Estatística com Apoio Computacional

Distribuições de probabilidade

Universidade Estadual Vale do Acaraú – UVA

Paulo Regis Menezes Sousa
paulo_regis@uvanet.br

- Em qualquer situação em que ocorrem diversos resultados, a **teoria da probabilidade** oferece métodos de **quantificação das chances** ou possibilidades de ocorrência associadas aos diversos resultados.
- Muitos livros se dedicam exclusivamente à probabilidade, mas **nosso objetivo** é cobrir apenas a parte da disciplina que tem maior ligação com os problemas de inferência estatística.

Experimento aleatório É qualquer experimento, ou processo, cujo resultado está sujeito à incerteza.

Espaço amostral O espaço amostral de um experimento, representado por S, é o **conjunto** de todos os resultados possíveis desse experimento.

Evento Evento é qualquer grupo (**subconjunto**) de resultados contidos no espaço amostral S.

- Sabemos que um evento pode ocorrer ou não.
- Sendo $P(E)$ a probabilidade de que ele ocorra (**sucesso**) e que $\bar{P}(E)$ a probabilidade de que ele não ocorra (**insucesso**), para o mesmo evento existe sempre a relação:

$$P(E) + \bar{P}(E) = 1 \implies \bar{P}(E) = 1 - P(E) \tag{1}$$

Dizemos que dois eventos são **independentes** quando a realização de um não afeta a probabilidade da realização do outro e vice-versa.

- Se dois eventos são independentes, a probabilidade de que eles se realizem **simultaneamente** é igual ao produto das probabilidades de realização dos dois eventos.
- Sendo $P(E_1)$ a probabilidade de realização do primeiro evento e $P(E_2)$ a probabilidade de realização do segundo evento, a probabilidade de que tais eventos se realizem simultaneamente é dada por:

$$P = P(E_1) \times P(E_2) \tag{2}$$

Dizemos que dois ou eventos são **mutuamente exclusivos** quando a realização de um exclui a realização do(s) outro(s).

- Se dois eventos são mutuamente exclusivos, a probabilidade de que um **ou** outro se realize é igual à soma das probabilidades de que cada um deles se realize:

$$P = P(E_1) + P(E_2) \tag{3}$$

Warning!

A utilização dos recursos da estatística descritiva requer que o espaço amostral não-numérico de um determinado experimento aleatório seja transformado em um espaço amostral numérico.

- **Tal transformação é feita por uma função** que confere um número real a cada resultado do experimento aleatório. A esta função dá-se o nome de **variável aleatória**.
- Em geral, as variáveis aleatórias são representadas pela letra maiúscula X, e elas podem ser de dois tipos: discretas ou contínuas.

- Em alguns experimentos, como por exemplo, o número de bits transmitidos que são recebidos com erro, tem-se uma medida limitada a **números inteiros**.
- Ou ainda, quando se deseja registrar que 0,0042 de 1000 bits foram recebidos com erro, tem-se uma medida fracionária mais ainda assim limitada na linha dos **números reais**.

A variável que representa uma medida com uma faixa finita de valores ou infinita contável é denominada **variável aleatória discreta** ou **VAD**.

- Como exemplo de variáveis aleatórias discretas, tem-se a quantidade de partes defeituosas em n testadas, número de bits transmitidos com erro, ou seja, tudo que estiver relacionado com a contagem de determinados elementos.

- Um experimento binomial diz respeito a um experimento aleatório que consiste em repetidas tentativas que apresentam apenas dois resultados possíveis (tentativas de Bernoulli) e possui as seguintes características:
  - As tentativas são independentes, ou seja, o resultado de uma não altera o resultado da outra;
  - Cada repetição do experimento admite apenas dois resultados: sucesso ou fracasso;
  - A probabilidade de sucesso (p), em cada tentativa, é constante.

