**Questions**

**Q1.** Suppose you are interested in estimating the effect of hours spent in an SAT preparation course (*hours*) on the total SAT score (*sat*). The population is all college-bound high school seniors graduating in a particular year.

1. Suppose you are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of *hours* on *sat*.
2. Consider the more realistic case where students choose the number of hours they spend on SAT preparation and whether or not to take an SAT-prep course. In this case, you can only randomly select *sat* and *hours* from the population of students. The population model is:

𝑠𝑎𝑡= 𝛽0+ 𝛽1ℎ𝑜𝑢𝑟𝑠=𝑢

where, because of the intercept, E(u) =0. List one factor that might explain *sat* that is likely included in *u*. Do you think this factor is correlated with *hours*? If this factor were correlated with *hours,* why might that be a problem?

c) In the equation from part (ii), what should the sign of 𝛽1 be if the preparation course is effective?

**Q2**. Use the results from L1 to calculate the following. Even though the regression output tells you the answers, show how they are calculated assuming you had the other output but you did not have the regression results. Please show what formula you used and the steps you used to get to the final answer. You can always use the output to see if you did it right!

a) The coefficient for beauty, using the formula. (Hint: What is the numerator? What is the denominator?)

b) Write down the null and alternative hypotheses to test if the coefficient on beauty is equal to zero.

c) Calculate the t-statistic to test the null hypothesis from part b). Note: In this part it is OK to use the standard error calculated by the regression output rather than having to calculate it.

d) To test the hypothesis in part b) with a confidence level of 95% level, then the critical t-value is 1.96. Would you reject or accept (fail to reject) the null hypothesis? Explain.

e) Calculate the R2 for this model. What does it represent?

f) Compare your results to a), c) and e) to the output from your regression in L1. Are they the same?

**Lab Problems**

**L1.**

--------------------------------------------------------------------------------

name: <unnamed>

log: C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab3\lab3\teaching.log

log type: text

opened on: 14 Feb 2013, 11:13:18

. /\*a. Create a graph that includes both a scatter plot of average course evalua

> tion (course\_eval) vs. instructor’s beauty

> (beauty) and the fitted line. Save this scatter plot as scattereval\_beauty.em

> f so it can be printed and turned in with

> your assignment.\*/

>

> twoway (lfit course\_eval beauty)(scatter course\_eval beauty);

. graph export scattereval\_beauty.emf, replace;

(file C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab3\lab3\scattereval\_beauty.e

> mf written in Enhanced Metafile format)

. graph export scattereval\_beauty.png, replace;

(file scattereval\_beauty.png written in PNG format)

. //Note: this is for my use: .emf is difficult to work with on Mac/Linux

>

> /\*b. Using the scatter plot and the fitted line from a), does it appear that c

> ourse evaluations and beauty are related?

> Explain why.

> //TODO Interpretation

>

> \*/

>

>

>

> /\*c. Use the summarize command to look at the description of the variables in

> this dataset.\*/

>

> summarize;

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

minority | 463 .1382289 .3455134 0 1

age | 463 48.36501 9.802742 29 73

female | 463 .4211663 .4942802 0 1

onecredit | 463 .0583153 .2345922 0 1

beauty | 463 4.75e-08 .7886477 -1.450494 1.970023

-------------+--------------------------------------------------------

course\_eval | 463 3.998272 .5548656 2.1 5

intro | 463 .3390929 .4739135 0 1

nnenglish | 463 .0604752 .2386229 0 1

. /\*d. Use the covariance option for the correlate command to get the covariance

> between course\_eval and beauty.\*/

>

> correlate course\_eval beauty, covariance;

(obs=463)

| course~l beauty

-------------+------------------

course\_eval | .307876

beauty | .082722 .621965

. /\*e. Use the regression command to estimate the following linear regression mo

> del

> CourseEvaluations\_i = Beta\_zero \* Beauty + u\_i\*/

>

> regress course\_eval beauty;

Source | SS df MS Number of obs = 463

-------------+------------------------------ F( 1, 461) = 17.08

Model | 5.08300731 1 5.08300731 Prob > F = 0.0000

Residual | 137.155613 461 .297517598 R-squared = 0.0357

-------------+------------------------------ Adj R-squared = 0.0336

Total | 142.23862 462 .307875801 Root MSE = .54545

------------------------------------------------------------------------------

course\_eval | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

beauty | .1330014 .0321775 4.13 0.000 .0697687 .1962342

\_cons | 3.998272 .0253493 157.73 0.000 3.948458 4.048087

------------------------------------------------------------------------------

. /\*f. What is the interpretation of the intercept term0ßin this regression mode

> l.

> //TODO Interpretation\*/

>

>

> /\*g. Using the regression output, what is the value of the estimated intercept

> 0ˆß?

> //TODO Interpretation\*/

>

>

> /\*h. What is the interpretation of the slope term1ßin this regression model.

