



## Probabilidade e Estatística (EAD)

*Tradução e adaptação:*

Priscilla Gnewuch, Márcia Helena Barbian e Maitê Mückler

Slides baseados no material desenvolvido por Mine Çetinkaya-Rundel of OpenIntro.

Tanto este material adaptado, quanto o original, podem ser copiados, editados e/ou compartilhados. O material adaptado está licenciado sob a Licença Creative Commons Atribuição 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by/4.0/>

# **Capítulo 3: Distribuições de Variáveis Aleatórias**

---

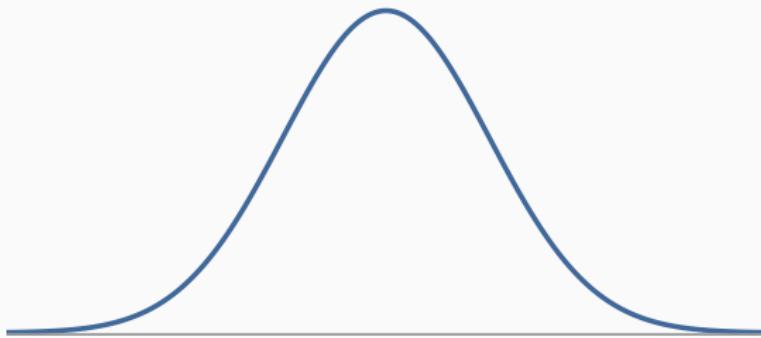
Slides desenvolvidos por Mine Çetinkaya-Rundel of OpenIntro.  
Os slides podem ser copiados, editados e / ou compartilhados via CC BY-SA licença.  
Algumas imagens podem ser incluídas em diretrizes de uso justo (propósitos educacionais).

### **3.1. Distribuição normal**

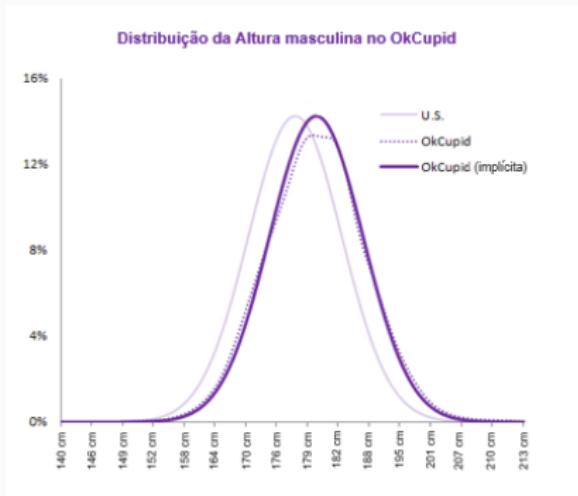
---

## Distribuição normal

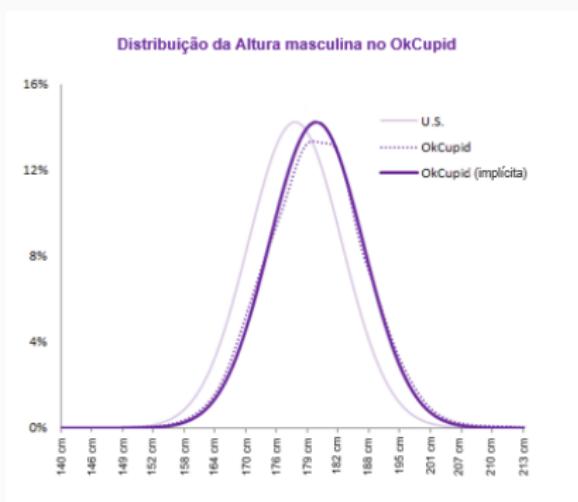
- Curva unimodal e simétrica em forma de sino
- Muitas variáveis são quase normais, mas nenhuma é exatamente normal
- Denotado como  $N(\mu, \sigma)$  → Normal com média  $\mu$  e desvio padrão  $\sigma$



# Alturas dos homens



# Alturas dos homens



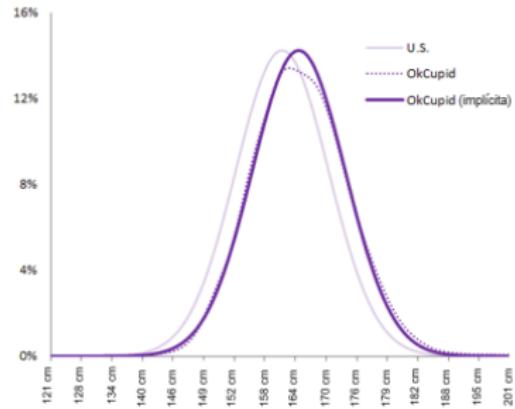
"As alturas dos homens no OkCupid quase seguem a distribuição normal esperada - exceto pelo fato de que o gráfico é mais deslocado para a direita do que na população geral de homens. Frequentemente os homens gostam de adicionar alguns centímetros à sua altura verdadeira."

"Uma vaidade discreta pode ser observada: a parte superior da curva pontilhada, que começa em cerca de 176 cm, inclina-se ainda mais para a direita. Isso significa que eles chegam mais perto de um 1 metro e 80 cm, ou seja, aumentando ainda mais o valor de referência da altura desejada pelos homens."

<http://blog.okcupid.com/index.php/the-biggest-lies-in-online-dating/>

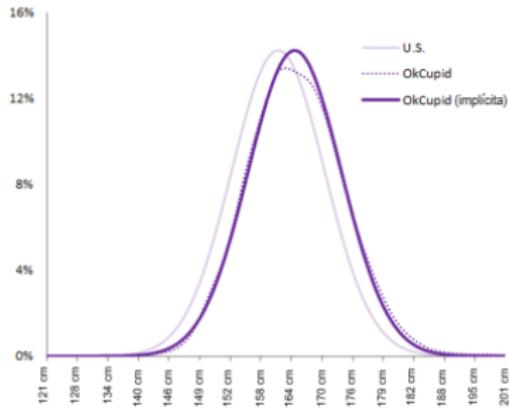
# Alturas das mulheres

Distribuição da Altura feminina no OkCupid



# Alturas das mulheres

Distribuição da Altura feminina no OkCupid



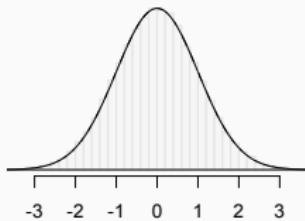
"Já quando analisamos os dados das mulheres, ficamos surpresos ao ver que o exagero da altura uma característica tão propagada, embora, no caso das mulheres, sem necessariamente uma altura de referência desejada.

<http://blog.okcupid.com/index.php/the-biggest-lies-in-online-dating/>

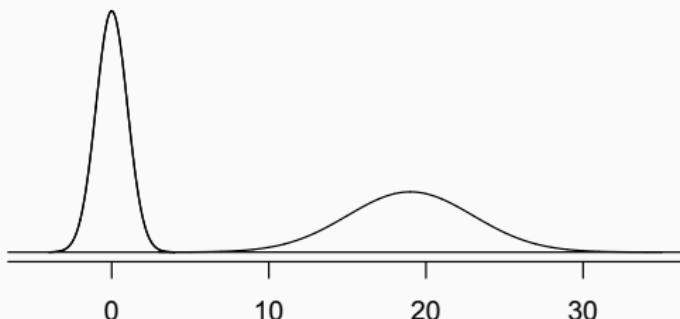
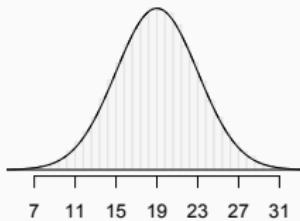
# Distribuições normais com diferentes parâmetros

$\mu$ : média,  $\sigma$ : desvio padrão

$$N(\mu = 0, \sigma = 1)$$



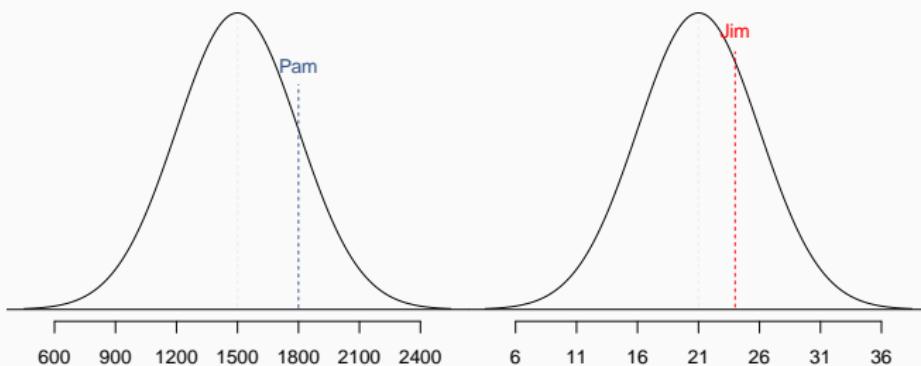
$$N(\mu = 19, \sigma = 4)$$



# Distribuições normais com diferentes parâmetros

Os escores do SAT (*Scholastic Aptitude Test* - um tipo de vestibular americano) possuem distribuição quase normal, com média 1.500 e desvio padrão 300. Os escores do ACT (*American College Testing* - outro vestibular americano) também são distribuídos quase que normalmente, com média 21 e desvio padrão 5.

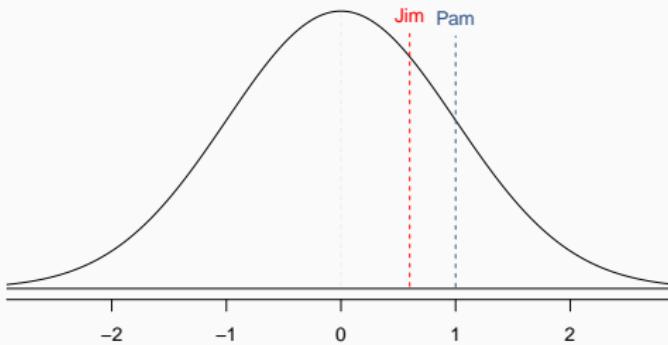
Um funcionário responsável pela aprovação dos candidatos em uma Universidade gostaria de determinar qual, de dois candidatos, obteve uma pontuação melhor em seu vestibular em relação aos outros vestibulandos: Pam, que obteve 1800 em seu SAT, ou Jim, que marcou 24 em seu ACT?



## Padronizando com escores Z

Como não podemos simplesmente comparar esses dois escores brutos, vamos verificar a quantos desvios-padrão além da média cada escore está.

- A pontuação de Pam é  $\frac{1800 - 1500}{300} = 1$  desvio padrão acima da média.
- A pontuação de Jim é  $\frac{24 - 21}{5} = 0.6$  desvios-padrão acima da média.



## Padronizando com escores Z (cont.)

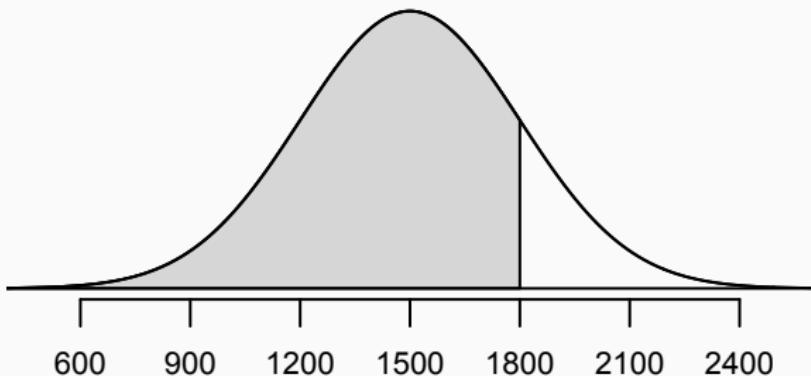
- Estes são chamados de *escores padronizado*, ou *escores Z*.
- O escore Z de uma observação é o número de desvios-padrão que esta observação está acima ou abaixo da média.

$$Z = \frac{\text{observacao} - \text{media}}{SD}$$

- Os escores Z são definidos para distribuições de qualquer formato, porém somente quando a distribuição é normal podemos usar escores Z para calcular *percentis*.
- Observações que estão a mais de 2 desvios-padrão da média ( $|Z| > 2$ ) são geralmente consideradas incomuns.

