LIVRO TEXTO

# ESTATÍSTICA

EAD.FAMINAS
EDUCAÇÃO A DISTÂNCIA

S586e Silva Filho, Augusto Souza da

Estatística / Augusto Souza da Silva Filho. - Muriaé: Faculdade de Minas, 2015.
77 p.

1. Estatística - Apostila. I. Santos, Érica Marques da Silva. II. Silva, Wanderley da. III.Título.

Bibliotecária responsável: Ana Paula – CRB-6/2...

Revisão e organização: Fernanda Cristina Abrão da Rocha

Editoração: Jéssica A. Corrêa do E. Santo

# Sum[redacted]

## APRESENTAÇÃO

A busca pelo conhecimento científico nos leva ao processo de investigação que, por sua vez sugere pesquisa, busca de informações e análise de dados. E isto nos leva a pensar em Estatística.

A Estatística está cada vez mais relacionada com as demais ciências. Por exemplo, a estatística auxilia a Genética nas questões de hereditariedade; é valiosa na Economia, na análise da produtividade, da rentabilidade e estudos de viabilidade; é básica para as Ciências Sociais nas pesquisas socioeconômicas; é de aplicação intensa na Engenharia Industrial, no controle de qualidade e na comparação de fabricações, é também muito aplicada na engenharia agrícola, entre outras.

A aplicação da Estatística cresceu nas últimas décadas, tornando-se o foco do estudo de especialistas para as áreas econômicas, sociais, culturais, políticas, educacionais, de saúde, meio ambiente, etc. Enfim, explícita ou implicitamente, a Estatística está presente em todos os aspectos da vida moderna, e essa presença só tende a crescer, pois o estudo estatístico colabora como indicardor para trabalhar com diversos produtos, possibilitando maiores estratégias na busca e no planejamento de soluções.

Atualmente um novo olhar está acontecendo em relação à estatásitica por exemplo, quando a estatística é aplicada a dados provenientes de observações realizadas em diferentes aspectos das Ciências da Vida, como: Medicina, Psicologia, Nutrição, Biologia, Farmácia, Enfermagem, Odontologia, Veterinária e Agronomia, é utilizado o termo bioestatístico para distingui-la da aplicação de outras áreas do conhecimento. Entretanto os conceitos e técnicas são os mesmos.

A estatística fornece-nos as técnicas para extrair informação de dados, os quais são muitas vezes incompletos, na medida em que nos dão informação útil sobre o problema em estudo.

Sendo assim, é um dos objetivos da Estatística extrair informação dos dados para obter uma melhor compreensão das situações que representam. Quando se aborda uma problemática envolvendo métodos estatísticos, estes devem ser utilizados mesmo antes de se recolher a amostra, isto é, deve-se planejar a experiência que nos vai permitir recolher os dados, de modo que, posteriormente, se possa extrair o máximo de informação relevante para o problema em estudo, ou seja, para a população de onde os dados provêm.

Quando de posse dos dados, procura-se agrupá-los e reduzi-los, sob forma de amostra, deixando de lado a aleatoriedade presente.

Seguidamente o objetivo do estudo estatístico pode ser o de estimar uma quantidade ou testar uma hipótese, utilizando-se técnicas estatísticas convenientes as quais realçam toda a potencialidade da Estatística na medida em que vão permitir tirar conclusões acerca de uma população, baseando-se numa pequena amostra, dando-nos ainda uma medida do erro cometido.

Concluindo, a estatística tornou-se uma poderosa ferramenta para a compreensão, análise e previsão de inúmeras situações na nossa vida. Para podermos nos situar de forma mais precisa possível nesse rápido processo de mudanças que enfrentamos, é necessária a utilização dessas ferramentas. Por isso, é necessário saber ler gráficos, interpretá-los, prever situações, analisar dados, etc. A estatística nos ajudará nesta tarefa.

Assim, estude cuidadosamente este material. Refaça os exemplos apresentados e busque apoio nas indicações fornecidas no tópico pesquisando.

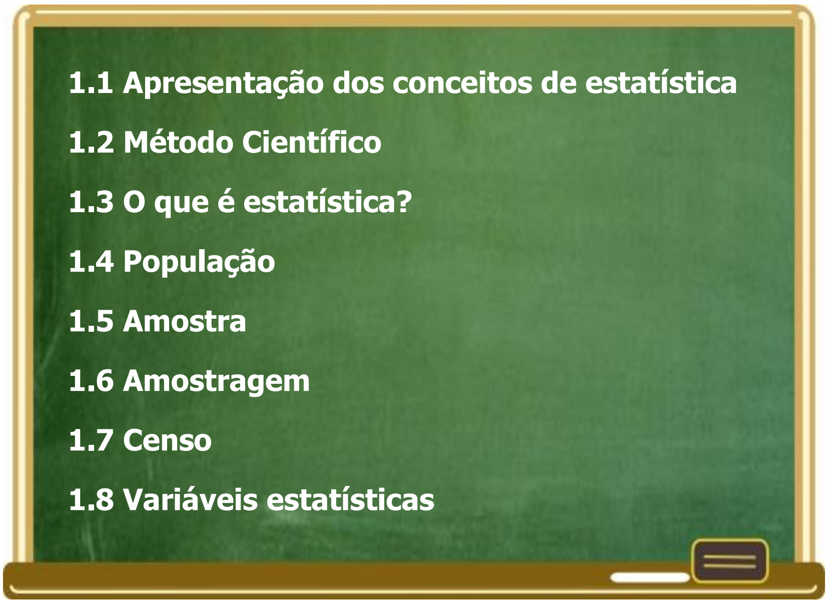
# UNIDADE I –INTRODUÇÃO À ESTATÍSTICA

## Objetivos

- Definir estatística e seus tipos.
- Reconhecer os processos relacionados ao estudo da população na estatística.
- Diferenciar e aplicar os métodos científicos.
- Diferenciar e determinar as variáveis quantitativas e qualitativas.

## CONSIDERAÇÕES INICIAIS

Vamos iniciair nossos estudos construindo a base para o desenvolvimento desta disciplina. Os conceitos abordados aqui, irão nos acompanhar ao longo de todo nosso estudo.




## 1.1– APRESENTAÇÃO DOS CONCEITOS DE ESTATÍSTICA

Ao realizarmos uma pesquisa científica, o procedimento comum e geral inciarmos nosso trabalho formulando hipóteses e em seguida partimos para o teste destas hipóteses.

Inicialmente, essas hipóteses são formuladas em termos científicos dentro da área de estudo. Em seguida, devem ser expressas em termos estatísticos. Existem muitas definições propostas por autores que objetivam estabelecer com clareza a estatística, mas de uma maneira geral, como veremos a seguir, ela visa elaborar métodos capazes e aplicáveis a todas as fases do estudo dos fenômenos de massa.

## 1.2 - MÉTODO CIENTÍFICO

Método é um conjunto de meios e rotinas dispostos convenientemente para chegar a um fim que se deseja de forma organizada e confiável.

Aqui iremos considerar dois tipos de métodos científicos:

### Método experimental:

- Este método que consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que possa achar seus efeitos.

### Método estatístico:

- Este método que admite todas as causas presentes variando-as, dada a impossibilidade de manter as causas constantes, registrando essas variações e procurando determinar as influências que cabem a cada uma delas.

## 1.3 - O QUE É ESTATÍSTICA?

Você sabe quantas pessoas existem na sua casa? Com certeza. Mas em toda a sua família, você sabe? Bem... Quantas pessoas existem na sua rua? E no seu bairro? E na sua cidade? E na sua Faculdade? E no seu estado? E no Brasil? E no mundo, afinal?

Bem, pode ser que você considere essas preocupações bastante exageradas, mas nem sempre o mundo foi tão populoso.

Se pararmos para pensar na população mundial de um tempo atrás, digamos, no século XV, veremos que a quantidade de pessoas era bem menor. Se voltássemos à Grécia Antiga, menor ainda. Pois bem, esse crescimento acelerado de habitantes foi verificado no mundo moderno, com a sociedade de massas. A partir daí, a Estatística se tornou, juntamente com a ciência da economia, a ciência social por excelência.

Por quê?

Porque lidamos com grandes números.

A Estatística ou métodos estatísticos, como é chamada algumas vezes, nasceu com os negócios do Estado, daí seu nome.

Mas, hoje, sua influência pode ser encontrada nas mais diversas atividades: agricultura, biologia, comércio, química, comunicações, economia, educação, medicina, ciências políticas e muitas outras.

A Estatística se interessa pelos métodos científicos para coleta, organização, resumo, apresentação e análise de dados, bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises. Algumas vezes, o termo Estatística é empregado para designar os próprios dados ou números, por exemplo, estatística de empregos, de acidentes etc. Observe que, na figura 1, são apresentados os elementos principais da definição de estatísitca

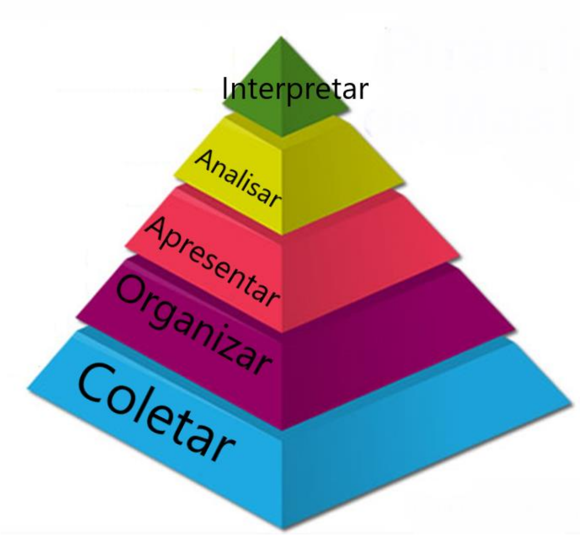



FIGURA 1 ELEMENTOS PRINCIPAIS DA DEFINIÇÃO DE ESTATÍSITCA

Assim, fique atento a importância desses verbos para nosso estudo, e para sua atualização profissional.

É comum que, as pessoas limitem o termo Estatística à organização e descrição dos dados, quando na verdade sua maior contribuição é:

**Proporcionar métodos inferenciais, que permitam conclusões que transcendam os dados obtidos inicialmente.**

Através da análise e interpretação dos dados estatísticos que é possível o conhecimento de uma realidade, de seus problemas, bem como, a formulação de soluções apropriadas por meio de um planejamento objetivo da ação, para além dos “achismos” e “casuismos” comuns.

Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados. Ela é dividida em:

**Estatística Descritiva:** A Estatística que lida com a organização, resumo e apresentação de dados numéricos é denominada de Estatística Descritiva. Assim, pode-se definir a Estatística Descritiva como sendo: Os procedimentos usados para organizar, resumir e apresentar dados numéricos.

Conjuntos de dados desorganizados são de pouco ou nenhum valor. Para que os dados se transformem em informação é necessário organizá-los, resumi-los e apresentá-los. O resumo de conjuntos de dados é feito através das medidas, a organização e apresentação através das distribuições de frequências, gráficos ou diagramas.

**Estatística Indutiva:** também conhecida como amostral ou inferencial, é aquela que partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões fundamentando-se na teoria das probabilidades.

## 1.4 – POPULAÇÃO

É todo conjunto, finito ou infinito, que possui ao menos uma característica em comum entre todos os seus elementos componentes.

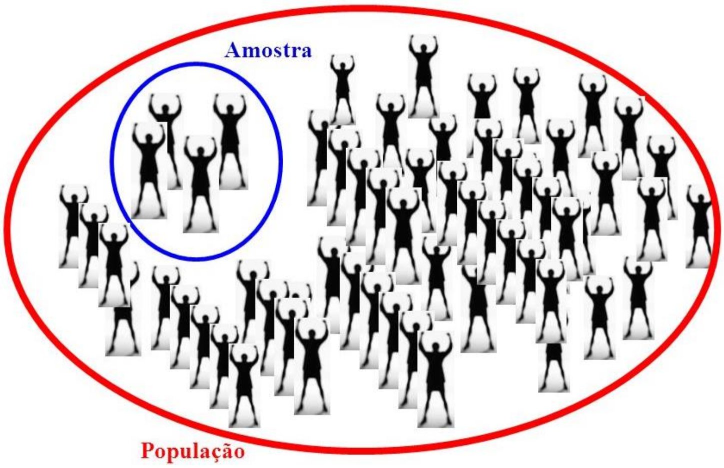
## 1.5 – AMOSTRA

É uma parte ou um subconjunto representativo de uma população, isto é, é um conjunto de elementos extraídos da população. Os dados de observação registrados na amostra fornecem informações sobre a população. O processo pelo qual são tiradas conclusões sobre a população, com base nos resultados obtidos na amostra, refere-se à inferência estatística vista na definição.

As estatísticas obtidas na amostra são denominadas estimativas. Portanto, toda a análise estatística será inferida a partir das características obtidas da amostra. É importante que a amostra seja representativa da população, isto é, que as suas características sejam, em geral, as mesmas que as do todo (população).

Enfim, a amostra é um subconjunto finito e representativo de uma população. Muitas vezes, por motivos práticos ou econômicos, limitam-se os estudos estatísticos somente a uma parte da população, à amostra.

As razões de recorrer a amostras são: menor custo e tempo para o levantamento de dados e melhor investigação dos elementos observados.




EXEMPLO

**Há apenas um verdadeiro amor para cada pessoa?**

Uma pesquisa nacional por telefone, nos Estados Unidos, realizada pela Pew Foundation,1 em outubro de 2010, perguntou a 2625 adultos com 18 anos ou mais "Algumas pessoas dizem que há apenas um verdadeiro amor para cada pessoa. Você concorda ou discorda?" Além de encontrar a proporção dos que concordavam com a afirmativa, a Pew Foundation intencionava, também, descobrir se tal quantitativo era diferente entre os homens e as mulheres, e se a proporção dos que concordavam variava com base em nível de educação (nenhum estudo superior, algum estudo superior, ou grau universitário). Os participantes da pesquisa foram selecionados aleatoriamente, por telefones fixos e celulares.

Para este caso : Qual é a amostra? Qual é a população?

**Solução**

A **amostra** são as 2625 pessoas observadas e entrevistadas.

A **população** são todos os adultos com 18 anos de idade, ou mais, que têm telefone fixo ou celular.

Viu como é simples! Mas, fique atento, pois, esses conceitos serão muito usados em nossos estudos.

## 1.6 – AMOSTRAGEM

**DEFINIÇÃO:** É a coleta das informações de parte da população chamada amostra mediante métodos adequados de seleção dessas unidades. Amostragem é considerada uma técnica especial de escolher amostras, de forma a garantir o acaso na escolha. Assim, cada elemento da população tem a mesma chance de ser escolhido, o que garante à amostra um caráter de representatividade da população.

### 1.6.1 TÉCNICAS DE AMOSTRAGEM PROBABILÍSTICAS

Existem técnicas adequadas para recolher amostras, de forma a garantir (o quanto possível) o sucesso da pesquisa e dos resultados.

Definidos os objetivos e a população a ser estudada, deve-se pensar em como será constituída a amostra dos dados e quais as características ou variáveis a serem estudadas.

a) Amostragem casual ou aleatória simples – este tipo de amostragem é baseado no sorteio da amostra. Numera-se a população de 1 a n, e, utilizando um dispositivo aleatório qualquer, por exemplo, sorteio, escolhem-se k números dessa sequência, os quais corresponderão aos elementos da amostra. Observe a figura 2,

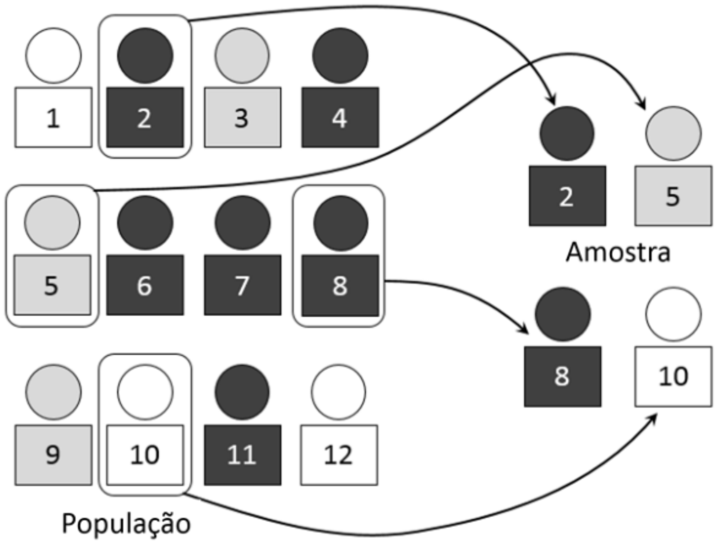



FIGURA 2 - AMOSTRAGEM SIMPLES ALEATÓRIA

**EXEMPLO**

Uma cidade turística tem 30 hotéis de três estrelas. Pretende-se conhecer o custo médio da diária para apartamento de casal. Os valores populacionais consistem nos seguintes preços diários (em dólares): 25, 20, 35, 21, 22, 24, 25, 30, 38, 24, 20, 20, 25, 20, 19, 25, 23, 24, 28, 24, 24, 22, 28, 26, 23, 25, 22, 27, 25, 23. Extraia uma amostra aleatória simples de tamanho 10 desta população.

***Vamos resolver este problema de duas formas: por sorteio e por tabela de números aleatórios.***

POR SORTEIO

Escrevemos os valores em papéis, então os colocamos em uma urna, misturamos e.sorteamos a amostra de n = 10

Resultado obtido: n= (20, 24, 22, 28, 23, 24, 21, 20, 25, 27)

Esse processo não é muito prático para grandes populações, nesse caso é preferível utilizar uma tabela de números aleatórios.

POR TABELA DE NÚMEROS ALEATÓRIOS:

*1º passo:* Elaborar a relação dos dados brutos da população, ordenando os números com uma numeração aleatória. Como dispomos de um conjunto de elementos de 30 números começaremos pelo 00 até o 29, usando dois dígitos, caso tivéssemos 1000 elementos, iniciaríamos pelo 000 até o 999, e assim sucessivamente, usando então três dígitos.

| **N.º** | **Hotel Custo** | **N.º** | **Hotel Custo** | **N.º** | **Hotel Custo** | **N.º** | **Hotel Custo** | **N.º** | **Hotel Custo** |
|---|---|---|---|---|---|---|---|---|---|
| 00 | $25 | 06 | $25 | 12 | $20 | 18 | $28 | 24 | $23 |
| 01 | $20 | 07 | $30 | 13 | $19 | 19 | $24 | 25 | $25 |
| 02 | $25 | 08 | $38 | 14 | $25 | 20 | $24 | 26 | $27 |
| 03 | $21 | 09 | $24 | 15 | $23 | 21 | $22 | 27 | $25 |
| 04 | $22 | 10 | $20 | 16 | $20 | 22 | $28 | 28 | $25 |
| 05 | $24 | 11 | $25 | 17 | $24 | 23 | $26 | 29 | $23 |

2º Passo: Agora iremos sortear o valor de **n**, aqui num tamanho igual a 10, utilizando a tabela de números aleatórios. Utilizaremos a tabela agrupando 2 em 2 números pois nossa amostra é de dois dígitos, começando de qualquer ponto na vertical ou na horizontal, até conseguirmos sortear o tamanho de **n** existente.

| | | |
|---|---|---|
| *09 - $24* | *13 - $19* | *19 - $24* |
| *11 - $25* | *20 - $24* | *18 - $28* |
| *12 - $20* | *21 - $22* | |
| *25 - $25* | *06 - $25* | |

3º Passo: Acima estão os números sorteados, os que não têm na amostra são descartados, e no nosso caso como não utilizaremos as repetições, pois queremos um sorteio sem reposição, então também serão descartadas as repetições.

