AULA 10
Estimativas e Tamanhos Amostrais

Ernesto F. L. Amaral

18 de setembro de 2012
Metodologia de Pesquisa (DCP 854B)

Fonte:
Triola, Mario F. 2008. "Introdução à estatística". 10 ª ed. Rio de Janeiro: LTC. Capítulo 7 (pp.250-303).

ESQUEMA DA AULA

– Estimação da proporção populacional.

– Estimação da média populacional: σ conhecido.

– Estimação da média populacional: σ desconhecido.

– Estimação da variância populacional.

OBJETIVO DO CAPÍTULO

– Neste capítulo, são usados dados amostrais para obter estimativas de parâmetros populacionais, o que é a essência da inferência estatística.

– As duas principais aplicações da inferência estatística envolvem o uso de dados amostrais para:

  – Estimar o valor de um parâmetro populacional (proporções, médias, variâncias).

  – Testar alguma afirmação (ou hipótese) sobre uma população.

– São ainda apresentados métodos para determinação dos tamanhos amostrais necessários para estimar esses parâmetros.

ESTIMAÇÃO DA PROPORÇÃO POPULACIONAL

ESTIMAÇÃO DA PROPORÇÃO POPULACIONAL

– A intenção é de usar uma proporção amostral para estimar o valor de uma proporção populacional com um intervalo de confiança.

– São apresentados métodos para encontrar o tamanho da amostra necessário para estimar a proporção populacional.

– É importante:

  – Entender o que são, o que fazem e por que são necessários os intervalos de confiança.

  – Desenvolver a habilidade de construir estimativas de intervalos de confiança de proporções populacionais.

  – Aprender como interpretar corretamente um intervalo de confiança.

REQUISITOS

– Serão considerados casos em que distribuição normal pode ser usada para aproximar distribuição amostral de proporções amostrais.

– Requisitos para métodos de estimação de proporções:

  – É utilizada amostra aleatória simples.

  – Condições para distribuição binomial são satisfeitas: (1) número fixo de tentativas; (2) tentativas independentes; (3) duas categorias de resultados; e (4) probabilidades permanecem constantes para cada tentativa.

  – Há pelo menos 5 sucessos e pelo menos 5 fracassos. Essa exigência é uma forma de garantir que $np \geq 5$ e $nq \geq 5$, permitindo usar distribuição normal como aproximação para a distribuição binomial.

NOTAÇÃO PARA PROPORÇÕES

– p = proporção populacional.

– $\hat{p} = x/n$ = proporção amostral de x sucessos em uma amostra de tamanho n.

– $\hat{q} = 1 - \hat{p}$ = proporção amostral de fracassos em uma amostra de tamanho n.

– Esta seção se concentra na proporção populacional p, que é o mesmo que trabalhar com probabilidades e porcentagens.

– Expresse porcentagens em forma decimal.

ESTIMATIVA PONTUAL

– Se desejamos estimar proporção populacional com único valor, a melhor estimativa é $\hat{p}$ (estimativa pontual).

– Estimativa pontual é um único valor usado para aproximar um parâmetro populacional.

– Proporção amostral $\hat{p}$ é a melhor estimativa pontual da proporção populacional p.

– A estimativa pontual é usada porque é não-viesado e é o mais consistente dos estimadores que poderiam ser usados:

  – Distribuição das proporções amostrais tende a centralizar em torno do valor de p.

  – Proporções amostrais não subestimam/superestimam p.

  – Desvio padrão das proporções amostrais tende a ser menor do que desvios padrões de outros estimadores.

POR QUE USAR INTERVALOS DE CONFIANÇA?

– Como a estimativa pontual não diz o quão precisa ela é, os estatísticos desenvolveram o intervalo de confiança (estimativa intervalar).

– **Intervalo de confiança (IC)** é uma faixa (ou intervalo) de valores usada para estimar o verdadeiro valor de um parâmetro populacional.

– A um intervalo de confiança é associado um nível de confiança, por exemplo, 0,95 (ou 95%).

– O **nível de confiança (NC)** apresenta a taxa de sucesso do procedimento usado para construir o intervalo de confiança.

– Nível de confiança é expresso como probabilidade ou área ($1–\alpha$), em que α é o complemento do nível de confiança.

– Quanto maior o NC, maior o IC.

NÍVEL DE CONFIANÇA

– **Nível de confiança** (grau de confiança ou coeficiente de confiança) é a proporção de vezes que o intervalo de confiança realmente contém o parâmetro populacional, supondo que processo seja repetido várias vezes.

– As escolhas mais comuns para nível de confiança são 90% (α = 0,10), 95% (α = 0,05) e 99% (α = 0,01).

– Escolha de 95% é mais comum porque resulta em bom equilíbrio entre **precisão** (largura do intervalo de confiança) e **confiabilidade** (nível de confiança).

– Precisão (exatidão) é a qualidade de que o resultado da amostra reflita o mundo real.

– Confiabilidade é a qualidade de uma determinada técnica produzir os mesmos resultados em várias aplicações.

INTERPRETAÇÃO DE INTERVALO DE CONFIANÇA

– Por exemplo: $n = 280$; $0,381 < p < 0,497$.

– **Correto:** estamos 95% confiantes de que o intervalo de 0,381 a 0,497 realmente contém o verdadeiro valor de *p*.

  – Se selecionássemos muitas diferentes amostras de tamanho 280 e construíssemos os intervalos de confiança correspondentes, 95% deles realmente conteriam o valor da proporção populacional *p*.

  – O nível de 95% se refere à taxa de sucesso do processo em uso para se estimar a proporção populacional, e não se refere à própria proporção populacional.

– **Errado:** como o valor de *p* é fixo, é incorreto dizer que há uma chance de 95% de que o verdadeiro valor de *p* esteja entre 0,381 e 0,497.

INTERPRETAÇÃO DE INTERVALO DE CONFIANÇA

– Em qualquer ponto no tempo, há um valor de *p* fixo e constante, e um intervalo de confiança construído a partir de uma amostra que inclui ou não inclui *p*.

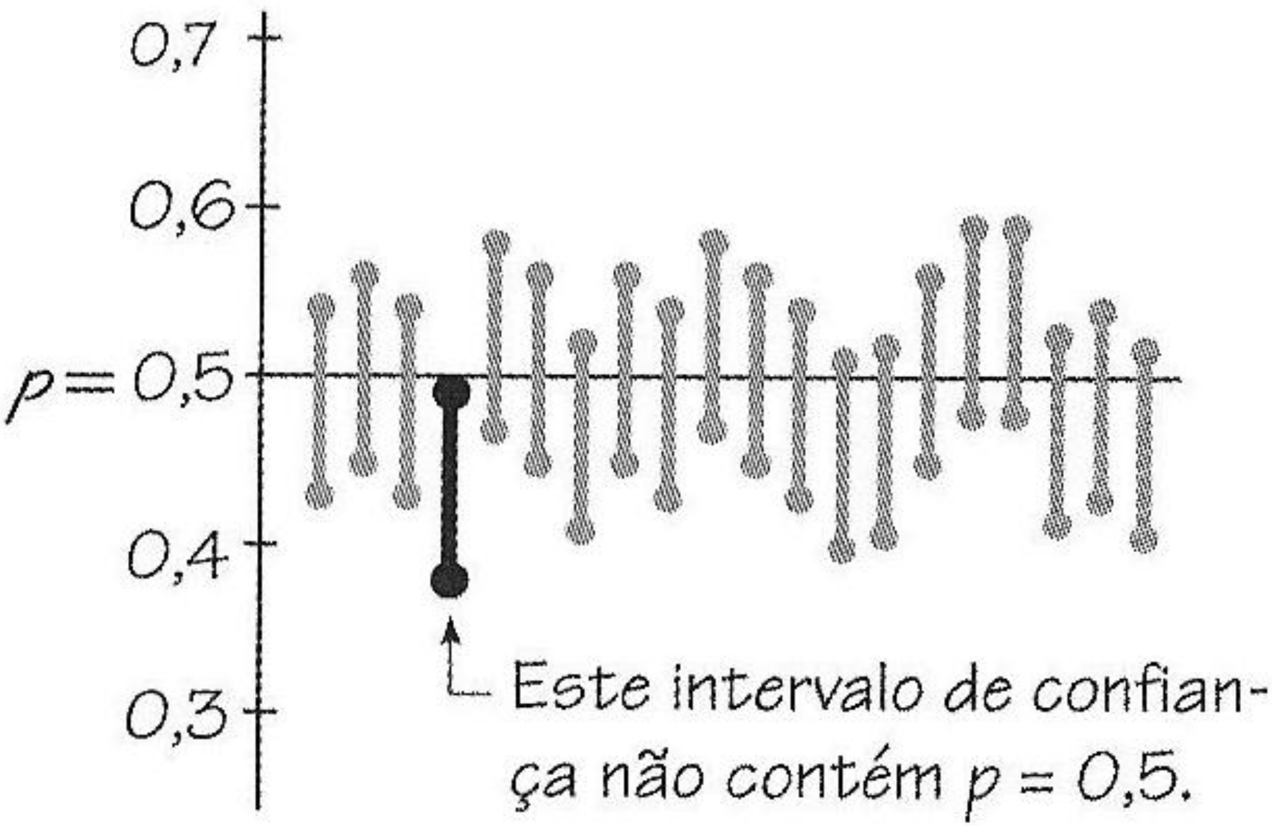
– O valor de *p* é fixo, de modo que os limites do intervalo de confiança ou contêm ou não contêm *p*, e é por isso que é errado dizer que há uma chance de 95% de que *p* esteja entre valores como 0,381 e 0,497.

