Lecture Notes

# Chapter 19: Inferential Statistics

## Learning Objectives

* 1. Explain the difference between a sample and a population.
  2. Explain the difference between a statistic and a parameter.
  3. Recognize the symbols used for the mean, variance, standard deviation, correlation coefficient, proportion, and regression coefficient.
  4. Provide the definition of sampling distribution.
  5. Compare and contrast point estimation and interval estimation.
  6. Explain how confidence intervals work over repeated sampling.
  7. List and explain the steps in hypothesis testing.
  8. Explain the difference between the null hypothesis and the alternative hypothesis.
  9. Explain the difference between a nondirectional and a directional alternative hypothesis.
  10. Explain the difference between a probability value and the significance level.
  11. Draw the hypothesis-testing decision matrix and explain the contents.
  12. State how to decrease the probability of Type I and Type II errors.
  13. Explain the purpose of hypothesis testing.
  14. Explain the basic logic of significance testing.
  15. Explain the different significance tests discussed in the chapter.
  16. Explain the difference between statistical and practical significance.
  17. Explain what an effect size indicator is.

## Chapter Summary

Chapter 19 focuses on inferential statistics. The chapter discusses sampling distributions, estimation, and hypothesis testing in theory and in practice.

## Annotated Chapter Outline

1. Introduction
   1. Chapter 18 focused on descriptive statistics that researchers use to describe the numerical characteristics of their data.
   2. Chapter 19 focuses on inferential statistics that researchers use to attempt to go beyond their data to make inferences about populations based on samples. The chapter discusses how these inferences are made and why.
   3. Inferential statistics: use of the laws of probability to make inferences and draw statistical conclusions about populations based on sample data. Four important points:
      1. Distinction between sample and population is essential
         1. **Sample:** a set of cases taken from a larger population
         2. **Population:** the complete set of cases
      2. A **statistic** is a numerical characteristic of a sample, and a **parameter** is a numerical characteristic of a population.
      3. In inferential statistics, researchers study samples when they are actually more interested in the population.
         1. Cannot study the population directly
         2. Sometimes conclusions based on sample are not applicable to population.
      4. Random sampling is assumed in inferential statistics.
      5. Discussion Question: Explain why researchers differentiate between the sample and the population.
   4. Table 19.1: Statisticians use Greek letters to symbolize population parameters (i.e., numerical characteristics of populations, such as means and correlations) and English letters to symbolize sample statistics (i.e., numerical characteristics of samples, such as means and correlations).
      1. Discussion Question: Why do you think that statisticians use Greek and English letters when dealing with statistics?
2. Sampling Distributions: the theoretical notion of a sampling distribution is what allows researchers to make probability statements about population parameters based on sample statistics.
   1. **Sampling Distribution:** the theoretical probability distribution of the values of a statistic that results when all possible random samples of a particular size are drawn from a population.
      1. Emerges from **repeated sampling** (drawing many or all possible samples from a population)
         1. Each sampling would provide slightly different values for the statistic due to chance.
         2. Use known rules of probability to make decisions with inferential statistics.
      2. A sampling distribution can be constructed for any sample statistic.
         1. Statisticians have developed sampling distributions for all common statistics so researchers do not need to.
         2. Researchers draw a sample from the population and then uses a computer program to analyze the data.
         3. The computer program uses sampling distributions to determine certain probabilities.
         4. Sampling distribution is a theoretical distribution because there is a sampling distribution underlying each inferential statistical procedure.
      3. **Sampling error:** the difference between the value of a sample statistic and the corresponding population parameter.
         1. Statistics from random samples vary because of chance fluctuations.
         2. **Standard error:** the standard deviation of a sampling distribution; a lot of sampling error leads to large standard errors and little sample error leads to small standard errors.
      4. In a constructed sampling distribution, the average of the values of the sample statistic is equal to the population parameter.
         1. Sometimes sample statistic will overestimate parameter and sometimes sample statistic will underestimate the population parameter.
      5. Discussion Question: Discuss how the information about sampling distributions, sampling error, and standard error are related to inferential statistics.
   2. **Sampling Distribution of the Mean:** the theoretical probability distribution of the means of all possible random samples of a particular size drawn from a population.
      1. Book example of annual income in multiple random samples of the population.
      2. See Figure 19.1.
      3. If you take all possible random samples and calculate the mean of each sample, the means will fluctuate randomly around the population mean, and they will form a normal distribution.
      4. The average of all of the sample means will equal the true population mean.
      5. Discussion Question: Why is the sampling distribution of the mean important in research.
3. Estimation: Researchers use inferential statistics to make an estimation of a population parameter: “Based on my random sample, what is my estimate of the population parameter?”
   1. **Point estimation:** the use of the value of the sample statistic as the estimate of the value of a population parameter.
      1. **Point estimate:** the estimated value of a population parameter
      2. Point estimates are rarely the same as the population parameter due to sampling error.
      3. Because of sampling error, many researchers recommend using interval estimation.
   2. Interval estimation: also known as confidence intervals; they are constructed.
      1. **Confidence interval:** A range of numbers inferred from the sample that has a certain probability or chance of including the population parameter.
      2. **Confidence limits:** the end points of a confidence interval
      3. **Lower limit:** the smallest number of a confidence interval
      4. **Upper limit:** the largest number of a confidence interval.
      5. **Level of confidence:** the probability that a confidence interval to be constructed from a random sample will include the population parameter.
      6. Figure 19.2: Hypothetical sampling distribution of the mean
         1. Ninteen of the 20 confidence intervals included the population mean or 95% of the time the confidence intervals captured the population parameter.
         2. 99% confidence interval is wider than 95% confidence interval and therefore less precise.
         3. Increasing the sample size leads to higher levels of confidence and a more narrow (or precise) confidence interval.
