MAE-212: Introdução à Probabilidade e à Estatística II

Aula 03

Inferência Estatística

Distribuição amostral de uma Proporção

X = 1, se o indivíduo for portador da característica;
= 0, caso contrário;

É fácil ver que $\mu_x = E(X) = p$,

$\sigma_x^2 = Var(X) = p(1-p)$

Retirada de uma AAS(amostragem aleatória simples) dessa população, e indicando por Y_n o total de indivíduos portadores da característica na amostra, então

$Y_n \sim b(n, p)$

$$Y_n = X_1 + X_2 + \ldots + X_n,$$

$$\hat{p} = \frac{Y_n}{n}.$$

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Exemplo 10.12 Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e calculamos $\hat{p}$ = proporção de mulheres na amostra. Qual a probabilidade de que $\hat{p}$ difira de p em menos de 0,01? Temos que essa probabilidade é dada por

$$P(|\hat{p} - p| < 0{,}01) = P(-0{,}01 < \hat{p} - p < 0{,}01).$$

Mas, $\hat{p} - p \sim N\left(0, \dfrac{p(1-p)}{n}\right)$ e como $p = 0{,}3$, temos que

$$\mathrm{Var}(\hat{p}) = (0,\,3)(0,\,7)/10 \ = 0{,}021,$$

e, portanto, a probabilidade pedida é igual a

$$P\left(\frac{-0{,}01}{\sqrt{0{,}021}} < Z < \frac{0{,}01}{\sqrt{0{,}021}}\right) = P\left(-0{,}07 < Z < 0{,}07\right) = 0{,}056.$$

Outras Distribuições Amostrais

Exemplo 10.13. Na Tabela 10.6 apresentamos a distribuição de três outras estatísticas; a variância da amostra,

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

a mediana amostral, *md*, e o estimador

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

Tabela 10.6: Distribuição amostral de algumas estatísticas obtidas de amostra de tamanho $n = 3$, retiradas da população {1, 3, 5, 5, 7} ($\mu = 4{,}2$, $\sigma^2 = 4{,}16$ e $Md = 5$).

Tipo de amostra	Freqüência (prob. × 125)	Soma	Soma dos quadrados	Média $\bar{x}$	Mediana md	Variância	
						s^2	$\hat{\sigma}^2$
111	1	3	3	1,00	1	0	0
113	3	5	11	1,67	1	4/3	8/9
115	6	7	27	2,33	1	16/3	32/9
117	3	9	51	3,00	1	12	8
133	3	7	19	2,33	3	4/3	8/9
135	12	9	35	3,00	3	4	8/3
137	6	11	59	3,67	3	28/3	56/9
155	12	11	51	3,67	5	16/3	32/9
157	12	13	75	4,33	5	28/3	56/9
177	3	15	99	5,00	7	12	8
333	1	9	27	3,00	3	0	0
335	6	11	43	3,67	3	4/3	8/9
337	3	13	67	4,33	3	16/3	32/9
355	12	13	59	4,33	5	4/3	8/9
357	12	15	83	5,00	5	4	8/3
377	3	17	107	5,67	7	16/3	32/9
555	8	15	75	5,00	5	0	0
557	12	17	99	5,67	5	4/3	8/9
577	6	19	123	6,33	7	4/3	8/9
777	1	21	147	7,00	7	0	0
Total	125						

Tabela 10.7: Distribuição amostral da variância S^2, para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

s^2	0,00	1,33	4,00	5,33	9,33	12,00
$P(S^2 = s^2)$	11/125	42/125	24/125	24/125	18/125	6/125

$$E(S^2) = 4{,}16, \quad \mathrm{Var}(S^2) = 11{,}28.$$

Tabela 10.8: Distribuição amostral da mediana da amostra md para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

md	1	3	5	7
Prob.	13/125	31/125	68/125	13/125

$$E(md) = 4{,}30, \quad \mathrm{Var}(md) = 2{,}54.$$

Tabela 10.9: Distribuição amostral da variância $\hat{\sigma}^2$, para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

