QE Problem Set 2

1 No.

HM Health Secretary appears to have concluded from the fact that individuals who ate chocolate very frequently weighed about 2.5kg less on average than those who consumed no chocolate at all that an increase in chocolate consumption would lead to a decrease in weight.

This ~~conclusion~~ inference is valid only if the other causal determinants of weight are not systematically related to chocolate consumption. We know that this is not the case because individuals who ate chocolate more frequently also tended to consume more calories per week, and calorie intake is causally related to weight.

The observation that individuals who ate chocolate very frequently weighed less than those who consumed no chocolate at all could be explained by the fact that the latter group ~~includes~~ consists disproportionately of diabetics who ~~on average weigh more than no~~ are more likely to be obese than non-diabetics.

2 $nchildren = \beta_0 + \beta_1 educ + u$

Other determinants of nchildren collected in u: income, country of residence, area of residence (urban/rural), year of birth, religious belief.

Income, country of residence, area of residence, and year of birth are likely to be correlated with educ.

Women with high income, in wealthy, developed countries, living in cities, and born later are likely to have a higher educ.

b ~~Whether regression using this data provides a consistent estimate of β_1 depends on the sampling methodology.~~ No.

Estimate $\hat{\beta}_1 = \hat{cov}(nchildren, educ)/\hat{var}(educ)$ is a consistent estimator of the population regression parameter $\beta_1 = cov(nchildren, educ)/var(educ)$ since sample covariance and sample variance are consistent estimators of population covariance and population variance.

$(\beta_0, \beta_1) = \text{argmin}_{b_0, b_1} E[\text{nchildren} - (b_0 + b_1 \text{educ})]^2$

β_1 is a consistent estimator of β_1 only if u is orthogonal to educ, such that $E(u) = 0$ and ~~E~~ $cov(X, u) = 0$, hence (β_0, β_1) solve $\text{argmin}_{b_0, b_1} E[\text{nchildren} - (b_0 + b_1 \text{educ})]^2$

From (a), determinants captured in u are likely to be correlated with educ, hence it is likely that $cov(X, u) \neq 0$, and $\hat{\beta}_1$ is not a consistent estimator of β_1.

3a $n = 1388$

$\text{bwght}_i = 3395 - 15 \text{cigs}_i + \hat{u}_i$

$E(\text{bwght} \mid \text{cigs} = 0) = 3395$, given $E(\hat{u}) = 0$
$E(\text{bwght} \mid \text{cigs} = 20) = 3095$
The predicted birthweight of a baby ~~falls by~~ given that the mother smokes one pack of cigarettes per day during the pregnancy is 300g (~9%) lower than that predicted given that the mother does not smoke during the pregnancy.

b No.

Regression in (1) gives a reliable ~~of~~ estimate of the causal effect of smoking on birthweight only if the causal model is linear and u, which captures all other determinants of birth weight is orthogonal to cigs: $E(u) = 0$ and ~~cov~~ $cov(\text{cigs}, u) = 0$, such that ~~estimate regression estimate (which is a consis~~ population regression parameters (which the ~~given estim~~ sample regression parameters consistently estimate) are equal to the ~~values~~ parameters in the causal model.

There is no reason to think that the causal model is linear.

It is not likely that u is orthogonal to cigs. Determinants of birthweight captured in u likely include income, eating habits, drinking habits. Plausibly, mothers with higher income are less likely to smoke and more likely to have better access to ~~healt~~ healthcare, hence more likely to deliver babies with higher birthweight. There would then be some correlation between cigs and u, the assumption of orthogonality fails, and the given regression does not give a reliable estimate of the causal effect

c $3500 = \cancel{\text{...}}$ bwght$_i$ = $3395 - 15\,cigs_i + \hat{u}_i$
$cigs_i = -7$

Since it is not possible to smoke -7 cigarettes a day, given only information on smoking ~~...~~ frequency during pregnancy, ~~I we would never expect~~ the predicted ~~b~~ birthweight never exceeds 3395. Given only information on smoking frequency during pregnancy, we would never predict a birthweight of 3500.

