Introduction to Statistical Investigations

Second Edition

Tintle, Chance, Cobb, Rossman, Roy, Swanson, VanderStoep

Chapter 1

Significance: How Strong is the Evidence?

Definitions so far

- **Sample (space):** The values we observe and write down.
  - **Example:** Harley the dog chooses the right cup or the wrong cup.
- **Sample size:** The number of observational units.
  - **Example:** Harley has 2 choices (right/wrong cup).
- **Statistic:** A number computed to summarize the variable we measured on a sample.
  - **Example:** Harley the dog chose the correct cup 9 times out of 10, so 9/10.
- **NEW! Parameter:** A long-run numerical property of the process. Long-run proportion

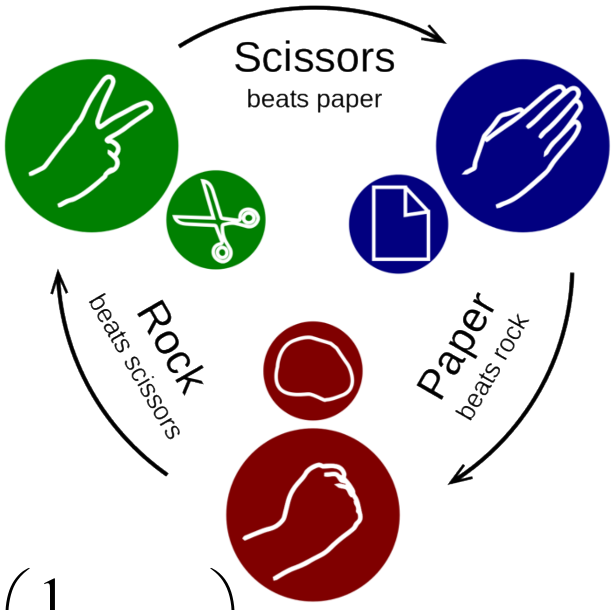
Definitions so far

- **Chance model**: using a random process to model the situation.
- **3S strategy:**
  - **Statistic:** a number that we observe from the data.
  - **Simulate:** Write up what a "by-chance" model would give us. Repeatedly conduct the experiment, and see how our observed statistic compares.
  - **Strength of evidence:**
    - Determine if the observed statistic is *unlikely/likely* to have occurred from the chance model.

Section 1.2: Measuring Strength of Evidence

- In the previous section we performed **tests of significance**.
- In this section we will make things slightly more complicated, formalize the process, and define new terminology.

Example 1.2: Rock-Paper-Scissors (1 of 2)



- Rock-Paper-Scissors Rules
  - Rock smashes scissors
  - Paper covers rock
  - Scissors cut paper
- **Are these choices used in equal proportions** $\left(\frac{1}{3}\text{ each}\right)$?

Example 1.2: Rock-Paper-Scissors (2 of 2)

- One study suggests, among novice players, that scissors are chosen less than $\frac{1}{3}$ of the time.
- Suppose we are going to test this with a friend (who hasn't played before) by playing 20 games against you.
- What are the observational units?
- What is the variable?

Example 1.2: Rock-Paper-Scissors

- Even though there are three outcomes, we are focusing on whether the player chooses scissors or not. This is called a **binary** variable because we are focusing on *two* outcomes (both not necessarily equally likely).

Example 1.2: Terminology: Hypotheses

- When conducting a test of significance, one of the first things we do is state the null and alternative hypotheses.
- The **null hypothesis** is the chance explanation.
- Typically the **alternative hypothesis** is what the researchers think is true.

Example 1.2: Null vs. Alternative Hypotheses

Example 1.2: Hypotheses from Buzz and Doris

- **Null Hypothesis:** Buzz will randomly choose a button. (He chooses the correct button 50% of the time, in the long run.)
- **Alternative Hypothesis:** Buzz understands what Doris is communicating to him. (He chooses the correct button more than 50% of the time, in the long run.)