      7. Confidence Interval = point estimate ± margin of error.
         1. **Margin of error:** one half the width of a confidence interval
         2. Typically calculated by statistical computer programs.
         3. Confidence interval is constructed by taking the point estimate and surrounding it by the margin of error.
      8. Discussion Question: Compare and contrast point and interval estimation. How are they both related to inferential statistics?
4. Hypothesis testing: The researcher states his or her null and alternative hypotheses and then uses inferential statistics on a new set of data to determine what decision needs to be made about these hypotheses.
   1. **Hypothesis testing**: the branch of inferential statistics that is concerned with how well the sample data support a null hypothesis and when the null hypothesis can be rejected.
      1. Key question that is answered in hypothesis testing: “Is the value of my sample statistic unlikely enough (assuming that the null hypothesis is true) for me to reject the null hypothesis and tentatively accept the alternative hypothesis?”
      2. Goal of hypothesis testing is to help the researcher make a probabilistic decision about the truth of the null and alternative hypotheses.
      3. Discussion Question: Discuss why hypothesis testing is important.
   2. Null and Alternative Hypotheses
      1. Starting point of hypothesis testing is stating the null and alternative hypotheses.
         1. **Null Hypothesis** (H0): a statement about a population parameter. In most educational research, the null hypothesis predicts no difference or relationship in the population.
         2. Null hypothesis is tested directly using probability theory so sometimes called “null hypothesis significance testing.”
         3. If results from research study are very different from what is expected under the assumption that the null hypothesis is true, the researcher rejects the null hypothesis and tentatively accepts the alternative hypothesis.
         4. **Alternative hypothesis** (H1): statement that the population parameter is some value other than the value stated by the null hypothesis.
         5. Null and alternative hypotheses are contradictory because they cannot both be true at the same time.
         6. Alternative hypothesis is almost always more consistent with the researcher’s research hypothesis, therefore the researcher hopes to support the alternative hypothesis.
         7. Discussion Question: Discuss the role of the null and alternative hypotheses in hypothesis testing.
      2. Table 19.2: several examples of research questions, null hypotheses, and alternative hypotheses.
         1. When students look at the table make sure they notice that the null hypothesis has the “equal to” sign in it but can also have “less than or equal to” or “greater than or equal to” and the alternative hypothesis has the “not equal to,” “less than,” or “greater than” sign in it.
         2. They can also see in the table that hypotheses can be tested for many different kinds of research questions such as questions about means, correlations, and regression coefficients.
   3. Directional Alternative Hypotheses
      1. **Nondirectional alternative hypothesis:** an alternative hypothesis that includes the not equal sign (≠).
      2. **Directional alternative hypothesis:** an alternative hypothesis that contains either a greater than sign (>) or a less than sign (<).
      3. Major drawback of using directional alternative hypotheses is that if the researcher uses a directional alternative hypothesis and a large difference is found in the opposite direction, the researcher must conclude that no relationship exists in the population.
         1. As a result, researchers may state directional research hypothesis, but they test nondirectional alternative hypothesis, so the discovery function of science is still intact.
         2. If a researcher uses a directional alternative hypothesis, they must tell the reader.
      4. Discussion Question: When would a researcher want to use a directional or nondirectional hypothesis?
   4. Examining the Probability Value and Making a Decision
      1. When researchers state a null hypothesis, they can use the principles of inferential statistics to construct the probability model about what would happen if the null hypothesis were true.
         1. This probability model is the sampling distribution that would result for the sample statistic over repeated sampling if the null hypothesis was true.
         2. Computer program automatically selects the correct sampling distribution.
         3. Computer analyzes the research data and provides a **probability value or *p* value** (the probability of the observed result of your research study or a more extreme result if the null hypothesis was true).
         4. This value is used to make a decision about the null hypothesis. If the probability values are very small, the researcher rejects the null hypothesis and accepts the alternative hypothesis and claim the finding is **statistically significant** (claim made when the evidence suggests an observed result was probably not due to chance). Researchers claim that a finding is statistically significant when they do not believe (based on the evidence of their data) that their observed result was due only to chance or sampling error
         5. If probability value is small (≤.05), sample result is unlikely (assuming that the null hypothesis is true). If probability value is large (≥.05), sample result is not unlikely.
      2. If probability value is ˃.05, researcher would fail to reject the null hypothesis. The researcher would also claim that the difference between the sample means (or other statistic) is not statistically significant.
      3. **Significance level or alpha level:** The cutoff the researcher uses to decide when to reject the null hypothesis, value with which the researcher compares the probability level.
         1. When the probability level is less than or equal to the significance level, the researcher rejects the null hypothesis.
         2. When the probability value is greater than the significance level, the researcher fails to reject the null hypothesis.
         3. Significance level does not have to be .05 but most researchers adopt it.
      4. Two rules of hypothesis testing
         1. Rule 1: If the probability value obtained from the computer printout and based on the research results is less than or equal to the significance level (usually set at .05), then reject the null hypothesis and tentatively accept the alternative hypothesis. Conclude that the observed relationship is statistically significant (the observed difference between the groups is not just due to chance fluctuations).
         2. Rule 2: If the probability value is greater than the significance level, then the researcher cannot reject the null hypothesis. The researcher can only claim to fail to reject the null hypothesis and conclusion that the relationship is not statistically significant (any observed difference between the groups is probably nothing but a reflection of chance fluctuations).
      5. Table 19.3: Steps in Hypothesis Testing.
      6. Discussion Question: Have students explain the process of hypothesis testing as discussed in the book and presented in Table 19.3.
   5. The Hypothesis-Testing Decision Matrix: Hypothesis testing is based on samples of data and relies on probability theory to inform decision-making process. As a result, decision-making errors will be made.
      1. Table 19.5 :The Four Possible Outcomes in Hypothesis Testing

