Econ 21020 - Problem Set 4

Problem 1

In class we said that, for an n-dimensional column vector X,

$$Var(X) = E[(X - E[X])(X - E[X])']$$

is an $n \times n$ dimensional matrix where the element in the ith row and jth column is $Cov(X_i, X_j)$. Show/explain why this is the case.

SOLUTION: Each object in parentheses:

$$(X - E[X]) = \begin{bmatrix} X_1 - E[X_1] \\ . \\ . \\ . \\ X_n - E[X_n] \end{bmatrix}$$

is an n-dimensional column vector. Multiplying $(X - E[X])$ by $(X - E[X])'$ will then result in an $n \times n$ dimensional matrix, for which the element in the ith row and jth column is equal to the ith component of $(X - E[X])$ multiplied by the jth component of $(X - E[X])$. When $i = j$, which occurs along the main diagonal of the matrix, this is equal to:

$$(X_i - E[X_i])^2$$

Then, when taking the expectation of the matrix, this becomes:

$$E[(X_i - E[X_i])^2] = Var(X_i) = Cov(X_i, X_i)$$

Whenever $i \neq j$, so we are off the main diagonal, we'll instead get:

$$(X_i - E[X_i])(X_j - E[X_j])$$

which, taking expectations, becomes:

$$E[(X_i - E[X_i])(X_j - E[X_j])] = Cov(X_i, X_j)$$

Problem 2

We're interested in the relationship between two random variables $X \in \{0,1\}$ and $Y \in \{0,1\}$. Specifically, we're interested in something called the "odds ratio." Define following notation:

$$p(y,x) = P(Y = y, X = x)$$
$$p(y|x) = P(Y = y|X = x)$$

Suppose that $p(y,x) > 0$ for all possible combinations of (y,x). Then, we'll define the odds ratio as:

$$OR = \frac{\frac{p(1|1)}{p(1|0)}}{\frac{p(0|1)}{p(0|0)}}$$

(a) Express OR in terms of $p(0,0)$, $p(0,1)$, $p(1,0)$, and $p(1,1)$.

SOLUTION: We learned early in the course that:

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}$$

Using our notation, we can then reexpress the odds ratio:

$$\begin{aligned} OR &= \frac{\frac{p(1|1)}{p(1|0)}}{\frac{p(0|1)}{p(0|0)}} \\ &= \frac{\frac{p(1,1)/p(1)}{p(1,0)/p(0)}}{\frac{p(0,1)/p(1)}{p(0,0)/p(0)}} \\ &= \frac{\frac{p(1,1)}{p(1,0)}\frac{p(0)}{p(1)}}{\frac{p(0,1)}{p(0,0)}\frac{p(0)}{p(1)}} \\ &= \frac{\frac{p(1,1)}{p(1,0)}}{\frac{p(0,1)}{p(0,0)}} \end{aligned}$$

we can rest assured that we never divided by 0 during this process because $p(x,y) > 0$ for all values of (x,y).

(b) Suppose we have a sample $(Y_1, X_1), ..., (Y_n, X_n)$ that are iid $\sim (X,Y)$. Define

$$\hat{p}_n(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Y_i = y, X_i = x\}$$
$$\hat{OR}_n = \frac{\frac{\hat{p}_n(1,1)}{\hat{p}_n(1,0)}}{\frac{\hat{p}_n(0,1)}{\hat{p}_n(0,0)}}$$

Show that $\hat{OR}_n$ is a consistent estimator for OR.

