The Normal Distribution and CLT

Nate Wells

Math 141, 4/4/22

Outline

In this lecture, we will. . .

Outline

In this lecture, we will...

- Investigate properties of the Normal Distribution
- Discuss the Central Limit Theorem and its role in statistics

Section 1

The Normal Distribution

The Normal Distribution

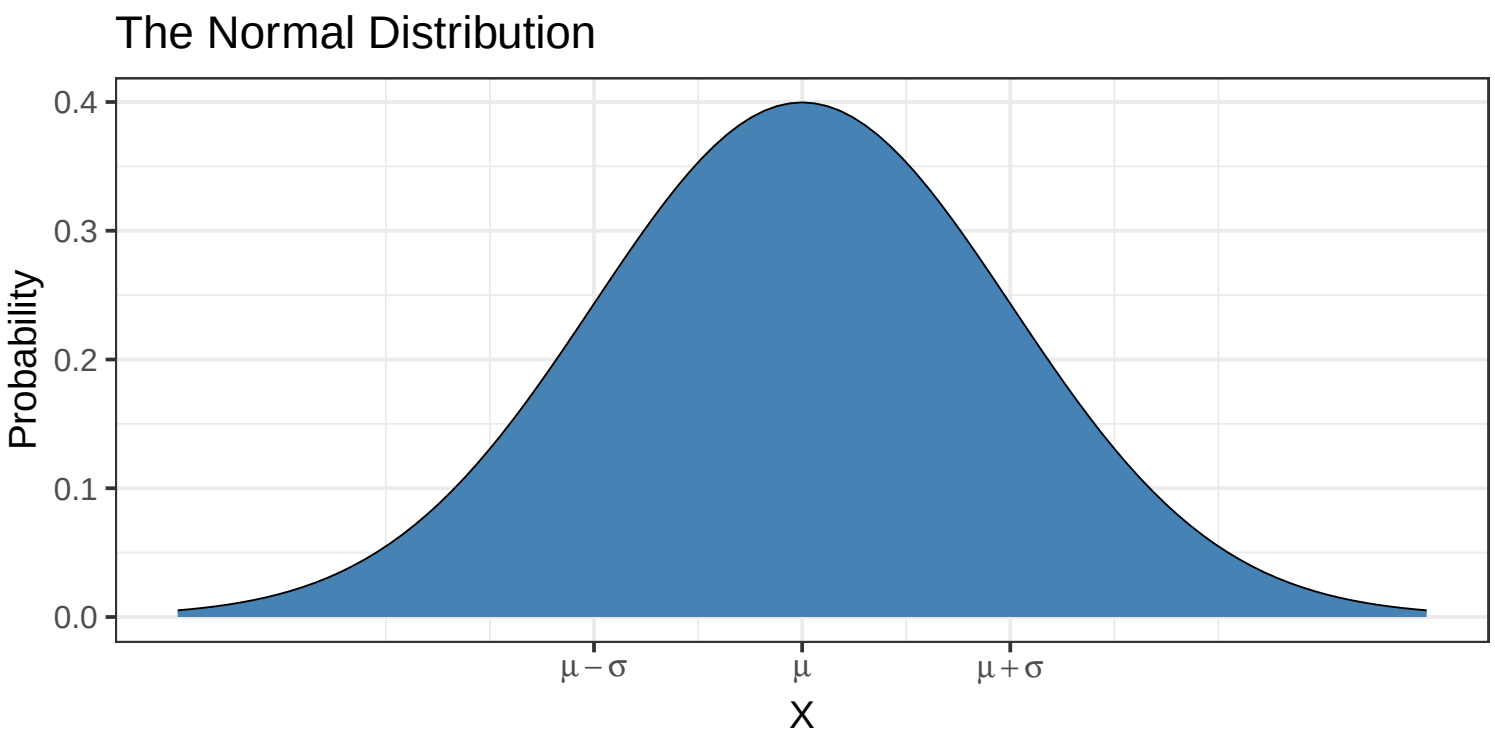
- The general Normal density curve with mean μ and standard deviation σ is given by the formula

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma} \qquad \text{Don't memorize this}$$

The Normal Distribution

- The general Normal density curve with mean μ and standard deviation σ is given by the formula

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma} \qquad \text{Don't memorize this}$$



Normal Probabilities

Recall that for a random variable which has a **continuous** distribution, we find probabilities by looking at areas under the density curve.

Normal Probabilities

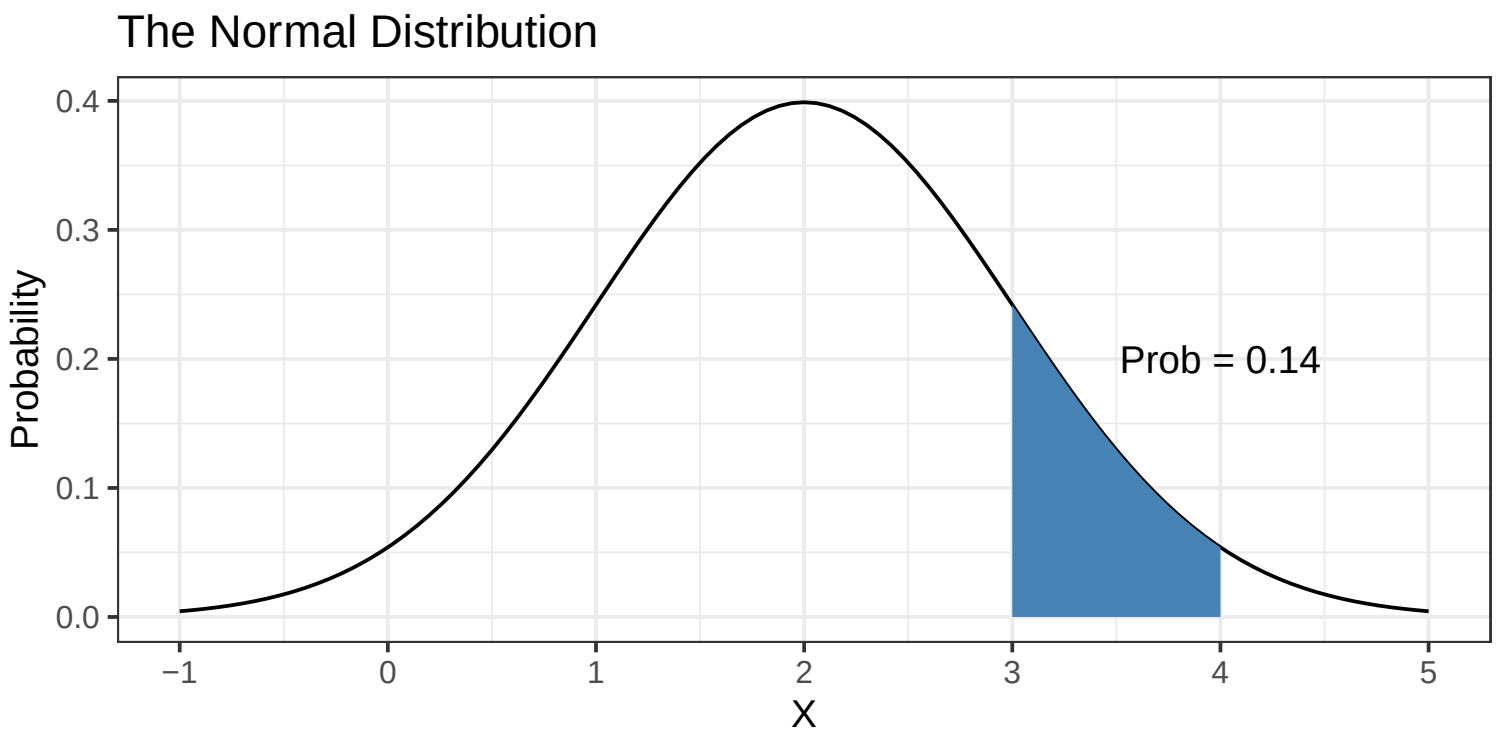
Recall that for a random variable which has a **continuous** distribution, we find probabilities by looking at areas under the density curve.

Suppose X is Normally distributed with mean 2 and standard deviation 1. What is the probability that X is between 3 and 4?

Normal Probabilities

Recall that for a random variable which has a **continuous** distribution, we find probabilities by looking at areas under the density curve.

Suppose X is Normally distributed with mean 2 and standard deviation 1. What is the probability that X is between 3 and 4?



Calculating Normal Probabilities in R

How do we actually find areas under the Normal density curve?

Calculating Normal Probabilities in R

How do we actually find areas under the Normal density curve?

- R has a built-in function for computing cummulative probabilites under Normal densities: `pnorm(q =... , mean =... , sd =... )`

Calculating Normal Probabilities in R

How do we actually find areas under the Normal density curve?

