

## 01c: Review of Pearson Correlation

### 1. Characteristics

Pearson  $r$

- $r$  = Pearson's correlation coefficient
- used to assess a linear relation between two quantitative variables
- ranges from -1.00 to 1.00
- $r = 0.00$  means no linear relation, but there may be a non-linear relation
- the closer  $r$  to 1.00 in absolute value, the stronger the relationship, the closer to 0.00, the weaker the relationship

### 2. General Interpretation

a. Found  $r = -.77$  between car horsepower and MPG. What does this tell us; what is the interpretation of this correlation in terms of the variables examined?

**Answer:** Negative relationship ---- The greater the car horsepower, the lower will be expected MPG.

b. Found  $r = .40$  between reading self-efficacy and reading test scores. What does this tell us; what is the interpretation of this correlation in terms of the variables examined?

**Answer:** Positive relationship --- The higher reading self-efficacy, the higher will be reading test scores, on average.

c. Found  $r = .00$  between student weight and interest in mathematics. What does this tell us?

**Answer:** No linear relationship --- student weight and interest in mathematics does not appear to be linearly related; one cannot predict interest in mathematics based upon one's weight.

### 3. Reading Published Correlation Tables

Example 1

Menon, ST (2001). Employee empowerment: An integrative psychological approach. Applied psychology: An international review, 50, 153-180. Source:

[http://www.bwgriffin.com/gsu/courses/edur9131/activities/Menon\\_ST\\_2001\\_employee\\_empowerment\\_Applied\\_Psychology.pdf](http://www.bwgriffin.com/gsu/courses/edur9131/activities/Menon_ST_2001_employee_empowerment_Applied_Psychology.pdf)

TABLE 1  
Means, Standard Deviations, and Intercorrelations

Item*	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<b>Perceived Control</b>																
1. PC1	4.58	1.18														
2. PC2	4.50	1.26	.57													
3. PC3	4.72	1.11	.44	.61												
4. PC4	4.47	1.29	.75	.57	.45											
5. PC5	4.87	1.01	.43	.48	.36	.46										
<b>Perceived Competence</b>																
6. COMP1	5.54	0.62	.15	.14	.25	.15	.06									
7. COMP2	5.53	0.69	.09	.10	.16	.11	.13	.45								
8. COMP3	5.56	0.57	.16	.13	.28	.18	.12	.66	.63							
9. COMP4	5.39	0.79	.24	.27	.32	.30	.15	.33	.37	.43						
10. COMP5	5.26	0.77	.19	.18	.37	.26	.15	.37	.38	.45	.47					
<b>Goal Internalisation</b>																
11. GI1	4.45	1.18	.29	.38	.40	.34	.42	.05	.20	.18	.23	.19				
12. GI2	4.67	1.11	.29	.39	.43	.38	.38	.17	.11	.17	.24	.20	.68			
13. GI3	4.79	0.92	.34	.39	.44	.30	.35	.16	.21	.20	.32	.27	.46	.55		
14. GI4	4.38	1.19	.29	.36	.43	.39	.40	.10	.22	.22	.28	.25	.73	.69	.52	
15. GI5	5.25	0.95	.29	.34	.38	.36	.41	.08	.16	.25	.32	.29	.45	.52	.43	.52

\* Item wordings are available in Table 2.

Correlations < .12 nonsignificant. Correlations .12 to .15,  $p < .05$ . Correlations .16 to .19,  $p < .01$ . All other correlations,  $p < .001$

What do the numbers across the top row of this table indicate?

**Answer:** Note that each of the 15 variables are numbered in the first column. The numbers on the first row represent each of the 15 variables. Using numbers rather than repeat the names of each variable saves space and allows for a larger correlation matrix to be printed.

What is the mean for Perceived Competence 3 (Comp3)?

**Answer:**  $M = 5.56$  (see column 2)

What is the standard deviation for Goal Internalization 1 (GI1)?

**Answer:**  $SD = 1.18$  (see column 3)

What is the correlation between Perceived Control 2 and 4 (PC2 and PC4)?

**Answer:**  $r = .57$

What is the correlation between Perceived Control 3 and Goal Internalization 4 (PC3 and GI4)?

