



By @kakashi_copiador

TEORIA DA AMOSTRAGEM

Introdução

Nesta aula, estudaremos as **etapas** da **Inferência Estatística** (também chamada de **Estatística Inferencial**), que nos ajuda a tirar conclusões a respeito de um **todo** (que chamamos de **população**) a partir das observações feitas em uma **parte** dessa população (que chamamos de **amostra**). A inferência é uma técnica importante porque normalmente **não** é possível conhecer a informação **exata** (por exemplo, altura, idade, salário, etc.) para **toda a população**.

Por exemplo, para conhecer a **altura média dos brasileiros**, os órgãos responsáveis por essa estatística NÃO verificam a altura de TODOS os brasileiros para conhecer exatamente a sua média. Em vez disso, é selecionada uma **amostra** de brasileiros, a partir da qual são feitas as **medições** necessárias. Por fim, com base nessas **medições da amostra**, são feitos os cálculos necessários para tirar **conclusões** a respeito da altura média da **população** de brasileiros.

Vamos destacar os seguintes conceitos do processo que acabamos de descrever:

- A **característica numérica da população** que se deseja conhecer (no nosso exemplo, a altura média da população de brasileiros) é chamada de **parâmetro populacional**;
- A medida correspondente feita na **amostra** (no caso, a altura média dos brasileiros da amostra selecionada) é chamada de **parâmetro de estimativa** ou **estatística da amostra**;
- As **conclusões** a respeito da população feitas a partir da amostra são chamadas de **inferência**.

As etapas desse processo de inferência estão representadas abaixo:



Esta aula lida com as maneiras de selecionarmos uma amostra, que é a primeira etapa.

CONCEITOS INICIAIS

A Teoria da Amostragem estuda a **extração de amostras** de uma dada **população**. Uma população é um **conjunto de elementos** que apresenta pelo menos uma **característica em comum**, enquanto uma amostra é um **subconjunto próprio** dessa população. O termo **próprio** significa que o subconjunto **não é vazio** e também **não contém todos os elementos** do conjunto original.

Nesse sentido, a amostra se distingue do **censo**, em que **todos os elementos** da população são examinados. Por não analisar todos os elementos, a amostragem **não** é uma técnica **exata**, sendo, portanto, recomendada quando o censo for inviável ou extremamente custoso.



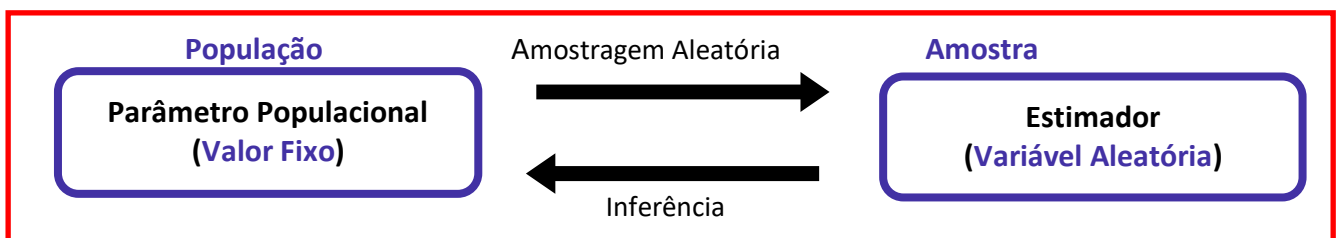
A característica da população que se deseja estudar, descrita numericamente, é chamada de **parâmetro populacional**. São exemplos de parâmetro populacional, a média populacional, a variância populacional, a proporção populacional, etc. Por ser uma característica da população, o parâmetro populacional é **fixo**, **não varia**.

O parâmetro populacional é inferido a partir dos resultados encontrados na amostra, que chamamos de **estimador** (ou **estimativa**) **do parâmetro populacional**, **parâmetro amostral** (ou **de estimativa**). Um estimador é uma **estatística da amostra**, ou seja, uma **função dos dados observados**, que apresenta determinadas características desejáveis (essas características serão estudadas na seção 4). Ou seja, os estimadores são calculados **após** a extração das amostras.

Por exemplo, a média amostral \bar{x} é o estimador adequado para estimar a média populacional μ , sendo calculado pela razão entre o somatório dos elementos da amostra x , dividido pelo número de elementos da amostra n (função da amostra):

$$\bar{x} = \frac{\sum x}{n}$$

Como os valores dos elementos da amostra x variam, pois a cada extração podemos obter elementos **diferentes**, o estimador também **varia**. Ou seja, o estimador é uma **variável aleatória**. Portanto, podemos representar o processo de amostragem e o de inferência como a seguir:



Os elementos da população a serem selecionados podem ser chamados de **unidades amostrais**. Por exemplo, vamos supor que estejamos interessados em estudar as condições socioeconômicas de uma população. Para isso, podemos selecionar diferentes pessoas dessa população, caso em que as unidades amostrais corresponderão a pessoas. Por outro lado, pode ser mais interessante estudarmos famílias como um todo - afinal, o salário de uma pessoa isolada pode não ser condizente com a sua condição socioeconômica, que pode estar mais relacionada com a situação familiar. Nesse caso, as unidades amostrais serão as famílias.

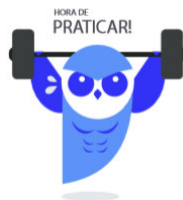
A **quantidade** de elementos da **amostra**, n , pode ser chamada de **tamanho amostral**; e a quantidade de elementos da população, N , de **tamanho populacional**. Também podemos adotar uma perspectiva diferente e considerar os elementos como **área**. Nesse caso, as unidades amostrais são **parcelas da área**; o tamanho amostral corresponde à **área total da amostra** (a); e o tamanho populacional corresponde à **área total da população** (A).

A razão entre o tamanho da amostra n e o tamanho da população N é chamada de **fração amostral** ou **intensidade amostral** e representa a proporção de indivíduos da população selecionados na amostra:

$$f = \frac{n}{N}$$

A razão **inversa** pode ser chamada de **fração de seleção** e corresponde ao número de indivíduos da população que cada indivíduo selecionado na amostra representa:

$$f_s = \frac{N}{n}$$



(2017 – Conter) Entre os casos em que se recomenda a utilização de amostragem, cita-se a situação em que

- a) haja necessidade de alta precisão.
- b) as características da população sejam de fácil mensuração, mesmo com população que não seja pequena.
- c) a população seja muito pequena e a amostra fique relativamente grande.
- d) a população seja muito pequena e suas características sejam totalmente desconhecidas.
- e) a população seja numerosa e com características bastante homogêneas.

Comentários:

A amostragem não fornece uma informação exata, logo, quando houver necessidade de alta precisão, a amostragem não é indicada. Portanto, a alternativa A está incorreta.

A amostragem é recomendada quando o censo for inviável ou extremamente custoso, logo não é necessária quando a população é pequena ou quando a característica é de fácil mensuração. Logo, as alternativas B, C e D estão incorretas.

A alternativa E está correta porque a amostragem é, de fato, indicada para uma população numerosa, quando há pelo menos uma característica em comum.

Gabarito: E

(2018 – Prefeitura de Raposa/MA) Dentro do contexto da Estatística, assinale a alternativa que preenche corretamente a lacuna do texto abaixo:

_____ é um valor aproximado do parâmetro e é calculado com o uso da amostra.

- a) Variável
- b) Estimativa
- c) Atributo
- d) População

Comentários:

O valor calculado a partir da amostra que aproxima o parâmetro populacional é a estimativa (ou estatística).

Gabarito: B

(CESPE/2013 – PC/DF) Julgue o item abaixo, a respeito de técnicas de amostragem.

Em uma amostragem sistemática cuja fração de seleção seja igual a 3 e o tamanho resultante da amostra seja igual a 125.000 observações, o tamanho da população será superior a 300.000 elementos.

Comentários:

Essa questão denota a razão entre o tamanho da população e o tamanho da razão por fração de seleção:

$$f_s = \frac{N}{n}$$

Sendo $f_s = 3$ e $n = 125.000$, o tamanho da população é:

$$N = f_s \times n = 3 \times 125.000 = 375.000$$

Que é superior a 300.000

Gabarito: Certo

Tipos de Erros

Os erros do processo de amostragem podem ser amostrais ou não amostrais, de modo que o erro total da amostragem corresponde à soma de ambos os erros.