- R has a built-in function for computing cummulative probabilites under Normal densities: `pnorm(q =... , mean =... , sd =... )`
- For example, the following code computes the area left of 1.5 in the Normal distribution with mean 0 and standard deviation 1:

```
pnorm(q =1.5 , mean =0 , sd =1 )
```

```
## [1] 0.9331928
```

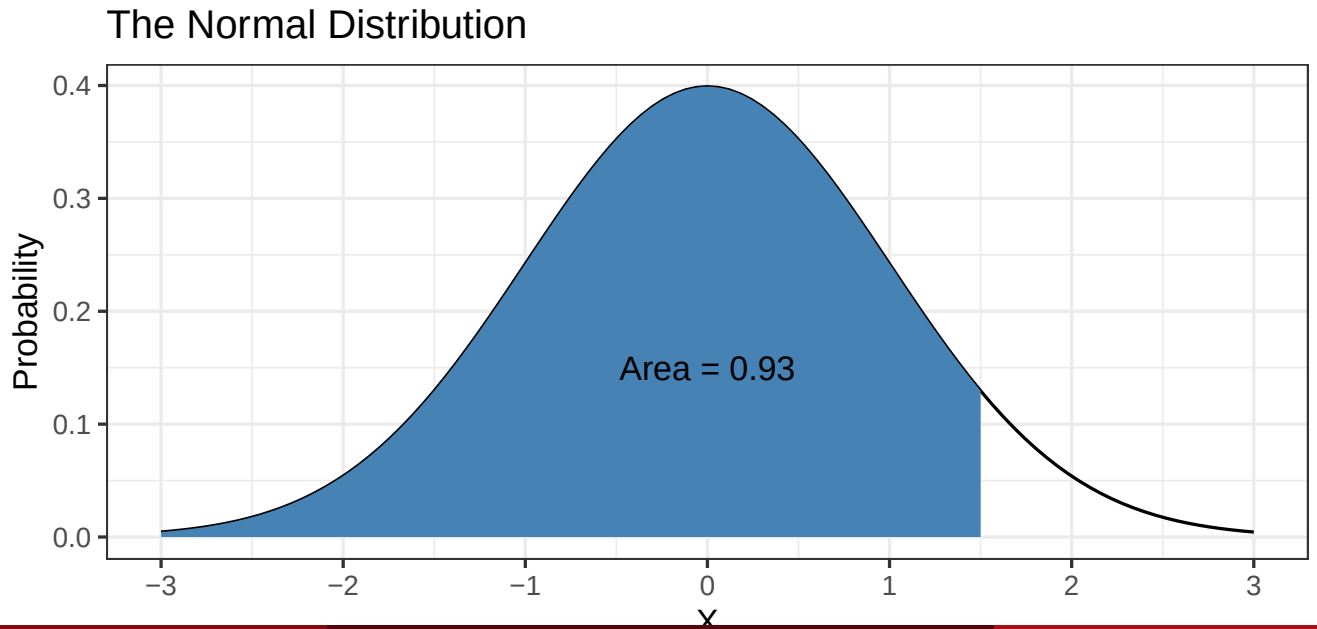
Calculating Normal Probabilities in R

How do we actually find areas under the Normal density curve?

- R has a built-in function for computing cummulative probabilites under Normal densities: `pnorm(q =... , mean =... , sd =... )`
- For example, the following code computes the area left of 1.5 in the Normal distribution with mean 0 and standard deviation 1:

```
pnorm(q =1.5 , mean =0 , sd =1 )
```

```
## [1] 0.9331928
```



Finding Areas of General Regions

The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

Finding Areas of General Regions

The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Finding Areas of General Regions

The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Find the area between -.25 and 1.5 under the Normal density with mean 0 and standard deviation 1.

```
pnorm(q =1.5 , mean =0 , sd =1 ) - pnorm(q = -.25 , mean =0 , sd =1 )
```

```
## [1] 0.5318991
```

Finding Areas of General Regions

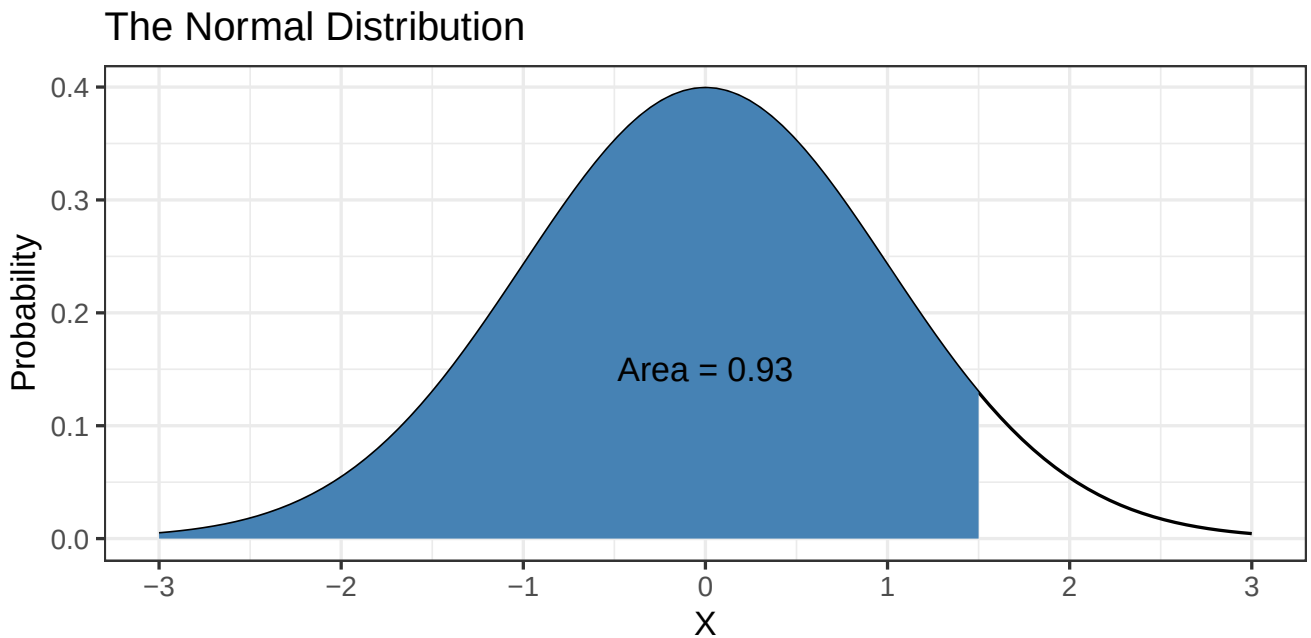
The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Find the area between -.25 and 1.5 under the Normal density with mean 0 and standard deviation 1.

```
pnorm(q =1.5 , mean =0 , sd =1 ) - pnorm(q = -.25 , mean =0 , sd =1 )
```

```
## [1] 0.5318991
```



Finding Areas of General Regions

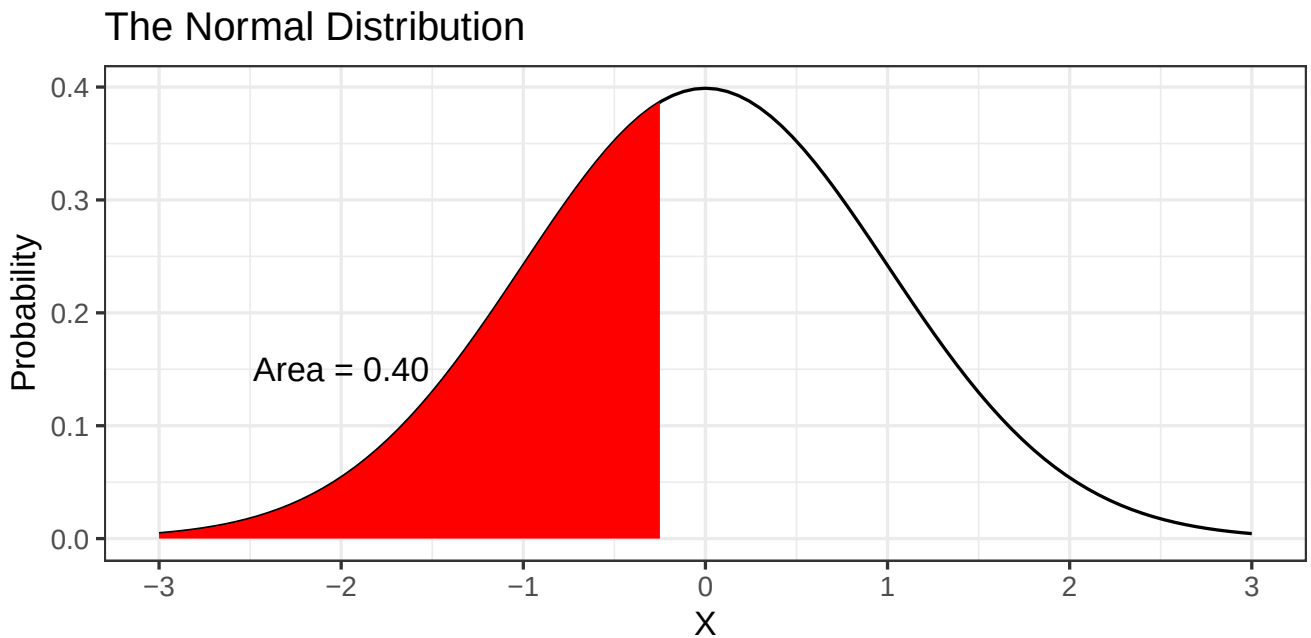
The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Find the area between -.25 and 1.5 under the Normal density with mean 0 and standard deviation 1.

```
pnorm(q =1.5 , mean =0 , sd =1 ) - pnorm(q = -.25 , mean =0 , sd =1 )
```

```
## [1] 0.5318991
```



