



ISSN 2359-5051

# Revista Diálogos Interdisciplinares

## GEPFIP/UFMS/CPAQ

Grupo de Estudos e Pesquisa em Formação Interdisciplinar  
de Professores

---

### INTRODUÇÃO À ESTATÍSTICA: HISTÓRIA, FINALIDADE E PROCESSO DE AMOSTRAGEM

### INTRODUCTION TO STATISTICS: HISTORY, PURPOSE AND SAMPLING PROCES

Antonio Sales<sup>1</sup>

Alessandra Pucci Mantelli Galhardo<sup>2</sup>

Leda Márcia de Araújo Bento<sup>3</sup>

#### RESUMO

Este trabalho tem finalidade didática e nele destacamos o papel da estatística e a sua relevância nas pesquisas quantitativas. Apresentamos um resumo histórico, a sua relevância atual e a contribuição dos softwares. O trabalho apresenta um mapa conceitual das variáveis, a diferença entre as pesquisas qualitativa e quantitativa, do ponto de vista da utilização da Estatística. Ele foi pensado para uma aula para pesquisadores iniciantes da área da saúde, mas os princípios expostos servem indistintamente para qualquer área, inclusive Educação e Sociologia. Indicamos procedimentos para o cálculo da amostra, distinguimos amostra probabilística de amostra por conveniência e apontamos formas de obter estimadores para o cálculo da amostra.

**Palavras-chave:** Estimadores Estatísticos. Softwares Estatísticos. Variáveis Estatísticas

#### ABSTRACT

This work has didactic purposes and in it we highlight the role of statistics and its relevance in quantitative research. We present a historical summary, its current relevance and the contribution of software. Presents a conceptual map of variables, the difference between qualitative and quantitative research, from the point of view of the use of Statistics. It was designed for a class for health researchers, but the principles exposed are applicable to any area, including Education and Sociology. We indicate procedures for sample calculation, distinguish probability samples from convenience

---

<sup>1</sup> Licenciado em Matemática, Bacharel em Filosofia, Mestre e Doutor em Educação. Docente do Curso de Medicina e do Mestrado em Ensino de Ciências e Matemática da Universidade Anhanguera-Uniderp. [profesales@hotmail.com](mailto:profesales@hotmail.com)

<sup>2</sup> Formada em Odontologia, Doutora em Ciências Odontológicas, Pós-Doutora e Pesquisadora Voluntária da USP. Professora Adjunta da Faculdade de Odontologia da Universidade Federal de Mato Grosso do Sul.

[alessandra.galhardo@ufms.br](mailto:alessandra.galhardo@ufms.br)

<sup>3</sup> Bióloga, Mestre e Doutora em Farmacologia. Docente do Curso de Medicina da Universidade Anhanguera-Uniderp. [leda.bento@cogna.com.br](mailto:leda.bento@cogna.com.br)



samples and point out ways to obtain estimators for sample calculation.

**Keywords:** Statistical Estimators. Statistical Software. Statistical Variables.

## 1 INTRODUÇÃO

Sentindo falta de um material didático, escrito em linguagem acessível, que trate da importância da Estatística, para as pesquisas, nos propusemos em elaborar este material que esclarece os processos de amostragem e indica o papel dos softwares e da planilha do Excel no tratamento estatístico. Os exemplos dados são da área da saúde, mas as orientações apresentadas se aplicam a pesquisas em todas as áreas. Destacamos que toda pesquisa envolvendo seres humanos e animais devem ser submetidas a um Comitê de Ética em Pesquisa com Seres Humanos ou a uma Comissão de Ética sobre Uso de Animais. Esses Comitês ou Comissões são compostos por pesquisadores experientes e por representantes da comunidade, todos com atuação voluntária. Costumam ser rigorosos, sempre visando o bem do participante e a segurança jurídica do pesquisador.

O Comitê deve ser consultado tanto por pesquisadores da área da saúde quanto da educação ou sociologia. Pesquisas qualitativas ou quantitativas, necessitam, igualmente, da apreciação de um Comitê. Para ter acesso a esses comitês ou submeter um projeto para apreciação o pesquisador precisa ter o Currículo Lattes e ser inscrito na Plataforma Brasil<sup>4</sup>. Atualmente a maioria das revistas exigem que os artigos submetidos tenham a aprovação de um Comitê de Ética.

Na área de educação há pouca produção de pesquisa quantitativa o que não significa que ela não tenha contribuição para essa área.

Recentemente avaliamos um trabalho muito interessante, envolvendo o resultado de um trabalho com estudantes da Educação Básica, submetido a uma revista. Lamentavelmente tivemos que devolver ao autor porque não fizera indicação de que fora aprovado por um Comitê de Ética e alguns princípios estatísticos não foram bem manipulados. São experiências como essa que impulsionam a produção de um trabalho nessa perspectiva.

Para a elaboração deste trabalho procedeu-se uma revisão bibliográfica, preferencialmente, em materiais produzidos com finalidade didática. Recorreu-se a livros-texto de Estatística Aplicada para as áreas de formação acadêmica nas quais a pesquisa quantitativa é a propulsora da produção de conhecimentos. Alguns artigos também foram consultados, especialmente os que explicitam a metodologia utilizada e justificam o procedimento adotado.

