Métodos Quantitativos

André Amorim

Finanças Corporativas

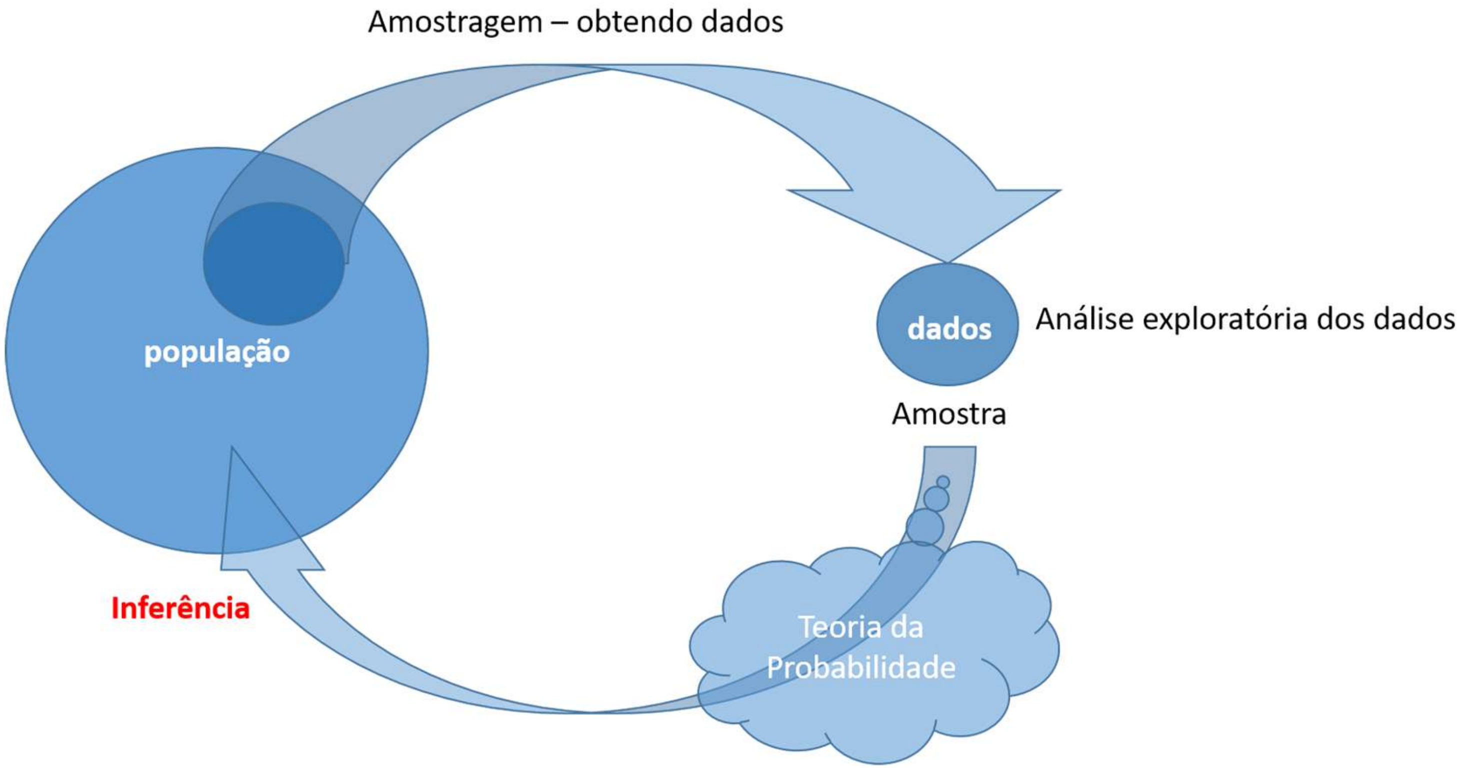
contato@andreamorim.com.br

www.andreamorim.com.br

Anhanguera

10 ª Aula – Distribuição dos estimadores

MQA



Teorema do Limite Central

Para entendermos o que significa distribuição de probabilidade da média, considere que ao observar uma variável X na população tenhamos obtido $\Omega = \{1,2,3,4\}$.

