Tuesday, March 12, 2019

- **Warm-up**
  - **Using either the calculator or the table, find the critical z (z^*) for each of the following confidence levels.**

invNorm

- **75%**

$z^* = 1.150$

- **80%**
- **98%**

- **More About Inference with means**

Objectives

Content Objectives: I will begin to work with t-values.

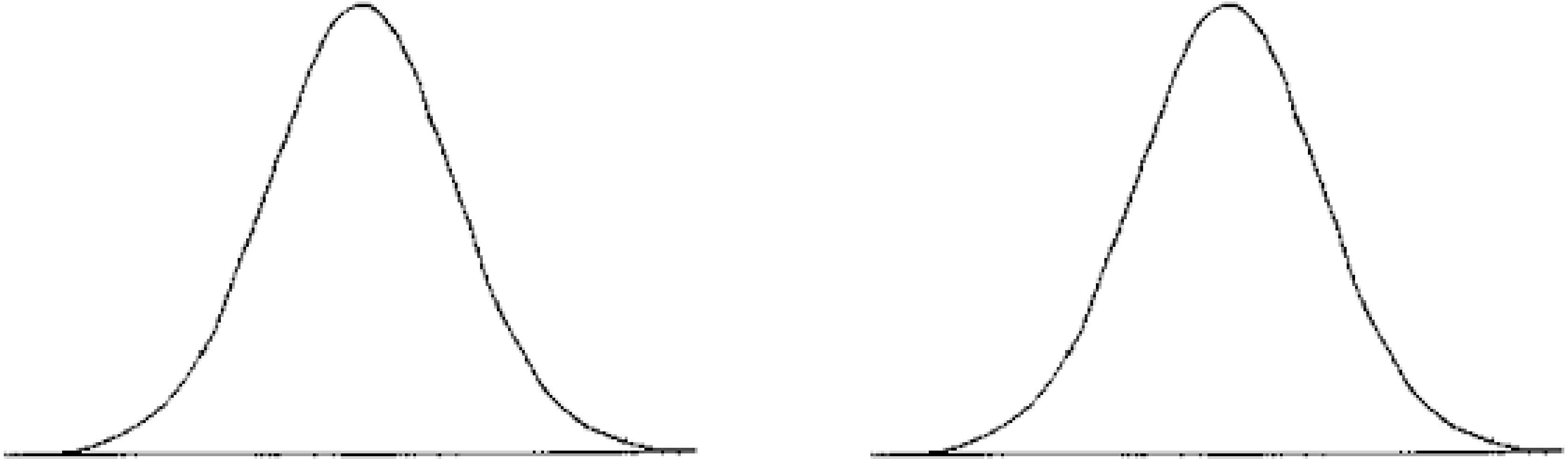
Social Objective: I will listen to others and participate in the class activity.

Language Objective: I will take clear notes that I will be able to refer to.

Getting Started

Now that we know how to create confidence intervals and test hypotheses about proportions, it'd be nice to be able to do the same for means.

Just as we did before, we will base both our confidence interval and our hypothesis test on the sampling distribution model.



The Central Limit Theorem told us that the sampling distribution model for means is Normal with mean μ and standard deviation

