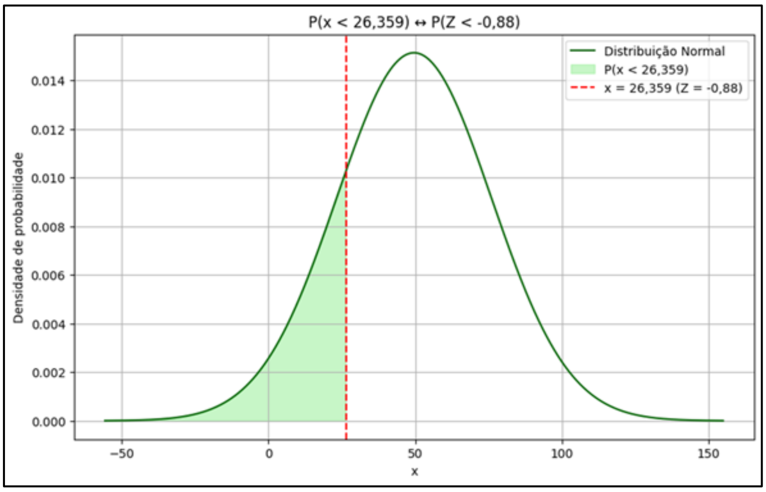
5- $P(x < S) \rightarrow P(x < 26{,}359) \leftrightarrow \boldsymbol{P(Z < -0{,}88)}$

$$Z = \frac{X - \mu}{S} = \frac{26{,}359 - 49{,}625}{26{,}359} = -0{,}88$$

$$Z = -0{,}88 = 0{,}3106$$

$$P(Z < -0{,}88) = 0{,}5 - 0{,}3106 = \mathbf{18{,}94}\%$$

Gráfico 14 - Probabilidade para Z < -0,88



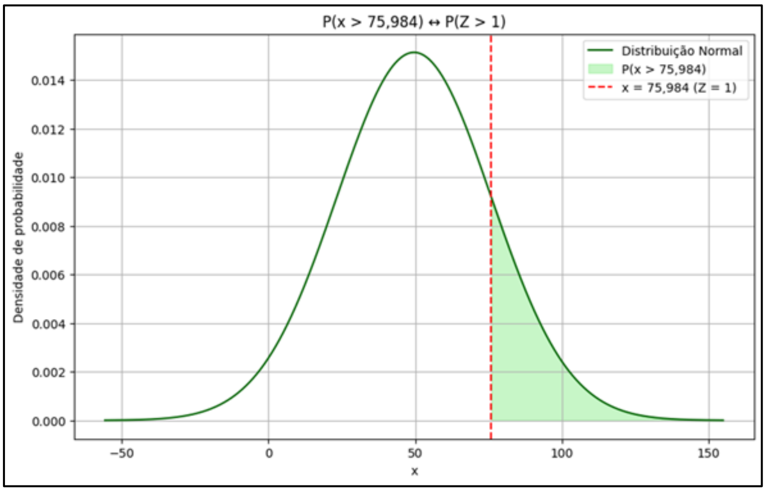
6- $P(x > Me + S) \rightarrow P(x > 75{,}984) \leftrightarrow \boldsymbol{P(Z > 1)}$

$$Z = \frac{X - \mu}{S} = \frac{75{,}984 - 49{,}625}{26{,}359} = 1$$

$Z = 1 = 0{,}3413$

$P(Z > 1) = 0{,}5 - 0{,}3413 = \mathbf{15{,}87}\%$

Gráfico 15 - Probabilidade para Z > 1



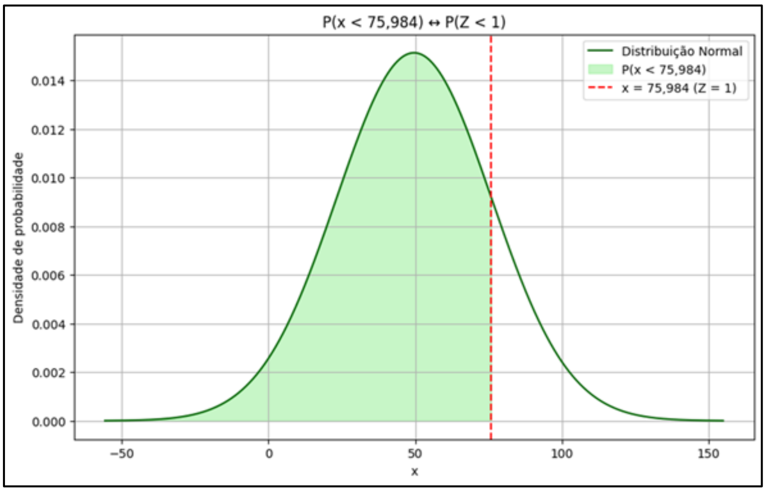
7- $P(x < Me + S) \rightarrow P(x < 75{,}984) \leftrightarrow \boldsymbol{P(Z < 1)}$

$$Z = \frac{X - \mu}{S} = \frac{75{,}984 - 49{,}625}{26{,}359} = 1$$

$Z = 1 = 0{,}3413$

$P(Z < 1) = 0{,}5 + 0{,}3413 = \mathbf{84{,}13}\%$

Gráfico 16 - Probabilidade para Z < 1



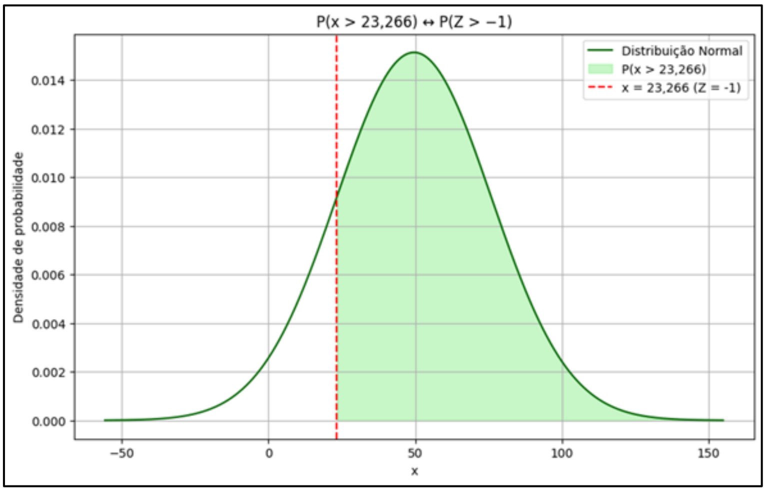
8- $P(x > Me - S) \rightarrow P(x > 23{,}266) \leftrightarrow \boldsymbol{P(Z > -1)}$

$$Z = \frac{X - \mu}{S} = \frac{23{,}266 - 49{,}625}{26{,}359} = -1$$

$Z = -1 = 0{,}3413$

$P(Z > -1) = 0{,}5 + 0{,}3413 = \mathbf{84{,}13}\%$

Gráfico 17 - Probabilidade para Z > -1



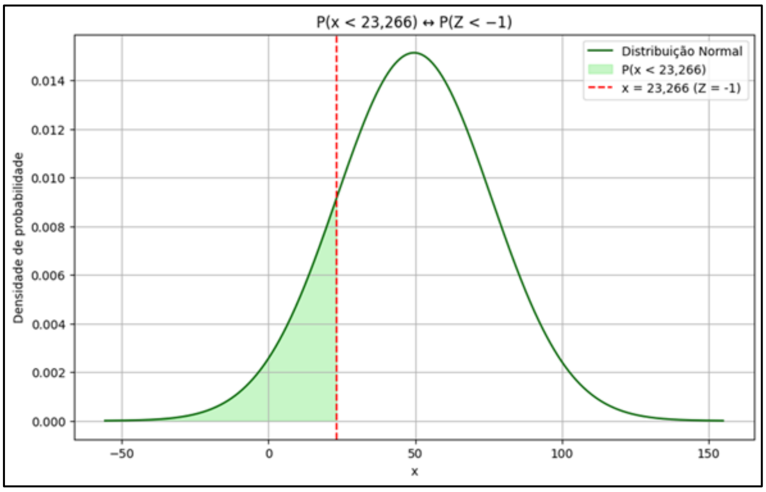
9- $P(x < Me - S) \rightarrow P(x < 23{,}266) \leftrightarrow \boldsymbol{P(Z < -1)}$

$$Z = \frac{X - \mu}{S} = \frac{23{,}266 - 49{,}625}{26{,}359} = -1$$

$Z = -1 = 0{,}3413$

$P(Z < -1) = 0{,}5 - 0{,}3413 = \mathbf{15{,}87}\%$

Gráfico 18 - Probabilidade para Z < -1



CURIOSIDADES:

Figura 21 - Tabela de Distribuição Normal Reduzida

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

AULA 11: Estimativa: Intervalo de Confiança da Média

Intervalo de Confiança

Intervalo de confiança é uma faixa de valores que tem uma certa probabilidade de conter o valor verdadeiro de um parâmetro populacional, como a média ou a proporção.

Explicando de forma simples:

Imagine que você quer saber a média de altura de todos os alunos de uma escola, mas só mede a altura de uma amostra (um grupo menor de alunos). Com base nessa amostra, você estima a média da escola inteira. Mas, como é só uma amostra, pode haver erro.

O intervalo de confiança é uma maneira de dizer:

"Com X% de confiança, acreditamos que a média verdadeira da população está entre A e B."

Exemplo:

Suponha que a média de altura da amostra seja 1,65 m e o intervalo de confiança de 95% seja de 1,62 m a 1,68 m. Isso significa:

"Estamos 95% confiantes de que a altura média de todos os alunos da escola está entre 1,62 m e 1,68 m."

Coisas importantes:

- O nível de confiança mais comum é 95%, mas também podem ser usados 90% ou 99%.
- Um intervalo mais estreito é mais preciso, mas normalmente exige uma amostra maior ou menor variabilidade nos dados.
- O intervalo de confiança não garante que o valor verdadeiro esteja dentro da faixa — ele só expressa a confiança com base na amostra e no modelo estatístico.

Estimativa do Intervalo de Confiança da Média

A estimativa do intervalo de confiança da média é uma ferramenta que permite estimar um intervalo dentro do qual podemos afirmar, com um certo nível de confiança, que a média populacional se encontra.

Passo a passo para calcular:

1. Coletar os dados: é necessária uma amostra aleatória da população.

n = tamanho da amostra

$\bar{x}$ = média da amostra

s = desvio padrão da amostra

2. Escolher o Nível de Confiança: determina a porcentagem de intervalos que conterão a média populacional se repetirmos a amostragem várias vezes.

3. Encontrar o Valor Crítico: dependendo do tamanho da amostra e do nível de confiança, utilizaremos uma distribuição normal (Z) ou t de student:

- Para amostras grandes ($n > 30$), geralmente usamos a distribuição normal.

- Para amostras pequenas ($n \leq 30$), usamos a distribuição t.

Os valores críticos podem ser encontrados em tabelas específicas ou calculadoras estatísticas.

