Lecture 5b: Practice Problem Solutions

John McGready
Johns Hopkins University

Hourly Wages

1. Recall the MLR relating hourly wages to years of formal education and worker's sex? (1 = female, 0 = male). Years of education in this sample ranged from 2 to 18 years.

```
regress wage edlevel sex

      Source |       SS       df       MS              Number of obs =     534
-------------+------------------------------           F(  2,   531) =   61.62
       Model |  2651.49936     2  1325.74968           Prob > F      =  0.0000
    Residual |  11425.1992   531  21.5163827           R-squared     =  0.1884
-------------+------------------------------           Adj R-squared =  0.1853
       Total |  14076.6985   533  26.4103162           Root MSE      =  4.6386

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     edlevel |   .7512834   .0768225     9.78   0.000     .6003701    .9021967
         sex |  -2.124057   .4028322    -5.27   0.000    -2.915397   -1.332716
       _cons |   .2178312   1.036322     0.21   0.834    -1.817962    2.253624
------------------------------------------------------------------------------
```


Hourly Wages

1. Recall the MLR relating hourly wages to years of formal education and worker's sex? (1 = female, 0 = male). Years of education in this sample ranged from 2 to 18 years.
   a) What does the model estimate for the mean difference in hourly wages for male workers with 16 years of education compared to male workers with 12 years of education?

   – *The slope estimate for years of education,* $\hat{\beta}_1$ *(edlevel), quantifies the sex adjusted relationship between wages and years of education, per one year increment. The comparison here is for a four-year difference in years of education, so the resulting mean difference in wages estimate is* $4\hat{a}_1 = 4 \times 0.75 = \$3.00/hr.$

Hourly Wages

1. Recall the MLR relating hourly wages to years of formal education and worker's sex? (1 = female, 0 = male). Years of education in this sample ranged from 2 to 18 years.
   b) Compute a 95% CI for the true mean difference in hourly wages for the same comparison in part a.

   – *Recall, the resulting estimated mean difference is given by* $4\hat{\beta}_1$. *A 95% for 4* β_1 *is* $4\hat{\beta}_1 \pm 2\hat{SE}(4\hat{\beta}_1) \rightarrow 4\hat{\beta}_1 \pm 2 \times 4\hat{SE}(\hat{\beta}_1)$. *Notice this is equivalent to* $4\left(\hat{\beta}_1 \pm 2\hat{SE}(\hat{\beta}_1\right)$.*, which can be easily obtained by multiplying the endpoints from the 95% CI for* β_1 *given in the Stata output: (4*0.60, 4*0.90) gives a 95% CI of ($2.40/hr, $3.60/hr).*

Hourly Wages

1. Recall the MLR relating hourly wages to years of formal education and worker's sex? (1 = female, 0 = male). Years of education in this sample ranged from 2 to 18 years.
   c) What does the model estimate for the mean difference in hourly wages for female workers with 16 years of education compared to female workers with 12 years of education?

   – *There are no extra computations needed; this is exactly the same as the answers to part a.*

Hourly Wages

d) What does the model estimate for the mean difference in hourly wages for *female* workers with 16 years of education compared to *male* workers with 12 years of education?

– *Notice that this cannot be answered using just one slope from the regression. Let's write out what the equation predictions for both of the groups we are considering.*

– *F, 16 years of education:* $\hat{y} = \hat{\beta}_o + \hat{\beta}_1 \times 16 + \hat{\beta}_2 \times 1$

– *M, 12 years of education:* $\hat{y} = \hat{\beta}_o + \hat{\beta}_1 \times 12$

– *The difference in these two estimates is as follows:*

$$\hat{\beta}_1 \times 4 + \hat{\beta}_2 = 0.75 \times 4 + -2.12 = \$0.88/hr$$

Hemoglobin

2. Recall the MLR relating Hb to PCV and age. In this sample, PCV ranges from 25% to 55%; age ranges from 20 years to 67 years.

```
. regress Hb PCV age

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  2,    18) =   48.26
       Model |  88.7923124     2  44.3961562           Prob > F      =  0.0000
    Residual |  16.5591129    18  .919950714           R-squared     =  0.8428
-------------+------------------------------           Adj R-squared =  0.8254
       Total |  105.351425    20  5.26757126           Root MSE      =  .95914

------------------------------------------------------------------------------
          Hb |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         PCV |   .1023427   .0312317     3.28   0.004     .0367274     .167958
         age |   .1013414   .0164271     6.17   0.000     .0668293    .1358534
       _cons |   5.516596   1.114866     4.95   0.000     3.174349    7.858842
------------------------------------------------------------------------------
```


Hourly Wages

2. Recall the MLR relative Hb to PCV and age. In this sample, PCV ranges from 25% to 55%; age ranges from 20 years to 67 years.
   a) What does the model estimate for the mean difference in Hb for 60 year old subjects with PCV of 40% compared to 60 year old subjects with PCV of 32%?

   – *The slope estimate for packed cell volume (PCV),* $\hat{\beta}_1$ *, quantifies the age adjusted relationship between Hb and PCV per 1% increment. The comparison here is for an 8-percent difference in PCV among persons of the same age (60), so the resulting mean difference in Hb estimate is as follows:*

$$8\hat{a}_1 = 8 \times 0.10 = 0.80g/dL.$$

Hourly Wages

2. Recall the MLR relative Hb to PCV and age. In this sample, PCV ranges from 25% to 55%; age ranges from 20 years to 67 years.
   b) Compute a 95% CI for the true mean difference in Hb levels for the same comparison in part a.

   – *Recall, the resulting estimated mean difference is given by* $8\hat{\beta}_1$. *A 95% for 8* β_1 *is* $8\hat{\beta}_1 \pm 2S\hat{E}(8\hat{\beta}_1) \rightarrow 8\hat{\beta}_1 \pm 8 \times 2S\hat{E}(\hat{\beta}_1)$. *Notice this is equivalent to* $8\left(\hat{\beta}_1 \pm 2S\hat{E}(\hat{\beta}_1)\right)$, *which can be easily obtained by multiplying the endpoints from the 95% CI for* β_1 *given in the Stata output: (8*0.04, 8*0.17) gives a 95% CI of 0.32g/dL, 1.36 g/dL).*

Hourly Wages

c) What does the model estimate for the mean difference in Hb for 60 year old subjects with PCV of 40% compared to 50 year old subjects with PCV of 32%?

– *Notice that this cannot be answered using just one slope from the regression. Let's write out what the equation predictions for both of the groups we are considering.*

– *60 year olds, PCV = 40%:* $\hat{y} = \hat{\beta}_o + \hat{\beta}_1 \times 40 + \hat{\beta}_2 \times 60$

– *50 year olds, PCV = 32%:* $\hat{y} = \hat{\beta}_o + \hat{\beta}_1 \times 32 + \hat{\beta}_2 \times 50$

– *The difference in these two estimates is as follows:*

$$8\hat{\beta}_1 + 10\hat{\beta}_2 = 0.10 \times 8 + 0.10 \times 10 = 1.80g/dL$$