É importante não confundir **erro** com **viés** do estimador. O **viés** (ou vício ou tendência) do estimador é a diferença entre o seu valor **esperado** (ou sua esperança) e o **verdadeiro valor** do parâmetro. Assim, um estimador não viesado (ou não viciado ou ainda não tendencioso) é aquele cujo valor esperado equivale ao verdadeiro parâmetro populacional.

Erros Amostrais

O **erro amostral** é definido como a **diferença** entre o estimador, que podemos representar por $\hat{\theta}$, e o parâmetro populacional sendo estimado, que podemos representar por θ :

$$e = \hat{\theta} - \theta$$

Esses erros são intrínsecos ao processo de amostragem (plano amostral), isto é, ocorrem pelo simples fato de os estimadores serem **variáveis aleatórias**.

Erros amostrais podem ser **controlados** e **mensurados** com base no plano amostral e tendem a **reduzir** com o **aumento** do número de elementos na **amostra**. Por exemplo, uma estimativa baseada em uma amostra de 100 elementos será mais precisa (ou seja, com **menor erro**) do que se tivesse baseada em uma amostra de apenas 20 elementos. No limite, caso toda a população fosse analisada (censo), o erro amostral seria nulo.

Adotando-se uma perspectiva diferente, podemos fixar o erro amostral máximo tolerado e, com base nele, calcular o tamanho da amostra necessário. Quanto menor o erro tolerado, maior será a amostra.

O tamanho da amostra adequado também pode variar de acordo com o **tamanho da população** - quanto maior a população, maior a amostra. Quando a população é infinita (ou muito grande em relação à amostra), o cálculo do tamanho da amostra **independe** do tamanho da população.

Erros Não Amostrais

Já os **erros não amostrais**, (ou **erros alheios à amostragem** ou **erros sistemáticos**) decorrem de **falhas no processo de amostragem**, **não** podendo ser **medidos** e nem totalmente **controlados**. Esses erros não estão relacionados com a variabilidade dos estimadores e existiriam ainda que toda a população fosse analisada. Dentre os erros não amostrais, podemos citar erros de especificação, erros de cobertura, não resposta, erros nas respostas e erros de processamento.

A **especificação incorreta** da população ocorre quando o pesquisador não entende a **população de interesse**.

Por exemplo, em uma pesquisa de preferências de alimentos para crianças, indagar apenas os pais pode produzir resultados incompletos, uma vez que as crianças podem ter alguma influência nas decisões dos pais.

O **erro de cobertura**, que também pode ser chamado de *frame error*, ocorre quando a amostra **não representa** bem a população de interesse, seja por **deixar de fora** determinados grupos de interesse ou **incluir** elementos que **não estão no escopo** da pesquisa. Por exemplo, se uma pesquisa atual voltada para a classe média utilizar como referência uma lista de telefones fixos, os resultados poderão ser inadequados, pois uma parcela significativa da população, que possui apenas telefones móveis, não estará representada na amostra.

A **não resposta** é um problema grave que ocorre quando uma parcela significativa da amostra **não responde** à pesquisa, o que pode gerar resultados tendenciosos. Pesquisas de satisfação, por exemplo, podem representar as opiniões de uma pequena parcela de clientes, quais sejam aqueles que se dispõem a responder a pesquisa.

Erros nas respostas dadas pelos entrevistados podem ocorrer quando as **perguntas** são **inadequadas**, sejam elas ambíguas, constrangedoras ou formuladas para o público errado (de modo que o entrevistado não saiba responder).

Por fim, **os erros de processamento** podem ocorrer em **qualquer fase** do processo de amostragem e estão associados a falhas humanas ou de softwares na codificação, digitação, transcrição, edição etc. das informações.



ESQUEMATIZANDO

Tipos de Erros

Erros Amostrais

- São intrínsecos ao plano amostral
- Podem ser controlados e mensurados
- Reduzem com o aumento do número de elementos na amostra

Erros não Amostrais (Sistemáticos)

- Decorrem de falhas no processo de amostragem
- Não podem ser controlados ou medidos
- Existem ainda que toda a população seja analisada



(FCC/2015 – CNMP – Adaptada) Julgue a seguinte afirmação.

Em qualquer tipo de amostragem, a não resposta é uma fonte potencial de erro.

Comentários:

A não resposta é um exemplo de erro não amostral, o qual pode ocorrer em qualquer tipo de amostragem.

Resposta: Certo

(CESPE/2014 – ANATEL) Em junho de 2014, o Brasil registrou 275,71 milhões de linhas ativas na telefonia móvel e teledensidade de 136,06 acessos por 100 habitantes. Além disso, nesse mesmo mês, houve um acréscimo de 255,08 mil linhas na telefonia móvel: os acessos pré-pagos totalizaram 212,27 milhões (76,99% do total) e os pós-pagos, 63,44 milhões (23,01% do total). A banda larga móvel totalizou 128,49 milhões de acessos, dos quais 3,27 milhões eram terminais 4G. Considerando as informações apresentadas no texto acima e supondo que um analista pretenda elaborar um plano amostral por meio de uma amostra aleatória simples sem reposição das linhas ativas na telefonia móvel com o objetivo de estimar a proporção de ligações não completas em junho de 2014, julgue o item a seguir.

Nos erros não amostrais que o analista poderá identificar incluem-se os erros sistemáticos.

Comentários:

Erros sistemáticos são, de fato, erros não amostrais, quais sejam, aqueles decorrentes de falhas no processo de amostragem.

Gabarito: Certo

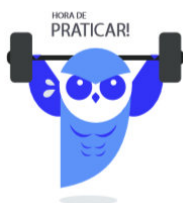
TIPOS DE AMOSTRAGEM

Para que a estimativa da população seja confiável, precisamos que a amostra extraída seja **representativa**, isto é, que as **proporções populacionais de interesse estejam refletidas na amostra**. Para isso, a extração da amostra precisa obedecer a determinados **processos**.

O processo de amostragem pode ser **probabilístico** ou **não probabilístico**. A amostragem **probabilística** (ou **estatística** ou **aleatória**) segue métodos rigorosamente científicos, em que cada elemento possui uma **probabilidade conhecida, maior que zero**, de pertencer à amostra.

Já na amostragem **não probabilística**, há uma **escolha deliberada** dos elementos da amostra, a depender do pesquisador/entrevistador. Logo, não é possível calcular a probabilidade de seleção dos elementos.

A amostragem probabilística produz resultados mais confiáveis, sendo possível o emprego de técnicas que permitem compensar falhas no processo de amostragem (erros não amostrais), como os **métodos de expansão**, que permitem generalizar os resultados encontrados na amostra. Essa técnica é empregada, por exemplo, para compensar a não resposta ou a inviabilidade de se analisar determinados elementos selecionados na amostra.



(2020 – Câmara de Xaxim/SC – Adaptada) Julgue as afirmativas abaixo sobre amostragem.

- I – Pela amostragem probabilística, o julgamento é usado para selecionar os itens da amostra.
- II – Pela amostragem não probabilística, os itens da amostra são selecionados de modo que cada unidade de amostragem tenha uma probabilidade conhecida de ser selecionada.

Comentários:

Vimos que o tipo de amostragem que depende do **julgamento** do entrevistador é a amostragem **não** probabilística; e que o tipo de amostragem em que todos os elementos apresentam uma probabilidade conhecida (maior que zero) é a amostragem **probabilística** (ou estatística).

Resposta: ambas as afirmativas estão incorretas.

(2017 – DEMAE) Na auditoria governamental, o método de amostragem é aplicado como forma de viabilizar a realização de ações de controle em situações nas quais o objeto alvo se apresenta em grandes quantidades e/ou distribui-se de maneira pulverizada. Os métodos de amostragem são classificados em:

- a) multivariado e univariado.
- b) fatorial e não fatorial.
- c) estatístico e aleatório.

d) probabilístico e não probabilístico.

Comentários:

Os métodos de amostragem são classificados em probabilístico e não probabilístico.

Gabarito: D.

