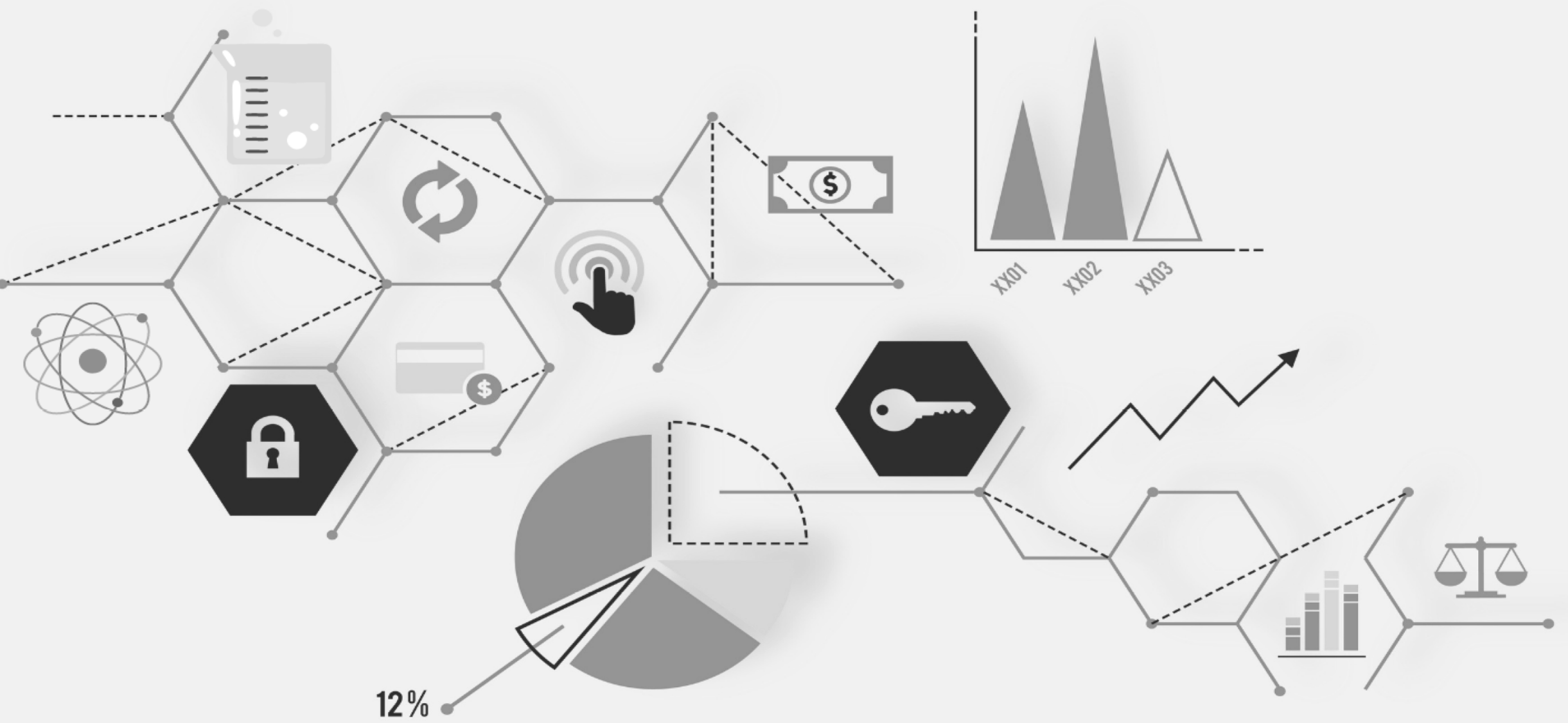
CURSO DE ESTATÍSTICA PARA TODOS

Uilla Fava Pimentel
Gildeir Lima Rabello
Autores

ARCO EDITORES

CURSO DE ESTATÍSTICA PARA TODOS



Uilla Fava Pimentel
Gildeir Lima Rabello
Autores

ARCO EDITORES

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Pimentel, Uilla Fava
Curso de estatística para todos [livro eletrônico] / Uilla Fava Pimentel, Gildeir Lima Rabello. -- Santa Maria, RS : Arco Editores, 2022.
PDF

ISBN 978-65-5417-065-9

1. Estatística - Estudo e ensino 2. Matemática - Estudo e ensino I. Rabello, Gildeir Lima. II. Título.

22-136884 CDD-519.507

Índices para catálogo sistemático:

1. Estatística : Estudo e ensino 519.507

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

10.48209/978-65-5417-065-9

Diagramação e Projeto Gráfico: Gabriel Eldereti Machado

Imagem capa: Designed by canva

Revisão: Organizadores e Autores(as)

ARCO EDITORES

Telefone: 5599723-4952

contato@arcoeditores.com

www.arcoeditores.com

Apresentação

Este livro foi baseado nas principais dificuldades dos alunos durante as monitorias de estatística básica na UFES e em experiências adquiridas na academia (IFF, IFES, UFES e UFRJ). O livro traz uma abordagem simples e clara, com ilustrações e exemplos. Além de conceitos matemáticos básicos que normalmente são utilizados em estatística.

Prefácio

A estatística é uma ciência que se dedica a organização, análise e interpretação de dados provenientes de um conjunto experimental. A estatística está inserida em diversas áreas do conhecimento, por conseguinte, é necessário o seu entendimento para utilizá-la como uma ferramenta eficaz na pesquisa. A estatística integra diferentes campos de estudo, por isso, faz parte da grade curricular de diversos cursos desde engenharias ao direito.

Por está presente em diversos cursos de graduação, os alunos de estatística básica apresentam diferentes níveis de conhecimento em matemática, assim é necessário a realização de um nivelamento de conceitos matemáticos para que os alunos consigam compreender a disciplina de estatística de forma eficaz. Dessa forma, observa-se a existência de diversas dúvidas em conceitos básicos de matemática e por consequência dificuldades em cursos de estatística básica proveniente da realização desses cálculos matemáticos. Por isso, é recomendado que as disciplinas de estatísticas básicas contemplem a apresentação de conceitos matemáticos básicos como somatório, produtos notáveis entre outros. Buscando, portanto, suprir esse empecilho, o primeiro capítulo apresenta conceitos, propriedades e exemplos de técnicas matemáticas básicas necessárias ao entendimento da estatística.

Os demais capítulos tratam de conteúdos de estatística, a estrutura dos capítulos irá apresentar conceitos, definições e exemplos práticos da utilização da estatística na área da química, engenharia química, engenharia de alimentos entre outras áreas que aplicam a estatística básica. O segundo capítulo é uma introdução ao campo da estatística, em que se apresenta definições e conceitos fundamentais ao entendimento da estatística. Além disso, esse capítulo apresente demonstração prática das formas de apresentação dos dados (tabulados e gráficos).

O terceiro capítulo trata da estatística descritiva (análise exploratória dos dados) e das suas medidas de posição e dispersão. São apresentados conceitos básicos, formulação e exemplo de aplicação.

O quarto capítulo dedica-se ao estudo da inferência estatística, apresentando conceitos sobre testes de hipótese. Para esse capítulo foi dado uma abordagem lógica com presentação dos testes, método do cálculo de ANOVA e um exemplo de resultados obtidos por pacotes estatísticos.

O quinto e último capítulo é uma introdução ao planejamento experimental, englobando conceitos básicos e um exemplo ilustrativo de planejamento fatorial.

Vale destacar, que vários exemplos apresentados nesse livro possuem arquivos disponíveis para downloads. Logo, espera-se que este livro possa facilitar o entendimento da estatística básica, pois busca tornar a linguagem acessível e prática.

Os autores.

Sumário

CAPÍTULO 1

REVISÃO DE CONCEITOS MATEMÁTICOS BÁSICOS

Em estatística básica algumas propriedades matemáticas são necessárias, principalmente operações que envolvem somatórios, produtórios e produtos notáveis. Com isso é fundamental a compreensões básicas dessas operações para um melhor aprendizado.

1.1 Somatório

Considere que uma criança deseja brincar empilhando copos plásticos (Figura 1.1.1).

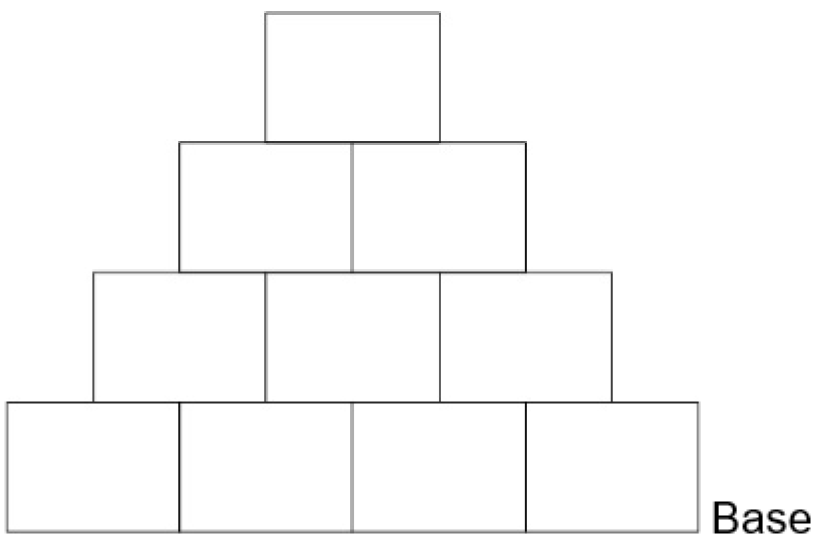


Figura 1.1.1. Representação de uma torre de copos

Para tal finalidade, qual é o mínimo de copos plásticos que precisam ser disponibilizados à criança para que a base da torre de copos contenha 4 copos? Observe que essa questão pode ser facilmente respondida se somar todos os números inteiros de 1 a 4, representado pela equação 1.

$$1+2+3+4 \tag{1.1}$$

Contudo, agora supõem que deseja saber quantos copos plásticos é necessário para construir uma torre com uma base de 100 copos. De forma análoga a Equação 1.1, podemos concluir que é necessário somar todos os números inteiros de 1 a 100 (Equação 1.2).

$$1+2+3+4+5+6+4+8+9+10+\ldots+98+99+100 \tag{1.2}$$

Apesar da Equação 1.2 conseguir representar o problema, é uma notação que não se adequa bem a maioria dos casos matemáticos e estatísticos, devido a sua extensão e não visualização de todos os termos. Por outro lado, caso a escrita da Equação 2 contenha todos os termos, iríamos conseguir descreve--los, porém uma dificuldade de visualização poderia ocorrer devido a números de termos escritos. Assim, é conveniente utilizar uma simbologia que consiga compactar a Equação 2, de forma a não injuriar a sua interpretação. A representação que normalmente é utilizada é a letra grega sigma (Σ), assim a Equação 2 pode ser compactada pela Equação 3.

$$\sum_{i=1}^{100} (i) \tag{1.3}$$

Nessa notação o número 1 representa o limite inferior do índice, o número 100 representa o limite superior do índice, a variável i localizada na parte inferior da letra Σ é o índice e a operação entre parênteses é o termo do somatório.

Nesses cenários, algumas propriedades de somatórios são normalmente utilizadas em operações matemáticas.

1. Termo constante (k)

$$\sum_{i=1}^{n}(k) = k + k + k + \cdots + k = n * k \qquad 1.4$$

Exemplo e Aplicação

Suponha que a criança gostaria de realizar 5 torres de copos com 4 copos em sua base. Sabendo que cada torre necessita de 10 copos, podemos representar essa questão pela Equação 1.5.

$$\sum_{i=1}^{5}(10) = 10 + 10 + 10 + 10 + 10 = 5 * 10 = 50 \qquad 1.5$$

2. Somatório do produto entre uma constante e um termo.

$$\sum_{i=1}^{n}(k * a_i) = k * a_1 + k * a_2 + \cdots + k * a_n = k * (a_1 + a_2 + \cdots + a_n) = k\sum_{i=1}^{n} a_i \qquad 1.6$$

Exemplo e Aplicação

Observe que o Exemplo 1.1 pode ser representado pela Equação 1.7.

$$\sum_{i=1}^{n}(5i) = 5 * 1 + 5 * 2 + 5 * 3 + 5 * 4 = 5 * (1 + 2 + 3 + 4) = 5\sum_{i=1}^{n}(i) = 50 \qquad 1.7$$

3. Soma de somatórios.

$$\sum_{i=1}^{n}(a_i) + \sum_{i=1}^{n}(b_i) = \sum_{i=1}^{n}(a_i + b_i) \qquad 1.8$$

Exemplo e Aplicação

Considere que a criança construiu 2 torres de copos com uma base de 4 copos. Essa questão pode ser representada pela Equação 1.9.

$$\sum_{i=1}^{4}(i) + \sum_{i=1}^{4}(i) = (1+2+3+4) + (1+2+3+4)$$
$$= (1+1) + (2+2) + (3+3) + (4+4) = \sum_{i=1}^{4}(i+i) \quad 1.9$$

1.2 Produtórios

Considere o produto de todos os números inteiros entre 1 a 5 (Equação 1.10).

$$1*2*3*4*5 \quad 1.10$$

Similar aos somatórios, a representação do produto descrito na Equação 10 é inviável quando se aumenta significativamente o número de termos. Por exemplo representar o produto de todos os números inteiros de 1 a 100. Assim é conveniente utilizar uma simbologia para a representação do produto de um termo. A representação normalmente utilizada é pela letra grega pi (Π) Dessa forma o produtório dos números inteiros até um valor pode ser compactado pela Equação 1.11.

$$\prod_{i=1}^{n}(i) \quad 1.11$$

Nessa notação o número 1 representa o limite inferior do índice, o número n representa o limite superior do índice, a variável i localizada na parte inferior da letra ∏ é o índice e a operação entre parênteses é o termo do produtório.

