How To Do Research

15 Labs for the Social & Behavioral Sciences

JANE F. GAULTNEY ▪ HANNAH D. PEACH

LAB 9

Correlational Design

Objective

This lab introduces correlational designs and explores analysis and interpretation of these designs. Additional material is included on statistically controlling covariates and the concept of partial correlations.

Target Article

Cavallo, A., Ris, M. D., & Succop, P. (2003). The night float paradigm to decrease sleep deprivation: Good solution or a new problem? *Ergonomics, 46*(7), 653–663. doi:10.1080/0014013031000085671

This study examined the impact of a "night float rotation" on the sleep, mood, and performance of medical residents. Night float rotation is a work schedule in which the individual works at night but has no daytime duties (as compared to being on duty for, say, 24 hours at a time). Even though the residents had enough time during the day to sleep, they got fewer hours of sleep than they did while on a typical daytime rotation. Measures of fatigue during the night float rotation were correlated with measures of attention. Greater sleepiness was associated with more errors of attention.

EXPERIMENTAL VERSUS NON-EXPERIMENTAL

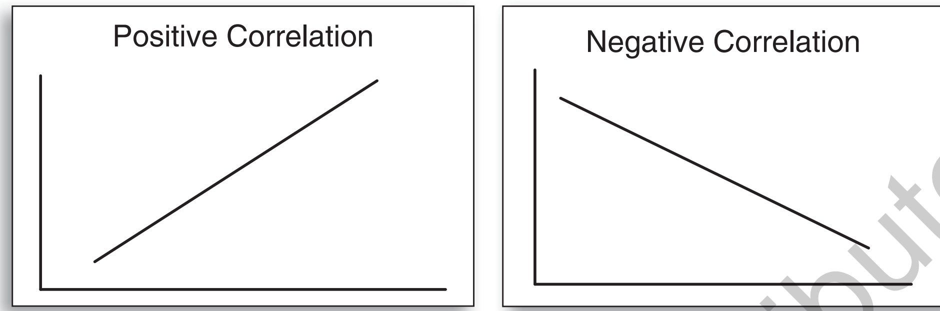

©iStock.com/geotrac

The next few labs can be divided into two groups: non-experimental methods and experimental methods. What is the difference between the two? Experimental studies use random assignment to groups and have an independent variable (IV) that is manipulated. Some research questions simply can't be studied using an experimental design, so a non-experimental one is chosen. Correlation designs are usually non-experimental. Regression, two-group, three-group, and factorial designs may be experimental if random assignment to groups is used, but the designs can also be used for quasi-experimental research in which random assignment to groups is not possible. One type of quasi-experimental design is an ex post facto or non-equivalent groups design in which some participant characteristic is used as an IV. In this situation, the design looks like a duck (my analogy here for an experimental design) and quacks like a duck, but isn't actually a duck. For example, if you want to look at gender differences in access to health care, you don't get to randomly assign participants to the male or female group; they come to you already in one group or the other. In that case, the data collection and analysis are the same as an experimental design. The difference comes in your interpretation of the findings. When you can't use random assignment to groups, your findings are shakier. You have not demonstrated cause and effect and need to proceed cautiously in your conclusions.

In this lab, we'll look at non-experimental correlational designs. This is appropriate to use when your hypothesis predicts an association between variables (the rank orders of the scores vary in some systematic way relative to each other). Do high scores on one variable tend to co-occur with high or low scores on another variable? There is no IV or DV—just two variables to compare. A correlation is reported as r and can vary from –1.00 to 1.00. The closer to 0, the less relationship there is between the two variables. The closer to –1 or 1, the stronger the correlation.

A positive correlation means that high scores on one variable go with high scores on the other variable . . .

. . . and a negative correlation means that high scores on one variable are associated with low scores on the other variable.



TRY IT YOURSELF

Do Sleepier People Have More Accidents at Work?

Impaired attention could cause more accidents and errors when working. For this lab, we're going to examine the relation (or association) between average hours of sleep and number of accidents on the job. We're going to compare the rank-ordering of these two variables. We won't be comparing group means. We want to know if the rank ordering varies in a systematic way. We are predicting that people who get the most sleep tend to have the fewest accidents. To do this, we will measure two variables—amount of sleep and number of accidents—for each participant and then correlate the variables.

©iStock.com/RTimages

You will set up a data file, run and interpret a correlation analysis, and write a short Results section for the analysis.

