Module 1 – Section 3

Sampling Distribution for a Sample Proportion

Outline

- Population Proportion p and Sample Proportion $\hat{p}$
- Sampling Distribution of $\hat{p}$
- Connection to the Binomial Distribution

Population Proportion p

- $p =$ proportion of population belonging to a particular category (category of interest or success)
  - Ex. Proportion of U.S. adults who believe the coronavirus situation is getting a little or a lot worse in the U.S. today.
- Value of p is generally unknown, but of interest to researchers.

Estimating p

- Take simple random sample of size n from population
- Assume simple random sample taken from large population
  - $n < 10\%$ of population size
  - We can treat outcome from simple random sample of size n the same as outcome of n independent and identical trials.

Estimating p

- Denote $Y =$ number of successes in the sample of size n
- Estimate p with $\hat{p} = \frac{Y}{n}$
- $\hat{p} =$ sample proportion of observations in category of interest

Summary of Sample

- General Relative Frequency Table

Outcome	Proportion
Success	$\hat{p}$
Failure	$1 - \hat{p}$
Total	1

Ex. Summary of Sample

- Gallup Poll: Out of 3,104 randomly selected adults in the U.S., 1,956 responded they believe the coronavirus situation is getting a little or a lot worse in the U.S. today.
  - $\hat{p} = \frac{1956}{3104} = 0.63$

Understanding behavior of $\hat{p}$

- Random variable Y varies from sample to sample
- Random variable $\hat{p} = \frac{Y}{n}$ varies from sample to sample
- Description of this variability = sampling distribution of $\hat{p}$

Sampling Distribution of $\hat{p}$

- Three components
  - Mean
  - Variance
  - Shape

Sampling Distribution of $\hat{p}$

- Mean

$$E(\hat{p}) = p$$

- On average, the sample proportion is equal to p.
- $\hat{p}$ is an unbiased estimator for p.

Sampling Distribution of $\hat{p}$

- Variance

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

- Variability of sample proportion $\hat{p}$ around p.
- Variability of $\hat{p}$ decreases as sample size increases.
- More variability when p is near 0.5; less variability when p is near 0 or 1.

Sampling Distribution of $\hat{p}$

- Standard Deviation

$$SE(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$$

Shape

- Depends on values of n and p.
- If $np \geq 10$ and $n(1-p) \geq 10$ then sampling distribution of $\hat{p}$ is close to normal distribution.

Ex. Sampling Distribution

- Suppose that 10% of all people are left-handed
  - $p = 0.1$
- Collect information from each person in sample of size $n = 250$ about dominant hand
  - $Y =$ number of people in sample who are left-handed
  - $\hat{p} = \frac{Y}{250} =$ proportion of left-handed people in sample

