

14 –Hypothesis Testing, Part I

Hypothesis testing is the statistical evaluation of a conjecture. That conjecture will state that a population statistic (mean, proportion, or standard deviation) is either not equal to, greater than, or less than some fixed value.

For example, we may speculate that the mean fuel efficiency for 2013 American mid-sized cars is 19.8 mpg. The technique we will use enables us to confirm this value, or to deny it. In the statistician's language, hypothesis testing results in either *rejecting* the speculation that the mean is 19.8 or *not rejecting* that it is 19.8. Here is the technique explained using this example:

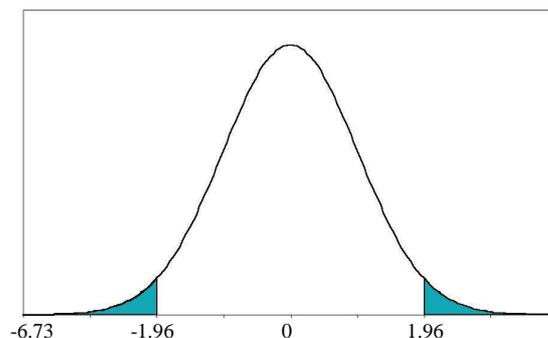
Example 1: A sales brochure suggests that the mean efficiency for all 2013 midsized American cars is 19.8 mpg. We believe that this may be inaccurate and wish to test the hypothesis that the mean is 19.8 mpg with a confidence level of 95%. The standard deviation for fuel efficiency is known to be 2.1 mpg, and a sample of 50 midsized 2013 American cars had a mean fuel efficiency of 17.8 mpg.

Solution & Procedure:

Step 1: Set up hypotheses and identify the problem as two-tailed, right-tailed, or left-tailed. There will be two hypotheses. The first, H_0 , is called the null hypothesis and will always state that the population statistic is equal to the fixed value given. In this case, the null hypothesis is $H_0: \mu = 19.8$. The second hypothesis, H_1 , is called the alternative hypothesis, and will state that the population statistic is not equal to, greater than, or less than the fixed value given. In this example, the alternative hypothesis is $H_1: \mu \neq 19.8$. When the alternative hypothesis uses the \neq symbol, it is called a two-tailed test. There will be different protocols for $>$ and $<$.

Step 2: Identify the tail(s) and its boundary values. In this problem, because we wish to be 95% confident of our conclusion, we find values in the z -distribution that have areas of 2.5% in the tails (just like we did in computing confidence intervals). In this case, those z scores are 1.96 and -1.96.

Step 3: Compute the test value. The test value is computed with a formula. There are several different formulas to consider and will look at each one separately. In this case, the formula is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{17.8 - 19.8}{2.1 / \sqrt{50}} \approx -6.73$.



Step 4: Reject or do not Reject H_0 . As a rule, if the test value is in the tail identified in step 2, we reject H_0 in favor of H_1 and if it is not in the tail, then we do not reject H_0 . Since -6.73 is in the left tail (-6.73 is less than -1.96), we reject the null hypothesis. Our conclusion is that we have sufficient evidence to indicate that the mean fuel efficiency of all 2013 midsized American cars is not 19.8 mpg. We make this claim with 95% confidence.

The steps will always be the same in this technique. The only part that may change is in the tails (steps 1 & 2) and the formula used (step 3). Here are the different scenarios:

- When the population standard deviation is unknown, or if the sample size is small (under 30), we use the t -distribution instead in $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$. The same decision making process from hypothesis testing applies here.
- When determining tails, if the alternative hypothesis uses the $<$ symbol, it is a left tailed test. When it uses $>$, it is a right tailed test. In these cases our confidence level corresponds to the area under the distribution out of the tail. The wording of the problem will discern whether the test is two-tailed, right-tailed, or left-tailed.
- Sometimes the term *significance* is used instead of confidence. It is simple: significance level = $1 - \text{confidence level}$.

Example 2: The life span of light bulbs is known to have a standard deviation of 120 hours. At random, a sample of 36 bulbs from a lot is selected and the mean life of the sample is determined to be 1500 hours. Is there evidence at the significance level of 0.01 to suggest that the life span is actually *longer* than the manufacturer's suggestion of 1350 hours?

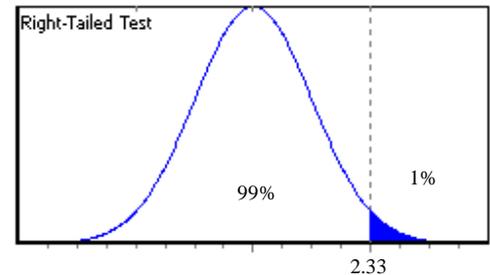
Solution:

Step 1: $\mu = 1350$; $H_1: \mu > 1350$. This is a right tailed test.