Qual o valor de μ? (Lembre-se de que μ e a media populacional)

Teorema do Limite Central

cálculo simples mostra que:

$$\mu = \frac{1+2+3+4}{4} = 2{,}5.$$

Teorema do Limite Central

Ao retirar uma amostra de tamanho 2 dessa população, conseguiríamos estimar precisamente μ por $\bar{x}$?

Ou, ainda, em todas as amostras o valor de $\bar{x}$ seria o mesmo?

Teorema do Limite Central

As respostas para essas perguntas são, respectivamente, "pouco provável" e "não".

Veja a seguir todas amostras possíveis de tamanho 2 e suas respectivas médias.

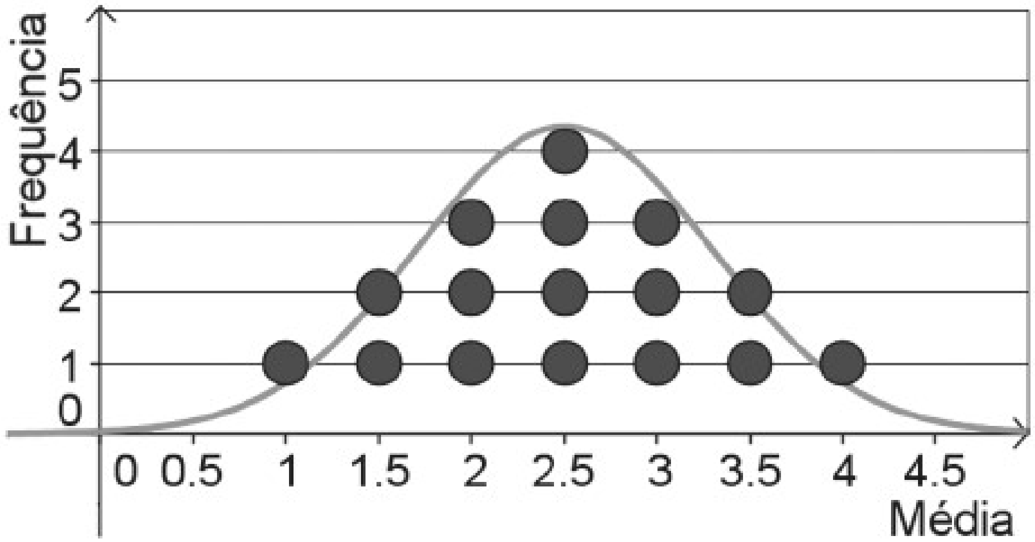
Amostra	$\bar{x}$	Amostra	$\bar{x}$	Amostra	$\bar{x}$	Amostra	$\bar{x}$
{1,1}	1,0	{2,1}	1,5	{3,1}	2,0	{4,1}	2,5
{1,2}	1,5	{2,2}	2,0	{3,2}	2,5	{4,2}	3,0
{1,3}	2,0	{2,3}	2,5	{3,3}	3,0	{4,3}	3,5
{1,4}	2,5	{2,4}	3,0	{3,4}	3,5	{4,4}	4,0

Teorema do Limite Central

Podemos montar um diagrama de dispersão com os valores das médias amostrais, como na Figura 3.5.

Observou algo de curioso na forma como os dados se distribuíram? A linha ajudou, mas esperamos que você tenha notado que os dados se distribuíram de forma semelhante a uma curva normal.

Figura 3.5 | Frequências das médias amostrais



Fonte: O autor (2015)

Teorema do Limite Central

A média amostral também pode ser considerada uma variável. Vamos calcular a média das médias amostrais ($\mu_{\bar{x}}$) e a variância das médias amostrais ($\sigma^2 \bar{x}$) para termos uma ideia quantitativa da distribuição? Observe

$$\mu_{\bar{x}} = \frac{1{,}0 + 1{,}5 + 2{,}0 + \cdots + 3{,}0 + 3{,}5 + 4{,}0}{16} = \frac{40}{16} = 2{,}5.$$

$$\sigma^2{}_{\bar{x}} = \frac{(1{,}0 - 2{,}5)^2 + (1{,}5 - 2{,}5)^2 + \cdots + (3{,}5 - 2{,}5)^2 + (4{,}0 - 2{,}5)^2}{16} = \frac{10}{16} = 0{,}625.$$