Tipos de Amostragem Probabilística

Nesta seção, veremos tipos de amostragem **probabilística**, isto é, aquelas que seguem métodos rigorosamente científicos, quais sejam, (1) a amostragem aleatória simples; (2) a amostragem sistemática; (3) a amostragem por estratificação; (4) a amostragem por conglomerados; e (5) a amostragem múltipla.

Amostragem Aleatória Simples

Na **Amostragem Aleatória Simples (AAS)**, também chamada de **casual** ou **acidental**, cada elemento da população tem a **mesma probabilidade** de ser selecionado.

Esse tipo de amostragem probabilística consiste basicamente em **sortear** os elementos que farão parte da amostra, sendo indicado para populações **homogêneas**.

A probabilidade de um elemento qualquer ser selecionado na amostra pode ser calculada utilizando-se a **definição clássica de probabilidade**, isto é, a razão entre o número de elementos favoráveis e o número de elementos totais.

Para uma amostra de tamanho n e uma população com N elementos no total, a probabilidade de um elemento ser selecionado é:

$$p = \frac{n}{N}$$

Por exemplo, para uma população com $N = 100$ pessoas e uma amostra de tamanho $n = 20$, a probabilidade de um elemento qualquer ser selecionado é:

$$p = \frac{20}{100} = 0,2$$



O **inverso da probabilidade** de um elemento ser incluído na amostra é chamado de **peso amostral**. Por exemplo, se a probabilidade de um elemento ser selecionado na amostra for $p = 0,2$, então o peso amostral será $w = \frac{1}{0,2} = 5$.

Esse conceito transmite a ideia de **quantas unidades da população estão representadas em cada elemento da amostra**. Para esse exemplo, cada elemento da amostra representa 5 elementos da população.

Uma AAS pode ser extraída **com** ou **sem reposição**, ou seja, os elementos selecionados para as amostras podem **ser** ou **não repostos** ao conjunto da população.

Se a população for **infinita**, as extrações com e sem **reposição** (ou **repetição**) são **equivalentes**, pois a retirada de elementos da população **não** influencia em seu tamanho, que continua sendo infinito. Na prática, temos aproximadamente essa situação quando a população é **muito grande**, em especial, quando comparada ao tamanho da amostra.

Por outro lado, se a população for **finita** (e não muito grande), as possibilidades das possíveis amostras extraídas por esses dois procedimentos (**com ou sem reposição**) são **distintas**.

Denotando o número de elementos da população por N e o número de elementos da amostra por n , o número de **amostras possíveis**, extraídas **com reposição**, é dado por:

$$N^n = \underbrace{N \times N \times \dots \times N}_{n \text{ vezes}}$$

Isso porque há N possibilidades para a seleção de cada elemento, multiplicado pelo número de elementos selecionados, n (princípio multiplicativo de análise combinatória).

Por exemplo, se a população for de 100 pessoas, então para cada extração há 100 possibilidades. Se a amostra for de 3 pessoas, então o número de amostras distintas possíveis com reposição é:

$$100 \times 100 \times 100 = 1.000.000$$

Consequentemente, a **probabilidade** de um elemento ser incluído na amostra, em qualquer extração, é sempre a **mesma**, igual a:

$$P = \frac{1}{N}$$

Por outro lado, se a amostra é extraída **sem reposição**, o número de amostras possíveis é:

$$C_{N,n} = \frac{N!}{(N-n)!n!}$$

Ou seja, corresponde às possibilidades de escolher n elementos, dentre N , no total. Lembre-se que também podemos denotar essa combinação por C_N^n ou por $\binom{N}{n}$.

Para a mesma população de 100 pessoas e a amostra de 3 pessoas, o número de amostras distintas possíveis **sem reposição** é:

$$C_{100,3} = \frac{100!}{(100-3)!3!} = \frac{100 \times 99 \times 98 \times 97!}{97! \times 3!} = \frac{100 \times 99 \times 98}{3 \times 2} = 100 \times 33 \times 49 = 161.700$$



Quando as amostras são extraídas **com reposição**, as amostras são consideradas **ordenadas**, ou seja, selecionar (A,B,C) é **diferente** (C,A,B).

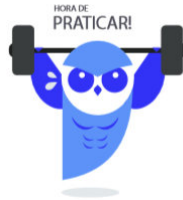
Quando as amostras são extraídas **sem reposição**, as amostras são consideradas **não ordenadas**, ou seja, (A,B,C) e (C,A,B) são consideradas a **mesma** possibilidade.

Para a AAS **sem reposição**, as probabilidades **variam** a cada extração:

- A probabilidade de um elemento ser selecionado na primeira extração é $P_1 = \frac{1}{N}$;
- A probabilidade de um elemento, dentre os $N - 1$ elementos não selecionados, ser selecionado na segunda extração é: $P_2 = \frac{1}{N-1}$
- A probabilidade de um elemento, dentre os $N - 2$ elementos não selecionados, ser selecionado na terceira extração é: $P_2 = \frac{1}{N-2}$
- ...

A AAS sem reposição é considerada mais eficiente do que a AAS com reposição, pois não são obtidas novas informações quando selecionamos o mesmo elemento para a amostra, mais de uma vez.

Pontue-se que, **quando a população é infinita (na prática, muito grande) OU quando as amostras são extraídas com reposição, as observações são independentes.** Por outro lado, quando a população é finita E as extrações são feitas com reposição, as observações são dependentes.



(FCC/2007 – ANS – Adaptada) Com relação à teoria geral da amostragem, julgue o item a seguir.

Na realização de amostragem aleatória simples, as unidades amostrais são selecionadas com probabilidade igual.

Comentários:

Na amostragem aleatória simples (AAS), todos os elementos possuem a mesma probabilidade de serem selecionados.

Resposta: Certo

(CESPE/2020 – TJ-PA) Muitos sorteios virtuais são realizados em uma plataforma que gera números de maneira aleatória, sendo cada número sorteado apenas uma vez com a mesma probabilidade. Essa técnica é denominada amostragem

- a) estratificada.
- b) aleatória simples com repetição.
- c) sistemática.
- d) aleatória simples sem repetição.
- e) por conglomerados.

Comentários:

A técnica de amostragem em que todos os elementos apresentam a mesma probabilidade de serem selecionados é a amostragem aleatória simples (AAS). Como os elementos podem ser sorteados apenas uma vez, então os elementos selecionados não são repostos. Logo, temos uma amostragem sem reposição (ou repetição).

Gabarito: D

(CESPE/2016 – TCE-PA) Suponha que o tribunal de contas de determinado estado disponha de 30 dias para analisar as contas de 800 contratos firmados pela administração. Considerando que essa análise é necessária para que a administração pública possa programar o orçamento do próximo ano e que o resultado da análise

deve ser a aprovação ou rejeição das contas, julgue o item a seguir. Sempre que necessário, utilize que $P(Z > 1,96) = 0,025$ e $P(Z > 1,645) = 0,05$, em que Z representa a variável normal padronizada.

Em uma amostra aleatória simples de 400 elementos, o peso amostral de cada elemento será maior ou igual a 2.

Comentários:

O peso amostral é o inverso da probabilidade um elemento ser incluído na amostra, que pode ser calculada pela definição clássica de probabilidade. Sabendo que a população possui $N = 800$ contas e a amostra terá tamanho $n = 400$, a probabilidade de um elemento ser selecionado é:

$$p = \frac{n}{N} = \frac{400}{800} = \frac{1}{2}$$

Logo, o peso amostral é o inverso:

$$w = \frac{1}{p} = \frac{1}{\frac{1}{2}} = 2$$

Gabarito: Certo

(FGV/2019 – DPE-RJ – Adaptada) Sobre os desenhos mais utilizados para a seleção da amostra e suas características, julgue a seguinte afirmativa:

A probabilidade “a priori” de seleção de um indivíduo numa AAS é diferente quando a amostra é sem ou com reposição.

Comentários:

A probabilidade de um elemento ser selecionado em qualquer extração de uma AAS com reposição é $P = \frac{1}{N}$, em que N é o tamanho da população.

Em uma AAS sem reposição, a probabilidade se altera a cada extração. Antes da primeira extração (a priori), a probabilidade de um elemento ser selecionado é também $P_1 = \frac{1}{N}$.

