AULA:

Inferência Estatística

Prof. Víctor Hugo Lachos Dávila

Inferência Estatística

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

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

Problemas da Inferência

Exemplo: Qual a distribuição da altura dos brasileiros adultos?. Parece razoável pensar num modelo Normal, a questão agora é identificar os parâmetros (μ e σ^2) para que ela fique completamente especificada. Como fazer isso?

- Medindo a altura de todos os Brasileiros adultos. Neste caso não é necessário usar Inferência Estatística!
- Escolher estrategicamente uma amostra $(X_1, X_2, \ldots, X_n)$ da população de adultos e através dessa amostra inferir sobre os parâmetros (μ e σ^2) da população.
- Os resultados dependeram da qualidade da amostra. Esta tem que ser representativa da população.
- Descrevemos aqui um dos problemas básicos da Inferência estatística: Estimação

Problemas da Inferência

Exemplo: suponha agora que desejamos saber se a média da altura dos brasileiros é maior que a dos argentinos (1,65m)?

- Para tomarmos uma decisão, escolhemos estrategicamente uma amostra $(X_1, X_2, \ldots, X_n)$ da população de adultos e analisamos se $\mu > 1{,}65$ com alta probabilidade.

- Descrevemos aqui um outro problema básico da Inferência estatística: Teste de Hipóteses

Estimação	Teste de Hipóteses
•Qual é a probabilidade de "cara"no lançamento de uma moeda?	•A moeda é honesta ou é desequilibrada?
•Qual é a proporção de votos que o candidato A tem nas eleições?	•O candidato A vencerá as eleições ?
•Qual é a proporção de motoristas que tiveram sua carteira apreendida após a vigência da nova lei de trânsito?	•Pelo menos 2% dos motoristas habilitados de SP tiveram suas carteiras apreendidas após a entrada da nova lei do trânsito ou não?

Como Selecionar uma Amostra

- Ex1: Análise da quantidade de glóbulos brancos na sangue de certo indivíduo. Uma gota do dedo seguramente será representativa para a análise. Caso Ideal!
- Ex2: Opinião sobre um projeto governamental. Se escolhemos uma cidade favorecida o resultado certamente conterá erro (viés).

Note que a maneira de se obter a amostra é muito importante. A Tecnologia da AMOSTRAGEM é uma das especialidades dentro da estatística que fornece procedimentos adequados.

Aqui trataremos o caso mais simples e que serve de base para procedimentos muito mais elaborados: Amostragem aleatória simples (AAS)

AAS

- Supomos que podemos listar todos os N elementos da população (população finita).
- Usando métodos de geração de números aleatórios, sorteia-se um elemento da população, sendo que todos os elementos tem a mesma chance de ser selecionados.
- Repete-se o procedimento até que sejam sorteadas as "n" unidades da amostra.
- Temos AAS com reposição e sem reposição.
- AAS com reposição implica que tenhamos independência entre as unidades selecionadas, facilitando o estudo das propriedades dos estimadores. Logo, nestas notas:

AAS ≈ AAS com reposição

Algumas Definições

Definição: Uma ***amostra aleatória simples*** (a.a) de tamanho n de uma v.a. X, é o conjunto de n v.a's independentes $(X_1, X_2, \ldots, X_n)$, cada uma com a mesma distribuição de X.

Definição: As quantidades da população, em geral desconhecidas, sobre as quais temos interesse, são denominadas ***parâmetros***. θ, μ, σ^2

Definição: A combinação de elementos da amostra, construída com a finalidade de estimar um parâmetro, é chamado de ***estimador***, exemplo, $\overline{X}$ Aos valores numéricos assumidos pelos estimadores chamamos de ***estimativas*** exemplo, $\overline{x}$

Definição: Chamamos de ***estatística*** a qualquer função T da amostra aleatória, i.e.

$$T = T(X_1, X_2, \ldots, X_n)$$

Exemplo: Estamos interessados na média (μ) e variância (σ²) das alturas de jovens com idade entre 15 e 18 anos de certa cidade. Vamos coletar uma amostra para tirar conclusões. Suponha que escolhemos ao acaso 10 jovens (AAS).

- Possíveis *estimadores para* μ *(que* por sua vez são estatísticas)

$$\hat{\mu}_1 = t_1(X_1,\dots,X_{10}) = \frac{(Min+Max)}{2}; \hat{\mu}_2 = t_2(X_1,\dots,X_{10}) = X_1; \hat{\mu}_3 = t_3(X_1,\dots,X_{10}) = \frac{X_1+\dots+X_{10}}{10} = \overline{X};$$

- Possíveis *estimadores para* σ²

$$\hat{\sigma}^2{}_1 = t_4(X_1,\dots,X_{10}) = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2; \hat{\sigma}^2{}_2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2; \hat{\sigma}^2{}_3 = \left(\frac{Max-Min}{2}\right)^2$$

- *Agora temos a amostra observada:* (em metros) 1,65;1.57;1,72;1,66;1,71;1,74;1,81;1,68;1,60;1,77. *As* ***estimativas*** seriam:

$$\hat{\mu}_1 = \frac{(1{,}57+1{,}81)}{2} = 1{,}69; \quad \hat{\mu}_2 = 1{,}65_1; \quad \hat{\mu}_3 = \frac{1{,}65+\dots+1{,}77}{10} = 1{,}69;$$

$$\hat{\sigma}^2{}_1 = 0{,}005; \quad \hat{\sigma}^2{}_2 = s^2 = 0{,}006; \quad \hat{\sigma}^2{}_3 = 0{,}014$$

Propriedades dos estimadores

- **Definição: Um estimador $\hat{\theta}$ é não viciado para um parâmetro θ se $E(\hat{\theta}) = \theta$**

- **Definição: Um estimador $\hat{\theta}$ é consistente, se, a medida que o tamanho de amostra aumenta, seu valor esperado converge para o parâmetro de interesse e sua variância converge para zero. i.e.**

$$i) \lim_{n \to \infty} E(\hat{\theta}) = \theta$$

$$ii) \lim_{n \to \infty} Var(\hat{\theta}) = 0$$

Observe que na definição de consistência estamos supondo que o estimador depende do tamanho de amostra *n*. Na definição de vício o resultado vale para qualquer que seja *n*.

