Two Independent Samples t test

Overview of Tests Presented

Three tests are introduced below: (a) t-test with equal variances, (b) t-test with unequal variances, and (c) equal variance test. Generally one would follow these steps to determine which t-test to use:

1. Perform equal variances test to assess homogeneity of variances between groups,
2. if group variances are equal, then use t-test with equal variances,
3. if group variances are not equal, then use t-test with unequal variances.

These notes begin with presentation of t-test with equal variances assumed, next is information about testing equality of variances, then presented is the t-test with unequal variances.

1. Purpose

The two independent samples t-test enables one to determine whether sample means for two groups differ more than would be expected by chance. The independent variable is qualitative with two categories and the dependent variable must be quantitative (ratio, interval, or sometimes ordinal).

Example 1: Is there a difference in mean systolic blood pressure between males and females in EDUR 8131?

(Note: IV = sex [male vs. female], DV = blood pressure.)

Example 2: Does intrinsic motivation differ between students who are given an opportunity to provide instructional feedback and students who not given an opportunity to provide instructional feedback?

(Note: IV = feedback opportunity [yes vs. no], DV = intrinsic motivation. Background: Two weeks into a semester an instructor asked students to provide written feedback on instruction with suggestions for improvement.)

2. Steps of Hypothesis Testing

Like with the one-sample t test, the two-sample t test follows the same steps for hypothesis testing:

a. Define both H_0 and H_1
b. Set alpha (α, probability of a Type 1 Error)
c. Identify decision rule (either for α, test statistic, or confidence interval)
d. Calculate the test statistic (t ratio)
e. Find df (degrees of freedom), confidence intervals, and p-values
f. Present inferential and interpretation of results (in APA style)

3. Hypotheses

(a) Example 1

Is there a difference in mean systolic blood pressure between males and females in EDUR 8131?

Null

Written

The mean systolic blood pressure for males and females in EDUR 8131 is equal.

Symbolic

H_0: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$

Alternative Non-directional

Written

The mean systolic blood pressure for males and females in EDUR 8131 differ.

Symbolic

H_1: $\mu_1 \neq \mu_2$ or H_1: $\mu_1 - \mu_2 \neq 0$

(b) Example 2

Does intrinsic motivation differ between students who are given an opportunity to provide instructional feedback and students who not given an opportunity to provide instructional feedback?

Null

Written

There is no difference in mean intrinsic motivation between those given the opportunity to provide feedback and those not given the opportunity to provide feedback.

Symbolic

H_0: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$

Alternative Non-directional

Written

Mean intrinsic motivation differs between those given the opportunity to provide feedback and those not given the opportunity to provide feedback.

Symbolic

H_1: $\mu_1 \neq \mu_2$ or H_1: $\mu_1 - \mu_2 \neq 0$

4. Decision Rules

Decision rules for the two-sample t test are listed below.

(a) p-values and α

As noted for the one-sample t test, the probability of committing a Type 1 error, α, is normally set to .10, .05, or .01.

Decision rule for p-values (this decision rule holds for all statistical tests in which p-values are present) states

If $p \leq \alpha$ reject H_0, otherwise fail to reject H_0

A p-value is calculated for the sampled data assuming the null hypothesis is true. A p-value for the two-sample t test may be interpreted as follows:

> *p-value*: The probability of obtaining, randomly, two sample means from the population that differ by the amount observed, or more, assuming the null hypothesis is true (i.e., H_0: $\mu_1 - \mu_2 = 0$) given the size of the samples selected.
>
> *p-value simplified*: The probability of randomly obtaining results this discrepant from the null hypothesis, or more discrepant, assuming the null hypothesis is true.

Assuming there is really no difference between group means in the population (i.e., $\mu_1 = \mu_2$), the p-value indicates the probability of obtaining random sample means that will deviate this far or further from each other.

(b) Critical t values (t_{crit}) and calculated t

An equivalent hypothesis approach is to compare calculated t-values against critical t-values:

> *If $-t \leq -t_{crit}$ or $t \geq t_{crit}$ reject H_0, otherwise fail to reject H_0*

This can be simplified by using absolute values as follows:

$$If\ |t| \geq |t_{crit}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

(c) Confidence Intervals

Since the null hypothesis usually specifies no difference between group means,

> H_0: $\mu_1 - \mu_2 = 0$

this indicates that the population difference between means should be 0.00 if the null is true. Note, it is possible to specify that means will be different in the null (e.g., H_0: $\mu_1 - \mu_2 = 6.35$) but this is rarely the case with hypothesis testing.

One may calculate a confidence interval for the mean difference, and this interval provides a range of possible values for the population mean difference. Given this, one may specify the following decision rule:

> *If 0.00 lies within the CI for the mean difference then fail to reject H_0, otherwise reject H_0*

The logic here is that if 0.00 is within the confidence interval this suggests that 0.00 is a possible value for the population mean difference, and if 0.00 is a possible value for the population mean difference then the data collected are consistent with the null hypothesis of no difference, i.e., the data do not contradict H_0: $\mu_1 - \mu_2 = 0$. If, however, 0.00 is not within the confidence interval, this suggests 0.00 is not a reasonable value for the population mean difference so the collected data are not consistent with the null hypothesis therefore one may reject the null.

If one specified a value other than 0.00 in the null, then one would perform hypothesis testing using the specified value. For example, if one specified H_0: $\mu_1 - \mu_2 = 6.35$, then if the value 6.35 does not lie within the confidence interval one would reject H_0.

5. t-test with Equal Variances: t-ratio, Assumptions, SE_d, and df

There are several ways to calculate the standard error of the mean difference, SE_d, in a two-group t-test. One approach assumes groups have equality of variances or homogeneous variances (i.e., $H_0\colon \sigma_1^2 = \sigma_2^2$), and the second approach relaxes this assumption and allows for unequal group variances. The second approach is presented later in these notes.

(a) t-ratio Formula

Like the one-sample t test, the two-sample t-test forms a ratio of mean differences divided by the standard error of that difference:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

Since the difference between population means is expected to be 0.00, the component μ_1-μ_2 drops from the equation leaving the reduced equation for the t-ratio:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{SE_d}$$

(Note: If the mean difference expected in the null hypothesis does not have to be 0.00 but could be any expected difference [e.g., H_0: $\mu_1 - \mu_2 = 3.00$]. While it is rare that one hypothesizes a mean difference value other than 0.00, be aware that the component μ_1-μ_2 would not drop from the equation when something other than 0.00 is expected.)

(b) Assumptions

For the independent samples t-test to provide accurate calculated t-ratios, p-values, and confidence intervals, three assumptions are required:

1. Scores sampled from population distributions that are normally distributed,
2. population scores are independent,
3. and both groups have population variances that are homogeneous, or equal (i.e., $\sigma_1^2 = \sigma_2^2$), .

