Confidence Intervals and Sampling Design

Lecture Notes VI

Statistics 112, Fall 2002
Announcements

• For homework question 3(b), assume that the true $p$ is expected to be about $\frac{179}{220} = 0.814$ in calculating the sample size required to estimate $p$ within $\pm 0.02$ with 95% confidence (see today’s lecture notes).

• Friday’s problem sessions:
  – 12-12:50. Huntsman Hall, F96 (except for Friday, October 11, Huntsman Hall, F94 and Friday, November 8, Huntsman Hall, F92).
Outline

- Confidence intervals for proportions.
- Choosing the sample size.
- Assumptions behind confidence intervals and hypothesis tests.
- Sampling design and study design.
Examples

- Effects of enzyme exposure on respiratory function.

- New York Times poll.
  - For a sample of 0's and 1's,

\[ s = \sqrt{\hat{p}(1 - \hat{p})}, \quad \frac{s}{\sqrt{n}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]
Choosing the Sample Size

- A wise user of statistics never plans data collection without at the same time planning the inference. You can arrange to have both high confidence and a small margin of error.

- The margin of error of the confidence interval $\bar{x} \pm z^* \sigma / \sqrt{n}$ for the mean of a normal population is

$$z^* \frac{\sigma}{\sqrt{n}}$$

To obtain a desired margin of error $m$, just set this expression equal to $m$, substitute the value of $z^*$ for your desired confidence level, and solve for the sample size $n$.

- The confidence interval for a population mean will have a specified margin of error $m$ when the sample size is

$$n = \left( \frac{z^* \sigma}{m} \right)^2$$

- Notes:
  - In order to halve the margin of error (i.e., double the accuracy), you need to quadruple the sample size.
As long as the population is much larger than the sample size (say 50 times as large as the sample size), only the size of the sample affects the margin of error, not the size of the population (i.e., a survey of 1000 people in Delaware is just as accurate as a survey of 1000 people in Pennsylvania.)

- If you are trying to obtain a confidence interval for a proportion $p$ and want to find the needed sample size to obtain a margin of error $m$, you can either (i) use $\sigma = 0.5$ (this is conservative because $\sigma$ cannot be greater than 0.5) or (ii) use $\sigma = \sqrt{p^* (1 - p^*)}$ where $p^*$ is the $p$ that you expect.

- Example: For the New York Times poll, how large a sample is required to estimate $p$ within $\pm 0.03$ with 95% confidence if $p$ is expected to be about 0.68?
Assumptions behind CIs and Hyp. Tests

- Both the confidence intervals and the error rates for the hypothesis tests we have developed are based on two assumptions:
  - The distribution of values in the population is normal.
  - The sample is a simple random sample from the population.

- The first assumption is not critical. For large sample sizes \((n \geq 40)\), the results will be approximately correct. For moderate sample sizes \((n \geq 15)\), the results will be approximately correct except in the presence of outliers or strong skewness. For small samples \((n < 15)\), the results will only be correct if the data are close to normal. You should always make graphs (e.g., boxplot, histogram, stem-and-leaf plot, normal quantile plot) to examine your data for skewness and outliers.

- The second assumption is critical. If the sample is not a random sample from the intended population, the results may be very untrustworthy.
Biased Samples

- The Literary Digest Poll. In the 1936 presidential election, the Literary Digest predicted an overwhelming victory for the Republican Alf Landon over the incumbent Democrat Franklin Delano Roosevelt. However, Roosevelt won the election by a landslide - 62% to 38%. What went wrong?

- The sample was taken by mailing questionnaires to 10 million people whose names and addresses came from sources like telephone books and club membership lists. 2.4 million people returned the questionnaires - the largest poll ever taken at the time.

- Selection bias: When the procedure for selecting a sample results in samples that are systematically different from the population, e.g., the mean of the sampling distribution does not equal the population mean.

- When a selection procedure is biased, taking a large sample does not help. This just repeats the basic mistake on a large scale.
The best methods for sampling involve the planned introduction of chance.

A probability sample gives each member of the population a known chance (greater than zero) to be selected.

A simple random sample (SRS) of size $n$ consists of $n$ individuals from the population chosen in such a way that every set of $n$ individuals has an equal chance to be the sample actually selected.

Random selection eliminates bias in the choice of a sample from a list of the population.

However, in practice, other sources of bias arise:
- Nonresponse bias.
- Undercoverage.

Review Section 3.3.
Even if the sample is representative of the intended population, the confidence intervals and hypothesis tests provide inferences for a population mean that may not answer the real question of interest.

- Do you think that the United States should *allow* public speeches against democracy?
- A random sample of 1000 Americans was taken and 62 percent said no (Rugg, 1941). A 95% confidence interval of 59 percent to 65 percent was found. Does this mean that the Americans are against free political speech? Not necessarily, when asked
  - Do you think the United States should *forbid* public speeches against democracy?
  - 46 percent said yes.

The Pepsi-Cola company carried out research to determine whether people tended to prefer Pepsi-Cola to Coca-Cola. Participants were asked to taste two glasses of cola and then state which they preferred. The two glasses were not labeled Pepsi or Coke for obvious reasons. Instead, the Coke glass was labeled Q and the Pepsi glass was labeled M.
• The results showed that “more than half choose Pepsi over Coke.” Is this conclusion warranted?

• Threats to validity of surveys/experiments for answering questions of real interest: wording of questions, response bias, placebo effect.

• Survey design: pay attention to wording of questions, develop good interviewing techniques.

• Study design to enhance validity of experiment/observational study: use the principles of comparative design to make sure that the treatment and control groups are treated the same in all respects except for the treatment, e.g., use placebos, double blinding.

• In the minimum wage study, Card and Krueger (1994) actually compared the mean difference in employment before and after the minimum wage increase between NJ and eastern PA fast food restaurants, reasoning that PA fast food restaurants are a control group that is affected by the same national and regional economic trends.
Question 7.28 in Moore and McCabe (third edition). The Acculturation Rating Scale for Mexican Americans (ARSMA) measures the extent to which Mexican Americans have adopted Anglo/English culture. During the development of ARSMA, the test was given to a group of 17 Mexicans. Their scores, from a possible range of 1.00 to 5.00, had $\bar{x} = 1.67$ and $s = 0.25$. Because low scores should indicate a Mexican cultural orientation, these results helped to establish the validity of the test (Based on I. Cuellar, L.C. Harris, and R. Jasso, “An acculturation scale for Mexican American normal and clinical populations,” *Hispanic Journal of Behavioral Sciences*, 2 (1980), pp. 199-217).

(a) Give a 95% confidence interval for the mean ARSMA score of Mexicans.

(b) What assumptions does your confidence interval require? Which of these assumptions is most important in this case?
A confidence interval is a range of plausible values for the population parameter. The confidence level of the interval indicates just how plausible this range is. The following is usually an approximate 95% confidence interval:

\[
\text{point estimate } \pm 2\text{se(estimate)},
\]

e.g., for the population mean in a one sample problem, \(\bar{x} \pm 2s/\sqrt{n}\) is an approximate 95% confidence interval.

Inferences from confidence intervals and hypothesis tests are based on the critical assumption that the sample is a random sample from the intended population. If the sample is biased, the inferences from a confidence interval or hypothesis test are not reliable.

The most reliable method of sampling involves the planned introduction of chance.

Consider carefully the connection between the question the study actually addresses and the question of real interest. Methods for enhancing the validity of studies include careful training of interviewers and comparative design.

Next time: Regression. How to predict \(Y\) based on \(X\). Read Section 2.3.