PROJECT 4: Reporting Patterns in National Values

GUIDEBOOK

For this project, you will be asked to observe, and critically reflect upon, patterns in social data in a representative sample of United States citizens (using data from the General Social Survey). First, you will perform a cross-tabulation of two ordinal/nominal variables. The second test, an ANOVA, will have you examine a causal relationship between an ordinal independent variable and an interval-ratio dependent variable. And for your third test, you will examine the relationship between two interval-ratio variables. This project covers learning goals and objectives #1, 2, 3, and 4 as listed in the syllabus.

TABLE OF CONTENTS

REQUIRED MATERIALS

- A word processing program to type up your final **REPORT** (e.g., Microsoft Word)
- The data analysis program **SPSS**
- The data **FILE** (.sav file located on Canvas) containing all of the variables and scores from the sample
- The **CODEBOOK** (SECTION B) containing a list of the variables and values you may use
- **ANALYSIS INSTRUCTIONS** (SECTION C) explain how to use SPSS
- Your **NOTES** from lessons 9, 10, and 11
- The **RUBRIC** (SECTION D)
- A project **EXAMPLE** (SECTION E)

SECTION A: PROJECT INSTRUCTIONS

Description. For this project, you will be asked to do a little research and perform two types of hypothesis tests using data from the Global Values Survey. First, you will perform a single-sample hypothesis test, where you will compare data from your class to the global population on a single variable. The second test, an independent-samples hypothesis test, will have you explore how two different group means compare on a single variable. You will also be asked to evaluate the work of your peers.

Organization. Your final report will be organized by test. The following is how your report will be organized:

PART ONE: CROSS-TABULATION and CHI-SQUARE

1. Description of independent and dependent variables (5 points)
2. Hypothesis statements (5 points)
3. Tables (5 points)
4. Annotated test statistics (2.5 points)
5. Hypothesis decision (5 points)
6. Statement of test strength (2.5 points)
7. Reflection (10 points)

PART TWO: ANALYSIS OF VARIANCE (ANOVA)

8. Description of independent and dependent variables (5 points)
9. Hypothesis statements (5 points)
10. Annotated test statistics (5 points)
11. Hypothesis decision (5 points)
12. Examination of multiple relationships (5 points)
13. Reflection (10 points)

PART THREE: REGRESSION AND CORRELATION

14. Description of independent and dependent variables (5 points)
15. Statement and explanation of regression equation (5 points)
16. Report and description of CORRELATION COEFFICIENT (5 points)
17. Report and description of COEFFICIENT OF DETERMINATION (5 points)
18. Reflection (10 points)

Stepwise instructions. For information on how to perform these tests in SPSS, please refer to the Analysis instructions (SECTION C). For the rubric and an example project, please refer to SECTIONS D and E.

PART ONE: CROSS-TABULATION and CHI-SQUARE

For PART 1, use SPSS to report a cross-tabulation between two variables, test a hypothesis with a chi-square, and reflect on that relationship. You will need to select two appropriate variables for this portion of the assignment. *If you select variables with inappropriate levels of measurement, you will lose credit on this portion of the assignment.*

1. **Introduction and descriptive statistics.** Describe the variables are you testing. Use SPSS to generate the following statistics about your variable but treat them as population parameters (hint – you can refer back to instructions from project 2 on how to obtain descriptive statistics using SPSS):
   - Population size (N=___)
   - mean, median, mode, standard deviation,

- standard error,
- the minimum/maximum values, and the range.

Then, in a small paragraph, critically reflect on these statistics. *Consider the following questions: Why does your independent variable influence your dependent variable? What is the best measure of central tendency for each variable and what do these values mean?*

2. **Hypothesis statement.** For this test, write a null and **one-tailed** research hypothesis using symbols AND using appropriate statistical language.
3. **Tables.** Perform a cross-tabulation of your variables in SPSS. Also have SPSS produce column percentages. In Word or Excel, produce your own cross-tabulation table with column percentages AND row/column totals table using the information gathered from SPSS. HINT: Your independent variable should be reported as the column variable and your dependent variable should be the row variable.
4. **Annotated test statistics.** Report and describe (in a list) the obtained Pearson chi-square statistic, the degrees of freedom, the P-value, and the chi-square critical value.
5. **Hypothesis decision.** Using appropriate statistical language, state your decision regarding your hypothesis in one sentence. Assume your α (alpha) is set at a 0.05 level.
6. **Statement of test strength.** Select and report the most appropriate measure of association for your variables. You can report lambda, Cramer's V, gamma, Kendall's tau-b or Kendall's tau-c, depending on the nature of your variables. What does this indicate about your test?
7. **Reflection.** In three to seven sentences, explain what the findings from the cross-tabulation, chi-square test, and the measures of association mean in words. Make sure you use appropriate statistical language and describe what the findings actually mean, and critically reflect on your population. You are not required to use resources, but they will strengthen your arguments.

PART TWO: ANALYSIS OF VARIANCE (ANOVA)

For this part, you will be asked to conduct a hypothesis test in SPSS using ANOVA and Tukey's HSD (Tukey's HSD is only covered in the class memo). For this portion of the project, you will need to select an ordinal or nominal independent variable and an interval-ration dependent variable.

8. **Description of independent and dependent variables.** Describe the variables are you testing. Use SPSS to generate the following statistics about your variable but treat them as population parameters (hint – you can refer back to instructions from project 2 on how to obtain descriptive statistics using SPSS):
   - Population size (N=___)
   - mean, median, mode, standard deviation,
   - standard error,
   - the minimum/maximum values, and the range.

Then, in a small paragraph, critically reflect on these statistics. *Consider the following questions: Why does your independent variable influence your dependent variable? What is the best measure of central tendency for each variable and what do these values mean?*

9. **Hypothesis statement.** For this test, write a null and **one-tailed** research hypothesis using symbols AND using appropriate statistical language.

10. **Annotated test statistics.** List and describe the SSB, SSW, dfb, dfw, MSb, MSw, the obtained F-statistic, the P-value, and the F-critical value.
11. **Hypothesis decision.** Using appropriate statistical language, state your decision regarding your hypothesis in one sentence. Assume your α (alpha) is set at a 0.05 level.
12. **Examination of multiple relationships.** Examine the table created from Tukey's HSD. In a list, report all statistically significant differences. Report the mean difference and the P-value for each statistically significant difference. *HINT: If there are none, you must clearly state there are no statistically significant differences between groups in order to receive credit for this portion of the assignment.*
13. **Reflection.** In three to seven sentences, explain what the findings from all of your statistics in words. Make sure you use appropriate statistical language and describe what the findings actually mean, and critically reflect on your population. You are not required to use resources, but they will strengthen your arguments.

PART THREE: REGRESSION AND CORRELATION

In this part, you will perform a **regression and correlation.** You will need to select two appropriate variables for this portion of the assignment. *NOTE: Be sure to select an independent and dependent variable. Also remember which levels of measurement are appropriate for regression. If you select variables with inappropriate levels of measurement, you may lose credit on this portion of the assignment.*

14. **Description of independent and dependent variables.** Describe the variables are you testing. Use SPSS to generate the following statistics about your variable but treat them as population parameters (hint – you can refer back to instructions from project 2 on how to obtain descriptive statistics using SPSS):
    - Population size (N=___)
    - mean, median, mode, standard deviation,
    - standard error,
    - the minimum/maximum values, and the range.

Then, in a small paragraph, critically reflect on these statistics. *Consider the following questions: Why does your independent variable influence your dependent variable? What is the best measure of central tendency for each variable and what do these values mean?*

15. **Statement and explanation of regression equation.** Using SPSS, conduct the regression and write the complete regression equation. Using appropriate statistical language, explain what the regression equation means in two to five sentences.
16. **Report and description of CORRELATION COEFFICIENT.** Report and label the correlation coefficient. Using appropriate statistical language, interpret the correlation coefficient in one to two sentences.
17. **Report and description of COEFFICIENT OF DETERMINATION.** Report and label the coefficient of determination. Using appropriate statistical language, interpret the coefficient of determination.
18. **Reflection.** In three to seven sentences, explain what the findings from all of your statistics in words. Make sure you use appropriate statistical language and describe what the findings actually mean, and critically reflect on your population. You are not required to use resources, but they will strengthen your arguments.

SECTION B: CODEBOOK

You must pay attention to the level of measurement for all projects.

A codebook will contain all possible variables in a given data set. Read over the descriptive statistics and explore the data (i.e., the .sav) file in SPSS in order to learn more. This data is a subset of the GENERAL SOCIAL SURVEY (GSS) from around 2016. The data has been cleaned (e.g., many cases omitted from the "original data") and the response categories edited for ease of use. For more on GSS, please visit: http://gss.norc.org/

Sample: n=743 Americans from who answered all questions from the GSS.

