Hypothesis Testing

Introduction

In many cases, it is necessary to evaluate the validity of a hypothesis, determining whether it is true or false. For instance, a pharmaceutical company may want to determine if a new drug is effective in treating a disease. In this scenario, two hypotheses are considered: the first hypothesis (H_0) states that the drug is not effective, while the second hypothesis (H_1) states that the drug is effective. The terms H_0 and H_1 are used to refer to these hypotheses, respectively. Another example involves a radar system that uses radio waves to detect aircraft. The system receives a signal and needs to decide whether an aircraft is present or not. In this case, there are two opposing hypotheses:

- H_0: No aircraft is present.
- H_1: An aircraft is present.

The hypothesis H_0 is known as the null hypothesis, while H_1 is referred to as the alternative hypothesis. The null hypothesis (H_0) is typically considered the default assumption, initially assumed to be true. On the other hand, the alternative hypothesis (H_1) contradicts H_0. Based on the observed data, a decision must be made to either accept H_0 or reject it, in which case H_1 is accepted. These are known as hypothesis testing problems. In this chapter, we will discuss how to approach such problems from a classical (frequentist) perspective. We will begin with an example and introduce some commonly used terminology. It is not necessary to focus too much on the terminology in this example, as more precise definitions will be provided later on.

Example 0.0.1. Suppose you have a coin and you want to determine whether it is fair or not. Specifically, let θ represent the probability of heads, $\theta = P(H)$. Two hypotheses are considered:

- H_0 (the null hypothesis): The coin is fair, i.e., $\theta = \theta_0 = \frac{1}{2}$.
- H_1 (the alternative hypothesis): The coin is not fair, i.e., $\theta \neq \frac{1}{2}$.

Solution

To evaluate whether the coin is fair or not, we conduct an experiment. We toss the coin 100 times and record the number of heads. Let X denote the number of heads observed, so X follows a binomial distribution with parameters 100 and θ ($X \sim \text{Binomial}(100, \theta)$).
If H_0 is true, then $\theta = \theta_0 = \frac{1}{2}$, and we expect the number of heads to be close to 50. Intuitively, if we observe a number of heads close to 50, we should accept H_0; otherwise, we should reject it. More specifically, the following criteria are suggested:

- If $|X - 50|$ is less than or equal to a certain threshold, we accept H_0.
- If $|X - 50|$ exceeds the threshold, we reject H_0 and accept H_1.

Let's denote this threshold as t.

- If $|X - 50| \leq t$, accept H_0.
- If $|X - 50| > t$, accept H_1.

However, the question arises: How do we choose the threshold t? To choose t appropriately, we need to define some requirements for our test. An important consideration is the probability of error. One type of error occurs when we reject H_0 while it is actually true, known as a type I error. Specifically, this refers to the event that $|X - 50| > t$ when H_0 is true. Therefore, we can express the probability of a type I error as $P(\text{type I error}) = P(|X - 50| > t|H_0)$. We interpret this as the probability that $|X - 50| > t$ when H_0 is true. (Note that in classical statistics, H_0 and H_1 are not treated as random events, so $P(|X-50| > t|H_0)$ is not a conditional probability. Another common notation is $P(|X - 50| > t \text{ when } H_0 \text{ is true})$.) To determine t, we can choose a desired value for $P(\text{type I error})$. For example, we might want a test where $P(\text{type I error}) \leq \alpha = 0.05$.

0.1 General Setting and Definitions

Introduced the fundamentals of hypothesis testing. In this section, we aim to establish a general framework for hypothesis testing problems and formally define the terminology used in such tests. While there may be unfamiliar phrases like null hypothesis, type I error, significance level, etc., there are no fundamentally new concepts or tools here. With the help of a few examples, these concepts should become clear.