It is also important, as with all statistical estimates, that samples not contain influential outliers, otherwise all parameter estimates and inferential statistics can be inaccurate.

The first assumption, normal distribution, tends to be of little importance with larger samples due to the central limit theorem, so violations of this assumption often have little impact on calculated t-ratios, p-values, or confidence intervals.

The second assumption is important; if scores are not independent then the "independent samples t-test" is likely the wrong test to use. Depending upon the nature of lack of independence, other statistical procedures may be more appropriate such as the correlated samples t-test. Violation of this assumption will also lead to inaccurately calculated standard errors, t-ratios, p-values, and confidence intervals.

Violations to the third assumption can also impact calculated standard errors, t-ratios, p-values, and confidence intervals. However, when sample sizes are equal or approximately equal, or when sample sizes are large (whether equal or not), then violations of this assumption typically have little impact. However, for small sample sizes unequal variance can greatly affect p-values, t-ratios, and confidence intervals.

The t-test can be corrected for violation to the homogeneous variance assumption. This correction is discussed below in the presentation of t-test for unequal variances. The next section, however, shows how to calculate a t-ratio when variances are assumed equal or approximately equal.

(c) Standard Error of Difference, SE_d (Equal Variances Assumed)

In most cases the following formulas for SE_d and df will be appropriate when sample sizes are equal or approximately equal, or when variances are equal or approximately equal. When equal or approximately equal variances exist between the two groups, one may take the weighted mean of variances and this mean is called the pooled estimate of the population variance and is symbolized by s_p^2 (where p = pooled). A weighted mean is an average that takes into account sample sizes for each group such that the larger group's variance will influence more the pooled variance estimate. The pooled variance is found by this formula:

$$s_p^2 = \left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)$$

Using the pooled variance, s_p^2, the standard error of the mean difference , SE_d or $s_{\bar{x}_1 - \bar{x}_2}$, is calculated as follows:

$$SE_d = s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

(d) Degrees of Freedom, df (Equal Variances Assumed)

When the pooled estimate of the SE_d is employed, one may calculate degrees of freedom as

df (or ν) = $n_1 + n_2 - 2$

6. Critical Values (Equal Variances Assumed)

If the two groups have approximately equal variances, or if they have approximately equal sample sizes (since equal sample sizes negate any problems due to lack of equal variances), then df = $n_1 + n_2 - 2$.

Like for the one-sample t-test, critical values can be found in a table of critical t-values for identified α and df. A table of critical t-values can be found on the course web page.

Example for Finding df

(a) Compare systolic blood pressure between males and females; there are seven males and seven females, what would be the critical t-value for this analysis if α = .05? What if α = .01?

df = $n_1 + n_2 - 2 = 7 + 7 - 2 = 12$

α = .05; critical t = ± 2.18
α = .01; critical t = ± 3.05

(b) Motivation differences are expected between those with instructional feedback (n = 7) and those without instructional feedback (n = 12). What would be the critical t-values for α = .10, α = .05, and α = .01?

df = $n_1 + n_2 - 2 = 7 + 12 - 2 = 17$

α = .10; critical t = ± 1.74
α = .05; critical t = ± 2.11
α = .01; critical t = ± 2.90

(c) Is there a difference in introductory statistics final grades for students on-line (n=48) vs. students in face-to-face classes (n = 53)? What would be the critical t-values for α = .10, α = .05, and α = .01?

df = $n_1 + n_2 - 2 = 48 + 53 - 2 = 99$

α = .10; critical t = ± 1.66
α = .05; critical t = ± 1.99
α = .01; critical t = ± 2.63

7. Example t-tests in Excel and SPSS (Equal Variances Assumed)

(a) Example 1

Is there a difference in mean systolic blood pressure between males and females in EDUR 8131? Raw data, descriptive statistics, and t-test results for both groups are provided below.

Table 1
Systolic Blood Pressure by Sex

Male		Female	
95	151	76	90
145	110	120	101
129	143	115	81
138		137	
M =	130.143	M =	102.857
s =	20.514	s =	22.267
n =	7	n =	7

If calculating by hand, follow these steps:

(a) Determine mean and sample size for both groups.

(b) Find s^2, variances, for both groups (use of standard deviation rather than variance is a common error):

$$s^2_{male} = 20.514^2 = 420.824$$
$$s^2_{female} = 22.267^2 = 495.819$$

(c) Find $s^2_{p,}$ pooled estimate of the population variance:

$$s_p^2 = \left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}\right) = \left(\frac{(7-1)420.824 + (7-1)495.819}{7+7-2}\right) = \left(\frac{(6)420.824 + (6)495.819}{12}\right)$$

$$= \left(\frac{2524.944 + 2974.916}{12}\right) = \left(\frac{5499.86}{12}\right) = 458.322$$

(d) Find SE_d or $s_{\bar{x}_1 - \bar{x}_2}$, standard error of the mean difference:

$$SE_d = \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{458.322\left(\frac{1}{7}+\frac{1}{7}\right)} = \sqrt{458.322\left(\frac{2}{7}\right)} = \sqrt{458.322(.2857)}$$

$$= \sqrt{130.843} = 11.443$$

(e) Find the t-ratio, t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE_d} = \frac{130.143 - 102.857}{11.443} = \frac{27.286}{11.443} = 2.385$$

(f) Find df and critical values for whichever significance level (α) is specified if using t-values for hypothesis testing:

df = $n_1 + n_2 - 2 = 7 + 7 - 2 = 12$

α = .05; critical t = ± 2.18
α = .01; critical t = ± 3.05

(g) Apply decision rule

$$If\ |t| \geq |t_{crit}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

$$If\ |\mathbf{2.385}| \geq |\mathbf{2.18}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since the calculated t-value of 2.385 is larger than the critical t-value of 2.18 reject Ho and conclude that males have higher blood pressure than females since males' blood pressure mean is higher.

Results from SPSS are presented below.

Figure 1
SPSS Output for Two-group t-test of Blood Pressure by Sex

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
blood_pressure	Female	7	102.8571	22.26678	8.41605
	Male	7	130.1429	20.51364	7.75343

Use this row for equal variance t-test results.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
blood_pressure	Equal variances assumed	.122	.733	-2.384	12	.034	-27.28571	11.44315	-52.21819	-2.35324
	Equal variances not assumed			-2.384	11.920	.035	-27.28571	11.44315	-52.23671	-2.33472

Questions

(a) Why does SPSS show a t-value of -2.38 while the calculated t-value above was positive, t= 2.38? Does this affect inference and interpretation?

Reason – simply which group is listed first in the t-ratio formula.
No – no, inference will remain the same (e.g., same p-value), interpretation – be clear which group has the higher mean.

