coggle
made for free at coggle.it

- Estatística Básica
  - Descritiva I
    - Medidas Descritivas
      - Tendência Central
        - Média
          - Aritmética
            - A média aritmética é a medida sintetizadora mais adequada quando não existem muitos Outliers.
            - O seu valor é obtido através da soma de todos os valores e dividido pelo tamanho da amostra ou população.
            - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$
            - Também conhecida como Esperança Matemática
          - Ponderada
          - Harmônica
        - Mediana
          - É o valor que divide o conjunto de dados ao meio
          - Para calcular a mediana é necessário que o conjunto de dados estaja organizado em ordem crescente
          - Se n é impar, a mediana é dada pelo valor que ocupa a posição $\frac{n+1}{2}$
          - Se n é par, a mediana será a média dos valores que ocuparn as posições $\frac{n}{2}$ e $\frac{n+2}{2}$
          - A mediana deve ser usada sempre que possível como medida representativa de distribuições fortemente assimétricas.
        - Moda
          - Dado um conjunto de valores, a Moda (Mo) desses valores será aquele que se repetir o maior número de vezes
          - Uma amostra que não tem moda que é chamada de amodal, duas modal é bimodal e três ou mais é chamada multimodal
        - Forma de Distribuição
          - Simétrica
            -
            - Média = Mediana = Moda
          - Assimétrica à direita
            -
            - Média > Mediana > Moda
          - Assimétrica à esquerda
            -
            - Média < Mediana < Moda
      - Dispersão
        - Amplitude
          - É definida como sendo a diferença entre o maior e o menor valor do conjunto de dados
        - Variância
          - Utilizada para identificar quanto próximo ou afastado o valor está da média
          - $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$
          - A variância tem a desvantagem de apresentar unidade de medida igual ao quadrado da unidade de medida dos dados.
        - Desvio Padrão
          - Utilizado para voltar a Variância para a medida original
          - $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$
        - Coeficiente de Variação
          - Classificação
            - < 10%: homogeneidade muito alta
            - de 10% a 20%: homogeneidade alta
            - de 20% a 30%: homogeneidade média
            - 30%: homogeneidade baixa
          - $CV_{(\%)} = \frac{s}{\bar{X}} \cdot 100\%$
          - É utilizado, quando se compara amostras com médias ou medidas desiguais.
        - Escore Padronizado (Escore Z)
          - O objetivo de se calcular o escore z é expressar em unidades de desvio padrão quanto um determinado número está distante da média
          - $\mathbf{z} = \frac{X - \bar{x}}{s}$
      - Posição Relativa
        - Quarts
        - Decis
        - Percents
      - Outliers
        - São valores muito grandes ou muito pequenos em relação aos demais
        - Alteram enormemente a média e a variabilidade dos dados
        - Outlier Moderado
          - $OS = Q_3 + 1,5 \times A$
          - $OI = Q_1 - 1,5 \times A$
        - Outlier Extremo
          - $OS = Q_3 + 3 \times A$
          - $OI = Q_1 - 3 \times A$
          - Amplitude no cálculo de Outlier é a diferença de Q3 - Q1
    - Descritiva II
      - Tipo de Dados
        - Qualitativos
          - Nominal
          - Ordinal
        - Quantitativos
          - Discretos
          - Continuos
      - Objetivo Estatística
        - Coletar
        - Organizar
        - Descrever
        - Analisar
        - Interpretar
      - Estudos Estatísticos
        - Fases
          1. Identificação do Problema
          2. Coleta de Dados
          3. Tratamento dos Dados
          4. Análise e interpretação
          5. Apresentação dos resultados
          6. Obtenção de relatório técnico
        - Tipos de Estudos
          - Observacional
            - Amostras Independentes
            - Características da População não são manipuladas
            - Ex. Censo, Pesquisa Eleitoral
          - Experimental
            - Amostras Pareadas
            - O grupo de individuos é manipulado para avaliar efeito
            - Ex. Eficácia de Tratamentos, Reações de Processos Químicos
      - Conceitos Básicos
        - Amostragem e Planejamento de Experimentos
          - Processo de escolha da amostra
          - Métodos de Amostragem
            - Aleatória Simples
            - Sistemática
            - Estratificada
            - Conglomerado
          - Parte inicial de qualquer estudo estatístico
          - Amostra é um subconjunto de elementos de uma população
        - Áreas da Estatística
          - Estatística Descritiva
          - Probabilidade
          - Estatística Inferencial
        - População
          - É o conjunto de todos os resultados, respostas, medidas ou contagens a serem estudados
          - Censo
            - É o conjunto de dados relativos a todos os elementos de uma população.
