S.3.2 Hypothesis Testing (P-Value Approach)

The *P*-value approach involves determining "likely" or "unlikely" by determining the probability — assuming the null hypothesis were true — of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed. If the *P*-value is small, say less than (or equal to) α, then it is "unlikely." And, if the *P*-value is large, say more than α, then it is "likely."

If the *P*-value is less than (or equal to) α, then the null hypothesis is rejected in favor of the alternative hypothesis. And, if the *P*-value is greater than α, then the null hypothesis is not rejected.

Specifically, the four steps involved in using the *P*-value approach to conducting any hypothesis test are:

1. Specify the null and alternative hypotheses.
2. Using the sample data and assuming the null hypothesis is true, calculate the value of the test statistic. Again, to conduct the hypothesis test for the population mean μ, we use the *t*-statistic $t^*=\frac{\bar{x}-\mu}{s/\sqrt{n}}$ which follows a *t*-distribution with *n* - 1 degrees of freedom.
3. Using the known distribution of the test statistic, calculate the ***P*-value**: "If the null hypothesis is true, what is the probability that we'd observe a more extreme test statistic in the direction of the alternative hypothesis than we did?" (Note how this question is equivalent to the question answered in criminal trials: "If the defendant is innocent, what is the chance that we'd observe such extreme criminal evidence?")
4. Set the significance level, α, the probability of making a Type I error to be small — 0.01, 0.05, or 0.10. Compare the *P*-value to α. If the *P*-value is less than (or equal to) α, reject the null hypothesis in favor of the alternative hypothesis. If the *P*-value is greater than α, do not reject the null hypothesis.

Example S.3.2.1

Mean GPA

Statistics Online

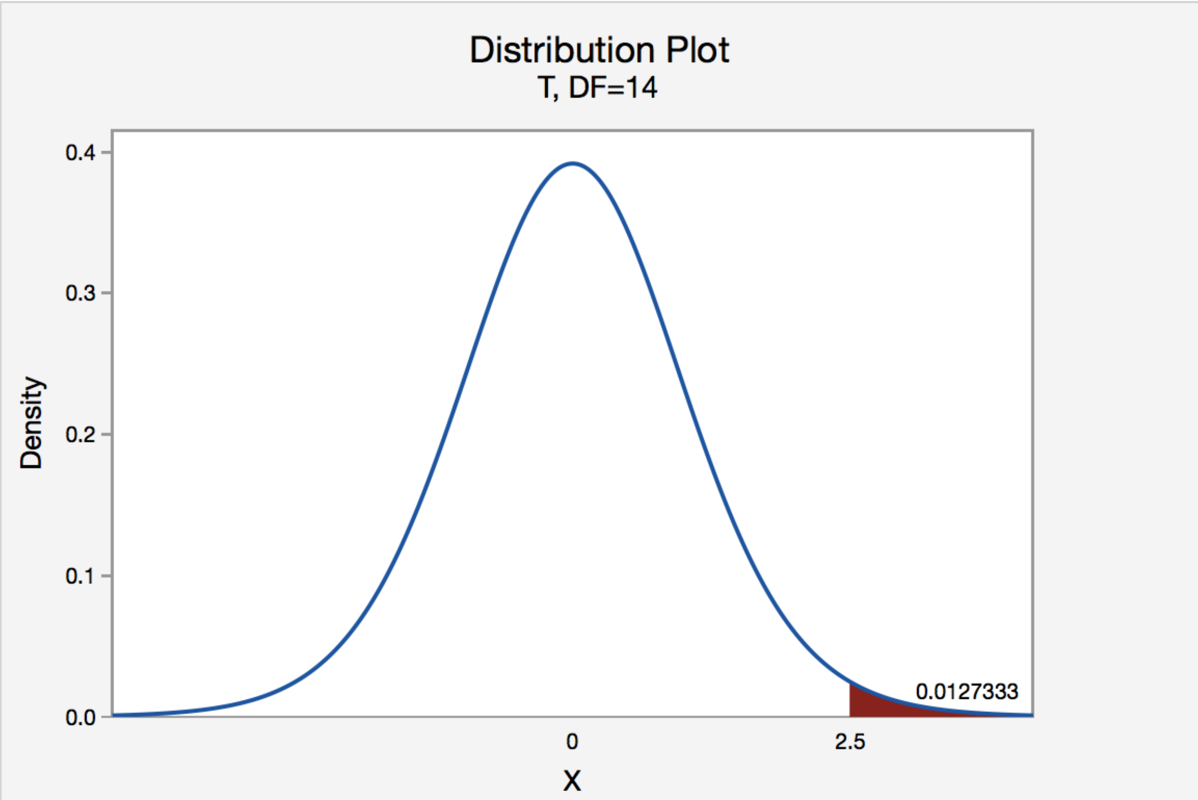
- For Students
  - Learning Online Orientation
  - Math & Stat Reviews
    - Algebra Review
    - Basic Statistical Concepts
      - S.1 Basic Terminology
      - S.2 Confidence Intervals
      - S.3 Hypothesis Testing
        - S.3.1 Hypothesis Testing (Critical Value Approach)
        - **S.3.2 Hypothesis Testing (P-Value Approach)**
        - S.3.3 Hypothesis Testing Examples
      - S.4 Chi-Square Tests
      - S.5 Power Analysis
      - S.6 Test of Proportion
      - S.7 Self-Assess
    - Calculus Review
    - Matrix Algebra Review
  - Ethics and Statistics

In our example concerning the mean grade point average, suppose that our random sample of n = 15 students majoring in mathematics yields a test statistic t^* equaling 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.

