

Section 1F – Normal Quantitative Data Analysis

Vocabulary

Quantitative data: Data in the form of numbers that measure or count something. They usually have units and taking an average makes sense. For example, height, weight, salary, or the number of pets a person has.

Normal Data: Data that is bell shaped, symmetric and unimodal. Also referred to as data that has a normal distribution.

Sample Size: Also called the total frequency.

Average: Also called the center of the data. A single number that represents a typical person or object in the data set.

Variability: Also called the spread. A measure of how spread out a data set is. A large spread tells us that the data is less consistent and the more difficult to predict. A small spread tells us that the data is more consistent and easier to predict.

Mean Average (\bar{x}): The balancing point for distances in a data set. The average for a data set that is normal.

Standard Deviation: The average or typical distance that points in a data set are from the mean. The measure of typical spread (typical variability) for a data set that is normal.

Maximum: The largest number in a data set.

Minimum: The smallest number in a data set.

Outliers: Unusual values in the data set.

Introduction

When analyzing numerical quantitative data, always start with finding the shape of the data set. Categorical data can be graphed, but does not have a shape. Categorical bar charts can be organized in a variety of ways depending on the order of the categories. Quantitative data is numerical measurement data and does have a shape.

Why should we find the shape?

The goal in analyzing quantitative data is to find the average, spread and unusual values. In statistics, there are many types of averages, many types of spreads. Shape helps us determine which averages and spreads are most accurate for the data.

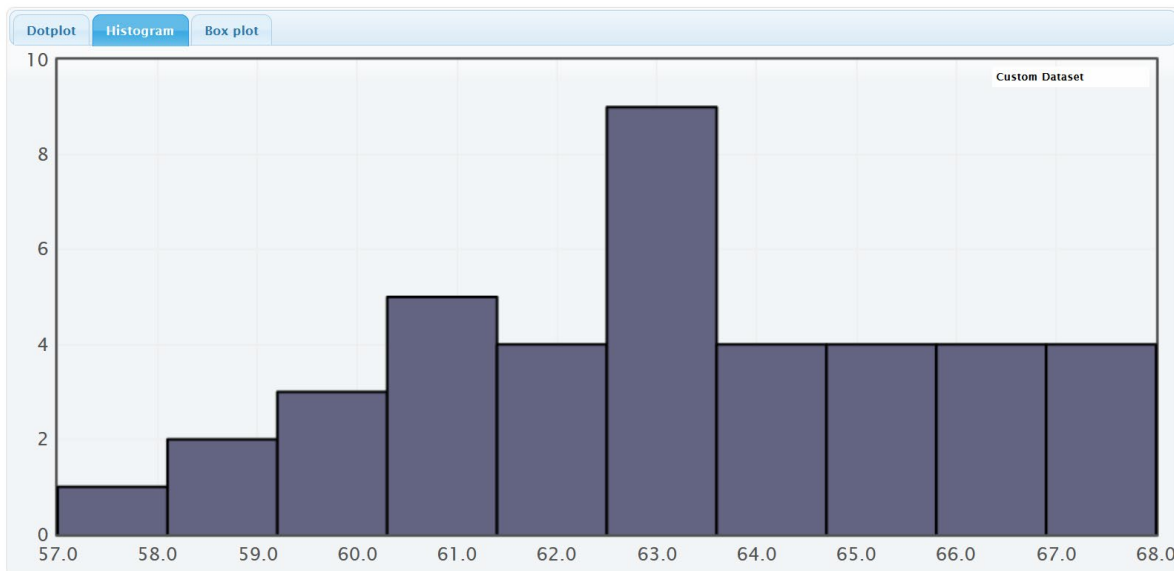
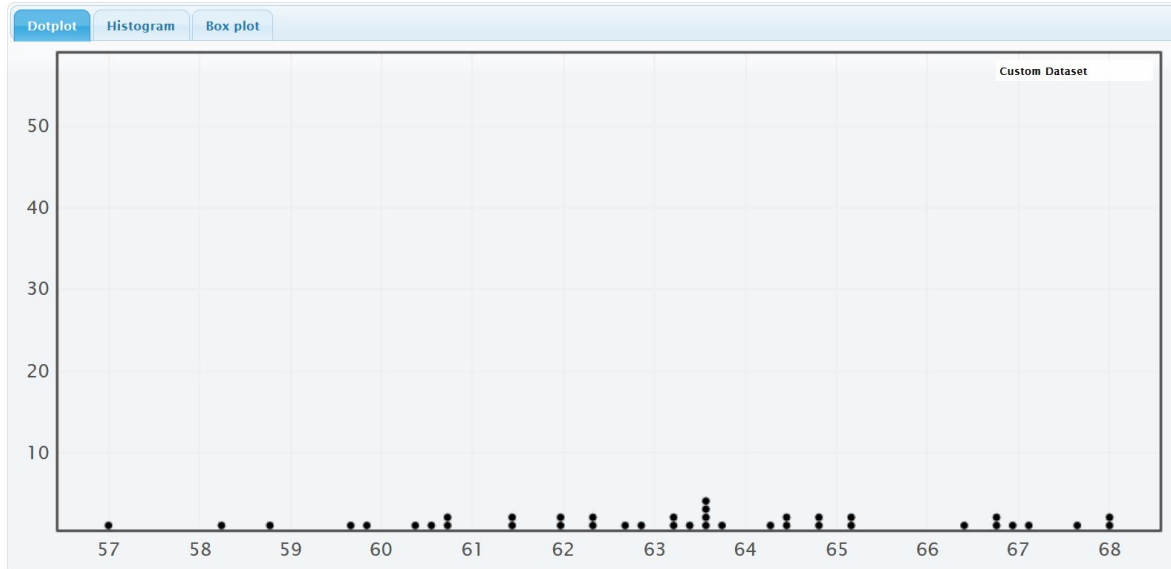
Quantitative Statistics and Graphs with StatKey

