

CE301 - Estatística Básica

Famílias de distribuições

Paulo Justiniano Ribeiro Jr

Departamento de Estatística
Setor de Ciências Exatas
Universidade Federal do Paraná

22 de maio de 2025



Como interpretar probabilidades?

- ▶ Informação prévia/disponível. *Definição clássica.*
- ▶ Obtida por observação/simulação. *Definição frequentista* (estimativa, limite de proporção)
- ▶ Atribuição “opinativa”. *Definição subjetiva.*

Em outras palavras, probabilidades podem ser atribuídas por:

- ▶ *modelos,*
- ▶ *dados,*
- ▶ *ou de forma subjetiva.*

Probabilidade(s) em (eventos de) interesse.
Revisando:

- ▶ Medida no intervalo $[0,1]$.
- ▶ Resultados por vezes contra-intuitivos.
- ▶ Formalização: fundamentos e notação.
- ▶ Propriedades (complementar, união, intersecção, condicional).
- ▶ Teorema de Bayes.

Problemas estilizados e analogias.

Atribuindo probabilidades

- ▶ Sair cara no lançamento de uma moeda.
- ▶ A soma de dois dados ser 6.
- ▶ Uma ninhada de cinco cães ter três fêmeas.
- ▶ Um jogador de basquete acertar todos lance-livres em um jogo.
- ▶ Uma seguradora registrar mais de 10 sinistros em um dia.
- ▶ Uma região registrar chuva (precipitação) em abril superior a 200 mm.
- ▶ Um paciente responder positivamente a certo tratamento.
- ▶ Uma pessoa avaliar positivamente um filme.

Atribuindo probabilidades

Modelos probabilísticos

- ▶ Padrões de comportamento.
- ▶ Situações *estilizadas* – modelos probabilísticos.
- ▶ Variáveis aleatórias
- ▶ Distribuições de probabilidades

Muitas vezes a sociedade "organiza" o mundo de forma a trazer previsibilidade a partir fenômenos de ocorrência imprevisível.

Como ter previsibilidade a partir da imprevisibilidade?

Variáveis aleatórias

São “simplificações” (funções) de interesse do espaço amostral.
São "dispositivos" para atribuir probabilidades em certos padrões de comportamento.

Um exemplo simples: quantos machos há em uma ninhada de três cães?
(estamos interessados somente no número de machos e não na ordem dos nascimentos)

Espaço amostral:

$$\Omega = \{(F,F,F), (F,F,M), (F,M,F), (M,F,F), (F,M,M), (M,F,M), (M,M,F), (M,M,M)\}$$

Variável aleatória: Y : número de machos

Domínio: $y \in \{0, 1, 2, 3\}$

Variáveis aleatórias

Organizando:

- ▶ Definindo a variável aleatória:

Y : número de machos em ninhada com 3 cães.

- ▶ Verificando seus possíveis valores (*domínio*):

$$y \in \{0, 1, 2, 3\}.$$

- ▶ Atribuindo probabilidades a seus possíveis valores:

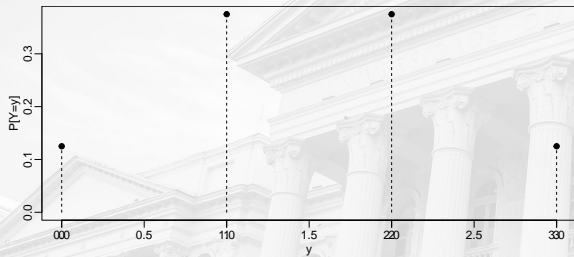
y	0	1	2	3
Probabilidades	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A tabela pode ser substituída por uma equação que produz seus valores:

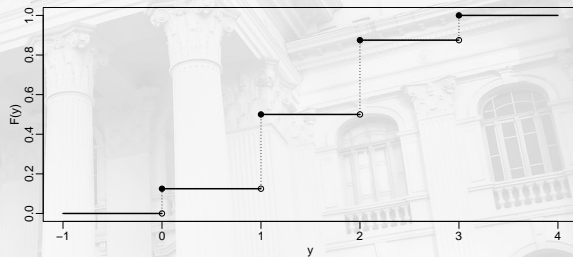
$$P[Y = y] = \binom{3}{y} (1/2)^y (1 - 1/2)^{3-y} = \binom{n}{y} (p)^y (1 - p)^{n-y}.$$

Probabilidades em gráficos

função de probabilidade



função (de probabilidade) acumulada



Revisitando a lista de problemas:

- ▶ Vamos formalizar as soluções
- ▶ Pode-se representar problemas *reais* parecidos ou análogos

Exercício 1

Considere o lançamento de um dado normal.

1. Quais os resultados possíveis?
2. Qual a probabilidade de sair a face 5?
3. Qual a probabilidade de cada possível resultado?
4. Qual a probabilidade de sair uma face que seja um número divisível por 3?

Exercício 1

Suposições e modelo teórico

- ▶ ocorre uma entre seis possíveis faces
- ▶ $\Omega = \{1,2,3,4,5,6\}$
- ▶ as faces são igualmente prováveis

Tabela 1. Distribuição de probabilidades da face no lançamento de um dado

Face (y)	1	2	3	4	5	6
Probabilidades $P[Y = y]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Prob. Acum. $P[Y \leq y] = F(y)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

$$P[Y = y_i] = \frac{1}{6} \quad \forall i$$

- ▶ Definimos uma **variável aleatória** (v.a.) que associa valores ao resultado do experimento. Neste último caso a associação é simples e direta, 1-1.
- ▶ **Distribuição de probabilidades**: como o total da probabilidade (1) se distribui entre os possíveis valores da v.a..
- ▶ A distribuição da v.a. pode ser descrita por uma tabela, ou, caso possível, por uma *fórmula* (uma **função de probabilidades**) que atribui seus valores.
- ▶ Neste caso dizemos ter uma **família de distribuição** especial ou conhecida.

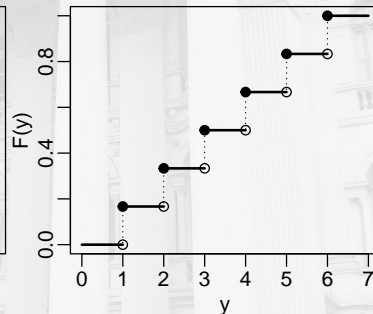
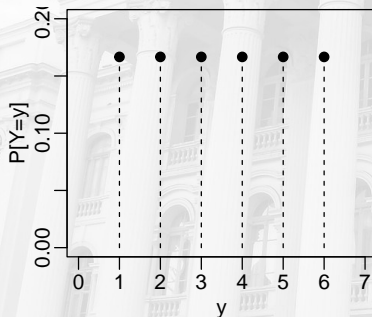
Generalizando a solução: Exercício 1

Y : face de um dado (v.a. discreta)

$y \in \{1,2,3,4,5,6\}$

$Y \sim U_d(n = 6)$

$$P[Y = y] = \frac{1}{n} = \frac{1}{6}$$



Exercício 2

Considere o lançamento de um dado não usual, no qual a probabilidade de cada face é proporcional ao seu valor.

1. Quais os resultados possíveis?
2. Qual a probabilidade de sair a face 5?
3. Qual a probabilidade de cada possível resultado?
4. Qual a probabilidade de sair uma face que seja um número divisível por 3?

Exercício 2

Suposições e modelo probabilístico (teórico)

- ▶ Ocorre uma entre seis possíveis faces.
- ▶ $\Omega = \{1,2,3,4,5,6\}$.
- ▶ As faces **não** são igualmente prováveis.