$$SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

n = 1

$$\frac{\sigma}{\sqrt{1}} = \sigma$$

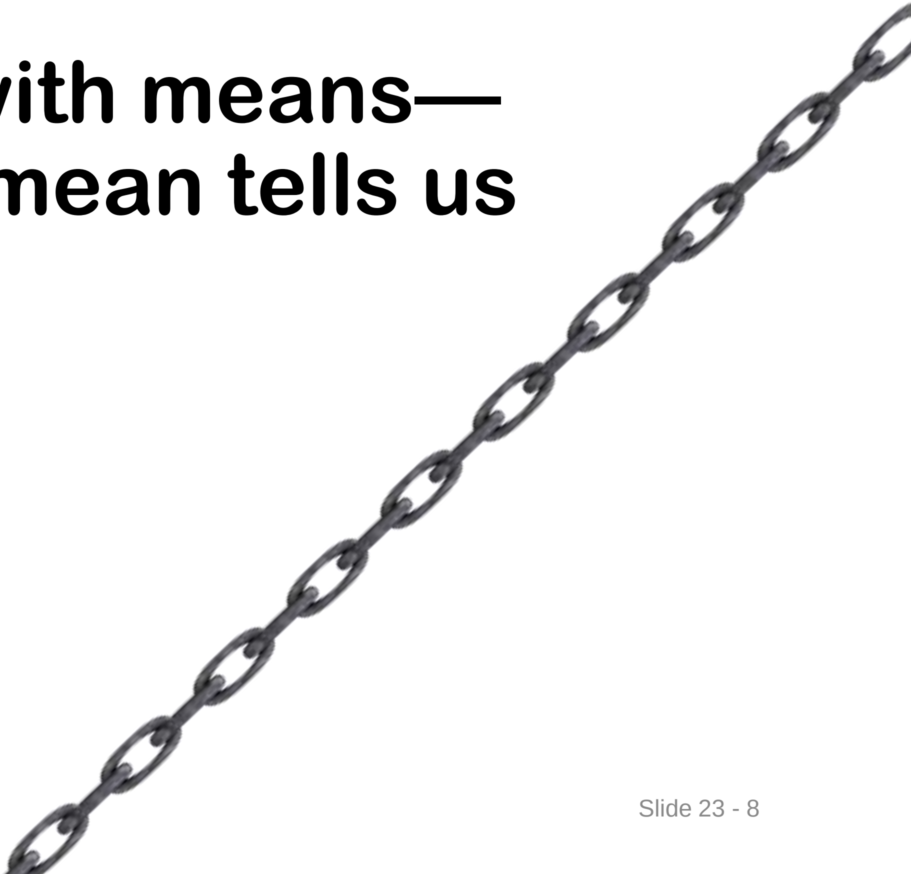
All we need is a random sample of quantitative data.

And the true population standard deviation, σ.

Well, that's a problem…

- **Proportions have a link between the proportion value and the standard deviation of the sample proportion.**

- **This is not the case with means—knowing the sample mean tells us nothing about** $SD(\bar{y})$



- We'll do the best we can: estimate the population parameter σ with the sample statistic s.

