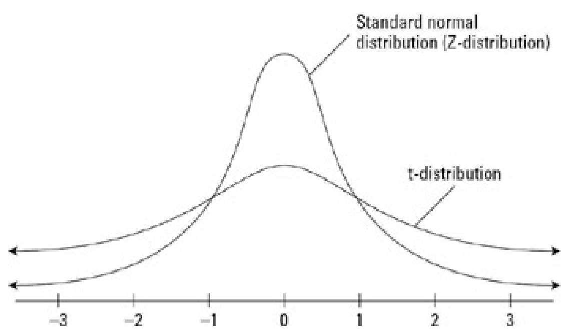
Section 7.4
Confidence Interval for a Population Mean

Recall the formula for a confidence interval is *statistic* $\pm$ *margin of error*. When we are making an inference about a population mean, the statistic will be our sample mean, $\bar{x}$.

The critical value we use to find the margin of error for our calculation will be based on whether the population or sample standard deviation is known. When the population standard deviation is known, we use the formula $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ and when it is unknown, we will need to find the sample standard deviation, *s*, and use the formula $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$, where t^* is the *t*-critical value based on $n - 1$ degrees of freedom.

The value t^* is like z^*, but for a *t*-distribution. Like the standard normal distribution, the *t*-distribution is centered at zero, symmetric and bell-shaped, but has heavier tails, meaning that it is more prone to producing values that fall far from its mean. The shape of it depends on the value of *n*. The higher the *n*, the more it'll look like a standard normal distribution.

t-distribution vs. Standard Normal Distribution



There are actually many different *t*-distributions. The particular form of the *t*-distribution is determined by its **degrees of freedom** (df) – the sample size minus one.

To find critical values for a *z*-distribution, the R command is: $\text{qnorm}\left(\frac{1+\text{CL}}{2}\right)$

To find critical values for a *t*-distribution, the R command is: $\text{qt}\left(\frac{1+\text{CL}}{2}, \text{df}\right)$

The assumptions for a population mean are:
1. The sample must be an SRS from the population of interest.
2. The data must come from a normally distributed population. If this is not the case or if we are unsure whether the population is normally distributed, the sampling distribution of $\bar{x}$ must be normally distributed. (Recall from section 4.4 that we can assume that the sampling distribution of $\bar{x}$ is normal for values of $n \geq 30$.)

Example 1: Suppose your class is investigating the weights of Snickers 1-ounce fun-size candy bars to see if customers are getting full value for their money. Assume that the weights are normally distributed with standard deviation σ = 0.005 ounces (note, this is the population standard deviation because it uses σ). Several candy bars are randomly selected and weighed with sensitive balances borrowed from the physics lab.

The weights are: 0.95 1.02 0.98 0.97 1.05 1.01 0.98 1.00

We want to determine a 90% confidence interval for the true mean, μ.

a. What is the sample mean?

Commands: Answer:

```
> weights=c(0.95,1.02,0.98,0.97,1.05,1.01,0.98,1.00)
```

b. Determine $z*$.

Command: Answer:

c. Determine the 90% confidence interval.

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Confidence Interval:

d. Write a sentence that explains the significance of the confidence interval.

But notice that the population claim, 1 oz, is not even in our confidence interval. This raises a flag! Maybe we need to question the weight or maybe we just need to take a bigger sample. Or maybe if we use a 95% confidence interval then the population claim would be in our interval (because the second number is almost 1).

Example 2: An SRS of 16 seniors from HISD had a mean SAT-math score of 500 and a standard deviation of 100. We know that the population of SAT-math scores for seniors in the district is approximately normally distributed.

$n =$ $\bar{x} =$ $s =$

a. Find the 90% confidence interval for the mean SAT-math score for the population of all seniors in the district. *You must first find t*.*
Command:

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

Confidence Interval:

b. Explain the meaning of the above confidence interval.

The confidence level means that if we repeat this process and take many samples of 16, then 90% of them would contain the true mean.

Like with population proportions, sometimes we are asked to find the minimum sample size needed to produce a particular margin of error given a certain confidence level. When working with a one-sample mean, the population standard deviation needs to be known (or estimated) and we can use the formula Maximum $ME \geq z^* \dfrac{\sigma}{\sqrt{n}}$.

Whenever n is unknown, always use z^*.

Example 3: An evaluator wishes to make a statement about the emotional maturity of the freshman population, so she decides to sample the population and administer an emotional maturity test. Putting aside any concerns about the validity of the test used, and considering sampling techniques only, how many students should she sample in order to be 95% confident that her estimate of freshman emotional maturity will be within 6 units of the true mean? (The test publishers indicate the population variance is 100 units.)

We'll need to apply: $ME \geq z^* \dfrac{\sigma}{\sqrt{n}}$

Careful!!! What is the standard deviation?

Summary of Section 7.2 – 7.4

Confidence Intervals

- These are centered about proportions ($\hat{p}$ or $\hat{p}_1 - \hat{p}_2$) or means (μ or $\bar{x}$).

  When working with proportions, you're given a percentage in the problem, such as, "A company estimates that 82% of households use their Squeaky Clean All Purpose Cleaner." or you're given something like, "In a survey it was found that 68 out of 100 men use Shaving Cream A.". The proportion can be p, $\hat{p}$ or if nothing is given, then 0.5 is used.

- The critical value is based on the confidence level (CL) and is
  $z^* = \text{qnorm}\left(\frac{1 + \text{confidence level}}{2}\right)$ or $t^* = \text{qt}\left(\frac{1 + \text{CL}}{2}, \text{df}\right)$

- *statistic* $\pm$ *margin of error* the margin of error is also called the standard error. The statistic is $\hat{p}$, $\hat{p}_1 - \hat{p}_2$, $\bar{x}$ or $\bar{x}_1 - \bar{x}_2$ and the margin or error is either z^* or t^*.
- If our statistic is $\hat{p}$, then *statistic* $\pm$ *margin of error* = $\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- If our statistic is $\bar{x}$, then *statistic* $\pm$ *margin of error* is either
  $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ (if we know the population standard deviation) or
  $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ (if we know the sample standard deviation).

- In either formula, notice that *n* is in the denominator. So the larger *n* is, the larger the mean and the smaller the *n*, the smaller the mean.

Example 4: The effect of exercise on the amount of lactic acid in the blood was examined in an article for an exercise and sport magazine. Eight males were selected at random from those attending a week-long training camp. Blood lactate levels were measured before and after playing three games of racquetball, as shown in the accompanying table. Use this data to estimate the mean increase in blood lactate level using a 95% confidence interval.

Player	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Before	13	20	17	13	13	16	15	16
After	18	37	40	35	30	20	33	19

Note: We don't know that this is a normal distribution nor do we have n > 30, so we must assume normality.

Since we want to estimate the mean increase, we must first find it.

Player	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mean Increase								

In order to use a 95% confidence interval, we need to find: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

So, let's find each piece:

n = degrees of freedom = df =

Mean:

Commands: Answer:

```
> lactic=c(5,17,23,22,17,4,18,3)
```

Standard Deviation:

Command: Answer:

$$t^* = \text{qt}\left(\frac{1+\text{CL}}{2}, \text{df}\right)$$

Command: Answer:

Hence, $\bar{x} \pm t^* \frac{s}{\sqrt{n}} =$

Confidence Interval: