

Estatística e Probabilidade

Conceitos básicos



Núcleo de Educação a Distância

www.unigranrio.com.br

Rua Prof. José de Souza Herdy, 1.160
25 de Agosto – Duque de Caxias - RJ

Reitor

Arody Cordeiro Herdy

Pró-Reitoria de Programas de Pós-Graduação

Nara Pires

Pró-Reitoria de Programas de Graduação

Lívia Maria Figueiredo Lacerda

Pró-Reitoria Administrativa e Comunitária

Carlos de Oliveira Varella

Núcleo de Educação a Distância (NEAD)

Lúcia Inês Kronemberger Andrade

Produção: Gerência de Desenho Educacional - NEAD

Desenvolvimento do material: Gregório Dalle Vedove Nosaki

1ª Edição

Copyright © 2022, Unigranrio

Nenhuma parte deste material poderá ser reproduzida, transmitida e gravada, por qualquer meio eletrônico, mecânico, por fotocópia e outros, sem a prévia autorização, por escrito, da Unigranrio.

Sumário

Conceitos básicos

Para Início de Conversa

Objetivos

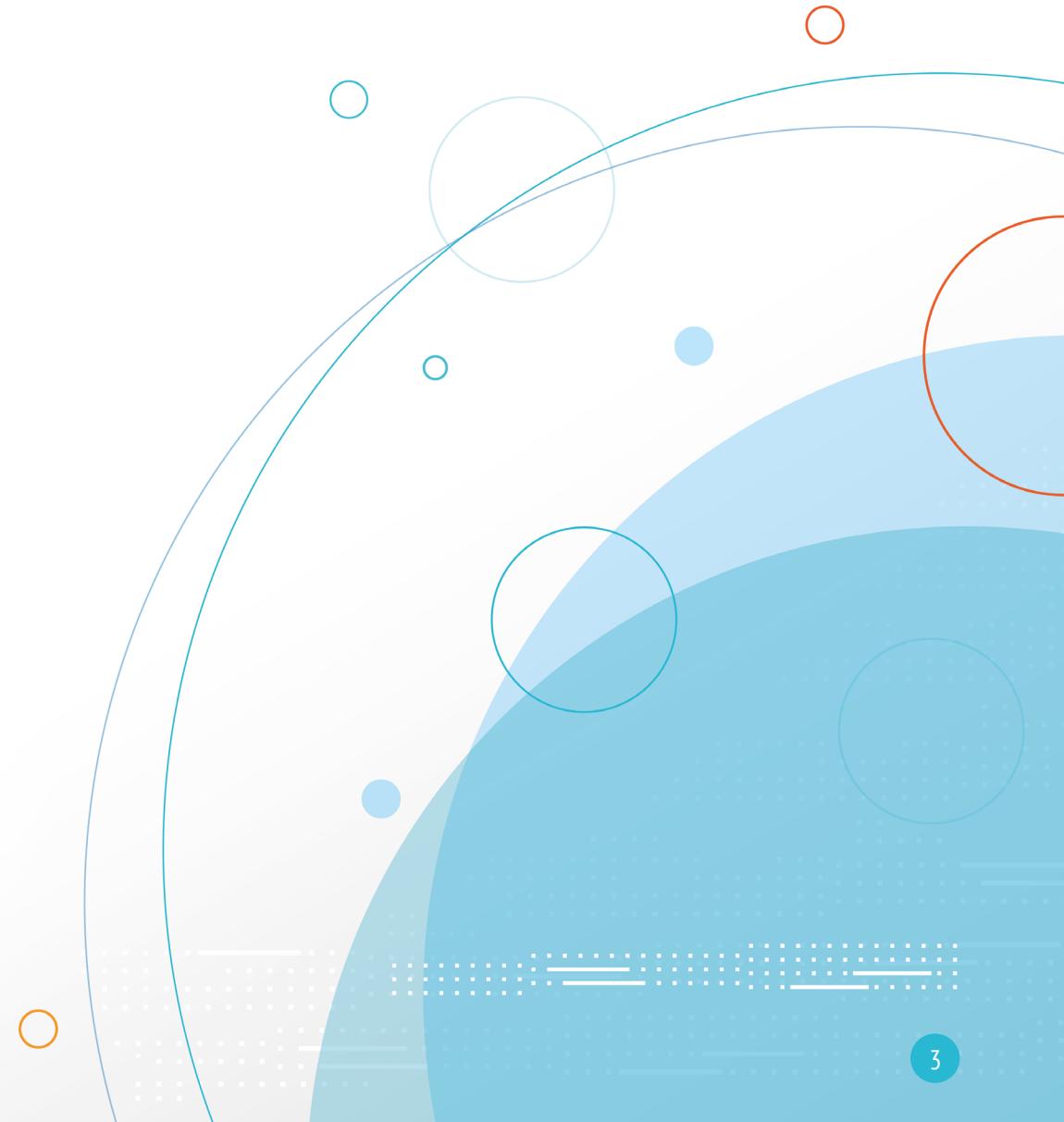
1. População

2. Amostra

3. Variáveis

Referências

4
4
5
6
8
12



Para Início de Conversa

Neste primeiro capítulo, começaremos a trabalhar com a análise exploratória de dados que é um dos principais objetos de estudo em Estatística. Abordaremos a coleta, organização, processamento e apresentação dos dados de um experimento ou de uma pesquisa. Definiremos os conceitos básicos de população e amostra e apresentaremos diversos exemplos dessas duas definições. Analisaremos a relação entre a população e uma amostra dessa população bem como esse processo de escolha, chamado de amostragem, pode ser realizado. Esses são elementos que percorrem todo o estudo estatístico e, por isso, são tão importantes em qualquer curso de Estatística. Apresentaremos, ainda, algumas classificações para os diferentes tipos de amostragem que podem ser realizadas e suas principais características, destacando vantagens e desvantagens. Trataremos também da definição de variáveis dentro de um banco de dados e seus diferentes tipos, além de apresentar diversos exemplos para ilustrar a aplicação desses conceitos.

Objetivos

- Conhecer e saber aplicar os conhecimentos da estatística como ferramenta para tomada de decisão e/ou pesquisa quantitativa.
- Aplicar técnicas de amostragem.
- Compreender conceitos básicos relacionados à estatística, como variáveis, estimadores, estimativas, parâmetros, população, amostras.



1. População

