Chapter 13
Inferential Statistics in Healthcare

Key Terms

Analysis of variance (ANOVA)
Chi square
Confidence interval
Inferential statistics
Null hypothesis
Standard error of the mean
***t* test**
Type I error
Type II error

Objectives

At the conclusion of this chapter, you should be able to:

- Define inferential statistics
- Interpret the standard error of the mean and confidence intervals
- Identify and describe the null hypothesis
- Understand the importance of *t* tests
- Interpret ANOVA
- Understand the significance of chi square

Inferential Statistics

Inferential statistics allow us to generalize from a sample to a population with a certain amount of confidence regarding our findings. Without inferential statistics, it would be very difficult, short of conducting a census, to describe the characteristics of a population. The ability to interpret inferential statistics requires a pre-existing understanding of descriptive statistical measures and computations.

Errors in sampling procedures are inevitable; even with random sampling, there is no guarantee that the sample drawn will be representative of the population. For this reason, inferential statistics are limited to certain amounts of confidence.

This chapter discusses the interpretation of common inferential statistical measures, including standard error of the mean, confidence intervals, the null hypothesis, *t* tests, ANOVA, and chi square.

It is not the intention of this chapter to replace your general statistics course requirement. This chapter merely provides a review of concepts you have learned in that course. An exhaustive explanation of these concepts, including mathematical computation, is beyond the scope of this book.

Standard Error of the Mean and Confidence Intervals

When trying to determine the characteristics of a population, the mean from one sample may not be sufficient because of random sampling error. A more accurate representation could be found by taking many large samples, calculating the mean for each sample, and then finding the standard deviation of all the sample means. This value is called the **standard error of the mean,** abbreviated SE_m, and comes from the standard deviation of many sample means. The graphical representation of all the sample means is a normal distribution when the sample size is reasonably large.

Finding the SE_m in the manner described above can be tedious, and it may not be realistic to find the SE_m of the sample means. For this reason statisticians have found a formula for approximating the SE_m from only one random sample ($SE_m = sd/sqrt(n)$). (A discussion of the theory leading to this formula is not covered in this book.) When the SE_m has been computed, it is generally reported in one of two ways.

1. The data may be displayed in research reports as follows:

   $$\overline{X} = 150,\ sd = 10,\ n = 75,\ SE_m = 1.15$$

   Where:

   $\overline{X}$ = mean

   sd = standard deviation

   n = number of observations

   SE_m = standard error of the mean

   Based on the characteristics of the normal curve and treating the SE_m as a margin of error, by adding the SE_m (1.15) to and subtracting it from the sample mean (150), it can be said that we are 68.3 percent sure that the interval

148.85 to 151.15 contains the population mean. This is called a **confidence interval** (abbreviated C.I.).

$$150 - 1.15 = 148.85$$

$$150 + 1.15 = 151.15$$

Similarly, it can be said that we are 95.5 percent sure that the interval 147.70 to 152.30 contains the population mean.

$$150 - (2 \times 1.15) = 147.70$$

$$150 + (2 \times 1.15) = 152.30$$

2. The data may be displayed in a research report as follows:

$$\overline{X} = 60,\ sd = 4,\ n = 16,\ SE_m = 1$$

$$68.3\%\ \text{C.I.} = 59 \text{ to } 61$$

$$95.5\%\ \text{C.I.} = 58 \text{ to } 62$$

$$99.7\%\ \text{C.I.} = 57 \text{ to } 63$$

In this example, the confidence intervals have already been computed.

In a normal distribution, confidence intervals are computed by adding one standard error of the mean to and subtracting one from the mean, then adding two standard errors of the mean to and subtracting two from the mean, and so on. The percentages above come from a normal distribution:

68.3 percent of all scores in a normal distribution fall within + or – 1 standard deviation away from the mean

95.5 percent of all scores in a normal distribution fall within + or – 2 standard deviations away from the mean

99.7 percent of all scores in a normal distribution fall within + or – 3 standard deviations away from the mean

A confidence interval is the range of scores in which we are estimating the population mean to be. In the example above, the confidence intervals (C.I.) let us know, with a certain amount of confidence, where the mean will lie.

It is important to note that the larger the sample size, the more confidence there is in the mean and the smaller the chance for sampling errors, meaning a smaller SE_m. Further, the more homogenous a population (less variation), the smaller the SE_m. In fact, if a population were all the same (no variation) the SE_m would be zero.

The Null Hypothesis

The **null hypothesis** states that the difference between two population means is zero. This statement is generally made based on two sample means. For example, let's suppose we obtain a sample of the heights of 15-year-old girls at high school A and a second sample of the heights of 15-year-old girls at school B. These are the means:

High school A: $\bar{x} = 5.7$ ft.

