ESTATÍSTICA

S237e Santos, Érica Marques da Silva
Estatística. / Érica Marques da Silva Santos; Mariana de Lazzari Gomes (rev.). – Muriaé: FAMINAS, 2024.
104p.

ISBN: 978-65-89983-27-9

1. Estatística. I. Santos, Érica Marques da Silva. II. Gomes, Mariana de Lazzari (rev.). III. Título.

CDD: 519.5

Ficha catalográfica elaborada pelo Sistema Integrado de Biblioteca FAMINAS

SUMÁRIO

APRESENTAÇÃO DA DISCIPLINA

Olá, alunos e alunas!

Sejam muito bem-vindos à nossa Disciplina ***Estatística***.

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 indicador para trabalhar com diversos produtos, possibilitando maiores estratégias na busca e no planejamento de soluções.

Atualmente um novo olhar está acontecendo em relação à estatística 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.

Vamos juntos nesse desafio?

UNIDADE I

CONCEITOS INICIAIS E MÉTODO ESTATÍSTICO

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

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

CONSIDERAÇÕES INICIAIS

- Apresentação dos conceitos de estatística
- Método Científico
- O que é estatística?
- População
- Amostra
- Amostragem
- Censo
- Variáveis estatísticas
- Fases do método Estatístico

APRENSENTAÇÃO DOS CONCEITOS DE ESTATÍSTICA

Ao se fazer uma pesquisa científica, devemos iniciar nosso trabalho elaborando hipóteses e, em seguida devemos realizar o teste destas hipóteses.

A princípio, essas hipóteses são elaboradas em termos científicos dentro da área de estudo. Em seguida, estas devem ser expressas em termos estatísticos. A Estatística possui diversas definições, mas de um modo geral podemos entender a estatística como um conjunto de métodos e técnicas que visam torna clara e objetiva todas as fases do estudo dos fenômenos de massa.

MÉTODO CIENTÍFICO

Podemos entender o método científico como um conjunto de meios e rotinas organizados 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.

O QUE É ESTATÍSTICA

Antes de definirmos formalmente a estatística, peço que respondam às seguintes perguntas.

- Você sabe quantas pessoas existem na sua casa? Tenho certeza que sim.
- 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?

Ufaaa!!! Acredito que você deve estar pensando que essas perguntas são detalhes bem exagerados, mas nem sempre o mundo foi tão populoso.

Se pararmos para analisar historicamente a população mundial de um tempo atrás, digamos, no século XV, vamos verificar que a quantidade de pessoas era bem menor. Se voltássemos um pouco mais no tempo por exemplo, à Grécia Antiga, a população seria ainda menor. Pois então, 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.

Assim, por quê?

Porque lidamos com grandes números.

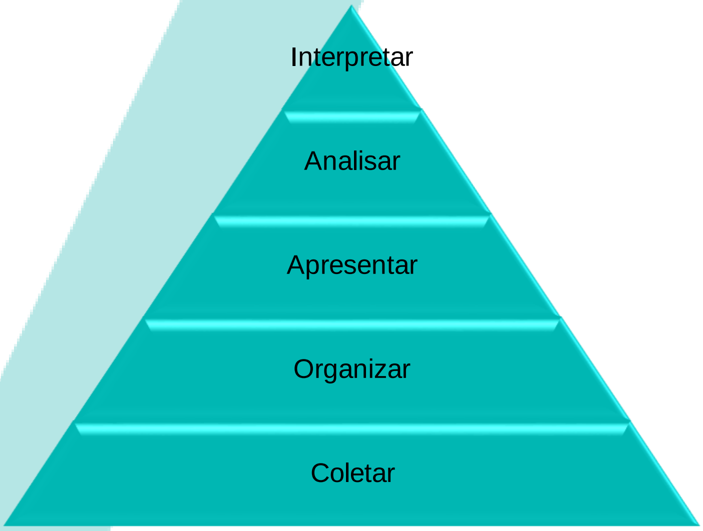
A Estatística ou métodos estatísticos, como é chamada algumas vezes, nasceu com os negócios do Estado, daí a origem do seu nome seu nome.

No entanto, na sociedade atual, 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.

(Fonte:http://www.gestaoescolar.diaadia.pr.gov.br/arquivos/File/profuncionario/cadernos/disc_ft_se_cad_16_estatistica_aplicada_a_educacao.pdf)

Pirâmide – Elementos Principais da definição de Estatística



É comum que, as pessoas limitem o termo Estatística à organização e descrição dos dados, este tipo de limitação é equivocado, pois a maior contribuição da estatística é proporcionar métodos inferenciais, que permitam conclusões que transcendam os dados obtidos inicialmente.

Através da correta 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 “casuísmos” 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.

POPULAÇÃO

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

AMOSTRA

É uma parte ou um subconjunto de uma população que deve ser representativa, ou seja, amostra é um conjunto de elementos pertencentes à população. Uma amostra deve sempre ser representativa, isto é, os dados que pertencem à amostra devem fornecer e representar as informações sobre a população.

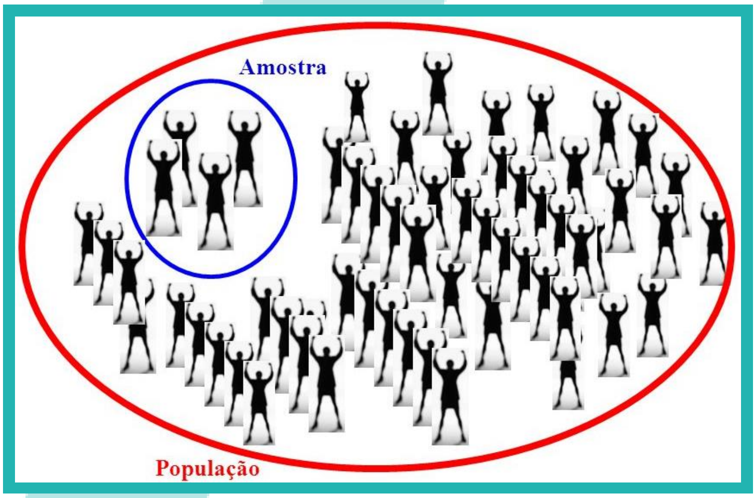
Quando o estudo é feito sobre a amostra, os resultados obtidos são usados para tirar conclusões sobre a população, este tipo de trabalho, refere-se à inferência estatística.

As estatísticas resultantes do estudo sobre uma amostra são denominadas estimativas. Assim, podemos dizer que toda análise estatística será inferida a partir das características obtidas da amostra. É importante que a amostra seja representativa da população, em outras palavras, que as suas características sejam, em geral, as mesmas que as do todo (população).

Enfim, pode-se afirmar que a amostra é um subconjunto finito e representativo de uma população. A utilização de amostras se dá em grande parte por questões práticas e/ou econômicas.

Dentre as razões de recorrer ao uso de amostras podemos citar: baixo custo e menor tempo para o levantamento de dados e melhor investigação dos elementos observados.

Figura 1 – Amostra e População



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.

AMOSTRAGEM

DEFINIÇÃO: É o processo de seleção de amostras que irão representar a população. A composição da amostra é feita através de métodos adequados de seleção dessas unidades. A Amostragem é considerada uma técnica especial de escolher amostras, de forma a garantir o acaso (aleatório) 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.

Técnicas de amostragem Probabilísticas

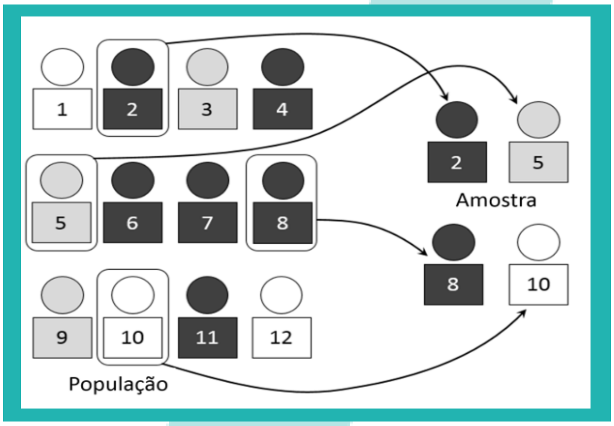
Existem técnicas apropriadas para obter amostras, de modo a garantir (o quanto possível) o sucesso da pesquisa e dos resultados.

Uma vez que os objetivos e a população a ser estudada já estão definidos, deve-se pensar na constituição a amostra, e quais as características ou variáveis a serem investigadas.

