Algumas Distribuições de Probabilidade Contínuas

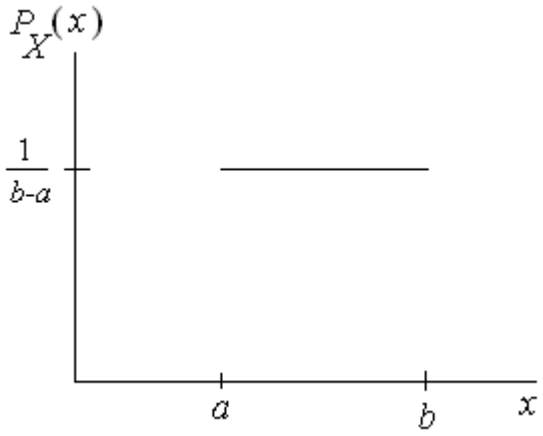
Vamos agora estudar algumas importantes distribuições de probabilidades para variáveis contínuas.

Distribuição Uniforme

Uma variável contínua X, definida dentro de um intervalo I limitado pelas constantes a e b ($I = [a, b]$, (a, b), $[a, b)$ ou $(a, b]$), é distribuída uniformemente dentro do intervalo se a sua função densidade de probabilidade $f_X(x)$ for

$$f_X(x) = \frac{1}{b-a}, \quad \text{para } x \in I.$$

O gráfico desta função densidade está mostrado abaixo.



A média (valor esperado) desta distribuição é dada por:

$$\mu = \int_a^b x f_X(x)\,dx = \int_a^b x\left(\frac{1}{b-a}\right)dx = \frac{\left(b^2-a^2\right)}{2(b-a)} = \frac{a+b}{2}.$$

A variância da distribuição uniforme é dada por (mostre como exercício)

$$\sigma^2 = \int_a^b (x-\mu)^2 f_X(x)dx = \int_a^b (x-\mu)^2 \left(\frac{1}{b-a}\right)dx =$$

$$= \left(\frac{1}{b-a}\right)\frac{(b-a)^3}{12} = \frac{(b-a)^2}{12}.$$

Um caso particular importante desta distribuição de probabilidades é aquele para o qual $a = 0$ e $b = 1$. Para este caso:

$$f_X(x) = 1$$

e

$$F_X(x) = \int_0^x f_X(x)dx = \int_0^x dx = x.$$

A média e a variância para este caso particular são, respectivamente:

$$\mu = \frac{1}{2} \quad \text{e} \quad \frac{1}{12}.$$

A Distribuição Normal

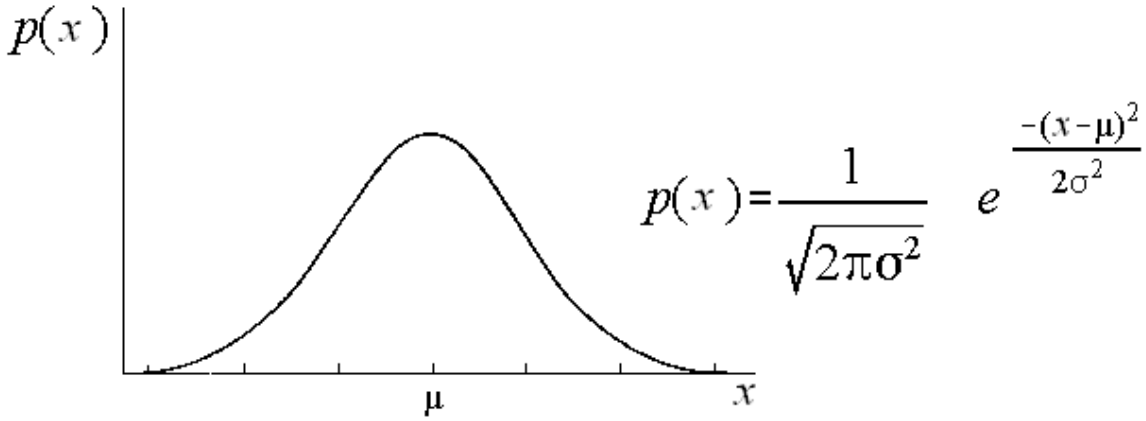
A distribuição contínua mais conhecida e mais usada é a chamada **distribuição normal**, ou **distribuição gaussiana**. A sua popularidade decorre do fato de que as distribuições empíricas de muitos fenômenos naturais (por exemplo, pressão arterial, altura, peso, valores de colesterol etc) são aproximadas pela distribuição normal.