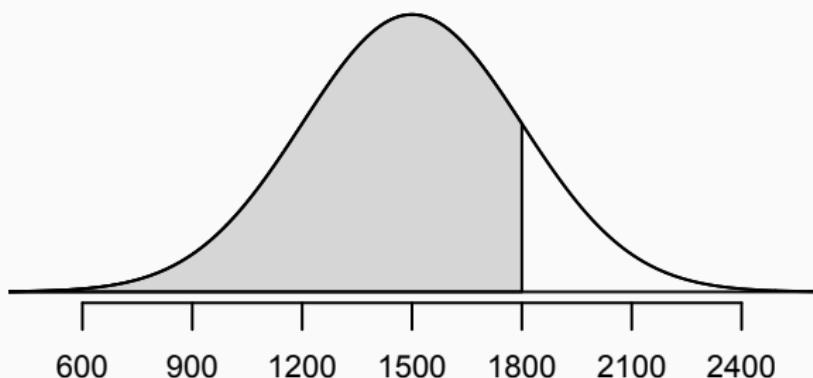
## Percentis

- *Percentil* é a porcentagem de observações que caem abaixo de um determinado valor da distribuição de dados.
- Graficamente, o percentil é a área abaixo da curva de distribuição de probabilidade à esquerda dessa observação.



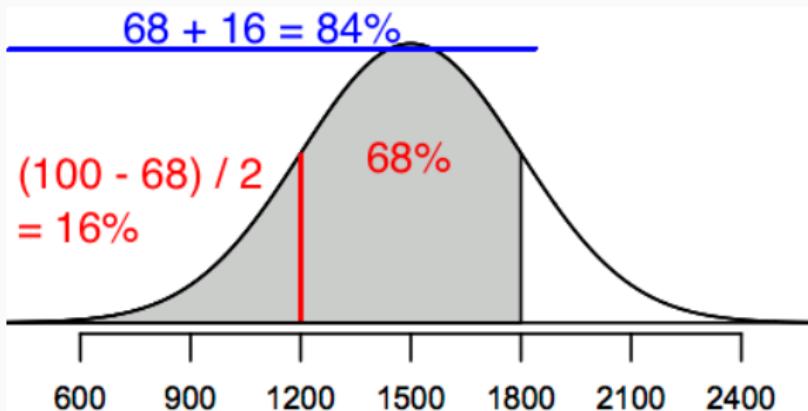
## Percentis

Aproximadamente, qual porcentagem de alunos tem pontuação abaixo de 1800 no SAT? (Dica: use a regra 68-95-99.7%).



## Percentis

Aproximadamente, qual porcentagem de alunos tem pontuação abaixo de 1800 no SAT? (Dica: use a regra 68-95-99.7%).



$$100 - 68 = 32\%$$

$$32/2 = 16\%$$

$$68 + 16 = 84\%$$

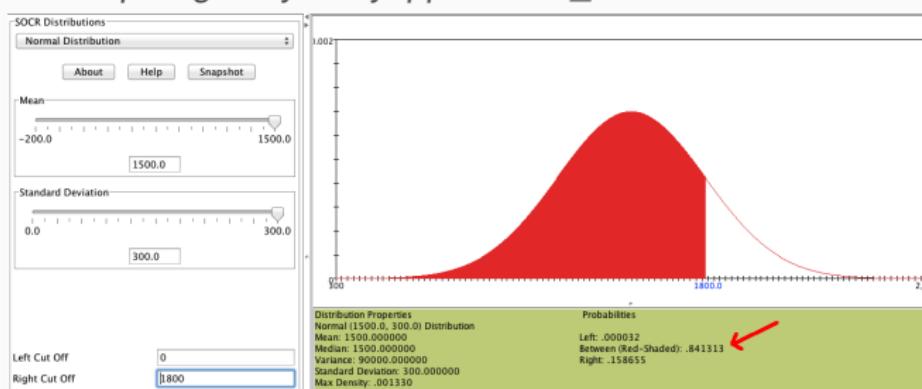
# Calculando percentis - usando computação

Existem muitas maneiras de calcular os percentis/áreas sob a curva:

- R:

```
> pnorm(1800, mean = 1500, sd = 300)  
[1] 0.8413447
```

- Applet: [https://gallery.shinyapps.io/dist\\_calc/](https://gallery.shinyapps.io/dist_calc/)



# Calculando percentis - usando tabelas

Z		Second decimal place of Z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	

## 6 sigma

“O termo *processo 6 sigma* vem da ideia de que, se houver seis desvios-padrão entre a média do processo e o limite de especificação mais próximo, como mostrado no gráfico, praticamente nenhum item deixará de atender às especificações.”

6 $\sigma$

[http://en.wikipedia.org/wiki/Six\\_Sigma](http://en.wikipedia.org/wiki/Six_Sigma)

## Controle de qualidade

Na fábrica de ketchup Heinz, as quantidades de ketchup inseridas nas garrafas devem ser normalmente distribuídas com uma média de 36g e desvio-padrão 0,11g. A cada 30 minutos, uma garrafa é selecionada da linha de produção e seu conteúdo é precisamente anotado. Se a quantidade de ketchup na garrafa estiver abaixo de 35,8g ou acima de 36,2g, então o frasco falha na inspeção de controle de qualidade. Qual a porcentagem de garrafas possue menos de 35,8g de ketchup?

## Controle de qualidade

Na fábrica de ketchup Heinz, as quantidades de ketchup inseridas nas garrafas devem ser normalmente distribuídas com uma média de 36g e desvio-padrão 0,11g. A cada 30 minutos, uma garrafa é selecionada da linha de produção e seu conteúdo é precisamente anotado. Se a quantidade de ketchup na garrafa estiver abaixo de 35,8g ou acima de 36,2g, então o frasco falha na inspeção de controle de qualidade. Qual a porcentagem de garrafas possue menos de 35,8g de ketchup?

*Seja  $X$  = quantidade de ketchup em uma garrafa:*

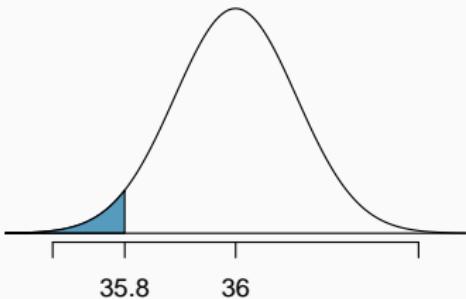
$$X \sim N(\mu = 36, \sigma = 0,11)$$

## Controle de qualidade

Na fábrica de ketchup Heinz, as quantidades de ketchup inseridas nas garrafas devem ser normalmente distribuídas com uma média de 36g e desvio-padrão 0,11g. A cada 30 minutos, uma garrafa é selecionada da linha de produção e seu conteúdo é precisamente anotado. Se a quantidade de ketchup na garrafa estiver abaixo de 35,8g ou acima de 36,2g, então o frasco falha na inspeção de controle de qualidade. Qual a porcentagem de garrafas possue menos de 35,8g de ketchup?

*Seja  $X$  = quantidade de ketchup em uma garrafa:*

$$X \sim N(\mu = 36, \sigma = 0,11)$$

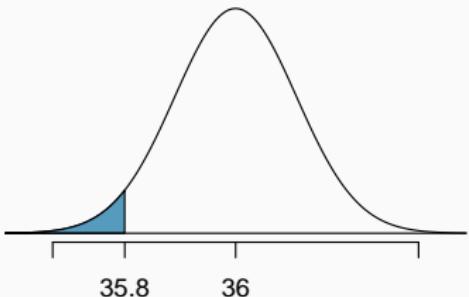


## Controle de qualidade

Na fábrica de ketchup Heinz, as quantidades de ketchup inseridas nas garrafas devem ser normalmente distribuídas com uma média de 36g e desvio-padrão 0,11g. A cada 30 minutos, uma garrafa é selecionada da linha de produção e seu conteúdo é precisamente anotado. Se a quantidade de ketchup na garrafa estiver abaixo de 35,8g ou acima de 36,2g, então o frasco falha na inspeção de controle de qualidade. Qual a porcentagem de garrafas possue menos de 35,8g de ketchup?

Seja  $X$  = quantidade de ketchup em uma garrafa:

$$X \sim N(\mu = 36, \sigma = 0,11)$$



$$Z = \frac{35.8 - 36}{0.11} = -1.82$$

## Encontrar a probabilidade exata - usando a tabela Z

Segunda casa decimal de Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	-2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	-2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	-2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	-2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	-2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	-2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	-2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	-2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	-2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	-2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5

# Encontrar a probabilidade exata - usando a tabela Z

Segunda casa decimal de Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	-2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	-2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	-2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	-2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	-2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	-2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	-2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	-2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	-2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	-2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5

## Encontrar a probabilidade exata - usando R

```
> pnorm(-1.82, mean = 0, sd = 1)  
[1] 0.0344
```

## Encontrar a probabilidade exata - usando R

```
> pnorm(-1.82, mean = 0, sd = 1)  
[1] 0.0344
```

ou

## Encontrar a probabilidade exata - usando R

```
> pnorm(-1.82, mean = 0, sd = 1)  
[1] 0.0344
```

ou

```
> pnorm(35.8, mean = 36, sd = 0.11)  
[1] 0.0345
```

## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

- (a) 1.82%
- (d) 93.12%
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%

## Prática

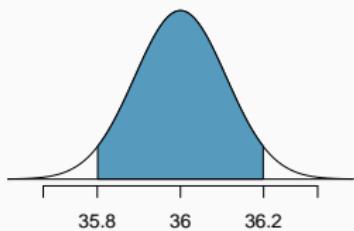
Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%

## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

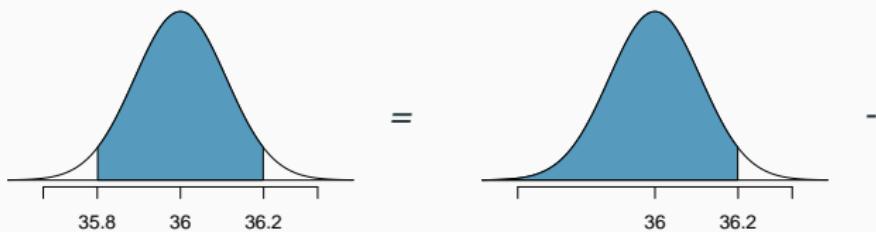
- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%



## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

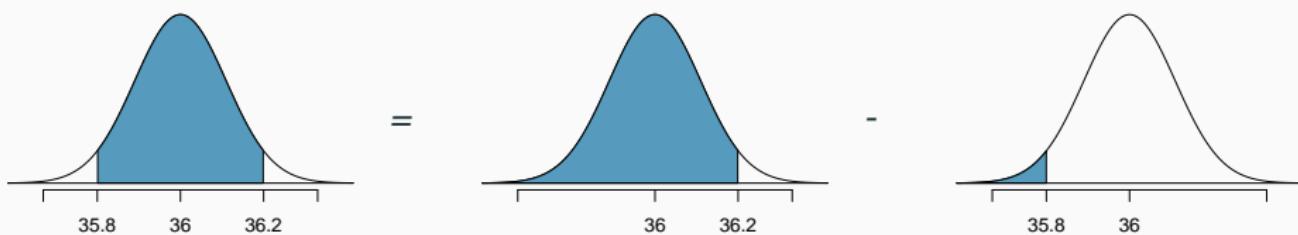
- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%



## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

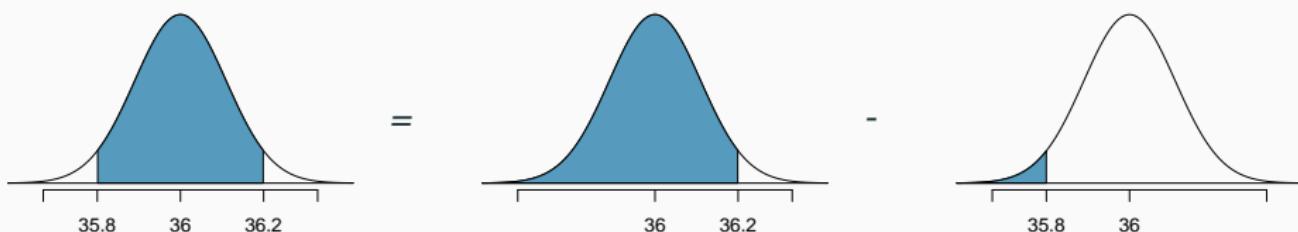
- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%



## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%

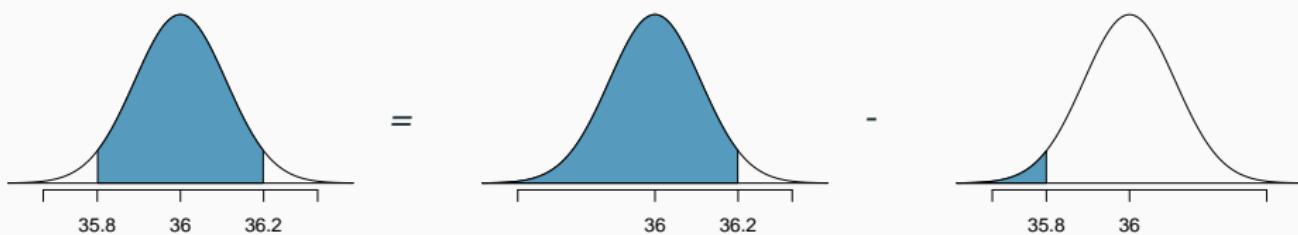


$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$

## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%



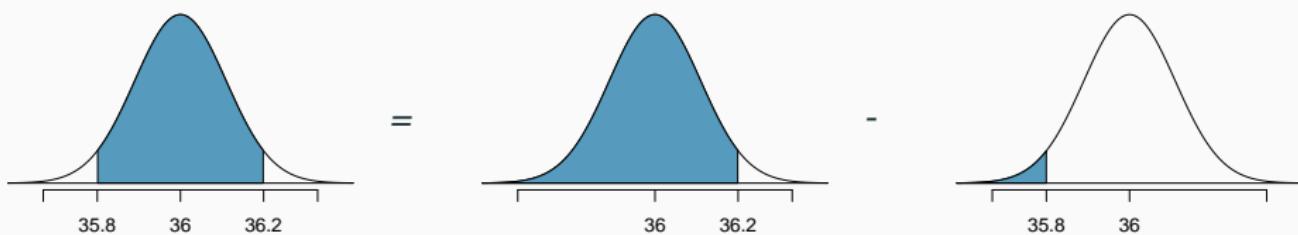
$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$

$$Z_{36.2} = \frac{36.2 - 36}{0.11} = 1.82$$

## Prática

Qual o percentual de garrafas que passa na inspeção de controle de qualidade?

