Introdução à Inferência Estatística

Prof. Dr. Francisco Marcelo M. da Rocha

10 de Setembro de 2019

Índice

Objetivo da aula

Entender Noções básicas da ***Inferência Estatística***, com os conceitos de **População**, **Amostra**, **Distribuição Amostral da Média**, **Distribuição Amostral da Proporção** e **Intervalo de Confiança**.

População e Amostra Aleatória Simples

População e Amostra

População é o conjunto de todos os elementos ou resultados sob investigação.
Amostra é qualquer subconjunto da população.

Objetivo da Inferência Estatística

O objetivo da **Inferência Estatística** é tirar conclusões sobre a população com base na informação fornecida por uma amostra.

Amostragem Aleatória Simples - AAS

Podemos ter uma AAS **com reposição, $\mathbf{AAS_c}$**, se for permitido que uma unidade possa ser sorteada mais de uma vez, e **sem reposição,$\mathbf{AAS_s}$**, se a unidade for removida da população.

População e Amostra Aleatória Simples

Amostragem Aleatória Simples - AAS

Como a amostragem aleatória simples com reposição $\mathbf{AAS_c}$, conduz a um tratamento teórico mais simples, este será o plano amostral considerado inicialmente. Nós iremos nos referir a esse plano como **AAS**.

Amostragem Aleatória Simples - AAS

Uma **amostra aleatória simples** de tamanho n de uma variável aleatória X, com dada distribuição, é o conjunto de n variáveis aleatórias independentes $X_1, X_2, \ldots, X_n$, cada uma com a mesma distribuição de X.

Parâmetro Populacional, Estatística, Estimador e Estimativa

Parâmetro

As quantidades da população, em geral desconhecidas, sobre as quais temos interesse, são denominadas **parâmetros** e são representadas por letras gregas θ, μ e σ entre outras.

Exemplos

- μ - média de var. quantitativa na população.
- σ^2 - variância na população.
- p - proporção na população.

Parâmetro Populacional, Estatística, Estimador e Estimativa

Estatística

Uma **Estatística** é uma característica da amostra $X_1, X_2, \ldots, X_n$, ou seja, uma estatística T é uma função de $X_1, X_2, \ldots, X_n$ e portanto é uma variável aleatória.

São exemplos de Estatísticas

- $T(X_1, X_2, \ldots, X_n) = \overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$: média amostral.
- $T(X_1, X_2, \ldots, X_n) = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})$: variância amostral.
- $T(X_1, X_2, \ldots, X_n) = X_{(1)} = Min(X_1, X_2, \ldots, X_n)$: o menor valor na amostra.

Parâmetro Populacional, Estatística, Estimador e Estimativa

Estimador Pontual

Um **Estimador Pontual** é uma **estatística** construída com a finalidade de representar, ou estimar, um parâmetro de interesse na população. Em geral, os estimadores são representados pelas letras gregas que representam o parâmetro de interesse na população com o acento circunflexo: $\widehat{\mu}$, $\widehat{\theta}$, $\widehat{\sigma}^2$ etc.

São exemplos de Estimadores Pontuais

- $\widehat{\mu} = \overline{X}$ a média amostral é estimador da média na população μ.
- $\widehat{\sigma}^2 = S^2$: a variância amostral é estimador de σ^2 (variância na população).
- $\widehat{P}$: a proporção amostral é estimador da proporção de unidades com a característica de interesse na população p.

Parâmetro Populacional, Estatística, Estimador e Estimativa

Propriedades de Estimadores

- **Vício**: um estimador $\widehat{\theta}$ é não viciado ou não viesado para o parâmetro θ se

$$\mathbb{E}(\widehat{\theta}) = \theta$$

  A média amostral $\bar{X}$ é estimador não viesado de μ e a variância amostral S^2 é estimador não viesado de σ^2.

- **Consistência**: Um estimador $\widehat{\theta}$ é consistente para θ, se, à medida que o tamanho da amostra aumenta, seu valor esperado converge para o parâmetro de interesse e sua variância converge para zero.

$$\text{i) } \lim_{n\to\infty} \mathbb{E}(\widehat{\theta}) = \theta;$$

$$\text{ii) } \lim_{n\to\infty} \mathbb{V}\text{ar}(\widehat{\theta}) = 0;$$

Parâmetro Populacional, Estatística, Estimador e Estimativa

Estimativa

Uma **Estimativa** é um valor numérico assumido pelo **estimador**.

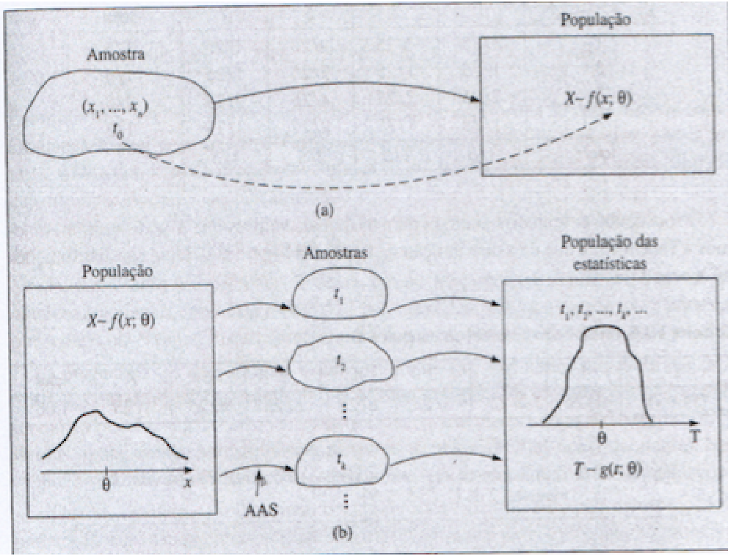
São exemplos de Estimativas

- $\overline{x} = 122$ é o valor de $\overline{X}$ para uma dada amostra observada.
- $s^2 = 144$ é o valor de S^2: para uma dada amostra observada.
- $\widehat{p} = 59\%$ é o valor de $\widehat{P}$ para uma dada amostra observada.

Distribuições Amostrais

Um estimador é uma função de variáveis aleatórias e portanto também é variável aleatória com uma distribuição de probabilidade denominada **Distribuição Amostral**.

Distribuições Amostrais - Bussab e Morettin (2013)



Fonte: Bussab e Morettin (2013).

Distribuições Amostrais - Bussab e Morettin

Exemplo

Consideremos uma população em que a variável X pode assumir um dos valores do conjunto $\{1, 3, 5, 5, 7\}$. A distribuição de probabilidade de X é

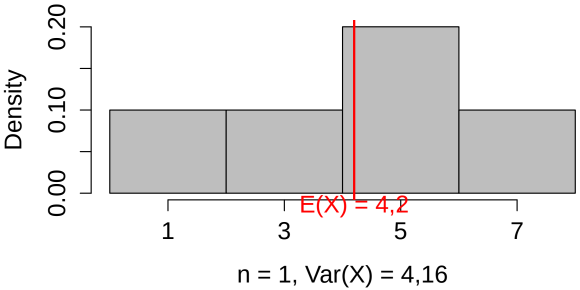
x	1	3	5	7
$\mathbb{P}(X = x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

Esperança e Variância

$$\mathbb{E}(X) = \mu_X = 4,2$$
$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X - \mu) = \sigma_X^2 = 4,16$$

Distribuições Amostrais

Figura 1: Distribuição amostral de $\overline{X}$ para amostras de tamanho 1.



Distribuições Amostrais

Exemplo

Vamos selecionar todas as amostras aleatórias simples de tamanho 2, $n = 2$, selecionadas ao acaso e com reposição da população X, e encontrar a distribuição do estimador pontual $\overline{X} = \widehat{\mu}_X$, ou seja, vamos encontrar a distribuição da média amostral.

$$\overline{X} = \frac{X_1 + X_2}{2}$$

em que

- X_1 é o valor selecionado na primeira extração.
- X_2 é o valor selecionado na segunda extração.

Amostra (X_1, X_2)	Probabilidade	Média Amostral
(1,1)	1/25	1
(1,3)	1/25	2
(1,5)	2/25	3
(1,7)	1/25	4
(3,1)	1/25	2
(3,3)	1/25	3
(3,5)	2/25	4
(3,7)	1/25	5
(5,1)	1/25	3
(5,3)	1/25	4
(5,5)	2/25	5
(5,7)	1/25	6
(7,1)	1/25	4
(7,3)	1/25	5
(7,5)	2/25	6
(7,7)	1/25	7

Distribuições Amostrais

Distribuição de $\overline{X}$ para $n = 2$

$\overline{x}$	1	2	3	4	5	6	7
$\mathbb{P}(\overline{X} = \overline{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25

Esperança e Variância de $\overline{X}$, $n = 2$.

$$\mathbb{E}(\overline{X}) = \mu_X = 4,2$$

$$\mathbb{V}\mathrm{ar}(\overline{X}) = 2,08 = \frac{\sigma_X^2}{2}$$

Distribuições Amostrais

Distribuição de $\overline{X}$ para $n = 3$

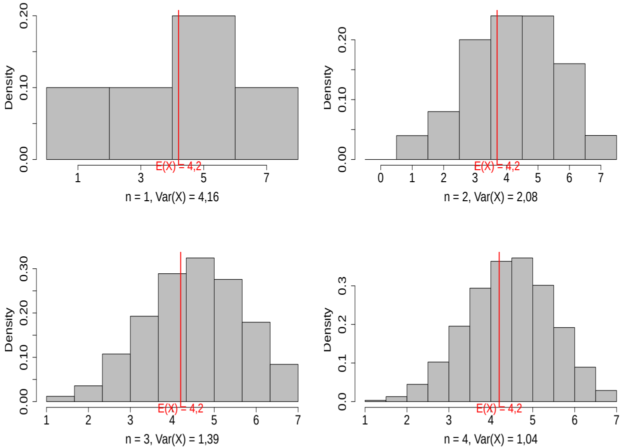
$\bar{x}$	$\mathbb{P}(\overline{X} = \bar{x})$
1	1/125
5/3	3/125
7/3	9/125
3	16/125
11/3	24/125
13/3	27/125
5	23/125
17/3	15/125
19/3	6/125
1	1/125

Esperança e Variância de $\overline{X}$, $n = 3$.

$$\mathbb{E}(\overline{X}) = \mu_X = 4,2 \;\text{ e }\; \mathbb{V}\mathrm{ar}(\overline{X})1,39 = \frac{\sigma_X^2}{3}$$

Distribuições Amostrais

Figura 2: Histogramas correspondentes às distribuições de X e $\overline{X}$ para diferentes amostras da população $\{1, 3, 5, 5, 7\}$.



