

9.1: The Normal Distribution

Learning Objectives

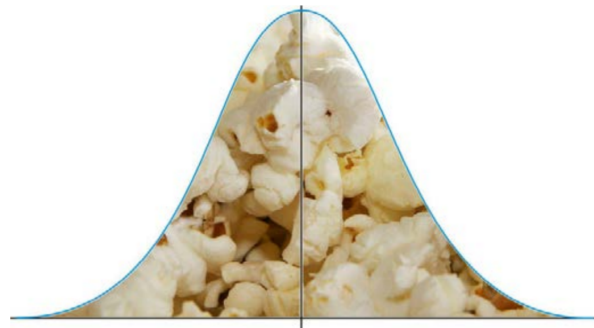
- Identify the characteristics of a normal distribution
- Apply the Empirical Rule (68-95-99.7 Rule) for normal distributions
- Calculate and interpret z-scores

Introduction

Most high schools have a set amount of time in between classes during which students must get to their next class. If you were to stand at the door of your math class and watch the students coming in, think about how the students would enter. Usually, one or two students enter early, then more students come in, then a large group of students enter, and finally, the number of students entering decreases again, with one or two students barely making it on time, or perhaps even coming in late!

Now consider this. Have you ever popped popcorn in a microwave? Think about what happens in terms of the rate at which the kernels pop. For the first few minutes, nothing happens, and then, after a while, a few kernels start popping. This rate increases to the point at which you hear most of the kernels popping, and then it gradually decreases again until just a kernel or two pop.

Here's something else to think about. Try measuring the height, shoe size, or the width of the hands of the students in your class. In most situations, you will probably find that there are a couple of students with very low measurements and a couple with very high measurements, with the majority of students centered on a particular value.



All of these examples show a typical pattern that seems to be a part of many real-life phenomena. In statistics, because this pattern is so pervasive, it seems to fit to call it normal, or more formally, the normal distribution. The normal distribution is an extremely important concept, because it occurs so often in the data we collect from the natural world, as well as in many of the more theoretical ideas that are the foundation of statistics. This chapter explores the details of the normal distribution.

The Characteristics of a Normal Distribution

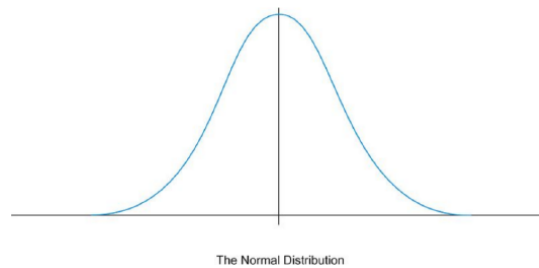
Normal Distribution

The **normal distribution** is a commonly occurring statistical distribution whose bell shape is determined by the mean and standard deviation of the distribution.

The normal distribution is a continuous probability distribution. The distribution is represented by a smoothed-out histogram. If you recall from the probability chapter, the sum of the probabilities of all outcomes in the sample space is 1. This is also true for the normal distribution: the total area under the normal distribution curve, which represents the probabilities of all outcomes in the sample space, is always 1.

Shape

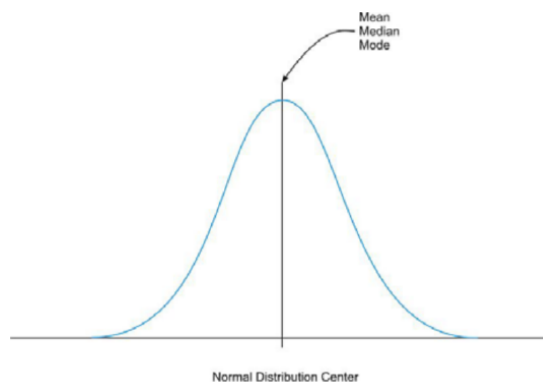
When graphing the data from each of the examples in the introduction, the distributions from each of these situations would be mound or bell-shaped and mostly symmetric. A normal distribution is a perfectly symmetric, bell-shaped distribution. It is commonly referred to as a normal curve, or bell curve.



