

Formas de amostragem

O sucesso de uma análise estatística envolve aspectos importantes sobre as formas de amostragem. Neste sentido, não basta que saibamos descrever convenientemente os dados de uma amostra e que dominemos perfeitamente as técnicas estatísticas para que possamos executar, com êxito, um trabalho estatístico completo. Antes de tudo, **é preciso garantir que a amostra ou amostras que serão usadas sejam obtidas por processos adequados**. Se erros grosseiros forem cometidos no momento de selecionar os elementos da amostra, o trabalho todo ficará comprometido e os resultados finais serão provavelmente bastante incorretos. Devemos, portanto, tomar especial cuidado quanto aos critérios que serão usados na seleção da amostra.

O que é necessário garantir, em suma, é que a amostra seja *representativa da população*. Isso significa que, a menos de certas pequenas discrepâncias inerentes à aleatoriedade sempre presente, em maior ou menor grau, no processo de amostragem, **a amostra deve possuir as mesmas características básicas da população, no que diz respeito à(s) variável(is) que desejamos pesquisar**.

A necessidade da representatividade da amostra não é, acreditamos, difícil de entender. O que talvez não seja tão fácil é saber quando temos uma amostra representativa ou não. Veremos adiante algumas recomendações sobre como proceder para garantir, da melhor forma possível, a representatividade da amostra.

Os problemas de amostragem podem ser mais ou menos complexos e sutis, dependendo das populações e das variáveis que se deseja estudar. Na indústria onde amostras são freqüentemente retiradas para efeito de controle da qualidade dos produtos e materiais, em geral os problemas de amostragem são mais simples de resolver. Por outro lado, em pesquisas sociais, econômicas ou de opinião, a complexidade dos problemas de amostragem é normalmente bastante grande.

Em tais casos, extremo cuidado deve ser tomado quanto à caracterização da população e ao processo usado para selecionar a amostra, a fim de evitar que os elementos desta constituam um conjunto com características fundamentalmente distintas das da população.

No caso de questionários serem distribuídos, muita atenção é também requerida em sua elaboração, visando evitar perguntas capciosas ou inibidoras, o que viria a distorcer os resultados.

Em resumo, a obtenção de soluções adequadas para o problema de amostragem exige, em geral, muito bom senso e experiência. Além disso, é muitas vezes conveniente que o trabalho do estatístico seja complementado pelo de um especialista do assunto em questão.

Amostragem probabilística

Distinguiremos dois tipos de amostragem: a **PROBABILÍSTICA** e a **NÃO-PROBABILÍSTICA**. **A amostragem será probabilística se todos os elementos da população tiverem probabilidade conhecida, e diferente de zero, de pertencer à amostra. Caso contrário, a amostragem será não-probabilística.**

Segundo essa definição, a amostragem probabilística implica um sorteio com regras bem determinadas, cuja realização só será possível se a população for finita e totalmente acessível.

Embora as técnicas de inferência estatística pressuponham que as amostras utilizadas sejam probabilísticas, muitas vezes não se pode conseguir. No entanto o bom-senso irá indicar quando o processo de amostragem, embora não sendo probabilístico, pode ser para efeitos práticos considerado como tal. Isso amplia consideravelmente as possibilidades de utilização do método estatístico em geral.

A utilização de uma amostragem probabilística é a melhor recomendação que se deve fazer no sentido de se garantir a representatividade da amostra, pois o acaso será o único responsável por eventuais discrepâncias entre população e amostra, o que é levado em consideração pelos métodos de análise da Estatística Indutiva (ou Inferência Estatística).

Damos a seguir algumas das principais técnicas de amostragem probabilística. Outras poderão também ser usadas, como combinação ou não das descritas.

1. Amostragem casual simples

Esse tipo de amostragem, também chamada *simples ao acaso*, *aleatória*, *casual*, *simples*, *elementar*, *randômica*, *etc.*, é equivalente a um sorteio lotérico. Nela, todos os elementos da população têm igual probabilidade de pertencer à amostra, e todas as possíveis amostras têm também igual probabilidade de ocorrer.

