Name: ________________________________ Spring 2017

EC200 Econometrics and Applications

# Unit 1 Quiz

You have 50 minutes to complete this quiz. There are 30 total points.

1. *(14 points)* Biddle and Hammermesh (1990) study whether hours worked affects sleep. One simple way to recreate the spirit of their results is to regress total *minutes* of sleep per week (*sleep*) on total *minutes* of worked per week (*totwork*), controlling for years of education (*educ*), as shown below:

   (a) Write the population model that underlies this OLS regression. *[2 points]*

   $sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + u$

   (b) Interpret $\widehat{\beta}_{totwrk}$, the coefficient on minutes of weekly work. *[2 points]*

   A one-minute increase in total weekly work is associated with a 0.15-minute reduction in total minutes of sleep per week.

(c) One sample member, Susan, has 12 years of schooling, works exactly 15 hours per week (900 minutes), and sleeps exactly 7 hours per night (2940 minutes). What is her predicted minutes of sleep per week? What is her residual? *[2 points]*

$$\widehat{sleep} = 3756.22 - 0.15(900) - 12(13.50) = 3459.45$$
$$u = sleep - \widehat{sleep} = 2940 - 3459.45 = -519.45$$

(d) Irene says that because Susan's actual sleep differs from the model's prediction, the model is biased. Explain why you do or do not agree with her. *[2 points]*

*Potential key points:*

- Bias has nothing to do with whether Susan's actual sleep equals her predicted sleep, and in fact, it may be the case that every observation differs from its predicted value without the presence of bias.
- To know whether the estimates are biased, we need to ask whether $E[\beta_{totwrk}] = \beta_{totwrk}$, and so on for the other coefficients
- It is possible that there are other, omitted variables, that are correlated with sleep and either work or schooling (or both), which would cause the results of the regression to be biased.

(e) Set up a hypothesis test of whether years of education affect sleep. Report the null hypothesis, alternative hypothesis, test statistic, and decision. Can you reject the null hypothesis at the 5% level? At the 1% level? *[3 points]*

Null hypothesis: $\beta_{educ} = 0$

Alternative hypothesis: $\beta_{educ} \neq 0$

Test statistic: $t = -2.38$

Decision at 5% level? Reject the null hypothesis, as $0.018 = p < 0.05$

Decision at 1% level? Do not reject the null hypothesis, as $0.018 = p > 0.01$

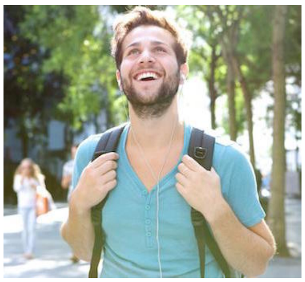
(f) Suppose you estimate the same model, but you use the log of minutes worked, $log(totwrk)$. Interpret the new coefficient on $log(totwrk)$. Does this model or the first one have a better fit? How do you know? *[3 points]*

```
.
. regress _hrs1  _doesnthurt childs  educ age if _female == 1 & childs > 0

      Source |       SS           df       MS      Number of obs   =     1,172
-------------+----------------------------------   F(4, 1167)      =     70.12
       Model |  102760.201         4  25690.0502   Prob > F        =    0.0000
    Residual |  427534.475     1,167  366.353449   R-squared       =    0.1938
-------------+----------------------------------   Adj R-squared   =    0.1910
       Total |  530294.676     1,171  452.856256   Root MSE        =     19.14

------------------------------------------------------------------------------
       _hrs1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _doesnthurt |   .5572165   1.120283     0.50   0.619    -1.640778     2.75521
      childs |  -.8703755    .426837    -2.04   0.042    -1.707829   -.0329217
        educ |   1.146882   .1924429     5.96   0.000     .7693097    1.524455
         age |  -.4831164   .0339084   -14.25   0.000    -.5496446   -.4165883
       _cons |    31.4043   3.456793     9.08   0.000     24.62208    38.18653
------------------------------------------------------------------------------
```

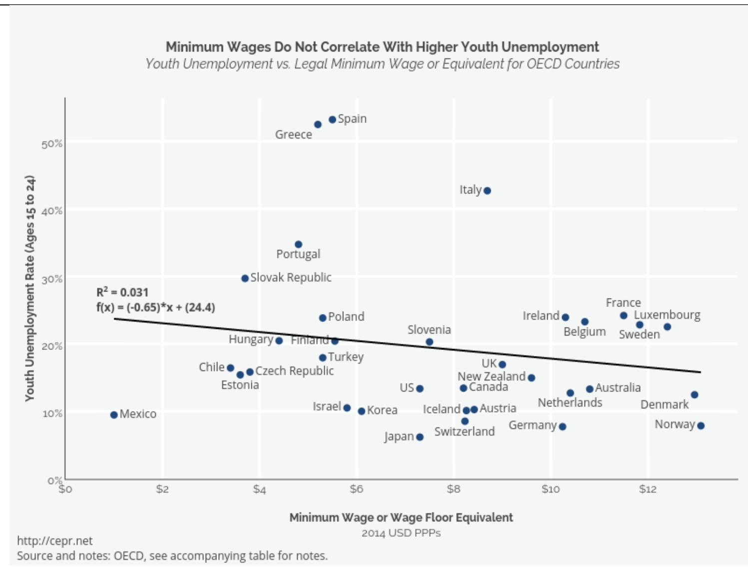
Interpretation: *A 1-percent increase in total minutes of weekly work is associated with a 2.32-minute decrease in total minutes slept that week*