Uma variável aleatória contínua X obedece a uma distribuição normal no intervalo $(-\infty, +\infty)$ se a sua função densidade de probabilidade for

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

As constantes μ e σ são os chamados *parâmetros* da distribuição. Pode-se mostrar que elas são iguais, respectivamente, à média e ao desvio padrão da distribuição normal.

O gráfico da densidade de probabilidade para a distribuição normal é dado abaixo.

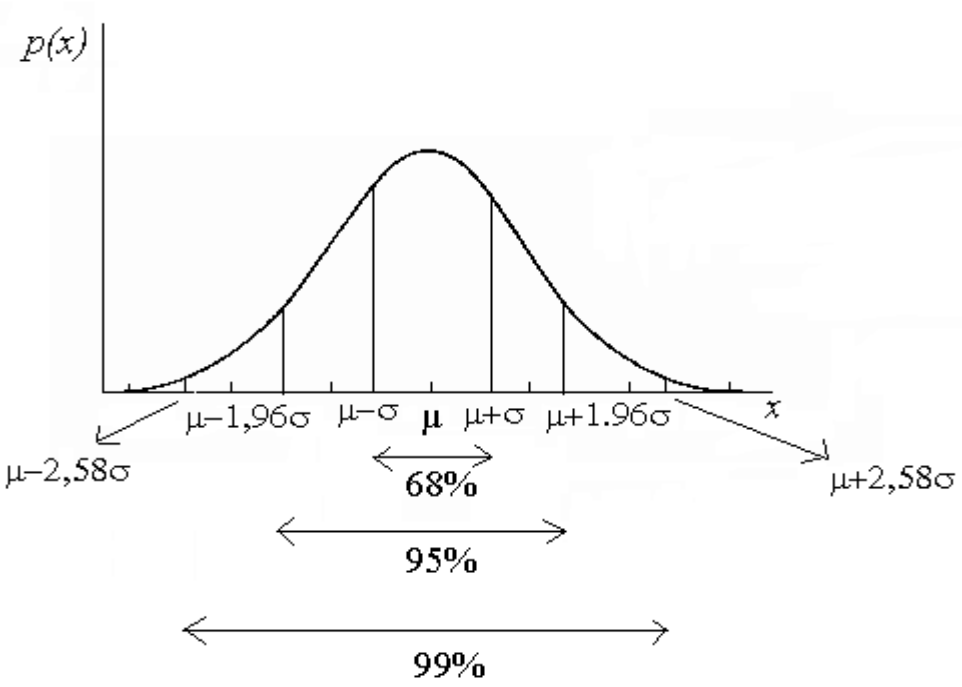


A Distribuição Normal

As características mais importantes da distribuição normal são as seguintes:

- Uma distribuição normal é unimodal, tem "forma de sino" e é simétrica em torno de sua média μ;
- A média, a mediana e a moda de uma distribuição normal são todas iguais;

- A área total abaixo da curva da distribuição normal é igual a 1 (pois ela é uma distribuição de probabilidades). Isto quer dizer que 50% das observações estão acima da média e 50% estão abaixo dela;
- Aproximadamente, 68% das observações estão dentro de uma região distante 1 desvio padrão (para ambos os lados) da média, ($\mu \pm 1\sigma$) e aproximadamente 95% das observações caem dentro de uma região distante 1,96 desvios padrões da média, ($\mu \pm 1{,}96\sigma$).



A figura acima mostra os valores das áreas abaixo de uma curva normal para alguns intervalos em torno da média. Note que essas áreas dão as probabilidades de que a variável *x* tenha um valor dentro do intervalo correspondente.

- 68% de todos os valores de *x* estão em um intervalo que vai de $x = \mu - \sigma$ a $x = \mu + \sigma$.
- 95% de todos os valores de *x* estão no intervalo $(\mu - 1{,}96\sigma \; ; \; \mu + 1{,}96\sigma)$.
- 99% de todos os valores de *x* estão no intervalo $(\mu - 2{,}58\sigma \; ; \; \mu + 2{,}58\sigma)$.

Por exemplo, se uma variável aleatória X estiver distribuída de acordo com uma distribuição normal, podemos dizer com certeza que 95% de todos os possíveis valores de X estarão dentro de uma faixa com $1{,}96\sigma$ de distância da média μ para cada lado. Dito de outro modo, podemos dizer que a probabilidade de observar um valor de X na faixa $\mu \pm 1{,}96\sigma$ é de 95%.

Para cada valor da média μ e do desvio padrão σ existe uma função (uma curva) normal diferente. Dizemos que μ e σ são os parâmetros da distribuição normal. No entanto, **todas** as funções de distribuição normais satisfazem as propriedades acima.

