MAE-212: Introdução à Probabilidade e à Estatística II

Aula 02

Inferência Estatística

Distribuições Amostrais

O problema da inferência estatística: fazer uma afirmação sobre os parâmetros da população θ (média, variância, etc) através da amostra.

Usaremos uma AAS de n elementos sorteados dessa população, nossa decisão será baseada na estatística $T = f(X_1, ..., X_n)$.

Colhida essa amostra, observamos um particular valor de T: t_0, e baseados nesse valor é que faremos a afirmação sobre o parâmetro populacional θ.

***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.*

$\boldsymbol{p}$: proporção de moradores do município de São Paulo que foram infectados pelo vírus Covid-19 no período de março a agosto de 2020.

População

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

$$X_1, X_2, ..., X_n,$$

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

Uma vez selecionada a amostra saberemos a renda de ***João*** **(**$\boldsymbol{x_1}$**)**

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

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

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

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

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

Ex.: $\bar{x}$ é o valor de $\overline{X}$ para a amostra observada.

Os estimadores $\overline{X}$ (**média amostral**) e $\hat{p}$ (**proporção amostral**) são intuitivos e têm boas propriedades.

Estimadores são funções de variáveis aleatórias e, portanto, eles também são variáveis aleatórias.

Consequentemente, têm uma distribuição de probabilidades, denominada ***distribuição amostral do estimador.***

População

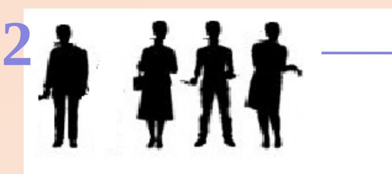
X - variável de interesse: ***Renda***

Por exemplo, obter a distribuição amostral da Média

Amostra 1

$\bar{x}_1$

Amostra 2



$\bar{x}_2$

...

...