First of all, let's review the use of correlation designs and the correlation statistic. Read the tutorial at http://www.wadsworth.com/psychology_d/templates/student_resources/workshops/stat_workshp/correlation/correlation_01.html. Now go to http://www.stat.berkeley.edu/~stark/Java/Html/Correlation.htm to review what scatter plots for positive and negative, strong and weak correlations look like. Click the regression line button to see the linear association of two variables. At the top of the figure, play around with correlations close to 0 (e.g., .08) and ones that are larger (e.g., .80). Try positive and negative correlations.

Canned Data

Entering and Analyzing the Data

Enter and save the following data into SPSS. The variable labeled *Avg_slp* is the average hours slept at night and *Accidents* indicates the number of accidents over the previous 12 months. Higher numbers indicate more sleep and more accidents, respectively.

Avg_slp	*Accidents*
7.00	2.00
6.00	4.00
4.00	3.00
7.00	3.00
4.00	6.00
5.00	6.00
4.00	7.00
9.00	2.00
9.00	4.00
8.00	3.00
7.00	4.00
4.00	5.00
6.00	3.00
9.00	1.00
8.00	2.00
5.00	5.00
6.00	4.00
8.00	5.00
5.00	6.00
7.00	6.00

Eyeball the data. Run descriptive and frequency analyses for your variables then check to make sure everything looks the way you expected it. When you are sure your data are accurate, proceed.

Run the correlational analysis.

1. Click *Analyze*
2. *Correlate*
3. *Bivariate*
4. Highlight the names of the two variables and click the arrow to the right to move them into the *Variable* box.
5. Now click *OK*.

Your output should look like this:

Correlations

		AVG_SLP	ACCIDENT
AVG_SLP	Pearson Correlation	1	-.634**
	Sig. (2-tailed)	.	.003
	N	20	20
ACCIDENT	Pearson Correlation	-.634**	1
	Sig. (2-tailed)	.003	.
	N	20	20

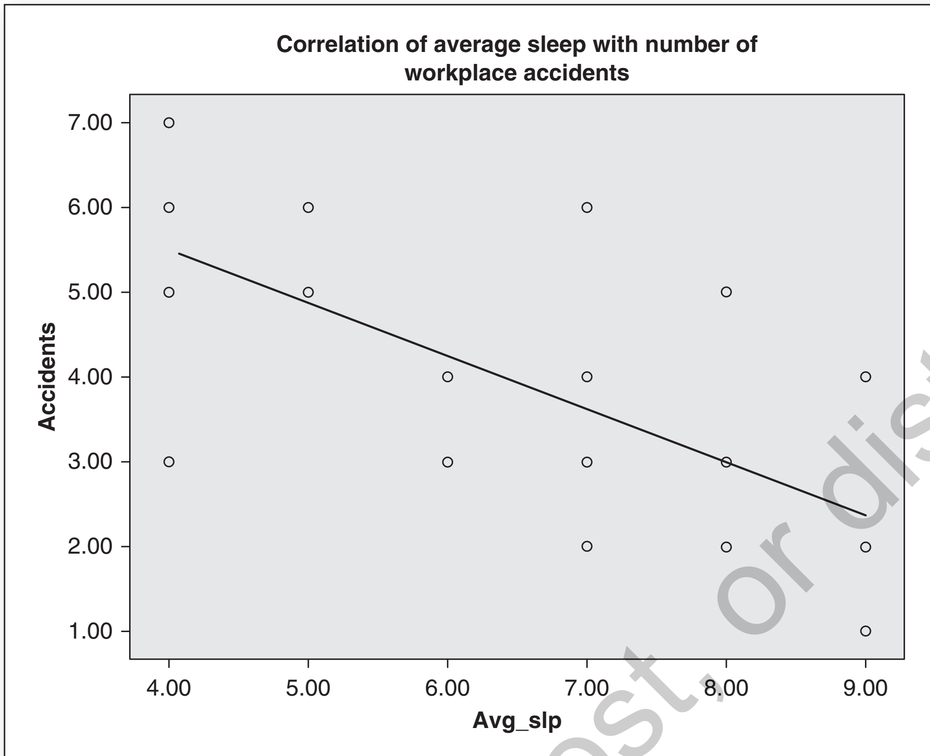
**. Correlation is significant at the 0.01 level (2-tailed).

Notice that we are using the Pearson correlation. This is used when both variables are interval or ratio scale measures. There are other measures that can be used when one or both variables use another scale.

The correlation between hours slept and number of accidents is –.634, the sample size was 20, and the significance level is .003. Is this a significant correlation? ___ The fact that it is negative means that people who get more sleep have ________ (more or fewer) accidents. ________ Does this support our prediction? ___ [yes, fewer, yes]

If you'd like to see a scatterplot of the data, go to Graphs, Legacy, Scatterplot*. Put average sleep on the *x*-axis and number of accidents on the *y*-axis.

