Introduction to Hypothesis Testing

January 25, 2022

In this class we will introduce hypothesis testing. We will be studying the proportion of shots that a player scores correct in a game of paper toss. We will consider this to be a fixed value, p, that is the parameter we are interested in estimating.

For the sake of this class a game of paper toss will be comprised of 10 shots from 3 meters away.

1. Suppose that your shooting percentage is claimed to be 90% or 0.9. We call this claim the 'null' value and denote it p_0. Do you agree with this value?

> **Solution:** 0.9 seems like an unusually high completion rate for this game. As statisticians we cannot rule out the claim being false, but we can formulate a test to speak about the claim probabilistically.

The goal of our hypothesis test is to test the claim that you score a proportion of p_0 correct. We call the original hypothesis our 'null' hypothesis. It can be stated as

$$H_0 : p = p_0$$

2. State the null hypothesis in this scenario

> **Solution:**
>
> $$H_0 : p = 0.9$$

Our alternate hypothesis asks the question: "What is the probability that our true completion rate is the same or more extreme than our data if we assume the null to be true." This can be stated in several different ways for different scenarios.

$$H_a : p \leq p_0$$
$$H_a : p \geq p_0$$
$$H_a : p \neq p_0$$

3. State the alternative hypothesis for this scenario?

> **Solution:** We suspect that the true value of p is lower than $p_0 = 0.9$.
>
> $$H_a : p \leq 0.9$$

To test our claim we would like to gather data. In practice data is collected using sampling, or experimentation. After playing the game once fill in the table below:

Shot Number	Hit/Miss	Count
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
	Total:	

The 'sample proportion', denoted $\hat{p}$, refers to the proportion of shots made in the sample/experiment. This is a **statistic**. Statistics are used to estimate parameters.

4. From our 'experiment' what proportion of shots did you complete?

Solution: When completing the paper toss experiment I scored 4 out the 10 correctly.

$$\hat{p} = 0.4$$

5. In this example, will we have $\hat{p} \sim \text{normal}\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$?

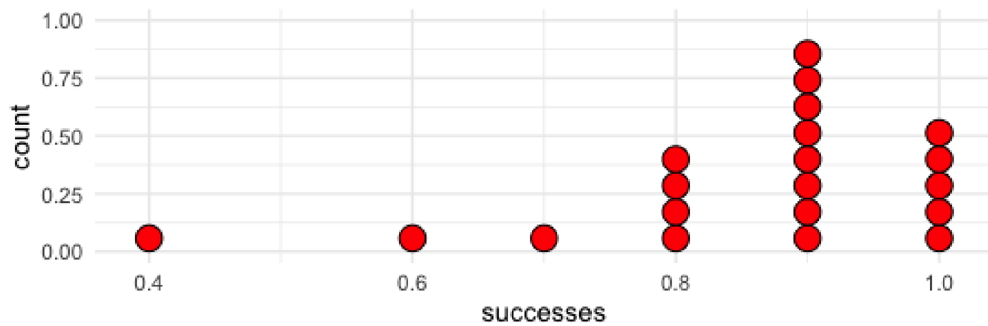
Solution: Here the normality condition is clearly violated. However, it may be reasonable to assume that the shots are independent and random.

To test our data against the null hypothesis we would like to ask the question "How likely is this data if the null hypothesis is true?" We will answer this question in two different ways **simulation** and **theoretical probability**.

Hypothesis Test Using Simulation

6. Assuming that the null hypothesis is true, use a random number generator (or dice) to simulate 20 iterations of the experiment. Plot the value of $\hat{p}$ obtained in each simulation using a dot-plot below

Solution: 20 simulations of the experiment are shown below



7. Looking at your simulated values for $\hat{p}$, what is the probability of observing a value that is equal to or more 'extreme' than your observed value of $\hat{p}$.

