Psychology Research Methods
Lab Session – Week 5

SPSS Correlation/Regression

Due at the Start of Lab: None

Rationale for Today's Lab Session

This tutorial is designed to ensure that you have mastered basic correlational analyses and have a basic understanding of more advanced correlational and regression analyses. You will need these skills for the lab assignment and papers. These skills are also essential for academic and employment pursuits in research. Today, you will go through this tutorial with your lab instructor. You can work collaboratively on this tutorial but must work independently on the graded lab assignment.

Instructions

Warning
SPSS periodically changes the visual display and organization of menus. The instructions presented in this tutorial may need to be augmented marginally depending on the version of SPSS you are using. If you get stuck, use Google, or ask the lab instructor for help.

Accessing SPSS
Once you log on, go to the Start menu in the lower left corner of the screen and find SPSS. If you have difficulty finding it, ask the lab assistant for help.

Data File
For this tutorial, you will use Data Set F (see Blackboard → Content → Data Files). The data set includes 509 participants who were Tulane students and their friends and family. Download the data file (DataF.sav) and the "data dictionary" that provides more detail on the variables that were included in the survey (DataF_Dictionary.xls). The files should open in SPSS and Excel, respectively. Double-click on them to open them, or open the programs and use the file menus to locate and open these files.

Review each of these files in detail. The Data Dictionary file in Excel provides added information beyond what is included in the SPSS Data file. Specifically, the columns in Excel indicate (1) the number and name of each construct measured, (2) whether the variable is categorical or continuous, (3) the specific question asked, and (4) the potential response options. Much of this information can also be found in Variable View of the SPSS file.

Review of Basic SPSS Skills – Practice Questions

Using the SPSS file, find the following information. If you get stuck, review the previous SPSS tutorial and/or seek help from the lab instructor. If your neighbor is stuck on this tutorial, feel free to help them.

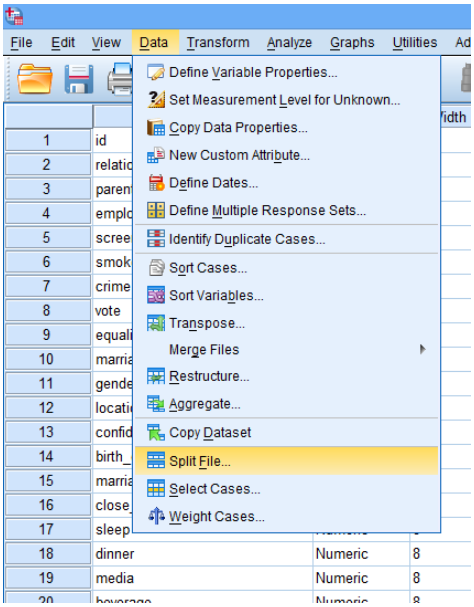
1. What was the median Age (variable #127) for our sample, and what does that value mean?
2. What percentage of our participants report that their favorite beverage is water in the Beverage variable (#20).
3. Examine participant 64 on variables 1 – 11. Is this person a single father?
4. What is the correlation between Fast Food Consumption (#32) and Worry about Death (#34)?
5. What is the correlation between Anger (#44) and Phone Throwing (#60)?
6. Of the "Big 5" personality traits (#108 - #112), which one correlates most strongly with Vocabulary (#129)?
7. Of the "Big 5" personality traits (#104 - #108), which one has the weakest correlation with Laughing Frequency (#28)?
8. There are two Happiness variables (#30 and #63). Correlate them with Anxiety (#24), Depression (#29), Excitement (#67), and Life Satisfaction (#100). Which Happiness variable has better construct validity?
9. There are two Satisfaction with Appearance variables (#38 and #54). Why aren't they perfectly correlated?
10. How would you describe the magnitude of the correlation (e.g. small/medium/large) between Delay of Gratification (#105) and Health (#106)?
11. How would you describe the magnitude of the correlation (e.g. small/medium/large) between Overall Political Orientation (#118) and Opposition to Reproductive Rights (#76)? Based on how the variables are coded, what does this mean?
12. Is the correlation between Overeating (#42) and Support for Marijuana Legalization (#61) statistically significant?
13. Is the correlation between Need for Social Media (#89) and Insomnia (#40) statistically significant?