4. Calcular o Erro Padrão da Média

O erro padrão é dado por:

$$EPM = \frac{s}{\sqrt{n}}$$

5. Calcular o Intervalo de Confiança

O intervalo de confiança é calculado usando a fórmula:

$$IC = \bar{x} \pm (valor\_crítico \times EPM)$$

IC = intervalo de confiança

O valor crítico é multiplicado pelo erro padrão para determinar a margem de erro.

Amplitude do Intervalo de Confiança da Média

A amplitude do intervalo de confiança de média é a diferença entre os limites superior e inferior do intervalo de confiança. Essa amplitude indica a precisão da estimativa da média populacional: quanto menor a amplitude, mais precisa é a estimativa

Cálculo:

$$amplitude = Limite\ Superior - limite\ inferior$$

Passo a passo:

1. Calcular o intervalo de confiança

2. Determinar os limites:

- Limite inferior: $\bar{x} - (valor\_crítico \times EPM)$

- Limite superior: $\bar{x} + (valor\_crítico \times EPM)$

3. Substituir os limites na fórmula da amplitude:

$Amplitude = [\bar{x} + (valor\_crítico \times EPM)] - [\bar{x} - (valor\_crítico \times EPM)]$

Simplificando: $Amplitude = 2 \times (valor\_crítico \times EPM)$

EXERCÍCIOS:

1- Obter ICM e fazer esboço dos gráficos para destacar área (%) sob a curva normal

Tabela 7 - Limite Confiança

LIMITE CONFIANÇA (%)	COEFICIENTE CONFIANÇA (Zc)
99,73	3,00
99	2,58
98	2,33
96	2,05
95,45	2,00
95	1,96
90	1,645
80	1,28
68,27	1,00
50	0,6745

N = n = $\sum fi = 440$ gols

$\mu = \text{Me} = 49{,}625$ minutos;

$\text{S} = 26{,}359$ minutos

A) 99,73%

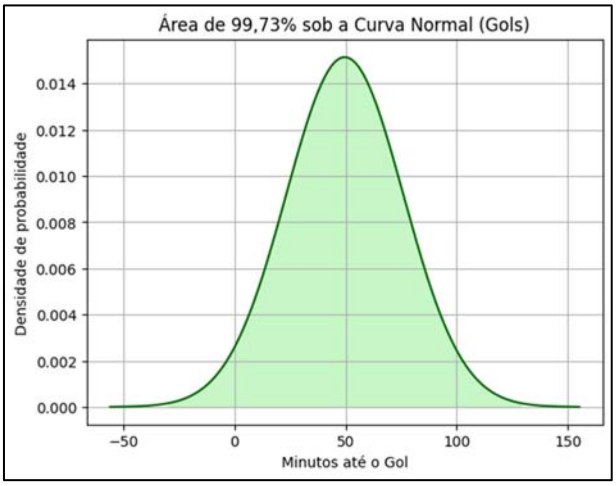
$99{,}73 = 3{,}00$

$49{,}625 \pm (3{,}00)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 3{,}7698$

$49{,}625 + 3{,}7698 = \mathbf{53,3948}\ \boldsymbol{minutos}$

$49{,}625 - 3{,}7698 = \mathbf{45,8552}\ \boldsymbol{minutos}$

Gráfico 19 - IMC 99,73%



B) 99%

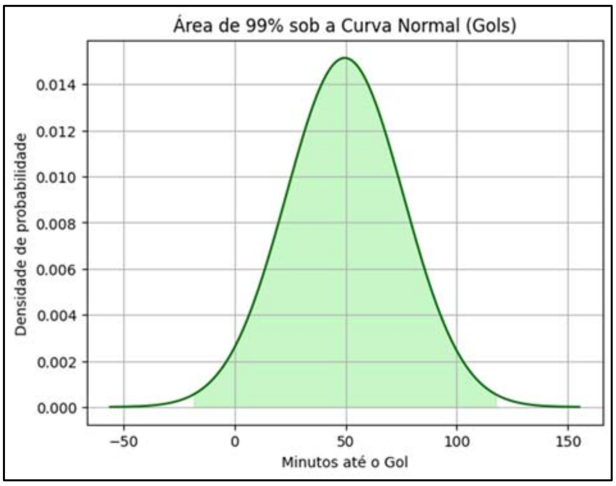
$99 = 2{,}58$

$49{,}625 \pm (2{,}58)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 3{,}2420$

$49{,}625 + 3{,}2420 =$**52,8670** $minutos$

$49{,}625 - 3{,}2420 =$**46,3830** $minutos$

Gráfico 20 - IMC 99%



C) 98%

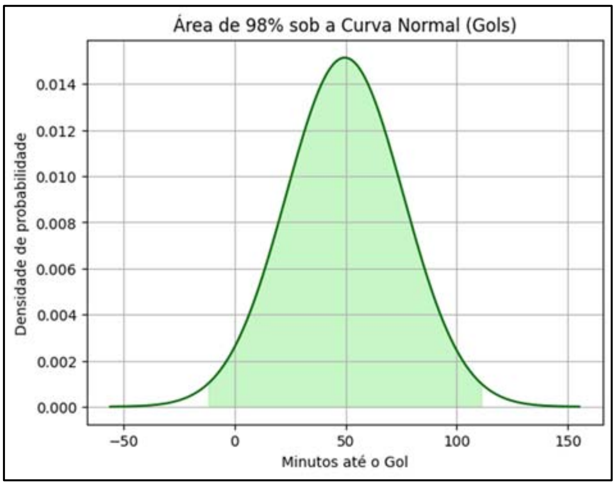
$$98 = 2{,}33$$

$$49{,}625 \pm (2{,}33)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 2{,}9278$$

$$49{,}625 + 2{,}9278 = \mathbf{52{,}5528}\ \boldsymbol{minutos}$$

$$49{,}625 - 2{,}9278 = \mathbf{46{,}6972}\ \boldsymbol{minutos}$$

Gráfico 21 - IMC 98%



D) 96%

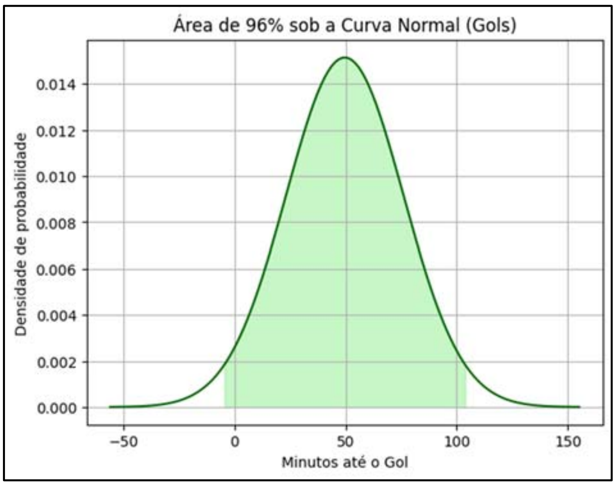
$$96 = 2{,}05$$

$$49{,}625 \pm (2{,}05)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 3{,}3066$$

$$49{,}625 + 3{,}3066 = \mathbf{52{,}9316}\ \boldsymbol{minutos}$$

$$49{,}625 - 3{,}3066 = \mathbf{46{,}3184}\ \boldsymbol{minutos}$$

Gráfico 22 - IMC 96%



E) 95,45%

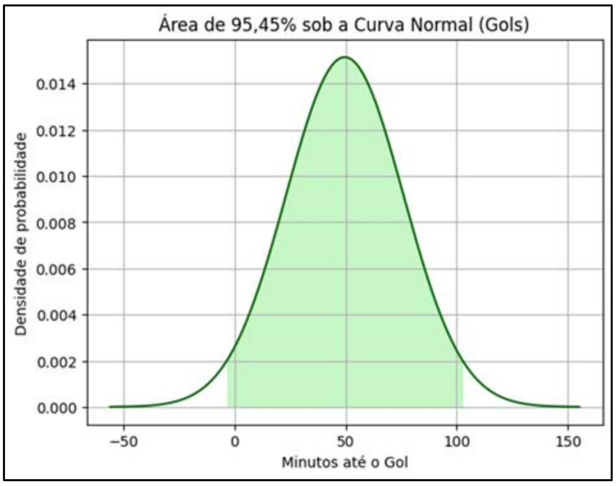
$95{,}45 = 2{,}00$

$49{,}625 \pm (2{,}00)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 2{,}5132$

$49{,}625 + 2{,}5132 = \mathbf{52{,}1382}\ \boldsymbol{minutos}$

$49{,}625 - 2{,}5132 = \mathbf{47{,}1118}\ \boldsymbol{minutos}$

Gráfico 23 - IMC 95,45%



F) 95%

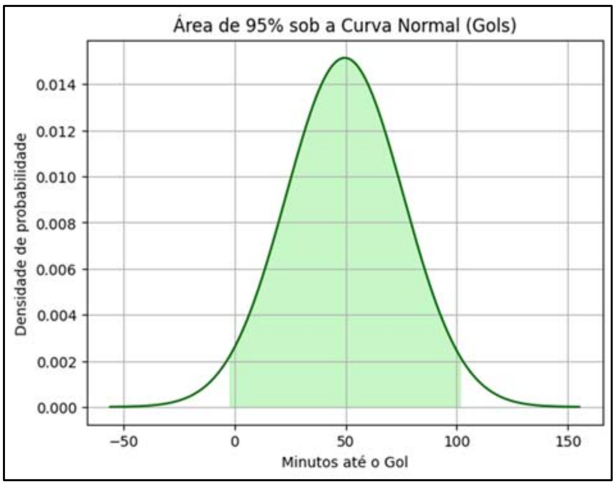
$95 = 1{,}96$

$49{,}625 \pm (1{,}96)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 2{,}4629$

$49{,}625 + 2{,}4629 = \mathbf{52{,}0879}\ \boldsymbol{minutos}$

$49{,}625 - 2{,}4629 = \mathbf{47{,}1621}\ \boldsymbol{minutos}$

Gráfico 24 - IMC 95%



G) 90%

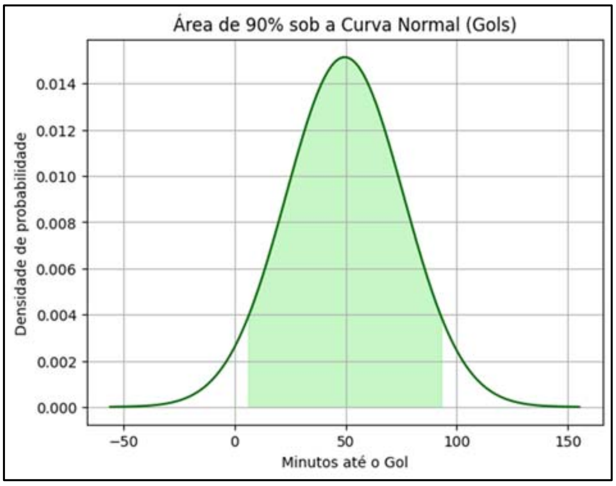
$90 = 1{,}645$

$49{,}625 \pm (1{,}645)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 2{,}0671$

