Amostragem

Introdução

- ***Censo ou recenseamento***:
  - ***Def***: Estudo baseado na análise de ***todos os elementos da população***.
  - *Nota*: Permite o cálculo exacto dos parâmetros populacionais (μ, σ, etc).

- ***Sondagem***:
  - ***Def***: Estudo baseado na análise de uma ***amostra da população***.
  - *Nota*: Fornece apenas estimativas dos parâmetros populacionais.

- ***Motivações para a sondagem***:
  - ***População infinita ou demasiado grande*** para a realização de um censo.
  - ***Tempo*** ou ***custo*** do censo ***excessivos***.
  - ***Recolha destrutiva*** dos dados.
  - Alguns ***elementos*** da população ***não acessíveis***.

Introdução

- ***Amostragem aleatória***:
  - ***Def***: Técnica de selecção dos elementos de uma amostra onde ***todos os elementos da população podem ser seleccionados,*** de acordo com uma probabilidade pré-definida.

- ***Amostra aleatória i.i.d.***:
  - ***Def***: Uma ***amostra*** diz-se ***aleatória i.i.d.*** quando as observações $X_1, X_2, ..., X_n$ que a constituem são *v. a.* ***independentes*** e ***identicamente distribuídas***.

- O ***tamanho da amostra*** influencia:
  - A ***fiabilidade*** da sondagem.
  - O ***tempo*** e o ***custo*** da sondagem.

Introdução

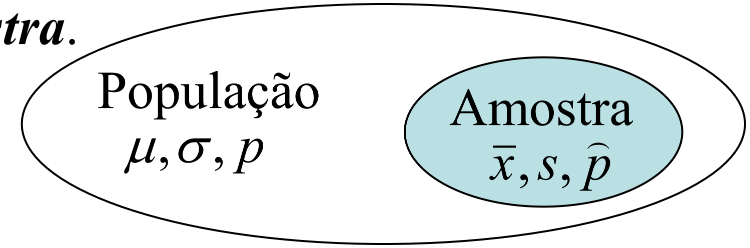
- ***Parâmetro, θ*:**
  - ***Def***: ***Constante*** que mede uma ***característica da população***.
  - ***Ex***: μ, σ, p

- ***Estatística,*** $\widehat{\Theta}$:
  - ***Def***: ***v. a.*** que representa uma ***função real*** das *v. a.*, $\mathbf{X_1}$, $\mathbf{X_2}$, **...**, $\mathbf{X_n}$, que constituem uma ***amostra***.

$$\widehat{\Theta} = G(X_1, X_2, \ldots, X_n)$$

  - ***Ex***: $\overline{X}, S, \widehat{P}$

- ***Estimativa,*** $\widehat{\theta}$:
  - ***Def***: ***Valor*** da estatística ***observado na amostra***.
  - ***Ex***: $\bar{x}, s, \widehat{p}$



Média amostral

- ***Média amostral,*** $\overline{X}$:
  - ***Def***: Seja X_1, X_2, ..., X_n uma amostra aleatória de tamanho n, de uma população X. A ***média amostral*** é a estatística

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

  - ***Utilidade***: É um estimador da média, μ, da população, X, porque

$$E\left(\overline{X}\right) = \mu$$

- ***Distribuição de*** $\overline{X}$:
  - ***Teor***: Seja X_1, X_2, ..., X_n uma amostra aleatória i.i.d. de tamanho n duma população X tal que E(X) = μ e V(X) = σ^2. Então:

$$i)\ X \sim N\left(\mu, \sigma^2\right) \underset{TAN}{\Rightarrow} \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$ii)\ n \geq 30 \underset{TLC}{\Rightarrow} \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Média amostral

	$X \sim \mathrm{N}(\mu, \sigma^2)$	$E(X) = \mu$ e $V(X) = \sigma^2$ mas X não é normal
$n \geq 30$	$\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$	$\bar{X}$ é aproximadamente $\mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$
$n < 30$	$\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$	$\bar{X}$ não é normal mas $E(\bar{X}) = \mu$ e $V(\bar{X}) = \sigma^2/n$

Exercício

A distância percorrida por um veículo a motor, com um litro de combustível, segue uma distribuição normal de média 72 km e de desvio padrão 23.13 km. Qual é a probabilidade da distância média percorrida por 45 veículos ser superior a 74 km?

X_i:"Distância percorrida pelo veículo *i*, em km" $\quad X_i \sim \mathrm{N}\left(72, 23.13^2\right)$

$\bar{X}$:"Distância ***média*** percorrida por 45 veículos, em km"

$$\bar{X} = \frac{1}{45}\sum_{i=1}^{45} X_i \sim \mathrm{N}\left(72, \frac{23.13^2}{45}\right)$$

TAN

$$P(\bar{X} > 74) = 1 - P(\bar{X} \leq 74) = 0.2809$$

$$1 - pnorm\left(74, 72, \frac{23.13}{\sqrt{45}}\right)$$

Proporção amostral

- ***Proporção amostral,*** $\hat{P}$:
  - ***Def***: Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória i.i.d. de tamanho n, tal que $X = \sum_{i=1}^{n} X_i \sim \mathrm{Bi}(n, p)$. A ***proporção amostral*** é a estatística

$$\hat{P} = \frac{X}{n}$$

  - ***Utilidade***: É um estimador da proporção de sucessos, p, numa população, porque

$$E(\hat{P}) = p$$

- ***Distribuição de*** $\hat{P}$:
  - ***Teor***: Se $X \sim \mathrm{Bi}(n,p)$, então

$$n \geq 30 \quad \underset{TLC}{\Rightarrow} \quad \hat{P} = \frac{X}{n} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

Exercício

Sabe-se que, numa fábrica que produz fios de cobre, a proporção de fios defeituosos é de 0.20. Calcule a probabilidade de, num lote de 100 desses fios, mais de 25% serem defeituosos.

$\widehat{P}$:"Proporção de fios defeituosos num lote de 100"

$$\widehat{P} \sim \mathrm{N}\left(p, \frac{pq}{n}\right), \quad \begin{cases} n = 100 \\ p = 0.2 \quad \Rightarrow \quad q = 0.8 \end{cases} \qquad \therefore \quad \widehat{P} \sim N\left(0.2, 0.04^2\right)$$

$\boxed{\text{TLC, porque } n \geq 30}$

$$P(\hat{P} > 0.25) = 1 - P(\hat{P} \leq 0.25) = 0.1056$$

$1 - pnorm(0.25, 0.2, 0.04)$

Variância amostral

- ***Variância amostral, S^2***:
  - ***Def***: Seja X_1, X_2, ..., X_n uma amostra aleatória de tamanho n. A ***variância amostral*** é a estatística

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

  - ***Utilidade***: É um estimador da variância, σ^2, numa população, porque

$$E\left(S^2\right) = \sigma^2$$

- ***Desvio padrão amostral, S***:
  - ***Def***:

$$S = \sqrt{S^2}$$

  - ***Utilidade***: É um estimador de σ.