

Exploring Diversity with Statistics

EXPLORING DIVERSITY WITH STATISTICS

Step-by-step JASP Guides

RUTH WALKER; ASHLYN A. MORAINÉ; DR. HANNAH J. OSBORN;
ASIA SYMONE PALMER; ERIN N. PRINCE; KELSEY L. HUMPHREY;
KENDRA E. SCOTT; AND KRISTEN J. BLACK

MARCELLA REA AND RACHEL FLEMING

University of Tennessee at Chattanooga
Chattanooga (Tenn.)



Exploring Diversity with Statistics Copyright © 2021 by Ruth Walker is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This book was produced with Pressbooks (<https://pressbooks.com>) and rendered with Prince.

CONTENTS

Part I. Main Body

1. Running and Interpreting an Independent Samples t test in JASP	1
Kelsey L. Humphrey; Ruth Walker; and Erin N. Prince	
2. Running and Interpreting a Paired Samples t test in JASP	23
Kendra E. Scott; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker	
3. Running and Interpreting a One-Way ANOVA in JASP	40
Ashlyn A. Moraine; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker	
4. Running and Interpreting a Two-Way ANOVA in JASP	64
Ashlyn A. Moraine; Erin N. Prince; and Ruth Walker	
5. Running and Interpreting a Chi-Square Test in JASP	86
Ashlyn A. Moraine; Dr. Hannah J. Osborn; and Ruth Walker	
6. Running and Interpreting a Correlation in JASP	101
Asia Symone Palmer; Ruth Walker; and Kristen J. Black	
7. Running and Interpreting Linear Regression in JASP	114
Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black	
8. Running and Interpreting Multiple Regression in JASP	132
Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black	
Instructor Development Resources for Integrating and Teaching EDI Topics	149
Ruth Walker	
Downloadable Materials	153

1.

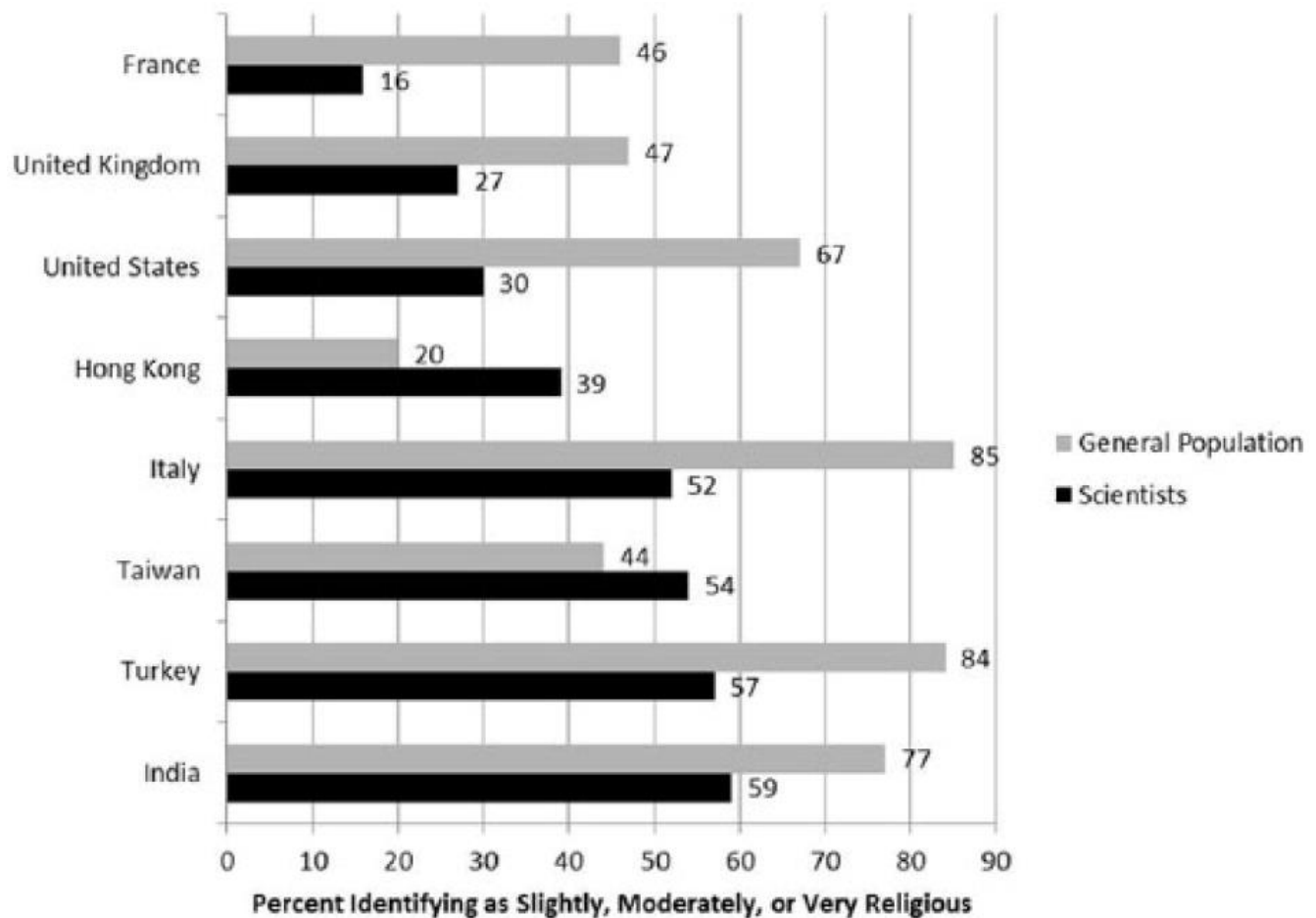
RUNNING AND INTERPRETING AN INDEPENDENT SAMPLES T TEST IN JASP

Kelsey L. Humphrey; Ruth Walker; and Erin N. Prince

Download Data Sets:

- JASP Data File Independent Samples t test
- CSV Data File Independent Samples t test

Researchers Elizabeth Barnes, Jasmine Truong, and colleagues (2020) conducted a series of studies at Arizona State University to understand if there is bias against Christian students in the natural sciences. Pew Research Center (2019) surveys have found that approximately 65% of the population in the United States describe themselves as Christian; however, a worldwide sample of over 9,000 scientists in biology and physics found lower levels of religiosity in scientists compared to the general population in the United States (see graph below; Howard Ecklund et al., 2016). Further, 22% of scientists in the United States said that science has made them “much less religious” (p. 5).



This led Barnes and colleagues (2020) to their research questions: Why are Christians underrepresented in the natural sciences? Do negative stereotypes about the scientific abilities of Christian students put them at a disadvantage in the natural sciences? Their first study found that it was common for biology students to report perceptions of bias against Christians in science (see table below).

Item	% who agreed	% who neither agreed nor disagreed	% who disagreed
Discrimination against Christians is not a problem in science.	19.6%	28.5%	52.0%
It is rare to see Christians discriminated against in the sciences.	24.2%	40.7%	35.1%
On average, people in science treat Christians and non-religious people equally.	41.1%	26.4%	32.5%
Society has reached a point where Christians and non-religious people have equal opportunities for achievement in science.	54.7%	26.2%	19.1%

<https://doi.org/10.1371/journal.pone.0226826.t001>

Once they gathered data to show that students *perceive* there is a bias against Christians in science, they wanted to conduct studies to help determine if academic scientists in the natural sciences exhibited biased behavior towards Christian students. To do this, they conducted two experimental studies. We will analyze Study 2 data when we discuss One-Way ANOVAs, for now we will look closer at Study 3. For Study 3, the researchers recruited 261 faculty in biology and randomly assigned them to one of two conditions:

1. Faculty in condition one read a graduate student application that signaled evangelism. Their application listed the student taking a mission trip for Campus Crusade for Christ, with a recommendation letter from a mentor in that ministry.
2. Faculty in condition two read a graduate student application that listed the student taking a service trip for the United Nations Children's Fund, with a recommendation letter for a mentor in that organization.

All other aspects of their application were kept consistent (e.g., same GPA and GRE scores). Faculty were asked to rate the student on competence, hireability, and likeability using a 7-point likert scale from 1 (*not at all*) to 7 (*very much*). Higher scores on each of these scales indicate higher levels of competence, hireability, and likeability.

Hypotheses

We are going to focus on two of the three outcome variables Barnes and colleagues (2020) measured in their study. Specifically, we will focus on faculty ratings of (1) student competence and (2) student likeability. This means that we will have three different sets of hypotheses.

The null hypotheses are:

- Student Competence
 - Conceptual H_0 : There are no differences in faculty ratings of graduate student competence between the Christian student condition and the Control student condition.
 - Mathematical H_0 : The mean score for graduate student competence in the Christian student condition is equal to the mean score for student competence in the Control student condition; $M_1 = M_2$.
- Student Likeability
 - Conceptual H_0 : There are no differences in faculty ratings of graduate student likeability between the Christian student condition and the Control student condition.
 - Mathematical H_0 : The mean score for graduate student competence in the Christian student condition is equal to the mean score for student likeability in the Control student condition; $M_1 = M_2$.

The alternative hypotheses are:

- Student Competence
 - Conceptual H_1 : There are significant differences in faculty ratings of graduate student competence between the Christian student condition and the Control student condition.
 - Mathematical H_1 : The mean score for graduate student competence in the Christian student condition is not equal to the mean score for student competence in the Control student condition;

$$M_1 \neq M_2.$$

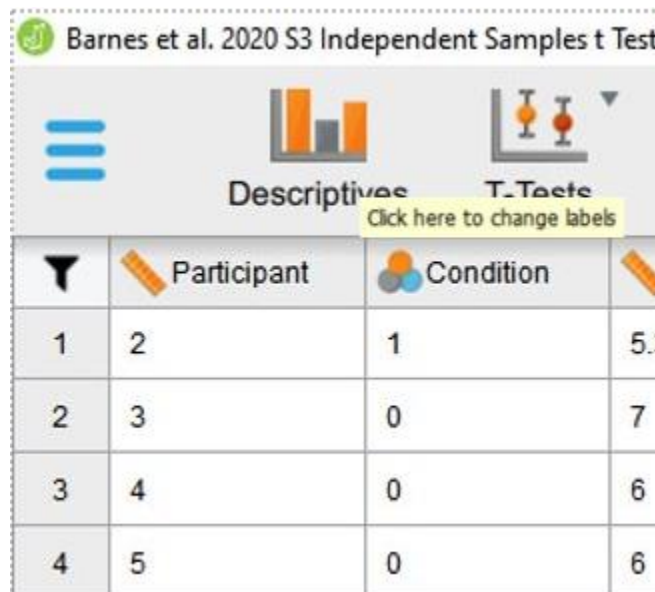
- Student Likeability
 - Conceptual H₂: There are significant differences in faculty ratings of graduate student likeability between the Christian student condition and the Control student condition.
 - Mathematical H₂: The mean score for graduate student competence in the Christian student condition is not equal to the mean score for student likeability in the Control student condition; $M_1 \neq M_2$.

JASP Analyses

In order to run analyses in JASP, the first thing we need to do is open the data set we will be working with. To do this, open JASP and follow the steps below.

File → Open → Computer → Browse → Pick the Independent t Practice Data (Barnes et al., 2020 Independent Samples *t* Test Data)

Once the data set is open in JASP, we will change the data labels for our condition variable to make interpretation easier. Currently, the Condition column has either a 0 or a 1 for each participant. To change the numerical data into our categorical labels (i.e., Christian Student or the Control Condition), you will take your cursor and hover over Condition. When you see a note pop up saying, “click here to change labels” click on it.



	Participant	Condition	
1	2	1	5.3
2	3	0	7
3	4	0	6
4	5	0	6

Condition		
Filter	Value	Label
✓	0	Christian Student
✓	1	Control Condition

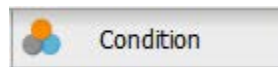
Barne's et al. (2020) codebook indicate that the appropriate label for 0 is "Christian student" and 1 is "Control condition." To make this change in our dataset, we will click on the 0 under the Label column and type 'Christian Student'. Then we will click on the 1 under the Label column and type 'Control Condition'. After you have changed the labels, you can close the window by clicking on the "x" button.

Assumption Testing

Prior to running our independent samples t test, we first have to check to make sure we meet the assumptions for this statistical test.

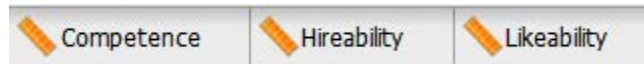
Assumption One: Is our independent variable dichotomous and measured at the categorical (i.e., nominal) level?

Yes. Our independent variable for this study is the level of religiosity in the graduate student applications faculty read. Participants either read the scenario featuring a "Christian Student" or a student with no religious affiliation given. This means our independent variable for this particular study is dichotomous (meaning, there are only two categories), and categorical. If you look at the data label icon in JASP, you can see it has the appropriate icon (three overlapping circles) for a nominal/categorical variable. We meet this assumption.



Assumption Two: Is the dependent variable continuous (i.e., ratio or interval)?

Yes. Our dependent variables for this study are faculty ratings of the graduate student applicants' likeability and competence. Each variable was assessed using a four-item scale. Faculty rated each item on a 7-point likert scale from 1 (*not at all*) to 7 (*very much*). This variable is continuous, meaning it was measured at the interval or ratio level. If you look at the data label icon in JASP, you can see it has the appropriate icon (a ruler) for a scale or continuous variable. We meet this assumption.



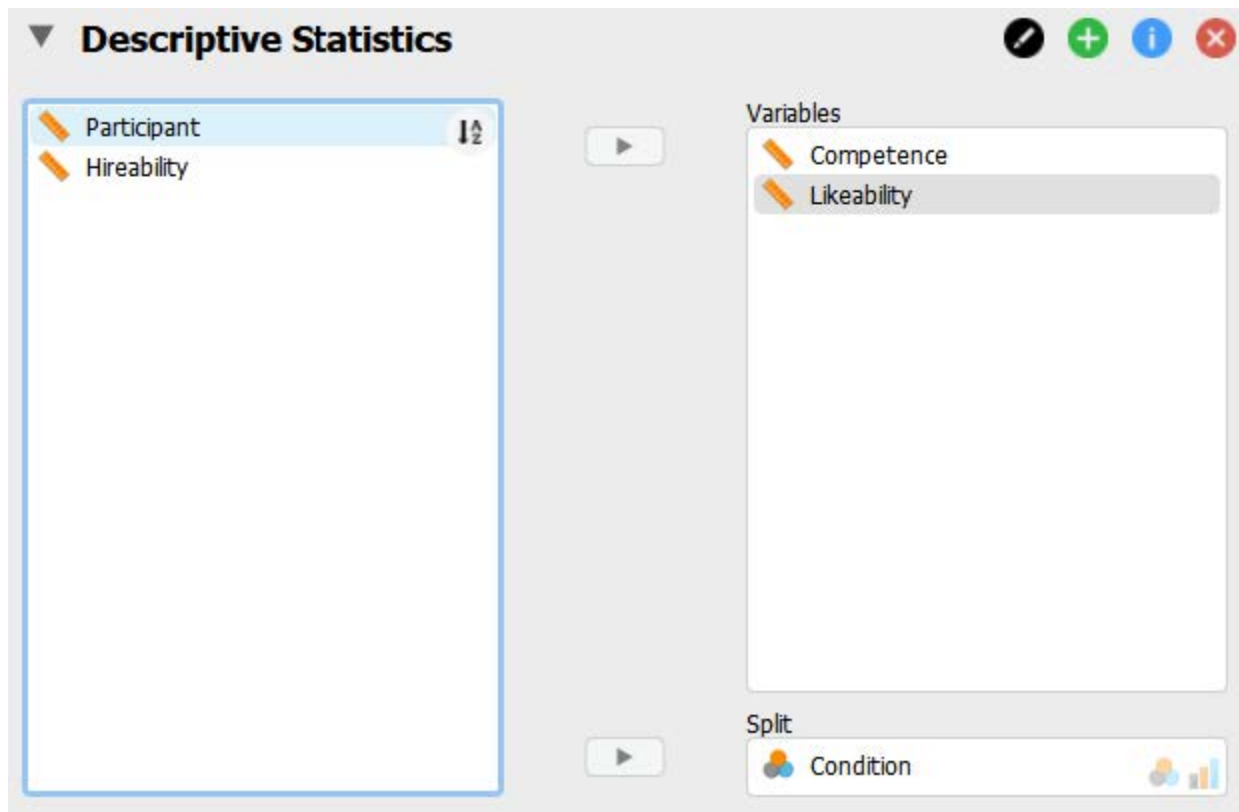
Assumption Three: Are the samples independent?

Yes. Looking at our data set, we can determine that our two groups (Christian Student condition and Control Student condition) are independent. If we look in the "Condition" column, we can see that all biology faculty participants were randomly sorted into either the Christian Student or Control Condition groups – there are no participants who have both labels or any other group value listed. Thus, we can conclude that participants in both groups are independent of one another. No participant is in more than one group. We meet this assumption.

Assumption Four: Is the dependent variable normally distributed for each group of the IV?

To check our data for the next two assumptions, we will use the Descriptive analysis tab. Click Descriptive. When the "Descriptive Statistics" window pops up, we will need to move our dependent variable to the

“Variables” box and our independent variable to the “Split” box using the arrows depicted in the photo below – this is so we can examine normality in the DV for both of our groups separately.



Once we have our variables in the appropriate boxes, we are going to ask JASP to run the various statistics and graphics we will need to interpret for our assumptions by clicking on the appropriate boxes in the test window. We will be asking JASP to provide us with all possible output we may want to look at for our assumptions; however, we will focus on interpreting the output that you will be expected to analyze for your statistical lab assignment.

1. Under the “Customizable Plots” tab we will check the “Boxplots,” “Boxplot element,” “Jitter element,” and “Label outliers” boxes. We will also check “Distribution plots” and “Q-Q plots” under the “Basic plots” tab.
2. Under the “Statistics” tab we will check the “Skewness,” “Kurtosis,” and “Shapiro-Wilk test” boxes.

See the image below to make sure your test window has all the appropriate boxes selected.

Customizable plots

Color palette Colorblind ▼

☒ Boxplots

- ☒ Boxplot element ☐ Use color palette
- ☐ Violin element ☒ Label outliers
- ☒ Jitter element

Basic plots

- ☒ Distribution plots ☐ Correlation plots ☐ Interval plots
- ☐ Display density ☒ Q-Q plots
- ☐ Display rug marks ☐ Pie charts
- Bin width type Sturges ▼ ☐ Dot plots
- Number of bins 30

Categorical plots

- ☐ Pareto plots
- ☐ Pareto rule 95.0 %
- ☐ Likert plots
- ☐ Assume all variables share the same levels
- Adjustable font size for vertical axis Normal ▼

Statistics

Sample size

☒ Valid
 ☒ Missing

Quantiles

☐ Quartiles
 ☐ Cut points for: equal groups
 ☐ Percentiles:

Central tendency

☐ Mode
 ☐ Median
 ☒ Mean

Distribution

☒ Skewness
 ☒ Kurtosis
 ☒ Shapiro-Wilk test
 ☐ Sum

Dispersion

☐ S.E. mean
 ☒ Std. deviation
 ☐ Coefficient of variation
 ☐ MAD
 ☐ MAD robust
 ☐ IQR
 ☐ Variance
 ☐ Range
 ☒ Minimum
 ☒ Maximum

To determine if we have normally distributed data, you can look at the histograms and Q-Q plots to visually inspect the data. For this course, however, we will focus on interpreting the skewness and kurtosis statistics. Specifically, we want skewness and kurtosis statistics that are between -2 and +2. Remember, we have two groups, so we have to check the skewness and kurtosis values for our dependent variable, biased behavior, for both the Christian Student condition and the Control Condition. Looking at the values on the output copied below, we can see our skewness values for competence (-0.41, -0.46) and likeability (-0.37, -0.35) are all within the acceptable range of -2 and +2. Additionally, the kurtosis values for competence (0.31, -0.32) and likeability (-0.36, -0.53) are also within the acceptable range of -2 and +2.

To report these results in APA format, we could write:

Competence and likeability for the Christian Student condition and the Control Condition were normally distributed, skewness and kurtosis statistics were between -2 and +2.

Descriptive Statistics

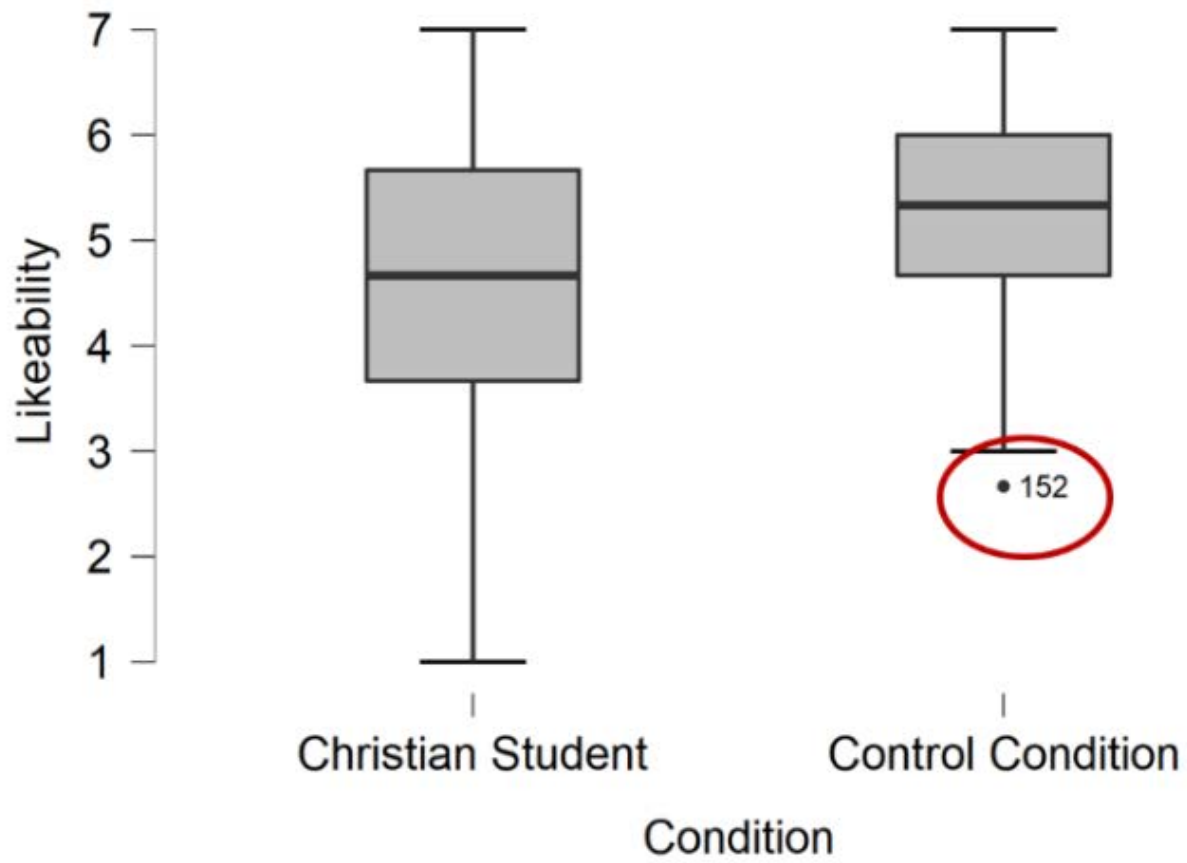
	Competence		Likeability	
	0	1	0	1
Valid	128	133	128	133
Missing	0	0	0	0
Mean	4.682	5.090	4.503	5.268
Std. Deviation	1.237	1.142	1.359	1.047
Skewness	-0.406	-0.460	-0.366	-0.345
Std. Error of Skewness	0.214	0.210	0.214	0.210
Kurtosis	0.309	-0.317	-0.363	-0.530
Std. Error of Kurtosis	0.425	0.417	0.425	0.417
Minimum	1.000	2.333	1.000	2.667
Maximum	7.000	7.000	7.000	7.000

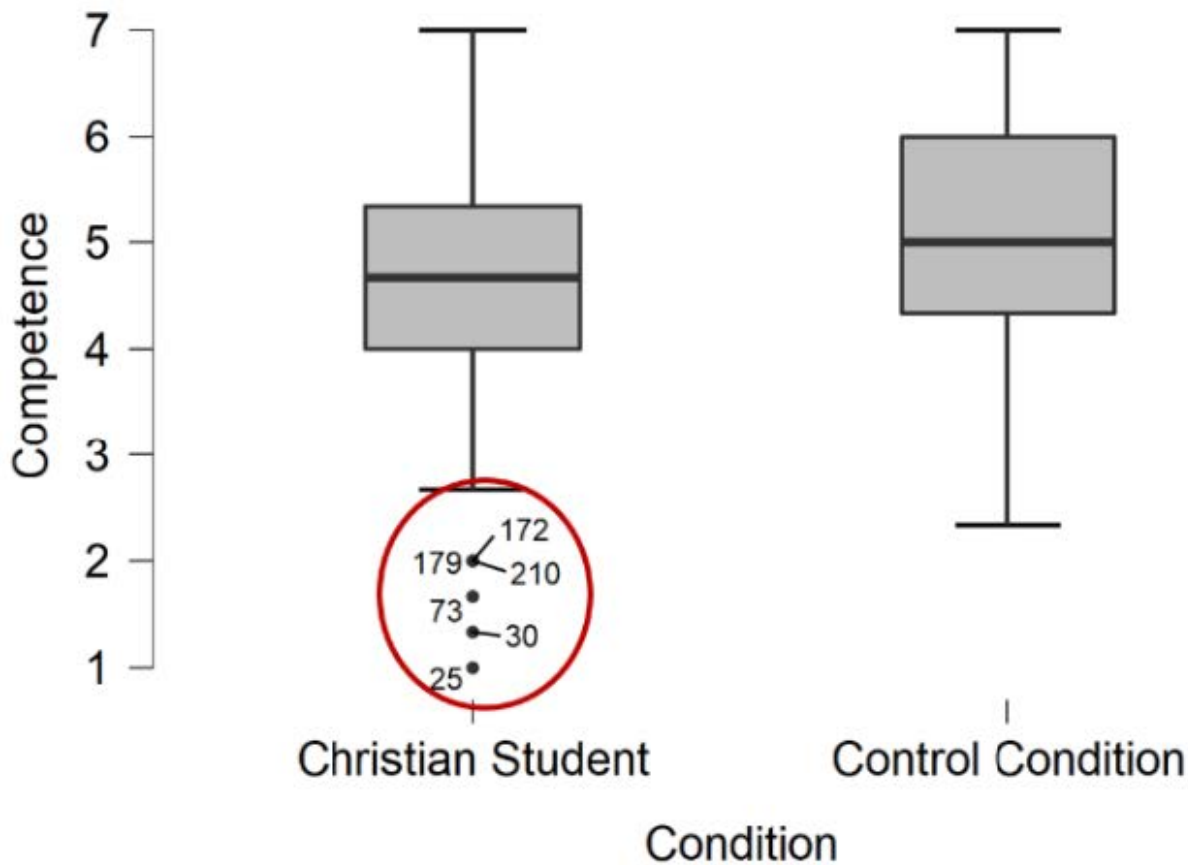
Assumption Five: Are there any outliers in the sample?

To assess for outliers, we will look at the boxplots in our JASP output. If we have outliers, they would be outside the top and bottom lines or whiskers. You can see we have one outlier identified in the boxplot for likeability for the Control Condition, below the lower quartile. Also, we have 6 outliers in the boxplot for competence for the Christian Student condition, below the lower quartile.

To report this in APA format, we could write

There was one outlier in the boxplot for likeability for the Control Condition and six outliers in the boxplot for the competence for the Christian Student condition, as assessed by the inspection of a boxplot.





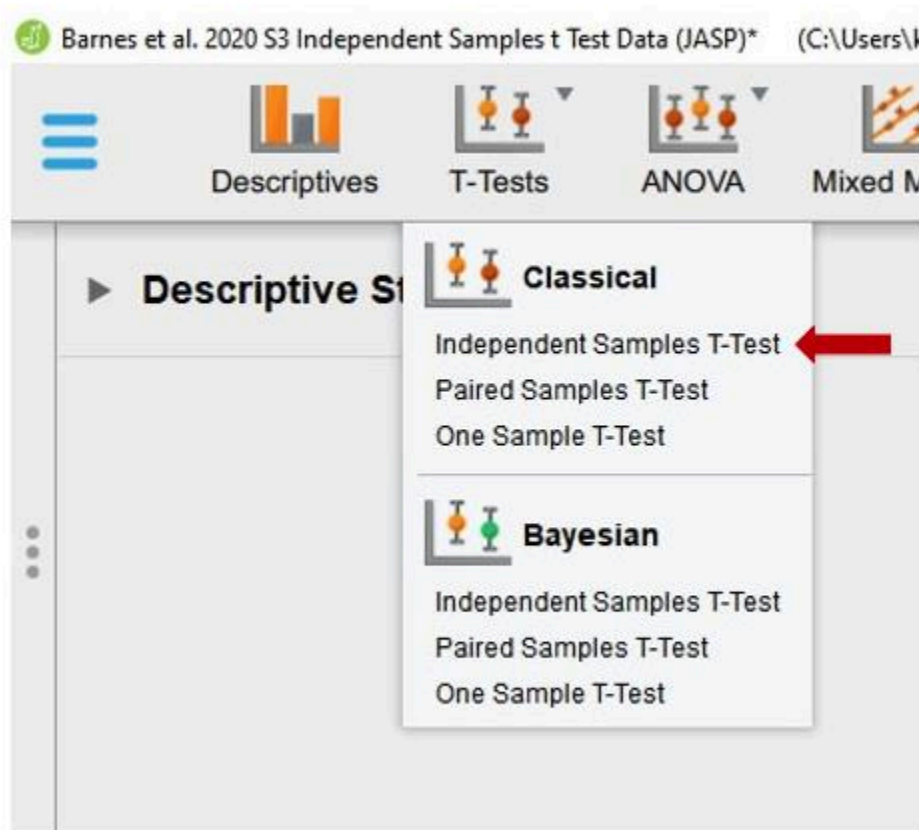
As we have discussed previously, when we have outliers we have to decide if they are going to correct it, keep it, delete it, or replace it. Although we did not collect this data personally, the majority of survey research is collected online through programs such as Qualtrics, QuestionPro, and Survey Monkey. Thus, we can assume the outliers for likeability and competence in the Control and Christian Student conditions were not a data entry error. Cleaning data for random responding and fast survey completion times would have occurred in the data cleaning process, before we got to our analyses, so we can rule those out as well. As a researcher, you would then have to decide if the outliers are exerting a significant enough impact on your results to warrant deletion. It is common for researchers to run their analyses with the outliers retained and with the outliers deleted to see if their exclusion changes the results. With either decision, you would need to add those details when you report your results. In this case, we will retain the outliers; however, we will discuss whether the results would be different if they were deleted when we cover the primary analyses.

Assumption Six: Are there homogeneity of variances?

To determine if we have homogeneity of variances, we need to ask JASP to run our independent samples t test. Checking for this assumption is a part of the overall independent samples t test. Let's move onto our primary analyses below, and complete checking this assumption in that section.

Primary Analyses

To run an independent samples t test, go to T-Tests at the top of the JASP screen and click on Independent Samples T-Test. See the image below.



1First, we need to move our dependent variables [Competence, Likeability] to the “Variables” box and our independent variable [Condition] to the “Grouping Variable” box. We are then going to ask JASP to run the various analyses we will need to interpret for our test by clicking on the appropriate boxes in the test window.

2Under “Tests” we will check the “Student” and “Welch” test boxes.

3Under “Assumption Checks” we will check the “Equality of variances” box.


4Next, we will check the “Effect Size” box and make sure that “Cohen’s d ” is selected.


5Other options that are helpful to ask JASP to provide you with include the “Descriptives” and “Descriptives Plots.”


See the image below to make sure your test window has all the appropriate boxes selected.


▼


Independent Samples T-Test












Participant



Hireability

12


↑

↓


▶

▶


Dependent Variables




Competence




Likeability



Grouping Variable



Condition



Tests <input checked="" type="checkbox"/> Student <input checked="" type="checkbox"/> Welch <input type="checkbox"/> Mann-Whitney Alt. Hypothesis <input checked="" type="radio"/> Group 1 \neq Group 2 <input type="radio"/> Group 1 > Group 2 <input type="radio"/> Group 1 < Group 2 Assumption Checks <input type="checkbox"/> Normality <input checked="" type="checkbox"/> Equality of variances <input type="radio"/> Brown-Forsythe <input checked="" type="radio"/> Levene's	Additional Statistics <input type="checkbox"/> Location parameter <input type="checkbox"/> Confidence interval 95.0 % <input checked="" type="checkbox"/> Effect size <input checked="" type="radio"/> Cohen's d <input type="radio"/> Glass' delta <input type="radio"/> Hedges' g <input type="checkbox"/> Confidence interval 95 % <input checked="" type="checkbox"/> Descriptives <input checked="" type="checkbox"/> Descriptives plots Confidence interval 95.0 % <input type="checkbox"/> Bar plots <input type="checkbox"/> Confidence interval 95.0 % <input type="checkbox"/> Standard error <input checked="" type="checkbox"/> Fix horizontal axis to 0 <input type="checkbox"/> Raincloud plots <input type="checkbox"/> Horizontal display <input type="checkbox"/> Vovk-Sellke maximum p-ratio Missing Values <input checked="" type="radio"/> Exclude cases per dependent variable <input type="radio"/> Exclude cases listwise
--	--

Now that we have asked JASP to run the appropriate analyses, we can begin our interpretation of the results. The first thing we need to do is go back to our last assumption and determine if we meet the assumption for homogeneity of variances.

Are there homogeneity of variances?

- If p is less than .05, you do not meet the homogeneity of variances assumption.
- If p is greater than .05, you do meet the homogeneity of variances assumption.

Assumption Checks

Test of Equality of Variances (Levene's)

	F	df	p
Competence	0.268	1	0.605
Likeability	8.313	1	0.004

To interpret the results of the Levene's test, we need to look at the significance or probability value. If p is less than .05, the test is significant, and we do *not* have homogeneity of variances. If p is greater than .05, the test is non-significant, and we *do* have homogeneity of variances. We meet this assumption.

- If homogeneity of variances was met, you can proceed with interpreting the “Student” independent samples t test results.
- If homogeneity of variances was violated, you can still continue conducting an independent samples t test, but will you need to interpret the “Welch” independent samples t test results instead. The Welch t test is referred to as the unequal variance t test, separate variances t test, or the Welch t test after its creator (Welch, 1947). This test can handle testing mean differences between samples with unequal variances. To include the Welch test in your results, check the “Welch” box under “Tests.” I often click this box when running an independent samples t test automatically, just in case the Levene's test is significant and there are not equal variances.

To report the results of the Levene's test in APA format, it might look something like this:

There was homogeneity of variances, as assessed by the Levene's test for equality of variances, for competence, $p > .05$. However, homogeneity of variances was violated for likeability, $p < .05$.

Interpreting the Statistical Significance of Independent Samples t Tests

The first thing we will interpret is the statistical significance. To do this, we are going to look at the “Independent Samples T-Test” table. Because we have three different dependent variables, we are going to do this one at a time. For our first dependent variable, likeability, we did not meet the assumption of homogeneity of variances. Thus, we are going to interpret the “Welch” t test results in the second row.

Independent Samples T-Test

	Test	Statistic	df	p	Cohen's d
Competence	Student	-2.771	259.000	0.006	-0.343
	Welch	-2.766	255.439	0.006	-0.343
Likeability	Student	-5.110	259.000	< .001	-0.633
	Welch	-5.085	238.676	< .001	-0.631

This is the Cohen's d value, the effect size, for this analysis. A Cohen's d value of .2 is a small effect, .5 is a moderate effect, and .8 is a large effect.

If we have homogeneity of variances, we interpret the results of the Student test. If we do not have homogeneity of variances, we would interpret the results of the Welch test instead.

This is the t statistic for this analysis.

This is the degrees of freedom.

- If p is less than .05, you will reject the null hypothesis.
- If p is greater than .05, you will fail to reject the null hypothesis.

To interpret the results, we want to look at the p value. If p is less than our alpha level of .05, we will reject the null hypothesis. There is a significant statistical difference between the group means. If p is greater than our alpha level of .05, we will fail to reject our null hypothesis. There is not a significant statistical difference between the group means. For our Likeability variable, you can see that our p value is < .001. This is less than our alpha level of .05, so we will reject our null hypothesis. There is a significant statistical difference between the Christian Student condition and the Control Condition on faculty ratings of likeability.

For our second dependent variable, competence, we did meet the assumption of homogeneity of variances. Thus, we are going to interpret the "Student" t test results in the first row. For our Competence variable, you can see that our p value is < .01. This is also less than our alpha level of .05, so we will reject our third null hypothesis. There is a significant statistical difference between the Christian Student condition and the Control Condition on faculty ratings of competence.

Independent Samples T-Test

	Test	Statistic	df	p	Cohen's d
Competence	Student	-2.771	259.000	0.006	-0.343
	Welch	-2.766	255.439	0.006	-0.343
Likeability	Student	-5.110	259.000	< .001	-0.633
	Welch	-5.085	238.676	< .001	-0.631

If we have homogeneity of variances, we interpret the results of the Student test. If we do not have homogeneity of variances, we would interpret the results of the Welch test instead.

This is the t statistic for this analysis.

- If p is less than .05, you will reject the null hypothesis.
- If p is greater than .05, you will fail to reject the null hypothesis.

This is the degrees of freedom.

This is the Cohen's d value, the effect size, for this analysis. A Cohen's d value of .2 is a small effect, .5 is a moderate effect, and .8 is a large effect.

Writing the basic results of the t test in APA format follows this general format:

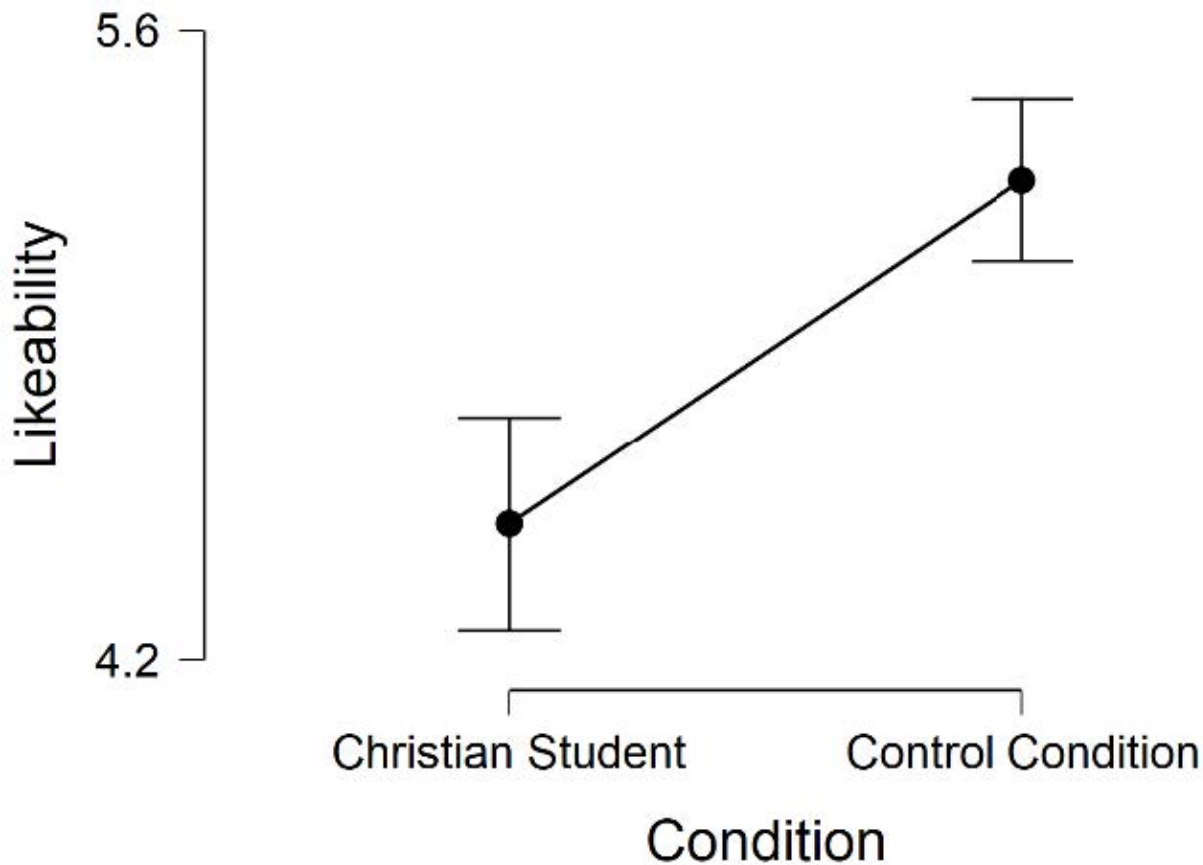
$$t(df) = t \text{ statistic}, p < .05 \text{ or } > .05, d = \text{Cohen's } d \text{ statistic}$$

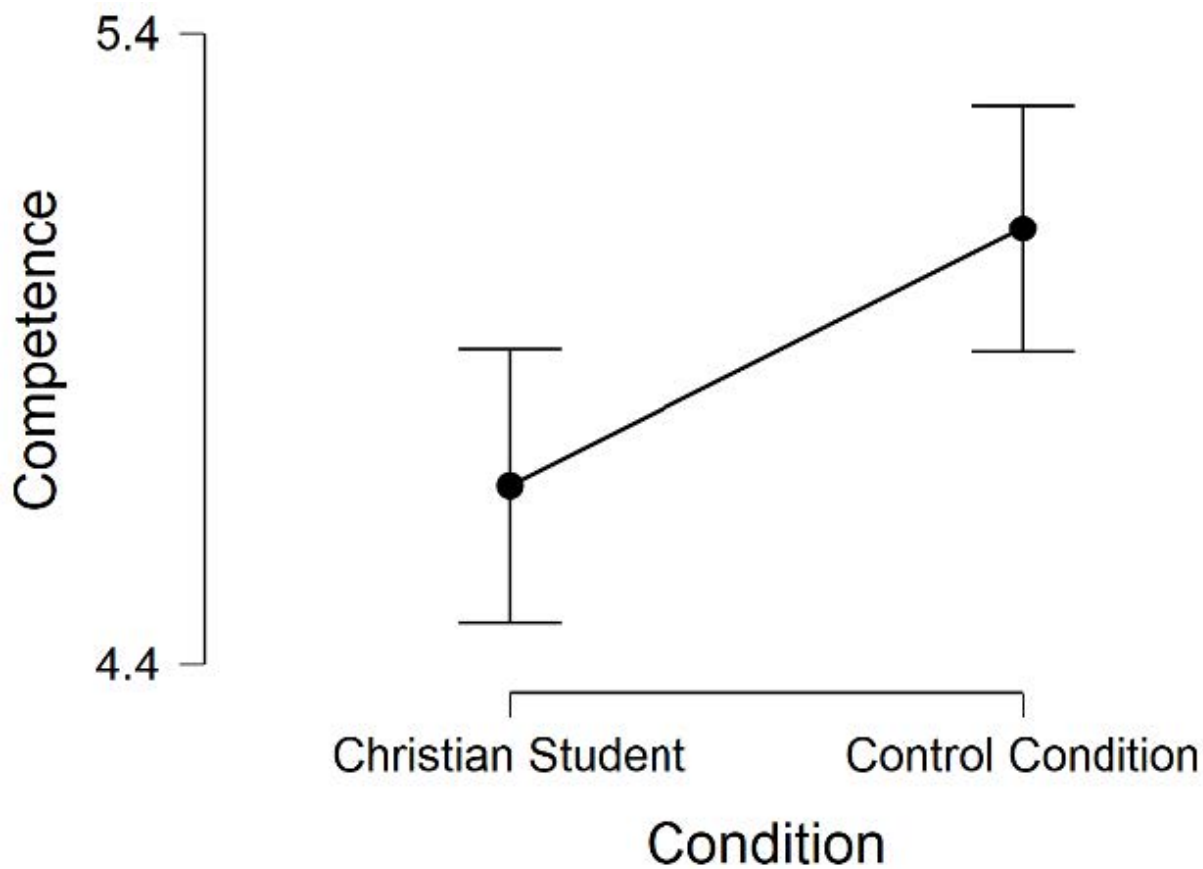
Plugging our results into this format, we have:

Variable	APA Reporting
Likeability	$t(238.68) = -5.09, p < .001, d = -.63$
Competence	$t(259) = -2.77, p < .01, d = -.34$

Note. Remember, with our Likeability variable, it did not have homogeneity of variances, so we needed to interpret and report the reports of the Welch statistic.