|  |  |  |
| --- | --- | --- |
|  | The True (But Unknown) Status of the Null Hypothesis | |
| The Null Hypothesis Is True. (It Should Not Be Rejected.) | The Null Hypothesis Is False. (It Should Be Rejected.) |
| Your Decision[\*](#s9781483391588.i5514) | | |
| Fail to reject the null hypothesis | Type A correct decision! | Type II error (false negative) |
| Reject the null hypothesis | Type I error (false positive) | Type B correct decision! |

* + - 1. Type A correct decisions occur when the null hypothesis is true and you do not reject it (you fail to reject the null hypothesis)
      2. Type B correct decisions occur when the null hypothesis is false and your reject it. This is what researchers hope for.
      3. **Type I Error:** rejecting a true null hypothesis also called false positives because falsely concluded that there is a relationship in the population, error in claiming statistical significance.
      4. **Type II Error:** failing to reject a false null hypothesis, false negatives because falsely concluded there is no relationship in the population.
    1. Traditionally, researchers are more concerned with avoiding Type I errors than Type II errors.
       1. Significance level is defined as the probability of making a Type I error, the researcher is willing to tolerate.
       2. Willing to claim, incorrectly, that there is an effect only 5% of the time.
    2. Discussion Question: Compare and contrast the four possible outcomes of hypothesis testing.
  1. Controlling the Risk of Error
     1. Type I and Type II errors are inversely related: If you decrease the likelihood of a Type I error, you usually increase the probability of a Type II error. If you choose a smaller significance level and make the possibility of a false positive less likely, you make a false negative more likely.
     2. Solution is to include more participants. Larger sample sizes provide a test that is more sensitive or has more **power** (the likelihood of rejecting the null hypothesis when it is false).
     3. If sample size is sufficiently large and the researcher obtains statistical significance, the finding must also have **practical significance** (a conclusion made when a relationship is strong enough to be of practical importance)
        1. With large samples, small differences are often significant.
        2. **Effect size indicator:** a measure of the strength or magnitude of a relationship between the independent and dependent variable.
        3. Cohen’s standardized effect size, eta squared, omega squared, Cramer’s *V*, correlation coefficient squared.
        4. Statistical significance is not enough, must know about effect size or practical importance of finding.
     4. Discussion Question: Discuss why effect size and practical significance are important in hypothesis testing.

