UNIVERSIDADE DA REGIÃO DE JOINVILLE - UNIVILLE

Bacharelado em Engenharia de Software (BES)

Estatística para computação

Professora Priscila Ferraz Franczak

Engenheira Ambiental - UNIVILLE

Mestre em Ciência e Engenharia de Materiais - UDESC

Doutora em Ciência e Engenharia de Materiais - UDESC

priscila.franczak@gmail.com

Plano de Aula

Probabilidades

1. Modelo Binomial
2. Modelo Poisson
3. Modelo Normal
4. Geração de números aleatórios
5. Exercícios

Podemos definir probabilidade de A como segue:

- Seja E um experimento, S o espaço amostral relativo a E e A um evento.
- Podemos associar um número real ao evento A, representado por P(A), denominado probabilidade de A ocorrer, definido por:

$$P(A) = \frac{\textit{número de casos favoráveis a } A}{\textit{número de casos possíveis no espaço } S}$$

$$P(A) = \frac{n(A)}{n(S)}$$

Regras básicas da probabilidade

1) Campo de variação das probabilidades

A probabilidade de um evento A deve ser um número maior ou igual a 0, porém menor ou igual a 1.

$$0 \leq P(A) \leq 1 \; ou \; 0\% \leq P(A) \leq 100\%$$

2) Probabilidade do espaço amostral

A probabilidade do espaço amostral S é igual a 1.

$$P(S) = 1 \; ou \; P(S) = 100\%$$

3) Regra da adição de probabilidades

A probabilidade de ocorrência do evento A **ou** do evento B (ou de ambos) é igual a:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Caso os eventos A e B sejam mutuamente exclusivos, isto é:

$A \cap B = \emptyset$, então:

$$P(A \cup B) = P(A) + P(B)$$

Essa regra pode ser estendida para n eventos mutuamente exclusivos: A_1, A_2, A_3,..., A_n.

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

4) Probabilidade de um evento complementar

- Sabemos que um evento pode ocorrer ou não.

- Sendo P(A) a probabilidade de que ele ocorra (sucesso) e P($\bar{A}$) a probabilidade de que ele não ocorra (insucesso), para um mesmo evento existe sempre a relação:

$$P(\bar{A}) = 1 - P(A)$$

Exemplo de aplicação das regras básicas:

- **Considere um experimento que consiste em retirar uma carta de um baralho de 52 cartas.**

1) Campo de variação das probabilidades

S = {a_1, a_2, ..., a_{52}}, onde qualquer carta tem a mesma probabilidade de ser sorteada, ou seja: $P(a_i) = \frac{1}{52}$

1) Campo de variação das probabilidades

A probabilidade, por exemplo, do evento A = {retirar uma carta com o número 16} é dada por:

$$P(A) = \frac{número\ de\ casos\ favoráveis\ ao\ evento\ A}{número\ de\ casos\ possíveis}$$

$$P(A) = \frac{0}{52} = 0$$

1) Campo de variação das probabilidades

Por outro lado, se A = {retirar uma carta qualquer} então:

$$P(A) = \frac{número\ de\ casos\ favoráveis\ ao\ evento\ A}{número\ de\ casos\ possíveis}$$

$$P(A) = \frac{52}{52} = 1$$

Então: $0 \leq P(A) \leq 1$

2) Nesse caso o evento A = S, assim: P(A) = P(S)=1

3) Regra da adição

Deseja-se calcular a probabilidade de ser retirada uma carta vermelha ou um rei.

Seja A = {carta vermelha} e B = {rei}

(não são mutuamente exclusivos!)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52}$$

$$P(A \cup B) = \frac{7}{13}$$ ou 53,8%

3) Regra da adição

Deseja-se calcular a probabilidade de ser retirada uma carta de espadas ou uma dama de ouros.

Seja A = {carta de espadas} e B = {dama de ouros}

(são mutuamente exclusivos! Por quê?)