– Um nível de confiança de 95% diz que o processo resultará, a longo prazo, em limites de intervalo de confiança que contenham a verdadeira proporção populacional 95% das vezes.

EXEMPLO DE INTERVALOS DE CONFIANÇA

– Intervalos de confiança a partir de 20 amostras diferentes.

– Com 95% de confiança, esperamos que 19 das 20 amostras resultem em intervalos de confiança que realmente contenham o verdadeiro valor de *p*.



VALORES CRÍTICOS

– O escore padrão z ou valor crítico ($z_{\alpha/2}$) separa proporções amostrais que têm chance de ocorrer das que não têm.

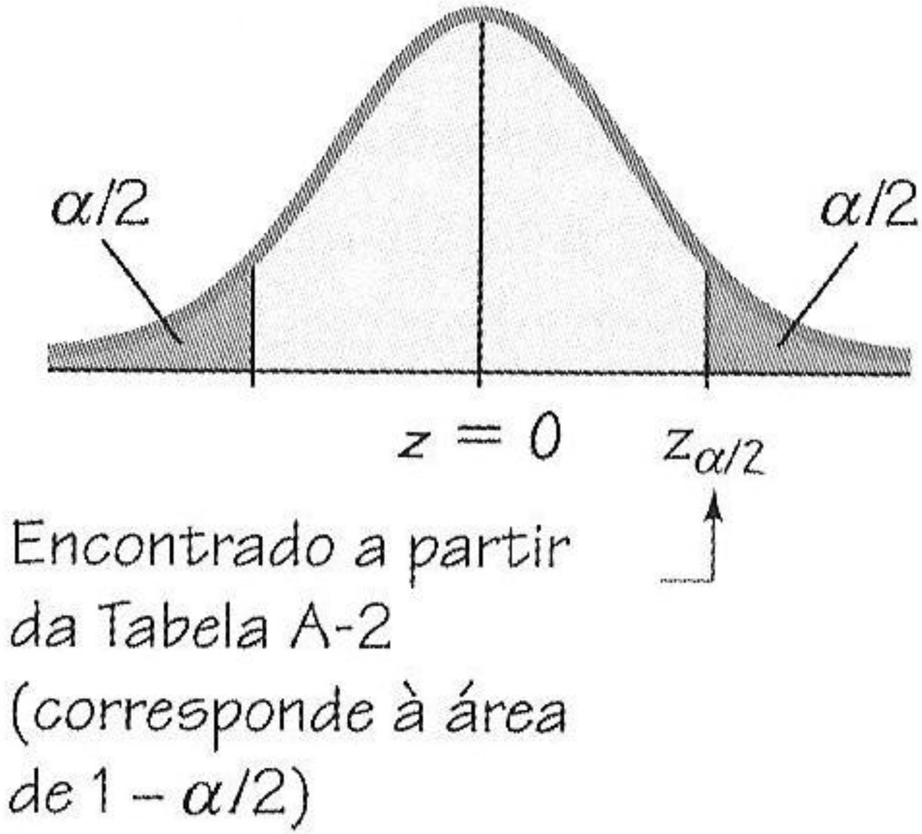
– Os valores críticos se baseiam nestas observações:

- – A distribuição amostral das proporções amostrais pode ser aproximada por uma distribuição normal.
- – Proporções amostrais têm uma chance relativamente pequena de cair em uma das caudas da curva normal.
- – Representando cada cauda por $\alpha/2$, há uma probabilidade total α de que uma proporção amostral caia em uma das duas caudas.
- – Há uma probabilidade de $1-\alpha$ de que uma proporção amostral caia na região entre os pontos críticos (+ e –).

VALORES CRÍTICOS NA DISTRIBUIÇÃO NORMAL PADRÃO

– Valor crítico é um número que separa estatísticas amostrais que têm chance de ocorrer daquelas que não têm.

– O número $z_{\alpha/2}$ é um valor crítico que separa uma área $\alpha/2$ na cauda direita da distribuição normal padronizada.

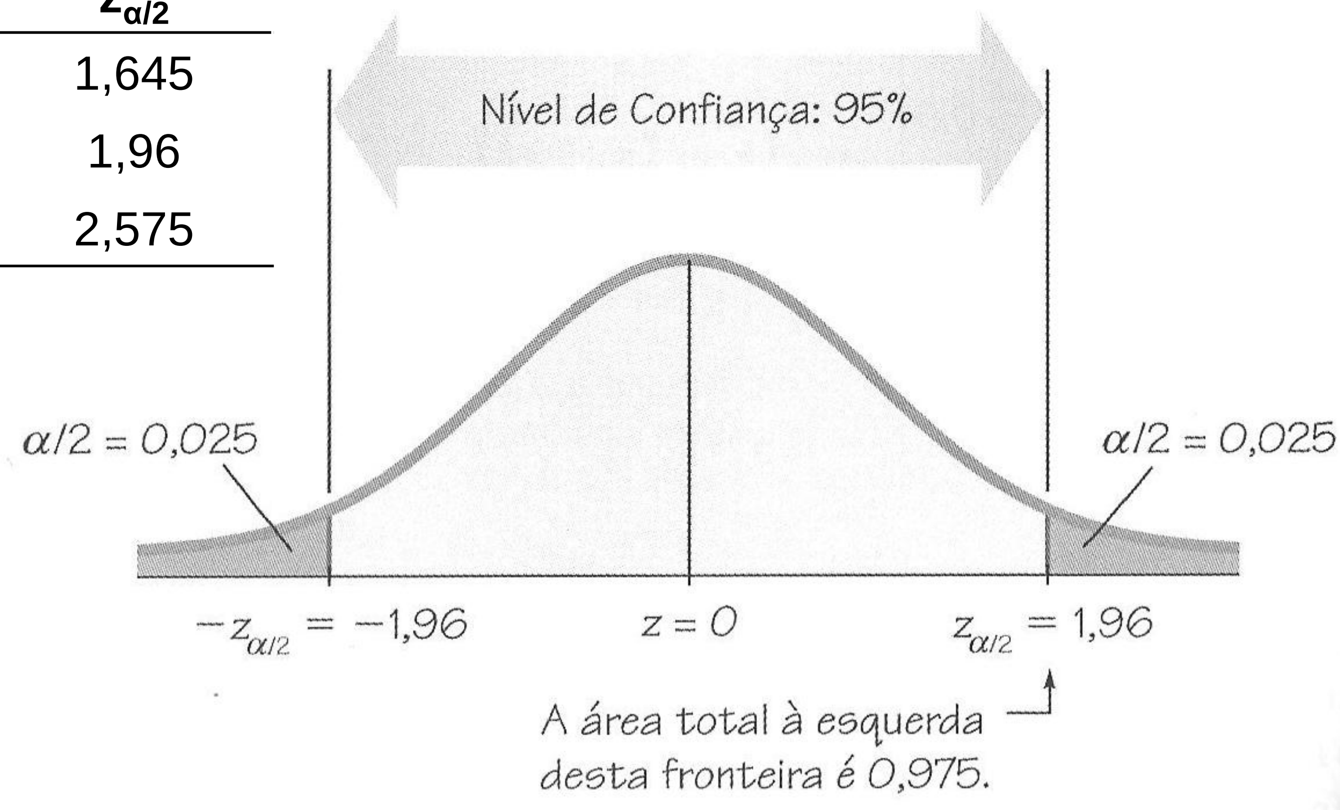


MAIS SOBRE VALORES CRÍTICOS

– O valor de $z_{\alpha/2}$ está na fronteira da cauda direita e o valor de $-z_{\alpha/2}$ está na fronteira da cauda da esquerda.

– Encontrando $z_{\alpha/2}$ para um nível de confiança específico...

Nível de confiança	α	Valor crítico $z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575



MARGEM DE ERRO

– Quando coletamos um conjunto de dados amostrais, podemos calcular a proporção amostral, a qual é tipicamente diferente da proporção populacional.