$49{,}625 + 2{,}0671 = \mathbf{51{,}6921}\ \boldsymbol{minutos}$

$49{,}625 - 2{,}0671 = \mathbf{47{,}5579}\ \boldsymbol{minutos}$

Gráfico 25 - IMC 90%



H) 80%

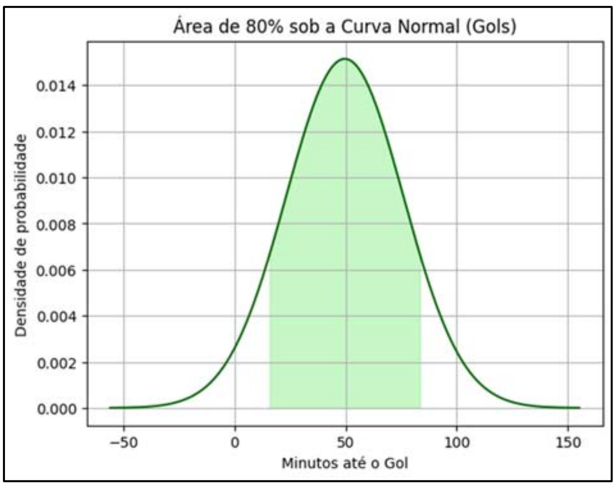
$80 = 1{,}28$

$49{,}625 \pm (1{,}28)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 1{,}6084$

$49{,}625 + 1{,}6084 = \mathbf{51{,}2334}\ \boldsymbol{minutos}$

$49{,}625 - 1{,}6084 = \mathbf{48{,}0166}\ \boldsymbol{minutos}$

Gráfico 26 - IMC 80%



I) 68,27%

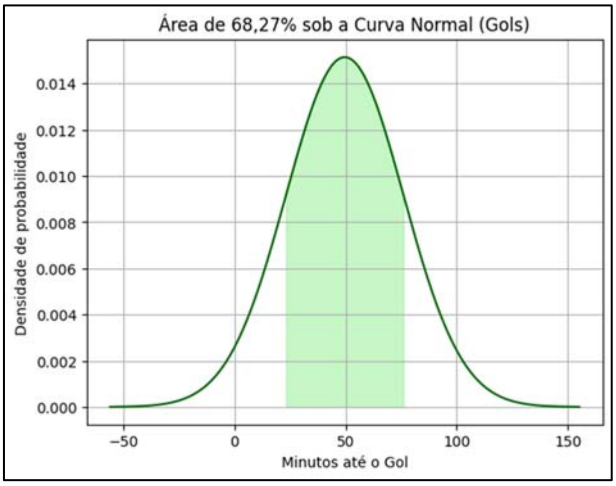
$68{,}27 = 1{,}00$

$$49{,}625 \pm (1{,}00)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 1{,}2566$$

$49{,}625 + 1{,}2566 = \mathbf{50{,}8816}\ \boldsymbol{minutos}$

$49{,}625 - 1{,}2566 = \mathbf{48{,}3684}\ \boldsymbol{minutos}$

Gráfico 27 - IMC 68,27%



J) 50%

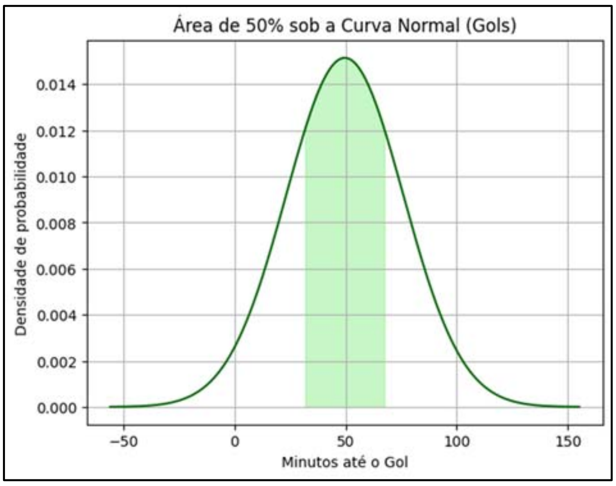
$50 = 0{,}6745$

$$49{,}625 \pm (0{,}6745)\left(\frac{26{,}359}{\sqrt{440}}\right) = 49{,}625 \pm 0{,}8475$$

$49{,}625 + 0{,}8475 = \mathbf{50{,}4725}\ \boldsymbol{minutos}$

$49{,}625 - 0{,}8475 = \mathbf{48{,}7775}\ \boldsymbol{minutos}$

Gráfico 28 - IMC 50%



CURIOSIDADES:

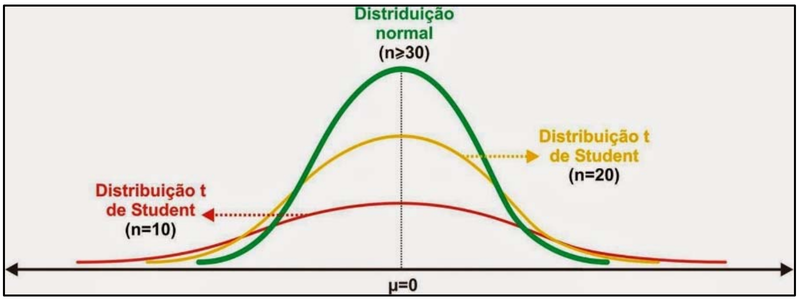
O intervalo de confiança é usado até em pesquisas eleitorais!

Quando vemos algo como “Candidato X tem 52% das intenções de voto com margem de erro de 2 pontos percentuais”, na verdade, isso é um intervalo de confiança! Significa que, com 95% de confiança, o verdadeiro valor está entre 50% e 54%. Mesmo com toda a tecnologia, a estatística é o que garante credibilidade nas previsões políticas.

AULA 12: Modelos de Distribuição: STUDENT e QUI-QUADRADO
Distribuição t de Student

A distribuição t de Student é uma distribuição de probabilidade que é utilizada principalmente em estatísticas inferenciais, especialmente para estimar a média de uma população quando a amostra é pequena (geralmente, $n < 30$) e a variância populacional é desconhecida. A forma da distribuição t se assemelha à distribuição normal, mas possui caudas mais longas, o que reflete a maior incerteza em amostras pequenas. À medida que o tamanho da amostra aumenta, a distribuição t se aproxima da distribuição normal.

Figura 22 - Exemplo de distribuições

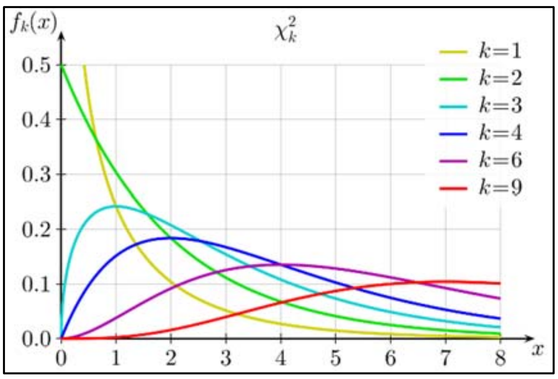


Fonte: Blogger (Estatística SPSS), 2025

Distribuição do Qui-quadrado

A distribuição do qui-quadrado é uma distribuição de probabilidade que se utiliza principalmente em testes de hipóteses e na construção de intervalos de confiança para variâncias. É frequentemente usada em análises de variância (ANOVA) e na avaliação da adequação de modelos estatísticos. A forma da distribuição do qui-quadrado depende dos graus de liberdade; quanto mais graus de liberdade, mais a curva se aproxima da normal.

Figura 23 - Distribuição qui-quadrado



Fonte: Quizlet, 2025

Pequenas Amostras

Em estatística, pequenas amostras referem-se geralmente a conjuntos de dados com um número reduzido de observações (n < 30). Quando se trabalha com pequenas amostras, as suposições sobre a população podem não ser tão confiáveis. Por isso, técnicas específicas, como o uso da distribuição t e métodos não paramétricos, são aplicadas para fazer inferências sobre a população.

Graus de Liberdade

Os graus de liberdade referem-se ao número independente de valores que podem variar ao calcular uma estatística. Em termos simples, eles representam a quantidade de informação disponível para estimar parâmetros em uma amostra. Por exemplo, ao calcular a variância em uma amostra, os graus de liberdade são geralmente dados por n - 1, onde n é o número total de observações na amostra. Isso ocorre porque estamos estimando a média da amostra ao calcular a variância.

EXERCÍCIOS:

1- Uma equipe de cientistas de dados está avaliando o tempo médio (em segundos) de resposta de um novo algoritmo de busca implementado em um sistema de banco de dados distribuído. Para isso, foram realizadas 12 execuções independentes do algoritmo sob as mesmas condições de carga e infraestrutura. A média amostral dos tempos de resposta obtida foi de 7,80 segundos, com um desvio padrão amostral de 1,50 segundos.

Considerando um nível de confiança de 95% e assumindo que os tempos de resposta seguem uma distribuição aproximadamente normal, utilize a distribuição t de Student para calcular o intervalo de confiança para o tempo médio de resposta do algoritmo.

Resposta:

$IC = 95\%;\ n = 12$

$Me = 7{,}80;\ S = 1{,}50$

$g.l. = 12 - 1 = 11$

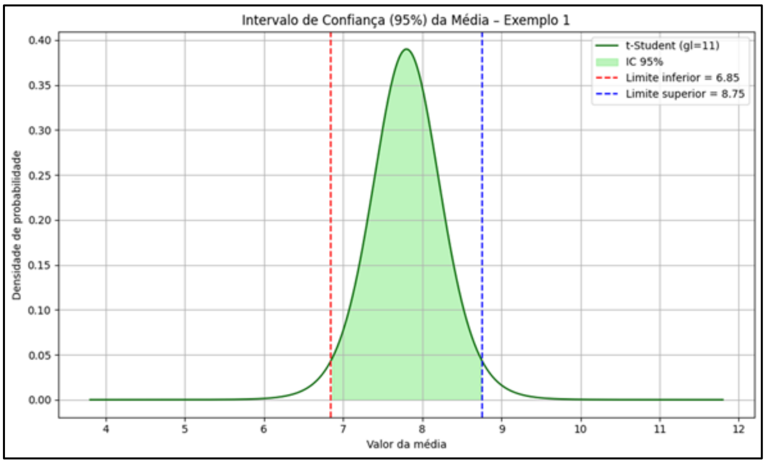
$$Me \pm tc\frac{S}{\sqrt{n}} = 7{,}80 \pm (2{,}201) \times \frac{1{,}5}{\sqrt{12}}$$

$7{,}80 + 0{,}9530 = 8{,}753$

$7{,}80 - 0{,}9530 = 6{,}847$

Conclusão: Com 95% de confiança, o tempo médio de resposta do algoritmo está entre 6,85 e 8,75 segundos.

