8.3
ESTIMATING A POPULATION MEAN

HW: P. 518 (55, 57, 59, 63, 65, 67, 71, 73, 75-79)

OVERVIEW…

- In this section, you will continue your study of confidence intervals by learning how to construct and interpret a confidence interval for the mean.
- The overall procedure is identical to that for a proportion, with one major difference: we will use a new distribution to determine critical values called a *t*-distribution.

CONDITIONS FOR INFERENCE ABOUT A POPULATION PROPORTION

- Random Sample
  - The data are a random sample from the population of interest.
- 10% Rule (Independent)
  - The sample size is no more than 10% of the population size: $n \leq \frac{1}{10}N$
- Large Counts (Normal)
  - If the sample size is large $(n \geq 30)$, then we can assume normality for any shape of a distribution.
  - When the sample is smaller than 30, the *t* procedures can be used except in the presence of outliers or strong skewness. Construct a quick graph of the data to make an assessment.

WHEN σ IS KNOWN: THE ONE-SAMPLE Z INTERVAL FOR POPULATION MEAN

Draw an SRS of size *n* from a large population having unknown mean μ and known standard deviation σ. As long as the Normal and Independent conditions are met, a level *C* confidence interval for μ is

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

The critical value z^* is found from the Normal distribution.

*This method isn't very useful and is **rarely used** to calculate a confidence interval for a population mean μ. **The only time z procedures are valid is when the population standard deviation is known.**

CHOOSING THE SAMPLE SIZE

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*:

$$z^* \frac{\sigma}{\sqrt{n}} \leq ME$$

EXAMPLE: HOW MUCH HOMEWORK?

Administrators at your school want to estimate how much time students spend on homework, on average, during a typical week. They want to estimate μ at the 90% confidence level with a margin of error of at most 15 minutes. A pilot study indicated that the standard deviation of time spent on homework per week is about 154 minutes.

How many students need to be surveyed?

$$z^* \frac{\sigma}{\sqrt{n}} \leq ME$$

1. Find z^* $(z^* = \pm 1.645)$
2. Substitute and solve for *n*.

$$1.645 \frac{154}{\sqrt{n}} \leq 15 \quad \rightarrow \quad 285.2 \leq n$$

The administrators need to survey at least 286 students.

CHECK YOUR UNDERSTANDING P. 501

To assess the accuracy of a laboratory scale, a standard weight known to weight 10 grams is weighed repeatedly. The scale readings are Normally distributed with unknown mean (this mean is 10 grams if the scale has no bias). In previous studies, the standard deviation of the scale reading has been about 0.0002 gram.

How many measurements must be averaged to get a margin of error of 0.0001 with 98% confidence? Show your work.

$$2.33\frac{0.0002}{\sqrt{n}} = 0.0001 \ \rightarrow \ \boldsymbol{n = 22}$$

T-DISTRIBUTIONS

T-DISTRIBUTIONS

- When the population standard deviation is unknown, we can no longer model the test statistic with the Normal distribution.
- We can no longer use the critical z^* values to determine the margin of error in our confidence interval.
- When the Normal condition is met, the test statistic calculated using the sample standard deviation s_x has a distribution similar in appearance to the Normal distribution, but with more area in the tails.

T-DISTRIBUTIONS

- The statistic t has the same interpretation as any standardized statistic: it says how far $\bar{x}$ is from its mean μ in standard deviation units.
- There is a different t distribution for each sample size.
- We specify a particular t distribution by giving its **degrees of freedom (df)**.
- The appropriate degrees of freedom are found by subtracting 1 from the sample size *n*, making $df = n - 1$. (Abbreviated as t_{n-1})