The most common quantitative statistics we like to look at are the mean, median, standard deviation, first quartile, third quartile, interquartile range, max, min, and range. The most basic kind of graph for quantitative data is the dot plot. The computer draws the numerical scale usually horizontally. It then draws a dot for every single number in the data set. Another type of graph is a histogram. This graph counts the number of data values in certain sections and makes a bar telling us how many numbers are in that section. The number of bars are also called “bins” or “buckets”. Another graph we like to look at is the boxplot. A boxplot is a graph of the first quartile, median, and third quartile as well as potential outliers.

All of these graphs and statistics can be made with StatKey. Let us look at an example. Go to www.matt-teachout.org and click on the “statistics” tab and then the “data sets” tab. Look for the “Health Data” excel file. Open the data set and copy the women’s heights data. Notice the data is quantitative. It measures the height in inches of the women and it seems reasonable to look for an average height of these women.



Go to www.lock5stat.com and click on the “StatKey” button. Under the “Descriptive Statistics and Graphs” menu, click on “One Quantitative Variable”. Click on the “Edit Data” button. Copy and paste the women’s height data into StatKey. Uncheck the box that says, “First column is an identifier”. An identifier is a word next to every number. This data set does not have that. Check the box that says, “Data has a header row”. This means the data set has a title. Now push “OK”. Notice StatKey gives you the sample statistics, a dot plot, a histogram and a boxplot.



On the right of this histogram, you will see a slider that can adjust the number of “buckets” or bins. The smaller the data set the less bins you should have. This data set only has 40 numbers, so we want only a few bars. If we slide it to 3 buckets, we get the following.