Distribuições Amostrais

Análise dos Histogramas

- Conforme o tamanho da amostra aumenta, $n \to \infty$, os valores de $\overline{X}$ tendem a concentrar-se cada vez mais em torno de $\mathbb{E}(\overline{X}) = \mu_X = 4,2$.
- A variância diminui à medida em que o tamanho da amostra aumenta.
- Para suficientemente grande, a forma do histograma aproxima-se de uma distribuição normal.

Figura 3: Histogramas correspondentes às distribuições de $\overline{X}$ para amostras de tamanho 1 de algumas populações.

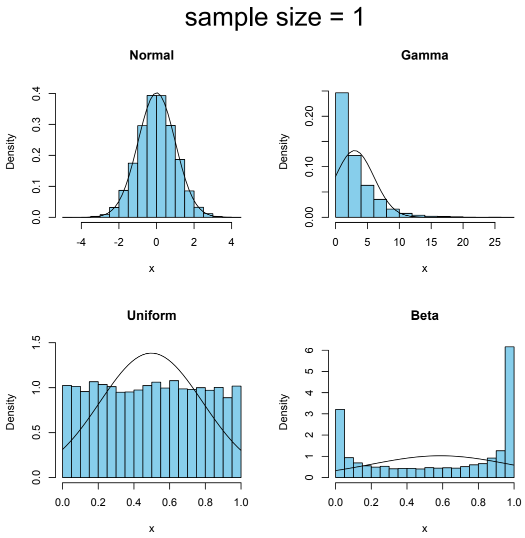
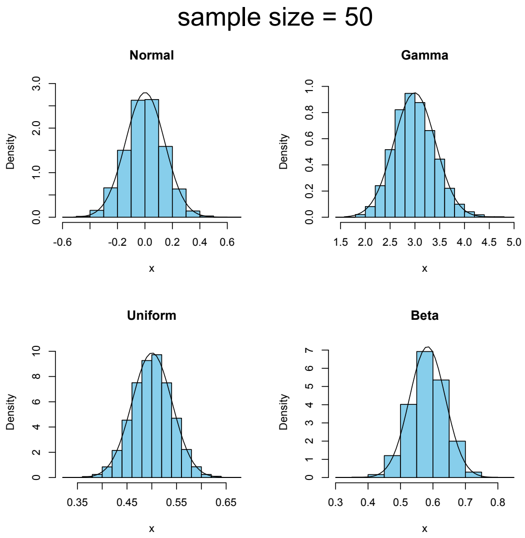


Figura 4: Histogramas correspondentes às distribuições de $\overline{X}$ para amostras de tamanho 50 de algumas populações.



Distribuições Amostrais

Os gráficos acima sugerem que,

- quando o tamanho da amostra aumenta, independentemente da forma da distribuição de X, a distribuição de probabilidade da média amostral $\overline{X}$ aproxima-se de uma distribuição normal.

Teorema Central do Limite - TC

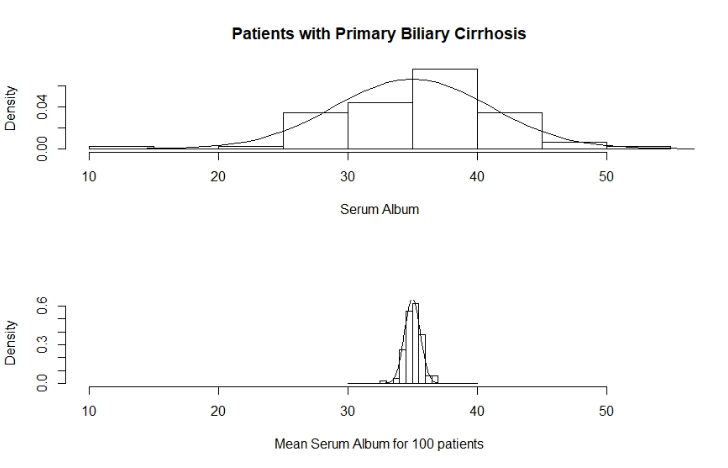
Para amostras aleatórias simples, $X_1, X_2, \ldots, X_n$, retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média $\overline{X}$ aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2/n, e

$$\frac{(\overline{X} - \mu)}{\sqrt{\sigma/n}} \xrightarrow[n \to \infty]{\text{em distribuição}} Z \sim \mathcal{N}(0; 1).$$

- A variável aleatória $e = \overline{X} - \mu$ é denominada **erro amostral da média**.
- O desvio padrão $\sigma/\sqrt{n}$ é denominado **erro padrão da média**.

