Distribuição T - Student

Prof. Herondino S. F.

Distribuição T-Student

- A **distribuição *T* de Student** é uma distribuição de probabilidade estatística, publicada por um autor que se chamou de *Student*, pseudônimo de William Sealy Gosset, que não podia usar seu nome verdadeiro para publicar trabalhos enquanto trabalhasse para a cervejaria Guinness.

T- Student

- Padronizar variável aleatória normal requer que o μ e σ sejam conhecidos. Na prática, porém, não podemos calcular $z = (x - \mu)/\sigma$ porque σ é desconhecido. Em vez disso, substituímos σ por s e calculamos a estatística t.

$$t = \frac{x - \mu}{s}$$

Distribuição Amostral da Média e da variância

- Se discrepâncias nas observações sobre a média são aleatórios e independentes, então a distribuição amostral da média tem μ e variância, σ^2/n.
- A quantidade σ^2/n é a variância da média.
- Sua raiz quadrada é chamada o ***erro padrão da média***:

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

- A estimativa do **erro padrão da média** é:

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

Distribuição t

- Normalmente, a variância da população, σ^2 não é conhecida e não podemos usar a distribuição normal como a distribuição de referência para a média da amostra. Em vez disso, substituir e usar a distribuição *t*.
- Se a distribuição de referência é normal e a variância da população é estimado por s^2, a quantidade:

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

- que é conhecido como a média padronizada ou como a estatística t, terá à distribuição com $v = n - 1$ graus de liberdade.

Exemplo:

- Para os dados de nitrato, a média da amostra de concentração é igual a 7,51 mg/L e encontra-se a uma distância considerável abaixo do verdadeiro valor de referência 8,00 mg/L (Figura 2). Se a verdadeira média da amostra é de 8,0 mg/L e o laboratório está medindo precisamente, um valor tão baixo quanto 7,51 que ocorrem por acaso apenas quatro vezes em 100. Sabe-se que o desvio padrão é 1,38 e 27 amostras. Qual será o valor T.

"Qual a probabilidade de se obter uma amostra tão pequenas com média = 7,51 mg/L a partir da análise das 27 amostras?"

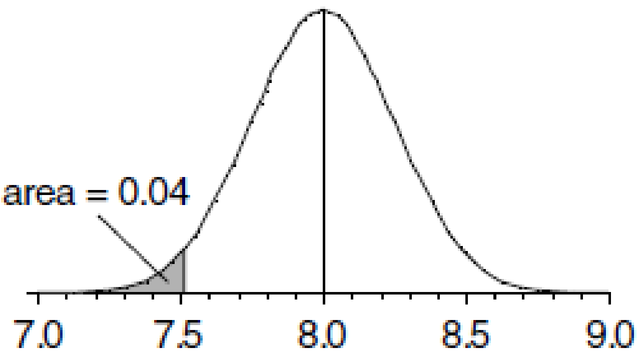
$$T = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