$\hat{\sigma}^2$	0,00	0,89	2,67	3,56	6,22	8,00
Prob.	11/125	42/125	24/125	24/125	18/125	6/125

$$E(\hat{\sigma}^2) = 2{,}77, \quad \mathrm{Var}(\hat{\sigma}^2) = 5{,}04.$$

Figura 10.6: Distribuição amostral de S^2 para amostras de tamanho $n = 3$ extraídas de $\{1, 3, 5, 5, 7\}$.

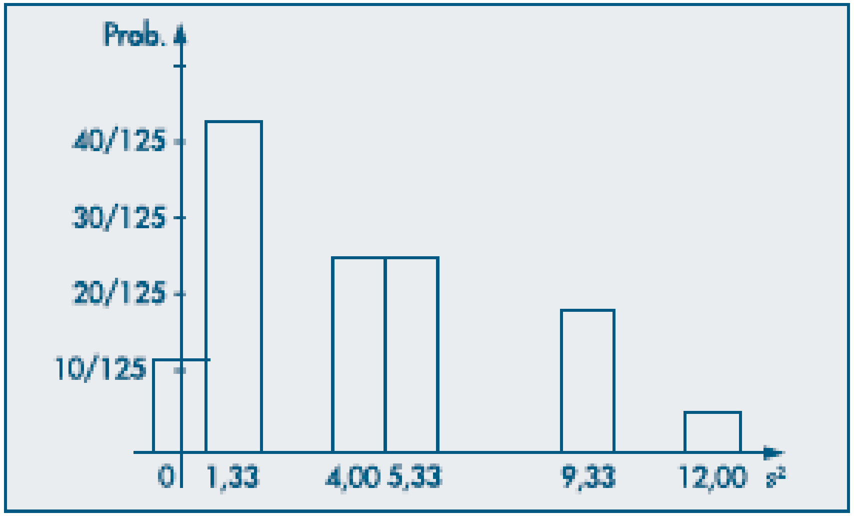


Figura 10.7: Distribuição amostral de md para amostras de tamanho $n = 3$ de $\{1, 3, 5, 5, 7\}$.