***Nossa Amostra então será: (24, 25, 20, 25, 19, 24, 22, 25, 24,28)***

b) Amostragem proporcional estratificada – quando as populações se dividem em subpopulações (estratos), pode ser razoável supor que a variável de interesse apresente comportamento distinto nos diferentes estratos. Assim, para que uma amostra seja representativa, é necessário utilizar-se uma amostragem proporcional estratificada, que considera os estratos (subgrupos) e obtém a amostragem proporcional a estes. Observe a figura 3,

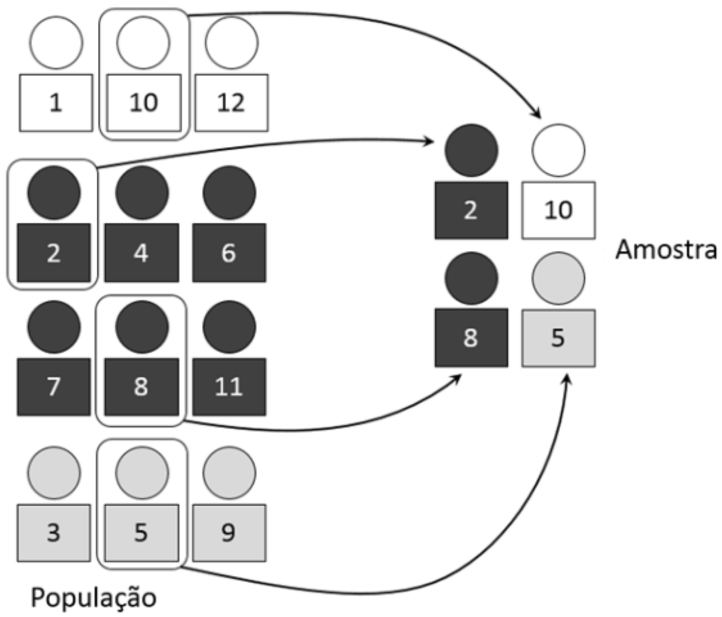



**FIGURA 3 AMOSTRAGEM PROPORCIONAL ESTRATIFICADA**

EXEMPLO

Considerando um grupo de 90 alunos de uma escola, onde 54 sejam meninos e 36 sejam meninas. Teremos dois estratos (sexo masculino e sexo feminino) e queremos uma amostra de 10% da população, assim:

• Definimos a amostra em estratos:

• Numeramos os alunos de 01 a 90, sendo que 01 a 54 correspondem aos meninos e de 55 a 90, meninas.

| Sexo | População | 10% | Amostra |
|---|---|---|---|
| **Meninos** | 54 | 5,4 | 5 |
| **Meninas** | 36 | 3,6 | 4 |
| **Total** | 90 | 9,0 | 9 |

Para determinar a amostra, efetuamos os sorteios até atingirmos 5 meninos (por exemplo, os de números 05, 17, 31, 46 e 53) e quatro meninas (por exemplo, as de números 63, 74, 75 e 90).

• Nesse caso, serão obtidas as características dos seguintes alunos:

Meninos: 05, 17, 31, 46 e 53.

Meninas: 63, 74, 75 e 90.

<u>c) Amostragem estratificada uniforme</u> – não utiliza o critério de proporcionalidade, pois se seleciona a mesma quantidade de elementos de cada estrato, devendo ser usada para comparar os estratos ou obter estimativas separadas para cada estrato.

EXEMPLO

Uma empresa de automação conta com 480 funcionários, dos quais 288 são do sexo feminino e os 192 restantes do sexo masculino. Considerando a variável “sexo” para estratificar essa população, vamos obter uma amostra estratificada uniforme de 50 funcionários.

Supondo que haja homogeneidade dentro de cada categoria, pode-se obter amostra estratificada uniforme de 50 funcionários com a seleção de 25 elementos de cada estrato.

| Estrato (por sexo) | População | Amostra estratificada uniforme |
|---|---|---|
| **Feminino** | 288 | $n_1 = 25$ |
| **Masculino** | 192 | $n_2 = 25$ |
| **Total** | 480 | 50 |

d) Amostragem sistemática – é um procedimento para a amostragem aleatória aplicada quando os elementos da população já estão ordenados. Assim, não é necessário construir um sistema de referência ou de amostragem. Agora, observe a figura 4:

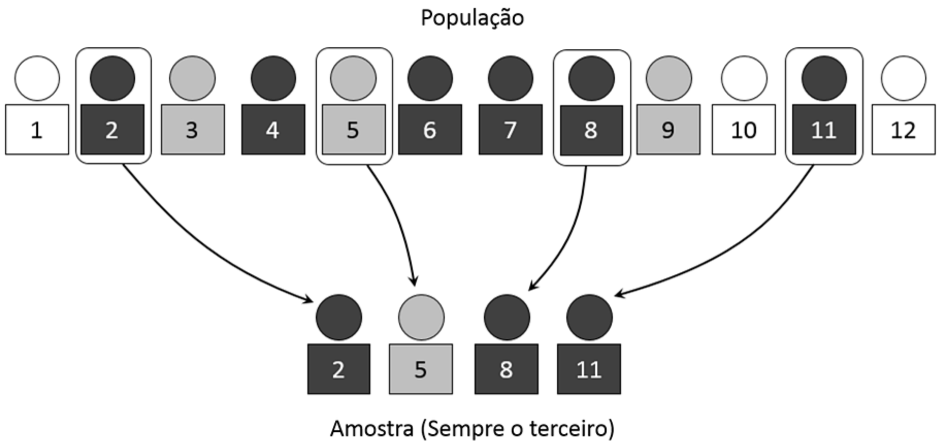



FIGURA 4 AMOSTRAGEM SISTEMÁTICA

EXEMPLO

Num estoque de 100 peças, para obtermos 10 amostras sistemáticas podemos retirar as peças de número 10, 20, 30, e assim por diante, até completarmos 10 amostras sistematicamente colhidas.

Para encontrarmos os pontos onde faremos as coletas sistemáticas das amostras, podemos seguir os seguintes passos:

1º) Define-se tamanho da população: N = 1.600.

2º) Define-se o tamanho da amostragem total: n = 100. $I = \frac{N}{n} = \frac{1600}{100} = 16$

Onde: I é igual ao intervalo de seleção

3º) Sorteia-se um número de 1 a 16, que será o primeiro número da amostra, logo, as próximas amostras serão retiradas de 16 em 16.

**Observação:**

O intervalo da seleção corresponde ao número de vezes que a amostra cabe na população.

<u>e) Amostragem por Conglomerados</u> – A amostra por conglomerados é uma técnica que explora existência de grupos (*clusters*) na população. Esses grupos representam adequadamente a população total em relação a característica que queremos medir. Em outras palavras, estes grupos contêm variabilidade da população inteira. Se isso acontecer, você pode selecionar apenas alguns desses conglomerados para realizar o estudo, conforme ilustra a figura 5.

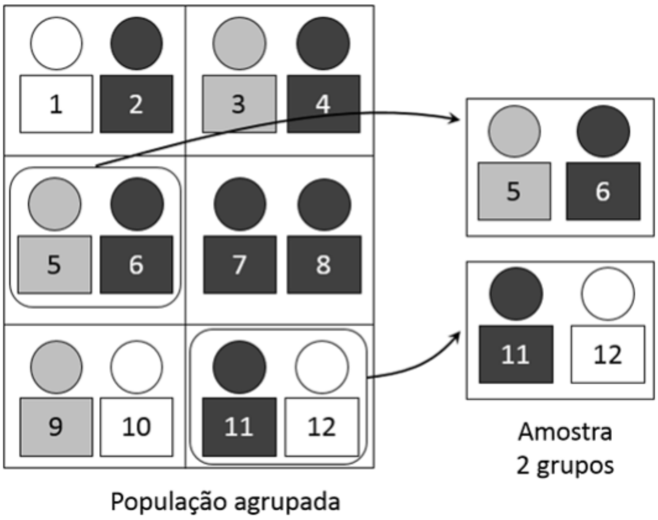



**FIGURA 5 AMOSTRAGEM POR CONGLOMERADOS**

Podemos ver esta técnica a partir de outro ponto de vista. Enquanto todas as técnicas de amostragem estudadas até agora as unidades da amostra coincidem com os indivíduos a serem estudados, na amostra por conglomerados, a unidades de amostra são grupos do estudo, o que pode ser muito benéfico em relação ao custo de amostragem em si. Em troca, é comum obter uma menor precisão ao utilizar esta técnica, causada pela falta de heterogeneidade dentro dos conglomerados.

Uma população estratificada em que o número de estratos é muito grande, ao invés de sortear uma amostra de cada estrato, o que poderia ser inviável devido à quantidade de estratos, o pesquisador poderia optar por sortear alguns estratos (conglomerados)

EXEMPLO

Sabemos que o Centro Universitário Unifaminas, possui diversos cursos de graduação, então imaginem que faremos uma pesquisa e utilizaremos amostragem por conglomerados para seleção da amostra.

Vamos proceder da seguinte forma: Separamos os cursos e rotulamos cada curso por um número, por exemplo:

| | | |
|---|---|---|
| **Administração – 01** | **Enfermagem – 07** | **Medicina – 13** |
| **Arquitetura e Urbanismo – 02** | **Engenharia Civil – 08** | **Nutrição – 14** |
| **Biomedicina – 03** | **Engenharia de Produção – 09** | **Odontologia – 15** |
| **Ciências Contábeis – 04** | **Farmácia – 10** | **Psicologia – 16** |
| **Direito – 05** | **Fisioterapia – 11** | |
| **Educação Física – 06** | **Gastronomia – 12** | |

O próximo passo é sorteamos ao acaso quatro números, supondo que os números sorteados sejam: 02; 12; 08; 05

Desta forma todos alunos que cursam: ***Arquitetura e Urbanismo, Gastronomia, Engenharia Civil e Direito***, farão parte da amostra.

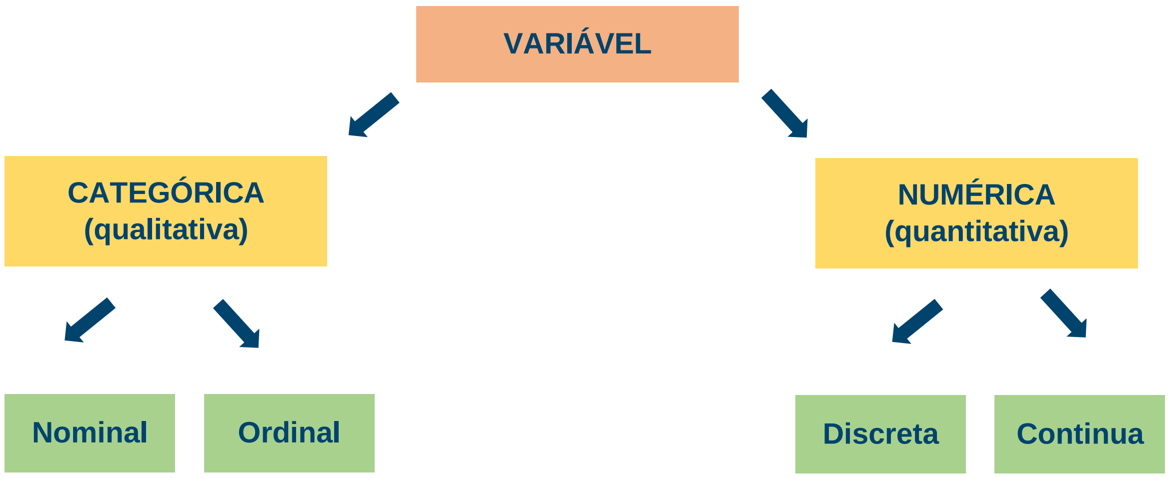
Observe que não selecionamos ao acaso os indivíduos (estudantes) e sim o curso (conglomerado), desta forma todos alunos dos cursos selecionados irão compor a amostra.

## 1.7 – CENSO

É o exame completo de toda população. Quanto maior a amostra, mais precisas e confiáveis deverão ser as induções feitas sobre a população. Logo, os resultados mais perfeitos são obtidos pelo Censo. Na prática, esta conclusão, muitas vezes não acontece, pois o emprego de amostras com certo rigor técnico pode levar a resultados mais confiáveis ou até mesmo melhores do que os que seriam obtidos através de um censo.

## 1.8 – VARIÁVEIS ESTATÍSTICAS

Em estatística nos deparamos com informações a serem trabalhadas que se apresentam de diversas formas, o conceito de variável é aplicado na estatística para classificar os dados quanto suas características. Assim temos:




### Variável Numérica Contínua

Refere-se a dados de mensuração, podem existir valores intermediários, ex: peso, altura, uréia, creatinina, hemoglobina.

### Variável Numérica Discreta

Só podem assumir valores numéricos inteiros, ex: número de consultas médicas, número de episódios de uma enfermidade.

### Variável Qualitativa Nominal

São dados que se definem exclusivamente por nomes (não são mensurados), ex: grupo sanguíneo (A, AB, B e O), estado civil (casado/viúvo/solteiro, etc), raça, sexo.

### Variável Qualitativa Ordinal

Os dados são ordenados de alguma maneira (incluem escalas). Ex: estadiamento de doença (avançada, moderada, branda, nenhuma), grau da dor (forte, moderada, branda, nenhuma), grau de escolaridade, categoria salarial.

### EXEMPLO

No site do Census Bureau, podem-se ver, em detalhes, os dados coletados pela Pesquisa da Comunidade Americana, embora, naturalmente, as identidades das pessoas e das unidades residenciais sejam protegidas. Se escolhermos o arquivo de dados sobre pessoas, os indivíduos são as pessoas que moram nas unidades residenciais contatadas pela pesquisa. Mais de 100 variáveis são registradas para cada indivíduo. Vamos analisar uma pequena parte dos dados.

- PESO (em quilogramas)
- RAÇA
- GRAU MAIS ELEVADO DE ESCOLARIZAÇÃO (no Ensino Fundamental completo, Ensino Fundamental Incompleto, Ensino Médio completo, Ensino Médio Incompleto, Ensino Superior completo, Ensino superio Incompleto e Pós graduação)
- NÚMERO DE PESSOAS QUE RESIDEM NA MESMA CASA.

Classifique cada variável apresentada.

Para classificarmos é importante verificar com calma do que se trata a informação e se, possuírem unidade de medida, como elas estãos sendo consideradas. Então vamos lá!

PESO – Variável qualitativa e contínua

RAÇA – Variável qualitativa e nominal

GRAU MAIS ELEVADO DE ESCOLARIZAÇÃO Variável qualitativa e ordinal

NÚMERO DE PESSOAS QUE RESIDEM NA MESMA CASA. Variável qualitativa e discreta

## ATIVIDADES DE APRENDIZAGEM

Após ter lido este material, experimente responder estas perguntas de acordo com o que você entendeu, depois pode verificar no texto se sua resposta está correta!

Bom Trabalho!!!!

1. Explique o que se entende por estatística.
2. Como a estatística se divide?
3. Quais as diferenças entre população, amostra e amostragem?
4. Defina os métodos de amostragem. 5. Qual o significado do censo?

## SUGESTÃO DE LEITURA

Aprenda mais sobre a história da estatística e os conceitos estudados, acessando os seguintes sites:

http://www.somatematica.com.br/estat/basica/pagina2.php

http://alea-estp.ine.pt/

http://www.leg.ufpr.br/~silvia/CE055/

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

SILVEIRA JÚNIOR, P., MACHADO, A. A., ZONITA, E.P., SILVA, J.B. da**. Curso de Estatística.** Pelotas: Universidade Federal de Pelotas, 1989. 135p. V1

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

NAZARETH, HELANALDA. **Curso Básico de Estatítica**. 12. ed. São Paulo: Ática, 2000.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelando o Poder dos dados** [VitalSource Bookshelf version]. Retrieved from vbk://9788521633440

S., M. D., I., N. W., A., F. M. (07/2017). **A Estatística Básica e sua Prática**, 7ª edição [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287

## UNIDADE 2 – FASES DO MÉTODO ESTATÍSTICO

### Objetivos

- Conhecer e analisar as fases do método estatístico.
- Empregar as fases do método estatístico.

Ao desenvolver um estudo estatístico completo, existem algumas fases do seu método que devem ser desenvolvidas em sequência, para chegar aos resultados finais do trabalho. Assim, nesta aula estudaremos essas fases.

## 2 - FASES DO MÉTODO ESTATÍSTICO

Quando se pretende empreender um estudo estatístico completo, existem diversas fases do trabalho que devem ser desenvolvidas para se chegar aos resultados finais de um estudo capaz de produzir resultados válidos. As fases principais são as seguintes:


## 2.1 - DEFINIÇÃO DO PROBLEMA

Iniciamos essa primeira fase do método estatístico com a pergunta: o que PESQUISAR? Nessa fase você deve saber exatamente aquilo que se pretende pesquisar.

## 2.2 - PLANEJAMENTO

Como levantar informações? Que dados deverão ser obtidos?Qual levantamento a ser utilizado? Censitário? Poramostragem? E o cronograma de atividades? Os custos envolvidos? etc.

## 2.3 - COLETA DE DADOS

O terceiro passo é essencialmente operacional, compreendendo a coleta das informações propriamente ditas. Nesta fase, do método estatístico é conveniente estabelecer uma distinção entre duas espécies de dados:

- Dados primários – quando são publicados ou coletados pelo próprio pesquisador ou organização que os escolheu;
- Dados secundários – quando são publicados ou coletados por outra organização. Um conjunto de dados é, pois, primário ou secundário em relação a alguém. As tabelas do Censo Demográfico são fontes primárias. Quando determinado jornal publica estatísticas extraídas de várias fontes e relacionadas com diversos setores industriais, os dados são secundários para quem desejar utilizar-se deles em alguma pesquisa que esteja desenvolvendo.

A coleta de dados pode ser realizada de duas maneiras:

- Coleta Direta – quando é obtida diretamente da fonte, como no caso da empresa que realiza uma pesquisa para saber a preferência dos consumidores pela sua marca;
- Coleta Indireta – quando é inferida a partir dos elementos conseguidos pela coleta direta, ou através do conhecimento de outros fenômenos que, de algum modo, estejam relacionados com o fenômeno em questão.

## 2.4 - CRÍTICA DOS DADOS

Resumo dos dados através de sua contagem e agrupamento. É a condensação e tabulação de dados.

## 2.5 - APRESENTAÇÃO DOS DADOS

Há duas formas de apresentação, que não se excluem mutuamente. A **apresentação tabular**, ou seja é uma apresentação numérica dos dados em linhas e colunas distribuídas de modo ordenado, segundo regras práticas fixadas pelo Conselho Nacional de Estatística. A **apresentação gráfica** dos dados numéricos constitui uma apresentação geométrica permitindo uma visão rápida e clara do fenômeno.

## 2.6 - ANÁLISE E INTERPRETAÇÃO DOS DADOS

Nesta última etapa, o interesse maior reside em tirar conclusões que auxiliem o pesquisador a resolver seu problema. A análise dos estatísticos está ligada essencialmente ao cálculo de medidas, cuja finalidade principal é descrever o fenômeno.

Assim, o conjunto de dados a ser analisado pode ser expresso por números resumo, as estatísticas que evidenciam as características particulares desse conjunto. O significado exato de cada um dos valores obtidos através do cálculo das várias medidas estatísticas disponíveis deve ser bem interpretado.