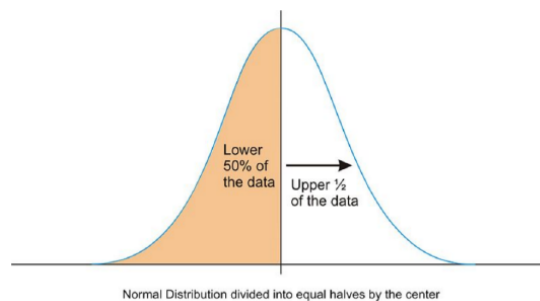
Because so many real data sets closely approximate a normal distribution, we can use the idealized normal curve to learn a great deal about such data. With a practical data collection, the distribution will never be exactly symmetric, so just like situations involving probability, a true normal distribution only results from an infinite collection of data.

Center

Due to the exact symmetry of a normal curve, the center of a normal distribution, or a data set that approximates a normal distribution, is located at the highest point of the distribution, and all the statistical measures of center we have already studied (the mean, median, and mode) are equal. The normal distribution is symmetric about the mean, median and mode. In statistics, the mean is represented by the Greek letter mu μ .

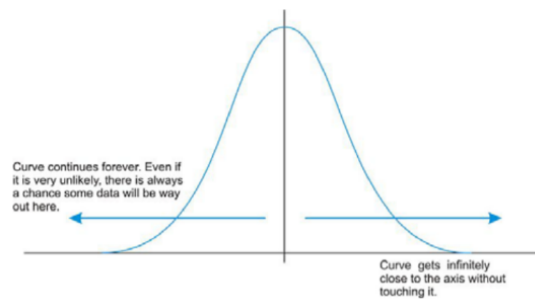


It is also important to realize that this center peak divides the data into two equal parts since it represents the median.

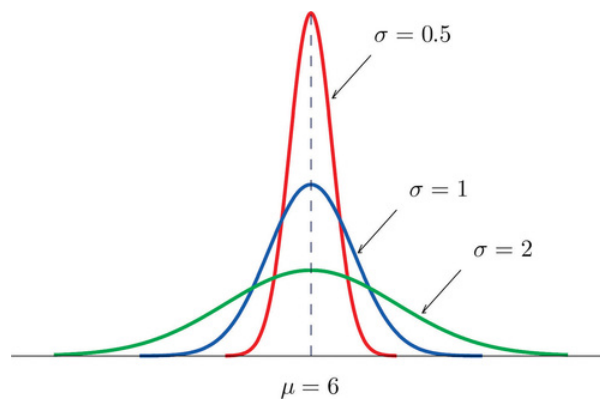


Spread

Let's go back to our popcorn example. The bag advertises a certain time, beyond which you risk burning the popcorn. From experience, the manufacturers know when most of the popcorn will stop popping, but there is still a chance that there are those rare kernels that will require more (or less) time to pop than the time advertised by the manufacturer. The directions usually tell you to stop when the time between popping is a few seconds, but aren't you tempted to keep going so you don't end up with a bag full of un-popped kernels? Because this is a real, and not theoretical, situation, there will be a time when the popcorn will stop popping and start burning, but there is always a chance, no matter how small, that one more kernel will pop if you keep the microwave going. In an idealized normal distribution, the distribution continues infinitely in both directions. This means that the curve never actually touches the x -axis. The curve approaches the x -axis, but technically never touches it.



Because of this infinite spread, the range would not be a useful statistical measure of spread. The most common way to measure the spread of a normal distribution is with the standard deviation, or the typical distance away from the mean. Because of the symmetry of a normal distribution, the standard deviation, which is denoted by the Greek letter sigma σ in statistics, indicates how far away from the maximum peak the data will be, on average. The value of σ determines whether the bell curve is tall and thin or short and squat, subject always to the condition that the total area under the curve be equal to 1. This is shown below, where we have arbitrarily chosen to center the curves at $\mu = 6$.



The distribution with $\sigma = 0.5$ pictured above has the smallest standard deviation, and more of the data are heavily concentrated around the mean than in the distribution with $\sigma = 1$ or $\sigma = 2$. Also, in the distribution with $\sigma = 0.5$, there are fewer data values at the extremes than in the other 2 distributions. Because the distribution with $\sigma = 2$ has the largest standard deviation, the data are spread farther from the mean value, with more of the data appearing in the tails.

