UNIDADE I

Conceitos Básicos de Estatística/Tipos de Estatística

Dr. José Roberto Andrade do Nascimento Junior

Aula 01

Conceitos Básicos de Bioestatística

Introdução

Desde os primórdios o ser humano procura investigar sobre tudo aquilo que o cerca. Investigação recomenda pesquisa, busca de informações e análise de dados, elementos estes que, de uma certa maneira, envolvem a utilização da estatística. A aplicação da estatística na área médica e biológica é denominada de bioestatística. A interação da estatística com as ciências da saúde tem se tornado cada vez mais intensa e importante com o passar dos anos, otimizando o sistema de investigação, desde o projeto geral, a amostra, a qualidade de informação e a prestação dos resultados. Veja-se, por exemplo, que a bioestatística auxilia a genética nas questões de hereditariedade; proporciona informações à Biologia para a manutenção da saúde ambiental; é valiosa na Medicina, no estudo das doenças e possíveis tratamentos; é aplicada na Farmacologia para o desenvolvimento de medicamentos; além de ser fundamental para as demais áreas da ciências da saúde, como a Fisioterapia, Nutrição e Educação Física.

Ao final desta aula, você será capaz de:

- compreender a estatística como ciência;
- reconhecer os principais conceitos utilizados na estatística;

- compreender as fases dos métodos estatísticos.

Considerações a Respeito da Estatística

Você sabia que desde a antiguidade o homem tem a necessidade de atribuir valores numéricos e ponderações quantitativas às diversas situações da vida? Nas primeiras civilizações, procedimentos matemáticos que envolvem o estabelecimento de datas, contagem de nascimentos e óbitos, de objetos com quantidades de pessoas e animais, e questões monetárias já eram realizados pelo homem. Na Idade Média, os procedimentos estatísticos envolviam as informações obtidas com finalidades tributárias ou bélicas. O termo “estatística” surgiu no século XVII originado do vocábulo latino “status”, que significa estado. O termo era utilizado para o levantamento de dados por parte do Estado, visando à tomada de decisões (MEMÓRIA, 2004).

SAIBA MAIS

Para saber mais, acesse o link a seguir e leia este texto sobre a história da estatística. Disponível em: <http://www.ufrgs.br/mat/graduacao/estatistica/historia-da-estatistica>. Acesso em: 26 abr. 2019.

A bioestatística teve sua origem no século XIX durante a Guerra da Crimeia, quando se descobriu que a taxa de mortes nos hospitais era maior do que nas batalhas por meio das informações obtidas a respeito das péssimas condições de higiene nos hospitais. Nos dias atuais, as informações numéricas são fundamentais para a tomada de decisões nos assuntos que envolvem a sociedade.

Atualmente, a estatística é considerada uma ciência que dispõe de processos apropriados para coleta, organização, classificação, análise, apresentação e interpretação de conjuntos de dados (GLANTZ, 2014). Os procedimentos estatísticos têm sido utilizados em quase todos os campos de investigação científica, possibilitando a avaliação de novas teorias, o desenvolvimento de medicamentos, a descoberta e a cura de doenças, o crescimento demográfico, previsão de acontecimentos naturais, entre outras aplicações.

SAIBA MAIS

Leia este texto para saber mais sobre a importância da estatística nas pesquisas clínicas.

Fonte: RODRIGUES, C. F de S.; LIMA, F. J. C. de; BARBOSA, F. T. Importância do uso adequado da estatística básica nas pesquisas clínicas. Brazilian Journal of Anesthesiology, v. 67, n. 6, p. 619-625, 2017.

O procedimento estatístico ocorre por meio da observação das características de um determinado fenômeno e das características coletadas.

Os Dados e as Variáveis

As variáveis se referem a qualquer quantidade ou característica de um determinado fenômeno que pode assumir diferentes valores numéricos. Já os dados podem ser definidos como um valor numérico, ou não numérico, atribuído a uma característica avaliada, que terá função de nortear os métodos estatísticos e, posteriormente, o tratamento estatístico.

O primeiro passo para a obtenção dos dados é a coleta de dados, que consiste na busca ou compilação dos dados das variáveis, componentes do fenômeno a ser estudado. A coleta de dados pode ocorrer de forma direta ou indireta.

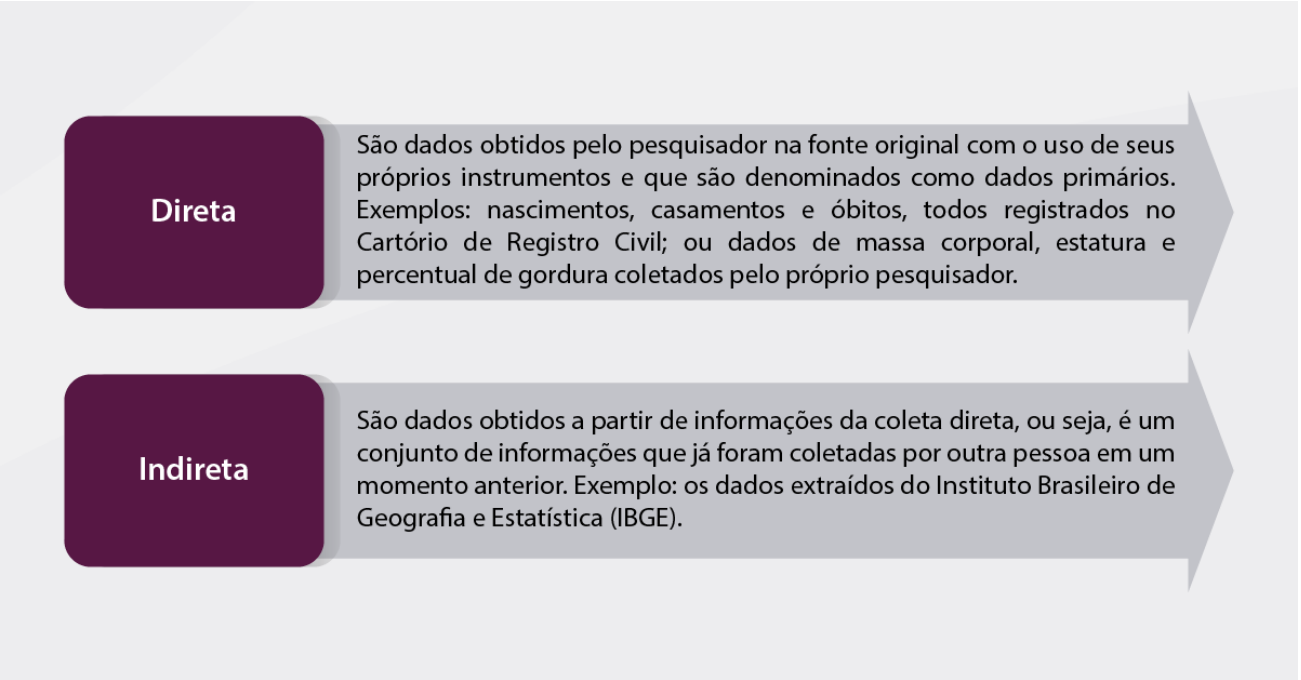


Figura 1 - Tipos de coleta de dados
Fonte: Elaborada pelo autor.

SAIBA MAIS

Dados primários são aqueles coletados a partir de uma amostra pelo próprio pesquisador ou sua equipe com a intenção de alcançar os objetivos da pesquisa. Dados secundários são aqueles oriundos de outros estudos e que estão à disposição do pesquisador. Algumas fontes de dados secundários são: internet, bancos de dados, jornais, revistas e filmes.

As características observadas de um fenômeno, que são denominadas de variáveis, podem ser classificadas em qualitativas/categóricas e quantitativas/numéricas (GLANTZ, 2014).

As variáveis quantitativas são aquelas características que podem ser expressas em valores numéricos. De uma forma geral, são dados obtidos de medições, contagens e experimentos. Alguns exemplos são: a massa corporal (quilogramas - kg); estatura (metros – m; centímetros – cm); idade (anos e meses) e Pressão Arterial (milímetros de mercúrio – mmHg). No entanto, algumas variáveis quantitativas podem ser mensuradas sem o emprego de unidades de medida e são denominadas adimensionais, como por exemplo, a nota atribuída em uma prova, na qual o aluno pode obter a nota 8,0 e não possui unidade de medida. As variáveis quantitativas são subdivididas em dois grupos, conforme a figura abaixo:

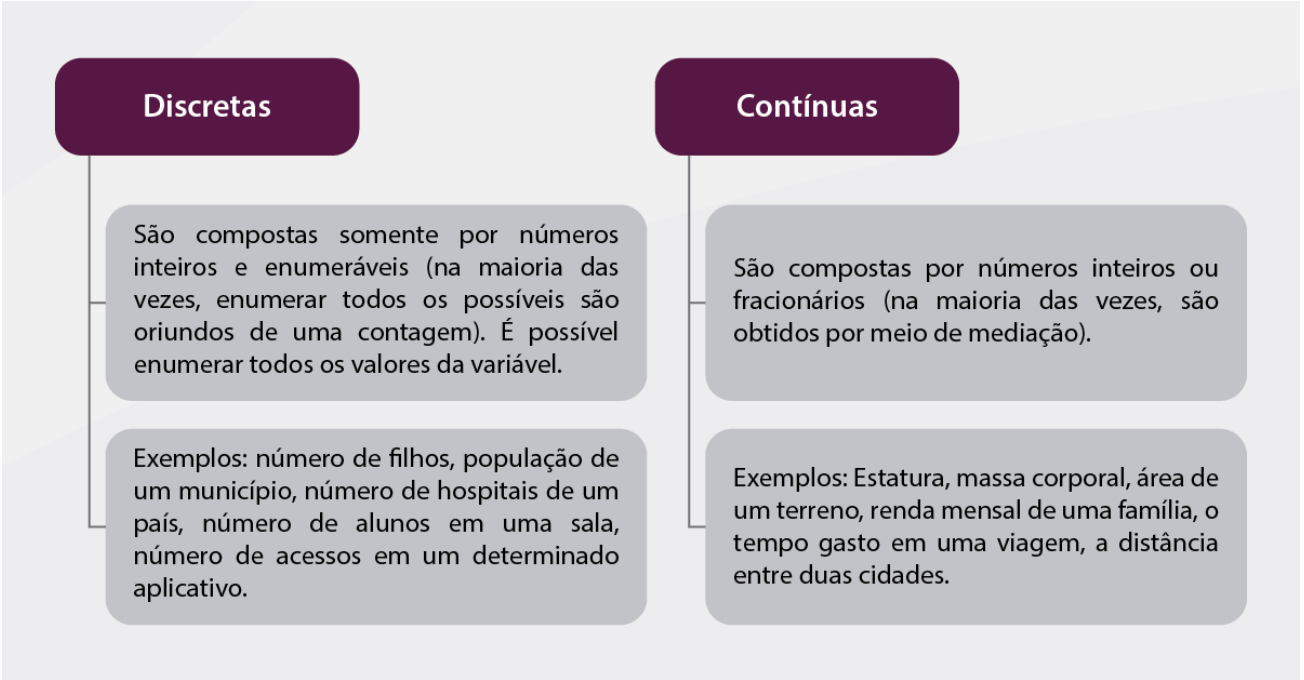


Figura 2 - Tipos de variáveis quantitativas
Fonte: Elaborada pelo autor.

As variáveis qualitativas envolvem as características de um fenômeno que podem ser aferidas por qualidades não numéricas, ou seja, quando o resultado da observação é apresentado na forma de qualidade ou atributo. Exemplos desse tipo de variável são o sexo das pessoas, a cor, o nível de escolaridade, o estado civil e a cor dos olhos. As variáveis qualitativas também são divididas em dois subgrupos:

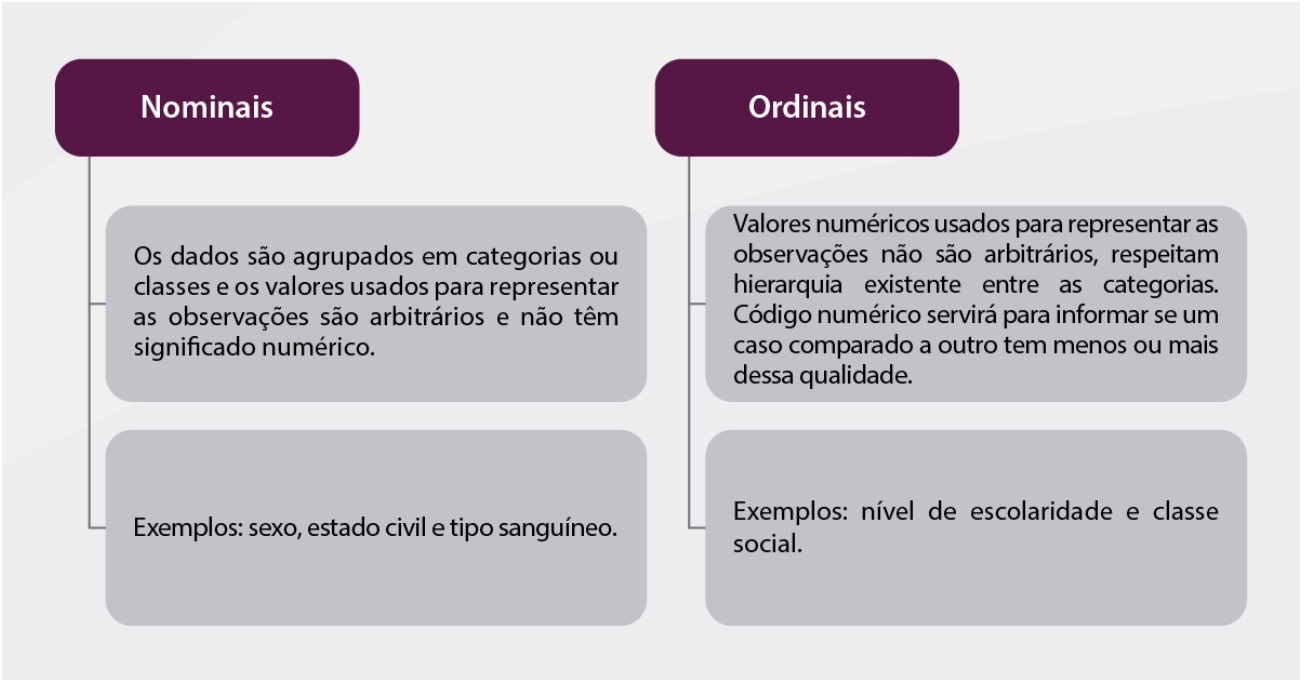


Figura 3 - Tipos de variáveis qualitativas
Fonte: Elaborada pelo autor.

Resumindo, as variáveis são classificadas da seguinte forma:

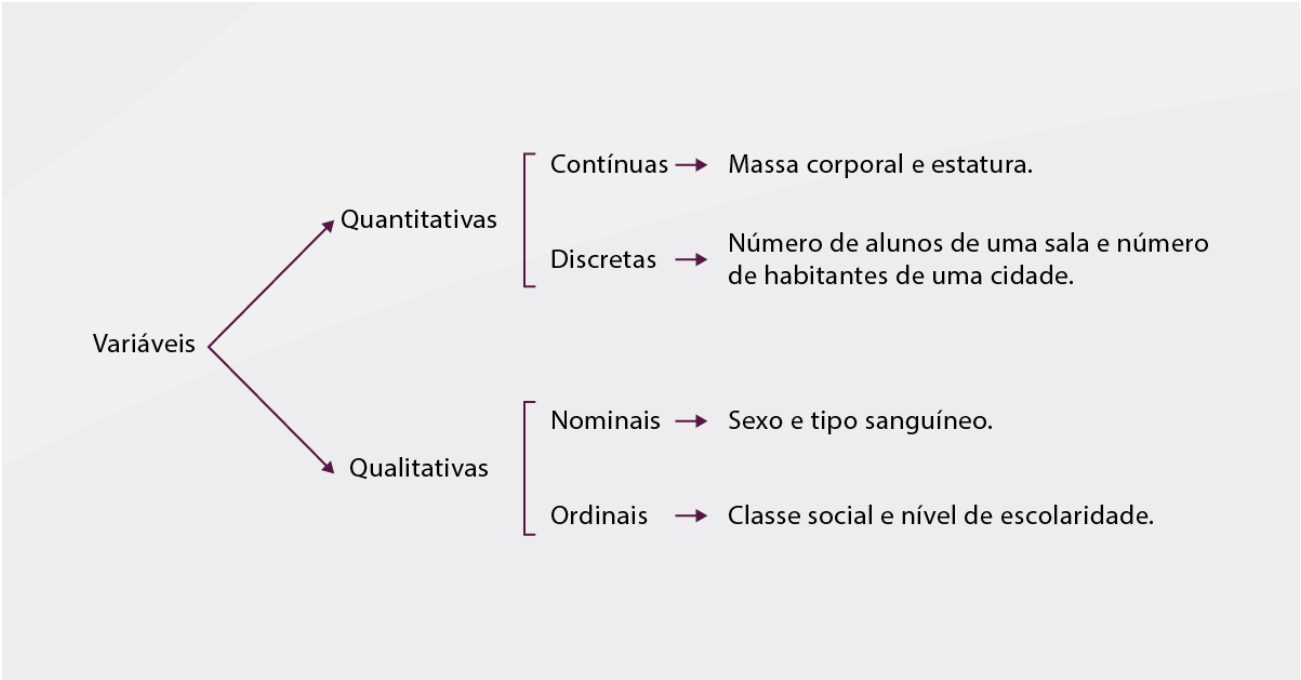


Figura 4 - Tipos de variáveis
Fonte: Elaborada pelo autor.