Nesse cenário, algumas propriedades de produtórios são normalmente utilizadas em operações matemáticas.

1. Produtório de uma constante

$$\prod_{i=1}^{n}(k) = k * k * k * \ldots * k = k^{n} \tag{1.12}$$

Exemplo e Aplicação

$$\prod_{i=1}^{3}(2) = 2 * 2 * 2 = 2^{3} = 8 \tag{1.13}$$

2. Produtório de um termo por uma constante

$$\prod_{i=1}^{n}(k * a_i) = (k * a_1) * \ldots * (k * a_n) = k^{n}(a_1 * a_2 * \ldots * a_n) = k\prod_{i=1}^{n}(a_i) \tag{1.14}$$

Exemplo e Aplicação

$$\prod_{i=1}^{3}(2 * i) = (2 * 1) * (2 * 2) * (2 * 3) = (2 * 2 * 2)(1 * 2 * 3) = 2^{3}\prod_{i=1}^{3}(i) \tag{1.15}$$

3. Produtório de dois termos

$$\prod_{i=1}^{n}(a_i * b_i) = a_1 * b_1 * \ldots * a_n * b_n = (a_1 * \ldots * a_n) * (b_1 * \ldots * b_n) = \prod_{i=1}^{n}(a_i)\prod_{i=1}^{n}(b_i) \tag{1.16}$$

Exemplo e Aplicação

$$\prod_{i=1}^{3}(i)\,(i^2) = (1*1^2)*(2*2^2)*(3*3^2) = (1*2*3)*(1^2*2^2*3^2) = \prod_{i=1}^{3}(i)\prod_{i=1}^{n}(i^2) \quad 1.17$$

1.3 Produtos notáveis

Uma terceira operação que normalmente é requerida nas operações de estatística básica, são os produtos notáveis. Nesse capítulo daremos ênfase nas três operações mais usuais: produtos notáveis da soma, produtos notáveis da diferença e diferença entre quadrados.

1. Produtos notáveis da soma.

$$(a+b)^2 = (a+b)(a+b) = a(a+b) + b(a+b) + aa + ab + ba + bb = a^2 + 2ab + b^2 \quad 1.18$$

Exemplo e Aplicação

$$(2+3)^2 = 2^2 + 2*2*3 + 3^2 \quad 1.19$$

2. Produtos notáveis da diferença.

$$(a-b)^2 = (a-b)(a-b) = a(a-b) - b(a-b) + aa - ab - ba + bb = a^2 - 2ab + b^2 \quad 1.20$$

Exemplo e Aplicação

$$(2-3)^2 = 2^2 - 2 * 2 * 3 + 3^2 \qquad 1.21$$

3. Diferença entre quadrados

$$(a+b)(a-b) = a * (a-b) + b(a-b) = aa - ab + ba - bb = a^2 - b^2 \qquad 1.22$$

Exemplo e Aplicação

$$(2+3)(2-3) = 2^2 - 3^2 \qquad 1.23$$

CAPÍTULO 2

DEFINIÇÕES E APRESENTAÇÃO DE DADOS

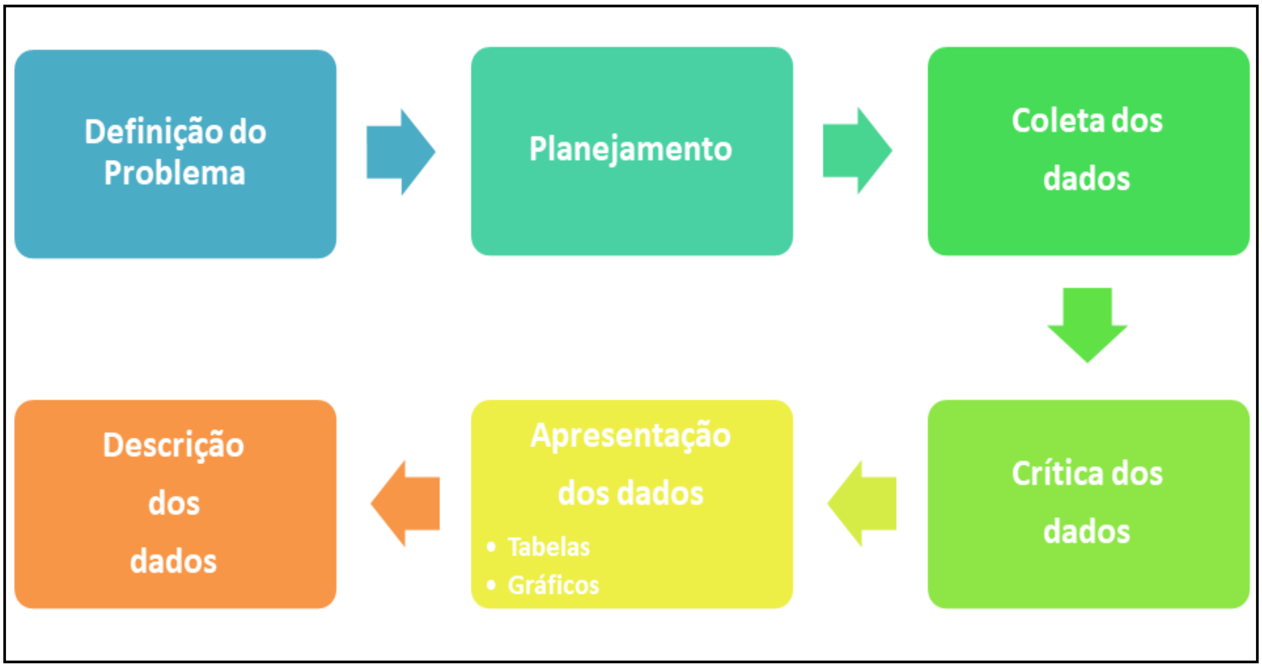
2.1 Definições

Nesse capítulo serão apresentadas algumas definições fundamentais para o entendimento da estatística básica. As definições foram baseadas nos livros Morettin e Bussab (2004); Vieira (2016) e em notas de aulas Bauer (2017).

2.1.1 **Estatística**: Ciência que se preocupa com a organização, descrição, análise e interpretação de dados experimentais. Ela pode ser dividida em duas: estatística indutiva ou inferência estatística que tem por objetivo fazer generalizações sobre uma população com base em dados amostrais; estatística descritiva que busca descrever e analisar certo grupo de dados, sem tirar quaisquer conclusões ou inferências sobre um grupo maior.

2.1.2 **Experimentos:** geram informações para ser analisadas e trabalhadas, em que se busca criar condições livres de qualquer interferência. Os passos para uma experimentação são apresentados na figura 2.1.1.

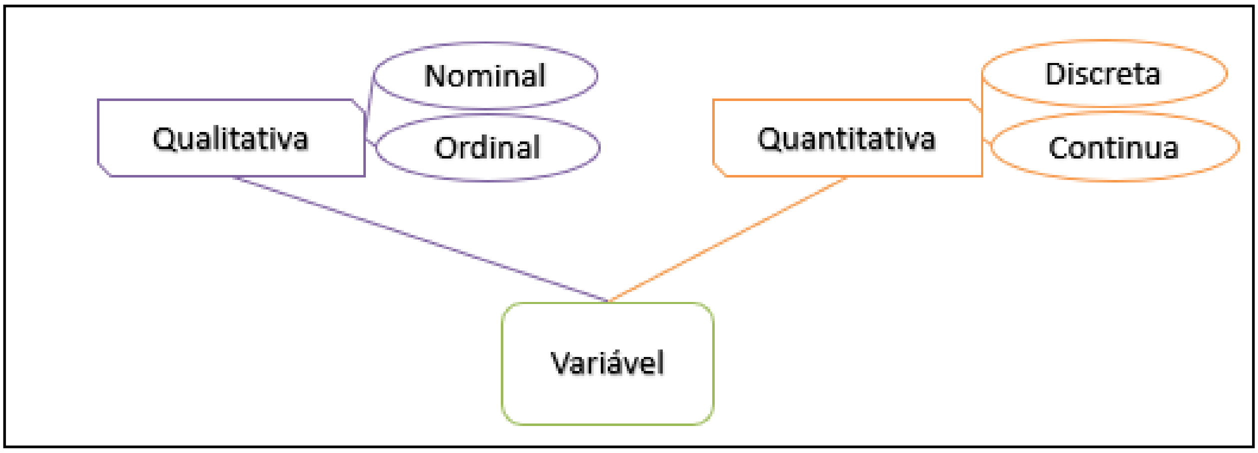
Figura 2.1.1. Passos para a realização de uma experimentação



Fonte: Morettin e Bussab (2004)

2.1.3 **Variáveis:** É a característica medida ou observada em determinado experimento. Classificação das variáveis: Observe a figura 2.1.2. Qualitativa - quando os valores são expressos por atributos: sexo; cor. Quantitativa: quando os valores são expressos em números.

Figura 2.1.2. Classificação das Variáveis.



Fonte: Morettin e Bussab (2004) e Vieira (2016)

2.1.4. **População estatística** e **amostra**: População se refere ao conjunto de portadores de pelo menos uma característica comum. Já amostra é o subconjunto finito de uma população (Figura 2.1.3).

Figura 2.1.3. Ilustração dos conceitos de população e amostra.

Fonte: Morettin e Bussab (2004) e Vieira (2016).

2.1.5 Amostragem: correspondem as técnicas utilizadas para obter dados representativos de uma população a partir de uma amostra.

2.2 Apresentação dos dados

Os dados em estatísticos podem ser trabalhados a fim de organizado-los para facilitar o estudo de algum fenômeno em estudo. Os dados podem ser apresentados de forma tabular e gráfica. A representação gráfica é uma forma muito eficiente de apresentar os dados, que consiste em figura construída a partir de dados tabulados. A apresentação tabular fornece uma inspeção mais precisa dos dados, já os gráficos fornecem um entendimento rápido sobre o mesmo. Como apresentado na Figura 2.2.1, a qual ilustra os dados tabulados e não tabulados.

Figura 2.2.1 Ilustra dados não tabulados e dados tabulados, demonstrado o resumo dos dados

Dados não tabulados

Posição	Empresa	Valor de Mercado em bilhões(US$)
1	Petrobras	83,98
2	Vale	82,03
3	Itaú	52,89
4	Banco Bradesco	43,16
5	Banco do Brasil	20,7
6	JBS	17,48
7	Banco Btg Pactual	56,14
8	Braskem	6,76
9	Gerdau	10,16
10	Suzano	14,66

Dados tabulados

classes	Li	Ls	fi
1	7	33	5
2	33	59	3
3	59	85	2
		Total	10

2.2.1 Apresentação tabular

A apresentação é realizada em forma de um quadro que traz as informações de forma resumida que seguem as normas estabelecidas, como o Conselho Nacional de Estatística (1967) e o IBGE (1993).

A tabela é composta pelos seguintes elementos: *Corpo* que traz as informações da variável em estudo, em suas colunas e linhas; *Cabeçario* é a parte superior da tabela, que informa o conteúdo das colunas; *Título* que deve trazer o maior número de informação sobre a representação, onde deve responde às perguntas (O que?, quando? e onde?).

As tabelas podem apresentar intervalor de classes ou não. Isso é, os dados podem ser divididos em intervalos. A tabela sem intervalo de classe corresponde a representação tabular em que aparece o dado e a frequência que o mesmo aparece, sem haver agrupamento dos mesmos. É interessante trabalhar com esse tipo de tabulação quando se tem interesse em observar cada dado separadamente Tabela 2.2.1. A tabela com intervalo de classe apresenta uma representação tabular em que aparece os dados agrupados em classes. Sua utilização ocorre quando há interesse em se trabalhar com muitos dados, minimizando a quantidade de colunas e facilitar a caracterização dos dados (Tabela 2.2.1).

Para a representação tabular em que se analisa determinada informação, pode-se realizar cálculos a fim de organizar a informação e facilitar a visualização da informação, organizando os dados em classes. Para isso, é necessário a definições de alguns conceitos que integram a tabela, são eles:

a) Intervalo de classe (i): É cada um dos intervalos internos (disjuntos) em que se subdivide o intervalo total dos dados analisados.

b) Número de classe (k): é a raiz do número de dados (n), alguns autores sugerem utilizar no mínimo igual a 4 classes.

$$k = \sqrt{n} \tag{2.1}$$

c) Frequência (f): A quantidade de vezes que um determinado dado se repete da distribuição.

d) Limites da distribuição (LD): São os termos da distribuição que representam: LDs limite superior (o maior valor do dado da distribuição); LDi- Limite Inferior (o menor valor do dado que está na distribuição).

e) Limite de classes: são os termos de cada classe que representam, Ls; limite superior (o maior valor do dado que pode entra na classe); Li- Limite Inferior (o menor valor do dado que está na classe).

f) Frequência absoluta de classe (fi): É a quantidade de dado que pertencem a um determinado intervalo de classe.

g) Frequência acumulada de classe (F_i): É a somas da frequência da classe com a frequência das classes anteriores.

h) Frequência absoluta relativa (fri): É a razão entre a f_i da classe e o número de dados total.

$$fri = \frac{fi}{\Sigma fi} \tag{2.2}$$

i) Frequência absoluta relativa ($\boldsymbol{F}_{ri}$i): É a razão entre a f_{ri} da classe e o número de dados total.

$$F_{ri} = \frac{fi}{n} \tag{2.3}$$

j) Amplitude de classe (hi): É a diferença entre o Li e Ls de cada classe. hi= Ls-Li dividido pelo número de classe; pode ser determinada também utilizando a seguinte equação:

$$h_i = \frac{\text{LDs} - \text{LDi}}{(k-1)} \tag{2.4}$$

k) Amplitude total: É a diferença entre o maior e o menor valor do conjunto de dados em análise.

