

3. AMOSTRAGEM

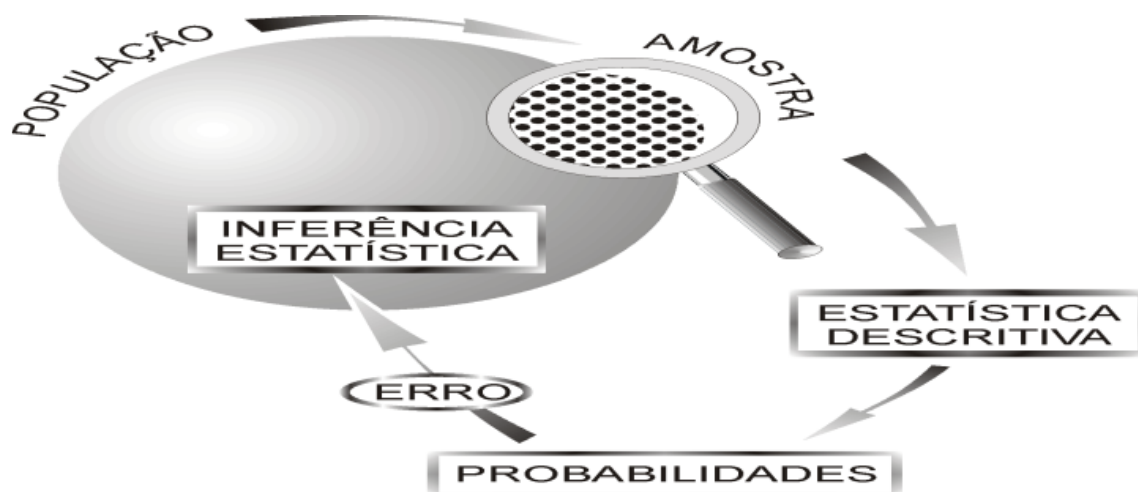
3.1. Considerações iniciais

É a parte da Teoria Estatística que define os procedimentos para os planejamentos amostrais e as técnicas de estimação utilizadas.

As técnicas de amostragem, tal como o planejamento amostral, são amplamente utilizados nas pesquisas científicas e de opinião para se conhecer alguma característica da população.

Nos **planejamentos amostrais**, a coleta dos dados deve ser realizada observando-se uma metodologia adequada para que os resultados possam ser extrapolados para a população como um todo.

Esse processo de **extensão dos resultados para a população** é o que, na estatística, chamamos de **INFERÊNCIA**.



A amostragem (processo de retirada de amostras de uma população) é uma das etapas fundamentais na tomada de decisões nos diversos níveis gerenciais, pois uma amostragem mal executada, com certeza, resultará em estatísticas pouco confiáveis e em uma tomada de decisão possivelmente imprecisa.

3.2. Conceitos e definições

Primeiramente, torna-se necessário definir população e amostra.

População é o conjunto de elementos que apresentam uma ou mais características em comum.

Uma parte desta população, que apresente as mesmas características, é denominada de **amostra**, ou seja, uma porção ou fração da população que preserva todas as características importantes dos elementos que a integram.

Parâmetro é um valor desconhecido associado a uma característica da população e que se deseja conhecer.

Estimador é uma função que permite estimar o valor de um parâmetro (estimativa) a partir de uma amostra da população.

Amostragem é o estudo das relações existentes entre a amostra, a população de onde ela foi extraída e a forma como ocorre esta extração. É útil na avaliação de grandezas desconhecidas da população (**parâmetros**) a partir do conhecimento de grandezas correspondentes das amostras (**estimativas** ou **estatísticas**). Também auxilia na verificação de diferenças observadas entre duas ou mais amostras (**tratamentos**), para que se saiba se estas diferenças são devidas a uma variação casual ou se são verdadeiramente relacionadas aos efeitos de tratamentos.

Representatividade corresponde à possibilidade de manter as mesmas características presentes na população.

Quando se obtém informações a partir das amostras para conhecer parâmetros de populações, está se realizando uma **inferência**.

As principais vantagens da utilização do estudo **por amostras representativas** (aquelas que mantêm as características da população de onde a amostra foi retirada) em relação ao **censo (avaliação de toda a população)** são:

- ocorre uma redução no custo, pois sendo os dados obtidos apenas de uma fração da população, os gastos são menores do que as oriundas de um censo;
- na prática das organizações, é necessário que os resultados sejam obtidos com a maior rapidez possível;
- outra vantagem corresponde a uma maior amplitude e flexibilidade;
- a última vantagem é a maior exatidão dos resultados.

Para que as conclusões da teoria de amostragem sejam válidas, as amostras devem ser escolhidas de modo a serem representativas da população. Isso significa que a amostra deve possuir as mesmas características básicas da população, no que diz respeito à(s) variável(eis) que desejamos estudar. Desta forma, o plano de amostragem deve ser formulado para garantir esta representatividade.

