Normal Distribution

cK-12 Study Guides
MADE BY STUDENTS!

Big Picture

The normal distribution is extremely important in statistics. It is a perfectly symmetric, mound-shaped distribution. This pattern appears so often that it is "normal" to see it in data for many real-life phenomena.

Key Terms

Normal Distribution: A continuous probability distribution that has a symmetric bell-shaped curve with a single peak.

Standard Normal Distribution: A normal distribution with $\mu = 0$ and $\sigma = 1$.

Random Variable: Numerical data observed or measured in an experiment.

Continuous Random Variable: A random variable that can take on a countless number of values in an interval.

Mean (also called the arithmetic mean)**:** The numerical balancing point of the data set. Calculated by adding all the data values and dividing the sum by the total number of data points.

Standard Deviation: A measure of how data points deviate from the mean.

Inflection Point: A point where the curve changes concavity (from concave up to concave down, or concave down to concave up).

Empirical Rule: States what percentages of data in a normal distribution lies within 1, 2, and 3 standard deviations of the mean.

Density Curve: A curve where the area under the curve equals exactly one.

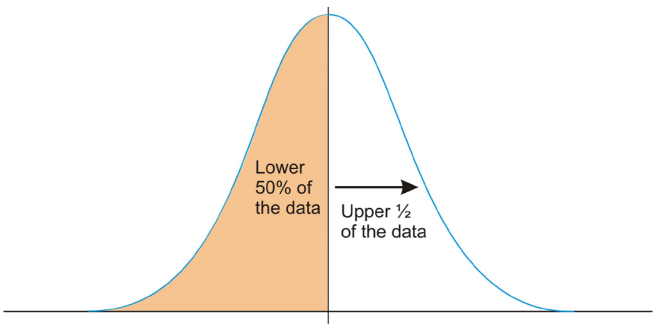
Normal Density Curve: A normal curve where the area under the curve is equal to exactly one.

z **-Score:** A measure of the number of standard deviations a particular data point is away from the mean.

Characteristics

All **normal distributions** have the same shape.

- Perfectly symmetric, mound-shaped distribution
- Also known as normal curve, or bell curve
- Describes **continuous random variable**
- Distribution continues infinitely in both directions – both sides of the curve are approaching 0, but they never reach 0



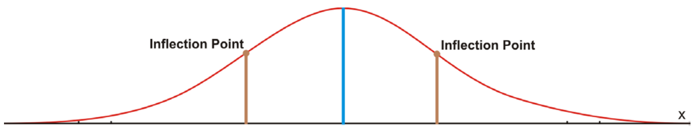
Normal Distribution divided into equal halves by the center

Center:

- Located at the highest point over the **mean** μ
- Mean, median, and mode are all equal
- Splits the data into two equal parts

Spread:

- Measured with **standard deviation** σ
- Larger standard deviations mean that the data is spread farther from the center



Inflection point:

- Curve changes shape at the inflection points – in other words, the curve changes concavity
  - A curve that is concave **u**p looks like a **u**-shape
  - A curve that is concave dow**n** looks like a **n**-shape
- The two inflection points occur ± 1 standard deviation away from the mean (μ - σ and $\mu + \sigma$)

Notes

This guide was created by Lizhi Fan and Jin Yu. To learn more about the student authors, visit http://www.ck12.org/about/about-us/team/interns.

Normal Distribution cont.

Empirical Rule

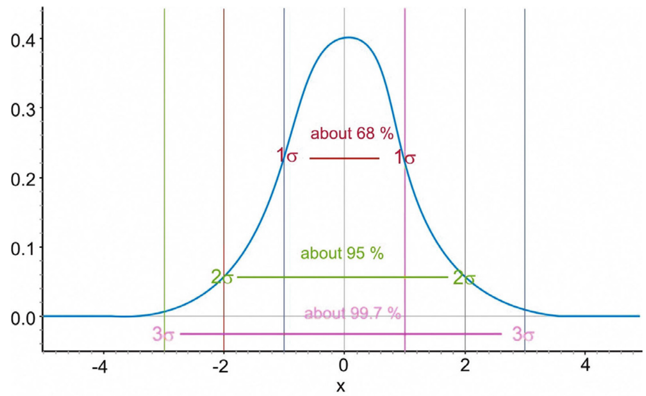
The standard deviation is a measure of the "typical" distance away from mean. How much data is actually within one standard deviation?

- The space under the whole curve contains 100% of the data

The **empirical rule** gives these values:

- The percent of data within 1 standard deviation away from the mean is 68%
- The percent of data within 2 standard deviations away from the mean is 95%
- The percent of data within 3 standard deviations away from the mean is 99.7%

This fact should be memorized as it helps with many calculations.