        - Parâmetro
          - O tamanho da população é denotado como N
          - Descreve numericamente uma característica da população
          - A média é denotada como $\mu$ para população e
      - Estimador ou Estatística
        - O tamanho da amostra é denotada como n
        - Descreve numericamente uma característica da amostra
        - O resultado de um parâmetro estimador é a estimativa
        - A média é denotada como $\bar{X}$
        - A mediana é denotada como $\tilde{X}$
        - A variância é denotada como $s^2$
        - O desvio padrão é denotado como S ou s
        - Graus de Liberdade (GL) é o número de elementos da amostra(n), menos o número de parâmetros (medidas da população) já estimados.
    - Inferência II
  - Tabelas
    - Distribuição de Frequência
      - 
    - Cruzadas
      - 
    - O símbolo $\vdash$ é utilizado para representar o intervalo aberto/fechado.
    - Uma maneira de determinar o número de classes é através do cálculo $K = \sqrt{n}$
  - Gráficos
    - Boxplot
      - 
      - É uma caixa que apresenta cinco medidas para se fazer análises estatísticas: a mediana, o primeiro quartil, o terceiro quartil, o menor valor e o maior valor no conjunto de dados.
      - Serve para visualizar o centro, a dispersão e a assimetria de um conjunto de dados
    - Gráfico de Dispersão (Scatter Plot)
      -
      - Os dados são exibidos como uma coleção de pontos, cada um com o valor de uma variável determinando a posição no eixo y e o valor da outra variável no eixo x.
      - O gráfico de dispersão é útil para verificar como dois conjuntos de dados compartilhem concordam entre si.
    - Linhas
      -
      - Um gráfico de linhas é muitas vezes usado para visualizar uma tendência nos dados em intervalos de tempo
    - Colunas ou Barras Vertical
      -
      - Os gráficos de colunas são úteis para mostrar alterações de dados durante um período ou para ilustrar a comparação entre itens.
      - Os gráficos de colunas podem ser agrupados, empilhados ou 100% empilhados (para visualizar distribuições).
    - Barras Horizontal
      -
      - O propósito é o mesmo do de barras vertical
      - Utilizasse quando se tem muitas categorias.
    - Setores ou Pizza
      -
      - Um diagrama circular em que os valores de cada categoria estatística representada são proporcionais às respectivas medidas dos ângulos
      - Evitar este gráfico quando houver mais de 4 categorias
    - Ramo e Folhas
      - 1 | 6 7 7 8 8 9 9 9
        2 | 6 8 8 1 1 2 3 3 4 4 4 4 8 9 9 9
        3 | 0 1 2 2 2 2 7 7 8 8
        4 | 2 3 6 6 6 6 8 8 9
        5 | 2 3 3 4 4 8
        6 | 0 2 5 6
        7 | 1 4
      - Dispositivo para apresentação de dados quantitativos em um formato gráfico
      - São úteis para mostrar a densidade relativa e a forma dos dados, dando ao leitor uma rápida visão geral da distribuição.
    - Histograma
      -
      - Representação gráfica em colunas de um conjunto de dados previamente tabulado e dividido em classes de valores uniformes ou não uniformes.
      - Pode ser Simétrico (ou Gaussiano ou Unimodal), Distorcido à Direita, Distorcido à Esquerda, Bimodal, Multimodal ou Plato.
  - Probabilidade II
    - Conceitos
      - Definições Básicas
        - Experimento Aleatório
          - São experimentos cujos resultados podem ser diferentes, ainda que sejam repetidos sob condições idênticas
          - É possível descrever todos os possíveis resultados.