$$t = \frac{7{,}51 - 8}{1{,}38 / \sqrt{27}}$$

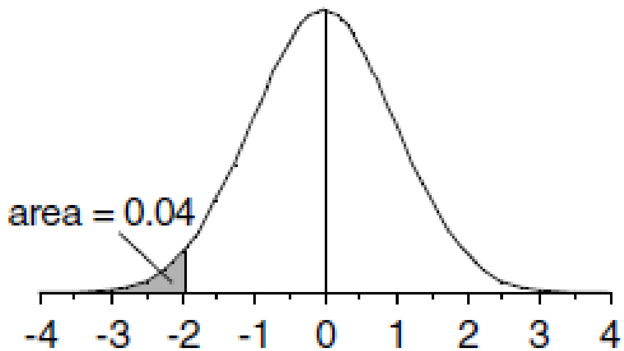
$$t = \frac{-0{,}49}{1{,}38 / 5{,}19}$$

$$t = \frac{-0{,}49}{0{,}2658} = -1{,}842$$

Utilizando a tabela



a) Referência de distribuição de $\overline{X}$
P($\overline{X}$ ≤ 7,51) = 0.04



b) Referência de distribuição T
P(t ≤ 1,853) = 0,04

n	$\alpha = 0.1$	0.05	0.025	0.01	0.005
2	1.886	2.920	4.303	6.965	9.925
4	1.533	2.132	2.776	3.747	4.604
6	1.440	1.943	2.447	3.143	3.707
10	1.372	1.812	2.228	2.764	3.169
20	1.325	1.725	2.086	2.528	2.845
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
40	1.303	1.684	2.021	2.423	2.704
∞	1.282	1.645	1.960	2.326	2.576

Grau de liberdade v = n - 1
Como são 27 amostras, temos:
v=27-1=26 grau de liberdade

$$t = -1{,}842$$

Análise

- Se este resultado é altamente improvável, pode ser que a amostra não representam a população, provavelmente porque o processo de medição foi tendenciosa para produzir concentrações abaixo do valor real.
- Ou, poderíamos decidir que o resultado, embora improvável, deve ser aceito como ocorrido devido ao acaso e não devido uma causa atribuível (como viés nas medições).
- Inferência estatística envolve fazer uma avaliação a partir de dados experimentais sobre um parâmetro desconhecido da população (por exemplo, uma média ou variância).

A Distribuição T

- A distribuição de referência é necessária, a fim de escolher se o resultado é facilmente explicada por mero acaso ou se é variação excepcional.
- A *distribuição T* é uma relevante referência que representa o conjunto de resultados que poderiam ocorrer por acaso.
- Um resultado que cai sobre a cauda da distribuição pode ser considerado excepcional.

T- Student

- A distribuição T é similar a distribuição Z, em que ambos são simétricas na média zero.
- Ambas as distribuições são em forma de sino, mas a T distribuição é mais variável em virtude dos T - valores depender das flutuações de duas quantidades, $\overline{X}$ e S^2, considerando que os valores-Z depende apenas das mudanças na $\overline{X}$ de amostra para amostra.

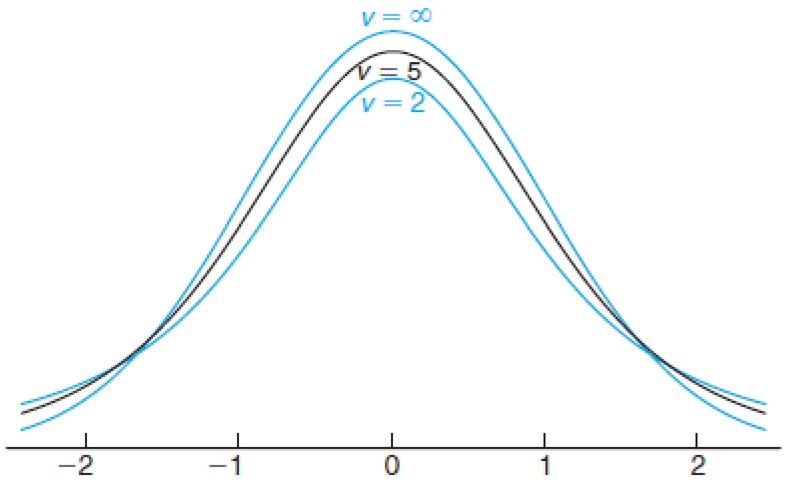


Figura 1 - A distribuição t para curvas v=2, v=5 e v=∞.

Na Figura 1, mostramos a relação entre a distribuição normal padrão (v = ∞) e distribuições *t* com 2 e 5 graus de liberdade

Graus de liberdade

- A distribuição de T diferente daquela de Z na variação de T depende do tamanho da amostra n e é sempre maior do que 1.
- Somente quando o tamanho da amostra $n \to \infty$ as duas distribuições se tornará o mesmo.
- A porcentagem da distribuição t é dada por Tabelas.