must use sample

- Our resulting standard error is

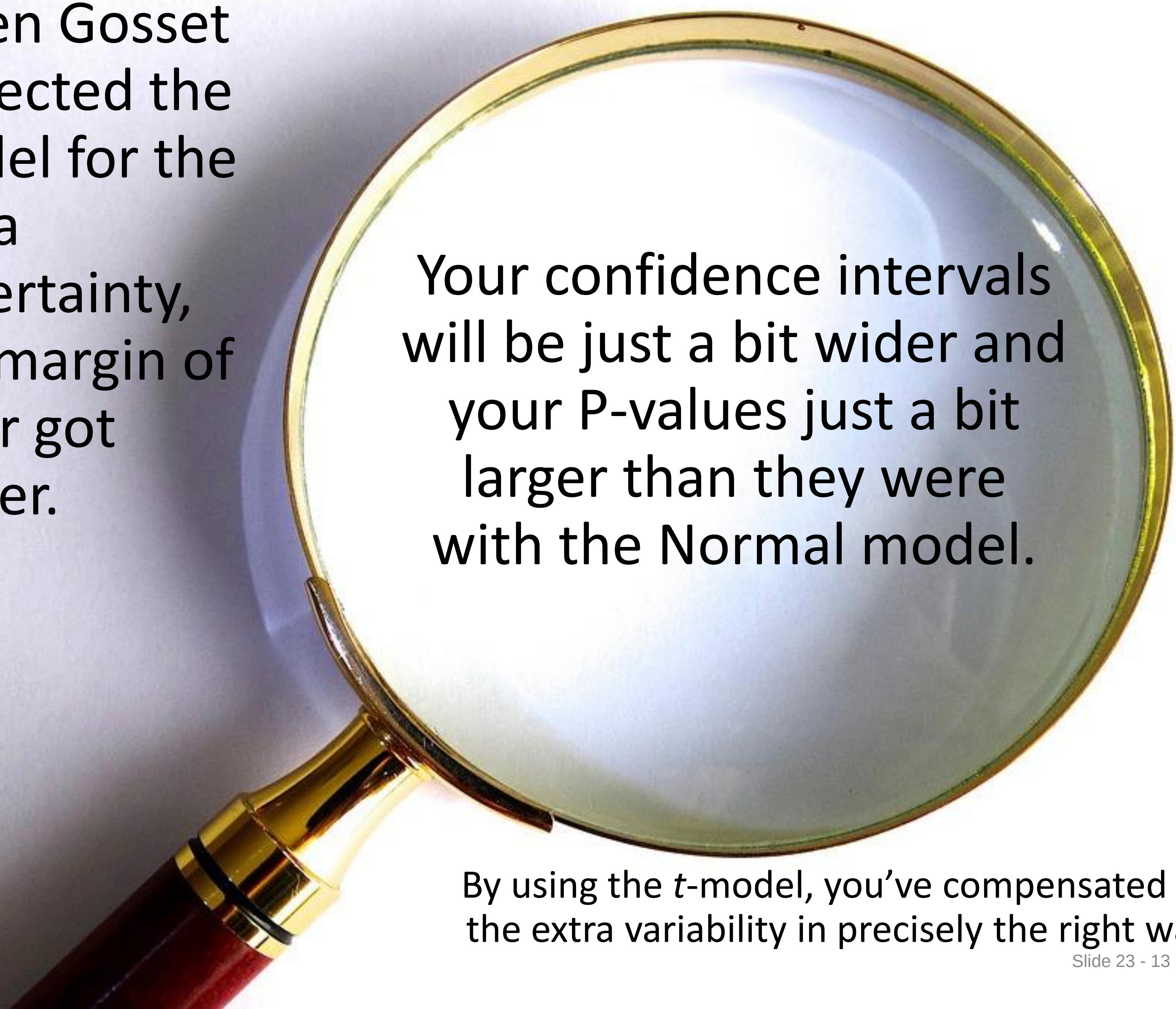
$$SE(\bar{y}) = \frac{s}{\sqrt{n}}$$

- **We now have extra variation in our standard error from *s*, the sample standard deviation.**
  - **We need to allow for the extra variation so that it does not mess up the margin of error and P-value, especially for a small sample.**

And, the *shape* of the sampling model changes—the model is no longer Normal. So, what is the sampling model?

Also called student's *t*-models, they are unimodal, symmetric, and bell shaped, just like the Normal.

But *t*-models with only a few degrees of freedom have much fatter tails than the Normal. (That's what makes the margin of error bigger.)



When Gosset corrected the model for the extra uncertainty, the margin of error got bigger.

Your confidence intervals will be just a bit wider and your P-values just a bit larger than they were with the Normal model.

By using the *t*-model, you've compensated for the extra variability in precisely the right way.

One-sample t-interval for the mean

- When the conditions are met, we are ready to find the confidence interval for the population mean, μ.

- The confidence interval is $$\bar{y} \pm t^{*}_{n-1} \times SE(\bar{y})$$

$$\hat{p} \pm z^{*} SE(\hat{p}) \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where the standard error of the mean is $$SE(\bar{y}) = \frac{s}{\sqrt{n}}$$

- The critical value t^{*}_{n-1} depends on the particular confidence level, *C*, that you specify and on the number of degrees of freedom, $n - 1$, which we get from the sample size.

Critical t

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

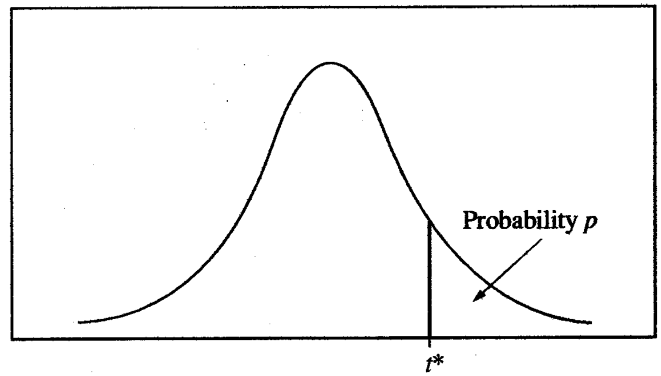


Table B *t* **distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Confidence level C

Assumptions & Conditions

Independence Assumption:

Independence Assumption. The data values should be independent.