É possível mesmo, nesta fase, arriscar algumas generalizações, as quais envolverão, como mencionado anteriormente, algum grau de incerteza, porque não se pode estar seguro de que o que foi constatado para aquele conjunto de dados (a amostra) se verificará igualmente para a população.

EXEMPLO

Uma professora decidiu realizar uma pesquisa sobre a altura média dos alunos do ensino médio da escola onde ela ministra aulas de biologia. Estabeleças as fases do método estatístico a ser realizado pela professora.

Então vejamos:

- Problema
  Altura média dos alunos do Ensino Médio .

- Planejamento
  A pesquisa será realizada com a população de alunos do Ensino Médio.

- Coleta
  Os dados serão primários, de modo que será medida a altura de cada estudante e registrados em uma tabela.

- Crítica
  Observar e analisar de forma crítica e cautelosa, os dados obtidos, pois no caso medida de altura, os valores não podem ser muito pequeno ou muito grande.

- Apresentação dos dados
  Os dados serão apresentados por meio de tabela.

- Análise e interpretação dos dados

  A partir dos dados analisados, investigar se a altura média dos estudantes estão dentro dos valores de referências determinados pelas agências de pesquisa e referência.

## ATIVIDADES DE APRENDIZAGEM

Vamos verificar o que você aprendeu!

Responda cada pergunta cuidadosamente e, reflita sobre os conceitos estudos até aqui!

1. Cite as fases do método estatístico.

2. Explique como ocorre a fase de coleta de dados.

3. Como se faz um planejamento através do método estatístico?

## *REFERÊNCIAS*

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

COSTA, Paulo Roberto. **Estatística**. Disponível em: https://www.ufsm.br/unidades-universitarias/ctism/cte/wpcontent/uploads/sites/413/2018/11/04_estatistica.pdf. Acesso em 18/07/2019.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelando o Poder dos dados** [VitalSource Bookshelf version]. Retrieved from vbk://9788521633440

S., M. D., I., N. W., A., F. M. (07/2017). **A Estatística Básica e sua Prática**, 7ª edição [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287.

### SUGESTÃO DE LEITURA

Para saber mais sobre a unidade que acabamos de estudar sugiro que pesquise:

- Acesse o site:

https://www.ufsm.br/unidades-universitarias/ctism/cte/wp-content/uploads/sites/413/2018/11/04_estatistica.pdf

## UNIDADE 3 – REPRESENTAÇÃO DOS DADOS ESTATÍSTICOS

### Objetivos

- Apresentar as formas de representação de dados estatísticos.
- Conhecer os tipos de séries estatísticas bem como suas aplicações;
- Conhecer os tipos de gráficos estatísticos bem como suas aplicações; e,
- Interpretar os dados obtidos através de sua representação gráfica.

Nesta unidade, trataremos da questão das tabelas e gráficos estatísticos. Também observaremos as séries estatísticas que são de fundamental importância no estudo descritivo. Pois, em todo estudo estatístico os dados observados necessitam serem organizados para que se faça a análise dos mesmos.

**3 Representação Tabular**
**3.1 Elementos de uma Tabela**
**3.2 Séries estatísticas**
**3.3 Representação gráfica**
**3.4 Gráficos Estatísticos**

## 2 – REPRESENTAÇÃO TABULAR

A apresentação tabular é uma apresentação numérica dos dados. Consiste em dispor os dados em linhas e colunas distribuídos de modo ordenado, segundo algumas regras práticas adotadas pelos diversos sistemas estatísticos. As regras que prevalecem no Brasil foram fixadas pelo Conselho Nacional de Estatística.

## 3.1 – ELEMENTOS DE UMA TABELA

É um quadro que resume um conjunto de observações. As tabelas têm a vantagem de conseguir expor, sinteticamente em um só local, os resultados sobre determinado assunto, de modo a se obter uma visão global mais rápida daquilo que se pretende analisar.

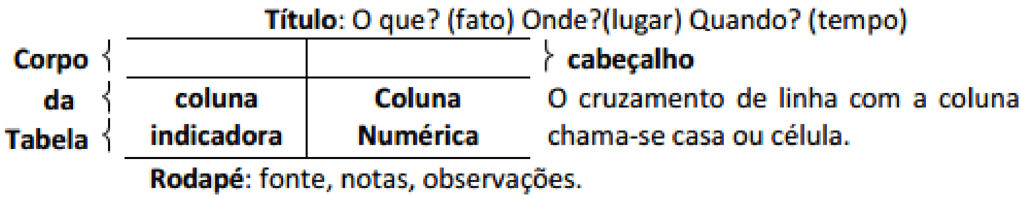



FIGURA 1- COMPONENTES DE UMA TABELA ESTATÍSTICA

**ATENÇÃO**

**1) Recomenda-se não delimitar (fechar) por traços verticais, os extremos da tabela, à direita e à esquerda;**

**2) Usa-se um traço horizontal ( - ) quando o dado for nulo, inexisti o fenômeno;**

**3) Usa-se (...) quando não se dispuser dos dados, embora ele possa ser quantificado;**

**4) Usa-se zero (0) quando o valor é muito pequeno para ser expresso pela unidade utilizada.**

**5) Usa-se uma interrogação (?) quando o valor é duvidoso.**

## 3.2 - SÉRIES ESTATÍSTICAS

Denomina-se série estatística toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local, ou da espécie (fenômeno).

Numa série estatística observa-se a existência de três elementos ou fatores: o tempo, o espaço e a espécie. Conforme varie um desses elementos, a série estatística classifica-se em temporal, geográfica e específica.

***Série Temporal ou Cronológica:*** Nesta série, o elemento de variação é o tempo (dia, mês, ano, etc). Também chamada de série temporal, série histórica, série evolutiva ou marcha, identifica-se pelo caráter variável do fator cronológico. Assim, deve-se ter:

Elemento variável: Época

Elementos Fixos: Local e Fenômeno.

| CESE | | | |
|---|---|---|---|
| **Ano Letivo** | Novas Inscrições | Total de Inscritos | Diplomados |
| **1993/94** | 38 | 38 | - |
| **1994/95** | 66 | 96 | 14 |
| **1995/96** | 53 | 111 | 9 |
| **1996/97** | 46 | 129 | 29 |
| **1997/98** | 46 | 133 | 13 |
| **1998/99** | a) | 86 | 24 |
| **1999/00** | a) | 42 | 21 |

Figura 2- Série temporal - Fonte: https://www.si.ips.pt/ests_si/web_base.gera_pagina?p_pagina=1183. Acesso em: 14 jul. 2019.

***Série Especificativa:*** Também chamada de série categórica ou série por categoria, identifica-se pelo caráter variável de fator especificativo. Assim, deve-se ter:

Elemento variável: Fenômeno

Elementos Fixos: Local e Época

| Matrícula por sexo – Penedo - 2000 | |
|---|---|
| Sexo | F |
| Masculino | 200 |
| Feminino | 1.000 |
| Total | 1.200 |

Figura 3-Série Especificativa - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

***Série Regional ou geográfica:*** Também chamada de série territorial, série espacial ou série de localização, identifica-se pelo caráter variável do fator geográfico. Assim, deve-se ter:

Elemento variável: Local

Elementos Fixos: Época e Fenômeno

| Matrícula por Município/AL - 2000 | |
|---|---|
| Municípios | F |
| Penedo | 1.200 |
| Piaçabuçu | 950 |
| Pariçonha | 700 |
| Total | 2.850 |

Figura 4- Série regional - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

***Série Mista:*** As tabelas apresentadas anteriormente são tabelas estatísticas simples, onde apenas uma série está representada. É comum, todavia, haver necessidade de apresentar, em uma única tabela, mais do que uma série. Quando as séries aparecem conjugadas, tem-se uma tabela de dupla entrada. Em uma tabela desse tipo são criadas duas ordens de classificação: uma horizontal (linha) e uma vertical (coluna).

EXEMPLO

**Especificativa x Temporal:** é a série estatística onde variam fenômeno e o tempo.

| Matrícula por Cursos, UFAL/2000-01 | | | |
|---|---|---|---|
| CURSOS | ANOS | | TOTAL |
| | 2000 | 2001 | |
| Medicina | 131 | 120 | 251 |
| Engenharia | 76 | 38 | 114 |
| Pedagogia | 92 | 147 | 239 |
| Economia | 34 | 86 | 120 |
| Serviço Social | 81 | 113 | 194 |

Figura 5- *Especificativa x Temporal* - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

EXEMPLO

Evasão Escolar por Estados – 1999/2000

| ESTADOS | ANOS | | TOTAL |
|---|---|---|---|
| | 1999 | 2000 | |
| São Paulo | 198 | 187 | 385 |
| Minas Gerais | 131 | 198 | 329 |
| Alagoas | 296 | 211 | 507 |
| Piauí | 341 | 131 | 472 |
| Sergipe | 121 | 148 | 269 |

Figura 6- Temporal x Geográfica - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

EXEMPLO

**Especificativa X Geográfica:** é a série estatística onde variam o fenômeno e o local.

Veículos Adquiridos por Regiões / 2001

| Veiculos | Regiões | | | | | Total |
|---|---|---|---|---|---|---|
| | Norte | Nordeste | Centro Oeste | Sul | Sudeste | |
| Kombi | 87 | 58 | 79 | 51 | 36 | 311 |
| Corsa | 56 | 71 | 86 | 88 | 92 | 393 |
| Ka | 93 | 84 | 71 | 81 | 62 | 391 |
| Escort | 48 | 76 | 90 | 75 | 81 | 370 |

Figura 7- Especificativa x Geográfica - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

EXEMPLO

**Especificativa X Especificativa:** é a série estatística onde o fenômeno varia mais de uma vez.

Distribuição de Material Escolar por Séries Alagoas/2011

| Materiais | Séries | | | Total |
|---|---|---|---|---|
| | 1ª | 2ª | 3ª | |
| Lapis | 53 | 21 | 39 | 113 |
| Borracha | 61 | 38 | 46 | 145 |
| Caneta | 32 | 71 | 60 | 163 |
| Lapiseira | 38 | 48 | 46 | 132 |

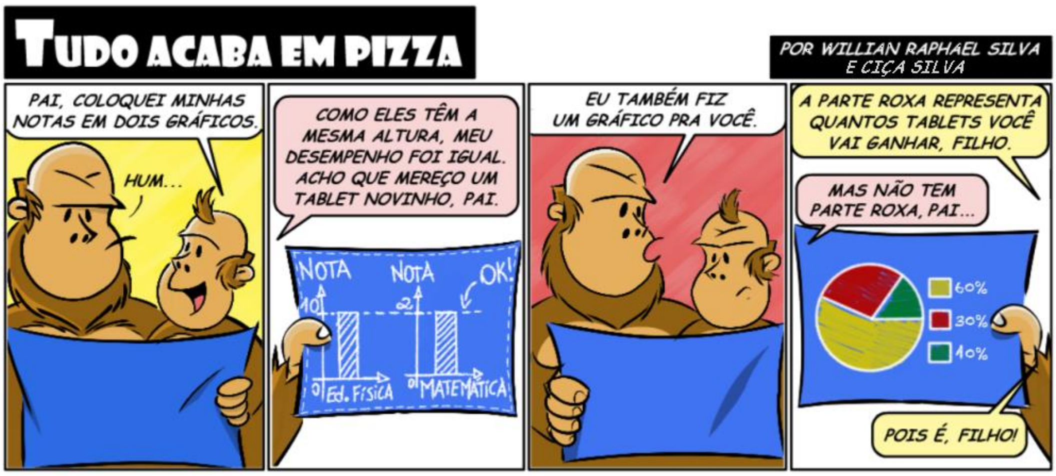
Figura 8- Especificativa x Especificativa - Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso em: 14 jul. 2019.

## ATIVIDADES DE APRENDIZAGEM

1. Qual a definição de série estatística e quais os fatores que a compõem?

2. Elabore um exemplo de série geográfica. Poste, lá no Fórum para que eu possa verificar se você compreendeu o conceito, ok!

## 3.3 - REPRESENTAÇÃO GRÁFICA




Fonte: https://blogdoprofh.wordpress.com/2014/10/15/notas-e-graficos-tirinha/. Acesso em: 14 jul. 2019.

Os gráficos são desenhos que envolvem formas e cores cuja construção utiliza técnicas de desenho. Os gráficos são de extrema importância na visualização e interpretação de informações e dados acerca de temas de aspectos naturais, sociais e econômicos.

Podemos dizer que os gráficos são representações visuais dos dados estatísticos cujo objetivo é produzir uma impressão mais rápida e viva do fenômeno em estudo. Os gráficos permitem a representação de uma relação entre variáveis e facilitam a compreensão de dados.

Além disso, os gráficos devem ser correspondentes às tabelas estatísticas, mas não devem substituí-las. Desse modo, a representação gráfica é um complemento importante da apresentação tabular.

A vantagem de um gráfico sobre a tabela está na possibilidade de uma rápida impressão visual da distribuição dos valores ou das frequências observadas. A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil, tais como:


**Deve possibilitar a análise rápida do fenômeno em estudo. Deve conter apenas o essencial.**


**Deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.**


**Deve expressar a verdade sobre o fenômeno em estudo.**

## 3.4 - FORMAS DE APRESENTAÇÃO DOS GRÁFICOS

Quando for levado em conta, o visual de acordo com a composição de formas, os gráficos podem ser:

3.4.1 **Gráficos de informação**: São gráficos destinados principalmente ao público em geral, objetivando proporcionar uma visualização rápida e clara. São gráficos tipicamente expositivos e dispensam comentários explicativos adicionais. As legendas podem ser omitidas, desde que as informações desejadas estejam presentes.

3.4.2 **Gráficos de análise**: São gráficos que servem principalmente ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser informativos. Os gráficos de análise frequentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes, um texto explicativo, que chama a atenção do leitor para os pontos principais revelados pelo gráfico.

## 3.5 - PRINCIPAIS TIPOS DE GRÁFICOS

### 3.5.1 – GRÁFICO EM LINHA OU EM CURVA:

Nesse tipo de gráfico utiliza-se uma linha poligonal para representar séries temporais, principalmente quando a série cobrir um grande número de períodos de tempo.

Neste sistema, faz-se uso de duas retas perpendiculares; as retas são os eixos coordenados e o ponto de intersecção, a origem. O eixo horizontal é denominado eixo das abscissas (ou eixo dos x) e o vertical, eixo das ordenadas (ou eixo dos y).

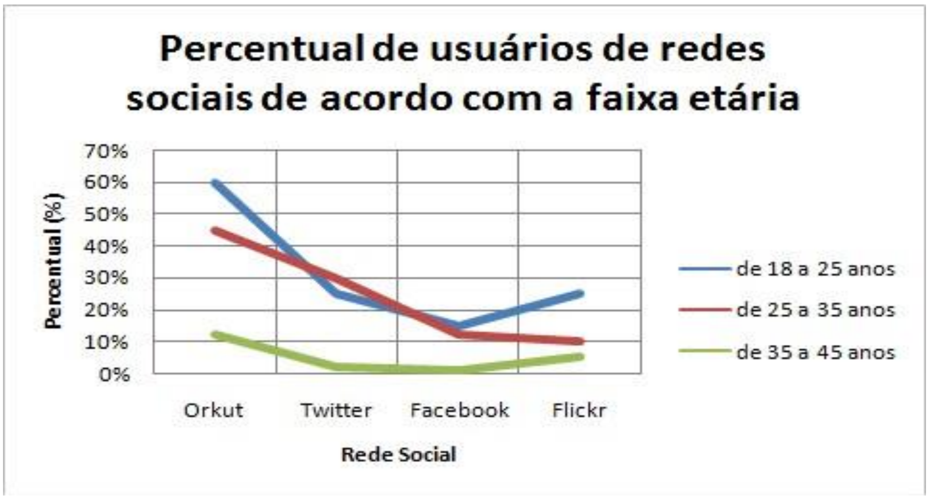
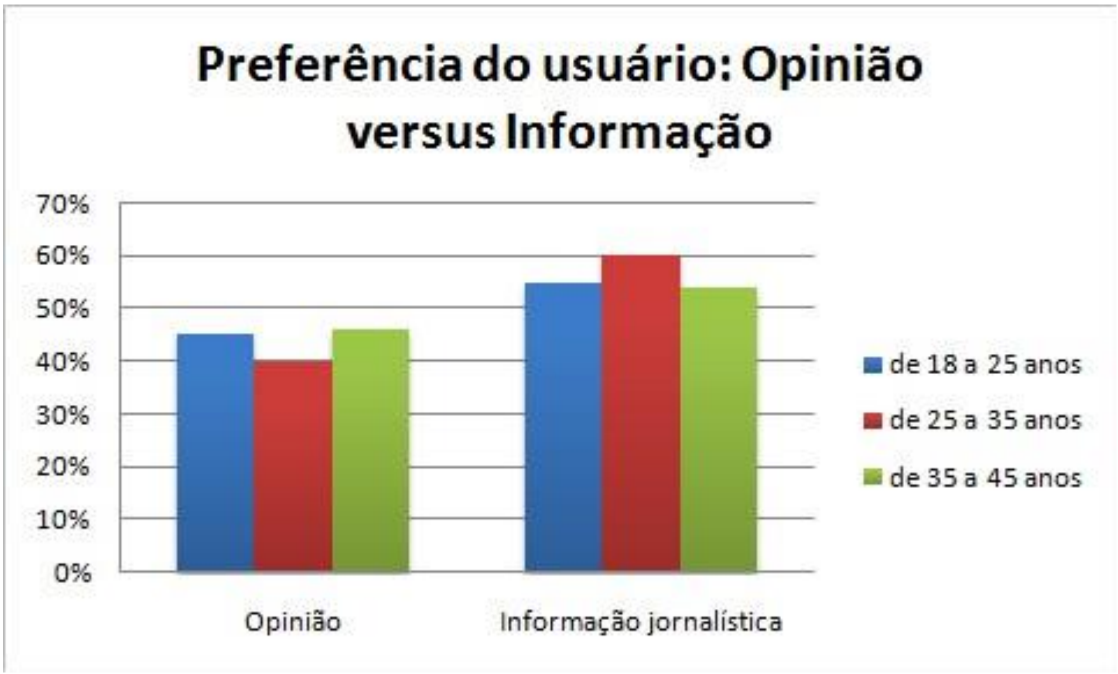



Figura 9- Gráfico em linha ou em curva - Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso em: 14 jul. 2019.

## 3.5.2 – GRÁFICO EM COLUNA OU EM BARRAS:

É a representação de uma série por meio de retângulos, dispostos verticalmente (em colunas) ou horizontalmente (em barras). Quando em colunas, os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados. E Quando em barras, os retângulos têm a mesma altura e os comprimentos são proporcionais aos respectivos dados.




Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso em: 24 jul. 2019.

## 2.4.3 – GRÁFICO EM SETORES:

Este gráfico é construído com base em um círculo, e é empregado sempre que deseja-se ressaltar a participação do dado no total. O total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes.

Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série. O total da série corresponde a 360º (total de graus de um arco de circunferência). O gráfico em setores representa valores absolutos ou porcentagens complementares.

As séries geográficas, específicas e as categorias em nível nominal são mais representadas em gráficos de setores, desde que não apresentem muitas parcelas (no máximo sete).

Para construir este gráfico, cada setor será expresso graficamente em graus (ângulo do setor) e a porcentagem calculada através de uma regra de três:

Total (%) $\rightarrow$ 360º
Parte (%) $\rightarrow$ $x$º


Figura 10- Gráfico em setores - Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso em: 24 jul. 2019.

## ATIVIDADES DE APRENDIZAGEM

1. Explique os tipos de gráficos e dê um exemplo de cada um deles.
2. Qual a principal vantagem de um gráfico em relação a uma tabela?
3. A representação gráfica deve obedecer a certos requisitos. Quais são eles?

