Point Estimates and Sampling Variability

Colby Community College

Definition

A **point estimate** is a single value used to estimate a parameter.

Definition

A **point estimate** is a single value used to estimate a parameter.

Note

The sample proportion, $\hat{p}$, the best point estimate of the population proportion p. But, is it a *good* estimate?

Definition

A **point estimate** is a single value used to estimate a parameter.

Note

The sample proportion, $\hat{p}$, the best point estimate of the population proportion p. But, is it a *good* estimate?

Definition

The difference between a point estimate and the true parameter is called the **error** in the estimate.

Definition

A **point estimate** is a single value used to estimate a parameter.

Note

The sample proportion, $\hat{p}$, the best point estimate of the population proportion p. But, is it a *good* estimate?

Definition

The difference between a point estimate and the true parameter is called the **error** in the estimate.

Note

In general, there are two sources of error: sampling error and bias.

Definition

Sampling error describes how much an estimate will tend to vary from one sample to the next.

Definition

Sampling error describes how much an estimate will tend to vary from one sample to the next.

Example 1

One sample may have $\hat{p} = 1\%$ and another sample may have $\hat{p} = 3\%$.

Definition

Sampling error describes how much an estimate will tend to vary from one sample to the next.

Example 1

One sample may have $\hat{p} = 1\%$ and another sample may have $\hat{p} = 3\%$.

Note

Much of statistics is focused on understanding and quantifying sampling error.

Definition

Bias describes a systematic tendency to over-estimate or under-estimate the true population value.

Definition

Bias describes a systematic tendency to over-estimate or under-estimate the true population value.

Example 2

If a university took a student poll asking about support for a new stadium, they'd get a biased response if they asked:

"Do you support your school by supporting funding for the new stadium?"

Definition

Bias describes a systematic tendency to over-estimate or under-estimate the true population value.

Example 2

If a university took a student poll asking about support for a new stadium, they'd get a biased response if they asked:

"Do you support your school by supporting funding for the new stadium?"

Note

We try to minimize bias by using thoughtful data collection procedures.

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

If we take a poll of 1000 American adults on this topic, the estimate would not be perfect. But, how close can we expect $\hat{p}$ to be to p?

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

If we take a poll of 1000 American adults on this topic, the estimate would not be perfect. But, how close can we expect $\hat{p}$ to be to p?

We can simulate such a sample:

1. As of 2021, there are about 258 million adults in America. Let us get 258 million slips of paper and write “support” on 88% of them and “not” on the remaining 12%.

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

If we take a poll of 1000 American adults on this topic, the estimate would not be perfect. But, how close can we expect $\hat{p}$ to be to p?

We can simulate such a sample:

1. As of 2021, there are about 258 million adults in America. Let us get 258 million slips of paper and write “support” on 88% of them and “not” on the remaining 12%.
2. Mix up the slips and pull out 1000, to represent our sample.

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

If we take a poll of 1000 American adults on this topic, the estimate would not be perfect. But, how close can we expect $\hat{p}$ to be to p?

We can simulate such a sample:

1. As of 2021, there are about 258 million adults in America. Let us get 258 million slips of paper and write “support” on 88% of them and “not” on the remaining 12%.
2. Mix up the slips and pull out 1000, to represent our sample.
3. Compute the fraction of the sample that says “support”.

Example 3

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$.

If we take a poll of 1000 American adults on this topic, the estimate would not be perfect. But, how close can we expect $\hat{p}$ to be to p?

We can simulate such a sample:

1. As of 2021, there are about 258 million adults in America. Let us get 258 million slips of paper and write “support” on 88% of them and “not” on the remaining 12%.
2. Mix up the slips and pull out 1000, to represent our sample.
3. Compute the fraction of the sample that says “support”.

Note

While this method seems silly, a compute can do these steps in a short amount of time.

Example 4

I wrote a short program to run this simulation, but one simulation isn't enough to get a sense of the distribution of the point estimates.

Example 4

I wrote a short program to run this simulation, but one simulation isn't enough to get a sense of the distribution of the point estimates.

So, I ran nine simulations:

$\hat{p}$	Error	$\hat{p}$	Error	$\hat{p}$	Error
0.867	-0.013	0.876	-0.004	0.883	0.003
0.889	0.009	0.887	0.007	0.874	-0.006
0.896	0.016	0.898	0.018	0.874	-0.006

Notice that they are all kinda close to $p = 0.88$, but there is variation.

Example 4

I wrote a short program to run this simulation, but one simulation isn't enough to get a sense of the distribution of the point estimates.

