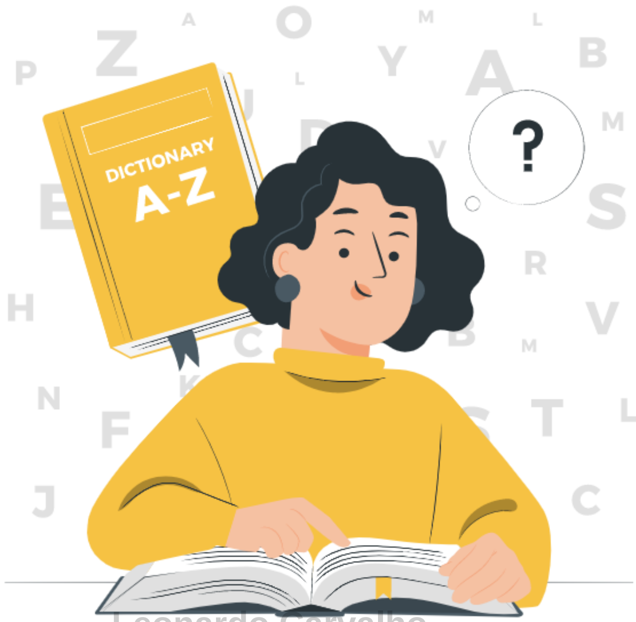
E.B.A - ESTATÍSTICA DO BÁSICO AO AVANÇADO

COM RENATA BIAGGI

TESTE DE HIPÓTESE E INTERVALO DE CONFIANÇA DE MÉDIAS E PROPORÇÕES

GLOSSÁRIO DE SÍMBOLOS



ITEM DE UMA SEQUÊNCIA

Se temos um conjunto de dados, cada item pode ser representado como sendo x1, x2, x3, etc. Por exemplo, temos o conjunto 10, 21, 32, 43, 58, então:

x1 = 10, x2 = 21, x3 = 32, x4= 43, x5= 58

SOMATÓRIO - ∑

Em matemática, somatório ou somatória é a adição de uma sequência de quaisquer tipos de números. O resultado é sua soma ou total.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

Renata Biaggi

O índice i significa a partir de qual elemento devemos começar a soma. O símbolo n indica que, nesse caso, vamos somar a sequencia toda - que tem um total de n elementos.

Vamos a alguns exemplos. Vamos supor que queremos somar os números 1, 2 e 3, ou seja, temos um total de 3 elementos para somar:

$$\sum_{i=1}^{3} xi = x1 + x2 + x3 = 1 + 2 + 3 = 6$$

Muitas vezes vemos o somatório em expressões como:

$$\sum_{n=1}^{3} 2n - 1$$

Esse somatório indica que devemos somar a expressão 2*n - 1, substituindo o n por cada um dos valores da série. Logo, para a série 1, 2 e 3 temos

$$\sum_{n=1}^{3} 2n - 1$$

$$= \underbrace{[2(1) - 1]}_{n=1} + \underbrace{[2(2) - 1]}_{n=2} + \underbrace{[2(3) - 1]}_{n=3}$$

$$= 1 + 3 + 5$$

$$= 9$$

PRODUTÓRIO - Π

O produtório é a multiplicação de uma sequência de objetos matemáticos (números, funções, vetores, matrizes, etc.), chamados fatores, que tem como resultado o produto. É uma operação análoga ao somatório, embora seja menos utilizada quanto esse último. É representado pela letra grega pi maiúscula (Π).

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times \ldots \times x_n$$

OUTROS SÍMBOLOS IMPORTANTES

Símbolo	**Parâmetro**
μ	Média populacional
σ	Desvio-padrão populacional
σ^2	Variância populacional
p	Proporção populacional
$\overline{x}$	Média amostral
s	Desvio-padrão amostral
s^2	Variância amostral
$\hat{p}$	Proporção amostral
α	Nível de significância

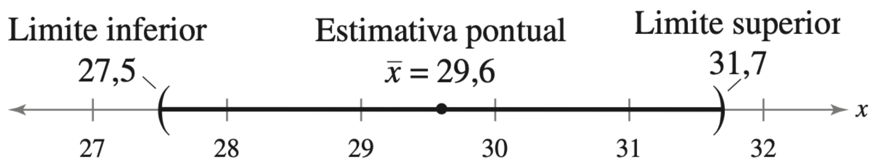
10. INTERVALO DE CONFIANÇA PARA MÉDIAS

Como dissemos anteriormente, o intervalo de confiança sempre vai concordar com um teste de hipótese, então nada mais justo do que começar com ele.

Suponha que por exemplo você queira estimar a média da altura de mulheres brasileiras. Nesse caso, você coleta uma amostra sem vieses sistemáticos e encontra uma altura média de 1,63 m. Essa estimativa é chamada de **estimativa pontual** (ou por ponto). Obviamente, quanto maior o tamanho da sua amostra, mais próximo você de fato se aproxima da média real da altura de sua população. Entretanto, a verdade é que tentar representar toda uma população por apenas 1 número é bastante complicado e carrega muitos erros de estimação (afinal, qual é a chance de selecionar uma mulher aleatória do Brasil e ela ter exatamente 1,63m?).

Por conta disso, surge a ideia de construirmos um intervalo em torno da estimativa por ponto, de modo que a esse intervalo tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro. Essa é a ideia da

estimativa intervalar. Para formar uma estimativa intervalar, usamos a **estimativa pontual** como o **centro** do intervalo e depois adicionamos e subtraímos uma margem de erro.



Chamamos de **intervalo de confiança** o intervalo que, com probabilidade conhecida, deverá conter o valor real do parâmetro. No caso da figura acima, a o intervalo é [27,5, 31,7].

A **margem de erro** é o valor que somamos e subtraímos da estimativa pontual para obter o intervalo - no caso da figura acima, a margem de erro é 2,1 (somando e subtraindo 2,1 de 29,6 obtemos 31,7 e 27,5 respectivamente).

Antes de continuarmos falando como medir o intervalo de confiança, é importante frisar aqui que os intervalos que vamos calcular são **simétricos** em relação ao seu parâmetro e que, portanto, pressupõem uma **distribuição normal**. Por exemplo, voltando novamente ao exemplo acima, somando 2,1 a 29,6 obtemos 31,7 e subtraindo 2,1 de 29,6 e obtermos 27,5. A margem de erro é a mesma "para mais" e "para menos", tornando o intervalo simétrico em relação à estimativa pontual.

Diferentes amostras aleatórias retiradas da mesma população podem produzir intervalos ligeiramente diferentes. Se você extrair muitas amostras aleatórias e calcular um intervalo de confiança para cada amostra, uma proporção específica dos intervalos conterá o parâmetro populacional. Essa porcentagem é o **nível de confiança (c)**.

Se considerarmos c = 100%, podemos ter certeza que o verdadeiro valor do parâmetro estará nesse intervalo. Por exemplo, teríamos certeza que, se escolhessemos aleatoriamente uma mulher brasileira, com certeza a altura dela estaria nesse intervalo com c = 100%. O problema é que esse intervalo seria infinito! É quase impossível garantirmos que temos em um intervalo uma população inteirinha, a menos que esse intervalo seja gigante - ou seja, que cubra todos os números de altura possíveis para um ser humano. Pois bem, isso não é muito útil para nós, certo? Ter um intervalo muito grande (infinito) não é uma aproximação, é praticamente uma constatação - afinal, é óbvio que a altura de um ser humano está entre 0 a infinito e nem precisaríamos de uma amostra para dizer isso.

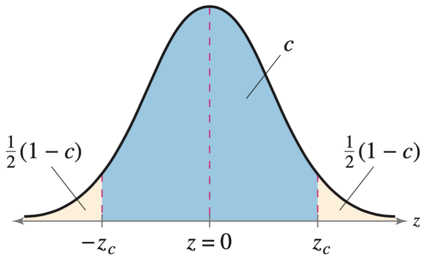
Por outro lado, se considerarmos um c muito pequeno (por exemplo, c = 5%), o seu intervalo vai ser bem menor, porém existe uma confiança muito pequena (5% nesse caso) do seu parâmetro não estar no seu intervalo. Suponhamos que o intervalo a 5% de confiança de altura das mulheres é 1,62 a 1,64. Essa altura não representa bem a maioria das mulheres no Brasil, certo? Apesar de termos um intervalo bem pequeno, ele não representa bem nossa população.

Grande parte dos estudos existentes considera c = 95%, podendo variar de 90 a 99% a depender da precisão necessária para seu estudo.

Como dissemos, para podermos aplicar o intervalo de confiança com as fórmulas a seguir, precisamos que a distribuição seja normal. Entretanto, você sabe do teorema do limite central que quando $n \geq 30$, a **média amostral** terá distribuição aproximadamente normal. Essa premissa é de suma importância quando queremos calcular um intervalo de confiança, pois tais intervalos são simétricos (mesma margem de erro para menos e para mais). Apenas podemos assumir intervalos de confiança simétricos se a própria distribuição é simétrica - ou seja, é uma normal.

INTERVALO DE CONFIANÇA PARA MÉDIAS QUANDO O DESVIO-PADRÃO POPULACIONAL É CONHECIDO

O **nível de confiança** corresponde à área sob a curva normal padrão entre os valores críticos. Aqui vamos chamar os **valores críticos** de **–zc e zc.**



Esses valores críticos, em geral, separam resultados prováveis (região central) de improváveis, ou incomuns (caudas).

Lembrando que a área total embaixo da curva é 1, e que **c é a área em azul**, então a área restante **(área bege) é 1 – c**.

Sabendo que a curva é simétrica, a área em cada cauda (cada uma das áreas beges) é (1 – *c*)/2.

Por exemplo, se *c* = 95%, então 2.5% da área está à esquerda e 2.5% está à direita. Dessa forma, **zc** deve ser a probabilidade para um determinado z em que a área seja azul + ½(1-c), ou seja, 95% + 2.5% = 97.5%. Logo, quando trabalhamos com 95% de confiança para encontrar um intervalo de confiança, a área acumulada até zc é 97.5%.

O z em questão aqui é o próprio z que corresponde a uma normal padrão, que vimos no capítulo de probabilidades e distribuição, o qual é tabelado. Quando o desvio-padrão da **população** é conhecido podemos usá-lo para representar os valores improváveis (a área bege do gráfico acima).

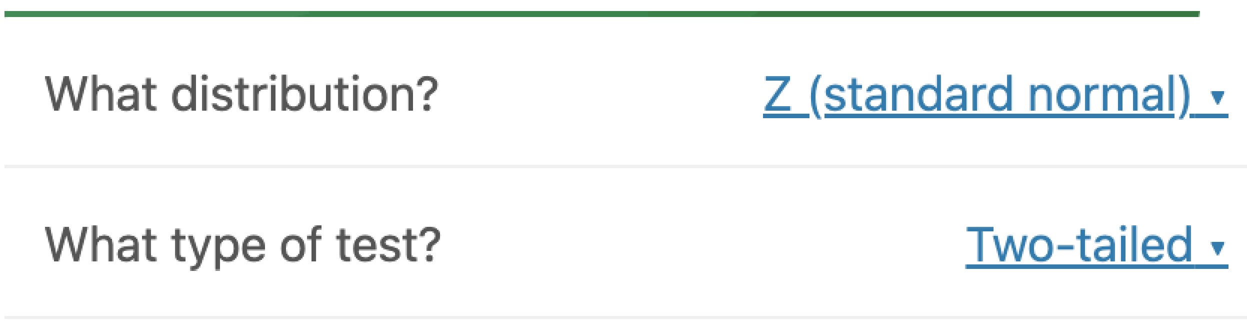
Para facilitar, abaixo segue o valor de z para os **intervalos de confiança** mais usuais:

Percentage Confidence	*z**-Value
80	1.28
90	1.645
95	1.96
98	2.33
99	2.58

Quando trabalhamos com 95% de confiança em um teste bicaudal, verificamos que $P(z<z_c) = 0.9750$ para um $z = 1.96$. Dessa forma, falamos que $z_c = 1.96$.