Exemplo e aplicação

O ranking anual da Forbes Global 2000 apresenta as maiores empresas de capital aberto do mundo, a classificação avalia as organizações baseadas em métricas que englobam vendas, lucro, ativos e o valor de mercado.

As 10 corporações de diversos setores que apresentaram as melhores classificações estão apresentadas na tabela 2.2.1, nela é possível perceber o valor de mercado das corporações, sua posição no Brasil e no mundo.

Tabela 2.2.1. As dez maiores corporações Brasileiras de 2022

Posição	Empresa	Valor de Mercado em bilhões (US$)	Posição mundial
1	Petrobras	83,98	65
2	Vale	82,03	118
3	Itaú	52,89	138
4	Banco Bradesco	43,16	182
5	Banco do Brasil	20,7	268
6	JBS	17,48	401
7	Banco Btg Pactual	56,14	774
8	Braskem	6,76	953
9	Gerdau	10,16	962
10	Suzano	14,66	992

Fonte: Forber, 2022.

Ao se trabalhar os dados buscando observar o quanto cada empresa representa em valor de mercado em relação as demais empresas, foi obtido a tabela de distribuição de frequências (Tabela 2.2.2) baseados no valor de mercado das empresas. Para isso, seguiu-se os seguintes passos:

1. Determinou o número de classe (K), como raiz de 10 é aproximadamente 3. Alguns autores sugerem utilizar no mínimo 4 classes, mas como trata-se de um conjunto de dados pequenos, optou-se por utilizar somente 3.

2. Obteve-se a amplitude de classe,

$$h_i = \frac{83{,}98 \; - \; 6{,}76}{(4-1)} = 25{,}74 \cong 26$$

3. Foi construído as classes e calculado as frequências absolutas (Os calculos estão disponíveis na Planilha Apresentação de dados).

Tabela 2.2.2. Tabela de distribuição de Frequência

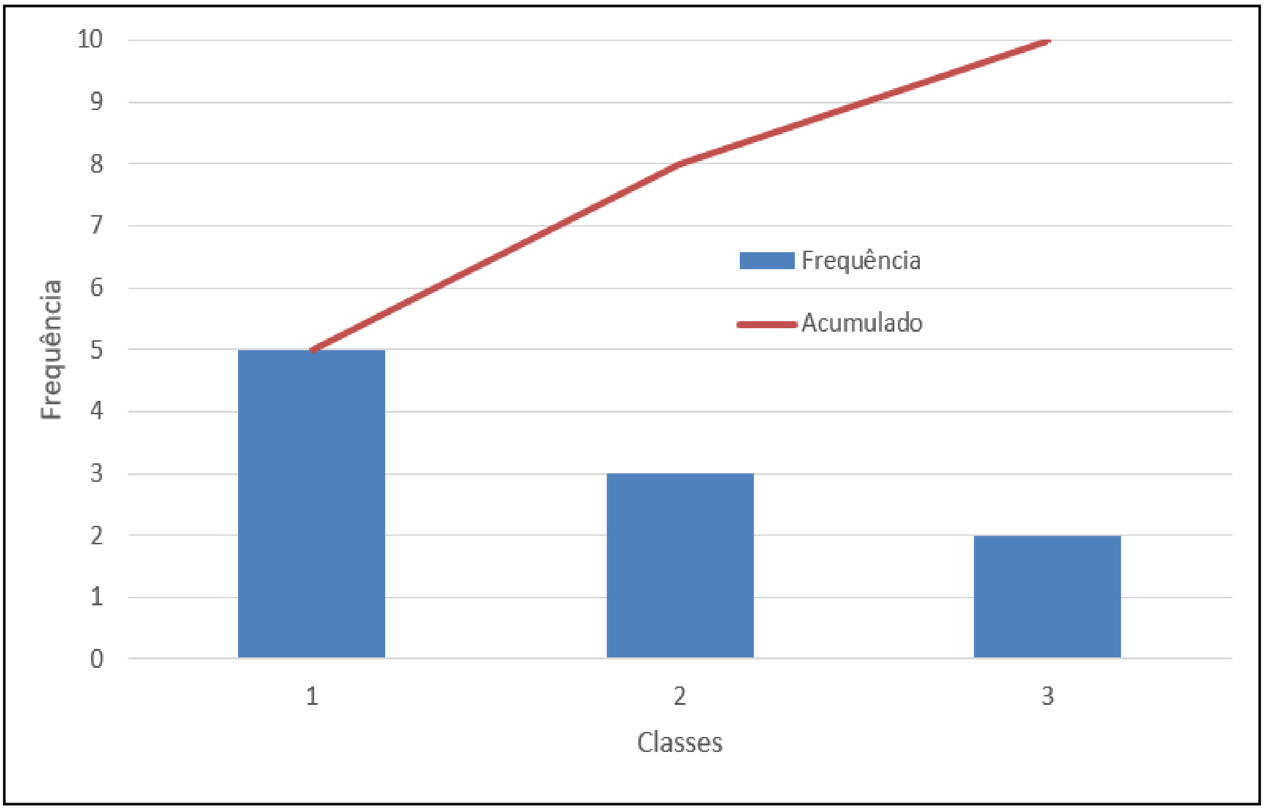
classes	L_i	L_s	f_i	f_{ri}	$f_{ri}(\%)$	F_i	$F_{ci}(\%)$
1	7	33	5	0,5	50	0,5	50
2	33	59	3	0,3	30	0,8	80
3	59	85	2	0,2	20	1	100
		Total	10	1	100		

2.2.2 Apresentação gráfica

A análise gráfica possibilita uma rápida e concisa leitura da variabilidade das variáveis, fornecendo muita informação sobre seu comportamento. Existem vários tipos de gráficos: Gráficos para variáveis qualitativas - barra e de composição em setores (pizza); gráficos para variáveis quantitativas – possuem maior variedade de gráficos, como o histograma e gráfico de dispersão, (MORETTIN; BUSSAB, 2004).

a) **Histograma** – É um gráfico onde a abscissa é representada pelos intervalos de classe e o eixo das ordenadas pela frequência absoluta. Na construção, primeiro deve-se ter uma tabela com representação em intervalos de casses e sua frequência. Tendo estes dados jogue os valores no gráfico onde uma classe inicia no termino da outra e cada classe estará associada a uma frequência. Observe o a figura 2.2.1 obtida a partir da tabela 2.2.1

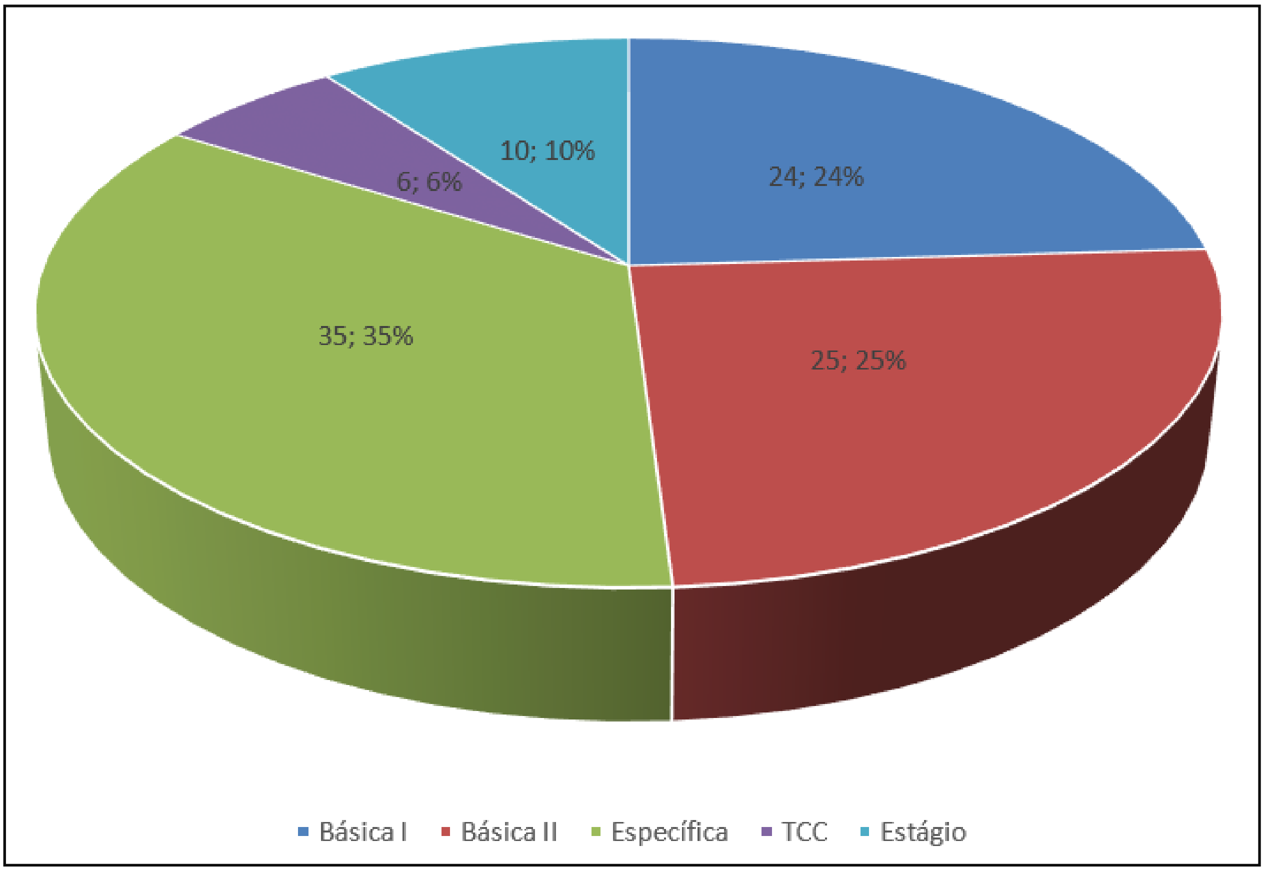
Figura 2.2.1. Classes pelas frequencias e % acumulada da distribuição das empresas a partir da tabulação baseada em seu valor de marcado.



b) **Pizza ou Setores**: É um diagrama circular em que os valores de cada categoria estatística representada são proporcionais às respectivas medidas dos ângulos (1% no gráfico de setor equivale a 3,6º aproximadamente). É muito utilizado para determinar porcentagens.

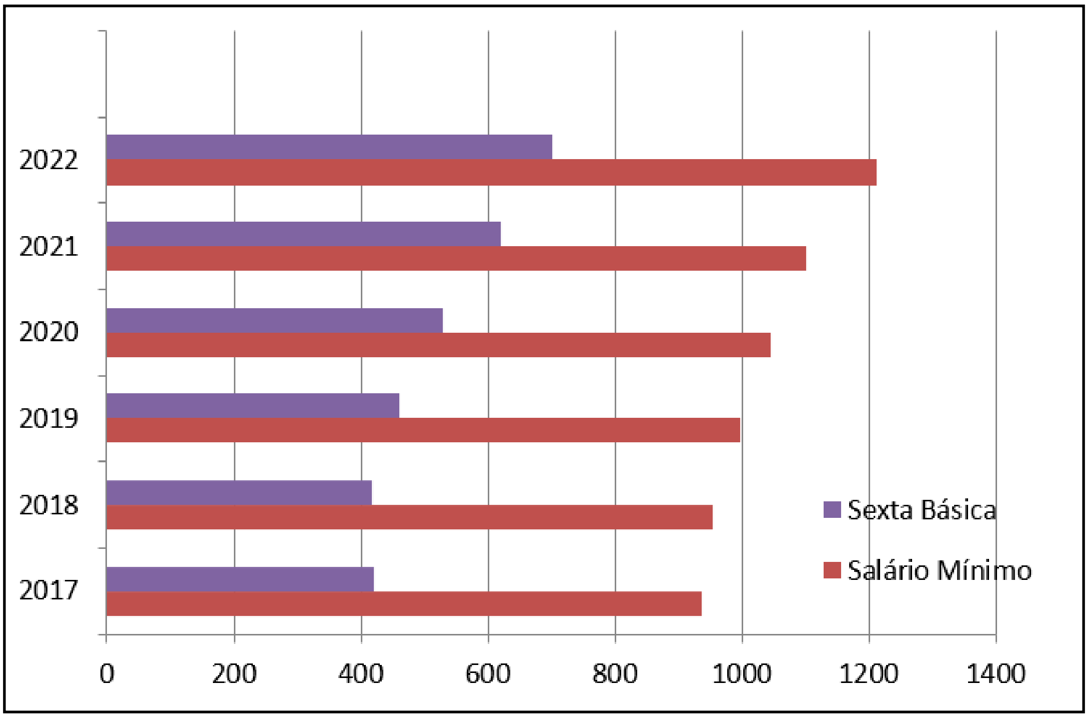
O gráfico a baixo representa a porcentagem de horas aulas distribuídas entre as disciplinas de ciclo básico, específicas, trabalho de conclusão de cursos e estágio de cursos de engenharia.