(b) What is the p-value specified by SPSS for this t-test and what would be the p-value decision rule?

p = .034 (see "Sig. (2-tailed)" in SPSS output.
If $p \leq \alpha$ reject H_0, otherwise fail to reject H_0
*If **.034 ≤ .05** reject H_0, otherwise fail to reject H_0* ***(p is less than alpha, so reject Ho)***

(b) Example 2

Does intrinsic motivation differ between students who are given an opportunity to provide instructional feedback to their instructor and students who not given an opportunity to provide instructional feedback? Intrinsic motivation is measured as a mean composite of several items each of which is on a scale that ranges from 1 = low to 5 = high. For example students rate the degree to which the subject matter is of interest to them, how easy it is for them to pay attention to the subject matter, and their ability to stay focused when studying this content.

Raw data and descriptive statistics for intrinsic motivation are presented below in Table 2.

Table 2
Intrinsic Motivation by Feedback Opportunity

Feedback Opportunity: Yes		Feedback Opportunity: No	
4.2	4.6	1.1	1.3
4.5		2.3	3.1
4.3		3.8	4.7
3.9		4.8	2.9
4.3		4.2	2.9
3.9		4.9	1.2
M =	4.243	M =	3.100
s =	0.270	s =	1.413
n =	7	n =	12

If calculating by hand, follow these steps:

(a) Determine mean and sample size for both groups.

(b) Find s^2, variances, for both groups (use of standard deviation rather than variance is a common error):

$$s_{yes}^2 = 0.27^2 = 0.0729$$
$$s_{no}^2 = 1.413^2 = 1.9966$$

(c) Find $s_{p,}^2$ pooled estimate of the population variance:

$$s_p^2 = \left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}\right) = \left(\frac{(7-1)0.0729 + (12-1)1.9966}{7 + 12 - 2}\right) = \left(\frac{(6)0.0729 + (11)1.9966}{17}\right)$$

$$= \left(\frac{0.4374 + 21.9626}{17}\right) = \left(\frac{22.4}{17}\right) = 1.3176$$

(d) Find SE_d or $s_{\bar{x}_1 - \bar{x}_2}$, standard error of the mean difference:

$$SE_d = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{1.3176\left(\frac{1}{7} + \frac{1}{12}\right)} = \sqrt{1.3176\left(\frac{19}{84}\right)} = \sqrt{1.3176(.22619)}$$

$$= \sqrt{0.298} = 0.5459$$

(e) Find the t-ratio, t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE_d} = \frac{4.243 - 3.1}{0.5459} = \frac{1.143}{0.5459} = 2.0937$$

(f) Find df and critical values for whichever significance level (α) is specified if using t-values for hypothesis testing:

df = $n_1 + n_2 - 2$ = 7 + 12 – 2 = 17

α = .05; critical t = ± 2.11
α = .01; critical t = ± 2.90

(g) Apply decision rule

$$If\ |t| \geq |t_{crit}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

$$If\ |\mathbf{2.094}| \geq |\mathbf{2.11}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since the calculated t-value of 2.094 is less than the critical t-value of 2.11 fail to reject Ho and conclude intrinsic motivation appears to be similar for students whether allowed or not allowed to provide instructional feedback.

Results from SPSS are presented below.

Figure 2
SPSS Output for Two-group t-test of Intrinsic Motivation by Feedback Opportunity

Group Statistics

	feedback	N	Mean	Std. Deviation	Std. Error Mean
motivation	1.00	7	4.2429	.26992	.10202
	.00	12	3.1000	1.41293	.40788

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
motivation	Equal variances assumed	10.712	.004	2.094	17	.052	1.14286	.54589	-.00888	2.29459
	Equal variances not assumed			2.718	12.331	.018	1.14286	.42044	.22951	2.05620

Question

What is the p-value specified by SPSS for this t-test and what would be the p-value decision rule?

SPSS calls the p-value for the t-test "Sig. (2-tailed)" and it equals 0.052 in this example.
If $p \leq \alpha$ *reject* H_0*, otherwise fail to reject* H_0
*If **.052** ≤ **.05** reject* H_0*, otherwise fail to reject* H_0 (since .052 is larger than .05 fail to reject Ho)

8. Results of Heterogeneity of Variance (Unequal Variances between Groups)

The formula for the standard of the mean difference, SE_d or $s_{\bar{x}_1 - \bar{x}_2}$, presented above assumes that both groups have the same or similar variances. If the variances are largely different, then the standard error for the mean difference can be over- or underestimated and result in incorrect inferential decisions due to inaccurate t-ratios, p-values, and confidence intervals.

Consider, for example, the SPSS t-test results for intrinsic motivation presented in Figure 3. Note that the variances for the two groups are 0.073 and 1.996 – the ratio of these two variances is 1.996 / 0.073 = 27.3, so one variance is 27x larger than the second variance.

Figure 3

SPSS Output for Two-group t-test of Intrinsic Motivation by Feedback Opportunity

Group Statistics

	feedback	N	Mean	Std. Deviation	Std. Error Mean
motivation	1.00	7	4.2429	.26992	.10202
	.00	12	3.1000	1.41293	.40788

Variances are:

sd = 0.2699 and var = 0.073

sd = 1.4129 and var = 1.996

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
motivation	Equal variances assumed	10.712	.004	2.094	17	.052	1.14286	.54589	-.00888	2.29459
	Equal variances not assumed			2.718	12.331	.018	1.14286	.42044	.22951	2.05620

When group variances are so different, and when sample sizes are unequal, the inequality of variances can produce misleading inferential results. For example, note the t-values and p-values for the two rows of t-test results in Figure 3. The top row assumes equal variances and the bottom row adjusts for unequal variances. With $\alpha = .05$, these results lead to different inferential conclusions: **fail to reject Ho** if equal variances assumed and **reject Ho** if equal variances are not assumed.

If sample sizes are unequal, or not very similar, and if group variances are largely different (i.e., one variance is about twice the size of the other), then one should test for homogeneity of variances and determine whether inferential results and confidence intervals differ between the two t-test results.

9. Assessing Homogeneity of Variances

A first step in conducting a t-test should include determining whether the assumption of homogeneity of variances (equal group variances) holds. Below are two methods for testing this assumption.

(a) Levene's Test for Equality of Variances (SPSS Default)

Levene's test assesses whether group variances differ more than would be expected by chance. The null and alternative hypotheses are

$$H_0: \sigma_a^2 = \sigma_b^2$$
$$H_1: \sigma_a^2 \neq \sigma_b^2$$

where subscripts *a* and *b* represent group A and B. In words, Ho indicates that the two samples come from populations with equal variances.

Example 1: Blood Pressure

Do males and females appear to have similar variances for blood pressure?

$$H_0: \sigma_{males}^2 = \sigma_{females}^2$$
$$H_1: \sigma_{males}^2 \neq \sigma_{females}^2$$

Figure 4 shows SPSS results for Levene's test (see area highlighted in red).

