

GSS Panel Data Release Notes

Released in April 2010

I. Overview

This GSS panel data set has two waves of interviews, first in 2006 and second in 2008. Among the 4,510 cases interviewed in 2006, we randomly selected 2,000 cases for re-interviews. We then screened out 59 respondents who did not meet our eligibility criteria for re-interview: living primarily in households in the United States at the time of the second interview. Among the eligible cases, 405 cases could not be re-interviewed. Thus, we ended up re-interviewing 1,536 cases for the second wave in 2008.

This data file contains those 2,000 respondents who were pre-selected among the 2006 samples and those variables that were asked in both 2006 and 2008.

II. Data File Organization

1. To denote the wave, we have added a suffix “_1” or “_2” to the existing GSS variable names. For example, EDUC_1 is the years of education in the first wave (2006) and EDUC_2 is education in the second wave (2008).
2. YEAR_1 is the GSS year of the first wave while YEAR_2 is the GSS year of the second wave. For those who were not re-interviewed in the second wave, YEAR_2 is coded to “Inapplicable (IAP)”.
3. ID_1 is the identification number used in the GSS 2006 data, and ID_2 is ID used in 2008.
4. PANSTAT_1 and PANSTAT_2 indicate panel selection status. PANSTAT_1 is coded to 4 for all 2,000 cases since they were newly sampled in 2006. With PANSTAT_2, users can identify those cases that were: (1) selected, eligible for re-interviews, and actually re-interviewed; (2) selected, eligible, but not re-interviewed; and (3) selected, but not eligible and not re-interviewed.
5. For those 464 cases that were not re-interviewed in the second wave, values in all variables are coded to IAP (actual codes vary by variables).
6. WTPAN_2 is a weight variable that adjusts the data to panel sampling in the second wave. WTNRPAN_2 makes an additional adjustment for non-response. Please refer to the next

section for more information. For the sampling method and design weights in 2006, please refer to the *GSS Codebook*.

III. Weight

Selection of respondents from the 2006 round for the panel was done in two phases: first we selected segments and second we selected cases within those segments. To calculate weights for the panel cases, we simply adjusted the 2006 design weights to account for these additional stages of selection. Let $WEIGHT_{2006}$ be the design weight of all panel cases calculated just as $WEIGHT$ is described above. Then $WEIGHT_{panel}$ is simply:

$$WEIGHT_{panel} = \frac{WEIGHT_{2006}}{\pi_{segment}^{panel} * \pi_{case}^{panel}}$$

The final weight is just this variable, rescaled to sum to the number of completed panel cases.

To adjust this weight for nonresponse in 2008, we use logistic regression to predict a response propensity score for each case. The independent variables in this regression are factual (not attitudinal) responses collected in 2006 for which there is no missing data for either respondents or nonrespondents.¹ The variables used were: born outside the US², living alone, gender, race (white / nonwhite) and Census division. We used the predicted response propensities from this model to divide the responding and nonresponding cases into five equal size adjustment cells and within each cell we inflated the weights of the responding cases by the inverse of the mean response propensity in that cell.

$$WEIGHTNR_{panel} = \frac{WEIGHT_{2006}}{(\pi_{segment}^{panel} * \pi_{case}^{panel})} * \frac{1}{mean_h(\hat{RP})}$$

where h is the nonresponse adjustment cell. This weight was then also scaled to sum to the number of completed panel cases.

¹ Missing data on the 40 ineligible cases was permitted.

² Of all variables studied, this one most strongly predicted response in 2008.