Mathematics 231

Lecture 31

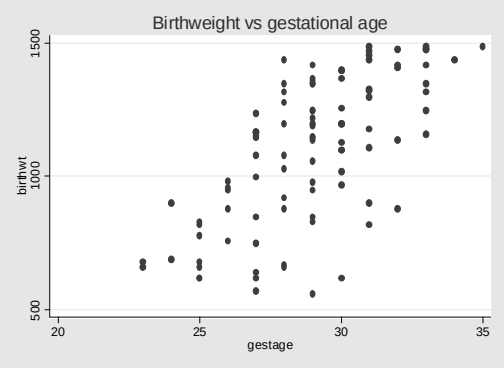
Liam O'Brien

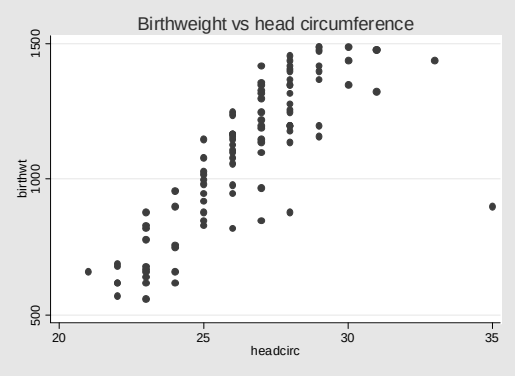
Announcements

- Reading
  - Today M&M 11

Example: Birthweights

- Consider the birthweight data that we used in class before.
- We predicted birthweight with infant length before, but now also consider two new variables.
- Data was also obtained on the presence of toxemia, gestational age, head circumference, and mother's age.
- We next examine the predictive power of head circumference and gestational age on birthweight.





Comparison of Predictors

- We see a positive relationship between both predictors and the response.
- This is common and we often look at the correlations among all variables (response and predictors).
- Highly correlated predictors often have overlapping predictive information about the response.

Example: Birthweights

```
. pwcorr birthwt gestage headcirc

             |  birthwt  gestage headcirc
-------------+---------------------------
     birthwt |   1.0000
     gestage |   0.6599   1.0000
    headcirc |   0.7988   0.7807   1.0000
```


Example: Birthweights

```
. regress birthwt gestage

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =   75.61
       Model |  3143019.07     1  3143019.07           Prob > F      =  0.0000
    Residual |  4073723.68    98   41568.609           R-squared     =  0.4355
-------------+------------------------------           Adj R-squared =  0.4298
       Total |  7216742.75    99  72896.3914           Root MSE      =  203.88

------------------------------------------------------------------------------
     birthwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     gestage |   70.30993   8.085854     8.70   0.000     54.26382    86.35604
       _cons |  -932.4039   234.4884    -3.98   0.000    -1397.738   -467.0693
------------------------------------------------------------------------------
```


Example: Birthweights

```
. regress birthwt headcirc

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =  172.82
       Model |  4605298.87     1  4605298.87           Prob > F      =  0.0000
    Residual |  2611443.88    98  26647.3866           R-squared     =  0.6381
-------------+------------------------------           Adj R-squared =  0.6344
       Total |  7216742.75    99  72896.3914           Root MSE      =  163.24

------------------------------------------------------------------------------
     birthwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    headcirc |   85.17802   6.479268    13.15   0.000     72.32013    98.03592
       _cons |  -1154.109   172.1523    -6.70   0.000    -1495.739    -812.478
------------------------------------------------------------------------------
```


Example: Birthweights

```
. regress birthwt gestage headcirc

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) =   86.79
       Model |  4629638.61     2  2314819.31           Prob > F      =  0.0000
    Residual |  2587104.14    97  26671.1767           R-squared     =  0.6415
-------------+------------------------------           Adj R-squared =  0.6341
       Total |  7216742.75    99  72896.3914           Root MSE      =  163.31

------------------------------------------------------------------------------
     birthwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     gestage |    9.90101   10.36436     0.96   0.342    -10.66938     30.4714
    headcirc |   77.44205   10.37285     7.47   0.000     56.85482    98.02929
       _cons |  -1235.533    192.166    -6.43   0.000    -1616.929   -854.1362
------------------------------------------------------------------------------
```


Example: Birthweights

- The usefulness of gestational age as a predictor of birthweight has now become statistically insignificant (at the 0.05 level).
- This is happening because both gestational age and head circumference provide overlapping information on birthweights.
- Notice what happened to R-squared – it increased from 0.63 to 0.64 with the addition of head circumference to the model.
- What about the interpretation of the coefficients?