These hypotheses represent the parameter (*long run behavior*) not the statistic (*the observed results*).

Example 1.2: Hypotheses for Rock Paper Scissors using Words

- **Null Hypothesis:** People playing Rock-Paper-Scissors will equally choose between the three options. (In particular, they will choose scissors one-third of the time, in the long run.)
- **Alternative Hypothesis:** People playing Rock-Paper-Scissors will choose scissors less than one-third of the time, in the long run.

Note the differences (and similarities) between these hypotheses and the hypotheses for Buzz and Doris.

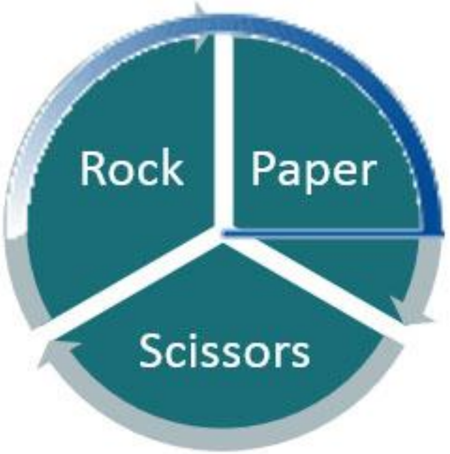
Example 1.2: Hypotheses for Rock Paper Scissors using Symbols

- $H_0: \pi = \frac{1}{3}$
- $H_a: \pi < \frac{1}{3}$

where π is players' true probability of throwing scissors

Example 1.2: Setting Up a Chance Model

- Because the Buzz and Doris example had a 50% chance outcome, we could use *a coin* to model the outcome from one trial.
- What could we use in the case of Rock-Paper-Scissors?



Example 1.2: Three S Strategy

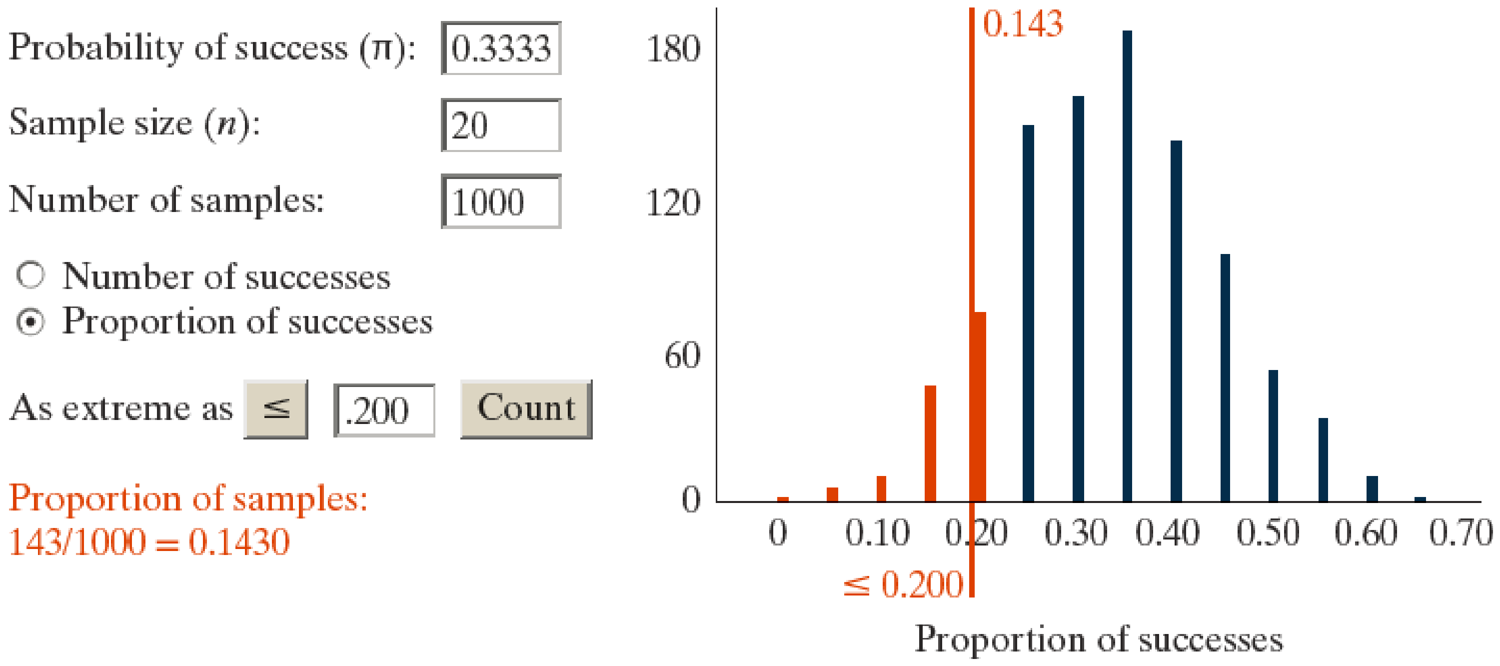
- **Statistic:** Compute the statistic from the observed data. **[In 20 games, scissors was chosen 4 times. This sample proportion can be described using the symbol $\hat{p}$ (p - hat)].**
- **Simulate**: Identify a model that represents a chance explanation. Repeatedly simulate values of the statistic that could have happened when the chance model is true and form a distribution.
- **Strength of evidence**: Consider whether the value of the observed statistic is unlikely to occur when the chance model is true.

