Two Group t-test with Excel

1. Description of t-test

The two-group t-test, also called the independent samples t-test, is designed to determine whether means of a dependent variable differ between two groups, the independent variable. Examples include learning whether vocabulary scores differ between females and males, and whether yearly income differs between high school and college graduates.

The two-group t-test is an inferential test used for hypothesis testing. Why conduct hypothesis tests? We know samples have variability and don't always match the population from which they were drawn. It is possible, for example, for a sample to indicate a relation exists between two variables, or for means to differ between two groups, when neither a relation nor a mean difference exists in the population. Hypothesis tests help us decide whether our sample results appear to be due to random sample variation or whether the results appear to be reflective of real population relations or differences.

With the independent samples t-test the null hypothesis states that two groups have the same mean on a dependent variable. For example, consider the variables memory test performance and student sex. The null states that mean levels of memory test performance will be the same for both females and males. A sample of 5^{th} grade students is taken and females score 70% correct and males score 65% correct. We don't know whether this 70% - 65% = 5 percentage point difference represents sample variation or a real difference in the population. The t-test helps us decide whether the observed mean difference is likely random or real.

The t-test calculates the mean difference between groups, then divides this mean difference by the standard error of the mean. The standard error of the mean is like the standard deviation for scores except that it is calculated for sample means rather than raw scores. This standard error provides an indicator of how means are likely to vary due to random chance, so the t-ratio (i.e., t = (mean difference) / (standard error of mean)) is a ratio of signal to noise and provides an indication of how rare or common the mean difference might be if the null hypothesis of no difference is true. Larger t-ratios, in absolute value, indicate very unusual mean differences if the differences are due to random chance. As a result, when large t-ratios are found, this is an indication that the results would be very unusual if pure sampling variation – chance variation – were the cause of this large mean difference. Thus, the t-test provides a way of gauging whether the mean difference observed is likely or unlikely if the null hypothesis is true; in fact, a probability value, called a p-value, is provided and is discussed and defined below.

2. Free Excel for GSU Students

Calculating the two-group t-test is not difficult, but formulas for the t-test won't be covered in this course. Instead, we will rely on Excel to calculate t-test results. If you don't have Excel, note that current Georgia Southern students may freely download Microsoft Office 365. The link is provided in Folio in the Technology Resources section (see Figure 1).

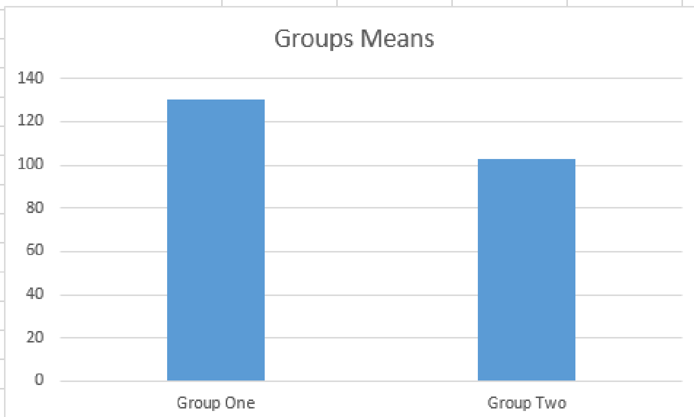
Figure 1: Link to Free MS Office 365

3. Calculating t-test results with Excel

The t-test spreadsheet appears as shown in Figure 2.

Figure 2: Excel Sheet to Calculate t-test Results

	A	B	C	D	E	F	G	H
1	Enter Data Below (up to 300)							
2	Group One	Group Two			Group One	Group Two		
3	95	76		Mean =	130.142857	102.857143		
4	145	120		Standard Deviation =	20.5136424	22.2667807		
5	129	115		Variance =	420.809524	495.809524		
6	138	137		Sample size, n =	7	7		
7	151	90						
8	110	101		Pooled Variance =	458.309524			
9	143	81		SE of Mean Difference =	11.4431455	Note: Standard error (SE) of mean differe		
10				Mean Difference =	27.2857143			
11				degrees of freedom =	12			
12								
13				t -ratio =	2.38445927			
14				p-value for t =	0.034482			



The spreadsheet has been protected to prevent users from accidentally erasing formulas, but if you wish to edit the sheet, it can be unprotected by clicking on the Unprotect icon under Review. There is no password.

One enters data in the two green columns labeled Variable 1 and Variable 2 (see Figure 3).

Figure 3: Data Entry in Green Columns

Enter Data Below (up to 300)	
Group One	Group Two
1	6
2	7
3	8
4	9
5	10

Results are presented to the right of the green columns (see Figure 4).

Figure 4: Results for the t-test

	Group One	Group Two	
Mean =	130.142857	102.857143	
Standard Deviation =	20.5136424	22.2667807	
Variance =	420.809524	495.809524	
Sample size, n =	7	7	
Pooled Variance =	458.309524		
SE of Mean Difference =	11.4431455	Note: Standard	
Mean Difference =	27.2857143		
degrees of freedom =	12		
t -ratio =	2.38445927		
p-value for t =	0.034482		

A histogram is also provided to allow visualization of the group means.