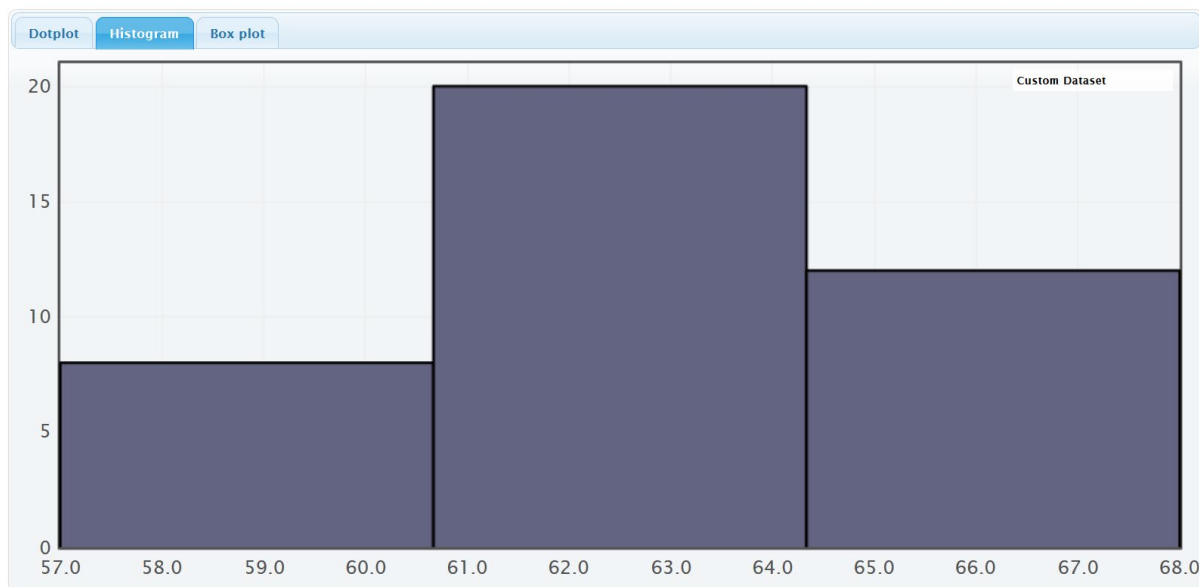
Histogram Controls

Set Limits

Number of buckets: 3



This chapter is from *Introduction to Statistics for Community College Students*, 1st Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a “CC-By” [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) – 10/1/18



This data has a very special shape. It is called bell shaped or normal. Normally distributed data has the highest bar in the middle and about equal number of bars decreasing from the middle. It looks like bell. We see that this data set is relatively normal (bell shaped) or “normally distributed”. StatKey has also given us summary statistics. Which statistics are most accurate for normal data?

Summary Statistics

Statistic	Value
Sample Size	40
Mean	63.195
Standard Deviation	2.741
Minimum	57
Q ₁	61.350
Median	63.350
Q ₃	64.900
Maximum	68

Mean and Standard Deviation

Important Note about Shape: The mean and standard deviation should only be used if the data set is normal. The mean and standard deviations are not accurate if the data does not have a normal shape.

Mean (\bar{x}): The mean is a type of average used for data that is normally distributed. The mean balances the distances between all the numbers in the data set and the mean. Think of it this way. If you took all the numbers in the data set below the mean, measured their distances from the mean, then added up those distances. That total distance for numbers below the mean would be equal to the total distance for numbers above the mean. The mean is calculated by adding up all the numbers in a data set ($\sum x$) and then dividing by how many numbers are in the data set (sample size “n”).

$$\bar{x} = \frac{\sum x}{n}$$



This chapter is from *Introduction to Statistics for Community College Students*, 1st Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a “CC-By” [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) – 10/1/18

Standard Deviation (S): We said that the mean balances the distances in a data set. The standard deviation calculates the average distance numbers are from the mean. It is the most accurate measure of typical spread for data sets that are normally distributed. To calculate the standard deviation, computer programs take every single number in the data set and subtract the mean. Since those differences can be negative sometimes, they computer squares all the differences and then adds up the squares. This is a famous calculation called “sum of squares”. Since we want the average distance, we divide by $n - 1$ (degrees of freedom) and take the square root at the end to undo all the squares. Never calculate this by hand. It is a long calculation that should be left to a computer program.

$$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Why do we study spread? Spread is a measure of how much variability is in the data set. Think of it this way. Suppose we were looking at exam scores in a history class that are normally distributed. If a data set were very spread out, then the standard deviation would be quite large. This would mean that the scores had a lot of variability. We had A's, B's, C's, D's, and F's. The exam scores are not consistent, and the history teacher will have a hard time predicting how her class will do. If the data set has a small spread, then the standard deviation would be quite small. The exam scores are very consistent. Maybe everyone in the class got an A or a high B. It is easier to predict how the class will do.