– A **margem de erro** (***E***) é a diferença máxima provável entre a proporção amostral observada e o verdadeiro valor da proporção populacional:

- – Isso ocorre quando dados de amostra aleatória simples são usados para estimar uma proporção populacional.
- – É também chamada de erro máximo da estimativa.
- – É encontrada pela multiplicação do valor crítico pelo desvio padrão das proporções amostrais.

MARGEM DE ERRO E INTERVALO DE CONFIANÇA

– Margem de erro para proporções é calculada por:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

– Há uma probabilidade *α* de que a proporção amostral tenha erro maior do que *E*.

– Ou seja, $\hat{p}$ terá probabilidade de 1 – *α* de estar a: $z_{\alpha/2}\sqrt{pq/n}$ de *p*.

– Intervalo de confiança para proporção populacional é representado por: $\hat{p} - E < p < \hat{p} + E$

$$\hat{p} \pm E$$

$$(\hat{p} - E; \hat{p} + E)$$

CONSTRUÇÃO DE INTERVALO DE CONFIANÇA

– Verifique se requisitos são satisfeitos: (1) amostra aleatória simples; (2) condições para distribuição binomial (tentativas fixas, independentes, duas categorias, probabilidade constante); e (3) há pelo menos 5 sucessos e 5 fracassos.

– Ache o valor crítico que corresponde ao nível de confiança desejado. Se nível de confiança é 95%, $z_{\alpha/2} = 1{,}96$.

– Calcule a margem de erro:

$$E = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$$

– Use o valor da margem de erro e o valor da proporção amostral para encontrar o intervalo de confiança:

$$\hat{p} - E < p < \hat{p} + E$$

– Arredonde os limites do intervalo de confiança.

EXEMPLO DE CÁLCULO

– Por exemplo, em 280 tentativas, houve 123 acertos:

- – n = 280
- – $\hat{p}$ = 123/280 = 0,439286
- – $\hat{q}$ = 1 – 0,439286 = 0,560714

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1{,}96\sqrt{\frac{(0{,}439286)(0{,}560714)}{280}} = 0{,}058133$$

$$\hat{p} - E < p < \hat{p} + E$$

$$0{,}439286 - 0{,}058133 < p < 0{,}439286 + 0{,}058133$$

$$0{,}381 < p < 0{,}497$$

– A taxa de sucesso é de 44%, com margem de erro de mais ou menos 6% e nível de confiança de 95% (geralmente resultados eleitorais omitem o nível de confiança).

FUNDAMENTOS PARA MARGEM DE ERRO

– Distribuição amostral das proporções é aproximadamente normal ($np \geq 5$ e $nq \geq 5$).

– Parâmetros da média e desvio padrão são relativos a *n* tentativas e são convertidos para a base por 1 tentativa pela divisão por *n*.

– Média das proporções amostrais:

$$\mu = \frac{np}{n} = p$$

– Desvio padrão das proporções amostrais:

$$\sigma = \frac{\sqrt{npq}}{n} = \sqrt{\frac{npq}{n^2}} = \frac{\sqrt{pq}}{n} = \frac{\sqrt{\hat{p}\hat{q}}}{n}$$

COMO DEFINIR O TAMANHO AMOSTRAL?

– Utilizando a fórmula da margem de erro, chegamos a:

$$n = \frac{\left[z_{\alpha/2}\right]^2 \hat{p}\hat{q}}{E^2}$$

– Se não conhecemos qualquer estimativa $\hat{p}$:

$$n = \frac{\left[z_{\alpha/2}\right]^2 0{,}5 * 0{,}5}{E^2} = \frac{\left[z_{\alpha/2}\right]^2 0{,}25}{E^2}$$

– Se o tamanho amostral calculado não for um número inteiro, arredonde-o para o inteiro maior mais próximo.

– Quando a amostragem é sem reposição, a partir de uma população finita relativamente pequena, utilize:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\sqrt{\frac{N-n}{N-1}} \qquad n = \frac{\left[z_{\alpha/2}\right]^2 N\hat{p}\hat{q}}{\hat{p}\hat{q}\left[z_{\alpha/2}\right]^2 + (N-1)E^2}$$

TAMANHO DA POPULAÇÃO

– Para o cálculo do tamanho da amostra, o tamanho da população é usado somente em casos em que fazemos amostragem sem reposição a partir de uma população relativamente pequena.

– **Outras observações:**

- – Se margem de erro desejada igual a 5%, E=0,05.
- – Se nível de confiança desejada é de 95%, $z_{\alpha/2}$=1,96.
- – Assim:

$$n = \frac{\left[z_{\alpha/2}\right]^2 0{,}25}{E^2} = \frac{(1{,}96)^2 * 0{,}25}{(0{,}05)^2} = \frac{0{,}9604}{0{,}025} = 384{,}16 \approx 385$$

DETERMINAÇÃO DE ESTIMATIVA PONTUAL E DE “E”

– Se conhecemos os limites do intervalo de confiança, a proporção amostral e a margem de erro podem ser encontradas desta forma:

– Estimativa pontual de *p*:

$$\hat{p} = \frac{(limite\ de\ confiança\ superior) + (limite\ de\ confiança\ inferior)}{2}$$

– Margem de erro:

$$E = \frac{(limite\ de\ confiança\ superior) - (limite\ de\ confiança\ inferior)}{2}$$

INTERVALO DE CONFIANÇA AJUSTADO DE WALD

– O intervalo de confiança ajustado de Wald tem um melhor desempenho por ter maior probabilidade de conter a verdadeira proporção populacional.

– Acrescente 2 ao número de sucessos x, acrescente 2 ao número de fracassos e, então, calcule o intervalo de confiança.

– Se x=10 e n=20:

  – Intervalo usual: $0{,}281 < p < 0{,}719$

  – Intervalo ajustado de Wald com x=12 e n=24:

  $0{,}300 < p < 0{,}700$

– A chance de que o intervalo $0{,}300<p<0{,}700$ contenha p é mais próxima de 95% do que a chance de $0{,}281<p<0{,}719$.

INTERVALO DE CONFIANÇA DO ESCORE DE WILSON

– Limite inferior do intervalo de confiança:

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}}{n}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

– O limite superior do intervalo de confiança se expressa pela mudança do sinal negativo pelo sinal positivo:

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}}{n}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

– Usando x=10 e n=20, o intervalo de confiança do escore de Wilson é $0{,}290<p<0{,}701$.

ESTIMAÇÃO DA MÉDIA POPULACIONAL: σ CONHECIDO

ESTIMAÇÃO DA MÉDIA POPULACIONAL: σ CONHECIDO

– Aqui são apresentados métodos para usar dados amostrais para se encontrar estimativa pontual e intervalo de confiança para uma **média populacional**.

– Requisitos:

  – Amostra aleatória simples (todas amostras de mesmo tamanho têm igual chance de serem selecionadas).

  – Valor do desvio padrão populacional (σ) é conhecido.

  – Uma ou ambas as condições seguintes são satisfeitas: população é normalmente distribuída ou $n>30$.

– Se $n \leq 30$, a população não precisa ter uma distribuição exatamente normal, mas deve ser próxima da normal.

– Os métodos dessa seção são robustos, não sendo fortemente afetados por afastamentos da normalidade.

SUPOSIÇÃO DE TAMANHO AMOSTRAL REQUERIDO

– Distribuição normal é utilizada como distribuição das médias amostrais.

– Se população original não é normalmente distribuída, as médias de amostras com n>30 têm uma distribuição próxima da normal.

– Não é possível identificar tamanho amostral mínimo que seja suficiente para todos casos.

– Tamanho amostral mínimo depende de como distribuição populacional se afasta de uma normal.

– É utilizado o critério simplificado de n>30 como justificativa para tratar distribuição das médias amostrais como distribuição normal.

MELHOR ESTIMATIVA DA MÉDIA POPULACIONAL

– A média amostral $\bar{x}$ é a melhor estimativa pontual da média populacional μ.

– Para todas populações, a média amostral é um estimador não-viesado da média populacional.

- – A distribuição das médias amostrais tende a se centralizar em torno do valor da média populacional.
- – Médias amostrais não tendem a superestimar ou subestimar o valor populacional.

– Para muitas populações, a distribuição das médias amostrais tende a ser mais consistente (menos variação) do que as distribuições de outras estatísticas amostrais.

INTERVALO E NÍVEL DE CONFIANÇA, MARGEM DE ERRO

– O **intervalo de confiança** permite compreender melhor a precisão da estimativa da média amostral.

– Este intervalo está associado a um **nível de confiança**, o qual indica a taxa de sucesso do procedimento usado para construção do intervalo (confiabilidade).