**Answer:**  $r = .43$

How do we determine which correlations are statistically significant?

**Answer:** Footnote at bottom of table tells us that any correlation larger than .12 (in absolute value) is significant at the .05 level or lower (e.g., .01 or .001).

Example 2

Source: Pintrich & De Groot (1990) Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Ed. Psychology*, 82, 33-40.

**Table 2**  
*Zero-Order Correlations Between Motivation and Self-Regulated Learning Variables and Performance*

Variable	Grade 1	Seat-work	Exams/Quizzes	Essays/Reports	Grade 2
<b>Motivation components</b>					
Intrinsic value	.25**	.21**	.20**	.27**	.30***
Self-efficacy	.34***	.19*	.24**	.25**	.36***
Test anxiety	-.24**	-.14	-.21**	-.14	-.23**
<b>Self-regulated learning components</b>					
Strategy use	.18*	.07	.20**	.19*	.20**
Self-regulation	.32***	.22**	.28**	.36***	.36***

Note.  $N = 173$ .

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

What is the correlation between Self-efficacy and Essays/Reports?

**Answer:**  $r = .25$

What is the correlation between Self-regulation and Grade 2?

**Answer:**  $r = .36$

What do the asterisks (\*) next to correlations tell us?

**Answer:** Any correlation marked with asterisk is significant, which means the null hypothesis of no relation is rejected for those two variables.

What is the interpretation of the correlation between Intrinsic Value and Seatwork?

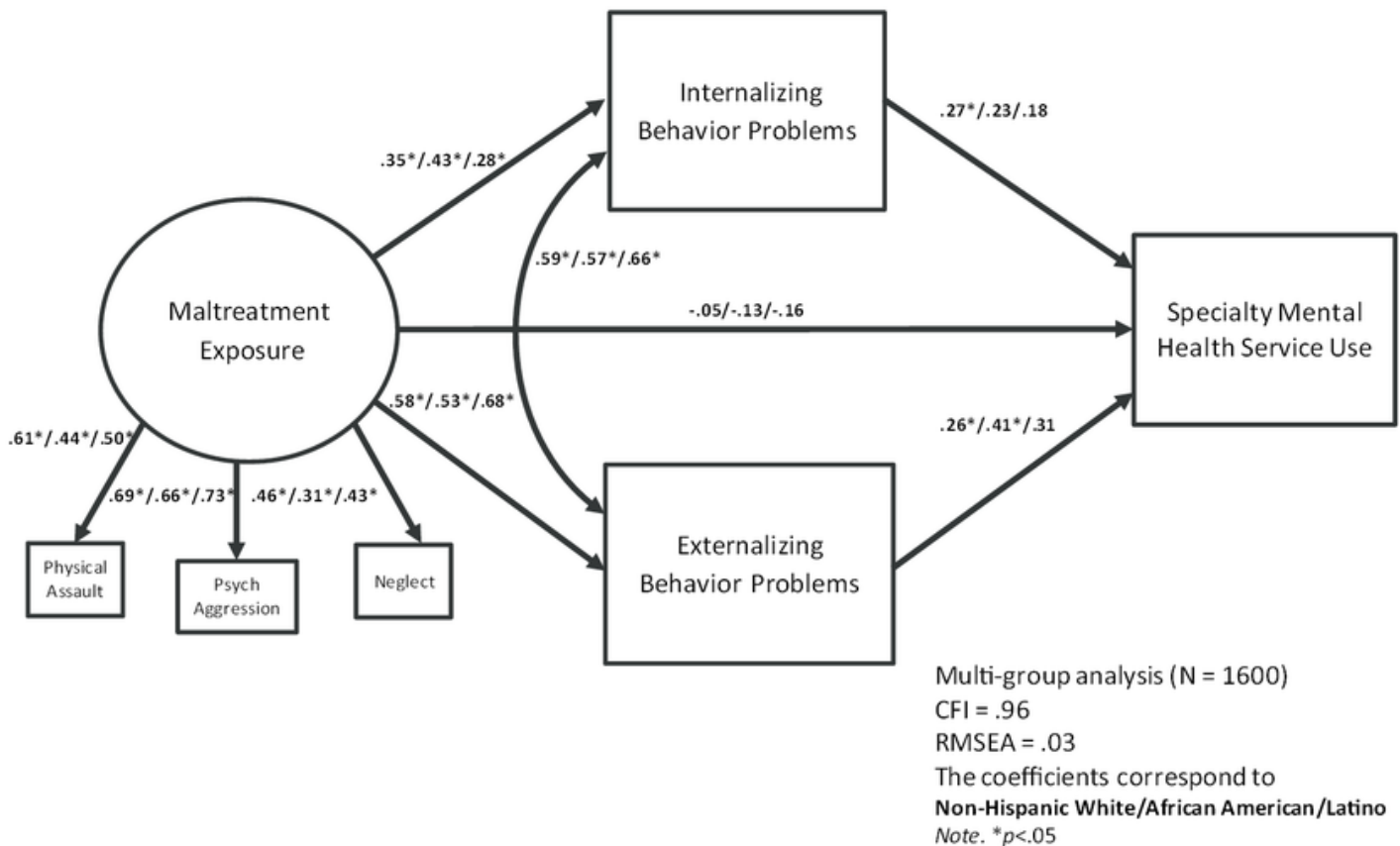
**Answer:** The correlation is  $r = .21$  and significant. This tells us there is a significant, positive correlation between intrinsic value and seatwork; the more students value learning intrinsically, the better they perform on seatwork.