1. Hypothesis Testing in Practice: Research is all about testing hypothesis and reporting on statistically significant findings.
   1. **Significance testing:** a commonly used synonym for hypothesis testing.
      1. When hypothesis testing, researchers are also checking for statistical significance.
      2. Set significance level, obtain probability level, determine whether Rule 1 (probability level is less than equal to significance level, reject null hypothesis, and conclude that finding is statistically significant) or Rule 2 (probability level is greater than significance level, fail to reject the null hypothesis, and conclude finding is not statistically significant).
      3. Discussion Question: Discuss the steps to be followed in significance testing.
      4. The book works through examples here using the data from the college students in Table 18.1.
   2. ***t* Test for Independent Samples:** statistical test used to determine whether the difference between the means of two groups is statistically significant.
      1. Used with quantitative dependent variable and dichotomous (two levels or groups) independent variable.
      2. *t* distribution similar to the normal curve but with smaller samples *t* distribution is flatter and more spread out than the normal curve, mean is 0.
      3. Reject the null hypothesis when *t* is large (falls in one of the two tails of the *t* distribution).
      4. Values greater the +2.00 or less than –2.00 are considered large *t* values.
      5. Discussion Question: Explain why *t* test scores greater than +2.00 and less than –2.00 are considered large *t* values.
   3. **One-Way Analysis of Variance:** statistical test used to compare two or more group means (ANOVA)
      1. Appropriate with quantitative dependent variable and one categorical independent variable. Two-way analysis of variance is used with two categorical independent variables, and so on.
      2. Based on a *F* distribution (see Figure 18.6c) which is skewed to the right.
      3. *F* test is the way ANOVA is reported.
      4. Discussion Question: Describe studies that could be analyzed by an ANOVA.
   4. Post Hoc Tests in Analysis of Variance
      1. One-way ANOVA tells the researcher whether the relationship between the independent and dependent variables is statistically significant but does not tell you which means are significantly different. If there are only two levels of the independent variable, you can tell the means are different. However, if there are three or more levels of the independent variable, you must conduct post hoc tests.
      2. **Post hoc test:** a follow-up test to the analysis of variance
      3. Multiple post hoc tests: Tukey test, Bonferroni test, Sidak test.
         1. If there are only three groups, the LSD test is the most powerful.
      4. Discussion Test: Explain why no post hoc tests are needed when there are only two levels of the independent variable.
   5. ***t* Test for Correlation Coefficients**: statistical test used to determine whether a correlation coefficient is statistically significant.
      1. Correlation coefficients are used to show the relationship between a quantitative dependent variable and a quantitative independent variable.
      2. Same *t* distribution as discussed before.
      3. Cohen’s rules for interpreting the size of correlation coefficients
         1. *r* = .1, weak correlation
         2. *r* = .3, medium correlation
         3. *r* = .5, large correlation
      4. Probability values from printout are investigated for significance.
      5. Discussion Question: Generate ideas when correlation coefficients might be used to test hypotheses.
   6. ***t* Test for Regression Coefficients**: Statistical test used to determine whether a regression coefficient is statistically significant.
      1. Simple regression is used to test the relationship between one quantitative dependent variable and one independent variable.
      2. Multiple regression is used to test the relationship between one quantitative dependent variable and two or more independent variables.
      3. Discussion Question: Compare and contrast the steps to take when trying to determine whether a regression coefficient is statistically significant for simple regression versus multiple regression.
   7. **Chi-Square Test for Contingency Tables:** statistical test used to determine whether a relationship observed in a contingency table is statistically significant.
      1. Used to see whether a relationship observed in a contingency table is statistically significant.
      2. Chi-square is used when both variables are categorical.
      3. Cramer’s *V* is the effect size indicator for chi-square that is interpreted similarly to a correlation coefficient.
      4. Discussion Question: Discuss the steps to take to see whether the relationship observed in a contingency table is statistically significant.
   8. Other Significance Tests
      1. The logic of significance testing: understanding and following the steps shown in Table 19.3.
      2. The processes used in this chapter apply to any significance testing.
         1. State the null and alternative hypothesis
         2. Determine the probability value and compare to significance level
         3. Decide whether the finding is statistically significant or not statistically significant.
         4. Obtain a measure of effect size, interpret the results, and make a judgment about practical significance.
      3. Discussion Question: Discuss the steps to be taken to determine statistical significance.