Math 124 Lecture 17 part II

Dr Ben Bolstad

bolstad_math124@bmbolstad.com

Oct 13, 2004

Question 1

A poll is taken to determine the proportion of registered voters in a certain city that intend to vote to re-elect the current mayor. If we increase the size of our poll from 500 to 1000 what will we gain? Why might we not increase the size?

Answer: Increasing the sample size will reduce the variability of our estimate of the proportion of voters who wish to re-elect the current mayor. We might not increase our sample size because of the economic costs of taking a larger sample.

Question 2

A political scientist wants to know how college students feel about social security. She obtains a list of 3456 undergraduates at her college and mails a questionnaire to 250 students selected at random from the list. Only 104 questionnaires are returned. What is the population in this study? What is the sample? Will the researcher have any difficulties analyzing her study?

Answer: The population is the set of all undergraduates at the researchers college. The sample is the 250 students selected for the survey. The research is going to have problems analyzing her study because of the many surveys were not returned. We refer to this problem as non-response bias. It is possible that only people who hold strong opinions on the matter will reply to the questionnaire and the researcher may not get an accurate view of college student opinion on social security.

Question 3

Right after the polls close in an election a carefully designed exit poll by CNN of 1000 voters determines 54% for the Democrat and 46% for the Republican. Simulatenously it is announced that of the first 100,000 votes counted 52% were for the Republican and 48% for the Democrat. Should the early returns or the exit pool be trusted more? Explain why.

Answer: The early returns is based upon 100,000 votes so we would expect our estimate of the proportion of votes for a candidate to have low variance. But because votes are counted in a non-random way it is likely to be biased. For example smaller precincts may be counted first and these might be primarily rural and lean more Republican than larger urban precincts that take longer to count. Provided the exit poll has been carefully designed it will likely have a lower bias and should be preferred. You may check the margin of error by computing the variability of the estimate for the Democrat in the exit poll and seeing if it included 50%, if it doesn't you can reasonably secure that the exit poll is going to predict the winner. A 95% CI would give us a margin of error.

Question 4

At a party there are 30 students 21 years old or older and 20 students under 21 years old. A research chooses 3 students from the 21 or older age group and 2 students from the under 21 age group to interview about their attitudes to alcohol. Does every student have equal probability of being selected? Is this a Simple Random Sample?

Answer: The probability of selecting any one of the over 21 year olds is $\frac{3}{30} = \frac{1}{10}$ and the probability of selecting any one of the under 21 year olds is $\frac{2}{20} = \frac{1}{10}$ so every individual has equal probability of being selected. One of the conditions of a Simple Random Sample is that every individual has equal probability of being selected. However a Simple Random Sample would allow us to pick any 5 students (ie we could, by chance, pick 5 from the over 21 age group). Instead we refer to our sampling situation as Stratified Random Sampling. The two strata in this case are the age groups. Within each age group we have simple random samples.

Question 5

Telephone surveys. Which of the following two method should be preferred and why? Is either perfect.

  a A sample of households in a community is selected at random from the telephone directory.

  b After selecting an exchanges (or set of exchanges) the final four digits are selected at random.

Answer: The problem with (a) is that it misses people who have unlisted numbers. This introduces a bias, in particular we would call this an under-coverage bias. Our sample is missing part of the population (the set of all households in the community. (b) is better because we no longer miss people with unlisted numbers. However we also have a second under-coverage bias, in particular household with no telephones. Neither (a) or (b) would include these households.

Historical example of sampling gone wrong

A famous example that illustrates the problems of the situation in Questions 3 and Questions 5 is the 1948 US Presidential election. There is a famous picture showing President Truman holding a newspaper with the headline "Dewey Defeats Truman". As we all know Truman won the election. Polling leading up to the election had been conducted using telephones and many Democratic households at the time did not have telephones and so were missed by the polling organizations. In addition the early returns were indicating a big win for Dewey so the Chicago Tribune published the ultimately incorrect and now famous headline.