Lecture 22/Chapter 19
Part 4. Statistical Inference Ch. 19
Diversity of Sample Proportions

- Probability versus Inference
- Behavior of Sample Proportions: Example
- Behavior of Sample Proportions: Conditions
- Behavior of Sample Proportions: Rules

Course Divided into Four Parts *(Review)*

1. Finding Data in Life:
   scrutinizing origin of data
2. Finding Life in Data: summarizing data yourself or assessing another's summary
3. Understanding Uncertainty in Life:
   probability theory (completed)
4. Making Judgments from Surveys and Experiments: statistical inference

Approach to Inference

- **Step 1 (Chapter 19):** Work *forward*---if we happen to know the population proportion falling in a given category, what behavior can we expect from sample proportions for repeated samples of a given size?
- **Step 2 (Chapter 20):** Work *backward*---if sample proportion for a sample of a certain size is observed to take a specified value, what can we conclude about the value of the unknown population proportion?

After covering Steps 1&2 for proportions, we'll cover them for means.

Understanding Sample Proportion

3 Approaches:

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

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

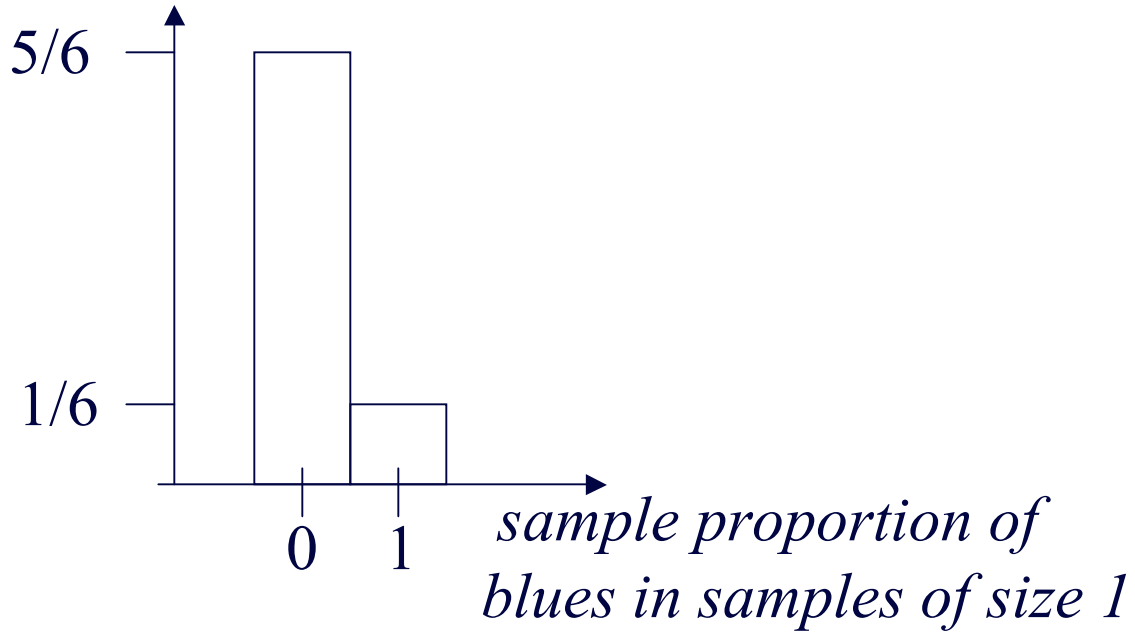
Example: *Intuit Behavior of Sample Proportion*

- **Background**: Population proportion of blue M&M's is 1/6=0.17.
- **Question:** How does sample proportion behave for repeated random samples of size 25 (a teaspoon)?
- **Response:** Summarize by telling
- Experiment: sample teaspoons of M&Ms, record sample proportion of blues on sheet **and** in notes (need a calculator)

Example: *Intuit Behavior of Sample Proportion*

- **Background**: Population proportion of blue M&M's is 1/6=0.17.

Note: The shape of the underlying distribution (sample size 1) will play a role in the shape of sample proportions for various sample sizes.



