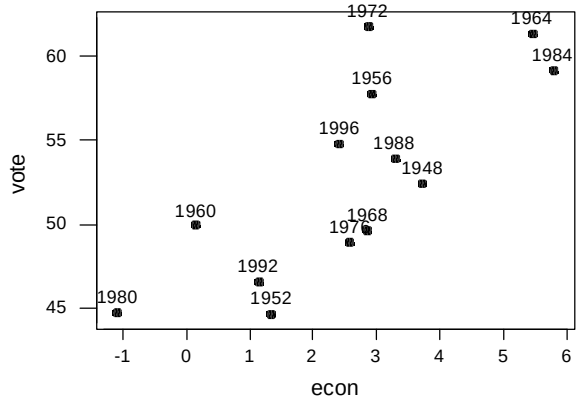
Statistics // Green

Solutions to Supplementary Problem Set 9

1. *What is real per capita disposable income? What does "net candidate advantage" measure?*
"Real" means corrected for inflation (using the consumer price index, which has its problems). "Per capita" mends divided by the size of the population. "Disposable income" refers to income that may be spent after taxes are removed. The index of net candidate advantage is an omnibus measure of how popular the 'incumbent' candidate is in comparison to the challenger.

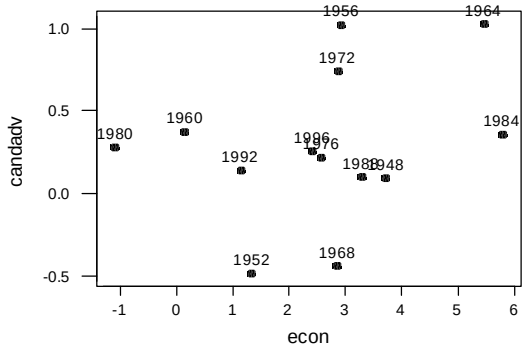
2. *Create a scatterplot of the national vote by ECON. Does this plot indicate a strong or weak electoral-economic connection? What is the correlation between these two variables? What are the apparent "outliers"?*




```
Correlation of vote and
econ = 0.762, P-Value =
0.002
```

1972 seems to be an outlier; Nixon won by a bigger landslide than the model would predict.

3. *Create a scatterplot of candidate advantage by ECON. Is there a connection here? What is the correlation, and what does it imply?*




```
Correlation of candadv and econ
= 0.303, P-Value = 0.314
```

This correlation is weakly positive, suggesting that incumbent candidates tend to be somewhat more popular when running amid a strong economy.

4. *Regress the vote on ECON and interpret the slope, the intercept, the standard errors of the slope and intercept, the standard error of the regression, and the R2. Explain how each of these statistics is calculated.*

```
The regression equation is vote = 46.6 + 2.39 econ

Predictor        Coef       StDev          T        P
Constant       46.564       1.937      24.05    0.000
econ           2.3863      0.6119       3.90    0.002

S = 4.051       R-Sq = 58.0%     R-Sq(adj) = 54.2%
```

The slope suggests that for every one unit increase in RDPCI, the expected vote percentage increases by 2.38. The intercept suggests that if RDPCI is zero, the expected vote is 46.564% (i.e., the incumbent loses if the economy is stagnant). The standard error for the slope suggests that the sampling distribution of this parameter estimate is centered at 2.38 and has a standard deviation of .61. The p-value of .002 tells us that we can easily reject the null that b=0 against a two-sided alternative. The standard error of the intercept tells us that under stagnant economic conditions, the expected vote is 46.6% with a standard deviation of 1.9%. The R-square indicates that the independent variable predicts 58% of the observed variability in Y. The standard error of the regression (s) indicates that 4.1 is our best guess of the standard deviation of the disturbances.

The slope is calculated as Cov(X,Y)/Var(X). The intercept is the mean of Y minus (b times the mean of X). S is the sum of squared residuals divided by n minus the number of parameter estimates (2 in this case). The standard error of the slope is s divided by the square root of (n-1 times Var(X)). The R-squared in this case is the squared correlation between X and Y.

5. *Regress the vote on ECON and candidate advantage. Substantively, how do the results differ from #4? What do you conclude?*

```
The regression equation is vote = 45.9 + 1.87 econ + 7.11 candadv

Predictor        Coef       StDev          T        P
Constant       45.883       1.214      37.78    0.000
econ           1.8661      0.3992       4.67    0.001
candadv         7.115       1.656       4.30    0.002

S = 2.518       R-Sq = 85.3%     R-Sq(adj) = 82.3%
```

The effects of ECON are somewhat diminished, but remain strong. Part of the influence of ECON on vote is mediated through candidate advantage. That is, one reason why economic times are good for incumbents is that they make the incumbent candidates more attractive (or scare off stronger competitors!).

6. *Show that the following procedure generates the same slope estimate obtained in #5 for the effect of ECON: regress ECON on CANDIDATE and compute the residuals; then, regress VOTE on these residuals. Show that an analogous approach can be used to calculate the slope estimate for CANDIDATE.*

```
The regression equation is econ = 2.22 + 1.26 candadv

Predictor        Coef       StDev          T        P
Constant       2.2201      0.6269       3.54    0.005
candadv         1.258       1.192       1.06    0.314

The regression equation is vote = 52.7 + 1.87 res_econ

Predictor        Coef       StDev          T        P
Constant       52.715       1.428      36.93    0.000
res_econ       1.8661      0.8160       2.29    0.043

S = 5.147       R-Sq = 32.2%     R-Sq(adj) = 26.1%

The regression equation is candadv = 0.096 + 0.0731 econ

Predictor        Coef       StDev          T        P
Constant       0.0958      0.2192       0.44    0.671
econ          0.07311     0.06927       1.06    0.314

S = 0.4585      R-Sq = 9.2%      R-Sq(adj) = 0.9%


The regression equation is vote = 52.7 + 7.11 res_adv

Predictor        Coef       StDev          T        P
Constant       52.715       1.479      35.63    0.000
res_adv         7.115       3.507       2.03    0.067

S = 5.334       R-Sq = 27.2%     R-Sq(adj) = 20.6%
```

7. *Use your results based on the period 1948-76 to forecast the outcomes for elections 1980-1996. Generate prediction intervals (you choose the width) for each year. Did the actual election results fall within these prediction intervals?*

```
     Fit  StDev Fit         95.0% CI             95.0% PI
   47.95       3.18   (   39.76,   56.13)  (   36.76,   59.13) 1980
   57.58       2.61   (   50.86,   64.30)  (   47.42,   67.74) 1984
   52.31       1.29   (   48.99,   55.63)  (   43.99,   60.63) 1988
   49.81       1.56   (   45.81,   53.82)  (   41.20,   58.43) 1992
   52.38       1.07   (   49.62,   55.14)  (   44.27,   60.49) 1996

Actual Vote Outcomes

44.70   1980
59.17   1984
53.90   1988
46.55   1992
54.74   1996
```

In each case, the observed outcome fell within the (very wide) 95% prediction intervals.