CE301 - Estatística Básica

Introdução às variáveis aleatórias (parte 2/2)

Paulo Justiniano Ribeiro Jr

Departamento de Estatística
Setor de Ciências Exatas
Universidade Federal do Paraná

9 de maio de 2026

Exercício 9

Alguns biólogos fizeram estudos de laboratório sobre o comportamento de animais quando submetidos a um estímulo, o quais poderiam ou não apresentar resposta (sensíveis). Considera-se que na população destes animais, 10% sejam sensíveis ao estímulo.

- **O biólogo** A possuia um grupo em que 10 animais eram sensíveis e 20 eram insensíveis. Ele selecionou ao acaso 8 animais para teste. *Qual a probabilidade encontrar ao menos 2 animais sensíveis?*
- **O biólogo** B submeteu 10 animais ao estímulo e registrou quantos eram sensíveis. *Qual a probabilidade de encontrar mais que 3 animais sensíveis?*
- **O biólogo** C dispunha de um grande número de animais e foi testando um a um até encontrar um sensível ao estímulo. *Qual a probabilidade de precisar testar mais que 3 animais?*
- **O biólogo** D dispunha de um grande número de animais e foi testando um a um até encontrar o terceiro sensível ao estímulo. *Qual a probabilidade de precisar testar no máximo 6 animais?*
- **O biólogo** E fez testes diários e encontrou uma média de 2,8 animais sensíveis a cada dia. *Qual a probabilidade de encontrar menos que dois sensíveis em um dia?*

Exercício 8

Distribuições discretas mais usuais:

- Hipergeométrica
- Bernoulli e Binomial
- Geométrica
- Binomial negativa (Pascal)
- Poisson

Biólogo A: Distribuição Hipergeométrica

O biólogo A possuia um grupo em que 10 animais eram sensíveis e 20 eram insensíveis. Ele selecionou ao acaso 8 animais para teste.
Qual a probabilidade do biólogo A encontrar ao menos 2 animais sensíveis?

Y : número de sensíveis entre oito testados.

$Y \sim \text{HG}(m, n, k)$

$y \in \{0, 1, 2, \ldots 8\}$

$$P[Y = y] = \frac{\binom{m}{y}\binom{n}{k-y}}{\binom{m+n}{k}} = \frac{\binom{10}{y}\binom{20}{8-y}}{\binom{30}{8}}$$

$$P[Y \geq 2] = 1 - P[Y = 0] - P[Y = 1] = 1 - \frac{\binom{10}{0}\binom{20}{8-0}}{\binom{30}{8}} - \frac{\binom{10}{1}\binom{20}{8-1}}{\binom{30}{8}} = 0.8460$$

Código **R**:

```
1 - phyper(1, m=10, n=20, k=8)
## [1] 0.8460308
```

Biólogo B: Distribuição Binomial

O biólogo B submeteu 10 animais ao estímulo e registrou quantos eram sensíveis..
Qual a probabilidade de encontrar mais que 3 animais sensíveis?

Y : número de sensíveis entre 10 testados.

$Y \sim \mathrm{B}(n, p)$

$y \in \{0, 1, 2, \ldots 10\}$

$$P[Y = y] = \binom{n}{y} p^y (1-p)^{n-y} = \binom{10}{y} (0.1)^y (0.9)^{10-y}$$

$$P[Y > 3] = 1 - P[Y \leq 3] = 1 - P[Y = 0] - P[Y = 1] - P[Y = 2] - P[Y = 3] = 0.0128$$

Código **R**:

```
1 - pbinom(3, size=10, prob=0.1)
## [1] 0.0127952
```

Biólogo C: Distribuição Geométrica

O biólogo C dispunha de um grande número de animais e foi testando um a um até encontrar um sensível ao estímulo.
Qual a probabilidade de precisar testar mais que 3 animais?

Y : número de não sensíveis testados até encontrar o primeiro sensível.

$$Y \sim \mathrm{G}(p)$$

$$y \in \{0, 1, 2, \ldots\}$$

$$P[Y = y] = (1 - p)^y p$$

$$P[Y \geq 3] = 1 - P[Y = 0] - P[Y = 1] - P[Y = 2] = 0.7290$$

Código **R**:

```
1 - pgeom(2, prob=0.1)
## [1] 0.729
```

Biólogo D: Distribuição binomial negativa

O biólogo D dispunha de um grande número de animais e foi testando um a um até encontrar o terceiro sensível ao estímulo.
Qual a probabilidade de precisar testar no máximo 6 animais?

Y : número de não sensíveis encontrados até encontrar o terceiro sensível.

$Y \sim \text{BN}(k, p)$

$y \in \{0, 1, 2, \ldots\}$

$$P[Y = y] = \binom{y+k-1}{k-1}(1-p)^y p^k = \binom{y+3-1}{3-1}(0.9)^y(0.1)^3$$

$$P[Y \leq 3] = P[Y = 0] + P[Y = 1] + P[Y = 2] + P[Y = 3] = 0.0158$$

Código **R**:

```
pnbinom(3, size=3, prob=0.1)
## [1] 0.01585
```

Biólogo E: Distribuição Poisson

O biólogo E fazia testes diários e encontrava uma média de 2,8 animais sensíveis a cada dia.
Qual a probabilidade de encontrar menos que dois sensíveis em um determinado dia?

Y : número de sensíveis encontrados por dia.

