8.4.1 Introduction

Often, we need to test whether a hypothesis is true or false. For example, a pharmaceutical company might be interested in knowing if a new drug is effective in treating a disease. Here, there are two hypotheses. The first one is that the drug is not effective, while the second hypothesis is that the drug is effective. We call these hypotheses H_0 and H_1 respectively. As another example, consider a radar system that uses radio waves to detect aircraft. The system receives a signal and, based on the received signal, it needs to decide whether an aircraft is present or not. Here, there are again two opposing hypotheses:

H_0: No aircraft is present.

H_1: An aircraft is present.

The hypothesis H_0 is called the *null hypothesis* and the hypothesis H_1 is called the *alternative hypothesis*. The null hypothesis, H_0, is usually referred to as the default hypothesis, i.e., the hypothesis that is initially assumed to be true. The alternative hypothesis, H_1, is the statement contradictory to H_0. Based on the observed data, we need to decide either to accept H_0, or to reject it, in which case we say we accept H_1. These are problems of *hypothesis testing*. In this section, we will discuss how to approach such problems from a classical (frequentist) point of view. We will start with an example, and then provide a general framework to approach hypothesis testing problems. When looking at the example, we will introduce some terminology that is commonly used in hypothesis testing. Do not worry much about the terminology when reading this example as we will provide more precise definitions later on.

Example 8.22

You have a coin and you would like to check whether it is fair or not. More specifically, let θ be the probability of heads, $\theta = P(H)$. You have two hypotheses:

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

We need to design a test to either accept H_0 or H_1. To check whether the coin is fair or not, we perform the following experiment. We toss the coin 100 times and record the number of heads. Let X be the number of heads that we observe, so

$$X \sim Binomial(100, \theta).$$

Now, if H_0 is true, then $\theta = \theta_0 = \frac{1}{2}$, so we expect the number of heads to be close to 50. Thus, intuitively we can say that if we observe close to 50 heads we should accept H_0, otherwise we should reject it. More specifically, we suggest the following criteria: If $|X - 50|$ is less than or equal to some threshold, we accept H_0. On the other hand, if $|X - 50|$ is larger than the threshold we reject H_0 and accept H_1. Let's call that threshold t.

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

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

But how do we choose the threshold t? To choose t properly, we need to state some requirements for our test. An important factor here is probability of error. One way to make an error is when we reject H_0 while in fact it is true. We call this *type I error*. More specifically, this is the event that $|X - 50| > t$ when H_0 is true. Thus,

$$P(\text{type I error}) = P(|X - 50| > t \mid H_0).$$

We read this as the probability that $|X - 50| > t$ <u>when</u> H_0 is true. (Note that, here, $P(|X - 50| > t \mid H_0)$ is not a conditional probability, since in classical statistics we do not treat H_0 and H_1 as random events. Another common notation is $P(|X - 50| > t \text{ when } H_0 \text{ is true})$.) To be able to decide what t needs to be, we can choose a desired value for $P\big(\text{type I error}\big)$. For example, we might want to have a test for which

$$P(\text{type I error}) \leq \alpha = 0.05$$

Here, α is called *the level of significance*. We can choose

$$P(|X - 50| > t \mid H_0) = \alpha = 0.05 \qquad (8.2)$$

to satisfy the desired level of significance. Since we know the distribution of X under H_0, i.e., $X|H_0 \sim Binomial(100, \theta = \frac{1}{2})$, we should be able to choose t such that Equation 8.2 holds. Note that by the central limit theorem (CLT), for large values of n, we can approximate a $Binomial(n, \theta)$ distribution by a normal distribution. More specifically, we can say that for large values of n, if $X \sim Binomial(n, \theta_0 = \frac{1}{2})$, then

$$Y = \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} = \frac{X - 50}{5}$$

is (approximately) a standard normal random variable, $N(0,1)$. Thus, to be able to use the CLT, instead of looking at X directly, we can look at Y. Note that

$$\begin{aligned} P(\text{type I error}) = P(|X-50| > t|H_0) &= P\left(\left|\frac{X-50}{5}\right| > \frac{t}{5} \mid H_0\right) \\ &= P\left(|Y| > \frac{t}{5} \mid H_0\right). \end{aligned}$$

For simplicity, let's put $c = \frac{t}{5}$, so we can summarize our test as follows:

If $|Y| \leq c$, accept H_0.

If $|Y| > c$, accept H_1.

where $Y = \frac{X-50}{5}$. Now, we need to decide what c should be. We need to have

$$\begin{aligned} \alpha &= P\left(|Y| > c\right) \\ &= 1 - P\left(-c \leq Y \leq c\right) \\ &\approx 2 - 2\Phi\left(c\right) \quad \big(\text{using } \Phi(x) = 1 - \Phi(-x)\big). \end{aligned}$$

Thus, we need to have

$$2 - 2\Phi(c) = 0.05$$

So we obtain

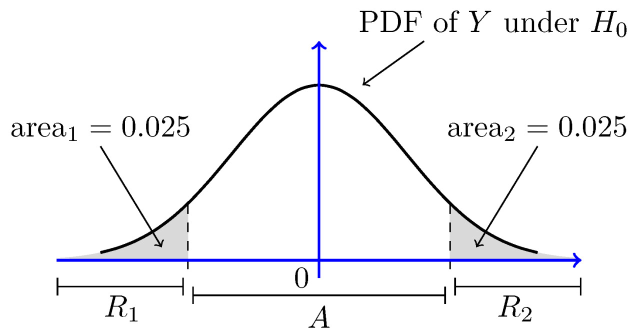
$$c = \Phi^{-1}(0.975) = 1.96$$

Thus, we conclude the following test

If $|Y| \leq 1.96$, accept H_0.

If $|Y| > 1.96$, accept H_1.

The set $A = [-1.96, 1.96]$ is called the *acceptance region*, because it includes the points that result in accepting H_0. The set $R = (-\infty, -1.96) \cup (1.96, \infty)$ is called the *rejection region* because it includes the points that correspond to rejecting H_0. Figure 8.9 summarizes these concepts.



A = Acceptance Region
$R = R_1 \cup R_2$ = Rejection Region
$\alpha = \text{P(type I error)} = \text{area}_1 + \text{area}_2 = 0.05$

Figure 8.9 - Acceptance rejection, rejection region, and type I error for Example 8.22

Note that since $Y = \frac{X-50}{5}$, we can equivalently state the test as

If $|X - 50| \leq 9.8$, accept H_0.

If $|X - 50| > 9.8$, accept H_1.

Or equivalently,

If the observed number of heads is in $\{41, 42, \cdots, 59\}$, accept H_0.

If the observed number of heads is in $\{0, 1, \cdots, 40\} \cup \{60, 61, \cdots, 100\}$, reject H_0 (accept H_1).

In summary, if the observed number of heads is more than 9 counts away from 50, we reject H_0.

Before ending our discussion on this example, we would like to mention another point. Suppose that we toss the coin 100 times and observe 55 heads. Based on the above discussion we should accept H_0. However, it is often recommended to say "we failed to reject H_0" instead of saying "we are accepting H_0." The reason is that we have not really proved that H_0 is true. In fact, all we know is that the result of our experiment was not statistically contradictory to H_0. Nevertheless, we will not worry about this terminology in this book.