



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 4: Fundamentos para inferência

Slides desenvolvidos por Mine Çetinkaya-Rundel of OpenIntro.

Os slides podem ser copiados, editados e / ou compartilhados via CC BY-SA license.

Algumas imagens podem ser incluídas em diretrizes de uso justo (propósitos educacionais).

4.1. Estimativas pontuais e variabilidade de amostragem

Estimativas pontuais e erro

- Frequentemente estamos interessados em *parâmetros populacionais*.
- É difícil coletar todos os dados de uma população (fazer um censo). Por isso usamos *estatísticas da amostra* como *estimativas pontuais* dos parâmetros desconhecidos da população de interesse.
- *Erro* na estimativa = diferença entre o parâmetro da população e a estatística da amostra.
- *Viés* é uma tendência sistemática que super ou subestima o verdadeiro parâmetro da população.
- *Erro de amostragem* descreve o quanto uma estimativa tende a variar de uma amostra para outra.
- Grande parte das estatísticas concentra-se em compreender e quantificar o erro de amostragem, e o *tamanho da amostra* é útil para quantificar esse erro.

Suponha que nós amostramos aleatoriamente 1.000 adultos de cada estado nos EUA. Você esperaria que as médias amostrais de suas alturas sejam as mesmas, um pouco diferentes ou muito diferentes?

Suponha que nós amostramos aleatoriamente 1.000 adultos de cada estado nos EUA. Você esperaria que as médias amostrais de suas alturas sejam as mesmas, um pouco diferentes ou muito diferentes?

Esperamos que as médias amostrais não sejam as mesmas, mas apenas um pouco diferentes.

Young, Underemployed and Optimistic

Coming of Age, Slowly, in a Tough Economy

Young adults hit hard by the recession. A plurality of the public (41%) believes young adults, rather than middle-aged or older adults, are having the toughest time in today's economy. An analysis of government economic data suggests that this perception is correct. The recent indicators on the nation's labor market show a decline in the

Tough economic times altering young adults' daily lives, long-term plans. While negative trends in the labor market have been felt most acutely by the youngest workers, many adults in their late 20s and early 30s have also felt the impact of the weak economy. Among all 18- to 34-year-olds, fully half (49%) say they have taken a job they didn't want just to pay the bills, with 24% saying they have taken an unpaid job to gain work experience. And more than one-third (35%) say that, as a result of the poor economy, they have gone back to school. Their personal lives have also been affected: 31% have postponed either getting married or having a baby (22% say they have postponed having a baby and 20% have put off getting married). One-in-four (24%) say they have moved back in with their parents after living on their own.

Tradução do texto: Young, Underemployed and Optimistic

Jovem, Subempregado e Otimista

Envelhecer, lentamente, em uma economia difícil.

Jovens adultos são duramente atingidos pela recessão. *Uma grande parte do público (41%) acredita que os jovens adultos, ao invés dos adultos de meia-idade ou mais os velhos, estão tendo o momento mais difícil na economia atual.* Uma análise dos dados econômicos do governo sugere que essa percepção está correta. Os indicadores recentes sobre o mercado de trabalho da nação mostram um declínio na taxa de desemprego. No entanto, desde 2010, a proporção de jovens adultos de 18 a 24 anos atualmente empregados (54%) é a menor desde que o governo começou a coletar esses dados em 1948. E a diferença no emprego entre os jovens e todos os adultos em idade ativa - 15 pontos percentuais - é o mais amplo da registrado na história. Além disso, os jovens adultos empregados em período integral tiveram uma queda maior nos seus ganhos semanais (menos 6%) do que qualquer outro grupo etário nos últimos quatro anos.

Tradução do texto: Young, Underemployed and Optimistic

Tempos econômicos difíceis que alteram a vida cotidiana dos jovens adultos e seus planos de longo prazo. Embora as tendências negativas no mercado de trabalho tenham sido mais fortemente sentidas pelos trabalhadores mais jovens, muitos adultos com cerca de 20 e 30 anos também sentiram o impacto da fraca economia. *Entre todas as pessoas de 18 a 34 anos, metade (49%) diz que aceitou um trabalho que não queria apenas para poder pagar as contas, com 24% dizendo que aceitou um emprego não remunerado para ganhar experiência profissional.* E mais de um terço (35%) dizem que, como resultado da economia pobre, voltaram à escola. Suas vidas pessoais também foram afetadas: 31% adiaram o casamento ou a gravidez (22% disseram que adiaram ter um bebê e 20% adiaram o casamento). Um em quatro (24%) dizem que voltaram a morar com os pais depois de viverem sozinhos.

Margem de erro

The general public survey is based on telephone interviews conducted Dec. 6-19, 2011, with a nationally representative sample of 2,048 adults ages 18 and older living in the continental United States, including an oversample of 346 adults ages 18 to 34. A total of 769 interviews were completed with respondents contacted by landline telephone and 1,279 with those contacted on their cellular phone. Data are weighted to produce a final sample that is representative of the general population of adults in the continental United States. Survey interviews were conducted under the direction of Princeton Survey Research Associates International, in English and Spanish. Margin of sampling error is plus or minus 2.9 percentage points for results based on the total sample and 4.4 percentage points for adults ages 18-34 at the 95% confidence level.

Tradução do texto: Margem de Erro

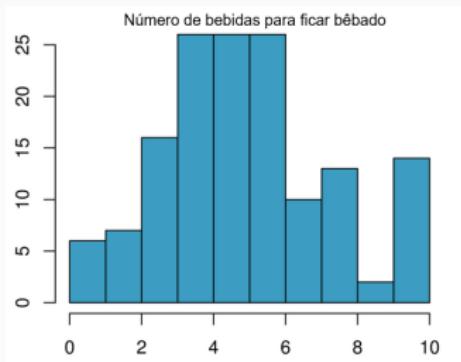
A pesquisa do público geral é baseada em entrevistas telefônicas realizadas de 6 a 19 de dezembro de 2011, com uma amostra nacionalmente representativa de 2.048 adultos com 18 anos ou mais vivendo nos Estados Unidos continental, incluindo uma sobreamostra de 346 adultos entre 18 e 34 anos. Um total de 769 entrevistas foram realizadas para os telefones fixos e 1.279 para os telefones celulares das pessoas. Os dados são ponderados para produzir uma amostra final que seja representativa da população geral de adultos nos Estados Unidos continental. As entrevistas da pesquisa foram conduzidas sob a direção da Princeton Survey Research Associates International, em inglês e espanhol. *A margem de erro de amostragem é de mais ou menos 2,9 pontos percentuais para os resultados com base na amostra total e 4,4 pontos percentuais para os adultos com idades entre 18 e 34 anos, com nível de confiança de 95%.*

Margem de erro

- $41\% \pm 2.9\%$: Estamos 95% confiantes de que entre 38.1% e 43.9% do público acreditam que os jovens adultos, em vez de adultos de meia-idade ou mais velhos, estão tendo o momento mais difícil na economia atual.
- $49\% \pm 4.4\%$: Estamos 95% confiantes de que entre 44.6% e 53.4% dos jovens entre 18 e 34 anos aceitaram um emprego que não queriam apenas para poder pagar as contas.

Exercício de aplicação

O histograma a seguir mostra a distribuição do quantidade de bebidas que um grupo de estudantes universitários precisa ingerir para ficarem bêbados. Vamos supor que esta é a nossa população de interesse. Se selecionarmos aleatoriamente observações desse conjunto de dados, quais valores têm maior probabilidade de serem selecionados? E quais são os menos prováveis?



Exercício de aplicação

1	7	16	3	31	5	46	4	61	10	76	6	91	4	106	6	121	6	136	6
2	5	17	10	32	9	47	3	62	7	77	6	92	0.5	107	2	122	5	137	7
3	4	18	8	33	7	48	3	63	4	78	5	93	3	108	5	123	3	138	3
4	4	19	5	34	5	49	6	64	5	79	4	94	3	109	1	124	2	139	10
5	6	20	10	35	5	50	8	65	6	80	5	95	5	110	5	125	2	140	4
6	2	21	6	36	7	51	8	66	6	81	6	96	6	111	5	126	5	141	4
7	3	22	2	37	4	52	8	67	6	82	5	97	4	112	4	127	10	142	6
8	5	23	6	38	0	53	2	68	7	83	6	98	4	113	4	128	4	143	6
9	5	24	7	39	4	54	4	69	7	84	8	99	2	114	9	129	1	144	4
10	6	25	3	40	3	55	8	70	5	85	4	100	5	115	4	130	4	145	5
11	1	26	6	41	6	56	3	71	10	86	10	101	4	116	3	131	10	146	5
12	10	27	5	42	10	57	5	72	3	87	5	102	7	117	3	132	8		
13	4	28	8	43	3	58	5	73	5.5	88	10	103	6	118	4	133	10		
14	4	29	0	44	6	59	8	74	7	89	8	104	8	119	4	134	6		
15	6	30	8	45	10	60	4	75	10	90	5	105	3	120	8	135	6		

Exercício de aplicação

Exemplo:

Lista de números aleatórios: 59, 121, 88, 46, 58, 72, 82, 81, 5, 10

1	7	16	3	31	5	46	4	61	10	76	6	91	4	106	6	121	6	136	6
2	5	17	10	32	9	47	3	62	7	77	6	92	0.5	107	2	122	5	137	7
3	4	18	8	33	7	48	3	63	4	78	5	93	3	108	5	123	3	138	3
4	4	19	5	34	5	49	6	64	5	79	4	94	3	109	1	124	2	139	10
5	6	20	10	35	5	50	8	65	6	80	5	95	5	110	5	125	2	140	4
6	2	21	6	36	7	51	8	66	6	81	6	96	6	111	5	126	5	141	4
7	3	22	2	37	4	52	8	67	6	82	5	97	4	112	4	127	10	142	6
8	5	23	6	38	0	53	2	68	7	83	6	98	4	113	4	128	4	143	6
9	5	24	7	39	4	54	4	69	7	84	8	99	2	114	9	129	1	144	4
10	6	25	3	40	3	55	8	70	5	85	4	100	5	115	4	130	4	145	5
11	1	26	6	41	6	56	3	71	10	86	10	101	4	116	3	131	10	146	5
12	10	27	5	42	10	57	5	72	3	87	5	102	7	117	3	132	8		
13	4	28	8	43	3	58	5	73	5.5	88	10	103	6	118	4	133	10		
14	4	29	0	44	6	59	8	74	7	89	8	104	8	119	4	134	6		
15	6	30	8	45	10	60	4	75	10	90	5	105	3	120	8	135	6		

Exercício de aplicação

Exemplo:

Lista de números aleatórios: 59, 121, 88, 46, 58, 72, 82, 81, 5, 10

1	7	16	3	31	5	46	4	61	10	76	6	91	4	106	6	121	6	136	6
2	5	17	10	32	9	47	3	62	7	77	6	92	0.5	107	2	122	5	137	7
3	4	18	8	33	7	48	3	63	4	78	5	93	3	108	5	123	3	138	3
4	4	19	5	34	5	49	6	64	5	79	4	94	3	109	1	124	2	139	10
5	6	20	10	35	5	50	8	65	6	80	5	95	5	110	5	125	2	140	4
6	2	21	6	36	7	51	8	66	6	81	6	96	6	111	5	126	5	141	4
7	3	22	2	37	4	52	8	67	6	82	5	97	4	112	4	127	10	142	6
8	5	23	6	38	0	53	2	68	7	83	6	98	4	113	4	128	4	143	6
9	5	24	7	39	4	54	4	69	7	84	8	99	2	114	9	129	1	144	4
10	6	25	3	40	3	55	8	70	5	85	4	100	5	115	4	130	4	145	5
11	1	26	6	41	6	56	3	71	10	86	10	101	4	116	3	131	10	146	5
12	10	27	5	42	10	57	5	72	3	87	5	102	7	117	3	132	8		
13	4	28	8	43	3	58	5	73	5.5	88	10	103	6	118	4	133	10		
14	4	29	0	44	6	59	8	74	7	89	8	104	8	119	4	134	6		
15	6	30	8	45	10	60	4	75	10	90	5	105	3	120	8	135	6		

Média da amostra: $(8+6+10+4+5+3+5+6+6+6) / 10 = 5.9$

Distribuição de amostragem

O que você acabou de construir é chamado de *distribuição de amostragem*.

Distribuição de amostragem

O que você acabou de construir é chamado de *distribuição de amostragem*.

Qual é a forma e o centro dessa distribuição? Com base nessa distribuição, qual você acha que é a média real da população?

Distribuição de amostragem

O que você acabou de construir é chamado de *distribuição de amostragem*.

Qual é a forma e o centro dessa distribuição? Com base nessa distribuição, qual você acha que é a média real da população?

Aproximadamente 5,39, a verdadeira média da população.

Teorema do limite central (TLC)

Teorema do limite central

A distribuição da média da amostra é bem aproximada por um modelo normal:

$$\bar{x} \sim N\left(\text{média} = \mu, SE = \frac{\sigma}{\sqrt{n}}\right),$$

onde SE representa *erro padrão*, que é definido como o desvio padrão da distribuição amostral. Se σ é desconhecido, use s .

- Não foi uma coincidência que a distribuição de amostragem que vimos anteriormente fosse simétrica e centralizada na média real da população.

Teorema do limite central (TLC)

- Nós não vamos passar por uma prova detalhada do porquê $SE = \frac{\sigma}{\sqrt{n}}$, mas note que conforme n aumenta, SE diminui.
 - À medida que o tamanho da amostra aumenta, esperaríamos que as amostras produzissem médias de amostras mais consistentes, portanto a variabilidade entre as médias da amostra seria menor.

TLC - condições

Certas condições devem ser atendidas para o TLC valer:

1. *Independência*: As observações da amostra devem ser independentes.

Isso é difícil de verificar, mas é mais provável termos independência

- quando amostragem/atribuição aleatória é realizada, e
- quando é feita amostragem sem reposição, $n < 10\%$ da população.

TLC - condições

Certas condições devem ser atendidas para o TLC valer:

1. *Independência*: As observações da amostra devem ser independentes.
Isso é difícil de verificar, mas é mais provável termos independência
 - quando amostragem/atribuição aleatória é realizada, e
 - quando é feita amostragem sem reposição, $n < 10\%$ da população.
2. *Tamanho da amostra/assimetria*: A distribuição da população é normal ou, se a distribuição da população for assimétrica, o tamanho da amostra precisa ser grande.
 - Quanto mais assimétrica a distribuição da população, maior o tamanho da amostra que precisamos para a aplicar TLC.
 - para distribuições moderadamente assimétrica $n > 30$ é uma regra prática amplamente usada

Isso também é difícil de verificar para a população, mas podemos verificar-la usando os dados da amostra e assumir que a amostra espelha a população.

4.2. Intervalos de confiança

Intervalos de confiança

- Um intervalo de valores plausível para um parâmetro da população é chamado de *intervalo de confiança*.
- Usar apenas uma estatística de amostra para estimar um parâmetro é como pescar em um lago escuro com uma lança, mas usar um intervalo de confiança é como pescar com uma rede.



Podemos jogar uma lança onde vimos um peixe, mas provavelmente o perdemos. Se atirarmos uma rede nessa área, temos uma boa chance de pegar o peixe.



- Se utilizarmos apenas uma estimativa pontual, provavelmente não atingiremos o parâmetro populacional exato. Se utilizarmos um intervalo de valores plausíveis, temos uma boa chance de alcançar o parâmetro.

Número médio de relacionamentos exclusivos

Uma amostra aleatória de 50 estudantes universitários foi questionada sobre quantos relacionamentos exclusivos eles mantinham até então. Esta amostra rendeu uma média de 3,2 e um desvio padrão de 1,74. Estime o verdadeiro número médio de relacionamentos exclusivos usando essa amostra.

Número médio de relacionamentos exclusivos

Uma amostra aleatória de 50 estudantes universitários foi questionada sobre quantos relacionamentos exclusivos eles mantinham até então. Esta amostra rendeu uma média de 3,2 e um desvio padrão de 1,74. Estime o verdadeiro número médio de relacionamentos exclusivos usando essa amostra.

$$\bar{x} = 3.2 \quad s = 1.74$$

Número médio de relacionamentos exclusivos

Uma amostra aleatória de 50 estudantes universitários foi questionada sobre quantos relacionamentos exclusivos eles mantinham até então. Esta amostra rendeu uma média de 3,2 e um desvio padrão de 1,74. Estime o verdadeiro número médio de relacionamentos exclusivos usando essa amostra.

$$\bar{x} = 3.2 \quad s = 1.74$$

O intervalo de confiança aproximado de 95% é definido como

$$\text{estimativa pontual} \pm 2 \times SE$$

Número médio de relacionamentos exclusivos

Uma amostra aleatória de 50 estudantes universitários foi questionada sobre quantos relacionamentos exclusivos eles mantinham até então. Esta amostra rendeu uma média de 3,2 e um desvio padrão de 1,74. Estime o verdadeiro número médio de relacionamentos exclusivos usando essa amostra.

$$\bar{x} = 3.2 \quad s = 1.74$$

O intervalo de confiança aproximado de 95% é definido como

$$\text{estimativa pontual} \pm 2 \times SE$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

Número médio de relacionamentos exclusivos

$$\bar{x} \pm 2 \times SE = 3.2 \pm 2 \times 0.25$$

Número médio de relacionamentos exclusivos

$$\begin{aligned}\bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5)\end{aligned}$$

Número médio de relacionamentos exclusivos

$$\begin{aligned}\bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5) \\ &= (2.7, 3.7)\end{aligned}$$

Qual das seguintes interpretações deste intervalo de confiança é a correta?

Estamos 95% confiantes de que

- (a) o número médio de relacionamentos exclusivos que os estudantes universitários da amostra estiveram é entre 2,7 e 3,7.
- (b) Em média, os estudantes universitários têm entre 2,7 e 3,7 relacionamentos exclusivos.
- (c) um estudante universitário escolhido aleatoriamente tem de 2,7 a 3,7 relacionamentos exclusivos.
- (d) 95% dos estudantes universitários têm de 2,7 a 3,7 relacionamentos exclusivos.

Prática

Qual das seguintes interpretações deste intervalo de confiança é a correta?

Estamos 95% confiantes de que

- (a) o número médio de relacionamentos exclusivos que os estudantes universitários da amostra estiveram é entre 2,7 e 3,7.
- (b) *Em média, os estudantes universitários têm entre 2,7 e 3,7 relacionamentos exclusivos.*
- (c) um estudante universitário escolhido aleatoriamente tem de 2,7 a 3,7 relacionamentos exclusivos.
- (d) 95% dos estudantes universitários têm de 2,7 a 3,7 relacionamentos exclusivos.

Um intervalo mais preciso

Intervalo de confiança, uma fórmula geral

$$\text{estimativa pontual} \pm z^* \times SE$$

Um intervalo mais preciso

Intervalo de confiança, uma fórmula geral

$$\text{estimativa pontual} \pm z^* \times SE$$

Condições para ter uma estimativa pontual = \bar{x} :

1. *Independência:* As observações na amostra devem ser independentes
 - amostragem/atribuição aleatória
 - se amostragem sem reposição, $n < 10\%$ da população
2. *Tamanho da amostra/assimetria:* $n \geq 30$ e a distribuição da população não deve ser extremamente assimétrica.

Um intervalo mais preciso

Intervalo de confiança, uma fórmula geral

$$\text{estimativa pontual} \pm z^* \times SE$$

Condições para ter uma estimativa pontual = \bar{x} :

1. *Independência:* As observações na amostra devem ser independentes
 - amostragem/atribuição aleatória
 - se amostragem sem reposição, $n < 10\%$ da população
2. *Tamanho da amostra/assimetria:* $n \geq 30$ e a distribuição da população não deve ser extremamente assimétrica.

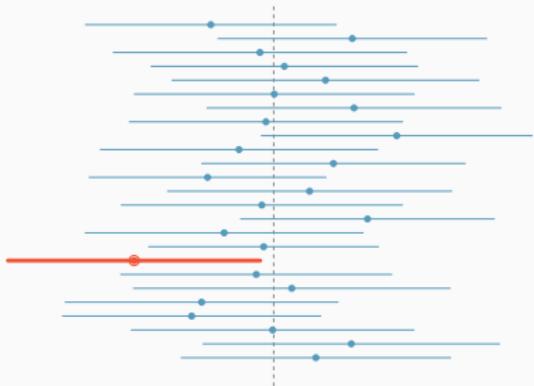
Nota: Vamos discutir sobre amostras onde $n < 30$ no próximo capítulo.

O que significa 95% de confiança?

- Suponha que tiramos muitas amostras e construímos um intervalo de confiança de cada amostra usando a equação:

$$\text{estimativa pontual} \pm z^* \times SE$$

- Então, cerca de 95% desses intervalos conteriam a média real da população (μ).
- A figura mostra esse procedimento com 25 amostras, em que 24 dos intervalos de confiança resultantes contém o número médio real de relacionamentos exclusivos e um intervalo não contém o número médio real.



Largura de um intervalo

Se queremos ter mais certeza de que capturamos o parâmetro população, ou seja, se queremos aumentar o nosso nível de confiança, devemos usar um intervalo maior ou um intervalo menor?