Statistics for Normal Data

Quantitative Variable and Units

Sample Size (n)

Maximum Value

Minimum Value

Average: Mean (\bar{x})

Spread: Standard Deviation (s)

Typical Values: One standard deviation from the mean. Here is a formula that is sometimes used.

$$\bar{x} - s \leq \text{typical values} \leq \bar{x} + s$$

Outliers (unusual values): More than two standard deviations from the mean. Here are formulas that are sometimes used.

$$\text{Unusually Low Values (Low outliers)} \leq \bar{x} - 2s$$

$$\text{Unusually High Values (High outliers)} \geq \bar{x} + 2s$$

Women's Height Example

Quantitative Variable and Units: Women's heights in inches

Sample Size (n): There were 40 women in the data set.

Maximum Value: The tallest woman in the data set was 68 inches.

Minimum Value: The shortest woman in the data set was 57 inches.

Average: Mean (\bar{x}). The average height of the women in the data was 63.195 inches.

Spread: Standard Deviation (s). The typical spread for this data was 2.741 inches. Typical women in the data were 2.741 inches from the mean.



Typical Values: Add and subtract the mean and standard deviation. Typical women in the data set have a height between 60.454 inches and 65.936 inches. We will see later that these values are the cutoffs for the middle 68% for normal data.

$$\bar{x} - s \leq \text{typical values} \leq \bar{x} + s$$

$$63.195 - 2.741 \leq \text{typical values} \leq 63.195 + 2.741$$

$$60.454 \leq \text{typical values} \leq 65.936$$

Outliers (unusual values): Add and subtract the mean and two standard deviations. Unusually tall women are 68.677 inches or higher. There are no unusually tall women in this data set. Unusually short women are 57.713 inches or lower. This means that the minimum value of 57 inches was unusually low. We will see later that these values are the cutoffs for the top and bottom 2.5% for normal data.

$$\text{Unusually Low Values (Low outliers)} \leq \bar{x} - 2s = 63.195 - (2 \times 2.741) = 57.713 \text{ inches}$$

$$\text{Unusually High Values (High outliers)} \geq \bar{x} + 2s = 63.195 + (2 \times 2.741) = 68.677 \text{ inches}$$

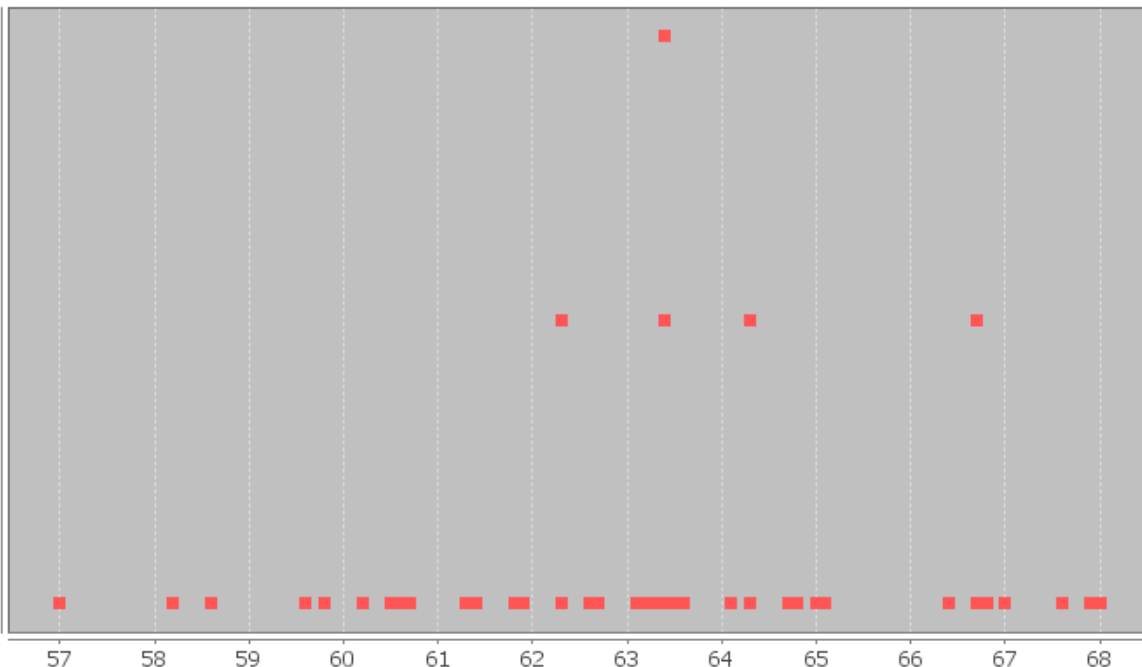
Quantitative Statistics and Graphs with Statcato

You can also make dot plots, histograms and sample statistics with Statcato. Copy and paste women's heights into a column of Statcato. The data set is only 40 values, so you will not need to add rows to Statcato. To make a dot plot, go to the graph menu and click on dot plot. Then click on the column of data you want to use. Then push ok.