$Y \sim \mathrm{P}(\lambda)$

$y \in \{0, 1, 2, \ldots\}$

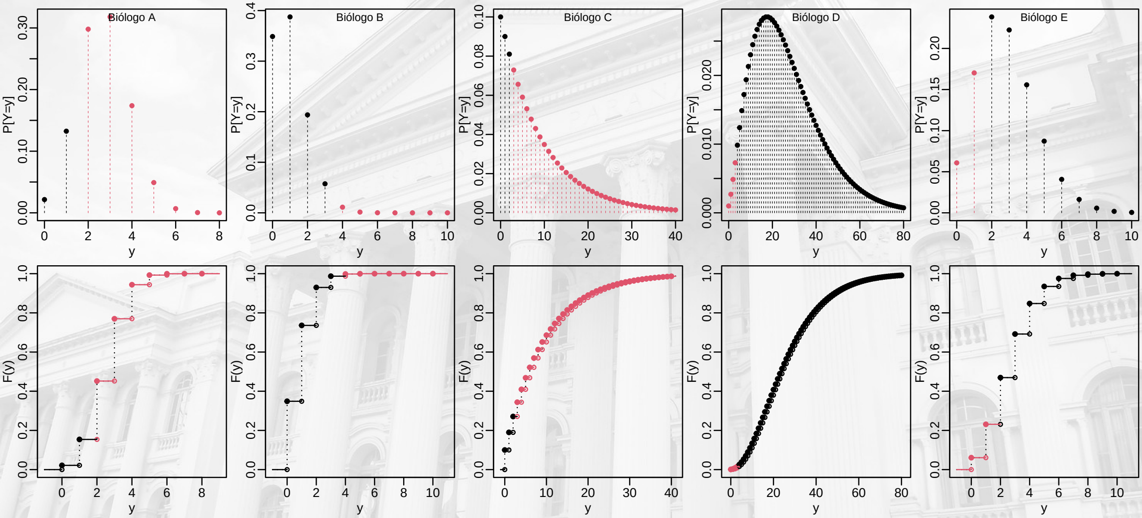
$$P[Y = y] = \frac{e^{-\lambda}\lambda^{y}}{y!} = \frac{e^{-2.8}2.8^{y}}{y!}$$

$$P[Y < 2] = P[Y = 0] + P[Y = 1] = \frac{e^{-2.8}2.8^{0}}{0!} + \frac{e^{-2.8}2.8^{1}}{1!} = 0.2311$$

Código **R**:

```
ppois(1, lambda=2.8)
## [1] 0.2310782
```

Modelos para os biólogos



Atenção: definição da variável

Geométrica:

Y: número de "falhas" até primeiro sucesso $P[Y = y] = (1-p)^y p, \text{ para } y = 0, 1, \ldots$

ou

Y: número de tentativas até primeiro sucesso $P[Y = y] = (1-p)^{y-1} p, \text{ para } y = 1, 2, \ldots$

Binomial negativa:

Y: número de "falhas" até o k-ésimo sucesso ...

ou

Y: número de tentativas até o k-ésimo sucesso ...

Modelos

Modelos de probabilidade e modelos estatísticos.

- Escolha de modelo.
- Constantes e parâmetros.
- Famílias e obtenção do(s) parâmetro(s).
- Modelagem estatística, regressão e generalizações.

Exercício 10

Considere o surgimento de um defeito na pista em um trecho de rodovia com extensão de 20 km.

1. Qual a probabilidade de que o defeito ocorra nos primeiros 5 km?
2. Qual a probabilidade de que o defeito ocorra entre os quilômetros 12 e 15?
3. Se o defeito ocorre na segunda metade do trecho, qual a probabilidade de seja nos últimos 3 km?
4. O custo de manutenção é de R$2.000,00 se ocorre nos primeiros 5 km, de R$5.000,00 se ocorre entre 5 e 16 kms e de R$10.000,00 se ocorre nos ultimos 4 km. Que custo espera-se ter em 20 manutenções?

Exercício 10

Considere o surgimento de um defeito na pista em um trecho de rodovia com extensão de 20 km.

1. Qual a probabilidade de que o defeito ocorra nos primeiros 5 km?
   Resposta: 1/4
2. Qual a probabilidade de que o defeito ocorra entre os quilômetros 12 e 15?
   Resposta: 3/20
3. Se o defeito ocorre na segunda metade do trecho, qual a probabilidade de seja nos últimos 3 km?
   Resposta: 3/10
4. O custo de manutenção é de R$2.000,00 se ocorre nos primeiros 5 km, de R$5.000,00 se ocorre entre 5 e 16 kms e de R$10.000,00 se ocorre nos ultimos 4 km. Que custo espera-se ter em 20 manutenções?
   Resposta: $\frac{5}{20}2000 + \frac{11}{20}5000 + \frac{4}{20}10000 = 5250$

Exercício 10

O que mudou?

Domínio de Y nos reais (ou subconjunto dos reais), no exemplo $y \in [0, 20]$

- a v.a. é contínua,
- não faz mais sentido em avaliar o valor em um ponto $P[Y = y]$
- Define-se uma função de densidade de probabilidade (fdp) $f(y)$
  - $f(y) \geq 0 \;\; \forall y$
  - $\int f(y) \mathrm{d}y = 1$
- E as probabilidades de interesse são áreas sob a função

$$P[a < Y < b] = \int_a^b f(y) \mathrm{d}y$$

Exercício 10 (cont)

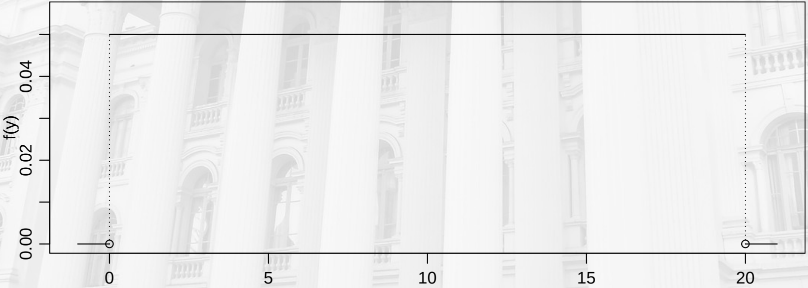
Considere o surgimento de um defeito na pista em um trecho de rodovia com extensão de 20 km.

