

# EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

# Overview

- The t-Distribution
  - ▣ Historical Moment: William Sealy Gosset
  - ▣ Characteristics
- z vs. t tables
  - ▣ Probability and Cut-Off Values
- The t-table
  - ▣ Level of Significance
  - ▣ Degrees of Freedom
  - ▣ One or Two tailed
- Confidence Intervals with t-distribution
- Critical t and Confidence Intervals in R

# Why do we need distributions?

- Distributions help us determine the probability of observing a sample mean of a certain value.
- Why does this matter?
  - ▣ It is not uncommon to observe one person that is very tall, or one building high lead content pipes. This may not be ground breaking news or require attention. But what if we saw an entire city or state of very tall people or high lead content? This is much more unlikely and may be worthy of further investigation...

But HOW unlikely does something have to be for us to conclude there IS something going on here?

# How Unlikely is Unlikely?

- A fairly common convention in statistics is to test how likely something is... If the probability of seeing, for example, this high lead content in water is less than 5%, we tend to call this “statistically significant.”
- There are three main levels of significance:
  - .10, which is equivalent to a less than 10% chance
    - p value < .10
  - .05, which is equivalent to a less than 5% chance
    - p value < .05
    - Most common.
  - .01, which is equivalent to a less than 1% chance
    - p value < .01

# How Unlikely is “Unlikely”?

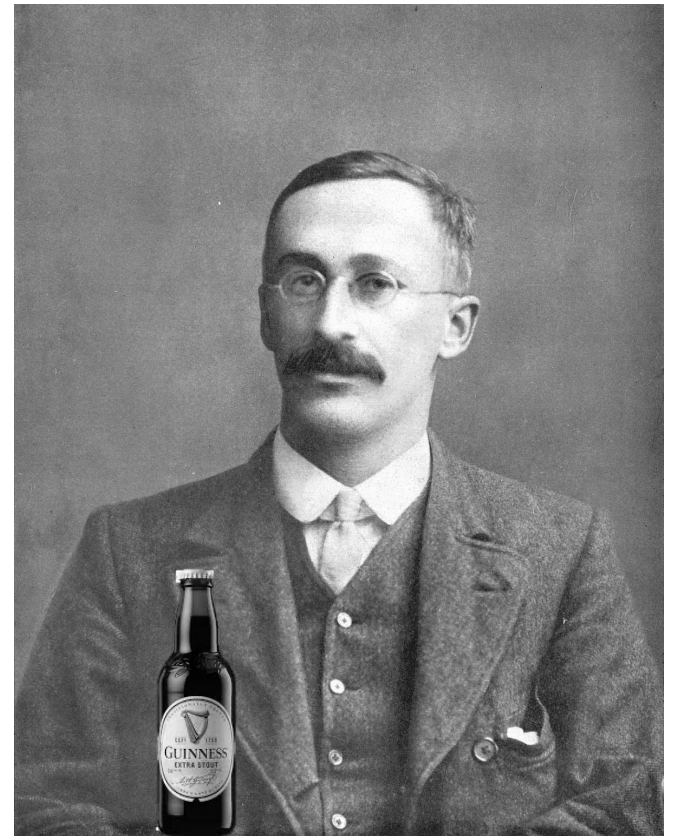
- When we were working with the z-distribution, the z-values associated with those chances are known and we can look them up on a z-table:
- 90% confidence has  $z = 1.65$ 
  - ▣ p value = .10
- 95% confidence has  $z = 1.96$ 
  - ▣ p value = .05
- 99% confidence has  $z = 2.58$ 
  - ▣ p value = .01

But now, we need to be a bit more realistic... rarely do we know  $\sigma$  and sample sizes can be small sometimes. We need a different distribution now...

# t-Distribution

# Historical Moment: William Sealy Gosset

- **William Sealy Gosset** (June 13, 1876 – October 16 1937)
- (AKA “Student”)
- A statistician, chemist, and Head Brewer for Guinness



# Historical Moment: William Sealy Gosset

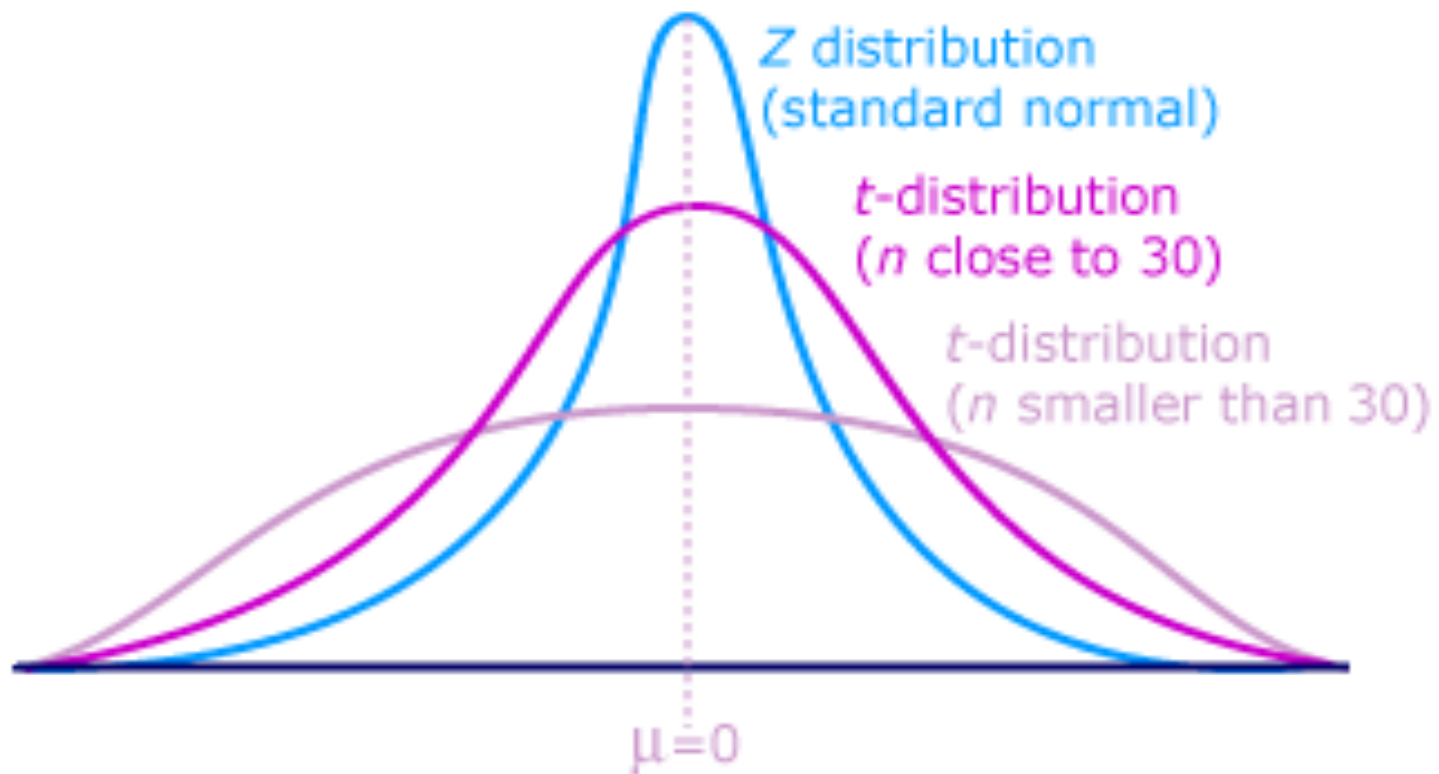
- He had to check the sugar content in the malt used to make the beer. It had to be within a certain range because sugar is related to alcohol content... And he didn't want beer batches to vary too much.
- But... He frequently had to use small sample sizes which has more uncertainty associated with it, and he didn't know the true  $\sigma$ . So, he created a new distribution, very similar to the normal distributions.
- It became known as "Student's t" because Guinness would not allow him to mention beer, Guinness, or his own last name. So he called himself, "Student". The "t" comes from people using Gosset's "Student test".