- A variável aleatória X denota o número de tentativas que resultaram em sucesso e possui uma distribuição binomial com parâmetros **p** e **n** $= 1,2,3,\ldots$ A função de probabilidade de X é:

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, 2, \ldots, n. \tag{4}$$

onde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5}$$

- **n** - número de tentativas; **p** - probabilidade de sucesso; **x** - número de sucessos;

- A esperança (média) e a variância são dadas por:

$$\mu = E(X) = np \tag{6}$$

$$\sigma^2 = V(x) = np(1-p) \tag{7}$$

- O nome binomial é devido ao fato de o resultado apresentar duas possibilidades.

Exemplo 1

Uma amostra de ar tem 10% de chance de conter certa molécula rara. Considere que as amostras sejam independentes com relação à presença da molécula rara. Encontre a probabilidade de que nas próximas 18 amostras, exatamente 2 contenham a molécula rara.

Solução: x é o número de amostras de ar que contenham a molécula rara nas próximas 18 amostras analisadas.

$p = 10\% = 0,1$

$n = 18$

$x = 2$

```
#Exemplo 1:
> p <- 0.1 #probabilidade
> n <- 18 #número de amostras
> x <- 2
#número de sucessos em 18 amostras
> dbinom(x, n, p)
[1] 0.2835121
```

Exemplo 2

A probabilidade de uma peça artesanal ser feita com perfeição por um artesão é de 50%. Considerando que o artesão produz, de maneira independente, 6 peças por dia, pede-se:

a) Obter a distribuição de probabilidades, ou seja, as probabilidades associadas aos possíveis valores da variável aleatórias discreta x, em que x = número de peças perfeitas produzidas pelo artesão num único dia.
Observação: $x = \{0, 1, 2, 3, 4, 5, 6\}, n = 6, p = 0.5$.

b) Plotar o gráfico com os valores da probabilidade calculada.

Solução:

```
#Exemplo 2:
> x <- 0:6
> n <- 6
> p <- 0.5
> bino <- dbinom(x, n, p)
> bino
[1] 0.015625 0.09375 0.234375 0.312500 0.234375 0.093750 0.015625
> plot(x,bino,type="h",xlab="Número de peças perfeitas",
+ ylab="Probabilidade",main="Distribuição binomial")
```

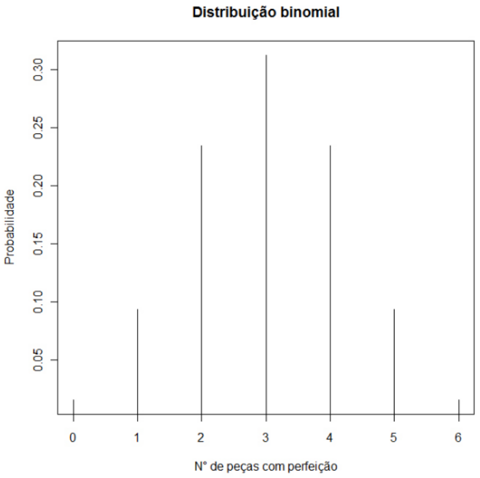

Figura: Gráfico do Exemplo 2.b: Distribuição binomial de probabilidades

- Em muitas situações nos deparamos com a situação em que o número de ensaios n é grande ($n \to \infty$) e p é pequeno ($p \to 0$), no cálculo da função binomial, o que nos leva a algumas dificuldades.
- Quando temos o número de observações de uma variável em um intervalo contínuo (tempo ou espaço). Para estes casos podemos usar a distribuição de Poisson.
- Pressupostos:
  - Os números de ocorrências em quaisquer intervalos são independentes.
  - A probabilidade de duas ou mais ocorrências simultâneas é zero.
  - O número médio de ocorrências λ é constante em todo o intervalo considerado.
- Exemplos:
  - Erros de digitação por um certo período de tempo.
  - Número de falhas em componentes por unidade de tempo.
  - Número de acidentes registrados durante um mês na BR 222.
  - Clientes chegando ao caixa de um supermercado.

