Lecture 6: Chapter 6

C C Moxley

UAB Mathematics

1 October 15

§6.1 Continuous Probability Distributions

Last week, we discussed the binomial probability distribution, which was discrete.

§6.1 Continuous Probability Distributions

Last week, we discussed the binomial probability distribution, which was discrete. The continuous analogue of a binomial distribution is the normal distribution!

§6.1 Continuous Probability Distributions

Last week, we discussed the binomial probability distribution, which was discrete. The continuous analogue of a binomial distribution is the normal distribution!

Definition (Normal Distribution)

If a continuous random variable has a symmetric, bell-shaped graph and can be described by the equation

$$y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}},$$

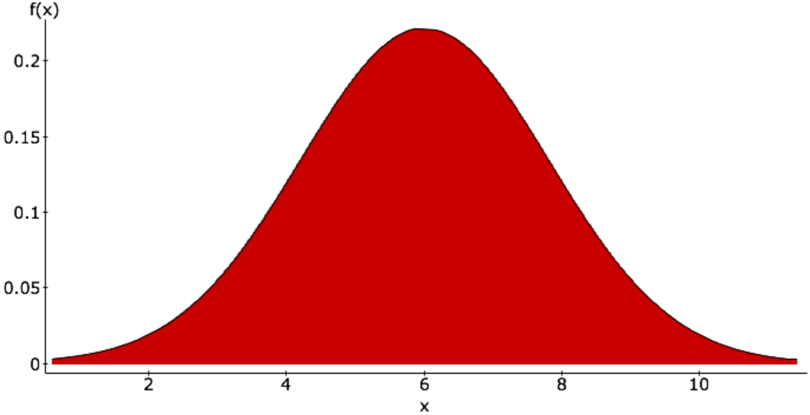
then we say that the random variable has a **normal distribution**.

§6.1 Normal Distribution

Here's the graph of a normal distribution with $\sigma = 1.8$ and $\mu = 6$.

§6.1 Normal Distribution

Here's the graph of a normal distribution with $\sigma = 1.8$ and $\mu = 6$.

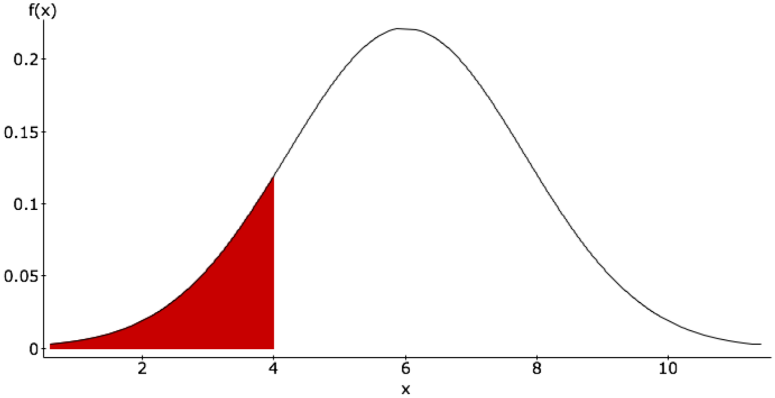


§6.1 Normal Distribution

The area under the curve between points a and b (where $a < b$) on the x-axis represents the probability that the random variable takes values between a and b, i.e. $P(a \leq x \leq b)$. In the image below, $a = -\infty$ and $b = 4$.

§6.1 Normal Distribution

The area under the curve between points a and b (where $a < b$) on the x-axis represents the probability that the random variable takes values between a and b, i.e. $P(a \leq x \leq b)$. In the image below, $a = -\infty$ and $b = 4$.



§6.1 Normal Distribution

There are two ways of calculating the probability $P(-\infty \leq x \leq 4)$, where x is a Normal(6,1.8).

§6.1 Normal Distribution

There are two ways of calculating the probability $P(-\infty \leq x \leq 4)$, where x is a Normal(6,1.8). You could either

- calculate the corresponding z-score and look up the value in a table or

§6.1 Normal Distribution

There are two ways of calculating the probability $P(-\infty \leq x \leq 4)$, where x is a Normal(6,1.8). You could either

- calculate the corresponding z-score and look up the value in a table or
- use StatCrunch to calculate the probability, ensuring that you have changed the mean and standard deviation in the normal calculator appropriately.

§6.2 Uniform Distribution

To help us see how the area under the curve relates to probabilities, lets consider the uniform random variable.

§6.2 Uniform Distribution

To help us see how the area under the curve relates to probabilities, lets consider the uniform random variable. If you recall, this distribution is flat between its highest and lowest values.