$$Y : \text{local onde ocorre o defeito}$$

$$y \in [0, 20]$$

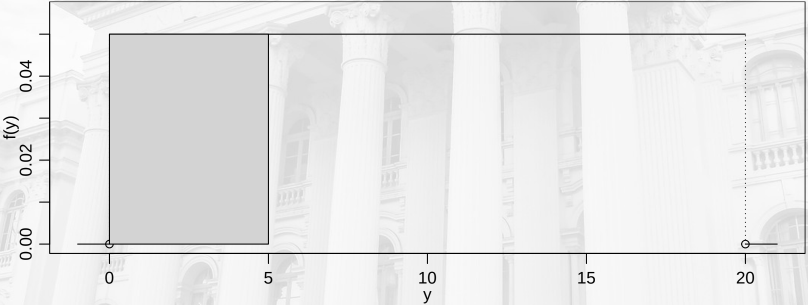
$$Y \sim U_c[0, 20]$$

$$f(y) = \frac{1}{20}$$



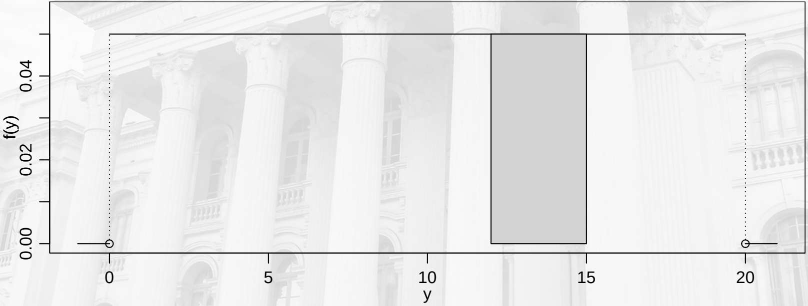
Exercício 10 (cont)

1. Qual a probabilidade de que o defeito ocorra nos primeiros 5 km?
   Resposta: 1/4 = 0,25



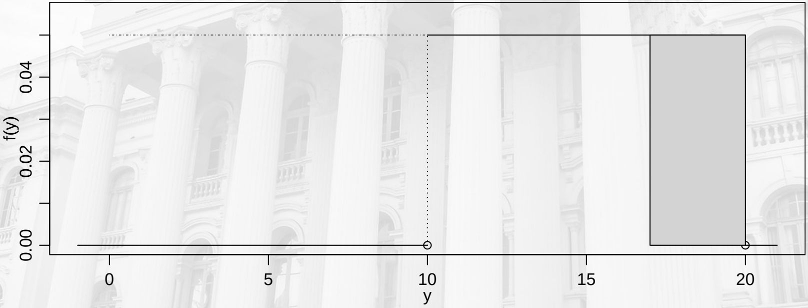
Exercício 10 (cont)

2. Qual a probabilidade de que o defeito ocorra entre os quilômetros 12 e 15?
   Resposta: 3/20 = 0,15



Exercício 10 (cont)

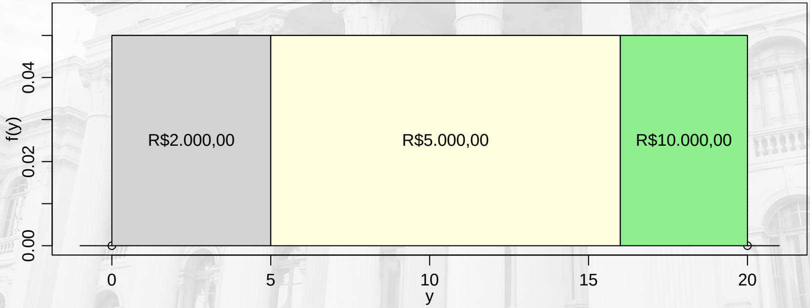
3 Se o defeito ocorre na segunda metade do trecho, qual a probabilidade de seja nos últimos 3 km?
Resposta: $\frac{3}{10}$



Exercício 10 (cont)

4. O custo de manutenção é de R$2.000,00 se ocorre nos primeiros 5 km, de R$5.000,00 se ocorre entre 5 e 16 kms e de R$10.000,00 se ocorre nos ultimos 4 km. Que custo espera-se ter em 20 manutenções?

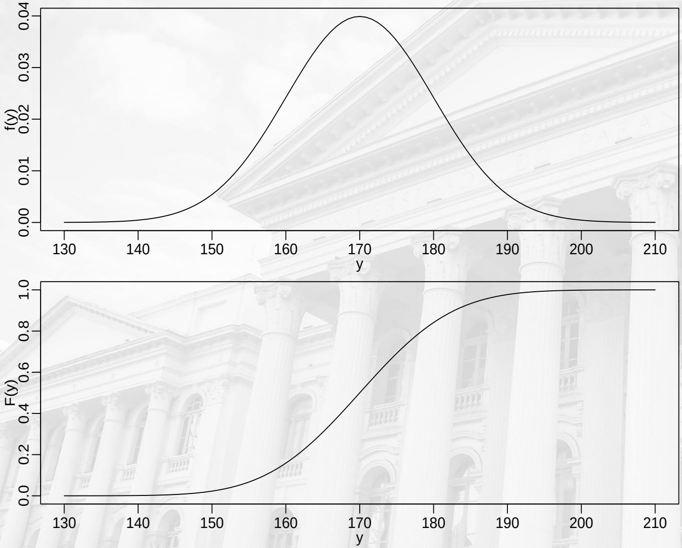
   Resposta: $\frac{5}{20}2000 + \frac{11}{20}5000 + \frac{4}{20}10000 = 5250$



Outras distribuições

Mas, …"o mundo" não é uniforme!