Gráfico 29 - ICM 95%



2- Uma equipe de desenvolvedores está analisando o tempo médio de execução (em milissegundos) de uma nova função de ordenação implementada em um sistema embarcado de tempo real. Foram realizadas 15 execuções em condições controladas, resultando em:

- Média amostral: 5,50 ms
- Desvio padrão amostral: 0,60 ms

Considerando um nível de confiança de 99%, calcule o intervalo de confiança para o tempo médio de execução dessa função. Use a distribuição t de Student e assuma que o tempo de execução segue uma distribuição aproximadamente normal.

Reposta:

$IC = 99\%;\ n = 15$

$Me = 5{,}50;\ S = 0{,}60$

$g.l. = 15 - 1 = 14$

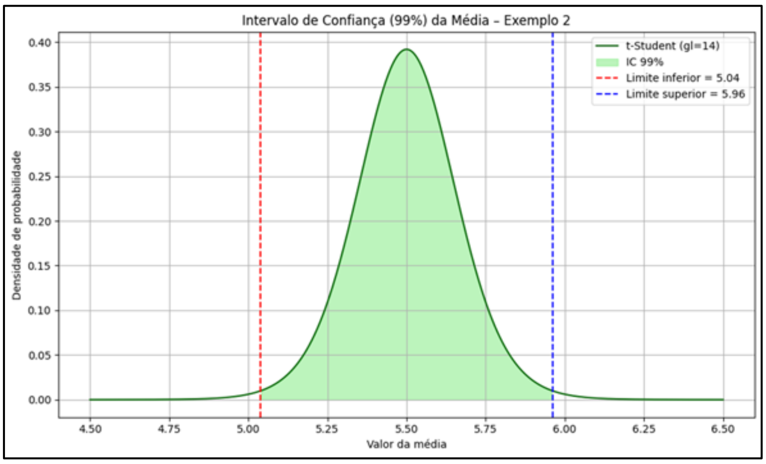
$$Me \pm tc\frac{S}{\sqrt{n}} = 5{,}50 \pm (2{,}977) \times \frac{0{,}60}{\sqrt{15}}$$

$5{,}50 + 0{,}4611 = 5{,}9611$

$5{,}50 - 0{,}4611 = 5{,}0389$

Conclusão: Com 99% de confiança, o tempo médio de execução da função está entre 5,04 ms e 5,96 ms.

Gráfico 30 - ICM 99%

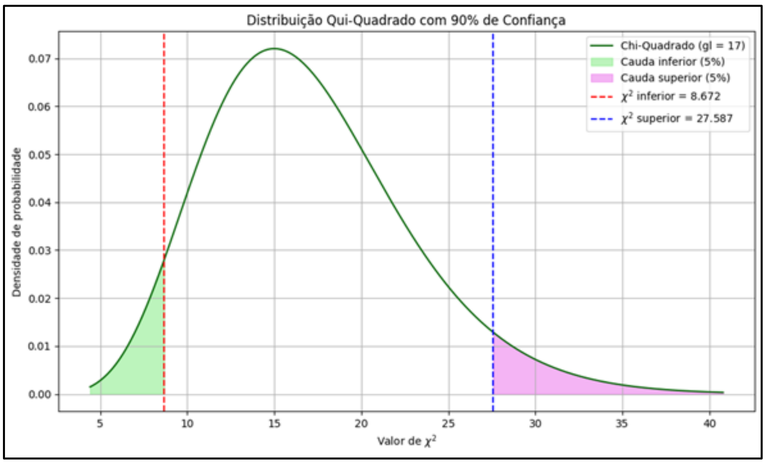


3- Dado um conjunto de dados categóricos com n = 18 observações e graus de liberdade (g.l.) = 17, deseja-se avaliar a independência entre duas variáveis por meio do teste de Qui-Quadrado (χ^2).

Considerando um nível de confiança (IC) de 90%, determine os valores críticos da distribuição Qui-Quadrado para as seguintes condições:

- χ^2 inferior (χ^2_i): Valor onde 5% da área está à esquerda ($\alpha/2 = 0{,}05$) → $\chi^2(17, 0{,}95)$ = 8,672
- χ^2 superior (χ^2_s): Valor onde 5% da área está à direita ($1 - \alpha/2 = 0{,}05$) → $\chi^2(17, 0{,}05)$ = 27,587

Gráfico 31 - Qui-Quadrado 90%



CURIOSIDADES:

Figura 24 - Tabela Distribuição Qui-Quadrado

Distribuição Qui-Quadrado

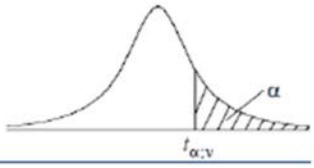
A tabela fornece os valores "c" tais que
$P(\chi^2 > c) = p$
onde "n" é o número de graus de liberdade e "p" é a probabilidade de sucesso.

gl	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672
35	17,192	18,509	20,569	22,465	24,797	34,336	46,059	49,802	53,203	57,342	60,275
40	20,707	22,164	24,433	26,509	29,051	39,335	51,805	55,758	59,342	63,691	66,766
45	24,311	25,901	28,366	30,612	33,350	44,335	57,505	61,656	65,410	69,957	73,166
50	27,991	29,707	32,357	34,764	37,689	49,335	63,167	67,505	71,420	76,154	79,490
55	31,735	33,570	36,398	38,958	42,060	54,335	68,796	73,311	77,380	82,292	85,749
60	35,534	37,485	40,482	43,188	46,459	59,335	74,397	79,082	83,298	88,379	91,952
70	43,275	45,442	48,758	51,739	55,329	69,334	85,527	90,531	95,023	100,425	104,215
80	51,172	53,540	57,153	60,391	64,278	79,334	96,578	101,879	106,629	112,329	116,321
90	59,196	61,754	65,647	69,126	73,291	89,334	107,565	113,145	118,136	124,116	128,299
100	67,328	70,065	74,222	77,929	82,358	99,334	118,498	124,342	129,561	135,807	140,169
110	75,550	78,458	82,867	86,792	91,471	109,334	129,385	135,480	140,917	147,414	151,948
120	83,852	86,923	91,573	95,705	100,624	119,334	140,233	146,567	152,211	158,950	163,648

Fonte: IME/UNICAMP, 2025

Figura 25 – Tabela Distribuição T Student

Tabela da T de Student



g.l.(ν)	α = 0.4	0.25	0.1	0.05	0.04	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	7.916	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	3.320	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	2.605	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.104	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.26	0.7	1.372	1.812	1.948	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.26	0.697	1.363	1.796	1.928	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.258	0.69	1.337	1.746	1.869	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.255	0.681	1.306	1.688	1.802	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.255	0.681	1.305	1.687	1.800	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.255	0.681	1.304	1.686	1.799	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.255	0.681	1.304	1.685	1.798	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	1.781	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.254	0.678	1.292	1.664	1.773	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.254	0.677	1.290	1.660	1.769	1.984	2.081	2.364	2.626	2.871	3.174	3.390
120	0.254	0.677	1.289	1.658	1.766	1.980	2.076	2.358	2.617	2.860	3.160	3.373
140	0.254	0.676	1.288	1.656	1.763	1.977	2.073	2.353	2.611	2.852	3.149	3.361
160	0.254	0.676	1.287	1.654	1.762	1.975	2.071	2.350	2.607	2.847	3.142	3.352
180	0.254	0.676	1.286	1.653	1.761	1.973	2.069	2.347	2.603	2.842	3.136	3.345
200	0.254	0.676	1.286	1.653	1.760	1.972	2.067	2.345	2.601	2.838	3.131	3.340
250	0.254	0.675	1.285	1.651	1.758	1.969	2.065	2.341	2.596	2.832	3.123	3.330
inf	0.253	0.674	1.282	1.645	1.751	1.960	2.054	2.326	2.576	2.807	3.090	3.290

Fonte: Portal UFRN, 2025

AULA 13: Testes e Hipóteses
Teste de Hipótese

Um teste de hipótese é um procedimento estatístico utilizado para tomar decisões sobre uma população com base em informações obtidas de uma amostra. O objetivo é avaliar se há evidências suficientes nos dados amostrais para rejeitar a hipótese nula em favor da hipótese alternativa.

Hipótese Nula (H_0)

A hipótese nula é uma afirmação que sugere que não há efeito ou diferença significativa entre grupos ou condições. É o ponto de partida do teste e, geralmente, representa a posição conservadora que busca testar. Por exemplo, H_0 pode afirmar que a média de duas populações é igual.

Hipótese Alternativa (H_1 ou H_a)

A hipótese alternativa é a afirmação que contraria a hipótese nula. Ela sugere que há um efeito ou diferença significativa entre os grupos ou condições. Dependendo do tipo de teste, a hipótese alternativa pode ser unilateral (indicando uma direção específica) ou bilateral (sem direção específica).

Nível de Significância do Teste (α)

O nível de significância é o critério usado para decidir se deve rejeitar a hipótese nula. Comumente, é definido como 0,05 (5%), significando que há uma probabilidade de 5% de rejeitar a hipótese nula quando ela é verdadeira. Esse valor representa o risco que o pesquisador está disposto a aceitar.

Região de Aceitação

A região de aceitação é o intervalo de valores da estatística do teste onde não rejeitamos a hipótese nula. Se o valor da estatística do teste cair dentro dessa região, aceitamos H_0 e concluímos que não há evidências suficientes para apoiar H_1.

Região de Rejeição

A região de rejeição é o intervalo onde rejeitamos a hipótese nula. Se o valor da estatística do teste cair nessa região, rejeitamos H_0 e aceitamos H_1. A localização dessa região depende do nível de significância e do tipo de teste realizado.

Região Crítica

A região crítica é um termo sinônimo à região de rejeição, referindo-se ao conjunto de valores para os quais a hipótese nula seria rejeitada com base nos dados amostrais coletados e no nível de significância estabelecido.

EXERCÍCIOS:

1- Uma pesquisa realizada por uma empresa especializada afirma que 45% dos profissionais de tecnologia nos Estados Unidos são a favor de substituir linguagens de programação tradicionais, como C e Java, por linguagens mais modernas como Python e Rust nos currículos universitários nos próximos 10 anos. Você decide testar essa informação e entrevista uma amostra de 200 desenvolvedores, dentre os quais 49% apoiam a substituição. Com 5% de significância, há evidência suficiente para apoiar a afirmação?

Resposta:

$p_0 = 0{,}45$

$n = 200$

$\hat{p} = 0{,}49$

$\alpha = 0{,}05$

$\boldsymbol{H_0}\ (\boldsymbol{nula})\colon p = 0{,}45\ (a\ proporção\ verdadeira\ \text{é}\ 45\%)$

$\boldsymbol{H_1}\ (\boldsymbol{alternativa})\colon p \neq 0{,}45\ (a\ proporção\ verdadeira\ \text{é}\ diferente\ de\ 45\%)$

Como estamos testando se a amostra difere da proporção da população **sem direção específica**, este é um **teste bicaudal** (bilateral).