Teorema do Limite Central

A média amostral também pode ser considerada uma variável. Vamos calcular a média das médias amostrais ($\mu_{\bar{x}}$) e a variância das médias amostrais ($\sigma^2\,\bar{x}$) para termos uma ideia quantitativa da distribuição? Observe

$$\mu_{\bar{x}} = \frac{1{,}0 + 1{,}5 + 2{,}0 + \cdots + 3{,}0 + 3{,}5 + 4{,}0}{16} = \frac{40}{16} = 2{,}5.$$

Observe que a media das medias amostrais e exatamente igual a media da população, ou seja, $\mu_x = \mu$.

$$\sigma^2{}_{\bar{x}} = \frac{(1{,}0 - 2{,}5)^2 + (1{,}5 - 2{,}5)^2 + \cdots + (3{,}5 - 2{,}5)^2 + (4{,}0 - 2{,}5)^2}{16} = \frac{10}{16} = 0{,}625.$$

Teorema do Limite Central

E quanto à variância, será que $\sigma^2_{\bar{x}} = \sigma^2$? Vejamos:

$$\sigma^2 = \frac{(1{,}0 - 2{,}5)^2 + (2{,}0 - 2{,}5)^2 + (3{,}0 - 2{,}5)^2 + (4{,}0 - 2{,}5)^2}{4} = \frac{5}{4} = 1{,}25.$$

- Note que $\sigma^2_{x} < \sigma^2$, resultado que pode ser mais bem compreendido com a leitura do Teorema do Limite Central (TLC)

Teorema do Limite Central

De acordo com Morettin (2010), “o TLC diz que para n amostras aleatórias simples, retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2/n”.

Teorema do Limite Central

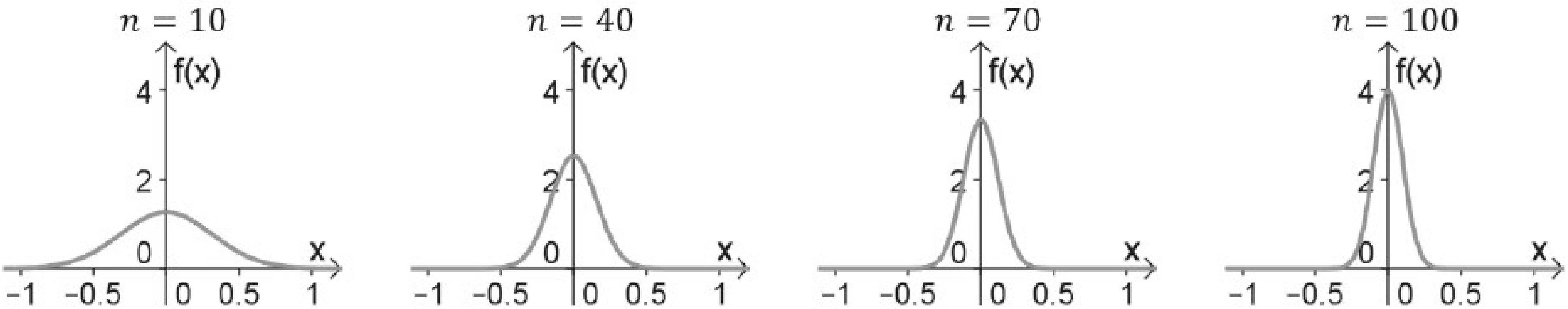
O TLC é de extrema importância para a estatística inferencial e tem implicações muito interessantes.

Observe que, apesar de ele não dizer nada a respeito da distribuição da população, afirma que a distribuição amostral da média aproximasse de uma curva normal, e, além disso, essa distribuição tem a mesma média que a população e variância σ^2/n, isto é, a mesma variância que a população, mas dividida por n.

Teorema do Limite Central

A partir desse resultado, concluímos que, quanto maior o número de amostras, mais precisão teremos para a média, pois σ^2/n diminui conforme n aumenta. Podemos visualizar esse resultado na Figura 3.6.

Figura 3.6 | Distribuição amostral da média $\bar{x}$ de uma população $X \sim N(0,1)$ para vários valores de n



Fonte: O autor (2015)

Teorema do Limite Central

A partir desse resultado, concluímos que, quanto maior o número de amostras, mais precisão teremos para a média, pois σ^2/n diminui conforme n aumenta. Podemos visualizar esse resultado na Figura 3.6.

