Continue

BrainMass: The Original Online Tutor Purchase from thousands of already prepared solutions and responses to specific questions. Just search and see if your question is already here. Browse the Solution Library Can't find it in the Solution Library? Ask your own question and buy a NEW step-by-step explanation, provided to you one-on-one by one of our subject Experts. Academic eBooks that clearly explain Everything You Need to Know about key concepts and issues. Search through and buy the ones that will help you! BrainMass is a community of academic subject Experts that provides online tutoring, homework help and Solution Library services across all subjects, to students of all ages at the University, College and High School levels. Our Academic Experts are all PhDs, Masters or post-graduate level tutors in their subjects, who have all been through our application and screening process. We have helped thousands of students through difficult assignments, courses and exam preparations since we started in 2002. We offer one-on-one Custom Help from our online tutors; just give us your question. We offer a Solution Library of ready-prepared step-by-step solutions for hundreds of thousands of cases, assignments and textbook questions that are available for instant download. And we offer an eBook Library containing our own "Everything You Need to Know" series; designed to help you learn and know everything about key academic concepts and issues. And to back it all up we have an amazing full time Customer Service team ready to help you with any questions you may have as you explore and use our services. Significance Level (α): The probability of incorrectly rejecting the null hypothesis (H0: θ = 0; where θ = μ1 - μ2), also known as the false positive rate or the Type I error rate. An α of 0.05 (5%) means that if we repeated an experiment where we drew samples from the same population many times, we would expect to incorrectly reject the null hypothesis in 5% of cases. α can also be thought of as a measure of how extreme the observed difference in sample means has to be before we reject the null hypothesis. With an α of 0.05, we would reject the null hypothesis when observing a difference that we would expect to see 5% (or less) of the time when drawing two samples from the same population.Statistical Power (1 - β): β is the probability that we will fail to reject the null hypothesis when the samples are drawn from different populations. This is also known as the false negative rate or the Type II error rate. Statistical power or 1 - β is therefore the probablity that we will correctly reject the null hypothesis. In the same way that we can draw samples with different means from the same population, there is also a risk that we draw samples with very similar means from two different populations.Effect Size (Cohen's d): A standardized measure of the difference in the means (can be sample or population means depending on the context). The difference in means is divided by the pooled standard deviation of the two samples/populations to provide a metric, in units of standard deviations, that can be compared across studies. It can also be used directly in some calculations instead of the means and standard deviations of the samples. The calculator below implements the most known statistical test, namely, the Independent Samples t-test or Two samples t-test. t-test, also known as Student's t-test, after William Sealy Gosset. "Student" was his pen name. The test deals with the null hypothesis such that the means of two populations are equal. To put it in other words, the difference we find between the means of the two samples should not significantly differ from zero. Again, the test works only if certain assumptions are met. These are: That the two samples are independently and randomly drawn from the source population(s). That the scale of measurement for both samples has the properties of an equal-interval scale. That the source population(s) can be reasonably supposed to have a normal distribution. And, for this particular implementation of the test, that the variance of each population is the same The calculator displays a level of confidence for both directional and non-directional tests. Let's say you get the result of 96%. Essentially this means that you have 96% confidence that the obtained difference shows something more than simple luck. The chance that you can get the obtained difference and the means of the two samples are the same is only 4%. This is the level of significance you calculate. Now, depending on your chosen level of significance, you can reject or fail to reject your null hypothesis. To estimate the confidence we need to calculate t-value, and then lookup the inverse of CDF of Student's t-distribution with degrees of freedom. is the size of sample A and is the size of sample B. To find t-value you start from calculating the mean and sum of squared deviations, or sum of squares for each sample. Then you estimate the variance of the source population as This estimation is called pooled variance, and it is a method for estimating the variance of several different populations when the mean of each population may be different. Still, one may assume that the variance of each population is the same. Then you estimate the standard deviation of the sampling distribution of sample-mean differences (the "standard error" of) as . Finally, you calculate t as If you care to find more, you can read excellent explanations here, starting from Chapter 9. 