A Estatística pode ser descrita como a ciência dos dados. Por meio dela que podemos estudar, coletar, organizar e analisar todo tipo de dados. Em razão da sua grande abrangência, ela está presente em diversos campos do conhecimento, fazendo parte do cotidiano de todos. Vamos iniciar o estudo dessa ciência definindo o conceito de população.

Definição: uma **população** é uma coleção de elementos do mesmo tipo que compartilham de uma determinada característica. Geralmente, chamamos cada elemento de uma população de **indivíduo**.



Figura 1: População. Fonte: Dreamstime.

Os alunos de uma determinada escola formam uma população, pois todos compartilham a característica de serem alunos dessa escola. Os usuários do Instagram formam uma população, sendo que cada usuário é considerado um indivíduo. Mas o sentido do conceito de população em estatística é mais abrangente do que apenas um grupo de pessoas que compartilham uma característica. Podemos trabalhar com carros produzidos em uma determinada fábrica; dessa forma, cada carro é um indivíduo da população total de carros produzidos nessa fábrica. Da mesma forma, se estivermos analisando os boletins de ocorrência (B.O.) registrados em uma cidade ou região, então, cada B.O. registrado configura um indivíduo dentro da totalidade de registros.

! Importante

É necessário destacar que, dentro de uma mesma população, todos os elementos pertencem ao mesmo tipo de classe, ou seja, se estamos trabalhando com uma população formada por pessoas, todo indivíduo da população será uma pessoa. Não pode haver uma população formada por pessoas e carros, por exemplo.

Após determinarmos a população que iremos trabalhar, estaremos interessados em um estudo sobre determinadas características dos indivíduos dessa população. O levantamento de dados de todos os indivíduos da população sobre uma determinada característica é chamado de **censo**. Contudo, nem sempre é possível realizar esse

tipo de coleta de dados, pois acessar as informações de cada um dos indivíduos pode não ser viável. Diante desse fato, surge o conceito de amostra de uma população.

Curiosidade

O censo demográfico ou recenseamento demográfico busca avaliar com precisão a população de um determinado país ou região, além de determinar características sociais e econômicas da população. Segundo a ONU:

Um recenseamento de população pode ser definido como o conjunto das operações que consistem em recolher, agrupar e publicar dados demográficos, econômicos e sociais relativos a um momento determinado ou em certos períodos, a todos os habitantes de um país ou território. (CAMARGO, 2021).

