Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

September 22, 2019

Topics

Objectives

Objectives:

- Interpret the probability density curve of a continuous random variable.
- Recognize a normal distribution, and interpret its mean and standard deviation.
- Recognize the standard normal distribution.
- Find areas under normal curves using Table II.

Continuous Random Variables and the Normal Distribution (6.1, 6.2, 6.3)

Probability Density Curves

- If X is a **continuous** random variable, its **probability distribution** is represented by a ***density curve***:

Continuous Random Variables and the Normal Distribution (6.1, 6.2, 6.3)

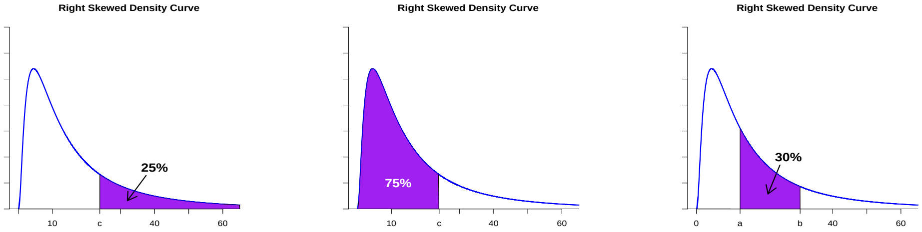
Probability Density Curves

- If X is a **continuous** random variable, its **probability distribution** is represented by a ***density curve***:
  - The curve can be thought of as an "idealized", smooth version of the **histogram** of a **large population.**

Continuous Random Variables and the Normal Distribution (6.1, 6.2, 6.3)

Probability Density Curves

- If X is a **continuous** random variable, its **probability distribution** is represented by a ***density curve***:
  - The curve can be thought of as an "idealized", smooth version of the **histogram** of a **large population.**
  - The **area** under the curve **above any interval of x values** represents the **proportion** of individuals whose x values lie in that interval (i.e. it represents the **probability** that a randomly selected individual's x value will fall in that interval).


Right Skewed Density Curve
25%
10
c
40
60
Right Skewed Density Curve
75%
10
c
40
60
Right Skewed Density Curve
30%
0
a
b
40
60

- (density curves, cont'd)
  - Every *density curve* has the following properties:
    1. The entire curve lies on or above the x-axis.
    2. The **total area** under the curve **equals 1.0** (representing 100% of the population).

The Normal Distribution

- By far the most widely applicable continuous probability distribution is the ***normal distribution***, which is represented by the ***normal density curve***.

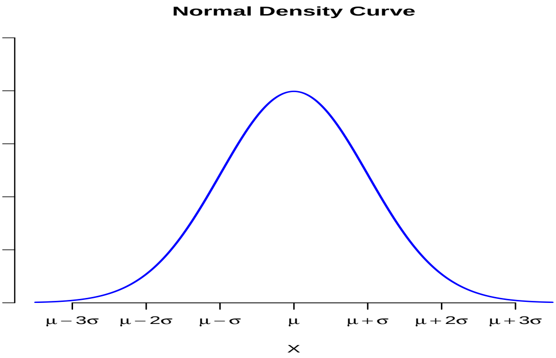


Figure: Normal density curve with mean μ and standard deviation σ.

- There are actually **many** normal curves, one for each pair of values for its **mean** μ and **standard deviation** σ.

- There are actually **many** normal curves, one for each pair of values for its **mean** μ and **standard deviation** σ.

  These **parameters** are interpreted as the **population mean** and **population standard deviation**.

- There are actually **many** normal curves, one for each pair of values for its **mean** μ and **standard deviation** σ.

  These **parameters** are interpreted as the **population mean** and **population standard deviation**.

- μ is the **center** of the distribution, and σ measures its **spread**.

