**Hypothesis Testing Exercises**

Se tiene un valor de referencia que puede ser lo que sea y una desviación standard . Tambien tiene una muestra donde la media es . La hipótesis nula siempre es H0: .

La hipótesis alternativa es lo que se observa. Si se ve que , entonces esa es su hipótesis alternativa. También puede ser que . Si se ve que los valores son mas o menos cercanos entonces se pone como hipótesis nula , para mayor seguridad.

Luego

H1 indica lo que se debe hacer:

Si H1: , se rechaza la hipótesis nula si . Quiere decir que la diferencia es de varias desviaciones estandar, o sea grande, lo que es improbable.

Si H1: ( o serán negativos), se rechaza la hipótesis nula si .

Si H1: , se rechaza la hipótesis nula si  o , acepta si . Esta probando si el estadístico cae dentro del intervalo de confianza, que tiene dos colas. Si cae, acepta la hipótesis nula, si no, rechaza.

Con el p.value es al revés, aunque al final, un poco diferente:

Si H1: , se rechaza la hipótesis nula si . Quiere decir que la probabilidad de una diferencia tan grande es muy baja.

Si H1: ( o serán negativos), se rechaza la hipótesis nula si   
.

Si H1: , rechaza la hipotesis nula si , es decir, que puede ser la mitad de las pruebas anteriores.

1. True or false: If the significance level of sample data, with respect to some null hypothesis, is 7%, then we could say, "The data is significant at the 10% level, but not at the 5% level."

**Answer**: True. This is purely terminology. You will sometimes hear people bracket the precise significance level instead of telling you the specific value. "Significant at x%" means the precise significance level is less than x%. "Not significant at y%" means the precise significance level is greater than y%.

Significance level is called . We reject Ho when .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Significance Level | Z from mean | Prob | Z estimated | Confidence in rejecting Ho |
|  | Z esta a 1.28 desv std |  | z small | Low |
|  | Z esta a 1.645 desv std |  | z median | Normal |
|  | Z esta a 2 desv std |  | z large | High |
|  | Z esta 2.326 desv std |  | z large | High |

Casos:

Que si los datos son significativos para , siempre los serán también para , pero no necesariamente para .

1. A company wonders if a new marketing campaign has increased demand for their product. They examine weekly demand prior to the campaign, and afterwards. If m1 is the old (true, long-term) average weekly demand, and m2 is the new (true, long-term) average weekly demand, and if the decision they must make is whether or not to allocate substantial further resources to the campaign, they should take as their alternative hypothesis that …

**Answer**: m1≥m2. This will force the campaign to "prove itself" before they throw new resources into it. Null hypothesis is that m1=m2.

1. A large distributor of cosmetics has kept his outstanding accounts receivable to a mean age of 18 days, over the past year. This average is considered a standard by which to measure the efficiency of the credit and collections department. During the current month, however, a random check of 100 accounts yields an average of 20 days, with a sample standard deviation of 9 days.

1. How significant is the difference from the standard?

**Answer**: The standard error of the mean in our study is , and so the actual data (20) is 2.22 standard deviations above what we'd expect to see if the standard (18) still held.

If the null hypothesis is that the truth is still 18, then a sample mean above 20 or below 16 would be at least as contradictory to the null hypothesis as is the data we're seeing here. So we do a two-tailed test (t-student), and find that the significance level of the data is 2\*p.value=0.0286=2.8544% (t.inv(2.22,99)=09857), so we multiply by 2 to account for the right tail also). The difference is significant for  
 , so we reject Ho. Cannot reject at .

We accept if

We reject if

1. If management has reason to believe that the collection of accounts is becoming slower, and is interested only in the possibility that the average has increased, how significant is the observed difference?

**Answer**: Here we'd do a one-tailed test, so the significance level of the (t-student) data is p.value=0.0143=1.4272%. We get a smaller number (I.e., stronger evidence against the null hypothesis) because, when the null hypothesis is that the true mean is less than or equal to 18, only data above 18 would contradict the null hypothesis … and that's exactly what we're seeing.