Podemos ter curvas com a mesma dispersão (mesmo σ), mas centradas em pontos diferentes (μs diferentes) e curvas com dispersões diferentes (σs diferentes) centradas no mesmo ponto (mesmo μ) (veja a figura abaixo):

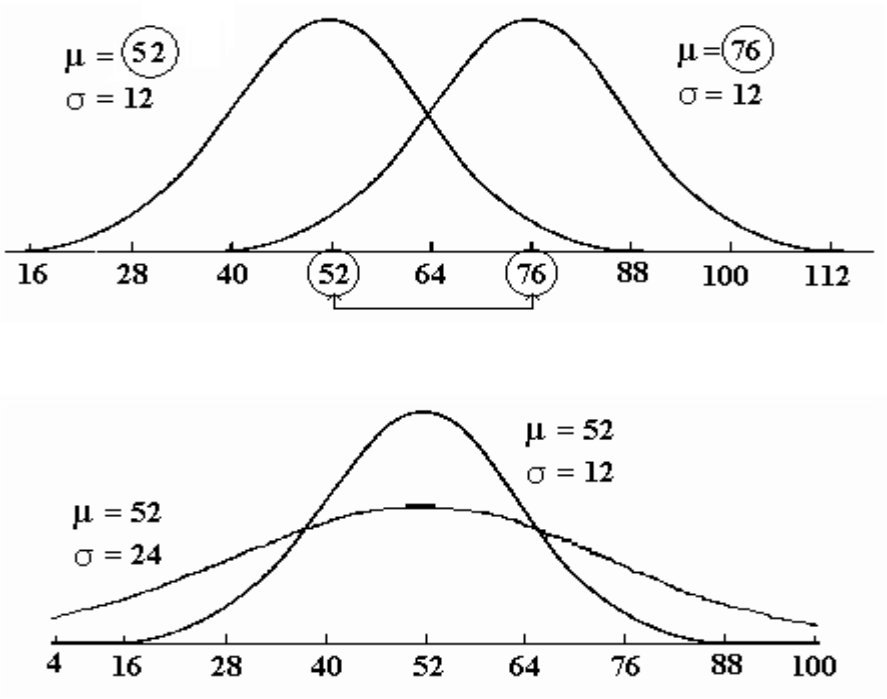


Figura adaptada de: David W. Stockburger, Introductory Statistics: Concepts, Models and Applications, 1996. http://www.psychstat.smsu.edu/introbook/sbk00.htm.

Para cada par de parâmetros μ e σ pode ser traçada uma curva normal. Além disso, apenas uma curva normal pode ser traçada para um par μ, σ. Se, para um conjunto de dados, soubermos sua média, seu desvio padrão e que eles satisfazem uma distribuição normal, então a distribuição dessa população estará totalmente caracterizada. Costuma-se denotar que uma dada variável aleatória tem distribuição normal com média μ e desvio padrão σ pela notação: $N(\mu, \sigma)$.

Para se calcular a probabilidade de que uma dada variável aleatória X tenha valor dentro de um intervalo (a, b) deve-se calcular a integral,

$$P(a < x < b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx .$$

Esta integral não pode ser calculada analiticamente de forma exata. No entanto, pode-se fazer aproximações numéricas para se obter o seu valor com grande precisão. Os valores das integrais para cada intervalo (a, b) podem, então, ser colocados em tabelas para uso quando necessário.

Seria impossível fazer um livro contendo tabelas para todas as distribuições normais possíveis (para cada par de valores μ e σ). Por causa disto, faz-se uma transformação na variável x levando-a para uma variável z que também satisfaz a uma distribuição normal, chamada de distribuição normal padrão. Desta forma, uma tabela apenas − aquela para a distribuição normal padrão − é suficiente para o cálculo de qualquer probabilidade para variáveis que satisfaçam a distribuições normais.

A distribuição normal padrão tem média zero e desvio padrão igual a 1. A área sob a curva da distribuição normal padrão, entre cada ponto no eixo horizontal *z* e a média 0, é dada pelos valores da tabela da próxima transparência. Por exemplo, para z = 0,73 a área é igual a 0,2673 (intersecção da linha de 0,7 com a coluna de 0,03).