26 21 22 26 19 22 26 25 24 21 23 23 18 29 22 18 23 21 20 20 29 20 16 20 26 21 25 17 18 19 Calculation precisionDigits after the decimal point: 1Non-directional hypothesisLevel of confidence for two-tail test of significanceLevel of confidence for one-tail test of significance Page 2 Project statisticsThe file is very large. Browser slowdown may occur during loading and creation. Yes, statistics are not yet impressive, as at the opening on April 25, 2008, there are just one calculator and an article:). However, as time passes, we guarantee improvements!Page 3 Project statisticsThe file is very large. Browser slowdown may occur during loading and creation. Yes, statistics are not yet impressive, as at the opening on April 25, 2008, there are just one calculator and an article:). However, as time passes, we guarantee improvements!Page 4 Project statisticsThe file is very large. Browser slowdown may occur during loading and creation. Yes, statistics are not yet impressive, as at the opening on April 25, 2008, there are just one calculator and an article:). However, as time passes, we guarantee improvements! Previously we have considered how to test the null hypothesis that there is no difference between the mean of a sample and the population mean, and no difference between the means of two samples. We obtained the difference between the means by subtraction, and then divided this difference by the standard error of the difference. If the difference is 196 times its standard error, or more, it is likely to occur by chance with a frequency of only 1 in 20, or less.With small samples, where more chance variation must be allowed for, these ratios are not entirely accurate because the uncertainty in estimating the standard error has been ignored. Some modification of the procedure of dividing the difference by its standard error is needed, and the technique to use is the t test. Its foundations were laid by WS Gosset, writing under the pseudonym “Student” so that it is sometimes known as Student’s t test. The procedure does not differ greatly from the one used for large samples, but is preferable when the number of observations is less than 60, and certainly when they amount to 30 or less.The application of the t distribution to the following four types of problem will now be considered.The calculation of a confidence interval for a sample mean.The mean and standard deviation of a sample are calculated and a value is postulated for the mean of the population. How significantly does the sample mean differ from the postulated population mean?The means and standard deviations of two samples are calculated. Could both samples have been taken from the same population?Paired observations are made on two samples (or in succession on one sample). What is the significance of the difference between the means of the two sets of observations?In each case the problem is essentially the same – namely, to establish multiples of standard errors to which probabilities can be attached. These multiples are the number of times a difference can be divided by its standard error. We have seen that with large samples 1.96 times the standard error has a probability of 5% or less, and 2.576 times the standard error a probability of 1% or less (Appendix table A). With small samples these multiples are larger, and the smaller the sample the larger they become.A rare congenital disease, Everley’s syndrome, generally causes a reduction in concentration of blood sodium. This is thought to provide a useful diagnostic sign as well as a clue to the efficacy of treatment. Little is known about the subject, but the director of a dermatological department in a London teaching hospital is known to be interested in the disease and has seen more cases than anyone else. Even so, he has seen only 18. The patients were all aged between 20 and 44.The mean blood sodium concentration of these 18 cases was 115 mmol/l, with standard deviation of 12 mmol/l. Assuming that blood sodium concentration is Normally distributed what is the 95% confidence interval within which the mean of the total population of such cases may be expected to lie?The data are set out as follows:To find the 95% confidence interval above and below the mean we now have to find a multiple of the standard error. In large samples we have seen that the multiple is 1.96 (Chapter 4). For small samples we use the table of t given in Appendix Table B.pdf. As the sample becomes smaller t becomes larger for any particular level of probability. Conversely, as the sample becomes larger t becomes smaller and approaches the values given in table A, reaching them for infinitely large samples.Since the size of the sample influences the value of t, the size of the sample is taken into account in relating the value of t to probabilities in the table. Some useful parts of the full t table appear in . The left hand column is headed d.f. for “degrees of freedom”. The use of these was noted in the calculation of the standard deviation (Chapter 2). In practice the degrees of freedom amount in these circumstances to one less than the number of observations in the sample. With these data we have 18 – 1 = 17 d.f. This is because only 17 observations plus the total number of observations are needed to specify the sample, the 18th being determined by subtraction.To find the number by which we must multiply the standard error to give the 95% confidence interval we enter table B at 17 in the left hand column and read across to the column headed 0.05 to discover the number 2.110. The 95% confidence intervals of the mean are now set as follows:Mean + 2.110 SE to Mean – 2.110 SEwhich gives us:115 – (2.110 x 283) to 115 + 2.110 x 2.83 or 109.03 to 120.97 mmol/l.We may then say, with a 95% chance of being correct, that the range 109.03 to 120.97 mmol/l includes the population mean.Mean + 2.898 SE to Mean – 2.898 SEwhich gives:115 – (2.898 x 2.83) to 115 + (2.898 x 2.83) or 106.80 to 123.20 mmol/l.Difference of sample mean from population mean (one sample t test)Estimations of plasma calcium concentration in the 18 patients with Everley’s syndrome gave a mean of 3.2 mmol/l, with standard deviation 1.1. Previous experience from a number of investigations and published reports had shown that the mean was commonly close to 2.5 mmol/l in healthy people aged 20-44, the age range of the patients. Is the mean in these patients abnormally high?We set the figures out as follows:t difference between means divided by standard error of sample mean. Ignoring the sign of the t value, and