Introdução à Estatística

Prof^{a}. Juliana Freitas Pires

Departamento de Estatística
Universidade Federal da Paraíba - UFPB
juliana@de.ufpb.br

Variáveis Aleatórias e Distribuições de Probabilidades

$\rightarrow$ Até aqui, vimos a teoria básica de probabilidade, a qual é importante para a compreensão do conceito da probabilidade e suas propriedades.

$\rightarrow$ Conhecer as probabilidades associadas ao experimento em estudo é saber como se comportam as chances de ocorrência de seus resultados, e isto é bastante útil na solução de problemas práticos.

$\rightarrow$ Já existem diversos modelos teóricos, chamados **distribuições de probabilidades**, os quais são adequados para modelar um experimento.

Definição: Uma variável aleatória (v.a.) é uma função que descreve os resultados de um experimento através de valores numéricos.

Uma variável aleatória pode ser classificada em dois tipos:

- Variável Aleatória Discreta
- Variável Aleatória Contínua

Variável Aleatória Discreta

Definição: Denomina-se X uma variável aleatória discreta se o número de valores possíveis de X for um conjunto de pontos finito ou infinito enumerável.

Exemplos:

- Número de ações vendidas de uma empresa.
- Número de erros de trasmissão em um processo.
- Número de aparelhos defeituosos em uma produção.

Variável Aleatória Contínua

Definição: Denomina-se X uma variável aleatória discreta se ela assume valores num intervalo de números reais.

Exemplos:

- resistência de um material;
- concentração de CO2 na água
- tempo de vida de um componente eletrônico;
- tempo de resposta de um sistema computacional;

Distribuição de Probabilidade

Definição: A distribuição de probabilidade descreve as probabilidades dos resultados numéricos de um experimento, ou seja, associa uma probabilidade de cada valor de uma variável aleatória.

Existem, na literatura, diversas distribuições de probabilidades para modelar variáveis aleatórias discretas e contínuas.

Distribuição de Probabilidade

Desta forma, na prática, quando desejamos investigar algum fenômeno, estamos na realidade interessados em estudar a distribuição que melhor modele uma variável aleatória.

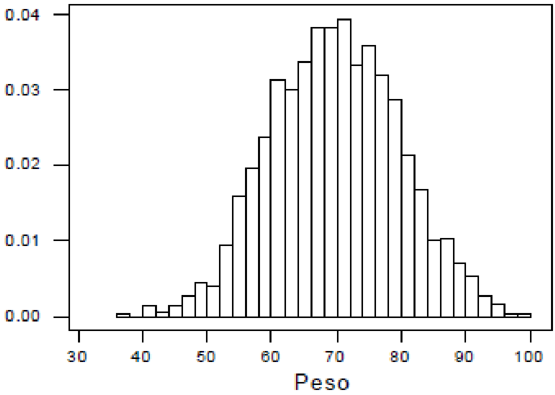
Aqui, estudaremos a distribuição de probabilidade Normal, a qual é apropriada para modelar diversas v.a's contínuas nas mais diferentes áreas.

Distribuição Normal

Motivação: Distribuição Normal

Exemplo: Observamos o peso (kg) de 1500 pessoas adultas selecionadas ao acaso em uma população.

O histograma é o seguinte:



Motivação: Distribuição Normal

A análise do histograma indica que:

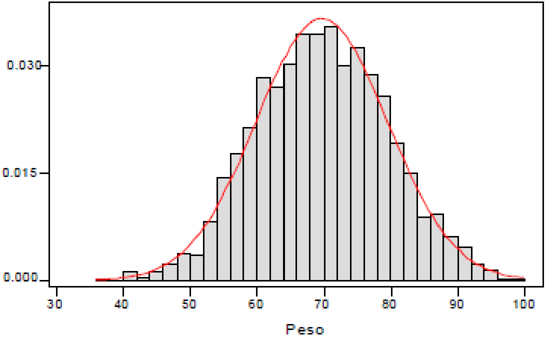
- a distribuição dos valores é aproximadamente simétrica em torno de 70kg;
- a maioria dos valores (88%) encontra-se em torno da média, no intervalo $(55; 85)$;
- a proporção das alturas vai diminuindo a medida que os valores se afastam da média. Existe apenas uma pequena proporção de valores abaixo de 48kg $(1,2\%)$ e acima de 92kg (1%).

Motivação: Distribuição Normal

Vamos definir a variável aleatória:

$X =$ peso, em kg, de uma pessoa adulta escolhida ao acaso da população.

