We will test if a given coin is fair or not (actually we will test if it is not a fair coin, i.e., if the probability of heads is different than 0.5). We can express the statement "the coin is a fair coin" mathematically as

$$\rho_H = 0.5$$

where ρ_H is the probability of getting heads in a single toss of the coin. If the coin is not a fair coin, i.e., if the statement $\rho_H = 0.5$ is not true then we must have $\rho_H \neq 0.5$. Hence we have two statements and we would like to decide which one more likely to be true. We will call the statement $\rho_H = 0.5$ the *null hypothesis* (comes from $\rho_H - 0.5 = 0$, i.e., the difference is "null") and the statement $p \neq 0.5$ the *alternative hypothesis* (it is an alternative to the null hypothesis). We will denote the null hypothesis with H_0 and the alternative hypothesis with H_1. With this notation we can simply write the statements we are interested with as

$$\begin{aligned} H_0 : &\quad \rho_H = 0.5 \\ H_1 : &\quad \rho_H \neq 0.5 \ . \end{aligned}$$

We assume that H_0 is true unless we have "sufficient" evidence to believe otherwise.

Note that H_0 and H_1 are statements about a population (actually a parameter). The population of interest in these hypotheses consists of the results obtained from all the past tosses of the coin and all the results that will be obtain in the future tosses of the coin (i.e., a sequence of heads and tails). The parameter in the hypotheses is the relative frequency of heads in this population (hence probability of a getting heads in a single toss). Also the null hypothesis and the alternative hypothesis should be mutually exclusive and together they should cover all possibilities.

Now we have to decide on an experiment to test these hypotheses. One reasonable experiment would be to toss the coin a large number of times. We will use n to denote the number of times we plan to toss the coin. For this case we will assume that we toss the coin 100 times, i.e., $n = 100$. By designing an experiment we are actually deciding how to take a sample from the population. The result of the experiment (i.e., the sample) will consists of a sequence of heads and tails. Now we have to decide how to summarize this sample and how to use this sample to decide whether we should reject H_0 or not. It looks reasonable to count the number of heads obtained (i.e., number of heads in the sample). This number summarizes the sample, hence it is a *statistic*. We will use this statistic to test the truth of H_0. Hence we will call this statistic our *test statistic*.

Now we have to decide how to use the test statistic (number of heads in the sample) to make a decision about H_0. Note that, if we have "too few" heads or "too many" heads then we would suspect the truth of H_0. This is so because, if H_0 is true, i.e., if the coin is fair, (and we assume it is unless we have strong evidence against it) then observing “too few” or “too many” heads is “almost impossible.”

Now we have to clarify what we mean by “almost impossible”, “too few”, and “too many”. First we will start with “almost impossible”. An event is almost impossible if it has a “very small probability.” But what “very small” means could change from person to person. Hence we will state what we mean by a “very small probability” before we conduct the experiment. We will do this by announcing a positive real number close to zero. We will call this the *significance level* and denote it with α. Hence, whatever α is, any event with a probability of α or less we accept as a “almost impossible event.” For this example we will take $\alpha = 0.05$.

Now we must clarify what “too few” and “too many” means. “Too few” will mean less than or equal to some integer L and “too many” will mean greater than or equal to some integer U. Thus if we observe L or less heads then we will say that we observed “too few” heads and if we observe U or more heads then we will say that we observed “too many” heads. Now we have to state what L and U are. The information we have about L and U is in the sentence

> If H_0 is true then observing “too few” or “too many” heads is “almost impossible.”

Recall that by “too few” we mean observing L or less heads, by “too many” we mean U or more heads, and by “almost impossible” we mean the probability is less than or equal to α. Hence if we use X_{n,ρ_H} to denote the number of heads when we toss the coin n times and the probability of heads is ρ_H at each toss, then we can rewrite the above sentence, in mathematical notation, as

$$\mathrm{P}((X_{n,\rho_H} \le L) \cup (X_{n,\rho_H} \ge U)) \le \alpha \ .$$

Since the events $\{X_{n,\rho_H} \le L\}$ and $\{X_{n,\rho_H} \ge U\}$ are disjoint, we can write the above inequality as

$$\mathrm{P}(X_{n,\rho_H} \le L) + \mathrm{P}(X_{n,\rho_H} \ge U) \le \alpha \ . \tag{1}$$

Thus we would like to find a pair of integers, L and U, which satisfies the above inequality. There can be more than one pair which satisfies the above inequality. If L and U satisfies the above inequality then any pair of numbers L' and U' where L' is less than L and U' is greater than U will also satisfy the inequality. Hence, we should find the largest integer L and the smallest integer U that satisfies the above inequality.

