

Name Key
Please ALSO write your name on the back of the last page of the test.

Ag 400 - Quiz 4
Fall 2008

- A. Suppose that the following data were collected for a sample of 25 emphysema patients: Number of years each patient smoked (X); and a physician's rating of the "health" of the patient's lungs. The latter variable was measured on a scale of 0 to 100.

	(X) Years Smoked	(Y) Lung Health
Sum of the scores	500	1,000
Mean scores	20	40
Sum of the squares of the scores	10,625	41,600
Sum of the squares of the deviations about the mean	625	1,600
Sum of the cross products (Σxy)		-400

- (20) 1. Use a two-tailed test and the .05 level to test the statistical significance of the linear relationship between number of years smoked and lung health. Specify the following:

- a. State the Null Hypothesis and the Alternative Hypothesis.

H_0 : There is no linear relationship between number of years a person smoked and his/her lung health

H_a : There is a linear relationship

- b. Show the necessary calculations.

$$r = \frac{-400}{\sqrt{(625)(1600)}} = -.40$$

- c. Indicate degrees of freedom.

$$df = 23$$

- d. Reject or do not reject the Null Hypothesis.

Reject H_0

- e. What is the probability that you have made a Type I error in (d) above?

$$.01 < \alpha < .05$$

- f. What is the probability that you have made a Type II error in (d) above?

$$\beta = 0.00$$

- g. Interpret the meaning of your conclusion in (d) above precisely in terms of the problem. Be sure to describe the nature of the relationship, if any, in your answer.

There is a linear relationship between number of years a person smoked and her/his lung health. The relationship is negative. As the number of years of smoking increased, health of lungs declines.

- (12) 2. Suppose that Mr. Seegar, a 50 year old man has been smoking for 35 years. Estimate his lung health rating using the information given above.

$$b_{yx} = \frac{-400}{625} = -.64$$

$$a_{yx} = \bar{Y} - b_{yx} \bar{X} \quad 40 - (.64)(20) = 52.8$$

$$Y = 52.8 - (.64)(35) = 30.4$$

- (12) 3. Suppose that Mrs. Selda Smoke has a lung health rating of 60. How many years would you estimate she has been smoking?

$$b_{xy} = \frac{-400}{1600} = -.25$$

$$a_{xy} = 20 - (-.25)(40) = 30$$

$$X = 30 - .25(60) = \underline{\underline{15}}$$

- (16) 4. Indicate whether each of the following statements is true (T) or false (F) in terms of the above data. If any part of a statement is untrue, it should be marked false (F).

F a. The statistical unit in this problem is a year.

F b. The coefficient of determination for the above data is negative.

T c. a_{yx} is the estimated Lung Health rating for people who have never smoked.

F d. If you had access to the original data on which the above statistics were based, you could test the statistical significance of the relationship between years of smoking and lung health, by a t-test for the difference between two sample means, using \bar{X} and \bar{Y} as the two means.

F e. In this problem $b_{yx} \neq b_{xy}$ and $r_{yx} \neq r_{xy}$

F f. In any correlation-regression analysis, if r_{yx} is negative, b_{yx} , b_{xy} , a_{yx} , and a_{xy} are also negative.

F g. Based on the above analysis, you know that the relationship between "Years Smoked" and "Lung Health" is linear and negative in the population.

F h. Since data for both "Years Smoked" and "Lung Health" are for the same subjects, a paired t-test could be used to test the statistical significance of the relationship between these two variables.

- (40) B. The attached output was obtained using data from a survey of a random sample of graduate students who took the final exam in Ag 400 over the last 5 years. The data set is called:

FINAL.sav

The file contains the following variable

Y = FINAL = score on the final exam

X1 = COURSES = Total number of previous stat courses

X2 = HOURS = number of hours the student reported that he/she had studied for the final

Indicate whether each of the following statements is true (T) or false (F) in terms of this output.

T a. The bivariate correlation between HOURS and FINAL is positive and statistically significant at the .001 level using a one-tailed test.

F b. Overall, approximately 17.7% of the variation in final exam scores is associated with the variation in COURSES.

