Understanding the Dimensions
Of the Normal Distribution

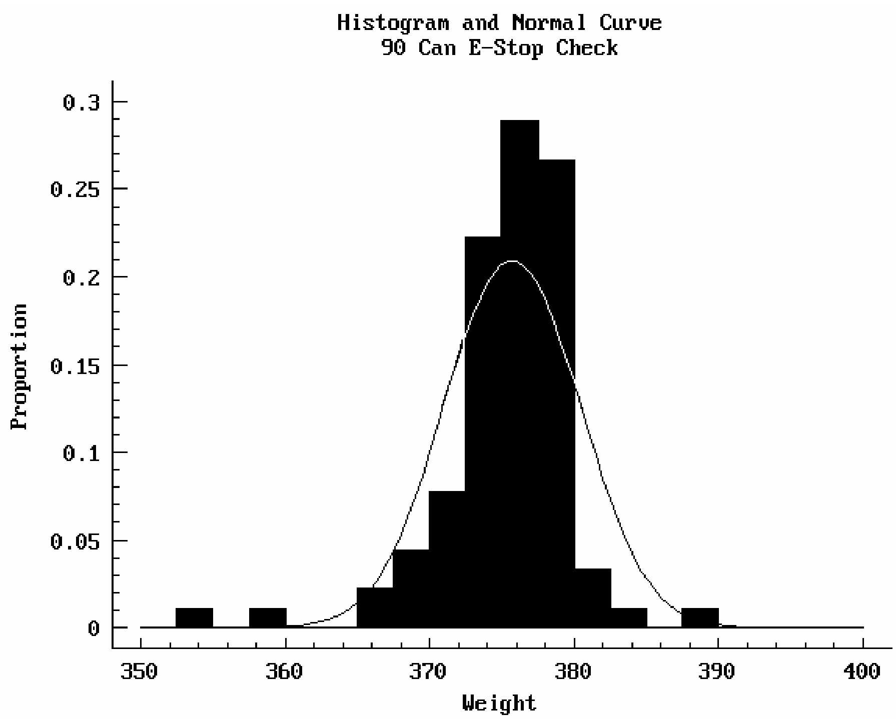
Misty River Consulting
Donald A. Kerper

Understanding the Dimensions Of The Normal Distribution

Normal Distribution

A normal distribution is an arrangement of a population's data that when put into a histogram form has a bell shape. This histogram, if made up of a large number of columns and a large sample size will have appear as a "bell" shaped curve with a peak at the average and tails that decrease in height, eventually perceptually leveling out, as the tails move away from the average.

This normal curve graph is a "picture" of one kind of a variation's distribution – a normal distribution. It usually is the kind of variation found within nature where only natural variation is present. Nothing special is happening to effect the results.



Normal Distribution Dimensions

The normal distribution has two basic dimensions: central tendency and range of dispersion. The central tendency is the dimension of the normal curve that shows where the data tends to be at. As we can see in looking at a normal curve, the central tendency tends to be centered around the peak of the curve.

The other dimension of the normal distribution is the "range of dispersion". The range of dispersion is a dimension of the normal curve that indicates how far and fast the distribution's data spreads out from its "central tendency". Data from a normal distribution therefore tends to centrally locate around the middle of the distribution but also tends to disperse from the middle in a decreasing fashion.

Central Tendency
Average

Several statistical tools can be used to describe central tendency that include:

1) Average or mean

2) Median

3) Mode

4) Histogram

The average of a normal distribution is found at the peak of the bell curve. The average is a statistical tool or descriptive statistic that describes the central tendency of the normal distribution. The meaning of average is found in the meaning of central tendency – "average" is where things tend to be.

In a normal distribution, you will find that half of the data in the distribution is on one side of the distribution and half of the data is on the other side. For non normal distributions, the average may or may not be at the peak of the distribution. In these non normal distributions, the meaning of average is lost – only the formula makes sense.

How To Calculate

Calculating averages can be accomplished by following the following steps:

1) Count how many data points you have

2) Add up the total of the data points

3) Divide the sum of the data points by the number of data points - this is the average

Example:

Given this set of data:

1, 4, 2, 5, 3, 6, 4, 7, 5, 8, 6, 9

The average would be:

$$\frac{1+4+2+5+3+6+4+7+5+8+6+9}{12} = \frac{60}{12} = 5 = \text{the average}$$

12 = the number of data points in the set of data.

Median

The median of a normal distribution is also found at the peak of the bell curve. The meaning of median is that half of the data is on one side of the median and half is on the other side. This is true whether the distribution is a normal distribution or not.

How to Calculate

To calculate the median of a set of data, follow the following steps.

1) Place the data in numerical order.

2) Count how many pieces of data there are.

3) If an even number of datum exist, select the two middle numbers and average them. This is the median.

4) If an odd number of data exist, select the middle number. This is the median.

Example

Given this set of data:

1, 4, 2, 5, 3, 6, 4, 7, 5, 8, 6, 9

The median would be:

1, 2, 3, 4, 4, 5, 5, 6, 6, 7, 8, 9 – placed in numerical order

12 = the number of data points in the set of data.