SOLUTION: Consider any

$$\hat{p}_n(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Y_i = y, X_i = x\}$$

The fact that the $(Y_1, X_1), ..., (Y_n, X_n)$ are iid $\sim (X, Y)$ ensures that the $\mathbb{1}\{Y_i = y, X_i = x\}$ are iid $\sim \mathbb{1}\{Y = y, X = x\}$. Also, $\mathbb{1}\{Y_i = y, X_i = x\}^2 \leq 1$, so $E[\mathbb{1}\{Y_i = y, X_i = x\}^2] \leq 1$. Thus, we can apply the Weak Law of Large Numbers to any $\hat{p}_n(y,x)$ to say that:

$$\hat{p}_n(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Y_i = y, X_i = x\} \xrightarrow{p} E[\mathbb{1}\{Y = y, X = x\}] = p(y,x)$$

. We could then consider

$$\hat{OR}_n = \frac{\frac{\hat{p}_n(1,1)}{\hat{p}_n(1,0)}}{\frac{\hat{p}_n(0,1)}{\hat{p}_n(0,0)}}$$

to be a continuous function of $\hat{p}_n(1,1)$, $\hat{p}_n(1,0)$, $\hat{p}_n(0,1)$, $\hat{p}_n(0,0)$, each of which converge in probability. Continuity is ensured because $p(x,y) > 0$ for all values of (x,y), so we're never dividing by 0. Thus, by the CMT:

$$\hat{OR}_n = \frac{\frac{\hat{p}_n(1,1)}{\hat{p}_n(1,0)}}{\frac{\hat{p}_n(0,1)}{\hat{p}_n(0,0)}} \xrightarrow{p} \frac{\frac{p(1|1)}{p(1|0)}}{\frac{p(0|1)}{p(0|0)}} = OR$$

(c) What will

$$\sqrt{n}(\begin{bmatrix} \hat{p}_n(1,1) \\ \hat{p}_n(1,0) \\ \hat{p}_n(0,1) \\ \hat{p}_n(0,0) \end{bmatrix} - \begin{bmatrix} p(1,1) \\ p(1,0) \\ p(0,1) \\ p(0,0) \end{bmatrix})$$

converge to in distribution as $n \to \infty$? (The things in square brackets are 4×1 column vectors). Make sure the variance of the limiting distribution is specified (a general element-wise description is sufficient - not need to lay out the entire matrix). (Hint: look at the multivariate version of one of our familiar statistics results).

SOLUTION: $\begin{bmatrix} \hat{p}_n(1,1) \\ \hat{p}_n(1,0) \\ \hat{p}_n(0,1) \\ \hat{p}_n(0,0) \end{bmatrix}$ is a random vector, where each element is a sample mean. Moreover, the sample that is going into each sample mean is iid, and $E[\mathbb{1}\{Y_i = y, X_i = x\}^2] < \infty$, as discussed above. We also know that:

$$\begin{bmatrix} p(1,1) \\ p(1,0) \\ p(0,1) \\ p(0,0) \end{bmatrix} = \begin{bmatrix} E[\mathbb{1}\{Y = 1, X = 1\}] \\ E[\mathbb{1}\{Y = 1, X = 0\}] \\ E[\mathbb{1}\{Y = 0, X = 1\}] \\ E[\mathbb{1}\{Y = 0, X = 0\}] \end{bmatrix}$$

Thus, we can apply the CLT to say that:

$$\sqrt{n}(\begin{bmatrix} \hat{p}_n(1,1) \\ \hat{p}_n(1,0) \\ \hat{p}_n(0,1) \\ \hat{p}_n(0,0) \end{bmatrix} - \begin{bmatrix} p(1,1) \\ p(1,0) \\ p(0,1) \\ p(0,0) \end{bmatrix}) \xrightarrow{d} N(0, Var(\begin{bmatrix} \mathbb{1}\{Y = 1, X = 1\} \\ \mathbb{1}\{Y = 1, X = 0\} \\ \mathbb{1}\{Y = 0, X = 1\} \\ \mathbb{1}\{Y = 0, X = 0\} \end{bmatrix}))$$

The variance is the variance of a 4×1 column vector, so it will be a 4×4 matrix, where the element in the ith row and jth column is equal to the covariance of the ith and jth elements of the vector. Along the main diagonal, when $i = j$, this are the variances:

$$\begin{aligned} Var(\mathbb{1}\{Y = y, X = x\}) &= E[\mathbb{1}\{Y = y, X = x\}^2] - E[\mathbb{1}\{Y = y, X = x\}]^2 \\ &= E[\mathbb{1}\{Y = y, X = x\}] - E[\mathbb{1}\{Y = y, X = x\}]^2 \\ &= p(y,x) - p(y,x)^2 \end{aligned}$$

