

1 Terminology

Write down a definition, in your own words, for the following terms:

The Null Hypothesis

A hypothesis that says that the data was generated at random under precisely-specified assumptions that *can be simulated on a computer*. The word **null** reinforces the idea that any difference in how the observed data (versus simulated data) looks like is due to *nothing* but chance.

The Alternative Hypothesis

This hypothesis says that some reason other than chance made the data differ from what was predicted by the null hypothesis; the observed difference between the simulated data and the observed data is "real".

The Test Statistic

Statistic: A function that summarizes takes the dataset and returns a number. A test statistic, is a statistic that is used to summarize data for hypothesis testing.

After we've defined our hypotheses, how do we go about testing them?

We simulate the dataset under the assumption that the null hypothesis is true many, many times, and compute a test statistic for each simulation, producing a histogram for the results of our simulations. The idea here is that the null hypothesis says that any difference in the observed sample was simply due to chance. Therefore, when we simulate the data, our simulations could come out differently as well.

2 Create Some Hypotheses

Suppose that you're at the casino, playing dice (with a six-sided die). You suspect that the die is loaded - the dice rolls you see are abnormally high. Define a test statistic, null, and alternative hypotheses.

Null Hypothesis:

The observed dice rolls are like the average of picking numbers at random from the range of 1 to 6.

Alternative Hypothesis:

No the observed average is too high. Something other than chance caused it.

Test Statistic:

The average of the die rolls.

After trying your luck with the dice to no avail, you're back at work as a spearmint gum quality control specialist. You begin to notice that a lot of the gum has minor defects. You suspect that it might be due to more than chance. How do we go about testing this hypothesis?

Null Hypothesis:

The defects are randomly produced, the sample that you observed simply has a large number of defects due to chance.

Alternative Hypothesis:

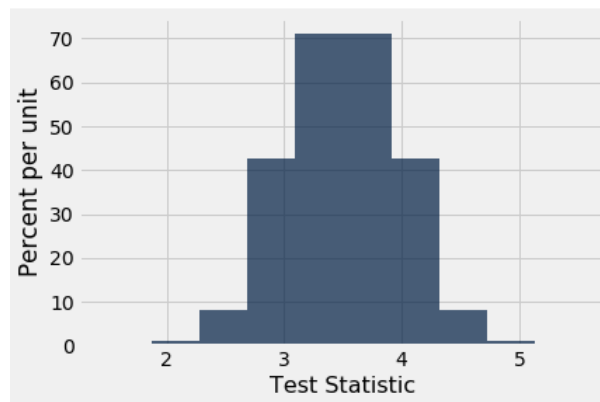
There's something other than chance causing a high level of defects.

Test Statistic:

The number of defects/the number of successes.

3 Evaluate the Hypotheses

After simulating the data for your dice rolls, you produce the following histogram:



If the mean of the dice rolls you observed was 3.923, what could you conclude from the histogram?

From the histogram, it looks like the higher mean from gambling was not at all that unusual - it certainly could have been that section 3's grades were just random samples from the possible rolls. A substantial fraction of the averages were greater than 3.923 if the null hypothesis were true.

If the mean of the dice rolls you observed was instead 5.4, what could you conclude from the histogram?

From the histogram, it now appears that the higher mean from gambling was very unusual - assuming the null hypothesis is true, there isn't much data that appears to be greater than or equal to a mean of 5.4. Therefore, we could suspect that there is something other than chance going on.

Extra Optional question! Given that our data is stored in a table called **dice_data**, how can we create one simulation of the data, and then get our test statistic for that data? **dice_data** has 1 column named "Rolls". Suppose that we also have a table named **possible_dice_rolls** that contains all the possible dice rolls (1 through 6), which has the same column name.

```
simulation = possible_dice_rolls.sample(len(dice_data.column("Rolls"))
test_statistic = np.mean(simulation.column("Rolls"))
```