Estimação da média populacional μ

Estudamos algumas distribuições teóricas de probabilidade: distribuição *binomial* e *normal*.

Probabilidade $\Rightarrow$ os parâmetros da distribuição são conhecidos $\Rightarrow$ calculamos probabilidades

Inferência $\Rightarrow$ os valores desses parâmetros são desconhecidos $\Rightarrow$ queremos estimá-los.

Parâmetro: quantidade desconhecida de uma característica **da população** e sobre a qual temos interesse.

Exemplos: μ - *média da característica da população*:

μ : taxa média de glicose de mulheres com idade superior a 60 anos, em certa localidade;

p – *proporção de "indivíduos" em uma população com determinada característica.*

p: proporção de pacientes com menos de 40 anos diagnosticados com câncer nos pulmões.

X: variável de interesse : ***Renda***

Vamos observar n elementos, extraídos ao acaso da população, de forma independente;

Para cada elemento selecionado, observamos o valor da variável X de interesse.

Obtemos, então, uma **amostra aleatória** ($a.a.$) de tamanho n de X, que representamos por

$$X_1, X_2, \ldots, X_n,$$

sendo X_i a variável de interesse para o i-ésimo indivíduo da amostra.

Uma vez selecionada a amostra saberemos a renda de ***João*** (x_1)

Estimador: função dos elementos da amostra, construída com a finalidade de representar, ou estimar, um parâmetro da característica de interesse X na população.

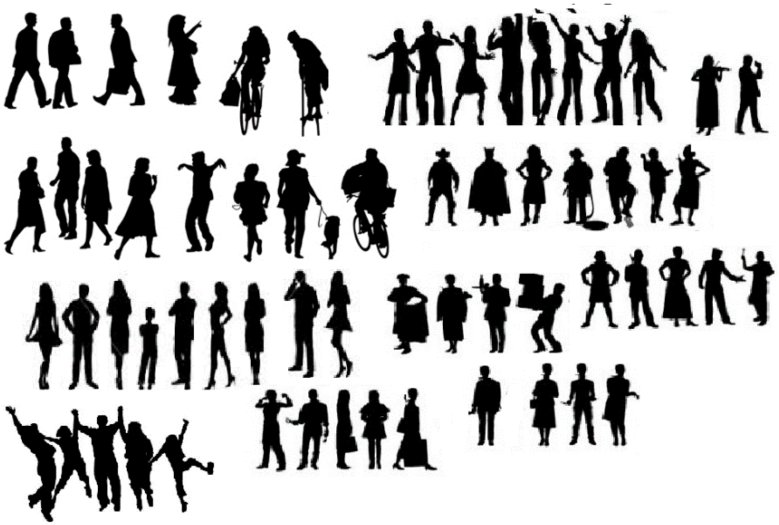
→ **Estimador** (ou estatística) $\Rightarrow \boldsymbol{f}\,(\boldsymbol{X_1}, \boldsymbol{X_2},\ldots,\boldsymbol{X_n})$.

Ex.: $\bar{X}$: média amostral (estimador da média μ da característica X da população).

$\hat{p}$: proporção amostral (estimador da proporção p populacional).

Vamos discutir estes dois exemplos mais importantes: $\bar{X}$ *e* $\hat{p}$

Estimativa: valor numérico assumido pelo estimador, para a amostra selecionada.



X - variável de interesse: ***Renda***

Por exemplo, obter a distribuição amostral da Média

Amostra 1

$\bar{x}_1$

Amostra 2

$\bar{x}_2$

...

...

Amostra *k*

$\bar{x}_k$

...

...

População das médias de amostras de tamanho n

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

Média amostral
=
estimador de μ

Exemplos:

μ : peso médio de homens na faixa etária de 20 a 30 anos, em uma certa localidade;

μ : salário médio dos empregados da indústria metalúrgica em São Bernardo do Campo;

μ : taxa média de glicose em indivíduos do sexo feminino com idade superior a 60 anos, em determinada localidade;