Largura de um intervalo

Se queremos ter mais certeza de que capturamos o parâmetro população, ou seja, se queremos aumentar o nosso nível de confiança, devemos usar um intervalo maior ou um intervalo menor?

Um intervalo maior.

Largura de um intervalo

Se queremos ter mais certeza de que capturamos o parâmetro população, ou seja, se queremos aumentar o nosso nível de confiança, devemos usar um intervalo maior ou um intervalo menor?

Um intervalo maior.

Você consegue ver alguma desvantagem em usar um intervalo maior?



Largura de um intervalo

Se queremos ter mais certeza de que capturamos o parâmetro população, ou seja, se queremos aumentar o nosso nível de confiança, devemos usar um intervalo maior ou um intervalo menor?

Um intervalo maior.

Você consegue ver alguma desvantagem em usar um intervalo maior?



Se o intervalo for muito largo, pode não ser muito informativo.

Image source: http://web.as.uky.edu/statistics/users/earo227/misc/garfield_weather.gif

Alterando o nível de confiança

$$\text{estimativa pontual} \pm z^* \times SE$$

- Em um intervalo de confiança, $z^* \times SE$ é chamado de *margem de erro* e, para uma determinada amostra, a margem de erro é alterada conforme o nível de confiança é alterado.
- Para alterar o nível de confiança, precisamos ajustar z^* na fórmula acima.
- Os níveis de confiança comumente usados são 90%, 95%, 98% e 99%.
- Para um intervalo de confiança de 95%, $z^* = 1.96$.
- No entanto, usando a distribuição normal padrão (z), é possível encontrar o z^* apropriado para qualquer nível de confiança.

Prática

Qual dos valores abaixo de Z é o z^* apropriado ao calcular um intervalo de confiança de 98%?

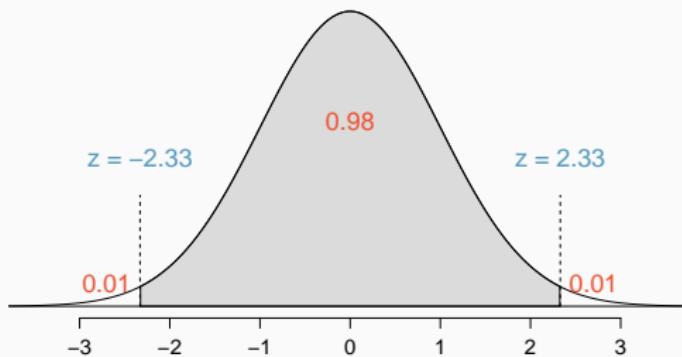
(a) $Z = 2.05$

(d) $Z = -2.33$

(b) $Z = 1.96$

(e) $Z = -1.65$

(c) $Z = 2.33$



Prática

Qual dos valores abaixo de Z é o z^* apropriado ao calcular um intervalo de confiança de 98%?

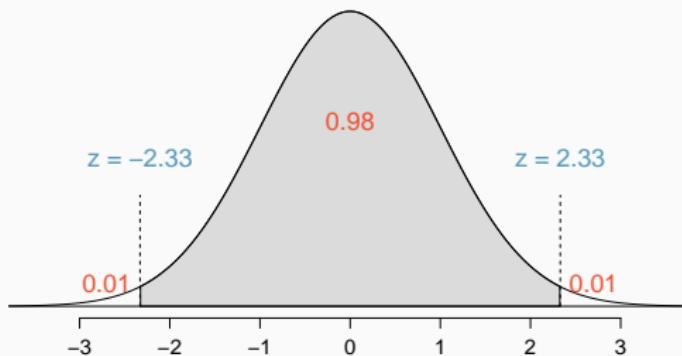
(a) $Z = 2.05$

(d) $Z = -2.33$

(b) $Z = 1.96$

(e) $Z = -1.65$

(c) **$Z = 2.33$**



4.3. Testando hipóteses

Lembre-se de quando...

Experimento sobre discriminação de gênero: Ser promovido no emprego independe do gênero?

Gênero	Promoção		Total
	Promovido	Não Promovido	
Masculino	21	3	24
Feminino	14	10	24
Total	35	13	48

Lembre-se de quando...

Experimento sobre discriminação de gênero: Ser promovido no emprego independe do gênero?

Gênero	Promoção		Total
	Promovido	Não Promovido	
Masculino	21	3	24
Feminino	14	10	24
Total	35	13	48

$$\hat{p}_{homens} = 21/24 \approx 0.88$$

$$\hat{p}_{mulheres} = 14/24 \approx 0.58$$

Lembre-se de quando...

Experimento sobre discriminação de gênero: Ser promovido no emprego independe do gênero?

Gênero	Promoção		Total
	Promovido	Não Promovido	
Masculino	21	3	24
Feminino	14	10	24
Total	35	13	48

$$\hat{p}_{homens} = 21/24 \approx 0.88$$

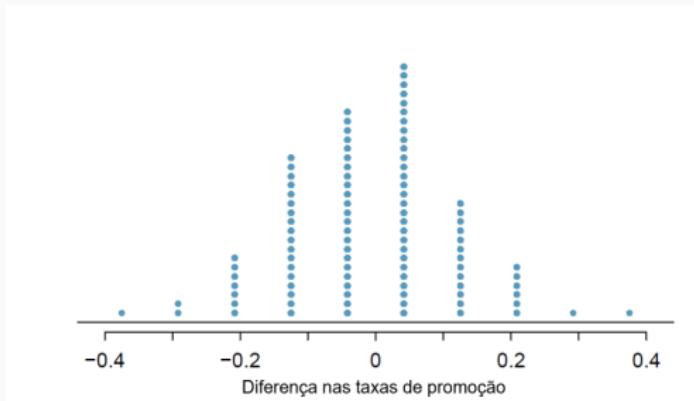
$$\hat{p}_{mulheres} = 14/24 \approx 0.58$$

Lembre-se de quando...

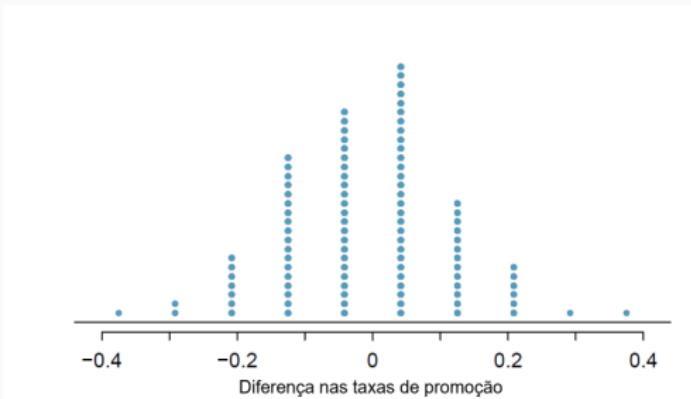
Explicações possíveis:

- Promoção e gênero são *independentes*, não há discriminação de gênero, a diferença observada nas proporções é simplesmente devido ao acaso. → *nulo* - (nada está acontecendo)
- Promoção e gênero são *dependentes*, há discriminação de gênero, a diferença observada nas proporções não se deve ao acaso. → *alternativa* - (algo está acontecendo)

Resultado



Resultado



Como era bastante improvável obter resultados como os dados reais ou algo mais extremo nas simulações (promoções masculinas sendo 30% ou mais frequentes às promoções femininas), decidimos rejeitar a hipótese nula em favor da alternativa.

Recapitulação: estrutura de teste de hipóteses

- Começamos com uma *hipótese nula (H_0)* que representa o status quo.

Recapitulação: estrutura de teste de hipóteses

- Começamos com uma *hipótese nula (H_0)* que representa o status quo.
- Também temos uma hipótese *alternativa (H_A)* que representa nossa questão de pesquisa, ou seja, o que estamos testando.

Recapitulação: estrutura de teste de hipóteses

- Começamos com uma *hipótese nula (H_0)* que representa o status quo.
- Também temos uma hipótese *alternativa (H_A)* que representa nossa questão de pesquisa, ou seja, o que estamos testando.
- Realizamos um teste de hipóteses sob o pressuposto de que a hipótese nula é verdadeira, seja através de simulação ou métodos tradicionais baseados no teorema do limite central (veremos logo mais...).

Recapitulação: estrutura de teste de hipóteses

- Se os resultados do teste sugerirem que os dados não fornecem evidências convincentes para a hipótese alternativa, nós ficaremos com a hipótese nula. Mas se o fizerem, rejeitamos a hipótese nula em favor da alternativa.

Recapitulação: estrutura de teste de hipóteses

- Se os resultados do teste sugerirem que os dados não fornecem evidências convincentes para a hipótese alternativa, nós ficaremos com a hipótese nula. Mas se o fizerem, rejeitamos a hipótese nula em favor da alternativa.

Vamos introduzir formalmente a estrutura de testes de hipóteses usando um exemplo para testar uma afirmação sobre uma média populacional.

Testando hipóteses usando intervalos de confiança

Anteriormente, calculamos um intervalo de confiança de 95% para o número médio de relacionamentos exclusivos em que os estudantes universitários estiveram (2,7, 3,7). Com base nesse intervalo de confiança, esses dados confirmam a hipótese de que os estudantes universitários, em média, já estiveram em mais de três relacionamentos exclusivos.

Testando hipóteses usando intervalos de confiança

Anteriormente, calculamos um intervalo de confiança de 95% para o número médio de relacionamentos exclusivos em que os estudantes universitários estiveram (2,7, 3,7). Com base nesse intervalo de confiança, esses dados confirmam a hipótese de que os estudantes universitários, em média, já estiveram em mais de três relacionamentos exclusivos.

- As hipóteses associadas são:

$H_0: \mu = 3$: Os estudantes universitários já estiveram em três relacionamentos exclusivos, em média.

$H_A: \mu > 3$: Os estudantes universitários já estiveram em mais de 3 relacionamentos exclusivos, em média.

Testando hipóteses usando intervalos de confiança

- Como o valor nulo está dentro do intervalo, não rejeitamos a hipótese nula em favor da alternativa.

Testando hipóteses usando intervalos de confiança

- Como o valor nulo está dentro do intervalo, não rejeitamos a hipótese nula em favor da alternativa.
- Esta é uma abordagem rápida (e um pouco tosca) para fazer testes de hipóteses. No entanto, ela não nos diz a probabilidade de certos resultados sob a hipótese nula, ou seja, o p-valor, com base no qual podemos tomar uma decisão sobre as hipóteses.

Número de pedidos de faculdade

Uma pesquisa semelhante perguntou para 206 alunos da Duke University para quantas faculdades eles aplicaram. Esta amostra rendeu uma média de 9,7 aplicações universitárias com um desvio padrão de 7. O site do College Board afirma que os orientadores recomendam que os alunos se inscrevam em aproximadamente 8 faculdades. Esses dados fornecem evidências convincentes de que o número médio de faculdades aos quais os alunos da Duke se candidatam é superior ao recomendado?