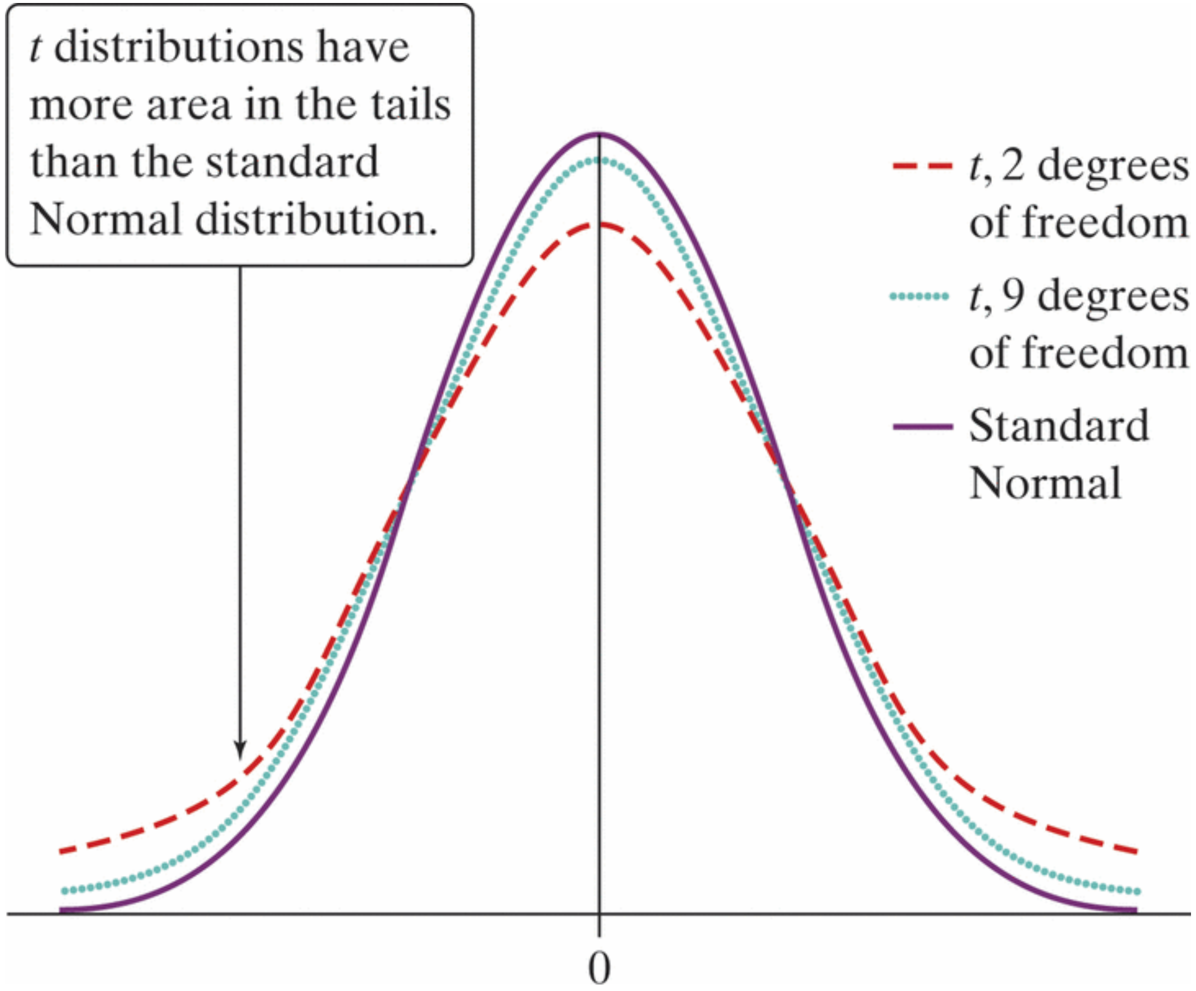
THE *T* DISTRIBUTIONS; DEGREES OF FREEDOM

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** $df = n - 1$. This statistic will have approximately a t_{n-1} distribution as long as the sampling distribution of $\bar{x}$ is close to Normal.

T DISTRIBUTIONS ILLUSTRATED



MORE ABOUT *T* DISTRIBUTIONS

- As the sample size (and degrees of freedom) increase, the t distribution approaches the standard Normal distribution more and more closely.
- We calculate standardized t values the same way we calculate z values. However, we must use the t table and consider degrees of freedom when determining critical values.
- When using the t distribution table, the areas shown are calculated to the **right** of t^*
- You should exercise caution in using t procedures when there is evidence of strong skewness or outliers in the sample data.
- You can use your calculator! 2nd→Vars→invT(

*Your calculator calculates the area to the **left** of the desired critical value.

LET'S TRY!