* The wording may be slightly different depending on the version of SPSS being used.



I've added the diagonal line so you can visualize the correlation.

Reporting the Results

Now we're going to write a very short Results section about these data. Since this may be your first try at a Results section, I'll give you an example. I'm going to round the value of the correlation.

> **Results**
>
> The average hours slept by participants were correlated with the number of accidents they had reported during the previous 12 months. The two variables were significantly correlated, $r(20) = -.634$, $p = .003$, indicating that participants who averaged more hours of sleep at night tended to have fewer accidents.

Notice several things here. You told the reader what statistic was used, what the variables were, and that they were significantly correlated, then you backed that up by reporting the statistics. The *r* and *p* are both lower case and italicized, and the sample size is indicated after the *r*. Finally, you helped the reader understand what the significant statistic meant.

What can we conclude based on these data? We can claim only that less sleep is associated with a higher accident rate. We don't know if lack of sleep *caused* the accident, if being accident-prone *caused* less sleep, or if they are really not connected with each other, but rather there is some other factor causing both of them. We can't speculate on cause based on these data.

Now . . . close this lab manual, look at the SPSS output, and write a short Results section. When you are done, compare what you wrote to the example above. Did you include all the relevant information? Did you report the statistics using proper format? Did you explain the correlation to the reader? Make corrections if needed. I'm not worried about plagiarism here; it's OK if your corrected writing is similar to what is modeled here. DO TRY to write this on your own, however; the day is coming when you will be expected to know how to do this without a model.

IF YOU WANT TO COLLECT YOUR OWN DATA . . .

If you'd like to analyze data generated by your class, you could collect data on the association of sleep duration and academic performance. You can use the chart below to keep track of how much time participants slept the night before an exam (perhaps all those in your class who are willing to contribute data can serve as participants). The two variables correlated will be number of hours slept and grade on the exam. What would you predict about a correlation between these two variables? Let's say you hypothesize that people who report more sleep the night before an exam will have a higher grade on the exam. Bear in mind that staying up at night to cram for a test is not a productive study strategy for at least two reasons. We know that distributed practice (smaller study sessions over several study periods) produces better recall than does massed practice (one intensive study session), and sleep deprivation may decrease your ability to do things like encode and consolidate new material, recall, pay attention, solve problems, organize your thoughts, and so on. Alternatively, it is possible that people who tend to skimp on sleep use poor study practices in general!

Hours slept the night before the exam: ______________

Grade in exam: ______________

Now you need to include your data with that of your other classmates. Pass around a spreadsheet and have everyone (anonymously) fill in their numbers or data they collected from others. Enter these data into an SPSS file and analyze as described above.

	Hours slept	Grade
Person 1	__________	__________
Person 2	__________	__________
Person 3	__________	__________

. . . and so on.

Notice that we aren't trying to demonstrate that one variable caused the other. We're just comparing rank order to see if two things co-occur in a discernible pattern.

Closure

Correlations do

1. compare the nature of association between two variables, indicating whether or not there is a relation.
2. indicate whether the relationship is strong, weak, or non-existent, and whether the factors measured vary in the same direction (positive correlations) or opposite directions (negative correlations).

Correlations do not

1. compare group means.
2. have an IV or DV.
3. indicate cause and effect. Variable A *may* have caused variable B, but it is just as possible that variable B caused variable A, or that there is some other variable, C, out there that is causing both A and B. We can't tell on the basis of the correlation.

IF YOU WANT TO GO FURTHER . . .

Partial Correlations

As stated above, we know that a correlation between variables A and B does not necessarily equal causation, and it is possible that a variable C may in fact be causing both A and B. What can you do with your statistical analysis when you know some extraneous variable may be responsible for a correlation between A and B? You can control those extraneous variables statistically by measuring them and entering them into your analysis as covariates. For the moment, we'll talk about covariates in correlation since that is the analysis du jour. Before we go further, let's note that it is better to control extraneous variables using design rather than statistics, but sometimes statistical control is your best option.

