Start on 3/9

Estimating a Population Mean

How can we construct a confidence interval for an unknown population mean μ when we don't know the population standard deviation σ? In the previous section, we made the unrealistic assumption that we knew the value of σ. In practice, σ is unknown. We must estimate σ from the data even though we are primarily interested in μ. The need to estimate σ changes some computational details of confidence intervals for μ, but not their interpretation.

As before, we need to verify three important conditions before we estimate a population mean. When we do inference in practice, verifying the conditions is often a bit more complicated.

Conditions for Inference about a Population Mean:

1. SRS: is necessary for us to be able to generalize our results

2. Normality: use sample size and CLT or if the population distribution is Normal, also look at graphs of the sample data to assess normality

3. Independence: $N \geq 10n$

The previous section taught you how to find confidence intervals. The problem that we never addressed in that section is that all the problems were based on the unrealistic assumption that we knew the population standard deviation σ. Our knowing that value is highly improbable and yet, knowing it allowed us to use *z*-confidence intervals. In this section, we are much more realistic and do not depend on σ. We are still going to sample via an SRS from a population with meanμ and standard deviation σ.
We'll perform inferential procedures on the mean without knowing the population standard deviation. However, to do so means we need to learn a new type of distribution.

Since we don't know the population standard deviation, we can no longer rely on the normal distribution for our calculations. Thankfully, William Sealy Gosset solved that problem for us...

William Sealy Gosset - The Student's t-Distribution

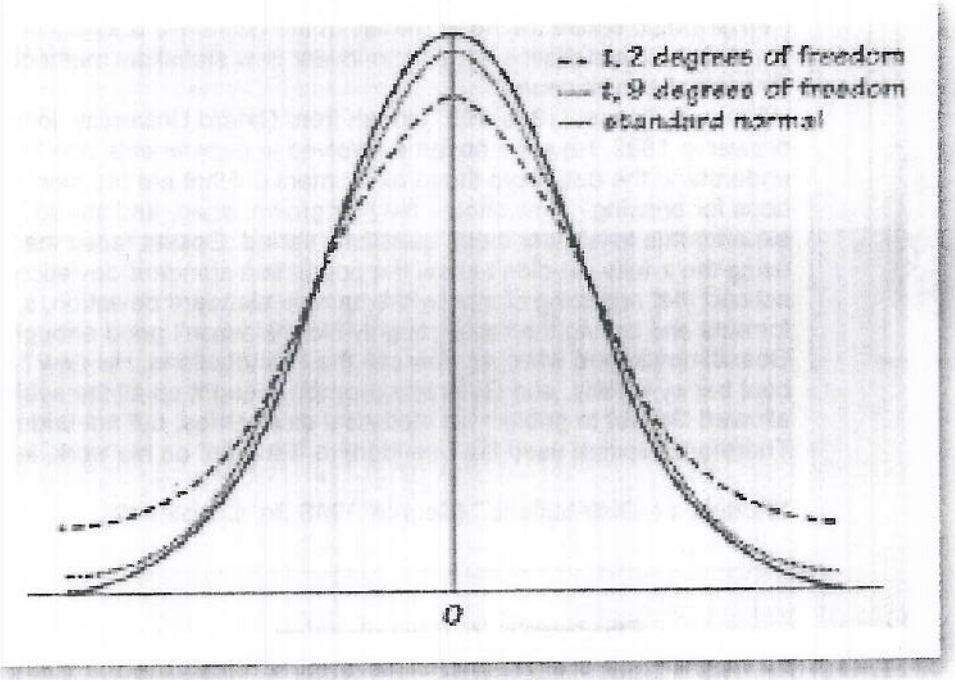
" What would cause the head brewer of the Guinness brewery in Dublin, Ireland, to not only use statistics but also to invent new statistical methods? Why...the search for better beer, of course!
William S. Gosset (1876-1937), fresh from Oxford University, joined Guinness as a brewer in 1899. He soon became involved in experiments and in statistics to understand the data from those experiments. What are the best varieties of barley and hops for brewing? How should they be grown, dried, and stored? The results of field experiments to answer these questions varied. Gosset faced the problem we noted in using the z-test--he didn't know the population standard deviation. . Further, he noticed that replacing sigma by the sample standard deviation, *s*, in the z-score formula and calling the result roughly Normal wasn't good enough. After much work, Gosset developed what we now call the *t*-distributions. His new *t* test identified the best barley variety, and Guinness promptly bought up all the available seed. Guinness allowed Gosset to publish his statistical discoveries, but not under his own name. Therefore, Gosset used the pseudonym "Student" on his work, leading to the name "