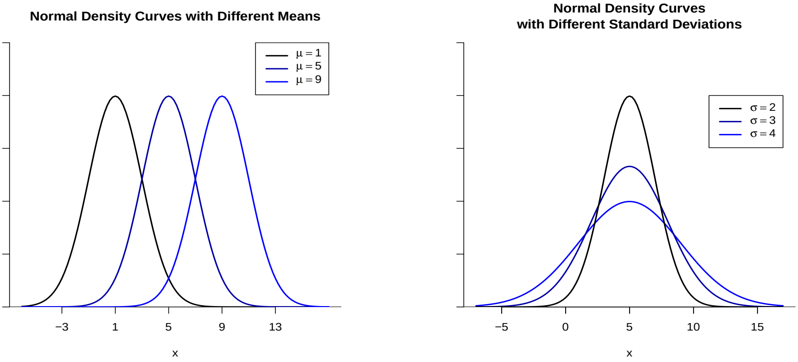


Figure: Normal distributions with different values of μ, but the same σ (left), and with the same μ, but different values of σ (right).

Example

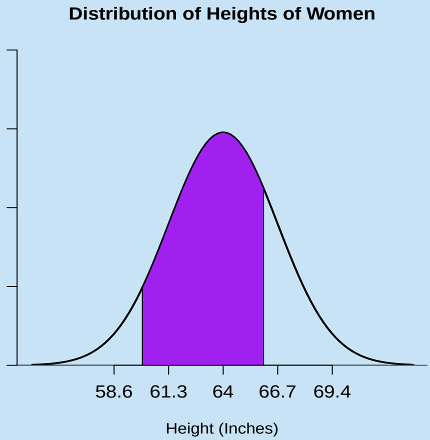
Heights of women aged 20 to 29 follow a **normal distribution** with **mean μ = 64** inches and **standard deviation σ = 2.7** inches.

Example

Heights of women aged 20 to 29 follow a **normal distribution** with **mean** μ **= 64** inches and **standard deviation** σ **= 2.7** inches.

This normal curve is shown on the next slide.

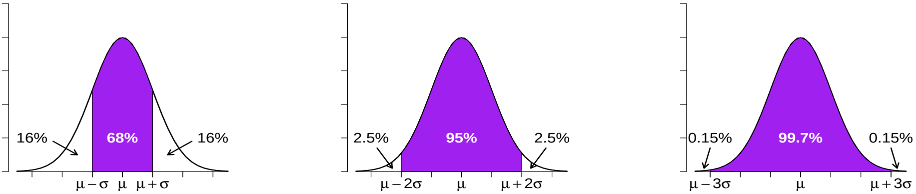
The **shaded area** corresponds to the **proportion** of women who are **between 60 and 66** inches tall.


Distribution of Heights of Women
58.6
61.3
64
66.7
69.4
Height (Inches)

The Empirical Rule (or 68-95-99.7 Rule) Revisited

- **The Empirical Rule (or 68-95-99.7 Rule)**: If a variable x in a population follows a **normal distribution** with mean μ and standard deviation σ, then

    1. **68%** of the population lies within **one σ of μ.**
    2. **95%** of the population lies within **two σ's of μ.**
    3. **99.7%** of the population lies within **three σ's of μ.**


16%
68%
16%
μ − σ
μ
μ + σ
2.5%
95%
2.5%
μ − 2σ
μ
μ + 2σ
0.15%
99.7%
0.15%
μ − 3σ
μ
μ + 3σ

The Standard Normal Curve

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

The Standard Normal Curve

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

  It's shown on the next slide.

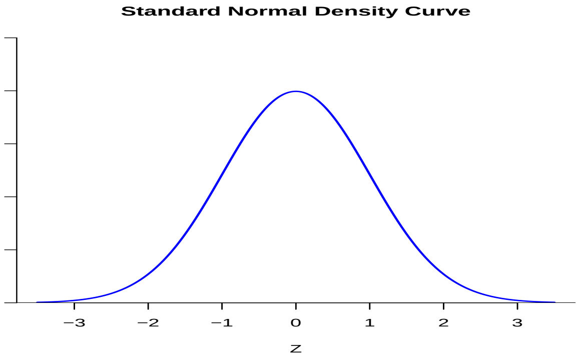


Figure: Standard normal density curve ($\mu = 0$ and $\sigma = 1$).

