



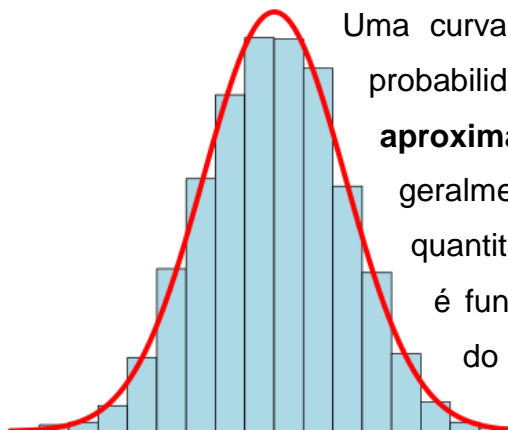
**Cruzeiro do Sul
Virtual**
Educação a Distância

AS DISTRIBUIÇÕES NORMAIS

Prof. Ismar Frango



Curvas de densidade



Fonte: Wikimedia Commons (Licença CC)

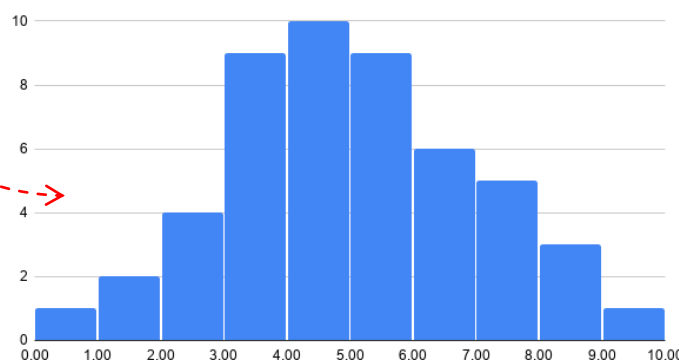
Uma curva de densidade é, basicamente, um gráfico de probabilidades. Ela é utilizada como uma forma de **aproximar** os dados contidos em um histograma, geralmente vindos de um *dataset* de variáveis quantitativas. Sua aplicação em aplicações de *Big Data* é fundamental, já que ela traz uma boa aproximação do comportamento geral dos dados, sem ter que analisar os dados individualmente.

5,57	3,65
5,91	8,46
5,46	5,02
4,90	3,50
3,84	8,20
3,80	7,10
4,40	6,54
1,77	2,75
4,90	3,52
6,68	7,58
4,15	6,55
3,49	3,20
0,25	4,60
2,50	5,57
7,34	5,70
6,79	3,82
9,50	6,85
1,49	4,81
2,82	3,62
5,34	7,90
6,95	4,83
7,16	5,22
4,78	5,28
2,90	8,44
4,84	4,24

Por exemplo, veja o *dataset* ao lado, com 50 notas (entre 0 e 10) de alunos em uma disciplina. O histograma que o representa segue abaixo (dez intervalos de tamanho 1):

Intervalo	Frequência
0-1	1
1-2	2
2-3	4
3-4	9
4-5	10
5-6	9
6-7	6
7-8	5
8-9	3
9-10	1

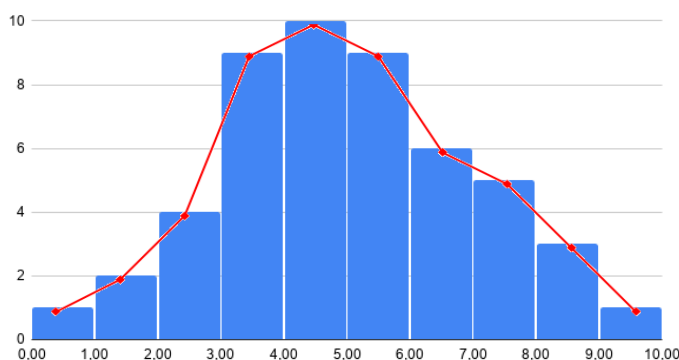
Histograma



Fonte: autor

Esse histograma pode vir representado junto de uma linha, geralmente chamada de **polígono de frequência**.