Neste material iremos abordar alguns procedimentos e tipos de seleção de amostra, que veremos a seguir.

a) **Amostragem casual ou aleatória simples** – este tipo de amostragem envolve o sorteio dos elementos da amostra. Para isto, podemos numerar a população de 1 a **n**, um dispositivo aleatório é utilizado para selecionar, por exemplo, sorteio, escolhem-se k números dessa sequência, os quais corresponderão aos elementos da amostra.

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

Uma das formas de realizarmos a amostragem aleatória por sorteio é registrar os valores em papéis, então depositamos os papeis em uma urna, misturamos e sorteamos a amostra.

Esse procedimento não é muito prático para populações numerosas, nesse caso é mais indicado utilizar uma tabela de números aleatórios.

POR TABELA DE NÚMEROS ALEATÓRIOS:

Para realizar a seleção da amostra por tabela de números aleatórios devem seguir alguns passos, são eles:

1º passo: Elaborar a relação dos dados brutos da população, ordenando os números com uma numeração aleatória. Por exemplo: Uma pesquisa obteve 30 informações sobre uma variável em estudo. Considerando que o total de dados brutos é 30, teremos uma tabela enumerada de 00 até o 29, usando dois dígitos, como segue.

Nº	HOTEL CUSTO	Nº	HOTEL CUSTO	Nº	HOTEL CUSTO	Nº	HOTEL CUSTO	Nº	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: Organizar a tabela, o próximo passo é sortear o valor de **n (Nº)**, neste exemplo iremos selecionar uma amostra de tamanho igual a 10, desta forma utilizando a tabela de números aleatórios, serão selecionados 10 números. Após o sorteio dos 10 números, verifica-se quais os valores da variável irão compor a amostra. Como segue.

09 - $24	25 - $25	21 - $22
11 - $25	13 - $19	06 - $25
12 - $20	20 - $24	

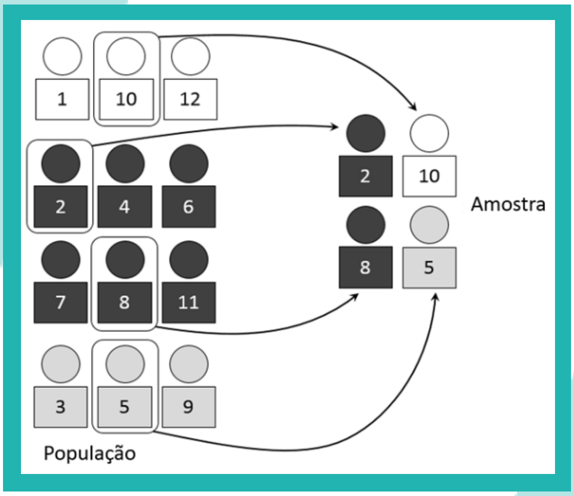
3º Passo: Os valores sorteados no 2º passo irão compor a amostra e os demais são descartados. É possível que o sorteio da amostra possa ser com reposição ou sem reposição. Quando falamos de reposição, significa que o número sorteado, volta para a urna e pode ser sorteado novamente. No exemplo que foi apresentado a seleção da amostra foi sem reposição.

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

Amostragem proporcional estratificada – Este método é usado quando as populações se dividem em subpopulações, chamados também de estratos, pode ser aceitável julgar que a variável de interesse apresente comportamento distinto nos diferentes estratos.

Quando ocorre de existir subpopulações, a amostragem proporcional estratificada é mais indicada para garantir a representatividade da amostra.

Figura 3 – Amostragem Proporcional Estratificada



EXEMPLO

Considerando um grupo de 90 alunos de uma escola de dança, 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

A seleção os meninos e meninas que irão compor a amostra, é feita por 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.

Amostragem estratificada uniforme – Esta se difere da amostragem proporcional estratificada por não utilizar critério de proporcionalidade, neste tipo de amostragem seleciona-se a mesma quantidade de elementos de cada estrato, devendo ser usada para comparar os estratos ou obter estimativas separadas para cada estrato.

EXEMPLO

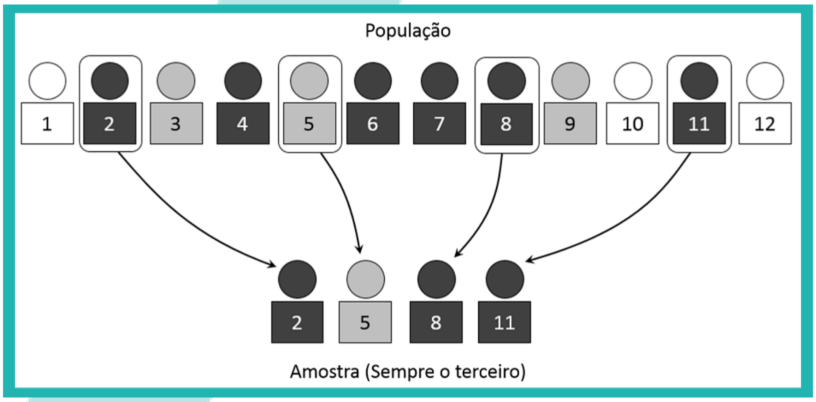
Uma empresa de Inovação Tecnológica possui com 480 funcionários, dos quais 288 são do sexo feminino e os 192 restantes do sexo masculino. Considerando a variável qualitativa nominal "sexo" para estratificar essa população, vamos obter uma amostra estratificada uniforme de 50 funcionários.

Considerando 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
Meninos	288	$n_1 = 25$
Meninas	196	$n_2 = 25$
Total	480	50

Amostragem sistemática – é um procedimento para a seleção aleatória da amostra aplicada quando os elementos da população já estão ordenados. Assim, não é necessário construir um sistema de referência ou de amostragem.

Figura 4 – Amostragem Sistemática



EXEMPLO

Uma empresa tem um 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.

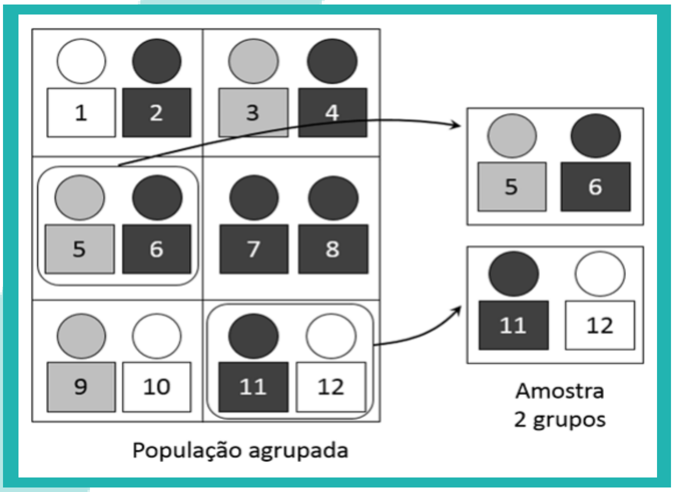
2º define-se o tamanho da amostragem total: n = 100. $I = \frac{N}{n} = \frac{1600}{100} = 16$

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.

Amostragem por Conglomerados – Esta técnica explora existência de grupos (*clusters*) na população. Esses grupos são representativos da população em relação a característica que se pretende medir. Em outras palavras, estes grupos contêm elementos distintos que representam toda população. Quando os grupos possuem esta representatividade, é possível selecionar apenas alguns desses conglomerados para realizar o estudo, conforme ilustra a figura 5.

Figura 5 – Amostragem por conglomerados



Podemos ver esta técnica por outra perspectiva. A demais técnicas que estudamos até o momento, a amostragem por conglomerados tem como unidade de amostra os grupos de estudo, o que pode ser visto como um benefício em relação ao custo do processo de amostragem. No entanto, na amostragem por conglomerados, aumenta-se a imprecisão dos resultados devido à baixa heterogeneidade dentro dos 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	Educação Física – 06	Gastronomia – 12
Arquitetura e Urbanismo – 02	Enfermagem – 07	Medicina – 13
Biomedicina – 03	Engenharia Civil – 08	Nutrição – 14
Ciências Contábeis – 04	Engenharia de Produção – 09	Odontologia – 15
Direito – 05	Farmácia – 10	Psicologia – 16
	Fisioterapia – 11	

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.