Uma distribuição especial, ou melhor, **normal**



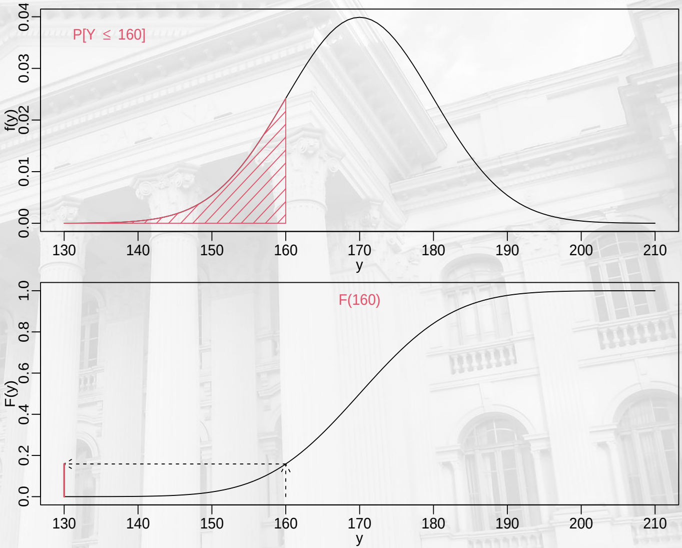
$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1. $P[Y \leq 160]$
2. $P[Y > 180]$
3. $P[155 \leq Y \leq 175]$
4. $P[Y \leq a] = 0{,}99$
5. $P[\mu - b \leq Y \leq \mu + b] = 0{,}80$
6. $P[Y \leq 160 | Y \leq 170]$
7. $P[Y > 180 | Y > 165]$
8. $P[Y \leq 185 | Y > 155]$
9. $P[Y > 175 | Y \leq 190]$

Probabilidade

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1. $P[Y \leq 160]$
2.
3.
4.
5.
6.
7.
8.
9.



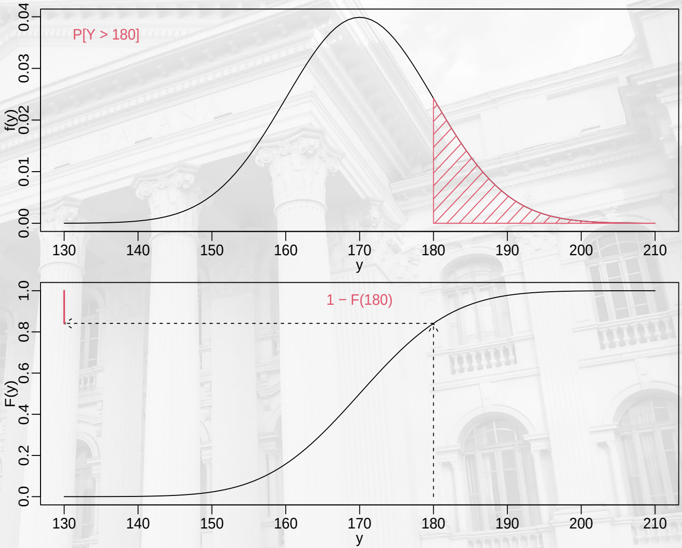
Código **R**:

```
pnorm(160, mean = 170, sd = 10)
## [1] 0.1586553
```

Probabilidade

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2. $P[Y > 180]$
3.
4.
5.
6.
7.
8.
9.



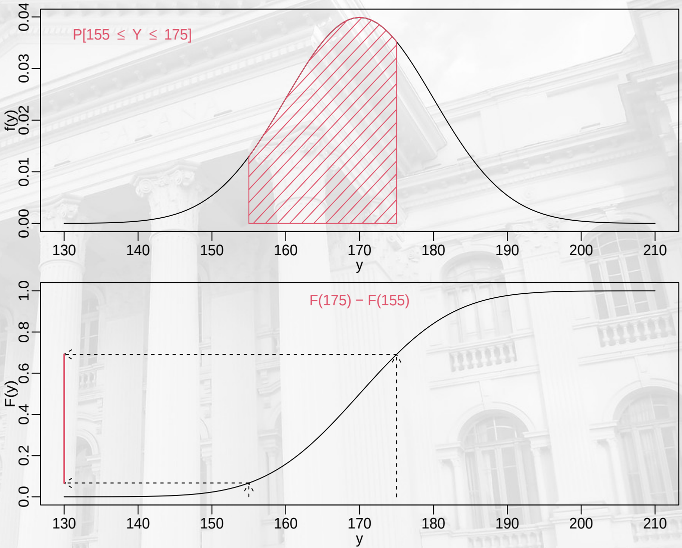
Código **R**:

```
1 - pnorm(180, 170, 10)
## [1] 0.1586553
```

Probabilidade

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3. $P[155 \leq Y \leq 175]$
4.
5.
6.
7.
8.
9.



Código **R**:

```
pnorm(175, mean = 170, sd = 10) -
    pnorm(155, mean = 170, sd = 10)
## [1] 0.6246553
```

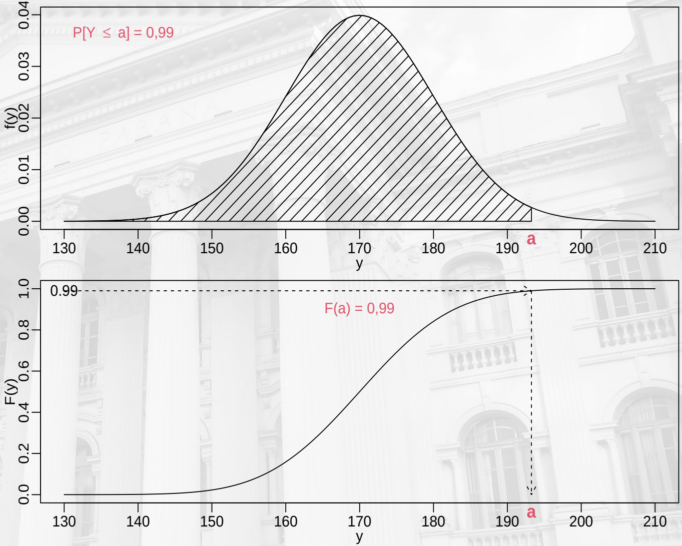
Quantis

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4. $P[Y \leq a] = 0{,}99$
5.
6.
7.
8.
9.