Histograma



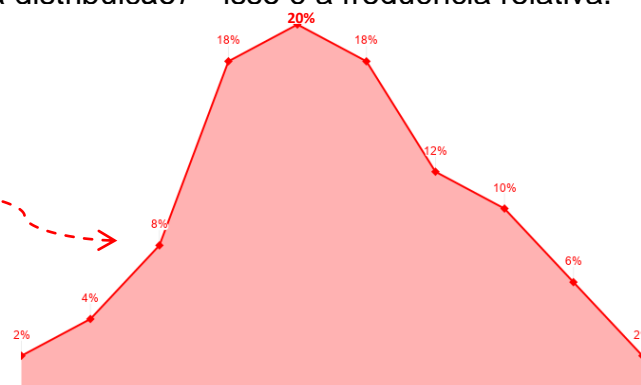
Fonte: autor

Tecnicamente, não poderia ser chamada de "polígono", pois não é uma forma fechada – é uma "polilinha". Mas "polígono" é o termo comumente usado.

É mais interessante trabalharmos com as frequências **relativas** do que as **absolutas**, para compreendermos melhor um histograma. Por exemplo, ao invés de dizermos que existem 10 notas no intervalo entre 4 e 5, podemos dizer que **20% das notas** estão nesse intervalo (já que 10 equivale a 20% de 50, que é a quantidade de notas da distribuição) – isso é a frequência relativa.

Intervalo	Frequência absoluta
0-1	1
1-2	2
2-3	4
3-4	9
4-5	10
5-6	9
6-7	6
7-8	5
8-9	3
9-10	1

Intervalo	Frequência relativa
0-1	2%
1-2	4%
2-3	8%
3-4	18%
4-5	20%
5-6	18%
6-7	12%
7-8	10%
8-9	6%
9-10	2%



Veja que a área compreendida abaixo desse “polígono de frequência” equivale a 100% das observações.

Imagine agora que você refaz o histograma com intervalos de tamanho 0,1 ao invés de 1 – você teria 100 intervalos. Ou com intervalos de tamanho 0,01 – o que daria 1000 intervalos. Ou, pensando em um tamanho de intervalo infinitamente pequeno, nosso polígono iria se transformar em uma curva suave:



A essa curva, damos o nome de **curva de densidade**.

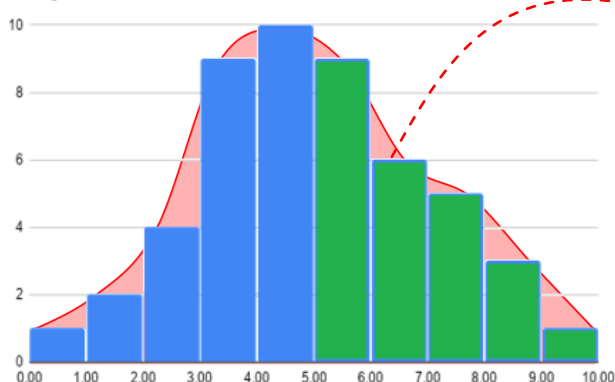
Aplicações da curva de densidade

Por que a curva de intensidade é útil?

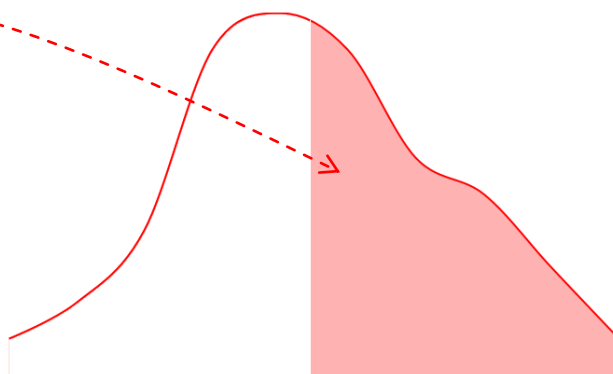
Quando analisamos grandes quantidades de dados, nem sempre vale a pena gastar poder computacional para processar todos os dados, todas as vezes que queremos extrair alguma informação do *dataset*. Muitas vezes, valores aproximados por algum modelo matemático (no nosso caso, a curva de densidade) são suficientes para os objetivos da análise que precisamos fazer dos dados.