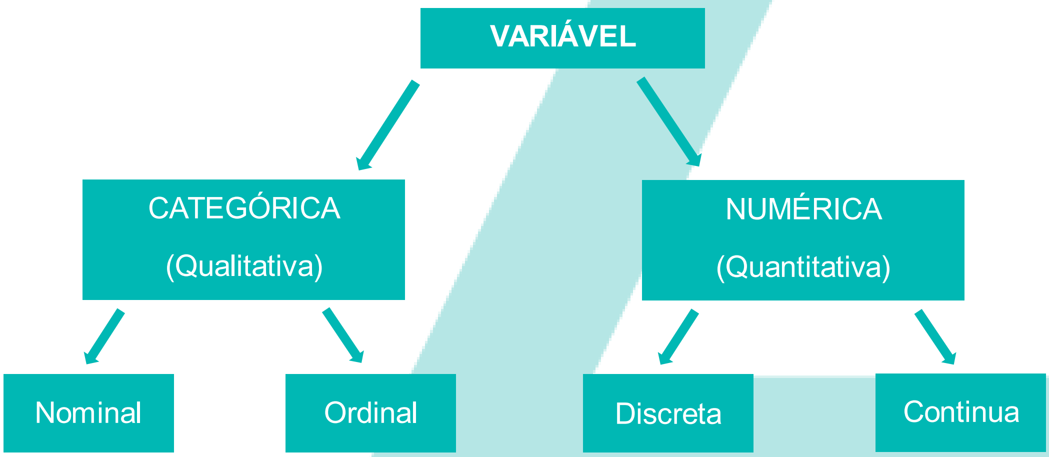
CENSO

É o estudo completo de toda população, ou seja, todos elementos que constituem a população são estudos. Sabe-se que, quanto maior a amostra, mais precisas e confiáveis deverão ser as induções feitas sobre a população. Assim, os resultados mais perfeitos são obtidos pelo Censo.

Cabe ressaltar, que quando o emprego de amostras ocorre com certo rigor técnico, a amostra pode levar a resultados mais confiáveis ou até mesmo melhores do que os que seriam obtidos através de um censo.

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, ureia, 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 superior Incompleto e Pós-graduação)
- **NÚMERO DE PESSOAS QUE RESIDEM NA MESMA CASA**.

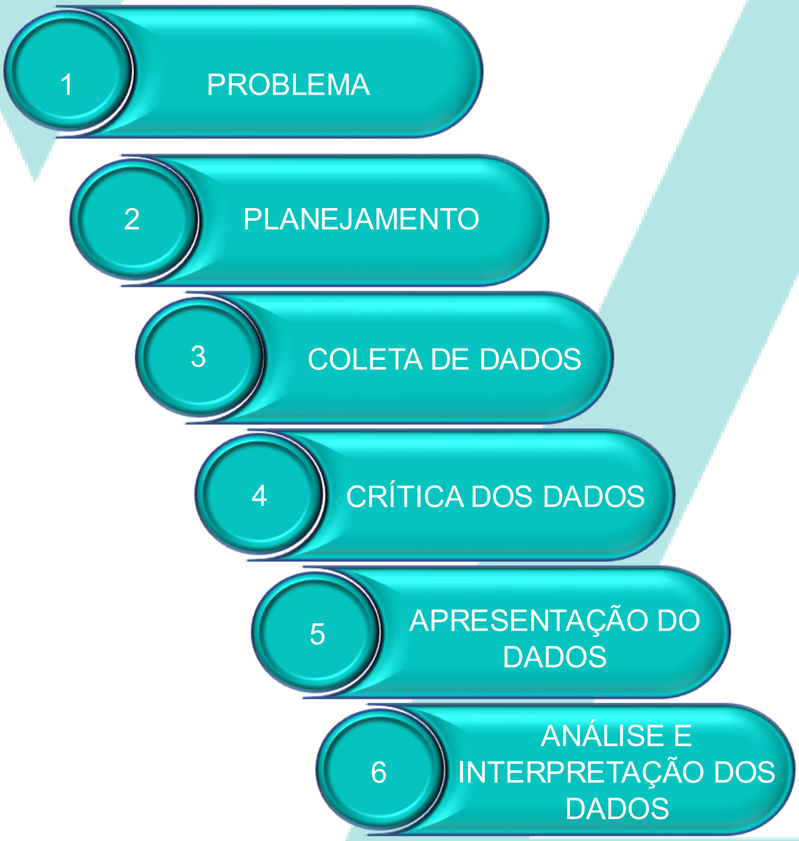
Classifique cada variável apresentada.

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

PESO - Variável **quantitativa** 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 **quantitativa** e **discreta**

FASES 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:



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.

PLANEJAMENTO

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

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.

CRÍTICA DOS DADOS

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

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.

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 pequenos ou muito grandes.

Apresentação dos dados

- Os dados serão apresentados por meio de tabela.

Analise 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.

LEITURA COMPLEMENTAR

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/

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

http://www.gestaoescolar.diaadia.pr.gov.br/arquivos/File/profuncionario/cadernos/disc_ft_se_cad_16_estatistica_aplicada_a_educacao.pdf. Acesso em 30/07/2021.

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

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.

UNIDADE II

TABELAS, SÉRIES E GRÁFICOS 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.**
- **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.

CONSIDERAÇÕES INICIAIS

- Representação Tabular
- Elementos de uma Tabela
- Séries estatísticas
- Representação gráfica
- Gráficos Estatísticos

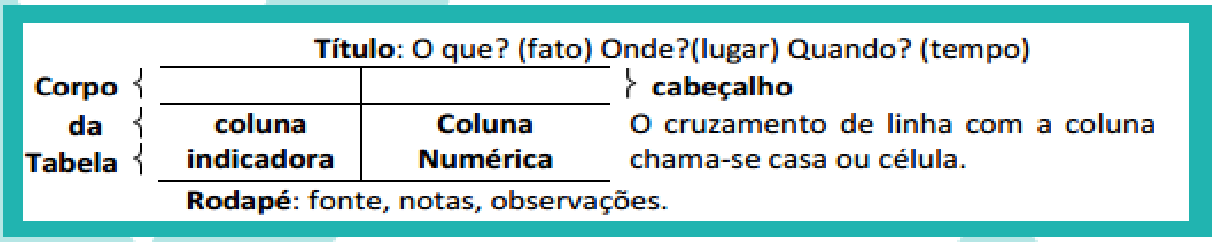
REPRESENTAÇÃO TABULAR

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

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.**

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 oumarcha, identifica-se pelo caráter variável do fator cronológico. Assim, deve-se ter:

Elemento variável:

Época Elementos Fixos: Local e Fenômeno

Tabela 1 – Série Temporal

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

a) Não foram aceitas mais inscrições

Fonte: https://www.si.ips.pt/ests_si/web_base.gera_pagina?p_pagina=1183. Acesso 14/07/2019.

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

Tabela 2 – Série Especificativa

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

Fonte:http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso14/07/2019.

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

Elemento variável: Local

Elementos Fixos: Época e Fenômeno

Tabela 3 – Série Regional

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

Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso 14/07/2019.

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.

Tabela 4 – Especificativa x Temporal

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

Fonte:http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html.Acesso 14/07/2016.

EXEMPLO

Temporal x Geográfica: é a série estatística onde variam o tempo e o local.

Figura 6 – Temporal x Geográfica

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

Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso14/07/2016.

EXEMPLO

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

Figura 7 - Especificativa X Geográfica

Veículos Adquiridos por Regiões / 2001						
Veículos	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

Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso14/07/2016.

EXEMPLO

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

Figura 8 - Especificativa X Especificativa

Distribuição de Material Escolar por Séries Alagoas/2011				
Materiais	Séries			Total
	1ª	2ª	3ª	
Lápis	53	21	39	113
Borracha	61	38	46	145
Caneta	32	71	60	163
Lapiseira	38	48	46	132

Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso14/07/2016.

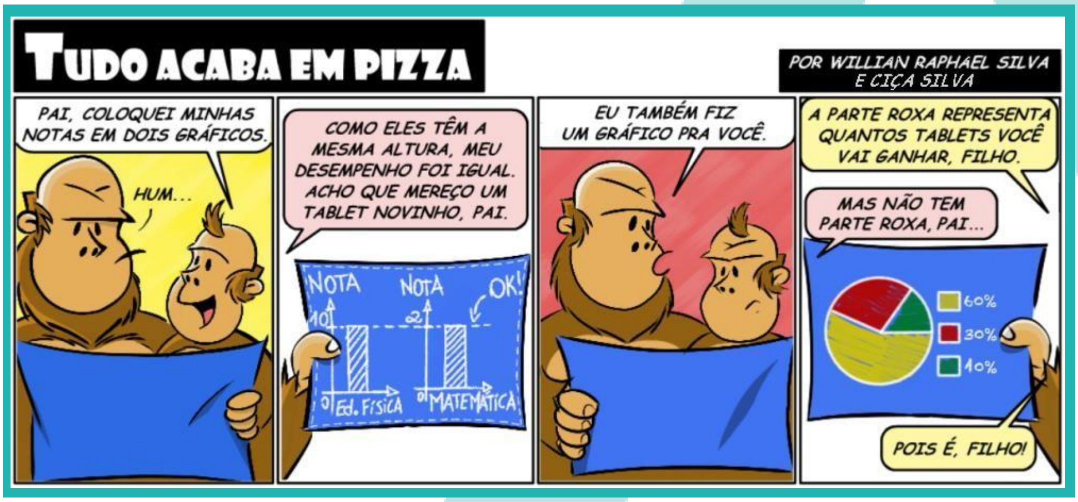
ATIVIDADES DE APRENDIZAGEM

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

2. Elabore um exemplo de série geográfica.

REPRESENTAÇÃO GRÁFICA

Figura 2 – Tudo acaba em pizza



Fonte: https://blogdoprofh.wordpress.com/2014/10/15/notas-e-graficos-tirinha/. Acessoem 22/07/2016.

