DEMONSTRATING MULTI-COLINEARITY WITH THE CONSUMPTION FUNCTION FOR THE U.S.

Here’s an example of multi-colinearity using the data from assignment 3 which also appeared on the second exam.

```
. use assg3

. summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     num |      20        11.5    5.91608          2         21
    year |      20      1977.5    5.91608       1968       1987
     gnp |      20    2347.765   1188.909      892.7     4486.2
  cnsmpt |      20    1506.075   792.4492      552.5       2966
  invest |      20       373.2   204.1194       15.2      716.4
 intrate |      20      9.6775   3.546692       5.25      18.87
deflator |      20      74.795   27.74539       39.3      118.9
    rgnp |      20    29.81999   4.608627   22.71501   37.73087
 rcnsmpt |      20    19.00449   3.291805   14.05852   24.94533
 rinvest |      20    4.661046    1.32795   .3707317   6.149861
rintrate |      20    .1370528   .0411115   .0691337   .1994715
     d80 |      20          .4   .5026247          0          1
  d80gnp |      20        1443   1851.497          0     4486.2
 d80intr |      20      4.9155   6.584569          0      18.87
d80invst |      20     229.795   296.4755          0      716.4
  d8rgnp |      20    13.68699   17.26466          0   37.73087
  d8rintr |      20     .048513   .0676758          0   .1994715
```

Here’s the basic regression relating consumption to gnp and interest rates

```
. regress cnsmpt gnp intrate

  Source |       SS       df       MS                  Number of obs =      20
---------+------------------------------               F(  2,    17) =15470.24
   Model |  11924985.5     2  5962492.77               Prob > F      =  0.0000
Residual |  6552.08842    17  385.416966               R-square      =  0.9995
---------+------------------------------               Adj R-square  =  0.9994
   Total |  11931537.6    19  627975.664               Root MSE      =  19.632

------------------------------------------------------------------------------
  cnsmpt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gnp |   .6752645   .0042736    158.008   0.000       .6662479     .684281
 intrate |  -6.621716   1.432583     -4.622   0.000      -9.644202   -3.599229
   _cons |  -15.20564   13.54168     -1.123   0.277      -43.77609    13.36481
------------------------------------------------------------------------------
```

Now we want to see whether there was a shift in the consumption function during the 1980s so we introduce a dummy variable d80=1 if year>1979, 0 otherwise.

```
. regress cnsmpt gnp intrate d80

  Source |       SS       df       MS                  Number of obs =      20
---------+------------------------------               F(  3,    16) =15491.70
```

```
   Model |  11927431.4     3  3975810.45             Prob > F      =  0.0000
Residual |  4106.26104    16  256.641315             R-square      =  0.9997
---------+------------------------------             Adj R-square  =  0.9996
   Total |  11931537.6    19  627975.664             Root MSE      =   16.02

------------------------------------------------------------------------------
  cnsmpt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gnp |   .6568772    .006902     95.172   0.000        .6422457    .6715088
 intrate |   -8.72962   1.353815     -6.448   0.000       -11.59958   -5.859661
     d80 |   56.75664   18.38515      3.087   0.007        17.78188    95.73141
   _cons |    25.6599   17.24354      1.488   0.156       -10.89478    62.21458
------------------------------------------------------------------------------
```

No problems to this point. The coefficient for d80 is significantly different from zero so it looks like there was an upward shift in the consumption function in the 1980s. Now we want to see if the marginal propensity to consume was different in the 80s than it was previously so we form d80gnp=d80*gnp and add it to the regression.

```
. regress cnsmpt gnp intrate d80 d80gnp

  Source |       SS       df       MS                Number of obs =      20
---------+------------------------------             F(  4,    15) =11673.12
   Model |  11927705.8     4  2981926.46             Prob > F      =  0.0000
Residual |  3831.78654    15  255.452436             R-square      =  0.9997
---------+------------------------------             Adj R-square  =  0.9996
   Total |  11931537.6    19  627975.664             Root MSE      =  15.983

------------------------------------------------------------------------------
  cnsmpt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gnp |   .6485717    .010565     61.389   0.000         .626053    .6710903
 intrate |   -6.99172   2.152975     -3.247   0.005       -11.58068   -2.402762
     d80 |  -13.35875   70.08501     -0.191   0.851       -162.7414    136.0239
  d80gnp |   .0221732   .0213911      1.037   0.316       -.0234208    .0677673
   _cons |   24.39106   17.24705      1.414   0.178       -12.37015    61.15228
------------------------------------------------------------------------------
```