Technology Tutorials

Right Tailed

The P-value for conducting the **right-tailed** test $H_0 : \mu = 3$ versus $H_A : \mu > 3$ is the probability that we would observe a test statistic greater than t^* = 2.5 if the population mean μ really were 3. Recall that probability equals the area under the probability curve. The P-value is therefore the area under a $t_{n-1} = t_{14}$ 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.

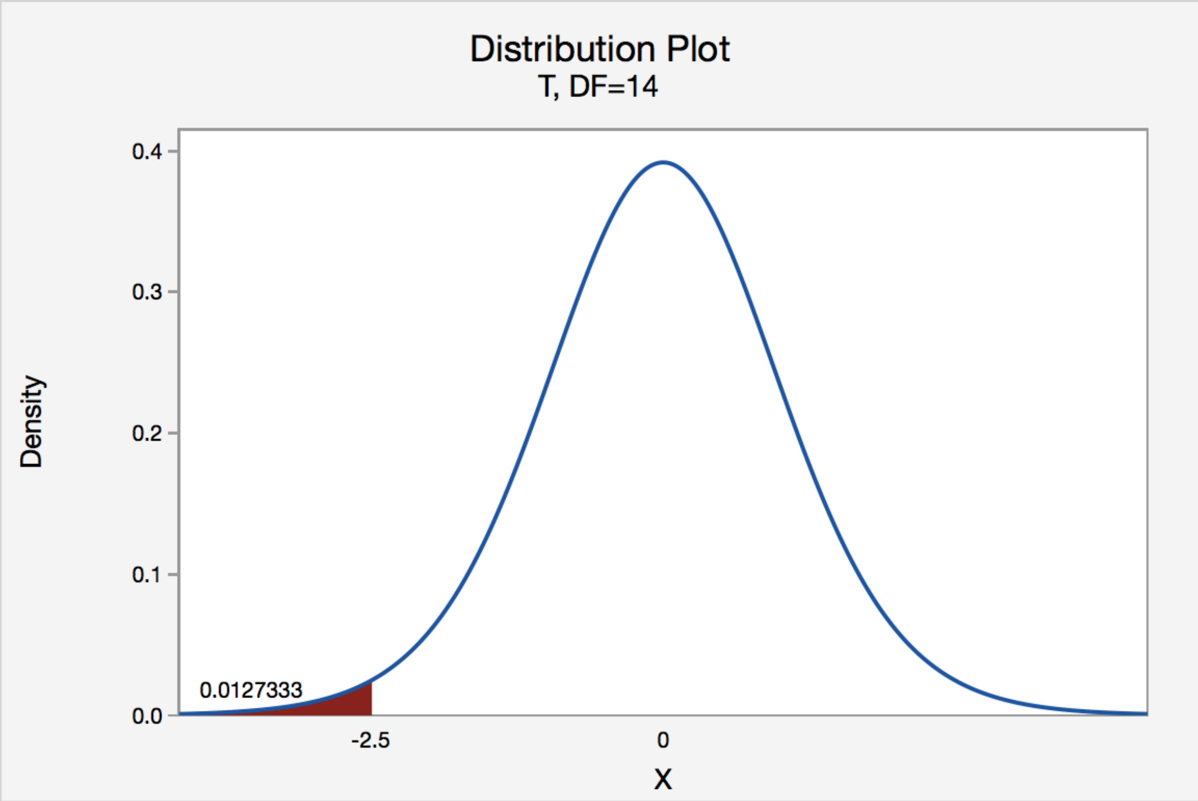


The P-value, 0.0127, tells us it is "unlikely" that we would observe such an extreme test statistic t^* in the direction of H_A if the null hypothesis were true. Therefore, our initial assumption that the null hypothesis is true must be incorrect. That is, since the P-value, 0.0127, is less than α = 0.05, we reject the null hypothesis $H_0 : \mu = 3$ in favor of the alternative hypothesis $H_A : \mu > 3$.

Note that we would not reject $H_0 : \mu = 3$ in favor of $H_A : \mu > 3$ if we lowered our willingness to make a Type I error to α = 0.01 instead, as the P-value, 0.0127, is then greater than α = 0.01.

Left Tailed

In our example concerning the mean grade point average, suppose that our random sample of n = 15 students majoring in mathematics yields a test statistic t^* instead equaling -2.5. The P-value for conducting the **left-tailed** test $H_0 : \mu = 3$ versus $H_A : \mu < 3$ 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 $t_{n-1} = t_{14}$ 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.