Better fit? How do you know? *To determine fit, compare the $R^2$ values between the two regressions. The first model explains 11 percent of the variation, while this second model explains 9 percent. For that reason, the first model provides a better fit.*

2. *[10 points]* Economics major Charles (who has never taken EC200) is studying the relationship between minimum wages and unemployment for his EC011 project. He finds a cool chart on the internet (see chart) and concludes that minimum wages do not affect youth unemployment.



Charles




Charles's Chart

The line on Charles's chart represents a the results of an OLS regression, where $x$ is minimum wage and $f(x) = y =$ youth unemployment rate.

(a) Suppose that the t-statistic for $\widehat{\beta}$, the coefficient on minimum wages, is 1.24. Construct a 95% confidence interval for $\beta_1$, the slope coefficient on minimum wages.
*[3 points]*

- *First, recall that* $t = \dfrac{\widehat{\beta}}{\widehat{se(\beta)}}$. *Rearranging for the standard error, we have that* $se(\widehat{\beta}) = \dfrac{\widehat{\beta}}{t}$
- Therefore, $se(\widehat{\beta}) = \dfrac{-0.65}{1.24} = -0.524$
- Plug into formula for confidence interval, $\widehat{\beta} \pm z_{\alpha/2} se(\widehat{\beta}) = -0.65 \pm 1.96(0.524) = -0.65 \pm 1.027$
- Confidence interval is $[0.377, -1.677]$

(b) Give two specific examples to help Charles understand why this chart may not demonstrate a *causal* relationship between minimum wages and youth unemployment. *[4 points]*

> *Potential answers include the following 1. There may be omitted factors that affect both minimum wages and youth unemployment - for example, countries with high minimum wages may have more liberal governments, which in turn might impose other policies (like high social transfers) that keep youth unemployment high. Or, countries with a relatively large youth population may experience higher youth unemployment, and those youth may put pressure on the government for high minimum wages.*
>
> *2. High youth unemployment may put pressure on governments to keep minimum wages low, so that there may be an actual (negative) causal relationship between minimum wages and youth unemployment, but it's confounded by this reverse causality.*

(c) You want to help Charles, and you have data on the inflation rate ($inflation$), mean annual temperature ($temp$), and the share of the population under 25 ($pop25$). Should you include all three in the population model to reduce omitted variable bias? Explain your reasoning and any assumptions you make in order to draw your conclusions. *[3 points]*

> *Answers should discuss whether factors are correlated with minimum wages and correlated with unemployment. One possible response is that temperature may affect unemployment, but it's unlikely to be correlated with minimum wages, so it won't help with omitted variable bias. If it's not related at all, it will only increase the standard errors of our estimates. However, it's likely that the other two may be related to both unemployment and minimum wages, and if that's the case, then including those would reduce OVB.*

3. *[6 points]* Consider the following population model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$. Suppose that $E[u|x_1, x_2, x_3] = 0$ and $Var(u|x_1, x_2, x_3) = \sigma^2_{x_2}$ (that is, the variance of the error term depends on $x_2$.)

   (a) Given this information, can OLS estimates of $\widehat{\beta}_2$ be BLUE? If yes, list any additional assumptions you would need. If no, explain why not. *[2 points]*

   *Because $Var(u|x_1, x_2, x_3) = \sigma^2_{x_2}$, there is heteroskedasticity in the model, which means OLS cannot be blue (violates the fifth Gauss-Markov assumption of homoskedasticity).*

   (b) Given this information, can OLS estimates of $\widehat{\beta}_2$ be unbiased? If yes, list any additional assumptions you would need. If no, explain why not. *[2 points]*

   *Homoskedasticity is not required to demonstrate that $\widehat{\beta}_2$ is unbiased. Because $E[u|x_1, x_2, x_3] = 0$, the zero conditional mean assumption already holds, and we know the population model is linear from the question. For OLS estimates to be unbiased, we need to assume (1) random sampling and (2) no multicollinearity.*

   (c) Given this information, can OLS estimates of the *variance* of $\widehat{\beta}_2$ be unbiased? If yes, list any additional assumptions you would need. If no, explain why not. *[2 points]*

   *No. Our OLS estimates of the variance are unbiased only with homoskedasticity, which we do not have.*