# Flatter and Wider

- The t-distribution is flatter and wider than the z-distribution, but gets closer to the z-distribution as sample size increases.



# t-distribution Characteristics

- Slight differences from z but same purposes:
  - ▣ *to standardize samples for comparisons,*
  - ▣ *determine probability of observing certain means,*
  - ▣ *and creating confidence intervals*
- t-distribution handles uncertainty (**unknown  $\sigma$** )
  - ▣ It also takes into account the sample size, n
- Similar to the z-distribution when sample size is large
  - ▣ Once sample size is large enough, it mimics the z-distribution with no real difference

# z vs. t-table

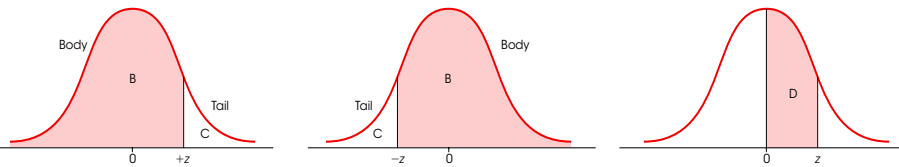
What difference do you notice?  
What similarities?

## z-table

## t-table

**TABLE B.1 The Unit Normal Table\***

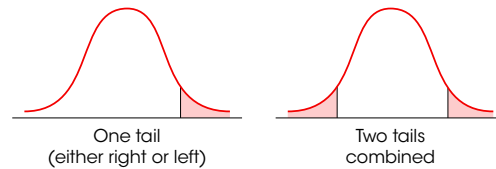
\*Column A lists z-score values. A vertical line drawn through a normal distribution at a z-score location divides the distribution into two sections. Column B identifies the proportion in the larger section, called the *body*. Column C identifies the proportion in the smaller section, called the *tail*. Column D identifies the proportion between the mean and the z-score.  
*Note:* Because the normal distribution is symmetrical, the proportions for negative z-scores are the same as those for positive z-scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

**TABLE B.2 The t Distribution**

Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

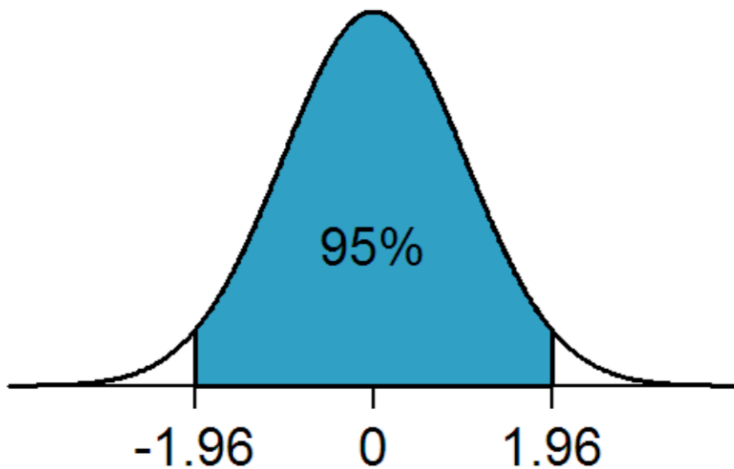
Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

# z vs. t-table

- Main purpose of z and t distributions is to standardize to determine how probable something is and to create confidence intervals
- The main difference between the z and t tables
  - z tables have:
    - Lists of z values to look up corresponding proportions
    - Or look up proportion to find the corresponding z-score
  - t tables have:
    - Degrees of freedom ( $n-1$ )
    - The level of significance you want
    - One or two tail options

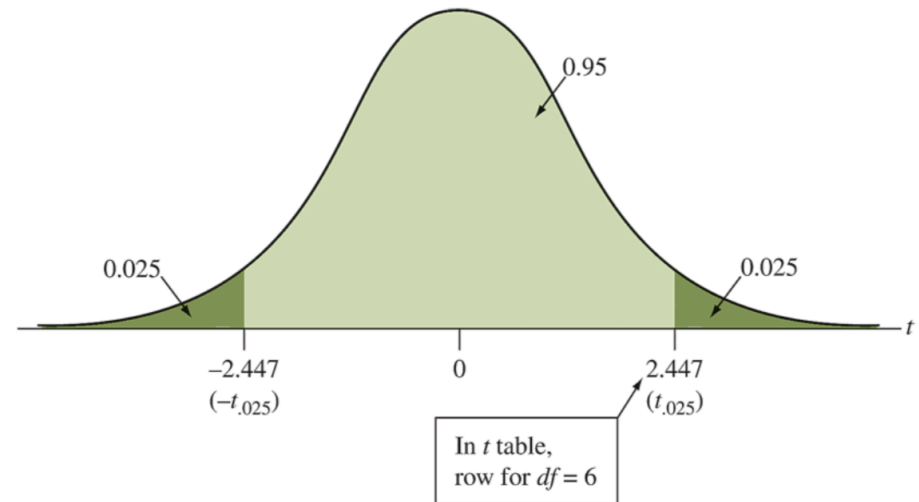
# .05 in z and .05 (df=6) in t

z-distribution



$$z_{.025} = \pm 1.96$$

t-distribution

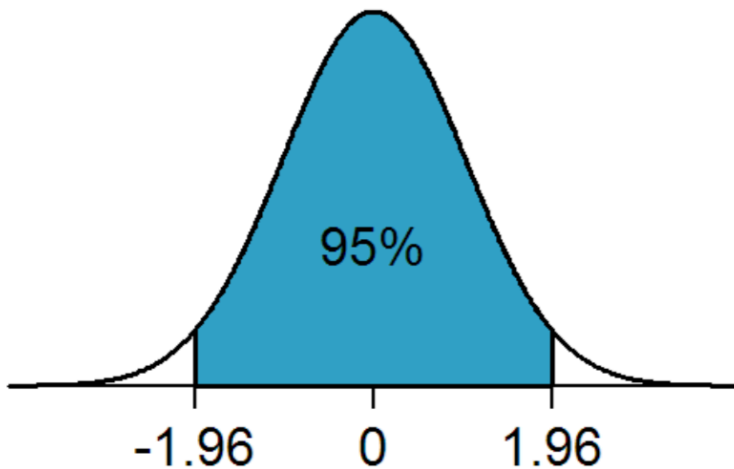


$$\text{When } df = 6, t_{.025} = \pm 2.447$$

This subscript is the level of significance, here it is .05 divided into two sides (.025)

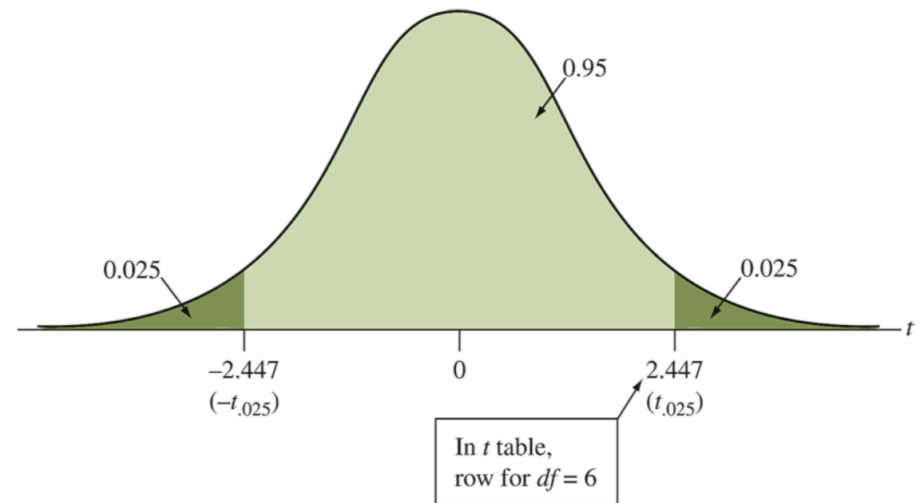
# .05 in z and .05 (df=6) in t

z-distribution



$$z_{.025} = \pm 1.96$$

t-distribution



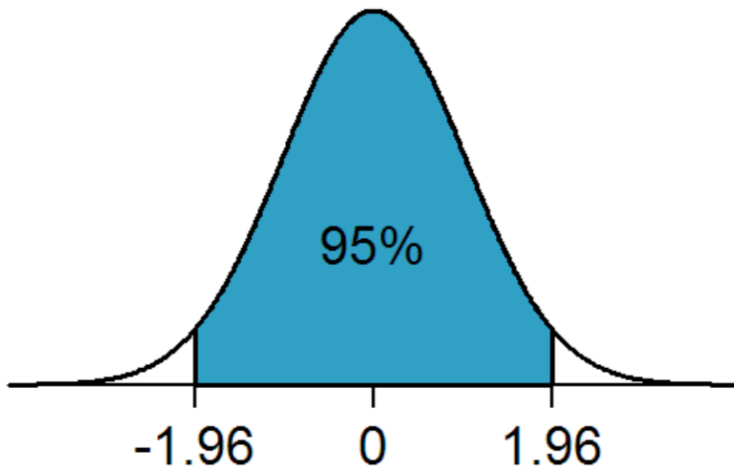
$$\text{When } df = 6, t_{.025} = \pm 2.447$$

Which will have a wider confidence interval?

1.96 vs. 2.447?

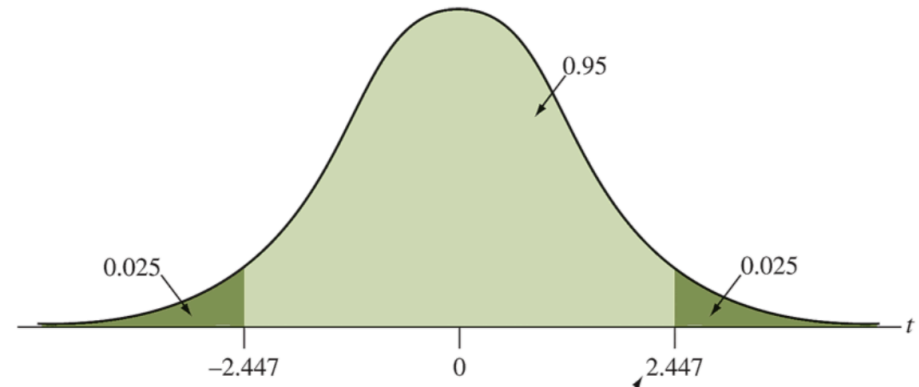
# 95% CI in z and in t (df=6)