Now we know that we have a significant difference between the mean levels of competence and likeability between faculty assigned to the Christian Student condition and the Control Condition in this study, but what does that mean? Who did faculty rate as having higher levels of likeability and competence in the science field? To answer this, let's go back to our "Group Descriptives" and "Descriptives Plots" to look at the means and standard deviations for our groups on our outcome variance.





If we look at the Descriptives Plot in our JASP output, we can see a graphical representation of our results. This graph plots the means for each of our groups on the dependent variable. Looking at the *y-axis*, you can see the label for our dependent variables “likeability” and “competence.” Looking at the *x-axis*, you can see the label for the levels of our independent variable “Christian Student” and “Control Condition.” Just looking at the means represented on this plot, it is easy to see that the Control Condition was rated as more likeable and competent than the Christian Student condition.

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE
Competence	0	128	4.682	1.237	0.109
	1	133	5.090	1.142	0.099
Likeability	0	128	4.503	1.359	0.120
	1	133	5.268	1.047	0.091

If we look at our Group Descriptives table, we can see the sample sizes (N), means (Mean), standard

deviations (SD), and standard errors (SE) for both groups. We will want to include this information in our results. Formatting this information into APA format, we might report:

There were 128 faculty randomly assigned to the Christian Student condition and 133 faculty randomly assigned to the Control Condition who participated in this study. Faculty in the Control Condition rated the applicant as having a higher level of likeability ($M = 5.27, SD = 1.05$), compared to faculty ratings in the Christian Student Condition ($M = 4.50, SD = 1.36$). Additionally, faculty in the Control Condition rated the applicant as having a higher level of competence ($M = 5.09, SD = 1.14$), compared to faculty ratings in the Christian Student Condition ($M = 4.68, SD = 1.24$).

Interpreting the Practical Significance of Independent Samples t Tests

Now that we have interpreted the statistical significance, we will look at the practical significance by looking at our effect size. A significant p value tells us that there is a difference in mean level of our dependent variables (likeability and competence) between the Christian Student and Control Conditions in this study, but the effect size tells us how big this difference is. Unlike p values, the Cohen's d effect size test is not impacted by the sample size.

Reminder: Cohen's d Effect Sizes

Effect Size	Strength
.2	Small
.5	Medium
.8	Large

We obtained Cohen's d values of $-.63$ for Likeability and $-.34$ for Competence. We obtained a small effect size for faculty Competence ratings and a medium effect size for faculty ratings of applicant Likeability.

Reporting in APA Format

What do you need to report in your results?

Value (Notation)	Levene's Statistic
Means (M)	Standard Deviations (SD)
Test Statistic (t)	Degrees of Freedom (df)
Probability (p)	Effect Size
Cohen's d	Confidence Interval
Skewness and Kurtosis	Boxplot

Bonferroni Correction

For this example, we had three separate hypotheses, one for each of our dependent variables. Because we are

conducting multiple comparisons with the same sample, we need to correct for experimentwise or familywise error to keep our Type I error rate at 5%. With an alpha level of .05, we have a 5% chance of making a Type I error (rejecting a null hypothesis that is true).

Hypothesis One	Comparing the mean faculty rating of applicant likeability for the Christian Student Condition with the Control Student Condition.	5% Type I Error Rate
Hypothesis Two	Comparing the mean faculty rating of applicant competence for the Christian Student Condition with the Control Student Condition.	5% Type I Error Rate
		10% Total Type I Error Rate

To reduce our chance of making a Type I error with multiple comparisons, we can conduct a Bonferroni Correction. To do this, we divide our alpha level (α) by the number of comparisons (k). In this example, our alpha level is .05 and we had a total of three comparisons.

$$P_{\text{Critical}} = \alpha \div k$$

$$P_{\text{Critical}} = .05 \div 2 = 0.025$$

Given the p critical results above, we would need to interpret our p values with an adjusted alpha level or Bonferroni Correction, of 0.025. This means, in order to reject our null hypotheses, all of our p values would need to be below 0.025 in order to be considered statistically significant. Prior to reporting your analyses, you would add to report:

Analyses were conducted with a Bonferroni adjustment of $p < .025$.

Putting it All Together

Now, let's write up the results of our test combining everything we've done so far, including the test of assumptions, and the results of the t test.

A set of independent samples t test were conducted to determine if there were mean differences in faculty ratings of a PhD applicant's likeability and competence between faculty assigned to read application materials for a Christian Student versus a Control Student condition. There was one outlier for likeability for the Control Condition and six outliers for the competence for the Christian Student condition, as assessed by the inspection of a boxplot. Likeability and competence were normally distributed, with skewness and kurtosis values between -2 and +2. There was homogeneity of variances, as assessed by the Levene's test for competence ($p > .05$), but not for likeability ($p < .05$).

Analyses were conducted with a Bonferroni adjustment of $p < .025$. Faculty in the Control condition rated PhD applicants as having higher levels of likeability and competence. Specifically, faculty in the Control Condition rated the applicant as having a higher level of likeability ($M = 5.27$, $SD = 1.05$), compared to faculty ratings in the Christian Student Condition ($M = 4.50$, SD

$= 1.36$), $t(238.68) = -5.09$, $p < .001$, $d = -.63$. Additionally, faculty in the Control Condition rated the applicant as having a higher level of competence ($M = 5.09$, $SD = 1.14$), compared to faculty ratings in the Christian Student Condition ($M = 4.68$, $SD = 1.24$), $t(259) = -2.77$, $p < .01$, $d = -.34$.

Real World Meaning

When we interpret the real world meaning of a study, we want to take out any statistical jargon and describe the results of the study in a way that is easy to understand by people who do not know anything about statistics. How would you describe the results to your roommate, sibling, parent, or neighbor? For this study, we want to not only say there are differences between faculty ratings when reading mock application materials for a Christian Student compared to a Control Student where their religiosity is not mentioned, but we want to describe which group has higher or lower scores. We also need to translate our outcome variables “likeability and competence” into terms that are easy to understand by others. Thankfully the names of these variables are fairly straightforward, but that is not always the case in psychological research. For example, if I was telling my brother the real world meaning of our results, I would say:

Researchers found that biology faculty members that read application materials for a fake potential PhD student where their religion was not discussed perceived the applicant as more likeable and more competent than biology faculty that read fake application materials for a Christian Student who discussed their religion and evangelical religious activities. The results of this study suggest that individuals within the science community may be more likely to perceive PhD Christian students as less likeable and less competent than PhD students who did not explicitly state their religious beliefs.

Barnes, Truong, and colleagues (2020) conducted additional analyses that showed that atheist faculty showed a stronger bias against the Christian Student condition compared to faculty who identified as Christian themselves. The authors suggested this may be due to a historical bias against science in fundamentalist and evangelical Christian beliefs (Marsden, 2015; Numbers, 2006). This includes, for example, movements by these groups to teach “creationism in US science classes in an attempt to discredit evolution to students” (Barnes et al., 2020, p. 12; Berkman & Plutzer, 2011). However, the scenario faculty read in this study did not include any mention of the fake graduate student applicant’s political beliefs and may have used stereotypes to judge the candidate. The authors conducted a total of three separate studies. We will analyze data from their second study in a future assignment to see if they observed similar results when comparing graduate student applications that mentioned activities that were either Christian, Atheist, or a Control condition.

References

Barnes, M. E., Truong, J. M., Grunspan, D. Z., & Brownell, S. E. (2020). Are scientists biased against Christians? Exploring real and perceived bias against Christians in academic biology. *PLoS ONE*, 15(1). <https://doi.org/10.1371/journal.pone.0226826>

Berkman, M. B. & Plutzer, E. (2011). Defeating creationism in the courtroom, but not in the classroom. *Science*, 331, 404-405.

Ecklund, E. H., Johnson, D. R., Scheitle, C. P., Matthews, K. R. W., & Lewis, S. W. (2016). Religion among Scientists in International Context: A New Study of Scientists in Eight Regions. *Socius : Sociological Research for a Dynamic World*, 2, 237802311666435–. <https://doi.org/10.1177/2378023116664353>

Marsden, G.M. (2015). Religious discrimination in academia. *Society*, 52(1), 19-22.

Numbers, R. L. (2006). *The Creationists: From scientific creationism to intelligent design*. Harvard University Press.

Pew Research Center (2019). In U.S., decline of Christianity continues at rapid pace. Retrieved from <https://www.pewforum.org/2019/10/17/in-u-s-decline-of-christianity-continues-at-rapid-pace/>

Authors

This guide was written and created by Kelsey L. Humphrey, Ruth V. Walker, PhD, and Erin N. Prince.

Acknowledgements

We would like to thank M. Elizabeth Barnes, Jasmine M. Truong, Daniel Z. Grunspan, and Sara E. Brownell for making their data available through PLoS One. Additionally, we appreciate Dr. Kristen Black's edits and suggestions during the creation of this guide.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting an Independent Samples t test in JASP Copyright © 2021 by Kelsey L. Humphrey; Ruth Walker; and Erin N. Prince is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

2.

RUNNING AND INTERPRETING A PAIRED SAMPLES T TEST IN JASP

Kendra E. Scott; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker

Download Data Sets:

- JASP Data File Paired Samples t test
- CSV Data File Paired Samples t test
- JASP Data File Paired Samples t test (Outliers Deleted)
- CSV Data File Paired Samples t test (Outliers Deleted)

The Association of American Universities conducted a Campus Climate Survey on Sexual Assault and Misconduct in 2019. After surveying over 181,000 students across 33 colleges and universities, they found that 13% reported experiencing non-consensual sexual contact. With women and transgender students reporting significantly higher rates than men. One way that colleges and universities have been working to reduce rates of sexual assault is through educational programs that teach students how to intervene to prevent sexual assault. Researchers Banyard, Moynihan, and Plante (2007) created a bystander intervention program to prevent sexual assault on college campuses. In their research article, they present data demonstrating the efficacy of effectiveness of their intervention program.

To study the efficacy of the program, participants were separated into three different groups: a control group that did not attend a prevention program, a group that attended a one-session prevention program, and a group that attended a three-session prevention program. For our analyses, we are going to analyze the pre-and-post-test bystander self-efficacy scores for the 1-Session Program group.

Bystander self-efficacy was measured in this study using a 14-item measure of participant confidence to perform bystander behavior. Participants were asked to rate how confident they are that they can perform 14 different types of bystander behavior from 0 (“*can’t do*”) to 100 (“*very certain can do*”). Sample questions included items such as, “How confident are you that you could ask a stranger who looks very upset at a party if they are ok or need help?” Higher scores mean the participant was more confident in their ability to perform bystander behaviors. The pre-test variable name is “PreBSE” and the post-test variable name is “PostBSE.”

Hypotheses

The null hypothesis is:

- Conceptual H_0 : There is no significant difference in bystander self-efficacy scores for the One-Session Prevention Group before and after the bystander intervention program.
- Mathematical H_0 : The population mean difference on bystander self-efficacy scores before and after the bystander intervention program is equal to zero; $M_D = 0$.

The alternative hypothesis is:

- Conceptual H_0 : There is significant difference in bystander self-efficacy scores for the One-Session Prevention Group before and after the bystander intervention program.
- Mathematical H_0 : The population mean difference on bystander self-efficacy scores before and after the bystander intervention program is not equal to zero; $M_D \neq 0$.

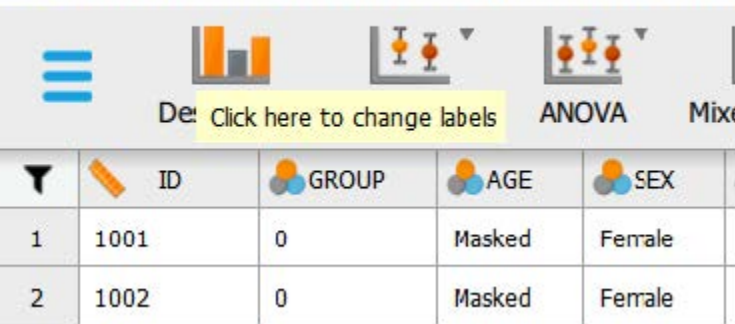
JASP Analyses

In order to run analyses, the first thing we need to do is open the data set we will be working with. To do this, open JASP and follow the steps below.

File → Open → Computer → Browse → Select the Paired Samples t Test Data (Banyard et al. 2007 Paired Samples t Test Data JASP) file wherever it's saved on your computer.

Once the data set is open in JASP, we will change the data labels for our group variable so that we can select for those just in the 1-Session Program group. Currently the group column has a 0, 1, or a 2 for each participant. We will need to change these labels so that 0 = 'control,' 1 = '1-Session Program,' and 2 = '3-Session Program.' To change the numerical data into our categorical labels, you will take your cursor and hover over the 'group' column. When you see a note pop up saying, "click here to change labels" click on it. To change the labels, click on the 0 under the Label column and type 'control'. Then click on the 1 under the Label column and type '1-Session Program'. Then click on the 2 under the Label column and type '3-Session Program'. After you have changed the labels, you can close the window by clicking on the X button.

Banyard et al. 2007 Paired Samples t Test Data (JASP) (C:\Users\




	ID	GROUP	AGE	SEX
1	1001	0	Masked	Female
2	1002	0	Masked	Female

GROUP		
Filter	Value	Label
✓	1	Control
✓	2	1-Session Program
✓	3	3- Session Program

Before we test our assumptions, we also need to filter out participants who were in the control condition and 3- Session Program, so that we're only looking at scores for those participants in the One-Session Program. One way to do this is to click on the Condition column to bring up the Values and Labels box. To filter out those in the control condition and Three-Session Program, under the "Filter" column, click the check mark – it will turn it into an X. Now, when we run any analysis, JASP will only consider those participants in the One-Session Program in the calculations!

GROUP		
Filter	Value	Label
✓	1	Control
✓	2	1-Session Program
✓	3	3- Session Program



Click on the check mark under the "Filter" Column for the row "control" to filter out those in the control condition.

GROUP		
Filter	Value	Label
✗	1	Control
✓	2	1-Session Program
✗	3	3-Session Program

Assumption Testing

Of the four assumptions, assumptions #3 and #4 must be tested prior to conducting a paired samples t test. Let's consider whether our example data meets all four assumptions:

Assumption 1: Is there one dependent variable that is measured at the continuous (i.e., ratio or interval) level?

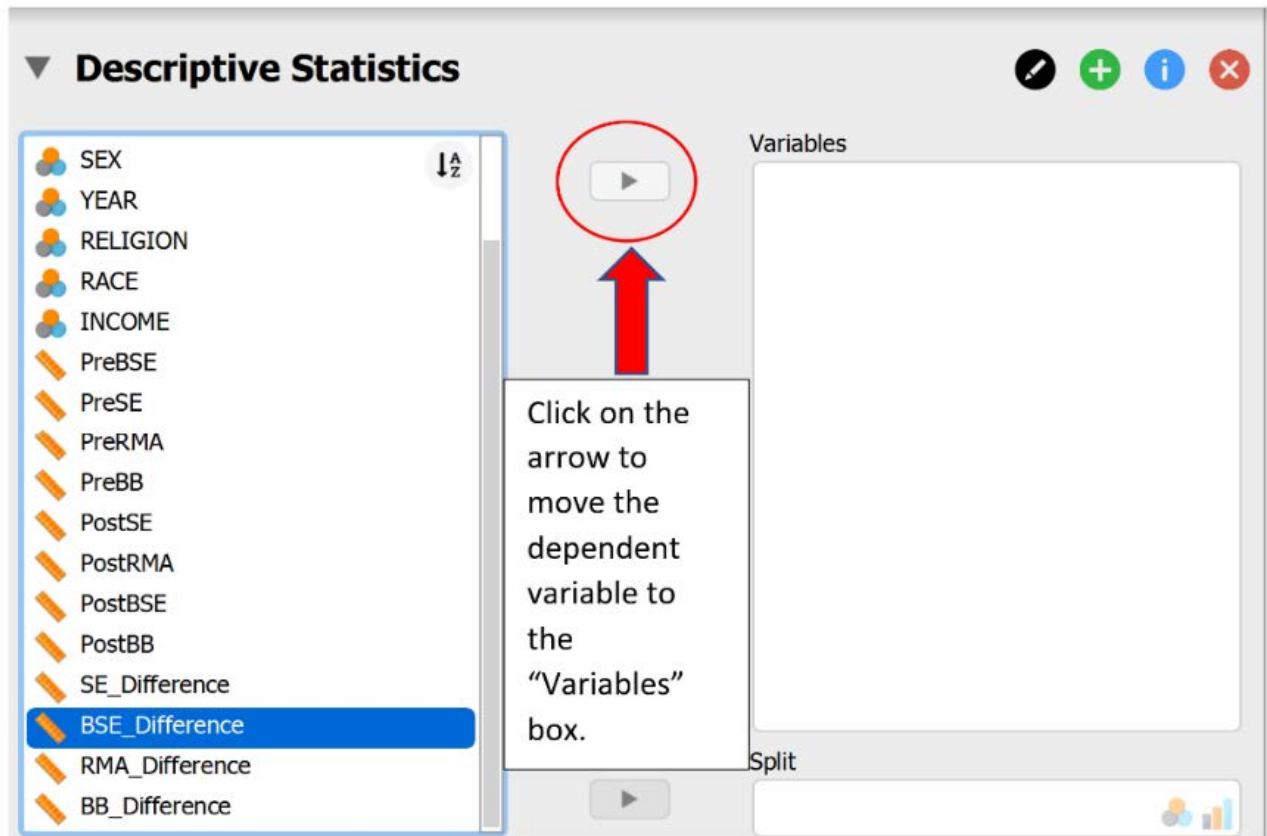
Yes. We have one dependent variable (bystander self-efficiency scores) and this dependent variable is a singular composite score of bystander self-efficiency based on 14 different types of bystander behaviors measured on a Likert-type scale. In psychological research, composite scores (i.e., averages of multiple individual questions) are treated as continuous data. Therefore, we meet this assumption.

Assumption 2: Is there one independent variable that consists of two categorical related groups or matched pairs?

Yes. We have a research design where each participant provides a score on the same dependent variable at two separate time points (before and after experiencing bystander intervention programs – “PreBSE” and “PostBSE.” in the dataset). Therefore, the scores are related across two (i.e., categorical) unique time points. Therefore, we meet this assumption.

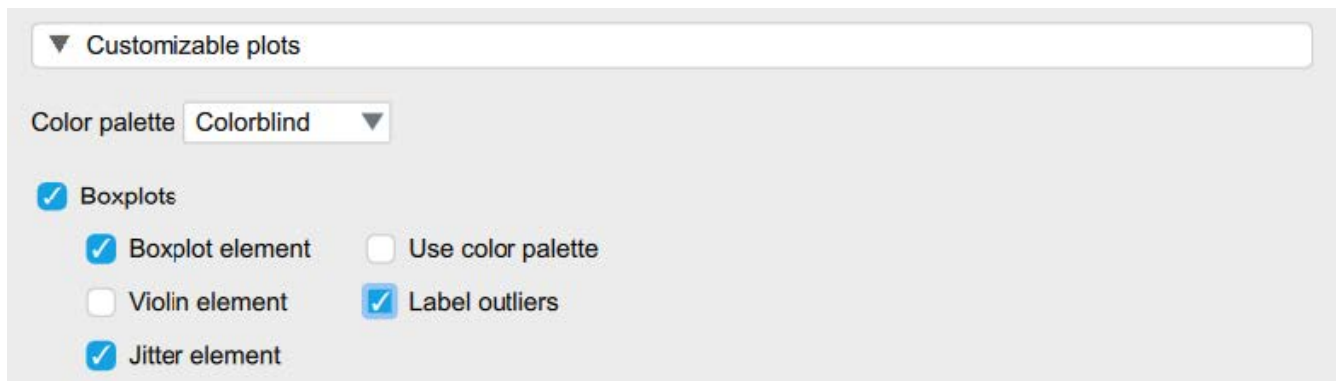
Assumption 3: Are there any significant outliers in the difference scores between the two paired groups?

To check our data for Assumption 3 (and Assumption 4 to follow), we will be working with the Descriptives tab. When the “Descriptive Statistics” window pops up, we will need to move the difference score of our dependent variable (labeled “BSE_ Difference”) to the “Variables” box. This difference score was calculated by subtracting the pretest scores from the posttest scores on the dependent variable.



Then, under the “Customizable Plots,” “Basic Plots,” and “Statistics” tabs, select the following check marks.

1. Under the “Customizable Plots” tab we will check the “Boxplots,” “Boxplot element,” “Jitter element,” and “Label outliers” boxes.
2. We will also check “Distribution plots” and “Q-Q plots” under the “Basic plots” tab.
3. Under the “Statistics” tab we will check the “Skewness,” “Kurtosis,” and “Shapiro-Wilk test” boxes.



▼ Basic plots

☒ Distribution plots
 ☐ Correlation plots
 ☐ Interval plots

☒ Display density
 ☒ Q-Q plots

☐ Display rug marks
 ☐ Pie charts

Bin width type: Sturges

Number of bins: 30

☐ Dot plots

▼ Statistics

Sample size
☒ Valid
☒ Missing

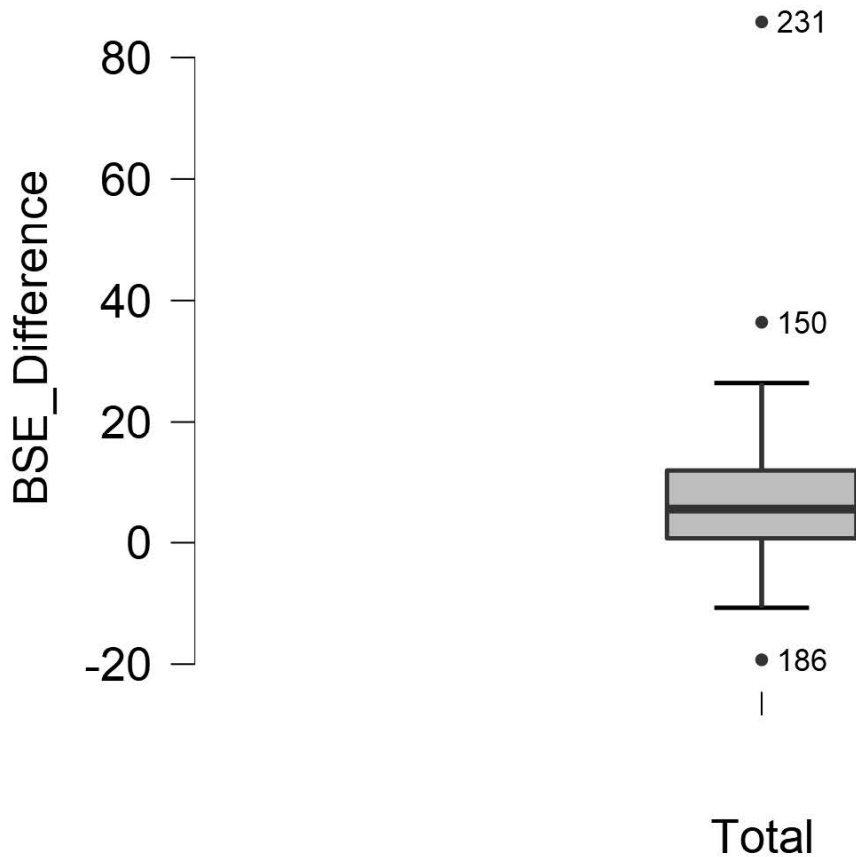
Quantiles
☐ Quartiles
☐ Cut points for: 4 equal groups
☐ Percentiles:

Central tendency
☐ Mode
☐ Median
☒ Mean

Distribution
☒ Skewness
☒ Kurtosis
☒ Shapiro-Wilk test
☐ Sum

Dispersion
☐ S.E. mean
☒ Std. deviation
☐ Coefficient of variation
☐ MAD
☐ MAD robust
☐ IQR
☐ Variance
☐ Range
☒ Minimum
☒ Maximum

To determine if we have any outliers, we will look at the boxplots in our JASP output. If we have any outliers, they would be outside the top and bottom lines or whiskers. To help see the labeled outliers more clearly, you may want to uncheck the “jitter element” option. Looking at the boxplot below, we can see there are a total of 3 outliers labeled on the boxplot.



To report this in APA format, we would write:

There are 3 outliers in the difference scores for bystander self-efficacy, as assessed by the inspection of a boxplot.

Assumption 4: Is the distribution of the difference scores between the two related groups approximately normally distributed?

To determine if the difference scores for bystander efficacy are approximately normally distributed, we want skewness and kurtosis values between -2 and +2. Looking at the values on the output copied below, we can see our skewness statistic is not within the acceptable range of -2 and +2 (skewness: 2.91); additionally, our kurtosis statistic is a whopping 18.99! Neither of these statistics are within the acceptable range. Because normality is impacted by outliers, we may want to consider deleting the outliers to improve our normality statistics.

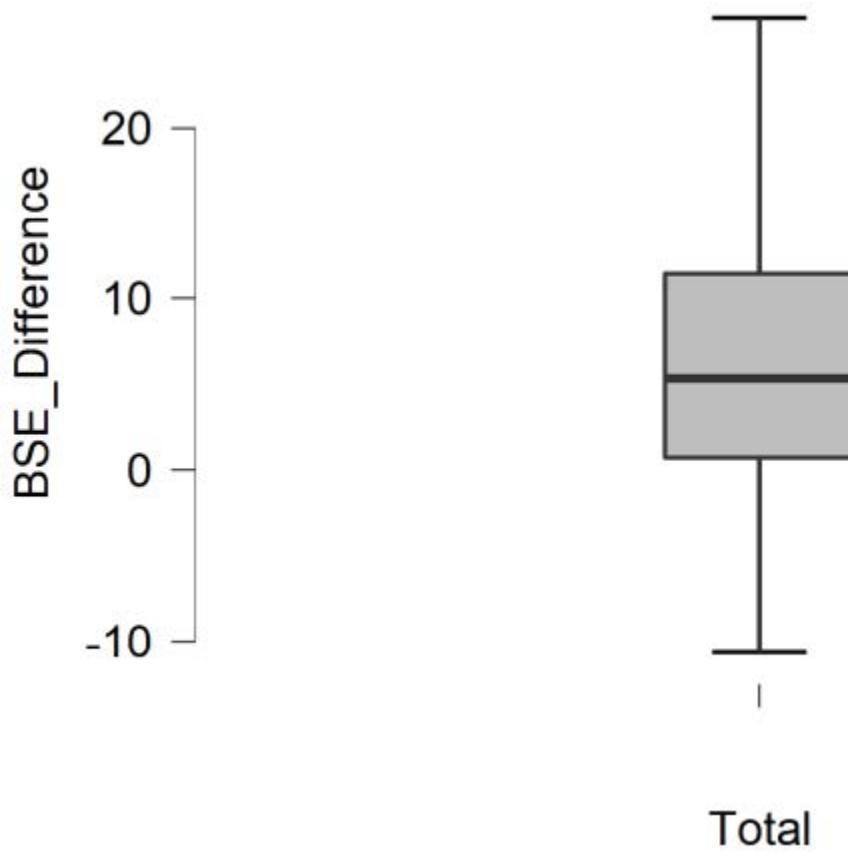
Descriptive Statistics

	BSE_Difference
Valid	128
Missing	9
Mean	7.333
Std. Deviation	11.098
Skewness	2.907
Std. Error of Skewness	0.214
Kurtosis	18.987
Std. Error of Kurtosis	0.425
Minimum	-19.286
Maximum	85.857

To report these results in APA format, we can write:

The difference scores for bystander self-efficacy were not normally distributed; skewness and kurtosis statistics were above the acceptable range of -2 and +2 at 2.91 and 18.99.

Let's delete our outliers in order to meet our assumptions and recheck our assumptions. You can open the data file in Excel and delete the outliers yourself, or simply open the "Banyard et al. 2007 Paired Samples t Test Data Outliers Deleted JASP" file. Looking at our boxplot after deleting the 3 outliers, we can see there are no longer any outliers visible on our boxplot.



Rerunning our descriptive statistics, you can also see our skewness and kurtosis statistics have improved dramatically with the removal of the outliers and are now within the acceptable range of -2 and +2 at .36 (skewness) and -.26 (kurtosis).

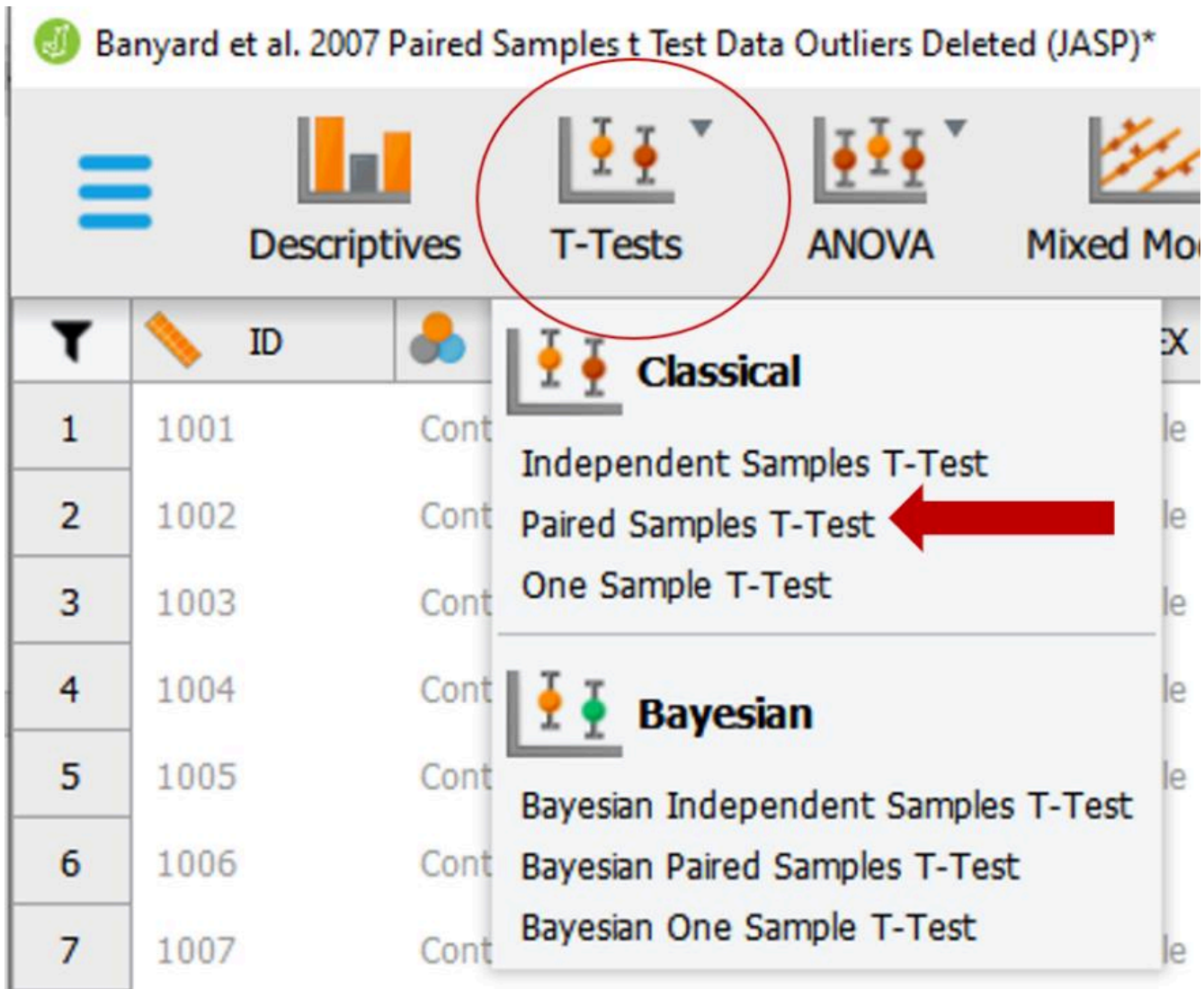
Descriptive Statistics

	BSE_Difference
Valid	125
Missing	9
Mean	6.685
Std. Deviation	7.966
Skewness	0.360
Std. Error of Skewness	0.217
Kurtosis	-0.259
Std. Error of Kurtosis	0.430
Minimum	-10.714
Maximum	26.429

If you were writing these results up in APA formatting, you would want to let your readers know what you have done with your data (i.e., deleting outliers) and how that improved normality statistics. Researchers will sometimes report whether their primary analyses are different with the exclusion or inclusion of outliers in the dataset. In case you were wondering, the results of our analyses would be the same either way, but our data are normally distributed with the outliers deleted.

Primary Analyses

To run a paired samples t test, go to T-Tests at the top of the JASP screen and click on Paired Samples T-Test. We are going to conduct our primary analyses with the dataset that does not include the 3 original outliers, ‘Banyard et al. 2007 Paired Samples t Test Data Outliers Deleted (JASP).’



To run the paired samples t test, we need to move our two variable pairs ('PreBSE' and 'PostBSE') into the "Variable pairs" box. Then, we need to select the following check boxes as depicted in the image below:

- Tests: Student
- Additional Statistics: Effect Size, Descriptives, Descriptive Plots

Paired Samples T-Test

Variable Pairs: PreBSE, PostBSE

Tests

- ☒ Student
- ☐ Wilcoxon signed-rank

Alt. Hypothesis

- ☒ Measure 1 \neq Measure 2
- ☐ Measure 1 > Measure 2
- ☐ Measure 1 < Measure 2

Additional Statistics

- ☐ Location parameter
 - ☐ Confidence interval 95.0 %
- ☒ Effect size
 - ☐ Confidence interval 95.0 %
- ☒ Descriptives
 - Confidence interval 95.0 %
- ☒ Descriptives plots
 - Confidence interval 95.0 %
 - ☐ Bar plots
 - ☐ Confidence interval 95.0 %
 - ☐ Standard error
 - ☒ Fix horizontal axis to 0
 - ☐ Raincloud plots
 - ☐ Raincloud difference plots
 - ☐ Horizontal display
 - ☐ Vovk-Sellke maximum p-ratio

Assumption Checks

- ☐ Normality

Missing Values

- ☒ Exclude cases per dependent variable

Interpreting the Statistical Significance of Paired Samples t Tests

The first thing we will examine is the statistical significance of the paired samples t test. To do this, we are going to look at the “Paired Samples T-Test” table from the output.

- If p is less than .05, you will reject the null hypothesis.
- If p is greater than .05, you will fail to reject the null hypothesis.

Paired Samples T-Test ▼

Measure 1		Measure 2	t	df	p	Cohen's d
PreBSE	-	PostBSE	-9.382	124	< .001	-0.839

Note. Student's t-test.

This is the t statistic for this analysis.

This is the degrees of freedom.

This is the Cohen's d value, the effect size, for this analysis. A Cohen's d value of .2 is a small effect, .5 is a moderate effect, and .8 is a large effect.

To interpret the results, we want to look at the p value. If p is less our alpha level of .05, we will reject the null hypothesis (indicating there is a significant statistical difference before and after the intervention on our dependent variable). If the p value is greater than our alpha level of .05, we will fail to reject the null hypothesis (indicating there is not a significant statistical difference before and after the intervention on our dependent variable).

In this example, you can see that our p value is <.001, which is less than the alpha of .05. Therefore, we will reject our null hypothesis – there is a significant difference in the bystander self-efficacy scores for One-Session Prevention Group before and after the bystander intervention program.

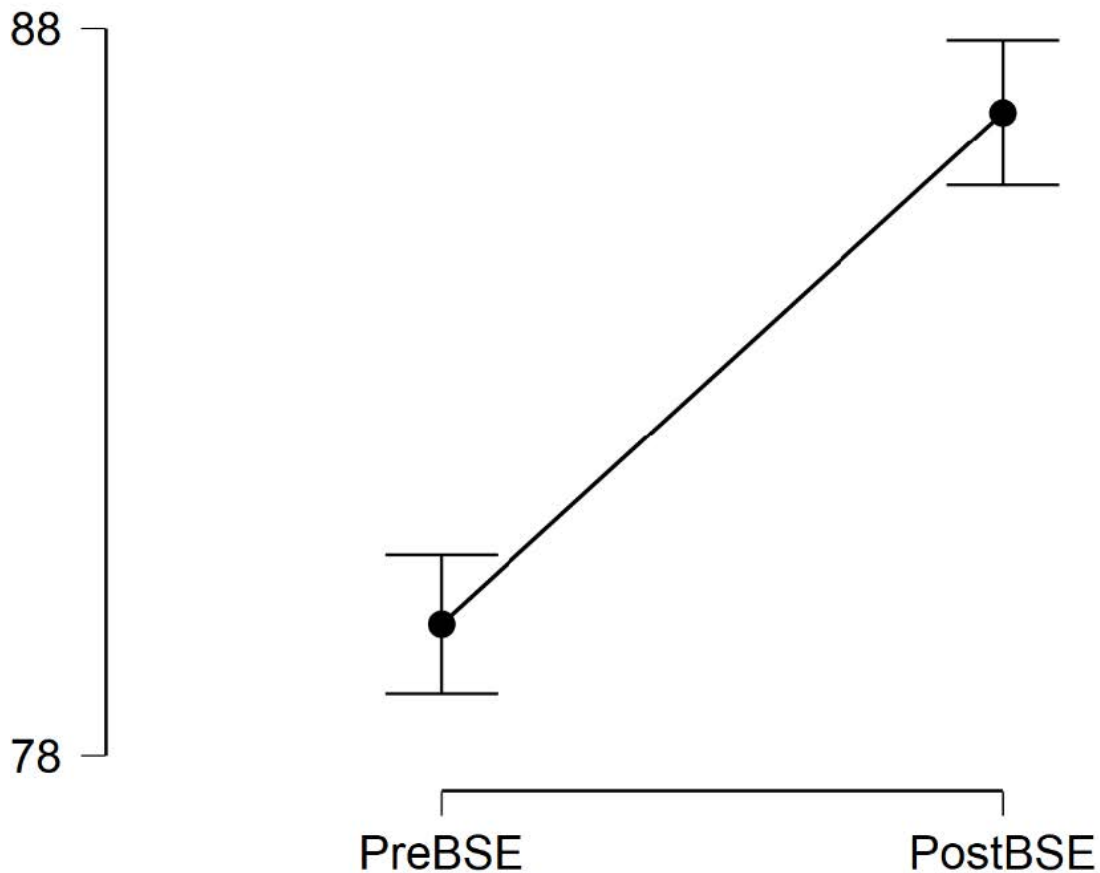
Writing the basic results of the t test in APA format following this general format:

$$t(df) = t \text{ statistic}, p < .05 \text{ or } p > .05, d = \text{Cohen's } d \text{ statistic}$$

Plugging in our results into this format should look like this:

$$t(124) = -9.38, p < .001, d = -.84$$

Now that we know there are significant statistical differences in the bystander self-efficacy scores for One-Session Prevention Group before and after the bystander intervention program – in what direction is this difference? Did people experience greater efficiency scores before, or after, the bystander intervention program? To answer this question, let's look at the “Descriptives Table” and “Descriptive Plots” from our output to look at the means and standard deviations for our two groups of scores.



Remember, the descriptive plots provide a graphical representation of our results. This plot graphs the mean levels of bystander self-efficiency scores at time 1 (i.e., before the bystander intervention program) and at time 2 (i.e., after the bystander intervention program). By inspecting this graph, it is easy to see that participants reported higher scores of bystander self-efficiency after attending the bystander intervention program than before attending it.

