Chapter 10 – Testing Claims Regarding a Parameter

The other branch of statistical inference is concerned with testing hypotheses about the value of parameters.

**Defn:** A **hypothesis** is a statement about the value of a population parameter.

**Defn:** In a hypothesis test, the hypothesis which the researcher is attempting to prove is called the **alternative (or research) hypothesis**, denoted $H_a$. The negation of the alternative hypothesis is called the **null hypothesis**, $H_0$. The null hypothesis usually represents current belief about the value of the parameter. The researcher suspects that the null hypothesis is false, and wants to disprove it.

**Note:** Hypotheses are tested in pairs (here the symbol $\mu_0$ represents some specified number):

- If the null hypothesis is $H_0$: $\mu = \mu_0$, then the corresponding alternative hypothesis is $H_a$: $\mu \neq \mu_0$.
- If the null hypothesis is $H_0$: $\mu \geq \mu_0$, then the corresponding alternative hypothesis is $H_a$: $\mu < \mu_0$.
- If the null hypothesis is $H_0$: $\mu \leq \mu_0$, then the corresponding alternative hypothesis is $H_a$: $\mu > \mu_0$.

In each case, the alternative hypothesis is a strict inequality; the null hypothesis is not. The above hypothesis statements are written in terms of a population mean, $\mu$. If we are testing hypotheses about a population proportion, $p$, or some other parameter, that parameter would appear instead of $\mu$. 
Example: If the researcher wants to prove that the proportion, \( p \), of the U.S. population who catch a cold between October 1 and March 31 is more than 40%, the two hypotheses would be stated as follows:

\[
H_0: \ p \leq 0.40 \quad H_a: \ p > 0.40
\]

Example: If the researcher is in charge of quality control for the Pepsi-Cola Company, and wants to determine whether an assembly line at a plant which produces 12-oz. cans of Pepsi is working properly, the hypotheses would be stated as follows in terms of the mean amount of Pepsi in the cans:

\[
H_0: \ \mu = 12 \text{ oz.} \quad H_a: \ \mu \neq 12 \text{ oz.}
\]

Note: We cannot prove the null hypothesis; we may disprove it, or we may fail to disprove it. After taking a sample and conducting a test, we will either reject \( H_0 \) and believe \( H_a \), or we will fail to reject \( H_0 \) because of insufficient evidence against it. As Dr. Carl Sagan once said, “Absence of evidence is not evidence of absence.”

There are four possibilities:
1) We reject the null hypothesis when the null hypothesis is, in fact, true. In this case, we make a Type I Error.
2) We reject the null hypothesis when the null hypothesis is, in fact, false. In this case, we make a correct decision.
3) We fail to reject the null hypothesis when the null hypothesis is, in fact, true. In this case, we make a correct decision.
4) We fail to reject the null hypothesis when the null hypothesis is, in fact, false. In this case, we make a Type II Error.

We want to maximize the probability of making a correct decision, and minimize the probability of making either a Type I Error or a Type II Error.
We use the following notation:
\[ P(\text{Type I Error}) = P(\text{Reject } H_0 \mid H_0 \text{ True}) = \alpha. \]
\[ P(\text{Type II Error}) = P(\text{Fail to Reject } H_0 \mid H_a \text{ True}) = \beta \]

We want both of these probabilities to be as small as possible. However, if we choose the value of \( \alpha \) to be smaller, the value of \( \beta \) will automatically become larger.

To see this, consider the analogy of the American criminal justice system. Suppose that you are the defendant in a murder trial. The hypotheses are
\[ H_0: \text{Defendant innocent} \quad \text{vs.} \quad H_a: \text{Defendant guilty} \]
The prosecutor is attempting to disprove the null hypothesis, thus proving the alternative hypothesis. To do this, the prosecutor presents evidence to the jury. The jury makes a decision on the basis of the evidence. The jury can decide to reject the null hypothesis, or to fail to reject the null hypothesis. In either case, the jury cannot know with absolute certainty that the decision is correct.

If the rules of evidence are changed to make it more difficult for the jury to convict (reject \( H_0 \)), then it becomes easier for the jury to acquit (fail to reject \( H_0 \)). Then it less likely that the jury will make a Type I Error, but more likely that the jury will make a Type II Error. Thus, \( \alpha \) has been made smaller, but \( \beta \) becomes larger.

If the rules of evidence are changed to make it easier for the jury to convict (reject \( H_0 \)), then it becomes more difficult for the jury to acquit (fail to reject \( H_0 \)). Then it more likely that the jury will make a Type I Error, but less likely that the jury will make a Type II Error. Thus, \( \alpha \) has been made larger, but \( \beta \) becomes smaller.
The only way that we can reduce both $\alpha$ and $\beta$ at the same time (make erroneous decisions less likely) is to gather more evidence.

In statistical hypothesis testing, “gather more evidence” means choose a larger sample from the population (choose a larger value of $n$).

