

Distribuição Normal

Prof. Eduardo Bezerra

(CEFET/RJ) - BCC - Inferência Estatística

16 de março de 2018

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal
- 3 Distribuição normal padrão
- 4 Distribuição normal - padronização
- 5 Distribuição normal bivariada

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal
- 3 Distribuição normal padrão
- 4 Distribuição normal - padronização
- 5 Distribuição normal bivariada

Definições iniciais - função densidade de probabilidade

Uma variável aleatória contínua tem **função densidade de probabilidade** $f(x)$ se f é uma função não-negativa integrável tal que a probabilidade no intervalo $[a, b]$ é dada por

$$\int_a^b f(x) dx$$

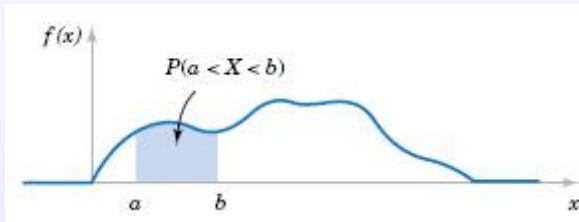
quaisquer que sejam a e b , e a probabilidade de todo o espaço amostral é 1:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

A função de densidade é zero para valores de x que não podem ocorrer e é considerada zero em pontos em que ela não for especificamente definida.

Definições iniciais - função densidade de probabilidade

A probabilidade de X estar entre a e b é igual à integral de $f(x)$ de a até b , isto é, igual à área sob a curva de $f(x)$ entre a e b .



$$\Pr(a < X < b) = \int_a^b f(x) dx$$

Definições iniciais - função de distribuição acumulada

A **função de distribuição acumulada** de uma variável aleatória contínua em um ponto x é definida como a integral da função de densidade deste $-\infty$ até o ponto x :

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^x f(y) dy$$

Definições iniciais - valor esperado e variância

Para uma variável aleatória contínua X , seu **valor esperado** é definido pela integral de todos os valores da função $xf(x)$:

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

e sua variância é definida como

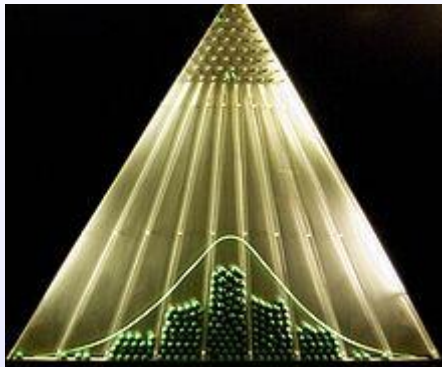
$$\sigma^2 = \text{Var}(X) = \int (x - \mu)^2 f(x) dx$$

Nas expressões acima, temos integrais definidas sobre o intervalo válido para X .

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal**
- 3 Distribuição normal padrão
- 4 Distribuição normal - padronização
- 5 Distribuição normal bivariada

Distribuição normal



Fonte da figura: http://en.wikipedia.org/wiki/File:Planche_de_Galton.jpg

Distribuição normal

Uma variável aleatória contínua X é dita ter **distribuição normal** se sua função de densidade de probabilidade é dada por:

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

Onde:

- π é a constante matemática ($\pi \approx 3,14159$).
- \exp é a função exponencial $\exp(y) = e^y$ ($e \approx 2,71828$).
- μ é a média da distribuição, denominado parâmetro de localização.
- σ é o desvio-padrão da distribuição, denominado parâmetro de forma (ou dispersão).

