

UNIDADE



Técnicas de amostragem

Objetivo

Nesta Unidade, você vai compreender em detalhes o que é amostragem, quando deve usá-la, as suas principais técnicas, a definição do plano de amostragem, e aprenderá a utilizar uma fórmula simplificada para cálculo do tamanho mínimo de amostra.

Técnicas e definições de Amostragem

Caro estudante!

Conforme vimos na Unidade 1, a amostragem é uma das formas de coleta de dados, e observamos também que se trata de uma das subdivisões da Estatística, cujo conhecimento é indispensável para o administrador. Tenha em mente que estamos interessados em obter dados confiáveis para a tomada de decisões, e muitas vezes precisaremos realizar pesquisas para coletar tais dados. Convidamos você a conhecer um pouco mais sobre esta técnica de pesquisa e seus diferentes métodos de aplicação.

Há vários argumentos para justificar a utilização da amostragem, mas há casos em que seu uso pode não ser a melhor opção. O administrador precisa conhecer tais argumentos, para que, confrontando com os recursos disponíveis e os objetivos da pesquisa, possa tomar a melhor decisão sobre a forma de coleta dos dados.

Se o administrador decidir por amostragem, é preciso delinear o plano de amostragem, indicando como ela será implementada e qual será o seu tamanho, item crucial e que vai influenciar muito nos custos da pesquisa. Vamos ver isso em detalhes nesta Unidade.

O que é amostragem?

Amostragem é a subdivisão da Estatística que reúne os métodos necessários para coletar adequadamente amostras representativas e suficientes para que os resultados obtidos possam ser generalizados para a população de interesse. A pressuposição básica é que todas as etapas prévias do planejamento da pesquisa (veja na Unidade 1) já

GLOSSÁRIO

***Censo** – forma de coleta de dados em que a pesquisa é realizada com todos os elementos da população. Fonte: Barbetta, Reis e Bornia (2004).

***Amostragem** – forma de coleta de dados em que apenas uma pequena parte, considerada representativa, da população é pesquisada. Os resultados podem ser, então, generalizados, usualmente através de métodos estatísticos apropriados, para toda a população. Fonte: Barbetta (2006).

***População** – é o conjunto de medidas da(s) característica(s) de interesse em todos os elementos que a(s) apresenta(m). Fonte: Andrade e Ogliari (2007).

***Testes destrutivos** – são ensaios realizados para avaliar a durabilidade, resistência ou conformidade com as especificações de determinados produtos, que causam a sua inutilização, impedindo a sua comercialização. Muitos testes destrutivos são previstos em legislação específica das mais diversas áreas. Fonte: elaborado pelo autor.

foram cumpridas e que o administrador agora precisa decidir se coletará os dados por **censo*** ou por **amostragem***.

O censo consiste simplesmente em estudar todos os elementos da **população***, e a amostragem pesquisa apenas uma pequena parte dela, suposta representativa do todo. Para realizar um estudo por amostragem, de maneira que seus resultados sejam válidos e possam ser generalizados para a população, algumas técnicas precisam ser empregadas. A essência deste processo é mostrada na Figura 7:

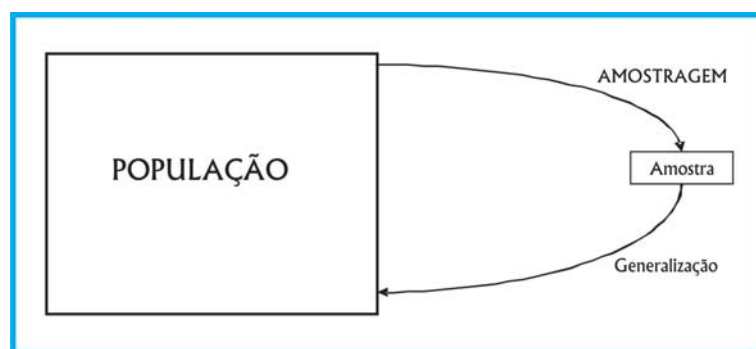


Figura 7: Processo de Amostragem e Generalização

Fonte: elaborada pelo autor

É importante saber avaliar os argumentos a favor de cada forma de coleta.

Quando devemos usar amostragem

Podemos enumerar, basicamente, três motivos para usar amostragem em uma pesquisa: economia, rapidez de processamento e quando há a necessidade de **testes destrutivos***.

- **Economia:** é muito mais barato levantar as características de uma pequena parcela da população do que de todos os seus integrantes, especialmente para grandes populações. O custo do censo demográfico do IBGE é tão colossal que somente pode ser feito cada dez anos.
- **Rapidez de processamento:** como a quantidade de dados coletada é muito menor do que a produzida em um censo, especialmente para grandes populações, o seu processamento

é mais rápido. Os resultados ficam disponíveis em pouco tempo, permitindo tomar decisões em seguida. Tal característica é especialmente importante em pesquisas de opinião eleitoral, cujo resultado precisa ser conhecido rapidamente, para que candidatos e partidos possam reavaliar suas estratégias.

- **Testes destrutivos:** se, para realizar a pesquisa, precisamos realizar testes destrutivos (de resistência, tempo de vida útil, entre outros), o censo torna-se impraticável, exigindo a utilização de amostragem. Em muitos casos, como no caso de produtos alimentícios e farmacêuticos, há normas legais que precisam ser cumpridas rigorosamente quando da realização dos ensaios.

A Figura 8 sintetiza os motivos:



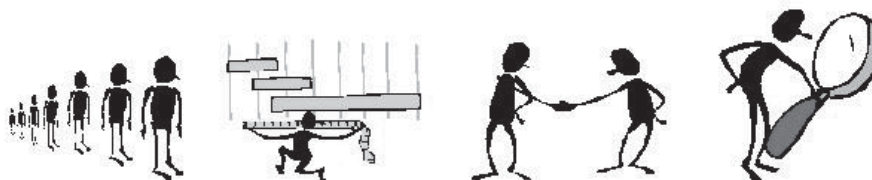
Figura 8: Os três motivos para usar amostragem em uma pesquisa
Fonte: Microsoft Office (2007)

Após reconhecer os motivos de utilizar a amostragem, pense em algumas situações em que seria recomendável utilizar esta técnica.