What is the interpretation of the correlation between Test Anxiety and Exams/Quizzes?

**Answer:** The correlation is  $r = -.21$  and significant. There is a negative correlation between test anxiety and exams/quizzes scores; those students with greater test anxiety tended to obtain lower grades on exams and quizzes.

## Example 3

Source: Martinez, Gudino, & Lau (2013) Problem-Specific Racial/Ethnic Disparities in Pathways from Maltreatment Exposure to Specialty Mental Health Service Use for Youth in Child Welfare. *Child Maltreatment*, 18, 98-107.



The graph above shows a structural model, sometimes called a path model, among four variables (Maltreatment Exposure, Internalizing Behavior Problems, Externalizing Behavior Problems, and Specialty Mental Health Service Use), and three indicators of Maltreatment Exposure (a latent variable).

Lines with single arrows, like this  $\rightarrow$ , indicate IV and DV where the IV is hypothesized to predict or influence the DV. Curves with double arrows, like this  $\leftrightarrow$ , indicate related or correlated variables, but does not specify an IV or DV.

Next to each line or curve are numbers. If the number is adjacent a single-arrow line, then it is a regression coefficient. If the number is next to double-arrow curve, it is a correlation coefficient. You will also notice that for each line or curve there are three numbers. The diagram footnote explains that the numbers represent coefficients for three different groups: Non-Hispanic White/African American/Latino.

Which variables are hypothesized to be correlated?

**Answer:** Given that only two variables show double-arranged curves, they are Internalizing and Externalizing Behavior Problems.

What is the correlation between internalizing and externalizing behavior problem for Latinos, and what is the interpretation of this correlation?

**Answer:**  $r = .66$ . This correlation tells us that for Latinos, Externalizing and Internalizing Behavior Problems are positively, and significantly, related. The more one externalizes a problem, the more likely they are also to internalize a problem behavior.

Which group demonstrated the weakest correlation between internalizing and externalizing behavior problems?

**Answer:** African American with  $r = .57$ .

#### 4. Worked Example

A data file is linked below to provide practice with finding correlations with SPSS.

[http://www.bwgriffin.com/gsu/courses/edur8132/tests/math\\_sat.sav](http://www.bwgriffin.com/gsu/courses/edur8132/tests/math_sat.sav)

The data file contains three variables:

- **math\_sat:** Average mathematics SAT scores in each state.
- **pupil\_teacher\_ratio :** Average ratio of students to teacher in each state; a higher number indicates more students per teacher.
- **average\_teacher\_salary:** Average salary per teacher in each state in thousands of dollars, thus a figure of 25.000 means the average salary per teacher is \$25,000 per year.

Before obtaining correlation estimates, first consider how these three variables may be related. Assume the DV is Math SAT and the two IV are Pupil-teacher ratio (class size) and Average Teacher Salary.

