Probabilidade e Estatística

Professores:

Otton Teixeira da Silveira Filho

Regina Célia Paula Leal Toledo

Probabilidade e Estatística

Livro Texto:

[1] "Noções de Probabilidade e Estatística"
Marcos Nascimento Magalhães e Antonio Carlos Pedroso de Lima, Edusp (2005).

[2] "Probabilidade: Um Curso Introdutório"
Carlos A. B. Dantas, Edusp (2004).

Aula 11

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Inferência Estatística - Estimação II

Conteúdo:

- 11.1 Distribuições Amostrais
- 11.2 Teorema Central do Limite
- 11.3 Estimação por Intervalo

$$\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\sigma_X^2 \qquad \bar{Y}=\frac{\sum_{i=1}^{n} y_i}{n}$$

$$A\cap B=\varnothing$$

11.0 Recapitulando

Uma definição:

Inferência Estatística é um conjunto de técnicas que objetiva estudar uma população através de evidências fornecidas por uma amostra.

11.1 Distribuições Amostrais

Estimadores são funções de variáveis aleatórias sendo, também, variáveis aleatórias. Mas....

Qual a distribuição de probabilidade dos estimadores?

11.1 Distribuições Amostrais

Para variar, um exemplo:

Com uma moeda equilibrada, jogamos três vezes. Se sai cara, você ganha um ponto, se sai coroa, você perde um ponto.

Vamos chamar de $\mathbf{X}$ a variável aleatória que toma os valores 1 e -1, com probabilidades iguais e vamos determinar as funções de probabilidade dos estimadores $\overline{X}$ e S^2

11.1 Distribuições Amostrais

O vetor amostral (X_1, X_2, X_3) é constituido de variáveis aleatórias independentes e identicamente distribuidas com função de probabilidade igual à de $\mathbf{X}$

(X_1, X_2, X_3)	probabilidade	$\overline{X}$	S^2
(-1, -1, -1)	1/8	-1	0
(-1, -1, 1)	1/8	-1/3	4/3
(-1, 1, -1)	1/8	-1/3	4/3
(-1, 1, 1)	1/8	1/3	4/3
(1, -1, -1)	1/8	-1/3	4/3
(1, -1, 1)	1/8	1/3	4/3
(1, 1, -1)	1/8	1/3	4/3
(1, 1, 1)	1/8	1	0

11.1 Distribuições Amostrais

Destes dados podemos tirar as informações

E(X) = 0 (média) Var(X) = 1 (variância)

Os valores tabelados são simples de calcular. Por exemplo, para a amostra (-1, 1, -1) escreveremos

$$\bar{x}_{obs} = \frac{-1+1-1}{3} = \frac{-1}{3}$$

$$s^2_{obs} = \frac{(-1)^2 + (1)^2 + (-1)^2 - 3(-1/3)^2}{(3-1)} = \frac{4}{3}$$

11.1 Distribuições Amostrais

Por meio da tabela, podemos construir as distribuições dos estimadores:

$\bar{X}$	-1	-1/3	1/3	1
p_i	1/8	3/8	3/8	1/8

S^2	0	4/3
p_i	1/4	3/4

que nos dão os valores esperados

$$E(\bar{X}) = (-1) \times \frac{1}{8} + (-1/3) \times \frac{1}{8} + 1/3 \times \frac{1}{8} + 1 \times \frac{1}{8} = 0 \qquad E(S^2) = 0 \times \frac{1}{4} + \frac{4}{3} \times \frac{3}{4} = 1$$

11.1 Distribuições Amostrais

Observe que:

$$E(\overline{X}) = 0 = E(\mathbf{X}) \qquad E(S^2) = 1 = Var(\mathbf{X})$$

Assim temos que ambos os estimadores são não viciados.

Mas assim é fácil, não? E num caso mais geral?

11.1 Distribuições Amostrais

Importante

No problema anterior é possível enumerar as amostras, o que facilita achar a função de probabilidade.

No entanto,

Se $\mathbf{X}$ tivesse distribuição Uniforme (portanto, contínua), entre -1 e 1, não seria viável enumerar todas as amostras possíveis.

Assim, em geral, obter a distribuição de probabilidade de estimadores é um problema importante e, muitas vezes, difícil.

11.1 Distribuições Amostrais

Vamos para o caso da distribuição Normal, ou seja,

$$\mathbf{X} \sim N(\mu, \sigma^2)$$

Temos que (X_1, X_2, X_3) representa uma amostra aleatória com elementos independentes e identicamente distribuídos, com média e variância μ, σ^2 .

Também já vimos que qualquer combinação linear de $\boldsymbol{X}_i$ tem também uma distribuição Normal, ou seja,

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

tem distribuição Normal.

11.1 Distribuições Amostrais

Com estes dados podemos obter a distribuição média amostral para a distribuição Normal se tomarmos no somatório $a_i = 1/n$ para $n = 1, ..., n$.

Usando as propriedades da Esperança e da Variância obtemos

$$\boldsymbol{\mu}_x = E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} a_i X_i\right) = \frac{1}{n} n\boldsymbol{\mu} = \boldsymbol{\mu} \qquad \boldsymbol{\sigma}_x^2 = Var\left(\frac{i}{n}\sum_{i=1}^{n} a_i X_i\right) = \frac{1}{n^2} n\boldsymbol{\sigma}^2 = \frac{\boldsymbol{\sigma}^2}{n}$$

Conclusão:

11.1 Distribuições Amostrais

Com estes dados podemos obter a distribuição média amostral para a distribuição Normal se tomarmos no somatório $a_i = 1/n$ para $n = 1, ..., n$.

