PS699 Problem Set 8: Some Regression Diagnostics

Let's return to one of our more successful sets of hypotheses, that the duration of governments in postwar democracies is a linear function of the number of parties in government, parliamentary support for the government, and party discipline. Re-estimate that model as I have done here:

```
reg dgovpw npgovpw psupgpw PD
```

```
  Source |       SS       df       MS                  Number of obs =      23
---------+------------------------------                F(  3,    19) =    4.10
   Model |  806.117215     3  268.705738                Prob > F      =  0.0211
Residual |  1244.90002    19  65.5210536                R-squared     =  0.3930
---------+------------------------------                Adj R-squared =  0.2972
   Total |  2051.01723    22  93.2280561                Root MSE      =  8.0945

------------------------------------------------------------------------------
  dgovpw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 npgovpw |  -3.520692   2.001078     -1.759   0.095      -7.708997    .6676123
 psupgpw |   .4338356   .2105621      2.060   0.053      -.0068759    .8745471
      PD |   9.242491   3.638877      2.540   0.020       1.626235    16.85875
   _cons |    1.24442   11.07908      0.112   0.912      -21.94436     24.4332
---------+--------------------------------------------------------------------
```

For each of the diagnostic plots and statistics below, print the results and answer the questions.
Answers might well be simply "no" or "nothing in particular".
[NOTE: syntax for these plots has changed a little from when problem set originally written. s([ctry]) no longer quite right to label graph points with 'ctry' name.]

```
. avplots , s([ctry])
```

1. This plots the estimated residuals from regressing **y** on all **X** but one against the estimated residuals from regressing that one on the others.
   - a1) Does anything you see in these three graphs alarm you as indicating a possible violation of the CNLRM assumptions?
   - a2) What would you expect to see if all was well on that score?
   - b1) Does anything you see in these three graphs alarm you as indicating potential inordinate influence of some country(ies) on the results?
   - b2) What would you expect to see if all was well on this score?
   - c) Anything else you think worth noting about this?

```
. cprplot npgovpw , s([ctry])

. cprplot psupgpw , s([ctry])

. cprplot PD , s([ctry])
```

2. These three plot the estimated residuals from regressing **y** on all **X** plus the line given by $b\mathbf{x}$ against that **x**.
   - a1) Does anything you see in these three graphs alarm you as indicating a possible violation of the CNLRM assumptions?
   - a2) What would you expect to see if all was well on that score?
   - b1) Does anything you see in these three graphs alarm you as indicating potential inordinate influence of some country(ies) on the results?
   - b2) What would you expect to see if all was well on this score?
   - c) Anything else you think worth noting about this?

```
. rvfplot , s([ctry])
```

3. This plots your estimated residuals versus fitted values.

a1) Does anything you see in this graph alarm you as indicating a possible violation of the CNLRM assumptions?
a2) What would you expect to see if all was well on that score?
b1) Does anything you see in this graph alarm you as indicating potential inordinate influence of some country(ies) on the results?
b2) What would you expect to see if all was well on this score?
c) Anything else you think worth noting about this?

```
. rvpplot PD , s([ctry])

. rvpplot psupgpw , s([ctry])

. rvpplot npgovpw , s([ctry])
```

4. These three plot your estimated residuals versus each of the **X** variables.
a1) Does anything you see in these three graphs alarm you as indicating a possible violation of the CNLRM assumptions?
a2) What would you expect to see if all was well on that score?
b1) Does anything you see in these three graphs alarm you as indicating potential inordinate influence of some country(ies) on the results?
b2) What would you expect to see if all was well on this score?
c) Anything else you think worth noting about this?

```
. dfbeta
DFnpgovp:  DFbeta(npgovpw)
DFpsupgp:  DFbeta(psupgpw)
DFPD:      DFbeta(PD)

. fpredict cooksd , c
```

5. Print and examine these three vectors of DFBetas and the Cook's distance summary score.
a1) Does anything you see in these four vectors alarm you as indicating a possible violation of the CNLRM assumptions?
a2) What would you expect to see if all was well on that score?
b1) Does anything you see in these four vectors alarm you as indicating potential inordinate influence of some country(ies) on the results?
b2) What would you expect to see if all was well on this score?
c) Anything else you think worth noting about this?

```
. vif

Variable |      VIF      1/VIF
---------+----------------------
 npgovpw |      1.24    0.807147
 psupgpw |      1.23    0.810485
      PD |      1.05    0.948403
---------+----------------------
Mean VIF |      1.18
```

5. Here are the variance inflation factors.
a1) Does anything you see in these values alarm you as indicating a possible violation of the CNLRM assumptions?
a2) What would you expect to see if all was well on that score?

b1) Does anything you see in these three values alarm you as indicating potential inordinate influence of some country(ies) on the results?
b2) What would you expect to see if all was well on this score?
c) Anything else you think worth noting about this?

```
. lvr2plot , s([ctry])
```

6. This graph plots leverage (potential influence) against normalized squared (estimated) residuals.
a1) Does anything you see in these values alarm you as indicating a possible violation of the CNLRM assumptions?
a2) What would you expect to see if all was well on that score?
b1) Does anything you see in these three values alarm you as indicating potential inordinate influence of some country(ies) on the results?
b2) What would you expect to see if all was well on this score?
c) Anything else you think worth noting about this?

```
. fpredict StudRes , rstu
```

7. Using these (estimated) studentized residuals, can you:
a) Specify a regression model to estimate to test the claim that variance is higher among Westminster (UK, Canada, Australia, and New Zealand) systems? What would you do? (Don't have to do it.)
b) Specify a regression model to estimate to test the claim that residuals in neighboring countries were correlated? What would you do? (Don't have to do it.)
c) In which (may be none, one, or several) of the above plots using regular residuals do you think Stata would have done better to use studentized residuals?