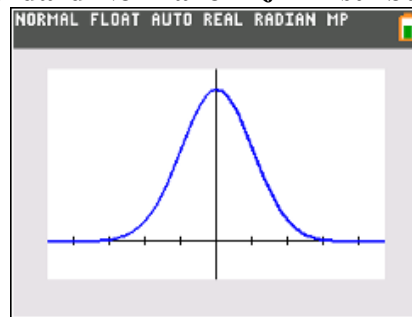


Unit 7: Hypothesis Testing and Confidence Intervals for Means

Introduction: We have done a lot of work with confidence intervals and hypothesis testing for population proportions, we will now be dealing with the same idea with population means, instead. Just like we used proportions when we had categorical data, we use means when we have numerical data. These ideas both build on the ideas discussed in Unit 5 – that is, the idea of the sampling distribution. Using the concept of the sampling distribution, we are going to explore how likely (or unlikely) it is that the mean of an individual sample is different from a population value. This will provide evidence (or not) about claims for hypothesis testing.

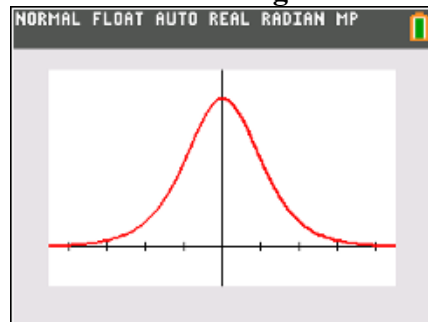
Functionally, this is going to be almost identical to process in the previous unit both for constructing confidence intervals and for hypothesis testing, but with the formulas modified to work for means rather than proportions. In addition, because we have samples that are small (especially compared to large, normally distributed populations) and don't know the population standard deviation, we will be looking at a "*t*-distribution" rather than a *z*-distribution. The key difference between these distributions is that the "tails" of the *t*-distribution have more area under them – this is because with small sample sizes, a single outlier (effectively) adds more "weight" to a tail than it would in a large population.

Standard Normal or "z-" Distribution

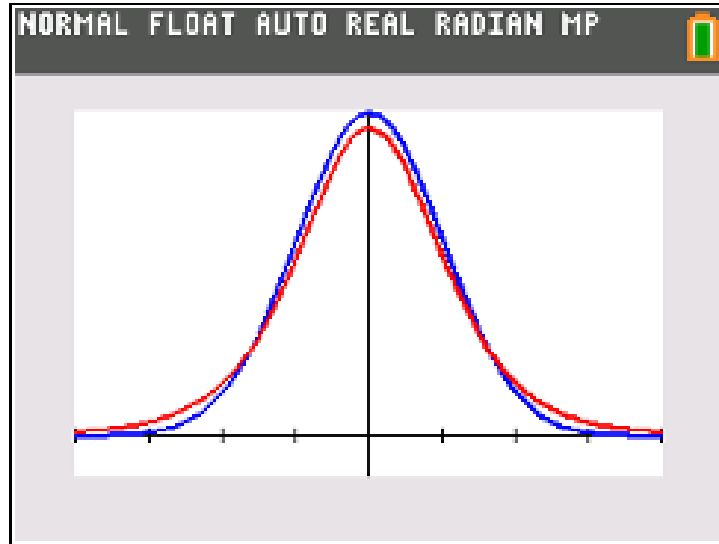


Note that the tails of the curve are essentially collapsed onto the axis by the third standard deviation – the area beneath that curve from 3 to infinity and from negative infinity to 3 is only 0.0027. If you had a normally distributed sample with 1000 individuals in it, you would expect less than three individuals would fall into the regions under both of those tails. With the *t*-distribution, you have a greater chance for extreme values – given small sample sizes, you have to adjust the distribution to attempt to correctly reflect reality.

t-Distribution with 5 degrees of freedom

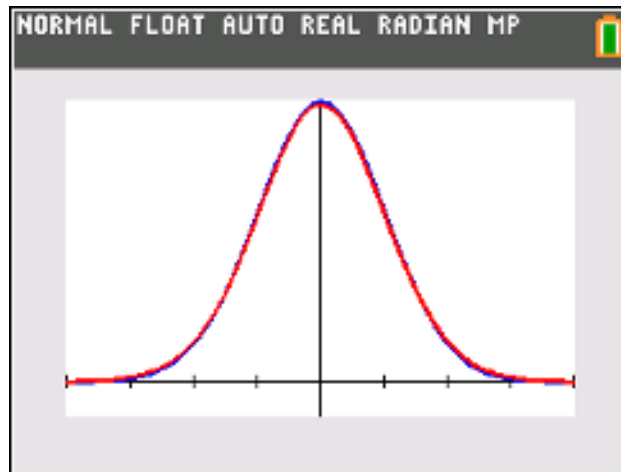


While you may struggle to see the difference between the standard normal curve and a t -distribution with 5 degrees of freedom (we will discuss what this means in section 1), it is critically important in approaching problems with small sample sizes. Below is a graph of both of the distributions (the normal distribution is in blue, and the t -distribution is in red).



Note that since the t -distribution has a lower peak and “fatter” tails, you will have fewer values close to the average and more values at the extremes than you would in a strictly normal distribution.

But as the degrees of freedom increase, the t -distribution approaches the normal distribution. Here is a graph of the normal distribution and a t -distribution with 30 degrees of freedom.



Effectively, it seems like one graph is overlaid directly on top of the other. While this is not strictly true, past 29 degrees of freedom the graphs are practically the same (though we still need the t -distribution if we do not have the population standard deviation – which, in reality, is most often the case).