We accept if

We reject if

1. In a brand-preference survey of 1,600 consumers in a given area, 760 express a preference for brand A, and 840 for all other brands combined.

1. Construct a 95%-confidence interval for the percentage of consumers preferring brand A

**Answer**: 47.50%±2.45%

1. Is the proportion of consumers who prefer brand A significantly less than 50%?

**Answer**: We take as our null hypothesis that the proportion preferring brand A is 50 and that that the proportion preferring brand A is less than 50%. The standard error of the proportion in our study is 1.25%. Our estimated proportion is precisely two standard deviations below the hypothesized proportion, so the significance level of our data is 2 .

H0:

H1:

We accept if

We reject if

For the following exercises, use the following data and the t distribution:

Let's suppose that there exists a population of 7 million college students in the United States today. (The actual number depends on how you define "college student") And, let's assume that the average GPA of all of these college students is 2.7 (on a 4-point scale). Standard deviation is 0.4648.

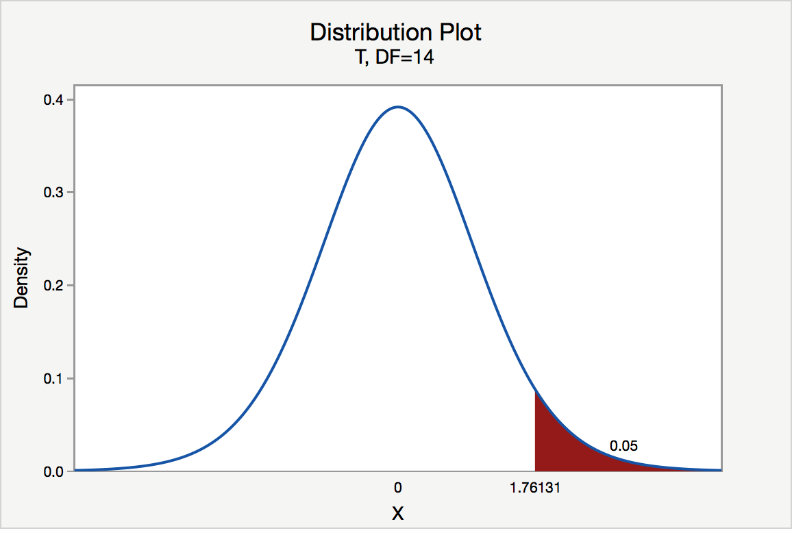
Suppose we take a random sample of *n* = 15 students majoring in mathematics. The mean was found to be 3. Since *n* = 15, our test statistic *t*\* has *n* - 1 = 14 degrees of freedom. Also, suppose we set our significance level α at 0.05, so that we have only a 5% chance of making a Type I error.

For the following examples, we are talking about the mean grade point average. Suppose that our random sample of n = 15 students majoring in mathematics yields a test statistic t\* equalling 2.5. Since n = 15, our test statistic t\* has n - 1 = 14 degrees of freedom. Also, suppose we set our significance level α at 0.05, so that we have only a 5% chance of making a Type I error.