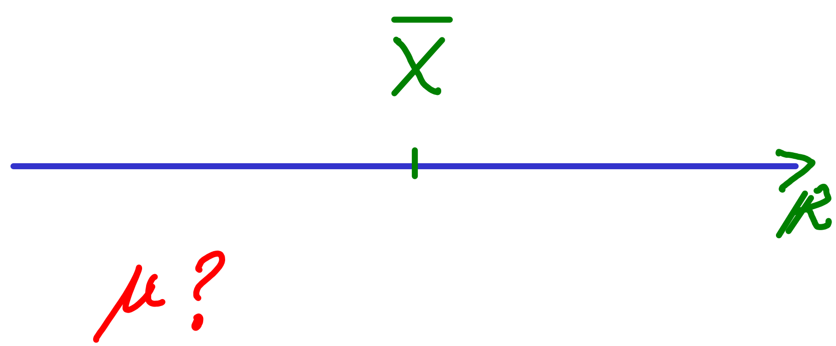
μ : comprimento médio de tartarugas adultas de uma certa espécie;

μ : pontuação média obtida no *ENEM* em 2014.

Um ***estimador pontual para*** μ, baseado numa amostra aleatória de tamanho n, é dado pela *média amostral*,

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}\ .$$

Se observamos os valores x_1, x_2, …x_n para as variáveis $X_1, \ldots, X_n$ obtemos $\overline{x} = \dfrac{x_1 + \ldots + x_n}{n}$, que denominamos ***estimativa pontual para*** μ**.**



Exemplo 1: Considere

X_i: taxa de glicose do indivíduo i do sexo feminino, com idade superior a 60 anos, em certa localidade, i = 1, …, n e

μ: taxa média de glicose de mulheres, com idade superior a 60 anos, em certa localidade;

Suponha que foram selecionadas n=10 mulheres, nessa faixa etária dessa localidade e suas taxas de glicose, em mg/dl, foram

102; 95; 110; 104; 123; 92; 112; 89; 97;101.

A ***estimativa pontual*** (***média amostral***) ***para*** μ é dada por:

$$\bar{x} = \frac{102 + 85 + 110 + 104 + 123 + 92 + 112 + 89 + 97 + 101}{10} = \frac{1015}{10} = 101{,}15\ mg/dl.$$

Note que outra amostra de mesmo tamanho pode levar a uma outra estimativa pontual para μ

$\bar{X}$ *é uma variável aleatória!*

Estimativa por intervalo ou intervalo de confiança

- Para uma amostra observada, os estimadores pontuais fornecem como estimativa um único valor numérico para o parâmetro.

- Os estimadores pontuais são variáveis aleatórias e, portanto, possuem uma distribuição de probabilidade, em geral, denominada ***distribuição amostral do estimador.***

Ideia: construir **intervalos de confiança**, que incorporem à estimativa pontual informações a respeito de sua variabilidade (erro amostral).

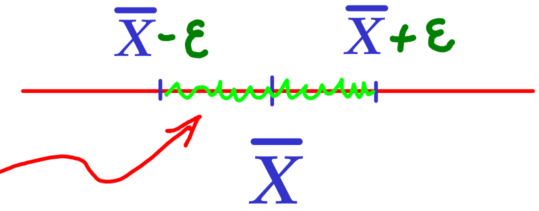
Intervalos de confiança são obtidos por meio da ***distribuição amostral do estimador pontual*.**

Um ***estimador intervalar*** ou ***intervalo de confiança para*** μ tem a forma

Intervalo ao redor de X

"vai conter μ com certa chance, fixada γ "

$$\left[\overline{X} - \varepsilon \; ; \; \overline{X} + \varepsilon\right]$$



Estimador pontual

sendo ε a margem de erro, calculada a partir da *distribuição de probabilidade de* $\overline{X}$

Como é a distribuição de probabilidade da média amostral ?

"Quão grande é o erro entre $\overline{X}$ e μ ?"