Exemplo: Considere que uma certa característica X, na população tem media μ e variância σ^2. Uma amostra aleatória simples (a.a.) de tamanho n, representado por $(X_1,...,X_n)$ é obtida para estimar μ. Estude as propriedades da media amostral.

Claro que $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ e que os X_i são independentes, $i = 1,\dots,n$.

O estimador da media populacional μ, es da forma $\hat{\mu} = \bar{X}$. *Logo*

$$E(\hat{\mu}) = E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{n\mu}{n} = \mu$$

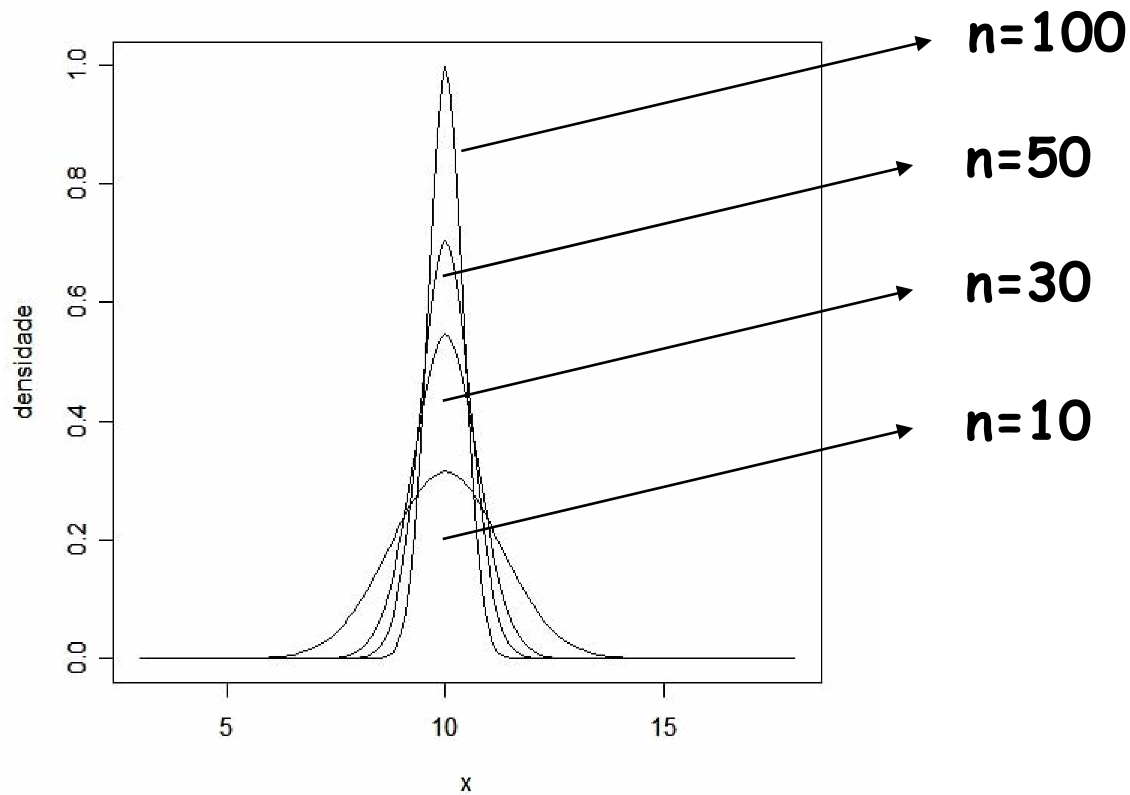
$$Var(\hat{\mu}) = Var(\bar{X}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

Portanto, a média amostral é um estimador não viciado para a média populacional μ e como sua variancia tende a zero conforme n cresce, concluímos também que é um estimador consistente para μ.

- Se o interesse é estimar σ^2. Estude as propriedades de

$$\hat{\sigma}^2{}_1 \quad \text{e} \quad \hat{\sigma}^2{}_2 = S^2$$

Exemplo: **Considere uma a.a. $(X_1,...,X_n)$ de uma variável X~N(10,16). Como se comporta $\overline{X}$ em função de n.**



À medida que *n* aumenta, a f.d.p. vai se concentrando ao redor da média populacional 10. Quanto maior o tamanho de amostra maior probabilidade que uma estimativa de $\overline{X}$ este próxima da média populacional.

Estimadores para a média, proporção e Variância

Parâmetro	Esimador	Propriedades
μ	$\overline{X}$	Não viciado e consistente
p	$\hat{p} = \frac{\text{No de casos favoraveis à carateristica}}{\text{n}}$	Não viciado e consistente
σ^2	$S^2 = \frac{1}{n-1}(\sum X^2{}_i - n\overline{X}^2)$	Não viciado e consistente
σ^2	$\hat{\sigma}^2 = \frac{1}{n}(\sum X^2{}_i - n\overline{X}^2)$	Viciado e consistente