Usando as propriedades da Esperança e da Variância obtemos

$$\boldsymbol{\mu}_{\mathrm{x}} = \mathrm{E}(\bar{\mathrm{X}}) = \mathrm{E}\left(\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i X_i}\right) = \frac{1}{\mathrm{n}}\mathrm{n}\boldsymbol{\mu} = \boldsymbol{\mu} \qquad \boldsymbol{\sigma}_{\mathrm{x}}^2 = \mathrm{Var}\left(\frac{\mathrm{i}}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i X_i}\right) = \frac{1}{\mathrm{n}^2}\mathrm{n}\boldsymbol{\sigma}^2 = \frac{\boldsymbol{\sigma}^2}{\mathrm{n}}$$

Conclusão: Uma coleção de variáveis aleatórias independentes, com uma mesma distribuição de probabilidade dada por um modelo Normal com média e variância μ, σ^2 ,terá a média amostral $\mathbf{X}$ também com distribuição Normal mas com média e variância $\mu, \dfrac{\sigma^2}{n}$.

11.1 Distribuições Amostrais

Observe:

Tomando o resultado anterior, podemos concluir que a medida que a amostra cresce, a probabilidade de a média amostral estar na vizinhança da média da população se torna maior.

11.1 Distribuições Amostrais

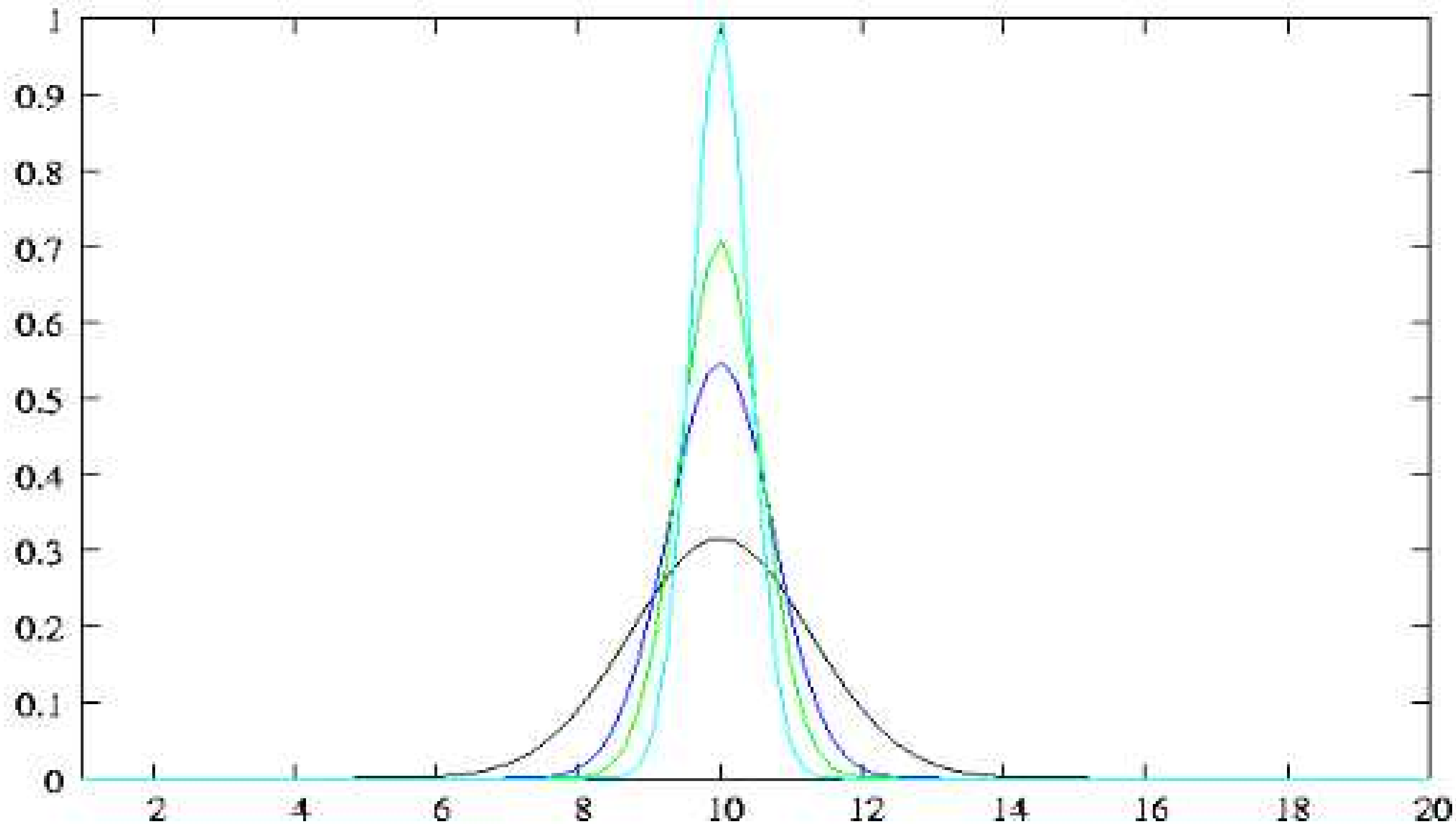
Mais um exemplo:

Considere uma amostra independente da tamanho ***n*** de uma variável com distribuição normal de média 10 e variância 16, ou seja, N(10, 16). Como se comporta $\overline{X}$ em função de ***n***?

A variável aleatória $\overline{X}$ tem distribuição N(10, 16/n), pelo resultado que achamos para a distribuição média amostral para a distribuição Normal.

