Hypothesis Testing (part 1, null models)

Ryan Miller

Introduction

- Last week, we focused on the *sample average* (proportion) as a *random variable*
  - Central Limit theorem gave us a Normal model for the *sampling distribution* of the sample average
  - This allowed us to come up with meaningful interval estimates (confidence intervals) of the *population average* (proportion)

Introduction

- Last week, we focused on the *sample average* (proportion) as a *random variable*
  - Central Limit theorem gave us a Normal model for the *sampling distribution* of the sample average
  - This allowed us to come up with meaningful interval estimates (confidence intervals) of the *population average* (proportion)
- This week, we'll remain focused on the sample average, but we will shift our attention to **hypothesis testing**, or using probability to statistically evaluate certain conjectures about a population

Example - Infants Choosing Toys

- On the first day of class, we discussed an experiment published in *Nature* investigating whether infants had preference for a "helper" character

Example - Infants Choosing Toys

- On the first day of class, we discussed an experiment published in *Nature* investigating whether infants had preference for a "helper" character
- 16 infants repeatedly watched demonstrations of two scenarios
  - The "helper" toy assisting the main character
  - The "hinderer" toy blocking the main character

Example - Infants Choosing Toys

- On the first day of class, we discussed an experiment published in *Nature* investigating whether infants had preference for a "helper" character
- 16 infants repeatedly watched demonstrations of two scenarios
  - The "helper" toy assisting the main character
  - The "hinderer" toy blocking the main character
- When given the choice, 14 of 16 infants chose the "helper" toy

Example - Infants Choosing Toys

Statisticians will approach an experiment like this by trying to rule out possible explanations:

- Could the majority have chosen the "helper" toy due to a *confounding variable* like the toy's color or shape?

Example - Infants Choosing Toys

Statisticians will approach an experiment like this by trying to rule out possible explanations:

- Could the majority have chosen the “helper” toy due to a *confounding variable* like the toy’s color or shape?
  - No, recall that these things were *randomly assigned*

Example - Infants Choosing Toys

Statisticians will approach an experiment like this by trying to rule out possible explanations:

- Could the majority have chosen the “helper” toy due to a *confounding variable* like the toy’s color or shape?
  - No, recall that these things were *randomly assigned*
- Could the majority have chosen the “helper” toy due to some type of *bias*?

Example - Infants Choosing Toys

Statisticians will approach an experiment like this by trying to rule out possible explanations:

- Could the majority have chosen the "helper" toy due to a *confounding variable* like the toy's color or shape?
  - No, recall that these things were *randomly assigned*
- Could the majority have chosen the "helper" toy due to some type of *bias*?
  - Probably not, measurement of this outcome is pretty clear-cut

Ideally, statisticians are left with only two viable explanations for an observed outcome: random chance/luck or a real relationship

Example - Infants Choosing Toys

The remaining step is now to rule out random chance/luck as a viable explanation. To do so, statisticians apply the following logic:

1) Identify a suitable **null model** for the outcome of interest
2) Calculate the probability of seeing the outcome that occurred in the sample data if the null model were true
3) If this probability is sufficiently small, rule out random chance as an explanation

Example - Infants Choosing Toys

- In this example, a suitable null model would be $p = 0.5$, which implies that each baby's choice is just a coin-flip
  - Under this null model, we can investigate what outcomes could occur by random chance alone

