

Distribuição Normal

1. Introdução

“O mundo é normal!” Acredite se quiser! Muitos dos fenômenos aleatórios que encontramos na prática apresentam uma distribuição muito peculiar, chamada Normal.

Um modelo probabilístico é aquele que nos diz, ou melhor, nos traduz na forma de números o comportamento de uma variável. Por exemplo, já fizemos algumas análises das variáveis altura e peso. Considerando o caso de grandes amostras aleatórias ou mesmo da população, tanto o peso como a altura tem um comportamento muito parecido. Vejamos, por exemplo, a variável altura, considerando-se o sexo masculino e sendo a população os habitantes do Brasil. Essa variável é dita Normal, conforme discutiremos mais adiante.

Essa distribuição de freqüência denominada curva normal, considerada um modelo teórico ou ideal que resulta muito mais de uma equação matemática do que de um real delineamento de pesquisa com coleta de dados.

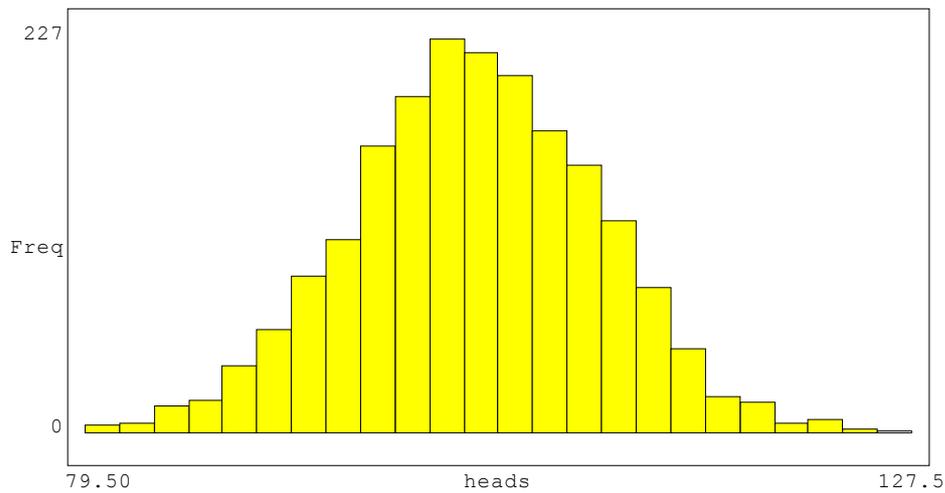
A curva normal é um tipo de curva simétrica, suave, cuja forma lembra um sino. Ela é unimodal, sendo seu ponto de freqüência máxima situado no meio da distribuição, em que a média, a mediana e a moda coincidem.

Para exemplificarmos, suponhamos 2000 lançamentos de 200 moedas honestas. Utilizando um software de simulação, obtivemos os seguintes resultados:

[80 ... 81]	5
[82 ... 83]	6
[84 ... 85]	16
[86 ... 87]	19
[88 ... 89]	39
[90 ... 91]	60
[92 ... 93]	91
[94 ... 95]	111
[96 ... 97]	165
[98 ... 99]	194
[100 ... 101]	227
[102 ... 103]	220
[104 ... 105]	206
[106 ... 107]	174
[108 ... 109]	155
[110 ... 111]	123
[112 ... 113]	84
[114 ... 115]	49
[116 ... 117]	21
[118 ... 119]	18
[120 ... 121]	6
[122 ... 123]	8
[124 ... 125]	2
[126 ... 127]	1

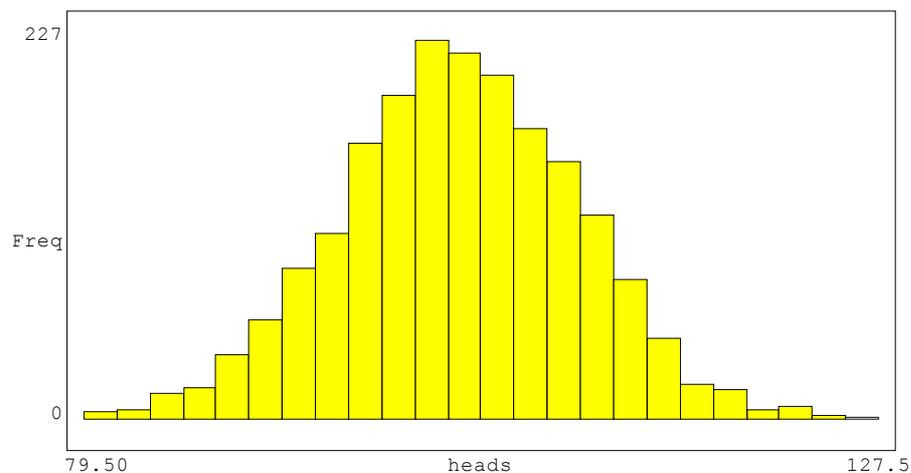
sample mean = 102.132
sample st dev = 7.238

A partir desses dados, construímos o histograma:

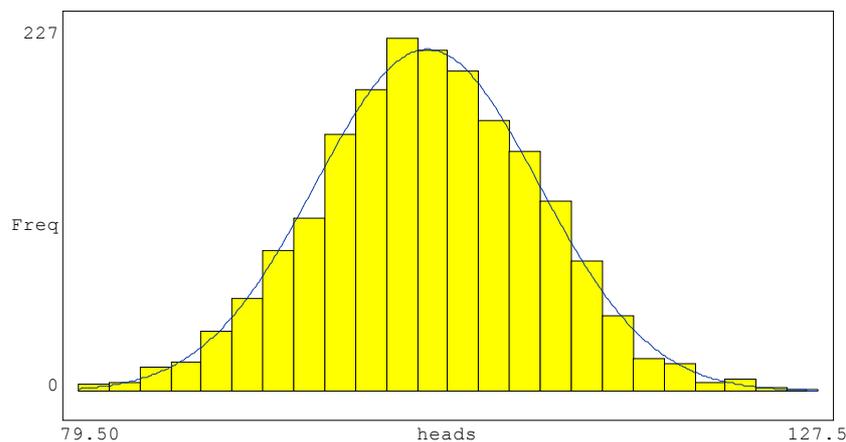


Observando tal histograma e a tabela anterior, notamos que a média está entre as duas classes com maior frequências. Além disso, considerando-se a média, percebemos que o histograma parece ser simétrico ao redor dela. A frequência é menor quanto mais nos afastamos da média, tanto para mais quanto para menos, sendo que as menores frequências ocorrem nas pontas do gráfico.

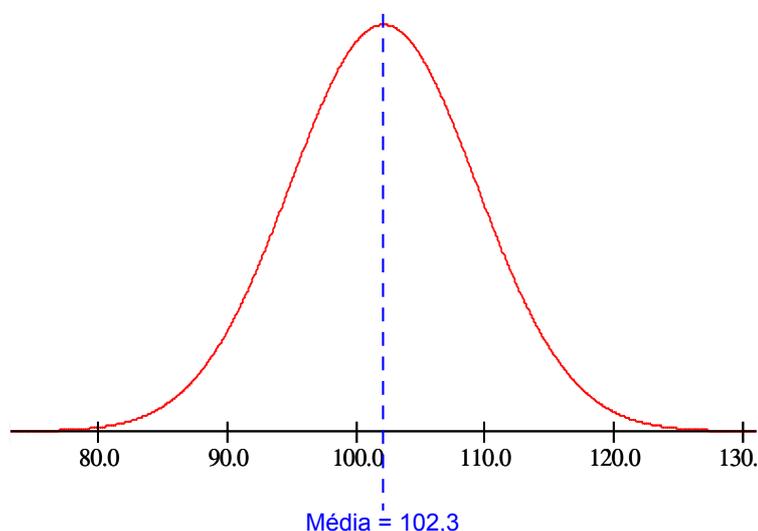
Se aumentarmos a quantidade de classes, por exemplo, elevando para 40, obtemos o seguinte histograma:



Note que quanto mais classes usamos, mais fácil fica identificarmos um formato criado pelas colunas do histograma. Esse formato faz lembrar um [sino](#), conforme a figura seguinte:



“Limpendo” o gráfico acima, temos:



Essa curva que chamamos de sino recebe um nome especial: **Curva Normal**.

A Curva Normal é a representante do modelo normal e é obtida a partir da função densidade que nada mais é do que uma função que origina o gráfico anterior.

Assim, se X é uma variável aleatória com distribuição Normal com média μ e variância σ^2 (notação: $X \sim N(\mu, \sigma^2)$) então a sua densidade é dada por

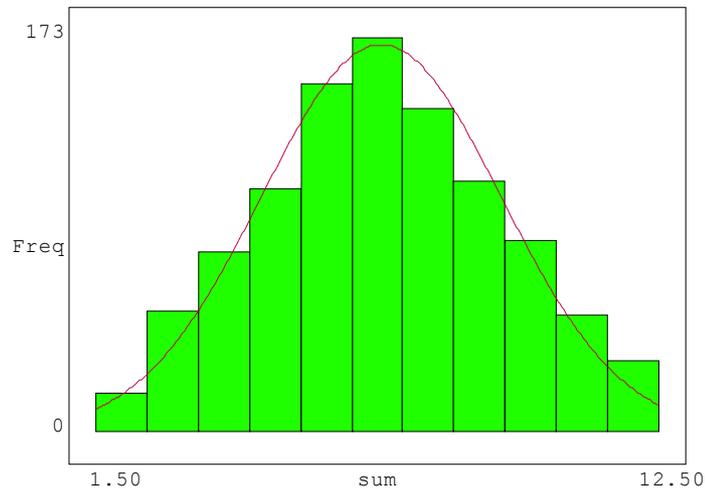
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A Normal apresenta as seguintes propriedades:

- é simétrica ao redor da média;
- a área sobre a curva é igual a 1;
- para valores muito grandes de x , tendendo a infinito (ou muito pequenos, tendendo a menos infinito), a curva tende a zero.

Note que conforme o caso, poderemos ter curvas com formatos diferentes, ou seja, mais para a direita, mais para a esquerda, mais ou menos achatadas... enfim, cada caso poderá gerar uma curva diferente. Vejamos mais um caso.

