#867 #DesenvolveDados Seg • Qua • Sex

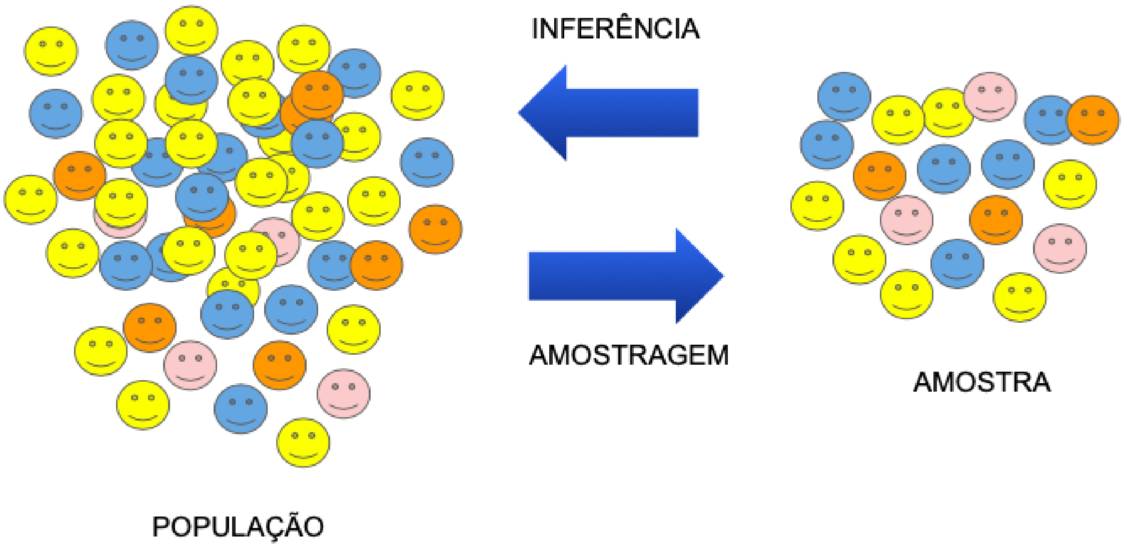
Estatística I

Conteúdo

5 Amostragem

Amostragem

1. Noções Básicas



Geralmente quando se prepara um macarrão, uma unidade desse é retirada para saber se o ponto de cozimento é o desejado. Quando um médico deseja identificar se um paciente está doente, alguns ml de sangue são retirados para análise. Note que nos dois casos, não seria conveniente analisar o todo, para chegar a uma conclusão satisfatória.

Em Estátistica, este procedimento de tirar uma parte do todo para validar alguma suposição é chamada de amostragem. Em outras palavras, o procedimento amostral visa obter informações sobre o todo baseando-se no resultado de uma amostra.

1.1 Definições

População: ou Universo é o conjunto de todas as unidades elementares de interesse. A população deve ser definida claramente e em termos da informação que se pretende conhecer.

Unidade: trata-se de qualquer elemento da populaçao.

Amostra: uma parte ou subconjunto da população

Censo: observação de todos os elementos da população.

Parâmetro Populacional: é o vetor correspondente a todos os valores de uma variável de interesse. Pode ser qualitativa (gosto musical, opnião sobre o governo, etc) ou quantitativa (média, proporção, quantidade, etc).

Função Paramétrica Populacional: é uma característica numérica da população, ou seja, uma expressão numérica que condensa os valores do vetor de parâmetro populacional. Por exemplo, média, total, proporção, dentre outros.

Exemplo 1: Considere uma população formada por 4 alunos de uma escola. Com as seguintes caracteristicas:

Variável	Valores			
Aluno	1	2	3	4
Nome	Ana	João	Lucas	Francisco
Idade	8	7	8	12
Sexo	F	M	M	M

Neste exemplo, cada aluno é um elemento da população. Com relação à amostragem os subconjuntos (Ana, João), (Francisco, Ana), (João) são **exemplos de amostra**. **Parâmetros populacionais**: idade = (8,7,8,12) e sexo = (F,M,M,M). Com relação às **funções paramétricas**, poderíamos definir:

- Idade média: fazendo idade = Y:

$$\mu = \overline{Y} = \frac{\sum_{i=1}^{4} Y_i}{4} = \frac{8+7+8+12}{4} = 8,75$$

- Idade máxima: max(Y)= max(8,7,8,12) = 12

- Porporção de meninas: sexo = Y = (F,M,M,M)

$$p(F) = \frac{1}{4} = 0,25$$

1.2 Tipos de Amostragem

- **Amostra probabilística**: todos os elementos da população apresentam probabilidade maior que zero de serem selecionados \[C_{2}^{6}= \frac{654!}{2 * 4!}=15\[

- **Amostra não probabilística**: quando não há probabilidade clara/conhecida de seleção dos elementos. Os elementos são escolhidos de forma julgamental.

2. Métodos de Amostragem

Neste material abordaremos apenas os métodos relacionados à amostragem probabilística, com o objetico de obter uma **amostra representativa**. Uma amostra é considerada representativa quando consegue **refletir as caracteristicas da população**.

2.1 Amostra Aleatória Simples

Este é o método mais simples e mais importante de seleção de uma amostra, pois pode ser usada em combinação com outros métodos. A premissa assumida é que a população é homogênea com relação à característica de interesse.