Estatística do Teste (Z):

Fórmula teste de proporção:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Substituindo:

$$Z = \frac{0{,}49 - 0{,}45}{\sqrt{\frac{0{,}45(0{,}55)}{200}}} = \frac{0{,}04}{\sqrt{\frac{0{,}2475}{200}}} = \frac{0{,}04}{\sqrt{0{,}0012375}} \cong \mathbf{1{,}137}$$

Valor Crítico (Z crítico):

$Z_{\alpha/2} = Z_{0{,}025} \cong \pm 1{,}96$

Decisão

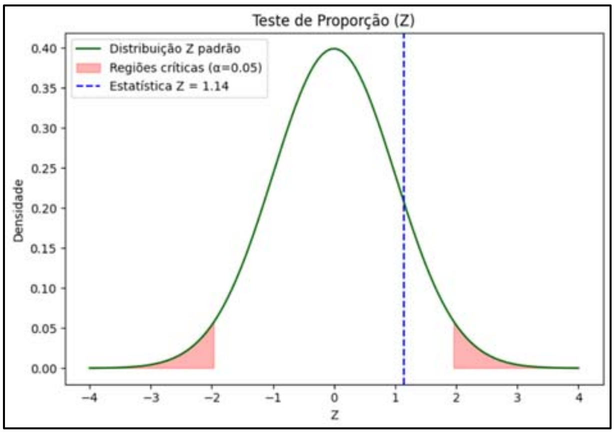
Regra: Rejeita-se H_0 se $|Z| > 1{,}96$

Como $|Z| = 1{,}137 < 1{,}96$, **não rejeitamos** H_0.

Conclusão:

Com 5% de significância, **não há evidência estatística suficiente para afirmar que a proporção de profissionais que apoiam a substituição das linguagens tradicionais difere de 45%**. Ou seja, os dados da amostra **não contradizem** a pesquisa original.

Gráfico 32 - Teste de Proporção



2- O custo X de manutenção mensal de servidores de um data center segue uma distribuição normal, $X \sim N(\mu, 400)$. Durante muito tempo, o valor médio μ tem sido considerado igual a 200. No entanto, há uma suspeita de que esse valor aumentou, e só nos interessa saber se o novo valor é superior a 210.

Desejamos planejar um teste com:

Nível de significância: $\alpha=5\%$ quando $\mu=200$

Erro tipo II desejado: $\beta=10\%$ quando $\mu=210$

a) Determinar o tamanho da amostra *n*

Fórmula para teste unilateral com variância conhecida:

$$n = \left(\frac{Z_\alpha + Z_\beta}{\frac{\mu_1 - \mu_0}{\sigma}} \right)^2$$

$\sigma^2 = 400 \rightarrow \sigma = \sqrt{400} = 20$

$\mu_0 = 200$

$\mu_1 = 210$

$Z_\alpha = Z_{0,05} \cong 1{,}645$ (à $direita$)

$Z_\beta = Z_{0,10} \cong 1{,}28$

Substituindo:

$$n = \left(\frac{1{,}645 + 1{,}28}{\frac{210 - 200}{20}} \right)^2 = \left(\frac{2{,}925}{0{,}5} \right)^2 = (5{,}85)^2 \cong 34{,}22$$

Conclusão: O tamanho mínimo da amostra é 35 servidores (arredondando para cima, pois o tamanho da amostra deve ser inteiro).

b) Determinar a região crítica (RC)

Como estamos testando:

H_0: μ = 200

H_1: μ > 210 (à direita)

A região crítica é definida para valores da média amostral $\bar{X}$ que levam à rejeição de H_0. A estatística do teste é:

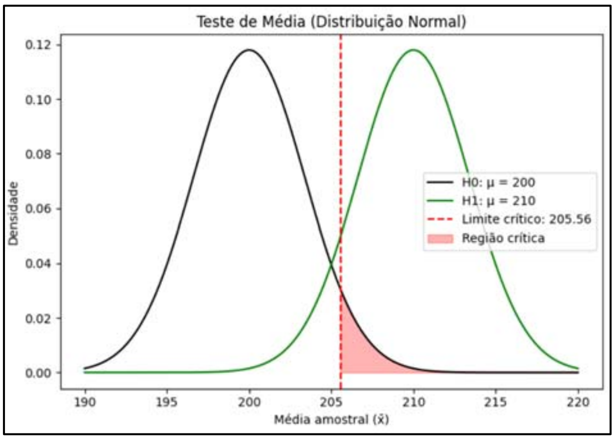
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 200}{20/\sqrt{35}} = \frac{\bar{X} - 200}{3{,}38}$$

O valor crítico de Z é Z_α = 1,645. Invertendo a fórmula para encontrar o valor crítico da média amostral:

$$\bar{X} > 200 + 1{,}645 \times \frac{20}{\sqrt{35}} = 200 + 1{,}645 \times 3{,}38 \cong \mathbf{205{,}56}$$

Conclusão: Região crítica: rejeitar H_0 se $\bar{X} > 205{,}56$.

Gráfico 33 - Teste de Média



CURIOSIDADES:

A aspirina só foi liberada graças aos testes de hipótese!

O medicamento foi testado com grupos de pessoas em estudos clínicos usando o famoso "teste de hipótese nula" para verificar se o efeito era realmente estatisticamente significativo e não apenas sorte. Hoje, praticamente toda pesquisa médica usa essa técnica para aprovar ou descartar tratamentos.

AULA 14: Correlação e Regressão Linear

Correlação Linear

A correlação linear é uma medida estatística que quantifica a força e a direção da relação linear entre duas variáveis. Ela indica se as variáveis tendem a aumentar ou diminuir juntas (correlação positiva), ou se uma aumenta enquanto a outra diminui (correlação negativa), e a intensidade dessa relação. O coeficiente de correlação de Pearson é uma das medidas mais comuns para avaliar a correlação linear, variando de -1 a +1, com 0 indicando ausência de correlação linear.

Em resumo:

- **Correlação Linear:** Uma relação entre duas variáveis que pode ser representada por uma linha reta.
- **Direção:** Pode ser positiva (as variáveis variam no mesmo sentido) ou negativa (as variáveis variam em sentidos opostos).
- **Força:** Medida pela proximidade do coeficiente de correlação com +1 ou -1 (forte) ou 0 (fraca).
- **Coeficiente de Correlação de Pearson:** A medida mais comum para avaliar a correlação linear, com valores entre -1 e +1.

Detalhamento:

- **Correlação Positiva:** Quando uma variável aumenta, a outra também tende a aumentar. Por exemplo, a altura e o peso de um indivíduo podem apresentar uma correlação positiva.
- **Correlação Negativa:** Quando uma variável aumenta, a outra tende a diminuir. Por exemplo, a idade e o número de cabelo podem apresentar uma correlação negativa.
- **Correlação Zero:** Não há relação linear entre as variáveis. Os pontos em um gráfico de dispersão não tendem a se alinhar em uma linha reta.
- **Coeficiente de Correlação de Pearson:** Este coeficiente quantifica a força e a direção da relação linear entre duas variáveis.

Valores:

+1: Correlação linear perfeita e positiva.

-1: Correlação linear perfeita e negativa.

0: Ausência de correlação linear.

Interpretação:

Quanto mais próximo de +1 ou -1, mais forte a correlação linear.

Quanto mais próximo de 0, mais fraca a correlação linear.

Notação Fórmula:

$$R_{x,y} = \frac{n \times \sum xi\, yi - \sum xi \times \sum yi}{\sqrt{n \times \sum xi^2 - \sum (xi)^2} \times \sqrt{n \times \sum yi^2 - \sum (yi)^2}}$$

Regressão Linear

Regressão Linear é uma técnica estatística usada para modelar a relação entre uma variável dependente (também chamada de variável resposta) e uma ou mais variáveis independentes (ou preditoras). O objetivo é encontrar uma equação que melhore a previsão ou explicação do comportamento da variável dependente com base nas variáveis independentes.

Tipos de Regressão Linear:

- **Regressão Linear Simples**

  Usa uma variável independente para prever uma variável dependente.

  Fórmula: $y = a + bx$

  Onde:

  y: variável dependente

  x: variável independente

  a: intercepto (valor de y quando x=0)

  b: coeficiente angular (inclinação da reta)

- **Regressão Linear Múltipla**

  Usa duas ou mais variáveis independentes.

  Fórmula: $y = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$

Cada b_i representa o impacto da variável x_i sobre y.

Objetivo

Encontrar os coeficientes que minimizam o erro entre os valores previstos e os valores reais, geralmente usando o método dos mínimos quadrados.

Métrica Comum

- **Erro Quadrático Médio (MSE):**

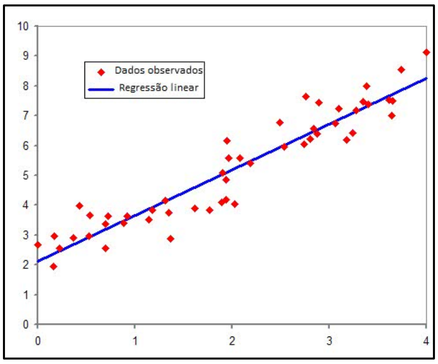
  Mede o erro médio ao quadrado entre os valores previstos e os reais. Quanto menor, melhor o modelo.

- **R^2 (Coeficiente de Determinação):**

  Mede quanta variabilidade da variável dependente é explicada pelo modelo (varia de 0 a 1).

Exemplo:

Figura 26 - Exemplo de Gráfico de Dispersão



Fonte: Medium, 2025.

Diagrama de Dispersão

Um Diagrama de Dispersão (ou Mapa de Dispersão, em inglês Scatter Plot) é uma representação gráfica usada para visualizar a relação entre duas variáveis numéricas.

Estrutura do Diagrama de Dispersão:

- Eixo X: representa a variável independente (ex: horas de estudo).
- Eixo Y: representa a variável dependente (ex: nota do aluno).
- Pontos: cada ponto representa uma observação (um par de valores x e y).

Finalidade:

- Visualizar tendências ou padrões (como correlações).
- Identificar possíveis relações lineares ou não lineares.
- Detectar outliers (valores fora do padrão).
- Auxiliar na análise de regressão linear.

Correlação Visual

Tipo de Padrão	**Interpretação**
Pontos sobem →	Correlação positiva
Pontos descem →	Correlação negativa
Pontos espalhados →	Sem correlação clara

Ligação com a Regressão Linear

Ao traçar uma linha de tendência (reta de regressão) sobre o diagrama de dispersão, você pode ver como a regressão se ajusta aos dados:

$$y = a + bx$$

Essa linha minimiza a distância quadrada entre ela e todos os pontos do gráfico.

EXERCÍCIOS:

1- Durante todo o mês de julho de 2018, a Sociedade Empresária Alfa realizou pesquisa diária visando medir a força relativa da relação linear entre o número de acessos ao seu site na Internet e o volume de vendas (em R$) de seu Produto "A". Sabe-se que os dados amostrais obtidos para os quatro primeiros dias de pesquisa foram:

Tabela 8 - Tabela para Ex1 Correlação Linear

Dia	Número de acessos ao site	Volume de Vendas do Produto "A" (em R$)
1	7	14,00
2	10	20,00
3	9	18,00
4	10	20,00

Com base nesses dados, calcule o coeficiente de correlação linear de Pearson (r) entre o número de acessos ao site e o volume de vendas do Produto "A" nos quatro primeiros dias da pesquisa.