Figure 4

SPSS Output for Two-group t-test of Blood Pressure by Sex

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
blood_pressure	Female	7	102.8571	22.26678	8.41605
	Male	7	130.1429	20.51364	7.75343

Test for equal variances.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
blood_pressure	Equal variances assumed	.122	.733	-2.384	12	.034	-27.28571	11.44315	-52.21819	-2.35324
	Equal variances not assumed			-2.384	11.920	.035	-27.28571	11.44315	-52.23671	-2.33472

The Levene's test reports an F ratio and a p-value (denoted "Sig." by SPSS). A large F ratio signifies large differences between the two group variances and small F ratio indicates little difference between variances. An F-ratio of 0.00 means the two group variances are equal; F-ratios cannot be less than 0.00. For the example provided in Figure 4, the F-ratio = 0.122 and the corresponding p-value is 0.733; this is a small F-ratio and indicates the group variances appear to be similar.

The F-ratio's p-value can be used to determine whether the null, $H_0: \sigma^2_{males} = \sigma^2_{females}$, should be rejected. As with all p-values, if $p \leq \alpha$ reject H_0 otherwise fail to reject. For tests of variance equality many statisticians recommend using a Type 1 error rate of 0.10 rather than 0.05. In this example the p-value is 0.733 which is larger than 0.10 so one would fail to reject Ho and conclude that group variances are similar (or equal).

Example 2: Intrinsic Motivation

Is there a difference in variances for intrinsic motivation between those given instructional feedback opportunities and those not given such opportunities?

$$H_0: \sigma^2_{feedback} = \sigma^2_{no\ feedback}$$
$$H_1: \sigma^2_{feedback} \neq \sigma^2_{no\ feedback}$$

Figure 5 shows SPSS results for Levene's test (see area highlighted in red) for equal variances for intrinsic motivation. The Levene's test reports an F-ratio = 10.712 and p-value = 0.004.

In this example the p-value of 0.004 is less than an alpha error rate of 0.10 so one would reject Ho and conclude that group variances are unequal.

Figure 5
SPSS Output for Two-group t-test of Intrinsic Motivation by Feedback Opportunity

Group Statistics

	feedback	N	Mean	Std. Deviation	Std. Error Mean
motivation	1.00	7	4.2429	.26992	.10202
	.00	12	3.1000	1.41293	.40788

Test for equal variances.

Use this row if variances are not equal.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
motivation	Equal variances assumed	10.712	.004	2.094	17	.052	1.14286	.54589	-.00888	2.29459
	Equal variances not assumed			2.718	12.331	.018	1.14286	.42044	.22951	2.05620

Since the groups appear to have unequal variances, one should use a t-test that corrects for unequal variances.

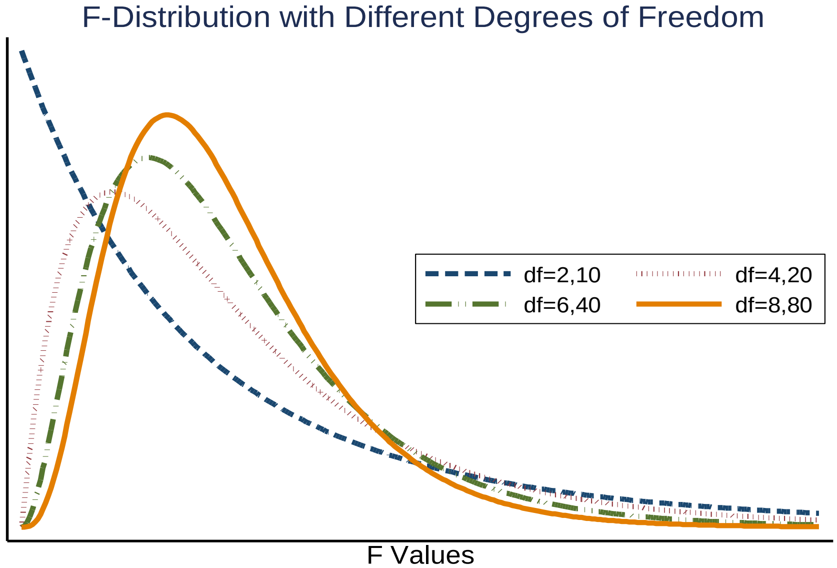
SPSS provides this correction with results denoted as "Equal variances not assumed" – see section highlighted in blue in Figure 5.

Note that with the corrected t-test (i.e., unequal variances t-test) results the inferential statistics are t = 2.718 and p = 0.018. Given these results, one would reject H_0: $\mu_1 = \mu_2$ and conclude that Intrinsic Motivation does appear to differ between Instructional Feedback groups. This inference differs from the inference provided above when equal group variances were assumed.

(b) F-max Test: Useful for Hand Calculation Assessment of Homogeneity of Variance

If homogeneity of variances is not automatically tested by software employed, then one may assess H_0: $\sigma_a^2 = \sigma_b^2$ by using the F-max test. This simple test requires only group variances and sample sizes. F-max is based upon the F distribution which is positively skewed, cannot be less than zero, and takes a different shape depending upon the degrees of freedom available. Examples of F distributions are presented in Figure 6.

Figure 6
Examples of Several F Distributions for Varying Degrees of Freedom



The F-ratio for the F-max test is calculated as

$$F = \frac{s^2_{largest}}{s^2_{smallest}}$$

where $s^2_{largest}$ is the group with the largest variance and $s^2_{smallest}$ is the group with the smallest variance. With all F tests there are two different degrees of freedom,

$$df_1(also\ df_{k1}\ and\quad df_{numerator}) = n_{large\ var.} - 1\ (n\ for\ the\ group\ with\ largest\ variance, s^2_{largest})$$
$$df_2(also\ df_{k2}\ and\ df_{denominator}) = n_{small\ var.} - 1\ (n\ for\ the\ group\ with\ smallest\ variance, s^2_{smallest})$$

One uses df_1 and df_2 to find critical F values against which are compared calculated F ratios for the F-max test.

<u>Critical F values</u> are located in Figure 7. To find critical values one matches the column (df_1) and row (df_2) and selects the F value that results. If the calculated df does not exist, round down to the nearest tabled df value.

Examples
(a) $df_1 = 20$ and $df_2 = 17$; the critical F value is 1.86.
(b) $df_1 = 9$ and $df_2 = 55$; since $df_2 = 55$ is not tabled select 50 for df_2; the critical F value is 1.76.
(c) $df_1 = 29$ and $df_2 = 221$; there is no $df_1 = 29$ tabled so select $df_1 = 24$; and for $df_2 = 221$ select $df_2 = 100$; critical F value is 1.46.