Propriedades

- As condições em que a quantidade $t = (x - \mu)/s$ tem a distribuição com graus de liberdade v são:

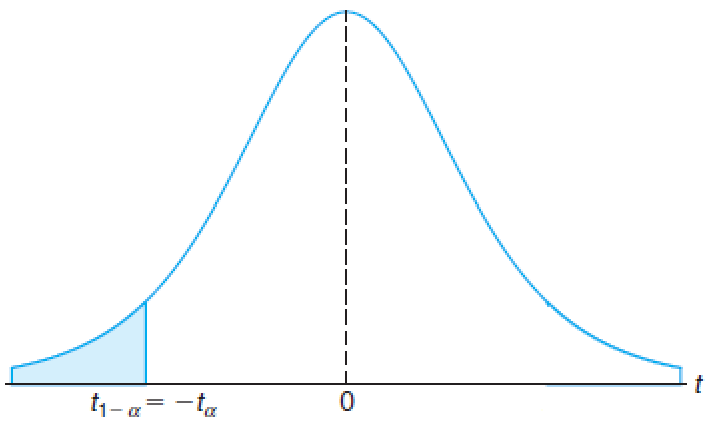
- 1) x é normalmente distribuído sobre μ com variância σ^2 ;

- 2) s é distribuído Independentemente da média, isto é, a variância da amostra não aumenta ou diminui à medida que aumenta ou diminui as médias;

- 3) O s^2 quantidade, que tem v graus de liberdade, é calculada a partir de observações normalmente distribuídos e têm variância independentemente σ^2.

Exemplo 2:

- O t-value com v = 14 graus de liberdade que deixa uma área de 0,025 para a esquerda, e, portanto, uma área de 0.975 para a direita, é:



$$t_{0,975} = -t_{0,025} = -2.145.$$

Exemplo 3:

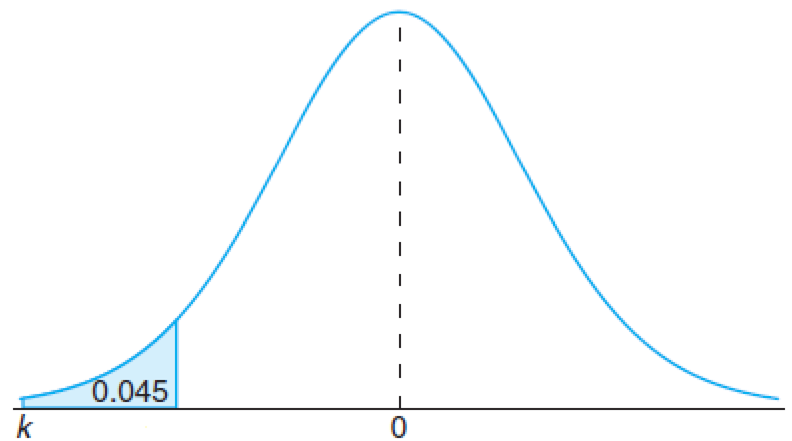
- Encontre $P(-t_{0,025} < T < t_{0,05})$.
- *Como* $t_{0,05}$ deixa uma área de 0,05 para a direita, e $-t_{0,025}$ deixa uma área de 0,025 à esquerda, encontramos uma área total de $1 - 0,05 - 0.025 = 0.925$ entre $-t_{0,025}$ e $t_{0,05}$.
- *Portanto,* $P(-t_{0,025} < T < t_{0,05}) = 0.925$.