ATENÇÃO

Saber classificar uma característica avaliada em variável quantitativa ou qualitativa é fundamental para a condução do procedimento estatístico, uma vez que os procedimentos a serem adotados durante a análise estatística variam de acordo com os tipos de variáveis. Além disso, a classificação da variável depende do contexto. Para fins cadastrais, a variável idade pode ser quantitativa discreta, enquanto na Medicina pode ser contínua, pois a parte fracionária também é considerada.

A investigação das variáveis quantitativas e qualitativas pode ocorrer de duas maneiras:

- Investigando todos os elementos da população;
- Amostragem, ou seja, selecionando alguns elementos da população.

População e Amostra

A população é o conjunto de indivíduos, objetos ou informações que apresentam, pelo menos, uma característica comum, também chamada de universo estatístico (DANCEY; REIDY; ROWE, 2017). Um exemplo são os habitantes de um município,

pois a característica comum é o fato de residirem no mesmo município. Os estudantes de uma escola também são um exemplo.

Todavia, na maioria das vezes não é possível realizar o levantamento dos dados referentes a todos os elementos de uma população. Dessa forma, é analisada apenas uma parte representativa da população, isto é, uma amostra.

A amostra corresponde ao subconjunto finito e representativo de uma população, ou seja, é uma fração da população que permita ser examinada com a finalidade de se tirar conclusões a respeito da população em estudo (DANCEY; REIDY; ROWE, 2017). Para se obter uma boa amostra, utilizamos a técnica da amostragem.

Amostragem

A amostragem se refere à coleta das informações de parte da população, a amostra, a partir de métodos adequados de seleção. Destacam-se três tipos de amostragem: a simples, a sistemática e a estratificada.

Na amostragem simples todos os itens da população têm igual chance de pertencer à amostra. Este tipo de amostragem também é conhecido como aleatório (normalmente feita por sorteio).

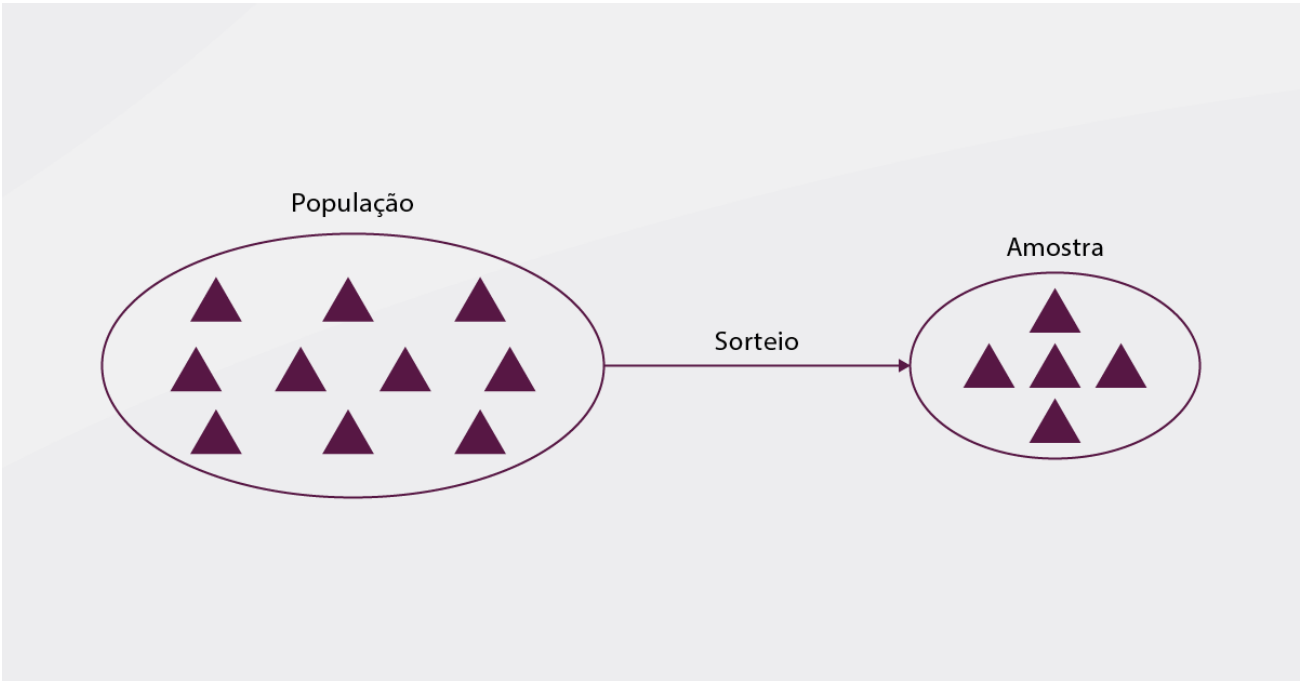


Figura 5 - Amostragem simples
Fonte: Elaborada pelo autor.

A amostragem sistemática é um processo no qual os itens se encontram ordenados e numerados e a seleção dos elementos da amostra é realizada periodicamente. No exemplo da figura a seguir, os elementos são selecionados em períodos de dois, iniciando do primeiro e em seguida coletando de dois em dois.

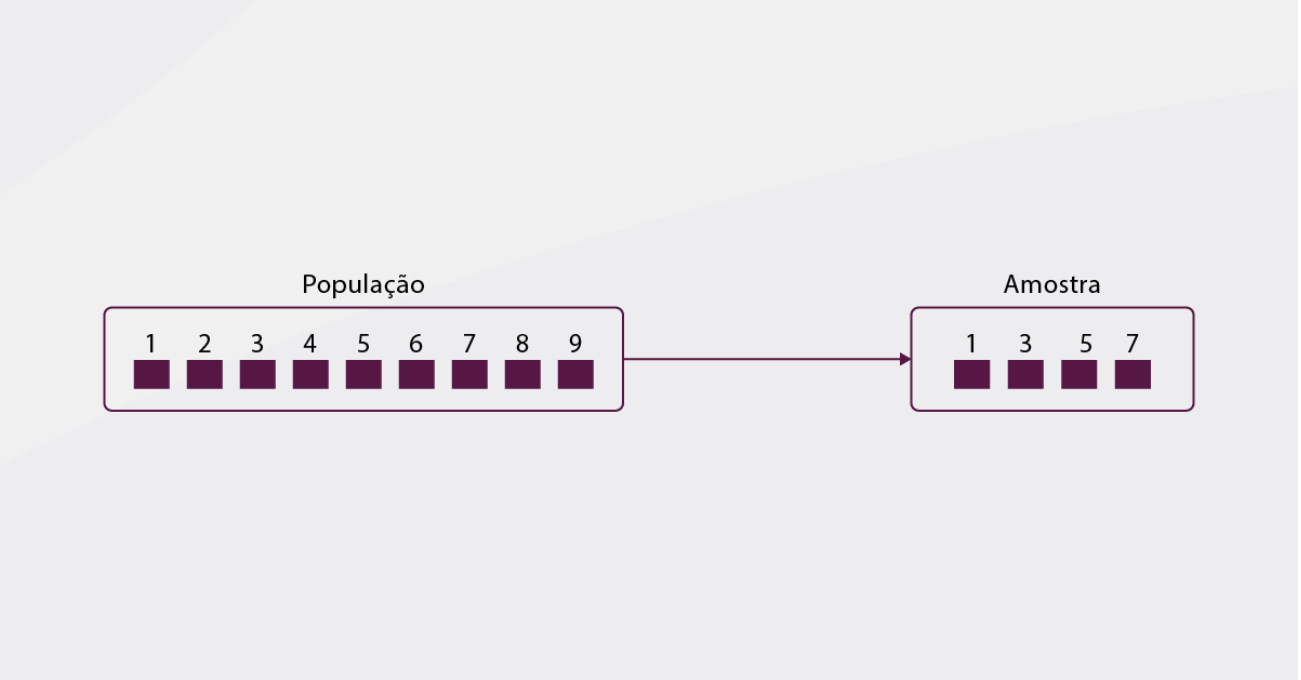


Figura 6 - Amostragem sistemática
Fonte: Elaborado pelo autor.

Já na amostragem estratificada, a população se encontra dividida em estratos e as amostras são selecionadas aleatoriamente de cada estrato. O estrato pode considerar uma série de fatores para serem definidos, tais como o sexo, a idade, a cor, o nível de escolaridade e outros.

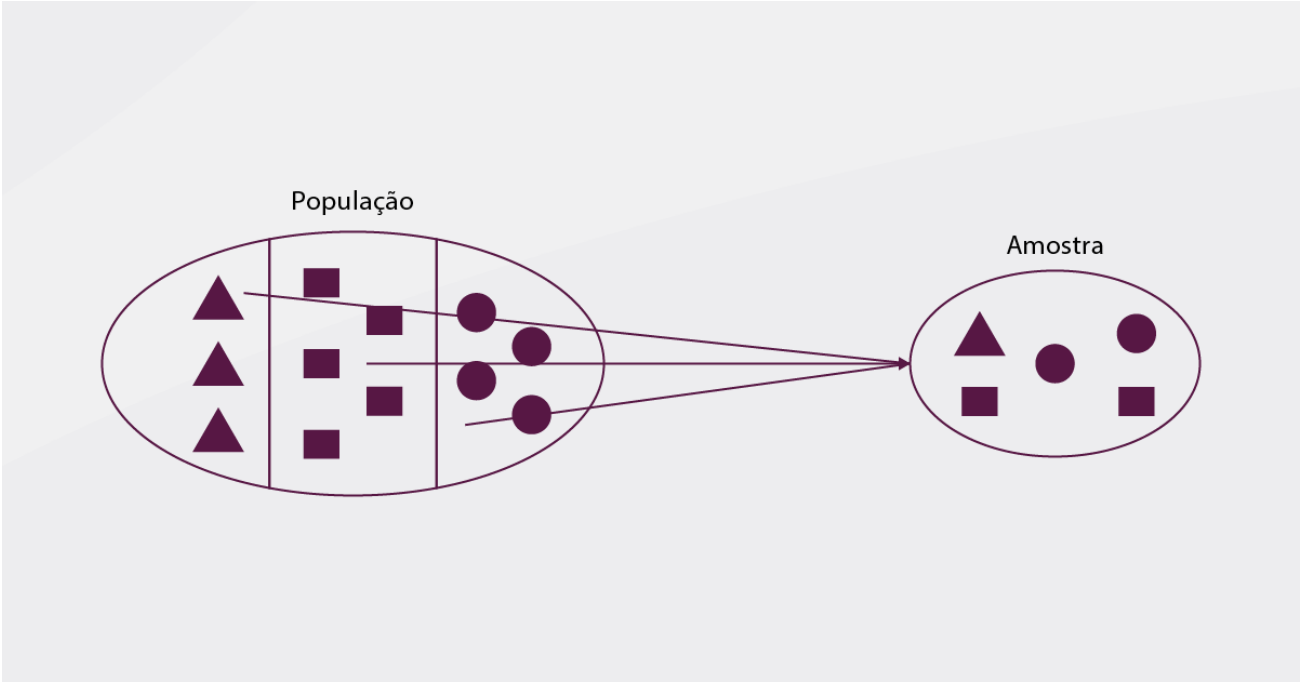


Figura 7 - Amostragem estratificada
Fonte: Elaborado pelo autor.

SAIBA MAIS

Após a leitura desta aula, você pode aprofundar o conhecimento a respeito das técnicas de amostragem assistindo o vídeo do link disponível em: <https://www.youtube.com/watch?v=vcTQCIfz9_M>. Acesso em: 26 abr. 2019.

Tipos de Estatística

A estatística consiste de métodos racionais para a obtenção de informações a respeito de um determinado fenômeno, favorecendo a obtenção de conclusões válidas para o fenômeno e permitindo a tomada de decisões por meio dos dados observados. Nessa perspectiva, a estatística pode ser dividida em dois ramos: a Estatística Descritiva e a Estatística Inferencial.

Estatística Descritiva

É o ramo da estatística que tem como finalidade a descrição dos dados observados. Geralmente, a estatística descritiva é a primeira a ser considerada em uma investigação científica e segue as seguintes etapas (DANCEY; REIDY; ROWE, 2017):

1. Obtenção dos dados estatísticos: realizada por meio de um questionário ou de observação/mensuração direta de uma população ou amostra;

2. Organização dos dados: ordenação dos dados e correção dos valores observados, falhas humanas e abandono de dados duvidosos;

3. Redução dos dados: compreensão dos dados por meio de simples leitura dos valores individuais;

4. Representação dos dados: engloba técnicas para a melhor visualização dos dados estatísticos, como os gráficos e tabelas;

5. Obtenção de informações para a descrição do fenômeno: são as informações obtidas dos dados que sumarizam os dados e facilitam a descrição dos fenômenos observados.

Estatística Inferencial ou Indutiva

É o ramo da estatística que trabalha com os dados de uma amostra de forma a estabelecer hipóteses, que podem ser assumidas ou rejeitadas, possibilitando a elaboração de conclusões científicas. A estatística inferencial tem como finalidade a análise e interpretação dos dados obtidos de um fenômeno.

SAIBA MAIS

Veja esta vídeoaula sobre os tipos de estatística. Além disso, você poderá adiantar o estudo das formas de representação da estatística descritiva, da aula 2. Disponível em: <https://www.youtube.com/watch?v=mvSFpCNgWAU>. Acesso em: 26 abr. 2019.

É importante ressaltar que a estatística descritiva e a estatística inferencial são abordagens complementares na análise de uma determinada característica de uma amostra.

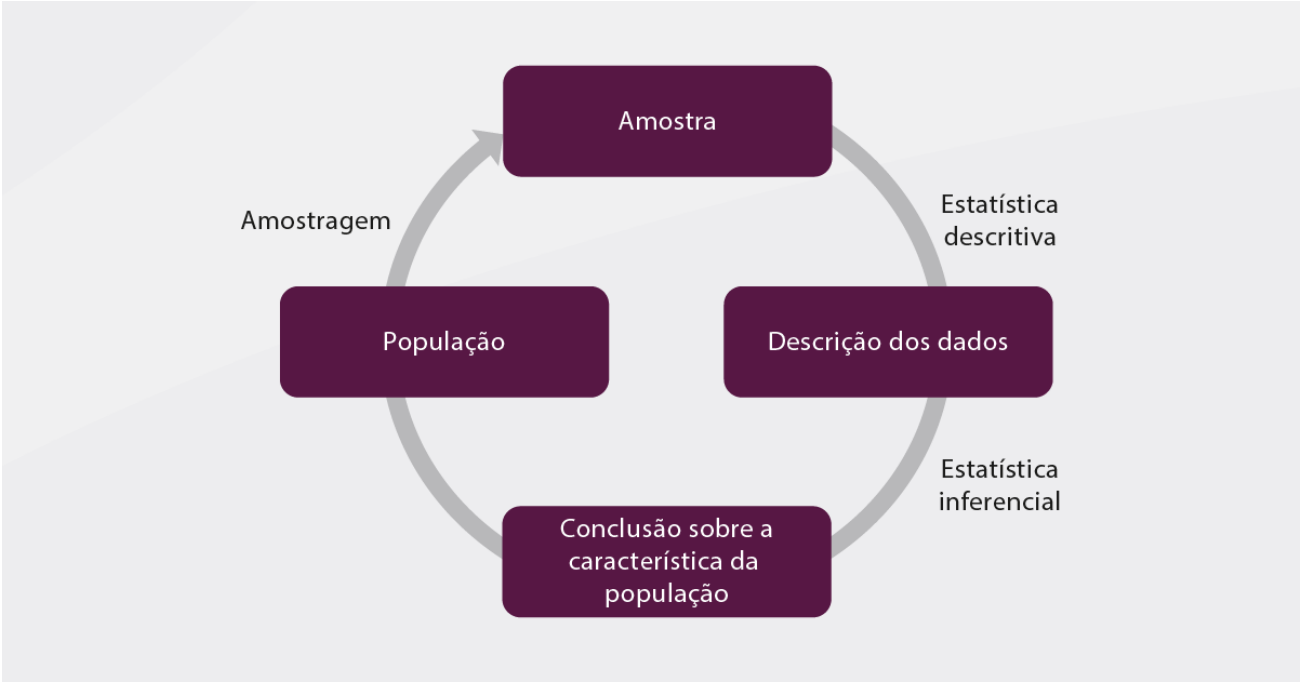


Figura 8 - Esquema de utilização da estatística descritiva e inferencial
Fonte: Elaborado pelo autor.