- (a) 1.82%
- (d) **93.12%**
- (b) 3.44%
- (e) 96.56%
- (c) 6.88%



$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$

$$Z_{36.2} = \frac{36.2 - 36}{0.11} = 1.82$$

$$P(35.8 < X < 36.2) = P(-1.82 < Z < 1.82) = 0.9656 - 0.0344 = 0.9312$$

## Encontrando pontos de corte

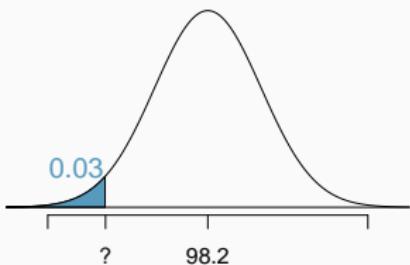
Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}$  F e desvio-padrão de  $0,73^{\circ}$  F.

Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?

## Encontrando pontos de corte

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}$  F e desvio-padrão de  $0,73^{\circ}$  F.

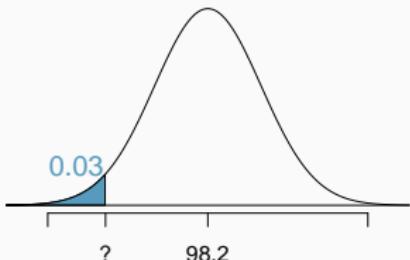
Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?



# Encontrando pontos de corte

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}$  F e desvio-padrão de  $0,73^{\circ}$  F.

Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?

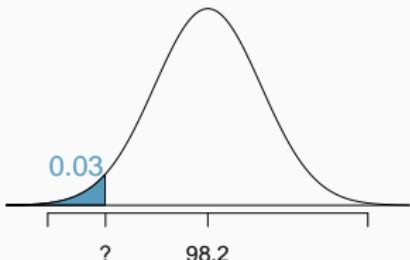


0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	-1.7

# Encontrando pontos de corte

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}$  F e desvio-padrão de  $0,73^{\circ}$  F.

Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?



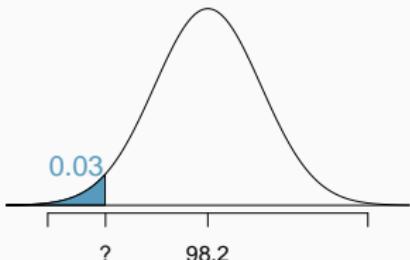
0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	-1.7

$$P(X < x) = 0.03 \rightarrow P(Z < -1.88) = 0.03$$

# Encontrando pontos de corte

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}$  F e desvio-padrão de  $0,73^{\circ}$  F.

Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?



0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	-1.7

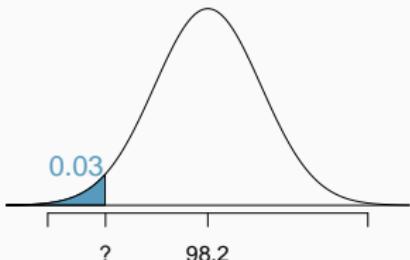
$$P(X < x) = 0.03 \rightarrow P(Z < -1.88) = 0.03$$

$$Z = \frac{obs - media}{SD} \rightarrow \frac{x - 98.2}{0.73} = -1.88$$

# Encontrando pontos de corte

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

Qual é o ponto de corte dos 3% que possuem temperaturas corporais mais baixas?



0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	-1.7

$$P(X < x) = 0.03 \rightarrow P(Z < -1.88) = 0.03$$

$$Z = \frac{\text{obs} - \text{media}}{SD} \rightarrow \frac{x - 98.2}{0.73} = -1.88$$

$$x = (-1.88 \times 0.73) + 98.2 = 96.8^{\circ}\text{F}$$

## Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

- (a)  $97,3^{\circ}\text{F}$
- (c)  $99,4^{\circ}\text{F}$
- (b)  $99,1^{\circ}\text{F}$
- (d)  $99,6^{\circ}\text{F}$

## Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

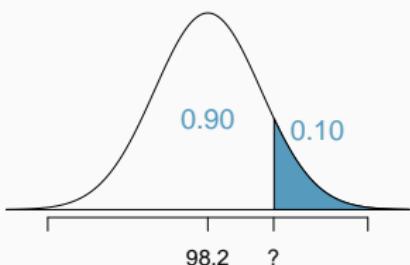
Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

(a)  $97.3^{\circ}\text{F}$

(b)  $99.1^{\circ}\text{F}$

(c)  $99.4^{\circ}\text{F}$

(d)  $99.6^{\circ}\text{F}$



## Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

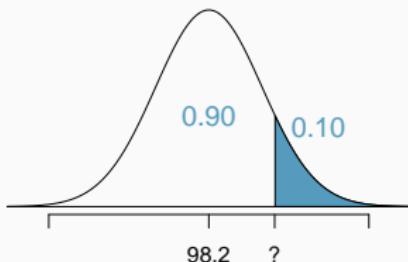
Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

(a)  $97.3^{\circ}\text{F}$

(b)  **$99.1^{\circ}\text{F}$**

(c)  $99.4^{\circ}\text{F}$

(d)  $99.6^{\circ}\text{F}$



Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

## Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

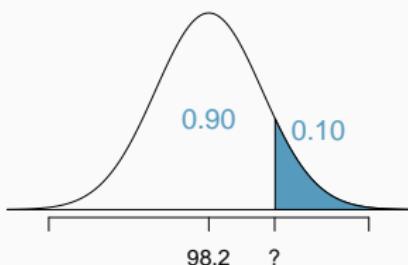
Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

(a)  $97.3^{\circ}\text{F}$

(b)  **$99.1^{\circ}\text{F}$**

(c)  $99.4^{\circ}\text{F}$

(d)  $99.6^{\circ}\text{F}$



Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

# Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

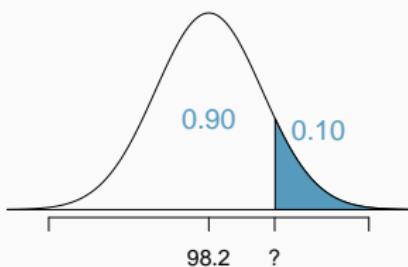
Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

(a)  $97.3^{\circ}\text{F}$

(b)  **$99.1^{\circ}\text{F}$**

(c)  $99.4^{\circ}\text{F}$

(d)  $99.6^{\circ}\text{F}$



Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

$$Z = \frac{obs - media}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

# Prática

Em seres humanos saudáveis, as temperaturas corporais são distribuídas quase que normalmente com média  $98,2^{\circ}\text{ F}$  e desvio-padrão de  $0,73^{\circ}\text{ F}$ .

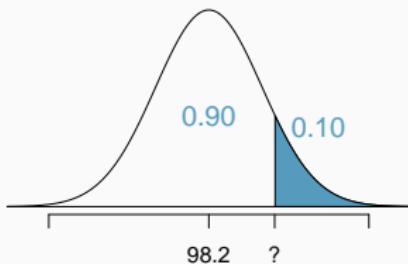
Qual é o ponto de corte dos 10% que possuem temperaturas corporais mais altas?

(a)  $97.3^{\circ}\text{F}$

(b)  **$99.1^{\circ}\text{F}$**

(c)  $99.4^{\circ}\text{F}$

(d)  $99.6^{\circ}\text{F}$



Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

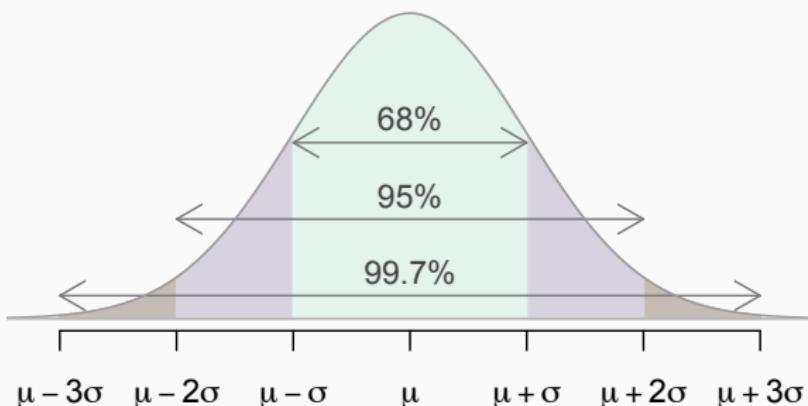
$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

$$Z = \frac{obs - media}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

$$x = (1.28 \times 0.73) + 98.2 = 99.1$$

## Regra 68-95-99.7

- Para dados quase normalmente distribuídos,
  - cerca de 68 % cai dentro de 1 SD da média,
  - cerca de 95 % cai dentro de 2 SD da média,
  - cerca de 99,7 % cai dentro de 3 SD da média.
- É possível que as observações caiam 4, 5 ou mais desvios-padrão da média, mas essas ocorrências são muito raras se os dados forem quase normais.



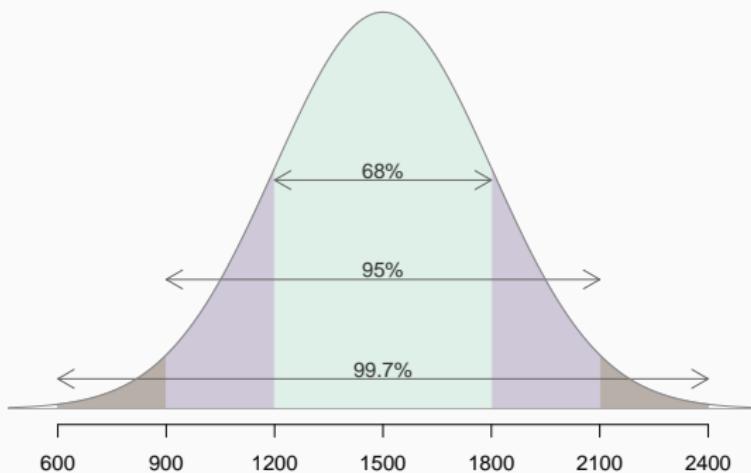
## Descrevendo a variabilidade usando a regra 68-95-99.7

Os escores do SAT são distribuídos quase que normalmente com média 1.500 e desvio padrão 300.

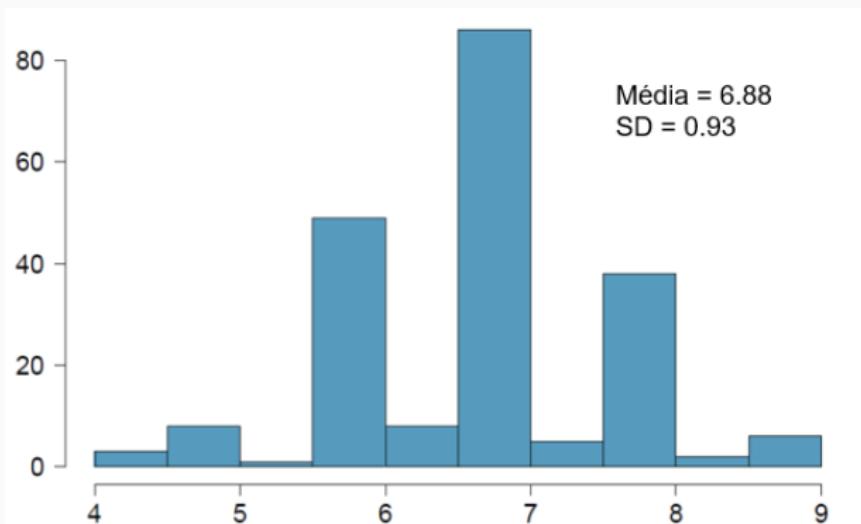
## Descrevendo a variabilidade usando a regra 68-95-99.7

Os escores do SAT são distribuídos quase que normalmente com média 1.500 e desvio padrão 300.