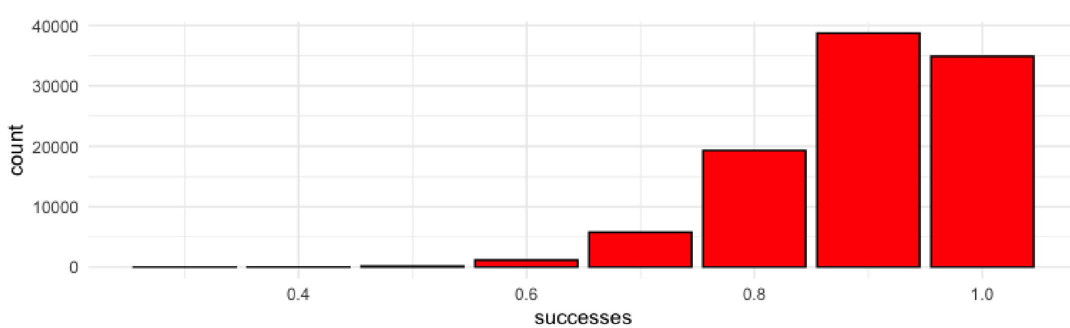
Solution: For our experiment $\frac{1}{20}$ simulations are the same or more extreme than our $\hat{p}$ value of 0.4.

We call this probability the p-value. It indicates the probability of observing more extreme results than our experimental results given we assume the null claim ($p_0 = 0.9$) to be true.

8. Interpret the probability calculated in the previous question.

Solution: Assuming that the true completion rate for my shots is 0.9, the probability of observing a completion rate of 0.4 or lower in future samples is 0.05, or 5%. This low probability gives fairly condemning evidence against the null hypothesis.

Our calculations become more accurate if we consider a larger number of simulations, shown below is a histogram of 100,000 simulations of the experiment.



Hypothesis Test Using Theoretical Probability

We may also calculate the p-value using theoretical probability.

9. Let X be the number of shots that you make assuming that your shots are independent with a completion rate of 0.9. What is the distribution of X?

10. Let x_0 be the number of shots made in your 'experiment'. What is $P(X \leq x_0)$?

> **Solution:** For our experiment we make 10 independent shots, and would like the probability of hitting 4 shots or less. This can be broken up into
>
> $$P(X \leq 4) = P(X = 0) + P(X = 1) + \cdots + P(X = 4)$$
>
> First consider the case with 0 shots made. This can only happen one way: Miss, Miss, Miss, ..., Miss. Since events are independent this probability is
>
> $$P(X = 0) = (0.1)^{10}$$
>
> The case where one shot is being made is very similar, but this can now happen $\binom{10}{1}$ different ways.
>
> $$P(X = 1) = \binom{10}{1}(0.9)(0.1)^9$$
>
> We can continue this line of reasoning up to the case where 4 shots out of 10 are made:
>
> $$P(X \leq 4) = \sum_{x=0}^{4} \binom{10}{x}(0.9)^x(0.1)^{10-x} = 0.0001469026$$

11. Interpret the probability calculated in the previous question.

> **Solution:** Assuming that our completion rate is 0.9, there is a 0.0001469026 chance of observing 4 shots or less in future iterations on the experiment. This extremely low probability is very condemning evidence against the null hypothesis. We say we reject the null hypothesis in favour of the alternative hypothesis.

What Could Go Wrong?

Consider the following table summarizing all possible results from our experiment:

	H_0 True	H_0 False
Reject H_0	Type I Error α	✓
Fail to reject H_0	✓	Type II Error β

It is important to note that our experiment could have produced an unlikely event just due to random chance. A person who is in reality a good shot could have made 10 poor shots. A person who in reality is a poor shot could have made 10 good shots. Notice that the probability of making a type I error is exactly the p-value.

12. What would a type I error be in the context of the paper toss experiment?

Solution: Here we reject H_0 in favour of the claim that the true completion rate is lower than 0.9 when in reality the true completion rate is indeed 0.9. This is also called a 'false positive'.

13. What would a type II error be in the context of the paper toss experiment?

Solution: Here we fail to reject the claim that the completion rate is 0.9 when in reality the completion rate is lower than 0.9. This is also called a 'false negative'.