Teorema Central do Limite - TC

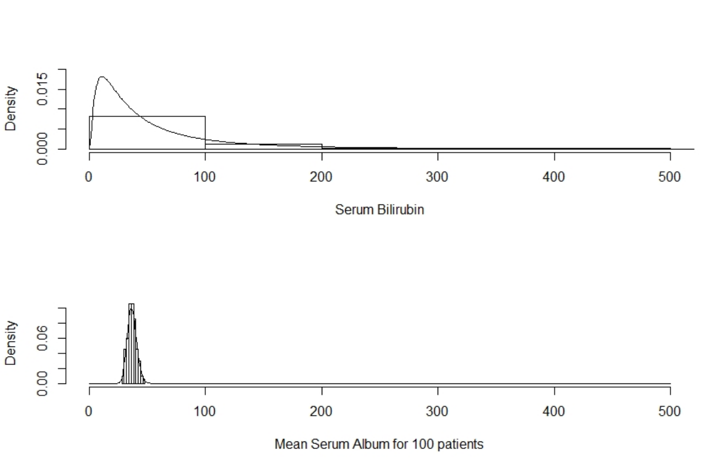
Altman p.153 = $\mu = 35$g/l, $\sigma = 6$ g/l



Teorema Central do Limite - TC

Altman p.155 - Serum Bilirubin $\mu = 60.73\mu mol/l$, $\sigma = 77.91\mu mol/l$

Log Serum Bilirubin parece $N(3.6, 1.1^2)$



Distribuição Amostral da Média

Exemplo 1.

Sabe-se que o gasto diário com internações segue uma distribuição de média 20 mil e desvio padrão 2 mil. Qual é a probabilidade de que num período de 60 dias, o gasto total ultrapasse R$ 1.230.000,00?

Seja X o gasto diário em milhares de reais e sejam os gastos diários independentes. Sabemos que $\mathbb{E}(X) = \mu_X = 20$ e $\sigma_X = \sqrt{\mathbb{V}\mathrm{ar}(X)} = 2$. Obtendo uma amostra de 60 valores, denotada por $X_1, X_2, \ldots, X_{60}$, com X_i representando o gasto no dia $i = 1, 2, 3, \ldots, 60$. Então

$$\mathbb{P}(X_1 + X_2 + \ldots + X_{60} > 1230) = \mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_{60}}{60} > \frac{1230}{60}\right) =$$

$$\mathbb{P}(\overline{X} > 20,5) \approx \mathbb{P}\left(Z > \frac{\sqrt{60}(20,5 - 20)}{2}\right) = 0,0262$$

Distribuição Amostral da Média

Exemplo 2. Bussab e Morettin, 2011, 7a ed. p.280,

A capacidade máxima de um elevador é de 500kg. Se a distribuição dos pesos (X) dos usuários é suposta N(70kg, 100kg^2):

a) Qual a probabilidade de 7 passageiros ultrapassarem esse limite?
b) E seis passageiros?

Estimação para a Média

Objetivo

Estimar a partir de uma AAS a média populacional desconhecida μ de uma variável aleatória X, que representa a característica de interesse da população.

Exemplos:

- μ: idade média das alunas da faculdade;
- μ: a renda média, em reais, dos habitantes de uma localidade;
- μ: o salário médio de clientes de uma seguradora;
- μ: a altura média.

Estimação para a Média

- Retiramos uma AAS, $X_1, \ldots, X_n$, de n elementos da população;
- Para cada elemento selecionado, observamos o valor da variável de interesse X;
- Um ***Estimador Pontual*** para μ é a média amostral

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

- Um ***Estimador Intervalar*** ou um ***Intervalo de Confiança*** para μ tem a forma

$$]\overline{X} - \epsilon; \overline{X} + \epsilon],$$

sendo ϵ o **erro amostral** ou **erro de estimação**, ou margem de erro, calculado a partir da distribuição de $\overline{X}$.

Intervalo de Confiança para a Média

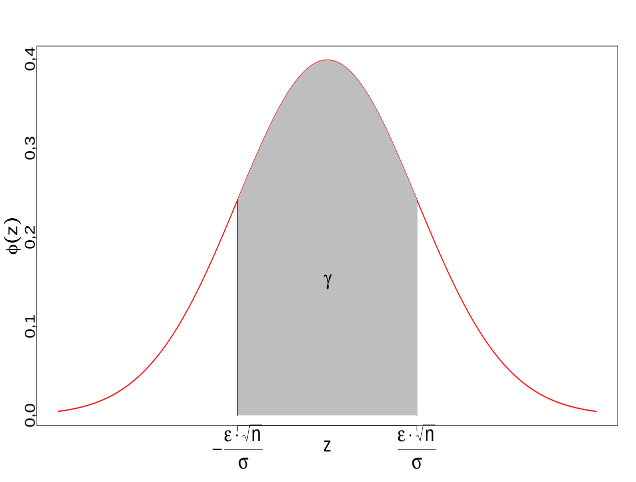
Como determinar ϵ ?