Making a dot plot in Statcato: *Graph => Dot plot => Pick a column => OK*

Here is the dot plot for the 40 women's heights.

Dot Plot Women's Heights (Inches)



To make a histogram in Statcato, go to the graph menu, and then click on histogram. Chose a column of data and how many bars (bins) you want. Then chose ok.

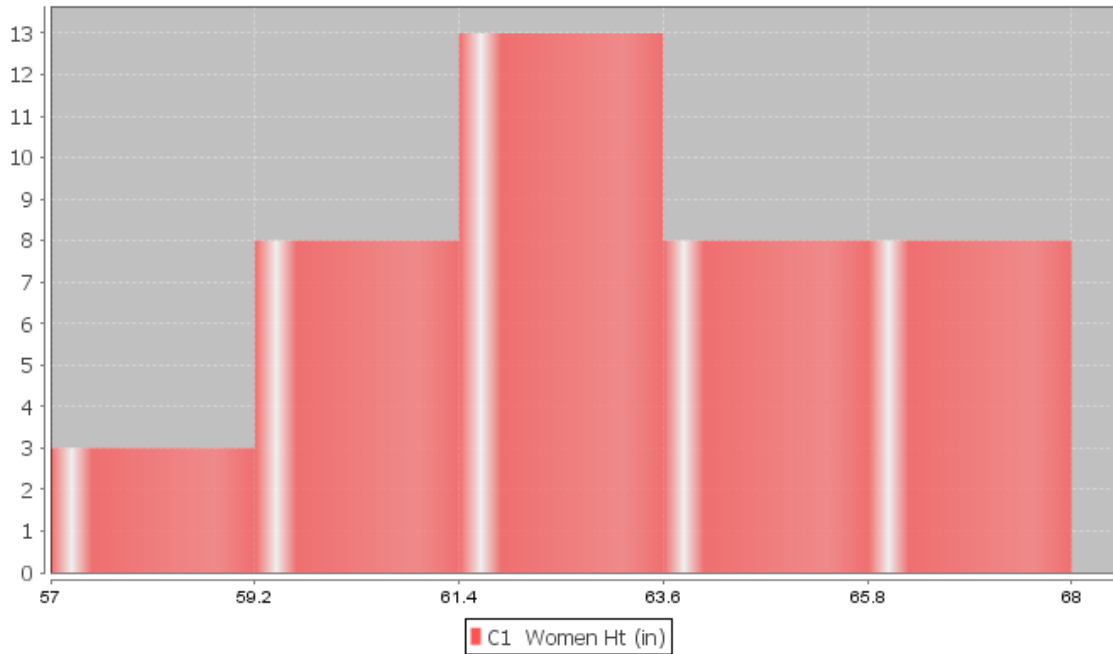
Making a histogram in Statcato: *Graph => Histogram => Pick a column => Chose number of bins => OK*



This chapter is from [Introduction to Statistics for Community College Students](#), 1st Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a ["CC-By" Creative Commons Attribution 4.0 International license](#) – 10/1/18

Note about bins: If you chose too many bars then the histogram starts to look very crazy and you will have a hard time seeing the shape. Remember the goal is to break the dots up into groups. For example, in this health data there are only 40 women. I would not want 40 bins since that would give me about one bar per dot. If it were a small data set like the health data, I would do about three bins. Remember, the more bins you have, the more difficult it is to see the shape. This graph has five bins.

Histogram of Women's Height (Inches)



Notice again that the highest bar is close to the middle and the bars get smaller as we move away from the middle. This is often called “Bell Shaped” or “Normal Data”. Some like to describe this shape as unimodal (1 hill) and symmetric (left and right side look about the same). I prefer to call it bell shaped or normal.