– Diferença entre a média amostral e a média populacional é um erro.

– **Margem de erro** para a média, baseada em *σ* conhecido:

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

– Com isso, calculamos os limites do intervalo de confiança:

$\bar{x} - E < \mu < \bar{x} + E$ ou $\bar{x} \pm E$ ou $(\bar{x} - E; \bar{x} + E)$

CONSTRUÇÃO DE INTERVALO DE CONFIANÇA

– Verifique se: (1) temos uma amostra aleatória simples; (2) σ é conhecido; e (3) população parece ser normal ou n>30.

– Encontre o valor crítico $z_{\alpha/2}$ que corresponde ao nível desejado de confiança (se nível de confiança=95%, z=1,96).

– Calcule margem de erro: $E = z_{\alpha/2} * \sigma/\sqrt{n}$

– Com valor da margem de erro e valor da média, ache valores dos limites do intervalo de confiança:

$$\bar{x} - E < \mu < \bar{x} + E$$

– Ao usar o conjunto original de dados, arredonde limites do intervalo para uma casa decimal a mais do que as originais.

– Ao usar estatísticas-resumo, arredonde limites para mesmo número de casas decimais usados na média amostral.

INTERPRETANDO UM INTERVALO DE CONFIANÇA

– Se temos $72{,}4<\mu<80{,}2$ com intervalo de confiança de 95%:

– **Correto:**

  – Estamos 95% confiantes de que o intervalo de 72,4 a 80,2 realmente contenha o verdadeiro valor de μ.

  – Se selecionamos muitas amostras diferentes de mesmo tamanho e construímos os intervalos de confiança correspondentes, 95% deles realmente conterão μ.

  – Essa é a taxa de sucesso do processo usado para estimar média populacional.

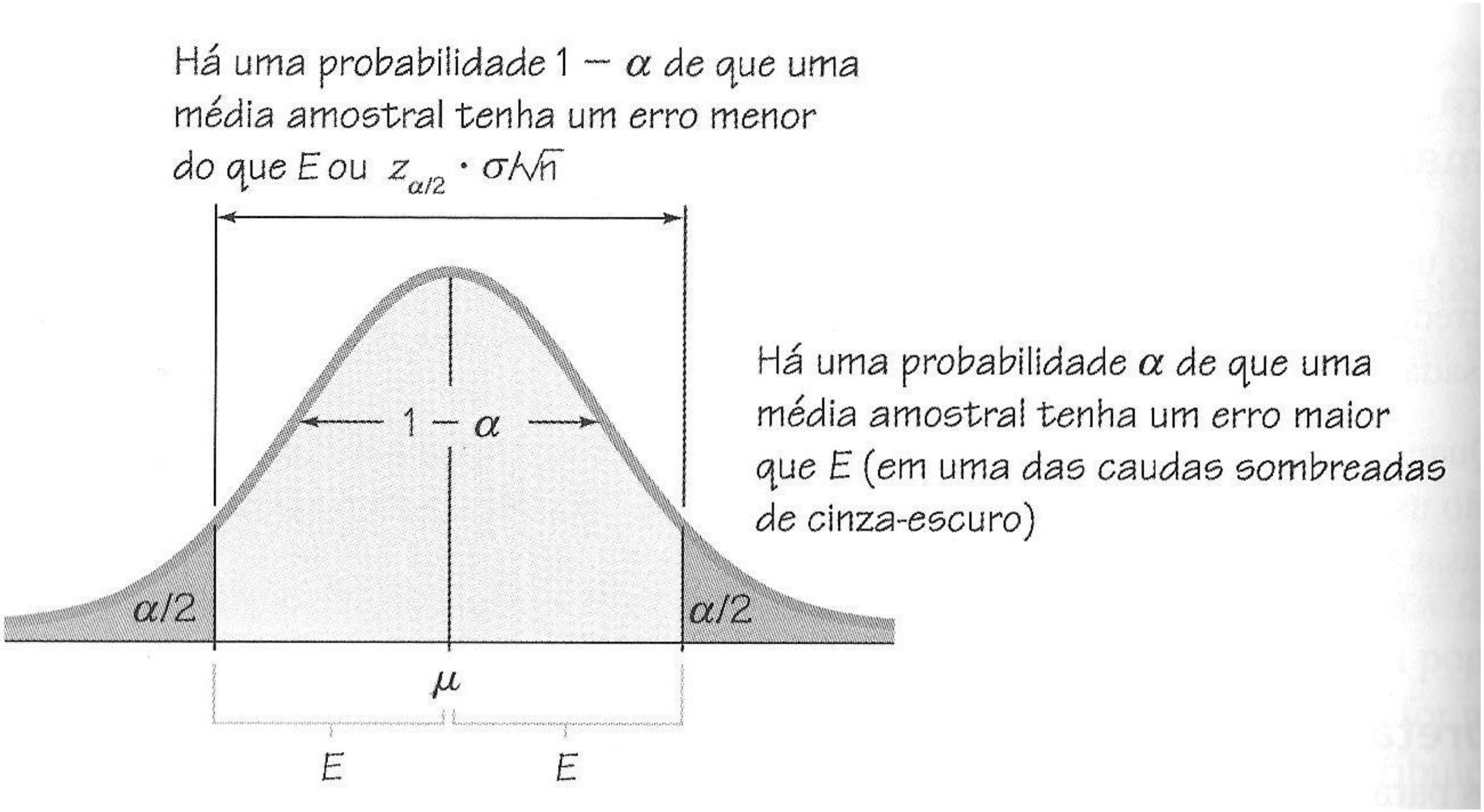
– **Errado:**

  – Como μ é constante fixa, é errado dizer que há uma chance de 95% de que μ esteja entre 72,4 e 80,2.

  – 95% das médias amostrais estão entre 72,4 e 80,2.

DISTRIBUIÇÃO DE MÉDIAS AMOSTRAIS

– Distribuição de médias amostrais com σ conhecido.



FUNDAMENTOS PARA INTERVALO DE CONFIANÇA

- Construção de intervalos de confiança está baseada no teorema central do limite, que diz que:
  - ao coletar amostras aleatórias simples de mesmo tamanho de uma população distribuída normalmente...
  - ... as médias amostrais são normalmente distribuídas com média μ e desvio padrão $\sigma/\sqrt{n}$.
- Formato do intervalo de confiança vem de equação do TCL:
  - Utilize: $z = (\bar{x} - \mu_{\bar{x}})/\sigma_{\bar{x}}$; $\sigma_{\bar{x}} = \sigma/\sqrt{n}$; $\mu_{\bar{x}} = \mu$.
  - Para obter: $\mu = \bar{x} \pm z\left(\sigma/\sqrt{n}\right)$.
- O uso de valores positivo e negativo de z resulta nos limites do intervalo de confiança com que estamos trabalhando.
- Com NC=95%, há probabilidade de 0,05 da média amostral estar a mais ou a menos de 1,96 DP da média populacional.

TAMANHO AMOSTRAL PARA ESTIMAR MÉDIA μ

– Determinação do tamanho de amostra aleatória simples é importante, porque amostras grandes gastam tempo e dinheiro, e amostra pequenas levam a resultados imprecisos.

– Fórmula do tamanho amostral não depende do tamanho da população (*N*):

$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

– $z_{\alpha/2}$ = escore *z* crítico com base no nível de confiança.

– *E* = margem de erro desejada.

– σ = desvio padrão populacional.

– Caso de amostra sem reposição de população finita:

$$n = \frac{N\sigma^2\left(z_{\alpha/2}\right)^2}{(N-1)E^2 + \sigma^2\left(z_{\alpha/2}\right)^2}$$

LIDANDO COM *σ* DESCONHECIDO

– Geralmente o desvio padrão populacional é desconhecido.

– Use a **regra empírica da amplitude** para estimar o desvio padrão ($\sigma \approx$ amplitude/4).

  – Esse valor é maior ou igual ao real σ pelo menos 95% das vezes.

– Realize **estudo piloto**: comece processo de coleta da amostra e com base nos primeiros valores, calcule o desvio padrão amostral (s) e use-o no lugar de σ.

  – Esse valor pode ser melhorado à medida que mais dados são obtidos.

– Estime valor de σ com resultados de **estudos anteriores**.

– Ao calcular n, **erros devem ser conservadores**, no sentido de aumentar tamanho amostral em vez de diminuir.

ESTIMAÇÃO DA MÉDIA POPULACIONAL: *σ* DESCONHECIDO

ESTIMAÇÃO DA MÉDIA POPULACIONAL: σ DESCONHECIDO

– São apresentados métodos para determinar intervalo de confiança de média populacional quando o desvio padrão da população não é conhecido.

– Requisitos:

  – Amostra aleatória simples (todas amostras de mesmo tamanho têm igual chance de serem selecionadas).