<http://www.collegeboard.com/student/apply/the-application/151680.html>

Definindo as hipóteses

- O *parâmetro de interesse* é o número médio de escolas aplicadas por todos estudantes da Duke.

Definindo as hipóteses

- O *parâmetro de interesse* é o número médio de escolas aplicadas por todos estudantes da Duke.
- Pode haver duas explicações para a nossa média de amostra ser maior do que as 8 faculdades recomendadas.
 - A verdadeira média populacional é diferente.
 - A média real da população é 8, e a diferença entre a média real da população e a média da amostra é simplesmente devido à variabilidade natural da amostragem.

Definindo as hipóteses

- O *parâmetro de interesse* é o número médio de escolas aplicadas por todos estudantes da Duke.
- Pode haver duas explicações para a nossa média de amostra ser maior do que as 8 faculdades recomendadas.
 - A verdadeira média populacional é diferente.
 - A média real da população é 8, e a diferença entre a média real da população e a média da amostra é simplesmente devido à variabilidade natural da amostragem.

Definindo as hipóteses

- Começamos com a suposição de que o número médio de faculdades em que os alunos da Duke se candidatam é 8 (como recomendado)

$$H_0 : \mu = 8$$

Definindo as hipóteses

- Começamos com a suposição de que o número médio de faculdades em que os alunos da Duke se candidatam é 8 (como recomendado)

$$H_0 : \mu = 8$$

- Testamos a alegação de que o número médio de faculdades em que os alunos da Duke se candidatam é maior que 8

$$H_A : \mu > 8$$

Número de candidaturas universitárias - condições

Qual das seguintes opções não é uma condição que precisa ser atendida para aplicar este teste de hipótese?

- (a) Os alunos da amostra devem ser independentes uns dos outros em relação a quantas faculdades se candidataram.
- (b) A amostragem deveria ter sido feita aleatoriamente.
- (c) O tamanho da amostra deve ser menor que 10 % da população de todos os alunos da Duke.
- (d) Deve haver pelo menos 10 sucessos e 10 falhas na amostra.
- (e) A distribuição do número de faculdades em que os alunos aplicaram não deve ser extremamente assimétrica.

Número de candidaturas universitárias - condições

Qual das seguintes opções não é uma condição que precisa ser atendida para aplicar este teste de hipótese?

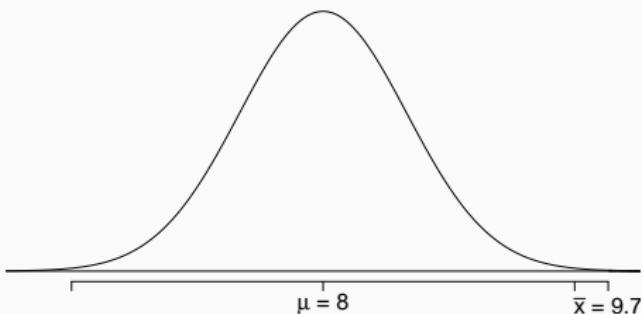
- (a) Os alunos da amostra devem ser independentes uns dos outros em relação a quantas faculdades se candidataram.
- (b) A amostragem deveria ter sido feita aleatoriamente.
- (c) O tamanho da amostra deve ser menor que 10 % da população de todos os alunos da Duke.
- (d) *Deve haver pelo menos 10 sucessos e 10 falhas na amostra.*
- (e) A distribuição do número de faculdades em que os alunos aplicaram não deve ser extremamente assimétrica.

Estatística de teste

Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.

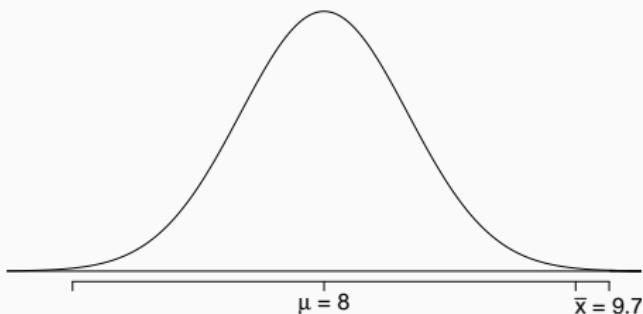
Estatística de teste

Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.



Estatística de teste

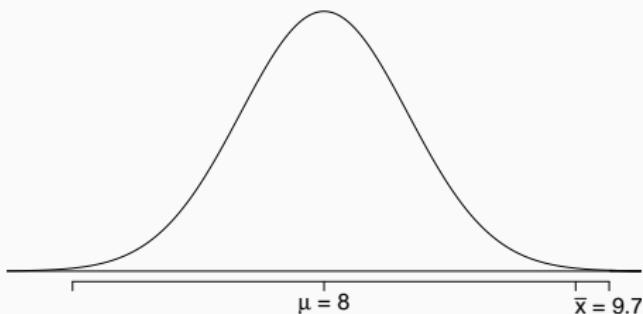
Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

Estatística de teste

Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.

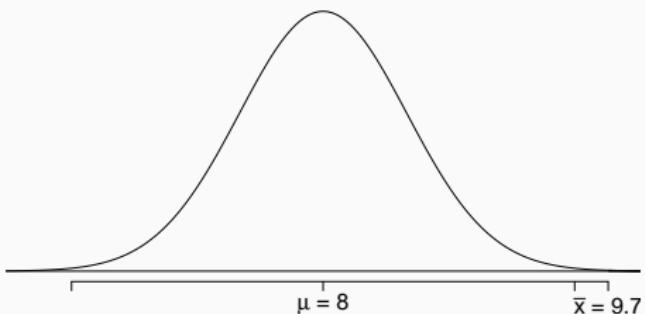


$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

Estatística de teste

Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.



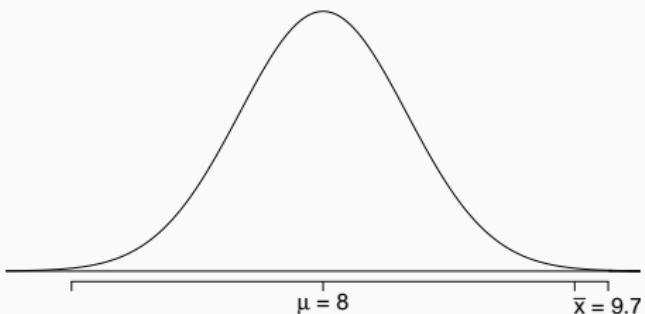
A média da amostra é de 3,4 erros padrão longe do valor hipotético. Isso é considerado excepcionalmente alto? Ou seja, é o resultado *estatisticamente significativo*?

$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

Estatística de teste

Para avaliar se a média da amostra observada é diferente da distribuição de amostragem hipotética, determinamos quantos erros-padrão podem estar fora da hipótese nula, o que também é chamado de *estatística de teste*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

A média da amostra é de 3,4 erros padrão longe do valor hipotético. Isso é considerado excepcionalmente alto? Ou seja, é o resultado *estatisticamente significativo*?

Sim, e podemos quantificar o quanto incomum é essa média usando um *p*-valor.

P-valor

- Em seguida, usamos essa estatística de teste para calcular o *p-valor*, ou seja, a probabilidade de observar dados tão favoráveis à hipótese alternativa quanto nosso conjunto de dados atual, se a hipótese nula fosse verdadeira.

P-valor

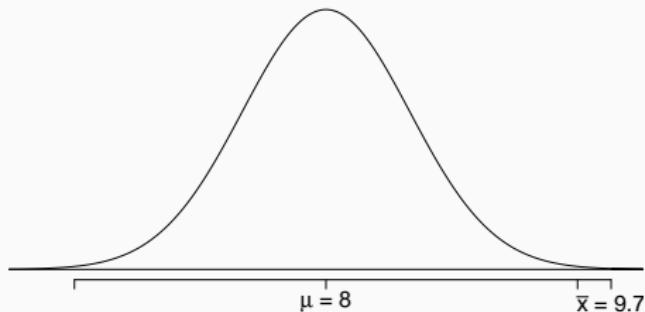
- Em seguida, usamos essa estatística de teste para calcular o *p-valor*, ou seja, a probabilidade de observar dados tão favoráveis à hipótese alternativa quanto nosso conjunto de dados atual, se a hipótese nula fosse verdadeira.
- Se o p-valor for *baixo* (menor que o nível de significância, α , que normalmente é de 5%), dizemos que seria muito improvável observar os dados se a hipótese nula fosse verdadeira, e daí *rejeitamos H_0* .

P-valor

- Em seguida, usamos essa estatística de teste para calcular o *p-valor*, ou seja, a probabilidade de observar dados tão favoráveis à hipótese alternativa quanto nosso conjunto de dados atual, se a hipótese nula fosse verdadeira.
- Se o p-valor for *baixo* (menor que o nível de significância, α , que normalmente é de 5%), dizemos que seria muito improvável observar os dados se a hipótese nula fosse verdadeira, e daí *rejeitamos H_0* .
- Se o p-valor for *alto* (maior que α), dizemos que é provável observar os dados mesmo que a hipótese nula seja verdadeira e, portanto, *não rejeitamos H_0* .

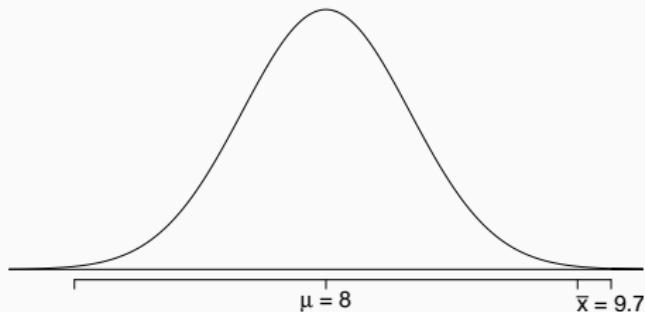
Número de inscrições em faculdades - p-valor

P-valor: probabilidade de observar dados pelo menos tão favoráveis a H_A como nosso conjunto de dados atual (uma média de amostra maior que 9,7), se de fato H_0 fosse verdadeira (a média real da população é 8).



Número de inscrições em faculdades - p-valor

P-valor: probabilidade de observar dados pelo menos tão favoráveis a H_A como nosso conjunto de dados atual (uma média de amostra maior que 9,7), se de fato H_0 fosse verdadeira (a média real da população é 8).



$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4) = 0.0003$$

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003
 - Se a média real do número de faculdades em que os alunos da Duke se candidataram é 8, há apenas 0,03% de chance de observar uma amostra aleatória de 206 alunos da Duke que, em média, aplicaram para 9,7 ou mais faculdades.