Example 1.2: Applet

- We will use the One Proportion applet for our test.
- This is the same applet we used last time except now we will change the probability under the null hypothesis.
- Let's go to the applet and run the test. (Notice the use of symbols in the applet and how they change when we change 0.50 to 0.3333.)

Example 1.2: Null Distribution

- The **null distribution** is the distribution of simulated statistics that could have happened in the study assuming the null hypothesis was true.



Example 1.2: *p*-value

- The ***p*-value** is the proportion of the simulated statistics in the null distribution that are at least as extreme (in the direction of the alternative hypothesis) as the value of the statistic actually observed in the research study.
- We should have seen something similar to this in the applet:
  - Proportion of samples: $\frac{143}{1000} = 0.143$

Example 1.2: What Can We Conclude? (1 of 2)

- Do we have strong evidence that less than $\frac{1}{3}$ of the time scissors is thrown?
- How small of a *p*-value would you say gives strong evidence?
- **Remember: the smaller the *p*-value, the stronger the evidence against the null.**

Example 1.2: Guidelines for Evaluating Strength of Evidence from *p*-values

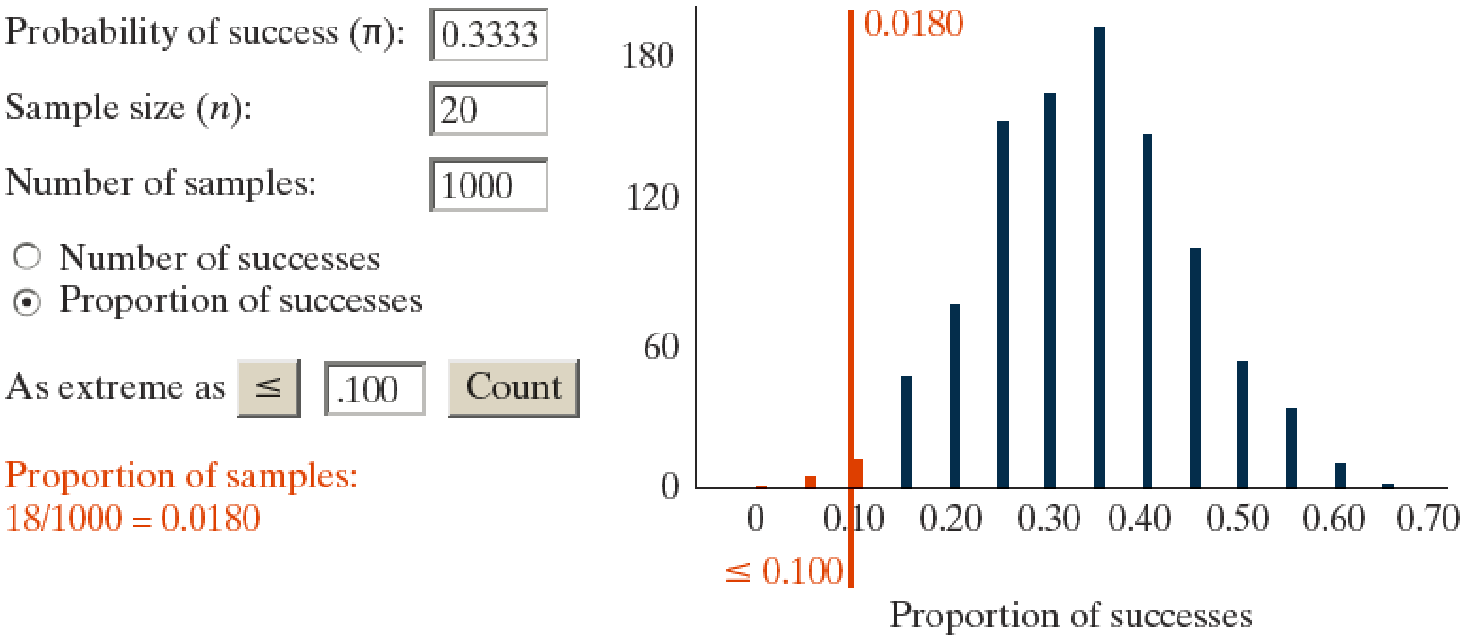
- p-value > 0.10, not much evidence against null hypothesis
- $0.05 < p$-value ≤ 0.10, moderate evidence against the null hypothesis
- $0.01 < p$-value ≤ 0.05, strong evidence against the null hypothesis
- p-value ≤ 0.01, very strong evidence against the null hypothesis

Example 1.2: What Can We Conclude? (2 of 2)

- So we do not have strong evidence that fewer than $\frac{1}{3}$ of the time scissors is thrown.