Quando NÃO devemos usar amostragem

Existem situações em que a utilização de amostragem pode não ser a melhor opção. Neste caso, podemos enumerar basicamente quatro motivos: população pequena, característica de fácil mensuração, necessidades políticas e necessidade de alta precisão.

- **População pequena:** quando é utilizada uma amostra probabilística (aleatória) e a população é pequena (digamos, menos de cem elementos), o tamanho mínimo de amostra para obter bons resultados será quase igual ao próprio tamanho da população (veremos isso mais adiante, ainda nesta Unidade). Vale a pena, então, realizar um censo.
- **Característica de fácil mensuração:** a característica pode não precisar de mecanismos sofisticados de mensuração, simplesmente resume-se em uma opinião direta – a favor ou contra uma proposta. Neste caso, a coleta dos dados seria bastante simples, possibilitando avaliar todos os elementos da população. Outro caso freqüente na indústria são os sistemas automatizados de medição, por exemplo, em uma fábrica de cubos de rodas de bicicletas, situada na zona franca de Manaus, os diâmetros de todos os cubos produzidos são medidos automaticamente por um sistema de telemetria a laser, dispensando a coleta por amostragem e um inspetor humano para realizar a medição.
- **Necessidades políticas:** muitas vezes, uma proposta vai afetar dramaticamente todos os elementos da população, como a adoção de um regime ou forma de governo, por exemplo, o que pode ensejar a realização de um censo, para que todos manifestem sua opinião.
- **Necessidade de alta precisão:** por que o IBGE conduz um censo cada dez anos? Porque as informações demográficas têm que ser precisas, para orientar políticas governamentais, e somente dessa maneira esse objetivo pode ser atingido. A Figura 9 sintetiza os motivos:



População pequena/Fácil mensuração/Necessidades políticas/Alta precisão

Figura 9: Situações em que a utilização de amostragem pode não ser a melhor opção

Fonte: Microsoft Office (2007)

Exercite a mente! Pense em algumas situações nas quais seja aconselhável usar um censo. Você deve se lembrar da pesquisa que esboçamos na Unidade 1: “o CRA de Santa Catarina está interessado em conhecer a opinião dos seus registrados sobre o curso em que se graduaram, desde que tal curso esteja situado em Santa Catarina”. Vimos que o número de registrados no CRA, com graduação em Santa Catarina, foi suposto igual a 9.000. Além disso, há uma listagem com os registrados, para fins de cobrança de anuidade inclusive, que contém informações sobre endereço, curso em que se graduou, entre outras. Para conhecer a opinião das pessoas, precisamos entrevistá-las (via correio, internet, telefone ou pessoalmente). Com base no que foi dito até agora, você sabe responder se a pesquisa deve ser conduzida por censo ou por amostragem? Vamos ver juntos, então!

Aspectos necessários para o sucesso da amostragem

Há três aspectos necessários para que uma pesquisa realizada por amostragem gere resultados confiáveis: representatividade, suficiência e aleatoriedade da amostra.

A **representatividade*** é o mais óbvio. A amostra precisa retratar a variabilidade existente na população: ela precisa ser uma “cópia reduzida” da população. Sendo assim, todas as subdivisões da população precisam ter representantes na amostra. A chave é avaliar se as subdivisões da população (por sexo, classe econômica, cidade, atividade profissional) podem influenciar nos resultados da pesquisa. Imagine uma pesquisa eleitoral para governador: devemos entrevistar eleitores em todas as regiões do Estado (assume-se que haja diferenças de opinião de região para região), pois, se escolhermos apenas uma delas, e ela for base política de um candidato, o resultado será distorcido.

GLOSSÁRIO

*Amostra representativa – é aquela que representa na sua composição todas as subdivisões da população, procurando retratar da melhor maneira possível a sua variabilidade. Fonte: elaborado pelo autor.

Vamos aprender ainda nesta Unidade uma fórmula simplificada para o cálculo do tamanho de amostra, e na Unidade 9 veremos uma expressão mais completa. Em ambos os casos, porém, veremos que o tamanho de amostra também dependerá da precisão que queremos para o nosso resultado.

GLOSSÁRIO

***Amostra suficiente** – é aquela que tem um tamanho tal que permite representar adequadamente a variabilidade da população. Fonte: elaborado pelo autor.

***Amostra aleatória, casual ou probabilística** – é a amostra retirada por meio de um sorteio não viciado, que garante que cada elemento da população terá uma probabilidade maior do que zero de pertencer à amostra. Fonte: Barbetta, Reis e Bornia (2004).

A **suficiência*** também é um aspecto relativamente óbvio. É necessário que a amostra tenha um tamanho suficiente para representar a variabilidade existente na população. Quanto mais homogênea for a população (menor variabilidade), menor poderá ser o tamanho da amostra, e quanto mais heterogênea (maior variabilidade), maior terá que ser o tamanho da amostra para representá-la.

A **aleatoriedade*** da amostra é o aspecto menos intuitivo, mas extremamente importante. Significa que os elementos da amostra serão selecionados da população por meio de sorteio não viciado: todos os elementos da população têm chance de pertencer à amostra. É necessária uma listagem com os elementos da população, permitindo a atribuição de números a cada um deles, e faz-se o sorteio. Idealmente, nós escreveríamos os números dos elementos da população em pequenos papéis, depositaríamos em uma urna, misturaríamos os papéis, e, de olhos vendados, escolheríamos os números, selecionando a amostra. Para grandes populações, esse procedimento é inviável, e com a disponibilidade de recursos computacionais, contraproducente.

O sorteio pode ser realizado através de **tabelas de números aleatórios** ou **algoritmos de geração de números pseudo-aleatórios***. (Algoritmos de geração de números pseudo-aleatórios são programas computacionais que geram números aleatórios (pseudo-aleatórios, pois têm uma regra de formação), procurando simular os sorteios manuais de números de 0 a 9 e garantir que todo número com a mesma quantidade de algarismos tenha a mesma probabilidade de ocorrência.)