4a) Let $C = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$

~~FOCs:~~

~~∂C~~

~~FOCs:~~

~~$\partial C/\partial b_0 = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-1)$~~

~~$= -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)$~~

~~$= 0$ (1)~~

~~$\partial C/\partial b_1 = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-X_i)$~~

~~$= -2X_i \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)(X_i)$~~

~~$= 0$ (2)~~

~~From (2),~~

~~$-2X_i(1/n)\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0$, $(X_i) = 0$~~

~~$-2X_i($~~

~~From (1),~~

~~$(1/n)\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0$,~~

~~$\bar{Y} - b_0 - b_1\bar{X} = 0$~~,

~~FOCs:~~

~~$\partial C/\partial \beta_0 =$~~

FOCs:

$\partial C/\partial b_0 = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-1)$

$= -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)$

~~$\partial C/\partial b_1 = \sum_{i=1}^{n}$~~

$-2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0$ (1)

$\partial C/\partial b_1 = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-X_i)$

$= -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)(X_i)$

$-2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)(X_i) = 0$ (2)

From (1),

$1/n \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0$

$\bar{Y} - \beta_0 - \beta_1\bar{X} = 0$

$\beta_0 = \bar{Y} - \beta_1\bar{X}$ (3)

From ②

$\frac{1}{n}\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i)(X_i)=0$ ④

Sub ③ into ④

$\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y}+\beta_1\bar{X}-\beta_1 X_i)(X_i)=0$,

$\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})X_i = \cancel{\frac{1}{n}\sum_{i=1}^{n}}$

$= \beta_1(\frac{1}{n})\sum_{i=1}^{n}(X_i-\bar{X})X_i$

$\beta_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i) / \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i)$

$= \hat{cov}(Y,X)/\hat{var}(X)$

not very sure how this step works, it would seem to imply $\bar{X}=0$ since $cov(Y,X) := E(Y-E(Y))(X-E(X))$

b $\hat{u}_i = Y_i-\beta_0-\beta_1 X_i$

From ①, $\sum_{i=1}^{n}\hat{u}_i=0$ ⑤

From ②, $\sum_{i=1}^{n}\hat{u}_i X_i=0$ ⑥

From ⑤, $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i=0$ ⑦

From ⑥, $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i X_i=0$ ⑧

From ⑦, ⑧,

$\hat{cov}(X_i,\hat{u}_i) = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i X_i - (\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i)(\frac{1}{n}\sum_{i=1}^{n}X_i)$

$= 0$

c $\hat{corr}(Y,X) = \hat{cov}(Y,X)/\sqrt{\hat{var}(Y)\hat{var}(X)}$

$= \beta_1\sqrt{\hat{var}(X)/\hat{var}(Y)}$

$= \beta_1 \hat{sd}(X)/\hat{sd}(Y)$

d $\hat{\beta}_0+\hat{\beta}_1\bar{X}$

$= \bar{Y}-\hat{\beta}_1\bar{X}+\hat{\beta}_1\bar{X}$

$= \bar{Y}$

$\hat{\beta}_0+\hat{\beta}_1 X$ passes through $(\bar{X},\bar{Y})$

5 $Y_i = \beta_0+\beta_1 X_i+u_i$

$cov(Y_i,X_i) = cov(\beta_0+\beta_1 X_i+u_i, X_i)$

$= \beta_1 cov(X_i,X_i)$

$= \beta_1 var(X_i)$

$(\rho_0,\rho_1) = \cancel{\text{argm}}\ \text{argmin}_{b_0,b_1} E(Y_i-b_0-b_1X_i)^2$

Let $C = E(Y_i-b_0-b_1X_i)^2$

~~FOCs:~~

FOCs:

$\partial C/\partial b_0 = E\,2(Y_i-b_0-b_1X_i)(-1)$

$= -2E(Y_i-b_0-b_1X_i)$

$E(Y_i-\rho_0-\rho_1X_i)=0$ ①

$\partial C/\partial b_1 = E\,2(Y_i-b_0-b_1X_i)(X_i)$

$= -2E(Y_i-b_0-b_1X_i)(X_i)$

$E(Y_i-\rho_0-\rho_1X_i)(X_i)=0$ ②

From ①,

$E(Y_i)-\rho_0-\rho_1E(X_i)=0$

$\rho_0 = E(Y_i)-\rho_1E(X_i)$ ③

~~$= \beta_0+\beta_1E(X_i)+E(u_i)-\rho_1E(X_i)$ ③~~

Sub ③ into ②

$E(Y_i-(E(Y_i)-\rho_1E(X_i))-\rho_1X_i)X_i=0$

$E(Y_i-E(Y_i))X_i-\rho_1E(X_i-E(X_i))X_i$

$\rho_1 = cov(Y,X)/var(X)$

$= \beta_1 var(X)/var(X)$

$= \beta_1$

Date

6 ~~co~~

~~It does not follow.~~

It does not follow.

(β_0, β_1) solves $\arg\min_{b_0, b_1} E(Y_i - b_0 - b_1 X_i)^2$

$\beta_1 = cov(Y, X)/var(X)$

~~Let β_0' and β_1' be the parament~~

~~Let $\beta_0' + \beta_1' Y$~~

Let $X_i = \beta_0' + \beta_1' Y_i$ be the population linear regression of X_i on Y_i.

