

Name _____

Correlation and Regression Lab, Phase 2

1. *Why can't you make causal claims as a result of analysis of correlational designs?*

- Causal arrow problem
- Third variable problem

It is because of the fact that nothing is being manipulated that you must be careful to interpret any relationship that you find. That is, you should understand that determining that there *is* a linear relationship between the two variables doesn't tell you anything about the causal relationship between the two variables—correlation does not imply causation. If you did not know anything at all about the person's IQ, your best guess about that person's GPA would be to guess the typical (average, mean) GPA. Finding that a correlation exists between IQ and GPA simply means that knowing a person's IQ would let you make a better prediction of that person's GPA than simply guessing the mean. You don't know for sure that it's the person's IQ that determined that person's GPA, you simply know that the two tend to covary in a predictable fashion.

If you find a relationship between two variables, A and B, it may arise because A directly affects B, it may arise because B affects A, or it may be because an unobserved variable, C, affects both A and B. In this specific example of IQ and GPA, it's probably unlikely that GPA could affect IQ, but it's not impossible. It's more likely that either IQ affects GPA or that some other variable (e.g., test-taking skill, self-confidence, patience in taking exams) affects both IQ and GPA.

Consider this report from BBC News 3/19/2008

'Healthier hearts' for cat owners

Cat owners appear to have a much lower risk of dying from a heart attack than their feline-spurning counterparts, a study suggests.

Researchers looked at nearly 4,500 adults and found that cat ownership was related to a 40% lower risk of suffering a fatal heart attack. The team speculated that having a cat may reduce stress and anxiety, and so protect against cardiovascular disease. The findings of the study were unveiled at the International Stroke Conference. The study, led by Professor Adnan Qureshi at the University of Minnesota, suggested that even those who no longer owned a cat benefited from these protective effects. But specifically, some 3.4% of those who owned a cat during the duration of the study died from a heart attack, compared with 5.8% of those who did not. The benefits held true even after the researchers adjusted for cardiovascular risk factors such as blood pressure, diabetes, smoking and high cholesterol.

However the authors warned against impulsive cat purchases. They said while cats may indeed have a calming effect, it was unclear whether the kind of people who opted for a cat in the first place may have a lower risk of heart attack. This study did not examine the advantages of having a dog, although previous research has suggested this too may have health benefits above and beyond taking them for walks. The Pet Health Council notes "there is an increasing amount of research proving that contact with animals can bring real physiological and psychological benefits including reducing stress, helping to prevent illness and allergies, lowering blood pressure, aiding recovery and boosting fitness levels. "Research has also shown that pet owners make fewer annual visits to the doctors than non pet owners proving the saying, 'a pet all day keeps the doctor away!'"

2. Computation of correlation coefficient and regression coefficient

A researcher has developed a new test of self-esteem. To evaluate the reliability of the test, the researcher takes a sample of $n = 8$ participants. Each individual takes the test on a Monday evening, then returns two weeks later to take the test again. The two scores for each individual are reported below. Are the participant's scores on the two tests related to one another? (This measure is often referred to as test-retest reliability.)

Note that you might think of this analysis as a single-factor repeated measures ANOVA, should you be interested in determining if the scores on the first test differed from scores on the second test. However, you should note that the question asks about a relationship between the two tests.

Statistical Hypotheses: $H_0: r = 0$ $H_1: r \neq 0$

Decision Rule: Set $\alpha = .05$, and with a sample of $n = 8$ students, your obtained r must exceed .707 to be significant (using Table B.6, $df = n - 2 = 6$, two-tailed test).

If the scores on the two tests were as illustrated below, you'd need to compute the product of the scores on the two tests for each participant, then sum those products.