Quando a variável aleatória *Y* em estudo tiver distribuição normal, mas com média diferente de zero e desvio padrão diferente de 1, ainda pode-se usar a tabela para a curva normal padrão. Para tal, é preciso converter os valores *y* da distribuição de *Y* para novos valores *z*, chamados de variáveis reduzidas ou escores-*z*. Esta conversão é dada por:

$$z = \frac{y - \mu}{\sigma} \qquad \frac{(\text{desvio em relação à média})}{(\text{desvio padrão})}.$$

De maneira pictórica, a transformação da variável original *y* para a variável reduzida *z* é mostrada na figura abaixo:

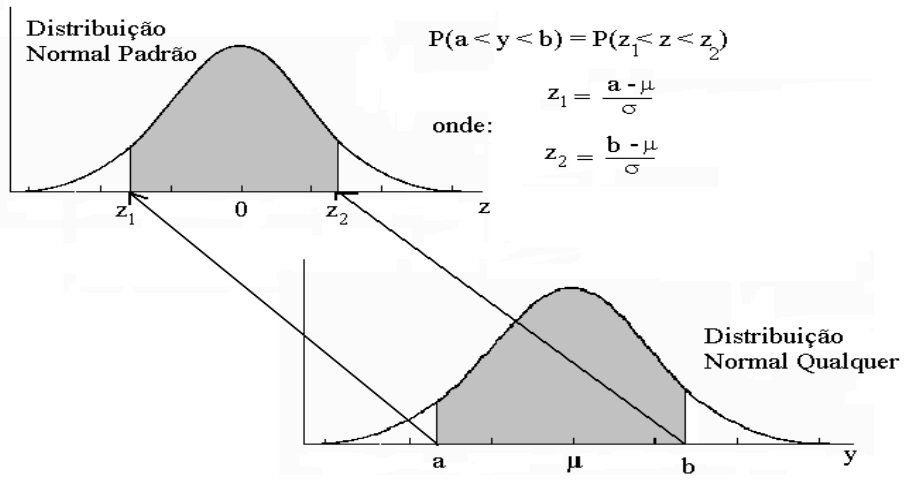
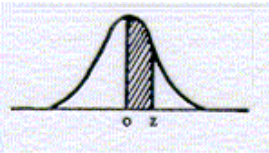


Tabela: Áreas de uma distribuição normal padrão

Cada casa na tabela dá a fração sob a curva inteira entre z=0 e um valor positivo de z. As áreas para os valores de z negativos são obtidas por simetria.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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	0,2703	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	0,3261	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
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Exemplo ilustrativo: Suponhamos que se tenha uma população de pessoas adultas para as quais foi medida a sua altura. A distribuição das alturas é normal com média de 172 cm e desvio padrão de 5 cm.

1) Que proporção dessa população tem estaturas menores que 177 cm?

Aplicando a fórmula $z = (y - \mu)/\sigma$, temos:

$$z = \frac{177 - 172}{5} = \frac{5}{5} = 1 \text{ (variável reduzida).}$$

Da tabela para a curva normal padrão, a área entre a média 0 e z=1,00 é 0,3413. Como queremos a proporção de alturas menores que 177 cm temos que considerar também os valores abaixo da média, de 172 cm para trás. A área sob a curva de cada lado da média vale 0,5 (0,5 de 0 a +∞ e 0,5 de –∞ a 0). Portanto, a porcentagem de alturas abaixo de 177 cm é igual a 0,3413 + 0,5000 = 0,8413 ⇒ 84,13%.

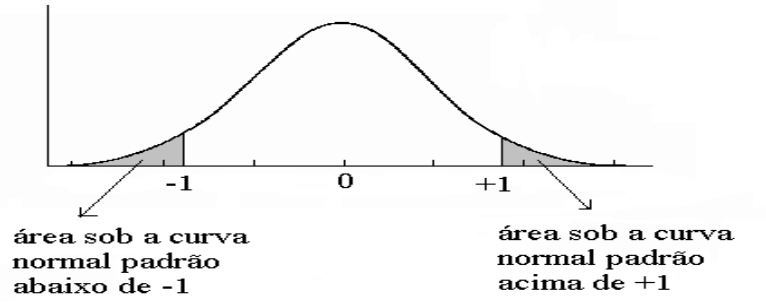
2) Que proporção dessa população tem estaturas maiores que 177 cm?

Basta tirar 0,8413 de 1, já que a área total sob a curva da distribuição normal padrão vale 1: 1 − 0,8413 = 0,1587 ⇒ 15,87%.

3) Que proporção dessa população tem estaturas abaixo de 167 cm?

De $z = (y - \mu)/\sigma$ temos: $z = \frac{167 - 172}{5} = \frac{-5}{5} = -1$.

A tabela não lista valores negativos de z, mas como a curva é simétrica (veja a figura a seguir), a área abaixo de –1 é igual à área acima de +1, ou seja, 15,87%.