Sabe-se que um pesquisador procura prever ou avaliar o resultado da aplicação de determinado método de trabalho sobre a população e da implementação de certo insumo agrícola na produção.

<sup>4</sup> Links de acesso: <https://lattes.cnpq.br/> ; <https://plataformabrasil.saude.gov.br/login.jsf>

Busca saber o que produz o aparecimento ou desaparecimento, de certos fenômenos sociais e prever as consequências. Um pesquisador em saúde, por exemplo, tem interesse em predizer o comportamento de certas patologias, seus efeitos sociais e a sua influência na qualidade de vida das pessoas. Mas, ele não se detém somente no estudo das patologias. Pode estudar também as condições de saúde, e de vida, em geral, de uma pessoa. Ele também faz previsões construtivas sobre a realidade de saúde de uma comunidade, de uma cidade ou de um povo (Levin; Fox, 2004).

Nesse processo de investigação, analisa variáveis, isto é, a frequência de certa patologia, impactos de uma pandemia sobre a população, comportamento humano resultante de certas doenças, características físicas ou sociais de uma faixa etária e, assim por diante. Avalia o nível de compreensão de um determinado problema de saúde, nível de comprometimento pessoal para o seu próprio bem-estar, fatores contributivos para que determinada patologia, ou mesmo epidemia possa se instalar em um corpo ou em uma comunidade. Estuda condições de vida em certa comunidades.

Na Sociologia também são estudadas as condições de vida de comunidades, envolvendo questões econômicas, escolaridade, parentesco, produtividade, satisfação com a vida, vitimização, incidência de certa opinião na comunidade, o uso de determinadas palavras, entre outros temas (Collares, 2013; Gonçalves; Santos 2023). Um educador pode ter o interesse em avaliar o efeito de uma sequência didática sobre o aprendizado da sua disciplina escolar.

Entretanto, deve também levar em conta que "Nem todas as características humanas variam. É um fato da vida, por exemplo, que o sexo de uma pessoa, que gera outra, é feminino. Portanto, o grupo de indivíduos que tem o sexo da mãe é a constante feminino" (Levin; Fox, 2004, p.2).

É possível que em alguma ciência (biologia, por exemplo) ou contexto, se interrogue como é possível que a fêmea gere outro ser de sexo igual ou diferente do seu. Apesar deste fato ser considerado como um caso resolvido (dado que não varia), por vezes ele pode ser investigado sob outro ponto de vista, como por exemplo, o que esse potencial gerativo de vida significa para a fêmea de dada espécie. A variável, nesse caso, pode ser o significado da maternidade. Ou, talvez, como uma mulher se torna mãe. Melhor dizendo: em que difere o apego de uma mãe biológica do apego de uma mãe afetiva?

Ao considerarmos como um caso resolvido o fato de a mãe ser sempre do sexo feminino, essa informação não pode ser usada para explicar um aborto espontâneo ou uma má formação congênita. Para um estudo como esse é necessário levar em conta outros fatores, isto é, outras variáveis (Levin; Fox, 2004).

Talvez possa se conjecturar que a idade da mãe ou o seu estado emocional durante a gravidez, ou ainda sua condição social com a presença de determinada patologia, nesse período, seja o fator desencadeante do aborto ou da má formação. Essas conjecturas são as hipóteses de trabalho do

pesquisador. A etnia, por exemplo, ou a genética da pessoa pode ser um fator desencadeante de uma patologia. Sabe-se de certas patologias que são mais comuns em algumas etnias. Na educação escolar pode-se perguntar, por exemplo, por que estudantes ficam mais tensos em véspera de prova de determinada disciplina.

Por vezes o pesquisador adota uma unidade de observação para sua pesquisa. Nesse caso cada pessoa entrevistada é a sua unidade. Entretanto, ele pode também focalizar a sua pesquisa em agregados, caso queira saber como as medidas variam entre as pessoas de um grupo ou de uma aldeia. Por exemplo, a relação entre idade média de um grupo de mães e o índice de má formação de seus bebês. Ou, ainda, o número médio de tempo de atividade física semanal e o rendimento acadêmico de estudantes. Desse modo, seja centrando a sua atenção em indivíduos ou grupos, a ideia que o pesquisador faz da natureza do problema exerce influência sobre a tomada de decisão, a escolha da variável e a sua hipótese (Levin; Fox, 2004).

Há pesquisas experimentais e não experimentais. Uma pesquisa experimental difere de uma pesquisa de levantamento ou observacional pelo grau de controle que o pesquisador pode ter sobre a situação de pesquisa. Quando se trata de seres humanos, porém, as pesquisas experimentais costumam ter mais limites do que quando realizadas com animais ou plantas. Mesmo que se crie o grupo controle ou grupo experimental, quando o participante é um ser humano, não se pode garantir que ele use o medicamento com a regularidade correta ou na quantidade estipulada, por exemplo.

Há ainda as pesquisas do tipo Survey (formulários) que são consideradas retrospectiva porque os registros são feitos após os fatos ocorridos. Elas apresentam a vantagem de não sofrer os efeitos da manipulação por parte do pesquisador, porém, por serem retrospectivas, dependem da memória. O pesquisador precisa levar em conta que aquilo que foi assinalado não são os únicos fatos que ocorreram, mas os que mais marcaram o participante (Levin; Fox, 2004).