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = \frac{13}{52} + \frac{1}{52}$$

$$P(A \cup B) = \frac{14}{52}$$ ou 26,9%

4) Probabilidade de um evento complementar

Se A = {carta de paus}, então $\bar{A}$= {qualquer carta exceto paus}. Assim:

$$P(A) = 1 - P(\bar{A}) \text{ ou}$$

$$\mathrm{P}(\bar{A}) = 1 - P(A)$$

$$\mathrm{P}(\bar{A}) = 1 - \frac{13}{52}$$

$$\mathrm{P}(\bar{A}) = \frac{3}{4} \; ou \; 75\%$$

Multiplicação de probabilidades

- Dois eventos serão estatisticamente independentes se a ocorrência de um deles não afetar a ocorrência do outro.

Exemplo: lançamento de duas moedas:

S = {cc, kk, ck, kc}

probabilidade de cada resultado = $\frac{1}{4}$

- Dados dois eventos independentes, A e B, a probabilidade da ocorrência conjunta é definida pela regra da multiplicação:

$$P(A.B) = P(A \cap B) = P(A).P(B)$$

Essa regra é válida para *n* eventos independentes.

Exemplo: Em uma experiência que consiste em lançar simultaneamente, um dado e duas moedas, qual a probabilidade de obter o número 5 e duas coroas em uma única jogada?

Os eventos são independentes

$$P(5KK) = P(5 \cap K \cap K) = P(5).P(K).P(K)$$

$$P(5KK) = P(5 \cap K \cap K) = \frac{1}{6}.\frac{1}{2}.\frac{1}{2}$$

$$P(5KK) = \frac{1}{24}\ ou\ 4{,}16\%$$

1. Modelo Binomial

- O modelo de distribuição binomial é também conhecido como modelo de Bernoulli.

- Tem essa denominação por levar em conta apenas as probabilidades de sucesso, representada por p

- E insucesso representada por q sendo

$$q = 1 - p$$

Aplicações:

- Respostas de teste com diversas questões V ou F;
- Escolher entre produto bom ou defeituoso;
- Sexo das crianças nascidas em uma maternidade;
- Fumantes ou não fumantes em um grupo;
- Alunos vacinados ou não.

Suponhamos que um experimento é repetido n vezes e sejam:

1. X a variável aleatória;
2. A frequência do valor k da variável X seja dada por: $f_k = \frac{n!}{k!(n-k)!}$ importante! Número de combinações de n elementos, tomados k a k.

3. p a probabilidade de sucesso e q = 1 - p a probabilidade insucesso;

4. k o número de sucessos esperados.

Então, a probabilidade de ocorrer k sucessos é dada pela fórmula:

$$p(X = k) = \left(\frac{n!}{k!\,(n-k)!}\right) p^k q^{n-k}$$

- A esperança matemática é $\mu = np$
- A variância é $\sigma^2 = npq$
- O desvio padrão é $\sigma = \sqrt{npq}$

Seguem esse modelo experimentos do tipo lançar uma moeda, (cara ou coroa), nascimentos, (menino ou menina), respostas afirmativas ou negativas, etc.

Exemplo: Um laboratório proclama que a eficácia de determinada droga antidepressiva é 80%. Tal droga foi administrada a 10 pacientes. Pergunta-se:

A eficácia proclamada pelo fabricante é a probabilidade de sucesso. No problema temos:

n = 10

p = 0,8

$$p(X=k)=\left(\frac{n!}{k!(n-k)!}\right)p^{k}q^{n-k}$$

q = 0,2

(a) Qual a probabilidade de que exatamente 6 pacientes fiquem curados?

O evento é A = {6 pacientes ficam curados}.

A probabilidade de que exatamente 6 pacientes fiquem curados é obtida usando a fórmula:

$$
\begin{aligned}
P(x=6) &= \left(\frac{10!}{6!(10-6)!}\right)(0.8)^6(0,2)^{10-6} \\
&= 210(0.262140)(0.0016) \\
&= 0.08807
\end{aligned}
$$

P (x=6) = 8,81%

- Resolvendo no R:

```
> p<-0.80
> n<-10
> x<-6
> dbinom(x,n,p)
[1] 0.08808038
```

(b) Qual a esperança média de pacientes que deverão ficar curados?