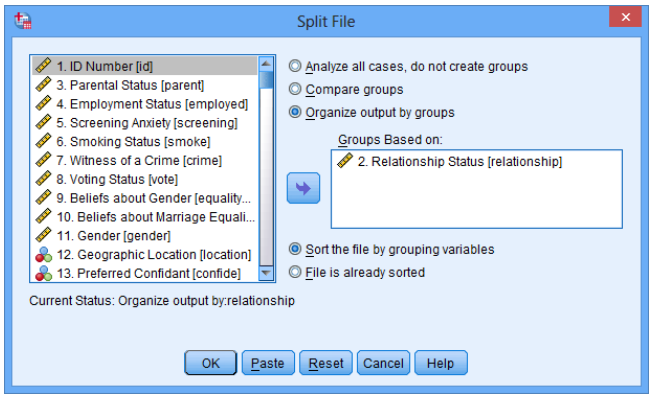
Split Analyses

SPSS allows people to run correlational analyses that are split by group. For example, it is possible to compare whether the correlation between two variables differs across groups (e.g. Is the correlation between Extraversion and Happiness different for males and females?).

To split the analyses by group, go to the Data menu, and choose Split file. In the window that pops up, choose Organize Output by Groups. Then, select a categorical variable (e.g. gender, parental status, relationship status, etc.) and move it to the "Groups Based on" area. For this example, move Relationship Status (#2) to the "Groups Based on" area. Then click OK.

From now on, any analyses we conduct will present separate results for depending on whether or not someone is in a relationship, rather than the entire sample.





As an example, now run a correlation between Neuroticism (#109) and Depression (#29). The Output (see below) should be separated by relationship status. Notice that the value of the correlations differs by group. In fact, the correlation is modestly higher for single people (r = .59) than for those in relationships (r = .38). This difference suggests neuroticism is closely aligned with depression in single people, but neuroticism it is less related to depression among people in relationships.

2. Relationship Status = No

Correlations[a]

		109. Neuroticism	29. Depression
109. Neuroticism	Pearson Correlation	1	.588**
	Sig. (2-tailed)		.000
	N	277	277
29. Depression	Pearson Correlation	.588**	1
	Sig. (2-tailed)	.000	
	N	277	277

**. Correlation is significant at the 0.01 level (2-tailed).

a. 2. Relationship Status = No

2. Relationship Status = Yes

Correlations[a]

		109. Neuroticism	29. Depression
109. Neuroticism	Pearson Correlation	1	.376**
	Sig. (2-tailed)		.000
	N	232	232
29. Depression	Pearson Correlation	.376**	1
	Sig. (2-tailed)	.000	
	N	232	232

**. Correlation is significant at the 0.01 level (2-tailed).

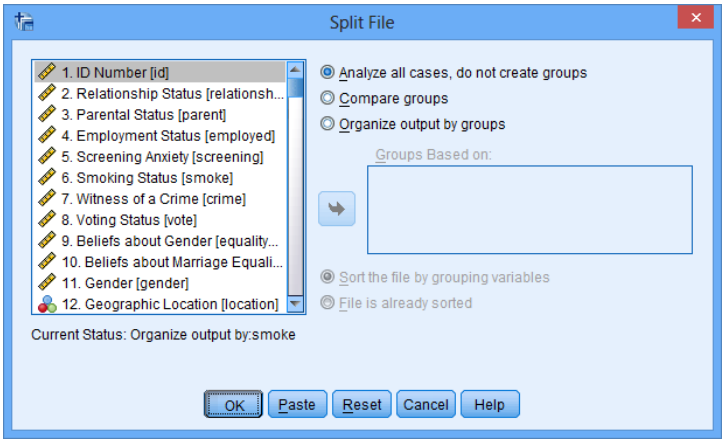
a. 2. Relationship Status = Yes

Split Analyses – Practice Questions

1. Set the file so it is split by Gender (#11). Find the correlation between self-reported Intelligence (#103) and Vocabulary (#129). Based on this information, which group views their intelligence as more closely tied to vocabulary skills?

2. Now, go set the file so it is split by the Smoking variable (#6) instead of gender. Find the correlation between being Anger (#44) and Cell Phone Throwing (#60). How would you explain the results in plain English that a non-statistics student could understand?

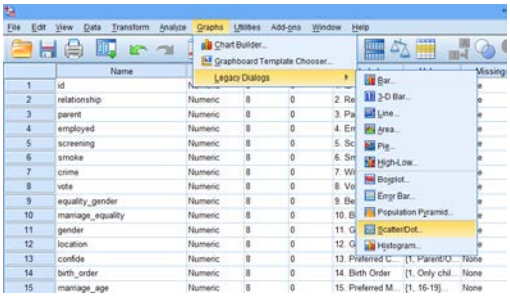
Before going onto the next section, reset the split file command so that results will not be split by category. In the Split File pop-up window, click Reset. Then, click OK.