Por exemplo, se quiséssemos saber quantos alunos tiraram nota maior ou igual a 5,0, ao invés de realizar a contagem diretamente no histograma, podemos analisar a área abaixo da curva a partir do ponto equivalente ao valor 5,0.

Histograma



Essa área abaixo da curva é uma boa aproximação da soma dos valores em verde no histograma.



Fonte: autor

Obviamente, tudo isto só faz sentido se o cálculo da área abaixo da curva for uma tarefa mais fácil do que o processamento dos dados no *dataset*. Neste exemplo, a curva de densidade não segue nenhuma forma conhecida e ela aparenta ser mais difícil de se obter o resultado esperado do que a simples contagem das frequências absolutas no histograma.

Porém, em aplicações do mundo real, veremos que muitas distribuições se comportam de uma maneira bastante regular, o que facilitará bastante a análise dos dados.

Medidas de centro em curvas de densidade

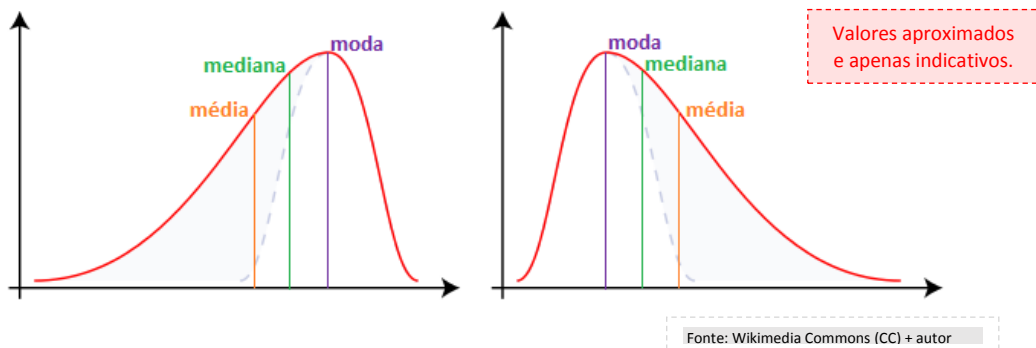
As medidas de centro em uma curva de densidade podem ser identificadas de maneira aproximada, considerando que:

A **Moda** é o ponto que corresponde ao valor mais alto da curva.

A **Mediana** é o ponto que divide a área em duas metades iguais:

A **Média** é o centro de massa da curva, isto é, o ponto onde se poderia equilibrar a curva se ela fosse feita de material sólido.

Como vimos, as curvas de densidade não têm que ser simétricas. Quando são, dizemos que a distribuição é **normal** e as três medidas de centro são iguais. Quando a curva é assimétrica à esquerda, a média fica à esquerda da mediana, que fica à esquerda da moda. Quando a curva é assimétrica à direita, a média fica à direita da mediana, que fica à direita da moda.



Distribuições Normais

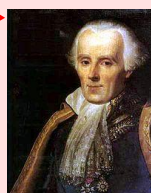
Uma Distribuição Normal, também chamadas de Distribuição Gaussiana ou de **Gauss**, ou ainda de Distribuição de **Laplace**-Gauss, é caracterizada pela **Curva Normal**. É o tipo de curva mais utilizado para representar fenômenos naturais.



Matemático, astrônomo e físico alemão.

Saiba mais sobre ele em:

https://pt.wikipedia.org/wiki/Carl_Friedrich_Gauss



Matemático, astrônomo e físico francês.

Saiba mais sobre ele em:

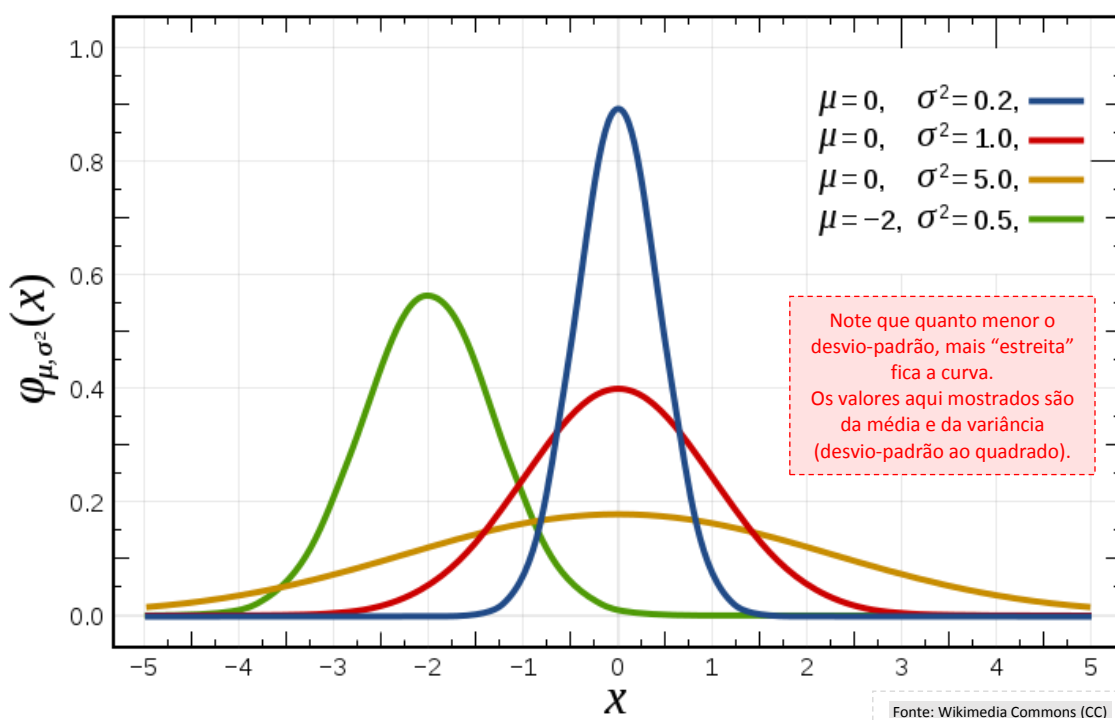
https://pt.wikipedia.org/wiki/Pierre-Simon_Laplace

Todas as curvas Normais são simétricas e têm um único pico. Dizemos que essas curvas são *bell-shaped*, o que quer dizer “em forma de sino”.

Toda curva Normal é caracterizada por sua média μ (que é igual à mediana e à moda) e seu desvio-padrão σ . Representamos essa curva por **$N(\mu;\sigma)$** .

O valor do desvio-padrão σ controla a **variabilidade** da curva normal:

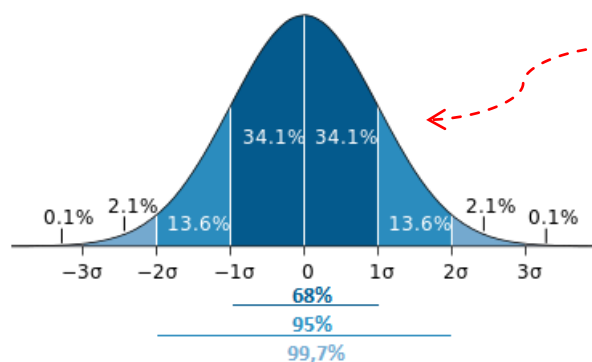
Como usamos **vírgula** para separar os decimais no Brasil, usamos um **ponto-e-vírgula** para separar μ e σ :
 $N(\mu;\sigma)$.
Poderia ser também:
 $N(\mu,\sigma)$.
Nesse caso, as casas decimais teriam que ser separadas por **ponto**.



As curvas normais têm uma propriedade interessante: para **qualquer** distribuição normal com média μ e desvio-padrão σ :

- aproximadamente **68%** das observações estão a uma distância de até σ de média μ 68,27%
- aproximadamente **95%** das observações estão a uma distância de até σ de média μ 95,45%
- aproximadamente **99,7%** das observações estão a uma distância de até σ de média μ 99,73%

Essa propriedade é conhecida como “**Regra 68 – 95 – 99,7**”.