Off of the main diagonal, when $i \neq j$, we'll have the covariance of some $\mathbb{1}\{Y = y, X = x\}$ with some $\mathbb{1}\{Y = y', X = x'\}$ where either $y \neq y'$, $x \neq x'$, or both. That covariance will look like:

$$\begin{aligned} &Cov(\mathbb{1}\{Y = y, X = x\}, \mathbb{1}\{Y = y', X = x'\}) \\ &\quad = E[\mathbb{1}\{Y = y, X = x\}\mathbb{1}\{Y = y', X = x'\}] - E[\mathbb{1}\{Y = y, X = x\}]E[\mathbb{1}\{Y = y', X = x'\}] \\ &\quad = 0 - p(x,y)p(x',y') \\ &\quad = -p(x,y)p(x',y') \end{aligned}$$

To see why the first expectation is equal to 0, note that $(Y = y, X = x)$ and $(Y = y', X = x')$ are disjoint events - it is impossible for them to occur simultaneously. Thus, at least one of those indicator functions will always be 0, the expectation is also 0.

Problem 3

Suppose we're interested in studying the association of wages with other variables. Sepcifically, we consider a regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$$

where

$$\begin{aligned} Y &= \text{wage} \\ X_1 &= \text{age in years} \\ X_2 &= \text{years of schooling} \\ X_3 &= \text{years of experience} \end{aligned}$$

It is the case that, among our population of interest, everyone starts school at the age of 6 and everyone works every year that they are not in school (and are older than 6).

(a) We cannot estimate this regression consistently (even under a descriptive interpretation). Why not? Explain specifically why the problem you identify arises.

SOLUTION: We see that, in the situation as described, age, years of schooling, and years of experience are perfectly multicolinear - because people spend every year after 6 either in school or working, we can represent any of those three variables as a linear function of the other two:

$$X_1 = 6 + X_2 + X_3$$

(b) Propose an alternative regression that avoids the problem mentioned in part (a).

SOLUTION: The easiest solution would simply be to drop any of the three variables. For instance, we can drop X_3, at which point we can no longer express the variables as a linear function of each other, as we did in part a). We also are not really "losing information" by doing so, as we can always infer years of experience from the other two. For instance, under a best linear predictor interpretation,

$$\hat{Y}(X_1 = 20, X_2 = 12) = \hat{\beta}_0 + \hat{\beta}_1 * 20 + \hat{\beta}_2 * 12$$

evaluated at the OLS estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ will tell us the estimated best linear prediction of wage for someone who is 20 years old and has 12 years of education. We can then infer that this person has 2 years of experience, so this is also the best linear predictor of wage for someone who has 2 years of experience (and the other characteristics).

(c) In our description of the context, we implicitly assume that there is no such thing as "unemployment" - everyone either works or is in school every year after the age of 6. Suppose we now say that unemployment is a possibility (it is possible for someone to neither work nor be in school). Will the issue identified in part (a) still apply?

SOLUTION: No, because now there is no longer a linear relationship between X_1, X_2, and X_3 *always*. If someone spends any years unemployed, then, it will be the case that:

$$X_1 \neq 6 + X_2 + X_3$$

so we no longer have a multicolinearity problem. We can include all of X_1, X_2, and X_3 in the regression and estimate it consistently.

Problem 4

One of the values of multivariate linear regression is that it allows us to specify more general types of relationships between variables than simple linear regression. In class we discussed interaction effects as an example of this. We'll now look at another type of example. We define Y as: $Y = X + X^2$ where $X \sim N(0,1)$. Then, $E[Y|X] = X + X^2$.

