PROJECT TWO: STATA PROBLEM SET ANSWERS
Professor Lemke
Fall 2008

```
. * PROBLEM ONE;


Contains data from wi_data.dta
  obs:           330
 vars:             9                          5 Feb 2005 20:10
 size:        11,880 (99.9% of memory free)
-------------------------------------------------------------------------------
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
bamin           int    %8.0g                  Minimum Salary for teacher w BA
bamax           long   %12.0g                 Maximum Salary for teacher w BA
mamin           long   %12.0g                 Minimum Salary for teacher w MA
mamax           long   %12.0g                 Maximum Salary for teacher w MA
enroll          long   %12.0g                 District Enrollment
lunch           long   %12.0g                 Percent Studs Rec Free Lunch
dropout         int    %8.0g                  Percent Students Dropout
act             float  %9.0g                  District Average ACT Score
totalexp        long   %12.0g                 Total District Expenditures
-------------------------------------------------------------------------------


-------------+--------------------------------------------------------
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       bamin |       330    25719.98    1301.497      22293      30815
       bamax |       330    36851.16     3648.96      27572      48476
       mamin |       330    28724.04    1746.136      24393      36032
       mamax |       330    44219.98    4048.267      32975      57209
      enroll |       330     2436.37    6150.913        204      99814
       lunch |       330    622.5879     3954.44         21      70928
     dropout |       330    19.34545    159.9305          0       2862
         act |       330    22.01364    1.012278       17.7       25.2
    totalexp |       330    1.82e+07    5.14e+07    1818099   8.53e+08
-------------+--------------------------------------------------------
```

```
. * PROBLEM TWO;

       size |      Freq.     Percent        Cum.
------------+-----------------------------------
      Small |         99       30.00       30.00
     Medium |        132       40.00       70.00
      Large |         99       30.00      100.00
------------+-----------------------------------
      Total |        330      100.00


-------------------------------------------------------------------------------
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
size            float  %9.0g       sizes      3 size categories for districts
small           float  %9.0g                  district <= 30th pctile in size
medium          float  %9.0g                  district bw 30th & 70th pctile
large           float  %9.0g                  district >= 70th pctile in size
-------------------------------------------------------------------------------


-------------+---------------------------------------------------------
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        size |       330           2     .775773          1          3
       small |       330          .3    .4589535          0          1
      medium |       330          .4    .4906419          0          1
       large |       330          .3    .4589535          0          1
-------------+---------------------------------------------------------
```

```
. * PROBLEM THREE;
```

variable name	type	format	label	variable label
bamin	int	%8.0g		Minimum Salary for teacher w BA
baminz	float	%9.0g		bamin in $1,000
bamax	long	%12.0g		Maximum Salary for teacher w BA
bamaxz	float	%9.0g		bamax in $1,000
mamin	long	%12.0g		Minimum Salary for teacher w MA
maminz	float	%9.0g		mamin in $1,000
mamax	long	%12.0g		Maximum Salary for teacher w MA
mamaxz	float	%9.0g		mamax in $1,000
ppexp	float	%9.0g		District Per Pupil Expenditures
ppexpz	float	%9.0g		ppexp in $1,000

Variable	Obs	Mean	Std. Dev.	Min	Max
bamin	330	25719.98	1301.497	22293	30815
baminz	330	25.71998	1.301497	22.293	30.815
bamax	330	36851.16	3648.96	27572	48476
bamaxz	330	36.85116	3.64896	27.572	48.476
mamin	330	28724.04	1746.136	24393	36032
maminz	330	28.72404	1.746136	24.393	36.032
mamax	330	44219.98	4048.267	32975	57209
mamaxz	330	44.21998	4.048267	32.975	57.209
ppexp	330	7377.191	875.6122	5807.12	12535.09
ppexpz	330	7.377191	.8756122	5.80712	12.53509

```
. * PROBLEM FOUR;


. reg act ppexpz dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  3,   326) =   11.39
       Model |  31.9967095     3  10.6655698           Prob > F      =  0.0000
    Residual |  305.131947   326  .935987567           R-squared     =  0.0949
-------------+------------------------------           Adj R-squared =  0.0866
       Total |  337.128656   329  1.02470716           Root MSE      =  .96746

------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ppexpz |   .0072515   .0642602     0.11   0.910    -.1191656    .1336685
     dropout |  -.1634897   .1551833    -1.05   0.293    -.4687767    .1417973
       lunch |  -.0233489   .0046311    -5.04   0.000    -.0324596   -.0142383
       _cons |   22.52049   .4595995    49.00   0.000     21.61633    23.42464
------------------------------------------------------------------------------
```

STATISTICALLY SIGNIFICANT PREDICTORS

The only statistically significant predictors of ACT is lunch, which proxies for poverty, and it is statistically significant at even the 1% level.