	First Test	Second Test	X*Y
	13	15	
	5	4	
	12	13	
	11	11	
	9	10	
	14	13	
	8	8	
	8	6	
Sum	80	80	
SS	64	100	

$$r = \hat{\rho} = \frac{SP}{\sqrt{SS_X SS_Y}} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n} \right) \left(\sum Y^2 - \frac{(\sum Y)^2}{n} \right)}}$$

Compute r , then determine if it's significant. If so, then determine the regression equation.

The regression equation is: $\hat{Y} = bX + a$

To compute the slope (b) and y-intercept (a) we would use the following simple formulas, based on quantities already computed for r (or easily computed from information used in computing r).

$$b = \frac{SP}{SS_X} \qquad a = \bar{Y} - b\bar{X}$$

3. The Coefficient of Determination (r^2) [your old measure of effect size]

The variability shared between the two variables is measured by the *coefficient of determination* (r^2). This measure is quite informative. For example, suppose that you have a significant correlation of .4 (which would be significant if $df = 23$). Those two variables share only .16 (16%) of their variability. That means that there is .84 of the variability in each variable that is not shared with the other variable. We call this statistic ($1 - r^2$) the *coefficient of alienation*. Thus, even though a correlation of .4 or less may be statistically significant with a sufficiently large sample size, it may not be practically significant (given the low coefficient of determination).

4. Learning how to read SPSS output

Here are the data:

	GPA	IQ	Study	HSGPA	SAT	Extra	Sleep	Motivate	Credits	SES	Drug	Time	StudSkill
1	3.20	.	30.00	3.70	.	15.00	7.50	8.00	15.00	7.00	4.00	9.00	7.00
2	2.90	.	3.00	.	1400.00	2.00	7.00	6.00	15.00	7.00	2.00	2.00	5.00
3	3.40	.	6.00	4.40	1680.00	1.00	7.00	8.00	16.00	4.00	3.00	8.00	8.00
4	3.48	.	30.00	3.20	.	15.00	8.00	9.00	15.00	6.00	4.00	10.00	9.00
5	3.70	.	10.00	3.90	1350.00	1.00	7.00	9.00	15.00	6.00	5.00	7.00	8.00
6	3.49	.	8.00	3.29	1360.00	3.00	6.50	7.00	15.00	8.50	8.00	6.00	7.00
7	3.40	.	10.00	3.60	1110.00	2.00	7.00	9.00	16.00	7.00	2.00	9.00	7.00
8	3.20	.	14.00	3.40	1350.00	18.00	8.00	7.00	15.00	8.00	7.00	6.00	9.00
9	3.40	130.00	6.00	3.80	1180.00	0.00	6.00	7.00	14.00	7.00	2.00	5.00	7.00
10	3.70	125.00	5.00	.	1340.00	9.00	8.00	8.00	18.00	7.00	2.00	10.00	10.00
11	3.40	150.00	30.00	3.10	1400.00	7.00	6.00	8.00	17.00	7.00	4.00	5.00	7.00
12	3.53	122.00	20.00	3.12	1340.00	4.00	7.00	9.10	15.00	8.00	2.00	6.00	9.00
13	3.30	.	10.00	3.67	.	5.00	7.00	7.00	15.00	7.00	3.00	7.00	7.00
14	3.74	.	12.00	3.76	1800.00	10.00	8.00	9.00	15.00	3.00	2.00	7.00	5.00
15	3.55	122.00	12.00	92.00	1200.00	18.00	7.00	8.00	16.00	6.00	2.00	7.00	7.00
16	3.15	120.00	9.00	3.70	1280.00	6.00	5.50	7.00	16.00	5.00	2.00	5.00	6.00
17	3.30	.	15.00	3.80	.	5.00	6.00	6.00	16.00	7.00	3.00	8.00	7.00
18	3.40	.	40.00	3.70	1130.00	20.00	7.00	10.00	17.00	1.00	2.00	8.00	8.00
19	3.78	.	15.00	3.70	1240.00	5.00	7.00	9.00	16.00	7.00	3.00	9.00	9.00
20	3.50	.	22.00	3.87	.	2.00	8.00	10.00	15.00	8.00	2.00	9.00	9.00
21	3.50	150.00	7.00	4.70	1330.00	1.00	9.00	10.00	16.00	6.00	5.00	9.00	10.00