A média de pacientes curados é

$$\mu = np$$
$$\mu = 10.0,80$$
$$\mu = 8\, pacientes$$

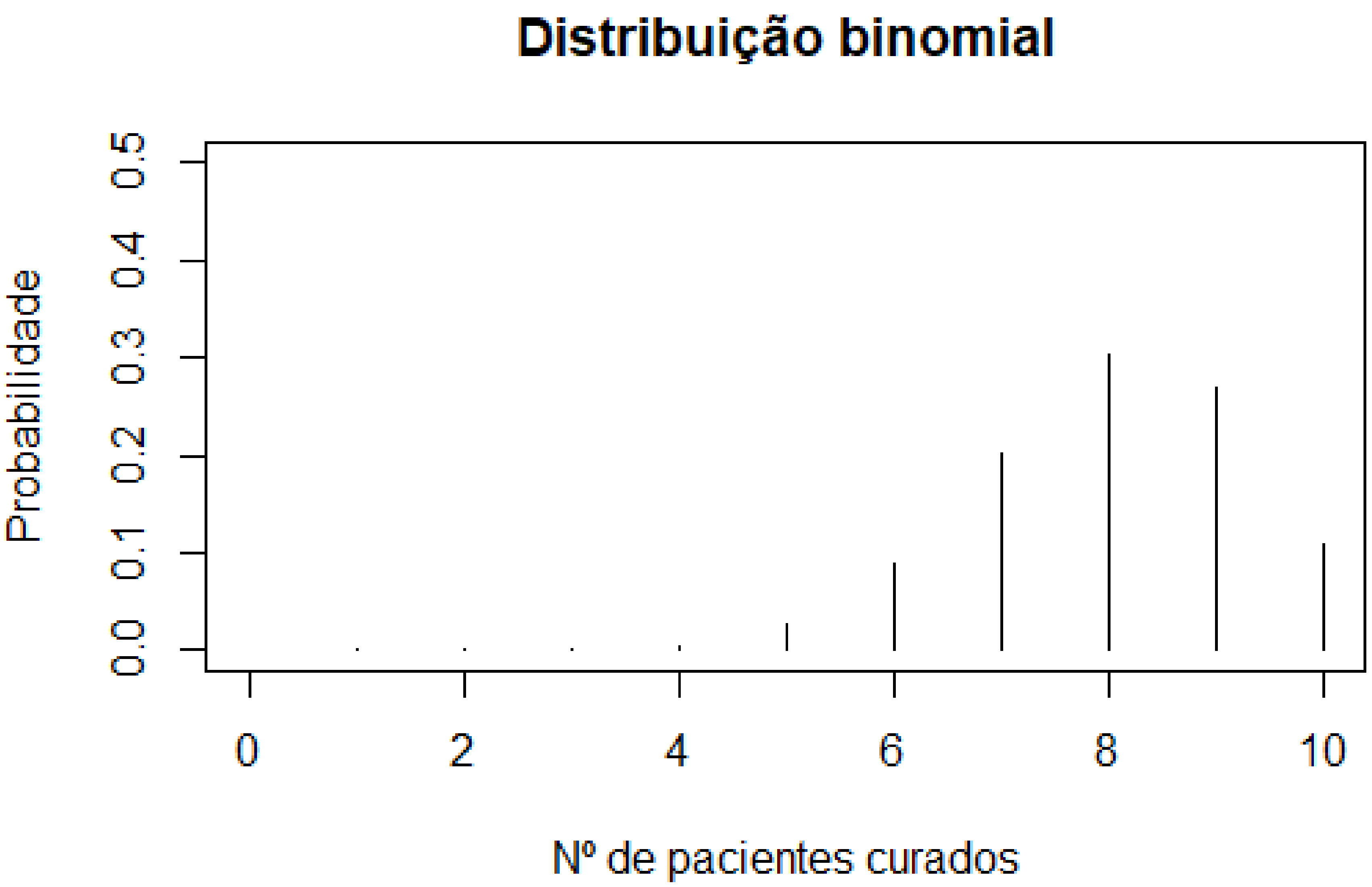
(c) Qual a variância e o desvio padrão?