As tabelas de números aleatórios são instrumentos usados para auxiliar na seleção de amostras aleatórias. São formadas por sucessivos sorteios de algarismos do conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, fazendo com que todo número com a mesma quantidade de algarismos tenha a mesma probabilidade de ocorrência. Quando o sorteio é realizado “manualmente”, a tabela é realmente chamada de tabela de números aleatórios. (Muitos estatísticos realizaram tais sorteios, registraram os resultados e os publicaram em livros e periódicos para uso geral). Se, porém, os números são obtidos mediante simulação computacional, passamos a ter uma tabela de números pseudo-aleató-

rios, pois os números são provenientes da execução de um **algoritmo** matemático, que tem uma lógica e uma **lei de formação** dos resultados. Não obstante, tal problema, caso o algoritmo seja bom, somente ocorre após milhões ou bilhões de sorteios, quantidade muitíssimo superior àquela usada nas nossas pesquisas. Alguns estatísticos construíram tabelas de números pseudo-aleatórios e as deixaram disponíveis para o público em geral.

Nos dias de hoje, com todas as facilidades da informática, são cada vez mais comuns bases de dados armazenadas em meio digital, desde uma simples planilha do Microsoft Excel até grandes bancos de dados.

Então, pergunta-se: por que não realizar também o processo de amostragem, em meio digital, com os algoritmos citados no parágrafo anterior: os algoritmos de geração de números pseudo-aleatórios?

Trata-se de programas computacionais que procuram simular os sorteios reais de números. As grandes vantagens do seu uso são a possibilidade de adaptar facilmente o sorteio ao tamanho da população envolvida e, obviamente, a velocidade de processamento. Veja um exemplo de **números aleatórios** de quatro dígitos (de 0001 a 9000) gerados pelo Microsoft Excel®:

3439	907	5369	8092	7962	8626	131	3667	7769	1248
2206	410	292	1478	1977	155	2566	3088	4983	3217
3347	3201	8193	4195	3836	2736	8781	7260	8921	2307

No caso da nossa pesquisa para o CRA de Santa Catarina, em que temos 9.000 registrados graduados em Santa Catarina e há uma listagem da população, pense: como seria o sorteio?

No caso mais simples de amostragem aleatória, o registrado de número 3.439 seria sorteado, seguido pelo 907 e pelo 5.369, e assim por diante, até completar o tamanho de amostra. Usualmente, cria-se automaticamente uma nova base de dados com os elementos sorteados.

Neste caso, há sempre o risco de os números se repetirem se a série for muito longa, descaracterizando a aleatoriedade.

Na seção “Para saber mais”, vamos disponibilizar um link que explica como gerar números pseudo-aleatórios com este aplicativo.

Veremos sobre a teoria da inferência estatística nas Unidades 8, 9 e 10.

Toda a teoria de inferência estatística pressupõe que a amostra, a partir da qual será feita a generalização (veja a Figura 1 desta Unidade), foi retirada de forma aleatória.

Agora que já conhecemos os aspectos principais para o sucesso da amostragem, podemos detalhar o plano de amostragem.

Plano de amostragem

Uma vez tendo decidido realizar a pesquisa selecionando uma amostra da população, é preciso elaborar o **plano de amostragem**, que consiste em definir as unidades amostrais, o modo como a amostra será retirada (o tipo de amostragem) e o próprio tamanho da amostra.

As **unidades amostrais** são as unidades selecionadas para chegar aos elementos da própria população. Podem ser os próprios elementos da população, quando há acesso direto a eles, ou qualquer outra unidade que possibilite chegar até eles: selecionar os domicílios como unidades de amostragem, para chegar até as famílias (que são os elementos da população); selecionar as turmas como unidades de amostragem, para chegar até os alunos (que são os elementos da população). No caso da pesquisa do CRA de Santa Catarina, as unidades amostrais são os próprios elementos da população, uma vez que temos a sua listagem. No caso da Pesquisa Nacional por Amostragem de Domicílios do IBGE, as unidades amostrais são os domicílios, através dos quais se chega às famílias.

O modo como a amostra será retirada é outra decisão importante, que precisa constar do plano de amostragem. Na Figura 10 a seguir, vemos o resumo dos diversos tipos de amostragem:

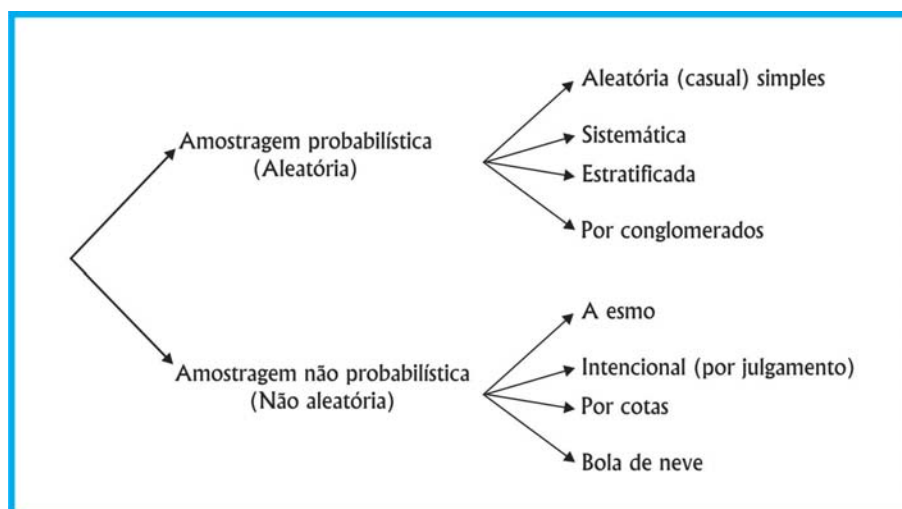


Figura 10: Tipos de Amostragem

Fonte: elaborada pelo autor

Amostragem probabilística (aleatória)

Amostragem probabilística, aleatória ou casual é aquela que garante que cada elemento da população tenha probabilidade de pertencer à amostra. Para que isso ocorra, é necessário que a amostra seja selecionada por sorteio não viciado, ou seja, exige-se aleatoriedade. Sua importância decorre do fato de que apenas os resultados provenientes de uma amostra probabilística podem ser generalizados estatisticamente para a população da pesquisa.

Você deve estar se perguntando: “mas, afinal, o que significa ‘estatisticamente’?”. Significa que podemos associar aos resultados uma probabilidade de que estejam corretos, ou seja, uma medida da confiabilidade das conclusões obtidas. Se a amostra não for probabilística, não há como saber se há 95% ou 0% de probabilidade de que os resultados sejam corretos, e as técnicas de inferência estatística porventura utilizadas terão validade questionável.

