Chapter 8.1: Hypotheses and Test Procedures

Instructor: Dr. Arnab Maity
In Chapter 7, we learned how to construct a confidence interval that provides an interval of values that we believe contains our parameter of interest. In this chapter, we will learn how to test if a specified value of the parameter is reasonable. This is called hypothesis testing.

**Hypothesis testing** is one statistical technique using sample data to determine which of two contradictory claims about the parameter is correct.

- Honda Civic Hybrid claims that its mean gas mileage is 38 MPG in city. We are interested in whether this claim is correct or not.
- For a customer bank, the average waiting time for cashier service used to be 5 minutes. We want to know whether the average waiting time is significantly reduced 3 months after the bank has adopted a new system.

A **statistical hypothesis** is a claim about parameters or distributions.

- The **null hypothesis** $H_0$: the claim that is initially assumed to be true, the “prior belief”.
- The **alternative hypothesis** $H_a$: the claim that is contradictory to $H_0$, the “researcher’s hypothesis”. It is the claim that researchers would really like to validate.

**Examples:**

1. Suppose $\mu$ is the average diameter of a certain type of PVC pipes manufactured by a company. The manufacturer claims that the average diameter is 0.75 inches. We would like to test this claim. In this case

   $$H_0 : \mu = 0.75 \quad H_a : \mu \neq 0.75.$$

2. Suppose $p$ is true proportion of defective circuit boards produced by a certain manufacturer. They claim that less than 10% of their circuit boards are defective. In this case

   $$H_0 : p \leq 0.10 \quad H_a : p > 0.10.$$

3. Suppose $\mu$ is the true average tread life of a certain type of tire. The manufacturer claims that the average tread life is at least 30,000 miles. We would like to test this claim. In this case

   $$H_0 : \mu \geq 30000 \quad H_a : \mu < 30000.$$
Possible conclusions of hypothesis testing

- Reject $H_0$
- Fail to reject $H_0$ (there is not enough evidence to reject the null hypothesis)

Errors in hypothesis testing

- **Type I error**: reject $H_0$ when $H_0$ is true.
  - $\alpha = P$(Type I error)
- **Type II error**: $P$(not reject $H_0$ when $H_0$ is false).
  - $\beta = P$(Type II error)

**Hypothesis Testing**

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td><img src="image" alt="Type I Error" /></td>
<td><img src="image" alt="Smiley" /></td>
</tr>
<tr>
<td>Not reject $H_0$</td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Type II Error" /></td>
</tr>
</tbody>
</table>

**Type I: Reject $H_0$ when $H_0$ is true**

- Type II: Not reject $H_0$ when $H_0$ is false

**U.S. Justice System**

<table>
<thead>
<tr>
<th></th>
<th>Innocent</th>
<th>Guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guilty Verdict</td>
<td><img src="image" alt="Type I Error" /></td>
<td><img src="image" alt="Smiley" /></td>
</tr>
<tr>
<td>Not Guilty Verdict</td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Type II Error" /></td>
</tr>
</tbody>
</table>

**Type I: Send an innocent person to jail**

- Type II: A guilty person goes free
1. For each of the following assertions, state whether it is a legitimate statistical hypothesis and why:
   (a) $H : \sigma = 100$.
   (b) $H : \bar{X} = 10$.
   (c) $H : \bar{X} - \bar{Y} = 10$.

2. For the following pairs of assertions, indicate which do not comply with our rules for setting up hypotheses and why (the subscripts 1 and 2 differentiate between quantities for two different populations or samples):
   (a) $H_0 : \mu = 100$ v.s. $H_a : \mu > 100$.
   (b) $H_0 : p \neq 0.25$ v.s. $H_a : p = 0.25$.
   (c) $H_0 : S^2_1 = S^2_2$ v.s. $H_a : S^2_1 \neq S^2_2$.

3. A certain type of light bulb is advertised as having an average lifetime of 750 hours. A potential customer likes the price and wants to purchase a large amount of them if it can be shown the average lifetime is higher than advertised. A random sample of 20 bulbs was selected and the lifetime of each bulb was determined. The mean was 768.4 hours. It is known that the lifetime of the light bulbs is normally distributed and the true standard deviation is 30.5 hours. Set up the null and alternative hypotheses.
4. The average age of a person on Facebook last year was 19.34 years. The standard deviation of the age of Facebook users is 8.2 years. Suppose an advertising agency is interested in seeing if the average age is different this year. He randomly selects 100 profiles and finds that the sample mean is 18.01 years old. Set up the null and alternative hypotheses.

