Noções de Estatística

Aula 7

O Modelo de distribuição Normal

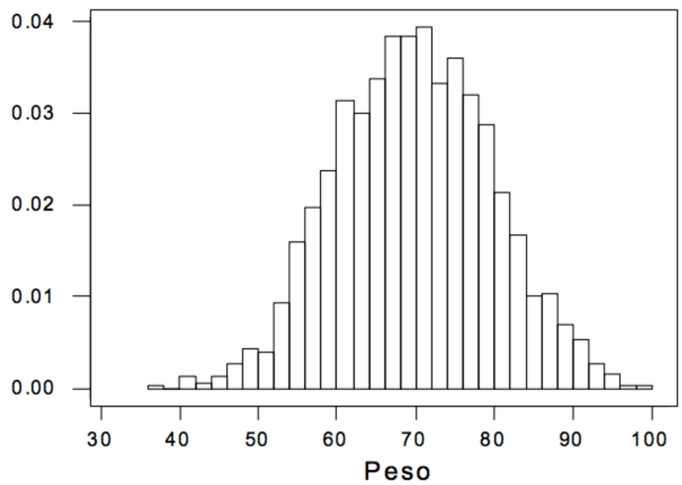
7.1 Introdução

O modelo de distribuição normal é um importante modelo probabilístico. Suas origens remontam a Gauss em seus trabalhos sobre erros de observações astronômicas, por volta de 1810, cujo modelo escolheu nomear de distribuição gaussiana. Esse modelo é muito importante pois se apresenta frequentemente como um modelo teórico excelente para representar diversos comportamentos na natureza, além de possibilitar o desenvolvimento de diversos teoremas que hoje nos permitem construir testes estatísticos.

A distribuição normal é uma distribuição que modela variáveis quantitativas contínuas.

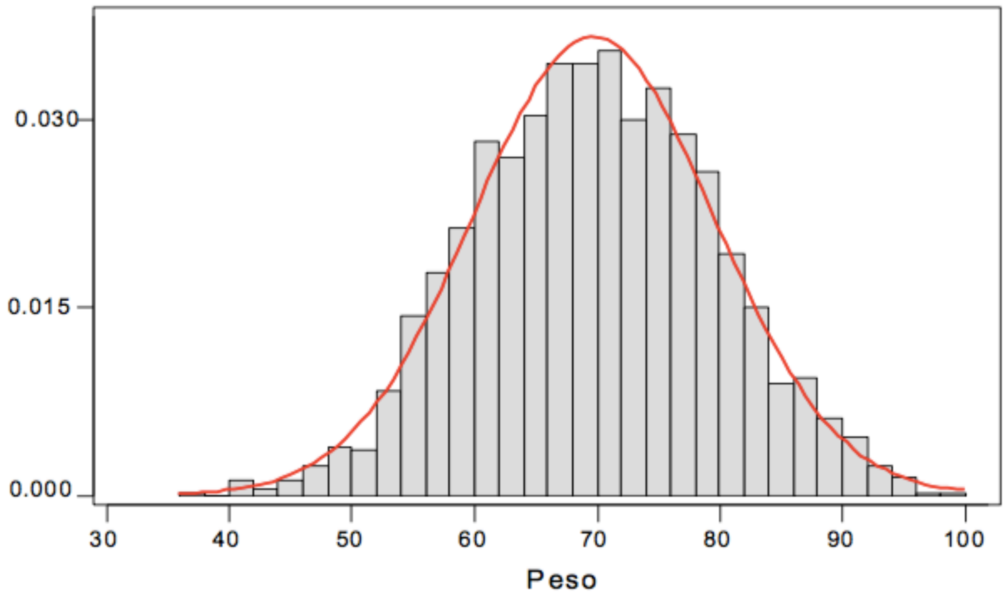
Exemplo:

Observe o histograma abaixo.



Esse tipo de comportamento pode ser observado muito frequentemente vezes na natureza. Sempre que houver uma concentração desse tipo (muitos valores em torno da média que vão diminuindo gradativamente) podemos supor um modelo de distribuição normal para os dados.

(Curva da normal teórica em vermelho)



Se há uma variável X com esse tipo de distribuição, dizemos que ela é uma variável X com distribuição normal com parâmetros: média μ e variância σ^2.

$$X \sim N(\mu, \sigma^2)$$

Quando temos $\mu = 0$ e $\sigma^2 = 1$, temos uma distribuição normal padrão.

Para quê serve a distribuição normal padrão?

Como vimos, frequentemente nos deparamos com variáveis que apresentam esse comportamento da Normal. Mas essas variáveis apresentam grandezas distintas, tornando o cálculo das probabilidades muito complicado.

Se quisermos calcular a probabilidade de X pertencer a um intervalo, teríamos que utilizar artifícios matemáticos muito avançados, como integral por exemplo.

Apenas por curiosidade, a função que produz o modelo da normal é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in (-\infty, \infty).$$

Para calcular a probabilidade de X pertencer a um intervalo desta função, precisaríamos integrar toda esta função com relação ao intervalo pedido.

Felizmente, a normal padrão é uma forma de facilitar esse cálculo.

Padronizando a variável normal, podemos compará-la à variável normal padrão, para a qual já existem tabelas prontas para o cálculo das probabilidades.

Para padronizar uma variável, basta criar uma nova variável Z.

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

Essa nova variável Z terá média 0 e desvio-padrão 1, possibilitando portanto a utilização de uma tabela que foi desenvolvida para o cálculo das probabilidades de uma distribuição Normal Padrão.

7.2 Cálculo de probabilidades para a distribuição normal

Retomando: Se X tem distribuição normal, X é uma variável contínua.

Se X é uma variável contínua, a probabilidade de X ser igual a um ponto específico é zero.

