36-303 Sampling, Surveys & Society
Homework 03 Solutions

February 29, 2012

1 1a. Groves 7.1, p. 255

This question measures people's attitudes toward president Obama, U.S military, energy independence and public lighting. I don't think the analytic goal is met because the sentence wording fall into the bipolar approach category (as discussed on page 249 in the textbook). It forces respondents to choose among positive and negative emotions.

2 1b. Groves 7.5, p. 255-256

The first problem with this question is that there is no "other " category. Second, the definitions of some of the choices are ambiguous. For example, what does a "tune-up" or "transmission overhaul" consist of? Third, there are overlap among some of the options. For example, "oil change" is technically part of a "fluid replacement". Lastly, the question only allows one or two answers, but an individual may have had more service work done so options lower down on the list may be possibly ignored.

3 1c. Groves 7.6, p. 256

- Ex. 1. It is probably difficult for the respondent to remember specific activity information from the previous four weeks. It may be better to ask what types of household activities they perform on a regular basis. Also, some of the options have overlapping meanings. One possible rewording can be " During the past four weeks, what type of household activities have you done on a regular basis?"

- Ex. 2. First, the question is asking information that may be sensitive to respondents. Secondly, "times" are a very ambiguous measure for the question purpose. One possible wording may be "How many days of the past week did you drink alcoholic beverages?"

- Ex. 3. I think income and finance type of questions may be sensitive to some respondents. It may be better to ask a self-administered survey type of question. Secondly, the sentence may be hard for respondents to understand and there are many information I consider repetitive. One possible wording may be " What would be the smallest income you family need to meet the expenses you consider necessary? "

- Ex. 4. I think the problem of this sentence is that the question covers too broad of a time period for respondents to remember easily. There is also redundant information in the question. One possible wording is "During the past 6 months, how many times did you get sick or injured that resulted in staying home or hospital more than half the day?"

4 1d. Groves 7.8, p. 256-257

- a) As hospitalizations are rare, respondents are likely to use recall-and-count estimation, tallying up the relevant instances and potentially "adjusting the answer upward to allow for forgotten incidents." Depending on whether omissions are under- or over-compensated for, this could lead to either under- or over-reporting.
- b) As eating in restaurants is for many people a relatively frequent occurrence, respondents are likely to employ rate-based estimation, perhaps counting the number of times eaten out in the last week and multiplying by four to scale up to the month level. Studies suggest rate-based estimation leads to overestimates of frequencies, though it could in principle lead to underestimates as well.
- c) We imagine that, like hospitalizations, a respondent's spouse's vacations over a summer are also relatively infrequent; this would suggest recall-and-count estimates, with their attendant reporting errors (see part a).

5 Question 2

5.1 a)

$$E[\hat{\pi}] = E\left[\frac{\hat{\lambda} - 1/2(1-p)}{p}\right] \quad (1)$$

$$= \frac{1}{p}E\hat{\lambda} - \frac{1}{2p}(1-p) \quad (2)$$

$\hat{\lambda}$ is defined as the fraction of "Yes" answers in the survey, therefore, $n\hat{\lambda} \sim Binomial(n, P(Yes))$ and $E[n\hat{\lambda}] = nP(Yes) = n\lambda$. Thus, $E\hat{\lambda} = \lambda$. Replacing,

$$\begin{aligned} E[\hat{\pi}] &= \frac{1}{p}\lambda - \frac{1}{2p}(1-p) && (3) \\ &= \frac{\lambda - 1/2(1-p)}{p} && (4) \\ &= \pi && (5) \end{aligned}$$

5.2 b)

$$\begin{aligned} V[\hat{\pi}] &= V\left[\frac{\hat{\lambda} - 1/2(1-p)}{p}\right] && (6) \\ &= V\left[\frac{\hat{\lambda}}{p}\right] && (7) \\ &= \frac{1}{p^2}V[\hat{\lambda}] && (8) \end{aligned}$$

¿From the last result,

$$\lim_{p\to 1} V[\hat{\pi}] = \lim_{p\to 1}\frac{1}{p^2}V[\hat{\lambda}] = V[\hat{\lambda}] \quad (9)$$

5.3 c)

Using a normal approximation, a 95% confidence interval for π can be constructed as $\hat{\pi} \pm 2 \cdot se(\hat{\pi})$. Therefore, the width of the interval is $w = 4 \cdot se(\hat{\pi})$. If we want the confidence interval be only 0.02 wide then, $se(\hat{\pi}) = 0.02/4 = 0.005$.

$$\begin{aligned} se(\hat{\pi})^2 &= V[\hat{\pi}] && (10) \\ &= \frac{1}{p^2}V[\hat{\lambda}] && (11) \end{aligned}$$

Since $n\hat{\lambda} \sim Binomial(n, P(Yes))$,

$$\begin{aligned} se(\hat{\pi})^2 &= \frac{1}{p^2}V[\frac{1}{n}n\hat{\lambda}] && (12) \\ &= \frac{1}{p^2n^2}V[n\hat{\lambda}] && (13) \\ &= \frac{1}{p^2n^2}n(1-\lambda)\lambda && (14) \\ &= \frac{\lambda}{np^2}(1-\lambda) && (15) \end{aligned}$$

Since we know $p = 1/2$ and we assume $\pi = 0.1$,

$$\lambda = \pi p + \frac{1}{2}(1 - p) = 0.3 \tag{16}$$

and

$$se(\hat{\pi})^2 = \frac{4}{n} \times 0.3(1 - 0.3)$$

Solving for n we get

$$n = \frac{4}{0.005^2} \times 0.3(1 - 0.3) = 33600 \tag{17}$$