Estatística paramétrica e não Paramétrica

O conceito de estatística paramétrica e não-paramétrica é importante, pois será utilizado posteriormente durante a análise de normalidade dos dados e testes de hipóteses. A estatística paramétrica é o tipo de estatística inferencial empregado para conjuntos de dados com distribuição normal e homogeneidade de variância, ao passo que a estatística não paramétrica é utilizada para amostras que não atendem aos pressupostos de normalidade e homogeneidade (BARROS et al., 2012). A principal vantagem dos testes paramétricos envolve o maior poder em relação aos equivalentes não-paramétricos.

Fases do Método Estatístico

O método estatístico é dividido em seis fases, conforme demonstra o Quadro 1:

Definição do problema	Definição e delimitação do fenômeno que se pretende pesquisar.
Planejamento	Envolve os instrumentos de coleta de dados, as técnicas de amostragem, o cronograma de atividades,

	os custos envolvidos, entre outros.
Coleta de dados	É a fase operacional do método estatístico que envolve o registro sistemático de dados, com um objetivo determinado.
Crítica dos dados	Síntese dos dados por meio de sua contagem e agrupamento. Em outras palavras, é o processo de condensação e tabulação de dados.
Apresentação dos dados	Os dados podem ser apresentados por meio de tabelas e gráficos.
Análise e interpretação dos dados	É a fase mais delicada, visto que envolve o cálculo de medidas e coeficientes, cuja finalidade principal é descrever o fenômeno (estatística descritiva). Na estatística inferencial, a interpretação dos dados se fundamenta na teoria da probabilidade.

Quadro 1 -Fases do método estatístico
Fonte: Elaborado pelo autor.

Fechamento

A estatística é a ciência que se preocupa com a coleta, organização, descrição, análise e interpretação de dados experimentais. A terminologia utilizada na estatística depende dos fins pelos quais são utilizados os conceitos. Na área da saúde e biológica, por exemplo, a disciplina é denominada Bioestatística e abrange os dados relativos ao corpo humano, patologias, questões epidêmicas e assuntos similares. Alguns conceitos básicos são importantes para a correta condução do método estatístico, como população, amostra, tipos de variáveis e tipos de estatística.

Nesta aula, você teve a oportunidade de:

- compreender o processo que levou a estatística a ser considerada como uma ciência;
- reconhecer os conceitos básicos utilizados na estatística;
- compreender as fases dos métodos estatísticos e a utilização dos conceitos básicos em cada fase.

Aula 02

Estatística Descritiva: Tabelas e Gráficos

Introdução

Pense na seguinte atividade: planejar e desenvolver o seu trabalho de conclusão de curso na faculdade. Para isso, você precisará definir o fenômeno e o problema a ser investigado, a população e a amostra a ser selecionada, a forma de coleta de dados, a técnica de análise de dados e a medida de apresentação dos resultados.

Nesta aula, você vai conhecer como podemos representar as informações obtidas dos dados coletados de uma respectiva amostra, identificando quando podemos utilizar tabelas e/ou gráficos para representar os diferentes tipos de variáveis. Fique atento aos conceitos desta aula, pois, por meio deles, poderemos compreender melhor a apresentação das tabelas e gráficos encontrados em jornais e revistas e analisar e interpretar as informações transmitidas.

Ao final desta aula, você será capaz de:

- compreender a importância da organização dos dados estatísticos;
- desenvolver tabelas e gráficos de dados com informações relevantes;
- interpretar os dados de uma tabela e um gráfico.

Tabelas

Durante o processo de coleta dos dados, muitas vezes os pesquisadores não destinam muita atenção para a organização dos dados. Uma das melhores formas de organizar e apresentar os dados é por meio da utilização de tabelas, entretanto, em algumas situações, as tabelas não são utilizadas adequadamente. As técnicas que serão estudadas nesta aula permitirão detectar e corrigir erros e inconsistências ocorridos durante um processo de organização e representação dos dados.

As tabelas são quadros (sem que se fechem por completo as linhas e colunas, pois do contrário seria uma grade) organizados em forma de linhas e colunas que procuram, de forma clara e simples, expor os dados coletados e apresentar de forma detalhada e objetiva os resultados obtidos. Uma tabela é composta de seis elementos básicos:

- Corpo: refere-se às linhas e colunas que compõem as informações sobre a variável em estudo;
- Cabeçalho: é a parte superior da tabela, que indica o conteúdo de cada coluna da tabela;
- Coluna indicadora: é a parte da tabela que indica o conteúdo de cada linha da tabela;
- Casa ou célula: é o espaço indicado para os valores e números obtidos por meio dos dados;
- Título: é localizada no topo da tabela e deve contemplar o conjunto de informações da investigação. Os elementos do título devem responder às perguntas: O que?, Quando?, Onde?;

SAIBA MAIS

Para saber mais informações sobre os valores a serem inseridos nas células de uma tabela em situações específicas, como quando se tem casos ausentes, dúvidas sobre os valores ou ausência de dados, leia a Resolução nº 886, de 26 de outubro de 1966, que altera as normas para a Apresentação Tabular da Estatística Brasileira.

Fonte: IBGE (1967).

Séries Estatísticas

As séries estatísticas se referem a qualquer tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época (fator temporal), do local (fator geográfico) ou da espécie (fenômeno investigado). Em outras palavras, é uma sequência de números que se refere a uma certa variável em um determinado período do tempo e em uma localização geográfica específica. Conforme varie um desses elementos, a série estatística é classificada em temporal, geográfica e específica.

Série Temporal

Neste tipo de série os dados variam de acordo com o tempo, enquanto o local e a espécie (fenômeno) são elementos fixos. Esta série também é conhecida como

histórica ou evolutiva.

Ano	Preço médio em reais
2015	5,45
2016	5,77
2017	5,83
2018	5,78
2019	5,80

Tabela 1 - Preço do produto “X” no país “Y” nos últimos anos
Fonte: Elaborado pelo autor.

Série Geográfica

Nesta série o elemento variável é o fator geográfico. O tempo e o fenômeno são elementos fixos. A série geográfica também é denominada de espacial, territorial ou de localização.

Bairro	Preço médio em reais
Centro	5,76
Zona Sul	5,65
Zona Norte	5,82
Zona Leste	5,53
Zona Oeste	5,40

Tabela 2 -Preço do produto “X” na cidade “Y” em 2019
Fonte: Elaborado pelo autor.

Série Específica

Nesta série os dados estão de acordo com a espécie, isto é, variam em função do fenômeno investigado. A série específica também é chamada de categórica.

Ano	Preço médio em reais
Marca A	5,85
Marca B	5,73
Marca C	5,75
Marca D	5,63
Marca E	5,58

Tabela 3 - Preço do produto “X” das diferentes marcas
Fonte:Elaborado pelo autor.

Série Mista

As séries citadas anteriores podem ser combinadas, formando novas séries que são denominadas séries compostas ou mistas, as quais são apresentadas em tabelas de dupla entrada. Esta série possui duas ordens de classificação: uma horizontal e outra vertical. O exemplo abaixo é de uma série geográfica-temporal.

Bairro	Preço médio em reais em 2018	Preço médio em reais em 2019
Centro	5,76	5,82
Zona Sul	5,65	5,68
Zona Norte	5,82	5,90
Zona Leste	5,53	5,75

Zona Oeste	5,40	5,47

Tabela 4 - Preço do produto “X” na cidade “Y” nos anos de 2018 e 2019
Fonte: Elaborado pelo autor.

SAIBA MAIS

Veja esta videoaula para ver mais exemplos dos diferentes tipos de série estatística. Disponível em: <https://www.youtube.com/watch?v=4blPmTVjq3U>. Acesso em: 26 abr. 2019.

Gráficos

Os gráficos são representações visuais que têm como finalidade a representação de dados quantitativos ou qualitativos de uma determinada amostra, possuindo uma aplicação variada no cenário científico e na comunicação de uma forma geral. A

apresentação dos dados por meio de gráfico é um complemento importante da apresentação tabular. Os gráficos apresentam as mesmas informações de uma tabela de maneira mais simples e dinâmica, promovendo uma rápida visualização da distribuição dos valores ou das frequências observadas.

É fundamental que os gráficos atendam alguns requisitos básicos para que os dados apresentados sejam relevantes. Primeiramente, um gráfico não deve conter informações e traços desnecessários, para minimizar o excesso de informações. O gráfico deve representar de forma correta e adequada os dados estatísticos do fenômeno investigado. Por último, um gráfico deve expressar informações verídicas a respeito do fenômeno estudado. Além disso, todo gráfico deve possuir um título e uma escala para que seus valores sejam interpretados corretamente.

É importante saber, também, que a escolha de um gráfico deve ser de acordo com o tipo de variável utilizada e objetivo da investigação. Algumas das principais formas de representação gráfica são apresentadas a seguir.

Gráficos de Linhas

Os gráficos de linhas têm como finalidade revelar tendências e progressos ao longo do tempo. Este gráfico é muito utilizado quando se quer representar séries temporais que cobrem grande período de tempo. Este gráfico é mais adequado para a representação de dados contínuos e que possuem diferentes categorias. O gráfico de linhas pode ser usado para mostrar o número de pessoas com Zika Vírus ao longo de quatro anos em três estados brasileiros.

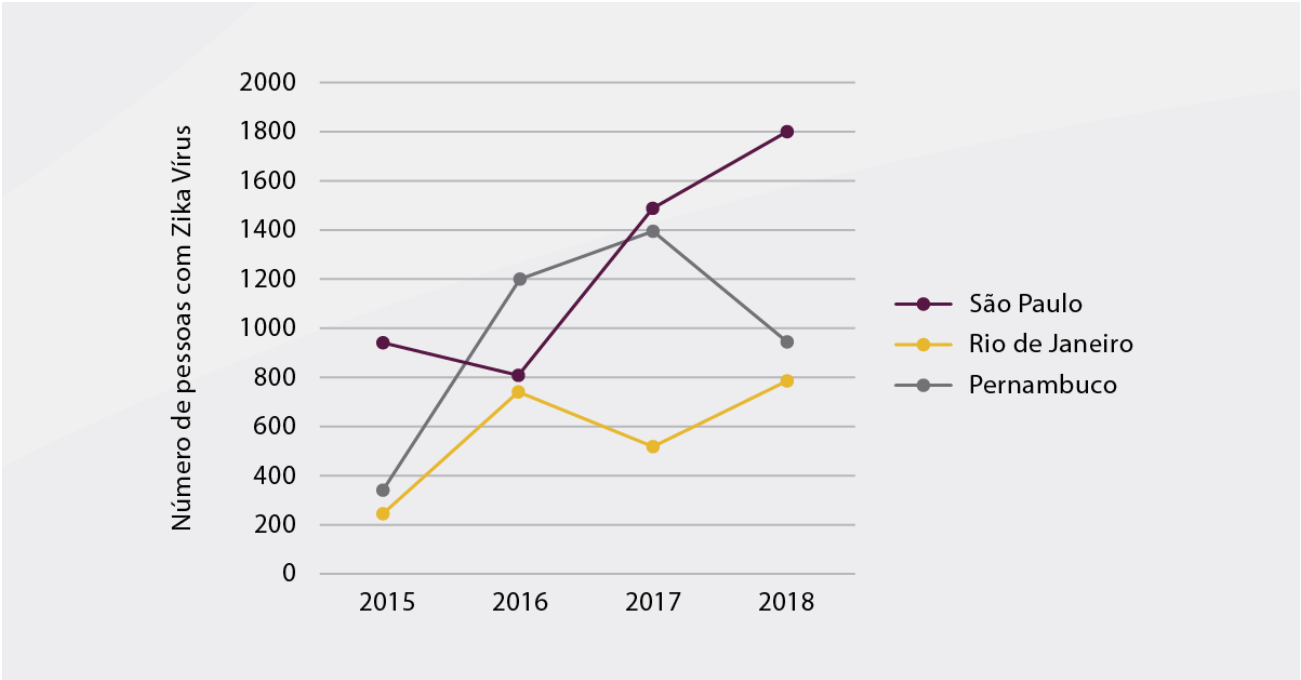


Figura 1 -Gráfico de linhas do número de pessoas com Zika Vírus ao longo de quatro anosem três estados brasileiros
Fonte:Elaborada pelo autor.

Gráficos em Colunas

O gráfico de colunas serve para representar séries estatísticas por meio de retângulos organizados em colunas verticais. Este tipo de gráfico pode ser utilizado para representar qualquer tipo de dado estatístico e chama mais atenção para os dados apresentados, uma vez que as colunas verticais podem ter grande espessura, sem que se perca a precisão na leitura e interpretação dos dados. No exemplo a seguir, é apresentada a taxa de mortalidade em três cidades diferentes no ano de 2018.

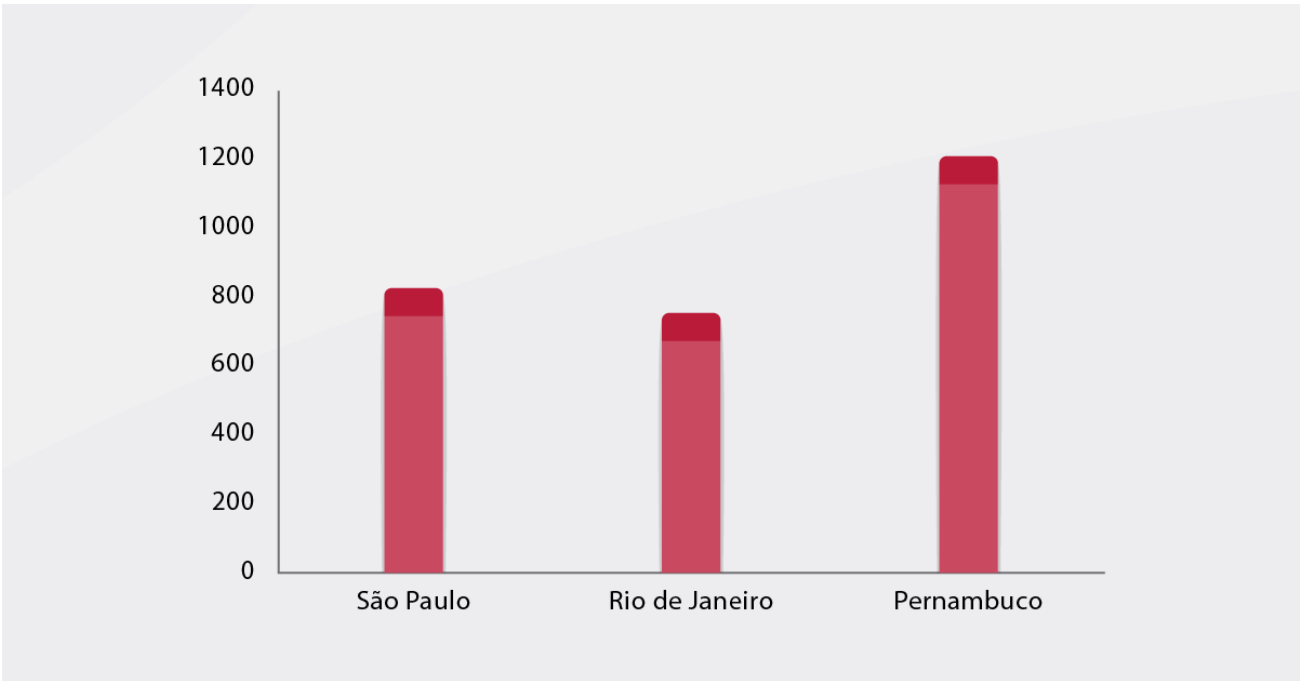


Figura 2 - Gráfico de colunas da taxa de mortalidade em três cidades diferentes no ano de 2018
Fonte:Elaborada pelo autor.

ATENÇÃO

A elaboração do gráfico de colunas deve respeitar algumas regras específicas. As bases das colunas devem ser iguais e as alturas proporcionais aos dados apresentados. Ainda, o espaço entre as colunas do gráfico deve variar entre ⅓ e ⅔ do tamanho da base da coluna, entretanto, esta medida depende do tipo de dado estatístico apresentado.

Gráficos em Barras

O gráfico de colunas tem características semelhantes às do gráfico de barras, sendo que a principal diferença é a disposição das barras, que devem estar na horizontal. Este modelo de gráfico é mais adequado para o realce da variação (mínimo e máximo) de duas ou mais variáveis no eixo vertical. Um exemplo pode ser a frequência de atendimentos na saúde pública, particular e por plano de saúde no ano de 2018 em uma cidade “X”.