## SUGESTÃO DE LEITURA

Para saber mais sobre a unidade que acabamos de estudar sugiro que pesquise:

- Acesse o site:

http://alea-estp.ine.pt
www.estatistica.ccet.ufrn.br

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e probabilidade.** 2.ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelando o Poder dos dados** [VitalSource Bookshelf version]. Retrieved from vbk://9788521633440

S., M. D., I., N. W., A., F. M. (07/2017). **A Estatística Básica e sua Prática**, 7ª edição [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287

## UNIDADE 4 – MEDIDAS DE TENDÊNCIA CENTRAL OU POSIÇÃO

### Objetivos

- Identificar as medidas de tendência central.
- Estabelecer uma comparação entre as medidas de posição identificando suas vantagens e desvantagens.
- Analisar exemplos que evidenciem o procedimento de obtenção das medidas de posição.

Nesta unidade, conheceremos as principais medidas de posição. Será nosso foco a obtenção dessas medidas através das distribuições de frequências que foram estudadas na unidade anterior. Também abordaremos as medidas separatrizes que são medidas que ocupam determinados lugares na distribuição de frequências.

## O QUE VAMOS ESTUDAR AGORA !!!!


## 4 - INTRODUÇÃO

Como definido na Unidade I, a Estatística busca “coletar, organizar, apresentar, analisar e interpretar dados numéricos”, desta forma torna-se necessário, após a tabulação dos resultados e da representação gráfica, encontrar valores que possam representar a distribuição como um todo.

As medidas de posição têm como finalidade representar o ponto de equilíbrio ou centro de uma distribuição. Na maioria das vezes, podem ser considerados valores típicos ou representativos do conjunto.

Vimos na **UNIDADE 3** que é possível agrupar os dados numéricos na forma de tabelas de distribuição de frequência, assim, temos três formas de representar um conjunto de dados, são elas: **Dados não agrupados, Dados agrupados sem intervalo de Classe** e **Dados agrupados com intervalo de Classe**. As medidas de tendência central que estudaremos agora serão abordadas considerando estas três formas de representação de dados.

## 4.1 - CONCEITO

As medidas de posição ou também conhecidas como medidas de tendência central compõem-se de um número que representa um conjunto particular de informações. Geralmente se localizam em torno do centro da distribuição, onde a maior parte das observações tende a concentra-se.

## 4.2 - MÉDIA ARITMÉTICA

É a medida mais conhecida e utilizada, isto se deve pela facilidade de cálculo e de compreensão aliadas às suas propriedades matemáticas. Existem dois tipos de média aritmética: simples ou ponderada.

### 4.2.1- MÉDIA SIMPLES

Consiste em somar todas as observações ou medidas dividindo-se o resultado pelo número total de valores.

$$\boldsymbol{\overline{x}} = \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}}$$

**Vejamos:**

**Número de observações**

Onde,

$\boldsymbol{x_i}$ (dados observados): são os dados medidos, valores medidos ou obtidos em uma pesquisa.

$\sum \boldsymbol{x_i}$ é a soma de todos valores medidos na pesquisa,

$\boldsymbol{n}$ é o número de dados medidos na pesquisa**.**

**Por Exemplo:** supondo que foram sorteados os seguinte números 2, 4, 6, 8 e 10. Temos

$\sum \boldsymbol{x_i} = \mathbf{2 + 4 + 6 + 8 + 10 = 30}$ e $\boldsymbol{n} = \mathbf{5}$ $\textit{que é a quantidade números sorteados}$

Uma das grandes paixões dos brasileiros é o futebol, e o Campeonato Brasileiro é a maior competição nacional, através da tabela abaixo é possível verificar o número de gols feitos nas últimas 5 edições deste Campeonato.

tabela 1- número de gols do campeonato brasileiro 2012-2016

| *Ano* | *Número de jogos* | *Número de Gols* |
|---|---|---|
| *2016 (em andamento)* | 150 | 387 |
| *2015* | 380 | 897 |
| *2014* | 380 | 860 |
| *2013* | 380 | 936 |
| *2012* | 380 | 940 |

Fonte: http://futpedia.globo.com/campeonato/campeonato-brasileiro. Adaptado. Acesso 14/07/2016.

A partir dos dados apresentados na tabela 1, podemos calcular a média de gols marcados por edição do Campeonato Brasileiro. Vejamos,

Para o ano de 2012 temos: $\bar{x} = \frac{\sum x_i}{n} = \frac{940}{380} \sim 2,47$

Podemos concluir que no ano de 2012 a média de gols nos campeonato Brasileiro foi de aproximadamente 2,47 gols por partida.

## 4.2.2 - MÉDIA PONDERADA

Uma média aritmética na qual é atribuído um peso a cada valor da série. Neste caso temos os dados agrupados sem intervalo de classe e com intervalo de classe.

**Vejamos**

$$\bar{x} = \sum \frac{f_i x_i}{n}$$

**Frequência Simples**

**Dados Observados**

**Número de observações**

**VEJAMOS ESTES EXEMPLOS:**

- **Dados agrupados sem intervalo de classe**

Um professor realizou uma entrevista com seus alunos a respeito da quantidade de irmãos que eles possuem, e o dados foram representados de acordo com a tabela de distribuição de frequência simples (tabela 2) abaixo:

tabela 2- distribuição de frequência simples

| Número de irmãos ($x_i$) | Número de alunos ($f_i$) |
|---|---|
| **0** | 7 |
| **1** | 21 |
| **2** | 8 |
| **3** | 5 |
| **4** | 4 |
| **5** | 3 |
| **6** | 2 |
| **Total (n)** | 50 |

Para calcularmos a média aplicamos a seguinte fórmula:

$$\bar{x} = \sum \frac{f_i x_i}{n}$$

A fim de facilitar nossos cálculos vamos adicionar à tabela 2 mais uma coluna que vai representar o produto - $f_i x_i$

| Número de irmãos (xi) | Número de alunos ($f_i$) | $f_i x_i$ |
|---|---|---|
| 0 | 7 | 0 |
| 1 | 21 | 21 |
| 2 | 8 | 16 |
| 3 | 5 | 15 |
| 4 | 4 | 16 |
| 5 | 3 | 15 |
| 6 | 2 | 12 |
| Total (n) | 50 | $\sum f_i x_i = 95$ |

Agora é só substituir na fórmula os valores obtidos em seus respectivos lugares:

$$\bar{x} = \sum \frac{f_i x_i}{n} \frac{95}{50} = 1,9$$

**Podemos então concluir que em média cada aluno tem 1,9 irmãos.**

- **Dados agrupados com intervalo de classes**

Um professor de estatística registrou a idade seus alunos em dias e os dados foram representados de acordo com a tabela de distribuição de frequência com intervalo de classe (tabela 3) abaixo:

TABELA 3- DISTRIBUIÇÃO DE FREQUÊNCIA COM INTERVALO DE CLASSE

| Idades | Número de alunos |
|---|---|
| **230 ⊢ 250** | 12 |
| **250 ⊢ 270** | 9 |
| **270 ⊢ 290** | 8 |
| **290 ⊢ 310** | 7 |
| **310 ⊢ 330** | 6 |
| **330 ⊢ 350** | 5 |
| **350 ⊢ 370** | 3 |
| Total | 50 |

$$\bar{x} = \sum \frac{f_i X_i}{n}$$

A fim de facilitar nossos cálculos vamos adicionar à tabela 3 mais uma coluna que vai representar o ponto médio do intervalo - $x_i$

| Idades | Número de alunos | $x_i$ |
|---|---|---|
| **230 ⊢ 250** | 12 | 240 |
| **250 ⊢ 270** | 9 | 260 |
| **270 ⊢ 290** | 8 | 280 |
| **290 ⊢ 310** | 7 | 300 |
| **310 ⊢ 330** | 6 | 320 |
| **330 ⊢ 350** | 5 | 340 |
| **350 ⊢ 370** | 3 | 360 |
| Total | 50 | ------ |

Agora vamos adicionar mais uma coluna referente ao produto **-** $f_i x_i$

| Idades | Número de alunos | $x_i$ | $f_i x_i$ |
|---|---|---|---|
| **230 ⊢ 250** | 12 | 240 | 2880 |
| **250 ⊢ 270** | 9 | 260 | 2340 |
| **270 ⊢ 290** | 8 | 280 | 2240 |
| **290 ⊢ 310** | 7 | 300 | 2100 |
| **310 ⊢ 330** | 6 | 320 | 1920 |
| **330 ⊢ 350** | 5 | 340 | 1700 |
| **350 ⊢ 370** | 3 | 360 | 1080 |
| Total | 50 | ------ | $\sum f_i x_i$ =14260 |

O próximo passo é substituir na fórmula os valores obtidos em seus respectivos lugares:

$\overline{x} = \sum \frac{f_i x_i}{n} \frac{14260}{50} = 285,20$ **Ou seja, 285 meses e 6 dias.**

## 4.3 - MEDIANA

É o valor central em um rol, ou seja, a mediana de um conjunto de valores ordenados, ou ainda a mediana divide a distribuição ao meio.

### 4.3.1 - MEDIANA DE VALORES BRUTOS

Ordenar os valores em ordem crescente (Rol);

Verifica se o número de elementos é par ou ímpar;

Se **n** for ímpar, posição da mediana no conjunto, será o valor localizado na posição dada por $P = \frac{n+1}{2}$

**Veja como faz!!!!**

Calcule a mediana dos valores: 2 ; 5; 7; 15; 13; 4; 10.

*Rol*: 2; 4; 5; 7; 10; 13; 15.

*n* = 7 (ímpar)

Posição da mediana: ***P* = (7 + 1) / 2 = 4**

***Me* = 7**

Se **n** for par, o conjunto terá dois valores centrais, neste caso, a mediana será igual à média aritmética dos valores centrais, cujas posições são dadas por:

P1 = n / 2 e P2 = (n / 2) + 1

**Veja como faz!!!!**

Em um grupo de 6 pessoas cujas as alturas medidas em centímetros fossem as seguintes:

183 cm, 170 cm, 165 cm, 180 cm, 185 e 160 cm, qual a altura mediana deste grupo de pessoas?

*Rol*: 160; 165; 170; 180; 183; 185.

*n* = 6 (par)

Posição da mediana:

$$P_1 = \left(\frac{6}{2}\right) = 3$$

$$P_2 = \left(\frac{6}{2}\right) + 1 = 4$$

A mediana será a média aritmética das posições *P*1 e *P*2 , então:

$$Me = \frac{170 + 180}{2} = 175$$

## 4.3.2 - MEDIANA DE DADOS AGRUPADOS SEM INTERVALO DE CLASSE

Neste caso calculamos a mediana a partir da posição mediana.

Vamos calcular a mediana da tabela 2, aqui representada novamente:

**TABELA 2- DISTRIBUIÇÃO DE FREQUÊNCIA SIMPLES**

| Número de irmãos ($x_i$) | Número de alunos ($f_i$) |
|---|---|
| 0 | 7 |
| 1 | 21 |
| 2 | 8 |
| 3 | 5 |
| 4 | 4 |
| 5 | 3 |
| 6 | 2 |
| Total (n) | 50 |

Primeiro passo, é determinar a posição mediana ($P_{me}$), que por sua vez é calculada da seguinte forma:

$$P_{me} = \frac{\sum f_i}{2} = \frac{n}{2} = \frac{50}{2} = 25$$

A classe que tiver a frequência acumulada que engloba o valor encontrado é a classe mediana. Então vamos calcular a frequência acumulada da tabela 4:

| **Número de irmãos ($x_i$)** | **Número de alunos ($f_i$)** | **Frequência acumulada** |
|---|---|---|
| 0 | 7 | 7 |
| 1 | 21 | 28 |
| 2 | 8 | 36 |
| 3 | 5 | 41 |
| 4 | 4 | 45 |
| 5 | 3 | 48 |
| 6 | 2 | 50 |
| Total (n) | 50 | ----- |

A mediana é a frequência simples da classe que contém a frequência acumulada calculada pela fórmula anterior.

Logo, a mediana para os dados agrupados sem intervalos de classe é: Me = 21

## 4.3.3 - MEDIANA DE DADOS AGRUPADOS COM INTERVALO DE CLASSE

**Amplitude do intervalo da classe mediana**

**Posição Mediana**

$$\boldsymbol{Me = l_i + h_i\left[\frac{\frac{n}{2} - F_{aa}}{f_i}\right]}$$

**Frequência Acumulada da classe anterior**

**Limite inferior da classe mediana**

**Frequência Simples da classe mediana**

Você deve estar se perguntando:


Vejamos o :

# PASSO A PASSO!!!

**1º passo** calcula-se a posição mediana: $P_{md} = \frac{n}{2}$

**2º passo**: identifica-se a classe *Mediana* pela coluna das Frequências Acumuladas;

**3º passo:** Aplica-se a fórmula:

$$\boldsymbol{Me = l_i + h_i \left[ \frac{\frac{n}{2} - F_{aa}}{f_i} \right]}$$

***Como fica na prática?***

Para entendermos melhor vamos utilizar os dados da tabela 3:

**TABELA 3- DISTRIBUIÇÃO DE FREQUÊNCIA COM INTERVALO DE CLASSE**

| Idades | Número de alunos ($f_i$) | $F_i$ |
|---|---|---|
| **230 ⊢ 250** | 12 | 12 |
| **250 ⊢ 270** | 9 | 21 |
| **270 ⊢ 290** | 8 | 29 |
| **290 ⊢ 310** | 7 | 36 |
| **310 ⊢ 330** | 6 | 42 |
| **330 ⊢ 350** | 5 | 47 |
| **350 ⊢ 370** | 3 | 50 |
| Total | 50 | |

**1º passo** calcula-se a posição mediana: $P_{md} = \frac{n}{2} = \frac{50}{2} = 25$

**2º passo**: identifica-se a classe *Mediana* pela coluna das Frequências Acumuladas;

| Idades | Número de alunos ($f_i$) | $F_i$ |
|---|---|---|
| **230 ⊢ 250** | 12 | 12 |
| **250 ⊢ 270** | 9 | 21 |
| **270 ⊢ 290** | **8** | **29** |
| **290 ⊢ 310** | 7 | 36 |
| **310 ⊢ 330** | 6 | 42 |
| **330 ⊢ 350** | 5 | 47 |
| **350 ⊢ 370** | 3 | 50 |
| Total | 50 | |

**Classe Mediana**

**3º passo:** Aplica-se a fórmula:

$$Me = l_i + h_i \left[ \frac{\frac{n}{2} - F_{aa}}{f_i} \right]$$

$$Me = 270 + 20 \left[ \frac{25 - 21}{8} \right]$$

$$Me = 270 + 20 \left[ \frac{4}{8} \right]$$

$$Me = 270 + 20 \left[ \frac{1}{2} \right]$$

$$Me = 270 + 10$$

$$Me = 280$$

Desta forma, a mediana desta distribuição é igual a 280.

**A mediana é muito empregada em pesquisas em que não interessam valores extremos, por terem pouca significação para o conjunto em geral.**

## 4.4 – MODA

É aquilo que está em evidência, o valor que mais aparece num conjunto de informações ou o de maior frequência em uma tabela. É a única medida que pode não existir e, existindo, pode não ser única.

### 4.4.1 – MODA DE VALORES BRUTOS

Basta observar o valor que mais aparece no conjunto.

**1º Exemplo:** 3 ; 3 ; 6 ; 8 ; 10 ; 10; 10; 11; 11; 12

*Mo* = 10.

**2º Exemplo:** Vamos analisar os seguintes valores: 0 3 3 5 8 8 8 9 13 13 13 14

Observando os valores acima podemos perceber que o valor 8 e 13 possuem o mesmo número de ocorrência (3 vezes), assim, **Mo = 8 e 13.** Neste caso, temos uma distribuição **BIMODAL.** Isso, pq temos dois números que aparecem mais de uma vez.

**3º Exemplo:** Agora vejamos a seguinte situação: 1 4 7 12 14 19 23 25 33

Observando os valores acima podemos perceber a distribuição não possui elementos com repetição, logo, tem uma distribuição **AMODAL.** Isso quer dizer que nenhum número se repete.

### 4.4.2 – MODA DE VALORES AGRUPADOS SEM INTERVALO DE CLASSE

Vejamos como fica a moda para os dados da tabela 2:

| Número de irmãos ($x_i$) | Número de alunos ($f_i$) |
|---|---|
| **0** | 7 |
| **1** | **21** |
| **2** | 8 |
| **3** | 5 |
| **4** | 4 |
| **5** | 3 |
| **6** | 2 |
| **Total (n)** | 50 |

**Maior frequência**

Assim a **Mo = 21**

## 4.4.3 – MODA DE VALORES AGRUPADOS COM INTERVALO DE CLASSE

Numa distribuição de frequência chamamos classe modal à classe que possui maior frequência. Como o ponto médio é representativo de qualquer classe de frequências, chamamos moda bruta ao ponto médio da classe modal.

### 4.4.3.1- MODA BRUTA

Para os dados apresentados na tabela 3 temos:

| Idades | Número de alunos ($f_i$) |
|---|---|
| **230 ⊢ 250** | 12 |
| **250 ⊢ 270** | 9 |
| **270 ⊢ 290** | 8 |
| **290 ⊢ 310** | 7 |
| **310 ⊢ 330** | 6 |
| **330 ⊢ 350** | 5 |
| **350 ⊢ 370** | 3 |
| Total | 50 |

A classe modal é a que possui maior frequência simples, que para este exemplo é a primeira classe. Assim,

$$Mo = \frac{230 + 250}{2} = \frac{480}{2} = 240$$

Portanto, a moda bruta é igual a: **Mo = 240**

## 4.4.3.2 - MODA PELO PROCESSO DE KING

Exitem alguns processos mais detalhados para calcularmos a moda para uma distribuição de frequência com intervalos, vamos estudar aqui o processo que chamamos de ***moda pelo processo de King***.

$$Mo = l_i + \left(\frac{f_{post}}{f_{ant} + f_{post}}\right) h_i$$

**Frequência simples da classe posterior à classe modal**

**Amplitude do intervalo da classe modal**

**Limite inferior da classe modal**

**Frequência simples da classe anterior à classe modal**

***moda pelo processo de King***..

**Pq esse nome?** Este nome é devido em homenagem ao estatístico, e como aqui não tratamos das deduções e parte histórica da estatísitica eu não coloquei.

**Para que serve?** Como citado no texto antes da fórmula, para processos mais detalhados.

**E quando deve ser udada?** Para Cálculos mais detalhados e minuciosos.

Se preferir eu faço toda uma introdução teórica e dedutiva, para detalhar este tema, porém acredito que os alunos irão se perder em meio a tantos cálculos e deuduções.

