Two terms that students often get confused in statistics are t-values and p-values.To understand the difference between these terms, it helps to understand t-tests.Broadly speaking, there are three different types of t-tests:One-sample t-test: Used to test whether a population mean is equal to some value.Two-sample t-test: Used to test whether two population means are equal.Paired samples t-test: Used to test whether two population means are equal when each observation in one sample can be paired with an observation in the other sample.We use the following steps to perform each test:Step 1: State the null and alternative hypothesis.Step 2: Calculate the t-value.Step 3: Calculate the p-value that corresponds to the t-value.For each test, the t-value is a way to quantify the difference between the population means and the p-value is the probability of obtaining a t-value with an absolute value at least as large as the one we actually observed in the sample data if the null hypothesis is actually true.If the p-value is less than a certain value (e.g. 0.05) then we reject the null hypothesis of the test.For each type of t-test, were interested in the p-value and we simply use the t-value as an intermediate step to calculating the p-value.The following example shows how to calculate and interpret a t-value and corresponding p-value for a two-sample t-test.Example: Calculate & Interpret T-Values and P-ValuesSuppose we want to know whether or not the mean weight between two different species of turtles is equal. We go out and collect a simple random sample of 12 turtles from each population with the following weights:Species #1: 301, 298, 295, 297, 304, 305, 309, 298, 291, 299, 293, 304Species #2: 302, 309, 324, 313, 312, 310, 305, 298, 299, 300, 289, 294Heres how to perform a two-sample t-test using this data:Step 1: State the null and alternative hypothesis.First, well state the null and alternative hypotheses:H0:1= 2 (the two population means are equal)H1:1 2 (the two population means are not equal)Step 2: Calculate the t-value.Next, well enter the weights of each sample of turtles into the Two Sample t-test Calculator and find that the t-value is -1.608761.Step 3: Calculate the p-value.We can also use the Two Sample t-test Calculator to find that the p-value that corresponds to a t-value of -1.608761 is 0.121926.Since this p-value is not less than .05, we fail to reject the null hypothesis.This means we do not have sufficient evidence to say that the mean weight of turtles between the two populations are different.Notice that we simply used the t-value as an intermediate step to calculating the p-value. The p-value is the true value that we were interested in, but we had to first calculate the t-value.Additional ResourcesThe following tutorials offer additional information on t-tests and p-values:An Introduction to the One Sample t-testAn Introduction to the Two Sample t-testAn Introduction to the Paired Samples t-testHow to Calculate a P-Value from a t-Test By Hand Problems with your data, such as skewness and outliers can adversely affect your results. Use the graphs to look for skewness (by examining the spread of each sample) and to identify potential outliers. When data are skewed, the majority of the data are located on the high or low side of the graph. Often, skewness is easiest to detect with a histogram or boxplot. The boxplot with right-skewed data shows wait times. Most of the wait times are relatively short, and only a few wait times are long. The boxplot with left-skewed data shows failure time data. A few items fail immediately, and many more items fail later. Data that are severely skewed can affect the validity of the p-value if your samples are small (either sample is less than 15 values). If your data are severely skewed and you have a small sample, consider increasing your sample size. In these boxplots, the data for Hospital B appear to be severely skewed. Outliers, which are data values that are far away from other data values, can strongly affect the results of your analysis. Often, outliers are easiest to identify on a boxplot. On a boxplot, asterisks (*) denote outliers. Try to identify the cause of any outliers. Correct any dataentry errors or measurement errors. Consider removing data values for abnormal, one-time events (also called special causes). Then, repeat the analysis. For more information, go to Identifying outliers. In these boxplots, the data for Hospital B have 2 outliers.How to use and interpret t-test resultsExplore t-tests: the statistical testing analysis that helps make data insights more reliable. Learn how to use t-tests for confident, data-driven decisions. .pdf version of this pageIn this review, well look at significance testing, using mostly the t-test as a guide. As you read educational research, youll encounter t-test and ANOVA statistics frequently. Part I reviews the basics of significance testing