Logo, a probabilidade é a mesma.

Resposta: Errado

(FCC/2019 – Banrisul) Uma população consiste nos 6 primeiros números inteiros estritamente positivos, ou seja, $\{1, 2, 3, 4, 5, 6\}$. Seja n_1 o número de amostras aleatórias possíveis de 2 elementos que podem ser extraídas da população com reposição e n_2 o número de amostras aleatórias possíveis de 2 elementos que podem ser extraídas da população sem reposição. O módulo de $(n_1 - n_2)$ é igual a

- a) 49
- b) 24
- c) 26
- d) 30
- e) 21

Comentários:

Se a população consiste em $N = 6$ elementos e a amostra consiste em $n = 2$ elementos, então o número de amostras possíveis extraídas com reposição é:

$$n_1 = N \times N = 6 \times 6 = 36$$

E o número de amostras possíveis extraídas sem reposição é:

$$C_{N,n} = \frac{N!}{(N-n)!n!}$$
$$n_2 = C_{6,2} = \frac{6!}{(6-2)!2!} = \frac{6 \times 5 \times 4!}{4! \times 2} = \frac{6 \times 5}{2} = 15$$

Logo, a diferença $n_1 - n_2$ é:

$$n_1 - n_2 = 36 - 15 = 21$$

Como 21 é positivo, o módulo de 21 é ele mesmo: $|21| = 21$.

Gabarito: E

Amostragem Sistemática

A Amostragem Sistemática é também uma amostragem probabilística, que consiste em **ordenar os elementos da população, seguindo qualquer critério** (por exemplo, ordem alfabética), para que possam ser identificados pela respectiva posição. Em seguida, as amostras são extraídas **periodicamente**, de x em x elementos (por exemplo, de 10 em 10 elementos, de 25 em 25 elementos, etc.). Essa técnica é indicada para populações **homogêneas e finitas**.

Para definir as posições dos elementos a serem extraídos, denominadas **pontos de extração ou pontos de coleta**, primeiro calculamos a **razão** entre o número de elementos da **população** e o número de elementos da **amostra**:

$$R \cong \frac{N}{n}$$

E utilizamos a parte inteira da razão, **R** . Os elementos serão então extraídos de R em R elementos. Por exemplo, para uma população de $N = 20$ elementos e uma amostra de $n = 5$ elementos, então, temos:

$$R = \frac{20}{5} = 4$$

Para definir a posição do primeiro elemento a ser extraído (e, conseqüentemente, os demais elementos), pode-se **sortear um número de 1 a R** . Vamos supor, para o nosso exemplo, que tenha sido sorteado o número **3**, as extrações serão realizadas nos seguintes pontos:

- i) **3**;
- ii) **3 + 4 = 7**;
- iii) **7 + 4 = 11**;
- iv) **11 + 4 = 15**;
- v) **15 + 4 = 19**

Pontue-se que os pontos de extração seguem uma **Progressão Aritmética**, com razão **R** . Assim, o ponto de extração do n -ésimo elemento selecionado pode ser calculado por:

$$A_n = A_1 + (n - 1) \times R$$

Para o nosso exemplo, tivemos $R = 4$ e $A_1 = 3$. Assim, a posição do 5º elemento selecionado é:

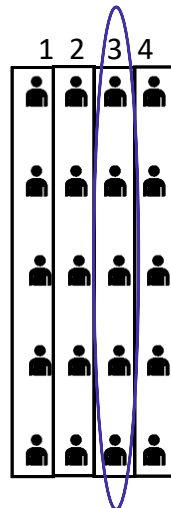
$$A_n = 3 + 4 \times 4 = 19$$

Ou seja, o 5º elemento selecionado será o 19º elemento populacional. Esse é o mesmo resultado que obtivemos acima.

Esse procedimento pode ser visto como a divisão dos $N = 20$ elementos da população em $n = 5$ grupos de $R = 4$ unidades, sendo selecionado o 3º elemento de cada grupo, como representado abaixo:

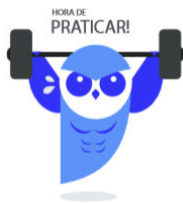


Também podemos visualizá-lo como a divisão dos $N = 20$ elementos da população em $R = 4$ grupos de $n = 5$ elementos, sendo selecionado um dos grupos (no caso, o 3º grupo), do qual todos os membros serão analisados (como na amostragem de conglomerados que veremos posteriormente):



Como as amostras são extraídas periodicamente, se houver uma **periodicidade nos dados**, os elementos selecionados podem não ser uma boa representação de toda a população.

Por exemplo, vamos supor que queremos estimar a quantidade esperada (média) de chuva anual em determinada região, com base em informações pretéritas. Se fizermos uma amostragem sistemática com intervalos de 12 meses, estaremos considerando as informações de apenas 1 mês no ano. Como a quantidade de chuva é sazonal, provavelmente não conseguiríamos uma boa estimativa da quantidade de chuvas no ano.



(FCC/2007 – ANS – Adaptada) Com relação à teoria geral da amostragem, julgue o item a seguir.

A amostragem sistemática é sempre menos precisa do que a amostragem aleatória simples.

Comentários:

A amostragem sistemática é uma técnica probabilística, assim como a amostragem aleatória simples. Logo, ambas são igualmente precisas.

Resposta: Errado.

(FGV/2019 – DPE-RJ – Adaptada) Sobre os desenhos mais utilizados para a seleção da amostra e suas características, julgue a seguinte afirmativa:

A amostragem sistemática tem como fragilidade o fato de que apenas a seleção do primeiro indivíduo é probabilística.

Comentários:

A amostragem sistemática é totalmente probabilística, e não somente para a seleção do primeiro elemento.

Resposta: Errado

(2021 – Prefeitura de Porto Alegre/RS) Assinale a alternativa que indica uma desvantagem da técnica de amostragem sistemática.

- a) Não ser probabilística.
- b) Exigir grandes tamanhos de amostragem.
- c) Podem ocorrer problemas se existir algum tipo de periodicidade oculta.
- d) Alta complexidade.
- e) Erro elevado

Comentários:

A amostragem sistemática é probabilística (logo, a alternativa A está errada) e pode ser aplicada para qualquer tamanho de amostra (logo, a alternativa B está errada). Ela também não é complexa e não apresenta erro elevado (alternativas D e E erradas). Porém, se houver alguma periodicidade nos dados, a amostra selecionada pode não representar bem a população.

Gabarito: C

(FGV/2022 – TJDF) Um Tribunal de Justiça deseja obter uma amostra de tamanho 3.000 de uma população de 60.000 ações. Esse Tribunal possui um cadastro em que cada ação está associada, sequencialmente, a um número (começando com o número 1 e terminando com o número 60.000).

De posse do referido cadastro e considerando o tamanho da amostra solicitada, o pesquisador utilizou o seguinte procedimento para a seleção da amostra:

1. Determinou o intervalo de seleção da amostra dividindo o total da população pelo tamanho da amostra: $60.000/3.000=20$;
2. Elegeu aleatoriamente um número inteiro, entre $[1, 20]$. Essa foi a primeira ação selecionada;
3. A próxima ação selecionada foi definida pela soma do intervalo de seleção ao número selecionado na etapa 2.

E, assim, sucessivamente, foram determinados os próximos elementos, acrescentando-se ao selecionado anteriormente o intervalo de seleção da amostra.

O número escolhido na etapa de número 2 foi 17; logo, a primeira ação selecionada foi a de número 17; a seguinte, a de número 37, seguida da de número 57, e assim sucessivamente.

O milésimo elemento selecionado nessa amostra foi a ação de número:

- a) 19.937;
- b) 19.957;
- c) 19.977;
- d) 19.997;
- e) 20.017.

Comentários:

O enunciado descreve a amostragem sistemática, em que primeiro calculamos a razão R entre o número de elementos da população e o número de elementos da amostra (aqui, temos $R = 20$); e, em seguida, selecionamos aleatoriamente a posição do primeiro elemento (aqui, temos $A_1 = 17$).