Graficamente, vemos o efeito do crescimento de ***n*** sobre a distribuição amostral a figura que se segue, para ***n*** igual a 10, 30, 50 e 100...

11.1 Distribuições Amostrais



A medida que ***n*** cresce, mais a distribuição se estreita em torno da média

11.1 Distribuições Amostrais

Mais um exemplo:

Numa empresa se aceitará um lote de 1000 peças somente se as mesmas tiverem comprimento médio entre 5 e 10 cm, tomando uma amostra de 10 peças. Sabe-se que o comprimento das peças é uma variável aleatória com distribuição Normal de média 7,5 cm e variância 20 cm.

O que podemos dizer da aceitação do lote de peças?

11.1 Distribuições Amostrais

Mais um exemplo:

Definindo X_i como o comprimento da i-ésima peça retirada, com i indo de 1 até 10, temos que a média das peças a serem retiradas, representada por $\overline{X}$, terá distribuição Normal com média 7,5 cm e variância 20/10 = 2 cm^2. Assim, a probabilidade de aceitar o lote será

$$P(5<\bar{X}<10)=P\left(\frac{5-7,5}{\sqrt{2}}<\frac{\bar{X}-\mu}{\sigma\sqrt{n}}<\frac{10-7,5}{\sqrt{2}}\right)=P(-1,77<Z<1,77)=0,9232$$

ou seja, mais de 92% de probabilidade do lote ser aceito.

Foi usada a tabela para N(0, 1).

11.1 Distribuições Amostrais

De uma maneira geral, não temos informações a respeito da distribuição das variáveis da amostra.

No entanto, sob certas condições, pode ser mostrado que:

Para um tamanho de amostra suficientemente grande, a distribuição de probabilidade média amostral pode ser aproximada por uma distribuição Normal.

Este fato constitui .....

11.2 Teorema Central do Limite

Teorema Central do Limite

Suponha uma amostra aleatória simples de tamanho ***n*** retirada de uma população de média e variância μ, σ^2 sem que se especifique o modelo da variável aleatória. Representando esta amostra por $(X_1, X_2, \cdots, X_n)$ e denotando sua média por $\overline{X}$, temos que

$$Z = \frac{\sum_{i=1}^{N} X_i - n\mu}{\sigma\sqrt{n}}$$

onde ***Z*** ~ N(0,1).

Traduzindo....

11.2 Teorema Central do Limite

Teorema Central do Limite

O teorema garante que para ***n*** suficientemente grande a distribuição da média amostral, devidamente padronizada, se comporta segundo o modelo Normal com média 0 e variância 1.

Com isto, quando estamos trabalhando com a média amostral, o teorema permite que estudemos $\overline{X}$ probabilisticamente usando a distribuição Normal.

Estudos feitos envolvendo simulações indicam que para ***n*** por volta de 30 já são obtidos bons resultados.

11.2 Teorema Central do Limite

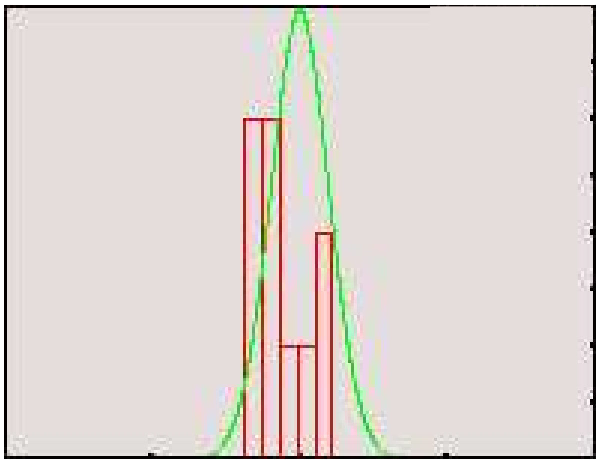
Não demonstramos o teorema no entanto, você pode fazer uma série de simulações que apoiam o resultado enunciado. Já que vocês são do curso de Tecnologia em Sistemas de Computação, usem o computador para fazerem o seguinte procedimento:

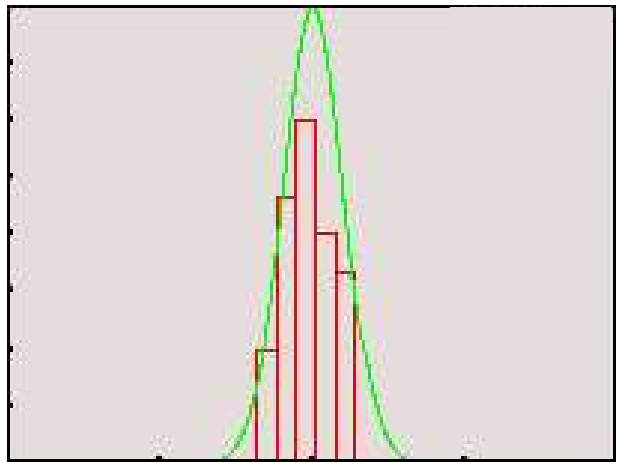
- Simule uma coleta de amostras de um modelo;
- Repita esta coleta um número grande de vezes, calcule as médias amostrais e crie um histograma com estes valores;
- Varie o tamanho da amostra;
- Varie o modelo;
- Compare os gráficos obtidos.