Exemplo 4:

- Encontre k sendo $P(k < T < -1.761) = 0.045$ *de uma amostra aleatória de tamanho 15 selecionado de uma distribuição normal e* $\dfrac{\overline{X}-\mu}{s/\sqrt{n}}$

Tabela 1- Distribuição t

v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131



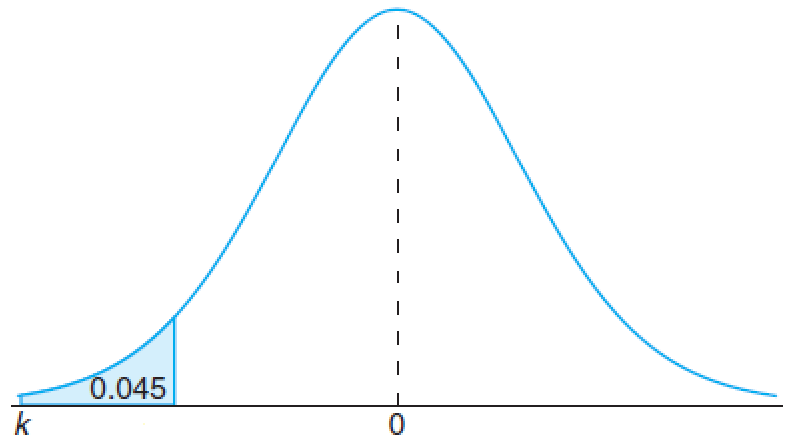
Os valores t do exemplo 4

Exemplo 4:

- Encontre k sendo *P(k < T < −1.761) = 0.045 de uma amostra aleatória de tamanho 15 selecionado de uma distribuição normal e* $\dfrac{\overline{X}-\mu}{s/\sqrt{n}}$

Tabela 1- Distribuição t

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131



Os valores t do exemplo 4

Como $k=-t_{\alpha}$ *então*

$0{,}05-\alpha=0{,}045$

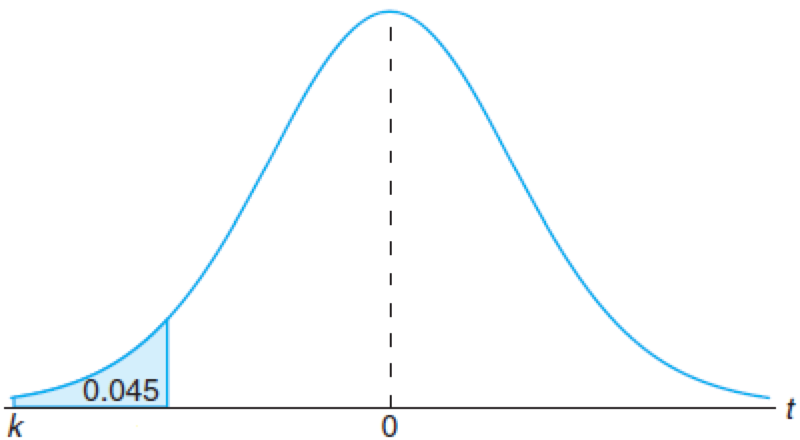
$\alpha=0{,}005$

Exemplo 4:

- Encontre k sendo *P(k < T < −1.761) = 0.045 de uma amostra aleatória de tamanho 15 selecionado de uma distribuição normal e* $\frac{\overline{X}-\mu}{s/\sqrt{n}}$

Tabela 2- Distribuição t

v	α 0.02	0.015	0.01	0.0075	0.005	0.0025
1	15.894	21.205	31.821	42.433	63.656	127.321
2	4.849	5.643	6.965	8.073	9.925	14.089
3	3.482	3.896	4.541	5.047	5.841	7.453
4	2.999	3.298	3.747	4.088	4.604	5.598
5	2.757	3.003	3.365	3.634	4.032	4.773
6	2.612	2.829	3.143	3.372	3.707	4.317
7	2.517	2.715	2.998	3.203	3.499	4.029
8	2.449	2.634	2.896	3.085	3.355	3.833
9	2.398	2.574	2.821	2.998	3.250	3.690
10	2.359	2.527	2.764	2.932	3.169	3.581
11	2.328	2.491	2.718	2.879	3.106	3.497
12	2.303	2.461	2.681	2.836	3.055	3.428
13	2.282	2.436	2.650	2.801	3.012	3.372
14	2.264	2.415	2.624	2.771	2.977	3.326
15	2.249	2.397	2.602	2.746	2.947	3.286



