

We do the same stuff here as we did with proportions. But remember we use the t-distribution for the means.

## Statistics

### Class Notes

#### Hypothesis Tests for Population Means (Section 10.3)

The mean height of American males is 69.5 inches. The mean height of American presidents is 70.78 inches. Could we conclude that American presidents are taller than the average American male?

According to the Crown ATM Network, the mean ATM withdrawal is \$67. If the company institutes improvements and the mean goes up to \$73, does that imply that the improvements increased the mean?

**Recall: Definition: Student's t-Distribution:** Suppose that a simple random sample of size  $n$  is taken from a population. If the population from which the sample is drawn follows a normal

distribution, the distribution of  $t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$  follows Student's t-distribution with  $n - 1$

degrees of freedom, where  $\bar{x}$  is the sample mean,  $s$  is the sample standard deviation, and  $\mu$  is the population mean.

We must verify that the following is true before continuing with hypothesis testing.

- sample data come from a simple random sample or randomized experiment,
- sample data has no serious outliers, ✓
- sample size is small relative to the population size ( $n \leq 0.05N$ ), and
- the data come from a population that is normally distributed, or the sample size is large ( $n \geq 30$ ).

Small departures from normality will *not* cause trouble. However, outliers are a bigger problem. If outliers exist, do not use these procedures.

Performing hypothesis tests for sample means follows the same logic as we used with proportions, except we will use the t-distribution here.

There are tables to look up  $P$ -values manually. However, we will rely on technology to find the  $P$ -values we need in this section. As a result, we will only study the  $P$ -value approach in this section.



## Summary of the $P$ -value Approach:

**Step 1:** Determine the null and alternative hypotheses. Again, the hypotheses can be structured in one of three ways:

1. Equal versus *not* equal hypothesis (**two-tailed test**)

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0 \quad \star$$

2. Equal versus less than (**left-tailed test**)

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0 \quad \star$$

3. Equal versus greater than (**right-tailed test**)

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0 \quad \star$$

We use  $\mu_0$  to denote the assumed population mean.

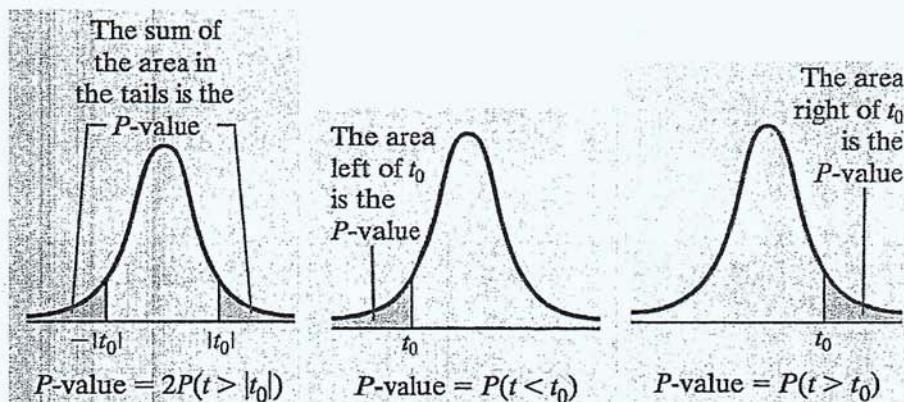
**Step 2:** Select a level of significance,  $\alpha$ , depending on the seriousness of making a Type I error.

**Step 3:** We could compute the test statistic  $t_0 = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$  and use Table VII to

approximate the  $P$ -value. (Table X in the book can be used for more accurate  $P$ -values.)

However, we will exclusively use the calculators or StatCrunch to perform the hypothesis testing. This calculation will be done for us.

**Step 4:** If the  $P$ -value  $< \alpha$ , reject the null hypothesis. For an understanding of the  $P$ -values, we will look quickly at these pictures.



**Step 5:** State the conclusion.



expl 1: Do students who learned English and another language simultaneously score worse on the SAT Critical Reading Exam than the general population of test takers? The mean score among all test takers is 501. A random sample of 100 test takers who learned English and another language simultaneously had a mean score of 485 with a standard deviation of 116. Answer the following questions.

a.) Determine the null and alternative hypotheses.

$$H_0: \mu = 501$$

$$H_1: \mu < 501$$

b.) Verify that the requirements to perform the test using the  $t$ -distribution are satisfied. In other words, is the sample size large? Is the sample random? Are the sample values independent? (It does not appear that we can check that there are no outliers.)

$$n = 100 \geq 30 \quad \text{random} \checkmark$$

$$n \leq 0.05N \quad ? \checkmark$$

c.) Perform a test at the  $\alpha = 0.01$  level. The calculator and StatCrunch instructions follow. What is your conclusion? Write a full sentence.

$$t \approx -1.379 \quad \text{test statistic}$$

$$p \approx 0.085$$

Our  $p$ -value is 0.085 which is not less than  $\alpha = 0.01$ .  
So, we do not reject the null hypothesis. We do not have sufficient evidence to say  $\mu < 501$ .