as related to the null hypothesis and p values. Part II shows you how to conduct a t-test, using an online calculator. Part III deal s with interpreting t-test results. Part IV is about reporting t-test results in both text and table formats and concludes with a guide to interpreting confidence intervals.What is Statistical Significance?The terms significance level or level of significance refer to the likelihood that the random sample you choose (for example, test scores) is not representative of the population. The lower the significance level, the more confident you can be in replicating your results. Significance levels most commonly used in educational research are the .05 and .01 levels. If it helps, think of .05 as another way of saying 95/100 times that you sample from the population, you will get this result. Similarly, .01 suggests that 99/100 times that you sample from the population, you will get the same result. These numbers and signs (more on that later) come from Significance Testing, which begins with the Null Hypothesis. Part I: The Null HypothesisWe start by revisiting familiar territory, the scientific method. Well start with a basic research question: How does variable A affect variable B? The traditional way to test this question involves:Step 1. Develop a research question. Step 2. Find previous research to support, refute, or suggest ways of testing the question.Step 3. Construct a hypothesis by revising your research question: HypothesisSummaryTypeH1: A = BThere is no relationship between A and BNullH2: A BThere is a relationship between A and B. Here, there is a relationship, but we dont know if it is positive or negative.AlternateH3: A < BThere is a negative relationship between A and B. Here, the < suggests that the less A is involved, the better B.AlternateH4: A > BThere is a positive relationship between A and B. Here, the > suggests that the more B is involved, the better A.Alternate Step 4. Test the null hypothesis. To test the null hypothesis, A = B, we use a significance test. The italicized lowercase p you often see, followed by > or < sign and a decimal (p .05) indicate significance. In most cases, the researcher tests the null hypothesis, A = B, because is it easier to show there is some sort of effect of A on B, than to have to determine a positive or negative effect prior to conducting the research. This way, you leave yourself room without having the burden of proof on your study from the beginning.Step 5. Analyze data and draw a conclusion. Testing the null hypothesis leaves two possibilities:OutcomeWordingTypeA = BFail to reject the null. We find no relationship between A and B.NullA =, BReject the null. We find a relationship between A and B.AlternateStep 6. Communicate results. See Wording results, below. Part II: Conducting a t-test (for Independent Means)So how do we test a null hypothesis? One way is with a t-test. A t-test asks the question,Is the difference between the means of two samples different (significant) enough to say that some other characteristic (teaching method, teacher, gender, etc.) could have caused it?To conduct a t-test using an online calculator, complete the following steps: Step 1. Compose the Research Question. Step 2. Compose a Null and an Alternative Hypothesis.Step 3. Obtain two random samples of at least 30, preferably 50, from each group.Step 4. Conduct a t-test:Go to #1, check Enter mean, SD and N.For #2, label your groups and enter data. You will need to have mean and SD. N is group size.For #3, check Unpaired t test.For #4, click Calculate now.Step 5. Interpret the results (see below).Step 6. Report results in text or table format (see below).Get p from P value and statistical significance: Note that this is the actual value.Get the confidence interval from Confidence interval:Get the t and df values from Intermediate values used in calculations:Get Mean, and SD from Review your data. Part III. Interpreting a t-test (Understanding the Numbers)ttells you a t-test was used.(98)tells you the degrees of freedom (the sample # of tests performed).3.09is the t statistic the result of the calculation.p .05is the probability of getting the observed score from the sample groups. This the most important part of this output to you.If this signIt means all these thingsp .05likely to be a result of chance (same as saying A = B)difference is not significantnull is correctfail to reject the nullThere is no relationship between A and B.If this signIt means all these things p .05not likely to be a result of chance (same as saying A B)difference is significantnull is incorrectreject the nullThere is a relationship between A and B. Note: We acknowledge that the average scores are different. With a t-test we are deciding if that difference is significant (is it due to sampling error or something else?).Understanding the Confidence Interval (CI)The Confidence Interval (CI) of a mean is a