The Empirical Rule for Normal Distributions

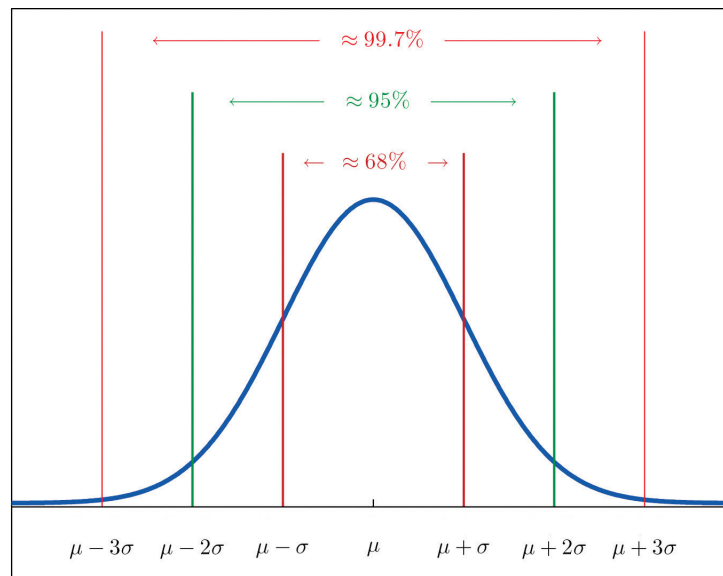
Since all normal distributions have the same shape and the total area under the curve is 1 or 100%, it turns out that the amount of data between the mean and the standard deviations for any approximately normal distribution is consistent.

Empirical Rule

For data sets that are approximately bell-shaped:

- approximately 68% of the data lie within one standard deviation of the mean, that is, in the interval with endpoints $\mu \pm \sigma$
- approximately 95% of the data lie within two standard deviations of the mean, that is, in the interval with endpoints $\mu \pm 2\sigma$
- approximately 99.7% of the data lie within three standard deviations of the mean, that is, in the interval with endpoints $\mu \pm 3\sigma$

The Empirical Rule is also referred to as the 68-95-99.7 Rule. The figure below illustrates the Empirical Rule.



Two key points in regard to the Empirical Rule are that the data distribution must be approximately bell-shaped and that the percentages are only approximately true. The Empirical Rule does not apply to data sets that are not bell-shaped, especially severely asymmetric distributions, and the actual percentage of observations in any of the intervals specified by the rule could be either greater or less than those given in the rule.

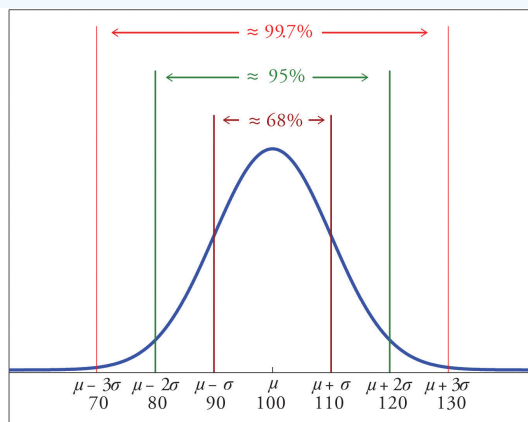
✓ ✓ Example 9.1.1

Scores on IQ tests have a bell-shaped distribution with mean $\mu = 100$ and standard deviation $\sigma = 10$. Discuss what the Empirical Rule implies concerning individuals with IQ scores of 110, 120, and 130.

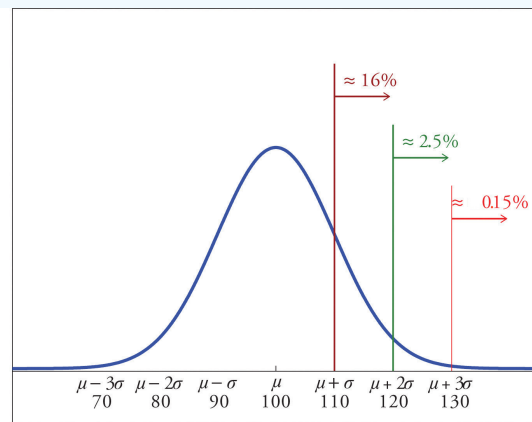
Solution:

A sketch of the IQ distribution is shown below. The Empirical Rule states that

- approximately 68% of the IQ scores in the population lie between 90 and 110
- approximately 95% of the IQ scores in the population lie between 80 and 120
- approximately 99.7% of the IQ scores in the population lie between 70 and 130.