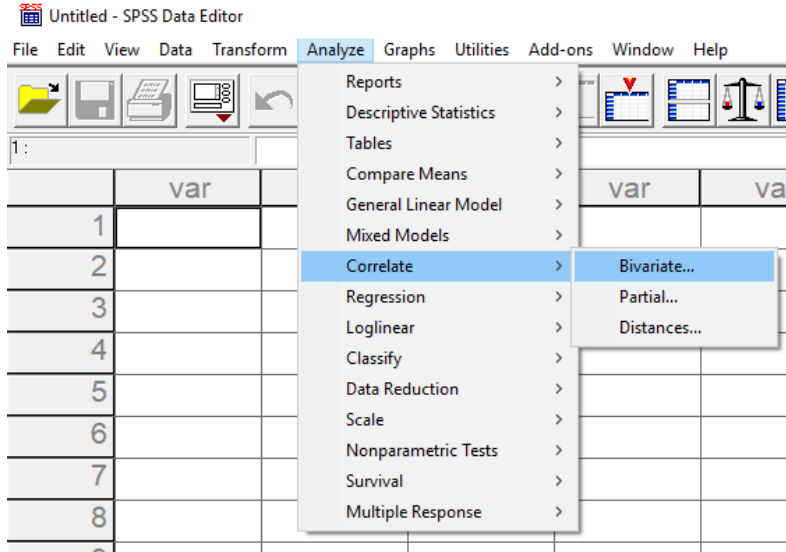
What might be the relation between Mat SAT and Pupil-teacher ratio (class size)?

**Answer:** Most assume that larger class sizes result in less learning, and if that is true, then these two variables should produce a negative relation; as class size increases, Math SAT scores should decline.

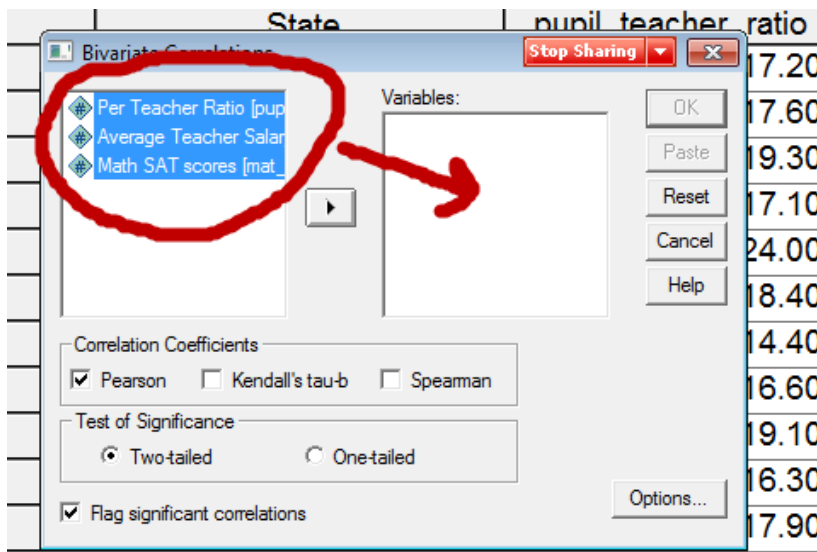
How might Average Teacher Salary and Math SAT be related?

**Answer:** Often educators argue that higher salaries attract better trained teachers, so one may expect that states with higher teacher salaries should also see higher Math SAT scores, a positive relation.

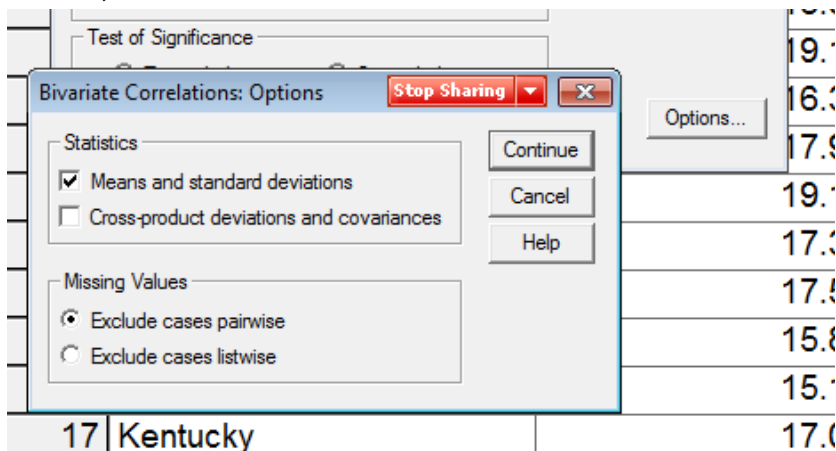
SPSS Commands to obtain Pearson correlations.



