Planejamento de Experimentos em Engenharia

Rodrigo R. Pescim

5 de maio de 2020

- Um engenheiro é aquele que resolve problemas pela utilização de **princípios científicos**
- O princípio científico é a abordagem para formular e resolver os problemas por meio do **método científico**, que possui as seguintes etapas:
  - Desenvolver uma descrição clara e concisa do problema
  - Identificar os fatores importantes que afetam esse problema
  - Propor um modelo (teórico) para o problema
  - Conduzir experimentos apropriados e coletar dados para testar ou validar o modelo proposto
  - Tirar conclusões ou fazer recomendações baseadas na solução do problema

- A **ciência estatística** tem por interesse a coleta, a apresentação, a análise e a utilização dos dados para tomar decisões, resolver problemas e planejar processos

- Os **métodos estatísticos** são utilizados para compreender o conceito de variabilidade

- **Variabilidade** $\rightarrow$ "Sucessivas observações de um fenômeno que não produz exatamente o mesmo resultado"

Exemplo 1

- Algumas estruturas nas áreas de construção e engenharia civil são expostas a "forças naturais".

- Essas estruturas podem sofrer, ao longo do tempo, **degradação** (deterioração, fadiga, deformação, etc.) e também sofrer do efeito de **fatores externos** (corrosão, sobrecarga ou riscos ambientais).

- Assim, **as respostas associadas a esses fatores** nas estruturas não devem ser consideradas constante, mas sim, como **uma variável aleatória ao longo do tempo**.

Exemplo 2

- Um artigo do periódico *Quality Engineering* apresenta dados de viscosidade de um processo químico em batelada. Uma amostra desses dados é apresentada a seguir.

1,570	1,720	1,900	2,400	2,522	2,700
2,720	2,750	2,800	3,125	3,200	3,250
3,400	3,450	3,600	3,720	4,100	4,600

- Como retirar as primeiras informações sobre a viscosidade de um fluido num processo químico ?

Estatística Descritiva

- Tem por objetivo resumir e apresentar os dados sob a forma de **tabelas e gráficos**, além de estudar os parâmetros e suas estimativas.

- **Organização**: Como "tratar" os dados a fim de extrair informações a respeito de uma ou mais características de interesse ?

- **Variáveis**: Uma variável, neste contexto, é uma medida ou classificação obtida de cada elemento da população ou da amostra.
- Exemplo:
  - Variável $X \rightarrow$ refere-se a viscosidade de um processo químico
  - x_i, $i = 1, \ldots, 18$ é o valor observado da viscosidade em cada unidade experimental
  - $x_1 = 1,570$; $x_2 = 1,720$; $x_3 = 1,900$; $\ldots$

Tipos de Variáveis

- Os tipos de variáveis mais comumente utilizadas na descrição dos dados são: Quantitativas (Discretas ou Contínuas) e Qualitativas (nominais ou ordinais)
- Exemplos de variáveis quantitativas discretas:
    - Número de peças defeituosas;
    - Número de bits transmitidos que foram recebidos com erros;
    - Número de lâmpadas queimadas
- Exemplos de variáveis quantitativas contínuas:
    - Intensidade de corrente elétrica;
    - Tempo de falha de algum componente mecânico;
    - Medidas de voltagem, comprimento, pressão, etc

- Exemplos de variáveis qualitativas nominais (ou categóricas):
    - Sexo;
    - Nacionalidade;
    - Área de atividade