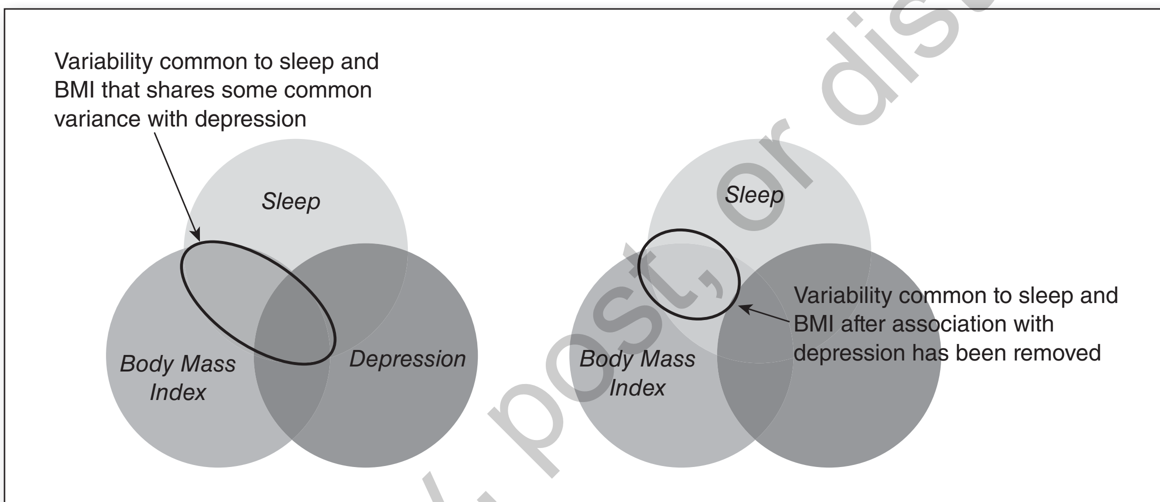
Pretend you found a significant correlation between ice cream consumption and murder rates—does this mean eating ice cream is related somehow to committing murder? Absolutely not. There could be a confounding variable, *hot temperature*, that is associated with both variables. Higher temperatures are associated with more ice cream eating (yum!) and more aggressive behaviors leading to homicides (not cool). If I correlated ice cream consumption and murder rates, I might get a significant correlation. However, this would be misleading, because ice cream and murder are not directly related. Instead, both are a byproduct of hot weather. If a partial correlation was analyzed controlling for temperature, the significant association between ice cream and murder rates would likely disappear.

©iStock.com/cheische

Let's say I am interested in examining the relationship between sleep quality and body mass index (BMI)—I first want to know if a relationship exists before I jumpstart a research project looking at poor sleep quality as a risk factor for obesity. Research has repeatedly shown associations between lower sleep quality and greater BMI (Cappuccio et. al., 2008); however, studies have also shown that depression is related to both higher BMI (e.g., Johnston, Johnson, McLeod, & Johnston, 2004) and worse sleep quality (e.g., Alvaro, Roberts, & Harris, 2013). How then can we know if sleep quality and BMI have a *unique* association, *above and beyond* shared relations with depression? In other words, we want to know if sleep and BMI are correlated even when we control for depression.

This is a common scenario in scientific studies, where researchers want to *control covariates.* A covariate is a third (or more) variable that can affect the relationship between two other variables of interest—let's call this our variable C. If we can separate out the **covariance** (how much the variables vary in a corresponding manner) of variable C from variables A and B, we can leave behind only the covariance of variables A and B. This can tell us if there is still a relationship left behind between variables A and B once we've controlled or accounted for variable C. See the figure below; if we can completely remove the association with depression, we are left with the *partial correlation* between sleep and BMI.

Figure 9.1 Partial Correlations



We can test our hypothesis that sleep and BMI are associated even after completely controlling for depression by calculating a partial correlation using SPSS. Let's pretend we've measured self-reported quality of sleep per night, BMI using the standard formula (which can range from > 15 to < 40 kg/m^2), and depression (on a scale of 0–30). Higher scores indicate better sleep quality, greater BMI, and more depressive symptoms. Enter and save the data below in SPSS.

Sleep	*BMI*	*Depress*
10	17	9
8	18	8

(Continued)

(Continued)

Sleep	*BMI*	*Depress*
7	26	13
9	21	9
9	19	8
8	25	5
6	33	17
5	34	22
11	16	4
7	24	12
10	18.5	8
5	35	14
7	26.6	15
6	32	14
9	20	7
5	40	21
8	17.5	11
9	22	10
7	16.5	13
10	18	9

As always, **eyeball the data.**

Once the data are entered, run descriptive and frequency analyses for your variables, then check to make sure everything looks the way you expected it. When you are sure your data are accurate, proceed.

Let's first run the correlational analysis (do you remember how?) to see if all three variables are significantly correlated with each other. Review instructions from earlier in the lab for running a bivariate correlation, but this time choose all three variables. The bivariate correlation (that includes no covariates) is also called the *zero-order* correlation.

