Assignment

The homework assignment was to read chapter 2 and hand in answers to the following problems at the end of the chapter: 2.1 – 2.5 and C2.1 – C2.3.

2.1

Let “kids” denote the number of children ever born to a woman, and let “educ” denote years of education for the woman. A simple model relating fertility to years of education is

\[ \text{kids} = \beta_0 + \beta_1 \text{educ} + u, \]

where \( u \) is the unobserved error.

(i)

What kinds of factors are contained in \( u \)? Are these likely to be correlated with level of education?

Education of the woman’s parents, number of siblings the woman has, her income, her marital status, her job, etc., are all likely to be contained in \( u \). That is, each of these factors could be thought of as “explanatory.” Each of them (and many others that you could come up with) help determine the number of children a woman has. Many of the factors that I just listed are likely to be correlated with a woman’s level of education. Education of the woman’s parents: certainly correlated. Number of siblings the woman has: arguable whether or not this is correlated with the woman’s level of education. Marital status: correlated. Job: correlated.

(ii)

Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.
Because the omitted explanatory variables (such as those above) are embodied in \( u \), and because these variables are also correlated with education, the regression will not reveal a causal relationship. The assumption \( E(u|x) = 0 \) is violated.

2.2

*In the simple linear regression model* \( y = \beta_0 + \beta_1 x + u \), *suppose that* \( E(u) \neq 0 \). *Leting* \( \alpha_0 = E(u) \), *show that the model can always be rewritten with the sample slope, but a new intercept and error, where the new error has a zero expected value.*

Suppose that \( E(u) = \alpha_0 \neq 0 \). We could write up a model with a new intercept. Instead of calling it \( \beta_0 \) like we did before, we’ll just call it \( \beta_0 + \alpha_0 \). Then we introduce a new “error” (or “unobservable”) term, call it \( \epsilon \), where \( \epsilon = u - \alpha_0 \). But you don’t have to know this to start with. You get it if you just subtract \( \alpha_0 \) from both sides of the equation (which you can always do — if you do the same thing to both sides of an equation, you are not disturbing the underlying relationship). You know you want to subtract \( \alpha_0 \) from both sides of the equation because you know you want the expectation of the unobservable to be 0. Subtracting \( \alpha_0 \) gets you that 0.

\[
E[y - \alpha_0] = E[(\beta_0 + \beta_1 x + u - \alpha_0) \\
= (\beta_0 + \alpha_0) + \beta_1 \bar{x} + E[u - \alpha_0] \\
= (\beta_0 + \alpha_0) + \beta_1 \bar{x} + E[u] - E[\alpha_0] \\
= (\beta_0 + \alpha_0) + \beta_1 \bar{x} + \alpha_0 - \alpha_0 \\
= (\beta_0 + \alpha_0) + \beta_1 \bar{x}.
\]

This is all exactly equivalent to writing a new model with an intercept = \( \beta_0 + \alpha_0 \) and a new unobservable \( \epsilon = u - \alpha_0 \).

2.3

*The following table contains the ACT scores and the GPA (grade point average) for eight college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.*

<table>
<thead>
<tr>
<th>Student</th>
<th>GPA</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>3.4</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>3.5</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>3.6</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>3.0</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>2.7</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>3.7</td>
<td>30</td>
</tr>
</tbody>
</table>
Estimate the relationship between GPA and ACT using OLS; that is, obtain the intercept and slope estimates in the equation

\[ \hat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT \]

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the GPA predicted to be if the ACT score is increased by five points?

The easiest way to get the answer to this question would be to enter the data by hand into Stata (or Excel). Then run a simple OLS regression. Since there are only 8 data points, this would be pretty quick. Or, again because there aren’t many data points, you could just compute the formula for the OLS coefficients using the formula on page 29 and a calculator. I used Stata (the easiest way to enter this number of data points is by using the edit command and just filling in the spreadsheet) and found the following coefficients (rounded to two decimal places):

\[ \hat{\beta}_0 = 0.57, \hat{\beta}_1 = 0.1. \]

The direction of the relationship is positive, as expected. Students with a higher ACT score have a higher GPA. Whether this has to do with intelligence, test-taking ability, or whatever, I am not sure. But the direction of the relationship is certainly expected. The literal interpretation of the intercept is “the GPA if the ACT score was 0.” Is that even possible? Can you score a 0 on your ACT? Or do you get 10 points for writing your name? I don’t know. But either way, my answer to the question is probably “no, the intercept does not have a useful interpretation.” All of the ACT scores in our data are well above 0, indicating to me that the notion of a 0 ACT score is well outside our sample, and thus well outside what I am comfortable predicting with this model.