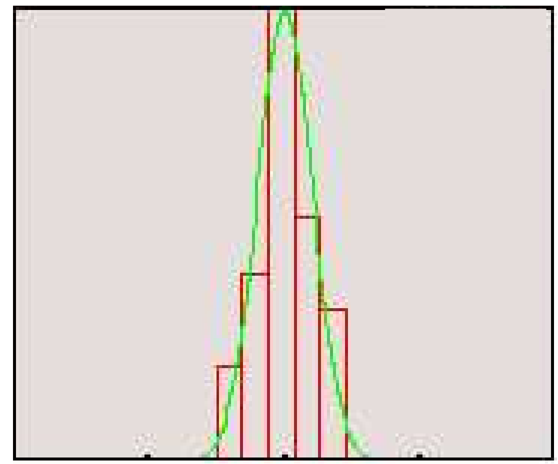
11.2 Teorema Central do Limite

Exemplo computacional

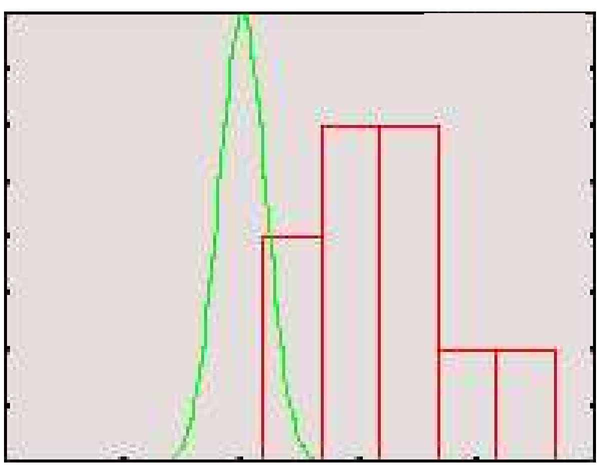
Exponencial

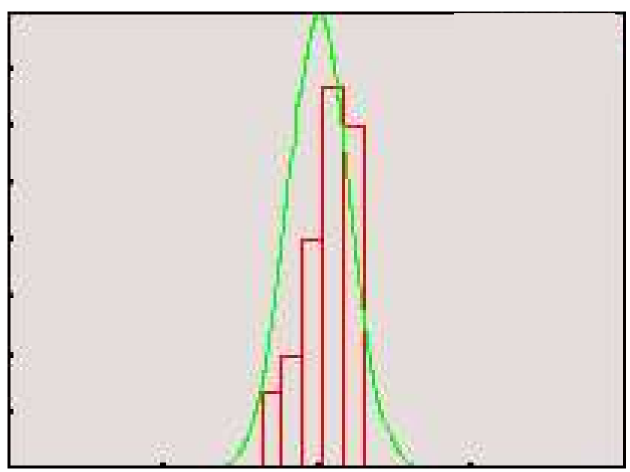

10 amostras

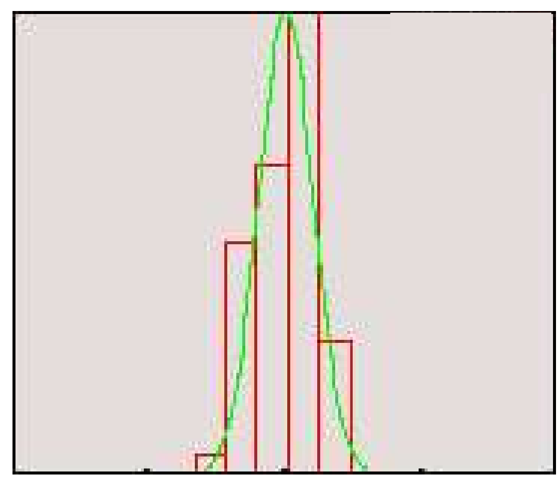

30 amostras


50 amostras

Uniforme


10 amostras

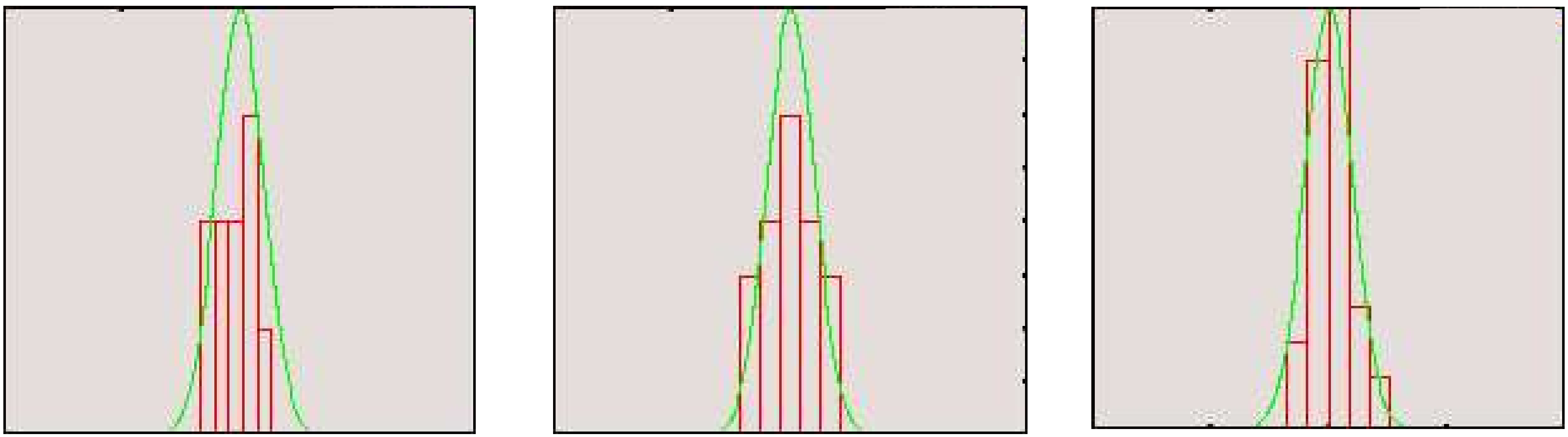

30 amostras


50 amostras

11.2 Teorema Central do Limite

Exemplo computacional

Binomial(10;0,1)