Code	Variable Description	Response Categories
marital	Respondent marital status	1=Married; 2=Widowed; 3=Divorced; 4=Separated; 5=Never Married
sex	Respondent sex	1=male; 2=female
race	Respondent race	0=no; 1=white; 2=black; 3=other
vote12	Respondent voted in 2012 Election	1=yes; 2=no; 3=ineligible
owngun	Respondent owns a gun or revolver	1=yes; 2=no; 3=refused
partyid	Respondent political spectrum ID	0=Strong Democrat… 3=Independent… 6=Strong Republican
happy	Respondent general happiness	1=very happy; 2=pretty happy; 3=not too happy
health	Respondent general health	1=Excellent; 2=good; 3=fair; 4=poor
satfin	Respondent satisfaction with financial situation	1=satisfied; 2=more or less; 3=not at all
tvhours	Respondent hours per day watching TV	0,1,2,3,4,5…24
wwwhr	Respondent number of hours respondent spends on internet per week	0,1,2,3,4,5…168
childs	Number of children	0,1,2,3,4,5,6,7, & 8+
age	Age of respondent	18-99
educ	Respondent highest year of education	0,1,2,3,4,5…20

[marital] Are you currently--married, widowed, divorced, separated, or have you never been married?

[sex] Respondent's sex

[race] What race do you consider yourself?

[vote12] In 2012, you remember that Obama ran for President on the Democratic ticket against Romney for the Republicans.
Do you remember for sure whether or not you voted in that election?

[owngun] Do you happen to have in your home (IF HOUSE: or garage) any guns or revolvers?

[partyid] Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

[happy] Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?

[health] Would you say your own health, in general, is excellent, good, fair, or poor?

[satfin] We are interested in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all?

[tvhours] On the average how many hours of TV do you watch per day?

[wwwhr] Not counting e-mail, about how many minutes or hours per week do you use the Web? (Include time you spend visiting regular web sites and time spent using interactive Internet services like chat rooms, Usenet groups, discussion forums, bulletin boards, and the like.)

[childs] How many children have you ever had? Please count all that were born alive at any time (including any you had from a previous marriage).

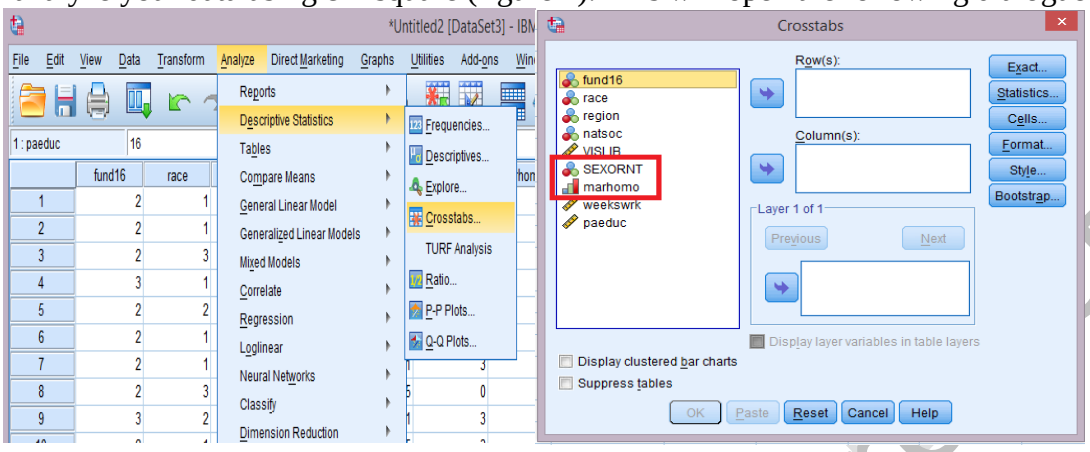
[age] Age of respondent at time of survey.

[educ] Highest year of education. Note: 1-12 = grades in school, 13-20 = indicate years spent in college.

SECTION C: ANALYSIS INSTRUCTIONS

Cross-Tabulation and Chi-Square: Background

Once your data are open, click **Analyze,** then **Descriptive Statistics,** and finally **Crosstabs** to create your cross-tabulation and analyze your data using chi-square (figure 1). This will open the following dialogue (figure 2):

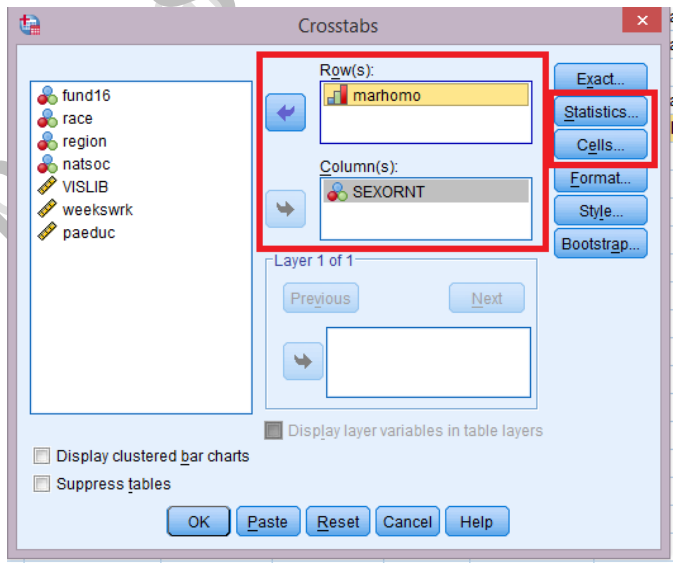


Next, select the variables you will use to create your crosstab. You should select your independent variable to be your **column variable**, and the dependent variable to be your **row variable.** Select the variable and then click the arrow next to the appropriate column or row box to move it to the appropriate section.

In this example, I selected **SEXORNT** as my column variable. SEXORNT is a nominal level variable created from the question, "Which of the following best describes you?" with the following categories: "gay, lesbian, or homosexual," "bisexual," and "heterosexual or straight." Respondents who reported "don't know," "refused," or "not applicable" were coded as missing.

For my dependent variable, I selected **marhomo** as my row variable. marhomo is an ordinal level variable created from responses to the statement, "homosexuals should have the right to marry." Respondents reported the following valid responses: "strongly agree," "agree," "neither agree nor disagree," "disagree," and "strongly disagree." Respondents who reported "cannot choose" or "not applicable" were coded as missing.

Figure 1

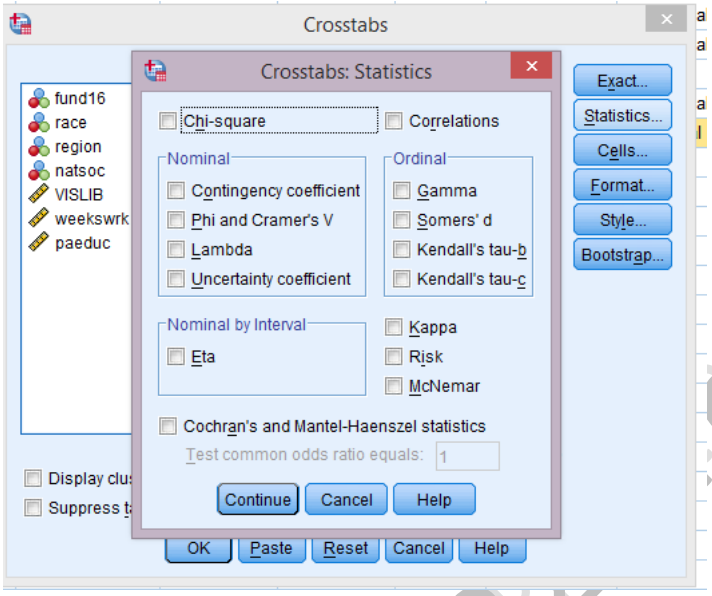


I will test the following hypotheses. The alpha has been set at 0.05.

- H_0: Sexual orientation and attitudes toward same sex marriage are statistically independent.
- H_1: Sexual orientation and attitudes toward same sex marriage are statistically dependent.

Next, you will need to click on the box labeled **Statistics** to tell SPSS which statistics you will calculate for your cross-tabulation. The following dialogue will appear:

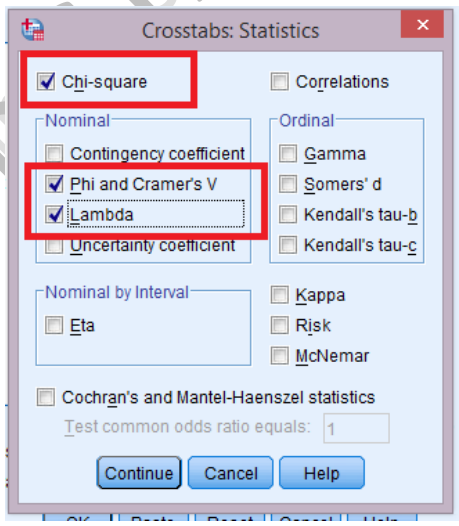
Figure 2



Next, you will click on the checkbox next to each of the statistics you will calculate. At bare minimum, you will select the box for chi-square.