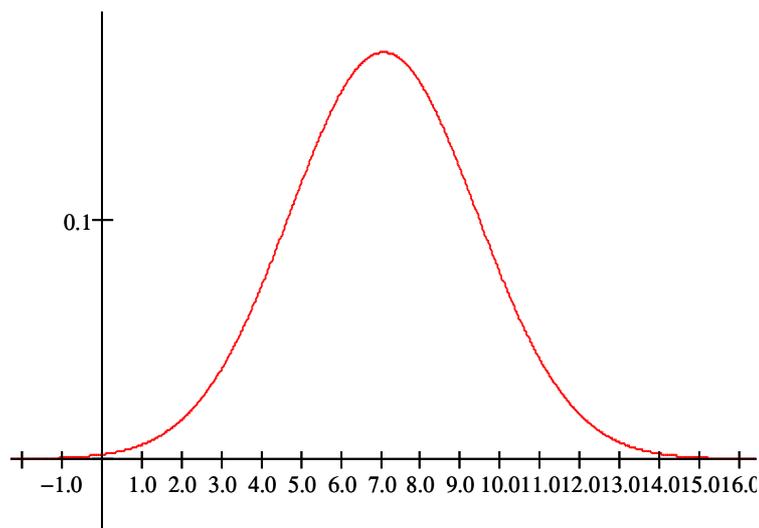
Consideremos o lançamento de dois dados não viciados. Estamos interessados em analisar a soma dos resultados obtidos em cada jogada. Realizando uma simulação e construindo o histograma dos resultados, obtemos:



No caso simulado, obtivemos:

sample mean = 7.07200
sample st dev = 2.34282

Ou seja, $\mu = 7,07$ e $s = 2,34$, onde μ indica a média e s o desvio padrão da amostra. Isolando a curva da normal temos:



Assim, conforme havíamos dito, existem diferentes curvas, que variam conforme os valores da média e do desvio-padrão.

Lembramos que a área abaixo desse gráfico vale 1. Ou seja, a **área corresponde a uma probabilidade**.

2. Área sob a Curva Normal

É aquela região do plano compreendida entre a curva e o eixo das abscissas, que corresponde em qualquer distribuição normal a 100% dos dados considerados.

A natureza simétrica da curva normal vai levar a concluir que qualquer distância medida em “sigmas” (desvio padrão), acima ou abaixo da média, contém a mesma porção da área sob a curva. Temos então:

- 34,13% da área total situam-se entre a média e 1 DP abaixo ou acima da média;
- 47,72% da área total situam-se entre a média e 2 DP abaixo ou acima da média;
- 49,87% da área total situam-se entre a média e 3 DP abaixo ou acima da média;

Exemplo 1: consideremos uma população de uma cidade A e de uma outra cidade B. Suponhamos que todas as pessoas tenham informado as respectivas alturas (em centímetros). E deseja-se fazer uma comparação entre tais populações. As principais estatísticas obtidas por essa pesquisa foram:

	População A	População B
Média (μ)	174	178
Variância (σ^2)	64	1

Um pesquisador deseja sortear aleatoriamente pessoas para fazer um teste sobre DNA e crescimento e, para isso, gostaria de coletar (aleatoriamente!) pessoas com mais de 1,80m. Em qual das duas populações será mais fácil achar pessoas com tais características?

Para isso, podemos dizer que estamos trabalhando com duas variáveis:

X_A : altura de 1 pessoa selecionada ao acaso na população A

X_B : altura de 1 pessoa selecionada ao acaso na população B

Nosso objetivo é calcular as seguintes probabilidades:

$P(X_A > 180)$ e $P(X_B > 180)$.

Mas antes, calculemos o valor do desvio-padrão (σ) de cada uma das populações. Lembrando que o desvio-padrão nada mais é do que a raiz quadrada da variância. Logo:

$$\sigma_A = \sqrt{64} = 8$$

$$\sigma_B = \sqrt{1} = 1$$

Utilizando um software é possível calcularmos tais probabilidades. Esse software calcula, na verdade, a área abaixo da curva:

$P(X_A > 180) = 0,2266$ ou **22,66%**

$P(X_B > 180) = 0,0228$ ou **2,28%**

Portanto, o pesquisador terá mais facilidade de achar pessoas com mais de 1,80m na população A. Aleatoriamente falando, tal probabilidade é de 22,66%.

Esse resultado pode parecer estranho, visto que a população B possui uma média de alturas superior ao da população A. Porém, tal fato é explicado através da variabilidade dos resultados, ou melhor, pelo desvio-padrão.

Exemplo 2: Consideremos, ainda, o exemplo anterior. Só que, agora, temos interesse em obter pessoas que tenham entre 1,75m e 1,80m. Qual população possui um número maior de habitantes nessa faixa de altura?

Aqui, queremos calcular as seguintes probabilidades:

$$P(175 < X_A < 180) \text{ e}$$

$$P(175 < X_B < 180) .$$

Novamente, usando um software, obtemos:

$$P(175 < X_A < 180) = 0,2236 \text{ e}$$

$$P(175 < X_B < 180) = 0,9759.$$

Portanto, é evidente que a população B, mais uma vez, representa a melhor escolha, pois a probabilidade de escolhermos uma pessoa nessa faixa de altura na população B é quase 100%, ou seja, quase todos habitantes possuem as alturas procuradas.



Meu computador “deu pau”! Como calcular as probabilidades da normal?

A NORMAL PADRÃO

Em muitos livros de Estatística podemos encontrar uma tabela da Normal Padrão. É uma tabela que nos fornece valores das áreas (= probabilidades) de acordo com o valor de x , exatamente como fizemos nos exemplos anteriores com o auxílio do Winstats.

A maioria dos softwares de Estatística possuem comandos que permitem calcularmos o valor das áreas sobre curvas como a Normal, dentre eles, o SAS, o Excel, o S-PLUS, e outros. Porém, se não temos acesso a um computador, podemos usar a **Tabela da Normal Padrão**. Mas, como dissemos anteriormente, existem diversas curvas da normal, que variam segundo a média e variância. Isso significa que teríamos que ter inúmeras tabelas da normal... o que não parece muito viável.

