Intro to Econometric Theory
Heinz School, Carnegie Mellon University
90-906, Spring 2003-4

Solution #4

The Medical Expenditure Panel Survey is an annual survey which collects information about medical expenditures, income, employment, demographics, health information, &c for a representative sample of Americans.

I have prepared an extract of these data for 1996, and it is available on the course website. The following are the columns in the data, in order:

Variable	Meaning
age	age of person in years
sex	sex of person, 1=male & 0=female
income	income in 1996 $
employed	1=employed, 0=not employed
insured	1=had health insurance, 0=not
health	perceived health status, higher is sicker
spending	spending on health care, 1996 $

To begin with, let's consider a model like the one we used on the midterm:

$$\begin{aligned} \text{spending}_i &= \beta_1 + \beta_2 \text{income} + \beta_3 \text{age} + \beta_4 \text{sex} \\ &+ \beta_4 \text{employed} + \beta_5 \text{insured} + \beta_6 \text{health} \end{aligned} \tag{1}$$

1. What do you think of the claim that income and sex do not belong in this model?

   To test this, we must test the null hypothesis that income and sex do not belong in the model, that is that both their coefficients are zero. We will conduct an F-test comparing the equation above to the restricted model with coefficients on sex and income constrained to be zero.

   The unrestricted model appears in the SAS output on page 3. The restricted model appears in the SAS output on page 5.

$$\begin{aligned} \text{F} - \text{stat} &= \frac{(5.6944 \times 10^{11} - 5.6932 \times 10^{11})/2}{(5.6932 \times 10^{11})/9504} \\ &= 1.00 \end{aligned}$$

Since an F-stat of 1 does not lead to rejection at a conventional significance level, we accept the null hypothesis and conclude that we do not have compelling evidence in these data to reject the theory that sex and income do not affect health care spending, once the other variables in the regression are controlled for.

Notice, on page 4 of the output, I use SAS's test command to perform a Wald test of the same null hypothesis, leading to the same conclusion (as we discussed in class).

2. Consider the health status variable. Respondents were asked to rate their health status; their choices were excellent, very good, good, fair, or poor. These were assigned the numerical values 1-5. Does it make sense to enter health status as a single continuous variable as in equation 1?

   Enter health status into the model as a set of dummies, and then test whether they belong.

   No, it does not make a lot of sense to enter this variable linearly. The underlying variable is ordinal (the order 1-5 has meaning) but not cardinal (the distance between the responses does not have meaning). Entering the variable linearly amounts to the assumption that the difference between spending between a poor & fair person is the same as the difference in spending between a fair & good person is the same as the difference in spending between a good & a very good person is the same as the difference in spending between a very good & excellent person, all controlling for the other variables in the model. There is no *a priori* reason to believe that this is true.

   Whether the health status variables do not belong can be tested by comparing (with an F-test) the regressions on output pages 6 and 9 or by examining the Wald test on page 7. The Wald-like test produces an F-stat of 33, which clearly rejects the null hypothesis that the health status variables do not belong. So, we conclude that we are very confident that they do belong.

3. How much more do people in very good health status spend than do people in excellent health status (estimate and CI).

   This question can be answered by looking at the coefficient on hsvryg in the regression on page 6. Since excellent health is the omitted contrast, the coefficient on hsvryg in the model estimated on this page is the difference in spending between people in very good and excellent health status, controlling for the other variables in the equation.

   Estimate and 95% CI:

$$\begin{aligned}\beta_{\text{hsvryg}} &= 286 \pm (1.96)202 \\ &= 286 \pm 396\end{aligned}$$

4. Test whether it was correct to enter health status linearly.

   As discussed above, entering health status linearly means assuming that the differences in spending between each adjacent health category (poor vs fair, fair vs good, etc) are equal. The null hypothesis that this is true can be written as a set of linear restrictions:

$$\begin{aligned}\beta_{\text{vryg}} &= \beta_{\text{good}} - \beta_{\text{vryg}} \\ \beta_{\text{good}} - \beta_{\text{vryg}} &= \beta_{\text{fair}} - \beta_{\text{good}} \\ \beta_{\text{good}} - \beta_{\text{vryg}} &= \beta_{\text{poor}} - \beta_{\text{fair}}\end{aligned}$$

   We can write these another way (substituting in the above):

$$\begin{aligned}2\beta_{\text{vryg}} &= \beta_{\text{good}} \\ 3\beta_{\text{vryg}} &= \beta_{\text{fair}} \\ 4\beta_{\text{vryg}} &= \beta_{\text{poor}}\end{aligned}$$

   There are a few ways to test this. For example, we can run the regression reported on output page 6 and examine the Wald-like F-test on page 8. Or, we can compare the regression on page 6 with the constrained regression on page 10. Or, we can compare the regression on page 6 with the constrained regression on page 3 — Notice that the sum of squared errors and many param estimates are identical for the regressions on pages 3 and 10: you should think about why this is so. I'll do the F-test comparing pages 6 and 10.

$$\begin{aligned}\text{F} - \text{stat} &= \frac{(5.6932 \times 10^{11} - 5.6724 \times 10^{11})/3}{(5.6724 \times 10^{1}1)/9497} \\ &= 11.61\end{aligned}$$

   This leads to rejection at any conventional significance level. We conclude that we are very confident that it is wrong to enter health status linearly here.

5. Test whether insurance affects spending for people of different health statuses differently and discuss.

   To get insurance to affect spending differently for people of different health statuses, we need to interact health status with insurance. This is done in the model estimated on output page 11. It looks as if insurance increases the health spending of sick people more than it does the health spending of well people. To test this, we can compare the regressions on pages 11 and 6 with an F-test or we can look at the Wald-like F-test on page 12. Looking at page 12, we can reject the null of no interactions at better than the 1% level, so we conclude that these interactions belong in the regression.