Figure 7

Critical Values for F Distribution for Type 1 Error Rate of α = .10

Critical values of the F distribution for $\alpha = 0.10$; $P\{F > F_{.10}(k_1,k_2)\} = 0.10$

Denomin df (k_2)	Numerator Degrees of Freedom (k_1) 1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	100
1	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7	61.2	61.7	62.0	62.3	62.5	62.8	63.0
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.13
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.75
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.50
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.19
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.09
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.01
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.94
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.07	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.76
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.73
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.95	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.70
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.93	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.65
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.90	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.63
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.61
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.87	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.58
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.84	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.83	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.55
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.82	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.54
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.53
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.80	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.52
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.51
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.43
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.70	1.68	1.63	1.57	1.54	1.50	1.46	1.42	1.39
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.36
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68	1.65	1.63	1.57	1.51	1.48	1.44	1.40	1.36	1.32
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.64	1.61	1.56	1.49	1.46	1.42	1.38	1.34	1.29

Note that the two different degrees of freedom are often reported as df = x,y where x and y are numbers, for example $df_1 = 20$ and $df_2 = 17$ could be reported as df = 20,17.

The decision rule for the F test is

$$If\ F \geq F_{crit}\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

So if the F-ratio calculated is larger than the critical F-value, reject H_0 and conclude that group variances appear to be unequal (heterogeneous).

Example 1: Blood Pressure

Do males and females in EDUR 8131 have systolic blood pressure with similar variances?

$$H_0: \sigma^2_{females} = \sigma^2_{males}$$

Table 3

Summary Statistics for Systolic Blood Pressure by Sex

Male		Female	
M =	130.143	M =	102.857
s =	20.514	s =	22.267
n =	7	n =	7

The variance for males is $20.514^2 = 420.824$ and the variance for females is $22.267^2 = 495.82$.

The female variance is larger so it will be the numerator in the F-max calculation and also $df_1 = 7 - 1 = 6$. The denominator degrees of freedom is $df_2 = 7 - 1 = 6$.

$$F = \frac{s_{largest}^2}{s_{smallest}^2} = \frac{495.82}{420.82} = 1.18$$

The critical F value, for α=.10, with df = 6,6 is 3.05.

$$If\ \mathbf{1.18 \geq 3.05}\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since 1.18 is less than 3.05 the decision is fail to reject $H_0: \sigma_{females}^2 = \sigma_{males}^2$, therefore we conclude that the two groups come from populations with similar variances, given the data collected.

If the variances are different then usually one would proceed with the t-test formula that corrects for heterogeneous variances.

Example 2

Do the two groups of students involved in feedback participation have similar intrinsic motivation variances?

$$H_0: \sigma_{feedback}^2 = \sigma_{no\ feedback}^2$$
$$H_1: \sigma_{feedback}^2 \neq \sigma_{no\ feedback}^2$$

Table 4
Intrinsic Motivation by Feedback Opportunity

Feedback Opportunity: Yes		Feedback Opportunity: No	
M =	4.243	M =	3.100
s =	0.270	s =	1.413
n =	7	n =	12

The group that provided feedback had an intrinsic motivation variance of 0.073 (n = 7) and the group who did not participate in feedback displayed an intrinsic motivation variance of 1.996 (n = 12).

The no feedback group's variance is larger so it will be the numerator in the F-max calculation and that group's $df_1 = 12 - 1 = 11$. The feedback group's variance will be the denominator in the F test and their corresponding degrees of freedom is $df_2 = 7 - 1 = 6$.

$$F = \frac{s_{largest}^2}{s_{smallest}^2} = \frac{1.996}{0.073} = 27.342$$

The critical F value, for α=.10, with df = 11,6 is 2.92.

$$If\ \mathbf{27.342 \geq 2.92}\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since 11.285 is larger than 2.92 one must reject H_0 and conclude the intrinsic motivation variances for these two groups differ therefore the unequal variance formulas, the un-pooled variances, for the two-group t test must be used.

10. t-test with Unequal Variances: t-ratio, SE_d, and df

When variances are unequal, one uses the standard formula for the t-ratio:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{SE_d}$$

The difference from the pooled variance t-test is how the standard error of the mean difference is calculated; when unequal variances exist one should use separate variances to determine the standard error. The standard error of the difference, using separate variances rather than the pooled variance, is found by

$$SE_d = s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The t ratio that results from employing the above standard error of the difference does not follow a t-distribution so p-values will be incorrect. One approach to addressing this deficiency is to use a correction to the degrees of freedom. The following formula represents the Satterthwaite approximated, or corrected, degrees of freedom, df_c, reported by most software:

$$df_c = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

This formula usually produces degrees of freedom that have decimal fractions (i.e., not a whole number). Critical t tables only provide critical values for whole numbered degrees of freedom. If one is using statistical software to calculate the standard error and df_c then the decimal fraction is not an issue since the software should handle this information correctly. However, if one wishes to find critical t values from a t table using df_c, then the decimal fraction is problematic because tables for critical t values assume df are whole numbers.

Recommendations for using df_c with critical t tables:

- replace df_c with $n_1 - 1 \text{ or } n_2 - 1$, whichever is smaller, and use this value as df for finding critical t values
- round down the df_c to the nearest whole number – this method is recommended – then find critical t values

11. Example t-tests for Unequal Variances in Excel and SPSS

(a) Example 1: Blood Pressure

Is there a difference in mean systolic blood pressure between males and females in EDUR 8131? Raw data, descriptive statistics, and t-test results for both groups are provided below.

The t-test with unequal variances is not needed for these data, nevertheless these data will be used as an example of how to calculate the t-test for unequal variances.

Table 1 (reposted)
Systolic Blood Pressure by Sex

Male		Female	
95	151	76	90
145	110	120	101
129	143	115	81
138		137	
M =	130.143	M =	102.857
s =	20.514	s =	22.267
n =	7	n =	7

If calculating by hand, follow these steps:

(a) First, perform F-max or Levene's test to assess homogeneity of variances. If $H_0\colon \sigma_1^2 = \sigma_2^2$ is rejected then proceed with the steps below for the unequal variances t-test; if If $H_0\colon \sigma_1^2 = \sigma_2^2$ is not rejected, the use the pooled-samples (equal variances) t-test steps and formulas discussed above in section 7.

- SPSS: Levene's F = 0.12; p = 0.73 ; since p > α= .10 fail to reject, variances are similar
- Excel: F-max F = 0.18; critical F = 3.05 (α= .10); since calculated F is smaller than critical, FTR Ho
- F-max test and Levene's test for these data show variances are similar so normally one would proceed using the equal variances assumed t-test.
- However, we will pretend that these groups have unequal variances to illustrate steps in the t-test with unequal variances.