Descriptives				
	N	Mean	SD	SE
PreBSE	134	79.816	13.422	1.160
PostBSE	125	86.842	11.923	1.066

What we see visually from the Descriptives Plot is provided numerically in the Descriptives table. Provided in this table is information about sample size (N), means (Mean), standard deviations (SD) and standard errors (SE) for both groups of scores (i.e., at both time points). This table is important, because we will want to provide this information in the reporting of our results. Formatting this information into APA format, we might report the following:

A total of 124 individuals participated in this study and were in the 1- Session Program group.

Participants reported significantly higher levels of bystander self-efficacy scores after ($M = 86.84$, $SD = 11.92$) as compared to before ($M = 79.82$, $SD = 13.42$) they attended the bystander intervention program.

Interpreting the Practical Significance of Paired Samples t Tests

Now that we have interpreted the statistical significance of the paired samples t test, let's consider the practical significance of these results by looking at the effect size. Remember, a significant p value tells us that there is a *statistically* significant difference in self-reported bystander efficacy, but the effect size will tell us how large this difference actually is. Unlike p values, the Cohen's d effect size measure is not impacted by sample size.

Cohen's d Effect Sizes

<i>Effect Size</i>	<i>Strength</i>
.2	Small
.5	Medium
.8	Large

The obtained Cohen's d value in this example was -.84. Given that this value falls above the cut-off of .8 for a large effect, we would say that our results suggest a large effect. What this tells us is that not only are there statistically significant differences in bystander self-efficacy scores before and after attending a bystander intervention program to prevent sexual assault, but this difference is *large*. Meaning, after attending one-session of a sexual assault bystander intervention program, participants felt significantly *more confident* in their ability to intervene to prevent a sexual assault from happening.

Reporting in APA Format

What do you need to report in your results?

Value (Notation)	Descriptives and Test Statistics
Mean (M)	Standard Deviation (SD)
Test Statistic (t)	Degrees of Freedom (df)
Effect Size	Cohen's d
Assumptions	Skewness and Kurtosis
Boxplot	

Putting it All Together

A paired samples t test was conducted to determine if there were significant differences in bystander self-efficacy scores before and after attending a bystander intervention program. To correct for issues related to

normality and outliers, three outliers were deleted from the dataset. The remaining data had no outliers, as assessed by the inspection of a boxplot and the difference scores for bystander efficacy were normally distributed, with skewness and kurtosis values within the acceptable range of -2 and +2. Participants reported significantly higher levels of bystander self-efficiency scores after ($M = 86.84$, $SD = 11.92$) as compared to before ($M = 79.82$, $SD = 13.42$) they attended the bystander intervention program, $t(124) = -9.38$, $p < .001$, $d = -.84$.

Real World Meaning

Recall that when we interpret the real world meaning of a study, we want to refrain from using any statistical jargon—that is to say, we want people who know nothing about statistics (a parent, your roommate, your friends, etc.) to be able to understand what the results of our statistical test are telling us about psychological phenomenon.

How would you describe the results of this study to someone who knows nothing about statistics? One thing that's helpful is to think about what important pieces of information do we want to come across to the audience? In this example, we want to be able to communicate that the participants, reported higher scores of bystander self-efficiency scores after attending one session of a bystander intervention program compared to before attending one session. If I were describing the results of this study to my uncle, I would say the following:

Bystander intervention programs teach people ways they can stand up and help intervene to prevent sexual assault. Participants who attended one-session of a bystander intervention program felt more confident in their ability to act to prevent sexual assault after the program.

It should be noted that Banyard and colleagues (2007) conducted more complex analyses than what we cover in this course. The researchers included a Control Group and a 3-Session Intervention Group as well to establish a stronger experimental research design. They also measured participants across additional time points at 2-months, 4-months, and 12 months after the intervention. Because they also had multiple dependent variables they were interested in, the analyses they conducted were called a multivariate analysis of variance (MANOVA) to see if there were differences between the three conditions (i.e., control, one-session, and three-session groups). Additionally, they conducted repeated measures analysis to look at change over time, while controlling for participant gender by conducting a repeated measure multivariate analysis of covariance (MANCOVA). You will be very happy to know both of those analyses are beyond the scope of this class. 😊

Overall, Banyard and colleagues (2007) found that participants in their one-and-three session groups reported significantly more knowledge of sexual violence, bystander attitudes, bystander behaviors, and bystander efficacy after attending the intervention. Additionally, they reported less rape myth acceptance. If you would like to learn more about bystander intervention programs, visit the National Sexual Violence Resources Center for more information.

References

Association of American Universities. (2019). AAU releases 2019 survey on sexual assault and misconduct.

Retrieved from <https://www.aau.edu/newsroom/press-releases/aau-releases-2019-survey-sexual-assault-and-misconduct>

Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology*, 35(4), 463–481. <https://doi.org/10.1002/jcop.20159>

Authors

This guide was written and created by Kendra E. Scott, PhD, Hannah J. Osborn, PhD, and Erin N. Prince.

Acknowledgements

We would like to thank Victoria L. Banyard, Elizabeth G. Plante, and Mary M. Moynihan for making their data available through the Inter-university Consortium for Political and Social Research (ICPSR). Additionally, we appreciate Dr. Kristen Black's edits and suggestions during the creation of this guide.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting a Paired Samples t test in JASP Copyright © 2021 by Kendra E. Scott; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

3.

RUNNING AND INTERPRETING A ONE-WAY ANOVA IN JSAP

Ashlyn A. Moraine; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker

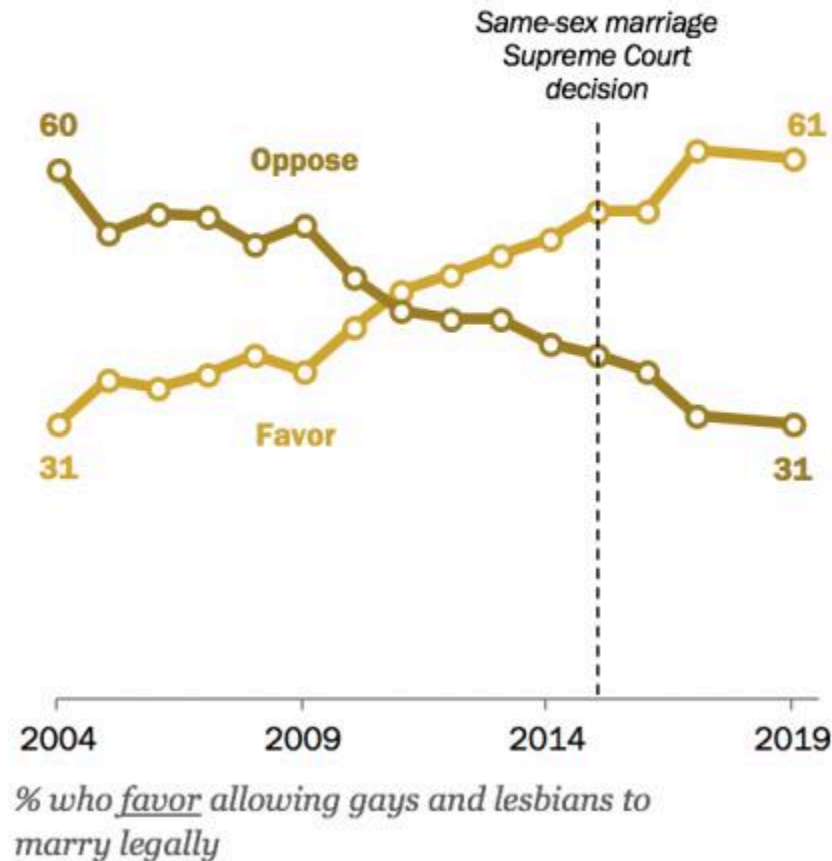
Download Data Sets:

- [JASP Data File One-Way ANOVA](#)
- [CSV Data File One-Way ANOVA](#)

Minority stress theory describes how experiences with stigma, prejudice, and discrimination connect to the physical and mental health of sexual minority people (Meyer, 2003). As societal attitudes toward sexual minorities (e.g., gay, lesbian, and bisexual people) continues to improve, this theory would suggest that physical and mental health outcomes for this population should also improve. As you can see from the following Pew Research Center (2019) graph, attitudes in favor of same-sex marriage, for example, have increased significantly since 2004. Additionally, the US Supreme Court banned employment discrimination based on gender identity or sexual orientation in 2020. Using minority stress theory, these improvements in social context (e.g., attitudes, legislation) would suggest that sexual minorities have improved physical and mental outcomes. Meyer and colleagues (2021) conducted a study to see if there is support for this. Specifically, they wanted to know if minority stress is reduced in younger generations of sexual minorities, compared to older generations of sexual minorities. They did this by comparing the outcomes of three different cohorts of sexual minorities: Equality Cohort, aged 18-25 years ($n = 670$), Visibility Cohort, aged 34-41 years ($n = 372$), and Pride Cohort, aged 52-59 years ($n = 476$).

Public remains supportive of same-sex marriage; wide partisan gap persists

% who ____ allowing gays and lesbians to marry legally



Meyer and colleagues (2021) examined numerous variables, both categorical and continuous. For the purposes of this guide, we will be using these two continuous variables as our dependent variables: *Connection with the LGBT community* and *life satisfaction*. We will need to run two One-way ANOVA's to see if either dependent variable has statistically significant differences between group means for our independent variable or factor (*cohort*).

Connection with the LGBT community

This variable measures the desire and/or strength of a participant's connection to the LGBT community. Scores can range from 1-4 and are based on a 4-point Likert scale ("agree strongly" to "disagree strongly"). Lower scores represent lower connection with the LGBT community.

Life satisfaction

This variable measures the degree to which participants are satisfied with their life. This variable is measured by asking participants to rate 5-items on a 7-point Likert scale ("disagree strongly" to "agree strongly"). Lower scores represent lower life satisfaction.

Hypotheses

We have two different dependent variables we are looking at in this example: (1) Connection with the LGBT Community and (2) Life satisfaction. This means that we will have two different sets of hypotheses.

The null hypotheses are:

- Connection with the LGBT Community
 - Conceptual H_0 : There are no significant group mean differences on connection with the LGBT community between the younger, middle-aged, and older adult groups.
 - Mathematical H_0 : There are no significant differences between the population means for the younger, middle-aged, and older adult groups on connection with the LGBT community; $M_{\text{younger}} = M_{\text{middle}} = M_{\text{older}}$
- Life satisfaction
 - Conceptual H_0 : There are no significant group mean differences on life satisfaction between the younger, middle-aged, and older adult groups.
 - Mathematical H_0 : There are no significant differences between the population means for the younger, middle-aged, and older adult groups on life satisfaction; $M_{\text{younger}} = M_{\text{middle}} = M_{\text{older}}$.

The alternative hypotheses are:

- Connection with the LGBT Community
 - Conceptual H_2 : There is at least one significant group mean difference on connection with the LGBT community between the younger, middle-aged, and older adult groups.
 - Mathematical H_2 : There is at least one significant difference between the population means for the younger, middle-aged, and older adult groups on connection with the LGBT community; $M_{\text{younger}} \neq M_{\text{middle}} \neq M_{\text{older}}$.
- Life satisfaction
 - Conceptual H_1 : There is at least one significant group mean difference on life satisfaction between the younger, middle-aged, and older adult groups.
 - Mathematical H_1 : There is at least one significant difference between the population means for the younger, middle-aged, and older adult groups on life satisfaction; $M_{\text{younger}} \neq M_{\text{middle}} \neq M_{\text{older}}$.

JASP Analyses

In order to run analyses, the first thing we need to do is open the data set we will be working with. To do this, open JASP and follow the steps below:

File → Open → Computer → Browse → Select the “One-Way ANOVA Practice Data Set (Meyer et al. 2021)” JASP file wherever it is saved on your computer.

Once the data set is open in JASP, we will change the data labels for our condition variable so that we can

see which groups are being compared when we run the omnibus and post hoc tests. Currently the condition column has either a 1, 2, or 3 for each participant. Based on Meyer et al.'s (2021) codebook, we will need to change these labels so that 1 = 'young cohort', 2 = 'middle cohort', 3 = 'older cohort.' To change the numerical data into our categorical labels, you will take your cursor and hover over the 'Condition' column. When you see a pop up saying, "click here to change labels" click on it. To change the labels, click on the 1 under the Label column and type 'younger.' Then click on the 2 under the Label column and type 'middle, and so-on until all conditions are appropriately labeled. After you have changed the labels, you can close the window by clicking on the X button.

One-Way ANOVA Practice Data Set (Meyer et al. 2021) (C:\Users\ptd539\Documents\One-Way ANOVA Practice Data Set (Meyer et al. 2021).csv)

Click here to change labels T-Tests ANOVA Mixed Models

	Cohort	Day_Discrimination	Felt_Stigma
1	3	1.66667	2.33333
2	2	2.11111	2.66667

Filter	Value	Label
✓	1	Young Cohort
✓	2	Middle Cohort
✓	3	Older Cohort

Assumption Testing

There are six statistical assumptions that data must meet in order to run a one-way ANOVA. Many of these assumptions will look familiar, as they are quite similar to those required to run an independent measures t test – we're just adding more groups! Assumptions 1 through 3 have to do more with the research design, and assumptions 4 through and 6 refer to those which data must meet in order to run the statistical test. Let's consider whether our data meet these assumptions:

Assumption 1: Categorical Independent Variable.

Yes. We have one independent variable (experimental condition), and participants have been assigned to one

of three levels of that independent variable based on their age. Therefore, participants are in one ‘category’ of the independent variable, so we meet this assumption.

Assumption 2: Continuous Dependent Variable.

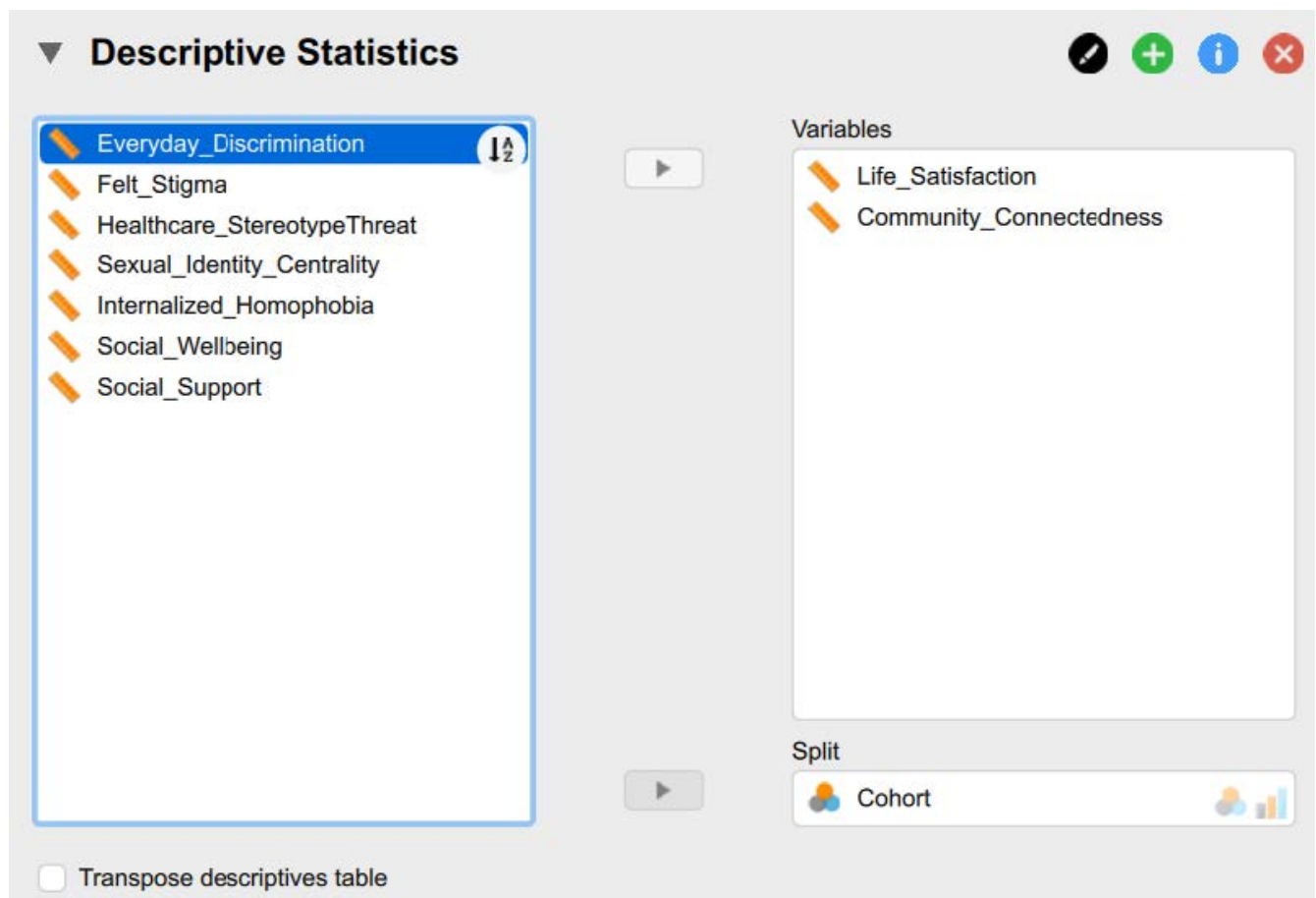
Yes. We have dependent variable(s) (‘life satisfaction’ and ‘community connectedness’) and these dependent variables are single items based on self-reports measured on Likert-type scales. In psychological research, Likert-type responses are treated as continuous data. Therefore, we meet this assumption.

Assumption 3: Independence of Observations.

Yes. Our participants were assigned to one of the three possible levels of the independent variable, and because no participant can be in more than one age group, we can assume that there is no relationship between the observations in each group. This assumption has been met.

Assumption 4: Normal Distribution.

To check our data for Assumptions 4 and 5, we will need to utilize the Descriptives tab. When the “Descriptive Statistics” window opens, move the dependent variables (‘life satisfaction’ and ‘community connectedness’) to the “Variables” box. Then, move our factor or independent variable (‘Cohort’) to the “Split” box so that we get plots for each group.



Under the “Plots” tab, select:

1. “Customizable Boxplots”: “Boxplots,” “Boxplot element” and “Label outliers”

▼ Customizable plots

Color palette Colorblind ▼

☒ Boxplots

☒ Boxplot element ☐ Use color palette

☐ Violin element ☒ Label outliers

☐ Jitter element

Under the “Statistics” tab, select:

1. Under “Central Tendency,” select “Mean”
2. Under “Distribution,” select “Skewness” and “Kurtosis”
3. Under “Dispersion,” select “Minimum,” “Maximum,” and “Std. deviation”

▼ Statistics

Sample size

☒ Valid

☒ Missing

Central tendency

☐ Mode

☐ Median

☒ Mean

Dispersion

☐ S.E. mean

☐ Coefficient of variation

☐ MAD robust

☐ Variance

☒ Minimum

☒ Std. deviation

☐ MAD

☐ IQR

☐ Range

☒ Maximum

Quantiles

☐ Quartiles

☐ Cut points for: equal groups

☐ Percentiles:

Distribution

☒ Skewness

☒ Kurtosis

☐ Shapiro-Wilk test

☐ Sum

Although there are multiple ways to determine if the dependent variable is normally distributed for each group of our independent variable, for this course we will focus on interpreting the skewness and kurtosis statistics. We want skewness and kurtosis values that are between -2 and +2. Looking at the values on the following output copied, we can see our skewness and kurtosis values for the dependent variable for all of our groups are within the acceptable range of -2 and +2 across both of our dependent variables.

Descriptive Statistics ▼

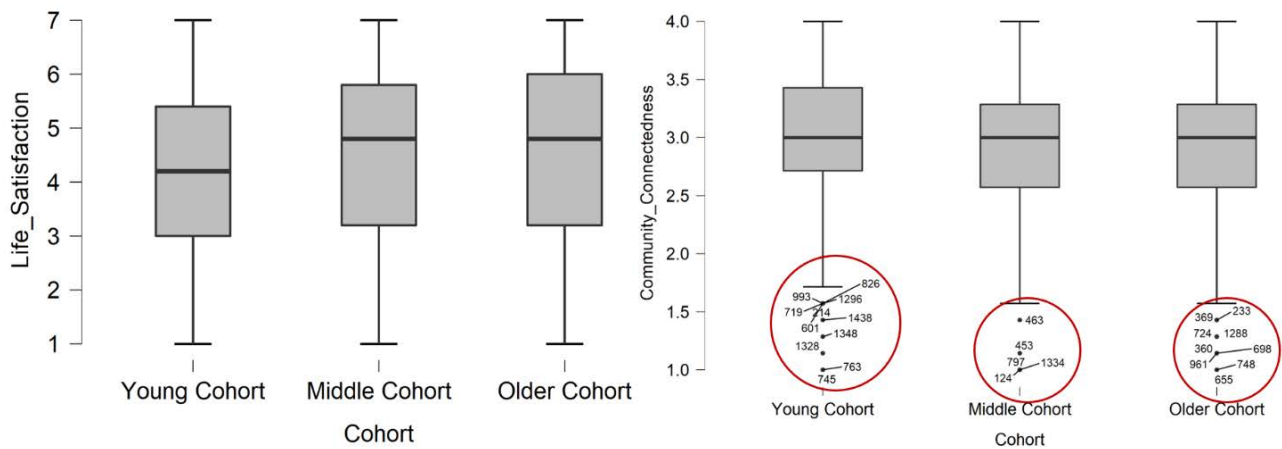
	Life_Satisfaction			Community_Connectedness		
	Younger Cohort	Middle Cohort	Older Cohort	Younger Cohort	Middle Cohort	Older Cohort
Valid	670	372	476	670	372	476
Missing	0	0	0	0	0	0
Mean	4.141	4.364	4.531	3.026	2.916	2.902
Std. Deviation	1.516	1.692	1.710	0.556	0.558	0.569
Skewness	-0.110	-0.396	-0.405	-0.333	-0.392	-0.367
Std. Error of Skewness	0.094	0.126	0.112	0.094	0.126	0.112
Kurtosis	-0.806	-0.951	-0.955	0.339	0.633	0.602
Std. Error of Kurtosis	0.189	0.252	0.223	0.189	0.252	0.223
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	7.000	7.000	7.000	4.000	4.000	4.000

To report these results in APA format, we can write:

The scores on both life satisfaction and community connectedness were normally distributed across all conditions, as skewness and kurtosis statistics fell within the acceptable range of -2 and +2.

Assumption 5: No Outliers.

To determine if we have any outliers, we will look at the boxplots in our JASP output. If we have any outliers, they would be outside the top and bottom lines or whiskers. As you can see from the boxplots, we do have outliers in our data, and we have violated this assumption. Which means we have to decide whether we want to keep the outliers, transform the outliers, or remove them. Given that outliers are especially common in datasets with large samples (as ours is; $N = 1,458$), and that we have so many, we will keep these outliers in our dataset in order to avoid removing a significant amount of datapoints from our model (something that is generally discouraged, Faraway, 2015). Instead, we will model how to run the analyses as normal with outliers present, and at the end of this guide, will run the analyses with those outliers removed to examine whether the results remained the same.



To report this using APA format, we would write:

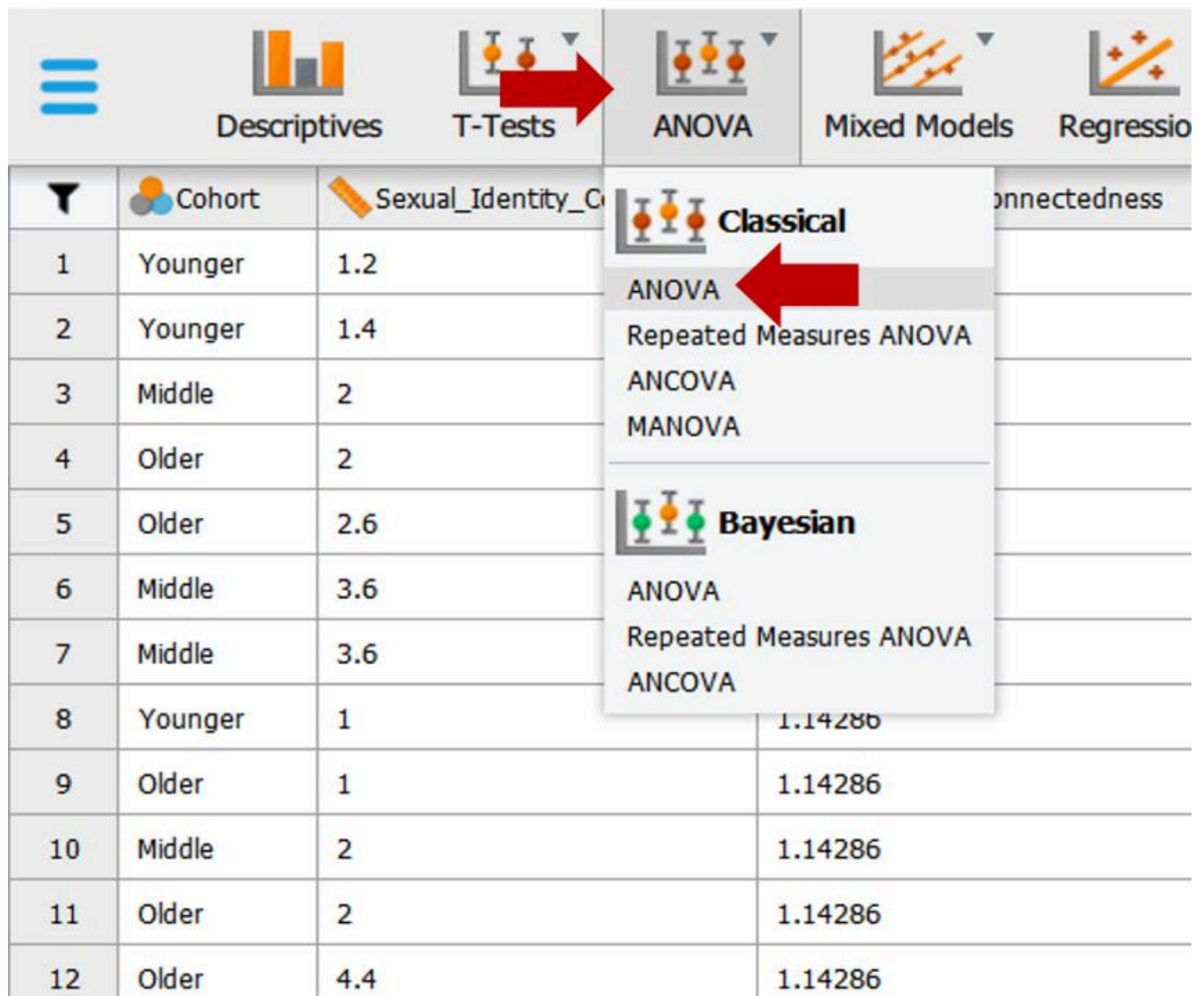
There were no outliers identified on the life satisfaction scores; however, there were several outliers across all three cohorts for the community connectedness variable, as assessed by the inspection of a boxplot. We decided to retain these outliers given the large sample size to avoid removing a large number of datapoints from the model.

Assumption 6: Homogeneity of Variance.

Checking for the homogeneity of variances assumption is part of the overall one-way ANOVA test. Let's move onto our primary analyses, and complete checking this assumption in that section.

Primary Analyses: Omnibus Test for Community Connectedness.

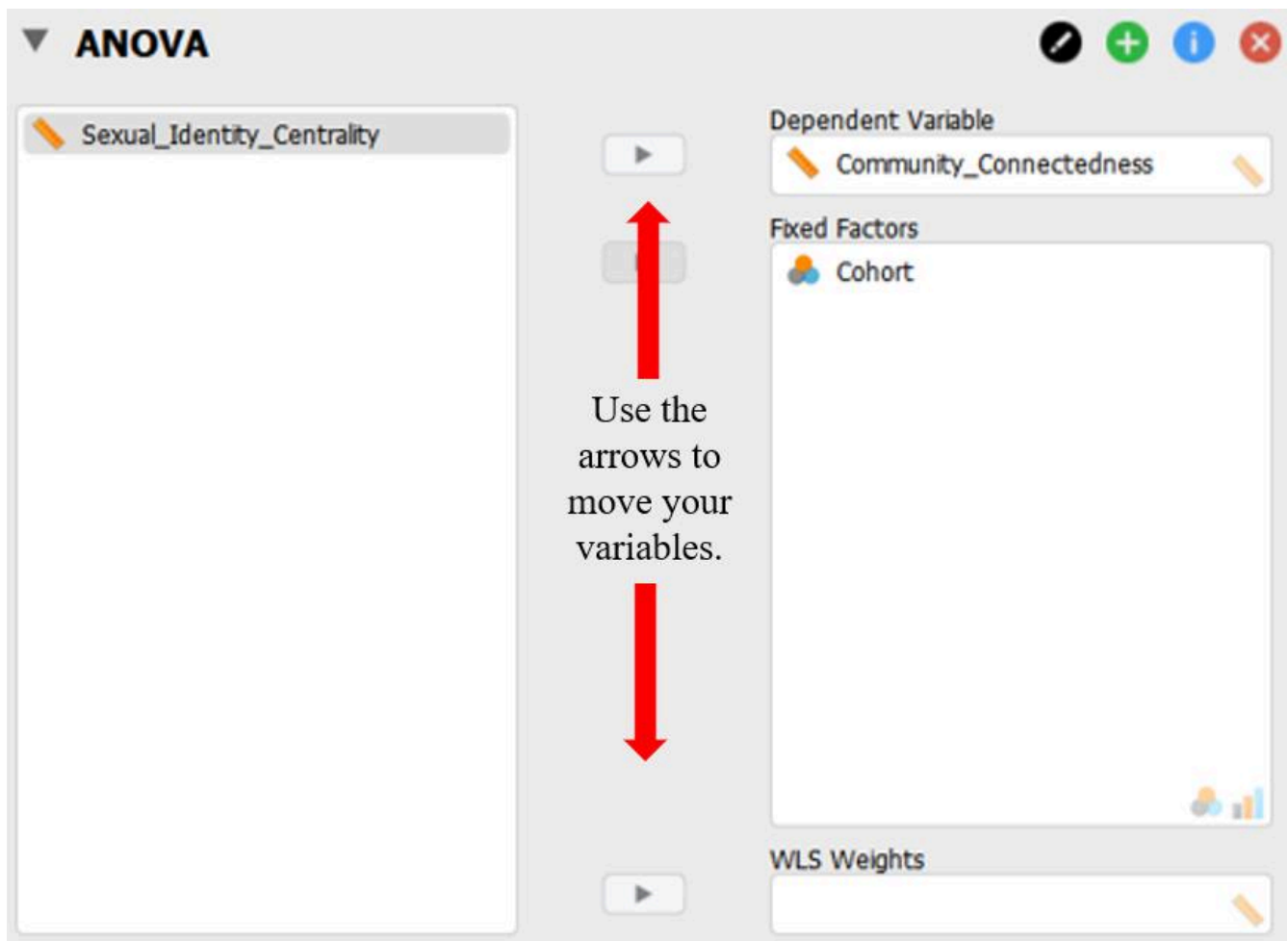
To run a one-way between subjects ANOVA, we will begin by testing the omnibus effect, and then follow up with a section on Post hoc analyses. To run the one-way ANOVA, go to ANOVA at the top of the JASP screen and select 'ANOVA'. See the image below:



The screenshot shows a statistical software interface with a menu bar at the top containing icons for Descriptives, T-Tests, ANOVA, Mixed Models, and Regression. A red arrow points from the T-Tests icon to the ANOVA icon. A dropdown menu is open under the ANOVA icon, showing two main sections: 'Classical' and 'Bayesian'. Under 'Classical', the options are ANOVA, Repeated Measures ANOVA, ANCOVA, and MANOVA. A red arrow points to the 'ANOVA' option. Under 'Bayesian', the options are ANOVA, Repeated Measures ANOVA, and ANCOVA. The background shows a data table with columns for Cohort, Sexual_Identity_C, and connectedness.

	Cohort	Sexual_Identity_C	connectedness
1	Younger	1.2	
2	Younger	1.4	
3	Middle	2	
4	Older	2	
5	Older	2.6	
6	Middle	3.6	
7	Middle	3.6	
8	Younger	1	1.14286
9	Older	1	1.14286
10	Middle	2	1.14286
11	Older	2	1.14286
12	Older	4.4	1.14286

First, we need to move our dependent variable [community connectedness] to the “Variables” box and our independent variable [cohort] to the “Fixed Factors” box.



- We are then going to ask JASP to run the various analyses we will need to interpret for our test by clicking on the appropriate boxes in the window.
 - Under “Display” we will want to check ‘Descriptive statistics’ and ‘Estimates of effect size’ – ‘eta-squared’ and ‘omega-squared’
 - Under “Assumption Checks” select ‘Homogeneity tests’; under “Assumption corrections” select ‘None’ and ‘Welch’

Display

☒ Descriptive statistics

☒ Estimates of effect size

☒ η^2 ☐ partial η^2 ☒ ω^2

☐ Vovk-Selke maximum p-ratio

► Model

▼ Assumption Checks

☒ Homogeneity tests

Homogeneity corrections

☒ None ☐ Brown-Forsythe ☒ Welch

☐ Q-Q plot of residuals

Now that we have asked JASP to run the appropriate analyses for the omnibus test, let's first look to see if we've met the assumption for homogeneity of variances. In your output, this is located under the "Assumption Checks" table and includes Levene's test for equality of variances.

- If p is less than .05, you do not meet the homogeneity of variances assumption.
- If p is greater than .05, you do meet the homogeneity of variances assumption.

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
0.064	2.000	1455.000	0.938

To interpret the results of the Levene's test, we need to look at the significance or probability value. If p is less than .05, the test is significant, and we do not have homogeneity of variances. If p is greater than .05, the test is non-significant, and we do have homogeneity of variances. Because our p value is .94, which is greater than .05, we meet this assumption.

- If homogeneity of variances was met, you can proceed with interpreting the omnibus test without any

homogeneity corrections.

To report the results of the Levene's test in APA format, it might look something like this:

There was homogeneity of variances between groups, as assessed by the Levene's test for equality of variances ($p > .05$).

Interpreting the Statistical Significance of the One-Way ANOVA Omnibus Test

The first thing we will interpret is the statistical significance of the omnibus test. To do this, we will look at the ANOVA table in the output. Because we have homogeneity of variances, we can interpret this without any homogeneity corrections as shown in the first row of this table.

These are the degrees of freedom for this analysis. The top row is df_{between} (df1) and the bottom is df_{error} (df2)

This is the F statistic for this analysis.

- If p is less than .05, you will reject the null hypothesis.
- If p is greater than .05, you will fail to reject the null hypothesis.

ANOVA - Community_Connectedness ▼

Homogeneity Correction	Cases	Sum of Squares	df	Mean Square	F	p	η^2	ω^2
None	Cohort	4.816	2.000	2.408	7.585	< .001	0.010	0.009
	Residuals	461.906	1455.000	0.317				
Welch	Cohort	4.816	2.000	2.408	7.568	< .001	0.010	0.009
	Residuals	461.906	857.414	0.539				

Note. Type III Sum of Squares

If we have homogeneity of variances, we interpret the results in the first row of the table. If we do not have homogeneity of variances, we would interpret the results of the Welch test instead (in the second row).

These are our effect size measures. You will just report one. For both, .01 is a small effect, .06 is a medium effect, and .15 is a large effect.

To interpret the results, we want to look at the p value. If p is less our alpha level of .05, we will reject the null hypothesis (indicating there is a significant statistical difference between the means of the groups on our dependent variable). If the p value is greater than our alpha level of .05, we will fail to reject the null hypothesis (indicating there is not a significant statistical difference between the group means on our dependent variable).

In this example, you can see that our p value is <.001, which is less than the alpha of .05. Therefore, we will reject our null hypothesis – there is at least one significant group mean difference on connection with the LGBT community between the younger, middle, and older cohorts.

Writing the basic results of the omnibus ANOVA test (F test) in APA format follows this general format:

$$F(df1, df2) = F \text{ statistic}, p < .05 \text{ or } p > .05, \eta^2 = \eta^2 \text{ value OR } \omega^2 = \omega^2 \text{ value}$$

So, plugging in our results into this format should look like this:

$$F(2, 1455) = 7.59, p < .05, \eta^2 = .01 \text{ OR } \omega^2 = .01$$

Now that we know we have a significant difference between the mean levels of community connectedness

between our three conditions, what does this mean? Which of the conditions, specifically, are different from one another in regard to community connectedness? To answer these questions, we will need to look at the Post Hoc Analyses section.

Post Hoc Analyses: Community Connectedness

Now that we have rejected our null hypothesis and determined there is at least one age group that is significantly different in their level of connection to the LGBT+ community, let's conduct post hoc tests to determine what groups are different. Post hoc analyses allow us to understand which of our groups/levels of the independent variable are different from one another. To do this, go back to the ANOVA test selection for the community connectedness outcome variable. Under "Post Hoc Tests" we will move our grouping variable [Cohort] over using the arrow and select:

- Under Type: select 'Standard' and 'Effect size' (Note: If we had violated the homogeneity of variances assumption, we would need to select the "Games-Howell" option here.).
- Under Correction: select 'Tukey'
- Under Display: select 'Flag significant comparisons'

▼ Post Hoc Tests

Use the arrow to move your grouping variable (IV) over.

Correction

☒ Tukey
☐ Scheffe
☐ Bonferroni
☐ Holm
☐ Šidák

Type

☒ Standard
☐ From 1000 bootstraps
☒ Effect size
☐ Games-Howell
☐ Dunnett
☐ Dunn

Display

☐ Confidence intervals 95.0 %
☒ Flag Significant Comparisons

If you violated the homogeneity of variances assumption, select "Games-Howell" as well.

Which of our groups are significantly difference from one another? To examine this question, let's look at both the "Descriptives" table and the "Post Hoc Tests" table from our output. The Descriptives table includes means, standard deviations, and sample size for each of our conditions. The Post Hoc Tests table details the results of all possible comparisons between our three groups.

Descriptives

Descriptives - Community_Connectedness

Cohort	Mean	SD	N
Middle	2.923	0.555	363
Older	2.908	0.574	448
Younger	3.030	0.560	647

Post Hoc Tests

Standard

This is the difference between the means for each comparison.

This is the t test statistic for the comparison between the two groups.

Post Hoc Comparisons - Cohort

		Mean Difference	SE	t	Cohen's d	p_{Tukey}
Young, Cohort	Middle, Cohort	0.110	0.036	3.033	0.198	0.007**
	Older, Cohort	0.123	0.034	3.672	0.220	< .001***
Middle, Cohort	Older, Cohort	0.013	0.039	0.347	0.024	0.936

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. Cohen's d does not correct for multiple comparisons.

Note. P -value adjusted for comparing a family of 3

This is Cohen's d for the effect size of the group differences. Small is .20, Medium is .50, and Large is .80.

- If p is less than .05, there is a significant difference between the two groups.
- If p is greater than .05, there is not a significant difference between the two groups.

The Tukey post hoc test is a good one to select if you haven't violated the homogeneity of variances assumption and you want to compare all possible combinations of group differences. The Post Hoc Tests table tells us not only whether the mean differences between the groups are statistically significant, they also tell us the effect size of this group difference by providing Cohen's d . For each row in the table above, there is a test of the comparison between each group and the reference group (located on the far left). For instance, let's look at the first two rows, each of which compare a given condition to the 'control' condition. Row one is the comparison between the means of the 'younger' group and the 'middle group'; row two is the comparison between the means of the 'younger' group and the 'older' group; row three is the comparison between the mean of the 'middle' group and the 'older' group.

The information in the table above has the following meaning:

- Mean Difference – the mean difference between the reference group ('younger') and the target group ('middle'), mathematically, it is $M_{\text{younger}} - M_{\text{middle}}$.
- Standard Error (SE) – the standard error of the mean difference between the reference group ('younger') and the target group ('middle')
- t – the test statistic from the independent measures t test comparing the reference group ('younger') and the target group ('middle')
- Cohen's d – the Cohen's d effect size for the comparison between the reference group ('younger') and the target group ('middle'). This is interpreted the same as we've seen before.
- p_{Tukey} – The statistical significance level of the mean difference between the reference group ('younger') and the target group ('middle') Notice that the p value says Tukey next to it, indicating that these p values are corrected or adjusted for the number of comparisons that we could possibly run. Therefore, this p value accounts for the fact that we had three separate two-group comparisons.

From here, we would interpret the statistical and practical significance of each group pair comparisons. For each group comparison that was statistically significant, we would provide information using the following format, ending with “no other groups comparisons were significantly different, all $p_{\text{Tukey}} > .05$ ”:

Group 1 was [higher/lower] in DV ($M = ?$, $SD = ?$) compared to Group 2 ($M = ?$, $SD = ?$), $p_{\text{Tukey}} < .05$, $d = \text{Cohen's } d \text{ value}$, indicating a [small/medium/large group] difference. No other group comparisons were significantly different, all $p_{\text{Tukey}} > .05$.

If there were significant group differences between *all* of the groups, we could write something using the following format:

All groups were significantly different from one another ($p_{\text{Tukey}} < .05$). Group 1 reported the highest DV ($M = ?$, $SD = ?$), which was significantly higher than Group 2 ($M = ?$, $SD = ?$), $p < .05$, $d = \text{Cohen's } d \text{ value}$, and significantly higher than group 3 ($M = ?$, $SD = ?$), $p < .05$, $d = ?$. The difference between group 2 and group 3 was also significant ($p < .05$, $d = ?$).