***Como fica na prática?***

Vejamos como fica a moda para a tabela de notas:

| Notas | Número de alunos ($f_i$) |
|---|---|
| **1 ⊢ 3** | 10 |
| **3 ⊢ 5** | 12 |
| **5 ⊢ 7** | **18** |
| **7 ⊢ 9** | 10 |
| Total(n) | 50 |

Aplicando a fórmula de King, temos:

$$Mo = 5 + \left(\frac{10}{12+10}\right)2$$

$$Mo = 5 + \left(\frac{10}{22}\right)2$$

$$Mo = 5 + (0{,}45)2$$

$$Mo = 5 + 0{,}9$$

$$Mo = 5{,}9$$

## QUADRO COMPARATIVO

| MÉDIA | MEDIANA | MODA |
|---|---|---|
| **Objetivo**: Usada para operações estatísticas avançadas | **Objetivo:** Eventualmente pode ser usada para operações estatísticas mais avançadas ou para separar distribuições em duas categorias | **Objetivo:** Medida de tendência central rápida simples, mas um tanto grosseira. |
| **Vantagens**<br>➢ No cálculo a média participa de todos os valores observados.<br>➢ É uma medida de fácil interpretação e presta-se muito bem a tratamentos estatísticos adicionais.<br>➢ É uma medida que sempre existe e é rigorosa e unicamente determinada.<br>➢ É um valor típico de um conjunto de dados podendo substituir todos os valores de um conjunto sem alterar o total.<br>➢ É o ponto de equilíbrio de uma distribuição, sendo tão mais eficiente quanto mais simétrica for a distribuição dos valores ao seu redor. | **Vantagens**<br>➢ Define exatamente o centro de uma distribuição, mesmo quando os valores se distribuem assimetricamente em torno da média.<br>➢ Pode ser determinada mesmo quando não se conhece todos os valores do conjunto de dados.<br>➢ É uma medida que sempre existe e é única.<br>➢ Esta medida pode ser utilizada para definir o meio de um número de objetos propriedades ou quantidades que possam de alguma forma ser ordenados.<br>➢ É uma medida resistente, ou seja, não sofre influência de valores discrepantes. | **Vantagens**<br>➢ É uma medida que têm existência real dentro do conjunto de dados e em grande número de vezes.<br>➢ Não exige cálculo, apenas uma contagem.<br>➢ Pode ser determinada também para variáveis qualitativas nominais. |
| **Desvantagens** | **Desvantagem**<br>✓ É uma medida que não se presta a cálculos matemáticos | **Desvantagens** |

| | | |
|---|---|---|
| ✓ É uma medida altamente influenciada por valores discrepantes.<br>✓ É afetada pelos valores extremos | | ✓ É uma medida que não se presta a cálculos matemáticos.<br>✓ Deixa sem representação todos os valores do conjunto de dados que não forem iguais a ela. |

## Sugestão de Leitura

http://alea-estp.ine.pt

http://www.somatematica.com.br/estatistica.php

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

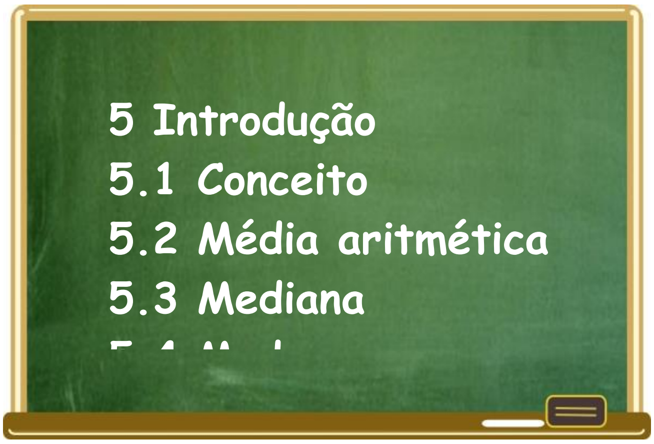
## UNIDADE 5 – MEDIDAS DE TENDÊNCIA CENTRAL OU POSIÇÃO

### Objetivos

- Identificar as medidas de tendência central
- Estabelecer uma comparação entre as medidas de posição identificando suas vantagens e desvantagens.
- Analisar exemplos que evidenciem o procedimento de obtenção das medidas de posição.

Nesta unidade, conheceremos as principais medidas de posição. Será nosso foco a obtenção dessas medidas através das distribuições de frequências que foram estudadas na unidade anterior. Também abordaremos as medidas separatrizes que são medidas que ocupam determinados lugares na distribuição de frequências.

## O QUE VAMOS ESTUDAR AGORA !!!!




## 5 - INTRODUÇÃO

Dados estatísticos são produzidos para podermos fazer alguma inferência sobre o tema abordado na pesquisa, nas fases do método estatístico que estudamos na Unidade II, vimos que os dados devem ser analisados e interpretados.

As medidas de posição, são utlilizadas com a finalidade de auxiliar na análise e interpretação dos dados. É importante saber que as medidas de posição, são apenas uma das ferramentas a serem utilizadas e que, sozinhas não produzem resultados completos.

Nesta unidade, aprenderemos sobre os principais tipos de medidas de posição e como devemos calcular seu valores.

## 5 - 1CONCEITO

As medidas de posição ou também conhecidas como medidas de tendência central compõem-se de um número que representa um conjunto particular de informações. Geralmente se localizam em torno do centro da distribuição, onde a maior parte das observações tende a concentra-se.

Para podemos obter as medidas de tendência central é importante organizarmos os dados, ou seja, trabalhamos com os dados em ROL.

## 5.2 - MÉDIA ARITMÉTICA

É a medida mais conhecida e utilizada, isto se deve pela facilidade de cálculo e de compreensão aliadas às suas propriedades matemáticas. Existem dois tipos de média aritmética: simples ou ponderada.

### 5.2.1- MÉDIA SIMPLES

Consiste em somar todas as observações ou medidas dividindo-se o resultado pelo número total de valores.

Dados observados

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \cdots x_n}{n} = \frac{\sum x_i}{n}$$

Número de observações

**EXEMPLO**

Uma das grandes paixões dos brasileiros é o futebol, e o Campeonato Brasileiro é a maior competição nacional. Através da tabela abaixo, é possível verificar o número de gols feitos nas últimas 5 edições deste Campeonato.

**TABELA 1- NÚMERO DE GOLS DO CAMPEONATO BRASILEIRO 2012-2016**

| *Ano* | *Número de jogos* | *Número de Gols* |
|---|---|---|
| *2016* | 150 | 387 |
| *2015* | 380 | 897 |
| *2014* | 380 | 860 |
| *2013* | 380 | 936 |
| *2012* | 380 | 940 |

Fonte: http://futpedia.globo.com/campeonato/campeonato-brasileiro. Adaptado. Acesso 15/07/2019.

A partir dos dados apresentados na tabela 1, podemos calcular a média de gols marcados por edição do Campeonato Brasileiro. Vejamos,

Para o ano de 2012 temos: $\overline{x} = \frac{\sum x_i}{n} = \frac{940}{380} \sim 2,47$

Podemos concluir que no ano de 2012, a média de gols no campeonato Brasileiro foi de aproximadamente 2,47 gols por partida.

## 5.2.2 - MÉDIA PONDERADA

Uma média aritmética na qual é atribuído um peso a cada valor da série. Neste caso, temos os dados agrupados sem intervalo de classe e com intervalo de classe.

$$\overline{x} = \sum \frac{f_i x_i}{n}$$

**Frequência Simples**

**Dados Observados**

**Número de observações**

**Agora vamos ver como a mesma fórmula é usada para dados agrupados de forma diferente. Para isso vamos analisar duas situações práticas.**

- **Dados agrupados sem intervalo de classe**

EXEMPLO

Os valores abaixo referem-se ao tempo (em dias) de cicatrização de cortes provenientes de cirurgia de 30 pacientes.

| Tempo de cicatrização (dias) | $f_i$ |
|---|---|
| 14,0 | 5 |
| 15,0 | 4 |
| 16,0 | 6 |
| 17,0 | 9 |
| 18,0 | 6 |
| **Total** | 30 |

Fonte: Fictícia

Para calcularmos a média aplicamos a seguinte fórmula:

$$\bar{x} = \sum \frac{f_i x_i}{n}$$

Onde **xi** é tempo de cicatrização e **fi** a frequência simples absoluta.

Para agilizar nossos cálculos vamos adicionar à tabela 2 mais uma coluna que vai representar o produto - $f_i x_i$

| Tempo de cicatrização (dias) | $f_i$ | $f_i x_i$ |
|---|---|---|
| 14,0 | 5 | 70 |
| 15,0 | 4 | 60 |
| 16,0 | 6 | 96 |
| 17,0 | 9 | 153 |
| 18,0 | 6 | 108 |
| **Total** | 30 | 487 |

Agora é só substituir na fórmula os valores obtidos em seus respectivos lugares:

$$\bar{x} = \sum \frac{f_i x_i}{n} \frac{487}{30} \sim 16,2$$

**Podemos concluir que em média são necessários 16,2 dias para a cicatrização de cortes cirúrgicos estejam completamente curados.**

- **Dados agrupados com intervalo de classes**

EXEMPLO

Em uma investigação dos fatores de risco para as doenças cardiovasculares, os níveis séricos de cotinina (produto metabólico da nicotina) foram registrados para um grupo de fumantes. Complete a tabela abaixo referente à distribuição de frequências correspondentes à pesquisa.

Tabela - níveis séricos de cotinina (produto metabólico da nicotina) para um grupo de fumantes

| Nível de cotinina (mg/ml) | Fumantes |
|---|---|
| 0 ⊢ 50 | 211 |
| 50 ⊢ 100 | 142 |
| 100 ⊢ 150 | 206 |
| 150 ⊢ 200 | 197 |
| 200 ⊢ 250 | 220 |
| 250 ⊢ 300 | 151 |
| 300 ⊢ 350 | 412 |
| Total | 1539 |

Fonte: http://people.ufpr.br/~jomarc/exerciciosestatistica2.pdf- Adaptado. Acesso em 22/07/2019

Para calcularmos a média aplicamos a seguinte fórmula:

$$\bar{x} = \sum \frac{f_i x_i}{n}$$

A primeira coisa a se fazer é calcular o ponto médio de cada intervalo. Caso, tenha dúvidas sobre esse assunto, sugiro que retorno a Unidade IV, onde este assunto foi tratado. Se sua dúvida, persistir, posto no **Fórum Tira Dúvidas** para auxilar você.Mas, vamos continuar com o nosso cálculo.

| Nível de cotinina (mg/ml) | Fumantes | $x_i$ |
|---|---|---|
| 0 ⊢ 50 | 211 | 25 |
| 50 ⊢ 100 | 142 | 75 |
| 100 ⊢ 150 | 206 | 125 |
| 150 ⊢ 200 | 197 | 175 |
| 200 ⊢ 250 | 220 | 225 |
| 250 ⊢ 300 | 151 | 275 |
| 300 ⊢ 350 | 412 | 325 |
| Total | 1539 | |

Para agilizar nossos cálculos vamos adicionar à tabela 2 mais uma coluna que vai representar o produto - $f_i x_i$

| Nível de cotinia (mg/ml | Fumantes | $x_i$ | $f_i x_i$ |
|---|---|---|---|
| **0 ⊢ 50** | 211 | 25 | 5275 |
| **50 ⊢ 100** | 142 | 75 | 10650 |
| **100 ⊢ 150** | 206 | 125 | 25750 |
| **150 ⊢ 200** | 197 | 175 | 34475 |
| **200 ⊢ 250** | 220 | 225 | 49500 |
| **250 ⊢ 300** | 412 | 275 | 113300 |
| **300 ⊢ 350** | 151 | 325 | 49075 |
| **Total** | 1539 | | 288025 |

Agora é só substituir na fórmula os valores obtidos em seus respectivos lugares:

**Podemos concluir que os indivíduos possuem em média 187,15 mg/ml de cotinina no organismos.**

$$\bar{x} = \sum \frac{f_i x_i}{n} \frac{288025}{1539} \sim 187,15$$

## 5.3 - MEDIANA

É o valor central em um rol, ou seja, a mediana de um conjunto de valores ordenados, ou ainda a mediana divide a distribuição ao meio.

## 5.3.1 - MEDIANA DE VALORES BRUTOS

Ordenar os valores em ordem crescente (Rol);

Verifica se o número de elementos é par ou ímpar;
Se **n** for ímpar, posição da mediana no conjunto, será o valor localizado na posição dada por $P = \frac{n+1}{2}$

**Veja como faz!!!!**

EXEMPLO

Abaixo estão listados os tamanhos dos tórax (em polegadas) e os pesos (em libras) de ursos selecionados aleatoriamente e que foram anestesiados e medidos, com base em dados de Gary Alt and Minitab, Inc.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Tórax** | 26 | 45 | 54 | 49 | 35 | 41 | 41 | 49 | 38 | 32 |
| **Peso** | 80 | 344 | 416 | 348 | 166 | 220 | 262 | 360 | 204 | 140 |

Primeiramente vamos colocar os dados em ROL, ou seja, vamos coloca-los em ordem crescente.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Tórax** | 26 | 32 | 35 | 38 | 41 | 41 | 45 | 49 | 49 | 54 |
| **Peso** | 80 | 140 | 166 | 204 | 220 | 262 | 344 | 348 | 360 | 416 |

O total de medidas de Tórax e Peso é par então, calculamos a mediana da seguinte forma

*n* = 10 (par)

***Tórax :*** $\boldsymbol{Me = \frac{41+41}{2} = \frac{82}{2} = 41}$

***Peso:*** $\boldsymbol{Me = \frac{220+262}{2} = \frac{482}{2} = 241}$

**Se n for par, o conjunto terá dois valores centrais, neste caso, a mediana será igual à média aritmética dos valores centrais**

Agora vamos supor que temos um número ímpar de medidas. Para isso vamos eliminar aleatoriamente uma medida para o peso e uma para o tórax.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Tórax** | 32 | 35 | 38 | 41 | 41 | 45 | 49 | 49 | 54 |
| **Peso** | 80 | 140 | 166 | 204 | 220 | 262 | 344 | 348 | 416 |

Neste caso como o total de medidas é ímpar, a mediana é o valor do meio da distribuição, assim temos:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Tórax** | 32 | 35 | 38 | 41 | 41 | 45 | 49 | 49 | 54 |
| **Peso** | 80 | 140 | 166 | 204 | 220 | 262 | 344 | 348 | 416 |

***Tórax :*** $\boldsymbol{Me = 41}$

***Peso:*** $\boldsymbol{Me = 220}$

## 5.3.2 - MEDIANA DE DADOS AGRUPADOS SEM INTERVALO DE CLASSE

Vamos utilizar o exemplo dado para o cálculo da média para dados sem intervalo de classe.

| Tempo de cicatrização (dias) | $f_i$ |
|---|---|
| 14,0 | 5 |
| 15,0 | 4 |
| 16,0 | 6 |
| 17,0 | 9 |
| 18,0 | 6 |
| **Total** | 30 |

Primeiro passo é determinar a posição mediana ($P_{me}$), que por sua vez é calculada da seguinte forma:

$$P_{me} = \frac{\sum f_i}{2} = \frac{n}{2} = \frac{30}{2} = 15$$

A posição mediana determina a posição da medida que representa a mediana, para encontrar este valor, devemos calcular primeiro a frequência acumulada da tabela.

| Tempo de cicatrização (dias) | $f_i$ | $F_a$ |
|---|---|---|
| 14,0 | 5 | 5 |
| 15,0 | 4 | 9 |
| 16,0 | 6 | 15 |
| 17,0 | 9 | 24 |
| 18,0 | 6 | 30 |
| **Total** | 30 | |

Em seguida obsevamos na coluna da freuqencia acumulada onde ocorre pela primeira vez o valor encontrado para a posição mediana.

| Tempo de cicatrização (dias) | $f_i$ | $F_a$ |
|---|---|---|
| 14,0 | 5 | 5 |
| 15,0 | 4 | 9 |
| 16,0 | 6 | 15 |
| 17,0 | 9 | 24 |
| 18,0 | 6 | 30 |
| **Total** | 30 | |

**Classe que contém a posição mediana, também conhecida como classe mediana**

A mediana é a medida observada na classe que contém na frequência acumulada o valor da posição mediana

**Logo, a mediana para os dados agrupados sem intervalos de classe é: Me = 16 dias.**

## 5.3.3 - MEDIANA DE DADOS AGRUPADOS COM INTERVALO DE CLASSE

**Amplitude do intervalo da classe mediana**

**Posição Mediana**

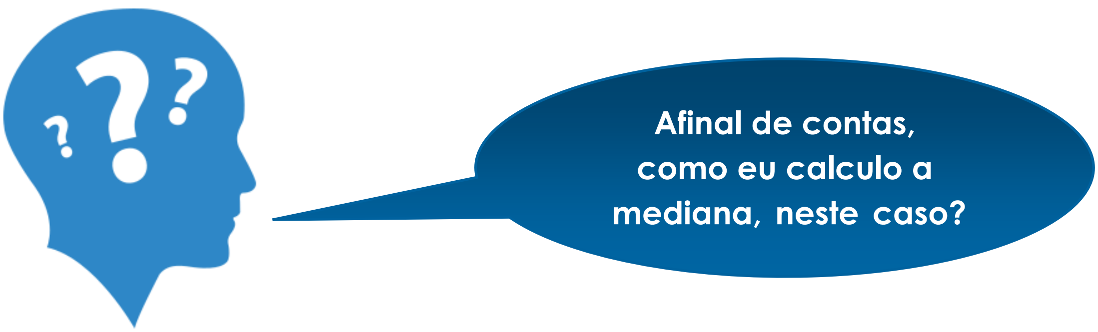
**Frequência Acumulada da classe anterior**

$$Me = l_i + h_i\left[\frac{\frac{n}{2} - F_{aa}}{f_i}\right]$$

**Limite inferior da classe mediana**

**Frequência Simples da classe mediana**

Você deve estar se perguntando:




Vejamos o :

# PASSO A PASSO!!!

**1º passo** calcula-se a posição mediana: $P_{md} = \frac{n}{2}$

**2º passo**: identifica-se a classe *Mediana* pela coluna das Frequências Acumuladas;

**3º passo:** Aplica-se a fórmula:

$$Me = l_i + h_i \left[ \frac{\frac{n}{2} - F_{aa}}{f_i} \right]$$

***Como fica na prática?***

EXEMPLO

Para entendermos melhor vamos considerar o exemplo para média de dados agrupados com intervalo de classe.

| Nível de cotinina (mg/ml) | Fumantes | $F_a$ |
|---|---|---|
| **0 $\vdash$ 50** | 211 | 211 |
| **50 $\vdash$ 100** | 142 | 353 |
| **100 $\vdash$ 150** | 206 | 559 |
| **150 $\vdash$ 200** | 197 | 756 |
| **200 $\vdash$ 250** | 220 | 976 |
| **250 $\vdash$ 300** | 412 | 1388 |
| **300 $\vdash$ 350** | 151 | 1539 |
| **Total** | 1539 | |

**1º passo** calcula-se a posição mediana: $P_{md} = \frac{1539}{2} = 769{,}5$

**2º passo**: identifica-se a classe *Mediana* pela coluna das Frequências Acumuladas;

| Nível de cotinina (mg/ml) | Fumantes | $F_a$ |
|---|---|---|
| **0 $\vdash$ 50** | 211 | 211 |
| **50 $\vdash$ 100** | 142 | 353 |
| **100 $\vdash$ 150** | 206 | 559 |
| **150 $\vdash$ 200** | 197 | 756 |
| **200 $\vdash$ 250** | 220 | 976 |
| **250 $\vdash$ 300** | 412 | 1388 |
| **300 $\vdash$ 350** | 151 | 1539 |
| **Total** | 1539 | |

**Classe Mediana**

**3º passo:** Aplica-se a fórmula:

$$Me = l_i + h_i \left[ \frac{\frac{n}{2} - F_{aa}}{f_i} \right]$$

$$Me = 200 + 50 \left[ \frac{769{,}5 - 756}{220} \right]$$

$$Me = 200 + 50 \left[ \frac{13{,}5}{220} \right]$$

**Desta forma, a mediana desta distribuição é igual a 203,5 mg/ml.**

$$Me = 200 + 50[0,061]$$

$$Me = 200 + 3,05$$

$$Me = 203,5$$

**A mediana é muito empregada em pesquisas onde não interessam valores extremos, por terem pouca significação para o conjunto em geral.**

## 5.4 – MODA

É aquilo que está em evidência, o valor que mais aparece num conjunto de informações ou o de maior frequência em uma tabela. É a única medida que pode não existir e, existindo, pode não ser única.