Your output should look like this:

Correlations

		Sleep	BMI	Depress
Sleep	Pearson Correlation	1	-.864**	-.856**
	Sig. (2-tailed)		.000	.000
	N	20	20	20
BMI	Pearson Correlation	-.864**	1	.802**
	Sig. (2-tailed)	.000		.000
	N	20	20	20
Depress	Pearson Correlation	-.856**	.802**	1
	Sig. (2-tailed)	.000	.000	
	N	20	20	20

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

Are the correlations significant? Yes—we see sleep quality is strongly correlated with both BMI [$r(20) = -.86$, $p < .001$] and depression [$r(20) = -.86$, $p < .001$], and depression and BMI are also strongly correlated [$r(20) = .80$, $p < .001$]. Now what happens if we control for depression (i.e., take away any covariation it has with the two variables of interest) using a partial correlation?

Run the partial correlational analysis.

1. Click *Analyze*
2. *Correlate*
3. *Partial*
4. Highlight the names of the two variables of interest (Sleep quality and BMI) and click the arrow to the top right to move them into the *Variables* box.
5. Highlight the name of the variable we want to control for (Depression) and click the arrow to the bottom right to move it to the *Controlling for:* box.
6. Now click *OK*.

Your output should look like this:

Correlations

Control Variables			Sleep	BMI
Depress	Sleep	Correlation	1.000	-.574
		Significance (2-tailed)	.	.010
		df	0	17
	BMI	Correlation	-.574	1.000
		Significance (2-tailed)	.010	.
		df	17	0

Think about the following questions (no peeking at the answers until you've given them a shot!):

1. Is the partial correlation still significant?
2. What does this information tell you about the relationship between sleep quality and BMI?
3. If the partial correlation had not been significant, what could we conclude?

ANSWERS

1. Is the partial correlation still significant?

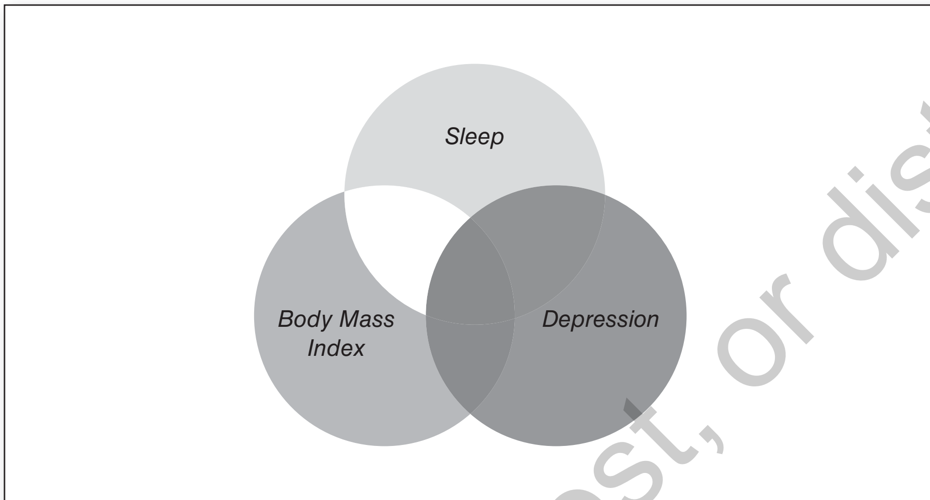
   Yes.

2. What does this information tell you about the relationship between sleep quality and BMI?

   Sleep quality and BMI are related above and beyond associations with depression, meaning the relationship is not solely due to the statistical influence of depression. While the partial correlation does not imply cause and effect, it provides evidence that a relationship may exist that is not due to a confounding effect of a third variable (or at least not the covariate we measured).

3. If the partial correlation had *not* been significant, what could we conclude?

That the zero-order (bivariate) correlation between sleep quality and BMI was due to associations with depression rather than a unique association. See the visual below to aid in understanding—we can see that the only "association" between sleep quality and BMI would be due to overlapping associations with depression. Pretty neat!



REFERENCES

Alvaro, P. K., Roberts, R. M., & Harris, J. K. (2013). A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. *Sleep*, *36*(7), 1059–1068. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3669059/pdf/aasm.36.7.1059.pdf

Cappuccio, F. P., Taggart, F. M., Kandala, N. B., Currie, A., Peile, E., Stranges, S., & Miller, M. A. (2008). Meta-analysis of short sleep duration and obesity in children and adults. *Sleep, 31*(5), 619–626. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2398753/pdf/aasm.31.5.619.pdf

Johnston, E., Johnson, S., McLeod, P., & Johnston, M. (2004). The relation of body mass index to depressive symptoms. *Canadian Journal of Public Health, 95*(3), 179–183. Retrieved from http://journal.cpha.ca/index.php/cjph/article/viewFile/489/489