Finding Areas of General Regions under Normal curve

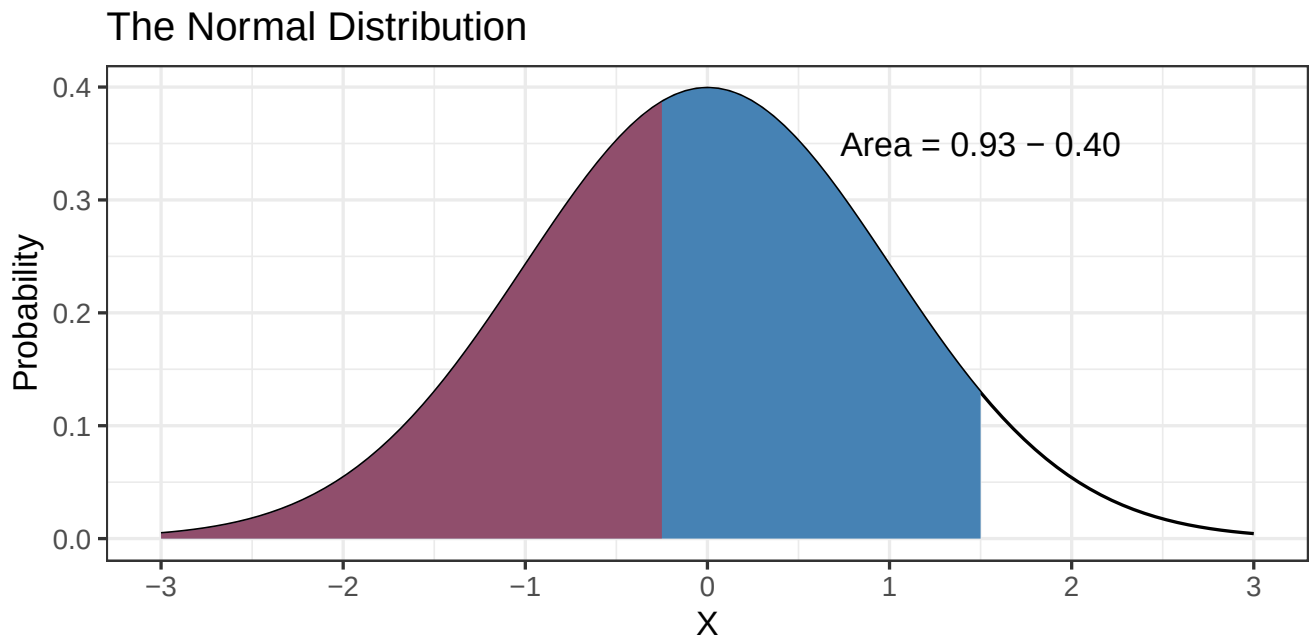
The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Find the area between -.25 and 1.5 under the Normal density with mean 0 and standard deviation 1.

```
pnorm(q =1.5 , mean =0 , sd =1 ) - pnorm(q = -.25 , mean =0 , sd =1 )
```

```
## [1] 0.5318991
```



Finding Areas of General Regions

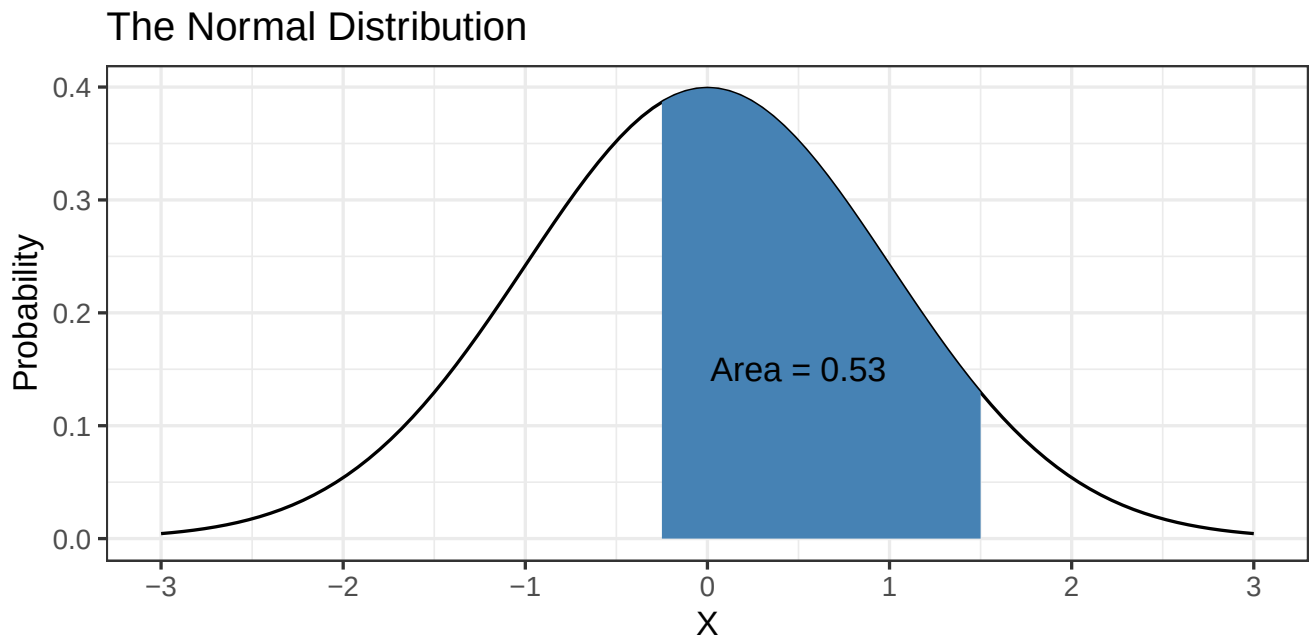
The `pnorm` function lets us compute cumulative areas (i.e. all area to the left of a given value). But how do we compute the area between two values?

- Answer: By computing two cumulative areas and subtracting the results!

Find the area between -.25 and 1.5 under the Normal density with mean 0 and standard deviation 1.

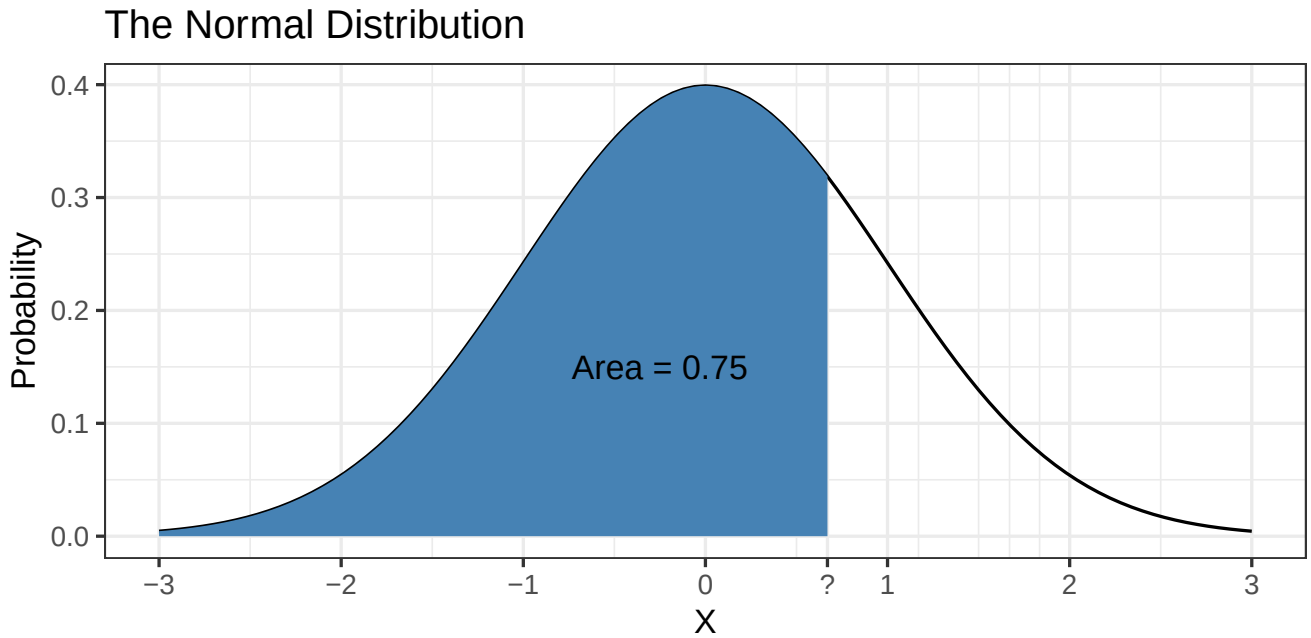
```
pnorm(q =1.5 , mean =0 , sd =1 ) - pnorm(q = -.25 , mean =0 , sd =1 )
```

```
## [1] 0.5318991
```



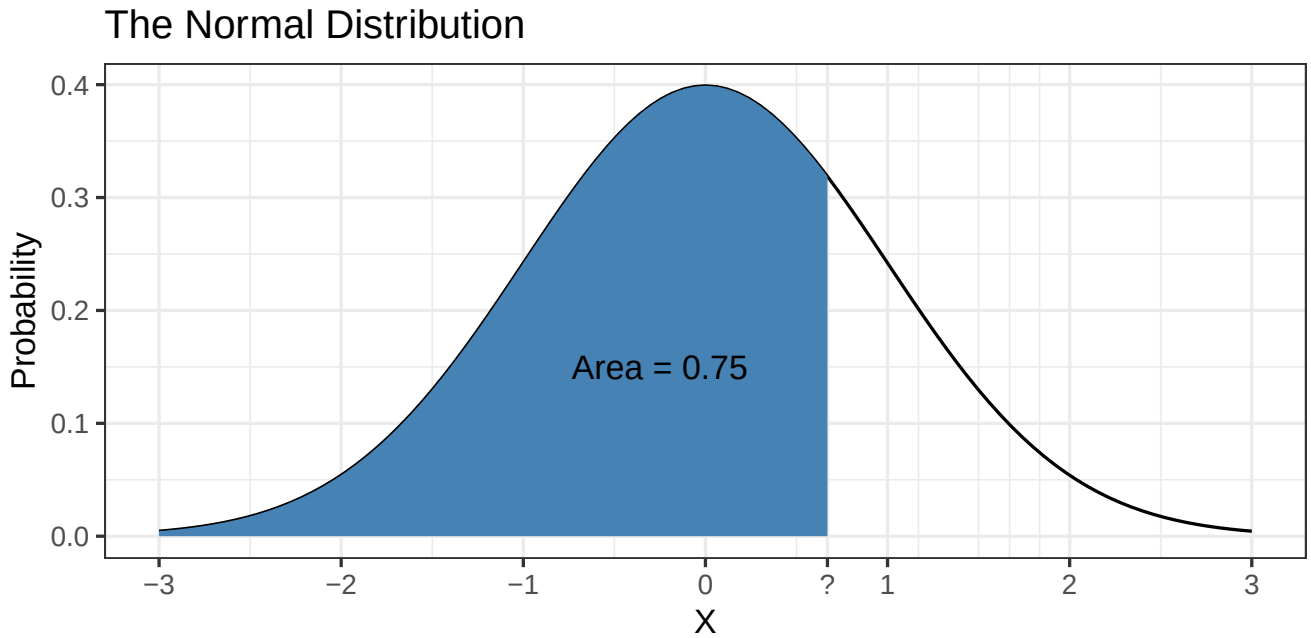
Finding Quantiles

Suppose we instead have the opposite problem: We want to FIND the value of X with a given cumulative area.



Finding Quantiles

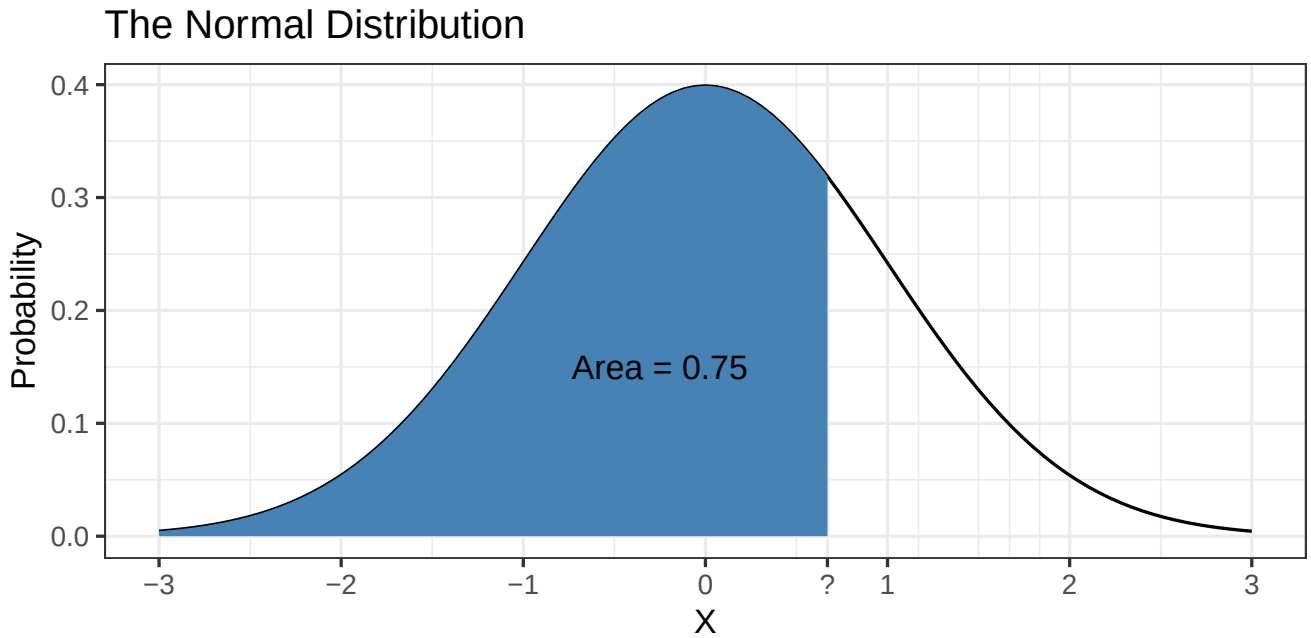
Suppose we instead have the opposite problem: We want to FIND the value of X with a given cumulative area.



- That is, we want to find the .75 quantile (i.e. the 75th percentile)

Finding Quantiles

Suppose we instead have the opposite problem: We want to FIND the value of X with a given cumulative area.

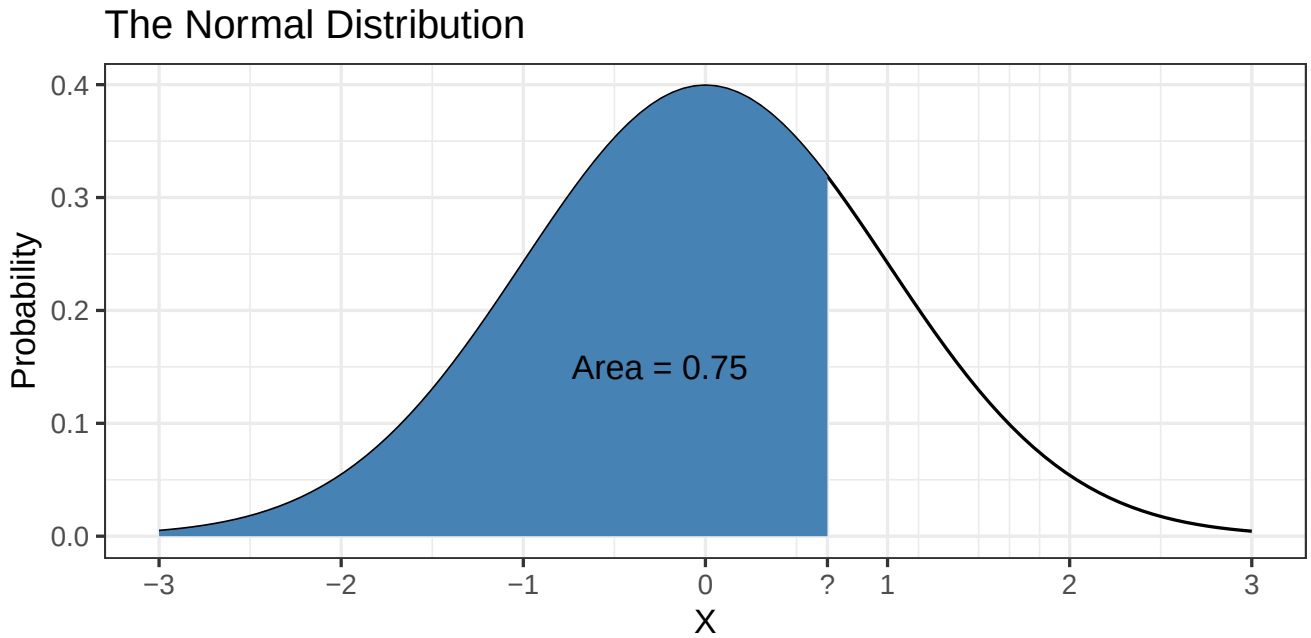


- That is, we want to find the .75 quantile (i.e. the 75th percentile)

R has a built-in function for that too! `qnorm(p =... , mean =... , sd =... )`

Finding Quantiles

Suppose we instead have the opposite problem: We want to FIND the value of X with a given cumulative area.



- That is, we want to find the .75 quantile (i.e. the 75th percentile)

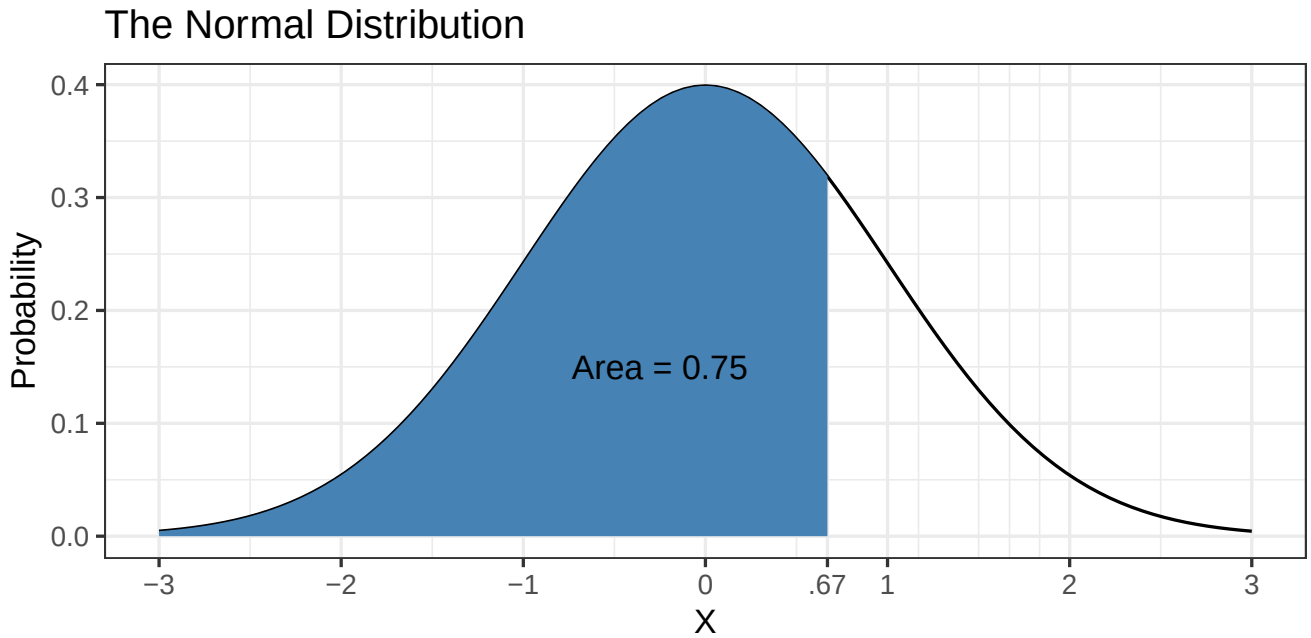
R has a built-in function for that too! `qnorm(p =... , mean =... , sd =... )`

```
qnorm(p =.75 , mean =0 , sd =1 )
```

```
## [1] 0.6744898
```

Finding Quantiles

Suppose we instead have the opposite problem: We want to FIND the value of X with a given cumulative area.