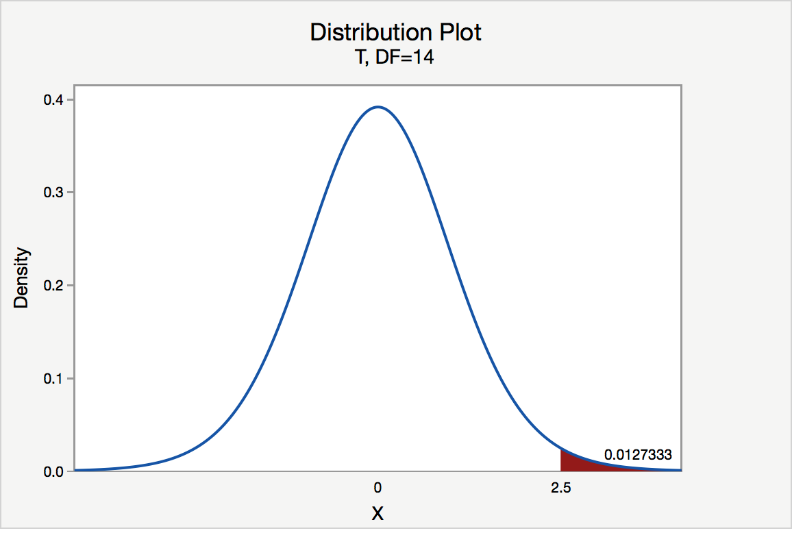
1. **Right-Tailed**

Assume  . Is it statistically significantly larger than 3? The **critical value** for conducting the right-tailed test H0: μ = 3 versus H1: μ > 3 is the t-value, denoted , such that the probability to the right of it is . It can be shown using either statistical software or a t-table that the critical value t0.05,14 is 1.7613. That is, we would reject the null hypothesis H0: μ = 3 in favor of the alternative hypothesis H1: μ > 3 if the test statistic t\* is greater than 1.7613. Visually, the rejection region is shaded red in the graph. In this case, . Since 2.5>1.7613 it falls in the rejection region.

***That is 3.3 is statistically significantly larger than 3 for a sample size of 15.***



The **P-value** for conducting the right-tailed test H0: μ = 3 versus H1: μ > 3 is the probability that we would observe a test statistic greater than t\* = 2.5 if the population mean really were 3. Recall that probability equals the area under the probability curve. The p.value is therefore the area under a tn - 1 = t14 curve and to the right of the test statistic t\* = 2.5. It can be shown using statistical software that the p.value is 0.0127. The graph depicts this visually.

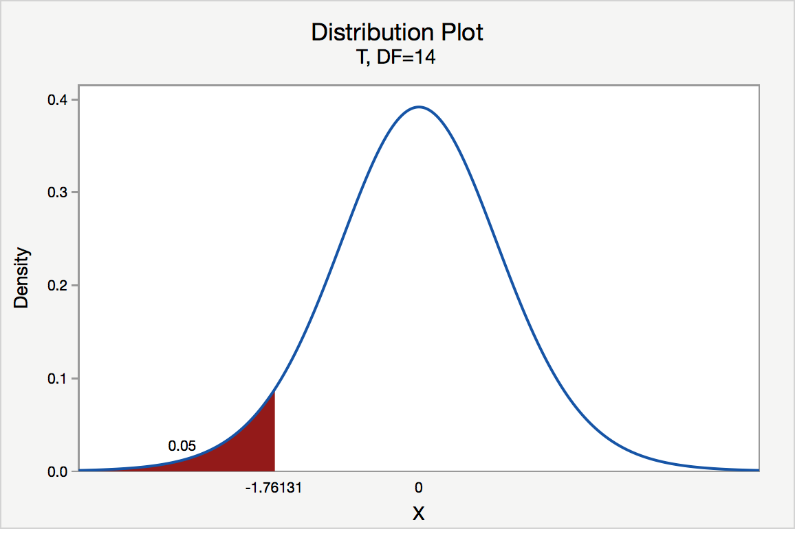


The p.value, 0.0127, tells us it is "unlikely" that we would observe such an extreme test statistic t\* in the direction of H1 if the null hypothesis were true. Therefore, our initial assumption that the null hypothesis is true must be incorrect. That is, since p.value, 0.0127, is less than  = 0.05, we reject the null hypothesis H0: μ = 3 in favor of the alternative hypothesis H1: μ > 3.

