Outline

1 Recap

2 Proportions

3 Sampling Distribution of \( \hat{p} \)
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 22.
Sampling Distribution of $\bar{X}$

- Suppose that $\bar{X}$ is the sample mean of a simple random sample of size $n$ from a large population with mean $\mu$ and standard deviation $\sigma$.

- $\bar{X}$ is a random variable because every time we take a random sample we will not get the same sample mean $\bar{X}$. Thus we want to know the distribution of the sample means $\bar{X}$.

- The center of the sample means (mean of the sample means) $\mu_{\bar{X}}$ is $\mu$. Also called the expected value.

- The spread of the sample means (standard deviation of the sample means) $\sigma_{\bar{X}}$ is $\sigma/\sqrt{n}$. 
Shape of the Sample Mean Distribution

If a population has a Normal distribution, then the sample mean $\bar{X}$ of $n$ independent observations also has a Normal distribution with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.

Central limit theorem: For any population, when $n$ is large ($n > 30$), the sampling distribution of the sample mean $\bar{X}$ is approximately a Normal distribution with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.
Sampling Distribution

Sampling Distribution for the sample proportion
Sample Proportions

- The population proportion is $p$ a parameter. In some cases we do not know the population proportion, thus we use the sample proportion, $\hat{p}$ to estimate $p$.

- The sample proportion is calculated by: $\hat{p} = \frac{X}{n}$

- $X$ = the number of observations of interest in the sample or the number of "successes" in the sample.

- $n$ = the sample size or number of observations.

- Recall that $X \sim \text{Bin}(n, p)$ and can be approximated by the Normal distribution with $\mu_X = E(X) = np$ and $\sigma_X = SD(X) = \sqrt{np(1 - p)}$ as long as $np \geq 10$ and $np(1 - p) \geq 10$.

- Now we want to know how is $\hat{p} = \frac{X}{n}$ distributed. Thus we want to know $\mu_{\hat{p}} = E(\hat{p})$ and $\sigma_{\hat{p}} = SD(\hat{p})$. 
Example

- According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes.

- A sample of 50 taxpayers was selected. What do we expect the sample proportion \( \hat{p} \) to be?

- If we take other samples will the sample proportions always be the same value?

- If not, what would \( \hat{p} \) be off by?
Shape of the distribution of $\hat{p}$

We can use the **Normal distribution** as long as

- $np \geq 10$ the number of successes are at least 10
- $n(1 - p) \geq 10$ the number of failures are at least 10.
Center of the distribution of $\hat{p}$

- The center is the mean (expected value): $\mu_{\hat{p}} = p$ the proportion of success.

- $\hat{p} = \frac{X}{n}$ where $X$ is the number of successes out of $n$ observations. Thus $X$ has a binomial distribution with parameters $n$ and $p$.

- The mean of $X$ is:
  \[ \mu_X = E(X) = np \]

- Thus the mean of $\hat{p}$ is:
  \[ \mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{\mu_X}{n} = \frac{np}{n} = p \]
Center of the distribution of $\hat{p}$

- The center is the mean (expected value): $\mu_{\hat{p}} = p$ the proportion of success.
- $\hat{p} = \frac{X}{n}$ where $X$ is the number of successes out of $n$ observations. Thus $X$ has a binomial distribution with parameters $n$ and $p$.
- The mean of $X$ is:
  \[ \mu_X = E(X) = np \]
- Thus the mean of $\hat{p}$ is:
  \[ \mu_{\hat{p}} = E(\hat{p}) = E \left( \frac{X}{n} \right) = \frac{\mu_X}{n} = \frac{np}{n} = p \]
Center of the distribution of $\hat{p}$

- The center is the mean (expected value): $\mu_{\hat{p}} = p$ the proportion of success.
- $\hat{p} = \frac{X}{n}$ where $X$ is the number of successes out of $n$ observations. Thus $X$ has a binomial distribution with parameters $n$ and $p$.
- The mean of $X$ is:
  \[ \mu_X = E(X) = np \]

Thus the mean of $\hat{p}$ is:
\[ \mu_{\hat{p}} = E(\hat{p}) = E\left( \frac{X}{n} \right) = \frac{\mu_X}{n} = \frac{np}{n} = p \]
The center is the mean (expected value): \( \mu_{\hat{p}} = p \) the proportion of success.

\( \hat{p} = \frac{X}{n} \) where \( X \) is the number of successes out of \( n \) observations. Thus \( X \) has a binomial distribution with parameters \( n \) and \( p \).