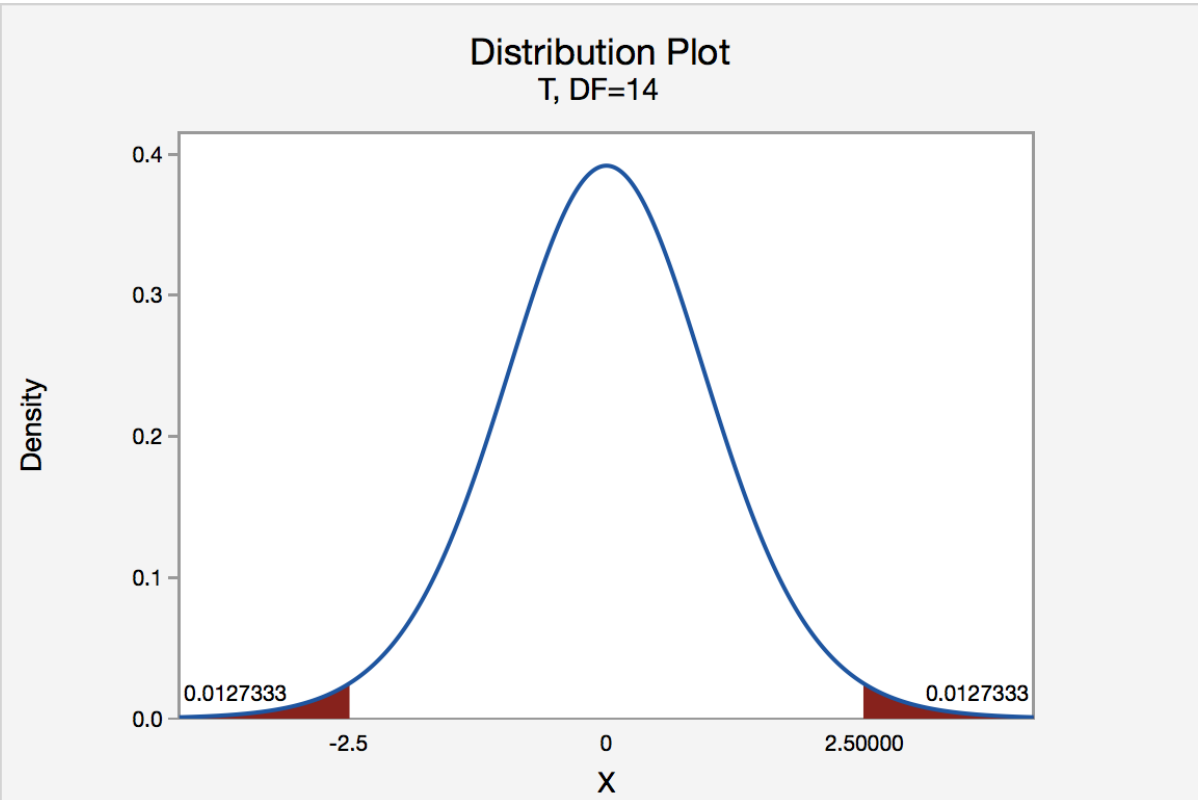


The *P*-value, 0.0127, tells us it is "unlikely" that we would observe such an extreme test statistic t^* in the direction of H_A if the null hypothesis were true. Therefore, our initial assumption that the null hypothesis is true must be incorrect. That is, since the *P*-value, 0.0127, is less than α = 0.05, we reject the null hypothesis $H_0 : \mu = 3$ in favor of the alternative hypothesis $H_A : \mu < 3$.

> Note that we would not reject $H_0 : \mu = 3$ in favor of $H_A : \mu < 3$ if we lowered our willingness to make a Type I error to α = 0.01 instead, as the *P*-value, 0.0127, is then greater than $\alpha = 0.01$.

Two Tailed

In our example concerning the mean grade point average, suppose again that our random sample of $n = 15$ students majoring in mathematics yields a test statistic t^* instead equaling -2.5. The *P*-value for conducting the **two-tailed** test $H_0 : \mu = 3$ versus $H_A : \mu \neq 3$ 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 $t_{n-1} = t_{14}$ 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 H_A 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 $H_0 : \mu = 3$ in favor of the alternative hypothesis $H_A : \mu \neq 3$.

> Note that we would not reject $H_0 : \mu = 3$ in favor of $H_A : \mu \neq 3$ if we lowered our willingness to make a Type I error to α = 0.01 instead, as the *P*-value, 0.0254, is then greater than $\alpha = 0.01$.

Now that we have reviewed the critical value and *P*-value approach procedures for each of three possible hypotheses, let's look at three new examples — one of a right-tailed test, one of a left-tailed test, and one of a two-tailed test.

The good news is that, whenever possible, we will take advantage of the test statistics and *P*-values reported in statistical software, such as Minitab, to conduct our hypothesis tests in this course.