(a) Whole Spectrum



(b) Higher End

1. Since 68% of the IQ scores lie *within* the interval from 90 to 110, it must be the case that $100\% - 68\% = 32\%$ lie *outside* that interval. By symmetry approximately half of that 32%, or 16% of all IQ scores, will lie above 110. If 16% lie above 110, then $100\% - 16\% = 84\%$ lie below. We conclude that the IQ score 110 is the 84th percentile.
2. The same analysis applies to the score 120. Since approximately 95% of all IQ scores lie within the interval from 80 to 120, only 5% lie outside it, and half of them, or 2.5% of all scores, are above 120. The IQ score 120 is thus higher than 97.5% of all

IQ scores, and is quite a high score.

3. By a similar argument, only $\frac{15}{100}$ of 1% of all adults, or about one or two in every thousand, would have an IQ score above 130. This fact makes the score 130 extremely high.

Try It 9.1.1

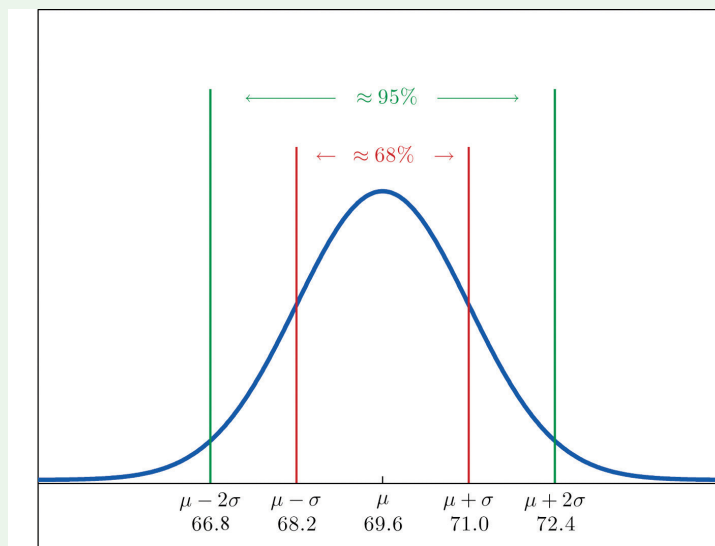
Heights of 18-year-old males have a bell-shaped distribution with mean 69.6 inches and standard deviation 1.4 inches.

- About what proportion of all such men are between 68.2 and 71 inches tall?
- What interval centered on the mean should contain about 95% of all such men?

Answer

A sketch of the distribution of heights is shown below.

- Since the interval from 68.2 to 71.0 has endpoints $\mu - \sigma$ and $\mu + \sigma$, by the Empirical Rule about 68% of all 18-year-old males should have heights in this range.
- By the Empirical Rule, the shortest such interval has endpoints $\mu - 2\sigma$ and $\mu + 2\sigma$. Since $\mu - 2\sigma = 69.6 - 2(1.4) = 66.8$ and $\mu + 2\sigma = 69.6 + 2(1.4) = 72.4$ the interval in question is the interval from 66.8 inches to 72.4 inches.



Calculating and Interpreting z-Scores

z-score

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

For example, let's say the mean score on a test for your math class was an 82, with a standard deviation of 7 points. If your score was an 89, it is exactly one standard deviation to the right of the mean; therefore, your z-score would be 1. If, on the other hand, you scored a 75, your score would be exactly one standard deviation below the mean, and your z-score would be -1 . All values that are below the mean have negative z-scores, while all values that are above the mean have positive z-scores. A z-score of -2 would represent a value that is exactly 2 standard deviations below the mean, so in this case, the value would be $82 - 2(7) = 82 - 14 = 68$.

To calculate a z-score for which the numbers are not so obvious, you take the deviation and divide it by the standard deviation.

$$z = \frac{\text{deviation}}{\text{standard deviation}}$$

You may recall that deviation is the mean value of the variable, μ , subtracted from the observed value, x , so in symbolic terms, the z-score would be:

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

As previously stated, since σ is always positive, z will be positive when x is greater than μ and negative when x is less than μ . A z-score of zero means that the term has the same value as the mean. The value of z represents the number of standard deviations the given value of x is above or below the mean and is usually rounded to 2 decimal places.