10 amostras 30 amostras 50 amostras

11.2 Teorema Central do Limite

Mais um exemplo:

Uma variável aleatória $\mathbf{X}$, toma os valores 3, 6, e 8 com probabilidades 0,4; 0,3 e 0,3 respectivamente. Uma amostra com 40 observações é sorteada. Não se conhece a distribuição de $\mathbf{X}$ e temos que a média e a variância tem os valores 5,4 e 4,44 respectivamente.Queremos saber qual a probabilidade da média amostral ser superior a 5, ou seja,

$$P(\bar{X} > 5) = P\left(\frac{\bar{X} - 5,4}{\sqrt{4,44/40}} > \frac{5 - 5,4}{\sqrt{4,44/40}}\right) \approx P(Z > -1,20) = 0,8849$$

onde o último resultado foi obtido usando a tabela da N(0, 1).

11.2 Teorema Central do Limite

E mais um exemplo:

Numa certa cidade, a duração das ligações telefônicas seguem o modelo Exponencial com parâmetro 1/3, ou seja X ~exp(1/3). Observando uma amostra de 50 chamadas, qual a probabilidade delas, em média, não ultrapassarem 4 minutos?

Recordando que para o modelo Exponencial

$$\mu = 1/\alpha \text{ e } \sigma^2 = 1/\alpha^2$$

temos que:

$$\mu = E(X) = 3 \text{ e } \sigma^2 = Var(X) = 9$$

11.2 Teorema Central do Limite

$\mathbf{X}$~exp(1/3) $\quad \boldsymbol{\mu}=\mathrm{E}(\mathrm{X})=3$ e $\boldsymbol{\sigma}^2=\mathrm{Var}(\mathrm{X})=9$

Supondo que esta amostra seja suficientemente grande, podemos usar o Teorema Central do Limite para calcular a probabilidade desejada, ou seja,

$$\mathrm{P}(\bar{\mathrm{X}}<4)=\mathrm{P}\left(\frac{\bar{\mathrm{X}}-3}{\sqrt{9/50}}>\frac{4-3}{\sqrt{9/50}}\right)\approx \mathrm{P}(\mathrm{Z}\leq 2{,}36)=0{,}9909$$

ou seja, é praticamente certo que a média amostral das ligações serão menores do que 4 minutos. Poderíamos usar aqui a distribuição de Erlang, mostrada na aula 9. Mas, neste caso, teríamos de ter mais um parâmetro estimado.

Veja a referência [2] Probabilidade: Um Curso Introdutório de Carlos A. B. Dantas.

11.2 Teorema Central do Limite

Distribuição de Proporção amostral

Lembrem que definimos a proporção amostral como

$$\hat{p} = \frac{\text{número de indivíduos na amostra com dada característica}}{n}$$

Se temos que o i-ésimo indivíduo de uma variável aleatória Y_i tal que esta valha 1 se o indivíduo tem esta característica e 0 caso não a tenha. Podemos reescrever a proporção amostral como

$$\hat{p} = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} = \sum_{i=1}^{n} \frac{Y_i}{n} = \bar{Y}$$

Ou seja, a proporção amostral é a média das variáveis aleatórias convenientemente definidas.

11.2 Teorema Central do Limite

Distribuição de Proporção amostral

Com a proporção de indivíduos com a característica na população escrita como ***p*** e que os indivíduos foram selecionados aleatoriamente, temos que $Y_1, \ldots, Y_n$ formam uma seqüência de variáveis aleatórias independentes com distribuição de Bernoulli. Então, como

$$E(Y_i) = p \quad \text{e} \quad Var(Y_i) = p(1-p)$$

$$E(\hat{p}) = E\left(\sum_{i=1}^{n} \frac{Y_i}{n}\right) = p \qquad Var(\hat{p}) = Var\left(\sum_{i=1}^{n} \frac{Y_i}{n}\right) = \frac{p(1-p)}{n}$$

Temos que $\hat{p}$ é um estimador não viciado e consistente com ***p***.

11.2 Teorema Central do Limite

Distribuição de Proporção amostral

Num lote de peças 40% estavam fora de especificação. Qual a probabilidade de acharmos uma proporção de peças menor que 50% numa amostra de 30 peças?

Este problema se encaixa perfeitamente no modelo Binomial e, aproximadamente, pelo modelo Normal. Representando como proporção amostral das peças defeituosas como $\hat{p}$, teremos

Modelo Binomial $W \sim b(30;0,40)$

Modelo Normal com o uso do Teorema Central do Limite

$$\hat{p} \sim N\left(0,40;\frac{0,40(1-0,40)}{30}\right)$$

Obs: O cálculo do modelo Binomial é bem mais complexo que o do modelo Normal.

11.2 Teorema Central do Limite

Distribuição de Proporção amostral

Modelo Binomial $W \sim b(30;0{,}40)$

$$P(\hat{p} < 0{,}50) = P(W/30 < 0{,}50) = P(W < 15) = \sum_{i=1}^{n} \binom{30}{i} 0{,}40^{i} 0{,}60^{30-i} = 0{,}8250$$

Modelo Normal com o uso do Teorema Central do Limite

$$\hat{p} \sim N\left(0{,}40; \frac{0{,}40(1-0{,}40)}{30}\right)$$

$$P(\hat{p} < 0{,}50) = P\left(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0{,}50-0{,}40}{\sqrt{\frac{0{,}40(1-0{,}40)}{30}}}\right) = P(Z < 1{,}12) = 0{,}8686$$

Com uma amostra de apenas tamanho 30 obtemos um resultado aproximado muito bom.