Estimador (variável aleatória)

Parâmetro (valor numérico desconhecido)

ALGUNS RESULTADOS IMPORTANTES

RESULTADO 1:

Para qualquer variável aleatória X, com média μ e variância σ^2, temos que, considerando uma amostra aleatória de tamanho n de X,

$$E(\overline{X}) = \mu \qquad \text{e} \qquad Var(\overline{X}) = \frac{\sigma^2}{n}$$

O desvio padrão $\sqrt{\dfrac{\sigma^2}{n}} = \dfrac{\sigma}{\sqrt{n}}$ é denominado *erro padrão da média amostral.*

RESULTADO 2: *(Teorema Limite Central)*

Se a variável aleatória $\boldsymbol{X}$, na <u>população,</u> tem distribuição com média μ e variância σ^2, então, para uma amostra aleatória de tamanho n de X,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Aproximadamente, para n grande

e, portanto

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

Aproximadamente, para n grande

Seja $P(\varepsilon) = \gamma$, a probabilidade da média amostral $\overline{X}$ estar a uma distância de, no máximo ε, da média populacional μ (desconhecida), ou seja,

$$\gamma = P(\varepsilon) = P(|\overline{X} - \mu| \leq \varepsilon).$$

"a distância entre $\overline{X}$ e μ é menor ou igual a ε "

"Intervalo contem μ "

A probabilidade $P(\varepsilon)$ é também denominada coeficiente de confiança do intervalo, que denotamos por γ (gama).

$\overline{X}-\varepsilon$ $\overline{X}+\varepsilon$ $\overline{X}$

$\overline{X} \geq \mu-\varepsilon$

$\overline{X}-\varepsilon \leq \mu \leq \overline{X}+\varepsilon$

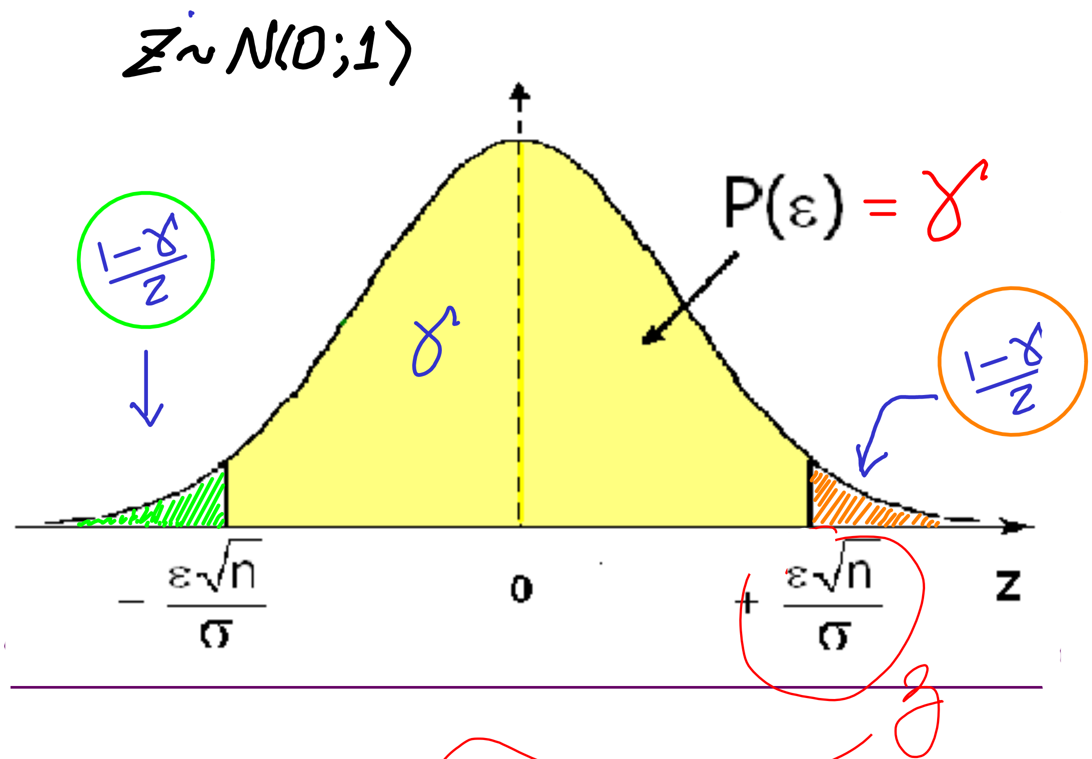
$\overline{X} \leq \mu+\varepsilon$

Desse modo, temos

$$P(\varepsilon) = P(|\overline{X} - \mu| \leq \varepsilon) = P(\mu - \varepsilon \leq \overline{X} \leq \mu + \varepsilon)$$

$$= P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \leq \underbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}_{= Z} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} \leq Z \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

sendo $Z \sim N(0,1)$.