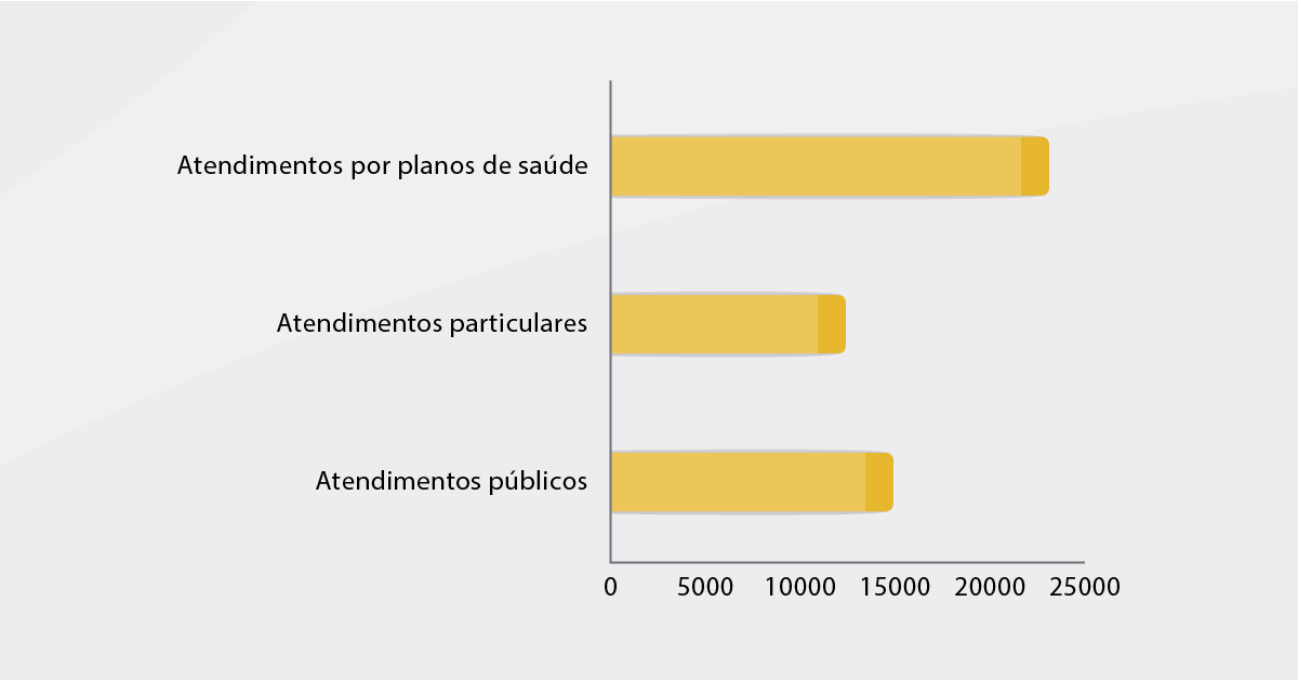


Figura 3 -Gráfico de barras da frequência de atendimentos na saúde pública, particular epor plano de saúde no ano de 2018 em uma cidade “X”
Fonte:Elaborado pelo autor.

Gráfico em Setores

Este modelo de gráfico também é conhecido como gráfico de “pizza” e a sua representação dos dados é simples e clara, uma vez que representa uma série estatística em um círculo por meio de setores com ângulos centrais proporcionais às frequências observadas. Este gráfico pode representar frequências absolutas e percentuais. Os principais tipos de séries estatísticas representados por gráficos de setores são as séries geográficas, específicas e as categorias em nível nominal, desde que não apresentem muitas categorias (no máximo sete). Os setores do gráfico são expressos em graus e o cálculo de cada setor é feito por meio de uma regra de três, conforme a fórmula a seguir:

$$Total -> 360^{o}$$

$$Setor -> X^{o}$$

Para se descobrir o setor de um percentual de 20,0% é realizado o seguinte cálculo:

$$100,0$$

$$20,0 -> X^{o}$$

$$100X = 7200$$

$$X = 7200/100$$

$$X = 72^{o}$$

Logo, um percentual de 20% representa 72° do total de 360° de um gráfico de setores.

Embora o cálculo de cada setor seja facilmente calculado pela regra de três, os softwares estatísticos fazem esse cálculo automaticamente a partir das frequências observadas da variável investigada. O exemplo a seguir apresenta os principais alimentos não saudáveis consumidos por crianças de uma cidade “X”.

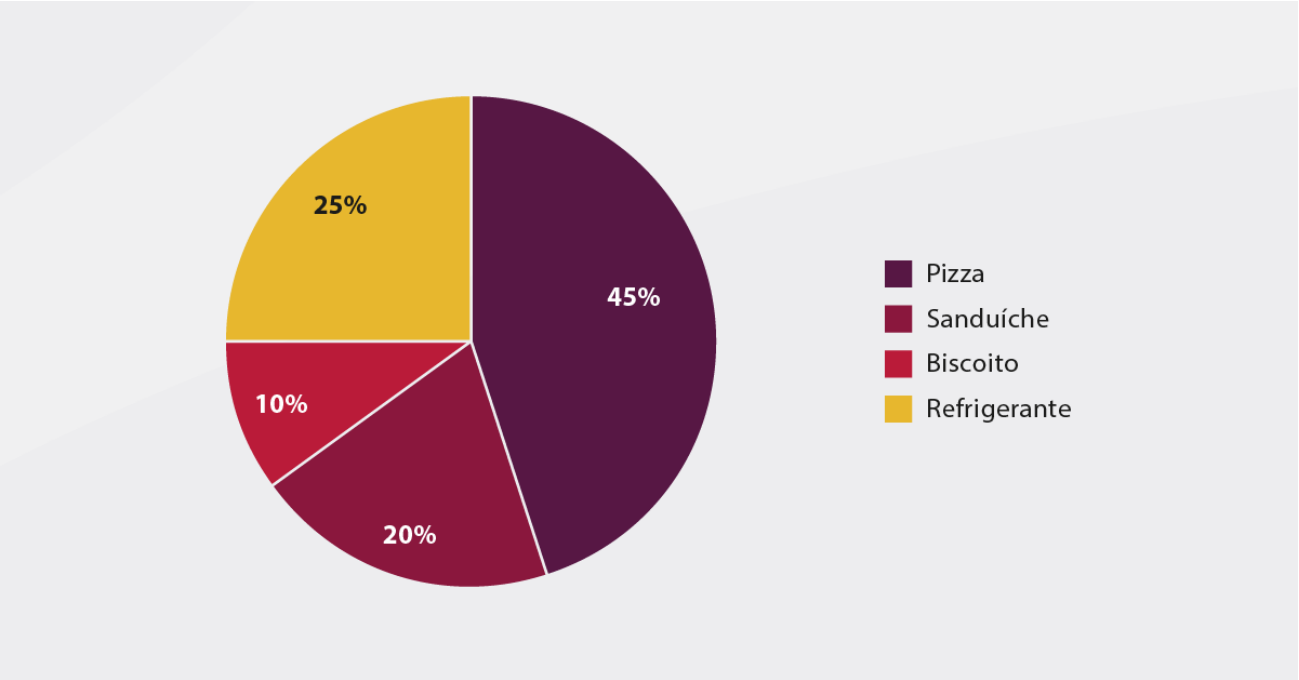


Figura 4 -Gráfico de setores da frequência alimentos não-saudáveis consumidos por crianças de uma cidade “X”
Fonte: Elaborada pelo autor.

SAIBA MAIS

Assista ao vídeo do link a seguir para aprender a elaborar os tipos de gráficos apresentados nesta aula no Microsoft Excel de forma fácil e prática. Disponível em: <https://www.youtube.com/watch?v=ExCPPp_YydE>. Acesso em 26 abr. 2019.

Estes são os principais tipos de gráficos existentes, entretanto, outros modelos de gráficos dinâmicos podem ser elaborados a partir de softwares estatísticos, os quais são, frequentemente, utilizados na área educacional e no cenário científico. É importante ressaltar que diferentes modelos de gráficos podem representar os mesmos dados estatísticos, entretanto, deve-se analisar cuidadosamente qual modelo representa melhor o dado estatístico investigado na pesquisa.

SAIBA MAIS

Para saber mais informações a respeito dos tipos de gráficos e como escolher o gráfico ideal para os diferentes dados estatísticos, leia o texto disponível no link disponível em: <https://www.matematica.pt/util/resumos/tipos-graficos-estatisticos.php>. Acesso em: 26 abr. 2019.

No exemplo a seguir é apresentada a opinião de uma amostra de estudantes universitários sobre qual o curso da saúde com maior nível de dificuldade, sendo obtidos os seguintes percentuais: Medicina (40%), Enfermagem (23%), Farmácia (18%), Educação Física (10%), Nutrição (8%) e Outros (1%). Estes dados podem ser apresentados por meio de diferentes tipos de gráficos, como o de setores e o de barras.

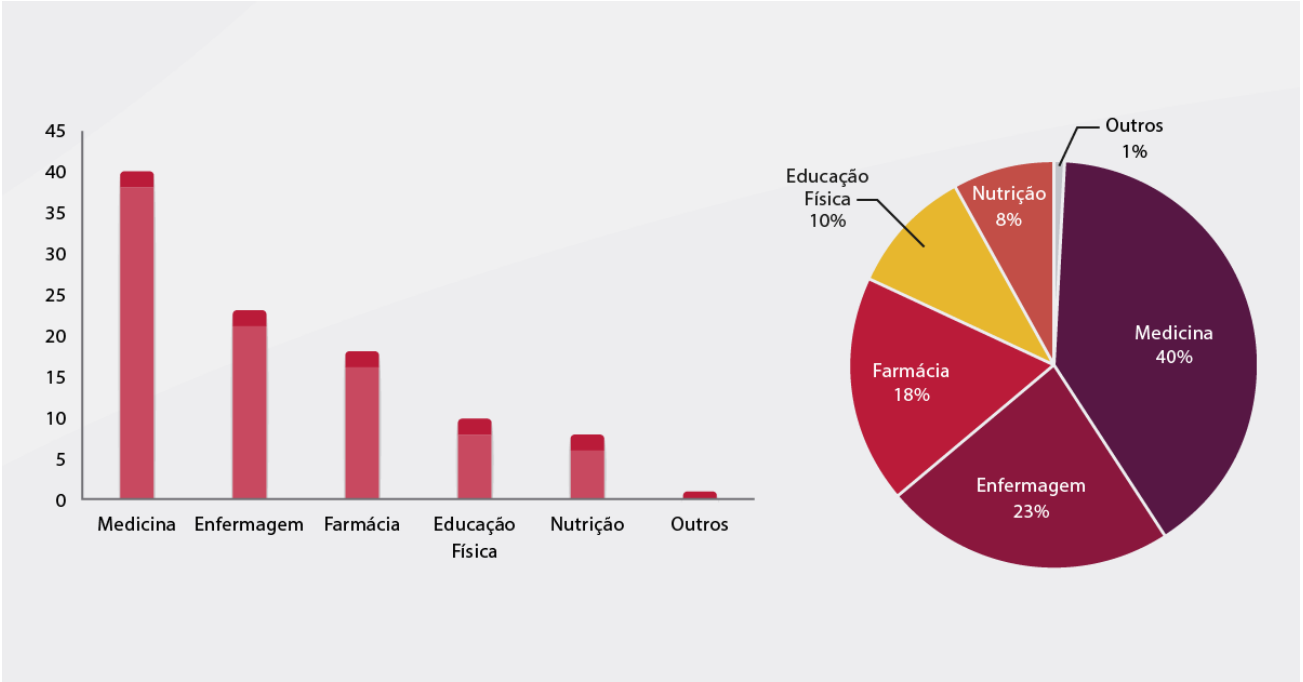


Figura 5 - Curso da área da saúde com maior nível de dificuldade de acordo com estudantes universitários de uma faculdade
Fonte: Elaborado pelo autor.

Nota-se que ambos os gráficos representam corretamente os mesmos dados estatísticos (opinião de estudantes universitários sobre qual o curso da saúde com maior nível de dificuldade), entretanto, é papel do pesquisador analisar criticamente qual modelo de gráfico é mais adequado para o dado estatístico investigado.

SAIBA MAIS

Apesar de ambos os gráficos utilizados (setores e barras) apresentarem corretamente os dados, qual deles representou melhor os dados obtidos pelo pesquisador?

Leve em consideração as características do gráfico, da série estatística e do fenômeno investigado para decidir qual o melhor gráfico para os dados apresentados.

Histograma

A única diferença do histograma para o gráfico de colunas é que o histograma não apresenta espaços entre as colunas, uma vez que é um gráfico para representar variáveis quantitativas (dados agrupados em intervalos). O intervalo de valores da variável (intervalos de classe) é representado pela largura da coluna, enquanto que a altura da coluna indica a frequência de dados dentro de cada intervalo de classe. Um exemplo pode ser a frequência de indivíduos com diferentes idades de uma sala de aula".

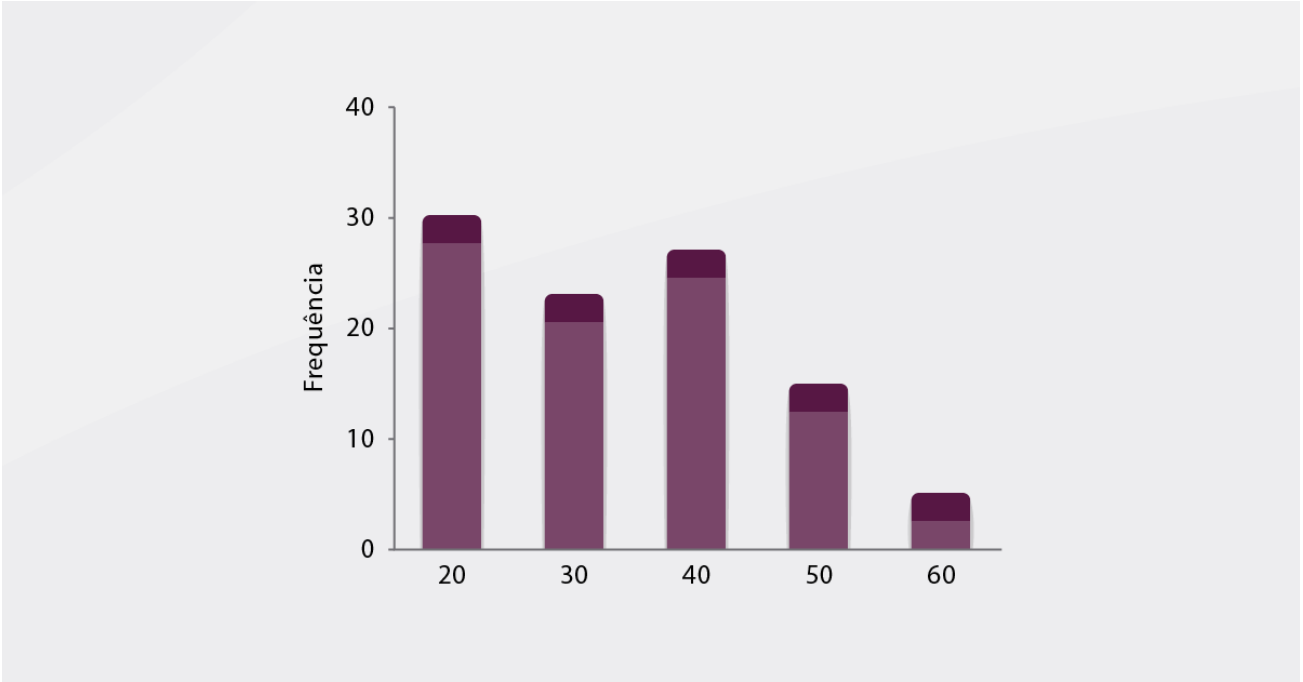


Figura 6 - Histograma da frequência de indivíduos com diferentes idades de uma sala de aula
Fonte:Elaborado pelo autor.

Fechamento

Nesta aula você aprendeu que a organização e apresentação de dados estatísticos é de extrema importância para que os resultados sejam interpretados de forma correta, sendo as tabelas e os gráficos as principais formas de apresentação dos dados estatísticos.

As tabelas são utilizadas para a melhor organização dos dados, além de proporcionarem a primeira visualização dos resultados. Ressalta-se que toda tabela apresenta a distribuição de um conjunto de dados estatísticos em função da época (fator tempo), local (fator geográfico) ou da espécie (fenômeno).

Já os gráficos são utilizados como um complemento das tabelas e devem ser simples e objetivos. Os modelos gráficos têm sido muito utilizados para a representação de dados quantitativos e qualitativos, visto que possuem melhor estética do que as tabelas e facilitam a interpretação dos dados. Além disso, existem diversos recursos da informática que facilitam a elaboração de gráficos dinâmicos.

Nesta aula, você teve a oportunidade de:

- compreender a importância da organização dos dados em uma pesquisa;
- desenvolver tabelas e gráficos com informações relevantes de acordo com o tipo de dado investigado;
- interpretar os dados de uma tabela e um gráfico.

Aula 03

Estatística Descritiva: Distribuição De Frequências

Introdução

Você sabia que os dados obtidos em uma pesquisa precisam ser devidamente organizados para que os resultados sejam claros e precisos? Muitos pesquisadores e estudantes não prestam atenção nesse importante detalhe que interfere diretamente nos resultados de uma pesquisa. Nesta aula veremos como um conjunto de dados de uma determinada amostra pode ser organizado de forma adequada com o intuito de proporcionar informações relevantes e claras para o pesquisador. Uma das técnicas mais eficientes na estatística descritiva, para organizar os dados e apresentar resultados claros, é a distribuição de frequência.