(b) Determine <u>mean</u> and <u>sample size</u> for both groups.

(c) Find s^2, <u>variances</u>, for both groups (use of standard deviation rather than variance is a common error):

$$s_{male}^2 = 20.514^2 = 420.824$$
$$s_{female}^2 = 22.267^2 = 495.819$$

(d) Find SE_d or $s_{\bar{x}_1-\bar{x}_2}$, <u>standard error of the mean difference</u>:

$$SE_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{420.824}{7} + \frac{495.819}{7}} = \sqrt{60.1177 + 70.8313} =$$

$$= \sqrt{130.949} = 11.443$$

(e) Find the <u>t-ratio</u>, t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE_d} = \frac{130.143 - 102.857}{11.443} = \frac{27.286}{11.443} = 2.385$$

(f) Find <u>df</u> and <u>critical values</u> for whichever significance level (α) is specified if using t-values for hypothesis testing:

$$df_c = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{420.824}{7} + \frac{495.819}{7}\right)^2}{\frac{1}{7-1}\left(\frac{420.824}{7}\right)^2 + \frac{1}{7-1}\left(\frac{495.819}{7}\right)^2}$$

$$= \frac{(130.949)^2}{0.1667 * (60.1177)^2 + 0.1667 * (70.8313)^2} = \frac{17147.6406}{602.3566 + 836.1785}$$

$$= \frac{17147.6406}{1438.5351} = 11.92 \downarrow 11.00$$

α = .05; critical t = ± 2.20
α = .01; critical t = ± 3.11

(g) Apply decision rule

$$If\ |t| \geq |t_{crit}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

$$If\ |\mathbf{2.385}| \geq |\mathbf{2.20}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since the calculated t-value of 2.385 is larger than the critical t-value of 2.20 reject Ho and conclude that males have higher blood pressure than females since males' blood pressure mean is higher.

Results from SPSS are presented below.

Figure 8
SPSS Output for Two-group t-test of Blood Pressure by Sex

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
blood_pressure	Female	7	102.8571	22.26678	8.41605
	Male	7	130.1429	20.51364	7.75343

Use this row for unequal variance t-test results.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
blood_pressure	Equal variances assumed	.122	.733	-2.384	12	.034	-27.28571	11.44315	-52.21819	-2.35324
	Equal variances not assumed			-2.384	11.920	.035	-27.28571	11.44315	-52.23671	-2.33472

Questions

1. Why is the t-value obtained by hand calculation positive (2.385) and the one produced by SPSS negative (-2.384)?

In SPSS's t-test command one has the option of identifying which group should be listed first in the t-test formula; if the group with the higher mean is listed first then the mean difference and t-ratio will be positive; if the group with the lower mean is listed first then the mean difference and t-ratio will be negative.

2. Recall that for blood pressure data we found that group variances were not different which means the t-test with equal variances can be used. This also means the results for both equal and unequal t-tests should be similar. Do you notice any differences in t-test results between the equal and unequal t-test results reported by SPSS in Figure 8?

(b) Example 2: Intrinsic Motivation

Does intrinsic motivation differ between students who are given an opportunity to provide instructional feedback to their instructor and students who not given an opportunity to provide instructional feedback?

Raw data and descriptive statistics for intrinsic motivation are presented below in Table 2.

Table 2 (reposted)
Intrinsic Motivation by Feedback Opportunity

Feedback Opportunity: Yes		Feedback Opportunity: No	
4.2	4.6	1.1	1.3
4.5		2.3	3.1
4.3		3.8	4.7
3.9		4.8	2.9
4.3		4.2	2.9
3.9		4.9	1.2
M =	4.243	M =	3.100
s =	0.270	s =	1.413
n =	7	n =	12

If calculating by hand, follow these steps:

(a) First, perform F-max or Levene's test to assess homogeneity of variances. If $H_0: \sigma_1^2 = \sigma_2^2$ is rejected then proceed with the steps below for the unequal variances t-test; if If $H_0: \sigma_1^2 = \sigma_2^2$ is not rejected, the use the pooled-samples (equal variances) t-test steps and formulas discussed above in section 7.

- SPSS: Levene's F = 10.71; p = 0.004 ; since p is less than α = .10 reject Ho
- Excel: F-max F = 27.39; critical F = 2.92; reject $H_0: \sigma_1^2 = \sigma_2^2$ since calculated F is larger than critical F

(b) Determine mean and sample size for both groups.

(c) Find s², variances, for both groups (use of standard deviation rather than variance is a common error):

$$s_{yes}^2 = 0.27^2 = 0.0729$$
$$s_{no}^2 = 1.413^2 = 1.9966$$

(d) Find SE_d or $s_{\bar{x}_1-\bar{x}_2}$, standard error of the mean difference:

$$SE_d = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{0.0729}{7}+\frac{1.9966}{12}} = \sqrt{0.0104+0.1664}$$

$$= \sqrt{0.1768} = 0.4205$$

(e) Find the t-ratio, t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE_d} = \frac{4.243 - 3.100}{0.4205} = \frac{1.143}{0.4205} = 2.7182$$

(f) Find df and critical values for whichever significance level (α) is specified if using t-values for hypothesis testing:

$$df_c = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{0.0729}{7} + \frac{1.9966}{12}\right)^2}{\frac{1}{7-1}\left(\frac{0.0729}{7}\right)^2 + \frac{1}{12-1}\left(\frac{1.9966}{12}\right)^2}$$

$$= \frac{(0.17679)^2}{0.16667 * (0.01041)^2 + 0.0909 * (0.16638)^2} = \frac{0.03125}{0.000018 + 0.002516}$$

$$= \frac{0.03125}{0.002534} = 12.33 \downarrow 12.00$$

α = .05; critical t = ± 2.18
α = .01; critical t = ± 3.05

(g) Apply decision rule

$$If\ |t| \geq |t_{crit}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

$$If\ |\mathbf{2.718}| \geq |\mathbf{2.18}|\ reject\ H_0, otherwise\ fail\ to\ reject\ H_0$$

Since the calculated t-value of 2.718 is larger than the critical t-value of 2.18 reject Ho and conclude that mean intrinsic motivation is higher for those students allowed to provide instructional feedback than for those not given opportunity to offer instructional feedback.

Results from SPSS are presented in Figure 9.

Figure 9
SPSS Output for Two-group t-test of Intrinsic Motivation by Feedback Opportunity

Group Statistics

	feedback	N	Mean	Std. Deviation	Std. Error Mean
motivation	1.00	7	4.2429	.26992	.10202
	.00	12	3.1000	1.41293	.40788

Use this row for unequal variances t-test.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
motivation	Equal variances assumed	10.712	.004	2.094	17	.052	1.14286	.54589	-.00888	2.29459
	Equal variances not assumed			2.718	12.331	.018	1.14286	.42044	.22951	2.05620

Notice the difference in inferential statistics (t-value, df, p-value, standard error of mean difference, and confidence intervals) between equal and unequal variance t-test results.