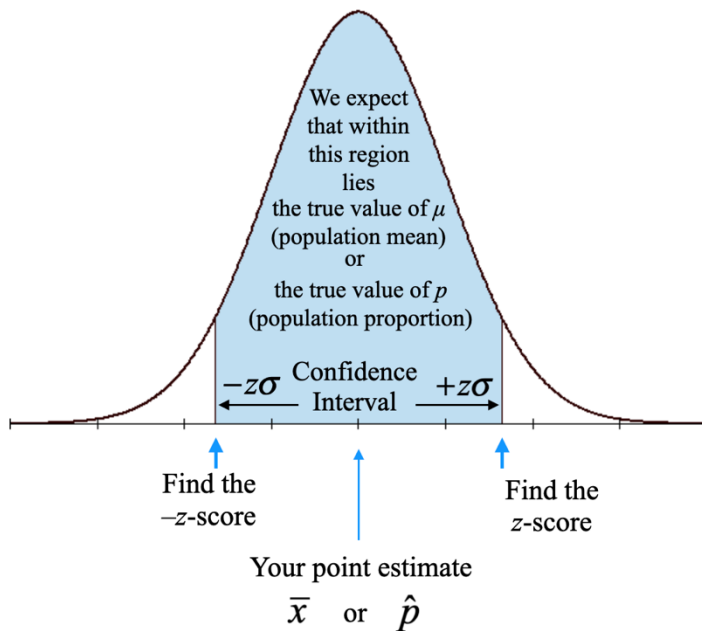
7.1 Confidence Intervals for Population Means

Objectives:

- Construct confidence intervals for population and sample means.
- Correctly interpret confidence intervals and confidence levels.
- Relate margin of error and sample size.
- Construct and interpret confidence intervals for differences of sample means.

Just like in Unit 6, we will be constructing **confidence intervals** for population parameters or sample statistics; however, here we will be exploring population and sample **means** rather than proportions. The process and reasoning are much the same as in the previous chapter, so we do not have to discuss the rationale as much here as we did there. However, you should recall the following:

Remember that unlike Unit 5, in real life we usually do not have the population parameter such as p or μ so we need to sample in order to get our best estimate.



In the example shown to the left, we have plotted a normal distribution with our point estimate as the mean. This is a value that we will have obtained from our sample.

Using z-scores we get from our calculator or a table, we can choose the appropriate size of our confidence interval as a percentage.

If we want an $x\%$ confidence interval, we can then say that we are $x\%$ confident that the true mean or proportion lies within that shaded interval

Here we also need to note the difference between a confidence interval and a confidence level:

- A **confidence interval (CI)** for a population characteristic is an *interval estimate* of plausible values for the characteristic.
- The **confidence level** associated with a confidence interval estimate is the success rate of the *method* used to construct the interval.

Since all of this is based on finding sample proportions and means as we did in the previous chapter, we still have to check our assumptions. Let's review them here and discuss how to apply them to confidence intervals.

We must always check our assumptions in working with population or sample means. These assumptions verify *independence* and *normality*.

Assumptions:

- Data should be collected using a random sample or in a randomized experiment.
 - When sampling without replacement, n must be less than or equal to 10% of the population.
- **Either** of the two following criteria are met:
 - The population distribution is **normal**
 - The sample size is sufficiently large ($n \geq 30$) – this is called the **central limit theorem**.

While we can construct confidence intervals for a population mean, in reality, this is an uncommon occurrence. What happens far more frequently is that we need to use our point estimate (\bar{x}) rather than using the population value (μ , which we will likely not know).

Recall from the previous unit, that our general process for any **confidence interval** is given by the following:

$$(\text{test statistic}) \pm (\text{critical value})(\text{standard deviation})$$

Up until now, the critical value has always been a z -score. However, in the absence of the population standard deviation (σ), we have to use a t value in place of the z -score. A t distribution is given by the following formula:

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

This is a t -distribution with $n - 1$ degrees of freedom.

- **Important:** the number of degrees of freedom is always one less than the sample size.

You might not immediately realize this, but the t -distribution is essentially generated exactly like a z -distribution, but s is used in lieu of σ and there is an additional \sqrt{n} in the denominator.

$$z = \frac{x - \mu}{\sigma}$$

It is interesting to note that the quantity $\frac{s}{\sqrt{n}}$ is known as the Standard Error. If we don't know the population standard deviation, σ , we are effectively using s , the sample standard deviation in place of σ to generate the standard error.

Finally, from all of this, the confidence interval for a sample mean is given below:

$$\bar{x} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

Where t^* is the t critical value associated with $n - 1$ degrees of freedom. This is similar to the z critical value from section 6-1, but instead of the fixed values for z for the 90%, 95%, and 99% confidence intervals, the values are different for each degree of freedom. We can generate these values using the **invT** function on the calculator (similar to the **invNorm** function). We could also use the table provided by the AP test.

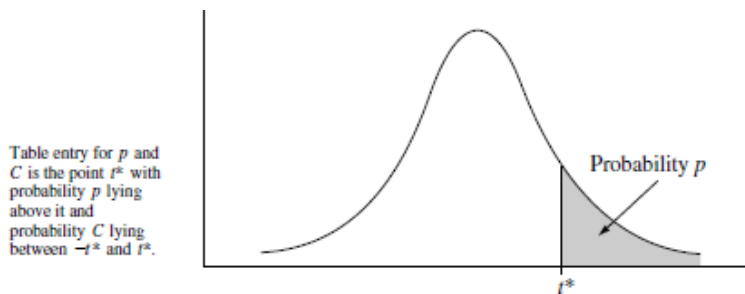


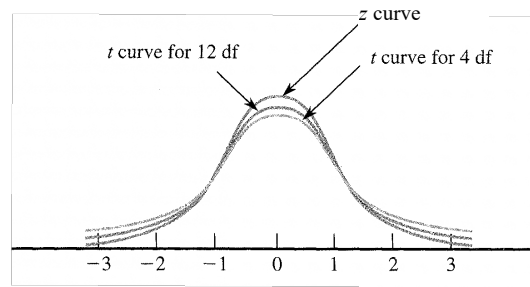
Table B t distribution critical values

| df | Tail probability p | | | | | | | | | | | |
|----------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | .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

Important Properties of t Distributions

1. The t curve corresponding to any fixed number of degrees is bell shaped and is centered at 0 (just like the standard normal (z) curve).
2. Each t curve is more spread out than the z curve.
3. As the number of degrees of freedom increases, the spread of the corresponding t curve decreases.
4. As the number of degrees of freedom increases, the corresponding sequence of t curves approaches the z curve.