z-distribution



$$CI = 10 \pm 1.96 * 2.5$$
$$[5.1, 14.9]$$

t-distribution

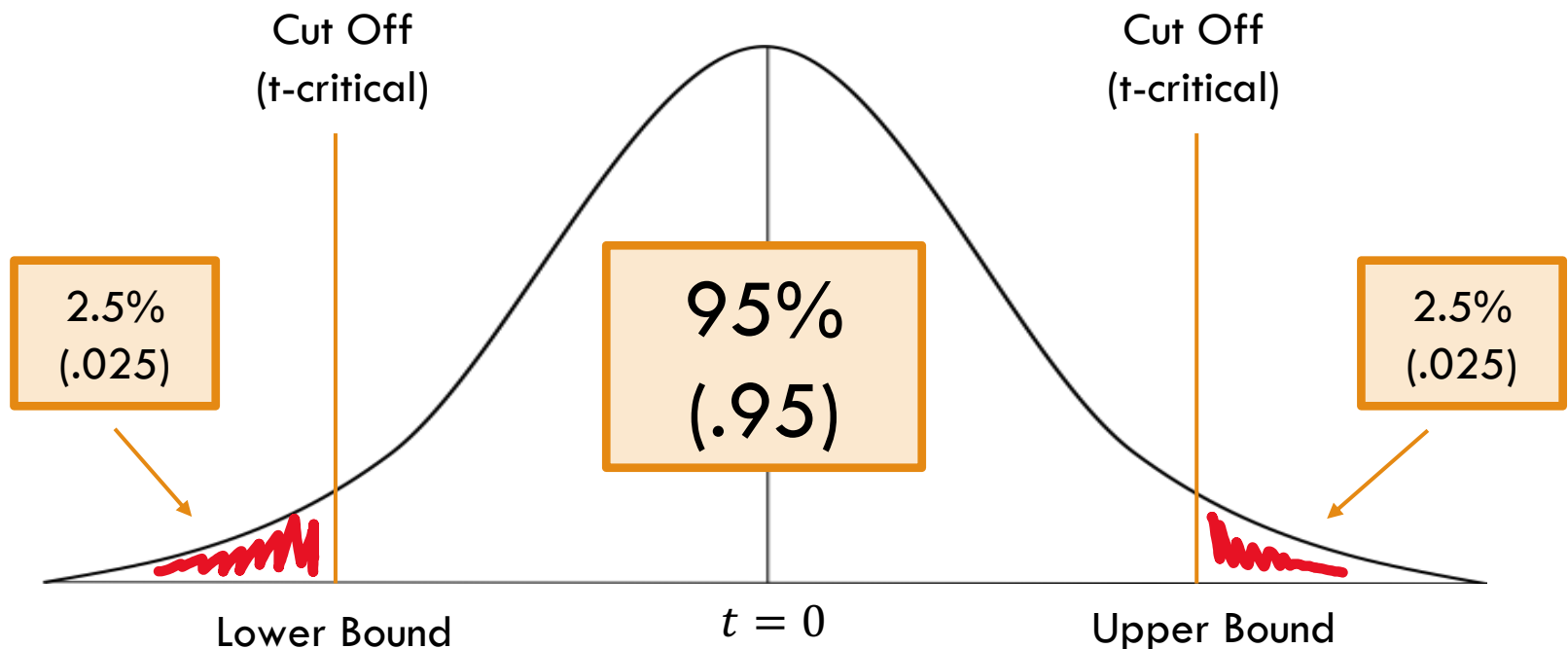


$$CI = 10 \pm 2.447 * 2.5$$
$$[3.9, 16.1]$$

All else being equal, because of the extra uncertainty of not knowing  $\sigma$ , the t-distribution will result in a *wider* confidence interval.

# How Do We Know Where It Is?

- Taking into account sample size and our desired level of confidence, we can look up on a table to determine our critical t-value(s)





# The t-table

# t-Table: New Things

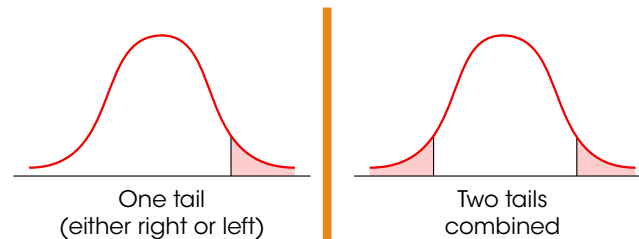
## □ One or Two Tail

- ▣ All 5% on one side vs. 2.5% on each side for different type so tests
- ▣ Confidence intervals are always two tailed

## □ degrees of freedom (df)

- ▣ This is how we take sample size into account

■  $df = n - 1$



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
df	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169

## □ Level of Significant

- ▣ How confident you want to be
- .05 = 95% confidence

# Only the Important Stuff

- With z-tables, we are looking for an exact proportion, i.e. probability. This is possible because  $\sigma$  is known. But now that  $\sigma$  is unknown and we might have a small sample size, there isn't one chart for every single combination of degrees of freedom and one vs. two tailed tests. There would be way too many charts. Instead, we use charts that focus on the levels of significance we are interested in as researchers. Those levels are typically .10, .05, and .01.
- With the t-table, instead of looking up a z-value and getting the proportion, we are going to look for the t-value that corresponds to the proportion we want, like .05.

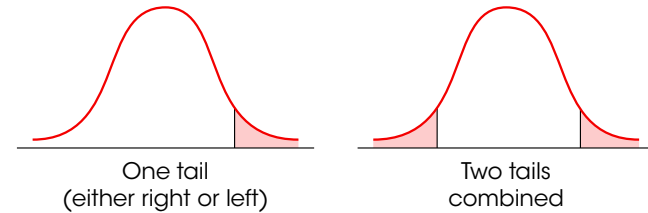
**TABLE B.2 The t Distribution**

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.

# t-table

- We need to know how far out that 5% (2.5% on either side) is and what t-critical corresponds it to...

Why can't we just use 1.96 anymore?



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

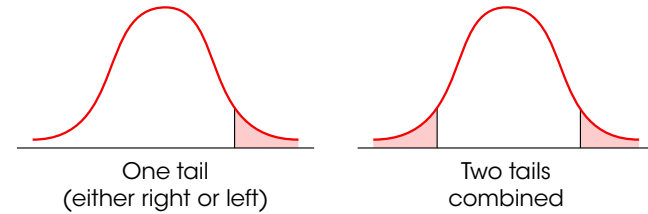
**TABLE B.2 The t Distribution**

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.

# t-table

Why can't we just use 1.96 anymore?

Because  $\sigma$  is unknown AND we have to take sample size into account...



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
df	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.766	1.771	2.777	4.179	6.578	9.348
4	0.741	1.734	2.706	4.115	6.349	9.000
5	0.727	1.711	2.678	4.075	6.256	8.846
6	0.718	1.696	2.658	4.045	6.181	8.716
7	0.713	1.688	2.645	4.025	6.119	8.601
8	0.709	1.683	2.637	4.015	6.067	8.501
9	0.706	1.679	2.632	4.008	6.024	8.413
10	0.703	1.676	2.628	4.002	5.987	8.337
11	0.699	1.673	2.625	3.998	5.955	8.271
12	0.697	1.671	2.623	3.995	5.927	8.213
13	0.695	1.669	2.621	3.993	5.901	8.162
14	0.694	1.668	2.620	3.992	5.877	8.117
15	0.693	1.667	2.619	3.991	5.854	8.077
16	0.692	1.666	2.618	3.990	5.833	8.040
17	0.691	1.665	2.617	3.990	5.813	8.006
18	0.690	1.664	2.616	3.989	5.794	7.975
19	0.689	1.663	2.615	3.989	5.776	7.946
20	0.688	1.662	2.614	3.988	5.759	7.919
21	0.688	1.661	2.613	3.988	5.743	7.894
22	0.687	1.660	2.612	3.987	5.728	7.870
23	0.687	1.659	2.611	3.987	5.713	7.847
24	0.685	1.658	2.610	3.986	5.699	7.825
25	0.684	1.657	2.609	3.986	5.685	7.804
26	0.684	1.656	2.608	3.985	5.672	7.784
27	0.684	1.655	2.607	3.985	5.659	7.765
28	0.683	1.654	2.606	3.984	5.647	7.747
29	0.683	1.653	2.605	3.984	5.635	7.730
30	0.683	1.652	2.604	3.983	5.624	7.714
40	0.681	1.649	2.601	3.980	5.599	7.670
60	0.679	1.646	2.599	3.977	5.579	7.630
120	0.677	1.644	2.597	3.975	5.562	7.600
$\infty$	0.674	1.642	2.596	3.973	5.548	7.576