$$s^2 = npq$$
$$s^2 = 10.0,80.0,20$$
$$s^2 = 1,6$$

$$s = \sqrt{npq}$$
$$s = 1,26$$

Plotando a distribuição binomial:

```
x<-1:10
p<-0.80
n<-10
binomial<-dbinom(x,n,p)
plot(x,binomial,type = "h",
      xlab = "Nº de pacientes curados",
      ylab = "Probabilidade",
      main = "Distribuição binomial",
      xlim = c(0,10),
      ylim = c(0,0.50))
```


Distribuição binomial
Probabilidade
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
Nº de pacientes curados

2. Modelo Poisson

A distribuição de Poisson é útil para descrever as probabilidades do número de ocorrências num campo ou intervalo contínuo.

- Exemplos:
  - defeitos por centímetro quadrado,
  - acidentes por dia,
  - reações alérgicas a uma vacina por 1.000.000 de pessoas vacinadas,
  - chamadas telefônicas (ou bip) por minuto (ou por dia),
  - número de pacientes que chegam a um pronto socorro por dia.

- Podemos enumerar os fatos que ocorreram, mas não é possível enumerar aqueles que deixaram de ocorrer.

Por exemplo, não podemos dizer quantos acidentes deixaram de acontecer em determinado dia e nem quantas reações alérgicas deixaram de ocorrer num dia de vacinação em massa.

As hipóteses para o uso da distribuição de Poisson são:

- A probabilidade de uma ocorrência é a mesma em todo o campo de observação.
- A probabilidade de mais de uma ocorrência num único ponto é aproximadamente zero.

- O número de ocorrências em qualquer intervalo (ou campo) é independente do número de ocorrências em outros intervalos (ou campos).
- A distribuição é binomial e o número de elementos da amostra é muito grande.

(Em estatística uma amostra é considerada grande quando tem mais de 30 elementos).

Seja X uma variável aleatória discreta tomando os seguintes valores: 0; 1; 2; 3...n... num intervalo, então a probabilidade de X ocorrer K vezes é:

$$p(X = K) = \frac{\lambda^K e^{-\lambda}}{K!}$$

- A média é $\lambda = np$
- *e* é a base dos logaritmos naturais = 2,718
- A variância é $\sigma^2 = np$
- O desvio padrão é $\sigma = \sqrt{\lambda}$

Exemplo 1: Uma central telefônica recebe uma média de 5 chamadas por minuto. Qual a probabilidade de receber nenhuma chamada durante um intervalo de 1 minuto?

- Temos a média λ = 5 chamadas por minuto. Queremos saber: p (K = 0). Então:

$$p(X = K) = \frac{\lambda^K e^{-\lambda}}{K!}$$

$$p(X = 0) = \frac{5^0 e^{-5}}{0!}$$

$$p(X = K) = \frac{1 * (2{,}718)^{-5}}{1}$$

$$p(X = K) = 0{,}0067$$

$$p(X = K) = 0{,}67\%$$

Resolvendo no R:

```
> dpois(0,5)
[1] 0.006737947
```

0,67%

Exemplo 2: O número de crianças que apresentaram reação alérgica a determinada vacina, durante os testes, foi 3 em 100.000.000 crianças observadas. Numa região serão vacinadas 100.000 crianças. Qual a probabilidade da ocorrência de:

(a) Ocorrer exatamente 2 casos de reações alérgicas?

(b) No máximo 2 casos?

(c) Pelo menos 3 casos?