11.3 Estimação por intervalo

Os estimadores vistos até agora fornessem um único valor numérico para os parâmetros de interesse. São estimadores pontuais.

É uma informação útil se soubermos a medida de precisão do valor obtido.

Veremos agora um método de estimação chamado intervalo de confiança que incorpora estimativa pontual do parâmetro e informações a respeito de sua variabilidade.

11.3 Estimação por intervalo

Intervalo de confiança

Vejamos o caso do intervalo de confiança para a média numa população Normal da qual conhecemos a média e a variância, μ, σ^2 .

Ainda supomos ter uma amostra de tamanho ***n*** dada por $(X_1, X_2, \cdots, X_n)$. Já vimos que a média amostral tem uma distribuição Normal com média e variância dadas por $\mu, \sigma^2/n.$

Assim temos,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Mas qual o intervalo de confiança deste valor?

11.3 Estimação por intervalo

Intervalo de confiança

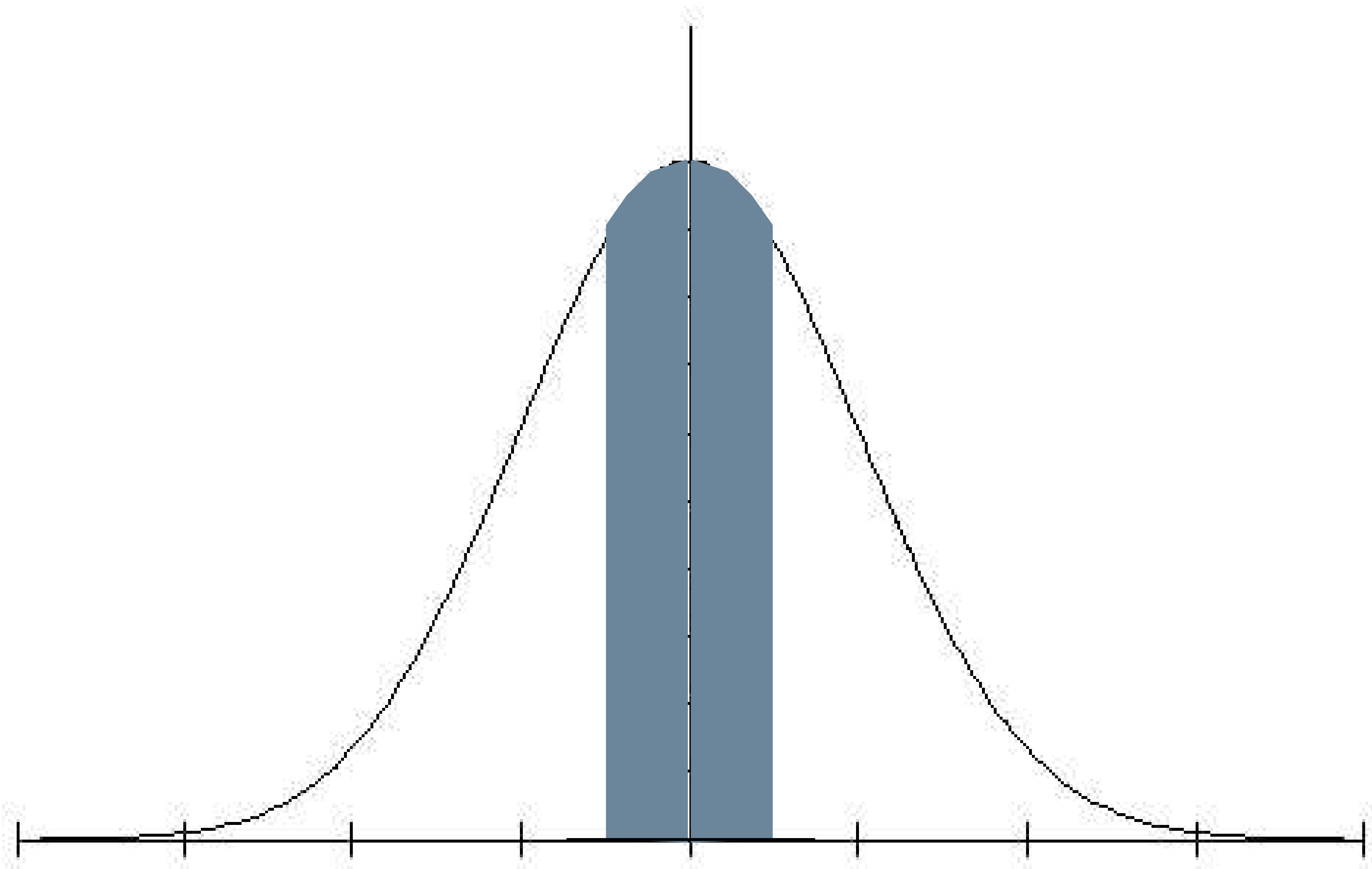
$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

Vamos introduzir um valor tal que $0<\gamma<1$. Podemos achar um valor $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ tal que

$$P(|Z|<z_{\gamma/2})=P(-z_{\gamma/2}<Z<z_{\gamma/2})=\gamma$$

O índice de $z_{\gamma/2}$ apresenta o valor de γ dividido por 2 uma vez que este valor deve ser igualmente distribuido em torno da origem. Ou num gráfico ....