A condição primordial para uso da amostragem probabilística é que todos os elementos da população tenham uma probabilidade

maior do que zero de pertencer à amostra. Tal condição é materializada se:

GLOSSÁRIO

* **Amostragem aleatória simples** – é o processo de amostragem em que todos os elementos da população têm a mesma probabilidade de pertencer à amostra, e cada elemento é sorteado. Fonte: Barbetta, Reis e Bornia (2004).

* **Tabelas de números aleatórios** – são instrumentos usados para auxiliar na seleção de amostras aleatórias, formadas por sucessivos sorteios de algarismos do conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, fazendo com que todo número com a mesma quantidade de algarismos tenha a mesma probabilidade de ocorrência. Fonte: Barbetta (2006).

- há acesso a toda a população. Ou seja, não há teoricamente problema em selecionar nenhum dos elementos, todos poderiam ser pesquisados. Concretamente, há uma lista da população, como no caso da pesquisa do CRA, que dispõe de uma relação com os 9.000 registrados que se graduaram em Santa Catarina; e
- os elementos da amostra são selecionados através de alguma forma de sorteio não viciado: tabelas de números aleatórios, números pseudo-aleatórios gerados por computador. Com a utilização de sorteio, elimina-se a ingerência do pesquisador na obtenção da amostra e garante-se que todos os integrantes da população têm probabilidade de pertencer à amostra.

Agora, vamos lhe apresentar os tipos de amostragem probabilística.

Amostragem aleatória (casual) simples

A **amostragem aleatória simples*** é o tipo de amostragem probabilística recomendável, somente, se a população for homogênea em relação aos objetivos da pesquisa, por exemplo, quando se admite que todos os elementos da população têm características semelhantes em relação aos objetivos da pesquisa. Há uma listagem dos elementos da população, atribuem-se números a eles, e através de alguma espécie de sorteio não viciado, por meio de **tabelas de números aleatórios*** ou números pseudo-aleatórios gerados por computador, os integrantes da amostra são selecionados. Neste tipo de amostragem probabilística, todos os elementos da população têm a mesma probabilidade de pertencer à amostra. Foi exatamente o que fizemos no final do tema “Aspectos necessários para o sucesso da amostragem” para a nossa pesquisa do CRA.

Amostragem sistemática

Quando a lista de respondentes for muito grande, a utilização de amostragem aleatória simples pode ser um processo moroso, ou se o tamanho de amostra for substancial, teremos que realizar um grande número de sorteios: caso estejamos utilizando números pseudo-aleatórios, aumenta o risco de repetição dos números. Utiliza-se, então, uma variação, a **amostragem sistemática***, que também supõe que a população é homogênea em relação à variável de interesse, mas que consiste em retirar elementos da população a intervalos regulares, até compor o total da amostra. A amostragem sistemática somente pode ser retirada se a ordenação da lista não tiver relação com a variável de interesse. Imagine que queremos obter uma amostra de idades de uma listagem justamente ordenada desta forma, neste caso a amostragem sistemática não seria apropriada, a não ser que reordenássemos a lista.

Veja a seguir o procedimento para a amostragem sistemática:

- obtém-se o tamanho da população (N);
- calcula-se o tamanho da amostra (n) – veremos isso mais adiante;
- encontra-se o intervalo de retirada $k = N/n$:
 - se k for fracionário, deve-se aumentar n até tornar o resultado inteiro; e
 - se N for um número primo, excluem-se *por sorteio* alguns elementos da população para tornar k inteiro;
- sorteia-se o ponto de partida (um dos k números do primeiro intervalo), usando uma tabela de números aleatórios ou qualquer outro dispositivo (isso precisa ser feito para garantir que todos os elementos da população tenham chance de pertencer à amostra); e
- a cada k elementos da população, retira-se um para fazer parte da amostra, até completar o valor de n.

O resumo deste processo é retratado na Figura 11, veja:

GLOSSÁRIO

***Amostragem sistemática** – é a variação da amostragem aleatória simples em que os elementos da população são retirados a intervalos regulares, até compor o total da amostra, sendo o sorteio realizado apenas no ponto de partida. Fonte: Barbetta (2006).

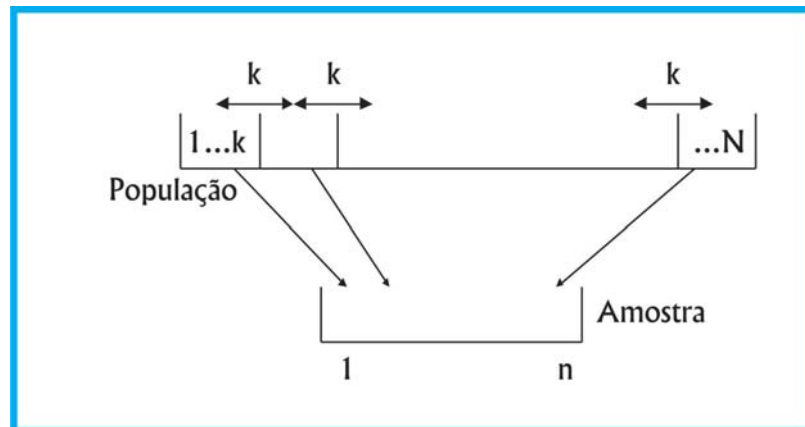


Figura 11: Processo de amostragem sistemática

Fonte: elaborada pelo autor

O exemplo a seguir ajudará você a entender melhor sobre o processo de amostragem sistemática. Leia com atenção!

Exemplo 1: uma operadora telefônica pretende saber a opinião de seus assinantes comerciais sobre seus serviços na cidade de Florianópolis. Supondo que há 25.037 assinantes comerciais, e a amostra precisa ter no mínimo 800 elementos, mostre como seria organizada uma amostragem sistemática para selecionar os respondentes.

A operadora dispõe de uma lista ordenada alfabeticamente com todos os seus assinantes. O intervalo de retirada será:

$$k = N/n = 25.037/800 = 31,2965.$$

Como o valor de k é fracionário, algo precisa ser feito. Aumentar o tamanho da amostra não resolverá o problema, porque 25.037 é um número primo. Como não podemos reduzir o tamanho de amostra, devendo permanecer igual a 800, se excluirmos por sorteio 237 elementos da população e refizermos a lista, teremos:

$$k = N/n = 24.800/800 = 31.$$

A cada 31 assinantes, um é retirado para fazer parte da amostra. Devemos sortear o ponto de partida: um número de 1 a 31 (do 1º ao 31º assinante). Imagine que o sorteio resultasse em 5, então a amostra seria (número de assinantes): {5, 36, 67, 98, ..., 24.774}.

