7) t-distribution

Vitor Kamada

June 2018

Reference

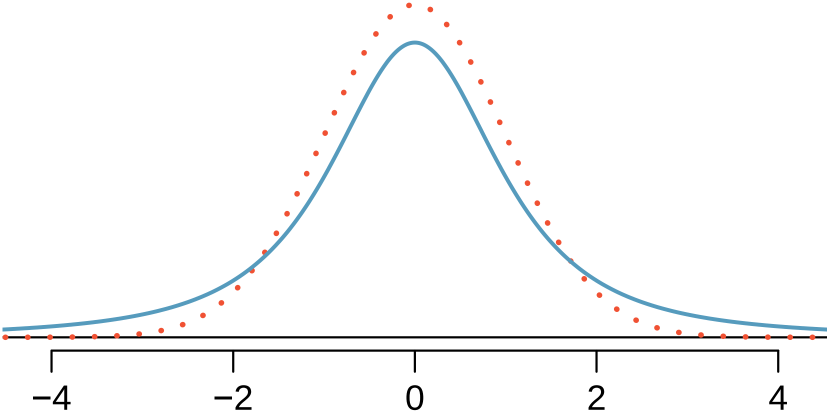
Tables, Graphics, and Figures from

Introductory Statistics with

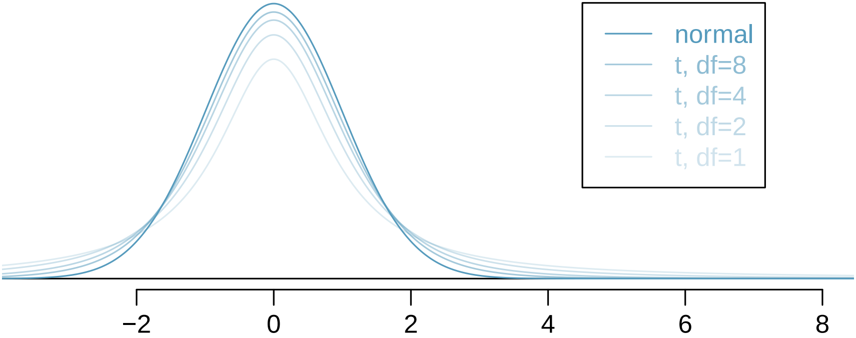
Randomization and Simulation

Diez et al. (2014): Chapter 4 - Inference for Numerical Data

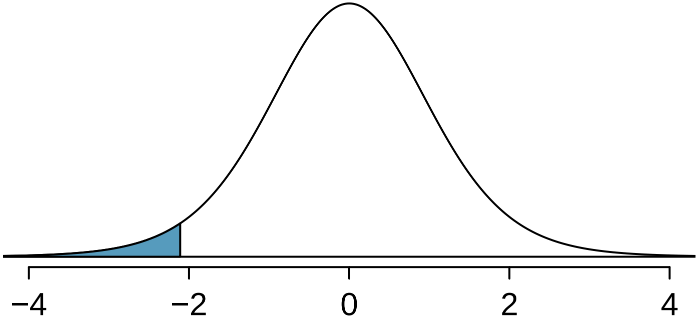
t Distribution (Blue Solid Line) vs Normal Distribution (Red Dotted Line)



t Distribution and Degrees of Freedom (df)



What is the area below $t_{18} = -2.10$?



import scipy.stats

scipy.stats.t.cdf(-2.1,18)

$$\text{Blue Area} = 2.5\%$$

t Table

scipy.stats.t.ppf(0.025,18)

one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
400	1.28	1.65	1.97	2.34	2.59
500	1.28	1.65	1.96	2.33	2.59
∞	1.28	1.65	1.96	2.33	2.58

Confidence Interval for μ

n	$\bar{x}$	s	minimum	maximum
19	4.4	2.3	1.7	9.2

$$\bar{x} \pm t_{df} \times \frac{s}{\sqrt{n}}$$

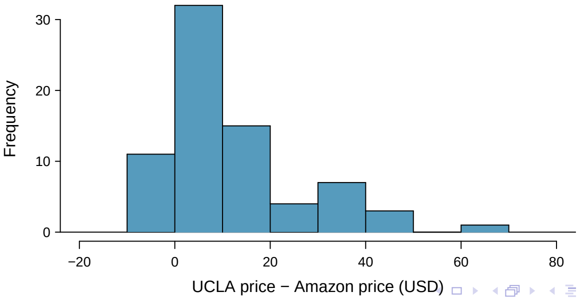
$$\bar{x} \pm t_{18} \times \frac{2.3}{\sqrt{19}}$$

$$4.4 \pm 2.10 \times 0.528 = [3.29, 5.51]$$

We are 95% confident the average mercury content of muscles in Risso's dolphins is between 3.29 and 5.51 g/wet gram