Qual a distribuição de probabilidade de X?

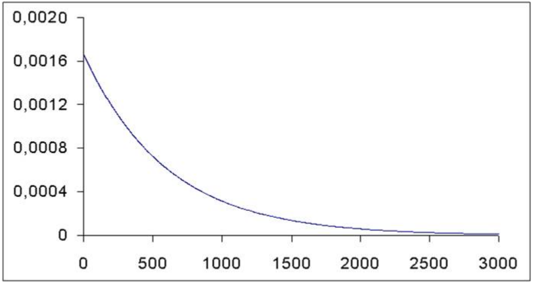


A curva contínua denomina-se curva Normal.

- A distribuição Normal é uma das mais importantes distribuições contínuas de probabilidade.

- Muitos fenômenos aleatórios comportam-se de forma próxima a essa distribuição. Exemplos:
  1. Altura;
  2. Pressão sanguínea;
  3. Peso.

Nem todos os fenômenos se ajustam à distribuição Normal. Exemplo:
Y = Duração, em horas, de uma lâmpada.



A distribuição de Y deve ser assimétrica. Em que, a grande proporção de valores está entre 0 e 500 horas e poucos valores acima de 1500 horas.

Uma variável aleatória X tem **distribuição Normal** com parâmetros μ e σ^2 se sua função densidade de probabilidade é dada por:

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

- Campo de variação: $-\infty < X < +\infty$;
- $\mathrm{E}(X) = \mu$;
- $\mathrm{Var}(X) = \sigma^2$ (e portanto, $\mathrm{DP}(X) = \sigma$).

Distribuição Normal

Notação: $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ indica que v.a. X tem distribuição Normal com parâmetros μ e σ^2.

Propriedades:

- É simétrica em torno da média μ;
- A média e a mediana são coincidentes;
- A área total sob a curva é igual a 1;
- $x = \mu$ é ponto de máximo de $f(x)$;
- $f(x) \to 0$ quando $x \to \pm\infty$.

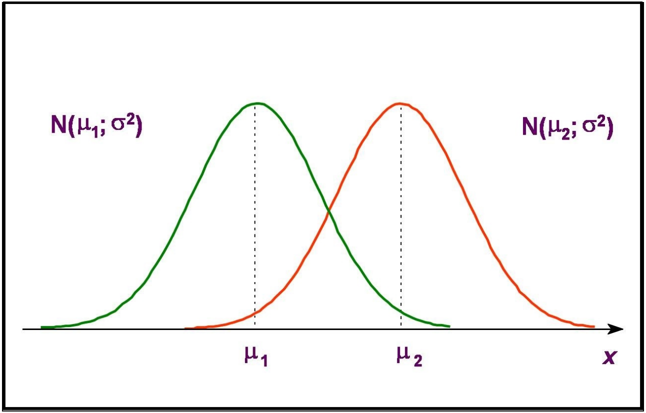
Distribuição Normal Padrão

Se $\mu = 0$ e $\sigma^2 = 1$, a distribuição é chamada de **distribuição normal padrão** e a função de densidade de probabilidade reduz-se a:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}$$

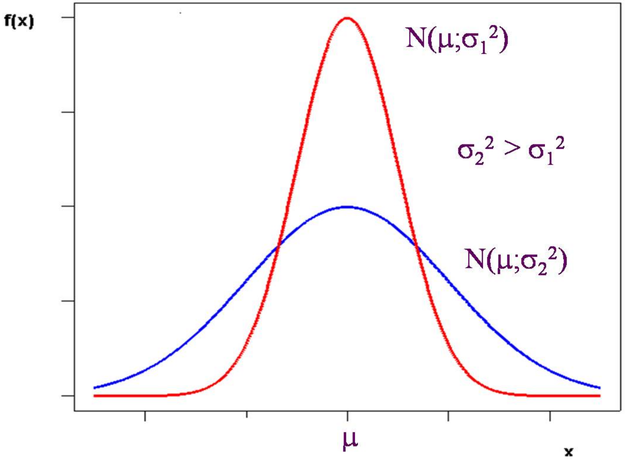
ou seja, $Z \sim \mathcal{N}(0, 1)$.