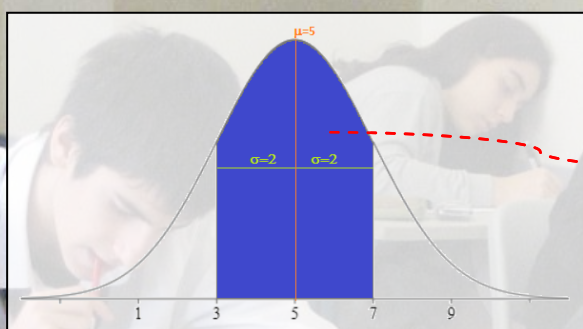


Nesse gráfico, você pode observar como funciona a Regra 68-95-99,7

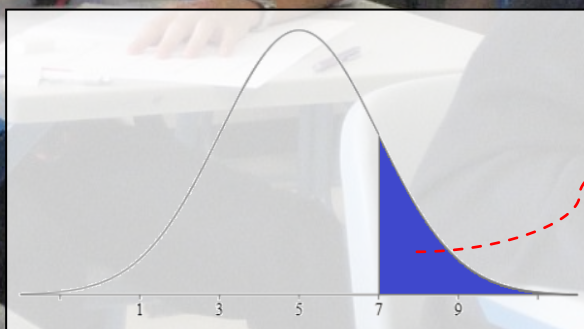
Vamos para alguns exemplos:

Considere uma prova aplicada a **1800** alunos de uma escola. Após corrigidas as provas, verificou-se que as notas, variando de 0 a 10, seguem uma distribuição Normal **$N(5;2)$** .

Significa que a média μ das notas é 5, com desvio-padrão σ igual a 2



Podemos afirmar que aproximadamente **68%** da distribuição (1224 alunos) tirou entre 3 e 7 ($\mu - \sigma = 5 - 2 = 3$ e $\mu + \sigma = 5 + 2 = 7$)



Segundo o mesmo raciocínio, sabemos que apenas 16% dos alunos (288) tirou mais que 7. Se sabemos que 68% tirou entre 3 e 7, então restam 32% que tiraram abaixo de 3 (16%) ou acima de 7 (16%).

Um outro exemplo:

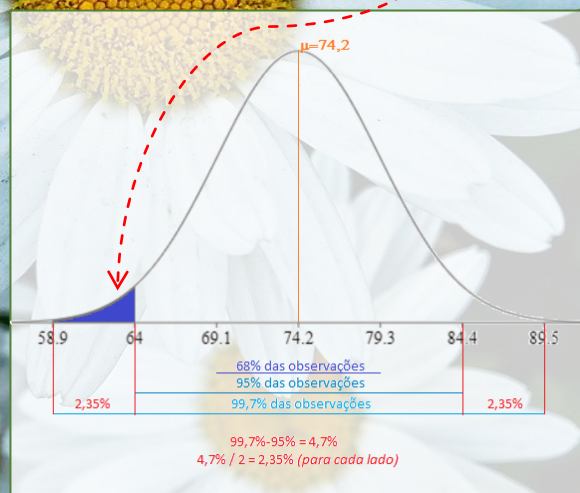
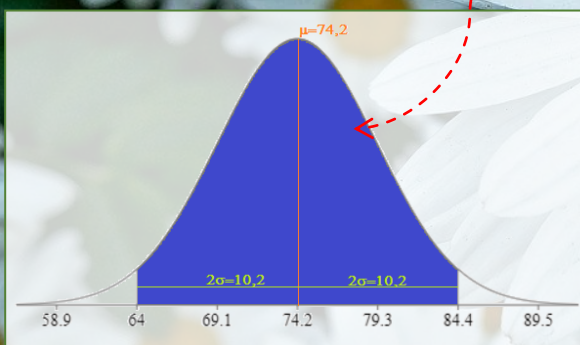
Uma empresa de flores ornamentais, especializada em margaridas, fez uma colheita de **24500** flores. Sabe-se que o tamanho do caule das margaridas (em cm) segue uma distribuição Normal **$N(74,2;5,1)$** .