If the ACT score increases by 5 points, I expect to observe a 0.5 point increase in GPA (= 0.1 * 5).

(ii)

Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

Again, this is very easy to do by hand, and even easier to do in stata. Here’s the code which will verify the output.

```stata
reg gpa act
predict gpaHat
gen resid = gpa - gpaHat
gen totalResid = sum(resid)
browse
```
Look at the data. The variable `totalResid` is the running total of the residuals. By the time you get to observation 8, the residuals sum to 2.38e-07, or 0.000000238. That’s close enough to 0 for me.

(iii)

*What is the predicted value of GPA when ACT = 20?*

The predicted value of GPA when ACT = 20 is

$$0.57 + 0.1 \times 20 = 2.57.$$  

Even though we don’t have an observation with ACT = 20, we can predict what would happen if we did, using our model.

(iv)

*How much of the variation in GPA for these eight students is explained by ACT? Explain.*

About 58% of the variation in GPA is explained by ACT. We see this in the R-squared statistic, which represents the proportion of the total variation in GPA that is explained by all the explanatory variables. Since we only have one explanatory variable, R-squared gives us exactly what the question is looking for. (This would not be true if there were multiple explanatory variables).

2.4

*The data set BWGHT.RAW contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (bwght), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (cigs). The following simple regression was estimated using data on \(n = 1,388\) births:*

\[
\hat{bwght} = 119.77 - 0.514cigs
\]

(i)

*What is the predicted birth weight when cigs = 0? What about when cigs = 20 (one pack per day)? Comment on the difference.*

The predicted birth weight when cigs = 0 is 119.77 ounces (7.5 lbs). When cigs = 20, the predicted birth weight is equal to 119.77 - 0.514 * 20 = 109.5 ounces (6.8 lbs). It seems clear to me from this equation that (shock!) smoking is nasty bad: it makes your breath stink, your teeth yellow, makes it hard to run a mile, and it makes your kid small (OK, only the last one is justified by the model, but you know what I mean).
(ii)

*Does this simple regression necessarily capture a causal relationship between the child’s birth weight and the mother’s smoking habits? Explain.*

The regression captures a causal relationship between the child’s birth weight and the mother’s smoking habits *unless* the mother’s smoking habits are positively related to other unobserved factors which also negatively influence birth weight. Off the top of my head, I would guess that income and education (which both influence nutrition and the amount of prenatal care that a woman gets) could be such variables. So if we wanted to be sure that smoking reduced birth weight, we’d have to make sure that these variables were accounted for. My guess is that we would still find a negative relationship between smoking and birth weight, but that the effect would be attenuated (i.e. the coefficient would be smaller in absolute value).

(iii)

*To predict a birth weight of 125 ounces, what would cigs have to be? Comment.*

Uh-oh. Gotta drop some algebra to get this one done. 125 is the $y$-variable, in this case.

\[ 125 = 119.77 - 0.514 \times cigs. \]

Solve for cigs:

\[ cigs = \frac{119.77 - 125}{0.514} = -10.2. \]

Impossible! You can’t smoke negative cigarettes (if anybody figures out how to do this please let me know. I’d like to begin negative smoking immediately).

(iv)

*The proportion of women in the sample who do not smoke while pregnant is about 0.85. Does this help reconcile your finding from part (iii)?*

I don’t really like this question. It’s not as illuminating as you might think. As I mentioned in class, this *sort of* helps reconcile the findings. If most of our sample is pegged at $cigs = 0$, then I expect that the intercept is going to be pretty close to the mean birth weight for the part of the sample that doesn’t smoke (think of the interpretation of the intercept and you’ll see why). So even though some children a certainly born with birth weights above 120, our model doesn’t say much about them. All of the stuff that causes children to be born heavier than about 120 ounces is unobservable in our model.
2.5

In the linear consumption function

\[ \hat{\text{cons}} = \hat{\beta}_0 + \hat{\beta}_1 \text{inc}, \]

the (estimated) marginal propensity to consume (MPC) out of income is simply the slope, \( \hat{\beta}_1 \), while the average propensity to consume (APC) is \( \hat{\text{cons}}_{\text{inc}} = \hat{\beta}_0 \text{inc} + \hat{\beta}_1 \). Using observations for 100 families on annual income and consumption (both measured in dollars), the following equation is obtained:

\[ \hat{\text{cons}} = -124.84 + 0.853\text{inc} \]
\[ n = 100, R^2 = 0.692 \]

(i)

Interpret the intercept in this equation, and comment on its sign and magnitude.