Figura 2.2.2. Pizza ou Setores de oras aulas distribuídas entre as disciplinas de ciclo básico, específicas, trabalho de conclusão de cursos e estágio de cursos de engenharia.



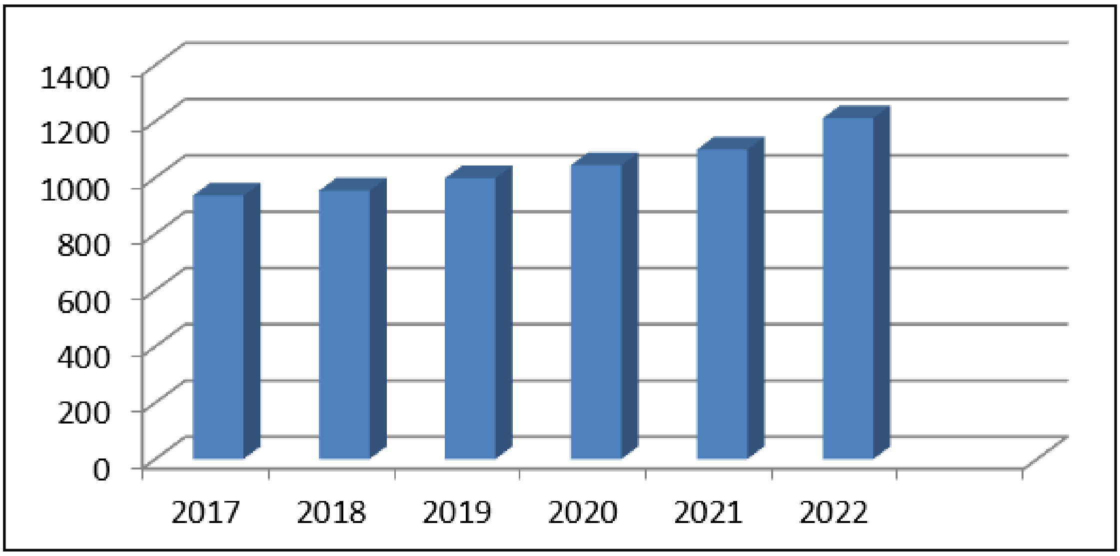
c) **Barra**: O gráfico de barras é composto por dois eixos, um vertical e outro horizontal. No eixo vertical são construídas as barras que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade.

Figura 2.2.3. Gráfico de Barras que apresentam o valor em reais do salário mínimo no Brásil e o valor gasto com a sexta básica.



d) **Colunas**: É composto por dois eixos, um vertical e outro horizontal. No eixo horizontal são construídas as colunas que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade. Essa intensidade é indicada pelo eixo vertical. As colunas devem sempre possuir a mesma largura e a distância entre elas deve ser constante.

Figura 2.2.4 Gráfico de Coluna que apresentam o valor em reais do salário mínimo no Brasil para os últimos 6 anos.



Referências

BAUER, M. O. Estatística Básica: curso introdutório, 2015. 100 f. Notas de Aula.

Conselho nacional de estatística. 1967. Normas de apresentação tabular. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv82497.pdf. Acesso: 05 de novembro de 2022

FORBES. The Global 2000. 2022. Disponível em: https://www.forbes.com/lists/global2000/?sh=32d7a5bc5ac0. Acesso: 05 do 11 de 2022

IBGE. 1993. Normas de apresentação tabular. Disponível em: https://www.btu.unesp.br/Home/sobre/biblioteca/normas_ibge-min.pdf. Acesso: 05 de novembro de 2022

MORETTIN, Pedro A; BUSSAB, Wilton de O. **Estatística Básica**. 5. ed. São Paulo: Saraiva, 2004. 537 p.

VIEIRA, Sonia. **Introdução à bioestatística**. 5. ed. Rio de Janeiro: Elsevier, 2016. 308p.

CAPÍTULO 3

ESTATÍSTICA DESCRITIVA

Em diversas áreas do conhecimento lidamos com situações nas quais dispomos de uma grande quantidade de dados, o que torna difícil absorver completamente a informação. É necessário, portanto, que as informações sejam minimizadas até o ponto em que se possa interpretá-las com maior clareza. Dessa maneira, torna-se indispensável resumi-las, através do uso de certas medidas-sínteses, comumente conhecidas como estatísticas descritivas ou simplesmente estatísticas. Consequentemente, a estatística descritiva é um número que sozinho descreve uma característica de um conjunto de dados. No entanto, ao sintetizar os dados através do uso de medidas descritivas, é evidente que muita informação será perdida e alguns resultados obtidos serão distorcidos.

Em um sentido mais amplo, a Estatística Descritiva pode ser interpretada como um ramo da estatística cujo objetivo básico é o de observar, coletar, organizar e classificar um conjunto de dados, permitindo dessa forma a interpretação de uma determinada variável através de uma amostra.

A descrição dos dados é feita por meio de medidas que representam os dados de forma sumária. As Medidas de Posição são usados para descrever o centro de um conjunto de dados.

Nesse capítulo serão apresentadas algumas definições fundamentais para o entendimento da estatística básica. As definições foram baseadas nos livros

Morettin e Bussab (2004); Vieira (2016) e em notas de aulas Bauer (2017)

Considerando um conjunto de dados com *n* termos $X_1 + X_2 + ... + X_n$ a média aritmética é obtida pela equação 3.1, existem ainda médias geométricas e harmônicas que são pouco utilizadas.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

A Moda (M_o) é o número da distribuição que mais aparece. Para obtê-la basta observar os dados. M_o pode ser classificada de acordo com a quantidade de modas que aparecem, são elas: Amodal (zerol modas); Unimodal (uma Moda); Bimodal (duas modas); Multimodal (Treis a mais modas).

A Mediana é o número que está no meio da amostra estando a mesma em ordem crescente.

As Separatrizes são números que dividem a sequência ordenada de dados em partes que contêm a mesma quantidade de elementos da série, como a mediana, quartil, quintil, decil e percentil. As medianas dividir um conjunto numérico em duas partes iguais. Já o quartil dividir um conjunto em quatro partes iguais. No qual, cada parte representará 25% do total. Considere n o número de dados, obtém-se, portando:

Primeiro Quartil $\rightarrow Q_1 = 0{,}25\ (n+1)$ 3.2

Segundo Quartil $\rightarrow Q_2 = 0{,}50\ (n+1)$ 3.3

Terceiro Quartil $\rightarrow Q_3 = 0{,}75\ (n+1)$ 3.4

O Quintil dividi a série ordenada em cinco partes iguais. Assim, cada parte representa 20%. O Decil divide o conjunto em 10 partes iguais. No qual, cada parte conterá 10% do conjunto. O percentil fornece a divisão do conjunto em

100 partes iguais. Para identificar a medida que se pretende obter com o percentil correspondente, , basta utilizarmos a equação 3.5

$$P_i = \frac{i\,x\,n}{100} \quad 3.5$$

Medidas de Dispersão são medidas que quantificam de algum modo a variabilidade dos dados geralmente utilizando como referência uma medida de posição. E são obtidas pelas equações a seguir.

A amplitude (AT) é a diferença entre o maior e o menor dado da distribuição, possui como vantagem uma rápida obtenção de dispersão, mas é pouco informativa

$$AT = x_{máximo} - x_{mínimo} \quad (3.6)$$

A variância amostral S^2 determina a dispersão dos valores entorno da média. O desvio padrão amostral S é a raiz quadrada da variância, possui como vantagem o fato de estar na mesma unidade que os dados em analise, assim, quanto maior o desvio mais afastado da média os dados estão

O coeficiente de variação $CV(\%)$ é a medida relativa da variabilidade em um conjunto de dados. Seus valores variam de mais a menos infinito, isso é, podem ser negativos e acima de 100. Quanto menor o $CV(\%)$ maior é a concentração dos dados entorno do valor central, assim, maior é a homogeneidade dos dados. O $CV(\%)$ é adimensional e pode ser utilizado para comparar variabilidade de conjuntos de dados com unidades diferentes.

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{(n-1)} \quad (3.7)$$

$$S = \sqrt{S^2} \quad (3.8)$$

$$CV(\%) = \frac{S_x}{\bar{x}} \times 100 \qquad (3.9)$$

O erro padrão da média $S\,\bar{x}$ é uma medida de dispersão das médias amostrais em torno da média da população. Quanto menor o valor $S\,\bar{x}$ mais provável será a chance de se obter a média amostral nas proximidades da populacional.

$$S\,\bar{x} = \frac{S}{\sqrt{n}} \qquad (3.9)$$

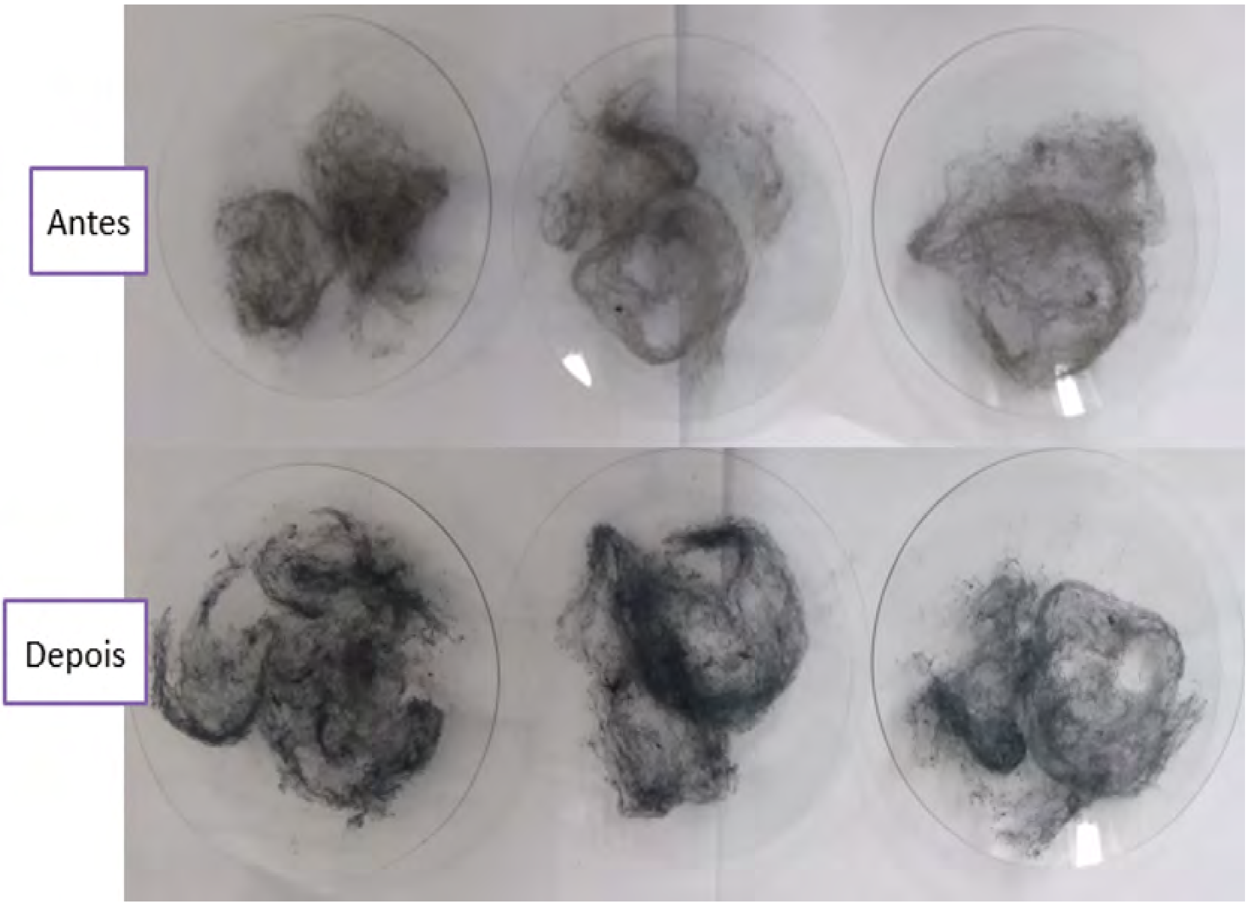
Existem ainda medidas estatísticas das distribuições que são o coeficiente de assimetria e coeficiente de curtose. O coeficiente de assimetria mede a simetria ou a assimetria da distribuição enquanto que a curtose mede o grau de achatamento.

Exemplo e aplicação

As esponjas de aço são produzidas a partir de arames de aço com baixo teor de carbono. Quando queimada, a lã de aço aumenta de massa, pois o ferro aquecido se combina com o oxigênio (SUPER INTERESSANTE, 2016).

A fim de observar o erro experimental para a quantificação do aumento de massa, foi realizado a pesagem em triplicata da massa da lã antes e após a queima. A Figura 2.2.5 apresenta a lãs de aço antes e após a queima, Tabela 3.1.

Figura 3.1. lãs de aço antes e após a queima.



Foi determinado a média, desvio padrão e coeficiente de variação do aumento da massa da lã, tabela 3.2. Com isso, foi possível observa que o aumento de massa foi de 0,28 ± 0,07g. Os resultados das repetições foram próximos as médias e podem ser considerados homogêneos uma vez que coeficiente de variação foi baixo (0,26).