(β_0', β_1') solves $\arg\min_{b_0', b_1'} E(X_i - b_0' - b_1' Y_i)^2$

$\beta_1' = cov(Y, X)/var(Y) \neq 1/\beta_1$

7 $\cancel{var(E(Y|X)) = E[(E(Y|X) - E(E(Y|X)))]^2}$

$\cancel{= E[E(Y|X) - E(Y)]^2}$

$\cancel{= E[E(Y|X)^2 + E(Y)^2 - 2E(Y)E(Y|X)]}$

$\cancel{= E(E(Y|X)^2) + E(Y)^2 - 2E(Y)^2}$

$\cancel{= E(E(Y|X)^2) - E(Y)^2}$

~~var~~

$\cancel{E(var(Y|X)) = E(E((Y - E(Y|X))^2|X))}$

$\cancel{= E(Y - E(Y|X))^2}$

$\cancel{= E(Y^2 + E(Y|X)^2 - 2YE(Y|X))}$

$\cancel{= E(Y^2) + E(E(Y|X)^2) - 2YE(Y)\ 2E(Y)^2}$

$\cancel{= E(Y^2) + E(E(Y|X)^2) - 2E(YE(Y|X))}$

$\cancel{= E(Y^2) + E(E(Y|X)^2) - 2E(Y)^2}$

$\cancel{var(E(Y|X))}$

$\cancel{E_x var_{Y|X}(Y|X) = E_x[E_{Y|X}(Y^2|X) - E^2_{Y|X}(Y|X)]}$

$var(E(Y|X)) + E(var(Y|X))$

$\cancel{= E_x[E^2_{Y|X}(Y|X)] - E_x}$

$= E_x E^2_{Y|X}(Y|X) - E^2_x E_{Y|X}(Y|X)$

$\quad + E_x[E_{Y|X}(Y^2|X) - E^2_{Y|X}(Y|X)]$

$= E_x E_{Y|X}(Y^2|X) - E^2_x E_{Y|X}(Y|X)$

$= E_Y(Y^2) - E^2_Y(Y)$

$= var(Y)$

8 Population linear regression: $Y = \rho_0 + \rho_1 X + u$

(ρ_0, ρ_1) solve $\arg\min_{b_0, b_1} E(Y - b_0 - b_1 X)^2$

$\rho_1 = cov(Y, X)/var(X)$

$\rho_0 = E(Y) - \rho_1 E(X)$

Let $C = E_x\{E_{Y|X}[Y|X] - (b_0 + b_1 X)\}^2$

FOCs:

$\partial C/\partial b_0 = E_x 2\{E_{Y|X}[Y|X] - (b_0 + b_1 X)\}(-1)$

$= -2(E_x E_{Y|X}[Y|X] - E_x(b_0 + b_1 X))$

$= -2E_Y(Y) + 2b_0 + 2b_1 E_x(X)$

$-2E(Y) + 2\rho_0 + 2\rho_1 E(X) = 0,$

$\rho_0 = E(Y) - \rho_1 E(X)$

Date

$$\partial C/\partial b_1 = E_X 2\{E_{Y|X}[Y|X] - (b_0 + b_1 X)\}(-X)$$

$$= \cancel{-2E_X\{XE_{Y|X}[Y|X] - (b_0 X + b_1 X^2)\}}$$

$$\cancel{E_X\{XE_{Y|X}[Y|X] - (\beta_0 X + \beta_1 X^2)\} = 0}$$

$$\cancel{E_X\{XE_{Y|X}[Y|X] - ((E(Y) - \beta_1 E(X))X + \beta_1 X^2)\} = 0},$$

$$\cancel{E_X\{E_{Y|X}(Y|X) - (\cdot(\beta_0 - \beta_1 E(X)}$$

$$E_X\{E_{Y|X}(Y|X) - ((E(Y) - \beta_1 E(X)) + \beta_1 X)\}(X) = 0$$

$$E_X\{E_{Y|X}(Y|X) - E(Y) - \beta_1(X - E(X))\}(X) = 0$$

$$E_X\{E_{Y|X}(Y|X) - E(Y)\}(X) = \beta_1 E_X(X - E(X))(X)$$

$$\beta_1 = E_X\{XE_{Y|X}(Y|X) - XE(Y)\} / E_X(X - E(X))(X)$$

$$= (E_X(XE_{Y|X}(Y|X)) - E(X)E(Y)) / \mathrm{var}(X)$$

How does this step work?

$$= (E_X E_{Y|X}(XY|X) - E(X)E(Y)) / \mathrm{var}(X)$$

$$= (E(XY) - E(X)E(Y)) / \mathrm{var}(X)$$

$$= \mathrm{cov}(X, Y) / \mathrm{var}(X)$$