Amostra

- Amostra é todo o subconjunto de unidades tirados da população
- Amostra Probabilística
  - Todos os elementos da população apresentam probabilidade maior que zero de serem selecionados (aleatória simples, estratificada, sistemática e por conglomerados)
- Amostra não Probabilística
  - Quando não há probabilidade clara/conhecida de seleção dos elementos. Os elementos são escolhidos de forma julgamental (acidental, intencional, por cotas)

Amostragem Probabilísticas (mesma chance para qualquer elemento da população)

- Amostra aleatória simples
  - Pode ser utilizada com reposição ou sem reposição

- Amostra Estratificada
  - População dividida em estratos homogêneos (grupos com elementos de características comuns) e é selecionada uma amostra aleatória de cada estrato.

- Amostra Sistemática
  - Os elementos são selecionados seguindo uma regra pré definida

- Amostra por conglomerados
  - A população (extensa) é dividida em miniaturas da população (não homogêneas) e seleciona-se uma amostra aleatória desses conglomerados.

Tipos de dados

- Dados Qualitativos:
  - Nominais: São categorias que não têm uma ordem específica (ex.: cores, tipos de frutas).
  - Ordinais: São categorias que têm uma ordem ou classificação (ex.: níveis de satisfação, classificações).

- Dados Quantitativos:
  - Discretos: São contáveis e geralmente são números inteiros. (ex.: número de filhos, quantidade de livros).
  - Contínuos: Podem assumir qualquer valor dentro de um intervalo e são mensuráveis (ex.: altura, peso, temperatura).

Gráficos

- **Linha**
  - É utilizado para demonstrar uma sequência numérica de um certo dado ao longo do tempo. É indicado para demonstrar evoluções (ou regressões) que ocorrem em sequência para que o comportamento dos fenômenos e suas transformações sejam observadas.
- **Coluna**
  - É composto por duas linhas ou eixos, um vertical e outro horizontal. Muitas vezes também é conhecido como gráfica de barras. Medir a frequência de um dado categórico.
- **Barra**
  - Tem a mesma finalidade que o gráfico de colunas mas pode ser apresentado através de barras horizontais e verticais.
- **Setores**
  - Também conhecido como pizza, é construído com base em um círculo e é empregado sempre que desejamos ressaltar a participação do dado no total.
- **Histograma**
  - Conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidem com os pontos médios dos intervalos de classe. A área de um histograma é proporcional à soma das frequências simples ou absolutas.
- **Boxplot**
  - É uma maneira gráfica de representar a alteração dos dados de uma variável por meio de quartis.
- **Dispersão**

- São utilizados na análise do relacionamento entre duas variáveis

Medidas

- **Posição**
  - Média (muito influenciada pelos outliers)
  - Moda
  - Mediana (tem que estar ordenado)
  - Quartis
    - Divide os dados em 4 partes iguais
    - 1 3 5 6 8 9 11 15 16 18 20
    - 5 → 1 quartil
    - 9 → 2 quartil
    - 16 → 3 quartil
- **Variabilidade**
  - Amplitude
    - valor final - valor inicial
    - sensível a outliers
  - Variância

