Experimental Psychology
Lab Session – Week 7

SPSS Correlation/Regression

Due at the Start of Lab: Lab 3

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 Lab Assignment 4 and Paper 1. 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 one of the lab assistants for help.

Pre-Existing Data File
For this tutorial, you will use data collected in the spring of 2009 by students like you who were enrolled in Univariate and Research Methods courses. The participants in that study were the students themselves as well as their friends and family. In total, 975 people completed their survey.

Accessing a Pre-Existing Data File
Log on to BlackBoard. Download the data file (Spring 2009 Data File: S2009_data.sav) and a "data dictionary" that provides details on the variables that were included in the survey (Spring 2009 Dictionary File: S2009_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.

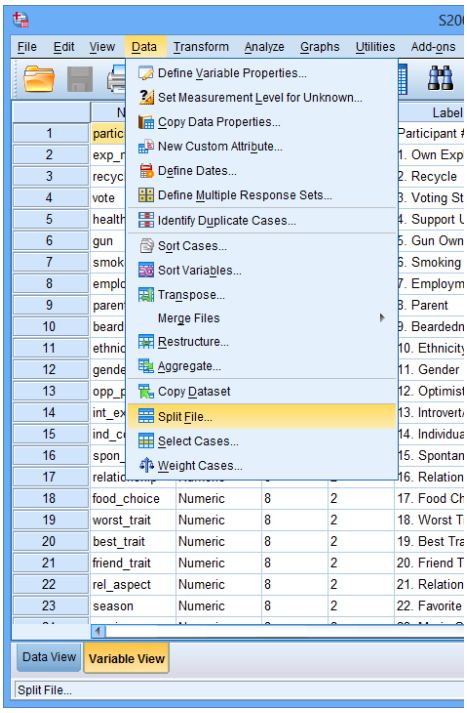
1. What was the mean Age (variable #122) for our sample?
2. What percentage of our participants report that their favorite season is winter or spring? See the Favorite Season variable (#22).
3. Examine participant 964 on variables 1 – 12. What is atypical about this participant?
4. What is the correlation between Watching Movies (#50) and Sports Participation (#56)?
5. What is the correlation between Road Rage (#57) and the 2nd Anger variable (#63)?
6. Of the "Big 5" personality traits (#104 - #108), which one correlates most strongly with Vocabulary (#123)?
7. Of the "Big 5" personality traits (#104 - #108), which one has the weakest correlation with Laughing (#52)?
8. There are two Anger variables (#34 and #63). Correlate them with Cell Phone Throwing (#100), Loving Parental Relationships (#93), and Regretfulness (#102). Which Anger variable has better construct validity?
9. There are two Confidence variables (#26 and #30). Why aren't they perfectly correlated?
10. How would you describe the magnitude of the correlation (e.g. small/medium/large) between ACT Scores (#113) and Vocabulary (#123)?
11. How would you describe the magnitude of the correlation (e.g. small/medium/large) between Shame (#62) and Body Satisfaction (#84)?
12. Is the correlation between Text Messaging (#43) and Academic Focus (#90) statistically significant?
13. Is the correlation between Text Messaging (#43) and Being Hopeful for Obama (#92) statistically significant?

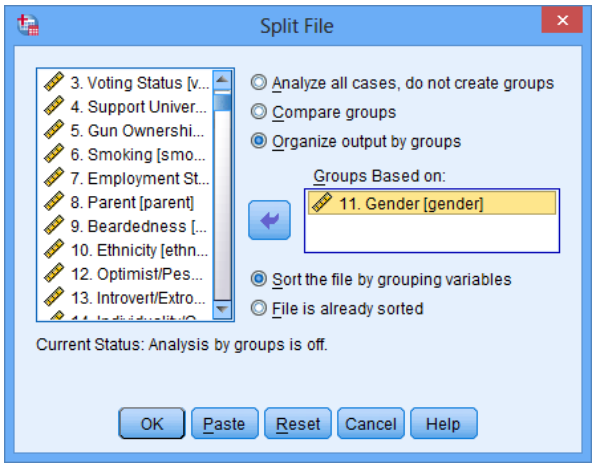
Split Analyses

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, ethnicity, relationship status, etc.) and move it to the "Groups Based on" area. For this example, move Gender (#11) to the "Groups Based on" area. Then click OK.