INTERPRETATION OF COEFFICIENT ON PPEXPZ:

For every $1,000 more spent per pupil, the district’s average ACT score is expected to increase by 0.007 points. This is not a statistically significant effect, nor is it an economically significant effect. Notice from the summary statistics that ppexpz ranges from a minimum of over 5 to a maximum of under 13. Thus, ppexpz does not differ by more than 8 across any two districts, and consequently the greatest possible predicted difference in ACT scores is 8 x 0.007, which is less than 0.056 of a point. Thus, we are left with the conclusion that ACT scores do not depend on public school expenditures.

```
. * PROBLEM FIVE;

. reg act bamin mamax enroll dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  5,   324) =   10.81
       Model |  48.1908753     5  9.63817505           Prob > F      =  0.0000
    Residual |  288.937781   324  .891783275           R-squared     =  0.1429
-------------+------------------------------           Adj R-squared =  0.1297
       Total |  337.128656   329  1.02470716           Root MSE      =  .94434
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bamin |  -.0000487   .0000493    -0.99   0.324    -.0001456    .0000483
       mamax |   .0000662   .0000171     3.88   0.000     .0000326    .0000998
      enroll |  -2.25e-06     .00001    -0.22   0.822     -.000022    .0000174
     dropout |  -.2587043    .168657    -1.53   0.126    -.5905054    .0730968
       lunch |  -.0173153   .0045258    -3.83   0.000     -.026219   -.0084116
       _cons |   20.81044   1.071905    19.41   0.000     18.70166    22.91921
------------------------------------------------------------------------------
```

STATISTICALLY SIGNIFICANT PREDICTORS

The maximum salary paid to a teacher with a masters degree and lunch are the only two statistically significant variables.

```
. * PROBLEM SIX;

. reg act baminz mamaxz enroll dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  5,   324) =   10.81
       Model |  48.1908778     5  9.63817555           Prob > F      =  0.0000
    Residual |  288.937779   324  .891783267           R-squared     =  0.1429
-------------+------------------------------           Adj R-squared =  0.1297
       Total |  337.128656   329  1.02470716           Root MSE      =  .94434
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0486525   .0492948    -0.99   0.324    -.1456309    .0483259
      mamaxz |   .0662345     .01708     3.88   0.000     .0326329    .0998361
      enroll |  -2.25e-06     .00001    -0.22   0.822     -.000022    .0000174
     dropout |  -.2587044    .168657    -1.53   0.126    -.5905054    .0730967
       lunch |  -.0173153   .0045258    -3.83   0.000     -.026219   -.0084116
       _cons |   20.81044   1.071905    19.41   0.000     18.70166    22.91921
------------------------------------------------------------------------------
```

DISCUSSION CONCERNING RESCALING:

Absolutely nothing changes between the two sets of regression results in problems 5 and 6 including the r-squared and adjusted r-squared except that the coefficient estimates, standard errors, and confidence intervals on baminz and mamaxz in problem six are 1000 times greater than their counterparts on bamin and mamax in problem five. Notice that the *t*-statistics and *p*-values associated with these variables are the same across regression. This must be the case, as rescaling variables can have no meaningful effect on the results and interpretations.

```
. * PROBLEM SEVEN;


. reg act baminz mamaxz enroll enrollsq dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  6,   323) =   12.12
       Model |  61.9559048     6  10.3259841           Prob > F      =  0.0000
    Residual |  275.172751   323  .851928023           R-squared     =  0.1838
-------------+------------------------------           Adj R-squared =  0.1686
       Total |  337.128656   329  1.02470716           Root MSE      =    .923

------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0579365   .0482361    -1.20   0.231     -.152833      .03696
      mamaxz |   .0465065   .0174004     2.67   0.008      .012274     .080739
      enroll |    .000086    .000024     3.58   0.000     .0000387    .0001332
    enrollsq |  -1.02e-09   2.53e-10    -4.02   0.000    -1.51e-09   -5.19e-10
     dropout |  -.3146059   .1654307    -1.90   0.058    -.6400637    .0108519
       lunch |  -.0148302   .0044665    -3.32   0.001    -.0236173    -.006043
       _cons |   21.71985   1.071828    20.26   0.000      19.6112     23.8285
------------------------------------------------------------------------------
```