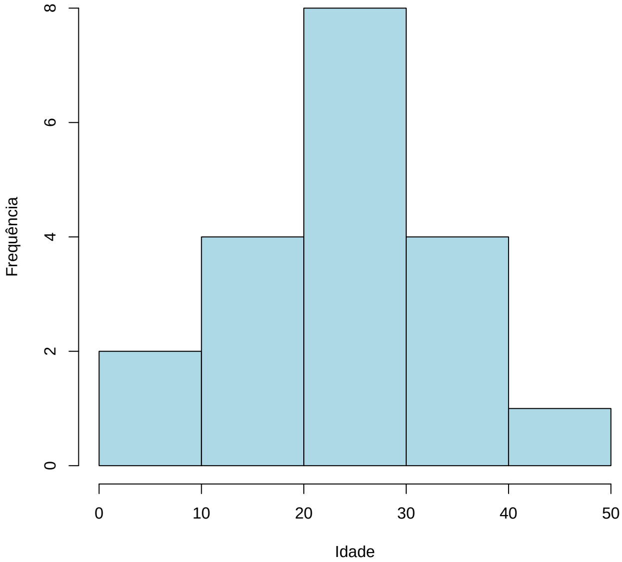
- Exemplos de variáveis qualitativas ordinais:
    - Classes sociais;
    - Nível de instrução;
    - Categoria de clientes (ouro, prata e bronze)

Histograma

- Quando os dados estão agrupados em classes - o histograma apresenta as frequências das classes em colunas
- As frequências representadas podem ser simples ou relativas
- As colunas possuem bases com mesma largura
- Não existe espaço entre as classes

Exemplo: Tabela para dados quantitativos contínuos organizados em classes de frequências

Classe	f_i
$0 \vdash 10$	2
$10 \vdash 20$	3
$20 \vdash 30$	9
$30 \vdash 40$	4
$40 \vdash 50$	1


Frequência
0
2
4
6
8
0
10
20
30
40
50
Idade

Medidas de Posição

- Às vezes, tem-se o interesse em resumir um conjunto de dados, apresentando um ou alguns valores que representem todo conjunto de dados.

- Para determinar esses valores, utiliza-se as **medidas de posição central**: Média, Mediana e Moda.

Média Amostral

- Sejam $X_1, \ldots, X_n$, n valores amostrais. A média amostral ($\bar{X}$) é definida por

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + \ldots + X_n}{n}$$

- Exemplo: Notas das provas dos alunos da UEL: 6,4,5,4,6
- A média é considerada a **melhor medida de centro**, pois possui boas propriedades matemáticas
- Entretanto, ela é **falha** quando alguns dos valores estão muito afastados da maioria dos dados

Mediana e Moda

- A Mediana (Md) divide um conjunto ordenado (ordem crescente) de valores em duas partes iguais. É definida por

$$Md = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se} \quad \text{n for ímpar} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se} \quad \text{n for par} \end{cases}$$

- A Moda (Mo) é definida como a realização mais frequente do conjunto de dados observados.

Medidas de Dispersão

- Às vezes, tem-se o interesse em resumir um conjunto de dados, apresentando um ou alguns valores que representem todo conjunto de dados
- Exemplos:
  - Aluno A (variável X): 6,4,5,4,6
  - Aluno B (variável Y): 5,5,5,5,5
  - Aluno C (variável Z): 10,10,5,0,0
- Observa-se $\bar{X} = \bar{Y} = \bar{Z} = 5$, porém nada importa sobre suas diferentes variabilidades
- Dessa forma, deve-se construir medidas que sumarizem a variabilidade nos dados. Tais medidas são: Variância, desvio-padrão e coeficiente de variação

Variância e desvio-padrão

- A variância (s^2) fornece a dispersão dos dados em torno do valor central (média)
- Sejam $X_1, \ldots, X_n$, n valores amostrais com média $\bar{X}$. É definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- O **desvio-padrão (s)** é definido como a raiz quadrada da variância
- O desvio-padrão indica, em média, qual será o erro (desvio) cometido ao tentar substituir cada observação pela média

- O CV é uma medida de dispersão relativa utilizada para comparar variáveis.

$$CV = \frac{s}{\bar{X}}$$

- Observa-se que quanto menor for o valor do CV, mais **homogêneo** é um conjunto de dados