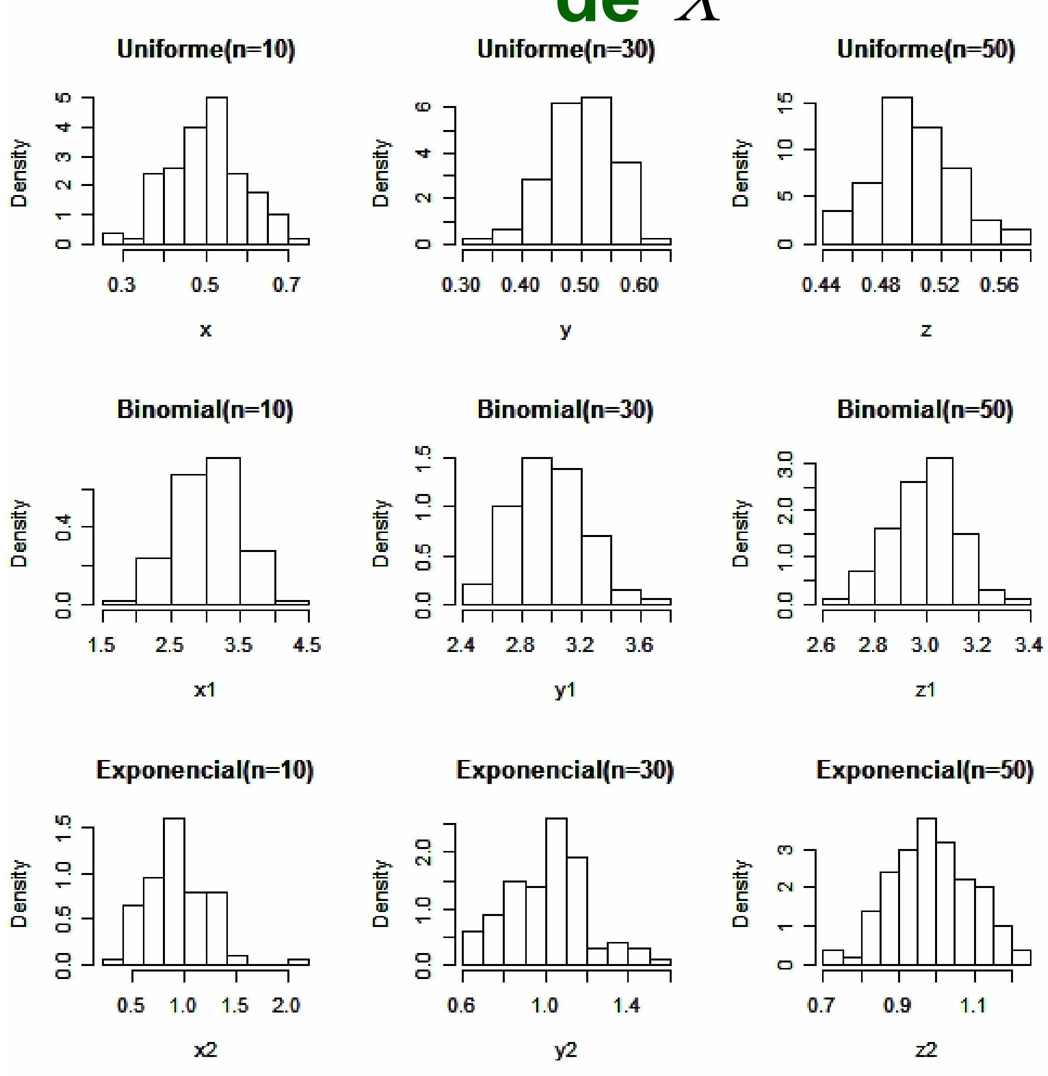
Teorema Limite Central (TLC)

Suponha que uma amostra aleatória simples (X_1,...X_n) é retirada de uma população com média μ e variância σ^2 . Então, temos que

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \approx N(0,1), \quad \text{quando n} \uparrow \infty$$

- Em palavras o TLC garante que para *n* grande a distribuição da média amostral, devidamente padronizada, se comporta segundo um modelo Normal padronizado (Z).
- Em casos onde a verdadeira distribuição dos dados é simétrica, boas aproximações são obtidas para n ao redor de 30.
- Um estudo de simulação descreve graficamente o comportamento de $\bar{X}$ para diferentes situações. X~U(0,1), X~Bin(10,0,3) e X~Exp(1)

Efeito do tamanho de amostra sobre a distribuição de $\bar{X}$



Exemplo: Numa certa cidade, a duração de conversas telefônicas em minutos, segue um modelo Exponencial com parâmetro 3. Observando-se uma amostra aleatória de 50 dessas chamadas, qual será a probabilidade de em média, a duração de conversas telefônicas não ultrapassarem 4 minutos.

$Seja\ X : duração\ das\ chamadas,\quad \mathrm{X \sim Exp(3)}.\quad \mathrm{Logo\ \ E(X) = 3\ e\ Var(X) = 9}$

Admitindo que n é grande o suficiente, podemos calcular a probabilidade desejada da seguinte forma:

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

O Caso da Proporção Amostral ($\hat{p}$)

Coletamos uma a.a. (X_1,...X_n) de X~Bernoulli(p), com o objetivo de estimar *p*. Definimos a proporção amostral (estimador de p) como sendo a fração de indivíduos com a característica X , i.e.,

$$\hat{p} = \frac{\text{No de casos favoraveis à carateristica}}{\text{n}}$$

Note que podemos escrever

$$\hat{P} = \frac{X_1 + X_2 + \ldots X_n}{n} = \overline{X}, \qquad \mathrm{X_i} = \begin{cases} 1, & \text{sucesso} \\ 0, & \text{fracaso} \end{cases}$$

Assim, temos que

$$E(\hat{P}) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n} = \frac{np}{n} = p \qquad Var(\hat{P}) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Pelo TLC

$$\frac{\overline{X} - E(\overline{X})}{Var(\overline{X})} = \frac{\overline{X} - p}{\sqrt{p(1-p)/n}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1)$$

Exemplo: A proporção de peças fora de especificação num lote é de 0,4. Numa amostra de tamanho 30, calcule a probabilidade de que a proporção de peças defeituosas seja menor do que 0,5.

Seja $\hat{p}$: *a proporção de peças defeituosas na amostra* (*proporção amostral*).