- Assim a variável aleatória X que segue uma distribuição de Poisson com parâmetro λ, $\lambda > 0$ é representada através da função:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \;\; x = 0, 1, 2, \ldots, n. \tag{8}$$

- A média e a variância de uma distribuição de Poisson são representadas por:

$$\mu = E(X) = \lambda \tag{9}$$

$$\sigma^2 = V(X) = \lambda \tag{10}$$

Exemplo 3

Em um fio delgado de cobre, o número de falhas no fio segue a distribuição de Poisson, com uma média de 2,3 falhas por milímetro.

a) Determine a probabilidade de existir exatamente 2 falhas em um milímetro de fio.

b) Sabendo que o número máximo de erros no teste de qualidade é de 10 erros/mm, verifique as probabilidades de que ocorram de 0 a 10 falhas no fio. Plote o gráfico da distribuição.

Solução (a): X representa o número de falhas em 1 milímetro de fio:
$E(X) = \mu = \lambda = 2,3$(falhas)/mm.

```
#Exemplo 3a:
> x<-2
> lambda<-2.3
> #distribuição de Poisson com parâmetros x e lambda:
> dpois(x,lambda)
[1] 0.2651846
```

Solução (b): Fazendo x variar de 0 a 10 erros/mm, temos:

```
#Exemplo 3b:
> x <- 0:10
> poisson <- dpois(x, lambda)
> plot(x,poisson, xlab= "Número de erros por milímetro",
+ ylab="Probabilidade de Poisson",main="Distribuição de Poisson")
> lines(x,poisson)
```

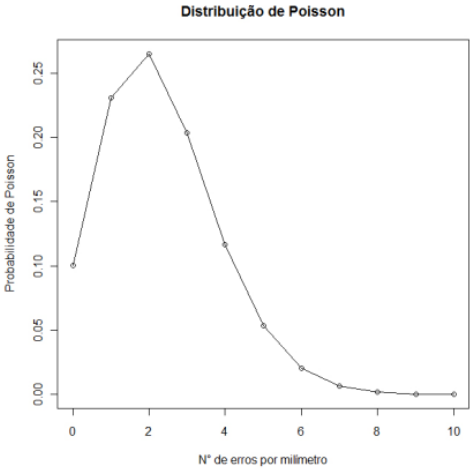

Figura: Gráfico do Exemplo 3.b: Distribuição Poisson de probabilidades

- A corrente elétrica em um fio condutor ou o comprimento de uma peça são exemplos de experimentos aleatórios que apresentam a medida de interesse como um número real.

> A variável aleatória que representa uma medida, com um intervalo finito ou infinito de números reais para sua faixa denomina-se **variável aleatória contínua** ou **VAC**.

- Como exemplos de variáveis aleatórias contínuas têm-se: a corrente elétrica, comprimento, pressão, temperatura, tempo, tensão, peso.

- É um dos mais importantes modelos de probabilidade para VACs. Aplicado em inúmeros fenômenos e muito utilizado no desenvolvimento teórico e na área da inferência estatística.
- O teorema central do limite explica esta propriedade, ele diz que: "*Toda a soma de variáveis aleatórias independentes de média finita e variância limitada é aproximadamente Normal, desde que o número de termos da soma seja suficientemente grande*".
- As variáveis aleatórias com diferentes médias e variâncias podem ser modeladas pelas funções densidade de probabilidade normal. $E(X) = \mu$ determina o centro do gráfico em forma de sino, e $V(X) = \sigma^2$ determina a largura da distribuição

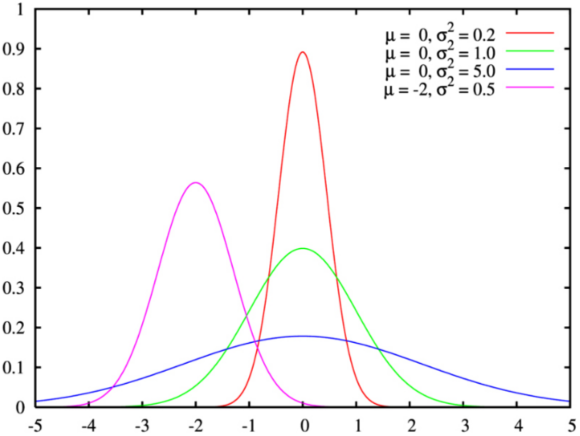

Figura: Distribuição Normal com diferentes parâmetros.