DISCUSSION OF THE SQUARED ENROLLMENT TERM:

As the coefficient estimates on enrollment squared is negative, the predicted shape of enrollment on ACT is an inverted U-shaped parabola. In particular, the predicted effect of enrollment on ACT scores is that ACT scores increase with enrollment but at a decreasing rate until it reaches a maximum at 0.000086 ÷ (2*1.02e-09) = 42,157. After this size of district, increased enrollment has a negative predicted effect on ACT score and increasingly so.

```
. * PROBLEM EIGHT;

. reg act baminz mamaxz small medium large dropout lunch;
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0561791   .0484584    -1.16   0.247    -.1515131    .0391548
      mamaxz |   .0376336   .0185962     2.02   0.044     .0010486    .0742186
       small |  -.5264869   .1741813    -3.02   0.003      -.86916   -.1838137
      medium |  -.5030562   .1381464    -3.64   0.000    -.7748366   -.2312758
       large |  (dropped)
     dropout |  -.4081188   .1551283    -2.63   0.009    -.7133082   -.1029294
       lunch |  -.0149708    .004566    -3.28   0.001    -.0239536   -.0059881
       _cons |   22.63049   1.142073    19.82   0.000     20.38365    24.87733
------------------------------------------------------------------------------

. reg act baminz mamaxz small        large dropout lunch;
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0561791   .0484584    -1.16   0.247    -.1515131    .0391548
      mamaxz |   .0376336   .0185962     2.02   0.044     .0010486    .0742186
       small |  -.0234307   .1355342    -0.17   0.863    -.2900718    .2432105
       large |   .5030562   .1381464     3.64   0.000     .2312758    .7748366
     dropout |  -.4081188   .1551283    -2.63   0.009    -.7133082   -.1029294
       lunch |  -.0149708    .004566    -3.28   0.001    -.0239536   -.0059881
       _cons |   22.12744   1.098737    20.14   0.000     19.96585    24.28902
------------------------------------------------------------------------------

DISCUSSION:

Compared to a comparable medium district, a small district is expected to score 0.0234
points lower on its average ACT score; however this effect is not statistically
significant. Similarly, a large district is expected to score 0.503 points higher on
its average ACT score compared to a comparable medium sized district; and this effect
is statistically significant.

Moreover, these results can be used to determine that a large district is expected to
score 0.5264 points higher on its average ACT score compared to a comparable small
district. This is found by subtracting -0.0234 from 0.503. This effect is also readily
apparent in the following regression, which omits small and includes medium in the
model.

. reg act baminz mamaxz       medium large dropout lunch;
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0561791   .0484584    -1.16   0.247    -.1515131    .0391548
      mamaxz |   .0376336   .0185962     2.02   0.044     .0010486    .0742186
      medium |   .0234307   .1355342     0.17   0.863    -.2432105    .2900718
       large |   .5264869   .1741813     3.02   0.003     .1838137      .86916
     dropout |  -.4081188   .1551283    -2.63   0.009    -.7133082   -.1029294
       lunch |  -.0149708    .004566    -3.28   0.001    -.0239536   -.0059881
       _cons |     22.104    1.08092    20.45   0.000     19.97747    24.23054
------------------------------------------------------------------------------
```

. * PROBLEM NINE;

ANSWER #1.

The best explanation is that we are statisticians, who value statistical significance. In particular, the estimated coefficient on bamin, which is consistently negative, is, more importantly, consistently statistically insignificant. Thus, the correct conclusion is that there is no effect of bamin on ACT. But to say that it has a negative effect would be wrong.

The question then becomes, why is it that bamin has no effect but mamax does matter? One possible answer is union power, which is not controlled for in the regression. In particular, consider that some districts have stronger unions than others. Moreover, unions want to give their bargained wage increases to their most entrenched members. This means giving more money at the higher-end of the salary schedule (i.e., to those teachers with Masters degrees). Next, if these expensive salary schedules are observed by teachers, teachers will have an incentive to pursue the most highly paying jobs (though they won't receive the high paying salaries for several years). Finally, if district administers can evaluate good vs. bad teaching, they will be able to fire the bad teachers and hire the good teachers as the lucrative pay schedule will be attractive to all teachers.