Sendo N o número de elementos da população e n o número de elementos da amostra, cada elemento da população tem probabilidade n/N de pertencer à amostra. A essa relação n/N denomina-se *fração de amostragem*. Por outro lado, sendo a amostragem feita sem reposição, a suposição que fazemos em geral, é que existem (n) possíveis amostras, todas igualmente prováveis.

Na prática, a amostragem simples ao acaso pode ser realizada numerando-se a população de 1 a N , sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, n números dessa seqüência, os quais corresponderão aos elementos sorteados para a amostra.

Um instrumento útil para realizar o sorteio acima descrito é a *tabela de números ao acaso*. Tal tabela é simplesmente constituída por inúmeros dígitos que foram obtidos por algum processo equivalente a um sorteio.

2. Amostragem sistemática

Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, temos uma *amostragem sistemática*. Assim, por exemplo, em uma linha de produção, podemos, a cada dez itens produzidos, retirar um para pertencer a uma amostra da produção diária.

Considere o exemplo: $N = 800$, $n = 50$ e a população já ordenada, poderemos adotar o seguinte procedimento: sortear um número de 1 a 16 (note-se que $800/50 = 16$), o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente retirados de 16 em 16. Equivalentemente, poder-se-iam considerar os números de 1 a 800 dispostos seqüencialmente em uma matriz com 50 linhas e 16 colunas, sorteando-se a seguir uma coluna, cujos números indicariam os elementos da amostra. Vemos que, nesse caso, cada elemento da população ainda teria probabilidade $50/800$ de pertencer à amostra, porém existem agora apenas 16 possíveis amostras.

A principal vantagem da amostragem sistemática está na grande facilidade na determinação dos elementos da amostra. O perigo em adota-la está na possibilidade da existência de ciclos de variação da variável de interesse, especialmente se o período desses ciclos coincidir com o período de retirada dos elementos da amostra. Por outro lado, se a ordem dos elementos na população não tiver qualquer relacionamento com a variável de interesse, então a amostragem sistemática terá efeitos equivalentes à causal simples, podendo ser utilizada sem restrições.

3. Amostragem por meio de conglomerados

Quando a população apresenta uma subdivisão em pequenos grupos, chamados conglomerados, é possível - e muitas vezes conveniente - fazer-se a *amostragem por meio desses conglomerados*, a qual consiste em sortear um número suficiente de conglomerados, cujos elementos constituirão a amostra. Ou seja, as unidades de amostragem, sobre as quais é feito o sorteio, passam a ser os conglomerados e não mais os elementos individuais da população. Esse tipo de amostragem é às vezes adotado por motivos de ordem prática e econômica.

4. Amostragem estratificada

Muitas vezes a população se divide em subpopulações ou estratos, sendo razoável supor que, de estrato para estrato, a variável de interesse apresente um comportamento substancialmente diverso, tendo, entretanto, comportamento razoavelmente homogêneo dentro de cada estrato. Em tais casos, se o sorteio dos elementos da amostra for realizado sem se levar em consideração a existência dos estratos, pode acontecer que os diversos estratos não sejam convenientemente representados na amostra, a qual seria mais influenciada pelas características da variável nos estratos mais favorecidos pelo sorteio. Evidentemente, a tendência da ocorrência de tal fato será tanto maior quanto menor o tamanho da amostra. Para evitar isso, pode-se adotar uma *amostragem estratificada*.

A amostragem estratificada consiste em especificar quantos elementos da amostra serão retirados em cada estrato. É costume considerar três tipos de amostragem estratificada: uniforme, proporcional e ótima. Na amostragem estratificada uniforme, sorteia-se igual número de elementos em cada estrato. Na proporcional, o número de elementos sorteados em cada estrato é proporcional ao número de elementos existentes no estrato. Evidentemente, a amostragem estratificada uniforme será, em geral, recomendável se os estratos da população forem pelo menos

aproximadamente do mesmo tamanho; caso contrário, será em geral preferível a estratificação proporcional, por fornecer uma amostra mais representativa da população.