Unidades amostrais correspondem às unidades selecionadas na amostragem para calcular as estatísticas. Uma vez tendo decidido realizar a pesquisa selecionando uma amostra da população, é preciso elaborar o plano de amostragem. O plano de amostragem consiste em definir as **unidades amostrais**, maneira pela qual a amostra será retirada (o tipo de amostragem), e o próprio tamanho da amostra.

Estas unidades amostrais podem corresponder aos próprios elementos da população, quando há acesso direto a eles, ou qualquer outra unidade que possibilite chegar até eles.

Tem-se dois tipos de amostragem:

- **Amostragem probabilística:** quando todos os elementos da população tem uma probabilidade conhecida e diferente de zero de pertencer à amostra. A realização deste tipo de amostragem só é possível se a população for finita e totalmente acessível.
- **Amostragem não probabilística:** quando não se conhece a probabilidade de um elemento da população pertencer à amostra.

A utilização de uma amostra probabilística é melhor para garantir a representatividade da amostra, pois o acaso será o único responsável por eventuais discrepâncias entre população e amostra. Estas discrepâncias são levadas em consideração nas inferências estatísticas.

3.3. Esquemas amostrais

Os principais esquemas amostrais (Aleatória Simples, Estratificada, por Conglomerados, Sistemática, a Esmo, Intencional e por Cotas) são apresentados a seguir.

3.3.1. Amostragem aleatória simples (probabilística)

A amostragem aleatória simples (AAS) deve ser utilizada somente quando a população for homogênea em relação à variável que se deseja estudar. Geralmente, atribui-se uma numeração a cada indivíduo da população, e através de um **sorteio aleatório**, os elementos que vão compor a amostra são selecionados. Todos os elementos da população têm a mesma probabilidade de pertencer à amostra.

3.3.2. Amostragem aleatória estratificada (probabilística)

Quando a variável de interesse apresenta uma heterogeneidade na população e esta heterogeneidade permite a identificação de grupos homogêneos, é possível dividir a população em grupos (estratos) e fazer uma amostragem dentro de cada estrato, garantindo, assim, a representatividade de cada estrato na amostra.

Como se está dividindo a população em estratos (grupos) que são homogêneos dentro de si, é possível, então, caracterizar a amostragem estratificada. Para determinar a amostragem estratificada de forma proporcional, é preciso definir a **proporção do estrato em relação à população**.

A proporção do estrato **h** será igual ao número de elementos presentes neste estrato (**N_h**) dividido pelo tamanho da população (**N**), ou seja, (**N_h/N**).

Após determinar esta proporção do estrato em relação à população, deve-se multiplicar o tamanho total da amostra (**n**) pela proporção de cada estrato na população (**N_h/N**). Assim, tem-se um tamanho de amostra em cada estrato, proporcional ao tamanho do estrato em relação à população.

3.3.3. Amostragem aleatória por conglomerados (probabilística)

A amostragem estratificada apresenta resultados satisfatórios, porém a sua implementação é dificultada pela falta de informações para fazer a estratificação da

população. Para contornar este problema, é possível utilizar o esquema de amostragem por conglomerados.

Os conglomerados são definidos em função da experiência do pesquisador. Geralmente, os conglomerados são estabelecidos por fatores geográficos (por exemplo, bairros e quarteirões). A utilização da amostragem por conglomerados possibilita uma redução significativa do custo do processo de amostragem. Portanto, um conglomerado é um subgrupo da população, que individualmente reproduz a população, ou seja, individualmente os elementos que o compõem são muito heterogêneos entre si. Este tipo de amostragem é muito útil quando a população é grande (por exemplo, uma pesquisa em nível nacional).