Dessa forma, criou-se uma maneira mais simples de se obter as áreas desejadas. Criou-se uma curva denominada **Normal Padrão**, que corresponde a uma **distribuição normal com média zero e desvio-padrão um**. Geralmente a variável aleatória associada à distribuição normal padrão é chamada de **Z**. Em notação:

$$\underline{Z \sim N(0,1). \text{ ou seja. } \mu_Z = 0 \text{ e } \sigma_Z = 1}$$

A grande vantagem de usarmos tal distribuição é o fato de trabalharmos apenas com uma distribuição e, portanto, com uma única tabela. Tudo é mais fácil!

Porém, como fazer para obtermos tal variável Z (**padronizada**) a partir de uma variável aleatória qualquer X tal que $X \sim N(\mu, \sigma^2)$?

Basta **padronizarmos** ou **normalizarmos** a variável X através da fórmula:

$$Z = \frac{X - \mu}{\sigma}$$

onde:

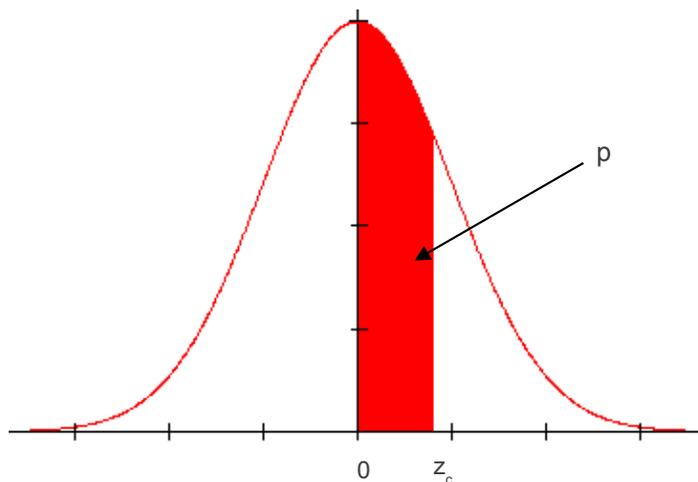
μ = média de X

σ = desvio-padrão de X

Usando a tabela da Normal Padrão

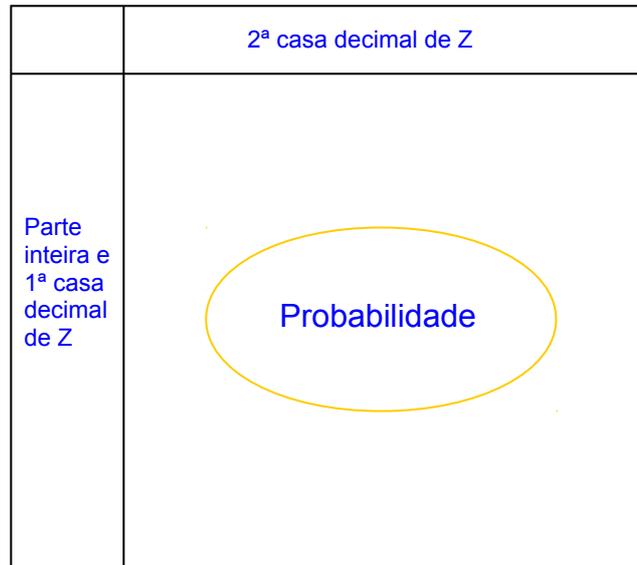
Existem algumas variações de apresentação da tabela. No nosso caso, utilizaremos uma tabela tal que:

$P(0 \leq Z \leq z_c) = p$, ou seja, a probabilidade fornecida pela tabela (p) corresponde ao intervalo que vai de 0 até um certo número z_c no eixo x. Esquemáticamente, a tabela da normal nos fornece a probabilidade correspondente à área a seguir:



Leitura da Tabela

Veja, esquematicamente, como deve ser a leitura da tabela:



Exemplo 3 – Padronização

Voltemos ao caso do Exemplo 3 onde tínhamos $\mu = 174$ e $\sigma = 8$. Queríamos calcular $P(X > 180)$. Vamos normalizar a variável X (ou seja, transformá-la em Z) e utilizar a tabela para obter a probabilidade desejada.

$$P(X > 180) = P\left(\frac{X - \mu}{\sigma} > \frac{180 - 174}{8}\right) = P(Z > 0,75)$$