We also begin our hypothesis test by choosing a value for $\alpha$, based on our assessment of the relative consequences of making a Type I Error, vs. the consequences of making a Type II Error. If we believe that the consequences of making a Type I Error are more serious than the consequences of making a Type II Error, we want to choose $\alpha$ to be rather small, perhaps 0.01 or 0.001. If we believe that the consequences of making a Type II Error are more serious than the consequences of making a Type I Error, we want to choose $\alpha$ to be rather large, perhaps 0.10. The most common choice for $\alpha$ is 0.05.

The evidence from the sample is used in the form of the value of a test statistic.

**Defn:** The test statistic is a random variable whose value depends on the sample data, such that the value of the statistic tells us whether to reject $H_0$ or to fail to reject $H_0$. We must choose the test statistic so that we know its sampling distribution under $H_0$.

**Defn:** The rejection region is the set of possible values of the test statistic which would lead us to reject $H_0$ and assert that we have proved $H_a$.

**Defn:** The significance level of the test is the probability, $\alpha$, of making a Type I Error.
**Defn:** The p-value of the test is the probability (under $H_0$) of obtaining a value of the test statistic at least as extreme as the one which is actually obtained from the data.

**Note:** To make our decision, we compare the computed p-value, obtained from the calculator, with our chosen value of $\alpha$. If the p-value is smaller than $\alpha$, we reject $H_0$ and say that we have proved $H_a$. If the p-value is larger than our chosen value of $\alpha$, then we fail to reject $H_0$.

**Types of Alternative Hypotheses: One-Tailed Tests or Two-Tailed Tests**

There are three forms for an alternative hypothesis:

1) $H_a$: $\mu \neq \mu_0$
2) $H_a$: $\mu < \mu_0$
3) $H_a$: $\mu > \mu_0$

The first form results in a two-tailed test, so called because the rejection region consists of two intervals, one associated with the left-hand tail of the probability distribution of the test statistic, and the other associated with the right-hand tail of the distribution.

The second form results in a one-tailed test, because the rejection region is an interval associated with the left-hand tail of the probability distribution of the test statistic.

The third form results in a one-tailed test, because the rejection region is an interval associated with the right-hand tail of the probability distribution of the test statistic.
**Steps in Hypothesis Testing**

**Step 1:** State the alternative hypothesis, $H_a$. (This is what we are trying to prove; it must be a strict inequality.) State the null hypothesis, $H_0$. (This is what we are trying to disprove; it cannot be a strict inequality. This statement will be the negation of the alternative hypothesis.)

**Step 2:** Choose the size of the random sample(s) to be selected from the population(s). Choose the significance level, $\alpha$, for the test. and the most common value is 0.05. In a real research situation, the researcher would choose the value based on an assessment of the relative seriousness of the consequences of making a Type I Error vs. a Type II Error.

**Step 3:** Choose the test statistic which will be used. In this course, if the hypotheses concern one or more population means, the test statistic will be a t-statistic. If the hypotheses concern a population proportion, the test statistic will be a Z-statistic. This step should be stated in the form of a sentence: “The test statistic is __________, which under $H_0$ has a _______ distribution (with ___ degrees of freedom).”

**Step 4:** Graph the distribution, indicating the rejection region, using the information from the previous three steps.

**Step 5:** Select the random sample from the population, gather the data, do the calculations of the value of the test statistic and the p-value. Make a decision based on a comparison of the p-value associated with the test statistic with the chosen level of significance. If $p$-value $< \alpha$, we reject $H_0$ and affirm $H_a$. Otherwise, we fail to reject $H_0$.

**Step 6:** State the conclusion of the hypothesis test.

If we reject the null hypothesis, the conclusion should have the following form: “We reject $H_0$ at the ($\alpha$) level of significance. We have sufficient evidence to conclude (statement of research conjecture).”
If we fail to reject the null hypothesis, the conclusion should have the following form: “We fail to reject $H_0$ at the ($\alpha$) level of significance. We do not have sufficient evidence to conclude (statement of research conjecture).”

**Example:** Suppose a consumer protection agency has received complaints that 24-oz. packages of Post Grape-Nuts contain less than 24 oz. of product. The agency wants to test this claim. They select a random sample of 100 boxes of cereal and measure the amount of product in each one. The sample average is $\bar{X} = 23.94$ oz., and the sample standard deviation is $S = 0.13$ oz. Test the claim at the 0.05 level of significance.

**Step 1:** $H_0$: $\mu \geq 24$ oz.  
$H_a$: $\mu < 24$ oz.

**Step 2:** $n = 100$, $\alpha = 0.05$

**Step 3:** The test statistic will be $t = \frac{\bar{X} - 24 \text{ oz.}}{\frac{S}{\sqrt{n}}}$. Under the null hypothesis, this statistic has a t-distribution with d.f. = 99.