The marginal propensity to consume when income is zero is -124.84 — when you aren’t making any money, you de-consume (you sell stuff). Its sign isn’t surprising; even though interpreting a negative number literally is dangerous, in this case. We don’t have the data in front of us, so I can’t be sure, but I’d guess there are no negative entries in consumption. So this is probably another case where the intercept doesn’t have much of an interpretation. If it really is possible (in the data) to observe negative consumption (selling your stuff) then the interpretation I gave is meaningful. But I doubt it. (Note: if you just saved and did nothing, without selling stuff, your MPC would be 0 — when you get money, you do nothing).

(ii)

What is the predicted consumption when family income is $30,000?

Consumption when income is $30,000 is given by

\[ -124.84 + 0.853 \times 30000 = 25465.16, \]

i.e. $25,465.16.

(iii)

With inc on the x-axis, draw a graph of the estimated MPC and APC.

The first thing Wooldridge asks you to do is very silly. Graphing MPC against inc is just graphing a constant number (0.853) against income. That is a flat line at 0.853.
Not worth doing. If you graphed consumption against income, that would be more interesting (but only slightly). You would have a line that passed through the origin at -$124 and increased with a slope of $0.85 per dollar of income. See below.

Graphing the APC is just slightly more interesting. See below.

The average propensity to consume increases steeply when income is low, then levels off significantly at around $2,000 at about 0.8. By the time a family is making around $100,000, their average propensity to consume is still only about 0.85. Graphing the two together is sort of cool.
C2.1

The data in 401K.RAW are a subset of data analyzed by Papke (1995) to study the relationship between participation in a 401(K) pension plan and the generosity of the plan. The variable “prate” is the percentage of eligible workers with an active account; this is the variable we would like to explain. The measure of generosity is the plan match rate, “mrate.” This variable gives the average amount the firm contributes to each worker’s plan for each $1 contribution by the worker. For example, if “mrate” = 0.50, then a $1 contribution by the worker is matched by a 50 cent contribution by the firm.

(i) Find the average participation rate and the average match rate in the sample of plans.

Use Stata code

```
sum prate mrate
```

to find that the averages are 87.4% participation and 0.73 match-rate, respectively.
Now, estimate the simple regression equation

\[ \hat{prate} = \hat{\beta}_0 + \hat{\beta}_1 \text{mrate} \]

and report the results along with the sample size and R-squared.

The following Stata code produces the results below.

```stata
reg prate mrate
\beta_0 = 83.07, \beta_1 = 5.86, n = 1,534, R-squared = 0.07.
```

(iii)

Interpret the intercept in your equation. Interpret the coefficient on “mrate.”

The intercept = 83.07. If the match-rate were 0, i.e. if employers contributed nothing to employees’ 401k plans, 83% of employees would still participate in a 401k. For every dollar increase in the match-rate, employee participation increases by almost 6%.

(iv)

Find the predicted “prate” when “mrate” = 3.5. Is this a reasonable prediction? Explain what is happening here.

The predicted \( prate \) when \( mrate \) is 3.5 is given by

\[ 83.07 + 5.86 \times 3.5 = 103.58. \]

That is, 103% of employees are predicted to participate. Of course this is unreasonable. See the graphic above to get a picture of what is going on. Basically, we are using a linear
model (which predicts out into infinity in both the positive and negative directions) to fit data that is constrained to be less than or equal to 100. That is, the linear model is capable of predicting stuff that is incapable of happening. If we want to use the linear model for this type of problem, we have to live with this tradeoff. We should be very skeptical of the model to begin with, and we should absolutely not interpret the model’s out-of-possibility predictions literally.

(v)

How much of the variation in “prate” is explained by “mrate?” Is this a lot in your opinion?