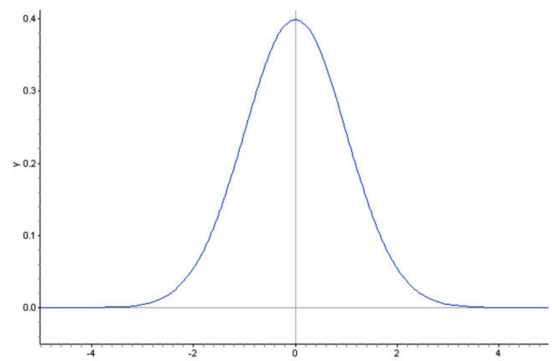
The Empirical Rule

Standard Normal Curve

This graph is an example of a **standard normal curve** where $\mu = 0$ and $\sigma = 1$.

- This means that the value on the x-axis equals the number of standard deviations from the mean.
- The inflection points are at $\pm\sigma$.

A **density curve** is a curve where the area underneath the curve (and above the x-axis) is equal to exactly one. The standard normal curve is also a **normal density curve**, meaning that the area under the normal distribution curve (and above the x-axis) equals exactly one.



z-Scores

***z*-scores** are useful for comparing data from different data sets and different normal distributions. To calculate a z-score, take the deviation and divide it by the standard deviation. The difference between a data value and the mean is called the deviation.

$$z = \frac{\text{Deviation}}{\text{Standard Deviation}} = \frac{x - \mu}{\sigma}$$

σ is always positive, so if the z-score is negative, x must be below the mean.

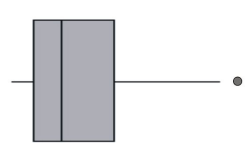
Assessing Normality

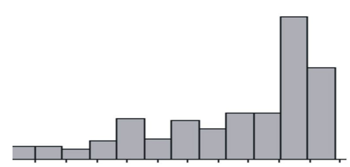
Many times, you can tell a normal distribution does not describe a data set just by looking at the graphs.

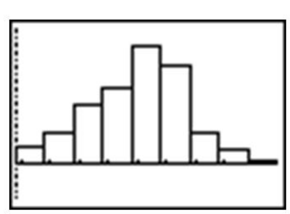
A skewed left distribution will have its "tail" on the left and most of the data points on the right. Skewed right means that the longer tail is on the right, with most of the data gathered on the left.

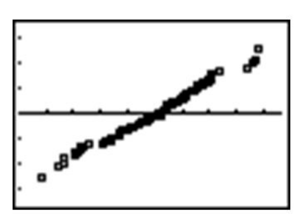
A normal probability plot can also be used to determine normality.

- Find the z-scores for the data set.
- Plot the z-scores against the actual values of the data.
- The closer the normal probability plot is to being linear, the more closely the data set approximates a normal distribution.


Skewed right distribution with outliers


Skewed left distribution





Normal Distribution cont.

Standardizing

Any normal distributions can be standardized so that $\mu = 0$ and $\sigma = 1$.

1. Recenter the curve by shifting it so that $\mu = 0$. All the values on the x-axis are moved to $x - \mu$.
2. Rescale the curve so that the horizontal axis displays the z-scores instead of x.

Calculating Density Curve Areas

The area under the density curve gives the probability that an event will occur. One way to calculate the area between any interval along the horizontal axis is to use a z-table.

- If needed, standardize the curve BEFORE calculating the area.
- There are several types of z-table. Make sure to read the table carefully before doing any calculations. Are the values calculated for the area to the left of a specific z-value? Or are the values calculated for the area between $z = 0$ and another z value?
- Remember that the area under the entire curve is equal to 1.

Approximation of Binomial Distribution

The binomial distribution is a discrete probability distribution of the number of successes of k independent trials where there are only two possible outcomes. In some cases, it is easier to approximate the binomial distribution with the normal distribution.

- A binomial experiment consists of n independent, identical trials
- Only two possible outcomes for each trial (success or failure)
- Probability of success remains constant throughout the whole experiment (denoted by p)
  - The probability of failure is denoted by $q = 1 - p$
  - Each trial results in either success or failure, so $p + q = 1$

If $np > 5$ and $nq > 5$, then k has a binomial distribution that can be approximated by a normal distribution. This normal distribution has:

- $\mu = np$
- $\sigma = \sqrt{npq}$

Graphing Calculator

In a graphing calculator, we can use a command to find the normal distribution. The command for the normal distribution is: normalpdf(x, μ, σ). x is the value, μ is the mean, σ is the standard deviation. This will give us the area under the normal curve at this value.

There is another similar equation called normalcdf, which requires us to plug in two values for x: one low and one high. This will give us the area under the normal curve between those two values.

If you can't find the commands, check the manual for your graphing calculator. For the TI-83/TI-84, the commands are found by pressing [2ND][DISTR].