Example 1: A study on the ability of individuals to walk in a straight line reported the following data on cadence (strides per second) for a sample of $n = 20$ randomly selected men.

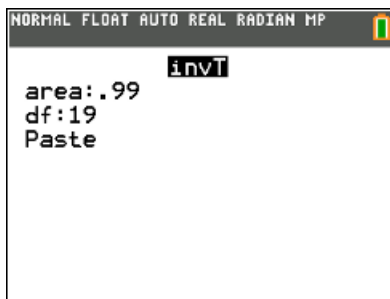
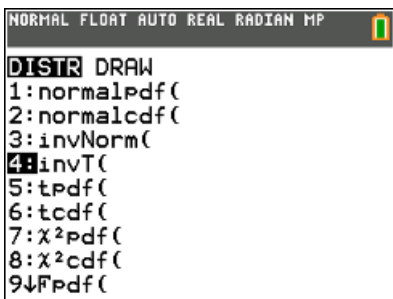
| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 0.95 | 0.85 | 0.92 | 0.95 | 0.93 | 0.86 | 1.00 | 0.92 | 0.85 | 0.81 |
| 0.78 | 0.93 | 0.93 | 1.05 | 0.93 | 1.06 | 1.06 | 0.96 | 0.81 | 0.96 |

Construct a 99% confidence interval. Interpret the interval and level.

Answer: First you have to recognize that we do not have any population values, so we will definitely be using a t distribution. Start by putting the list into your calculator and generating 1-Variable Stats (from the Stats menu on the calculator).

$$n = 20, \text{degrees of freedom} = 19, s = 0.081, \bar{x} = 0.9255$$

Use **invT** or the chart to find the t^* value.



$$\bar{x} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

$$\text{Confidence Interval: } 0.9255 \pm 2.539 \left(\frac{0.081}{\sqrt{20}} \right)$$

Confidence Interval: (0.873,0.977)

(Note: we could use the "TInterval" function in the **stat** Menu, under TESTS on your calculator to find this interval)

Interpretation of Interval:

We are 99% confident that μ , the true mean of cadence, is between 0.873 strides per second and 0.977 strides per second.

Interpretation of Level:

We used a method to construct this estimate that in the long run will successfully capture the true value of μ 99% of the time.

Don't forget that the interpretation follows the same format that it always has:

Interpretation for Large Sample Means Confidence Intervals

"We are _____% confident that μ , the true mean of _____, is between _____ (units) and _____ (units)."

Interpretation for the Confidence Level of a Large Sample Mean Confidence Intervals

"We used a method to construct this estimate that, in the long run, will successfully capture the true value of μ _____% of the time."

Example 2: Based on a sample of workers at Diablo Canyon Nuclear Power Plant, the Diablo's Unit 2 reactor was rated as the 28th lowest in the country (out of 92 reactors). They reported a mean annual radiation exposure of 0.481 rem (roentgen equivalent in man) for a random sample of 100 Unit 2 workers. Let μ denote the true mean radiation exposure for Unit 2 workers at Diablo Canyon Nuclear Power Plant, and assume that $\sigma = 0.35$. Construct a 99% confidence interval for μ .

Answer: We have a σ , so we can use a z distribution, not a t distribution, so we will be using the z value associated with 99% ($z = 2.58$), and we can just use the generic formula (you should memorize this generic formula):

$$\begin{aligned} \text{Confidence Interval} &= (\text{test statistic}) \pm (\text{critical value})(\text{standard deviation}) \\ &= \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \\ &= 0.481 \pm 2.58 \left(\frac{0.35}{\sqrt{100}} \right) \end{aligned}$$

Confidence Interval: (0.391,0.571)

Example 3: Suppose the actual mean radiation exposure is the upper limit of your 99% confidence interval is the true population mean (that is, $\mu = 0.571$ rem). If the safety limit allowed for nuclear plant workers in the U.S. is 5.000 rem per year¹, what is the probability that a sample with this mean has an exposure that exceeds the safety limit?

Answer: This is not a confidence interval problem; this is a straightforward normal distribution interpretation - you could calculate a z -score and look it up, or you could just use the normalcdf function on your calculator:

$$\text{normalcdf}(\text{lower bound} = 5.000, \text{upper bound} = 1E99, \mu = 0.571, \sigma = 0.35)$$

$$P(x > 5.000) = 5.48 \times 10^{-37} \approx 0$$

¹Note: It is interesting to note that even the "lowest ranked" nuclear reactors in the U.S. are incredibly safe and have very low relative radiation exposure - by way of comparison, an average airline pilot is exposed to 0.307 rem - and that industry has the highest mean radiation exposure (from cosmic rays from outside the atmosphere). The average exposure to radiation for every person from natural sources is 0.033 rem.

Example 4: The average unweighted GPA of a randomly selected group of 18 AP Stats students at SI (out of 153 total AP Stats students). Construct a 90% confidence interval for μ . Interpret the interval and level.

Again, much of statistics deals with comparisons of data, so it is frequent that we will be comparing means to one another (and most often this comparison is a difference of means - we usually want to know how far apart two samples or two populations might be). Recall from Unit 5 the formulas for differences of means:

If two random variables X and Y are independent and have distributions that are approximately normal then we can say that $\mu_{X \pm Y} = \mu_X \pm \mu_Y$ and their sampling distributions would have this relationship:

$$\mu_{\bar{X} - \bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}}$$

The standard deviations are not as simple however.