Quanto à forma de abordar ou analisar o material nas pesquisas, pode-se dizer que há duas maneiras de fazê-lo: a forma qualitativa e a forma quantitativa. Não se pode confundir abordagem com variáveis. As variáveis também podem ser qualitativas ou quantitativas, contudo, tanto uma como a outra pode ser tratada qualitativa ou quantitativamente.

Para diferenciar uma abordagem da outra concentremos a nossa atenção, primeiramente, na perspectiva qualitativa. Ela leva em conta a subjetividade do participante uma vez que, conforme Esteban (2010), nessa perspectiva, o significado não resulta da interação sujeito-objeto, mas de uma imposição do primeiro. Procura saber qual o sentido que essa pessoa dá a determinado fenômeno. Por exemplo: o que significa, para uma mulher, ter um aborto espontâneo? O que significa, para um casal, a constatação de que um dos dois é estéril?

A abordagem quantitativa sobre esses mesmos fenômenos, por sua vez, leva em conta número



de abortos espontâneos ocorridos em determinada população, em determinado intervalo de tempo. O número de casais que se divorciam porque um dos dois é estéril, e assim por diante. Ela quantifica a presença do fenômeno. A abordagem quantitativa fundamenta-se na estatística, enquanto a abordagem qualitativa não tem na estatística a sua ferramenta principal (Dalfovo; Lana; Silveira, 2008).

Segundo Gabriel (2014, p. 352) os que defendem a abordagem qualitativa são os que se opõem “ao chamado positivismo e defendem a superioridade do construtivismo, idealismo, relativismo, humanismo, hermenêutica e, em alguns casos, pós-modernismo”. Admitem realidades múltiplas. Para esse mesmo autor, as pesquisas quantitativas podem ser enquadradas nas categorias: experimental e não experimental. Por sua vez, as pesquisas quantitativas são classificadas nas categorias: a interativas e as não interativas.

Callegari-Jacques (2003, p.13) inicia o seu trabalho afirmando que a Estatística não é mais uma complicação matemática porque ela vem se mostrando “um instrumento útil na organização e interpretação dos dados”. Ressalta ainda que “essa ciência propicia uma avaliação adequada da variabilidade observada nos processos biológicos”. A mesma autora vai além e afirma que: “é sabido que existem diferenças entre indivíduos e que eles reagem de forma diferente a estímulos idênticos; por outro lado, o mesmo indivíduo apresenta variações de um momento para outro”. Tendo isso em mente o pesquisador deseja saber “qual o grau de confiabilidade dos resultados”. A estatística é a ferramenta que permite a aferição dos dados trabalhados.

A estatística, “ciência que estuda o comportamento dos fenômenos chamados coletivos” (Batanero, 2013, p. 1), está dividida em duas grandes áreas, a saber: estatística descritiva, que envolve o resumo e a apresentação dos dados (gráfico, tabelas, porcentagens), e a estatística inferencial que permite concluir sobre conjuntos maiores de dados, ou seja, sobre populações, quando apenas uma parte desses conjuntos (amostra) foi estudado. As hipóteses são verificadas pela estatística inferencial. É com ela que se procedem testes de hipóteses científicas que influenciam na tomada de decisões sobre populações.

Procurou-se, neste trabalho, justificar a razão da sua produção destacando a carência de material que seja, ao mesmo tempo, simples na linguagem, completo em relação ao fim a que se destina, contextualizado e detalhado. A motivação que norteou a sua produção se resume na forma como esse conteúdo aparece disperso em capítulos de livros destinados a um público específico: o professor que ministra a disciplina em cursos de graduação nas áreas biológicas, da saúde, da terra e exatas. Também aparecem, resumidamente, em roteiros de aula de pós-graduação das áreas mencionadas. Textos estes que, embora tenham finalidade didática, não têm redação acessível.

O contato com professores, que orientam Trabalho de Conclusão de Curso de Graduação em Licenciatura (e até dissertação de mestrado), revelou que é frequente vê-los abonar trabalhos com



pequenos números chamando-os de “trabalhos estatísticos”<sup>5</sup> desacreditizando uma ciência respeitada no meio acadêmico.

Tendo em vista o exposto, a carência de um material voltado para os que pretendem se inserir na arte da pesquisa, o objetivo deste artigo é apresentar, em linguagem didática, noções fundamentais de estatística descritiva e inferencial, com ênfase em variáveis, tipos de amostragem e estimadores para cálculo amostral.

O leitor encontrará informações sobre a importância da Estatística, diferença entre pesquisa qualitativa e quantitativa, pesquisa experimental e não experimental, na introdução. A história da Estatística, o processo de amostragem, fórmulas de cálculo e a distinção entre diversas variáveis, conhecimentos estes fundamentais para qualquer pesquisador, se encontram ao longo do texto a seguir.