Você também pode encontrar os valores de z para cada confiança nesse site: https://www.omnicalculator.com/statistics/critical-value

No site acima, escolha:



O significance level (nível de significância) deve ser 1-c. Para 95%, deverá ser 0.05.

Disclaimer: Quando escolhemos o parâmetro "two-tailed" ele já distribui os 5% em ambas as caudas e, por isso, precisamos apenas passar 5% ao invés de 2,5%.

Dado um nível de confiança c, a **margem de erro E** (às vezes chamada também de erro máximo da estimativa ou tolerância de erro) é a maior distância possível entre a estimativa pontual e o valor do parâmetro que ela está estimando. Sua fórmula é dada por:

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

Usando uma estimativa pontual e uma margem de erro, você pode construir uma estimativa intervalar de um parâmetro populacional. Essa estimativa intervalar é chamada de **intervalo de confiança** e é representada por

$$\text{Confidence Interval} = \overline{X} \pm Z \times \frac{\sigma}{\sqrt{n}}$$

Em que CI é o intervalo de confiança, $\overline{x}$ é a média amostral, σ é o desvio-padrão populacional e n é o tamanho da amostra.

Exemplo: O diretor de admissões de uma faculdade deseja estimar a idade média de todos os estudantes atualmente matriculados. Em uma amostra aleatória de 20 estudantes, a idade média encontrada é de 22,9 anos e a distribuição é normal. De estudos anteriores, o desvio padrão populacional conhecido é de 1,5 ano, e a população é normalmente distribuída. Construa um intervalo de confiança de 90% da idade média da população.

Resposta: Primeiro, nossa distribuição é normal. Segundo, o desvio-padrão **populacional** é conhecido. Terceiro, a amostra é aleatória. Logo, podemos usar o intervalo de confiança com z.

Usando n = 20, $\overline{x}$ = 22,9, s = 1,5 e zc = 1,645 (99% de confiança), a margem de erro no intervalo de confiança de 90% é:

$$E = z_c \frac{\sigma}{\sqrt{n}} = 1{,}645 \cdot \frac{1{,}5}{\sqrt{20}} \approx 0{,}6.$$

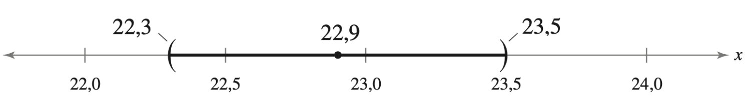
Limite inferior

$$\overline{x} - E \approx 22{,}9 - 0{,}6 = 22{,}3$$

Limite superior

$$\overline{x} + E \approx 22{,}9 + 0{,}6 = 23{,}5$$

$$22{,}3 < \mu < 23{,}5$$



Ou seja, se coletarmos amostras de 100 estudantes, 90 intervalos de confiança conterão a média populacional real. Como nosso caso temos apenas 1 amostra, podemos dizer que estamos 90% confiantes que nosso intervalo contém a média real. Em outras palavras, com 90% de confiança, você pode dizer que a idade média de todos os estudantes está entre 22,3 e 23,5 anos.

DISTRIBUIÇÃO T-STUDENT

Em muitas situações da vida real, o **desvio padrão da população é desconhecido.** Então, como podemos construir um intervalo de confiança para uma média populacional? Para uma variável aleatória que é normalmente distribuída (ou aproximadamente normalmente distribuída), a variável média amostral comporta-se tal qual outro modelo, a **distribuição t,** também chamado de **t-student.**

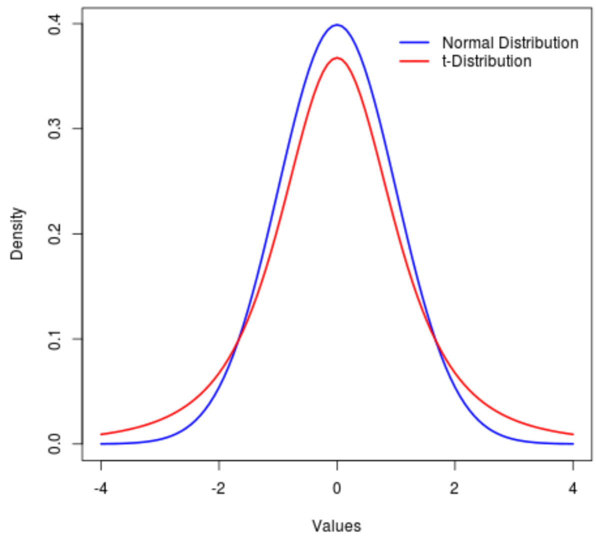
A distribuição t é uma **família de curvas**, cada uma determinada por um parâmetro chamado de **graus de liberdade**. Os graus de liberdade são o número de escolhas livres deixadas depois que uma estatística amostral tal como x é calculada. Quando usamos a distribuição t para estimar uma média populacional, os graus de liberdade são iguais ao tamanho da amostra menos um (n-1).

A distribuição t distribui-se também simetricamente com média 0, porém não é uma normal. O valor de t é dado por:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Em que s é o desvio-padrão **amostral** (note que não é o populacional!). Quando nossa amostra é muito grande, s tende a se aproximar do desvio-padrão populacional e, portanto, t passa a assumir exatamente a mesma fórmula de z - nesse caso, a distribuição passa a ser normal.

A distribuição t ao primeiro olhar realmente se parece com a distribuição normal, mas não é (a não ser para uma quantidade de amostras grandes, quando s se aproxima de σ). Em geral, t tem causas mais grossas do que a normal.



Da mesma forma que o z, os valores de t também são tabelados. A partir do grau de liberdade e de um nível de confiança desejável, conseguimos verificar o t-tabelado a partir da tabela-t.

Exemplo: Encontre o valor crítico tc para um nível de confiança de 95% quando o tamanho da amostra é 15.

Resposta:

Como n = 15, os graus de liberdade são:

Grau liberdade = n – 1 = 15 – 1 = 14.

Usando GL. = 14 e c = 0,95, você pode encontrar o valor crítico tc, como mostrado pelas áreas destacadas na tabela

	Nível de confiança, c	0,80	0,90	0,95	0,98	0,99
	Unilateral, α	0,10	0,05	0,025	0,01	0,005
g.l.	Bilateral, α	0,20	0,10	0,05	0,02	0,01
1		3,078	6,314	12,706	31,821	63,657
2		1,886	2,920	4,303	6,965	9,925
3		1,638	2,353	3,182	4,541	5,841
12		1,356	1,782	2,179	2,681	3,055
13		1,350	1,771	2,160	2,650	3,012
14		1,345	1,761	2,145	2,624	2,977
15		1,341	1,753	2,131	2,602	2,947
16		1,337	1,746	2,120	2,583	2,921

Da tabela, você pode ver que *tc* = 2,145.

Porém, ao longo do curso não usaremos a tabela! Vamos usar Sheets e Python para coletar esses valores quando necessário.

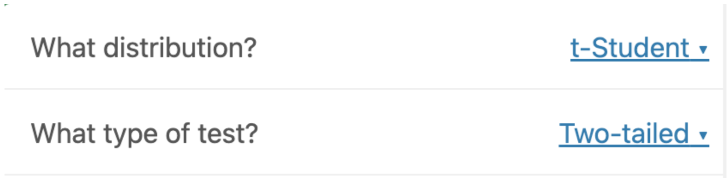
INTERVALO DE CONFIANÇA PARA MÉDIAS QUANDO O DESVIO-PADRÃO POPULACIONAL NÃO É CONHECIDO

Construir um intervalo de confiança quando o desvio-padrão populacional **não** é conhecido usando a distribuição *t* é similar a construir um intervalo de confiança quando o desvio-padrão é conhecido usando a distribuição normal — mudando apenas o parâmetro z para t. A fórmula abaixo é usada para encontrar a margem de erro

$$E = t_c \frac{s}{\sqrt{n}}$$

Exemplo: Você seleciona aleatoriamente 16 cafeterias e mede a temperatura do café vendido em cada uma delas. A temperatura média da amostra é 162,0 oF (fahrenheit) com desvio padrão de 10,0 oF. Construa um intervalo de confiança de 95% para a temperatura média da população de cafés vendidos. Suponha que as temperaturas tenham distribuição aproximadamente normal.

Resposta: Como o desvio-padrão populacional é desconhecido, a amostra é aleatória e as temperaturas têm distribuição aproximadamente normal, use a distribuição t. Sendo n = 16, x = 162,0, s = 10,0, c = 0,95 e g.l. = 15, você pode os valores de t nesse site:https://www.omnicalculator.com/statistics/critical-value. Escolhemos os seguintes parâmetros:



Em "degrees of freedom" devemos colocar nossos graus de liberdade (em nosso exercício, será 15) e em significance level devemos colocar 1-c, nesse caso será 0.05 (uma vez que c é 95%).

Logo, t será aproximadamente 2,131

What distribution?	t-Student ▾
What type of test?	Two-tailed ▾
Degrees of freedom (d)	15
Significance level	0.05

The test statistic follows the t-distribution with 15 degrees of freedom.

Critical value: ±2.1314

Critical region:

(-∞, -2.1314] ∪ [2.1314, ∞)

Voltando a fórmula, temos:

$$E = t_c \frac{s}{\sqrt{n}} = 2{,}131 \cdot \frac{10{,}0}{\sqrt{16}} \approx 5{,}3.$$

Com isso, conseguimos calcular nosso intervalo de confiança:

Limite inferior	Limite superior
$\bar{x} - E \approx 162 - 5{,}3 = 156{,}7$	$\bar{x} + E \approx 162 + 5{,}3 = 167{,}3$

$$156{,}7 < \mu < 167{,}3$$

Com 95% de confiança, você pode dizer que a temperatura média da população de cafés vendidos está entre 156,7 oF e 167,3 oF.

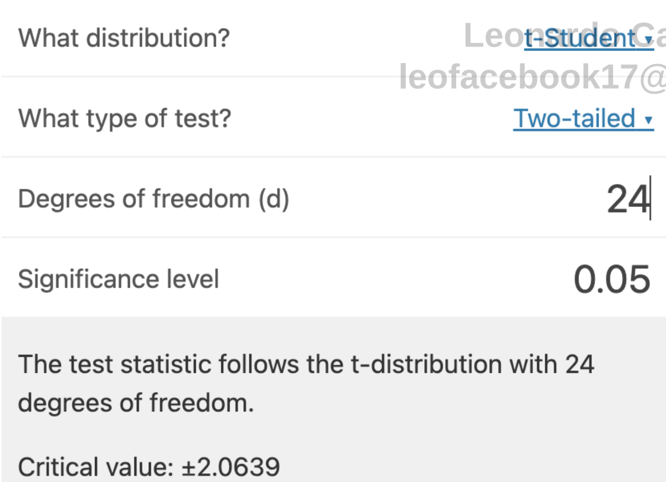
O INTERVALO DE CONFIANÇA CONCORDA MESMO COM O TESTE DE HIPÓTESE?