For any two **independent** random samples: $\sigma_{X \pm Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$ Note that we **always** add variances.

For the difference in sample means of any two **independent** random samples we have

$$\sigma_{\bar{X} - \bar{Y}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Of course, we are dealing with samples and t distributions, (so we do not know the true population standard deviation, σ), so we will make a substitution of s for σ (that is the sample statistic for the population parameter), and we will come up with the formulas on the following page.

Assuming that our two distributions are x_1 and x_2 , the formulas are as follows:

Difference of Means

$$\mu_{\bar{x}_1 - \bar{x}_2} = \bar{x}_1 - \bar{x}_2$$

Standard Error

$$S.E._{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

Of course, our confidence interval is always generated by the same formula:

$$\text{Confidence Interval} = (\text{test statistic}) \pm (\text{critical value})(\text{standard error})$$

$$C.I. = \bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

It is useful to note that while the formulas change, they follow the same basic template - differences of means may be more complicated to plug in the numbers, but it is still just test statistic +/- the margin of error (critical value time standard error).

Degrees of Freedom can be difficult to calculate with different samples - there is a long formula to generate degrees of freedom for samples of mixed (or even the same) sizes - your calculator will generate this for you, and you will never be asked to do this by hand.

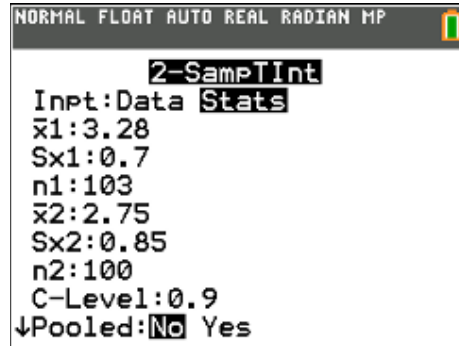
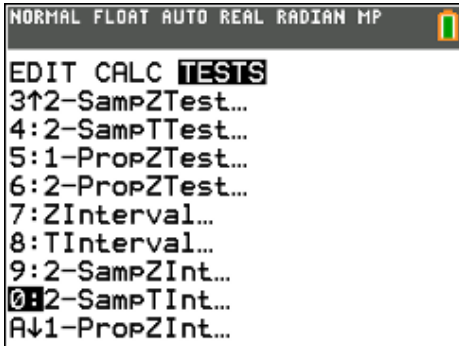
Example 5: First year nursing students enrolled in a science course were randomly assigned to one of two groups. Students in the control group were provided with conventional study-skill guidelines. Students in the treatment group were provided with homework prescriptions based on their identified learning style as well as with the conventional study-skill guidelines. Construct a 90% confidence interval and interpret the interval and level.

| Group | Sample Size | Sample Mean | Sample SD |
|-----------|-------------|-------------|-----------|
| Control | 100 | 2.75 | 0.85 |
| Treatment | 103 | 3.28 | 0.70 |

Answer: Since we only have samples and sample standard deviation, we will be using the t -distribution. By convention, we will usually use the larger mean as \bar{x}_1 .

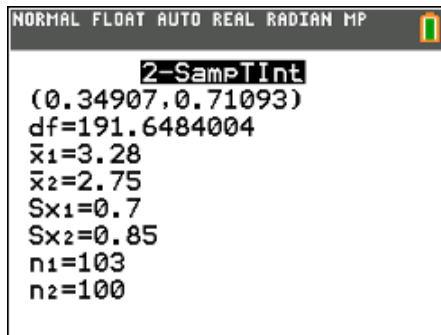
$$\text{Confidence Interval} = \bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

To determine t^* , we have to use the calculator – again, go to the **stat** Menu, under TESTS on your calculator; we will be using a 2-SampTInt (2 sample t interval).



Always leave “pooled” set to NO!

Then simply scroll down and click on **calculate**:



Note that we have 191.65 degrees of freedom.

Confidence Interval: (0.349, 0.711)

Interpretation of Interval:

We are 90% confident that $\mu_1 - \mu_2$, the true mean of the difference in GPA, is between 0.349 and 0.711 grade points.

Interpretation of Level:

We used a method to construct this estimate that in the long run will successfully capture the true value of $\mu_1 - \mu_2$ 90% of the time.

Example 6: Does talking elevate blood pressure, contributing to the tendency for blood pressure to be higher when measured in a doctor's office than when measured in a less stressful environment? Patients with high blood pressure were randomly assigned to one of two groups. Those in the first group (the talking group) were asked questions about their medical history and the sources of stress in their life in the minutes before their blood pressure was measured. Those in the second group (the counting group) were asked to count aloud from 1 to 100 four times before their blood pressure was measured. The following data values for diastolic blood pressure (in mm Hg) are consistent with the summary quantities appearing in the paper:

| | | | | | | | | |
|----------|-----------|-----|----------------------|-----|-----|--------------|-----|-----|
| Talking | 104 | 110 | 107 | 112 | 108 | 103 | 108 | 118 |
| | $n_1 = 8$ | | $\bar{x}_1 = 108.75$ | | | $s_1 = 4.74$ | | |
| Counting | 110 | 96 | 103 | 98 | 100 | 109 | 97 | 105 |
| | $n_2 = 8$ | | $\bar{x}_2 = 102.25$ | | | $s_2 = 5.39$ | | |

Use this to construct a 95% confidence interval for $\mu_1 - \mu_2$.

Margin of Error

Margin of Error is exactly what it was in previous sections - when you hear this term, think of the end piece of your confidence interval, the part you "plus or minus". That is essentially how far off you may be from the true value. (Note we wrote it with standard error in context of this situation, but it is standard deviation for the z distribution, standard error for the t distribution.)

$$\text{Margin of Error} = (\text{critical value})(\text{standard error})$$

$$= t^* \left(\frac{s}{\sqrt{n}} \right)$$

Any time you are asked for margin of error, simply calculate the +/- part of the formula. If you are given a margin of error, you simply solve for n , the sample size (exactly as we did in Unit 5). Note that the margin of error formula is for a single sample – we typically will not be asked a margin of error question for difference of sample means.