Let's consider an unknown parameter θ. In hypothesis testing problems, we need to decide between two contradictory hypotheses. Specifically, let S represent the set of possible values for θ. We can partition S into two disjoint sets, S_0 and S_1. We define H_0 as the hypothesis that θ belongs to S_0, and H_1 as the hypothesis that θ belongs to S_1.

- H_0 (null hypothesis): θ belongs to S_0.
- H_1 (alternative hypothesis): θ belongs to S_1.

In Example 0.0.1, we had $S = [0, 1]$, $S_0 = \left\{\frac{1}{2}\right\}$, and $S_1 = [0, 1] - \left\{\frac{1}{2}\right\}$. In this case, H_0 is an example of a simple hypothesis because S_0 contains only one value of θ. On the other hand, H_1 is an example of a composite hypothesis since S_1 contains more than one element. It is common for the null hypothesis to be a simple hypothesis.

To make a decision between H_0 and H_1, we often examine a function of the observed data. For example, in Example 0.0.1, we looked at the random variable Y defined as:

$$Y = \frac{X - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}},$$

where X was the total number of heads observed. Here, X is a function of the observed data (sequence of heads and tails $X = \sum_{i=1}^{n}\{Head\}$), and thus Y is also a function of the observed data. We refer to Y as a statistic.

Reminder: Let $X_1, X_2, \ldots, X_n$ be a random sample of interest. A *statistic* is a real-valued function of the data. For instance, the sample mean defined as $W(X_1, X_2, \ldots, X_n) = \frac{X_1+X_2+\ldots+X_n}{n}$ is a statistic. A *statistic test* is a statistic on which we base our hypothesis test.

To choose between H_0 and H_1, we select a test statistic $W = W(X_1, X_2, \ldots, X_n)$. Assuming H_0, we define the set A as the set of possible values of W for which we would accept H_0. A is called the *acceptance region*, while the set $R = S - A$ is referred to as the *rejection region*. In Example 0.0.1, the acceptance region was determined as $A = [-1.96, 1.96]$, and the rejection region was $R = (-\infty, -1.96) \cup (1.96, \infty)$.

There are two possible errors that can occur. We define a *type I error* as the event of rejecting H_0 when H_0 is actually true. Note that the probability of a type I error generally depends on the true value of θ. More specifically:

$$P(\text{type I error}|\theta) = P(\text{Reject } H_0|\theta) = P(W \in R|\theta), \text{ for } \theta \in S_0.$$

If the probability of a type I error satisfies:

$$P(\text{type I error}) \leq \alpha, \text{ for all } \theta \in S_0,$$

we say the test has a significance level α or that the test is a level α test.

it's important to note that the significance level α is chosen by the experimenter and represents the maximum tolerable probability of committing a type I error.

The other possible error is a *type II error*, which occurs when we fail to reject H_0 when H_0 is false. The probability of a type II error is denoted by β, and it depends on the true value of θ as well as the specific value of θ under consideration as an alternative hypothesis. More formally:

$$P(\text{type II error}|\theta') = P(\text{Do not reject } H_0|\theta') = P(W \in A|\theta'), \text{ for } \theta' \in S_1.$$

The *power* of a test is defined as $1 - \beta$, which represents the probability of correctly rejecting H_0 when H_0 is false.

In summary, hypothesis testing involves making decisions between two contradictory hypotheses H_0 and H_1. A test statistic is used to make this decision, and an acceptance region A and a rejection region R are defined based on the test statistic. The significance level α determines the maximum tolerable probability of a type I error, while the power of a test represents the probability of correctly rejecting H_0 when H_0 is false.

0.2 One-sided tests and Two-sided tests:

In hypothesis testing, the choice between a one-sided (or one-tailed) test and a two-sided (or two-tailed) test depends on the nature of the research question and the direction of interest.