- That is, we want to find the .75 quantile (i.e. the 75th percentile)

R has a built-in function for that too! `qnorm(p =... , mean =... , sd =... )`

```
qnorm(p =.75 , mean =0 , sd =1 )
```

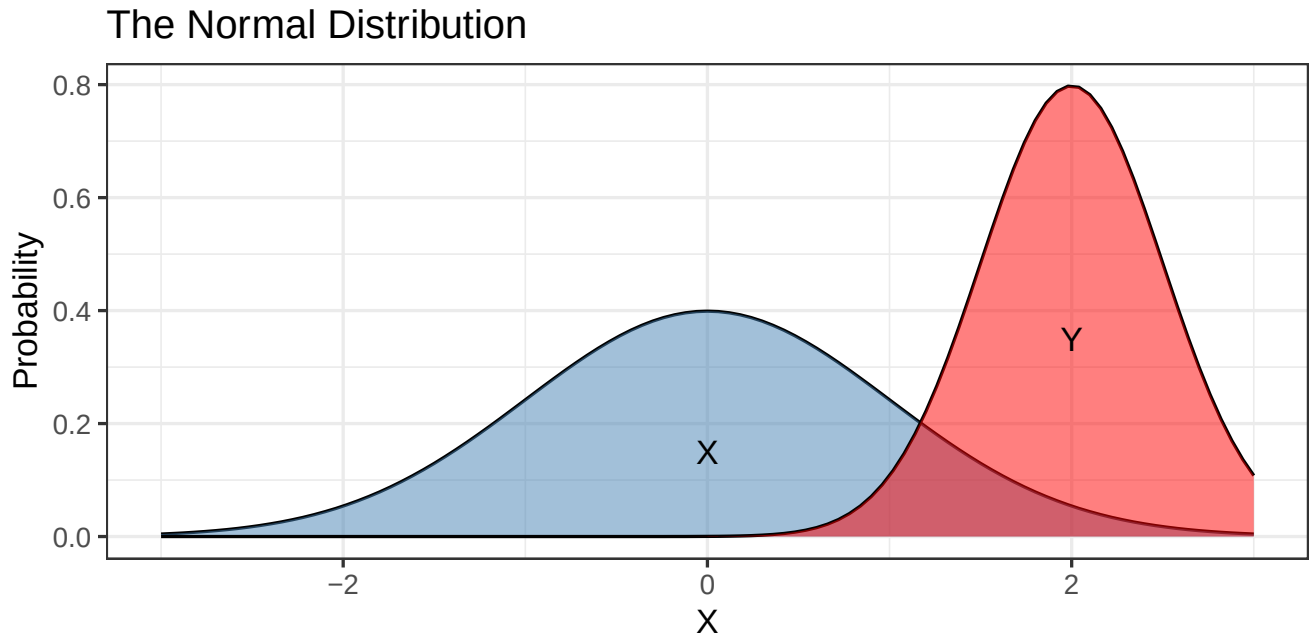
```
## [1] 0.6744898
```

Scale and Translation Invariance

- Consider a Normal variable X with $\mu = 0$ and $\sigma = 1$, and another Normal variable Y with mean $\mu = 2$ and $\sigma = .25$.

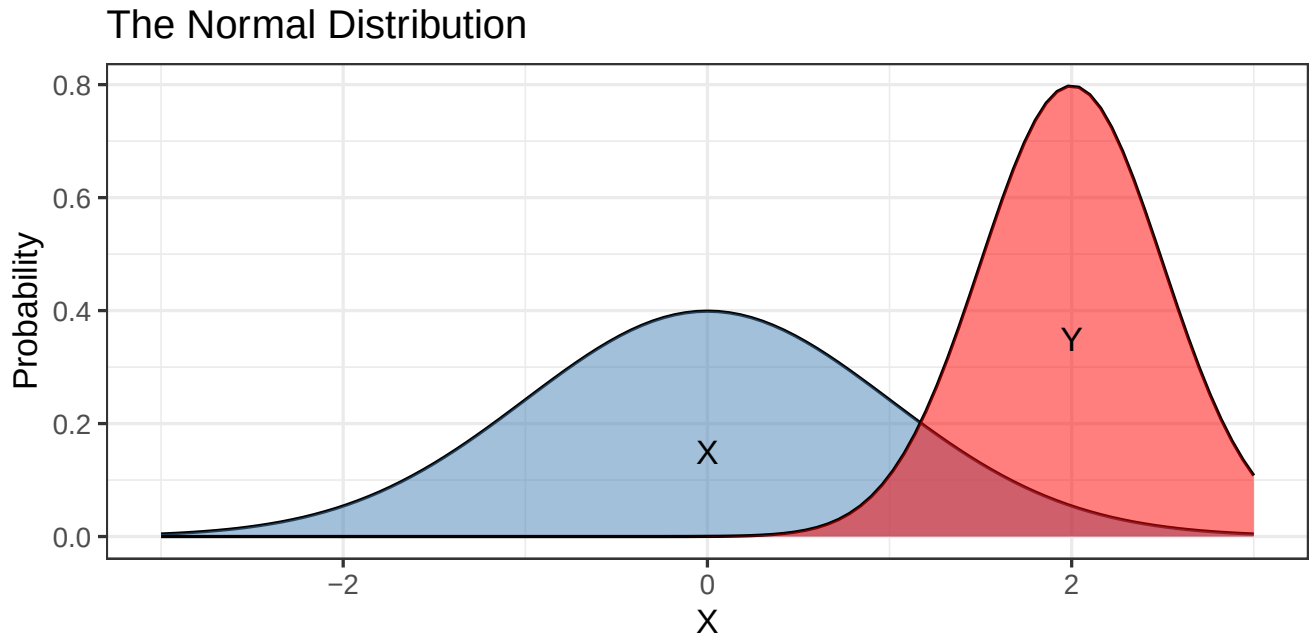
Scale and Translation Invariance

- Consider a Normal variable X with $\mu = 0$ and $\sigma = 1$, and another Normal variable Y with mean $\mu = 2$ and $\sigma = .25$.



Scale and Translation Invariance

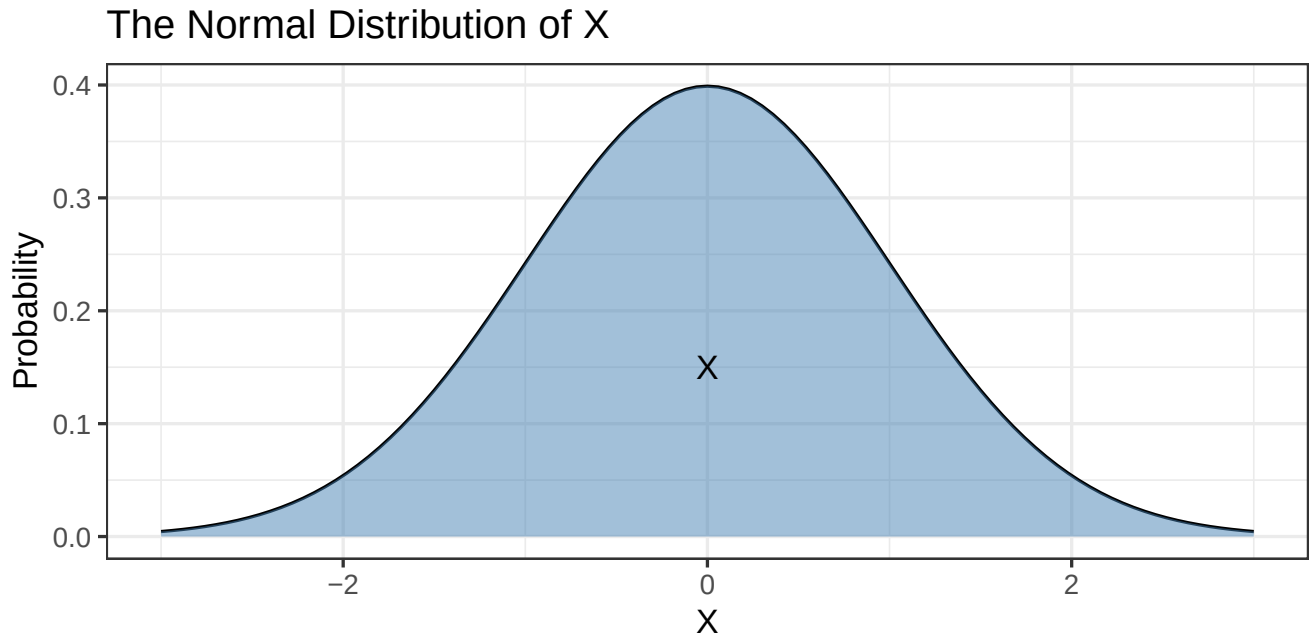
- Consider a Normal variable X with $\mu = 0$ and $\sigma = 1$, and another Normal variable Y with mean $\mu = 2$ and $\sigma = .25$.



- The distributions for X and Y have different means and different heights and widths. . .
  - But otherwise have identitical shapes!

Scale and Translation Invariance

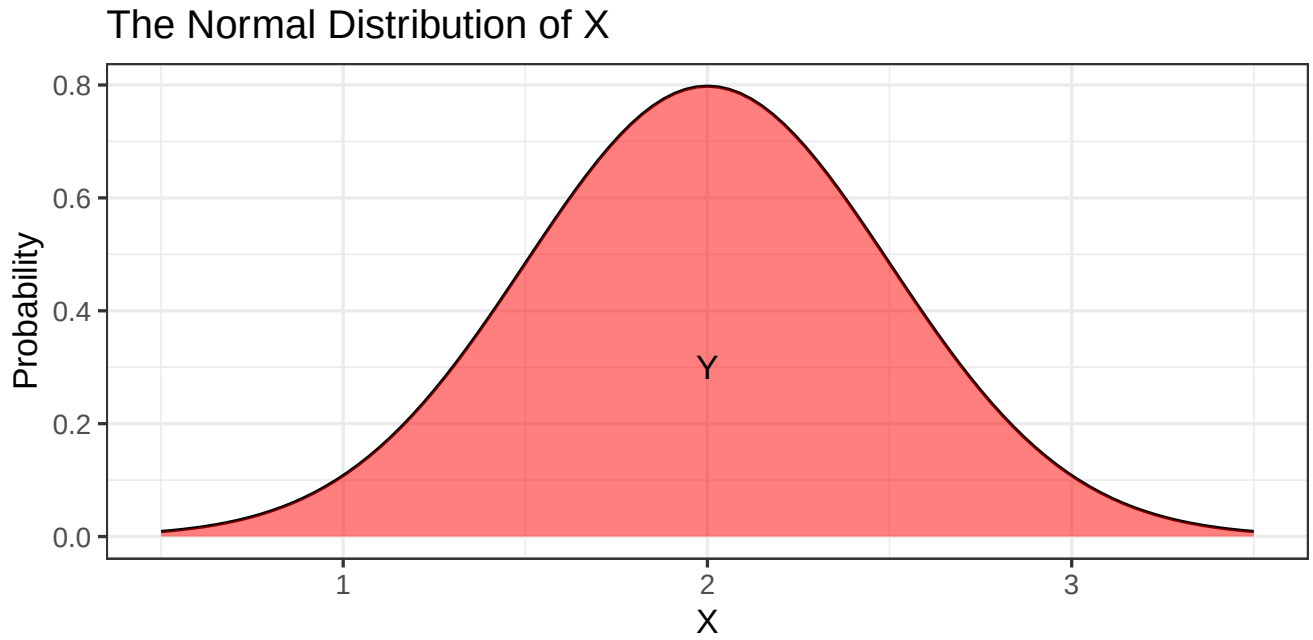
- Consider a Normal variable X with $\mu = 0$ and $\sigma = 1$, and another Normal variable Y with mean $\mu = 2$ and $\sigma = .25$.



