

Intervalos de Confiança - Amostras Pequenas

Prof. Eduardo Bezerra

CEFET/RJ

20 de Abril de 2018

- 1 Distribuição t de Student
- 2 Funções relevantes no R
- 3 Intervalos de confiança

1 Distribuição t de Student

2 Funções relevantes no R

3 Intervalos de confiança

Distribuição t de Student

- Quando o tamanho da amostra é pequeno ($n < 30$), não há bons métodos gerais para encontrar intervalos de confiança, nem para médias nem para proporções.
- No entanto, quando a população é aproximadamente normal, a **distribuição t de Student** pode ser usada para calcular os intervalos de confiança para uma média populacional.
- Vamos descrever essa distribuição e mostrar como usá-la.

Distribuição t

Caso o tamanho da amostra seja menor que 30 e a população seja aproximadamente normal, o intervalo de confiança deve ser calculado com o uso da estatística T :

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

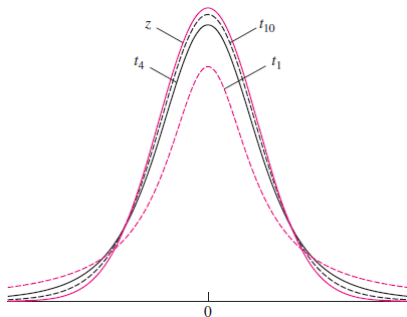
em que s é o estimador do desvio padrão populacional σ . Essa estatística tem distribuição t de Student, com $(n - 1)$ **graus de liberdade**.

Características da distribuição t de Student:

- a) é simétrica em relação ao 0;
- b) é semelhante à distribuição normal padrão, porém, apresenta maior espalhamento do que a normal padrão;
- c) quando $n > 30$, a distribuição t tende para a normal padrão;
- d) as curvas de t estão mais espalhadas do que a da normal padrão, mas o espalhamento diminui à medida que o número de graus de liberdade aumenta.

Distribuição t

Representação gráfica da fdp da curva t de Student para vários graus de liberdade. A curva normal padrão é também desenhada para comparação.



1 Distribuição t de Student

2 Funções relevantes no R

3 Intervalos de confiança

Função qt

$qt(.95,30)$ retornará 1.69, que é o valor do percentil 95 desta distribuição.

$qt(.95,30)$ retornará 1.69, que é o valor do percentil 95 desta distribuição.

Esse valor significa que

- 95% de todos os números em nossa distribuição é inferior a 1,69, e
- apenas 5% é maior.

Isso é inverso da CDF.

Função pt

A chamada $pt(1.69,30)$, retornará um resultado próximo de 95%. Esta função retorna a CDF, que é a probabilidade de obter um número menor ou igual ao argumento. Já que 1,69 corresponde ao 95^o percentil, o valor da CDF é de 95%.

Função dt

$dt(x, 30)$ produzirá o valor da função de densidade de probabilidade em x .

Função dt

$dt(x, 30)$ produzirá o valor da função de densidade de probabilidade em x .

Para 1,69, o valor é 0,096, que é bastante baixo, enquanto que para 0 é 50%.

Exercício 01

Encontre os percentis 2.5 e 97.5 da distribuição Student t com 5 graus de liberdade.

Exercício 01

Encontre os percentis 2.5 e 97.5 da distribuição Student t com 5 graus de liberdade.

Solução:

$qt(c(.025, .975), df=5)$

Exercício 02

Uma amostra aleatória de tamanho 10 é colhida a partir de uma população com distribuição normal com média 4. A estatística t de Student $t = (\bar{X} - 4)/(s/10)$ deve ser calculada. Qual é a probabilidade de que $t > 1,833$?

Exercício 02 - solução

Essa estatística t tem $10 - 1 = 9$ graus de liberdade. A partir da tabela t , $\Pr(t > 1,833) = 0,05$.



Exercício 03

Encontre o valor para a distribuição t_{12} cuja probabilidade da parte superior da cauda é 0,025.

Exercício 03 - solução

Basta procurar de cima para baixo na coluna intitulada 0,025 para a linha correspondente a 12 graus de liberdade. O valor para t_{12} é então 2,179.