Influência de μ na curva da Distribuição Normal



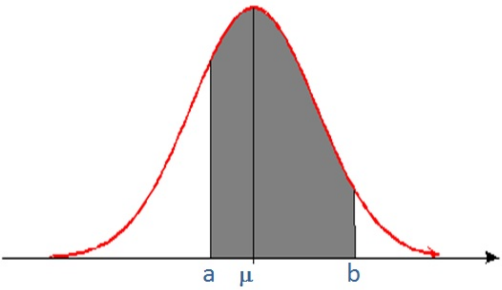
Curvas normais com mesma variância, porém com médias diferentes $(\mu_2 > \mu_1)$.

Influência de σ^2 na curva da Distribuição Normal



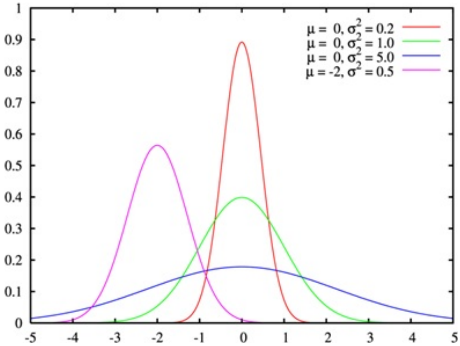
Curvas normais com mesma média, porém com variâncias diferentes ($\sigma_2^2 > \sigma_1^2$).

P$(a < X < b) =$ área sob a curva e acima do eixo horizontal (X) entre a e b.

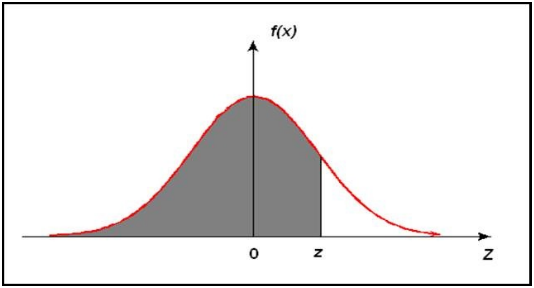


Os cálculos dessas áreas (probabilidades) já foram obtidos e registrados em tabelas.

Pergunta: Se $f(x)$ depende de μ e σ^2, então temos disponíveis uma infinidade de Tabelas, uma para cada par μ e σ^2??

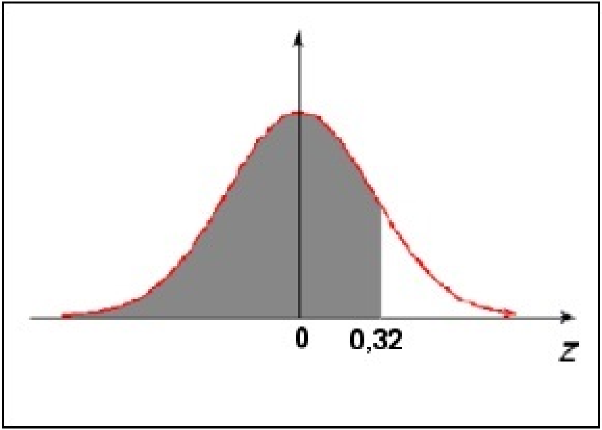


Essas probabilidades que estão registradas em tabelas são para variáveis que tem distribuição normal padrão ($Z \sim \mathcal{N}(0, 1)$).



Essa é a área fornecida pela tabela e é denotada por $\mathrm{P}(Z \leq z)$.

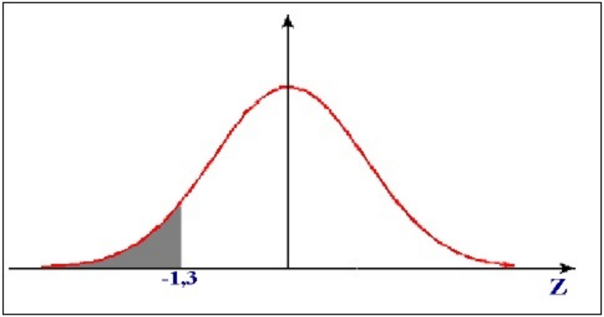
- Calcular $\mathrm{P}(Z \leq 0,32)$.