- The distributions for X and Y have different means and different heights and widths. . .
  - But otherwise have identitical shapes!

Scale and Translation Invariance

- Consider a Normal variable X with $\mu = 0$ and $\sigma = 1$, and another Normal variable Y with mean $\mu = 2$ and $\sigma = .25$.



- The distributions for X and Y have different means and different heights and widths. . .
  - But otherwise have identical shapes!

Location-Scale Transformations

The previous example suggest that if we shift and rescale a Normal random variable, we should still get a Normal random variable

Location-Scale Transformations

The previous example suggest that if we shift and rescale a Normal random variable, we should still get a Normal random variable

Theorem

Suppose X is a Normal random variable with mean μ and standard deviation σ. Then $Z = \frac{X-\mu}{\sigma}$ is a Normal random variable with mean 0 and standard deviation 1.

Location-Scale Transformations

The previous example suggest that if we shift and rescale a Normal random variable, we should still get a Normal random variable

Theorem

Suppose X is a Normal random variable with mean μ and standard deviation σ. Then $Z = \frac{X-\mu}{\sigma}$ is a Normal random variable with mean 0 and standard deviation 1.

The Normal variable with mean 0 and standard deviation 1 is given a special name: **the standard Normal**.

Location-Scale Transformations

The previous example suggest that if we shift and rescale a Normal random variable, we should still get a Normal random variable

Theorem

Suppose X is a Normal random variable with mean μ and standard deviation σ. Then $Z = \frac{X-\mu}{\sigma}$ is a Normal random variable with mean 0 and standard deviation 1.

The Normal variable with mean 0 and standard deviation 1 is given a special name: **the standard Normal**.

The process of subtracting off the mean from a random variable and dividing by the standard deviation is called **standardizing.**

Location-Scale Transformations

The previous example suggest that if we shift and rescale a Normal random variable, we should still get a Normal random variable

Theorem

Suppose X is a Normal random variable with mean μ and standard deviation σ. Then $Z = \frac{X-\mu}{\sigma}$ is a Normal random variable with mean 0 and standard deviation 1.

The Normal variable with mean 0 and standard deviation 1 is given a special name: **the standard Normal**.

The process of subtracting off the mean from a random variable and dividing by the standard deviation is called **standardizing.**

It's often useful to standardize a variable so that we only need to consider a single density function (the *standard* Normal density) rather than many (one for each choice of μ and σ)

Section 2

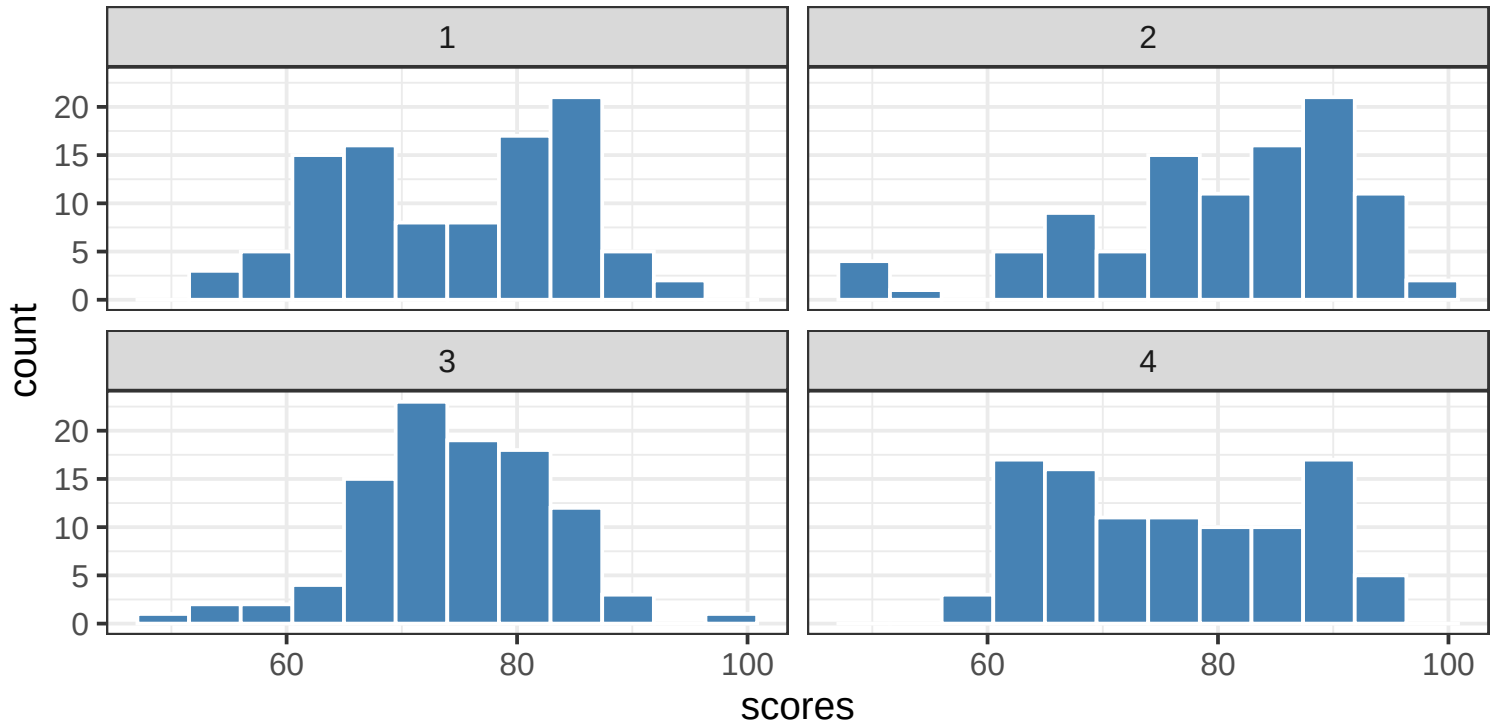
The Central Limit Theorem

Exam scores

Consider the following distributions for scores on a statistics exam for 4 classes of 100 students:

Exam scores

Consider the following distributions for scores on a statistics exam for 4 classes of 100 students:



Random Sample Means

Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

Random Sample Means

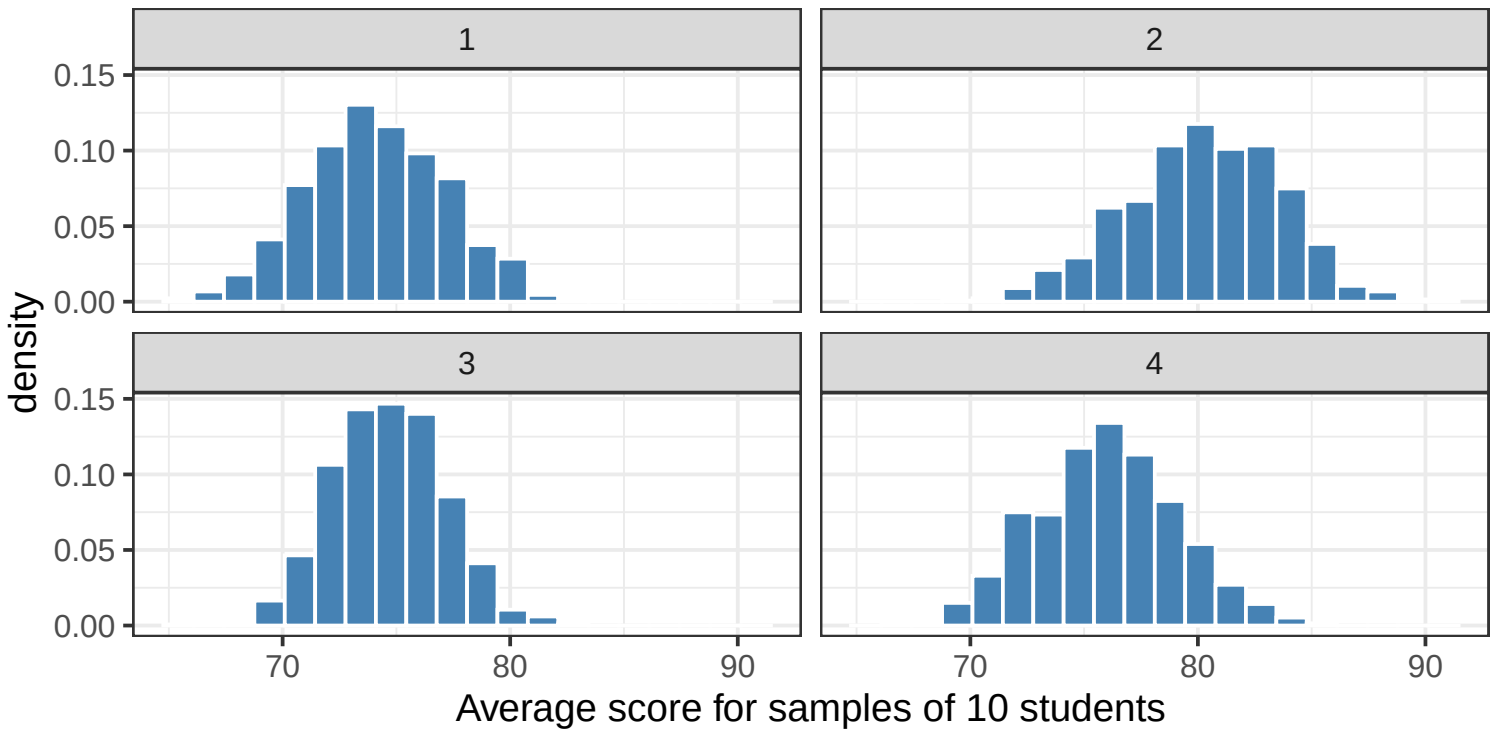
Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

- What does the distribution of sample means $\bar{x}$ look like?

