Hypothesis Testing: Coin example, with background explanation

We will carry out a hypothesis test to test whether a coin is fair.

The null and alternative hypotheses are as follows:

H0: p = 0.50
H1: p != 0.50

where p = probability of heads on a single coin toss

After carrying out an experiment, a coin is flipped 100 times, and lands heads up 62 times. Is there evidence the coin is biased?

From the Central Limit Theorem, if p is the probability of success, and we expect at least 15 successes and failures, then a sample proportion obtained from $n \geq 30$ independent samples has the following probability distribution:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

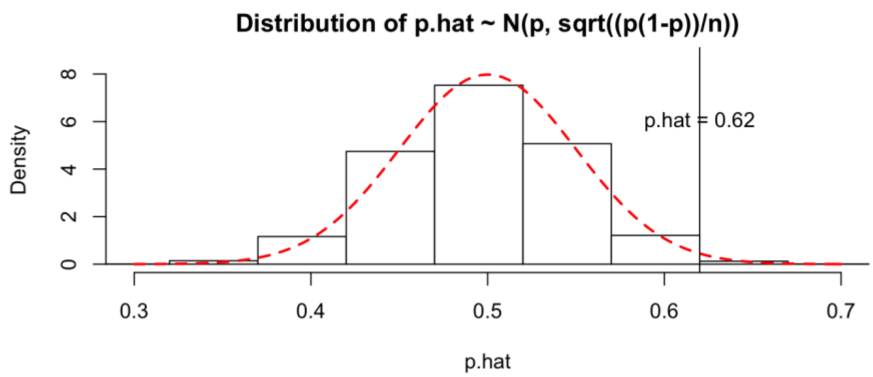
To carry out a hypothesis test, calculate a test statistic

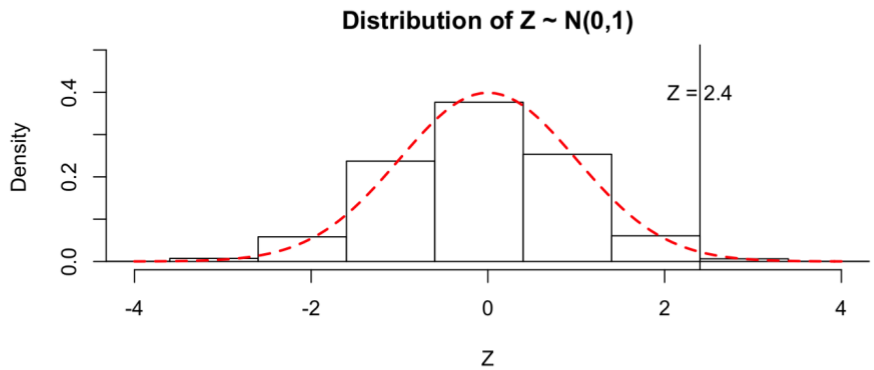
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

and $Z \sim N(0,1)$ if p_0 is the true probability of success according to H_0.

If the observed $\hat{p}$ (or observed test statistic) is unlikely under H_0, then we will reject H_0 (our assumptions are incorrect).

If we assume H_0 is true (that p = 0.50), how unlikely is our observed sample proportion of 0.62?





Context behind the *p-value*: our observed proportion (0.62) differs from the expected value (which is 0.50) by 0.12. How likely is it, under H_0, that out observed proportion would differ by the null (expected) proportion by at least 0.12? What is the probability that the sample proportion differs from the expected proportion by at least the observed amount?

Z-scale interpretation:

Our Z test statistic, calculated using the formula above, is 2.4. Remember Z is the number of standard deviations from the mean. So a question equivalent to the above one is: How likely is it that our observed proportion is at least 2.4 standard deviations from the mean?

In either case, we look at the area under the curve in *both* tails, and find that this p-value is 0.0164.

It is very *unlikely* that we our observation is at least 2.4 standard deviations from the mean – in fact, this only happens about 1.6% of the time. It is equally unlikely that our observed sample proportion would differ from 0.50 by at least a value of 0.12.

Therefore, we reject H_0 and accept the H_1. There is sufficient evidence that this coin is biased (and in this case, biased towards heads).

Conclusions regarding H_0.

- If p-value < 0.05 (or the specified significance level), then reject H_0 and accept H_1. There is sufficient evidence in favor of the alternative hypothesis.

- If p-value is NOT < 0.05, we fail to reject H_0. We say that there is not sufficient evidence to reject H_0.

In both cases, you must state the conclusion in the context of the problem. For example if you are testing that a coin is biased, and your p-value is 0.23, you would conclude that there is NOT sufficient evidence to conclude that the coin is biased.

Note: You can never conclude that H_0 is true! Hypothesis testing is only designed to test for evidence against H_0.

Why?

For example, you flip a coin 30 times and get 20 heads. What do you conclude?