Resposta:

Tabela auxiliar:

Tabela 9 - Tabela Auxiliar Ex1

x	y	xy	x^2	y^2
7	14	98	49	196
10	20	200	100	400
9	18	162	81	324
10	20	200	100	400
36	**72**	**660**	**330**	**1320**

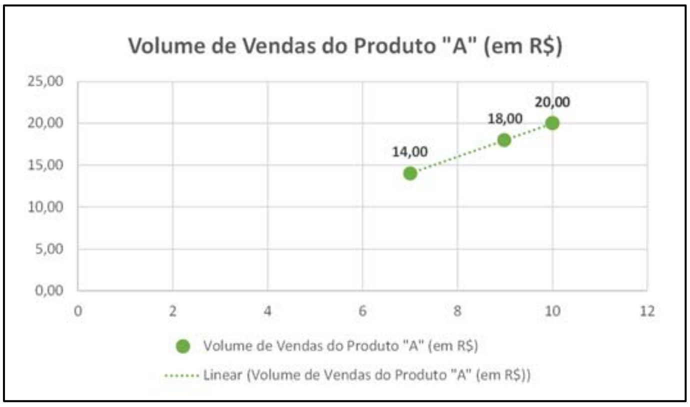
n = 4

$$R_{x,y} = \frac{4 \times (660) - (36) \times (72)}{\sqrt{4 \times (330) - 36^2} \times \sqrt{4 \times (1320) - 72^2}}$$

$$R_{x,y} = \frac{2640 - 2592}{\sqrt{1320 - 1296} \times \sqrt{5280 - 5184}} = \frac{48}{\sqrt{24 \times 96}} = \frac{48}{48} = \mathbf{1}$$

Conclusão: O coeficiente de correlação é 1, ou seja, uma correlação linear perfeita.

Gráfico 34 - Gráfico de Dispersão Ex1



2- A tabela abaixo mostra o tempo médio de resposta de um servidor em diferentes horários do dia:

Tabela 10 - Tabela para Ex2 Correlação Linear

Horário (X)	2	4	6	8	10	12	14	16
Tempo (ms) (Y)	112	104	100	92	88	86	84	80

Verifique se há uma correlação linear entre o horário de acesso e o tempo médio de resposta do servidor.

Resposta:

Tabela auxiliar:

Tabela 11 - Tabela Auxiliar Ex2

x	y	xy	x²	y²
2	112	224	4	12544
4	104	416	16	10816
6	100	600	36	10000
8	92	736	64	8464
10	88	880	100	7744
12	86	1032	144	7396
14	84	1176	196	7056
16	80	1280	256	6400
72	**746**	**6344**	**816**	**70420**

n = 8

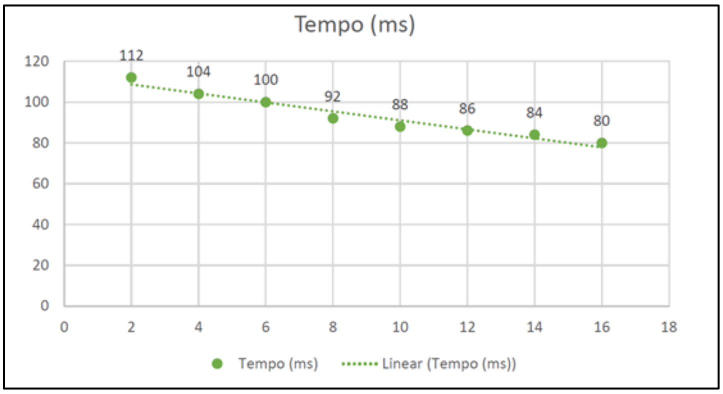
$$R_{x,y} = \frac{8 \times (6344) - (72) \times (746)}{\sqrt{8 \times (816) - 72^2} \times \sqrt{8 \times (70420) - 746^2}}$$

$$R_{x,y} = \frac{50752 - 53712}{\sqrt{6528 - 5184} \times \sqrt{563360 - 556516}} = \frac{-2960}{\sqrt{1344} \times \sqrt{6844}} = \frac{-2960}{36{,}66 \times 82{,}72} =$$

$$= \frac{-2960}{3032{,}51} = -0{,}976 \cong \mathbf{-1}$$

Conclusão: Isso indica uma **forte correlação negativa**: conforme o horário avança, o tempo médio de resposta do servidor tende a diminuir.

Gráfico 35 – Gráfico de Dispersão Ex2



CURIOSIDADES:

Correlação não implica causalidade – e isso já causou muita confusão!

Um site chamado *Spurious Correlations* mostra exemplos engraçados como: o consumo de queijo *per capita* nos EUA correlacionado com mortes por emaranhamento em lençóis. A relação existe nos números, mas obviamente não há uma ligação causal real. Isso mostra a importância de interpretar a estatística com cuidado.

Link do site mencionado: https://www.tylervigen.com/spurious-correlations

AULA 15: Revisão de Conteúdo
1- Distribuição Binomial (Aula 9)

Em um experimento de probabilidade, uma moeda honesta é lançada 6 vezes. Cada lançamento é independente dos demais. Seja X a variável aleatória que representa o número de vezes que a moeda caiu com o lado "cara" voltado para cima.

Com base nesse experimento, calcule:

a) A probabilidade de não sair nenhuma "cara":

$P(X = 0)$

$$\binom{6}{0} = \frac{6!}{(6-0)! \times 0!} = \frac{720}{720 \times 1} = 1$$

$$\binom{6}{0} \times 0{,}5^0 \times 0{,}5^{6-0} = 1 \times 1 \times 0{,}015625 = 0{,}015625 = \mathbf{1{,}5625}\%$$

b) A probabilidade de sair exatamente 1 "cara":

$P(X = 1)$

$$\binom{6}{1} = \frac{6!}{(6-1)! \times 1!} = \frac{720}{120 \times 1} = 6$$

$$\binom{6}{1} \times 0{,}5^1 \times 0{,}5^{6-1} = 6 \times 0{,}5 \times 0{,}03125 = 0{,}09375 = \mathbf{9{,}375}\%$$

c) A probabilidade de sair exatamente 2 "caras":

$P(X = 2)$

$$\binom{6}{2} = \frac{6!}{(6-2)! \times 2!} = \frac{720}{24 \times 2} = 15$$

$$\binom{6}{2} \times 0{,}5^2 \times 0{,}5^{6-2} = 15 \times 0{,}25 \times 0{,}0625 = 0{,}234375 = \mathbf{23{,}4375}\%$$

d) A probabilidade de sair exatamente 3 "caras":

$P(X = 3)$

$$\binom{6}{3} = \frac{6!}{(6-3)! \times 3!} = \frac{720}{6 \times 6} = 20$$

$$\binom{6}{3} \times 0{,}5^3 \times 0{,}5^{6-3} = 20 \times 0{,}125 \times 0{,}125 = 0{,}3125 = \mathbf{31{,}25}\%$$

e) A probabilidade de sair no máximo 1 “cara”:

$P(X \leq 1) = P(X = 0) + P(X = 1)$

$0{,}015625 + 0{,}09375 = 0{,}109375 = \mathbf{10{,}9375}\%$

f) A probabilidade de sair pelo menos 1 “cara”:

$P(X \geq 1) = 1 - P(X = 0)$

$1 - 0{,}015625 = 0{,}984375 = \mathbf{98{,}4375}\%$

g) A probabilidade de sair no máximo 2 “caras”:

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$0{,}015625 + 0{,}09375 + 0{,}234375 = 0{,}34375 = \mathbf{34{,}375}\%$

h) A probabilidade de sair pelo menos 2 “caras”:

$P(X \geq 2)1 - [P(X = 0) + P(X = 1)]$

$1 - (0{,}015625 + 0{,}09375) = 0{,}890625 = \mathbf{89{,}0625}\%$

i) A probabilidade de sair no máximo 3 “caras”:

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$0{,}34375 + 0{,}3125 = 0{,}65625 = \mathbf{65{,}625}\%$

j) A probabilidade de sair pelo menos 3 “caras”:

$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$1 - (0{,}34375) = 0{,}65625 = \mathbf{65{,}625}\%$

2- Distribuição Poisson (Aula 9)

Em uma empresa de tecnologia, o setor de controle de qualidade identificou que, em média, ocorrem 1,5 erros de digitação por página de relatório técnico revisado.

Assumindo que o número de erros por página segue uma Distribuição de Poisson, com média λ=1,5, responda, para $e^{-1,5}$ = 0,223:

Qual a probabilidade de uma página...

a) não conter nenhum erro de digitação?

$P(X = 0)$

$$\frac{1{,}5^0 \times 0{,}223}{0!} = \frac{1 \times 0{,}223}{1} = 0{,}223 = \mathbf{22{,}30}\%$$

b) conter exatamente 1 erro de digitação?

$P(X = 1)$

$$\frac{1{,}5^1 \times 0{,}223}{1!} = \frac{1{,}5 \times 0{,}223}{1} = 0{,}3345 = \mathbf{33,45}\%$$

c) conter exatamente 2 erros de digitação?

$P(X = 2)$

$$\frac{1{,}5^2 \times 0{,}223}{2!} = \frac{2{,}25 \times 0{,}223}{2} = \frac{0{,}50175}{2} = 0{,}250875 = \mathbf{25,09}\%$$

d) conter exatamente 3 erros de digitação?

$P(X = 3)$

$$\frac{1{,}5^3 \times 0{,}223}{3!} = \frac{3{,}375 \times 0{,}223}{6} = \frac{0{,}752625}{6} = 0{,}1254375 = \mathbf{12,54}\%$$

e) conter no máximo 1 erro de digitação?

$P(X \leq 1) = P(X = 0) + P(X = 1)$

$0{,}223 + 0{,}3345 = 0{,}5575 = \mathbf{55,75}\%$

f) conter pelo menos 1 erro de digitação?

$P(X \geq 1) = 1 - P(X = 0)$

$1 - 0{,}223 = 0{,}777 = \mathbf{77,70}\%$

g) conter no máximo 2 erros de digitação?

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$0{,}5575 + 0{,}2509 = 0{,}8083 = \mathbf{80,83}\%$

h) conter pelo menos 2 erros de digitação?]

$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$1 - 0{,}5575 = 0{,}4425 = \mathbf{44,25}\%$

i) conter no máximo 3 erros de digitação?

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$0{,}8083 + 0{,}1254 = 0{,}9337 = \mathbf{93,37}\%$

j) conter pelo menos 3 erros de digitação?

$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$1 - 0{,}9337 = 0{,}0663 = \mathbf{6{,}63}\%$

3- Distribuição Normal Padrão (Aula 10)

Um estudo foi realizado com 100 estudantes universitários para analisar o peso corporal médio dos participantes. A média amostral encontrada foi de 65,5Kg, com um desvio padrão de 5,5Kg. Considerando que os pesos dos estudantes seguem uma distribuição normal e que a amostra é suficientemente grande, responda:

Qual a probabilidade de um estudante ter peso:

a) igual à média amostral de 65,5Kg?