(a) Consider a simple linear regression:

$$Y = \beta_0 + \beta_1 X + U$$

Under the best linear predictor interpretation, β_0 and β_1 will satisfy the following first-order conditions:

$$\begin{aligned} E[Y - \beta_0 - \beta_1 X] &= 0 \\ E[X(Y - \beta_0 - \beta_1 X)] &= 0 \end{aligned}$$

Solve this system of equations for β_0 and β_1. It may be useful to note that for $X \sim N(0,1)$, $E[X^2] = 1$ and $E[X^3] = 0$.

SOLUTION: We have a system of two equations and two unknowns. As such, let's solve the first equation for one of the unknowns:

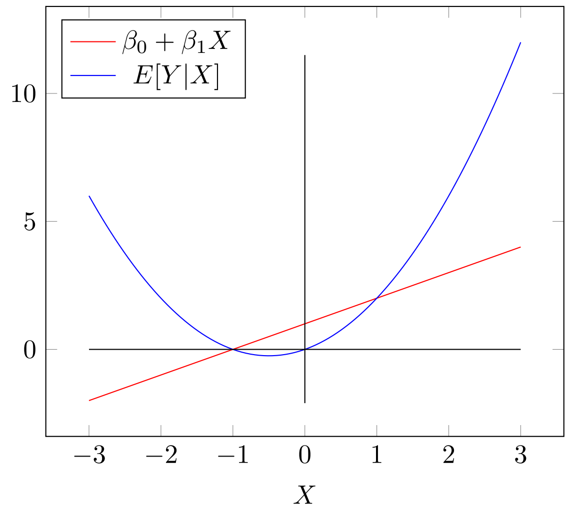
$$\begin{aligned} E[Y - \beta_0 - \beta_1 X] &= 0 \\ E[X + X^2 - \beta_0 - \beta_1 X] &= 0 \\ E[X] + E[X^2] - \beta_0 - \beta_1 E[X] &= 0 \\ \Rightarrow \beta_0 &= 1 \end{aligned}$$

And then we just need to solve for β_1 using the second equation:

$$\begin{aligned} E[X(Y - \beta_0 - \beta_1 X)] &= 0 \\ E[X(X + X^2 - \beta_0 - \beta_1 X)] &= 0 \\ E[X^2 + X^3 - \beta_0 X - \beta_1 X^2] &= 0 \\ E[X^2] + E[X^3] - \beta_0 E[X] - \beta_1 E[X^2] &= 0 \\ \Rightarrow \beta_1 &= 1 \end{aligned}$$

(b) Draw pictures of the best linear approximation to $E[Y|X]$, $\beta_0 + \beta_1 X$, and the actual $E[Y|X]$ on the same graph. The graph need not be extremely precise - it just needs to capture the major features of the functions.

SOLUTION:



(c) Now consider the multivariate linear regression:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + U$$

Under the best linear predictor interpretation, β_0, β_1, and β_2 will satisfy the following first-order conditions:

$$E[Y - \beta_0 - \beta_1 X - \beta_2 X^2] = 0$$
$$E[X(Y - \beta_0 - \beta_1 X - \beta_2 X^2)] = 0$$
$$E[X^2(Y - \beta_0 - \beta_1 X - \beta_2 X^2)] = 0$$

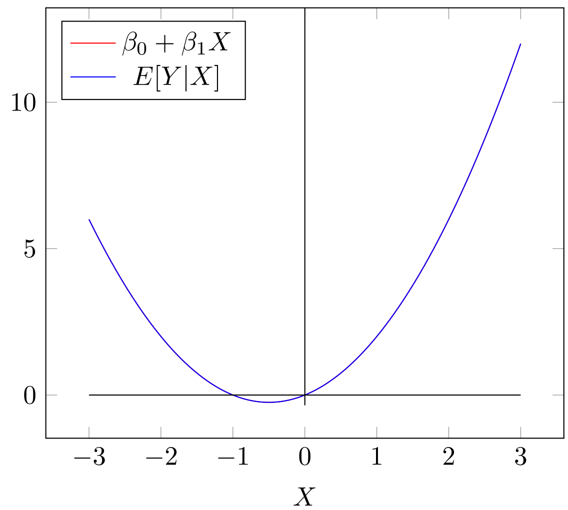
Solve this system of equations for β_0, β_1, and β_2. Make use of the previously given moments of a standard normal, and that $E[X^4] = 3$ for $X \sim N(0,1)$.