Finding Areas Under Normal Curves (6.1, 6.2, 6.3)

- Recall that the ***z*-score** for an individual's x value is:

$$z = \frac{x - \mu}{\sigma}.$$

Finding Areas Under Normal Curves (6.1, 6.2, 6.3)

- Recall that the ***z-score*** for an individual's x value is:

$$z = \frac{x - \mu}{\sigma}.$$

- In order to find **areas** under **any normal curve**, we'll need the facts on the next slide.

If a variable x follows a **normal distribution** with **mean** μ and **standard deviation** σ, then

1. The **standardized** version of that variable,

$$z = \frac{x - \mu}{\sigma},$$

   follows a **standard normal** distribution.

2. The **proportion** of individuals whose x **values** lie between two numbers a and b is the ***same*** as the **proportion** whose z**-scores** lie between

$$z = \frac{a - \mu}{\sigma} \qquad \text{and} \qquad z = \frac{b - \mu}{\sigma},$$

   and this proportion is the **area** between these two z values under the **standard normal curve**, which can be obtained from **Table II**.

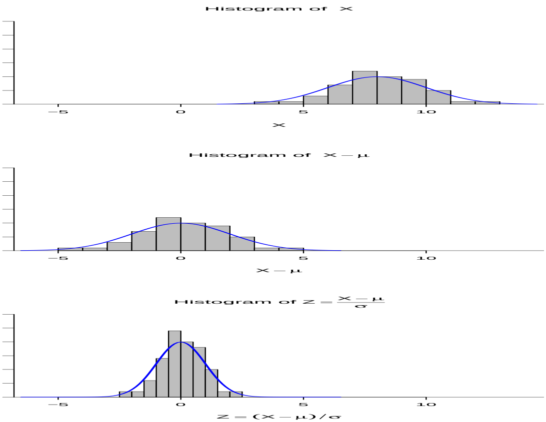


Figure: Histogram of a variable x following a normal distribution with $\mu = 8$ and $\sigma = 2$ (top); histogram after subtracting $\mu = 8$ from each x value (center); histogram of z, the standardized version of x (bottom). The bottom plot is a standard normal distribution.

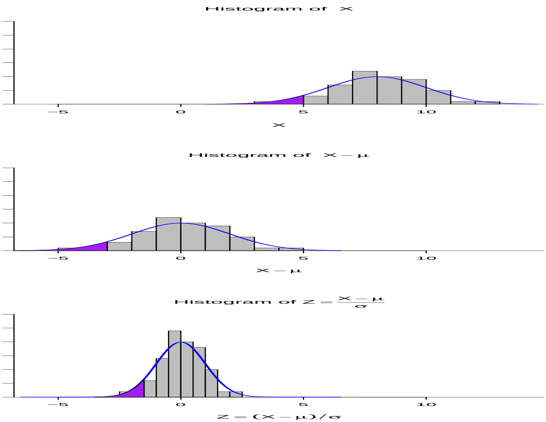


Figure: Histogram of a variable x following a normal distribution with $\mu = 8$ and $\sigma = 2$ (top); histogram after subtracting $\mu = 8$ from each x value (center); histogram of z, the standardized version of x (bottom). The shaded areas in the three plots are the same.

- Thus to find the **area** under ***any*** **normal curve**, use the following.

Finding Areas Under a Normal Curve: To find the area between two values a and b under a normal curve whose mean and standard deviation are μ and σ:

1. **Sketch the normal curve** associated with the variable x.
2. **Shade the area** of interest and mark its delimiting x value(s).

