Repositório de Estatistica e Probabilidade

UFSC Joinville - EMB5010
Artur Gemaque

16 de outubro de 2024

Resumo

Este documento tempo como principal funcionalidade registrar os contudos ensinados em sala de aula pelo professor Jaimes, ademais servirá como fonte de estudo para as provas referentes há Matéria.

1 Estatistica

Vamos começar com as definições mais simplórias da matéria que são resultado do cálculo dos dados.

1.1 Média

A média aritmética de um conjunto de n números é a soma desses números dividida por n.

$$\tilde{x} = \sum_{i=1}^{n} \frac{X_i}{n}$$

1.2 Variância

A variância é um conceito fundamental em estatística que mede a dispersão de um conjunto de dados em relação à sua média.

$$(S)^2 = \frac{\sum_{i=1}^{n} (X_i - \tilde{x})^2}{n-1}$$

1.3 Desvio padrão

O desvio padrão é uma medida de dispersão estatística que indica o quão afastados os dados de um conjunto estão da sua média. Ele é a raiz quadrada da variância, o que o torna mais interpretável.

$$S = \sqrt[2]{S^2}$$

1.4 Valores de análise

Partindo agora para valores que são obtidos apartir dos dados ordenados em formato crescente. Observação que as seguintes fórmulas serão para que possamos obter as posição dos referidos valores.

1.4.1 Mediana

Mediana é o número no centro de um grupo de números.

$$Md = X_{(\frac{n+1}{2})}$$

1.4.2 Quartis

Os quartis são valores que dividem uma amostra de dados em quatro partes iguais e são usados para avaliar a dispersão e a tendência central de um conjunto de dados. Eles são os valores contidos nas posições de $n * 25\%$ entre outras porcentagens, caso "n"dê um valor quebrado o Quartil vai ser a média entre os dois valores.

$$Q_1 = X_{(\frac{n+1}{4})}$$
$$Q_3 = X_{(\frac{3(n+1)}{4})}$$

1.4.3 Moda

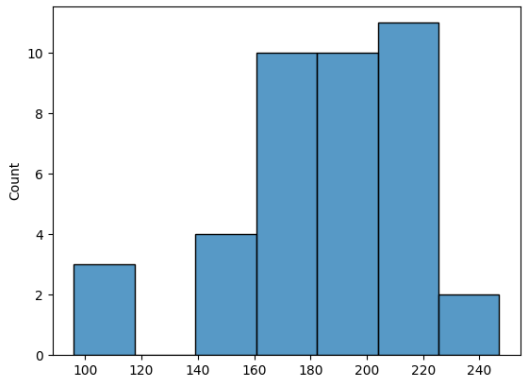
A moda é o valor que mais aparece em um conjunto de dados, ou seja, o valor que tem maior frequência.

1.4.4 Amplitude

A amplitude de um conjunto de dados é a diferença entre o maior e o menor valor. Para calcular a amplitude, subtrai-se o menor valor do maior.

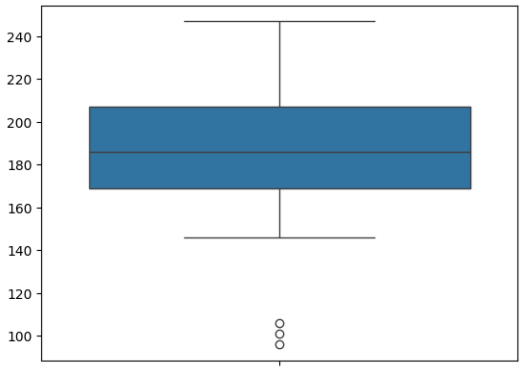
1.5 Histograma

Um histograma é uma espécie de gráfico de barras que demonstra uma distribuição de frequências. No histograma, a base de cada uma das barras representa uma classe e a altura representa a quantidade ou frequência absoluta com que o valor de cada classe ocorre.



1.6 Boxplot

O Box Plot, que estudamos no curso Green Belt, é uma ferramenta gráfica que ajuda a identificar a existência de possíveis outliers no conjunto de dados. Em um boxplot são apresentadas 5 estatísticas: o mínimo, o primeiro quartil (Q1), a mediana, o terceiro quartil (Q3) e o máximo.



2 Axiomas da Probabilidade

Sejam A_i num espaço amostral Ω :

1. $0 \leqslant P(A_i) \leqslant 1$
2. $P(\Omega) = 1$
3. $P(A_1 \bigcup A_2) = P(A_1) + P(A_2)$
   $\Leftrightarrow P(A_1 \cap A_2) = 0$

Algumas propriedades decorrentes dos axiomas: spacing

- $P(\varnothing) = 0$
- $P(A) + P(\bar{A}) = 1 \ \rightarrow P(\bar{A}) = 1 - P(A)$
- $P(A_1 \bigcup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

2.1 Teorema da Probabilidade Total e Teorema de Bayes

Exemplo 1

Um indivíduo possui 3 contas de e-mail diferentes.

Do total de mensagens que ele recebe:

- 70% na conta 1
- 20% na conta 2
- 10% na conta 3

Mensagens que são SPAM

- 1% das mensagens da conta 1
- 2% das mensagens da conta 2
- 5% das mensagens da conta 3

Questão:

1. Qual a Probabilidade de uma mensagem selecionada aleatoriamente ser SPAM ?

Partindo que C_1 representa a probabilidade de uma mensagem chegar na conta 1 e S_1 a probabilidade de receber SPAM

- $P(S) = [C_1 \bigcap S] \bigcup [C_2 \bigcap S] \bigcup [C_3 \bigcap S]$
- $P(S) = P(C_1 \bigcap S) + P(C_2 \bigcap S) + P(C_3 \bigcap S)$
- $P(S) = P(C_1)P(\frac{S}{C_1}) + P(C_2)P(\frac{S}{C_2}) + P(C_3)P(\frac{S}{C_3})$
- $P(S) = (0,7) * (0,01) + (0,2) * (0,02) + (0,1) * (0,05)$
- $P(S) = 0,0160 \rightarrow 1,6\%$ De se receber um SPAM.

2. Sabendo que uma mensagem selecionada aleatoriamente é SPAM qual a probabilidade de que ela tenha sido recebida pela conta 3?

Primeiro redusimos nosso espaço amostral para as mensagens SPAM $1,6\%$ e ultilizamos a definição de probabilidade

$$P(\frac{C_3}{S}) = \frac{P(C_3 \bigcap S)}{P(S)} \rightarrow \frac{10\% * 5\%}{1,6\%} = 31,25\%$$

Portanto, a probabilidade total é determinada apartir da fórmula

$$P(S) = P(E_1)P(\frac{F}{E_1}) + P(E_2)P(\frac{F}{E_2}) + \cdots + P(E_k)P(\frac{F}{E_k})$$

Exemplo 2

Uma doença "rara"acontece 1 em 1000 adultos. Um teste diagnóstifico foi desenvolvido, o qual tem o seguinte desempenho:

- Se o indivíduo testado tiver a doença, o teste resulta positivo 99% das vezes
- Se o indivíduo testado **Não** tiver a doença, o teste resulta positivo 2% das vezes

Questão

1. Se um indivíduo selecionado aleatoriamente foi testado, e o resultado for positivo, qual a probabilidade de ele de fato ter a doença?

Faz ae!

$$P(p) = 4,72\% \text{ resultado!}$$

2.2 Variável Aleatória