But notice how if you sample everyone (infinity), the corresponding 95% confidence t-value is 1.96 (like the z-value associated with 95% confidence)

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

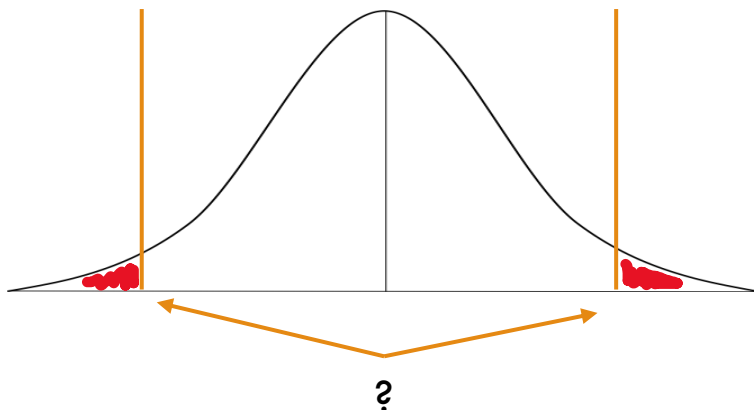
# Degrees of Freedom = $n - 1$

- Degrees of Freedom are the number of scores allowed to vary when calculating a statistic...
  - Suppose you want an average score of 85 ( $\bar{x} = 85$ ) on all three exams.
    - There are no restrictions for the first two exam scores
      - Ex. You could get a 75 and a 85 (or a 65 and a 95, etc) for an average of 80, but that last score has to be a 95 to get that 85 average
- As soon as you know your first two exam scores, the mean ( $\bar{x} = 80$ ) dictates the score that you need on the third exam (ex. a 95).
  - Notice that the first two exam scores are free to vary, but the third score is fixed by the mean (i.e.  $n - 1$  scores are free)

For now, what is most important is that you know how to calculate the degrees of freedom for a sample so you can look up the correct t-critical.

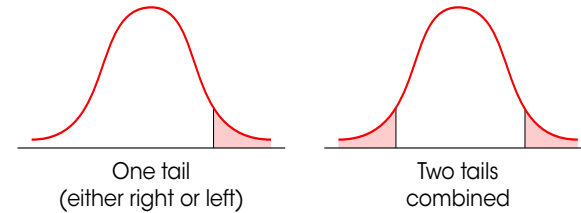
# Quick Practice

- I have a sample size of  $n = 10$ , what are my  $df$ ?
- If I want to construct a 95% CI, which is two tailed, what is the critical  $t$ -value?



**TABLE B.2 The  $t$  Distribution**

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.

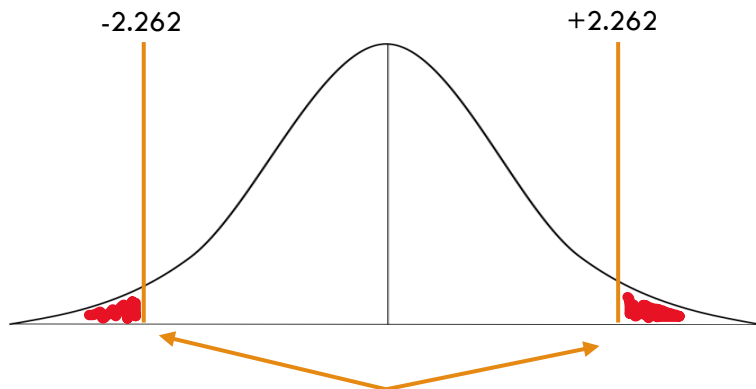


$df$	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

# Quick Practice

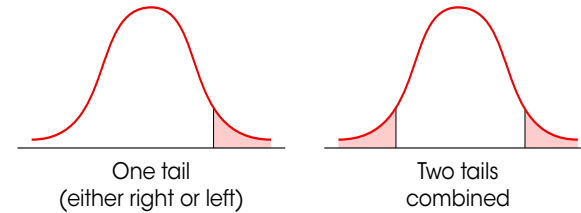
- I have a sample size of  $n = 10$ , what are my  $df$ ?
  - ▣  $10 - 1 = 9$
- If I want to construct a 95% CI, two tailed, what are the critical  $t$ -values?
  - ▣ 2.262



$df = 9, t = \pm 2.262$

**TABLE B.2 The t Distribution**

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.



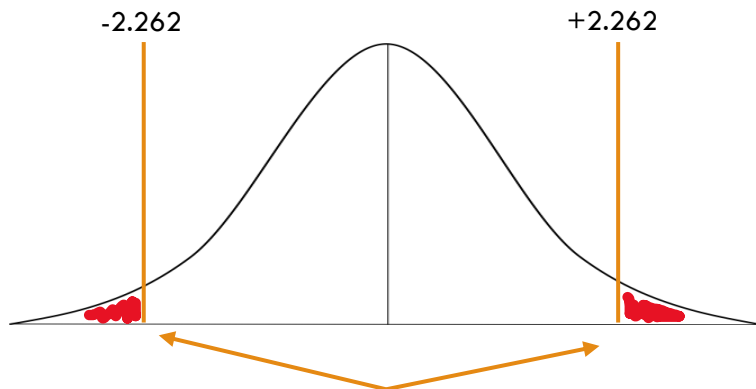
	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
$df$	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.



# Quick Practice

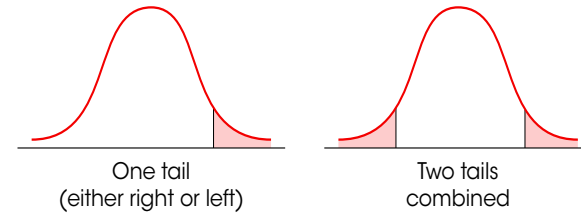
- So the standardized t-value to construct a 95% CI for a sample with 10 people when  $\sigma$  is unknown is 2.262.



$df = 9, t = \pm 2.262$

TABLE B.2 The t Distribution

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

# Confidence Interval, Again

Confidence interval for a mean when  $\sigma$  is unknown:

$$df = n - 1$$
$$CI = \bar{x} \pm t_{critical} * \frac{s}{\sqrt{n}}$$

What are the main difference when calculating a CI when  $\sigma$  is unknown?