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003
 - Se a média real do número de faculdades em que os alunos da Duke se candidataram é 8, há apenas 0,03% de chance de observar uma amostra aleatória de 206 alunos da Duke que, em média, aplicaram para 9,7 ou mais faculdades.
 - Esta é uma probabilidade muito baixa para acreeditarmos que uma média amostral de 9,7 ou mais faculdades acontecerá por acaso.

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003
 - Se a média real do número de faculdades em que os alunos da Duke se candidataram é 8, há apenas 0,03% de chance de observar uma amostra aleatória de 206 alunos da Duke que, em média, aplicaram para 9,7 ou mais faculdades.
 - Esta é uma probabilidade muito baixa para acreeditarmos que uma média amostral de 9,7 ou mais faculdades acontecerá por acaso.
- Como o p-valor é *baixo* (menor que 5%), nós *rejeitamos H_0* .

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003
 - Se a média real do número de faculdades em que os alunos da Duke se candidataram é 8, há apenas 0,03% de chance de observar uma amostra aleatória de 206 alunos da Duke que, em média, aplicaram para 9,7 ou mais faculdades.
 - Esta é uma probabilidade muito baixa para acreeditarmos que uma média amostral de 9,7 ou mais faculdades acontecerá por acaso.
- Como o p-valor é *baixo* (menor que 5%), nós *rejeitamos H_0* .
- Os dados fornecem evidências convincentes de que os alunos da Duke aplicaram para mais de 8 faculdades, em média.

Número de inscrições em faculdades - tomada de decisão

- p-valor = 0.0003
 - Se a média real do número de faculdades em que os alunos da Duke se candidataram é 8, há apenas 0,03% de chance de observar uma amostra aleatória de 206 alunos da Duke que, em média, aplicaram para 9,7 ou mais faculdades.
 - Esta é uma probabilidade muito baixa para acreeditarmos que uma média amostral de 9,7 ou mais faculdades acontecerá por acaso.
- Como o p-valor é *baixo* (menor que 5%), nós *rejeitamos H_0* .
- Os dados fornecem evidências convincentes de que os alunos da Duke aplicaram para mais de 8 faculdades, em média.
- A diferença entre a hipótese nula de 8 faculdades e a média amostral observada de 9,7 faculdades *não é por acaso* ou devido à variabilidade amostral.

Prática

Uma pesquisa da National Sleep Foundation descobriu que os estudantes universitários têm em média 7 horas de sono por noite. Uma amostra de 169 estudantes universitários que fizeram uma aula de estatística introdutória rendeu uma média de 6,88 horas de sono por noite, com um desvio padrão de 0,94 horas. Supondo que esta seja uma amostra aleatória representativa de todos os estudantes universitários (*com um pouco de fé?*), um teste de hipótese foi realizado para avaliar se os estudantes universitários em média dormem menos que 7 horas por noite. O p-valor para este teste de hipótese é 0,0485. Qual das seguintes opções está correta?

Prática

- (a) Não rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários dormem em média menos de 7 horas.
- (b) Rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários dormem menos de 7 horas em média.
- (c) Rejeitamos H_0 , os dados provam que os estudantes universitários dormem mais de 7 horas em média.
- (d) Não rejeitamos H_0 , os dados não fornecem evidências convincentes de que os estudantes universitários dormem menos de 7 horas em média.
- (e) Rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários nesta amostra dormem menos de 7 horas em média.

Prática

- (a) Não rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários dormem em média menos de 7 horas.
- (b) *Rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários dormem menos de 7 horas em média.*
- (c) Rejeitamos H_0 , os dados provam que os estudantes universitários dormem mais de 7 horas em média.
- (d) Não rejeitamos H_0 , os dados não fornecem evidências convincentes de que os estudantes universitários dormem menos de 7 horas em média.
- (e) Rejeitamos H_0 , os dados fornecem evidências convincentes de que os estudantes universitários nesta amostra dormem menos de 7 horas em média.

Teste de hipóteses bilateral e p-valor

- Se a pergunta de pesquisa for "Os dados fornecem evidências convincentes de que a média de sono dos estudantes universitários por noite é *diferente* que a média nacional?", a hipótese alternativa seria diferente.

$$H_0 : \mu = 7$$

$$H_A : \mu \neq 7$$

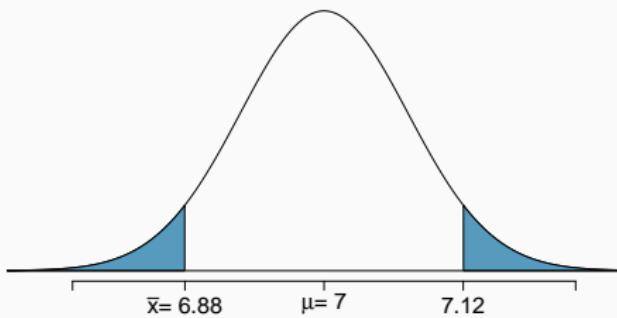
Teste de hipóteses bilateral e p-valor

- Se a pergunta de pesquisa for "Os dados fornecem evidências convincentes de que a média de sono dos estudantes universitários por noite é **diferente** que a média nacional?", a hipótese alternativa seria diferente.

$$H_0 : \mu = 7$$

$$H_A : \mu \neq 7$$

- Portanto, o p-valor também mudaria:



$$\begin{aligned}\text{valor-p} \\ &= 0.0485 \times 2 \\ &= 0.097\end{aligned}$$

Erros de decisão

- Testes de hipóteses não são perfeitos.
- No sistema judiciário, pessoas inocentes são às vezes erroneamente condenadas e os culpados às vezes saem impunes.
- Da mesma forma, podemos tomar uma decisão errada nos testes de hipóteses estatísticas.
- A diferença é que temos as ferramentas necessárias para quantificar a frequência com que cometemos erros nas estatísticas.

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro		
	H_A verdadeiro		

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro	✓	
	H_A verdadeiro		

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro	✓	
	H_A verdadeiro		✓

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro	✓	<i>Erro tipo 1</i>
	H_A verdadeiro		✓

- O *erro tipo 1* é quando rejeitamos a hipótese nula quando, na verdade, H_0 é verdadeiro.

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro	✓	<i>Erro tipo 1</i>
	H_A verdadeiro	<i>Erro tipo 2</i>	✓

- O *erro tipo 1* é quando rejeitamos a hipótese nula quando, na verdade, H_0 é verdadeiro.
- O *erro tipo 2* é quando não conseguimos rejeitar a hipótese nula quando, na verdade, H_A é verdadeiro.

Erros de decisão (cont.)

Existem duas hipóteses concorrentes: a nula e a alternativa. Em um teste de hipótese, tomamos uma decisão sobre o que pode ser verdade, mas nossa escolha pode estar incorreta.

		Decisão	
		não rejeitar H_0	rejeitar H_0
Verdade	H_0 verdadeiro	✓	<i>Erro tipo 1</i>
	H_A verdadeiro	<i>Erro tipo 2</i>	✓

- O *erro tipo 1* é quando rejeitamos a hipótese nula quando, na verdade, H_0 é verdadeiro.
- O *erro tipo 2* é quando não conseguimos rejeitar a hipótese nula quando, na verdade, H_A é verdadeiro.
- Nós (quase) nunca sabemos se H_0 ou H_A é a verdade, mas precisamos considerar todas as possibilidades.

Teste de Hipótese como um julgamento

Se pensarmos novamente em um teste de hipótese como um processo criminal, então faz sentido enquadrar o veredito em termos das hipóteses nula e alternativa:

$$H_0 : \text{Réu é inocente}$$

$$H_A : \text{Réu é culpado}$$

Qual tipo de erro está sendo cometido nas seguintes circunstâncias?

- Declarando o réu inocente quando ele é culpado.
- Declarando o réu culpado quando ele é inocente.

Teste de Hipótese como um julgamento

Se pensarmos novamente em um teste de hipótese como um processo criminal, então faz sentido enquadrar o veredito em termos das hipóteses nula e alternativa:

$$H_0 : \text{Réu é inocente}$$

$$H_A : \text{Réu é culpado}$$

Qual tipo de erro está sendo cometido nas seguintes circunstâncias?

- Declarando o réu inocente quando ele é culpado.

Erro tipo 2

- Declarando o réu culpado quando ele é inocente.

Teste de Hipótese como um julgamento

Se pensarmos novamente em um teste de hipótese como um processo criminal, então faz sentido enquadrar o veredito em termos das hipóteses nula e alternativa:

$$H_0 : \text{Réu é inocente}$$

$$H_A : \text{Réu é culpado}$$

Qual tipo de erro está sendo cometido nas seguintes circunstâncias?

- Declarando o réu inocente quando ele é culpado.

Erro tipo 2

- Declarando o réu culpado quando ele é inocente.

Erro tipo 1

Teste de Hipótese como um julgamento

Se pensarmos novamente em um teste de hipótese como um processo criminal, então faz sentido enquadrar o veredito em termos das hipóteses nula e alternativa:

$$H_0 : \text{Réu é inocente}$$

$$H_A : \text{Réu é culpado}$$

Qual tipo de erro está sendo cometido nas seguintes circunstâncias?

- Declarando o réu inocente quando ele é culpado.

Erro tipo 2

- Declarando o réu culpado quando ele é inocente.

Erro tipo 1

Qual erro você acha que é pior?

Teste de Hipótese como um julgamento

Se pensarmos novamente em um teste de hipótese como um processo criminal, então faz sentido enquadrar o veredito em termos das hipóteses nula e alternativa:

$$H_0 : \text{Réu é inocente}$$

$$H_A : \text{Réu é culpado}$$

Qual tipo de erro está sendo cometido nas seguintes circunstâncias?

- Declarando o réu inocente quando ele é culpado.

Erro tipo 2

- Declarando o réu culpado quando ele é inocente.

Erro tipo 1

Qual erro você acha que é pior?

“melhor dez pessoas culpadas escaparem do que um inocente sofrer”

Taxa de erro de tipo 1

- Como regra geral, rejeitamos H_0 quando o p-valor é menor que 0,05, ou seja, usamos um *nível de significância* de 0,05, $\alpha = 0.05$.

Taxa de erro de tipo 1

- Como regra geral, rejeitamos H_0 quando o p-valor é menor que 0,05, ou seja, usamos um *nível de significância* de 0,05, $\alpha = 0.05$.
- Isso significa que, nos casos em que H_0 é realmente verdadeira, não queremos rejeitá-la incorretamente mais de 5% dessas vezes.