11.3 Estimação por intervalo



11.3 Estimação por intervalo

O valor $z_{\gamma/2}$ pode ser obtido da tabela da Normal padrão, localizando o valor de $\gamma/2$ na tabela e obtendo o valor de $z_{\gamma/2}$ nas margens correspondentes. Feito isto, temos o intervalo

$$- z_{\gamma/2} < z < z_{\gamma/2} \Rightarrow - z_{\gamma/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_{\gamma/2}$$

que pode ser reescrito como

$$\bar{X} - z_{\gamma/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\gamma/2} \frac{\sigma}{\sqrt{n}}$$

11.3 Estimação por intervalo

Daqui então

$$\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}$$

podemos tirar o intervalo de confiança para a média com coeficiente de confiança γ, sendo dado por

$$IC(\mu, \gamma) = \left[\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right]$$

Como interpretar esta definição?

11.3 Estimação por intervalo

Intervalo de confiança

Como interpretar esta definição?

$$IC(\mu, \gamma) = \left[\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right]$$

- Envolve $\overline{X}$ que é variável aleatória logo, o intervalo também é aleatório;
- A probabilidade deste intervalo conter μ é dada por γ;
- Ao coletar a amostra, $\overline{X}$ torna-se $\overline{X}_{obs}$ e, conhecidos

σ, $\boldsymbol{n}$, $z_{\gamma/2}$ o intervalo passa a ser numérico.

No resumo....

11.3 Estimação por intervalo

Intervalo de confiança

Se obtivermos várias amostras de mesmo tamanho e, para cada uma delas, calcularmos os correspondentes intervalos de confiança com coeficiente de confiança γ, esperamos que a proporção de intervalos que contenham o valor de μ seja igual a γ.

11.3 Estimação por intervalo

Suponha que os comprimentos de jacarés adultos de uma determinada raça siga o modelo Normal com média μ desconhecida e variância igual a 0,01 m^2. Uma amostra de 10 animais deu a média 1,69m. Estime γ.

A probabilidade de $\overline{X}$ é normal com a média μ e a variância é $\sigma^2 = 0{,}01/10 = 0{,}001\ m^2$.

Vamos obter o intervalo de confiança para μ com coeficiente γ.

Estabelecendo $\gamma = 95\%$ obtemos da tabela Normal

$$z_{\gamma/2} = z_{0{,}475} = 1{,}96$$

11.3 Estimação por intervalo

Intervalo de confiança

$$IC(\boldsymbol{\mu},95\%)=\left[1,69-1,96\sqrt{\frac{0,01}{10}};1,69+1,96\sqrt{\frac{0,01}{10}}\right]=[1,63;1,75]$$

Adotando a interpretação dada na tela anterior temos que estes cálculos indicam que:

- ao construirmos 100 intervalos, 95 conteriam a verdadeira média;
- a verdadeira média estaria dentro do intervalo [1,63; 1,75], uma amplitude de 12 cm, relativamente pequena em comparação ao valor presumido da média.

11.3 Estimação por intervalo

Intervalo de confiança

$$IC(\mu, 95\%) = \left[1{,}69 - 1{,}96\sqrt{\frac{0{,}01}{10}}\,;\, 1{,}69 + 1{,}96\sqrt{\frac{0{,}01}{10}}\right] = [1{,}63; 1{,}75]$$

Adotando a interpretação dada na tela anterior temos que estes cálculos indicam que:

- ao construirmos 100 intervalos, 95 conteriam a verdadeira média;
- a verdadeira média estaria dentro do intervalo [1,63; 1,75], uma amplitude de 12 cm, relativamente pequena em comparação ao valor presumido da média.

Obs: O tamanho de jacarés, e de quaisquer seres vivos, não seguem rigidamente o modelo Normal. Não existe a possibilidade de acharmos um jacaré de 1km de comprimento, embora seja possível (mas improvável) pelo modelo Normal.

11.3 Estimação por intervalo

A amplitude de confiança é dada formalmente como

$$\bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}} - \left(\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right) = 2z_{\gamma/2}\frac{\sigma}{\sqrt{n}}$$

Observe que esta expressão depende da:

- confiança γ ;
- do desvio padrão σ ;
- do tamanho da amostra ***n***.

Logo......

11.3 Estimação por intervalo

Amplitude de confiança

$$\bar{X}+z_{\gamma/2}\frac{\sigma}{\sqrt{n}}-\left(\bar{X}-z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right)=2z_{\gamma/2}\frac{\sigma}{\sqrt{n}}$$

A medida que aumentamos a confiança, o valor de $z_{\gamma/2}$ fica maior e a amplitude do intervalo aumenta pois com intervalos maiores temos maior probabilidade de incluirmos a verdadeira média.

A medida que aumentamos o número de amostras, diminui o intervalo, aumentando assim a probabilidade de termos um valor mais próximo da média verdadeira.

11.3 Estimação por intervalo

Um exemplo

Queremos saber a vida média de uma bateria de uma certa marca. Baseado em estudos de outros fabricantes há indícios que a vida de baterias seguem o modelo Normal com desvio padrão de 4,5 meses.

Qual o tamanho da amostra deverá ter para que a amplitude de confiança seja de 90% para uma vida média de 3 meses?

Para achar ***n*** vamos usar a equação

$$2z_{\gamma/2}\frac{\sigma}{\sqrt{n}}=3$$

11.3 Estimação por intervalo

Um exemplo

Para achar ***n*** vamos usar a equação

$$2z_{\gamma/2}\frac{\sigma}{\sqrt{n}}=3$$

Para $2z_{\gamma/2}$ = 1,64, correspondente à γ = 0,90 e como $\sigma = 4{,}5$, temos

$$\sqrt{n}=2z_{\gamma/2}\frac{\sigma}{3}=\frac{2\times 1{,}64\times 4{,}5}{3}=4{,}92\approx 5$$

Assim a amostra deverá ser de tamanho 25 e como sobrevalorizamos n, a amplitude do intervalo será ligeiramente menor que 3 e, portanto, mais informativo.