entering table B at 17 degrees of freedom, we find that 2.69 comes between probability values of 0.02 and 0.01, in other words between 2% and 1% and so It is therefore unlikely that the sample with mean 3.2 came from the population with mean 2.5, and we may conclude that the sample mean is, at least statistically, unusually high. Whether it should be regarded clinically as abnormally high is something that needs to be considered separately by the physician in charge of that case.Here we apply a modified procedure for finding the standard error of the difference between two means and testing the size of the difference by this standard error (see Chapter 5 for large samples). For large samples we used the standard deviation of each sample, computed separately, to calculate the standard error of the difference between the means. For small samples we calculate a combined standard deviation for the two samples.The assumptions are:that the data are quantitative and plausibly Normalthat the two samples come from distributions that may differ in their mean value, but not in the standard deviationthat the observations are independent of each other.The third assumption is the most important. In general, repeated measurements on the same individual are not independent. If we had 20 leg ulcers on 15 patients, then we have only 15 independent observations.The following example illustrates the procedure.The addition of bran to the diet has been reported to benefit patients with diverticulosis. Several different bran preparations are available, and a clinician wants to test the efficacy of two of them on patients, since favourable claims have been made for each. Among the consequences of administering bran that requires testing is the transit time through the alimentary canal. Does it differ in the two groups of patients taking these two preparations?The null hypothesis is that the two groups come from the same population. By random allocation the clinician selects two groups of patients aged 40-64 with diverticulosis of comparable severity. Sample 1 contains 15 patients who are given treatment A, and sample 2 contains 12 patients who are given treatment B. The transit times of food through the gut are measured by a standard technique with marked pellets and the results are recorded, in order of increasing time, in Table 7.1 .Table 7.1These data are shown in figure 7.1 . The assumption of approximate Normality and equality of variance are satisfied. The design suggests that the observations are indeed independent. Since it is possible for the difference in mean transit times for A-B to be positive or negative, we will employ a two sided test.Figure 7.1With treatment A the mean transit time was 68.40 h and with treatment B 83.42 h. What is the significance of the difference, 15.02h?The procedure is as follows:Obtain the standard deviation in sample 1: Obtain the standard deviation in sample 2: Multiply the square of the standard deviation of sample 1 by the degrees of freedom, which is the number of subjects minus one:Repeat for sample 2Add the two together and divide by the total degrees of freedomThe standard error of the difference between the means iswhich can be writtenWhen the difference between the means is divided by this standard error the result is t. Thus, The table of the tdistribution Table B (appendix) which gives two sided P values is entered at degrees of freedom.For the transit times of table 7.1, shows that at 25 degrees of freedom (that is (15 – 1) + (12 – 1)), t= 2.282 lies between 2.060 and 2.485. Consequently, this degree of probability is smaller than the conventional level of 5%. The null hypothesis that there is no difference between the means is therefore somewhat unlikely.A 95% confidence interval is given by This becomes83.42 – 68.40 2.06 x 6.58215.02 – 13.56 to 15.02 + 13.56 or 1.46 to 18.58 h.If the standard deviations in the two groups are markedly different, for example if the ratio of the larger to the smaller is greater than two, then one of the assumptions of the ttest (that the two samples come from populations with the same standard deviation) is unlikely to hold. An approximate test, due to Sattherwaite, and described by Armitage and Berry, (1)which allows for unequal standard deviations, is as follows.Rather than use the pooled estimate of variance, compute This is analogous to calculating the standard error of the difference in two proportions under the alternative hypothesis as described in Chapter 6We now compute We then test this using a t statistic, in which the degrees of freedom are:Although this may look very complicated, it can be evaluated very easily on a calculator without having to write down intermediate steps (see below). It can produce a degree of freedom which is not an integer, and so not available in the tables. In this case one should round to the nearest integer. Many statistical packages now carry out this test as the default, and to get the equal variances I statistic one has to specifically ask for it. The unequal variance t test tends to be less powerful than the usual t test if the variances are in fact the same, since it uses fewer assumptions. However, it should not be used indiscriminantly because, if the standard deviations are different, how can we interpret a nonsignificant difference in means, for example? Often a better strategy is to try a data transformation, such as taking logarithms as described in Chapter 2. Transformations that render distributions closer to Normality often also make the standard deviations similar. If a log transformation is successful use the usual t test on the logged data. Applying this method to the data of Table 7.1 , the calculator method (using a Casio fx-350) for calculating the standard error is:When the effects of two alternative treatments or experiments are compared, for example in cross over trials, randomised trials in which randomisation is between matched pairs, or matched case control studies (see Chapter 13), it is sometimes possible to make comparisons in