Então,

Como conseqüência do TLC, temos que

$$\frac{\hat{p} - E(\hat{p})}{\sqrt{Var(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1), \quad \text{quando} \quad n \uparrow \infty$$

$$\hat{p} \sim N(0{,}40, \frac{0{,}40(0{,}6)}{30}), \quad \text{Assim,}$$

$$P(\hat{p} < 0{,}5) = P\left(\frac{\hat{p} - 0{,}4}{\sqrt{\frac{0{,}40(0{,}6)}{30}}} < \frac{0{,}5 - 0{,}4}{\sqrt{\frac{0{,}40(0{,}6)}{30}}}\right) \approx P(Z \le 1{,}12) = 0{,}8686$$

Definição**[Intervalo de Confiança]** Seja $X_1,\ldots,X_n$ uma amostra aleatória de uma população com a característica $X\sim f(x,\theta)$. Seja $T_1=G(X_1,\ldots,X_n)$ e $T_2=H(X_1,\ldots,X_n)$ duas estatísticas tais que $T_1< T_2$ e que

$$P(T_1 < \theta < T_2) = 1-\alpha.$$

O intervalo (T_1, T_2) é chamado de intervalo de $100(1-\alpha)\%$ de confiança para θ.

Notação: $IC(\mu,1-\alpha)= (T_1, T_2)$, onde T_1 e T_2 são os limite inferior superior respectivamente e $1-\alpha$ é o coeficiente (ou nível) de confiança

Intervalo de confiança para uma média populacional

Suponha que $X_1, \cdots X_n$ é uma amostra aleatória de tamanho n, de uma população normal com média μ (desconhecida) e variância σ^2(conhecida). Vimos que a média amostral $\overline{X}$, tem distribuição normal com média μ e variância σ^2/n. Isto é

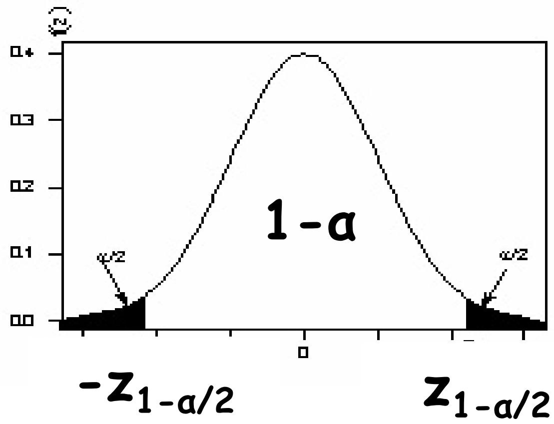
$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Logo, fixando um nível de confiança $(1-\alpha)$, pode-se determinar $z_{\alpha/2}$ de tal forma:.

$$P(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Ou que é equivalente

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



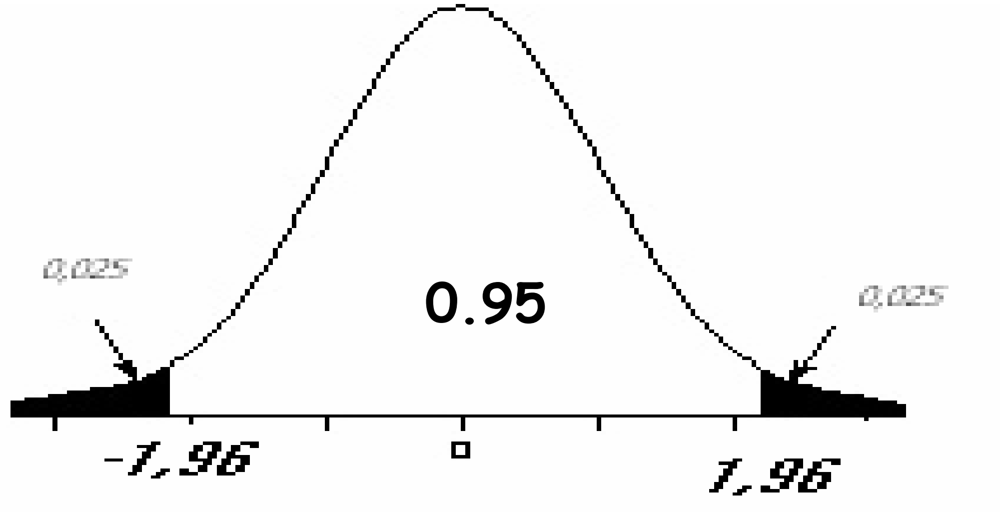
$$-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}} \Leftrightarrow \overline{X} - \overbrace{z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}^{E} \leq \mu \leq \overline{X} + \overbrace{z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}^{E}$$

Logo, intervalo de 100 (1-α)% de confiança para μ é dado por:.

$$IC(\mu, 1-\alpha) = \left(\overline{X} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} ; \overline{X} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \right) = \left(\overline{X} - E ; \overline{X} + E \right)$$

Exemplo 1: Em uma industria de cerveja, a quantidade de cerveja inserida em latas tem-se comportado como uma distribuição normal com média 350 ml e desvio padrão 3 ml. Após alguns problemas na linha de produção, suspeita-se que houve alteração na média. Uma amostra de 20 latas acusou uma média 346 ml. Obtenha um intervalo de 95% para a quantidade média μ de cerveja inserida em latas, supondo que não tenha ocorrido alteração na variabilidade.

Já que 1-α=0,95, temos da tabela normal padrão $z_{0,975}$=1,96.



$$IC(\mu,0{,}95) = \left(\bar{X} - 1{,}96 \times \frac{\sigma}{\sqrt{n}} ; \bar{X} + 1{,}96 \times \frac{\sigma}{\sqrt{n}}\right)$$

$$IC(\mu,0{,}95) = \left(346 - 1{,}96 \times \frac{3}{\sqrt{20}} ; 346 + 1{,}96 \times \frac{3}{\sqrt{20}}\right) = (346 - 1{,}31; 346 + 1{,}31)$$

$$= (344{,}69 ; 347{,}31)$$

O erro máximo de estimação na estimação de μ é dado por

$$E = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow n = \frac{z^2_{1-\alpha/2}\sigma^2}{E^2}$$

No caso de população finita de N elementos é introduzida o fator de correção de população finita

$$E = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad \Rightarrow n = \frac{Nz^2_{1-\alpha/2}\sigma^2}{E^2(N-1) + z^2_{1-\alpha/2}\sigma^2}$$

Exemplo: Uma firma construtora deseja estimar a resistência média das barras de aço utilizadas na construção de casas. Qual o tamanho amostral necessário para garantir que haja um risco de 0,001 de ultrapassar um erro de 5 kg ou mais na estimação ? O desvio padrão da resistência para este tipo de barra é de 25 kg.