4. Example 1: Systolic Blood Pressure by Sex

To illustrate use of this spreadsheet, we will test whether there is a different in systolic blood pressure between females and males. The blood pressure data are presented below.

Male	Female
95	76
145	120
129	115
138	137
151	90
110	101
143	81

Enter the data in the green columns in the t-test sheet. See Figure 5 below. Group one represents males and group two females.

Figure 5: Example 1 Systolic Blood Pressure

	A	B	C
1	**Enter Data Below (up to 300)**		
2	**Group One**	**Group Two**	
3	95	76	
4	145	120	
5	129	115	
6	138	137	
7	151	90	
8	110	101	
9	143	81	

Results and descriptive statistics are presented to the left of the green columns.

Figure 6: Example 1 Results

	Group One	Group Two	
Mean =	130.142857	102.857143	
Standard Deviation =	20.5136424	22.2667807	
Variance =	420.809524	495.809524	
Sample size, n =	7	7	
Pooled Variance =	458.309524		
SE of Mean Difference =	11.4431455	Note: Standard	
Mean Difference =	27.2857143		
degrees of freedom =	12		
t -ratio =	2.38445927		
p-value for t =	0.034482		

Figure 6 shows that males, group one, have a mean systolic blood pressure of 130.142 and females have a mean of 102.857. The mean difference is 130.142 – 102.857 = 27.28, and this mean difference also appears in Figure 6. Measures of variability, standard deviation and variance, are also presented for both groups. The inferential test information is presented in the last two rows with t-ratio and p-value for t.

The null hypothesis for these variables follows:

> Null: There is no mean difference in systolic blood pressure between females and males.

We wish to know whether the sample data provide enough evidence to determine whether a mean difference is possible in the population. Our sample data show a mean difference of about 27.28 points, with males demonstrating the higher mean blood pressure (130.14 vs. 102.85). What we don't yet know, however, is whether this mean difference is likely the result of random sampling variation or representative of a real difference in the population. The t-test results, specifically the t-ratio and p-value, provide the information needed to make this judgement given this sample. The t-ratio is 2.384 with a p-value of 0.034.

To help decide whether the data are consistent, or inconsistent, with the stated null hypothesis, we use the p-value which is derived from the t-test. The reported p-value, shown in Figure 6, is 0.034. What does this mean?

> p-value = .034: The probability, through random chance, of obtaining a t-ratio for the mean difference this large or larger (in absolute value) if the null hypothesis of no mean difference is true in the population.

What does this mean? To calculate this p-value we must assume that the null is true for the population, so we assume there is no mean systolic blood pressure difference between males and females. If the null is true, then what are the chances of taking a random sample of 20 people and finding a t-ratio of 2.38 or larger in absolute value (i.e., $t \geq 2.38$ or $t \leq -2.38$); the p-value tells us this chance, or probability, and it is .034. If we were to randomly sample 20 people from the population, the chance of getting a sample like we observed, or with a more extreme mean difference, is about 3.4%. The key here is the assumption that the null hypothesis – no differences between females and males – is true in the population. We don't know if it is true, but we do know that if it were true, we could expect to see mean differences this large or larger only about 3.4% of the time. So, if the null is true, then the mean difference we observed is a rare event.

When collecting sample data, like the data in this example, random variability is expected from sample to sample. When random variation occurs, it may create random mean differences among groups. The point of conducting hypothesis tests and examining p-values is to help us decide whether the mean difference we observed is likely due to random chance, or due to something real. The fact that we observed a mean difference 130.142 – 102.857 = 27.28 in a sample does not mean such a difference exists in the population; the hypothesis test helps us decide whether that difference could be due to random variation or due to a real difference in the population.

In hypothesis testing small p-values lead researchers to reject the null and conclude that a difference exists; if the p-value is large, researchers won't reject the null hypothesis and therefore will conclude that no difference exists. Since this p-value is 0.034, which is small, it appears that the sample data are inconsistent with the stated null, so our sample data seem to indicate there is a difference in blood pressure between males and females.

Recall the discussion of hypothesis testing errors. Two error probabilities were presented, alpha (α) and beta (β). Alpha is the probability of a Type 1 error (rejecting the null and claiming there is a difference when in fact there is no difference in the population), and beta is the probability of a Type 2 error (failing to reject a false null and claiming there is no difference when in fact there is a difference in the population).

Researchers use a decision rule when deciding to reject or not reject the null hypothesis:

Decision Rule: If p-value $\leq \alpha$ reject the null; if p-value $> \alpha$ do not reject the null

A common alpha level used for hypothesis testing decisions is .05 which means there is a 5% chance of making a Type 1 error in hypothesis testing.

If alpha = .05, would the null hypothesis of no difference in mean blood pressure be rejected? The p-value reported above for these data was .034, so complete the decision rule:

Decision Rule: If $.034 \leq .05$ reject the null; if $.034 > .05$ do not reject the null

Since .034 is less than .05, the null is rejected and we claim there is a "statistically significant" difference between mean blood pressure between females and males.