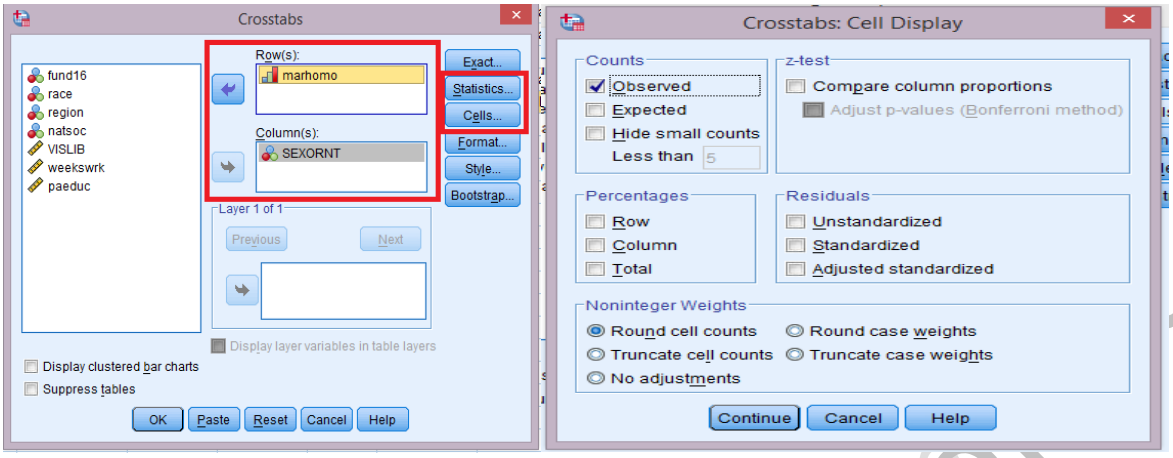
You will also select a check-box to calculate an appropriate **measure of association** to test the strength of the relationship between your two variables. For our purposes, I will check the boxes for **lambda** and **phi and Cramer's V.** Later in the guide, I will show you an example using two ordinal level variables so you can effectively read the output.

Figure 3



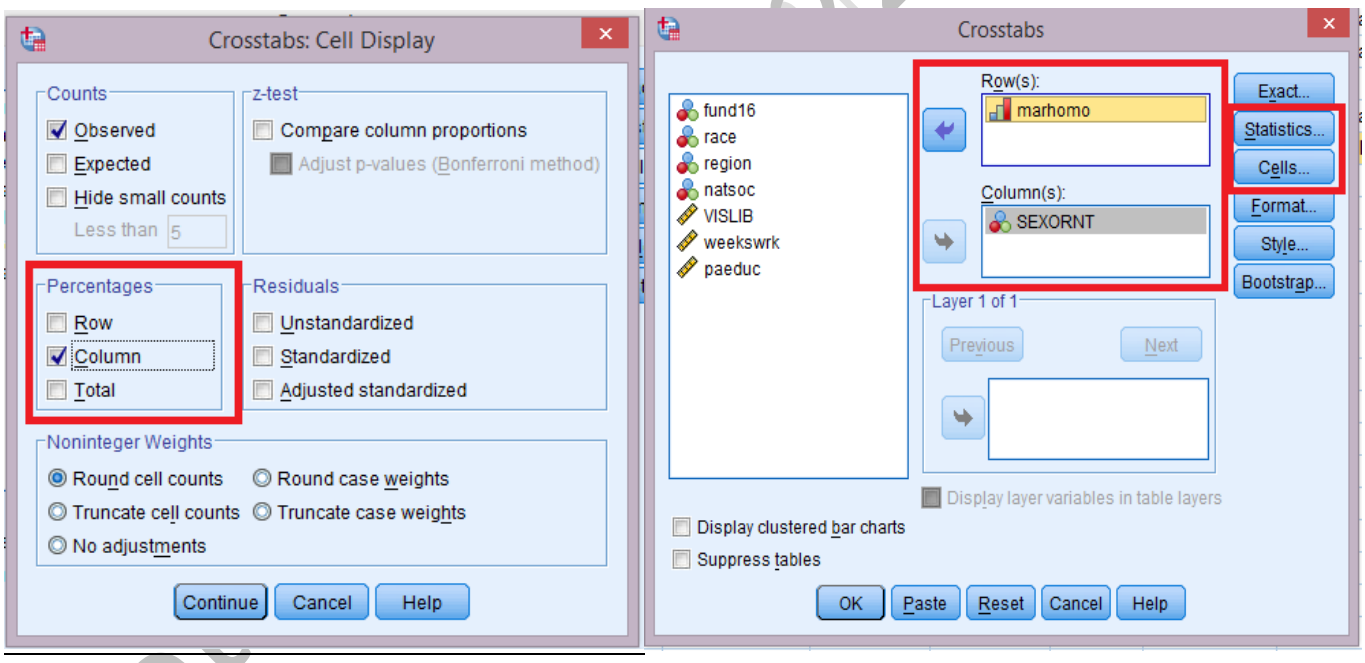
Click **Continue** and you will return to this screen (figure 6). Click **Cells.** The following **cell display** dialogue will appear (figure 7).

Figure 4 & Figure 5



By default, only **Observed** should be checked. This will report the number of people who fall into each cell. In addition to **observed,** you should also check the box for **column percentages** (Figure 8). *When you create your own table in Word or Excel, I only want to see the column percentages in your new table.* Click **Continue.** This will return you to the **crosstabs** screen (figure 9).

Figure 6 & Figure 7



Now you are ready to click **OK.** This will produce the output.
The **output screen** will show several boxes to you. Let's go through them one by one.

Output

The first box simply shows you how many **valid and missing cases** you have. For your purposes, the **Valid percent** should be 100%, and **missing** should be 0%.