So, I ran nine simulations:

$\hat{p}$	Error	$\hat{p}$	Error	$\hat{p}$	Error
0.867	-0.013	0.876	-0.004	0.883	0.003
0.889	0.009	0.887	0.007	0.874	-0.006
0.896	0.016	0.898	0.018	0.874	-0.006

Notice that they are all kinda close to $p = 0.88$, but there is variation.

The mean of all these $\hat{p}$ values is 0.8827, which is pretty close to p.

Example 4

I wrote a short program to run this simulation, but one simulation isn't enough to get a sense of the distribution of the point estimates.

So, I ran nine simulations:

$\hat{p}$	Error	$\hat{p}$	Error	$\hat{p}$	Error
0.867	-0.013	0.876	-0.004	0.883	0.003
0.889	0.009	0.887	0.007	0.874	-0.006
0.896	0.016	0.898	0.018	0.874	-0.006

Notice that they are all kinda close to $p = 0.88$, but there is variation.

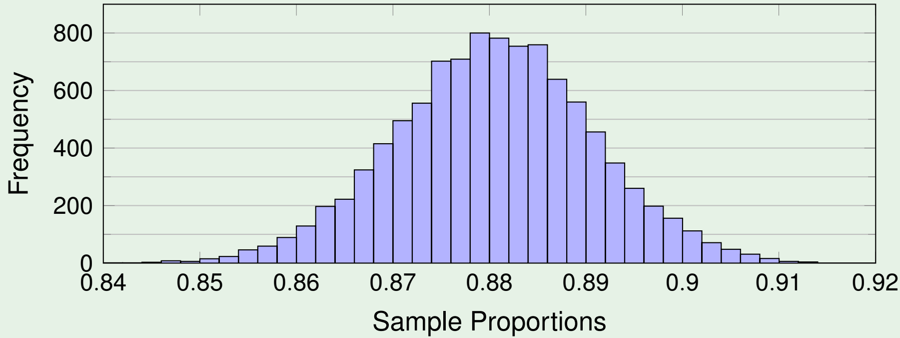
The mean of all these $\hat{p}$ values is 0.8827, which is pretty close to p.

Definition

The **sampling distribution** is the distribution of sample proportions.

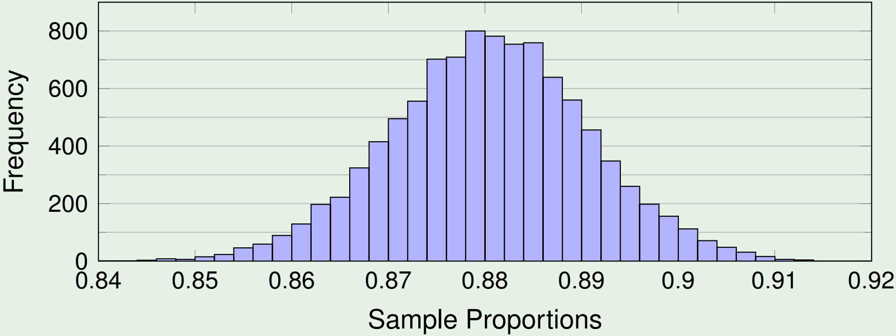
Example 5

Running the simulation 10,000 times gives the histogram:



Example 5

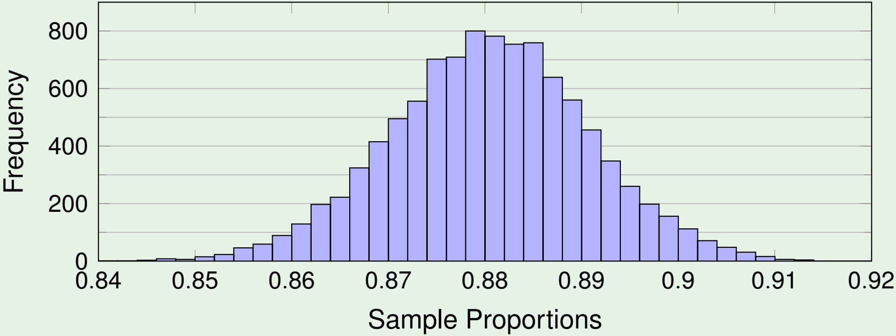
Running the simulation 10,000 times gives the histogram:



Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8799$, which is very close to $p = 0.88$.