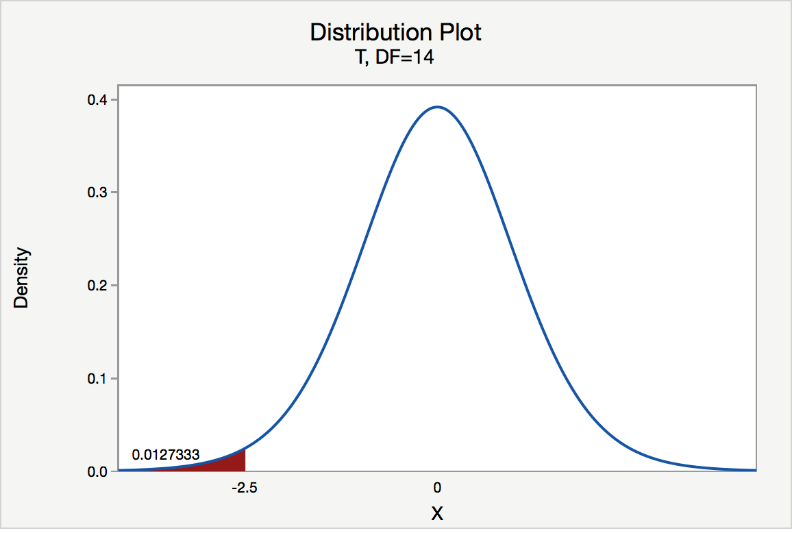
Note that we would not reject H0: μ = 3 in favour of H1 μ > 3 if we lowered our willingness to make a Type I error to  = 0.01 instead, as the p.value, 0.0127, is then greater than  = 0.01.

1. **Left-Tailed**

Assume  . Is it statistically significantly smaller than 3? The **critical value** for conducting the left-tailed test H0: μ = 3 versus H1: μ < 3 is the t-value, denoted , such that the probability to the left of it is . It can be shown using either statistical software or a t-table that the critical value -t0.05,14 is -1.7613. That is, we would reject the null hypothesis H0: μ = 3 in favor of the alternative hypothesis H1: μ < 3 if the test statistic t\* is less than -1.7613. Visually, the rejection region is shaded red in the graph. In this case, . Since -2.5<1.7613 it falls in the rejection region.



The **p.value** for conducting the left-tailed test H0: μ = 3 versus H1: μ < 3 is the probability that we would observe a test statistic less than t\* = -2.5 if the population mean μ really were 3. The p.value is therefore the area under a tn - 1 = t14 curve and to the left of the test statistic t\* = -2.5. It can be shown using statistical software that the p.value is 0.0127. The graph depicts this visually.



The p.value, 0.0127, tells us it is "unlikely" that we would observe such an extreme test statistic t\* in the direction of H1 if the null hypothesis were true. Therefore, our initial assumption that the null hypothesis is true must be incorrect. That is, since the p.value, 0.0127, is less than α = 0.05, we reject the null hypothesis H0: μ = 3 in favour of the alternative hypothesis H1: μ < 3.

Note that we would not reject H0: μ = 3 in favour of H1: μ < 3 if we lowered our willingness to make a Type I error to  = 0.01 instead, as the P-value, 0.0127, is then greater than  = 0.01.