- Seja γ a probabilidade da média amostral $\overline{X}$ estar a uma distância de, no máximo ϵ, da média populacional μ

$$\begin{aligned}
\gamma &= \mathbb{P}(|\overline{X} - \mu| \leq \epsilon) \\
&= \mathbb{P}\left(\mu - \epsilon \leq \overline{X} \leq \mu + \epsilon\right) \\
&= \mathbb{P}\left(-\frac{\epsilon}{\sigma/\sqrt{n}} \leq \frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}} \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right) \\
&= \mathbb{P}\left(-\frac{\epsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right),
\end{aligned}$$

com $Z \sim \mathcal{N}(0, 1)$.

Intervalo de Confiança para a Média



Intervalo de Confiança para a Média

- Fazendo $\dfrac{\epsilon\sqrt{n}}{\sigma} = z_\gamma$, obtemos o erro amostral

$$\epsilon = \frac{z_\gamma \sigma}{\sqrt{n}},$$

com $\mathbb{P}(z_\gamma \leq Z \leq z_\gamma) = \gamma$.

- Conhecendo-se o coeficiente de confiança γ obtemos z_γ.

Exercício

Considerando o nível de confiança $\gamma = 0,95$. Determine z_γ.

Intervalo de Confiança para a Média com Variância Conhecida

- O **intervalo de confiança para a média** μ, com coeficiente de confiança γ é dado por

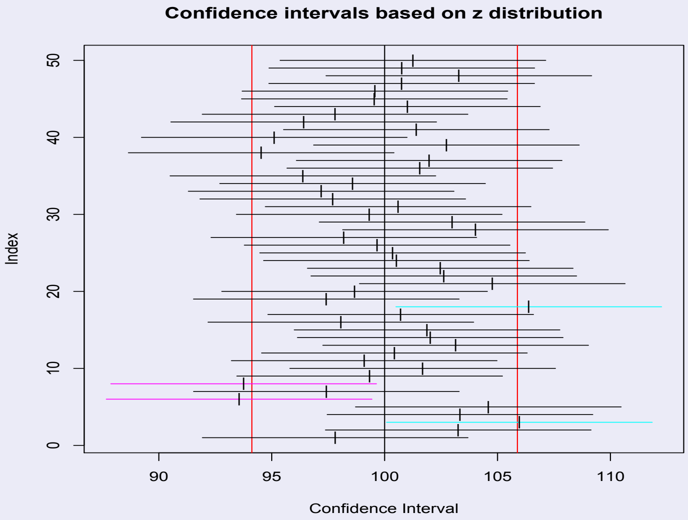
$$IC(\mu;\gamma) = \left]\overline{X} - z_{\gamma}\frac{\sigma}{\sqrt{n}};\overline{X} + z_{\gamma}\frac{\sigma}{\sqrt{n}}\right[,$$

sendo σ sendo o desvio padrão de X.

- Interpretação do intervalo de confiança:

Se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $\left]\overline{X} - z_{\gamma}\frac{\sigma}{\sqrt{n}};\overline{X} + z_{\gamma}\frac{\sigma}{\sqrt{n}}\right[$***, todos baseados em amostras de tamnho*** n***,*** γ***100% deles conteriam o parâmetro*** μ***.***

$\mu = 100, \sigma = 15, n = 25$ e N^o de amostras$= 100$



Intervalo de Confiança para a Média com Variância Conhecida

Exemplo 2.

Não se conhece o consumo médio de combustível de automóveis da marca T. Sabe-se, no entanto, que o desvio padrão do consumo de combustível de automóveis dessa marca é 10 km/l. Na análise de 100 automóveis da marca T, obteve-se consumo médio de combustível de 8 km/l. Encontre um intervalo de confiança para o consumo médio de combustível dessa marca de carro. Adote um coeficiente de confiança igual a 95%.

Intervalo de Confiança para a Média com Variância Conhecida

Exemplo 2.

- X: consumo de combustível de automóveis da marca T.
- $\sigma_X = 10\text{km}/\text{l}$, $n = 100$, $\bar{x} = 8\text{km}/\text{l}$ e $\gamma = 0,95 \Rightarrow z_\gamma = 1,96$.
- Obtenha o intervalo de confiança.

Intervalo de Confiança para a Média com Variância Conhecida

Exercício

Deseja-se estimar o tempo médio de estudo (em anos) da população adulta de um município. Sabe-se que o tempo de estudo tem distribuição normal com desvio padrão $\sigma = 2,6$ anos. Foram entrevistados $n = 50$ indivíduos, obtendo-se para essa amostra, um tempo médio de estudo igual a 10,5 anos. Obter um intervalo de 90% de confiança para o tempo médio de estudo populacional.

Dimensionamento da Amostra

Dimensionamento da Amostra

A partir da expressão do erro de estimação (erro amostral) podemos determinar o tamanho da amostra

$$\epsilon = \frac{z_\gamma \sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_\gamma}{\epsilon}\right)^2 \sigma^2.$$

- Para determinarmos o tamanho da amostra precisamos fazer afirmações sobre ϵ, γ e σ^2.
- Quando não conhecemos σ^2 podemos usar uma pequena amostra piloto para estimar σ^2 usando o estimador S^2 e, uma vez observada a amostra piloto, obtemos uma estimativa da variância s^2.
- Os intervalos de confiança e a maneira de dimensionarmos a amostra é baseada no fato de que a população é infinita.