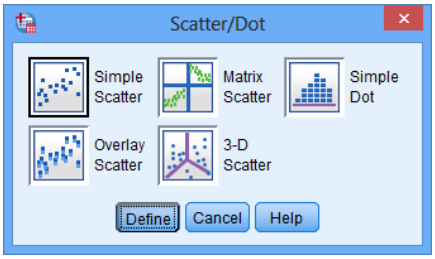


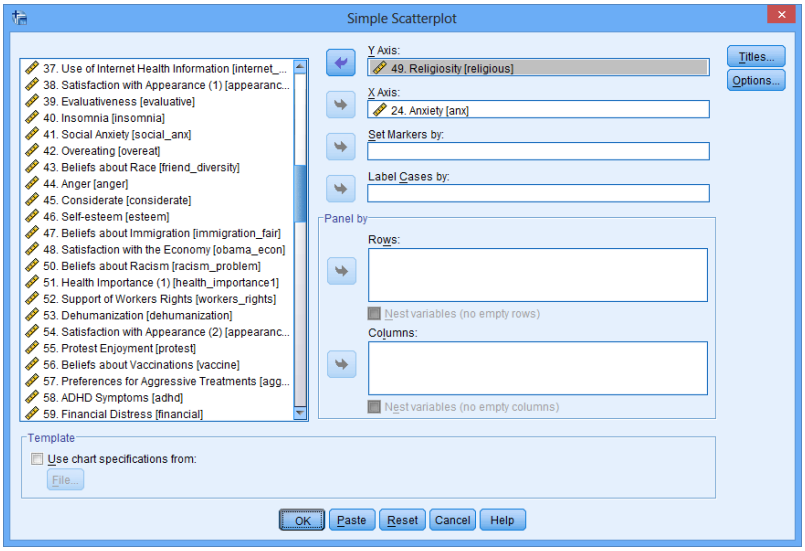
Scatterplots

Although SPSS has many statistical features, it is also useful for generating various graphs. To make a scatterplot, go to the Graphs menu, point to Legacy Dialogs, and choose Scatter/Dot (note different versions of SPSS organize menus differently, so simply find the Scatter/Dot command).

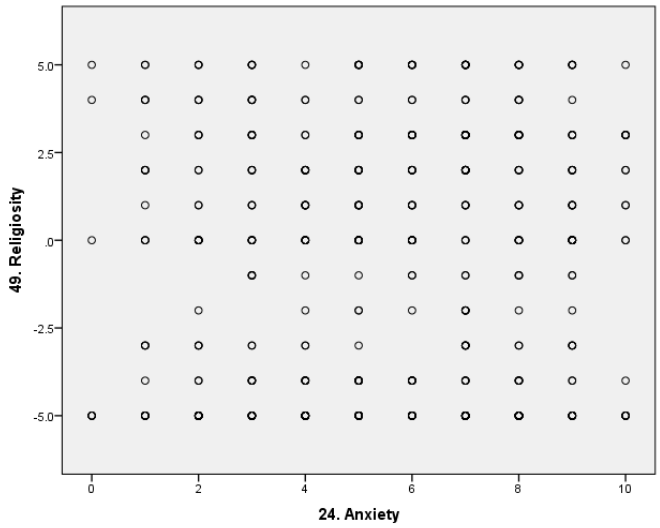
A pop-up window will appear. Select Simple Scatter and click the Define button. A new window pops up. To make a scatterplot, move one variable to the X Axis area and one to the Y axis area, then click the OK button. For example, move Anxiety (#24) to the X Axis area and Religiosity (#49) to the Y Axis area; click OK.







Your Output should look something like this:



There does not appear to be much of a correlation between anxiety and religiosity.

Scatterplots – Practice Questions

1. Make a scatterplot with Anxiety (#24) on the X Axis and Depression (#29) on the Y Axis. By estimating (see p. 267 of the textbook), what is the approximate correlation between the two variables? Run the correlational analyses to see how close your guess was.

2. At Taco Bell, one of your friends remarks that if you keep eating so many chalupas, you're going to be digging yourself an early grave. Make a scatterplot comparing Fast Food Consumption (#32) with Health (#106). Are there many people with very high fast food consumption and very good health? What's the correlation?