A amostragem estratificada ótima, por sua vez, toma, em cada estrato, um número de elementos proporcional ao número de elementos do estrato e também à variação da variável de interesse no estrato, medida pelo seu desvio-padrão. Pretende-se assim otimizar a obtenção de informações sobre a população, com base no princípio de que, onde a variação é menor, menos elementos são necessários para bem caracterizar o comportamento da variável. Dessa forma, com um menor número total de elementos na amostra, conseguir-se-ia uma quantidade de informação equivalente à obtida nos demais casos. As principais dificuldades para a utilização desse tipo de amostragem residem nas complicações teóricas relacionadas com a análise dos dados e em que, muitas vezes, não podemos avaliar de antemão o desvio-padrão da variável nos diversos estratos.

Constituem exemplos em que uma amostragem estratificada parece ser recomendável a estratificação de uma cidade em bairros, quando se deseja investigar alguma variável relacionada à renda familiar; a estratificação de uma população humana em homens e mulheres, ou por faixas etárias; a estratificação de uma população de estudantes conforme suas especializações, etc.

5. Amostragem múltipla

Numa *amostragem múltipla*, a amostra é retirada em diversas etapas sucessivas. Dependendo dos resultados observados, etapas suplementares podem ser dispensadas. Esse tipo de amostragem é, muitas vezes, empregado na inspeção por amostragem, sendo particularmente importante a amostragem dupla. Sua finalidade é diminuir o número médio de itens inspecionados a longo prazo, baixando assim o custo da inspeção.

Um caso extremo de amostragem múltipla é a *amostragem seqüencial*. A amostra vai sendo acrescida item por item, até se chegar a uma conclusão no sentido de se aceitar ou rejeitar uma dada hipótese. Com a amostragem seqüencial pretende-se tornar mínimo o número médio de itens inspecionados a longo prazo.

Amostragem não-probabilística

Amostras não-probabilísticas são também, muitas vezes, empregadas em trabalhos estatísticos, por simplicidade ou por impossibilidade de se obterem amostras probabilísticas, como seria desejável. Como em muitos casos os efeitos da utilização de uma amostragem não-probabilística podem ser considerados equivalentes aos de uma amostragem probabilística, resulta que os processos não-probabilísticos de amostragem têm também sua importância. Apresentamos a seguir alguns casos de amostragem não-probabilística.

1. Inacessibilidade a toda a população

Essa situação ocorre com muita freqüência na prática. Somos então forçados a coletar a amostra na parte da população que nos é acessível. Surge aqui, portanto, uma distinção entre *população-objeto* e *população amostrada*. A população-objeto é aquela que temos em mente ao realizar o trabalho estatístico. Apenas uma parte dessa população, porém, está acessível para que dela retiremos a amostra. Essa parte é a população amostrada.

Se as características da variável de interesse forem as mesmas na população-objeto e na população amostrada, então esse tipo de amostragem equivalente a uma amostragem probabilística.

Uma situação muito comum em que ficamos diante da inacessibilidade a toda a população é o caso em que parte da população não tem existência real, ou seja, uma parte da população é ainda hipotética. Assim, por exemplo, seja a população que nos interessa constituída por todas as peças produzidas por certa máquina. Ora, mesmo estando a máquina em funcionamento normal, existe uma parte da população que é formada pelas peças que ainda vão ser produzidas. Ou, então, se nos interessar a população de todos os portadores de febre tifóide, estaremos diante de um caso semelhante. Deve-se notar que, em geral, estudos realizados com base nos elementos da população amostrada terão, na verdade, seu interesse de aplicação voltado para os elementos restantes da população-objeto. Esse fato realça a importância de se estar convencido de que as duas populações podem ser consideradas como tendo as mesmas características.

O presente caso de amostragem não-probabilística pode ocorrer também quando, embora se tenha a possibilidade de atingir toda a população, retiramos a amostra de uma parte que seja prontamente acessível. Assim, se fôssemos colher uma amostra de um monte de minério poderíamos, por simplificação, retirá-la de uma camada próxima à superfície exterior do monte, pois o acesso às porções interiores seria problemático.

2. Amostragem a esmo ou sem norma

A amostragem em que o amostrador, para simplificar o processo procura ser aleatório sem, no entanto, realizar propriamente o sorteio usando algum dispositivo aleatório confiável. Por exemplo, se desejarmos retirar uma amostra de 100 parafusos de uma caixa contendo 10.000, evidentemente não faremos uma amostragem casual simples, pois seria extremamente trabalhosa, mas procederemos à retirada simplesmente a esmo.

