

## Amostragem

Como podemos determinar quantas pessoas em uma população apresentam certa característica? Por exemplo, quantos eleitores apóiam um candidato à presidência? Ou então, da população de determinado estado, quantas pessoas são crianças, quantas vivem em centros urbanos, quantas estão desempregadas?

Uma forma de responder a essas questões consiste em entrevistar todas as pessoas. Mas este é um processo demorado e caro.

Outro processo possível consiste então em consultar um grupo de pessoas, que constituem uma amostra. Se a amostra representa de fato toda a população, podemos utilizar as características dos seus elementos para estimar as características de toda a população.

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.

*Exemplo:* Numa empresa deseja-se escolher 3 diretores entre seus chefes executivos. A escolha é aleatória e não depende do prestígio, da capacidade, dos anos de serviço, etc. Temos uma amostragem probabilística.

As técnicas da estatística pressupõem que as amostras utilizadas sejam probabilísticas, o que 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.

Apresentamos a seguir algumas técnicas de amostragem probabilística.

### ▪ Amostragem por conglomerado

A população é dividida em diferentes conglomerados (grupos), extraído-se uma amostra apenas dos conglomerados selecionados, e não de toda a população. O ideal seria que cada conglomerado representasse tanto quanto possível o total da população. Na prática, selecionam-se os conglomerados geograficamente. Escolhem-se aleatoriamente algumas regiões, em seguida algumas sub-regiões e finalmente, alguns lares. Esse processo possibilita ao pesquisador entrevistar apenas poucas pessoas.

### ▪ Amostragem Estratificada

Se a população pode ser dividida em subgrupos que consistem, todos eles, em indivíduos bastante semelhantes entre si, pode-se obter uma amostra aleatória de pessoas em cada grupo. Esse processo pode gerar amostras bastante precisas, mas só é viável quando a população pode ser dividida em grupos homogêneos.

### ▪ **Amostragem Aleatória Simples**

A amostragem aleatória simples é a maneira mais fácil para selecionarmos uma amostra probabilística de uma população. Começamos introduzindo o conceito de AAS de uma população finita, para a qual temos uma listagem de todas as unidades elementares. Podemos obter uma amostra nessas condições, escrevendo cada elemento num cartão, misturando-os numa urna e sorteando tantos cartões quantos desejarmos na amostra. Esse procedimento torna-se inviável quando a população é muito grande. Nesse caso, usa-se um processo alternativo, no qual os elementos são numerados e em seguida sorteados por meio de uma tabela de números aleatórios.

Utilizando-se um procedimento aleatório, sorteia-se um elemento da população, sendo que todos os elementos têm a mesma probabilidade de ser selecionados. Repete-se o procedimento até que sejam sorteadas as unidades da amostra.

Podemos ter uma AAS com reposição, se for permitido que uma unidade possa ser sorteada mais de uma vez, e sem reposição, se a unidade sorteada for removida da população.

Do ponto de vista da quantidade de informação contida na amostra, amostrar sem reposição é mais adequado. Contudo, a amostragem com reposição conduz a um tratamento teórico mais simples, pois ela implica que tenhamos independência entre as unidades selecionadas. Essa independência facilita o desenvolvimento das propriedades dos estimadores que serão considerados.

Se a população for infinita então as retiradas com e sem reposição serão equivalentes, isto é, se a população for infinita (ou então muito grande), o fato de se recolocar o elemento retirado de volta na população não vai afetar em nada a probabilidade de extração do elemento seguinte.

Se, no entanto, a população for finita (e pequena) será necessário fazer uma distinção entre os dois procedimentos, pois na extração com reposição as diversas retiradas serão independentes, mas no processo sem reposição haverá dependência entre as retiradas, isto é, o fato de não recolocar o elemento retirado afeta a probabilidade do elemento seguinte ser retirado. A amostragem sem reposição é mais eficiente que a amostragem com reposição e reduz a variabilidade uma vez que não é possível retirar elementos extremos mais do que uma vez.

### ▪ **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.

Amostras não-probabilísticas são também, muitas vezes, empregados em trabalhos estatísticos, por simplicidade ou por impossibilidade de se obterem amostras probabilísticas, como seria desejável. No entanto processos não-probabilísticos de amostragem têm também sua importância. Sua utilização, entretanto, deve ser feita com cuidado.

Apresentamos a seguir algumas técnicas de amostragem não-probabilística.

### ▪ **Inacessibilidade a toda população**

Esta situação ocorre com muita frequência na prática. Por exemplo, seja a população que nos interessa constituída de 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.

Este 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 recolher uma amostra de um monte de minério, poderíamos por simplificação retirar a amostra de uma camada próxima da superfície do monte, pois o acesso as porções interiores seria problemático.

- **Amostragem a esmo**

É 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 AAS, pois seria muito trabalhosa, mas retiramos simplesmente a esmo.

Os resultados da amostragem a esmo são, em geral, equivalentes aos da 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.

- **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. O perigo desse tipo de amostragem é grande, pois o amostrador pode facilmente se enganar em seu pré-julgamento.

- **Amostragem por voluntários**

Ocorre, por exemplo, no caso da aplicação experimental de uma nova droga em pacientes, quando a ética obriga que haja concordância dos escolhidos.

## **Distribuições Amostrais**

O conceito de distribuição de probabilidade de uma variável aleatória será agora utilizado para caracterizar a distribuição dos diversos valores de uma variável em uma população.

Ao retirar uma amostra aleatória de uma população estaremos considerando cada valor da amostra como um valor de uma variável aleatória cuja distribuição de probabilidade é a mesma da população no instante da retirada desse elemento para a amostra.

Em consequência do fato de os valores da amostra serem aleatórios, decorre que qualquer quantidade calculada em função dos elementos da amostra também será uma variável aleatória.

*Parâmetros* – são valores teóricos correspondentes a população.

*Estatísticas* – são funções dos valores amostrais.

As estatísticas, sendo variáveis aleatórias, terão alguma distribuição de probabilidade, com uma média, variância, etc. A distribuição de probabilidade de uma estatística chama-se comumente distribuição amostral ou distribuição por amostragem.

## ESTIMAÇÃO

A inferência estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de amostra. Um dos itens básicos nesse processo é a estimação de parâmetros. A estimação pode ser por ponto ou por intervalo.

- ❖ *Estimativa por Ponto*: é a estimativa de um parâmetro populacional por um único valor.
- ❖ *Estimativa por Intervalo*: consiste em um intervalo em torno da estimativa por ponto de tal forma que ele possua probabilidade conhecida (nível de confiança  $(1-\alpha)$ ) de conter o verdadeiro valor do parâmetro. Este intervalo é conhecido por intervalo de confiança (IC).

### Intervalo de confiança para a média $\mu$ de uma população

Os intervalos de confiança para a média são tipicamente construídos com o estimador  $\bar{X}$  no centro do intervalo.

#### 1- Quando $\sigma$ é conhecido:

Quando o uso da distribuição normal está garantido, o intervalo de confiança para a média é determinado por:

$$\mathbf{IC = ( \bar{x} - z \frac{\sigma}{\sqrt{n}} ; \bar{x} + z \frac{\sigma}{\sqrt{n}} ) \text{ ou}}$$
$$\mathbf{IC = ( \bar{x} - z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} ; \bar{x} + z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} )}$$

**no caso de população finita de tamanho N e amostragem sem reposição.**

Os intervalos de confiança mais freqüentemente utilizados são os de 90%, 95% e 99%.

<b>z</b>	<b>(1-<math>\alpha</math>)</b>
1,65	0,90
1,96	0,95
2,58	0,99