- **One-sided test:**
  - Focuses on detecting an effect in one specific direction (greater than or less than).
  - Null hypothesis (H_0) is typically stated as no effect or a specific value.
  - Example: $H_0 : \mu \leq cst$ versus $H_1 : \mu > cst$ (testing if the mean is greater than cst).
- **Two-sided test:**
  - Examines whether there is a significant difference in any direction.
  - Null hypothesis (H_0) often states no effect or equality.
  - Example: $H_0 : \mu = cst$ versus $H_1 : \mu \neq cst$ (testing if the mean is different from 10).

The choice between one-sided and two-sided tests depends on the specific hypotheses being tested and the research question's requirements.

Example 0.2.1. Suppose you are conducting a hypothesis test on the average height (μ) of a certain population.

- **One-sided test:**
  - Null hypothesis (H_0): $\mu \leq 65$ inches
  - Alternative hypothesis (H_1): $\mu > 65$ inches
  - This one-sided test aims to determine if the average height is greater than 65 inches.

- **Two-sided test:**
  - Null hypothesis (H_0): $\mu = 70$ inches
  - Alternative hypothesis (H_1): $\mu \neq 70$ inches
  - This two-sided test aims to determine if the average height is different from 70 inches.

For the one-sided test, you would be interested in detecting whether the average height is significantly greater than the specified value (65 inches). In the two-sided test, the interest is in detecting any significant difference in the average height, whether it is greater or less than the specified value (70 inches).

The choice between one-sided and two-sided tests depends on the specific hypothesis and the research question you want to address.

0.3 The different types of errors

In hypothesis testing, there are two types of possible errors: type I errors and type II errors. Understanding these error types is important for evaluating and designing statistical tests.

Type I Error

A type I error occurs when the null hypothesis (H_0) is rejected even though it is true. In other words, a statistically significant result is found when there is really no effect. The probability of making a type I error is denoted by alpha (α) and is preset at the beginning of the test, usually at 5% or 1%.

Type II Error

A type II error occurs when the null hypothesis (H_0) fails to be rejected even though it is false. In other words, no statistically significant result is found even though there is a real effect. The probability of making a type II error is denoted by beta (β). The power of a test is equal to $1 - \beta$ and indicates the probability of correctly rejecting the null hypothesis when it is false.

Relationship Between Type I and Type II Errors

The type I and type II error rates are intrinsically related. As the significance level (α) decreases, making it harder to reject the null, the chance of type I error decreases but the chance of type II error increases. Conversely, increasing the significance level makes it easier to reject the null and lowers type II errors but raises type I errors. There is always a tradeoff between the two error types.

Understanding the different types of errors in hypothesis testing is crucial for properly evaluating statistical tests and avoiding misleading conclusions from data analysis. Both type I and type II errors should be considered when designing experiments and setting significance levels.

0.4 Power of a Statistical Test

Statistical hypothesis testing is a fundamental practice in data analysis and research. It allows to make inferences about populations based on experimental data and results from samples. A crucial consideration in the planning and evaluation of hypothesis tests is *the power of the statistical tests* being used. Properly assessing power helps ensure meaningful and conclusive results can be obtained from studies.

Definition 0.4.1. *The power of a statistical test gives the probability of rejecting the null hypothesis when it is false. Just as the significance level (alpha) gives the probability of rejecting the null hypothesis when it is true, power quantifies the chance of correctly rejecting the null hypothesis when it is false. Thus, power represents a test's ability to correctly reject the null hypothesis.*

Calculating power beforehand is important to ensure the sample size is sufficient for the test objectives. Otherwise, the test may be inconclusive, wasting resources. Power should generally not be calculated after the test, except to determine an adequate sample size for a follow-up study.

Example 0.4.1. Consider testing whether the average time per week spent watching TV is 4 hours versus the alternative that it is greater than 4 hours. We will calculate the power of this test for a specific value under the alternative hypothesis of 7 hours.