Paired Data

	dept	course	ucla	amazon	diff
1	Am Ind	C170	27.67	27.95	-0.28
2	Anthro	9	40.59	31.14	9.45
3	Anthro	135T	31.68	32.00	-0.32
4	Anthro	191HB	16.00	11.52	4.48
⋮	⋮	⋮	⋮	⋮	⋮
72	Wom Std	M144	23.76	18.72	5.04
73	Wom Std	285	27.70	18.22	9.48



Inference for Paired Data

$$H_o : \mu_{diff} = 0 \text{ vs } H_A : \mu_{diff} \neq 0$$

n_{diff}	$\bar{x}_{diff}$	s_{diff}
73	12.76	14.26

$$SE_{\bar{x}_{diff}} = \frac{s_{diff}}{\sqrt{n_{diff}}} = \frac{14.26}{\sqrt{73}} = 1.67$$

$$t = \frac{(\bar{x}_{diff} - 0)}{SE_{\bar{x}_{diff}}} = \frac{12.76 - 0}{1.67} = 7.59$$

p-value of $t > 7.50$ is .00001

Does treatment using embryonic stem cells (ESCs) help improve heart function following a heart attack?

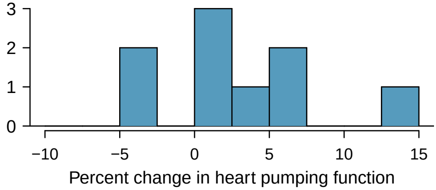
	n	$\bar{x}$	s
ESCs	9	3.50	5.17
control	9	-4.33	2.76

$$\bar{x}_{esc} - \bar{x}_{control} = 3.50 - (-4.33) = 7.83$$

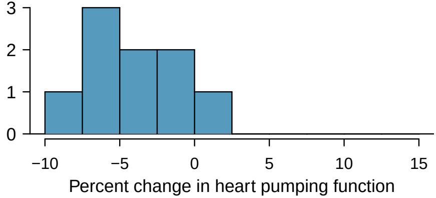
Higher values are associated with greater improvement

Histograms

Embryonic stem cell transplant



Control (no treatment)



95% Confidence Interval for $\mu_1 - \mu_2$

$$SE_{\mu_1-\mu_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cong \sqrt{\frac{5.17^2}{9} + \frac{2.76^2}{9}} = 1.95$$

$$\bar{x}_{esc} - \bar{x}_{control} \pm t^* \times SE_{esc-control}$$

$$7.83 \pm 2.31 \times 1.95 \rightarrow [3.38, 12.38]$$

*To calculate the df, use software or the smaller of $n_1 - 1$ and $n_2 - 1$

Absorption of Phosphorus by Rumex Acetosa

import pandas as pd

df =
pd.read_table('http://www.stat.umn.edu//~gary//book//fcdae.data//e
header=10, delim_whitespace=True)

15 Days				28 Days			
4.3	4.6	4.8	5.4	5.3	5.7	6.0	6.3

Two-Sample t-Test

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_A : \mu_1 < \mu_2$$

$$s_p = \sqrt{\frac{\sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^2+\sum_{i=1}^{n_2}(y_{2i}-\bar{y}_2)^2}{n_1+n_2-2}}$$

$$t = \frac{\bar{y}_1-\bar{y}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{4.775-5.825}{.446\sqrt{\frac{1}{4}+\frac{1}{4}}}$$

t = -3.3273, df = 6, p-value = 0.00793

Two-Sample t-Test (Equal Variance)

a = df[df['days'] == 15]

b = df[df['days'] == 28]

TwoTail=scipy.stats.ttest_ind(a['y'],b['y'], equal_var=True)

```
Ttest_indResult(statistic=-3.3273307180250296, pvalue=0.015859198
```

OneTail = TwoTail.pvalue/2

```
0.007929599172545378
```