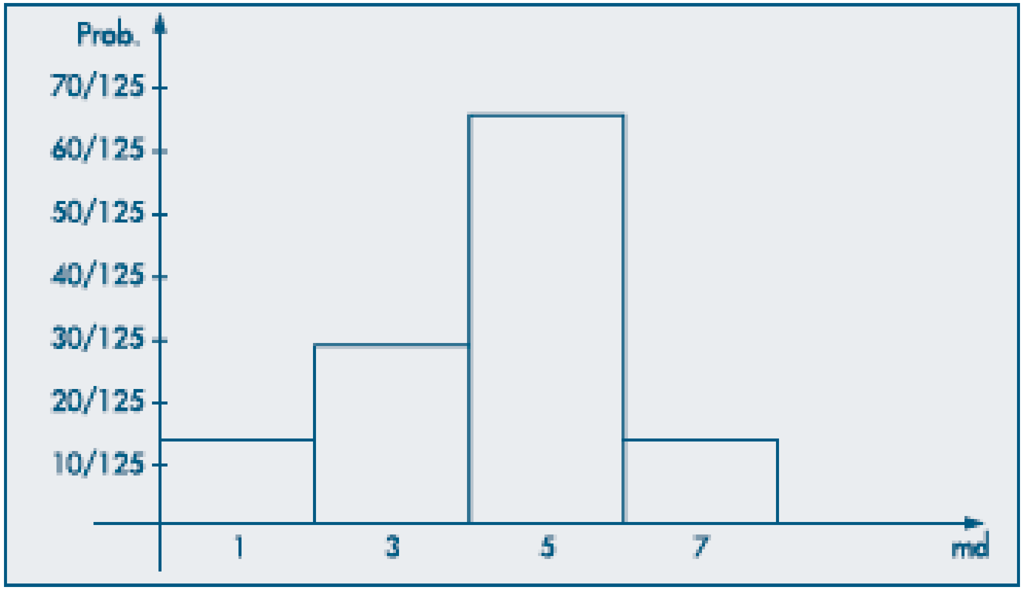
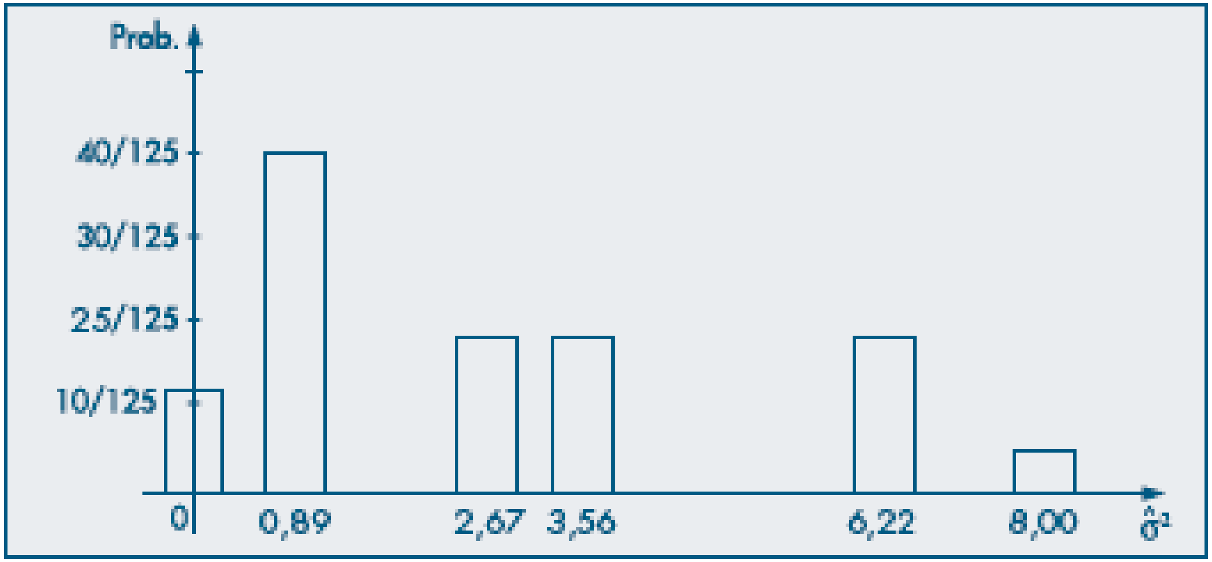


Figura 10.8: Distribuição amostral de $\hat{\sigma}^2$ para amostras de tamanho $n = 3$ extraídas de $\{1, 3, 5, 5, 7\}$.



Por exemplo, note que $E(S^2) = 4{,}16 = \sigma^2$, logo S^2 satisfaz uma propriedade análoga a $E(\overline{X}) = \mu$; dizemos que $\overline{X}$ e S^2 são estimadores *não-viesados* dos respectivos parâmetros μ e σ^2. Esta propriedade já não vale para md e $\hat{\sigma}^2$, pois $E(md) = 4{,}3$, enquanto $Md = 5{,}0$ e $E(\hat{\sigma}^2) = 2{,}77$ e não 4,16. Vemos que $\hat{\sigma}^2$ sub-estima a verdadeira variância.

Dimensionamento da amostra

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

$$\gamma = P\left(\left|\bar{X} - \mu\right| \leq \varepsilon\right) = P\left(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon\right)$$

$$= P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \cong P\left(\frac{-\varepsilon\sqrt{n}}{\sigma} \leq Z \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right),$$

sendo Z ~ N(0,1) .