# Confidence Intervals w/ t instead of z

Are confidence interval wider or narrower when  $\sigma$  is unknown?

Does uncertainty make your CI more or less precise?

Does precision mean wider or narrower?

$$df = n - 1$$

$$95\% CI = \bar{x} \pm t_{critical} * \frac{s}{\sqrt{n}}$$

# Confidence Intervals w/ t instead of z

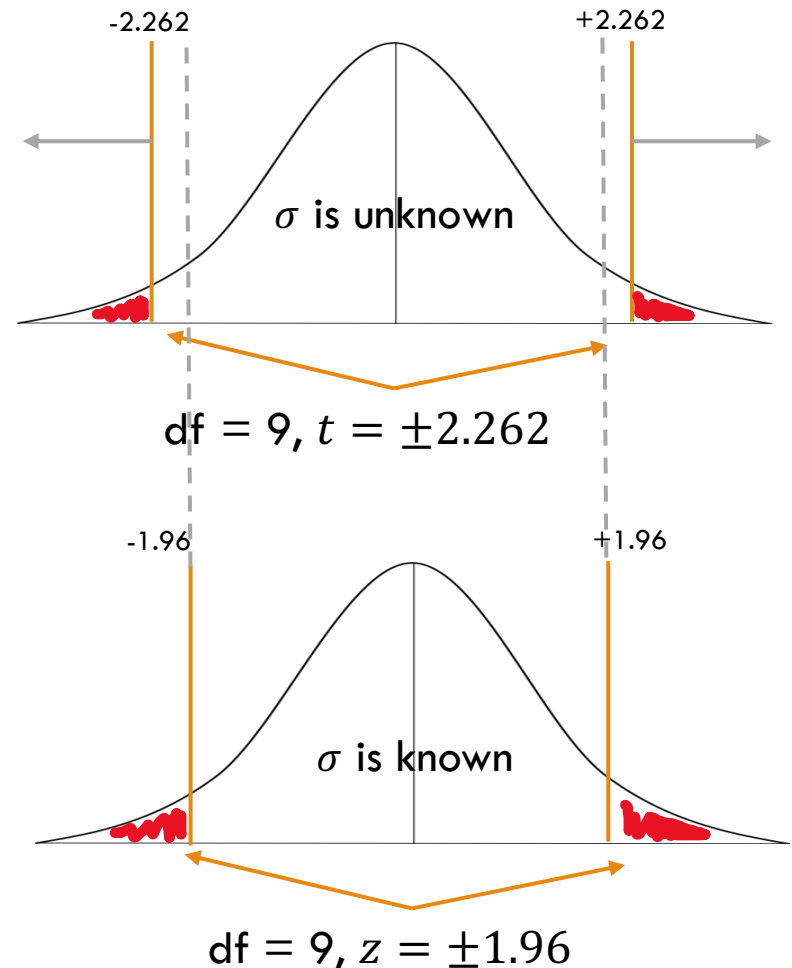
- What affects the width of our CI?
  - ▣ Margin of Error (MOE) which is determined by:
    - Sample size, standard deviation, and the critical t value
- We know that using t is always going to result in a wider CI because it “penalizes” small sample sizes and unknown  $\sigma$ 
  - ▣ For 95% CI, z is always 1.96 while t depends on sample size but is always equal to (when n is large) or *greater* than z
  - ▣ The larger the critical value, the *wider* your CI

$$df = n - 1$$

$$95\% CI = \bar{x} \pm t_{critical} * \frac{s}{\sqrt{n}}$$

# Comparing t to z

- The standardized t-critical to construct a 95% CI for a sample with 10 people when  $\sigma$  is unknown is  $\pm 2.262$ .
- Notice how the “cut offs”, the critical t-value, moves further outward, away from the mean. Compare this to the same situation but assume  $\sigma$  is known where the z-value is  $\pm 1.96$
- The larger the critical t-value, the wider your CI will be, meaning less precise.



# Try it.

- A random sample of 25 college graduates revealed that they worked an average of 6 years at a job before being promoted, with a standard deviation of 1.3 years. Compute and interpret a 99% confidence interval for the mean number of years worked at a job before being promoted.

$$\bar{x} = 6, \quad s = 1.3$$

$$n = 25 \rightarrow df = 24$$

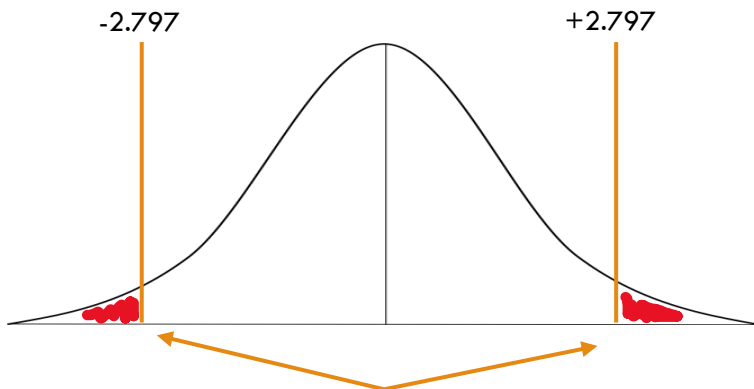
*Level of Significance = 99% or .01*

# Quick Practice

□  $\bar{x} = 6, \quad s = 1.3$

□  $n = 25 \rightarrow$   
 $df = 24$

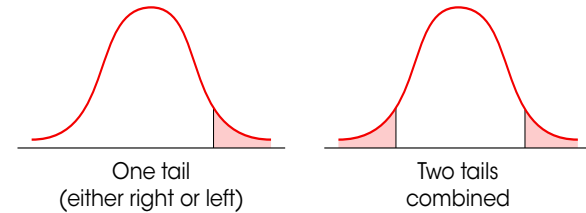
□  $t_{.005, 24} = 2.797$



$df = 24, t = \pm 2.797$

TABLE B.2 The t Distribution

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.