## 2 A ORIGEM E ATUALIDADE DA ESTATÍSTICA

Fontelles (2012) informa que a palavra Estatística deriva da mesma palavra latina que deu origem à palavra estado, porque nos primórdios eram os estados que se ocupavam de “observar e descrever os aspectos inerentes a uma dada população”, e que:

hoje, essa ciência tem por objetivo a observação, a coleta, a classificação, o resumo, a organização, a análise e a interpretação de dados, bem como a análise de fenômenos coletivos ou de massa, e, também, a indução das leis que tais fenômenos obedecem, tornando-se, portanto, uma poderosa ferramenta a fornecer um conjunto de técnicas de análise de dados e auxiliar na tomada de decisões científicas (Fontelles, 2012, p. 16).

Atualmente ela está presente em todos os centros de pesquisa que envolve a saúde e a economia, enfim que se ocupam de pesquisar dados relativos a populações. Dispomos de softwares especializados que popularizaram a Estatística nos meios acadêmicos trazendo agilidade e facilidade na sua aplicação. Esses softwares são fundamentados em fórmulas matemáticas o que nos informa que a Estatística tem na Matemática o seu fundamento.

A aplicação dos métodos estatísticos à saúde recebeu o nome de Bioestatística e, falando dela, Fontelles (2012, p. 18) afirma que não se limita a questões matemáticas como cálculo de médias, desvio-padrão ou ao uso de gráficos e tabelas porque ela ajuda o profissional de saúde a “avaliar o método utilizado e os resultados apresentados em artigos científicos publicados por literatura especializada”. Convém, não obstante, observar que nem sempre estão reunidos num mesma pessoa um pesquisador e um estatístico. Os softwares agilizam o cálculo, mas nem sempre indicam o procedimento inicial recomendado ou fornecem ajuda suficiente para a interpretação dos

<sup>5</sup> A propósito, está prevista a submissão de um artigo didático que discute, entre outras coisas, a pertinência ou não de graficos e tratamentos percentuais com pequenos números.

resultados.

Não cabe aqui discutir qual o software mais recomendado ou mais eficiente porque todos se fundamentam nos mesmos princípios matemáticos. O que os diferencia é a quantidade e a espécie de serviços que disponibilizam. A maioria deles importa dados diretamente do Excel. Dessa forma, a escolha do usuário recai na espécie de serviço que se precisa, naquele momento e a familiaridade com a interface do software.

Memória (2004) destaca que o simples uso de dados numéricos não se constitui em um tratamento estatístico. Os registros de batismo, casamentos e óbitos pela Igreja Católica Romana, não são, necessariamente, informações estatísticas. Da mesma forma, o levantamento populacional por um governante pode não ser uma atividade estatística. Atividade estatística é o tratamento sistematizado dessas informações. É a tentativa de tirar conclusões, estabelecer relações quantitativas dos dados entre si e entre os dados e a realidade vivida, na época em que foram produzidos ou na atualidade.

Conforme exposto em parágrafos precedentes, este é texto didático. O seu público-alvo é composto por estudantes de pós-graduação de programas que se alinham, preferencialmente, com a pesquisa quantitativa ou pesquisadores, que trabalham com pesquisas qualitativas e são desafiados a orientar Iniciação Científica com dados quantitativos. Os parágrafos seguintes entram em detalhes desse processo.

### **3 AMOSTRAGEM UM IMPORTANTE PROCESSO ESTATÍSTICO**

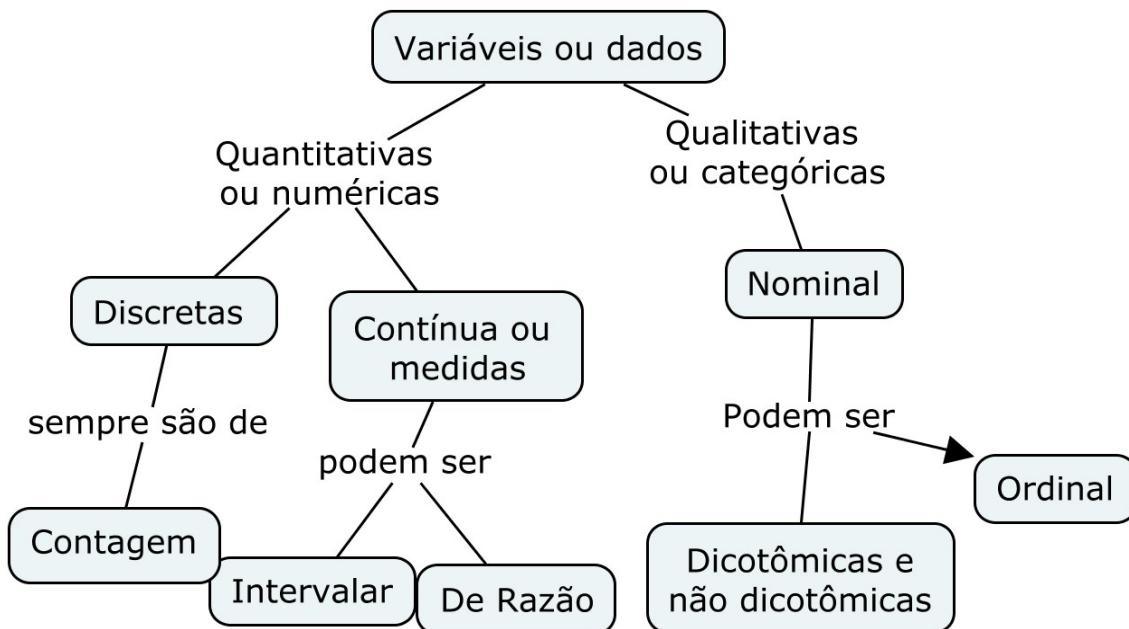
Marotti *et al.* (2008, p.187) afirmam que “Na realização de qualquer estudo, é praticamente impossível examinar todos os elementos da população de interesse”. Segundo esses autores usa-se a inferência para generalizar os dados da amostra e concluem que “o cálculo do tamanho da amostra deve fazer parte de qualquer projeto de pesquisa clínica. A principal meta é estabelecer, objetivamente, qual o número de indivíduos que necessitam ser estudados” (Marotti *et al.*, 2008, p. 187). Os autores citados estão se referindo a amostras que representam a população. Entretanto, há amostras que não representam a população. Vem, desse fato, a importância de um cálculo para a determinação da amostra e um tratamento tal que a torne probabilística, isto é, aleatória.

