Module 5 Homework

There are a total of 40 points for this homework assignment.

1. Which of the following statements about Pearson’s correlation coefficient is **false**?
   1. It can be used as a measure of effect size
   2. It varies between –1 and +1.
   3. It cannot be used with binary variables (those taking on a value of 0 or 1).
   4. It should be used on ranked data.
2. Which of the following best describes covariance?
3. It is an unstandardized version of the correlation coefficient.
4. It is a measure of the strength of relationship between two variables.
5. It is dependent on the units of measurement of the variables.
6. All of the above.
7. The correlation between two variables *A* and *B* is 0.12 with a significance of *p* < .01. What can we conclude?
8. There is a substantial relationship between A and B
9. There is a weak relationship between A and B
10. Variable A causes variable B
11. Variable A causes a significant effect in variable B
12. How much variance is shared by two variables with a correlation of .90?
13. 81%
14. 18%
15. 9%
16. None of the above
17. When interpreting a correlation coefficient, it is important to look at:
18. The significance of the correlation coefficient.
19. The magnitude of the correlation coefficient.
20. The sign of the correlation coefficient.
21. All of the above.
22. If there was a perfect positive correlation between two interval/ratio variables, Pearson's *r* test would give a correlation coefficient of:
23. –0.33
24. +1.00
25. +0.88
26. –1.00
27. What is the name of the test that is used to assess the relationship between two ordinal variables?
28. Spearman's rho
29. Phi
30. Cramér's V
31. Chi-square
32. A negative correlation shows that:
33. Two variables are unrelated.
34. As one score increases, so does the other.
35. As one score increases, the other decreases. We can never say for certain whether two variables are or are not related.
36. Both a and c.
37. In order for accurate measures of the linear relationship between two variables to be achieved, what type of data is required if using Pearson’s correlation coefficient?
38. Interval/Ratio
39. Ordinal
40. Nominal
41. Under what conditions should Spearman’s correlation coefficient be used?
42. When the variables are distributed normally.
43. When the variables are distributed parametrically.
44. When the variables are not normally distributed
45. When there are many tied ranks

Now let’s do some problems. I’ll lead you through them and simply work through the steps.

Access the file titled Exam Anxiety.sav in this week’s folder. When you open it, you will see that there are 4 variables: exam performance, exam anxiety, and time spent revising. (There is also a gender variable that you can ignore for now). You will be looking for a relationship between **exam performance** (score out of 100) and **exam anxiety** (out of 100). [This is a rather timely example in view of your first statistics exam coming up!]

1. Since you are looking for a relationship between two variables that are interval/ratio data we are going to calculate a Pearson’s correlation coefficient. Since we are interested in the correlation between 2 variables, you should choose the “bivariate” analysis. Even if we had 3 or 4 variables to correlate, bivariate correlation gives you the correlation between 2 variables *at a time*. As your textbook says on page 257, the two assumptions you should consider are linearity and normality. Let’s start with linearity. Create a simple scatterplot to make sure the data points are arranged in a *mostly* linear way and paste your scatterplot for #11. Consider **exam anxiety** as the independent/predictor variable so it will be on the horizontal axis (x), and **exam performance** as the dependent/outcome variable on the vertical axis (y).Be sure to include a title for your scatterplot.

Chart, scatter chart

Description automatically generated

1. Since the data looked linear we are OK for that assumption. For normality we should look at two P-P plots (one for exam performance, and one for exam anxiety) as shown on page 258. We discussed P-P plots earlier when we covered testing for assumptions (Chapter 6). To create a P-P plot, you need to go to Analyze- Descriptive Statistics. If you don’t remember, go to page 184.

Paste your two Normal P-P plots for #12. Note: you are probably going to create 4 graphs in your output file but only paste the 2 Normal P-P plots (Exam anxiety and Exam Performance).

Chart, scatter chart

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

1. How would you interpret the two P-P plots? Type your answer for #13. Keep in mind though, that *a P-P plot is simply a visual interpretation and not a statistical test*.

Both variables seem to deviate from the line of normality in at least one section on each P-P plot because the dots are not close to their respective lines on the plots.

1. From the P-P results the data might not be perfectly normal. In fact, if you ran a K-S or Shapiro-Wilk test of normality, you would find that the data for both variables deviates significantly from normality. However, since there are 103 subjects, we can assume that the central limit theorem gives us confidence to go forward with the Pearson’s *r* in SPSS. Go ahead and run a correlation analysis between “exam anxiety” and “exam performance”. What *r*-value did you get?