Tabela 2. Distribuição de probabilidades da face no lançamento de um dado

Face (y)	1	2	3	4	5	6
$P[Y = y]$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$
$P[Y \leq y]$	$\frac{1}{21}$	$\frac{3}{21}$	$\frac{6}{21}$	$\frac{10}{21}$	$\frac{15}{21}$	$\frac{21}{21}$

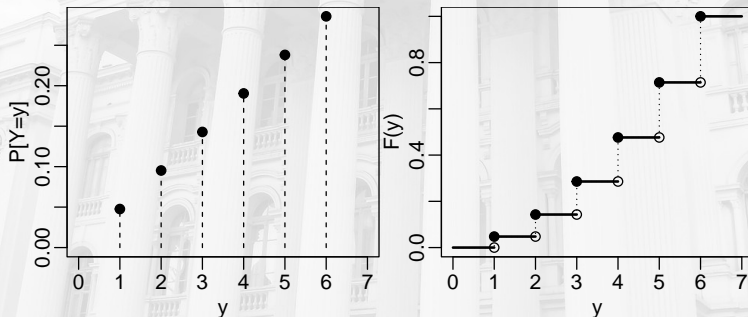
Generalizando a partir do Exercício 2

Notação:

Y : face de um dado (v.a. discreta)

$y \in \{1,2,3,4,5,6\}$

$$P[Y = y_i] = \frac{y_i}{\sum_i y_i}$$



Exercício 3

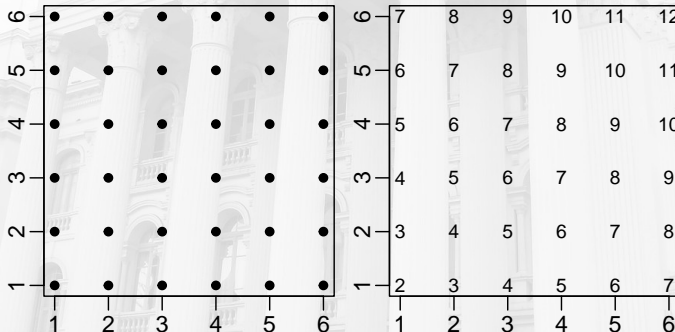
Considere o lançamento de dois dados e o interesse está na soma das faces.

1. Quais os resultados possíveis?
2. Qual a probabilidade da soma ser 5?
3. Qual a probabilidade de cada possível resultado?
4. Qual a probabilidade que a soma das faces seja um número divisível por 3?

Exercício 3

Suposições e modelo teórico

- ▶ $\Omega = \{(1,1),(1,2), \dots (2,1),(2,2), \dots (6,5),(6,6)\}$ $n(\Omega) = 36$
- ▶ Y : soma das faces no lançamento de dois dados
- ▶ $y \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

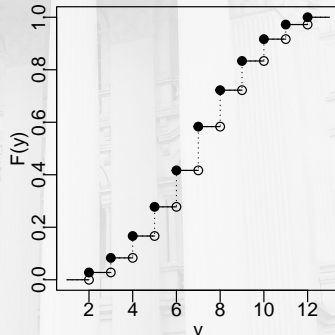
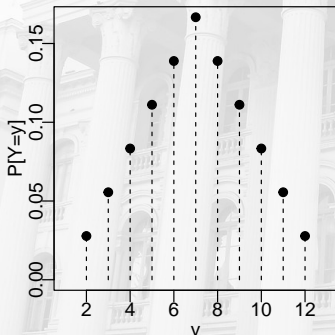


Exercício 3 (cont)

Tabela 3. Distribuição de probabilidades da soma das faces no lançamento de dois dados

y	2	3	4	5	6	7	8	9	10	11	12
$P[Y = y]$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(y) = P[Y \leq y]$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

A tabela pode ser expressa pela equação: $P[Y = y] = \frac{6-|y-7|}{36}$.



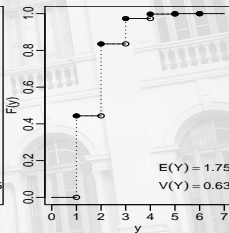
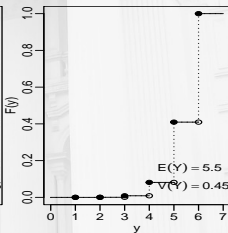
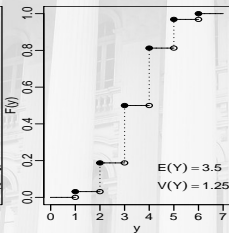
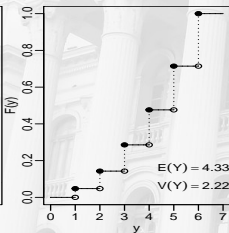
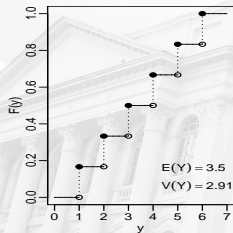
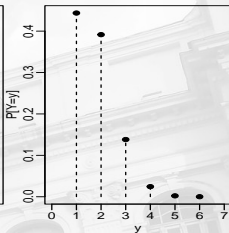
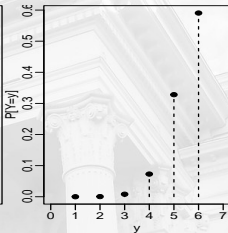
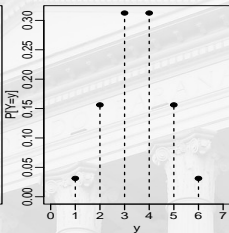
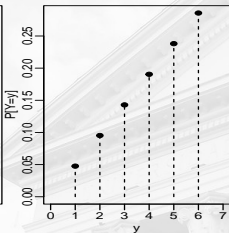
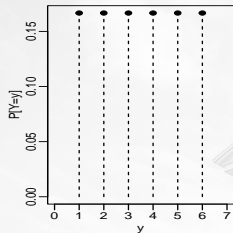
Variáveis aleatórias discretas

- ▶ Y tem valores em um conjunto **enumerável**
- ▶ Possui um **função de probabilidades** $P[Y = y]$ com valores expressos por uma tabela ou fórmula
- ▶ Pode tb ser caracterizada por uma **função acumulada**

$$F(y) = \sum_i P[Y = y_i]$$

- ▶ Pode-se calcular medidas que expressam características da distribuição tais como:
 - ▶ Média (valor esperado, esperança): $E[Y] = \sum_i y_i \cdot P[Y = y_i]$
 - ▶ Variância : $\text{Var}[Y] = \sum_i (y_i - E[Y])^2 \cdot P[Y = y_i]$
 - ▶ mediana, quantis, etc

Variáveis aleatórias discretas



Exercício 4

Um dado foi fabricado com o centro em madeira leve e cada face com uma chapa metálica porém de diferentes características (espessura/densidade) em cada face?

1. Quais os resultados possíveis?
2. Como calcular a probabilidade de sair a face 5?
3. Como calcular a probabilidade de cada possível resultado?
4. Como calcular a probabilidade de sair uma face que seja um número divisível por 3?