### 5.4.1 – MODA DE DADOS SIMPLES

EXEMPLO

Abaixo estão listados os tamanhos dos tórax (em polegadas) e os pesos (em libras) de ursos selecionados aleatoriamente e que foram anestesiados e medidos, com base em dados de Gary Alt and Minitab, Inc.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Tórax** | 26 | 32 | 35 | 38 | 41 | 41 | 45 | 49 | 49 | 54 |
| **Peso** | 80 | 140 | 166 | 204 | 220 | 262 | 344 | 348 | 360 | 416 |

**Tórax:** *Mo* = 41 e 49 (os valores repetiram 2x enquanto os outros valores apareceram apenas 1x).

Quando uma distribuição possui dois valores com a mesma frequência de repetição dizemos que a distribuição é **bimodal**

**Peso:** A variável peso não possui medidas com repetição, portanto, a distribuição é **Amodal**.

## 5.4.2 – MODA DE VALORES AGRUPADOS SEM INTERVALO DE CLASSE

| Tempo de cicatrização (dias) | $f_i$ |
|---|---|
| 14,0 | 5 |
| 15,0 | 4 |
| 16,0 | 6 |
| 17,0 | 9 |
| 18,0 | 6 |
| **Total** | 30 |

Maior frequência

Quando estamos falando de dados agrupados sem intervalo de classe, **a moda é o valor que possui maior frequência simples**, então neste exemplo temo que a moda é:

**Mo = 17 dias, ou seja, a maioria dos pacientes investigados apresentaram cicatrização completa 17 dias após a cirurgia.**

## 5.4.3.1- MODA BRUTA

Numa distribuição de frequência chamamos classe modal à classe que possui maior frequência. Como o ponto médio é representativo de qualquer classe de frequências, podemos calcular a moda por dois processos distintos.

## 5.4.3 – MODA DE VALORES AGRUPADOS COM INTERVALO DE CLASSE

Para os dados apresentados na tabela abaixo temos:

| Nível de cotinina (mg/ml) | Fumantes |
|---|---|
| **0 ⊢ 50** | 211 |
| **50 ⊢ 100** | 142 |
| **100 ⊢ 150** | 206 |
| **150 ⊢ 200** | 197 |
| **200 ⊢ 250** | 220 |
| **250 ⊢ 300** | 412 |
| **300 ⊢ 350** | 151 |
| **Total** | 1539 |

Maior frequência

A classe modal é a que possui maior frequência simples, que para este exemplo é a primeira classe. Assim,

$$\boldsymbol{Mo} = \frac{\mathbf{250 + 300}}{\mathbf{2}} = \frac{\mathbf{550}}{\mathbf{2}} = \mathbf{225}$$

Portanto, a moda bruta é igual a: **Mo = 225 mg/ml**

## 5.4.3.2 - MODA PELO PROCESSO DE KING

Exitem alguns processos mais detalhados para calcularmos a moda para uma distribuição de frequência com intervalos, vamos estudar aqui o processo que chamamos de ***moda pelo processo de King***.

Frequência simples da classe posterior à classe modal

$$Mo = l_i + \left(\frac{f_{post}}{f_{ant} + f_{post}}\right) h_i$$

Amplitude do intervalo da classe modal

Limite inferior da classe modal

Frequência simples da classe anterior à classe modal

***Como fica na prática?***

EXEMPLO

Vejamos como fica a moda para a tabela abaixo

| Nível de cotinina (mg/ml) | Fumantes | $F_a$ |
|---|---|---|
| **0 ⊢ 50** | 211 | 211 |
| **50 ⊢ 100** | 142 | 353 |
| **100 ⊢ 150** | 206 | 559 |
| **150 ⊢ 200** | 197 | 756 |
| **200 ⊢ 250** | 220 | 976 |
| **250 ⊢ 300** | 412 | 1388 |
| **300 ⊢ 350** | 151 | 1539 |
| **Total** | 1539 | |

Aplicando a fórmula de King, temos:

$$Mo = 250 + \left(\frac{151}{220 + 151}\right)50$$

$$Mo = 250 + \left(\frac{151}{371}\right)50$$

$$Mo = 250 + (0{,}407)50$$

$$Mo = 250 + 20{,}35$$

$$Mo = 270{,}35$$

É interessante observar que a moda calculada por processo diferentes para uma mesma distribuição pode gerar valores diferentes, isso se deve a falta de precisão no resultado da moda calculada pelo processo bruto. Desta forma, a moda calculada pelo processo de King é mais precisa.

## QUADRO COMPARATIVO

| Medida | Objetivo | Vantagens | Desvantagens |
|---|---|---|---|
| **Média** | Usada para operações estatísticas avançadas | ➢ No cálculo a média participa de todos os valores observados.<br>➢ É uma medida de fácil interpretação e presta-se muito bem a tratamentos estatísticos adicionais.<br>➢ É uma medida que sempre existe e é rigorosa e unicamente determinada.<br>➢ É um valor típico de um conjunto de dados podendo substituir todos os valores de um conjunto sem alterar o total.<br>➢ É o ponto de equilíbrio de uma distribuição, sendo tão mais eficiente quanto mais simétrica for a distribuição dos valores ao seu redor. | É uma medida altamente influenciada por valores discrepantes.<br>É afetada pelos valores extremos. |
| **Mediana** | Eventualmente pode ser usada para operações estatísticas mais | ➢ Define exatamente o centro de uma distribuição, mesmo quando | É uma medida que se torna impraticável para conjuntos de dados muito grande. |

| | | | |
|---|---|---|---|
| | avançadas ou para separar distribuições em duas categorias | os valores se distribuem assimetricamente em torno da média.<br>➢ Pode ser determinada mesmo quando não se conhece todos os valores do conjunto de dados.<br>➢ É uma medida que sempre existe e é única.<br>➢ Esta medida pode ser utilizada para definir o meio de um número de objetos propriedades ou quantidades que possam de alguma forma ser ordenados.<br>➢ É uma medida resistente, ou seja, não sofre influência de valores discrepantes. | |
| **Moda** | Medida de tendência central rápida simples, mas um tanto grosseira. | ➢ É uma medida que têm existência real dentro do conjunto de dados e em grande número de vezes.<br>➢ Não exige cálculo, apenas uma contagem.<br>➢ Pode ser determinada também para variáveis qualitativas nominais. | É uma medida que não se presta a cálculos matemáticos, pois pode não existir para conjuntos de dados. |

## SUGESTÃO DE LEITURA

http://alea-estp.ine.pt

http://www.somatematica.com.br/estatistica.php

# REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

COSTA, Paulo, R. **Estatítica.** Disponível em <https://www.ufsm.br/unidades-universitarias/ctism/cte/wp content/uploads/sites/413/2018/11/04_estatistica.pdf. > Acesso em 18/07/2019

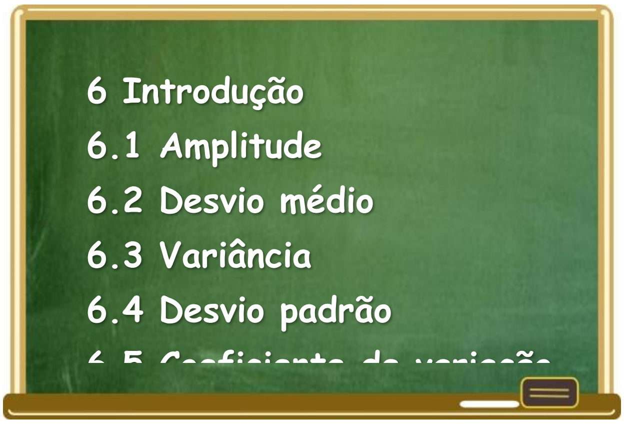
## UNIDADE 6 – MEDIDAS DE DISPERSÃO OU VARIABILIDADE

### Objetivos

- Identificar as medidas de dispersão; e,
- Analisar exemplos que evidenciem o procedimento de obtenção das medidas de dispersão.

Agora, estudaremos as medidas de dispersão ou variabilidade que tem um papel importantíssimo na análise dos dados, pois avaliam a variabilidade em torno da média, ou seja as variações que ocorrem com os dados em relação à média.

Sendo assim veremos ...




## 6 – INTRODUÇÃO

Uma análise estatística correta deve considerar o comportamento dos dados de forma total e precisa, assim as medidas de posição são insuficientes para fornecer informações que possam contribuir de forma significativa para um processo de inferência estatística.

Para podermos considerar os dados confiáveis para uma análise e inferência estatística é necessário que as medidas de posição estejam acompanhadas das medidas de dispersão ou variabilidade.

As medidas de dispersão servem para medir o grau de variabilidade ou dispersão dos valores observados em torno da média aritmética. Elas medem a representatividade da média e proporcionam o conhecimento do nível de homogeneidade ou heterogeneidade dentro de cada grupo analisado.

A dispersão ou variabilidade representa um dos mais importantes grupos de medidas da estatística. Para o conhecimento pleno e adequado de uma série ou de uma distribuição de frequências, é necessário determinar não apenas a posição central dos valores, através das medidas de posição, mas é preciso conhecer o real grau de afastamento de um conjunto de números em relação a sua média.

Agora vamos entender a dispersão do ponto de vista estatístico. A dispersão mede o quão próximo uns dos outros estão os valores de um conjunto de dados. Para isto vamos analisar a figura a seguir:

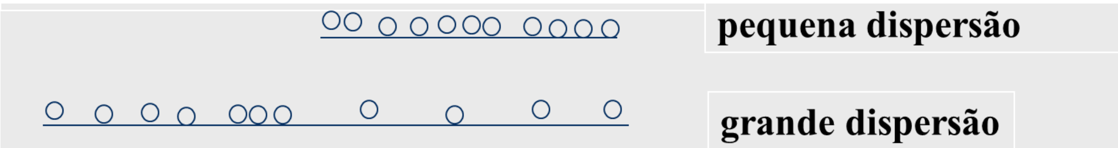



FIGURA 1- REPRESENTAÇÃO PICTÓRICA DA DISPERSÃO DOS CONJUNTOS DE DADOS A E B RESPECTIVAMENTE.

Assumindo que a representação pictórica feita na Fig. 1 seja correspondente aos seguintes dados, podemos perceber que embora a média seja igual para os dois conjuntos de dados, o conjunto B apresenta maior dispersão.

$$A = (25, 28, 31, 34, 37) \qquad B = (17, 23, 30, 39, 46)$$
$$\bar{x}_A = 31 \qquad \bar{x}_B = 31$$

Desta forma, fica claro que utilizar apenas uma medida de posição para representar um conjunto de dados não é muito seguro e confiável, logo podemos pensar que,

**Uma boa representação de dados** = **Uma medida de posição** (quase sempre a média) + **Uma medida de dispersão** (quase sempre o desvio padrão)

Nesta unidade, iremos discutir e aprender sobre as medidas de dispersão conhecidas como:

- Medidas de dispersão absoluta: Amplitude, Varância e Desvio Padrão; e,
- Medida de dispersão relativa: Coeficiente de Variação de Pearson.

## 6.1 – AMPLITUDE

Um modo mais simples de se ter uma indicação da dispersão dos valores de uma amostra ou população é comparar o valor máximo com o mínimo. Entretanto a Amplitude Total não nos fornece qualquer indicação do que ocorre no interior do conjunto.

Calculamos a Amplitude Total como,

**AT = Valor máximo – Valor mínimo**

Vamos entender melhor...

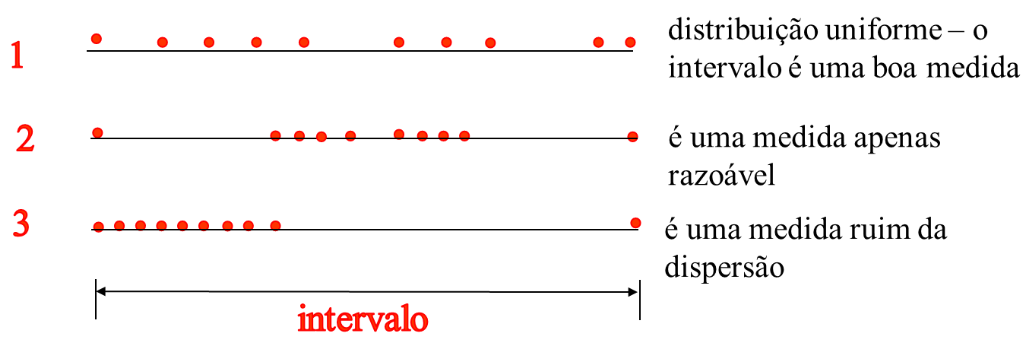



FIGURA 2- TIPOS DE DISPERSÃO

A **distribuição 1** é considerada uniforme, visto que os dados estão aproximadamente equidistantes entre si, ou seja, a distância entre eles é quase igual. Na **distribuição 2**, a maioria dos dados estão concentrados bem próximos um dos outros, entretanto o valor mínimo e máximo estão muito distantes, o que prejudica a dispersão. E por fim, a **distribuição 3** tem apenas um dado distante dos demais fazendo com que a dispersão seja ruim.

Podemos concluir que,

**A Amplitude total não é muito confiável, uma vez que utiliza apenas os valores dos extremos de uma distribuição. Também por esta razão é extremamente influenciada por valores discrepantes. É utilizada quando apenas uma ideia rudimentar da variabilidade dos dados é suficiente.**

## 6.2 – VARIÂNCIA

É a média quadrática das somas dos desvios em relação à média aritmética. É uma medida de dispersão bastante estudada no meio científico. Quando o estudo for feito na amostra a variância é simbolizada por: $S^2$. E quando estudamos a variância de uma população, o símbolo usado é $\sigma^2$.

### 6.2.1 – VARIÂNCIA PARA DADOS NÃO AGRUPADOS

**Amostra**

Valores observados

Média

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

**População**

Total de observações

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

Observem que há uma diferença no denominador das fórmulas, isso se deve ao fato de que, quando trabalhamos com amostra apenas uma parte da população é utilizada nos cálculos, desta forma, podemos dizer que o valor "1" representa 100%, ou seja, a população. E, quando trabalhamos com amostra deste 100% ou "1", são retirados "n" dados para serem investigados.

EXEMPLO

Os dados a seguir são os tempos (em minutos) que 8 competidores levaram para completar um circuito de *crossfit*:

55 58 46 58 49 46 41 60

Iniciamos calculando a média, para isto vamos utilizar o que aprendemos na Unidade V

$\bar{x} = \frac{\sum x_i}{n} = \frac{413}{8} = 51{,}62$

Para facilitar vamos montar uma tabela:

| $x_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| 55 | 3.38 | 11,42 |
| 58 | 6,28 | 39,44 |
| 46 | -5,62 | 31,58 |
| 58 | 6,38 | 40,70 |
| 49 | -2,62 | 6,86 |
| 46 | -5,62 | 31,58 |
| 41 | -10,62 | 112,78 |
| 60 | 8,38 | 70,22 |
| | | $\sum (x_i - \bar{x})^2 = \mathbf{344,58}$ |

Agora é só aplicar a fórmula,

**Para uma amostra**

$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{344{,}58}{7} \sim 49{,}22$

**Para uma população**

$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{344{,}58}{8} \sim 43{,}07$

Você deve estar se perguntando: ***Mas como eu vou saber se calculo para população, para amostra ou se devo calcular os dois?***

Então vamos esclarecer sua dúvida.

*Ao fazer um exercício leia atentamente o enunciado, nele deve conter de forma clara e objetiva se o problema se trata de um estudo feito sobre toda população ou em uma amostra. É comum os problemas mencionarem de forma clara a palavra amostra* *quando esta está sendo utilizada, quando se trata de população, a palavra "população" não é mencionda.*

*Assim, caso ao ler um exercício se a palavra amostra aparecer, vc deve calcular seus dados como amostra, se não for mencioada a palavra "amostra" o exercício deve ser resolvido como população.*

## 6.2.2 – VARIÂNCIA PARA DADOS AGRUPADOS

**Amostra**

**Valores observados** **Média**

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

**Frequência Simples** **Total de observações**

**População**

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n}$$

**EXEMPLO**

O conjunto de dados LongevidadeMamíferos inclui informação sobre longevidade (tempo de vida típico), em anos, para 40 espécies de mamíferos, bem como tempo de gestação desses mesmos mamíferos. Os dados relacionados com a longevidade são apresentados na Tabela 3. *Fonte: H., LOCK, R., LOCK, Frazer, MORGAN, Lock, LOCK, F., LOCK, F.. Estatística - Revelando o Poder dos dados. LTC, 01/2017. VitalBook file.*

**Tabela 3** – Longevidade de 40 espécies de maníferos

| Longevidade (anos) | $f_i$ |
|---|---|
| **1 ⊢ 9** | 10 |
| **9 ⊢ 17** | 23 |
| **17 ⊢ 25** | 2 |
| **25 ⊢ 33** | 2 |
| **33 ⊢ 41** | 3 |
| **Total** | 40 |

Fonte: Fictícia

Iniciamos calculando a média, para isto vamos utilizar o que aprendemos na Unidade V

$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{560}{40} = 14$

Para facilitar, vamos montar uma tabela:

| $x_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $f_i(x_i - \bar{x})^2$ |
|---|---|---|---|
| **5** | -9 | 81 | 810 |
| **13** | -1 | 1 | 23 |
| **21** | 7 | 49 | 98 |
| **29** | 15 | 225 | 550 |
| **37** | 23 | 529 | 1587 |
| **Total** | | | $\sum f_i(x_i - \bar{x})^2 = 3068$ |

Agora é só aplicar a fórmula,

Notem no enunciado que em momento algum foi citado a palavra “amostra”, portanto iremos realizar nossos cálculos considerando como população.

$$\sigma^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n} = \frac{3068}{40} \sim 76{,}7$$

**Propriedades da variância:**

A variância absoluta de uma constante é igual a zero;

Somando-se ou diminuindo-se a todos os valores da série um valor constante *K*≠ 0, a nova variância será igual à anterior, isto é, não se altera.

Multiplicando-se ou dividindo-se todos os valores de uma série por um valor constante, *K* ≠0, a nova variância calculada será igual à variância absoluta original multiplicada ou dividida pelo quadrado da constante utilizada.

**Desvantagens da variância:**

- Como a variância é calculada a partir da média, é uma medida pouco resistente, ou seja, muito influenciada por valores discrepantes.
- Como a unidade de medida fica elevada ao quadrado, a interpretação da variância se torna mais difícil.

## 6.3 – DESVIO PADRÃO

É a raiz quadrada da variância. É a medida mais informativa da variação dos dados. O Desvio Padrão nos fornece uma indicação do que ocorre entre os dois extremos. Portanto, o Desvio Padrão é a medida de quanto os valores observados variam em torno da média.

**O Desvio Padrão amostral é dado por:**

$$s = \sqrt{s^2}$$

**O desvio padrão populacional é dado por:**

$$\sigma = \sqrt{\sigma^2}$$

**EXEMPLO**

Utilizando a Variância da calculada anteriormente nas seções **6.2.1 e 6.2.2** temos:

**Cálculos da Seção 6.2.1**

**Para uma amostra**

$$s^2 = 49{,}22$$

$$s = \sqrt{49{,}22} \sim 7{,}02$$

**Para uma população**

$$\sigma^2 = 43{,}07 = \sqrt{43{,}07} \sim 6{,}56$$

**Cálculos da Seção 6.2.2**

**Para uma população**

$$\sigma^2 = 76{,}7$$

$$s = \sqrt{76{,}7} \sim 8{,}76$$

## 6.4 – COEFICIENTE DE VARIAÇÃO

O coeficiente de variação Pearson ou somente coeficiente de variação é a medida mais utilizada quando existe interesse em comparar variabilidades de diferentes conjuntos de dados. Embora esta comparação possa ser feita através de outras medidas de variação, nas situações em que as médias dos conjuntos comparados são muito desiguais ou as unidades de medidas são diferentes, devemos utilizar o coeficiente de variação.