#### Instructions for TI Calculators:

1. If needed, enter the raw data in L1.

2. Press **STAT**. Arrow over the **TESTS**. Select **2: T-Test**.

3. If you entered data, select **Data**. If not, select **Stats**. Press **ENTER**. If you selected **Stats**, enter the assumed mean as  $\mu_0$ . Enter the sample mean as  $\bar{x}$ . Enter the sample's standard deviation as **Sx**. Enter the sample size as **n**. The final line will give you a spot to tell it you want a two-tailed, left-tailed, or right-tailed test. (If you selected **Data**, you'll enter  $\mu_0$ , **L1** for **Lists** and 1 for **Freq**.)

4. Finally, select **Calculate** and press **ENTER**. The calculator will output a  $t$ -value (the test statistic in step 3 on the previous page), the  $P$ -value we need (shown as  $p$ ), the sample mean and standard deviation, as well as the sample size, presumably to check.



expl 2: Calcium is essential to tree growth. In 1990, the concentration of calcium in precipitation in a city in New York was 0.11 milligrams per liter (mg / L). A random sample of ten precipitation dates in 2014 results in the following data. Answer the following questions.

Concentration of Calcium in Precipitation, Sample Data 2014				
0.065	0.087	0.070	0.262	0.126
0.183	0.120	0.234	0.313	0.108

A normal probability plot shows the data to be roughly normal. A boxplot shows no outliers.

Does the sample evidence show that calcium levels have changed since 1990? Use the  $\alpha = 0.01$  level.

$$H_0: \mu = 0.11$$

$$H_1: \mu \neq 0.11$$

test statistic  $t \approx 1.707$

pvalue  $\approx 0.122$

Since pvalue  $\neq \alpha = 0.01$ , we do not reject null hypothesis. We do not have sufficient evidence that the mean is no longer 0.11.

#### Instructions for StatCrunch:

1. If you have raw data, enter it in the spreadsheet.
2. Select Stat > T Stats > One Sample. Then choose With Data or With Summary.
3. If you chose With Data, select the column with the data. If you chose With Summary, enter the mean, standard deviation, and sample size. Choose the hypothesis test radio button. Enter the value of the mean stated in the null hypothesis and choose the direction of the alternative hypothesis from the pull-down menu. You can also tell it you want a QQPlot (normal probability plot) and a boxplot.) Click Compute!



When is statistically significant *not* significant?:

Consider this example.

expl 3: A math teacher has developed a course that she says will improve a student's performance on the SAT. Based on data from the College Board, SAT scores are normally distributed and the mean score is 515. This math teacher conducts a simple random sample to test her new course. The mean score of the 1800 students who took the course was 519 with a standard deviation of 111. Answer the following questions.

a.) State the null and alternative hypotheses.

$$H_0: \mu = 515 \quad H_1: \mu > 515$$

b.) Test the hypothesis at the 0.10 level of significance. What is your conclusion?

$$p\text{value} \approx 0.0632 < 0.10 = \alpha$$

we reject that  $\mu = 515$ . There is sufficient evidence to say the mean is greater than 515.

c.) Do you think a student would want to go through this course for that small an increase of their score? In other words, is this increase significant to the student?

no, maybe not. This increased mean score is not high enough to really make me take the course.

d.) Retest the null hypothesis assuming a sample size of 400 at the 0.10 level of significance. What is your conclusion?

$$p\text{value} \approx 0.2358$$

No longer significant at 10% level.

e.) How did the large sample size affect the conclusion?

it made the small difference in score appear significant.

**Definition: Practical significance** refers to the idea that, while small differences between the statistic and parameter stated in the null hypothesis are statistically significant, the difference may *not* be large enough to cause concern or be considered important.

When a large sample size is used in a hypothesis test, the results could be statistically significant even though the difference between the sample statistic and mean stated in the null hypothesis may have *no practical significance*.