"Student's *t*- Distributions." {Source: YMS 3e, Chapter 12}

Student's *t*-Distribution

In probability and statistics, the t-distribution (or Student's *t*-distribution) is a probability distribution that arises in the problem of estimating the mean of a normally distributed population when the sample size is small. Student's t distribution arises when the population standard deviation is unknown and has to be estimated from the data--as is the case in nearly all practical statistical work. If the population standard deviation is unknown, we must estimate it using the sample standard deviation, *s*. The value of *s* may not be close to sigma--especially if *n* is small. Further, *s* may vary from sample to sample. As a result, the use of s in place of sigma introduces extra variability into our problem. Due to the extra variability, the *t*-distribution is more spread out than the normal (*z*) distribution.

The *t*-distribution looks very much like the normal distribution (for which we use *z*) that follows the 68-95-99.7 rule. However, since we do not know σ, we had to estimate it by using *n*. The larger *n* is, the closer our *t* distribution looks like the *z* (normal) distribution. We describe this with $n - 1$ degrees of freedom which is abbreviated as (df). The picture below shows 3 curves. The highest curve is the standard normal (z) distribution. The innermost curve is the t distribution with 2 degrees of freedom and the middle curve is the t-distribution with 9 degrees of freedom. Note that the t-distribution shapes are symmetric but have greater probability in the tails than the normal distribution. Also note that has the number of degrees of freedom increases; the t distribution approaches the normal distribution. This happens because s estimates σ more accurately as the sample size n increases. So using s in place of σ causes little problem when the sample is large.



Since we don't know σ, we are going to estimate it. When we sample from a population, we will get a mean $\bar{x}$ and a standard deviation s. We estimate the standard deviation of $\bar{x}$ by using $s/\sqrt{n}$

Standard Error:

When the standard deviation of a statistic is estimated from the data, the result is called the standard error, SE, of the statistic.

$SE_{x\text{-bar}} = s/\sqrt{n}$

When we knew the value of σ. we found our z-statistic by the formula $z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$. Now that we don't know σ and are going to estimate it using our standard error $\dfrac{s}{\sqrt{n}}$, we create a new statistic called the t-statistic.

Draw an SRS of size n from a population with unknown μ and standard deviation σ.

The **one-sample t-statistic** is defined as $t = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$.

The t-distribution has $n - 1$ degrees of freedom (which we will discuss shortly).

The good news is that the procedures for constructing a CI are the same for a t statistic as they are for a z statistic. So, rather than going through our procedures for confidence intervals using *t* rather than *z*, I am going to lay it all out for you on this page and then we will do a number of examples illustrating them.

However, remember, for right now, you decide on *z* vs. *t* based on one criterion: are you given σ?

If you are not (which is usual), you use the *t*-procedures outlined below. (Inference Toolbox!!!)

Draw an SRS of size *n* from a population with unknown mean; find $\bar{x}$ and s for the sample.

To find the level C confidence interval, use the formula:

$\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right)$ with $n-1$ degrees of freedom (df). $t^*\frac{s}{\sqrt{n}}$ is your margin of error.

To find t^*, use the *t*-distribution table at the back of the book. Go to the row for your degrees of freedom (df) … that is $n-1$. Now, move to the column for your desired confidence level C – those numbers are at the bottom. The intersection of the correct row and column is your t^*. Note that if n is above 30, the rows skip 10. It won't matter much whether you use the lower or upper row as these values are very close.

use lower df - will widen CI & err on side of caution!