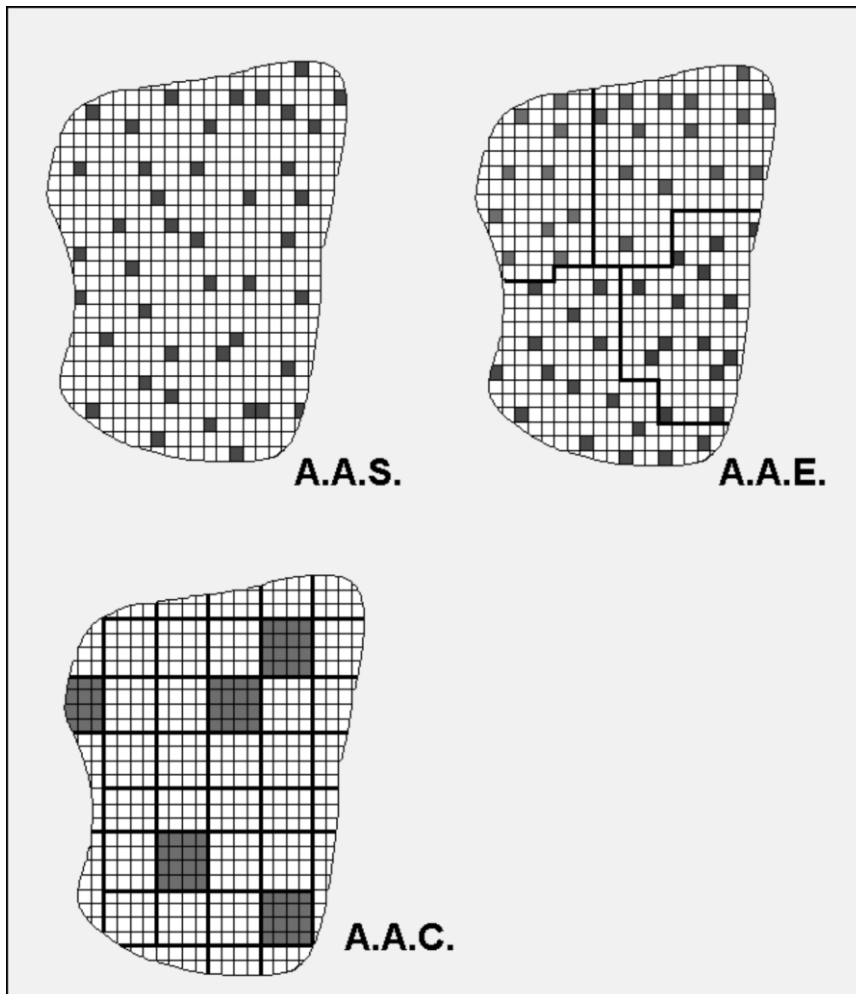
Na amostragem por conglomerados, primeiramente define-se o conglomerado e assim divide-se a população nos conglomerados. Sorteiam-se os conglomerados por meio de um processo aleatório e avaliam-se todos os indivíduos presentes nos conglomerados. Este processo é denominado de amostragem por conglomerados em um estágio. Caso se decida fazer um sorteio dos elementos dentro de cada conglomerado, tem-se uma amostragem por conglomerados em dois estágios.

3.3.4. Comparação entre AAS, AAE e AAC

A A.A.C. produz **resultados menos precisos** do que a A.A.S. com o mesmo tamanho de amostra e, por consequência, do que a A.A.E. É **mais barata** por agrupar os elementos da população. O tamanho da amostra **n** será determinado a posteriori, pelo número total de elementos observados nos conglomerados (no estágio final de amostragem).

Quadro comparativo entre os três métodos de amostragem

A.A.E.	Mais precisa do que a A.A.S., porém mais cara. Considera a heterogeneidade da população.
A.A.S.	Planejamento ideal. Pode ser muito cara. Não considera a heterogeneidade da população.
A.A.C.	Menos precisa do que a A.A.S. e A.A.E., porém mais barata. Resolve o problema do custo.



3.3.5. Amostragem aleatória sistemática (probabilística)

Em algumas situações, é conveniente retirar da população os elementos que vão compor a amostra de forma cíclica (em períodos), por exemplo, quando os elementos da população se apresentam ordenados. Porém, é de fundamental importância que a variável de interesse não apresente ciclos de variação coincidente com os ciclos de retirada, pois este fato tornará a amostragem não aleatória. Esta técnica de amostragem se chama amostragem sistemática.

O procedimento a ser adotado é o seguinte:

- determina-se o intervalo de seleção, que é dado por $R = N/n$.
- sorteia-se um elemento (ou item) dentre os R primeiros da relação.
- a partir daí, seleciona-se os indivíduos sistematicamente a cada intervalo de tamanho R .

3.3.6. Amostragens não probabilísticas

Em uma amostragem não probabilística, não se conhece, a priori, a probabilidade que um elemento da população tem de pertencer à amostra. Neste caso, não é possível calcular o erro decorrente da generalização dos resultados das análises estatísticas da amostra para a população de onde a amostra foi retirada.

Utiliza-se a amostragem não probabilística por simplicidade ou por impossibilidade de se obter uma amostra probabilística.

Os principais tipos de amostragem não probabilística que temos são:

- **Amostragem a esmo:**

Seja uma caixa com 1.000 parafusos. A enumeração destes parafusos é muito difícil, e a amostragem aleatória simples se torna inviável. Então, em situações deste tipo, supondo que a população de parafusos seja homogênea, escolhe-se a esmo a quantidade relativa ao tamanho da amostra. Quanto mais homogênea for a população, mais podemos supor a equivalência com uma AAS. Os parafusos serão escolhidos para compor a amostra de um determinado tamanho sem nenhuma norma ou a esmo.

