

## AP Statistics: 9.3 : Tests about a Population Mean

Name: \_\_\_\_\_

You Will Learn:

- Carrying Out a Significance Test for  $\mu$
- One Sample t test
- Two Sided tests and confidence intervals
- Inference for Means: Paired Data
- Using Tests Wisely

- 1) Confidence intervals and Significance tests for a population proportion p are based on z-values for the standard Normal Distribution.
- 2) When population standard deviation  $\sigma$  is unknown, Inference about a population mean  $\mu$  uses a t distribution with n-1 degrees of freedom.

### Carrying Out a Significance Test for $\mu$

Recall from Chapter 8: Guidelines for Normal conditions ( for t- procedures)

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

### Refer to the example from P- 566 and do the following problem:

A classic rock radio station claims to play an average of 50 minutes of music every hour. However, every time you turn to this station it seems like there is a commercial playing. To investigate their claim, you randomly select 12 different hours during the next week and record what the radio station plays in each of the 12 hours. Here are the number of minutes of music in each of these hours:

44 49 45 51 49 53 49 44 47 50 46 48

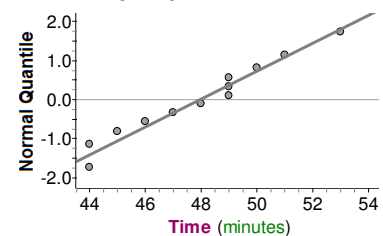
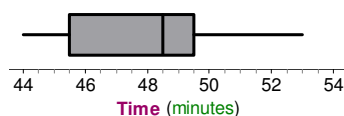
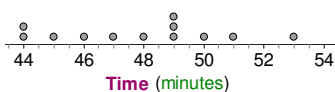
**Problem:** Check the conditions for carrying out a significance test of the company's claim that it plays an average of at least 50 minutes of music per hour.

**Solution:**

Random:

- **A random sample of hours was selected.**

**Normal: ( Make all the three graphs).** We don't know if the population distribution of music times is approximately Normal and we don't have a large sample size, so we will graph the data and look for any departures from Normality.



The dotplot and boxplot are roughly symmetric with no outliers, and the Normal probability plot is close to linear, so it is reasonable to use  $t$  procedures for these data.

- **Independent:** There are more than  $10(12) = 120$  hours of music played during the week.

In the previous example, you wanted to test  $H_0: \mu = 50$  versus  $H_a: \mu < 50$ .

**Problem:** Compute the test statistic and  $P$ -value for these data.

**Solution:**  $\bar{x} = \underline{\hspace{2cm}}$ ,  $s_x = \underline{\hspace{2cm}}$ , Test Statistic is  $t =$

d.f. =  $\underline{\hspace{2cm}}$ , . Using Table B and a  $t$  distribution for  $P(t < \underline{\hspace{2cm}}) = \underline{\hspace{2cm}}$

So, the  $P$ -value is between  $\underline{\hspace{2cm}}$  and  $\underline{\hspace{2cm}}$ .

The sample mean for these data is  $\bar{x} = 47.9$  with a standard deviation of  $s_x = 2.81$ .

Thus, the test statistic is  $t = \frac{47.9 - 50}{\frac{2.81}{\sqrt{12}}} = -2.59$ . Using Table B and a  $t$  distribution with

$12 - 1 = 11$  degrees of freedom, the  $P$ -value is  $P(t < -2.59)$ . This probability is the same as  $P(t > 2.59)$  so the  $P$ -value is between 0.01 and 0.02. [ Calculator:  $\text{tcdf}(-1E99, -2.59, 11) = .012$  ]

#### More practice with Table B

**Problem:**

Find the  $P$ -value for a test of  $H_0: \mu = 10$  versus  $H_a: \mu > 10$  that uses a sample of size 75 and has a test statistic of  $t = 2.33$ .

Since Table B does not include a line for  $df = 75$ , we will be conservative and use  $df = 60$ .

Since  $t = 2.33$  is between 2.099 and 2.390, the  $P$ -value is between 0.01 and 0.02.