Solution:

1. State the null and alternative hypotheses
   - Null Hypothesis (H_0): The average time spent watching TV per week (μ) equals 4 hours
   - Alternative Hypothesis (H_1): The average time spent watching TV per week (μ) equals 6 hours
2. Define the parameters
   - μ_0 = Average time under the null hypothesis = 4 hours
   - μ_1 = Average time under the alternative hypothesis = 6 hours
3. Specify additional information
   - The standard deviation from past data is known to be 2 hours
   - The sample size is 4
4. Calculate the power of this test for a sample size of 4. Show the step-by-step working.

1. At the 5% significance level, the decision criterion for the test is to reject H_0 if $Z > 1.645$, where

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X} - 4}{\frac{2}{\sqrt{4}}} = \overline{X} - 4.$$

   The 5% critical value from the standard normal distribution is 1.645. Equating the critical Z-value to the calculated Z gives the corresponding (hypothetical) sample mean value:

$$\overline{X} = 5.645.$$

2. Calculate the Z-statistic assuming the alternative hypothesis is true, i.e., $\mu_1 = 6$:

$$Z = \frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{5.645 - 6}{\frac{2}{\sqrt{4}}} = -0.355.$$

3. $P(Z > -0.355) = 0.6387$. The power of the test is approximately 64%. In general, tests with 80% power and higher are considered to be statistically powerful.

To increase power, one may:

- Increase effect size difference
- Increase sample size(s)
- Decrease variability
- Increase significance level (but increases type I error risk)

0.5 P-Values

In the previous discussions, we only provided an "accept" or "reject" decision as the outcome of a hypothesis test. However, we can offer more information by using a measure called P-values. In essence, P-values indicate how close the decision was. To elaborate, if we reject the null hypothesis H_0 at a significance level $\alpha = 0.05$, we can inquire about the outcome at a different significance level, such as $\alpha = 0.01$. Can we still reject H_0? More precisely, we can ask the following question:

What is the smallest significance level α that leads to the rejection of the null hypothesis?

The response to this query is known as the P-value. The P-value is the minimum significance level α that results in rejecting the null hypothesis. In simple terms, if the P-value is small, it implies that the observed data is highly unlikely to occur under H_0, thereby providing stronger evidence for rejecting the null hypothesis. How do we determine P-values? Let's examine an example.

Example 0.5.1. Suppose you have a coin and you want to investigate whether it is fair or biased. Specifically, let θ denote the probability of obtaining heads, where $\theta = P(H)$. You need to choose between the following hypotheses:

H_0 (null hypothesis): The coin is fair, i.e., $\theta = \theta_0 = \frac{1}{2}$.

H_1 (alternative hypothesis): The coin is biased, i.e., $\theta > \frac{1}{2}$.

You toss the coin 100 times and observe 60 heads. Can we reject H_0 at a significance level $\alpha = 0.05$? Can we reject H_0 at a significance level $\alpha = 0.01$? What is the P-value?

Solution

Let X be the random variable representing the number of observed heads. In our experiment, we observed $X = 60$. Since $n = 100$ is relatively large, assuming H_0 is true, the random variable

$$W = \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$$

is approximately a standard normal random variable, $N(0, 1)$. If H_0 is true, we expect X to be close to 50, whereas if H_1 is true, we anticipate X to be larger. Therefore, we can propose the following test: choose a threshold c. If $W \leq c$, we accept H_0; otherwise, we accept H_1. To calculate the P(type I error), we can express it as

$$P(\text{type I error}) = P(\text{Reject } H_0 | H_0) = P(W > c | H_0).$$

Since W follows a standard normal distribution under H_0, we need to select $c = z_\alpha$ to ensure a significance level α. In this example, we find that

$$W = \frac{X - 50}{5} = \frac{60 - 50}{5} = 2.$$

If we require a significance level $\alpha = 0.05$, then

$$c = z_{0.05} = 1.645.$$

The value above can be obtained in MATLAB using the command `normինv(1 - 0.05)`. As $W = 2 > 1.645$, we reject H_0 and accept H_1.