Figure 8

Crosstabs

```
[DataSet3]
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY * SEXUAL ORIENTATION	1121	56.8%	853	43.2%	1974	100.0%

The next box shows you the **cross-tabulation.** We'll go through it one by one.

Figure 9

HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY * SEXUAL ORIENTATION Crosstabulation

			SEXUAL ORIENTATION			
			Gay, lesbian, or homosexual	Bisexual	Heterosexual or straight	Total
HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY	STRONGLY AGREE	Count	14	12	280	306
		% within SEXUAL ORIENTATION	66.7%	42.9%	26.1%	27.3%
	AGREE	Count	3	7	270	280
		% within SEXUAL ORIENTATION	14.3%	25.0%	25.2%	25.0%
	NEITHER AGREE NOR DISAGREE	Count	2	3	126	131
		% within SEXUAL ORIENTATION	9.5%	10.7%	11.8%	11.7%
	DISAGREE	Count	1	3	145	149
		% within SEXUAL ORIENTATION	4.8%	10.7%	13.5%	13.3%
	STRONGLY DISAGREE	Count	1	3	251	255
		% within SEXUAL ORIENTATION	4.8%	10.7%	23.4%	22.7%
Total		Count	21	28	1072	1121
		% within SEXUAL ORIENTATION	100.0%	100.0%	100.0%	100.0%

First, just look at the table and make sure your data are displayed appropriately. Ours are good! Notice the **cells:**

Figure 10

HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY * SEXUAL ORIENTATION Crosstabulation

			SEXUAL ORIENTATION			
			Gay, lesbian, or homosexual	Bisexual	Heterosexual or straight	Total
HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY	STRONGLY AGREE	Count	14	12	280	306
		% within SEXUAL ORIENTATION	66.7%	42.9%	26.1%	27.3%
	AGREE	Count	3	7	270	280
		% within SEXUAL ORIENTATION	14.3%	25.0%	25.2%	25.0%
	NEITHER AGREE NOR DISAGREE	Count	2	3	126	131
		% within SEXUAL ORIENTATION	9.5%	10.7%	11.8%	11.7%
	DISAGREE	Count	1	3	145	149
		% within SEXUAL ORIENTATION	4.8%	10.7%	13.5%	13.3%
	STRONGLY DISAGREE	Count	1	3	251	255
		% within SEXUAL ORIENTATION	4.8%	10.7%	23.4%	22.7%
Total		Count	21	28	1072	1121
		% within SEXUAL ORIENTATION	100.0%	100.0%	100.0%	100.0%

Just from looking at the cells, it looks like people who identify as gay, lesbian, or homosexual seem more likely than bisexuals or heterosexuals to agree that LGBT people should be able to marry. Bisexuals seem more

divided, but lean more heavily on agreeing that they should be allow to marry. Heterosexuals seem even more divided: about half agreeing, 11.8% do not agree or disagree, and more than a third disagree.
Take a look at the **totals:**

Figure 11

HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY * SEXUAL ORIENTATION Crosstabulation

			SEXUAL ORIENTATION			
			Gay, lesbian, or homosexual	Bisexual	Heterosexual or straight	Total
HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY	STRONGLY AGREE	Count	14	12	280	306
		% within SEXUAL ORIENTATION	66.7%	42.9%	26.1%	27.3%
	AGREE	Count	3	7	270	280
		% within SEXUAL ORIENTATION	14.3%	25.0%	25.2%	25.0%
	NEITHER AGREE NOR DISAGREE	Count	2	3	126	131
		% within SEXUAL ORIENTATION	9.5%	10.7%	11.8%	11.7%
	DISAGREE	Count	1	3	145	149
		% within SEXUAL ORIENTATION	4.8%	10.7%	13.5%	13.3%
	STRONGLY DISAGREE	Count	1	3	251	255
		% within SEXUAL ORIENTATION	4.8%	10.7%	23.4%	22.7%
Total		Count	21	28	1072	1121
		% within SEXUAL ORIENTATION	100.0%	100.0%	100.0%	100.0%

By now you should also notice that the sample is disproportionately heterosexual. If we only looked at the raw totals, we would not have been able to infer much about the relationship between sexual orientation and attitudes toward same sex marriage. **You will need to replicate this table on your own, using either Excel or Word to create the table for you.**
The next box shows you the chi-square statistics. We are interested in the information contained in the row labeled **Pearson Chi-Square.**

Figure 12 & Figure 13

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.355[a]	8	.004
Likelihood Ratio	20.985	8	.007
Linear-by-Linear Association	15.962	1	.000
N of Valid Cases	1121		

a. 5 cells (33.3%) have expected count less than 5. The minimum expected count is 2.45.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.355[a]	8	.004
Likelihood Ratio	20.985	8	.007
Linear-by-Linear Association	15.962	1	.000
N of Valid Cases	1121		

a. 5 cells (33.3%) have expected count less than 5. The minimum expected count is 2.45.

Our obtained chi-square statistic is 22.355. We have eight degrees of freedom. The P-value (labeled **asymp. sig. (2-sided)** is 0.004. With this information, we can *reject the null hypothesis.* Sexual orientation and attitudes toward same sex marriage are statistically dependent.
The next box shows us the **directional measures.** These are the first set of **measures of association** we calculated. In this box, we can find **lambda.** Since our dependent variable is **marhomo**, we need to refer to that section of the table.

Figure 14 & Figure 15

Directional Measures

		Value	Asymp. Std. Error[a]	Approx. T	Approx. Sig.
Lambda	Symmetric	.000	.000	.[b]	.[b]
	HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY Dependent	.000	.000	.[b]	.[b]
	SEXUAL ORIENTATION Dependent	.000	.000	.[b]	.[b]
Goodman and Kruskal tau	HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY Dependent	.006	.003		.000[c]
	SEXUAL ORIENTATION Dependent	.013	.006		.000[c]

ull hypothesis.

d because the asymptotic standard error equals zero.

e approximation

Directional Measures

		Value	Asymp. Std. Error[a]	Approx. T	Approx. Sig.
Lambda	Symmetric	.000	.000	.[b]	.[b]
	HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY Dependent	.000	.000	.[b]	.[b]
	SEXUAL ORIENTATION Dependent	.000	.000	.[b]	.[b]
Goodman and Kruskal tau	HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY Dependent	.006	.003		.000[c]
	SEXUAL ORIENTATION Dependent	.013	.006		.000[c]

ull hypothesis.

d because the asymptotic standard error equals zero.

e approximation

Lambda is zero! Why is that?

lambda will always be zero when the mode for *each category* of the independent variable falls into the same category of the independent variable – *even if other measures of association tell us that the two variables actually are related.* If the two variables seem related, based on the chi-square statistic or observations of the differences in percentages, we need to try a different measure of association to measure the strength of the relationship.

Lambda is not an adequate measure of association for our relationship. Here, we need to look at the row labeled **Cramer's V.** So let's take a look at the final table in our output:

Figure 16 & Figure 17

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.141	.004
	Cramer's V	.100	.004
N of Valid Cases		1121	

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.141	.004
	Cramer's V	.100	.004
N of Valid Cases		1121	

Do not worry about information in the column labeled **approx. sig.** Our Cramer's V is 0.10. Interpret appropriately. ☺

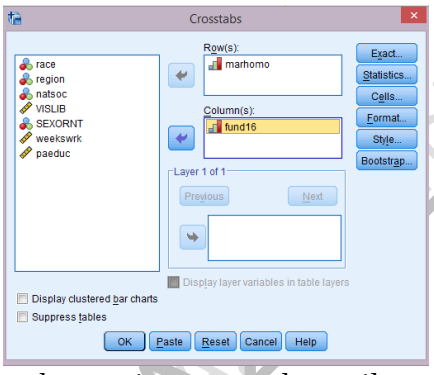
Chi-Square and Measures of Association for Two Ordinal Variables

If you select two ordinal level variables to complete this portion of the assignment, you will need to select a different measure of association. For this next test, I will continue to use **marhomo** as the dependent variable. The variable **fund16** is the independent variable. **Fund16** is an ordinal level variable that reports the "fundamentalism/liberalism of religion [the] respondent was raised in." The categories "fundamentalist" (1) means the religion was categorized as a very conservative denomination. "Moderate" (2) means the religion was categorized as being a more moderate religion. "Liberal" (3) means the religion was considered to be a liberal religious group. We can think of this as a continuum. Higher scores indicate more liberal religious beliefs at 16; lower scores indicate more conservative religious beliefs.

We will test the following hypotheses:

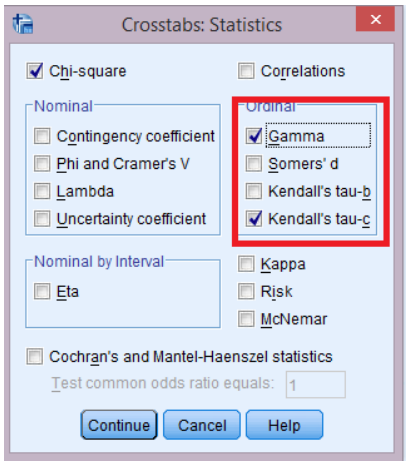
- H_0: Religious fundamentalism at 16 and attitudes toward same sex marriage are statistically independent.
- H_1: Religious fundamentalism at 16 and attitudes toward same sex marriage are statistically dependent.

Figure 18



We set up our chi-square test similarly to the previous example, until we get to the **statistics** box. We will still select a chi-square test to test the significance of the relationship, but with respect to the association, we need to select **gamma** and either **Kendall's tau-b** or **Kendall's tau-c.** Since there are five categories on our dependent variable, and only three for our column variable, we should select **Kendall's tau-c** in addition to gamma. This is because we use Kendall's tau-c when the cross-tabulation is a **rectangle.** If we had the same number of categories on both variables, we would use **Kendall's tau-b.**