Se $X \sim N(0,1)$, a f.d.p. da variável $\bar{x}$ pode ser escrita como

$$f(x;\ 0, 1/n) = \sqrt{\frac{n}{2\pi}}\, e^{-nx^2/2}.$$

Com base no TLC ha ainda dois resultados interessantes que podemos enunciar.

Teorema do Limite Central

Assimile

De acordo com Morettin (2010), "sendo X uma variável com média μ e variância σ^2 finita, e $\bar{x}$ a variável média amostral, então a variável $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$ tem distribuição normal com média 0 e variância 1, ou seja, $Z \sim N(0,1)$".

Podemos ainda definir a variável e como a diferença entre o estimador $\bar{x}$ e o parâmetro μ, ou seja, $e = \bar{x} - \mu$.

Determinando o tamanho de uma amostra

Vamos supor que o erro máximo que estipulamos para estimar a média populacional seja ε.

Desse modo, qualquer valor $\bar{x}$ no intervalo $[\mu - \varepsilon, \mu + \varepsilon]$ nos deixara satisfeitos para essa estimativa.

Para assimilar melhor, suponha que queiramos estimar a verdadeira média populacional μ = 1,70 m da altura de certo grupo de atletas e, para isso, queiramos cometer um erro máximo de ε = 2 cm.

Portanto, qualquer valor de $\bar{x}$ pertencente ao intervalo [1,68 m; 1,72 m] servirá. Além disso, para acompanhar essa estimativa, suponha que queiramos ter uma probabilidade de acerto de y (95%, por exemplo), uma margem de segurança.

Determinando o tamanho de uma amostra

Matematicamente, afirmar que $\bar{x}$ pertence ao intervalo $[\mu - \varepsilon, \mu + \varepsilon]$ implica $\mu - \varepsilon \leq \bar{x} \leq \mu + \varepsilon$ ou, $|\bar{x} - \mu| \leq \varepsilon$. Além disso, ter uma probabilidade de acerto de y que $|\bar{x} - \mu| \leq \varepsilon$ pode ser traduzido matematicamente por $P(|\bar{x} - \mu| \leq \varepsilon) \geq \gamma$.

Determinando o tamanho de uma amostra

Com base nos resultados obtidos do TLC, temos:

$$P(|\bar{x}-\mu| \leq \varepsilon) \geq \gamma \Rightarrow P(-\varepsilon \leq \bar{x}-\mu \leq +\varepsilon) \geq \gamma \Rightarrow P\left(-\frac{\sqrt{n}\,\varepsilon}{\sigma} \leq \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma} \leq +\frac{\sqrt{n}\,\varepsilon}{\sigma}\right) \geq \gamma.$$

Lembre-se de que $Z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$, logo:

$$P\left(-\frac{\sqrt{n}\,\varepsilon}{\sigma} \leq \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma} \leq +\frac{\sqrt{n}\,\varepsilon}{\sigma}\right) \geq \gamma \Rightarrow P\left(-\frac{\sqrt{n}\,\varepsilon}{\sigma} \leq Z \leq +\frac{\sqrt{n}\,\varepsilon}{\sigma}\right) \geq \gamma.$$

Determinando o tamanho de uma amostra

Com base nos resultados obtidos do TLC, temos:

Dado um valor y podemos obter na tabela Z um valor z_γ tal que $P(-z_\gamma \leq Z \leq +z_\gamma) \geq \gamma$ e ainda:

$$z_\gamma = \frac{\sqrt{n}\,\varepsilon}{\sigma} \Rightarrow z_\gamma \sigma = \sqrt{n}\,\varepsilon \Rightarrow \sqrt{n} = \frac{z_\gamma \sigma}{\varepsilon} \Rightarrow n = \frac{{z_\gamma}^2 \sigma^2}{\varepsilon^2}.$$

Determinando o tamanho de uma amostra

Observe que, se tivermos o conhecimento de σ^2, podemos estimar n em função de y e ε, prefixados, ou estimar ε em função de y e n. Com base na última igualdade podemos justificar a afirmativa feita na Unidade 2 de que o erro diminui à medida que o tamanho da amostra aumenta, pois:

$$n = \frac{z_{\gamma}{}^{2}\sigma^2}{\varepsilon^2} \Rightarrow \varepsilon^2 = \frac{z_{\gamma}^2\sigma^2}{n} \Rightarrow \varepsilon = \frac{\sqrt{z_{\gamma}^2\sigma^2}}{\sqrt{n}}.$$

Podemos agora, observando a última igualdade, ver claramente que, se n aumenta ($n \to \infty$), o erro diminui ($\varepsilon \to 0$).