(cont'd):

3. **Find the z-score(s)** for the delimiting x values of Step 2. For example, if the shaded area of Step 2 is delimited by two values a and b, convert them to z-scores using:

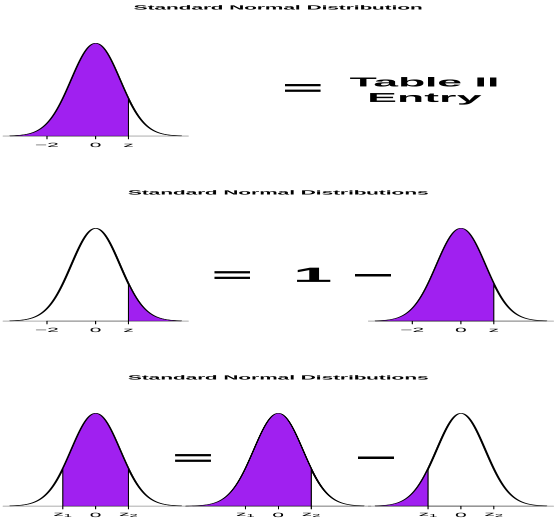
$$z = \frac{a - \mu}{\sigma} \qquad \text{and} \qquad z = \frac{b - \mu}{\sigma}$$

4. **Refer to Table II** to find the **area** under the ***standard normal*** curve delimited by the z-score(s) of Step 3.

- **Table II** only gives areas to the ***left*** of a z-score.

- **Table II** only gives areas to the ***left*** of a z-score.

  The figures below and the exercise ahead show how to find the area to the ***right*** of a z-score and ***between*** two z-scores.


Standard Normal Distribution
= Table II Entry
−2 0 z
Standard Normal Distributions
= 1 −
−2 0 z
−2 0 z
Standard Normal Distributions
= −
z1 0 z2
z1 0 z2
z1 0 z2

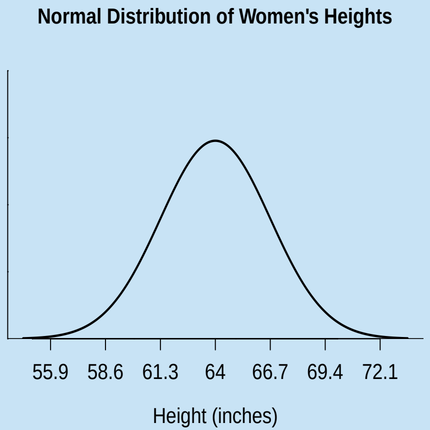
Example

Recall that **heights** of **women** aged 20 - 29 follow a **normal distribution** with **mean** μ **= 64** inches and **standard deviation** σ**= 2.7** inches.

Example

Recall that **heights** of **women** aged 20 - 29 follow a **normal distribution** with **mean** μ **= 64** inches and **standard deviation** σ**= 2.7** inches.

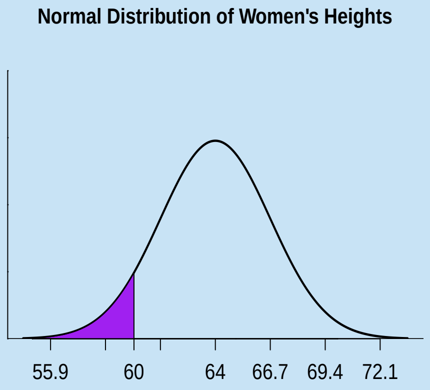
This is the **normal density curve** on the next slide.

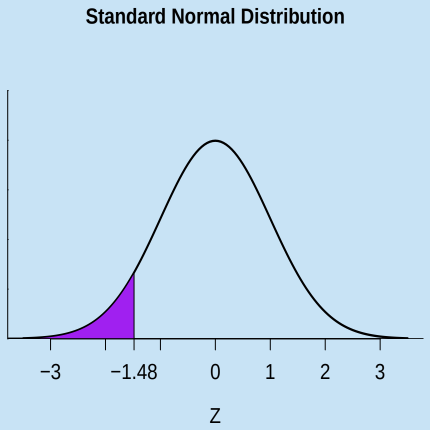

Normal Distribution of Women's Heights
55.9
58.6
61.3
64
66.7
69.4
72.1
Height (inches)

First we'll find the **percentage** of women that are **shorter** than **60** inches tall.

First we'll find the **percentage** of women that are **shorter** than **60** inches tall.