Summary:

- Use a t -interval if you **do not know** population standard deviation. Use a z -interval if you know the population standard deviation.
 - Standard Error is generated using a sample standard deviation, s .
 - A t -distribution always has $n - 1$ **degrees of freedom**.

- Generally, **confidence intervals** are generated as follows:

$$\text{Confidence Interval} = (\text{test statistic}) \pm (\text{critical value})(\text{standard error})$$

- Interpret the interval and level using the following phrases:
 - **Interval:** "We are _____% confident that μ , the true mean of _____, is between _____ (units) and _____ (units)."
 - **Level:** "We used a method to construct this estimate that, in the long run, will successfully capture the true value of μ _____% of the time."
- **Margin of Error** is simply the "+/-" portion of the confidence interval.
 - If you are given a margin of error, you are using that to solve for n .

7.1 Checkpoint

Multiple Choice:

- 1) Suppose we compare the data on two independent and random samples, A and B , and come up with the following data: $n_A = n_B = 10$, $\bar{x}_A = 25$, $s_A = 3.21$, $\bar{x}_B = 22.2$, $s_B = 3.09$. Find a 99% confidence interval for the difference between the means. Assume the samples are independent.
 - (a) (-1.26, 6.86)
 - (b) (-1.78, 7.38)
 - (c) (-1.67, 7.27)
 - (d) (-0.83, 6.43)
 - (e) None of the above

- 2) A random sample of 32 games is chosen for a professional basketball team, team A, and their results are recorded. The team averages 88 points per game with a standard deviation of 8. The same is done for a second team, team B, with this team averaging 90 points per game with a standard deviation of 6. A 95% confidence interval is constructed for the difference in points scored per game between the two teams. What do the results of the confidence interval show?
 - (a) We can be 95% confident that on average team A scores between 1.54 and 5.54 more points per game than team B.
 - (b) We can be 95% confident that on average team A scores between 1.54 and 5.54 fewer points per game than team B.
 - (c) We can be 95% confident that on average team A scores between 1.54 points fewer than and 5.54 points more than team B.
 - (d) We can be 95% confident that on average team A scores between 5.54 points fewer than and 1.54 points more than team B.
 - (e) The conditions necessary to find a 95% confidence interval have not been met.

7.1 Homework:

- 1) A study collected the mean resting pulse rate in beats per minute of 20 adult sedentary males. The group was randomly selected.

95 85 94 91 74 86 90 102 85 82
78 93 83 95 73 106 86 86 81 96

Construct a 99% confidence interval. Interpret the interval and level.

- 2) A random sample of 26 high school track athletes found that their mean 100 meter dash time was 11.72 seconds with a standard deviation of 0.42 seconds. Assuming a normal distribution of times in the overall population of high school track athletes, construct a 95% confidence intervals for the mean 100 meter dash time. Interpret the interval and level.
- 3) Senior high school students taking the SAT were randomly assigned to one of two groups. Students in the control group were provided with conventional study-skill guidelines for the SAT. Students in the treatment group were provided with private tutoring sessions as well as with the conventional study-skill guidelines. The data is displayed in the table below. Construct a 90% confidence interval for the difference of the means and interpret the interval and level.

| Group | Sample Size | Sample Mean | Sample SD |
|-----------|-------------|-------------|-----------|
| Control | 68 | 1046 | 104 |
| Treatment | 72 | 1098 | 92 |

- 4) A random sample of 60 high school student finds a mean GPA of 2.7 with a standard deviation of 0.45. What is the margin of error for a 95% confidence interval?

7.2 Hypothesis Testing for Population Means

Objectives:

- Conduct and interpret a hypothesis test for a population mean.
- Conduct and interpret a hypothesis test for a difference of population means.

This section revisits the steps and conclusions of **Hypothesis Testing**. This process closely follows the procedure for hypothesis testing for proportions, but with different test statistics (and with the potential for degrees of freedom).

We have two basic cases, the first, where we know σ , the population standard deviation. The second is when we do not know the population standard deviation (this second case is much more common in the real world application of statistics).

Case 1: σ is known. Use the z -test

$$\text{Test Statistic: } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\mu =$ hypothesized value

Case 2: σ is unknown. Use the t -test
(we know s rather than σ)

$$\text{Test Statistic: } t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$\mu =$ hypothesized value

The P -value for each test will be based on either the z -score or the t -score, as appropriate. It is much more likely that we will be dealing with **Case 2** rather than **Case 1**.

We still have to test our assumptions for both cases, and they are the same as the assumptions for sample means from chapter 5.

Assumption (only with difference of means): The two samples are **independent** of each other.

Assumption #1: When the population distribution is normal, the sampling distribution of \bar{x} is also normal for any sample size n .

Assumption #2 (Central Limit Theorem): When n is sufficiently large, the sampling distribution of \bar{x} is well approximated by a normal curve, even when the population distribution is not itself normal.

Review: STEPS IN HYPOTHESIS TESTING

1. Define the population characteristic (i.e. parameter) about which hypotheses are to be tested.
2. State the null hypothesis, H_0 .
3. State the alternative hypothesis, H_a .
4. State the significance level α for the test.
5. Check all assumptions.
6. State the name of the test.
7. *State degrees of freedom if applicable (only applicable with t-tests, $df = n - 1$).*
8. Display the test statistic to be used without any computation at this point.