Os resultados da amostragem a esmo serão, em geral, equivalentes aos de uma amostragem probabilística se a população é homogênea e se não existe a possibilidade de o amostrador ser inconscientemente influenciado por alguma característica dos elementos da população.

3. A população é formada por material contínuo

Nesse caso é impossível realizar amostragem probabilística, devido à impraticabilidade de um sorteio rigoroso. Se a população for líquida ou gasosa, o que se costuma fazer, com resultado satisfatório, é homogeneizá-la e retirar a amostra a esmo. Tal procedimento pode às vezes, também, ser usado no caso de material sólido.

Outro procedimento que pode ser empregado nesses casos, especialmente quando a homogeneização não é praticável, é a *enquartação*, a qual consiste em subdividir a população em diversas partes (a origem do nome pressupõe a divisão em quatro partes), sorteando-se uma ou mais delas para constituir a amostra ou para delas retirar a amostra.

4. Amostragens intencionais

Enquadram-se aqui os diversos casos em que o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população.

O perigo desse tipo de amostragem é obviamente grande, pois o amostrador pode facilmente se equivocar em seu pré-julgamento. Apesar disso, o uso de amostragens intencionais, ou parcialmente intencionais, é bastante freqüente, ocorrendo em vários tipos de situações reais que poderíamos tentar identificar e classificar. Não o faremos, porém, por fugir à nossa finalidade neste texto.

Delineando a amostra aleatória simples

Nesta seção iremos nos preocupar com o processo de seleção de uma amostra aleatória simples. Embora não seja necessariamente o procedimento de amostragem probabilística mais econômico ou eficiente, tem sido a base de evolução dos procedimentos mais sofisticados.

A chave para uma seleção de amostra adequada é obter e manter uma lista atualizada de todos os indivíduos ou itens a partir dos quais a amostra será extraída. Tal lista é conhecida como a **estrutura de população (universo)** ou simplesmente **população**. A lista da população irá servir como **população-alvo**, de modo que se muitas amostras probabilísticas diferentes forem retiradas desta lista, é de se esperar que cada amostra seja a **representação** em miniatura da população e que ela produza estimativas razoáveis de suas características. Se a lista for inadequada porque certos grupos de indivíduos ou itens na população não foram incluídos apropriadamente, as amostras de probabilidade aleatórias irão apenas oferecer estimativas das características da população-alvo (e não da *verdadeira* população), os resultados sofrerão tendenciosidade.

Dois métodos básicos poderiam ser utilizados para selecionar a amostra: a amostra poderia ser obtida **com reposição ou sem reposição** da população finita. O método aplicado deve ser claramente estabelecido, uma vez que várias formas utilizadas posteriormente com o propósito de inferência estatística são dependentes do método de seleção.

Sejam N o tamanho da população e n o tamanho da amostra. Para retirar uma amostra aleatória simples de tamanho n , poderíamos supostamente registrar os nomes dos N membros em cartões separados individualmente, do mesmo tamanho, colocar estes cartões numa grande cesta, misturá-los bem e então aleatoriamente selecionar os n sujeitos da amostra de dentro da cesta.

Quando se realiza amostragem com reposição, a chance de que um membro em particular da população seja selecionado na primeira retirada da cesta é de $1/N$. Independentemente de quem seja selecionado na primeira retirada, informações pertinentes são registradas num arquivo principal, e então este cartão específico é recolocado na cesta (amostragem com reposição). Os N cartões na cesta são então bem embaralhados e o segundo cartão é retirado. Uma vez recolocado o primeiro cartão, a chance de seleção de cada membro em particular na segunda retirada incluindo

aquele primeiro indivíduo, independentemente de ele ter sido previamente selecionado é ainda $1/N$. Novamente, as informações pertinentes são registradas num arquivo principal e o cartão é recolocado para se preparar para a terceira retirada. Tal processo é repetido até que n , o tamanho desejado da amostra, seja alcançado. Assim, ao realizar amostragem com reposição, cada indivíduo ou item, em cada retirada, terá sempre a mesma chance $1/N$ de ser selecionado.