Taxa de erro de tipo 1

- Como regra geral, rejeitamos H_0 quando o p-valor é menor que 0,05, ou seja, usamos um *nível de significância* de 0,05, $\alpha = 0.05$.
- Isso significa que, nos casos em que H_0 é realmente verdadeira, não queremos rejeitá-la incorretamente mais de 5% dessas vezes.
- Em outras palavras, ao usar um nível de significância de 5%, há cerca de 5% de chance de gerar um erro de Tipo 1 se a hipótese nula for verdadeira.

$$P(\text{Erro tipo 1} \mid H_0 \text{ Verdade}) = \alpha$$

Taxa de erro de tipo 1

- Como regra geral, rejeitamos H_0 quando o p-valor é menor que 0,05, ou seja, usamos um *nível de significância* de 0,05, $\alpha = 0.05$.
- Isso significa que, nos casos em que H_0 é realmente verdadeira, não queremos rejeitá-la incorretamente mais de 5% dessas vezes.
- Em outras palavras, ao usar um nível de significância de 5%, há cerca de 5% de chance de gerar um erro de Tipo 1 se a hipótese nula for verdadeira.

$$P(\text{Erro tipo 1} \mid H_0 \text{ Verdade}) = \alpha$$

- É por isso que preferimos valores pequenos de α – aumentando α aumenta a taxa de erro do tipo 1.

Escolhendo um nível de significância

- Escolher um nível de significância para um teste é importante em muitos contextos, e o nível tradicional é de 0,05. No entanto, geralmente é útil ajustar o nível de significância com base no que se está testando.
- Podemos selecionar um nível menor ou maior que 0,05, dependendo das consequências de quaisquer conclusões obtidas no teste.
- Se as consequências de um Erro Tipo 1 forem perigosas ou caras, devemos escolher um nível de significância pequeno (por exemplo, 0,01). Sob este cenário, queremos ser muito cautelosos ao rejeitar a hipótese nula, então exigimos evidências muito fortes favorecendo H_A antes de rejeitarmos H_0 .

Escolhendo um nível de significância

- Se um Erro Tipo 2 for relativamente mais perigoso ou muito mais dispendioso do que um Erro Tipo 1, então devemos escolher um nível de significância mais alto (por exemplo, 0,10). Aqui, queremos ter cautela ao não rejeitar H_0 quando a hipótese nula for realmente falsa.

As próximas páginas serão um breve resumo dos testes de hipóteses..

Recapitulação: estrutura de testes de hipóteses

1. Defina as hipóteses.
2. Verifique as suposições e condições.
3. Calcule uma *estatística de teste* e um p-valor.
4. Tome uma decisão e interprete-a no contexto da questão de pesquisa.

Recapitulação: teste de hipóteses para uma média populacional

1. Defina as hipóteses

- $H_0 : \mu = \text{valor nulo}$
- $H_A : \mu < \text{ou} > \text{ou} \neq \text{valor nulo}$

2. Calcular a estimativa pontual

3. Verifique as suposições e condições

- Independência: amostra/atribuição aleatória, condição de 10% quando amostragem é sem substituição.
- Normalidade: população quase normal ou $n \geq 30$, sem assimetria extrema - ou use a distribuição t.

4. Calcule uma *estatística de teste* e um p-valor (desenhe uma figura!)

$$Z = \frac{\bar{x} - \mu}{SE}, \text{ onde } SE = \frac{s}{\sqrt{n}}$$

5. Tome uma decisão e interprete-a no contexto

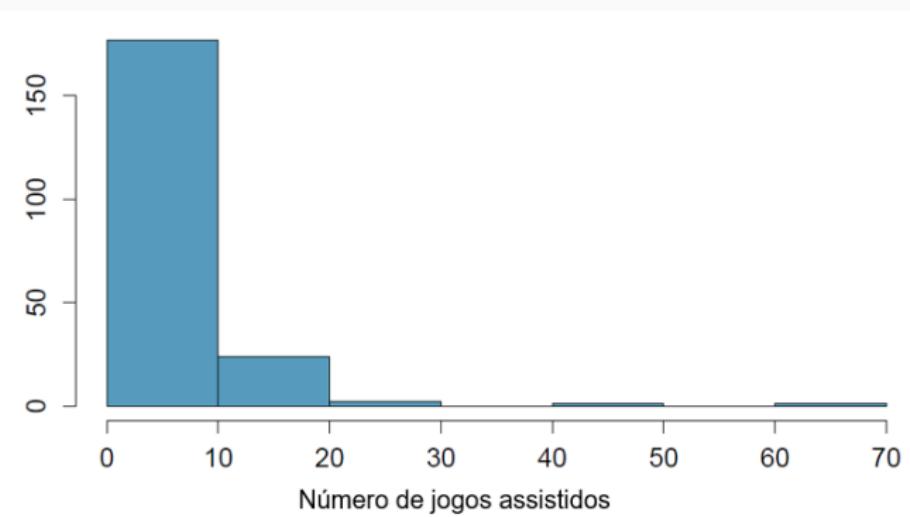
Recapitulação: teste de hipóteses para uma média populacional

- Se o p-valor for $p < \alpha$, rejeitamos H_0 , os dados fornecem evidências à favor de H_A .
- Se o p-valor for $p > \alpha$, não rejeitamos H_0 , os dados não fornecem evidências à favor de H_A .

4.4. Teorema do Limite Central

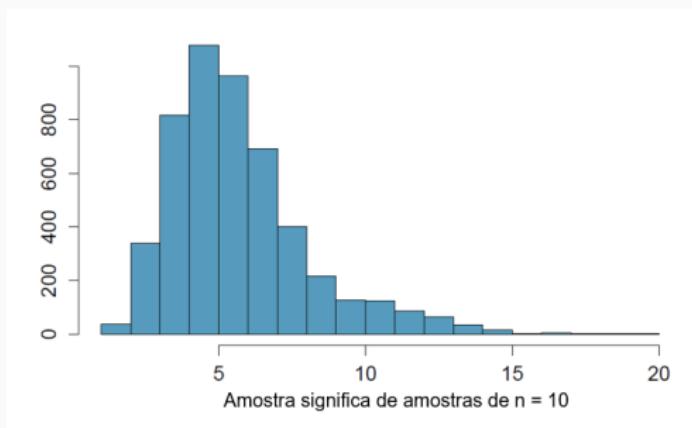
Número médio de jogos de basquete

Em seguida, vamos ver os dados da população para o número de jogos de basquete que assistiram:



Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 10$:

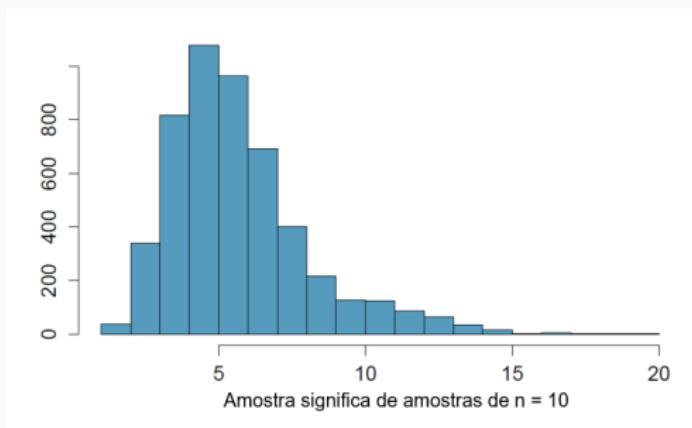


O que cada observação
nesta distribuição repre-
senta??

Média da amostra (\bar{x}) de
amostras de tamanho $n =$
 $|0|$.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 10$:

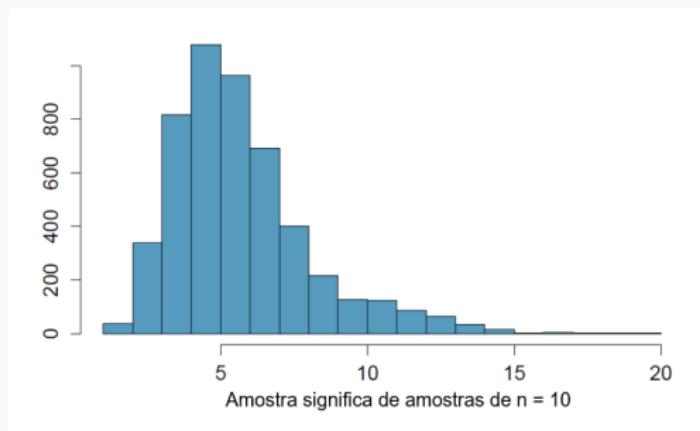


O que cada observação
nesta distribuição repre-
senta??

Média da amostra (\bar{x}) de
amostras de tamanho $n =$
10.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 10$:

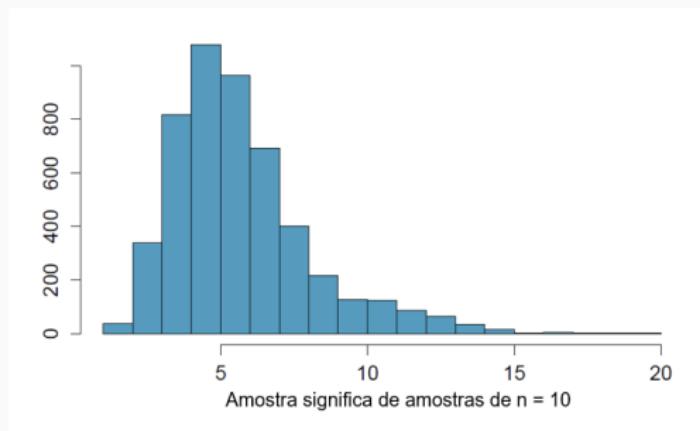


A variabilidade da distribuição amostral é menor ou maior que a variabilidade da distribuição populacional? Por quê?

Menor a média das amostras irá variar menos do que as observações individuais.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 10$:

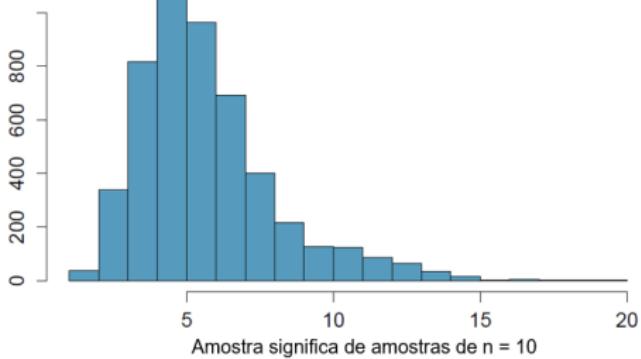


A variabilidade da distribuição amostral é menor ou maior que a variabilidade da distribuição populacional? Por quê?