  – Amostra provém de população normalmente distribuída ou n>30.

– Uma população pode ser considerada normalmente distribuída se dados amostrais não tiverem valores extremos (*outliers*) e histograma for próximo de normal.

– O tamanho da amostra depende de quanto a distribuição se afasta de uma distribuição normal.

MELHOR ESTIMATIVA DA MÉDIA POPULACIONAL

– A média amostral $\bar{x}$ continua sendo a melhor estimativa pontual da média populacional μ.

– Se σ não é conhecido, mas requisitos são satisfeitos, usamos **distribuição *t* de Student** (em vez de distribuição normal).

– O valor de σ é estimado com o valor do desvio padrão amostral (*s*), mas isso introduz fonte de não-confiabilidade, principalmente quando amostras são pequenas.

– Isso é compensado fazendo o intervalo de confiança um pouco mais largo, com os valores críticos $t_{\alpha/2}$ que são maiores do que os valores críticos $z_{\alpha/2}$.

DISTRIBUIÇÃO *t* DE STUDENT

– Se uma população tem distribuição normal, então a distribuição *t* de Student para todas amostras de tamanho *n* é representada por:

$$t = \frac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$$

– Para encontrar o valor crítico de $t_{\alpha/2}$, precisamos saber o número apropriado de graus de liberdade.

– O número de **graus de liberdade** para um conjunto de dados amostrais é o número de valores amostrais que podem variar depois que certas restrições (como a média) tiverem sido impostas aos dados amostrais:

graus de liberdade = *n* – 1

EXEMPLO

– Uma amostra de tamanho *n*=23 é uma amostra aleatória simples selecionada de uma população normalmente distribuída. Ache o valor crítico $t_{\alpha/2}$ correspondente a um nível de confiança de 95%.

– No Stata:

  – *di invttail(n,α/2) = di invttail*(23,0.025) = 2.0686576 = $t_{\alpha/2}$

– Se tivéssemos *n*=23 e $t_{\alpha/2}$=2.07, podemos calcular *α/2*:

  – *di ttail(n,*$t_{\alpha/2}$*) = di ttail*(23,2.07) = 0.02493172 = *α*/2

MARGEM DE ERRO E INTERVALO DE CONFIANÇA

– Para calcular margem de erro *E* para estimativa de μ com σ desconhecido, onde $t_{\alpha/2}$ tem n–1 graus de liberdade:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

– Intervalo de confiança para estimativa de μ com σ desconhecido:

$$\bar{x} - E < \mu < \bar{x} + E$$

CONSTRUÇÃO DE INTERVALO DE CONFIANÇA

– Verifique se os requisitos são satisfeitos: (1) amostra aleatória simples; e (2) população próxima de distribuição normal ou $n>30$.

– Usando $n-1$ graus de liberdade, ache valor crítico $t_{\alpha/2}$, correspondente ao nível de confiança.

– Calcule margem de erro: $E = t_{\alpha/2}\, s/\sqrt{n}$

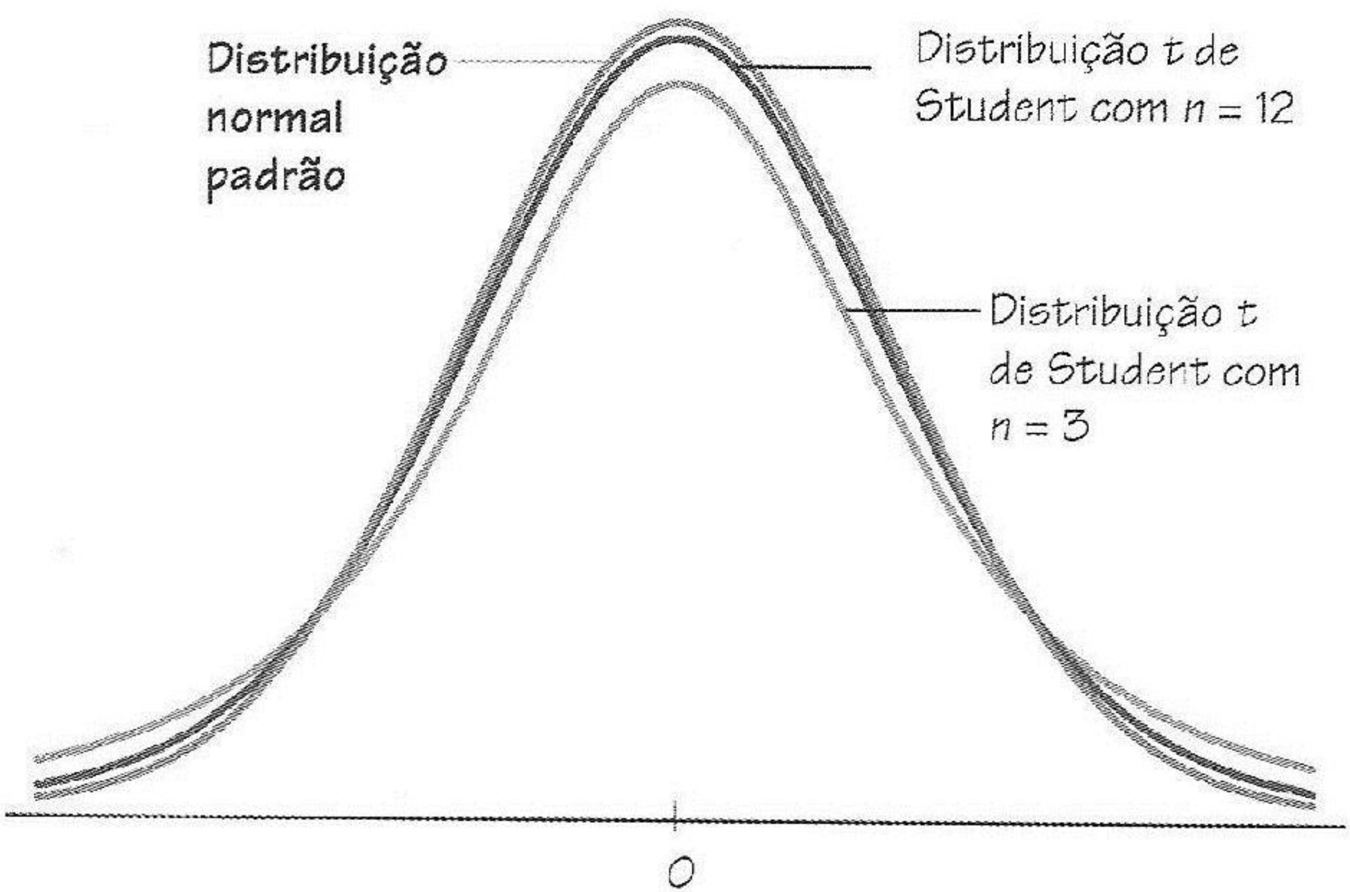
– Use valor da margem de erro e valor da média amostral e ache os valores dos limites do intervalo de confiança:

$$\bar{x} - E < \mu < \bar{x} + E$$

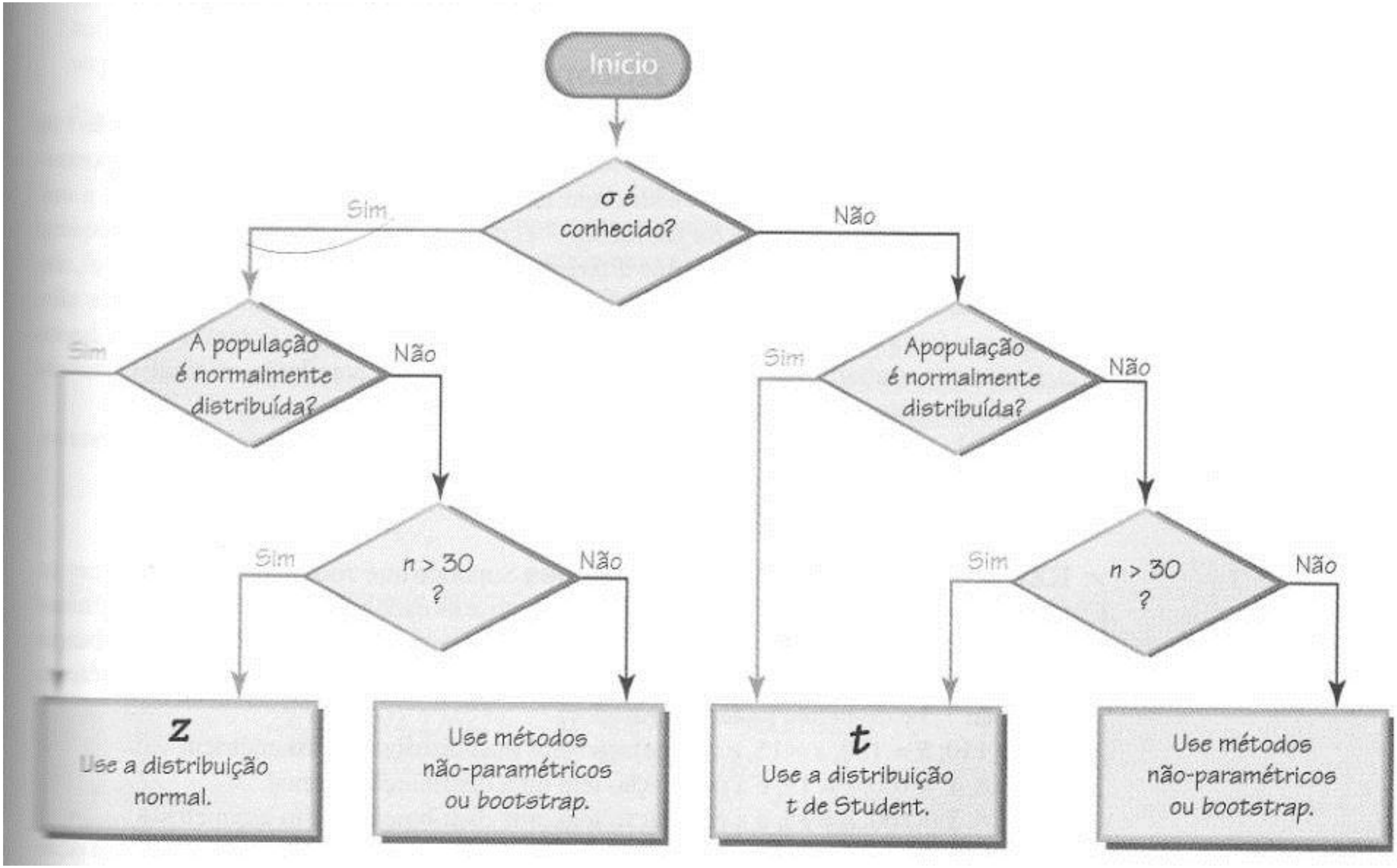
– Arredonde os limites do intervalo de confiança resultante.

DISTRIBUIÇÃO *t* DE STUDENT PARA *n*=3 E *n*=12

– Distribuição *t* de Student tem a mesma forma geral da distribuição normal padrão, mas reflete a maior variabilidade que se espera com amostras pequenas.



ESCOLHA DA DISTRIBUIÇÃO APROPRIADA



– Métodos não-paramétricos e *bootstrap* não fazem suposições sobre população original.

DETERMINAÇÃO DE ESTIMATIVA PONTUAL E DE "E"

– Se conhecemos os limites do intervalo de confiança, a média amostral e a margem de erro podem ser encontradas desta forma:

– Estimativa pontual de *μ*:

$$\bar{x} = \frac{(limite\ de\ confiança\ superior) + (limite\ de\ confiança\ inferior)}{2}$$

– Margem de erro:

$$E = \frac{(limite\ de\ confiança\ superior) - (limite\ de\ confiança\ inferior)}{2}$$

USO DE INTERVALOS DE CONFIANÇA

– Intervalo de confiança pode ser usado para:

  – Estimar o valor de um parâmetro populacional.

  – Descrever, explorar ou comparar conjuntos de dados.

```
. proportion x001

Proportion estimation                    Number of obs    =    79946
```

	Proportion	Std. Err.	[95% Conf.	Interval]
x001				
male	.4969604	.0017683	.4934945	.5004264
female	.5030396	.0017683	.4995736	.5065055

– Porém, intervalos de confiança não devem ser usados para se tirarem conclusões finais sobre igualdade de médias.

ESTIMAÇÃO DA VARIÂNCIA POPULACIONAL

ESTIMAÇÃO DA VARIÂNCIA POPULACIONAL

– São apresentados métodos para:
  - – Encontrar intervalo de confiança para um desvio padrão ou variância populacional.
  - – Determinar tamanho amostral necessário para estimativa do desvio padrão (σ) ou variância populacional (σ^2).

– Requisitos:
  - – Amostra aleatória simples.
  - – População deve ter valores normalmente distribuídos, mesmo que amostra seja grande.

– Afastamento da distribuição normal pode levar a erros grosseiros.

– Distribuição qui-quadrado é usada para encontrar intervalo de confiança para σ ou σ^2.

DISTRIBUIÇÃO QUI-QUADRADO

– Suponha que população:

– Seja normalmente distribuída.

– Tenha variância populacional (σ^2).

– Desta população:

– São selecionadas amostras aleatórias independentes de tamanho n.

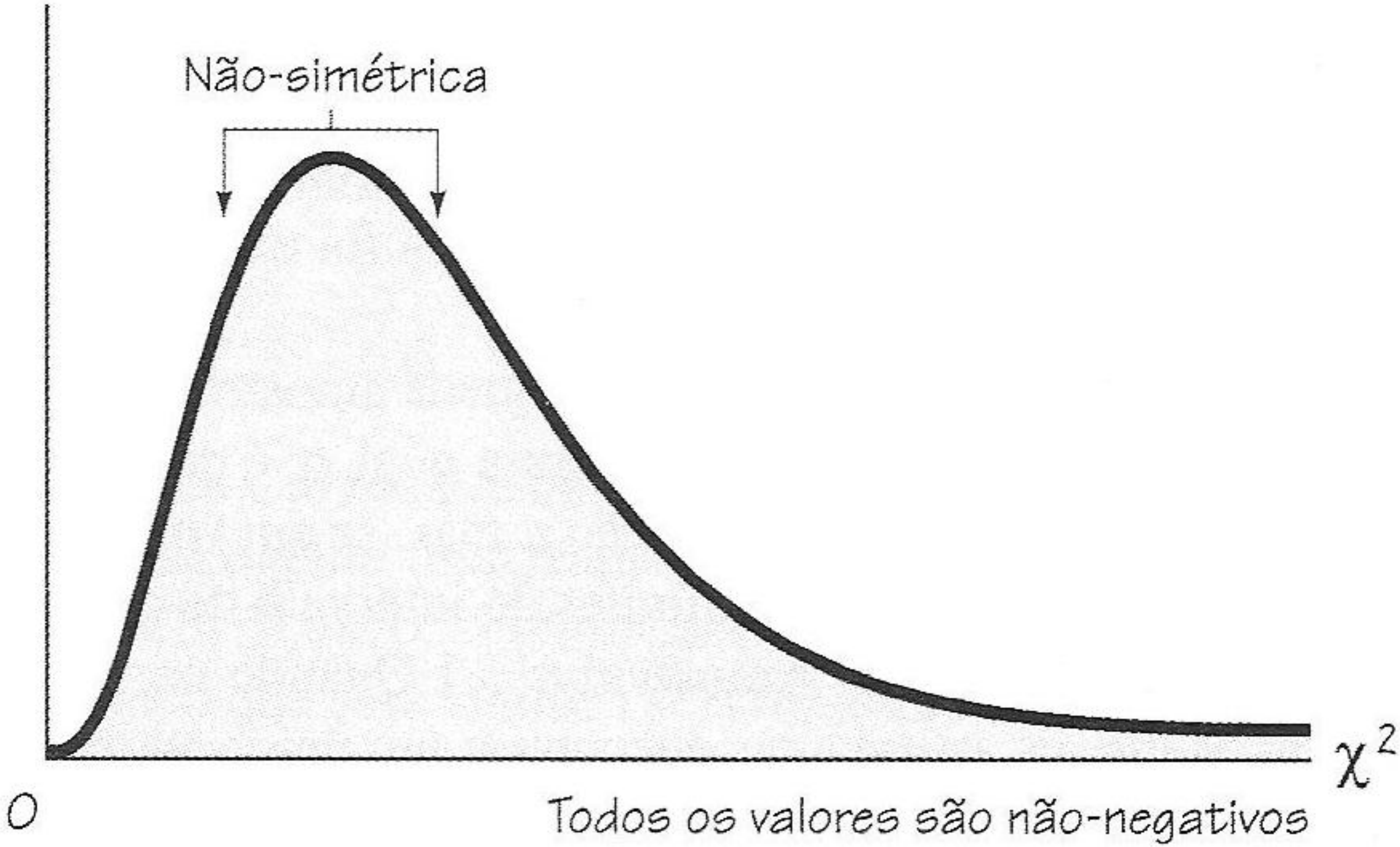
– São calculadas a variância amostral (s^2).