SOLUTION: We've now got a system of three equations with three unknowns. Solving this:

$$
\begin{aligned}
E[Y - \beta_0 - \beta_1 X - \beta_2 X^2] &= 0 \\
E[X + X^2 - \beta_0 - \beta_1 X - \beta_2 X^2] &= 0 \\
E[X] + E[X^2] - \beta_0 - \beta_1 E[X] - \beta_2 E[X^2] &= 0 \\
1 - \beta_2 &= \beta_0 \\
E[X(Y - \beta_0 - \beta_1 X - \beta_2 X^2)] &= 0 \\
E[X(X + X^2 - \beta_0 - \beta_1 X - \beta_2 X^2)] &= 0 \\
E[X^2 + X^3 - \beta_0 X - \beta_1 X^2 - \beta_2 X^3)] &= 0 \\
E[X^2] + E[X^3] - \beta_0 E[X] - \beta_1 E[X^2] - \beta_2 E[X^3] &= 0 \\
\Rightarrow \beta_1 &= 1 \\
E[X^2(Y - \beta_0 - \beta_1 X - \beta_2 X^2)] &= 0 \\
E[X^2(X + X^2 - \beta_0 - \beta_1 X - \beta_2 X^2)] &= 0 \\
E[X^3 + X^4 - \beta_0 X^2 - \beta_1 X^3 - \beta_2 X^4] &= 0 \\
E[X^3] + E[X^4] - \beta_0 E[X^2] - \beta_1 E[X^3] - \beta_2 E[X^4] &= 0 \\
3 - 3\beta_2 &= \beta_0 \\
\Rightarrow 3(1 - \beta_2) &= 1 - \beta_2 \\
\Rightarrow \beta_2 &= 1 \\
\Rightarrow \beta_0 &= 0
\end{aligned}
$$

(d) Draw pictures of the new best linear approximation to $E[Y|X]$, $\beta_0 + \beta_1 X + \beta_2 X^2$, and the actual $E[Y|X]$ on the same graph. The graph need not be extremely precise - it just needs to capture the major features of the functions.

SOLUTION:



We now see that the best "linear" approximation to the conditional expectation is simply the same as the conditional expectation.

(e) We see that the approximation improves by cleverly allowing for the nonlinearity in Y. Think of a real-life Y and X where allowing for nonlinearities in such a way may be useful. That is, think of a Y and X that you think might be (very approximately) described by $Y = X + X^2$ Neither math nor accuracy to the real world are required - just think of an example that you think might fit and give some economic intuition.

SOLUTION: Any variable that could plausibly have an exponential relationship with another variable could satisfy this. For instance, wages often increase by an increasing amount as you move "up" to higher level positions in a company. Thus, we might have Y as wages and X as some kind of measure of how high-ranking someone is within in their company. Then, we might expect to see that Y increases at a greater than linear rate with respect to X (on the domain where $X > 0$).

Problem 5

Let's consider another example of the selection on observables identification strategy (based on Fagereng et al (2021)). This paper considers the question of why wealthy parents tend to have wealthy children. Specifically, the paper is interested in the extent to which wealthiness passes from parents to children due to favorable genetic characteristics versus monetary endowments (buying stuff for the kid after they are born, like better schooling, or simply giving the kid money).

(a) Consider a causal model:

$$Y = \beta_0 + \beta_1 W + U$$

where:

$$Y = \text{child's wealth (upon reaching adulthood)}$$
$$W = \text{parents' wealth}$$

(We assume that W will here stand in for the totality of parental characteristics). Fagereng et al were concerned that they could not consistently estimate the causal parameter β_1 for this model. Explain the specific concern given the description of the research question.