- ~68% dos alunos possuem pontuação entre 1200 e 1800 no SAT.
- ~95% dos alunos possuem pontuação entre 900 e 2100 no SAT.
- ~99.7% dos alunos possuem pontuação entre 600 e 2400 no SAT.



## Número de horas de sono de estudantes



- Média = 6.88 horas, SD = 0.92 horas

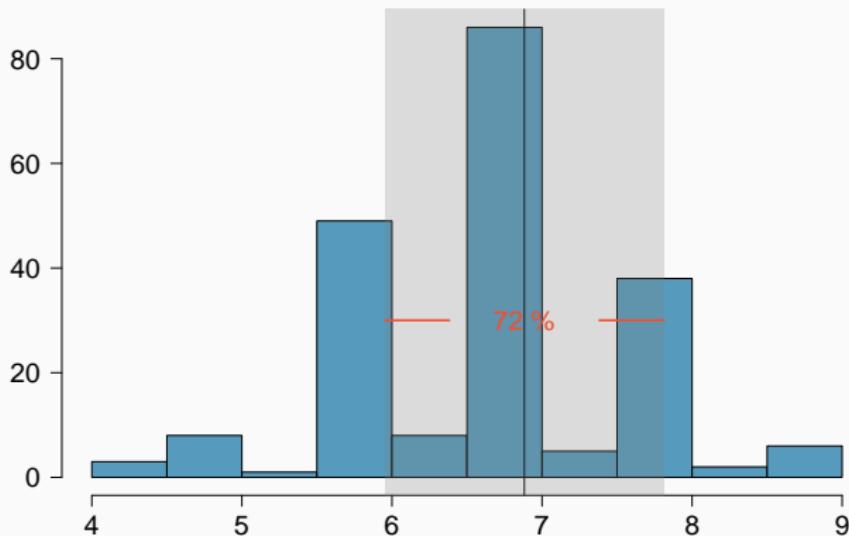
72% dos dados estão dentro de 1 SD da média,  $6.88 \pm 0.93$

92% dos dados estão dentro de 1 SD da média,

$6.88 \pm 2 \times 0.93$

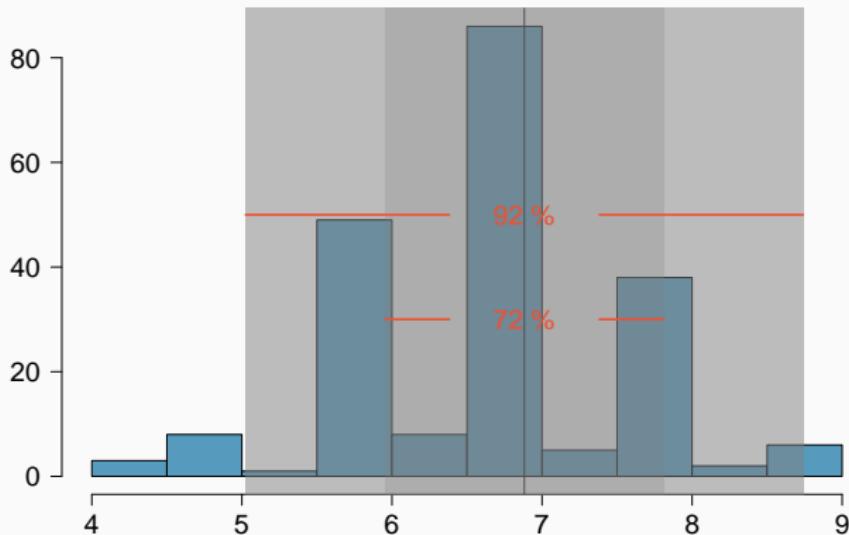
99% dos dados estão dentro de 1 SD da média:

## Número de horas de sono de estudantes



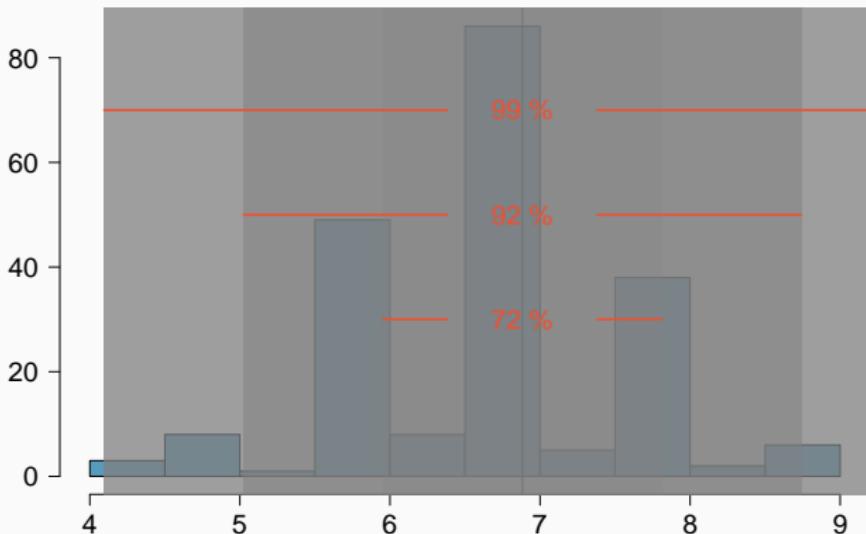
- Média = 6.88 horas, SD = 0.92 horas
- 72% dos dados estão dentro de 1 SD da média:  $6.88 \pm 0.93$
- 95% dos dados estão dentro de 2 SD da média:  $6.88 \pm 2 \times 0.93$
- 99% dos dados estão dentro de 3 SD da média:  $6.88 \pm 3 \times 0.93$

## Número de horas de sono de estudantes



- Média = 6.88 horas, SD = 0.92 horas
  - 72% dos dados estão dentro de 1 SD da média:  $6.88 \pm 0.93$
  - 92% dos dados estão dentro de 1 SD da média:  $6.88 \pm 2 \times 0.93$
- 99% dos dados estão dentro de 3 SD da média:  $6.88 \pm 3 \times 0.93$

## Número de horas de sono de estudantes



- Média = 6.88 horas, SD = 0.92 horas
- 72% dos dados estão dentro de 1 SD da média:  $6.88 \pm 0.93$
- 92% dos dados estão dentro de 1 SD da média:  $6.88 \pm 2 \times 0.93$
- 99% dos dados estão dentro de 1 SD da média:  $6.88 \pm 3 \times 0.93$

Qual das seguintes opções é falsa?

- (a) A maioria dos escores Z de uma distribuição assimétrica à direita são negativos.
- (b) Em distribuições assimétricas, o escore Z da média pode ser diferente de 0.
- (c) Para uma distribuição normal, o IQR é inferior a  $2 \times SD$ .
- (d) Os escores Z são úteis para determinar o quanto incomum é um ponto de dados quando comparado ao restante dos dados da distribuição.

Qual das seguintes opções é falsa?

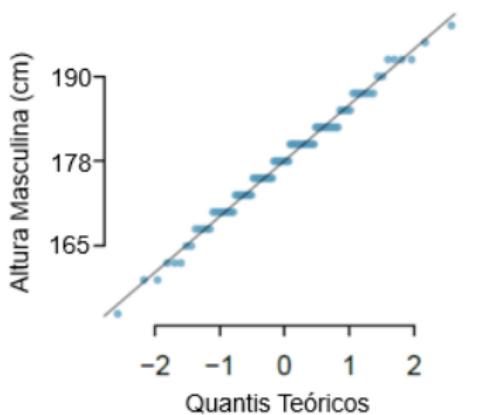
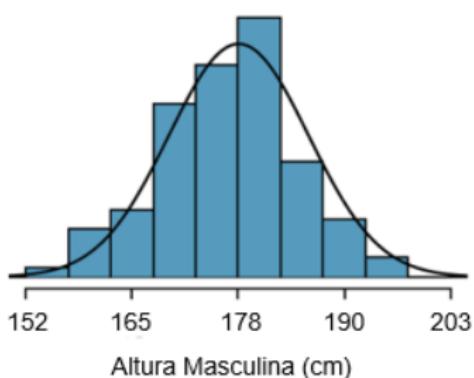
- (a) A maioria dos escores Z de uma distribuição assimétrica à direita são negativos.
- (b) *Em distribuições assimétricas, o escore Z da média pode ser diferente de 0.*
- (c) Para uma distribuição normal, o IQR é inferior a  $2 \times SD$ .
- (d) Os escores Z são úteis para determinar o quanto incomum é um ponto de dados quando comparado ao restante dos dados da distribuição.

## **3.2. Avaliando a aproximação normal**

---

## Gráfico de probabilidade normal

Um histograma e um *gráfico de probabilidade normal* de uma amostra de 100 alturas masculinas.

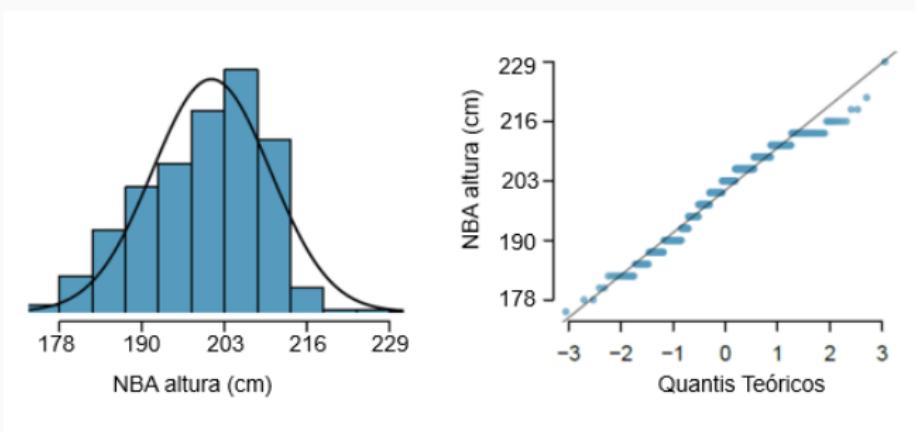


## Anatomia de um gráfico de probabilidade normal

- Os dados são plotados no eixo y de um gráfico de probabilidade normal e quantis teóricos (seguindo uma distribuição normal) no eixo x.
- Se houver um relacionamento linear no gráfico, os dados seguem uma distribuição quase normal.
- A construção de um gráfico de probabilidade normal requer o cálculo dos percentis e dos escores z correspondentes para cada observação, o que pode ser entediante. Portanto, geralmente dependemos de um software para fazer esses gráficos.

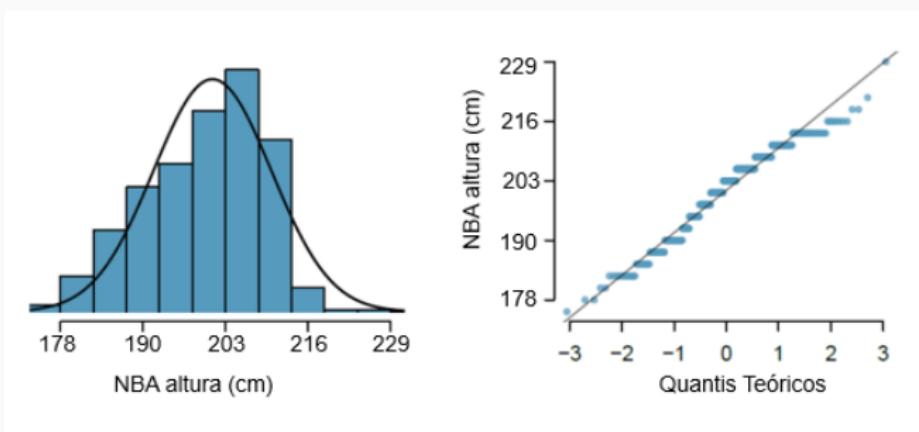
## Exemplo

Abaixo está um histograma e um gráfico de probabilidade normal para as alturas da NBA da temporada 2008-2009. Esses dados parecem seguir uma distribuição normal?



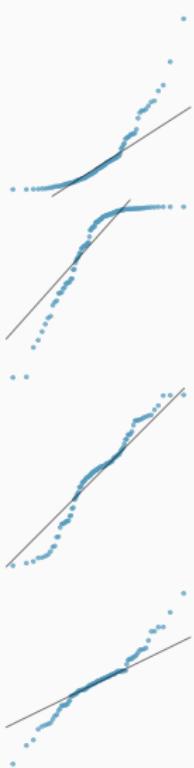
## Exemplo

Abaixo está um histograma e um gráfico de probabilidade normal para as alturas da NBA da temporada 2008-2009. Esses dados parecem seguir uma distribuição normal?