(a) Ocorrer exatamente 2 casos de reações alérgicas?

- Como durante os testes foram observadas 3 casos de reações alérgicas em 100.000.000 crianças, a probabilidade de ocorrer uma reação alérgica será:

$$p = \frac{3}{100.000.000}$$

Já a média de reações alérgicas possíveis em 100.000 crianças a serem vacinadas será:

$$\lambda = np$$

$$\lambda = 100.000\left(\frac{3}{100.000.000}\right)$$

$$\lambda = 0{,}003$$

(a) Ocorrer exatamente 2 casos de reações alérgicas?

- Logo, para X = 2 teremos a probabilidade dada pela fórmula:

$$p(X = K) = \frac{\lambda^{K} e^{-\lambda}}{K!}$$

$$p(X = 2) = \frac{0{,}003^{2} e^{-0{,}003}}{2!}$$

$$p(X = 2) = 0{,}000004$$

$$p(X = 2) = 0{,}0004\%$$

Resolvendo no R:

$$K = 2 \; e \; \lambda = 0{,}003$$

```
> dpois(2,0.003)
[1] 4.48652e-06
```

0,0004%

(b) A probabilidade de ocorrer no máximo X = 2 casos de reações alérgicas em 100.000 crianças é dado por:

- Logo, para X = 2 teremos a probabilidade dada pela fórmula:

$$p(x < 3) = p(X = 0) + p(X = 1) + p(X = 2)$$

$$p(X = K) = \frac{\lambda^{K} e^{-\lambda}}{K!}$$

$$p(X < 3) = \frac{0{,}003^{0} e^{-0{,}003}}{0!} + \frac{0{,}003^{1} e^{-0{,}003}}{1!} + \frac{0{,}003^{2} e^{-0{,}003}}{2!}$$