If we require a significance level $\alpha = 0.01$, then

$$c = z_{0.01} = 2.33.$$

The value above can be obtained in MATLAB using the command `norminv(1 - 0.01)`. As $W = 2 \leq 2.33$, we fail to reject H_0, so we accept H_0.

The P-value is the minimum significance level α that leads to the rejection of H_0. In this case, since $W = 2$, we reject H_0 only if $c < 2$. Note that $z_\alpha = c$, thus

$$\alpha = 1 - \Phi(c).$$

If $c = 2$, we obtain

$$\alpha = 1 - \Phi(2) = 0.023.$$

Therefore, we rejectH_0 at a significance level of $\alpha = 0.023$. The P-value is 0.023.

0.6 Hypothesis Testing for the Mean

Here, we would like to discuss some common hypothesis testing problems. We assume that we have a random sample $X_1, X_2, ..., X_n$ from a distribution and our goal is to make inference about the mean of the distribution μ. We consider three hypothesis testing problems. The first one is a test to decide between the following hypotheses:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0.$$

In this case, the null hypothesis is a simple hypothesis and the alternative hypothesis is a two-sided hypothesis (i.e., it includes both $\mu < \mu_0$ and $\mu > \mu_0$). We call this hypothesis test a two-sided test. The second and the third cases are one-sided tests. More specifically, the second case is

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0.$$

Here, both H_0 and H_1 are one-sided, so we call this test a one-sided test. The third case is very similar to the second case. More specifically, the third scenario is

$$H_0 : \mu \geq \mu_0, \quad H_1 : \mu < \mu_0.$$

In all of the three cases, we use the sample mean $\overline{X} = \frac{1}{n}(X_1+X_2+...+X_n)$ to define our statistic. In particular, if we know the variance of the X_i's, $Var(X_i) = \sigma^2$, then we define our test statistic as the normalized sample mean (assuming H_0):

$$W(X_1, X_2, ..., X_n) = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

If we do not know the variance of the X_i's, we use

$$W(X_1, X_2, ..., X_n) = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

where S is the sample standard deviation,

$$S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2}.$$

In any case, we will be able to find the distribution of W, and thus we can design our tests by calculating error probabilities. Let us start with the first case.

0.6.1 Two-sided Tests for the Mean

Here, we are given a random sample $X_1, X_2, ..., X_n$ from a distribution. Let $\mu = E(X_i)$. Our goal is to decide between

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0.$$

Example 0.0.1, which we saw previously, is an instance of this case. If H_0 is true, we expect $\overline{X}$ to be close to μ_0, and so we expect $W(X_1, X_2, ..., X_n)$ to be close to 0 (see the definition of W above).

Therefore, we can suggest the following test. Choose a threshold, and call it c. If $|W| \leq c$, accept H_0, and if $|W| > c$, accept H_1. How do we choose c? If α is the required significance level, we must have

$$P(\text{type I error}) = P(\text{Reject } H_0|H_0) = P(|W| > c|H_0) \leq \alpha.$$

Thus, we can choose c such that $P(|W| > c|H_0) = \alpha$. Let us look at an example.

Example 0.6.1. Let $X_1, X_2, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, where μ is unknown but σ is known. Design a level α test to choose between

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0.$$

Solution As discussed above, we let

$$W(X_1, X_2, ..., X_n) = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Under the assumption of H_0, we can observe that W follows a standard normal distribution, $W \sim N(0, 1)$.

To choose a threshold value c, we aim to satisfy the condition $P(|W| > c|H_0) = \alpha$. Since the standard normal distribution is symmetric around 0, we have $P(|W| > c|H_0) = 2P(W > c|H_0)$. Hence, we conclude that $P(W > c|H_0) = \alpha/2$. Therefore, $c = z_{\alpha/2}$, where $z_{\alpha/2}$ represents the $(1-\alpha/2)$ percentile of the standard normal distribution.