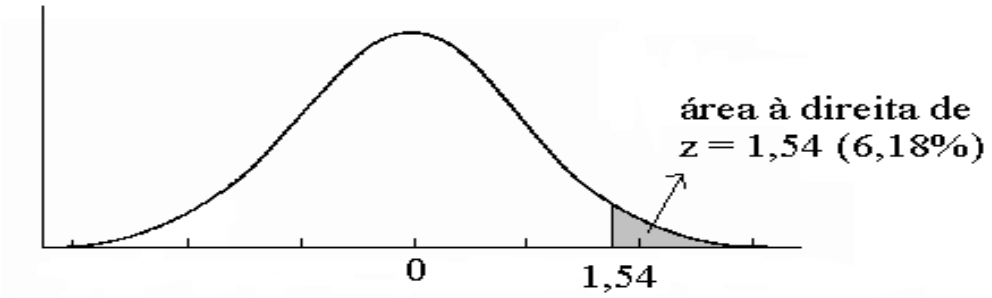
4) Que proporção dessa população tem estaturas dentro do intervalo entre 167 cm e 177 cm?

Dos resultados anteriores, temos que: para $y = 167$, $z = -1$; e para $y = 177$, $z = +1$. A área entre $z = 0$ e $z = 1$ é igual a 0,3413 (34,13%). Deste modo, a área entre $z = -1$ e $z = +1$ será igual a duas vezes este valor, ou seja, 0,6826 (68,26%).

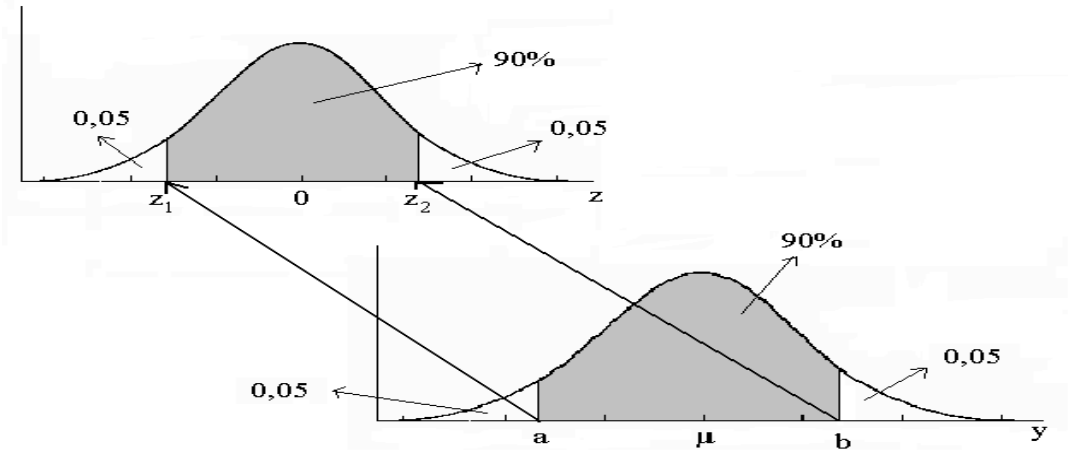
5) Que proporção dessa população tem estaturas maiores que 179,7 cm?

O escore-z para este caso vale: $z = \dfrac{179{,}7 - 172}{5} = 1{,}54$.

Da tabela, vemos que a área entre $z = 0$ e $z = 1{,}54$ é igual a 0,4382. Subtraindo este valor de 0,5000, obtemos a área à direita de $z = 1{,}54$, que é 0,0618 $\Rightarrow$ 6,18%.



6) Entre quais valores de estaturas estarão 90% de todas as estaturas medidas, de maneira que 5% delas estejam acima do valor superior e 5% estejam abaixo do valor inferior?



A área entre z_1 e z_2 é igual a 0,90. Portanto, a área entre 0 e Z_2 deve ser igual a 0,45. Olhando na tabela, o valor de *z* para o qual a área entre 0 e *z* está mais próxima de 0,45 é *z* = 1,65 (o valor correto seria 1,645, pois 0,45 está a meio caminho entre 0,4495 e 0,4505, mas para os propósitos deste exemplo *z* = 1,65 basta). Logo, $z_1 = 1{,}65$ e $z_2 = -1{,}65$. Voltando para a variável *y*, obtemos os valores limites do intervalo:

$$z = \frac{y-\mu}{\sigma} \qquad \begin{aligned} &\Rightarrow \; -1{,}65 = \frac{y_1 - 172}{5} \Rightarrow y_1 = 163{,}75 \\ &\Rightarrow \; +1{,}65 = \frac{y_2 - 172}{5} \Rightarrow y_2 = 180{,}25 \end{aligned} .$$