Menor a média das amostras irá variar menos do que as observações individuais.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 10$:

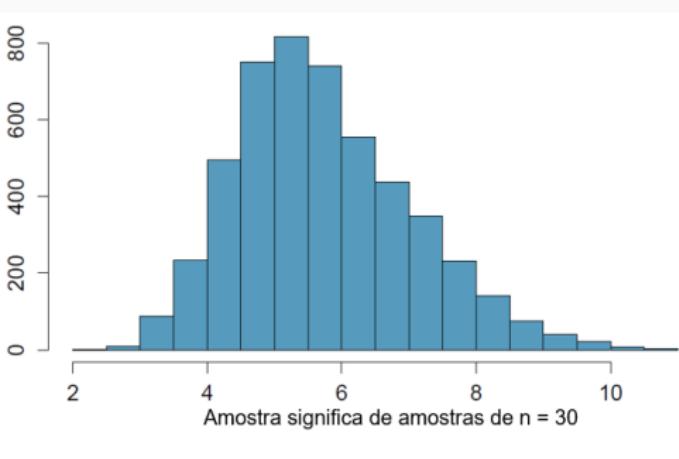


A variabilidade da distribuição amostral é menor ou maior que a variabilidade da distribuição populacional? Por quê?

Menor, a média das amostras irá variar menos do que as observações individuais.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 30$:

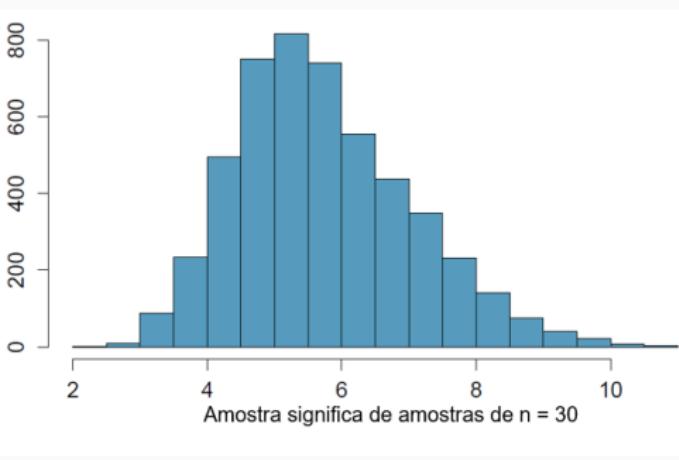


Como o formato, o centro e a disposição da distribuição amostral mudaram de $n = 10$ para $n = 30$?

O formato é mais simétrico, o centro é aproximadamente o mesmo, a disposição é menor.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 30$:

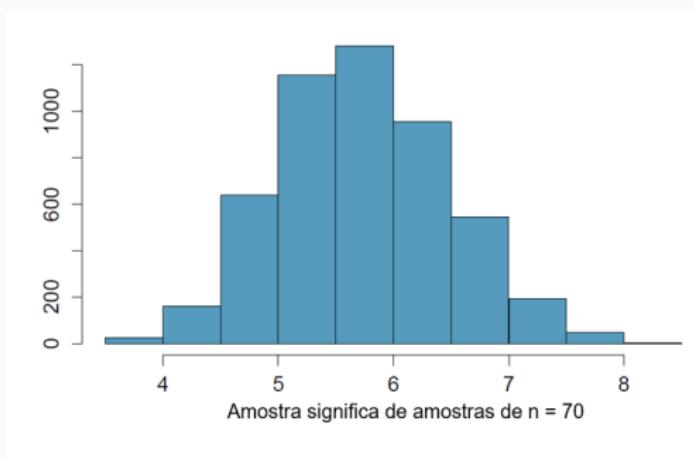


Como o formato, o centro e a disposição da distribuição amostral mudaram de $n = 10$ para $n = 30$?

O formato é mais simétrico, o centro é aproximadamente o mesmo, a disposição é menor.

Número médio de jogos de basquete (cont.)

Distribuição de amostras, $n = 70$:



Número médio de jogos de basquete (cont.)

A média da distribuição amostral é de 5,75 e o desvio padrão da distribuição amostral (também chamado de *erro padrão*) é 0,75. Qual dos seguintes opções é o palpite mais razoável para o intervalo de confiança de 95% para o número médio real de jogos de basquete assistidos?

- (a) 5.75 ± 0.75
- (b) $5.75 \pm 2 \times 0.75$
- (c) $5.75 \pm 3 \times 0.75$
- (d) não podemos afirmar nada com as informações dadas.

Número médio de jogos de basquete (cont.)

A média da distribuição amostral é de 5,75 e o desvio padrão da distribuição amostral (também chamado de *erro padrão*) é 0,75. Qual dos seguintes opções é o palpite mais razoável para o intervalo de confiança de 95% para o número médio real de jogos de basquete assistidos?

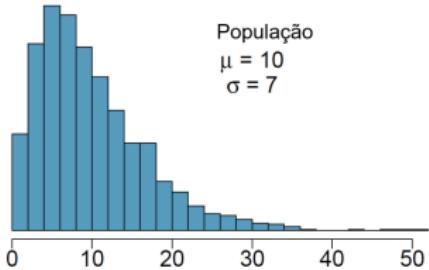
- (a) 5.75 ± 0.75
- (b) $5.75 \pm 2 \times 0.75 \rightarrow (4.25, 7.25)$
- (c) $5.75 \pm 3 \times 0.75$
- (d) não podemos afirmar nada com as informações dadas.

Prática

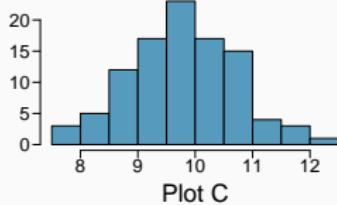
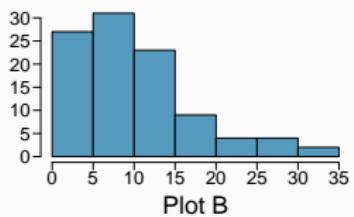
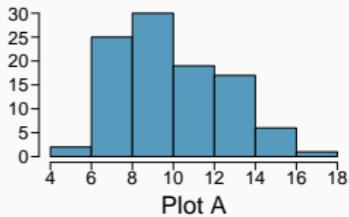
Quatro gráficos: determine qual gráficos (A, B ou C) é qual.

- (1) No topo: distribuição para uma população ($\mu = 10, \sigma = 7$),
- (2) uma única amostra aleatória de 100 observações dessa população,
- (3) uma distribuição de 100 amostras de amostras aleatórias com tamanho 7, e
- (4) uma distribuição de 100 amostras de amostras aleatórias com tamanho 49.

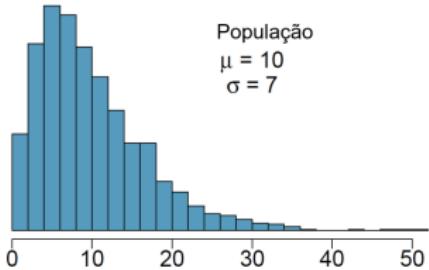
Prática



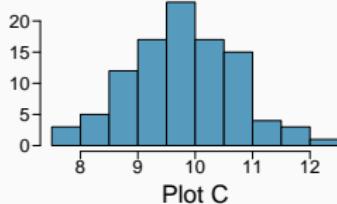
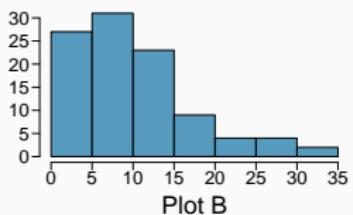
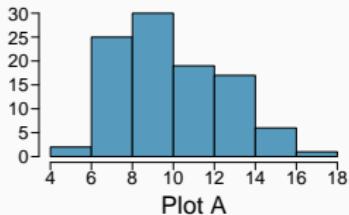
- População
 $\mu = 10$
 $\sigma = 7$
- (a) A - (3); B - (2); C - (4)
 - (b) A - (2); B - (3); C - (4)
 - (c) A - (3); B - (4); C - (2)
 - (d) A - (4); B - (2); C - (3)



Prática



- (a) A - (3); B - (2); C - (4)
- (b) A - (2); B - (3); C - (4)
- (c) A - (3); B - (4); C - (2)
- (d) A - (4); B - (2); C - (3)



4.5. Inferência para outras estimativas

Inferência para outras estimativas

- A média amostral não é a única estimativa pontual para a qual a distribuição amostral é quase normal. Por exemplo, a distribuição amostral das proporções da amostra também é quase normal quando n é suficientemente grande.
- Um pressuposto importante sobre as estimativas pontuais é que elas são *imparciais*, ou seja, a distribuição amostral da estimativa é centrada no parâmetro populacional verdadeiro que se está estimando.
 - Ou seja, uma estimativa imparcial não sobrestima ou subestima o parâmetro. Em vez disso, ele tende a fornecer uma estimativa "boa".
 - A média da amostra é um exemplo de uma estimativa pontual imparcial, assim como cada um dos exemplos que introduzimos nesta seção.

Inferência para outras estimativas

- Algumas estimativas pontuais seguem distribuições diferentes da distribuição normal, e alguns cenários requerem técnicas estatísticas que ainda não foram abordadas - discutiremos estas no final desta seção.

Intervalos de confiança para estimativas pontuais quase normais

Um intervalo de confiança baseado em uma estimativa pontual imparcial e quase normal é

$$\text{estimativa pontual } z^* \text{ SE}$$

onde z^* é selecionado para corresponder ao nível de confiança e SE representa o erro padrão.

Lembre-se que o valor $z^* \text{ SE}$ é chamado de *margem de erro*.

Prática

Um dos primeiros exemplos de assimetria comportamental é uma preferência em humanos por virar a cabeça para a direita, e não para a esquerda, durante as últimas semanas de gestação e nos primeiros 6 meses após o nascimento. Acredita-se que isso influencie o desenvolvimento subsequente das preferências perceptivas e motoras. Um estudo de 124 casais descobriu que 64,5% viravam a cabeça para a direita quando se beijavam. O erro padrão associado a essa estimativa é de aproximadamente 4%. Qual das opções abaixo é falsa?

- (a) O intervalo de confiança de 95% para a porcentagem de casais que durante o beijo viram suas cabeças para a direita é aproximadamente $64.5\% \pm 4\%$.
- (b) Um tamanho de amostra maior produziria um erro padrão menor.
- (c) A margem de erro para um intervalo de confiança de 95% para a porcentagem de casais que durante o beijo viram a cabeça para a direita é de aproximadamente 8%.
- (d) O intervalo de confiança de 99,7% para a porcentagem de casais que durante o beijo viram a cabeça para a direita é aproximadamente $64.5\% \pm 12\%$.

Güntürkün, O. (2003) Adult persistence of head-turning asymmetry. *Nature*. Vol 421.