Código **R**:

```
qnorm(0.99, mean = 170, sd = 10)
## [1] 193.2635
```



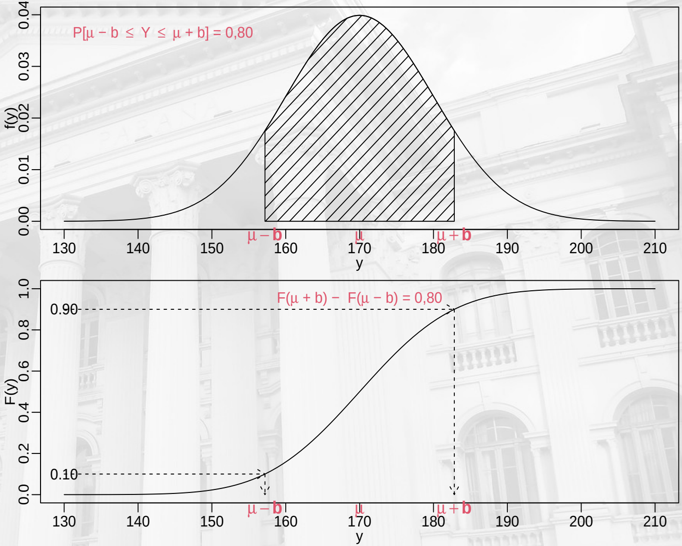
Quantis

$$Y \sim \text{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4.
5. $P[\mu - b \leq Y \leq \mu + b] = 0{,}80$
6.
7.
8.
9.

Código **R**:

```
qnorm(0.90, 170, 10) - 170
## [1] 12.81552
```



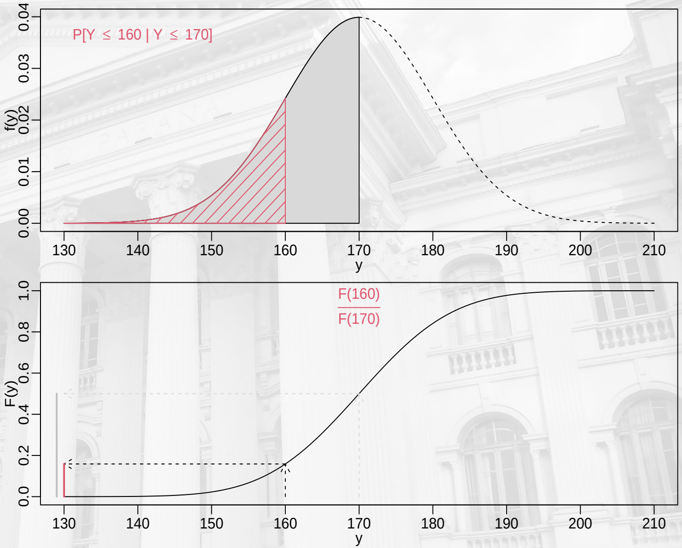
Probabilidade condicional

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4.
5.
6. $P[Y \leq 160 | Y \leq 170]$
7.
8.
9.

Código **R**:

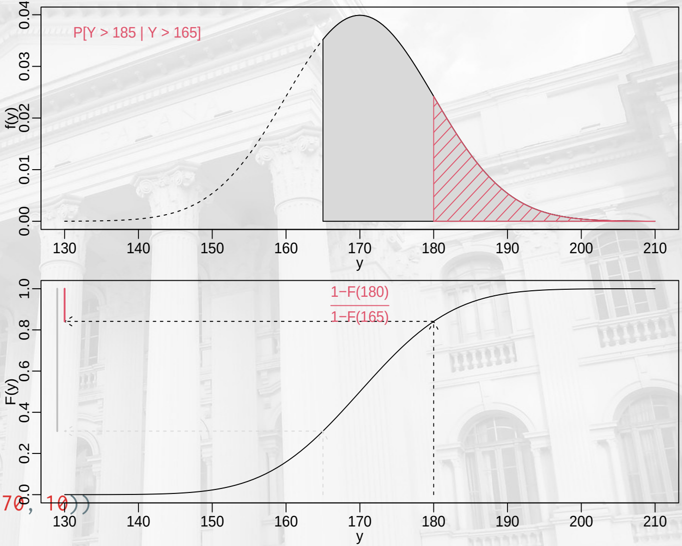
```
pnorm(160, 170, 10)/pnorm(170, 170, 10)
## [1] 0.3173105
```



Probabilidade condicional

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4.
5.
6.
7. $P[Y > 180 | Y > 165]$
8.
9.



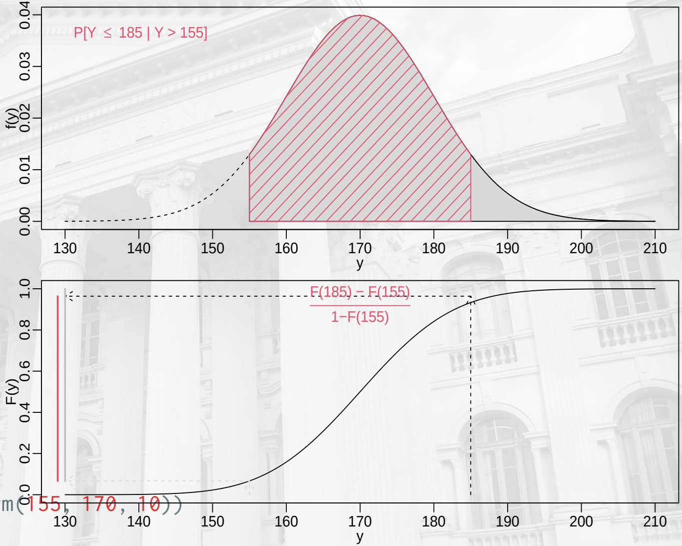
Código **R**:

```
(1 - pnorm(180, 170, 10))/(1 - pnorm(165, 170, 10))
## [1] 0.2294488
```

Probabilidade condicional

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4.
5.
6.
7.
8. $P[Y \leq 185 | Y > 155]$
9.