Amostra *k*

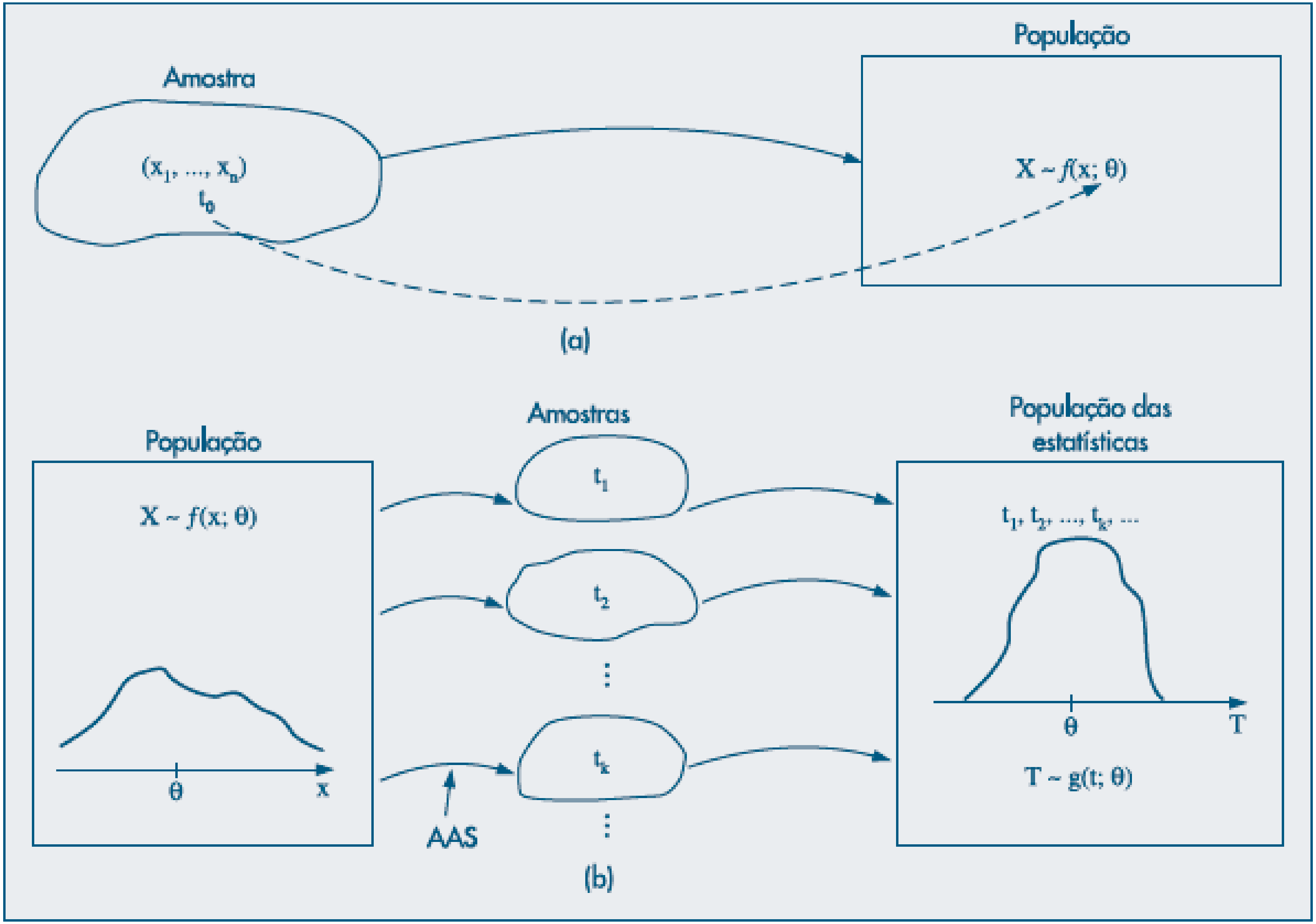
$\bar{x}_k$

...

...

População das médias de amostras de tamanho n

Figura 10.1: (a) Esquema de inferência sobre θ.
(b) Distribuição amostral da estatística T.



Distribuição amostral da média

Exemplo 1: Considere uma população em que uma variável X assume um dos valores do conjunto {1, 3, 5, 5, 7}. A distribuição de probabilidade de X é dada por

x	1	3	5	7
P(X=x)	1/5	1/5	2/5	1/5

É fácil ver que $\mu_x = E(X) = 4{,}2$,

$$\sigma_x^2 = Var(X) = 4{,}16.$$

Vamos relacionar <u>todas</u> as amostras possíveis de tamanho *n* = 2, selecionadas ao acaso e com reposição dessa população, e encontrar a distribuição da média amostral

$$\bar{X} = \frac{X_1 + X_2}{2},$$

sendo

X_1: valor selecionado na primeira extração; e

X_2: valor selecionado na segunda extração.