Do enunciado tem-se σ=25, α=0,001, e E=5, $z_{0,9995}$=3,29

$$E = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = 5 \qquad \Rightarrow n = \frac{(3,29)^2 (25)^2}{5^2} = 270,6025 \approx 271$$

Intervalo de confiança para uma média populacional quando σ é desconhecido

A distribuição t-Student

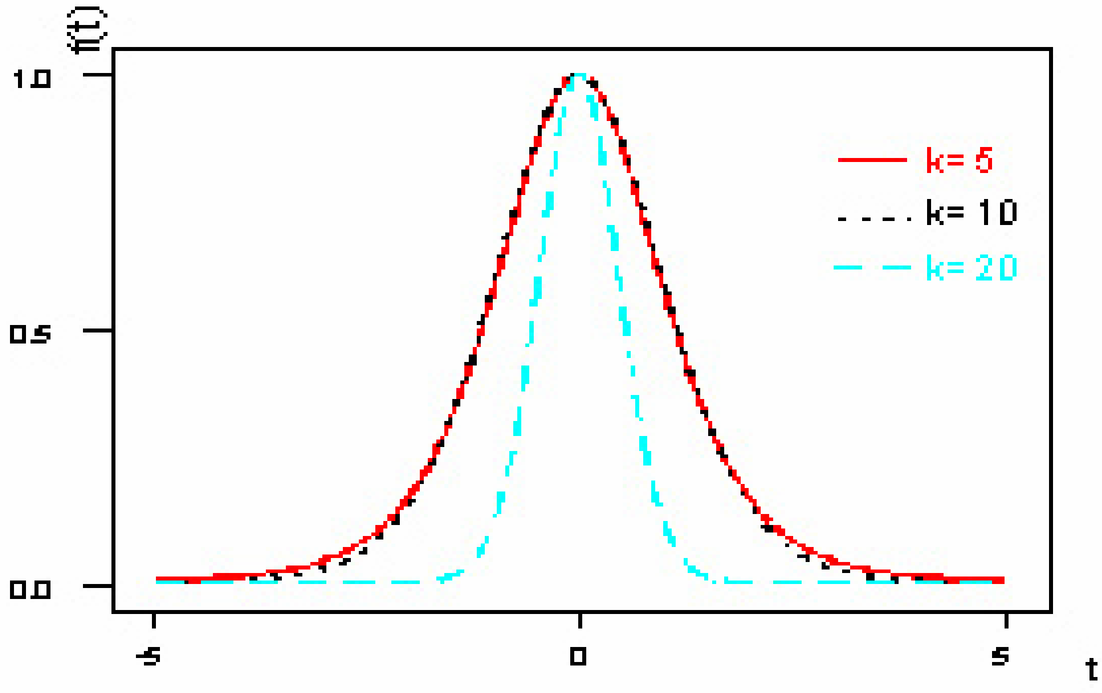
Supondo que a característica de interesse da população é normal, a estatística

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \qquad (1)$$

tem distribuição de probabilidade conhecida com distribuição t de Student com n-1 graus de liberdade.

A função de densidade de um v.a t-Student com k graus de liberdade é dado por:.

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)(\pi k)^{1/2}}\left(1+\frac{t^2}{k}\right)^{-(k+1)/2} \quad ; \quad t \in R$$



Notação; T~t(k), indica que v.a tem distribuição t-Student com k graus de liberdade.

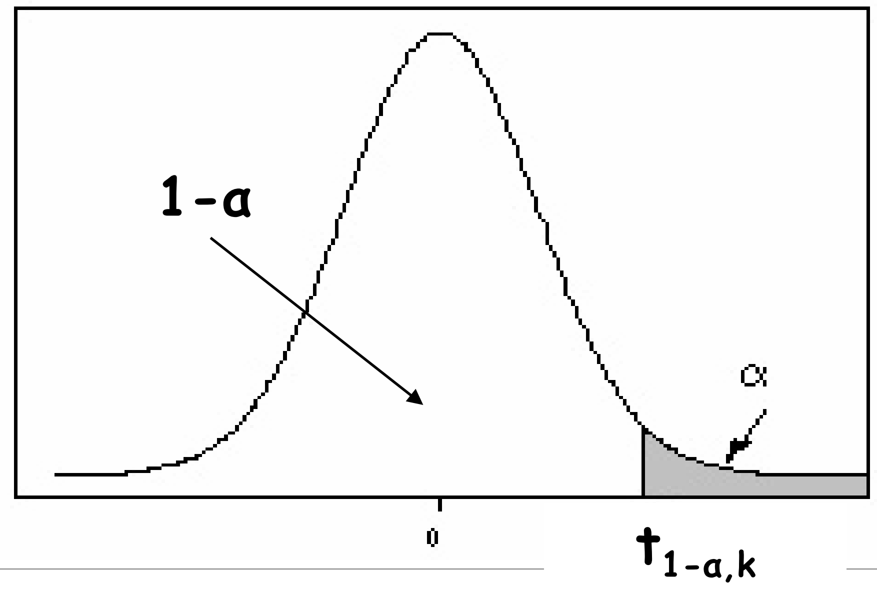
*Propriedades: se **T~t(k)***

$$(i)\ E(T) = 0;\ Var(T) = \frac{k}{k-2},\ k > 2$$

$$(ii) k \to \infty \Rightarrow T \sim N(0,1)$$

Uso Da Tabela Distribuição t-Student

$$P(T \leq t_{1-\alpha,k}) = 1 - \alpha$$



Considerando a estatística dada em (1), pode-se mostrar que um intervalo de 100(1-α)% de confiança para μ é dado por:

$$IC(\mu, 1-\alpha) = \left(\bar{X} - \underbrace{t_{1-\alpha/2, n-1} \times \frac{S}{\sqrt{n}}}_{E} ; \bar{X} + \underbrace{t_{1-\alpha/2, n-1} \times \frac{S}{\sqrt{n}}}_{E} \right) = \left(\bar{X} - E; \bar{X} + E \right)$$