pairs. Matching controls for the matched variables, so can lead to a more powerful study.The test is derived from the single sample t test, using the following assumptions.The data are quantitativeThe distribution of the differences (not the original data), is plausibly Normal.The differences are independent of each other.The first case to consider is when each member of the sample acts as his own control. Whether treatment A or treatment B is given first or second to each member of the sample should be determined by the use of the table of random numbers Table F (Appendix). In this way any effect of one treatment on the other, even indirectly through the patient’s attitude to treatment, for instance, can be minimised. Occasionally it is possible to give both treatments simultaneously, as in the treatment of a skin disease by applying a remedy to the skin on opposite sides of the body.Let us use as an example the studies of bran in the treatment of diverticulosis discussed earlier. The clinician wonders whether transit time would be shorter if bran is given in the same dosage in three meals during the day (treatment A) or in one meal (treatment B). A random sample of patients with disease of comparable severity and aged 20-44 is chosen and the two treatments administered on two successive occasions, the order of the treatments also being determined from the table of random numbers. The alimentary transit times and the differences for each pair of treatments are set out in Table 7.2Table 7.2In calculating t on the paired observations we work with the difference, d, between the members of each pair. Our first task is to find the mean of the differences between the observations and then the standard error of the mean, proceeding as follows:Entering Appendix Table B.pdf at 11 degrees of freedom (n – 1) and ignoring the minus sign, we find that this value lies between 0.697 and 1.796. Reading off the probability value, we see that 0.1A 95% confidence interval for the mean difference is given byIn this case t 11 at P = 0.05 is 2.201 (table B) and so the 95% confidence interval is:-6.5 – 2.201 x 4.37 to -6.5 + 2.201 x 4.37 h. or -16.1 to 3.1h.This is quite wide, so we cannot really conclude that the two preparations are equivalent, and should look to a larger study.The second case of a paired comparison to consider is when two samples are chosen and each member of sample 1 is paired with one member of sample 2, as in a matched case control study. As the aim is to test the difference, if any, between two types of treatment, the choice of members for each pair is designed to make them as alike as possible. The more alike they are, the more apparent will be any differences due to treatment, because they will not be confused with differences in the results caused by disparities between members of the pair. The likeness within the pairs applies to attributes relating to the study in question. For instance, in a test for a drug reducing blood pressure the colour of the patients’ eyes would probably be irrelevant, but their resting diastolic blood pressure could well provide a basis for selecting the pairs. Another (perhaps related) basis is the prognosis for the disease in patients: in general, patients with a similar prognosis are best paired. Whatever criteria are chosen, it is essential that the pairs are constructed before the treatment is given, for the pairing must be uninfluenced by knowledge of the effects of treatment.Suppose we had a clinical trial with more than two treatments. It is not valid to compare each treatment with each other treatment using t tests because the overall type I error rate will be bigger than the conventional level set for each individual test. A method of controlling for this to use a one way analysis of variance .(2)Should I test my data for Normality before using the t test?It would seem logical that, because the t test assumes Normality, one should test for Normality first. The problem is that the test for Normality is dependent on the sample size. With a small sample a non-significant result does not mean that the data come from a Normal distribution. On the other hand, with a large sample, a significant result does not mean that we could not use the t test, because the t test is robust to moderate departures from Normality – that is, the P value obtained can be validly interpreted. There is something illogical about using one significance test conditional on the results of another significance test. In general it is a matter of knowing and looking at the data. One can “eyeball” the data and if the distributions are not extremely skewed, and particularly if (for the two sample t test) the numbers of observations are similar in the two groups, then the t test will be valid. The main problem is often that outliers will inflate the standard deviations and render the test less sensitive. Also, it is not generally appreciated that if the data originate from a randomised controlled trial, then the process of randomisation will ensure the validity of the I test, irrespective of the original distribution of the data.Should I test for equality of the standard deviations before using the usual t test?The same argument prevails here as for the previous question about Normality. The test for equality of variances is dependent on the sample size. A rule of thumb is that if the ratio of the larger to smaller standard deviation is greater than two, then the unequal variance test should be used. With a computer one can easily do both the equal and unequal variance t test and see if the answers differ.Why should I use a paired test if my data are paired? What happens if I don’t?Pairing provides information about an experiment, and the more information that can be provided in the analysis the more sensitive the test. One of the major sources of variability is between subjects variability. By repeating measures within subjects, each subject acts as its own control, and the between subjects variability is removed. In general this means that if there is a true difference between the pairs the paired test is more likely to pick it up: it is more powerful. When the pairs are generated by matching the matching criteria may not be important. In this case, the paired and unpaired tests should give similar results.Armitage P, Berry G. Statistical Methods in Medical Research. 