Ao final desta aula, você será capaz de:

- reconhecer os tipos de distribuição de frequência;
- compreender as características da distribuição de frequência com intervalos de classe;
- utilizar os diferentes tipos de frequência para apresentar dados estatísticos.

Conceitos Básicos de Distribuição de Frequências

Quando se investiga uma variável (qualitativa ou quantitativa), o principal objetivo do pesquisador é conhecer a distribuição dessa variável por meio dos possíveis valores dela. Nessa perspectiva, a distribuição de frequências se apresenta como uma das principais maneiras de se representar um conjunto de valores. Trata-se de uma série estatística específica em que os dados se encontram dispostos em classes ou categorias juntamente com suas respectivas frequências. Neste caso, todos os elementos da tabela (tempo, local e fenômeno) são fixos, entretanto, os dados relacionados aos fenômenos são apresentados de acordo com sua magnitude.

Dois conceitos são importantes quando se estuda as distribuições de frequência, os dados brutos e o rol. Os dados brutos se referem ao conjunto dos dados numéricos obtidos após a crítica dos valores coletados e que não foram numericamente organizados. Os valores a seguir poderiam ser os dados brutos de um conjunto de dados: 25, 23, 28, 30, 24, 20, 35.

Já o rol é o arranjo dos dados brutos em ordem de frequência crescente ou decrescente, ou seja, é a tabela obtida após a ordenação dos dados. Os dados brutos citados anteriormente ficariam assim em forma de rol: 20, 23, 24, 25, 28, 30, 35. A contagem dos valores do rol para a tabela de frequências deve ser feita cuidadosamente, visto que um erro na contagem pode gerar análises equivocadas e valores errados de todas os resultados da tabela.

SAIBA MAIS

Assista o vídeo do link a seguir para saber mais sobre os conceitos de dados brutos e a importância do rol para os diferentes tipos de distribuição de frequência. Disponível em: <https://www.youtube.com/watch?v=QZGQk56aHOw>. Acesso em: 26 abr. 2019.

As distribuições de frequência podem ser divididas em dois tipos: sem intervalos de classe e com intervalos de classe.

Distribuição de Frequência sem Intervalos de Classe

Esta distribuição de frequência engloba a simples condensação dos dados de acordo com as repetições de seus valores. Geralmente é utilizada para variáveis qualitativas ou quantitativas discretas com poucos valores diferentes. Um exemplo pode ser a quantidade de demissões em pequenas empresas em uma cidade no ano de 2018.

Número de demissões (Xi)	Número de empresas (fi)
0	5
1	5

2	10
3	30
4 ou mais	50
Total	100

Tabela 1 -Número de demissões em pequenas empresas da cidade “X” em 2018
Fonte:Elaborado pelo autor.

Nota:

Xi = são as categorias em que o fato se subdivide.

fi = é a frequência absoluta, isto é, o número de vezes que cada uma das categorias ocorre.

N = soma dos fi = total de elementos observados na população.

n = soma dos fi = total de elementos observados na amostra.

Distribuição de Frequência com Intervalos de Classe

Quando o tamanho da amostra é elevado, é mais viável efetuar o agrupamento dos valores em vários intervalos de classe. É muito utilizada para variáveis quantitativas contínuas ou discretas com muitos valores diferentes. Um exemplo pode ser a nota da disciplina de Estatística de 50 estudantes de uma faculdade.

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

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

X = rol de notas dos 50 estudantes da faculdade.

A partir destes dados, o pesquisador deve organizar os dados em um rol para facilitar a contagem dos valores e a posterior distribuição de frequência. Assim, os dados brutos apresentados ficariam da seguinte forma quando organizados em rol:

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

X = rol de notas dos 50 estudantes da faculdade.

Com os dados organizados em rol, faz-se a contagem dos valores e os dados são expressos pela seguinte tabela:

Notas)	fi
0 ⊢ 1,0	3
1,0 ⊢ 2,0	6
2,0 ⊢ 3,0	6

3,0 ⊢ 4,0	8
4,0 ⊢ 5,0	11
5,0 ⊢ 6,0	8
6,0 ⊢ 7,0	4
7,0 ⊢ 8,0	4
Total	50

Tabela 2 - Notas finais dos estudantes da disciplina de Estatística –2009
Fonte: Elaborada pelo autor.

Nota: fi = frequência absoluta das classes.

ATENÇÃO

O símbolo que representa o intervalo de classes possui informações importantes a respeito do intervalo, conforme mostra a figura abaixo:

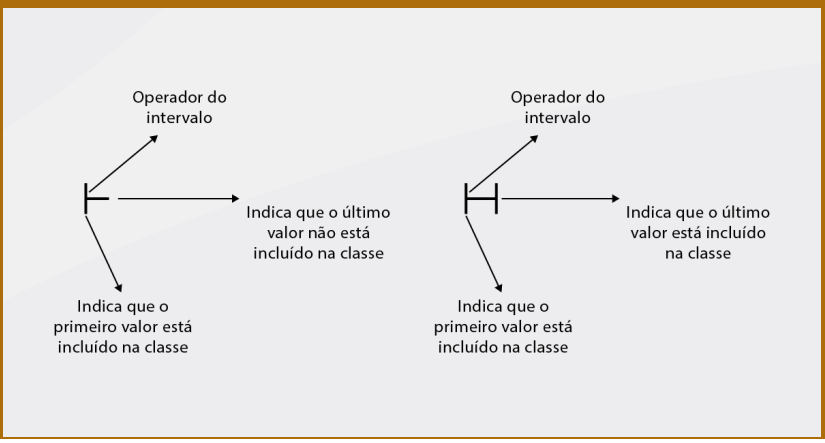


Figura 1 - Significados de cada parte do símbolo que representa o intervalo de classes
Fonte: Elaborado pelo autor.

SAIBA MAIS

Assista ao vídeo do link a seguir para saber mais sobre os intervalos de classes e adiantar o estudo a respeito dos elementos de uma distribuição de frequência com classes. Disponível em: <https://www.youtube.com/watch?v=SFWw6PiMxRw>. Acesso em: 26 abr. 2016.

Elementos de uma distribuição de frequência (com intervalos de classe)

Classes

As classes correspondem aos grupos de valores em que se subdivide a amplitude total do conjunto de valores observados da variável. Uma classe é representada pelo símbolo “i” (i = 1, 2, 3, 4, 5...) e o total de classes da distribuição de frequências é representado pelo símbolo “k”. Se levarmos em consideração o exemplo apontado na Tabela 2, podemos dizer que a distribuição é formada por oito classes (k = 8) e o intervalo 4,0 ⊢ 5,0 representa a quinta classe (i = 5).

Conforme ilustrado na Figura 1, quando o símbolo do intervalo de classes está fechado à esquerda e aberto à direita, significa que o número à esquerda pertence à classe e o número à direita não pertence. Quando o símbolo está fechado à direita e à esquerda, significa que tanto o número à esquerda quanto o número à direita fazem parte da classe.

Limite de Classe

O limite de classes representa os extremos de cada classe de uma distribuição de frequência, sendo que o menor valor é o limite inferior (li) e o maior valor é o limite superior da classe (ls). A partir do exemplo da Tabela 2, destacam-se os seguintes limites da segunda classe $(i = 2) : li_2 = 1,0 e ls_2 = 2,0.$

Amplitude de um Intervalo de Classe e Amplitude Total

A amplitude de um intervalo de classe (h) é a diferença entre os limites superior e inferior da classe. O cálculo da amplitude é realizado a partir da seguinte fórmula:

$$h = ls - li$$

A amplitude da classe apresentada acima (i = 2) é 1,0, pois é calculado a diferença entre o limite superior (ls = 2,0) e o limite inferior (li = 1,0).

ATENÇÃO

Uma distribuição de frequência com intervalos de classe sempre terá a mesma amplitude para todas as classes. Note que para todos os intervalos do exemplo da Tabela 2 a amplitude é 1,0.

Já a amplitude total da distribuição de frequência é a diferença entre o valor máximo e o valor mínimo da amostra. Para isso, é só identificar o menor e o maior valor do rol. Note que o menor valor da amostra do exemplo da Tabela 2 é 0,5 e o maior valor é 7,9. Logo, a amplitude total é 7,4.

Ponto Médio de uma Classe (xi)

O ponto médio de uma classe é o ponto que divide o intervalo de classe em duas partes iguais, e que pode ser calculado a partir da fórmula abaixo:

$$xi = \frac{li = ls}{2}$$

Vamos calcular o ponto médio da sétima classe (i = 7) do exemplo apresentado na Tabela 2. Assim, temos: li = 6,0 e ls = 7,0 e o cálculo do ponto médio fica da seguinte forma:

$$xi = 6,0 + 7,0/2$$

$$xi = 13,0/2$$

$$xi = 6,5$$

Número de Classes

Definir o número de classe em uma distribuição de frequências é uma etapa importante, visto que ao adotar poucas classes, você perderá detalhe de informação, enquanto que se você utilizar um número muito grande de classes, pode ser que alguma das classes apresente frequência nula ou muito baixa.

Existem vários critérios e fórmulas que podem ser utilizados para determinar o número de classes em uma distribuição de frequências, porém todos eles servem apenas como uma indicação, uma vez que o pesquisador é o responsável por tal definição.

Uma solução é a definição a partir do tamanho da amostra, adotando o seguinte critério:

$$n \leq 25 \rightarrow K = 5$$

$$n > 25 \rightarrow K\sqrt{n}$$

Para este critério, se a amostra for de 22 sujeitos, o número de classe é 5. No entanto, se a amostra for de 125 sujeitos, número de classe para a distribuição de frequência é calculado por meio da seguinte fórmula:

$$K\sqrt{125}$$

$$K = 11,18, arredondando:$$

$$K = 11$$

Outra solução é utilizar a Fórmula de Sturges, que estabelece que o número de classes a partir da seguinte fórmula:

$$K = 1 + 3,3log\ n$$

$$onden = tamanhodaamostra$$

No mesmo exemplo utilizado anteriormente para uma amostra de 125 sujeitos, teríamos o seguinte cálculo:

$$K = 1 + 3,3log(125)$$

$$K = 1 + 3,3(2,09)$$

$$K = 1 + 6,90$$

$$K = 7,90, arredondando:$$

$$K = 8 classes$$

SAIBA MAIS

Assista ao vídeo do link a seguir para saber mais sobre a aplicação da fórmula de Sturges para determinar o número de de classes em uma distribuição de frequência. Disponível em: <https://www.youtube.com/watch?v=J8W2Kj5xnsU>. Acesso em: 26 abr. 2019.

ATENÇÃO

Apesar de ser um bom parâmetro para o cálculo do número de classes, a fórmula de Sturges propõe um número demasiado de classes para uma amostra pequena e, relativamente, poucas classes quando a amostra é muito grande.

Amplitude do Intervalo de Classes

A amplitude do intervalo de classes (Ai) se refere ao comprimento de cada classe. O cálculo da amplitude do intervalo é feito por meio da fórmula abaixo:

$$Ai = \frac{H}{K}$$

Ao considerar novamente o exemplo da Tabela 2 nesta aula, vimos que o menor valor é 0,5 e o maior valor é 7,9 e a amplitude total é 7,4. Utilizando a fórmula de Sturges, teríamos sete classes (k = 7). Assim, a fórmula da amplitude ficaria da seguinte forma:

$$Ai = 7,4/7$$

$$Ai = 1,05$$

Logo o intervalo de cada classe seria 1,0.

Tipos de Frequências

Frequência Absoluta ou Simples

A frequência absoluta (fi) é o número de ocorrências para cada uma das classes, obtida por meio da contagem no rol. A soma das frequências simples representa o total dos dados da distribuição.

Frequência relativa

A frequência relativa (fri) é a divisão da frequência simples com a soma das frequências da classe, que fornece o percentual de cada classe em relação ao número total de observações. A soma das frequências relativas varia entre 0,0% e 100%.

Frequência Simples e Relativa Acumulada

A frequência simples acumulada (faci) corresponde à soma das frequências até a classe indicada, enquanto que a frequência relativa acumulada (fraci) se refere à divisão da frequência acumulada da classe pela frequência total da amostra.

SAIBA MAIS

Para mais informações sobre distribuição de frequência e a utilização dos diferentes tipos de frequência para a organização e apresentação de dados estatísticos, leia o livro Estatística Fácil, de Antônio Arnot Crespo.

Fechamento

Nesta aula percebemos que a organização dos dados por meio da distribuição de frequências facilita o entendimento dos dados estatísticos de uma pesquisa, evidenciando ser uma ferramenta importante para ser utilizada em trabalhos científicos sempre que o objetivo for organizar e sintetizar dados de uma pesquisa.

Destaca-se que muitos conceitos devem ser levados em consideração para elaborar uma distribuição de frequências de forma correta, como os intervalos, limites e amplitudes de classe, os métodos para determinar o número de classes e os tipos de frequências. Os diferentes tipos de frequências são medidas úteis para a apresentação dos dados de forma objetiva e simples, favorecendo a compreensão dos dados da pesquisa.

Nesta aula, você teve a oportunidade de:

- reconhecer a importância dos tipos de distribuição de frequência para a organização e apresentação dos dados;
- compreender as características da distribuição de frequência com intervalos de classe;
- utilizar os diferentes tipos de frequência para apresentar dados estatísticos.

Aula 04

Estatística Descritiva: Medidas de Tendência Central e Dispersão

Introdução

O termo “média” é um dos mais utilizados no dia a dia pelas pessoas. Todos os dias utilizamos este termo em frases, como: “eu durmo em média 8 horas por noite” ou “eu pratico, em média, 30 minutos de atividade física por dia”. Em todas essas situações utilizamos a média aritmética, entretanto, veremos nessa aula que também existem a média ponderada, a média harmônica, a média geométrica, além da mediana e da moda.

Sempre que quisermos representar e caracterizar um determinado conjunto de dados que seja constituído de variáveis quantitativas, nós utilizaremos uma dessas medidas, as quais são chamadas de “Medidas de Tendência Central”. Além disso, é possível verificar as diferenças e a homogeneidade entre os valores de variáveis quantitativas por meio da utilização das “Medidas de Dispersão”, que são medidas complementares às medidas de tendência central.

Ao final desta aula, você será capaz de:

- reconhecer as principais medidas de tendência central e dispersão;
- compreender a utilização de cada medida de tendência central e dispersão;

- aplicar as principais medidas de tendência central e dispersão.

Medidas de Tendência Central

Após abordar a sintetização dos dados sob a forma de tabelas, gráficos e distribuições de frequências, vamos aprender a calcular as medidas que possibilitam representar um conjunto de dados de uma variável quantitativa. As principais medidas de tendência central são as Médias, a Moda e a Mediana.

Médias

As médias são as medidas de tendência central mais utilizadas para representar dados quantitativos. No entanto, é importante saber que existem diferentes tipos de média, como a média aritmética simples, a média aritmética ponderada, a média harmônica e a média geométrica (DANCEY; REIDY; ROWE, 2017). Embora seja amplamente utilizada, a média é uma medida de tendência central que, por uniformizar os valores de um conjunto de observações, não representa bem os conjuntos de valores extremos, ou seja, com valores muito altos ou muito baixos.

ATENÇÃO

Quando valores extremos e discrepantes influenciarem o valor da média, é possível utilizar outras medidas de tendência central, como a mediana. Mais a frente, nesta aula, você verá de forma detalhada estes conceitos.

Média Aritmética Simples

A média aritmética é a medida de tendência central mais frequente nas representações de um conjunto de observações. A média simples de um conjunto de observações corresponde à soma de todos os dados da amostra dividida pelo número de elementos da amostra, conforme demonstra a fórmula a seguir:

$$Media = \frac{soma\ dos\ dados\ (valores\ observados)}{número\ total\ de\ observações}$$

Exemplo prático: a partir da Tabela 1, calcule a média aritmética da massa corporal dos estudantes de enfermagem de uma faculdade.

Alunos (massa corporal)	Frequência
55	2

60	2
65	2
70	1
75	1
Total	8

Tabela 1 - Massa corporal dos estudantes de enfermagem de uma faculdade
Fonte: Elaborado pelo autor.

Resolução:

$$média = \frac{55 + 55 + 60 + 60 + 65 + 65 + 70 + 75}{8}$$

$$média = \frac{505}{8}$$

$$média = 63,125$$

Média Aritmética Ponderada

A média ponderada é utilizada quando se pretende sintetizar valores que têm diferentes graus de importância. Este tipo de média é obtido pela soma de todos os valores do conjunto de observações multiplicado pelos seus respectivos pesos ou ponderações (p) dividida pelo somatório dos pesos ou ponderações (p), conforme a fórmula a seguir:

$$Mp = \frac{a_1p_1 + a_2p_2 + a_3p_3 + \ldots\ldots a_np_n}{p_1 + p_2 + p_3 + \ldots\ldots p_n}$$

Exemplo prático: o professor da disciplina de epidemiologia de uma faculdade aplicou duas avaliações nos alunos, entretanto, a primeira avaliação tinha peso 3 e a segunda avaliação tinha peso 2. A média mínima para ser aprovado é 6,0. Um dos alunos tirou 5,5 na primeira avaliação e 7,0 na segunda avaliação. Sabendo dessas informações, é possível afirmar que o aluno foi aprovado sem a necessidade de exame final?

