

# Assignment 1

Michael Cai

February 10, 2016

## 1 Problem One

a)  $-X$  and  $-Y$  are positively correlated. **True.**

Recall that  $\beta_1$  is the change in  $Y_i$  given a one unit change in  $X_i$ . Thus if there is a negative one unit change in  $X_i$  then there will also be a negative change in  $Y_i$ .

b)  $-X$  and  $Y$  are positively correlated. **False.**

The logic of part a) proves that b) is a false statement. If there is a positive one unit change in  $-X$ , there will also be a negative change in  $Y$ , thus the two variables in this question are negatively correlated.

c) The covariance of  $-X$  and  $-Y$  is positive. **True.**

The covariance of two random variables is the measure of how two random variables change together. As is stated in part a), if there is a positive correlation between  $-X$  and  $-Y$  and thus these variables must positively change with one another. This is also further confirmed by the definition of  $\rho(X, Y) = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$ .

d) The covariance of  $-X$  and  $Y$  is positive. **False.**

False for the same reasoning as in b) and c). Correlation and covariance are really restatements of one another just with different units.

e/f) The slope coefficient obtained by regressing  $Y$  on  $X$  would be positive, and  $X$  on  $Y$  would be positive. **Both true.**

Because  $X$  and  $Y$  are positively correlated random variables when they are regressed upon another they produce the same result - a positive slope coefficient. This is because for each unit change in both  $X$  and  $Y$  there is a corresponding increase in  $Y$  and  $X$  respectively. Furthermore, if these random variables were economic indicators, for instance, interest rates and savings rates, then there would be some endogenous selection as well thus further confirming the conclusion.

## 2 Problem Two

a) Write the population regression function.

$$Y_i = \beta_0 + \beta_1 X_i$$

b) What will be the value of  $R^2$ ?

$R^2$  is the fraction of the sample variance of  $Y_i$  that is explained by  $X_i$ . But since we are working with variables that have a binary and completely determined relation, i.e. If there are a given number of correct answers ( $Y$ ) then the number of incorrect answers ( $X$ ) is always the same and thus completely determined. Therefore  $R^2 = 1$ .

c) What will be the value of the standard error of the regression?

Also because the value pairs of  $(X, Y)$  are completely determined by each other then there is no standard error as all of the observations will lie on the population regression function.

## 3 Problem Three

a) Suppose you know  $\beta_0 = 0$ . Derive a formula for the least-squares estimator of  $\beta_1$ .

Find the least-squares estimator of  $\beta_1$  by minimizing  $\Sigma(Y_i - (\beta_0 + \beta_1 X_i))^2$  with respect to  $\beta_0$  and  $\beta_1$ .

We already know that  $\beta_0 = 0$  so we just need to solve the partials for  $\frac{\partial Y}{\partial \beta_0}$  and  $\frac{\partial Y}{\partial \beta_1}$  set equal to 0.

$$\frac{\partial Y}{\partial \beta_0} = -2\Sigma(Y_i - \beta_0 - \beta_1 X_i) = 0$$

We divide by  $-2n$  to get:

$$\frac{\partial Y}{\partial \beta_0} = \frac{1}{n}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$\frac{\partial Y}{\partial \beta_0} = \frac{1}{n}\Sigma Y_i - \frac{1}{n}\Sigma \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\Sigma X_i$$

$$= \bar{y}_n - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_n = 0$$

$$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$$

Now to solve for  $\hat{\beta}_1$

$$\frac{\partial Y}{\partial \beta_1} = -2\Sigma X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$$

Divide by  $-2n$  to get

$$= \frac{1}{n}\Sigma(Y_i - \beta_0 - \beta_1 X_i)X_i = 0$$

$$= \frac{1}{n}\Sigma Y_i X_i - \hat{\beta}_0 \frac{1}{n}\Sigma X_i - \hat{\beta}_1 \frac{1}{n}\Sigma X_i^2$$

$$= S_{xy} + \bar{y}_n \bar{x}_n - \hat{\beta}_0 \bar{x}_n - \hat{\beta}_1 (S_x^2 + \bar{x}_n^2)$$

Plugging in what we got for  $\hat{\beta}_0$  we get:

$$= S_{xy} + \bar{y}_n \bar{x}_n - (\bar{y}_n - \hat{\beta}_1 \bar{x}_n) \bar{x}_n - \hat{\beta}_1 (S_x^2 + \bar{x}_n^2)$$

$$\text{This simplifies to: } S_{xy} - \hat{\beta}_1 S_x^2 = 0$$

Therefore,  $\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$

**\*\* So I feel that although this is the derivation of the standard formula, but because  $\beta_0 = 0$  there is a different derivation.** Let's try this:

Because  $\beta_0 = 0$  then we can just derive the partial for  $\beta_1$  immediately.

$$\frac{\partial Y}{\partial \beta_1} = -2X_i \Sigma(Y_i - \beta_1 X_i) = 0$$

Divide by  $-2n$  to get:

$$= \frac{1}{n} \Sigma(Y_i - \beta_1 X_i) X_i = 0$$

$$= \frac{1}{n} \Sigma X_i Y_i - \hat{\beta}_1 \frac{1}{n} \Sigma X_i^2$$

$$= S_{xy} + \bar{y}_n \bar{x}_n - \hat{\beta}_1 (S_x^2 + \bar{x}_n^2)$$

$$\hat{\beta}_1 = \frac{S_{xy} + \bar{y}_n \bar{x}_n}{S_x^2 + \bar{x}_n^2}$$

\*This is what I think is the correct answer, but I can't figure out how to prove that it is consistent because without the term cancellation from the general derivation, there is not such a direct way to justify probabilistic convergence.

b) Show your estimator from part a) is consistent.

We prove consistency by showing that the Law of Large Numbers is valid.

In other words, to show that  $\hat{\beta}_1 \rightarrow_p \beta_1$ .

By the LLN,  $\bar{x}_n \rightarrow_p \mu_x$  and  $\bar{y}_n \rightarrow_p \mu_y$ . Therefore,  $\hat{\beta}_1 \rightarrow_p \frac{\mu_y}{\mu_x}$  which equals  $\beta_1$  or in words, the estimator for the slope of the regression line converges with probability to the slope of the population regression function.

## 4 Problem Four

a) Write the population regression function.

$$TS_i = \beta_0 + \beta_1 EXPENDITURE_i$$

b) Interpret  $\beta_0$  and  $\beta_1$

$\beta_0$  is the average math and reading test score for school  $i$  where the expenditure is equal to zero dollars per student.

$\beta_1$  is the change in the average math and reading test score for school  $i$  when the expenditure per student (in dollars) is increased by one unit.

c) Estimate the above regression model and write the sample regression line.

R-Code:

```
CAS$exps ← CAS$expenditure/CAS$students
```

```
fm ← lm(ts ~exps, data=CAS)
```

$$\widehat{TS}_i = \hat{\beta}_0 + \hat{\beta}_1 EXPENDITURE_i$$

$$\widehat{TS}_i = 652.45510 + 0.15463 EXPENDITURE_i$$

d) Interpret  $\hat{\beta}_0$  and  $\hat{\beta}_1$

$\hat{\beta}_0 = 652.45510$  means that if the expenditure per student (in dollars) was zero then the average math and reading test score for school  $i$  would be 652.45510.

$\hat{\beta}_1 = 0.15463$  means that for every one unit increase in expenditure per student of school  $i$  then there will be a 0.15463 increase in average math and reading test score.

e) Is the sign of  $\hat{\beta}_1$  what you would expect? Explain.

Yes, I expected that  $\hat{\beta}_1$  would be positive because you would expect that if you spend more money per student then that student should probably perform better in school.

f) For a \$1000 increase in per-student expenditure, you can expect an estimated average change of  $1000 * 0.15463 = 154.63$  point change in test score.

## 5 Problem Five

a) Generate the variable  $CS$  from the variables **computer** and **students**.

$CAS\$cs \leftarrow CAS\$computer / CAS\$students$

b) Write the population regression function.

$$TS_i = \beta_0 + \beta_1 CS_i$$

c) Interpret  $\beta_0$  and  $\beta_1$

$\beta_0$  is the average math and reading test score for school  $i$  when the number of computers per student is equal to zero.

$\beta_1$  is the increase in average math and reading test score for school  $i$  for every one unit increase in computers per student.

d) Estimate the above regression model and write the sample regression line.

$$TS_i = \hat{\beta}_0 + \hat{\beta}_1 CS_i$$

$$TS_i = 643.36 + 79.41 CS_i$$

e) Interpret  $\hat{\beta}_0$  and  $\hat{\beta}_1$

$\hat{\beta}_0 = 643.36$  is the average math and reading test score for school  $i$  when the number of computers per student is equal to zero.

$\hat{\beta}_1 = 79.41$  is the increase in average math and reading test score for school  $i$  for every one unit increase in the number of computers per student.

f) Is the sign of  $\hat{\beta}_1$  what you would expect? Explain.

Yes, because with an increasing number of computers per student or more generally with increasing access to computers, students will have more learning opportunities and resources at hand, which I assume would positively correlate with an increase in testing ability.