Exercício 4

Suposições e procedimento empírico

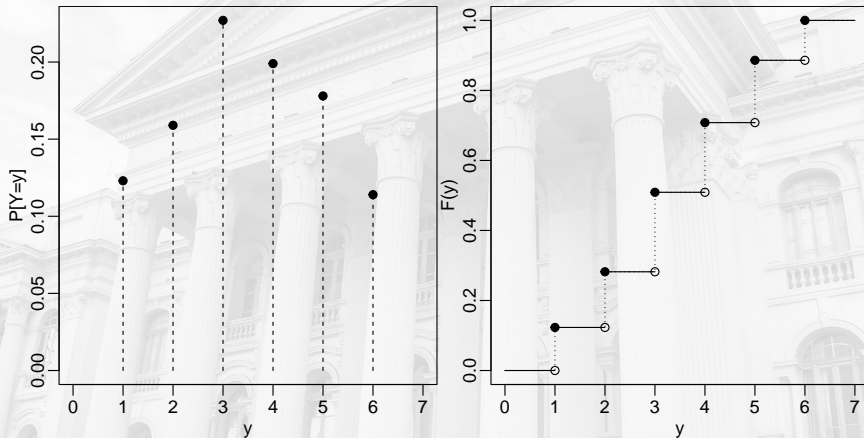
- ▶ ocorre uma entre seis possíveis faces
- ▶ $\Omega = \{1,2,3,4,5,6\}$
- ▶ as faces podem não ser igualmente prováveis
- ▶ experimento: lançou-se o dado 1000 vezes obtendo-se

Tabela 4. Distribuição de probabilidades da face no lançamento de um dado

Face (y)	1	2	3	4	5	6
Probabilidades	$\frac{123}{1000}$	$\frac{159}{1000}$	$\frac{227}{1000}$	$\frac{199}{1000}$	$\frac{178}{1000}$	$\frac{114}{1000}$

Exercício 4 (cont)

Distribuição de probabilidades obtida *empiricamente*

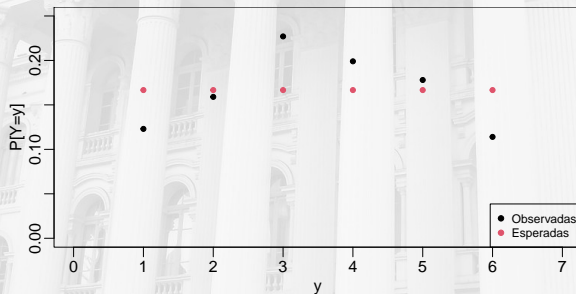


Exercício 4 (cont)

Uma pergunta válida:

Os valores observados são incompatíveis com a hipótese de equiprobabilidade?

Face (y)	1	2	3	4	5	6
Probs Observadas	$\frac{123}{1000}$	$\frac{159}{1000}$	$\frac{227}{1000}$	$\frac{199}{1000}$	$\frac{178}{1000}$	$\frac{114}{1000}$
Probs "Esperadas"	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{166}{1000}$	$\frac{166}{1000}$



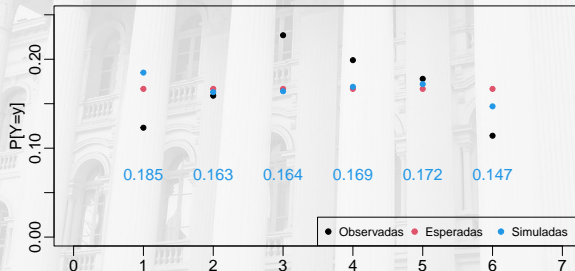
Exercício 4 (cont)

Uma pergunta válida:

Os valores observados são incompatíveis com a hipótese de equiprobabilidade?

Quanto pode variar mesmo sendo equiprováveis?

Face (y)	1	2	3	4	5	6
Probs Observadas	$\frac{123}{1000}$	$\frac{159}{1000}$	$\frac{227}{1000}$	$\frac{199}{1000}$	$\frac{178}{1000}$	$\frac{114}{1000}$
Probs "Esperadas"	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{166}{1000}$	$\frac{166}{1000}$
Probs Simuladas	$\frac{185}{1000}$	$\frac{163}{1000}$	$\frac{164}{1000}$	$\frac{169}{1000}$	$\frac{172}{1000}$	$\frac{147}{1000}$



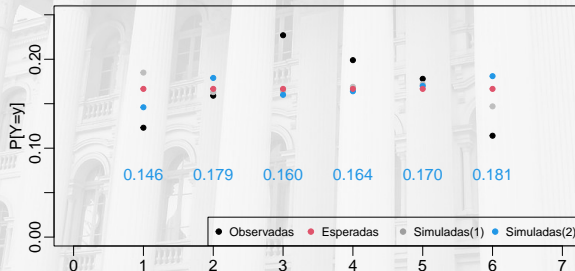
Exercício 4 (cont)

Uma pergunta válida:

Os valores observados são incompatíveis com a hipótese de equiprobabilidade?

Quanto pode variar mesmo sendo equiprováveis?

Face (y)	1	2	3	4	5	6
Probs Observadas	$\frac{123}{1000}$	$\frac{159}{1000}$	$\frac{227}{1000}$	$\frac{199}{1000}$	$\frac{178}{1000}$	$\frac{114}{1000}$
Probs "Esperadas"	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{167}{1000}$	$\frac{166}{1000}$	$\frac{166}{1000}$
Probs Simuladas (2)	$\frac{146}{1000}$	$\frac{179}{1000}$	$\frac{160}{1000}$	$\frac{164}{1000}$	$\frac{170}{1000}$	$\frac{181}{1000}$

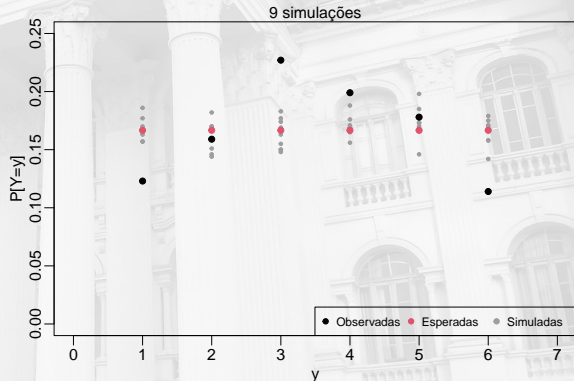
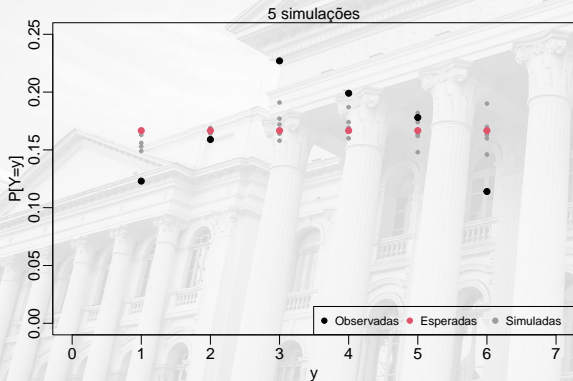


Exercício 4 (cont)

Uma pergunta válida:

Os valores observados são incompatíveis com a hipótese de equiprobabilidade?

Quanto pode variar mesmo sendo equiprováveis?

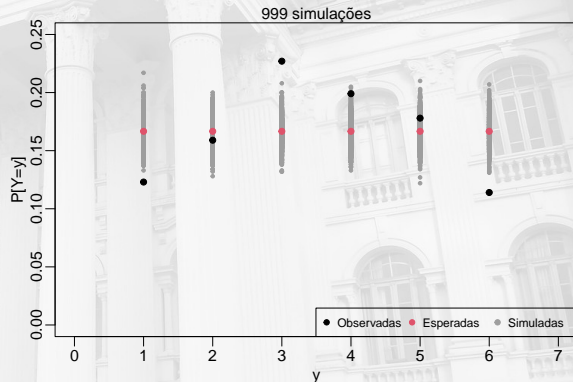
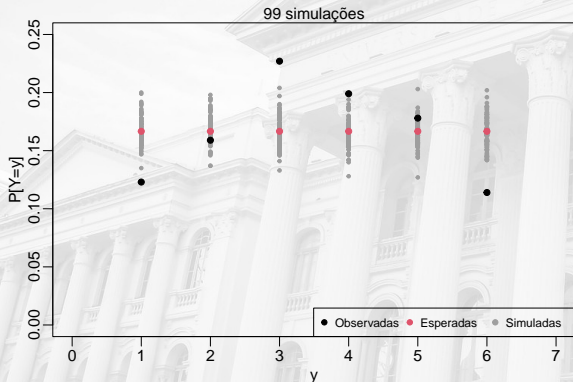


Exercício 4 (cont)

Uma pergunta válida:

Os valores observados são incompatíveis com a hipótese de equiprobabilidade?