High school B: $\bar{x} = 5.6$ ft.

From these results, it would appear that 15-year-old girls at high school A are taller than 15-year-old girls at high school B. However, this may not be true; in fact, the difference may be due to errors in sampling and the mean of the heights of girls at high school A (population mean) may equal the heights of girls at high school B (population mean). This is a statement of a null hypothesis.

The null hypothesis can be expressed as follows:

H_0: $\mu_1 - \mu_2 = 0$

H_0 represents the null hypothesis.

μ_1 represents the population mean for group 1.

μ_2 represents the population mean for group 2.

When conducting research, most researchers are searching for differences in population means and are therefore not looking to confirm the null hypothesis. In fact, most research is carried out to reject the null hypothesis. Therefore, researchers may develop an alternative hypothesis to test against the null. The alternative hypothesis can be expressed in one of three ways:

$H_1 = \mu_1 > \mu_2$

H_1 represents the alternative hypothesis.

μ_1 represents the population mean for group 1, hypothesized to be larger than group 2.

μ_2 represents the population mean for group 2, hypothesized to be smaller than group 1.

or

$H_1 = \mu_1 < \mu_2$

H_1 represents the alternative hypothesis.

μ_1 represents the population mean for group 1, hypothesized to be smaller than group 2.

μ_2 represents the population mean for group 2, hypothesized to be larger than group 1.

or

$H_1 = \mu_1 \neq \mu_2$

H_1 represents the alternative hypothesis.

μ_1 represents the population mean for group 1, hypothesized to be different from group 2 without sufficient information to determine which group is larger.

μ_2 represents the population mean for group 2, hypothesized to be different from group 1 without sufficient information to determine which group is larger.

The first alternative hypothesis states that group 1 is larger than group 2; this would be the same as the samples would imply, that 15-year-old girls from high school A are taller than 15-year-old girls from high school B. The second alternative hypothesis states that group 1 is smaller than group 2; this would indicate the samples do not represent the population, and that 15-year-old girls from high school A are not taller than 15-year-old girls from high school B. The third alternative hypothesis states that group 1 and group 2 are different, but there is not enough information to determine which population of girls has the higher mean height.

Because there is always a chance that the null hypothesis is true, testing it will lead to a probability (p) that it is true. The smaller the p value (probability that the null hypothesis is true), the more evidence that we should reject the null hypothesis. As a rule of thumb accepted by most statisticians in hypothesis testing, if the probability that the null hypothesis is true is 5 percent or less, the null hypothesis is rejected.

In hypothesis testing, rejecting the null hypothesis with this level of confidence is, in effect, describing the relationship between the means as statistically significant. This is generally the manner in which the null hypothesis is discussed in academic journals.

Probabilities of the null hypothesis are expressed as follows:

$p < 5\%$ The probability that the null hypothesis is true is less than 5 percent.

$p < 1\%$ The probability that the null hypothesis is true is less than 1 percent.

$p < 0.1\%$ The probability that the null hypothesis is true is less than 0.1 percent.

Where: p = the probability that the null hypothesis is true given the values present in the sample. Remember, we are generalizing from a sample to make conclusions about the population.

These different probabilities represent different statistical significance levels; the lower the percentage that the null hypothesis is true, the more significant the finding. Thus, a $p \leq 0.1\%$ is more significant a finding than a $p \leq 5\%$.

Dealing with levels of uncertainty in hypothesis testing creates two types of errors: Type I errors and Type II errors. A **Type I error** occurs when the null hypothesis is rejected, yet it is actually true. A **Type II error** occurs when the null hypothesis is not rejected, yet it is false. Following are examples of both types of errors.

Type I Error

According to the National Cancer Institute, a woman's chance of being diagnosed with breast cancer from age 50 through age 59 is 2.38 percent. With this knowledge, going to see a doctor because of a lump on her breast presents a risk that the doctor will diagnose the patient with breast cancer even if she does not have breast cancer. Consider the following null hypothesis: there is no difference between women diagnosed with breast cancer and women not diagnosed with breast cancer. Rejecting the null hypothesis in this instance would be assuming that the doctor will not diagnose her with breast cancer as there is a difference between the two populations (women diagnosed with breast cancer and women not diagnosed with breast cancer). However, if she is diagnosed with breast cancer, this would be an example of a Type I error. Simply said, the patient could receive a positive test result when, in fact, she does not have breast cancer.