5. Suppose the Michelin tire company interested in showing that one of their competitors tires lasts less than 30,000 miles. They take a random sample of size \( n = 45 \). The sample mean is 28,540 miles and the sample standard deviation is 1930 miles. Set up the null and alternative hypotheses.

6. It has been determined that as long as the mean temperature of the discharged water is at most 150°F, there will be no negative effects on the river’s ecosystem. We want to investigate whether the plant is in compliance with regulations that prohibit a mean discharge-water temperature above 150°F. Suppose we construct \( H_0 : \mu = 150 \) against \( H_a : \mu > 150 \).

   (a) Describe the Type I error.

   (b) Describe the Type II error.

   (c) Which error do you view as more serious?
We make our decision using sample data. We need

- the test statistic: a function of the sample data, for example, \( \bar{X}, S \)
- the rejection region (RR): the set of the test statistic values such that \( H_0 \) will be rejected.

**Example:** Suppose a cigarette manufacturer claims that the average nicotine content (\( \mu \)) of brand A is at most 1.5 mg per cigarette. We are skeptical of this claim and want to test the validity. Here the null and alternative hypotheses are

\[
H_0 : \mu \leq 1.5 \quad H_a : \mu > 1.5.
\]

Now we take a sample of size \( n \) and compute the sample mean \( \bar{X} \).

- If \( H_0 \) is indeed true, we expect \( E(\bar{X}) = 1.5 \).
- If \( H_0 \) is false, we expect \( \bar{X} \) to exceed 1.5. If the observed value of \( \bar{X} \) is considerably larger than 1.5, then that will provide strong evidence against \( H_0 \).

  - Suppose for a specific sample of cigarettes, we observe \( \bar{x} = 4 \). What seems like a reasonable conclusions to make?

- Thus we can use \( \bar{X} \) as our test statistic and construct a rule: “if \( \bar{x} > 4 \), then reject \( H_0 \)”. Here a rejection region is \( \bar{x} \geq 4 \).

- What if you use \( \bar{x} > 1.7 \) as a rejection region? Which region to choose?

While the test statistic \( \bar{X} \) and the rejection region makes sense intuitively, the choice of cut-off value (4 in our example) is arbitrary. How to formalize these concepts in a statistical framework?

- Once the test statistic and sample size \( n \) are fixed, there is no rejection region that will simultaneously make both type I error level \( \alpha \) and type II error level \( \beta \) small. A region must be chosen to effect a compromise between these two quantities.

- In general, we first specify the largest value of type I error level \( \alpha \) that can be tolerated, and then find an rejection region that has the specified value of \( \alpha \). This makes \( \beta \) as small as possible subject to the bound on \( \alpha \).
**Example:** The drying time of a certain type of paint (in minutes) under specified test conditions is known to follow a $N(75, 81)$. Chemists have designed a new additive to decrease average drying time. Let $Y$ denote the drying time of this new additive. Let us assume that $Y \sim N(\mu, 81)$, where $\mu$ is unknown. We want to determine if there is strong evidence to suggest an improvement in average drying time. Suppose a random sample of 25 drying times is taken.

1. State the null and the alternative hypotheses.

2. Suppose that the 25 drying times are denoted by $X_1, \ldots, X_{25}$. What is the distribution of $\bar{X}$ in general? If $H_0$ is indeed true, what is the distribution of $\bar{X}$?

3. Suppose we use $\bar{X}$ as our test statistic and use the rejection region $\{\bar{x} : \bar{x} \leq 70.8\}$, that is, we reject $H_0$ if the observed value of the sample mean of our sample is less or equal to than 70.8. Draw the distribution of $\bar{X}$ when $H_0$ is true and shade the rejection region.

4. Compute the type I error rate for this rejection region.
5. Assume now that in actuality the average drying time is 72 minutes. Find the probability of a type II error. (We denote this as $\beta(72)$.) Redraw the above plot with shaded rejection region along with a plot depicting the type II error.

6. Assume now that in actuality the average drying time is 70 minutes. Find the probability of a type II error $\beta(70)$. Redraw the above plot with this information included.

7. What do parts (4-6) tell you?