If we were reporting in APA format using this formula, we would write the following:

The younger cohort was significantly higher in community connectedness ($M = 3.03$, $SD = .56$) compared to the older cohort ($M = 2.91$, $SD = .57$), $p_{\text{Tukey}} < .05$, $d = .21$, indicating a small-to-moderate group difference. The younger cohort was significantly higher in community connectedness ($M = 3.03$, $SD = .56$) compared to the middle cohort ($M = 2.92$, $SD = .56$), $p_{\text{Tukey}} < .05$, $d = .20$, indicating a small-to-moderate group difference. No other group comparisons were significantly different, $p_{\text{Tukey}} > .05$.

Primary Analyses: Omnibus Test for Life Satisfaction

Now let's repeat the above steps, but for our second dependent variable or factor, 'life satisfaction.' Spoiler alert: We are running another set of analyses so we can practice how to handle violations to the homogeneity of variances assumption.

- First, we need to move our dependent variable [life satisfaction] to the “Variables” box and our independent variable [cohort] to the “Fixed Factors” box. We are then going to ask JASP to run the various analyses we will need to interpret for our test by clicking on the appropriate boxes in the window.
 - Under “Display” we will want to check ‘Descriptive statistics’ and ‘Estimates of effect size’ – ‘eta-squared’ and ‘omega-squared’
 - Under “Assumption Checks” select ‘Homogeneity tests’; under “Assumption corrections” select ‘None’ and ‘Welch’

ANOVA

Dependent Variable: Life_Satisfaction

Fixed Factors: Cohort

WLS Weights:

Display

- ☒ Descriptive statistics
- ☒ Estimates of effect size
 - ☒ η^2 ☐ partial η^2 ☒ ω^2
- ☐ Vovk-Sellke maximum p-ratio

Model

Assumption Checks

- ☒ Homogeneity tests

Homogeneity corrections

 - ☒ None ☐ Brown-Forsythe ☒ Welch
- ☐ Q-Q plot of residuals

Now that we have asked JASP to run the appropriate analyses for the omnibus test, let's first look to see if we've met the assumption for homogeneity of variances. In your output, this is located under the "Assumption Checks" table and includes Levene's test for equality of variances.

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
9.825	2.000	1515.000	< .001

- If p is less than .05, you do not meet the homogeneity of variances assumption.
- If p is greater than .05, you do meet the homogeneity of variances assumption.

To interpret the results of the Levene's test, we need to look at the significance or probability value. If p is less than .05, the test is significant, and we do not have homogeneity of variances. If p is greater than .05, the test is non-significant, and we do have homogeneity of variances. Our p value is $< .001$, which is less than .05. This means we do not meet this assumption.

- If homogeneity of variances was violated, you can continue conducting the one-way ANOVA, but will you need to have selected the "Welch" check box under 'homogeneity corrections' and interpret the "Welch" ANOVA results instead. As with independent measures t tests, the Welch homogeneity correction is calculated without pooling the variances. For this reason, it is often easier to select the "Welch" check box when running the omnibus test just in case the homogeneity assumption is violated. Additionally, if you violate the homogeneity assumption, there is an alternative post hoc selection which will be discussed in more detail in the Post Hoc Analyses section.

To report the results of the Levene's test in APA format, it might look something like this:

There was not homogeneity of variances, as assessed by the Levene's test for equality of variances ($p < .05$).

Interpreting the Statistical Significance of the One-Way ANOVA Omnibus Test

The first thing we will interpret is the statistical significance of the omnibus test. To do this, we will look at the ANOVA table in the output. Because we do not have homogeneity of variances, we interpret this with homogeneity corrections as shown in the second row in the table labeled 'Welch'.

ANOVA - Life_Satisfaction

These are the degrees of freedom for this analysis. The top row is df_{between} (df1) and the bottom is df_{error} (df2)

This is the F statistic for this analysis.

Homogeneity Correction	Cases	Sum of Squares	df	Mean Square	F	p	η^2	ω^2
None	Cohort	43.331	2.000	21.665	8.230	< .001	0.011	0.009
	Residuals	3988.169	1515.000	2.632				
Welch	Cohort	43.331	2.000	21.665	8.282	< .001	0.011	0.009
	Residuals	3988.169	859.934	4.638				

Note. Type III Sum of Squares

If we have homogeneity of variances, we interpret the results in the first row of the table. If we do not have homogeneity of variances, we would interpret the results of the Welch test instead (in the second row).

- If p is less than .05, you will reject the null hypothesis.
- If p is greater than .05, you will fail to reject the null hypothesis.

These are our effect size measures. You will just report one. For both, .01 is a small effect, .06 is a medium effect, and .15 is a large effect.

To interpret the results, we want to look at the p value. In this example, you can see that our p value is $<.001$, which is less than the alpha of .05. Therefore, we will reject our null hypothesis – there are significant group mean differences on life satisfaction between the younger, middle, and older cohorts.

Writing the basic results of the omnibus ANOVA test (F test) in APA format follows this general format:

$F(df1, df2) = F \text{ statistic}, p < .05 \text{ or } p > .05, \eta^2 = \eta^2 \text{ value OR } \omega^2 = \omega^2 \text{ value}$

So, plugging in our results into this format should look like this:

$F(2, 859.93) = 8.28, p < .001, \eta^2 = .01 \text{ or } \omega^2 = .01$

Now that we know we have a significant difference between the mean levels of life satisfaction between our three age groups, what does this mean? Which of the age groups, specifically, are different from one another regarding life satisfaction? To answer these questions, we will need to look at the Post Hoc Analyses section.

Post Hoc Analyses: Life Satisfaction

Now that we have rejected our null hypothesis and determined there is at least one age group that is significantly different in their level of life satisfaction, let's conduct post hoc tests to determine what groups are different. To do this, go back to the ANOVA test selection for the life satisfaction outcome variable. Under "Post Hoc Tests" we will move our grouping variable [Cohort] over using the arrow and select:

- Under Type: select 'Standard' and 'Effect size' (Note: If we had violated the homogeneity of variances assumption, we would need to select the "Games-Howell" option here.).
- Under Correction: select 'Tukey'
- Under Display: select 'Flag significant comparisons'

▼ Post Hoc Tests

Use the arrow to move your grouping variable (IV) over.

Because we violated the homogeneity of variances assumption, select “Games-Howell.”

Type

- ☒ Standard
 - ☐ From bootstraps
- ☒ Effect size
- ☒ Games-Howell
- ☐ Dunnett
- ☐ Dunn

Display

- ☐ Confidence intervals %
- ☒ Flag Significant Comparisons

Correction

- ☒ Tukey
- ☐ Scheffe
- ☐ Bonferroni
- ☐ Holm
- ☐ Šidák

Cohort

Which of our groups are significantly difference from one another? To examine this question, let’s look at both the “Descriptives” table and the “Post Hoc Tests” table from our output.

Descriptives

Descriptives - Life_Satisfaction

Cohort	Mean	SD	N
Middle Cohort	4.364	1.692	372
Older Cohort	4.531	1.710	476
Young Cohort	4.141	1.516	670

Looking at the preceding Descriptives table image, we can see the older cohort has the highest mean level of life satisfaction ($M = 4.53$), the middle cohort has the second highest mean level of life satisfaction ($M = 4.36$), and the young cohort has the lowest mean level of life satisfaction ($M = 4.14$). We still don’t know which

groups are statistically different, but we can see a trend of the mean levels of life satisfaction increasing for each age group.

Games-Howell

Games-Howell Post Hoc Comparisons - Cohort

Comparison	Mean Difference	SE	t	df	p _{Tukey}
Young Cohort - Middle Cohort	-0.223	0.105	-2.115	698.190	0.088
Young Cohort - Older Cohort	-0.390	0.098	-3.984	944.321	< .001***
Middle Cohort - Older Cohort	-0.167	0.118	-1.417	800.942	0.333

*** p < .001

This is the *t* test statistic for the comparison between the two groups.

This is the difference between the means for each comparison.

- If *p* is less than .05, there is a significant difference between the two groups.
- If *p* is greater than .05, there is not a significant difference between the two groups.

The Games-Howell post hoc test is preferred when your data violate the homogeneity of variances assumption. As you can see from the table above, it is set up a little differently than the Tukey post hoc test results. The Games-Howell Post Hoc Comparisons table still tells us if the mean differences between the groups are statistically significant; however, Cohen's *d* effect size statistics are not available for this post hoc option. [Note: If you need to report an effect size for your post hoc comparisons, you can check the "Confidence Intervals" box under "Post Hoc Tests" and report the 95% confidence intervals for each comparison.] Unlike the Tukey post hoc test table, the Games-Howell table tells us what groups are being compared in each row rather than using a reference group. The top row is comparing the young cohort and the middle cohort, the second row is comparing the young cohort and middle cohort, and the third row is comparing the middle cohort and the older cohort.

From here, we would interpret the statistical significance of each group pair comparisons. If we were reporting in APA format using this formula, we would write the following:

The older cohort ($M = 4.53$, $SD = 1.71$) had significantly higher levels of life satisfaction compared to the young cohort ($M = 4.14$, $SD = 1.52$), $p_{\text{Tukey}} < .001$. No other group comparisons were significantly different, $p_{\text{Tukey}} > .05$.

Interpreting the Practical Significance of the One-Way ANOVA Omnibus Test

Now that we have interpreted the statistical significance of our ANOVA Omnibus tests, let's consider the practical significance. Remember, a significant *p* value tells us that there is a significant difference in the mean levels of community connectedness and life satisfaction between our three age groups. The effect size tells us how meaningful the difference between the age cohorts is for our dependent variable.

η^2 and ω^2 Effect Size Value	Strength
.01	Small
.06	Medium
.15	Large

Eta-squared (η^2) or omega-squared (ω^2) are both a representation of effect sizes for the omnibus test for an ANOVA. Because we are not yet able to say that, for example, Group 1 is different than Group 2 for the omnibus ANOVA test, you have to estimate how much the difference is across *all* the groups. Therefore, eta- and omega-squared are used to determine the amount of variance accounted for (out of 100% variance) that the IV explains in the DV. These values are listed as proportions, so they never go over the value of one. To get this interpretation, we would multiply the value of eta-squared or omega-squared by 100 and interpret it as a percentage. In this example, therefore, we could say that 1.3% (eta-squared) or 1.1% (omega-squared) of the variance in community connectedness is accounted for by the experimental condition. Eta-squared is the more commonly reported effect size, while omega-squared is often used to help represent the true population effect size (Maxwell et al., 1981).

Let's interpret eta-squared for our variables. We obtained eta-squared values of .01 – this is a small effect. So, although we have a statistical difference in the mean levels of community connectedness and life satisfaction between cohorts, that difference is *small* (negligible, even). This can often happen in studies that have a very large sample size. Finding a significant difference becomes easier, even if it is really small.

Value (Notation)	
Omnibus Test	Test Statistic (F)
Degrees of Freedom ($df1, df2$)	Probability (p)
Assumptions	Skewness and Kurtosis
Boxplot	Levene's Test
Post Hoc Analyses	Mean (M)
Standard Deviation (SD)	Probability of comparison (p_{Tukey})
Effect Size (Cohen's d)	Effect Size (Eta- or omega-squared)

Putting it All Together

We ran the test for both community connectedness and life satisfaction, but for simplicity, let's just look at the full write-up for the community connectedness analysis.

A one-way ANOVA was conducted to examine whether there were any mean differences in community connectedness between cohorts. There were outliers identified on community

connectedness scores across conditions, as assessed by the inspection of a boxplot. We decided to retain these outliers given the large sample size and so as to avoid removing a large number of datapoints from the model. The scores on community connectedness were normally distributed across all conditions, as skewness and kurtosis statistics fell within the acceptable range of -2 and +2. There was homogeneity of variances, as assessed by the Levene's test for equality of variances ($p > .05$).

The omnibus test was statistically significant, $F(2, 1455) = 7.59, p < .05, \eta^2 = .01$ OR $\omega^2 = .01$. Post hoc analyses revealed that the younger cohort was significantly higher in community connectedness ($M = 3.03, SD = .56$) compared to the older cohort ($M = 2.91, SD = .57$), $p_{\text{TUkey}} < .05, d = .21$, indicating a small-to-moderate group difference. The younger cohort was also significantly higher in community connectedness ($M = 3.03, SD = .56$) compared to the middle cohort ($M = 2.92, SD = .56$), $p_{\text{TUkey}} < .05, d = .20$, indicating a small-to-moderate group difference. No other group comparisons were significantly different, $p_{\text{TUkey}} > .05$.

Real World Meaning

Now that the statistical jargon is out of the way – what do our results mean? We rejected the null hypothesis for community connectedness and found differences between the younger cohort and the middle/older cohorts. This means the younger cohort feels a stronger connection to the LGBT community compared to the middle and older cohorts; however, given the small effect size we obtained, this difference is may not be practically meaningful. We also rejected the null hypothesis for life satisfaction and found differences between the younger cohort and the older cohort. This means the older members of the LGBT community included in this study felt more satisfied with their life than the younger cohort.

Thinking back to Meyer and colleagues (2021) original hypothesis using minority stress theory, they measured other variables to address their original research question. Specifically, they thought the theory would mean that societal improvements in attitudes toward sexual minorities as well as increased legislative protections, would result in improved mental health outcomes for younger cohorts of sexual minorities in the United States. Certainly, younger cohorts' feeling a greater connection to the LGBT community might suggest that is true as well. Unfortunately, the data did not support their hypothesis. If the data supported the minority stress theory, we would have found higher levels of life satisfaction in the younger cohort; however, we found the highest levels of life satisfaction in the older cohort. Looking across all the outcomes Meyer and colleagues analyzed, they found that younger cohorts reported more experiences with physical violence, sexual violence, and verbal abuse. They also reported higher levels of felt stigma and psychological distress.

References

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin*, 129(5), 674–697. <https://doi.org/10.1037/0033-2909.129.5.674>

Meyer, I. H., Russell, S. T., Hammack, P. L., Frost, D. M., & Wilson, B. D. M. (2021). Minority stress,

distress, and suicide attempts in three cohorts of sexual minority adults: A U.S. probability sample. *PloS One*, 16(3), e0246827–e0246827. <https://doi.org/10.1371/journal.pone.0246827>

Pew Research Center (2019). Majority of public favors same-sex marriage, but division persist. Retrieved from <https://www.pewresearch.org/politics/2019/05/14/majority-of-public-favors-same-sex-marriage-but-divisions-persist/>

The Office of Equity, Diversity, and Inclusion (2020). The Supreme Court bans employment discrimination towards sexual and gender minorities. Retrieved from <https://www.edi.nih.gov/blog/news/supreme-court-bans-employment-discrimination-towards-sexual-and-gender-minorities>

Authors

This guide was written and created by Ashlyn A. Moraine, PhD, Hannah J. Osborn, PhD, and Erin N. Prince.

Acknowledgements

We would like to thank Ilan H. Meyer, Stephen T. Russell, Phillip L Hammack, David M. Frost, and Bianca D. M. Wilson for making their data available through the Inter-university Consortium for Political and Social Research (ICPSR). Additionally, we appreciate Dr. Kristen Black's edits and suggestions during the creation of this guide.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting a One-Way ANOVA in JSAP Copyright © 2021 by Ashlyn A. Moraine; Dr. Hannah J. Osborn; Erin N. Prince; and Ruth Walker is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

4.

RUNNING AND INTERPRETING A TWO-WAY ANOVA IN JASP

Ashlyn A. Moraine; Erin N. Prince; and Ruth Walker

Download Data Sets:

- [JASP Data File Two-Way ANOVA](#)
- [CSV Data File Two-Way ANOVA](#)

The Pew Research Center collected and published survey data on attitudes toward women in leadership as part of their American Trends Panel Survey in 2018. This report detailed the findings from a sample of 4,587 participants who completed an online survey. As part of their report, Parker, Horowitz, and Igielnik (2018) discussed the disparities in attitudes toward women in leadership positions – both in political and corporate settings – by gender, political party, and age. For example, they reported that the majority of women said gender discrimination is a major reason there aren't more women in high political offices and top executive business positions; however, only 36% of men agreed it was an issue for high political offices and 44% for top executive business positions. When they looked at gender and age together, the majority of women agreed gender discrimination is a barrier to leadership across all ages, while only 35% of men 18-49 years old agreed and 38% of men 50 years and older agreed. We are going to follow up on their analyses by taking a more nuanced look at the interaction between age and gender on attitudes toward women in leadership using their data set – specifically, we are going to focus on women in political positions.

Our first factor, gender, has two levels: (1) woman and (2) man. Our second factor, age, will have three levels: (1) younger adult (30-49 years older), (2) middle-aged adult (50-64 years old), and (3) older adults (65 years and older). Our dependent variable, attitudes toward women in politics, was created by creating an average of participant responses to 23 items that accessed attitudes toward women in political positions. Higher scores on this variable mean more favorable attitudes toward women in politics.

This study utilized a quasi-experimental research design. We do not have a control group and are not randomly assigning participants to a particular gender or age. We are simply grouping participants into the gender they identified with and the age group based on their reported age when the data were collected. The table below shows our 2×3 factorial ANOVA study design.

	Factor Two: Age		
Factor One: Gender	30-49 years old (Level 1)	50-64 years old (Level 2)	65+ years old (Level 3)
Women (Level 1)	<i>30-49 years old, women</i>	<i>50-64 years old, women</i>	<i>65+ years old, women</i>
Men (Level 2)	<i>30-49 years old, men</i>	<i>50-64 years old, men</i>	<i>65+ years old, men</i>

Hypotheses

The null hypotheses:

Factor One (Main Effect of Gender)

- Conceptual H_{0A} : There are no significant mean differences in favorable attitudes toward women in politics between women and men.
- Mathematical H_{0A} : There are no significant differences between the population means for women and men on favorable attitudes toward women in politics. $M_1 = M_2$

Factor Two (Main Effect of Age Group)

- Conceptual H_{0B} : There are no significant mean differences in favorable attitudes toward women in politics between the three age groups.
- Mathematical H_{0B} : There are no significant differences between the population means for the 30-39 years old group, 50-64 years old group, and 65+ years old group on favorable attitudes toward women in politics. $M_1 = M_2 = M_3$

Factor One x Factor Two (Interaction between Gender and Age Group)

- Conceptual H_{0AB} : There is not a significant interaction between gender and age on favorable attitudes toward women in politics.
- Mathematical H_{0AB} : Gender x Age Group = 0 for each cell.

The alternative hypotheses:

Factor One (Main Effect of Gender)

- Conceptual H_{1A} : There is a significant mean difference in favorable attitudes toward women in politics between women and men.
- Mathematical H_{1A} : There is a significant difference between the population means for women and men on favorable attitudes toward women in politics. $M_1 \neq M_2$

Factor Two (Main Effect of Age Group)

- Conceptual H_{1B} : There is a significant mean difference in favorable attitudes toward women in politics between the three age groups.
- Mathematical H_{1B} : There is a significant difference between the population means for the 30-39 years old group, 50-64 years old group, and 65+ years old group on favorable attitudes toward women in politics. $M_1 \neq M_2 \neq M_3$

Factor One x Factor Two (Interaction between Gender and Age Group)

- Conceptual H_{1AB} : There is a significant interaction between gender and age on favorable attitudes toward women in politics.
- Mathematical H_{1AB} : Gender x Age Group $\neq 0$ for at least one cell.

JASP Analyses

In order to run analyses in JASP, the first thing we need to do is open the data set we will be working with. To do this, open JASP and complete the following steps.

Click on File → Open → Computer → Browse → Choose the Two-Way ANOVA Practice Data (Pew Research Center 2019 W36)

Before we get started with our analyses, we are going to filter our AgeGroup variable to filter out participants who were in the 18-29-year-old age group. Unfortunately, the sample size in this group was too small to compare to the other age groups in this analysis. Click the arrow in the filter column for the 18-29-year-old label. To do this, click on “AgeGroup” and click the checkmark to filter out participants who are in the 18-29-year-old age group (Value = 1). When there is an X instead of check mark, then you have filtered this group out appropriately. See the image below.

AgeGroup		
Filter	Value	Label
✗	1	18-29
✓	2	30-49
✓	3	50-64
✓	4	65+

Assumption Testing

Prior to running our two-way ANOVA, we first must check to make sure we meet the assumptions for this statistical test.

Assumption One: Are our factors measured at the categorical (i.e., nominal) level with two or more groups?

Yes. Our independent variables for this study are gender and age. Our first factor, participant gender, had two levels: women and men. Participants who participated in this survey either identified themselves as a woman or a man. Thus, this factor is dichotomous and categorical. If you look at the data label icon in JASP, you can see it has the appropriate icon (three overlapping circles) for a nominal/categorical variable. Our second factor, age, had three levels: 30-39 years old, 50-64 years old, and 65+ years old. This factor is categorical and has the appropriate icon in JASP to indicate it is a nominal variable. We meet this assumption.

Assumption Two: Is the dependent variable continuous (i.e., ratio or interval)?

Yes. Our dependent variable for this study is participants' mean score for their responses to the 23-items measuring their level of favorable attitudes toward women in politics. This variable is continuous, meaning it was measured at the interval or ratio level. If you look at the data label icon in JASP, you can see it has the appropriate icon (a ruler) for a scale or continuous variable. We meet this assumption.

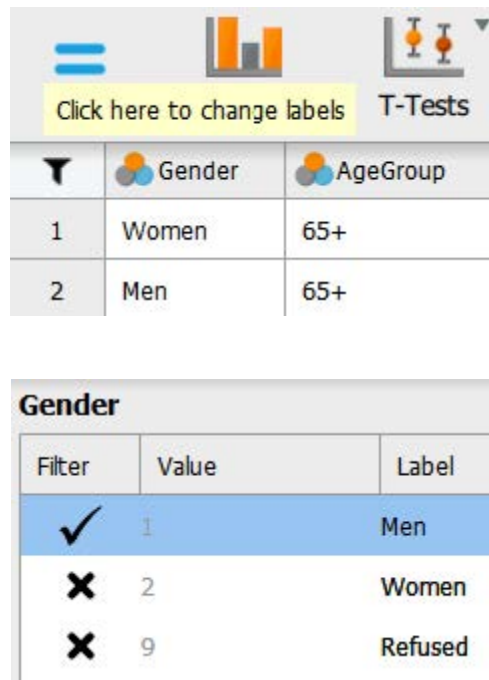
Assumption Three: Are the samples independent?

Yes. Looking at our data set, we can determine that the groups of participants in each cell of our research design are independent. In our dataset, you can see that all participants identified as either a woman or a man for factor one. You can also see that participants reported that their age group was one of the following: 30-39 years old, 50-64 years old, and 65+ years old. When we display the levels of both of our factors in a table, as you see below, we have six independent cells representing the possible combinations of our two factors for our 2×3 ANOVA. Thus, knowing the research design did not allow for participants to be exposed to more than one condition or level of either factor, we can conclude that participants in all cells are independent of one another. No participant is in more than one group. We meet this assumption.

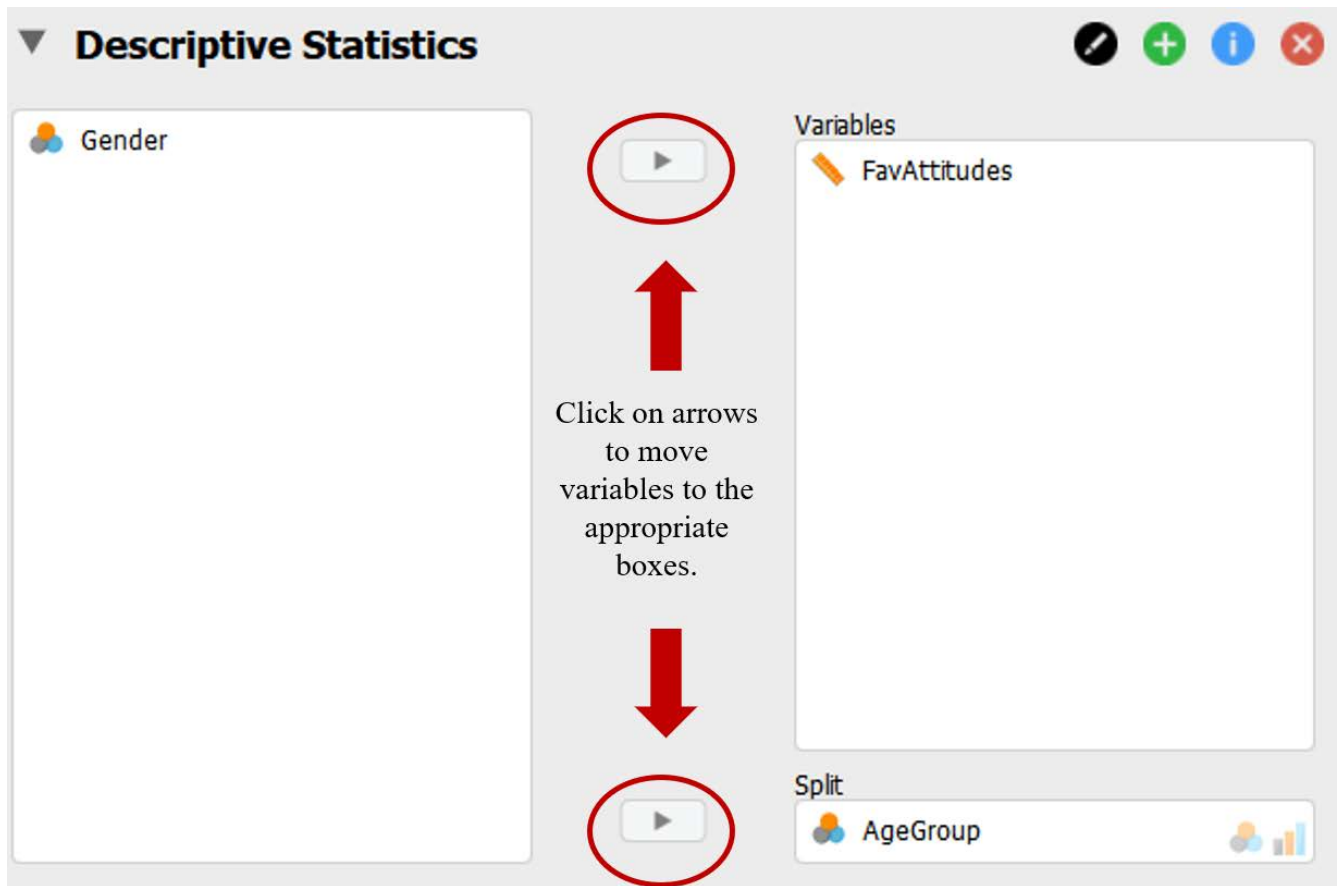
	30-49 years old	50-64 years old	65+ years old
<i>Women</i>	Cell 1 Women, 30-49 years	Cell 2 Women, 50-64 years	Cell 3 Women, 65+ years
<i>Men</i>	Cell 4 Men, 30-49 years	Cell 5 Men, 50-64 years	Cell 6 Men, 65+ years

Assumption Four: Is the dependent variable normally distributed for each group of the IV?

Before we run our descriptive statistics, we first have to filter our first factor. To do this, click on “Gender” and click the arrow to filter out participants who refused to answer (value = 9) and women (value = 2). After you click the arrow you will see an x instead of a check mark (see image below). Then click on the x in the bottom righthand corner to exit out of that screen. This will filter out women participants and allow us to calculate the descriptive statistics for the top row of our 2×3 ANOVA.



To check our data for the next two assumptions, we will use the Descriptives analysis tab. Click Descriptives. When the “Descriptive Statistics” window pops up, we will need to move our dependent variable to the “Variables” box and our independent variable to the “Split” box using the arrows depicted in the photo below – this is so we can examine normality in the DV for the three levels of our second factor separately.



Once we have our variables in the appropriate boxes and our first factor filtered, we are going to ask JASP to run the various statistics and graphics we will need to interpret for our assumptions by clicking on the appropriate boxes in the test window. We will be asking JASP to provide us with all possible output we may want to look at for our assumptions; however, we will focus on interpreting the output that you will be expected to analyze for your course assignments.

1. Under the “Basic Plots” tab we will check “Distribution plots.”
2. Under the “Customizable Plots” tab we will check the “Boxplots,” “Boxplot element,” and “Label outliers” boxes.
3. Under the “Statistics” tab we will check the “Skewness,” and “Kurtosis boxes.”

See the following image to make sure your test window has all the appropriate boxes selected.

▼ Basic plots

- ☒ Distribution plots
 ☐ Correlation plots
 ☐ Interval plots
- ☒ Display density
 ☐ Q-Q plots
- ☐ Display rug marks
 ☐ Pie charts
- Bin width type Sturges
☐ Dot plots
- Number of bins 30

▼ Customizable plots

- Color palette Colorblind
- ☒ Boxplots
- ☒ Boxplot element
 ☐ Use color palette
- ☐ Violin element
 ☒ Label outliers
- ☐ Jitter element

▼ Statistics

- Sample size**
- ☒ Valid
 ☐ Quartiles
- ☒ Missing
 ☐ Cut points for: 4 equal groups
- ☐ Percentiles:
- Central tendency**
- ☐ Mode
 ☒ Skewness
- ☐ Median
 ☒ Kurtosis
- ☒ Mean
 ☐ Shapiro-Wilk test
- ☐ Sum
- Dispersion**
- ☐ S.E. mean
 ☒ Std. deviation
- ☐ Coefficient of variation
 ☐ MAD
- ☐ MAD robust
 ☐ IQR
- ☐ Variance
 ☐ Range
- ☒ Minimum
 ☒ Maximum

As a reminder, for this course, we will focus on interpreting the skewness and kurtosis statistics to determine if our data are normally distributed. We want our skewness and kurtosis statistics to be between -2 and +2. For this study design, we have six different cells that we have to check to determine if each of those six groups of participants have a normal distribution of attitudes toward women in politics.

Descriptive Statistics

	FavAttitudes		
	30-49	50-64	65+
Valid	671	655	607
Missing	0	0	0
Mean	2.473	2.479	2.522
Std. Deviation	0.561	0.549	0.570
Skewness	-0.421	-0.237	-0.307
Std. Error of Skewness	0.094	0.095	0.099
Kurtosis	-0.597	-0.672	-0.649
Std. Error of Kurtosis	0.188	0.191	0.198
Minimum	1.077	1.125	1.000
Maximum	3.500	3.500	3.500

Looking at the values on the output above, we can see our skewness values for the 30-49-year-old group (-0.42), 50-64-year-old group (-0.24), and 65+ group (-0.31) are all within the acceptable range of -2 and +2. Additionally, the kurtosis for the 30-49-year-old group (-0.59), 50-64-year-old group (-0.67), and 65+ group (-0.65) are also within the acceptable range of -2 and +2.

Now that we have analyzed the descriptive statistics for participants in the participants who identified as a man, we have to conduct the same steps for participants who identified as a woman. To do this, click on “Gender.” Next, you will click on the check mark in the “Filter” column for Men so it is now an X. You will also need to click on the X next to Women to make sure it is now a check mark. Once this is completed, click on the x in the bottom righthand corner to exit out of that screen. This will filter out participants who identified as men and allow us to calculate the descriptive statistics for the bottom row of our 2x3 ANOVA.

Gender		
Filter	Value	Label
✗	1	Men
✓	2	Women
✗	9	Refused

Once this is completed, you will want to repeat the steps above to run Descriptives again with our remaining three groups.

Descriptive Statistics ▼

	FavAttitudes		
	30-49	50-64	65+
Valid	626	720	600
Missing	0	0	0
Mean	2.711	2.629	2.634
Std. Deviation	0.478	0.538	0.532
Skewness	-0.616	-0.548	-0.484
Std. Error of Skewness	0.098	0.091	0.100
Kurtosis	0.310	-0.045	-0.213
Std. Error of Kurtosis	0.195	0.182	0.199
Minimum	1.125	1.000	1.000
Maximum	3.500	3.500	3.500

Looking at the values on the output above, we can see our skewness values for the 30-49-year-old group (-0.62), 50-64-year-old group (-0.55), and 65+ group (-0.48) are all within the acceptable range of -2 and +2. Additionally, the kurtosis for the 30-49-year-old group (-0.31), 50-64-year-old group (-0.05), and 65+ group (-0.21) are also within the acceptable range of -2 and +2. The skewness and kurtosis statistics for each of our six cells are also reported in the table below.

	30-49 years old	50-64 years old	65+ years old
<i>Women</i>	Cell 1 Skewness: -0.62 Kurtosis: 0.31	Cell 2 Skewness: -0.55 Kurtosis: -0.05	Cell 3 Skewness: -0.48 Kurtosis: -0.21
<i>Men</i>	Cell 4 Skewness: -0.42 Kurtosis: -0.59	Cell 5 Skewness: -0.24 Kurtosis: -0.67	Cell 6 Skewness: -0.31 Kurtosis: -0.65

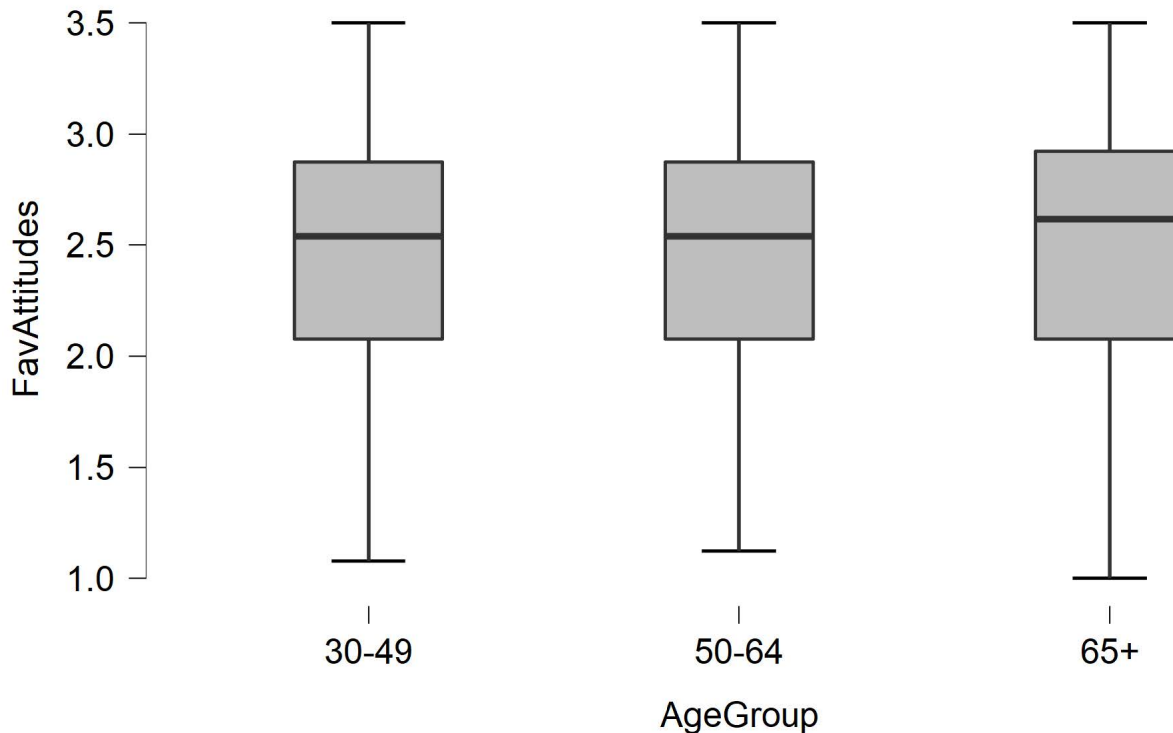
To report these results in APA format, we could write:

Attitudes toward women in politics was normally distributed within all six cells, with skewness and kurtosis statistics between -2 and +2.

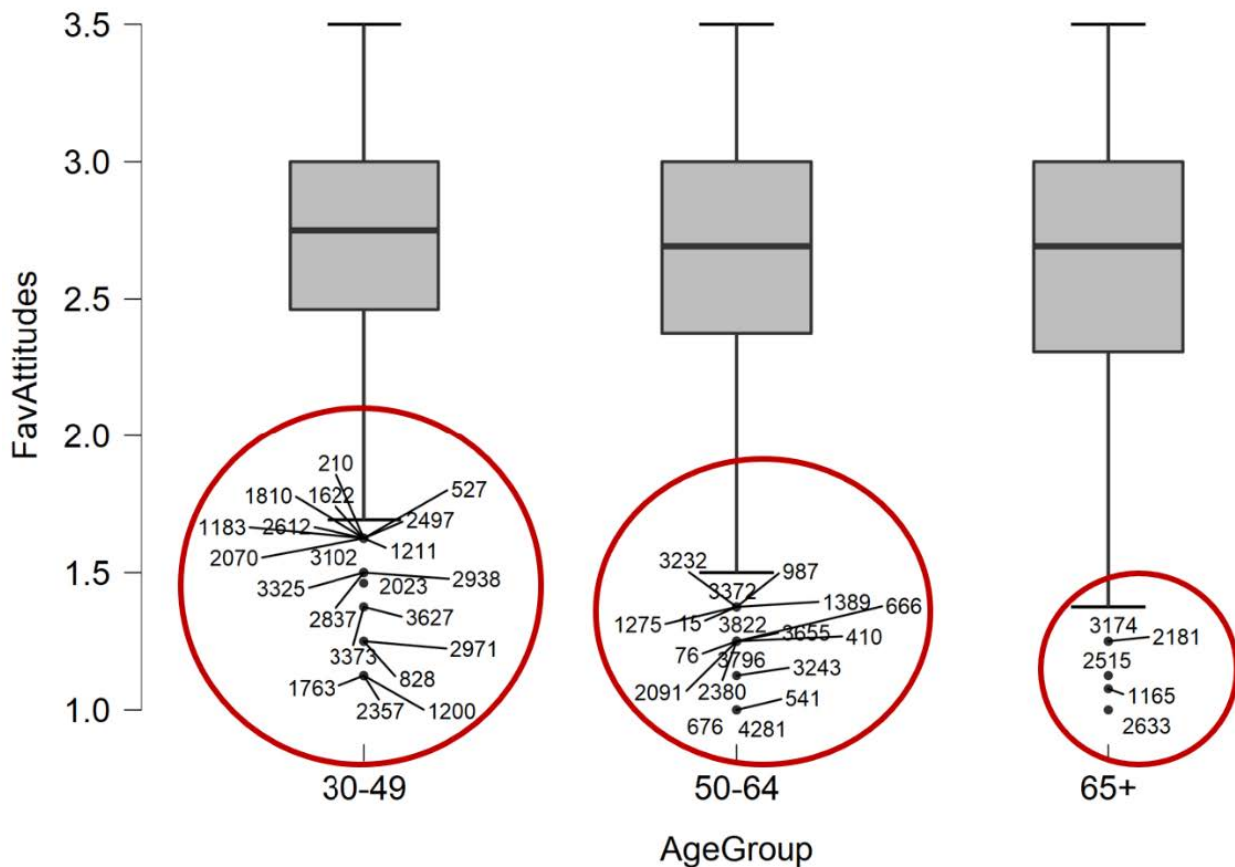
Now that we have checked our data for normality, let's take a look at our boxplots for each cell to see if we have outliers in any of our study cells.

Assumption Five: Are there any outliers in the sample?

To assess for outliers, we will look at the boxplots in our JASP output. If we have outliers, they would be outside the top and bottom lines or whiskers. When we look at the boxplot for participants who identified as a man, you can see there are no outliers visible in 30-49-year-old age group, 50-64-year-old age group, or the 65+ age group.



However, if we look at the boxplot for participants who identified as women, you can see we have a lot of outliers across all three age groups, below the lower quartile.



To report this in APA format, we could write

There were five outliers in the 65+, women group and several outliers in both the 50-64-year-old, women group and in the 30-49-year-old, women group as assessed by the inspection of a boxplot.

As we discussed in previous chapters, when we have outliers we have to decide if they are going to correct it, keep it, delete it, or replace it. This data has already been checked and cleaned by researchers with the Pew Research Center; thus, we can assume these outliers are not due to a data entry error. As a researcher, you would then have to decide if the outliers are exerting a significant enough impact on your results to warrant deletion. Given that outliers are especially common in datasets with large samples (as ours is; $N = 4,337$), and that we have so many, we will keep these outliers in our dataset in order to avoid removing a significant amount of datapoints from our model (something that is generally discouraged, Faraway, 2015). Instead, we will model how to run the analyses as normal with outliers present, and at the end of this chapter, will run the analyses with those outliers removed to examine whether the results remained the same.

Assumption Six: Are there homogeneity of variances?

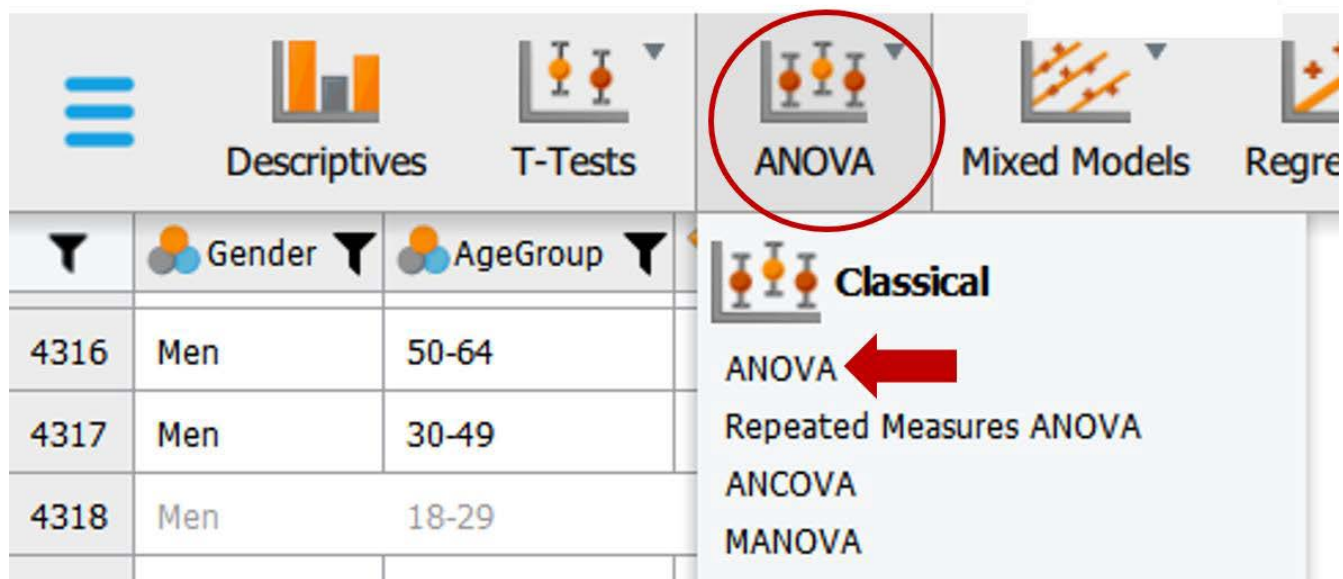
To determine if we have homogeneity of variances, we need to ask JASP to run our two-way ANOVA. Checking for this assumption is a part of the overall two-way ANOVA. Let's move onto our primary analyses below, and complete checking this assumption in that section.