> //TODO Interpretation\*/

>

>

> /\*i. Using the regression output, what is the value of the estimated slope 1ˆß

> ?

> Is it significantly different from zero? How do you know?

> //TODO Interpretaion\*/

>

>

> /\*j. Suppose Professor Smith is of average beauty, while Professor Miller is o

> ne standard deviation above the average beauty.

> Using the regression output and the summary statistics, predict Professor Smi

> th’s and Professor Miller’s course evaluations.

> //TODO Interpretaion\*/

>

> /\*k. Does this model explain a large fraction of the variance in course evalua

> tions

> //TODO Interpretaion\*/

>

> log close;

name: <unnamed>

log: C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab3\lab3\teaching.log

log type: text

closed on: 14 Feb 2013, 11:13:19

--------------------------------------------------------------------------------

/\* A. Shawn Bandy

Lab #3

February 12th, 2013

\*/

/\* close previous run do-files \*/

cap log close

set more 1

clear

#delimit ;

cd "C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab3\lab3";

use "TeachingRatings";

log using teaching.log , replace;

/\*a. Create a graph that includes both a scatter plot of average course evaluation (course\_eval) vs. instructor’s beauty

(beauty) and the fitted line. Save this scatter plot as scattereval\_beauty.emf so it can be printed and turned in with

your assignment.\*/

twoway (lfit course\_eval beauty)(scatter course\_eval beauty);

graph export scattereval\_beauty.emf, replace;

graph export scattereval\_beauty.png, replace; //Note: this is for my use: .emf is difficult to work with on Mac/Linux

/\*b. Using the scatter plot and the fitted line from a), does it appear that course evaluations and beauty are related?

Explain why.

//TODO Interpretation

\*/

/\*c. Use the summarize command to look at the description of the variables in this dataset.\*/

summarize;

/\*d. Use the covariance option for the correlate command to get the covariance between course\_eval and beauty.\*/

correlate course\_eval beauty, covariance;

/\*e. Use the regression command to estimate the following linear regression model

CourseEvaluations\_i = Beta\_zero \* Beauty + u\_i\*/

regress course\_eval beauty;

/\*f. What is the interpretation of the intercept term0ßin this regression model.

//TODO Interpretation\*/

/\*g. Using the regression output, what is the value of the estimated intercept 0ˆß?

//TODO Interpretation\*/

/\*h. What is the interpretation of the slope term1ßin this regression model.

//TODO Interpretation\*/

/\*i. Using the regression output, what is the value of the estimated slope 1ˆß?

Is it significantly different from zero? How do you know?

//TODO Interpretaion\*/

/\*j. Suppose Professor Smith is of average beauty, while Professor Miller is one standard deviation above the average beauty.

Using the regression output and the summary statistics, predict Professor Smith’s and Professor Miller’s course evaluations.

//TODO Interpretaion\*/

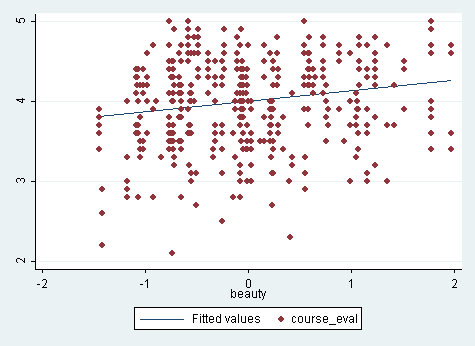
/\*k. Does this model explain a large fraction of the variance in course evaluations

//TODO Interpretaion\*/

log close;

*b. Using the scatter plot and the fitted line from a), does it appear that course evaluations and beauty are related?*

*Explain why.*



Although I cannot say that there is no relationship at all, I do not see any linear – or any other – relation in this scatter plot. If there were a linear relation, I would expect to see the *course\_eval* plots near the fitted line. In this case the plots are truly scattered over the area of the graph.

*f. What is the interpretation of the intercept term ß0 in this regression model.*

ß0 is the expected value for course\_eval when beauty = 0.

*g. Using the regression output, what is the value of the estimated intercept β0?*

According to the regression output, ß0 equals 3.998272.

*h. What is the interpretation of the slope term ß1 in this regression model?*

For each unit change in *beauty*, *course\_eval* changes by 0.1330014 units.

*i. Using the regression output, what is the value of the estimated slope β1-hat ? Is it significantly different from zero? How do you know?*

The value of the estimated slope of *β1-hat* is 0.1330014 and is significantly different than zero at the 95% confidence level because 0 is between 3.948458 and 4.048087.

*j. Suppose Professor Smith is of average beauty, while Professor Miller is one standard deviation above the average beauty. Using the regression output and the summary statistics, predict Professor Smith’s and Professor Miller’s course evaluations.*

*k. Does this model explain a large fraction of the variance in course evaluations?*