Lecture 19: Chapter 8, Section 1 Sampling Distributions: Proportions

- Typical Inference Problem
- Definition of Sampling Distribution
- 3 Approaches to Understanding Sampling Dist.
- Applying 68-95-99.7 Rule

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-4)
  - Displaying and Summarizing (Lectures 5-12)
  - Probability
    - Finding Probabilities (discussed in Lectures 13-14)
    - Random Variables (discussed in Lectures 15-18)
    - Sampling Distributions
      - Proportions
      - Means
  - Statistical Inference

Typical Inference Problem

If sample of 100 students has 0.13 left-handed, can you believe population proportion is 0.10?

Solution Method: Assume (temporarily) that population proportion is 0.10, find probability of sample proportion as high as 0.13. If it's too improbable, we won't believe population proportion is 0.10.

Key to Solving Inference Problems

For a given population proportion p and sample size n, need to find probability of sample proportion $\hat{p}$ in a certain range:

Need to know **sampling distribution** of $\hat{p}$.

Note: $\hat{p}$ can denote a single statistic or a random variable.

Definition

Sampling distribution of sample statistic tells probability distribution of values taken by the statistic in repeated random samples of a given size.

Looking Back: *We summarize a probability distribution by reporting its* ***center, spread, shape.***

Behavior of Sample Proportion *(Review)*

For random sample of size n from population with p in category of interest, sample proportion $\hat{p} = \frac{X}{n}$ has

- mean p
- standard deviation $\sqrt{\frac{p(1-p)}{n}}$
- shape approximately normal for large enough n

Looking Back: *Can find normal probabilities using 68-95-99.7 Rule, etc.*

Rules of Thumb *(Review)*

- Population at least 10 times sample size n
  (formula for standard deviation of $\hat{p}$ approximately correct even if sampled without replacement)
- np and $n(1-p)$ both at least 10
  (guarantees $\hat{p}$ approximately normal)

Understanding Dist. of Sample Proportion

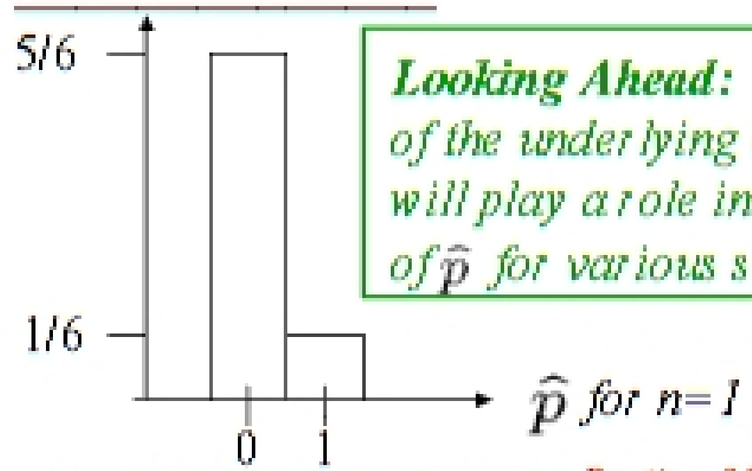
3 Approaches:

1. Intuition
2. Hands-on Experimentation
3. Theoretical Results

Looking Ahead: *We'll find that our* ***intuition*** *is consistent with* ***experimental*** *results, and both are confirmed by mathematical* ***theory****.*

Example: *Shape of Underlying Distribution (n=1)*

- **Background**: Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** How does the probability histogram for sample proportions appear for samples of size 1?
- **Response:** ____________



***Looking Ahead:** The shape of the underlying distribution will play a role in the shape of $\hat{p}$ for various sample sizes.*

Example: *Sample Proportion as Random Variable*

- **Background**: Population proportion of blue M&Ms is 0.17.
- **Questions:**
  - Is the underlying variable categorical or quantitative?
  - Consider the behavior of sample proportion $\hat{p}$ for repeated random samples of a given size. What type of variable is sample proportion?
  - What 3 aspects of the distribution of sample proportion should we report to summarize its behavior?
- **Responses:**
  - Underlying variable ____________________
  - ____________________
  - Summarize with ________, ________, ________

Example: *Center, Spread of Sample Proportion*

- **Background**: Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** What can we say about center and spread of $\hat{p}$ for repeated random samples of size $n = 25$ (a teaspoon)?
- **Response:**
  - **Center:** Some $\hat{p}$'s more than ___, others less; should balance out so mean of $\hat{p}$'s is p = ______.
  - **Spread** of $\hat{p}$'s: s.d. depends on ____.
    - For n=6, could easily get $\hat{p}$ anywhere from ___ to ___.
    - For n=25, spread of $\hat{p}$ will be ____ than it is for $n = 6$.

Example: *Intuit Shape of Sample Proportion*

- **Background**: Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** What can we say about the shape of $\hat{p}$ for repeated random samples of size $n = 25$ (a teaspoon)?
- **Response:**

  $\hat{p}$ close to ____ most common, far from _____ in either direction increasingly less likely →

  ____