Hence, we accept H_0 if $\left|\overline{X} - \mu_0\right| / (\sigma/\sqrt{n}) \leq z_{\alpha/2}$ and reject it otherwise.

Relation to Confidence Intervals: It is interesting to examine the above acceptance region. Here, we accept H_0 if $|\bar{X} - \mu_0 \frac{\sigma}{\sqrt{n}}| \leq z_{\alpha/2}$. We can rewrite the above condition as $\mu_0 \in [\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$. The above interval should look familiar to you. It is the $(1 - \alpha) \times 100\%$ confidence interval for μ_0. This is not a coincidence as there is a general relationship between confidence interval problems and hypothesis testing problems.

Example 0.6.2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, where μ is unknown but σ is known. Design a level α test to choose between

- $$H_0 : \mu = \mu_0$$

- $$H_1 : \mu \neq \mu_0.$$

,

find β, the probability of type II error, as a function of μ.

Solution: We have

$$\begin{aligned}\beta(\mu) &= P(\text{type II error}) \\ &= P(\text{accept } H_0|\mu) \\ &= P(|\bar{X} - \mu_0 \frac{\sigma}{\sqrt{n}}| < z_{\alpha/2}|\mu).\end{aligned}$$

If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Thus,

$$\begin{aligned}\beta(\mu) &= P(|\bar{X} - \mu_0 \frac{\sigma}{\sqrt{n}}| < z_{\alpha/2}|\mu) \\ &= P(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \\ &= \Phi(z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}).\end{aligned}$$

Unknown variance: The above results (Example 0.6.2) can be extended to the case when we do not know the variance using the t-distribution. More specifically, consider the following example.

Example 0.6.3. Let $X_1, X_2, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, where μ and σ are unknown. Design a level α test to choose between

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

Solution: Let S^2 be the sample variance for this random sample. Then, the random variable W defined as $W(X_1, X_2, ..., X_n) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ has a t-distribution with $n-1$ degrees of freedom, i.e., $W \sim T(n-1)$. Thus, we can repeat the analysis of Example 8.24 here. The only difference is that we need to replace σ by S and $z_{\alpha/2}$ by $t_{\alpha/2,n-1}$. Therefore, we accept H_0 if $|W| \leq t_{\alpha/2,n-1}$, and reject it otherwise. Let us look at a numerical example of this case.

Example 0.6.4. Consider the following scenario: Let $X_1, X_2, \ldots, X_n$ represent a random sample drawn from a normal distribution with mean μ and variance σ^2. In this case, the value of μ is unknown while σ is known. Our objective is to design a test with a significance level α to make a decision between the null hypothesis $H_0 : \mu \leq \mu_0$ and the alternative hypothesis $H_1 : \mu > \mu_0$.

To accomplish this, we can define a test statistic $W(X_1, X_2, \ldots, X_n)$ as $\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$. If H_0 is true (meaning $\mu \leq \mu_0$), we anticipate that $\bar{X}$ (and consequently W) will be relatively small. Conversely, if H_1 is true, we expect $\bar{X}$ (and thus W) to be larger. Based on this observation, we can establish the following test: Select a threshold value, denoted as c. If $W \leq c$, we accept H_0; otherwise, if $W > c$, we accept H_1.

The question now arises: How do we determine the appropriate value for c? To ensure that the probability of committing a type I error (rejecting H_0 when it is true) is at most α, we need to examine the relationship between c and α. The probability of a type I error is contingent on the value of μ. More precisely, for any $\mu \leq \mu_0$, we can express the probability of a type I error as follows: $P(\text{type I error}|\mu) = P(\text{Reject } H_0|\mu) = P(W > c|\mu)$.

By employing properties of the normal distribution, we can simplify the expression above: $P(W > c|\mu) = P\left(\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > c|\mu\right) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > c + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}|\mu\right) \leq P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > c|\mu\right)$ (since $\mu \leq \mu_0$) $= 1-\Phi(c)$ (since $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ follows a standard normal distribution).