*r* = -0.441

1. According to your text, would this *r*-value be considered a small, medium or large effect size?

The *r*-value would be considered a medium effect size since it is near -0.5.

1. What is your coefficient of determination (R2) value?

R^2= 0.194= 19.4%

1. Type a statement of conclusion for your R2 results (that is, what does it tell you)?

The results of R-squared tells me that exam anxiety accounts for only 19.4% of the variability in exam performance.

1. Was the correlation statistically significant and what is the p-value?

**(**Remember….if SPSS says the p-value is .000 it’s really <.001. Also remember….significant doesn’t always mean meaningful….because you can get a significant correlation if you just have a lot of subjects. So, you might find a correlation of r= .21 to be significant if your sample size is large enough, but is that really meaningful?)

The p level is p < 0.001, and the correlation is statistically significant.

1. If the hypotheses were written as a relationship or no relationship between these two variables….write a sentence that could go in the results section of a manuscript. (Be sure to use Field’s tips on “How to report correlation coefficients” in your textbook.) “This study found that \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.”

This study found that there is a complex inter-relationship between the two variables because as exam anxiety increases, the grade obtained in that exam decreases.

1. Now, let’s bring gender into the picture. You are interested in knowing whether the relationship between exam performance and exam anxiety is the same for men and women. To determine this, run a separate test of correlation for men and women by using the “split file” function. Record the Pearson’s *r* for men and for women here.

Men *r*-value= -0.506

Women *r*-value= -0.381

OK. On to our next problem…

You are interested in knowing the relationship between the variables “student satisfaction” and “CI satisfaction” responses to a question regarding an acute care clinical rotation. The possible item responses were on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Please set the alpha at .05 and perform a two-tailed test.

|  |  |
| --- | --- |
| **Student Satisfaction** | **CI satisfaction** |
| **1** | **2** |
| **1** | **3** |
| **2** | **4** |
| **5** | **5** |
| **3** | **4** |
| **2** | **2** |
| **5** | **3** |
| **1** | **4** |
| **3** | **3** |

I’ve received 9 student/CI sets of data so far. Here are the data:

1. The responses are ordinal data (since rating was on a Likert scale) so we know that we should go with a Spearman’s rho correlation. You can quickly enter this data into a new SPSS file, or you could calculate by hand if you are especially energetic!

Record the Spearman’s rho that you got here. Be sure to use the appropriate designation for the Spearman’s rho. *rs*= 0.362

1. Was it significant? Why?

No, it was not significant because our p-value is greater than 0.05 (p-value= 0.339).

1. Write a short results sentence describing your findings.

There was no statistical significance found between a student’s satisfaction at a clinic and the clinic’s care satisfaction, meaning a student’s satisfaction at the clinic had no noticeable affect on how satisfied their patients were with their care at the clinic.

1. Would it be correct to use a Kendall’s tau in this case? Why or why not?

Yes, because satisfaction, for both variables, is being ranked.

1. Re-do the analysis as a Kendall’s tau….what value of tau did you get? Indicate your value here using the appropriate designation for Kendall’s tau. *­T*= 0.305
2. Was it significant? Why?

No, it was not significant because our p-value is greater than 0.05 (p-value= 0.309).

Next problem:

I just left a meeting in Alexandria and the APTA is reporting that women did considerably better on the last board exam nationwide as compared to men. I am giving you a random sample of 20 students from last year’s class. Can you tell me the relationship between these two variables? For gender, the “0” are men and the “1” are women and I’m giving you their scores. Please set the alpha at .05 and perform a two-tail test. Use dataset “Gender and Exam” in Module 5 Homework Materials.

1. Since one variable is a discrete dichotomy (gender) and the other is an interval/ratio variable, which correlation coefficient should you calculate?

You should calculate Pearson’s correlation coefficient because one variable is a interval/ratio variable.

1. What is your calculated correlation value? Please use the appropriate designation.

*r*= 0.227

1. Was it significant? Why?

No, it was not significant because the p-value is greater than 0.05 (p-value= 0.337).

1. Now recode gender such that “0”s are women and “1”s are men and re-run your analysis. What is your calculated correlation value? \_\_\_\_\_\_\_\_

*r*= 0.227

1. Compare this correlation value with the one you got for #28 and explain your results.

The r-values found in questions 28 and 30 were the same, insignificant value (*r*= 0.227), meaning an individual’s gender did not affect how well they did on the board exam (ie: there was no change in *r* even after the genders swapped scores).