Move variables to be correlated to variables box.



Select Options then place a check next to Means and Standard deviations to obtain descriptive statistics for each variable,



## SPSS Results

## Descriptive Statistics

	Mean	Std. Deviation	N
Per Teacher Ratio	16.8580	2.26635	50
Average Teacher Salary in Thousands of Dollars	35.2800	5.96603	50
Math SAT scores	508.7800	40.20473	50

## Correlations

		Per Teacher Ratio	Average Teacher Salary in Thousands of Dollars	Math SAT scores
Per Teacher Ratio	Pearson Correlation	1	.010	.095
	Sig. (2-tailed)		.945	.510
	N	50	50	50
Average Teacher Salary in Thousands of Dollars	Pearson Correlation	.010	1	-.403**
	Sig. (2-tailed)	.945		.004
	N	50	50	50
Math SAT scores	Pearson Correlation	.095	-.403**	1
	Sig. (2-tailed)	.510	.004	
	N	50	50	50

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

What is the correlation between Math SAT and student-teacher ratio (find correlation in SPSS output). Do we reject or fail to reject the null hypothesis,  $H_0$ , at the .05 level of significance ( $\alpha = .05$ )? How would this correlation be interpreted?

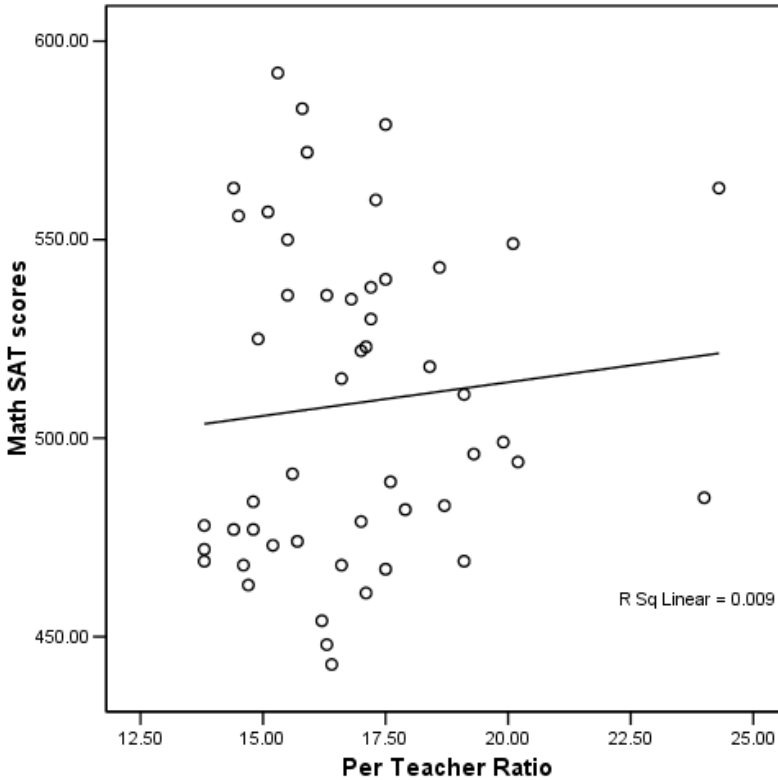
Recall hypothesis testing decision rule for p-values:

If  $p \leq \alpha$  reject  $H_0$ ; if  $p > \alpha$  fail to reject  $H_0$

## Answer:

- $r = .095$ , there is a weak, positive correlation between Math SAT and class size (Student-teacher ratio)
- Since the p-value for this correlation is  $p = .51$ , and since  $.51$  is larger than  $\alpha = .05$ , we fail to reject the null hypothesis and conclude there is no relation between Math SAT and class size.
- Interpretation: There is no association between class size ratio and SAT math scores; on average class size does not appear to predict or be related to SAT math scores.