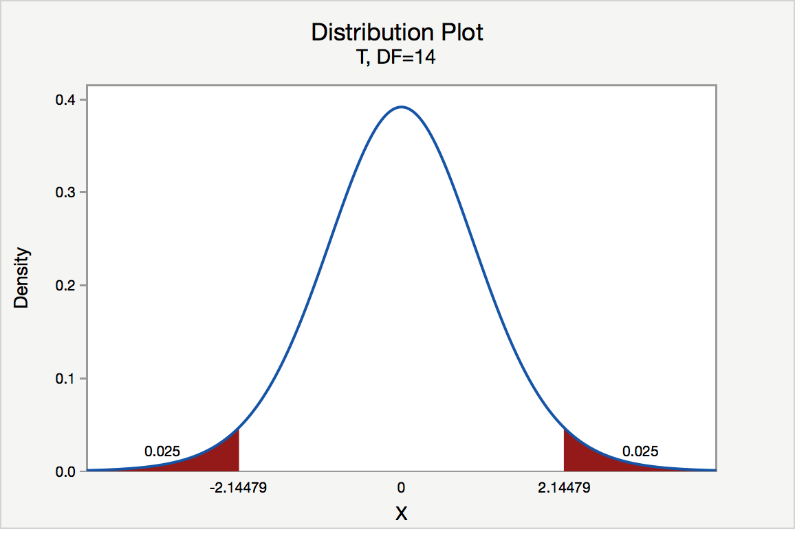
***That is 2.7 is statistically significantly smaller than 3 for a sample size of 15.***

1. **Two-Tailed**

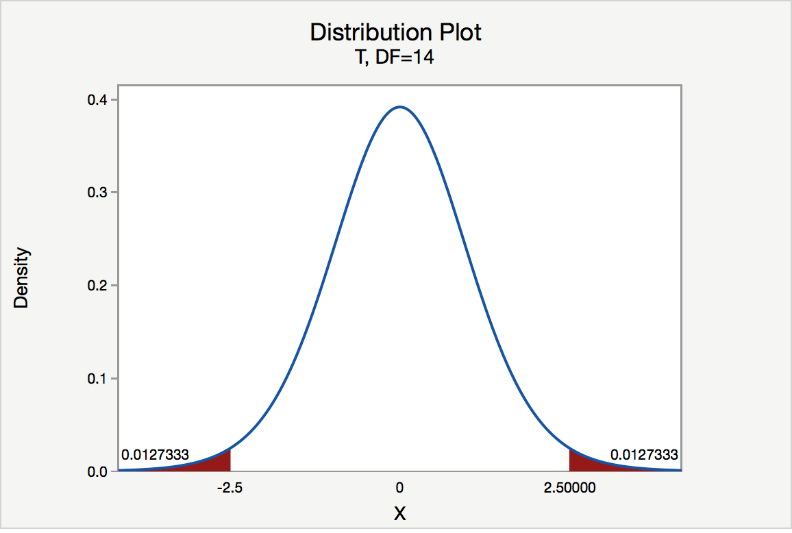
There are two **critical values** for the two-tailed test H0: μ = 3 versus H1: μ ≠ 3, one for the left-tail denoted and one for the right-tail denoted . The value is the t-value such that the probability to the left of it is , and the value is the t-value such that the probability to the right of it is . It can be shown using either statistical software or a t-table that the critical value -t0.025,14 is -2.1448 and the critical value t0.025,14 is 2.1448. That is, we would reject the null hypothesis   
H0: μ = 3 in favor of the alternative hypothesis H1: μ ≠ 3 if the test statistic t\* is less than -2.1448 or greater than 2.1448. Visually, the rejection region is shaded red in the graph. In this case, . Since 2.5>2.1448 it falls in the rejection region. And . Since -2.5<-2.1448 it also falls in the rejection region

Note that the t-value for a two-tailed test always uses for the t-value for either of the one-tailed tests.

***Notice how this is equivalent to testing if he test statistic falls inside the confidence interval***.



The **p.value** for conducting the two-tailed test H0: μ = 3 versus H1: μ ≠ 3 is the probability that we would observe a test statistic less than -2.5 or greater than 2.5 if the population mean μ really were 3. That is, the two-tailed test requires taking into account the possibility that the test statistic could fall into either tail (and hence the name "two-tailed" test). The p.value is therefore the area under a tn - 1 = t14 curve to the left of -2.5 and to the right of the 2.5. It can be shown using statistical software that the p.value is 0.0127 + 0.0127, or 0.0254. The graph depicts this visually.