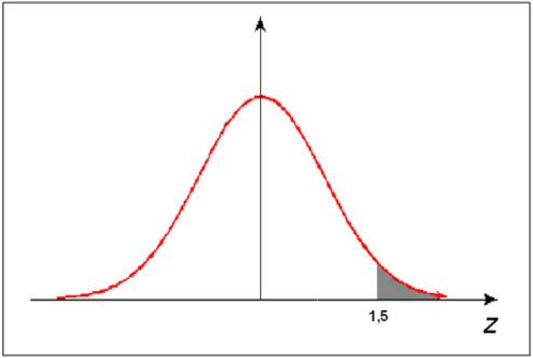
z	0	1	2
0,0	0,5000	0,5039	0,5079
0,1	0,5398	0,5437	0,5477
0,2	0,5792	0,5831	0,5870
0,3	0,6179	0,6217	0,6255

$$P(Z \leq 0,32) = 0,6255.$$

$\mathrm{P}(Z \leq -1,3) = 0,0968.$



$P(Z \geq 1,5)$



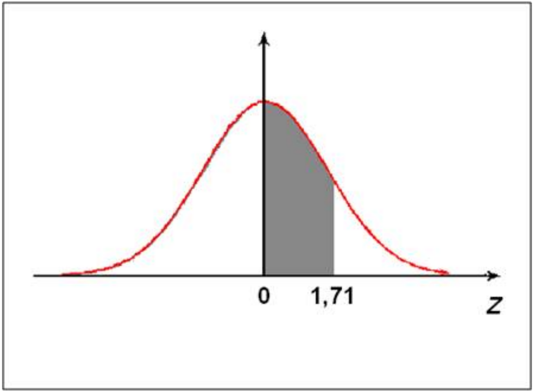
$$\begin{aligned} P(Z \geq 1,5) &= 1 - P(Z < 1,5) \\ &= 1 - 0,9332 \\ &= 0,0668 \end{aligned}$$

ou

$$P(Z \geq 1,5) = P(Z \leq -1,5) = 0,0668$$

obs: $P(Z \geq z) = P(Z \leq -z)$, pela simetria

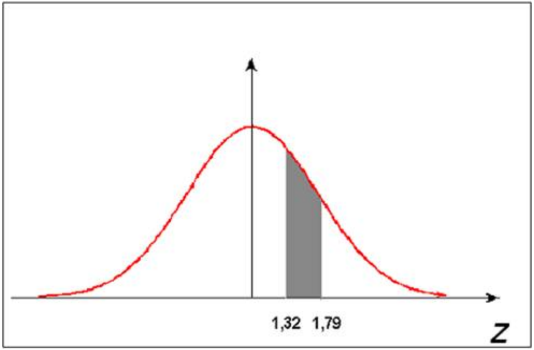
$\mathrm{P}(0 < Z \leq 1,71)$



$$
\begin{aligned}
\mathrm{P}(0 < Z \leq 1,71) &= \mathrm{P}(Z \leq 1,71) - \mathrm{P}(Z < 0) \\
&= 0,9564 - 0,5 \\
&= 0,4564
\end{aligned}
$$

obs: $\mathrm{P}(Z < 0) = \mathrm{P}(Z > 0) = 0,5$, pela simetria

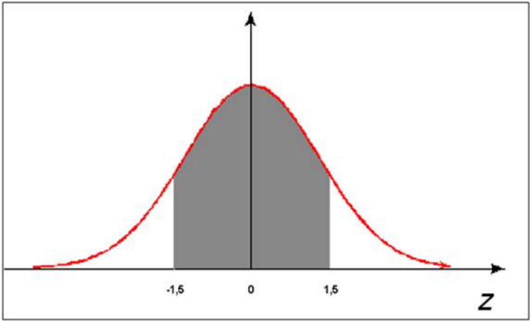
$$\mathrm{P}(1,32 < Z \leq 1,79)$$



Exemplo

$$
\begin{aligned}
\mathrm{P}(1,32 < Z \leq 1,79) &= \mathrm{P}(Z \leq 1,79) - \mathrm{P}(Z < 1,32) \\
&= 0,9633 - 0,9066 \\
&= 0,0567
\end{aligned}
$$

$$P(-1,5 < Z < 1,5)$$



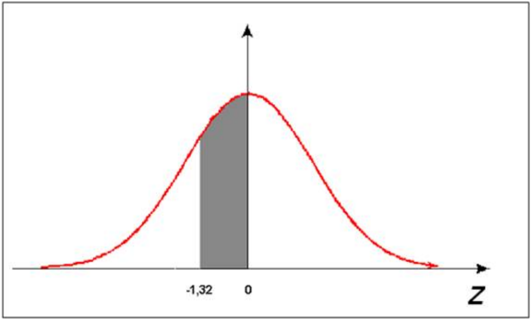
$$
\begin{aligned}
P(-1,5 < Z < 1,5) &= P(Z < 1,5) - P(Z < -1,5) \\
&= 0,9332 - 0,0668 \\
&= 0,8664
\end{aligned}
$$