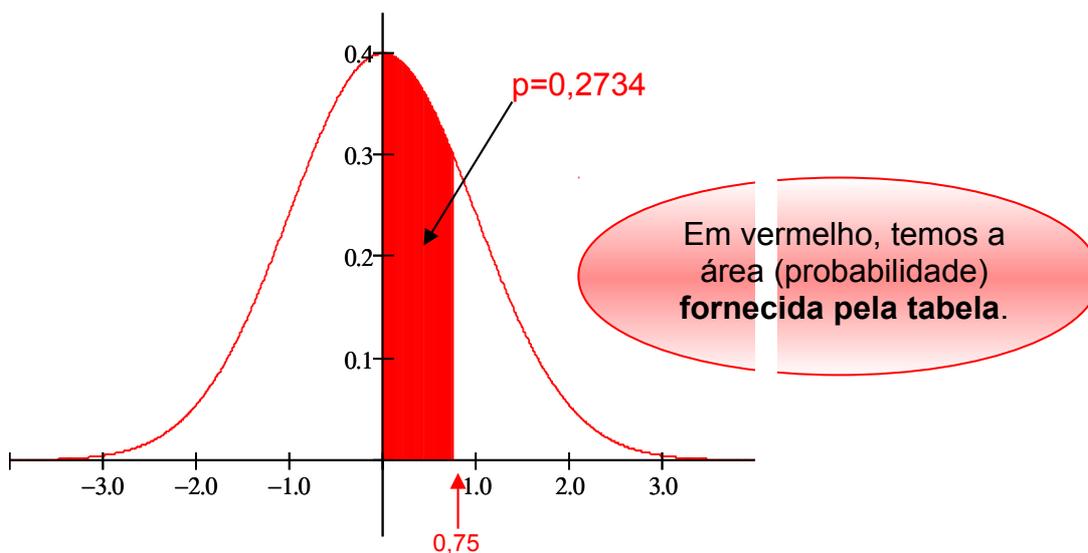
Para obtermos a probabilidade desejada, devemos lembrar que a nossa tabela nos fornece a probabilidade de 0 até um certo valor.

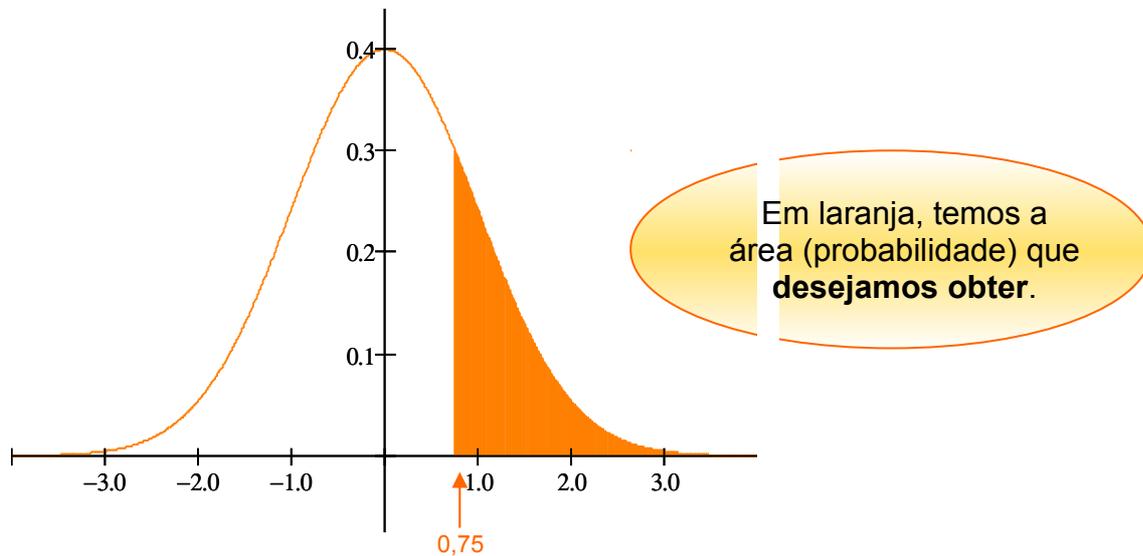
Consultado a tabela da Normal

Dado $Z = 0,75$, vejamos como obter a probabilidade a partir da tabela da Normal Padrão. Na coluna mais à esquerda, em verde, tomamos a parte inteira e a primeira decimal de Z , no caso, $0,7$. Na linha superior, em azul, observamos o valor da segunda casa decimal, no caso 5 (lembre-se que o número é $0,75$). A célula correspondente á linha do número $0,7$ e da coluna de 5 é o valor da probabilidade. No caso, $p=0,2734$.

Parte inteira e primeira decimal de z	Segunda casa decimal de z									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621

Então:



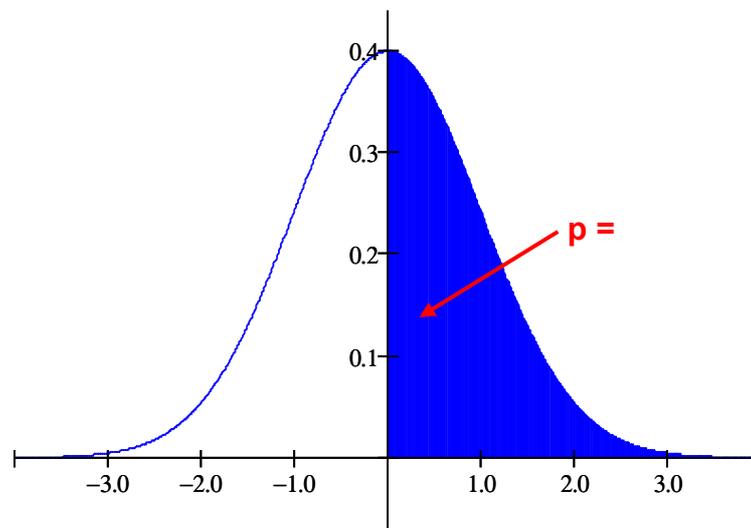


Como fazer para calcular o valor desejado a partir do valor da tabela?

Bem, é preciso relembrar as propriedades da curva da normal, em especial duas delas:

- 1) a área abaixo da curva é igual a 1;
- 2) a curva é simétrica em torno da média (no caso da normal padrão, em torno do zero).

