Lecture Notes

# Chapter 7: Estimation

## Learning Objectives

1. Explain the concept of estimation, point estimates, confidence level, and confidence interval.
2. Calculate and interpret confidence intervals for means.
3. Describe the concept of risk and how to reduce it.
4. Calculate and interpret confidence intervals for proportions.

## Chapter Outline

1. Introduction
   1. Procedures involved in estimating population means and proportions based on the principles of sampling and statistical inference.
   2. Knowledge about sampling distribution allows us to estimate population means and proportions from sample outcomes and to assess the accuracy of these estimates.
   3. **Estimation** is a process whereby we select a random sample from a population and use a sample statistic to estimate a population parameter.
   4. We can use sample proportions as estimates of population proportions, sample means as estimates of population means, or sample variances as estimates of population variances.
   5. The major objective of sampling theory and statistical inference is to provide estimates of unknown population parameters from sample statistics.
2. Point and Interval Estimation
   1. Estimates of population characteristics divided into two types:
      1. **Point estimates**: Sample statistics used to estimate the exact value of a population parameter.
      2. **Interval estimates**: Ranges of values within which the population parameter may fall; more accurate than point estimates.
   2. In interval estimation, we identify **confidence interval (CI)**, the range of values within which the population parameter may fall, that we identify in interval estimation. They are sometimes referred to in terms of **margin of error**, whichis simplythe radius of a confidence interval.
   3. Confidence interval (CI) defined in terms of confidence levels.
      1. **Confidence level:** While using confidence intervals to estimate population parameters, we can evaluate their accuracy by assessing the likelihood that any given interval will contain the mean. This likelihood, expressed as a percentage or a probability, is called a confidence level.
      2. Confidence intervals can be constructed for many different parameters based on their corresponding sample statistics.
3. Confidence Intervals for Means
   1. Assessing the needs of commuter students:
      1. Survey a random sample of 500 students to estimate the average commuting time of all 15,000 commuters on our campus--the population parameter.
         1. To obtain this estimate, we calculate the average commuting time for the sample. Suppose the sample average is *Y* = 7.5 hr/week.
         2. Because it is based on a sample, this estimate is subject to sampling error. It is unlikely that our sample mean, *Y* = 7.5 hr/week deviates much from the true population mean (Central limit theorem: For a large enough sample size, most sample means will tend to be close to the true population mean.)
      2. Sampling distribution of the mean = , with a population mean μ and a standard error .
4. Allows us to use the normal distribution to determine the probability that a sample mean will fall within a certain distance measured in standard deviation (standard error) units or *Z* scores of μor .
   * 1. Assumptions:
5. A total of 68% of all random sample means will fall within ±1 standard error of the true population mean.
6. A total of 95% of all random sample means will fall within ±1.96 standard errors of the true population mean.
7. A total of 99% of all random sample means will fall within ±2.58 standard errors of the population mean.
   1. Determining the Confidence Interval
      1. Calculate the standard error of the mean.
      2. Decide on the level of confidence, and find the corresponding *Z* value.
      3. Calculate the confidence interval.
      4. Interpret the results.
   2. Reducing Risk
      1. One way to reduce the risk of being incorrect is by increasing the level of confidence: This is a trade-off between achieving greater confidence in an estimate and the precision of that estimate, as the width of the confidence interval increases.
         1. Although using a higher level of confidence increases our confidence that the true population mean is included in our confidence interval, the estimate becomes less precise as the width of the interval increases.
      2. Review Table 7.1: Confidence Levels and Corresponding *Z* Values: It lists three commonly used confidence levels along with their corresponding *Z* values.
      3. Review Figure 7.4: It illustrates the relationship between the confidence level and the precision of the confidence interval.
   3. Estimating Sigma
      1. To calculate confidence intervals, we need to know the standard error of the sampling distribution.
      2. Standard error: . It is a function of the population standard deviation and the sample size.
      3. Standard error using sample standard deviation (*s*): .
   4. Sample Size and Confidence Intervals
      1. Researchers can increase the precision of their estimate by increasing the sample size.
         1. A more tightly clustered sampling distribution means that our confidence intervals will be narrower and more precise.
      2. Review Table 7.2: Ninety-Five Percent Confidence Interval and Width for Mean Number of Hours per Day Watching Television for Three Different Sample Sizes.