§6.2 Uniform Distribution

To help us see how the area under the curve relates to probabilities, lets consider the uniform random variable. If you recall, this distribution is flat between its highest and lowest values.

Definition (Uniform Distribution)

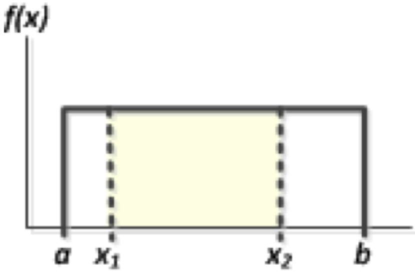
A continuous random variable as a **uniform distribution** if its values are spread evenly over the range of possibilities. The graph of a uniform distribution results in a rectangular shape.

§6.2 Uniform Distribution

Here, the lower bound of values is $a = 1$, and the upper bound is $b = 15$. The height of the graph is $\frac{1}{b-a}$, so the area under the graph between b and a is 1. Then the area between $x_1 = 3$ and $x_2 = 8$ is $\frac{1}{b-a} \cdot (x_2 - x_1) = \frac{5}{14} \approx 0.357$.

§6.2 Uniform Distribution

Here, the lower bound of values is $a = 1$, and the upper bound is $b = 15$. The height of the graph is $\frac{1}{b-a}$, so the area under the graph between b and a is 1. Then the area between $x_1 = 3$ and $x_2 = 8$ is $\frac{1}{b-a} \cdot (x_2 - x_1) = \frac{5}{14} \approx 0.357$.



§6.2 What Makes a Density Curve?

In order for a curve to be a density curve, i.e. to describe a continuous probability distribution, the following must occur.

- The total area under the curve must equal 1.
- Every point on the curve must have a vertical height that is 0 or greater, i.e. it must lie above the x-axis.

§6.2 Standard Normal Distribution

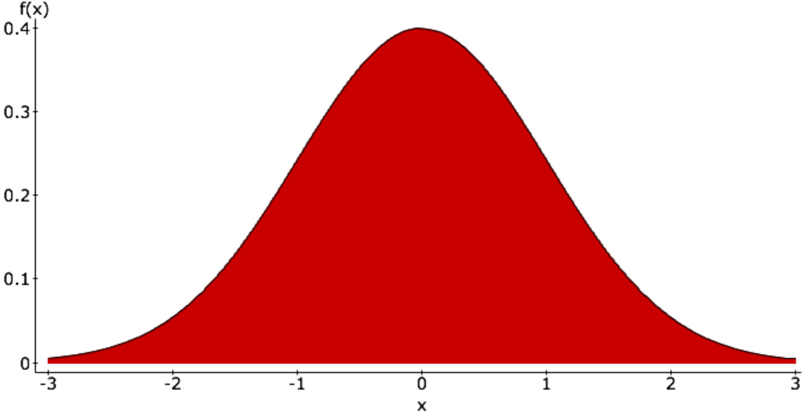
Definition (Standard Normal Distribution)

The **standard normal distribution** is the normal distribution with parameters $\mu = 0$ and $\sigma = 1$.

§6.2 Standard Normal Distribution

Definition (Standard Normal Distribution)

The **standard normal distribution** is the normal distribution with parameters $\mu = 0$ and $\sigma = 1$.



§6.2 Finding Probabilities Given z-Scores

It's simple to calculate $P(x < z)$.

§6.2 Finding Probabilities Given z-Scores

It's simple to calculate $P(x < z)$.

- Use the calculator in StatCrunch.

§6.2 Finding Probabilities Given z-Scores

It's simple to calculate $P(x < z)$.

- Use the calculator in StatCrunch.
- Look up the value in a table - pay careful attention to whether you are looking at the smaller or larger portion!

§6.2 Finding Probabilities Given z-Scores

It's simple to calculate $P(x < z)$.

- Use the calculator in StatCrunch.
- Look up the value in a table - pay careful attention to whether you are looking at the smaller or larger portion!

Once this has been accomplished, you can calculate other probabilities $P(x > z)$, $P(z_1 < x < z_2)$, and $P(z_1 > x > z_2)$.

§6.2 Example

A bone density test reading between -1.00 and -2.50 indicates that a patient has osteopenia. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

§6.2 Example

A bone density test reading between -1.00 and -2.50 indicates that a patient has osteopenia. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

We must calculate $P(-2.50 < x < -1.00)$.