We can also calculate all of the sample statistics with Statcato. Go to the “Statistics” menu. Then click “Basic Statistics” and “Descriptive Statistics”. I had pasted the data into column 1, so type in “C1” under “input variable”. Check the boxes for statistics that you want and push “OK”.



Descriptive Statistics ×

Help F1

<p>Inputs</p> <p>Input Variable(s): C1</p> <p>Enter valid column names separated by space. For a continuous range of columns, separate using dash (e.g. C1-C30).</p> <p>By Variable (optional): <input type="text" value=""/></p> <p>Results</p> <p>Store Results in: <input type="checkbox"/> New datasheet</p>	<p>Statistics</p> <p><input type="checkbox"/> Select all statistics</p> <table border="0" style="width: 100%;"> <tr> <td><input checked="" type="checkbox"/> Mean</td> <td><input type="checkbox"/> Trimmed mean: cutoff % <input type="text" value=""/> % of values to be trimmed (between 0 and 100)</td> </tr> <tr> <td><input type="checkbox"/> SE of mean</td> <td><input type="checkbox"/> Sum</td> </tr> <tr> <td><input checked="" type="checkbox"/> Standard deviation</td> <td><input checked="" type="checkbox"/> Minimum</td> </tr> <tr> <td><input type="checkbox"/> Variance</td> <td><input checked="" type="checkbox"/> Maximum</td> </tr> <tr> <td><input type="checkbox"/> Coefficient of variation</td> <td><input type="checkbox"/> Range</td> </tr> </table> <table border="0" style="width: 100%;"> <tr> <td><input checked="" type="checkbox"/> First quartile</td> <td><input type="checkbox"/> N nonmissing</td> <td><input type="checkbox"/> Sum of squares</td> </tr> <tr> <td><input checked="" type="checkbox"/> Median</td> <td><input type="checkbox"/> N missing</td> <td><input type="checkbox"/> Skewness</td> </tr> <tr> <td><input checked="" type="checkbox"/> Third quartile</td> <td><input checked="" type="checkbox"/> N total</td> <td><input type="checkbox"/> Kurtosis</td> </tr> <tr> <td><input checked="" type="checkbox"/> Interquartile range</td> <td><input type="checkbox"/> Cumulative N</td> <td><input type="checkbox"/> MSSD</td> </tr> <tr> <td><input type="checkbox"/> Mode</td> <td><input type="checkbox"/> Percent</td> <td></td> </tr> <tr> <td><input type="checkbox"/> Percentile: <input type="text" value=""/></td> <td><input type="checkbox"/> Cumulative Percent</td> <td></td> </tr> </table> <p>e.g. 10 for the 10th percentile</p> <p style="text-align: right;"> <input type="button" value="OK"/> <input type="button" value="Cancel"/> </p>	<input checked="" type="checkbox"/> Mean	<input type="checkbox"/> Trimmed mean: cutoff % <input type="text" value=""/> % of values to be trimmed (between 0 and 100)	<input type="checkbox"/> SE of mean	<input type="checkbox"/> Sum	<input checked="" type="checkbox"/> Standard deviation	<input checked="" type="checkbox"/> Minimum	<input type="checkbox"/> Variance	<input checked="" type="checkbox"/> Maximum	<input type="checkbox"/> Coefficient of variation	<input type="checkbox"/> Range	<input checked="" type="checkbox"/> First quartile	<input type="checkbox"/> N nonmissing	<input type="checkbox"/> Sum of squares	<input checked="" type="checkbox"/> Median	<input type="checkbox"/> N missing	<input type="checkbox"/> Skewness	<input checked="" type="checkbox"/> Third quartile	<input checked="" type="checkbox"/> N total	<input type="checkbox"/> Kurtosis	<input checked="" type="checkbox"/> Interquartile range	<input type="checkbox"/> Cumulative N	<input type="checkbox"/> MSSD	<input type="checkbox"/> Mode	<input type="checkbox"/> Percent		<input type="checkbox"/> Percentile: <input type="text" value=""/>	<input type="checkbox"/> Cumulative Percent	
<input checked="" type="checkbox"/> Mean	<input type="checkbox"/> Trimmed mean: cutoff % <input type="text" value=""/> % of values to be trimmed (between 0 and 100)																												
<input type="checkbox"/> SE of mean	<input type="checkbox"/> Sum																												
<input checked="" type="checkbox"/> Standard deviation	<input checked="" type="checkbox"/> Minimum																												
<input type="checkbox"/> Variance	<input checked="" type="checkbox"/> Maximum																												
<input type="checkbox"/> Coefficient of variation	<input type="checkbox"/> Range																												
<input checked="" type="checkbox"/> First quartile	<input type="checkbox"/> N nonmissing	<input type="checkbox"/> Sum of squares																											
<input checked="" type="checkbox"/> Median	<input type="checkbox"/> N missing	<input type="checkbox"/> Skewness																											
<input checked="" type="checkbox"/> Third quartile	<input checked="" type="checkbox"/> N total	<input type="checkbox"/> Kurtosis																											
<input checked="" type="checkbox"/> Interquartile range	<input type="checkbox"/> Cumulative N	<input type="checkbox"/> MSSD																											
<input type="checkbox"/> Mode	<input type="checkbox"/> Percent																												
<input type="checkbox"/> Percentile: <input type="text" value=""/>	<input type="checkbox"/> Cumulative Percent																												