Primary Analyses

Before running our primary analyses, we need to turn off the filter for Gender. To do this, hover over Gender and click when you see “click here to change labels or inspect filter.” Then click on the x icon for men. You should now see a check mark next to values 1 and 2, for men and women. Click on the X in the righthand corner to exit that screen. [You want an X to remain for Value 9].

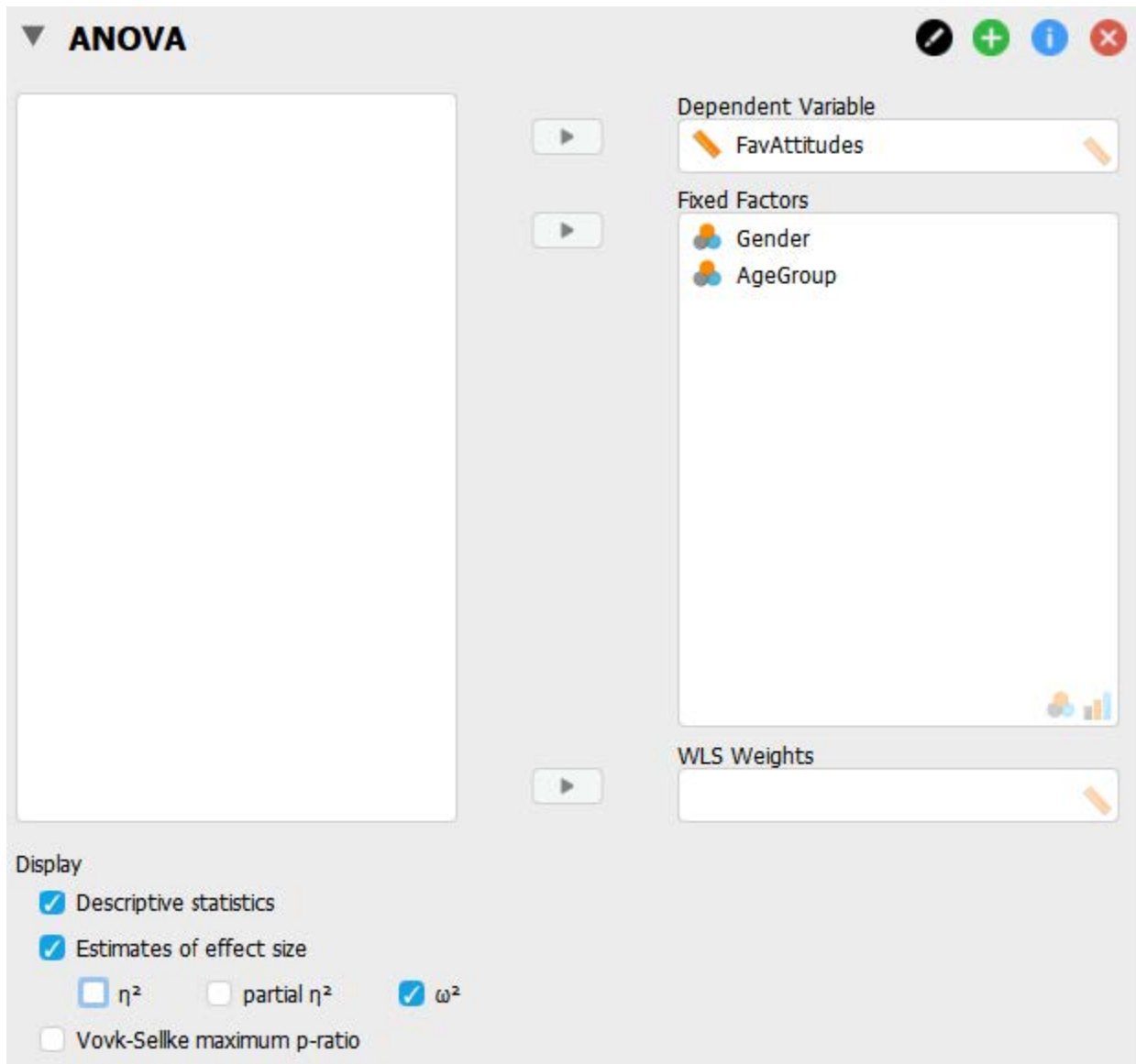
Filter	Value	Label
✗	1	Men
✓	2	Women
✗	9	Refused

To run a two-way ANOVA, go to ANOVA at the top of the JASP screen and click on ANOVA. See the image below.



1. First, we need to move our dependent variable FavAttitudes to the “Variables” box and our factors Gender and AgeGroup to the “Fixed Factors” box. We are then going to ask JASP to run our omnibus test by clicking on the appropriate boxes in the test window.
2. Under “Display” we will check the “Descriptive statistics” and “Estimates of effect size” test boxes.
3. Under “Estimates of effect size” we will check the omega squared “ ω^2 ” box.

See the image below to make sure your test window has all the appropriate boxes selected.



ANOVA

Dependent Variable: FavAttitudes

Fixed Factors: Gender, AgeGroup

WLS Weights:

Display

- ☒ Descriptive statistics
- ☒ Estimates of effect size
 - ☐ η^2 ☐ partial η^2 ☒ ω^2
- ☐ Vovk-Sellke maximum p-ratio

Next, we are going to ask JASP to give us the statistics for our final assumption – homogeneity of variances.

4. Under “Assumption Checks” we will check the “Homogeneity tests” box.



Assumption Checks

- ☒ Homogeneity tests

Homogeneity corrections

- ☒ None ☐ Brown-Forsythe ☐ Welch
- ☐ Q-Q plot of residuals

Are there homogeneity of variances?

Now that we have asked JASP to run our first set of analyses, we can begin our interpretation of the results.

The first thing we need to do is go back to our last assumption and determine if we meet the assumption for homogeneity of variances. As you can see from the output below, the p value for our Levene's test is $< .001$. Because p is $< .05$, we can conclude that we do not meet this assumption. Although this is a bummer, we are in luck! As long as the number of participants in each group is approximately equal and not less than five per group, ANOVA is generally "robust" to a violation of this assumption. If we look at the descriptives box in our output, we can see the sample sizes in each cell are approximately equal: 671, 655, 607, 626, 720, and 600. Thus, we can proceed with our analyses. If we did not have equal sample sizes in each of our groups, we could potentially conduct a nonparametric test equivalent instead, the Scheirer-Ray-Hare test.

- If p is less than .05, you do not meet the homogeneity of variances assumption.
- If p is greater than .05, you do meet the homogeneity of variances assumption.

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
8.140	5.000	3873.000	$< .001$

To report the results of the Levene's test in APA format, it might look something like this:

There was not homogeneity of variances, as assessed by the Levene's test for equality of variances ($p < .05$); however, because there was an approximately equal number of participants in each cell, the two-way ANOVA is considered robust to this violation (Maxwell & Delaney, 2004).

Interpreting the Statistical Significance of Two-Way ANOVAs

Now that we have completed the assumption checking for this statistical test, we can proceed with interpreting the primary analyses. The first thing we will interpret is the statistical significance of the omnibus test results. To do this, we are going to look at the "ANOVA – FavAttitudes" table. We have a total of three different F tests that we will interpret for each of our hypotheses.

ANOVA - FavAttitudes ▼

These are our degrees of freedom (*df*) for our analyses. (1) *df* 1, for Factor One. (2) *df* 1, for Factor Two. (3) *df* 1, for the interaction. (4) *df* 2, for both main effects and the interaction.

These are our *p* values for our analyses. (1) *p* value for Factor One. (2) *p* value for Factor Two. (3) *p* value for the interaction. Remember, if $p < .05$, we reject the null hypothesis; if $p > .05$, we fail to reject the null hypothesis.

Cases	Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>	ω^2
Gender	26.843	(1) 1	26.843	92.399	< .001	0.023
AgeGroup	0.993	(2) 2	0.496	1.709	0.181	3.563e -4
Gender * AgeGroup	2.651	(3) 2	1.325	4.562	0.010	0.002
Residuals	1125.159	(4) 3873	0.291			

Note. Type III Sum of Squares

These are our *F* statistics for our analyses. (1) *F* value for Factor One. (2) *F* value for Factor Two. (3) *F* value for the interaction.

These are our omega-squared values, the effect size we are looking at for these analyses. An omega-squared value of .01 is a small effect, .06 is a medium effect, and .14 is a large effect.

To interpret the results, we want to look at the *p* values for the main effect for factor one (hypothesis one), the main effect for factor two (hypothesis two), and the interaction (hypothesis three). If *p* is less than our alpha level of .05, we will reject the null hypothesis.

1. There is a significant statistical difference between the group means of our first factor, gender, because $p < .001$. Because our *p* value is less than our alpha level of .05, we will reject our first null hypothesis. Thus, we would say we have a main effect of gender on participant attitudes toward women in politics.
2. There is not a significant statistical difference between the group means of our second factor, age group, because $p = .18$. Because our *p* value is greater than our alpha level of .05, we will fail to reject our second null hypothesis. Thus, we would say we do not have a main effect of age on participant attitudes toward women in politics.
3. There is a significant interaction between gender and age group because $p = .01$. Because our *p* value is less than our alpha level of .05, we will reject our third null hypothesis. Thus, we would say we do have a significant interaction between gender and age group on participant attitudes toward women in politics.

Writing the basic results of a two-way ANOVA in APA format follows this general format:

$$F(df1, df2) = F \text{ statistic}, p < .05 \text{ or } > .05, \omega^2 = \text{omega-squared statistic}$$

However, with a two-way ANOVA, we have **three** *F* statistics to report: (1) main effect for our first factor, (2) main effect for our second factor, and (3) our interaction. Plugging our results above into this format, we have:

$$\text{Main Effect for Factor One: } F(1, 3873) = 92.39, p < .001, \omega^2 = .02$$

$$\text{Main Effect for Factor Two: } F(2, 3873) = 1.71, p > .05, \omega^2 = .00$$

$$\text{Main Effect for the Interaction: } F(2, 3873) = 4.56, p < .05, \omega^2 = .002$$

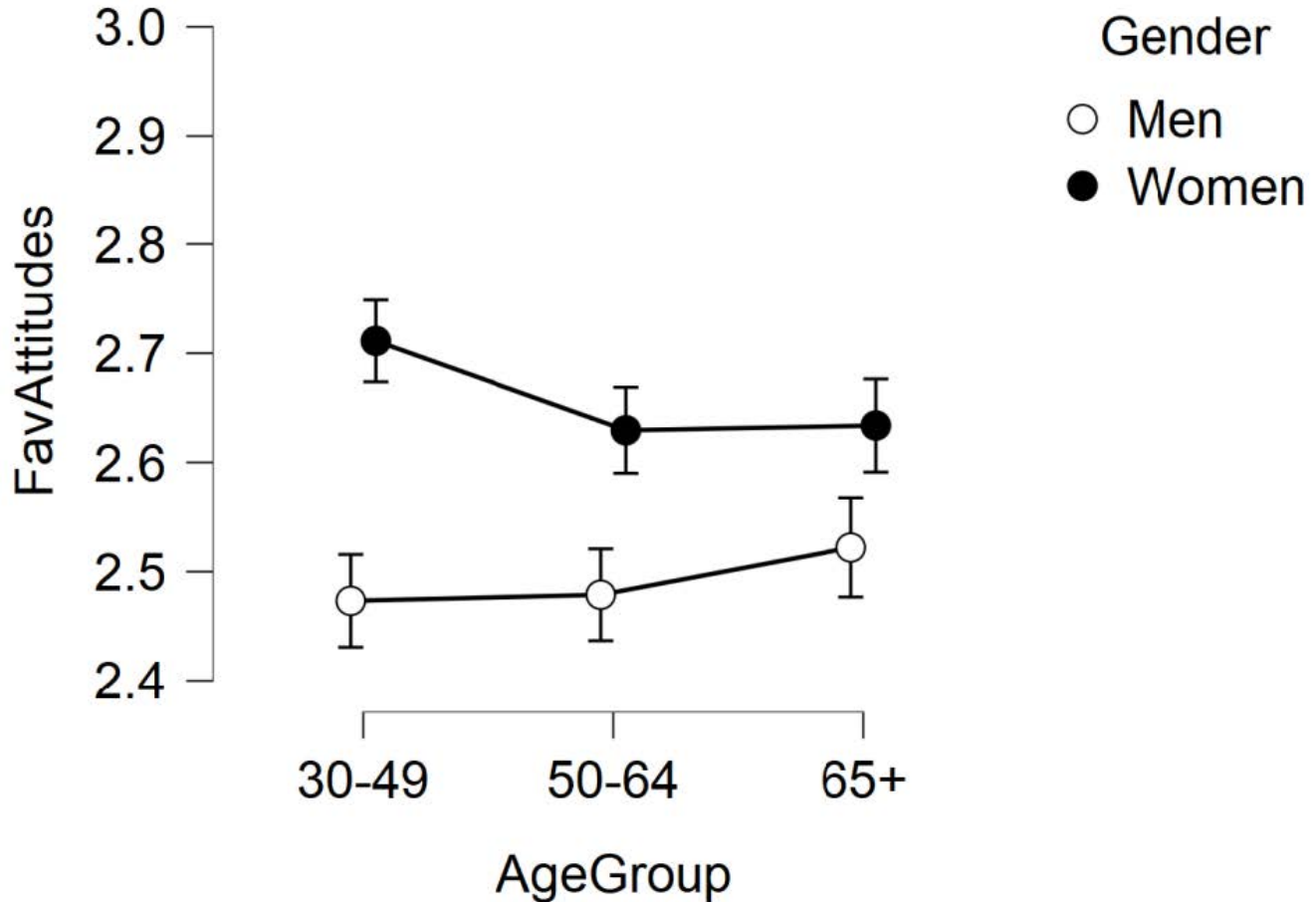
Now we know that we have a significant interaction between our two factors – gender and age group – but what does that mean? To answer this, we are going to do two things: (1) graph the means of each cell in our

interaction to help us visualize what is happening with the data and (2) conduct post hoc tests called simple main effects.

We can graph the means of the interaction using JASP or using Excel. For now, let's take the easy way out and ask JASP to provide us with a graph of the interaction.

1. In the ANOVA analysis window, click on “Descriptive Plots.”
2. Move Gender to the “Separate Lines” box. This will graph our first factor, gender, as two separate lines on our graph.
 1. I choose this variable to have separate lines for because it will make it easier to understand two separate lines rather than three. However, you can always play around with different options to see what method works best for your study.
3. Move AgeGroup to the “Horizontal Axis” box. This will group our second factor, age group, on the x-axis of the graph.
4. Click “Display error bars.”

The screenshot shows the 'Descriptives Plots' window in JASP. On the left, under 'Factors', there is an empty box. In the center, there are three right-pointing arrow buttons. On the right, there are three boxes: 'Horizontal Axis' containing 'AgeGroup', 'Separate Lines' containing 'Gender', and 'Separate Plots' which is empty. At the bottom, under 'Display', the 'Display error bars' checkbox is checked. Below it, the 'Confidence interval' radio button is selected, with a value of '95.0' and a '%' symbol. The 'Standard error' radio button is unselected.



Looking at the Descriptive plots or line graph above, we can see that we have an ordinal interaction – the groups means do not cross or overlap in our graph. By graphing the means of each of our study cells, we can visualize that is happening with our data. For example, looking at this graph we can see women in the younger (30-49-year-old) age group reported the most favorable attitudes toward women in politics. Alternately, men in the younger age group reported the least favorable attitudes toward women in politics. This graph is limited because it does not tell us which groups are significantly different from one another. However, by asking JASP to display error bars using the confidence intervals, we are able to take an educated guess as to where the statistical differences will be. The dots on the graph indicate the group means and the error bars indicate the 95% confidence intervals around the means. A general indication of how spread out or variable participant ratings of attitudes toward women in politics were in each group. Groups that do not have overlapping error bars are more likely to be significantly different from one another; however, this is not guaranteed – so post hoc tests are needed to confirm. Looking at the line graph above, it seems likely that women across all age groups had significantly more favorable attitudes toward women in politics than men. The group with the smallest gender differences in attitudes toward women in politics was the older adult age group. To confirm our interpretation of the graph is correct, we will conduct post hoc tests. A commonly used post hoc test to help interpret significant interactions with two-way ANOVAs are simple main effects.

Our next step is to run tests of simple effects.

1. In the ANOVA analysis window, click on “Simple Main Effects.”
2. Move AgeGroup to the “Simple Effect Factor” box.
3. Move Gender to the “Moderator Factor 1” box.

Simple Main Effects

Factors

Simple Effect Factor

AgeGroup

Moderator Factor 1

Gender

Moderator Factor 2

This will test to see if there are differences in our second factor, age group, at each level of the first factor (women, men). Our results are copied below.

Simple Main Effects ▼

Simple Main Effects - AgeGroup ▼

Level of Gender	Sum of Squares	df	Mean Square	F	p
Men	0.895	2	0.448	1.541	0.214
Women	2.717	2	1.359	4.677	0.009

These are our p values for our analyses. The top p value is for the simple effect of Age for Men. The bottom p value is for the simple effect of Age for Women.

To interpret simple effects, we will focus on the p values for each comparison.

- The p value on top ($p = .21$), tells us that when we compare participant attitudes toward women in politics between young, middle-aged, and older *men*, there is no statistical difference because p is greater than .05.
- The p value on the bottom ($p = .009$), tells us that when we compare participant attitudes toward women in politics between young, middle-aged, and older *women*, there is a statistical difference because p is less than .05.

We can repeat these steps to run simple main effects to see if there are differences in our first factor, gender, at

each level of the second factor (younger, middle-aged, older). We will follow the same instructions as before but switch the boxes we move Gender and AgeGroup into.

1. In the ANOVA analysis window, click on “Simple Main Effects.”
2. Move Gender to the “Simple Effect Factor” box.
3. Move AgeGroup to the “Moderator Factor 1” box.

Simple Main Effects

Factors

Simple Effect Factor: Gender

Moderator Factor 1: AgeGroup

Moderator Factor 2:

The results of the additional simple main effects analyses are copied below.

Simple Main Effects ▼

Simple Main Effects - Gender ▼

Level of AgeGroup	Sum of Squares	df	Mean Square	F	p
30-49	18.343	1	18.343	63.140	< .001
50-64	7.770	1	7.770	26.747	< .001
65+	3.758	1	3.758	12.937	< .001

These are our p values for our analyses. The top p value is for the simple effect of Gender for 30-49 year old participants. The middle p value is for the simple effect of gender for 50-64 year old participants. The bottom p value is for the simple effect of gender for participants 65 years and older.

To interpret simple effects, we will focus on the p values for each comparison.

- The p value on top ($p < .001$), tells us that when we compare attitudes toward women in politics for women and men who are 30-49 years old, there is a significant difference because p is less than .05.
- The p value in the middle ($p < .001$), tells us that when we compare attitudes toward women in politics for women and men who are 50-64 years old, there is a significant difference because p is less than .05.
- The p value on the bottom ($p < .001$), tells us that when we compare attitudes toward women in politics for women and men who are 65 years and older, there is a significant difference because p is less than .05.

To report the results of our simple main effects analyses, we could write:

After conducting analyses of simple main effects, we found a significant difference in attitudes toward women in politics between women and men across in the younger adult ($F(1) = 63.14$, $p < .001$), middle-aged adult ($F(1) = 26.75$, $p < .001$), and older adult age groups ($F(1) = 12.94$, $p < .001$). Across all age groups, women had significantly more favorable views toward women in politics compared to men. Further, there was not a simple main effect of age for men, $F(2) = 1.54$, $p > .05$. There were no differences in attitudes toward women in politics across age groups for men. However, there was a simple main effect of age for women, $F(2) = 4.68$, $p < .01$. Younger women held more favorable attitudes toward women in politics than the middle-aged and older age groups.

Interpreting the Practical Significance of Two-Way ANOVAs

Now that we have interpreted the statistical significance, we will look at the practical significance by looking at our effect size. A significant p value tells us that there is a significant interaction between gender and age on attitudes toward women in politics, but the effect size tells us how much of our dependent variable we are explaining with these factors.

Reminder: Omega Squared Effect Sizes (Cohen, 1988; Kirk, 1996)

Effect Size	Strength
.01	Small
.06	Medium
.14	Large

We obtained an omega-squared value of .002 for our interaction, this is less than a small effect. There is a statistical interaction between gender and age on attitudes toward women in politics, but the variance explained by that interaction is very *small*. Meaning, the interaction between gender and age only explains .2% of the variance in participant attitudes toward women in politics. If we look at our main effect for gender, however, the omega-squared value was .02. This means that the variance in participant attitudes toward women in politics explained by gender alone was 2.3%, still a *small* amount, but larger than the interaction.

Reporting in APA Format

What do you need to report in your results?

Value (Notation)		
Means (M)	Standard Deviations (SD)	Test Statistics (F)
Degrees of Freedom (df)	Probabilities (p)	ω^2 or η^2
Confidence Intervals	Assumptions	Skewness and Kurtosis
Boxplot	Levene's Statistic	Post Hoc Test
Simple Main Effect	Effect Size	

Putting it All Together

Now, let's write up the results of our test combining everything we've done so far, including the test of assumptions, and the results of the two-way ANOVA.

A two-way ANOVA was conducted to determine if there was an interaction between gender and age on attitudes toward women in politics. Attitudes toward women in politics was normally distributed within all six cells, with skewness and kurtosis statistics between -2 and +2. There were five outliers in the 65+, women group; 19 outliers in the 50-64-year-old, women group; and 21 outliers in the 30-49-year-old, women group as assessed by the inspection of a boxplot. There was not homogeneity of variances, as assessed by the Levene's test for equality of variances ($p < .05$); however, because there was an approximately equal number of participants in each cell, the two-way ANOVA is considered robust to this violation (Maxwell & Delaney, 2004).

There was a statistically significant interaction between gender and age on attitudes toward women in politics, $F(2, 3873) = 4.56, p < .05, \omega^2 = .002$. There was also a main effect of gender on attitudes toward women in politics, $F(1, 3873) = 92.39, p < .001, \omega^2 = .02$. There was not a main effect of age group, $F(2, 3873) = 1.71, p > .05, \omega^2 = .00$. After conducting analyses of simple main effects, we found a significant difference in attitudes toward women in politics between women and men across in the younger adult ($F(1) = 63.14, p < .001$), middle-aged adult ($F(1) = 26.75, p < .001$), and older adult age groups ($F(1) = 12.94, p < .001$). Across all age groups, women had significantly more favorable views toward women in politics compared to men. Further, there was not a simple main effect of age for men, $F(2) = 1.54, p > .05$. There were no differences in attitudes toward women in politics across age groups for men. However, there was a simple main effect of age for women, $F(2) = 4.68, p < .01$. Younger women held more favorable attitudes toward women in politics than the middle-aged and older age groups.

Real World Meaning

Remember, when we interpret the real world meaning of a study, we want to take out any statistical jargon and describe the results of the study in a way that is easy to understand by people who do not know anything about statistics. Think about how you would describe the results to your roommate, sibling, parent, or neighbor. For example, if I was telling my neighbor the real world meaning of our results, I would say:

When researchers studied the impact of gender and age on attitudes toward women in politics

they found a few things: (1) men have less favorable attitudes toward women in politics across all age groups, (2) there are differences in attitudes toward women in politics between young, middle-aged, and older women – but not between different ages for men, and (3) younger women have the most favorable attitudes toward women in politics. Younger women had more favorable attitudes toward women in politics compared to all of the men in the study, and more favorable attitudes compared to the middle-aged and older women as well.

References

Parker, K., Horowitz, J., & Pumphrey, J. (2018). Women and leadership 2018. Pew Research Center. Retrieved from https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2018/09/Gender-and-leadership-for-PDF_updated-10.1.pdf

Authors

This guide was written and created by Ruth V. Walker, PhD, Ashlyn A. Moraine, and Erin N. Prince.

Acknowledgements

We would like to thank the Pew Research Center for generously making their data available for secondary analysis. Additionally, we appreciate Dr. Kristen Black's edits and suggestions during the creation of this guide.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting a Two-Way ANOVA in JASP Copyright © 2021 by Ashlyn A. Moraine; Erin N. Prince; and Ruth Walker is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

5.

RUNNING AND INTERPRETING A CHI-SQUARE TEST IN JASP

Ashlyn A. Moraine; Dr. Hannah J. Osborn; and Ruth Walker

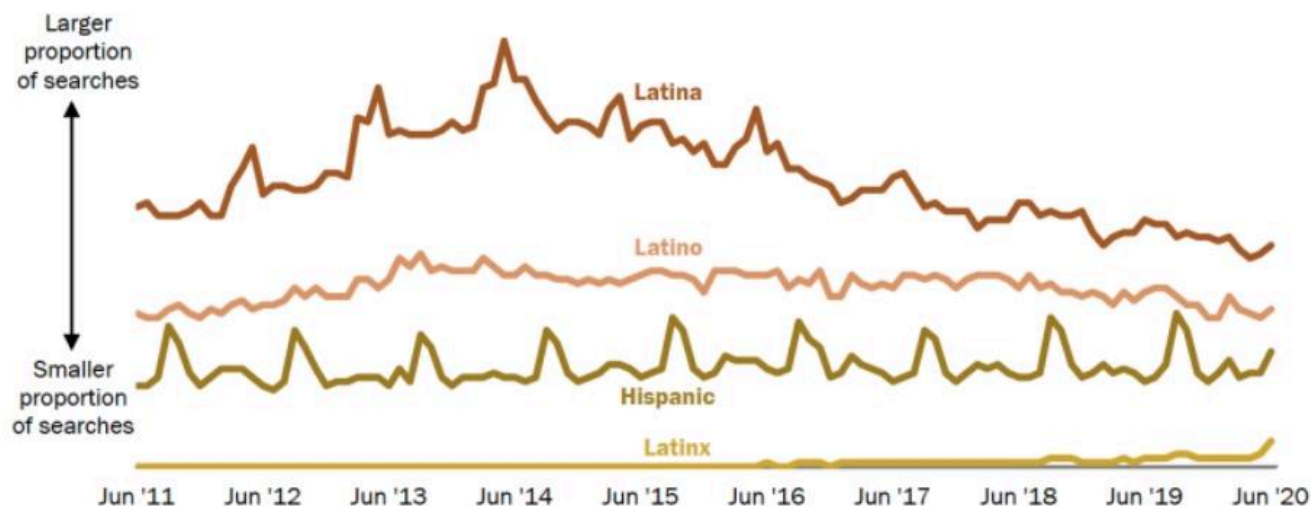
Download Data Sets:

- JASP Data File Chi-Square Test
- CSV Data File Chi-Square Test

The terms used to describe their ethnic identity by people who can trace their origins back to South America, Latin America, and Spain varies between individuals and have changed over time. In a national survey of 5,103 Hispanic adults in 2013, the Pew Research Center found that the majority of participants defined their ethnic identity in terms of their specific Hispanic origin, particularly those who immigrated to the U.S. from other countries. For example, using terms that indicated a geographical origin such as Dominican, Cuban, South American, Puerto Rican, or Salvadoran. The rest of the participants used the terms Hispanic, Latino, or American, with higher proportions of native-born participants using these terms – particularly amongst third generation members of the Hispanic community. Recently, a gender-neutral term, Latinx, has been gaining momentum as a new way of describing the Hispanic population in the U.S. Why use the term Latinx instead of Latino, Latina, or Hispanic? As one younger participant in a national survey explained, “Latinx is a more inclusive term to use for those who do not choose to identify with a certain gender. The terms Latino and Latina are very limiting for certain people” (Noe-Bustamante, Mora, & Hugo Lopez, 2020).

Although Google trend data between 2011-2020 shows a rise in the prevalence of Latinx, the terms Latina, Latino, and Hispanic continue to be more widely used.

Proportion of all monthly Google searches in the U.S. that contain the term ...



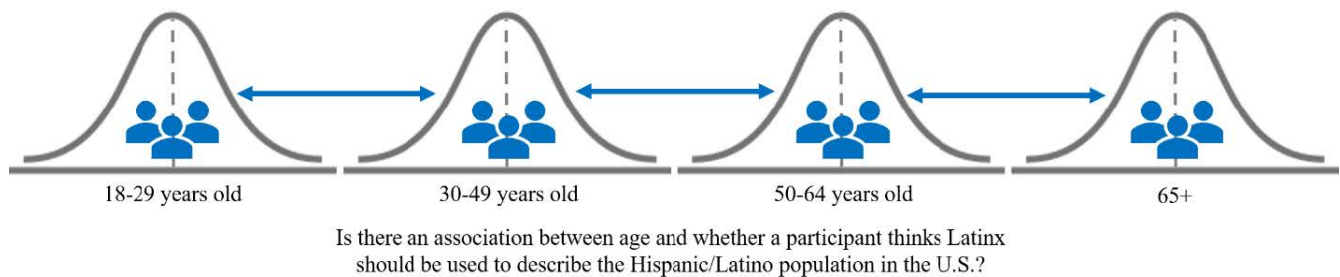
Note: Google Trends data does not indicate the number of searches but instead are standardized data, displaying the relative change in searches over the time period on a 0 to 100 scale. Google Trends are monthly averages based on weekly volume.

Source: Pew Research Center analysis of Google Trends data for June 2011 to June 2020, accessed July 24, 2020.

Pew Research Center Graph (2020)

A recent survey by researchers at the Pew Research Center found that 76% of U.S. Hispanic/Latino adults surveyed in late 2019 had not heard of the term Latinx (Noe-Bustamante et al., 2020); however, of the 23% who had heard of the term, only 3% reported using it to describe themselves. Additionally, younger participants between 18-29 years old were the most likely to have heard of the term Latinx, with 42% saying they were familiar with the term compared to only 7% of older adults 65 years and older. When asked if they thought Latinx should be adopted as a pan-ethnic term for U.S. Hispanics, the majority of participants (61%) said they prefer the term Hispanic, 29% preferred Latino, and only 4% said they preferred Latinx. Knowing there are age differences in terms of how aware participants are of the term Latinx, are there also age differences in whether participants think the term Latinx should be adopted by the Hispanic/Latino population? Let's find out!

For this walk-through example, we are going to focus on two of these variables: "Age" and "Latinx." We want to see whether there is an association between age groups (18-29, 30-49, 50-64, and 65+) and whether a person thinks Latinx should be used to describe the Hispanic or Latino population (Yes, No). To answer this question, we will conduct a chi-square test.

Noc-Bustamante et al., (2020) Study*Acceptance of Latinx to describe Hispanic population by age***Hypotheses**

Let's review our hypotheses for this example we are about to run in JASP.

The null hypothesis is:

Conceptual H_0 : There is no significant association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population

Mathematical H_0 : The observed frequencies are equal to the expected frequencies.

The alternative hypothesis is:

Conceptual H_1 : There is a significant association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population

Mathematical H_1 : The observed frequencies are not equal to the expected frequencies.

JASP Analyses

In order to run analyses, the first thing we need to do is open the data set we will be working with. To do this, open JASP and follow the steps below:

File → Open → Computer → Browse → Select the 'Chi Square Class Practice Data (Pew Research Center, 2020)' csv file wherever it is saved on your computer.

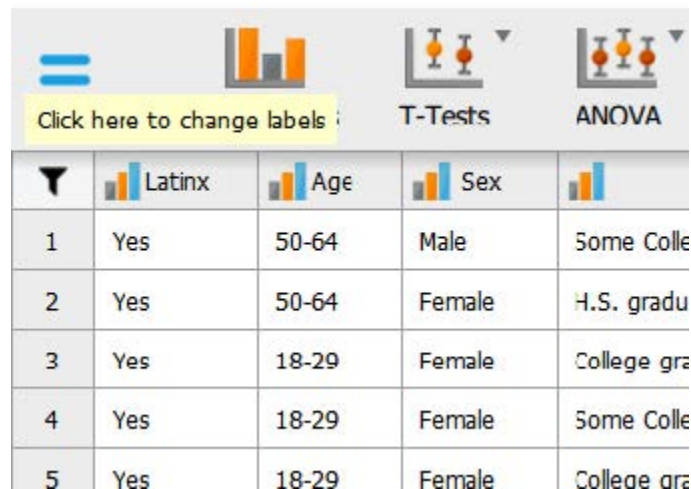
Once the data is open in JASP, we will need for first change the data labels for our variables so that we know what the values mean when we run analyses. There are multiple variables in the dataset, however, we are going to focus on Latinx and Age for this example. See the following bulleted list for a description of how each variable was coded.

- **Latinx:** 1 = 'Yes', 2 = 'No'
- **Age:** 1 = '18-29', 2 = '30-49', 3 = '50-64', 4 = '65+', 99 = 'Refused'
- **Sex:** 1 = 'Male', 2 = 'Female', 99 = 'Refused'
- **Education:** 1 = 'College graduate', 2 = 'Some college', 3 = 'H.S. graduate or less', 99 = 'Refused'

Before we do anything else, let's get our dataset to reflect the participants we want to analyze in this chi-square test. Remember, we are comparing participants in different age groups and whether they believe the term Latinx should be used to describe the Hispanic or Latino population. To do this, we want to first make sure

we filter out any participants who did not respond to either prompt. For both Latinx and Age we are going to filter out participants with the “Refused” label.

To change the numerical data into our categorical labels, you will take your cursor and hover over the column for the variable you want to change. Let’s do ‘Latinx’ together, and then you can do the others on your own. Hover over the variable column ‘Latinx’. When you see a pop up saying, “click here to change labels” click on it. To change the labels, click on the 1 under the Label column and type ‘Yes’ and hit enter. Then click on the 2 under the Label column and type ‘No’ and hit enter. This will change your labels for the variable ‘Latinx’.



	Latinx	Age	Sex	
1	Yes	50-64	Male	Some Colle
2	Yes	50-64	Female	H.S. gradu
3	Yes	18-29	Female	College gra
4	Yes	18-29	Female	Some Colle
5	Yes	18-29	Female	College ara

To filter out participants, we will click the check mark next to the responses we want filtered out. In the variable labels box, click the check mark next to the ‘99’ label, and it will turn the button into an X button.

Latinx		
Filter	Value	Label
✓	1	Yes
✓	2	No
✗	99	Refused

Click the check mark to filter out participants who indicated 'Refused'. The check mark should change to an X button.

Repeat the above steps to filter out participants who refused to provide their age for our "Age" variable.

Age		
Filter	Value	Label
✓	1	18-29
✓	2	30-49
✓	3	50-64
✓	4	65+
✗	99	99

Assumption Testing

There are just three statistical assumptions for a chi-square test. The first two deal more with the research design, and the third assumption we will test when we actually run the chi-square test.

Assumption 1: Two Categorical Variables.

Yes. In this example, we have two categorical, nominal variables. Our first variable, 'Latinx' is a dichotomous categorical variable, where participants responded that they either think this term should be used to describe the Hispanic or Latino population, or not. Our second variable, 'Age group', is also a categorical variable,

because participants either indicate their age as '18-29' or '30-49' or '50-64' or '65+.' Because both of our variables are categorical variables, we meet this assumption.

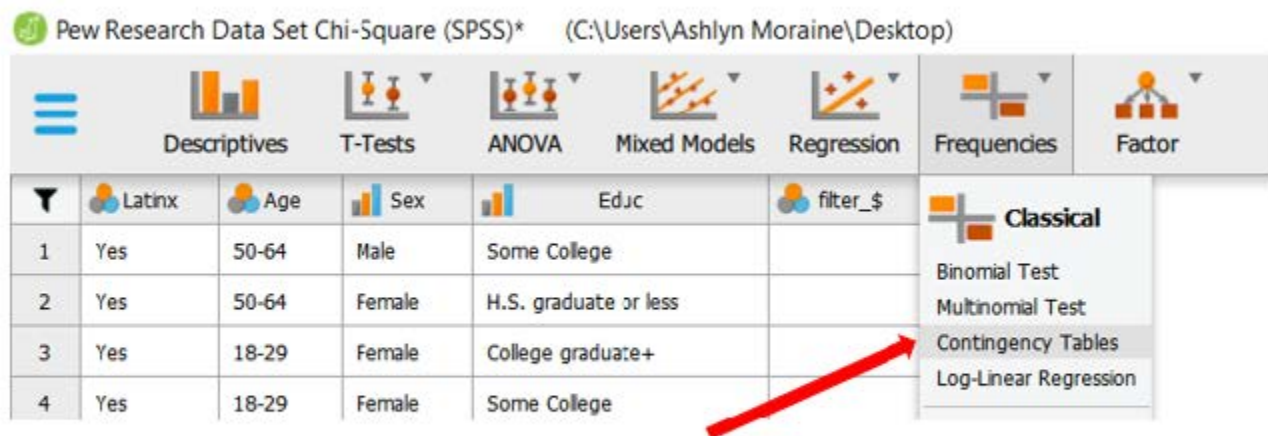
Assumption 2: Independence of Observations.

Yes. We can assume that we have independence of observations because each participant represents a unique combination of the different levels of each variable. That means that for any given participant, they are only in one of the total possible “cells” in the contingency table. For instance, a participant is in one of eight possible cells in the following contingency table. Because participants can only be in one of these cells based on their specific categories, we can assume independence of observations. So, we have met this assumption.

	18-29 years old	30-49 years old	50-64 years old	65+
Yes Latinx should be used to describe the Hispanic/Latino population.	18-29-year olds who said “Yes”	30-49-year olds who said “Yes”	50-64-year olds who said “Yes”	65+ year olds who said “Yes”
No Latinx should NOT be used to describe the Hispanic/Latino population.	18-29-year olds who said “No”	30-49-year olds who said “No”	50-64-year olds who said “No”	65+ year olds who said “No”

Assumption 3: Expected Counts Greater than Five.

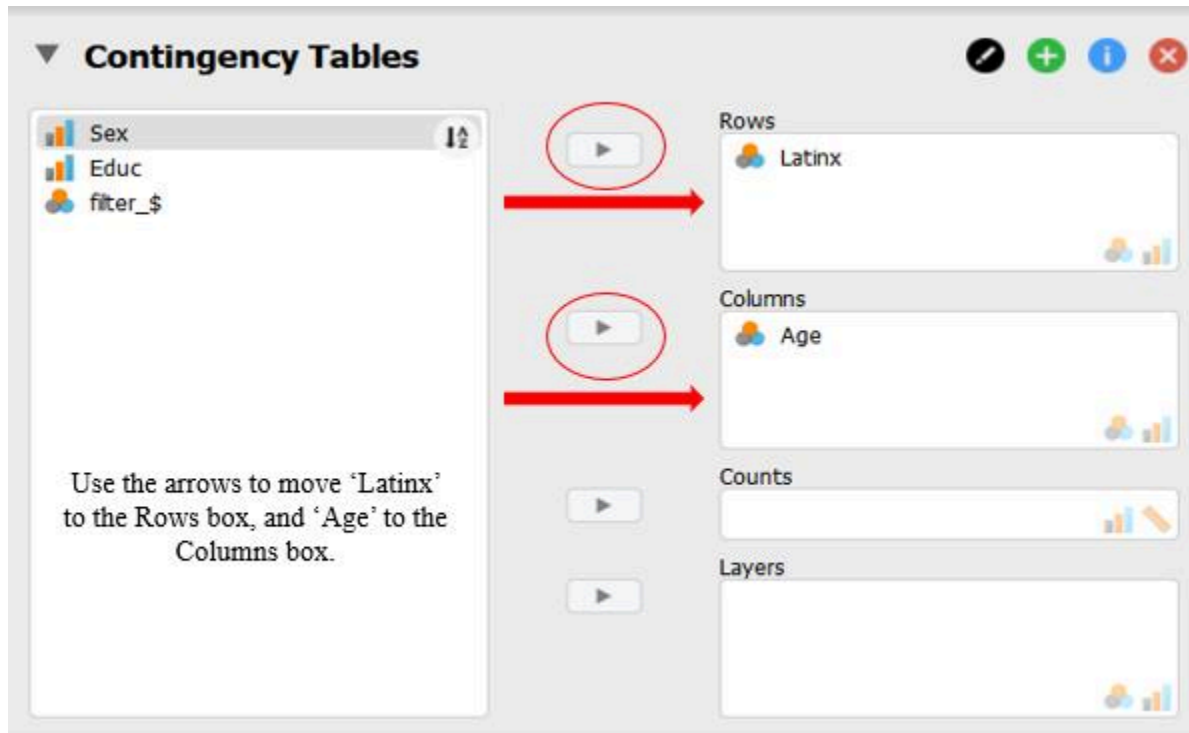
To determine whether our expected counts are greater than five, we need to have JASP run the chi-square test. To run the chi-square test, go to ‘Frequencies’ at the top of the JASP window and select ‘Contingency Tables’, as shown in the following image.