Vocês lembram que comentamos que o intervalo de confiança sempre concorda com o teste de hipótese? Pois bem, vamos mostrar isso com um exemplo!

Voltando ao nosso exemplo de média de gastos de combustível, temos:

Variável	**Valor**
$\bar{x}$	330.6
s	154.2
n	25

Usando o 95% de confiança, para g.l. = 25 - 1 = 24 o valor de t é 2,0639



Com esse valor de t, temos:

$$E = t * \frac{s}{\sqrt{n}} = 2,0639 * \frac{154,2}{\sqrt{25}} = 63,6$$

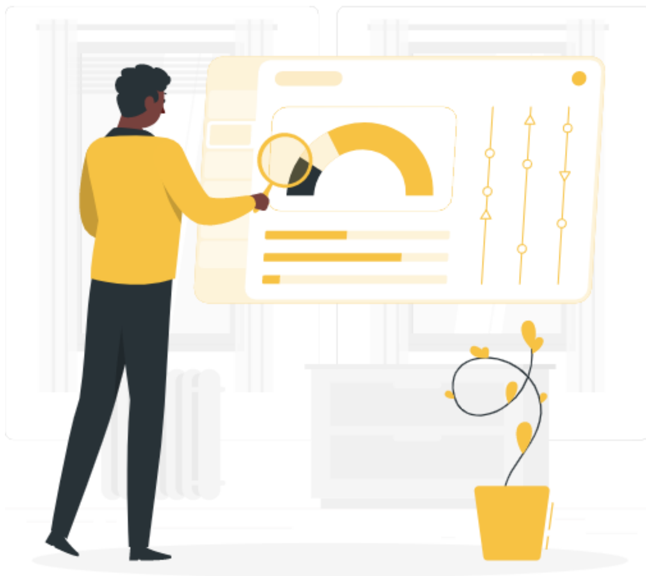
Logo, o intervalo de confiança é:

Limite inferior = 330,6 - 63,6 = 267
Limite superior = 330,6 + 63,6 = 394,2
Intervalo = [267   394,2]

Nossa ideia inicial era tentar ver, a 95%, se a hipótese nula estava correta, ou seja, se a média de gastos era de 260. Nos capítulos anteriores fizemos um teste de hipótese (que não mostramos os cálculos, mas obtivemos um p-valor) que rejeitava a hipótese nula a 95% de certeza.

Aqui podemos ver que o limite inferior ainda é superior a 260 a 95% de confiança. Como 260 não está dentro desses limites, dizemos que de fato a média dessa amostra não é 260. Porém, já deixo um spoiler, muito cuidado ao interpretar o intervalo de confiança! Vocês verão mais sobre isso no capítulo "Usos e abusos do intervalo de confiança"

11. TESTE DE HIPÓTESE PARA MÉDIAS



E finalmente chegamos nele, o tão aguardado teste de hipótese! Se vocês bem se lembram, usamos o teste de hipótese quando temos apenas amostras e nosso intuito é fazer afirmações sobre toda uma população. Devido aos erros aleatórios intrínsecos que acontecem quando usamos uma amostra, não podemos fazer afirmações diretamente em uma amostra sem um teste de hipótese adequado (ou o intervalo de confiança!).

Para realizar um teste de hipótese seguimos os seguintes passos:

Até agora falamos sobre os passos 1 e 2. Depois de construir as hipóteses nula e alternativa e especificar o nível de significância, o próximo passo em um teste de hipótese é obter uma **amostra aleatória** da população e calcular as estatísticas amostrais de interesse para aquele teste (tais como *média, proporção, desvio-padrão, etc*), correspondentes aos parâmetros na hipótese nula. Vamos a alguns casos de comparações que podemos usar um teste de hipótese:

1) Médias
   a) Queremos comprovar se nossa média de gastos de combustível é US$ 260. Não conseguimos coletar a população toda, por isso coletamos uma amostra e a média de gastos foi US$ 330,6. Será que podemos afirmar que a média populacional não é US$ 260? Precisamos de um teste de hipótese (ou intervalo de confiança) para comparar e ver se devemos rejeitar a hipótese nula ou não.
   b) Queremos testar 2 métodos de ensino e, para isso, selecionamos aleatoriamente 2 grupos de pessoas. No grupo 1 aplicamos o método A e no grupo 2 aplicamos o método B. Depois, esses 2 grupos fazem uma prova e calculamos a média da nota de cada um dos grupos. Nesse cálculo, vimos que a média do grupo 1 é maior do que a do grupo 2. Será que podemos afirmar que o método A é melhor que o B? Para esses grupos, com certeza podemos! Mas será que isso é aplicável a toda uma população de interesse? Precisamos de um teste de hipótese ou comparar os intervalos de confiança para tomarmos essa decisão.

2) Proporções
   a) Temos 2 grupos de pessoas escolhidas aleatoriamente e queremos testar uma vacina. O grupo A toma a vacina e o grupo B não toma. Medimos então a proporção de infectados no grupo A e no grupo B e vimos que a proporção é menor no grupo que tomou a vacina (grupo A). Será que podemos estender esse resultado à população? Precisamos de um teste de hipótese ou comparar os intervalos de confiança para tomarmos essa decisão.
   b) Fizemos uma melhoria em um site de e-commerce e agora queremos comprovar que a proporção de pessoas que efetivam a compra agora é de 20% . Podemos coletar uma amostra, medir a proporção de efetivação de compra e rodar um um teste de hipótese ou comparar os intervalos de confiança para vermos se de fato essa efetivação é de 20% para a população com base apenas na amostra.

3) Desvio-padrão
    a) Temos um processo industrial com alta variabilidade, que faz com que os produtos não sigam um padrão. Fizemos uma melhoria para reduzir a variabilidade, testamos uma amostra e agora precisamos comprovar se essa redução se estende a população.

Notem que em todos os exemplos podemos comparar uma amostra a um valor fixo ou 2 amostras. Mais para frente vocês também verão que também poderemos comparar mais que duas amostras.

Independente se queremos comparar médias, proporções ou desvios, temos uma denominação chamada **teste de hipótese 1 amostra** (hypothesis testing 1 sample) quando temos uma 1 amostra e queremos compará-la a um valor fixo. Também podemos comparar 2 amostras e, nesse caso, temos o que chamamos de **teste de hipótese 2 amostra** (hypothesis testing 2 samples/)

A estatística amostral de interesse é chamada de **estatística de teste**. Sob a suposição de que a hipótese nula é verdadeira, o valor específico da variável de teste (a média, por exemplo) é então transformada em uma estatística de teste, tal como z, t, f ou χ^2. A estatística de teste padronizada é usada na tomada de decisão sobre a rejeição ou não da hipótese nula.

Nessa seção vamos falar sobre a **média** especificamente (1 sample e 2 samples), mas nas próximas sessões falaremos sobre comparações de proporção e desvio-padrão. Para a média, nossas estatísticas de teste são a **z** e a **t**, a depender do caso.

1 SAMPLE

DESVIO-PADRÃO POPULACIONAL CONHECIDO

Quando temos um desvio-padrão populacional conhecido, padronizamos nossa média com o **teste z.**

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Podemos calcular a área embaixo da curva para valores mais extremos de z (lembrem-se que os valores de z são tabelados). Essa área para valores mais extremos corresponde ao nosso p-valor.

Vamos direto a um exemplo para deixar isso mais claro.

Exemplo: Em corrida de carros, o pit stop é o local em que um veículo vai para trocar pneus, abastecer, efetuar reparos e outros ajustes mecânicos. A eficiência de uma equipe que realiza esses ajustes pode afetar o resultado de uma corrida. Uma equipe afirma que seu tempo médio no pit stop (para 4 trocas de pneus e abastecimento) é **menor que 13 segundos**. Uma amostra aleatória de **32 tempos** de pit stop tem uma **média de 12,9 segundos**. Suponha que o **desvio padrão populacional é de 0,19 segundos.** Há evidência suficiente para concordar com a afirmação para 99% de confiança?

Resposta:

Para 99% de confiança, $\alpha = 1\text{-}0.99 = 0.01$

Como o desvio-padrão **populacional** é conhecido (desvio = 0,19), a amostra é aleatória e $n = 32 \geq 30$ (garantindo a normalidade pelo teorema do limite central), você pode usar o **teste *z*.** A afirmação é "o tempo médio no pit stop é menor que 13 segundos". Então, as hipóteses nula e alternativa são:

H0: média ≥ 13 segundos

Ha: média < 13 segundos

Notem que aqui estamos falando de um **teste unicaudal à esquerda!**

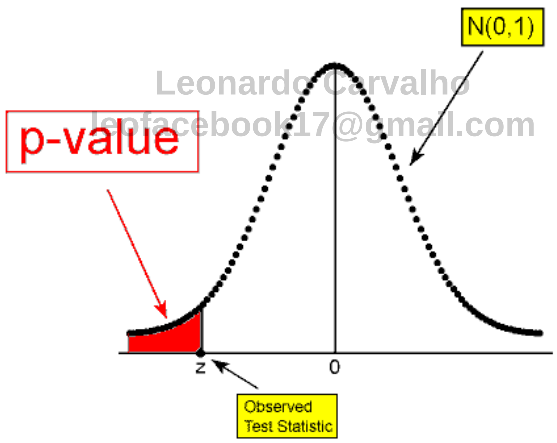
Logo, nosso **z calculado** é:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{12{,}9 - 13}{0{,}19 / \sqrt{32}}$$

$$\approx -2{,}98.$$

Vamos relembrar rapidamente o que é o p-valor. Para isso, vamos usar a figura abaixo para exemplificar.



A curva acima representa o cenário em que H0 seria verdadeiro (lembrem-se, esse cenário é fictício pois ainda veremos se vamos aceitar ou rejeitar H0). O **z calculado é o "observed test statistic"** e o p-valor seria a área vermelha. Dessa forma, dizemos que o p-valor é a probabilidade (área embaixo da curva) de obtermos o valor amostral que tivemos (no caso, 13 segundos) ou mais extremo (no caso, < 13 segundos) **SE** a hipótese nula estiver correta. Quando essa probabilidade é muito pequena, nós rejeitamos H0.

