DGI
2019

7.1. População, Amostra e Inferência Estatística

O objectivo fundamental da estatística é inferir as características de uma população, a partir de resultados observados numa ou mais amostras extraídas dessa população.

- ***O processo de obtenção/extracção de amostras é designado de amostragem.***

Exemplo

Tirar conclusões sobre as alturas (ou pesos) dos estudantes do IPB *(população)* observando apenas um grupo de 100 estudantes *(amostra)* seleccionados na população.

7.2. Amostragem Aleatória

Para se poderem tirar ***conclusões representativas*** acerca de uma população a partir da análise de uma amostra, a ***recolha da amostra a utilizar deve ser feita de acordo com certas regras***.

Considere-se uma população constituída por M elementos (ex. alunos do IPB) e a variável aleatória X que representa uma característica da população a estudar (ex. altura).

- Para obter uma amostra de tamanho $\boldsymbol{n}$ $(n < M)$ deve escolher-se aleatoriamente um indivíduo da população. Dessa forma, esse indivíduo (altura) pode tomar qualquer valor entre todas as alturas possíveis.

x_1 – altura do 1º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_1
x_2 – altura do 2º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_2
x_3 – altura do 3º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_3
$\vdots$ $\vdots$
x_n – altura do n indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_n

- ***O processo de amostragem diz-se aleatório*** (e as amostras aleatórias) ***quando as funções de probabilidade de X_1, X_2, ..., X_n*** *são idênticas à função* $p(x)$:

$$\forall_X : p(x_1) = p(x_2) = \ldots = p(x_n) = p(x)$$

- A justificação básica para para se utilizar a ***amostragem aleatória*** é assegurar que *as* ***inferências*** feitas a partir dos dados amostrais ***não são distorcidas por um enviezamento na selecção***. Tal enviezamento existirá sempre que ocorrer uma tendência sistemática para ***sobrerepresentar ou subrepresentar*** uma parte da população.

Como obter uma amostra aleatória de dimensão n?

(1) Numerar de 1 até M todos os elementos da população;

(2) Colocar bolas numeradas de 1 até M numa urna;

(3) Retirar uma bola, seleccionar o elemento da população com o número correspondente e registar a sua altura;

(4) Repor a bola na urna;

(5) Repetir os passos (3) e (4) até terem sido registadas as alturas de n elementos.

- Quando a *amostragem é feita com reposição as variáveis* X_1, X_2, ..., X_n *são independentes.*

- Um processo de amostragem aleatória com estas características diz-se *aleatório simples*, verificando-se:

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = p(x_1) \times p(x_2) \times \ldots \times p(x_n)$$

- Quando $M \to \infty$ (e a dimensão da amostra, n, se mantém finita), a dependência entre as variáveis X_1, X_2, ..., X_n tende a desaparecer. ***Quando a população é infinita, é indiferente realizar a amostragem aleatória com ou sem reposição*** (a amostragem é aleatória simples em qualquer dos casos).

DGI
2019

7.3. Distribuições Amostrais

parâmetros populacionais $\neq$ estatísticas amostrais

↓

Para uma dada população e uma variável aleatória sobre ela definida, ***os parâmetros da distribuição correspondente são fixos enquanto que as medidas estatísticas variam de amostra para amostra***.

- Uma ***estatística amostral*** calculada a partir de X_1, X_2, ..., X_n é uma função de várias variáveis aleatórias e é, portanto, ***ela própria uma variável aleatória***. ***A distribuição de probabilidade de uma estatística amostral designa-se distribuição amostral da estatística***.

- Naturalmente, podemos calcular, para uma distribuição amostral, a média, a variância, desvio padrão, etc.

- ***O conceito de distribuição amostral de estatísticas será abordado considerando apenas o caso de estas serem obtidas com base em amostragens aleatórias.***

7.3.1. Distribuição da Média Amostral

Valor esperado	$E\left(\bar{X}\right) = \mu_{\bar{X}} = \mu_X$	
Variância	$VAR\left(\bar{X}\right) = \sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n}$	(população infinita ou finita com reposição)
	$VAR\left(\bar{X}\right) = \sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n} \times \left(\dfrac{M-n}{M-1}\right)$	(população finita, sem reposição)

Justificação:

$$E\left(\bar{X}\right) = E\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times E\left(X_1 + X_2 + \ldots + X_n\right) =$$

$$= \frac{1}{n} \times \left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right] = \frac{1}{n} \times \sum_{i=1}^{n} E\left(X\right) = \frac{1}{n} \times \left[n \times E\left(X\right)\right] = E\left(X\right)$$

$$Var\left(\bar{X}\right) = Var\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times Var\left(X_1 + X_2 + \ldots + X_n\right) =$$

$$= \frac{1}{n^2} \times \left[Var\left(X_1\right) + Var\left(X_2\right) + \ldots + Var\left(X_n\right)\right] =$$

$$= \frac{1}{n^2} \times \sum_{i=1}^{n} Var\left(X\right) = \frac{1}{n^2} \times \left[n \times Var\left(X\right)\right] = \frac{Var\left(X\right)}{n}$$ (população infinita ou finita com reposição)

Exemplo

Considere-se uma população com quatro elementos aos quais se associam os seguintes valores da variável aleatória X: $\{2,\ 4,\ 6,\ 6\}$

(a) Qual a função de probabilidade $p(x)$?