Logo, concluímos, que a área correspondente à metade da curva é igual a 0,5:



Então, podemos observar que:

$$P(0 \leq Z \leq 0,75) + P(Z > 0,75) = 0,5 \Rightarrow$$

$$P(Z > 0,75) = 0,5 - P(0 \leq Z \leq 0,75) \Rightarrow$$

$$P(Z > 0,75) = 0,5 - 0,2734 \Rightarrow$$

$$P(Z > 0,75) = \mathbf{0,2266}$$

Logo, a probabilidade procurada é de 0,2266 que corresponde ao mesmo valor obtido utilizando-se o *software* (exemplo 1).

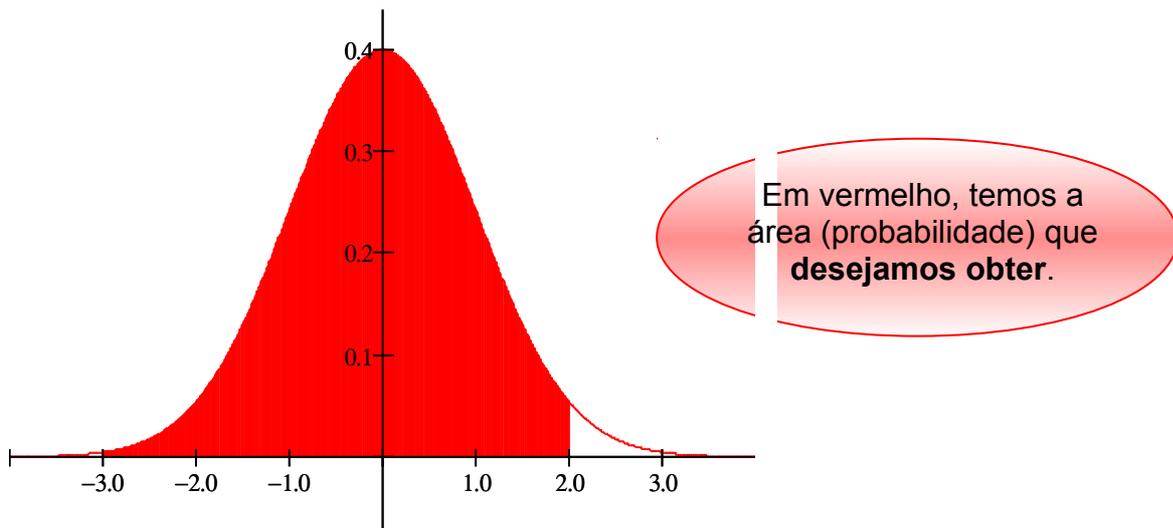
Exemplo 4 – Padronização

Consideremos a mesma situação do Exemplo 4. Ou seja, $\mu = 178$, $\sigma = 1$ e queremos calcular a probabilidade $P(175 < X < 180)$.

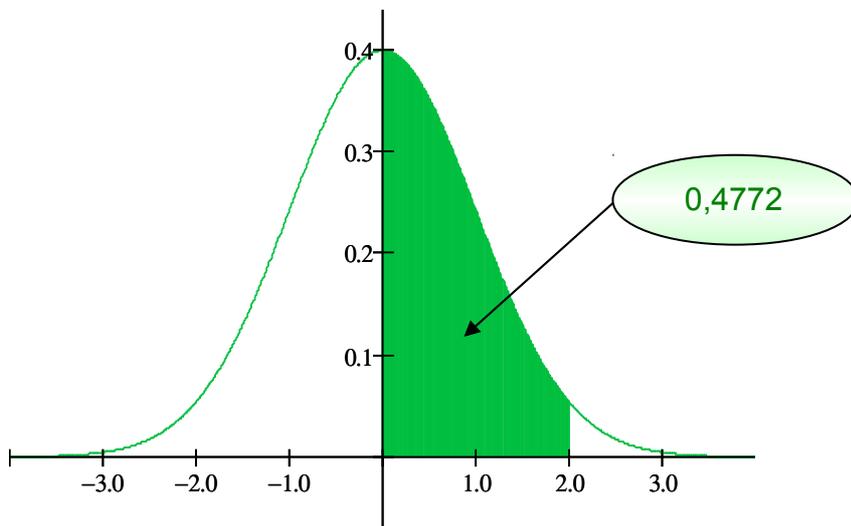
Normalizando a variável X temos:

$$P(175 < X < 180) = P\left(\frac{175-178}{1} < \frac{X-\mu}{\sigma} < \frac{180-178}{1}\right) = P(-3 < Z < 2)$$

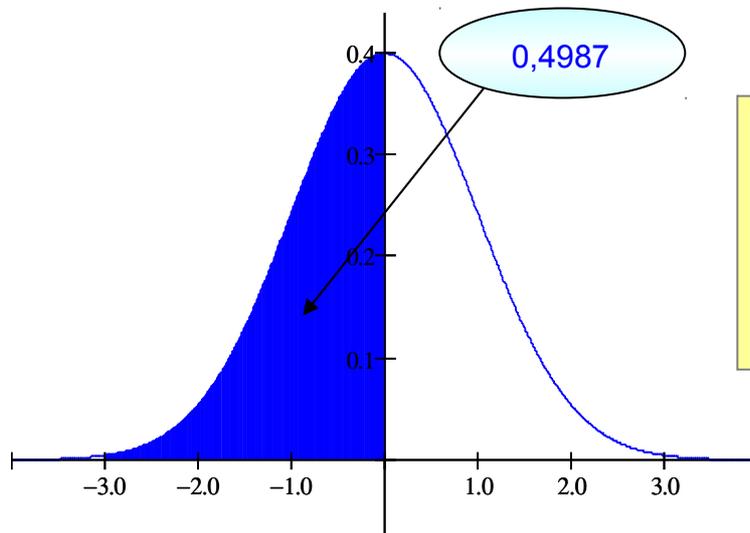
Graficamente:



Porém, a nossa tabela só nos fornece os valores:



E, também, fornece o valor de:



Lembre-se que, como a curva é **simétrica**, a área correspondente ao intervalo de **-3 à 0** é **igual** a área de **0 à 3!**

Logo, a área procurada corresponde à soma das áreas parciais:

$$P(-3 < Z < 2) = 0,4772 + 0,4987 = 0,9759,$$

que é igual à probabilidade obtida através do *software* (*exemplo 2*).