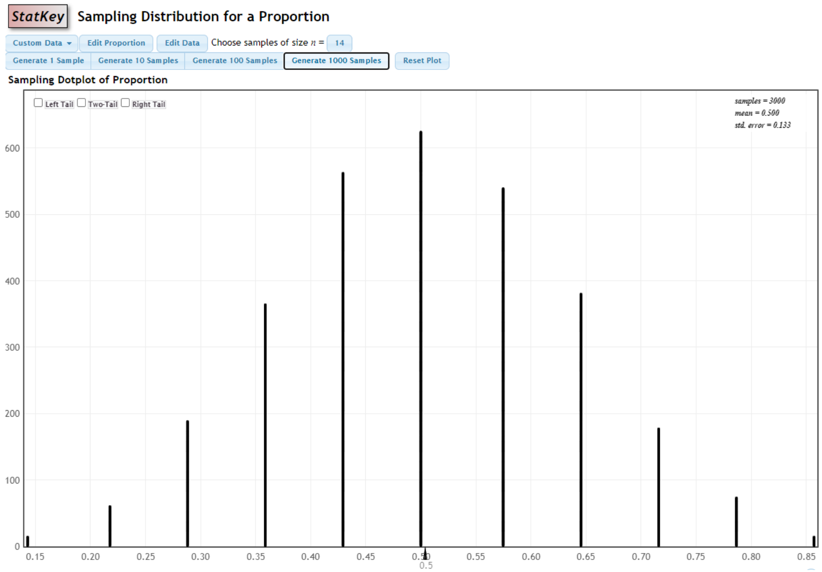
Example - Infants Choosing Toys

- In this example, a suitable null model would be $p = 0.5$, which implies that each baby's choice is just a coin-flip
    - Under this null model, we can investigate what outcomes could occur by random chance alone
- One approach is to use simulation, or flip sets of 16 coins on StatKey

Example - Infants Choosing Toys

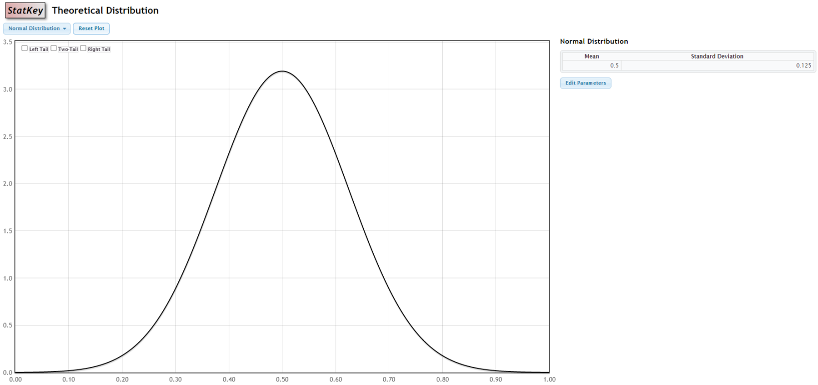
- In this example, a suitable null model would be $p = 0.5$, which implies that each baby's choice is just a coin-flip
  - Under this null model, we can investigate what outcomes could occur by random chance alone
- One approach is to use simulation, or flip sets of 16 coins on StatKey
- Another approach is to realize that Central Limit theorem can be used to determine the distribution of sample proportions under the null model

The Simulation Approach



The CLT Approach

Note that according to CLT, $SE = \sqrt{.5 * .5/16} = 0.125$



The Null Distribution

- Both of these distributions are examples of a **null distribution**, or a display of outcomes that could have been observed in the study *if the null model were true*

The Null Distribution

- Both of these distributions are examples of a **null distribution**, or a display of outcomes that could have been observed in the study *if the null model were true*
- The null distribution is used to gauge how compatible the actual sample data are with the outcomes we'd expect if the null model were true
  - In our example, we observed a sample proportion of $\hat{p} = 14/16 = 0.875$, which appears to be a very unlikely outcome under the null model

The Null Distribution

- Both of these distributions are examples of a **null distribution**, or a display of outcomes that could have been observed in the study *if the null model were true*
- The null distribution is used to gauge how compatible the actual sample data are with the outcomes we'd expect if the null model were true
  - In our example, we observed a sample proportion of $\hat{p} = 14/16 = 0.875$, which appears to be a very unlikely outcome under the null model
- In the next presentation, we'll introduce the *p*-value as a statistical tool to more precisely measure the amount of evidence against a null model