A posição do n -ésimo elemento pode ser calculada, utilizando-se a fórmula da Progressão Aritmética:

$$A_n = A_1 + (n - 1) \times R$$

Para $A_1 = 17$, $n = 1000$ e $R = 20$, temos:

$$A_{1000} = 17 + 999 \times 20 = 17 + 19980 = 19997$$

Gabarito: D

Amostragem por Estratificação

A amostragem por estratificação (ou estratificada) é aplicável para populações **heterogêneas**.

Ela consiste inicialmente em **segmentar** a população em grupos, denominados **estratos** (ou **subpopulações**), de modo que os elementos de um **mesmo estrato** sejam **homogêneos** (apresentem **baixa variabilidade**) e que os **diferentes estratos** sejam **heterogêneos** (apresentem alta variabilidade).

Em seguida, é realizada uma técnica de amostragem probabilística, normalmente a amostragem simples, para selecionar um número de elementos de cada estrato. Com isso, essa técnica de amostragem busca obter estimativas mais precisas para populações com tais características.

Por exemplo, para uma população de $N_T = 20$ pessoas, sendo 10 homens e 10 mulheres, teríamos um estrato para os 10 homens e outro para as 10 mulheres. Em seguida, poderíamos **sortear** 2 elementos, por exemplo, do estrato dos homens e 2 elementos do estrato das mulheres, como ilustrado abaixo:



Observe que os elementos de um mesmo grupo são homogêneos e que os diferentes grupos são heterogêneos. A figura acima ilustra, ainda, a seleção de alguns elementos (no caso, 2) de cada estrato.

Mas os tamanhos dos estratos (também chamados de **área**) não precisam ser iguais, como no exemplo acima. Vamos supor agora uma população de $N_T = 100$ pessoas, das quais $N_{Alta} = 10$ possuem alta renda, $N_{Baixa} = 50$ possuem baixa renda e $N_{Média} = 40$ são de classe média. Nesse caso, como definimos o número de elementos extraídos de cada estrato?

A amostra pode ser de **mesmo tamanho** para todos os estratos, o que chamamos de **amostragem uniforme**. Para o nosso exemplo, em uma **amostragem uniforme**, poderíamos selecionar $n_{Alta} = n_{Baixa} = n_{Média} = 3$ elementos de cada estrato.

Alternativamente, a amostra pode ser **proporcional** ao tamanho do estrato, o que chamamos de **amostragem proporcional**.

Para o nosso exemplo, podemos definir que queremos uma amostra total de $n_T = 10$ pessoas, extraída **proporcionalmente** de cada estrato. Para isso, primeiro calculamos a proporção entre a amostra total desejada n_T e a população total N_T :

$$p = \frac{n_T}{N_T} = \frac{10}{100} = 0,1$$

Esse percentual será mantido para todos os estratos. Logo, o tamanho da amostra para cada estrato alta renda é:

$$n_{Alta} = N_{Alta} \times p = 10 \times 0,1 = 1$$

$$n_{Baixa} = N_{Baixa} \times p = 50 \times 0,1 = 5$$

$$n_{Média} = N_{Média} \times p = 40 \times 0,1 = 4$$

Generalizando para um estrato x qualquer e substituindo p por $p = \frac{n_T}{N_T}$, o tamanho da amostra n_x pode ser calculado como:

$$n_x = N_x \times \frac{n_T}{N_T}$$

Também podemos escrever esse produto como:

$$n_x = n_T \times \frac{N_x}{N_T}$$

Fizemos um pequeno rearranjo dos fatores, mas, agora, estamos entendendo a amostra do estrato x como a razão entre a população do estrato x e a população total, multiplicada pelo tamanho total da amostra.

Para o nosso exemplo, a amostra da população de baixa renda, n_{Baixa} , pode ser vista como a razão entre o tamanho da população de baixa renda, $N_{Baixa} = 50$, e o tamanho da população total, $N_T = 100$, que é $\frac{N_{Baixa}}{N_T} = 0,5$, multiplicada pelo tamanho total da amostra $n_T = 10$:

$$n_{Baixa} = n_T \times \frac{N_{Baixa}}{N_T} = 10 \times \frac{50}{100} = 10 \times 0,5 = 5$$

O resultado é o mesmo, como era de se esperar. Então, por que você está falando isso?

Bem, mudamos um pouquinho o ponto de vista da amostragem proporcional porque vamos utilizar esse segundo raciocínio para obter o tamanho da amostra de um terceiro tipo de amostragem estratificada, que é **a amostragem ótima (de Neyman)**.

Nesse tipo de amostragem, o tamanho amostral de cada estrato depende tanto do **tamanho populacional**, quanto da sua **variabilidade**, medida pelo **desvio padrão**. Ou seja, para estratos com **maior variabilidade as amostras são maiores** para uma **melhor representação** das diferentes características.

Para calcular o tamanho da amostra por estrato, temos uma fórmula parecida com a que acabamos de ver para a amostragem proporcional, mas precisamos multiplicar o tamanho de cada estrato pelo respectivo desvio padrão:

$$n_x = n_T \times \frac{N_x \times D_x}{\sum_i N_i \times D_i}$$

Supondo que a variabilidade da população de alta renda seja $D_{Alta} = 4$, da população de baixa renda seja $D_{Baixa} = 2$ e da população de classe média seja $D_{Média} = 1$. Lembrando que há $N_T = 100$ pessoas no total, das quais $N_{Alta} = 10$ possuem alta renda, $N_{Baixa} = 50$ possuem baixa renda e $N_{Média} = 40$ são de classe média, vamos calcular o denominador da fração acima, qual seja, o somatório dos produtos entre o tamanho populacional de cada estrato e o desvio padrão de cada estrato:

$$\sum_i N_i \times D_i = (N_{Alta} \times D_{Alta}) + (N_{Baixa} \times D_{Baixa}) + (N_{Média} \times D_{Média})$$

$$\sum_i N_i \times D_i = (10 \times 4) + (50 \times 2) + (40 \times 1) = 40 + 100 + 40 = 180$$

Considerando que a amostra total será de $n_T = 10$ pessoas, então os tamanhos das amostras de cada estrato são:

$$n_{Alta} = n_T \times \frac{N_{Alta} \times D_{Alta}}{\sum_i N_i \times D_i} = 10 \times \frac{10 \times 4}{180} = 10 \times \frac{40}{180} \cong 2$$

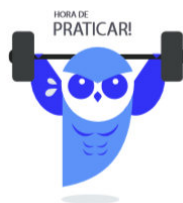
$$n_{Baixa} = n_T \times \frac{N_{Baixa} \times D_{Baixa}}{\sum_i N_i \times D_i} = 10 \times \frac{50 \times 2}{180} = 10 \times \frac{100}{180} \cong 6$$

$$n_{Média} = n_T \times \frac{N_{Média} \times D_{Média}}{\sum_i N_i \times D_i} = 10 \times \frac{40 \times 1}{180} = 10 \times \frac{40}{180} \cong 2$$

A amostragem ótima possui o **menor** erro amostral, dentre as técnicas de amostragem estratificada que vimos, enquanto a amostragem uniforme possui o **maior** erro amostral delas (a amostragem **proporcional** apresenta erro amostral **intermediário**).

Amostragem Estratificada	Uniforme	Mesmo tamanho de amostra para todos os estratos	Mais Erro
	Proporcional	Amostras proporcionais aos tamanhos dos estratos	Erro Médio
	Ótima	Amostra depende de tamanho e variabilidade dos estratos	Menos Erro

Pontue-se, ainda, que quanto mais **homogêneo** forem os elementos **dentro** de cada estrato, com relação às características estudadas, e mais **heterogêneos** forem os **diferentes** estratos, mais eficiente será o plano amostral, o que significa que o erro amostral será menor.



(FCC/2013 – SEFAZ/SP – Adaptada) Julgue o item a seguir.

Na amostragem aleatória estratificada, a população é dividida em estratos, usualmente, de acordo com os valores ou categorias de uma variável, e, depois, uma amostragem aleatória simples é utilizada na seleção de uma amostra de cada estrato.

Comentários:

Na amostragem aleatória estratificada, a população é dividida em estratos, de acordo com algum critério (por exemplo, idade, sexo, etc.) e uma amostragem aleatória é utilizada para selecionar uma amostra de cada estrato.