ou

$$
\begin{aligned}
P(-1,5 < Z < 1,5) &= 1 - 2 \times P(Z < -1,5) \\
&= 1 - 2 \times 0,0668 \\
&= 0,8664
\end{aligned}
$$

obs: $P(Z > 1,5) = P(Z < -1,5)$, pela simetria

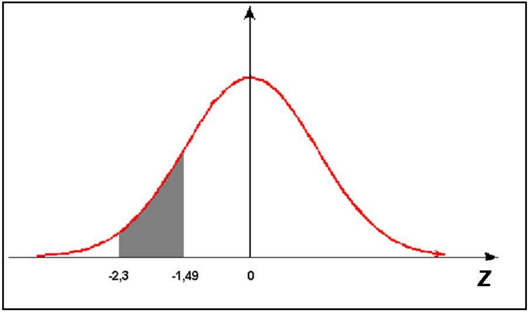
$\mathrm{P}(-1,32 < Z < 0)$



Exemplo

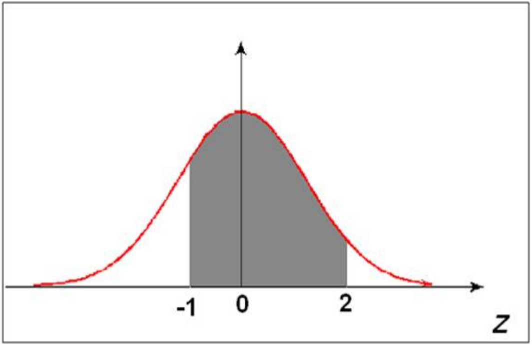
$$\begin{aligned} \mathrm{P}(-1,32 < Z < 0) &= \mathrm{P}(Z < 0) - \mathrm{P}(Z < -1,32) \\ &= 0,5 - 0,0934 \\ &= 0,4066 \end{aligned}$$

$$P(-2,3 < Z < -1,49)$$



$$
\begin{aligned}
\mathrm{P}(-2,3 < Z < -\,1,49) = & \\
&= \mathrm{P}(Z < -1,49) - \mathrm{P}(Z < -2,3) \\
&= 0,0681 - 0,0107 \\
&= 0,0574
\end{aligned}
$$

$$P(-1 < Z < 2)$$

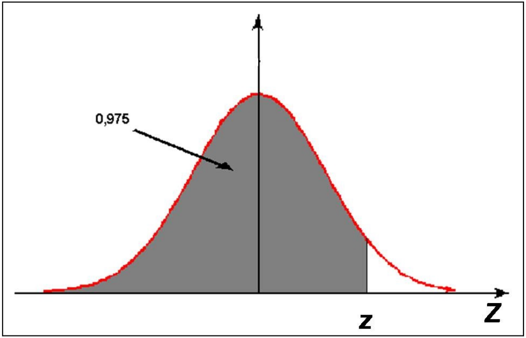


Exemplo

$$
\begin{aligned}
\mathrm{P}(-1 < Z < 2) &= \mathrm{P}(Z < 2) - \mathrm{P}(Z < -1) \\
&= 0,9773 - 0,1587 \\
&= 0,8186
\end{aligned}
$$

- Em alguns momentos nós sabemos a probabilidade de ocorrência de determinado evento e estamos interessados em saber quem é esse evento.
- A seguir veremos como fazemos para resolver esse tipo de situação.

- Como encontrar o valor z da distribuição $\mathcal{N}(0,1)$, tal que $\mathrm{P}(Z \leq z) = 0,975$?



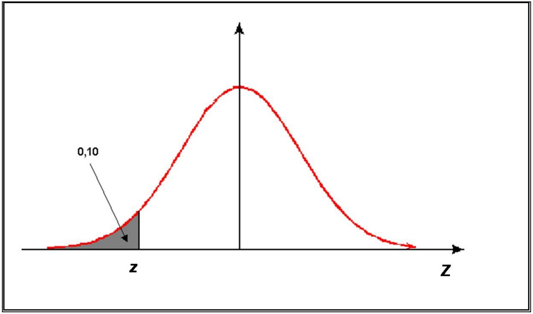
z	4	5	6
1,6	0,9495	0,9505	0,9515
1,7	0,9591	0,9599	0,9608
1,8	0,9671	0,9678	0,9686
1,9	0,9738	0,9744	0,9750

Exemplo

- $\mathrm{P}(Z \leq z) = 0,975$
- Pela tabela, $z = 1,96$.