Amostragem estratificada

É bastante comum que a população de uma pesquisa seja heterogênea em relação aos objetivos da pesquisa. No caso de uma pesquisa eleitoral para governador, por exemplo, podemos esperar que a opinião deva ser diferente dependendo da região onde o eleitor mora, classe social e mesmo profissão dos entrevistados. Contudo, podemos supor que haja certa homogeneidade de opinião dentro de cada grupo. Então, supõe-se que haja heterogeneidade entre os estratos, mas homogeneidade dentro dos estratos, e que eles sejam mutuamente exclusivos (cada elemento da população pode pertencer a apenas um estrato). Para garantir que a amostra seja **representativa*** da população, precisamos garantir que os diferentes estratos sejam nela representados: deve-se usar a **amostragem estratificada***, como representada a Figura 12:

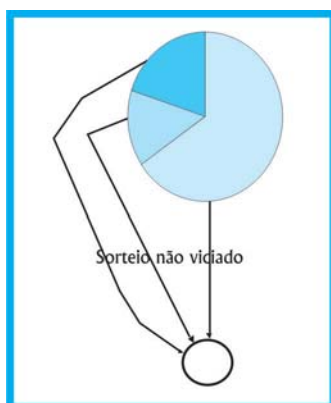


Figura 12: Amostragem estratificada
Fonte: elaborada pelo autor

Veja que a seleção dos elementos de cada estrato pode ser feita usando amostragem aleatória simples ou sistemática.

A amostragem estratificada pode ser:

- proporcional, quando o número de elementos selecionados de cada estrato é proporcional ao seu tamanho na população (por exemplo, se o estrato representa 15% da população, 15% da amostra deverá ser retirada dele); e

GLOSSÁRIO

***Amostra representativa** – aquela que representa na sua composição todas as subdivisões da população, procurando retratar da melhor maneira possível a sua variabilidade. Fonte: elaborado pelo autor.

*** Amostragem estratificada** – é a amostragem probabilística usada quando a população for heterogênea em relação aos objetivos da pesquisa (as opiniões tendem a variar muito de subgrupo para subgrupo), e a amostra precisa conter elementos de cada subgrupo da população para representá-la adequadamente. Fonte: Barbetta (2006).

- uniforme, quando os mesmos números de elementos são selecionados de cada estrato.

A amostragem estratificada proporcional possibilita resultados melhores, mas exige um grande conhecimento da população (para saber quantos são e quais são os tamanhos dos estratos). A amostragem estratificada uniforme é mais utilizada em estudos comparativos.

No caso da pesquisa do CRA, você acredita que a população é heterogênea em relação aos objetivos da pesquisa? Será que a região do Estado, o fato de ter estudado em faculdade pública ou particular pode influenciar as opiniões dos registrados sobre os cursos nos quais se graduaram?

Amostragem por conglomerados

Teoricamente, a amostragem estratificada proporcional apresenta os melhores resultados possíveis. Sua grande dificuldade de uso deve-se ao grau de conhecimento necessário sobre a população, que geralmente não existe ou é impraticável de obter. Uma alternativa consiste no uso de **conglomerados***.

Os conglomerados também são grupos mutuamente exclusivos de elementos da população, mas são definidos de forma mais arbitrária do que os estratos: é bastante comum definir os conglomerados geograficamente. Por exemplo, os bairros de uma cidade, que constituiriam conglomerados de domicílios.

O procedimento para a amostragem por conglomerados ocorre da seguinte forma:

- divide-se a população em conglomerados;
- sorteiam-se os conglomerados (usando tabela de números aleatórios ou qualquer outro método não viciado);
- pesquisam-se todos os elementos dos conglomerados sorteados ou sorteiam-se elementos deles.

GLOSSÁRIO

***Amostragem por conglomerados** – é a amostragem probabilística em que a população é subdividida em grupos definidos por conveniência (usualmente geográfica), e alguns destes grupos são selecionados por sorteio, e elementos dos grupos sorteados podem também ser sorteados para compor a amostra. Fonte: Barbetta (2006).

A utilização de amostragem por conglomerados permite uma redução substancial nos custos de obtenção da amostra, sem comprometer demasiadamente a precisão, e em alguns casos é a única alternativa possível. Veja a Figura 13 e entenda como ocorre essa amostragem:

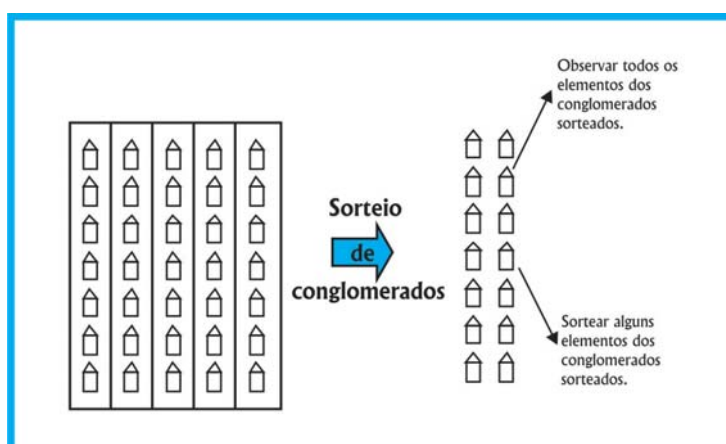


Figura 13: Amostragem por conglomerados

Fonte: elaborada pelo autor

A Pesquisa Nacional por Amostra de Domicílios (PNAD) do IBGE coleta informações demográficas e socioeconômicas sobre a população brasileira. Utiliza amostragem por conglomerados em três estágios:

- **primeiro estágio:** amostras de municípios (conglomerados) para cada uma das regiões geográficas do Brasil;
- **segundo estágio:** setores censitários sorteados em cada município (conglomerado sorteado); e
- **terceiro estágio:** domicílios sorteados em cada setor censitário.