Example 14: What critical value t^* from the table should be used for a confidence interval for the mean of the population in each of the following situations?

(a) A 90% confidence interval based on $n = 12$ observations.

INVT (% under t^*, df)

$t^* = 1.796$

(b) A 95% confidence interval from an SRS of 30 observations.

$t^* = 2.045$

(c) An 80% confidence interval from a sample of size 18.

$t^* = 1.333$

Example 15: In 1996, the U.S. Agency for International Development provided 238,300 metric tons of corn-soy blend (CSB) for development programs and emergency relief in countries throughout the world. CSB is a highly nutritious, low-cost fortified food that is partially precooked and can be incorporated into different food preparations by the recipients. As part of a study to evaluate appropriate vitamin C levels in this commodity, measurements were taken on samples of CSB produced in a factory.[7] The following data are the amounts of vitamin C, measured in milligrams per 100 grams (mg/100 g) of blend (dry basis), for a random sample of size 8 from a production run:

26	31	23	22	11	22	14	31

(a) What conditions must be satisfied in order to make inferences about μ, the mean vitamin C content of the CSB produced during this run? Determine whether each of the conditions is met in this case.

Population: all CSB produced that day
parameter: true mng of Vitamin C.

appears to fairly symmetric so assume normality of $\bar{x}$. Not an SRS, Not true ζ independent

(b) If the conditions are satisfied, construct and interpret a 95% confidence interval for μ using the Inference Toolbox. If not, explain why it would not be wise to calculate the interval.

$\bar{x} = 22.5$ $\quad s = 7.19$

$n = 8$

$t^* = 2.365$

$22.5 \pm 2.365\left(\frac{7.19}{\sqrt{8}}\right)$

22.5 ± 6.01

$(16.49, 28.51)$

I am 95% confident the true Vit C content is 16.49 - 28.51 mg/100g.

start 3/12

Paired t Procedures

Comparative studies are more convincing than single-sample investigations. For that reason, one-sample inference is less common than comparative inference. A common design to compare two treatments makes use of one-sample procedures. Recall that in a *matched pairs design,* subjects are matched in pairs and each treatment is given to one subject in each pair. Alternatively, each subject receives both treatments in some order. A coin toss can be used to assign the treatments to the two subjects in each pair or to determine the order in which an individual subject receives the two treatments. Another situation calling for ***paired t procedures*** is before-and-after observations on the same subjects.

> To compare the responses to the two treatments in a matched pairs design or before-and-after measurements on the same subjects, apply one-sample *t* procedures to the observed differences.

The parameter μ in a paired *t* procedure is

- the mean difference in the responses to the two treatments within matched pairs of subjects in the entire population (when subjects are matched in pairs), or
- the mean difference in response to the two treatments for individuals in the population (when the same subject receives both treatments), or
- the mean difference between before-and-after measurements for all individuals in the population (for before-and-after observations on the same individuals).

Some typical two-sample problems:

- Is one drug more effective than another?
- Do girls have more social insight than boys?
- Does one incentive plan promote the use of one credit card over another?
- Do students do better in courses when calculators are allowed compared to when they are not?
- Do people prefer white or dark chocolate?
- Does sunlight promote growth of plants?

Again, remember you must decide whether the experiment is a true two-sample experiment or a matched pairs design. The first problem, comparing drugs is most probably two-sample. You give drug A to one set of subjects and drug B to another set of subjects and compare the results. In the chocolate problem, you have a choice; You could give one set of people chocolate A to taste and another set of people chocolate B to taste, take the average rating and compare them. That would be a two-sample design. However, you could give each subject both chocolates to taste, get the ratings for both, and look at the difference between the ratings. That would be a one-sample matched pairs design.

Conditions for a two-sample design:

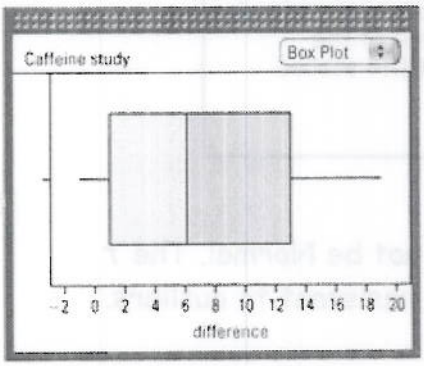
1. SRS
2. Normality: remember that we are working with differences not the raw data
3. Independence: Are the differences independent?