Example 5

Running the simulation 10,000 times gives the histogram:

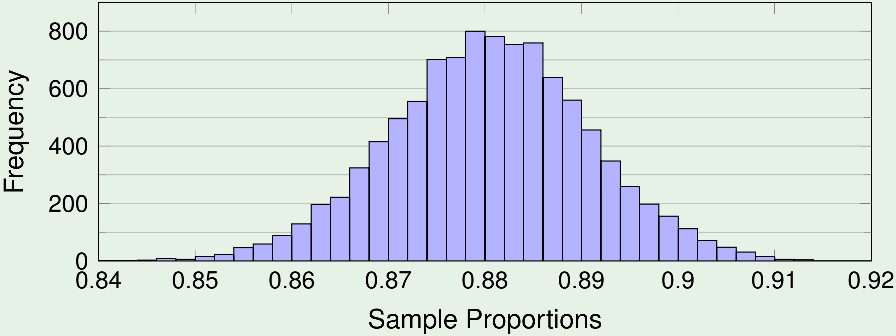


Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8799$, which is very close to $p = 0.88$.

Spread : The standard deviation of this distribution is $s_{\hat{p}} = 0.0102$.

Example 5

Running the simulation 10,000 times gives the histogram:

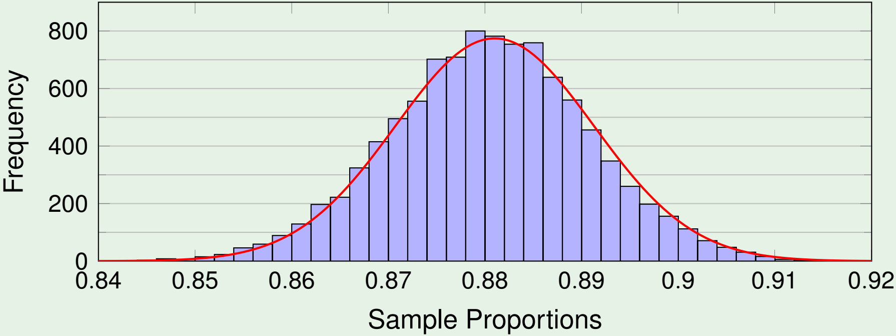


Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8799$, which is very close to $p = 0.88$.

Spread : The standard deviation of this distribution is $s_{\hat{p}} = 0.0102$. This is often called the **standard error**.

Example 5

Running the simulation 10,000 times gives the histogram:



Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8799$, which is very close to $p = 0.88$.

Spread : The standard deviation of this distribution is $s_{\hat{p}} = 0.0102$. This is often called the **standard error**.

Shape : This distribution is approximately normal.

Example 6

What if we used a much smaller sample size of $n = 50$?

Example 6

What if we used a much smaller sample size of $n = 50$?

Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8791$, which is still very close to $p = 0.88$.

Example 6

What if we used a much smaller sample size of $n = 50$?

Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8791$, which is still very close to $p = 0.88$.

Spread : The standard deviation of this distribution is $s_{\hat{p}} = 0.0462$, which is much bigger.

Example 6

What if we used a much smaller sample size of $n = 50$?

Center : The center of this distribution is $\bar{x}_{\hat{p}} = 0.8791$, which is still very close to $p = 0.88$.

Spread : The standard deviation of this distribution is $s_{\hat{p}} = 0.0462$, which is much bigger.

Note

This highlights an important property: a bigger sample tends to provide a more precise point estimate than a smaller sample.

Example 6

What if we used a much smaller sample size of $n = 50$?

- **Center**: The center of this distribution is $\bar{x}_{\hat{p}} = 0.8791$, which is still very close to $p = 0.88$.
- **Spread**: The standard deviation of this distribution is $s_{\hat{p}} = 0.0462$, which is much bigger.

Note

This highlights an important property: a bigger sample tends to provide a more precise point estimate than a smaller sample.

Note

In real-world applications, we never actually observe the sampling distribution, yet it is useful to always think of a point estimate as coming from a hypothetical distribution.

Central Limit Theorem

When observations are independent and the sample size, n, is sufficiently large, the sample proportions $\hat{p}$ will tend to follow a normal distribution with the following mean and standard deviation:

$$\mu_{\hat{p}} = p \qquad \text{and} \qquad \sigma_{\hat{p}} = SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Central Limit Theorem

When observations are independent and the sample size, n, is sufficiently large, the sample proportions $\hat{p}$ will tend to follow a normal distribution with the following mean and standard deviation:

$$\mu_{\hat{p}} = p \qquad \text{and} \qquad \sigma_{\hat{p}} = SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

In order for the Central Limit Theorem to hold, the sample size is typically considered sufficiently large when

$$np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10$$

which is called the **success-failure conditions.**

Central Limit Theorem

When observations are independent and the sample size, n, is sufficiently large, the sample proportions $\hat{p}$ will tend to follow a normal distribution with the following mean and standard deviation:

$$\mu_{\hat{p}} = p \qquad \text{and} \qquad \sigma_{\hat{p}} = SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

In order for the Central Limit Theorem to hold, the sample size is typically considered sufficiently large when

$$np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10$$

which is called the **success-failure conditions**.

