Assignment 4

1.   Suppose that the population of Stony Brook political science majors contains 35%

        males and 65% females. A survey you conduct of Stony Brook political science

        majors contains 100 male and 100 female students. These students were surveyed

        to determine their interest in a minor in International Relations. It was determined

        that 25% of male students and 60% of female students would be interested in such

        a minor.

(a)   Compute sample proportion of political science majors interested in the

        International Relations minor? Does the sample proportion provide a good

        estimate of the population proportion in this case? Explain.

(b)   Compute the weighted sample proportion of political science majors interested in

        the International Relations minor. Does the weighted sample proportion provide a

        good estimate of the population proportion in this case? Explain.

2.   For this question, consider the data file 'cbs04.sav', which is available on the

        course website.

(a)   Compute the sample proportion of respondents who have a favorable opinion of

        John Kerry (using the variable 'q2').

(b)   Compute the sample proportion of respondents who have advanced degrees (using

the variable 'educ').

(c)   We suspect that we have too many highly educated people in our sample. To

correct for this problem, we use the weights provides by the survey (the weight

variable is 'wght'). Recompute the frequency table for education.

(d)   Interpret the difference between the results in parts (b) and (c).

(e)   Using the same weights, compute the weighted proportion of respondents who

have a favorable opinion of John Kerry. Does this number change? Why does it

change?

3.   The 2004 exits polls were widely off. The exit polls predicted a 5 percent Kerry

win, but in reality by over 2 percent. In this problem, you will create our own

weights which will correct for this problem by weighting the results of the exit

poll to the election outcome. Consider the data file "exit_poll_US_04.sav" on the

course webpage.

(a)   Make sure that the weights are turned off and verify that Kerry's vote share is 5

percent higher than Bush's vote share. Include SPSS output in your answer.

(b)   Make sure that the weights "wgt" are turned on and Verify that Bush and Kerry's

vote shares are correctly estimated (the actual election results were 50.7% for

Bush and 48.3% for Kerry, and 1% for other candidates). Include SPSS output in

your answer.

(c)   Follow the procedure of the tutorial to generate a new set of weights (call these

new weights "new_wgt"). For simplicity, treat respondents with DK, missing

values, others, and "didn't vote" in the survey as having voted for other

candidates. Turn the new weights on to verify that you have successfully

corrected for the imbalance in the Bush-Kerry vote totals. Include SPSS output in

your answer.

(d)   Consider George W. Bush's approval rating as measured in the survey. Compute

estimates based on (i) no weighting, (ii) the weight included in the survey

("wgt"), and (iii) the new weight you created ("new_wgt"). What accounts for the

differences between these estimates? Which of these three estimates is likely to be

most accurate? Which is likely to be least accurate? Explain.

4.   For this problem, consider the data file wp04.sav on the course website. The

Washington Post/ABC News poll does not include a likely voter weight in their

data set, so we will create one of our own.

(a)   Take a look at the data set and identify variables that would be useful in

constructing a likely voter index. Explain why these variables are useful (hint: use

the variables q1, q2, q905, and q906).

(b)   Create a four point likely voter index using the variables you identified in part (a).

Report the frequencies of this variable with the demographic weight (WEIGHT)

turned on. (hint: use the transform command to create the variable "idx" with the

code, (q1=1)+(q2=1)+(q905=1)+(q906=1))

(c)   Create a likely voter weight with a target turnout rate of 55%. What percentage of

4's, 3's, 2's, 1's, and 0's should be included in the likely voter sample? Report a

histogram for the likely voter weights. (hint: use the transform command to create

the variable "lvw55" with the code WEIGHT*(idx=4)+(.55-

.424)/.411*WEIGHT*(idx=3))

(d)   Compute the percentage of respondents who plan to vote for Bush (q3) using both

the demographic weight and the likely voter weight.

(e)   Why are these two numbers different? Which number would you prefer to use to

predict how well Bush will do in the 2004 election? Explain.