**Step 4:** The alternative hypothesis says “less than.” Therefore the rejection region is a left-hand tail of the t-distribution. The area of this left-hand tail is 0.05, our chosen significance level. Draw the rejection region on the graph.

**Step 5:** We choose our random sample of 100 boxes, weigh the contents of each one. We obtain $\bar{X} = 23.94$ oz., and $S = 0.13$ oz.
Using the TI-83, choose STAT, then TESTS, then 2:T-Test. Choose Stats, rather than Data, since in this case we know the sample mean and sample standard deviation. Input the value of μ₀ appearing in the null hypothesis, in this case, 24. Input the sample mean and the sample standard deviation, and the sample size. Also, we must tell the calculator which of the three forms of alternative hypothesis we are using. In this case, we are using the second form. Choose Calculate. The value of the test statistic is –4.6154. The p-value associated with this test is 0.00000589. This is less than α, so we reject the null hypothesis.

Step 6: We reject H₀ at the 0.05 level of significance. We have sufficient evidence to conclude that the average amount of Grape Nuts in a 24-oz. box is less than 24 oz.

Example: p. 488, Exercise 16.

Tests of Hypotheses Concerning Population Proportions

The types of alternative hypotheses are as follows:

- Hₐ: p ≠ p₀
- Hₐ: p < p₀
- Hₐ: p > p₀

In order to conduct a test of hypotheses about the value of a population proportion, we must do a binomial experiment. We will choose a random sample of n members of the population; this will guarantee that the n trials in our experiment are independent of each other. For each member of the population we want to know the same information – does this member have the characteristic of interest; hence the n trials are also identical. For each trial, the information sought is in the form of a yes-or-no question – does the member have the characteristic of interest; hence there are only two
possible outcomes for each trial. Finally, since we are randomly sampling from the population, the probability that any one member of the sample will have the characteristic of interest is simply the proportion, p, of the entire population who have the characteristic of interest. Hence the probability of success is the same for each trial, namely p.

From the Central Limit Theorem, if our sample size is large enough, then the following random variable has an approximate standard normal distribution: 

\[ Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \]

The normal approximation will be reasonable if \( n > 20 \), and \( n \hat{p} > 5 \) and \( n(1-\hat{p}) > 5 \).

In the formula above \( \hat{p} \) is the proportion of the sample who have the characteristic of interest. Under the null hypothesis, this random variable is a statistic,

\[ Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \]

This will be the test statistic for our hypothesis test.

The hypothesis test will have the same steps as those used for testing hypotheses about population means:

**Step 1:** We state the null and alternative hypotheses.

**Step 2:** We state our chosen sample size and significance level.

**Step 3:** We state the form of the test statistic and its probability distribution under the null hypothesis.

**Step 4:** State the decision rule (drawing a graph of the distribution and indicating the rejection region).

**Step 5:** We choose our random sample, do our measurements, and compute the value of the test statistic and the associated p-value. We compare the p-value to make a decision.

**Step 6:** We state our conclusion, in terms of the statement we were trying to prove, and including the level of significance of the test.
Example: The U.S. Substance Abuse and Mental Health Services Administration conducts surveys on drug use by type of drug and age group. Results are published in National Household Survey on Drug Abuse. According to that publication, 12.0% of 18-25 year-olds were current users of marijuana or hashish in 1995. A recent poll of 1283 randomly selected 18-25 year-olds revealed that 146 of them currently use marijuana or hashish. At the 10% significance level, do the data provide sufficient evidence to conclude that the percentage of 18-25 year-olds who currently use marijuana or hashish has changed from the 1995 percentage of 12.0%?
Step 1:  \( H_0: \ p = 0.12 \)
\( H_a: \ p \neq 0.12 \)

Step 2:  \( n = 1283 \) and \( \alpha = 0.10 \)

Step 3:  The test statistic is \( Z = \frac{\hat{p} - 0.12}{\sqrt{\frac{(0.12)(0.88)}{1283}}} \), which under the null hypothesis has an approximate standard normal distribution.

Step 4:  The alternative hypothesis says “not equal to.” Therefore the rejection region is the union of a left-hand tail and a right-hand tail of the t-distribution. The area of each of these tails is 0.05, one half of our chosen significance level. Draw the rejection region on the graph.

\[ \text{Graph showing a standard normal distribution with rejection regions.} \]

Step 5:  \( X = 146 \). We choose \textbf{STAT, TESTS, 1-PropZTest}. We enter the null-hypothesized value, 0.12, the value of \( X \), and the sample size. We tell the calculator which of the three alternatives we are using. We find \( z = -0.683861 \), and \( p\)-value = 0.4940628. \( p\)-value > \( \alpha \). We don’t reject the null hypothesis.

Step 6:  We fail to reject \( H_0 \) at the 0.10 level of significance. We do not have sufficient evidence to conclude that the proportion of 18-25 year-olds currently using marijuana or hashish differs from the proportion found in 1995.