Resolução:

$$média = \frac{5,5.3 + 7,0.2}{5}$$

$$média = \frac{30,5}{5}$$

$$média = 6,1$$

A média ponderada do aluno foi 6,1, sendo aprovado na disciplina sem a necessidade de exame final.

Média Harmônica

A média harmônica corresponde ao inverso da média aritmética dos inversos e é calculada a partir da seguinte fórmula:

$$Mh = \frac{n(númerodeelementosdaamostra)}{1/a_1 + 1/a_2 + 1/a_3 + \ldots 1/a_n}$$

Exemplo prático: determine a média harmônica da idade dos estudantes de enfermagem de uma faculdade.

Idade dos alunos	Identificação
25	A
22	B
31	C

Tabela 2 - Idade dos estudantes de enfermagem de uma faculdade
Fonte: Elaborada pelo autor.

Resolução:

$$média = \frac{3}{1/25 + 1/22 + 1/31}$$

$$média = \frac{3}{0,04 + 0,045 + 0,032}$$

$$média = 3/0,117$$

$$média = 25,64$$

SAIBA MAIS

Assista ao vídeo do link a seguir para saber mais sobre os diferentes tipos de médias e ver mais exemplos de cálculo de cada uma delas. Disponível em: <https://www.youtube.com/watch?v=7SeCSogbDQc>. Acesso em: 26 abr. 2019.

Mediana

A mediana é uma medida de tendência central que corresponde ao valor localizado no centro de um conjunto de dados organizado de forma crescente ou decrescente. Quando o total de observações é ímpar, a mediana é exatamente o valor único no centro dos conjuntos de dados. Todavia, quando o total de observações é par, a mediana é a média aritmética simples dos dois valores centrais.

Exemplo prático 1: identifique a mediana do conjunto de dados a seguir: 12, 10, 15, 25, 20, 8, 15.

Resolução:

O primeiro passo é ordenar os dados em rol: 8, 10, 12, 15, 15, 20, 25.

O segundo passo é destacar o valor central do conjunto de dados:

8, 10, 12, 15, 15, 20, 25.

Logo, a mediana é 15.

Exemplo prático 2: agora determine a mediana do seguinte conjunto de dados: 20, 12, 12, 18, 16, 13, 13, 15,

Resolução:

O primeiro passo é ordenar os dados em rol: 12, 12, 13, 13, 15, 16, 18, 20.

O segundo passo é destacar os valores centrais (metade à esquerda, metade à direita): 12, 12, 13, 13, 15, 16, 18, 20.

O terceiro passo é efetuar a média aritmética: 13 + 15 / 2 = 28 / 2 = 14.

Logo, a mediana do conjunto de dados é 14.

SAIBA MAIS

Além da mediana, que é uma medida tanto de tendência quanto de posição, existem outras medidas importantes, como o quartil, o decil e o percentil. Para saber mais sobre estas medidas, leia o livro Estatística Fácil, de Antônio Arnot Crespo.

Moda

A moda se refere ao valor que ocorre com maior frequência em conjunto de dados. Esta medida estatística remete ao conceito de moda representado na Figura 1, já que é algo que mais se repete em um determinado conjunto, entretanto, é uma medida pouco utilizada no cenário científico. A moda pode ser utilizada com dados quantitativos e qualitativos. É importante ressaltar que a moda não existe em conjuntos de dados que nenhum dado se repete.

Figura 1 - Moda
Fonte: Katya Ulitina / 123RF.

Exemplo prático, identifique a moda no conjunto de dados apresentado a seguir: 20, 30, 40, 80, 10, 10, 20, 30, 20.

Resolução:

O primeiro passo é ordenar os dados em rol: 10, 10, 20, 20, 20, 30, 30, 40, 80.

O segundo passo é destacar os valores que mais se repetem: 10, 10, 20, 20, 20, 30, 30, 40, 80.

Logo, a moda do conjunto de dados é 20.

SAIBA MAIS

Alguns conjuntos de dados apresentam mais de uma moda, podendo ser bimodal, quando possui duas modas, e plurimodal, quando possui mais de duas modas. Preste atenção no conjunto de dados a seguir, que possui três modas:

(1, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5) → Mo = 3, Mo = 4 e Mo = 5 (valores mais frequentes).

Quando é Melhor utilizar a Média, a Mediana e a Moda?

Nota-se que a média, a mediana e a moda são medidas de tendência central que possuem vantagens e desvantagens, conforme mostra ao quadro a seguir:

Medida	Definição	Vantagens	Desvantagens
Média	Soma dos valores dividida pela quantidade de casos.	É atraente e reflete cada valor do conjunto de dados.	Os valores extremos afetam seu valor.

Mediana	Valor localizado no centro do conjunto de dados.	Sofre menor influência dos valores extremos.	Mais difícil de ser identificada em grandes conjuntos de dados.
Moda	Valor que mais se repete no conjunto de dados.	Maior frequência de valores concentrados no mesmo ponto.	Não tem utilidade na análise estatística.

Quadro 1 - Vantagens e desvantagens da utilização da média, mediana e moda
Fonte: Elaborado pelo autor.

Concluindo, a média deve ser utilizada quando existe uma forte concentração dos dados na região central da série, enquanto a mediana deve ter a preferência do pesquisador quando os dados estão concentrados no início ou no final do conjunto de dados. Já a moda pode ser apenas em conjuntos de dados em que a frequência de um determinado valor é muito superior à frequência dos outros valores.

Exemplo prático:

Calcule a média aritmética simples, a mediana e a moda do conjunto de dados a seguir: 10, 8, 15, 13, 20, 20, 38, 23, 20, 40.

O primeiro passo é ordenar os dados em rol: 8, 10, 13, 15, 20, 20, 20, 23, 38, 40.

Vamos calcular a média:

Média = 18 + 10 + 13 + 15 + 20 + 20 + 20 + 23 + 38 + 30 / 10

Média = 20,7

Agora, vamos identificar a mediana: 18, 10, 13, 15, 20, 20, 20, 23, 38, 30

Como ao dividir o conjunto de dados em duas partes iguais, a mediana fica entre os dois valores 20.

Logo, a mediana é 20.

Como a moda é o valor que mais se repete, a moda do conjunto de dados é 20.

(18, 10, 13, 15, 20, 20, 20, 23, 38, 30)

SAIBA MAIS

Vimos que a média, a mediana e a moda são fundamentais para representar os valores de um conjunto de dados, entretanto, cada uma tem vantagens e desvantagens. Considerando o exercício prático anterior, qual medida de tendência central é a mais adequada para representar este conjunto de dados?

Medidas de Dispersão

Embora a média, a mediana e a moda proporcionem informações relevantes a respeito de um conjunto de dados, essas medidas de tendência central não são suficientes para resumir o conjunto de dados de uma forma geral. A partir de agora vamos apresentar as medidas que indicam o quanto os dados estatísticos variam em torno da região central, que são as medidas de dispersão ou variabilidade (DANCEY; REIDY; ROWE, 2017).

Considere a situação abaixo para perceber que uma medida de tendência central, neste caso, a média, não é suficiente para descrever satisfatoriamente o conjunto de dados. Temos a idade de dois grupos de pessoas (A e B):

A = 25 29 31 34 35

B = 15 23 32 36 48

Ambos os grupos apresentaram a mesma média aritmética: 30,8. No entanto, analisando os dados de cada grupo, percebe-se que a idade das pessoas do grupo A varia apenas de 25 a 35 anos, enquanto a idade do grupo B varia de 15 a 48 anos, revelando que a idade do grupo A é mais homogênea do que a idade do grupo B.

Esta situação evidencia a importância de medidas que indiquem a variabilidade dos dados de uma amostra, uma vez que utilizando apenas a média, os dois grupos apresentaram um perfil de idade semelhante. Diante disso, veremos algumas das principais medidas de dispersão e que permitem identificar as diferenças existentes entre os dois grupos, tais como: a amplitude, o desvio, a variância, o desvio-padrão e o coeficiente de variação.

SAIBA MAIS

Além dessas medidas de dispersão que veremos nesta aula, existem outras medidas que também são importantes, como o desvio-médio, entretanto, não é muito utilizada quanto às medidas estudadas nesta aula. Para saber mais sobre esta medida, assista ao vídeo no link disponível em: <https://www.youtube.com/watch?v=S-NnnCnjeSI>. Acesso em: 26 abr. 2019.

Amplitude Total

Como a próprio nome diz, a amplitude total se refere à diferença entre o maior valor e o menor valor do conjunto de observações. Apesar de ser uma medida simples e fácil de ser calculada, a amplitude total é relativamente imprecisa em relação à variação no interior do conjunto de dados, uma vez que considera apenas os valores extremos (mínimo e máximo).

Exemplo: Calcule a amplitude total do conjunto de dados a seguir: 60, 80, 70, 62, 85, 57.

O primeiro passo é organizar os dados em rol e identificar o valor mínimo e máximo: 57, 60, 62, 70, 80, 85.

Amplitude = 85 - 57

Amplitude = 28.

Desvio

O desvio de um conjunto de dados numéricos é obtido por meio da diferença entre cada um dos valores e a média aritmética da amostra. Considere a tabela a seguir acerca das idades dos estudantes de enfermagem de uma faculdade para calcular o desvio da amostra.

Idade dos alunos	Estudante
25	A
22	B
31	C
26	D
20	E
30	F

Tabela 3 - Idade dos estudantes de enfermagem de uma faculdade
Fonte: Elaborada pelo autor.

Média aritmética = 25 + 22 + 31 + 28 + 20 + 30 / 2

Média aritmética = 26

Para encontrar o desvio do conjunto de dados, é preciso fazer o seguinte cálculo:

D1 = 25 - 26 = -1

D2 = 22 - 26 = -4

D3 = 31 - 26 = 5

D4 = 28 - 26 = 2

D5 = 20 - 26 = -6

D6 = 30 - 26 = 4

Soma dos desvios = 0

Note que a soma dos desvios é igual a zero (0) e esta medida é útil para se ter uma orientação a respeito dos valores que estão acima da média (desvios com sinal positivo) e os valores que estão abaixo da média (desvios com sinal negativo)

Variância e Desvio-padrão

As medidas de dispersão mais utilizadas nos procedimentos estatísticos e com maior grau de confiabilidade são a variância (s^2) e o desvio-padrão (s), uma vez que ambas analisam a variabilidade dos dados de uma amostra levando em consideração todos os valores do conjunto de dados.

A variância de uma amostra é obtida pela soma dos quadrados dos desvios da amostra em relação à média dos desvios, com posterior divisão pelo número de elementos da amostra (n) subtraído da unidade, conforme a fórmula abaixo:

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

Portanto, o desvio-padrão do exemplo anterior éobtido a partir da seguinte fórmula:

$$s = \sqrt{19,6}$$

$$s = \sqrt{4,43}$$

Segue abaixo algumas observações importantes a respeito da variância e do desvio-padrão:

- Quando os valores da amostra são repetidos, a variância será zero;
- A variância é fundamental na estatística inferencial, entretanto, não tem muita utilidade na estatística descritiva;

- Quando os valores da amostra estão espalhados, a variância será um valor elevado e que indica uma grande dispersão dos dados em relação à média;
- Quanto menor o desvio-padrão, mais próximos estão os valores da média;
- Quanto maior o desvio-padrão, mais distantes estão os valores da média.

Coeficiente de Variação

O coeficiente de variação (CV) é uma medida relativa de variabilidade que permite identificar a dispersão dos valores do conjunto de dados em relação à média aritmética. Este coeficiente é obtido pela divisão do desvio-padrão pela média multiplicado por 100, conforme mostra a fórmula a seguir:

$$CV = \frac{Desvio\ Padrão\ Amostral}{Média\ Amostral} \times 100 \rightarrow CV = \frac{S}{\underline{x}} \times 100$$

Considerando o exemplo anterior, o coeficiente de variação é obtido da seguinte forma:

$$CV = 4,43/26x100 =$$

$$CV = 0,17x100 = 17$$

A principal utilid.ade do coeficiente de variação é permitir a comparação da dispersão de diferentes conjuntos de dados, sendo que quanto menor for o coeficiente de variação, mais homogêneos são os valores da amostra.

SAIBA MAIS

Para saber mais sobre as medidas de dispersão estudadas nesta aula, leia o artigo científico abaixo:

Fonte: BASTOS, J. L. D.; DUQUIA, R. P. Medidas de dispersão: os valores estão próximos entre si ou variam muito. Scientia Medica, v. 17, n. 1, p. 40-44, 2007. Acesso em: 26 abr. 2019.

QUESTÃO OBJETIVA

Um indivíduo realizou diversas provas até ser aprovado em um concurso público. As notas do indivíduo em suas provas de concurso foram as seguintes: 8,4; 9,1; 7,2; 6,8; 8,7 e 7,2.

Sabendo de tais informações, a nota média, a nota mediana e a nota modal desse indivíduo, são respectivamente:

- a) 7,9; 7,8; 7,2.
- b) 7,2; 7,8; 7,9.
- c) 7,8; 7,8; 7,9.
- d) 7,2; 7,8; 7,9.
- e) 7,8; 7,9; 7,2.

QUESTÃO OBJETIVA

Das diversas medidas de dispersão apresentadas nesta aula, a amplitude e o desvio são medidas muito utilizadas nas mais diversas pesquisas. A respeito destas medidas, assinale a alternativa correta:

- a) O desvio é uma medida de dispersão calculada sobre cada um dos valores de um conjunto de informações.
-

b) A amplitude é uma medida de tendência central obtida a partir dos valores do conjunto de observações.

c) O desvio se refere a uma medida de dispersão utilizada para identificar a dispersão total de conjunto de dados.

d) A amplitude é uma medida de dispersão que corresponde ao valor localizado no centro do conjunto de dados.

e) Pode-se dizer que o desvio e a amplitude são medidas de dispersão utilizadas para identificar a mesma variabilidade de um conjunto de dados.

Fechamento

Nesta aula vimos que existem medidas adequadas para representar os valores de um conjunto de dados, que são as medidas de tendência central e as medidas de dispersão. A medida de tendência central mais utilizada pelos estatísticos é a média aritmética, entretanto, as outras medidas também são úteis em alguns casos. Além disso, percebemos que as medidas de tendência central não são capazes de revelar a totalidade das informações a respeito de um conjunto de dados, e que medidas de dispersão, como a variância, o desvio-padrão e o coeficiente de variação são úteis para apontar o grau de variabilidade dos valores de um conjunto de dados. Tanto as medidas de tendência central quanto as medidas de dispersão são amplamente utilizadas durante o processo de análise de dados estatísticos.

Nesta aula, você teve a oportunidade de:

- reconhecer as principais medidas de tendência central e dispersão;
- compreender a utilização de cada medida de tendência central e dispersão;
- aplicar as principais medidas de tendência central e dispersão em diferentes situações.

Aula 05

Curva de Gauss e a Distribuição Paramétrica

Introdução

Na estatística, a análise da distribuição normal é um dos principais pressupostos para a utilização dos testes inferenciais e, frequentemente, testar as hipóteses que foram estabelecidas. Com isso, uma das principais perguntas quando vamos analisar um conjunto de dados é: o meu conjunto de dados apresenta distribuição normal? Este questionamento é importante, uma vez que esta análise mostra como os dados estão distribuídos ao longo do conjunto de dados. Esta etapa é fundamental para decidir entre os testes paramétricos e não-paramétricos. Nesta aula veremos as características da distribuição normal, as diferentes formas de analisar a distribuição normal e o que fazer quando os dados não apresentam distribuição normal.

Ao final desta aula, você será capaz de:

- reconhecer as características teóricas da distribuição normal;
- analisar a normalidade a partir de diferentes medidas estatísticas;
- identificar as estratégias adotadas quando os dados não apresentam distribuição normal.

Distribuição Normal

A maior parte das inferências em pesquisas médicas e biológicas é baseada em dados com distribuição normal, razão pela qual para muitos pesquisadores uma das suposições mais importantes na estatística é a análise da normalidade dos dados, uma vez que é um pressuposto fundamental para selecionar o teste estatístico que será empregado para a análise dos dados. Muitos dos procedimentos estatísticos são testes paramétricos, os quais requerem que os dados sejam retirados de uma população normalmente distribuída (BARROS et al., 2012).

A análise da normalidade dos dados é um procedimento importante para o pesquisador decidir se vai adotar uma estatística paramétrica ou uma estatística não-paramétrica. Segundo Field (2009), quando um teste paramétrico é utilizado em dados com distribuição não normal, os resultados podem não estar de acordo com a realidade dos dados. Este erro pode afetar as conclusões quanto à aplicabilidade prática e clínica dos resultados, levando a tomadas de decisões imprecisas ou até mesmo catastróficas.