$$s^2 = \frac{\sum (X-\overline{X})^2}{N-1}$$

  - Desvio padrão
    - Raiz quadrada da variância
    - Amostral

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

- Populacional

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

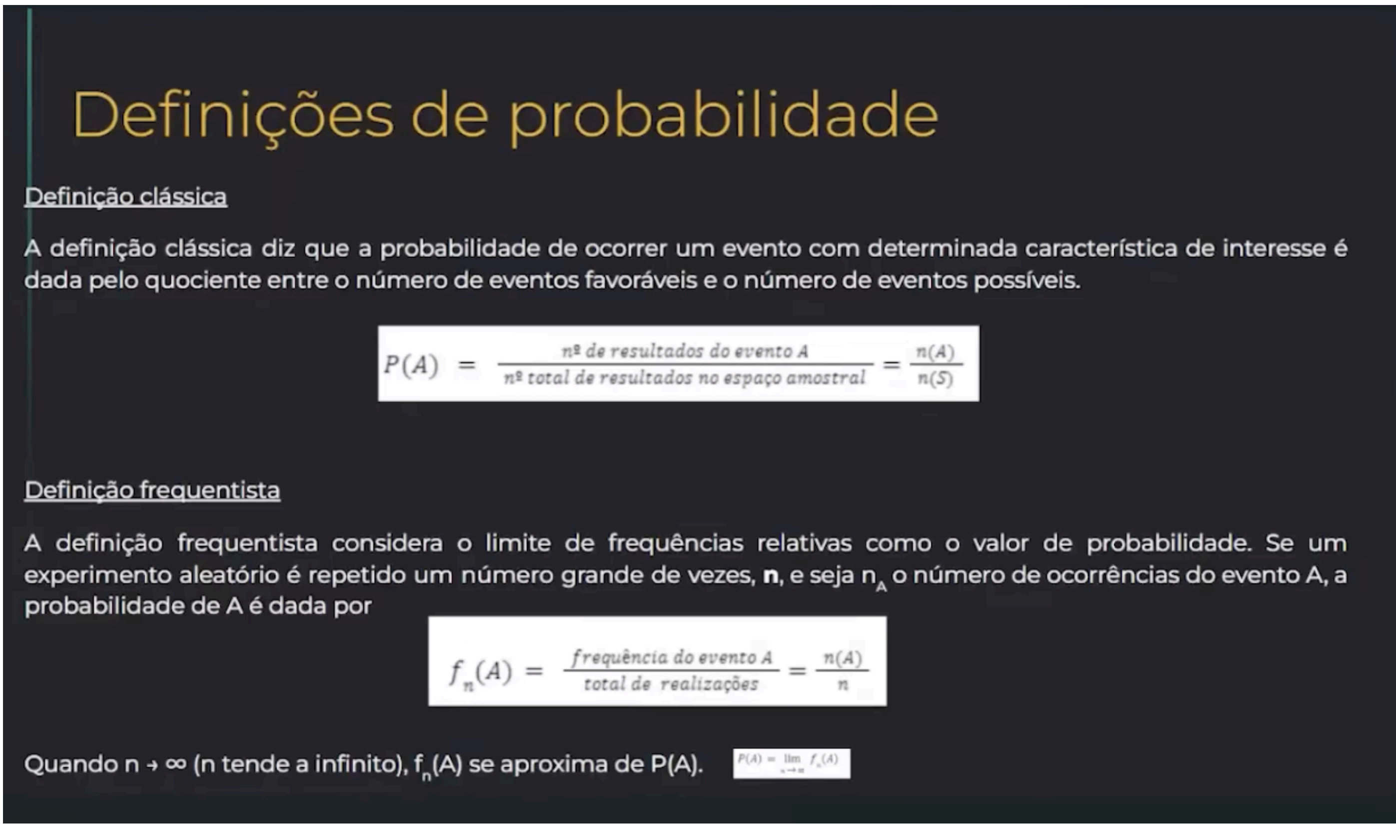
Correlação

- **Coeficiente de correlação**

- Também conhecido como correlação de Pearson ® é um grau de relação entre duas variáveis quantitativas e indica o grau de intensidade da correlação, além do sentido dessa correlação (positiva ou negativa).
- Seus limites vão de -1 a +1.
- se r = +1 há uma correlação perfeita e positiva entre as variáveis
- se r = -1 há uma correlação perfeita e negativa entre as variáveis
- se r = 0 nã0 há correlação entre as variáveis ou a relação que porventura existe, não é linear.

- Correlação não implica em causalidade.

Probabilidade



Definição axiomática

A probabilidade de um evento A é definida como sendo um número P(A) satisfazendo os axiomas:

(i) $0 \leq P(A) \leq 1$, para todo evento A contido em S.

(ii) $P(S) = 1$

(iii) $P\left(\bigcup_{i=1}^{\infty}\right) = \sum_{i=1}^{\infty} P(A_i)$, se $A_1, A_2, \ldots$, são eventos mutuamente exclusivos.

Decorrem da definição, as seguintes propriedades:

- $P(\phi) = 0$
- para todo evento A de S, $P(A) = 1 - P(\overline{A})$
- se A e B são dois eventos quaisquer de S, então: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Eventos independentes

Dizemos que dois eventos são independentes quando a realização ou não-realização de um dos eventos não afeta a probabilidade da realização do outro e vice-versa.

Dois eventos são independentes se e somente se P(AnB)=P(A)*P(B)

Eventos mutuamente exclusivos

Eventos mutuamente exclusivos (excludentes) ou eventos disjuntos. Dizemos que dois eventos são disjuntos ou mutuamente excludentes quando a realização de um exclui a realização do(s) outro(s).

Assim, no lançamento de uma moeda, o evento "tirar cara" e o evento "tirar coroa" são mutuamente excludentes, já que, ao se realizar um deles, o outro não se realiza.

Se dois eventos são mutuamente excludentes, AnB=vazio e portanto P(AnB)=0.

Probabilidade condicional

Probabilidade condicional é um dos conceitos mais importantes da teoria de probabilidades, pois frequentemente se quer calcular seu valor quando se tem alguma informação parcial a respeito do resultado de um experimento aleatório.

Além disso, mesmo quando não se tem essa informação parcial, as probabilidades condicionais muitas vezes são utilizadas para computar mais facilmente valores de probabilidades que se tem interesse.

Mais importante do que isso, o conceito por trás de probabilidades condicionais é de que nós conseguimos atualizar a nossa medida de incerteza a partir do momento que conhecemos uma informação nova.

Sejam A e B dois eventos de um espaço amostral S, com P(A)>0.

A probabilidade condicional de B, dado A, denotada por P(BIA), é definida por:

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

Análise Combinatória

Arranjo (Ordenados)

- Usamos o arranjo para calcular o número de possibilidades para formar um grupo ordenado de p elementos, dentre n elementos disponíveis. O número de possibilidade para um arranjo é:

$$A_{n,p} = \frac{n!}{(n-p)!}$$

- Sendo n! fatorial
  - n! = n . (n-1) . (n-2) . (n-3) . (n-4) … 3 . 2 . 1
  - 5! = 5 . 4 . 3 . 2 . 1

Combinação (Não importa a ordem)

- Usamos a combinação para formar grupos não ordenados de p elementos, dentre os n disponíveis. A principal diferença para o arranjo é que na combinação a ordem não importa. A combinação é dada por:

$$C_{n,p} = \frac{n!}{p!(n-p)!}$$

Probabilidade Total

- O Teorema da Probabilidade Total diz que a probabilidade de um evento B ocorrer é a soma das probabilidades das intersecções do evento B com as partições A de outro evento A:

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Ou seja, a probabilidade de B ocorrer é a somatória da probabilidade de B dada a partição vezes a probabilidade da partição.

Teorema de Bayes

- O teorema de bayes deriva da probabilidade condicional. Ele relaciona as probabilidades condicionais entre dois eventos da seguinte forma:

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