Exercício 04

Determine o valor para a distribuição t_{14} cuja probabilidade da cauda inferior é de 0,01.

Exercício 04 - solução

Basta procurar de baixo para cima na coluna intitulada 0,01 pela linha correspondente a 14 graus de liberdade. O valor para t_{14} é 2,624. Este valor corta uma área, ou a probabilidade, de 1% na cauda superior. O valor cuja probabilidade da cauda inferior é de 1% é portanto -2,624.

- 1 Distribuição t de Student
- 2 Funções relevantes no R
- 3 Intervalos de confiança**

Intervalos de confiança usando a distribuição t

Quando o tamanho da amostra é grande, não importa a distribuição da população para a determinação de uma estimativa para a média populacional,

- Isso porque o TLC garante que \bar{X} será aproximadamente normalmente distribuída.

Porém, quando a amostra é pequena ($n \leq 30$), a distribuição da população deve ser aproximadamente normal para que se possa fazer estimativas adequadas.

- Nesse caso, intervalos de confiança são construídos de forma semelhante ao caso em que $n > 30$.
- A diferença é que o z-score é substituído por um valor da distribuição t de Student.

Procedimento geral

O procedimento geral para a determinação de intervalos de confiança para \bar{X} , para amostras pequenas, é resumido a seguir.

Procedimento geral

Seja X_1, X_2, \dots, X_n uma amostra aleatória pequena colhida de uma população normal cuja média é μ . Então um intervalo de confiança no nível $100(1 - \alpha)\%$ para μ é

$$\bar{X} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

em que $t_{n-1, \alpha/2}$ é o valor que corta uma área de $\alpha/2$ na cauda do lado direito da distribuição t.

Exercício 05

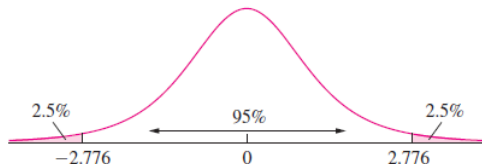
Um técnico em metalurgia está estudando um novo processo de soldagem. Ele fabrica 5 juntas soldadas usando esse processo e mede a resistência à deformação de cada uma. Os cinco valores (em ksi) são

56,3 65,4 58,7 70,1 63,9

Suponha que esses valores são uma amostra aleatória de uma população aproximadamente normal. Determine um intervalo de confiança no nível de 95% para a resistência média das soldaduras forjadas nesse processo.

Exercício 05 - solução

A figura a seguir mostra a distribuição t_4 . Por outro lado, 95% da área sob a curva está contida no intervalo $[-2,776; 2,776]$.



Segue-se que, para 95% de todas as amostras que poderiam ter sido escolhidas,

$$-2,776 < \frac{\bar{X} - \mu}{s/\sqrt{n}} < 2,776$$

Exercício 05 - solução (cont)

Ao manipular essa desigualdade, acabamos por obter

$$\bar{X} - 2,776 \frac{s}{\sqrt{n}} < \mu < \bar{X} + 2,776 \frac{s}{\sqrt{n}}$$

A média e o desvio padrão da amostra são, respectivamente, $\bar{X} = 62,88$ e $s = 5,4838$. O tamanho da amostra é $n = 5$. Ao fazermos a substituição de valores, descobrimos que um intervalo de confiança de 95% para μ é

$$62,88 - 6,81 < \mu < 62,88 + 6,81$$

ou $(56,07; 69,69)$.

Exercício 06

Sabendo-se que uma amostra tem 25 elementos, que a sua média 150 e desvio padrão igual a 10. Represente um intervalo de confiança em nível de 90%.

Exercício 06 - solução

- Como a amostra é menor que 30 elementos, então iremos usar a distribuição t de Student. Se desejamos um intervalo de confiança de 90%, temos:
- Para trabalharmos com a tabela, encontramos o número de graus de liberdade, que é $(n - 1)$. Logo, $(25 - 1) = 24$.
- O nível de confiança desejado é $(1 - \alpha) = 1 - 0,1 = 0,9 = 0,9$.
- Conhecendo o número de graus de liberdade e o nível de confiança desejado vamos a tabela e encontramos o valor t , neste caso igual a 1,7109.

$$\bar{X} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$