This is the **shaded area** under the density curves on the next slides.


Normal Distribution of Women's Heights
55.9
60
64
66.7
69.4
72.1


Standard Normal Distribution
−3
−1.48
0
1
2
3
Z

The z-**score** for a **60** inch tall woman is

$$z = \frac{60 - 64}{2.7} = \mathbf{-1.48},$$

The z**-score** for a **60** inch tall woman is

$$z = \frac{60 - 64}{2.7} = \mathbf{-1.48},$$

so the **percentage** of women **shorter** than **60** inches tall is the same as the **percentage** whose z**-scores** are **below -1.48**.

The z**-score** for a **60** inch tall woman is

$$z = \frac{60 - 64}{2.7} = \mathbf{-1.48},$$

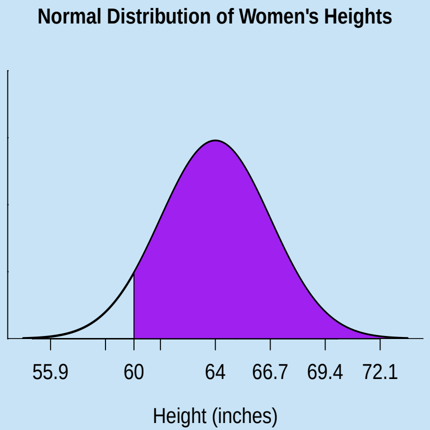
so the **percentage** of women **shorter** than **60** inches tall is the same as the **percentage** whose z**-scores** are **below -1.48.**

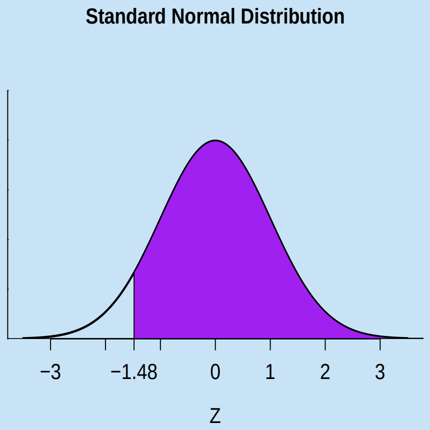
From **Table II**, a ***proportion*** **0.0694** is **shorter** than **60** inches, so the **percentage** is **6.94%.**

Now we'll find the **percentage** of women that are **taller** than **60** inches.

Now we'll find the **percentage** of women that are **taller** than **60** inches.

This is the **shaded area** under the density curves on the next slides.


Normal Distribution of Women's Heights
55.9
60
64
66.7
69.4
72.1
Height (inches)


Standard Normal Distribution
−3
−1.48
0
1
2
3
Z

Because the *total area* under a normal curve is **1.0** (corresponding to **100%** of women's heights), and a ***proportion*** **0.0694** is ***shorter*** than **60** inches (found previously), the ***proportion taller*** than **60** inches is

$$1 - 0.0694 = \mathbf{0.9306}.$$

Because the *total area* under a normal curve is **1.0** (corresponding to **100%** of women's heights), and a ***proportion* 0.0694** is ***shorter*** than **60** inches (found previously), the ***proportion taller*** than **60** inches is