Dimensionamento da Amostra

Exercício

A renda per-capita domiciliar numa certa região tem distribuição normal com desvio padrão $\sigma = 250$ reais e média μ desconhecida. Se desejamos estimar a renda mdéia μ com erro $\epsilon = 50$ reais e com uma confiança $\gamma = 95\%$, quantos domicílios devemos consultar?

Intervalo de Confiança para a Média com Variância Desconhecida

- O **intervalo de confiança para a média** μ, com coeficiente de confiança γ é dado por

$$IC(\mu;\gamma) = \left]\overline{X} - t_{\gamma}\frac{s}{\sqrt{n}}; \overline{X} + z_{\gamma}\frac{s}{\sqrt{n}}\right[,$$

sendo s sendo o desvio padrão amostral de X e
t é quantil da distribuição t de Student (Student, 1908).

- Interpretação do intervalo de confiança:

> ***Se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $\left]\overline{X} - t_{\gamma}\frac{s}{\sqrt{n}}; \overline{X} + t_{\gamma}\frac{s}{\sqrt{n}}\right[$, todos baseados em amostras de tamnho n, γ100% deles conteriam o parâmetro μ.***

Distribuição Amostral de uma Proporção

Distribuição Amostral de uma Proporção

Vamos considerar uma população em que a proporção de elementos portadores de certa característica é **p**. Logo, podemos definir uma v.a. X, da seguinte maneira:

$$X = \begin{cases} 1, & \text{se o indivíduo possuir a característica} \\ 0, & \text{se o indivíduo não possuir a característica.} \end{cases}$$

Qual é a distribuição de probabilidade de X?. A v.a. X tem

$$\mathbb{E}(X) = \mu_X = p$$
$$\mathbb{V}\mathrm{ar}(X) = \sigma_X^2 = p(1-p).$$

- Retirada uma AAS dessa população, considere a v.a. Y_n: número de indivíduos portadores da característica de interesse na amostra.

Distribuição Amostral de uma Proporção

- $Y_n = X_1 + \ldots + X_n$ e $Y_n \sim Bin(n;p)$.
- A proporção de indivíduos portadores da característica é definida como $\widehat{P} = Y_n/n = \overline{X}$
- A distribuição amostral exata de $\widehat{P}$ é obtida da distribuição de Y_n.

$$\mathbb{P}(Y_n = k) = \mathbb{P}\left(\frac{Y_n}{n} = \frac{k}{n}\right) = \mathbb{P}\left(\widehat{P} = \frac{k}{n}\right).$$

- Pelo TLC $\widehat{P}$ terá distribuição aproximadamente normal

$$\widehat{P} \sim \mathcal{N}\left(p; \frac{p(1-p)}{n}\right).$$

- Uma amostra **grande** não é suficiente para a normalidade. Para garantir que a distribuição de $\widehat{P}$ seja aproximadamente simétrica $np \geq 5$ e $n(1-p) \geq 5$.

Distribuição Amostral de uma Proporção

Exemplo 1

Visitantes estrangeiros nos Estados Unidos são cuidadosamente monitorados por organizações de segurança e pela indústria do turismo. Durante o mês de março de 2008 foram monitorados 4,7 milhões de visitantes internacionais nos EUA. Quarenta por cento de todos os visitantes da Europa Ocidental eram do Reino Unido (RU). Suponha que sejam selecionados aleatoriamente 120 visitantes do mês de março, vindos da Europa Ocidental, e que se determine o número dos visitantes que vieram do RU.

a) Ache a distribuição da proporção amostral de visitantes do RU, $\widehat{P}$.
b) Qual é a probabilidade de que a proporção amostral seja maior do que 0,50?
c) Ache a probabilidade de que a proporção amostral esteja entre 0,32 e 0,37.

Distribuição Amostral de uma Proporção

Exemplo 1

a) A amostra selecionada foi de 120 turistas, $n = 120$. A proporção de visitantes da Europa Ocidental, $p = 0,40$. Os critérios de não assimetria está satisfeito.

$$np = 120 \times 0,40 = 48 \text{ e } n(1-p) = 120 \times 0,60 = 72.$$

Portanto dizer que $\widehat{P} \sim \mathcal{N}\left(p = 0,40, \dfrac{p(1-p)}{n} = \dfrac{0,40 \times 0,60}{120}\right)$.

$$\widehat{P} \sim \mathcal{N}(0,40; 0,002).$$

Distribuição Amostral de uma Proporção

Exemplo 1

b)

$$\mathbb{P}(\widehat{P} > 0,50) = \mathbb{P}\left(\frac{\widehat{P} - 0,40}{\sqrt{0,002}} > \frac{0,50 - 0,40}{\sqrt{0,002}}\right)$$
$$= \mathbb{P}(Z > 2,24) = 1 - \mathbb{P}(Z \leq 2,24)$$
$$= 1 - 0,9875 = 0,0125.$$

c)

$$\mathbb{P}(0,32 \leq \widehat{P} \leq 0,37) = \mathbb{P}(-1,67 \leq Z \leq -0,67)$$
$$= \mathbb{P}(Z \leq -0,67) - \mathbb{P}(Z \leq -1,67)$$
$$= 0,2514 - 0,0367 = 0,2147.$$

Estimação para a Proporção Populacional p

Objetivo

Estimar uma proporção p (desconhecida) de elementos em uma população, apresentando certa caractersítica de interesse, a partir da informação fornecida por uma amostra.

