David Youngberg
BSAD 210—Montgomery College

LECTURE 14: CONFIDENCE INTERVALS II

I. Calculating the Margin of Error with Unknown σ

a. We often don't know σ. This requires us to rely on the sample's standard deviation, *s*. This changes everything.

b. Recall that *s* is influenced by sample size, *n*. Bigger sample sizes means we can more reasonably estimate the standard deviation. When *n* gets really large, there's no practical difference between knowing and not knowing σ.

c. But most the time, the sample is good but not enormous. Thus we cannot use our perfect normal distribution. We have to use a different distribution: Student's t-distribution.

   i. Like a normal distribution, the t-distribution is bell-shaped and symmetric around the mean.

   ii. The t-distribution is flatter and wider than the normal distribution.

   iii. The t-distribution depends on the *degrees of freedom*, or the number of values that are free to vary given that certain information is known.

      1. One bit of information that's known for any sample is the sample mean. Therefore, the degrees of freedom (df) for our use here is equal to n – 1.

   iv. The t-distribution is a family of distributions which change based on the degrees of freedom.

d. Here's the equation:

$$\widehat{CI}_{\bar{x}} = \bar{x} \mp t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

   i. The hat over CI reminds us that this is an approximation. Note how similar this equation is to the previous one.

e. Below is a table for the t-distribution with single tail alpha on top and two tail confidence levels on the bottom. Note that as df increases, the critical values approach the z-score.

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											

II. Critical T-values

a. Excel has a CONFIDENCE.T function for calculating the interval with a t distribution. Like CONFIDENCE.NORM, it outputs the margin of error.

b. It also has the whole table for t-values. The command for this is "=T.INV.2T"

   i. *Probability* is the alpha value, assuming two tails.

   ii. *Deg_freedom* is the degrees of freedom: n – 1.

c. So if you want to know the critical t-value for an alpha of 0.05 with six degrees of freedom, you'd type "=T.INV.2T(0.05,6)" and press ENTER.

i. You should get about 2.447, the same as in the table under 95% confidence with 6 degrees of freedom.

ii. Note there is a version without the 2T; the 2T stands for two tails. Without it, you're assuming alpha is only in one tail.

1. If you think the table claims the result should be 1.943 (or more accurately it displays -1.943), it's because you're looking at the one-tailed version and not the two-tailed version. (In the two-tailed version, that 0.05 is equally split between two sides of the distribution, thus a single tail would have a value of 0.25.)

III. Example

a. Recall Theo who works at a steel mill figured out the 95% confidence interval of scrap metal use was between 6,649.4 to 7,350.6 tons, based on an average of 7,000 tons of use from 20 weeks, with a σ (population standard deviation) of 800.

b. Suppose he doesn't know the population standard deviation of scrap metal use per week is 800 tons. Production can vary for all kinds of reasons so he has to estimate sigma using the *sample* standard deviation, calculating it based on the sample of 20 weeks he took.

c. ***To illustrate the effect of switching to the t distribution, assume the sample standard deviation happens to also be 800.*** Therefore, the only think that changes in the calculation is we're using t scores instead of z scores.

d. Using CONFIDENCE.T(0.05,800,20), we get a new margin of error: 374.4 tons, 6.8 percent higher than the 350.6 we found earlier. This higher margin of error reflects that we have less information than before. Because we don't *know* σ, we have to *estimate* σ, and that's not as accurate.

e. Note this distinction shrinks as n increases because the amount of information we have increases. Thus the sample standard deviation (s) is likely to be close to population standard deviation (σ). To illustrate, suppose n was 200 weeks instead of 20 weeks.

i. CONFIDENCE.NORM(0.05,800,200) = 110.9

ii. CONFIDENCE.T(0.05,800,200) = 111.6

iii. Switching to the t distribution when n was so high made the margin of error only 0.6 percent larger.