- **Amostragem intencional:**

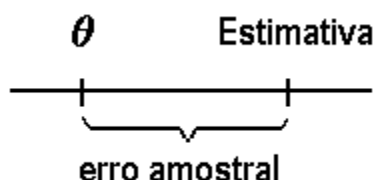
A amostragem intencional corresponde àquela em que o pesquisador escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população.

- **Amostragem por cotas:**

Neste tipo de amostragem, a população é dividida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, e sim os elementos são procurados até que a cota de cada grupo seja cumprida. Em pesquisas eleitorais, a divisão de uma população em grupos (considerando, por exemplo, o sexo, o nível de escolaridade, a faixa etária e a renda) pode servir de base para a definição dos grupos, partindo da suposição de que estas variáveis definem grupos com comportamentos diferenciados no processo eleitoral.

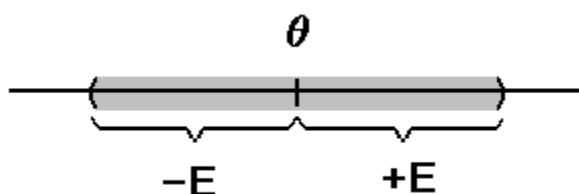
3.3. Erro amostral

O **erro amostral** é definido como sendo a diferença entre a estimativa obtida para um parâmetro e o seu verdadeiro valor (θ). É decorrente da variabilidade natural das unidades amostrais (**é aleatório**). Assim, erro amostral = estimativa – θ .



O erro decorrente da coleta dos dados é chamado de erro não amostral. Os planejamentos e a execução da pesquisa devem ser feitos com muita cautela para evitar os erros não amostrais. Alguns erros em amostragem: (i) população acessível diferente da população alvo; (ii) falta de resposta; (iii) erros de mensuração.

A determinação do tamanho da amostra é, talvez, o grande dilema dos pesquisadores, pois deve levar em conta um erro tolerável e a probabilidade de se cometer tal erro. O **erro tolerável** é uma margem de erro das estimativas em relação ao parâmetro θ , para mais ou para menos, o qual o pesquisador está disposto a aceitar.



O tamanho da amostra é determinado tal que a probabilidade de que a estimativa do parâmetro θ esteja dentro da margem de erro seja alta, por exemplo, de 95%.

3.4. Tamanho amostral

Antes de iniciar o cálculo do tamanho da amostra é imprescindível definir qual o erro amostral tolerável para o estudo que será realizado.

Após definir o erro amostral, determina-se a primeira aproximação do tamanho da amostra:

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

onde: **n₀** é a primeira aproximação do tamanho da amostra

E₀ é o erro amostral tolerável (Ex.: 2% = 0,02)

Na sequência, faz-se a correção do tamanho da amostra com a fórmula:

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

onde: **N** é o número de elementos da população

n é o tamanho da amostra

Considerando, agora, uma empresa que contém 2.000 colaboradores. Deseja-se fazer uma pesquisa de satisfação. Quantos colaboradores devem ser entrevistados para tal estudo?

$$N = 2.000$$

$$E_0 = 0,02 \text{ (erro amostral tolerável em 2\%)}$$

$$n_0 = 1 / (E_0)^2$$

$$n_0 = 1 / (0,02)^2$$

$$n_0 = 2.500$$

$$n = (N \cdot n_0) / (N + n_0)$$

$$n = (2.000 \times 2.500) / (2.000 + 2.500) = \mathbf{1.111 \text{ colaboradores}}$$
 (com o erro amostral tolerável em 2%, 1.111 colaboradores devem ser entrevistados para a pesquisa).

Considerando, agora, o erro amostral tolerável em 4%.

$$N = 2.000$$

$$E_0 = 0,04$$

$$n_0 = 1 / (E_0)^2$$

$$n_0 = 1 / (0,04)^2$$

$$n_0 = 625$$

$$n = (N \cdot n_0) / (N + n_0)$$

$$n = (2.000 \times 625) / (2.000 + 625)$$

$$\mathbf{n = 476 \text{ colaboradores}}$$

E se houvesse 300.000 colaboradores na empresa?

$$N = 300.000$$

$$E_0 = 0,04$$

$$n_0 = 1 / (E_0)^2$$

$$n_0 = 1 / (0,04)^2$$

$$n_0 = 625$$

$$n = (N \cdot n_0) / (N + n_0)$$

$$n = (300.000 \cdot 625) / (300.000 + 625)$$

$$n = 623 \text{ colaboradores}$$

Observa-se que a diferença entre **n** e **n₀**, neste último cálculo, é muito pequena. Portanto, se o número de elementos da população (N) é muito grande, a primeira aproximação do tamanho da amostra já é suficiente.