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

9. Compute the value of the test statistic, showing specific numbers used.
10. Calculate the P – value.
11. Sketch a picture of the situation.
12. State the conclusion in two sentences -
 - A. Summarize in theory discussing H_0 . Always start by stating the P – value compared to the significance level, α , of the test
 - If the P – value is **less than** α , then we **reject the null hypothesis (H_0)** at the significance level we tested.
 - If the P – value is **greater than** α , then we **fail to reject the null hypothesis (H_0)** at the significance level we tested.
 - B. Summarize in context discussing H_a .
 - **If we reject H_0** state that “we have evidence that the mean of _____ is ..., therefore, the (*initial claim*) is incorrect.”
 - **If we fail to reject H_0** state that “we have insufficient evidence that the mean of _____ is ..., therefore, we cannot reject the (*initial claim*).”

YOU MUST MEMORIZE THESE 12 STEPS!!!

Important Note: We NEVER accept the null hypothesis, we either reject or fail to reject it.

Example 1: One concern employers have about the use of technology is the amount of time that employees spend each day making personal use of company technology, such as personal phone calls, non-business related e-mail, Internet use, and computer games. The AP reported that a management consultant believes that, on average, workers spend 75 minutes a day making personal use of company technology and that this population is normally distributed. Suppose that the CEO of a large corporation wants to determine whether the average amount of time spent on personal use of company technology for her employees is greater than the reported value of 75 min. Each person in a random sample of 10 employees was contacted and asked about daily personal use of company technology. The resulting data are as follows:

66 70 75 88 69 89 71 63 86 71

Do these data provide evidence that the mean for this company is greater than 75 min.? Carry out a hypothesis test with $\alpha = 0.05$

Answer:

μ = mean time employees spend making personal use of company technology

$$H_0: \mu = 75 \text{ minutes} \quad H_a: \mu > 75 \text{ minutes} \quad \alpha = 0.05$$

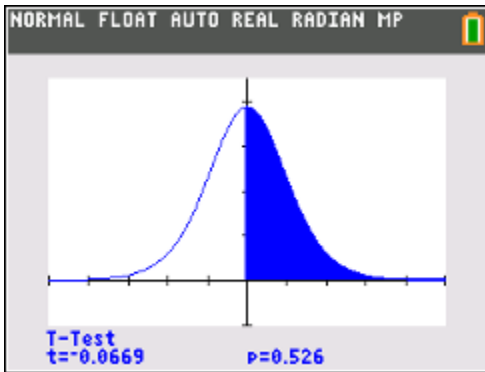
Assumption: population is normally distributed

Single sample t -test (since we do not know the population standard deviation)

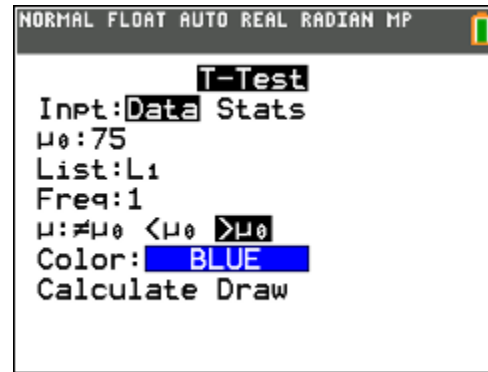
$$df = 9 \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad t = \frac{74.8 - 75}{9.449/\sqrt{10}} = -0.067$$

We got these numbers from plugging in a list and running 1-Var Stats on the calculator.

$$P(t > -0.067) = 0.527$$



You could run this as a T -test on the calculator:



Since $P > 0.05$, we fail to reject the null hypothesis at the 0.05 significance level.

We have insufficient evidence that the mean amount of time spent using company technology for personal use is greater than 75 minutes, therefore we cannot reject the initial claim of an average of 75 minutes of time spent using company technology for personal use.

Note that since this was a single sample t -test, we just used the “T-test” function on the calculator (Stats Menu, tests, option 2). You have a choice between “Data” and “Stats” for this test – I chose data because I had entered the list of data into the calculator, and it will simply import the data from the list – if I had the statistics given to me (or I had calculated them in another way), I would have used the “Stats” choice and typed them in manually.

Example 2: An article reported that female crickets are attracted to males that have high chirp rates and hypothesized that chirp rate is related to nutritional status. The usual chirp rate for male field crickets was reported to vary around a mean of 60 chirps per second. To investigate whether chirp rate was related to nutritional status, investigators fed male crickets a high protein diet for 8 days, after which chirp rate was measured. The mean chirp rate for the crickets on the high protein diet was reported to be 109 chirps per second. Is this convincing evidence that the mean chirp rate for crickets on a high protein diet is greater than 60 (which would then imply an advantage in attracting the ladies)? Suppose that the sample size and sample standard deviation are $n = 32$ and $s = 40$. We test the relevant hypotheses with $\alpha = 0.01$.

Answer:

μ = mean chirp rate for male field crickets

$H_0: \mu = 60$ chirps per minute $H_a: \mu > 60$ chirps per minute $\alpha = 0.01$

Assumption: Since the sample size is greater than 30 ($n = 32$) we can assume our sampling distribution is normally distributed by the Central Limit Theorem

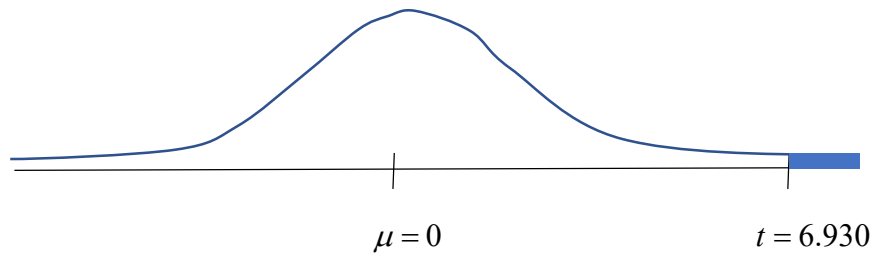
Single sample t -test (since we do not know the population standard deviation)

$$df = 31 \quad t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$$t = \frac{109 - 60}{40 / \sqrt{32}} = 6.930$$

We were given these numbers:
 $\bar{x} = 109, s = 40$

$$P(t > 6.930) = 4.5 \times 10^{-8}$$



Since $P < 0.01$, we reject the null hypothesis at the 0.01 significance level.

