Review of Midterm Concepts, Indicator Variables

Due in class 3/22/07

This problem set is in two parts. Please have two parts to turn in next week. I will grade Part 1 and Amy will grade Part 2.

Part 1: In case you missed it…
(This is less exciting than the Washington Post’s recap of American Idol…)
Some of the concepts from your midterm exam are very important, and I want to ensure that you understand them before we get too far in the second half of class. In this section, I am asking you to correct everything you got wrong on your midterm. For certain questions missed by a bunch of you, I am asking you to do some additional questions.

1. Go through your midterm exam. Write up correct answers for any question you got wrong, using your course notes. (Remember I have office hours Tuesday at 6:30 and check my email regularly. If you do not understand something, feel free to meet me or email me. The whole point of this exercise is to make sure we are all on the same page with the tools of the class—don’t be shy!)

The full output for every question in Part II and the formula sheet are available on blackboard—so you can check your work. You must show your work.

2. “Choose your own adventure…” Consult your midterm exam. If you had any points taken off of the following questions, answer these auxiliary questions.

   a. Part II, #6
      Many of you used output from the incorrect specifications. If you missed this question, write out the formula for the standard error. Annotate the formula with where, in general, you obtain the pieces of the formula. In addition, tell me which specification on your “Output for the Midterm” document contains that information.

   b. Part II, #7
      Many of you confused “variance” with “variation” in interpreting R-square. Write the definition of variance. Is “variance” one specific measure of variation?

   c. Part II, #8.d
      Many of you used output from the incorrect specifications. If you missed this question, write out the formula for the F-statistic. Annotate the formula with where, in general, you obtain the pieces of the formula. In addition, tell me which specification on your “Output for the Midterm” document contains that information (e.g., R-square from Restricted Model=Specification Y).
d. True/False, #4
The answer to this question will be more clear if you think about the
definitions of the following three items.
Write the definition of the error term. Write the definition of the residual (the
loose definition we use in class is OK). Write the definition of the sum of
squared residuals. Write one sentence about how these three things differ.

Part II: Exercises with Indicator Variables

A. Wooldridge 7.2
The following equations were estimated using the data in BWGHT.RAW:
Dependent Variable=L(Bwght)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.66</td>
<td>4.65</td>
</tr>
<tr>
<td></td>
<td>(.22)</td>
<td>(.38)</td>
</tr>
<tr>
<td>Cigs</td>
<td>-0.0044</td>
<td>-0.0052</td>
</tr>
<tr>
<td></td>
<td>(.0009)</td>
<td>(.0010)</td>
</tr>
<tr>
<td>Log(Faminc)</td>
<td>0.0093</td>
<td>0.0110</td>
</tr>
<tr>
<td></td>
<td>(.0059)</td>
<td>(.0085)</td>
</tr>
<tr>
<td>Parity</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.006)</td>
</tr>
<tr>
<td>Male</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(.010)</td>
<td>(.011)</td>
</tr>
<tr>
<td>White</td>
<td>0.055</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.015)</td>
</tr>
<tr>
<td>Mother’s Education</td>
<td></td>
<td>-0.0030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0030)</td>
</tr>
<tr>
<td>Father’s Education</td>
<td></td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0026)</td>
</tr>
<tr>
<td>N</td>
<td>1,388</td>
<td>1,191</td>
</tr>
<tr>
<td>R-square</td>
<td>0.0472</td>
<td>0.0493</td>
</tr>
</tbody>
</table>

The variables are defined as in Example 4.9, but we have added a dummy variable for
whether the child is male and a dummy variable indicating whether the child is
classified as white.

(i) In the first equation, interpret the coefficient on the variable cigs. In
particular, what is the effect on birth weight from smoking 10 more
cigarettes per day?

(ii) How much more is a white child predicted to weigh than a nonwhite child,
holding the other factors in the first equation fixed? Is the difference
statistically significant?

(iii) Comment on the estimated effect and statistical significance of motheduc.

(iv) From the given information, why are you unable to compute the F statistic
for joint significance of motheduc and fatheduc? What would you have to
do to compute the F statistic?
B. Wooldridge 7.5
In example 7.2, let \( noPC \) be a dummy variable equal to one of the student does not own a PC, and zero otherwise.

(i) If \( noPC \) is used in place of \( PC \) in equation (7.6), what happens to the intercept in the estimated equation? What will be the coefficient on \( noPC \)? (Hint: Write \( PC=1-noPC \) and plug this into the equation

\[
\text{colGPA} = \hat{\beta}_0 + \hat{\delta}_1PC + \hat{\beta}_2hsGPA + \hat{\beta}_3ACT
\]

(ii) What will happen to the R-squared if \( noPC \) is used in place of \( PC \)?

(iii) Should \( PC \) and \( noPC \) both be included as independent variables in the model? Explain.

C. (Use Wooldridge dataset wage2 for the following question)
1) Create an indicator variable that distinguishes those with less than a high school education (i.e., less than 12 years) from those with at least a high school education and respond to a-d below.

   a. How many observations and what proportion of the sample have less than 12 years of education?
   b. Run a t-test in SAS to determine whether average wages differ between those with less than 12 years of education and those with more education. Interpret the result.
   c. Run a SLR of wages on the indicator variable you created. Compare the result you obtain to that in b above, including a brief explanation for any similarities/differences.
   d. Re-run the model you ran in c above, but replace \( wage \) with \( lwage \). Interpret the coefficient for the indicator variable in this model.