region within which a score (like mean test score) may be said to fall with a certain amount of confidence. The CI uses sample size and standard deviation to generate a lower and upper number that you can be 95% sure will include any sample you take from a set of data.Consider Georgias AYP measure, the CRCT. For a science CRCT score, we take several samples and compare the different means. After a few calculations, we could determine something like. . .the average difference (mean) between samples is -7.5, with a 95% CI of -22.08 to 6.72. In other words, among all students science CRCT scores, 95 out of 100 times we take group samples for comparison (for example by year, or gender, etc.), one of the groups, on average will be 7.5 points lower than the other group. We can be fairly certain that the difference in scores will be between -22.08 and 6.72 points. Part IV. Wording Results Wording Results in TextIn text, the basic format is to report: population (N), mean (M) and standard deviation (SD) for both samples, t value, degrees freedom (df), significance (p), and confidence interval (CI.95)* .Example 1: p .05, or Significant ResultsAmong 7th graders in Lowndes County Schools taking the CRCT reading exam (N = 336), there was a statistically significant difference between the two teaching teams, team 1 (M = 818.92, SD = 16.11) and team 2 (M = 828.28, SD = 14.09), t(98) = 3.09, p .05, CI.95 -15.37, -3.35. Therefore, we reject the null hypothesis that there is no difference in reading scores between teaching teams 1 and 2.Example 2: p .05, or Not Significant ResultsAmong 7th graders in Lowndes County Schools taking the CRCT science exam (N = 336), there was no statistically significant difference between female students (M = 834.00, SD = 32.81) and male students (841.08, SD = 28.76), t(98) = 1.15 p .05, CI.95 -19.32, 5.16. Therefore, we fail to reject the null hypothesis that there is no difference in science scores between females and males. Wording Results in APA Table Format Table 1. Comparison of CRCT 7th Grade Science Scores by Gender Gender n Mean SD t df p 95% Confidence Interval Female50834.0032.81Male50841.0828.76Total100837.5430.901.1498.2540-19.32 5.16Note: On the Web site, this appears blocked and should not be. See the .pdf for the correct format. When we compared mean GCSE scores between boys and girls above, we saw a slight difference. On average, girls have higher mean GCSE scores than boys do. Is this difference large enough to be interesting? Alternatively, you may want to consider whether the difference in means across both sexes is simply due to chance. Because we calculated these means from data from a sample of the population, it may be that the difference in means across both sexes is due to chance. In order to investigate this, we can run a t test to see whether this difference in means is statistically significant. In a t test, like in most tests of significance, the significance threshold is traditionally set at p = 0.05. A p-value is basically the likelihood of finding a mean difference by chance if indeed there is no difference in the population. If in the population there is no difference in GCSE score for males and females, we may have caused there to be a different mean for males and females just by randomly selecting the sample. We can work out the chances of the result we have obtained happening by chance. If a p-value reported from a t test is less than 0.05, then that result is said to be statistically significant. If a p-value is greater than 0.05, then the result is insignificant. Because we have already run frequencies and used a histogram to confirm the normal distribution of our sample, we can run a t test to check for significance. We are going to use the Independent-Samples T Test , because we are interested in comparing the mean GCSE scores across the two unrelated categories male and female in the variable s1gender . Select Analyze , Compare Means , and then Independent-Samples T Test . Move our variable s1gcseptsnew to the Test Variable(s) box. Move s1gender to the Grouping Variable box. Click on Define Groups and enter 1 in the Group 1 box and 2 in the Group 2 box, because 1=Male and 2=Female in the dataset. (We are using the datasets codes for Male and Female. You can check the category codes for s1gender , or any other variable you choose, by finding the variables row in Variable View and clicking to open the Values cell.) After youve entered the information into Define Groups , click Continue and then OK . You should get output like the tables on the left and right. T Test Group Statistics Output From the Group Statistics table on the left, you can see that SPSS has again calculated the number of male and female respondents and the mean scores for both groups. Weve also been provided with the standard deviations for both men and women. In the Independent Samples Test table, SPSS gives us the significance