Quartis

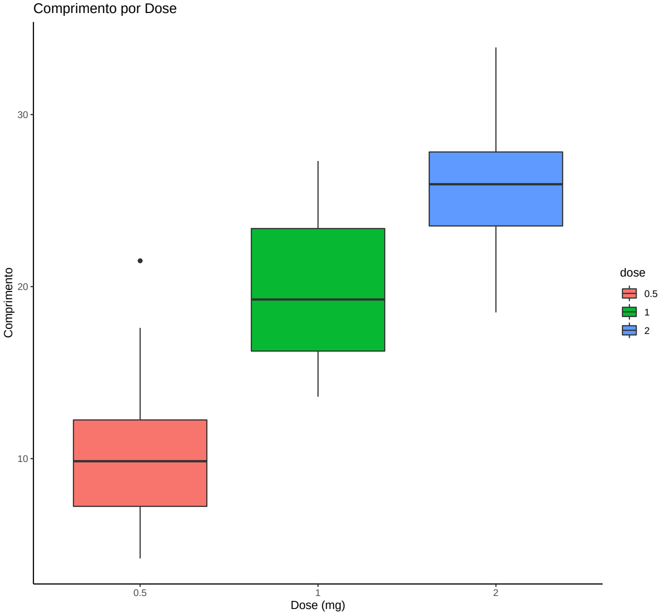
- Dê forma análoga a mediana, um conjunto de dados pode ser dividido em 4, 10, 100, etc ...
- Quartis: Divide o conjunto de dados em quatro partes iguais

$$\text{Quartis} : \begin{cases} Q_1 & \rightarrow & 1^o \text{ Quartil } (25\%) \\ Q_2 & \rightarrow & 2^o \text{ Quartil } (50\%) \\ Q_3 & \rightarrow & 3^o \text{ Quartil } (75\%) \end{cases}$$

- $Q_1 = X_{\left(\frac{n}{4}+\frac{1}{2}\right)}$ e $Q_3 = X_{\left(\frac{3n}{4}+\frac{1}{2}\right)}$

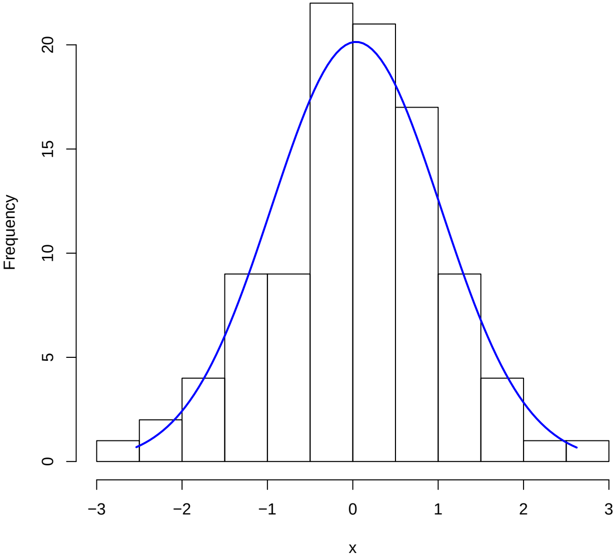
Box-plot

- É um gráfico construído com base no resumo de cinco medidas, constituído por:
  - Limite mínimo (L_1)
  - 1^o Quartil (Q_1)
  - 2^o Quartil (Q_2)
  - 3^o Quartil (Q_3)
  - Limite máximo (L_1')
- Para o cálculo do limite mínimo e máximo do boxplot, temos:
  - $L_1 = \text{máx}\{x_{min} \; ; \; Q_1 - 1,5(Q_3 - Q_1)\}$
  - $L_1' = \min\{x_{max} \; ; \; Q_3 + 1,5(Q_3 - Q_1)\}$
- Por meio do box-plot pode-se verificar: Assimetria nos dados, discutir pontos influentes (possíveis outliers), estudar a variabilidade do conjunto de dados.


