Last Time: Calculating power is a two-step process.
1) First, we need the distribution of the statistic under the null hypothesis. From that distribution and our level of significance, we determine the rejection region, the values of the statistic that produce a p-value smaller than the level of significance.
2) Then we look at the distribution specified by some alternative value(s) of the parameter(s). From that alternative distribution, we determine the probability of obtaining a statistic in the rejection region.
- Power is influenced by the difference between the null and alternative parameters, the sample size(s), and the level of significance.
- Power calculations are used to determine the sample size(s) necessary before you collect data.

Example 1: Suppose I thought that in the long run the 6 M&M colors were equally likely
<table>
<thead>
<tr>
<th>Blue</th>
<th>Brown</th>
<th>Green</th>
<th>Orange</th>
<th>Red</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

(a) Take a random sample of 50 M&M candies and record the color distribution of your sample.

(b) What are the observational units of this study? How many variables do we have? Quantitative or categorical? Binary?

(c) Let X count the number of blue M&Ms, would it be reasonable to model X with a binomial distribution? [Hint: What are the 4 conditions of a binomial process? (p. 28)] How would you define “success” and “failure”?

(d) If the null hypothesis is true, how many blue did you expect to find in your sample of 50 candies?

(e) Suppose I wanted to test whether $\pi_{\text{blue}} = 1/6$. Write the formula for the z test-statistic and show how it can be considered a comparison of the observed count and the expected count.
Definition: A multinomial process has the following conditions

- There are $n$ trials
- There are $k$ possible outcomes to each trial
- The probabilities of the $k$ outcomes ($\pi_1, \ldots, \pi_k$) are constant from trial to trial and they sum to one ($\sum \pi_i = 1$)
- The trials are independent (the outcome of one trial does not impact the outcome of any other trial)

Then the counts $X_1, \ldots, X_k$ follow a “multinomial” distribution.

(f) Does the multinomial process appear valid for this study? What is $k$?

(g) State the null and alternative hypotheses for testing my entire claim about the color distribution. [Hint: Think of the alternative hypothesis in terms of “not the null hypothesis”]

(h) To assess the evidence against my claim, we first need a statistic. Suggest a possible statistic we could use to summarize our data (i.e., a formula to end up with one number that you could use to assess the evidence against the null hypothesis). What kinds of values will your statistic have if the null is true? What kinds of values of your statistic will you consider evidence against the null hypothesis?

But how do I decide whether my observed statistic provides strong evidence against the null hypothesis? I need to know what kinds of values my statistic has when the null hypothesis is true.

(i) Explain how we could use simulation to generate some “could have been” results, assuming the null hypothesis is true. [Would you use pennies? Spinners? Coins? Dice? Cards?]

(j) Suppose you simulate some “could have been” results for your statistic, how will you compute a p-value? [i.e., what values of your statistic will you consider “more extreme” than what you observed?]

(k) Carry out a test of the null hypothesis using your observed data and the Goodness of Fit applet.