Statistical Models: Theory and Practice
DA Freedman
www.stat.berkeley.edu
Replicating Gibson on McCarthy: Weighted Regression
October 2005

The caption to figure 1 in Gibson's paper suggests the regression was run on the 26 states with complete data. According to note 9, observations were weighted by "the square root of the number of respondents within the state." He reports the unstandardized equation as

$$Y = 7.31 - .14 \times \text{mass opinion} - 1.11 \times \text{elite opinion}$$

I am unable to replicate this. If I take respondents as masses plus elites, and assume the error variance for a state to be inversely proportional to the number of respondents (his table A1), I get

$$Y = 9.10 - .55 \times \text{mass opinion} - 1.10 \times \text{elite opinion},$$

neither coefficient being significant. Unweighted, I get

$$Y = 7.39 + .19 \times \text{mass opinion} - 1.38 \times \text{elite opinion},$$

the coefficient of elite opinion being just significant. Note 11 suggests that the interaction (product) of the two opinion variables is insignificant. I find it significant, whether weighted or unweighted. With the weights, for instance, I get

$$\begin{array}{l} Y = \underset{26.3}{-55.8} + \underset{7.4}{17.8} \times \text{mass opinion} + \underset{5.4}{12.2} \times \text{elite opinion} - \underset{1.5}{3.8} \times \text{interaction} \end{array}$$

the numbers under the coefficients being the estimated standard errors. In these equations, Y is the predicted repression score for a state. The version of the data set I used (tables 1 and A1 in the paper) is on the web as a text file,

http://www.stat.berkeley.edu/users/census/gibson.txt

States are in alphabetical order, identified by fips code; missing data are entered as -1. In order, the columns are

(i) fips code for state
(ii) state average response—masses
(iii) number of mass respondents in state
(iv) state average response—elites
(v) number of elite respondents in state
(vi) state score on repressive legislation

Column (vi) comes from table 1 in the paper; columns (ii)–(v), from table A1. Column (i) may be useful in cross-checking the data against the source. The fips codes (DC is irrelevant) are shown in

http://www.stat.berkeley.edu/users/census/fips.txt

With my setup, the weights might be construed as number of respondents n_i in each state, since we are minimizing

$$\sum_i n_i (Y_i - X_i \beta)^2$$

Another idea is this. Take the error variance for state i as inversely proportional to $1/\sqrt{n_i}$, where n_i is the number of mass respondents in the state. So the weights are literally $\sqrt{n_i}$. We would choose β to minimize

$$\sum_i \sqrt{n_i}(Y_i - X_i \beta)^2$$

Of course, constant$/\sqrt{n_i}$ is not an obvious choice for the variance—even the variance for estimated mass opinion. In any event, we get

$$Y = 7.5965 - .2578 \times \text{mass opinion} - 1.0418 \times \text{elite opinion}$$

No replication, and neither coefficient is significant. (Their SEs are both about .73.) If we standardize the unweighted variables, but weight the regression as above, we get

$$Y = -0.0677 \times \text{mass opinion} - 0.3575 \times \text{elite opinion}$$

If we truncate instead of rounding, this replicates Gibson's path diagram. However, nothing is significant.