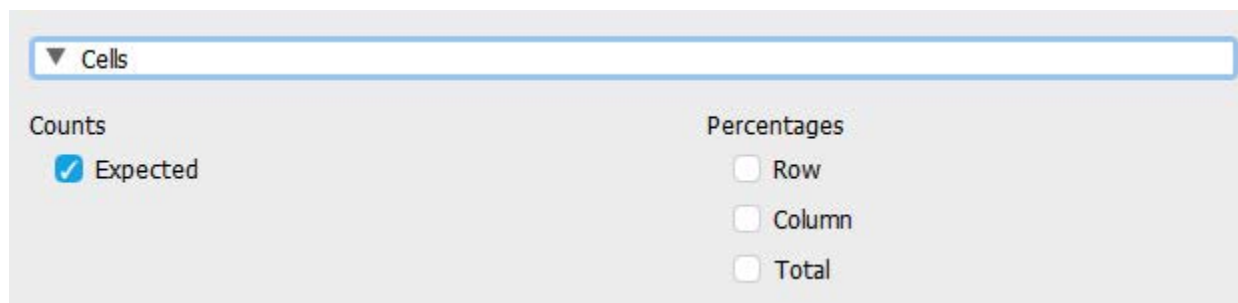
Under the 'Frequencies' option,
select 'Contingency Tables' from
the drop-down list.

Primary Analyses

Now we need to choose which variables to move into the ‘Row’ and ‘Column’ boxes. Ultimately, it does not matter how you choose to organize your table, so choose the format that will work the best for you. For this dataset we will use the top arrow to move the variable ‘Latinx’ into the ‘Row’ box. Click the other arrow to move the ‘Age’ variable to the ‘Column’ box.



To test our third assumption (expected counts greater than five), click on the ‘Cells’ and select the “Expected” check mark under the “Counts” heading.



In the output window, take a look at the table labeled “Contingency Table”. The rows that are labeled “Count” refer to the observed counts based on participants responses. Take a look at the expected counts highlighted in the image below with red boxes. What we can see for each of our cells is that our expected counts are greater than five for each cell. This means that we have satisfied assumption three and can move on with the chi-square test!

Can You Read the Contingency Table?

Let's make sure you can read the contingency table below. How many people in our sample are middle-aged (50-64 years old) and think Latinx should be used to describe the Hispanic/Latino population in the United States? To figure this out, we need to make sure we're looking at the right cell. We'll want to look at the row labeled "Yes", and within this row, pay attention to the row labeled 'Count'. Now that we're in the right row, we need to make sure we're in the right column. We want people who are 50-64 years old. If we look for the cell based on that combination of categories, you should end up at a value of 45. This means that 45 50-64-year olds indicated they thought Latinx should be used to describe the Hispanic/Latino population in the U.S.

Contingency Tables ▼

		Age				Total
		18-29	30-49	50-64	65+	
Yes	Count	110.000	86.000	45.000	7.000	248.000
	Expected count	90.765	92.114	48.250	16.871	248.000
No	Count	159.000	187.000	98.000	43.000	487.000
	Expected count	178.235	180.886	94.750	33.129	487.000
Total	Count	269.000	273.000	143.000	50.000	735.000
	Expected count	269.000	273.000	143.000	50.000	735.000

To report this assumption test in APA format, you could say:

A chi-square test for association was conducted between age group and whether people think the term Latinx should be used to describe the Hispanic or Latino population. All expected cell frequencies were greater than five.

To run the chi-square test and select all other pieces of information we'll need, click on the 'Statistics' option in the test menu. You'll see that the chi-square test is already selected. All we'll need to add now is Phi and Cramer's V to get the effect size. To do this, select the checkbox labeled "Phi and Cramer's V."

▼ Statistics

☒ χ^2

☐ χ^2 continuity correction

☐ Likelihood ratio

☐ Vovk-Sellke maximum p-ratio

☐ Odds ratio (2x2 only)

☒ Log Odds Ratio

Confidence interval %

Alt. Hypothesis (Fisher's exact test)

☐ Group one \neq Group two

☐ Group one $>$ Group two

☐ Group one $<$ Group two

Nominal

☐ Contingency coefficient

☒ Phi and Cramer's V

☐ Lambda

Ordinal

☐ Gamma

☐ Kendall's tau-b

Interpreting and Reporting the Statistical Significance of the Chi-Square Test

To determine whether the chi-square test is statistically significant, let's look at the following chi-square test output.

Chi-Squared Tests

This is the degrees of freedom for this analysis.

	Value	df	p
χ^2	15.811	3	0.001
N	735		

This is the chi-square statistic for this analysis (denoted χ^2)

This is the p-value for this analysis.

- If p is less than .05, then you will reject the null hypothesis.
- If p is greater than .05, then you will fail to reject the null hypothesis.

To interpret the statistical significance of the chi-square test, we want to look at the p value. If the p value is less than our alpha level of .05, we will reject the null hypothesis (indicating that there is a significant association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino

population). If the p value is greater than our alpha level of .05, then we will fail to reject the null hypothesis (indicating there is not a significant association between our variables).

In this example, you can see from the “Chi-Squared” table that our p value is $< .001$, which is less than the alpha of .05. Therefore, we will reject our null hypothesis – indicating that there is a significant association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population.

Next we will look at the “Nominal” table to report our effect size. Phi is only available when you have a 2×2 contingency table. Because we have a 2×4 contingency table, we will report the Cramer’s V value depicted in the following table.

Nominal	
	Value ^a
Phi-coefficient	NaN
Cramer's V	0.147

These are our effect size measures of Phi and Cramer's V. You will just report one. For Cramer's V, refer to the table for guidelines on the strength of association.

Writing the results of the chi-square test in APA format follows this general format:

$X^2(df) = \text{chi-square statistic}, p < .05 \text{ or } p > .05, \phi = \text{Phi coefficient OR Cramer's V} = \text{Cramer's V value}$

So, plugging in our results into this format should look like this:

$$X^2(3) = 15.81, p < .05, \text{Cramer's V} = .15$$

Interpreting the Practical Significance of the Chi-Square Test

Now that we have interpreted the statistical significance of the chi-square test, let's consider the practical significance of this test. Remember, a significant p value tells us that there is a significant age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population. The effect size, then, tells us how meaningful or strong this association is. That is, the effect size tells us how strongly related our two variables are to one another.

For the purposes of this analysis, because we are interpreting a 2×4 crosstabulation, we will interpret Cramer's V. Let's refer back to the Cramer's V guidelines for strength of association for our degrees of freedom = 3. A small effect is .06, a medium effect is .17, and a large effect is .29. We obtained a value of .15 (*almost* a medium effect), so we would say that the strength of association age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population is small to medium.

Guidelines for Interpreting Cramer's V

	Small Effect Size	Medium Effect Size	Large Effect Size
$df_{RC} = 1$.10	.30	.50
$df_{RC} = 2$.07	.21	.35
$df_{RC} = 3$.06	.17	.29
$df_{RC} = 4$.05	.15	.25
$df_{RC} = 5$.05	.13	.22

In addition to reporting Phi and Cramer's V effect size statistics as a measure of practical significance, odds ratios are also common when making a 2×2 comparison (e.g., you have two dichotomous variables). Because odds ratios can only be calculated for 2×2 contingency tables, we are unable to use that method to gain additional understanding from our current data since we are working with a 2×4 design with four levels of our age variable. However, we could potentially run *planned comparisons*. Planned comparisons are comparisons that need to be chosen *a priori* or before we run our test as part of our initial study hypotheses. For example, given the data shared in the Pew Research Report about awareness of the Latinx term being higher in younger adults, we might plan to compare the 18-29 year group to the 30-49 year old group, the 50-64 year old group, and the 65+ year old group to see if there is a significant association when broken down and compared individually. To do this, we would have to go back to our "Age" variable and filter out the groups we are not currently comparing.

Planned Comparison 1 Filter			Planned Comparison 2 Filter			Planned Comparison 3 Filter		
Age			Age			Age		
Filter	Value	Label	Filter	Value	Label	Filter	Value	Label
<input checked="" type="checkbox"/>	1	18-29	<input checked="" type="checkbox"/>	1	18-29	<input checked="" type="checkbox"/>	1	18-29
<input checked="" type="checkbox"/>	2	30-49	<input type="checkbox"/>	2	30-49	<input type="checkbox"/>	2	30-49
<input type="checkbox"/>	3	50-64	<input checked="" type="checkbox"/>	3	50-64	<input type="checkbox"/>	3	50-64
<input type="checkbox"/>	4	65+	<input type="checkbox"/>	4	65+	<input checked="" type="checkbox"/>	4	65+
<input type="checkbox"/>	99	Refused	<input type="checkbox"/>	99	Refused	<input type="checkbox"/>	99	Refused

We will then look at our Chi-Square results for each of the planned comparisons, illustrated in the following images. The first image includes the Chi-Square test results for our first comparison (18-29-year-old group compared to the 30-49-year-old group). Then we see our second comparison (18-29-year-old group compared to the 50-64-year-old group), and our third comparison (18-29-year-old group compared to the 65+ year old group). Remember, we want a *p* value that is less than .05 in order to conclude there is a significant association. Our probability value is significant, or less than .05, for the first and the third comparison but not for the second.

Chi-Squared Tests 18-29 year olds & 30-49 year olds

	Value	df	p
X ²	5.175	1	0.023
N	542		

p is less than .05, so there is a significant association.

Chi-Squared Tests 18-29 year olds & 50-64 year olds

	Value	df	p
X ²	3.533	1	0.060
N	412		

p is greater than .05, so there is NOT a significant association.

Chi-Squared Tests 18-29 year olds & 65+ year olds

	Value	df	p
X ²	13.129	1	< .001
N	319		

p is less than .05, so there is a significant association.

We will now look at our effect size results for each of the planned comparisons, illustrated in the following images. The first image includes the effect size results for our first comparison (18-29-year-old group compared to the 30-49-year-old group), our second comparison (18-29-year-old group compared to the 50-64-year-old group), and our third comparison (18-29-year-old group compared to the 65+ year old group). The first two comparisons have *small* effect sizes, and the third comparison between the 18-29-year-old group and 65+ year old group is *small to medium*.

Nominal 18-29 year olds & 30-49 year olds

	Value
Contingency coefficient	0.097
Phi-coefficient	0.098
Cramer's V	0.098

A Cramer's V of .09, with *df* of 1, is a *small* effect size.

Nominal 18-29 year olds & 50-64 year olds

	Value
Contingency coefficient	0.092
Phi-coefficient	0.093
Cramer's V	0.093

A Cramer's V of .09, with *df* of 1, is a *small* effect size.

Nominal 18-29 year olds & 65+ year olds

	Value
Contingency coefficient	0.199
Phi-coefficient	0.203
Cramer's V	0.203

A Cramer's V of .20, with *df* of 1, is a *small to medium* effect size.

Looking back at our original contingency table, we would be able to conclude that the young adult group (18-29-year olds) are significantly more likely to believe that Latinx should be used to describe the Hispanic/Latino population in the United States compared to 30-49 year olds and older adults 65 years and older; however, this difference in preferences is relatively *small*.

Reporting in APA Format

What do you need to report in your results?

Value (Notation)	
Chi-Square Test	Chi-Square Value (X^2)
Degrees of Freedom (df)	Probability (p)
Effect Size	Phi (ϕ) or Cramer's V
Odds Ratio	Assumptions
Expected Counts > 5	

Putting it All Together

Participants were current American Trends Panel (ATP) panel members who identified as being Hispanic. A chi-square test was conducted to examine whether there was an association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population. All expected cell frequencies were greater than five. There was a statistically significant association between age groups and whether a person thinks Latinx should be used to describe the Hispanic or Latino population, $X^2(3) = 15.81, p < .05$, Cramer's V = .15. This was a small to medium association. Planned comparisons were completed to compare the 18-29-year-old age group with the 30-49-year-old group, the 50-64-year-old group, and the 65+ year old group. There was a significant association between age group and whether a person thinks Latinx should be used to describe the Hispanic or Latino population for the first planned comparison (18-29-year-old group and 30-49-year-old group), $X^2(1) = 5.18, p < .05$, Cramer's V = .09, a small association. There was also a significant association between age group and whether a person thinks Latinx should be used to describe the Hispanic or Latino population for the third planned comparison (18-29-year-old group and 65+ year old group), $X^2(1) = 13.13, p < .001$, Cramer's V = .20, a small to medium association. However, there was not a significant association between age group and whether a person thinks Latinx should be used to describe the Hispanic or Latino population for the second planned comparison (18-29-year-old group and 50-64-year-old group), $X^2(1) = 3.53, p > .05$, Cramer's V = .09.

Real-World Meaning

Remember, when we're explaining the results of statistical analyses in real-world terms, we want to try our best to explain this to someone who has never taken statistics. We want them to be able to understand the results of the analyses without knowing how to do them. An effective way to do this is to briefly put the statistical test in the context of the research question, explain what the researchers did, and then describe what they found. It can also be helpful to put the results of the analyses back into the context of the original research question and attempt to answer a "so what?" question. If I was telling my friends about the results of these analyses, I would tell them that different age groups feel differently about using the term Latinx to describe the Hispanic/Latino population in the United States. Recent data suggests younger adults are more likely to believe Latinx term should be used compared to older generations.

References

Hugo Lopez, M. (2013). Hispanic Identity. Pew Research Center. Retrieved from <https://www.pewresearch.org/hispanic/2013/10/22/3-hispanic-identity/>

Noe-Bustamante, L., Mora, L., & Hugo Lopez, M. (2020). About one-in-four U.S. Hispanics have heard of Latinx, but just 3% use it. Pew Research Center. Retrieved from <https://www.pewresearch.org/hispanic/2020/08/11/about-one-in-four-u-s-hispanics-have-heard-of-latinx-but-just-3-use-it/>

Authors

This guide was written and created by Ashlyn A. Moraine, Ruth V. Walker, PhD, and Hannah J. Osborn, PhD.

Acknowledgements

We would like to thank the Pew Research Center for generously making their data available for secondary analysis. Additionally, we appreciate Dr. Kristen Black's edits and suggestions during the creation of this guide.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting a Chi-Square Test in JASP Copyright © 2021 by Ashlyn A. Moraine; Dr. Hannah J. Osborn; and Ruth Walker is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

6.

RUNNING AND INTERPRETING A CORRELATION IN JASP

Asia Symone Palmer; Ruth Walker; and Kristen J. Black

Download Data Sets:

- JASP Data File Pearson Correlation
- CSV Data File Pearson Correlation

Researchers Li, Wu, and Xiong (2021) published a study in the open-access journal, PLoS ONE, to understand the relationship between employee levels of cultural intelligence and their innovative behaviors in the workplace. Past researchers have found that an employee's ability to adapt to cross-cultural environments is associated with better job performance in culturally diverse settings (Ang et al., 2006). There has been a lack of research into the relationship between cultural intelligence and sustainable innovation behavior; however, past research in this area has found a relationship between cultural intelligence and employee adaptive work behavior (Pandey & Charoensukmongkol, 2019), employee creativity (Hu et al., 2017), and an organizations' ability to create and adopt innovations (Gölgeci et al., 2017). Li and colleagues (2021) focused on studying the relationship between employee cultural intelligence and sustainable innovation behavior, defined as "The act of generating new ideas and translating those ideas into practice" (p. 4).

The researchers measured Sustainable Innovation Behavior (SIB) by asking participants to answer six questions rated on a 7-point likert scale. Higher scores on this scale indicate higher levels of innovative behavior. A sample item includes, "I often come up with creative ideas."

The researchers measured Cultural Intelligence (CQ) by asking participants to answer 12 questions rated on a 7-point likert scale. Higher scores on this scale indicate higher levels of Cultural Intelligence. A sample item includes, "I vary the rate of my speaking when a cross-cultural situation requires it."

Hypotheses

We will test one of hypotheses Li and colleagues (2021) proposed, namely that there will be a positive relationship between employees' cultural intelligence and sustainable innovation behavior. This means they hypothesized a positive correlation, but for now, we will just hypothesize that correlation exists and conduct a non-directional (two-tailed) test. Two-tailed tests are the default approach in most applications of data analysis, even when we may suspect one direction over another. It is the most conservative approach and allows us to look for results opposite to our expectations.

The null and alternative hypothesis would be:

- H_0 : Employee cultural intelligence will be unrelated to sustainable innovation behavior; or $r = 0$.
- H_1 : Employee cultural intelligence will be significantly related to sustainable innovation behavior; or $r \neq 0$

Assumption Testing

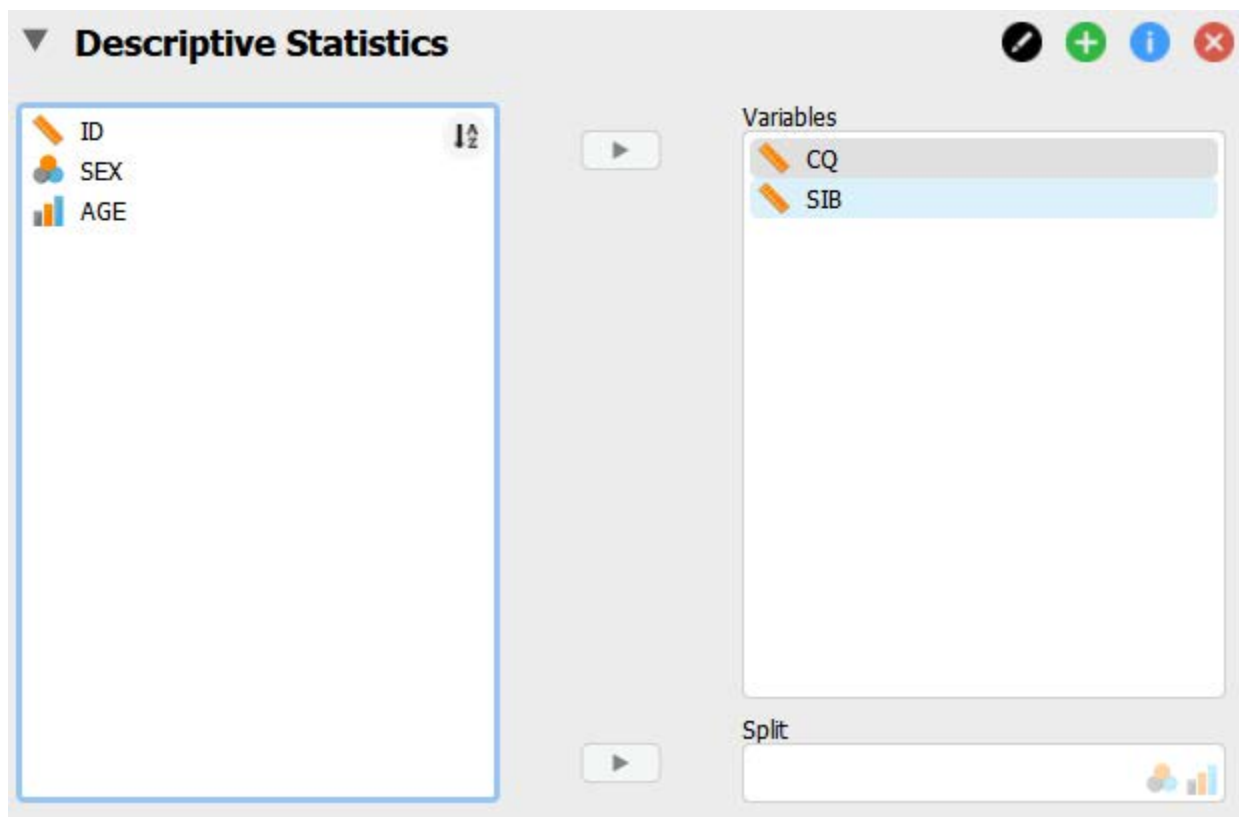
We need to test several assumptions to get started.

Assumption One: Are our variables continuous?

Yes. All our variables – sustainable innovation behavior and cultural intelligence – are numerical values. They use multiple Likert scale items to create a total score on each of those measures. This means these variables are continuous, interval level data. This data is appropriate for analysis with a correlational study. They have also been labeled appropriately in JASP, because we see the ruler icon in the corner of each variable name.

Assumption Two: Are our variables normally distributed?

Let's tackle these next two assumptions with some data visuals and descriptive statistics. We can get these by clicking the Descriptive Statistics menu. Then select the two variables we need for our hypothesis tests (CQ and SIB) and move them over.



To check for normality and outliers, we will look at our options under “Customizable Plots.” Let’s check the boxes for “Boxplots,” “Label outliers.” Under “Basic Plots,” we will check the box “Distribution plots.”

▼ Customizable plots

Color palette Colorblind ▼

☒ Boxplots

☒ Boxplot element
☐ Use color palette

☐ Violin element
☒ Label outliers

☐ Jitter element

▼ Basic plots

☒ Distribution plots
☐ Correlation plots
☐ Interval plots

☐ Display density
☐ Q-Q plots

☐ Display rug marks
☐ Pie charts

Bin width type Sturges ▼

Number of bins 30

☐ Dot plots

Categorical plots

☐ Pareto plots

☐ Pareto rule
95.0 %

☐ Likert plots

☐ Assume all variables share the same levels

Adjustable font size for vertical axis Normal ▼

We will also request skewness and kurtosis numerical estimates in our “Statistics” section of the descriptive statistics menu. Under the “Statistics” section, check the boxes for skewness and kurtosis.

Statistics

Sample size

☒ Valid
 ☒ Missing

Quantiles

☐ Quartiles
 ☐ Cut points for: 4 equal groups
 ☐ Percentiles:

Central tendency

☐ Mode
 ☐ Median
 ☒ Mean

Distribution

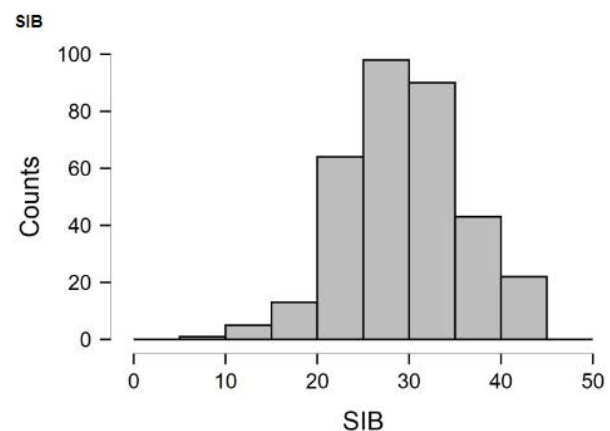
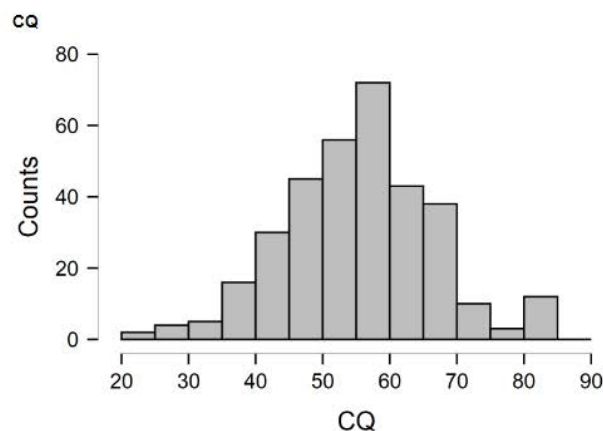
☒ Skewness
 ☒ Kurtosis
 ☐ Shapiro-Wilk test
 ☐ Sum

Dispersion

☐ S.E. mean
 ☒ Std. deviation
 ☐ Coefficient of variation
 ☐ MAD
 ☐ MAD robust
 ☐ IQR
 ☐ Variance
 ☐ Range
 ☒ Minimum
 ☒ Maximum

Looking at our results, the histograms give us a sense of whether our distribution is normal (symmetrical). The histograms for both CQ and SIB look approximately symmetrical, so we're happy with both of those.

Distribution Plots



As another check for normally distributed data, in addition to our histogram, we can interpret the skewness and kurtosis statistics. We want skewness and kurtosis statistics that are between -2 and +2. Remember, we have

two variables, so we have to check the skewness and kurtosis values for both sustainable innovative behavior and cultural intelligence.

Descriptive Statistics

	CQ	SIB
Valid	336	336
Missing	0	0
Mean	56.018	29.970
Std. Deviation	11.330	6.242
Skewness	0.048	-0.221
Std. Error of Skewness	0.133	0.133
Kurtosis	0.309	0.346
Std. Error of Kurtosis	0.265	0.265
Minimum	24.000	6.000
Maximum	84.000	42.000

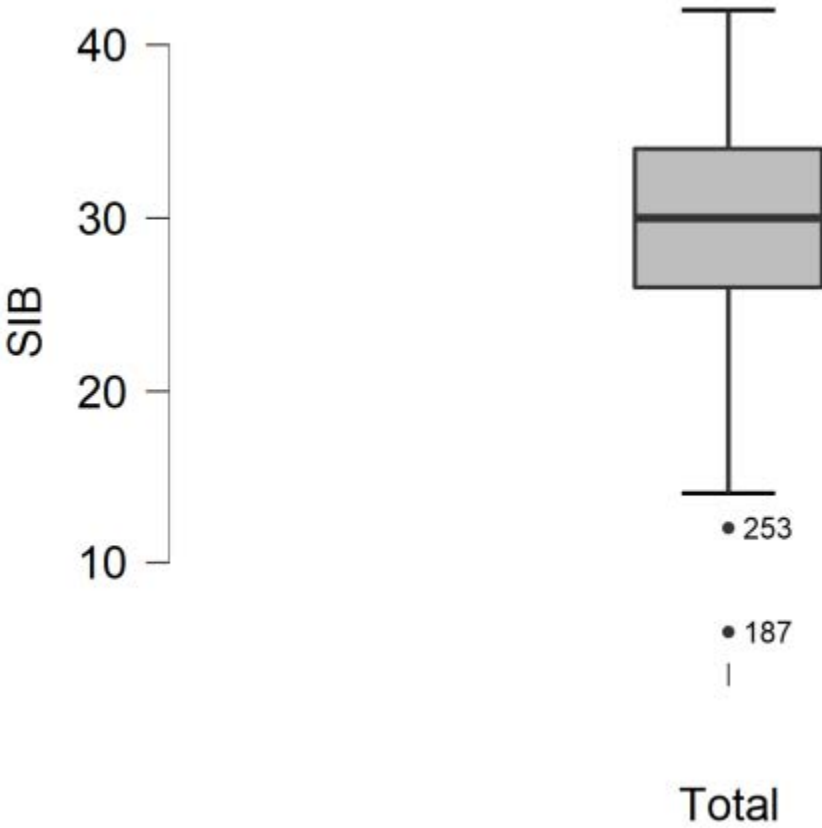
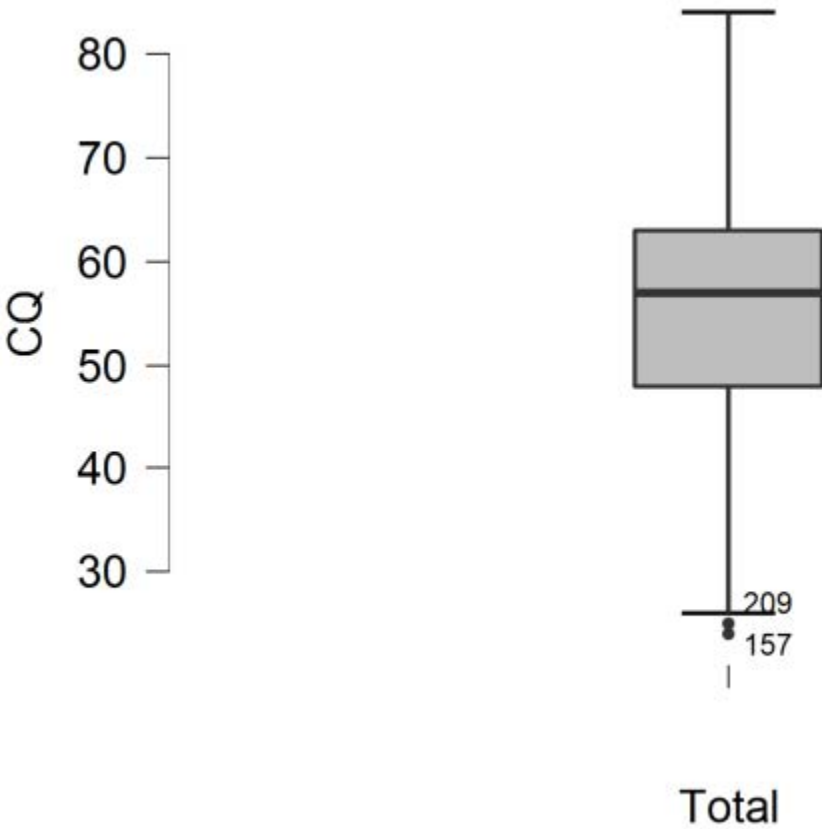
Looking at the values on the output copied above, we can see our skewness values for CQ (0.05) and SIB (-0.22) are both within the acceptable range of -2 and +2. Additionally, the kurtosis values for CQ (0.31) and SIB (0.35) are also within the acceptable range of -2 and +2.

To report these results in APA format, we could write:

CQ and SIB scores were normally distributed, skewness and kurtosis statistics were between -2 and +2.

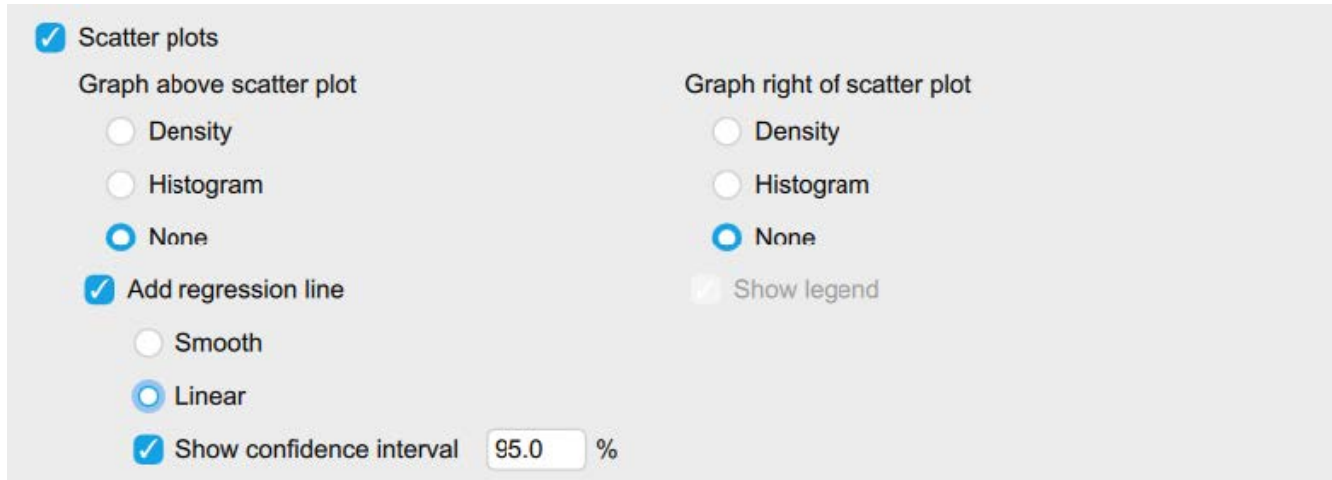
Assumption Three: Are there any outliers?

To check that we meet this assumption, we will look at the boxplots in our results. The boxplots show us that both cultural intelligence (CQ) and substantial innovation behaviors (SIB) have two outliers. However, since our normality statistics look good, we will go ahead and proceed with our analyses.



Assumption Four: Is the relationship between our variables linear?

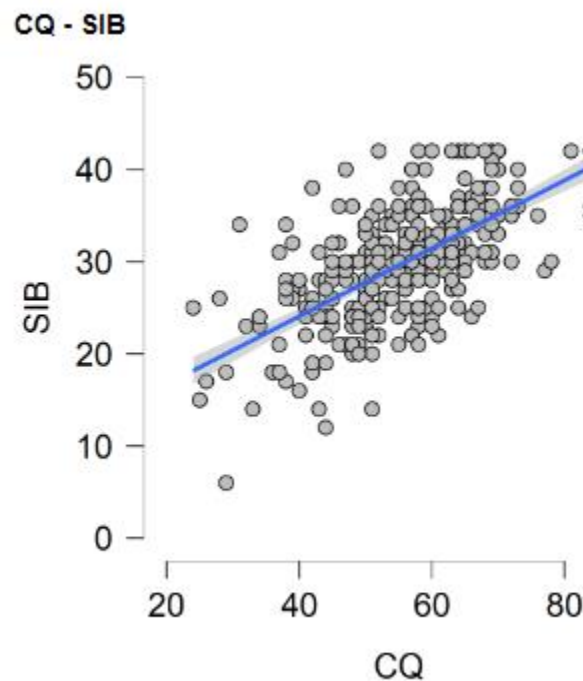
Going back to having our two original variables selected, we will check the box for a scatterplot from the “Customizable plots” menu. You will see I have requested the simplest scatterplot by selecting “none” for the graphs above and to the right (these options would show us the distributions of each variables outside of our primary scatterplot) and requesting a linear regression line, rather than smooth, to identify the overall trend.



The screenshot shows a settings panel for a scatterplot. It is divided into two columns: 'Graph above scatter plot' and 'Graph right of scatter plot'. In the left column, 'Scatter plots' is checked. Under 'Graph above scatter plot', 'None' is selected with a radio button. Under 'Add regression line', which is checked, 'Linear' is selected with a radio button. 'Show confidence interval' is checked, and the value '95.0' is entered in the adjacent text box. In the right column, 'Density', 'Histogram', and 'None' are all unselected. 'Show legend' is checked with a checkbox.

As we look at our scatterplots using the raw variables, it is clear that lots of the data is clustered at the middle of our x axis but are still evenly distributed along the line (because our variables are not skewed), so the pattern of the relationship does appear to follow a rough line. So, we can confirm that the assumption for linearity is ok.

Scatter Plots



Primary Analyses

We are going to start with how to run a basic correlation in JASP. We will find correlation nested under the regression menu.

	ID	SEX	AGE	SIB	CQ
1	1	1	4	32	65
2	2	1	6	35	54
3	3	2	1	36	65

Regression

- Classical
 - Correlation
 - Linear Regression
 - Logistic Regression

We will first move over our two variables that are in our hypotheses. Taking a look at the extra options, here's what we will (and can) specify:

- Additional options: we will check the boxes for “flag significant correlations”, “confidence intervals”, and “sample size”.
- Alternative Hypothesis: Notice, you can specify a positive or negative correlation in the JASP menu, but

between two variables by looking at where the row for that variable intersects with the column for the other variable. At the intersecting point, we see a few pieces of information reported: 1) the sample size used to compute the correlation, 2) the actual Pearson's r value, 3) the associated p -value for interpreting significance, 4) the confidence interval estimate around that value of r . Confidence intervals provide a range of values that the population correlation is likely to fall between. With the example of the correlation between employee cultural intelligence and sustainable innovation behavior, our estimated correlation is $r = .67$, but the range of likely values that the true correlation could be are between .61 and .72. The main pieces of information we typically report, are highlighted in the figure below.

Pearson's Correlations

Variable		SIB	CQ
1. SIB	n	—	—
	Pearson's r	—	—
	p -value	—	—
	Upper 95% CI	—	—
	Lower 95% CI	—	—
2. CQ	n	336	—
	Pearson's r	0.669***	—
	p -value	< .001	—
	Upper 95% CI	0.724	—
	Lower 95% CI	0.605	—

* $p < .05$, ** $p < .01$, *** $p < .001$

- This is the 95% confidence interval for our r value (0.67). This tells us that we can be 95% confident that the true correlation between these variables is somewhere between the lower interval (0.61) and the upper interval (0.72).

- This is the p -value. If the p value is less than .05, the null hypothesis will be rejected. If p is greater than .05, the null hypothesis will fail to be rejected. Based on the data, p is < .001, which is below the alpha level of .05, indicating we will reject the null hypothesis

This is the r statistic for this analysis. The Pearson's r value is 0.67 indicating a strong/significant correlation between CQ and SIB

As we think about interpreting our results in line with our hypotheses, we would say that our hypothesis was supported because the p -values associated with each correlation are smaller than our alpha value of .05. Employee Cultural Intelligence was related to sustainable innovation behavior. The general format for our APA statements for a correlation are:

$$r(df) = r \text{ statistics}, p < .05 \text{ or } > .05$$

$$df = N - 2$$

Note, sometimes r^2 is also included at the end.

Applying this format to our study hypotheses:

- $r(336) = .67, p < .05$

If we wanted to include confidence intervals, we would report it as

- $r(336) = .67, p < .05, 95\% \text{ CI } [0.61, 0.72]$

Interpreting the practical significance

We have general guidelines that tell us our correlation of .67 would be considered a relatively strong correlation. They are over the borderline of our criteria for a strong relationship, which is .50.

Coefficient Value	Strength of Association
0.1 – < 0.3	Small correlation
0.3 – < 0.5	Medium/moderate correlation
> 0.5	Large/strong correlation

We can also square our r values to interpret them in terms of variance shared or explained. Here's how we would summarize those relationships using r^2 .

- Squaring the correlation of .67 indicates that employee cultural intelligence explained 45% of the variance in sustainable innovation behavior.

Putting it all together

Now it is time to summarize our results in APA format. We will start with a basic interpretation, if we were to use just a standard correlation analysis with our raw variables.

What do you need to report in your analyses?

Value (Notation)	Assumptions
Correlations (r or p)	Skewness and Kurtosis
Degrees of Freedom (df)	Histogram/Boxplot
Probability (p)	Scatterplot
Effect Size (r^2)	

Prior to hypothesis testing, assumptions related to outliers, normality, and linearity were checked. There were two outliers in both cultural intelligence and sustainable innovation behavior, as assessed by the inspection of a boxplot; however, all after checking normality statistics and plots, we made the decision to retain all outliers. Cultural intelligence (CQ) and sustainable innovation behavior (SIB) were normally distributed, with

skewness and kurtosis values between -2 and +2. A review of scatterplots showed a linear relationship between QC and SIB.

We conducted a Pearson correlation to see if there is a relationship between cultural intelligence and sustainable innovation behavior. We found there is a significant, positive relationship between cultural intelligence and sustainable innovation behavior, ($r = .67, p < .001, r^2 = .45$).

The r-squared value was 0.45, indicating that 45% of the variance in cultural intelligence is explained by its relationship with sustainable innovation behavior.

Real World Meaning

A correlational analysis is one that is fairly intuitive to share about the real world meaning. Most individuals understand these relational tests pretty easily. But we still want to be careful in applying our results. We know that correlation is *not* causation, so we want to keep that in mind as we think about the real world meaning.

In the case of this study, we would simply say that the higher an employee's cultural intelligence is, the higher their levels of sustainable innovation behaviors are. This could be because cultural intelligence contributes to sustainable innovation behavior, or perhaps there is a third variable that explains this relationship. We cannot be completely sure, but the relationship is strong enough to suggest there is some connection between employee's cultural intelligence and innovation behavior.

References

Gölgeci, I., Swiatowiec-Szczepanska, J., & Raczkowski, K. (2017). How does cultural intelligence influence the relationships between potential and realised absorptive capacity and innovativeness? Evidence from Poland. *Technology Analysis & Strategic Management*, 29(8), 857–871. <https://doi.org/10.1080/09537325.2016.1245858>

Hu, S., Gu, J., Liu, H. & Huang, Q. (2017). The moderating role of social media usage in the relationship among multicultural experiences, cultural intelligence, and individual creativity. *Information Technology & People*, 30(2), pp. 265-281. <https://doi.org/10.1108/ITP-04-2016-0099>

Li, J., Wu, N., & Xiong, S. (2021). Sustainable innovation in the context of organizational cultural diversity: The role of cultural intelligence and knowledge sharing. *PloS One*, 16(5), e0250878–e0250878. <https://doi.org/10.1371/journal.pone.0250878>

Pandey, A. & Charoensukmongkol, P. (2019). Contribution of cultural intelligence to adaptive selling and customer-oriented selling of salespeople at international trade shows: does cultural similarity matter? *Journal of Asia Business Studies*, 13(1), pp. 79-96. <https://doi.org/10.1108/JABS-08-2017-0138>

Authors

This guide was written and created by Asia Symone Palmer, Ruth V. Walker, PhD, Kristen J. Black, PhD, and Erin N. Prince.

Acknowledgements

We would like to thank Jinlong Li, Na Wu, and Shengxu Xiong for generously publishing their data through PloS One.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting a Correlation in JASP Copyright © 2021 by Asia Symone Palmer; Ruth Walker; and Kristen J. Black is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

7.

RUNNING AND INTERPRETING LINEAR REGRESSION IN JASP

Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black

Download Data Sets:

- JASP Data File Regression
- CSV Data File Regression

We are going to use a portion of the data that was used in the study by Li, Wu, and Xiong (2021). This is the same study we used when we conducted a Pearson Correlation in the previous guide and found a strong, positive relationship between employee levels of cultural intelligence (e.g., ability to adapt to cross-cultural environments) and their innovative behaviors in the workplace. For this guide, we will be building upon that analysis by first conducting a simple linear regression to determine if Cultural Intelligence *predicts* innovative behaviors in the workplace. As we learned in the previous guide, Sustainable Innovation Behavior refers to how well an individual creates and enacts new ideas.

The researchers measured Sustainable Innovation Behavior (SIB) by asking participants to answer six questions rated on a 7-point Likert scale. Higher scores on this scale indicate higher levels of Sustainable Innovative Behavior. A sample item includes, “I often come up with creative ideas.”

The researchers measured Cultural Intelligence (CQ) by asking participants to answer 12 questions rated on a 7-point Likert scale. Higher scores on this scale indicate higher levels of Cultural Intelligence. A sample item includes, “I adjust my cultural knowledge as I interact with people from a culture that is unfamiliar to me.”

Now we are ready to try out an analysis of our own. Open up the “Li et al 2021” file. You have both a .csv file you can upload and a JASP file.

Hypotheses

We know there is a relationship between Cultural Intelligence and Sustainable Innovation Behaviors after completing a Pearson correlation; however, now we are going to build on those results to determine if Cultural Intelligence *predicts* Sustainable Innovation Behaviors using a linear regression model.