Estimação para a Proporção Populacional p

Exemplos:

- p: proporção de alunos da EPPEN que foram ao teatro no último mês;
- p: proporção de consumidores satisfeitos com os serviços prestados por uma empresa telefônica;
- p: proporção de pessoas no município de São Paulo favoráveis à diminuição da maioridade penal.

Estimação para a Proporção Populacional p

Procedimento de Estimação:

- **Estimação Pontual**. O estimador pontual para p é a proporção amostral definida como
$$\widehat{P} = \frac{Y_n}{n}.$$
Se observamos k indivíduos com as características de interesse numa amostra de tamnho n, uma estimativa potual para p é $\widehat{p} = k/n$.
- Um ***Estimador Intervalar*** ou um ***Intervalo de Confiança*** para p tem a forma
$$]\widehat{P} - \epsilon; \widehat{P} + \epsilon],$$
sendo ϵ o **erro amostral** ou **erro de estimação**, ou margem de erro, calculado a partir da distribuição aproximada de $\widehat{P}$.

Intervalo de Confiança para p

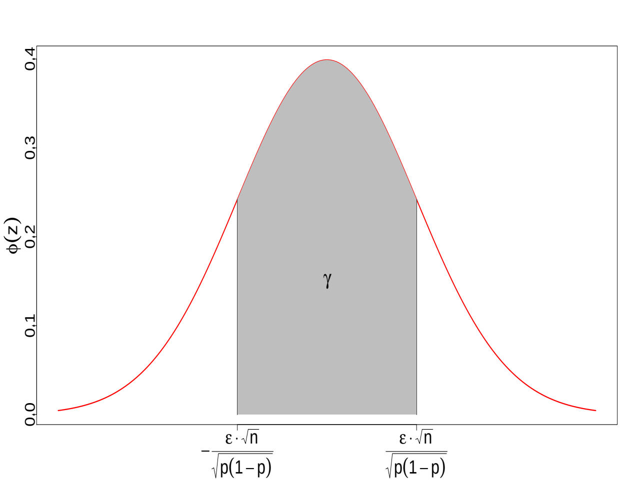
Como determinar ϵ ?

- Seja γ a probabilidade da média amostral $\overline{X}$ estar a uma distância de, no máximo ϵ, da média populacional μ

$$
\begin{aligned}
\gamma &= \mathbb{P}(|\widehat{P} - p| \leq \epsilon) \\
&= \mathbb{P}\left(p - \epsilon \leq \widehat{P} \leq P + \epsilon\right) \\
&= \mathbb{P}\left(-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} \leq \frac{\sqrt{n}(\widehat{P} - p)}{\sqrt{p(1-p)}} \leq \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \\
&= \mathbb{P}\left(-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} \leq Z \leq \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right),
\end{aligned}
$$

com $Z \sim \mathcal{N}(0,1)$.

Intervalo de Confiança para p



Intervalo de Confiança para p

- Fazendo $\dfrac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} = z_\gamma$, obtemos o erro amostral

$$\epsilon = z_\gamma \sqrt{\frac{p(1-p)}{n}},$$

com $\mathbb{P}(z_\gamma \leq Z \leq z_\gamma) = \gamma$.

- Conhecendo-se o coeficiente de confiança γ obtemos z_γ.

Exercício

Considerando o nível de confiança $\gamma = 0,95$. Determine z_γ.

Intervalo de Confiança para a p

- O **intervalo de confiança para a p**, com coeficiente de confiança γ é dados por

$$IC(p;\gamma) = \left]\widehat{p} - z_{\gamma}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}};\widehat{p} + z_{\gamma}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right[.$$

- Interpretação do intervalo de confiança:

Se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $IC(p;\gamma)$, todos baseados em amostras de tamnho n, γ100% deles conteriam o parâmetro p.

Intervalo de Confiança para a p

- O **intervalo de confiança, conservador, para a p**, com coeficiente de confiança γ é dados por

$$IC(p;\gamma) = \left]\widehat{p} - \frac{z_\gamma}{\sqrt{4n}}; \widehat{p} + \frac{z_\gamma}{\sqrt{4n}}\right[. \quad (1)$$

- O intervalo (2) é chamado de conservador, pois se p não for igual a $1/2$ e estiver próximo de zero ou de um, então ele fornece um intervalo desnecessariamente maior, porque substituímos $p(1-p)$ pelo seu valor máximo, $1/4$.

Intervalo de Confiança para a p

Exercício

Uma bem-sucedida companhia tem, em geral, seu nome e logomarca com alto nível de reconhecimento pelos consumidores. Por exemplo, os produtos da Coca-Cola estão disponíveis para 98% da população mundial e, portanto, deve ter o maior índice de reconhecimento de logomarca do que qualquer outra companhia. Uma empresa de tecnologia, que desenvolve certo produto, gostaria de estimar a proporção de pessoas que reconhecem a logomarca de pinguim do Linux. Dos 952 consumidores pesquisados, selecionados aleatoriamente, 132 puderam identificar o produto associado ao pinguim.

a) A distribuição de $\widehat{P}$ tem distribuição aproximadamente normal?
b) Ache um intervalo de confiança de 95% para a verdadeira proporção de consumidores que reconhecem o pinguim do Linux.