(b) Defina a função de probabilidade de $\bar{X}$ e a média amostral, calculadas para amostras aleatórias de dimensão 2 (amostras sem reposição).

Forma da Distribuição da Média Amostral quando a População é Normal

Quando a população é normal, a distribuição da média amostral também é normal.

(1) A média amostral é uma combinação linear das variáveis X_i.

(2) Se a população for normal, então

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

é uma combinação linear de variáveis normais.

(3) Qualquer combinação linear de variáveis normais é, ela própria, uma variável normal.

- Sendo a *população normal*, presume, em rigor, que a *população é infinita*. Nesta situação:

$$\bar{X} \rightarrow N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$ *população infinita*

- Se se admitir que a *população é finita mas de grande dimensão*, com uma *distribuição aproximadamente normal*, e que a *amostra é também de grande dimensão*, então na definição de $\bar{X}$ deve-se entrar em *consideração com o factor de redução*, vindo:

$$\bar{X} \rightarrow N\left(\mu_X, \frac{\sigma_X^2}{n} \times \frac{M-n}{M-1}\right)$$ *população finita e amostra de grande dimensão*

7.4. Teorema do Limite Central

(1) Sejam X_1, X_2, ..., X_n ***variáveis independentes com distribuições idênticas*** (com valor esperado μ_X e variância σ_X^2 finitos).

- Qualquer que seja a forma da distribuição comum aos X_i's, a distribuição da variável soma

$$S = \sum_{i=1}^{n} X_i$$

tende para uma distribuição normal à medida que n aumenta.

- *Qualquer que seja a distribuição de uma população* (com valor esperado μ_X e variância σ_X^2 finitos), então *a distribuição da média amostral*

$$\bar{X} = \frac{1}{n} \times S$$

tende para uma distribuição normal quando a dimensão da amostra, n, tende para infinito.

(presupõe-se que a amostragem é aleatória simples, para que as variáveis X_i sejam independentes)

(2) As condições apresentadas em (1) podem ser relaxadas de duas formas:

- As variáveis X_1, X_2, ..., X_n ***podem ter distribuições diferentes*** umas das outras (desde que se verifique a condição de que ***a contribuição da variância de cada uma delas para a variância da soma seja relativamente pequena***).

- As variáveis X_1, X_2, ..., X_n ***podem ser dependentes desde que a correlação seja fraca***.

Questão: Quando é que o valor de n é suficientemente grande para que a distribuição Normal seja efectivamente uma boa aproximação da média amostral?

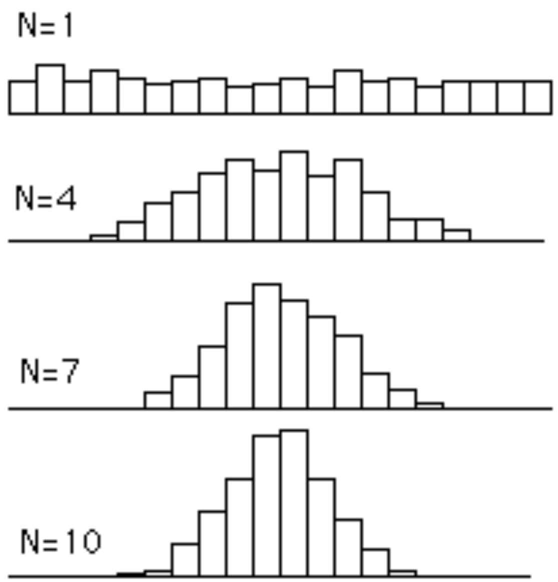
R: Para um dado valor de n, ***o rigor da aproximação depende da forma da distribuição original sendo, em particular, tanto menor quanto maior for a assimetria desta***.

Regra Prática:	$n \geq 10$	*se a distribuição original for simétrica*
	$n \geq 50$	*se a distribuição original for muito assimétrica*

DGI
2019

Teorema do Limite Central

Para uma qualquer população (caracterizada por uma qualquer distribuição) com variância finita, a distribuição da média amostral calculada com base numa amostra aleatória simples tende para uma distribuição normal, à medida que a dimensão da amostra aumenta.



DGI
2019

Exemplo

O custo de transporte de um lote de determinada matéria-prima depende da localização do fornecedor. Sabe-se que o custo é caracterizado por uma distribuição normal com valor médio igual a 575 u.m. e desvio padrão igual a 100 u.m..

a) Se forem adquiridos 10 lotes da referida matéria-prima a diferentes fornecedores, qual a média e a variância do custo médio desses lotes? Qual a probabilidade desse custo ultrapassar 6000 u.m.?

b) Determine a probabilidade do custo médio dos 10 lotes adquiridos ser superior a 600 u.m..