A amostra aleatória simples pode ser realizada com ou sem reposição. No caso em que há reposição, cada elemento pode ser sorteado mais de uma vez. Para exemplificar, suponha que se queira sortear um número aleatório de uma urna, se for uma AAS com preposição, este número voltará para urna para participar do próximo sorteio. Se não houver reposição, cada elemento só poderá ser selecionado uma vez para compor a amostra.

Considere uma população formada por N elementos (conhecido e finito). Este método consiste em selecionar n elementos, sendo que cada elemento tem a mesma probabilidade de ser selecionado

```
N = list(range(1,21,1))
print("População: ", N, "\n")

n = random.sample(N,5)
print("Amostra: ",n)
```

2.1.1 Estimadores

Temos os parâmetros da população:

- **Total** :

$$\sum_{j=1}^{N} Y_j$$

- **média** :

$$\frac{\sum_{j=1}^{N} Y_j}{N}$$

2.2 Amostra Sistemática

Usada quando os elementos população estão ordenados (população de lista telefônica, casas em uma rua).

Considere uma população de tamanho N e que se queira uma amostra de tamanho n. O processo de amostragem deste método consiste em:

- Dividir o tamanho populacional em K partes:

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

- Definir a posição de início da amostragem (que também será o primeiro elemento da amostra). Para tal fim, é sorteado i com o uso da amostra aleatória simples no intervalo, em que \(i \in [1, k]\[
- A partir do elemento selecionado aleatoriamente, é realizada sucessão aritimética para selecionar os $n-1$ indivíduos restantes

$$i, i+k, i+2k, i+3k, \ldots, i+(n-1)k$$

2.3 Amostra Estratificada

Trata-se do método em que a população é dividida em grupos (estratos) segundo alguma(s) característica(s) conhecida(s) na população sob estudo. São exemplos de estrato o gênero, faixa etária, região geográfica, profiissão. No geral, é usada quanto a população é heterogênea sob a ótica das características analisadas. Procedimento de amostragem:

- Dividir as N unidades da população em $N_1, N_2, \cdots, N_j$ estratos distintos e homogêneos
- Selecionar, ao acaso, uma amostra de tamanhos $n_1, n_2, \cdots, n_j$, de modo que o tamanho da amostra seja $n = n_1 + n_2 + \cdots + n_j$. O tamanho amostral pode ser proporcional à representatividade do estrato

3. Tamanho Amostral

Ao se realizar uma amostra para inferir uma determinada função paramétrica (média, máximo ou outra função de um parâmetro), há um erro associado ao planejamento amostral. A medida que o tamanho da amostra aumenta, o erro do estimador decresce. Vale ressaltar que uma amostra muito grande pode implicar em custos desnecessários, enquanto que uma amostra pequena pode tornar a pesquisa inconclusiva. Deste modo, o ponto chave de um levantamento amostral é determinar o tamanho da amostra.

3.2 Cálculo do tamanho amostral baseado na estimativa da média populacional

3.2.1 Populaçao Infinita

Uma população é considerada infinita quando seu tamanho é muito grande.

Ao realizar o calculo do tamanho da amostra n, deve-se levar em consideração o erro ϵ máximo que deseja-se assumir (ao estimar a função parâmetrica) e o nível de confiança do resultado (probabilidade). Sendo assim, o problema consiste em determinar **n** de forma que:

$$P(\mid \overline{X} - \mu \mid \leq \epsilon) \simeq 1 - \alpha$$

Mas pelo Teorema Central do Limite, a equação acima pode ser reescrita como:

$$P\left(\mid \overline{X} - \mu \mid \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \simeq 1 - \alpha$$

Sendo assim, dados um erro máximo e nível de confiança, calcular o tamanho amostral consiste em:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \epsilon \Longrightarrow n = \left(\frac{z_{\alpha/2}\sigma}{\epsilon}\right)^2$$

3.2.2 Populaçao Finita

No caso em que o tamanho populacional não é tão grande, a consideramos finita. Caso a amostra tenha um tamanho n maior ou igual a 5% do tamanho da população N, considera-se que a população é **finita**. Neste caso, aplica-se um fator de correção à fórmula vista anteriormente:

$$n = \frac{N(z_{\alpha/2}\sigma)^2}{(N-1)\epsilon^2 + (z_{\alpha/2}\sigma)^2}$$

3.2.4 Variância populacional desconhecida

No caso em que a variância populacional é desconhecida, pode-se realizar uma amostragem aleatória preliminar (ao menos 30 elementos) para estimar a variancia amostral e usa-la na equaçao acima.

$$\widehat{\sigma^2} = s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{N-1}$$

3.3 Cálculo do tamanho amostral baseado na estimativa da proporção populacional

3.3.1 População infinita

De maneira geral, em muitas situações, existe interesse em estudar a proporção de elementos em certa população que possuem determinada característica, como ser ou não um item defeituoso, ser ou não eleitor de determinado partido político e assim por diante. Neste caso a média populacional equivale à proporção (percentual) P de elementos com a característica e a variância populacional é dada por:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - P)^2 = P(1-P)$$

Sendo assim, o cálculo do tamanho amostral é dado por:

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{\epsilon^2}$$

3.3.2 População finita

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{(N-1)\epsilon^2 + z_{\alpha/2}^2 P(1-P)}$$

3.3.3 Variância populacional desconhecida

Caso a variância popupacional seja desconhecida, o tamanho da amostra pode ser calculada considerando o pior caso. Observe no grácico abaixo que o pior caso acontece quando $P = 0,5$, pois se tem a maior variância.

‹ Tópico anterior

Próximo Tópico ›