E o que é "muito pequeno"? É aí que entra o nosso α, que podemos entender como uma probabilidade limite para aceitarmos o H0. Se o p-valor é menor do que α, a probabilidade de ocorrer um valor igual ou mais extremo quando

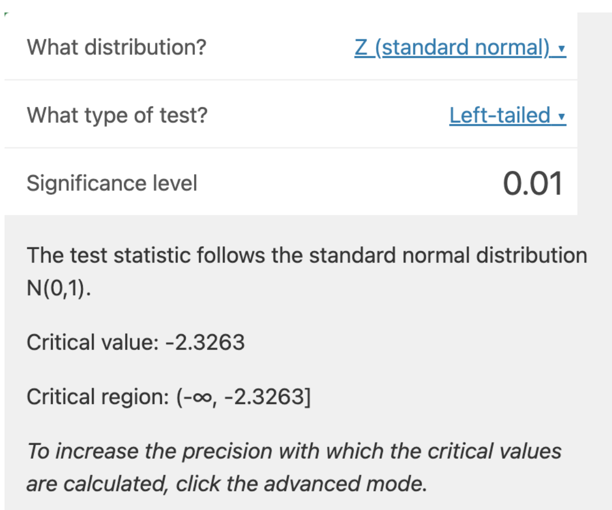
H0 é verdadeiro é TÃO pequena que não podemos aceitar H0 e, portanto, dizemos que H0 é mentira (em palavras mais bonitas, rejeitamos H0).

Dito isso, vocês devem estar pensando "ok, mas eu só tenho o z (observed test statistic) e não o p-valor (área embaixo da curva). E agora?". Agora a estatística te dá 2 opções:

1) Você pode encontrar o z que corresponde a α (esse z é chamado de **z crítico)** e compará-lo com z calculado. Se z calculado for menor que o z crítico, dizemos que a probabilidade mínima para aceitarmos H0 (α) não foi atendida, e, portanto, rejeitamos o H0.
2) Você pode encontrar a probabilidade (área embaixo da curva) que corresponde o z calculado. Essa área é justamente o seu **p-valor**! Com o p-valor, você pode compará-lo com α e seguir o mesmo raciocínio do item anterior.

Vamos fazer ambas as formas para demonstrar.

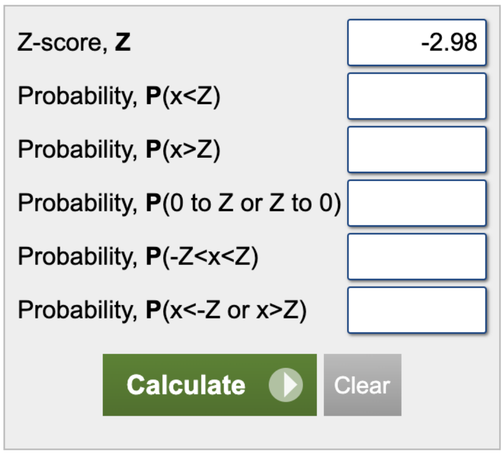
1) Nosso z crítico é o z que corresponde a α = 1%, unicaudal à esquerda (left-tail). Usando o site que já conhecemos https://www.omnicalculator.com/statistics/critical-value temos:



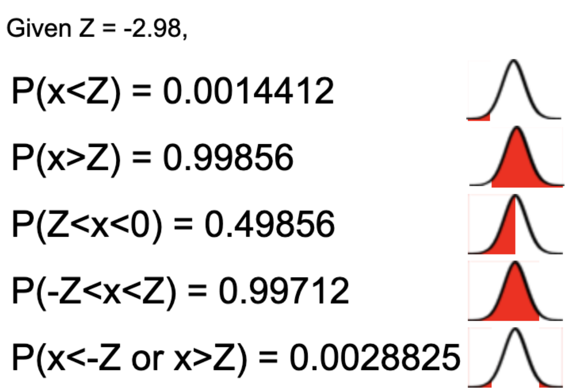
Z crítico = -2,3263. Portanto, z crítico é maior do que z calculado (-2,98) e, portanto, o z calculado cai na zona de rejeição. Ou seja, há evidência suficiente ao **nível de significância de 1%** para **concordar** com a afirmação de que o tempo médio no pit stop é menor que 13 segundos (Ha).

2) Aqui vamos usar um novo site:
https://www.calculator.net/z-score-calculator.html

Apontando z = -2,98, podemos usar o item "Z-score and Probability Converter"para calcular a área embaixo da curva. Nesse caso, basta substituir z por -2,98 e clicar em "calculate"



O resultado será



A primeira figura é exatamente o que estamos procurando: valores mais extremos que z e apenas menores do que z.

Nesse caso, temos que a área embaixo da curva é 0,0014412 = 0,14412%. Esse é o nosso p-valor.

Uma vez que o p-valor é **menor** que 1%, você **rejeita** a hipótese nula. Ou seja, há evidência suficiente ao **nível de significância de 1%** para **concordar** com a afirmação de que o tempo médio no pit stop é menor que 13 segundos (Ha).

DESVIO-PADRÃO POPULACIONAL DESCONHECIDO

Da mesma forma que acontece para o intervalo de confiança, é muito mais comum os casos de não termos o desvio-padrão populacional conhecido e termos apenas o desvio-padrão amostral. Semelhante ao intervalo de confiança, nesses casos também usamos a distribuição t-student para nossos cálculos. Nesse caso, temos que calcular o t por:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Em que s representa o desvio-padrão amostral.

Vamos a um exemplo

Exemplo: Uma indústria afirma que o nível médio do pH da água em um rio próximo é de 6,8. Você seleciona aleatoriamente 39 amostras de água e mede o pH de cada uma. A média amostral e o desvio padrão são de 6,7 e 0,35, respectivamente. Há evidência suficiente para rejeitar a afirmação da indústria considerando nível de significância α = 0,05?

Resposta:

Notem que aqui só temos o desvio padrão amostral! Também sabemos que n>30 e estamos querendo fazer os cálculos de médias amostrais, então podemos usar o teorema do limite central que garante normalidade na distribuição de médias amostrais. Portanto, podemos usar o teste t.

H0 : média = 6,8 (Afirmação)

Ha : média ≠ 6,8

O teste é bilateral (two-tail), o nível de significância é 0,05 e os graus de liberdade são g.l. = 39 – 1 = 38.

Da mesma forma que antes, temos 2 opções de cálculo

1) Você pode encontrar o t que corresponde a **α** (esse t é chamado de **t crítico)** e compará-lo com t calculado. Se t calculado for menor que o t crítico, dizemos que a probabilidade mínima para aceitarmos H0 (α) não foi atendida, e, portanto, rejeitamos o H0.
2) Você pode encontrar a probabilidade (área embaixo da curva) que corresponde o t calculado. Essa área é justamente o seu **p-valor**! Com o p-valor, você pode compará-lo com **α** e seguir o mesmo raciocínio do item anterior.

Agora, vamos usar apenas a abordagem do p-valor.

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Como σ é desconhecido e $n \geq 30$, use o teste t.

$$= \frac{6{,}7-6{,}8}{0{,}35/\sqrt{39}}$$

Suponha que $\mu = 6{,}8$.

$$\approx -1{,}784.$$

Arredonde para três casas decimais.

Agora vamos usar o seguinte site para ver qual seria o p-valor correspondente a esse t: https://www.socscistatistics.com/pvalues/tdistribution.aspx

P Value from T Score Calculator

This should be self-explanatory, but just in case it's not: your *t*-score
degrees of freedom in the *DF* box (*N - 1* for single sample and depen
independent samples), select your significance level and whether yo
hypothesis (if you're not sure, go with the defaults), then press the b

If you need to derive a T Score from raw data, then you can find t tes

Report a T-Test Result (APA)

T Score: -1.784

DF: 39

Significance Level:

- ○ .01
- ◉ .05
- ○ .10

One-tailed or two-tailed hypothesis?:

- ○ One-tailed
- ◉ Two-tailed

The *p*-value is .082206.

The result is *not* significant at *p* < .05.

Calculate

Não esqueçam que o teste é bicaudal (two-tailed).

Dessa forma, como p-valor = 0,0822 dizemos que devemos rejeitar a hipótese nula. Ou seja, não podemos afirmar que a média da população é 6,8.

2 SAMPLE

Na seção anterior você estudou métodos para testar uma afirmação sobre o valor de um parâmetro populacional. Agora você aprenderá como testar uma afirmação comparando 2 médias de duas populações. Antes de aprender como testar a diferença entre dois parâmetros, você precisa entender a diferença entre amostras independentes e **amostras dependentes.**

Duas **amostras são independentes** quando a amostra selecionada de uma população não é relacionada à amostra selecionada da segunda população. Duas **amostras são dependentes** quando cada elemento de uma amostra corresponde a um elemento da outra amostra. Amostras dependentes também são chamadas de amostras pareadas ou amostras emparelhadas.

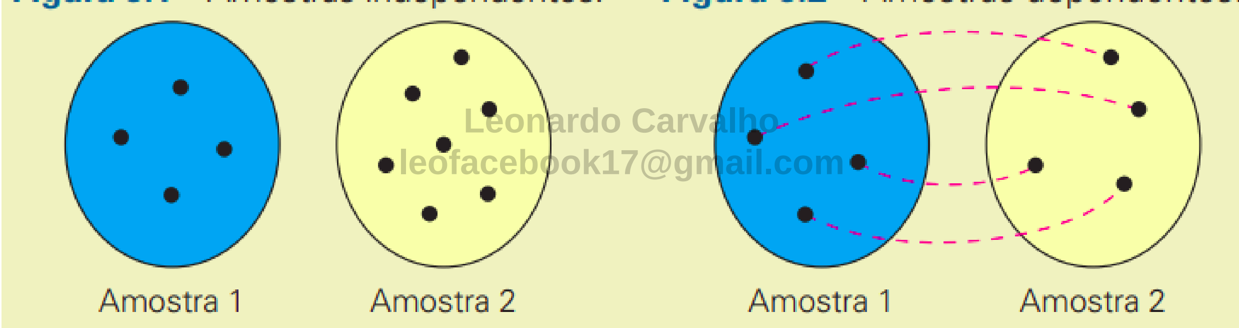


Figura 8.1 Amostras independentes. **Figura 8.2** Amostras dependentes.

Por exemplo

1) Amostra 1: pesos de 65 calouros universitários antes do início das aulas.
   Amostra 2: pesos dos mesmos 65 calouros após o primeiro ano.

   As amostras são dependentes. Como os pesos dos mesmos estudantes são medidos, as amostras são relacionadas. As amostras podem ser pareadas em relação a cada estudante.

2) Amostra 1: pontuações de 38 homens adultos em um teste psicológico para transtorno do déficit de atenção com hiperatividade.
   Amostra 2: pontuações de 50 mulheres adultas em um teste psicológico para transtorno do déficit de atenção com hiperatividade.

As amostras são independentes. Não é possível formar pares entre os elementos das amostras, pois os tamanhos das amostras são diferentes e os dados representam pontuações para indivíduos diferentes

UMA VISÃO GERAL DO TESTE DE HIPÓTESE USANDO DUAS AMOSTRAS

Nesta seção, você aprenderá como testar uma afirmação comparando as médias de duas populações usando amostras independentes.

Por exemplo, um provedor de serviço de internet está desenvolvendo um plano de marketing para determinar se há diferença nos tempos que estudantes universitários do sexo masculino e feminino passam conectados à internet por dia. A única maneira de se concluir com certeza que há diferença é fazendo um censo de todos os universitários, calculando os tempos médios diários que os estudantes do sexo masculino e do sexo feminino ficam conectados e encontrando a diferença. É claro que não é prático fazer esse censo. No entanto, é possível determinar com algum grau de certeza se tal diferença existe.

Para determinar se existe uma diferença, o provedor de serviço de internet começa assumindo que não há diferença no tempo médio das duas populações. Isto é:

$\mu 1 - \mu 2 = 0$.