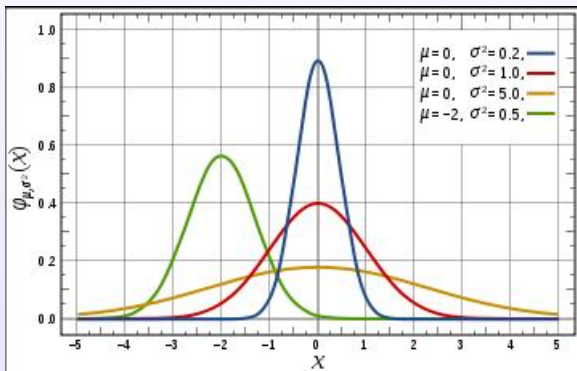
Distribuição normal

Outros dois termos associados à distribuição normal são

- **curva do sino** (*bell curve*), em razão da semelhança geométrica entre o gráfico da função de densidade e um sino.
- **distribuição gaussiana**, em homenagem a Carl Friedrich Gauss.

Distribuição normal

Gráficos de 4 funções de densidade para algumas combinações dos parâmetros μ e σ .



Distribuição normal

- A função de distribuição acumulada $F(X)$ de uma v.a. normal não possui uma forma fechada.
- Média:

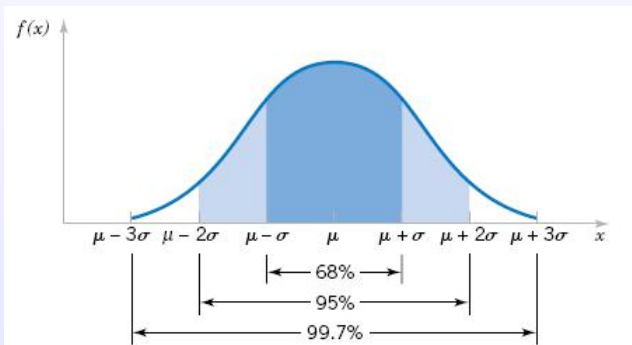
$$E(X) = \mu$$

- Variância:

$$\text{Var}(X) = \sigma^2$$

Distribuição normal

Se $X \sim N(\mu, \sigma^2)$, a faixa $\mu \pm 3 \times \sigma$, ocorre a quase totalidade (99,73%) dos valores. Por esse motivo, o valor 6σ é denominado **largura da distribuição normal**.



Distribuição normal

As probabilidades associadas a distribuições gaussianas são obtidas de 2 maneiras alternativas:

- 1 Se as probabilidades são necessárias como parte de um programa de computador, a integral de $f(x)$ pode ser resolvida por algoritmos de cálculo numérico.
- 2 Se apenas algumas probabilidades são necessárias, então podemos fazer uso de **tabelas** para calculá-las. Essas tabelas são encontradas na maioria dos livros de estatística, e apresentam valores para $F(x) = \Pr(X < x)$.

Distribuição normal

- De acordo com as propriedades gerais aplicáveis a uma função de densidade qualquer,
 - 1) a área total sob a curva é igual a 1 porque indica a probabilidade de todo o espaço amostral.
 - 2) a área sob a curva entre dois valores quaisquer de X indica a probabilidade de seu valor estar entre esses valores.
- A análise da **curva normal** nos permite concluir que as ocorrências de X tendem a se concentrar em torno de μ e se tornam mais raras (i.e., menos prováveis) na medida em que dela se afastam.

Distribuição normal

- Por meio de **integração numérica** da função de densidade normal, é possível calcular a probabilidade de ocorrência em função do afastamento da média segundo o número de desvios-padrão.
- Os valores apresentados a seguir representam as probabilidades de se encontrar um valor de X na faixa respectiva.
 - 0,6827 ou 68,27% para a faixa $\mu \pm 1 \times \sigma$
 - 0,9545 ou 95,45% para a faixa $\mu \pm 2 \times \sigma$
 - 0,9973 ou 99,73% para a faixa $\mu \pm 3 \times \sigma$

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal
- 3 Distribuição normal padrão**
- 4 Distribuição normal - padronização
- 5 Distribuição normal bivariada

Distribuição normal padrão

- Damos o nome de **distribuição normal padrão** (*standard normal distribution*) à distribuição normal com média $\mu = 0$ e desvio-padrão $\sigma = 1$.
- Escrevemos $X \sim N(\mu, \sigma^2)$ para denotar que uma variável aleatória X tem distribuição normal de média μ e variância σ^2 .
- Em particular, se uma variável aleatória Z tem distribuição normal padrão, então usamos $Z \sim N(0, 1)$.