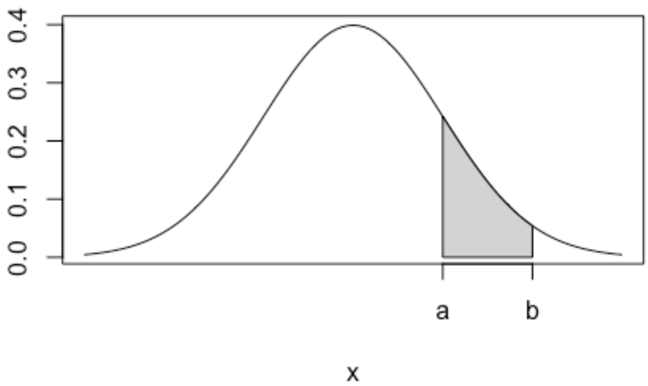
$$P(X = a) = 0$$

Quando tratamos de variáveis contínuas, sempre iremos pensar em probabilidades intervalares.

Ou seja,

$$P(a < X \leq b)$$
$$P(X < b)$$

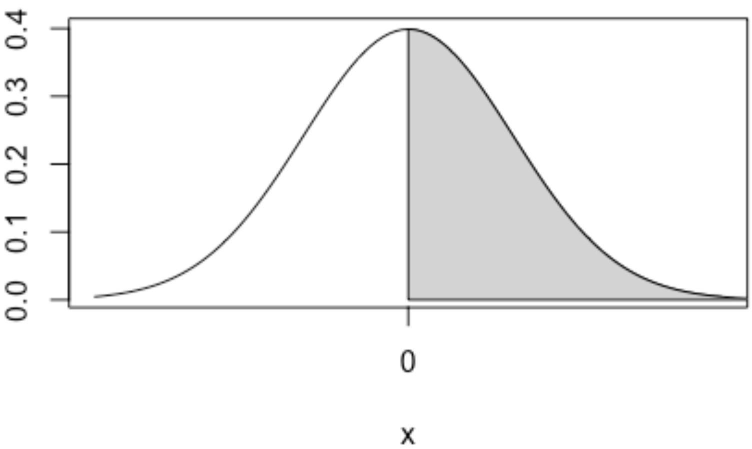
etc...

Após  a

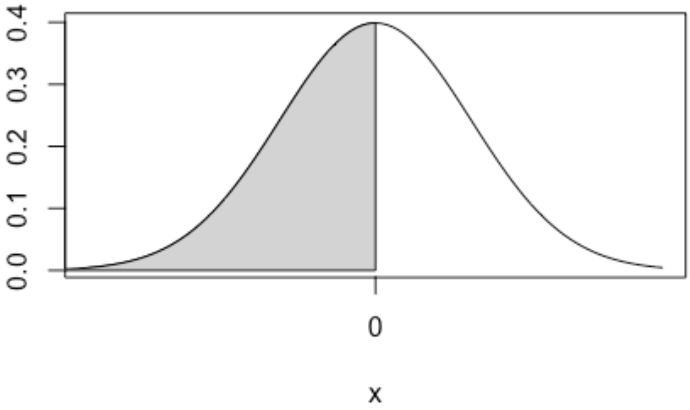
padronização da variável, deve-se fazer uso de uma tabela para o cálculo das probabilidades.

A distribuição normal padronizada tem sempre média 0. A probabilidade de $X < 0 = 0.5$, e claro, a probabilidade de $X > 0 = 0.5$ também.

$$P(X > 0) = 0.5$$

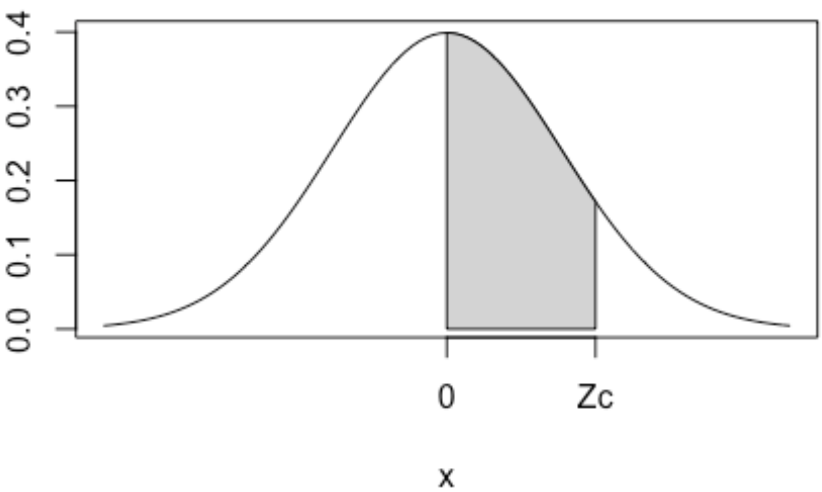


$$P(X < 0) = 0.5$$



Como utilizar a tabela?

A tabela da Normal foi elaborada para nos fornecer a probabilidade de P(0 < X < Zc), ou seja, um intervalo do tipo:



Se por um acaso for do nosso interesse calcular a probabilidade de X ser maior que um Zc (um Zc maior que 0), devemos calcular inicialmente a probabilidade de X ser menor que o Zc e, em seguida, calculamos 0.5 – a probabilidade obtida por meio da tabela.

Se por um acaso precisarmos calcular $P(a < X < b)$, $(a,b) > 0$.

Precisaremos calcular:

- $P(0 < X < a)$
- $P(0 < X < b)$

E em seguida:

$$P(0 < X < b) - P(0 < X < a)$$

Como exercício, como calcularíamos probabilidades para valores menores que 0?

E para o seguinte caso?