Assim, mostramos que a tabela pode ser muito útil, principalmente quando não temos um computador ou software adequado por perto.

Exemplo 5 – Teste de aptidão

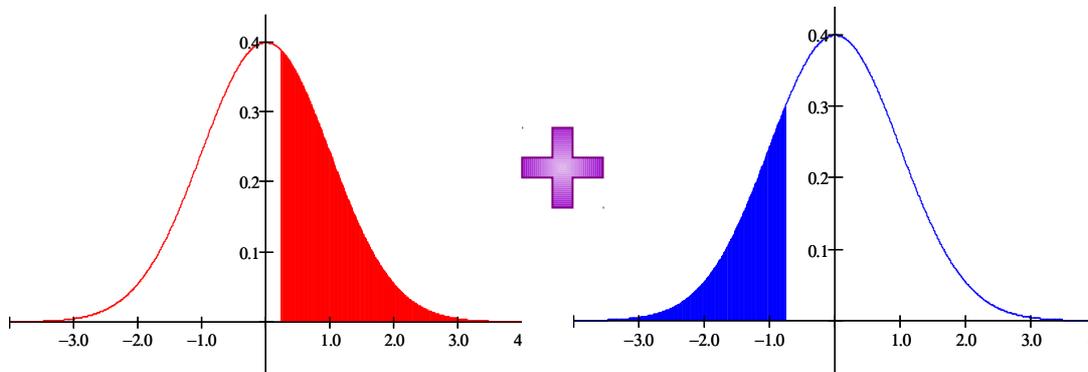
Em um certo teste de aptidão para contratação de determinada empresa, os candidatos devem realizar uma seqüência de tarefas no menor tempo possível. Suponhamos que o tempo necessário para completar esse teste tenha uma distribuição Normal com média 45 minutos e desvio-padrão de 20 minutos. Suponhamos que, numa primeira etapa, esse teste foi aplicado com uma amostra de 50 candidatos. Qual a probabilidade de encontrarmos algum candidato que tenha um tempo superior a 50 minutos (candidato muito lento) ou inferior a 30 minutos (que seria impossível completar o teste)? Qual o número aproximado de candidatos com tal perfil?

Inicialmente, seja X uma variável que indique o tempo de execução das tarefas tal que $X \sim N(45, 20^2)$. Desejamos calcular:

$$P(X > 50) + P(X < 35) =$$

$$= P\left(\frac{X - \mu}{\sigma} > \frac{50 - 45}{20}\right) + P\left(\frac{X - \mu}{\sigma} < \frac{30 - 45}{20}\right) =$$

$$= P(Z > 0,25) + P(Z < -0,75) =$$



$$= (0,5 - 0,0987) + (0,5 - 0,2734) =$$

$$= 0,6279 \text{ ou } 62,79\%$$

Como $0,6279 \cdot 50 = 31,39$, temos que o número de pessoas aproximado que contenham tais características é de 32 pessoas. Então, nesse teste a empresa já exclui 32 candidatos, restando apenas 18 para continuarem no processo de seleção.

Observação: uma outra notação para os Escores Padronizados é:

$$z = \frac{X - \bar{X}}{DP}$$

X = escore

\bar{x} = média

DP = desvio padrão

z = escore padronizado

4. Resumo das Propriedades da Distribuição Normal

1ª) A variável aleatória X pode assumir todo e qualquer valor real.

2ª) A representação gráfica da distribuição normal é uma curva em forma de sino, simétrica em torno da média, que recebe o nome de **curva normal ou de Gauss**.

3ª) A área total limitada pela curva e pelo eixo das abscissas é igual a 1, já que essa área corresponde à probabilidade de a variável aleatória X assumir qualquer valor real.

4ª) A curva normal é assintótica em relação ao eixo das abscissas, isto é, aproxima-se indefinidamente do eixo das abscissas sem, contudo, alcançá-lo.

5ª) Como a curva é simétrica em torno da média, a probabilidade de ocorrer valor maior que a média é igual à probabilidade de ocorrer valor menor do que a média, isto é, ambas as probabilidades são iguais a 0,5 ou 50%. Cada metade da curva representa 50% de probabilidade.

Quando temos em mãos uma variável aleatória com distribuição normal, nosso principal interesse é obter a probabilidade de essa variável aleatória assumir um valor em um determinado intervalo. Vejamos como proceder, por meio de um exemplo concreto.

Exemplo 6: Qual a probabilidade de escolher-se de forma aleatória, numa só tentativa, uma pessoa que tenha renda anual entre US\$ 5.000 e US\$ 7.000, morador de uma cidade. Sendo a renda média desta cidade US\$ 5.000 e o desvio padrão de US\$ 1.500?