Z-scores

In normal data, we often want to find out how many standard deviations a number (X-value) is from the mean. This is called a “Z-score”. Here is a common formula. In later chapters, we will see that we can also use the Z-score as a test statistic to measure significance.

$$Z = \frac{(X \text{ value} - \text{Mean})}{\text{Standard Deviation}}$$

Example: In the last example, we saw that the women’s height data was normally distributed with a mean of 63.195 inches and a standard deviation of 2.741 inches. Suppose a woman is 72 inches tall. What would be the Z-score for her height? Is she unusually tall?

It is important when calculating a Z-score that you subtract the X value and the mean first. Then divide by the standard deviation. Most people in statistics round Z-scores to the hundredths place (two numbers to the right of the decimal).

$$Z = \frac{(X \text{ value} - \text{Mean})}{\text{Standard Deviation}} = \frac{(72 - 63.195)}{2.741} = +3.21233 \approx +3.21$$

If the X-value is below the mean, the Z-score will be negative. If the X-value is above the mean, the Z-score will be positive. This Z-score was positive. So the woman that is 72 inches tall is 3.21 standard deviations above the mean. Is this unusual?

Remember the formula above for finding the cutoff for unusual values for normal data. Notice it is two standard deviations above and below the mean. Two standard deviations above the mean would be a Z-score of +2. Two standard deviations below the mean would be a Z-score of -2. So a common way to judge if a number is unusual (outlier) for normal data is to look at the Z-score.

Unusual High Values for Normal Data: $Z \geq +2$

Unusual Low Values for Normal Data: $Z \leq -2$

Hence, since the woman’s Z-score was greater than or equal to +2, she is unusually tall compared to the women in the data set.



Example: The women's height data was normally distributed with a mean of 63.195 inches and a standard deviation of 2.741 inches. One woman in the data set was 57 inches tall and we said was unusually short. If you recall, her height was below the unusual low cutoff of 57.713 inches. What would be the Z-score for her height?

$$Z = \frac{(X \text{ value} - \text{Mean})}{\text{Standard Deviation}} = \frac{(57 - 63.195)}{2.741} \approx -2.26$$

Since the X-value is below the mean, the Z-score will be negative. So the woman that is 57 inches tall is 2.26 standard deviations below the mean. Remember if the Z-score is less than -2, it is unusually low. This confirms what we already knew.

Typical Z-scores: Remember that typical values are within one standard deviation from the mean. This would mean that typical Z-scores are between -1 and +1.

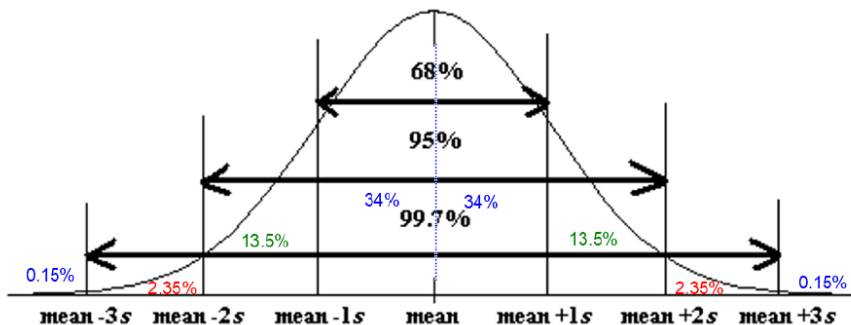
$$-1 \leq \text{Typical Z-scores} \leq +1$$

A woman with a height of 61 inches would have a Z-score of -0.80. Notice that this Z-score is between -1 and +1 on the number line. Therefore, 61 inches is a typical height for women in this data set.