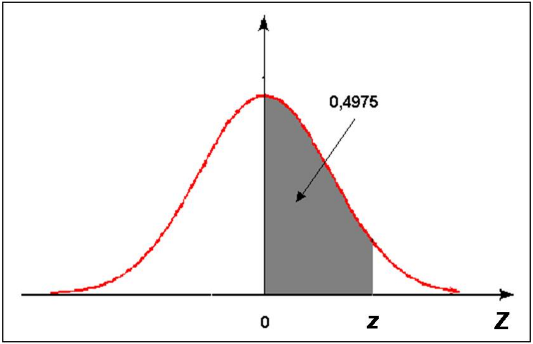
Exemplo

- Como encontrar o valor z da distribuição $\mathcal{N}(0, 1)$ tal que $\mathrm{P}(Z \leq z) = 0,10$?



- $P(Z \leq z) = 0,10$
- Pela tabela, $z = -1,28$.

- Qual o valor z da distribuição $\mathcal{N}(0,1)$ tal que $\mathrm{P}(0 < Z \leq z) = 0,4975$?

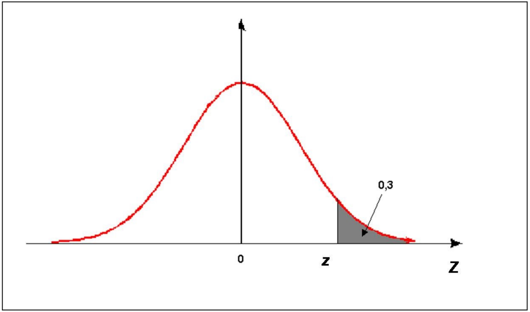


- Note que

$$P(Z < z) = P(Z \leq 0) + P(0 < Z \leq z)$$
$$P(Z < z) = 0,5 + 0,4975$$
$$P(Z < z) = 0,9975$$

- Pela tabela, $z = 2,81$.

- Como encontrar o valor z da distribuição $\mathcal{N}(0,1)$ tal que $\mathrm{P}(Z \geq z) = 0,3$?

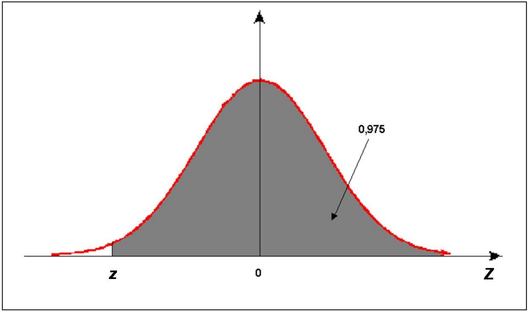


- Note que

$$\begin{aligned} \mathrm{P}(Z \geq z) &= 0,3 \\ 1 - \mathrm{P}(Z < z) &= 0,3 \\ \mathrm{P}(Z < z) &= 1 - 0,3 \\ \mathrm{P}(Z < z) &= 0,7 \end{aligned}$$

- Pela tabela, $z = 0,53$.

- Como encontrar o valor z da distribuição $\mathcal{N}(0,1)$ tal que $\mathrm{P}(Z \geq z) = 0,975$?

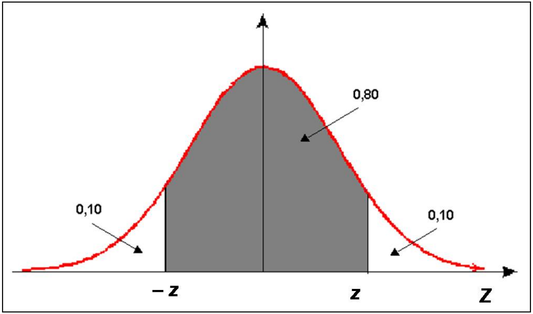


- Note que

$$
\begin{aligned}
\mathrm{P}(Z \geq z) &= 0,975 \\
1 - \mathrm{P}(Z < z) &= 0,975 \\
\mathrm{P}(Z < z) &= 1 - 0,975 \\
\mathrm{P}(Z < z) &= 0,025
\end{aligned}
$$

- Pela tabela, $z = -1,96$.

- Como encontrar o valor z da distribuição $\mathcal{N}(0,1)$ tal que $\mathrm{P}(-z \leq Z \leq z) = 0,8$?