- Use the t table to determine the critical value t^* that you would use for a confidence interval for a population mean μ in the following situation:
- An 80% confidence interval from a sample with size $n = 19$.
  - $t^* = 1.330$
- Now try with your calculator
  - $invT(0.1, 18) = 1.330$ (use the absolute value so it's a positive value)

CHECK FOR UNDERSTANDING P. 507

Use the t table to determine the critical value t^* that you would use for a confidence interval for a population mean μ in each of the following situations. Check your answer with technology.

a) A 98% confidence interval based on $n = 22$ observations. $t^* = 2.518$
b) A 90% confidence interval from an SRS of 10 observations. $t^* = 1.833$
c) A 95% confidence interval from a sample of size 7. $t^* = 2.447$

CONSTRUCTING A CONFIDENCE INTERVAL FOR μ

STANDARD ERROR OF THE SAMPLE MEAN

- The standard error of the sample mean $\bar{x}$ is $\mathrm{SE} = \frac{s_x}{\sqrt{n}}$, where s_x is the sample standard deviation.
- It describes how far $\bar{x}$ will be from μ, on average, in repeated SRSs of size *n*.

FORMULA:

$$statistic \pm (critical\ value) \cdot (standard\ deviation\ of\ sample)$$

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

FOUR STEP PROCESS:

- **State** the parameter you want to estimate and at what confidence level.
- **Plan** which confidence interval you will construct and verify that the conditions have been met (Random, Independent, Normal).
  - *If the sample size is less than 30, construct a boxplot, histogram, dotplot, etc. to check for strong skewness or outliers.
- **Do** the actual construction of the interval using the formula $\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$
- **Conclude** by interpreting the interval in the context of the problem.

EXAMPLE

The amount of sugar in soft drinks is increasingly becoming a concern. To test sugar content, a researcher randomly sampled 8 soft drinks from a particular manufacturer and measured the sugar content in grams/serving. The following data were produced:

26 31 23 22 11 22 14 31

Use these data to construct and interpret a 95% confidence interval for the mean amount of sugar in this manufacturer's soft drinks.

ANSWERS

State: We want to find the mean amount of sugar in this manufacturer's soft drinks with a 95% confidence level.

Plan:

- Random: The drinks were randomly selected.
- Normal: sample size is less than 30, so we construct a boxplot of the sample data to check for strong skewness or outliers. The boxplot of the sample data does not suggest strong skewness or outliers.
- Independent: We can assume there are more than $8(10) = 80$ softdrinks manufactured by this company, so our 10% condition is satisfied.

ANSWERS

Do:

$t^* = 2.365, df = 7$

Use calculator to find s_x and $\bar{x}$

$\bar{x} = 22.5,\ s_x = 7.191$

$$\bar{x} = t^* \frac{s_x}{\sqrt{n}} \rightarrow 22.5 = 2.365 \frac{7.191}{\sqrt{8}} \rightarrow (\mathbf{16.488, 28.512})$$

Conclude: We are 95% confident the interval from 16.488 to 25.812 captures the true mean amount of sugar for this manufacturer's soft drinks.

CALCULATOR

You can use your calculator to find this confidence interval.

Using raw data

Enter data in L_1

Press STAT $\rightarrow$ TESTS

TInterval

- Highlight "Data"
- List: L_1
- Freq: 1
- C-Level:

Using summary statistics

Press STAT $\rightarrow$ TESTS

TInterval

- Highlight "Stats"
- $\bar{x}$:
- s_x:
- n:
- C-Level:

CHECK YOUR UNDERSTANDING P. 511

Biologists studying the healing of skin wounds measured the rate at which new cells closed a cut made in the skin of an anesthetized newt. Here are data from a random sample of 18 newts, measured in micrometers (millionths of a meter) per hour:

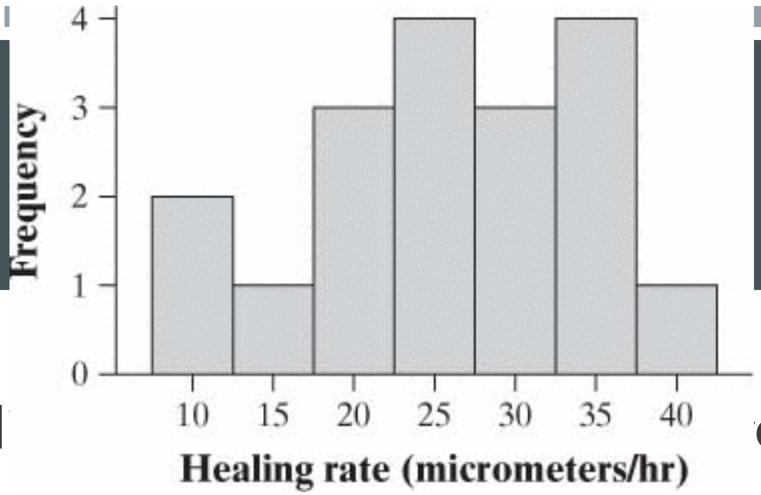
29 27 34 40 22 28 14 35 26 35 12 30 23 18 11 22 23 33

We want to estimate the mean healing rate μ with a 95% confidence interval.