Figure 19



Output

Let's take a look at the cross-tabulation:

Figure 20

HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY * HOW FUNDAMENTALIST WAS R AT AGE 16 Crosstabulation

			HOW FUNDAMENTALIST WAS R AT AGE 16			
			FUNDAMENTALIST	MODERATE	LIBERAL	Total
HOMOSEXUALS SHOULD HAVE RIGHT TO MARRY	STRONGLY AGREE	Count	63	164	83	310
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	17.9%	26.6%	30.2%	24.9%
	AGREE	Count	61	181	67	309
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	17.4%	29.3%	24.4%	24.9%
	NEITHER AGREE NOR DISAGREE	Count	44	67	36	147
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	12.5%	10.9%	13.1%	11.8%
	DISAGREE	Count	62	84	32	178
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	17.7%	13.6%	11.6%	14.3%
	STRONGLY DISAGREE	Count	121	121	57	299
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	34.5%	19.6%	20.7%	24.1%
Total		Count	351	617	275	1243
		% within HOW FUNDAMENTALIST WAS R AT AGE 16	100.0%	100.0%	100.0%	100.0%

It seems that something *may* be going on here. A little over one-third of people raised in fundamentalist religions agree that same-sex couples should be allowed to marry. Twelve and a half percent do not agree or disagree, while over half disagree with the statement that same-sex couples should be allowed to marry. Over one-half of those raised in moderate and liberal denominations agreed that same-sex couples should be allowed to marry. So there might be something going on here. Keep in mind that on the **fund16** variable, *lower scores = more conservative beliefs; higher scores = more liberal beliefs*. On our **marhomo** variable, *lower scores = agreement that same-sex couples should be allowed to marry* and *higher scores = disagreement that same-sex couples should be allowed to marry.* With this in mind, and with the evidence presented above, it seems that if the two variables are statistically dependent, we are likely to have a *negative relationship.* As the independent variable score increases, the dependent variable score decreases.

The chi-square test suggests that we should reject our null hypothesis. Religious fundamentalism and attitudes toward same-sex marriage are statistically dependent. Now we can examine the box labeled **symmetric measures.** Here we can find our gamma and the Kendall's tau-c.

Figure 21 & Figure 22

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	50.991[a]	8	.000
Likelihood Ratio	50.623	8	.000
Linear-by-Linear Association	31.330	1	.000
N of Valid Cases	1243		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 32.52.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-c	-.143	.026	-5.576	.000
	Gamma	-.193	.034	-5.576	.000
N of Valid Cases		1243			

a. Not assuming the null hypothesis.

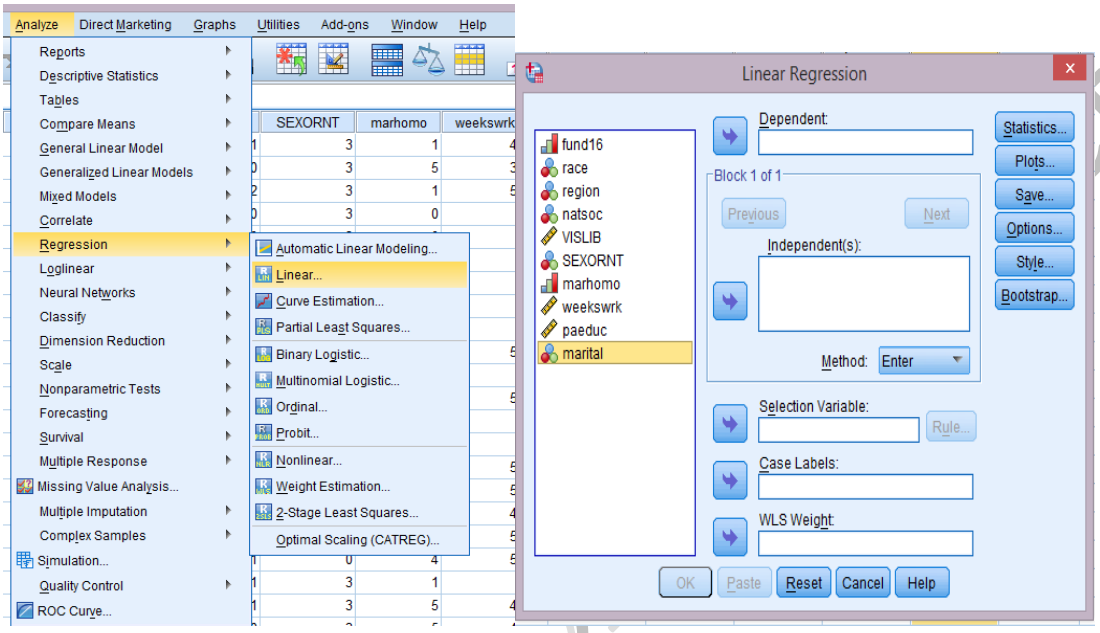
b. Using the asymptotic standard error assuming the null hypothesis.

We are only interested in the information contained in the **value** column. Our Kendall's tau-c is -0.143 and our gamma is -0.193. Interpret appropriately. ☺

Regression and Correlation

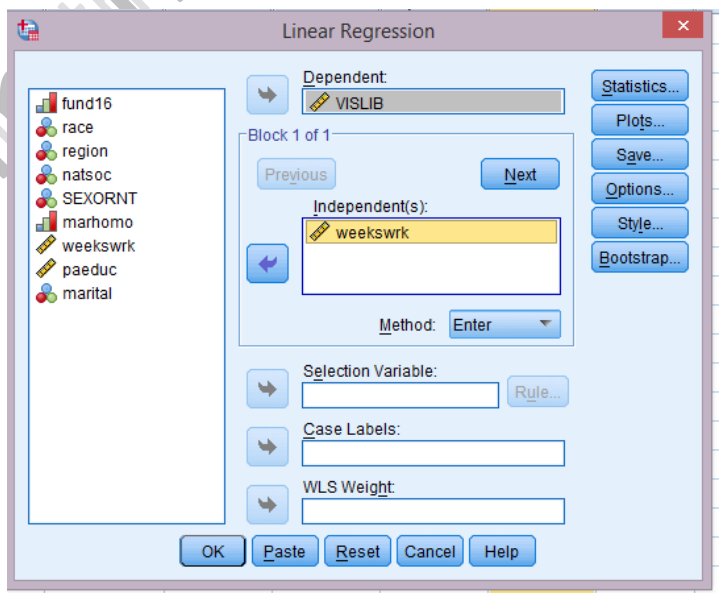
To begin, click **Analyze,** then **Regression,** and finally **linear** to complete the regression and correlation portion of Project 4. The following **linear regression** dialogue will appear:

Figure 23 & Figure 24



You will need to select two appropriate variables to obtain your regression line equation. Remember, your independent variable is the one you think will predict a change in the dependent variable.
In this example, **weekswrk** is an interval-ratio level variable reporting the number of weeks a respondent worked last year. It is the independent variable. **VISLIB** is an interval-ratio level variable reporting the number of times a respondent visited a public library last year.

Figure 25



All we need to do to get the information to report our linear regression equation, the correlation coefficient, and the coefficient of determination, is click **OK.** Your output will appear! I'll go through each box one by one.

Figure 26

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	WEEKS R. WORKED LAST YEAR[b]	.	Enter

a. Dependent Variable: HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR

b. All requested variables entered.

The first box just shows the variables in the equation and the method used to enter them (you don't need to worry about this). Double check that the **variables entered** matches your independent variable and the dependent variable **below the box** matches what you intended to do. We're all good!

The **Model Summary** box reports Pearson's correlation coefficient (**R**) and the coefficient of determination ($\mathbf{R^2}$).

Figure 27

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.036[a]	.001	.000	12.061

a. Predictors: (Constant), WEEKS R. WORKED LAST YEAR

Hold off on interpreting the correlation coefficient for now. We can also see that our regression equation demonstrates that only 0.1% of the variation in visits to the library last year is explained by how many weeks worked last year.

The next box shows us **ANOVA.** ANOVA and regression are related! For this portion of the assignment, do not worry about this box. This box shows us if there is a significant relationship in our regression (we did not cover this in our class – we can conduct hypothesis tests with regression, too!). In short: there isn't a statistically significant relationship.

Figure 28

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	184.299	1	184.299	1.267	.261[b]
	Residual	141406.083	972	145.480		
	Total	141590.382	973			

a. Dependent Variable: HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR

b. Predictors: (Constant), WEEKS R. WORKED LAST YEAR

The final box, labeled **coefficients,** contains the information you need to report your regression line equation.

Figure 29

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.027	.630		11.146	.000
	WEEKS R. WORKED LAST YEAR	-.019	.016	-.036	-1.126	.261

a. Dependent Variable: HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR

The information contained under the column labeled **B** shows us both the slope (***b***) and the **y-intercept** (***a***). The row labeled **(Constant)** shows us the *constant* of the regression line. This is different language than you are already familiar. The number in the cell where B and (Constant) meet is the **y-intercept.**
The row labeled **WEEKS R WORKED LAST YEAR** shows us various statistics relevant to how the independent variable is related to the dependent variable. In the cell where B and WEEKS R WORKED LAST YEAR meet, we are shown **the slope of the line** (b). For each unit increase in the number of weeks someone worked last year, we expect a decrease in the number of library visits last year of 0.019.
Let's return to the correlation box I showed you earlier.

Figure 30

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.036[a]	.001	.000	12.061

a. Predictors: (Constant), WEEKS R. WORKED LAST YEAR

The correlation coefficient presented here is *only going to show a positive figure (this is what you should expect; the explanation is beyond the scope of this class).* However, based on the slope of our regression line, we know that we actually have a *negative relationship* – as X increases, Y decreases. **When you report your correlation coefficient, make sure you report the appropriate sign. In this case, we know we have a very weak negative relationship (R = -0.036).**
Just to confirm, you can *but do not have to* calculate the **bivariate correlation** (click **Analyze,** then **Correlation**, and then **Bivariate Correlation**). Take a look at what we find:

Figure 31

Correlations

Correlations

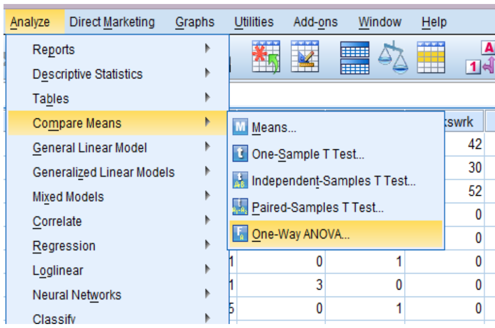
		HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR	WEEKS R. WORKED LAST YEAR
HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR	Pearson Correlation	1	-.036
	Sig. (2-tailed)		.261
	N	978	974
WEEKS R. WORKED LAST YEAR	Pearson Correlation	-.036	1
	Sig. (2-tailed)	.261	
	N	974	1963

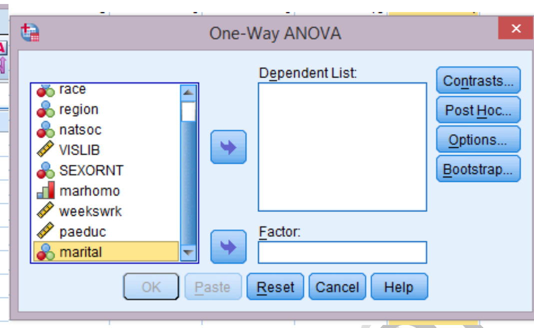
There it is! We can see there is a *negative correlation of -0.36!*

ANALYSIS OF VARIANCE (ANOVA)

To begin our ANOVA, you will first click **Analyze,** then **Compare Means,** and finally **One-way ANOVA.** The following **One-way ANOVA** dialogue will appear:

Figure 32 & Figure 33



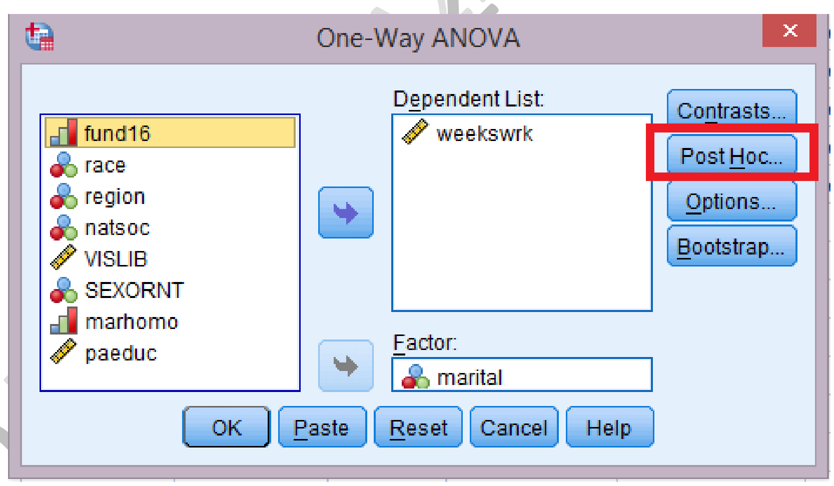