O coeficiente de variação é definido como a proporção da média representada pelo desvio padrão e dado por:

**Para uma amostra**

$$CV = \frac{s}{\bar{x}} \times 100$$

**Para uma população**

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

**ATENÇÃO**

- **Será considerada a série mais homogênea, aquela que apresentar menor valor do coeficiente de variabilidade.**

- **É uma medida estatística que serve para avaliar a homogeneidade de séries estatísticas, que é o grau de concentração dos valores observados em torno da sua média aritmética.**

**EXEMPLO**

Utilizando valores já calculados nas seções **6.2.1 e 6.2.2 e 6.3** temos:

**Cálculos da Seção 6.2.1**

**Para uma amostra**

$\bar{x} = 51{,}62$

$s \sim 7{,}02$

$$CV = \frac{7{,}02}{51{,}62} \times 100 \sim 13{,}60\%$$

**Para uma população**

$\bar{x} = 51{,}62$

$s \sim 6{,}56$

$$CV = \frac{6{,}56}{51{,}62} \times 100 \sim 12{,}71\%$$

### Cálculos da Seção 6.2.2

**Para uma população**

$$\bar{x} = 14$$

$$s \sim 8{,}76$$

$$CV = \frac{8{,}76}{14} \times 100 \sim 62{,}67\%$$

Observação

O coeficiente de variação de Pearson ou apenas coeficiente de variação (CV), geralmente é expresso em percentagem. Alguns analistas consideram:

- Baixa dispersão – CV < 15%
- Média dispersão – 15% < CV < 30%
- Alta dispersão – CV > 30%

Um coeficiente de variação maior ou igual a 30% revela que a série é heterogênea e a média tem pouco significado. Se o coeficiente de variação for menor que 30%, a série será homogênea.

## SUGESTÃO DE LEITURA

http://educacao.uol.com.br/disciplinas/matematica/media-desvio-padrao-e-variancia-nocoes-de-estatistica.htm

http://estatisticax.blogspot.com.br/2008/02/medidas-de-disperso.html

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

COSTA, Paulo, R. **Estatítica.** Disponível em <https://www.ufsm.br/unidades-universitarias/ctism/cte/wp content/uploads/sites/413/2018/11/04_estatistica.pdf. > Acesso em 18/07/2019

# UNIDADE 7 – PROBABILIDADE E ESTATÍSTICA

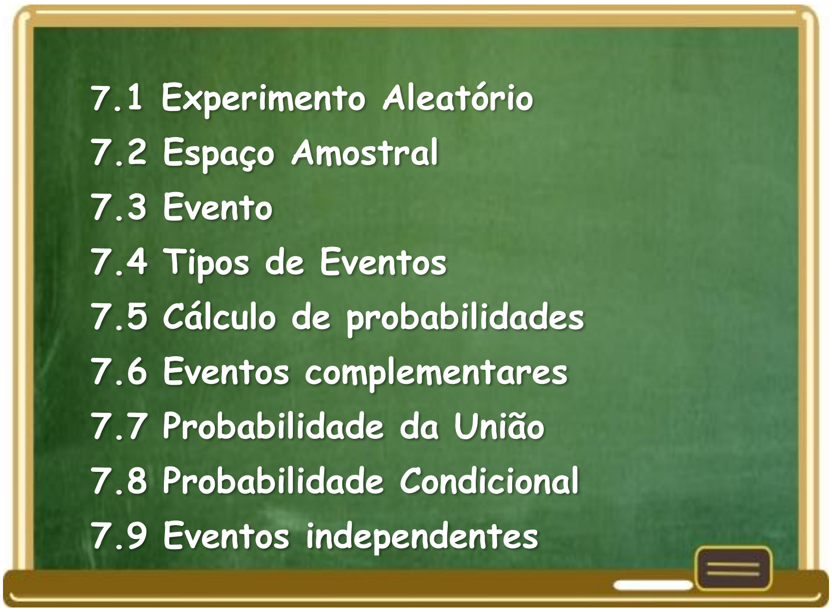
## Objetivos

- Identificar os principais elementos da probabilidade.
- Definir probabilidade condicional.
- Classificar os tipos de eventos.

Já sabemos que para se obter informações sobre alguma característica da população, o tamanho amostral é de fundamental importância. Estudaremos agora a probabilidade, que é uma ferramenta usada e necessária para se fazerem ligações entre a amostra e a população, de modo que a partir de informações da amostra se possam fazer afirmações sobre características da população.

Assim, pode-se dizer que a probabilidade é a ferramenta básica da Estatística Inferencial.

Estudaremos então ...




## 7.1 EXPERIMENTO ALEATÓRIO

É aquele experimento que, quando repetido em iguais condições, pode fornecer resultados diferentes, ou seja, são resultados explicados ao acaso, isto é, não podem ser previamente determinados, dependem exclusivamente do acaso. Se o fenômeno seguir um modelo não determinístico, temos um experimento aleatório que possui as seguintes características:

**EXEMPLO**

Lançamento de dois dados, lançamento de uma moeda, sorteio de um cupom dentre cem mil cupons, sorteio de uma peça dentre 200 fabricadas, etc...

## 7.2 ESPAÇO AMOSTRAL (S)

É o conjunto formado por todos os resultados possíveis de um experimento aleatório.

**EXEMPLO**

Num jogo de futebol, entre duas equipes, uma das equipes pode obter resultados tais como: vitória (v), empate (e) ou derrota (d). Tem-se então: S = {v, e, d}. Portanto: n (S) = 3

## 7.3 EVENTO (E)

É um conjunto qualquer de resultados de um experimento aleatório. Pode-se dizer que um evento é um subconjunto do espaço amostral.

**EXEMPLO**

No lançamento de 2 moedas apareçam faces iguais. Os elementos do evento são: E = {(K, K), (C, C)}.

## 7.4 TIPOS DE EVENTOS

**- Evento certo** – é o próprio espaço amostral.

**EXEMPLO**

Lançamento de um dado e ocorrência de um número menor ou igual a 6 na face superior.

**- Evento impossível –** é o subconjunto vazio do espaço amostral.

EXEMPLO

Lançamento de um dado e ocorrência de um número maior do que 6 na face superior.

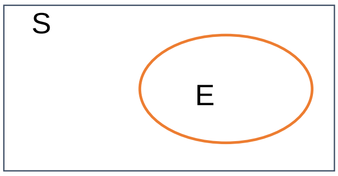
**- Eventos elementares –** são aqueles que têm um só elemento.

EXEMPLO

Lançamento de um dado e ocorrência de um número ímpar maior do que 4 na face superior.

## 7.5 CÁLCULO DA PROBABILIDADE DE UM EVENTO OCORRER

Podemos definir o cálculo da probabilidade de um evento como a razão (divisão) entre o número de elementos do evento e o número de elementos do espaço amostral.




$$\boldsymbol{P(E) = \frac{n(E)}{n(S)}}$$

Onde: **n** (E) = o número de elementos do evento

**n** (S) = o número de elementos do espaço amostral

**P** (E) = a probabilidade de ocorrer o evento

Na prática, calcular a probabilidade é dividir:

$$\boldsymbol{P(E) = \frac{\text{número de resultados favoráveis}}{\text{número de resultados possíveis}}}$$

Observação: Percentualmente, a probabilidade varia de 0% a 100%, ou seja, 0% P (E) 100% ou 0 P (E) 1.

EXEMPLO

Fazendo-se inspeção em um lote de 240 peças de motor, o departamento de controle de qualidade constatou que 20 peças estavam com defeito. Retirando-se ao acaso uma das 240 peças, a probabilidade de esta peça **NÃO** ser defeituosa é: • Sendo S o conjunto dos elementos do espaço amostral, casos possíveis, e n (S) o número de elementos deste conjunto.

• Sendo $\bar{A}$o conjunto dos elementos das peças defeituosas, e n $(\bar{A})$ o número de elementos deste conjunto.

• Sendo E' o conjunto dos elementos das peças não defeituosas, e n (E') o número de elementos deste conjunto. Neste caso, é o conjunto dos casos favoráveis.

$$n(S) = 240 \quad n(\bar{A}) = 20 \quad n(A) = 220$$

Para calcular a probabilildade de retirada de uma peça que seja não defeituosa, faça assim:

$$P(E') = \frac{n(A)}{n(S)} = \frac{220}{240} = \frac{11}{12} = 0{,}916 \ldots$$

Isso significa que a probabilidade de retirar uma peça não defeituosa é de 91,6% aproximadamente.

EXEMPLO

Uma recente pesquisa da Harris com 1.010 adultos nos EUA mostrou que 202 deles fumavam. Ache a probabilidade de que um adulto selecionado aleatoriamente nos EUA seja fumante.

Vamos considerar

$$n(S) = 1.010 \qquad n(F) = 202$$

$$P(F) = \frac{Número\ de\ fumantes}{Número\ total\ de\ pessoas\ pesquisadas} = \frac{202}{1.010} = 0{,}200\ ou\ 20\%$$

Note que, neste exemplo a abordagem clássica não pode ser usada, uma vez que os dois resultados (fumante e não fumante) não são igualmente prováveis.

## 7.6 EVENTOS COMPLEMENTARES P (A)

A probabilidade de não ocorrer o evento A é igual a 1 menos a probabilidade de ocorrer A, que pode ser representada por:

$$P(\bar{A}) = 1 - P(A)$$

EXEMPLO

Analisando um lote de 360 peças para computador, o departamento de controle de qualidade de uma fábrica constatou que 40 peças estavam com defeito.
Retirando-se uma das 360 peças, ao acaso, a probabilidade de esta peça NÃO ser defeituosa é:

**- Sendo S =** conjunto dos elementos do espaço amostral, casos possíveis, e n(s) o número de elementos deste conjunto.

**- Sendo A** = conjunto de elementos das peças defeituosas, e n(A) o número de elementos deste conjunto.

**- Use $\bar{A}$ =** conjunto dos elementos das peças não defeituosas, e n($\bar{A}$) o número de elementos deste conjunto. Neste caso, é o conjunto dos casos favoráveis.

- n(S) = 360, n(A) = 40 e n ($\bar{A}$) = 320

Para calcular a probabilidade de retirada uma peça que seja não defeituosa, podemos resolver de duas formas, vejamos então

- $\boldsymbol{P(\bar{A}) = \frac{n(E)}{n(S)} = \frac{320}{360} = \frac{8}{9}}$ **ou**
- $\boldsymbol{P(A) = \frac{n(A)}{n(S)} = \frac{40}{360} = \frac{1}{9}}$ **como** $P(\bar{A}) = 1 - P(A) \therefore P(\bar{A}) = 1 - \frac{1}{9} = \frac{8}{9}$

EXEMPLO

Uma recente pesquisa da Harris com 1.010 adultos nos EUA mostrou que a probabilidade de que um adulto selecionado aleatoriamente nos EUA seja fumante é de 0,200. Ache a probabilidade de se selecionar aleatoriamente um adulto nos EUA e ele não ser um fumante.

Considerando : $A = Evento\ de\ ser\ fumante$ e $\bar{A} = evento\ de\ não\ ser\ fumante$ temos que a probabilidade de não ocorrer o evento A é igual a 1 menos a probabilidade de ocorrer A, que pode ser representada por:

$$P(\bar{A}\,) = 1 - P(A)$$

Assim

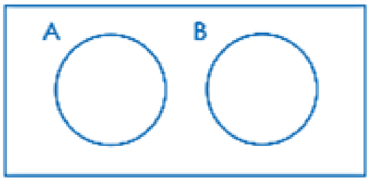
$$P(\bar{A}\,) = 1 - 0{,}200 = 0{,}800 \; ou \; 80\%$$

## 7.7 PROBABILIDADE DA UNIÃO P (A U B) = P (A OU B)

Nesse caso, existem dois tipos possíveis de situação

### 7.7.1 Eventos mutuamente exclusivos

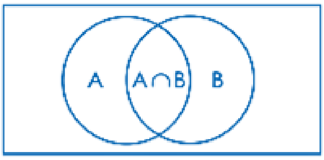
Dois eventos são mutuamente exclusivos se $A \cap B = \emptyset$, neste caso:




$$\boldsymbol{P(A \cup B) = P(A) + P(B)}$$

### 7.7.2 EVENTOS NÃO MUTUAMENTE EXCLUSIVOS

Dois eventos não são mutuamente exclusivos, se A ∩ B = Ø, neste caso:




$$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

**EXEMPLO**

No lançamento de um dado, determine a probabilidade de se obter um número par ou maior que 3.

- Espaço amostral: S = {1, 2, 3, 4, 5, 6} → n (S) = 6

- Evento A (números pares): A = {2, 4, 6} → n (A) = 3

- Evento B (números maiores que 3): B = {4, 5, 6} → n (B) = 3

- Evento de A B: A B = {4, 6} → n (A B) = 2

Calculando a probabilidade, temos:

$$P\,(A\,U\,B) = P\,(A) + P(B) - P\,(A \cap B) \rightarrow P\,(A\,U\,B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$$

## 7.8 PROBABILIDADE CONDICIONAL

Sejam dois eventos A e B associados a um espaço amostral S. A probabilidade de A ocorrer dado que o evento B ocorreu é definida por:

$\boldsymbol{P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}}$ **onde P(B) ≠ 0**

Portanto, quando calculamos P(A/B), tudo se passa como se o evento B fosse um novo espaço amostral reduzido dentro do qual queremos calcular a probabilidade do evento.

EXEMPLO

Considere o conjunto de números inteiros {1,2,3,4,5, ...,18,19,20}, e, por meio de um sorteio aleatório, retire um número. Se o número sorteado for ímpar, qual a probabilidade de o número sorteado ser o 13?

- Espaço amostral $S = \{1,2,3,\ldots,19,20\} \rightarrow n(S) = 20$

- Evento $A = \{13\} \rightarrow n(A) = 1$

- Evento B: Condição para ocorrência do evento $A = \{1,3,5,7,9,11,13,15,17,19\} \rightarrow n\,(B) = 10$

$$(A \cap B) = \{13\} \rightarrow n\,(A \cap B) = 1$$

$$\boldsymbol{P\left(\frac{A}{B}\right) = \frac{1}{10} = 0{,}1 = 10\%}$$

## 7.9 EVENTOS INDEPENDENTES

Diz-se que dois ou mais eventos são independentes, quando a ocorrência de um não depende (ou não é condicionada, ou não se vincula) da ocorrência do outro, isto é, a informação adicional de que um dos eventos já ocorreu em nada altera a probabilidade de ocorrência do outro.

Dados dois eventos independentes A e B, a probabilidade de que ocorram os eventos A e B é dado por:

$$P(A \cup B) = P(A) \times P(B)$$

EXEMPLO

Num grupo de jovens estudantes a probabilidade de que um jovem, escolhido ao acaso, tenha média acima de 7,0 é 1/5. Nesse mesmo grupo, probabilidade de que um jovem saiba jogar futebol é 5/6. Qual a probabilidade de escolhermos um jovem (ao acaso) que tenha média maior que 7,0 e saiba jogar futebol?

Resolvendo...

A: ter média acima de 7,0.

B: saber jogar futebol.

A e B: ter média acima de 7,0 e saber jogar futebol.

Logo podemos considerar, $P(A) = \frac{1}{5}$ e $P(B) = 5/6$

Assim, o fato de ter média maior que 7,0, não depende do fato de saber jogar futebol, e vice-versa. Quando isso ocorre, dizemos que os eventos são independentes.

$$P(A \cup B) = P(A) \times P(B)$$

$$P(A \cup B) = \frac{1}{5} \times \frac{5}{6} = \frac{5}{30} = \frac{1}{6} = 0{,}166\ldots \sim 16{,}7\%$$

## ATIVIDADES DE APRENDIZAGEM

1. Faça um resumos de todos conceitos abordados aqui.

2. Escreva as formulas apresentadas e indique sua aplicação.

## Sugestão de Leitura

http://arquivoescolar.org/bitstream/arquivo-e/97/1/IPE%202005.pdf
https://esquadraodoconhecimento.wordpress.com/matematica/probabilidade-e-estatistica/

http://www.matematiques.com.br/materiais.php

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

COSTA, Paulo, R. **Estatítica.** Disponível em <https://www.ufsm.br/unidades-universitarias/ctism/cte/wp content/uploads/sites/413/2018/11/04_estatistica.pdf. > Acesso em 18/07/2019.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelando o Poder dos dados** [VitalSource Bookshelf version]. Retrieved from vbk://9788521633440

S., M. D., I., N. W., A., F. M. (07/2017). **A Estatística Básica e sua Prática**. 7. ed. [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287.

# UNIDADE 8 – DISTRIBUIÇÃO NORMAL

## Objetivos

- Definir Distribuição de probabilidade
- Identificar as características de uma distribuição normal.
- Realizar cálculos de probabilidade para uma distribuição normal.

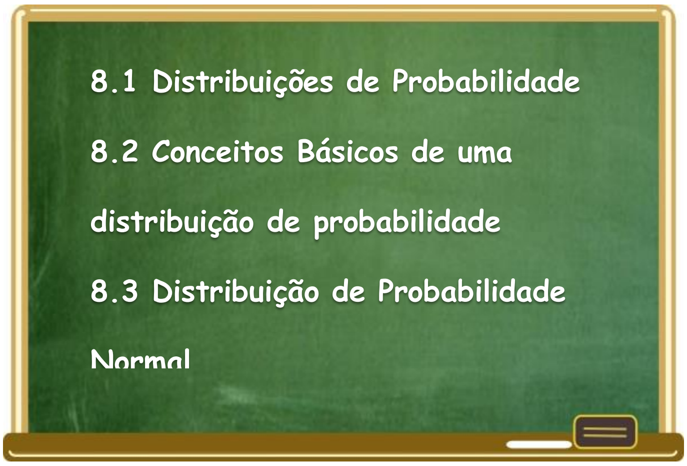
Uma distribuição estatística é uma função que define uma curva, e a área sob essa curva determina a probabilidade de ocorrer o evento por ela correlacionado. E o que é distribuição normal?

Aqui vamos responder a esta pergunta, aprendendo um pouco sobre uma das mais importantes distribuições estatísticas que temos na literatura.

A distribuição normal é conhecida também como como distribuição gaussiana, isto porque em meados do século XIX, Frederick Gauss, com seus estudos sobre eventos da natureza, observou um comportamento padrão entre as amostras estudadas por ele. Esse comportamento, posteriormente foi apresentado como a Curva de Gauss. Que mostrava que grande parte dos eventos ficam em torno de um valor médio, com uma certa variabilidade.

A Distribuição Normal é uma curva simétrica em torno do seu ponto médio, apresentando assim seu famoso formato de sino. A curva de distribuição normal representa o comportamento de diversos processos nas empresas e muitos fenômenos comuns, como por exemplo, altura ou peso de uma população, a pressão sanguínea de um grupo de pessoas, o tempo que um grupo de estudantes gasta para realizar uma prova.

Nesta unidade, estudaremos...




## 8 – INTRODUÇÃO

Vamos utilizar a situação esquematizada na Figura 1 para compreendermos o que este capítulo irá abordar.