Multiple Regression

When conducting correlational analyses, you may be disappointed to see that correlation values are often fairly small. The main reason for this is that behavior is multidetermined. Usually several different factors combine to make people who they are and behave in certain ways.

Multiple regression allows us to examine how well several factors combine to predict a single variable. Instead of the symbol *r*, we use *R* to represent a correlation when using multiple regression. *R* values are interpreted the same way as *r* values for the most part, but *R* simply shows how well multiple variables combine to predict some outcome. *R* ranges from 0 to 1.

Multiple regression has three steps.

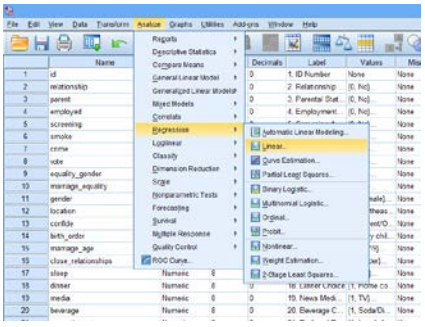
1. Come up with a theory. For example, we might think that parental warmth (#82), a sense of responsibility (#90), and investment in work or school (#126) are important for happiness (#30).
2. Test that theory with correlations (just like you learned to conduct previously). For example, see whether these three hypothesized variables actually correlate with happiness (then, compare to the correlation table below). Our theory was partially correct. Parental warmth and having a sense of responsibility were related to happiness, but investment in work/school was not.
3. After figuring out which variables correlate with the desired outcome, see how well they *combine* to predict the outcome using multiple regression. For example, we can examine the combined effect of parental warmth and responsibility on happiness, using one big correlation. We ignore the work/school variable because it was unrelated to happiness. For an explanation of how to run multiple regression, see below.

Correlations

		30. Happiness (1)	82. Parental Warmth	90. Responsibility	126. Work/School Participation
30. Happiness (1)	Pearson Correlation	1	.277**	.383**	.057
	Sig. (2-tailed)		.000	.000	.196
	N	509	509	509	509
82. Parental Warmth	Pearson Correlation	.277**	1	.099*	.119**
	Sig. (2-tailed)	.000		.025	.007
	N	509	509	509	509
90. Responsibility	Pearson Correlation	.383**	.099*	1	.047
	Sig. (2-tailed)	.000	.025		.286
	N	509	509	509	509
126. Work/School Participation	Pearson Correlation	.057	.119**	.047	1
	Sig. (2-tailed)	.196	.007	.286	
	N	509	509	509	509

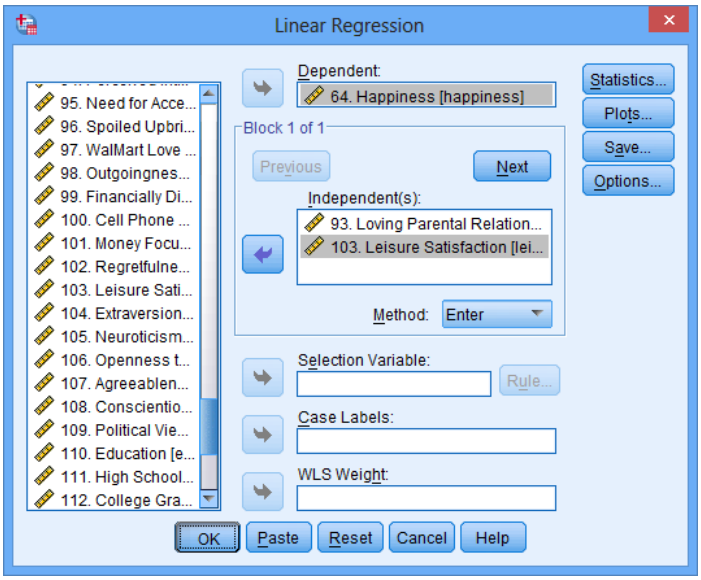
**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

The multiple regression analyses are not very difficult. Simply go to the Analyze menu, point to Regression, and choose Linear.



A window pops up. Where it says Independent(s), we enter our Independent variables, the predictors or causes (usually there are several). Where it says Dependent, we enter the single dependent variable, which is also known as the outcome variable or effect. To practice using our example, enter Happiness (#30) for the Dependent variable. Enter Parental Warmth (#82) and

Responsibility (#90) in the Independents section. Leave out the work/school variable because we already know it's unrelated from running the correlations. Then, press OK.