✓ Example 9.1.2

What is the z-score for an A on the test described above, which has a mean score of 82? (Assume that an A is a 93.)

Solution

For this problem, $x = 93$, $\mu = 82$ and $\sigma = 7$. The z-score can be calculated as follows:

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ z &= \frac{93 - 82}{7} \\ z &\approx 1.57 \end{aligned}$$

A score of 93 on the test is about 1.57 standard deviations above the average test score.

If we know that the test scores from the last example are distributed normally, then a z-score can tell us something about how our test score relates to the rest of the class. From the Empirical Rule, we know that about 68% of the students would have scored between a z-score of -1 and 1 , or between 75 and 89 , on the test. 95% of the students scored between a z-score of -2 and 2 , or between 68 and 96 .

✎ Try It 9.1.2

On a nationwide math test, the mean was 65 and the standard deviation was 10. If Robert scored 81, what was his z-score?

Answer

For this problem, $x = 81$, $\mu = 65$ and $\sigma = 10$. The z-score can be calculated as follows:

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ z &= \frac{81 - 65}{10} \\ z &= 1.60 \end{aligned}$$

Robert scored 1.6 standard deviations above the mean.

z-scores are useful when comparing data values that come from different data sets.

✓ Example 9.1.3

Two students, John and Ali, from different high schools, wanted to find out who had the highest GPA when compared to his school. Which student had the highest GPA when compared to his school?

Student	GPA	School Mean GPA	School Standard Deviation
John	2.85	3.0	0.7
Ali	77	80	10

Answer

For each student, determine how many standard deviations his GPA is away from the average, for his school. This is the z-score. Pay careful attention to signs when comparing and interpreting the answer.

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

For John,

$$z = \frac{2.85 - 3.0}{0.7} = -0.21$$

For Ali,

$$z = \frac{77 - 80}{10} = -0.3$$

John has the better GPA when compared to his school because his GPA is 0.21 standard deviations **below** his school's mean while Ali's GPA is 0.3 standard deviations **below** his school's mean.

John's z-score of -0.21 is higher than Ali's z-score of -0.3 . For GPA, higher values are better, so we conclude that John has the better GPA when compared to his school.

✓ Example 9.1.4

On a college entrance exam, the mean was 70, and the standard deviation was 8. If Helen's z-score was -1.5 , what was her exam score?

Solution

Since $z = \frac{x - \mu}{\sigma}$, then we can rewrite this formula solving for x :

$$x = \mu + z \cdot \sigma$$

Now, we can obtain Helen's exam score with the given parameters $\mu = 70$, $\sigma = 8$ and $z = -1.5$:

$$x = \mu + z \cdot \sigma$$

$$x = 70 + (-1.5)(8)$$

$$x = 58$$

Thus, Helen's exam score was 58. Notice a score of 58 is below the mean and this makes sense since her z-score was negative.

Notes:

1. Data available on the College Board Website: <http://professionals.collegeboard.com/data-reportsresearch/ap/archived/2007>

9.1: The Normal Distribution is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **11.1: The Standard Normal Probability Distribution** by Darlene Diaz is licensed CC BY-NC-SA 4.0. Original source: [https://www.sccollege.edu/OER/Documents/MATH 105/Math For Liberal Art Students \(2017\).pdf](https://www.sccollege.edu/OER/Documents/MATH%20105/Math%20For%20Liberal%20Art%20Students%20(2017).pdf).
- **11: Normal Distribution** by Darlene Diaz is licensed CC BY-NC-SA 4.0. Original source: [https://www.sccollege.edu/OER/Documents/MATH 105/Math For Liberal Art Students \(2017\).pdf](https://www.sccollege.edu/OER/Documents/MATH%20105/Math%20For%20Liberal%20Art%20Students%20(2017).pdf).
- **2.5: The Empirical Rule and Chebyshev's Theorem** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: <https://2012books.lardbucket.org/books/beginning-statistics>.
- **2.8: Measures of the Spread of the Data** by OpenStax is licensed CC BY 4.0. Original source: <https://openstax.org/details/books/introductory-statistics>.