levels of the differences in means. The first test, Levenes Test for Equality of Variances , tests whether the two means have been taken from sample populations that have equal variance. If, for instance, the mean GCSE scores for boys and girls in Sweep 1 of the YCS didnt vary, there would be no mean difference to test. For this test, a p-value of less than 0.05 indicates that there is, in fact, enough variance in the sample to account for possible mean differences. The p-value reported for Levenes Test for Equality of Variance in the table above is p = 0.000, which is well below the 0.05 threshold. So, we can say that equal variance is not assumed for this sample and go on to check the significance level reported in the t test for Equality of Means section. Because we have decided that equal variances are not assumed, we are going to use the p-value presented in the second row of the table. T Test Output Table The p-value for the t test for Equality of Means is 0.000, much lower than the p-value significance threshold of 0.05. This tells us that there is indeed a statistically significant difference in the mean GCSE scores for boys and girls in Sweep 1 of the YCS. Run another two-sample t test to see if there is a statistically significant difference in mean GCSE score between respondents in the variable s1q1e , which concerns whether or not a respondent is enrolled in full time education. (Before you run the test, use the Frequencies function to make sure there are no coded missing values in s1q1e . If there are, recode them.) What are the results of your t test? Are they what you might expect? Summary Youve just run a two sample t test to determine whether or not the differences in mean GCSE scores between girls and boys are statistically significant. You used the Levenes Test for Equality of Means to determine that equal variance was not assumed, and therefore the difference in mean scores is not due to chance and is statistically significant. Note: as we are making changes to a dataset well continue using for the rest of this section, please make sure to save your changes before you close down SPSS. This will save you having to repeat sections youve already completed. If youre not a statistician, looking through statistical output can sometimes make you feel a bit like Alice in Wonderland. Suddenly, you step into a fantastical world where strange and mysterious phantasms appear out of nowhere. For example, consider the T and P in your t-test results.Curiouser and curiouser! you might exclaim, like Alice, as you gaze at your output.What are these values, really? Where do they come from? Even if youve used the p-value to interpret the statistical significance of your resultsumpteen times, its actual origin may remain murky to you.T & P: The Tweedledee and Tweedledum of a T-testT and P are inextricably linked. They go arm in arm, like Tweedledee and Tweedledum. Here's why.When you perform a t-test, you're usually trying to find evidence of a significant difference between population means (2-sample t) or between the population mean and a hypothesized value (1-sample t). The t-value measures the size of the difference relative to the variation in your sample data. Put another way, T is simply the calculated difference represented in units of standard error. The greater the magnitude of T, the greater the evidence against the null hypothesis. This means there is greater evidence that there is a significant difference. The closer T is to 0, the more likely there isn't a significant difference.Remember, the t-value in your output is calculated from only one sample from the entire population. It you took repeated random samples of data from the same population, you'd get slightly different t-values each time, due to random sampling error (which is really not a mistake of any kindit's just the random variation expected in the data).How different could you expect the t-values from many random samples from the same population to be? And how does the t-value from your sample data compare to those expected t-values?You can use a t-distribution to find out.Using a t-distribution to calculate probabilityFor the sake of illustration, assume that you're using a 1-sample t-test to determine whether the population mean is greater than a hypothesized value, such as 5, based on a sample of 20 observations, as shown in the above t-test output.In Minitab, choose Graph > Probability Distribution Plot.Select View Probability, then click OK.From Distribution, select t.In Degrees of freedom, enter 19. (For a 1-sample t test, the degrees of freedom equals the sample size minus 1).Click Shaded Area. Select X Value. Select Right Tail.In X Value, enter 2.8 (the t-value), then click OK.The highest part (peak) of the distribution curve shows you where you can expect most of the t-values to fall. Most of the time, youd expect to get t-values close to 0. That makes sense, right? Because