Os gráficos são desenhos que envolvem formas e cores cuja construção utiliza técnicasde 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 cujoobjetivo 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 impressãovisual 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.

FORMAS DE APRESENTAÇÃO DOS GRÁFICOS

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

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.

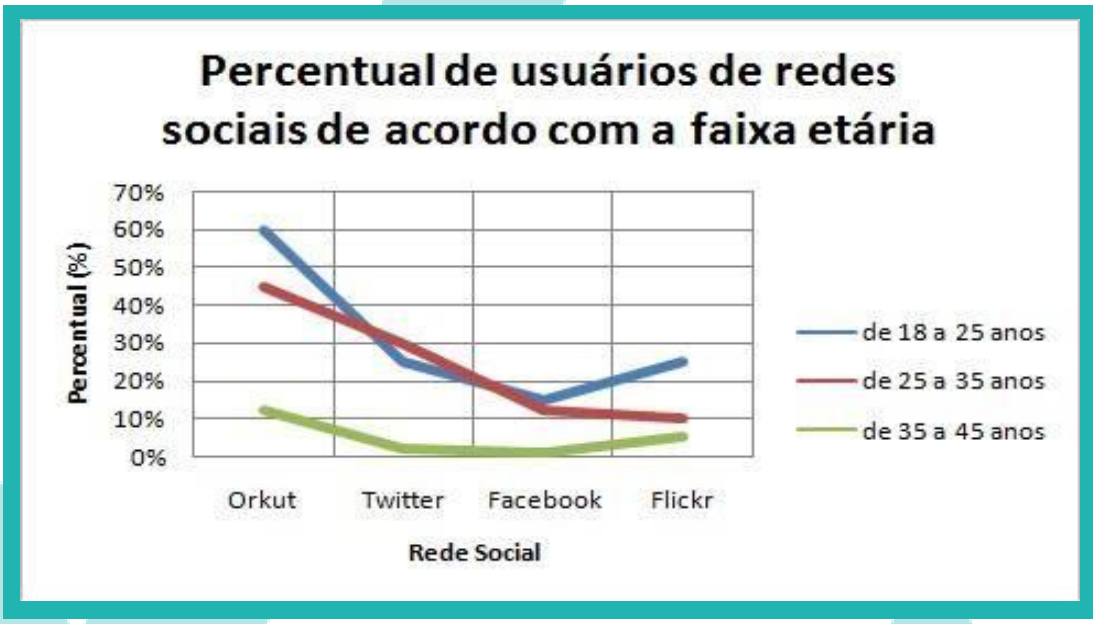
Gráficos de análise: São gráficos que servem principalmente ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser informativos.Os gráficos de análise frequentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes, um texto explicativo,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

PRINCIPAIS 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 3 – Gráficos em linha ou em Curva

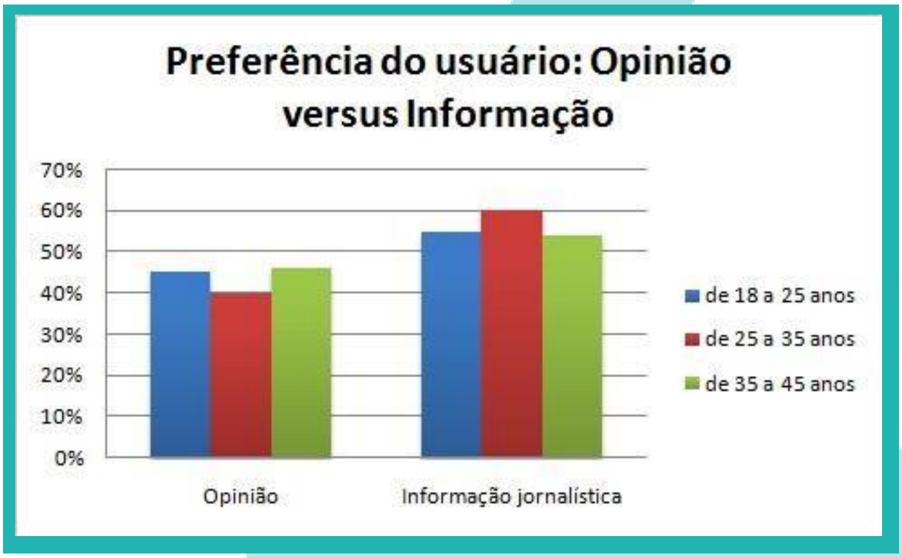


Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso 24/07/2019

GRÁFICOS 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 respectivosdados.

Figura 4 – Preferência do Usuário



Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso 24/07/2019

GRÁFICOS 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 ficadividido 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 maisrepresentadas 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 dosetor) e a porcentagem calculada através de uma regra de três:

$$\begin{array}{l} \text{Total (\%)} \rightarrow 360^{\circ} \\ \text{Parte (\%)} \rightarrow x^{\circ} \end{array}$$

Figura 5 – Gráficos em setores

Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso 24/07/2019

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?

LEITURA COMPLEMENTAR

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.** 2ed. Belo Horizonte. PUCMinas Virtual, 2003. 116p.

H., L. R., Frazer, L. P., Lock, M. K., F., L. E., F., L. D. (01/2017). **Estatística - Revelandoo 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 III

DISTRIBUIÇÃO DE FREQUÊNCIA

OBJETIVOS:

- **Compor uma distribuição de frequência com ou sem intervalos de classe;**
- **Determinar o quadro de frequências, eles são úteis para condensar grandes conjuntos de dados, facilitando a sua utilização;**
- **Representar uma distribuição de frequência através de histograma, polígono e ogiva.**

Na estatística trabalha-se, habitualmente, com grande número de informações, resultados de medições realizadas. Que podem ser dados discretos (o valor inteiro que não pode ser partido) ou contínuo (em intervalos).

CONSIDERAÇÕES INICIAIS

- Conceitos Fundamentais
- Tipos de distribuição de frequência
- Elementos de uma distribuição de frequência
- Gráficos de distribuição de frequência

CONCEITOS FUNDAMENTAIS

Dados brutos – são os dados originais, que ainda não se encontram prontos para análise, por não estarem numericamente organizados. (Também são conhecidos como Tabela Primitiva).

Exemplo: Número mensal de aparelhos defeituosos na Empresa X.

Figura 1 – Dados Brutos

	J	F	M	A	M	J	J	A	S	O	N	D
1995	6	2	5	1	0	3	2	1	3	5	5	3
1996	5	4	2	1	3	4	1	4	5	4	0	1
1997	3	1	2	4	3	1	4	1	0	3	0	2
1998	2	2	0	3	1	4	2	0	1	1	5	2

Fonte:http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_4.pdf. Acesso 14/07/2016.

Rol – são os dados brutos, organizados em ordem crescente ou decrescente.

Exemplo: Considerando o exemplo anterior temos:

Figura 2 - Rol

0	0	0	0	0	0	1	1	1	1	1	1
1	1	1	1	1	2	2	2	2	2	2	2
2	2	3	3	3	3	3	3	3	3	4	4
4	4	4	4	4	5	5	5	5	5	5	6

Fonte:http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_4.pdf. Acesso 14/07/2016.

TIPOS DE DISTRIBUIÇÃO DE FREQUÊNCIA

Dados Tabelados não agrupados em classes (dados agrupados sem intervalo de classes) – os valores da variável aparecem individualmente.
Exemplo, considerando os dados da tabela anterior:

Figura 3 – Dados agrupados sem intervalos de classes

Nº de aparelhos com defeitos	Nº de meses
0	06
1	11
2	09
3	08
4	08
5	05
6	01
Total	**48**

Fonte:http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_4.pdf. Acesso 14/07/2016.

Dados Tabelados agrupados em classes (dados agrupados com intervalo de classes) - os valores da variável não aparecem individualmente, mas agrupados em classes.

