**Key to exam 2, eco 311, spring 2014**

Q1: We need to make sure ceteris paribus holds, that is, everything else should be equal except the fertilizer. We can plant two lots of corns, giving them the same amount of water, same quality of soil, same sunshine, and so on. But apply the fertilizer to only one lot. Then we can compare the outputs in the two lots. The fertilizer is effective (having causal effect) in output if the difference in output is statistically significant.

Q2: The key assumption is $E(u|x) = 0$ or $\text{cov}(x, u) = 0$. In words, the regressor $x$ should be exogenous by being uncorrelated with the error term. Simple regression puts many variables in the error term. Chance is high $x$ can be correlated with at least some of those (omitted) variables. In other words, it is very likely that $\text{cov}(x, u) \neq 0$ for the simple regression, so it is unsuitable for inducing causality.

Q3: *On average*, a person smokes 7.14 cigarettes if his annual income is zero.

Q4: 

$$|t| = \left| \frac{0.0798664 - 0.1}{0.0528204} \right| = |-0.38| = 0.38 < 1.96$$

so we do *not* reject the null hypothesis at 0.05 level. We compute the absolute value of the t test because this is a two-tailed (or two-sided) test.

Q5: Note that $\text{dollarincome} = 1000 \times \text{income}$. So 

$$\hat{c}_1 = \frac{\hat{\beta}_1}{1000} = \frac{0.0798664}{1000} = 0.0000799$$

Q6: First note that $\beta_2 < 0$ since education has negative effect on cigs. Second, 

$$\text{cov}(x_1, x_2) = \text{cov}(\text{income}, \text{education}) > 0$$

as more educated people tend to have higher income. So 

$$\hat{\beta}_1 \rightarrow \beta_1 + \beta_2 \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)}$$

which implies that 

$$\hat{\beta}_1 < \beta_1$$
because $\beta_2 \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)} < 0$. That means $\hat{\beta}_1$ underestimates the true effect of income on cigs.

Q7:

$\hat{\beta}_1 \pm 1.645 \text{se} = 0.0798664 \pm 1.645(0.0528204) = (-0.0071155, 0.1668484)$

With 90% probability, the true value of $\beta_1$ is inside this interval.

Q8: Basically we switch the dependent and independent variables. Recall that

\[
\text{coefficient estimate} = \frac{\text{sample covariance}}{\text{sample variance of regressor}},
\]

so

\[
\hat{c}_1 = \frac{S_{cigs, income}}{S^2_{cigs}}, \quad \hat{\beta}_1 = \frac{S_{cigs, income}}{S^2_{income}}.
\]

As a result

$\hat{c}_1 \neq \frac{1}{\hat{\beta}_1}$

Q9: *Holding income and age constant*, one more year of schooling is associated with reducing the number of cigarettes smoked by 0.3775954.

Q10: False. Age does not matter not because its coefficient is close to zero, but because its t value is less than 1.96 in absolute value, or its p value is greater than 0.05. Remember, $\hat{\beta}$ can be manipulated, but $t$ and $p$ values cannot. It is possible that a variable has coefficient close to zero, but is still significant (having big t value).

Q11: First we need to keep the RSS of unrestricted regression 150175.667. Next we need to keep the RSS of restricted regression

```
reg cigs
```

Here, the restricted regression only includes the intercept term, which is equal to the sample mean of cigs (see Q9 in HW1). Thus the RSS, is the TSS in the unrestricted regression 151753.683. Now the F test is

\[
F = \frac{(151753.683 - 150175.667)/3}{150175.667/(807 - 3 - 1)} = 2.81,
\]
where the number of restriction is 3, and the denominator degree of freedom is 803

Q12: \( \hat{y} = \hat{\beta}_0 + \hat{\beta}_{\text{income}}(20) + \hat{\beta}_{\text{educ}}(10) + \hat{\beta}_{\text{age}}(40) = 9.75 \) The stata commands (not required for this problem) are

```
. dis 12.85394+.1171126*20-.3775954*10-.0416932*40
9.75251
```

Q13: In step 1 we need to regress income onto educ and age, and save the residual. In step 2 we regress cigs onto that residual obtained in step 1. The stata commands (not required for this problem) are

```
reg income educ age
predict rhat, r
reg cigs rhat
```

Q14: According to the formula

\[
\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{SST_{\text{educ}}(1 - R^2_{\text{educ}})}
\]

and

\[
SST_{\text{educ}} = (n - 1)S^2_{\text{educ}},
\]

if the variance of educ \( S^2_{\text{educ}} \) gets smaller, \( SST_{\text{educ}} \) will get smaller; the variance and standard error of \( \hat{\beta}_1 \) will get bigger.

Q15: The model can explain 1.04% total variation of cigs.

Q16: Multicollinearity is present if regressors are strongly correlated with each other. That will cause, for example, \( R^2_{\text{educ}} \) in the above formula to be close to one. As a result, the standard error of \( \hat{\beta}_1 \) will be very big, t value will be very small, and \( \hat{\beta}_1 \) is estimated imprecisely.
Q17:

\[ \sum y_i y_i = \sum \hat{y}_i (\hat{y}_i + \hat{u}_i) \quad (1) \]
\[ = \sum \hat{y}_i^2 + \sum \hat{y}_i \hat{u}_i \quad (2) \]
\[ = \sum \hat{y}_i^2 + \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i) \hat{u}_i \quad (3) \]
\[ = \sum \hat{y}_i^2 + \hat{\beta}_0 \sum \hat{u}_i + \hat{\beta}_1 \sum x_i \hat{u}_i \quad (4) \]
\[ = \sum \hat{y}_i^2 \quad (5) \]

since the first order conditions (FOC) of the OLS imply that the last two terms in (4) are both zero.