3rd ed. Oxford: Blackwell Scientific Publications, 1994:112-13.Armitage P, Berry G. Statistical Methods in Medical Research. 3rd ed. Oxford: Blackwell Scientific Publications, 1994:207-14.7.1 In 22 patients with an unusual liver disease the plasma alkaline phosphatase was found by a certain laboratory to have a mean value of 39 King-Armstrong units, standard deviation 3.4 units. What is the 95% confidence interval within which the mean of the population of such cases whose specimens come to the same laboratory may be expected to lie?7.2 In the 18 patients with Everley’s syndrome the mean level of plasma phosphate was 1.7 mmol/l, standard deviation 0.8. If the mean level in the general population is taken as 1.2 mmol/l, what is the significance of the difference between that mean and the mean of these 18 patients?7.3 In two wards for elderly women in a geriatric hospital the following levels of haemoglobin were found:Ward A: 12.2, 11.1, 14.0, 11.3, 10.8, 12.5, 12.2, 11.9, 13.6, 12.7, 13.4, 13.7 g/dl;Ward B: 11.9, 10.7, 12.3, 13.9, 11.1, 11.2, 13.3, 11.4, 12.0, 11.1 g/dl.What is the difference between the mean levels in the two wards, and what is its significance? What is the 95% confidence interval for the difference in treatments?7.4 A new treatment for varicose ulcer is compared with a standard treatment on ten matched pairs of patients, where treatment between pairs is decided using random numbers. The outcome is the number of days from start of treatment to healing of ulcer. One doctor is responsible for treatment and a second doctor assesses healing without knowing which treatment each patient had. The following treatment times were recorded.Standard treatment: 35, 104, 27, 53, 72, 64, 97, 121, 86, 41 days;New treatment: 27, 52, 46, 33, 37, 82, 51, 92, 68, 62 days.What are the mean difference in the healing time, the value of t, the number of degrees of freedom, and the probability? What is the 95% confidence interval for the difference? Welcome to the blog Stats and R. As the name suggests, this blog is about statistics and its applications in R (an open source statistical software program). An objective of this blog is to make statistics accessible to everyone! Throughout different articles, I share, explain and illustrate statistical concepts, and when appropriate, apply them in R. From time to time, I also present some work related to data science & data visualization using R, news about my research and, to a smaller extent, my journey in the blogging world. If you are new to this blog, here is what you can do: Follow us onTwitterFacebookYouTubeRSSContent linksAbout usResourcesExplore BMJMy accountEmail alertsActivate subscriptionInformationCopyright © 2025 BMJ Publishing Group Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved. Independent Samples T-Test calculator offers a convenient way to perform Independent Samples T-Tests, accommodating both Equal Variance and Unequal Variance scenarios. Ideal for researchers, data scientists, students, and professionals in various fields, it provides an intuitive interface for inputting sample data, including means, standard deviations, and sample sizes. With just a few clicks, users can effortlessly calculate and view the t-statistics and degrees of freedom. Whether you’re conducting academic research, analyzing business data, or exploring scientific datasets, this tool ensures accuracy and ease in statistical calculations. Independent Two-Sample T-Test Definition The independent two-sample t-test, also known as the independent samples t-test, is used to check if two populations have different means based on two independent samples. The student's t-test is most well known. The student's t-test was developed by William Gosset to help monitor the quality of stout at the Guinness brewery in Dublin, Ireland. The Welch's t-test is the default t-test in the well known statistical software R. The major difference between student's t-test and Welch's t-test is that student's t-test assumes equal variance and Welch's t-test assumes unequal variance. This calculator performs calculations for both student's t-test and Welch's t-test, and also provides recommendation based on the F test for equal variance. The Student’s t-Test is a statistical hypothesis test based on the t-distribution which assesses whether the means of two different groups are statistically different from each other. Thus, the Student’s t-Test can be used to determine if two different sets of data are similar or in fact very different. If the scaling term of the data is known, then normally the z-Test would be utilized to assess the differences; however, if the scaling term is not known, then the t-Test can be used instead. This phenomenon occurs because the t-distribution takes the shape of the normal distribution without standardizing the mean to 0 and the standard deviation to 1 with the scaling term. In statistics, the Student’s t-Test can be used in a variety of ways as long as the data collected originates from two different sources. For example, the t-Test can be used to assess the differences between one sample and a population, two different samples from two different populations (where the variances are equal) and two samples from the same population (one before a particular treatment and one after it). The actual test statistic for the t-distribution is very similar for the z-statistic and it is as follows: t = (x- μo)/(s/(n^1/2)) Where, X is the sample mean μo is the population mean s is the sample standard deviation n is the sample size Thus, understanding the very basics of the Student’s t-Test is critical for accurately assessing the differences between two data sets. © BrainMass Inc. brainmass.com May 11, 2025, 12:28 