Então, retirando uma amostra aleatória de cada população, um teste de hipótese baseado nas duas amostras é realizado usando a estatística de teste:

$x1 - x2$

Lembrando que:

- A hipótese nula H0 é uma hipótese estatística que geralmente diz que não há diferença entre os parâmetros de duas populações. A hipótese nula sempre contém o símbolo de igualdade
- A hipótese alternativa Ha é uma hipótese estatística que é verdadeira quando H0 é falsa.

Para cada cenário, temos uma fórmula para calcular o teste de hipótese de 2 médias.

DESVIO-PADRÃO POPULACIONAL CONHECIDO E AMOSTRAS INDEPENDENTES

Podemos usar um teste z para a diferença entre duas médias populacionais μ1 e μ2, quando as amostras são independentes. As condições a seguir são necessárias para realizar tal teste.

1. Os desvios padrão populacionais são conhecidos.

2. As amostras são selecionadas aleatoriamente.

3. As amostras são independentes.

4. As populações são normalmente distribuídas ou cada tamanho de amostra é de pelo menos 30.

Quando esses requisitos são satisfeitos, a distribuição amostral para x1 – x2 , ou seja, a diferença das médias das amostras, é uma distribuição normal com média e erro padrão conforme mostrado abaixo:

Em palavras	Em símbolos
A média da diferença das médias amostrais é a diferença presumida entre as duas médias populacionais. Quando nenhuma diferença é presumida, a média é 0.	Média $= \mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$
A variância da distribuição amostral é a soma das variâncias das distribuições amostrais individuais para $\bar{x}_1$ e $\bar{x}_2$. O erro padrão é a raiz quadrada dessa soma.	Erro padrão $= \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

O z a ser calculado nesse caso é:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Vamos explicar melhor a fórmula e o teste através de um exemplo.

Um grupo de analistas afirmou que existe diferença entre as médias dos gastos em cartões de crédito em São Paulo e no Rio de Janeiro. Eles se basearam em amostras (independente) de 250 pessoas de cada um desses estados e tiverem esse resultado:

Média de gastos no cartão para 250 pessoas no Rio de Janeiro:

x1 = R$ 4.777

Média de gastos no cartão para 250 pessoas em São Paulo:

x2 = R$ 4.866

Esses analistas, como não estudaram estatística a fundo, viram apenas a média e disseram que o pessoal de São Paulo gasta mais que o pessoal do Rio.

Você, como está estudando comigo, desconfia desse resultado. Então você descobre, de estudos anteriores, que você pode aproximar o desvio-padrão populacional para o Rio para R$ 1.045 e para São Paulo para R$ 1.350. Ve

As condições a seguir são necessárias para realizar testes de hipótese de amostras independentes:

1. Os desvios padrão populacionais são conhecidos
2. As amostras são selecionadas aleatoriamente
3. As amostras são independentes
4. As populações são normalmente distribuídas ou cada tamanho de amostra é de pelo menos 30

Seguindo todos os requisitos necessários, você pode usar o teste z para testar a diferença entre duas médias:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Sabendo que o problema do gasto médio no cartão cobre todos os requisitos acima e que agora você tem os desvios-padrão de cada uma das amostras, você decide aplicar um teste a 95% de confiança com a seguinte hipótese (lembrem-se que a hipótese de "igualdade" deve estar sempre na hipótese nula):

$$H_0: \mu_1 = \mu_2 \qquad \text{e} \qquad H_a: \mu_1 \neq \mu_2.$$

Na fórmula acima, $\mu_1 - \mu_2$ é nossa “premissa da hipótese nula”. Dessa forma, usamos a hipótese nula como valor $\mu_1 - \mu_2$, ou seja, se $\mu_1 - \mu_2 = 0$ pois, em nossa hipótese nula, estamos afirmando que $\mu_1 = \mu_2$ (e vamos provar se isso é verdade já já). Notem que esse teste é bicaudal!

O nível de significância é 0,05 (ou seja, 1-0.95)

Usando a fórmula, temos:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(4.777 - 4.866) - 0}{\sqrt{\frac{1.045^2}{250} + \frac{1.350^2}{250}}}$$

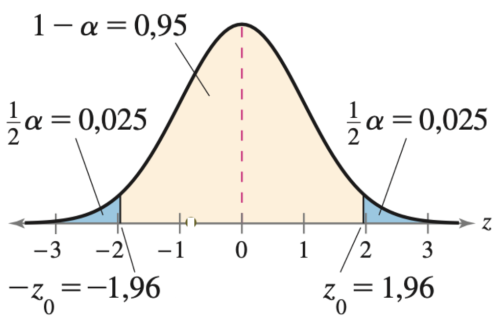
$$\approx -0{,}82.$$

Calcule no site caso prefira:

https://www.statology.org/two-sample-z-test-calculator/

A área que compreende z = -0.82 é 0,4108 (bicaudal), que corresponde ao nosso p-valor. O p-valor calculado é **maior** do que nosso nível de significância (0.05), ou seja **não** devemos rejeitar a hipótese nula. Portanto, a 95% de confiança, dizemos que a média é igual em ambos os grupos. Portanto, São Paulo não tem maiores gastos do que o Rio de Janeiro.

Outra forma de ver o resultado é pela **zona de rejeição**. Como o teste é bilateral (hipótese nula tem um "≠"), temos:



- Área azul deve ter 0.025 pois a confiança é de 95% e o teste é bilateral
- Pela tabela de z, devemos calcular um acumulado. Ou seja, a área azul à esquerda + área bege. Dessa forma, a área acumulada é 0.975.
- O valor de z que corresponde a essa área é 1,96 (procurado na tabela)
- Dada a simetria da distribuição normal, sabemos que -z é -1.96.

As áreas de rejeição da hipótese são as áreas em azul. Sabemos que a área bege vai de -1.96 até 1.96 e que o nosso z calculado é de -0.82. Como o z calculado não está dentro das áreas de rejeição **não** devemos rejeitar a hipótese nula. Portanto, a 95% de confiança, dizemos que a média é igual em ambos os grupos. Portanto, São Paulo não tem maiores gastos do que o Rio de Janeiro.

DESVIO-PADRÃO POPULACIONAL DESCONHECIDO E AMOSTRAS INDEPENDENTES

Em muitas situações da vida real os desvios padrão populacionais ***não*** são conhecidos. Logo, usamos o **teste *t*** para testar a diferença entre duas médias populacionais com desvio-padrão populacional desconhecido. Para usar um teste *t*, as condições a seguir são necessárias:

1. Os desvios padrão populacionais são desconhecidos

Renata Biaggi

2. As amostras são selecionadas aleatoriamente
3. As amostras são independentes
4. As amostras são normalmente distribuídas *ou* cada tamanho de amostra é de pelo menos 30

O teste t para 2 médias é escrito como:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

Em que S_{x1-x2} é definido de 2 formas:

- Se as **variâncias populacionais são consideradas iguais**, então as variâncias das duas amostras são combinadas para se calcular uma estimativa conjunta do desvio padrão. Nesse caso, S_{x1-x2} é:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

e $g.l = n_1 + n_2 - 2$.

- Se as **variâncias populacionais não são iguais ou se não sabemos**, então S_{x1-x2} é:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

O grau de liberdade nesse caso é dado por:

$$v = \frac{(A + B)^2}{A^2/(n - 1) + B^2/(m - 1)}$$

Em que A $= s_1^2/n_1$ e B $= s_2^2/n_2$ e n = n_1 e m = n_2 ` `

Como esse valor é geralmente fracionário, arredonde para o inteiro mais próximo para obter o número de graus de liberdade

Exemplo

Os resultados de um teste estadual de matemática para amostras aleatórias de estudantes ensinados por dois professores diferentes na mesma escola. Podemos concluir que há diferença nas pontuações médias dos testes de matemática para todos os estudantes dos dois professores? Use confiança de 90%. Suponha que as populações são normalmente distribuídas e que as variâncias populacionais não são iguais.

Professor 1	Professor 2
$\bar{x}_1 = 473$	$\bar{x}_2 = 459$
$s_1 = 39{,}7$	$s_2 = 24{,}5$
$n_1 = 8$	$n_2 = 18$

Resposta:

Nossa hipótese deve ser:

$$H_0\colon \mu_1 = \mu_2 \qquad \text{e} \qquad H_a\colon \mu_1 \neq \mu_2.$$

Para variâncias não iguais:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ Use o teste t (as variâncias *não* são iguais).

$$= \frac{(473 - 459) - 0}{\sqrt{\frac{(39{,}7)^2}{8} + \frac{(24{,}5)^2}{18}}}$$ Suponha que $\mu_1 = \mu_2$, então $\mu_1 - \mu_2 = 0$.

$$\approx 0{,}922.$$ Arredonde para três casas decimais.

O grau de liberdade deve ser calculado por:

$$v = \frac{(A + B)^2}{A^2/(n - 1) + B^2/(m - 1)}$$

A $=39,7^2/8$ = 4,96

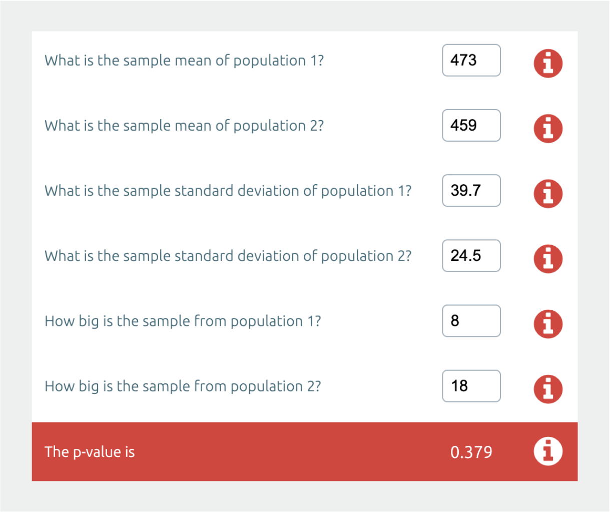
B $=24,5^2/18$ = 33,34

n = 8

m = 18

Logo, gl $=(4,96 + 33,34)^2 / [4,96^2/(8\text{-}1) + 33,34^2/(18\text{-}1)]$ = 1466.89/[3.51+65.38] = 21.29 = aprox. 21

Podemos procurar na nossa tabela t o valor de t = 0.922 com gl = 21 para encontrar a área acumulada ou então usar nossa calculadora: https://select-statistics.co.uk/calculators/two-sample-t-test-calculator/



P-valor é maior que alpha (0.1). Ou seja, não rejeitamos a hipótese nula ao nível de significância de 10%, portanto dizemos que a esse nível de confiança, as médias são estatisticamente iguais.

DESVIO-PADRÃO POPULACIONAL DESCONHECIDO E AMOSTRAS DEPENDENTES

Para realizar um teste de hipótese usando duas amostras dependentes, você usará uma técnica diferente. Você calculará primeiro a diferença *d* entre os elementos de cada par de dados:

d = (valor do dado na primeira amostra) – (correspondente valor do dado na segunda amostra).

Para esse teste, alguns critérios são importantes:

1. As amostras são selecionadas aleatoriamente.
2. As amostras são dependentes (emparelhadas)
3. As populações são normalmente distribuídas *ou* o número *n* de pares de dados é pelo menos 30.