Note that any pair of integers, L and U, which satisfies

$$\mathrm{P}(X_{n,\rho_H} \le L) \le \frac{\alpha}{2} \quad \text{and} \quad \mathrm{P}(X_{n,\rho_H} \ge L) \le \frac{\alpha}{2}$$

will also satisfy (1). The above inequalities simplifies the task of finding L and U: we have to find the largest integer L which satisfies

$$\mathrm{P}(X_{n,\rho_H} \le L) \le \frac{\alpha}{2} \ , \tag{2}$$

and the smallest integer U which satisfies

$$\mathrm{P}(X_{n,\rho_H} \geq U) \leq \frac{\alpha}{2} \ . \tag{3}$$

We will start by finding L. In order to find the largest integer L which satisfies (2), we have to calculate $\mathrm{P}(X_{n,\rho_H} \leq k)$ for different values of k. But to calculate $\mathrm{P}(X_{n,\rho_H} \leq k)$ we have to know the distribution of the random variable X_{n,ρ_H}. Note that X_{n,ρ_H} is the number of heads in n tosses of a coin where the probability of heads is ρ_H. Tossing a coin n times and observing the number of heads is a binomial experiment. Thus, the distribution of X_{n,ρ_H} is the binomial distribution with n trial and probability of success ρ_H. Recall that we decided to toss the coin 100 times, hence $n = 100$. Now we have to decide on the value of ρ_H. Note that, we have assumed that H_0 is true unless we have “sufficient” evidence against it. Since we do not have any evidence against H_0 yet we must assume H_0 is true. But this implies that $\rho_H = 0.5$. Hence the distribution of X_{n,ρ_H}, i.e., $X_{100,0.5}$, is $\mathrm{Bin}(100, 0.5)$.

Now we will calculate the probabilities $\mathrm{P}(X_{100,0.5} \leq k)$ for different values of k. Since we do not have the table for $\mathrm{Bin}(100, 0.5)$ we will use `EXCEL`. Using `EXCEL`’s `binomdist` function we construct the following table:

k	$\mathrm{P}(X_{100,0.5} \leq k)$
0	$7.89 \cdot 10^{-31}$
$\vdots$	$\vdots$
36	0.003319
37	0.006016
38	0.010489
39	0.017600
40	0.028444
41	0.044313
$\vdots$	$\vdots$
58	0.955687
59	0.971556
60	0.982400
61	0.989511
62	0.993984
63	0.996681
$\vdots$	$\vdots$
100	1.000000

Recall that we assumed $\alpha = 0.05$. Hence $\alpha/2 = 0.025$. So, from the above table we observe that $\mathrm{P}(X_{100,0.5} \leq 39) \leq 0.025$ but $\mathrm{P}(X_{100,0.5} \leq 40) > 0.025$. Hence the largest integer k which satisfies $\mathrm{P}(X_{100,0.5} \leq k) \leq \alpha/2$ is 39. Therefore $L = 39$.

To find U we can, again, use `EXCEL`. Note that $\mathrm{P}(X_{100,0.5} \geq k) = 1 - \mathrm{P}(X_{100,0.5} \leq k - 1)$. Thus finding the smallest integer k such that

$\mathrm{P}(X_{100,0.5} \geq k) \leq \alpha/2$ is equivalent to finding the smallest integer k which satisfies $\mathrm{P}(X_{100,0.5} \leq k-1) \geq 1-(\alpha/2) = 0.975$. From the table we see that $\mathrm{P}(X_{100,0.5} \leq 59) < 0.975$ and $\mathrm{P}(X_{100,0.5} \leq 60) \geq 0.975$. Thus, we must have $k-1 = 60$. Therefor the smallest integer, k, which satisfies $\mathrm{P}(X_{100,0.5} \geq k) \leq \alpha/2$ is 61. Hence $U = 61$. Note that for this special case (the case where $\rho_H = 0.5$) we can find U simply by using the fact that $\mathrm{Bin}(100, 0.5)$ is symmetric with respect to its expected value. Therefore, if $L = 39$ is the largest integer which satisfies $\mathrm{P}(X_{100,0.5} \leq L) \leq \alpha/2$, then $U = 100-39 = 61$ is the smallest integer which satisfies $\mathrm{P}(X_{n,0.5} \geq U) \leq \alpha/2$.

We conclude that if H_0 is true observing L or less heads, or U or more heads is "almost impossible". Hence we will *reject* H_0 if the number of heads is less than or equal to $L(=39)$, or more than or equal to $U(=61)$, otherwise we will *fail to reject* H_0. We will call this last statement the *decision rule* (the rule that we use to decide whether we should reject H_0 or fail to reject H_0).

Now we can conduct the experiment, i.e., toss the coin 100 times. If we observe 39 or less, or 61 or more heads then we say that we reject the hypothesis that "the coin is a fair coin" otherwise we will fail to reject it.

The idea of choosing a statement is as follows: We have to choose one of the following statements:

1. H_0 is true and an event which is almost impossible has occurred, or
2. it is very likely that H_0 is false.

If H_0 is true observing 39 or less, or 61 or more heads is almost impossible. Thus we either think that H_0 is true and a miracle has happened or we believe that H_0 is probably not true.