O censo pode se restringir a determinadas áreas ou especificidades, dependendo do tipo de pesquisa e da população a ser avaliada como, o censo agropecuário ou o censo eleitoral.

2. Amostra

Definição: **amostra** de uma população é um grupo mensurável e manejável de indivíduos, que é estudado de modo a revelar um comportamento ou uma característica que, mais tarde, será generalizado para toda a população.

No exemplo de alunos de uma escola, podemos selecionar um grupo de 20 alunos para avaliar quais são suas matérias favoritas. A partir dessa amostra, podemos inferir qual é a matéria favorita dos alunos dessa escola. Note que estamos **inferindo** esse resultado, dado que não foi realizada uma pesquisa na totalidade de alunos da escola. Essa é uma das principais atividades dentro da Estatística, também conhecida como **inferência estatística**, em que, a partir do estudo de uma amostra, as conclusões são generalizadas para toda a população.

Vejamos alguns exemplos de populações e amostras:

- Em uma pesquisa de satisfação sobre um modelo de refrigerador, a companhia responsável pela produção selecionou 20 clientes que adquiriram esse modelo nos últimos dois anos. Neste caso, a população são os proprietários desse modelo específico de refrigerador, e a amostra são os 20 clientes selecionados para responder à pesquisa.
- Os alunos da rede pública do estado de São Paulo serão avaliados quanto ao desempenho na disciplina de Matemática, por meio de uma prova a ser aplicada em 50 alunos de cada escola estadual. Aqui, a população seria todos os alunos da rede pública de ensino, e a amostra, todos os alunos selecionados para realizar a prova diagnóstica.
- Comparando as multas aplicadas por excesso de velocidade registradas por um radar em uma rodovia, deseja-se estudar se a maioria delas foram cometidas por homens. Para isso, são

selecionadas, aleatoriamente, 100 multas aplicadas no último ano e consultado o nome do proprietário do veículo. A população corresponde a todas as multas emitidas por esse radar, e a amostra, as 100 multas selecionadas.

Baseado nesses exemplos podemos compreender que a Estatística tem diversas aplicações em todas as áreas do conhecimento e pode ser utilizada para generalizar e inferir informações sobre uma população.

! Importante

Uma das características mais importantes na Estatística aplicada a uma determinada população ou amostra é a formulação das questões ou mesmo na seleção de quais tipos de dados serão coletados. No caso das multas aplicadas pelo radar do último exemplo, não necessariamente é o proprietário do veículo que estaria dirigindo naquele momento; portanto, analisar apenas o sexo do proprietário do veículo não geraria um levantamento de dados correspondente à realidade.

O processo de selecionar a amostra dentro de uma população torna-se fundamental para o processo de inferência estatística. Esse processo é chamado de **amostragem**, e existem diversas formas de se determinar a amostragem de uma população. Vamos estudar algumas dessas técnicas a seguir.

Um primeiro tipo de amostragem é chamado de **amostragem de conveniência**, em que os indivíduos da população selecionados são os mais acessíveis para a coleta de dados. Por exemplo, em uma fábrica que produz bolas de futebol, podemos recolher uma determinada quantidade para servir de amostra que esteja de mais fácil acesso dentro do depósito onde essas bolas são armazenadas. Esse tipo de amostragem não é muito confiável, pois podemos selecionar indivíduos que não representem de maneira significativa a população.

Outro tipo de técnica de amostragem é a **amostragem aleatória simples**, em que uma amostra de tamanho n é definida de maneira aleatória e cada indivíduo da população tem a mesma chance de ser selecionado. Se estivermos interessados na intenção de votos para o cargo de prefeito de um município, podemos selecionar aleatoriamente um grupo de eleitores e questioná-los sobre suas intenções de voto. Esse tipo de amostragem é umas das mais utilizadas, mas ainda está sujeita a riscos de não corresponder ao comportamento total da população.