Trataremos da determinação da amostra. Vale lembrar que quando se fala em tratamento estatístico, faz-se referência à amostra, uma vez que a observação ou coleta de dados de toda uma população recebe o nome de censo. Conforme já foi dito, a amostra tem a ver com a quantidade de pessoas, animais ou casos que serão observados ou analisados. A variável é o fenômeno que será observado. É o que está acontecendo e que despertou a atenção do pesquisador. Gianini (2008, p.

1) define variável como “um atributo mensurável que tipicamente varia entre indivíduos”

A figura (fig. 1) a seguir sintetiza os tipos de variáveis possíveis de serem encontradas.

**Figura1- Mapa conceitual das variáveis**



Fonte: Resumido de vários autores, dentre eles, Crespo (2002).

Exemplos de variáveis de acordo com a categoria:

### 3.1 Variável intervalar (zero relativo)

Temperatura, Conhecimento e Qualidade de Vida.

### 3.2 Variável de Razão (zero absoluto)

Massa Corporal, Estatura, Número de Itens no Estoque, Pressão Arterial, Batimentos Cardíacos, Dimensão da Parte Lesionada em um acidente, Idade. É tudo o que pode ser medido ou contado.

### 3.3 Nominal (categórica) ordinal

Nível Salarial, Faixa Etária, Nível de Atividade Física, Nível de Escolaridade, Nível de Colesterol (Baixo, Normal, Alto). A taxa de colesterol é variável de razão.

### 3.4 Nominal não ordinal

Sexo, Tipo Sanguíneo, Estado Civil, Cor da Embalagem, Cor dos Olhos, Etnia. Dependendo da pergunta formulada algumas variáveis nominais podem se tornar dicotômicas. Por exemplo, se perguntarmos pelo cor dos olhos de alguém termos quatro opções de respostas: verde, azul, castanho ou preto. Porém, se a pergunta for direcionada para uma cor, a resposta será sim ou não, logo dicotônica.

## 4 TIPOS DE AMOSTRAS

Seção destinada a discutir o processo de amostragem.

### 4.1 Amostra aleatória e amostra não aleatória

Dada uma população, uma amostra aleatória (randomizada) é uma amostra em que qualquer elemento da população *tem alguma probabilidade* de ser selecionado para participar. Se é uma amostra que tem por base um rigoroso cálculo amostral e os participantes são escolhidos aleatoriamente então é uma amostra probabilística. Numa amostra não aleatória, alguns elementos da população podem ficar propositadamente sem ser selecionados para participar (Fontelles, 2012).

Considerando que se vai fazer inferência sobre a população a partir da mostra, esta deve ser representativa. Um processo para determinar o tamanho da amostra é fazer um estimador (Martins, 2006). Existem fórmulas para o cálculo da amostra. Mas, de onde saem os valores para as fórmulas?

Quando a variável principal é um número, uma quantidades (idade, altura, massa corpórea, pressão arterial, taxa glicêmica), usa-se um tipo de estimador: a variância ou desvio-padrão. Calculase a variância de uma pequena amostra e com ela se determina o tamanho da amostra. Pode-se usar também o desvio-padrão aproximado (Fontelles, 2012, Triola, 1999).

Quando a variável é qualitativa, usa-se outro tipo de estimador: a proporção. Um **estimador** é um dado importante para estimar ou tamanho da amostra. Essa regra é baseada em dados observados. Seja a presença de uma patologia, um hábito, etnia ou presença da gravidez.

Pode-se falar em **estimador centrado** que é aquele em que o seu valor médio coincide com o parâmetro a estimar (Martins, 2006). Quanto maior o tamanho da amostra menor o enviesamento. Para variáveis quantitativas ele é determinado pela variância populacional  $\sigma^2$  ou pelo desvio-padrão aproximado. Para variáveis qualitativas ele é determinado pela proporção da presença (prevalência) do fenômeno na população.

Onde encontrar esses estimadores?

a) Na literatura. Alguma pesquisa já revelou essa proporção ou o desvio-padrão em algum

outro lugar.

b) Projeto piloto. Numa amostra aleatória (de  $n > 30$ ) busca-se encontrar a proporção presente na população.