Significa que a média μ dos caules é **74,2cm**, com desvio-padrão σ igual a 5,1cm

Como usamos vírgula para separar os decimais no Brasil, usamos um ponto-e-vírgula entre μ e σ :
 $N(74,2;5,1)$
só para não confundir.
Poderia ser também: $N(74.2,5.1)$

Podemos afirmar que 95% do total (23275 margaridas) têm o caule entre 64cm e 84,4cm:
 $\mu - 2\sigma = 74,2 - 10,2 = 64\text{cm}$
 $\mu + 2\sigma = 74,2 + 10,2 = 84,4\text{cm}$

Também podemos afirmar que apenas 2,35% das observações (aproximadamente 57 margaridas) têm caules entre 58,9 ($\mu - 3\sigma$) e 64cm ($\mu - 2\sigma$)



Fonte: autor + Wikimedia Commons–
Licença CC

Distribuições padronizadas

Como indica a Regra 68–95–99,7, todas as distribuições Normais possuem propriedades em comum. Na verdade, todas as distribuições Normais se comportam da mesma maneira, considerando seu desvio-padrão σ em relação à média μ .

Porém, as aplicações reais exigem que saibamos muito mais do que os valores relativos a 68%, 95% e 99,7% da distribuição. Como sabemos que todas as distribuições normais são iguais em termos de comportamento, é conveniente trabalharmos com a **padronização** dessas distribuições.

Um valor padronizado é chamado de um **escore z** (ou *z score*, em inglês). Assim, se x é uma observação de uma distribuição com média μ e desvio-padrão σ , o **valor padronizado z** para x significa a distância, em desvios-padrões, de x em relação a μ . Resumidamente:

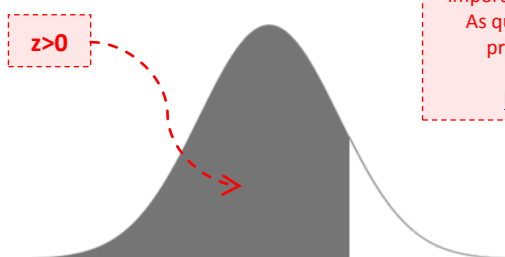
$$z = \frac{x - \mu}{\sigma}$$

Por exemplo: se tomarmos a situação das notas dos alunos descrita anteriormente, que seguia uma distribuição $N(5;2)$, se dissermos que uma nota tem o valor padronizado $z=1$, significa que ela está à distância de 1σ de μ , ou seja, é a nota 7 ($\mu+1\sigma = 5+2 = 7$). Da mesma maneira, se dissermos que uma outra observação corresponde a uma nota com escore $z=-0,5$, significa que ela está à distância de meio desvio-padrão da média, à esquerda (por causa do sinal negativo), Logo, é a nota 4 ($\mu+0,5\sigma = 5-1 = 4$).

A padronização de uma distribuição Normal leva a uma nova distribuição, chamada **distribuição Normal padrão**, que é a distribuição Normal

$N(0;1)$, com $\mu=0$ e $\sigma=1$.

Para trabalhar com o conceito de distribuição Normal padronizada, a maioria dos softwares calcula o que chamamos de **proporções acumuladas**. A ideia de “acumulada” é “tudo o que veio antes”. Além disso, é importante conhecer e se familiarizar com o uso de **tabelas de escore z**, principalmente quando não temos acesso a algum software (por exemplo, quando fazemos alguma prova ou estamos em um trabalho de campo, desconectados da Internet). Essas duas tabelas trazem valores aproximados de **proporções acumuladas** para cada valor de z , para z negativo (à esquerda da média) ou positivo (à direita da média).



Apesar de hoje existirem vários apps e sites para cálculo do escore z , é importante sabermos usar as tabelas. As que estão aqui disponíveis nas próximas páginas podem ser encontradas em <https://www.ztable.net/>

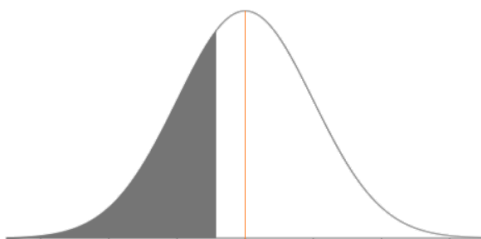


Tabela de escore z para $z < 0$

Como usar:

4. Calcule o valor de z com duas casas decimais. Por exemplo, $z = -0,65$ (significa que a observação está a 0,65s à esquerda da média).
5. Use os dois primeiros algarismos (0,6) para localizar a linha, olhando na primeira coluna
6. Use o último algarismo (5) para localizar a coluna correspondente (0,05).