Now we have a problem. The coefficient on the new variable, d80gnp, is not significant and the d80 coefficient is now insignificant as well. Notice that R^2 and Adj R^2 did not change from the previous regression to this one. This is a problem of multi-colinearity. If we do an F test to test
$H_0=b_{c,d80.i,gnp,d80gnp}=b_{c,d80gnp.i,gnp,d80}=0$

by comparing this regression, as the unrestricted regression to the base regression, as the restricted regression, we get:
$F_{szmple}=(.9997-.9995)/(1-.9997)*(20-5)/(5-3)=5.0$ which is > $F_{critical}=F_{2,15}=3.68$
So we can clearly reject H_0.

Notice the steps necessary to establish the presence of the multi-colinearity problem.
1.The regression with one of the variables added (in this case d80) yields a significant coefficient.
2.We add another variable (in this case d80gnp) and the first variable becomes insignificant and the new variable is insignificant as well. Also R^2 hardly increases at all when the second variable is added.

3.This step isn't really necessary but just as insurance we do an F test of the hypothesis that the two co-linear variables have coefficients that are jointly insignificantly different from zero and we reject this hypothesis.

The nature of the co-linearity in this case is complex. It involves not only the presence of d80 and d80gnp but also the inclusion of the interest rate variable. For example if we eliminate the interest rate variable and run the regression we get:

```
 regress cnsmpt gnp d80 d80gnp

  Source |       SS       df       MS                  Number of obs =      20
---------+------------------------------               F(  3,    16) = 9745.94
   Model |  11925011.8     3  3975003.94               Prob > F      =  0.0000
Residual |  6525.80171    16  407.862607               R-square      =  0.9995
---------+------------------------------               Adj R-square  =  0.9994
   Total |  11931537.6    19  627975.664               Root MSE      =  20.196

------------------------------------------------------------------------------
  cnsmpt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gnp |   .6315815   .0115979     54.457   0.000       .6069951    .6561678
     d80 |  -203.2679   48.80912     -4.165   0.001      -306.7386   -99.79718
  d80gnp |   .0762695   .0169569      4.498   0.000       .0403225    .1122166
   _cons |  -5.479687   18.43501     -0.297   0.770      -44.56016    33.60079
------------------------------------------------------------------------------
```

Both coefficients for d80 and for d80gnp are now significant, but R^2 is lower. Further if we delete d80 rather than the interest rate we get:

```
. regress cnsmpt gnp intrate d80gnp

  Source |       SS       df       MS                  Number of obs =      20
---------+------------------------------               F(  3,    16) =16561.64
   Model |  11927696.6     3  3975898.85               Prob > F      =  0.0000
Residual |  3841.06748    16  240.066717               R-square      =  0.9997
---------+------------------------------               Adj R-square  =  0.9996
   Total |  11931537.6    19  627975.664               Root MSE      =  15.494

------------------------------------------------------------------------------
  cnsmpt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gnp |   .6497492   .0083082     78.206   0.000       .6321367    .6673618
 intrate |  -7.334137   1.150334     -6.376   0.000      -9.772736   -4.895538
  d80gnp |    .018238   .0054272      3.360   0.004       .0067328    .0297432
   _cons |   25.27508   16.10374      1.570   0.136      -8.863326    59.41349
------------------------------------------------------------------------------
```

Now d80gnp coefficient is significant. So the multi-colinearity involves the three variables, intrate, d80 and d80gnp. Entering any two of the three yields significant coefficients for them but if all three are entered then we get multi-colinearity.