Tabela 3.1 Resultados experimentais da queima de lã de aço e medidas estatísticas de posição e dispersão.

repretição	Massa (g) da Lã	Massa (g) da Lã após queima	Aumento de massa
1	1	1,31	0,31
2	1,2	1,32	0,16
3	0,98	1,32	0,34

Tabela 3.2 Resultados do da queima de lã de aço e medidas estatísticas de posição e dispersão

Média	**0,28**
Desvio-padrão amostral	**0,07**
Coeficiente de variação (%)	**26,0**

Referências

BAUER, M. O. Estatística Básica: curso introdutório, 2015. 100 f. Notas de Aula.

MORETTIN, Pedro A; BUSSAB, Wilton de O. **Estatística Básica**. 5. ed. São Paulo: Saraiva, 2004. 537 p.

SUPER INTERESSANTE. **Se a palha de aço é de metal, por que pega fojo?**. 2016. Disponível em: https://super.abril.com.br/comportamento/se-a-palha-de--aco-e-de-metal-por-que-pega-fojo/. Acesso 05 de novembro de 2022.

VIEIRA, Sonia. **Introdução à bioestatística**. 5. ed. Rio de Janeiro: Elsevier, 2016. 308p.

CAPÍTULO 4

ESTATÍSTICA INFERENCIAL

A estatística indutiva ou inferência estatística busca fazer generalizações sobre uma população com base em dados amostrais. Para isso, são utilizados testes de hipóteses. Um teste de hipótese é o procedimento padrão que testa uma afirmativa a respeito da população. Esses testes fornecem metodologias que permitem verificar se os dados amostrais trazem evidências que sustentam ou não uma afirmação sobre a população (usualmente denominados parâmetros) (MORETTIN; BUSSAB, 2004).

Para a realização de testes, se faz necessário definir duas hipóteses: Hipótese de nulidade (H_o) é a hipótese a ser testada, ou seja, uma afirmação de igualdade sobre o parâmetro de interesse; e hipótese alternativa (Ha) que é a hipótese que contraria H_0. (CECON *et al.*, 2012)

Qualquer decisão tomada em relação às hipóteses formulada, há probabilidade de ocorrer erros. Os erros são definidos como: Erro tipo I caracterizado por rejeitar H_0 rejeitar H_0 dado que ela é verdadeira. Sua probabilidade é representada por α (nível de significância do teste).

O erro tipo II é caracterizado por não rejeitar H_0 quando H_0 é falsa. A probabilidade de ocorrência deste erro é representada por β (Cecon *et al.*, 2012)

$$\alpha = P(\text{erro tipo I}) = P(\text{rejeitar } H_0/H_0 \text{ verdadeira}) \quad (4.1)$$

$$\beta = P(\text{erro tipo II}) = P(\text{não rejeitar } H_0/H_0 \text{ falsa}) \quad (4.2)$$

O nível de confiança para um intervalo de confiança é definido como a probabilidade 1- α. Sendo α de 0.05 o mais comum (TRIOLA, 2008).

O poder do teste (Valor P) é a probabilidade de rejeitar H_0 quando esta é falsa (CECON *et al.*, 2012).

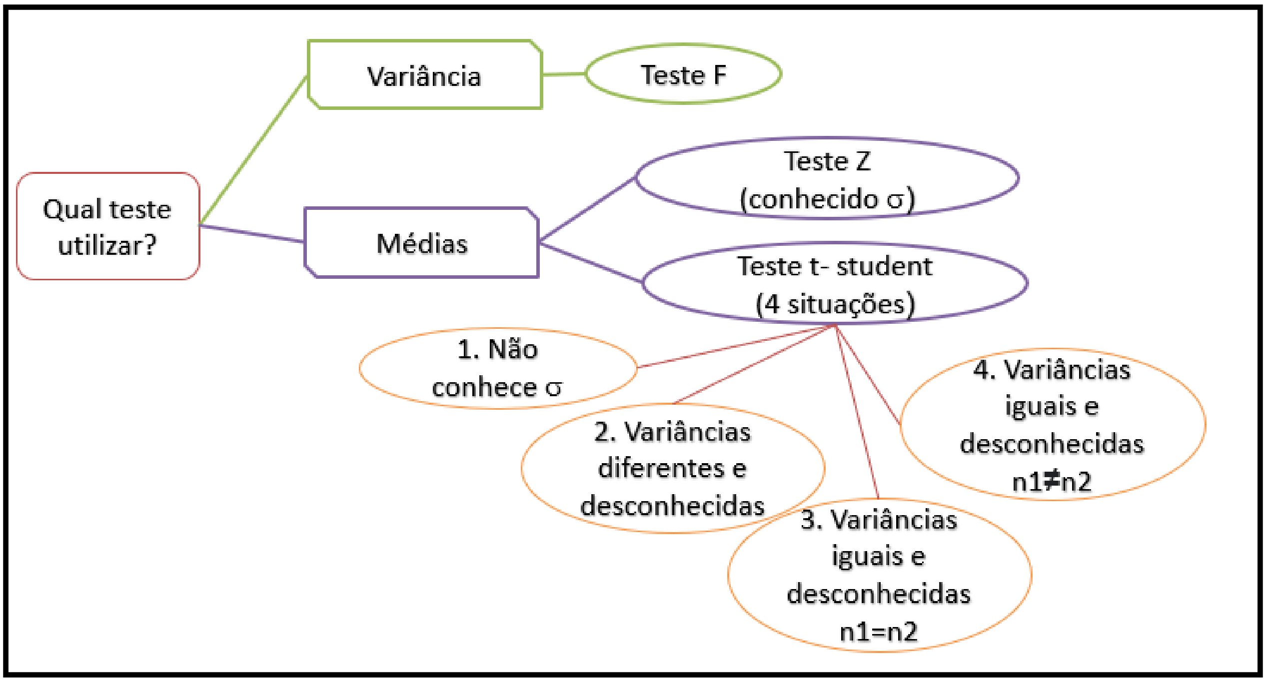
$$\text{Poder} = 1 - \beta \quad (4.3)$$

A estatística do teste é o valor usado para aceitar ou rejeitar a hipótese de nulidade, é encontrado pela conversão estatística amostral (como média amostral, desvio amostra, proporção) em um escore (como z, t, qui-quadrado) (TRIOLA, 2008).

A escolha do método é algo que deve ser realizado cuidadosamente e exige que se analise cada todas as informações que se possui sobre o conjunto de dados em análise.

A figura 4.1 ilustra os testes, quando busca-se comprar duas variâncias ou médias. O teste F verifica se existe diferença estatística significativa entre a variância de dois conjuntos de dados. Já a comparação entre as médias é realizado utilizando teste Z ou teste t-student. O teste Z é utilizado quando se compara as médias entre uma população e amostra, em que se conhece o desvio-padrão populacional (σ). Já para o teste t-student existem 4 situações para verificar se existe diferença significativa entre as médias em análises. No primeiro caso, busca-se comparar as médias entre uma população e amostra, em que não se conhece (σ). No segundo caso, se compara a média de duas amostragens com variâncias diferentes e desconhecidas. Já o terceiro e o quarto, se compara a média de duas amostragens com variâncias iguais e desconhecidas, sendo que na primeira os números de dados amostrais são iguais (n1=n2), e a segunda, diferentes (n1≠ n2). A fundação Venzolini (2017) reúne em uma tabela os testes de comparação, indicando qual teste usar a partir dos parâmetros populacionais conhecidos. Disponível do drive intitulada por síntese, pertence ao curso gratuito de estatística e probabilidade do professor Zancul (2017).

Figura 4.1 – Ilustração dos testes estatísticos a serem utilizados baseados no conhecimento dos dados em análises.



Em cursos de estatística básica ou bioestatística, os cálculos são realizados, inicialmente, sem auxilio de recursos computacionais a fim de fazer com que o aluno se familiarize com a utilização das fórmulas e manuseio de tabelas. Mas, atualmente, os cálculos são realizados por pacotes estatísticos, como Minitab, Splus, R e outros (MORETTIN; BUSSAB, 2004; RABELLO; PIMENTEL, 2017). Esses programas fornecem a probabilidade do valor da estatística do teste ser, na distribuição teórica, maior que o valor obtido. Essa probabilidade é conhecida como p-valor. Onde ser rejeita a H_0 toda vez que o p-valor for menor do que o nível de significância (VIEIRA, 2006).

A análise de Variância (ANOVA) é um método utilizado para testar a igualdade de três ou mais médias populacionais por meio da análise de variância (TRIOLA, 2008). Sendo, portanto, uma extensão do teste t-sdudent que compara um ou duas médias. A ANOVA permite que o pesquisador compare qualquer número de médias. A tabela 4.1 apresenta um conjunto de dados genéricos para um delineamento inteiramente casualizado. A partir da tabela 4.1 utilizando as equações é possível obter a ANOVA. Em que gl é grau de liberdade; n é número de dados da amostra r é repetição; SQt é a soma do quadrado total (4.1) SQt_r é a soma do quadrado do tratamento (4.2); SQ_R soma do quadrado do resíduo (4.3). Sendo C, conhecido como correção, calculado pela equação 4.4.

Tabela 4.1. Conjunto de dados genéricos para um delineamento inteiramente casualizado.

	Tratamento				**Total**
	1	2	3	k	
	...	...	...	...	
	...	...	...	...	
Total	T_1	T_2	T_3	Tk	$\sum Y$
n. de repetições	r	r	r		$n = K \times r$

Tabela 4.2. ANOVA genérica para DIC

Causa de Variação	gl	SQ	QM	F
Tratamento	(k-1)	SQt_r	QMt_r	$F=\frac{QMtr}{QMR}$
Resíduo	(n-k)	SQ_R	QM_R	
Total	n-1			

$$SQT = \sum Y^2 - C \qquad 4.1$$

$$SQTr = \sum T^2 - C \qquad 4.2$$

$$SQR = SQT - SQTr \qquad 4.3$$

$$c = \frac{(\sum Y)^2}{n} \qquad 4.4$$

O QMTr corresponde ao quadrado médio do tratamento (4.5); QMR é o quadrado médio do resíduo (4.6). E F corresponde a estatística do teste calculado, que é, portanto, encontrado pela conversão estatística amostral, isso, obtidos a partir dos dados amostrais, também denominado $F_{calulado}$. O $F_{crítico}$ é obtido a partir da tabela de distribuição F.

$$QMTr = \frac{SQTr}{k-1} \qquad 4.5$$

$$QMR = \frac{SQR}{k-1} \qquad 4.6$$

$$F = \frac{QMTr}{QMR} \qquad 4.7$$

Assim, o F corresponde a estatística do teste, que consiste em um valor usado para se tomar decisões sobre a hipótese nula ao ser comparado com o valor tabelado do teste. Caso o valor de $F_{calulado} > F_{crítico}$ pode-se afirmar que aumenos um dos tratamentos difere dos demais. Logo, para determinar essa diferença é necessário realizar a comparação de médias.

Para fazer uma análise de variância é preciso pressupor que: os erros são variáveis aleatórias independentes; existe homocedasticidade (variâncias são constantes); seguem uma distribuição normal. A análise de variância é apenas o primeiro passo na análise de dados, o passo seguinte consiste no exame das médias e das diferenças entre elas (VIEIRA, 2006).

Desta forma, a estatística é uma ferramenta para análise de dados e experimentação que conta com diversas técnicas, e que acompanha o crescente desenvolvimento cientifico tecnológico.

Exemplo e aplicação

Para analisar o desempenho computacional de ambientes de desenvolvimento integrado (IDE), Rabello e Pimentel (2017) e Pimentel, Rabello e Poubel (2021), apresentaram a tabela 4.3 com as medias descritiva dos tempos computacionais, mas a fim de verificar se havia ou não diferença estatística significativa, realizaram testes de hipótese (teste F e t).

Tabela 4.3. Medias descritiva do tempo computacionais.