Under **dependent list,** you will select your dependent variable. For our purposes, I'll use our **weekswrk** variable from earlier. Under **factor,** you will select your grouping variable. I'm selecting **marital,** which is a nominal level variable reporting respondents' marital status: married, widowed, divorced, separated, or never married. This is the variable you will use to see if there are differences in the mean number of weeks worked last year *by* **marital** status**.**

I'll test the following hypotheses:

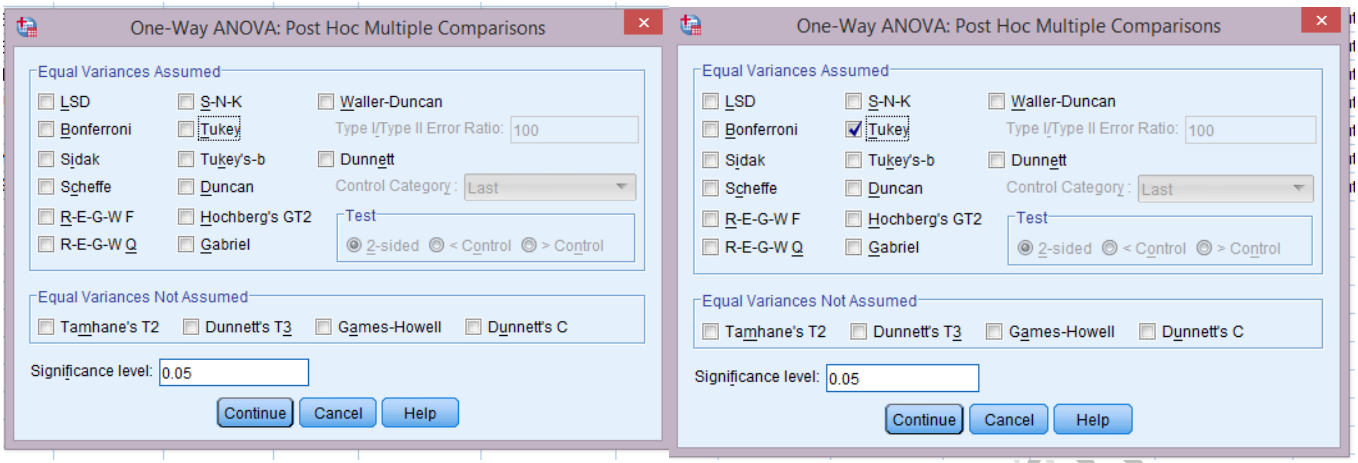
- H_0: There are no differences in mean weeks worked last year by marital status.
- H_1: There is at least one difference in the mean number of weeks worked last year by marital status.

Figure 34



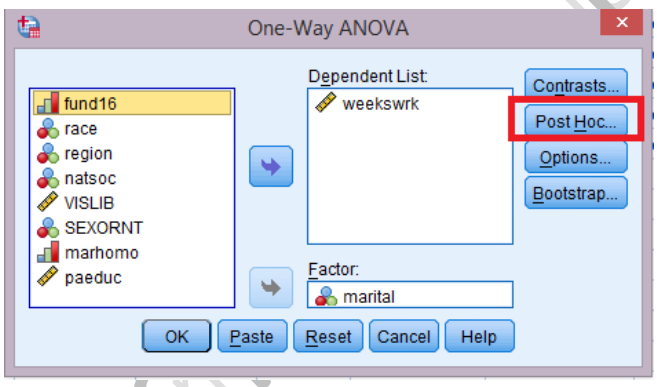
From here, click on the box that says **Post Hoc.** Your book doesn't discuss post-hoc tests, but they are very useful in figuring out *which* groups differ and which do not. The following **Post Hoc Multiple Comparisons** dialogue will appear:

Figure 35 & Figure 36



There are lots of different tests we can use to see which groups differ. Click the box labeled simply **Tukey.** Then click **Continue.** This will bring you back to the **One-Way ANOVA** dialogue. Click **OK.**

Figure 37



Now your output will appear. We'll go through it one by one. The first box shows you the sum of squares, mean squares, degrees of freedom, the obtained F-statistic, and the p-value. Everything you need and are already comfortable with calculating by hand. ☺

Figure 38

Oneway

ANOVA

WEEKS R. WORKED LAST YEAR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	81869.007	4	20467.252	40.814	.000
Within Groups	981881.356	1958	501.472		
Total	1063750.363	1962			

In Project 4, I've asked you to report **df_b, df_w, MSB, MSW,** the obtained **F-statistic,** and the **p-value.** It's all right here! We can *reject the null hypothesis* because our ANOVA demonstrates there is at least *one* difference in mean number of weeks worked last year. **If you did not find a significant relationship,** you're pretty much done. **If you did find a significant relationship,** you need to take a look at the next bit of output. The **Post Hoc Tests** output comes next.

Figure 39

Post Hoc Tests

Multiple Comparisons

Dependent Variable: WEEKS R. WORKED LAST YEAR

Tukey HSD

(I) MARITAL STATUS	(J) MARITAL STATUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
MARRIED	WIDOWED	23.317*	1.907	.000	18.11	28.52
	DIVORCED	1.543	1.469	.832	-2.47	5.55
	SEPARATED	6.034	2.817	.203	-1.66	13.72
	NEVER MARRIED	-.468	1.233	.996	-3.83	2.90
WIDOWED	MARRIED	-23.317*	1.907	.000	-28.52	-18.11
	DIVORCED	-21.775*	2.162	.000	-27.68	-15.87
	SEPARATED	-17.283*	3.233	.000	-26.11	-8.46
	NEVER MARRIED	-23.785*	2.009	.000	-29.27	-18.30
DIVORCED	MARRIED	-1.543	1.469	.832	-5.55	2.47
	WIDOWED	21.775*	2.162	.000	15.87	27.68
	SEPARATED	4.491	2.995	.563	-3.69	12.67
	NEVER MARRIED	-2.010	1.599	.717	-6.38	2.35
SEPARATED	MARRIED	-6.034	2.817	.203	-13.72	1.66
	WIDOWED	17.283*	3.233	.000	8.46	26.11
	DIVORCED	-4.491	2.995	.563	-12.67	3.69
	NEVER MARRIED	-6.502	2.887	.161	-14.38	1.38
NEVER MARRIED	MARRIED	.468	1.233	.996	-2.90	3.83
	WIDOWED	23.785*	2.009	.000	18.30	29.27
	DIVORCED	2.010	1.599	.717	-2.35	6.38
	SEPARATED	6.502	2.887	.161	-1.38	14.38

*. The mean difference is significant at the 0.05 level.

The first column shows us all five relationship status categories. Notice that it is labeled **(I) MARITAL STATUS.** The next column shows us *each of the other four relationship status categories, relative to the category reported in the first column.* It is labeled **(J) MARITAL STATUS.** The third column labeled **Mean Differences (I-J)** shows us the *mean difference in weeks worked last year, subtracting the mean number of weeks worked for the first category (labeled I) from the mean number of weeks worked for the second category (labeled J).* Let's examine the first row:

Figure 40

Post Hoc Tests

Multiple Comparisons

Dependent Variable: WEEKS R. WORKED LAST YEAR

Tukey HSD

(I) MARITAL STATUS	(J) MARITAL STATUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
MARRIED	WIDOWED	23.317*	1.907	.000	18.11	28.52
	DIVORCED	1.543	1.469	.832	-2.47	5.55
	SEPARATED	6.034	2.817	.203	-1.66	13.72
	NEVER MARRIED	-.468	1.233	.996	-3.83	2.90
WIDOWED	MARRIED	-23.317*	1.907	.000	-28.52	-18.11
	DIVORCED	-21.775*	2.162	.000	-27.68	-15.87
	SEPARATED	-17.283*	3.233	.000	-26.11	-8.46
	NEVER MARRIED	-23.785*	2.009	.000	-29.27	-18.30
DIVORCED	MARRIED	-1.543	1.469	.832	-5.55	2.47
	WIDOWED	21.775*	2.162	.000	15.87	27.68
	SEPARATED	4.491	2.995	.563	-3.69	12.67
	NEVER MARRIED	-2.010	1.599	.717	-6.38	2.35
SEPARATED	MARRIED	-6.034	2.817	.203	-13.72	1.66
	WIDOWED	17.283*	3.233	.000	8.46	26.11
	DIVORCED	-4.491	2.995	.563	-12.67	3.69
	NEVER MARRIED	-6.502	2.887	.161	-14.38	1.38
NEVER MARRIED	MARRIED	.468	1.233	.996	-2.90	3.83
	WIDOWED	23.785*	2.009	.000	18.30	29.27
	DIVORCED	2.010	1.599	.717	-2.35	6.38
	SEPARATED	6.502	2.887	.161	-1.38	14.38

*. The mean difference is significant at the 0.05 level.

First, we are looking at the mean difference in the number of weeks worked last year between **married respondents** and **widowed respondents.** The value reported in the first cell under the **Mean Difference (I-J) column** is calculated using the following formula:

$$Married - Widowed = mean\ difference\ in\ weeks\ worked$$

In this equation, I = Married and J = Widowed. The mean difference is 23.317 weeks.
This means that **married respondents worked an average of 23.317 more weeks than widowed respondents did last year.** Now take a look at the value under the **sig.** column. This is the P-value! The P-value is 0.000.
This means that at the alpha = 0.05 level, we can confirm that **married respondents worked significantly more weeks last year than widowed respondents.**
You need to do this for each row. Notice there is **repeated information in the table.** Let's look at the row where I = WIDOWED and J = MARRIED.

Figure 41

Post Hoc Tests

Multiple Comparisons

Dependent Variable: WEEKS R. WORKED LAST YEAR

Tukey HSD

(I) MARITAL STATUS	(J) MARITAL STATUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
MARRIED	WIDOWED	23.317*	1.907	.000	18.11	28.52
	DIVORCED	1.543	1.469	.832	-2.47	5.55
	SEPARATED	6.034	2.817	.203	-1.66	13.72
	NEVER MARRIED	-.468	1.233	.996	-3.83	2.90
WIDOWED	MARRIED	-23.317*	1.907	.000	-28.52	-18.11
	DIVORCED	-21.775	2.162	.000	-27.68	-15.87
	SEPARATED	-17.283*	3.233	.000	-26.11	-8.46
	NEVER MARRIED	-23.785*	2.009	.000	-29.27	-18.30
DIVORCED	MARRIED	-1.543	1.469	.832	-5.55	2.47
	WIDOWED	21.775*	2.162	.000	15.87	27.68
	SEPARATED	4.491	2.995	.563	-3.69	12.67
	NEVER MARRIED	-2.010	1.599	.717	-6.38	2.35
SEPARATED	MARRIED	-6.034	2.817	.203	-13.72	1.66
	WIDOWED	17.283*	3.233	.000	8.46	26.11
	DIVORCED	-4.491	2.995	.563	-12.67	3.69
	NEVER MARRIED	-6.502	2.887	.161	-14.38	1.38
NEVER MARRIED	MARRIED	.468	1.233	.996	-2.90	3.83
	WIDOWED	23.785*	2.009	.000	18.30	29.27
	DIVORCED	2.010	1.599	.717	-2.35	6.38
	SEPARATED	6.502	2.887	.161	-1.38	14.38

*. The mean difference is significant at the 0.05 level.

From this, we can see that **widowed respondents worked significantly fewer weeks last year than married respondents did.** They worked 23.317 weeks fewer, on average. **Hey! That's the reciprocal value!** ☺
Let's identify each set that significantly differed:

Figure 42

Post Hoc Tests

Multiple Comparisons

Dependent Variable: WEEKS R. WORKED LAST YEAR

Tukey HSD

(I) MARITAL STATUS	(J) MARITAL STATUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
MARRIED	WIDOWED	23.317*	1.907	.000	18.11	28.52
	DIVORCED	1.543	1.469	.832	-2.47	5.55
	SEPARATED	6.034	2.817	.203	-1.66	13.72
	NEVER MARRIED	-.468	1.233	.996	-3.83	2.90
WIDOWED	MARRIED	-23.317*	1.907	.000	-28.52	-18.11
	DIVORCED	-21.775*	2.162	.000	-27.68	-15.87
	SEPARATED	-17.283*	3.233	.000	-26.11	-8.46
	NEVER MARRIED	-23.785*	2.009	.000	-29.27	-18.30
DIVORCED	MARRIED	-1.543	1.469	.832	-5.55	2.47
	WIDOWED	21.775*	2.162	.000	15.87	27.68
	SEPARATED	4.491	2.995	.563	-3.69	12.67
	NEVER MARRIED	-2.010	1.599	.717	-6.38	2.35
SEPARATED	MARRIED	-6.034	2.817	.203	-13.72	1.66
	WIDOWED	17.283*	3.233	.000	8.46	26.11
	DIVORCED	-4.491	2.995	.563	-12.67	3.69
	NEVER MARRIED	-6.502	2.887	.161	-14.38	1.38
NEVER MARRIED	MARRIED	.468	1.233	.996	-2.90	3.83
	WIDOWED	23.785*	2.009	.000	18.30	29.27
	DIVORCED	2.010	1.599	.717	-2.35	6.38
	SEPARATED	6.502	2.887	.161	-1.38	14.38

*. The mean difference is significant at the 0.05 level.