Quando essas condições são satisfeitas, a distribuição amostral para *d* é aproximada por uma distribuição *t* com *n* – 1 graus de liberdade, em que *n* é o número de pares de dados.

Para esses casos, precisamos seguir os seguintes passos:

1. Calcular $\bar{d}$ por:

$$\bar{d} = \frac{\Sigma d}{n}$$

2. Calcular sd

$$s_d = \sqrt{\frac{\Sigma d^2 - \left[\frac{(\Sigma d)^2}{n}\right]}{n-1}}$$

3. Calcular t por

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

Exemplo

Um fabricante de tênis afirma que os atletas podem aumentar a altura de seus saltos verticais quando usam seu tênis por alguns meses. Os saltos de 8 atletas aleatoriamente selecionados são medidos. Após usarem os calçados por 8 meses, os saltos são novamente medidos. Com a = 0,10, há evidência suficiente para aceitar a afirmação do fabricante? Suponha que os saltos são normalmente distribuídos

Resposta:

Atleta	1	2	3	4	5	6	7	8
Altura do salto vertical (antes de usar o calçado)	24	22	25	28	35	32	30	27
Altura do salto vertical (após usar o calçado)	26	25	25	29	33	34	35	30

Como as amostras são aleatórias e dependentes, pois são as mesmas pessoas sendo avaliadas em um antes x depois, e as populações, normalmente distribuídas, você pode usar o teste *t*.

A afirmação é que "os atletas podem aumentar a altura de seus saltos verticais". Em outras palavras, o fabricante afirma que a altura do salto vertical de um atleta antes de usar o calçado será menor que a altura após usar o calçado. Cada diferença é dada por:

d **= (altura do salto antes do calçado) – (altura do salto após o calçado). As hipóteses nula e alternativa são:**

$H_0: \mu_d \geq 0$ e $H_a: \mu_d < 0$. (Afirmação.)

Logo, aqui estamos falando de um teste unilateral à esquerda (Ha contém "<").

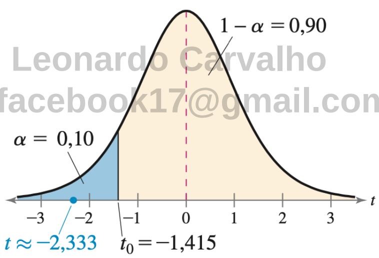
Como o teste é unilateral à esquerda, a = 0,10 e g.l. = 8 – 1 = 7, o valor crítico é $t0$ = –1,415 (vide site - teste unilateral). A região de rejeição é t < –1,415.

Usando as fórmulas anteriores:

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-14}{8} = -1,75.$$

$$s_d = \sqrt{\frac{\Sigma d^2 - \left[\frac{(\Sigma d)^2}{n}\right]}{n-1}} = \sqrt{\frac{56 - \frac{(-14)^2}{8}}{8-1}} \approx 2,1213.$$

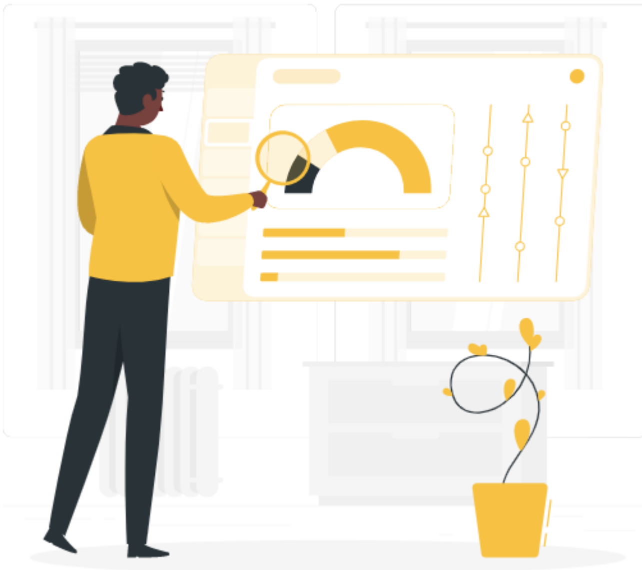
$$\begin{aligned} t &= \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \\ &\approx \frac{-1,75 - 0}{2,1213 / \sqrt{8}} \\ &\approx -2,333. \end{aligned}$$



Logo, como t calculado (-2,33) está na região de rejeição, há evidência suficiente, ao nível de significância de 10%, para concordar com a afirmação do fabricante de calçados de que os atletas podem aumentar a altura de seus saltos verticais usando o calçado de treinamento do fabricante.

Você pode chegar no mesmo resultado encontrando o p-valor para o t calculado e comparando com alpha = 0.1.

12. USOS E ABUSOS DO INTERVALO DE CONFIANÇA



As seções anteriores mostraram diferentes maneiras de comparar médias usando testes t e z.

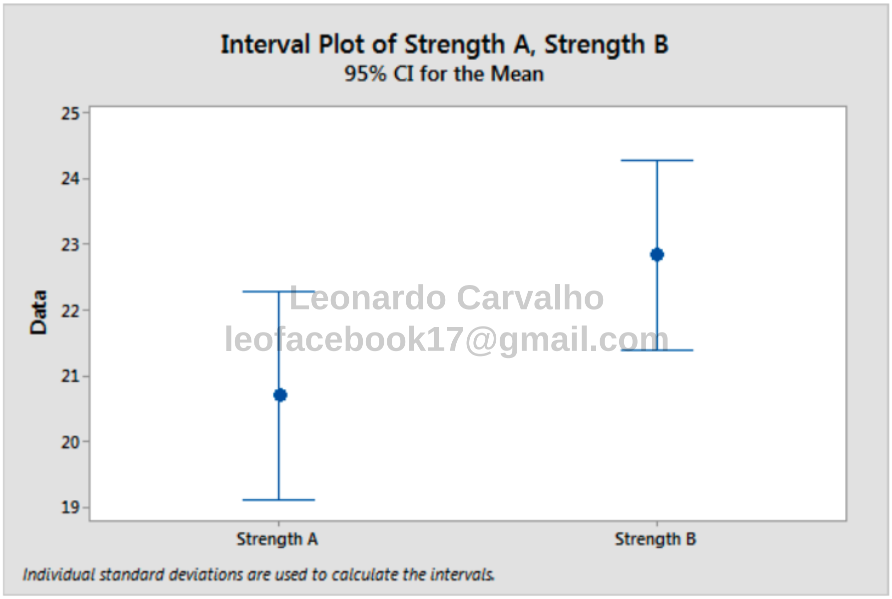
Agora, quero mostrar uma maneira de NÃO comparar dois meios que já vi pessoas usarem com muita frequência.

Muitos frequentemente comparam os intervalos de confiança para **duas** amostras (2 samples) para determinar se a diferença entre duas médias ou proporções (vocês verão isso a seguir) é estatisticamente significativa. Se esses intervalos se sobrepõem, eles concluem que a diferença entre os grupos não é estatisticamente significativa. Se não houver sobreposição, a diferença é significativa. A ideia por trás disso é: se um intervalo de confiança de uma amostra contém valores do intervalo de confiança de outra amostra, está provado que não há diferença estatística entre esses dois grupos.

Embora esse método visual de avaliar a sobreposição seja fácil de executar, ele reduz sua capacidade de detectar diferenças. Felizmente, existe uma solução simples para esse problema que permite realizar uma avaliação visual simples e ainda assim não diminuir o poder de sua análise.

Vou começar mostrando o problema em ação e explicar por que isso acontece. Em seguida, prosseguiremos para um método alternativo fácil que evita esse problema.

Determinar se os intervalos de confiança se sobrepõem é uma abordagem **excessivamente conservadora** para identificar diferenças significativas entre os grupos. É verdade que **quando os intervalos de confiança não se sobrepõem, a diferença entre os grupos é estatisticamente significativa. No entanto, quando há alguma sobreposição, a diferença ainda pode ser significativa.**



Ao ver como esses intervalos se sobrepõem, você conclui que a diferença entre as médias dos grupos não é estatisticamente significativa. Afinal, se eles estão sobrepostos, eles não são diferentes, certo?

Essa conclusão parece lógica, mas não é necessariamente verdadeira. Os resultados do teste t de 2 amostras são estatisticamente significativos com um valor p de 0,044. Apesar dos intervalos de confiança sobrepostos, a diferença entre essas duas médias é estatisticamente significativa.

Este exemplo mostra como o método de sobreposição de IC falha em rejeitar a hipótese nula com mais frequência do que o teste de hipótese correspondente. O uso desse método diminui sua capacidade de detectar diferenças, fazendo com que você perca descobertas essenciais.

Essa aparente discrepância entre os intervalos de confiança e os resultados dos testes de hipóteses pode te deixar surpreso, já que dissemos anteriormente que o intervalo de confiança e o teste de hipótese sempre concordam.

O problema ocorre porque **não estamos comparando os intervalos de confiança corretos com o resultado do teste de hipóteses.** Os resultados do teste se aplicam à **diferença entre as médias**, enquanto os ICs se **aplicam à estimativa da média de cada grupo**, não à diferença entre as médias. Estamos comparando maçãs com bananas, então não é surpresa que os resultados sejam diferentes.

Para obter resultados consistentes, devemos usar intervalos de confiança para **diferenças entre as médias dos grupos**.

Esse tipo de IC sempre concordará com o teste de 2 amostras – apenas certifique-se de usar a combinação equivalente de nível de confiança e nível de significância (por exemplo, 95% e 5%). Agora estamos comparando maçãs com maçãs!

Usando o mesmo conjunto de dados, o intervalo de confiança abaixo apresenta uma faixa de valores que provavelmente contém a diferença média para toda a população. A interpretação continua sendo uma simples avaliação visual. Zero representa nenhuma diferença entre as médias. O intervalo contém zero? Se não incluir zero, a diferença é estatisticamente significativa porque o intervalo não exclui nenhuma diferença. De relance, podemos dizer que a diferença é estatisticamente significativa.

Vamos a um exemplo:

Amostra A

Média 22.84
Desvio-padrão amostral: 3.08
n: 20

Amostra B

Média 20.69
Desvio-padrão amostral: 3.41
n: 20

Vamos supor que as populações têm distribuições normais, as amostras são independentes, o desvio-padrão populacional é desconhecido e que:

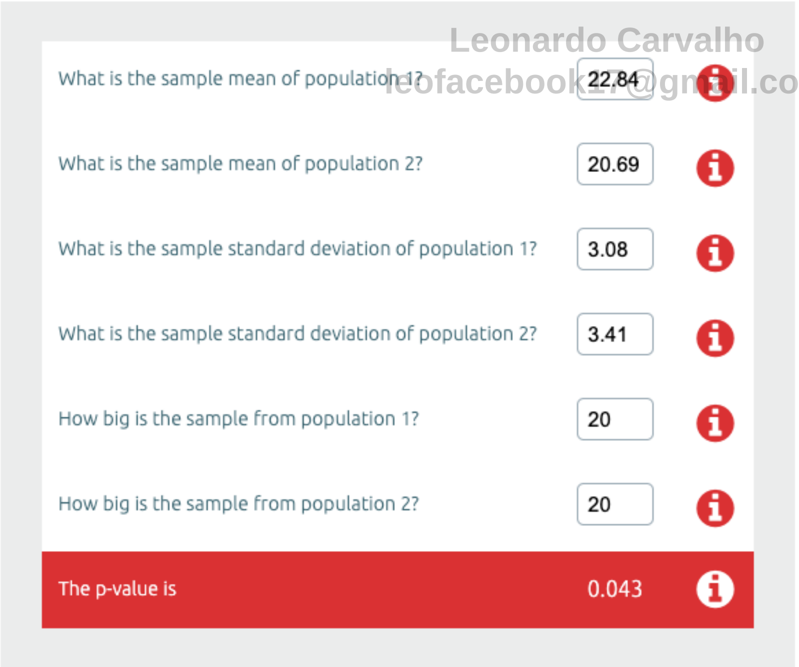
H0: Médias são iguais
Ha: Média são diferentes

Teste de hipótese

Nesse caso, usamos o teste t para variâncias populacionais não iguais

Usando o site para cálculo do p-valor (2 sample t-test)

https://select-statistics.co.uk/calculators/two-sample-t-test-calculator/



Ou seja, p-valor menor que alpha (0.05) e, portanto, rejeitamos a hipótese nula (médias não são iguais).