Pronto! Você achou o valor de probabilidade acumulada (0,25785), que corresponde à probabilidade acumulada de valores abaixo de z , ou seja, neste exemplo, que 25,785% dos valores são menores que o valor da observação.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

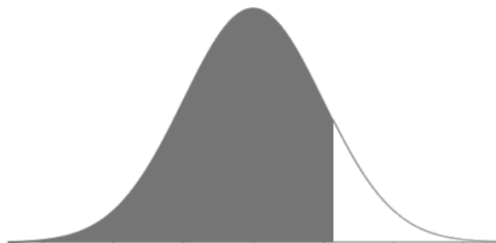


Tabela de escore z para $z > 0$

Como usar:

1. Calcule o valor de z com duas casas decimais. Por exemplo, $z=+2,09$ (significa que a observação está a 2,09s à direita da média).
2. Use os dois primeiros algarismos (2,0) para localizar a linha, olhando na primeira coluna
3. Use o último algarismo (9) para localizar a coluna correspondente (0,09).

Pronto! Você achou o valor de probabilidade acumulada (0,98169), que corresponde à probabilidade acumulada de valores abaixo de z , ou seja, neste exemplo, que 98,169% dos valores são menores que o valor da observação.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997
+4	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998

Vamos detalhar com alguns exemplos:

O peso dos bois da raça Zebu, sub-raça Brahman segue uma distribuição Normal $N(60;6)$, em arrobas (@) – uma arroba, no Brasil, equivale a 15kg.

ESTAREMOS OFERTANDO PRENHEZ
SEXADA DE FÊMEA DA DOADORA QERJ-28

Ou seja, o peso médio desses bois é de 60@, com desvio-padrão de 6@.

O boi da foto é o Queixada, que pesa 69@. O seu dono o está levando para uma exposição e quer saber se o Queixada estará entre os 10% dos bois mais pesados.

Como saber isso?

Vamos calcular o valor de z para $x=75@$:

$$z = \frac{x - \mu}{\sigma} = \frac{69 - 60}{6} = 1,50$$

Vamos olhar a tabela de escore z ($z>0$) e procurar, na primeira coluna, onde está a linha $z=+1,5$; em seguida, a linha equivalente a 0,00. Achamos 0,93319!

Ou seja, o dono do Queixada pode ficar tranquilo, pois aproximadamente 93,3% dos bois pesam menos que Queixada – ou seja, Queixada fica tranquilamente entre os 7% bois mais pesados.

z	0
+0	.50000
+0.1	.53983
+0.2	.57926
+0.3	.61791
+0.4	.65542
+0.5	.69146
+0.6	.72575
+0.7	.75804
+0.8	.78814
+0.9	.81594
+1	.84134
+1.1	.86433
+1.2	.88493
+1.3	.90320
+1.4	.91924
+1.5	.93319

Fonte: autor + Wikimedia Commons / Pixabay - Licenças CC

O raciocínio inverso pode ser aplicado à seguinte situação: qual seria o menor peso para que um boi desta raça esteja entre os 10% mais pesados?

Isso significa, obviamente, que ele tem que pesar mais que os 90% menos pesados!

Para “despadronizar” (achar x), basta multiplicar z pelo desvio-padrão e somar com a média, ou seja,

$$x = z\sigma + \mu = 1,29 \cdot 6 + 60 = 67,74$$

Então, para estar entre os 10% dos bois mais pesados, o peso mínimo é de 67,74@

Fazemos uma “busca inversa” na tabela e vemos que a primeira vez que temos um valor maior que 90% (0,9) é 0,90147, que aparece para $z=+1,29$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147



Vamos ver um outro exemplo?

Fonte: Flickr - Licença CC-BY-SA
www.flickr.com/photos/pfala/2822132026/

O verão de 2019 foi um dos **mais quentes** em toda a Europa. Paris, por exemplo, chegou à temperatura de 42,6 graus Celsius. Jornais disseram ter sido **o 25/7 mais quente de toda a História**.
Será mesmo verdade?