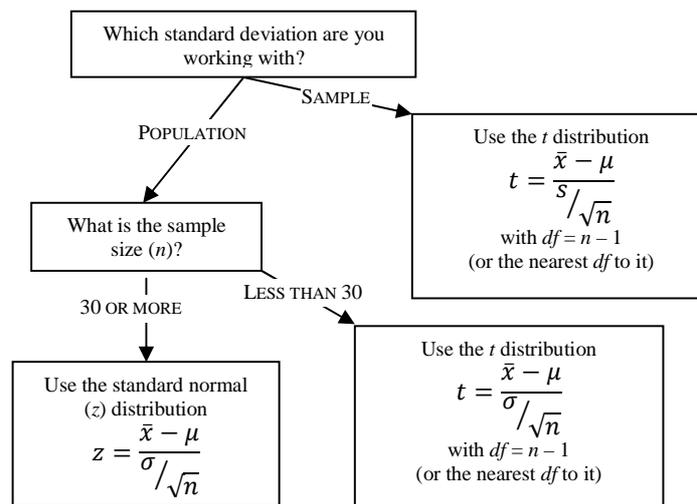
Step 2: Since this is a right-tailed test we need to find the z -score that creates an area of 0.01 in the right tail. So, we look up 0.9900 in the z table and get $z = 2.33$.

Step 3: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{1500 - 1350}{120 / \sqrt{36}} \approx 7.5$

Step 4: Since 7.5 is in the right tail (7.5 is larger than 2.33), we reject the null hypothesis. Our conclusion is that we have sufficient evidence to indicate that the mean life span of all light bulbs is longer than 1350 hours. And, we can make this claim with 99% certainty.



Here is the decision making process, adapted to the context of hypothesis testing:



Example 3: A local hospital administrator claims that on average 16.3 infections are contracted each week at their hospital. To test this claim, data was collected over a ten week period and the mean number of infections was 17.0 with standard deviation of 1.8. Is there enough evidence with 90% certainty to conclude that the mean number is actually larger than 16.3?

Solution:

Step 1: $\mu = 16.3$; $H_1: \mu > 16.3$. This is a right tailed test.

Step 2: Based upon the flow chart above, we need to use the t distribution. Since this is a right-tailed test we need to find the value of t that is the boundary for an area of 0.1 in the right tail. That value is $t = 1.383$ ($df = 9$).

Step 3: $t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{17.0 - 16.3}{1.8 / \sqrt{10}} \approx 1.22$

Step 4: Since 1.22 is not in the right tail (1.22 is smaller than 1.383), we do not reject the null hypothesis. Our conclusion is that we do not have sufficient evidence to indicate that the number of infections is larger than 16.3. And, we can make this claim with 90% certainty.

There is another approach that stems from the idea that each of the test values in examples 1 and 2 were substantially far into the tails. So, our confidence level could have been higher.

Rather than specifying a level of confidence, we can compute the test value and then give a threshold value of certainty above which H_0 would be rejected. This threshold is called a p -value. The smaller the p value, the higher the certainty.

Here are the p -values from the examples above:

Example 1, continued: The test value is $z = -6.73$ for a two-tailed test. The nearest value to this is -3.49 , which corresponds to an area of 0.0002 in the left tail. By symmetry, the right tail would match and so $p < .0004$. This means we can reject the null hypothesis with at least 99.96% certainty.

Example 2, continued: The test value is $z = 7.5$ for a right-tailed test. The nearest value to this is 3.49 , which corresponds to an area of 0.002 in the right tail. $p < .0002$. This means we can reject the null hypothesis with at least 99.98% certainty.

Example 3, continued: The test value is $t = 1.22$ with $df = 9$ for a right-tailed test. The nearest value to this is 1.383 , which would correspond to an area of 0.10 in the right tail. $p > 0.10$. This means we cannot reject the null hypothesis with any reasonable certainty.

Here is a complete example using the p value.

Example 4: The mean cost for 30-second nationally aired television commercials in 2012 was \$7284. This year, the mean cost for 20 commercials has been \$6572, with standard deviation of \$1291. Is there evidence to suggest that the cost has dropped? Use the p -value to draw a conclusion.

Solution:

Step 1: $\mu = 7284$; $H_1: \mu < 7284$. This is a left tailed test.

Step 2: We need to use the t distribution. $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{6572 - 7284}{1291/\sqrt{20}} \approx -2.466$ ($df = 19$)

Step 3: Because of symmetry, this is the same as a right-tailed test with $t = 2.466$. In the $df = 19$ row of the t distribution table, this value is between the values that correspond to 0.025 and 0.001 in the right tail. So, $.001 < p < .025$. In other words, we would reject $H_0: \mu = 7284$ in favor of $H_1: \mu < 7284$ with 97.5% certainty, but not with 99% certainty.