– Esta estatística amostral tem **distribuição qui-quadrado**:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

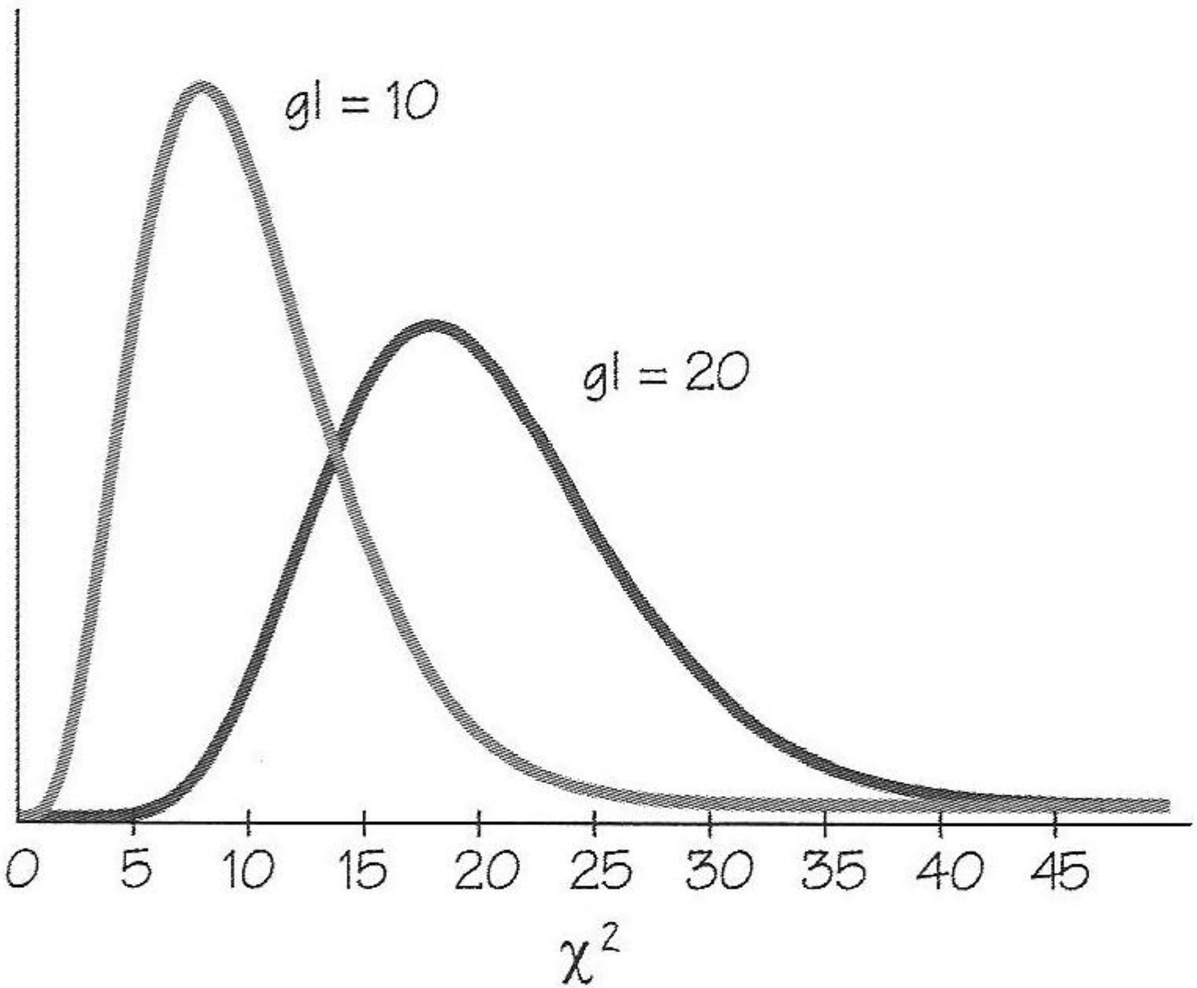
– A distribuição (χ^2) é determinada pelos graus de liberdade, por enquanto, calculada como $n - 1$.

PROPRIEDADES DA DISTRIBUIÇÃO QUI-QUADRADO



MAIS PROPRIEDADES

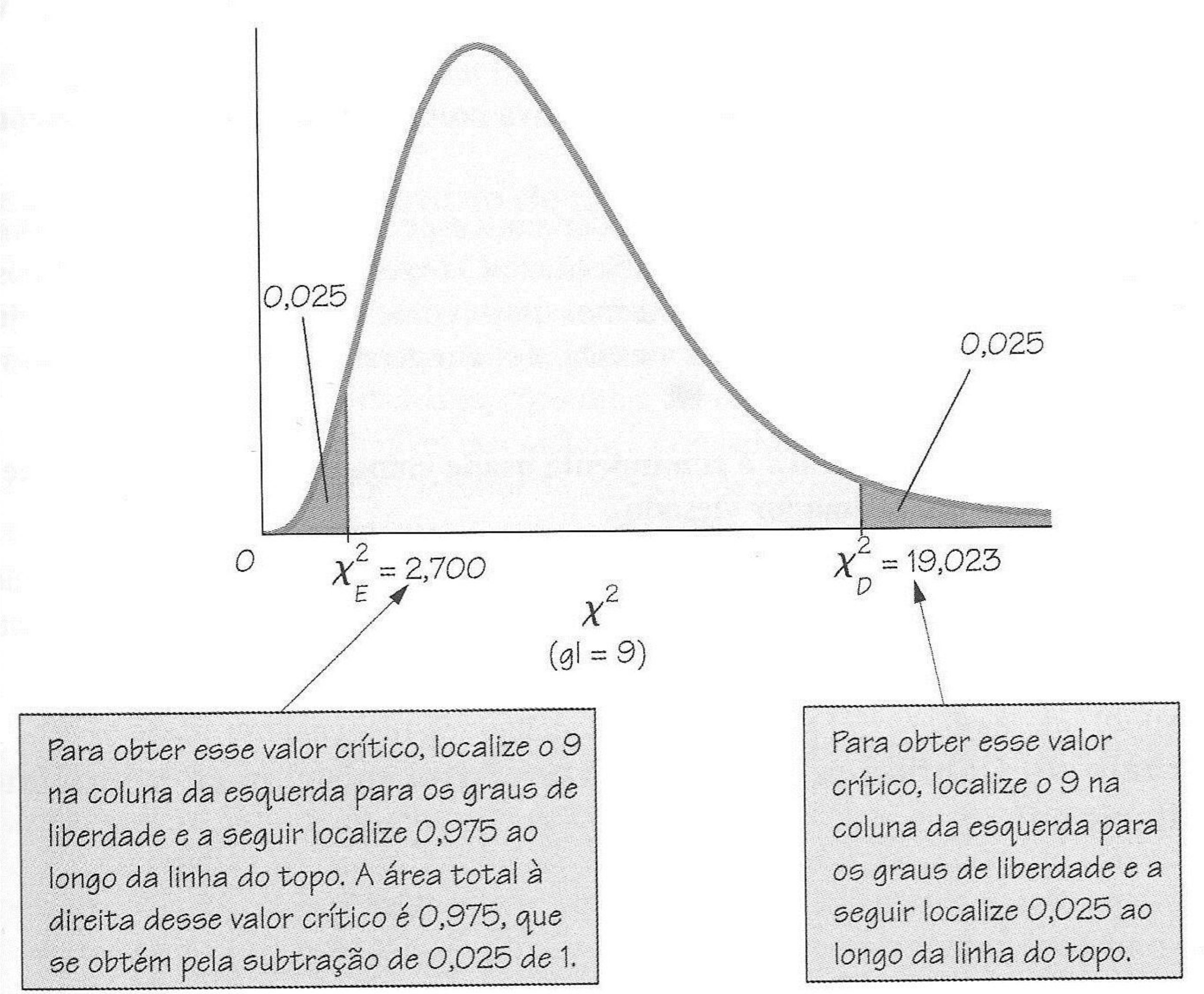
– À medida que graus de liberdade aumentam, distribuição qui-quadrado se aproxima de distribuição normal



VALORES CRÍTICOS DA DISTRIBUIÇÃO QUI-QUADRADO

– Na Tabela A-4, cada valor crítico de χ^2 corresponde à área acumulada à direita do valor crítico (ex.: n=10; área=0,025).

– Para amostra de tamanho n=10, extraída de população normalmente distribuída, a estatística χ^2 tem probabilidade 0,95 de estar entre valores críticos de 2,700 e 19,023.



ESTIMADORES DE σ^2 E σ

– A **variância amostral s^2** é a melhor estimativa pontual da variância populacional.

– Intervalo de confiança para variância populacional:

$$\frac{(n-1)s^2}{\chi^2_D} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_E}$$

– O **desvio padrão amostral s** é comumente usado como estimativa pontual de σ, mesmo sendo estimador viesado.