The null hypothesis is:

- Conceptual H_0 : Cultural Intelligence will not predict Sustainable Innovation Behavior.
- Mathematical H_0 : $b = 0$

The alternative hypothesis is:

- Conceptual H₁: Cultural Intelligence will predict Sustainable Innovation Behavior.
- Mathematical H₁: $b \neq 0$

Linear Regression Equation

$$Y' = bX + a$$

$$\text{Sustainable Innovative Behavior}' = (b \times \text{Cultural Intelligence}) + a$$

Here is a breakdown of the components of the regression question:

- Y' is the predicted value of your outcome or dependent variable (Y)
- b is the slope of the regression line (estimated by regression)
 - May also be written using this symbol instead: β_1
- X is the value of your predictor variable or independent variable
- a is the Y-intercept of the regression line (estimated by regression)
 - May also be written using this symbol instead: β_0

Assumption Testing

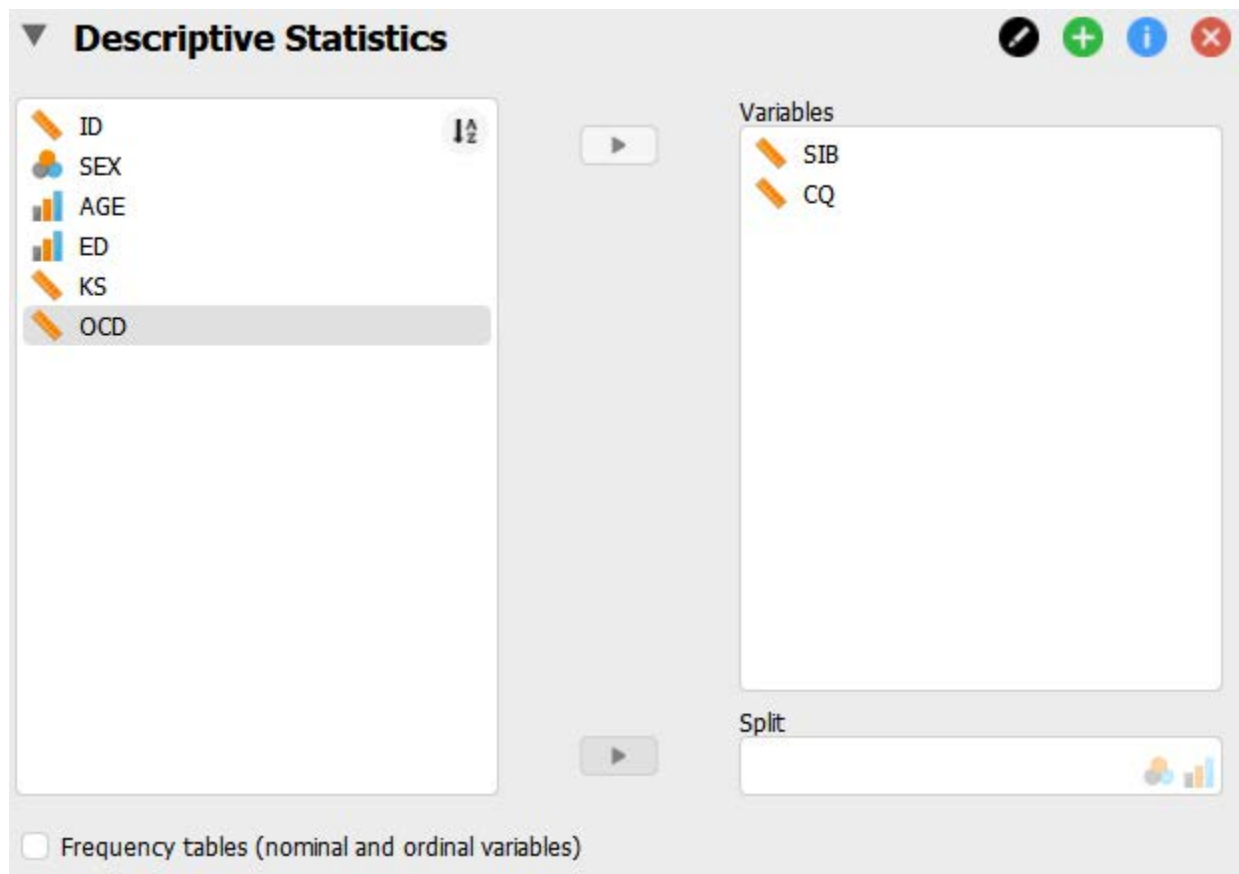
Let's walk through testing all six of our assumptions.

Assumption One: Are our variables continuous?

Yes. Our predictor, Cultural Intelligence is a continuous variable measured using a 12-item Likert scale questionnaire. Our dependent variable is also a continuous variable measured using a 6-item Likert scale questionnaire.

Assumption Two: Are the variables normally distributed?

To continue our assumption testing, click Descriptives. When the "Descriptive Statistics" window pops up, we will need to move the following study variables to the "Variables" box on the right: Cultural Intelligence and Sustainable Innovation Behavior.



In the Statistics drop down menu, check the boxes next to Skewness and Kurtosis under Distribution.

Statistics

Sample size

☒ Valid
 ☒ Missing

Central tendency

☐ Mode
 ☐ Median
 ☒ Mean

Dispersion

☐ S.E. mean
 ☐ Coefficient of variation
 ☐ MAD robust
 ☐ Variance
 ☒ Minimum

Quantiles

☐ Quartiles
 ☐ Cut points for: 4 equal groups
 ☐ Percentiles:

Distribution

☒ Skewness
 ☒ Kurtosis
 ☐ Shapiro-Wilk test
 ☐ Sum

☒ Std. deviation
 ☐ MAD
 ☐ IQR
 ☐ Range
 ☒ Maximum

Under Basic Plots select Distribution Plots. In the Customizable Plots drop down menu you want to select several options for the next set of assumptions.

1. Under Boxplots select Label outliers
2. Under Scatterplots
 1. Select None under Graph above scatter plot
 2. Select None under Graph right of scatter plot
 3. Select Add regression line
 1. Under Add regression line select Linear and Show confidence interval 95.0%

▼ Basic plots

☒ Distribution plots
☐ Correlation plots
☐ Interval plots

☐ Display density
☐ Display rug marks
☐ Q-Q plots
☐ Pie charts

Bin width type Sturges
☐ Dot plots

Number of bins 30

▼ Customizable plots

Color palette Colorblind

☒ Boxplots

☒ Boxplot element
☐ Use color palette
☐ Violin element
☒ Label outliers
☐ Jitter element

☒ Scatter plots

Graph above scatter plot
Graph right of scatter plot

☐ Density
☐ Density
☐ Histogram
☐ Histogram
☒ None
☒ None

☒ Add regression line
☒ Show legend

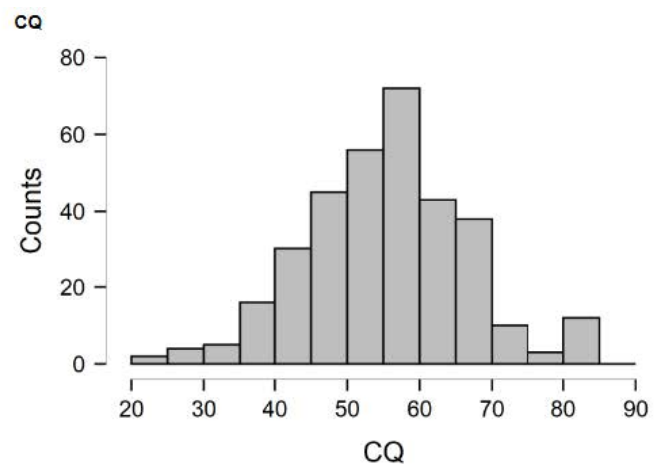
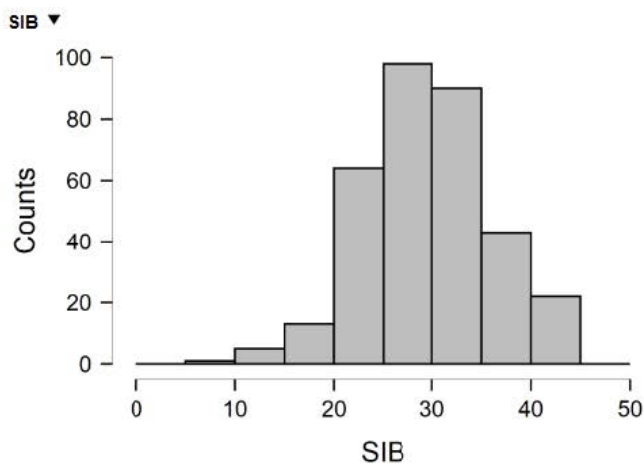
☐ Smooth
☒ Linear

☒ Show confidence interval 95.0 %

Now let's look at our results. We want histograms that look approximately symmetrical as well as skewness and kurtosis statistics between -2 and +2 for each of our study variables.

Descriptive Statistics

	SIB	CQ
Valid	336	336
Missing	0	0
Mean	29.970	56.018
Std. Deviation	6.242	11.330
Skewness	-0.221	0.048
Std. Error of Skewness	0.133	0.133
Kurtosis	0.346	0.309
Std. Error of Kurtosis	0.265	0.265
Minimum	6.000	24.000
Maximum	42.000	84.000



Looking at our results in the preceding images, we see that our skewness and kurtosis values are all within the acceptable range of -2 to +2. Our histograms also confirm that our variables are relatively normally distributed, with approximately symmetrical distributions; however, the way the histograms for each of the variables have

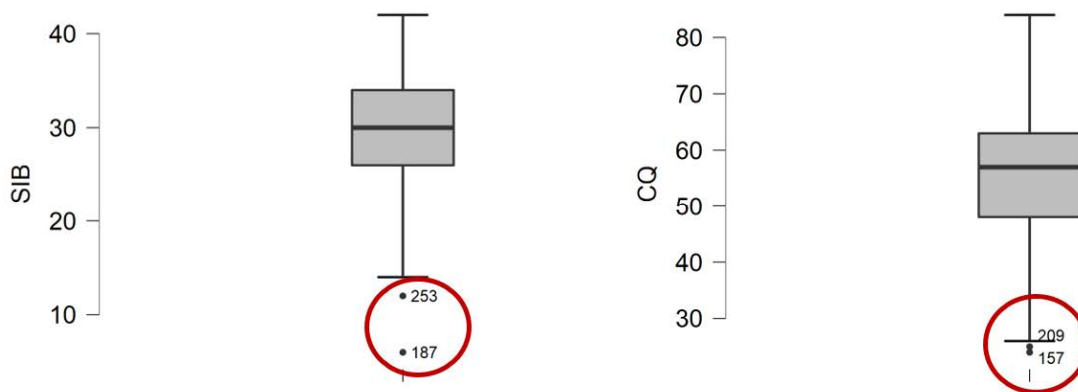
a slight tail to the left indicates we may have outliers present in the bottom quartiles of these distributions. To confirm, let's move to our next assumption.

To report these results in APA format, we can write:

The scores on both of our study variables were normally distributed, with skewness and kurtosis statistics within the acceptable range of -2 and +2.

Assumption Three: Are there any outliers?

To determine if there are any outliers within any of our study variables, we will look at the boxplot output in the following image. As suspected from the preceding histograms, there are outliers present in the bottom quartiles of both variables. There are two outliers in the Sustainable Innovative Behaviors variable and two outliers present in the Cultural Intelligence variable. As we have discussed before, researchers have to decide whether they are going to keep or delete outliers from their analyses. We will complete our analyses with our outliers for now, but it is good practice to rerun analyses with outliers deleted afterwards to determine if the results of your analyses would be different.



To report this using APA format, we would write:

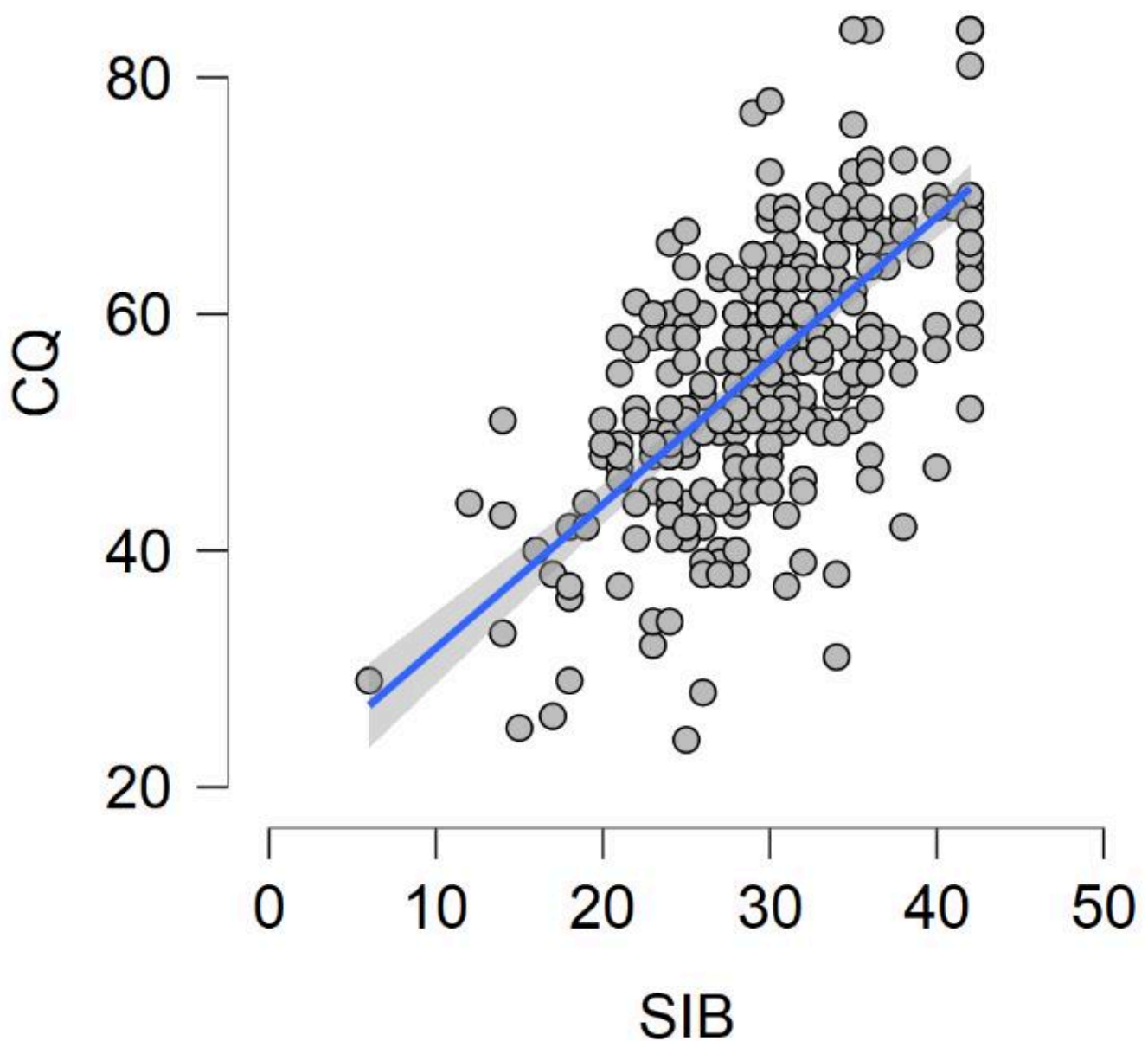
There were outliers identified in the bottom quartile of the cultural intelligence and sustainable innovative behaviors distributions, as assessed by an inspection of boxplots. There were two outliers in the sustainable innovative behaviors distribution and two outliers present in the cultural intelligence distribution.

Assumption Four: Are the relationships of interest linear?

Now we need to look at the scatterplot we requested in our output. Looking at the following scatterplot, our predictor variable (Cultural Intelligence) seems to have a linear relationship with our outcome variable, Sustainable Innovation Behavior.

To report this using APA format, we would write:

There is a linear relationship between cultural intelligence and sustainable innovative behavior, as visualized using a scatterplot.



Assumption Five: Are there concerns with heteroscedasticity?

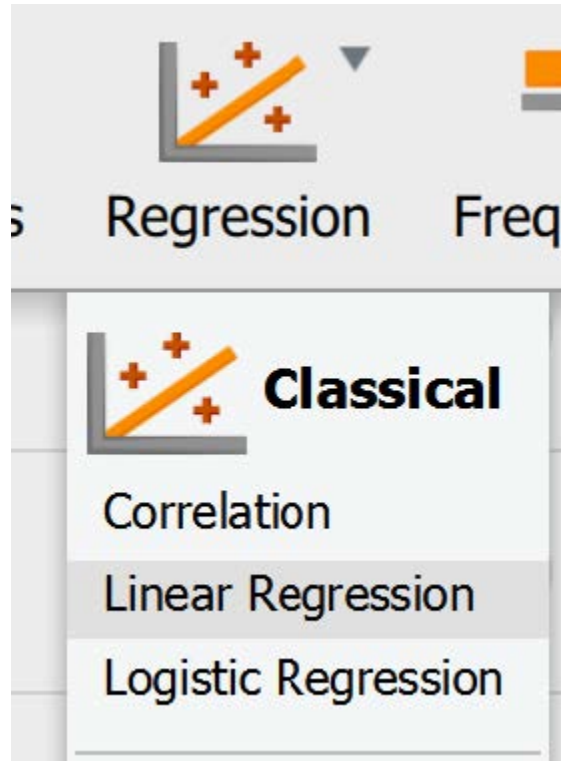
Looking back at the preceding scatterplot, there is not a concern for any cone-shaped distributions. Though there is certainly variability along our best-fit line, there is not one area of the line that has drastically different levels of variability than another. In other words, our error variances or residuals seem to be relatively similar across all values of our predictors.

To report this using APA format, we would write:

There are no concerns regarding heteroscedasticity, after a visual inspection of a scatterplot.

Primary Analyses

Now we are ready to run our linear regression. We can find the regression analysis with the regression tab, which we also used for correlations. We will select the option for linear regression.



In the main regression menu, we will move over Sustainable Innovation Behavior as our dependent variable and Cultural Intelligence as our covariate (or predictor variable).

Linear Regression

Variables: ID, SEX, AGE, ED, **KS**, OCD

Dependent Variable: SIB

Method: Enter

Covariates: CQ

Factors:

WLS Weights (optional):

There is nothing we need to adjust in the “model” drop down, so we will move on to “statistics”. In addition to the default settings, we will click:

- “Confidence intervals” under regression coefficients to get a range of likely slope values.
- “Statistics” under residuals to look at the typical amount of error in our model.
- “Descriptives” in the right column of options, just to look at typical values in our dataset.
- “Durbin-Watson” to check assumption six, that we have independence of errors.

▼ Statistics

Regression Coefficients

☒ Estimates ☐ From 5000 bootstraps ☒ Model fit

☒ Confidence intervals 95.0 % ☐ R squared change

☐ Covariance matrix ☒ Descriptives

☐ Vovk-Sellke maximum p-ratio ☐ Part and partial correlations

☐ Collinearity diagnostics

Residuals

☒ Statistics

☒ Durbin-Watson

☐ Casewise diagnostics

☐ Standard residual > 3

☐ Cook's distance > 1

☐ All

We will also skip over the model specification dropdown and move onto plots. In the plot dropdown, we will click on Residuals vs. histogram to request residuals plotted as a histogram.

► Method Specification

▼ Plots

Residuals Plots

☐ Residuals vs. dependent

☐ Residuals vs. covariates

☐ Residuals vs. predicted

☒ Residuals vs. histogram

☒ Standardized residuals

☐ Q-Q plot standardized residuals

☐ Partial plots

☐ Confidence intervals 95.0 %

☐ Prediction intervals 95.0 %

Now we're ready to look at our output! First let's begin by checking our remaining assumption.

Assumption Five: Do we have independence of errors?

One important consideration in regression analysis is whether the residuals (the differences between the actual values and predicted values) are correlated with each other or not. The Durbin-Watson statistic is a measure of this correlation. When the Durbin-Watson statistic is between 1.5 and 2.5, it suggests that there is not a significant correlation between the residuals or errors in our model. This is important because if there is a high degree of correlation between residuals, it can suggest that the model is not capturing all the relevant information about the relationship between the variables. In other words, the model may not be a good fit for the data. A Durbin-Watson statistic outside of the 1.5-2.5 range may indicate that there is a problem with the model, and further investigation may be needed to determine the cause of the correlation between residuals.

The Durbin-Watson statistic is a part of the Model Summary results. If this value is between 1.5-2.5, we are not concerned there is a significant amount of correlation between the residuals.

Model Summary - SIB ▼

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	6.242	0.060	1.871	0.236
H ₁	0.669	0.448	0.446	4.646	0.028	1.920	0.461

This is our Durbin-Watson statistic. We want this value to be between 1.5-2.5.

Looking at the preceding figure, we can determine that our Durbin-Watson statistic is within the acceptable range of 1.5-2.5. To report this in APA format:

The model has acceptable independence of errors, Durbin-Watson = 1.92.

Now that we have finished checking our assumptions, we can move onto the main results displayed in the first three tables. Notice the tables have a line for H₀ and H₁. This estimates the variance in the outcome with no predictors (H₀) and then adds a predictor in line with the alternative hypothesis (H₁). We will be focusing on H₁ output.

The model summary table tells us about the variance explained by our predictor variable, Cultural Intelligence. Specifically, we want to look at the R and R² values. The R value is a correlation coefficient that tells us that Cultural Intelligence has a strong, positive correlation with Sustainable Innovation Behavior ($r = .669$). Looking at our effect size, r^2 , we can also say that Cultural Intelligence explains 44.8% of the variance in Sustainable Innovation Behavior.

Another way of saying that is that 44.8% of the variance in Sustainable Innovation Behavior can be accounted for by Cultural Intelligence.

Model Summary - SIB ▼

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	6.242	0.060	1.871	0.236
H ₁	0.669	0.448	0.446	4.646	0.028	1.920	0.461

Cultural Intelligence is correlated with Sustainable Innovation Behavior at .669.

This is our r^2 value. Cultural Intelligence explains 44.8% of the variance in Sustainable Innovation Behavior.

The next part of the results we will be looking at is the ANOVA table. The ANOVA table tells us if our best-fit line explained a big enough portion of variance in the relationship between our variables to be statistically significant. For the variables we have in our model, it is a significant F value because our probability value is smaller than .05 ($F = 270.69, p < .001$).

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	5843.461	1	5843.461	270.687	< .001
	Residual	7210.242	334	21.588		
	Total	13053.702	335			

Note. The intercept model is omitted, as no meaningful information can be shown.

df 1

df 2

F statistic

p value. Because p is $< .05$ (our alpha level), this tells us our model is statistically significant.

To report this using APA format, we would write:

Cultural Intelligence significantly predicts Sustainable Innovation Behavior, $F(1, 334) = 270.69, p < .001, r^2 = .45$. Cultural Intelligence explains 44.8% of the variance in Sustainable Innovation Behavior.

Typically, we care more about the individual regression coefficients (slopes) than the overall model significance. To find that information, we move to the last table, labeled coefficients. In the coefficients table, the first column, labeled unstandardized, provides the intercept in the first row of the H₁ output and the slope for Cultural Intelligence in the second row of output. We see that our intercept is 9.31, meaning that if a person scored 0 on CQ, the average Sustainable Innovation Behavior score of that person would be around 9.31. Our slope is .37, meaning that for every one-unit increase in Cultural Intelligence, people are expected to score .37 more on Sustainable Innovation Behavior.

Putting this together, we could write out our regression equation:

$$\text{Sustainable Innovation Behavior} = .37 (\text{Cultural Intelligence score}) + 9.31$$

We are also given an estimate of error around our slope and intercept in the second column. This error term is used to compute the t value. For example, if you divide the slope by the standard error ($-.369 / .022$) you would get our t value of 16.45. Our t value of 16.45 has an associated p value of $< .001$. Because this is less than our alpha level ($.05$), this tells us that the slope is significantly different from zero. This means there is a significant relationship between Cultural Intelligence and Sustainable Innovation Behavior, $b = .37$, $SE = .02$, $t(334) = 16.45$, $p < .001$. [Note: the df for t is the same as the degree of freedom reported in the residual row on the omnibus ANOVA table.]

The standardized column shows us Beta, the standardized slope. This value is identical to the correlation in a model with just one variable.

Model	Coefficients	Unstandardized	Standard Error	Standardized	t	p	95% CI	
							Lower	Upper
H ₀	(Intercept)	29.970	0.341		88.007	< .001	29.300	30.640
H ₁	(Intercept)	9.321	1.280		7.279	< .001	6.802	11.840
	CQ	0.369	0.022	0.669	16.453	< .001	0.325	0.413

a or the Y -intercept of the regression line (9.32).

b or the slope of the regression line (.37).

SE or the standard error.

t statistic and p value. If the associated p value is less than .05, then we can conclude our slope is significantly different from zero.

Confidence intervals. This gives us the likely range of slope values.

Some additional pieces of output to notice are the standardized estimate and the confidence intervals. The standardized column gives us beta, which is the standardized slope. These values range from 0 to 1, just like a correlation. You may have noticed the standardized slope is the same as our R value in the Model Summary. These values will be the same if you only have one predictor variable. The confidence interval tells us a range of likely slope values within the population. If we sampled 100 times, 95 of those times, we would expect a slope estimate between .33 and .41.

The next couple of tables can provide some helpful context. Our descriptive statistics table tells us about typical levels of Cultural Intelligence and Sustainable Innovation Behavior in our sample. For instance, the average Sustainable Innovation Behavior in the whole sample was 29.9.

Descriptives ▼

	N	Mean	SD	SE
SIB	336	29.970	6.242	0.341
CQ	336	56.018	11.330	0.618

The residuals table tells us about the typical levels of discrepancy between the expected value of Sustainable Innovation Behavior (Y) given the Cultural Intelligence score, and the actual value reported in our data. We can see that our average residual is very small and there's a range of prediction error, with the maximum

underestimate (shown as the minimum) as -14.12 and our biggest overestimate, being off by about 13.51. The average standardized residual value tells us that we were often off by a very small amount, technically .000046 standard deviations. This information can be informative, but a plot is more helpful for viewing our residuals.

Prediction interval (\hat{Y}) gives us the range of the Y scores we would predict in our sample, given our regression equation.

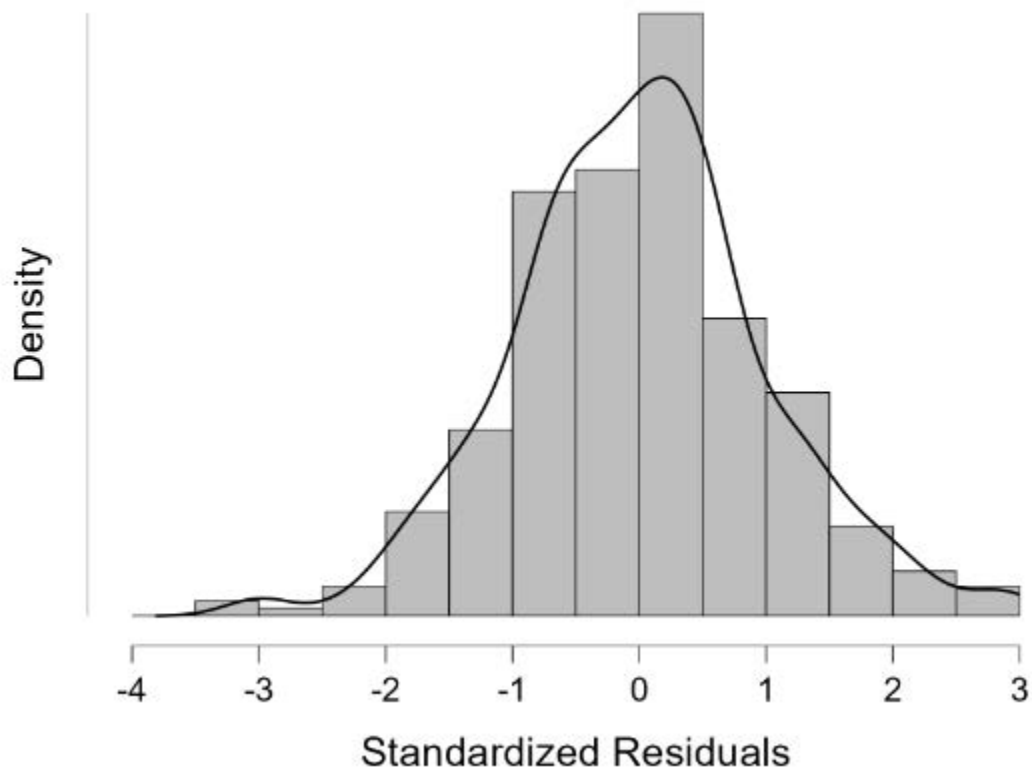
Residual interval ($Y - \hat{Y}$) gives us the range of how far our predicted values deviate from the actual participant scores.

Residuals Statistics	Minimum	Maximum	Mean	SD	N
Predicted Value	18.168	40.285	29.970	4.177	336
Residual	-14.121	13.511	-2.784e -17	4.639	336
Std. Predicted Value	-2.826	2.470	2.323e -16	1.000	336
Std. Residual	-3.046	2.913	4.616e -5	1.002	336

Average standardized residual. This tells us how much our predicted values deviated, on average, from the actual values across participants.

When you see numbers in your output with an “e” followed by “-#”, to convert it to decimal form, move the decimal over the number of times indicated after the minus sign. For example, “4.616e -5” would convert to .00004616.”

Our plot of the standardized residuals (differences between predicted Y and actual Y score) show a relatively normal distribution. This visual, helps us to be confident that are residuals are normally distributed, which is one of the assumptions we need to be true as we examine our prediction error.

Standardized Residuals Histogram

Putting it all Together

What do you need to report in your results?

Value (Notation)	Assumptions
slope (b)	Skewness and Kurtosis
standard error of slope (SE)	Histogram or Boxplot
t value (t)	Scatterplot
Degrees of Freedom (df)	VIF/Tolerance
Probability (p)	Plot of residuals
Effect Size (r^2 or R^2)	

It's time to summarize the results of our regression analysis in APA format.

Prior to conducting a simple linear regression, we tested several assumptions. We examined our predictor (Cultural Intelligence) and outcome (Sustainable Innovation Behavior) variables for normality and outliers. Skewness and kurtosis values were all in the normal range and visualizing the variables with boxplots and histograms revealed four outliers that were retained for analysis. Using a scatterplot, we also confirmed a linear

relationship between the variables with no evidence of heteroscedasticity. The residual histogram showed the residuals were approximately normally distributed.

For our primary analyses, we examined how Cultural Intelligence predicted Sustainable Innovation Behavior. The overall model was significant, $F(1, 334) = 270.69$, $p < .001$, $r^2 = .45$. There was a significant relationship between Cultural Intelligence and Sustainable Innovation Behavior, $b = .37$, $SE = .02$, $t(336) = 16.45$, $p < .001$. Our slope is .37, meaning that for every one-unit increase in Cultural Intelligence, people are expected to score .37 higher on Sustainable Innovation Behavior. Cultural Intelligence explained 44.8% of the variance in Sustainable Innovation Behavior.

References

Gölgeci, I., Swiatowiec-Szczepanska, J., & Raczkowski, K. (2017). How does cultural intelligence influence the relationships between potential and realised absorptive capacity and innovativeness? Evidence from Poland. *Technology Analysis & Strategic Management*, 29(8), 857–871. <https://doi.org/10.1080/09537325.2016.1245858>

Hu, S., Gu, J., Liu, H. & Huang, Q. (2017). The moderating role of social media usage in the relationship among multicultural experiences, cultural intelligence, and individual creativity. *Information Technology & People*, 30(2), pp. 265-281. <https://doi.org/10.1108/ITP-04-2016-0099>

Li, J., Wu, N., & Xiong, S. (2021). Sustainable innovation in the context of organizational cultural diversity: The role of cultural intelligence and knowledge sharing. *PloS One*, 16(5), e0250878–e0250878. <https://doi.org/10.1371/journal.pone.0250878>

Pandey, A. & Charoensukmongkol, P. (2019). Contribution of cultural intelligence to adaptive selling and customer-oriented selling of salespeople at international trade shows: does cultural similarity matter? *Journal of Asia Business Studies*, 13(1), pp. 79-96. <https://doi.org/10.1108/JABS-08-2017-0138>

Authors

This guide was written and created by Ruth V. Walker, PhD, Ashlyn A. Moraine, and Kristen J. Black, PhD.

Acknowledgements

We would like to thank Jinlong Li, Na Wu, and Shengxu Xiong for generously publishing their data through PloS One.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.





Running and Interpreting Linear Regression in JASP Copyright © 2021 by Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

8.

RUNNING AND INTERPRETING MULTIPLE REGRESSION IN JASP

Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black

Download Data Sets:

- JASP Data File Regression
- CSV Data File Regression

Now that we have the basics down, let's add additional predictors used in Li and colleagues (2021) original model to conduct a multiple regression. To do this, we will be conducting a multiple regression to examine how and if Cultural Intelligence, Organizational Culture Differences, and Knowledge Sharing have an impact on our dependent variable, Sustainable Innovation Behavior. Why are we adding additional variables? Most of the time our outcomes are affected by more than just one thing, so we can be more accurate in predicting if we include more than one factor that may be relevant. Organizational Culture Differences refers to a company's corporate culture and measures three aspects: values, systems, and management behaviors. Knowledge Sharing refers to an individual's willingness to share their cultural knowledge with others in the organization.

The researchers measured Organizational Culture Behavior (OCB) by asking participants to answer 5 questions rated on a 7-point Likert scale. Higher scores on this scale indicate higher levels of Organizational Culture Behavior. A sample item includes, "Organizations uses consensus seeking rather than authoritarian decision making."

The researchers measured Knowledge Sharing (KS) by asking participants to answer 4 questions rated on a 7-point Likert scale. Higher scores on this scale indicate higher levels of Knowledge Sharing. A sample item includes, "The more knowledgeable members will provide the other members with knowledge or skills that are difficult to acquire for free."

Hypotheses

Our hypothesis concerns whether we can *predict* Sustainable Innovation Behavior using three separate predictor variables: Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences.

The null hypothesis is:

- Conceptual H_0 : Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences will not predict Sustainable Innovation Behavior.

- Mathematical $H_0: b = 0$

The alternative hypothesis is:

- Conceptual H_1 : Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences will predict Sustainable Innovation Behavior.
- Mathematical $H_1: b \neq 0$

Assumption Testing

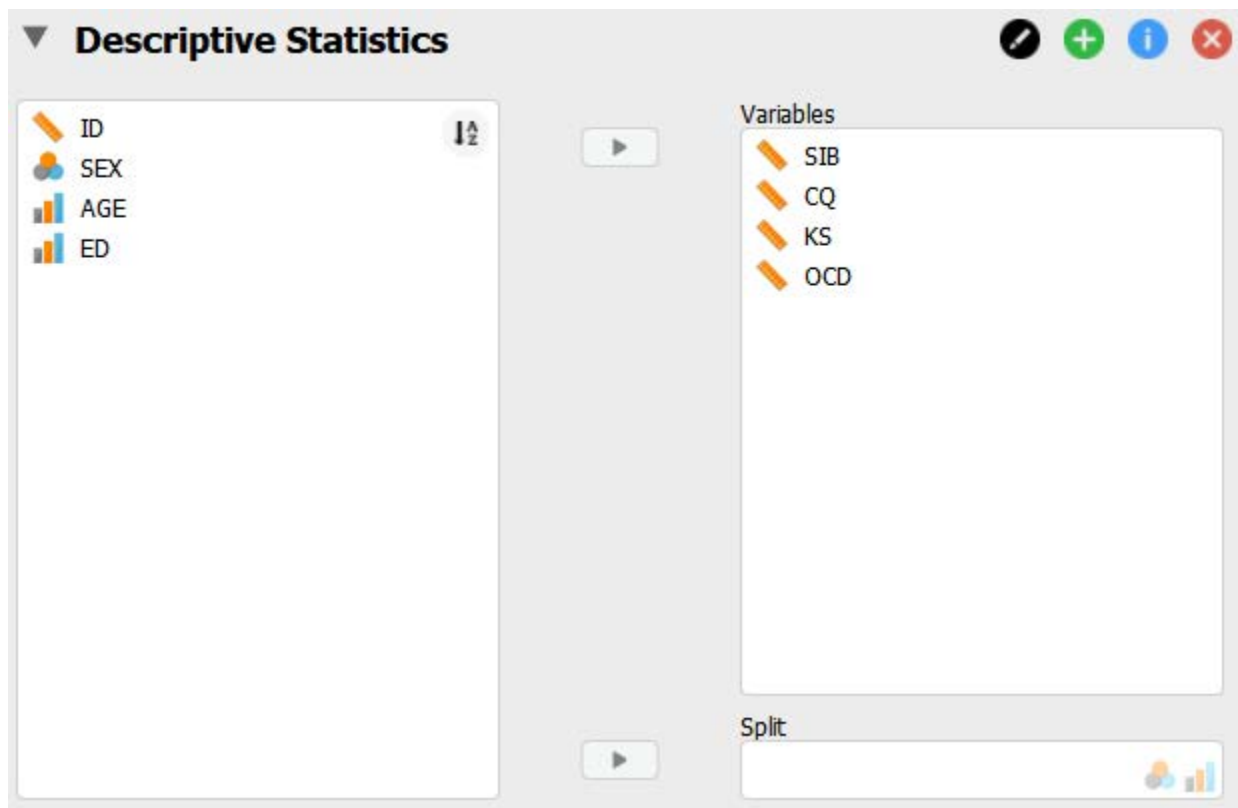
Let's walk through testing all six of our assumptions.

Assumption One: Are our variables continuous?

Yes. Our predictors, Cultural Intelligence, Knowledge Sharing, and Organization Culture are all continuous variables measured using multiple item Likert scale questionnaires. Our outcome of Sustainable Innovation Behavior is also continuous.

Assumption Two: Are the variables normally distributed?

To continue our assumption testing, click Descriptives. When the “Descriptive Statistics” window pops up, we will need to move the following study variables to the “Variables” box on the right: Cultural Intelligence, Knowledge Sharing, Organizational Culture Differences, and Sustainable Innovation Behavior.



In the Statistics drop down menu, check the boxes next to Skewness and Kurtosis under Distribution.

▼ Statistics

Sample size
☒ Valid
 ☒ Missing

Quantiles
☐ Quartiles
 ☐ Cut points for: 4 equal groups
 ☐ Percentiles:

Central tendency
☐ Mode
 ☐ Median
 ☒ Mean

Distribution
☒ Skewness
 ☒ Kurtosis
 ☐ Shapiro-Wilk test
 ☐ Sum

Dispersion
☐ S.E. mean
 ☒ Std. deviation
 ☐ Coefficient of variation
 ☐ MAD
 ☐ MAD robust
 ☐ IQR
 ☐ Variance
 ☐ Range
 ☒ Minimum
 ☒ Maximum

Under Basic Plots select Distribution Plots. In the Plots drop down menu you want to select several options for the next set of assumptions.

1. Under Boxplots select Label outliers
2. Under Scatterplots
 1. Select None under Graph above scatter plot
 2. Select None under Graph right of scatter plot
 3. Select Add regression line
 1. Under Add regression line select Linear and Show confidence interval 95.0%

▼ Basic plots

☒ Distribution plots
 ☐ Correlation plots
 ☐ Interval plots

☐ Display density
 ☐ Q-Q plots

☐ Display rug marks
 ☐ Pie charts

☐ Dot plots

Bin width type Sturges

Number of bins 30

Customizable plots

Color palette
Colorblind

☒ Boxplots

☒ Boxplot element
☐ Use color palette
☐ Violin element
☒ Label outliers
☐ Jitter element

☒ Scatter plots

Graph above scatter plot

☐ Density
☐ Histogram
☒ None

☒ Add regression line

☐ Smooth
☒ Linear

☒ Show confidence interval
95.0
%

Graph right of scatter plot

☐ Density
☐ Histogram
☒ None

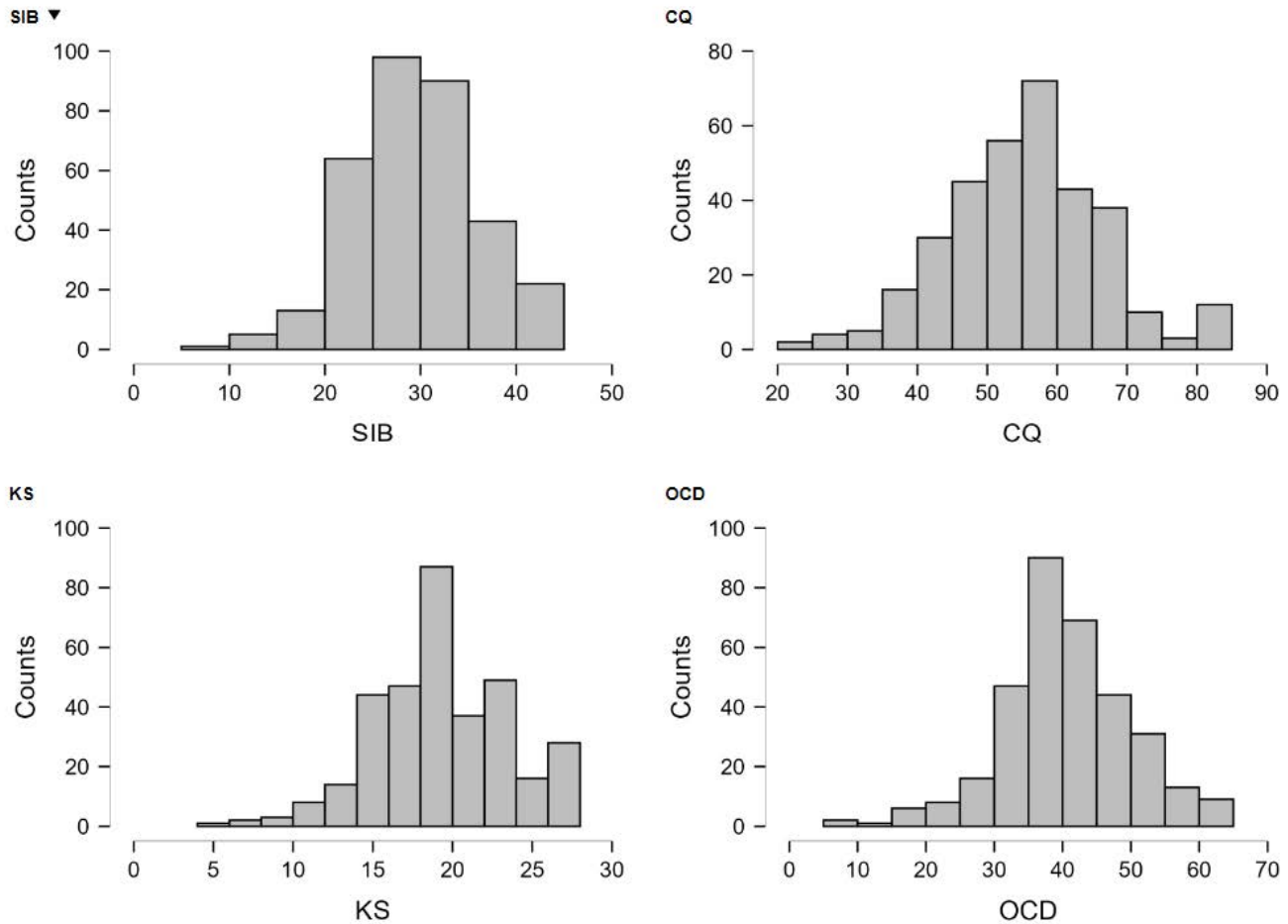
☐ Show legend

Now let's look at our results. We want histograms that look approximately symmetrical as well as skewness and kurtosis statistics between -2 and +2 for each of our study variables.