Você deve estar se perguntando: “e quando não for possível garantir a probabilidade de todo elemento da população pertencer à amostra?” Então, este é o momento de partirmos para a amostragem não probabilística.

Mais informações em

[http://
www.ibge.gov.br/
home/estatistica/
populacao/
trabalhoerendimento/
pnad98/saude/
metodologia.shtm](http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad98/saude/metodologia.shtm)

Amostragem não probabilística

GLOSSÁRIO

***Amostragem não probabilística** – é o processo de amostragem em que nem todos os elementos da população têm chance de pertencer à amostra, pois a seleção não é feita por sorteio não viciado.

Fonte: Barbetta (2006).

***Erro amostral** – é o valor máximo que o pesquisador admite errar na estimativa de uma característica da população a partir de uma amostra aleatória desta mesma população.

Fonte: Barbetta (2006).

A obtenção de uma amostra probabilística exige uma listagem com os elementos da população. Em suma, exige acesso a todos os elementos da população. Nem sempre é possível obter tal listagem na prática, o que teoricamente inviabilizaria a retirada de uma amostra probabilística. Então, pode-se recorrer à **amostragem não probabilística***.

Ao usar a amostragem não probabilística, o pesquisador não sabe qual é a probabilidade de que um elemento da população tem de pertencer à amostra. Portanto, os resultados da amostra não podem ser estatisticamente generalizados para a população, porque não se pode estimar o **erro amostral***.

Alguns dos usos habituais da amostragem não probabilística são os seguintes:

- a etapa preliminar em projetos de pesquisa;
- em projetos de pesquisa qualitativa; e
- em casos nos quais a população de trabalho não pode ser enumerada.

Veja que existem ainda vários tipos de amostragem não probabilística e que serão descritos na sequência.

Amostragem a esmo

Na amostragem a esmo, o pesquisador procura ser o mais aleatório possível, mas sem fazer um sorteio formal. Imagine um lote de 10.000 parafusos, do qual queremos tirar uma amostra de cem. Se fôssemos realizar uma amostragem aleatória simples, o processo talvez fosse trabalhoso demais. Então, simplesmente retiramos os elementos a esmo. Este tipo de amostragem também pode ser utilizado quando a população for formada por material contínuo (gases, líquidos, minérios), bastando homogeneizar o material e retirar a amostra.

Amostragem por julgamento (intencional)

Na amostragem por julgamento, o pesquisador deliberadamente escolhe alguns elementos para fazer parte da amostra, com base no seu julgamento de que aqueles seriam representativos da população. Este tipo de amostragem é bastante usado em estudos qualitativos. Obviamente, o risco de obter uma amostra viciada é grande, pois se baseia totalmente nas preferências do pesquisador, que pode se enganar (involuntária ou “voluntariamente”).

Amostragem por cotas

A amostragem por cotas parece semelhante a uma amostragem estratificada proporcional, da qual se diferencia por não empregar sorteio na seleção dos elementos. A população é dividida em vários subgrupos. Na realidade, é comum dividir em um grande número para compensar a falta de aleatoriedade, e seleciona-se uma cota de cada subgrupo, proporcional ao seu tamanho.

Em uma pesquisa de opinião eleitoral, por exemplo, poderíamos dividir a população de eleitores por sexo, nível de instrução, faixas de renda, entre outros aspectos, e obter cotas proporcionais ao tamanho dos grupos (que poderia ser obtido através das informações do IBGE). Na amostragem por cotas, os elementos da amostra são escolhidos pelos entrevistadores (de acordo com os critérios), geralmente em pontos de grande movimento, o que sempre acarreta certa subjetividade (e impede que qualquer um que não esteja passando pelo local no exato momento da pesquisa possa ser selecionado).

Na prática, muitas pesquisas são realizadas utilizando amostragem por cotas, incluindo as **polêmicas pesquisas eleitorais**.

No exemplo apresentado no Quadro 1, imagine que queremos saber a opinião dos eleitores do bairro Goiaba sobre o governo municipal. Supõe-se que as principais variáveis que condicionariam as respostas seriam sexo, idade e classe social. O bairro apresenta a seguinte composição demográfica para as variáveis:

Leia um texto muito interessante sobre o tema, que se encontra disponível em: <http://www.ime.unicamp.br/~dias/falaciaPesquisaEleitoral.pdf>

Sexo	Idade (faixa etária)	Classe social	% populacional
Masculino	18 -- 35	A	1%
Masculino	18 -- 35	B	4%
Masculino	18 -- 35	C	10%
Feminino	18 -- 35	A	1%
Feminino	18 -- 35	B	2%
Feminino	18 -- 35	C	9%
Masculino	35 -- 60	A	5%
Masculino	35 -- 60	B	8%
Masculino	35 -- 60	C	12%
Feminino	35 -- 60	A	4%
Feminino	35 -- 60	B	8%
Feminino	35 -- 60	C	10%
Masculino	Mais de 60	A	1%
Masculino	Mais de 60	B	9%
Masculino	Mais de 60	C	3%
Feminino	Mais de 60	A	3%
Feminino	Mais de 60	B	7%
Feminino	Mais de 60	C	3%

Quadro 1: Esquema de amostragem por cotas

Fonte: adaptado pelo autor de Lakatos e Marconi (2003)

Se, por exemplo, o tamanho de nossa amostra fosse igual a 200 (200 pessoas serão entrevistadas), o número de pessoas deveria ser dividido de forma proporcional: 1% do sexo masculino, com idade entre 18 e 25 anos, da classe A, totalizando duas pessoas; 4% do sexo masculino, com idade entre 18 e 25 anos, da classe B, totalizando oito pessoas, e assim por diante. Os entrevistadores receberiam suas cotas e deveriam escolher pessoas, em pontos de movimento do referido bairro, que se aproximassem dos critérios e entrevistá-las, recolhendo suas opiniões sobre o governo municipal. Usualmente, os resultados são generalizados estatisticamente para a população, empregando as técnicas que serão vistas na Unidade 9 deste livro-texto, mas rigorosa-

mente os resultados da amostragem por cotas **não têm validade estatística**, visto que não contemplam o princípio de aleatoriedade na seleção da amostra.