No entanto, será que desejaríamos ter o mesmo indivíduo ou item possivelmente selecionado mais de uma vez? Ao compor amostras com seres humanos, é em geral mais apropriado ter uma amostra com pessoas diferentes do que permitir medições repetidas da mesma pessoa. Assim sendo, empregariamos o método de amostragem sem reposição, de modo que, uma vez retirado determinado indivíduo, o mesmo não poderia ser selecionado novamente. Como antes, ao fazer amostras sem reposição a chance de que qualquer membro da população seja selecionado na primeira retirada da cesta é de $1/N$. Quem quer que seja selecionado, as informações pertinentes são registradas em um arquivo principal, e então o cartão específico é colocado de lado, em vez de recolocado na cesta (amostragem *sem reposição*). Os $N-1$ cartões remanescentes na cesta são então bem embaralhados e o segundo cartão é retirado. A chance de qualquer indivíduo não selecionado previamente ser escolhido na segunda retirada é agora de 1 sobre $N-1$. Este processo de selecionar um cartão, registrar a informação em um arquivo principal, embaralhar os cartões remanescentes e então retirar novamente continua até que seja obtida a amostra desejada de tamanho n .

Cuidados com a pesquisa

1. **Sub abrangência:** Quando alguns grupos da população são deixados fora do processo de escolha da amostra
2. **Não resposta:** não há meios de se obter a observação da característica de interesse. Quando o indivíduo escolhido para integrar a amostra não pode ser contatado ou se recusa a cooperar
3. **Questões mal formuladas:** questões confusas ou dirigidas podem introduzir viés (tendência) aos resultados, e mesmo, pequenas modificações no fraseado podem mudar o resultado da pesquisa
4. **Erro amostral:** é a diferença entre o resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais aleatórias. Os erros amostrais ocorrem única e exclusivamente em função do número de elementos da amostra e do processo de seleção desses elementos.
5. **Erros não amostrais:** são os cometidos durante o processo de pesquisa de marketing que não sejam oriundos do tamanho e do processo de seleção da amostra. Não são mensuráveis e tendem a crescer com o aumento do tamanho da amostra.
Ex. 1: escolha de amostra não aleatória e tendenciosidade
Ex. 2: utilização de instrumentos de mensuração defeituoso
Ex. 3: questão formulada de modo tendencioso
Ex. 4: não respostas
Ex. 5: definição errada do problema de pesquisa
Ex. 6: entrevistadores com nível inadequado para a pesquisa
Ex. 7: definição errada da população de pesquisa

Etapas envolvidas no processo de amostragem:

1. Identificação da população-alvo
2. Determinação da estrutura da amostra
3. Reconciliação da população, solução de problemas quanto às diferenças nas estruturas de amostragem
4. Seleção do procedimento de amostragem
5. Determinação do tamanho relevante da amostra
6. Obtenção de informações sobre os respondentes
7. Trato com o problema de não resposta
8. Geração de informação para propósito de tomada de decisão

População alvo:

- A definição da população alvo deve conter informações sobre elementos da amostragem, unidades de amostra e área de cobertura.
- Objetivos da pesquisa
- Unidades de amostragem
- Critérios de exclusão
- Restrições excessivas.

Razões para o uso de amostragem não probabilística

1. Não existir outra alternativa (não fazer a pesquisa ou fazer com aqueles que se dispõem a colaborar?)
2. Problemas que ocorrem na prática podem enfraquecer a superioridade de uma amostra probabilística tornando-a não probabilística
3. Não há intenção de generalizar os dados obtidos na amostra para a população (representatividade pode ser irrelevante)
4. Falta de disponibilidade de recursos financeiros, materiais e humanos para realização de uma amostragem probabilística.

Bibliografia:

NETO, Pedro L. C. **Estatística**. São Paulo: Ed. Blucher Ltda, 1977.

MOORE, D. S. **Estatística básica e sua prática**. Rio de Janeiro: Livros Técnicos e Científicos, 3. Ed., 2004.

BUSSAB, W.; BOLFARINE, H. **Elementos de amostragem**, São Paulo: Edgar Blucher, 2005.