The Output should look something like this:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	90. Responsibility , 82. Parental Warmth	.	Enter

a. All requested variables entered.
b. Dependent Variable: 30. Happiness (1)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.452[a]	.205	.201	1.611

a. Predictors: (Constant), 90. Responsibility, 82. Parental Warmth

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	337.895	2	168.948	65.082	.000[a]
	Residual	1313.543	506	2.596		
	Total	1651.438	508			

a. Predictors: (Constant), 90. Responsibility, 82. Parental Warmth
b. Dependent Variable: 30. Happiness (1)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.231	.330		9.781	.000
	82. Parental Warmth	.184	.030	.242	6.068	.000
	90. Responsibility	.303	.034	.359	9.012	.000

a. Dependent Variable: 30. Happiness (1)

In this entire section of Output, we can actually ignore most of the information. Everything we need is in the 2nd and 3rd boxes.

- The box I have shaded blue (where it says "R") is the *R* value. It is similar to the *r*-values you're already familiar with; however, it indicates the combined effect of both predictors. In this case, parental warmth and responsibility *combine* to correlate $R = .45$ with happiness, which is bigger than either of their singular effects. Together they modestly predict happiness.
    - Side note: You should be aware that although *r* values can be negative or positive (-1 to +1), *R* values only range from 0 to 1 (no negatives). The reason for this is that multiple regression lets you examine the combined effects of several variables; some of those variables could have positive correlations, some negative correlations, so the overall effect size (*R*) just puts everything on a positive scale.
- The R Square value in the red box stands for R^2 and is similar to r^2. It tells how much of the variability in the outcome variable we're able to account for. In this example, parental warmth and responsibility account for 21% of the variability in happiness.
- Finally, in the green box is a *p*-value, similar to the *p*-values you've already learned about. If $p<.05$, the finding is trustworthy. It is obviously lower than .05, so the finding is trustworthy. (Note that the ".000" value is due to rounding. It's very close to zero, but not exactly zero. By convention, we usually just write "$p < .001$").
- [Advanced Students: The final box of the Output contains some information often used in more advanced analyses. For example, the "Standardized Coefficient Betas" basically tell you each independent variable's correlation with the dependent variable, while controlling for the influence of any other independent variables in the model. The corresponding "Sig." values tell you whether the coefficient is statistically significant. This information is not required for lab assignments or the papers.]

<u>Multiple Regression – Practice Questions</u>

1) Your friend has a theory that religious people, conformists, and extraverts were more likely to engage in the charity "ice bucket" challenge. Test this hypothesis by examining

whether Religiosity (#49), Conformity (#98), and Extraversion (#108) are correlated with Charitable Giving (#31). If any of these variables significantly correlate with Charitable Giving ($p < .05$), incorporate them into a multiple regression. What is the *R* value using the significant predictors to predict charitable giving? What is the R^2 value? What does the R^2 value mean? What do you tell your friend?

2) Your friend argues that someone you know has very little "Emotional Intelligence" (#104) due to several factors, including having an exaggerated sense of Superiority (#66), low Self-Esteem (#46), being Depressed (#29), and having little Openness to Criticism (#80). Examine these correlations. If any of these variable significantly correlate with emotional intelligence ($p<.05$), incorporate them into a multiple regression. What is the *R* value using the significant predictors to predict emotional intelligence? If one of the correlations (*r*) was negative, why was *R* positive? What is the R^2 value, and what does it mean?

APA Style

On the next page, you will find examples of how to write-up results in APA-style. Refer back to these examples when working on Paper 1.

Lab Assignment

You are now ready to begin working independently on your next homework assignment, "Lab Assignment 2" (see "Due" column on the Course Calendar).

Dismissal

Students can be dismissed early if they complete LA2 in its entirety.

APA Style Guide

Note: You have my permission to copy any or all of this writing for this or future assignments.

Style and Rounding

Rules governing rounding vary considerably from discipline to discipline. These guidelines reflect the current norms in psychology.

p-values. Historically, published articles either reported statistically significant findings as "$p < .05$" and non-significant findings as "*ns*" – this was a very imprecise way of reporting the results. All major psychology journals now advocate reporting actual *p*-values when they are provided in text (in tables and figures asterisks are still common). In general, *p*-values should be reported rounded to two decimals (e.g., $p = .08$ or $p = .02$). However, if the *p*-value is less than .01, report three decimals (e.g., $p = .009$ or $p = .002$). If the p-value is less than .001, simply report as "$p < .001$" (note, SPSS strangely reports these as .000, but it is impossible to have a probability of zero, so do not report it that way).