Na **amostragem aleatória estratificada**, dividimos a população em grupos similares e, a partir dessa divisão, realizamos uma amostragem aleatória simples em cada um desses grupos. No caso da pesquisa de intenção de votos, podemos subdividir a população pelos bairros ou zonas da cidade, para que tenhamos certeza de consultar eleitores de diferentes regiões do município. Esse tipo de amostragem pode ser

mais conveniente, pois estaremos garantindo uma representatividade de cada grupo da população dentro da amostra, mas, para isso, é preciso escolher de maneira adequada a estratificação a ser realizada na população. Por exemplo, subdividir os eleitores em grupos definidos pela cor dos olhos não faz com que a amostra se torne mais homogênea ou mais representativa da população. Além disso, não parece haver uma relação lógica entre a cor dos olhos e a preferência por um determinado candidato (estudaremos esse conceito chamado de correlação nos próximos capítulos).

3. Variáveis

Existem basicamente dois tipos de variáveis (ou parâmetros) que consideramos em Estatística, e elas são divididas pelo tipo dos dados que são analisados. As **variáveis quantitativas** têm valores numéricos e têm uma relação de ordem bem estabelecida entre os elementos, já as **variáveis qualitativas** ou **variáveis categóricas** classificam os indivíduos em grupos ou categorias que não possuem uma relação de ordem entre elas. Com base na análise dessas variáveis dentro de uma amostra, podemos determinar estimativas para esses mesmos parâmetros dentro da população no procedimento de inferência estatística. Vejamos um exemplo para ilustrar todos os elementos apresentados até agora.

Exemplo 1: imagine um estudo sobre os gatos de rua de uma cidade. Nesse caso, estamos considerando como população todos os gatos de rua da cidade. Tomando uma amostragem por conveniência, podemos coletar dados sobre os indivíduos (gatos) que sejam mais dóceis e fáceis de capturar. Os dados coletados de cada indivíduo são:

- Peso
- Altura
- Cor da pelagem
- Cor dos olhos
- Local da captura
- Temperatura corporal

Cada um dos dados coletados é uma variável a ser considerada, sendo que as variáveis de peso, altura e temperatura corporal são variáveis quantitativas, enquanto que as variáveis de cor da pelagem e dos olhos, assim como o local da captura são variáveis categóricas.



Observe a tabela abaixo com os dados coletados de uma amostra de cinco indivíduos dessa população.

Gato	Peso	Altura	Cor da pelagem	Cor dos olhos	Local da captura	Temperatura
1	4kg	25cm	preta	amarelo	centro	38,6°C
2	3,1kg	22cm	preta	amarelo	centro	38,4°C
3	3,4kg	20cm	malhada	amarelo	Boa vista	39,1°C
4	2,8kg	20cm	malhada	amarelo	centro	38,6°C
5	3,2kg	23cm	malhada	amarelo	Pinheiros	38,9°C

A partir dessa amostra, podemos propor diversas questões como, por exemplo:

1. É possível concluir que todos os gatos de rua da cidade têm olhos amarelos?
2. Mais da metade da população de gatos se encontra no centro da cidade?
3. Todos os gatos de rua da cidade são pretos ou malhados?
4. A maioria dos gatos tem acima de 3kg?
5. A altura média dos gatos da cidade é de 22cm?

Como dito anteriormente, essa amostra foi tomada por conveniência e, portanto, pode não ser representativa do total da população. Além disso, com uma amostra com um número reduzido de indivíduos, fica ainda mais difícil inferir características para toda a população. Por isso, é sempre importante saber dimensionar o tamanho da amostra para que os resultados sejam significativos.

Os dados coletados de uma amostra ou de um experimento são geralmente apresentados como no exemplo anterior. Cada linha representa um indivíduo, e cada coluna, uma variável (ou característica). Abordaremos mais sobre esse e outros tipos de representação dos dados no próximo capítulo.

Exemplo 2: uma pizzaria decide realizar uma pesquisa sobre o comportamento de seus clientes, para que possa atualizar seu cardápio e incluir promoções que atendam às demandas dos pedidos. Para isso, uma planilha foi criada, em que cada pedido era analisado sobre as seguintes características:

1. Quantidade de pizzas no pedido.
2. O pedido foi acompanhado de bebida?
3. Alguma alteração no sabor original de alguma das pizzas foi requisitada?
4. Qual a forma de pagamento?
5. Qual a zona em que o pedido foi feito?
6. Esse foi o primeiro pedido do cliente?



Mostramos a seguir uma tabela com alguns dos dados coletados dos pedidos realizados em um final de semana.