Código **R**:

```
diff(pnorm(c(155,185), 170, 10))/(1 - pnorm(155, 170, 10))
## [1] 0.9284101
```

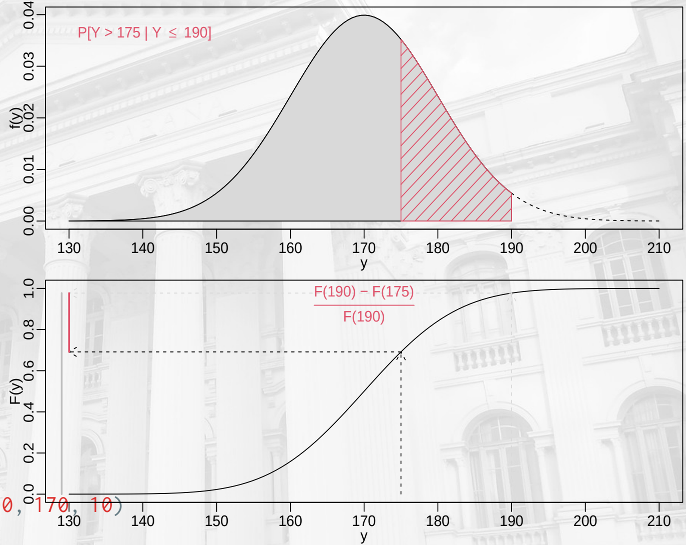
Probabilidade condicional

$$Y \sim \mathrm{N}(\mu = 170, \sigma^2 = 10^2)$$

1.
2.
3.
4.
5.
6.
7.
8.
9. $P[Y > 175 | Y \leq 190]$

Código **R**:

```
diff(pnorm(c(175,190), 170, 10))/pnorm(190, 170, 10)
## [1] 0.2924405
```



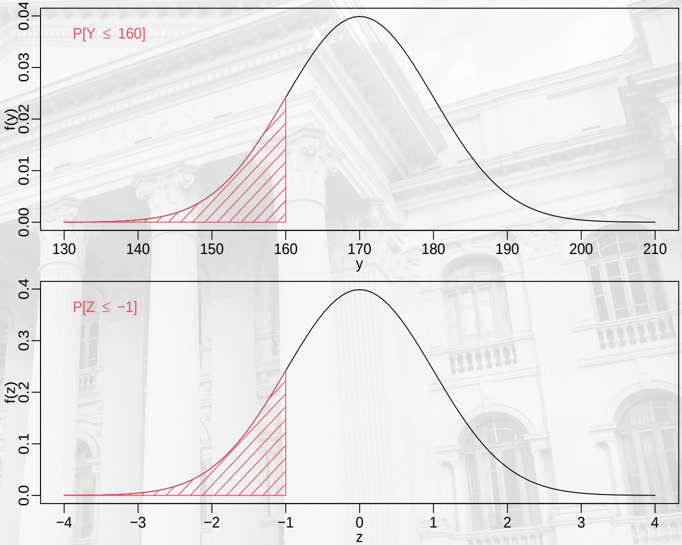
Escore Z e probabilidade

$$Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$$

$$Z = \frac{Y - \mu}{\sigma} \sim \mathrm{N}(\mu_z = 0, \sigma_z^2 = 1)$$

- Z desvios padrão acima (ou abaixo) da média.
- Probabilidades equivalentes em Y ou Z.
- Usado para cálculo de probabilidades utilizando "tabelas da distribuição normal".
- No exemplo:

$$P[Y \leq 160] = P[Z \leq \frac{160 - 170}{10}] = P[Z \leq -1]$$



Escore Z e quantil

$$Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$$

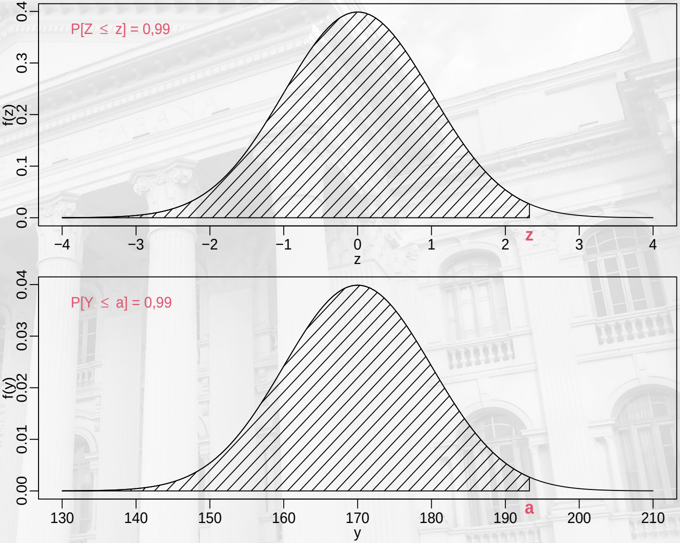
$$Z = \frac{Y - \mu}{\sigma} \sim \mathrm{N}(\mu_z = 0, \sigma_z^2 = 1)$$