$P(X = Me) \rightarrow P(Z = 0)$

$$Z = \frac{65{,}5 - 65{,}5}{5{,}5} = 0{,}00$$

b) igual ou maior a média?

$P(X \geq Me) \rightarrow P(Z \geq 0)$

$0{,}5 + 0{,}00 = \mathbf{50}\%$

c) igual ou menor que a média?

$P(X \leq Me) \rightarrow P(Z \leq 0)$

$0{,}5 + 0{,}00 = \mathbf{50}\%$

d) entre 65,5Kg e 70Kg?

$P(Me \leq X \leq 70) \rightarrow P(0 \leq Z \leq 0{,}81)$

$$Z = \frac{70 - 65{,}5}{5{,}5} = 0{,}81$$

$0{,}81 = 0{,}2910$

$0{,}00 + 0{,}2910 = \mathbf{29{,}10}\%$

e) entre 60Kg e 65,5Kg?

$P(60 \leq X \leq Me) \rightarrow P(-1 \leq Z \leq 0)$

$$Z = \frac{60 - 65{,}5}{5{,}5} = -1$$

$-1 = 0{,}3413$

$0{,}00 + 0{,}3413 = \mathbf{34{,}13}\%$

f) entre 60Kg e 70Kg?

$P(60 \leq X \leq 70) \rightarrow P(-1 \leq Z \leq 0{,}81)$

$-1 = 0{,}3413$

$0{,}81 = 0{,}2910$

$0{,}3413 + 0{,}2910 = 0{,}6323 = \mathbf{63{,}23}\%$

g) maior ou igual a 60Kg?

$P(X \geq 60) \rightarrow P(Z \geq -1)$

$0{,}5 + 0{,}3413 = 0{,}8413 = \mathbf{84{,}13}\%$

h) menor ou igual a 70Kg?

$P(X \leq 70) \rightarrow P(Z \leq 0{,}81)$

$0{,}5 + 0{,}2910 = 0{,}7910 = \mathbf{79{,}10}\%$

i) maior ou igual a 70Kg?

$P(X \geq 70) \rightarrow P(Z \geq 0{,}81)$

$0{,}5 - 0{,}2910 = 0{,}2090 = \mathbf{20{,}90}\%$

j) menor ou igual a 60Kg?

$P(X \leq 60) \rightarrow P(Z \leq -1)$

$0{,}5 - 0{,}3413 = 0{,}1587 = \mathbf{15{,}87}\%$

k) entre 55Kg e 65Kg?

$P(55 \leq X \leq 65) = P(-1{,}91 \leq Z \leq -0{,}09)$

$$Z = \frac{55 - 65{,}5}{5{,}5} \cong -1{,}91$$

$$Z = \frac{65 - 65{,}5}{5{,}5} \cong -0{,}09$$

$1{,}91 = 0{,}4719$

$0{,}09 = 0{,}0359$

$0{,}4719 - 0{,}0359 = 0{,}4360 = \mathbf{43{,}60}\%$

l) entre 75Kg e 85Kg?

$P(75 \leq X \leq 85) = P(1{,}73 \leq Z \leq 3{,}54)$

$Z = \frac{75 - 65{,}5}{5{,}5} \cong 1{,}73$

$Z = \frac{85 - 65{,}5}{5{,}5} \cong 3{,}54$

$1{,}73 = 0{,}4582$

$3{,}54 = 0{,}4998$

$0{,}4998 - 0{,}4582 = 0{,}0416 = \mathbf{4{,}16}\%$

4- Estimativa: ICM e EP (Aula 11)

De acordo com o enunciado do exercício anterior, calcule o intervalo de confiança:

a) 99,73%

$65{,}5 \pm (3{,}00) \times \frac{5{,}5}{\sqrt{100}} =$

$65{,}5 + 1{,}65 = \boldsymbol{67{,}15Kg}$

$65{,}5 - 1{,}65 = \boldsymbol{63{,}85Kg}$

b) 99%

$65{,}5 \pm (2{,}58) \times \frac{5{,}5}{\sqrt{100}} =$

$65{,}5 + 1{,}419 = \boldsymbol{66{,}919Kg}$

$65{,}5 - 1{,}419 = \boldsymbol{64{,}081Kg}$

c) 98%

$65{,}5 \pm (2{,}33) \times \frac{5{,}5}{\sqrt{100}} =$

$65{,}5 + 1{,}2815 = \boldsymbol{66{,}7815Kg}$

$65{,}5 - 1{,}2815 = \boldsymbol{64{,}2185Kg}$

d) 96%

$65{,}5 \pm (2{,}05) \times \frac{5{,}5}{\sqrt{100}} =$

$65{,}5 + 1{,}1275 = \boldsymbol{66{,}6275Kg}$

$65{,}5 - 1{,}1275 = \mathbf{64{,}3725\,Kg}$

e) 95,45%

$$65{,}5 \pm (2{,}00) \times \frac{5{,}5}{\sqrt{100}} =$$

$65{,}5 + 1{,}10 = \mathbf{66{,}60\,Kg}$

$65{,}5 - 1{,}10 = \mathbf{64{,}40\,Kg}$

f) 95%

$$65{,}5 \pm (1{,}96) \times \frac{5{,}5}{\sqrt{100}} =$$

$65{,}5 + 1{,}078 = \mathbf{66{,}578\,Kg}$

$65{,}5 - 1{,}078 = \mathbf{64{,}422\,Kg}$

g) 90%

$$65{,}5 \pm (1{,}645) \times \frac{5{,}5}{\sqrt{100}} =$$

$65{,}5 + 0{,}90475 = \mathbf{66{,}40475\,Kg}$

$65{,}5 - 0{,}90475 = \mathbf{64{,}59525\,Kg}$

h) 80%

$$65{,}5 \pm (1{,}28) \times \frac{5{,}5}{\sqrt{100}} =$$

$65{,}5 + 0{,}704 = \mathbf{66{,}204\,Kg}$

$65{,}5 - 0{,}704 = \mathbf{64{,}796\,Kg}$

i) 68,27%

$$65{,}5 \pm (1{,}00) \times \frac{5{,}5}{\sqrt{100}} =$$

$65{,}5 + 0{,}55 = \mathbf{66{,}05\,Kg}$

$65{,}5 - 0{,}55 = \mathbf{64{,}95\,Kg}$

j) 50%

$$65{,}5 \pm (0{,}6745) \times \frac{5{,}5}{\sqrt{100}} =$$

$$65{,}5 + 0{,}371 = \mathbf{65{,}871}\boldsymbol{Kg}$$

$$65{,}5 - 0{,}371 = \mathbf{65{,}129}\boldsymbol{Kg}$$

CONSIDERAÇÕES FINAIS

Este estudo permitiu uma análise detalhada e aplicada dos principais conceitos estatísticos, utilizando como fio condutor a carreira do jogador Neymar. Ao explorar desde os fundamentos da estatística descritiva até as inferências estatísticas com distribuições de probabilidade, intervalos de confiança e testes de hipótese, foi possível não apenas entender a teoria por trás de cada método, mas também visualizar sua aplicação prática em dados esportivos.

A análise dos gols marcados por Neymar ao longo de sua carreira mostrou como medidas de tendência central e dispersão podem oferecer insights sobre consistência e desempenho. Além disso, a aplicação de conceitos como correlação e regressão linear demonstrou como essas ferramentas podem ser úteis para entender relações entre diferentes variáveis no contexto esportivo.

Essas análises contribuem para uma compreensão mais profunda dos dados esportivos e podem ser aplicadas em diversas áreas, como avaliação de desempenho de atletas, desenvolvimento de estratégias de jogo e tomada de decisões baseadas em dados.

A estatística se mostra, portanto, uma ferramenta indispensável no esporte, permitindo análises mais precisas e embasadas, que podem impactar diretamente o sucesso das equipes e atletas.

ANEXOS

Site criado pelo grupo:

REFERÊNCIAS

ALEXANDRE PROFESSOR. **Gráficos.** [s.d.]. Disponível em: https://alexandreprofessor.blogspot.com/p/graficos.html. Acesso em: 18 fev. 2025.

ALURA. **Histograma: o que é, exemplos, gráficos e tipos.** [s.d.]. Disponível em: https://www.alura.com.br/artigos/o-que-e-um-histograma. Acesso em: 18 fev. 2025.

ANDERSON, D. B. **Distribuição normal padrão.** Statorials, [s.d.]. Disponível em: https://statorials.org/pt/distribuicao-padrao-normal/. Acesso em: 14 abr. 2025.

ARTES INSPER, R. **Coeficiente de assimetria.** [s.d.]. Disponível em: https://www.ime.usp.br/~mbranco/MedidasdeAssimetria_2014.pdf. Acesso em: 24 mar. 2025.

BRASIL ESCOLA. **Coeficiente de variação.** [s.d.]. Disponível em: https://brasilescola.uol.com.br/matematica/coeficiente-variacao.htm. Acesso em: 18 fev. 2025.

BRASIL ESCOLA. **Fatorial: o que é, como resolver, simplificação.** [s.d.]. Disponível em: https://brasilescola.uol.com.br/matematica/fatorial.htm. Acesso em: 18 fev. 2025.

BRASIL ESCOLA. **Método binomial.** [s.d.]. Disponível em: https://brasilescola.uol.com.br/matematica/metodo-binomial.htm. Acesso em: 7 abr. 2025.

BRASIL ESCOLA. **Moda, média e mediana: como calculá-las.** [s.d.]. Disponível em: https://brasilescola.uol.com.br/matematica/moda-media-mediana.htm. Acesso em: 18 fev. 2025.

COSTA, A. **Estatística - Assimetria e curtose.** [s.d.]. Disponível em: https://www.professoraluizio.com.br/Blog/1291/estatistica-assimetria-e-curtose/. Acesso em: 24 mar. 2025.

CUEMATH. **Euler's number.** [s.d.]. Disponível em: https://www.cuemath.com/numbers/eulers-number/. Acesso em: 18 fev. 2025.

DESCRITIVA, E. **Associação entre Variáveis.** s.d. Disponível em: www.ime.usp.br/~sandoval/mae116psi/Aula%203%20-%20Descritiva%20III.pdf. Acesso em: 13 maio 2025.

DNC, E. **Entendendo a diferença entre amostra e população na estatística inferencial.** Escola DNC, [s.d.]. Disponível em: https://www.escoladnc.com.br/blog/entendendo-a-diferenca-entre-amostra-e-populacao-na-estatistica-inferencial/. Acesso em: 18 fev. 2025.

Estatística e probabilidade para computação - CC. Recife: UFPE, [202-]. Disponível em: https://www.cin.ufpe.br/~cin0142/lista.html. Acesso em: 25 maio 2025.

ESTATÍSTICA FÁCIL. **O que é: coeficiente de assimetria.** [s.d.]. Disponível em: https://estatisticafacil.org/glossario/o-que-e-coeficiente-de-assimetria/. Acesso em: 24 mar. 2025.