Descriptive Statistics

	SIB	CQ	KS	OCD
Valid	336	336	336	336
Missing	0	0	0	0
Mean	29.970	56.018	19.964	41.074
Std. Deviation	6.242	11.330	4.235	9.452
Skewness	-0.221	0.048	-0.148	-0.149
Std. Error of Skewness	0.133	0.133	0.133	0.133
Kurtosis	0.346	0.309	0.237	0.515
Std. Error of Kurtosis	0.265	0.265	0.265	0.265
Minimum	6.000	24.000	4.000	9.000
Maximum	42.000	84.000	28.000	63.000

Distribution Plots ▼



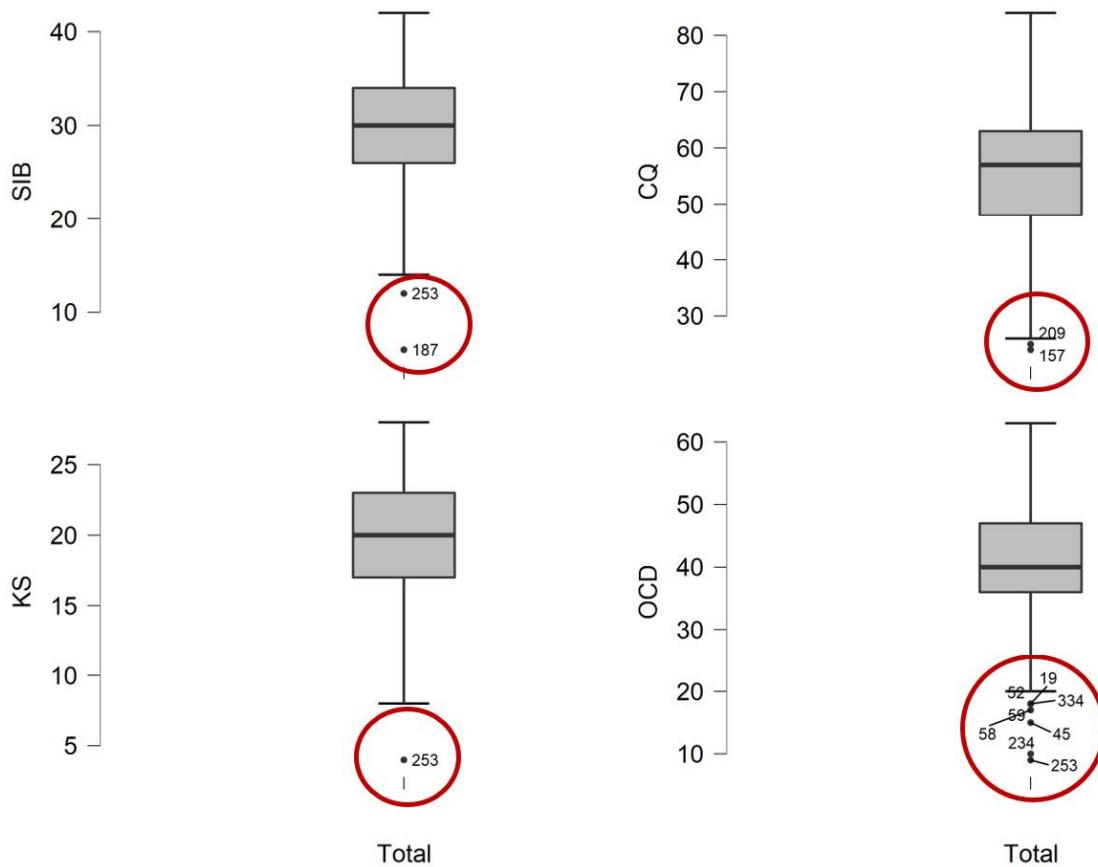
Looking at our results in the preceding images, we see that our skewness and kurtosis values are all within the acceptable range of -2 to +2. Our histograms also confirm that our variables are relatively normally distributed, with approximately symmetrical distributions; however, as we pointed out with the simple regression, the way the histograms for each of the variables have a slight tail to the left indicates we may have outliers present in the bottom quartiles of these distributions. To confirm, let's move to our next assumption.

To report these results in APA format, we can write:

The scores on all four of our study variables were normally distributed, with skewness and kurtosis statistics within the acceptable range of -2 and +2.

Assumption Three: Are there any outliers?

To determine if there are any outliers within any of our study variables, we will look at the boxplot output in the following image. As suspected from the preceding histograms, there are outliers present in the bottom quartiles of each of the four variables. There are two outliers in the Sustainable Innovative Behaviors variable, two outliers present in the Cultural Intelligence variable, one outlier in the Knowledge Sharing variable, and several outliers present in the Organizational Cultural Differences variable.



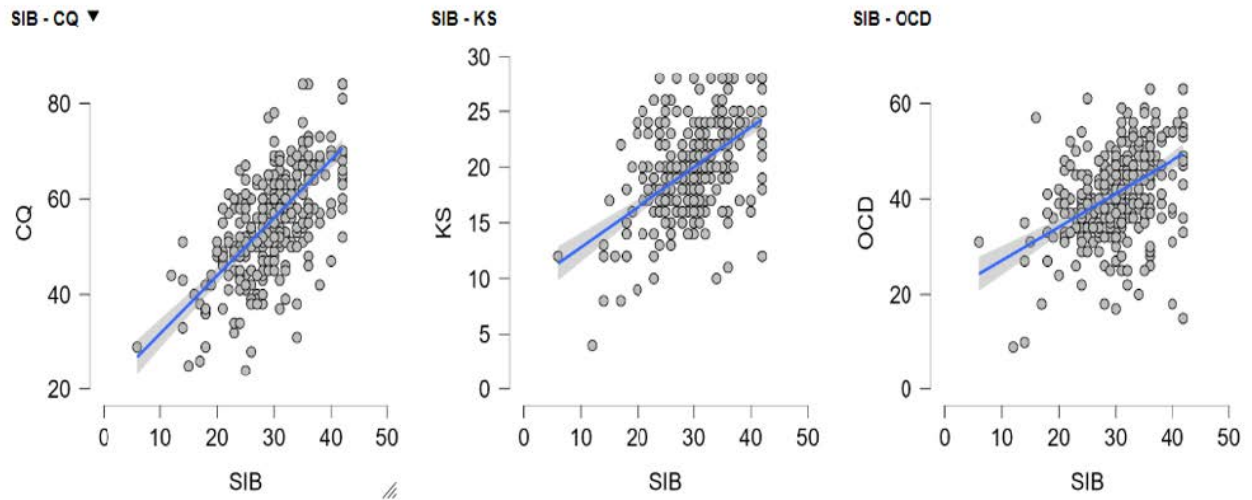
To report this using APA format, we would write:

There were outliers identified across all four study variables in the bottom quartile of the distributions. There were two outliers in the Sustainable Innovation Behavior distribution, two outliers present in the Cultural Intelligence distribution, one outlier in the Knowledge Sharing distribution, and several outliers present in the Organizational Culture Differences distribution.

Assumption Four: Are the relationships of interest linear?

Now we need to look at the scatterplots we requested in our output. Looking at our scatterplots, all of our predictors (Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences) seem to have a linear relationship with our outcome variable, Sustainable Innovation Behavior.

Scatter Plots ▼

***Assumption Five: Are there concerns with heteroscedasticity?***

Again, looking at our scatterplots, there is not a big concern for any cone-shaped distributions. Though there is certainly variability along our best-fit line, there is not one area of the line that has drastically different levels of variability than another. In other words, our error variances or residuals seem to be relatively similar across all values of our predictors.

Related, we will also take a closer look at our residuals to see if they are normally distributed when we get to our actual regression analysis and output.

Assumption Six: (For the multiple regression only) Are our predictor variables highly correlated?

If our predictor variables are highly correlated with each other, this can create problems in interpreting the results of the analysis. Collinearity is the term used to describe this issue of high correlation between predictor variables. One way to check for collinearity is by looking at the correlations between the variables. Another method is by checking the Variance Inflation Factor (VIF) and Tolerance statistics. Let's begin by using what we learned in the correlation exercise, let's click the regression menu, select correlation, and then select our four variables. We are interested in looking at the Pearson's r values for our study variables, so we won't select any fancy options this time.

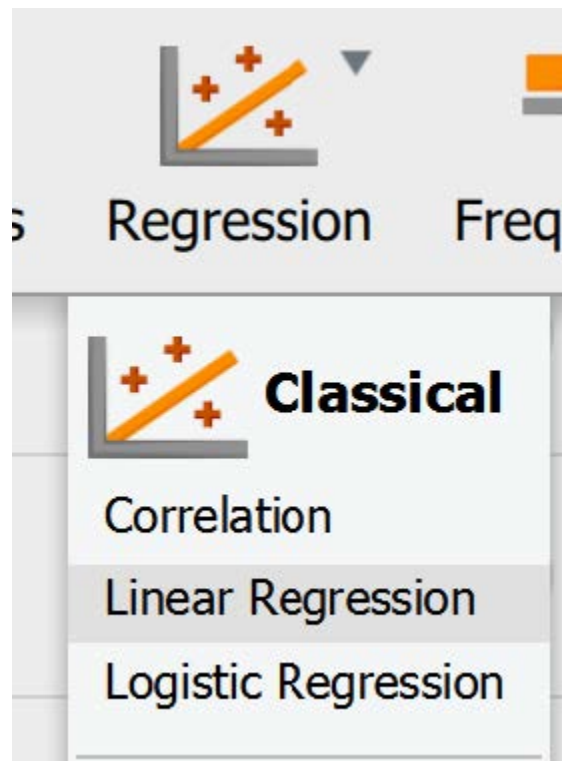
In our output table, we see that all three predictors are related to our outcome. We will consider those relationships in more detail with our actual regression analysis. For now, we are concerned with whether our three predictors are too highly correlated. In this case, the Cultural Intelligence and Knowledge Sharing are correlated at .63, meaning that the more cultural intelligence, the more knowledge sharing there is. This correlation is strong, so it is high enough that we may be concerned about multicollinearity in our data. Cultural Intelligence and Organization Culture Difference are moderately correlated at .52 and Knowledge Sharing and Organization Culture Difference are also highly correlated at .62. Keep in mind, we will also look at the VIF and tolerance values in the regression output as additional information about this assumption.

Pearson's Correlations

Variable		CQ	KS	OCD	SIB
1. CQ	Pearson's r	—			
	p-value	—			
2. KS	Pearson's r	0.628			
	p-value	< .001	—		
3. OCD	Pearson's r	0.524	0.624		
	p-value	< .001	< .001	—	
4. SIB	Pearson's r	0.669	0.529	0.461	
	p-value	< .001	< .001	< .001	—

Primary Analyses

Now we are ready to run our multiple regression model. We will go back up to the regression tab at the top of the JASP menu bar and select the option for linear regression.



In the main regression menu, we will move over Sustainable Innovation Behavior as our dependent variable and Cultural Intelligence (CQ), Knowledge Sharing (KS), and Organizational Culture Differences (OCD) as our covariates (or predictor variables).

Linear Regression

Variables: ID, SEX, AGE, ED

Dependent Variable: SIB

Method: Enter

Covariates: CQ, KS, OCD

Factors:

WLS Weights (optional):

There is nothing we need to adjust in the “model” drop down, so we will move on to “statistics”. In addition to the default settings, we will click:

- “Confidence intervals” under regression coefficients to get a range of likely slope values.
- “Statistics” under residuals to look at the typical amount of error in our model.
- “Descriptives” in the right column of options, just to look at typical values in our dataset.
- “Collinearity diagnostics” to continue checking assumption six, that we do not have multicollinearity.
- “Durbin-Watson” to check assumption seven, that we have independence of errors.

▼ Statistics

Regression Coefficients

☒ Estimates ☐ From 5000 bootstraps ☒ Model fit

☒ Confidence intervals 95.0 % ☐ R squared change

☐ Covariance matrix ☐ Descriptives

☐ Vovk-Sellke maximum p-ratio ☐ Part and partial correlations

☒ Collinearity diagnostics

Residuals

☒ Statistics

☒ Durbin-Watson

☐ Casewise diagnostics

☐ Standard residual > 3

☐ Cook's distance > 1

☐ All

We will also skip over the model specification dropdown and move onto plots. In the plot dropdown, we will click on Residuals vs. histogram to request residuals plotted as a histogram.

▼ Plots

Residuals Plots

☐ Residuals vs. dependent

☐ Residuals vs. covariates

☐ Residuals vs. predicted

☒ Residuals histogram

☒ Standardized residuals

☐ Q-Q plot standardized residuals

☐ Partial plots

☐ Confidence intervals 95.0 %

☐ Prediction intervals 95.0 %

Other Plots

☐ Marginal effects plots

☐ Confidence intervals 95.0 %

☐ Prediction intervals 95.0 %

Now we're ready to look at our output! First let's begin by checking our remaining assumptions.

Assumption Six: Are our predictor variables highly correlated?

As mentioned previously, if our predictor variables are highly correlated with each other, this can create problems in interpreting the results of the analysis. Collinearity is the term used to describe this issue of high

correlation between predictor variables. One way to check for collinearity is by looking at the correlations between the variables. Another method is by checking the Variance Inflation Factor (VIF) and Tolerance statistics. The VIF measures how much of the variance in our model is increased due to collinearity (or correlation) between our predictor variables. The Tolerance statistic is the reciprocal of the VIF and measures the proportion of variation in one predictor variable that is not explained by the other predictor variables in the model. In general, we want to see VIF values below 5 and Tolerance statistics above 0.2. If the VIF values are higher than 5, it suggests that there may be too much collinearity between the predictor variables, and this can lead to inaccurate estimates of the regression coefficients. In other words, it can be difficult to determine which predictor variables are contributing to the outcome variable. If our VIF values are above 5, we may need to: (1) remove one or more predictor variables or (2) combine variables into composite scores.

Assumption Seven: Do we have independence of errors?

One important consideration in regression analysis is whether the residuals (the differences between the actual values and predicted values) are correlated with each other or not. The Durbin-Watson statistic is a measure of this correlation. When the Durbin-Watson statistic is between 1.5 and 2.5, it suggests that there is not a significant correlation between the residuals or errors in our model. This is important because if there is a high degree of correlation between residuals, it can suggest that the model is not capturing all the relevant information about the relationship between the variables. In other words, the model may not be a good fit for the data. A Durbin-Watson statistic outside of the 1.5-2.5 range may indicate that there is a problem with the model, and further investigation may be needed to determine the cause of the correlation between residuals.

The Durbin-Watson statistic is a part of the Model Summary results. If this value is between 1.5-2.5, we are not concerned there is a significant amount of correlation between the residuals.

Model Summary - SIB

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	6.242	0.060	1.871	0.236
H ₁	0.688	0.473	0.468	4.552	0.029	1.922	0.469

This is our Durbin-Watson statistic. We want this value to be between 1.5-2.5.

Looking at the preceding figure, we can determine that our Durbin-Watson statistic is within the acceptable range of 1.5-2.5. To report this in APA format:

The model has acceptable independence of errors, Durbin-Watson = 1.92.

Let's look at our new output with three predictors. In the Model Summary table, we see that our three variables together are correlated with Sustainable Innovation Behavior at .688 (Multiple R). We also see that, together, Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences explain 46.8%

(R^2) of the Sustainable Innovation Behavior. That number grew from 44.8% in our simple regression with only Cultural Intelligence. That tells us that Knowledge Sharing and Organizational Culture Differences do help explain a little bit more about Sustainable Innovation Behavior, but only about 2% more when we add all three predictors together.

Model Summary - SIB

Model	R	R^2	Adjusted R^2	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	6.242	0.060	1.871	0.236
H ₁	0.688	0.473	0.468	4.552	0.029	1.922	0.469

This is our r^2 value. Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences explains 47.3% of the variance in Sustainable Innovation Behavior.

Our ANOVA table shows us that our linear model, including the three predictors, is explaining a significant amount of variance in Sustainable Innovation Behavior. The F value of 99.36 has a p value less than .001, which is below our .05 cutoff. This tells us that our overall model is significant.

ANOVA ▼

Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	6174.708	3	2058.236	99.336	< .001
	Residual	6878.994	332	20.720		
	Total	13053.702	335			

Note. The intercept model is omitted, as no meaningful information can be shown.

df 1

df 2

F statistic

p value. Because p is < .05 (our alpha level), this tells us our model is statistically significant.

Now that we have determined that we have a significant overall model predicting Sustainable Innovation Behavior in the workplace, we would like to gain a better understanding of the relationship between each of our predictor variables with the dependent variable. To do this, we are going to look at the slopes of each of our predictors, which we'll find in the coefficients table. Our intercept now reflects if you are zero on all three predictors. So, we can say if a person has no Cultural Intelligence, no Knowledge Sharing, and no Organizational Culture Differences, the average Sustainable Innovation Behavior is 6.91. The slope for Cultural Intelligence has gotten smaller now. That's because Knowledge Sharing and Organizational Culture Differences are explaining some of the variance now too, making the effect of Cultural Intelligence a bit weaker. The slope of Cultural Intelligence accounting for Knowledge Sharing and Organizational Culture Differences,

is .294. Our associated t value of 10.187 has a p value of $<.001$, which is less than .05. We would say that Cultural Intelligence is a significant predictor of Sustainable Innovation Behavior when you also account for Knowledge Sharing and Organizational Culture Differences. Knowledge Sharing does have a significant slope ($b = .196$, $t = 2.322$, $p < .02$). For every one-unit increase in Knowledge Sharing, you can expect Sustainable Innovation Behavior to increase by .196. Organizational Culture Differences does not have a significant slope ($b = .065$, $t = 1.881$, $p = .06$). For every one-unit increase in Organizational Culture Differences, you can expect Sustainable Innovation Behavior to increase by .065.

Our standardized slopes further show us that Cultural Intelligence (beta = .53) is a stronger predictor than Knowledge Sharing (beta = .13) and Organizational Culture Differences (beta = .10).

Coefficients table from multiple regression output in JASP.

a or the Y -intercept of the regression line (9.32).

The standardized column shows us Beta, the standardized slopes for our predictors.

Confidence intervals. This gives us the likely range of slope values.

Coefficients		Unstandardized	Standard Error	Standardized	t	p	95% CI		Collinearity Statistics	
Model							Lower	Upper	Tolerance	VIF
H ₀	(Intercept)	29.970	0.341		88.007	< .001	29.300	30.640		
H ₁	(Intercept)	6.912	1.392		4.965	< .001	4.173	9.650		
	CQ	0.294	0.029	0.534	10.187	< .001	0.237	0.351	0.577	1.733
	KS	0.196	0.084	0.133	2.322	0.021	0.030	0.361	0.486	2.058
	OCD	0.065	0.035	0.098	1.881	0.061	-0.003	0.133	0.581	1.720

b or the slopes of the regression line for each of our predictors.

SE or the standard error.

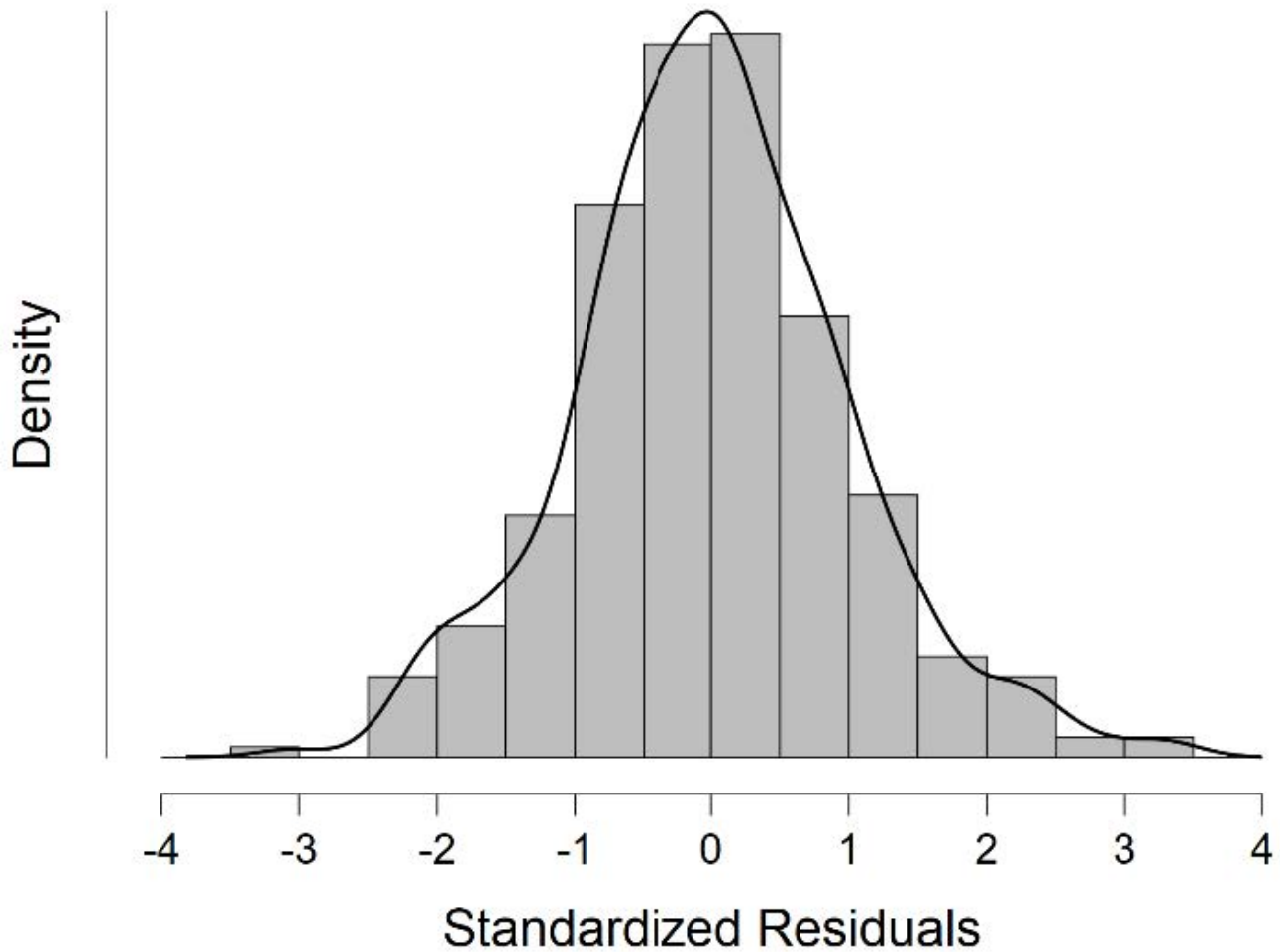
t statistics and p values for our predictor variables. If the associated p value is less than .05, then we can conclude our slope is significantly different from zero. The p value for Organizational Culture Differences is greater than .05, meaning this is not a significant predictor in our model.

VIF statistics. Values above 10 indicate multicollinearity issues.

The other piece of information we are interested in is the collinearity statistics in the last two columns. The VIF measures how much of the variance in our model is increased due to collinearity (or correlation) between our predictor variables. The Tolerance statistic is the reciprocal of the VIF and measures the proportion of variation in one predictor variable that is not explained by the other predictor variables in the model. In general, we want to see VIF values below 5 and Tolerance statistics above 0.2. If the VIF values are higher than 5, it suggests that there may be too much collinearity between the predictor variables, and this can lead to inaccurate estimates of the regression coefficients. In other words, it can be difficult to determine which predictor variables are contributing to the outcome variable. If our VIF values are above 5, we may need to: (1) remove one or more predictor variables or (2) combine variables into composite scores.

Our tolerance values range from .49 to .58, meaning that between 49% and 58% of the variance in the predictors is unshared. Our largest VIF value is 2.06, which is below 10, so there is not much concern for multicollinearity.

Related to our assumptions, we also want to make sure we take another look at our residuals plot for our multiple regression model. Again, we see that our residuals for our model (based on the regression equation with three predictors) approximate a fairly normal distribution.



Putting it all Together

It's time to summarize the results of our multiple regression analyses.

Prior to conducting our multiple regression, we tested several assumptions. We examined our predictors (Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences) and outcome (Sustainable Innovation Behavior) variables for normality and any concerns with outliers. All skewness and kurtosis values were all in the normal range. Upon inspection of variable boxplots and histograms, four outliers were revealed that were retained for analysis. Using a scatterplot, we also visualized the relationships between the variables. Upon viewing the scatterplots, we observed that the relationships appeared linear and that there were no concerns with heteroscedasticity. Plots of the residuals also showed that the residuals were approximately normally distributed. We also looked at the correlation between our predictor variables, finding that the variables were moderately correlated (r range = .52 to .63). The variance inflation factor values for our model were all less than 2.05, providing support that there were not substantial concerns with multicollinearity. The model has acceptable independence of errors, Durbin-Watson = 1.92.

In our multiple regression, we examined how the Cultural Intelligence, Knowledge Sharing, and Organization Culture affected Sustainable Innovation Behavior. We found that together, these three predictors

explained a significant amount of variance in Sustainable Innovation Behavior, $F(3, 336) = 99.3, p < .001$. Together, the predictors explained 46.8% of the variance in Sustainable Innovation Behavior. Organizational Culture Differences, Cultural Intelligence, and Knowledge Sharing were significant predictors of Sustainable Innovation Behavior, when controlling for the other predictors. Knowledge Sharing was positively related to Sustainable Innovation Behavior ($b = .20, SE = .08, t(332) = 2.32, p = .02$). Cultural Intelligence was also positively related to Sustainable Innovation Behavior ($b = .29, SE = .03, t(332) = 10.19, p < .001$). Organizational Culture Differences was not a significant predictor of Sustainable Innovation Behavior ($b = .07, SE = .04, t(332) = 1.88, p = .06$).

Real World Meaning

From this study, we could share that Cultural Intelligence, Knowledge Sharing, and Organization Culture Differences together impact Sustainable Innovation Behavior, but the strongest relationship exists between Sustainable Innovation Behavior and Cultural Intelligence, where the more cultural intelligence someone has, the more likely they are to also engage in sustainable innovation behaviors in the workplace. The same pattern is true with knowledge sharing, that sharing more information is related to more innovation, just not as strongly as cultural intelligence relates to innovation behaviors. Together, Cultural Intelligence, Knowledge Sharing, and Organizational Culture Differences can explain over half of the variance in Sustainable Innovation Behavior across a sample of individuals in organizations.

In their study, Li et al. (2021) found similar results. They examined how Cultural Intelligence impacted Sustainable Innovation Behavior with Knowledge Sharing as an intermediary variable and Organizational Culture Differences as a moderator. Their results showed that Cultural Intelligence positively impacts Sustainable Innovation Behavior, but this relationship is mediated or linked by Knowledge Sharing. They also found that Organizational Culture Differences did not significantly moderate or change the relationship between Cultural Intelligence and Sustainable Innovation Behavior, but it did impact the relationship between Cultural Intelligence and Knowledge Sharing.

The results of our statistical analysis and that of Li et al. (2021) both suggest that Organizational Culture Differences does not really have a significant impact on our main variable, Sustainable Innovation Behavior. It can also be said that Sustainable Innovation Behavior and Cultural Intelligence do have a strong relationship together whereas the relationship between Sustainable Innovation Behavior and Knowledge Sharing is not as strong.

References

- Gölgeci, I., Swiatowiec-Szczepanska, J., & Raczkowski, K. (2017). How does cultural intelligence influence the relationships between potential and realised absorptive capacity and innovativeness? Evidence from Poland. *Technology Analysis & Strategic Management*, 29(8), 857–871. <https://doi.org/10.1080/09537325.2016.1245858>
- Hu, S., Gu, J., Liu, H. & Huang, Q. (2017). The moderating role of social media usage in the relationship among multicultural experiences, cultural intelligence, and individual creativity. *Information Technology & People*, 30(2), pp. 265-281. <https://doi.org/10.1108/ITP-04-2016-0099>

Li, J., Wu, N., & Xiong, S. (2021). Sustainable innovation in the context of organizational cultural diversity: The role of cultural intelligence and knowledge sharing. *PloS One*, 16(5), e0250878–e0250878. <https://doi.org/10.1371/journal.pone.0250878>

Pandey, A. & Charoensukmongkol, P. (2019). Contribution of cultural intelligence to adaptive selling and customer-oriented selling of salespeople at international trade shows: does cultural similarity matter? *Journal of Asia Business Studies*, 13(1), pp. 79-96. <https://doi.org/10.1108/JABS-08-2017-0138>

Authors

This guide was written and created by Ruth V. Walker, PhD, Ashlyn A. Moraine, and Kristen J. Black, PhD.

Acknowledgements

We would like to thank Jinlong Li, Na Wu, and Shengxu Xiong for generously publishing their data through PloS One.

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



Running and Interpreting Multiple Regression in JASP Copyright © 2021 by Ruth Walker; Ashlyn A. Moraine; and Kristen J. Black is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, except where otherwise noted.

INSTRUCTOR DEVELOPMENT RESOURCES FOR INTEGRATING AND TEACHING EDI TOPICS

Ruth Walker

Instructor Development Resources for Integrating and Teaching EDI Topics

Adams, M. & Bell, L. A. (Eds.). (2016). *Teaching for diversity and social justice: A sourcebook* (3rd ed). New York, NY: Routledge Press.

American Psychological Association. (2022, July 19). *Transforming introductory psychology: Integrating diversity, equity, and inclusion* [Video]. Youtube. <https://youtu.be/dFv2Z6Jf0cc>

American Psychological Association. (2021, September 20). *The Science of race, sex and gender, and inclusivity* [Video]. Youtube. <https://youtu.be/DgTHYTJwfNo>

American Psychological Association. (2020-present). *Psychology and advocacy playlist*. [Videos]. Youtube. <https://www.youtube.com/watch?v=YFJaA4uCh8E&list=PLxf85IzktYWLRVaV2Y7a0LLKswP5nBwbZ>

American Psychological Association. (2013-present). *Training videos on gender and sexual minority issues playlist*. [Videos]. Youtube. <https://www.youtube.com/watch?v=yPRdrs4ZPNY&list=PLxf85IzktYWLJz-Xgz47o3b0M8jd4ErND>

Barnes, C. D. & Slaton, C. R. (2020). *Cultural humility: Expanding our view*. In cultural competence in higher education. In T. Puckett & N. S. Lind (Eds.), *Cultural Competence in Higher Education* (Vol. 28, pp. 53–63). Emerald Publishing Limited. <https://doi.org/10.1108/S2055-364120200000028007>

Bhatia, S. (2017). *Decolonizing psychology: Globalization, social justice, and Indian youth identities*. Oxford, UK: Oxford University Press.

Comas-Díaz, T. & Torres Rivera, E. (Eds.) (2020). *Liberation psychology: Theory, method, practice, and social justice*. In *Liberation psychology: Theory, method, practice, and social justice*. American Psychological Association. <https://doi.org/10.1037/0000198-000>

Etengoff, C. (2022). *Reframing psychological research methods courses as tools for social justice education*. *Teaching of Psychology*, 9862832210974–. <https://doi.org/10.1177/00986283221097404>

Fish, J. (2022). *Interview with a psychologist: Representing marginalized psychologists and diversity science in psychology coursework*. *Teaching of Psychology*, 9862832110690–. <https://doi.org/10.1177/00986283211069076>

Freire, P. (1970). *Pedagogy of the oppressed*. London, England: Penguin Books.

Freire, P. (1974). *Education for critical consciousness*. New York: Continuum.

Freire, P. (2014). *Pedagogy of hope: Reliving pedagogy of the oppressed*. New York, NY: Bloomsbury Academic.

Fuentes, M. A. & Shannon, C. R. (2016). The state of multiculturalism and diversity in undergraduate psychology training. *Teaching of Psychology*, 43(3), 197–203. <https://doi.org/10.1177/0098628316649315>

Gannon, K. M. (2020). *Radical hope: A teaching manifesto*. Morgantown, WV: West Virginia University Press.

Garlington, T., Ryan, V. M., Nolty, C., Ilagan, H., & Kunicki, Z. J. (2021). Bringing social justice into the statistics classroom. *Teaching of Psychology*, 48(3), 269–274. <https://doi.org/10.1177/0098628320979879>

Goldstone, R. (Executive Producer). (2020-present). *Under the cortex* [Audio podcast]. Association for Psychology Science. <https://psychologicalscience.podbean.com/>

Gómez, J. M. (2022). Diversity wanted! Utilizing transdisciplinary scholarship on structural inequality to educate psychology graduate students. *Teaching of Psychology*, 9862832110616–. <https://doi.org/10.1177/00986283211061687>

Hammond, Z. (2014). *Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students*. Thousand Oaks, CA: Corwin Press.

Haselau, T., & Saville Young, L. (2022). Co-constructing defensive discourses of service-learning in psychology: A psychosocial understanding of anxiety and service-learning, and the implications for social justice. *Teaching of Psychology*, 9862832210772–. <https://doi.org/10.1177/00986283221077206>

Hicks, E. T., Alvarez, M. de la C., & Domenech Rodríguez, M. M. (2022). Impact of difficult dialogues on social justice attitudes during a multicultural psychology course. *Teaching of Psychology*, 9862832211040–. <https://doi.org/10.1177/00986283221104057>

Hill Collins, P. (2012). *On intellectual activism*. Philadelphia, PA: Temple University Press.

Hogan, K. A., & Sathy, V. (2022). *Inclusive teaching: Strategies for promoting equity in the college classroom*. Morgantown, WV: West Virginia University Press.

hooks, b. (1994). *Teaching to transgress*. New York, NY: Routledge.

Huston, T. (2012). *Teaching what you don't know*. Cambridge, MA: Harvard University Press.

Jack, A. A. ((2019). *The Privileged poor: How elite colleges are failing disadvantaged students*. Cambridge, MA: Harvard University Press.

Jordan, L. S. (2022). Integrating qualitative inquiry and critical whiteness in psychology research methods courses. *Teaching of Psychology*, 9862832110568–. <https://doi.org/10.1177/00986283211056886>

Kadaba, M. L., Chow, J. S., & Briscoe-Smith, A. (2022). Recommendations for creating and teaching a graduate psychology course exclusively for Black, Indigenous, and People of Color Students. *Teaching of Psychology*, 9862832110700–. <https://doi.org/10.1177/00986283211070012>

Kernahan, C. (2019). *Teaching about race and racism in the college classroom: Notes from a white professor*. Morgantown, WV: West Virginia University Press.

Kowalski. (2000). Including gender, race, and ethnicity in psychology content courses. *Teaching of Psychology*, 27(1), 18–24. https://doi.org/10.1207/S15328023TOP2701_3

Love, B. L. (2019). *We want to do more than survive: Abolitionist teaching and the pursuit of educational freedom*. Boston, MA: Beacon Press.

Lund, E. M. (2021). Addressing the leaking pipeline: Supporting disabled graduate student teachers in psychology. *Teaching of Psychology*, 9862832110364–. <https://doi.org/10.1177/00986283211036420>

Maimon, M. R., Howansky, K., & Sanchez, D. T. (2021). Fostering inclusivity: Exploring the impact of identity safety cues and instructor gender on students' impressions and belonging. *Teaching of Psychology*, 9862832110437–. <https://doi.org/10.1177/00986283211043779>

Morgan, M. L., & Marin, P. (2022). “I was born to do this”: Faculty experiences teaching graduate-level diversity courses. *Teaching of Psychology*, 9862832110669–. <https://doi.org/10.1177/00986283211066943>

Mustafaa, F. N. & Martinez, M. N. (2022). A visionary working model for pursuing social justice praxis through educational psychology courses. *Teaching of Psychology*, 9862832110657–. <https://doi.org/10.1177/00986283211065732>

Pickering, R. M. (2022). Structures of inequity: Teaching privilege and oppression with a tower-building activity. *Teaching of Psychology*, 9862832110568–. <https://doi.org/10.1177/00986283211056888>

Prieto, L. R. (2018). Incorporating diversity content into courses and concerns about teaching culturally diverse students. *Teaching of Psychology*, 45(2), 146–153. <https://doi.org/10.1177/0098628318762875>

Prieto, L. R., Whittlesey, V., Herbert, D., Ocampo, C., Schomburg, A., & So, D. (2009). Dealing with diversity issues in the classroom: A survey of the STP membership. *Teaching of Psychology*, 36(2), 77–83. <https://doi.org/10.1080/00986280802529236>

Rains, F. (1998). Is the benign really harmless? Deconstructing some “benign” manifestations of operationalized white privilege. In J. L. Kincheloe, S. R. Steinberg, N. M. Rodriguez, & R. E. Chennault (Eds.), *White reign: Deploying whiteness in America*. New York: St. Martin's.

Sue, D. W., Rivera, D. P., Watkins, N. L., Kim, R. H., Kim, S. & Williams, C. D. (2011). Racial dialogues: Challenges faculty of color face in the classroom. *Cultural Diversity and Ethnic Minority Psychology*. 17, 331-340. <https://doi.org/10.1037/a0024190>

Sue, D. W., Alsaidi, S., Awad, M. N., Glaeser, E., Calle, C. Z., & Mendez, N. (2019). Disarming racial microaggressions: Microintervention strategies for targets, White allies, and bystanders. *American Psychologist*, 74(1), 128-142. <http://dx.doi.org/10.1037/amp0000296>

TEDxTalks. (2021, June 11). Affirming diversity in the classroom why it matters to your students | Nadiyah Herron | TEDxCSUSB [Video]. Youtube. <https://youtu.be/fqUPSSTFKE>

TEDxTalks. (2015, May 5). Practical diversity: taking inclusion from theory to practice | Dawn Bennett-Alexander | TEDxUGA [Video]. Youtube. <https://youtu.be/ExcDNly1DbI>

Tervalon, M., & Murray-Garcia, J. (1998). Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education. *Journal of Health Care for the Poor and Underserved*, 9(2), 117-125. <https://doi.org/10.1353/hpu.2010.0233>

Tuitt, F., Haynes, C., & Stewart, S. (2016). *Race, equity, and the learning environment: The global relevance of critical and inclusive pedagogies in higher education*. Sterling, VA: Stylus Publishing, LLC.

Society for the Teaching of Psychology Facebook Group. <https://www.facebook.com/groups/teachpsych/>

Stetsenko, A. (2020). Transformative-activist and social justice approaches to the history of psychology. In: *Oxford Research Encyclopedia of Psychology*.

Stetsenko, A. (2020). Research and activist projects of resistance: The ethical-political foundations for a transformative ethico-onto-epistemology. *Learning, Culture and Social Interaction*, 26, 100222 <https://doi.org/10.1016/j.lcsi.2018.04.002>

Vianna, E., & Stetsenko, A. (2014). Research with a transformative activist agenda: Creating the future through education for social change. In J. A. Vadeboncouer (Ed.), *Learning in and across contexts: Reimagining education* (Vol. 113, pp. 575–602). New York, NY: Teachers College Press.

Walter, M., & Andersen, C. (2016). *Indigenous statistics: A quantitative research methodology*. New York, NY: Routledge.

Wong, M.S., Weiner, L., Cerniak, J., & Yee, L.T.S. (Eds.). (2021). *Incorporating diversity in classroom settings: Real and engaging examples for various psychology courses*. (Vol 1: Ability, age, culture, ethnicity/ race, gender, religion, sexual orientation, and socioeconomic status). Retrieved from the Society for the Teaching of Psychology website: <http://teachpsych.org/ebooks/diverse1>

Wong, M.S., Weiner, L., Cerniak, J., & Yee, L.T.S. (Eds.). (2021). *Incorporating diversity in classroom settings: Real and engaging examples for various psychology courses*. (Vol 2: Intersectionality). Retrieved from the Society for the Teaching of Psychology website: <http://teachpsych.org/ebooks/diverse2>

Yoder, J., Mills, A. S., & Raffa, E. R. (2016). An effective intervention in research methods that reduces psychology majors' sexist prejudices. *Teaching of Psychology*, 43(3), 187–196. <https://doi.org/10.1177/0098628316649314>

Authors

This list of instructor resources was created by Ruth V. Walker, PhD. If you have additional resources you would like to suggest be added to this resource list, please send them to ruth-walker@utc.edu

Copyright

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.



DOWNLOADABLE MATERIALS

Instructor PowerPoint Slides

Course Introduction Instructor PPT Slides

Independent Samples t Test Instructor PPT Slides

Paired Samples t Test Instructor PPT Slides

One-Way ANOVA Instructor PPT Slides

Two Way ANOVA Instructor PPT Slides

Chi-Square Instructor PPT Slides

Pearson Correlation Instructor PPT Slides

Regression Instructor PPT Slides