***Exemplo 3:* Deseja-se avaliar a dureza esperada μ do aço produzido sob um novo processo de têmpera. Uma amostra de 10 corpos de prova de aço produziu os seguintes resultados, em HRc:**

36,4 35,7 37,2 36,5 34,9 35,2 36,3 35,8 36,6 36,9

Construir um intervalo de confiança para μ, com nível de confiança de 95%.

$$\overline{X}=\frac{1}{10}\sum_{i=1}^{10}X_i=36{,}5;\quad S=\sqrt{\frac{\sum_{i=1}^{10}\left(X_i-\overline{X}\right)^2}{n-1}}=0{,}7352;\ \frac{S}{\sqrt{n}}=0{,}2325$$

Já que, n=10 (1-α)=0,95,$\rightarrow$ α=0,05, temos: $t_{0,975,\,9}$=2,26

$$E=(2{,}26)(0{,}2325)=0{,}53$$

$$IC(\mu,0{,}95)==\left(\overline{X}-E;\overline{X}+E\right)$$

$$IC(\mu,0{,}95)==\left(36{,}5-0{,}53;\ 36{,}5+0{,}53\right)=\left(35{,}97;\ 37{,}03\right)$$

A distribuição Qui-quadrado

Supondo que a característica de interesse da população é normal, a estatística

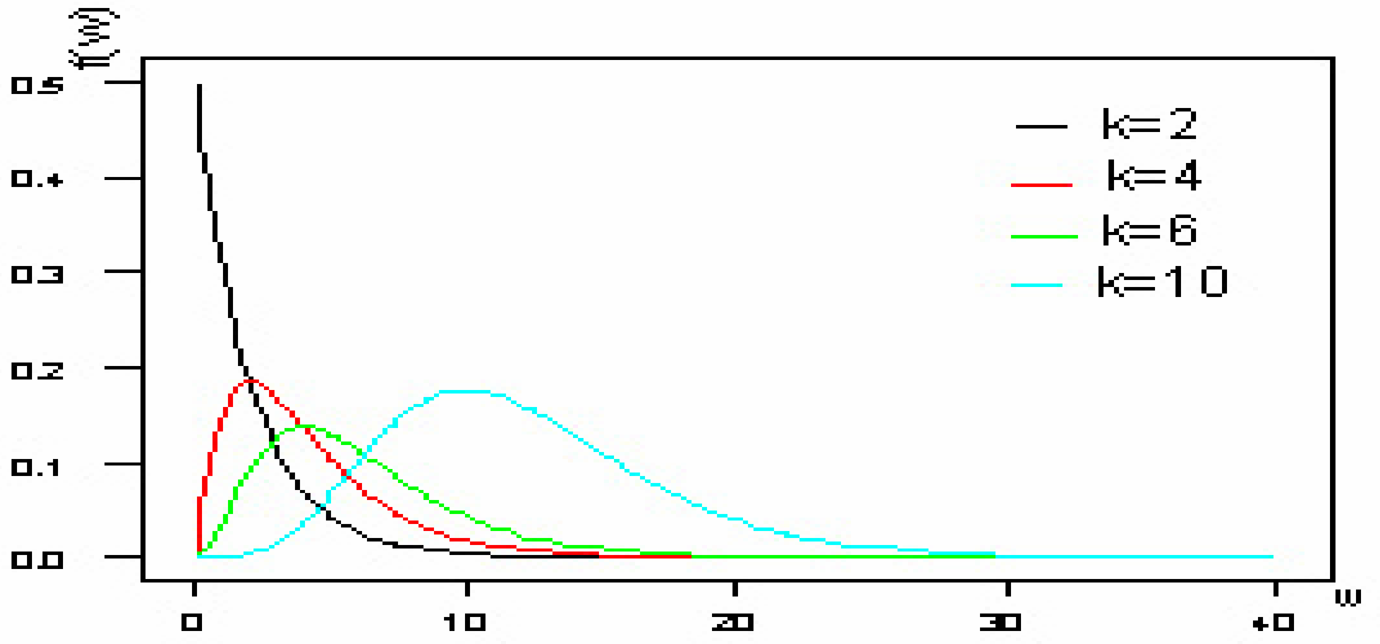
$$W = \frac{(n-1)S^2}{\sigma^2} \qquad (2)$$

tem distribuição de probabilidade conhecida com distribuição qui-quadrado com n-1 graus de liberdade.

A função de densidade de um v.a qui-quadrado com k graus de liberdade é dado por:.

$$f(w) = \frac{k}{\Gamma\left(\frac{k}{2}\right)(2)^{k/2}} w^{\frac{k}{2}-1} e^{-\frac{w}{2}}; \quad w > 0$$

Notação: $W \sim \chi^2_{(k)}$

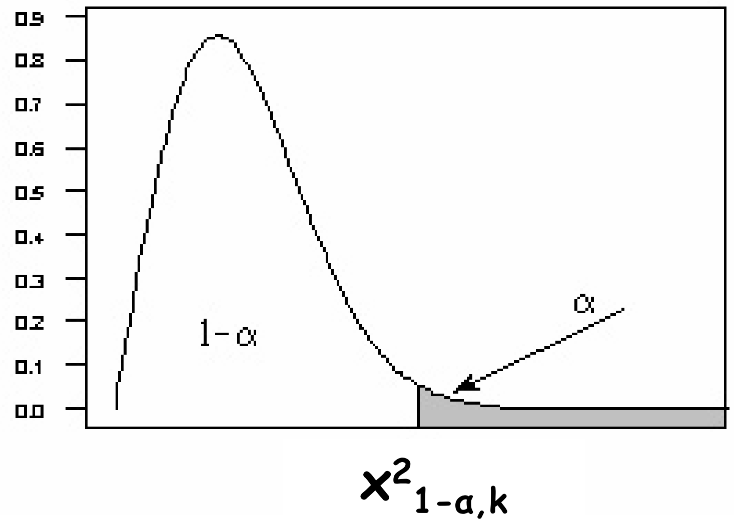


Se W tem distribuição qui-quadrado com k graus de liberdade então:

- **E(W)=k, Var(W)=2k;**
- **A distribuição é assimétrica á direita;**
- **A medida que os graus de liberdade aumenta a distribuição torna-se simétrica.**

Uso Da Tabela Distribuição Qui-Quadrado

$$\text{Se W} \sim \chi^2_{(k)} \Rightarrow P(\text{W} \leq \chi^2_{1-\alpha,\,k}) = 1 - \alpha$$



Exemplo 4: Suponha que W é uma v.a com 10 graus de liberdade determinar:

(a) P(W>2,56);

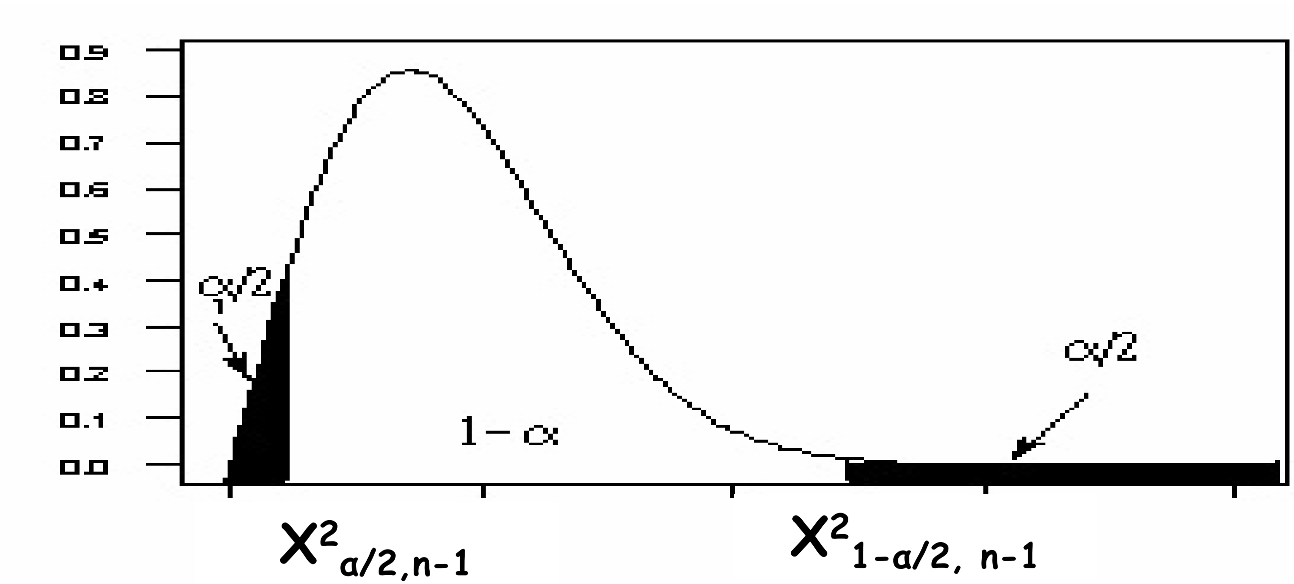
(b) P(2,56<W<4,87)

(c) O valor de k tal que, P(W<k)=0,95.

Da Estatística dada em (2) temos:

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Para uma nível de confiança 100(1-α)% fixado pode-se determinar $\chi^2_{\frac{\alpha}{2},n-1}$ *e* $\chi^2_{1-\frac{\alpha}{2},n-1}$ *da distribuição qui-quadrado como mostra a figura::*



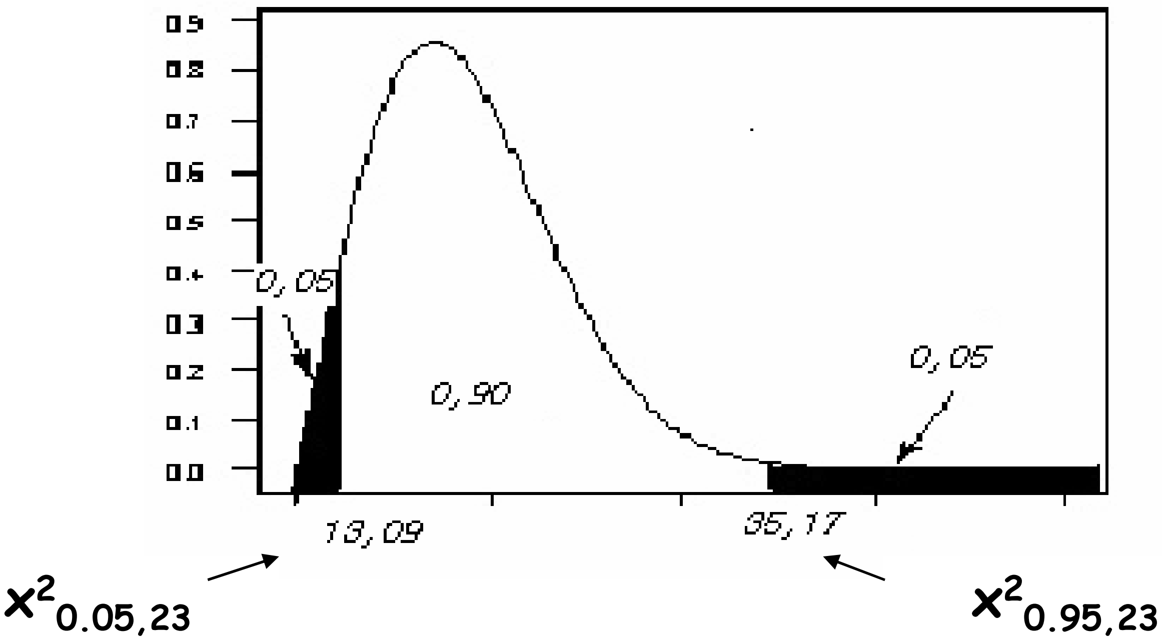
$$P\left(\chi^2_{\alpha/2,n-1} \le W \le \chi^2_{1-\alpha/2,\,n-1}\right) = P\left(\chi^2_{\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\alpha/2,\,n-1}\right) = 1-\alpha$$