12. Confidence Intervals about the Mean Difference

Recall the CI formula for a sample mean

$$(1-\alpha)CI = \bar{X} \pm ({}_{1-(\frac{\alpha}{2})}t_{df})(s_{\bar{x}})$$

which can be simplified to

$$(1-\alpha)CI = \bar{X} \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

The CI for the mean difference requires only one change, replace $\bar{X}$ with $\bar{X}_1 - \bar{X}_2$. The CI for the mean difference is

$$(1-\alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (critical\ t\ value)(standard\ error\ of\ mean\ difference)$$

(a) Example 1: Blood Pressure

Find a 95% and 99% CIs about the mean difference in systolic blood pressure between females and males.

Below are summary statistics for blood pressure:

Mean Blood Pressure for Males = 130.143
Mean Blood Pressure for Females = 102.857
Mean difference = 130.143 – 102.857 = **27.286**
SE_d = **11.443** (assuming equal variances)
df = 7 + 7 – 2 = 12
95% t critical = ± **2.18**
99% t critical = ± **3.05**

One needs only the mean difference, SE_d, and critical t-values to calculate confidence intervals for mean differences.

The 95% CI formula:

$$(1-\alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (critical\ t\ value)(standard\ error\ of\ mean\ difference)$$

$$.95\ CI = 27.286 \pm (2.18)(11.443)$$

Upper Limit = $27.286 + (2.18)(11.443) = 27.286 + 24.94574 = 52.23$
Lower Limit = $27.286 - (2.18)(11.443) = 27.286 - 24.94574 = 2.34$

Interpretation:

One may be 95% confident that the interval 2.34 to 52.23 includes the population mean difference of systolic blood pressure between females and males in EDUR 8131.

Since 0.00 does not appear within this interval, one may reject Ho since 0.00 is not one of the likely values for the mean difference in blood pressure between males and females.

The 99% CI formula (note the only change from above is the critical t-value, 2.18 to 3.05):

$$(1-\alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (critical\ t\ value)(standard\ error\ of\ mean\ differene)$$

$$.99\ CI = 27.286 \pm (\mathbf{3.05})(11.443)$$

Upper Limit = $27.286 + (\mathbf{3.05})(11.443) = 27.286 + 34.901 = 62.187$
Lower Limit = $27.286 - (\mathbf{3.05})(11.443) = 27.286 - 34.901 = -7.615$

Interpretation:

One may be 99% confident that the interval -7.615 to 62.187 includes the population mean difference of systolic blood pressure between females and males in EDUR 8131.

Since 0.00 does appear within this interval, one would fail to reject Ho since 0.00 is one of the likely values for the mean difference in blood pressure between males and females. Since 0.00 is one of the likely values for the mean difference, this suggests that no difference is possible.

Figure 1, reposted here, shows the 95% confidence interval created by SPSS.

Question

Why does SPSS provide a 95% confidence interval with negative values while the 95% confidence interval calculated above has positive values?

Whether negative or positive depends on which group is specified to be first when calculating the mean difference. See reply to Question 1 on page 19 of these notes.

Illustrate how to obtain 99% CI with SPSS, and how to change comparison groups to obtain a 95% CI with positive values to match the 95% CI calculated above.

Figure 1 (Reposted)
SPSS Output for Two-group t-test of Blood Pressure by Sex

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
blood_pressure	Female	7	102.8571	22.26678	8.41605
	Male	7	130.1429	20.51364	7.75343

Use this row for equal variance t-test results.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
blood_pressure	Equal variances assumed	.122	.733	-2.384	12	.034	-27.28571	11.44315	-52.21819	-2.35324
	Equal variances not assumed			-2.384	11.920	.035	-27.28571	11.44315	-52.23671	-2.33472

(b) Example 2: Intrinsic Motivation

Find a 95% and 99% CIs about the mean difference in systolic blood pressure between females and males.

Below are summary statistics for intrinsic motivation:

> Mean Intrinsic Motivation for the "Yes" Feedback Group = 4.243
> Mean Intrinsic Motivation for the "No" Feedback Group = 3.100
> Mean difference = 4.243 – 3.1 = **<u>1.143</u>**
> SE_d = **<u>0.4205</u>** (assuming unequal variances)
> df = $12.33 \downarrow 12.00$
> 95% t critical = ± **<u>2.18</u>**
> 99% t critical = ± **<u>3.05</u>**

One needs only the <u>mean difference</u>, <u>SE_d</u>, and <u>critical t-values</u> to calculate confidence intervals for mean differences.

The 95% CI formula:

$$(1-\alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (critical\ t\ value)(standard\ error\ of\ mean\ difference)$$

$$.95\ CI = 1.143 \pm (2.18)(0.4205)$$

Upper Limit = $1.143 + (2.18)(0.4205) = 1.143 + 0.91669 = 2.0597$
Lower Limit = $1.143 - (2.18)(0.4205) = 1.143 - 0.91669 = 0.2263$

Interpretation:

> One may be 95% confident that the interval 0.2263 to 2.0597 includes the population mean difference in intrinsic motivation between those given the opportunity to provide feedback and those not given opportunity to provide feedback.

Since 0.00 does not appear within this interval, one may <u>reject</u> Ho since 0.00 is not one of the likely values for the mean difference in intrinsic motivation between the two groups studied.

The 99% CI formula (note the only change from above is the critical t-value, 2.18 to 3.05):

$$(1-\alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (critical\ t\ value)(standard\ error\ of\ mean\ differene)$$

$$.95\ CI = 1.143 \pm (\mathbf{3.05})(0.4205)$$

Upper Limit = $1.143 + (\mathbf{3.05})(0.4205) = 1.143 + 1.2825 = 2.426$
Lower Limit = $1.143 - (\mathbf{3.05})(0.4205) = 1.143 - 1.2825 = -0.1336$

Interpretation:

> One may be 99% confident that the interval -0.1336 to 2.426 includes the population mean difference in intrinsic motivation between those given the opportunity to provide feedback and those not given opportunity to provide feedback.

Since 0.00 does appear within this interval, one would fail to reject Ho since 0.00 is one of the likely values for the mean difference in intrinsic motivation between the two groups studied. If 0.00 is one of the likely values for the mean difference, this suggests that no difference is possible.

Figure 9, reposted here, shows the 95% confidence interval created by SPSS.