Os valores t do exemplo 4

Como $k=-t_{\alpha}$ *então*

$0{,}045=0{,}05-\alpha$

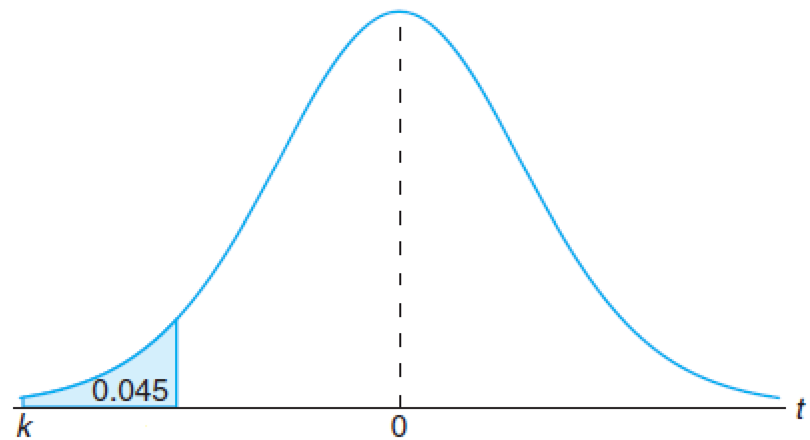
$\alpha=0{,}005$

Exemplo 4:

- Encontre k sendo $P(k < T < -1.761) = 0.045$ *de uma amostra aleatória de tamanho 15 selecionado de uma distribuição normal e* $\dfrac{\bar{X}-\mu}{s/\sqrt{n}}$

Tabela 2- Distribuição t

v	α 0.02	0.015	0.01	0.0075	0.005	0.0025
1	15.894	21.205	31.821	42.433	63.656	127.321
2	4.849	5.643	6.965	8.073	9.925	14.089
3	3.482	3.896	4.541	5.047	5.841	7.453
4	2.999	3.298	3.747	4.088	4.604	5.598
5	2.757	3.003	3.365	3.634	4.032	4.773
6	2.612	2.829	3.143	3.372	3.707	4.317
7	2.517	2.715	2.998	3.203	3.499	4.029
8	2.449	2.634	2.896	3.085	3.355	3.833
9	2.398	2.574	2.821	2.998	3.250	3.690
10	2.359	2.527	2.764	2.932	3.169	3.581
11	2.328	2.491	2.718	2.879	3.106	3.497
12	2.303	2.461	2.681	2.836	3.055	3.428
13	2.282	2.436	2.650	2.801	3.012	3.372
14	2.264	2.415	2.624	2.771	2.977	3.326
15	2.249	2.397	2.602	2.746	2.947	3.286