Mas você deve estar se perguntando: o que é distribuição normal? A distribuição normal também é conhecida como curva normal ou distribuição Gaussiana. A curva normal é também conhecida como distribuição Gaussiana devido à suposição de que Gauss foi o primeiro a fazer uso de suas propriedades para aplicações práticas (Figura 1).

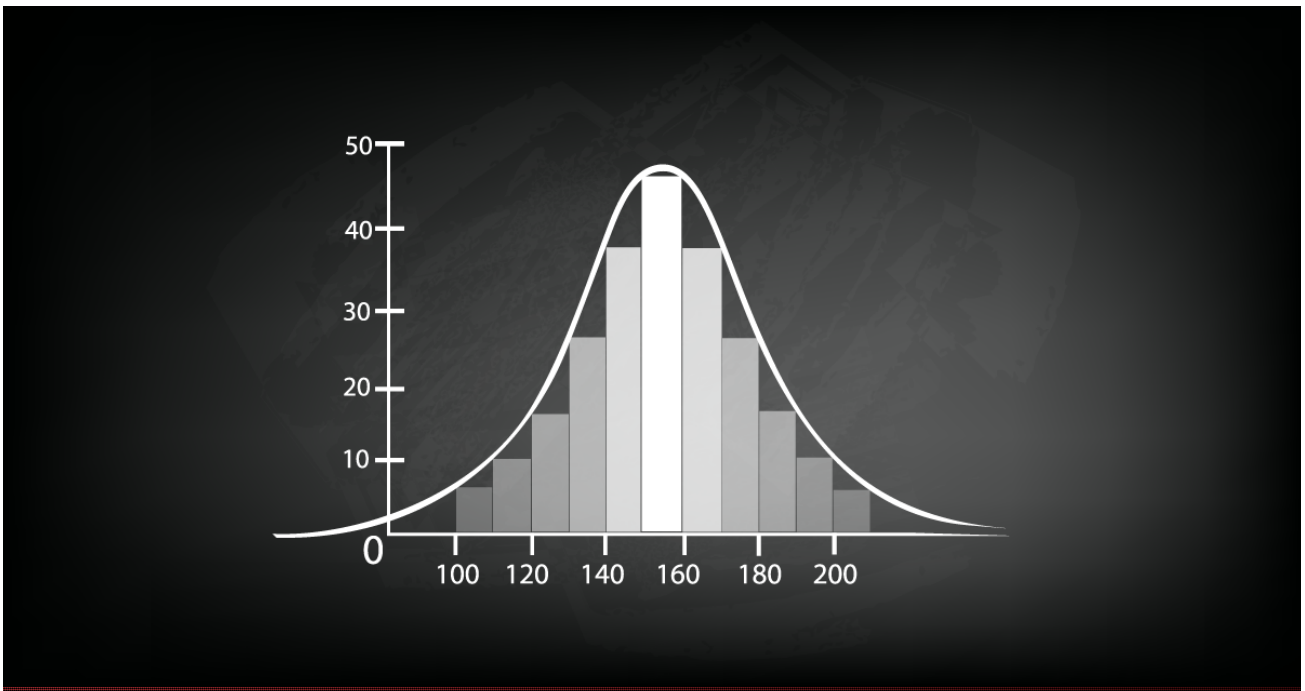


Figura 1 - A Curva normal perfeita de Gauss
Fonte: iamnee / 123RF.

A normalidade se refere à distribuição simétrica ou normal dos dados, indicando que a distância entre os valores de uma variável quantitativa deve ser igual em todas as partes ao longo da escala, em que o comportamento de um participante não influencia no comportamento de outro (NASCIMENTO et al., 2015).

SAIBA MAIS

Para saber mais sobre a importância da normalidade em pesquisas nas ciências da saúde, leia o artigo científico abaixo:

Fonte: NASCIMENTO, D. C. et al. Testes de normalidade em análises estatísticas: uma orientação para praticantes em ciências da saúde e atividade física. **Revista Mackenzie de Educação Física e Esporte**, v. 14, n. 2, 2015. Disponível em: <http://editorarevistas.mackenzie.br/index.php/remef/article/view/6583/6653>. Acesso em: 26 abr. 2019.

A função matemática que representa a distribuição normal envolve dois parâmetros (média e variância), a curva que a descreve tem forma de "sino" e sua principal propriedade é a simetria dos dados em torno da média. De acordo com Barros et al. (2012), algumas das principais características da distribuição normal são as seguintes e que também estão ilustradas na Figura 2:

- A distribuição normal pode ser completamente descrita pela média e pela variância;
- A curva tem o formato de um sino;
- Os dados apresentam distribuição simétrica em relação à média;
- A média, a moda e a mediana apresentam o mesmo valor;
- Ao assumir uma variância constante, a distribuição (curva) se desloca à direita conforme a média aumenta e se desloca à esquerda na medida em que a média diminui;
- A curva sofre achatamento na medida em que a variância aumenta e sofre alongamento na medida em que a variância diminui;
- As caudas da curva encontram o eixo x no infinito;
- 100% dos dados estão simetricamente distribuídos em relação à média;
- Quando a distribuição é normal (simétrica), a média representa bem os dados, enquanto que quando a distribuição é assimétrica, a mediana é mais representativa.

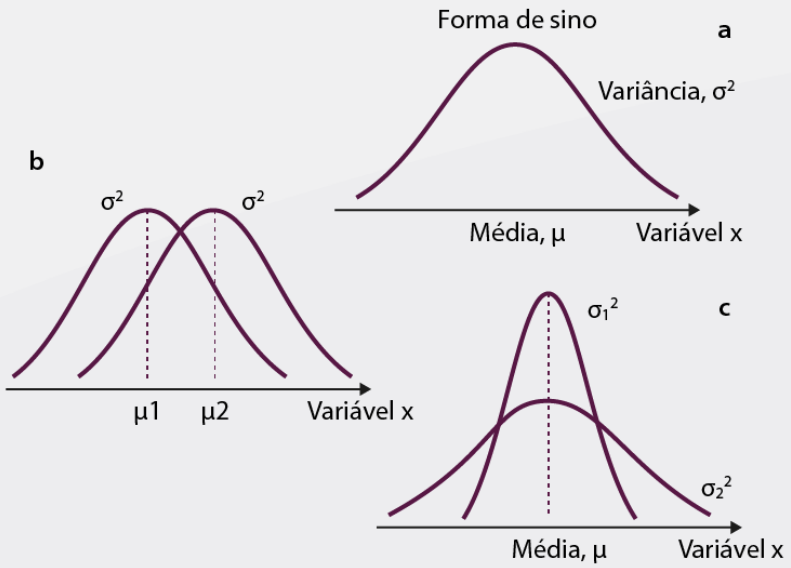


Figura 2 - Exemplos de curvas de distribuição normal: a) simétrica em relação à média; b) com médias diferentes (m2 > m1) e a mesma variância; c) variâncias diferentes e a mesma média
Fonte: Barros et al. (2012, p.73).

SAIBA MAIS

A distribuição normal é obtida quando a distribuição dos valores do conjunto de dados apresenta um pico na região central, apresentando uma forma de sino. A distribuição normal perfeita acontece quando a média, a mediana e a moda coincidem com o ponto do pico da curva, conforme mostra a Figura 3.

Fonte: Barros et al. (2012).

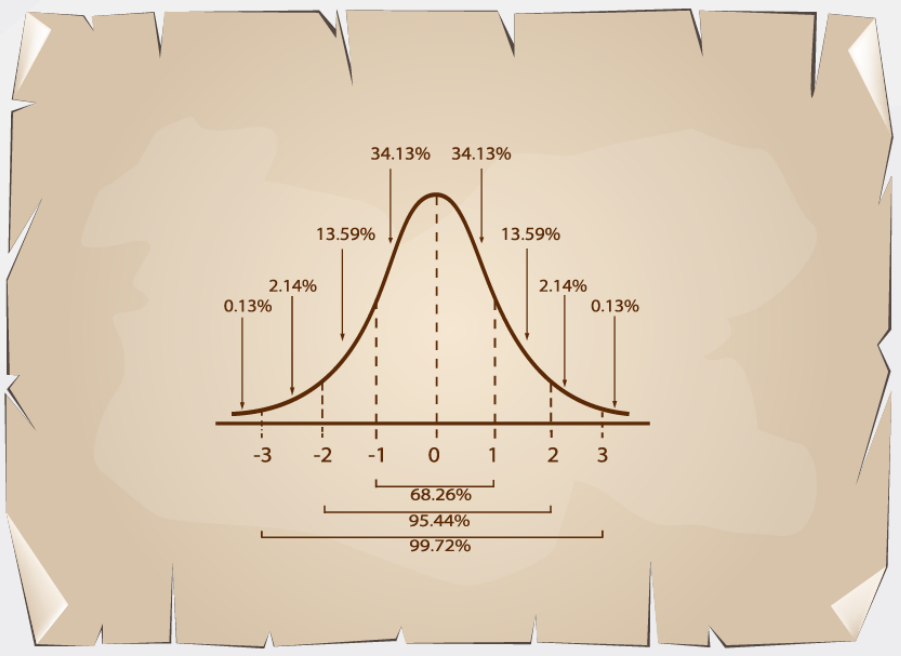
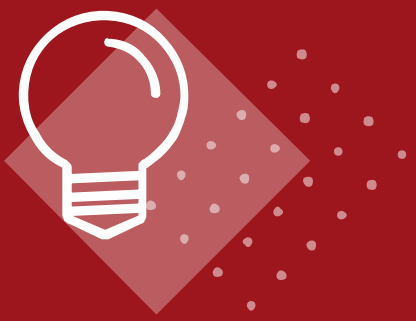


Figura 3 - Gráfico que representa a distribuição normal dos dados
Fonte: iamnee / 123RF.

Conforme já visto, a curva de Gauss é obtida a partir dos valores da média e da variância (e, consequentemente, o desvio-padrão). Sabendo os valores dessas medidas, torna-se possível desenhar a curva por meio de uma fórmula matemática. Esta fórmula não será apresentada nessa aula, mas é importante saber que a curva pode ser desenhada a partir dos valores da média e do desvio-padrão.

SAIBA MAIS



Para saber mais sobre as características da curva normal, como a distribuição normal padronizada e a utilização do escore Z, leia os artigos científicos:

A Questão da Não Normalidade: uma revisão, disponível em: <http://www.iea.sp.gov.br/ftpiea/publicar/rea2014-2/rea2-2014.pdf>. Acesso em: 26 abr. 2019.

Desvio-padrão ou erro padrão: qual utilizar?, disponível em: <http://www.ufjf.br/ppgsaude/files/2018/11/Desvio-Padra%CC%83o-e-Erro-Padra%CC%83o.-Qual-a-diferenc%CC%A7a.pdf>. Acesso em: 26 abr. 2019.

Desvios à Normalidade

A busca pela distribuição normal é recorrente por pesquisadores, pois os métodos estatísticos mais utilizados na área da saúde e na epidemiologia, como o teste t de Student, as ANOVAS, a correlação linear simples e a regressão linear e intervalos de confiança são testes paramétricos e requerem a distribuição normal dos dados. Quando os dados não apresentam distribuição normal, a estatística não-paramétrica deve ser adotada para os procedimentos estatísticos inferenciais.

No entanto, dificilmente um conjunto de dados apresenta as mesmas características da distribuição normal teórica. Com isso, sempre se procura verificar se a distribuição de um conjunto de dados tem uma distribuição próxima da normalidade. Para tal verificação, a distribuição dos dados é baseada nas medidas de assimetria e curtose da curva dos dados em comparação à curva normal (BARROS et al., 2012).

Assimetria

A assimetria se refere ao grau de afastamento da distribuição dos dados em relação ao seu eixo de referência ou simetria, isto é, na medida em que a distribuição se afasta do eixo, a distribuição fica mais assimétrica. Quando a distribuição não é simétrica, os valores da média, mediana e moda também são diferentes.

Quanto mais o afastamento da curva ocorre para o lado direito, a assimetria é considerada negativa, ao passo que o afastamento da curva para o lado esquerdo é chamado de assimetria positiva. Na assimetria negativa, a média é menor que a mediana que, por sua vez, é menor que a moda. Já na assimetria positiva, a moda é menor que a mediana que, por sua vez, é menor que a média (BARROS et al., 2012). A Figura 4 ilustra a curva normal (média = mediana = moda) e as assimetrias à direita e à esquerda.

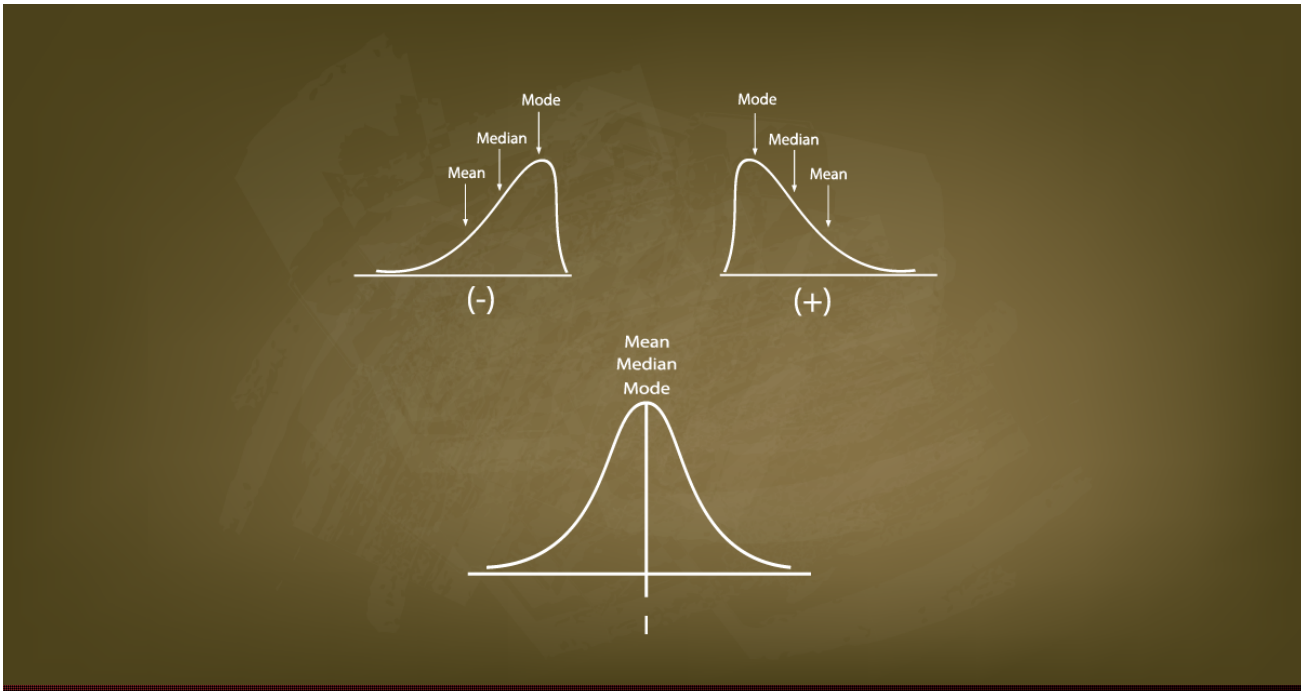


Figura 4 - Curva normal e assimetrias negativa (esquerda) e positiva (direita)
Fonte: iamnee / 123RF.

A assimetria é um pressuposto importante para a utilização da média como medida de tendência central, uma vez que na medida que a distribuição é mais assimétrica, os valores dispostos nas caudas podem distorcer o valor da média. Nestes casos, o mais recomendado é utilizar a mediana para representar os dados da amostra.

ATENÇÃO

As distribuições assimétricas são aquelas em que o pico da curva está deslocado para a direita ou para a esquerda e a cauda estendida para o lado oposto. Logo, na distribuição assimétrica negativa, o pico está deslocado à direita e a cauda à esquerda, ao passo que na distribuição assimétrica positiva o pico está deslocado à esquerda e a cauda à direita (Figura 4).

Curtose

A curtose corresponde ao grau de achatamento da distribuição dos dados, demonstrando o quanto a curva será achatada em comparação à curva normal. Assim, em relação à curtose (achatamento), a curva pode ser classificada em mesocúrtica, platicúrtica e leptocúrtica. A curva mais fechada ou alongada na parte superior é denominada de leptocúrtica, enquanto que a mais achatada e aberta é conhecida como platicúrtica. Já a curva normal é chamada de mesocúrtica (DANCEY; REIDY, 2019). A Figura 5 apresenta a curva normal, achatada e alongada.