- Does this mean we can conclude that $\frac{1}{3}$ of the time scissors is thrown?
- Is it plausible that $\frac{1}{3}$ of the time scissors is thrown?
- Are other values plausible? Which ones?
- What could we do to have a better chance of getting strong evidence for our alternative hypothesis?

Example 1.2: Rock Paper Scissors Redo

- What if only 2 out of 20 times (10%) scissors was played? How would this change the *p*-value?
- Values of the statistic that are even farther from the hypothesized parameter give smaller *p*-values and stronger evidence against the null.



Example 1.2: Summary (1 of 2)

- The **null hypothesis** (H_0) is the chance explanation. (=)
- The **alternative hypothesis** (H_a) is what you are trying to show is true. (< or >)
- A **null distribution** is the distribution of simulated statistics that represent the chance outcome.
- The ***p*-value** is the proportion of the simulated statistics in the null distribution that are at least as extreme as the value of the observed statistic.

Example 1.2: Summary (2 of 2)

- The smaller the p-value, the stronger the evidence against the null.
- p-values less than 0.01 provide strong evidence against the null.
- π represents the population parameter
- $\hat{p}$ represents the sample proportion
- n represents the sample size

Learning Objectives for Section 1.2 (1 of 2)

- Use appropriate symbols for parameter and statistic.
- State the null and the alternative hypotheses in words and in terms of the symbol π, the long-run proportion.
- Explain how to conduct a simulation using a null hypothesis probability that is not 50-50.
- Use the One Proportion applet to obtain the *p*-value after carrying out an appropriate simulation.