Exemplificando

Seja uma variável X~N(μ,4) observada em dada população. Com precisão de:

95%, qual o erro máximo que cometemos ao estimar a verdadeira média dessa população com base em uma amostra de tamanho n = 30?

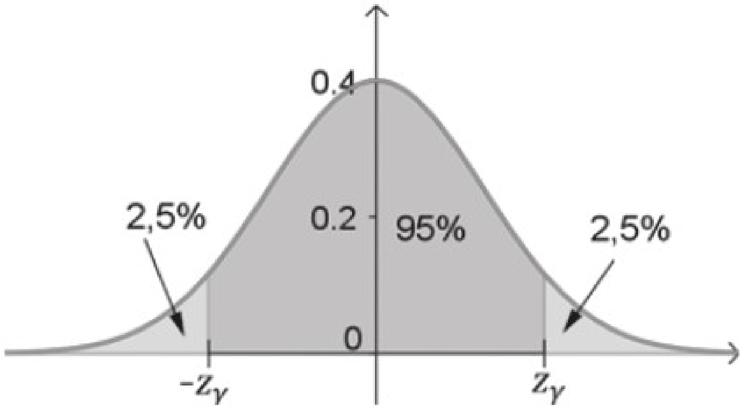
90%, qual o tamanho da amostra que deve ser coletada para que o erro seja de, no máximo, ε = 1?

Resolução:

a) Observe que a fórmula do erro $\varepsilon = \sqrt{z_\gamma{}^2\sigma^2} \big/ \sqrt{n}$ depende de z_γ, σ^2 e n. O parâmetro $\sigma^2 = 4$ foi dado e $n = 30$. Resta determinar z_γ, em que $\gamma = 95\% = 0{,}95$, para que tenhamos $P(-z_\gamma \leq Z \leq +z_\gamma) \geq \gamma = 0{,}95$. Observe a Figura 3.7.

Exemplificando

Figura 3.7 | Região correspondente a $P(-z_\gamma \leq Z \leq +z_\gamma) \geq 0{,}95$

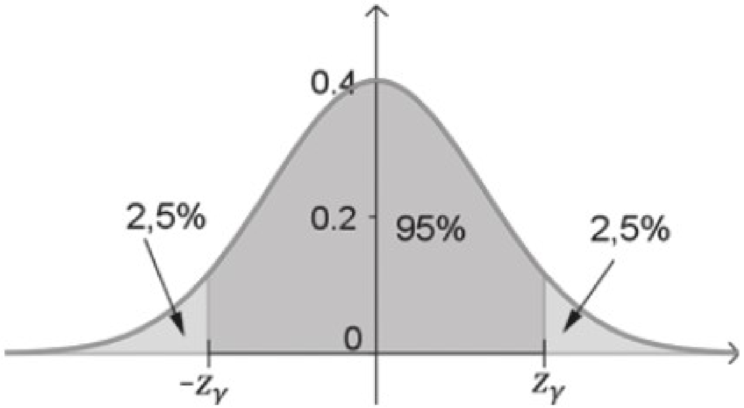


Fonte: O autor (2015)

Veja que o valor z_γ deve ser tal que $P(Z \leq z_\gamma) \geq 0{,}025 + 0{,}95 = 0{,}975$.
Consultando a tabela Z, temos $z_\gamma = 1{,}96$. Logo

Exemplificando

Figura 3.7 | Região correspondente a $P(-z_\gamma \leq Z \leq +z_\gamma) \geq 0{,}95$