The two middle numbers are 5 and 5

(5 + 5) / 2 = 5 – the average of the two middle numbers

5 = the median

Mode

Regardless of what distribution the data is sampled from, the mode is the data that occurs most often.

In a normal distribution, the mode is at the peak of the curve along with the average and the median. In a non normal distribution, the mode again is at the peak of the distribution.

How to Calculate

To calculate the median of a set of data, follow the following steps.

1) Place the data in numerical order.

2) Count how often each data value occurs. For situations where the data set is large, consider constructing a frequency table.

3) The data that occurs most often is the mode. Sometimes you might find more than one value that occurs most often – this would create a bimodal situation.

Example

Given this set of data:

1, 4, 2, 5, 3, 6, 4, 7, 5, 8, 6, 9

The mode would be:

1, 2, 3, 4, 4, 5, 5, 6, 6, 7, 8, 9 – placed in numerical order

Value	Frequency
1	1
2	1
3	1
4	2
5	2
6	2
7	1
8	1
9	1

The numbers 4, 5, and 6 occur most often.

The mode is a trimodal situation with 4, 5, and 6 all being modes.

Range of Dispersion

The two dimensions of the normal curve include not only the "central tendency", but also the "range of dispersion". Several statistical tools can be used to describe range of dispersion that include:

1) Range

2) Standard Deviation

3) Variance

4) Histogram

Range

A range tells how much the variable or attribute varies around the average. The range allows us to understand the spread or width of the variation that is found in the variable or attribute we are measuring or observing.

How To Calculate

Calculating ranges can be accomplished by following the following steps:

1) Identify the largest number

2) Identify the smallest number

3) Subtract the smallest from the largest number - this is the range.

Example

Given this set of data:

1, 4, 2, 5, 3, 6, 4, 7, 5, 8, 6, 9;

The range would be:

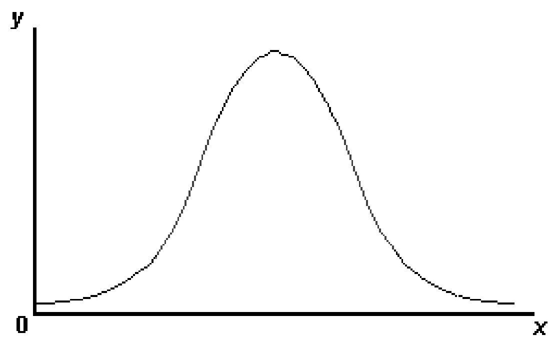
9 is the highest number and 1 is the lowest number.

9 - 1 = 8 = the range.

Standard Deviation

Standard deviation is a tool that describes the "range of dispersion" found in a normal distribution. It is a tool that allows us to understand where data is located as it disperses from the average of the distribution.

Standard Deviation is a unit of measure along the x-axis of the normal distribution chart. Starting from the average of the normal distribution, standard deviation units indicate how data disperses from the average- it allows us to understand how the data is arranged in a normal distribution.

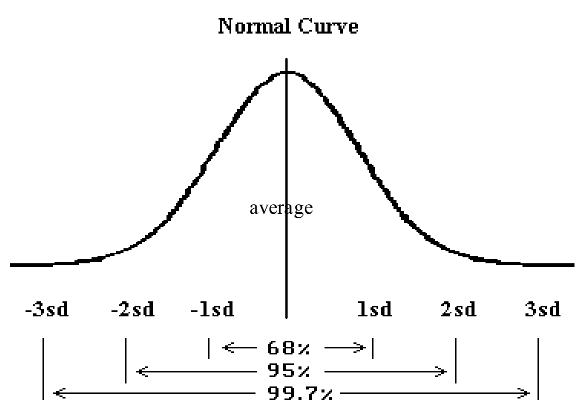


Standard deviation's meaning is only valid when this tool is applied to normal distributions. Distributions that are not normal cannot be described directly by the standard deviation tool. In a normal distribution, data is always arranged in the same exact way. It will always have a shape of a bell when it is plotted out in a histogram format. In all normal distributions we would find:

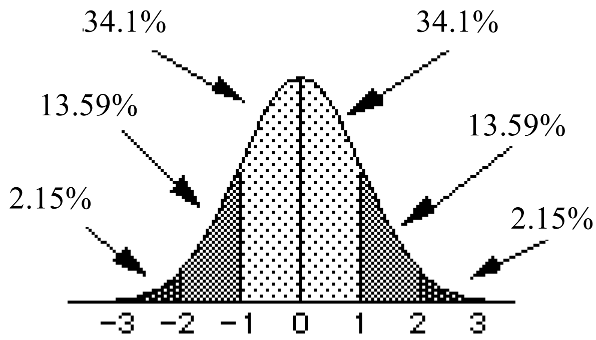
1) 68.3% of the data lies between –1 and 1 standard deviation unit or distance from the average.

2) 95.4% of the data lies between –2 and 2 standard deviation units or distance from the average.