$$P[Z \leq z] = 0{,}99$$

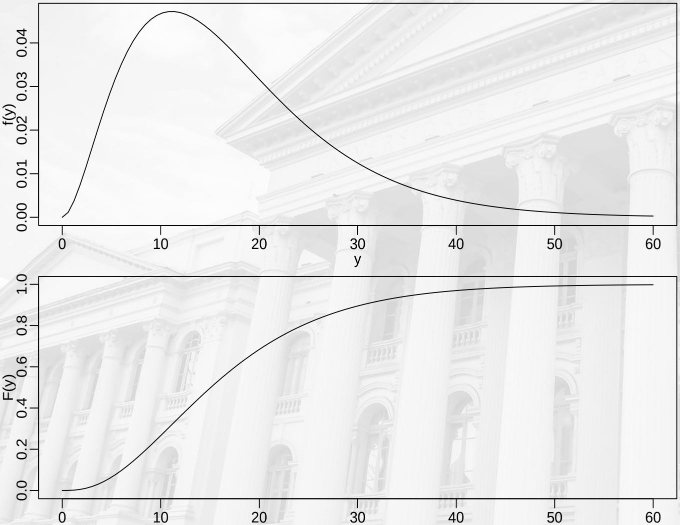
$$z = 2.326$$

$$\frac{a - 170}{10} = 2.326$$

$$a = 193.263$$



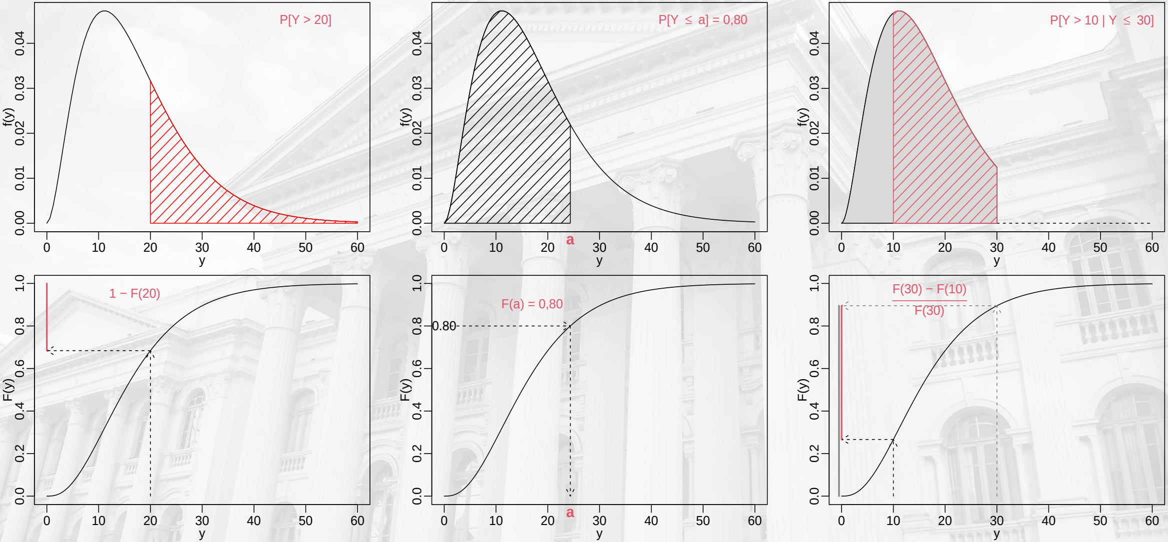
Uma outra distribuição de variável contínua: **gamma**



$$Y \sim G(\mu = 170, \sigma^2 = 10^2)$$
$$Y \sim G(\alpha = 2{,}89, \beta = 0{,}10)$$

1. $P[Y > 20]$
2. $P[Y \leq a] = 0{,}80$
3. $P[Y > 10 | Y \leq 30]$

Distribuição Gama



Distribuição Gama

$\mu = 17$: média e $\sigma^2 = 100$: variância

Parametrização utilizada: "shape" (forma) e "rate" (taxa)

$$\text{(shape)} = \alpha = \frac{\mu^2}{\sigma^2} = \frac{17^2}{100} = 2.89$$

$$\text{(rate)} = \beta = \frac{\mu}{\sigma^2} = \frac{17}{100} = 0.17$$

Códigos **R**:

```
pgamma(20, sh = (17/10)^2, rate = 17/10^2, lower = FALSE)
## [1] 0.3160481
qgamma(0.80, sh = (17/10)^2, rate = 17/10^2)
## [1] 24.35641
diff(pgamma(c(10, 30), sh = (17/10)^2, rate = 17/10^2))/pgamma(30, sh = (17/10)^2, rate=17/10^2)
## [1] 0.7024844
```

Comparando v.a.'s discretas e contínuas

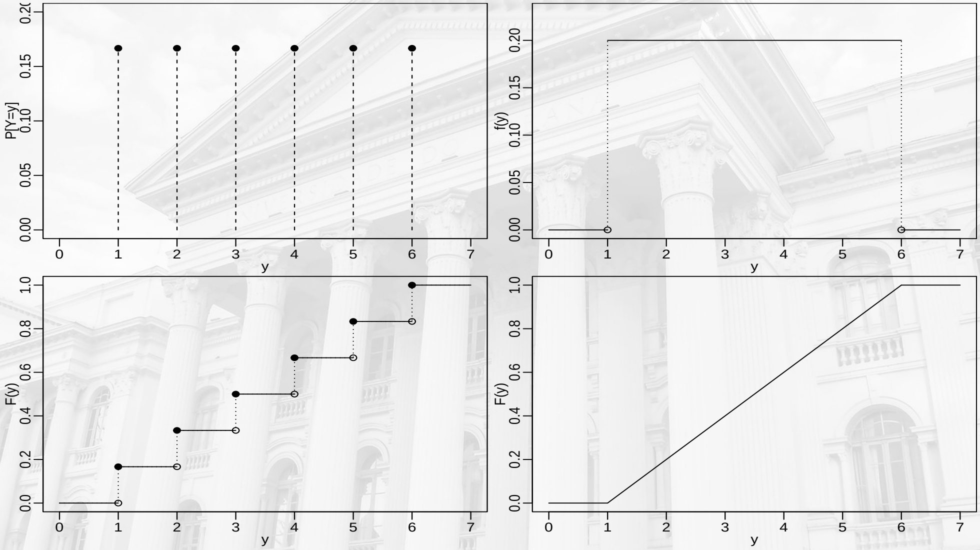
Uniforme discreta:

$$Y_1 \sim U_d\{1,6\}$$
$$y \in \{1,2,3,4,5,6\}$$
$$P[Y_1 = y] = \frac{1}{6}$$
$$F_{Y_1}(y) = \frac{\lfloor y \rfloor}{6}$$