$$p(X < 3) = 0{,}999994$$

$$p(X < 3) = 99{,}9994\%$$

Resolvendo no R:

```
> dpois(0,0.003)+dpois(1,0.003)+dpois(2,0.003)
[1] 1
```

100%

3. Modelo de distribuição normal

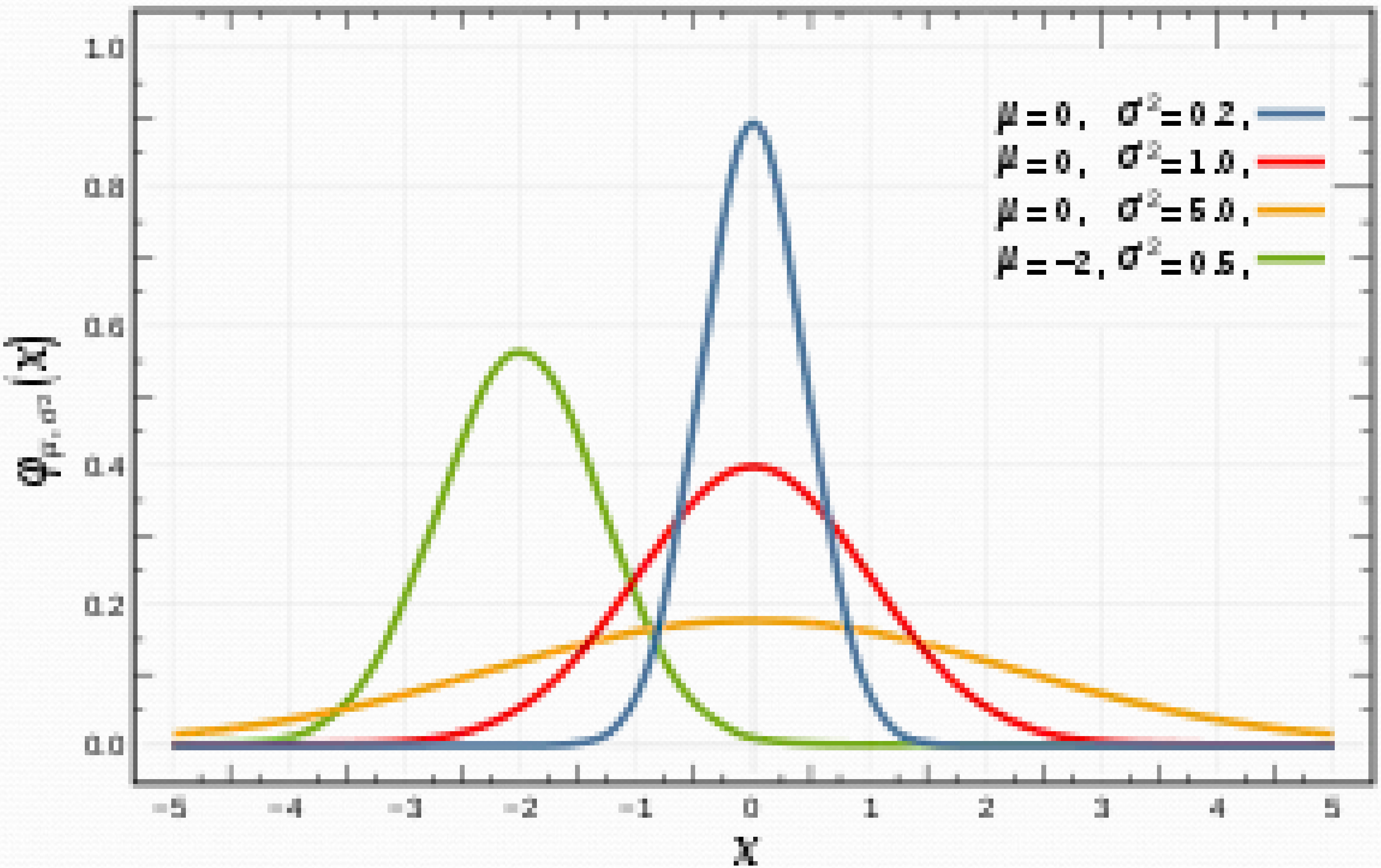
- Inúmeras varáveis contínuas que descrevem fenômenos naturais e sociais apresentam distribuições de probabilidades próximas da distribuição normal.

- O nome normal deve-se ao fato de que muitas distribuições de frequências de erros de observações e mensurações podem ser descritas por uma distribuição dessa natureza.

- Seja X uma variável aleatória que tome todos os pontos x tal que x ∈ R. A função densidade de probabilidade de uma variável X com distribuição normal é dada por:

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

- Uma notação frequentemente usada para denotar esta distribuição é N (μ,σ^2). Assim, dizer que a variável X tem distribuição N (2,9) significa que X tem distribuição normal com média $\mu = 2$ e variância $\sigma^2 = 9$.



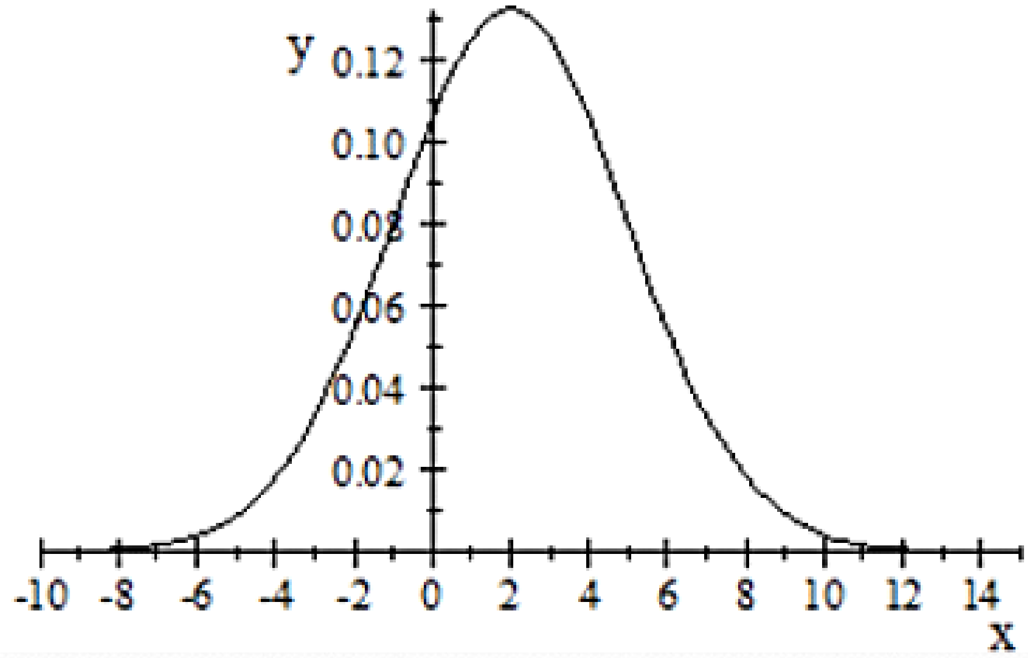
Função densidade de probabilidade de uma variável X

