Exploring The Web

Chapter 22

22.45 Where are the smokers? The Behavioral Risk Factor Surveillance System (BRFSS) is an ongoing data collection program designed to measure behavioral risk factors for the adult population (18 years of age or older) living in households. Data are collected from a random sample of adults (one per household) through a telephone survey. Go to the website `apps.nccd.cdc.gov/BRFSS/` and under Category, go to TOBACCO USE. Under the topic Adults Who are Current Smokers, you will find the percentage of smokers in each state.

(a) Which state has the highest percentage of smokers, and what is the reported value? Which state has the lowest percentage, and what is its value? Are the reported percentages statistics or parameters?

(b) Choose a state of interest to you and click on the link. In the table that opens, there is a line for n and the entries are the numbers who answered Yes and No. Find the percentage in the sample who answered Yes. Notice that it is different than the percentage reported in the table. The table estimates are weighted to try to reduce the bias. If it is determined that certain portions of the population are underrepresented in the sample, then that portion of the sample receives more weight when computing the estimate of the percentage. The assumptions for an SRS are rarely met in practice, and more complicated methods are often necessary to estimate proportions.

22.46 More on weighting of estimates. The website for the American Association for Public Opinion Research discusses several issues about polls. Read the discussion of weighting at `www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Weighting.aspx`. The necessity of weighting arises in Exercise 22.28. (page 524). Here are the important points.

Researchers want to estimate the proportion of Greenville country residents who have used a paved greenway trail in the last six months. It is known from U.S. Census Bureau records that 13% of the adult population of Greenville county is over 65 years old. A random sample of 2461 residential phone numbers are contacted and 726 surveys are completed, with 689 including data on the respondents age. Among the 689 surveys that included age data, 36% were from respondents over 65, and those over 65 tend to use the trail less than those under 65. Because those over 65 are overrepresented in the sample compared with the population of Greenville county and they also tend to use the trail less, the proportion in the sample that use the trail will tend to be smaller than that in the population, producing bias in the estimate.

(a) The online discussion of weighting gives three uses of weighting to adjust poll results. Which of these three uses is relevant in this situation?

Here is a breakdown of the respondents.

	Used the Trail in the Last 6 Months		
	Yes	No	Total
Under 65	152	284	436
Over 65	27	226	253
Total	179	510	689

(b) What is the proportion of adults under 65 in the sample, $\hat{p}_{<65}$, that have used the trail in the last six months and the proportion of adults over 65 in the sample, $\hat{p}_{>65}$, that have used the trail?

(c) If the sample proportions $\hat{p}_{<65}$ and $\hat{p}_{>65}$ were equal to the true proportions of those under and over 65 who use the trail, what would be the true proportion of all adults in Greenville county who use the trail? This is a weighted estimator. (Hint: You will need to use the fact that 87% of the population is under 65 and 13% is over in your computation.)

(d) The weighted estimator is larger than the unweighted estimator $\hat{p} = 179/689 = 0.26$. Why is this the correct direction to adjust the estimator?