Quanto pode variar mesmo sendo equiprováveis?



Exercício 5

Voce vai a um cassino em uma mesa que tem um jogo no qual se lançam dois dados como em um problema anterior. A regra é a de que se a soma for 6, 7 ou 8 voce ganha, valor igual ao apostado, caso contrário, perde o apostado.

1. Qual sua opinião sobre suas chances de ganhar?
2. Quais os resultados possíveis?
3. Qual sua opinião sobre a probabilidade da soma ser 5?
4. Qual sua opinião sobre a probabilidade de cada possível resultado?
5. Qual sua opinião sobre a probabilidade de sair uma face que seja um número divisível por 3?

Exercício 5

Qual sua **opinião** sobre a probabilidade?

Revisitando **Interpretações de probabilidade**

- ▶ Clássica: obtenção a partir de modelo teórico (ex. 1 a 3)
- ▶ Frequentista: obtenção a partir de procedimento empírico (ex. 4)
- ▶ Subjetiva: opinião do indivíduo (ex. 5)

Exercício 6

Em um teste múltipla escolha de quatro questões, deve-se marcar uma alternativa em cada questão.

Cada questão é composta de cinco alternativas das quais apenas uma é correta. Qual a probabilidade de um indivíduo acertar por mero acaso alguma questão?

Exercício 6

A resposta é mais simples se calcularmos pelo complementar: $P[\cdot] = 1 - (0,8)^4 = 0.59$, mas
 Organizando o problema ...

- ▶ $\Omega = \{(CCCC), (CCCE), \dots, (EEEE)\}$ $n(\Omega) = 16$
- ▶ Y : número de questões certas
- ▶ $y \in \{0, 1, 2, 3, 4\}$
- ▶ $P[Y = y] = \binom{4}{y} (1/5)^y (4/5)^{4-y} = \binom{n}{y} p^y (1-p)^{n-y}$ (distribuição Binomial!!)
- ▶ Evento de interesse:

$$P[Y \geq 1] = 1 - P[Y = 0] = 1 - \binom{4}{0} (1/5)^0 (4/5)^4 = 0.59$$

Código R:

```
1 - dbinom(0, size=4, prob=1/5)
## [1] 0.5904
```

Exercício 8

Alguns biólogos fizeram estudos de laboratório sobre o comportamento de animais quando submetidos a um estímulo, o quais poderiam apresentar ou não resposta positiva. Em particular estavam interessados nas respostas positivas ao estímulo. Considera-se que na população destes animais, 10% sejam sensíveis ao estímulo.

O biólogo *A* possuía um grupo em que 10 animais eram sensíveis e 20 eram insensíveis. Ele selecionou ao acaso 8 animais para teste.

O biólogo *B* dispunha de um grande número de animais e foi testando um a um até encontrar o terceiro sensível ao estímulo.

O biólogo *C* fazia testes diários e encontrava uma média de 2,8 animais sensíveis a cada dia.

O biólogo *D* submeteu 10 animais ao estímulo e registrou quantos eram sensíveis.

O biólogo *E* dispunha de um grande número de animais e foi testando um a um até encontrar um sensível ao estímulo.

Exercício 8

Distribuições discretas mais usuais:

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

Biólogo A: Distribuição Hipergeométrica

O biólogo A possuía 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.

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

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

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

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

Código **R**:

```
1 - hyper(1, m=10, n=20, k=8)
```

```
## [1] 0.8460308
```

Distribuição binomial negativa

O biólogo B 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.

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

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

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

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

Código **R**:

```
pnbinom(3, size=3, prob=0.1)
```

```
## [1] 0.01585
```

Distribuição Poisson

O biólogo C 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.

$Y \sim P(\lambda)$

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

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

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

Código **R**:

```
ppois(1, lambda=2.8)
```

```
## [1] 0.2310782
```

Distribuição Binomial

O biólogo D 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.

$Y \sim B(n, p)$

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

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

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

Código **R**:

```
1 - pbinom(3, size=10, prob=0.1)
```

```
## [1] 0.0127952
```

Distribuição Geométrica

O biólogo E 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.

$$Y \sim G(p)$$

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

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

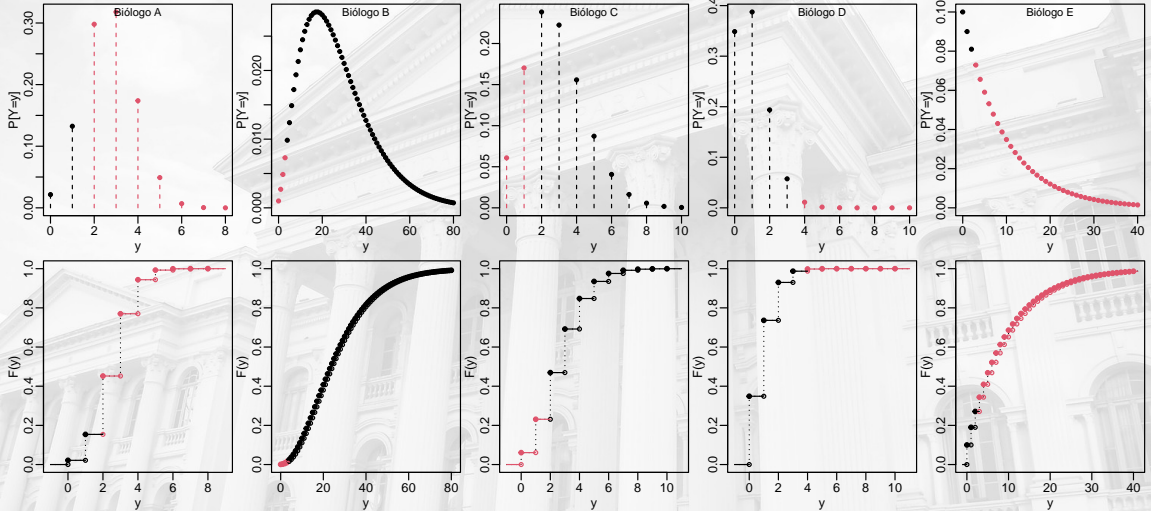
$$P[Y \geq 3] = 1 - P[Y = 0] - P[Y = 1] - P[Y = 2] = 0.7290$$

Código R:

```
1 - pgeom(2, prob=0.1)
```

```
## [1] 0.729
```


Os modelos dos biólogos



Atenção: definição da variável

Geométrica:

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

ou

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

Binomial negativa:

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

ou

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

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) $f(y)$
 - ▶ $f(y) \geq 0 \quad \forall y$
 - ▶ $\int f(y)dy = 1$
- ▶ E as probabilidades de interesse são áreas sob a função

$$P[a < Y < b] = \int_a^b f(y)dy$$

Exercício 10 (cont)

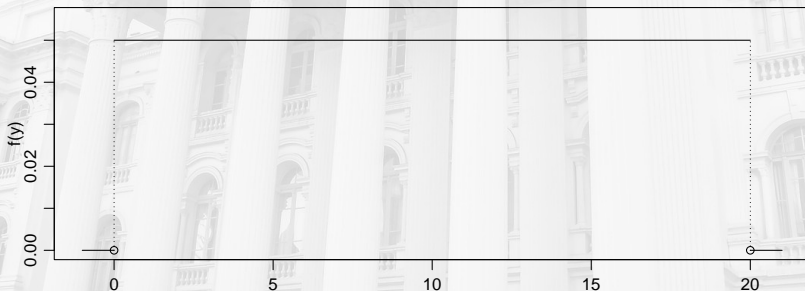
Considere o surgimento de um defeito na pista em um trecho de rodovia com extensão de 20 *km*.