Note

The Central Limit Theorem is incredibly important, and provides a foundation for much of statistics.

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np =$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$

$$n(1-p) =$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$\begin{aligned} np &= 1000 \cdot 0.88 = 880 \geq 10 \checkmark \\ n(1-p) &= 1000(1-0.88) \end{aligned}$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$\begin{aligned} np &= 1000 \cdot 0.88 = 880 \geq 10 \checkmark \\ n(1-p) &= 1000(1-0.88) = 1000 \cdot 0.12 \end{aligned}$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} =$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$

$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \quad \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$

$$SE_{\hat{p}} =$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.12)}{1000}}$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.12)}{1000}} = 0.0103$$

Example 7

In Example 3, we had a sample size of $n = 1000$ and $p = 0.88$.

Before we can apply the Central Limit Theorem, we need to check the success-failure conditions:

$$np = 1000 \cdot 0.88 = 880 \geq 10 \checkmark$$
$$n(1-p) = 1000(1-0.88) = 1000 \cdot 0.12 = 120 \geq \ \checkmark$$

Applying the Central Limit Theorem gives:

$$\mu_{\hat{p}} = p = 0.88$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.12)}{1000}} = 0.0103$$

This is very close to the observed standard error, 0.0102.

How to Verify Sample Observations are Independent

- Subjects in an experiment are considered independent if they undergo random assignment to the treatment groups.

How to Verify Sample Observations are Independent

- Subjects in an experiment are considered independent if they undergo random assignment to the treatment groups.
- If the observations are from a simple random sample, then they are independent.

How to Verify Sample Observations are Independent

- Subjects in an experiment are considered independent if they undergo random assignment to the treatment groups.
- If the observations are from a simple random sample, then they are independent.
- If a sample is from a seemingly random process, e.g. an occasional error on an assembly line, checking independence is more difficult. In this case, use your best judgment.

How to Verify Sample Observations are Independent

- Subjects in an experiment are considered independent if they undergo random assignment to the treatment groups.
- If the observations are from a simple random sample, then they are independent.
- If a sample is from a seemingly random process, e.g. an occasional error on an assembly line, checking independence is more difficult. In this case, use your best judgment.

Note

If a sample is larger than 10% of the population, the methods we discuss tend to overestimate the sampling error slightly. In these cases more advanced methods are needed.

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} =$$

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103}$$

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942$$

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} =$$

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} = \frac{0.90 - 0.88}{0.0103}$$

Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} = \frac{0.90 - 0.88}{0.0103} = 1.942$$

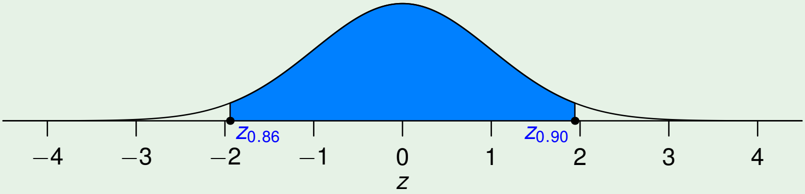
Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} = \frac{0.90 - 0.88}{0.0103} = 1.942$$



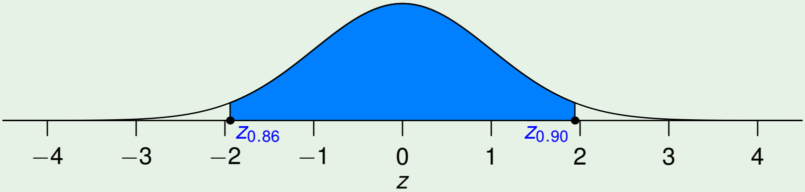
Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} = \frac{0.90 - 0.88}{0.0103} = 1.942$$



Using technology gives:

$$P(-1.94175 \leq z \leq 1.94175) = 0.947833$$

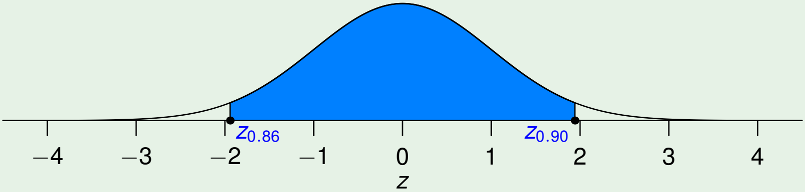
Example 8

Using $n = 1000$ and $p = 0.88$ from Example 3, let us find out how often $\hat{p}$ is within 0.02 (2%) of the population value $p = 0.88$.