Distribuição normal padrão

Exemplo 01 (Montgomery, pp. 112): Seja $Z \sim N(0, 1)$. Calcule

- a) $\Pr(Z < -0,86)$
- b) $\Pr(Z > 1,26)$
- c) $\Pr(Z > -1,37)$
- d) $\Pr(-1,25 < Z < 0,37)$
- e) $\Pr(Z < -4,6)$

Distribuição normal padrão

Resposta - Exemplo 01:

- a) $\Pr(Z < -0,86) = 0,1949$.
- b) $\Pr(Z > 1,26) = 1 - \Pr(Z \leq 1,26) = 1 - 0,8962 = 0,1038$.
- c) $\Pr(Z > -1,37) = \Pr(Z < 1,37) = 0,9147$.
- d) $\Pr(-1,25 < Z < 0,37) = \Pr(Z < 0,37) - \Pr(Z < -1,25) = 0,6443 - 0,1056 = 0,5387$.
- e) O menor valor da v.a. normal padrão Z encontrado na tabela de probabilidades é $-3,49$. Portanto, apenas podemos afirmar que $\Pr(Z < -4,6) < \Pr(Z < -3,49) = 0,0002$.

Distribuição normal padrão

Exemplo 02:

- Seja $Z \sim N(0, 1)$.
- Encontre o valor z tal que $\Pr(Z > z) = 0,05$.

Distribuição normal padrão

Exemplo 03:

- Seja $Z \sim N(0, 1)$.
- Encontre o valor z tal que $\Pr(-z < Z < z) = 0,99$.

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal
- 3 Distribuição normal padrão
- 4 Distribuição normal - padronização**
- 5 Distribuição normal bivariada

Distribuição normal - padronização

- Seja X uma v.a. com distribuição normal genérica, $X \sim N(\mu, \sigma^2)$. Então

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

é também um variável aleatória e possui distribuição normal padrão, isto é, $Z \sim N(0, 1)$

- Isso significa que, a partir de valores tabelados da distribuição normal padrão, é possível o cálculo de probabilidades para quaisquer valores dos parâmetros μ e σ^2 (i.e., para qualquer outra v.a. normal).

Distribuição normal - padronização

Esta **padronização** é conveniente, já que permite o cálculo da função de densidade e da função acumulada de uma distribuição normal qualquer, uma vez que temos acesso aos valores da função de densidade e da função acumulada da distribuição normal padrão. Essas funções são relacionadas da seguinte forma

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad f_X(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right).$$

Distribuição normal - padronização

Exemplo 04: as alturas dos alunos de determinada escola são normalmente distribuídas com média 1,6m e desvio padrão 0,30m.

- 1) Encontre a probabilidade de um aluno medir:
 - a) entre 1,50m e 1,80m
 - b) mais de 1,75m
 - c) menos de 1,48m
- 2) Qual deve ser a altura mínima para 10% dos mais altos?

Distribuição normal - padronização

Resposta - exemplo 04: Primeiro, note que

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1,50 - 1,60}{0,3} = -0,33$$

e que

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{1,80 - 1,60}{0,3} = 0,67.$$

Então

$$\begin{aligned} \Pr(x_1 < X < x_2) &= \Pr(z_1 < Z < z_2) = \\ \Pr(-0,33 < Z < 0,67) &= \Pr(Z < 0,67) - \Pr(Z < -0,33) = \\ &= 0,7486 - 0,3707 = \\ &= 0,3779 = 37,79\%. \end{aligned}$$

Roteiro

- 1 Distribuições contínuas
- 2 Distribuição normal
- 3 Distribuição normal padrão
- 4 Distribuição normal - padronização
- 5 Distribuição normal bivariada**

Distribuição normal bivariada