df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

# Quick Practice

□  $\bar{x} = 6, \quad s = 1.3$

□  $n = 25 \rightarrow$   
 $df = 24$

□  $t_{.005, 24} = 2.797$

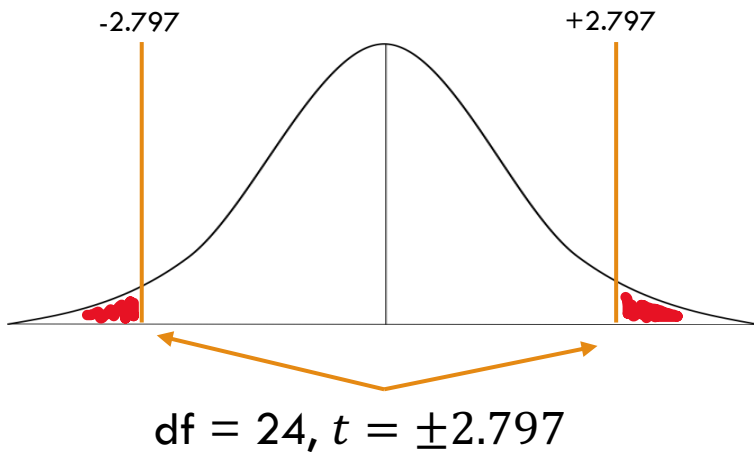
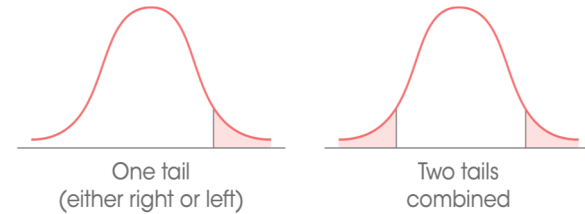


TABLE B.2 The t Distribution

Table entries are values of  $t$  corresponding to proportions in one tail or in two tails combined.



	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
					0.01	
						63.657
						9.925
						5.841
						4.604
						4.032
						3.707
						3.499
						3.355
						3.250
						3.169
						3.106
						3.055
						3.012
						2.977
						2.947
						2.921
						2.898
						2.878
						2.861
						2.845
						2.831
						2.819
						2.807
24						2.797
25						2.787
26						2.779
27						2.771
28						2.763
29						2.756
30						2.750
40						2.704
60						2.660
120						2.617
$\infty$						2.576

Why is the critical  $t$  so large?  
 Because we want to be 99% confident this time... So we have to *widen* our confidence interval to be 99% confident we capture true  $\mu$ .

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.



# Try it.

Information:

$$\bar{x} = 6, \quad s = 1.3$$

$$n = 25 \rightarrow df = 24$$

$$t_{.005, 24} = 2.797$$

Computing the 99% CI:

$$99\% CI = \bar{x} \pm t_{.005} * \frac{s}{\sqrt{n}}$$

$$99\% CI = 6 \pm 2.797 * \frac{1.3}{\sqrt{25}}$$

$$99\% CI = 6 \pm 2.797 * .26$$

$$99\% CI = 6 \pm .727$$

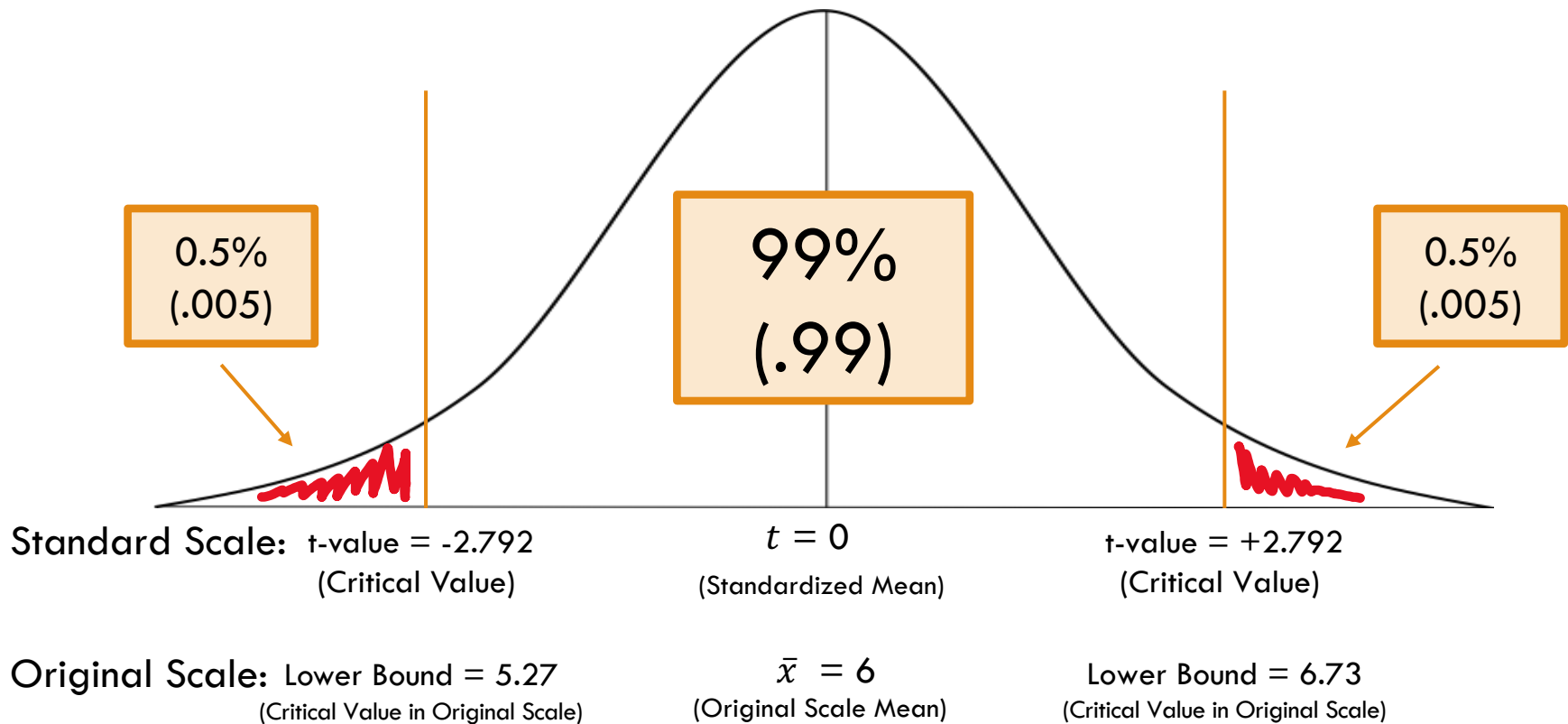
$$99\% CI = [5.27, 6.73]$$

How would you  
interpret this  
Confidence Interval,  
what is it saying?

# Interpretation

- A random sample of 25 college graduates revealed that they worked an average of 6 years at a job before being promoted, with a standard deviation of 1.3 years. Compute and interpret a 99% confidence interval for the mean number of years worked at a job before being promoted.
  - ▣ “If we take repeated samples of size 25 and compute a 99% confidence interval each time, approximately 99% of the intervals would contain the true mean number of years worked at a job before being promoted.”
  - ▣ The average number of years it takes to get promoted is somewhere between 5.27 year and 6.73 years.”
  - ▣ 1 out of 100 times we’ll get a confidence interval that is too high or too low

# Translating Numbers



$$99\% CI = [5.27, 6.73]$$

# That 1 in 100 Miss

\*Made up numbers for example.