Fazem-se pequenas amostras ( $n > 30$ ) para conseguir o estimador (Triola, 1999; Martins, 2006). Luchesa e Chaves Neto justificam essa opção afirmando que:

Ou então, poderemos usar o desvio padrão amostral,  $s$ , desde que seja tomada uma amostra igual ou maior do que 31 elementos; desde que  $n \geq 31$ . O Teorema do Limite Central garante que amostras maiores do que 30 são normais, independentemente do formato da distribuição de probabilidade da população da qual está sendo retirada a amostra (Luchesa; Chaves Neto, 2011, p.11).

Escolhe-se uma variável importante e estima a proporção (em percentual) dela ou o desvio-padrão. Este estimador depende muito da variabilidade da população subjacente.

Exemplo: Supondo que a variável mais importante seja o uso rotineiro de água mineral gaseificada. Deseja-se saber se ela tem a ver com o aparecimento de certo desconforto estomacal na população. Nesse caso, pode-se usar como estimador a proporção ( $p$ ). Onde  $p^6$  é um percentual que se expresse por número decimal.

Em outros casos usa-se o desvio-padrão aproximado que se determina “dividindo a amplitude total por quatro” (Triola, 1999, p. 149):  $\sigma = \frac{\text{amplitude total}}{4}$

Exemplos de desvios-padrão aproximados.

Idade numa classe de jovens, 19 a 25 anos, varia muito pouco e a amostra pode ser pequena. Desvio-padrão aproximado, dessa turma, segundo Triola, é  $(\frac{25-19}{4})=1,5$ .

Para a idade numa classe noturna, que varia de 20 a 70 anos, deve-se ter uma amostra grande. Nesse caso, o desvio-padrão aproximado é  $(\frac{70-20}{4})=12,5$

#### 4.2 Há entretanto, que calcular a dimensão de uma amostra

Há expressões matemáticas usadas para o cálculo das amostras e os softwares fazem isso com muita agilidade. Essas fórmulas são apresentadas aqui para que sejam conhecidas, ainda que de passagem, pelo pesquisador. Supomos que ele queira ter alguma informação sobre o que entra em jogo na determinação de uma amostra. Poderá necessitar discutir isso com os pares ou com o editor de um periódico científico. O software utilizado vai lhe pedir alguns dados que são apresentados a seguir.

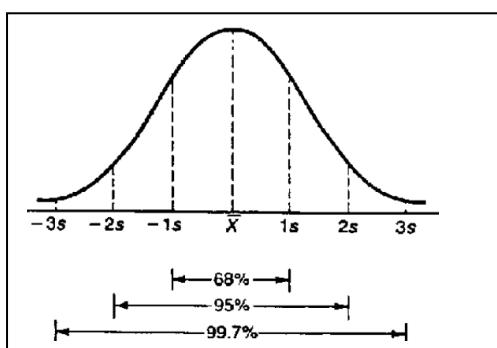
Nessas expressões matemáticas usam-se os seguintes valores:

<sup>6</sup>Estamos usando  $p$  para indicar a proporção, ou índice de prevalência, usada como estimador ( $0 \leq p \leq 1$ ).

a) Para um nível de confiança de 95%, tem  $-z=1,96$  (valor fixo); b) A margem de erro ( $e$ ) é escolha do pesquisador e normalmente varia de 2% (0,02) a 5% (0,05); c) o sigma  $\sigma$  representa o desvio-padrão populacional e pode ser aproximado; d) N, por sua vez, representa quantidade de elementos da população; e) n representa o número amostral; f) p é a proporção, em decimal sabendo que  $1=100\%$

Para a determinação dos valores de z os estudiosos levaram em conta a curva normal (ver figura), ou curva de Gauss, onde se define que a média  $\pm \sigma$  ( $\mu-\sigma; \mu+\sigma$ ) contempla 68,2% dos dados, média  $\pm 1,96\sigma$  ( $\mu-1,96\sigma; \mu+1,96\sigma$ ) contempla 95% dos dados e média  $\pm 2,58\sigma$  ( $\mu-2,58\sigma; \mu+2,58\sigma$ ) contempla 99% dos dados (fig. 2).

**Figura 2- Curva normal com valores aproximados**



Fonte: Bertolo (2012)

Segundo Fonseca e Martins (1996), para populações finitas e variável quantitativa intervalar usa-se a expressão matemática (fig. 3):

**Figura 3- Determinação da amostra para populações finitas com variável quantitativa**

$$n = \frac{z^2 \cdot \sigma^2 \cdot N}{e^2 \cdot (N-1) + z^2 \cdot \sigma^2}$$

E, se a variável for qualitativa, aplica-se a seguinte fórmula (fig 4):



**Figura 4- Determinação da amostra para populações finitas com variável qualitativa**

$$n = \frac{N \cdot p \cdot (1 - p) \cdot z^2}{p \cdot (1 - p) \cdot z^2 + (N - 1) \cdot e^2}$$