SOLUTION: You would not be able to estimate the causal parameter β_1 if unobserved determinants of child wealth, U are correlated with parental wealth. In the Fagerang et al example, the authors are specifically concerned that child genetics could be part of U. "Good" genetics may be associated with higher wages and may be more likely to be inherited from wealthier parents, which would induce a correlation between U and W.

(b) To answer this question, the paper considers a situation in the 20th century in which many Norwegian families adopted Korean children through a centralized agency. The agency did not allow the adoptive families to request any kind of characteristics of their adoptive children. Instead, the agency would simply match families with the next child in line for adoption, in the order that families were approved for adoption (where the order depends on when the family applied to adopt and how long it took them to get approved).

Define new variable

$$T = \text{Measure of when the adoptive family's application was approved}$$

What assumption can we make about this new variable in order to enable us to identify β_1 from part (a)? Interpret this assumption in words. (Hint: Follow the class example of using a control variable to identify a causal parameter).

SOLUTION: Following the example from class, we can make an assumption of conditional mean independence:

$$E[U|W,T] = E[U|T]$$

this would say that, if you know when the family's adoption request was accepted, you can't "learn" anything else about U from knowing a family characteristic like W. In particular, it could be plausible in this context that child genetics did not correlate at all with W once conditioning on time of approval. Families were matched to adoptees based solely on when the family was approved and where the child was in the "queue" to get adopted. Conditioning on when the match was made, there should be no association between child characteristics, which are part of U, and family characteristics, like W.

We will also assume that $E[U|T]$ is linear:

$$E[U|T] = \alpha_0 + \alpha_2 T$$

(c) Consider the new regression equation:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 W + \tilde{\beta}_2 T + \tilde{U}$$

Using the assumption that you made in part (b), show that $\tilde{\beta}_1$ will consistently estimate the causal β_1 from part 1. (Hint: Again, follow the example from class).

SOLUTION: We want to estimate the causal model:

$$Y = \beta_0 + \beta_1 W + U$$

but are concerned that $E[WU] \neq 0$. However, in the previous part, we assumed that:

$$E[U|W,T] = E[U|T] = \alpha_0 + \alpha_2 T$$

Define a new error term:

$$\begin{aligned}
\tilde{U} &= U - E[U|T] \\
\Rightarrow U &= \tilde{U} - \alpha_0 - \alpha_2 T
\end{aligned}$$

Then, we can rewrite our causal model as:

$$\begin{aligned}
Y &= \beta_0 + \beta_1 W + \tilde{U} - \alpha_0 - \alpha_2 T \\
Y &= \beta_0 - \alpha_0 + \beta_1 W - \alpha_2 T + \tilde{U}
\end{aligned}$$

Then, for ease of exposition, we can relabel the parameters of the above to look like a typical regression equation:

$$\begin{aligned}
Y &= \underbrace{\beta_0 - \alpha_0}_{=\tilde{\beta}_0} + \underbrace{\beta_1}_{=\tilde{\beta}_1} W + \underbrace{-\alpha_2}_{=\tilde{\beta}_2} T + \tilde{U} \\
Y &= \tilde{\beta}_0 + \tilde{\beta}_1 W + \tilde{\beta}_2 T + \tilde{U} \qquad (1)
\end{aligned}$$

We see that we have defined $\tilde{\beta}_1 = \beta_1$. Thus, our new regression will estimate the causal β_1, so long as regression (1) is consistently estimable.

We can show that $\tilde{U}$ is mean independent of the collection of dependent variables in our new regression:

$$\begin{aligned} E[\tilde{U}|W,T] &= E[U - E[U|W,T]|W,T] \\ &= E[U|W,T] - E[U|W,T] \\ &= 0 \end{aligned}$$

Mean independence implies uncorrelatedness, so given the above, we know that:

$$E[\begin{bmatrix} 1 \\ W \\ T \end{bmatrix} \tilde{U}] = 0$$

Along with assuming no perfect collinearity and $E[\begin{bmatrix} 1 \\ W \\ T \end{bmatrix} \begin{bmatrix} 1 & W & T \end{bmatrix}] < \infty$, which seem innocuous in this setting, the above tells us that we can consistently estimate the parameters of regression (1), which includes $\tilde{\beta}_1 = \beta_1$.