pm ad1c9bdddf One of the most important test within the branch of inferential statistics is the Student’s t-test.1 The Student’s t-test for two samples is used to test whether two groups (two populations) are different in terms of a quantitative variable, based on the comparison of two samples drawn from these two groups. In other words, a Student’s t-test for two samples allows to determine whether the two populations from which your two samples are drawn are different (with the two samples being measured on a quantitative continuous variable).2 The reasoning behind this statistical test is that if your two samples are markedly different from each other, it can be assumed that the two populations from which the samples are drawn are different. On the contrary, if the two samples are rather similar, we cannot reject the hypothesis that the two populations are similar, so there is no sufficient evidence in the data at hand to conclude that the two populations from which the samples are drawn are different. Note that this statistical tool belongs to the branch of inferential statistics because conclusions drawn from the study of the samples are generalized to the population, even though we do not have the data on the entire population. To compare two samples, it is usual to compare a measure of central tendency computed for each sample. In the case of the Student’s t-test, the mean is used to compare the two samples. However, in some cases, the mean is not appropriate to compare two samples so the median is used to compare them via the Wilcoxon test. This article being already quite long and complete, the Wilcoxon test is covered in a separate article, together with some illustrations on when to use one test or the other. These two tests (Student’s t-test and Wilcoxon test) have the same final goal, that is, compare two samples in order to determine whether the two populations from which they were drawn are different or not. Note that the Student’s t-test is more powerful than the Wilcoxon test (i.e., it more often detects a significant difference if there is a true difference, so a smaller difference can be detected with the Student’s t-test) but the Student’s t-test is sensitive to outliers and data asymmetry. Furthermore, within each of these two tests, several versions exist, with each version using different formulas to arrive at the final result. It is thus necessary to understand the difference between the two tests and which version to use in order to carry out the appropriate analyses depending on the question and the data at hand. In this article, I will first detail step by step how to perform all versions of the Student’s t-test for independent and paired samples by hand. The analyses will be done on a small set of observations for the sake of illustration and easiness. I will then show how to perform this test in R with the exact same data in order to verify the results found by hand. Reminders about the reasoning behind hypothesis testing, interpretations of the p-value and the results, and assumptions of this test will also be presented. Note that the aim of this article is to show how to compute the Student’s t-test by hand and in R, so we refrain from testing the assumptions and we assume all of them are met for this exercise. For completeness, we still mention the assumptions, how to test them and what other tests exist if one is not met. Interested readers are invited to have a look at the end of the article for more information about these assumptions. Before diving into the computations of the Student’s t-test by hand, let’s recap the null and alternative hypotheses of this test: \(H_0\): \(\mu_1 = \mu_2\) \(H_1\): \(\mu_1 e \mu_2\) where \(\mu_1\) and \(\mu_2\) are the means of the two populations from which the samples were drawn. As mentioned in the introduction, although technically the Student’s t-test is based on the comparison of the means of the two samples, the final goal of this test is actually to test the following hypotheses: \(H_0\): the two populations are similar \(H_1\): the two populations are different This is in the general case where we simply want to determine whether the two populations are different or not (in terms of the dependent variable). In this sense, we have no prior belief about a particular population mean being larger or smaller than the other. This type of test is referred as a two-sided or bilateral test. If we have some prior beliefs about one population mean being larger or smaller than the other, the Student’s t-test also allows to test the following hypotheses: \(H_0\): \(\mu_1 = \mu_2\) \(H_1\): \(\mu_1 > \mu_2\) or \(H_0\): \(\mu_1 = \mu_2\) \(H_1\): \(\mu_1 < \mu_2\) In the first case, we want to test if the mean of the first population is significantly larger than the mean of the second, while in the latter case, we want to test if the mean of the first population is significantly smaller than the mean of the second. This type of test is referred as a one-sided or unilateral test. Some authors argue that one-sided tests should not be used in practice for the simple reason that, if a researcher is so sure that the mean of one population is larger (smaller) than the mean of the other and would never be smaller (larger) than the other, why would she needs to test for significance at all? This a rather philosophical question and it is beyond the scope of this article. Interested readers are invited to see part of the discussion in Rowntree (2000). In statistics, many statistical tests is in the form of hypothesis tests. Hypothesis tests are used to determine whether a certain belief can be deemed as true (plausible) or not, based on the data at hand (i.e., the sample(s)). Most hypothesis tests boil down to the following 4 steps:3 State the null and alternative hypothesis. Compute the test statistic, denoted t-stat. Formulas to compute the test statistic differ among the different versions of the Student’s t-test but they have the same structure. See scenarios 1 to 5 below to see the different