We have convincing evidence that the mean rate of chirping for male crickets with a high protein diet is higher than the mean rate of chirping for male crickets generally, therefore, we can reject the initial claim that they are the same.

If this had been done as an experiment, where we collected data from male crickets on a control diet and male crickets on a high protein diet, then we would have been doing a 2-sample z -test. Since we were comparing a random sample of crickets with a treatment to a known population parameter – that’s how you could tell if you had two samples or just one.

Remember, for difference of two samples, simply subtract the means, but for the standard deviation, we have to add the variances and take the square root. This gives us the following null hypothesis and test statistic:

Two-Sample Means Hypothesis Tests:

$$H_0: \mu_1 - \mu_2 = \text{hypothesized value (often 0)} \quad \text{Test Statistic: } t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Degrees of Freedom are complicated to calculate and the AP test will not make us do that by hand – we can use the calculator test to generate the *degrees of freedom* for us.

Example 3: You design and run an experiment comparing the rate of chirping for male field crickets on a control diet vs. the rate of chirping for male field crickets. You collect simple random samples of field crickets, apply the diets and collect the data, which is displayed on the table below:

| Group | n | \bar{x} | s |
|--------------|-----|-----------|-----|
| Control | 44 | 63 | 21 |
| High Protein | 46 | 98 | 32 |

Does this study provide convincing evidence at the 0.01 significance level that the mean chirp rate for male field crickets is higher on a high-protein diet?

Answer:

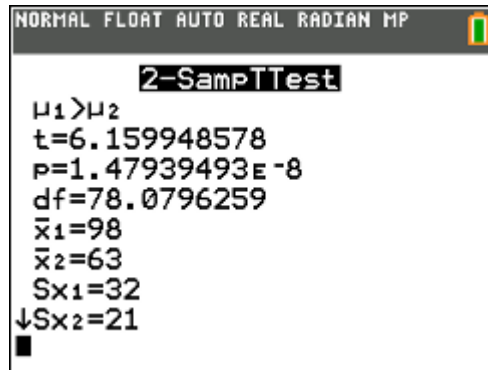
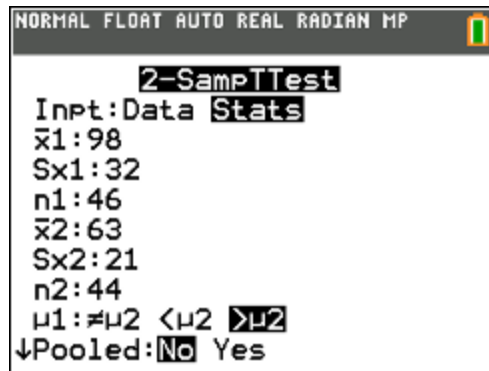
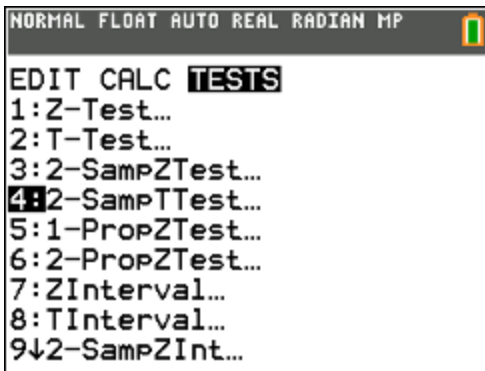
μ_1 = mean chirp rate for male field crickets on a high protein diet

μ_2 = mean chirp rate for male field crickets on the control diet

$$H_0: \mu_1 = \mu_2 \text{ (or } \mu_1 - \mu_2 = 0) \quad H_a: \mu_1 > \mu_2 \quad \alpha = 0.01$$

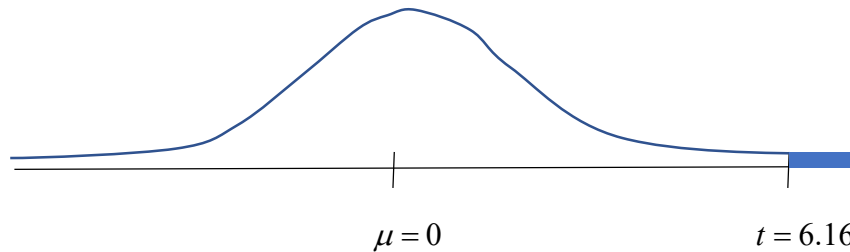
Assumptions: The two samples are independently selected random samples. Since the sample sizes are greater than 30 ($n_1 = 44$ and $n_2 = 46$) we can assume our sampling distribution is normally distributed by the Central Limit Theorem.

Single sample t -test (since we do not know the population standard deviation)



$$df = 78.1 \quad t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} \quad t = \frac{98 - 63 - (0)}{\sqrt{\frac{(32)^2}{46} + \frac{(21)^2}{44}}} = 6.16$$

$$P(t > 6.16) = 1.48 \times 10^{-8}$$



Since $P < 0.01$, we reject the null hypothesis at the 0.01 significance level.

We have convincing evidence that the mean rate of chirping for male crickets with a high protein diet is higher than the mean rate of chirping for male crickets generally, therefore, we can reject the initial claim that they are the same.

