36-303 Sampling, Surveys & Society
Homework 03 Solutions

February 15, 2011

1 Question 1

On page 171 of the textbook, you can find discussion about the question order effect. Question order are significantly less likely to occur in a self-administered survey than in a telephone or face-to-face survey.

2 Question 2

Let's call $\sum_i^n p_{ij} = p_{+j}$ and $\sum_j^n p_{ij} = p_{i+}$. Note that $\Pr[X = x_i] = p_{i+}$ and that $\Pr[Y = y_j] = p_{+j}$. Note also that $\sum_{i=1}^n \sum_{j=1}^n p_{ij} = 1$

a)

$$
\begin{aligned}
E\left[aX + bY + c\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(ax_i + by_i + c\right)p_{ij} && (1)\\
&= a\sum_{i=1}^{n}\sum_{j=1}^{n}x_i p_{ij} + b\sum_{i=1}^{n}\sum_{j=1}^{n}y_j p_{ij} + c\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij} && (2)\\
&= a\sum_{i=1}^{n}x_i p_{i+} + b\sum_{j=1}^{n}y_j p_{+j} + c && (3)\\
&= aEX + bEY + c && (4)
\end{aligned}
$$

b)

$$
V\left[aX + bY + c\right] = \underbrace{E\left[(aX + bY + c)^2\right]}_{\alpha} - \underbrace{E\left[aX + bY + c\right]^2}_{\beta} \quad (5)
$$

$$\alpha = a^2 \sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 p_{ij} + b^2 \sum_{i=1}^{n}\sum_{j=1}^{n} y_j^2 p_{ij} + c^2 + 2ab \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j p_{ij} \quad (6)$$

$$+2ac \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j p_{ij} + 2bc \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j p_{ij} + 2c^2 \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}$$

$$= a^2 EX^2 + b^2 EY^2 + 2abEXY + 2acEXY + 2c^2 \quad (7)$$

$$\beta = a^2 E[X]^2 + b^2 E[Y]^2 + c^2 + 2abEXEY + 2acEX + 2c^2 \quad (8)$$

Finally,

$$Var[aX + bY + c] = \alpha - \beta \quad (9)$$

$$= a^2 \left[EX^2 - E[X]^2\right] + b^2 \left[EY^2 - E[Y]^2\right] \quad (10)$$

$$2ab\left[EXY - EXEY\right]$$

$$= a^2 V[X] + b^2 V[Y] + 2ab\mathrm{Cov}[X, Y] \quad (11)$$

c)

If X and Y are independent $p_{ij} = p_i \cdot p_j = p_{i+} \cdot p_{+j}$,

$$E[X|Y = y_j] = \sum_{i=1}^{k} x_i P(X = x_i | Y = y_j) \quad (12)$$

$$= \sum_{i=1}^{k} x_i \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \quad (13)$$

$$= \sum_{i=1}^{k} x_i \frac{p_{i+} \cdot p_{+j}}{p_{+j}} \quad (14)$$

$$= \sum_{i=1}^{k} x_i \cdot p_{i+} \quad (15)$$

$$= E[X] \quad (16)$$

3 Question 3

3.1 a)

$$E[\hat{\pi}] = E\left[\frac{\hat{\lambda} - 1/2(1-p)}{p}\right] \quad (17)$$

$$= \frac{1}{p} E\hat{\lambda} - \frac{1}{2p}(1-p) \quad (18)$$

$\hat{\lambda}$ 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,

$$\begin{aligned} E[\hat{\pi}] &= \frac{1}{p}\lambda - \frac{1}{2p}(1-p) && (19)\\ &= \frac{\lambda - 1/2(1-p)}{p} && (20)\\ &= \pi && (21) \end{aligned}$$

3.2 b)

$$\begin{aligned} V[\hat{\pi}] &= V\left[\frac{\hat{\lambda} - 1/2(1-p)}{p}\right] && (22)\\ &= V\left[\frac{\hat{\lambda}}{p}\right] && (23)\\ &= \frac{1}{p^2}V[\hat{\lambda}] && (24) \end{aligned}$$

¿From the last result,

$$\lim_{p\to 1} V[\hat{\pi}] = \lim_{p\to 1} \frac{1}{p^2}V[\hat{\lambda}] = V[\hat{\lambda}] \quad (25)$$

3.3 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, $se(\hat{\pi}) = 0.02/4 = 0.005$.

$$\begin{aligned} se(\hat{\pi})^2 &= V[\hat{\pi}] && (26)\\ &= \frac{1}{p^2}V[\hat{\lambda}] && (27) \end{aligned}$$

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

$$\begin{aligned} se(\hat{\pi})^2 &= \frac{1}{p^2}V[\frac{1}{n}n\hat{\lambda}] && (28)\\ &= \frac{1}{p^2n^2}V[n\hat{\lambda}] && (29)\\ &= \frac{1}{p^2n^2}n(1-\lambda)\lambda && (30)\\ &= \frac{\lambda}{np^2}(1-\lambda) && (31) \end{aligned}$$

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

$$\lambda = \pi p + \frac{1}{2}(1-p) = 0.3 \tag{32}$$

and

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

Solving for n we get

$$n = \frac{4}{0.005^2} \times 0.3(1-0.3) = 33600 \tag{33}$$

4 Question 4

part a) 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.
part b) 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.
part c) 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?"