Uniforme contínua:

$$Y_2 \sim U_c[1,6]$$
$$y \in [1,6]$$
$$f_{Y_2}(y) = \frac{1}{6-1} = \frac{1}{5}$$
$$F_{Y_2}(y) = \frac{y-1}{6-1} = \frac{y-1}{5}$$

V.A.'s (uniforme) discreta e contínua



Probabilidades para v.a.'s discretas e contínuas

Uniforme discreta:

$$Y_1 \sim U_d\{1,6\}$$

$$P[Y_1 = 4] = \frac{1}{6}$$

$$P[Y_1 \leq 4] = F_{y_1}(4) = P[Y_1 = 1] + P[Y_1 = 2] + P[Y_1 = 3] + P[Y_1 = 4] = \frac{4}{6} = 0{,}67$$

Uniforme contínua:

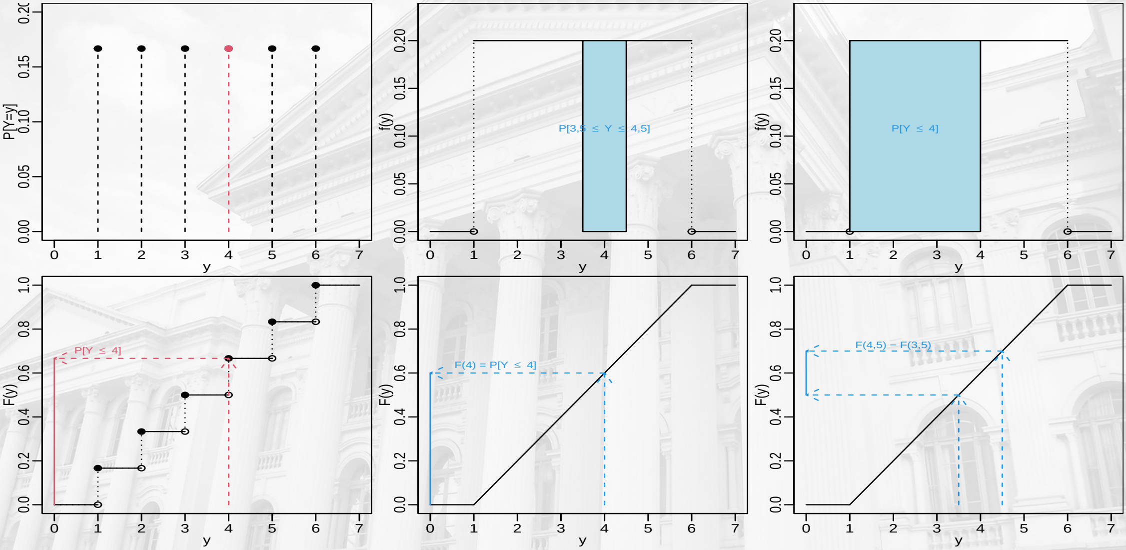
$$Y_2 \sim U_c[1,6]$$

$$P[Y_2 = 4] = f(4) = 0$$

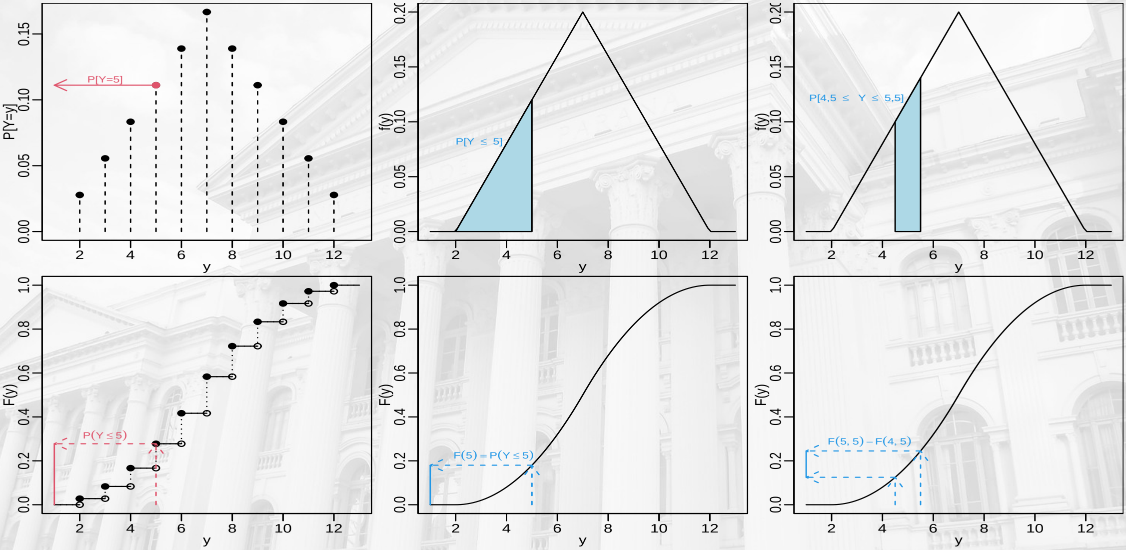
$$P[3{,}5 < Y_2 \leq 4{,}5] = \int_{3.5}^{4.5} f(y)\mathrm{d}y = 0.2$$

$$P[Y_2 \leq 4] = F_{y_2}(4) = \int_{1}^{4} f(y)\mathrm{d}y = 0.6$$

V.A.'s (uniforme) discreta e contínua



V.A.'s discretas e contínuas



Comentários finais

- Variáveis aleatórias são um tipo de "dispositivo" que nos ajuda a atribuir probabilidades a eventos de interesse.
- A distribuição de probabilidade define como a probabilidade “se espalha” entre os possíveis valores da v.a..
- Existem distribuições discretas e contínuas que acomodam diferentes comportamentos da variável de interesse.
- Aqui tratamos de distribuições univariadas, os conceitos se estendem para distribuições multivariadas.