Type II Error

A Type I error is controlled by the researcher setting the acceptable error rate. A Type II error is driven by the sample size and the particular test used. Suppose a drug company has developed a new drug for a serious disease. Suppose that, in reality, the new drug is effective. If, however, the null hypothesis is not rejected because the drug company selected a level of significance that is too high, the results of the study will have to be described as insignificant and the drug may not receive government approval (Pyrczak, 2003, 83). This is an example of a Type II error.

When describing the null hypothesis, it is never "accepted." Rather, we say "reject the null hypothesis" or "fail to reject the null hypothesis."

The *t* Test

A ***t* test** is an example of a test of the null hypothesis to determine if a set of results is statistically significant. As mentioned previously, for the results to be considered statistically significant, the probability that the null hypothesis is true should be 5 percent or less. The results of a *t* test can be written several ways. Consider the following example.

Example:

Group	$\bar{x}$	**Standard Deviation (*sd*)**	***n***
A	5.3	1.6	10
B	6.6	1.0	10

With this group of data, the results of the *t* test (computed using computer software or a calculator) can be written as follows:

The null hypothesis: There is no difference between the groups.

H_0: $\mu_1 - \mu_2 = 0$ (the difference between the means is zero)

1. Statistically significant ($t = 2.18$, $df = 18$, $p < 0.05$)

2. Statistically significant at the 0.05 level ($t = 2.18$, $df = 18$)

3. Reject H_0 at the 0.05 level ($t = 2.18$, $df = 18$)

Where: t = the value produced by the *t* test

df = The degrees of freedom are found by taking $n - 1$ for the number of groups. In the sample above, there are two groups, therefore:

$df = (n_1 - 1) + (n_2 - 1)$

$df = (10 - 1) + (10 - 1) = 18$

$df = 18$

Statements 1 through 3 above all reject the null hypothesis and say that there is a statistically significant difference between the population means of Group A and Group B ($5.3 - 6.6 = -1.3$).

However, the sample data may not support the conclusion that there is a statistically significant difference between the means. Consider the following example.

Example:

Group	$\bar{x}$	**Standard Deviation (*sd*)**	***n***
A	25	3	8
B	24	3	8

H_0: $\mu_1 - \mu_2 = 0$ (the difference between the means is zero)

1. Not statistically significant ($t = 0.67$, $df = 14$, $p > 0.05$)

2. H_0 is not rejected at the 0.05 level ($t = 0.67$, $df = 14$)

These statements do not reject the null hypothesis and say that the difference between the sample means is not statistically significant (25 – 24 = 1). The null hypothesis is not rejected because the probability of the null hypothesis actually occurring is greater than 5 percent; for this reason, the difference between the means is not statistically significant.

Example:

Group	$\bar{x}$	Standard Deviation (sd)	n
A	18	2.5	7
B	17	8	7

The null hypothesis: There is no difference between the groups.

H_0: $\mu_1 - \mu_2 = 0$ (the difference between the means is zero)

($t = 0.32$, $df = 12$, $p > 0.05$) Do not reject the null hypothesis.

Exercise 13.1

Given the information in the example above, would you reject the null hypothesis or not? How would you state your findings?

It is important to note that understanding the computation involved in testing the null hypothesis is reserved for an advanced statistics course and is not part of the daily activities of a healthcare practitioner. How to interpret the results of these tests, however, is important. For this reason, the next two sections on ANOVA and chi square are briefly introduced.

ANOVA

Although t tests are used to test the difference between two means, an **analysis of variance** (ANOVA) test is used to test the differences among more than two means. Sometimes ANOVA is referred to as an F test. In the special case when only two means are tested, the F test will yield a p-value that is the same as a t test for testing the difference between two population means.

In the following example, the means from four samples are listed. Using ANOVA will test to determine if any of six differences among the means is statistically significant.

	LOS Group 1: Patients with Private Insurance	LOS Group 2: Patients with Medicare/Medicaid
Males	$\bar{x} = 10$	$\bar{x} = 23$
Females	$\bar{x} = 15$	$\bar{x} = 17$

Males with Private Insurance (Group A): $\bar{x} = 11$

Males with Medicare/Medicaid (Group B): $\bar{x} = 19$

Females with Private Insurance (Group C): $\bar{x} = 17$

Females with Medicare/Medicaid (Group D): $\bar{x} = 20$

Differences tested by ANOVA:

1. Between A and B
2. Between A and C
3. Between A and D
4. Between B and C
5. Between B and D
6. Between C and D

This is the power of ANOVA, the ability to test many differences among means at once. If the results from an ANOVA are statistically significant (reject the null hypothesis that at least one of the pairs of means is different), the researcher's next step is to figure out what differences are significant. However, if the results are not statistically significant (the null hypothesis is not rejected), the work is done.