11.3 Estimação por intervalo

Mais um exemplo

Um provedor de acesso à Internet está monitorando a duração do tempo das conexões de seus clientes. O que se pretende é avaliar a necessidade de expansão dos equipamentos. Não se conhece a média e a distribuição de probabilidade mas o desvio padrão, por analogia a outros serviços, é igual a $\sqrt{50}$ minutos.

Uma amostra de 500 conexões resultou num valor médio observado de 25 minutos.

Qual a verdadeira média com confiança de 92%?

11.3 Estimação por intervalo

Mais um exemplo

Como não se conhece a média e a distribuição de probabilidade, usaremos o Teorema Central do Limite já que a amostra é de tamanho 500.

Já vimos exemplos onde tamanhos menores nos davam resultados úteis. Assim, para um coeficiente de confiança de 92%, obtemos o intervalo de confiança como abaixo

$$\mathrm{IC}(\boldsymbol{\mu}, 92\%) \approx \left[\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right] \approx$$

$$\approx \left[25 - 1{,}75\sqrt{\frac{50}{500}}; 25 + 1{,}75\sqrt{\frac{50}{500}}\right] = [24{,}45; 25{,}55]$$

com uma amplitude de apenas 1,1!

11.3 Estimação por intervalo

Um exemplo médico

Num experimento, pretende-se estimar a proporção ***p*** de cura pelo uso de um medicamento. Aleatoriamente, foi selecionada uma amostra de 200 pacientes e pretende-se que 160 pacientes estejam curados depois do tratamento. Como estimar a partir do experimento a proporção ***p*** da população como um todo?

Uma estimativa pontual para ***p*** é dada por $\hat{p}_{obs}=160/200=0,8$

Embora não haja indícios sobre a distribuição da proporção pontual, o Teorema Central do Limite garante que, para um número grande o suficiente, podemos aproximar a distribuição desconhecida com a distribuição Normal.

11.3 Estimação por intervalo

Um exemplo médico

Temos, então, como aproximação,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Continuando com nossas hipóteses, vamos examinar um intervalo de confiança com coeficiente $\gamma = 0,95$ ($z_{\gamma/2}$=1,96) e , daí, calcularmos o intervalo de confiança,

$$IC(\boldsymbol{\mu},95\%) = \left[\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}};0,8+1,96\sqrt{\frac{p(1-p)}{n}}\right] =$$

$$= \left[\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}};0,8+1,96\sqrt{\frac{p(1-p)}{n}}\right]$$

Obs: ***p*** é desconhecido e, portanto, o cálculo é impossível. Temos que fazer mais hipóteses!

11.3 Estimação por intervalo

Um exemplo médico

Mais uma hipótese: Vamos substituir ***p***(1 - ***p***) por $\hat{p}_{obs}(1-\hat{p}_{obs})$, ou seja, estamos usando a estimativa pontual para substituir ***p***.

Com mais esta suposição, podemos calcular o intervalo como:

$$IC_2(p, 95\%) = \left[0{,}8 - 1{,}96\sqrt{\frac{0{,}8}{200}}; 0{,}8 + 1{,}96\sqrt{\frac{0{,}8\times 0{,}2}{200}}\right] = [0{,}745; 0{,}855]$$

Mas esta não é a única abordagem possível.....

11.3 Estimação por intervalo

Um exemplo médico

Outra hipótese: Observando que ***p***(1 - ***p***) tem valor máximo ¼ para 0 ≤ ***p*** ≤ 1, usando este valor numérico, obtemos

$$IC_2(p, 95\%) = \left[0{,}8 - 1{,}96\sqrt{\frac{1/4}{200}} ; 0{,}8 + 1{,}96\sqrt{\frac{1/4}{200}}\right] = [0{,}731 ; 0{,}869]$$

que tem amplitude superior ao caso anterior.

Como interpretar estes valores?

11.3 Estimação por intervalo

Um exemplo médico

Como interpretar estes valores?

IC_1 é usualmente denominada de abordagem otimista. Parte da crença de que a estimativa de ***p*** é próxima o suficiente da realidade para termos também uma boa aproximação da variância ***p***(1 - ***p***)/***n***.

IC_2 é denominada de abordagem conservativa porque preferimos substituir a variância por um valor claramente superior ao real. Assim estamos assegurando que o coeficiente de segurança será no mínimo.

11.3 Estimação por intervalo

Um exemplo médico

Como interpretar estes valores?

Lembrando que a variância de um estimador é uma medi-da de sua precisão...

Ao adotarmos IC_2 (conservativo) estamos aceitando uma amplitude maior para o intervalo.

11.3 Estimação por intervalo

Num resumo:
Intervalos de confiança para a média e proporção

μ $$\left[\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right]$$

p $$\left[\hat{p} - z_{\gamma/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \hat{p} + z_{\gamma/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$ (otimista)

p $$\left[\hat{p} - z_{\gamma/2}\sqrt{\frac{1}{4n}} \hat{p} + z_{\gamma/2}\sqrt{\frac{1}{4n}}\right]$$ (conservativo)

Aula 11

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Inferência Estatística - Estimação II

Conteúdo:

- 11.1 Distribuições Amostrais
- 11.2 Teorema Central do Limite
- 11.3 Estimação por Intervalo

$$\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\sigma_X^2 \qquad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$A \cap B = \varnothing$$

Fundação CECIERJ Consórcio cederj