Comprimento por Dose
Comprimento
30
20
10
0.5
1
2
Dose (mg)
dose
0.5
1
2

Modelos de Probabilidade

- Por meio da Estatística Descritiva, pode-se descrever o **comportamento** de uma característica (variável)
- Sob o ponto de vista da teoria da probabilidade, o comportamento de uma variável refere-se a **distribuição dos dados**
- Consequência: O histograma é de grande importância na identificação dos **modelos adequados aos dados**
- Dizemos, então, que variáveis que apresentam um mesmo padrão de comportamento seguem uma mesmo **modelo (ou distribuição) de probabilidade**


Frequency
0
5
10
15
20
−3
−2
−1
0
1
2
3
x

Distribuição de Probabilidade

- Distribuição de probabilidade é definida como a descrição de um fenômeno aleatório (ou variável aleatória)
- As principais distribuições de probabilidade são:
  - Distribuição Bernoulli
  - Distribuição Binomial
  - Distribuição de Poisson
  - **Distribuição Normal**

Distribuição Normal (ou Gaussiana)

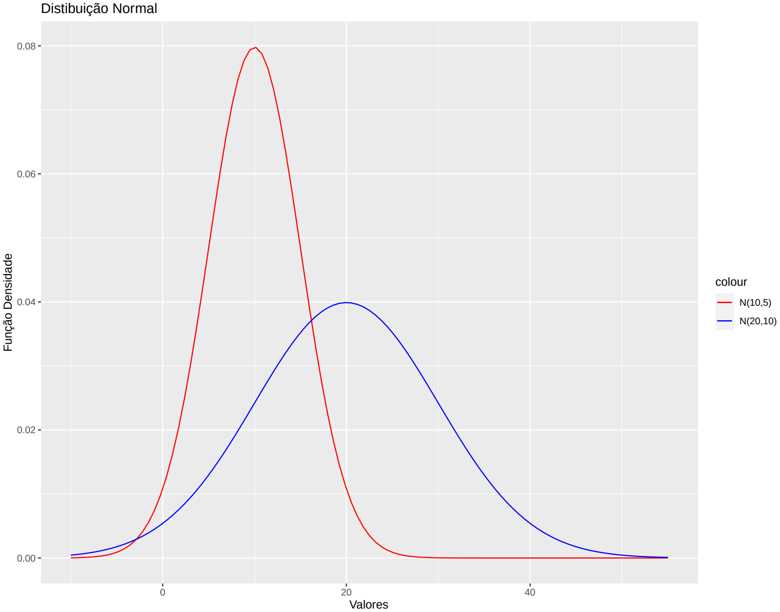
- É a distribuição mais importante para variáveis aleatórias, tanto do ponto de vista teórico como prático

- Uma variável aleatória X segue distribuição normal se sua função densidade é representada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ \mathrm{e}^{-\frac{1}{2}\left(\frac{\mathrm{x}-\mu}{\sigma}\right)^2},$$

  em que $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ é o parâmetro de locação e $\sigma > 0$ é o parâmetro de escala

- Notação : $X \sim N(\mu, \sigma^2)$

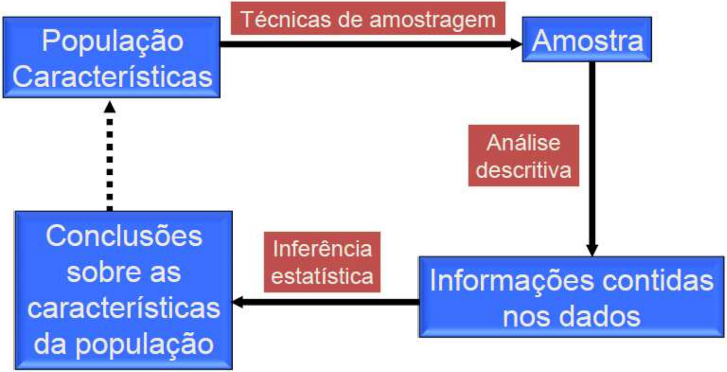

Distibuição Normal
0.08
0.06
0.04
0.02
0.00
Função Densidade
0
20
40
Valores
colour
N(10,5)
N(20,10)

Aplicações da Distribuição Normal

1. Experimentos astronômicos;
2. Experimentos meteorológicos e Precipitações;
3. Medições de peças manufaturadas;
4. Erros em medições científicas;
5. Quantidades como resistência à compressão, temperatura, pressão, volume, densidade, etc, podem ser descritas pela distribuição normal

