Normal approximation, standard error and sampling distribution
The empirical rule

If the data follow the normal curve, then

- About 2/3 (68%) of the data fall within one standard deviation of the mean.
- About 95% fall within 2 standard deviations of the mean.
- About 99.7% fall within 3 standard deviations of the mean.

Galton's measurements of heights of fathers have $\bar{x} = 68.3$ in and $s = 1.8$ in. Therefore about 95% of all heights are between $68.3 - 2 \times 1.8 = 64.7$ in and $68.3 + 2 \times 1.8 = 71.9$ in.
The empirical rule

If the data follow the normal curve, then

- about 2/3 (68%) of the data fall within one standard deviation of the mean
- about 95% fall within 2 standard deviations of the mean
- about 99.7% fall within 3 standard deviations of the mean
The empirical rule

If the data follow the normal curve, then

- about 2/3 (68%) of the data fall within one standard deviation of the mean
- about 95% fall within 2 standard deviations of the mean

Galton’s measurements of heights of fathers have $\bar{x} = 68.3$ in and $s = 1.8$ in. Therefore about 95% of all heights are between $68.3 - 2 \times 1.8 = 64.7$ in and $68.3 + 2 \times 1.8 = 71.9$ in.
The empirical rule

If the data follow the normal curve, then

- about 2/3 (68%) of the data fall within one standard deviation of the mean
- about 95% fall within 2 standard deviations of the mean
- about 99.7% fall within 3 standard deviations of the mean
The empirical rule

If the data follow the normal curve, then

- about 2/3 (68%) of the data fall within one standard deviation of the mean
- about 95% fall within 2 standard deviations of the mean
- about 99.7% fall within 3 standard deviations of the mean

Galton’s measurements of heights of fathers have $\bar{x} = 68.3$ in and $s = 1.8$ in. Therefore about 95% of all heights are between $68.3 \text{ in} - 2 \times 1.8 \text{ in} = 64.7 \text{ in}$ and $68.3 \text{ in} + 2 \times 1.8 \text{ in} = 71.9 \text{ in}$. 
Standardizing data

A normal curve is determined by $\bar{x}$ and $s$: If the data follow the normal curve, then knowing $\bar{x}$ and $s$ means knowing the whole histogram.
Standardizing data

A normal curve is determined by $\bar{x}$ and $s$: If the data follow the normal curve, then knowing $\bar{x}$ and $s$ means knowing the whole histogram.

To compute areas under the normal curve, we first standardize the data by subtracting off $\bar{x}$ and then dividing by $s$:

$$z = \frac{\text{height} - \bar{x}}{s}$$
Standardizing data

A normal curve is determined by $\bar{x}$ and $s$: If the data follow the normal curve, then knowing $\bar{x}$ and $s$ means knowing the whole histogram.

To compute areas under the normal curve, we first standardize the data by subtracting off $\bar{x}$ and then dividing by $s$:

$$z = \frac{\text{height} - \bar{x}}{s}$$

$z$ is called the standardized value or z-score.
Standardizing data

A normal curve is determined by \( \bar{x} \) and \( s \): If the data follow the normal curve, then knowing \( \bar{x} \) and \( s \) means knowing the whole histogram.

To compute areas under the normal curve, we first **standardize** the data by subtracting off \( \bar{x} \) and then dividing by \( s \):

\[
z = \frac{\text{height} - \bar{x}}{s}
\]

\( z \) is called the **standardized value** or **z-score**.

\( z \) has no unit (height, \( \bar{x} \) and \( s \) all have the unit ‘inches’).
Standardizing data

A normal curve is determined by $\bar{x}$ and $s$: If the data follow the normal curve, then knowing $\bar{x}$ and $s$ means knowing the whole histogram.

To compute areas under the normal curve, we first **standardize** the data by subtracting off $\bar{x}$ and then dividing by $s$: 

$$z = \frac{\text{height} - \bar{x}}{s}$$

$z$ is called the **standardized value** or z-score.

$z$ has no unit (height, $\bar{x}$ and $s$ all have the unit ‘inches’)

For example, $z = 2$ means the height is 2 standard deviations above average.
Standardizing data

A normal curve is determined by $\bar{x}$ and $s$: If the data follow the normal curve, then knowing $\bar{x}$ and $s$ means knowing the whole histogram.

To compute areas under the normal curve, we first standardize the data by subtracting off $\bar{x}$ and then dividing by $s$:

$$z = \frac{\text{height} - \bar{x}}{s}$$

$z$ is called the standardized value or z-score.

$z$ has no unit (height, $\bar{x}$ and $s$ all have the unit ‘inches’)

For example, $z = 2$ means the height is 2 standard deviations above average.

$z = -1.5$ means the height is 1.5 standard deviations below average.
The standard normal curve

Standardized data have mean 0 and standard deviation equal to 1 – this is the point of standardizing.
The standard normal curve

Standardized data have mean 0 and standard deviation equal to 1 – this is the point of standardizing.

Fathers’ heights follow the normal curve with $\bar{x} = 68.3$ in and $s = 1.8$ in. Therefore the standardized values follow the **standard normal curve** with mean 0 and standard deviation 1.
The standard normal curve

Standardized data have mean 0 and standard deviation equal to 1 – this is the point of standardizing.

Fathers’ heights follow the normal curve with $\bar{x} = 68.3$ in and $s = 1.8$ in. Therefore the standardized values follow the **standard normal curve** with mean 0 and standard deviation 1.

This curve is given by the function

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
Normal approximation

Finding areas under the normal curve is called normal approximation.
Normal approximation

Finding areas under the normal curve is called **normal approximation**.

What percentage of fathers have heights between 67.4 in and 71.9 in?
Normal approximation

Finding areas under the normal curve is called normal approximation.

What percentage of fathers have heights between 67.4 in and 71.9 in?

1. Standardize:
Normal approximation

Finding areas under the normal curve is called normal approximation.
What percentage of fathers have heights between 67.4 in and 71.9 in?

1. Standardize:

\[
\frac{67.4 \text{ in} - 68.3 \text{ in}}{1.8 \text{ in}} = -0.5
\]
Normal approximation

Finding areas under the normal curve is called \textbf{normal approximation}.

What percentage of fathers have heights between 67.4 in and 71.9 in?

1. **Standardize:**

\[
\frac{67.4 \text{ in} - 68.3 \text{ in}}{1.8 \text{ in}} = -0.5 \quad \frac{71.9 \text{ in} - 68.3 \text{ in}}{1.8 \text{ in}} = 2
\]
Finding areas under the normal curve is called **normal approximation**. What percentage of fathers have heights between 67.4 in and 71.9 in?

1. **Standardize:**

   \[
   \frac{67.4 \text{ in} - 68.3 \text{ in}}{1.8 \text{ in}} = -0.5 \quad \frac{71.9 \text{ in} - 68.3 \text{ in}}{1.8 \text{ in}} = 2
   \]

2. **Mark the area under the normal curve:**
Normal approximation

3. Write the desired area in a form that can be computed by software or looked up in a table:

\[
> \text{pnorm}(2)-\text{pnorm}(-0.5) \\
[1] 0.6687123
\]

This can also be computed directly, without standardizing:

\[
> \text{pnorm}(71.9,68.3,1.8)-\text{pnorm}(67.4,68.3,1.8) \\
[1] 0.6687123
\]
Normal approximation

3. Write the desired area in a form that can be computed by software or looked up in a table:

   Typically we can look up the area to the left of a given value.
Normal approximation

3. Write the desired area in a form that can be computed by software or looked up in a table:
   Typically we can look up the area to the left of a given value.

4. Use software or a table to find these values:
3. Write the desired area in a form that can be computed by software or looked up in a table:
   Typically we can look up the area to the left of a given value.

4. Use software or a table to find these values:
   > pnorm(2) - pnorm(-0.5)
   [1] 0.6687123
Normal approximation

3. Write the desired area in a form that can be computed by software or looked up in a table:
   Typically we can look up the area to the left of a given value.

4. Use software or a table to find these values:
   \[ \text{\textgreater{} pnorm(2)-pnorm(-0.5)} \]
   [1] 0.6687123

   This can also be computed directly, without standardizing:
   \[ \text{\textgreater{} pnorm(71.9,68.3,1.8)-pnorm(67.4,68.3,1.8)} \]
   [1] 0.6687123
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

\[
qnorm(0.3, 68.3, 1.8)
\]

\[
\begin{align*}
\text{Reasoning with a standard normal curve is as follows:} \\
\text{From software or from a normal table:} \\
z &= -0.52 \\
\text{Recall } z &= \frac{\text{height} - \mu}{\sigma} \\
\text{Solve for height} &= \mu + zs \\
\text{Or:} \\
z &= -0.52 \\
\text{means that the height is 0.52 standard deviations below average.} \\
\text{So the height is} &= \mu - 0.52s \\
\text{in} &= 68.3 - (0.52)(1.8) = 67.4\text{ in.}
\end{align*}
\]
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

> qnorm(0.3,68.3,1.8)
[1] 67.35608
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?
> qnorm(0.3,68.3,1.8)
[1] 67.35608

Reasoning with a standard normal curve is as follows:
From software or from a normal table: $z = -0.52$
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

> qnorm(0.3, 68.3, 1.8)

[1] 67.35608

Reasoning with a standard normal curve is as follows:

From software or from a normal table: \( z = -0.52 \)

Recall \( z = \frac{\text{height} - \bar{x}}{s} \).
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

> qnorm(0.3,68.3,1.8)
[1] 67.35608

Reasoning with a standard normal curve is as follows:
From software or from a normal table: \( z = -0.52 \)

Recall \( z = \frac{\text{height} - \bar{x}}{s} \). Solve for height = \( \bar{x} + zs \)
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

> qnorm(0.3,68.3,1.8)
[1] 67.35608

Reasoning with a standard normal curve is as follows:

From software or from a normal table: \( z = -0.52 \)

Recall \( z = \frac{\text{height} - \bar{x}}{s} \). Solve for height = \( \bar{x} + zs \)

Or: \( z = -0.52 \) means that the height is 0.52 standard deviations below average.
Computing percentiles for normal data

What is the 30th percentile of the fathers’ heights?

> qnorm(0.3,68.3,1.8)

[1] 67.35608

Reasoning with a standard normal curve is as follows:

From software or from a normal table: \( z = -0.52 \)

Recall \( z = \frac{\text{height} - \bar{x}}{s} \). Solve for height = \( \bar{x} + zs \)

Or: \( z = -0.52 \) means that the height is 0.52 standard deviations below average. So the height is \( \bar{x} - 0.52s = 68.3 \text{ in} - (0.52)(1.8 \text{ in}) = 67.4 \text{ in.} \)
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$. 

Outliers:

How far off from $\mu$ will $\bar{x}_n$ be?
The standard error (SE) of a statistic tells roughly how far off the statistic will be from its expected value.
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The expected value of one random draw is the population average $\mu$. 
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The expected value of one random draw is the population average $\mu$.

How about $\bar{x}_n$, the average of $n$ draws?
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The **expected value** of one random draw is the population average $\mu$.

How about $\bar{x}_n$, the average of $n$ draws?

The **expected value of the sample average**, $E(\bar{x}_n)$, is the population average $\mu$. 

But remember that $\bar{x}_n$ is a random variable because sampling is a random process. So $\bar{x}_n$ won’t be exactly equal to $\mu = 69.3$ in.: We might get, say, $\bar{x}_n = 70$. Taking another sample of size $n$ might result in $\bar{x}_n = 69.1$ in.
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The expected value of one random draw is the population average $\mu$.

How about $\bar{x}_n$, the average of $n$ draws?

The expected value of the sample average, $E(\bar{x}_n)$, is the population average $\mu$.

But remember that $\bar{x}_n$ is a random variable because sampling is a random process.
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The expected value of one random draw is the population average $\mu$.

How about $\bar{x}_n$, the average of $n$ draws?

The expected value of the sample average, $E(\bar{x}_n)$, is the population average $\mu$.

But remember that $\bar{x}_n$ is a random variable because sampling is a random process. So $\bar{x}_n$ won’t be exactly equal to $\mu = 69.3$ in: We might get, say, $\bar{x}_n = 70.1$ in. Taking another sample of size $n$ might result in $\bar{x}_n = 69.1$ in.
The expected value

If we sample an adult male at random, then we expect his height to be around the population average \( \mu \), give or take about one standard deviation \( \sigma \).

The expected value of one random draw is the population average \( \mu \).

How about \( \bar{x}_n \), the average of \( n \) draws?

The expected value of the sample average, \( E(\bar{x}_n) \), is the population average \( \mu \).

But remember that \( \bar{x}_n \) is a random variable because sampling is a random process. So \( \bar{x}_n \) won’t be exactly equal to \( \mu = 69.3 \) in: We might get, say, \( \bar{x}_n = 70.1 \) in. Taking another sample of size \( n \) might result in \( \bar{x}_n = 69.1 \) in.

How far off from \( \mu \) will \( \bar{x}_n \) be?
The expected value

If we sample an adult male at random, then we expect his height to be around the population average $\mu$, give or take about one standard deviation $\sigma$.

The **expected value** of one random draw is the population average $\mu$.

How about $\bar{x}_n$, the average of $n$ draws?

The **expected value of the sample average**, $E(\bar{x}_n)$, is the population average $\mu$.

But remember that $\bar{x}_n$ is a random variable because sampling is a random process. So $\bar{x}_n$ won’t be exactly equal to $\mu = 69.3$ in: We might get, say, $\bar{x}_n = 70.1$ in. Taking another sample of size $n$ might result in $\bar{x}_n = 69.1$ in.

How far off from $\mu$ will $\bar{x}_n$ be?

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.

So the SE for a statistic plays the same role that the standard deviation $\sigma$ plays for one observation drawn at random.
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.

So the SE for a statistic plays the same role that the standard deviation $\sigma$ plays for one observation drawn at random.

The **square root law** is key for statistical inference:

$$SE\left(\bar{x}_n\right) = \frac{\sigma}{\sqrt{n}}$$

▶ It shows that the SE becomes smaller if we use a larger sample size $n$.

▶ The formula for the standard error does not depend on the size of the population, only on the size of the sample.
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.

So the SE for a statistic plays the same role that the standard deviation $\sigma$ plays for one observation drawn at random.

The **square root law** is key for statistical inference:

$$SE(\bar{x}_n) = \frac{\sigma}{\sqrt{n}}$$

The importance of the square root law is twofold:

- It shows that the SE becomes smaller if we use a larger sample size $n$. 
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.

So the SE for a statistic plays the same role that the standard deviation $\sigma$ plays for one observation drawn at random.

The **square root law** is key for statistical inference:

$$\text{SE} (\bar{x}_n) = \frac{\sigma}{\sqrt{n}}$$

The importance of the square root law is twofold:

- It shows that the SE becomes smaller if we use a larger sample size $n$. We can use the formula to determine what sample size is required for a desired accuracy.
The standard error for the sample average

The **standard error (SE)** of a statistic tells roughly how far off the statistic will be from its expected value.

So the SE for a statistic plays the same role that the standard deviation $\sigma$ plays for one observation drawn at random.

The **square root law** is key for statistical inference:

$$\text{SE} \left( \bar{x}_n \right) = \frac{\sigma}{\sqrt{n}}$$

The importance of the square root law is twofold:

- It shows that the SE becomes smaller if we use a larger sample size $n$. We can use the formula to determine what sample size is required for a desired accuracy.
- The formula for the standard error **does not depend on the size of the population**, only on the size of the sample.
Expected value and standard error for the sum

What if we are interested in the sum of the \( n \) draws, \( S_n \), rather than the average \( \bar{x}_n \)?
What if we are interested in the sum of the $n$ draws, $S_n$, rather than the average $\bar{x}_n$? The sum and the average are related by $S_n = n\bar{x}_n$. 

Expected value and standard error for the sum
What if we are interested in the sum of the $n$ draws, $S_n$, rather than the average $\bar{x}_n$?

The sum and the average are related by $S_n = n\bar{x}_n$.