INTERVALO DE CONFIANÇA PARA DIFERENÇA ENTRE DUAS MÉDIAS

Dependendo dos tipos de amostra e se o desvio padrão da população é conhecido ou não, usaremos um teste z ou um teste t. Porém, a grande chave aqui é que precisamos construir o intervalo de confiança para a diferença.

Um intervalo de confiança (C.I.) para uma diferença entre médias é um intervalo de valores que provavelmente contém a verdadeira diferença entre duas médias populacionais com um certo nível de confiança.

Para estimar essa diferença, coletamos uma amostra aleatória de cada população e calculamos a média para cada amostra. No entanto, não sabemos com certeza se a diferença nas médias da amostra corresponde à verdadeira diferença nas médias da população e é por isso que eles podemos criar um intervalo de confiança para a diferença entre as duas médias. Isso fornece um intervalo de valores que provavelmente conterá a verdadeira diferença entre as médias da população.

Supondo um teste t, teríamos que:

$$\text{Lower Limit} = M_1 - M_2 - (t_{CL})(S_{M_1 - M_2})$$
$$\text{Upper Limit} = M_1 - M_2 + (t_{CL})(S_{M_1 - M_2})$$

Onde M1 - M2 é a diferença entre as médias amostrais, tCL é o t para o nível de confiança desejado e S_{m1-m2} é o erro padrão estimado da diferença entre as médias amostrais. O erro padrão é o desvio-padrão dividido por raiz de n.

Isso também pode ser escrito como (M1 - M2) ± tCL*S_{m1-m2}

O primeiro passo é calcular a estimativa do erro padrão da diferença entre médias S_{m1-m2}.

Para variâncias populacionais iguais:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Para variâncias populacionais não iguais:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Em ambos os casos, o grau de liberdade é n1+n2-2.

Caso suas amostras sejam dependentes, seu intervalo de confiança será :

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}} \right)$$

Em que:

$$s_d = \sqrt{\frac{\Sigma d^2 - \left[\frac{(\Sigma d)^2}{n}\right]}{n-1}}$$

Onde d é a diferença de cada amostra (antes e depois).

E t* é o tcrítico para g.l = n - 1 com o nível de significância que escolher.

Voltando ao exemplo entre a força de dois materiais, temos:

Amostra A

Média 22.84
Desvio-padrão amostral: 3.08
n: 20

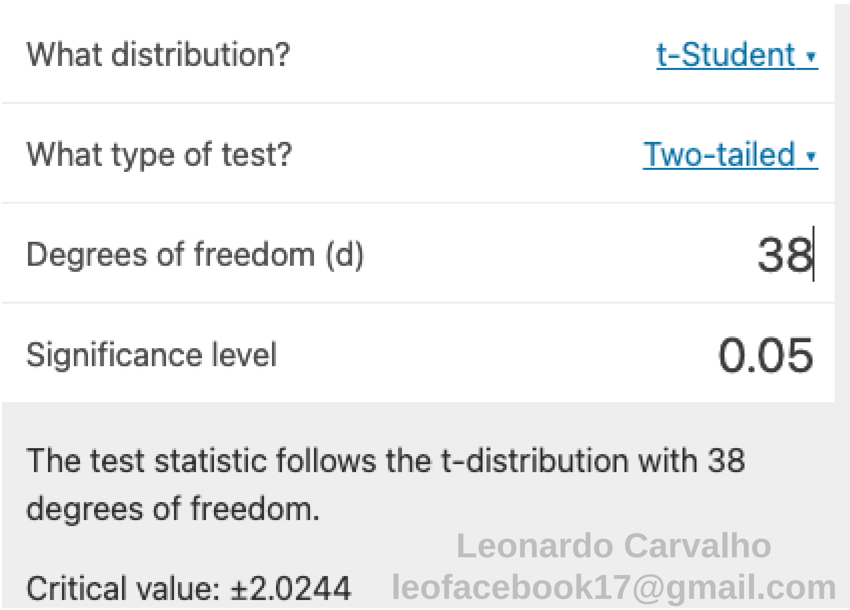
Amostra B

Média 20.69
Desvio-padrão amostral: 3.41
n: 20

Lembrando que são 2 amostras independentes.

g.l = 20+20-2 = 38

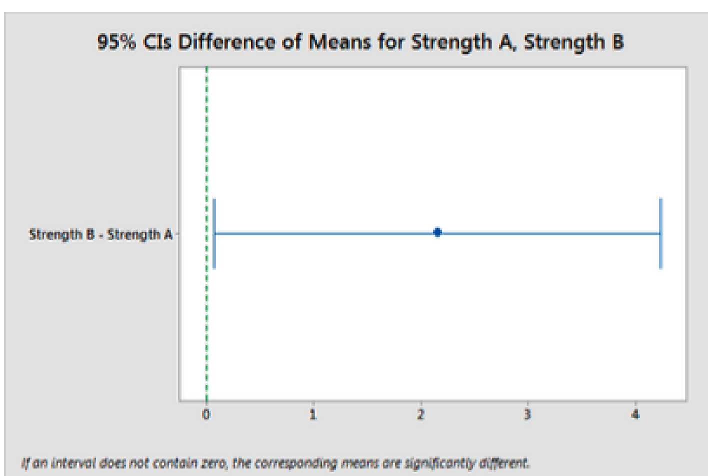
Usando nossa calculadora, temos que tcrítico é 2.0244



Logo, o intervalo de confiança é:

$$(M1\text{-}M2) \pm tc^*\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}} = (22.84\text{-}20.69) \pm \ 2.0244^*\sqrt{\frac{3.08^2}{20} + \frac{3.41^2}{20}} \ = 2{,}15 \pm 2{,}079$$

= [0.071 4.229]



A interpretação continua sendo uma simples avaliação visual. Zero representa nenhuma diferença entre as médias. O intervalo contém zero? S**e não incluir zero, a diferença é estatisticamente significativa** porque o intervalo não exclui nenhuma diferença.

Além de fornecer uma avaliação visual simples, o intervalo de confiança da diferença apresenta informações cruciais que nem os intervalos de confiança do grupo individuais nem o p-valor fornecem. Ele responde à pergunta, com base em nossa amostra, **quão grande é a diferença entre as duas populações?** Como qualquer estimativa, há uma margem de erro em torno da estimativa pontual da diferença. É importante levar em consideração essa margem de erro antes de agir com base nas descobertas.

Para o nosso exemplo, a estimativa pontual da diferença média é de 2,15 e podemos ter 95% de confiança de que a diferença da população está dentro do intervalo de 0,071 a 4,229.

Como em todos os intervalos, a largura do intervalo para a diferença média revela a precisão da estimativa. Intervalos mais estreitos sugerem uma estimativa mais precisa. E você pode avaliar se toda a gama de valores é praticamente significativa.

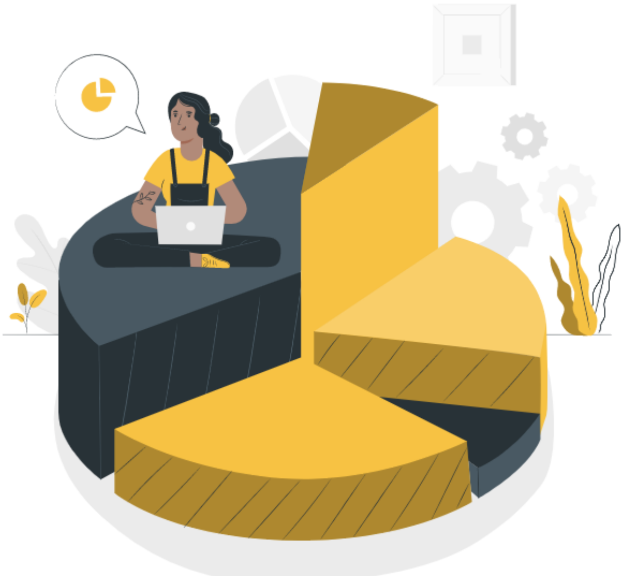
Quando o intervalo é muito amplo (impreciso) para ser útil e/ou o intervalo inclui diferenças que não são significativas na prática, você tem motivos para hesitar antes de tomar decisões com base nos resultados. Esses tipos de resultados de IC indicam que você pode não obter benefícios significativos, mesmo que a diferença seja estatisticamente significativa.

Não existe um método estatístico para responder a perguntas sobre quão precisa uma estimativa deve ser ou quão grande um efeito deve ser para ser útil na prática. Você precisará aplicar seu conhecimento da área de assunto ao intervalo de confiança da diferença para responder a essas perguntas.

Para o exemplo nesta seção, é importante observar que a extremidade inferior do IC está muito próxima de zero. Não será surpreendente se a diferença populacional real cair perto de zero, o que pode não ser significativo na prática, apesar do resultado estatisticamente significativo. Se você está pensando em mudar para o Grupo B para um produto mais forte, a melhoria média pode ser muito pequena para ser significativa.

Ao comparar grupos, avalie os intervalos de confiança dessas diferenças em vez de comparar os intervalos de confiança de cada grupo. Esse método é simples e ainda fornece informações valiosas adicionais.

13. INTERVALO DE CONFIANÇA PARA PROPORÇÃO



A **estimativa pontual para p**, a proporção populacional de sucessos, é dada pela proporção de sucessos em uma amostra e é denotada por:

$$\hat{p} = \frac{x}{n}$$

Em que x é o número de "sucessos" em uma amostra e n é o tamanho da amostra. A estimativa pontual para a proporção populacional de não sucessos é q =1 - p.

A **margem de erro** para esse caso será de:

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Exemplo: Em uma pesquisa com 1.000 adolescentes americanos, 372 disseram que possuem smartphones. 1) Encontre uma **estimativa pontual** para a proporção populacional de adolescentes americanos que **não** possuem smartphones; 2) Construa um intervalo de confiança a 95% para a proporção populacional de adolescentes americanos que possuem smartphones.

Resposta:

1) $p = 372/1000 = 0{,}372$ ou 37,2%

Logo, a proporção de adolescente que **não** possui smartphones é de:

$q = 1-0{,}372 = 0{,}628 = 62{,}8\%$

2) $E = zc*\sqrt{(p*q/n)} = 1.96*\sqrt{(0.628*0.372)/1000} = 0{,}030$

Logo,

Limite inferior = $p - E \approx 0{,}372 - 0{,}030 = 0{,}342$

Limite superior = $p + E \approx 0{,}372 + 0{,}030 = 0{,}402$

Portanto, com 95% de confiança, você pode dizer que a proporção populacional de adolescentes americanos que possuem smartphones está entre 34,2% e 40,2%.