Por que os pontos de probabilidade normal têm saltos?

## Gráfico de probabilidade normal e assimetria



Inclinação à direita - Os pontos são curvados para cima e para a esquerda da linha.

Inclinação da esquerda - Pontos curvados para baixo e para a direita da linha.

Inclinação mais acentuada (mais estreitas que a distribuição normal) - Os pontos seguem uma curva em forma de S.

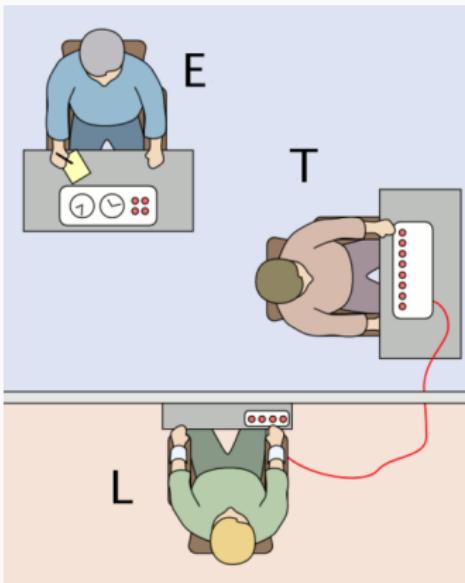
Inclinação menos acentuada (mais largas que a distribuição normal) - Os pontos começam abaixo da linha, passando a seguir a linha e terminam acima dela.

### **3.3. Distribuição geométrica**

---

# Experimento de Milgram

- Stanley Milgram, psicólogo da Universidade de Yale, conduziu uma série de experimentos sobre obediência à autoridade a partir de 1963.
- O pesquisador (E) ordena que o professor (T), que é o sujeito do experimento, aplique choques elétricos a um aluno (L) toda vez que este responder a uma pergunta incorretamente.



[http://en.wikipedia.org/wiki/File:Milgram\\_Experiment\\_v2.png](http://en.wikipedia.org/wiki/File:Milgram_Experiment_v2.png)

*Milgram\_Experiment\_v2.png*

## Experimento de Milgram (cont.)

- O aluno é na verdade um ator, e os choques elétricos não são reais, mas um som pré-gravado é tocado toda vez que o professor administra um choque elétrico.
- Esse experimento mediu a vontade dos participantes do estudo em obedecer a uma figura de autoridade que os instruiu a realizar atos que conflitassem com sua consciência pessoal.
- Milgram descobriu que cerca de 65% das pessoas obedeciam à autoridade e dariam tais choques.
- Ao longo dos anos, outras pesquisas sugeriram que esse número é aproximadamente consistente entre as comunidades e o tempo.

## Variáveis aleatórias de Bernoulli

- Na experiência de Milgram cada pessoa pode ser tratada como uma *tentativa*.
- A pessoa é considerada como *sucesso* se ela se recusar a administrar um choque e como *falha* se ela administrar o choque.
- Como apenas 35% das pessoas se recusaram a administrar o choque, a *probabilidade de sucesso* é  $p = 0.35$ .
- Quando uma particular tentativa tem apenas dois resultados possíveis, ela é chamada de *variável aleatória de Bernoulli*.

## Distribuição geométrica

A Dra. Smith quer repetir os experimentos de Milgram, mas ela só quer testar pessoas até encontrar alguém que não esteja disposto a aplicar choques. Qual é a probabilidade de ela parar o experimento depois da primeira pessoa?

$$P(1^{\text{a}} \text{ pessoa recusa}) = 0.35$$

## Distribuição geométrica

A Dra. Smith quer repetir os experimentos de Milgram, mas ela só quer testar pessoas até encontrar alguém que não esteja disposto a aplicar choques. Qual é a probabilidade de ela parar o experimento depois da primeira pessoa?

$$P(1^{\text{a}} \text{ pessoa recusa}) = 0.35$$

... a terceira pessoa?

$$P(1^{\text{a}} \text{ e } 2^{\text{a}} \text{ choque, } 3^{\text{a}} \text{ recusa}) = \frac{S}{0.65} \times \frac{S}{0.65} \times \frac{R}{0.35} = 0.65^2 \times 0.35 \approx 0.15$$

## Distribuição geométrica

A Dra. Smith quer repetir os experimentos de Milgram, mas ela só quer testar pessoas até encontrar alguém que não esteja disposto a aplicar choques. Qual é a probabilidade de ela parar o experimento depois da primeira pessoa?

$$P(1^{\text{a}} \text{ pessoa recusa}) = 0.35$$

... a terceira pessoa?

$$P(1^{\text{a}} \text{ e } 2^{\text{a}} \text{ choque, } 3^{\text{a}} \text{ recusa}) = \frac{S}{0.65} \times \frac{S}{0.65} \times \frac{R}{0.35} = 0.65^2 \times 0.35 \approx 0.15$$

# Distribuição geométrica

... a décima pessoa?

# Distribuição geométrica

... a décima pessoa?

$$P(9 \text{ choque}, 10^0 \text{ recusa}) = \underbrace{\frac{S}{0.65} \times \cdots \times \frac{S}{0.65}}_{9 \text{ destes}} \times \frac{R}{0.35} = 0.65^9 \times 0.35 \approx 0.0072$$

## Distribuição geométrica (cont.)

A *Distribuição geométrica* descreve o tempo de espera até um sucesso para variáveis aleatórias Bernoulli *independentes e identicamente distribuídas (iid)*.

## Distribuição geométrica (cont.)

A *Distribuição geométrica* descreve o tempo de espera até um sucesso para variáveis aleatórias Bernoulli *independentes e identicamente distribuídas (iid)*.

- independência: os resultados dos ensaios não afetam uns aos outros.
- idêntico: a probabilidade de sucesso é a mesma para cada tentativa.

## Distribuição geométrica (cont.)

A *Distribuição geométrica* descreve o tempo de espera até um sucesso para variáveis aleatórias Bernoulli *independentes e identicamente distribuídas (iid)*.

- independência: os resultados dos ensaios não afetam uns aos outros.
- idêntico: a probabilidade de sucesso é a mesma para cada tentativa.

### Probabilidades geométricas

Se  $p$  representa a probabilidade de sucesso, então  $(1 - p)$  representa a probabilidade de falha e  $n$  representa número de tentativas independentes

$$P(\text{sucesso na } n^{\text{esima}} \text{ tentativa}) = (1 - p)^{n-1} p$$

## Exemplo

Podemos calcular a probabilidade de obter um 6 pela primeira vez na sexta jogada de um dado utilizando a distribuição geométrica? Note que o sucesso (sair um 6) e o fracasso (não sair um 6) estão claramente definidos e um ou outro deve acontecer para cada tentativa.

- (a) não, no lançamento de um dado há mais de dois resultados possíveis
- (b) sim, pois

## Exemplo

Podemos calcular a probabilidade de obter um 6 pela primeira vez na sexta jogada de um dado utilizando a distribuição geométrica? Note que o sucesso (sair um 6) e o fracasso (não sair um 6) estão claramente definidos e um ou outro deve acontecer para cada tentativa.

- (a) não, no lançamento de um dado há mais de dois resultados possíveis
- (b) *sim, pois*

$$P(6 \text{ na } 6^{\text{a}} \text{ jogada}) = \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right) \approx 0.067$$

## Valor esperado

Quantas pessoas a Dra. Smith espera testar antes de encontrar a primeira que se recusará a administrar o choque?

## Valor esperado

Quantas pessoas a Dra. Smith espera testar antes de encontrar a primeira que se recusará a administrar o choque?

O valor esperado, ou a média, de uma distribuição geométrica é definida como  $\frac{1}{p}$ .

$$\mu = \frac{1}{p} = \frac{1}{0.35} = 2.86$$

## Valor esperado

Quantas pessoas a Dra. Smith espera testar antes de encontrar a primeira que se recusará a administrar o choque?

O valor esperado, ou a média, de uma distribuição geométrica é definida como  $\frac{1}{p}$ .

$$\mu = \frac{1}{p} = \frac{1}{0.35} = 2.86$$

Ela deve testar 2,86 pessoas antes de encontrar o primeiro que se recusa a administrar o choque.

## Valor esperado

Quantas pessoas a Dra. Smith espera testar antes de encontrar a primeira que se recusará a administrar o choque?

O valor esperado, ou a média, de uma distribuição geométrica é definida como  $\frac{1}{p}$ .

$$\mu = \frac{1}{p} = \frac{1}{0.35} = 2.86$$

Ela deve testar 2,86 pessoas antes de encontrar o primeiro que se recusa a administrar o choque.

Mas como ela pode testar um número não-inteiro de pessoas?

## Valor esperado e sua variabilidade

Média e desvio padrão da distribuição geométrica

$$\mu = \frac{1}{p} \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

## Valor esperado e sua variabilidade

Média e desvio padrão da distribuição geométrica

$$\mu = \frac{1}{p} \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

- Voltando à experiência da Dra. Smith:

$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.35}{0.35^2}} = 2.3$$

## Valor esperado e sua variabilidade

Média e desvio padrão da distribuição geométrica

$$\mu = \frac{1}{p} \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

- Voltando à experiência da Dra. Smith:

$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.35}{0.35^2}} = 2.3$$

- Espera-se que a Dra. Smith teste 2,86 pessoas antes de encontrar o primeiro que se recusa a administrar o choque, mais ou menos 2,3 pessoas.

## Valor esperado e sua variabilidade

### Média e desvio padrão da distribuição geométrica

$$\mu = \frac{1}{p} \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

- Voltando à experiência da Dra. Smith:

$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.35}{0.35^2}} = 2.3$$

- Espera-se que a Dra. Smith teste 2,86 pessoas antes de encontrar o primeiro que se recusa a administrar o choque, mais ou menos 2,3 pessoas.
- Esses valores só fazem sentido no contexto de repetir o experimento muitas vezes.

## **3.4. Distribuição binomial**

---

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

Vamos chamar essas pessoas de Allen (A), Bretanha (B), Caroline (C) e Damian (D). Cada um dos quatro cenários abaixo irá satisfazer a condição de "exatamente 1 deles se recusa a administrar o choque":

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

Vamos chamar essas pessoas de Allen (A), Bretanha (B), Caroline (C) e Damian (D). Cada um dos quatro cenários abaixo irá satisfazer a condição de "exatamente 1 deles se recusa a administrar o choque":

$$\text{Cenário 1: } \frac{0.35}{(\text{A}) \text{ recusa}} \times \frac{0.65}{(\text{B}) \text{ choque}} \times \frac{0.65}{(\text{C}) \text{ choque}} \times \frac{0.65}{(\text{D}) \text{ choque}} = 0.0961$$

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

Vamos chamar essas pessoas de Allen (A), Bretanha (B), Caroline (C) e Damian (D). Cada um dos quatro cenários abaixo irá satisfazer a condição de "exatamente 1 deles se recusa a administrar o choque":

Cenário 1:  $\frac{0.35}{(\text{A}) \text{ recusa}} \times \frac{0.65}{(\text{B}) \text{ choque}} \times \frac{0.65}{(\text{C}) \text{ choque}} \times \frac{0.65}{(\text{D}) \text{ choque}} = 0.0961$

Cenário 2:  $\frac{0.65}{(\text{A}) \text{ choque}} \times \frac{0.35}{(\text{B}) \text{ recusa}} \times \frac{0.65}{(\text{C}) \text{ choque}} \times \frac{0.65}{(\text{D}) \text{ choque}} = 0.0961$

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

Vamos chamar essas pessoas de Allen (A), Bretanha (B), Caroline (C) e Damian (D). Cada um dos quatro cenários abaixo irá satisfazer a condição de "exatamente 1 deles se recusa a administrar o choque":

Cenário 1:  $\frac{0.35}{(A) \text{ recusa}} \times \frac{0.65}{(B) \text{ choque}} \times \frac{0.65}{(C) \text{ choque}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

Cenário 2:  $\frac{0.65}{(A) \text{ choque}} \times \frac{0.35}{(B) \text{ recusa}} \times \frac{0.65}{(C) \text{ choque}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

Cenário 3:  $\frac{0.65}{(A) \text{ choque}} \times \frac{0.65}{(B) \text{ choque}} \times \frac{0.35}{(C) \text{ recusa}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

## Exemplo

Suponha que selecionamos aleatoriamente quatro indivíduos para participar desta experiência. Qual é a probabilidade de que exatamente um deles se recuse a administrar o choque?