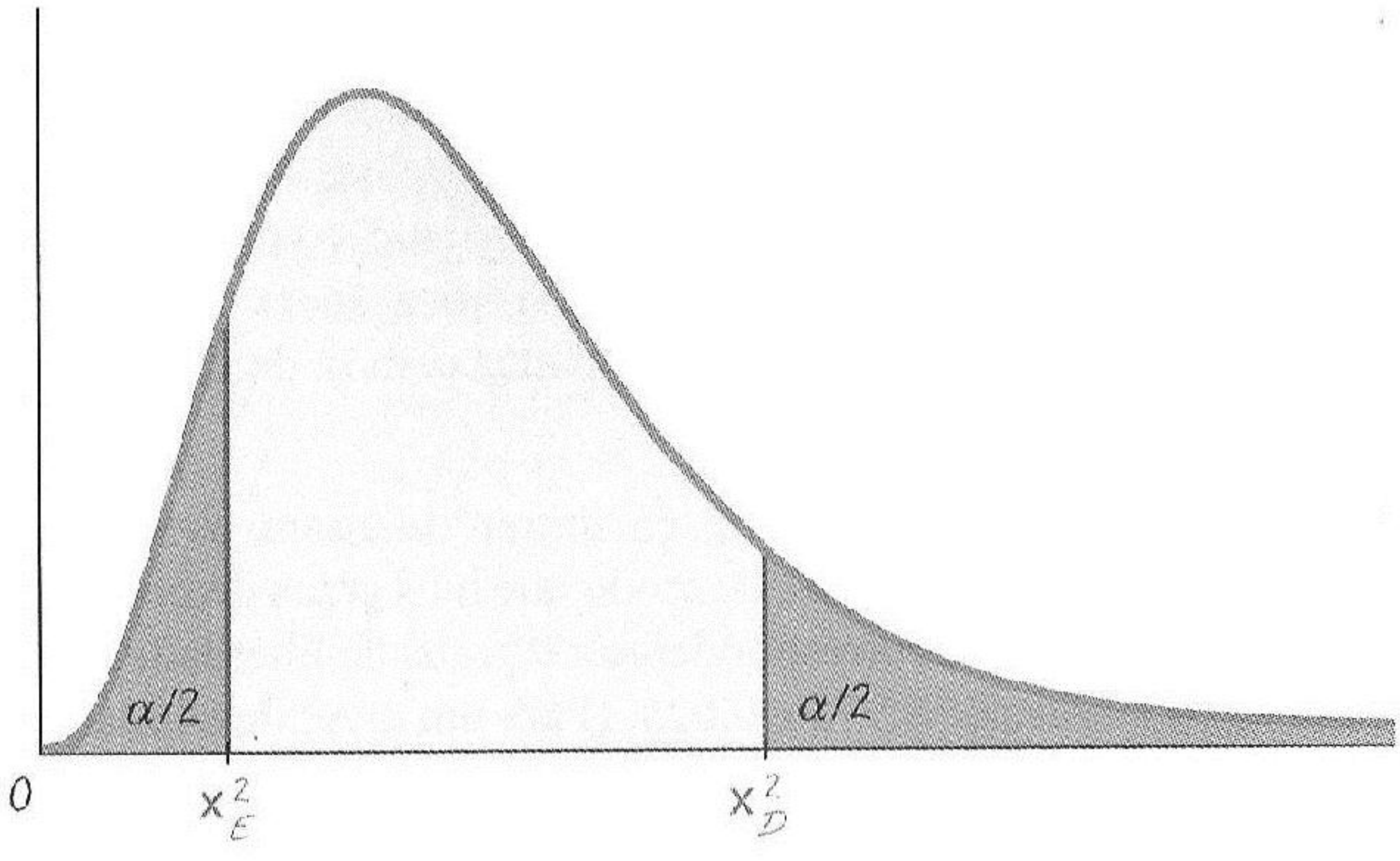
– Intervalo de confiança para desvio padrão populacional:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_D}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_E}}$$

– Sendo: χ^2_E (valor crítico da cauda esquerda) e χ^2_D (valor crítico da cauda direita).

DISTRIBUIÇÃO QUI-QUADRADO

– Valores críticos χ^2_E e χ^2_D separam áreas extremas que correspondem às variâncias amostrais que são improváveis, com probabilidade α.



CONSTRUÇÃO DE INTERVALO DE CONFIANÇA

– Verifique se requisitos são satisfeitos: (1) amostra aleatória simples; e (2) histograma ou gráfico dos quantis normais sugere população muito próxima da distribuição normal.

– Usando n – 1, ache valores críticos χ^2_E e χ^2_D, que correspondem ao nível de confiança desejado.

– Calcule os limites superior e inferior do intervalo de confiança:

$$\frac{(n-1)s^2}{\chi^2_D} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_E}$$

– Faça o mesmo para o desvio padrão (raiz quadrada).

– Arredonde limites do intervalo de confiança resultantes.

– Superposição de intervalos de confiança não deve ser usada para tirar conclusões sobre igualdade de variâncias.

TAMANHO AMOSTRAL

Tamanho amostral para σ^2		Tamanho amostral para σ	
Para se ter 95% de confiança de que s^2 está a menos de	do valor de σ^2, o tamanho amostral n deve ser, no mínimo	Para se ter 95% de confiança de que s está a menos de	do valor de σ, o tamanho amostral n deve ser, no mínimo
1%	77.207	1%	19.204
5%	3.148	5%	767
10%	805	10%	191
20%	210	20%	47
30%	97	30%	20
40%	56	40%	11
50%	37	50%	7
Para se ter 99% de confiança de que s^2 está a menos de	do valor de σ^2, o tamanho amostral n deve ser, no mínimo	Para se ter 99% de confiança de que s está a menos de	do valor de σ, o tamanho amostral n deve ser, no mínimo
1%	133.448	1%	33.218
5%	5.457	5%	1.335
10%	1.401	10%	335
20%	368	20%	84
30%	171	30%	37
40%	100	40%	21
50%	67	50%	13

APLICAÇÃO NO STATA

INSTALANDO MÓDULO STATAQUEST NO STATA

– Veja dois arquivos pdf no site para utilizar o módulo StataQuest, o qual instala o comando “ztesti”:

net cd http://www.stata.com

net cd quest7

net install quest1

net install quest2

net install quest3

PROPORÇÕES (*cii* ; *ci*)

– **Proporção para parâmetro binomial:**

cii #obs #succ, level(#)

cii 100 40, level(95)

cii 100 40, level(95) wilson

– Sendo: tamanho amostral (*obs*); número de sucessos (*succ*); e nível de confiança (*level*).

– **Para analisar variável (*x001*) de banco de dados:**

– Antes, variável binomial precisa ser codificada em 0/1:

gen homem=x001

replace homem=0 if x001==2

ci homem, binomial level(95)

ci homem, binomial wilson level(95)

– Erro padrão acima (Std.Err.) é igual a (Std.Dev.)/√*n* abaixo:

sum homem

MÉDIAS COM σ CONHECIDO (*ztesti*)

– **Média para população normal, σ conhecido:**

ztesti #obs #mean #σ #h0, level(#)

ztesti 16 24.3 6 0, level(95)

– Sendo: tamanho amostral (*obs*); média amostral (*mean*); desvio padrão populacional (*σ*); média sob hipótese nula (*h0*), mas intervalo de confiança não será afetado por H_0; e nível de confiança (*level*).

– Certos comandos terminados em "i" não utilizam informações do banco, mas sim dados digitados manualmente. É uma espécie de calculadora.

– O comando "ztesti" é instalado pelo módulo StataQuest.

MÉDIAS COM *σ* DESCONHECIDO (*cii* ; *ci*)

– **Média para população normal, *σ* desconhecido:**

cii #obs #mean #sd, level(#)

cii 9 4 3, level(95)

– Sendo: tamanho amostral (*obs*); média amostral (*mean*); desvio padrão amostral (*sd*); e nível de confiança (*level*).

– **Para analisar variável (*tradrat5*) do banco de dados:**

ci varname, level(#)

ci tradrat5, level(95)

ou

mean tradrat5

– Sendo: nome da variável (*varname*); e nível de confiança (*level*).

DESVIOS PADRÕES (*sdtesti* ; *sdtest*)

– **Desvio padrão para população normal:**

sdtesti #obs #mean #sd #val, level(#)

sdtesti 40 18 5 3, level(95)

– Sendo: tamanho amostral (*obs*); média amostral (*mean*); desvio padrão amostral (*sd*); desvio padrão populacional (*val*); e nível de confiança (*level*).

– **Para analisar variável (*tradrat5*) do banco de dados:**

sdtest varname==#, level(#)

sdtest tradrat5==1, level(95)

– Testa se desvio padrão da variável é igual a #, sendo: nome da variável (*varname*); e nível de confiança (*level*).