Random Sample Means

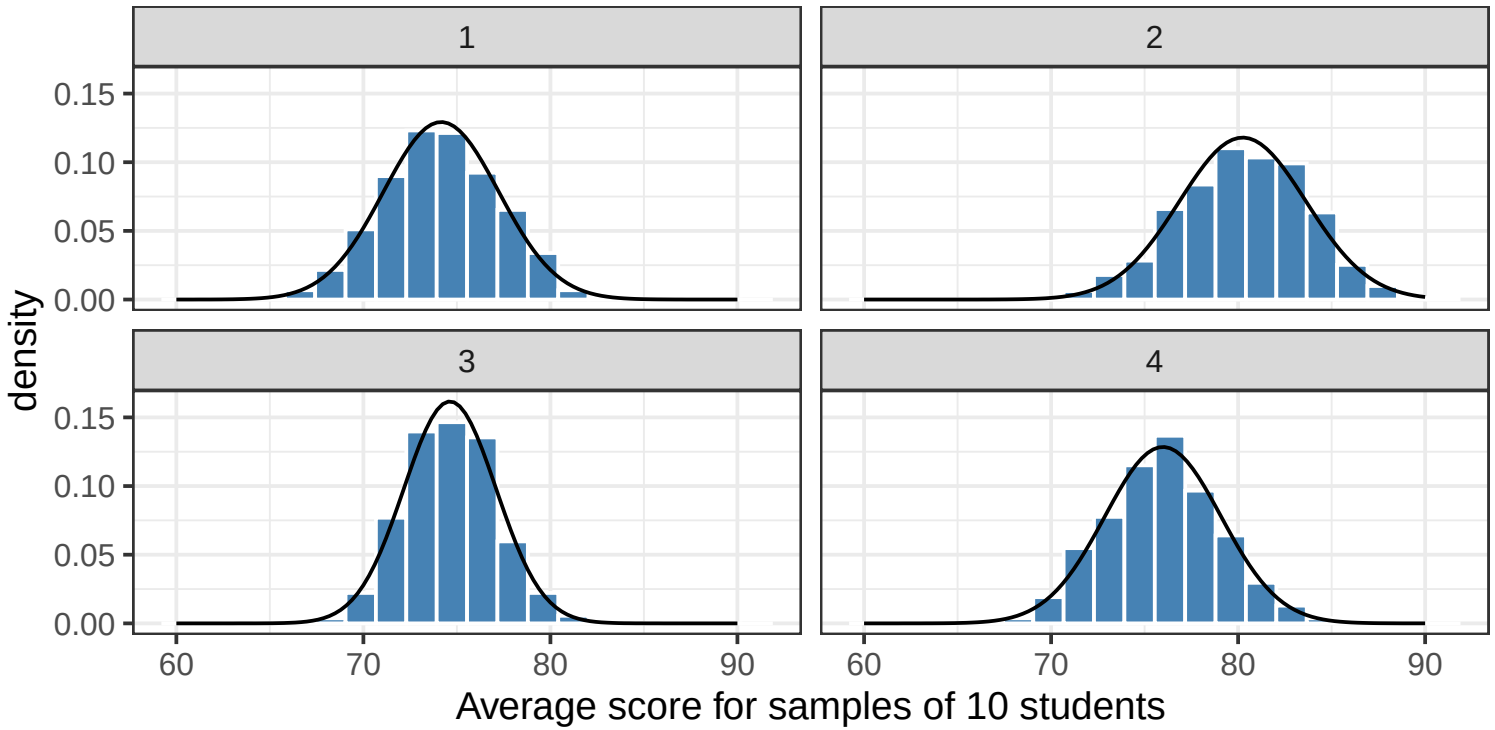
Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

- What does the distribution of sample means $\bar{x}$ look like?