	Média	Desvio-Padrão	CV (%)
Geany	2015,0539	39,57758	1,9609
Dev C/C++ 5.11	2443,7233	82,75372	3,38637

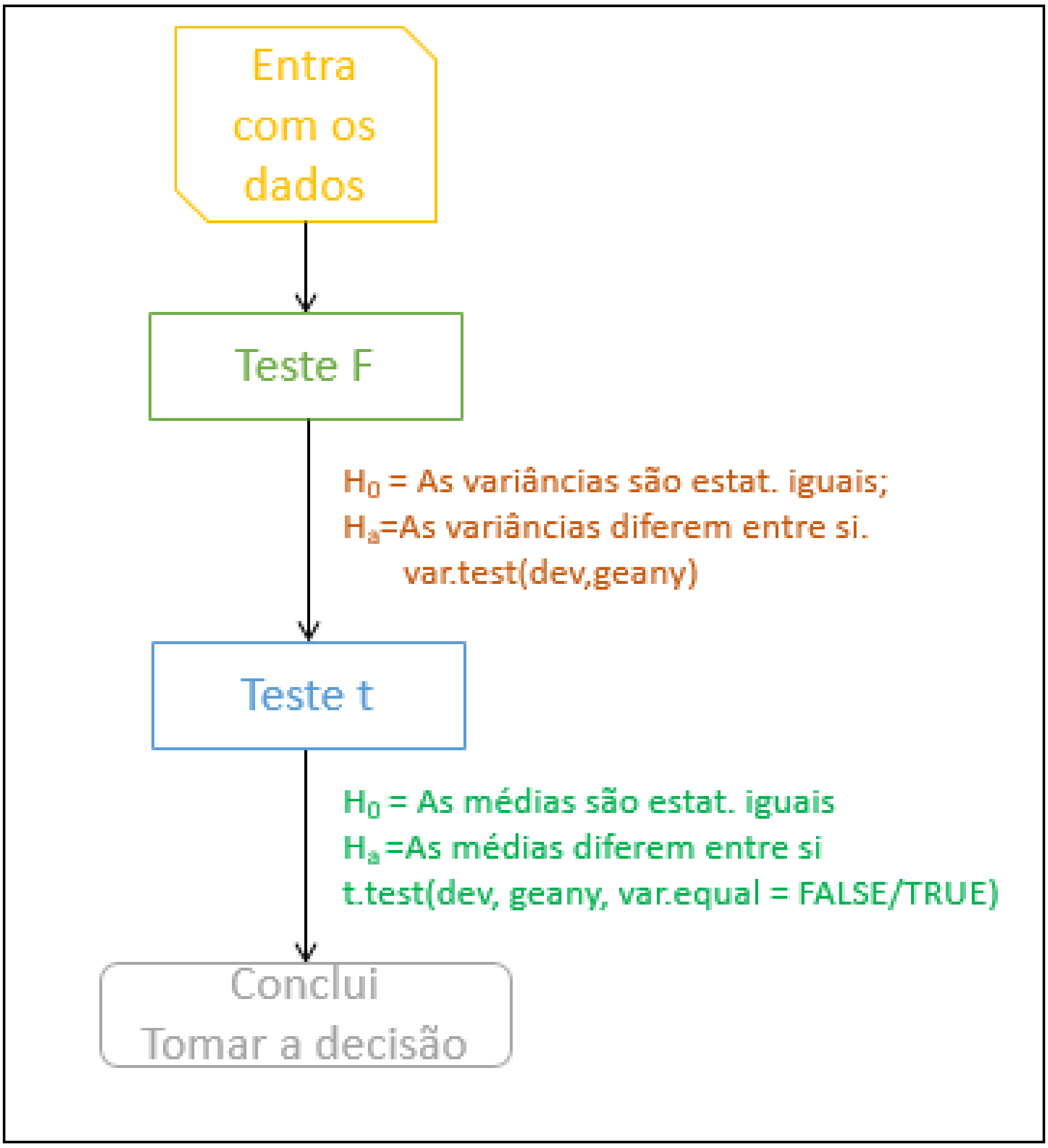
Fonte: Rabello e Pimentel (2017)

As medidas descritivas dos dados, representados na tabela 4.3 indicam que o maior tempo computacional é o Dev-C++, e que esse apresenta maior variabilidades dos dados, pois possui maior coeficiente de variação, já que quanto menor o valor de CV mais homogêneo serão os dados. As diferenças entre as médias foram de 428.6694s.

Foi realizado, também, teste estatístico de comparação entre médias de acordo com a ilustração apresentada na Figura 4.2. E assim, foram obtidos os valores de p: para F, p-valor foi de 0.002317; e para t p-valor foi menor que 2.2e-16. Sabendo que para p-valor menor que o nível de significância assumido (α=0.05) rejeita-se H_0. E assim, pode-se concluir que: as variâncias entre os tempos computacionais diferem estatisticamente entre si; e que as médias são estatisticamente diferentes, sendo a médias do Dev-C++ maior que a médias da Geany.

Assim, IDE Geany apresenta um menor custo computacional.

Figura 4.2 Comparação estatística entre duas médias.



Fonte: Rabello e Pimentel (2017).

Referências

CECON, Paulo Roberto et al. **Métodos Estatísticos**: Série Didática. Viçosa: Ufv, 2012. 229 p.

VANZOLINI. Curso de estatística e probabilidade. 2017. Disponivel em: https://vanzolini.org.br/educacao/dados/data-analytics/ . Acesso: 05 de maio de 2017

NORMANDO, David; TJÄDERHANE, Leo; QUINTÃO, Cátia Cardoso Abdo. A escolha do teste estatístico – um tutorial em forma de apresentação em PowerPoint. **Dental Press J. Orthod**, Rio de Janeiro, v. 15, p.101-106, jan. 2010. Disponível em: <https://www.dentalpress.com.br/portal/bioestatistica/>. Acesso em: 22 nov. 2017.

Rabello, G. L.; Pimentel, U. F. (2017). Modelagem e Simulação de um secador de leito fixo para grãos de café utilizando programação C/C++. Trabalho de conclusão de curso. Engenharia Química. Universidade Federal do Espírito Santo, Alegre.

PIMENTEL, U. F.; RABELLO, G. L; POUBEL, W. M. Desempenho computacional de ambientes de desenvolvimento integrado livre (IDE) aplicados ao processo de secagem. Anais do Congresso Online Nacional de Química.2021.

TRIOLA, Mario F. **Introdução à Estatística**. 10. ed. Rio de Janeiro: Ltc, 2008. 700 p.
TSTUDIO. **Take control of your R code**. Disponível em: <https://www.rstudio.com/products/rstudio/>. Acesso em: 24 nov. 2017.

VIEIRA, Sonia. **Análise de Variância**: ANOVA. São Paulo: Atlas, 2006. 2004 p.

CAPÍTULO 5

INTRODUÇÃO AO PLANEJAMENTO EXPERIMENTAL

O planejamento Experimental é uma técnica que possibilita ao pesquisador identificar variáveis que apresentem influência em determinados processos (CALADO; MONTGOMERY, 2003). Alguns exemplos: avaliação do efeito de adubos no crescimento de plantas; efeito do uso de catalizadores em reações e outros. O delineamento experimental é a forma como os tratamentos (níveis de um fator ou combinações de níveis de fatores) são alocadas as unidades experimentais.

Os delineamentos experimentais envolvem um ou mais fatores, cada fator com n níveis. Os principais delineamentos são: DIC (delineamento inteiramente ao acaso); DBC (delineamento inteiramente casualizados) e DQL (delineamento em quadrado latino). Em todos os delineamentos é obrigatório a utilização dos princípios básicos de experimentação que são: Casualização e Repetição. A casualização consiste em distribuir ao acaso os tratamentos as unidades experimentais, buscando obter uma estimativa válida do erro experimental, ao utilizar a casualização os tratamentos apresentaram as mesmas probabilidades de serem alocados a qualquer unidade experimental. Já a repetição consiste em aplicar o mesmo tratamento a várias unidades experimentais, portanto, consiste na reprodução do experimento base, com a repetição é possível estimar o erro experimental (VIEIRA, 2006).

Outro princípio básico de experimentação é o controle local, ele consiste em aplicar aos tratamentos blocos visando minimizar o efeito da variação no material experimentais ou nas condições para dividir condições experimentais heterogenias e tornar o delineamento mais eficiente (VIEIRA, 2006).

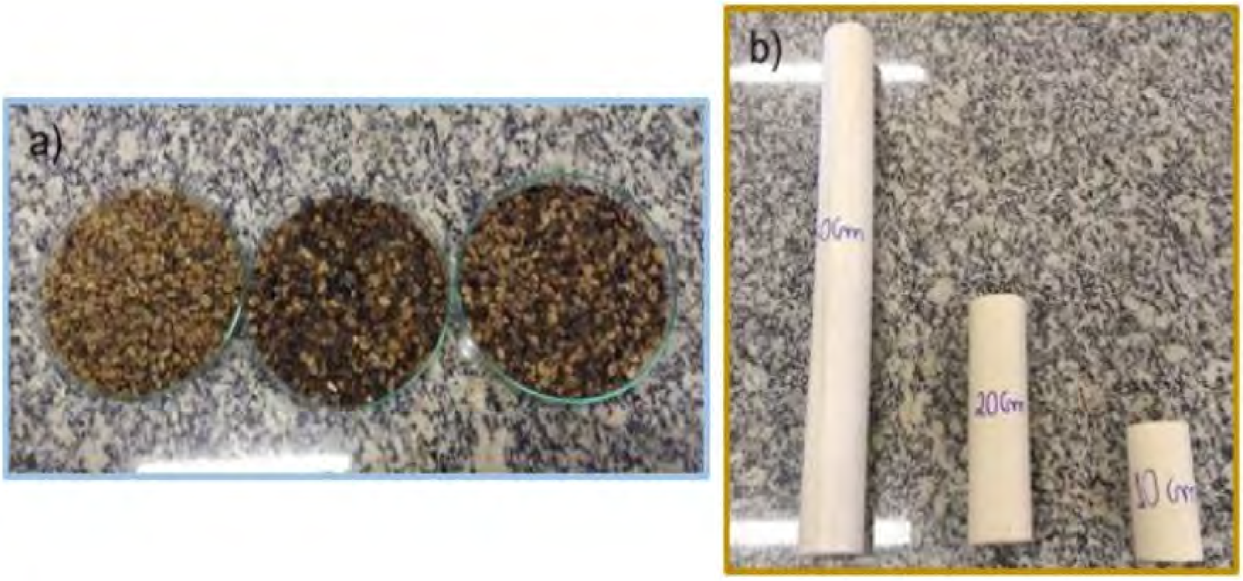
Os experimentos fatoriais não constituem um delineamento, ele é definido como um arranjo dos tratamentos os quais serão alocados na parcela de acordo com o delineamento, como apresentado anteriormente. Os fatores correspondem a tipos distintos de condições ou características que são manipuladas nas unidades experimentais, cuja influência sobre a variável está sendo investigada.

Os tipos de planejamentos fatoriais são: Planejamento fatorial completo; Planejamento fatorial fracionário; planejamento experimental saturado; planejamento experimental composto central e planejamento experimental de misturas. A seguir será apresentado um exemplo ilustrativo da utilização de um planejamento fatorial.

Exemplo e aplicação

Rabello e Pimentel (2017) e Pimentel, Rabello e Poubel (2018) realizaram uma análise do efeito da altura e composição do leito do café sob o parâmetro porosidade do leito. Para isso, conduziram um planejamento fatorial de 2 fatores. A Figura 5.1 apresenta o leito utilizado e sua altura.

Figura 5.1 Meio do leito poroso. Café seco a esquerda; Palha de café seca ao centro; Mistura de café seco com palha de café a direita; b) Canos de PVC de 50mm a alturas de 50, 20 e 10 cm. Fonte: Pimentel, Rabello e Poubel (2018)



Para a análise estatística dos resultados, eles fixaram α=0,05. Uma análise de variância foi realizada a fim de comprovar estatisticamente se os dados apresentavam diferenças significativas entre os tratamentos (TRIOLA, 2008).

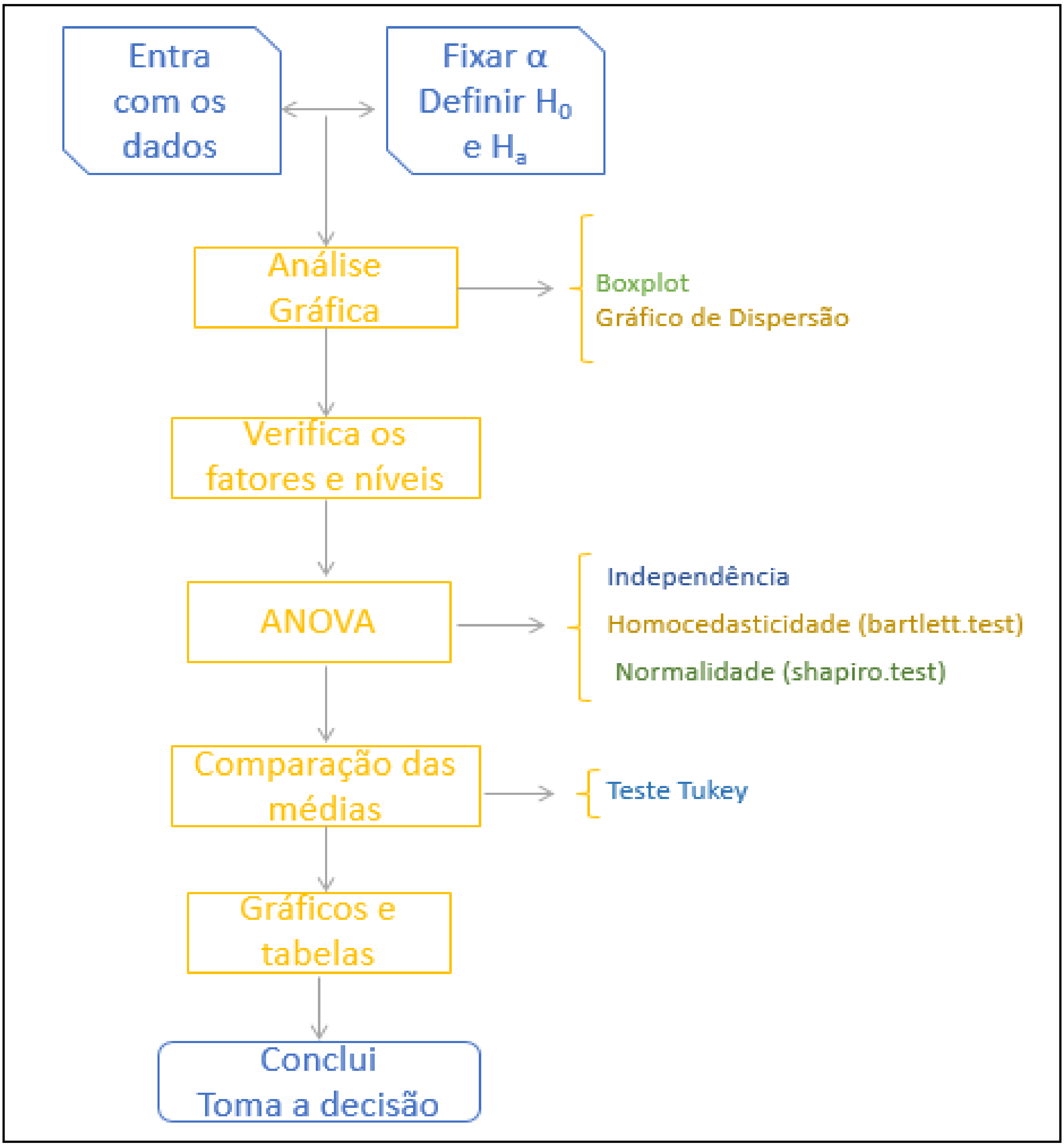
A análise de variância baseou-se no modelo estatístico apresentado na equação 5.1.

$$Y_{ijk} = \mu + \tau_i + \nu_j + \delta_{ij} + \sigma_{ijk} \tag{5.1}$$

Sendo: Y_{ijk} = Observação tomada na i-ésima variedade. J-ésimo teor de água e k-ésima repetição;μ = média da observação, constante do modelo; τ_i= efeito da i-ésima meio; ν_j= efeito da j-ésima altura; δ_{ij}=efeito da interação do i-ésimo meio a j-ésima altura; σ_{ijk}=erro experimental associado à observação de Yijk Porosidade.

