Economics 311 Fall 2017
Daily Problem #26 December 1

This problem uses a dataset designed to examine the effect of seat-belt usage on traffic fatalities. It contains annual state-level data (for the 50 states and the District of Columbia) for 1983–97. Some states are missing data for some years, so there is a total of 556 available observations.

The variables in these regressions are defined below:

Variable	Definition
fatalityrate	Number of fatalities per billion of traffic miles
sb_usage	Seat belt usage rate
speed65	Binary variable for 65 mile per hour speed limit
speed70	Binary variable for 70 or higher mile per hour speed limit
ba08	Binary variable for blood alcohol limit ≤ .08%
drinkage21	Binary variable for age 21 drinking age
income	Per-capita income (in thousands of dollars)
age	Mean age
state	State
year	Year
fips	State numeric ID Code

A few of explanatory notes about the dummy variables, for those who weren't alive during the sample:

- The omitted category for speed limit is 55, so both 65 and 70 are high speed limits relative to the omitted category.
- The omitted category for drinking age is < 21, so the state/years with *drinkage21* = 1 have a higher drinking age than the state where it is 0.
- The legal blood alcohol limit in many states (including Oregon) is 0.08, but in some states it is lower, so states with *ba08* = 1 have a more strict requirement.

We begin by defining the dimensions of the panel data set. We must use the *fips* variable instead of *state* because you are not allowed to use a non-numeric variable as the cross-section index.

```
. xtset fips year
       panel variable:  fips (strongly balanced)
        time variable:  year, 1983 to 1997
                delta:  1 unit
```

Our first regression is simple OLS, which yields:

```
. reg fatalityrate sb_usage speed65 speed70 drinkage21 ba08 income age

      Source |       SS           df       MS      Number of obs   =       556
-------------+----------------------------------   F(7, 548)       =     81.56
       Model |  7163.59121         7  1023.37017   Prob > F        =    0.0000
    Residual |  6875.81485       548  12.5471074   R-squared       =    0.5102
-------------+----------------------------------   Adj R-squared   =    0.5040
       Total |  14039.4061       555  25.2962271   Root MSE        =    3.5422

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |   2.044971   1.230365     1.66   0.097    -.3718378     4.46178
     speed65 |   -.055393   .4221364    -0.13   0.896    -.8845965    .7738104
     speed70 |   2.220152   .5322322     4.17   0.000     1.174687    3.265617
  drinkage21 |  -.9306173   .9060268    -1.03   0.305    -2.710328    .8490933
        ba08 |  -1.984619   .4636014    -4.28   0.000    -2.895272   -1.073966
      income |  -821829.8   46995.31   -17.49   0.000    -914142.8   -729516.8
         age |  -.0550734   .1134823    -0.49   0.628    -.2779869    .1678402
       _cons |   37.71707   3.822357     9.87   0.000      30.2088    45.22533
------------------------------------------------------------------------------
```

1. Describe the results of this regression. Are they what you would expect? Explain.

Now we estimate the same equation with state fixed effects:

```
. xtreg fatalityrate sb_usage speed65 speed70 drinkage21 ba08 income age , fe

Fixed-effects (within) regression               Number of obs     =       556
Group variable: fips                            Number of groups  =        51

R-sq:                                           Obs per group:
     within  = 0.6693                                         min =         8
     between = 0.2355                                         avg =      10.9
     overall = 0.4075                                         max =        15

                                                F(7,498)          =    143.96
corr(u_i, Xb)  = 0.0070                         Prob > F          =    0.0000

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |  -7.427945   1.143347    -6.50   0.000    -9.674322   -5.181567
     speed65 |   -.729366   .3393736    -2.15   0.032    -1.396147   -.0625854
     speed70 |   .9725035    .334633     2.91   0.004     .3150369    1.62997
  drinkage21 |    .442645   .5298421     0.84   0.404    -.5983564    1.483646
        ba08 |   -1.41711   .3833831    -3.70   0.000    -2.170358   -.6638626
      income |  -.5013766   .0650037    -7.71   0.000     -.629092   -.3736613
         age |   .2601212   .3673732     0.71   0.479    -.4616711    .9819136
       _cons |    24.5302   11.98775     2.05   0.041     .9773989     48.083
-------------+----------------------------------------------------------------
     sigma_u |  3.6363728
     sigma_e |  1.8364121
         rho |  .79678915   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(50, 498) = 30.82                     Prob > F = 0.0000
```

2. How do the results of this regression differ from OLS? Why do you think this happens? Which set of results do you think are more reliable and why?