1 in 100 times you will get an average that is out here

$$P(t < -2.792) = .005, 0.5\%$$

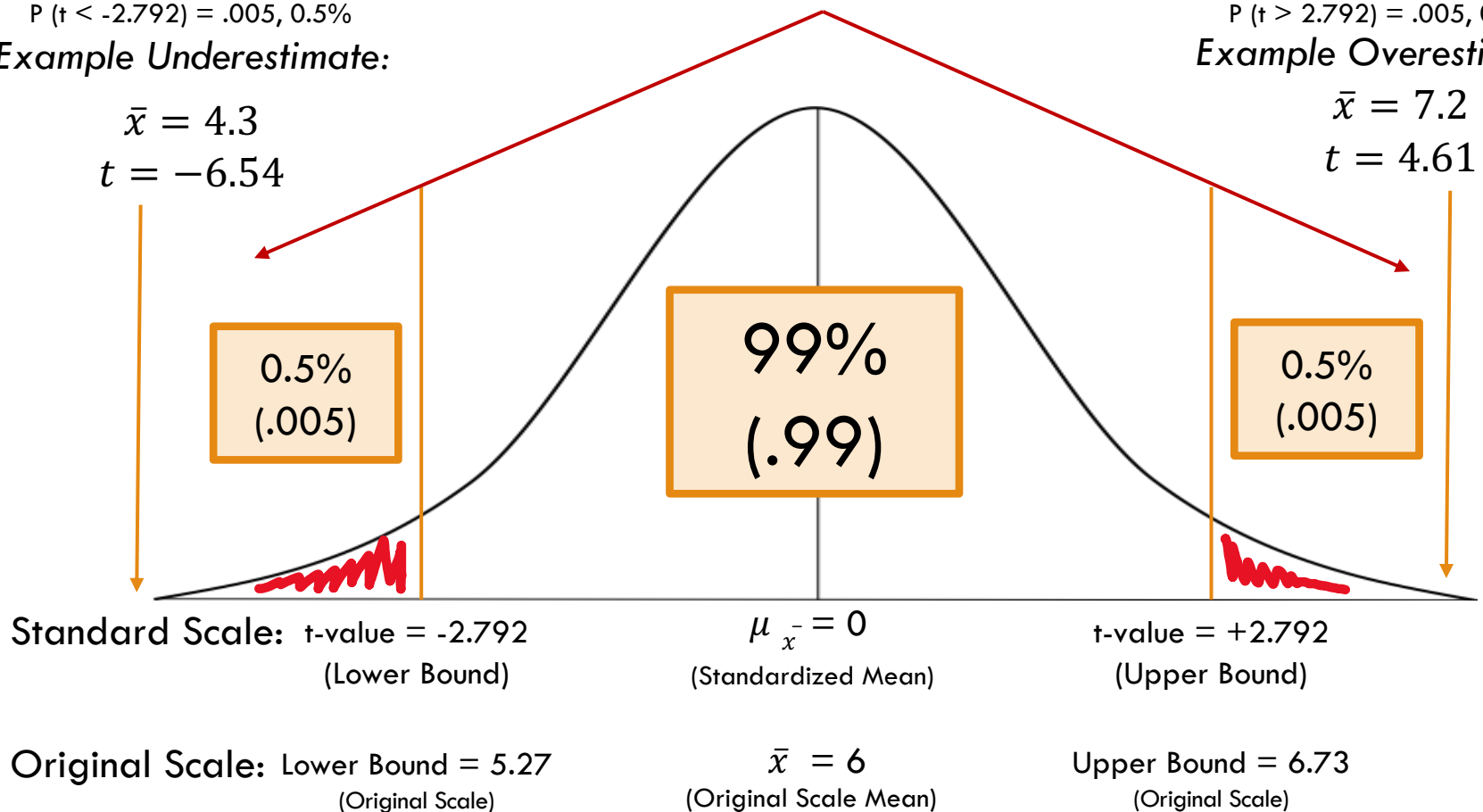
Example Underestimate:

$$\bar{x} = 4.3$$
$$t = -6.54$$

$$P(t > 2.792) = .005, 0.5\%$$

Example Overestimate

$$\bar{x} = 7.2$$
$$t = 4.61$$



# Which one?

- If I ask you to calculate a Confidence Interval, how do you know which method, which table, to use?
  - Hints:
    - Is this a sample within an KNOWN  $\sigma$ ?
      - If you see Greek,  $\sigma$ , you know you can use the **z-distribution**
    - Is this a sample with UNKNOWN  $\sigma$ ?
    - Do you see an “s” or “ $\sigma$ ”?
      - If you see “s” or “unknown variance”, you will use the **t-distribution**

# Why are we doing this...?

- So far, we've been calculating confidence intervals for statistic of interest like the mean
  - ▣ Point estimates aren't actually that precise... To be more confident, we must be more humble and give a range (confidence interval) for our estimates
  - ▣ And we have to acknowledge our shortcomings in regards to sample size (i.e. df, using t instead of z)
    - Getting boring, I know... But...
- This is all working up to being able to test bigger, more interesting ideas...
- Up next...

# Up Next...



# Hypothesis Testing

This is what we're really working towards!

# Critical $t$ and Confidence Intervals in R



# Using R to Look Up Critical T

```
#####  
##### Critical t-values #####  
#####  
  
# You can look up the critical t value using the "qt()" function in R  
# Gives the critical values based on the confidence level you want and df  
# p = probabilities, i.e. the confidence level you want, ex. 95% p = c(.025, .975)  
  
# 95% Confidence Interval is associated with 0.025 and .975 (.975 - .025 = .95)  
qt(p = c(.025, .975), df = )  
  
# 99% Confidence Interval is associated with 0.005 and 0.995 (0.995 - .005 = .99)  
qt(p = c(.005, .995), df = )  
  
# Example with n = 10  
n <- 10  
df <- n-1  
  
# Critical t-value associated with 95% Confidence  
qt(p = c(.025, .975), df = df) #critical t is 2.26
```

# Using R to Create CI, Unknown $\sigma$

```
#####  
## Confidence Intervals for Unknown Sigma ##  
#####  
##### Promotion Example #####  
#####  
  
average <- 6  
sd <- 1.3  
n <- 25  
df <- n - 1  
  
# Get the critical t value associated with a certain level of confidence based on df  
  
# 99% Confidence Interval is associated with 0.005 and 0.995 (0.995 - .005 = .99)  
critical_t_99 <- qt(p = c(.005, .995), df = df)  
  
# Calculate the Margin of Error  
# This will be two numbers, the +MOE and the -MOE  
moe <- critical_t_99 * sd/sqrt(n)  
  
# You only have to add the MOE to the average to get back the upper and lower bound  
confidence_interval <- average + moe
```