Independence Assumption:

Randomization Condition: The data arise from a random sample or suitably randomized experiment. Randomly sampled data (particularly from an SRS) are ideal.

Independence Assumption:

10% Condition: When a sample is drawn without replacement, the sample should be no more than 10% of the population.

Normal Population Assumption:

We can never be certain that the data are from a population that follows a Normal model, but we can check the

Nearly Normal Condition: The data come from a distribution that is unimodal and symmetric.

Check this condition by making a histogram, box plot or Normal probability plot.

30

Sample size

More About the Nearly Normal Condition:

The smaller the sample size *(n* < 15 or so*)*, the more closely the data should follow a Normal model.

For moderate sample sizes (*n* between 15 and 40 or so), the *t* works well as long as the data are unimodal and reasonably symmetric.

For larger sample sizes, the *t* methods are safe to use unless the data are extremely skewed.

Practice Problem

Hallux abducto valgus (call it HAV) is a deformation of the big toe that is fairly uncommon in youth and often requires surgery. Doctors used X-rays to measure the angle (in degrees) of deformity in a random sample of patients under the age of 21 who came to a medical center for surgery to correct HAV. The angle is a measure of the seriousness of the deformity. For these 21 patients, the mean HAV angle was 24.76 degrees and the standard deviation was 6.34 degrees. A dotplot of the data revealed no outliers or strong skewness.

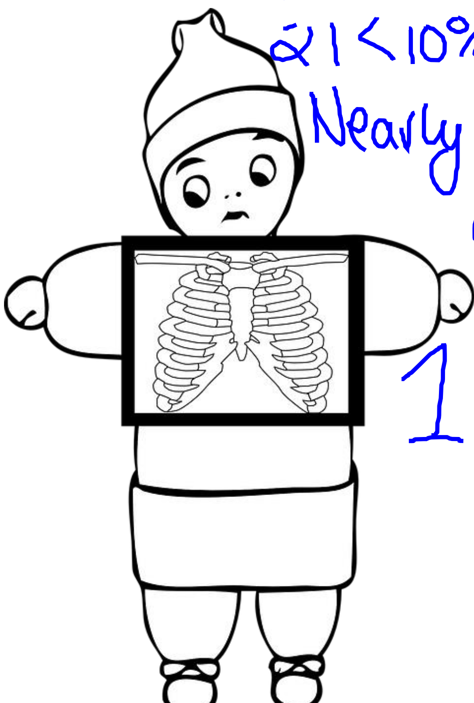
Construct and interpret a 90% confidence interval for the mean HAV angle in the population of all patients.

Random Sample Stated

21 < 10% of all HAV patients

Nearly normal - "dotplot has no outliers nor strong skewness"

1 sample t-interval



$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right)$

$24.76 \pm 1.725\left(\frac{6.34}{\sqrt{21}}\right)$

24.76 ± 2.38

$(22.374, 27.146)$

We are 90% confident that the mean HAV angle in the population of all patients is $24.76° \pm 2.38°$

- Remember that interpretation of your confidence interval is key.
- What NOT to say:
  - "*90% of all the vehicles* on Triphammer Road drive at a speed between 29.5 and 32.5 mph."
    - The confidence interval is about the *mean* not the individual values.
  - "We are 90% confident that *a randomly selected vehicle* will have a speed between 29.5 and 32.5 mph."
    - Again, the confidence interval is about the *mean* not the individual values.

- What NOT to say:
  - "The mean speed of the vehicles is 31.0 mph *90% of the time.*"
    - The true mean does not vary—it's the confidence interval that would be different had we gotten a different sample.
  - "*90% of all samples* will have mean speeds between 29.5 and 32.5 mph."
    - The interval we calculate does not set a standard for every other interval—it is no more (or less) likely to be correct than any other interval.

DO SAY:

- *"90% of intervals that could be found in this way would cover the true value."*

- Or make it more personal and say, *"I am 90% confident that the true mean is between 29.5 and 32.5 mph."*

Homework

- Page 556 (13, 14)