1. Define the parameter of interest.

   Population mean healing rate.

CHECK YOUR UNDERSTANDING P. 511



2. What inference method will you use? Check that the cond e are met.

One-sample t interval for μ.

Random: The description says that the newts were randomly chosen.

Normal: We do not know if the data are Normal and there are fewer than 30 observations, so we graph the data. The histogram shows that the data are reasonably symmetric with no outliers, so this condition is met.

Independent: We have data on 18 newts. There are clearly more than 180 newts, so this condition is met.

CHECK YOUR UNDERSTANDING P. 511

3. Construct a 95% confidence interval for μ. Show your method.

I will use the T Interval Test on my calculator.

$$t^* = 2.110, \qquad df = 17$$

$$\bar{x} = 25.667, \qquad s_x = 8.324$$

$$(\mathbf{21.527, 29.806})$$

4. Interpret your interval in context.

We are 95% confident that the interval from 21.53 to 29.81 micrometers per hour captures the true mean healing time for newts.

ROBUST PROCEDURES

- An inference procedure is called **robust** if the probability calculations involved in that procedure remain fairly accurate when a condition for using the procedure is violated.
- For confidence intervals, "robust" means that the stated confidence level is still pretty accurate. That is, if we use the procedure to calculate many 95% confidence intervals, about 95% of them would capture the true population mean μ.
- If the procedure is NOT robust, then the actual capture rate might be very different from 95%.
- If outliers are present in the sample, then the population may not be Normal. The t procedures are *not* robust against outliers, because $\bar{x}$ and s_x are not resistant to outliers.

PRACTICAL GUIDELINES FOR THE NORMAL CONDITION AND POPULATION MEAN:

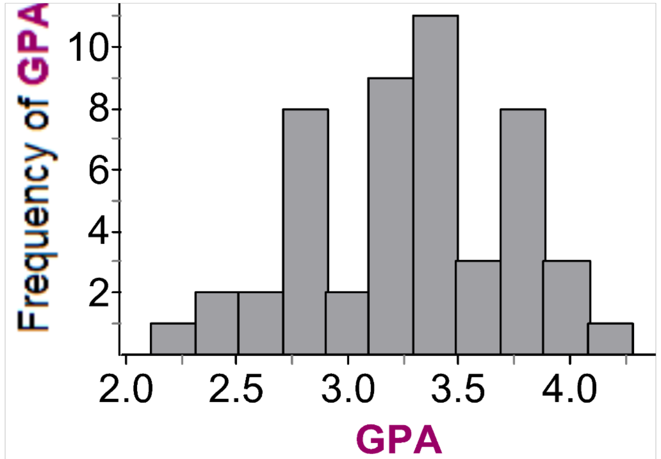
Always make a plot to check for skewness and outliers before you use the *t* procedures for small samples!

- Sample size less than 15: Use *t* procedures if the data appear close to Normal (roughly symmetric, single peak, no outliers). If the data are clearly skewed or if outliers are present, do not use *t*.
- Sample size at least 15: The *t* procedures can be used except in the presence of outliers or strong skewness.
- Large samples: The *t* procedures ca be used even for clearly skewed distributions when the sample is large, roughly $n \geq 30$.

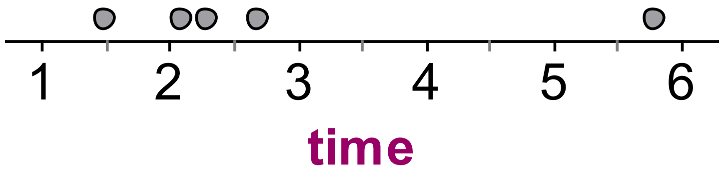
EXAMPLE:

Determine whether we can safely use a one-sample *t* interval to estimate the population mean in each of the following settings.

To estimate the average GPA of students at your school, you randomly select 50 students from classes you take. Here is a histogram of their GPAs:



The dotplot below shows the amount of time it took (in minutes) to order and receive a regular coffee in five visits to a local coffee shop.



The boxplot below shows the SAT Math scores for a random sample of 20 students at your high school.