3) 99.7% of the data lies between –3 and 3 standard deviation units or distance from the average.



In a nutshell, the meaning of standard deviation is found in these percentages. All normal distributions have data arranged as found in the above and below graphs. More specific information about how data is arranged can be found in a "Z Table" (found at the end of this article).



In the above diagram, the average is marked with “0”. The average is zero standard deviation units from itself and consequently that why the average is marked zero on the chart.

When we consider that one half of a normal distribution is the mirror image of the other have, we can get more details about how data is arranged in a normal curve.

Normal Distribution Equation

Although we can get a basic idea of how data is arranged from the above graphs and from the “Z Table”, the exact data arrangement found in a normal distribution is described by the following formula:

$$y = f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}$$

Where:

$\pi = 3.14159$

$e = 2.718282$

μ = the mean of the data

σ = the standard deviation of the data

x = the value of the x axis and

y = the resulting height found on the y-axis.

“x” can take on any value from -infinity to +infinity. “y” is very close to 0 if x is more than three standard deviations from the mean (less than -3 or greater than +3 standard deviations from the mean). As “x” approaches the mean, “y” increases and reaches its highest peak at the average or mean of the normal distribution.

Calculating the Standard Deviation

The calculation of the standard deviation is found in the following technical definition: take the distance of each number from the mean, square it, average the result, then take the square root. In short, calculating the standard deviation is taking the root mean square of the distances (or differences) from mean.

Calculating the standard deviation of a sample traditionally has been a very laborious task. In recent years, the task is much easier due to the availability of cheap yet sophisticated scientific calculators and the increasing availability of computers. The calculation is done using one of two versions of the standard deviation formula. The first version uses "n" in the denominator and is appropriate for sample sizes of 30 or more. The second version uses "n -1" in the denominator and is used for sample sizes less than 30.

For sample sizes less than 30:

$$s = \left(\frac{\sum \left(x - \bar{x} \right)}{n-1} \right)^{1/2} = \left(\frac{\sum x^2 - \frac{\left(\sum x \right)^2}{n}}{n-1} \right)^{1/2}$$

For sample sizes equal to or greater than 30:

$$s = \left(\frac{\sum \left(x - \bar{x} \right)}{n} \right)^{1/2} = \left(\frac{\sum x^2 - \frac{\left(\sum x \right)^2}{n}}{n} \right)^{1/2}$$

However, if one really does want to calculate the standard deviation by hand, one can use the following procedures to ease the complexity of the calculation. First set up a table of x and x^2. Second, calculate the sum of squares of the deviation from the mean, and finally, use the result in the standard deviation formula. For example, given 85, 70, 60, 90, 81, what is x and the standard deviation?

First, set up the table to determine the sum of squares of deviations:

	x	**x^2**
	85	7225
	70	4900
	60	3600
	90	8100
	81	6561
Total (Sum)	**386**	**30386**

The mean is determined by the following:

$$\bar{x} = \frac{\sum x}{n} = \frac{386}{5} = 77.2$$

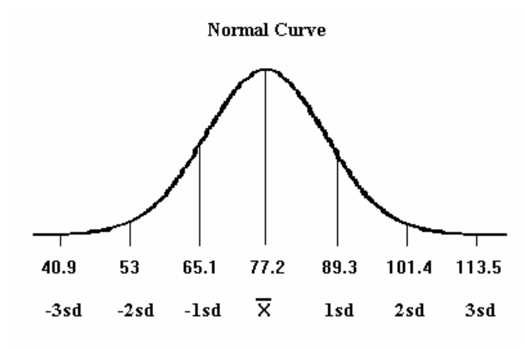
The sum of squares of deviations is determined with the following procedure:

$$\begin{aligned}\sum(x - x^2) &= \sum x^2 - \frac{(\sum x)^2}{n} \\ &= 30386 - \frac{386^2}{5} \\ &= 30386 - 29799.2 \\ &= 586.8\end{aligned}$$

The standard deviation would use the "n -1" version and would then be:

$$s = \left(\frac{\sum(x-\bar{x})^2}{(n-1)}\right)^{\frac{1}{2}} = \left(\frac{586.8}{4}\right)^{\frac{1}{2}} = (146.7)^{\frac{1}{2}} = 12.1$$

In the following normal curve diagram, the calculated average of the above example is plotted along with the calculated standard deviation locations.



Variance

Variance is a term that refers to the average squared distance from the average of the distribution being studied. What this means is that given all the data and each data's distance from the average of the distribution, if we square all those differences and then determine the average of those distances, we now know the variance. Complicated? Not too bad. Useable? Not really.

How to Calculate

The easiest way to calculate the variance is to simply square the standard deviation. The variance is the square of the standard deviation.

Example

The standard deviation in the above example was calculated to be 12.1. The variance would be the square of that or 146.41.

Histograms

Histograms provide a graphical way to show the central tendency of the data and the range of dispersion of the data simultaneously. Because of the complexity of the histogram, it is discussed elsewhere.