Percentages. Usually percentages are rounded to one decimal place (e.g. 88.6% or 1.1%).

Other statistics. In general, all other statistics are rounded to two decimal places (e.g., $M = 1.46$ or $r = -.33$)

Leading zero. If a statistic is a decimal, people usually include a leading zero only if the statistic can commonly exceed 1.0. For example, most descriptive statistics, as well as *t*-scores, *Z*-scores, and *F*-scores commonly exceed 1.0, so even if an observed value is a decimal, a leading zero is included (e.g., $Z = 0.23$ or $t = 0.96$). In contrast, correlations cannot exceed 1.0, so no leading zero is included (e.g., $r = .23$ or $r = .96$).

Italics. Statistical symbols should be italicized (e.g., *M*, *SD*, *r*, *t*, *d*, *F*, *p*, etc.) but not the numbers following them.

Descriptive Statistics

On the 1-9 depression severity scale, the sample reported a mean score of 4.66 (*SD* = 1.59), with 12.3% reporting a "1" (not at all depressed) and 2.1% reporting a "9" (completely depressed).

The sample was predominantly white (94.5%) and college-educated (86.6%), more often female (61.3%), and distributed relatively evenly across the U.S. (North: 24%, South: 30%, Midwest: 20%, West: 26%).

Participants varied considerably in age (*M* = 35.5 years, *SD* = 10.2, ranging from 18 to 77).

Participants identified as Democrats (30.2%), Republicans (19.8%), or Independents (50.0%).

Correlation (Significant, $p < .05$)

Note: Include the correlation, *p*-value, a description of the direction (more, less, positively, negatively, directly, inversely, etc.), and a description of the effect size (e.g., near-zero/marginal, small/slight, medium/moderate/modest, strong/large/sizeable). If the finding might be confusing to a non-statistician, include a second sentence explaining the finding in simpler terms.

Participants who were more neurotic reported exercising moderately less often, which was statistically significant, $r = -.35$, $p = .02$.

Quarterbacks who were taller had marginally better completion rates, $r = .09$, $p = .04$. Thus, tall quarterback throw completed (caught) passes more often than short quarterbacks.

Anxiety and depression were strongly positively correlated ($r = .71$, $p = .007$). Therefore, it could be difficult to distinguish between whether someone's primary diagnosis should be an anxiety disorder or a mood disorder.

Correlation (Non-Significant, $p > .05$)

Age was not significantly associated with income ($r = .13$, $p = .23$), political views ($r = .01$, $p = .99$), or vocabulary ($r = .06$, $p = .62$).

The present study failed to find an association between wealth and happiness, $r = .08$, $p = .64$.

Several Correlations, followed by Multiple Regression

Note: First, describe the correlational results, where you compare each of the predictors to the dependent variable. Then, provide a rationale for the regression analyses. In reporting the results, people usually include *R*, R^2, or both, followed by the *p*-value. Then, describe the results in plain English, if needed.

Family stress ($r = .48$, $p = .008$), work stress ($r = .56$, $p < .001$), and school stress ($r = .21$, $p = .04$) all significantly predicted overall life stress. However, social support did not predict level of life stress, $r = .03$, $p = .64$. Thus, although social support was not related to life stress, one's level of school stress was slightly related, family stress was modestly related, and work stress was strongly related to level of life stress. To examine the overall contribution of the three significant predictors (school stress, family stress, and life stress) in accounting for life stress, multiple regression was used. The results of the multiple regression analysis indicated that these three predictors accounted for a large proportion of the variance in life stress, $R^2 = .40$, $p < .001$. Thus, school stress, family stress, and work stress together account for 40% of the differences in overall life stress.

Several factors were hypothesized to predict college GPA. Being encouraged to read ($r = .19$, $p = .002$) and conscientiousness ($r = .26$, $p < .001$) had small positive relationships with college GPA. ADHD symptoms had a small negative relationship ($r = -.17$, $p = .007$). Hours of work per week was not correlated with GPA ($r = .08$, $p = .22$). Thus, being encouraged to read and being conscientious are related to better grades, but having ADHD symptoms is related to lower grades. The number of hours people spend on employment was not related to grades. Multiple regression was used to examine the combined effect of being encourages to read, conscientiousness, and ADHD symptoms on college GPA. These three predictors combined to modestly predict GPA, $R = .33$, $R^2 = .11$, $p < .001$. Therefore, being encouraged to read, conscientiousness, and ADHD symptoms explain 11% of the differences in college grades.