Sullivan Section 7.1

Gene Quinn

The Normal Distribution

Many continuous random variables have relative frequency histograms that are symmetric and bell-shaped.

The Normal Distribution

Many continuous random variables have relative frequency histograms that are symmetric and bell-shaped.

If you were to draw a smooth curve through the tops of the bars in the histogram, that curve would have a shape that is called the **normal curve**.

The Normal Distribution

Many continuous random variables have relative frequency histograms that are symmetric and bell-shaped.

If you were to draw a smooth curve through the tops of the bars in the histogram, that curve would have a shape that is called the **normal curve**.

The normal distribution plays a central role in statistics.

The reason for this is that, even if the population we are working with does not follow a normal distribution, important statistics like means **will** have a distribution that is approximately normal.

The Normal Distribution

Later in the semester, we will discuss the important theoretical result that establishes this fact, which is known as the *central limit theorem*.

The Normal Distribution

Later in the semester, we will discuss the important theoretical result that establishes this fact, which is known as the *central limit theorem*.

The central limit theorem makes life much easier for the practicing statistician, because instead of developing dozens of different procedures for different underlying population distributions, in many circumstances we can invoke the central limit theorem and treat the data as if the underlying distribution was actually normal.

The Normal Distribution

Definition: A continuous random variable is said to be **normally distributed** or said to have a **normal probability distribution** if its relative frequency histogram has the shape of a normal curve (i.e., bell-shaped and symmetric).

The Normal Distribution

Definition: A continuous random variable is said to be **normally distributed** or said to have a **normal probability distribution** if its relative frequency histogram has the shape of a normal curve (i.e., bell-shaped and symmetric).

As the number of subdivisions in the histogram becomes larger and larger, the tops of the bars begin to resemble a smooth curve rather than a series of steps.

The Normal Distribution

This theoretical smooth curve is known as a **probability density function**.

The word "density" implies that because values in a section of the real line where the curve is far above zero are more likely to occur than values where the curve is close to zero.

The Normal Distribution

This theoretical smooth curve is known as a **probability density function**.

The word "density" implies that because values in a section of the real line where the curve is far above zero are more likely to occur than values where the curve is close to zero.

So, roughly speaking, the values in the population are more "dense" where the curve has a value much larger than zero than they are where it is close to zero.

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1
- The area under the curve to the *right* of μ is $1/2$.

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1
- The area under the curve to the *right* of μ is $1/2$.
- The area under the curve to the *left* of μ is $1/2$.

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1
- The area under the curve to the *right* of μ is $1/2$.
- The area under the curve to the *left* of μ is $1/2$.
- For positive values of x, as x increases the curve approaches zero.

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1
- The area under the curve to the *right* of μ is $1/2$.
- The area under the curve to the *left* of μ is $1/2$.
- For positive values of x, as x increases the curve approaches zero.
- For negative values of x, as x moves to the left, the curve approaches zero.

The Normal Distribution

The curve associated with the normal probability density has the following properties:

- It is symmetric about its mean μ
- Its highest point occurs at $x = \mu$
- The area under the curve is 1
- The area under the curve to the *right* of μ is $1/2$.
- The area under the curve to the *left* of μ is $1/2$.
- For positive values of x, as x increases the curve approaches zero.
- For negative values of x, as x moves to the left, the curve approaches zero.
- The Empirical rule applies.

The Standard Normal Distribution

We have encountered the concept of a z-score or **standardized** score for a data value.

For a normal random variable X, we can define a new random variable Z by the formula:

$$Z = \frac{X - \mu}{\sigma}$$

The Standard Normal Distribution

We have encountered the concept of a z-score or **standardized** score for a data value.

For a normal random variable X, we can define a new random variable Z by the formula:

$$Z = \frac{X - \mu}{\sigma}$$

The random variable Z is said to have the **standard normal distribution**.

The Standard Normal Distribution

We have encountered the concept of a z-score or **standardized** score for a data value.

For a normal random variable X, we can define a new random variable Z by the formula:

$$Z = \frac{X - \mu}{\sigma}$$

The random variable Z is said to have the **standard normal distribution**.

The standard normal distribution always has its mean $\mu = 0$ and its standard deviation $\sigma = 1$.

The Standard Normal Distribution

When you see a table of the normal distribution in the appendix of a text, it is almost always the table of the standard normal distribution.