Amostragem “bola de neve”

A amostragem “bola de neve” é particularmente importante quando é difícil identificar respondentes em potencial. A cada novo respondente que é identificado e entrevistado, pede-se que identifique outros que possam ser qualificados como respondentes. Pode levar a amostras compostas apenas por “amigos” dos primeiros entrevistados, o que pode causar distorções nos resultados finais.

Agora que você já conhece sobre o importante e interessante tema do cálculo do tamanho de amostra, passaremos para uma amostra probabilística.

Cálculo do tamanho de uma amostra probabilística (aleatória)

A determinação do tamanho de amostra é um dos aspectos mais controversos da técnica de amostragem e envolve uma série de conceitos (probabilidade, inferência estatística e a própria teoria da amostragem). Nesta seção, apresentaremos uma visão simplificada para obter o tamanho mínimo de uma amostra aleatória simples que atenda aos seguintes requisitos:

- o interesse na proporção de ocorrência de um dos valores de uma variável qualitativa na população;
- a confiabilidade dos resultados da amostra deve ser aproximadamente igual a 95% (ou seja, há 95% de probabilidade de que a proporção populacional do valor da variável qualitativa esteja no intervalo definido pelos resultados da amostra);

- estamos fazendo uma estimativa exagerada do tamanho de amostra;
- não vamos nos preocupar com aspectos financeiros relacionados ao tamanho da amostra (embora, obviamente, seja uma consideração importante).

O primeiro passo para calcular o tamanho da amostra é definir o **erro amostral** tolerável, que será chamado de E_0 . Este erro é o valor máximo que o pesquisador admite ter na estimativa de uma característica da população.

Lembre-se das pesquisas de opinião eleitoral: “o candidato Fulano está com 18% de intenção de voto, a margem de erro da pesquisa é de 2% para mais ou para menos”. O 2% é o valor do erro amostral tolerável; então, o percentual de pessoas declarando o voto no candidato Fulano é igual a $18\% \pm 2\%$. Além disso, há uma probabilidade de que este intervalo não contenha o valor real do parâmetro, ou seja, o percentual de eleitores que declaram o voto no candidato, pelo fato de que estamos usando uma amostra, embora isso raramente seja dito na mídia, especialmente na televisão.

É razoável imaginar que, quanto menor o erro amostral tolerável escolhido, maior será o tamanho da amostra necessário para obtê-lo. Isso fica mais claro ao ver a fórmula para obtenção da primeira estimativa do tamanho de amostra:

$$n_0 = \frac{1}{E_0^2}$$

Onde E_0 é o erro amostral tolerável, e n_0 é a primeira estimativa do tamanho de amostra. Se o tamanho da população, N , for conhecido, podemos corrigir a primeira estimativa:

$$n = \frac{N \times n_0}{N + n_0}$$

Pense, com esse exemplo, em como obter o tamanho mínimo de uma amostra aleatória simples, admitindo com alto grau de confiança um erro amostral máximo de 2%, supondo que a população tenha:

- a) 200 elementos; e
- b) 200.000 elementos.

Observe a diferença entre os tamanhos das duas populações: a da letra b é mil vezes maior do que a da letra a. Como a primeira estimativa, n_0 não depende do tamanho da população, e o erro amostral é 2% para ambas, podemos calculá-lo apenas uma vez. Devemos dividir o 2% por 100 antes de substituir na fórmula:

$$n_0 = \frac{1}{E_0^2} = \frac{1}{(0,02)^2} = 2.500$$

Então, nossa primeira estimativa, para um erro amostral de 2%, é retirar uma amostra de 2.500 elementos.

- Obviamente, precisamos corrigir a primeira estimativa, pois a população conta com apenas 200 elementos. Então:

$$n = \frac{N \times n_0}{N + n_0} = \frac{200 \times 2.500}{200 + 2.500} = 185,185$$

Precisamos arredondar, sempre para cima, o tamanho mínimo da amostra. Então, a amostra deverá ter pelo menos 186 elementos para garantir um erro amostral de 2%. Observe que a amostra representa 93% da população. Será que um censo não seria mais aconselhável neste caso?

- Corrigindo a primeira estimativa com o tamanho da população:

$$n = \frac{N \times n_0}{N + n_0} = \frac{200.000 \times 2.500}{200.000 + 2.500} = 2.469,136$$

Arredondando, a amostra deverá ter no mínimo 2.470 elementos para garantir um erro amostral de 2%. Observe que a amostra representa 1,235 % da população. Claríssimo caso em que a amostragem é a melhor opção de coleta.

Poderíamos ter usado diretamente a primeira estimativa, 2.500 elementos, pois a correção não causou grande mudança. Este exemplo prova que não precisamos de grandes amostras para obter uma boa precisão nos resultados.

A Figura 14 mostra um gráfico relacionando tamanhos de amostra para diferentes tamanhos de população, considerando um erro amostral tolerável igual a 2%.

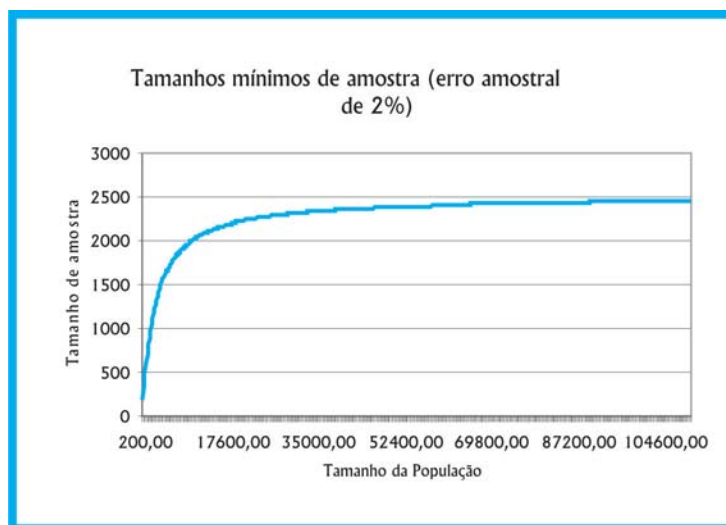


Figura 14: Tamanho de amostra x tamanho da população ($e_0 = 2\%$)
Fonte: elaborada pelo autor a partir de Microsoft

Observe na Figura 14 que, a partir de um determinado tamanho de população, para o mesmo erro amostral, o ritmo de crescimento do tamanho da amostra vai diminuindo; para 70.000 elementos ou mais, praticamente não há mais aumento. Isso mostra que não há necessidade de retirar, por exemplo, 50% da população para ter uma boa amostra.