The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

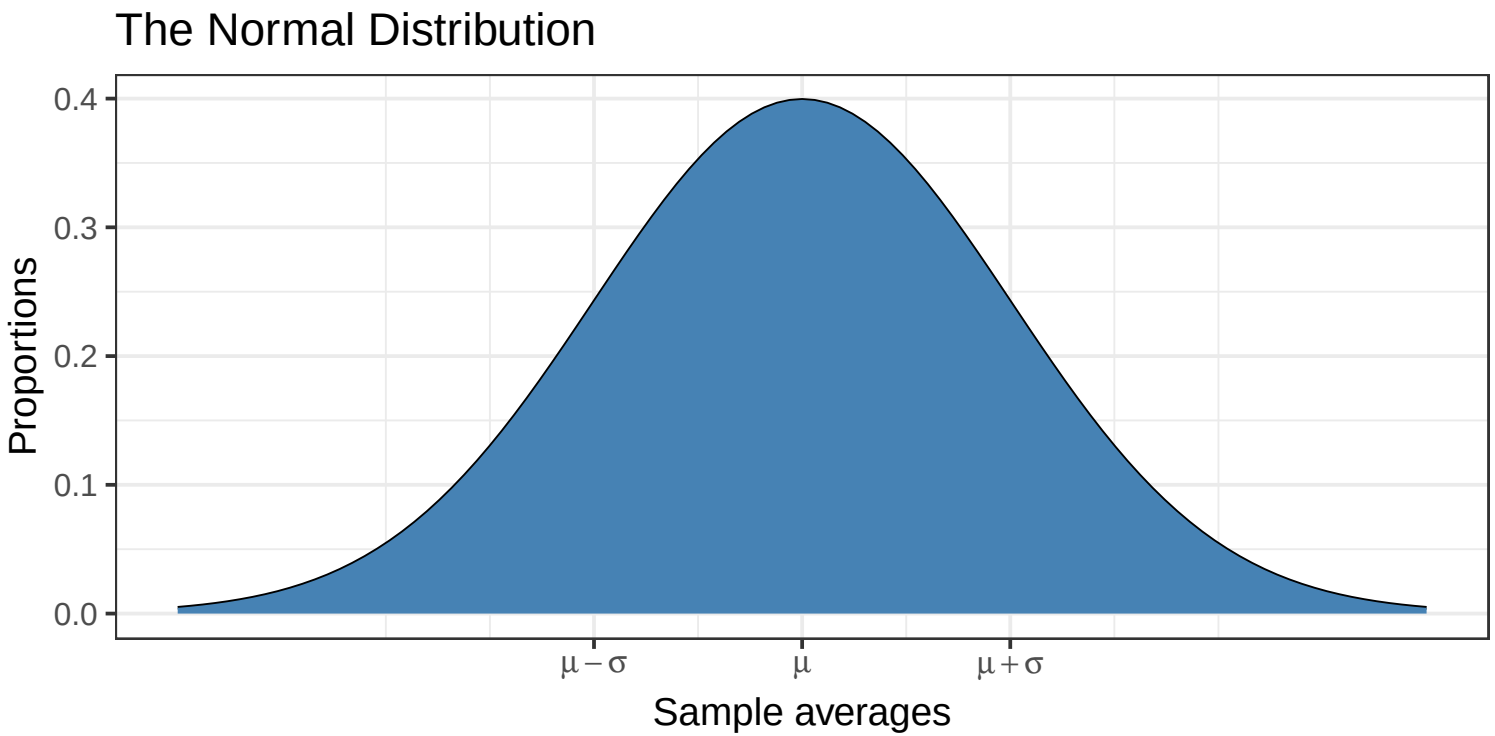


The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

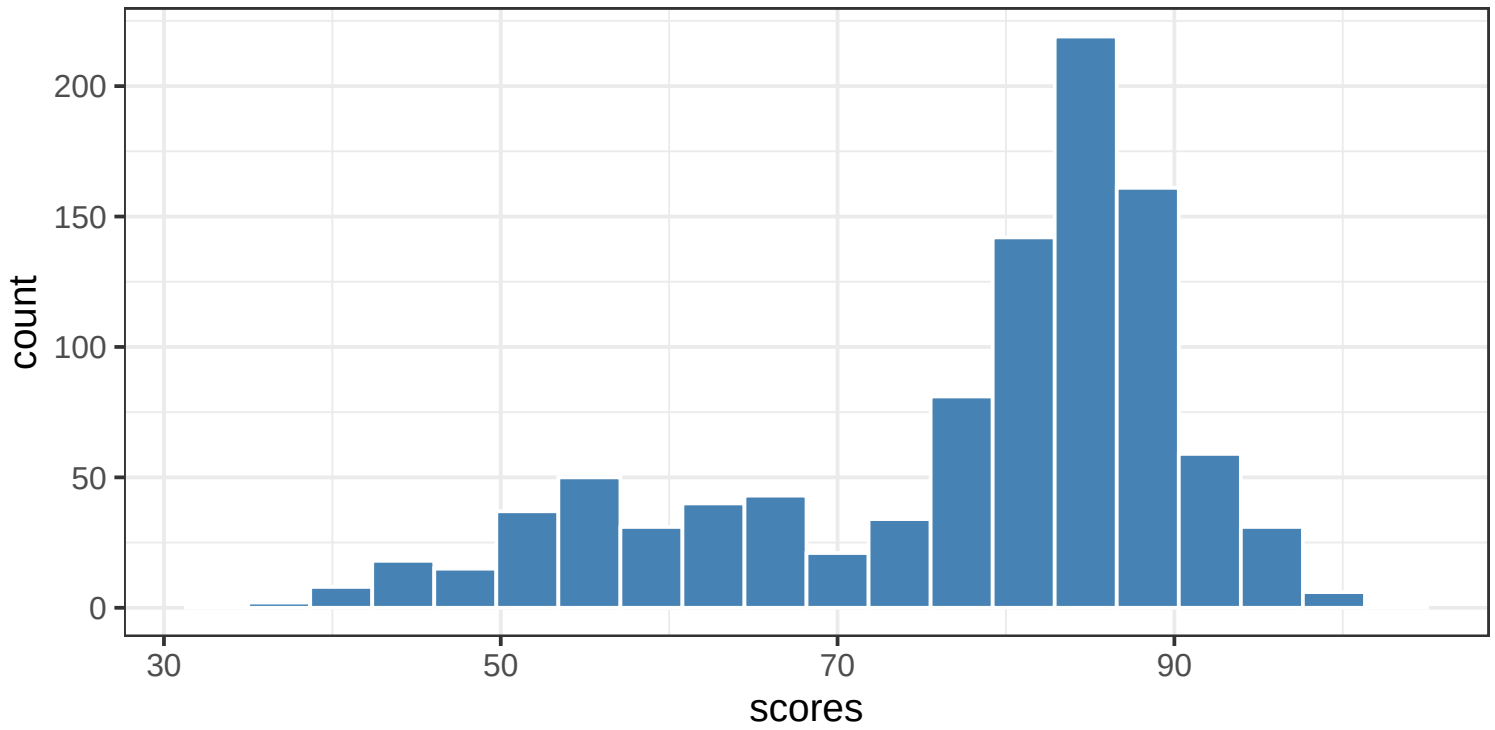


Effect of Sample Size

Suppose we have a class of 1000 students with the following score distribution

Effect of Sample Size

Suppose we have a class of 1000 students with the following score distribution

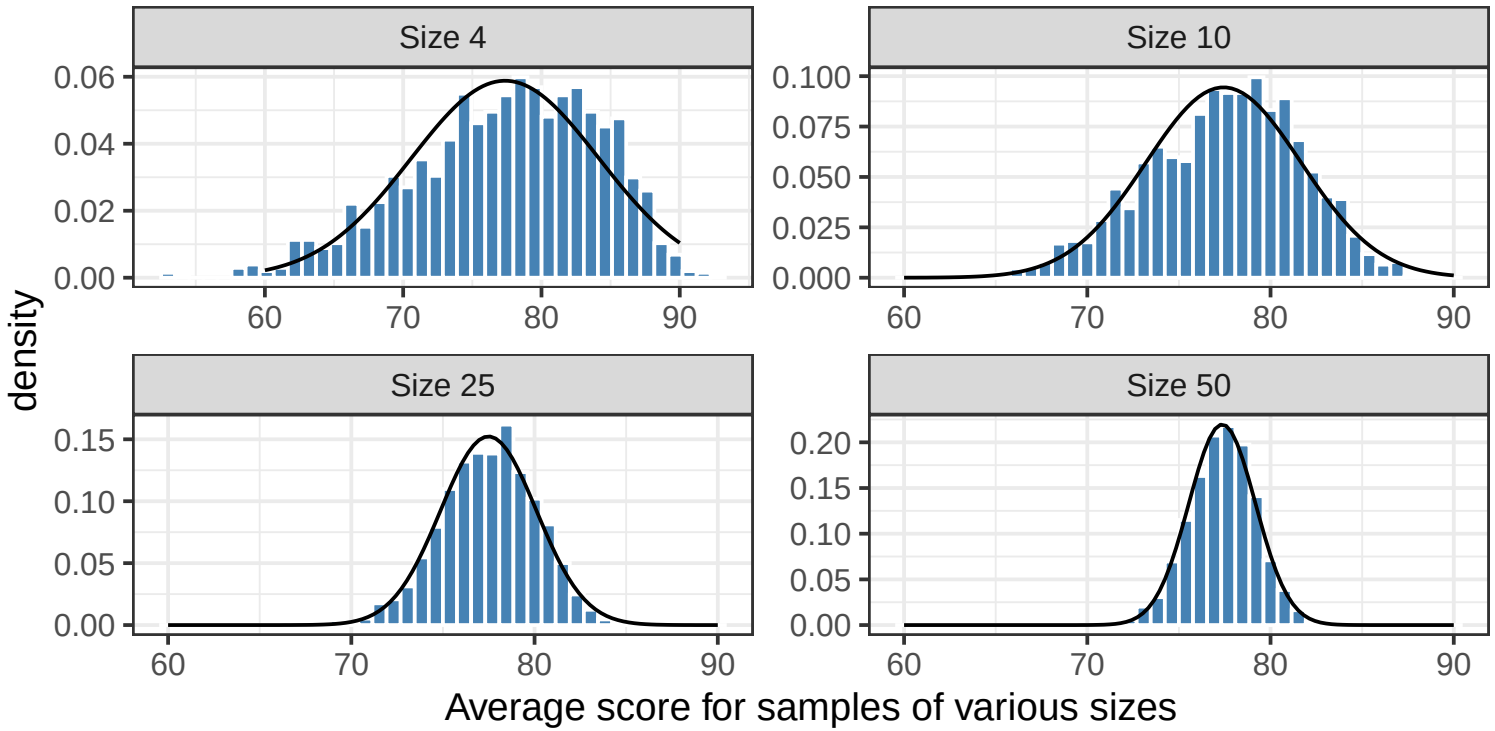


Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?

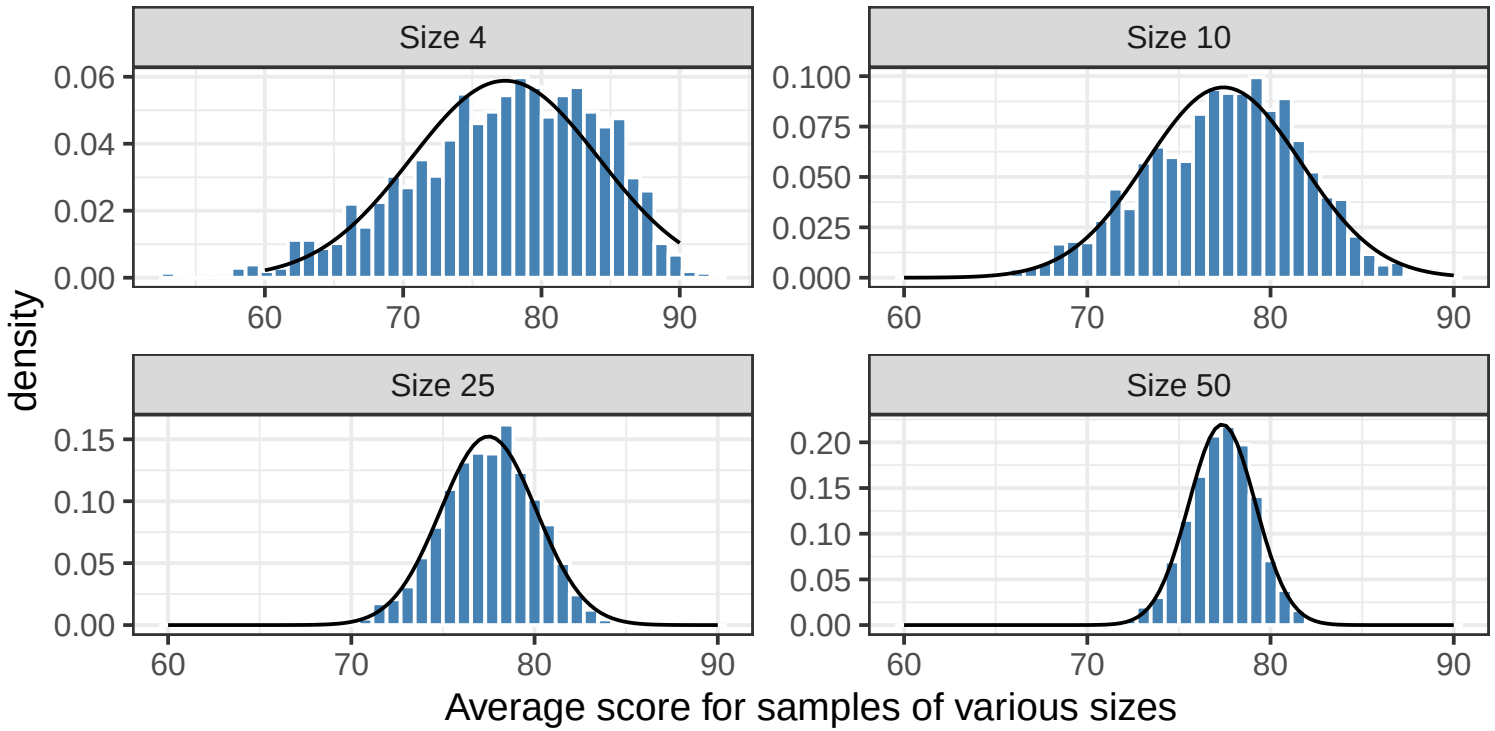
Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?



Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?



- As sample size increases, sampling distribution becomes **more** Normal, with **decreasing** variance

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

A proof of the CLT requires more advanced techniques in probability (See Math 391). But intuitively. . .

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

A proof of the CLT requires more advanced techniques in probability (See Math 391). But intuitively...

A sample mean is obtained by adding together INDEPENDENT values from the population.

In order to get a very large or very small value, nearly ALL of the independent values need to be extreme.

To get a moderate value, many can be extreme in the opposite direction, or many can be moderate (or several variations in between).

There are more ways to obtain moderate values in an average than to obtain extreme values

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases
- We can use properties of the Normal distribution to determine probabilities of obtaining extreme sample statistics

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases
- We can use properties of the Normal distribution to determine probabilities of obtaining extreme sample statistics
- Statistical inference can be performed using theoretical density functions, in addition to using simulation and bootstrapping