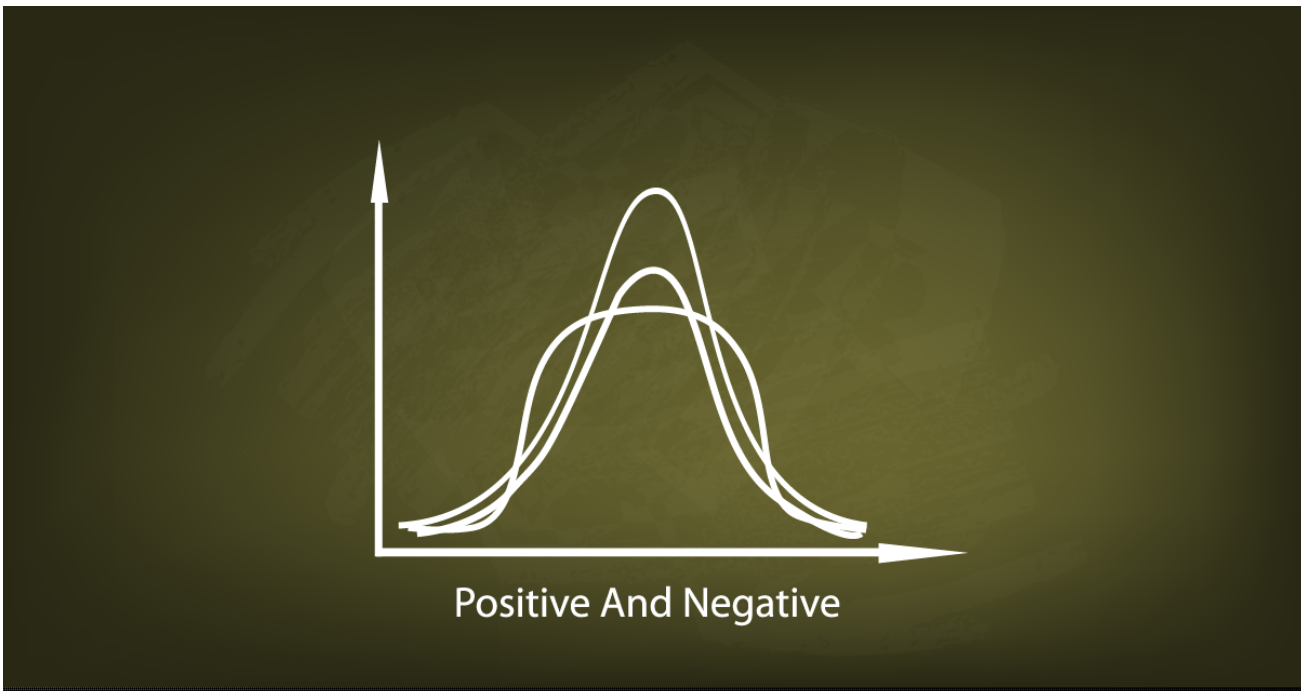


Figura 5 - Curva normal, curva achatada e curva alongada
Fonte: iamnee / 123RF.

Distribuições Bimodais

Em alguns casos é possível que os dados apresentem um tipo diferente de distribuição, com dois picos de curva iguais e, consequentemente, duas modas. Este tipo de distribuição é denominado distribuição bimodal. Quando este tipo de distribuição for encontrado é importante examinar o conjunto de dados, visto que é possível que os dados sejam provenientes de duas populações diferentes (DANCEY; REIDY, 2019). A Figura 6 ilustra um exemplo de distribuição bimodal, com dois picos de curva e duas modas.

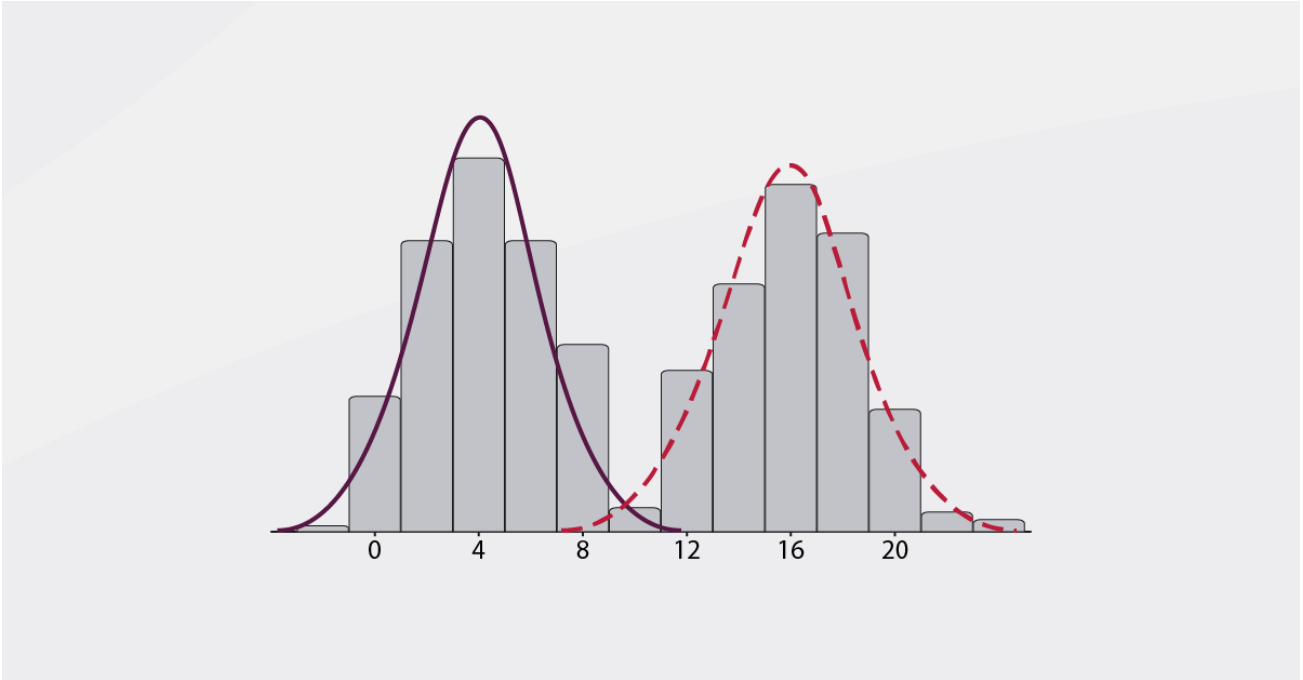


Figura 6 - Exemplo de distribuição bimodal
Fonte: Minitab (2019).

Testes de Normalidade

A análise da normalidade dos dados é efetuada de diversas formas, como o cálculo da assimetria e da curtose, métodos gráficos e testes estatísticos. O histograma pode ser utilizado para analisar a distribuição dos dados na medida em que a representação da curva normal sobre os dados do gráfico permite facilmente a identificação de assimetria e curtose, conforme ilustra a Figura 7.

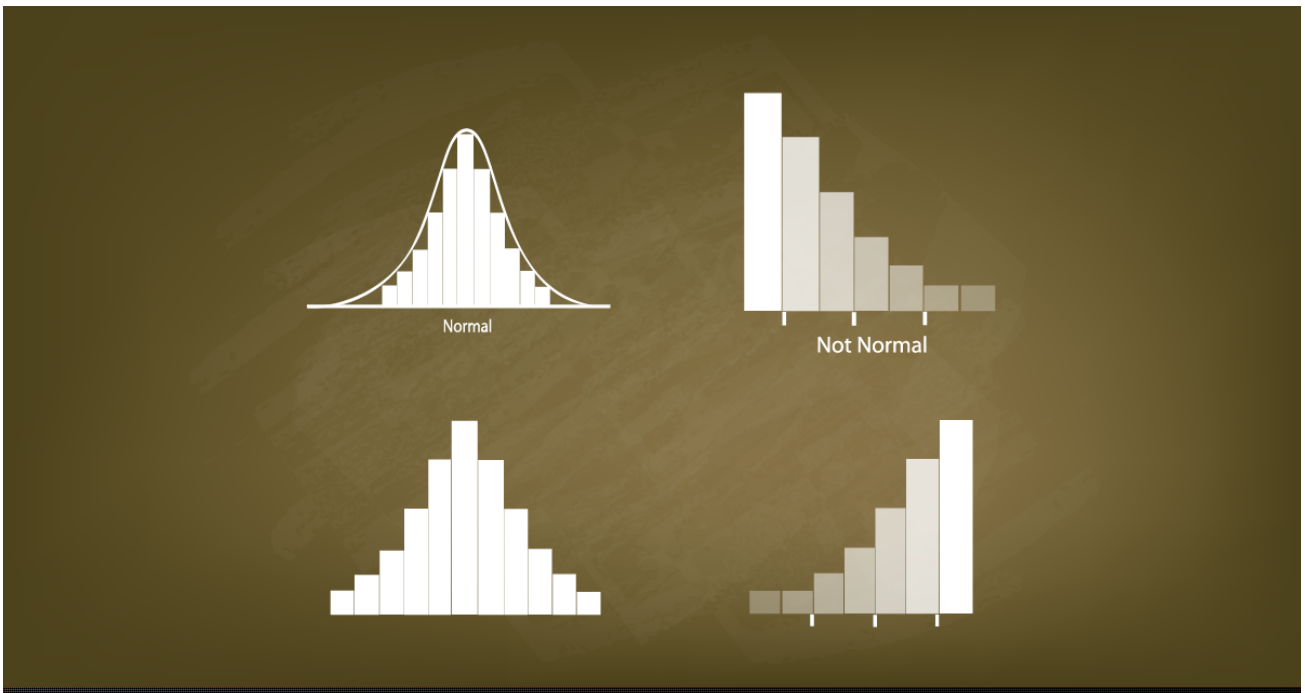


Figura 7 - Curva normal sobre a distribuição dos dados em um histograma
Fonte: iamnee / 123RF.

Para ilustrar a análise da curva de normalidade por meio da distribuição da frequência dos dados da variável idade por meio do histograma, preste atenção no infográfico a seguir.

INFOGRÁFICO INTERATIVO

Para consultar o Infográfico Interativo, acesse a **versão digital** deste material

Contudo, o histograma não fornece a medida quantitativa da assimetria e da curtose. Para isso, deve-se utilizar equações matemáticas, entretanto, não vamos apresentá-las nessa aula. Tais cálculos podem ser facilmente obtidos nos softwares estatísticos, como o SPSS, o Stata, o R e até o Excel.

SAIBA MAIS

Para saber mais sobre os cálculos padronizados da assimetria e da curtose, além de exemplos de softwares para a análise da normalidade dos dados, leia o livro Análise de Dados em Saúde, de Mauro V. G. Barros et al.

Ainda existem os testes estatísticos que confirmam se um conjunto de dados apresenta distribuição normal. Para amostras pequenas (até 50 indivíduos), o teste de Shapiro-Wilk é mais recomendado, enquanto o teste de Kolmogorov-Smirnov é recomendado para amostras acima de 50 sujeitos. Para que os dados apresentem distribuição normal nesses testes, o valor da significância deve ser maior do que 0,05 (BARROS et al., 2012).

SAIBA MAIS

Para ver como analisar a normalidade dos dados por meio dos testes de normalidade e histograma, assista ao vídeo disponível em: <https://www.youtube.com/watch?v=fkqbycVzSpw&t=33s>. Acesso em: 26 abr. 2019.

No entanto, o resultado destes testes é afetado pelo tamanho da amostra, visto que em grandes amostras os testes tendem a indicar uma distribuição não normal apesar de os dados seguirem uma distribuição simétrica. Dessa forma, em grandes amostras, também é recomendável a verificação da normalidade por meio do histograma e dos valores padronizados da assimetria e da curtose.

SAIBA MAIS

Para saber mais sobre a questão da normalidade dos dados em estudos clínicos e experimentais na área da saúde, leia o artigo disponível em: <https://bit.ly/30EVjQA>. Acesso em: 26 abr. 2019.

QUESTÃO OBJETIVA

Um dos principais pressupostos para decidir entre a estatística paramétrica e não-paramétrica é a distribuição normal dos dados. Esta análise é realizada por meio da curva normal, que também é conhecida como curva de Gauss ou distribuição Gausssiana. Considere as afirmações abaixo sobre as características da distribuição normal:

I. A curva tem forma de sino;

II. A média, mediana e moda apresentam valores diferentes em uma distribuição normal;

III. A curva pode apresentar assimetria para a esquerda ou para a direita;

IV. Os dados apresentam distribuição simétrica em torno da média;

V. As caudas devem se encontrar devem se estender até o infinito.

Agora, assinale a alternativa com as afirmativas corretas:

- a) I, IV e V, apenas.
- b) I e IV, apenas.
- c) II, II e V, apenas.
- d) I e V, apenas.
- e) I, II e IV, apenas.

QUESTÃO OBJETIVA

Ao analisar a normalidade dos dados é possível que você perceba que seus dados apresentem uma distribuição não normal, com curva com características diferentes da curva normal, e que apresentam valores extremos. Nesses casos, qual é a medida de tendência central mais adequada para ser utilizada?

- a) Mediana
- b) Média
- c) Moda
- d) Mediana e Média
- e) Nenhuma das alternativas anteriores

Fechamento

Nesta aula vimos um dos principais pressupostos antes de se iniciar a estatística inferencial, que é a análise da distribuição dos dados. Esta análise é realizada por meio da curva normal, que também é conhecida como curva de Gauss ou distribuição Gausssiana. A análise da normalidade se refere ao grau com que os valores de um conjunto de dados se dispersam ao longo da curva. Dados com distribuição normal apresentam diversas características, como: distribuição simétrica dos dados em torno da média, curva em forma de sino e média, mediana e moda com valores semelhantes. As principais formas de analisar a normalidade envolvem o histograma,

análise de assimetria e curtose e os testes de normalidade nos softwares estatísticos. A análise da normalidade irá determinar se o caminho a ser percorrido será o da estatística paramétrica ou não-paramétrica.

Nesta aula, você teve a oportunidade de:

- reconhecer as características teóricas da distribuição normal;
- analisar a distribuição normal a partir de diferentes medidas estatísticas;
- identificar as estratégias adotadas quando os dados não apresentam distribuição normal.

ATIVIDADE COMPLEMENTAR

Para complementar o aprendizado e aprofundar os conhecimentos em relação aos assuntos estudados na Unidade I, você pode ler os capítulos 1 a 6 do livro Análise de Dados em Saúde, que consta na referência ao final da unidade. Durante a leitura, preste atenção nos principais conceitos estatísticos que você aprendeu e em mais exemplos que são apresentados no livro. Além disso, se atente ao processo de organização de dados e às formas de representação e descrição de dados, com destaque para a distribuição de frequência, as medidas de tendência central e dispersão e a representação gráfica. Por último e não menos importante, preste muita atenção às explicações relacionadas às características da distribuição normal dos dados. Após a leitura, tente responder as questões a seguir:

1 - Qual a diferença entre as variáveis quantitativas e qualitativas? Cite exemplos de cada uma delas.

2 - Descreva a diferença entre a média, a mediana e a moda.

3 - Qual o papel das medidas de dispersão? Quais as principais medidas de dispersão?

4 - Quais os principais tipos de gráficos e quando devemos utilizar cada um deles?

5 - Qual a importância da análise da distribuição normal?

Teoria e Prática

Você sabia que o estresse no casamento afeta mais a saúde física e mental da mulher? “Um estudo desenvolvido nos Estados Unidos demonstrou que mulheres em casamentos problemáticos têm mais chances de sofrer problemas de saúde como obesidade, hipertensão e colesterol alto - sintomas de uma "síndrome do metabolismo" que pode levar a doenças cardíacas, diabetes e derrame. A pesquisa da Universidade de Utah mostra ainda que os homens são menos afetados por estes sintomas, mas têm os mesmos riscos que as mulheres de sofrer de estresse e depressão. Os pesquisadores entrevistaram 276 casais, com idades entre 40 e 70 anos, e que estavam casados há uma média de 20 anos. Eles avaliaram os aspectos positivos e negativos de cada casamento, além de monitorar a saúde dos voluntários” (BBC Brasil, 2009).

Esta situação mostra claramente o papel da estatística para as descobertas e avanços na área da saúde. Para se chegar a esta conclusão, foi necessário selecionar uma amostra a partir de uma população, coletar e organizar os dados, e analisar de forma descritiva e inferencial os dados obtidos.

ESTUDO DE CASO

Sabe-se que a taxa de suicídios tem aumentado substancialmente nos últimos anos em todo o mundo. Para traçar estratégias para reduzir a taxa de suicídios, um grupo de pesquisadores procurou compreender as razões das pessoas decidirem acabar com suas vidas. Especificamente, tiveram como meta fazer um levantamento do sexo da pessoa que comete o suicídio e o método escolhido para fazê-lo. Para se chegar a uma conclusão e ter informações para tomada de decisões em relação às estratégias para reduzir as taxas de suicídio, a condução de uma pesquisa utilizando a ferramenta estatística é fundamental.

O pesquisador tem um problema a ser estudado (suicídio) e precisa selecionar instrumentos para obter os dados, como, por exemplo, a elaboração de um questionário ou esquema de uma entrevista, ou até mesmo obtenção de dados de prontuários de pessoas que cometeram suicídio. Em seguida, os dados precisam ser coletados para, assim, se realizar a análise dos dados e chegar às conclusões para as tomadas de decisões.

Mapa Conceitual