formulas. Find the critical value given the theoretical statistical distribution of the test, the parameters of the distribution and the significance level \(\alpha\). For a Student’s t-test and its extended version, it is either the normal or the Student’s t distribution (t denoting the Student distribution and z denoting the normal distribution). Conclude by comparing the t-stat (found in step 2.) with the critical value (found in step. 3). If the t-stat lies in the rejection region (determined thanks to the critical value and the direction of the test), we reject the null hypothesis, otherwise we do not reject the null hypothesis. These two alternatives (reject or do not reject the null hypothesis) are the only two possible solutions, we never “accept” an hypothesis. It is also a good practice to always interpret the decision in the terms of the initial question. For the interested reader, see these 4 steps of hypothesis testing in more details in this article. There are several versions of the Student’s t-test for two samples, depending on whether the samples are independent or paired and depending on whether the variances of the populations are (un)equal and/or (un)known: On the one hand, independent samples means that the two samples are collected on different experimental units or different individuals, for instance when we are working on women and men separately, or working on patients who have been randomly assigned to a control and a treatment group (and a patient belongs to only one group). On the other hand, we face paired samples when measurements are collected on the same experimental units, same individuals. This is often the case, for example in medical studies, when testing the efficiency of a treatment at two different times. The same patients are measured twice, before and after the treatment, and the dependency between the two samples must be taken into account in the computation of the test statistic by working on the differences of measurements for each subject. Paired samples are usually the result of measurements at two different times, but not exclusively. Suppose we want to test the difference in vision between the left and right eyes of 50 athletes. Although the measurements are not made at two different time (before-after), it is clear that both eyes are dependent within each subject. Therefore, the Student’s t-test for paired samples should be used to account for the dependency between the two samples instead of the standard Student’s t-test for independent samples. Another criteria for choosing the appropriate version of the Student’s t-test is whether the variances of the populations (not the variances of the samples!) are known or unknown and equal or unequal. This criteria is rather straightforward, we either know the variances of the populations or we do not. The variances of the populations cannot be computed because if you can compute the variance of a population, it means you have the data for the whole population, then there is no need to do a hypothesis test anymore… So the variances of the populations are either given in the statement (use them in that case), or there is no information about these variances and in this case, it is assumed that the variances are unknown. In practice, the variances of the populations are most of the time unknown and the only thing to do in order to choose the appropriate version of the test is to check whether the variances are equal or not. However, we still illustrate how to do all versions of this test by hand and in R in the next sections following the 4 steps of hypothesis testing. Note that the data are artificial and do not represent any real variable. Furthermore, remind that the assumptions may or may not be met. The point of the article is to detail how to compute the different versions of the test by hand and in R, so all assumptions are assumed to be met. Moreover, assume that the significance level \(\alpha = 5\)% for all tests. If you are interested in applying these tests by hand without having to do the computations yourself, here is a Shiny app which does it for you. You just need to enter the data and choose the appropriate version of the test thanks to the sidebar menu. There is also a graphical representation that helps you to visualize the test statistic and the rejection region. I hope you will find it useful! For the first scenario, suppose the data below. Moreover, suppose that the two samples are independent, that the variances \(\sigma^2 = 1\) in both populations and that we would like to test whether the two population means are different. 0.9 1 -0.8 1 0.1 1 -0.3 1 0.2 1 0.8 2 -0.9 2 -0.1 2 0.4 2 0.1 2 So we have: 5 observations in each sample: \(n_1 = n_2 = 5\) mean of sample 1: \(\bar{x}_1 = 0.02\) mean of sample 2: \(\bar{x}_2 = 0.06\) variances of both populations: \(\sigma^2_1 = \sigma^2_2 = 1\) Following the 4 steps of hypothesis testing we have: \(H_0: \mu_1 = \mu_2\) and \(H_1: \mu_1 - \mu_2 e 0\). (\(e\) because we want to test whether the two means are different, we do not impose a direction in the test.) Test statistic: \[z_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}}\] \[= \frac{0.02-0.06-0}{0.632} = -0.063\] Critical value: \(\pm z_{\alpha / 2} = \pm z_{0.025} = \pm 1.96\) (see a guide on how to read statistical tables if you struggle to find the critical value) Conclusion: The rejection regions are thus from \(-\infty\) to -1.96 and from 1.96 to \(+\infty\). The test statistic is outside the rejection regions so we do not reject the null hypothesis \(H_0\). In terms of the initial question: At the 5% significance level, we do not reject the hypothesis that the two population means are the same, or there is no sufficient evidence in the data to conclude that the two populations considered are different. For the second scenario, suppose the data below. Moreover, suppose that the two samples are independent, that the variances in both populations are unknown but equal (\(\sigma^2_1 = \sigma^2_1\)) and that we would like to test whether the mean of population 1 is larger than the mean of population 2. 