Economics 312 — Spring 2014
Daily Problem #17 — February 27

We have used the instructor's Econ 201 grade data several times in class demonstrations. Here, we consider how we might learn about the shape of a functional relationship between grade and reader rating by using a set of dummy variables.

Recall that the (inverse) reader rating is a rating of academic promise at the time of admission assigned by the Reed admission office. The dependent variable is gpoints, defined as grade points earned in Econ 201, on a scale of 0 (F) to 4 (A or A+).

A linear regression of gpoints on irdr yields

```
. reg gpoints irdr

      Source |       SS       df       MS              Number of obs =     839
-------------+------------------------------           F(  1,   837) =   74.64
       Model |  49.0522912     1  49.0522912           Prob > F      =  0.0000
    Residual |   550.04747   837  .657165435           R-squared     =  0.0819
-------------+------------------------------           Adj R-squared =  0.0808
       Total |  599.099761   838  .714916182           Root MSE      =  .81066

------------------------------------------------------------------------------
     gpoints |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        irdr |   .4354099   .0503971     8.64   0.000     .3364902    .5343295
       _cons |   1.386878   .1701852     8.15   0.000     1.052838    1.720918
------------------------------------------------------------------------------
```

However, I am not convinced that the effect of reader rating on grade should be linear, or indeed that it should follow any simple parametric form. To allow the "response surface" (the graph of the predicted value of gpoints against irdr) to be determined flexible, I created four dummy variables with the following commands:

```
. g irdr_45=irdr>4.25

. g irdr_40 = irdr>3.75 & irdr<=4.25

. g irdr_35=irdr>3.25 & irdr<=3.75

. g irdr_30=irdr>2.75 & irdr<=3.25
```

The summary statistics for irdr and the dummies are

```
. summarize irdr irdr_30 irdr_35 irdr_40 irdr_45

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        irdr |       865    3.331075    .5547127        1.5          5
```

```
    irdr_30 |       865    .3560694    .4791131          0          1
    irdr_35 |       865     .265896    .4420647          0          1
    irdr_40 |       865    .1676301    .3737535          0          1
    irdr_45 |       865    .0358382     .185994          0          1
```

Regressing grade points on the set of dummies yields

```
. reg gpoints irdr_30 irdr_35 irdr_40 irdr_45

      Source |       SS       df       MS              Number of obs =     839
-------------+------------------------------           F(  4,   834) =   17.79
       Model |  47.1034509     4  11.7758627           Prob > F      =  0.0000
    Residual |   551.99631   834  .661866079           R-squared     =  0.0786
-------------+------------------------------           Adj R-squared =  0.0742
       Total |  599.099761   838  .714916182           Root MSE      =  .81355

------------------------------------------------------------------------------
     gpoints |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     irdr_30 |    .259464   .0818113     3.17   0.002     .0988837    .4200443
     irdr_35 |   .4137226   .0862559     4.80   0.000     .2444184    .5830267
     irdr_40 |   .6561783   .0960921     6.83   0.000     .4675675    .8447891
     irdr_45 |   .9875545   .1606941     6.15   0.000     .6721421    1.302967
       _cons |   2.489865   .0668735    37.23   0.000     2.358605    2.621125
------------------------------------------------------------------------------
```

1. Based on the summary statistics, what proportion of Econ 201 students fall into the inverse reader rating categories >4.25, 3.75–4.25, 3.25–3.75, 2.75–3.25, <2.75?

2. What, precisely, does each of the dummy-variable coefficients in the regression table tell you?

3. If the relationship between irdr and gpoints was linear, what would you expect from the dummy coefficients?

4. Graph (crudely) the estimated response surface for grade as a function of irdr implied by the coefficients on the dummies as estimated in the regression. Although the function is formally a step function, you can plot each interval at its midpoint, using 4.5 for the top interval and 2.5 for the bottom (omitted) one.

5. Based on your graph, does a linear relationship seem plausible? If not, what other functional form might be better?

6. What set of linear restrictions on the coefficients of the dummies could be tested to see if a linear relationship is rejected?

7. What are the advantages and disadvantages of this approach compared with, for example, including a quadratic term in the relationship?