Figura 4 – Dados com intervalos de classes

Notas	Nº de alunos
0 ⊢20	020
20 ⊢ 40	065
40 ⊢ 60	230
60 ⊢ 80	160
80 ⊢ 100	025
Total	580

Fonte:http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_4.pdf. Acesso 14/07/2016.

ELEMENTOS DE UMA DISTRIBUIÃO DE FREQUÊNCIA

Convenções

⊢ Intervalo fechado à esquerda e aberto à direita: apenas o limite inferior pertence ao intervalo;

⊣ Intervalo aberto à esquerda e fechado à direita: apenas o limite superior pertence ao intervalo;

⊢⊣Intervalo fechado de ambos os lados: os dois limites pertencem ao intervalo;

— Intervalo aberto em ambos os lados: os dois limites não pertencem ao intervalo.

Limites de classe

Limite inferior da distribuição de frequência (L_I): é o valor a partir do qual são contadas as observações na distribuição de frequências.

Limite superior da distribuição de frequência (L_S): é o valor até o qual são contadas as observações na distribuição de frequências.

Amplitude total

Amplitude total da distribuição de frequência (AT): é a diferença existente entre o maior e o menor valor observado da distribuição de frequência.

$AT = L_S - L_I$

Classes de uma distribuição de frequência: são os subintervalos nos quais são contadas as observações da variável.

O número de classes (k) é calculado a partir de uma das expressões mostradas abaixo.

$k = 1 + 3{,}322 \log N$ (Fórmula de STURGES)

Método Prático: se n < 25 utilize k = 5 se n ≥ 25 utilize $k = \sqrt{n}$.

Observação: existem diversas maneiras de calcularmos o número de classes. Depende da sensibilidade do pesquisador.

Limite Inferior de Classe (l_i): é o valor a partir do qual são contadas as observações dentro da classe.

Limite Superior de Classe (l_s): é o valor até o qual são contadas as observações dentro da classe.

Amplitude de Classe (h ou a_t): é a diferença entre o maior e o menor valor observado dentro da classe.

Observação: A amplitude de classe é obtida através da seguinte equação: $at = l_s - i_i$

Frequência Simples ou Frequência Absoluta da Classe (f_i): é o número de observações contadas dentro da classe.

Frequência Absoluta acumulada de Classe (F_i): é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer, das frequências simples ou absoluta das classes.

Frequência Relativa Absoluta de Classe (f_r): é a relação existente entre a frequência absoluta ou simples de classe e o número de observações da variável.

Obtém-se a frequência relativa de cada classe a partir da seguinte equação: $f_r = \dfrac{f_i}{\sum f_i}$

Frequência Relativa Acumulada (F_r): é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer das frequências relativas das classes.

Ponto Médio de Classe (x_i): é a média aritmética calculada entre o limite inferior e o superior da classe.

Obtém-se o ponto médio de cada classe a partir da seguinte equação: $xi = \dfrac{l_s + l_i}{2}$

Intervalo de Classe ou Amplitude do intervalo de Classe (h): é o comprimento da classe.

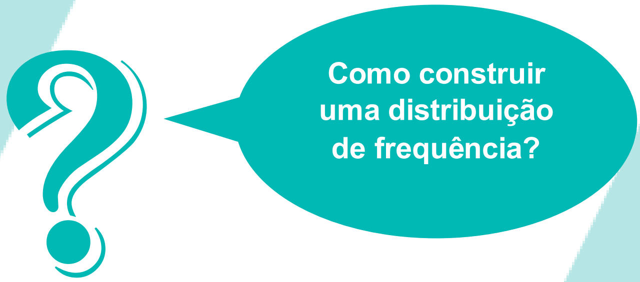
Obtém-se o intervalo de cada classe a partir da seguinte equação: $h = \dfrac{AT}{k}$

Convém arredondar o número correspondente à amplitude do intervalo de classe para facilitar os cálculos.

As séries de dados grupados (distribuição de frequências por intervalos e por pontos) são também chamadas de “séries de magnitude de variável”.

CONSTRUÇÃO DA DISTRIBUIÇÃO DE FRENQUÊNCIA

Você deve estar se perguntando:



Vamos considerar os passos que precisamos seguir para construir uma distribuição de frequência.

1º Passo

Colocar os dados em forma de Rol. Isto é, organizá-los de forma crescente ou decrescente. Aqui se recomenda colocá-los em ordem crescente

2º Passo

Identificar o valor máximo e o valor mínimo do conjunto de dados e encontrar a Amplitude Total (AT). $AT = L_S - L_I$

Calcular o comprimento ou a amplitude que deve ter o Intervalo de Classe (h). $h = \dfrac{AT}{K}$

3º Passo

Determinar o Número de Classes (K) que irão formar uma distribuição de frequências. $k = \sqrt{n}$

VAMOS PRATICAR?

A partir dos dados brutos abaixo realizaremos os passos acima citados:

10	5	0	8	7,5
10	5	1	6	4,5
9,5	6	3	5	4
8	1	7	8	9
5	0	10	5	8

1º Passo: Organizar em Rol

0	3	5	7,5	8,5
0	4	5	8	9
1	4,5	6	8	9,5
1	5	6,5	8	10
2	5	7	8	10

2º Passo: Obter a Amplitude Total.

AT = 10 – 0 = 10

3º Passo: Calcular o número de classes (k). $k = \sqrt{25} = 5$

4º Passo: obter Intervalo de Classe (h) e escrever os intervalos da tabela. $h = \frac{10}{5} = 2$

Agora que já concluímos os 4 passos importantes, vamos construir a distribuição de frequência com classes. Fique atento aos detalhes!!!!

> Importante!!! O símbolo $\sum$, é conhecido como somatório, ele representa a soma de todos os termos indicados à sua frente. No caso da frequência Relativa Absoluta de classe o termo $\sum f_i$ significa que devemos somar todas as frequências simples da distribuição

ENTENDENDO COMO SE FAZ UMA DISTRIBUIÇÃO DE FREQUÊNCIA COM CLASSES

Figura 5 –
Esquema de resolução

1) Comece pela 1ª classe, e escreva o menor valor observado.

2) agora some o menor valor com o h=2, assim encontramos o limite superior da classe. [0+2=2].

3) A partir da 2ª classe usa-se a regra, o limite superior da classe anterior será o limite inferior da classe subsequente, e o limite superior é o resultado da soma do limite inferior com o h (intervalo de classe).

4) Agora temos que contar os elementos que pertencem a cada intervalo. Por exemplo, na 1ª classe temos: 0 |-- 2, que significa, intervalo fechado à esquerda e aberto à direita. Os valores deste intervalo se aproximam de 2, isto é, não pertencem ao intervalo.de classe).

5) Os elementos: 0-0-1-1, pertencem a 1ª classe, então colocamos o valor 4 na 2ª coluna (frequência), pois, temos quatro elementos que pertencem ao intervalo. Já os elementos: 2-3, pertencem a 2ª classe, então na 2ª coluna (Frequencia) colocamos o valor 2. Assim é feito para as demais classes, observa-se o ROL.

Notas	Frequência
0 ⊢ 2	4
2 ⊢ 4	2
4 ⊢ 6	6
6 ⊢ 8	4
8 ⊢ 10	9
Total	25

Fonte: www.ebah.com.br/content/ABAAAfWSEAG/modulo-estatistica-ead. Acesso em 15/07/2016.

CALCULANDO FREQUÊNCIA E PONTO MÉDIO

Notas	Ponto Médio (xi)	Frequência simples (fi)	Frequência Relativa (fr)	Frequência Acumulada (Fr)
0 ⊢ 2	(2+0)/2 = 1	4	4/25 = 0,16	4
2 ⊢ 4	(2+4)/2 = 3	2	2/25 = 0,08	6
4 ⊢ 6	(4+6)/2 = 5	6	6/25 = 0,24	12
6 ⊢ 8	(6+8)/2 = 7	4	4/25 = 0,16	16
8 ⊢ 10	(8+10)/2 = 9	9	9/25 = 0,36	25
Total		25	$\sum fr = 1$	

GRÁFICOS REPRESENTATIVOS DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA

Feita a coleta de dados, seja através de censo, de levantamento por amostragem, ou de experimento, geralmente esses dados apresentam-se de maneira desorganizada, devendo ser organizados e resumidos para possibilitar a obtenção de informações úteis. Uma maneira de organizar os dados é a distribuição de frequência, que pode ser representada além da forma de tabela, através de gráficos.