In Example 7, we applied the Central Limit Theorem, getting $\mu_{\hat{p}} = 0.88$ and $SE_{\hat{p}} = 0.0103$.

We start by calculating the z-values:

$$z_{0.86} = \frac{0.86 - 0.88}{0.0103} = -1.942 \quad \text{and} \quad z_{0.90} = \frac{0.90 - 0.88}{0.0103} = 1.942$$



Using technology gives:

$$P(-1.94175 \leq z \leq 1.94175) = 0.947833$$

We expect $\hat{p}$ to be within 0.02 of 0.88 about 94.78% of the time.

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} =$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

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

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

$$n(1 - \hat{p}) = 1000(1 - 0.887)$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

$$n(1-\hat{p}) = 1000(1-0.887) = 1000 \cdot 0.113$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

$$n(1-\hat{p}) = 1000(1-0.887) = 1000 \cdot 0.113 = 113$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

> "Does the sample proportion from the poll approximately follow a normal distribution?"

Independence: Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions: Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

$$n(1-\hat{p}) = 1000(1-0.887) = 1000 \cdot 0.113 = 113 \geq 10 \checkmark$$

Example 9

We do not actually know the population proportion unless we conduct a full census of the entire population.

The value $p = 0.88$ was based on a Pew Research poll of 1000 adults that found $\hat{p} = 0.887$ of them favored expanding solar energy.

A question the researchers might have asked is:

"Does the sample proportion from the poll approximately follow a normal distribution?"

Independence Pew Research is a well known non-profit think tank, so we can believe that the poll is a simple random sample, and hence the observations are independent.

Success-Failure Conditions Since we don't actually know p, the next best thing we have is $\hat{p}$.

$$n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10 \checkmark$$

$$n(1-\hat{p}) = 1000(1-0.887) = 1000 \cdot 0.113 = 113 \geq 10 \checkmark$$

Because $n\hat{p}$ and $n(1-\hat{p})$ are both well above 10, we can conclude that $\hat{p}$ is a reasonable substitute for p.

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} =$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

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

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}}$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$

$$SE_{\hat{p}} \approx$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$
$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$

$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.887(1-0.887)}{1000}}$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$
$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$
$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.887(1-0.887)}{1000}} = 0.010012$$

Substitution Approximation

When np and $n(1-p)$ are much larger than 10, we can use $\hat{p}$ in place of p and the Centeral Limit Theorem becomes:

$$\mu_{\hat{p}} = p \approx \hat{p}$$

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 10

For $n = 1000$, $p - 0.88$, $\hat{p} = 0.887$:

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.88(1-0.88)}{1000}} = 0.010276$$

$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.887(1-0.887)}{1000}} = 0.010012$$

These values are the same to three decimal places, so using the substitution approximation won't make a major difference.

Trends in the Sampling Distribution

- The smaller either np or $n(1-p)$ is, the more discrete the sampling distribution is.

Trends in the Sampling Distribution

- The smaller either np or $n(1-p)$ is, the more discrete the sampling distribution is.
- When np or $n(1-p)$ is smaller than 10, the distribution is skewed.

Trends in the Sampling Distribution

- The smaller either np or $n(1-p)$ is, the more discrete the sampling distribution is.
- When np or $n(1-p)$ is smaller than 10, the distribution is skewed.
- The larger both np and $n(1-p)$ are, the more normal the sampling distribution is.

Trends in the Sampling Distribution

- The smaller either np or $n(1-p)$ is, the more discrete the sampling distribution is.
- When np or $n(1-p)$ is smaller than 10, the distribution is skewed.
- The larger both np and $n(1-p)$ are, the more normal the sampling distribution is.
    - This may be harder to see for larger sample sizes, as the variability also becomes smaller.

Trends in the Sampling Distribution

- The smaller either np or $n(1-p)$ is, the more discrete the sampling distribution is.
- When np or $n(1-p)$ is smaller than 10, the distribution is skewed.
- The larger both np and $n(1-p)$ are, the more normal the sampling distribution is.
    - This may be harder to see for larger sample sizes, as the variability also becomes smaller.
- When np and $n(1-p)$ are both very large, the distributions discreteness is hardly evident, and the distribution looks much more like a normal distribution.