

# Lab 5 - Math 58: binomial

*your name here*

*due Feb 25, 2020*

## Lab Goals

Computing a confidence interval for a single proportion

- calculating binomial probabilities
- comparing binomial and normal probabilities

## Getting started

The lab is designed to help think about simulating probabilities as compared to normal probabilities. *Which do you trust more: a simulation-based or normal-based analysis of an inference question?* In other words, if a simulation analysis and normal approximation give noticeably different p-values, which would you believe to be closer to the correct p-value?<sup>1</sup>

## Load packages

In this lab we will continue to use `infer` and the `xpnorm` function which is in the `mosaic` package.

Let's load the packages.

```
library(tidyverse)
library(mosaic)
library(infer)
```

## The data

Stemming from concern over childhood obesity, researchers investigated whether children might be as tempted by toys as by candy for Halloween treats<sup>2</sup>. Test households in five Connecticut neighborhoods offered two bowls to trick-or-treating children: one with candy and one with small toys. For each child, researchers kept track of whether the child selected the candy or the toy. The research question was whether trick-or-treaters are equally likely to select the candy or toy. More specifically, we will investigate whether the sample data provide strong evidence that trick-or-treaters have a tendency to select either the candy or toy more than the other.

## Background questions (and answers)

- What are the observational units?
- What is the variable, and what type of variable is it?
- What is the population of interest?
- What is the sample?
- Was the sample selected randomly from the population?
- What is the parameter of interest?
- What is the null hypothesis, in words?

---

<sup>1</sup>taken from Allan Rossman, <https://askgoodquestions.blog/>

<sup>2</sup>Schwartz et al, "Trick, Treat, or Toy: Children Are Just as Likely to Choose Toys as Candy on Halloween", *Journal of Nutrition Education and Behavior*, 2003, [https://www.jneb.org/article/S1499-4046\(06\)60335-7/abstract](https://www.jneb.org/article/S1499-4046(06)60335-7/abstract)

- What is the alternative hypothesis, in words?
- What is the null hypothesis, in symbols?
- What is the alternative hypothesis, in symbols?
- What are the observational units? Trick-or-treaters are the observational units.
- What is the variable, and what type of variable is it? The variable is the kind of treat selected by the child: candy or toy. This is a binary, categorical variable.
- What is the population of interest? The population is all trick-or-treaters in the U.S. Or perhaps we should restrict the population to all trick-or-treaters in Connecticut, or in this particular community.
- What is the sample? The sample is the trick-or-treaters in these Connecticut neighborhoods whose selections were recorded by the researchers.
- Was the sample selected randomly from the population? No, it would be very difficult to obtain a list of trick-or-treaters from which one could select a random sample. Instead this is a convenience sample of trick-or-treaters who came to the homes that agreed to participate in the study. We can hope that these trick-or-treaters are nevertheless representative of a larger population, but they were not randomly selected from a population.
- What is the parameter of interest? The parameter is the population proportion of all trick-or-treaters who would select the candy if presented with this choice between candy and toy. Alternatively, we could define the parameter to be the population proportion who would select the toy. It really doesn't matter which of the two options we designate as the "success," but we do need to be consistent throughout our analysis. Let's stick with candy as success.
- What is the null hypothesis, in words? The null hypothesis is that trick-or-treaters are equally likely to select the candy or toy. In other words, the null hypothesis is that 50% of all trick-or-treaters would select the candy.
- What is the alternative hypothesis, in words? The alternative hypothesis is that trick-or-treaters are not equally likely to select the candy or toy. In other words, the alternative hypothesis is that the proportion of all trick-or-treaters who would select the candy is not 0.5. Notice that this is a two-sided hypothesis.
- What is the null hypothesis, in symbols? First we have to decide what symbol to use for a population proportion, here we will use  $p$ . The null hypothesis is  $H_0 : p = 0.5$ .
- What is the alternative hypothesis, in symbols? The two-sided alternative hypothesis is  $H_A : p \neq 0.5$ .

### To Turn In

So, how did the data turn out? The researchers found that 148 children selected the candy and 135 selected the toy. The value of the sample proportion who selected the candy is therefore  $148/283 = 0.523$ .

Let's not lose sight of the research question here: *Do the sample data provide strong evidence that trick-or-treaters have a tendency to select either the candy or toy more than the other?*

How can we investigate whether the observed value of the sample statistic (.523 who selected the candy) would be very surprising under the null hypothesis that trick-or-treaters are equally likely to select the candy or toy? We now have THREE ways to find a p-value for the research questions.

- Simulate
  - Normal approximation
  - Binomial probabilities
1. Let's simulate first. Using the `infer` syntax, provide a histogram which describes the variability of the sample statistics under the null hypothesis.
    - Describe the shape, center, and variability of the distribution of these simulated sample proportions.

- What do we look for in the graph, in order to assess the strength of evidence about the research question?
  - Well, does it appear that 0.523 is unusual?
  - So, what do we conclude about the research question, and why?
  - What is the p-value of the test?
2. Now let's try applying the normal distribution.
    - Is the normal distribution appropriate as an approximation to the simulation above?
    - What is the corresponding p-value associated with the hypotheses above?
    - Notice that the p-value for the simulation and the normal approximation are different, which one is right?
  3. Can we calculate the *exact* p-value? YES! Through the binomial distribution.
    - In a sample of size 283, what are the numbers of children that would be “more extreme” than what we saw (148 children choosing toy). Your answer should say: a sample would be more extreme if it had \_\_\_\_\_ or more children choosing toys OR \_\_\_\_\_ or fewer children choosing toys.
    - Use the binomial distribution (`xpbinom`) to find the probability that a random sample (with  $p=0.5$ ) would have results that were more extreme than the observed data.
    - Which is closer to the exact binomial p-value, the simulation or the normal approximation? Explain.
  4. What does the analysis say about the trick-or-treaters? In particular, are you convinced that trick-or-treaters are equally likely to choose a toy or candy?