5. Example 2: Doctoral Student Anxiety toward the Dissertation Process

Doctoral students were asked to respond to a questionnaire designed to assess two constructs: anxiety toward the dissertation process and efficacy toward the dissertation process. The items represent an attempt to measure students' anxiety (concern, fear, nervousness) and efficacy (confidence, certainty) about undertaking and completing the process of developing and defending a dissertation.

The questionnaire is presented in Display 1. The even-numbered items measure anxiety and the odd-numbered items measure efficacy. For each respondent the mean of responses to the five anxiety items was used to form a composite score for anxiety, and the mean for the efficacy items was used to form a composite score for efficacy.

Display 1: Dissertation Process Questionnaire

The purpose of the questionnaire is to ascertain doctoral students' thoughts about the dissertation process. Your honest responses will help provide a better understanding of doctoral students' experience with this process. In the context of this questionnaire, dissertation process means the entire process students experience to construct and defend the dissertation. This includes, for example, developing the research idea, developing and defending the prospectus, collecting and analyzing data, writing the dissertation, and defending the dissertation before of a committee.

The following 10 statements refer to the dissertation process that you will soon experience. There are no right or wrong answers, so please answer as accurately as possible. Use the scale below to respond to each statement. If you think the statement is very true of you, circle 7; if the statement is not at all true of you, circle 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you.

	not at all true of me						very true of me
1. I believe I will do well on the dissertation.	1	2	3	4	5	6	7
2. I feel uneasy or uncomfortable with the dissertation process as a whole.	1	2	3	4	5	6	7
3. I am confident that I can address even the hardest aspects of the dissertation process.	1	2	3	4	5	6	7
4. Thinking about the upcoming dissertation process makes me feel anxious.	1	2	3	4	5	6	7
5. The process of writing and defending the dissertation may be difficult or hard, but I think I will be successful anyway.	1	2	3	4	5	6	7
6. I am worried about how well I will do during the dissertation defense.	1	2	3	4	5	6	7
7. I know that I have learned the literature and theories that will be necessary to report in the dissertation.	1	2	3	4	5	6	7
8. I feel my heart beating faster as I start to think about the dissertation.	1	2	3	4	5	6	7
9. I am sure that I will be able to answer some of the more challenging or difficult questions posed by the dissertation committee.	1	2	3	4	5	6	7
10. Thinking about the consequences of failing some component of the dissertation process makes me uptight.	1	2	3	4	5	6	7

Below are mean dissertation process anxiety scores for females and males.

Female	Male
4.8	3.8
5.8	4
6	3.8
3.4	4.4
2.4	4.4
6.8	
5.8	
2.6	
4.2	
1.0	
4.6	
1.0	
5.2	
4.6	

Is there any evidence that anxiety differs between female and male doctoral students? The entered data appear below in Figure 7; Group one is female and group two is male.

Figure 7: Dissertation Process Anxiety Scores for Females (Group 1) and Males (Group 2)

	A	B	C
1	Enter Data Below (up to 300)		
2	Group One	Group Two	
3	4.8	3.8	
4	5.8	4	
5	6	3.8	
6	3.4	4.4	
7	2.4	4.4	
8	6.8		
9	5.8		
10	2.6		
11	4.2		
12	1		
13	4.6		
14	1		
15	5.2		
16	4.6		
17			

Descriptive statistics and t-test results are reported in Figure 8.

Figure 8: Results for Dissertation Process Anxiety by Sex (Female - Group One, Male = Group Two)

	Group One	Group Two
Mean =	4.15714286	4.08
Standard Deviation =	1.83333	0.30331502
Variance =	3.3610989	0.092
Sample size, n =	14	5
Pooled Variance =	2.59189916	
SE of Mean Difference =	0.8387583	Note: Standarc
Mean Difference =	0.07714286	
degrees of freedom =	17	
t -ratio =	0.09197269	
p-value for t =	0.927795	

Mean anxiety toward the dissertation process is 4.157 for females and 4.08 for males for a mean difference of 4.157 – 4.08 = 0.077. Does the difference in means appear to reflect a real difference, or could it be due to sampling variation?

The t-ratio for this mean difference is t = 0.091 with a p-value of 0.927. Since this value is greater than alpha of .05, i.e.,

Decision Rule: If .927 ≤ .05 reject the null; if .927 > .05 do not reject the null

we fail to reject the null hypothesis (null: there is no difference in mean dissertation process anxiety between females and males) and conclude, based upon this sample of 19, that there appears to be no difference in levels of anxiety toward the dissertation process between females and males.

6. Review of Hypothesis Testing with t-test in Excel

(a) Enter data in green columns
(b) Examine descriptive statistics and t-test results to assess the difference found, if any
(c) If p-value is less than alpha (e.g., .05), reject the null and claim a difference was identified; if the p-value is greater than .05 do not reject the null and state that the sample data indicate no difference was identified
(d) Interpret the results – if the results are significant (reject null) explain what this means for the two groups; if the results are not significant (null not rejected) explain by indicating the groups appear have similar means scores on the dependent variable, and that any mean difference observed is not large enough to rule out sampling variation.