Dimensionamento da amostra

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

o tamanho da amostra *n* é determinado por

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

conhecendo-se o desvio padrão σ de X, o erro ε da estimativa e o coeficiente de confiança γ do intervalo, sendo z tal que

$$\gamma = \mathbf{P}\,(\mathbf{-z} \le \mathbf{Z} \le \mathbf{z}) \quad \mathbf{e} \quad \mathbf{Z} \sim \mathbf{N(0,1)}.$$

Exemplo 4:

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

$X \sim N(\mu\,;\,250^2)$

n = ?? tal que $\varepsilon = 50$ reais,

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

Então,

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

$$= \left(\frac{1,96}{50}\right)^2 (250)^2$$

$$= 96,04$$

Aproximadamente 97 domicílios devem ser consultados.

Exemplo 5:

A quantidade de colesterol X no sangue das alunas de uma universidade segue uma distribuição de probabilidades com desvio padrão σ = 50 mg/dl e média μ desconhecida. Se desejamos estimar a quantidade média μ de colesterol com erro ε = 20 mg/dl e confiança de 90%, quantas alunas devem realizar o exame de sangue?

X: quantidade de colesterol no sangue das alunas da universidade

σ = 50 mg/dl

n = ?? tal que ε = 20 mg/dl

$\gamma = 0,90 \Rightarrow z = 1,65$

Supondo que o tamanho da amostra a ser selecionada é suficientemente grande, pelo Teorema do Limite Central temos:

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

$$= \left(\frac{1{,}65}{20}\right)^2 (50)^2$$

$$= 17{,}02$$

Assim, aproximadamente 18 alunas devem realizar o exame de sangue.

No caso de proporção, usando a aproximação normal para proporção amostral, temos

$$n = \left(\frac{z}{\varepsilon} \right)^2 p(1-p)$$

Como não conhecemos p, podemos usar o fato de que p(1-p)<=1/4, para todos p.

Dimensionamento da amostra

Da relação $\varepsilon = z\sqrt{\dfrac{p(1-p)}{n}}$,

segue que o **tamanho amostral *n***, dados γ e a margem de erro ε, tem a forma

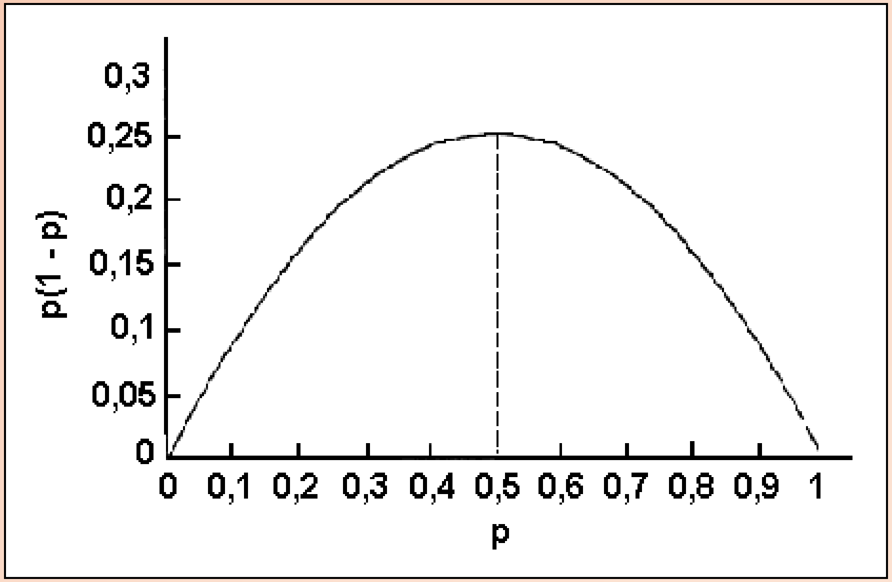
$$n = \left(\frac{z}{\varepsilon}\right)^2 p(1-p),$$

onde *z* é tal que $\gamma = P(-z \le Z \le z)$ e $Z \sim N(0,1)$.

