Chapter 7 The Normal Probability Distribution

7.1 Properties of the Normal Distribution

Objectives

1. Use the uniform probability distribution
2. Graph a normal curve
3. State the properties of the normal curve
4. Explain the role of area in the normal density function

1 Use the Uniform Probability Distribution

EXAMPLE Illustrating the Uniform Distribution

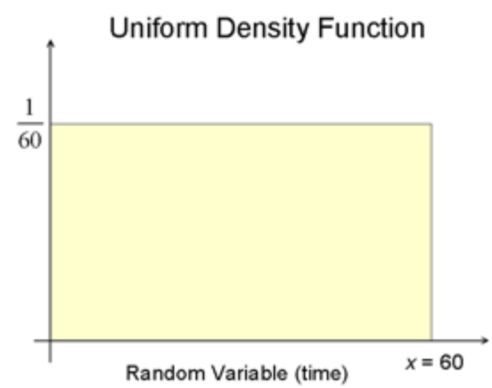
Suppose that United Parcel Service is supposed to deliver a package to your front door and the arrival time is somewhere between 10 am and 11 am. Let the random variable X represent the time from 10 am when the delivery is supposed to take place. The delivery could be at 10 am ($x = 0$) or at 11 am ($x = 60$) with all 1-minute interval of times between $x = 0$ and $x = 60$ equally likely. That is to say your package is just as likely to arrive between 10:15 and 10:16 as it is to arrive between 10:40 and 10:41. The random variable X can be any value in the interval from 0 to 60, that is, $0 \leq X \leq 60$. Because any two intervals of equal length between 0 and 60, inclusive, are equally likely, the random variable X is said to follow a **uniform probability distribution**.

■

A **probability density function (pdf)** is an equation used to compute probabilities of continuous random variables. It must satisfy the following two properties:

1. The total area under the graph of the equation over all possible values of the random variable must equal 1.
2. The height of the graph of the equation must be greater than or equal to 0 for all possible values of the random variable.

The graph illustrates the properties for the "time" example. Notice the area of the rectangle is one and the graph is greater than or equal to zero for all x between 0 and 60, inclusive. Values of the random variable X less than 0 or greater than 60 are impossible, thus the function value must be zero for X less than 0 or greater than 60.



The area under the graph of a density function over an interval represents the probability of observing a value of the random variable in that interval.

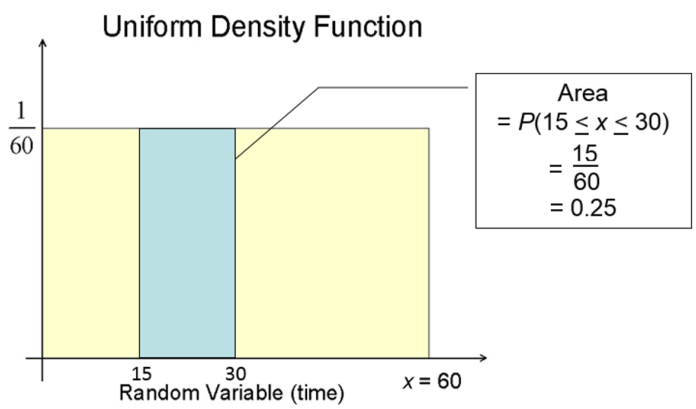
EXAMPLE Area as a Probability

Consider the UPS example just presented.
(a) What is the probability your package arrives between 10:15 am and 10:30 am?

(b) Suppose it is 10 am. There is a 30% probability your package will arrive within the next _______ minutes.

SOLUTION

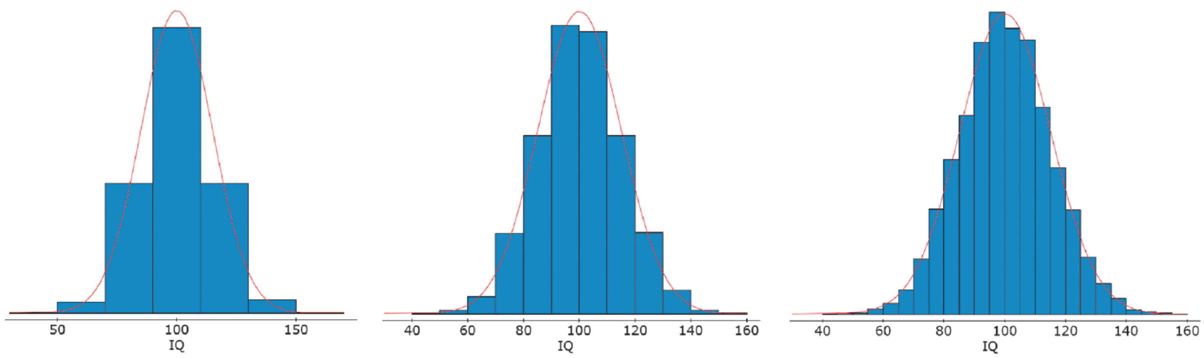
(a)