From now on, any analyses we conduct will present separate results for males and females, rather than the entire sample.





As an example, now run a correlation between Extraversion (#104) and Happiness (#64). The Output (see below) should be separated by gender. Notice that the value of the correlations differs by group. In fact, the correlation is slightly higher for males (r = .29) than for females (r = .21). This difference is relatively small, but extraversion probably plays a greater role in happiness for males than females.

11. Gender = female

Correlations[a]

		104. Extraversion	64. Happiness
104. Extraversion	Pearson Correlation	1	.209**
	Sig. (2-tailed)		.000
	N	676	676
64. Happiness	Pearson Correlation	.209**	1
	Sig. (2-tailed)	.000	
	N	676	676

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

a. 11. Gender = female

11. Gender = male

Correlations[a]

		104. Extraversion	64. Happiness
104. Extraversion	Pearson Correlation	1	.285**
	Sig. (2-tailed)		.000
	N	299	299
64. Happiness	Pearson Correlation	.285**	1
	Sig. (2-tailed)	.000	
	N	299	299

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

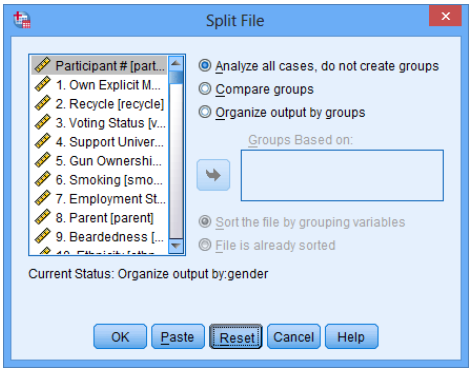
a. 11. Gender = male

Split Analyses – Practice Questions

1. Set the file so it is split by gender. Find the correlation between ACT Scores (#113) and College Grades (#112). Based on this information, do the ACTs appear strongly biased against females?

2. Now, go set the file so it is split by the Smoking variable (#6) instead of gender. Find the correlation between being Anxiety (#59) and Cell Phone Throwing (#100). How would you explain the results in plain English that a non-statistics student could understand?

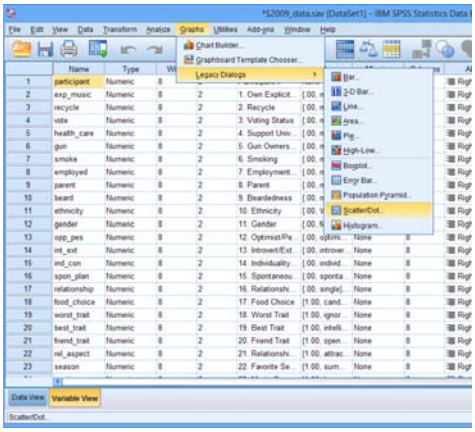
Before going onto the next section, reset the split file command so that results will not be split by category. In the Split File pop-up window, click Reset. Then, click OK.

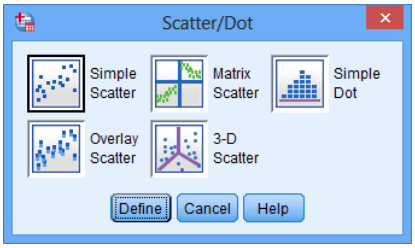


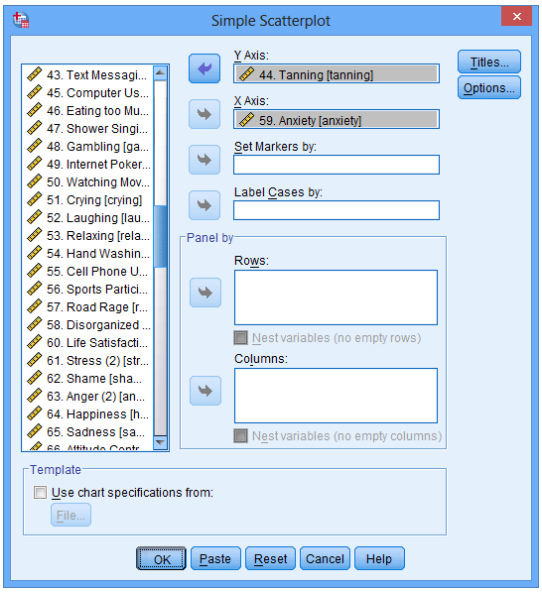
Scatterplots

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).