Y : 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*?

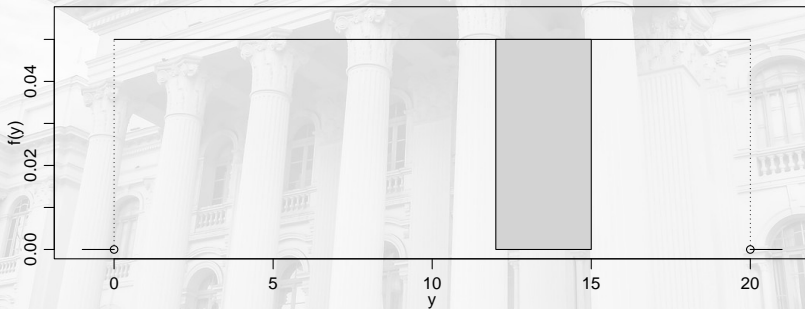
Resposta: $1/4 = 0,25$



Exercício 10 (cont)

2. Qual a probabilidade de que o defeito ocorra entre os quilômetros 12 e 15?

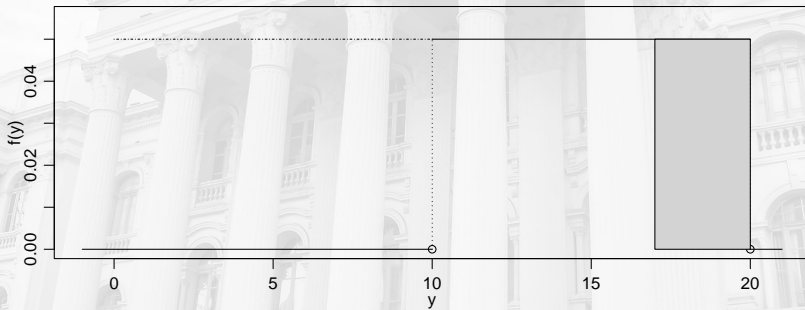
Resposta: $3/20 = 0,15$



Exercício 10 (cont)

- 3 Se o defeito ocorre na segunda metade do trecho, qual a probabilidade de seja nos últimos 3 km?

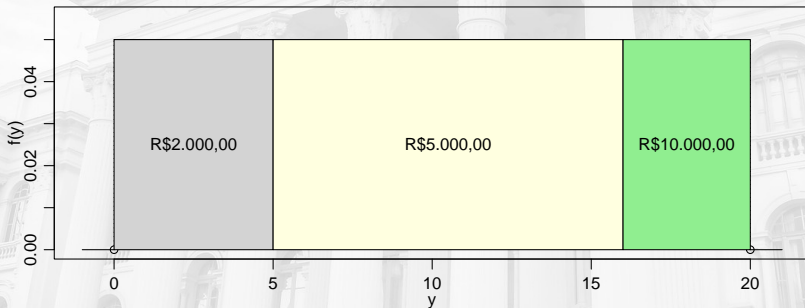
Resposta: $\frac{3}{10}$



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?

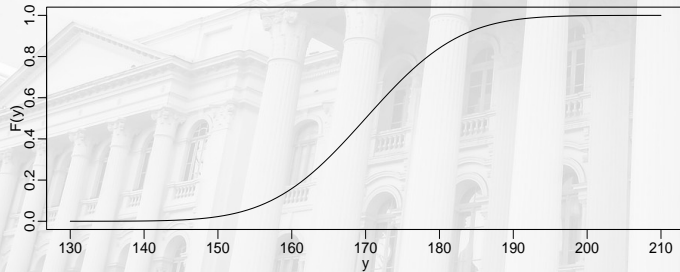
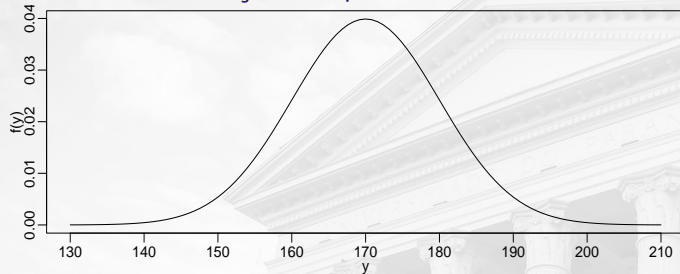
Resposta: $\frac{5}{20}2000 + \frac{11}{20}5000 + \frac{4}{20}10000 = 5250$



Outras distribuições

Mas, ... "o mundo" não é uniforme!

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



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

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

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

1. $P[Y \leq 160]$

2.

3.

4.

5.

6.

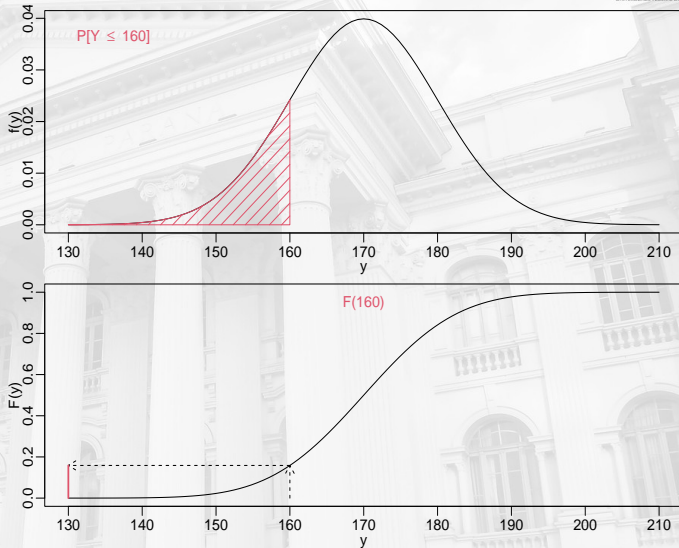
7.

8.

9.

Código **R**:

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



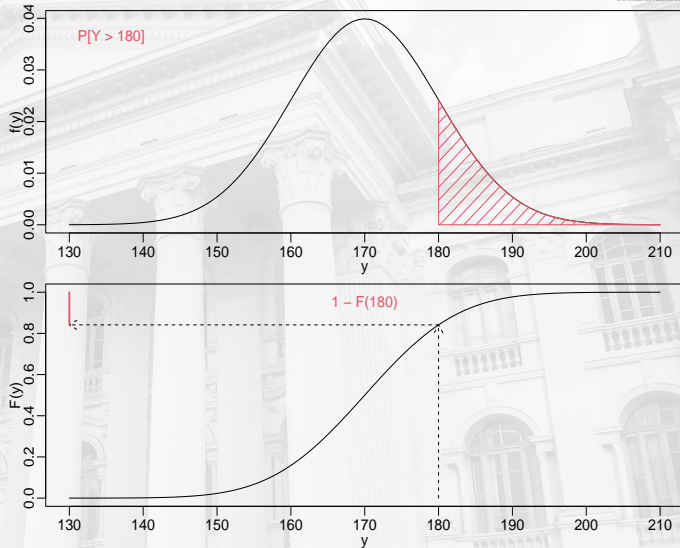
Probabilidade

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

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

Código **R**:

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



Probabilidade

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

1.