        - Espaço Amostral
          - O conjunto de todos os possíveis resultados de um experimento aleatório
          - Ex. [Ouro, Paus, Copas, Espadas]
          - É representado pela letra Omega Ω
        - Eventos
          - Subconjunto do espaço amostral
          - São enumeráveis no casos de dados discretos
          - São colocados na forma de intervalo no caso de dados continuos
          - É representado pela letra A
          - Eventos Equiprovaveis
            - São eventos on P(e1) = P(e2)=...=P(en)). logo que possuem a mesma probabilidade de ocorrer
            - A probabilidade de cada evento é 1/n.
        - Operações com Eventos Aleatórios
          - Complementação
            -
          - União
            - 
          - Intersecção
            - 
        - Axiomas (de Kolmogorov)
          - P(A) ≥ 0, para qualquer evento A em Ω
          - P(Ω) = 1
          - Se A e B são dois eventos de Ω e mutuamente exclusivos então: P(A ∪ B) = P(A) + P(B)
      - Clássico
        - Probabilidade que calcula analiticamente as chances de determinado fenômeno acontecer
        - O conceito clássico só pode ser utilizado em situações onde o espaço amostral é enumerável, finito e equiprovável.
        - P(A) = Número de casos favoráveis / Número de casos possíveis
      - Frequentista
        - Probabilidade de um evento como a proporção do número de vezes que o evento ocorre
        - A medida que o número de repetições do experimento cresce, a frequência relativa se aproxima da probabilidade P(A).
        - P(A) = Número de vez que A ocorreu / Número de Experimentos realizados
    - Teoremas
      - Probabilidade Total
        - Se os eventos A1, A2, A3...An formam uma partição do espaço amostral, então: $\bigcup_{i=1}^{n} A_i = \Omega$
      - Se ∅ é um conjunto vazio, então P(∅) = 0
      - Evento Complementar
        - Se Ac é o complemento de A, então P(Ac) = 1 - P(A)
      - Se A e B são dois eventos quaisquer e Bc é o complemento de B, então: P(A ∩ Bc ) = P(A) - P(A ∩ B)
      - Soma das Probabilidades
        - Se A e B são dois eventos quaisquer, então: P(A ∪ B) = P(A) + P(B) - P(A ∩ B)
      - Se A ⊂ (é subconjunto) B, então P(A) ≤ P(B)
      - Para um evento A qualquer: 0 ≤ P(A) ≤ 1
      - Produto das Probabilidades
        - P(A ∩ B) = P(A|B).P(B)
        - P(A ∩ B ∩ C) = P(A|(B ∩ C)).P(B|C).P(C)
      - Teoremla Bayes
        - As probabilidades condicionais P(A|B) e P(B|A) expressam diferentes valores de probabilidades
        - $P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$
        - onde: P(A|B).P(B) = P(A ∩ B) = P(B ∩ A) = P(B|A).P(A)
        - $P(B) = \sum_{i=1}^{n} P(A) \cdot P(B|A)$
    - Probabilidade Condicional
      - Definimos como a probabilidade de A ocorrer dado que B ocorre
      - P(A|B) = P(A ∩ B)/P(B), onde P(B) ≠ 0
    - Independência Estocástica ou Eventos Independentes
      - P(A|B) = P(A) ou P(B|A) = P(B)
      - Dois eventos A e B são ditos independentes se: P(A ∩ B ) = P(A).P(B)
      - Probabilidade de um evento ocorrer sem que dependa do fato de os outros eventos terem ou não ocorrido
  - Variáveis Aleatórias
    - Definições
      - Em matemática, é definida como uma função que associa a todo evento pertencente a uma partição do espaço amostral Ω a um único número real (R)
      - É comum a representação das variáveis aleatórias por letras maiúsculas e dos valores assumidos por letras minúsculas
      - O uso da variável aleatória permite descrever resultados de um experimento aleatório por meio de número ao invés de palavras
      - $X : \Omega \to \mathbb{R} \Leftrightarrow \{\omega : X(\omega) \leq a\} \in \mathcal{F}, \ \forall a \in \mathbb{R}$
    - Discretas
      - Assume qualquer valor dentro de um intervalo, ou seja, se houver um número incontável de resultados possíveis
      - Funções
        - Univariada
          - Unidimensional
          - Função de massa de probabilidade
            - $P(x_i) \geq 0$ para todo $x_i$
            - $\sum_{i=1} P(x_i) = 1$
          - Multivariada
            - Vetor Aleatório
            - Função de distribuição acumulada
              - F(x) = P(X ≤ x)
              - $F(x) = \sum_{x_i \leq x} p(x_i)$