8. In Table 7.2, we summarize the 95% confidence intervals for the mean number of hours watching television for these three sample sizes: *N* = 150, *N* = 778, and *N* = 1,556.
9. Although precision of estimates increases steadily with sample size, the gains would appear to be rather modest after *N* reaches 1,556. Researchers have to consider at what point the increase in precision is too small to justify the additional cost associated with a larger sample.
   * 1. Inverse relationship between sample size and width of confidence interval.
     2. Increase in sample size linked with increased precision of confidence interval. Researchers can increase precision of estimate and CI by increasing sample size.
10. Statistics in Practice: Hispanic Migration and Earnings
    1. Marta Tienda and Franklin Wilson’s study of discrepancy in earnings between Hispanic population groups.
       1. The gap in earnings has been attributed mainly to differences in migration status and in the level of education.
       2. Marta Tienda and Franklin Wilson argued that Mexicans, Puerto Ricans, and Cubans varied markedly in socioeconomic characteristics because of differences in the timing and circumstances of their immigration to the United States.
       3. Tienda and Wilson also noted persistent differences in educational levels among Mexicans and Puerto Ricans compared with Cubans, which were likely to be reflected in disparities in earnings among the three groups.
       4. We would anticipate that the earnings of Cubans would be higher than the earnings of Mexicans and Puerto Ricans.
       5. Review Figure 7.6: The confidence intervals for mean annual income of Cuban, Puerto Rican, and Mexican immigrants are illustrated in this figure. We can say with 95% confidence interval that the true income mean for each Hispanic group lies somewhere within the corresponding confidence interval. Note that the confidence intervals do not overlap, thus revealing great disparities in earnings among the three groups. Highest interval estimates are for Cubans, followed by Mexicans and then Puerto Ricans.
11. Confidence Intervals for Proportions
    1. Sample proportions or percentages are usually reported along with a margin error, plus or minus a particular value.
    2. Margin of error is the confidence interval and is used to estimate population proportions or percentages.
    3. The same conceptual foundations of sampling and statistical inference that are central to the estimation of population means, the selection of random samples and the special properties of the sampling distribution, are also central to the estimation of population proportions.
    4. Sampling distribution of proportions underlies the estimation of population proportions from sample proportions.
    5. With sufficient sample size:
       1. Sampling distribution of proportions is approximately normal.
       2. Mean μ*p* equal to the population proportion π*.*
       3. Standard error of proportions: Here = the standard error of proportions, π= the population proportion, and *N* = the population size.
       4. Estimated standard error:  Here, *s*p= the estimated standard error of proportions, *p* = the sample proportion, and *N* = the sample size.
    6. Determining the Confidence Interval
       1. General formula for constructing CI for proportions for any level of confidence: CI = *p* ± *Z*(*s*p):
          1. Here, CI = the confidence interval, p = the observed sample proportion, *Z* = the *Z* corresponding to the confidence level, and *s*p = the estimated standard error of proportions.
          2. To obtain a confidence interval at a certain level, we take the sample proportion and add to or subtract from it the product of a *Z* value and the standard error.
       2. The *Z* value we choose depends on the desired confidence level. We want the area between the mean and the selected ±*Z* to be equal to the confidence level.
       3. Steps to determine the confidence interval for a proportion:
12. Calculate the estimated standard error of the proportion.
13. Decide on the desired level of confidence, and find the corresponding *Z* value.
14. Calculate the confidence interval.
15. Interpret the results.
16. Reading the Research Literature: Women Victims of Intimate Violence
    1. Janet Fanslow and Elizabeth Robinson studied help-seeking behavior and motivation among women victims of intimate partner violence in New Zealand:
       1. Relied on data from the New Zealand Violence against Women Study to document the reasons why victims sought help or left their partner due to domestic violence.
       2. Identified the categories with the highest percentages for each sample: for seeking help (could not endure more, encouraged by friends or family, children suffering and badly injured) and for leaving (could not endure more, he threatened to kill her, children suffering and badly injured).
       3. Concern for children suffering was also identified as an important reason for female victims to seek help and/or to leave their abuser.
       4. When estimates are reported for subgroups, the confidence intervals are likely to vary. Even when a confidence interval is reported only for the overall sample, we can easily compute separate confidence intervals for each of the subgroups if the confidence level and the size of each of the subgroups are included.