Consequently, we can select $\alpha = 1 - \Phi(c)$, which implies $c = z_\alpha$. Therefore, we accept H_0 if $\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_\alpha$, and we reject it otherwise.

The above analysis can be extended to other cases as well. In general, suppose we are given a random sample $X_1, X_2, \ldots, X_n$ drawn from a distribution, and let $\mu = E(X_i)$. Our objective is to make a decision between the null hypothesis $H_0 : \mu \leq \mu_0$ and the alternative hypothesis $H_1 : \mu > \mu_0$.

We can define the test statistic W as follows: $W(X_1, X_2, \ldots, X_n) = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ if σ (the variance of X_i) is known, and $W(X_1, X_2, \ldots, X_n) = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ if σ is unknown. If H_0 is true (i.e., $\mu \leq \mu_0$), we expect $\bar{X}$ (and thus W) to be relatively small. Conversely, if H_1 is true"' (i.e., $\mu > \mu_0$), we anticipate $\bar{X}$ (and thus W) to be larger. Based on this expectation, we can establish the following test: Choose a threshold c. If $W \leq c$, we accept H_0; otherwise, if $W > c$, we accept H_1.

To determine the value of c, note that $P(\text{type I error}) = P(\text{Reject } H_0|H_0) = P(W > c|\mu \leq \mu_0) \leq P(W > c|\mu = \mu_0)$. The last inequality holds because increasing μ can only increase the probability of $W > c$. In other words, we assume the worst-case scenario, where $\mu = \mu_0$, to compute the probability of error. Hence, we can select c such that $P(W > c|\mu = \mu_0) = \alpha$. By following this procedure, we obtain the acceptance regions depicted in Table 8.3.

Example 0.6.5. The average adult male height in a certain country is 170 cm. We suspect that the men in a certain city in that country might have a different average height due to some environmental factors. We pick a random sample of size 9 from the adult males in the city and obtain the following values for their heights (in cm):

176.2, 157.9, 160.1, 180.9, 165.1, 167.2, 162.9, 155.7, 166.2

Table 1: One-sided hypothesis testing for the mean: $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$.

Case	Test Statistic	Acceptance Region
$X_i \sim N(\mu, \sigma^2)$, σ known	$W = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$W \leq z_\alpha$
n large, X_i non-normal	$W = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	$W \leq z_\alpha$
$X_i \sim N(\mu, \sigma^2)$, σ unknown	$W = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	*(remaining part of the sentence is missing)*

Assume that the height distribution in this population is normally distributed. Here, we need to decide between
$H_0 : \mu = 170$
$H_1 : \mu \neq 170$
Based on the observed data, is there enough evidence to reject H_0 at significance level $\alpha = 0.05$?

Solution: Let's first calculate the sample mean and sample standard deviation:

- Sample mean, $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = 166.44$
- Sample standard deviation, $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} = 8.548$
- The test statistic, $W = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \frac{166.44-170}{8.548/\sqrt{9}} = -1.478$

Since we have a two-sided alternative hypothesis, we need to find the critical values for the t-distribution with $n - 1 = 9 - 1 = 8$ degrees of freedom. For a significance level of $\alpha = 0.05$, the critical values are $t_{\alpha/2,n-1} = t_{0.025,8} = 2.306$.

Since $|W| = 1.478 < t_{\alpha/2,n-1} = 2.306$, we do not have enough evidence to reject H_0 at the significance level of $\alpha = 0.05$. Therefore, based on the observed data, there is not enough evidence to conclude that the average height of men in the city is different from 170 cm.

0.7 Likelihood Ratio Tests

In this section, we will explore the concept of Likelihood Ratio Tests, which is a general hypothesis testing procedure. Before diving into the details, let's review the definition of the likelihood function, which we have previously discussed.

