8.3
Estimating a Population Mean

- STATE and CHECK the Random, 10%, and Normal/Large Sample conditions for constructing a confidence interval for a population mean.
- EXPLAIN how the t distributions are different from the standard Normal distribution and why it is necessary to use a *t* distribution when calculating a confidence interval for a population mean.
- DETERMINE critical values for calculating a *C*% confidence interval for a population mean using a table or technology.
- CONSTRUCT and INTERPRET a confidence interval for a population mean.
- DETERMINE the sample size required to obtain a *C*% confidence interval for a population mean with a specified margin of error.

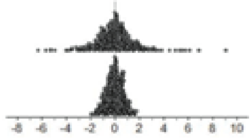
When σ Is Unknown: The *t* Distributions

When we standardize based on the sample standard deviation s_x, our statistic has a new distribution called a ***t* distribution**.

- ✓It has a *different shape* than the standard Normal curve:
- ✓It is symmetric with a single peak at 0,
- ✓However, it has much more area in the tails.

$t = \frac{\bar{x}-\mu}{S_x/\sqrt{n}}$ $\quad t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$

$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ $\quad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$



SE

Like any standardized statistic, *t* tells us how far $\bar{x}$ is from its mean μ in standard deviation units.

There is a different *t* distribution for each sample size, specified by its **degrees of freedom (df)**.

The *t* Distributions; Degrees of Freedom

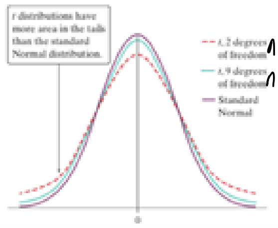
Conditions for Constructing a Confidence Interval About a Proportion

Draw an SRS of size *n* from a large population that has a Normal distribution with mean μ and standard deviation σ. The statistic

$$t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$$

has the ***t* distribution** with **degrees of freedom** df $= n - 1$. When the population distribution isn't Normal, this statistic will have approximately a t_{n-1} distribution if the sample size is large enough.

When comparing the density curves of the standard Normal distribution and *t* distributions, several facts are apparent:



n = 3

n = 10

- ✓The density curves of the *t* distributions are similar in shape to the standard Normal curve.
- ✓The spread of the *t* distributions is a bit greater than that of the standard Normal distribution.
- ✓The *t* distributions have more probability in the tails and less in the center than does the standard Normal.
- ✓As the degrees of freedom increase, the *t* density curve approaches the standard Normal curve ever more closely.

Problem: What critical value t^* from Table B should be used in constructing a confidence interval for the population mean in each of the following settings?

(a) A 95% confidence interval based on an SRS of size $n = 12$.

$t^* = 2.201$

TABLE C *t* distribution critical values

df = 11

	Upper-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	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	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*											

Conditions for Estimating μ

As with proportions, you should check some important conditions before constructing a confidence interval for a population mean.

Conditions For Constructing A Confidence Interval About A Mean

- **Random:** The data come from a well-designed random sample or randomized experiment.
  - **10%:** When sampling without replacement, check that $n \le \frac{1}{10}N$ Independent { we can use $\frac{S_x}{\sqrt{n}}$
- **Normal/Large Sample:** The population has a Normal distribution or the sample size is large ($n \ge 30$). If the population distribution has unknown shape and $n < 30$, use a graph of the sample data to assess the Normality of the population. Do not use *t* procedures if the graph shows strong skewness or outliers.

One-Sample *t* Interval for a Population Mean

The one-sample *t* interval for a population mean is similar in both reasoning and computational detail to the one-sample *z* interval for a population proportion

One-Sample *t* Interval for a Population Mean

When the conditions are met, a *C*% confidence interval for the unknown mean μ is

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

where t^* is the critical value for the t_{n-1} distribution with *C*% of its area between $-t^*$ and t^*.

t-interval

Choosing the Sample Size

We determine a sample size for a desired margin of error when estimating a mean in much the same way we did when estimating a proportion.

Choosing Sample Size for a Desired Margin of Error When Estimating μ

To determine the sample size *n* that will yield a level *C* confidence interval for a population mean with a specified margin of error *ME*:

- Get a reasonable value for the population standard deviation σ from an earlier or pilot study.
- Find the critical value z^* from a standard Normal curve for confidence level *C*.
- Set the expression for the margin of error to be less than or equal to *ME* and solve for *n*: $z^* \frac{\sigma}{\sqrt{n}} \le ME$

$$n = \left(\frac{z^* \cdot \sigma}{ME}\right)^2$$

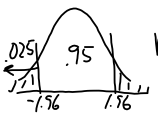
Example: Determining sample size from margin of error

Researchers would like to estimate the mean cholesterol level μ of a particular variety of monkey that is often used in laboratory experiments.

They would like their estimate to be within 1 milligram (ME) per deciliter (mg/dl) of the true value of μ at a 95% confidence level. $z^* = 1.96$

A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl. $\sigma = 5$

Problem: Obtaining monkeys is time-consuming and expensive, so the researchers want to know the minimum number of monkeys they will need to generate a satisfactory estimate.



$$n = \left(\frac{1.96 \cdot 5}{1}\right)^2 = 97$$

65 $n = 47$ $\bar{x} = -.03587$

$s = .02506$

a) Conditions:

Random - Stated

10%: $47 \le \frac{1}{10}$(All BF Women) yes

Normal: $47 \ge 30$ yes

→ Large Counts by CLT

$(-.0457, -.026)$

We are 99% confident that the interval $-.0457$ to $-.026$ capture the true Mean Bone mineral loss for breast feeding mothers.

b) yes, the interval is all negative indicating a loss in bone mineral