11 Supplemental Topics

11.4 Random-Response Model

Even in an anonymous survey, some respondents still might not want to answer truthfully, for fear that the survey is not truly anonymous. To mitigate this problem, consider using a *random-response model.*

Below is one version of the *random-response model.*

Example: Suppose a survey question is: "In the past 30 days, did you use marijuana at least once?" Let p be the true unknown population proportion of students at your university who *truly* did use marijuana at least once in the past 30 days. Out of four cards, three cards are labeled "I did NOT use marijuana," and the fourth card is labeled "I used marijuana." Each respondent selects one of the four cards WITH replacement, and is asked to respond if the card is *correct* or *incorrect.* Let θ be the probability that the respondent selects the card labeled "I used marijuana." Suppose that n_1 out of the n respondents state that the card is *correct.* Estimate p.

(a) Determine the probability that a respondent states the the card is correct.

(b) How should we estimate $P(C)$?

(c) Determine $\hat{p}$, the true sample proportion of students who used marijuana.

$$\hat{p} = \frac{1}{(2\,\theta - 1)} \left(\frac{n_1}{n}\right) - \left(\frac{1-\theta}{2\,\theta - 1}\right), \text{ for } \theta \neq \frac{1}{2} \tag{11.11}$$

(d) In a sample of 100 students, suppose that 71 said that their cards were *correct.* Estimate p.

(e) Determine $\hat{V}(\hat{p})$.

$$\hat{V}(\hat{p}) = \frac{1}{(2\,\theta - 1)^2}\,\frac{1}{n}\,\left(\frac{n_1}{n}\right)\left(1 - \frac{n_1}{n}\right) \tag{11.12}$$

(f) Determine the bound on the error of estimation.

(g) Construct a 95% confidence interval on p, the true population proportion of students who used marijuana in the past 30 days.

Interpretation: We are 95% confident that the population proportion of students who used marijuana in the past 30 days is less than 26%.

(h) Suppose absolutely none of the students at the university used marijuana in the past 30 days. Then, n_1/n should be around what number? Is it possible that $n_1/n = 77/100$? Estimate p if $n_1/n = 77/100$.

(i) Suppose that in a simple random sample of size 100 **withOUT using a random-response model,** we find that 8 students admitted to using marijuana in the past 30 days. Determine the appropriate estimator of p along with its bound on the error of estimation.

□

Homework p. 367: Exercise 11.14

11.7 Imputation

Many surveys have missing data. Why is this a problem?

Types of missingness are:

- Missing Completely at Random (MCAR) - The events that lead to any particular item being missing are independent both of observable variables and of unobservable parameters of interest.

- Missing at Random (MAR) - What caused the data to be missing does not depend upon the actual missing data.

- Not Missing at Random (NMAR) - Data are missing for a specific reason.

□

When a respondent omits some but not all of the questions on the questionnaire, sometimes the missing data might be **imputed,** rather than having the entire respondent's record discarded.

For example, consider a respondent who forgets to list his/her gender, but lists his/her occupation as a nurse.

A useful aid in *R* is the `mice` package; multiple imputation by chained equations.