0.7.1 Review of the Likelihood Function

Consider a random sample $X_1, X_2, X_3, ..., X_n$ from a distribution with a parameter θ. The likelihood function is defined differently for discrete and continuous random variables:

- For discrete random variables, the likelihood function is denoted as $L(x_1, x_2, ..., x_n; \theta)$ and represents the probability mass function $P_{X_1 X_2 \ldots X_n}(x_1, x_2, ..., x_n; \theta)$.

- For continuous random variables, the likelihood function is denoted as $L(x_1, x_2, ..., x_n; \theta)$ and represents the probability density function $f_{X_1 X_2 \dots X_n}(x_1, x_2, ..., x_n; \theta)$.

0.7.2 Likelihood Ratio Tests

Likelihood Ratio Tests are used in hypothesis testing when both the null and alternative hypotheses are simple. Suppose we have two hypotheses:

- Null Hypothesis: $H_0 : \theta = \theta_0$
- Alternative Hypothesis: $H_1 : \theta = \theta_1$

To decide between these hypotheses, we compare the likelihood functions:

$$l_0 = L(x_1, x_2, ..., x_n; \theta_0) \quad \text{(likelihood under } H_0\text{)}$$

$$l_1 = L(x_1, x_2, ..., x_n; \theta_1) \quad \text{(likelihood under } H_1\text{)}$$

If l_0 is significantly larger than l_1, we accept H_0. Conversely, if l_1 is significantly larger, we tend to reject H_0. The likelihood ratio $\frac{l_0}{l_1}$ is used to make the decision.

0.7.3 Likelihood Ratio Test for Simple Hypotheses

Consider a random sample $X_1, X_2, X_3, ..., X_n$ from a distribution with parameter θ. Suppose we want to test between two simple hypotheses:

- Null Hypothesis: $H_0 : \theta = \theta_0$
- Alternative Hypothesis: $H_1 : \theta = \theta_1$

We define the likelihood ratio as:

$$\lambda(x_1, x_2, ..., x_n) = \frac{L(x_1, x_2, ..., x_n; \theta_0)}{L(x_1, x_2, ..., x_n; \theta_1)}$$

To perform a Likelihood Ratio Test (LRT), we choose a constant c. We reject H_0 if $\lambda < c$ and accept it if $\lambda \geq c$. The value of c is determined based on the desired significance level α.

Example

Let's consider an example to illustrate how to perform a Likelihood Ratio Test. We revisit the radar problem, where we observe the random variable X given by $X = \theta + W$, with $W \sim N(0, \sigma^2 = \frac{1}{9})$. We want to test between the following hypotheses:

- Null Hypothesis: $H_0 : \theta = \theta_0 = 0$
- Alternative Hypothesis: $H_1 : \theta = \theta_1 = 1$

o design a level 0.05 test ($\alpha = 0.05$) to decide between H_0 and H_1, we calculate the likelihood ratio and determine the threshold value c. The decision rule is then defined based on the observed value of X.

0.7.4 Generalization to Non-Simple Hypotheses

If the hypotheses are not simple, meaning θ is an unknown parameter, we can still perform a Likelihood Ratio Test by partitioning the set of possible values for θ into two disjoint sets S_0 and S_1. The test involves finding the likelihood ratio for each possible value of θ and choosing the value that maximizes the likelihood ratio.

$$\lambda(x_1, x_2, ..., x_n) = \frac{\sup_{\theta \in S_0} L(x_1, x_2, ..., x_n; \theta)}{\sup_{\theta \in S_1} L(x_1, x_2, ..., x_n; \theta)}$$

To perform the Likelihood Ratio Test, we compare the likelihood ratio λ to a threshold value c. If $\lambda < c$, we reject H_0, and if $\lambda \geq c$, we fail to reject H_0.

The threshold value c is determined based on the desired significance level α. It is chosen such that the probability of rejecting H_0 when it is true (Type I error) is limited to α. In other words, we control the probability of falsely rejecting H_0.