Entretanto, nesta expressão, *n* depende de *p(1-p)*, que é desconhecido.

- ***Como calcular o valor de n?***

Gráfico da função *p(1-p)*, para $0 \le p \le 1$.



Pela figura observamos que:

- **a função *p(1-p)* é uma parábola simétrica em torno de *p* = 0,5;**
- **o máximo de *p(1-p)* é 0,25, alcançado quando *p* = 0,5.**

Assim, na prática, substituímos *p(1-p)* por seu valor máximo, obtendo

$$n = \left(\frac{z}{\varepsilon}\right)^2 0{,}25\,,$$

que pode fornecer um valor de *n* maior do que o necessário.

Exemplo 6:

No exemplo da USP (Exemplo 1) suponha que nenhuma amostra foi coletada. Quantos estudantes precisamos consultar de modo que a estimativa pontual esteja, no máximo, a 0,02 da proporção verdadeira *p*, com uma probabilidade de 0,95?

Dados do problema:

ε = 0,02 (erro da estimativa);

$P(\varepsilon) = \gamma = 0{,}95 \Rightarrow z = 1{,}96$.

$$n = \left(\frac{1{,}96}{0{,}02}\right)^2 p(1-p) \leq \left(\frac{1{,}96}{0{,}02}\right)^2 0{,}25 = 2401 \text{ estudantes.}$$

Pergunta: ***É possível reduzir o tamanho da amostra quando temos alguma informação a respeito de p?***

Por exemplo, sabemos que:

- ***p*** **não é superior a 0,30, ou**
- ***p*** **é pelo menos 0,80, ou**
- ***p*** **está entre 0,30 e 0,60.**

Resposta: ***Depende do tipo de informação sobre p.***

Em alguns casos, podemos substituir a informação *p(1-p)*, que aparece na expressão de *n*, por um valor menor que 0,25.

Redução do tamanho da amostra

Vimos que, se nada sabemos sobre o valor de *p*, no cálculo de *n*, substituímos *p(1-p)* por seu valor máximo, e calculamos

$$n = \left(\frac{z}{\varepsilon}\right)^2 0{,}25\,.$$

Se temos a informação de que *p é no máximo 0,30* ($p \leq 0{,}30$), então o valor máximo de *p(1-p)* será dado por 0,3x0,7 = 0,21.

Logo, reduzimos o valor de *n* para

$$n = \left(\frac{z}{\varepsilon}\right)^2 0{,}21.$$

Agora, se *p é pelo menos 0,80* ($p \geq 0{,}80$), então o máximo de *p(1-p)* é 0,8 x 0,2 = 0,16 e temos

$$n = \left(\frac{z}{\varepsilon}\right)^2 0{,}16\,.$$

Mas, se *$0{,}30 \leq p \leq 0{,}60$* , o máximo de *p(1-p)* é 0,5x0,5 = 0,25 e, neste caso, não há redução, ou seja,

$$n = \left(\frac{z}{\varepsilon}\right)^2 0{,}25.$$

Exemplo 7:

No Exemplo 6, suponha que temos a informação de que no máximo 30% dos alunos da USP foram ao teatro no último mês. Portanto, temos que $p \leq 0,30$ e, como vimos, o máximo de *p(1-p)* neste caso é 0,21.

Assim, precisamos amostrar

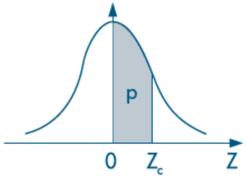
$$n = \left(\frac{z}{\varepsilon}\right)^2 0,21 = \left(\frac{1,96}{0,02}\right)^2 0,21 = 2017 \text{ estudantes},$$

conseguindo uma redução de 2401 – 2017 = 384 estudantes.

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 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	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 $P(Z \leq z) = A(z)$

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