

Math 336

Spring 2008

This question relates to survey research methods. Researchers almost always promise confidentiality to subjects in a survey setting and sometimes they will even take steps to assure complete anonymity. With particularly sensitive questions – such as “Have you had unsafe sex in the past year?”, or “Have you cheated on an exam since being in college?” – even these safeguards may not give the researcher confidence that honest answers are likely to be given. A method that goes even further in allaying a subject’s fear that their responses will be somehow identified with them is called the *randomized response technique*. I will describe this method and then ask some questions about the method.

A simple random sample with replacement is taken of a population of individuals. Of interest is the proportion s of the population with a certain characteristic – a characteristic that is of a sensitive or, perhaps, stigmatizing nature. First each subject, in private, flips a penny and also flips a dime. Assume that the penny and dime are each “fair,” i.e., the probability of “heads” is .5 as is the probability of “tails.” Only the subject knows the outcome of his/her coin flips. Now the instructions to each subject are this: “if your penny landed ‘heads’ ” answer this question: “Did your dime land ‘heads’?” But if the penny landed “tails,” answer this question: “Have you engaged in unsafe sex in the past year?”

Define the random variable X as 1 if a subject’s final response to the researcher is “yes” and 0 if it is “no.” Notice that with this scheme, the researcher cannot tell if a “yes” is because the subject had unsafe sex or because the subject’s dime landed heads. This is the extra assurance of anonymity in the method. Because we have sampled *with replacement*, you can assume here that X_1, X_2, \dots, X_n are independent and identically distributed Bernoulli trials. (With replacement is a bit unrealistic, but it makes the problem easier.)

Answer these questions, where s now denotes the proportion in the population who have engaged in unsafe sex in the past year, and assuming that all subject responses are truthful.

- (a) Let $P(X_i = 1) = p$. Find an expression for p in terms of s .
- (b) Solve the equation found in (a) for s .
- (c) Use (b) and a best estimator for p to find an unbiased estimator for s , and prove that the estimator is unbiased.
- (d) Find the variance for the estimator found in (c) in terms of p .
- (e) Make an argument, through sufficiency, that the estimator you found in (c) is a best estimator.
- (f) R. Wheeler and S. Bacher, two Oberlin students, conducted this very survey using this very method in 1990 to a random sample of 49 students. There were 29 “yes” responses in their 49. Estimate the proportion of Oberlin students at the time who had engaged in unsafe sex during the previous year. Use (d) and the value of p that would make the variance as large as possible to compute a conservative estimate for the 95% margin of error of your estimate.