Notice anything? Widowed respondents worked significantly fewer weeks last year than married, divorced, separated, and never married respondents. Perhaps even more interesting, the only significant differences involved widowed respondents. *What might explain this? It is probable that most of the widowed respondents are elderly, and thus at retirement age already.* Make sure you report all significant differences appropriately. ☺

SECTION D: RUBRIC

PART ONE: CROSS-TABULATION and CHI-SQUARE				
1. Description of independent and dependent variables (5 points) FULL CREDIT: Correct variables are selected and explanation is logical and clear. HALF CREDIT: Correct variables are selected but explanation is less logical or incomplete. NO CREDIT: Incorrect variables selected.	5	2.5	0	0
2. Hypothesis statements (5 points) FULL CREDIT: Null and research hypotheses are correct and clearly labeled. Appropriate statistical language is used. HALF CREDIT: Null and research hypotheses contain muddled language or may confuse a symbol or word. NO CREDIT: Erroneous and incorrect.	5	2.5	0	0
3. Tables (5 points) This criterion is linked to a Learning Outcome 3. Report cross-tabulation. FULL CREDIT: Cross-tabulation created using Word or Excel. Column percentages reported. Total row and total column reported. HALF CREDIT: Table isn't straightforward or maybe missing a component. NO CREDIT: Table is entirely inaccurate, missing, or copied	5	2.5	0	0
4. Annotated test statistics (2.5 points) FULL CREDIT: Report and explain the obtained Pearson chi-square statistic, the degrees of freedom, the P-value, and the chi-square critical value. HALF CREDIT: Report statistics with lax or nonsensical annotations. NO CREDIT: Missing, incorrect, or mislabeled statistics.	2.5	1.5	0	0
5. Hypothesis decision (5 points) FULL CREDIT: Correct decision reported using appropriate statistical language. HALF CREDIT: Logical decision based on stats that aren't appropriate or muddled statistical language.	5	2.5	0	0
6. Statement of test strength (2.5 points) FULL CREDIT: Reported at least one applicable measure of association for variables. OR, if appropriate, one was not reported. In this case, you must state one you would have reported if it were appropriate. NO CREDIT: Reports inappropriate measure of association or none at all.	2.5		0	

7. Reflection (10 points) This is your opportunity to demonstrate what you know about the data, the statistics you performed, and the information you've inferred from those statistics (that is, what do they SAY?). CHECK-PLUS. Above-and-beyond effort. Response demonstrates mastery of subject matter. Excellent explanation of accurate results in line with the information provided. The interpretation is between 5 and 7 sentences length. CHECK. Typical effort. Response demonstrates average, though correct, understanding of subject matter. Appropriate, though superficial explanation of the results. They are in line with the information provided. The interpretation is between 5 and 7 sentences in length. CHECK-MINUS. Below-average effort. Response demonstrates incomplete or incorrect understanding of subject matter. The explanation is not effective. It is short or demonstrates significant misunderstandings of the results.	10	7.5	5	0
PART TWO: ANALYSIS OF VARIANCE (ANOVA)				
8. Description of independent and dependent variables (5 points) FULL CREDIT: Correct variables are selected and explanation is logical and clear. HALF CREDIT: Correct variables are selected but explanation is less logical or incomplete. NO CREDIT: Incorrect variables selected.	5	2.5	0	0
9. Hypothesis statements (5 points) FULL CREDIT: Correct decision reported using appropriate statistical language. HALF CREDIT: Logical decision based on stats that aren't appropriate or muddled statistical language.	5	2.5	0	0
10. Annotated test statistics (5 points) FULL CREDIT: All of the following are reported and labeled: SSB, SSW, dfb, dfw, MSb, MSw, obtained F-statistic, P-value, F-critical HALF CREDIT: Missing or mislabeled statistics (less than half).	5	2.5	0	0
11. Hypothesis decision (5 points) FULL CREDIT: Null and research hypotheses are correct and clearly labeled. Appropriate statistical language is used. HALF CREDIT: Null and research hypotheses contain muddled language or may confuse a symbol or word. NO CREDIT: Erroneous and incorrect.	5	2.5	0	0
12. Examination of multiple relationships (5 points) FULL CREDIT: All groups with statistically significant mean differences and their corresponding P-values have been reported and labeled. If none of the groups significantly differed, this is stated. HALF CREDIT: All statistically significant differences have been reported, but issues with labeling OR p-values were omitted.	5	2.5	0	0
13. Reflection (10 points) This is your opportunity to demonstrate what you know about the data, the statistics you performed, and the information you've inferred from those statistics (that is, what do they SAY?). CHECK-PLUS. Above-and-beyond effort. Response demonstrates mastery of subject matter. Excellent explanation of accurate results in line with the information provided. The interpretation is between 5 and 7 sentences length. CHECK. Typical effort. Response	10	7.5	5	0

demonstrates average, though correct, understanding of subject matter. Appropriate, though superficial explanation of the results. They are in line with the information provided. The interpretation is between 5 and 7 sentences in length. CHECK-MINUS. Below-average effort. Response demonstrates incomplete or incorrect understanding of subject matter. The explanation is not effective. It is short or demonstrates significant misunderstandings of the results.				
PART THREE: REGRESSION AND CORRELATION				
14. Description of independent and dependent variables (5 points) FULL CREDIT: Correct variables are selected and explanation is logical and clear. HALF CREDIT: Correct variables are selected but explanation is less logical or incomplete. NO CREDIT: Incorrect variables selected.	5	2.5	0	0
15. Statement and explanation of regression equation (5 points) FULL CREDIT: The regression equation is effectively explained in two to five sentences. HALF CREDIT: Regression equation explained superficially or using muddled language.	5	2.5	0	0
16. Report and description of CORRELATION COEFFICIENT (5 points) FULL CREDIT: Interpretation of the correlation coefficient is effective and correct. HALF CREDIT: Interpretation, though correct, does not demonstrate statistical literacy. NO CREDIT: The interpretation is not effective and does not use appropriate statistical language.	5	2.5	0	0
17. Report and description of COEFFICIENT OF DETERMINATION (5 points) FULL CREDIT: Interpretation of the coefficient of determination is effective and correct. HALF CREDIT: Interpretation, though correct, does not demonstrate statistical literacy. NO CREDIT: The interpretation is not effective and does not use appropriate statistical language.	5	2.5	0	0
18. Reflection (10 points) This is your opportunity to demonstrate what you know about the data, the statistics you performed, and the information you've inferred from those statistics (that is, what do they SAY?). CHECK-PLUS. Above-and-beyond effort. Response demonstrates mastery of subject matter. Excellent explanation of accurate results in line with the information provided. The interpretation is between 5 and 7 sentences length. CHECK. Typical effort. Response demonstrates average, though correct, understanding of subject matter. Appropriate, though superficial explanation of the results. They are in line with the information provided. The interpretation is between 5 and 7 sentences in length. CHECK-MINUS. Below-average effort. Response demonstrates incomplete or incorrect understanding of subject matter. The explanation is not effective. It is short or demonstrates significant misunderstandings of the results.	10	7.5	5	0

SECTION E: EXAMPLE PROJECT

The following example demonstrates how a PROJECT might look. The data in this sample is derived from another source and some information (values and explanations) have been redacted to reduce copying.

NOTE: Your final report will contain more detail (and use different data) than what is presented here -- mostly so I can assess your statistical literacy.

PART ONE: CROSS-TABULATION and CHI-SQUARE

1. **Introduction and descriptive statistics.**