(b)

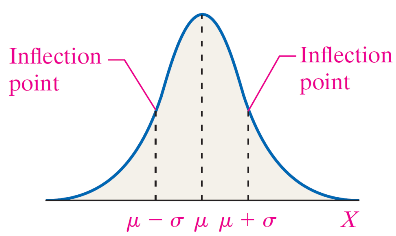
2 Graph a Normal Curve

Not all continuous random variables follow the uniform probability distribution. Some continuous random variables, such as birth weights or IQ scores, follow a bell-shaped distribution. Consider the following histograms. Notice as the class width decreases, the red curve closely approximates the histogram. The red curve is a **model** called the normal curve.



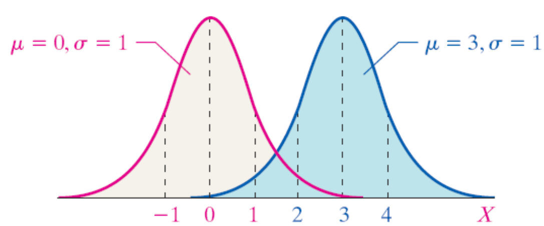
(a) (b) (c)

A continuous random variable is **normally distributed**, or has a **normal probability distribution**, if its relative frequency histogram has the shape of a normal curve.

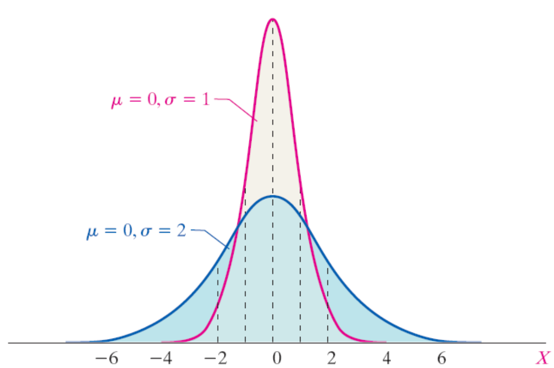


Exploration Go to StatCrunch. Select Applets > Distribution Demos. Select the Normal Curve. Use the sliders to see the role of the population mean and population standard deviation in the graph of the normal curve.

The role of the population mean, μ, on the graph of the normal curve:



The role of the population standard deviation, σ, on the graph of the normal curve:



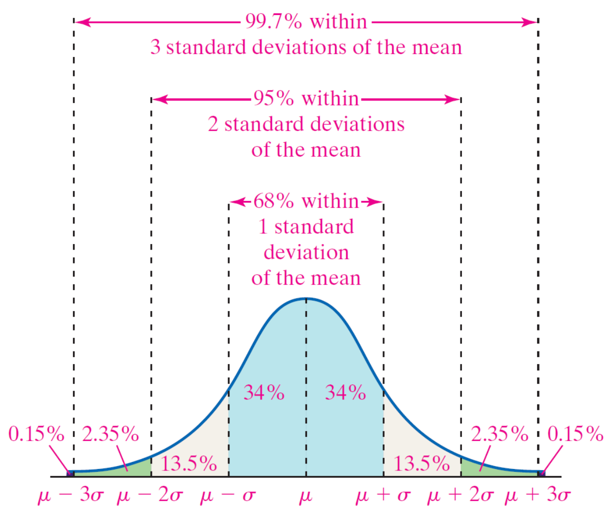
3 State the Properties of the Normal Curve

Properties of the Normal Density Curve

1. The normal curve is symmetric about its mean, μ.
2. Because mean = median = mode, the normal curve has a single peak and the highest point occurs at $x = \mu$.
3. The normal curve has inflection points at $\mu - \sigma$ and $\mu + \sigma$.
4. The area under the normal curve is 1.
5. The area under the normal curve to the right of μ equals the area under the curve to the left of μ, which equals $\frac{1}{2}$.
6. As x increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis.

7. The Empirical Rule:
   - Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$;
   - approximately 95% of the area is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$;
   - approximately 99.7% of the area is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$.