      - Fórmulas Importantes
        - Variância
          - $Var(x) = \sum_{i=1}^{n} x_i^2 \cdot p(x_i) - E(X)^2$
          - Propriedades
            - A variância também possui a propriedade de linearidade do valor médio, isto é, Var(aX+b)=a²Var(a(X))+b. Em vez disso, teremos Var(aX+b)=a²Var(X)
            - Se c é uma constante, então Var(X+c) = Var(X)
            - Suponha-se que c seja uma constante e X seja uma variável aleatória, então Var(cX) = c²Var(X)
            - Se (X , Y) for uma variável aleatória bidimensional, e se X e Y forem independentes. Então, Var(X+Y) = Var(X) + Var(Y)
        - Média ou Esperança
          - $E(\Sigma) = \sum_{i=1}^{n} x_i \cdot p(x_i)$
          - Propriedades
            - Se X = c, onde c é uma constante, então E(X) = c
            - Suponha-se que c seja uma constante e X seja uma variável aleatória. Então E(cX) = cE(X)
            - Sejam X e Y duas variáveis aleatórias quaisquer. Então, E(X+Y) = E(X) + E(Y)
            - Para funções lineares, se Y = aX + b, onde a e b são constantes, então E(Y) = aE(X) + b
      - Modelos Probabilísticos
        - Bernoulli
          - Probabilidade de Sucesso ou Fracasso
          - Ex. Acertar chute ao gol
          - Função
            - P(sucesso) = p
          - Média
            - p
          - Variância
            - p(1-p)
        - Binomial
          - Nro de sucessos (p) dentre n repetições de um experimento
          - Acertar 2 de 5 chutes ao gol
          - Função
            - $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$
          - Média
            - n.p
          - Variância
            - n.p(1-p)
        - Binomial Negativa
          - Nro de tentativas (k) para acontecer o r-ésimo sucesso
          - Acertar 2 chutes na tentativa 4
          - Função
            - $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
          - Média
            - $\frac{r}{p}$
          - Variância
            - $\frac{r(1-p)}{p^2}$
        - Geométrica
          - Nro de tentativas (k) até acontecer o primeiro sucesso
          - Acertar o primeiro chuta na terceira tentativa
          - Função
            - $P(X = k) = p(1-p)^{k-1}$
          - Média
            - $\frac{1}{p}$
          - Variância
            - $\frac{1-p}{p^2}$
        - HiperGeométrica
          - Nro. de elementos (k) de uma amostra com n elementos, obtida sem reposição de um universo de N elementos com r elementos com determinada característica
          - Ter 3 celulares com defeito de uma amostra de 10 celulares retirados sem reposição de um lote de 50 celulares que tinham 10 com defeito
          - Função
            - $P(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$
          - Média
            - $np$
          - Variância
            - $np(1-p)\frac{N-n}{N-1}$
        - Poisson
          - Nro. de sucessos (k) dentro do um intervalo de tempo, volume, espaço
          - Um callcenter receber 3 chamas em 1 minuto
          - Função
            - $P(X = k) = \frac{e^{\lambda} \cdot \lambda^k}{k!}$
          - Média
            - $\lambda$
          - Variância
            - $\lambda$
    - Contínuas
      - Assume valores que podem ser contados
      - Funções
        - Univariada
          - Unidimensional
          - Função de massa de probabilidade
            - f(x) ≥ 0, para todo x
            - A área definida por f(x) é igual a 1, ou seja, $\int_{-\infty}^{\infty} f(x)dx = 1$
            - Para quaisquer a e b, com −∞ < a < b < +∞, teremos $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$
          - Multivariada
            - $F(x) = \int_{-\infty}^{x} f(t)\,dt$
            - Função densidade de probabilidade conjunta
            - Unidimensional
      - Fórmulas Importantes
        - Média ou Esperança
          - $E(\Sigma) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$
          - Propriedades =
        - Variância
          - $Var(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - [E(X)]^2$
          - Propriedades =
    - Modelos Probabilísticos
      - Uniforme e Contínua
      - Distribuição Exponencial
      - Distribuição Normal
      - Gama
      - Weibull
      - Beta
      - Lognormal
  - Inferencial II