Both the expected value and the standard error can likewise be obtained by multiplying by $n$, therefore

$$\mathbb{E}(S_n) = n\mu \quad \text{SE}(S_n) = \sqrt{n}\sigma$$
What if we are interested in the sum of the $n$ draws, $S_n$, rather than the average $\bar{x}_n$?

The sum and the average are related by $S_n = n\bar{x}_n$.

Both the expected value and the standard error can likewise be obtained by multiplying by $n$, therefore

$$E(S_n) = n\mu \quad \text{SE}(S_n) = \sqrt{n}\sigma$$

So the variability of the sum of $n$ draws increases at the rate $\sqrt{n}$. 

Expected value and standard error for the sum
Expected value and standard error for percentages

What percentage of likely voters approve of the way the US President is handling his job?
What percentage of likely voters approve of the way the US President is handling his job?
The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:
What percentage of likely voters approve of the way the US President is handling his job?

The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:

- The population consists of all likely voters (about 140 million).
What percentage of likely voters approve of the way the US President is handling his job?

The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:

- The population consists of all likely voters (about 140 million).
- Each likely voter falls into one of two categories: approve or not approve.
Expected value and standard error for percentages

What percentage of likely voters approve of the way the US President is handling his job?

The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:

- The population consists of all likely voters (about 140 million).
- Each likely voter falls into one of two categories: approve or not approve.
- Put the label ‘1’ on each likely voter who approves, and ‘0’ on each who doesn’t.
Expected value and standard error for percentages

What percentage of likely voters approve of the way the US President is handling his job?

The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:

- The population consists of all likely voters (about 140 million).
- Each likely voter falls into one of two categories: approve or not approve.
- Put the label ‘1’ on each likely voter who approves, and ‘0’ on each who doesn’t.
- Then the number of likely voters who approve equals the sum of all 140 million labels.
What percentage of likely voters approve of the way the US President is handling his job?

The ‘percentage of likely voters’ is an average. This becomes clear by using the framework for counting and classifying:

- The population consists of all likely voters (about 140 million).
- Each likely voter falls into one of two categories: approve or not approve.
- Put the label ‘1’ on each likely voter who approves, and ‘0’ on each who doesn’t.
- Then the number of likely voters who approve equals the sum of all 140 million labels.
- The percentage of likely voters who approve is the percentage of 1s among the labels.
Expected value and standard error for percentages

In a sample of \( n \) likely voters

- the number of voters in the sample who are approving is the sum \( S_n \) of the draws
Expected value and standard error for percentages

In a sample of $n$ likely voters

- the number of voters in the sample who are approving is the sum $S_n$ of the draws
- the percentage of voters approving is the percentage of 1s, which is $\frac{S_n}{n} \times 100\% = \bar{x}_n \times 100\%$

$E(\text{percentage of 1s}) = \mu \times 100\%$

$SE(\text{percentage of 1s}) = \sigma \sqrt{\frac{1}{n}} \times 100\%$

where $\mu$ is the population average (= proportion of 1s) and $\sigma$ is the standard deviation of the population of 0s and 1s.

All of the above formulas are for sampling with replacement. They are still approximately true when sampling without replacement if the sample size is much smaller than the size of the population.
Expected value and standard error for percentages

In a sample of \( n \) likely voters

- the number of voters in the sample who are approving is the sum \( S_n \) of the draws
- the percentage of voters approving is the percentage of 1s, which is

\[
\frac{S_n}{n} \times 100\% = \bar{x}_n \times 100\%
\]

Therefore

\[
E(\text{percentage of 1s}) = \mu \times 100\%
\]

\[
\text{SE(percentage of 1s)} = \frac{\sigma}{\sqrt{n}} \times 100\%
\]

where \( \mu \) is the population average (\( = \)proportion of 1s) and \( \sigma \) is the standard deviation of the population of 0s and 1s.
Expected value and standard error for percentages

In a sample of \( n \) likely voters

- the number of voters in the sample who are approving is the sum \( S_n \) of the draws
- the percentage of voters approving is the percentage of 1s, which is
\[
\frac{S_n}{n} \times 100\% = \bar{x}_n \times 100\%
\]

Therefore

\[
\text{E(percentage of 1s)} = \mu \times 100\% \quad \text{SE(percentage of 1s)} = \frac{\sigma}{\sqrt{n}} \times 100\%
\]

where \( \mu \) is the population average (\( \equiv \) proportion of 1s) and \( \sigma \) is the standard deviation of the population of 0s and 1s.

All of the above formulas are for sampling with replacement. They are still approximately true when sampling without replacement if the sample size is much smaller than the size of the population.
Expected value and standard error when simulating

All of the above formulas are also true when the data are simulated, i.e. generated according to a probability histogram.
Expected value and standard error when simulating

All of the above formulas are also true when the data are simulated, i.e. generated according to a probability histogram. What are \( \mu \) and \( \sigma \) in that case?
Expected value and standard error when simulating

All of the above formulas are also true when the data are simulated, i.e. generated according to a probability histogram.
What are \( \mu \) and \( \sigma \) in that case?

If the random variable \( X \) that is simulated has \( K \) possible outcomes \( x_1, \ldots, x_K \), then

\[
\mu = \sum_{i=1}^{K} x_i P(X = x_i) \quad \sigma^2 = \sum_{i=1}^{K} (x_i - \mu)^2 P(X = x_i)
\]
Expected value and standard error when simulating

All of the above formulas are also true when the data are simulated, i.e. generated according to a probability histogram. What are \( \mu \) and \( \sigma \) in that case?

If the random variable \( X \) that is simulated has \( K \) possible outcomes \( x_1, \ldots, x_K \), then

\[
\mu = \sum_{i=1}^{K} x_i P(X = x_i) \quad \sigma^2 = \sum_{i=1}^{K} (x_i - \mu)^2 P(X = x_i)
\]

If the random variable \( X \) has a density \( f \), such as when \( X \) follows the normal curve, then

\[
\mu = \int_{-\infty}^{\infty} x f(x) \, dx \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx
\]
Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, \ldots, 100.
The sampling distribution

Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, \ldots, 100.

How likely is each outcome?
Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, ..., 100.
How likely is each outcome?

The number of tails has the binomial distribution with \( n = 100 \) and \( p = 0.5 \).
The sampling distribution

Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, \ldots, 100.

How likely is each outcome?

The number of tails has the binomial distribution with $n = 100$ and $p = 0.5$.

So if the statistic of interest is $S_n = \text{`number of tails'}, then the probability histogram of $S_n$ is given by the binomial distribution.
The sampling distribution

Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, \ldots, 100.

How likely is each outcome?

The number of tails has the binomial distribution with \( n = 100 \) and \( p = 0.5 \).

So if the statistic of interest is \( S_n = \text{number of tails} \), then the probability histogram of \( S_n \) is given by the binomial distribution. This is called the **sampling distribution** of the statistic \( S_n \).
The sampling distribution

Toss a coin 100 times. The number of tails has the following possible outcomes: 0, 1, 2, ..., 100.

How likely is each outcome?

The number of tails has the binomial distribution with \( n = 100 \) and \( p = 0.5 \).

So if the statistic of interest is \( S_n = \) ‘number of tails’, then the probability histogram of \( S_n \) is given by the binomial distribution. This is called the sampling distribution of the statistic \( S_n \).

The sampling distribution of \( S_n \) provides more detailed information about the chance properties of \( S_n \) than the summary numbers given by the expected value and the standard error.
There are three histograms

The chance process of tossing a coin 100 times comes with three different histograms:
There are three histograms

The chance process of tossing a coin 100 times comes with three different histograms:

1. The probability histogram for producing the data:
There are three histograms

The chance process of tossing a coin 100 times comes with three different histograms:

1. The probability histogram for producing the data:

2. The histogram of the 100 observed tosses. This is an empirical histogram of real data:
There are three histograms

3. The probability histogram of the statistic $S_{100} = \text{‘number of tails’}$, which shows the sampling distribution of $S_{100}$:
There are three histograms

3. The probability histogram of the statistic $S_{100} = \text{‘number of tails’}$, which shows the sampling distribution of $S_{100}$:

When doing statistical inference it is important to carefully distinguish these three histograms.