First of all, here is a matrix of r values from your data collected earlier:

Correlations

	Pearson Correlation Sig. (2-tailed) N	College GPA	IQ	Hours/Week Study	High School GPA	Math+ Verbal SAT	Hours Extracurricular	Hours Sleep	Motivation for GPA	Credits/Semester	Family SES	Drug/Alcohol Use	Time Management	Study Skills
College GPA		1.000	.015	.009	.149	.130	-.044	.297	.608**	.203	-.117	-.016	.544*	.434*
			.974	.970	.541	.632	.849	.191	.003	.378	.612	.946	.011	.049
		21.000	7	21	19	16	21	21	21	21	21	21	21	21
IQ		.015	1.000	.352	-.354	.491	-.376	.313	.465	.176	.050	.944**	.096	.232
		.974	7	.439	.492	.263	.406	.495	.293	.705	.915	.001	.838	.616
		7	7.000	7	6	7	7	7	7	7	7	7	7	7
Hours/Week Study		.009	.352	1.000	-.109	-.232	.594**	.036	.432	.155	-.272	-.025	.279	.159
		.970	.439	21.000	.657	.387	.005	.878	.051	.504	.232	.914	.220	.491
		21	7	7	19	16	21	21	21	21	21	21	21	21
High School GPA		.149	-.354	-.109	1.000	-.203	.385	-.016	-.050	.150	-.037	-.196	-.050	-.129
		.541	.492	.657	.486	.486	.103	.950	.839	.539	.880	.421	.838	.598
		19	6	19	19.000	14	19	19	19	19	19	19	19	19
Math+ Verbal SAT		.130	.491	-.232	-.203	1.000	-.148	.265	.552**	-.003	-.233	.087	-.099	-.258
		.632	.263	.387	.486	16.000	.586	.322	.781	.607	.385	.747	.714	.334
		16	7	16	14	16	16	16	16	16	16	16	16	16
Hours Extracurricular		-.044	-.376	.594**	.385	-.148	1.000	.213	.097	.227	-.342	.056	.201	.058
		.849	.406	.005	.103	.586	.353	.353	.676	.323	.129	.810	.382	.802
		21	7	21	16	16	21.000	21	21	21	21	21	21	21
Hours Sleep		.297	.313	.036	-.016	.265	.213	1.000	.552**	-.003	-.034	.162	.527*	.529*
		.191	.495	.878	.950	.322	.353	.010	.010	.990	.885	.483	.014	.014
		21	7	21	19	16	21	21.000	21	21	21	21	21	21
Motivation for GPA		.608**	.465	.432	-.050	-.075	.097	.552**	1.000	.179	-.343	-.139	.589**	.513*
		.003	.293	.051	.839	.781	.676	.010	.010	.438	.128	.548	.005	.017
		21	7	21	19	16	21	21	21.000	21	21	21	21	21
Credits/Semester		.203	.176	.155	.150	-.139	.227	-.003	.179	1.000	-.304	-.210	.324	.279
		.378	.705	.504	.539	.607	.323	.990	.438	.181	.361	.361	.152	.220
		21	7	21	19	16	21	21	21	21.000	21	21	21	21
Family SES		-.117	.050	-.272	-.037	-.233	-.342	-.034	-.343	-.304	1.000	.348	-.112	.209
		.612	.915	.232	.880	.385	.129	.885	.128	.181	.122	.122	.628	.364
		21	7	21	19	16	21	21	21	21	21.000	21	21	21
Drug/Alcohol Use		-.016	.944**	-.025	-.196	.087	.056	.162	-.139	-.210	.348	1.000	-.021	.222
		.946	.001	.914	.421	.747	.810	.483	.548	.361	.122	.122	.929	.333
		21	7	21	19	16	21	21	21	21	21.000	21	21	21
Time Management		.544*	.096	-.279	-.050	-.099	.201	.527*	.589**	.324	-.112	-.021	1.000	.626**
		.011	.838	.220	.838	.714	.382	.014	.005	.152	.628	.929	1.000	.002
		21	7	21	19	16	21	21	21	21	21	21	21.000	21
Study Skills		.434*	.232	.159	-.129	-.258	.058	.529*	.513*	.279	.209	.222	.626**	1.000
		.049	.616	.491	.598	.334	.802	.014	.017	.220	.364	.333	.002	.002
		21	7	21	19	16	21	21	21	21	21	21	21	21.000

