DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS

Posc/Uapp 816

Assignment 11

TIME SERIES REGRESSION

Name_____
(Printed)
Student Number_____
(Social Security Number)
E-mail_____

I would like you to try modeling time series data. The purpose is to build, test, and modify a simple (no doubt unrealistic) intervention model. There may be a fair amount of bookkeeping, as suggested in class, but you should have relatively little trouble if you stay organized. Each time you use multiple regression pick the variance inflation factor and Durbin-Watson statistic options.

The problem deals with several important issues in American politics, the most important of which is the apparent decline in voter turnout (in both presidential and congressional elections) during most of the 20th century. The web site has turnout figures defined as percent of eligible voters who actually voted in presidential elections from 1824 to 1996. (T =N = 44.)

Political scientists and historians have identified and described various electoral eras in American history. Mass participation in electoral politics really got underway in the 1820s and 1830 with the coming of Jacksonian democracy. By the end of the Civil War (1860 to 1864) the vast majority of the eligible electorate voted in presidential and congressional elections. After about 1900, however, turnout declined for several years until the 1930s when it picked up again. But after 1964 it began to fall once more. Many, many factors caused these changes in trend so it is impossible to attribute the collapse of any particular era to a single "event." Nevertheless, political scientists argue that the presidential campaign of 1896 was a "critical" election in that it marked (but did not cause) a long-term shift in partisan loyalties and changes in political behavior. (It is also about this time that the "mass media" began displacing the partisan press, reforms changed voter eligibility rules; and certain sections of the country became overwhelmingly one-party in their politics. All of these changes, it is suggested, had an adverse effect on participation. So why not attempt to construct a formal model of the event.

1. First retrieve the data set (Turnout in presidential elections) from the web site or enter them manually from the attached sheet.

2. There are 44 time periods (elections) but the first 4 will, I think, complicate the analysis. So eliminate them.
   A. By far the easiest way to do so is to copy the both the year and vote columns omitting the first 4 rows.
      i. In MINITAB go to **Manipulate**, **Copy columns**, fill in the dialogue box including the option that allows you to omit rows.
   B. If you just change the data to missing values with asterisks, some of the graphing programs may not work.
   C. You might eliminate the originals in order to keep a neat worksheet.
   D. **But keep a copy of the original file on your disk.**

3. Attach a carefully and fully labeled time series plot or graph of vote versus year or time order. By now you should be able to label the axes and add explanatory titles.

4. You need to create dummy variables to mark off the time periods.
   A. Let X_1 = 1, 2, 3,...,40. (There are total of 40 time periods for the new data.) I sometimes call this **Counter** in what follows.
   B. Let X_2 = 0 for each of the 14 periods before 1896; and 1 for each of the 26 others.
      i. The easiest way to do this is to go to the session window and at the prompt enter the **set** command.
         (1) Example: set c5
      ii. When you see the data prompt, enter number of "pre-intervention" periods (0) number of post-intervention periods (1).
         (1) Example: if you had 5 pre- and 7 post-intervention periods the command at the data prompt would be 5(0) 7(1).
      iii. Press enter and type **end** at the next data prompt. You should then see the regular MINITAB prompt.
      iv. I sometimes call this **Dummy**.
   C. Let $X_3 = X_2 X_1$. You can use mathematical expressions or calculator to create the variable. (I sometimes call this **Interaction**.)
   D. I suggest that you name the variables something meaningful such as "Counter" and "Dummy."
      i. If you want, attach a copy of the data sheet.

5. Okay, ready for some analysis? Use ANOVA to test the hypothesis that mean turnout out rate in period 1 (pre-intervention) differs from the rate in period 2. (Remember: use the dummy variable, as we have done in the past.)

   A. Mean for period 1 _____ Mean for period 2 _____

   B. Difference between mean 1 and mean 2 _____

   C. Observed F is _____ with degrees of freedom _____

D. Probability of F? _____

6. Now use regression to make the same analysis.

A. Regression equation: _____

B. Evaluate the model: R^2 _____ s _____ F_{obs} = _____

C. What is the expected (predicted) turnout rate for the pre-intervention period? _____

D. What is the "estimated" numerical "effect" of the intervention on the level of participation or turnout? _____

E. Is this effect statistically significant? Explain with the observed test results.

_____

_____

_____

_____

7. Now develop a model that includes changes in level and slope (trend).

A. Estimated equation? _____

B. Evaluation: R^2 _____ s _____ F_{obs} = _____

C. What is the model for the pre-intervention period? _____

D. What is the model for the post-intervention period? _____

E. What estimated numerical effect did the intervention (the critical election of 1896) have on the rate of participation in the 20^{th} century? _____

F. Attach a plot of the residuals against time order. (If you are using the Student version, you will have to store the residuals and then use time series plot.) What pattern, if any, do you see?

_____

_____

G. What is the Durbin-Watson statistic? _____

8. Let's see if we can't deal with autocorrelation. Obtain the lagged residuals. (You have to have stored them somewhere. Then use Time series and Lag or the simple Lag command at the session window.)

   A. What is the estimated value of the autocorrelation parameter, ρ? _____

9. Now lag the dependent and independent variables. Why not give them short names to help keep track of their meaning such as "Lag-vote," "Lag-count," and so forth.

10. Now use the original variables, their lags, and the estimated autocorrelation parameter to create new or transformed variables, as described in class.
   A. Example: New-vote = Vote - (.222*Lag-vote)
      i. Note the punctuation and symbols.
      ii. This assumes that the estimated ρ is .222, which it is not. Use the correct value.

      iii. The calculator or mathematical expressions boxes are useful for creating these variables.

   B. Again, name them so you can keep track of your work.

   C. What is the estimated model using the new or transformed variables?
      _____

   D. What this Durbin-Watson statistic? _____

11. I can tell you right now that multicolinearity is wreaking havoc on the estimates and the significance tests.

   A. What is the observed F for this model? _____

   B. What is the t for the trend or counter variable? _____ Probability? _____

   C. Look that the other individual coefficients and their levels of significance and variance inflation factors. None of them will be significant. The overall model is highly significant but the coefficients are not. That's a tipoff.

12. Try dropping the new dummy variable that allows you to test for change in level. Instead, just include the trend (i.e., modified X_1 and the modified X_3) variables.

A. What is the estimated equation? _____

B. What is R^2? _____ The observed F? _____ DW? _____

C. Describe the effect of the intervention on the rate or trend in participation.

_____

_____

_____

_____

D. Attach a labeled plot of residuals for this model versus order of observations.

13. What is your conclusion? Did turnout decline significantly (in a substantive sense) after 1896?

_____

_____

_____

_____

14. Could you construct a multiple intervention model for these data? Think about it.

Go to Assignment page

Go to Statistics page

Turnout in Presidential Elections, 1824 to 1996

Year	*Vote*	*Year*	*Vote*
1824	26.9	1916	61.6
1828	57.6	1920	49.2
1832	55.4	1924	48.9
1836	57.8	1928	56.9
1840	80.2	1932	56.9
1844	78.9	1936	61.0
1848	72.7	1940	62.5
1852	69.6	1944	55.9
1856	78.9	1948	53.0
1860	81.2	1952	63.3
1864	73.8	1956	60.6
1868	78.1	1960	64.0
1872	71.3	1964	61.7
1876	81.8	1968	60.6
1880	79.4	1972	55.2
1884	77.5	1976	53.6
1888	79.3	1980	52.8
1892	74.7	1984	53.1
1896	79.3	1988	50.2
1900	73.2	1992	55.6
1904	65.2	1996	49.0
1908	65.4		
1912	58.8		