Os valores t do exemplo 4

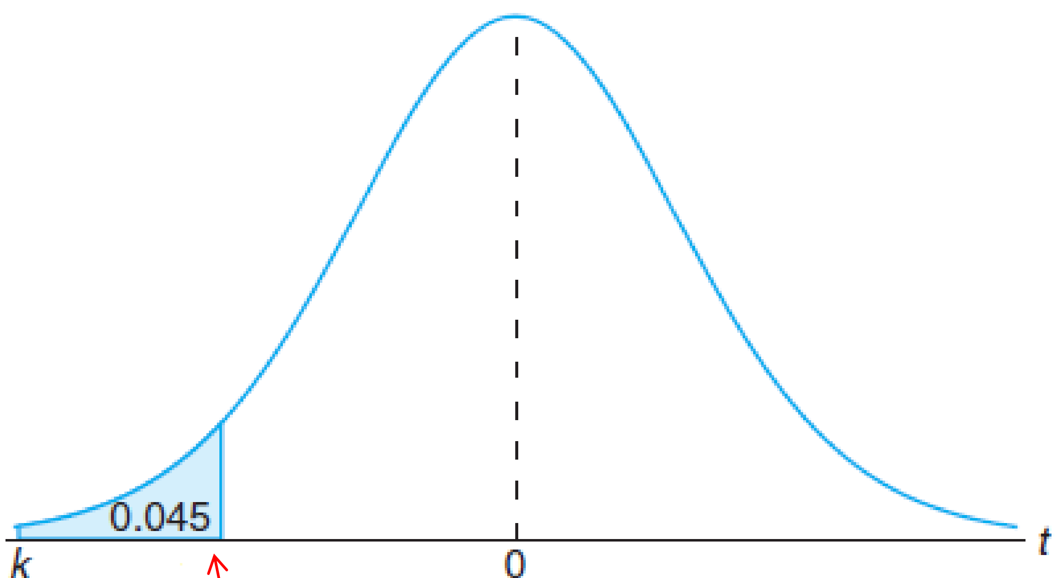
Como $k=-t_{\alpha}$ *então*

$0{,}045=0{,}05-\alpha$

$\alpha=0{,}005$

$-t_{0,005}=-2{,}977$

Exemplo 4:



$P(-2{,}977 < T < -1.761) = 0.045$

Distribuição t - Análise

- Exactamente 95% dos valores de uma distribuição-t com *$v = n\text{-}1$* graus de liberdade situar-se entre $-t_{0,025}$ e $t_{0,025}$. Claro, existem outras t-valores que contêm 95% da distribuição, como por exemplo $t_{0,03}$ e $-t_{0,02}$, mas esses valores não aparecem na Tabela t, e, além disso, o intervalo mais curto possível é obtido pela escolha t-valores isso deixa exatamente a mesma área nas duas caudas da nossa distribuição.

Distribuição t - Análise

- Esse t-valor que está abaixo de $-t_{0,025}$ ou acima $t_{0,025}$ os faria acreditar que qualquer um evento muito raro ocorreu ou a nossa suposição sobre μ está em erro.
- Caso isto aconteça, vamos tomar a decisão do pressuposto de que o nosso valor de μ é um erro
- Na verdade, uma t-valor cair abaixo $-t_{0,01}$ ou acima $t_{0,01}$ daria evidência ainda mais forte que o nosso valor assumido de μ ser bastante improvável.

Exemplo

- Um engenheiro químico afirma que a média da população de rendimento de um processo em lote é de 500 gramas por mililitro de matéria-prima. Para verificar essa afirmação tem amostras de 25 lotes de cada mês. Se o t-valor calculado cai entre $-t_{0,05}$ e $t_{0,05}$, dar-se por satisfeito com esta reivindicação.
- Qual conclusão poderia encontrar a partir de uma amostra que tem média de $\bar{x} = 518$ gramas por mililitro e um desvio padrão amostral de $s = 40$ gramas? Assumir a distribuição do rendimento aproximadamente normal.

Resolução

- Da tabela encontramos que $t_{0,05}$ = *1.711 para 24 graus de liberdade.*
- PORTANTO, o engenheiro pode estar satisfeito com a sua afirmação se uma amostra de 25 lotes rende um valor de t entre -1,711 e 1,711.
- Se μ= *500, então:*

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2{,}25$$

- A probabilidade de se obter um valor de t, com v = 24, igual a ou maior do que 2,25 é de aproximadamente 0,02. Se $\mu > 500$, o valor de T calculado a partir da amostra é mais razoável

Tabela 1- Distribuição t

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Tabela 2- Distribuição t

v	α 0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290

Referência Bibliográfica

- BERTHOUEX, Paul Mac; BROWN, Linfield C.. **Statistics for Environmental Engineers.** 2ª Boca Raton London New York Washington, D.c: Lewis Publishers, 2002. 10-13 p.
- Walpole, Ronald E *et al*. **Probability & statistics for engineers & scientists/Ronald E. Walpole . . . *[et al.]*** — *9th. Ed.* ISBN 978-0-321-62911-1.*Boston-USA/2011*
- TRIOLA, Mario F. et al. **Introdução a Estatística.** 7ª Edição.Livros Técnicos e Científicos Editora S.A. Rio de Janeiro,1998.