- Conjunto de métodos utilizados para tomar decisões ou tirar conclusões acerca de uma **população** utilizando a informação contida numa **amostra**
- **Situação 1**: Considere que um engenheiro civil esteja analisando a resistência à compressão do concreto.
  - Observa-se que o interesse está na estimação da **resistência média** à compressão.
  - Na pratica, ele utiliza **dados de uma amostra para calcular um valor razoável da verdadeira média (populacional)**
  - Esse valor é chamado de **estimativa!**

Etapas da Análise Estatística



- **Situação 2** : Considere duas temperaturas de reação, t_1 e t_2 tais que $t_1 \neq t_2$, que possam ser utilizadas num processo químico
    - O pesquisador conjectura que t_1 resulta em **rendimentos maiores** que t_2
    - Observa-se que, nesse caso, o interesse não está na estimação de rendimentos!
    - O foco da pesquisa está na tomada de decisão acerca de uma hipótese estabelecida (t_1 tem maior rendimento que t_2)

Intervalos de Confiança

- Em muitas situações, uma estimativa pontual não fornece informação completa sobre um parâmetro desconhecido

- Alternativa: Obter um intervalo (de confiança) para expressar o grau de incerteza associado com uma estimativa

- Esse processo de estimação, é chamado de **estimação intervalar ou intervalos de confiança (IC)**, que incorpora à estimativa pontual e fornece informações a respeito de sua variabilidade

- **Definição**: Um intervalo de confiança representa uma amplitude de valores que tem alta probabilidade (grau de confiança) de conter o verdadeiro valor do parâmetro

- O grau de confiança (ou nível de confiança) é uma medida que representa a probabilidade do intervalo conter o parâmetro populacional. Tal probabilidade é denotada por $(1-\alpha)$

- Logo, α (nível de significância) será a probabilidade de erro ao se afirmar que o intervalo contém o verdadeiro valor do parâmetro

- Se $\alpha = 0,05$ então $(1-\alpha) = 0,95$

- Isso significa que o procedimento de construção do intervalo é tal que em 95% das possíveis amostras, o intervalo de confiança obtido conterá o verdadeiro valor do parâmetro

IC para a média populacional (μ) com variância desconhecida (σ^2)

- As pressuposições para o uso desse intervalo de confiança são:
- σ^2 desconhecida, e
- $n < 30$
- O intervalo dé confiança é calculado utilizando-se a estatística

$$T = \frac{\overline{Y} - \mu}{s/\sqrt{n}} \sim t_{n-1},$$

t_{n-1} tem distribuição t de Student, com $(n-1)$ graus de liberdade

- Se o desvio padrão populacional σ for desconhecido, σ pode ser estimado a partir do desvio padrão amostral s

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

$$P\left[\overline{x} - t_{\left(\frac{\alpha}{2}, n-1\right)}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\left(\frac{\alpha}{2}, n-1\right)}\frac{s}{\sqrt{n}}\right] = 1 - \alpha$$

ou

$$IC[\mu; (1-\alpha)\%] = \left[\overline{x} - t_{\left(\frac{\alpha}{2}, n-1\right)}\frac{s}{\sqrt{n}}\, ; \,\overline{x} + t_{\left(\frac{\alpha}{2}, n-1\right)}\frac{s}{\sqrt{n}}\right]$$

  - em que $\overline{x}$ é a média amostral, s é desvio padrão amostral, n é tamanho amostral e $t_{[(\alpha/2), n-1]}$ com $(n-1)$ graus de liberdade, pode ser encontrado na tabela da distribuição t de Student

Exercício

- Um pesquisador deseja realizar um estudo sobre a força de remoção de oito conectores de náilon. Para isso, foi realizado um experimento e os resultados estão descritos na Tabela 1.

Tabela 1: Força de remoção (libras-pés) dos conectores de náilon

Conector	1	2	3	4	5	6	7	8
Força de Remoção	12,6	12,9	13,4	12,3	13,6	13,5	12,6	13,1

- Construir um intervalo de confiança para a média da força de remoção dos conectores de náilon, ao nível de confiança de 95%