HISTOGRAMA

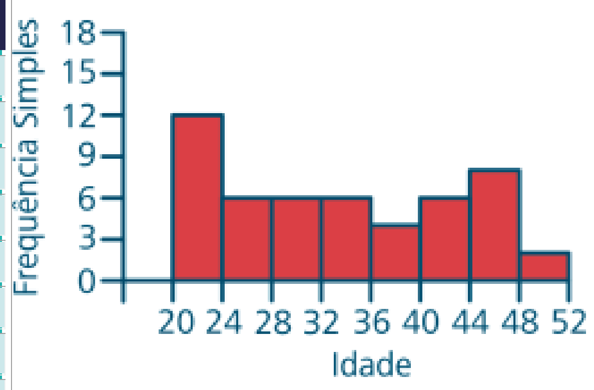
O histograma é formado por um conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal (eixo x). Sobre esse eixo são representados os intervalos de classe numa escala contínua, não sendo necessário que a escala inicie no zero, de tal modo que os seus pontos médios coincidam com os pontos médios dos intervalos de classe e seus limites coincidam com os limites da classe.

O número de retângulos encontrados em um histograma é igual ao número de intervalos de classe, sendo a largura de cada retângulo igual à amplitude do intervalo de classe, enquanto sua altura representa a frequência do intervalo de classe, e a área é proporcional à soma das frequências.

EXEMPLO

O histograma para a idade de 50 trabalhadores da indústria A, fica assim:

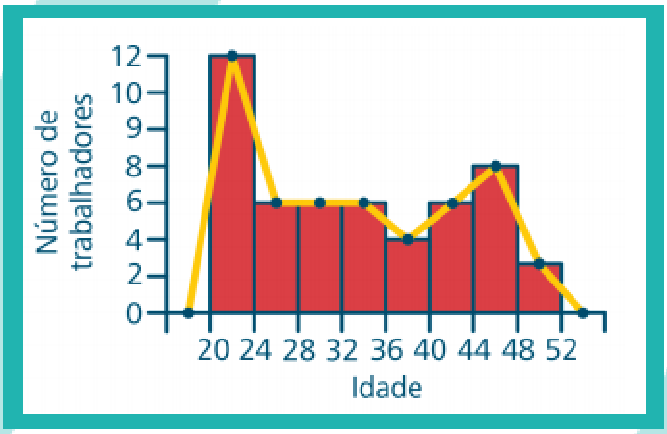
Idade (anos)	Frequência Simples
20 ⊢24	12
24 ⊢ 28	6
28 ⊢ 32	6
32 ⊢ 36	6
36 ⊢ 40	4
40 ⊢ 44	6
44 ⊢ 48	8
48 ⊢ 52	2
Total	50



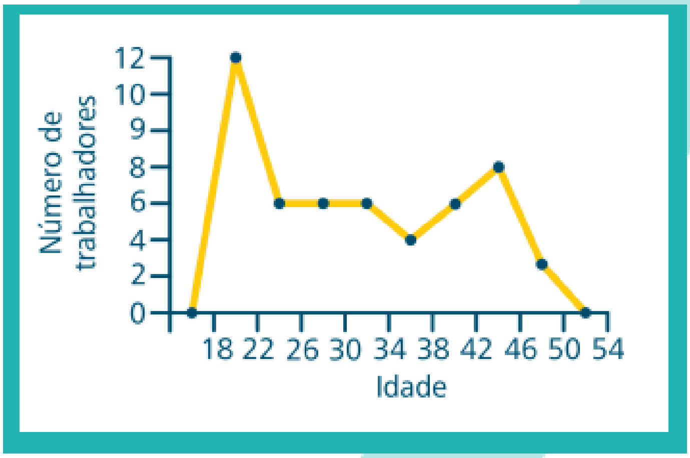
POLÍGONO DE FREQUÊNCIA

É um gráfico, em linha, de uma distribuição de frequências de configuração linear. Esse gráfico é obtido unindo-se, por segmentos de reta, os pontos médios das bases superiores dos retângulos de um histograma. Pode ser feito da mesma forma para frequências acumuladas.

Vamos construir o polígono de frequência para o exemplo anterior.



Se excluirmos o histograma, ficamos apenas com o polígono de frequência, assim:



Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso 14/07/2019.

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).

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

Figura 5 – Especificativa x Temporal

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

Fonte:http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html.Acesso 14/07/2016.

EXEMPLO

Temporal x Geográfica: é a série estatística onde variam o tempo e o local.

Figura 6 – Temporal x Geográfica

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

Fonte:http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso 14/07/2016.

EXEMPLO

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

Figura 7 - Especificativa X Geográfica

Veículos Adquiridos por Regiões / 2001						
Veículos	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

Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso 14/07/2016.

EXEMPLO

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

Figura 8 - Especificativa X Especificativa

Distribuição de Material Escolar por Séries Alagoas/2011				
Materiais	Séries			Total
	1ª	2ª	3ª	
Lápis	53	21	39	113
Borracha	61	38	46	145
Caneta	32	71	60	163
Lapiseira	38	48	46	132

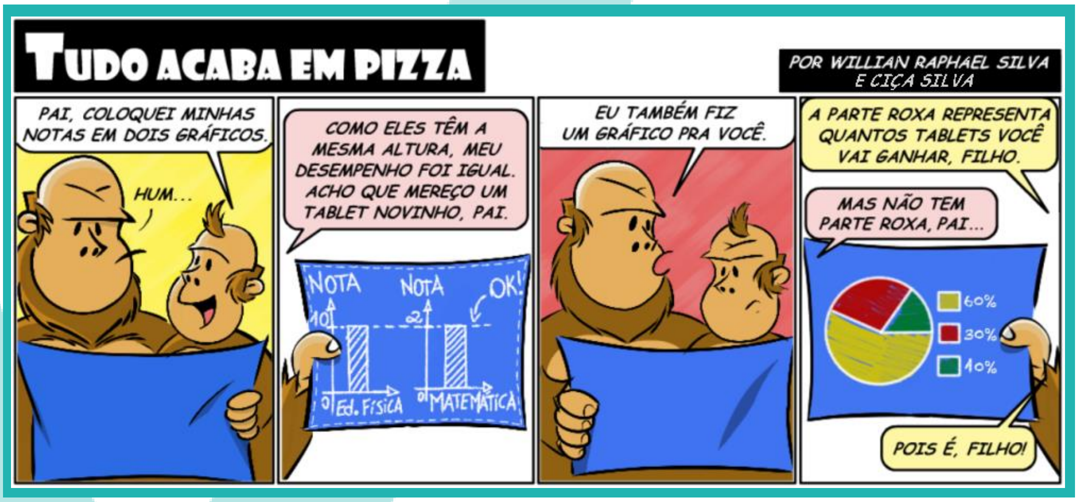
Fonte: http://estatisticaufal.blogspot.com.br/2012_04_01_archive.html. Acesso 14/07/2016.

ATIVIDADES DE APRENDIZAGEM

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

2. Elabore um exemplo de série geográfica.

REPRESENTAÇÃO GRÁFICA



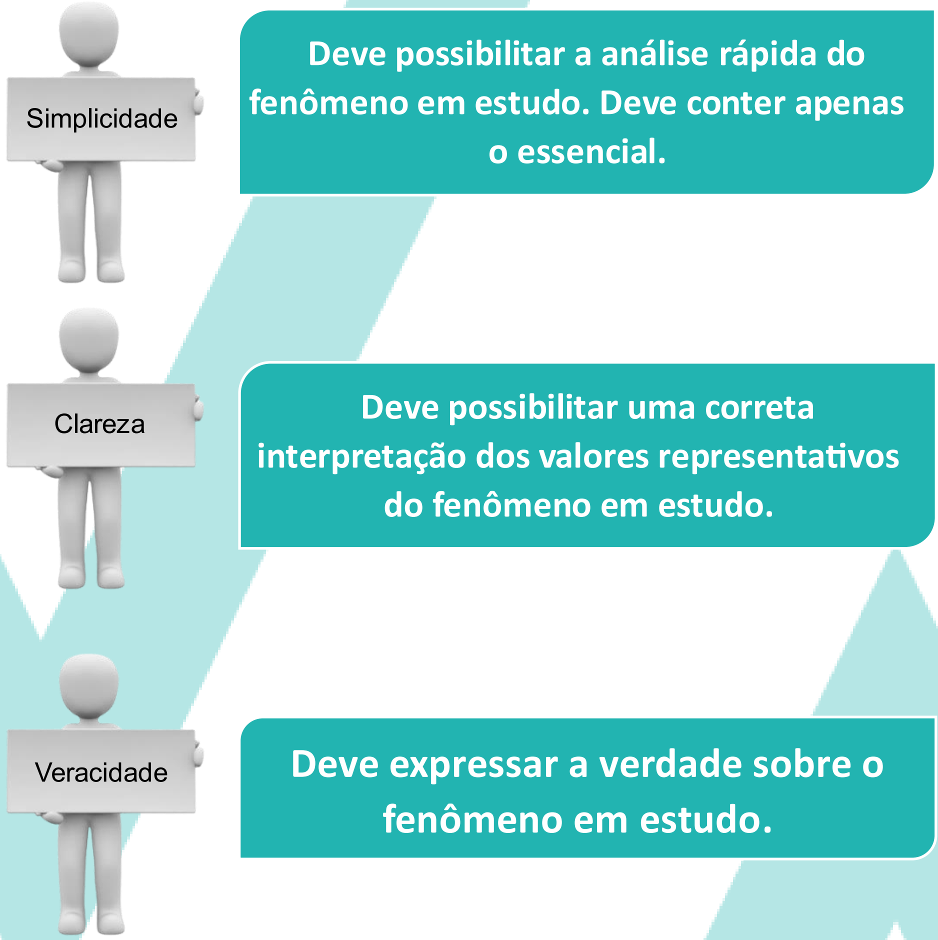
Fonte: https://blogdoprofh.wordpress.com/2014/10/15/notas-e-graficos-tirinha/. Acesso em 22/07/2016.