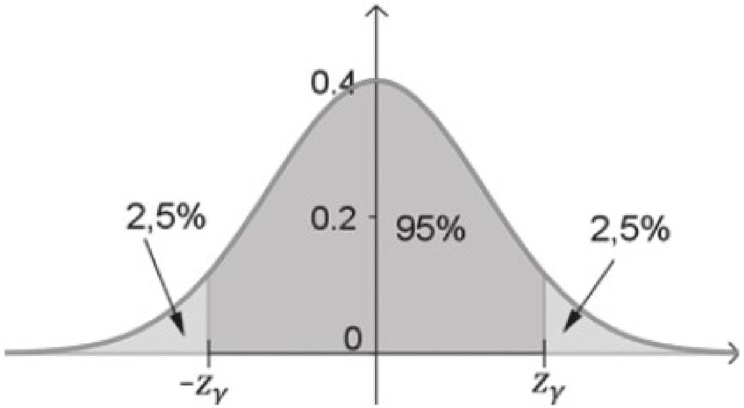


Fonte: O autor (2015)

$$\varepsilon = \sqrt{z_\gamma{}^2\sigma^2}\Big/\sqrt{n} = \sqrt{1{,}96^2 \cdot 4}\Big/\sqrt{30} \cong 0{,}72.$$

Exemplificando

Figura 3.7 | Região correspondente a $P(-z_\gamma \leq Z \leq +z_\gamma) \geq 0{,}95$



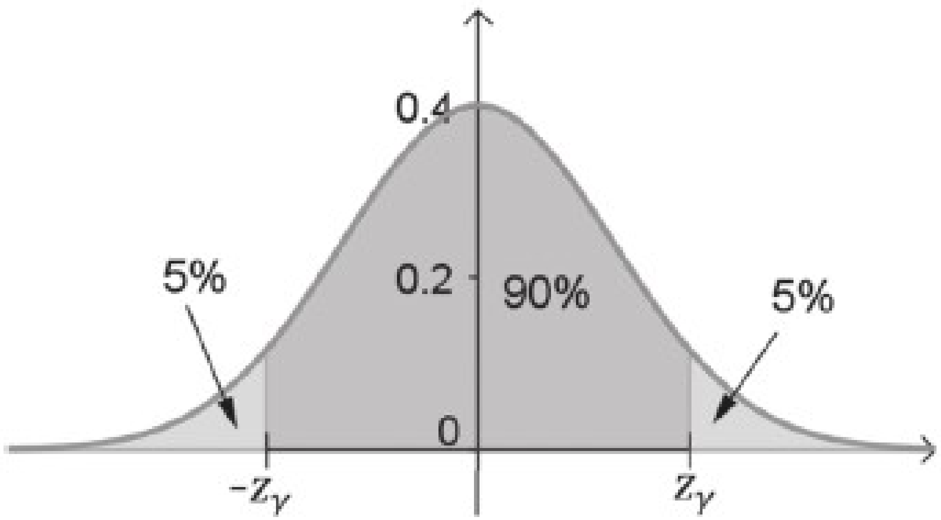
Fonte: O autor (2015)

Portanto, com precisão de 95%, o erro máximo que cometemos ao estimar a verdadeira média dessa população com base em uma amostra de tamanho $n = 30$ é $\varepsilon = 0{,}72$.

Exemplificando

b) Observe que, para determinar o tamanho da amostra, devemos utilizar a fórmula $n = \frac{z_\gamma{}^2 \sigma^2}{\varepsilon^2}$, em que σ^2 e ε foram dados, e z_γ deve ser consultado na tabela Z para $\gamma = 90\% = 0{,}90$. Veja a Figura 3.8.

Figura 3.8 | Região correspondente a $P(-z_\gamma \le Z \le +z_\gamma) \ge 0{,}90$



Fonte: O autor (2015)

Exemplificando

Veja que o valor z_γ deve ser tal que $P(Z \leq z_\gamma) \geq 0{,}05 + 0{,}90 = 0{,}95$. Consultando a tabela Z, temos $z_\gamma = 1{,}65$. Logo $n = \frac{{z_\gamma}^2 \sigma^2}{\varepsilon^2} = \frac{1{,}65^2 \cdot 4}{1^2} = 10{,}89 \cong 11$.

Portanto, com precisão de 90%, para ter erro máximo $\varepsilon = 1$, temos de obter uma amostra de tamanho $n = 11$ para estimar a verdadeira média da população.

FIM