Número do pedido	Pizzas	Bebida	Alteração	Pagamento	Zona	Primeiro pedido
#536	1	sim	sim	cartão	centro	não
#537	1	sim	não	cartão	sul	não
#538	2	sim	sim	cartão	sul	não
#539	1	não	não	cartão	oeste	sim
#540	1	não	não	dinheiro	oeste	não

#541	3	sim	não	cartão	sul	não
#542	1	não	sim	dinheiro	centro	sim
#543	2	sim	sim	cartão	centro	não
#544	2	sim	sim	cartão	leste	não

Neste caso, temos que cada pedido realizado na pizzaria é um elemento (ou indivíduo) dentro da população total. Aqui, a única variável quantitativa é a referente ao primeiro dado da pesquisa que determina o número de pizzas de cada pedido. Todas as outras variáveis são categóricas, sendo que três delas são agrupadas em apenas duas categorias (sim ou não). Esse tipo de variável é também chamada de **variável binária**, pois pode assumir apenas dois valores.

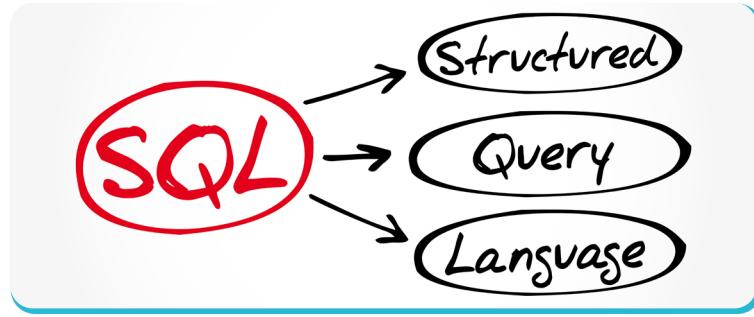
Após uma análise criteriosa e detalhada feita pelo dono da pizzaria, três novos sabores de pizzas foram adicionados ao cardápio e outras promoções foram criadas para atrair novos clientes e estimular pedidos com mais de uma pizza.



CURIOSIDADE

Esse tipo de pesquisa, que visa aumentar o lucro ou atualizar o tipo de serviço oferecido, é muito popular em marketing, publicidade e propaganda. Diversos bancos de dados são contratados e consultados por grandes empresas, e essa é uma área de atuação que vem crescendo muito no mundo todo, tanto a coleta e a organização de dados como a

análise de tais dados. A linguagem de programação SQL (*Structured Query Language*), criada na década de 1970, é um exemplo de linguagem que lida com bancos de dados gigantescos.



Os conceitos de população e amostra formam a base de qualquer coleção de dados e o princípio de toda análise estatística. Trabalhamos com diversos exemplos desses dois conceitos, assim como os diferentes processos de amostragem que podem ser utilizados para selecionar uma amostra de uma determinada população. Cada elemento da população é chamado de indivíduo e cada um deles possui características específicas que formalmente são classificadas como variáveis ou parâmetros. Vimos também as diferenças entre as variáveis quantitativas e qualitativas e como elas podem ser organizadas em simples tabelas. Destacamos a importância da coleta de dados e como um estudo pode ser tendencioso, caso não tomemos uma amostra adequada ou mesmo uma amostra que não seja tão significativa para o tamanho da população.



Referências

- CAMARGO, O. **Censo, a contagem da população**. Brasil Escola, [2021]. Disponível em: <https://brasilescola.uol.com.br/sociologia/censo-contagem-populacao.htm>. Acesso em: 1 jun. 2021.
- FONSECA, J. S. da; MARTINS, G. de A. **Curso de estatística**. 6. ed. São Paulo: Atlas, 2012.
- MOORE, David; NOTZ, William; FLIGNER, Michael: **A estatística básica e sua prática**. 7. ed. Rio de Janeiro: LTC, 2017.
- MORETTIN, P.; BUSSAB, W. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.
- SPIEGEL, M.; SCHILLIER, J.; SRINIVASAN, A. **Probabilidade e estatística**. 3. ed. Porto Alegre: Bookman, 2013.
- TRIOLA, M. **Introdução à estatística**. 12. ed. Rio de Janeiro: LTC, 2017.
- VIEIRA, S. **Fundamentos de Estatística**. 6. ed. São Paulo: Atlas, 2019.