Example 16: Our subjects are 11 people diagnosed as being dependent on caffeine. Each subject was barred from coffee, colas, and other substances containing caffeine. Instead, they took capsules containing their normal caffeine intake. During a different time period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. Table 10.3 contains data on two of several tests given to the subjects. "Depression" is the score on the Beck Depression Inventory. Higher scores show more symptoms of depression. "Beats" is the number of beats per minute the subject achieved when asked to press a button 200 times as quickly as possible. We are interested in whether being deprived of caffeine affects these outcomes. Let's construct and interpret a 90% confidence interval for the mean change in depression score. As always, we follow the Inference Toolbox format.

Placebo - Caff

Subject	Depression (caffeine)	Depression (placebo)		Number of beats (caffeine)		Number of beats (placebo)
1	5	16	11	281	80	201
2	5	23	18	284	22	262
3	4	5	1	300	17	283
4	3	7	4	421	131	290
5	8	14	6	240	-19	259
6	5	24	19	294	3	291
7	0	6	6	377	23	354
8	0	3	3	345	1	346
9	2	15	13	303	20	283
10	11	12	1	340	-51	391
11	1	0	-1	408	-3	411

Source: E. C. Strain et al., "Caffeine dependence syndrome: evidence from case histories and experimental evaluation," *Journal of the American Medical Association*, 272 (1994), pp. 1604–1607.



Population: caffeine dependent people
Parameter: true average change in depression scores

SRS: Not sure. Proceed with caution are probably volunteers but at least treatments are randomized

Ind: differences are Ind Not ind scores

Normality: No reason to doubt it

$\bar{x} = 7.36$

$s = 6.92$

$df = 10$

$7.36 \pm 1.812\left(\frac{6.92}{\sqrt{11}}\right)$

7.36 ± 3.78

$(3.58, 11.14)$

We are 96% confident the true avg change in depression scores is between 3.58 to 11.14 points.

Example 17 shows how to turn paired data into single-sample data by taking differences within each pair. We are making inferences about a single population, the population of differences for all caffeine-dependent individuals. It would be incorrect to ignore the pairing and analyze the data as if we had two samples, one from subjects who took a placebo pill and the other from subjects who took a caffeine pill. Inference procedures for two samples assume that the samples are selected independently of each other. This assumption does not hold when the same subjects are measured twice. **The proper analysis depends on the design used to produce the data.**

Robustness of t procedures

The *t* confidence interval is exactly correct when the distribution of the population is exactly Normal. No real data are exactly Normal. The usefulness of the *t* procedures in practice therefore depends on how strongly they are affected by lack of Normality. Procedures that are not strongly affected are called ***robust.***

Robust Procedures

An inference procedure is called **robust** if the probability calculations involved in that procedure remain fairly accurate when a condition for use of the procedure is violated. For confidence intervals, this means that the stated confidence level is still fairly accurate.