Normal Distribution



4 Explain the Role of Area in the Normal Density Function

EXAMPLE A Normal Random Variable

Open the data set "HomeRuns2017" from the SullyStats group in StatCrunch. The data represents the measurement of variables for every home run hit during the 2017 Major League baseball season.

(a) Draw a histogram of the variable "Distance", which represents the distance (in feet) the home run traveled. Use a lower class limit of the first class of 300 and a class width of 10. Do you believe the variable "Distance" can be modeled by a normal curve?
(b) Find the mean and standard deviation of the variable "Distance". Note that this is population data.
(c) Draw a normal curve on the histogram using the mean and standard deviation found in part (b). Is the area under the normal curve close to the areas of the rectangles in the histogram? NOTE: Technically, we need to create something called a density function to compute area under a curve.

Area under a Normal Curve

Suppose that a random variable X is normally distributed with mean μ and standard deviation σ. The area under the normal curve for any interval of values of the random variable X represents either

- the proportion of the population with the characteristic described by the interval of values or
- the probability that a randomly selected individual from the population will have the characteristic described by the interval of values.

EXAMPLE Interpreting the Area Under a Normal Curve

The distance (in feet) a home run traveled in 2017 approximately normally distributed with mean $\mu =$ 400.3 feet and standard deviation $\sigma = 25.6$ feet.

(a) Draw a normal curve with the parameters labeled.

(b) Shade the area under the normal curve to the right of $x = 450$ feet.

(c) Suppose that the area under the normal curve to the right of $x = 450$ feet is 0.026. Provide two interpretations of this result.

(d) Among all the home runs hit, what proportion exceeded 450 feet?

7.2 Application of the Normal Distribution

Objectives

1. Find and interpret the area under a normal curve

2. Find the value of a normal random variable

1 Find and Interpret the Area Under a Normal Curve

Now the question is, "How do I find the area under the normal curve?". There are two approaches – use of a table and Z-scores, or technology. We will focus on the use of technology (namely, StatCrunch), to find the area under a normal curve.

EXAMPLE Finding the Area Under a Normal Curve

The diameters of ball bearings produced at a factory are approximately normally distributed. Suppose the mean diameter is 1 centimeter (cm) and the standard deviation is 0.002 cm. The product specifications require that the diameter of each ball bearing be between 0.995 and 1.005 cm.
a. What proportion of ball bearings can be expected to have a diameter under 0.999 cm?
b. What is the probability a randomly selected ball bearing will have a diameter over 1.0055 cm?
c. What proportion of ball bearings can be expected to have a diameter between 0.995 and 1.005 cm? That is, what proportion of ball bearings can be expected to meet the specifications?

Because the area under the normal curve represents a proportion, we can also use the area to find percentile ranks of scores. Recall that the kth percentile divides the lower k% of a data set from the upper $(100 - k)$%.

EXAMPLE Using the Normal Model to Find Percentiles

The distance (in feet) a home run traveled in 2017 approximately normally distributed with mean $\mu = 400.3$ feet and standard deviation $\sigma = 25.6$ feet. Find the percentile rank of a home run that traveled 380 feet.

Some Cautionary Thoughts

The normal curve extends indefinitely in both directions. For this reason, there is no range of values of a normal random variable for which the area under the curve is 1. For example, if asked to find the area under a normal curve to the left of $x = 40$ with $\mu = 15$ and $\sigma = 2$, StatCrunch (as well as other software and calculators) will state the area is 1, because it can only compute a limited number of decimal places. However, the area under the curve to the left of $x = 40$ is not 1; it is some value slightly less than 1. So we will follow the practice of reporting such areas as >0.9999. Similarly, if software reports an area of 0, we will report the area as <0.0001.

2 Find the Value of a Normal Random Variable

Now we are going to reverse the process. Rather than give the value of a normal random variable and use the model to find a probability, proportion, or percentile rank, we are going to give the probability, proportion, or percentile rank, and find the corresponding value of the normal random variable, X.

EXAMPLE Finding the Value of a Normal Random Variable

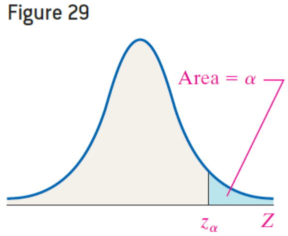
The distance (in feet) a home run traveled in 2017 approximately normally distributed with mean $\mu = 400.3$ feet and standard deviation $\sigma = 25.6$ feet. Find the distance of a home run that is at the 90^{th} percentile.