(d) What will $\tilde{\beta}_2$ estimate? Is it causal?

SOLUTION: In the previous part, we defined $\tilde{\beta}_2 = -\alpha_2$. α_2 was not a causal parameter, it was a descriptive parameter giving the shape of the conditional expectation of U given T. Thus, it tells us, descriptively, how non-parental wealth determinants of child wealth depend on the timing of the adoption.

Problem 6

Lets return to the data from Angrist and Krueger (1999), as was used in Problem 6 of the previous problem set. We will again refer to X as years of education and Y as log wage. We will continue to assume that $E[X^4], E[Y^4] < \infty$.

(a) The regression equation:

$$Y = \beta_0 + \beta_1 X + U \tag{1}$$

is likely difficult to interpret causally. Pick one "component" of causally-defined U that you would expect to be correlated with X (thereby preventing us from consistently estimating causal β_1). For your chosen component of U, guess what direction you think that omitting that variable will "bias" the estimate of OLS $\hat{\beta}_1$ (relative to causal β_1), appealing to the formula for omitted variable bias and your economic intuition.

SOLUTION: U includes all of the non-education determinants of wage. We want to pick one of these that might be correlated with education.

For instance, we could think of intelligence, and call it Z. The omitted variable bias formula tells us that the OLS estimate of regression (1) will converge to:

$$\beta_1 + \beta_2 \frac{Cov(X, Z)}{Var(X)}$$

where β_2 is the causal effect of intelligence on wage. I would expect both β_2 and $Cov(X, Z)$ to be positive - being intelligent might directly increase your wage by making you better at certain jobs while also encouraging you to remain in school for a longer time. That would suggest that omitting this variable cause an OLS estimate to converge to an something *above* the actual β_1.

(b) Suppose someone proposed using variable: $A =$ year of birth ("year_of_birth" in the data set) as a control variable, and claims that including this in the regression:

$$Y = \beta_0 + \beta_1 X + \beta_2 A + U \tag{2}$$

will allow us to estimate β_1 as a causal parameter consistently. Do you think this idea makes sense? Why or why not?

SOLUTION: As discussed in question 5, control variables need to meet a mean independence assumption in order to allow us to recover causal parameters. For us, that would mean:

$$E[U|X, A] = E[U|A]$$

which would say that, if we already know someone's age, knowing how much schooling they have would not cause us to update our "best guess" as to the other causal determinants of their wage. In this setting this makes little sense - age doesn't tell us much about how the amount of schooling decision was made, so it likely can't create mean independence. For instance, the association between intelligence and years of education discussed in part a) would likely apply for people of any age.

(c) Perform regressions according to equations (1) and (2) using R or another language of your choice.

(d) Appealing to knowledge rather than computational outputs, what will happen to the R^2 going from (1) to (2)? Will the same thing necessarily happen to the adjusted R^2?

SOLUTION: We know that the R^2 will weakly increase with the addition of extra regressors, so going from regression (1) to (2) must (weakly) increase the R^2. This property doesn't apply to the adjusted R^2, so it might increase or decrease from regression (1) to (2).

(e) (5 Points Bonus) Calculate the R^2 and adjusted R^2 for (1) and (2). Interpret what you find (pay close attention to the formulae for those two statistics).

SOLUTION: You'll find that the R^2's are extremely similar (likely identical to the number of decimal places outputted by your software). Compare the R^2 and adjusted R^2 formulas:

$$R^2 = 1 - \frac{SSR}{TSS}$$
$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

The adjustment made to adjusted R^2 depends not only on the number of regressors but also on the sample size. When you have a large sample, as we do here, it will be the case that:

$$\frac{n-1}{n-1-1} \approx \frac{n-1}{n-2-1} \approx 1$$

Thus, for both regressions (1) and (2), the two statistics will be very nearly identical. The adjustment matters more when you add many regressors (or if you have a small sample size).