if you randomly select representative samples from a population, the mean of most of those random samples from the population should be close to the overall population mean, making their differences (and thus the calculated t-values) close to 0. Ready for a demo of Minitab Statistical Software? Just ask!T values, P values, and poker handsT values of larger magnitudes (either negative or positive) are less likely. The far left and right "tails" of the distribution curve represent instances of obtaining extreme values of t, far from 0. For example, the shaded region represents the probability of obtaining a t-value of 2.8 or greater. Imagine a magical dart that could be thrown to land randomly anywhere under the distribution curve. What's the chance it would land in the shaded region? The calculated probability is 0.005712.....which rounds to 0.006...which is...the p-value obtained in the t-test results!In other words, the probability of obtaining a t-value of 2.8 or higher, when sampling from the same population (here, a population with a hypothesized mean of 5), is approximately 0.006.How likely is that? Not very! For comparison, the probability of being dealt 3-of-a-kind in a 5-card poker hand is over three times as high (0.021).Given that the probability of obtaining a t-value this high or higher when sampling from this population is so low, whats more likely? Its more likely this sample doesnt come from this population (with the hypothesized mean of 5). It's much more likely that this sample comes from different population, one with a mean greater than 5.To wit: Because the p-value is very low (< alpha level), you reject the null hypothesis and conclude that there's a statistically significant difference.In this way, T and P are inextricably linked. Consider them simply different ways to quantify the "extremeness" of your results under the null hypothesis. You cant change the value of one without changing the other.The larger the absolute value of the t-value, the smaller the p-value, and the greater the evidence against the null hypothesis.(You can verify this by entering lower and higher t values for the t-distribution in step 6 above).Try this two-tailed follow up...The t-distribution example shown above is based on a one-tailed t-test to determine whether the mean of the population is greater than a hypothesized value. Therefore the t-distribution example shows the probability associated with the t-value of 2.8 only in one direction (the right tail of the distribution).How would you use the t-distribution to find the p-value associated with a t-value of 2.8 for two-tailed t-test (in both directions)?Hint: In Minitab, adjust the options in step 5 to find the probability for both tails. If you don't have a copy of Minitab, download a free 30-day trial version. Ever scratched your head over what a p-value really tells you? Or why t-tests seem to pop up in every statistical analysis? You're not alone. These concepts are fundamental tools in data analysis, especially in experiments like A/B testing, but they can feel a bit daunting at first.Let's dive into t-tests and p-values in a way that's easy to grasp. We'll break down what they are, how they work together, and how you can interpret them to make informed decisions. By the end, you'll see how these statistical tools are not just formulas, but valuable assets in understanding your dataand how platforms like Statsig can help make sense of it all.Understanding t-tests and p-valuesT-tests are all about comparing means between groups to see if the differences are statistically significant. They're commonly used in A/B testing and other experiments where you want to test if one group differs from another.On the other hand, p-values indicate the probability of observing results as extreme as those obtained, assuming the null hypothesis is true. Lower p-values suggest stronger evidence against the null hypothesis.Interpreting p-values correctly is crucial. They represent the probability of the data given the null hypothesisnot the probability of the null hypothesis being true (a common misconception). This distinction is key to drawing valid conclusions from your data.It's easy to misinterpret p-values by equating them with error rates or the probability of hypotheses. That's why looking at precise p-values offers more insight into the strength of evidence than just stating significance levels.When you're interpreting p-values in t-tests, remember to consider factors like sample size, variability, and practical significance. Larger samples and effect sizes increase the likelihood of detecting true differences.Setting up a t-test: formulating hypotheses and selecting significance levelsBefore diving into a t-test to interpret p-values, it's important to clearly define your null and alternative hypotheses. The null hypothesis typically states there's no difference between groups or variables, while