Para análise estatística dos resultados foi seguido a metodologia apresentada na figura 5.2. Na qual pode-se observar que consta de uma análise completa, envolvendo tabulação dos dados, apresentação gráfica e testes de hipóteses.

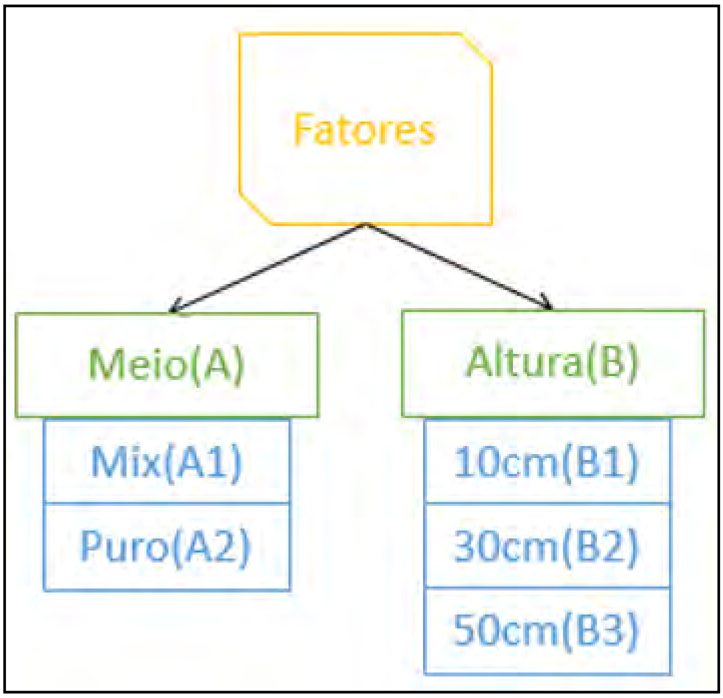
Figura 5.2 Metodologia para análise de variância e comparação de médias do esquema fatorial. Fonte: Rabello e Pimentel (2017)



Fonte: Rabello e Pimentel (2017)

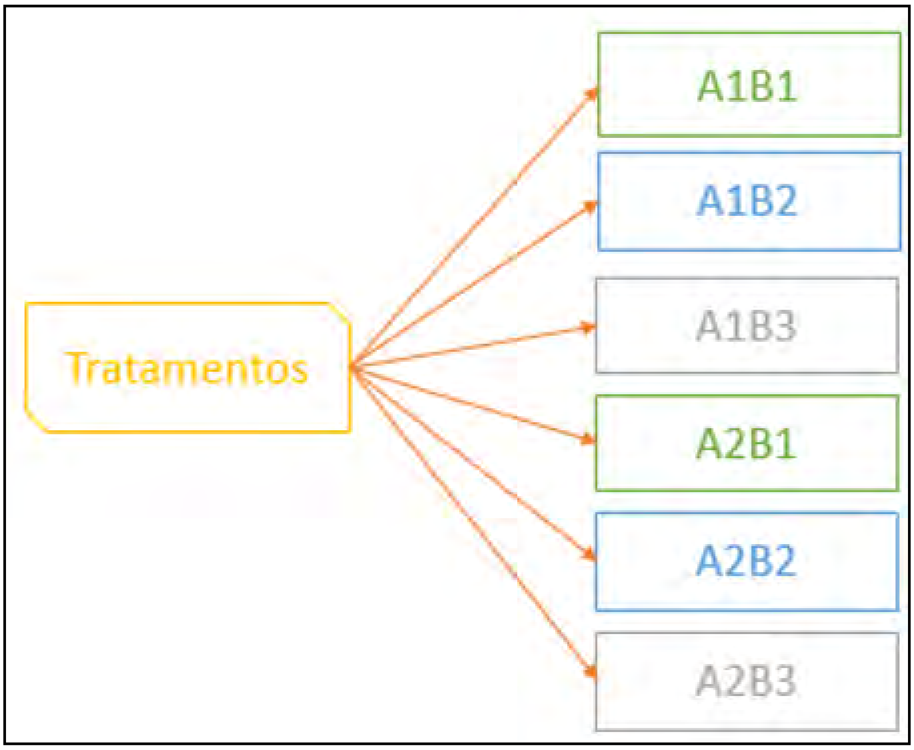
A figura 5.3 apresenta o esquema fatorial. Os fatores estão em verde e níveis em azul. Já a figura 5.4 ilustra os tratamentos e a figura 5.5 ilustra as hipóteses.

Figura 5.3. Esquema fatorial com 2 fatores, 2 níveis para o leito (meio) e 3 níveis para altura.



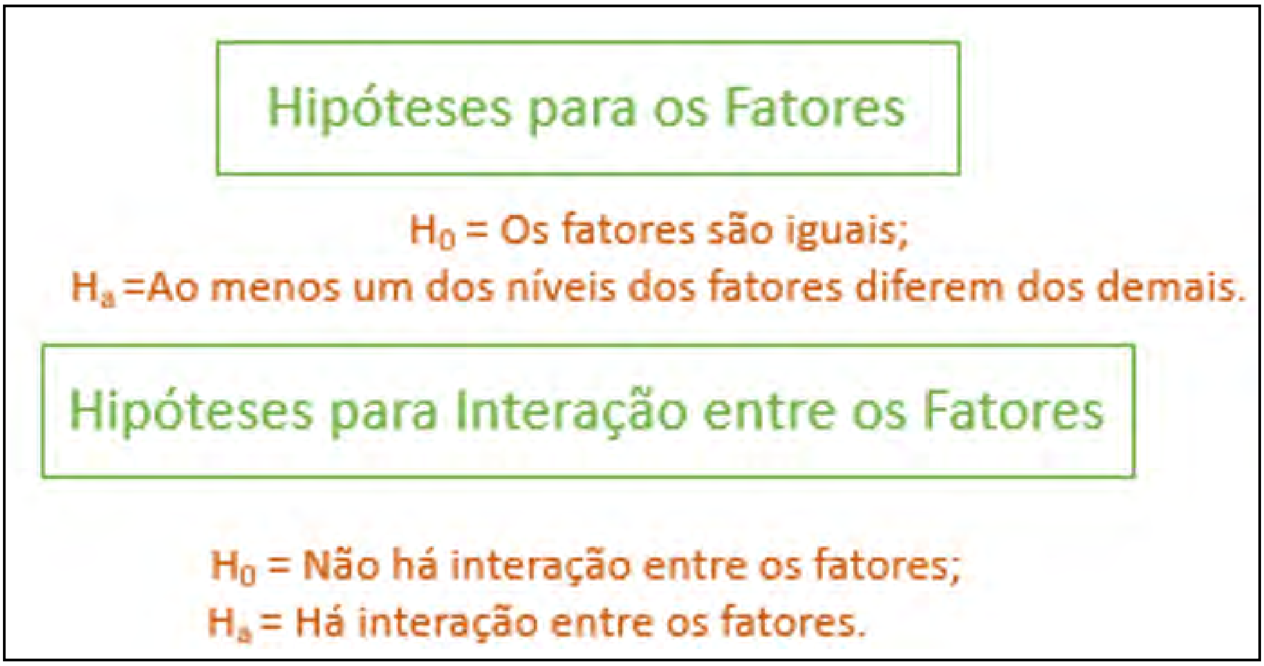
Fonte: Rabello e Pimentel (2017)

Figura 5.4. Os tratamentos.



Fonte: Rabello e Pimentel (2017)

Figura 5.5. Hipóteses.



Fonte: Rabello e Pimentel (2017)

Os valores de porosidade calculados para cada tratamento pela repetição correspondente encontram-se tabela 5.1. A tabela 5.2 foi obtida a partir da tabela 5.1, sendo está uma tabela auxiliar para o cálculo das as médias de cada nível (tabela 5.3).

Tabela 5.1 Resultado Experimental para determinação da porosidade.

	Tratamento					
Repetição	**A1B1**	**A1B2**	**A1B3**	**A2B1**	**A2B2**	**A2B3**
1	0,4094	0,4129	0,4202	0,5720	0,5549	0,5320
2	0,4197	0,4179	0,4045	0,5501	0,5674	0,5288
3	0,4224	0,4103	0,4033	0,5494	0,5532	0,5341

Fonte: Rabello e Pimentel (2017)

Tabela 5.2: Totais para tratamentos.

	B1	**B2**	**B3**	**Totais**
A1	1,2515	1,2412	1,2279	3,7207
A2	1,6715	1,6755	1,5950	4,9419
Totais	2,9230	2,9167	2,8229	8,6626

Fonte: Rabello e Pimentel (2017)

Tabela 5.3: Médias dos níveis

Nível	**Média**
mA1	0,413406
mA2	0,549106
mB1	0,487174
mB2	0,486109
mB3	0,470484

Fonte: Rabello e Pimentel (2017)

Cabe destacar que essa forma de organização de dados auxilia na realização dos cálculos de ANOVA, mas pode ser dispensada na maioria dos softwares estatísticos atuais.

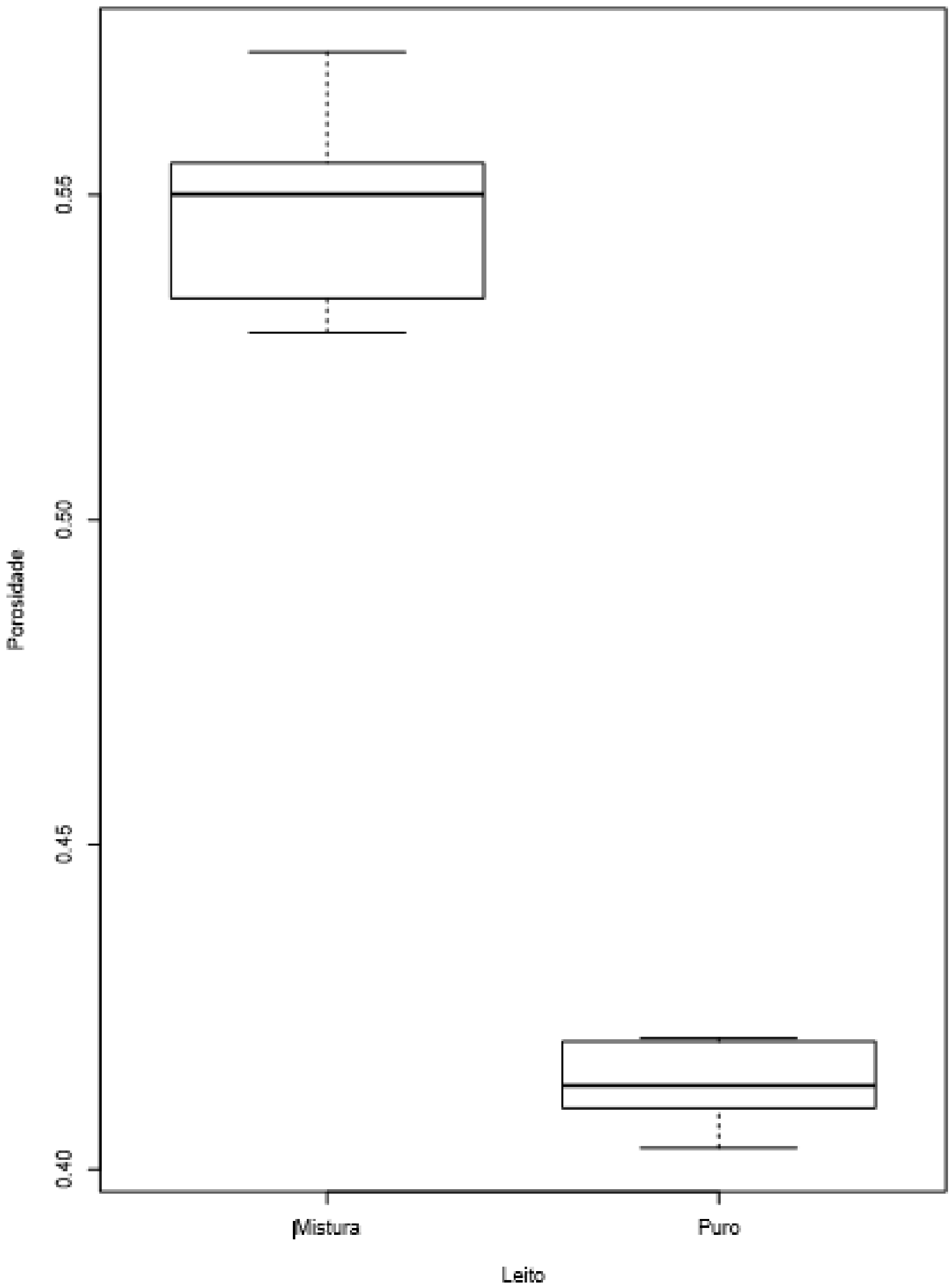
Para a análise, os autores utilizaram os gráficos bloxplot e gráfico de dispersão combinado. As figuras 5.6 e 5.7 representados gráficos. Os gráficos

Boxplot permitem visualizar a distribuição e posição dos dados (MORETTIN; BUSSAB, 2004).