ANSWER #2.

Another (less attractive) explanation for why minimum BA salary is coming in negative and maximum MA salary is coming in positive is not that spending more on new teachers is unproductive and spending more on veteran teachers is productive; rather, the best explanation is probably model mis-specification. In particular, OLS doesn't produce great results when two or more variables are highly correlated. The problem is that the effect of bamin and mamax are so highly related, that OLS can produce wildly wrong guesses at the true effect of each. Put differently, OLS struggles to separately identify the true effect of each.)

```
. * PROBLEM TEN;


. reg act baminz medium large dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  5,   324) =   13.09
       Model |  56.6691734     5  11.3338347           Prob > F      =  0.0000
    Residual |  280.459483   324  .865615688           R-squared     =  0.1681
-------------+------------------------------           Adj R-squared =  0.1553
       Total |  337.128656   329  1.02470716           Root MSE      =  .93038
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0067322    .042047    -0.16   0.873    -.0894518    .0759873
      medium |   .1078645   .1295672     0.83   0.406    -.1470347    .3627637
       large |   .6996199   .1524466     4.59   0.000     .3997097    .9995302
     dropout |  -.3976226   .1557804    -2.55   0.011    -.7040914   -.0911538
       lunch |  -.0160799   .0045546    -3.53   0.000    -.0250401   -.0071196
       _cons |   22.43021   1.073927    20.89   0.000     20.31746    24.54296
------------------------------------------------------------------------------

. reg act mamaxz medium large dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  5,   324) =   13.75
       Model |  59.0283121     5  11.8056624           Prob > F      =  0.0000
    Residual |  278.100344   324  .858334395           R-squared     =  0.1751
-------------+------------------------------           Adj R-squared =  0.1624
       Total |  337.128656   329  1.02470716           Root MSE      =  .92646
------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mamaxz |   .0267631   .0160678     1.67   0.097    -.0048472    .0583735
      medium |   .0327229   .1353688     0.24   0.809    -.2335897    .2990356
       large |   .5271935   .1742727     3.03   0.003     .1843445    .8700424
     dropout |  -.3935447   .1547002    -2.54   0.011    -.6978883   -.0892011
       lunch |  -.0153746   .0045551    -3.38   0.001    -.0243359   -.0064134
       _cons |   21.13878   .6897213    30.65   0.000     19.78188    22.49568
------------------------------------------------------------------------------
```

DISCUSSION:

Even when bamin is included by itself, the effect is still negative and statistically insignificant. This is what makes the second answer to question 9 unappealing.

```
. * PROBLEM ELEVEN;


. reg act baminz payincz medium large dropout lunch;

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  6,   323) =   11.70
       Model |  60.1807212     6  10.0301202           Prob > F      =  0.0000
    Residual |  276.947935   323  .857423948           R-squared     =  0.1785
-------------+------------------------------           Adj R-squared =  0.1632
       Total |  337.128656   329  1.02470716           Root MSE      =  .92597

------------------------------------------------------------------------------
         act |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      baminz |  -.0185455   .0422527    -0.44   0.661    -.1016708    .0645798
     payincz |   .0376336   .0185962     2.02   0.044     .0010486    .0742186
      medium |   .0234307   .1355342     0.17   0.863    -.2432105    .2900718
       large |   .5264869   .1741813     3.02   0.003     .1838138      .86916
     dropout |  -.4081188   .1551283    -2.63   0.009    -.7133082   -.1029294
       lunch |  -.0149708    .004566    -3.28   0.001    -.0239536   -.0059881
       _cons |     22.104    1.08092    20.45   0.000     19.97747    24.23054
------------------------------------------------------------------------------
```

DISCUSSION:

Although there remains no evidence that starting salaries matter, there is evidence that annual pay increases are positively related to test scores. In particular, a district that has a $10,000 greater gap between its starting BA salary and maximum MA salary compared to a comparable district is expected to have an average ACT score that is 0.376 points higher. (Notice that as I used a change of $10,000, I need to multiply the coefficient on payincz by 10.)

If I were going to report these results in a paper, I would report the results using baminz and payincz, because the coefficient estimates are larger and therefore easier to decipher immediately. For example, if bamin and payinc were used, the coefficients would be -0.00001854 and 0.0000376. This highlights that you want to choose your units of measurement to help your reader.