The "t" test

- When doing quantitative research or analysis, we are most often interested in a large *population;*
  - The average GPA of a university freshman class
  - The prevalence of depression among those age 65 or older
  - The number of round-trip miles driven to work by staff members
  - The number of customers served a McDonald's between Noon and 1PM
- However due to time and cost, we almost always use *sample data* to represent the larger population
- But sample data is always (at best) an *estimate or approximation* of the larger population from which it is selected

- As our sample size n becomes small, we are *less certain* that it is representative of our entire population; there is greater risk of error
- Additionally, we often know nothing about our population; its mean, variance, or standard deviation

So it seems like we are flying blind; we are using sample data from a larger population about which we have little-to-no information.

- Hopefully it is intuitive that the larger our sample size, the more confident we are that it is representative of the entire population
- In a larger sample, we are likely to capture the natural variation and diversity in the population data
- A small sample increases the chance that we either miss that variation or over-emphasize it
- However sample size, in most cases, does adhere to the "Law of Diminishing Returns"
- There is a point when increasing the sample size offers no more statistical benefits (which is good because it makes research more "doable")

- When our sample size is $n \leq 30$ and/or we do not know the variance/standard deviation of the *population* we use the t-distribution instead of the z-distribution (standard normal distribution)
- What does this do for us?
  - The t-distribution allows us to use small samples; $n \leq 30$
  - But to do so we sacrifice some certainty in our calculations; margin-of-error trade-off
  - It takes sample size into account using $n - 1$ *degrees of freedom*; there is a different t-distribution for any given sample size
  - The bell curve shape is "squished" in the middle and "fatter" on the ends (tails); squishier and fatter the smaller the sample size
  - However as $n > 30$ and definitely by $n \geq 100$, the t-distribution and standard normal z-distribution become indistinguishable

From
http://mathworld.wolfram.com/Studentst-Distribution.html

$$t \equiv \frac{\bar{x} - \mu}{s / \sqrt{N}},$$

Where $\bar{x}$= sample mean   μ = population mean   s= standard deviation   N sample size

The t score produced by this transformation can be associated with a unique cumulative probability.

Statisticians use t_α to represent the t-score that has a cumulative probability of (1 - α). For example, suppose we were interested in the t-score having a cumulative probability of 0.95. In this example, α would be equal to (1 - 0.95) or 0.05. We would refer to the t-score as $t_{0.05}$

Note: Because the t distribution is symmetric about a mean of zero, the following is true.

$t_\alpha = -t_{1 - \text{alpha}}$ And $t_{1 - \text{alpha}} = -t_\alpha$

Thus, if $t_{0.05} = 2.92$, then $t_{0.95} = -2.92$.

When to Use the t Distribution

The t distribution can be used

- The population distribution is normal.
- The population distribution is symmetric, unimodal, without outliers, and the sample size is at least 30.

Sample Problem

Acme Corporation manufactures light bulbs. The CEO claims that an average Acme light bulb lasts 300 days (population mean). A researcher randomly selects 15 bulbs (sample size) for testing. The sampled bulbs last an average of 290 days (sample mean), with a standard deviation (of sample) of 50 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 290 days?

Calculate t $t \equiv \frac{\bar{x} - \mu}{s/\sqrt{N}}$,

$$t = \frac{290 - 300}{50/SQRT(15)} = -0.7745$$

From t tables find 0.7745 for df = 14 (in blue in table) α: 0.774

Therefore cumulative probability is 1-.774 =0.226

Hence, if the true bulb life were 300 days, there is a 22.6% chance that the average bulb life for 15 randomly selected bulbs would be less than or equal to 290 days.

Hence, there is a 22.6% chance that the average sampled light bulb will burn out within 290 days.

See table on next slide

Problem 2:

The Acme Chain Company claims that their chains have an average breaking strength of 20,000 pounds, with a standard deviation of 1750 pounds. Suppose a customer tests 14 randomly-selected chains. What is the probability that the average breaking strength in the test will be no more than 19,800 pounds?

Sample mean x = 19,800, Population mean u = 20,000 N = 14, sample s =1750
degrees of freedom n-1 =13

t = -0.427

From table 0.427 at df=13 you get α = 0.662

Therefore the cumulative probability:
(1-.662) = 0.338

Thus, there is a 33.8% probability that an Acme chain will snap under 19,800 pounds of stress.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300

ERROR BARS

Figures with error bars can, if used properly, give information describing the data (descriptive statistics), or information about what conclusions, or inferences, are justified (inferential statistics). These two basic categories of error bars are depicted in exactly the same way, but are actually fundamentally different.

Table I. **Common error bars**

Error bar	Type	Description	Formula
Range	Descriptive	Amount of spread between the extremes of the data	Highest data point minus the lowest
Standard deviation (SD)	Descriptive	Typical or (roughly speaking) average difference between the data points and their mean	$SD = \sqrt{\frac{\sum (X - M)^2}{n-1}}$
Standard error (SE)	Inferential	A measure of how variable the mean will be, if you repeat the whole study many times	$SE = SD/\sqrt{n}$
Confidence interval (CI), usually 95% CI	Inferential	A range of values you can be 95% confident contains the true mean	$M \pm t_{(n-1)} \times SE$, where $t_{(n-1)}$ is a critical value of t. If n is 10 or more, the 95% CI is approximately $M \pm 2 \times SE$.