14. TESTE DE HIPÓTESE PARA PROPORÇÃO

1 SAMPLE

Nas seções anteriores você aprendeu como realizar um teste de hipótese para uma média populacional. Nesta seção, aprenderá como testar uma proporção populacional p.

Testes de hipótese para proporções podem ser usados, por exemplo, quando políticos querem saber a proporção de seus eleitores que são a favor de certo projeto de lei, quando engenheiros de qualidade testam a proporção de peças defeituosas, quando queremos saber a proporção de pessoas a mais que compram em um novo site, e assim por diante.

O teste z para uma proporção p é um teste estatístico para uma proporção populacional. O teste z pode ser usado quando uma **distribuição binomial** é dada tal que $n*p \geq 5$ e $n*(1-p) \geq 5$ (condição para assumir normalidade). A estatística de teste é a proporção amostral p e a estatística de teste padronizada é :

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

Em que $\hat{p}$ é a proporção de uma amostra, p seria a proporção de uma população, q é (1-p) e n é o tamanho da amostra. Vamos a um exemplo:

Um pesquisador afirma que menos de 40% dos proprietários de celular nos Estados Unidos usam seus aparelhos para a maioria de suas navegações online. Em uma amostra aleatória de 100 adultos, 31% dizem que usam seus aparelhos para a maioria de suas navegações online. Considerando o nível de significância a = 0,01, há evidência suficiente para concordar com a afirmação do pesquisador?

Resposta:

Os produtos np = 100(0,40) = 40 e nq = 100(0,60) = 60 são ambos maiores que 5. Então, você pode usar um teste z. A afirmação é: “menos de 40% usam seus aparelhos para a maioria de suas navegações online”. Então, as hipóteses nula e alternativa são:

$H_0: p \geq 0{,}4$ e $H_a: p < 0{,}4$ (Afirmação.)

Esse é um teste unilateral à esquerda (sinal de Ha é "< ").

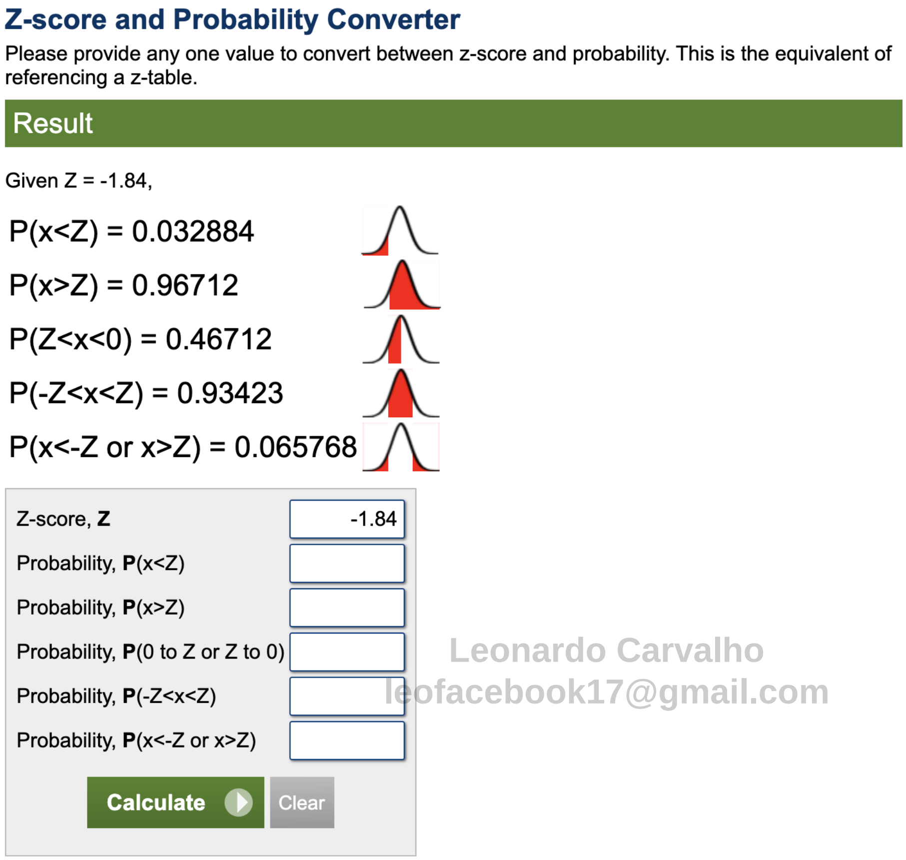
Calculando o z:

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$ Como $np \geq 5$ e $nq \geq 5$, você pode usar o teste z.

$$= \frac{0{,}31 - 0{,}4}{\sqrt{(0{,}4)\,(0{,}6)/100}}$$ Suponha que $p = 0{,}4$.

$$\approx -1{,}84.$$ Arredonde para duas casas decimais.

Vamos usar o site para entender qual seria o p-valor equivalente a esse z. https://www.calculator.net/z-score-calculator.html



Como o teste é unilateral a esqueda, estamos interessados no primeiro P(x<Z) dado, ou seja, p-valor = 0,032884. Lembrando que alpha = 0,01

Como p-valor > alpha, não rejeitamos a hipótese nula. Ou seja, não há evidência suficiente, ao nível de significância de 1%, para concordar com a afirmação de que menos de 40% dos proprietários de telefone celular nos Estados Unidos usam seus aparelhos para a maioria de suas navegações online.

2 SAMPLES

Se uma afirmação é feita sobre dois parâmetros populacionais *p*1 e *p*2, então os possíveis pares de hipóteses nula e alternativa são:

$$\begin{cases} H_0: p_1 = p_2 \\ H_a: p_1 \neq p_2 \end{cases}, \qquad \begin{cases} H_0: p_1 \leq p_2 \\ H_a: p_1 > p_2 \end{cases}, \qquad \text{e} \qquad \begin{cases} H_0: p_1 \geq p_2 \\ H_a: p_1 < p_2 \end{cases}.$$

Independentemente de quais hipóteses você use, para esse teste **sempre assuma em H0 que não há diferença entre as proporções populacionais (*p1* = *p*2).** Isto é, o teste será realizado assumindo que *H*0 é verdade, ou seja, sempre vamos assumir ***p1* = *p*2**.

As condições a seguir são necessárias para usar um teste *z* para testar tal diferença.

1. As amostras são selecionadas aleatoriamente
2. As amostras são independentes
3. As amostras são grandes o suficiente para usar uma distribuição amostral normal.

Z pode então ser calculado por:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Em que $\hat{p}1$ e $\hat{p}2$ são as proporções amostrais, $p1$ - $p2$ é a diferença a ser comprovada. Se a hipótese nula declara $p1$ = $p2$, então $p1$ - $p2$ é igual a 0. Lembrem-se que para esse teste, H0 sempre declarará igualdade, então em qualquer que seja o caso, $p1$ - $p2$ será 0.

Além disso, $\bar{p}$ é dado por

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Em que x1 e x2 é número de sucessos em cada amostra (número de casos da proporção) e n1 e n2 é o tamanho de cada amostra

E e $\bar{q}$ é dado por:

$$\bar{q} = 1 - \bar{p}$$

Exemplo:

Um estudo com 150 proprietários de carros de passageiros e 200 proprietários de caminhonetes, selecionados aleatoriamente, mostra que 86% dos ocupantes de carros de passageiros e 74% dos ocupantes de caminhonetes usam cinto de segurança. Com a = 0,10, você pode rejeitar a afirmação de que a proporção de pessoas que usam cinto de segurança é a mesma para os carros de passageiros e as caminhonetes?

Carros de passageiros	Caminhonetes
$n_1 = 150$	$n_2 = 200$
$\hat{p}_1 = 0{,}86$	$\hat{p}_2 = 0{,}74$
$x_1 = 129$	$x_2 = 148$

Resposta:

Da tabela acima, x1 e x2 é o total de passageiros que dizem usar cinto de segurança (note que **sempre** $n1 * \hat{p}1 = x1$ e $n2 * \hat{p}2 = x2$).

De acordo com as fórmulas acima:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{129 + 148}{150 + 200} = \frac{277}{350} \approx 0{,}7914$$

$$\bar{q} = 1 - \bar{p} \approx 1 - 0{,}7914 = 0{,}2086$$

De acordo com o enunciado:

$$H_0: p_1 = p_2 \quad \text{(afirmação)} \qquad \text{e} \qquad H_a: p_1 \neq p_2$$

Como queremos provar que as diferenças de proporções é 0 na população

$p1$ - $p2$ = 0

Logo,

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \frac{(0{,}86 - 0{,}74) - 0}{\sqrt{(0{,}7914)(0{,}2086)\left(\frac{1}{150} + \frac{1}{200}\right)}} \approx 2{,}73.$$

Usando o site https://www.calculator.net/z-score-calculator.html temos que:

Z-score and Probability Converter

Please provide any one value to convert between z-score and probability. This is the equivalent of referencing a z-table.

Result

Given Z = 2.73,

P(x<Z) = 0.99683

P(x>Z) = 0.0031667

P(0<x<Z) = 0.49683

P(-Z<x<Z) = 0.99367

P(x<-Z or x>Z) = 0.0063334

Z-score, **Z**	2.73
Probability, **P**(x<Z)	
Probability, **P**(x>Z)	
Probability, **P**(0 to Z or Z to 0)	
Probability, **P**(-Z<x<Z)	
Probability, **P**(x<-Z or x>Z)	

Calculate Clear

Como o teste é bicaudal, p-valor é nossa última imagem, ou seja, 0.0063334.

Como p-valor é menor que alpha, rejeitamos a hipótese nula. Ou seja, há evidência suficiente, ao nível de significância de 10%, para rejeitar a afirmação de que a proporção de pessoas que usam cinto de segurança é a mesma para carros de passeio e caminhonetes.

INTERVALO DE CONFIANÇA PARA A DIFERENÇA DE DUAS PROPORÇÕES

Da mesma forma que para a média, também podemos calcular o intervalo de confiança para a diferença de proporções. Nesse caso temos:

$$(\text{sample difference}) \pm (\text{critical value}) \begin{pmatrix} \text{standard error} \\ \text{of difference} \end{pmatrix}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}}$$

REFERÊNCIAS BIBLIOGRÁFICAS

Bussan, W., Morettin, P. - Estatística Básica - Editora Saraiva - 2010 - 6th ed

Frost, J. - Hypothesis testing - An intuitive guide for making data driven decisions - Jim Frost - 2020 - 1st ed

Huyen, C. - Designing machine learning systems - Editora O'Reilly - 2022 - 1st ed

Knaflic, C.N - Storytelling com dados: Um guia sobre visualização - Editora Alta Books - 2019

Larson, R., Farber B. - Estatística Aplicada - Editora Pearson - 2016 - 6th ed

Pinheiro, J., Cunha, S., S. Santiago, Gomes, G. - Probabilidade e Estatística: Quantificando a incerteza - Elsevier Editora Ltda - 2012