É importante alertar que, ao calcular o tamanho de amostra para amostragem estratificada, deve-se fazê-lo para cada estrato, e o tamanho total será a soma dos valores. Se isso não for feito, não podemos garantir o erro amostral dentro de cada estrato: se calcularmos um valor geral e dividirmos o tamanho da amostra por estrato (mesmo proporcionalmente), a margem de erro dentro de cada estrato será maior do que a prevista.

Saiba mais...

- Sobre amostragem, consulte BARBETTA, P. A. *Estatística Aplicada às Ciências Sociais*. 6. ed. Florianópolis: Ed. da UFSC, 2006, capítulo 3.
- Sobre características de fácil mensuração, consulte em LAGO NETO, J.C. *O efeito da autocorrelação em gráficos de controle para variável contínua: um estudo de caso*. 1999. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC. Florianópolis.
- Sobre pesquisas eleitorais, consulte SOUZA, J. *Pesquisas eleitorais: críticas e técnicas*. Brasília: Centro Gráfico do Senado Federal, 1990.
- Sobre como gerar números pseudo-aleatórios ou obter amostras aleatórias simples no Microsoft Excel, leia o texto “*Como gerar uma amostra aleatória simples com o Microsoft Excel®*”, no Ambiente Virtual de Ensino-Aprendizagem.
- Sobre Amostragem a esmo, leia COSTA NETO, P. L. da O. *Estatística*. 2. ed. São Paulo: Edgard Blücher, 2002.

RESUMO

O resumo desta Unidade está esquematizado na Figura 15. Veja:

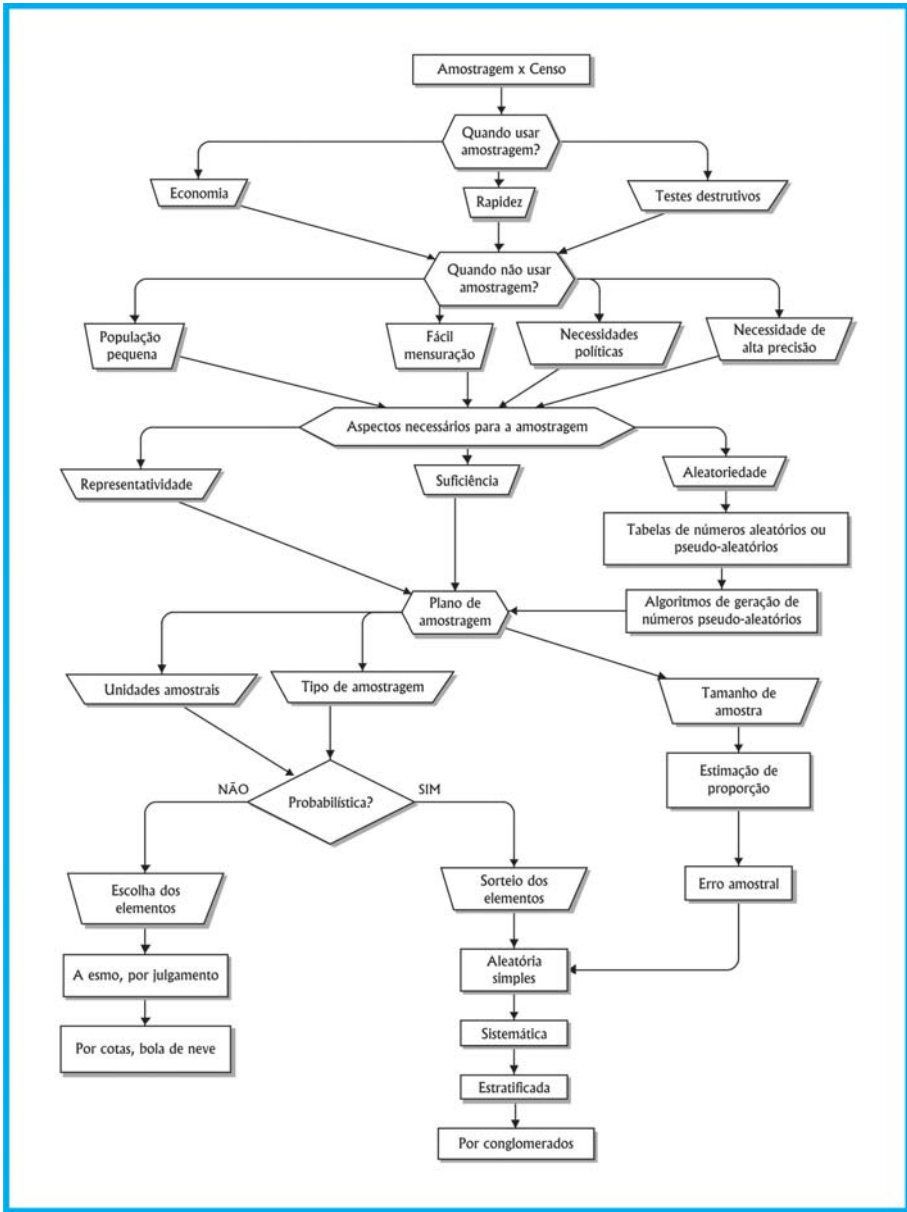


Figura 15: Resumo da Unidade 2
Fonte: elaborada pelo autor

Atividades de aprendizagem

As atividades de aprendizagem estão disponíveis no Ambiente Virtual de Ensino-Aprendizagem. Não deixe de respondê-las.

Caro estudante!

Chegamos ao final da Unidade 2. Nela estudamos amostragem e censo, e suas formas de utilização, habilidades necessárias para um bom administrador. Esta Unidade foi repleta de figuras, quadros, representações, e exemplos de utilização das técnicas e das diferentes formas de utilização, na íntegra de suas especificidades, e deu sustentação para as discussões das próximas unidades. Releia, caso necessário, todos os exemplos, leia as indicações do Saiba mais e discuta com seus colegas. Na realização das atividades de aprendizagem, você colocará em prática os ensinamentos repassados. Conte sempre com o acompanhamento da tutoria e das explicações do professor. Lembre-se que não está sozinho. Conte com a gente!