Example 4: To assess the impact of oral contraceptive use on bone mineral density (BMD), researchers in Canada carried out a study comparing bone mineral density for women who had used oral contraceptives for at least 3 months to bone mineral density for women who had never used oral contraceptives. Data on bone mineral density (in grams per cubic centimeter) is given below.

| | | | | | |
|---------------------------------|------|------|------|------|------|
| Never used oral contraceptives: | 0.82 | 0.94 | 0.96 | 1.31 | 0.94 |
| | 1.21 | 1.26 | 1.09 | 1.13 | 1.14 |
| Used oral contraceptives: | 0.94 | 1.09 | 0.97 | 0.98 | 1.14 |
| | 0.85 | 1.30 | 0.89 | 0.87 | 1.01 |

The authors of the paper believed that it was reasonable to view the samples used in the study as randomly selected representatives of the two populations of interest (which are both assumed to be normally distributed). Use the given information and a significance level of 0.05 to determine whether there is evidence that women who use oral contraceptives have a lower BMD than women who have never used oral contraceptives.

Summary:

- Use a t -test if you **do not know** population standard deviation. Use a z -test if you know the population standard deviation.
 - Standard Error is generated using a sample standard deviation, s .
 - A t -distribution always has $n - 1$ *degrees of freedom*.
- Memorize the 12 step process for hypothesis testing – you must do this regardless of what type of hypothesis test you are doing (1-sample, 2-sample, proportion, mean)

7.2 Checkpoint

- 1) An automotive company executive claims that a mean of 48.3 cars per dealership are being sold each month. A major stockholder believes this claim is high and runs a test by sampling 30 dealerships. What conclusion is reached if the sample mean is 45.4 cars with a standard deviation of 15.4?
 - (a) There is sufficient evidence to support the executive's claim.
 - (b) There is sufficient evidence to refute the executive's claim.
 - (c) The stockholder has sufficient evidence to reject the executive's claim.
 - (d) The stockholder does not have sufficient evidence to reject the executive's claim.
 - (e) There is not sufficient data to reach any conclusion.

- 2) The Acme Whoopee Cushion Company claims that each of its unbreakable whoopee cushions consists of 18 oz. of rubber. As a quality control engineer you want to know if the population mean is consistent with the claim of 18 oz. Which of the following would be appropriate for testing the company's claim?
 - I. A significance test with a one-sided alternative hypothesis
 - II. A significance test with a two-sided alternative hypothesis
 - III. A confidence interval

(a) I only (b) II only (c) III only (d) I and III (e) II and III

- 3) Given a hypothesized population mean $\mu = 58.4$, a set of sample data (with all three assumptions met) gives a P -value of 0.0154 if the alternate hypothesis is $\mu < 58.4$. Which of these statements is true about our hypothesis test if we use a significance level $\alpha = 0.02$?
 - I. We reject the null hypothesis if the alternate hypothesis is $\mu < 58.4$
 - II. We fail to reject the null hypothesis if the alternate hypothesis is $\mu \neq 58.4$
 - III. We reject the null hypothesis if the alternate hypothesis is $\mu > 58.4$

(a) I only
(b) I and II .
(c) I and III.
(d) II and III
(e) III only

- 4) Two professors, A and B, got into an argument about who grades tougher. Professor A insisted that his grades were lower than those for Professor B. In order to test this theory, each professor took a random sample of 25 student grades and conducted a test of significance. The graphical displays showed each grade distribution was approximately normal. The results are recorded below.

H_0 : Population mean of Professor A *equals that* of Professor B

H_a : Population mean of Professor A *less than that* of Professor B

| | Professor A | Professor B |
|------------|--------------------|--------------------|
| Count: | 25 | 25 |
| Mean: | 79 | 82 |
| Std dev: | 6 | 4 |
| Std error: | 1.2 | 0.8 |

Student's $t = -2.08$ $DF = 41.844$ $P\text{-value} = 0.022$

Which of the following conclusions is/are supported by the results of the significance test?

- I. At the 0.05 level, we have evidence to show that every student in Professor A's class scored lower than every student in Professor B's class.
- II. If there were no difference in grades between the two professors, then we could get results as extreme as those from the samples approximately 2.2% of the time.
- III. The test results are not valid, since the conditions necessary to perform the test were not met.

- (a) I only
- (b) II only
- (c) III only
- (d) I and II
- (e) II and III

7.2 Homework

- 1) A researcher is conducting a study on the health of a population of red-eared sliders (a small pond turtle). She knows that an average adult red-eared slider has a mean shell diameter of 16 cm, and that this is a normally distributed population. She takes a sample of 17 adult red-eared sliders from an ecosystem and finds the following:

$$\bar{x} = 16.7 \text{ cm}, s = 1.4 \text{ cm}$$

Perform a hypothesis test at the 0.10 significance level to determine if this sample is larger than average.

Would the result be different if we tested at the 0.05 significance level? Explain.

- 2) To assess the impact of oral contraceptive use on bone mineral density (BMD), researchers in Canada carried out a study comparing bone mineral density for women who had used oral contraceptives for at least 3 months to bone mineral density for women who had never used oral contraceptives. Data on bone mineral density (in grams per cubic centimeter) is given below.

Never used oral contraceptives: 0.82 0.94 0.96 1.31 0.94
 1.21 1.26 1.09 1.13 1.14

Used oral contraceptives: 0.94 1.09 0.97 0.98 1.14
 0.85 1.30 0.89 0.87 1.01

The authors of the paper believed that it was reasonable to view the samples used in the study as randomly selected representatives of the two populations of interest (which are both assumed to be normally distributed). Use the given information and a significance level of 0.05 to determine whether there is evidence that women who use oral contraceptives have a lower BMD than women who have never used oral contraceptives.

- 3) First year nursing students enrolled in a science course were randomly assigned to one of two groups. Students in the control group were provided with conventional study–skill guidelines. Students in the treatment group were provided with homework prescriptions based on their identified learning style as well as with the conventional study–skill guidelines. Does the treatment (the homework prescriptions based on learning style) improve science GPA? We will test the relevant hypothesis using a significance level of 0.01.

| Group | Sample Size | Sample Mean | Sample SD |
|-----------|-------------|-------------|-----------|
| Control | 100 | 2.75 | 0.85 |
| Treatment | 103 | 3.28 | 0.70 |