Note that the p.value for a two-tailed test is always **two times** the P-value for either of the one-tailed tests. The p.value, 0.0254, tells us it is "unlikely" that we would observe such an extreme test statistic t\* in the direction of HA if the null hypothesis were true. Therefore, our initial assumption that the null hypothesis is true must be incorrect. That is, since the p.value, 0.0254, is less than α = 0.05, we reject the null hypothesis H0: μ = 3 in favor of the alternative hypothesis H1: μ ≠ 3.

Note that we would not reject H0: μ = 3 in favour of H1: μ ≠ 3 if we lowered our willingness to make a Type I error to  = 0.01 instead, as the p.value, 0.0254, is then greater than  = 0.01.

***That is 3.3 is statistically significantly larger than 3 for a sample size of 15. And That is 2.7 is statistically significantly smaller than 3 for a sample size of 15.***

1. **Weight Loss for Diet vs Exercise**. Did dieters lose more fat than the exercisers?

Diet Only:

sample mean= 5.9 kg

sample standard deviation = 4.1 kg

sample size = n = 42

standard error = SEM1 = 4.1/ √42 = 0.633

Exercise Only:

sample mean= 4.1 kg

sample standard deviation = 3.7 kg

sample size = n = 47

standard error = SEM2 = 3.7/ √47 = 0.540

**Answer**:

measure of variability = (since variances add)

Determine the null and alternative hypotheses:

**Null hypothesis**: No difference in average fat lost in population for two methods. Population mean difference is zero.

**Alternative hypothesis**: There is a difference in average fat lost in population for two methods. Population mean difference is not zero. Then double tail test.

Test statistic.

So the test statistic:

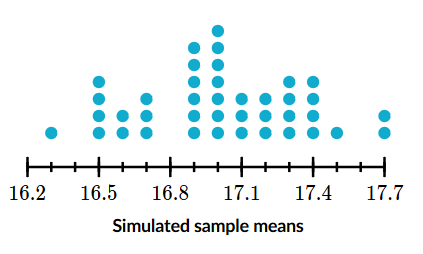
p-value=2x0.015=0.03

The p-value of 0.03 is less than or equal to 0.05, so ...If really there is no difference between dieting and exercise as fat loss methods, would see such an extreme result only 3% of the time, or 3 times out of 100. Prefer to believe truth does not lie with null hypothesis. We conclude that there is a statistically significant difference between average fat loss for the two methods.

1. Test with Simulation of Random Variable

An employee at an aquarium monitors how much their sea otters eat. The amount of food a particular otter eats daily is approximately normally distributed with a mean of 17 pounds and a standard deviation of 1 pound. They suspected this otter was not eating enough, so they took a random sample n=10 days and observed a sample mean of   pounds of food per day.

To see how likely a sample like this was to occur by random chance alone, the employee performed a simulation. They simulated 40 samples of n=10 values from a normal population with a mean of 17 pounds and a standard deviation of 1 pound. They recorded the mean of the values in each sample. Here are the sample means from their 40 samples:



They want to test:

H0: μ=17 lbs

Ha: μ<17 lbs

where μ is the true mean amount of food per day.

Based on these simulated results, what is the approximate p-value of the test?

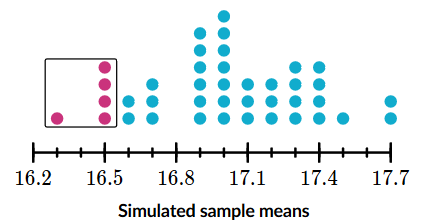
Note: The sample result was xˉ=16.5 lbs.

**Answer**:

The n=10 days in the sample had a mean of xˉ=16.5 lbs.

Since the alternative hypothesis is Ha:μ<17 lbs, we can find the approximate p-value of this result by looking at how often a sample result as low or lower than 16.5 lbs occurred in the simulation.

The simulation produced a sample mean at or below 16.5 lbs in 5 out of 40 samples:



p-value=