Ao investigarmos o número de meninas em famílias com exatemente dois filhos, podemos usar as duas abordagens a seguir:

- **Usar dados amostrais reais:** A abordagem nas unidades 1, 2, 3 e 4 é a de coletar dados amostrais de famílias reais, resumir os resultados em uma tabela que representa a distribuição de frequência e, então, encontrar estatísticas, tais como média amostral $\bar{x}$ e o desvio padrão amostral $s$.
- **Usar probabilidades para encontrar os resultados que são esperados:** Usando os princípios de probabilidade, podemos encontrar a probabilidade de cada possível número de meninas. Então, podemos resumir os resultados em uma tabela que represente uma distribuição de frequência.

## 8.1 – DISTRIBUIÇÃO DE PROBABILIDADE

Até aqui, as variáveis aleatórias consideradas não possuíam, necessariamente, qualquer sentido de aplicação. Entretanto algumas variáveis aleatórias são muito importantes e, devido a esa importância, surge o interesse em estudar suas distribuições de probabilidade.

Um distribuição de probabilidade é essencialmente um modelo de descrição probabilistítca de uma população, entendo por população o conjunto de todos os valores de uma variável aletória.

As ideias de população e distribuição de probabilidade são deste modo indissociáveis e serão, a partir de agora tratadas como sinônimos. As distribuições de probabilidade forma a espinha dorsaç da metodologia estatística, uma vez que pela sua natureza, a estatística somente trabalha com variáveis cujos valores não ocorrem de modo determinístico.

Existem vários modelos descrevendo o comportamento probabilístico de variáveis aleaórias discretas e contínuas. Nesta unidade, vamos discutir sobre a distribuição de probabilidade Normal.

Podemos definir distribuição de probabilidades como uma descrição que dá a probabilidade para cada valor da variável aleatória. Ela é frequentemente expressa na forma de uma tabela, de uma fórmula ou de um gráfico.

## 8.2 – CONCEITOS BÁSICOS DE UMA DISTRIBUIÇÃO DE PROBABILIDADE - DEFINIÇÕES

Uma **variável aleatória** é uma variável (normalmente representada por x) que assume um único valor numérico, determinado pelo acaso, para cada resultado de um experimento.

Uma **distribuição de probabilidade** é uma descrição que dá a probabilidade para cada valor da variável aletória. Ela é frequentemente expressa na forma de uma tabela, de uma fórmula ou de um gráfico.

Uma **variável aletória discreta** possui uma coleção de valores que é finita ou enumerável. (Se há infinitos valores, o número de valores é enumerável se é possível contá-los) individualmente, tal como o número de jogadas de uma moeda aantes de se obter coroa).

Uma **variável aletória contínua** tem infinitos valores, e a coleção desses valores não é enumerável. (Isto é, é impossível contar os itens individuais porque pelo menos alguns deles estão em uma escala contínua.)

## 8.3 – DISTRIBUIÇÃO DE PROBABILIDADE NORMAL

A distribuição normal( ou distribuição de Gauss) é uma distribuição especialmente importante na metodologia estatística. Sua importância advém das suas propriedades, do número de fenômenos ( variáveis) que podem, pelo menos aproximadamente, ser modelados através dela e da quantidade de métodos e técincias que são derivados tendo-a como pressuposição básica. Esse conjunto d emétodos e técnicas froma a chamada Estatística clássica ou Estatística Paramétrica.

É uma distribuição teórica de frequências, onde a maioria das observações se situa em torno da média ( centro de distribuição) e diminui gradual e simetricamente no sentido dos extremos. A distribuição normal é representada graficamente pela curva normal ( também chamada curva de Gauss) que tem forma se sino e é simétrica em relação ao centro, onde se localiza a média $\mu$.

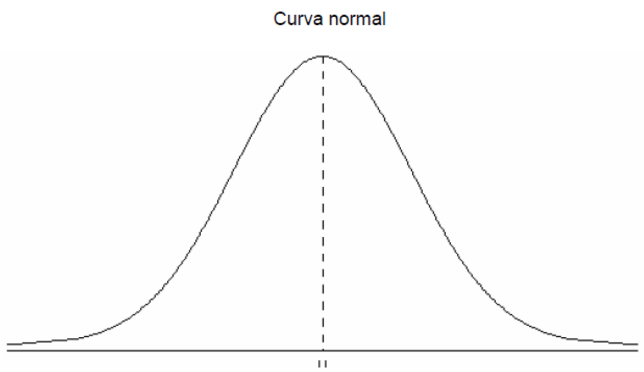



FIGURA 1 - GRÁFICO DA DISTRIBUIÇÃO NORMAL

### 8.3.1 - Função de densidade de probabilidade normal

A função que define a probabilidade normal é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\frac{x-\mu}{\sigma})^2}, x \in \mathbb{R}$$

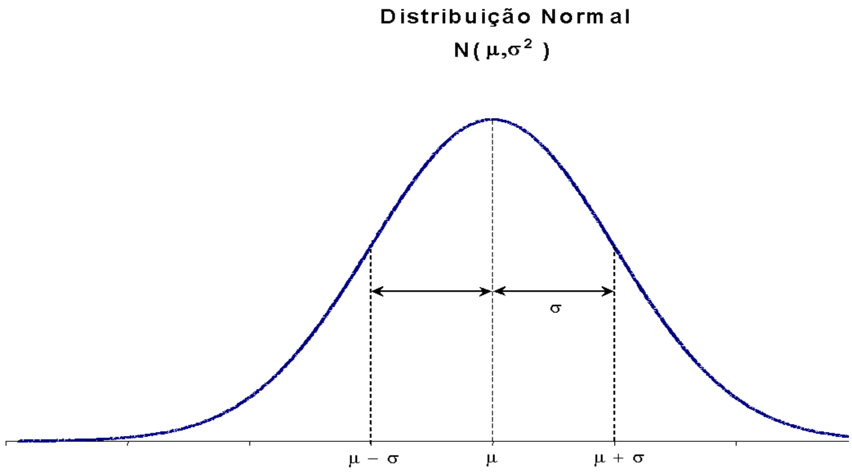
Onde,

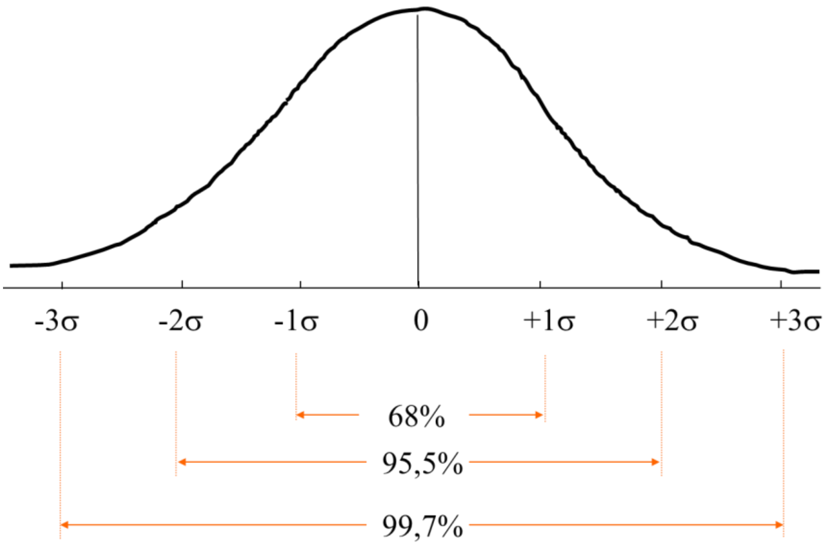
x – ponto considerado da distribuição

$\mu$ – média da distribuição

$\sigma$ – desvio padrão da distribuição

$$Notação: X \sim N(\mu, \sigma^2)$$




**8.3.2 – Propriedades da Distribuição Normal**

1 – O máximo da função densidade de probabilidade se dá no ponto $x = \mu$

2 – A distribuição é simétrica em relação ao centro onde coincidem a média, a moda e a mediana. $\mu = Mo = Md$

3 – Os pontos de inflexão ( onde a curva passa de convexa para côncava) são exatamente $\mu - \sigma$ e $\mu + \sigma$

4 – Verifica-se na distribuição normal que:

$$P(\mu - \sigma < X < \mu + \sigma) \sim 0{,}68 \; ou \; 68\%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \sim 0{,}955 \; ou \; 95{,}5\%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \sim 0{,}997 \; ou \; 99{,}7\%$$

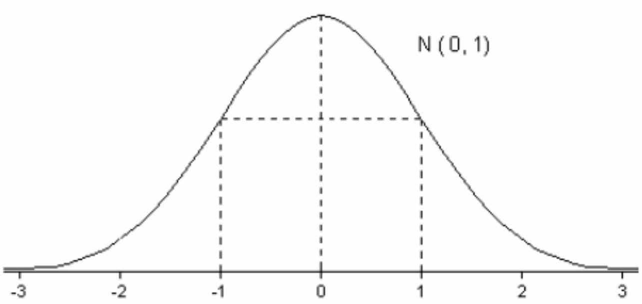
A distribuição normal depende da média e do desvio padrão, assim podemos dizer que para cada valor de $\mu$ e $\sigma$, existe uma distribuição normal diferente. Desta forma podemos afirmar que existem infinitas distribuições ( e curvas) normais. A fim de minimizar os cálculos para este tipo de distribuição, foi determinada uma distribuição normal padrão ou reduzida. Através da distribuição normal padrão é possível estudar qualquer variável que tenha distribuição normal, com quaisquer valores de $\mu$ e $\sigma$.

**8.3.3 –Distribuição Normal Reduzida**

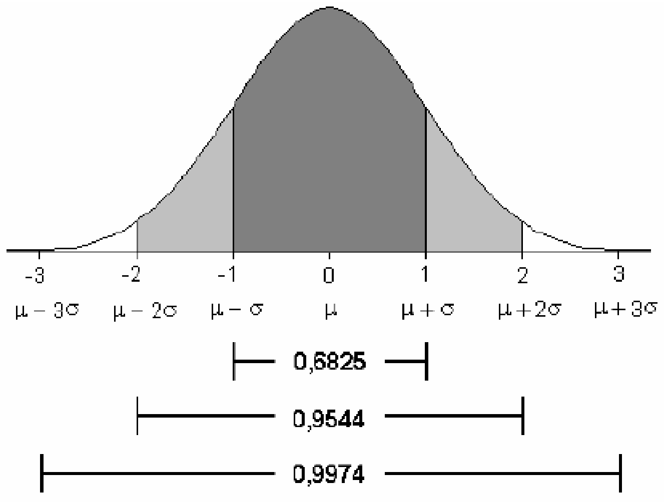
É uma distribuição normal de uma vairável Z que tem média igual a zero $(\mu = 0)$ e desvio padrão igual a 1. Para a variável Z, a função densidade de probabilidade resulta em

$$f(xz) = \frac{1}{\sqrt{2\pi}} e^{\frac{Z^2}{2}}, -\infty < Z <= +\infty$$

A função densidade de probabilidade mais simplificada da distribuição normal padrão facilitou o cálculo das áreas sob a sua curva. Assim, a curvaq norma padrão foi dividida em pequenas tiras, cujas áreas foram calculadas e apresentadas numa tabela. Na tabela da distribuição norma padrão (Anexo 1), são encontradas as áreas correspondentes aos intervalos de 0 a Z.




Os valores negativos não são apresentados na tabela porque a curva é simétrica; portanto, as áreas correspondentes a estes valores são exatamente iguais às dos seus simétricos positivos. Os valores de Z na tabela (Anexo 1) vão de 0 a 3,99; este é o limite estabelecido com base na quarta propriedade da distribuição normal, como mostra a figura a seguir.



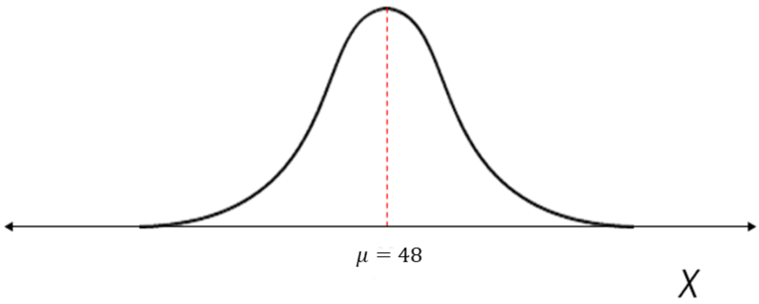

A distribuição de uma variável X, com quaisquer valores para média e desvio padrão, pode ser obtida pela transformação da vairável X na variável Z, através da expressão

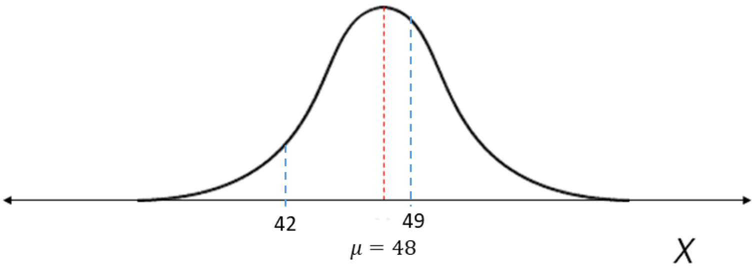
$$Z = \frac{X - \mu}{\sigma}$$

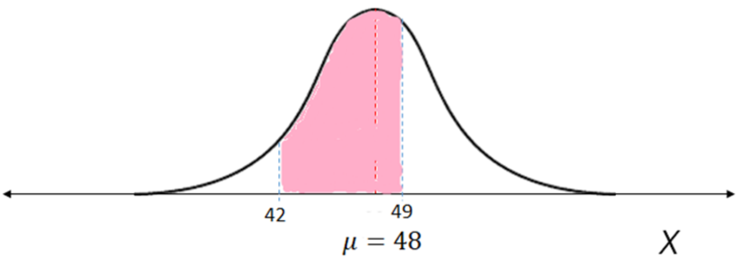

Suponha que a estatura de recém-nascidos do sexo feminino é uma variável com distribuição normal de média $\mu = 48\ cm$ e $\sigma = 3\ cm$. Detemine:

**a) A probabilidade de um recém-nascido ter estatura entre 42 e 49 cm;**

Vamos começar a resolver fazendo um esboço da curva normal, e identicando o ponto onde estamos calculando a probabilidade.




Para encontrar a probabilidade é preciso calcular o valor de Z, duas vezes, pois a fórmula é em torno da média. Assim, vamos calcular da seguinte forma:

$$P(42 < X < 48)$$

$$Z = \frac{42 - 48}{3} = -\frac{6}{3} = -2$$

Agora vamos olha na tabela do Anexo 1, o valor de Z para 2,00.

| Z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2518 | 0.2549 |
| 0.7 | 0.2580 | 0.2612 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |

Notem que, para Z= -2 olhamos o valor positivo, devido a simetria da curva em relação à média.

Assim

$$P(42 < X < 48) = 0{,}4772$$

Agora faremos o mesmo cálculo para:

$$P(48 < X < 49)$$

$$Z = \frac{49 - 48}{3} = \frac{1}{3} = 0{,}33 \ldots$$

Para Z = 0,33.. o valor na tabela é de

| Z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |

Então

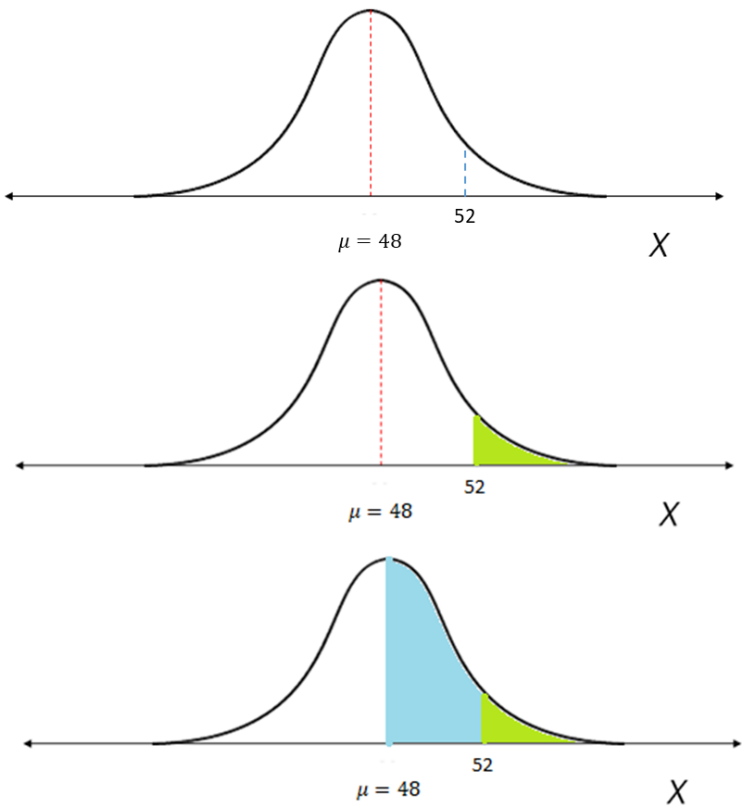
$$P(48 < X < 49) = 0{,}1293$$

Como procuramos a probabilidade de 42 a 49 cm, devemos somar os dois valores encontrados

$$P(42 < X < 49) = 0{,}4772 + 0{,}1293 = 0{,}6065 \; ou \; 60{,}65\%$$

**b) A probabilidade de um recém nascido ter estatura superior a 52cm.**

Neste caso, procuramos a probabilidade de ser maior que 52cm, como mostra parte hachura de verde na figura a seguir. Entretanto, os valores aqui calculados para Z são compreendidos entre um ponto e a média.




Para resolvermos este problema, faremos o uso da tabela para calcularmos inicialmente o valor de Z entre 48 e 52cm.

$$P(48 < X < 52)$$

$$Z = \frac{52 - 48}{3} = \frac{4}{3} = 1{,}33 \ldots$$

De acordo com a tabela do Anexo 1, temos

$$P(48 < X < 52) = 0{,}4082$$

O que procuramos é a probabilidade superior a 52cm, assim devemos considerar que a curva por ser simétrica em relação à média tem 0,5 ou 50% de probabilidade para a direita da média e para a esquerda da média.

Desta forma, para calcularmos a região hachurada em verde devemos fazer,

$$P(X > 52) = 0{,}50 - P(48 < X < 49)$$

$$P(X > 52) = 0{,}50 - 0{,}4082$$

$$P(X > 52) = 0{,}0918 \; ou \; 9{,}18\%$$

## REFERÊNCIAS

SILVA, J.G.C. da. **Estatística Experimental**: análise estatística de experimentos (Apostila) 2000. 318p.

CRESPO, ANTÔNIO ARNOT. **Estatística Fácil**. São Paulo: Saraiva, 2002.

COSTA, FABRÍCIO M. **Estatística.** Pará: Universidade Federal do Pará. 77p.

CORREA, SONIA M. B. B. **Estatística e Probabilidade.** 2.ed. Belo Horizonte. PUC Minas Virtual, 2003. 116p.

https://www.voitto.com.br/blog/artigo/distribuicao-normal. Acesso em 31/07/2019.

TRIOLA, Mário F. **Introdução à Estatística**. 12ed. Rio de Janeiro. LTC, 2017.

COSTA, Paulo, R. **Estatítica.** Disponível em <https://www.ufsm.br/unidades-universitarias/ctism/cte/wp content/uploads/sites/413/2018/11/04_estatistica.pdf. > Acesso em 18/07/2019.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelando o Poder dos dados** [VitalSource Bookshelf version]. Retrieved from vbk://9788521633440

S., M. D., I., N. W., A., F. M. (07/2017). **A Estatística Básica e sua Prática**, 7. ed. [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287