Range and standard deviation (SD) are used for descriptive error bars because they show how the data are spread (Fig. 1).

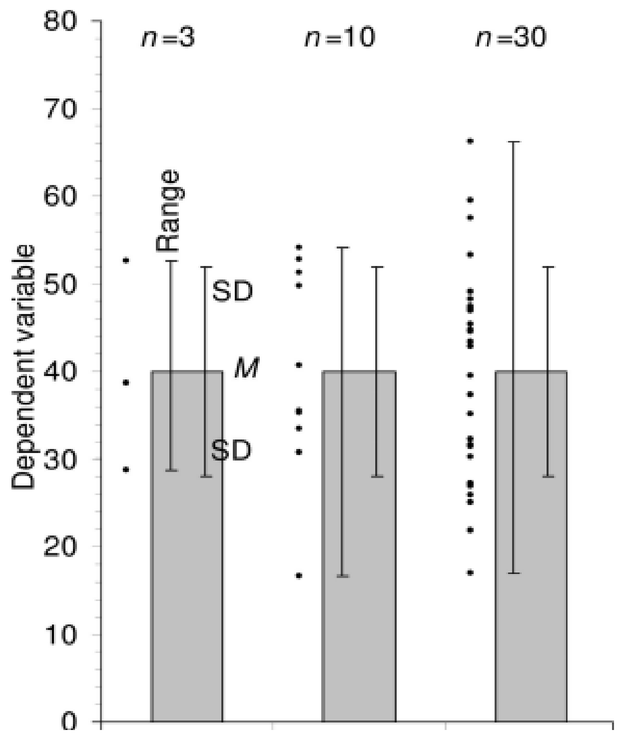

Figure 1. **Descriptive error bars.** Means with error bars for three cases: n = 3, n = 10, and n = 30. The small black dots are data points, and the column denotes the data mean M. The bars on the left of each column show range, and the bars on the right show standard deviation (SD). M and SD are the same for every case, but notice how much the range increases with n. Note also that although the range error bars encompass all of the experimental results, they do not necessarily cover all the results that could possibly occur. SD error bars include about two thirds of the sample, and 2 x SD error bars would encompass roughly 95% of the sample.

Cumming G, Fidler F, Vaux DL. Error bars in experimental biology. *The Journal of Cell Biology*. 2007;177(1):7-11.

SE (Standard Error) and CI (Confidence Interval) Error Bars

To make inferences from the data (i.e., to make a judgment whether the groups are significantly different, or whether the differences might just be due to random fluctuation or chance), a different type of error bar can be used. These are standard error (SE) bars and confidence intervals (CIs).

The mean of the data, M, with SE or CI error bars, gives an indication of the region where you can expect the mean of the whole possible set of results, or the whole population, μ, to lie. The interval defines the values that are most plausible for μ.

Standard Error (SE=SD/√n) - dependent on sample size n; decreases as n increases

Confidence Interval (CI)

By convention, if $P < 0.05$ you say the result is statistically signifi cant, and if $P < 0.01$ you say the result is highly signifi cant and you can be more confi dent you have found a true effect. Independent of sample size n; easier to understand; preferred by most scientists

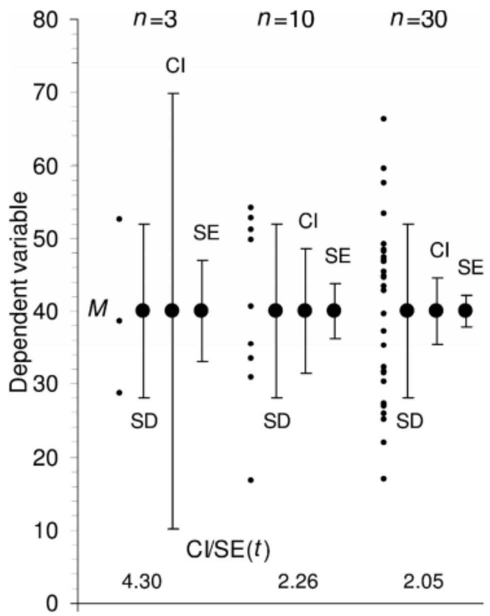

Figure 4. **Inferential error bars.** Means with SE and 95% CI error bars for three cases, ranging in size from $n = 3$ to $n = 30$, with descriptive SD bars shown for comparison. The small black dots are data points, and the large dots indicate the data mean *M*. For each case the error bars on the left show SD, those in the middle show 95% CI, and those on the right show SE. Note that SD does not change, whereas the SE bars and CI both decrease as *n* gets larger. The ratio of CI to SE is the *t* statistic for that *n*, and changes with *n*. Values of *t* are shown at the bottom. For each case, we can be 95% confident that the 95% CI includes μ, the true mean. The likelihood that the SE bars capture μ varies depending on *n*, and is lower for $n = 3$ (for such low values of *n*, it is better to simply plot the data points rather than showing error bars, as we have done here for illustrative purposes).

Cumming G, Fidler F, Vaux DL. Error bars in experimental biology. *The Journal of Cell Biology*. 2007;177(1):7-11.

Error Bars and the Difference Between two Means		
Error Bar	Bars Do Not Overlap	Bars Overlap
Standard Error of the Mean	No statistical inference can be made	Means ARE NOT different in a statistically significant way
95 % Confidence Interval	Means ARE different in a statistically significant way	No statistical inference can be made