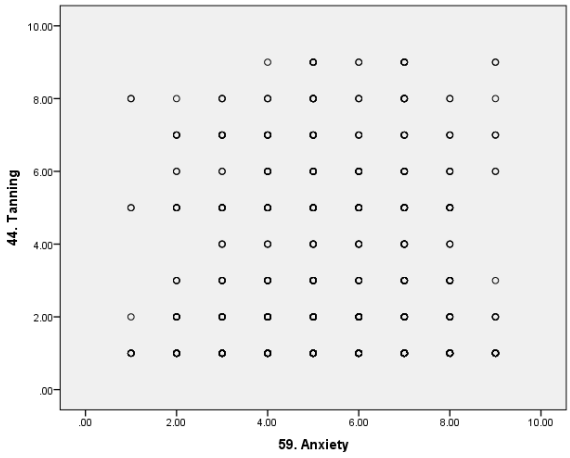
A pop-up window will appear. Select Simple Scatter and click the Define button. A new window pops up. To make a scatterplot, move one variable to the X Axis area and one to the Y axis area, then click the OK button. For example, move Anxiety (#59) to the X Axis area and Tanning (#44) to the Y Axis area; click OK.







Your Output should look something like this:



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

Scatterplots – Practice Questions

1. Make a scatterplot with High School Grades (#111) on the X Axis and ACT Scores (#113) on the Y Axis. By estimating, what is the approximate correlation between the two variables? Run the correlational analyses to see how close your guess was.

2. A friend of yours says that if you keep working so hard, you're going to get stressed out. Make a scatterplot comparing Hours of Work (#121) to the 1st Stress variable (#32). Are the two variables closely related?

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 having loving parental relationships (#93), time for leisure (#103) and a good education (#110) all combine to make people happy (#64).
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. Loving parental relationships and satisfaction with leisure time both correlated with happiness. However, one's level of education did not correlate significantly.

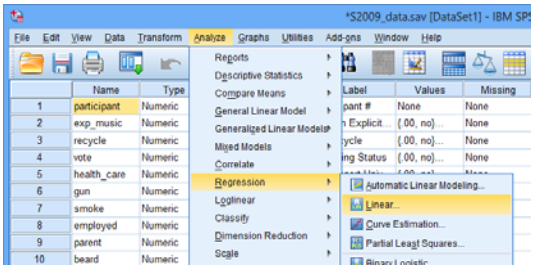
3. After figuring out which variables correlate with the desired outcome, see how well they *combine* to predict the outcome using multiple regression. For example, we can examine the combined effect of loving parental relationships and leisure satisfaction on happiness, using one big correlation. We ignore the education variable because it was unrelated to happiness. For an explanation of how to run multiple regression, see below.

Correlations

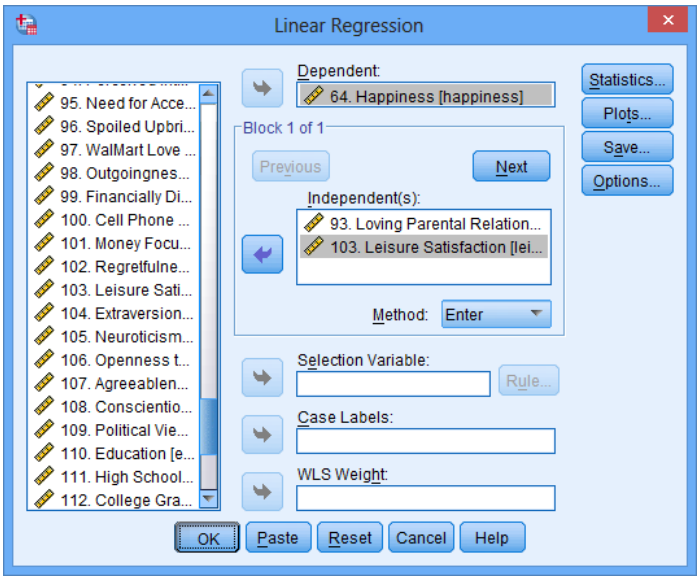
		64. Happiness	93. Loving Parental Relationships	103. Leisure Satisfaction	110. Education
64. Happiness	Pearson Correlation	1	.214**	.299**	.025
	Sig. (2-tailed)		.000	.000	.437
	N	975	975	975	975
93. Loving Parental Relationships	Pearson Correlation	.214**	1	.157**	.014
	Sig. (2-tailed)	.000		.000	.671
	N	975	975	975	975
103. Leisure Satisfaction	Pearson Correlation	.299**	.157**	1	.018
	Sig. (2-tailed)	.000	.000		.575
	N	975	975	975	975
110. Education	Pearson Correlation	.025	.014	.018	1
	Sig. (2-tailed)	.437	.671	.575	
	N	975	975	975	975