the alternative hypothesis suggests there is a significant difference.Picking the right significance level (alpha) is also crucial. Common alpha values are 0.05 or 0.01, representing the probability of rejecting a true null hypothesis. A lower alpha value means a more stringent test.To get valid results when interpreting p-values in a t-test, your data needs to meet certain assumptions:Independence: Observations should be independent of each other.Normality: Data should follow a normal distribution.Homogeneity of variance: The variance of the groups should be equal.If these assumptions aren't met, you might end up with inaccurate p-values and incorrect conclusions. In such cases, consider alternative tests or data transformations to proceed with your analysis.Calculating the p-value step-by-step in a t-testTo interpret p-values in a t-test, start by calculating the t-statistic. This involves using your sample data and the t-test formula specific to your test type (one-sample, two-sample, or paired). Essentially, the formula compares the observed difference to the variation in the data.Next, use the t-distribution to find the p-value that corresponds to your t-statistic. The t-distribution is a probability distribution that looks similar to a normal distribution but adjusts for small sample sizes. You can use a t-table or statistical software to find the p-value based on your t-statistic and degrees of freedom.Degrees of freedom (df) are about the number of independent values that can vary in your dataset. In a t-test, df is usually the sample size minus one. Smaller df result in a wider t-distribution, which means you'll need a larger t-statistic to reach the same level of significance.For example, suppose you conduct an independent two-sample t-test with 20 participants in each group. You calculate a t-statistic of 2.5 with 38 degrees of freedom (20 + 20 - 2). Using a t-table or statistical software, you find that the corresponding two-tailed p-value is approximately 0.017. This means that if the null hypothesis is true, there's only a 1.7% chance of observing a difference as extreme as the one in your data.By comparing the p-value to your chosen significance level (usually 0.05), you can decide whether to reject the null hypothesis. In this example, since the p-value (0.017) is less than 0.05, you would conclude that the difference between the two groups is statistically significant. Understanding how to interpret p-values is key for making informed decisions based on your t-test results.Interpreting the p-value and making informed decisionsWhen you're interpreting p-values, it's all about comparing the p-value to your chosen significance level (alpha). If the p-value is less than alpha, you reject the null hypothesis. If it's not, you fail to reject it.A small p-value suggests strong evidence against the null hypothesis, indicating a statistically significant result. But keep in mindstatistical significance doesn't always mean practical significance. So, consider the effect size and the context when interpreting your results.Be cautious of common misconceptions about p-values. A p-value doesn't represent the probability that the null hypothesis is true, nor does it indicate the likelihood of making a Type I error. It only measures how compatible your data is with the null hypothesis.When interpreting p-values in a t-test, remember that the p-value is influenced by factors like sample size, variability, and effect size. A larger sample size, smaller variability, or larger effect size can lead to a smaller p-value, increasing the likelihood of rejecting the null hypothesis.Visualizing the distribution of p-values can give you valuable insights. A p-value histogram with a peak near 0 suggests the presence of significant results, while a uniform distribution indicates mostly null hypotheses. Unusual patterns might point to issues with your test or data.At Statsig, we recognize the importance of correctly interpreting p-values to drive meaningful conclusions. Our platform is designed to help you navigate these statistical waters with confidence, ensuring your experiments yield actionable insights.Closing thoughtsGrasping t-tests and p-values doesn't have to be a chore. With a clear understanding of how they work and how to interpret them, you can make more informed decisions based on your data. Remember to consider the assumptions behind the tests, the context of your findings, and the practical significancenot just the statistical one.If you're eager to dive deeper, check out our other resources on statistical significance and hypothesis testing. At Statsig, we're here to help you turn data into decisions. Happy analyzing!Hope you found this useful!

T value and p value interpretation. T statistic and p value. Difference between p value and t value. T test and p value explained. When to use t value and p value. When to use p value or t test.