Denotando $\dfrac{\varepsilon\sqrt{n}}{\sigma} = z$,

Determinado a partir da "Tabela da Normal-Padrão"

temos que $\gamma = P(-z \le Z \le z)$.

$A(z) = \gamma + \dfrac{(1-\gamma)}{2}$

(estava errado nos slides comentados)

Assim, conhecendo-se o coeficiente de confiança γ obtemos z.

Erro na estimativa intervalar

Da igualdade $z = \varepsilon \dfrac{\sqrt{n}}{\sigma}$, segue que a margem de erro ε é dada por

$$\varepsilon = z \frac{\sigma}{\sqrt{n}},$$

sendo z tal que $\gamma = P(-z \le \boldsymbol{Z} \le z)$, com $\boldsymbol{Z} \sim N(0, 1)$.

O intervalo de confiança para a média μ, com coeficiente de confiança γ está dado por

$$\left[\overline{X} - z \frac{\sigma}{\sqrt{n}} \; ; \; \overline{X} + z \frac{\sigma}{\sqrt{n}} \right],$$

sendo σ o desvio padrão (conhecido) de X.

"amosta aleatória"

Antes de selecionarmos uma $a.a.$, a probabilidade de que o intervalo

$$\left[\overline{X} - z\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z\frac{\sigma}{\sqrt{n}}\right],$$

contenha a média μ verdadeira da população é γ.

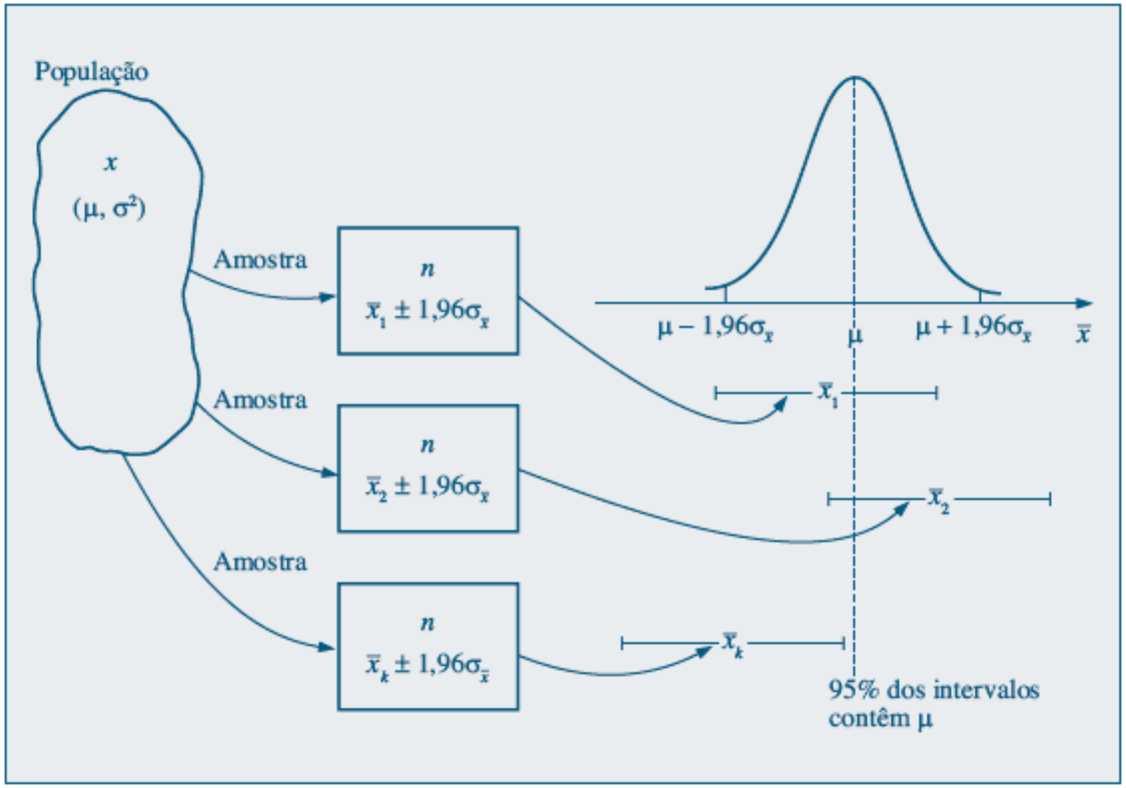
Para o valor observado $\bar{x}$ de $\overline{X}$, o intervalo de 95% de confiança será $[\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}};\ \ \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}]$;

$\gamma = 0{,}95 \Rightarrow z = 1{,}96$

(Verifique!)