Vamos chamar essas pessoas de Allen (A), Bretanha (B), Caroline (C) e Damian (D). Cada um dos quatro cenários abaixo irá satisfazer a condição de "exatamente 1 deles se recusa a administrar o choque":

Cenário 1:  $\frac{0.35}{(A) \text{ recusa}} \times \frac{0.65}{(B) \text{ choque}} \times \frac{0.65}{(C) \text{ choque}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

Cenário 2:  $\frac{0.65}{(A) \text{ choque}} \times \frac{0.35}{(B) \text{ recusa}} \times \frac{0.65}{(C) \text{ choque}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

Cenário 3:  $\frac{0.65}{(A) \text{ choque}} \times \frac{0.65}{(B) \text{ choque}} \times \frac{0.35}{(C) \text{ recusa}} \times \frac{0.65}{(D) \text{ choque}} = 0.0961$

## Exemplo

Cenário 4:  $\frac{0.65}{(\text{A}) \text{ choque}} \times \frac{0.65}{(\text{B}) \text{ choque}} \times \frac{0.65}{(\text{C}) \text{ choque}} \times \frac{0.35}{(\text{D}) \text{ recusa}} = 0.0961$

## Exemplo

$$\text{Cenário 4: } \frac{0.65}{(\text{A) choque}} \times \frac{0.65}{(\text{B) choque}} \times \frac{0.65}{(\text{C) choque}} \times \frac{0.35}{(\text{D) recusa}} = 0.0961$$

A probabilidade de exatamente 1 de 4 pessoas se recusarem a administrar o choque é a soma de todas essas probabilidades.

$$0.0961 + 0.0961 + 0.0961 + 0.0961 = 4 \times 0.0961 = 0.3844$$

## Distribuição binomial

A pergunta anterior pedia a probabilidade de um determinado número de sucessos,  $k$ , em um determinado número de tentativas,  $n$ , ( $k = 1$  sucesso em  $n = 4$  tentativas). Essa probabilidade é calculada como

$$\# \text{ de cenários} \times P(\text{único cenário})$$

## Distribuição binomial

A pergunta anterior pedia a probabilidade de um determinado número de sucessos,  $k$ , em um determinado número de tentativas,  $n$ , ( $k = 1$  sucesso em  $n = 4$  tentativas). Essa probabilidade é calculada como

$$\# \text{ de cenários} \times P(\text{único cenário})$$

- $\# \text{ de cenários}$ : há um modo mais direto de descobrir isso, discutiremos em breve...

## Distribuição binomial

A pergunta anterior pedia a probabilidade de um determinado número de sucessos,  $k$ , em um determinado número de tentativas,  $n$ , ( $k = 1$  sucesso em  $n = 4$  tentativas). Essa probabilidade é calculada como

$$\# \text{ de cenários} \times P(\text{único cenário})$$

- $\# \text{ de cenários}$ : há um modo mais direto de descobrir isso, discutiremos em breve...
- $P(\text{único cenário}) = p^k (1 - p)^{(n-k)}$

onde  $p$  é a probabilidade de sucesso elevada a quantidade de sucessos  $k$  e  $(1-p)$  é a probabilidade de fracasso elevada a quantidade de fracassos  $(n-k)$

## Distribuição binomial

A pergunta anterior pedia a probabilidade de um determinado número de sucessos,  $k$ , em um determinado número de tentativas,  $n$ , ( $k = 1$  sucesso em  $n = 4$  tentativas). Essa probabilidade é calculada como

$$\# \text{ de cenários} \times P(\text{único cenário})$$

- $\# \text{ de cenários}$ : há um modo mais direto de descobrir isso, discutiremos em breve...
- $P(\text{único cenário}) = p^k (1 - p)^{(n-k)}$

onde  $p$  é a probabilidade de sucesso elevada a quantidade de sucessos  $k$  e  $(p-1)$  é a probabilidade de fracassos elevada a quantidade de fracassos  $(n-k)$

A *Distribuição binomial* descreve a probabilidade de ter exatamente  $k$  sucessos em  $n$  tentativas de Bernoulli independentes com probabilidade de sucesso  $p$ .

## Contando o # de cenários

Anteriormente escrevemos todos os cenários possíveis que se encaixam na condição de exatamente uma pessoa se recusando a administrar o choque. Se  $n$  for maior e/ou  $k$  for diferente de 1, por exemplo,  $n = 9$  e  $k = 2$ :

## Contando o # de cenários

Anteriormente escrevemos todos os cenários possíveis que se encaixam na condição de exatamente uma pessoa se recusando a administrar o choque. Se  $n$  for maior e/ou  $k$  for diferente de 1, por exemplo,  $n = 9$  e  $k = 2$ :

**RRSSSSSSS**

## Contando o # de cenários

Anteriormente escrevemos todos os cenários possíveis que se encaixam na condição de exatamente uma pessoa se recusando a administrar o choque. Se  $n$  for maior e/ou  $k$  for diferente de 1, por exemplo,  $n = 9$  e  $k = 2$ :

$RR$ SSSSSSS

S $RR$ SSSSSS

## Contando o # de cenários

Anteriormente escrevemos todos os cenários possíveis que se encaixam na condição de exatamente uma pessoa se recusando a administrar o choque. Se  $n$  for maior e/ou  $k$  for diferente de 1, por exemplo,  $n = 9$  e  $k = 2$ :

$RR$ SSSSSSS

$SRR$ SSSSSS

SS $RR$ SSSSS

...

SS $RSSR$ SSS

...

SSSSSSS $RR$

Escrever todos os cenários possíveis seria extremamente tedioso e propenso a erros.

# Calculando o # de cenários

## Fórmula da combinação

A *fórmula da combinação* é útil para calcular o número de maneiras de obter  $k$  sucessos em  $n$  tentativas.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

# Calculando o # de cenários

## Fórmula da combinação

A *fórmula da combinação* é útil para calcular o número de maneiras de obter  $k$  sucessos em  $n$  tentativas.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- $k = 1, n = 4: \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4 \times 3 \times 2 \times 1}{1 \times (3 \times 2 \times 1)} = 4$

# Calculando o # de cenários

## Fórmula da combinação

A *fórmula da combinação* é útil para calcular o número de maneiras de obter  $k$  sucessos em  $n$  tentativas.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- $k = 1, n = 4: \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4 \times 3 \times 2 \times 1}{1 \times (3 \times 2 \times 1)} = 4$
- $k = 2, n = 9: \binom{9}{2} = \frac{9!}{2!(9-1)!} = \frac{9 \times 8 \times 7!}{2 \times 1 \times 7!} = \frac{72}{2} = 36$

---

> choose(9, 2)

[1] 36

**Note:** Você também pode usar R para esses cálculos.

## Propriedades da função escolher

Qual das seguintes afirmações é falsa?

- (a) Existem  $n$  maneiras de obter 1 sucesso em  $n$  tentativas,  $\binom{n}{1} = n$ .
- (b) Existe apenas uma maneira de obter  $n$  sucessos em  $n$  tentativas,  $\binom{n}{n} = 1$ .
- (c) Existe apenas uma maneira de obter fracassos de  $n$  em  $n$  tentativas,  $\binom{n}{0} = 1$ .
- (d) Existem  $n-1$  maneiras de obter  $n-1$  sucessos em  $n$  tentativas,  $\binom{n}{n-1} = n - 1$ .

## Propriedades da função escolher

Qual das seguintes afirmações é falsa?

- (a) Existem  $n$  maneiras de obter 1 sucesso em  $n$  tentativas,  $\binom{n}{1} = n$ .
- (b) Existe apenas uma maneira de obter  $n$  sucessos em  $n$  tentativas,  $\binom{n}{n} = 1$ .
- (c) Existe apenas uma maneira de obter fracassos de  $n$  em  $n$  tentativas,  $\binom{n}{0} = 1$ .
- (d) *Existem  $n-1$  maneiras de obter  $n-1$  sucessos em  $n$  tentativas,*  $\binom{n}{n-1} = n - 1$ .

## Distribuição binomial (cont.)

### Probabilidades Binomiais

Se  $p$  representa probabilidade de sucesso,  $(1 - p)$  representa probabilidade de fracasso,  $n$  representa número de tentativas independentes e  $k$  representa número de sucessos

$$P(k \text{ sucessos em } n \text{ tentativas}) = \binom{n}{k} p^k (1 - p)^{(n-k)}$$

## Prática

Qual das seguintes opções não é uma condição que precisa ser atendida para que a distribuição binomial seja aplicável?

- (a) os ensaios devem ser independentes.
- (b) o número de tentativas,  $n$ , deve ser fixo.
- (c) cada resultado de uma tentativa deve ser classificado como *sucesso* ou *fracasso*.
- (d) o número de sucessos desejados,  $k$ , deve ser maior que o número de tentativas.
- (e) a probabilidade de sucesso,  $p$ , deve ser a mesma para cada tentativa.

## Prática

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos. Entre uma amostra aleatória de 10 americanos, qual é a probabilidade de exatamente 8 serem obesos?

- (a) muito alta
- (b) muito baixa

Gallup: <http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx>, Janeiro 23, 2013.

## Prática

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos. Entre uma amostra aleatória de 10 americanos, qual é a probabilidade de exatamente 8 serem obesos?

- (a) muito alta
- (b) *muito baixa*

Gallup: <http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx>, Janeiro 23, 2013.

## Prática

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos. Entre uma amostra aleatória de 10 americanos, qual é a probabilidade de exatamente 8 serem obesos?

- (a)  $0.262^8 \times 0.738^2$
- (b)  $\binom{8}{10} \times 0.262^8 \times 0.738^2$
- (c)  $\binom{10}{8} \times 0.262^8 \times 0.738^2$
- (d)  $\binom{10}{8} \times 0.262^2 \times 0.738^8$

## Prática

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos. Entre uma amostra aleatória de 10 americanos, qual é a probabilidade de exatamente 8 serem obesos?

- (a)  $0.262^8 \times 0.738^2$
- (b)  $\binom{8}{10} \times 0.262^8 \times 0.738^2$
- (c)  $\binom{10}{8} \times 0.262^8 \times 0.738^2 = 45 \times 0.262^8 \times 0.738^2 = 0.0005$
- (d)  $\binom{10}{8} \times 0.262^2 \times 0.738^8$

## O problema do aniversário

Qual é a probabilidade de que 2 pessoas escolhidas aleatoriamente façam aniversário no mesmo dia?

## O problema do aniversário

Qual é a probabilidade de que 2 pessoas escolhidas aleatoriamente façam aniversário no mesmo dia?

Muito baixa,  $\frac{1}{365} \approx 0.0027$ .

## O problema do aniversário

Qual é a probabilidade de que 2 pessoas escolhidas aleatoriamente façam aniversário no mesmo dia?

Muito baixa,  $\frac{1}{365} \approx 0.0027$ .

Qual é a probabilidade de que pelo menos 2 pessoas de 366 pessoas façam aniversário no mesmo dia?

## O problema do aniversário

Qual é a probabilidade de que 2 pessoas escolhidas aleatoriamente façam aniversário no mesmo dia?

Muito baixa,  $\frac{1}{365} \approx 0.0027$ .

Qual é a probabilidade de que pelo menos 2 pessoas de 366 pessoas façam aniversário no mesmo dia?

Exatamente 1! (Excluindo a possibilidade de um aniversário de ano bissexto.)

## O problema do aniversário (cont.)

Qual é a probabilidade de que pelo menos 2 pessoas (1 correspondência) de 121 pessoas façam aniversário no mesmo dia?

## O problema do aniversário (cont.)

Qual é a probabilidade de que pelo menos 2 pessoas (1 correspondência) de 121 pessoas façam aniversário no mesmo dia?

Um pouco complicado de calcular, mas podemos pensar no complemento da probabilidade de não haver correspondências em 121 pessoas.

## O problema do aniversário (cont.)

$$\begin{aligned} P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\ &\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \end{aligned}$$

## O problema do aniversário (cont.)

$$\begin{aligned}P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\&\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}}\end{aligned}$$

## O problema do aniversário (cont.)

$$\begin{aligned}P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\&\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\&= \frac{365!}{365^{121} \times (365 - 121)!}\end{aligned}$$

## O problema do aniversário (cont.)

$$\begin{aligned}P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\&\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\&= \frac{365!}{365^{121} \times (365 - 121)!} \\&= \frac{121! \times \binom{365}{121}}{365^{121}}\end{aligned}$$

## O problema do aniversário (cont.)

$$\begin{aligned}P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\&\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\&= \frac{365!}{365^{121} \times (365 - 121)!} \\&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0\end{aligned}$$

## O problema do aniversário (cont.)

$$\begin{aligned}P(\text{não compartilham}) &= 1 \times \left(1 - \frac{1}{365}\right) \\&\quad \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\&= \frac{365!}{365^{121} \times (365 - 121)!} \\&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0\end{aligned}$$

$$P(\text{pelo menos 1 compartilha}) \approx 1$$

## Valor esperado

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos.