Chi Square

Chi square (represented by the symbol χ^2) is a test of significance that deals with nominal data and frequencies, specifically data where the standard deviation and mean are not meaningful descriptions. Note that calculation is possible, just not meaningful. Using computer software to conduct a chi square test is simple and quickly yields results. Consider the following example.

Example: Employees at a large healthcare facility were randomly asked whether they smoked and whether they had parents who smoked. The results of the 150 employees sampled were:

	Parents Who Smoke	Parents Who Do Not Smoke
Employee smokes	45	25
Employee does not smoke	30	50

These data in this table suggest that those employees with parents who did not smoke were likelier not to smoke (50 vs. 25) than were those employees with parents who smoked (30 vs. 45). From this sample alone, there appears to be a relationship between smoking and having parents who smoke. However, there is a possibility that the null hypothesis, that the relationship does not exist in this population, is true. A chi square test can determine whether the relationship is statistically significant. A chi square test using the values above yielded these results:

$$\chi^2 = 10.71,\ df = 1,\ p < 0.01$$

Based on the results, there is a less than 1 percent chance that the null hypothesis is true. Said another way, the results are statistically significant; there is a relationship between employees who do not smoke having parents who do not smoke.

Notice below that the degrees of freedom are determined differently for chi square. The degrees of freedom are found by taking $(r - 1) \times (c - 1)$ for the number of categories where r = the number of rows and c = the number of columns.

$$df = (r - 1)$$
$$df = (2 - 1) = 1$$
$$df = 1$$

Exercise 13.2

In the following two questions, identify which choices would be considered descriptive statistics and which would be considered inferential statistics.

1. Of 500 randomly selected people in New York City, 210 people had O+ blood.
   a. "42 percent of the people in New York City have O+ blood." Is the statement descriptive statistics or inferential statistics?
   b. "58 percent of the people of New York City do not have type O+ blood." Is the statement descriptive statistics or inferential statistics?
   c. "42 percent of all people living in New York State have type O+ blood." Is the statement descriptive statistics or inferential statistics?
2. On the last three Friday evenings, City Hospital diagnosed a number of heroin overdoses. There were four on January 14, 20XX, two on January 21, 20XX, and six on January 28, 20XX.
   a. "City Hospital averaged four heroin overdoses in their ER for the last three Friday evenings."
   b. "City Hospital never has more than six heroin overdoes on Friday evenings."
   c. "Friday nights are the busiest time for heroin overdoses at City Hospital."

Chapter 13 Test

1. Define inferential statistics.
2. Explain the difference between descriptive and inferential statistics.
3. Create a null hypothesis for the following research questions:
   a. What are the differences between emergency room shifts on medication errors?
   b. On a clinical trial of a new drug, what will be the effects over a currently used drug?
4. How could you reword these research hypotheses to null hypotheses?
   a. Smoking during pregnancy increases the risk of a child being born prematurely.
   b. Regular exercise in older adults decreases blood pressure levels.
5. Based on the results below, how would you describe the relationship between heart attack patients being treated with aspirin and whether or not they lived?

	Male Heart Attack Patients Treated with Aspirin	**Male Heart Attack Patients Not Treated with Aspirin**
Males who lived	15	4
Males who did not live	2	8

$$\chi^2 = 8.34, df = 1, p \le 0.01$$

(continued on next page)

Chapter 13 Test (continued)

6. Using the following information, determine if the differences between the means are significant.

   Medical Terminology Class

   Group A = Number of word elements remembered by students using flash cards

   Group B = Number of word elements remembered by students not using flash cards

Group	$\bar{x}$	**Standard Deviation (*sd*)**	***n***
A	75	2.0	10
B	50	3.5	10

   The null hypothesis: There is no difference between the groups.

   H_0: $\mu_1 - \mu_2 = 0$ (the difference between the means is zero)

   ($t = 2.50$, $df = 18$, $p < 0.05$)

7. Given the following p values, which would be considered more significant?
   a. $p \leq 0.3$
   b. $p \leq 0.02$
   c. $p \leq 0.25$

8. Given the following null hypothesis, give an example of a Type I error.

   H_0: There is no difference between the number of males or females who go to their primary care physician for an annual exam.

9. Given the following null hypothesis, give an example of a Type II error.

   H_0: There is no difference in the level of understanding of this chapter between students who have previously taken a statistics course and students who have not previously taken a statistics course.

10. Given the following information, determine the 68.3 percent, 95.5 percent, and 99.7 percent confidence intervals.

    $\bar{x} = 4.33$, $SE_m = 3$