Exemplo 7: Os pesos de 600 estudantes são normalmente distribuídos com média 65,3 kg e desvio padrão 5,5 kg. Determine o número de estudantes que pesam:

- a) entre 60 kg e 70 kg
- b) mais que 63,2 kg
- c) menos que 68 kg

Exemplo 8: Calcule as seguintes probabilidades:

- a) $P(0 < z < 1,44)$
- b) $P(-1,48 < z < 2,05)$
- c) $P(0,72 < z < 1,89)$
- d) $P(z > -2,03)$

Exemplo 9: X é $N(20; 49)$. Calcular $P(X < 30)$.

Exercícios

- 1) O tempo necessário em uma oficina para o conserto da transmissão de um tipo de carro segue distribuição normal com média de 45 minutos e desvio-padrão de 8 minutos. O mecânico comunicou a um cliente que o carro estará pronto em 50 minutos. Qual probabilidade do mecânico atrasar o seu serviço?
- 2) A duração de certo componente eletrônico pode ser considerada normalmente distribuída com média de 850 dias e desvio padrão de 45 dias. Calcular a probabilidade de um componente durar:
 - a) Entre 700 e 1000 dias
 - b) Mais de 800 dias
 - c) Menos de 750 dias
- 3) Determinado atacadista efetua suas vendas por telefone. Após alguns meses, verificou-se que os pedidos se distribuem normalmente com média de 3.000 pedidos e desvio-padrão de 180 pedidos. Qual a probabilidade de que um mês selecionado ao acaso esta empresa venda menos de 2700 pedidos
- 4) O conteúdo líquido das garrafas de 300 ml de um refrigerante é normalmente distribuído com média de 300 ml e desvio padrão de 2 ml. Determine a probabilidade de uma garrafa selecionada ao acaso apresentar conteúdo líquido:

- a) inferior a 306 ml
- b) Superior a 305 ml
- c) entre 302 e 304 ml

5) O lucro mensal obtido com ações de determinada empresa tem distribuição normal com média de 12 mil reais e desvio padrão de 5 mil reais. Qual a probabilidade de que em determinado mês o lucro desta empresa seja:

- a) superior a 18 mil reais
- b) inferior a 8 mil reais
- c) entre 10 e 15 mil reais

6) Durante o mês de dezembro aumenta a procura por concessão de crédito para pessoa física. De acordo com dados históricos é possível verificar que a procura segue uma distribuição aproximadamente normal com média de 12,8 milhões e desvio padrão de 15 milhões. Se as instituições de crédito reservarem 25 milhões para concessão de crédito, qual a probabilidade de faltar dinheiro para emprestar?

7) Suponha que a renda média anual de uma grande comunidade tenha distribuição normal com média de 15 mil reais e com um desvio-padrão de 3 mil reais. Qual a probabilidade de que um indivíduo aleatoriamente selecionado deste grupo apresente uma média salarial anual superior a 18 mil reais?

8) O escore de um estudante no vestibular é uma variável com distribuição normal com média de 550 pontos e desvio padrão de 30 pontos. Se a admissão em certa faculdade exige um escore mínimo de 575 pontos, qual é a probabilidade de um aluno ser admitido nesta faculdade?

9) As vendas de determinado produto têm apresentado distribuição normal com média de 600 unidades e desvio padrão de 40 unidades. Se a empresa decide fabricar 700 unidades naquele mesmo mês, qual é a probabilidade dela não poder atender a todos os pedidos desse mês por estar com o estoque esgotado?

10) O volume de enchimento de uma máquina automática usada para encher latas de bebidas gasosas é distribuído normalmente com uma média de 12,4 onças e um desvio padrão de 0,1 onça. Qual a probabilidade do volume de enchimento ser:

- a) inferior a 12 onças
- b) entre 12,1 e 12,6 onças
- c) superior a 12,3 onças

11) O tempo de reação de um motorista para o estímulo visual é normalmente distribuído com uma média de 0,4 segundos com um desvio-padrão de 0,05 segundos. Qual a probabilidade de que uma reação de um motorista requeira:

- a) mais de 0,5 segundos
- b) entre 0,4 e 0,5 segundos

12) O período de falta de trabalho em um mês por causa de doenças dos empregados é normalmente distribuído com uma média de 100 horas e desvio padrão de 20 horas. Qual a probabilidade desse período no próximo mês estar:

- a) entre 50 e 80 horas
- b) superior a 90 horas
- c) inferior a 60 horas

13) X é $N(10; 100)$. Calcular $P(12 \leq X \leq 20)$.

14) X é $N(30; 16)$. Calcular $P(X \leq 19)$.

15) X é $N(20; 25)$. Calcular $P(X \leq 30)$.

16) X é $N(50; 81)$. Calcular $P(40 \leq X \leq 60)$.

17) X é $N(10; 16)$. Calcular $P(X \geq 5)$.

Respostas

1. 0,2643

2. a) 0,9998 b) 0,8665 c) 0,0132

3. 0,4325

4. a) 0,9987 b) 0,0062 c) 0,1359

5. a) 0,1151 b) 0,2119 c) 0,7175

6. 0,2090 7. 0,1587 8. 0,2033

9. 0,0062

10. a) 0,0001 b) 0,9759 c) 0,8413

11. a) 0,0228 b) 0,4472

12. a) 0,1525 b) 0,6915 c) 0,0228

13. 26,2% 14. 0,3% 15. 97,72%

16. 73,3% 17. 89,44%