Entre uma amostra aleatória de 100 americanos, quantos você esperaria serem obesos?

## Valor esperado

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos.

Entre uma amostra aleatória de 100 americanos, quantos você esperaria serem obesos?

- é bastante fácil,  $100 \times 0.262 = 26.2$ .

## Valor esperado

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos.

Entre uma amostra aleatória de 100 americanos, quantos você esperaria serem obesos?

- é bastante fácil,  $100 \times 0.262 = 26.2$ .
- Ou mais formalmente,  $\mu = np = 100 \times 0.262 = 26.2$ .

## Valor esperado

Uma pesquisa de 2012 do Gallup sugere que 26,2 % dos americanos são obesos.

Entre uma amostra aleatória de 100 americanos, quantos você esperaria serem obesos?

- é bastante fácil,  $100 \times 0.262 = 26.2$ .
- Ou mais formalmente,  $\mu = np = 100 \times 0.262 = 26.2$ .
- Mas isso não significa que, em cada amostra aleatória de 100 pessoas, exatamente 26,2 serão obesas. Na verdade, isso nem é possível. Em algumas amostras, esse valor será menor e, em outros, maior. Quanto esperamos que esse valor varie?

# Valor esperado e sua variabilidade

Média e desvio padrão da distribuição binomial

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

# Valor esperado e sua variabilidade

Média e desvio padrão da distribuição binomial

$$\mu = np \qquad \sigma = \sqrt{np(1 - p)}$$

- Voltando à taxa de obesidade:

$$\sigma = \sqrt{np(1 - p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

# Valor esperado e sua variabilidade

## Média e desvio padrão da distribuição binomial

$$\mu = np \qquad \sigma = \sqrt{np(1 - p)}$$

- Voltando à taxa de obesidade:

$$\sigma = \sqrt{np(1 - p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

- Esperaríamos que 26,2 dos 100 americanos amostrados aleatoriamente fossem obesos, com um desvio-padrão de 4,4.

---

**Note:** A média e o desvio-padrão de uma binomial podem nem sempre ser números inteiros, e está tudo bem, esses valores representam o que esperamos ver em média.

## Observações incomuns

Usando o fato de que *observações com mais de 2 desvios-padrão da média são consideradas incomuns*, além da média e do desvio-padrão que acabamos de calcular, podemos obter um intervalo para o número plausível de obesos em amostras aleatórias de 100 americanos.

$$26.2 \pm (2 \times 4.4) = (17.4, 35)$$

# Prática

Em agosto de 2012, a pesquisa Gallup sugere que 13% dos americanos acham que o *home schooling* é uma excelente opção de educação para crianças. Uma amostra aleatória de 1.000 americanos, onde apenas 100 compartilham essa opinião, é considerada incomum?

(a) Não

(b) Sim

	Excelentes	Boas	Razoáveis	Ruins	Total Excelentes / Boas
	%	%	%	%	%
Escolas Privadas Independentes	31	47	13	02	78
Escolas relacionadas a religiões	21	48	18	05	69
Escola pública com lógica de gestão privada	17	43	23	05	60
Ensino por correspondência	13	33	30	14	46
Escolas Públicas	05	32	42	19	37

Gallup, Aug. 9-12, 2012

## Prática

Em agosto de 2012, a pesquisa Gallup sugere que 13% dos americanos acham que o *home schooling* é uma excelente opção de educação para crianças. Uma amostra aleatória de 1.000 americanos, onde apenas 100 compartilham essa opinião, é considerada incomum?

(a) Não

(b) *Sim*

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

<http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx>

## Prática

Em agosto de 2012, a pesquisa Gallup sugere que 13% dos americanos acham que o *home schooling* é uma excelente opção de educação para crianças. Uma amostra aleatória de 1.000 americanos, onde apenas 100 compartilham essa opinião, é considerada incomum?

(a) Não

(b) *Sim*

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Método 1: Intervalo de observações usuais:

$$130 \pm 2 \times 10.6 = (108.8, 151.2)$$

100 está fora desse intervalo, portanto, seria considerado incomum.

<http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx>

## Prática

Em agosto de 2012, a pesquisa Gallup sugere que 13% dos americanos acham que o *home schooling* é uma excelente opção de educação para crianças. Uma amostra aleatória de 1.000 americanos, onde apenas 100 compartilham essa opinião, é considerada incomum?

(a) Não

(b) *Sim*

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Método 1: Intervalo de observações usuais:

$$130 \pm 2 \times 10.6 = (108.8, 151.2)$$

100 está fora desse intervalo, portanto, seria considerado incomum.

Método 2: Z-score de observação:

$$Z = \frac{x - \text{mean}}{\text{SD}} = \frac{100 - 130}{10.6} = -2.83$$

100 é mais do que 2 SD abaixo da média, portanto, seria considerado incomum.

# Formatos de distribuições binomiais

## Formatos de distribuições binomiais

Para esta atividade, você usará um applet da web. Vá para [https://gallery.shinyapps.io/dist\\_calc/](https://gallery.shinyapps.io/dist_calc/) e escolha *Binomial coin experiment* no menu suspenso à esquerda.

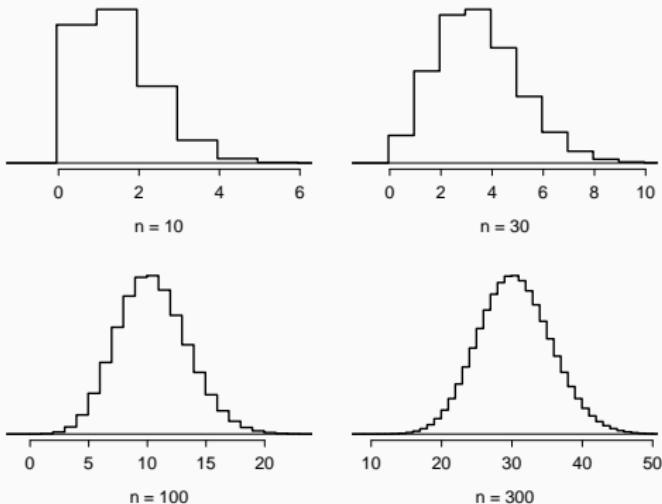
- Defina o número de tentativas para 20 e a probabilidade de sucesso para 0,15. Descreva a forma da distribuição do número de sucessos.
- Mantendo  $p$  constante em 0,15, determine o tamanho mínimo de amostra necessário para obter uma distribuição unimodal e simétrica do número de sucessos.

## Formatos de distribuições binomiais

- Outras considerações:
  - O que acontece com o formato da distribuição quando  $n$  permanece constante e  $p$  muda?
  - O que acontece com o formato da distribuição quando  $p$  permanece constante e  $n$  muda?

# Distribuições do número de sucessos

Histogramas de amostras do modelo binomial em que  $p = 0,10$  e  $n = 10, 30, 100$  e  $300$ . O que acontece quando  $n$  aumenta?



## Baixo grande é grande o suficiente?

O tamanho da amostra é considerado grande o suficiente se o número esperado de sucessos e fracassos forem pelo menos 10.

$$np \geq 10 \quad \text{e} \quad n(1 - p) \geq 10$$

## Baixo grande é grande o suficiente?

O tamanho da amostra é considerado grande o suficiente se o número esperado de sucessos e fracassos forem pelo menos 10.

$$np \geq 10 \quad \text{e} \quad n(1 - p) \geq 10$$

$$10 \times 0.13 = 1.3; 10 \times (1 - 0.13) = 8.7$$

## Prática

Abaixo estão quatro pares de parâmetros de distribuição binomial.  
Qual distribuição pode ser aproximada pela distribuição normal?

- (a)  $n = 100, p = 0.95$
- (b)  $n = 25, p = 0.45$
- (c)  $n = 150, p = 0.05$
- (d)  $n = 500, p = 0.015$

## Prática

Abaixo estão quatro pares de parâmetros de distribuição binomial.  
Qual distribuição pode ser aproximada pela distribuição normal?

- (a)  $n = 100, p = 0.95$
- (b)  $n = 25, p = 0.45 \rightarrow 25 \times 0.45 = 11.25; 25 \times 0.55 = 13.75$
- (c)  $n = 150, p = 0.05$
- (d)  $n = 500, p = 0.015$

## Uma análise dos usuários do Facebook

Um estudo recente descobriu que "os usuários do Facebook recebem mais do que dão". Por exemplo:

- 40% dos usuários do Facebook em nossa amostra fizeram uma solicitação de amizade, mas 63% receberam pelo menos uma solicitação de amizade.
- Os usuários em nossa amostra deram like para o conteúdo dos seus amigos em média 14 vezes, mas tiveram seu conteúdo "curtido" em média 20 vezes.
- Os usuários enviaram 9 mensagens pessoais, mas receberam 12.

## Uma análise dos usuários do Facebook

- 12% dos usuários marcaram um amigo em uma foto, mas 35% foram marcados em uma foto.

*Algum palpite de como esse padrão pode ser explicado?*

<http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx>

## Uma análise dos usuários do Facebook

- 12% dos usuários marcaram um amigo em uma foto, mas 35% foram marcados em uma foto.

*Algum palpite de como esse padrão pode ser explicado?*

*Usuários ativos contribuem com muito mais conteúdo do que o usuário padrão.*

<http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx>

## Prática

Este estudo também descobriu que aproximadamente 25% dos usuários do Facebook são considerados usuários ativos. O mesmo estudo descobriu que o usuário médio do Facebook tem 245 amigos. Qual é a probabilidade de que o usuário médio do Facebook com 245 amigos tenha 70 ou mais amigos que seriam considerados usuários ativos? Anote todas as suposições que você fizer.

Temos que  $n = 245$ ,  $p = 0.25$  e queremos saber a probabilidade  $P(K \geq 70)$ . Para prosseguir, precisamos de independência, o que vamos assumir, mas podemos tentar obter mais dados do Facebook.

## Prática

Este estudo também descobriu que aproximadamente 25% dos usuários do Facebook são considerados usuários ativos. O mesmo estudo descobriu que o usuário médio do Facebook tem 245 amigos. Qual é a probabilidade de que o usuário médio do Facebook com 245 amigos tenha 70 ou mais amigos que seriam considerados usuários ativos? Anote todas as suposições que você fizer.

Temos que  $n = 245$ ,  $p = 0.25$  e queremos saber a probabilidade  $P(K \geq 70)$ . Para prosseguir, precisamos de independência, o que vamos assumir, mas podemos tentar obter mais dados do Facebook.

## Prática

$$\begin{aligned}P(X \geq 70) &= P(K = 70 \text{ ou } K = 71 \text{ ou } K = 72 \text{ ou } \dots \text{ ou } K = 245) \\&= P(K = 70) + P(K = 71) + P(K = 72) + \dots + P(K = 245)\end{aligned}$$

## Prática

$$\begin{aligned}P(X \geq 70) &= P(K = 70 \text{ ou } K = 71 \text{ ou } K = 72 \text{ ou } \dots \text{ ou } K = 245) \\&= P(K = 70) + P(K = 71) + P(K = 72) + \dots + P(K = 245)\end{aligned}$$

Isso parece um monte de trabalho ...

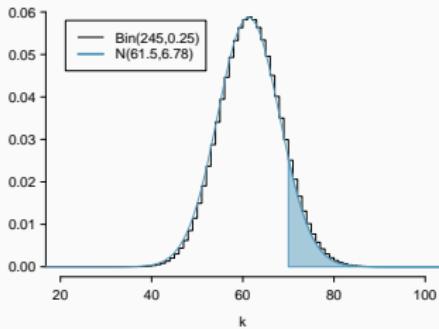
## Aproximação normal ao binômio

Quando o tamanho da amostra é grande o suficiente, a distribuição binomial de parâmetros  $n$  e  $p$  pode ser aproximada pelo modelo normal com os parâmetros  $\mu = np$  e  $\sigma = \sqrt{np(1 - p)}$ .

- No caso dos usuários avançados do Facebook,  $n = 245$  e  $p = 0.25$ .

$$\mu = 245 \times 0.25 = 61.25 \quad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$$

- $Bin(n = 245, p = 0.25) \approx N(\mu = 61.25, \sigma = 6.78)$ .

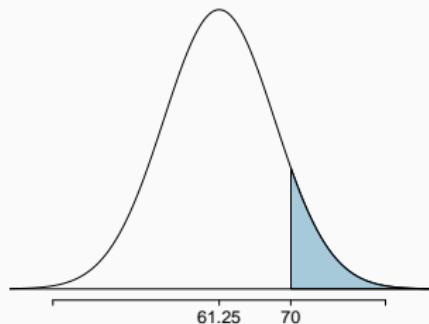


## Prática

Qual é a probabilidade de o usuário médio do Facebook com 245 amigos ter 70 ou mais amigos que seriam considerados usuários avançados?