Exemplo: Suponhamos que a variável X tem distribuição N (2,9). Então, X tem média μ = 2, $\sigma^2 = 9$, onde o desvio padrão é σ = 3 e f (x) dada por:

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

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-2}{3}\right)^2}}{3\sqrt{2\pi}}$$

O gráfico da distribuição X será:



Distribuição normal N(2,9)

- A probabilidade de X ocorrer em determinado intervalo será dada pela área sob o gráfico.
- Essas probabilidades estão padronizadas.
- As probabilidades associadas à distribuição normal padronizada são encontradas em tabelas, não havendo necessidade de serem calculadas.

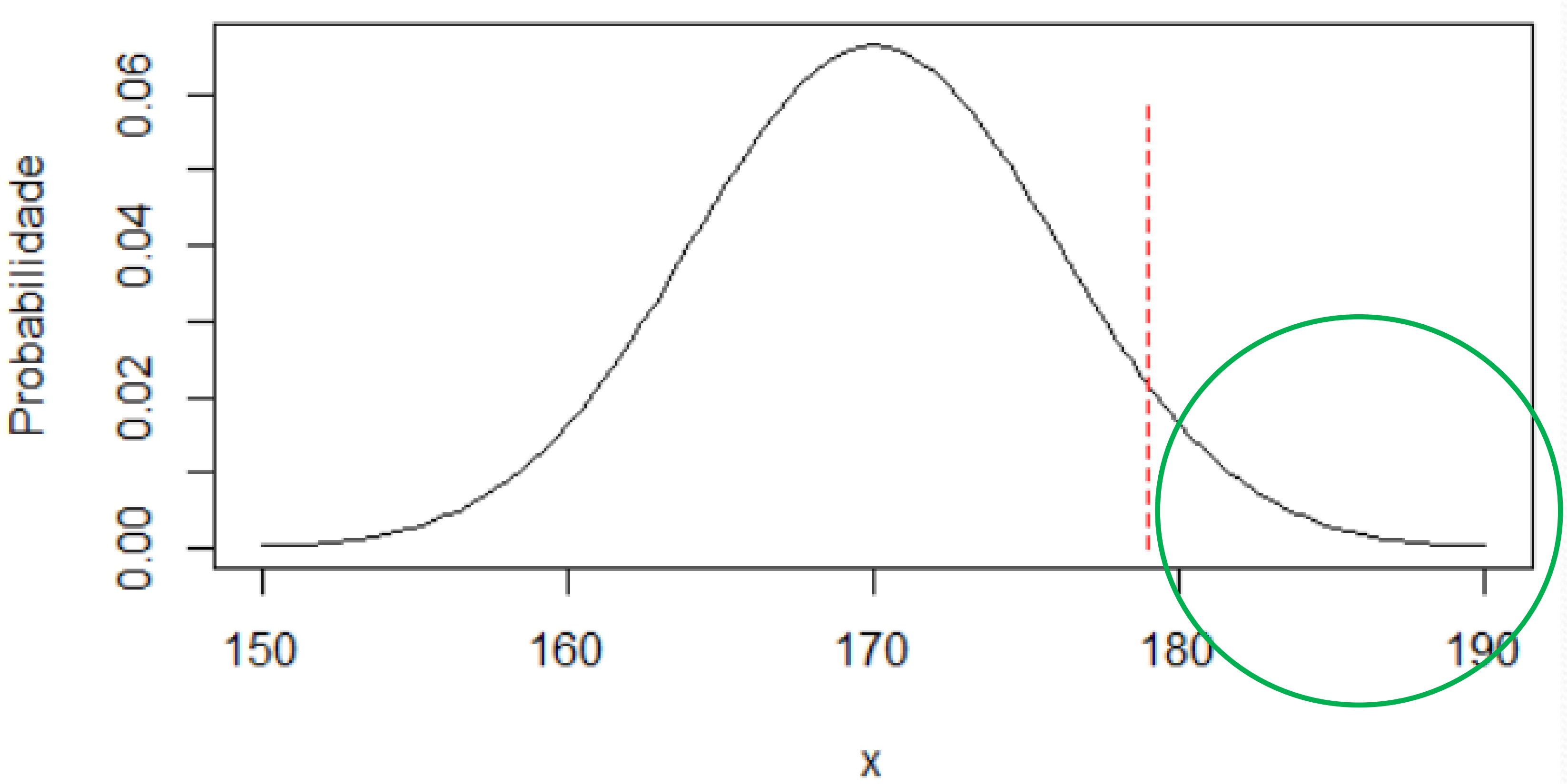
Exemplo:

Suponha que um pesquisador coletou dados de estatura de jovens em idade de alistamento militar.

Sabe-se que a estatura de certa população segue a distribuição normal, com média 170cm de variância de 36cm². Pede-se;

a) Qual a probabilidade de encontrarmos um jovem com mais de 179cm de altura?

```
> 1-pnorm(179,170,sqrt(36))
[1] 0.0668072
>
> curve(dnorm(x,170,sqrt(36)), #dist. normal
+       150,190,                #limites de x
+       ylab = "Probabilidade")
>       lines(c(179,179),       #linha do eixo x
+             c(0,0.06),        #linha do eixo y
+             col=2,            #linha vertical vermelha
+             lty=2)            #linha tracejada
```



4. Geração de números aleatórios

- O R pode gerar números aleatórios de várias formas.
- Pode-se gerar um número qualquer de um conjunto pré-estabelecido de valores ou atributos (baseado em sorteio pseudo aleatório),
- Ou quantos forem necessários de uma distribuição de interesse.

- Exemplo:

- Para simular o lançamento de um dado 100 vezes, podemos utilizar o comando sample( ), em que o primeiro argumento são quais valores podem ser assumidos, depois quantas vezes queremos jogar o dado.

```
> set.seed(128)       #fixando a semente
> x<-c(1,2,3,4,5,6) #espaço amostral
> amostra<-sample(x, #sortear os valores de x
+                   100, #sortear 100 vezes
+                   re=T) #com reposição
> amostra
  [1] 5 6 5 6 6 6 3 2 6 1 5 2 2 2 4 1 4 3 6 4 4 1 5 1 1 1
 [27] 5 6 5 3 2 4 5 3 5 1 2 6 1 3 4 1 1 5 3 5 4 5 5 3 1 2
 [53] 6 4 2 4 5 2 6 4 1 2 1 4 1 2 6 2 1 4 2 1 3 1 2 5 4 6
 [79] 6 2 6 3 4 4 3 6 6 1 6 5 4 6 6 4 4 2 4 4 2 5
> table(amostra)
amostra
 1  2  3  4  5  6 
18 17 10 20 16 19 
```

5. Exercícios