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:



FORMAS DE APRESENTAÇÃO DOS GRÁFICOS

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

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.

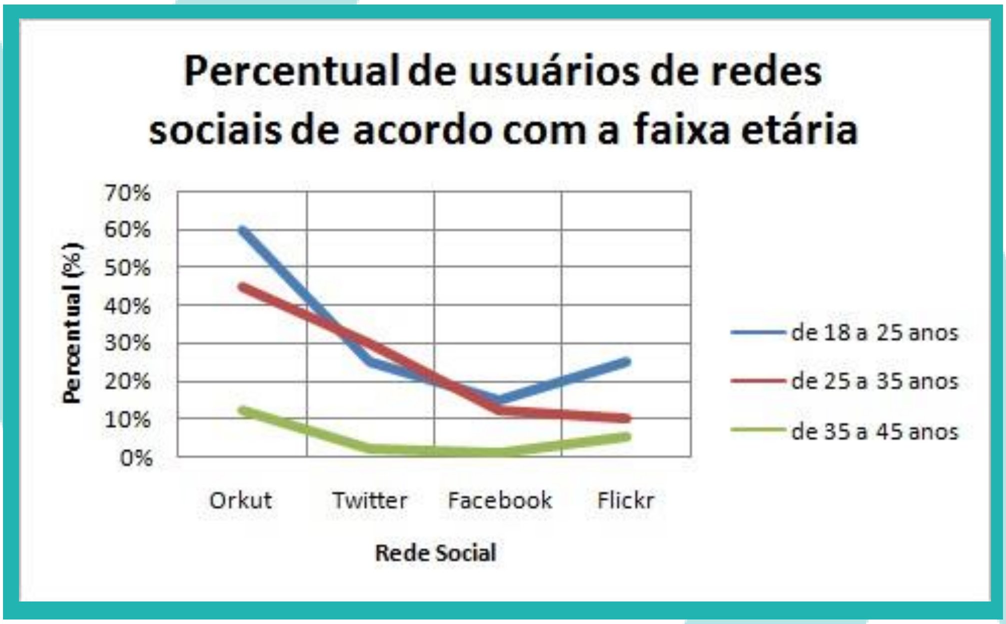
Gráficos de análise: São gráficos que servem principalmente ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser informativos. Os gráficos de análise frequentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes, um texto explicativo,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

PRINCIPAIS 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).

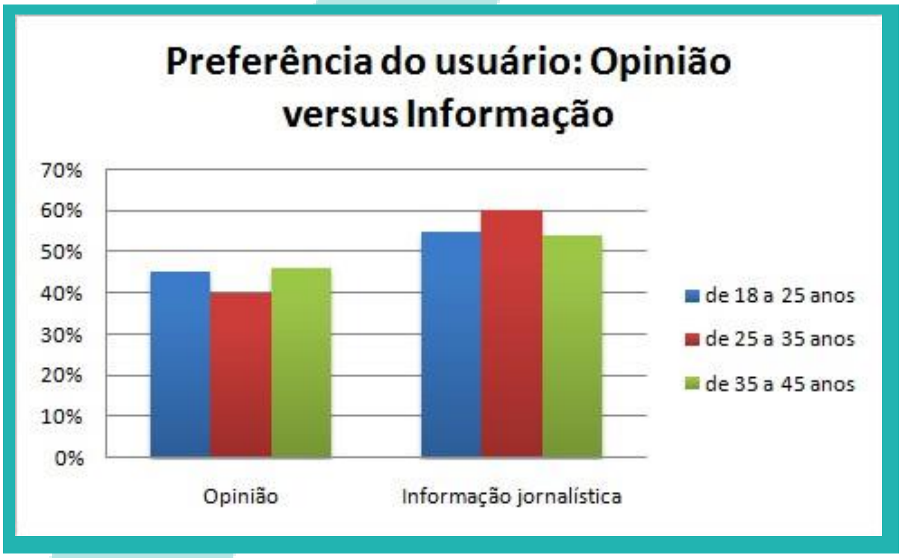
Figura 9 – Gráficos em linha ou em Curva



Fonte: http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso 24/07/2019

GRÁFICOS 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 24/07/2019

GRÁFICOS 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:

$$\begin{aligned} \text{Total (\%)} &\longrightarrow 360^\circ \\ \text{Parte (\%)} &\longrightarrow x^\circ \end{aligned}$$

Figura 10 – Gráficos em setores

Fonte:http://www.cursosprime.com.br/blogprime/2011/08/saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-excel-2007/. Acesso 24/07/2019

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?

LEITURA COMPLEMENTAR

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.** 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 IV

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.

CONSIDERAÇÕES INICIAIS

- Introdução
- Conceitos
- Média aritmética
- Mediana
- Moda

INTRODUÇÃO

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

As medidas de posição, são utilizadas 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 seus valores.

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 concentrar-se.

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

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.

MÉDIA DADOS SIMPLES OU EM ROL

Consideramos dados simples aqueles que não estão agrupados em forma de distribuição de frequência. Neste caso, o cálculo da média consiste em somar todas as observações ou medidas dividindo-se o resultado pelo número total de valores.

Dados observados

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

Números de observações

EXEMPLO

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: $\overline{\boldsymbol{x}} = \frac{\sum x_i}{n} = \frac{940}{380} \sim \mathbf{2,47}$

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

MÉDIA DISTRIBUIÇÃO DE FREQUÊNCIA - DADOS AGRUPADOS SEM INTERVALO DE CLASSE

No caso de dados resumidos em uma Tabela de distribuição de frequência simples, ou seja, sem intervalo de classe, o cálculo segue o seguinte modelo matemático.

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

Frequência Simples

Dados Observados

Número de observações

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.

Tabela 2 – Tempo de cicatrização (dias)

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:

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

Onde x_i é tempo de cicatrização e f_i a frequência simples absoluta.

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

Tabela 3 – Tempo de cicatrização (dias)

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.

MÉDIA DISTRIBUIÇÃO DE FREQUÊNCIA - DADOS AGRUPADOS COM INTERVALO DE CLASSE

Agora vamos ver como a mesma fórmula é usada para dados agrupados com intervalo de classe. O cálculo segue o seguinte modelo matemático semelhante, mas devemos nos atentar para uma diferença entre eles, veja a seguir:

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

Frequência Simples

Ponto Médio

Número de observações

Para os dados agrupados com intervalo de classes, consideramos $\boldsymbol{x_i}$ como ponto médio porque não sabemos os valores exatos que caem em determinada classe, de modo que, para tornar possíveis os cálculos, consideramos que, em cada classe, todos os valores amostrais são iguais ao ponto médio da classe.

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 4 - 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. (Unidade III).

Tabela 5 – Nível de cotinina

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 6 mais uma coluna que vai representar o produto - $f_i x_i$

Tabela 6 – Nível de cotinina

Nível de cotinina (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:

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

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

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.

MEDIANA DE DADOS SIMPLES

sendo a mediana uma medida que divide a distribuição ao meio, é imprescindível que os dados estejam ordenados em ordem crescente ou descrescente (Rol).

Veja como faz!!!!

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 $P = \frac{n+1}{2}$

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

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

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 : $Me = 41$

Peso: $Me = 220$

MEDIANA DISTRIBUIÇÃO DE FREQUÊNCIA - DADOS AGRUPADOS SEM INTERVALO DE CLASSE

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

Tabela 7 – Tempo de cicatrização (dias)

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_{md}), que por sua vez é calculada da seguinte forma:

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

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

Tabela 8 – Tempo de cicatrização (dias)

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 observamos na coluna da frequência acumulada onde ocorre pela primeira vez o valor encontrado para a posição mediana.