Na figura 5.6 foi identificado a diferença entre o valor da porosidade para cada leito (Mistura e Puro) e o comportamento de posição dos dados. Para a mistura os autores perceberam (e é possível visualizar no gráfico) que os dados estão mais distribuídos devido ao comprimento da cauda dos gráficos. Em ambas as figuras, percebe-se que não há pontos foras (outliers). Já a Figura 5.7 permite visualizar o comportamento dos dados para cada nível do fator altura. Na qual se identificou pequena variação nos dados devido ao tamanho das caudas.

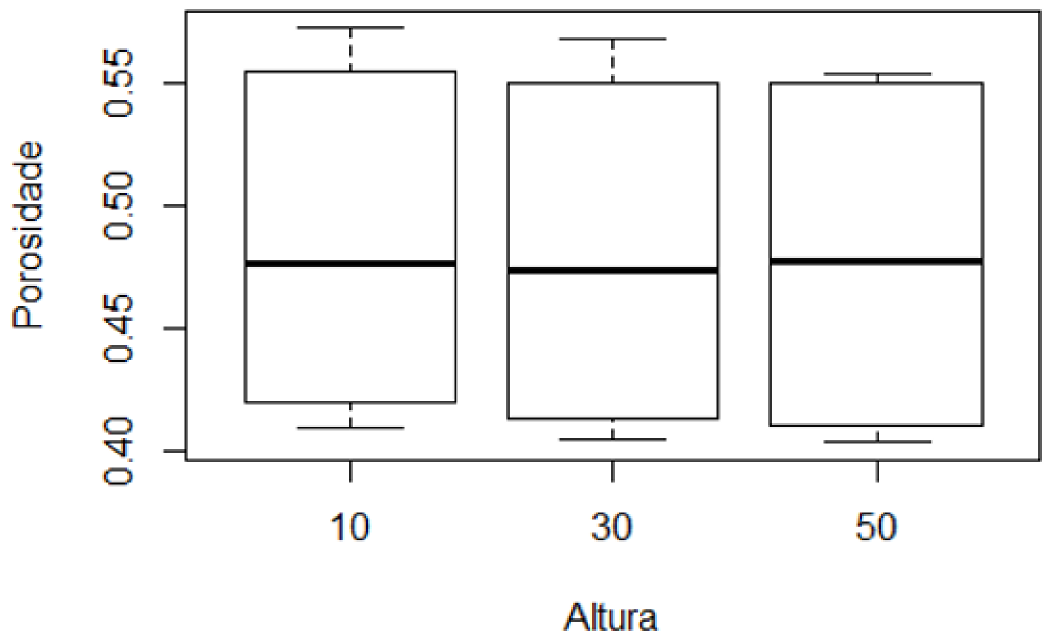
Na figura 5.8 foi observado o comportamento dos dados nos dois meios em diferente altura, no qual fez com que percebessem que os resultados obtidos de porosidade em cada repetição foram bem próximos, isso indica que erro experimental (o que corresponde a execução do experimento) foi pequeno. Além disso, foi possível verificar que os valores da porosidade diferem para os meios, o que não ocorre de forma tão expressiva para as alturas.

Figura 5.6 Bloxplot dos dados experimentais da porosidade em relação ao tipo de leito.



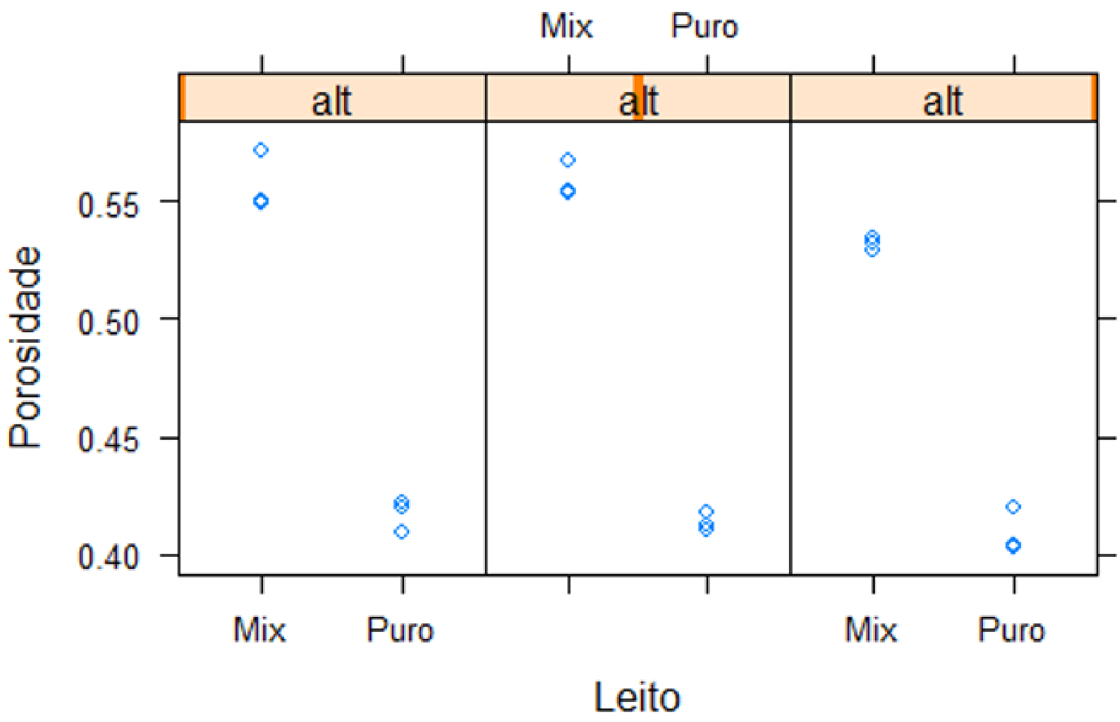
Fonte: Rabello e Pimentel (2017)

Figura 5.7 Bloxplot dos dados experimentais da porosidade em relação ao tipo de leito.



Fonte: Rabello e Pimentel (2017)

Figura 5.8: Gráfico de dispersão combinado dos valores da porosidade.



Fonte Rabello e Pimentel (2017)

Assim, análise visual realizada não garante que os fatores apresentam diferença estatística significativa, eles apenas fomentam tendências. A análise de Variância (ANOVA) é um método utilizado para testar a igualdade de três ou mais médias populacionais por meio da análise de variância (TRIOLA, 2008). Para tal análise, considerou que houve independência entre as amostras, ou seja, um valor de porosidade obtido não foi influenciado pela anterior ou pela próxima. Esse pressuposto garante que os dados foram coletados aleatoriamente dentro do espaço amostral (PIMENTEL; RABELLO; POUBEL, 2018).

A partir do valor de p obtido no teste de Bartlett e Shapiro-Wilk, $p > 0.05$, foi possível assumir que os resíduos se distribuem normalmente, e a variância dos erros são constantes, ou seja, são homocedasticos (PIMENTEL; RABELLO; POUBEL, 2018).

Os autores se basearam em p-valor (informa a probabilidade de ocorrência da hipótese de nulidade), quanto menor o p-valor obtido maior confiança em rejeitar a hipótese nula e assumir a hipótese alternativa.

De acordo com a tabela 5.4, os níveis dos meios influenciam significativamente no valor da porosidade. O mesmo ocorrer para a altura. Já para a interação entre os fatores, aceita-se H_0.

Buscando comparar as médias dos níveis dos fatores, foi realizado o teste de Tukey, resultados da Tabelas 5.5 verificou que há diferença estatística significativa entre os meios, e que o maior valor de porosidade corresponde ao nível A2 (Mix). Tal afirmação condiz com a análise gráfica.

Já para os níveis do Fator 2 pode-se afirmar que os níveis B1 e B2 são iguais e possuem os maiores valores de porosidade; e que o nível B3 difere estatisticamente dos demais e possui o menor valor de porosidade. Tal resultado condiz com o esperado, uma vez que B3 possuem a maior altura e assim sofrerá maior compactação do leito, devido ao efeito da gravidade.

Tabela 5.4 Tabela 2- Valor p das fontes de variação da ANOVA para o fatorial em DIC, obtido no RStudio.

	Meio	**Altura**	**Interação Meio e altura**
F	1297,122	8,192	3,297
p	1,34exp(-13)	0,0057	0,0722

Fonte: Pimentel, Rabello e Poubel (2018)

Tabela 5.5 Médias dos níveis, obtido no RStudio

Nível	**Média**
mA1	$0{,}4134^{a}$
mA2	$0{,}5431^{b}$
mB1	$0{,}4872^{A}$
mB2	$0{,}1861^{A}$
mB3	$0{,}4705^{B}$

Legenda: Médias com letra minúscula refere-se ao fator meio e a maiúsculas ao fator altura.
Fonte: Pimentel, Rabello e Poubel (2018)

Referências

CALADO, Verônica; MONTGOMERY, Douglas. **Planejamento de Experimentos usando o Statistica**. Editora E-papers, 2003.

CECON, Paulo Roberto et al. **Métodos Estatísticos**: Série Didática. Viçosa: Ufv, 2012. 229 p.

NORMANDO, David; TJÄDERHANE, Leo; QUINTÃO, Cátia Cardoso Abdo. A escolha do teste estatístico – um tutorial em forma de apresentação em PowerPoint. **Dental Press J. Orthod**, Rio de Janeiro, v. 15, p.101-106, jan. 2010. Disponível em: <https://www.dentalpress.com.br/portal/bioestatistica/>. Acesso em: 22 nov. 2017.

RABELLO, G. L.; PIMENTEL, U. F. (2017). Modelagem e Simulação de um secador de leito fixo para grãos de café utilizando programação C/C++. Trabalho de conclusão de curso. Engenharia Química. Universidade Federal do Espírito Santo, Alegre.

Pimentel, U. F., Rabello, G. L.,; Poubel, W. M. (2018). ANÁLISE EXPERIMENTAL DA POROSIDADE DO LEITO DE CAFÉ EM SECADORES ESTÁTICOS. No Anais do XVIII Encontro Latino Americano de Pós-graduação, XII Inic Jr da UNIVAP, VII INID (pp. 1-6). São José dos Campos: UNIVAP.

TRIOLA, Mario F. **Introdução à Estatística**. 10. ed. Rio de Janeiro: Ltc, 2008. 700 p.
TSTUDIO. **Take control of your R code**. Disponível em: <https://www.rstudio.com/products/rstudio/>. Acesso em: 24 nov. 2017.

VIEIRA, Sonia. **Análise de Variância**: ANOVA. São Paulo: Atlas, 2006. 2004 p.

Anexo

Link para acesso do material complementar:

Link ou https://drive.google.com/drive/folders/1Xx0eI-dJ4oOBn1ZT_gLfhs-b6wwE_kRgm?usp=sharing

SOBRE OS AUTORES

Uilla Fava Pimentel

Doutoranda em Processos Químicos e Bioquímicos pelo Programa da Escola de Química (EQ/EPQB) na Universidade Federal do Rio de Janeiro (UFRJ). Mestre em Processos Químicos e Bioquímicos (EQ/ EPQB/ UFRJ). Engenheira química pela Universidade Federal do Espírito Santo (UFES) e licenciada em Química pelo Instituto Federal do Espírito Santo (IFES). Atuou como monitora de estatística básica e bioestatística por mais de 2 anos. Possui experiência nas áreas de processos químicos, termodinâmica, estatística e planejamento experimental.

Gildeir Lima Rabello

Graduado pela Universidade Federal do Espírito Santo (Alegre) em Engenharia Química, mestre pelo Programa de Pós Graduação Em Engenharia Química na Universidade Federal Rural do Rio de Janeiro na área de Tecnologia Química, e atualmente é aluno de doutorado em Engenharia Química pelo Programa de Pós-Graduação em Engenharia de Processos Químicos e Bioquímicos (EPQB) da Universidade Federal do Rio de Janeiro. Técnico em agroindústria Integrado ao Ensino Médio no Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo - Campus de Alegre

CURSO DE ESTATÍSTICA PARA TODOS

ARCO EDITORES

www.arcoeditores.com
contato@arcoeditores.com
(55)99723-4952