R^2 and Adjusted R^2

- The R^2 value will always increase with the addition of a new predictor.
- Having too many predictors is BAD, so we penalize the R^2 value based on the number of predictors in the model – this is the adjusted R^2.
- Note that adjusted R^2 decreased slightly.
- This indicates that adding head circumference did not help us predict birthweight much.
- Note that the adjusted R^2 does not have the same interpretation as R^2.

R^2 and Adjusted R^2

- Use of the adjusted R^2 in evaluating regression model is highly controversial.
- It is safe to say, however, that too many predictors (i.e., the kitchen sink model) is NOT a good thing.
- We only have a finite amount of information and want to put as much as possible to work in estimating the error.
- However, we lose a degree of freedom for each predictor.

13

Examining Correlations

```
. pwcorr birthwt gestage length headcirc momage

             |  birthwt  gestage   length headcirc   momage
-------------+---------------------------------------------
     birthwt |   1.0000
     gestage |   0.6599   1.0000
      length |   0.8156   0.6752   1.0000
   headcirc |   0.7988   0.7807   0.7127   1.0000
     momage |   0.1546   0.2658   0.2180   0.1321   1.0000
```

14

Example: Birthweights

```
. regress birthwt toxemia

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =    0.01
       Model |  1005.40823     1  1005.40823           Prob > F      =  0.9072
    Residual |  7215737.34    98  73629.9729           R-squared     =  0.0001
-------------+------------------------------           Adj R-squared = -0.0101
       Total |  7216742.75    99  72896.3914           Root MSE      =  271.35

------------------------------------------------------------------------------
     birthwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     toxemia |    7.78481   66.61992     0.12   0.907    -124.4203    139.9899
       _cons |   1097.215   30.52908    35.94   0.000     1036.631    1157.799
------------------------------------------------------------------------------
```

15

Example: Birthweights

```
. regress birthwt gestage toxemia

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) =   51.91
       Model |  3730907.03     2  1865453.51           Prob > F      =  0.0000
    Residual |  3485835.72    97  35936.4507           R-squared     =  0.5170
-------------+------------------------------           Adj R-squared =  0.5070
       Total |  7216742.75    99  72896.3914           Root MSE      =  189.57

------------------------------------------------------------------------------
     birthwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     gestage |   84.05796   8.250832    10.19   0.000     67.68234    100.4336
     toxemia |  -206.5908   51.07771    -4.04   0.000     -307.966   -105.2157
       _cons |   -1286.2   234.9178    -5.48   0.000    -1752.447   -819.9535
------------------------------------------------------------------------------
```

16

Example

- The regression equation is given by:

  E(birthwt)=-1286+84.1(gestage)-206.6(toxemia)
- The variable "toxemia" is an indicator variable (or dummy variable) that indicates presence of toxemia.
- For those with toxemia:

  E(birthwt)=-1492+84.1(gestage)
- For those without toxemia:

  E(birthwt)=-1286+84.1(gestage)

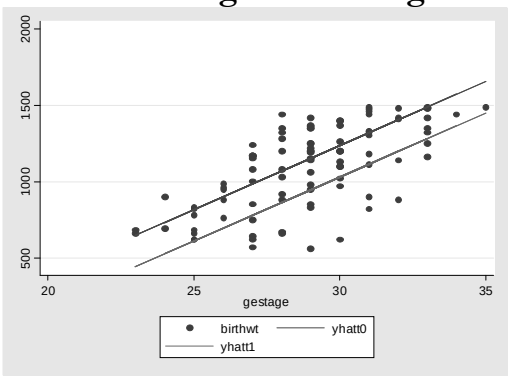
17

Interpreting Coefficients

- Note that the interpretation of any one coefficient is conditional on holding all the other coefficients constant.
- In the case of a binary predictor, we can create two "separate" regression lines based on setting the indicator to 0 or 1.

18

Birthweight vs Gestage



19