Amostra (X_1,X_2)	Probabilidade	Média Amostral
(1,1)	1/25	1
(1,3)	1/25	2
(1,5)	2/25	3
(1,7)	1/25	4
(3,1)	1/25	2
(3,3)	1/25	3
(3,5)	2/25	4
(3,7)	1/25	5
(5,1)	2/25	3
(5,3)	2/25	4
(5,5)	4/25	5
(5,7)	2/25	6
(7,1)	1/25	4
(7,3)	1/25	5
(7,5)	2/25	6
(7,7)	1/25	7
	1	

A distribuição de probabilidade de $\bar{X}$ para n = 2 é

$\bar{x}$	1	2	3	4	5	6	7
$P(\bar{X} = \bar{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25

Neste caso, $E(\bar{X}) = 4{,}2 = \mu_x$ **e**

$$\mathrm{Var}(\bar{X}) = 2{,}08 = \frac{\sigma_x^2}{2}.$$

Repetindo o mesmo procedimento, para amostras de tamanho *n* = 3, temos a seguinte distribuição de probabilidade de $\overline{X}$,

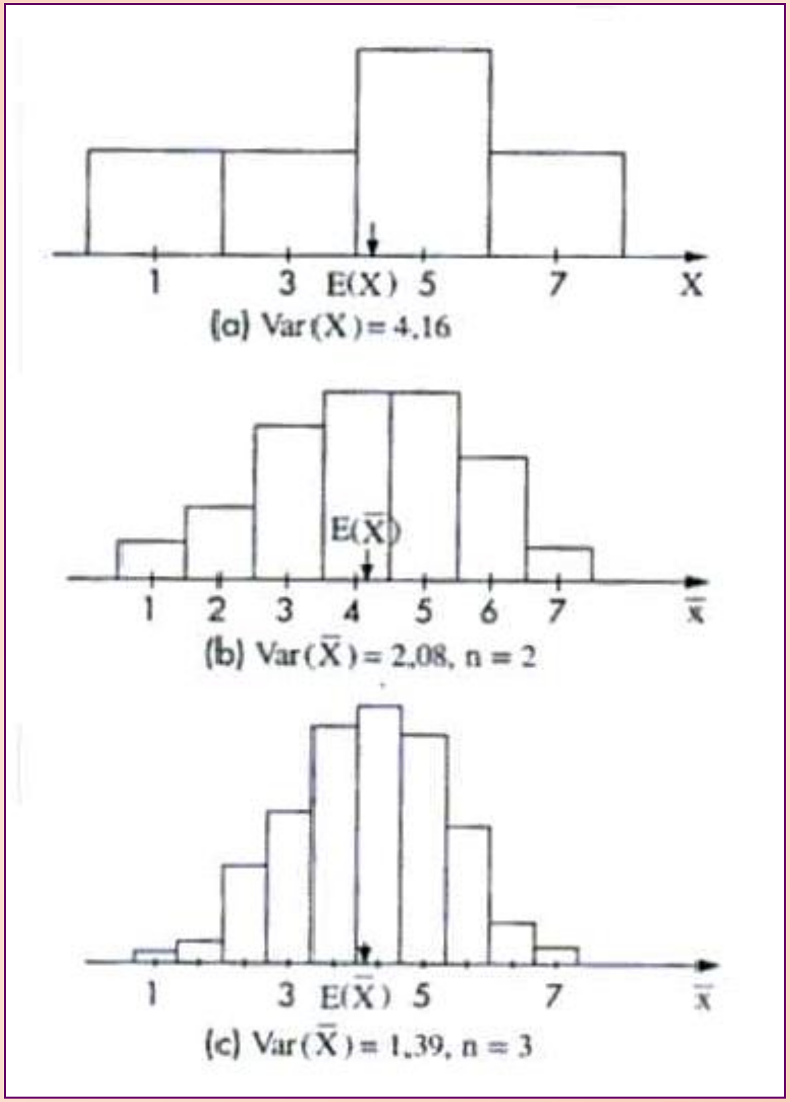
$\bar{x}$	$P(\overline{X} = \bar{x})$
1	1/125
5/3	3/125
7/3	9/125
3	16/125
11/3	24/125
13/3	27/125
5	23/125
17/3	15/125
19/3	6/125
7	1/125