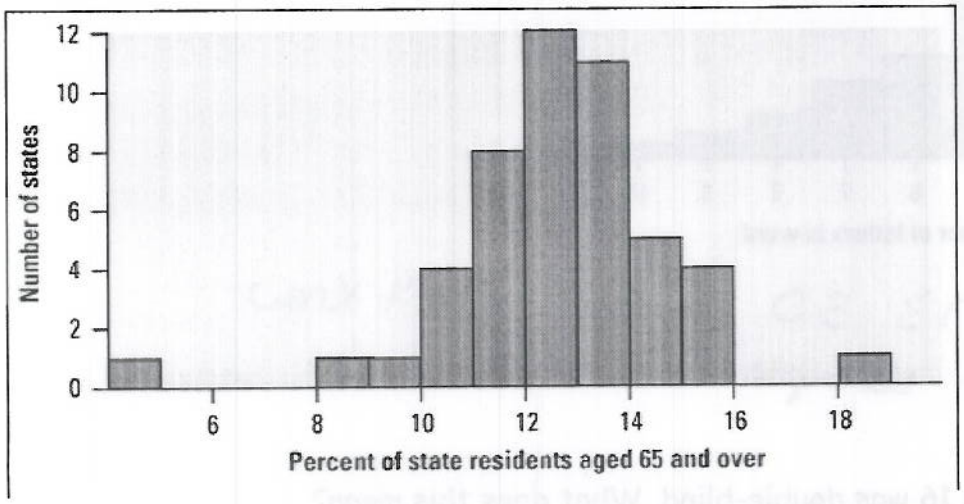
If outliers are present in the sample data, then the population may not be Normal. The *t* procedures are *not* robust against outliers, because $\bar{x}$ and s are not resistant to outliers.

Using the t procedures:

1. Except in the case of small samples, the assumption that the data are an SRS from the population of interest is more important than the assumption that the population distribution is Normal.
2. *Sample size less than 15.* Use *t* procedures if the data are close to Normal. If the data are clearly non-Normal or if outliers are present, do not use *t* procedures.
3. *Sample size at least 15.* The *t* procedures can be used except in the presence of outliers or strong skewness.
4. *Large samples.* The *t* procedures can be used even for clearly skewed distributions when the sample is large, say $n \geq 30$.

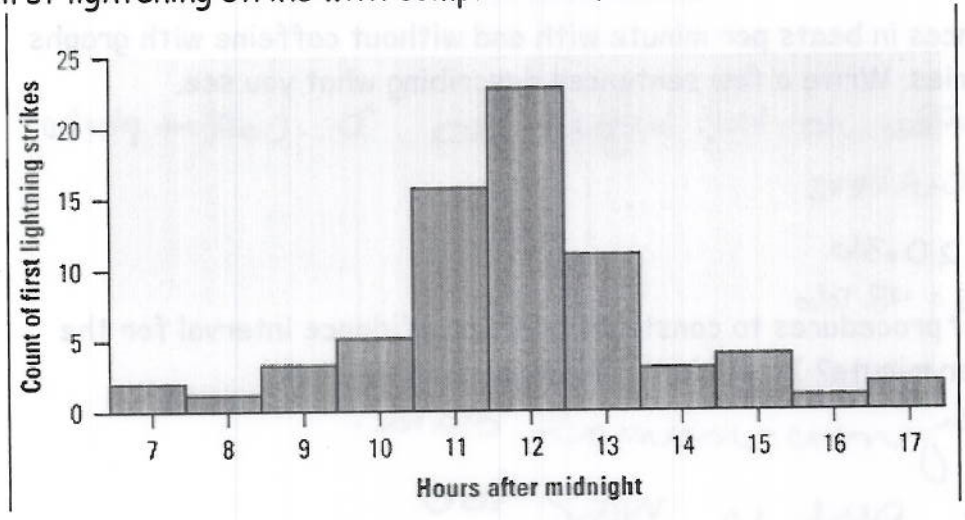
If your sample data would give a biased estimate for some reason, then you shouldn't bother computing a *t* interval. Or if the data you have *is* the entire population of interest, then there's no need to perform inference.

Example 18: The figure shows 3 histograms. The first one represents the percent of each state's residents who are at least 65% old. Why would using a CI not make sense for μ?