## Prática

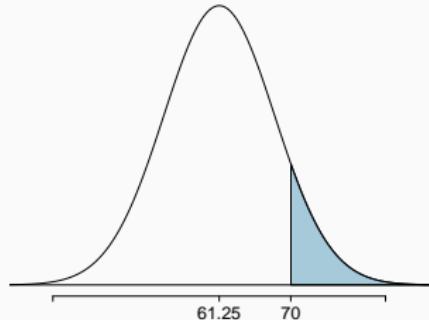
Qual é a probabilidade de o usuário médio do Facebook com 245 amigos ter 70 ou mais amigos que seriam considerados usuários avançados?



## Prática

Qual é a probabilidade de o usuário médio do Facebook com 245 amigos ter 70 ou mais amigos que seriam considerados usuários avançados?

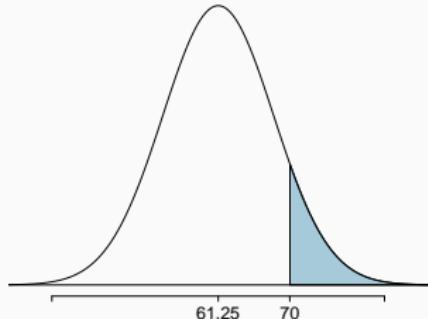
$$Z = \frac{obs - média}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$



## Prática

Qual é a probabilidade de o usuário médio do Facebook com 245 amigos ter 70 ou mais amigos que seriam considerados usuários avançados?

$$Z = \frac{obs - média}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

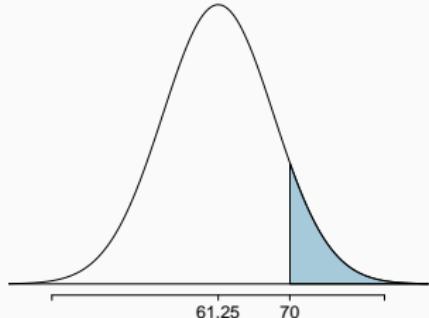


Z	Segunda casa decimal de Z				
	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015

## Prática

Qual é a probabilidade de o usuário médio do Facebook com 245 amigos ter 70 ou mais amigos que seriam considerados usuários avançados?

$$Z = \frac{obs - média}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$



Z	Segunda casa decimal de Z				
	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015

$$P(Z > 1.29) = 1 - 0.9015 = 0.0985$$

### **3.5. Outras distribuições discretas**

---

## Distribuição binomial negativa

- O *distribuição binomial negativa* descreve a probabilidade de observar o  $k^{\text{esimo}}$  sucesso na  $n^{\text{esimo}}$  tentativa.
- As quatro condições a seguir são úteis para identificar um caso binomial negativo:
  1. As tentativas são independentes.
  2. Cada resultado do estudo pode ser classificado como um sucesso ou fracasso.
  3. A probabilidade de sucesso ( $p$ ) é a mesma para cada tentativa.
  4. A última tentativa deve ser um sucesso.

Observe que as três primeiras condições são comuns à distribuição binomial.

# Distribuição binomial negativa

## Distribuição binomial negativa

$P(k^{\text{esimo}} \text{ sucesso na } n^{\text{esimo}} \text{ tentativa}) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$ ,  
onde  $p$  é a probabilidade de que uma tentativa individual seja um sucesso. Todas as tentativas são consideradas independentes.

## Prática

Uma estudante universitária que trabalha em um laboratório de psicologia é solicitada a recrutar 10 casais para participar de um estudo. Ela decide ficar do lado de fora do centro estudantil e perguntar a cada 5 pessoas que saem do prédio se eles estão em um relacionamento e, em caso afirmativo, se eles gostariam de participar do estudo com o seu parceiro. Suponha que a probabilidade de encontrar tal pessoa seja 10%. Qual é a probabilidade da estudante precisar perguntar a 30 pessoas antes de atingir seu objetivo?

## Prática

Uma estudante universitária que trabalha em um laboratório de psicologia é solicitada a recrutar 10 casais para participar de um estudo. Ela decide ficar do lado de fora do centro estudantil e perguntar a cada 5 pessoas que saem do prédio se eles estão em um relacionamento e, em caso afirmativo, se eles gostariam de participar do estudo com o seu parceiro. Suponha que a probabilidade de encontrar tal pessoa seja 10%. Qual é a probabilidade da estudante precisar perguntar a 30 pessoas antes de atingir seu objetivo?

Dado:  $p = 0,10$ ,  $k = 10$ ,  $n = 30$ . Queremos encontrar a probabilidade do 10º sucesso na 30ª tentativa, portanto usamos a distribuição binomial negativa.

## Prática

Uma estudante universitária que trabalha em um laboratório de psicologia é solicitada a recrutar 10 casais para participar de um estudo. Ela decide ficar do lado de fora do centro estudantil e perguntar a cada 5 pessoas que saem do prédio se eles estão em um relacionamento e, em caso afirmativo, se eles gostariam de participar do estudo com o seu parceiro. Suponha que a probabilidade de encontrar tal pessoa seja 10%. Qual é a probabilidade da estudante precisar perguntar a 30 pessoas antes de atingir seu objetivo?

Dado:  $p = 0,10$ ,  $k = 10$ ,  $n = 30$ . Queremos encontrar a probabilidade do 10º sucesso na 30ª tentativa, portanto usamos a distribuição binomial negativa.

## Prática

$$P(10^{\text{o}} \text{ sucesso na } 30^{\text{a}} \text{ tentativa}) = \binom{29}{9} \times 0.10^{10} \times 0.90^{20}$$

## Prática

$$\begin{aligned}P(10^{\text{o}} \text{ sucesso na } 30^{\text{a}} \text{ tentativa}) &= \binom{29}{9} \times 0.10^{10} \times 0.90^{20} \\&= 10,015,005 \times 0.10^{10} \times 0.90^{20}\end{aligned}$$

## Prática

$$\begin{aligned}P(10^{\text{o}} \text{ sucesso na } 30^{\text{a}} \text{ tentativa}) &= \binom{29}{9} \times 0.10^{10} \times 0.90^{20} \\&= 10,015,005 \times 0.10^{10} \times 0.90^{20} \\&= 0.00012\end{aligned}$$

## **Binomial vs. binomial negativo**

Como a distribuição binomial negativa é diferente da distribuição binomial?

## Binomial vs. binomial negativo

Como a distribuição binomial negativa é diferente da distribuição binomial?

- No caso binomial, normalmente temos um número fixo de tentativas ao invés de considerar o número de sucessos.
- No caso da binomial negativa, examinamos quantas tentativas são necessárias para observar um número fixo de sucessos, exigindo que a última observação seja um sucesso.

## Prática

Qual das seguintes opções descreve um caso em que usaríamos a distribuição binomial negativa para calcular a probabilidade desejada?

- (a) Probabilidade de que um menino de 5 anos seja mais alto que 106 centímetros.
- (b) Probabilidade de que 3 de 10 arremessos de softbol sejam bem sucedidos.
- (c) Probabilidade de receber uma sequência de mesmo naipe no poker.
- (d) Probabilidade de perder 8 tiros antes do primeiro acerto.
- (e) Probabilidade de acertar a bola pela 3<sup>a</sup> vez na 8<sup>a</sup> tentativa .

## Prática

Qual das seguintes opções descreve um caso em que usaríamos a distribuição binomial negativa para calcular a probabilidade desejada?

- (a) Probabilidade de que um menino de 5 anos seja mais alto que 106 centímetros.
- (b) Probabilidade de que 3 de 10 arremessos de softbol sejam bem sucedidos.
- (c) Probabilidade de receber uma sequência de mesmo naipe no poker.
- (d) Probabilidade de perder 8 tiros antes do primeiro acerto.
- (e) *Probabilidade de acertar a bola pela 3<sup>a</sup> vez na 8<sup>a</sup> tentativa .*

## Distribuição de Poisson

- A *distribuição de Poisson* é útil para estimar o número de eventos raros por unidade de tempo para uma população grande e fixa, dado que os indivíduos dentro dessa população são independentes.
- A *taxa* para uma distribuição de Poisson é o número médio de ocorrências em uma população fixa por unidade de tempo, e é denotada por  $\lambda$ .
- Usando a taxa, podemos descrever a probabilidade de observar exatamente  $k$  eventos por unidade de tempo.

# Distribuição de Poisson

## Distribuição de Poisson

$$P(\text{observar } k \text{ eventos}) = \frac{\lambda^k e^{-\lambda}}{k!},$$

onde  $k$  pode ser 0, 1, 2, e assim por diante, e  $k!$  representa  $k$ -fatorial. A letra  $e \approx 2.718$  é a base do logaritmo natural.

A média e o desvio-padrão dessa distribuição são  $\lambda$  e  $\sqrt{\lambda}$ , respectivamente.

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que, em uma determinada semana, a eletricidade falhe apenas uma vez.

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que, em uma determinada semana, a eletricidade falhe apenas uma vez.

Dado  $\lambda = 2$ .

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que, em uma determinada semana, a eletricidade falhe apenas uma vez.

Dado  $\lambda = 2$ .

$$P(\text{apenas 1 falha em uma semana}) = \frac{2^1 \times e^{-2}}{1!}$$

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que, em uma determinada semana, a eletricidade falhe apenas uma vez.

Dado  $\lambda = 2$ .

$$\begin{aligned} P(\text{apenas 1 falha em uma semana}) &= \frac{2^1 \times e^{-2}}{1!} \\ &= \frac{2 \times e^{-2}}{1} \end{aligned}$$

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que, em uma determinada semana, a eletricidade falhe apenas uma vez.

Dado  $\lambda = 2$ .

$$\begin{aligned} P(\text{apenas 1 falha em uma semana}) &= \frac{2^1 \times e^{-2}}{1!} \\ &= \frac{2 \times e^{-2}}{1} \\ &= 0.27 \end{aligned}$$

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que em um determinado dia a eletricidade falhe três vezes.

## Prática

Suponha que falhas de energia elétrica ocorram em uma região rural de um país em desenvolvimento, seguindo uma distribuição de Poisson com média de 2 falhas a cada semana.

Calcule a probabilidade de que em um determinado dia a eletricidade falhe três vezes.

Temos a taxa de falha semanal, mas para responder a essa pergunta precisamos primeiro calcular a taxa média de falha em um determinado dia:  $\lambda_{dia} = \frac{2}{7} = 0.2857$ . Note que estamos assumindo que a probabilidade de falha de energia é a mesma em qualquer dia da semana, ou seja, assumimos independência.

## Prática

$$P(3 \text{ falhas em um determinado dia}) = \frac{0.2857^1 \times e^{-0.2857}}{3!}$$

## Prática

$$\begin{aligned} P(3 \text{ falhas em um determinado dia}) &= \frac{0.2857^1 \times e^{-0.2857}}{3!} \\ &= \frac{0.2857 \times e^{-0.2857}}{6} \end{aligned}$$

## Prática

$$\begin{aligned}P(3 \text{ falhas em um determinado dia}) &= \frac{0.2857^1 \times e^{-0.2857}}{3!} \\&= \frac{0.2857 \times e^{-0.2857}}{6} \\&= 0.0358\end{aligned}$$

## É Poisson?

- Uma variável aleatória pode seguir uma distribuição de Poisson se o evento que está sendo considerado for raro, a população for grande e os eventos ocorrerem independentemente um do outro.
- No entanto, podemos pensar em situações em que os eventos não são realmente independentes. Por exemplo, se estivermos interessados na probabilidade de um certo número de casamentos durante um verão, devemos levar em consideração que fins de semana são mais populares para casamentos.
- Nesse caso, um modelo de Poisson pode às vezes ainda ser razoável se permitirmos que ele tenha uma taxa diferente para tempos diferentes; poderíamos modelar a taxa como maior nos finais de semana do que nos dias da semana.

## É Poisson?

- A ideia de modelar taxas para uma distribuição de Poisson em relação a uma segunda variável (dia da semana) forma a base de alguns métodos mais avançados chamados *modelos lineares generalizados*. Isso vai além do escopo deste curso, mas discutiremos uma base de modelos lineares nos Capítulos 7 e 8.

## Prática

Qual variável aleatória que segue as seguintes distribuições pode assumir valores diferentes de inteiros positivos?

- (a) Poisson
- (b) Binomial negativo
- (c) Binomial
- (d) Normal
- (e) Geométrico

## Prática

Qual variável aleatória que segue as seguintes distribuições pode assumir valores diferentes de inteiros positivos?

- (a) Poisson
- (b) Binomial negativo
- (c) Binomial
- (d) *Normal*
- (e) Geométrico