<u>Não podemos dizer</u> que há uma probabilidade de 95% de que o valor de μ pertença a esse intervalo de números; μ é fixo e está ou não nesse intervalo.

Figura 11.3 Significado de um IC para μ, com $\gamma = 0{,}95$ e σ^2 conhecido.



Fonte: "Estatística Básica", W. O. Bussab e P. A. Morettin, Ed. Saraiva, 5a. edição, 2002.

Interpretação frequentista:

Se extrairmos 100 *a.a.* de tamanho n da população e, para cada uma delas, construirmos um intervalo de confiança de 95%, esperamos que, aproximadamente, 95 dos intervalos contenham a média μ verdadeira da população e 5 não.

Exemplo

Deseja-se estimar o tempo médio de estudo (em anos) da população adulta de um município. Sabe-se que o tempo de estudo tem distribuição normal com desvio padrão σ = 2,6 anos. Foram entrevistados n = 25 indivíduos, obtendo-se para essa amostra, um tempo médio de estudo igual a 10,5 anos. Obter um intervalo de 90% de confiança para o tempo médio de estudo na população.

X : tempo de estudo, em anos, então $X \sim N(\mu;\, 2{,}6^2)$

n = 25 $\Rightarrow$ $\bar{x}$ = 10,5 anos

γ = 0,90 $\Rightarrow$ z = 1,65

A estimativa intervalar com 90% de confiança é dada por:

$$\left[\bar{X} - z\frac{\sigma}{\sqrt{n}}\,;\,\bar{X} + z\frac{\sigma}{\sqrt{n}}\right] =$$

$$\left[10{,}5 - 1{,}65\frac{2{,}6}{\sqrt{25}}\,;10{,}5 + 1{,}65\frac{2{,}6}{\sqrt{25}}\right] =$$

$$\left[10{,}5 - 0{,}86\,;\; 10{,}5 + 0{,}86\right] =$$

$$\left[9{,}64\,;\; 11{,}36\right]$$

Dimensionamento da amostra

"Qual é o n?"

A partir da relação $\varepsilon = z \dfrac{\sigma}{\sqrt{n}}$

o tamanho da amostra $\boldsymbol{n}$ é determinado por

$$n = \left(\frac{z}{\varepsilon}\right)^2 \sigma^2,$$

onde ε é a margem de erro, z é tal que

$$\gamma = P(-z \leq Z \leq z), \quad Z \sim N(0, 1)$$

e σ dé o desvio padrão (conhecido) de X.

Exemplo

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

X : renda per-capita domiciliar na região $\Rightarrow X \sim N(\mu; 250^2)$

$\varepsilon = 50$

$\gamma = 0,95 \Rightarrow z = 1,96$

$n = ??$

Então,

$$n = \left(\frac{z}{\varepsilon}\right)^2 \sigma^2 = \left(\frac{1,96}{50}\right)^2 \times (250)^2 = 96,04$$

Aproximadamente, 97 domicílios devem ser consultados.

Na prática, em geral, não conhecemos a variância populacional σ^2.

(Ou seja, tanto μ quanto σ são desconhecidos)