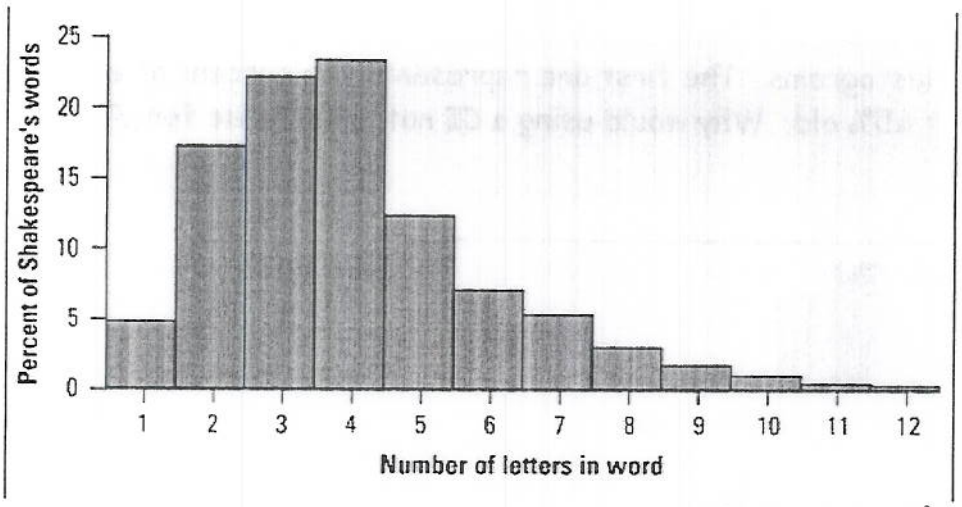
have the population of all 50 states so use it!

The second represents the time of the first lightening strike each day in a mountain region in Colorado. Why can we use t procedures to draw conclusions about the mean time iof a day's first lightening strike with complete confidence?



Seems fairly Symmetric so approx Normal & $n \geq 30$ so can safely use CLT

The third shows the distribution of word lengths in Shakespeare's plays. When can we use the t procedures?



only if $n \geq 30$ ↓ we don't know size of sample!

Example 19:

(a) The study in example 16 was double-blind. What does this mean?

Neither subjects nor researchers know what treatment was being used

(b) Examine the differences in beats per minute with and without caffeine with graphs and numerical summaries. Write a few sentences describing what you see.

Expect coffee to be higher so D= coffee-placebo

3 Suspected Outliers

$\bar{x} = 20.36$

$s = 48.66$

(c) Is it appropriate to use *t* procedures to construct a 90% confidence interval for the mean difference in beats per minute? If so, do it. If not, explain why not.

NO- Clearly non-normal data!

and Sample size is way too small for CLT

t-intervals on the TI calculator

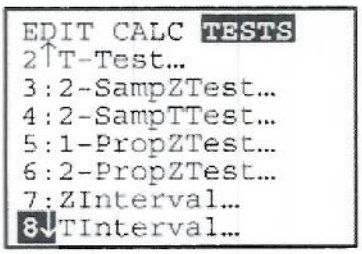
Confidence intervals using *t* procedures can be constructed on the TI-83/84/89, thus avoiding the use of tables. Here is a brief summary of the techniques using the healing-rates data. For reference, the differences in healing rates for the 14 newts (in micrometers per hour) are

−1	10	3	−3	−31	4	−12	−3	−7	−10	−22	−4	−1	−3

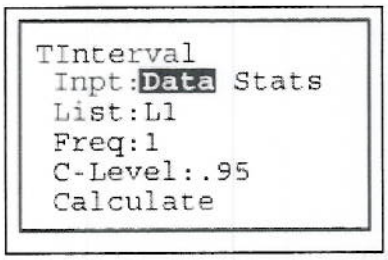
Enter these data in L_1. On the TI-83/84, all inference routines are found under STAT/TESTS

To determine a confidence interval for these data:

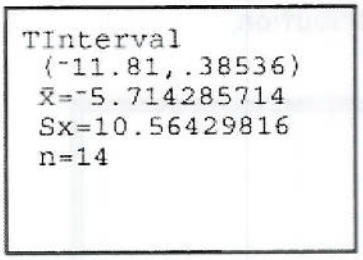
- Choose 8:TInterval.



Choose "Data" (not "Stats") and adjust the TInterval screen as shown.



Select "Calculate" and press ENTER



The results tell us that the 95% confidence interval for the true mean population difference in healing rate is between −11.81 and 0.385 micrometers per hour. If the researchers wanted to keep the 95% confidence level but wanted a shorter, more precise confidence interval, they would need to use more newts in the experiment (that is, increase the sample size *n*).