EXAMPLE Finding the Value of a Normal Random Variable

The diameters of ball bearings produced at a factory are approximately normally distributed. Suppose the mean diameter is 1 centimeter (cm) and the standard deviation is 0.002 cm. Determine the diameters of ball bearings that make up the middle 99% of all diameters.

Important Notation for the Future

Figure 29



The notation z_α (pronounced "z sub alpha") is the z-score such that the area under the standard normal curve to the right of z_α is α. Figure 29 illustrates the notation.

EXAMPLE Finding the Value of z_α

Find the value of $z_{0.25}$.

7.3 Assessing Normality

Objective

1 Use Normal Probability Plots to Assess Normality

1 Use Normal Probability Plots to Assess Normality

Suppose that we obtain a simple random sample from a population whose distribution is unknown. Many of the statistical tests that we perform on small data sets (sample size less than 30) require that the population from which the sample is drawn be normally distributed. Up to this point, we have said that a random variable *X* is normally distributed, or at least approximately normal, provided the histogram of the data is symmetric and bell-shaped. This method works well for large data sets, but the shape of a histogram drawn from a small sample of observations does not always accurately represent the shape of the population. For this reason, we need additional methods for assessing the normality of a random variable *X* when we are looking at sample data.

A **normal probability plot** plots observed data versus normal scores. A **normal score** is the expected *Z*-score of the data value if the distribution of the random variable is normal. The expected *Z*-score of an observed value will depend upon the number of observations in the data set.

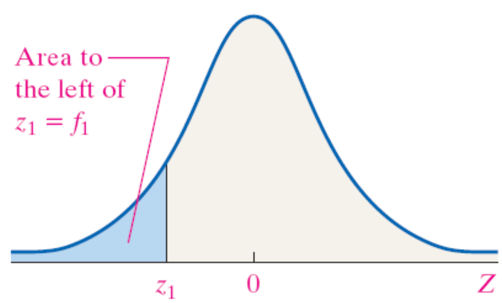
Drawing a Normal Probability Plot

Step 1 Arrange the data in ascending order.

Step 2 Compute $f_i = \frac{i - 0.375}{n + 0.25}$,* where *i* is the index (the position of the data value in the ordered list) and *n* is the number of observations. The expected proportion of observations less than or equal to the *i*th data value is f_i.

Step 3 Find the *z*-score corresponding to f_i from Table V.

Step 4 Plot the observed values on the horizontal axis and the corresponding expected *z*-scores on the vertical axis.



The idea behind finding the expected *z*-score is that, if the data comes from normally distributed population, we could predict the area to the left of each of the data value. The value of f_i represents the expected area left of the i^{th} observation when the data come from a population that is normally distributed. For example, f_1 is the expected area to the left of the smallest data value, f_2 is the expected area to the left of the second smallest data value, and so on.

If sample data is taken from a population that is normally distributed, a normal probability plot of the actual values versus the expected *Z*-scores will be approximately linear.

It is difficult to determine whether a normal probability plot is "linear enough." Basically, if the linear correlation coefficient between the observed values and expected *z*-scores is greater than the critical value found in Table VI in Appendix A, then it is reasonable to conclude that the data could come from a population that is normally distributed. Normal probability plots are typically drawn using graphing calculators or statistical software.

EXAMPLE Interpreting a Normal Probability Plot

The following data represent the time between eruptions (in seconds) for a random sample of 15 eruptions at the Old Faithful Geyser in California. Is there reason to believe the time between eruptions is normally distributed?

728	678	723	735	735
730	722	708	708	714
726	716	736	736	719

EXAMPLE Interpreting a Normal Probability Plot

Suppose that seventeen randomly selected workers at a detergent factory were tested for exposure to a Bacillus subtillis enzyme by measuring the ratio of forced expiratory volume (FEV) to vital capacity (VC). NOTE: FEV is the maximum volume of air a person can exhale in one second; VC is the maximum volume of air that a person can exhale after taking a deep breath. Is it reasonable to conclude that the FEV to VC (FEV/VC) ratio is normally distributed?

Source: Shore, N.S.; Greene R.; and Kazemi, H. "Lung Dysfunction in Workers Exposed to *Bacillus subtillis* Enzyme," *Environmental Research*, 4 (1971), pp. 512 - 519.

0.61	0.7	0.76	0.84
0.63	0.72	0.78	0.85
0.64	0.73	0.82	0.85
0.67	0.74	0.83	0.87
0.88			