Resposta: Certo

(FCC/2007 – ANS – Adaptada) Com relação à teoria geral da amostragem, julgue o item a seguir.

Amostragem estratificada consiste na divisão de uma população em grupos segundo alguma característica conhecida. Os estratos da população devem ser mutuamente exclusivos.

Comentários:

Na amostragem aleatória estratificada, a população é dividida em grupos, chamados estratos, sem interseção entre eles (mutuamente exclusivos).

Resposta: Certo

(FCC/2015 – CNMP – Adaptada) Julgue a afirmação a seguir.

Na amostragem estratificada fica assegurado que cada extrato esteja representado na amostra global mas não fica assegurado que todas as unidades de estudo tenham a mesma probabilidade de serem selecionadas.

Comentários:

Na amostragem aleatória estratificada, cada extrato da população será representado na amostra, mas nem todos os elementos de toda a população possuem a mesma probabilidade de serem selecionados.

Resposta: Certo

(CESPE/2020 – TJ-PA) Um professor de educação física realizou uma pesquisa a respeito das alturas dos estudantes da instituição de ensino onde trabalha. A instituição possui 1.285 estudantes, dos quais 535 são homens e 750 são mulheres. Para realizar essa pesquisa, foi selecionada uma amostra de 257 estudantes pelo método de amostragem estratificada com alocação proporcional, considerando-se os estratos homem e mulher. Nessa situação, foram selecionados

- a) 107 homens e 150 mulheres
- b) 128 homens e 129 mulheres
- c) 110 homens e 147 mulheres
- d) 150 homens e 107 mulheres
- e) 129 homens e 128 mulheres

Comentários:

Para calcular o tamanho da amostra de cada estrato, em uma amostragem estratificada proporcional, primeiro calculamos a proporção do tamanho total da amostra em relação ao tamanho total da população:

$$p = \frac{n}{N} = \frac{257}{1285} = 0,2$$

Essa é a proporção que será mantida para ambos os estratos. Logo, o tamanho da amostra de cada estrato é:

$$n_{Homens} = p \times N_{Homens} = 0,2 \times 535 = 107$$

$$n_{Mulheres} = p \times N_{Mulheres} = 0,2 \times 750 = 150$$

Gabarito: A

(CESPE/2020 – TJ-PA) Uma população é dividida nos estratos I, II e III. O estrato I é composto por 400 elementos; o II, por 600 elementos; e o III, por 1.000 elementos. Conforme um estudo piloto, os desvios padrão da variável de interesse nos estratos I, II e III são, respectivamente, 10, 20 e 8.

Caso um pesquisador pretenda retirar uma amostra aleatória de 240 elementos dessa população utilizando a locação ótima de Neyman, os tamanhos das amostras a serem extraídas dos estratos I, II e III devem ser, respectivamente,

- a) 40, 30 e 170
- b) 40, 120 e 80
- c) 48, 72 e 120
- d) 79, 81 e 80
- e) 50, 75 e 115

Comentários:

Essa questão trabalha com a alocação estratificada ótima, em que o tamanho amostral do estrato x é dado por:

$$n_x = n_T \times \frac{N_x \times D_x}{\sum_i N_i \times D_i}$$

O enunciado informa que:

- O tamanho da população do estrato I é $N_I = 400$
- O tamanho da população do estrato II é $N_{II} = 600$
- O tamanho da população do estrato III é $N_{III} = 1000$
- O desvio padrão da população do estrato I é $D_I = 10$
- O desvio padrão da população do estrato II é $D_{II} = 20$
- O desvio padrão da população do estrato III é $D_{III} = 8$

Com essas informações, podemos calcular o denominador da fórmula, que corresponde ao somatório dos produtos dos tamanhos populacionais de cada estrato, multiplicados pelos respectivos desvios-padrão:

$$\sum_i N_i \times D_i = (N_I \times D_I) + (N_{II} \times D_{II}) + (N_{III} \times D_{III})$$

$$\sum_i N_i \times D_i = (400 \times 10) + (600 \times 20) + (1000 \times 8) = 4000 + 12000 + 8000 = 24000$$

Considerando que o tamanho da amostra é $n_T = 240$, então os tamanhos amostrais são:

$$n_I = n_T \times \frac{N_I \times D_I}{\sum_i N_i \times D_i} = 240 \times \frac{400 \times 10}{24000} = \frac{4000}{100} = 40$$

$$n_{II} = n_T \times \frac{N_{II} \times D_{II}}{\sum_i N_i \times D_i} = 240 \times \frac{600 \times 20}{24000} = \frac{12000}{100} = 120$$

$$n_{III} = n_T \times \frac{N_{III} \times D_{III}}{\sum_i N_i \times D_i} = 240 \times \frac{1000 \times 8}{24000} = \frac{8000}{100} = 80$$

Gabarito: B

Amostragem por Conglomerados

Assim como a amostragem estratificada, a **amostragem por conglomerados** também consiste em dividir a população em **grupos** e utilizar uma técnica de **amostragem probabilística**, normalmente a AAS, em cada grupo.

Porém, ao contrário da primeira, os **diferentes grupos** formados na **amostragem por conglomerados**, são **homogêneos** entre si (com baixa variabilidade), enquanto os elementos de um **mesmo grupo** são **heterogêneos** (com alta variabilidade). Assim, cada grupo, que chamamos de **conglomerado** ou **cluster**, pode ser considerado uma pequena **representação** de toda população.

Por exemplo, para estudar os colaboradores que trabalham em montadoras, poderíamos agrupá-los de acordo com suas respectivas empresas. Assim, as diferentes empresas (conglomerados) teriam características similares (homogêneas) e, dentro de cada empresa, teríamos colaboradores com características muito distintas (heterogêneas).

Outra diferença entre as duas técnicas de amostragem é que, na **amostragem estratificada**, é aplicada uma técnica de amostragem para selecionar **alguns elementos de cada estrato**. Por outro lado, na **amostragem por conglomerados**, é aplicada uma técnica de amostragem para selecionar **alguns conglomerados**, dos quais **todos os elementos** são analisados. Para o nosso exemplo, selecionaríamos algumas empresas para analisar todos os seus colaboradores.

A figura a seguir ilustra a amostragem por conglomerados, para uma população de 10 conglomerados, dos quais 2 são selecionados para a amostra.



Observe que os elementos de cada grupo são heterogêneos e que os diferentes grupos são homogêneos. A figura acima ilustra, ainda, a seleção de alguns grupos (no caso, 2), dos quais **todos** os elementos são analisados.

Essa amostragem, em que analisamos todos os elementos de alguns conglomerados selecionados, é chamada de **amostragem por conglomerados em um estágio**. Se não for possível analisar conglomerados inteiros, então podemos subdividir cada conglomerado em novos grupos, preservando as mesmas

características de homogeneidade entre grupos e heterogeneidade dentro de cada grupo. Podemos efetuar essa subdivisão sucessivamente, até que todos os elementos de um grupo possam ser analisados. Essa técnica é chamada de amostragem por conglomerados em múltiplos estágios.

Para o nosso exemplo, em que as empresas montadoras de carros são multinacionais, seria inviável analisar todos os colaboradores de algumas empresas. Para viabilizar a análise, poderíamos subdividir as empresas selecionadas de acordo com os países em que operam. Em seguida, poderíamos subdividir os países de acordo com os estados de operação, e depois de acordo com as cidades, e assim sucessivamente, até que, no último estágio, todos os colaboradores sejam analisados.

Analogamente ao que vimos para a amostragem estratificada, quanto mais heterogêneos forem os elementos dentro de cada conglomerado, e mais homogêneos forem os conglomerados, menor o erro amostral, ou seja, mais eficiente o plano amostral.

Para utilizar a amostragem por conglomerado, deve haver uma lista dos conglomerados, não sendo necessária uma lista identificando todos os elementos da população, como nas demais técnicas de amostragem, e, por isso, tende a apresentar custos mais baixos. Além disso, quando os custos de análise da amostra aumentarem conforme a distância entre os elementos (por exemplo, quando a análise exigir a presença física do elemento), então a amostragem por conglomerados tende a apresentar custos ainda menores do que as demais técnicas probabilísticas.