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

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

Let's look at the relationship between GPA and Motivation for GPA:

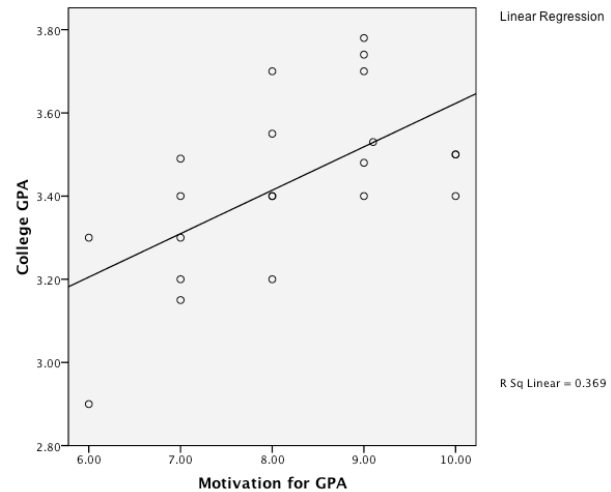
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.608 ^a	.369	.336	.17362

a. Predictors: (Constant), Motivation for GPA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.335	1	.335	11.126	.003 ^a
	Residual	.573	19	.030		
	Total	.908	20			

a. Predictors: (Constant), Motivation for GPA

b. Dependent Variable: College GPA



Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.579	.258		9.999	.000
	Motivation for GPA	.104	.031	.608	3.336	.003

a. Dependent Variable: College GPA

As you can see, there is a significant positive linear relationship between GPA and motivation for GPA, $r(19) = .608, p = .003$.

One can think of r^2 as simply that (r squared). Alternatively, you can think of this formula, which will yield the same result:

$$r^2 = \frac{SS_{Regression}}{SS_{Total}} = \frac{.335}{.908} = .369$$

$$r^2 = 1 - \frac{SS_{Error}}{SS_{Total}} = 1 - \frac{.573}{.908} = .369$$

You'll note that SPSS also provides an adjusted r^2 . That's a measure that we won't make use of, but it corrects for the complexity of the model being used (and we'll always use a simple model with one predictor variable. Nonetheless, here's the formula being used:

$$Adj\ r^2 = 1 - \frac{SS_{Error}/df_{Error}}{SS_{Total}/df_{Total}} = 1 - \frac{.573/19}{.908/20} = .336$$

The regression equation would be:

$$\text{Predicted GPA} = (.104) * (\text{Motivation for GPA}) + 2.579$$

Thus, if a person had a score of 10 on the motivation scale, you would predict that person would have a GPA of 3.62. If a person had a score of 1 on the motivation scale, you would predict that person would have a GPA of 2.68.

SPSS shows that the standard error of estimate is .174. If you applied your formula to the data, you'd also get .174:

$$SEE = \sqrt{\frac{SS_{Residual}}{df}} = \sqrt{\frac{.573}{19}} = .174$$

That is to say that the estimates from the regression equation should be fairly accurate predictors of actual data. If the data departed substantially from the line of best fit, the standard error of estimate would be much larger.