Tabela 9 – Tempo de cicatrização (dias)

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: **Md = 16 dias**

MEDIANA DE DADOS AGRUPADOS COM INTERVALO DE CLASSE

Amplitude do intervalo da classe mediana

Posição Mediana

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

Frequência Acumulada da classe anterior

Limite inferior da classe mediana

Frequência Simples da classe mediana

Você deve estar se perguntando:

Afinal de contas, como eu calculo a Mediana neste caso?

Vejamos:

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:

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

Como fica na prática?

EXEMPLO

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

Tabela 10 – Nível de cotinina

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	

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 Acumulada

Tabela 11 – Nível de cotinina

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	

Classe Mediana

3º passo: Aplica-se a fórmula:

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

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

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

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

$$Me = 200 + 3,05$$

$$Me = 203,5$$

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

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

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.

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**.

MODA DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA - DADOS AGRUPADOS SEM INTERVALO DE CLASSE

Quando estamos falando de dados agrupados sem intervalo de classe, a moda é o valor que possui maior frequência simples, então no exemplo abaixo temos que a moda é:

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

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

MODA DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA - DADOS 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, podemos calcular a moda por dois processos distintos.

MODA BRUTA

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**

MODA PELO PROCESSO DE KING

Existem 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***.

Diferença entre Frequência simples da modal e a classe posterior à classe modal

$$Mo = l_i + \left(\frac{\Delta_2}{\Delta_1 + \Delta_2}\right) h_i$$

Amplitude do intervalo da classe modal

Limite inferior da classe modal

Diferença entre Frequência simples da classe modal e a 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	

Maior frequência – classe Modal

A partir da classe modal, temos:

$$\Delta_1 = 412 - 220 = 192$$

$$\Delta_2 = 412 - 151 = 261$$

$$l_i = 250$$

$$h_i = 300 - 250 = 50$$

Aplicando a fórmula de King, temos:

$$Mo = 250 + \left(\frac{261}{192 + 261}\right)50$$

$$Mo = 250 + \left(\frac{261}{453}\right)50$$

$$Mo = 250 + (0{,}576)50$$

$$Mo = 250 + 28{,}8$$

$$Mo = 278{,}8$$

É interessante observar que a moda calculada por processos diferentes para uma mesma distribuição pode gerar valores diferentes, isso se deve à 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. - É uma medida de fácil interpretação e presta-se muito bem a tratamentos estatísticos adicionais. - É uma medida que sempre existe e é rigorosa e unicamente determinada. - É um valor típico de um conjunto de dados podendo substituir todos os valores de um conjunto sem alterar o total. - 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. É afetada pelos valores extremos
Mediana	Eventualmente pode ser usada para operações estatísticas mais avançadas ou para separar distribuições em duas categorias	- Define exatamente o centro de uma distribuição, mesmo quando os valores se distribuem assimetricamente em torno da média. - Pode ser determinada mesmo quando não se conhece todos os valores do conjunto de dados. - É uma medida que sempre existe e é única. - Esta medida pode ser utilizada para definir o meio de um número de objetos propriedades ou quantidades que possam de alguma forma ser ordenados. - É uma medida resistente, ou seja, não sofre influência de valores discrepantes.	É uma medida altamente influenciada por valores discrepantes. É afetada pelos valores extremos
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. - Não exige cálculo, apenas uma contagem. - 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.

LEITURA COMPLEMENTAR

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. **Estatística.** 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 V

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...

CONSIDERAÇÕES INICIAIS

- Introdução
- Amplitude
- Desvio médio
- Variância
- Desvio padrão
- Coeficiente de variação

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.

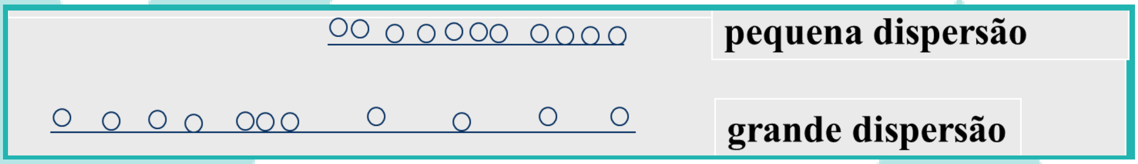
Para podermos considerar os dados confiáveis para uma análise e inferência estatística é necessário que as medidas de posição estejam acompanhadas das medidas de dispersão ou variabilidade.

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.

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

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, Variância e Desvio Padrão
- **Medida de dispersão relativa**: Coeficiente de Variação de Pearson.

AMPLITUDE

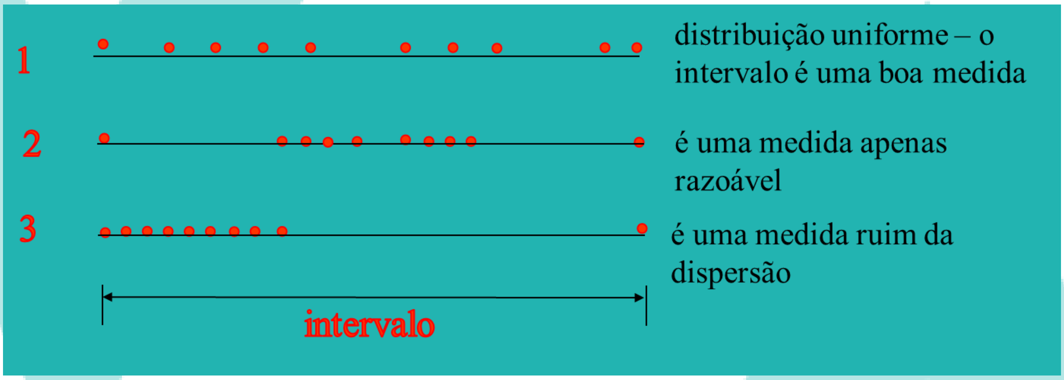
Um modo mais simples de se ter uma indicação da dispersão dos valores de uma amostra ou população é comparar o valor máximo com o mínimo. Entretanto a Amplitude Total não nos fornece qualquer indicação do que ocorre no interior do conjunto.

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.

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 σ^2.

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

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
		$\sum(x_i - \bar{x})^2 = 344{,}58$

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

VARIÂNCIA PARA DADOS AGRUPADOS

AMOSTRA

Valores observados

Média

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

Frequência Simples

Valores observados

POPULAÇÃO

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

EXEMPLO

O conjunto de dados Longevidade Mamíferos 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 1 – Longevidade de 40 espécies de mamíferos

Longevidade (anos)	f_i
1 $\vdash$ 9	10
9 $\vdash$ 17	23
17 $\vdash$ 25	2
25 $\vdash$ 33	2
33 $\vdash$ 41	3
Total	40

Fonte - Fictícia

Para facilitar, vamos montar uma tabela:

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

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.

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 temos:

Cálculos dados simples ou não agrupados

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 dados agrupados com intervalo de classe

Para uma população

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

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

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

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

- 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 anteriormente temos:

<u>Cálculos dados simples ou não agrupados</u>

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

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

$s \sim 6{,}56$

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

<u>Cálculos dados agrupados com intervalo de classe</u>

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.

LEITURA COMPLEMENTAR

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. **Estatística.** 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 VI

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 ...

CONSIDERAÇÕES INICIAIS

- Experimento Aleatório
- Espaço Amostral
- Evento
- Tipos de Eventos
- Cálculo de probabilidade
- Eventos complementares
- Probabilidade da União
- Probabilidade Condicional
- Eventos independentes

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...

ESPAÇO AMOSTRAL

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

EVENTO

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)}.

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.

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.

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

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:

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

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

EXEMPLO

O histograma para a idade de 50 trabalhadores da indústria A, fica assim:

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 probabilidade 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.

EVENTOS COMPLEMENTAR

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

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

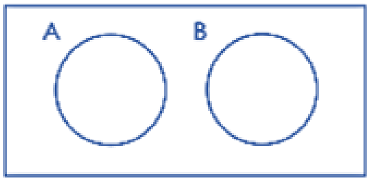
$$P(\bar{A}) = 1 - 0{,}200 = 0{,}800\ ou\ 80\%$$

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

Nesse caso, existem dois tipos possíveis de situação

EVENTOS MUTUAMENTE EXCLUSIVOS

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



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

EXEMPLOS

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:

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

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.

EXEMPLOS

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 $A = \{1,3,5,7,9,11,13,15,17,19\} \rightarrow n\,(B) = 10$

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

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

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:

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

EXEMPLOS

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.

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

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

PARA REFLETIR

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. **Estatística.** 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**, 7ª edição [VitalSource Bookshelf version]. Retrieved from vbk://9788521634287