Para ilustrar, vamos considerar novamente o nosso exemplo dos colaboradores. Poderíamos aplicar a amostragem por estratificação, dividindo a população nos diferentes grupos salariais e, em seguida, selecionando aleatoriamente elementos de cada grupo. Os elementos selecionados poderiam ser dos mais diversos países existentes, o que corresponderia a altos custos, caso a análise do elemento selecionado precisasse ser *in loco*. Já para a amostragem por conglomerados, podemos selecionar cidades específicas e efetuar as análises somente nesses locais.



ESQUEMATIZANDO

Amostragem Estratificada	Amostragem por Conglomerados
<ul style="list-style-type: none">Grupos de elementos homogêneos (chamados estratos)São selecionados alguns elementos de cada estrato, conforme o tipo de alocação (uniforme, proporcional ou ótima)As unidades amostrais são os elementos	<ul style="list-style-type: none">Grupos de elementos heterogêneos (chamados conglomerados)São selecionados alguns grupos, dos quais todos os elementos são analisadosAs unidades amostrais são os conglomeradosBaixo custo, principalmente quando este aumenta com a distância



(FCC/2007 – ANS – Adaptada) Com relação à teoria geral da amostragem, julgue o item a seguir.

Em uma amostra por conglomerados, a população é dividida em subpopulações distintas.

Comentários:

Na amostragem por conglomerados, os grupos (conglomerados) são compostos de elementos distintos (heterogêneos).

Resposta: Certo

(2017 – CRP 7ª Região – Adaptada) Em estudos numéricos, somente as amostragens probabilísticas permitem a correta generalização para a população dos resultados amostrais. Considerando essa informação, julgue a alternativa a seguir.

Na amostragem aleatória por conglomerados, se os elementos mostrais são os alunos, os conglomerados podem ser as respectivas escolas.

Comentários:

Na amostragem por conglomerados, a população é dividida em subgrupos formados por **elementos heterogêneos**, mas sendo os subgrupos homogêneos entre si, como ilustrado a seguir:



As escolas podem ser consideradas conglomerados (os retângulos ilustrados acima), pois são homogêneas entre si, enquanto os alunos de cada escola são elementos heterogêneos (são de anos/idades distintos, por exemplo).

Resposta: Certo.

(FGV/2019 – DPE-RJ – Adaptada) Sobre os desenhos mais utilizados para a seleção da amostra e suas características, julgue a seguinte afirmativa:

A amostra por conglomerados tem como principal restrição o custo de extração, que costuma ser elevado.

Comentários:

A amostragem por conglomerados tende a ser menos custosa, principalmente quando os custos da análise aumentam com a distância entre os elementos selecionados.

Resposta: Errado

Amostragem Múltipla

Na **amostragem múltipla**, as amostras são extraídas em **etapas sucessivas** e, dependendo dos resultados observados, as etapas posteriores podem ser **dispensadas**. Essa técnica é utilizada na **inspeção por amostragem**, isto é, quando é necessário tomar uma decisão sobre aceitar ou não um lote de produtos, com base na inspeção de apenas alguns elementos.

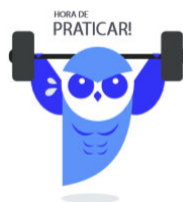
Por exemplo, quando uma empresa recebe um lote de 1000 produtos do seu fornecedor, os quais devem atender a determinadas exigências de qualidade. Essa empresa irá aceitar o lote se essas exigências forem atendidas e rejeitá-lo, caso contrário. Se não for viável inspecionar todos os produtos do lote, a empresa poderá inspecionar apenas alguns itens e, com base nos resultados encontrados, decidir se irá aceitar ou rejeitar o lote.

A empresa poderia decidir utilizar uma **amostragem única**, selecionando, digamos, uma amostra de 50 elementos para serem inspecionados. Alternativamente, a empresa poderia utilizar uma **amostragem dupla** (que é um **tipo de amostragem múltipla**), optando por 2 amostras de 25 elementos, extraídas em etapas **sucessivas** (uma após a outra). A partir dos resultados da primeira amostra, a empresa irá decidir se vai **aceitar o lote, rejeitá-lo ou extrair a segunda amostra**.

Essa técnica pode ser estendida para **múltiplas amostras sequenciais**. Ou seja, a empresa pode optar pela extração sequencial de 5 amostras com 10 elementos cada. Após a análise de cada amostra, a empresa decide se irá **aceitar o lote, rejeitá-lo ou extrair uma nova amostra**.

O objetivo da amostragem múltipla é **diminuir o número de elementos inspecionados a longo prazo, reduzindo, assim, os custos de inspeção**.

Um caso **extremo** dessa técnica é a **amostragem sequencial**, em que **um único elemento** é selecionado e inspecionado **por vez**, até que seja tomada a decisão de aceitar ou rejeitar o lote, com o objetivo de tornar mínimo o número de elementos inspecionados a longo prazo.



(FGV/2017 – IBGE) Dentre os métodos de amostragem probabilística, podem ser destacados os de amostras aleatórias simples, sistemáticas, por estratos, por cluster e múltipla.

Sobre cada um desses métodos, e nessa mesma ordem, poderiam ser associadas, relativamente, as seguintes palavras-chave ou expressões:

- a) equiprováveis, periodicidade, grupos homogêneos, grupos heterogêneos e sequencial;
- b) uniforme, sem cadastro, maior variância por grupo, menor variância por grupo e parada endógena;
- c) uniforme, periodicidade, menor custo, maior custo e geométrica;

- d) hipergeométrica, cadastral, grupos homogêneos, grupos heterogêneos e geométrica;
- e) equiprováveis, periodicidade, grupos heterogêneos, grupos homogêneos e “por etapas”.

Comentários:

Na **amostragem aleatória simples** (AAS), a probabilidade de seleção de um elemento é a **mesma** para todos os elementos, ou seja, os elementos são **equiprováveis**. Na **amostragem sistemática**, os elementos são selecionadas em **períodos iguais**, ou seja, a seleção é periódica (**periodicidade**). Na **amostragem por estratos**, os grupos são compostos de elementos **homogêneos**; e na **amostragem por conglomerados ou cluster**, os grupos são compostos de elementos **heterogêneos**. Na **amostragem múltipla**, as amostras são extraídas em etapas **sequenciais**.

Gabarito: A

Resumo

Tipos de Amostragem Probabilística

Segue **métodos científicos**; probabilidade de qualquer elemento ser selecionado é **maior que zero**

- **Amostragem Aleatória Simples**: os elementos possuem a **mesma** probabilidade de seleção
- **Amostragem Sistemática**: elementos ordenados e selecionados **periodicamente**
- **Amostragem Estratificada**: divide população em **grupos compostos de elementos similares**; são selecionados **alguns elementos de cada grupo** (seguindo método probabilístico)
- **Amostragem por Conglomerados**: divide população em **grupos compostos de elementos distintos**; são selecionados **alguns grupos** (seguindo método probabilístico), dos quais **todos os elementos são analisados**
- **Amostragem Múltipla**: amostras extraídas em sequência; com base nos resultados de cada amostra, toma-se decisão de aceitar o lote, rejeitá-lo ou extrair nova amostra

TIPOS DE AMOSTRAGEM NÃO PROBABILÍSTICA

Amostragem por Conveniência

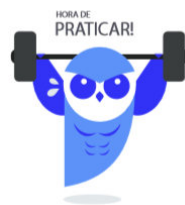
A amostragem por conveniência, também chamada de amostragem por acessibilidade, depende da **conveniência** do pesquisador, **não** havendo **critérios** científicos pré-definidos. Ela é utilizada devido a sua **praticidade**.

Um exemplo clássico desse método é a pesquisa em um local de grande circulação de pessoas, em que as pessoas selecionadas para a entrevista (ou seja, a amostra selecionada), são aquelas que passam perto do entrevistador e concordam com a entrevista.



Pelo senso comum, poderíamos dizer que a seleção de uma pessoa nessa situação (por um entrevistador em um local de grande circulação) seria uma seleção *aleatória*, no sentido de incerto, imprevisível.

Porém, como vimos anteriormente, uma amostragem exige que todos os elementos tenham a mesma chance de ser selecionado, o que não é o caso da amostragem por conveniência.