Scatterplot below of Student-teacher Ratio and Math SAT scores by state.



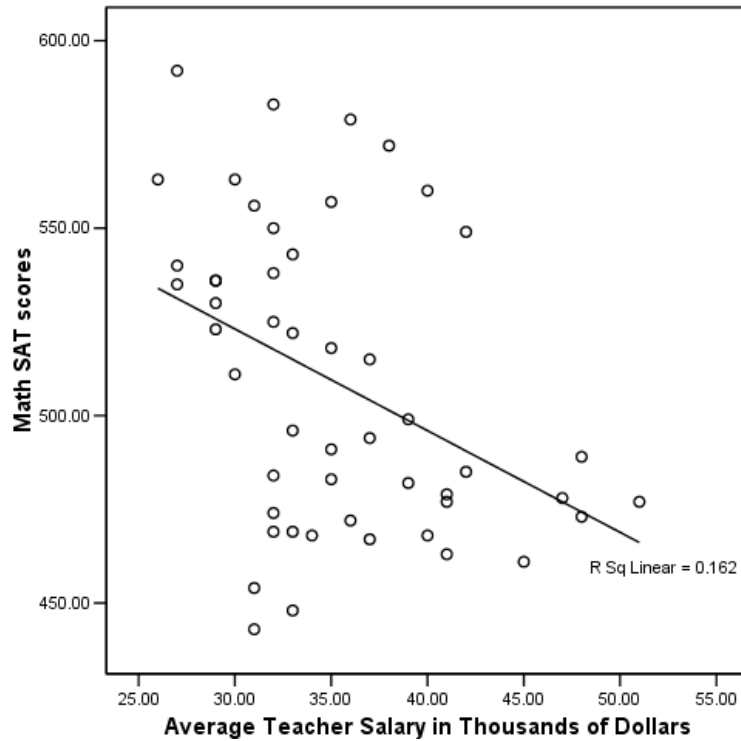
What is the correlation between Math SAT and teacher salary (find correlation in the SPSS output). Do we reject or fail to reject the null hypothesis,  $H_0$ , at the .05 level of significance ( $\alpha = .05$ )? How would this correlation be interpreted?

**Answer:**

- $r = -.403$ , there is a negative correlation between Math SAT and mean teacher salary
- Since the p-value for this correlation is  $p = .004$ , and since  $.004$  is less than  $\alpha = .05$ , we reject the null hypothesis and conclude there is a negative relation between Math SAT and teacher salary.
- Interpretation: There is a negative association between state mean teacher salary and SAT math scores; as mean teacher salary increases, mean Math SAT scores tend to decline.

Scatterplot below of Mean Teacher Salary and Math SAT scores by state.





## 5. APA Style Presentation

For those interested, below is an example of APA style presentation for the correlations obtained above.

Table 1: Correlations and Descriptive Statistics for State-level Mean Scores for Math SAT, Student-teacher Ratio, and Teacher Salary

	1	2	3
1. Math SAT	---		
2. Student-Teacher Ratio	.095	---	
3. Teacher Salary	-.403*	.010	---
M	508.78	16.86	35.28
SD	40.20	2.27	5.97

Note.  $n = 50$ .

\*  $p < .05$ .

Written report has two components – inference (was  $H_0$  rejected) and interpretation (what does result mean):

There is a statistically significant association between state-level mean mathematics SAT scores and teacher salary. There is not, however, an association at the .05 level between state-level mean mathematics SAT scores and student-teacher class ratio. Results show that states with higher salaried teachers tend to have lower mathematics SAT scores, while states with lower salaried teachers tend to have higher mathematics SAT scores; stated differently, there is a negative association between mathematics SAT scores and mean teacher salary. Results also show that mathematics SAT scores are unrelated to student-teacher class ratio, and this suggests that mathematics SAT scores are similar for both large and small sized classes across the states.