- Note que

$$P(Z \leq -z) = 0,1$$
$$\text{ou}$$
$$P(Z \leq z) = 0,9$$

- Pela tabela, $-z = -1,28$ e $z = 1,28$.

Mas como eu faço para calcular as probabilidades quando a variável não tem distribuição normal padrão ($\mathcal{N}(0,1)$)?

- Nesses casos as áreas (probabilidades) são calculadas através de uma transformação.
- Essa transformação transforma qualquer variável normal que tenha qualquer média e qualquer variância em uma variável normal padrão (com média zero e variância um).

- Essa transformação é feita da seguinte maneira:

  Seja $X \sim \mathcal{N}(\mu, \sigma^2)$, definimos $Z = \dfrac{X - \mu}{\sigma}$

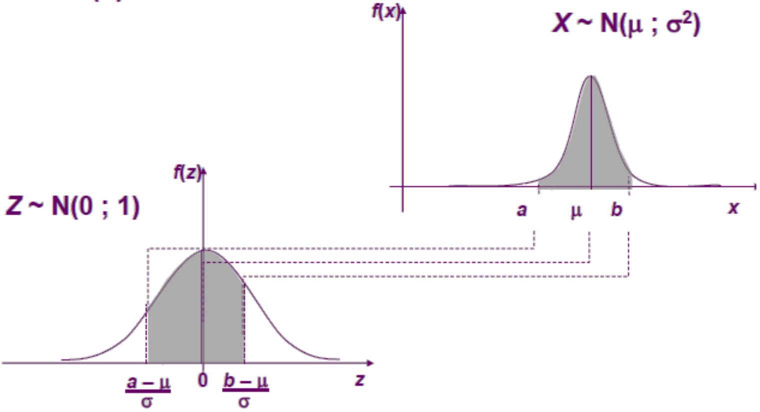
  Temos então: $\mathrm{E}(Z) = 0$ e $\mathrm{Var}(Z) = 1$.

- Com essa transformação $Z \sim \mathcal{N}(0, 1)$ (distribuição normal padrão).

Se $X \sim N(\mu\,;\sigma^2)$, definimos $\boxed{Z = \dfrac{X-\mu}{\sigma}}$

$E(Z) = 0$

$Var(Z) = 1$

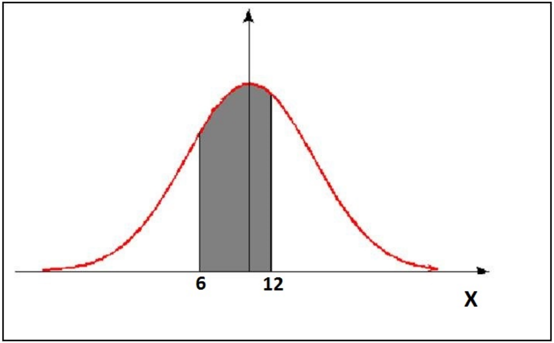


Com isso, podemos determinar as probabilidades de uma v.a. $X \sim \mathcal{N}(\mu, \sigma^2)$, com base na tabela da distribuição normal padronizada. Portanto,

$$\begin{aligned}
\mathrm{P}\left(a < X < b\right) &= \mathrm{P}\left(a - \mu < X - \mu < b - \mu\right) \\
&= \mathrm{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\
&= \mathrm{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)
\end{aligned}$$

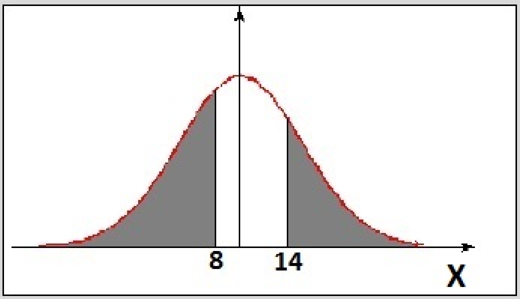
Seja $X \sim \mathcal{N}(10; 64)$ $\big(\mu = 10,\ \sigma^2 = 64 \text{ e } \sigma = 8\big)$

- Calcular $\mathrm{P}(6 \leq X \leq 12)$



$$
\begin{aligned}
\mathrm{P}(6 \leq X \leq 12) &= \mathrm{P}\left(\frac{6-10}{8} \leq \frac{X-10}{8} \leq \frac{12-10}{8}\right) \\
&= \mathrm{P}\left(-0,5 \leq Z \leq 0,25\right) \\
&= \mathrm{P}(Z < 0,25) - \mathrm{P}(Z < -0,5) \\
&= 0,5987 - 0,3085 \\
&= 0,2902
\end{aligned}
$$