§6.2 Example

A bone density test reading between -1.00 and -2.50 indicates that a patient has osteopenia. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

We must calculate $P(-2.50 < x < -1.00)$. To do this, we calculate: $P(x < -1.00) - P(x < -2.50)$.

§6.2 Example

A bone density test reading between -1.00 and -2.50 indicates that a patient has osteopenia. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

We must calculate $P(-2.50 < x < -1.00)$. To do this, we calculate: $P(x < -1.00) - P(x < -2.50)$. This yields $P(-2.50 < x < -1.00) \approx 0.1587 - 0.0062 = 0.1525$.

§6.2 Example

A bone density test reading between -1.00 and -2.50 indicates that a patient has osteopenia. Find the probability that a randomly selected subject has a reading between -1.00 and -2.50.

We must calculate $P(-2.50 < x < -1.00)$. To do this, we calculate: $P(x < -1.00) - P(x < -2.50)$. This yields $P(-2.50 < x < -1.00) \approx 0.1587 - 0.0062 = 0.1525$.

Let's draw the pictures for this to understand why it makes sense.

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

To do this, we must solve for z the equation $P(x < z) = 0.95$.

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

To do this, we must solve for z the equation $P(x < z) = 0.95$. We can also do this in the StatCrunch calculator or through the tables.

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

To do this, we must solve for z the equation $P(x < z) = 0.95$. We can also do this in the StatCrunch calculator or through the tables.

This yields $P(x < 1.645) \approx 0.95$. So the z-score of 1.645 marks the 95th Percentile.

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

To do this, we must solve for z the equation $P(x < z) = 0.95$. We can also do this in the StatCrunch calculator or through the tables.

This yields $P(x < 1.645) \approx 0.95$. So the z-score of 1.645 marks the 95th Percentile. What, then, marks the 5th Percentile?

§6.2 Example

Find the z-score corresponding to the 95th Percentile.

To do this, we must solve for z the equation $P(x < z) = 0.95$. We can also do this in the StatCrunch calculator or through the tables.

This yields $P(x < 1.645) \approx 0.95$. So the z-score of 1.645 marks the 95th Percentile. What, then, marks the 5th Percentile? Well, it would be -1.645 **by symmetry**.

§6.2 Critical Values

For the standard normal distribution, **critical values** separate typical from untypical values.

§6.2 Critical Values

For the standard normal distribution, **critical values** separate typical from untypical values. Also, z_α denotes the z-score with an area of α to its right.

§6.2 Critical Values

For the standard normal distribution, **critical values** separate typical from untypical values. Also, z_α denotes the z-score with an area of α to its right. When calculating a z_α, be careful of how you read the tables! The tables often give areas to the left of the z-score, not the right.

§6.3 Applications of the Normal Distribution

If you would like to use a table to calculate the probability that a normal random variable with mean μ and standard deviation σ takes a value less than X, you can convert it to a standard normal distribution:

§6.3 Applications of the Normal Distribution

If you would like to use a table to calculate the probability that a normal random variable with mean μ and standard deviation σ takes a value less than X, you can convert it to a standard normal distribution:

$$P(x < X) =$$

§6.3 Applications of the Normal Distribution

If you would like to use a table to calculate the probability that a normal random variable with mean μ and standard deviation σ takes a value less than X, you can convert it to a standard normal distribution:

$$P(x < X) = P\left(z < \frac{X - \mu}{\sigma}\right),$$

§6.3 Applications of the Normal Distribution

If you would like to use a table to calculate the probability that a normal random variable with mean μ and standard deviation σ takes a value less than X, you can convert it to a standard normal distribution:

$$P(x < X) = P\left(z < \frac{X - \mu}{\sigma}\right),$$

where x is the original random variable and z is a standard normal random variable.

§6.3 Example

British Airways requires its in-flight crew to have heights between 62 and 73 inches. Given that men have normally distributed heights with a mean of 69.5 inches and a standard deviation of 2.4 inches, find the percentage of men who could be in-flight crew for British Airways.

§6.3 Example

British Airways requires its in-flight crew to have heights between 62 and 73 inches. Given that men have normally distributed heights with a mean of 69.5 inches and a standard deviation of 2.4 inches, find the percentage of men who could be in-flight crew for British Airways.

We need to calculate

$$P\left(\frac{62 - 69.5}{2.4} < x < \frac{73 - 69.5}{2.4}\right)$$

$$\approx$$

§6.3 Example

British Airways requires its in-flight crew to have heights between 62 and 73 inches. Given that men have normally distributed heights with a mean of 69.5 inches and a standard deviation of 2.4 inches, find the percentage of men who could be in-flight crew for British Airways.