**. Correlation is significant at the 0.01 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 (#64) for the Dependent variable. Enter Loving Parental Relationships (#93) and Leisure Satisfaction (#103) in the Independents section. Leave out the Education 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	103. Leisure Satisfaction, 93. Loving Parental Relationshi...	.	Enter

a. All requested variables entered.

b. Dependent Variable: 64. Happiness

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.344[a]	.118	.116	1.23064

a. Predictors: (Constant), 103. Leisure Satisfaction, 93. Loving Parental Relationships

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	197.006	2	98.503	65.041	.000[a]
	Residual	1472.064	972	1.514		
	Total	1669.071	974			

a. Predictors: (Constant), 103. Leisure Satisfaction, 93. Loving Parental Relationships

b. Dependent Variable: 64. Happiness

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.076	.173		29.334	.000
	93. Loving Parental Relationships	.113	.020	.171	5.620	.000
	103. Leisure Satisfaction	.168	.019	.272	8.917	.000

a. Dependent Variable: 64. Happiness

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, loving parental relationships and leisure satisfaction *combine* to correlate $R = .34$ with happiness. That is, 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, loving parental relationships and leisure satisfaction account for 12% 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.

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

1) Your friend says that Religious Fundamentalism (#77), Leadership (#89), and Intelligence (#94) all lead someone to do more Volunteering (#38). Examine these correlations. If any of these variable significantly correlate with volunteering ($p<.05$), incorporate them into a multiple regression. What is the *R* value using the significant predictors to predict volunteering? What is the R^2 value? What does the R^2 value mean?

2) Your friend argues that someone you know has very little Family Closeness (#27) due to several factors, including problems Expressing Love (#37), being Financially Distressed (#99), and Neuroticism (#105). Examine these correlations. If any of these variable significantly correlate with family closeness ($p<.05$), incorporate them into a multiple regression. What is the *R* value using the significant predictors to predict family

closeness? If two of the correlations (*r*) were 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 4" (see "Due" column on the Course Calendar).

APA Style Guide

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

Correlation Only (Significant, $p < .05$):

Example 1: The correlation between IQ and hours of television watched was significant, $r = -.35$, $p = .02$. That is, people who were smarter watched moderately less television.

Example 2: The correlation between IQ and hours of television watched was significant, $r = -.35$, $p < .05$. That is, people who were smarter watched moderately less television.

Include the correlation. When significant, say "$p < .05$" or provide the exact *p*-value. Then describe the results in plain English.

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

Example 1: IQ and number of hours of television watched were not significantly related, $r = .08$, $p = .67$. Thus, one's level of intelligence was not related to time spent watching television.

Example 2: IQ and number of hours of television watched were not sizably related, $r = .08$, *ns*. Thus, one's level of intelligence was not related to time spent watching TV.

Include the correlation. When non-significant, say "ns" for non-significant, or include the exact p-value. Then describe the results in plain English.

Several Correlations, followed by Multiple Regression:

Example 1: Family stress ($r = .48$, $p < .05$), work stress ($r = .56$, $p < .05$), and school stress ($r = .21$, $p < .05$) all significantly predicted overall life stress. However, social support did not predict level of life stress, $r = .03$, *ns*. 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 indicate that these three predictors accounted for a large proportion of the variance in life stress, $R^2 = .40$, $p < .05$. Thus, school stress, family stress, and work stress together account for 40% of the differences in overall life stress.

Example 2: 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.

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.