- (a) *O intervalo de confiança de 95% para a porcentagem de casais que durante o beijo viram suas cabeças para a direita é aproximadamente $64.5\% \pm 4\%$.*
- (b) Um tamanho de amostra maior produziria um erro padrão menor.
- (c) A margem de erro para um intervalo de confiança de 95% para a porcentagem de casais que durante o beijo viram a cabeça para a direita é de aproximadamente 8%.
- (d) O intervalo de confiança de 99,7% para a porcentagem de casais que durante o beijo viram a cabeça para a direita é aproximadamente $64.5\% \pm 12\%$.

Güntürkün, O. (2003) Adult persistence of head-turning asymmetry. *Nature*. Vol 421.

Teste de hipóteses para estimativas pontuais quase normais

A terceira Pesquisa Nacional de Exame de Saúde e Nutrição coletou dados do percentual de gordura corporal (BF %) e gênero de 13.601 indivíduos com idades entre 20 e 80 anos. A média de BF% para os 6.580 homens da amostra foi de 23,9 e esse valor foi de 35,0 para as 7.021 mulheres. O erro padrão para a diferença entre a média do BF% de homens e mulheres foi de 0,114. Esses dados fornecem evidências convincentes de que homens e mulheres têm médias diferentes de BF%. Você pode assumir que a distribuição da estimativa pontual é quase normal.

Teste de hipóteses para estimativas pontuais quase normais

1. Definir hipóteses
2. Calcular estimativa pontual
3. Verificar condições
4. Desenhar distribuição amostral, sombrear o p-valor
5. Calcular estatísticas de teste e o p-valor, tomar uma decisão

Teste de hipóteses para estimativas pontuais quase normais

1. A hipótese nula é que homens e mulheres têm a mesma média de BF% e a alternativa é que esses valores são diferentes.

$$H_0 : \mu_{homem} = \mu_{mulher} \quad H_A : \mu_{homem} \neq \mu_{mulher}$$

Teste de hipóteses para estimativas pontuais quase normais

1. A hipótese nula é que homens e mulheres têm a mesma média de BF% e a alternativa é que esses valores são diferentes.

$$H_0 : \mu_{homem} = \mu_{mulher} \quad H_A : \mu_{homem} \neq \mu_{mulher}$$

2. O parâmetro de interesse é a diferença média nas médias populacionais de BF% para homens e mulheres, e a estimativa pontual para esse parâmetro é a diferença entre as duas médias amostrais:

$$\bar{x}_{homem} - \bar{x}_{mulher} = 23.9 - 35.0 = -11.1$$

Teste de hipóteses para estimativas pontuais quase normais

1. A hipótese nula é que homens e mulheres têm a mesma média de BF% e a alternativa é que esses valores são diferentes.

$$H_0 : \mu_{homem} = \mu_{mulher} \quad H_A : \mu_{homem} \neq \mu_{mulher}$$

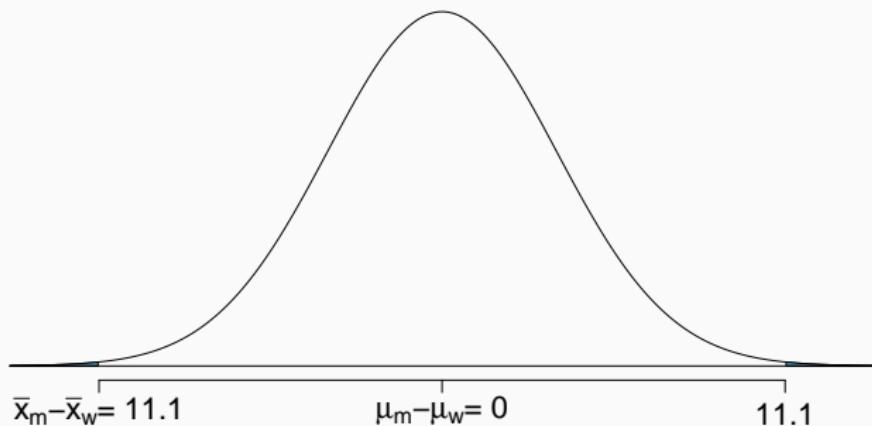
2. O parâmetro de interesse é a diferença média nas médias populacionais de BF% para homens e mulheres, e a estimativa pontual para esse parâmetro é a diferença entre as duas médias amostrais:

$$\bar{x}_{homem} - \bar{x}_{mulher} = 23.9 - 35.0 = -11.1$$

3. Estamos assumindo que a distribuição da estimativa pontual é quase normal (discutiremos detalhes para verificar essa condição no próximo capítulo, no entanto, dado o grande tamanho de amostra, a suposição de normalidade aparece adequada).

Teste de hipóteses para estimativas pontuais quase normais

4. A distribuição amostral será centralizada na hipótese nula ($\mu_{homem} - \mu_{mulher} = 0$), e o p-valor é a área além da diferença observada nas médias amostrais em ambas as caudas (menor que -11,1 e maior que 11,1).



5. A estatística de teste é calculada como a diferença entre a estimativa pontual e o valor nulo ($-11.1 - 0 = -11.1$), escalada pelo erro padrão.

$$Z = \frac{11.1 - 0}{0.114} = 97.36$$

O valor de Z é enorme! E, portanto, o p-valor será pequeno, permitindo-nos rejeitar H_0 em favor de H_A .

5. A estatística de teste é calculada como a diferença entre a estimativa pontual e o valor nulo ($-11.1 - 0 = -11.1$), escalada pelo erro padrão.

$$Z = \frac{11.1 - 0}{0.114} = 97.36$$

O valor de Z é enorme! E, portanto, o p-valor será pequeno, permitindo-nos rejeitar H_0 em favor de H_A .

Esses dados fornecem evidências convincentes de que a média de BF% de homens e mulheres é diferente.

Estimativas pontuais não normais

- Podemos aplicar as ideias de intervalos de confiança e testes de hipóteses aos casos em que a estimativa pontual ou a estatística de teste não é necessariamente normal. Existem muitas razões pelas quais tal situação pode surgir:
 - o tamanho da amostra é muito pequeno para que a aproximação normal seja válida;
 - a estimativa de erro padrão pode ser ruim; ou
 - a estimativa pontual tende para alguma distribuição que não é a distribuição normal.
- Para cada caso em que a aproximação normal não é válida, nossa primeira tarefa é sempre entender e caracterizar a distribuição amostral da estimativa pontual ou estatística de teste. Em seguida, podemos aplicar as estruturas gerais para intervalos de confiança e testes de hipóteses a essas distribuições alternativas.

Quando recuar

- As ferramentas estatísticas dependem das duas condições principais a seguir:
 - *Independência* Uma amostra aleatória de menos de 10% da população assegura a independência das observações. Em experimentos, isso é garantido por amostragem aleatória. Se a independência falhar, técnicas avançadas devem ser usadas e, em alguns casos, a inferência pode não ser possível.
 - *Tamanho e viés da amostra* Por exemplo, se o tamanho da amostra for muito pequeno, o viés muito forte ou valores extremos estiverem presentes, o modelo normal da média da amostra falhará.

Quando recuar

- Sempre que as condições não forem satisfeitas para uma técnica estatística:
 1. Aprenda novos métodos apropriados para os dados.
 2. Consulte um estatístico.
 3. ~~Ignorar o fracasso das condições.~~ Esta última opção efetivamente invalida qualquer análise e pode desacreditar descobertas novas e interessantes.

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}}$$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad p\text{-valor} = 0.0062$$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad p\text{-valor} = 0.0062$$

$$Z_{n=10000} = \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}}$$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad p\text{-valor} = 0.0062$$

$$Z_{n=10000} = \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}} = \frac{50 - 49.5}{\frac{2}{100}} = \frac{0.5}{0.02} = 25, \quad p\text{-valor} \approx 0$$

Prática

Todo o resto é igual, o p-valor será menor se $n = 100$ ou $n = 10,000$?

- (a) $n = 100$
- (b) $n = 10,000$

Suponha $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, e $H_A : \mu > 49.5$.

$$\begin{aligned} Z_{n=100} &= \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad p\text{-valor} = 0.0062 \\ Z_{n=10000} &= \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}} = \frac{50 - 49.5}{\frac{2}{100}} = \frac{0.5}{0.02} = 25, \quad p\text{-valor} \approx 0 \end{aligned}$$

Como n aumenta - $SE \downarrow$, $Z \uparrow$, $p\text{-valor} \downarrow$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$\beta - \text{Pr}(Z \geq 1.97) = 0.15$	$\beta - \text{Pr}(Z \geq 1.99) = 0.09$	$\beta - \text{Pr}(Z \geq 2.29) = 0.29$
$n = 5000$	$\beta - \text{Pr}(Z \geq 1.99) = 0.0012$	$\beta - \text{Pr}(Z \geq 2.01) = 0.0012$	$\beta - \text{Pr}(Z \geq 2.29) = 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p\text{-valor} = 0.45$	$p\text{-valor} = 0.59$	$p\text{-valor} = 0.29$
$n = 5000$	$p\text{-valor} = 0.99$	$p\text{-valor} = 0.9999$	$p\text{-valor} = 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.55$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0012$	$p - valor = 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.39$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0012$	$p - valor = 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.39$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0002$	$p - valor = 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.39$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0002$	$p - valor < 10^{-10}$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.39$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0002$	$p - valor \approx 0$

Prática

Teste a hipótese $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ para as 8 amostras seguintes. Assuma $\sigma = 2$.

\bar{x}	10.05	10.1	10.2
$n = 30$	$p - valor = 0.45$	$p - valor = 0.39$	$p - valor = 0.29$
$n = 5000$	$p - valor = 0.04$	$p - valor = 0.0002$	$p - valor \approx 0$

Quando n é grande, mesmo pequenos desvios da hipótese nula (tamanhos de efeito pequenos), que podem ser considerados praticamente insignificantes, podem produzir resultados estatisticamente significativos.

Significado estatístico vs. prático

- Diferenças reais entre a estimativa pontual e a hipótese nula são mais fáceis de detectar com amostras maiores.
- No entanto, amostras muito grandes resultarão em significância estatística mesmo para pequenas diferenças entre a média da amostra e a hipótese nula (*efeito de tamanho*), mesmo quando a diferença não for praticamente significativa.
- Isso é especialmente importante para a pesquisa: se conduzirmos um estudo, queremos nos concentrar em encontrar resultados significativos (queremos que as diferenças observadas sejam reais, mas também grandes o bastante para serem importantes).
- O papel de um estatístico não é apenas na análise de dados, mas também no planejamento e projeto de um estudo.

Significado estatístico vs. prático

“Chamar o estatístico após o término do experimento pode não ser mais do que pedir que ele faça um exame post-mortem: ele pode ser capaz de lhe dizer porque o experimento morreu.” – R.A. Fisher