F c. Using this output, if you reject the following Null Hypothesis in favor of the following Alternative Hypothesis, $\alpha = .042$; $\beta = 0.00$

H_0 : There is no bivariate linear relationship between the number of previous courses in statistics and score on the final.

H_A : There is a bivariate linear relationship between the number of previous statistics courses and score on the final exam. .

2-tailed
 H_A

T d. The interrelationship between HOURS and COURSES in the sample can be described as follows:

The more previous courses a student had in statistics, the fewer hours he/she was likely to study for the final.

F e. Degrees of freedom for testing the statistical significance of each of the correlations on Page 1 of the output equals 97.

- F f. Overall, in the sample data, as number of hours of study increased by 1 hour, scores on the final exam increased by .354 points.
- T g. The bivariate correlation between COURSES and FINAL is suppressed somewhat when HOURS is not controlled. When the effect of HOURS is adjusted for, the strength of the linear relationship between COURSES and FINAL increases.
- T h. The Multiple Correlation Coefficient for the relationship of HOURS and COURSES to FINAL is statistically significant at the .001 level.
- T i. Both the Multiple Correlation Coefficient and the Multiple Coefficient of Determination range between zero and 1.00 for any data set. The Multiple Coefficient of Determination is never greater than the Multiple Correlation Coefficient.
- F j. The zero-order correlation of COURSES to FINAL is positive and statistically significant at the .05 level (one-tailed test). However, when HOURS studied is controlled, COURSES is not significantly correlated with FINAL exam score.
- T k. Multiple correlation coefficients are never negative, although they can equal 0.00.
- T l. Based on this analysis, if a student studied for one hour and had no previous stat courses, he/she would score only about 46 points on the test.
- F m. $b_{y1.2}$ measures the number of points change on the final exam score for every 1 hour that a student prev stat course studied. Here $b_{y1.2}$ equals +3.876. ~~studied. Here $b_{y1.2}$ equals +3.876.~~
- F n. The zero-order correlation of .177 on Page 2 of the output is statistically significant ($p=.023$).
- T o. The output shows that, as the number of previous statistics courses increases by one course, score on the final exam increases by about 5.8 points when the effect of HOURS is adjusted for.
- F p. In this analysis the partial correlation coefficients are higher than their corresponding zero-order correlations. Partial correlations are never lower than their corresponding zero-order corrections.

T q. The following model accounts for about 17% of the variation in FINAL scores.

$$\text{FINAL} = 42.753 + (3.876 * \text{HOURS}) + (5.812 * \text{COURSES})$$

T r. If COURSES is deleted from the regression equation on Page 2, the variation in FINAL accounted for by the resulting equation would equal about 12.5%.

F s. Partial correlations coefficients are never negative, although they can equal 0.00.

F t. $b_{y1.2}$ is larger than $b_{y2.1}$ so we can conclude that X1 (COURSES) has a greater impact on FINAL than does X2 (HOURS).

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
FINAL	70.71	14.87	97
HOURS	5.32	1.45	97
COURSES	1.26	.56	97

Correlations

		FINAL	HOURS	COURSES
Pearson Correlation	FINAL	1.000	.354	.177
	HOURS	.354	1.000	-.111
	COURSES	.177	-.111	1.000
Sig. (1-tailed)	FINAL	.	.000	.042
	HOURS	.000	.	.140
	COURSES	.042	.140	.
N	FINAL	97	97	97
	HOURS	97	97	97
	COURSES	97	97	97

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	COURSES, HOURS ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: FINAL

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.416 ^a	.173	.155	13.67

a. Predictors: (Constant), COURSES, HOURS

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3668.717	2	1834.358	9.814	.000 ^a
	Residual	17569.201	94	186.906		
	Total	21237.918	96			

a. Predictors: (Constant), COURSES, HOURS

b. Dependent Variable: FINAL

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	42.753	6.482		6.596	.000			
	HOURS	3.876	.966	.379	4.010	.000	.354	.382	.376
	COURSES	5.812	2.510	.219	2.316	.023	.177	.232	.217

a. Dependent Variable: FINAL