ESTATÍSTICA FÁCIL. **O que é: standard normal distribution (distribuição normal padrão).** [s.d.]. Disponível em: https://estatisticafacil.org/glossario/o-que-e-standard-normal-distribution-distribuicao-normal-padrao/. Acesso em: 14 abr. 2025.

ESTATÍSTICA PARA CONCURSO. **Coeficiente percentílico de curtose: 3 interpretações.** [s.d.]. Disponível em: https://estatisticaparaconcurso.com/coeficiente-percentilico-de-curtose/. Acesso em: 24 mar. 2025.

ESTATÍSTICA SPSS. **Distribuição t Student**. 2014. Disponível em: https://projetos-estatisticaspss.blogspot.com/2014/05/distribuicao-t-student.html. Acesso em: 29 abr. 2025.

ESTRATÉGIA CONCURSOS. **Resumo sobre Correlação Linear e Regressão para ISS-BH - Estatística**. s.d. Disponível em: www.estrategiaconcursos.com.br/blog/resumo-sobre-correlacao-linear-e-regressao-para-iss-bh-estatistica/. Acesso em: 12 jul. 2024.

EXCEL EASY. **Gráfico de linhas no Excel | Como criar?** [s.d.]. Disponível em: https://exceleasy.com.br/grafico-de-linhas-no-excel/. Acesso em: 18 fev. 2025.

FM2S. **Amostragem: o que é? Quais são os tipos?** FM2S, [s.d.]. Disponível em: https://www.fm2s.com.br/blog/amostragem. Acesso em: 18 fev. 2025.

FM2S. **O que é e para que serve a Estatística descritiva básica?** [s.d.]. Disponível em: https://www.fm2s.com.br/blog/estatistica-descritiva-basica-e-centralidade. Acesso em: 18 fev. 2025.

GOUVEIA, R. **Significado de Intervalo de Confiança (O que é, Conceito e Definição).** Significados, s.d. Disponível em: www.significados.com.br/intervalo-de-confianca/. Acesso em: 21 abr. 2025.

GUEDES, T. et al. **Aprender fazendo estatística estatística descritiva 1 introdução.** [s.d.]. Disponível em: https://www.ime.usp.br/~rvicente/Guedes_etal_Estatistica_Descritiva.pdf. Acesso em: 18 fev. 2025.

HASHTAG TREINAMENTOS. **Gráfico de barras no Excel: visual e dinâmico.** [s.d.]. Disponível em: https://www.hashtagtreinamentos.com/grafico-de-barras-no-excel. Acesso em: 18 fev. 2025.

IBGE. **Censo 2010 | materiais | guia do Censo | apresentação.** [s.d.]. Disponível em: https://censo2010.ibge.gov.br/materiais/guia-do-censo/apresentacao.html. Acesso em: 18 fev. 2025.

IFMG. **Apostila introdução à estatística.** [s.d.]. Disponível em: https://www.ifmg.edu.br/conselheirolafaiete/noticias/anexos-noticias/apostila-introducao-a-estatistica-ifmg-cl.pdf. Acesso em: 18 fev. 2025.

INSTITUTO DE BIOCIÊNCIAS, LETRAS E CIÊNCIAS EXATAS - CÂMPUS DE SÃO JOSÉ DO RIO PRETO. **Estatística: definição e conceitos preliminares.** [s.d.]. Disponível em: https://www.ibilce.unesp.br/Home/Departamentos/CiencCompEstatistica/Adriana/estatistica%20defini%C3%A7%C3%A3o%20e%20conceitos%20preliminares.pdf. Acesso em: 18 fev. 2025.

KHAN ACADEMY. **Revisão sobre regressão linear (artigo)**. s.d. Disponível em: pt.khanacademy.org/math/statistics-probability/describing-relationships-quantitative-data/introduction-to-trend-lines/a/linear-regression-review. Acesso em: 12 jul. 2024.

KUMON. **Número elevado a zero: por que o resultado é sempre 1?** [s.d.]. Disponível em: https://www.kumon.com.br/blog/matematica/numero-elevado-a-zero/. Acesso em: 9 abr. 2025.

LEMOS, V. Entenda o que são graus de liberdade. [S.l.]: Blog Psicometria Online, [20--]. Disponível em: https://www.blog.psicometriaonline.com.br/o-que-sao-graus-de-liberdade/. Acesso em: 29 abr. 2025.

LIMA, M. **Estatística descritiva e estatística inferencial: entenda as diferenças.** Psicometria Online, [s.d.]. Disponível em: https://www.blog.psicometriaonline.com.br/estatistica-descritiva-e-estatistica-inferencial-o-que-sao-e-quais-as-diferencas/. Acesso em: 18 fev. 2025.

LOCAWEB. **Aprenda a calcular a frequência absoluta e a frequência relativa.** Locaweb, [s.d.]. Disponível em: https://www.locaweb.com.br/blog/temas/codigo-aberto/aprenda-a-calcular-a-frequencia-absoluta-e-a-frequencia-relativa/. Acesso em: 18 fev. 2025.

MACIEL, F. **Assimetria e curtose dos dados.** [s.d.]. Disponível em: https://blog.proffernandamaciel.com.br/assimetria-e-curtose-dos-dados/. Acesso em: 24 mar. 2025.

MINITAB, E. **Como compreender os testes de hipóteses: níveis de significância (alfa) e valores-p na estatística.** State College: Minitab Blog, 2021. Disponível em: https://blog.minitab.com/pt/como-compreender-os-testes-de-hipoteses-niveis-de-significancia-alfa-e-valores-p-na-estatistica. Acesso em: 10 maio 2025.

MOKI SISTEMAS. **Diagrama de Dispersão: o que é, como interpretar e mais**. s.d. Disponível em: site.moki.com.br/pt-br/blog/post/diagrama-de-dispersao. Acesso em: 13 maio 2025.

MORETTIN, L. G. **Estatística básica: probabilidade e inferência: volume único**. São Paulo: Pearson Prentice Hall, 2010.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica.** 10. ed. São Paulo: Saraiva Uni, 2023.

MUNDO EDUCAÇÃO. **Desvio-padrão: o que é e como calcular.** [s.d.]. Disponível em: https://mundoeducacao.uol.com.br/matematica/desvio-padrao.htm. Acesso em: 18 fev. 2025.

MUNDO EDUCAÇÃO. **Tipos de gráficos. Os principais tipos de gráficos.** [s.d.]. Disponível em: https://mundoeducacao.uol.com.br/geografia/tipos-graficos.htm. Acesso em: 18 fev. 2025.

NASCIMENTO, C. **Aprenda como fazer gráfico de área no Excel.** Ninja do Excel, [s.d.]. Disponível em: https://ninjadoexcel.com.br/aprenda-como-fazer-grafico-de-area-no-excel/. Acesso em: 18 fev. 2025.

NEWCASTLE UNIVERSITY. **Numeracy, Maths and Statistics – Academic Skills Kit**. Newcastle upon Tyne: Newcastle University, [20--]. Disponível em: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/hypothesis-testing/critical-region-and-confidence-interval.html. Acesso em: 10 maio 2025.

NEYMAR - perfil de jogador 2025. [S.l.]: Transfermarkt, 2025. Disponível em: https://www.transfermarkt.com.br/neymar/profil/spieler/68290. Acesso em: 25 maio 2025.

O que é um gráfico de região de aceitação? State College: Minitab Support, [202-]. Disponível em: https://support.minitab.com/pt-br/minitab/help-and-how-to/quality-and-process-improvement/acceptance-sampling/supporting-topics/what-is-an-acceptance-region-plot/. Acesso em: 10 maio 2025.

OLIVEIRA, J. A. Distribuição t de Student. [S.l.: s.n.], [20--]. Disponível em: https://www.fcav.unesp.br/Home/departamentos/cienciasexatas/JOAOADEMIRDEOLIVEIRA/12%20aula.doc. Acesso em: 29 abr. 2025.

OLYMPICS. **Olympics | Olympic Games, Medals, Results & Latest News.** Lausanne: International Olympic Committee, [2023]. Disponível em: https://www.olympics.com/en/. Acesso em: 25 maio 2025.

ORTEGA, C. **Estatística inferencial: o que é, importância e exemplos.** QuestionPro, [s.d.]. Disponível em: https://www.questionpro.com/blog/pt-br/estatistica-inferencial/. Acesso em: 18 fev. 2025.

ORTEGA, C. **Tabela de frequência: o que é, elementos e como criar.** QuestionPro, [s.d.]. Disponível em: https://www.questionpro.com/blog/pt-br/tabela-de-frequencia/. Acesso em: 10 fev. 2025.

PEQUENAS AMOSTRAS. [S.l.: s.n.], [20--]. Disponível em: https://alexandreprofessor.blogspot.com/p/blog-page_18.html. Acesso em: 29 abr. 2025.

PROFESSOR GURU. **Gráficos estatísticos.** [s.d.]. Disponível em: https://www.professorguru.com.br/estatistica/graficos-estatisticos.html. Acesso em: 18 fev. 2025.

REJECTION region - an overview. ScienceDirect Topics. Amsterdam: Elsevier, [202-]. Disponível em: https://www.sciencedirect.com/topics/mathematics/rejection-region. Acesso em: 10 maio 2025.

SALGADO, D. **Intervalo de Confiança: o que é e como calcular**. Opinion Box, s.d. Disponível em: blog.opinionbox.com/intervalo-de-confianca/. Acesso em: 21 abr. 2025.

SÓ MATEMÁTICA. **Distribuição normal.** [s.d.]. Disponível em: https://www.somatematica.com.br/estat/basica/normal.php. Acesso em: 14 abr. 2025.

TESTE DE HIPÓTESES. Florianópolis: UFSC, [202-]. Disponível em: https://www.inf.ufsc.br/~andre.zibetti/probabilidade/teste-de-hipoteses.html. Acesso em: 10 maio 2025.

TODA MATÉRIA. **Média, moda e mediana.** [s.d.]. Disponível em: https://www.todamateria.com.br/media-moda-e-mediana/. Acesso em: 18 fev. 2025.

TODA MATÉRIA. **Variância e desvio padrão: o que são, fórmulas, como calcular e exercícios.** [s.d.]. Disponível em: https://www.todamateria.com.br/variancia-e-desvio-padrao/. Acesso em: 18 fev. 2025.

UFSC. **Distribuição normal (gaussiana).** [s.d.]. Disponível em: https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html. Acesso em: 18 fev. 2025.

VIEIRA, S. **O que é assimetria (ou distorção) e como se mede?** [s.d.]. Disponível em: https://soniavieira.blogspot.com/2018/05/por-conta-de-umapergunta-sobre-questao_83.html. Acesso em: 24 mar. 2025.

ZANETTA, D. M. T. **2 conceitos básicos de inferência estatística.** [s.d.]. Disponível em: https://midia.atp.usp.br/plc/plc0503/impressos/plc0503_02.pdf. Acesso em: 18 fev. 2025.