For the next problem, open the file in your Module 5 Homework Materials titled Data Set Part and Partial. [the name of the dataset should give you a hint as to the type of correlation you will run!] If you have not viewed the youTube video by Dr. Grande in your Module 5 Homework materials, you should do so now.

1. OK….let’s say you were interested in the relationship between length of ICU stay and ICU mobility score at initial evaluation (1 is low score and 10 is high score) in 20 patients following lung transplantation. But you are thinking that age may play a role in this relationship. If you select “Zero-order correlations” as an option to include in your analysis, you will get correlation coefficients between ICU mobility and length of ICU stay when you do not account for age (eg. Zero-order), and when you do account for age. You will also get the correlation coefficient between age and mobility score and age and length of stay.

What is the correlation between ICU stay and ICU mobility when age is NOT considered? Hint: The part of your output that contains zero-order correlations mean that you are not controlling for any variable. The next part of your output (although not labelled as such is called first order correlation because it controls for **one** variable…age.] = -0.884

1. What is the correlation between LOS and mobility score when you DO control for age? = -0.597
2. Is this a part or partial correlation? Partial
3. Was the correlation between LOS and mobility score when you DO NOT control for age significant?

Yes, it was significant because the p-value is less than 0.05 (p-value < 0.001).

1. Was the correlation between LOS and mobility score when you DO control for age significant?

Yes, it was still significant because the p-value is less than 0.05 (p-value= 0.007).

1. What is your coefficient of determination (R2) value for the zero order correlation?

R2= 0.781

1. What is your coefficient of determination (R2) value for the first order correlation?

R2= 0.356

1. How would you interpret the answers you got for #32 thru #38? Please explain these results. Do they make sense to you?

Age does have a relationship with and influences the mobility and length of stay of ICU patients, and there is a strong relationship between the mobility and length of stay of ICU patients even when age is not a factor to consider. This is due to the correlation, with age and without age being a factor, between ICU length of stay and ICU mobility both being statistically significant. Yes, this does make sense to me because I could see how a patients’ ability to move would affect how long they stay in the ICU (the less mobile one is= the longer they have to stay in ICU and vice versa), regardless of age (although age does/can play a factor; the older they are, the less mobile an ICU patient will probably be, leading them to stay in ICU longer; whereas, the younger an ICU patient is, the faster they can heal to become more mobile, leading to a shorter ICU stay).

1. Write a results sentence or two reporting your findings? Note that although Dr. Field reports confidence intervals when reporting results of correlations, SPSS does not calculate these for you when conducting a correlation analysis. If you wish to calculate these yourself (and earn 1 extra credit point!), use the method described in the narrative, described as “tricking” SPSS.

Age does significantly affect the mobility of an ICU patient and their length of stay within the ICU (p-value= 0.007). In general, however, the mobility of an ICU patient was found to have a significant relationship with how long a patient stays in the ICU (p-value < 0.001; confidence intervals (-1.115, -0.652)), even without age being a factor.

Regarding power, there are differing opinions on whether or not you should conduct a power analysis for correlations. Later, when we study power in reference to statistical tests of differences, we will discuss in greater detail.

G\*Power is a statistical package that calculates power and or sample sizes needed to achieve a specified level of power. It can be downloaded from this website. It’s free. Here is the website: <http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/>

Remember that power is the ability to detect differences when they exist. With correlation, we are not testing differences so the sample size needed to have adequate power in this context is in reference to whether a correlation coefficient differs from zero.

If you are conducting a simple bivariate correlation there isn’t an option for using the G\*Power program. But you should probably assume that you need at least 30 subjects for a robust analysis. I located a website that will calculate it for you. Here is the website link: [Correlation sample size | Sample Size Calculators](http://www.sample-size.net/correlation-sample-size/)

Go ahead and play with it. Note that when you put a large anticipated correlation (Ex. r=0.80, you need a very small sample size (n=10), but if your anticipated correlation is small (r=0.25), you need a very large sample size (n=123). That makes sense, but why would you ever want a sample size that would give you significance when the strength (r-value) of the relationship is only 0.23?? Also, even if your power analysis indicates you only need 10 subjects to show significance, I would still strive for a sample of 30 because, 1) you are more likely to have normal distribution and thus be justified in using parametric analyses and, 2) a sample of 30 is more likely to resemble your target population.

Personally, I don’t see the relevance of a power analysis for a correlation study but I MIGHT be convinced otherwise by a strong argument about why it is important to do so. I just haven’t heard one yet. ☺ Note that Dr. Field makes no reference to power when discussing correlations.