Figure 9 (Reposted)
SPSS Output for Two-group t-test of Intrinsic Motivation by Feedback Opportunity

Group Statistics

	feedback	N	Mean	Std. Deviation	Std. Error Mean
motivation	1.00	7	4.2429	.26992	.10202
	.00	12	3.1000	1.41293	.40788

Use this row for unequal variances t-test.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
motivation	Equal variances assumed	10.712	.004	2.094	17	.052	1.14286	.54589	-.00888	2.29459
	Equal variances not assumed			2.718	12.331	.018	1.14286	.42044	.22951	2.05620

13. Pooled t-test vs. Unequal Variance t-test: Comparison of Possible Results

Given the presentation of both the pooled variances t-test and the unequal variances t-test, it is possible to arrive at different inferential conclusions if the conditions (i.e., sample sizes, variances) are sufficiently discrepant. The following examples should help clarify the effect of sample size and variance differences on statistical estimates for these two t-test approaches.

For each scenario presented below observe differences in estimates of the following statistics for both pooled and unequal variance t-test calculations:

- SE_d
- calculated t-ratio
- df
- p-value
- critical t-value
- 95% confidence interval

Work each example with the Excel file for calculating t-test results; use the “Summary Statistics” sheet.

(a) Larger variance with larger sample size:

Group A		Group B	
M =	65	M =	50
s =	5	s =	25
n =	5	n =	25

(b) Larger variance with smaller sample size:

Group A		Group B	
M =	65	M =	50
s =	25	s =	5
n =	5	n =	25

(c) Different variances but equal sample sizes:

Group A		Group B	
M =	65	M =	50
s =	25	s =	5
n =	25	n =	25

Note that when sample sizes are the same or nearly the same, there is usually little difference between t-test results based upon equal and unequal variances even when variances differ greatly (e.g., example c above).

14. APA Style Results Presentation

It is possible to report two-independent samples t-test results in either textual or tabular format, but the tabular format is preferred. Both options are provided below.

(a) Table Format

Three examples are provided:

- Table 5 shows a t-test with equal variances;
- Table 6 presents a t-test with unequal variances using the Satterthwaite approximation; and
- Table 7 provides results from three t-tests reported simultaneously in one table.

As with the one-sample t-test, written presentation should include first <u>inferential</u> information (whether Ho was rejected) and this should be followed with <u>interpretational</u> information (what the results mean in simple language).

Table 5 demonstrates APA style for the t-test with pooled variance (i.e., equal variances t-test).

Table 5
Results of t-test and Descriptive Statistics for Systolic Blood Pressure by Sex

	Sex						95% CI for Mean Difference		
	Female			Male					
	M	SD	n	M	SD	n		t	df
Sys. Blood Pressure	102.86	22.27	7	130.14	20.51	7	-52.22, -2.35	-2.38*	12

* p < .05.

There is a statistically significant mean difference in systolic blood pressure between males and females. Results show that males tend ot have a higher systolic blood pressure than do females.

Table 6 demonstrates presentation of t-test results with unequal variances using the Satterthwaite approximation for degrees of freedom. This is the most common adjustment used in statistical software and the adjustment found in SPSS.

Table 6

Results of t-test and Descriptive Statistics for Intrinsic Motivation by Instructional Feedback Opportunity

	Instructional Feedback Opportunity						95% CI for Mean Difference		
	Yes			No					
	M	SD	n	M	SD	n		t	df
Intrinsic Motivation	4.24	0.27	7	3.10	1.41	12	0.23, 2.06	2.72*	12.33

Note: Satterthwaite approximation employed due to unequal group variances.
* p < .05.

Results of the t-test show a statistically significant mean difference in intrinsic motivation between students given an opportunity to provide instructional feedback and students not given such an opportunity. Students allowed to provide instructional feedback demonstrated greater average levels of intrinsic motivation when compared to their peers not allowed to provide instructional feedback.

Table 7 shows how to combine results from several t-tests performed for the same groups.

Table 7

Results of t-tests and Descriptive Statistics SAT Verbal, SAT Math, and GPA by Sex

Outcome	Sex						95% CI for Mean Difference		
	Male			Female					
	M	SD	n	M	SD	n		t	df
SAT-Verbal	463.81	98.89	45	532.21	101.23	44	-110.56, -26.24	-3.22*	87
SAT-Math	515.43	99.56	44	483.31	98.97	44	-9.95, 74.20	1.52	86
College GPA	2.71	1.32	45	3.16	1.16	44	-0.97, 0.07	-1.71	87

* p < .05.

There are statistically significant differences, at the .05 level of significance, between male and female college students in SAT verbal scores, but not with SAT mathematics or college GPA. Results show that females had higher verbal scores, but no statistical difference exists between males and females in terms of GPA or SAT mathematics scores.

There is a mistake in Table 7 – can you identify that mistake? Hint:

- Logic – what cannot occur?
- Use the Excel spreadsheet for t-tests by summary statistics to identify which is incorrect.

Watch the instructional video for this section to learn about the problem with Table 7.

(b) Text Format

Some may prefer to report t-test results directly within text especially if printed space is limited (such as due to per-page publication costs). Below is an example illustration this presentation approach.

Results of the two-independent samples t-test shows that mean systolic blood pressure differs between males (M = 130.14, SD = 20.51, n = 7) and females (M = 102.86, SD = 22.27, n = 7) at the .05 level of significance (t = 2.38, df = 12, p < .05, 95% CI for mean difference 2.35 to 52.22). On average males tend to have higher systolic blood pressure than females.

15. Topics not Covered

(a) Directional tests

Directional (group 1 has higher mean than group 2)

The experimental group will show a higher level of achievement.

H_0: $\mu_1 \le \mu_2$
H_1: $\mu_1 > \mu_2$

or

H_0: $\mu_1 - \mu_2 \le 0.00$
H_1: $\mu_1 - \mu_2 > 0.00$

Directional (group 2 has higher mean than group 1)

The experimental group will show a lower level of achievement.

H_0: $\mu_1 \ge \mu_2$
H_1: $\mu_1 < \mu_2$

or

H_0: $\mu_1 - \mu_2 \ge 0.00$
H_1: $\mu_1 - \mu_2 < 0.00$

These are not covered because I think non-directional is better since it less likely one will "cheat" to obtain statistical significance (i.e., adopts in a post-hoc fashion directional tests since they offer more power compared with non-directional tests). Also statistical software tends to provide p-values only for non-directional tests.

(b) Effect Size d

The effect size d represents the mean difference divided by the common or pooled standard deviation:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

Like a Z score, this value indicates the size of the mean difference in standard deviation units. The d formula may appear to be the same as the t formula, but note the different denominator. In the d formula the denominator is s, the sample standard deviation, and in the t formula the denominator is $s_{\bar{x}}$, standard error of the mean.

(c) Sample Size

One may calculate required sample size for a two-sample t test to meet specifications for Type 1 error rate (α), power (1-β), and effect size (d). This requires use of tabled values or software.

16. Exercises

See the course web page and the course text for exercises for two-sample t test.