- A representação da função densidade de probabilidade é dada por:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\sigma)^2}{2\sigma^2}} \tag{11}$$

Em que:

$$-\infty < x < \infty \tag{12}$$

$$\sigma > 0 \tag{13}$$

$$E(X) = \mu \tag{14}$$

$$V(X) = \sigma^2 \tag{15}$$

Exemplo 4

Um pesquisador coletou os dados da estatura de jovens em idade de alistamento militar. Sabe-se que a estatura de uma população segue a distribuição normal, com média 170cm e variância 36cm^2 (desvio padrão de 6 cm^2).

a) Qual a probabilidade de se encontrar um jovem com mais de 1,79m de altura?

b) Qual a altura em que a probabilidade de encontrarmos valores menores que ela seja de 80%?

c) Represente graficamente a curva da distribuição normal para este problema e identifique as respostas dos itens **a** e **b**.

- A cada variável aleatória discreta ou contínua estão associadas duas funções: a função de probabilidade e a função de distribuição acumulada.

A **função de probabilidade** p correspondente à variável aleatória discreta X associa a cada número real x a probabilidade de que a variável X assuma aquele valor x (Ou seja, $x \to p(x) = P[X = x]$).

A **função de distribuição acumulada** F correspondente à variável aleatória discreta X é definida por $F(x) = P[X \leq x]$, para todo x real.

- Sendo um ambiente para análise de dados, o R dispõe de um grande conjunto de funções para trabalhar com *Distribuições Estatísticas*.
- Há quatro funções para se trabalhar com a distribuição normal no R:
  - **dnorm** – retorna a densidade probabilística para um dado valor da variável;
  - **pnorm** – retorna a probabilidade acumulada para um dado valor da variável;
  - **qnorm** – retorna o quantil para um dado valor de probabilidade acumulada;
  - **rnorm** – retorna valores (números aleatórios) gerados a partir da distribuição;

Solução

```
#Exemplo 4:
> #Item a)
> 1-pnorm(179,170,6) #pnorm (função de probabilidade)
[1] 0.0668072
> #Item b)
> qnorm(0.8,170,6) #qnorm (função de distrib. probabilidade)
[1] 175.0497
> #Item c)
> curve(dnorm(x,170,6),170-3*6,170+3*6,xlab="Alturas (cm)",
+ ylab="Probabilidade de se encontrar a altura x",
+ main="Distribuição Normal")
> lines(c(179,179),c(0,0.022),col="red")
> lines(c(175.0497,175.0497),c(0,0.0465),col="blue")
```

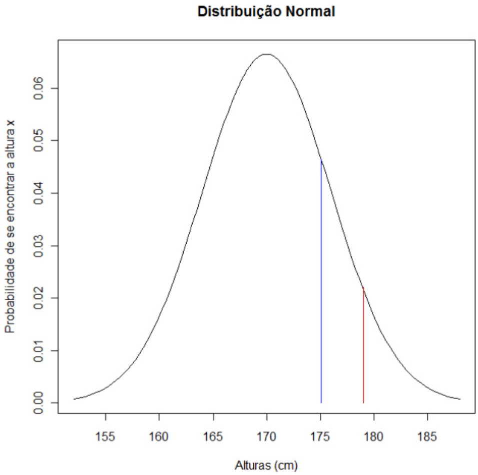

Figura: Distribuição normal do Exemplo 4.c