Hypothesis Testing for the Population Mean: Summary of t test

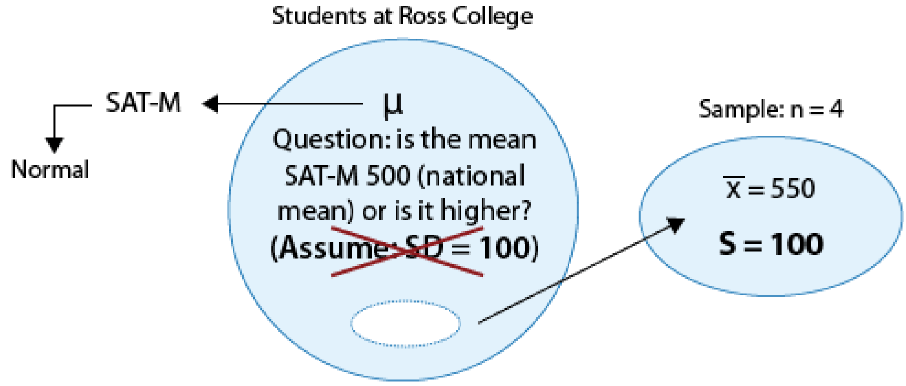
Learning Objective: Carry out hypothesis testing for the population proportion and mean (when appropriate), and draw conclusions in context.

For comparison purposes, we use a modified version of the two problems we used in the previous case. We first introduce the modified versions and explain the changes.

Example: 1

The SAT is constructed so that scores have a national average of 500. The distribution is close to normal. The dean of students of Ross College suspects that in recent years the college attracts students who are more quantitatively inclined. A random sample of 4 students entering Ross college had an average math SAT (SAT-M) score of 550, and a sample standard deviation of 100. Does this provide enough evidence for the dean to conclude that the mean SAT-M of all Ross College students is higher than the national mean of 500?

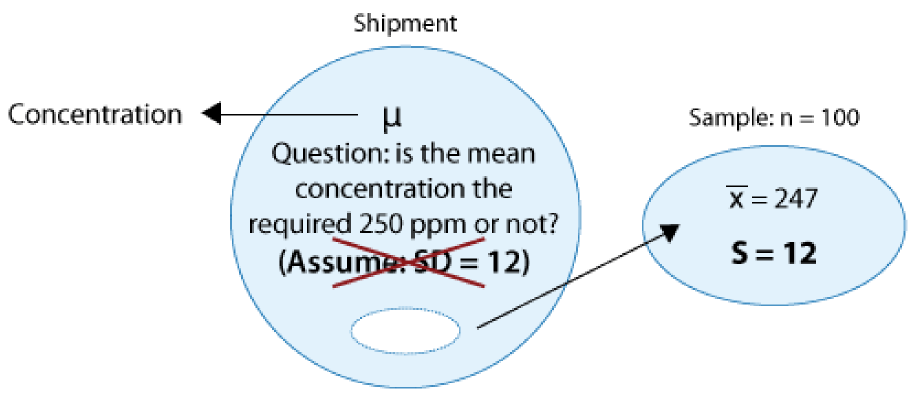
Here is a figure that represents this example where the changes are marked in blue:



Note that the problem was changed so that the population standard deviation (which was assumed to be 100 before) is now unknown, and instead we assume that the sample of 4 students produced a sample mean of 550 (no change) and a sample standard deviation of s=100. (Sample standard deviations are never such nice rounded numbers, but for the sake of comparison we left it as 100.) Note that due to the changes, the z-test for the population mean is no longer appropriate, and we need to use the t-test.

Example: 2

A certain prescription medicine is supposed to contain an average of 250 parts per million (ppm) of a certain chemical. If the concentration is higher than this, the drug may cause harmful side effects; if it is lower, the drug may be ineffective. The manufacturer runs a check to see if the mean concentration in a large shipment conforms to the target level of 250 ppm or not. A simple random sample of 100 portions is tested, and the sample mean concentration is found to be 247 ppm with a sample standard deviation of 12 ppm. Again, here is a figure that represents this example where the changes are marked in blue:



The changes are similar to example 1: we no longer assume that the population standard deviation is known, and instead use the sample standard deviation of 12. Again, the problem was thus changed from a z-test problem to a t-test problem.

However, as we mentioned earlier, due to the large sample size (n = 100) there should not be much difference whether we use the z-test or the t-test. The sample standard deviation, s, is expected to be close enough to the population standard deviation σ . We'll see this as we solve the problem.

Let's carry out the t-test for both of these problems:

Example 1:

1. There are no changes in the hypotheses being tested:

$$H_0 : \mu = 500$$
$$H_a : \mu > 500$$

2. The conditions that allow us to use the t-test are met since:

(i) The sample is random.

(ii) SAT-M is known to vary normally in the population (which is crucial here, since the sample size is only 4).

In other words, we are in the following situation:

Conditions: z-test for a population mean	Small sample size	Large sample size
Variable varies normally in the population	✓	✓
Variable doesn't vary normally in the population	✗	✓

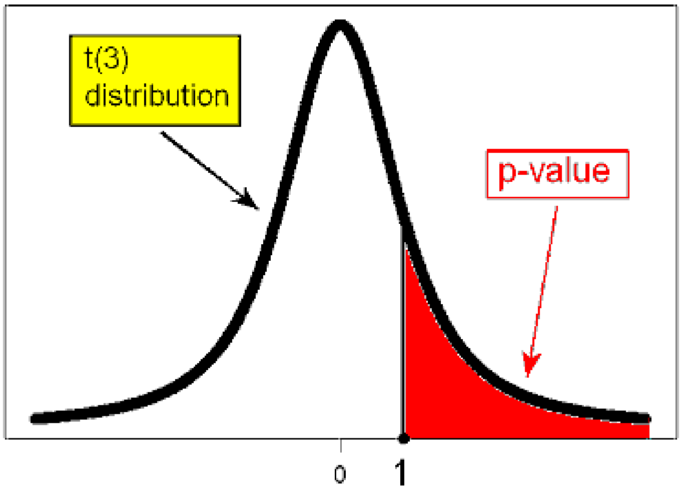
The test statistic is $t = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{550-500}{\frac{100}{\sqrt{4}}} = 1$

The data (represented by the sample mean) are 1 standard error above the null value.

3. Finding the p-value.

Recall that in general the p-value is calculated under the null distribution of the test statistic, which,

in the t-test case, is t(n-1). In our case, in which n = 4, the p-value is calculated under the t(3) distribution:



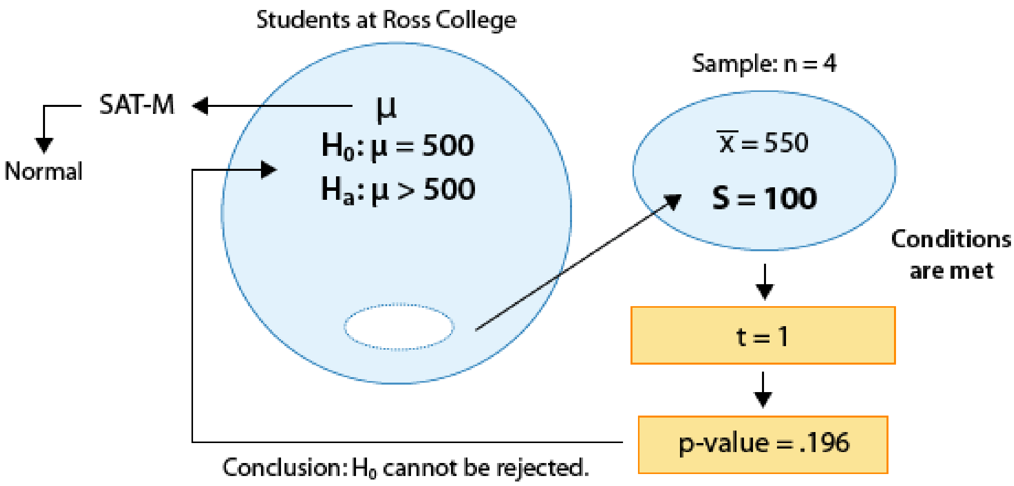
Using statistical software, we find that the p-value is 0.196. For comparison purposes, the p-value that we got when we carried out the z-test for this problem (when we assumed that 100 is the known σ rather the calculated sample standard deviation, s) was 0.159.

It is not surprising that the p-value of the t-test is larger, since the t distribution has fatter tails. Even though in this particular case the difference between the two values does not have practical implications (since both are large and will lead to the same conclusion), the difference is not trivial.

4. Making conclusions.

The p-value (0.196) is large, indicating that the results are not significant. The data do not provide enough evidence to conclude that the mean SAT-M among Ross College students is higher than the national mean (500).

Here is a summary:



Example 2:

1. There are no changes in the hypotheses being tested:

$$H_0 : \mu = 250$$
$$H_a : \mu \neq 250$$

2. The conditions that allow us to use the t-test are met:

(i) The sample is random

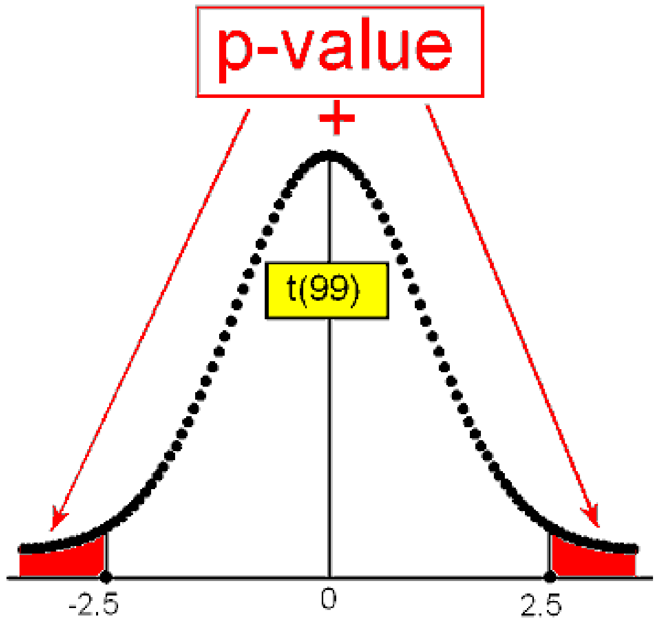
(ii) The sample size is large enough for the Central Limit Theorem to apply and ensure the normality of $\overline{X}$. In other words, we are in the following situation:

Conditions: z-test for a population mean	Small sample size	Large sample size
Variable varies normally in the population	✓	✓
Variable doesn't vary normally in the population	✗	✓

The test statistic is: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{247-250}{\frac{12}{\sqrt{100}}} = -2.5$

The data (represented by the sample mean) are 2.5 standard errors below the null value.

3. Finding the p-value.



To find the p-value we use statistical software, and we calculate a p-value of 0.014 with a 95% confidence interval of (244.619, 249.381). For comparison purposes, the output we got when we carried out the z-test for the same problem was a p-value of 0.012 with a 95% confidence interval of (244.648, 249.352).

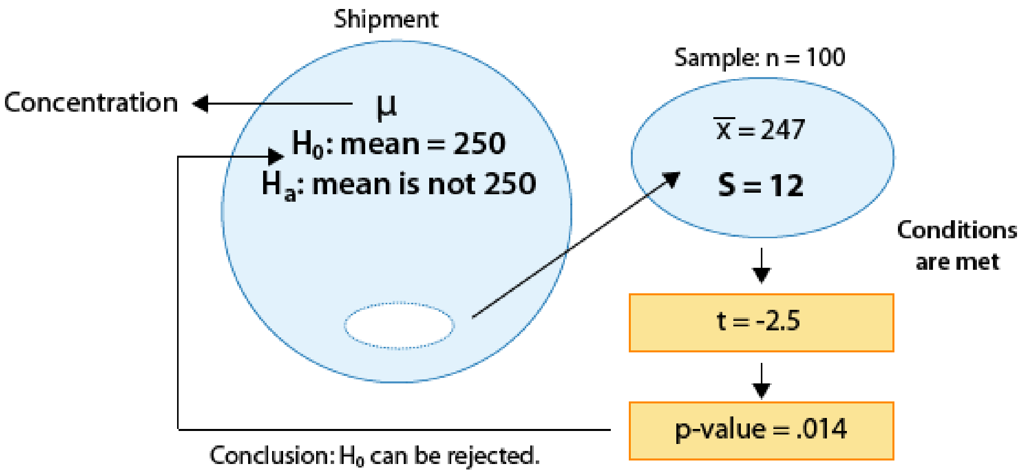
Note that here the difference between the p-values is quite negligible (0.002). This is not surprising, since the sample size is quite large (n = 100) in which case, as we mentioned, the z-test (in which we are treating s as the known σ) is a very good approximation to the t-test. Note also how the two 95%

confidence intervals are similar (for the same reason).

4. Conclusions:

The p-value is small (0.014) indicating that at the 5% significance level, the results are significant. The data therefore provide evidence to conclude that the mean concentration in entire shipment is not the required 250.

Here is a summary:



Comments

1. The 95% confidence interval for μ can be used here in the same way it is used when σ is known: either as a way to conduct the two-sided test (checking whether the null value falls inside or outside the confidence interval) or following a t-test where H_0 was rejected (in order to get insight into the value of μ).

2. While it is true that when σ is unknown and for large sample sizes the z-test is a good approximation for the t-test, since we are using software to carry out the t-test anyway, there is not much gain in using the z-test as an approximation instead. We might as well use the more exact t-test regardless of the sample size.

However, it is always worthwhile knowing what happens behind the scenes.

Scenario: Internet Use Among Users Age 50-65

A group of Internet users 50-65 years of age were randomly chosen and asked to report the weekly number of hours they spend online. The purpose of the study was to determine whether the mean weekly number of hours that Internet users in that age group spend online differs from the mean for

Internet users in general, which is 12.5 (as reported by "The Digital Future Report: Surveying the Digital Future, Year Four"). The following information is available:

Did I Get This

1/1 point (graded)

The population standard deviation is apparently:

- (•) unknown ✔
- ( ) 3.214
- ( ) 0.287

Answer

Correct:

Indeed, the problem description does not assume that the population standard deviation is known, and, therefore, the t-test is used.

Submit

Did I Get This

1/1 point (graded)

Note that the upper limit of the 95% confidence interval has been edited out. Which of the following is the number that has been edited out?

- ( ) 12.375
- (•) 12.577 ✔
- ( ) 13.439

Answer

Correct:

Indeed, since, based on the p-value, we cannot reject H_o at the 0.05 significance level, the 95% confidence interval must include 12.5 as a plausible value. Also the sample mean 12.008 is exactly the midpoint between 11.439 (the lower limit of the confidence interval) and 12.577, as it should be.

Submit

Did I Get This

1/1 point (graded)

If we suspected that the mean were less than 12.5, and carried out a one-sided test at the 0.05 significance level, we would conclude that the mean weekly number of hours that Internet users 50-65 years of age spend online:

- equals 12.5.
- may equal 12.5.
- is apparently different from 12.5.
- is apparently less than 12.5. ✔

Answer

Correct:

Since the p-value of the one-sided test is 0.09 / 2 = 0.045 (half the p-value of the two-sided test), therefore, at the 0.05 significance level, H_o can be rejected and we can conclude that the mean weekly number of hours that Internet users 50-65 years of age spend less than 12.5.

Submit

Let's Summarize

1. In hypothesis testing for the population mean (μ), we distinguish between two cases:

I. The less common case when the population standard deviation (σ) is known.

II. The more practical case when the population standard deviation is unknown and the sample standard deviation (s) is used instead.

2. In the case when σ is known, the test for μ is called the z-test, and in case when σ is unknown and s is used instead, the test is called the t-test.

3. In both cases, the null hypothesis is: $H_0 : \mu = \mu_0$

and the alternative, depending on the context, is one of the following:

$H_a : \mu < \mu_0$, or $H_a : \mu > \mu_0$, or $H_a : \mu \neq \mu_0$

4. Both tests can be safely used as long as the following two conditions are met:

(i) The sample is random (or can at least be considered random in context).

(ii) Either the sample size is large (n > 30) or, if not, the variable of interest can be assumed to vary normally in the population.

5. In the z-test, the test statistic is:

$$z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

whose null distribution is the standard normal distribution (under which the p-values are calculated).

6. In the t-test, the test statistic is:

$$t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

whose null distribution is t(n - 1) (under which the p-values are calculated).

7. For large sample sizes, the z-test is a good approximation for the t-test.

8. Confidence intervals can be used to carry out the two-sided test

$$H_0 : \mu = \mu_0 \quad vs.$$

$H_a : \mu \neq \mu_0$, and in cases where H_o is rejected, the confidence interval can give insight into the value of the population mean (μ).

9. Here is a summary of which test to use under which conditions:

		Sigma Known?	
		Known	Unknown
Situation	Large sample size (regardless of whether the population is normal or not)	**z-test**	**t-test** (z-test is a good approx.)
	Small sample size, population Normal* (footnote)	**z-test**	**t-test**
	Small sample size, population shape not Normal* or unknown (footnote)	**Neither z-test nor t-test**	

*by "population normal" we mean that either the population is known to be normal, or else that the population can be reasonably assumed to be normal as judged by the shape of the data histogram.

The following activity will reinforce the topics illustrated in the graphic above.

Scenario: Circuit Board Thickness

The Intel Corporation is conducting quality control on its circuit boards. Thickness of the manufactured circuit boards varies unavoidably from board to board. Suppose the thickness of the boards produced by a certain factory process varies normally. The distribution of thickness of the circuit boards is supposed to have the mean $\mu = 12$ mm if the manufacturing process is working correctly. A random sample of five circuit boards is selected and measured, and the average thickness is found to be 9.13 mm, and the standard deviation for the sample is computed to be 1.11 mm.

Learn By Doing

1/1 point (graded)

Which of the following is the proper output to use for testing whether or not the manufacturing process is working correctly, based on the study?

```
One-Sample T                                        A

Test of mu = 12 vs < 12

N      Mean     StDev   SE Mean      T       P
5  11.13000  1.11000  0.49641  -1.75  0.078
```

```
One-Sample Z                                        B

Test of mu = 12 vs < 12
The assumed standard deviation = 1.11

N      Mean   SE Mean      Z      P
5  11.13000  0.49641  -1.75  0.040
```

A

B

Answer

Correct:

In this case, σ (the population standard deviation) is unknown; the only standard deviation available was the sample standard deviation (it says "computed for the sample"), so the test could not be a z-test. Then, even though the sample size is relatively small (only five boards tested), the t-test is justified, because the population is known to be normal.

Submit

Learn By Doing

1/1 point (graded)

Which of the following would have made the z-test appropriate in the preceding scenario?

If the sample size had been 15 instead of 5, but the standard deviation of 1.11 had still been computed for the sample (with the rest of the scenario unchanged).

If the sample size had still been 5, but the standard deviation of 1.11 had been for the population of all circuit boards produced by the process (with the rest of the scenario unchanged).

Answer
Correct:
If the standard deviation had been for the population (so that σ had been known), then a z-test would be justified despite the small sample size.

Submit

Learn By Doing

1/1 point (graded)
Now, suppose that Intel is testing a brand new manufacturing process, for which prior information wasn't available. In particular, for this new process, **the population distribution's shape isn't known**. Use the following histograms to help you answer the question below.

To test whether or not the mean circuit board thickness is 12 mm with the new process, for which one of the following would **neither the z-test nor the t-test** be justified?

- ○ A random sample of 9 circuit boards is tested; the only available standard deviation is the one that is computed for the sample; and the histogram of the data is as shown above in histogram A.
- ○ A random sample of 40 circuit boards is tested; the only available standard deviation is the one that is computed for the sample; and the histogram of the data is as shown above in histogram B.
- ○ A random sample of 35 circuit boards is tested; the population standard deviation σ is known; and the histogram of the data is as shown above in histogram C.
- ◉ A random sample of 9 circuit boards is tested; the only available standard deviation is the one that is computed for the sample; and the histogram of the data is as shown above in histogram D. ✔

Answer
Correct:
In this case, the sample size is relatively small (only 9 boards), σ is unknown, and the population cannot be reasonably assumed to vary normally (because the data distribution is noticeably skewed). So in this case, nether the z-test nor the t-test would be justified to conduct formal statistical inference. (A different sort of test called a "non-parametric" test might be justified, but that is beyond the scope of this course.)

Submit