42,6 C à Paris, chaleur historique également à Lille, Strasbourg et une cinquantaine d'autres villes, nuit la plus chaude jamais mesurée en France... (Le Monde, 25/7/2019)

42,6 C em Paris, calor histórico também em Lille, Estrasburgo e cinquenta outras cidades, a noite mais quente já medida na França ...

Considerando que a temperatura média em 25 de julho para Paris, medida ao longo de vários anos, segue uma distribuição Normal (em graus Celsius):

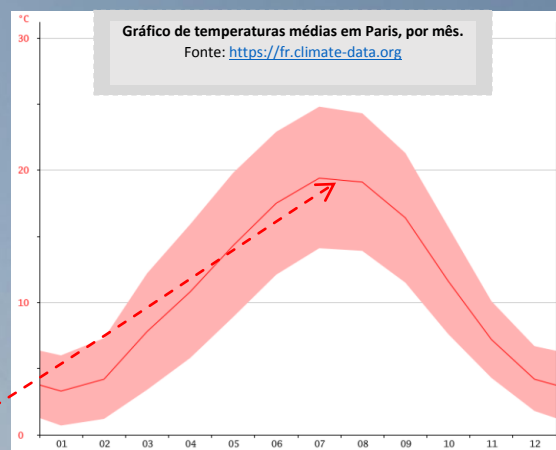
$N(19,5;4,95)$

o que podemos dizer desse pico de temperatura em 25/7/2019?

Ou seja, o temperatura média neste dia é de 19,5 graus Celsius, com desvio-padrão de 4,95 graus.

Vamos “despadronizar” esse valor de $z=+4,09$:

$x = z\sigma + \mu$
 $= 4,09 \cdot 4,95 + 19,5 = 39,75$
Ou seja, em **99,98%** das vezes, a temperatura **não passou de 39,75** graus!



Vamos calcular o valor de z para $x=42,6$ para a distribuição $N(19,5;4,95)$:

$$z = \frac{x - \mu}{\sigma} = \frac{42,6 - 19,5}{4,95} = 4,67$$

Mas, quando olhamos na tabela de escore z , **não achamos** o valor $z=4,67$ (o maior valor é para $z=+4,09$, que é de **0,9998!**)

E agora? 🤔

Apenas **0,02%** das temperaturas históricas **passaram** de 39,75 graus. Ou seja, é **muito provável** que 42,6 graus tenha realmente sido uma das maiores temperaturas já registradas no dia 25 de julho ao longo dos anos, senão a maior!

0,02% equivale a 1 dia em cada 5.000!



Vejamos mais um exemplo, similar ao anterior:

A média das temperaturas mínimas diárias na cidade de São Paulo atingiu o valor de 9,6 graus Celsius em julho de 2019. Muitos acham que este foi o **mês de julho mais frio** dos últimos 50 anos.

Considerando que, nestes 50 anos, a temperatura em julho na cidade de São Paulo (em graus Celsius) seguiu uma distribuição Normal de:

$N(12,1;3,4)$

Será que este foi mesmo o inverno mais frio?

Vamos calcular o valor de z para $x=9,6$ para a distribuição $N(12,1;3,4)$:

$$z = \frac{x - \mu}{\sigma} = \frac{9,1 - 12,1}{3,4} = -0,88$$

Na tabela de escore z , achamos o valor de **0,18943**

Ou seja, o mês de julho de 2019 está entre os **18,94%** julhos mais frios dos últimos 50 anos. Provavelmente, foi um dos 10 julhos mais frios dos últimos 50 anos, mas **não** podemos afirmar que foi o **mais** frio.

Fonte: Flickr - Licença CC-BY-SA
<https://www.flickr.com/photos/lubasi/6945191504>



Para saber mais, leia o capítulo 3 do e-book:

MOORE, David S.; NOTZ, William I.; FLINGER, Michael A. A
Estatística Básica e sua Prática. 7 ed. Rio de Janeiro: LTC,
2017 – Capítulo 3