The mean of \( X \) is:

\[ \mu_X = E(X) = np \]

Thus the mean of \( \hat{p} \) is:

\[ \mu_{\hat{p}} = E(\hat{p}) = E \left( \frac{X}{n} \right) = \frac{\mu_X}{n} = \frac{np}{n} = p \]
Spread of the distribution of $\hat{p}$

- The spread is the standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

- The variance of $X$ is:
  \[ \sigma^2_X = \text{Var}(X) = np(1 - p) \]

- The variance of $\hat{p}$ is:
  \[ \sigma^2_{\hat{p}} = \text{Var}(\hat{p}) = \text{Var} \left( \frac{X}{n} \right) = \frac{\text{Var}(X)}{n^2} = \frac{np(1 - p)}{n^2} = \frac{p(1 - p)}{n} \]
Spread of the distribution of $\hat{p}$

- The spread is the standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- The variance of $X$ is:

$$\sigma^2_X = \text{Var}(X) = np(1 - p)$$

- The variance of $\hat{p}$ is:

$$\sigma^2_{\hat{p}} = \text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{\text{Var}(X)}{n^2} = \frac{np(1 - p)}{n^2} = \frac{p(1 - p)}{n}$$
Spread of the distribution of $\hat{p}$

- The spread is the standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- The variance of $X$ is:
  \[ \sigma_X^2 = \text{Var}(X) = np(1 - p) \]

- The variance of $\hat{p}$ is:
  \[ \sigma_{\hat{p}}^2 = \text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{\text{Var}(X)}{n^2} = \frac{np(1 - p)}{n^2} = \frac{p(1 - p)}{n} \]
Sample Distribution of $n = 50$.

Summary Statistics

- Mean of Sample Proportions: 0.34244
- Std Dev of Sample Proportions: 0.06987
- No. of Samples: 1000
Sample Distribution of $n = 125$

Summary Statistics

Mean of Sample Proportions: 0.34082
Std Dev of Sample Proportions: 0.04375
No. of Samples: 1000
Assumptions

- The sampled values must be random and independent of each other. This can be tested by **10% Condition**: The sample size must be no larger than 10% of the population.

- The sample size, $n$ must be large enough. This can be be tested by **Success / Failure Condition**: The sample size has to be big enough so that both $np$ and $n(1 – p)$ at least 10.
Example for distribution of $\hat{p}$

According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the distribution of $\hat{p}$, the sample proportion of the 125 taxpayers that used computer software to do their taxes?

1. Check if we can use the Normal distribution.
   - $p = 0.34$, $n = 125$
   - $np = 125(0.34) = 42.5$
   - $n(1 - p) = 125(1 - 0.34) = 125(0.66) = 82.5$
   - Both $np$ and $n(1 - p)$ are greater than 10 so we can use the Normal distribution.

2. The mean is: $\mu_{\hat{p}} = p = 0.34$. If we take a sample we "expect" 34% to have used computer software to do their taxes.

3. The standard deviation is:
   \[ \sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{0.34(1 - 0.34)}{125}} = 0.0424 \]
According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the probability that between 28% and 40% of the taxpayers from the sample of 125 used computer software to do their taxes?

1. We want: \( P(0.28 < \hat{p} < 0.40) \)
The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

1. What is the shape of the sampling distribution of $\hat{p}$.
Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, \( \hat{p} \) that are "friends" with their children on Facebook.

2. What is the mean of the sampling distribution of \( \hat{p} \).
The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, \( \hat{p} \) that are "friends" with their children on Facebook.

3. What is the standard deviation of the sampling distribution of \( \hat{p} \).
Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

4. What is the probability that the sample proportion of 140 parents is greater than 72%?
Sample Proportions

- The population proportion is $p$ a parameter. In some cases we do not know the population proportion, thus we use the sample proportion, $\hat{p}$ to estimate $p$.

- The sample proportion is calculated by: $\hat{p} = \frac{X}{n}$

- $X =$ the number of observations of interest in the sample or the number of "successes" in the sample.

- $n =$ the sample size or number of observations.

- The mean of $\hat{p}$ is $\mu_{\hat{p}} = E(\hat{p}) = p$.

- The standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$.

- If $np \geq 10$ and $n(1 - p) \geq 10$, then $\hat{p}$ has an approximate Normal distribution.
Voting Questions

Suppose that 52% voted for a certain candidate. We take a random sample of 1500 likely voters. Determine the following probabilities.

1. What is the probability that the sample proportion is less than 0.50?

2. What is the probability that the sample proportion is within 3% of the population proportion?