Example: *Intuit Behavior of Sample Proportion*

- **Response:** (continued)
  - **Center:** some sample proportions will be less than 0.17 and others more; the mean of all sample proportions should be
  - **Spread**: depends on sample size; if we'd sampled only 5, we'd easily get sample proportions ranging from 0 to 0.6 or 0.8. For samples of 25, proportions
  - **Shape:** proportions close to would be most common, and those far from increasingly less likely---shape

Example: *Intuit Behavior of Sample Proportion*

- **Background**: Population proportion of blue M&M's is 1/6=0.17.
- **Question:** How does sample proportion behave for repeated random samples of size 75 (a Tablespoon)?
- **Response:** Again, we summarize by telling

  - Now sample Tablespoons of M&Ms, record sample proportion of blues on sheet **and** in notes (need a calculator)

Example: *Intuit Behavior of Sample Proportion*

- **Response:** (samples of size 75)
  - **Center:** The mean of all sample proportions should be _____ regardless of sample size.
  - **Spread**: should be _____ than what it would be for samples of size 25.
  - **Shape:** should bulge more close to 0.17, taper more at the ends, less right-skewness: it should be _____

Conditions for Rule of Sample Proportions

- Randomness [affects center]
  - Can't be biased for or against certain values
- Independence [affects spread]
  - If sampling without replacement, sample should be less than 1/10 population size
- Large enough sample size [affects shape]
  - Should sample enough to expect at least 5 each in and out of the category of interest.

Example: *Checking Conditions for Rule*

- **Background**: Population proportion of blue M&M's is 1/6=0.17. Students repeatedly take random samples of size 1 teaspoon (about 25) and record the proportion that are blue.
- **Question:** Are the 3 Conditions met?
- **Response:**

1.
2.
3.

Example: *Checking Conditions (larger sample)*

- **Background**: Population proportion of blue M&M's is 1/6=0.17. Students repeatedly take random samples of size 1 Tablespoon (about 75) and record the proportion that are blue.
- **Question:** Are the 3 Conditions met?
- **Response:**

1.
2.
3.

Rule for Sample Proportions

- **Center:** The mean of sample proportions equals the true population proportion.
- **Spread:** The standard deviation of sample proportions is standard error =

$$\sqrt{\frac{\text{population proportion} \times (1 - \text{population proportion})}{\text{sample size}}}$$

- **Shape:** (Central Limit Theorem) The frequency curve of proportions from the various samples is approximately normal.

Example: *Applying Rules for Sample Proportions*

- **Background**: Proportion of blue M&Ms is 1/6=0.17.
- **Question:** What does the Rule tell us about sample proportions that are blue in teaspoons (about 25)?
- **Response:**
  - Center: the mean of sample proportions will be
  - Spread: the standard deviation of sample proportions will be

    standard error =
  - Shape:

Example: *Applying Rules for Sample Proportions*

- **Background**: Proportion of blue M&Ms is 1/6=0.17.
- **Question:** What does the Rule tell us about sample proportions that are blue in Tablespoons (about 75)?
- **Response:**
  - Center: the mean of sample proportions will be
  - Spread: the standard deviation of sample proportions will be

    standard error =
  - Shape:

Empirical Rule *(Review)*

For any normal curve, approximately

- 68% of values are within 1 sd of mean
- 95% of values are within 2 sds of mean
- 99.7% of values are within 3 sds of mean

Example: *Applying Empirical Rule to M&Ms*

- **Background**: Population proportion of blue M&M's is 1/6=0.17. Students repeatedly take random samples of size 1 Tablespoon (about 75) and record the proportion that are blue.
- **Question:** What does the Empirical Rule tell us?
- **Response:**
  - 68% of the sample proportions should be within : in [0.127, 0.213]
  - 95% of the sample proportions should be within : in [0.084, 0.256]
  - 99.7% of the sample proportions should be within : in [0.041, 0.299]

How well did our sampled proportions conform?

Proportions then Means, Probability then Inference

Next time we'll establish a parallel theory for means, when the variable of interest is quantitative (number on dice instead of color on M&M). After that, we'll

- Perform inference with **confidence intervals**
  - For proportions (Chapter 20)
  - For means (Chapter 21)
- Perform inference with **hypothesis testing**
  - For proportions (Chapters 22&23)
  - For means (Chapters 22&23)