Very little. About 7% according to the R-squared statistic. Most of the participation decision appears to be determined by factors other than mrate. (Maybe new employees are enrolled in company 401k plans by default, e.g.)

(C2.2)

The data set in CEOSAL2.RAW contains information on chief executive officers for U.S. corporations. The variable “salary” is annual compensation, in thousands of dollars, and “ceoten” is prior number of years as company CEO.

(i)

Find the average salary and the average tenure in the sample.

Use Stata code

```
sum salary ceoten
```

to find that the average salary of a CEO in our sample is $865,860, and the average tenure is 7.95 years.

(ii)

How many CEOs are in their first year as CEO (that is, “ceoten” = 0)? What is the longest tenure as a CEO?

There are lots of ways to find this information. There are a total of 177 ceos in the sample. Since there are 177 observations of ceoten (no missing values in that particular variable), we can just count the number of observations that meet the criteria. A command such as

```
sum ceoten if ceoten == 0
```

will do the trick. You will see in the output that the number of observations fitting the criteria is 5. Alternatively, you could create a variable to do the calculation directly. For instance, you could create an indicator variable using the code

```
gen newGuy = 0
```
replace newGuy = 1 if ceoten == 0
then summarize this indicator variable with the code
sum newGuy.
You could do two things at this point to find the answer. You could multiply the number
of observations (177) by the mean (.0282486) to find the answer (5.0000022). Or you
could use the direct way.

This is a useful Stata tip: Every estimation command (even the very simple command
sum) saves some calculations in the background without printing them to the screen. To
see what results are saved and available to you (though not printed to the screen), look
at the bottom of the help file for the estimation command. Type help summarize to
see the help file for summarize. Scroll to the bottom and look at the material below the
“Saved results” heading. One of the things you will see is that the sum of the variable
being summarized (in this case, the newGuy variable) is stored in a scalar referred to in
Stata as r(sum). A scalar is simply a single number. A constant. A regular old number.
In Stata, scalars are referred to as such to distinguish them from data. Most stuff in
Stata is data, i.e. it is a whole column of information. If you want to refer to a single
entity (like a number equal to the sum of the newGuy indicator variable) then you preface
it with the scalar command. Run the following commands to see what happens.
sum newGuy
scalar count = r(sum)
scalar list count
You should see that the result (5) is printed to the screen. You have generated a scalar
that is equal to the sum of a variable which takes on the value 1 when a ceo is tenured
for zero years.

This is just one way to do it. You could successfully do it a number of different ways.

To find the longest tenured ceo, simply summarize the ceoten variable and look at
the maximum value. The longest tenured ceo has been on the job for 37 years.

(iii)

Estimate the simple regression model

$log(salary) = \beta_0 + \beta_1 ceoten + u,

and report your results in the usual form. What is the (approximate) predicted percent-
age increase in salary given one more year as CEO?

gen logSalary = log(salary)
reg logSalary ceoten
We see that for every year increase in ceo tenure, salary increases by about 1% (the
coefficient on ceoten is .0097236, which we multiply by 100 to obtain the percentage
increase).
(C2.3)

Use the data in SLEEP75.RAW from Biddle and Hamermesh (1990) to study whether there is a tradeoff between the time spent sleeping per week and the time spent in paid work. We could use either variable as the dependent variable. For concreteness, estimate the model

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + u,$$

where “sleep” is minutes spent sleeping at night per week and “totwork” is total minutes worked during the week.

(i)

Report your results in equation form along with the number of observations and $R^2$. What does the intercept in this equation mean?

My results “in equation form” are:

$$\hat{\text{sleep}} = 3586.38 - 0.15 \times \text{totwrk}.$$ 

The number of observations used to estimate the model is 706, and the R-squared is 0.1 (rounded to the 100ths place). The intercept of 3,586 is the number of minutes of night-sleep (as opposed to nap-sleep) per week that an unemployed person gets.

(ii)

If “totwrk” increases by 2 hours, by how much is “sleep” estimated to fall? Do you find this to be a large effect?

The variable totwrk is measured in minutes, so a change of two hours would be equal to an increase in totwrk of 120. An increase of totwrk by 120 would generate a decrease in sleep of 18 minutes per week ($120 \times (-0.15) = -18$). Doesn’t seem like much. Work an extra 2 hours per week and sleep 18 minutes less. Would you do it? Depends on your wage, I suppose.