Exemplos:

a) Supondo que se queira determinar a massa corporal (peso) média de uma turma de 200 estudantes em que o projeto piloto (pequena amostra de 31 pessoas) revelou que essa massa corpórea varia de 50kg a 100kg. Nesse caso, o desvio-padrão aproximado será de  $(100-50)/4=50/4= 12,5$ . Admitindo um erro de 0,05 e aceitando um nível de confiança de 95% ( $z=1,96$ ) tem-se  $n=((1,96)^2 \times (12,5)^2 \times 200) / [(0,05)^2 \times 199 + (1,96)^2 \times (12,5)^2] = 120.050 / (0,4975 + 600,25) = 120.050 / 600,7475 = 199,8$ , o que mostra que para populações pequenas e variabilidade grande a amostra pode coincidir com a população.

b) Supondo que se queira determinar a incidência de obesos nessa população e que o projeto piloto (pequena amostra de 31 pessoas) indicou que essa proporção é de 25%. Nesse caso a amostra será de  $(200 \times 0,25 \times 0,75 \times 1,96^2) / (0,25 \times 0,75 \times 1,96^2 + 199 \times 0,05^2) = 144,6 / 1,22 = 118$ .

c) No caso de dificuldade para se obter uma pequena amostra opta-se por fazer  $p=50\%=0,5$ . Isso significa que se supõe que 50% da população apresenta a característica em análise<sup>7</sup>.

Enfatizamos que o objetivo aqui não é ensinar a calcular uma amostra, mas mostrar a origem dos dados que serão necessários para que um software ou um estatístico calcule-a corretamente .

Quando se tem uma população muito grande, ou de difícil precisão do número exato, pode-se considerá-la população infinita e, nesse caso, Agranomik e Hirata (2011) indicam o uso da seguinte fórmula, para variáveis qualitativas (fig. 5):

**Figura 5- Para população muito grande e variável qualitativa**

$$n = \frac{z^2 \cdot p \cdot (1 - p)}{e^2}$$

<sup>7</sup>  $p=0,5$  é uma prevalência presumida para os casos em que se desconhece a prevalência real (Medronho, 2009). Considerando que no cálculo usa-se o produto  $p(1-p)$  e que  $0 \leq p \leq 1$ , fazendo  $p=0,5$  obtém-se maior amostra.

E a fórmula para variáveis quantitativas, apresentada por eles, é (fig. 6):

**Figura 6- Para população muito grande e variável quantitativa**

$$n = \left( z \cdot \frac{\sigma}{e} \right)^2$$

Segundo Martins (2006) uma população é considerada muito grande, ou infinita, quando o número estimado da população for, pelo menos, igual a 20 vezes o tamanho da amostra ( $N \geq 20 \times n$ ).

#### **4.3 Amostra de Conveniência**

Por último, pode-se falar em amostra de conveniência. Ela não possui o poder preditivo da amostra probabilística, mas pode ser usada como um indicador. Segundo Motta e Wagner:

trata-se de um procedimento não-probabilístico, mas isso não significa que não seja útil e até válido. Em pesquisas clínicas, muitas vezes, é inviável realizar um sorteio entre todos os componentes da população-alvo. Nesses casos, a amostra de conveniência pode representar uma solução razoável, sendo mais rápida e de menor custo. A inclusão sequencial (consecutiva) é preferível porque tende a menores distorções. O julgamento sobre o fato de uma amostra de conveniência ser ou não ser representativa da população-alvo de interesse é totalmente subjetivo (Motta; Wagner, 2003, p. 27).

Conveniência, neste caso, não deve ser tomada com o sentido de permitir que se enviesse, escolhendo os sujeitos conforme convier ao pesquisador, ou que se escolham aqueles cujos resultados são previstos. Conveniência no sentido de que não se fixa previamente o tamanho da amostra, mas prefixa-se o tempo de coleta de dados, por exemplo. Pode ter também o sentido de não se desconsiderar os dados porque o total da amostra não foi atingido por fatores alheios à vontade do pesquisador. Pode ser conveniente estipular, como amostra, os sujeitos que comparecem em determinado local em determinado dia ou época do ano.

As pesquisas qualitativas trabalham com amostras de conveniência e têm a sua importância reconhecida no mundo acadêmico. Elas atendem demandas que a pesquisa quantitativa não consegue contemplar. Pesquisas qualitativas e quantitativas são complementares entre si e não oposições.

#### **5 PALAVRAS FINAIS**



A Estatística é uma ciência que, quando bem utilizada, é ferramenta fundamental para tratar dados, comparar resultados de experimentos, avaliar trabalhos acadêmicos, fazer inferências e concluir sobre populações a partir de amostras.

Com a disponibilidade de softwares, tornou-se obsoleto o trabalho manual de calcular amostras, correlações, e testes diversos. Tudo isso hoje é feito com a ajuda da informática, cada vez mais aparelhada. Desse modo, as fórmulas aqui apresentadas são apenas para informação de quem tem a curiosidade de saber como funciona ou o que se oculta sob a tela do computador quando um software estatístico está sendo utilizado.

## 6 REFERÊNCIAS

AGRANOMIK, M.; HIRATA, V. N. Cálculo de tamanho de amostra: proporções. **Revista HCPA**, v.31, n.3, p.382- 388, 2011. Disponível em: <https://seer.ufrgs.br/hcpa/article/view/23574> Acesso em: 12 nov. 2024.