Ex. Sampling Distribution

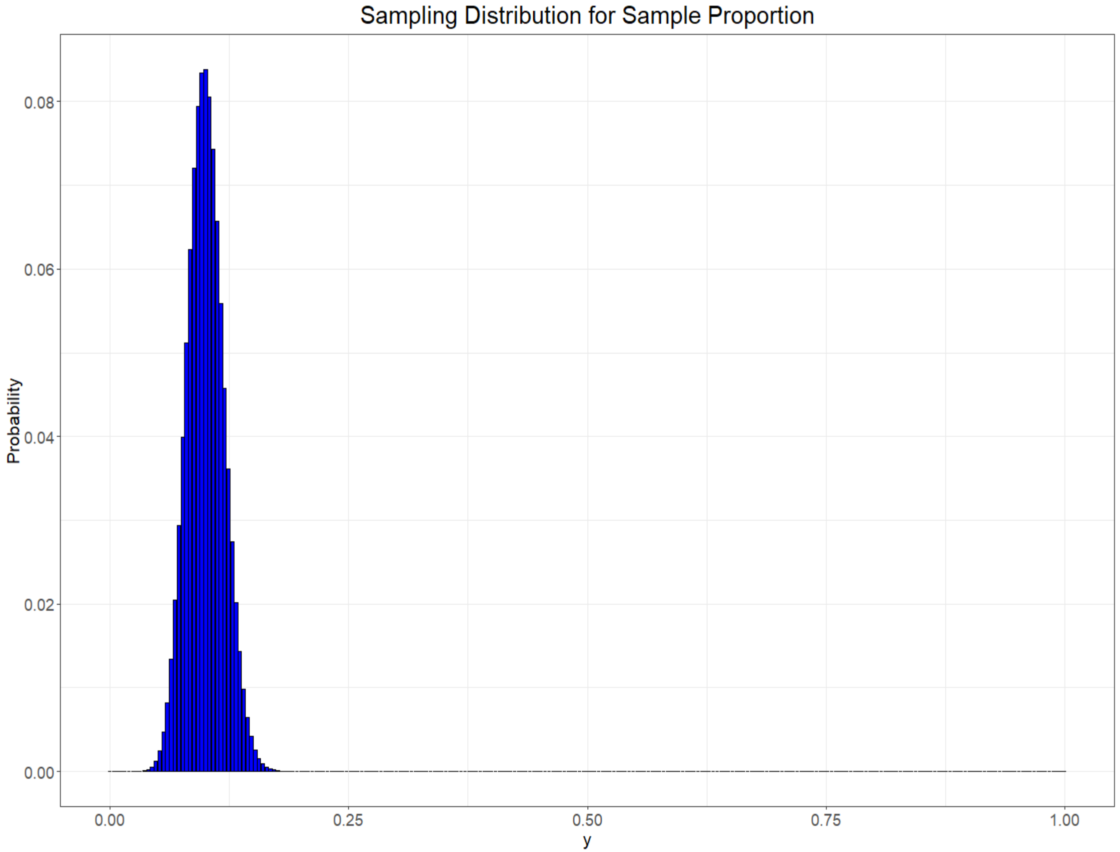
- Mean
  - $E(\hat{p}) = 0.1$
  - On average, the sample proportion of left-handed people will be 0.1, the same as the population proportion.

Ex. Sampling Distribution

- Standard Deviation
  - $SE(\hat{p}) = \sqrt{\frac{0.1(0.9)}{250}} = 0.0190$
  - The sample proportion will vary from $p = 0.1$ by around 0.0190

Ex. Sampling Distribution

- Shape
  - $250(0.1) = 25$
  - $250(0.9) = 225$
  - Approximately Normal Distribution



Probabilities from Sampling Distribution

- Determine probabilities about sample proportions using Normal sampling distribution
- Appropriate if $np \geq 10$ and $n(1-p) \geq 10$

Ex. Probabilities

- Probability less than 8% (0.08) of a sample of $n = 250$ people will be left-handed.

Ex. Probabilities

- Probability more than 13% (0.13) of a sample of $n = 250$ people will be left-handed.

Comparison

Binomial Distribution

- Random Variable
  - $Y =$ number of successes in n independent and identical trials
- Values
  - $n + 1$ possible values from 0 to n $(0, 1, 2, \dots, n)$

Sampling Distribution of $\hat{p}$

- Random Variable
  - $\hat{p} = \frac{Y}{n} =$ proportion of successes in sample of size n
- Values
  - $n + 1$ possible values from 0 to 1 $\left(0, \frac{1}{n}, \frac{2}{n}, \dots, 1\right)$

Comparison

- Sampling Distribution of $\hat{p}$ and Binomial Distribution
  - Same distribution – both based on Y
  - Different scale (0 to 1 vs. 0 to n)

Implications

- Sampling Distribution of $\hat{p}$
  - Discrete Distribution
    - Values only possible at j/n where j is an integer from 0 to n.
  - Approximated using continuous Normal Distribution
    - Shape
    - Discrete nature of values of $\hat{p}$

Add

- Want to show how we can use result to get probabilities of outcomes.
- Add this information to the Class Activities page.
- Will need to introduce pnorm in the Tech Guide.