We need to calculate

$$P\left(\frac{62-69.5}{2.4} < x < \frac{73-69.5}{2.4}\right)$$

$$\approx P\left(-3.125 < x < 1.458\right) \approx 0.9267.$$

§6.3 Example

British Airways requires its in-flight crew to have heights between 62 and 73 inches. Given that men have normally distributed heights with a mean of 69.5 inches and a standard deviation of 2.4 inches, find the percentage of men who could be in-flight crew for British Airways.

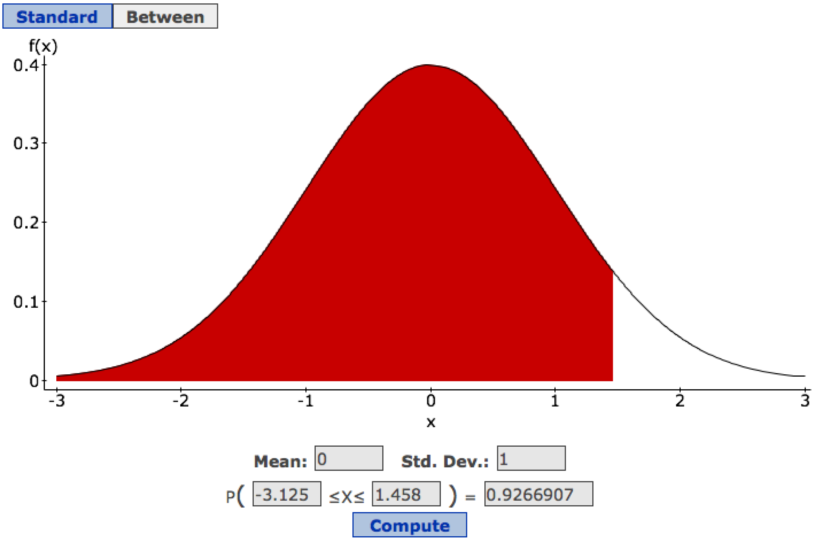
We need to calculate

$$P\left(\frac{62-69.5}{2.4} < x < \frac{73-69.5}{2.4}\right)$$

$$\approx P\left(-3.125 < x < 1.458\right) \approx 0.9267.$$

This area is recorded in the next slide.

§6.3 Example


Standard
Between
f(x)
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
x
Mean: 0
Std. Dev.: 1
P(-3.125 ≤X≤ 1.458) = 0.9266907
Compute

§6.3 Example

A local gym offers three running groups for their members - a fast group, and medium-paced group, and a slow group. If gym-going runners have a normally distributed pace with an average of 7.5 minutes per mile with a standard deviation of 1.25 minutes per mile, what should be the two cut-off times for the three groups?

§6.3 Example

A local gym offers three running groups for their members - a fast group, and medium-paced group, and a slow group. If gym-going runners have a normally distributed pace with an average of 7.5 minutes per mile with a standard deviation of 1.25 minutes per mile, what should be the two cut-off times for the three groups?

We need about one third of the runners in each group, so we need to find X_1 so that $P(x < X_1) = 0.33333$ and X_2 so that $P(x < X_2) = 0.66666$.

§6.3 Example

A local gym offers three running groups for their members - a fast group, and medium-paced group, and a slow group. If gym-going runners have a normally distributed pace with an average of 7.5 minutes per mile with a standard deviation of 1.25 minutes per mile, what should be the two cut-off times for the three groups?

We need about one third of the runners in each group, so we need to find X_1 so that $P(x < X_1) = 0.33333$ and X_2 so that $P(x < X_2) = 0.66666$. We get (for x begin a standard normal random variable)

$$X_1 = -0.431 \text{ and } X_2 = 0.431$$

§6.3 Example

A local gym offers three running groups for their members - a fast group, and medium-paced group, and a slow group. If gym-going runners have a normally distributed pace with an average of 7.5 minutes per mile with a standard deviation of 1.25 minutes per mile, what should be the two cut-off times for the three groups?

We need about one third of the runners in each group, so we need to find X_1 so that $P(x < X_1) = 0.33333$ and X_2 so that $P(x < X_2) = 0.66666$. We get (for x begin a standard normal random variable)

$$X_1 = -0.431 \quad \text{and} \quad X_2 = 0.431$$

This means that the cut-off times should be $\overline{X_1} = 7.5 + (-0.431)(1.25) = 6.96$ and $\overline{X_2} = 7.5 + (0.431)(1.25) = 8.04$.