Um intervalo de 100(1-α)% de confiança para σ^2 é ado por

$$IC(\sigma^2, 1-\alpha) = \left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right)$$

Exemplo: pretende-se avaliar a variabilidade associada ao resultado de um determinado método de análise química. Com esse objetivo, efetuaram-se 24 análises a uma determinada substância em que se segui o referido método, em condições perfeitamente estabilizadas. A variância amostral dos resultados (expressados numa determinada unidade) foi de 4,58. Admitindo que o resultado das análises segue uma distribuição normal. Obtenha um intervalo de 90% de confiança para variância.

Para 1-α=0,90 → α=0,10, da distribuição qui-quadrado com n-1=24-1=23 graus de liberdade temos:



$$IC(\sigma^2, 0{,}9) = \left(\frac{(24-1)(4{,}58)}{35{,}17}, \frac{(24-1)(4{,}58)}{13{,}09} \right) = (2{,}995; 8{,}047)$$

Intervalo de confiança para uma proporção populacional

Suponha que tem-se uma população dicotômica, constituída apenas por elementos de dois tipos , isto é, cada elemento pode ser classificado com *sucesso* ou *fracasso*, suponha que probabilidade de sucesso é ***p*** e de fracasso é **q=1-p,** e desta população se retira uma amostra aleatória, $X_1 \ldots, X_n$ de n observações. Vimos

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Para um nível confiança fixando em 100(1-α)%,um intervalo para **p**, para uma amostra suficientemente grande.

$$IC(p, 1-\alpha) = \left(\hat{p} - z_{1-\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} ; \hat{p} + z_{1-\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} \right)$$

Abordagem otimista

substituir $p(1\text{-}p)$ por $\hat{p}(1\text{-}\hat{p})$

$$IC(p,1-\alpha)=\left(\hat{p}-z_{1-\alpha/2}\times\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\hat{p}+z_{1-\alpha/2}\times\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \qquad (1a)$$

Abordagem conservativa

substituir $p(1\text{-}p)$ por 1/4

$$IC(p,1-\alpha)=\left(\hat{p}-z_{1-\alpha/2}\times\frac{1}{\sqrt{4n}};\hat{p}+z_{1-\alpha/2}\times\frac{1}{\sqrt{4n}}\right) \qquad (1b)$$

Exemplo: Um estudo foi feito para determinar a proporção de famílias em uma comunidade que tem telefone (p). Uma amostra de 200 famílias é selecionada, ao acaso, e 160 afirmam ter telefone. Que dizer de p com 95% de confiança?

Uma estimativa pontual de p é $\hat{p} = \frac{160}{200} = 0{,}8\ (80\%)$

Já que 1-α=0,95, temos da tabela normal padrão $z_{0975.}$=1,96. Substituindo em (1a)

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

Em (1b)

$$IC(p, 0{,}95) = \left(0{,}8 - 1{,}96 \times \frac{1}{\sqrt{4 \times 200}} ; 0{,}8 - 1{,}96 \times \frac{1}{\sqrt{4 \times 200}} \right) = (0{,}731; 0{,}869)$$

O erro máximo de estimação na estimação de p é dado por

$$E = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{p(1-p)}{n}} \qquad \Rightarrow n = \frac{z^2_{1-\alpha/2}\left(p(1-p)\right)}{E^2}$$

Quando não se tem informação de p:

$$\Rightarrow n = \frac{z^2_{1-\alpha/2}(0{,}25)}{E^2}$$

No caso de população finita de N elementos é introduzida o fator de correção de população finita

$$E = z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} \qquad \Rightarrow n = \frac{Nz^2_{1-\alpha/2}\,p(1-p)}{E^2(N-1)+z^2_{1-\alpha/2}\,p(1-p)}$$

Quando não se tem informação de p:

$$\Rightarrow n = \frac{Nz^2_{1-\alpha/2}(0{,}25)}{E^2(N-1)+z^2_{1-\alpha/2}(0{,}25)}$$

Exemplo: O serviço social de um município deseja determinar a proporção de famílias com uma renda familiar inferior a R$ 200,00. Estudos anteriores indicam que esta proporção é de 20%.

(a) Que tamanho de amostra se requer para assegurar uma confiança de 95% que o erro máximo de estimação desta proporção não ultrapasse o 0,05?

(b) Em quanto variara o tamanho da amostra se o erro máximo permissível é reduzido a 0,01.?

Dos dados temos p=0,20 e 1-α=0,95. Da tabela normal padrão $z_{0,975}$.=1,96.

(a) O erro máximo de estimação E=0,05.

$$\Rightarrow n = \frac{(1{,}96)^2 (0{,}2 \times 0{,}8)}{0{,}05^2} = 245{,}86 \approx 246$$

(b) O erro máximo de estimação E=0,01.

$$\Rightarrow n = \frac{(1{,}96)^2\left(0{,}2 \times 0{,}8\right)}{0{,}01^2} = 6146{,}56 \approx 6147$$

No caso de estarmos usando nível de confiança de 95% , temos que $z_{0,975}$.=1,96 ≅2, então temos:

$$n_0 = \frac{1}{E^2}$$

A expressão anterior é muito usado no planejamento de pesquisa de levantamento, com o objetivo de estimar várias proporções como nos exemplos seguintes:

• ***Numa pesquisa eleitoral,*** em que é comum a necessidade de avaliar a proporção de cada candidato;

• **Na pesquisa de mercado**, em que normalmente desejam-se avaliar as proporções de várias características dos consumidores.

No caso de população finita de N elementos é introduzida o fator de correção de população finita:

$$n = \frac{Nn_0}{N + n_0 - 1}$$