1.78 1 1.5 1 0.9 1 0.6 1 0.8 1 1.9 1 0.8 2 -0.7 2 -0.1 2 0.4 2 0.1 2 So we have: 6 observations in sample 1: \(n_1 = 6\) 5 observations in sample 2: \(n_2 = 5\) mean of sample 1: \(\bar{x}_1 = 1.247\) mean of sample 2: \(\bar{x}_2 = 0.1\) variance of sample 1: \(s^2_1 = 0.303\) variance of sample 2: \(s^2_1 = 0.315\) Following the 4 steps of hypothesis testing we have: \(H_0: \mu_1 = \mu_2\) and \(H_1: \mu_1 - \mu_2 > 0\). (> because we want to test if the mean of the first population is larger than the mean of the second population.) Test statistic: \[t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\] where \[s_p = \sqrt{\frac{(n_1-1)s^2_1+ (n_2 - 1)s^2_2}{n_1 + n_2 - 2}} = 0.555\] so \[t_{obs} = \frac{1.247-0.1-0}{0.555 * 0.606} = 3.411\] (Note that as it is assumed the variances of the two populations are equal, a pooled (common) variance, denoted \(s_p\), is computed.) Critical value: \(t_{\alpha, n_1 + n_2 - 2} = t_{0.05, 9} = 1.833\) Conclusion: The rejection region is thus from 1.833 to \(+\infty\) (there is only one rejection region because it is a one-sided test). The test statistic lies within the rejection region so we reject the null hypothesis \(H_0\). In terms of the initial question: At the 5% significance level, we conclude that the mean of population 1 is larger than the mean of population 2. For the third scenario, suppose the data below. Moreover, suppose that the two samples are independent, that the variances in both populations are unknown and unequal (\(\sigma^2_1 e \sigma^2_1\)) and that we would like to test whether the mean of population 1 is smaller than the mean of population 2. 0.8 1 0.7 1 0.1 1 0.4 1 0.1 1 1.78 2 1.5 2 0.9 2 0.6 2 0.8 2 1.9 2 So we have: 5 observations in sample 1: \(n_1 = 5\) 6 observations in sample 2: \(n_2 = 6\) mean of sample 1: \(\bar{x}_1 = 0.42\) mean of sample 2: \(\bar{x}_2 = 1.247\) variance of sample 1: \(s^2_1 = 0.107\) variance of sample 2: \(s^2_1 = 0.303\) Following the 4 steps of hypothesis testing we have: \(H_0: \mu_1 = \mu_2\) and \(H_1: \mu_1 - \mu_2 < 0\). (< because we want to test if the mean of the first population is smaller than the mean of the second population.) Test statistic: \[t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}\] \[= \frac{0.42-1.247-0}{0.268} = -3.084\] Critical value: \(-t_{\alpha, \upsilon}\) where \[\upsilon = \frac{\bigg(\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2} \bigg)^2}{\frac{\bigg(\frac{s^2_1}{n_1}\bigg)^2}{n_1 - 1} + \frac{\bigg(\frac{s^2_2}{n_2}\bigg)^2}{n_2 - 1}} = 8.28\] so \[-t_{0.05, 8.28} = -1.851\] Note: The degrees of freedom 8.28 does not exist in the standard Student distribution table, so simply take 8, or compute it in R with qt(p = 0.05, df = 8.28). For simplicity, this number of degrees of freedom is sometimes approximated as \(df = min(n_1 - 1, n_2 - 1)\), so in this case it would be \(df = 4\). Conclusion: The rejection region is thus from \(-\infty\) to -1.851. The test statistic lies within the rejection region so we reject the null hypothesis \(H_0\). In terms of the initial question: At the 5% significance level, we conclude that the mean of population 1 is smaller than the mean of population 2. Student’s t-test with paired samples are a bit different than with independent samples, they are actually more similar to one sample Student’s t-test. Here is how it works. We actually compute the difference between the two samples for each pair of observations, and then we work on these differences as if we were doing a one sample Student’s t-test by computing the test statistic on these differences. In case it is not clear, here is the fourth scenario as an illustration. Suppose the data below. Moreover, suppose that the two samples are dependent (matched), that the variance of the differences in the population is known and equal to 1 (\(\sigma^2_D = 1\)) and that we would like to test whether the mean difference between the two populations is different than 0. 0.9 0.8 -0.8 -0.9 0.1 -0.1 -0.3 0.4 0.2 0.1 The first thing to do is to compute the differences for all pairs of observations: 0.9 0.8 -0.1 -0.8 -0.9 -0.1 0.1 -0.1 -0.2 -0.3 0.4 0.7 0.2 0.1 -0.1 So we have: number of pairs: \(n = 5\) mean of the difference: \(\bar{D} = 0.04\) variance of the difference in the population: \(\sigma^2_D = 1\) standard deviation of the difference in the population: \(\sigma_D = 1\) Following the 4 steps of hypothesis testing we have: \(H_0: \mu_D = 0\) and \(H_1: \mu_D e 0\) Test statistic: \[z_{obs} = \frac{\bar{D} - \mu_0}{\frac{\sigma_D}{\sqrt{n}}} = \frac{0.04-0}{0.447} = 0.089\] (This formula is exactly the same than for one sample Student’s t-test with a known variance, except that we work on the mean of the differences.) Critical value: \(\pm z_{\alpha/2} = \pm z_{0.025} = \pm 1.96\) Conclusion: The rejection regions are thus from \(-\infty\) to -1.96 and from 1.96 to \(+\infty\). The test statistic is outside the rejection regions so we do not reject the null hypothesis \(H_0\). In terms of the initial question: At the 5% significance level, we do not reject the hypothesis that the mean difference between the two populations is equal to 0. For the fifth and final scenario, suppose the data below. Moreover, suppose that the two samples are dependent (matched), that the variance of the differences in the population is unknown and that we would like to test whether a treatment is effective in increasing running capabilities (the higher the value, the better in terms of running capabilities). The first thing to do is to compute the differences for all pairs of observations: 9 16 7 8 11 3 1 15 14 3 12 9 2 9 7 So we have: number of pairs: \(n = 5\) mean of the difference: \(\bar{D} = 8\) variance of the difference in the sample: \(s^2_D = 16\) standard deviation of the difference in the sample: \(s_D = 4\) Following the 4 steps of hypothesis testing we have: \(H_0: \mu_D = 0\) and \(H_1: \mu_D > 0\) (> because we would like to test whether the treatment is effective,