Neste caso,

$$E(\overline{X}) = 4{,}2 = \mu_x \quad \text{e}$$

$$Var(\overline{X}) = 1{,}39 = \frac{\sigma_x^2}{3}.$$

Figura 1: Histogramas correspondentes às distribuições de X e de $\overline{X}$, para amostras de {1,3,5,5,7}.



Dos histogramas, observamos que

- **conforme *n* aumenta, os valores de $\overline{X}$ tendem a se concentrar cada vez mais em torno de**

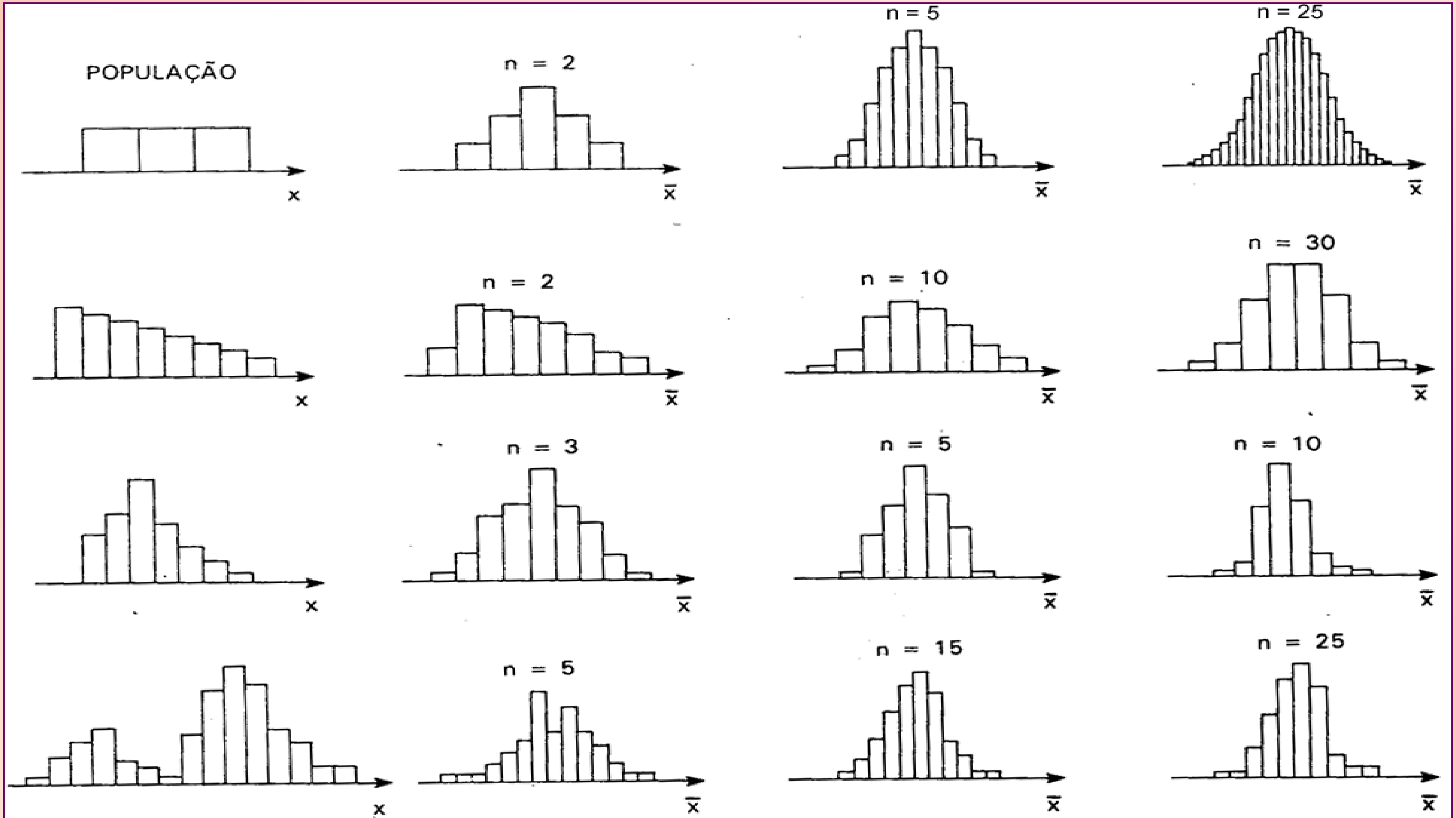
$$E(\overline{X}) = 4{,}2 = \mu_X \ ,$$

uma vez que a variância vai diminuindo;

- **os casos extremos passam a ter pequena probabilidade de ocorrência;**

- **para *n* suficientemente grande, a forma do histograma *aproxima-se de uma distribuição normal.***

Figura 2: Histogramas correspondentes às *distribuições de* $\overline{X}$ para amostras de algumas populações.



Fonte: BUSSAB & MORETTIN. ***Estatística Básica***. São Paulo, Atual, 4ª edição, 2002, pp. 273.

Esses gráficos sugerem que,

quando n aumenta, *independentemente da forma da distribuição* de X , a distribuição de probabilidade da média amostral $\bar{X}$ *aproxima-se* de uma distribuição normal.

Teorema do Limite Central

Seja X uma v. a. que tem média μ e variância σ^2. Para uma amostra $X_1, X_2, ..., X_n$, retirada ao acaso e com reposição de X, a distribuição de probabilidade da média amostral $\overline{X}$ *aproxima-se*, *para n grande*, de uma distribuição normal, com média μ e variância σ^2 / n , ou seja,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \text{ para n grande, aproximadamente.}$$

Comentários:

- **Se a distribuição de X é normal, então $\bar{X}$ tem distribuição normal <u>exata, para todo *n*.</u>**

- **O desvio padrão** $$\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$ **, que é**

o desvio padrão da média amostral, também é denominado <u>erro padrão.</u>

Corolário 10.1 Se $(X_1, \ldots, X_n)$ for uma amostra aleatória simples da população X, com média μ e variância σ^2 finita, e $\bar{X} = (X_1 + \ldots + X_n)/n$, então

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1). \tag{10.2}$$

Basta notar que se usou a transformação usual de reduzir a distribuição de $\bar{X}$ a uma normal padrão. Observe, também, que (10.2) pode ser escrita como

$$Z = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \sim N(0,1). \tag{10.3}$$

Chamemos de e a v.a. que mede a diferença entre a estatística $\bar{X}$ e o parâmetro μ, isto é, $e = \bar{X} - \mu$; e é chamado o *erro amostral da média*. Então, temos o

Corolário 10.2 A distribuição de e aproxima-se de uma distribuição normal com média 0 e variância σ^2/n, isto é,

$$\frac{\sqrt{n}\, e}{\sigma} \sim N(0,1). \tag{10.4}$$

Exemplo 2:

Uma maquina enchia pacotes de café cujos pesos seguiam uma distribuição N(500, 100). Colhendo-se uma amostra de n=100 pacotes e pesando-os. Se a máquina estiver regulada, qual a probabilidade de encontrarmos a média de 100 pacotes diferindo de 500g de menos de 2 gramas?

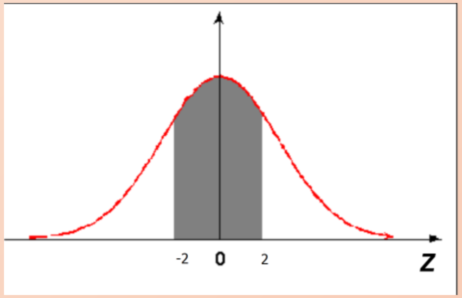
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \text{para n grande, aproximadamente.}$$

$\bar{X}$ terá uma distribuição normal com média 500 e variância 100/100 = 1.

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

$\mu \rightarrow 500$

$\sigma/\sqrt{n} \rightarrow 1$

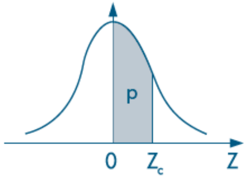


$$P(|\bar{X} - 500| < 2) = P(498 < \bar{X} < 502) = P(-2 < Z < 2) \approx 95\%.$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

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

Segunda decimal de z

Parte inteira e primeira decimal de 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