- Calcular $\mathrm{P}(X \leq 8 \text{ ou } X > 14)$



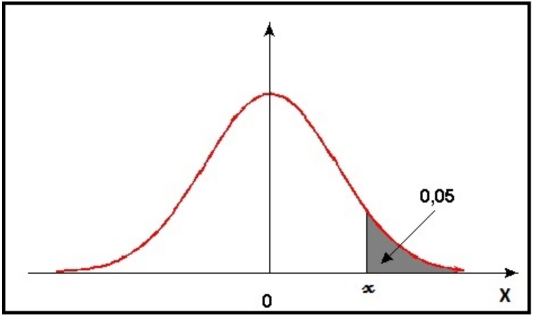
$$
\begin{aligned}
&\mathrm{P}(X \leq 8) + \mathrm{P}(X > 14) = \\
&= \mathrm{P}(Z \leq \frac{8-10}{8}) + \mathrm{P}(Z > \frac{14-10}{8}) \\
&= \mathrm{P}(Z \leq -0,25) + \mathrm{P}(Z > 0,5) \\
&= \mathrm{P}(Z \leq -0,25) + \mathrm{P}(Z < -0,5) \\
&= 0,7098
\end{aligned}
$$

Um problema inverso quando $X \sim \mathcal{N}(\mu; \sigma^2)$

- Existem situações em que sabemos a probabilidade de ocorrência de determinado evento de uma v.a. $X \sim \mathcal{N}(\mu; \sigma^2)$ e estamos interessados em saber quem é esse evento.
- Em tais momentos, podemos obter a v.a. $X \sim \mathcal{N}(\mu; \sigma^2)$ por meio da v.a. $Z \sim \mathcal{N}(0; 1)$, através da transformação inversa:

$$X = \mu + Z\sigma.$$

- Como encontrar o valor x da distribuição $X \sim \mathcal{N}(10; 64)$ tal que $\mathrm{P}(X \geq x) = 0,05$?



$$P(X \geq x) = 0,05 \Rightarrow P\left(Z \geq \frac{x-10}{8}\right) = 0,05$$

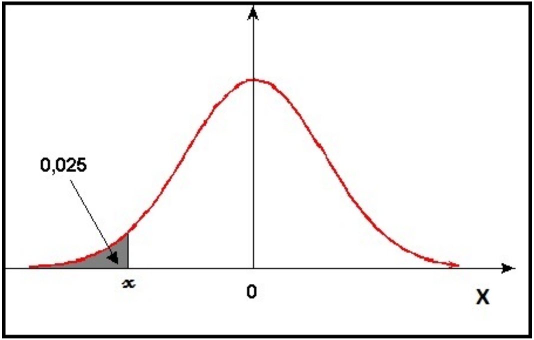
- temos que

$$\begin{aligned} P(Z \geq z) &= 0,05 \\ 1 - P(Z < z) &= 0,05 \\ P(Z < z) &= 1 - 0,05 \\ P(Z < z) &= 0,95 \end{aligned}$$

- Pela tabela, $z = 1,64$. Então,

$$\frac{x-10}{8} = 1,64 \quad \Rightarrow \quad x = 10 + 1,64 \times 8 = 23,12$$

- Como encontrar o valor x da distribuição $X \sim \mathcal{N}(10; 64)$ tal que $\mathrm{P}(X \leq x) = 0,025$?



$$P(X \leq x) = 0,025 \Rightarrow P\left(Z \leq \tfrac{x-10}{8}\right) = 0,025$$

- temos que $P(Z \leq z) = 0,025$
- Pela tabela, $z = -1,96$. Então,

$$\frac{x-10}{8} = -1,96 \quad \Rightarrow \quad x = 10 - 1,96 \times 8 = -5,68$$

O tempo de instalação de um software tem distribuição normal com média de 6 minutos e variância de 4 minutos.

a) Qual a probabilidade de que um software leve entre 5 e 7 minutos para ser instalado?

b) Qual a probabilidade de que um software leve mais que 6,5 minutos para ser instalado?

c) Qual a probabilidade de que um software leve menos que 5 minutos para ser instalado?

d) Qual o intervalo de tempo, simétrico em torno da média, que detém uma probabilidade 95% para que o software seja instalado?