Find the  $P$ -value for a test of  $H_0: \mu = 10$  versus  $H_a: \mu \neq 10$  that uses a sample of size 10 and has a test statistic of  $t = -0.51$ . ( Use two sided test) P- 568

Using the line for  $df = 9$ , we find that the value 0.51 is smaller than any  $t$ -value provided in that line. This means that the area under the  $t$ -curve to the right of 0.51 is greater than 0.25.

Since this is a two-sided test, the  $P$ -value must be at least 0.50.

#### One Sample $t$ test ( Refer to P- 570 , 571)

Every road has one at some point—construction zones that have much lower speed limits. To see if drivers obey these lower speed limits, a police officer used a radar gun to measure the speed (in miles per hour, or mph) of a random sample of 10 drivers in a 25 mph construction zone. Here are the results:

27 33 32 21 30 30 29 25 27 34

**Problem:**

(a) Can we conclude that the average speed of drivers in this construction zone is greater than the posted 25 mph speed limit?

**State:** We want to test the following hypotheses at the  $\alpha = 0.05$  significance level:

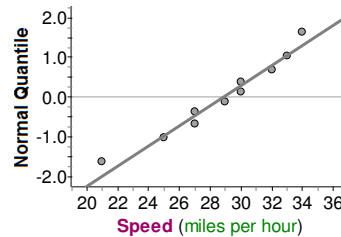
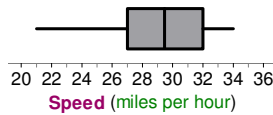
$$H_0: \mu = 25$$

$$H_a: \mu > 25$$

where  $\mu$  = the true mean speed of drivers in this 25 mph construction zone

**Plan:** If conditions are met, we will perform a one-sample  $t$  test for  $\mu$ .

- **Random:** A random sample of drivers was selected.
- **Normal:** We don't know if the population distribution of speeds is approximately Normal and we don't have a large sample size, so we will graph the data and look for any departures from Normality.



The dotplot and boxplot are only slightly skewed to the left with no outliers, and the Normal probability plot looks roughly linear, so it is reasonable to use  $t$  procedures for these data.

- **Independent:** There are more than  $10(10) = 100$  drivers that go through this construction zone.

**Do:** The sample mean speed is  $\bar{x} = 28.8$  mph with a standard deviation of  $s_x = 3.9$  mph.

- **Test statistic:**  $t = \frac{28.8 - 25}{3.94 / \sqrt{10}} = 3.05$
- **P-value:**  $P(t > 3.05)$  using the  $t$  distribution with  $10 - 1 = 9$  degrees of freedom. Using technology, the  $P$ -value =  $\text{tcdf}(3.05, 100, 9) = 0.0069$ .

**Conclude:** Since the  $P$ -value is less than  $\alpha$  ( $0.0069 < 0.05$ ), we reject the null hypothesis. There is convincing evidence that the true average speed of drivers in this construction zone is greater than 25 mph.

(b) Given your conclusion in part (a), which kind of mistake—a Type I or a Type II error—could you have made? Explain what this mistake means in this context.

Since we rejected the null hypothesis, it is possible that we made a Type I error. In other words, it is possible that we concluded that the average speed of drivers in this construction zone is greater than 25 mph when in reality it isn't.

### Two- Sided Tests and Confidence Intervals ( P- 574)

In the children's game Don't Break the Ice, small plastic ice cubes are squeezed into a square frame. Each child takes turns tapping out a cube of "ice" with a plastic hammer hoping that the remaining cubes don't collapse. For the game to work correctly, the cubes must be big enough so that they hold each other in place in the plastic frame but not so big that they are too difficult to tap out. The machine that produces the plastic ice cubes is designed to make cubes that are 29.5 millimeters (mm) wide, but the actual width varies a little. To make sure the machine is working well, a supervisor inspects a random sample of 50 cubes every hour and measures their width. The Fathom output below summarizes the data from a sample taken during one hour.

Collection 1		
Width	29.4874 mm	S1 = mean ( )
	50	S2 = count ( )
	0.0934676 mm	S3 = stdDev ( )
	0.0132183 mm	S4 = stdError ( )
	29.2717 mm	S5 = min ( )
	29.4225 mm	S6 = Q1 ( )
	29.4821 mm	S7 = median ( )
	29.5544 mm	S8 = Q3 ( )
	29.7148 mm	S9 = max ( )

**Problem:** (a) Interpret the standard deviation and the standard error provided by the computer output.

**Standard deviation:** The widths of the cubes are about 0.093 mm from the mean width, on average.

**Standard error:** In random samples of size 50, the sample mean will be about 0.013 mm from the true mean, on average.

(b) Do these data give convincing evidence that the mean width of cubes produced this hour is not 29.5 mm?  
( State, Plan, Do Conclude)

**State:** We want to test the following hypotheses at the  $\alpha = 0.05$  significance level:

$$H_0 : \mu = 29.5$$

$$H_a : \mu \neq 29.5$$

where  $\mu$  = the true mean width of plastic ice cubes

**Plan:** If conditions are met, we will perform a one-sample *t* test for  $\mu$ .

- **Random:** A random sample of plastic ice cubes was selected.  
**Normal:** We have a large sample size ( $n = 50 > 30$ ), so it is OK to use  $t$  procedures.
- **Independent:** It is reasonable to assume that there are more than  $10(50) = 500$  cubes produced by this machine each hour.

Do:

- **Test statistic:**  $t = \frac{29.4874 - 29.5}{\frac{0.09347}{\sqrt{50}}} = -0.95$
- **P-value:**  $2P(t < -0.95)$  using the  $t$  distribution with  $50 - 1 = 49$  degrees of freedom. Using technology, the  $P$ -value =  $2\text{tcdf}(-100, -0.95, 49) = 2(0.1734) = 0.3468$ .

**Conclude:** Since the  $P$ -value is greater than  $\alpha$  ( $0.3468 > 0.05$ ), we fail to reject the null hypothesis. There is not convincing evidence that the true width of the plastic ice cubes produced this hour is different from 29.5 mm.

( Refer to the example on P- 576)

Here is Fathom output for a 95% confidence interval for the true mean width of plastic ice cubes produced this hour. Would you make the same conclusion with the confidence interval as you did with the significance test in the previous example?

(a) Would you make the same conclusion with the confidence interval as you did with the significance test in the previous example?

Estimate of Collection 1		Estimate Mean
Attribute (numeric): Width		
Interval estimate for population mean of <b>Width</b>		
Count:	50	
Mean:	29.4874 mm	
Std dev:	0.0934676 mm	
Std error:	0.0132183 mm	
Confidence level:	95.0 %	
Estimate:	29.4874 mm +/- 0.0265632 mm	
Range:	29.4609 mm to 29.514 mm	

We are 95% confident that the interval from 29.4609 mm to 29.514 mm captures the true mean width of plastic ice cubes produced this hour. Since the interval includes 29.5 as a plausible value for the true mean width, we do not have convincing evidence that the true mean is not 29.5 mm. This is the same conclusion we made in the significance test earlier.

(b) Interpret the confidence level.

If we were to take many random samples of 50 plastic ice cubes and make 95% confidence intervals for the true mean width of the cubes, then about 95% of the intervals we construct will include the true mean width.

**Inference For Means: Paired Data ( P- 577)** ( Refer to the example on P- 577: Is Caffeine Dependence Real?)

For their second semester project in AP Statistics, Libby and Kathryn decided to investigate which line was faster in the supermarket, the express lane or the regular lane. To collect their data, they randomly selected 15 times during a week, went to the same store, and bought the same item. However, one of them used the express lane and the other used a regular lane. To decide which lane each of them would use, they flipped a coin. If it was heads, Libby used the express lane and Kathryn used the regular lane. If it was tails, Libby used the regular lane and Kathryn used the express lane. They entered their randomly assigned lanes at the same time and each recorded the time in seconds it took them to complete the transaction.

**Problem:** Carry out a test to see if there is convincing evidence that the express lane is faster.

**Solution:** Since these data are paired, we consider the differences in time ( regular-express). Here are 15 differences. In this case, a positive difference means that the express lane was faster.

5    246   -46   121   30   55   79   -94  
 -17   95   20   14   129   -39   42

**State:** We want to test the following hypotheses at the  $\alpha = 0.05$  significance level:

$H_0 : \mu_d = 0$

$H_a : \mu_d > 0$

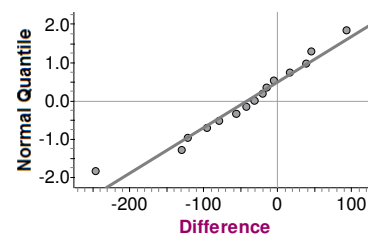
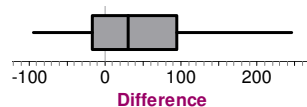
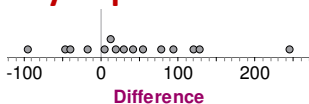
where  $\mu_d$  = the true mean difference (regular – express) in time required to purchase an item at the supermarket.

**Plan:** If conditions are met, we will perform a paired t test for  $\mu_d$ .

**Random:** A random sample of times was selected to make the purchases and the students were assigned to lanes at random.

**Normal:** We don't know if the population distribution of differences is approximately Normal and we don't have a large sample size, so we will graph the differences and look for any departures from Normality.

Time in Express Lane (seconds)	Time in Regular Lane (seconds)
337	342
226	472
502	456
408	529
151	181
284	339
150	229
357	263
349	332
257	352
321	341
383	397
565	694
363	324
85	127



The dotplot and boxplot are slightly skewed to the right with no outliers, and the Normal probability plot looks roughly linear, so it is reasonable to use  $t$  procedures for these data.

- **Independent:** Since we randomly selected the times to conduct the study from an infinite number of possible times, the differences should be independent.

**Do:** The sample mean difference is  $\bar{x} = 42.7$  seconds with a standard deviation of  $s_x = 84.0$  seconds.

- **Test statistic:**  $t = \frac{42.7 - 0}{84.0 / \sqrt{15}} = 1.97$
- **P-value:**  $P(t > 1.97)$  using the  $t$  distribution with  $15 - 1 = 14$  degrees of freedom. Using technology, the  $P$ -value =  $\text{tcdf}(1.97, 100, 14) = 0.034$ .

**Conclude:** Since the  $P$ -value is less than  $\alpha$  ( $0.034 < 0.05$ ), we reject the null hypothesis. There is convincing evidence that express lane is faster than the regular lane.

#### Using Tests Wisely:

1) Suppose that a large school district implemented a new math curriculum for the current school year. To see if the new curriculum is effective, the district randomly selects 500 students and compares their scores on a standardized test after the curriculum change to their scores on the same test before the curriculum change. The mean improvement was  $\bar{x}_d = 0.9$  with a standard deviation of  $s_d = 12.0$ .

#### **Problem:**

(a) Calculate the test statistic and  $P$ -value for a test of  $H_0: \mu_d = 0$  versus  $H_a: \mu_d > 0$ .

$$t = \frac{0.9 - 0}{12 / \sqrt{500}} = 1.68 \qquad P\text{-value} = \text{tcdf}(1.68, 100, 499) = 0.047$$

(b) Are the results significant at the 5% level?

**Since the  $P$ -value is less than 0.05, the results are significant at the 5% level. The improvement was larger than what we could expect to happen by chance. However, an increase of 0.9 points may or may not be practically significant.**

(c) Can we conclude that the new curriculum is the cause of the apparent increase in scores?

**No. Even though the increase was statistically significant, we cannot conclude that the new curriculum was the cause of the increase since there was no control group for comparison and no random assignment of treatments.**

- 2) Suppose that you wanted to know the average GPA for students at your school who are enrolled in AP Statistics. Since this isn't a large population, you conduct a census and record the GPA for each student. Is it appropriate to construct a one-sample  $t$  interval for the population mean GPA?

**No. If we have GPAs for every student in the population, we can calculate  $\mu$  exactly. We only use a confidence interval (or significance test) when we need to account for the variability caused by random sampling or random assignment to treatments.**

- 3) Suppose that 20 significance tests were conducted and in each case the null hypothesis was true. What is the probability that we avoid a Type I error in all 20 tests? ( Use 5% Significance Level)

**If we are using a 5% significance level, each individual test has a 0.95 probability of avoiding a Type I error.**

**Assuming that the results of the tests are independent, the probability of avoiding Type I errors in each of the tests is  $(0.95)(0.95) \dots (0.95) = (0.95)^{20} = 0.36$ . This means that there is a 64% chance we will make at least 1 Type I error in these 20 tests. So, finding one significant result in 20 is definitely not a surprise.**