

# Resumo Probabilidade e Estatística



Variáveis → características de interesse da pesquisa

- Variáveis:
  - Qualitativas: não numéricas
    - Nominais: não é possível estabelecer ordem natural entre seus valores (A ou B, sim ou não)
    - Ordinais: o atributo tem ordenação natural (ruim, média ou boa, pequeno, médio ou grande)
  - Quantitativas: numéricas
    - Discretas: obtidas a partir de contagem (número de filhos)
    - Contínuas: obtidas por mensuração (altura)

Tabelas de frequência: forma resumida da tabela dos dados brutos.

- Variáveis discretas: Consiste em listar os possíveis valores da variável e fazer a contagem do número de ocorrências na tabela de dados brutos.
  - $n_i$  = frequência de ocorrência da variável  $i$  e  $n$  é a frequência total.
  - Frequência relativa:  $f_i = \frac{n_i}{n}$
- Variáveis ordinais: acrescentar a frequência acumulada para termos pontos de corte com uma determinada frequência nos valores das variáveis
  - $f_{ac}$  = somatório das frequências de todos os valores da variável menores ou iguais ao valor considerado

Idade	$n_i$	$f_i$	$f_{ac}$
17	9	0,18	0,18
18	22	0,44	0,62
19	7	0,14	0,76

Classes ou faixas: para variáveis quantitativas contínuas ou discretas

- Ex: de 40kg (inclusive) até 50kg (exclusive) =>  $40 \leq 50$  ou  $[40,50)$

Medidas de resumo: informações numéricas sobre um conjunto de dados

- Medidas de posição (tendência central)

- Média:

$$\bar{x}_{obs} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

- Mediana: valor central depois que está ordenado  
Se n for ímpar:  $md_{obs} = valorpos.\left(\frac{n+1}{2}\right)$   
Se n for par:  $md_{obs} = \frac{valorpos.\left(\frac{n}{2}\right) + valorpos.\left(\frac{n+2}{2}\right)}{2}$
- Moda: é o valor mais frequente, com maior ocorrência.
  - Todos os valores com a mesma frequência de ocorrência: não tem moda
  - K valores tem a mesma frequência de ocorrência: tem k modas
- Medidas de dispersão
  - Amplitude: diferença entre o maior e o menor valor ( $\Delta$ )
  - Desvio mediano: somatório dos módulos da distância de cada valor até a mediana

$$desvio\ mediano = \frac{1}{n} \sum_{i=1}^n |x_i - md_{obs}|$$

- Desvio médio: somatório dos módulos da distância de cada valor até a média

$$desvio\ médio = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}_{obs}|$$

- Variância: somatório dos quadrados da distância de cada valor até a média:

$$var_{obs} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_{obs})^2$$

- Desvio padrão: raiz quadrada do somatório dos quadrados da distância de cada valor até a média, ou seja, raiz quadrada da variância

$$dp_{obs} = \sqrt{var_{obs}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_{obs})^2}$$

## Probabilidade

Fenômeno aleatório → Situação cujos resultados não podem ser previstos com certeza

Espaço amostral → conjunto de todos os resultados possíveis. Representado por  $\Omega$

↳ Subconjuntos → eventos

- União de eventos: ocorrência de pelo menos um dos eventos A ou B:  $A \cup B$
- Intersecção de eventos: ocorrência simultânea de A e B:  $A \cap B$
- Eventos disjuntos: quando não tem elementos em comum:  $A \cap B = \emptyset$
- Eventos complementares: se sua união é o espaço amostral e intersecção vazia:
- $A \cup B = \Omega$  e  $A \cap B = \emptyset$
- Evento A ocorre mas o B não:  $A \cap B^c$
- Nenhum deles ocorre:  $A^c \cap B^c$

- Exatamente um deles ocorre:  $(A^c \cap B) \cup (A \cap B^c)$
- Probabilidade da união de dois eventos:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidade condicional: uma informação anterior influencia em posteriores.

- Probabilidade de A dado que ocorreu B:  $P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$ 
  - Se  $P(B) = 0$ , então  $P(A|B) = P(A)$
- Eventos independentes: se a ocorrência ou não de B não altera a probabilidade de A:  

$$P(A \cap B) = P(A)P(B)$$

Probabilidade total:  $P(A) = \sum_{i=1}^n P(F_i) * P(A|F_i)$

Teorema de Bayes: basicamente é a probabilidade condicional de um cara acontecer dividido pela probabilidade total.