For this project I decided to assess the relationship between number of children, the independent variable, and the respondent's perception of government support for daycare services, the dependent variable. I believe that the more children a person has, the more inclined they will be to support help with daycare because it's terribly expensive! My sample size approaches 1000, which makes it pretty representative of the US population (source cite here). It looks like the data is slightly skewed because <redacted> and the typical person has about 2 kids, which is what we tend to hear people have (http://www.pewsocialtrends.org/2015/05/07/family-size-among-mothers/). The best measure of central tendency is _____ because _____. But the other ones work better/worse than _____*For the purposes of illustration, I re-coded number of children, usually an **interval-ratio** variable, into an **ordinal** variable. The categories are 0, 1, 2, and 3 or more children.

- Population size (n=919)
- Mean=2.89; median=2; mode= 0; standard deviation= <redacted>
- standard error= <redacted>
- Minimim=0; Maximum=3; range=3

2. Hypothesis statement.

H_0: There is no association between number of children and daycare support in the population.
H_1: The number of children and daycare support are statistically dependent.

3. Table

Table 1: Perception of Childcare Support by Number of Children

	Number of Children				
Support	0	1	2	3 or More	Total
Too Little	138	72	125	106	441
	47.4%	51.8%	45.5%	49.5%	48.0%
About Right	124	60	118	88	390
	42.6%	43.2%	42.9%	41.1%	42.4%
Too Much	29	7	32	20	88
	10.0%	5.0%	11.6%	9.3%	9.6%
Total	291	139	275	214	919

	100.0%	100.0%	100.0%	100.0%	100.0%

4. Annotated test statistics.

- Pearson chi-square statistic: 5.305. This statistic is the _____.
- Degrees of freedom: 6. The df allows us _____.
- P-value: 0.505. This p-value tells us _____.
- Chi-square critical value: 12.592 (Alpha = .05; 6 df) The critical value is _____.

5. Hypothesis decision.

I failed to reject the null hypothesis: The number of children a person has and their perception of childcare support by the government isn't related.

6. Statement of test strength.

Here you need to choose the CORRECT measure of Association and Statistical Test. Please refer to the How-to instructions and the memo for bivariate relationships for instructions on how to assess crosstabs with ordinal and/or nominal data.

For this test, I used Gamma and Kendall's Tau-C.

Gamma: 0.000

There is no relationship between number of children a person has and their perception of childcare support by the government ($\gamma = 0.00$).

Kendall's Tau-C: 0.000

There is no association between number of children a person has and their perception of childcare support by the government ($\tau_b = 0.00$). Using information about respondents' number of children and their perception of childcare support by the government, we did not reduce our prediction error.

7. Reflection.

Here, I expect a full paragraph describing the relationship you thought you would find, the information in the table, (i.e., the relationship you thought you had and the one you observed in the data. I also want a somewhat detailed discussion of significance and measures of association. Mine will be brief so that you don't copy it in your answers. So, I expected that groups with greater numbers of kids would be associated with perceptions of "not enough" childcare support. For example, I thought that folks with 0 or 1 kid would perceive that there is just enough or too much support compared with people that had 2 or 3 kids -- I was wrong. Nearly across the board, citizens perceived that the government provides either too little or just enough support for childcare. I think the categories of children might have been coded with greater distance, there might have been a difference -- that is, would you think support would differ that much for people with 1 versus 2 kids? Unlikely. For more information on childcare resources and spending, please visit: https://www.usa.gov/child-care. HINT: It's not much.

PART TWO: ANALYSIS OF VARIANCE (ANOVA)

8. Description of independent and dependent variables

In a few sentences, explain why you think your independent variable might influence the dependent variable. You would need to elaborate more here, but overall, I think that if people are married, they will likely spend less time on the internet compared with those that are widowed, divorced, separated, or never married. Include a description of the relevant statistics, including:

- Population size (N=___)
- mean, median, mode, standard deviation,
- standard error,
- the minimum/maximum values, and the range.

9. Hypothesis statement.

- H_0: $\mu 1 = \mu 2 = \ldots \mu k$: There is no difference in the mean hours spent on the Internet by marital status.
- H_1: There is at least one mean difference in the hours spent on the internet by marital status.

10. Annotated test statistics.

- SSB: 4119.326 This statistic is the _____. (explain all)
- SSW: 2000916.674
- df_b: 4
- df_w: 920
- MS_b: 1029.831
- MS_w: 218.388
- Obtained F-statistic: 4.716
- P-value: .001
- F-critical value: 2.37

11. Hypothesis decision.

My F-obtained exceeds my F-critical, which means I can reject my null hypothesis -- there is at least one mean difference in the amount of hours spent on the internet by marital status.

12. Examination of multiple relationships.

Examine the table created from Tukey's HSD. In a list, report all statistically significant differences (i.e., p<.05). Report the mean difference and the P-value for each statistically significant difference. HINT: If there are none, you must clearly state there are no statistically significant differences between groups in order to receive credit for this portion of the assignment.

Difference between Married and Never Married: -4.080 hours; p value: 0.003

13. Reflection.

You would need to elaborate more here, but I found out there were mean differences in the hours spent on the internet among marital status groups; however the relationship was not as clear as I suspected.

Only one group significantly differed: Married and non-married. Married people spend a little over 4 fewer hours on the internet compared with never married respondents. Other groups were not significantly different from each other; however, some were close (e.g., widowed and never married and divorced and never married). What this finding also means is that married, divorced, separated, and widowed people spend about the same amount of time on the internet!

PART THREE: REGRESSION AND CORRELATION

14. Description of independent and dependent variables.

In a few sentences, explain why you think your independent variable might influence the dependent variable. You would need to elaborate more here, but overall, I think the more time people spend on the internet, would mean that there is less time spent on household labor (dependent variable) Include a description of the relevant statistics, including:

- Population size (N=___)
- mean, median, mode, standard deviation,
- standard error,
- the minimum/maximum values, and the range.

15. Statement and explanation of regression equation.

y-hat = 9.711 =+ 0.31 (x)

This equation means that a one-hour increase in time spent on the Internet (independent variable) results in a 0.31 hour *increase* in time spend on household work. Essentially, the more time spent on the internet, the more time spent on household labor. I did not expect this relationship at all -- I thought that more time on the Internet would result in less time performing household labor.

16. Report and description of CORRELATION COEFFICIENT.

The correlation coefficient is .045. NOTE: You will have to report the CORRECT NAME OF THIS COEFFICIENT. This _____ coefficient indicates that there is a weak, yet positive association between the hours spent on the internet and hours spent performing housework.

17. Report and description of COEFFICIENT OF DETERMINATION

The coefficient of determination is .002. NOTE: You will have to report the CORRECT NAME OF THIS COEFFICIENT. Nearly NONE of the variation in hours spent performing housework is explained by hours spent on the internet. NOTE: You will need to use NUMBERS for your interpretation as well as a line or two on what that means for your analysis generally.

18. Reflection.

In three to seven sentences, explain what the findings from all of your statistics in words. Make sure you use appropriate statistical language and describe what the findings actually mean, and critically reflect on your population. You are not required to use resources, but they will strengthen your arguments. Here, I would reflect on why it is that I think that time spend on the internet is positively related to time spent doing household labor!