RESULTADO

Se X tem distribuição normal, com média μ e variância σ^2, então, para uma amostra aleatória de tamanho n de X, temos que

t_{n-1}

Distribuição "t de Student"

$$T = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1},$$

A distribuição t depende de n, MAS

$T \sim N(0,1)$

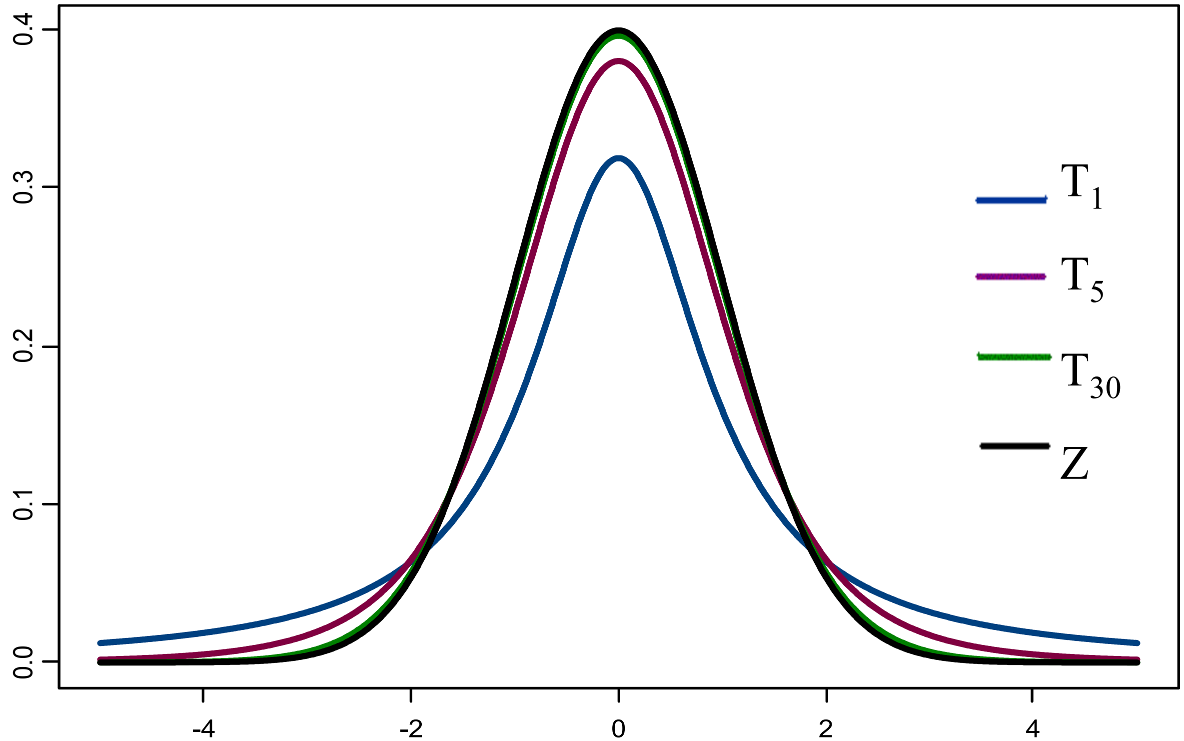
"aproximadamente se n for grande"

Estimador da variância σ^2:

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

é a variância amostral.

Distribuição *t* de *Student*



- tem caudas mais densas do que a distribuição normal;
- valores extremos são mais prováveis de ocorrer com a distribuição *t* do que com a normal padrão;

Exemplo

A quantidade de colesterol X no sangue das alunas de uma universidade tem uma distribuição aproximadamente normal com média e desvio padrão desconhecidos.

Para estimar a quantidade média de colesterol μ é selecionada uma amostra de 60 alunas. A média e o desvio padrão amostrais encontrados são 182 mg/dl e 50 mg/dl, respectivamente. Determine um intervalo de confiança com coeficiente de confiança de 90% para μ.

X: quantidade de colesterol no sangue das alunas da universidade $\Rightarrow X \sim N(\mu, \sigma^2)$

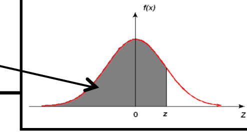
$n = 60 \Rightarrow \bar{x} = 182$ e $s = 50$

$\gamma = 0{,}90 \Rightarrow$ ***z=1,65, usando a tabela da Normal Padrão***

Substituindo os valores:

$$IC(\mu; 90\%) = (182 - 1{,}65\,\frac{50}{\sqrt{60}}\ ;\ 182 + 1{,}65\,\frac{50}{\sqrt{60}}) = \boldsymbol{(\ 171{,}35\ ;\ 192{,}25\ )}$$

Distribuição Normal : Valores de P(Z ≤ z) = A(z)



Segunda decimal de z

Parte inteira e primeira decimal de z

z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Volta