2.

3. $P[155 \leq Y \leq 175]$

4.

5.

6.

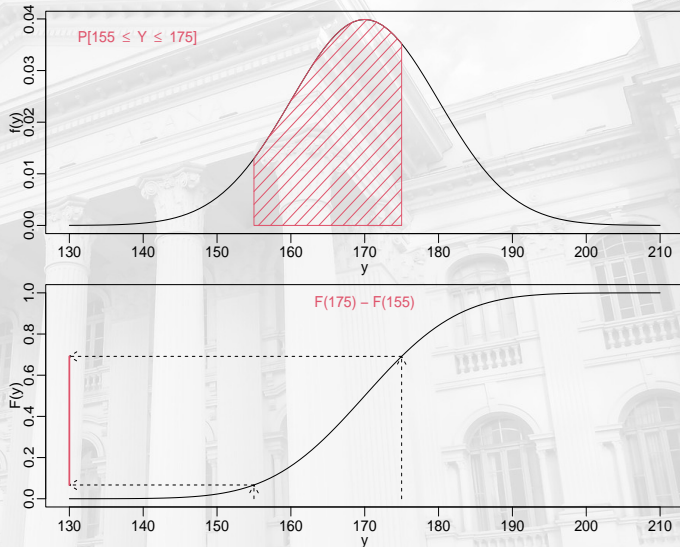
7.

8.

9.

Código **R**:

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



Quantis

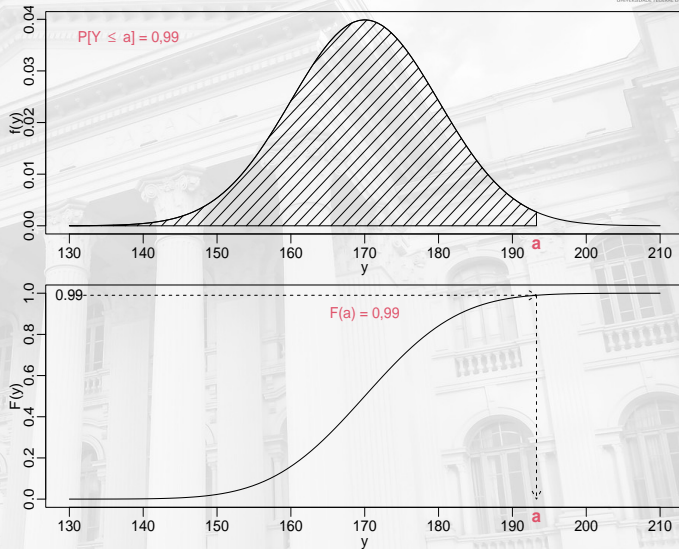
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

$$P[Y \leq a] = 0,99$$

Código **R**:

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



Quantis

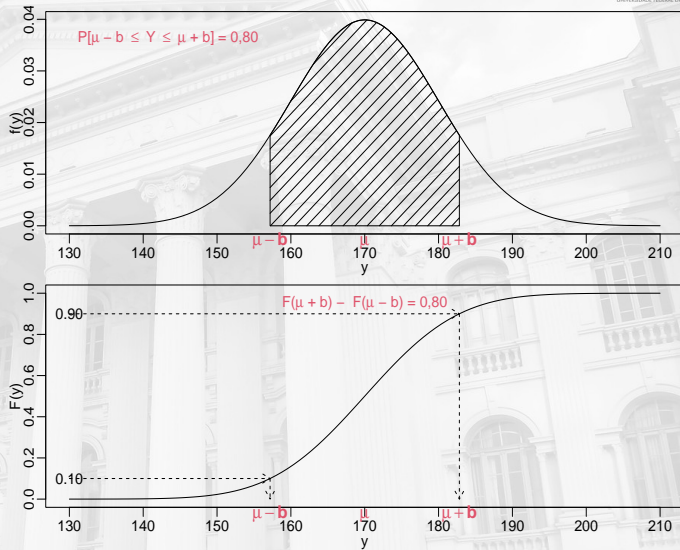
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

$$P[\mu - b \leq Y \leq \mu + b] = 0,80$$

Código **R**:

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



Probabilidade condicional

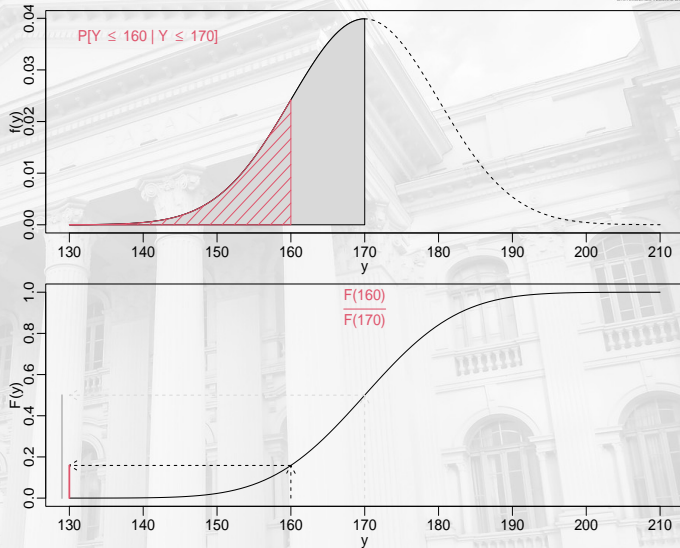
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

$$P[Y \leq 160 | Y \leq 170]$$

Código **R**:

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



Probabilidade condicional

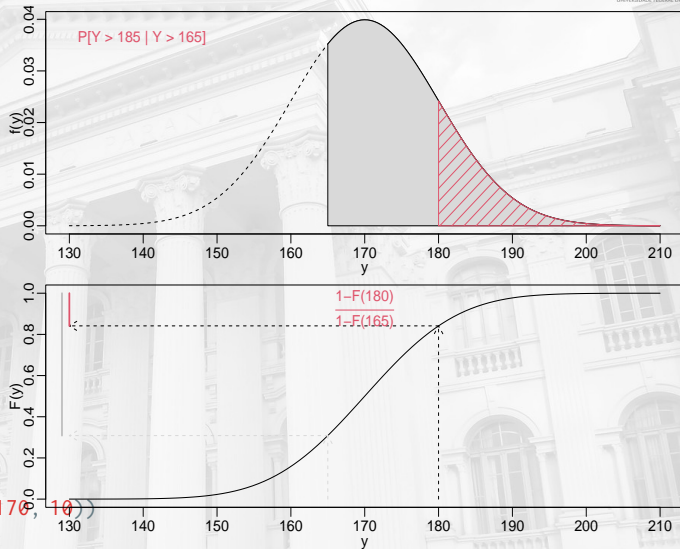
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

$$P[Y > 180 | Y > 165]$$

Código **R**:

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



Probabilidade condicional

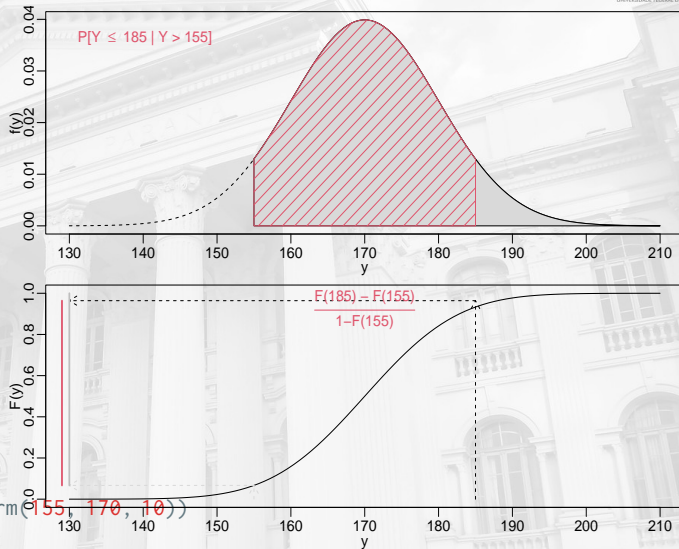
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

$$P[Y \leq 185 | Y > 155]$$

Código **R**:

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



Probabilidade condicional

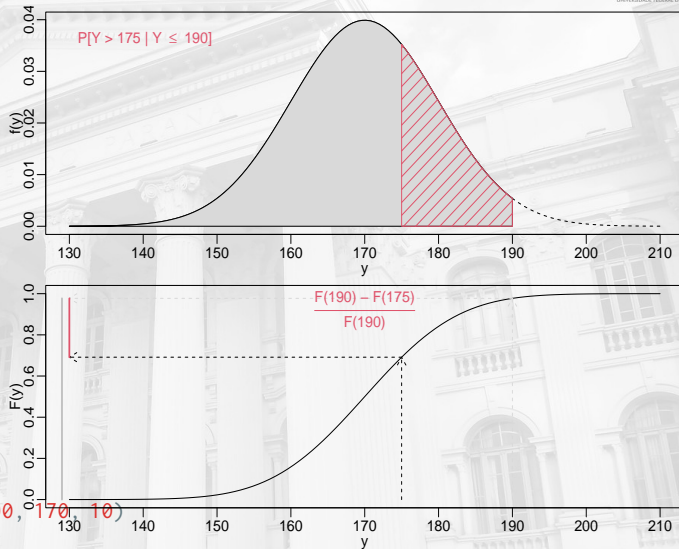
$$Y \sim N(\mu = 170, \sigma^2 = 10^2)$$

- 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
```



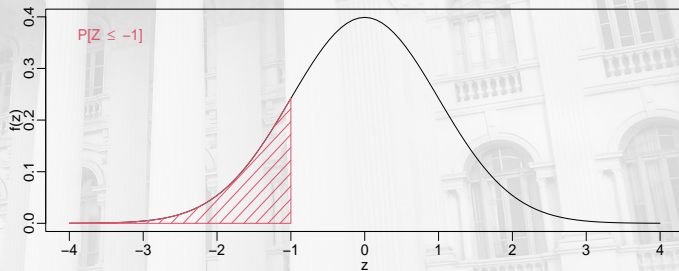
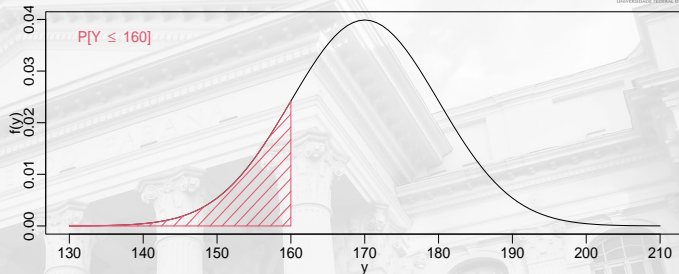
Escore Z e probabilidade

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

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

- 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:

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



Escore Z e quantil

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

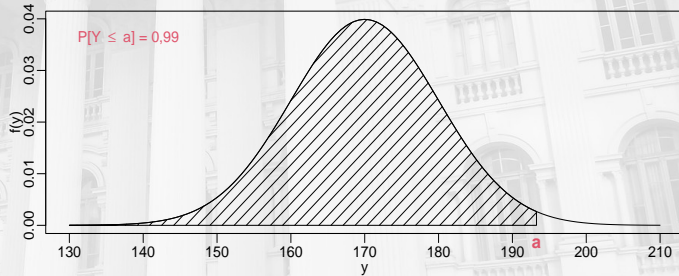
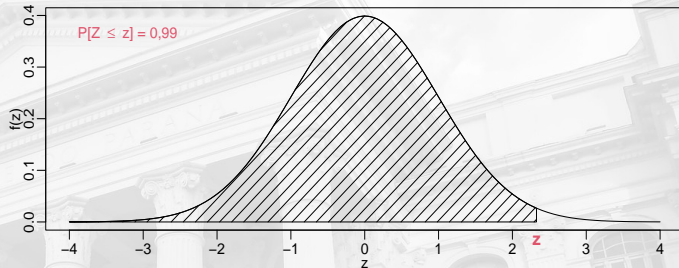
$$Z = \frac{Y - \mu}{\sigma} \sim N(\mu_Z = 0, \sigma_Z^2 = 1)$$

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

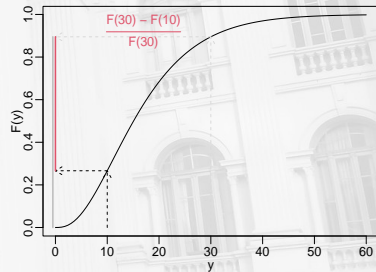
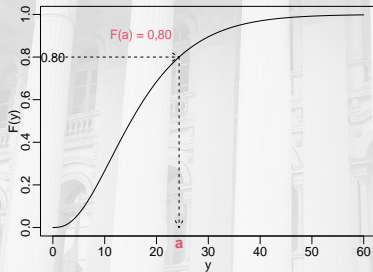
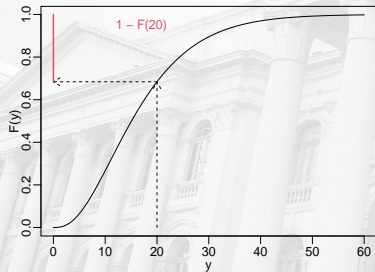
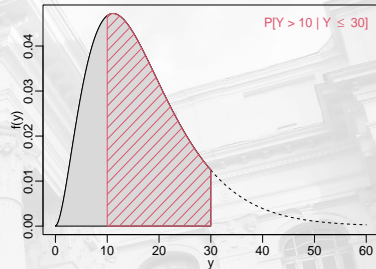
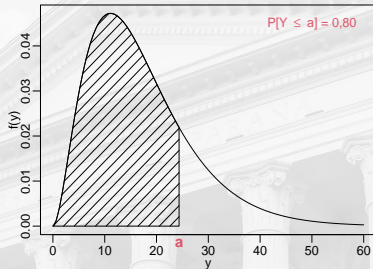
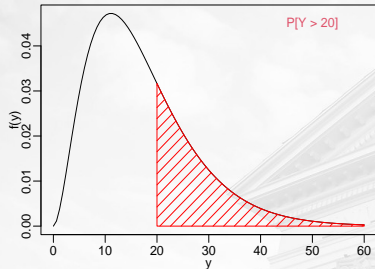
$$z = 2.326$$

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

$$a = 193.263$$



Distribuição Gama



Distribuição Gama

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

Parametrização utilizada: "shape" (forma) e "scale" (escala)

$$(\text{shape}) \text{ sh} = \frac{\mu^2}{\sigma^2}$$

$$(\text{scale}) \text{ sc} = \frac{\sigma^2}{\mu}$$

Códigos **R**:

```
pgamma(20, sh = (17/10)^2, sc = 10^2/17, lower = FALSE)
```

```
## [1] 0.3160481
```

```
qgamma(0.80, sh = (17/10)^2, sc = 10^2/17)
```

```
## [1] 24.35641
```

```
diff(pgamma(c(10, 30), sh = (17/10)^2, sc = 10^2/17))/pgamma(30, sh = (17/10)^2, sc=10^2/17)
```

```
## [1] 0.7024844
```

Comparando v.a.'s discretas e contínuas

Uniforme discreta:

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

$$y \in \{1,2,3,4,5,6\}$$

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

$$F_{Y_1}(y) = \frac{\lfloor y \rfloor + 1}{6}$$

Uniforme contínua:

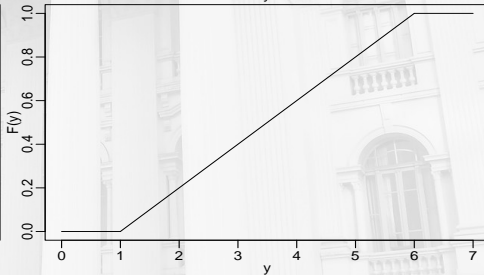
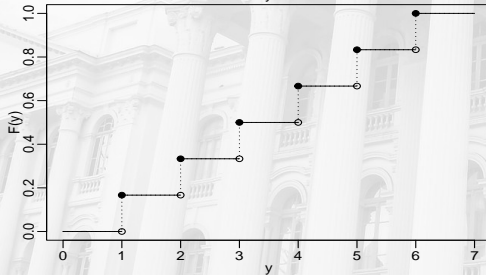
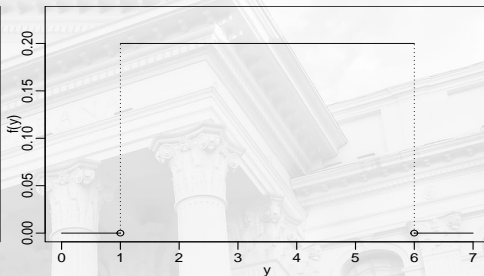
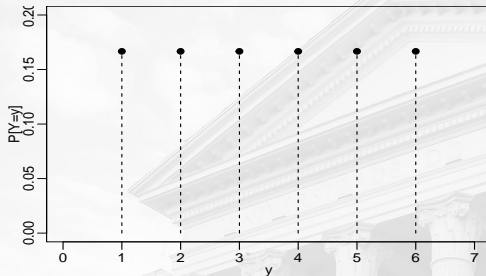
$$Y_2 \sim U_c[1,6]$$

$$y \in [1,6]$$

$$f(Y_2 = y) = \frac{1}{6 - 1} = \frac{1}{5}$$

$$F_{Y_2}(y) = \frac{y - 1}{6 - 1} = \frac{y - 1}{5}$$

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:

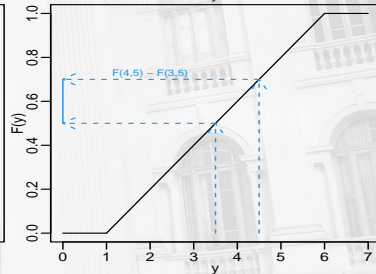
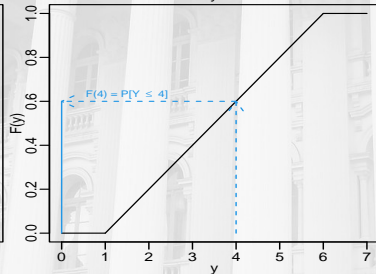
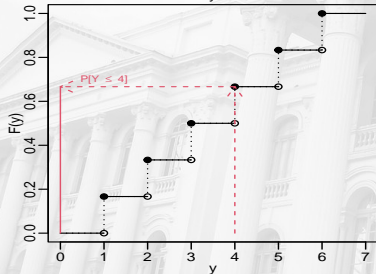
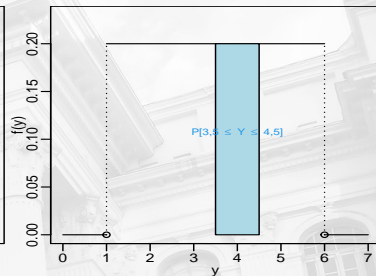
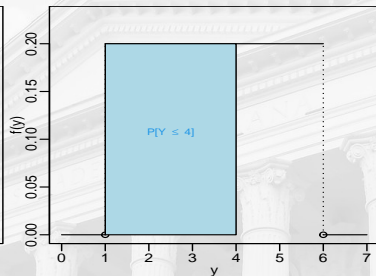
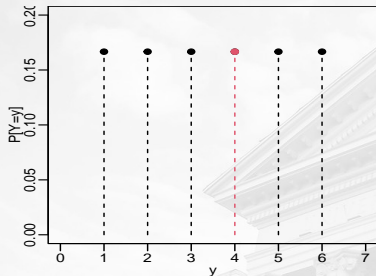
$$Y_2 \sim U_c[1,6]$$

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

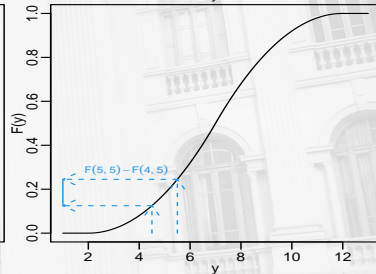
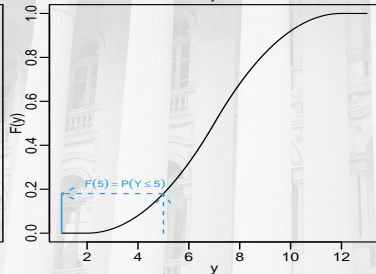
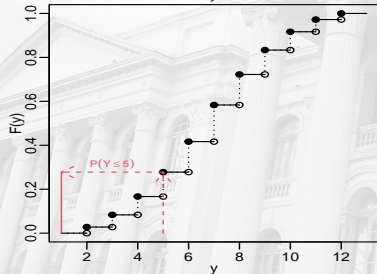
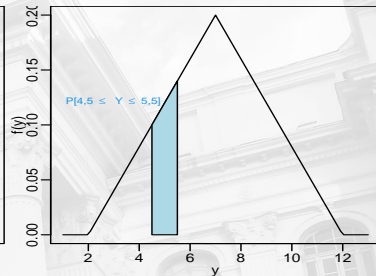
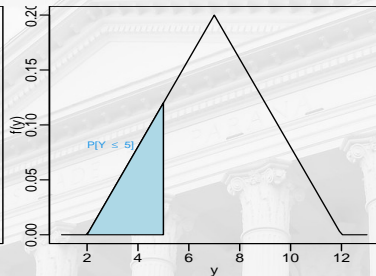
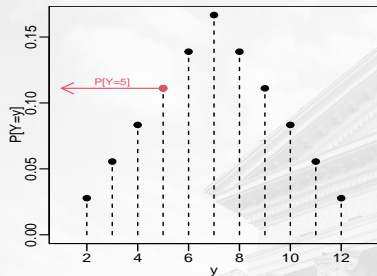
$$P[3,5 < Y_2 \leq 4,5] = \int_{3,5}^{4,5} f(y)dy = 0.2$$

$$P[Y_2 \leq 4] = F_{y_2}(4) = \int_1^4 f(y)dy = 0.6$$

V.A.'s (uniforme) discreta e contínua



V.A.'s discretas e contínuas



- ▶ **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.