(VUNESP/2018 – Prefeitura de Serrana/SP) De modo geral, as pesquisas sociais abrangem um universo de elementos tão grande que se torna impossível considerá-los em sua totalidade. Por isso, nesse tipo de pesquisa, é muito frequente o trabalho com os diversos tipos de amostragem. A amostragem por acessibilidade ou por conveniência

- a) é uma variação da amostragem aleatória simples.
- b) pode ser proporcional ou não proporcional.
- c) é destituída de qualquer rigor estatístico.

d) é o procedimento básico de amostragem científica.

e) compara vários estratos de uma população.

Comentários:

A amostragem por acessibilidade ou conveniência é uma amostragem não probabilística, sem qualquer critério científico pré-definido.

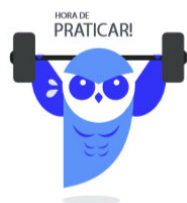
Gabarito: C

Amostragem por Julgamento

Na amostragem por julgamento, a **seleção** dos elementos para a amostra é feita com base nos **critérios definidos pelo julgamento do entrevistador**. Por exemplo, o entrevistador pode decidir selecionar pessoas com determinados aspectos visuais (como cor de pele ou altura), algum comportamento específico (pessoas que estejam se exercitando ou na frente de vitrine de alguma loja), frequência em determinados lugares (pessoas que estejam em determinada feira ou em algum parque), etc.

Amostragem por Cotas

Na amostragem por cotas, as amostras seguem as mesmas **proporções** da população, assim como a amostragem estratificada proporcional. Porém, a diferença é que neste último caso, utilizamos uma amostragem probabilística para selecionar os elementos de cada estrato, enquanto que para a amostragem por cotas, utiliza-se um método **não probabilístico**.



(FGV/2019 – DPE-RJ – Adaptada) Sobre os desenhos mais utilizados para a seleção da amostra e suas características, julgue a seguinte afirmativa:

A amostragem por cotas é uma forma de seleção estratificada, também apoiada em critérios probabilísticos.

Comentários:

Na amostragem por cotas, as amostras seguem as mesmas proporções da população, como na amostragem estratificada, porém, a amostragem por cotas **não** segue critérios **probabilísticos**.

Resposta: Errado.

(CESPE/2011 – FUB) Com relação às técnicas de amostragem de populações finitas, julgue o seguinte item.

As amostragens aleatórias simples, sistemática, estratificada e por cotas representam planos de amostragem probabilísticos.

Comentários:

As amostragens aleatória simples, sistemática e estratificada são probabilísticas, mas a amostragem por cotas é **não** probabilística.

Gabarito: Errado.

(2021 – Prefeitura de Porto Alegre/RS) As técnicas de amostragem podem ser classificadas como amostras probabilísticas e amostras não probabilísticas. São, respectivamente, uma técnica probabilística e uma técnica não probabilística:

- a) Amostra aleatória simples e amostra sistemática.
- b) Amostra por cotas e amostra por julgamento.
- c) Amostra por conglomerado (cluster) e amostra por cotas.
- d) Amostra aleatória simples e amostra por conglomerado (cluster).
- e) Amostra por cotas e amostra por conveniência.

Comentários:

O enunciado pede uma amostragem probabilística primeiro e uma não probabilística depois.

Em relação à alternativa A, tanto a amostra aleatória simples quanto a amostra sistemática são probabilísticas, logo, a alternativa A está incorreta.

Em relação à alternativa B, tanto a amostra por cotas quanto a amostra por julgamento são não probabilísticas, logo a alternativa B está errada.

Em relação à alternativa C, a amostra por conglomerado é probabilística e a amostra por cotas é não probabilística, logo a alternativa C está certa.

Em relação à alternativa D, tanto a amostra aleatória simples quanto a amostra por conglomerado são probabilísticas, logo a alternativa D está errada.

Em relação à alternativa E, tanto a amostra por cotas quanto a amostra por conveniência são não probabilística, logo a alternativa E está errada.

Gabarito: C

Amostragem por Tipicidade

Na **amostragem por tipicidade**, o pesquisador seleciona elementos que ele considera **representativos** dos subgrupos da população. Por exemplo, em uma população de colaboradores, o pesquisador pode selecionar um funcionário que recebe 1 salário mínimo, um funcionário que recebe 5 salários mínimos e um que recebe 10 salários mínimos, para representar as diferenças salariais dessa população. Logo, **é necessário que o pesquisador tenha algum conhecimento prévio da população para aplicar essa amostragem não probabilística.**

Amostragem por Voluntários

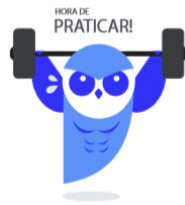
Na **amostragem por voluntários**, os próprios indivíduos da população se voluntariam para participar da pesquisa. Isso é comum nas pesquisas pelas redes sociais sobre questões como política, religião, etc. Os indivíduos não são contatados diretamente pelo pesquisador; eles decidem participar voluntariamente da pesquisa. Embora seja de fácil coleta e com baixo custo, o não há uma boa representatividade nesse tipo de amostragem.

Amostragem por Bola de Neve

Na **amostragem por bola de neve**, o pesquisador seleciona alguns indivíduos iniciais e esses indivíduos **convidam novos participantes**, dentre os seus conhecidos. Assim como uma bola de neve vai crescendo ao descer uma ladeira repleta de neve, essa amostra selecionada com base nessa técnica também cresce conforme os indivíduos selecionados convidam novos participantes.

Essa amostragem não probabilística é indicada para **populações de baixa incidência**, isto é, quando a característica que estiver sendo buscada for rara; ou de **difícil acesso**. Para estudar uma doença rara, por exemplo, pode ser difícil encontrar indivíduos selecionando-os aleatoriamente. Porém, ao selecionar alguns indivíduos com tal doença, é bem possível que esses conheçam outras pessoas com a mesma doença e possam convidá-los a participar do estudo.

Esse processo **apresenta baixo custo**, pois são os próprios participantes que selecionam a amostra. Porém, essa característica **leva a uma falta de controle** sobre as características e o tamanho da amostra. Assim, a amostra pode **não** ser uma boa **representação da população** como um todo.



(FGV/2017 – IBGE) Dentre os métodos de amostragem não probabilística, podem ser destacados os realizados por conveniência, por cotas, por julgamento, por tipicidade e as bolas de neve.

Sobre cada um dos métodos, e nessa exata ordem, poderiam ser associadas às seguintes palavras-chave ou expressões:

- a) praticidade, efeito de estratificação, arbitragem, para um subgrupo e indicações técnicas;
- b) proximidade, juízo de valor, para um subgrupo, avaliações em sequência e baixíssimo custo;
- c) baixo custo, arbitragem, seleção endógena, efeito cluster e para um subgrupo;
- d) seleção endógena, para um subgrupo, indicações técnicas, efeito de estratificação e efeito cluster;
- e) proximidade, avaliações em sequência, baixíssimo custo, efeito de estratificação e longa duração.

Comentários:

A amostragem por conveniência não segue critério específico, sendo utilizado por sua **praticidade**. Na amostragem por cotas, as amostras são proporcionais aos **estratos** da população. Na amostragem por julgamento, o pesquisador seleciona a amostra com base em seu julgamento, **arbítrio**. Na amostragem por tipicidade, o pesquisador seleciona elementos que ele entende representar os **subgrupos** da população. Na amostragem por bola de neve, a amostra é paulatinamente selecionada com base nas **indicações** dos participantes.

Logo, a alternativa A está correta.

Gabarito: A.

Resumo

Tipos de Amostragem Não Probabilística

Dependem de **escolhas** do pesquisador

- **Amostragem por conveniência**: a seleção da amostra **não** segue qualquer critério científico
- **Amostragem por julgamento**: o pesquisador **arbitra os critérios** para seleção da amostra
- **Amostragem por cotas**: o tamanho da amostra é **proporcional** aos estratos, mas a seleção da amostra não segue método probabilístico
- **Amostragem por tipicidade**: são selecionados elementos **representativos** das subpopulações
- **Amostragem por voluntários**: indivíduos decidem participar da pesquisa
- **Amostragem por bola de neve**: indivíduos selecionados **convidam** novos participantes