

## 6.3: Introduction to Hypothesis Testing

---

In this section, we begin a new type of statistical inference known as hypothesis testing. Hypothesis testing can seem awkward at first, but when you really understand it, you see that it's actually how your mind makes decisions after being convinced by sufficient evidence. We will consider the outcomes of flipping a coin versus spinning a coin.

1. What is the theoretical probability that a penny is flipped and lands on tails?
2. What if we flipped the coin 100 times? You said the probability of it landing on tails is \_\_\_\_\_. If that is true, if you flipped a penny 100 times, would you get EXACTLY 50 tails? Explain.
3. Daquan says, "I flipped a penny 100 times for a school project and the penny came up tails \_\_\_\_\_ number of times." Fill in the blank with a number that would make you think Daquan is lying or there was something wrong with his experiment. Then explain your thinking.
4. You and Daquan then have this conversation:  
Daquan: "You know spinning a penny is different than flipping a penny?"  
You: "You mean flipping where you throw it in the air, and spinning where you spin it on a table?"  
Daquan: "Yes! If you spin a penny it usually lands tails side up."  
How would you respond?
5. You try it yourself. You spin a penny 100 times and record the results. The penny lands:  
Heads side up 33 times  
Tails side up 67 times.  
Are you surprised by this result? Do you believe Daquan's claim?

You might be surprised by this result or you might think that the observation may have just happened by chance. Either way, you made a decision about a population (all penny spins) which you cannot observe in its entirety. Your decision was based on a small sample (100 penny spins). In statistics, sample data help us make decisions about populations through a process known as hypothesis testing. A hypothesis is an assumption or claim. We need a formal process to test Daquan's claim.

### Step 1: Determine the null and alternative hypotheses

Let's see if what we observed was unusual enough to convince us that Daquan is correct. We will test Daquan's claim, that a spinning penny lands tails side up the majority of the time.

We will begin by writing some hypotheses:

**The null hypothesis** is the statement of no change (the dull hypothesis). In this context, the proportion of coin spins that land tails up is 50% (the same as flipping a penny). In mathematical symbols,

$$H_0 : p = 0.5$$

Daquan's claim is what we call **the alternative hypothesis**. The proportion of coin spins that land tails up is actually more than 50% (a majority). In mathematical symbols,

$$H_a : \text{---} \text{---} \text{---}$$

### Step 2: Collect Sample Data

We want to know if our observation is unusual, therefore, we want to know if our sample proportion is unusual. We must take a look at the sampling distribution of sample proportions and we hope that it is approximately normal. In our sample, we spin a penny 100 times and it lands tails side up 67 times. This is the number of observed successes in the sample.

a. What is the number of expected successes in our sample (assuming our null hypothesis is true)?

b. What is the number of expected failures in our sample (assuming our null hypothesis is true)?

We want to know if our observation is unusual, therefore, we want to know if our sample proportion is unusual. We must take a look at the sampling distribution of sample proportions. We found that the sampling distribution is approximately normal because there were at least 10 expected successes and failures in our sample.

In our sample, we spin a penny 100 times and it lands tails side up 67 times. What is the sample proportion?

$$\hat{p} = \frac{\text{number of observed successes}}{\text{sample size}} =$$

### Step 3: Assess the Evidence

We want to know if our observation is unusual, therefore, we want to know if our sample proportion is unusual. We should find a Z-score. What is the Z-score for the observed sample proportion?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{67}{100} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} =$$

The correct Z-score test value is \_\_\_\_\_. That means the difference between what I got (67/100 tails) and what most people think I should get (50/100 tails) is 3.4 standard deviations away from what is expected. That is far away from what is expected. Next, we explore probability.

Use the Z-score to find the probability of observing a sample proportion as high or higher than the one we observed. Write the desmos function you used to do the computation.

$$P(\hat{p} \geq \frac{67}{100}) = P(Z \geq 3.4) =$$

If the null hypothesis is true (that the probability of tails is 0.5 when spinning a penny) were true, the probability of getting our result of 67 tails out of 100 spins just by chance is \_\_\_\_\_.

### Step 4: State a Conclusion

Statisticians use a rule about how small a probability should be in order for us to consider an event unusual, or **statistically significant**. We often consider an event unusual if the probability of its occurrence is less than or equal to 5%. This is called the **level of significance**. Other levels of significance can be used.

When the P-value is **less than or equal** to the level of significance, we reject the null hypothesis and support the alternative hypothesis.

Since the P-value is less than or equal to the level of significance, we reject the null hypothesis and support the alternative hypothesis. The sample data support the claim that the proportion of spins of a penny that result in tails is more than 50%. We can support the claim that spins land tails up a majority of the time.

What would be some reasons why spinning a penny results in so many more tails than flipping a penny?

## The Four Step Hypothesis Testing Process

### Step 1. Determine the null and alternative hypotheses

The null hypothesis is a mathematical sentence that makes an assumption of fairness. The alternative hypothesis is a mathematical sentence that represents an opposing or alternative belief.

### Step 2. Collect Sample Data

Compute or record the sample statistic and check that the sampling distribution is normally distributed.

### Step 3. Assess the Evidence

We determine the strength of our evidence through probability. This probability is called a **P-value**, not to be confused with  $p$  which represents a population proportion. The P-value is computed using the assumption made in the null hypothesis. A P-value is the probability of observing a sample statistic that is at least as extreme as the one we observed, assuming the null hypothesis is true.

If our sample proportion differs significantly from the assumed population proportion, then it likely did not occur just by chance.

### Step 4. State a Conclusion

Statisticians use a rule about how small a probability should be in order for us to consider an event unusual, or **statistically significant**. We often consider an event unusual if the probability of its occurrence is less than or equal to 5%. This is called the **level of significance**. Other levels of significance can be used.

When the P-value is **less than or equal** to the level of significance, we reject the null hypothesis and support the alternative hypothesis.

When the P-value is **greater than** the level of significance, we do not reject the null hypothesis and we cannot support the alternative hypothesis.

Lastly, write a conclusion in context in plain language.

## Tips for Writing Conclusions

1. Notice that *we do not support or accept the null hypothesis*. We assume fairness to begin with.
2. *We do not reject the alternative hypothesis*. We either have strong evidence to support it or not.
3. Always say something that makes it clear that your *evidence is based on sample data*. Always include a word that indicates the conclusion is about a *population parameter*. The parameter (proportion, mean, mean difference, standard deviation, etc.) should be included in the statement of the conclusion.