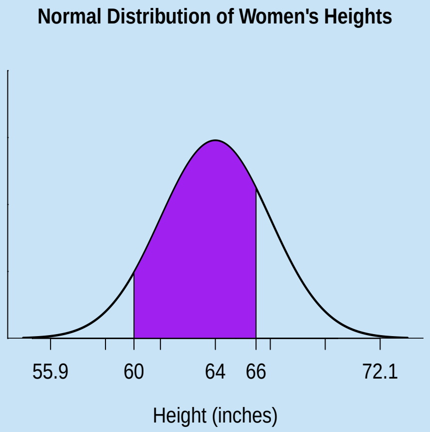
$$1 - 0.0694 = \mathbf{0.9306}.$$

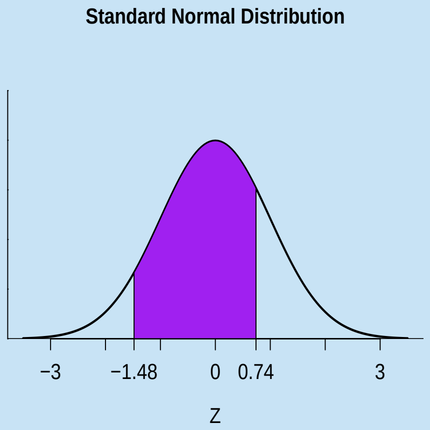
Thus the the **percentage** of women **taller** than **60** inches is **93.06%.**

Now we'll find the **percentage** of women that are **between 60** and **66** inches tall.

Now we'll find the **percentage** of women that are **between 60** and **66** inches tall.

This is the **shaded area** under the density curves on the next slides.


Normal Distribution of Women's Heights
55.9
60
64
66
72.1
Height (inches)


Standard Normal Distribution
−3
−1.48
0
0.74
3
Z

The z**-score** for a **66** inch tall woman is

$$z = \frac{66 - 64}{2.7} = \mathbf{0.74},$$

The z-**score** for a **66** inch tall woman is

$$z = \frac{66 - 64}{2.7} = \mathbf{0.74},$$

so the ***proportion*** of women ***shorter*** than **66** inches tall is the same as the ***proportion*** whose z-**scores** are **below 0.74**.

The z**-score** for a **66** inch tall woman is

$$z = \frac{66 - 64}{2.7} = \mathbf{0.74},$$

so the ***proportion*** of women ***shorter*** than **66** inches tall is the same as the ***proportion*** whose z**-scores** are **below 0.74.**

From **Table II**, this ***proportion*** is **0.7704**.

Thus, because a ***proportion*** **0.0694** is ***shorter*** than **60** inches (found previously), the ***proportion between*** **60** and **66** inches is

$$0.7704 - 0.0694 = \mathbf{0.7010}.$$

The z**-score** for a **66** inch tall woman is

$$z = \frac{66 - 64}{2.7} = \mathbf{0.74},$$

so the ***proportion*** of women ***shorter*** than **66** inches tall is the same as the ***proportion*** whose z**-scores** are **below 0.74.**

From **Table II**, this ***proportion*** is **0.7704**.

Thus, because a ***proportion*** **0.0694** is ***shorter*** than **60** inches (found previously), the ***proportion between*** **60** and **66** inches is

$$0.7704 - 0.0694 = \mathbf{0.7010}.$$

Therefore the **percentage between 60** and **66** inches is **70.1%.**

Exercise

Refer to the previous example.

a) **Sketch** the **density curve** and **shade the area** corresponding to the **percentage** of women **shorter** than **62** inches.

Exercise

Refer to the previous example.

a) **Sketch** the **density curve** and **shade the area** corresponding to the **percentage** of women **shorter** than **62** inches.

b) Find the **percentage** of women that are **shorter** than **62** inches.

Exercise

Refer to the previous example.

a) **Sketch** the **density curve** and **shade the area** corresponding to the **percentage** of women **shorter** than **62** inches.

b) Find the **percentage** of women that are **shorter** than **62** inches.

c) **Shade the area** under the density curve corresponding to the **percentage** of women that are **taller** than **67** inches.

Exercise

Refer to the previous example.

a) **Sketch** the **density curve** and **shade the area** corresponding to the **percentage** of women **shorter** than **62** inches.

b) Find the **percentage** of women that are **shorter** than **62** inches.

c) **Shade the area** under the density curve corresponding to the **percentage** of women that are **taller** than **67** inches.

d) Find the **percentage** of women that are **taller** than **67** inches.

e) **Shade the area** corresponding to the **percentage** of women that are **between 62** and **67** inches tall.

e) **Shade the area** corresponding to the **percentage** of women that are **between 62** and **67** inches tall.

f) Find the **percentage** of women that are **between 62** and **67** inches tall.