Dimensionamento da Amostra

Dimensionamento da Amostra

A partir da expressão do erro de estimação (erro amostral) podemos determinar o tamanho da amostra

$$\epsilon = z_{\gamma} \frac{\sqrt{p(1-p)}}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\gamma}}{\epsilon}\right)^2 p(1-p).$$

- Para determinarmos o tamanho da amostra precisamos fazer afirmações sobre ϵ, γ e p.
- Uma maneira de procedermos é substituir p por sua estimativa pontual $\widehat{p}$, ou podemos usar o valor 0.5, nesse caso obteremos o maior tamanho de amostra possível (mais conservador).
- Os intervalos de confiança e a maneira de dimensionarmos a amostra é baseada no fato de que a população é infinita.

Amostragem Sem Reposição em Pequenas Populações

Fração Amostral

A ***fração amostral*** é a relação entre o tamanho da amostra **n** e o tamanho da população **N**, ou seja, é n/N.

Intervalo de Confiança para a Média μ

- Se o tamanho da amostra, n, for maior do que 5% da população, de tamanho N, o **intervalo de confiança para a média** μ, com coeficiente de confiança γ é dados por

$$IC(\mu;\gamma) = \left]\overline{X} - z_{\gamma}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}};\ \overline{X} + z_{\gamma}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right[,$$

sendo σ sendo o desvio padrão de X.

Amostragem Sem Reposição em Pequenas Populações

Dimensionamento da Amostra

A partir da expressão do erro de estimação (erro amostral), ϵ, podemos determinar o tamanho da amostra

$$\epsilon = z_\gamma \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \Rightarrow n = \frac{z_\gamma \sigma^2 N}{\epsilon^2 (N-1) + z_\gamma \sigma^2}.$$

- Para determinarmos o tamanho da amostra precisamos fazer afirmações sobre ϵ, γ, σ^2 e sabermos o tamanho da população N.
- Quando não conhecemos σ^2 podemos usar uma pequena amostra piloto para estimar σ^2 usando o estimador S^2 e, uma vez observada a amostra piloto, obtemos uma estimativa da variância s^2.
- Os intervalos de confiança e a maneira de dimensionarmos a amostra é baseada no fato de que a população é finita.

Intervalo de Confiança para a Proporção Populacional p

- O **intervalo de confiança para a p**, com coeficiente de confiança γ é dados por

$$IC(p;\gamma) = \left]\widehat{p} - z_{\gamma}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\sqrt{\frac{N-n}{N-1}};\ \widehat{p} + z_{\gamma}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\sqrt{\frac{N-n}{N-1}}\right[.$$

Intervalo de Confiança Conservador para a Proporção Populacional p

- O **intervalo de confiança, conservador, para a p**, com coeficiente de confiança γ é dados por

$$IC(p;\gamma) = \left]\widehat{p} - \frac{z_{\gamma}}{\sqrt{4n}}\sqrt{\frac{N-n}{N-1}};\ \widehat{p} + \frac{z_{\gamma}}{\sqrt{4n}}\sqrt{\frac{N-n}{N-1}}\right[. \tag{2}$$

- O intervalo (2) é chamado de conservador, pois se p não for igual a $1/2$ e estiver próximo de zero ou de um, então ele fornece um intervalo desnecessariamente maior, porque substituímos $p(1-p)$ pelo seu valor máximo, $1/4$.

Dimensionamento da Amostra

Dimensionamento da Amostra

A partir da expressão do erro de estimação (erro amostral) podemos determinar o tamanho da amostra

$$\epsilon = z_\gamma \frac{\sqrt{p(1-p)}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \Rightarrow n = \frac{z_\gamma^2 p(1-p)N}{z_\gamma^2 p(1-p) + \epsilon^2 (N-1)}.$$

- Para determinarmos o tamanho da amostra precisamos fazer afirmações sobre ϵ, γ, p e sabermos o tamanho da população N.
- Uma maneira de procedermos é substituir p por sua estimativa pontual $\widehat{p}$, ou podemos usar o valor 0.5, nesse caso obteremos o maior tamanho de amostra possível (mais conservador).
- Os intervalos de confiança e a maneira de dimensionarmos a amostra é baseada no fato de que a população é finita.

Leitura

Leia.

- Capítulo 10, seções 10.1 até 10.9, 10.11.
- Capítulo 11, seções 11.6.
- Além da lista tente fazer os exercícios das seções dos capítulos.

Referências Bibliográficas

- Bussab, W.O. e Morettin, P.A. (2013). **Estatística Básica**. 8ª edição. São Paulo: Editora Saraiva.
- Magalhães, M.N. e Lima, A.C.P. (2013). **Noções de Probabilidade e Estatística**. 7ª edição. São Paulo: EDUSP.