BATANERO, C. Sentido estadístico: componentes y desarrollo. **I Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria**. Granada, 2013. Disponível em: <https://www.ugr.es/~batanero/pages/ARTICULOS/Sentidoestad%C3%ADstico.pdf> Acesso em: 14 dez. 2024.

BERTOLO, L.A. **Estatística Descritiva**: aula 3. Publicada em 2012. Disponível em: <https://www.bertolo.pro.br/FinEst/Estatistica/EstatisticaNosNegocios/l3.html> Acesso em: 22 Fev. 2024.

CALLEGARI-JACQUES, S. M. **Estatística**: princípios e aplicações. Porto Alegre: Artmed, 2003.

CRESPO, Antônio Arnot. **Estatistica Fácil** 18.ed. São Paulo: Saraiva: 2002.

COLLARES, A. C. M. Uma Questão de Método: desafios da pesquisa quantitativa na sociologia. **Idéias**, Edição especial, p. 109-135, 2013. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/ideias/article/download/8649415/15970> Acesso em: 15 nov. 2024.

DALFOVO, M. S.; LANA, R. A.; SILVEIRA, A. Métodos Quantitativos e Qualitativos: Um Resgate Teórico. **Revista Interdisciplinar Científica Aplicada**, Blumenau, v.2, n.4, p.01- 13, Sem II. 2008. Disponível em: <https://portaldeperiodicos.animaeducacao.com.br/index.php/rica/article/view/17591> Acesso em: 15 fev. 2024.

ESTEBAN, M.P.S. **Pesquisa qualitativa em educação**: fundamentos e tradições. Porto Alegre: AMGH, 2010

FONSECA, J. S.; MARTINS, G. A. **Curso de Estatística**. 6.ed. São Paulo: Atlas, 1996.

FONTELLES, M. J. **Bioestatística aplicada à pesquisa experimental**: volume 1. São Paulo; Livraria da Física, 2012.



GABRIEL, M. L. D. Métodos Quantitativos em Ciências Sociais: Sugestões para Elaboração do Relatório de Pesquisa. **Desenvolvimento em Questão**, v.12, n. 28, p. 348-369, 2014.

GIANINI, R. J. Critérios de Avaliação da Qualidade de Resultados de Pesquisas em Saúde: a escolha do teste estatístico. **Revista da Faculdade de Ciências Médicas de Sorocaba**, v. 10, n. 3,p. 1- 4, 2008. Disponível em: [https://scholar.google.com.br/scholar?hl=pt-BR&as\\_sd=0%2C5&q=o+poder+de+um+teste+estat%C3%ADstico&btnG=](https://scholar.google.com.br/scholar?hl=pt-BR&as_sd=0%2C5&q=o+poder+de+um+teste+estat%C3%ADstico&btnG=) Acesso em: 18 nov. 2024.

GONÇALVES, A. P. V.; SANTOS, R. S. P. A Pesquisa Quantitativa em Sociologia: recursos e dilemas da realização de surveys online com trabalhadores durante a Pandemia de Covid-19. **Sociologia e Antropologia**, v. 13, n.02p. 1-21, 2023. Disponível em: <https://www.scielo.br/j/sant/a/dW3qwyVNdfFcwnCKHSNMYBq/?format=pdf&lang=pt> Acesso em 20 nov. 2023.

LEVIN, J.; FOX, J. A. **Estatística para Ciências Humanas**. 9.ed. São Paulo: Prentice Hall, 2004.

LUCHESA, C. J.; CHAVES NETO, A. **Cálculo do tamanho da amostra nas pesquisas em Administração**. Curitiba: Edição do autor, 2011(Unicuritiba). Ebook disponível em:. [https://silo.tips/download/calculo-do-tamanho-da-amostra-nas-pesquisas-em-administraao#google\\_vignette](https://silo.tips/download/calculo-do-tamanho-da-amostra-nas-pesquisas-em-administraao#google_vignette) Acesso em: 16 fev. 2024.

MAROTTI, J.; GALHARDO, A. P. M.; FURUYAMA, R. J. PIGOZZO, M. N.; CAMPOS, T. N.; LAGANÁ, D. C. Amostragem em Pesquisa Clínica: tamanho da amostra. **Revista de Odontologia da Universidade Cidade de São Paulo**. v. 20, n.2, p. 186-194, 2008.

MARTINS, M. E. G. **Introdução à Inferência Estatística**. Lisboa: Faculdade de Ciências da Universidade de Lisboa, 2006. Disponível em: <http://homepage.ufp.pt/cmanso/ALEA/introInfEstat.pdf> Acesso em: 17 fev. 2024.

MEMÓRIA, J. M. P. **Breve história da estatística**. Brasília, DF: Embrapa Informação Tecnológica, 2004.

MEDRONHO, R.A.; BLOCH, K.V.; LUIZ, R. R.; WERNECK, G.L.(org.). **Epidemiologia**. 2.ed. São Paulo: Atheneu, 2009.

MOTTA, V. T.; WAGNER, M. B. **Bioestatística**. Caxias do Sul: Educs; São Paulo: Robe Editorial, 2003.

TRIOLA, M. F. **Introdução à Estatística**. 7.ed. Rio de Janeiro: LTC,1999.