§6.3 Example

A local gym offers three running groups for their members - a fast group, and medium-paced group, and a slow group. If gym-going runners have a normally distributed pace with an average of 7.5 minutes per mile with a standard deviation of 1.25 minutes per mile, what should be the two cut-off times for the three groups?

We need about one third of the runners in each group, so we need to find X_1 so that $P(x < X_1) = 0.33333$ and X_2 so that $P(x < X_2) = 0.66666$. We get (for x begin a standard normal random variable)

$$X_1 = -0.431 \quad \text{and} \quad X_2 = 0.431$$

This means that the cut-off times should be $\overline{X_1} = 7.5 + (-0.431)(1.25) = 6.96$ and $\overline{X_2} = 7.5 + (0.431)(1.25) = 8.04$. Draw these areas.

§6.4 Sampling Distributions and Estimators

Definition (Samling Distribution of a Statistic)

The **sampling distribution of a statistic** is the distribution of all values of the statistic when all possible samples of the same size *n* are taken from the population. (This is usually represented as a table, probability histogram, or formula.)

Definition (Samling Distribution of a Statistic)

The **sampling distribution of a statistic** is the distribution of all values of the statistic when all possible samples of the same size *n* are taken from the population. (This is usually represented as a table, probability histogram, or formula.)

A sampling distribution could be made for any statistic - e.g. a mean, variance, standard deviation, mean, median, etc.

§6.4 Example

Construct the sampling distribution for the mean number of children from population of 5 families with the following number of children where the sample size is 2.

Family	Number of Children
A	1
B	0
C	2
D	2
E	1

§6.4 Example

We need to find all the ways of choosing a pair from the 5 families. They are: AB(1), AC(3), AD(3), AE(2), BC(2), BD(2), BE(1), CD(4), CE(3), and DE(3). Each of these has a $\frac{1}{10}$ chance of being the sample. So we have

§6.4 Example

We need to find all the ways of choosing a pair from the 5 families. They are: AB(1), AC(3), AD(3), AE(2), BC(2), BD(2), BE(1), CD(4), CE(3), and DE(3). Each of these has a $\frac{1}{10}$ chance of being the sample. So we have

Sample Mean	Probability of Sample Mean
0.5	$\frac{1}{5}$
1	$\frac{3}{10}$
1.5	$\frac{2}{5}$
2	$\frac{1}{10}$

§6.4 Example

We need to find all the ways of choosing a pair from the 5 families. They are: AB(1), AC(3), AD(3), AE(2), BC(2), BD(2), BE(1), CD(4), CE(3), and DE(3). Each of these has a $\frac{1}{10}$ chance of being the sample. So we have

Sample Mean	Probability of Sample Mean
0.5	$\frac{1}{5}$
1	$\frac{3}{10}$
1.5	$\frac{2}{5}$
2	$\frac{1}{10}$

Notice, the mean of the sample mean given this probability distribution is 1.2, which matches the population mean! This makes mean an unbiased estimator!

§6.4 Example

You may repeat the above process for any parameter/statistic pair! Some statistics do not target the parameter they estimate, though.

§6.4 Example

You may repeat the above process for any parameter/statistic pair! Some statistics do not target the parameter they estimate, though. These do:

- mean
- variance
- proportions

We call these unbiased estimators.

§6.4 Example

You may repeat the above process for any parameter/statistic pair! Some statistics do not target the parameter they estimate, though. These do:

- mean
- variance
- proportions

We call these unbiased estimators. These do not:

- median
- range
- standard deviation

We call these biased estimators.

§6.5 The Central Limit Theorem

Theorem (Central Limit Theorem)

For all samples of the same size n with $n > 30$, the sampling distribution of $\overline{x}$ can be approximated by a normal distribution with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

§6.5 The Central Limit Theorem

Theorem (Central Limit Theorem)

For all samples of the same size n with $n > 30$, the sampling distribution of $\overline{x}$ can be approximated by a normal distribution with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

Note: You can use the CLT when the samples are coming from a normally distributed population even when the sample size is smaller than 30.

§6.5 Example

We believe the mean weight of a population of 2000 men is 160lbs and that the standard deviation for these weights is 30lbs. We take a sample of 36 of these men and find that their average weight is 171lbs. Does this agree with our assumption that the average weight of the population of men was 160lbs with a standard deviation of 30lbs?

§6.5 Example

We believe the mean weight of a population of 2000 men is 160lbs and that the standard deviation for these weights is 30lbs. We take a sample of 36 of these men and find that their average weight is 171lbs. Does this agree with our assumption that the average weight of the population of men was 160lbs with a standard deviation of 30lbs?

No! Why?