Note: Not all values are typical or unusual. A person that is 1.5 standard deviations from the mean would be neither typical (Z-score not between -1 and +1) nor unusual (Z-score not greater than +2 or less than -2).

Empirical Rule

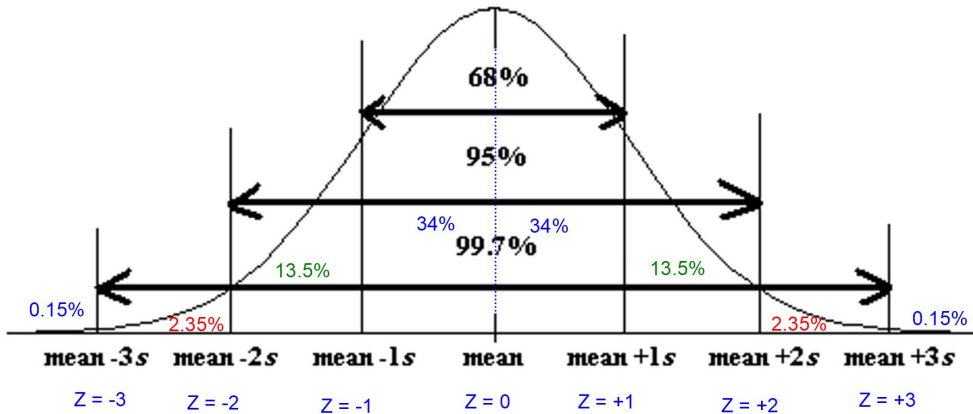
There is common percentages that go with normal (bell-shaped) data. Usually about 68% of normal data will be within one standard deviation of the mean (typical). About 95% of normal data will be within two standard deviations of the mean. About 99.7% of normal data will be within three standard deviations of the mean. These percentages are often referred to as the "Empirical Rule" or the "68-95-99.7 Rule".



Notice that we can use the 68%, 95% and 99.7% to figure out the sections. Since 68% makes up the middle two symmetric sections, we know each section is about 34%. Similarly, the middle four sections make up about 95%. Subtract out the middle two sections (68%) gives 27%. Divide that in half and you get two sections each making up 13.5% of the normal data. The middle six sections make up about 99.7%. Subtract out the middle four sections (95%) gives 4.7%. Divide that in half and you get two sections each making up 2.35% of the normal data. The end sections are calculated in a similar manner (100%-99.7% = 0.3%). Divide that into two symmetric tails and we get that each tail should be about 0.15%.

Remember the number of standard deviations from the mean is the Z-score. You can write the Z-scores for the bottom values in the Empirical rule. This is often called the "Standard Normal Curve". Notice the center of the curve is the mean (Z-score of zero) and the standard deviation of this curve is exactly one. When a computer program refers to a normal curve with a mean of zero and a standard deviation of one, they are talking about Z-scores and the Standard Normal Curve.





Many data sets are normal. We will see in the next chapter that many sampling distributions have a normal shape as well. It is therefore important to be able to calculate percentages associated with normal data and normal curves. Confidence Intervals and P-value are both extremely important topics that we will cover in chapter 3 and chapter 4 that involve the empirical rule and calculating percentages associated with normal curves.

Calculating Percentages for Normal Curves with StatKey

Computer software programs can calculate percentages associated with normal quantitative data. Go to www.lock5stat.com and click on “StatKey”. Under the “Theoretical Distributions” menu click on “Normal”. Notice the parameters are set at a mean of zero and a standard deviation of one. Remember this means it is set up to find Z-scores or to find percentages associated with Z-scores. The curve is sometimes called a “density curve”. The idea is that the total area under the curve is 100%, so to find a percentage you find the area under the curve.

Normal Distribution

Mean	Standard Deviation
0	1

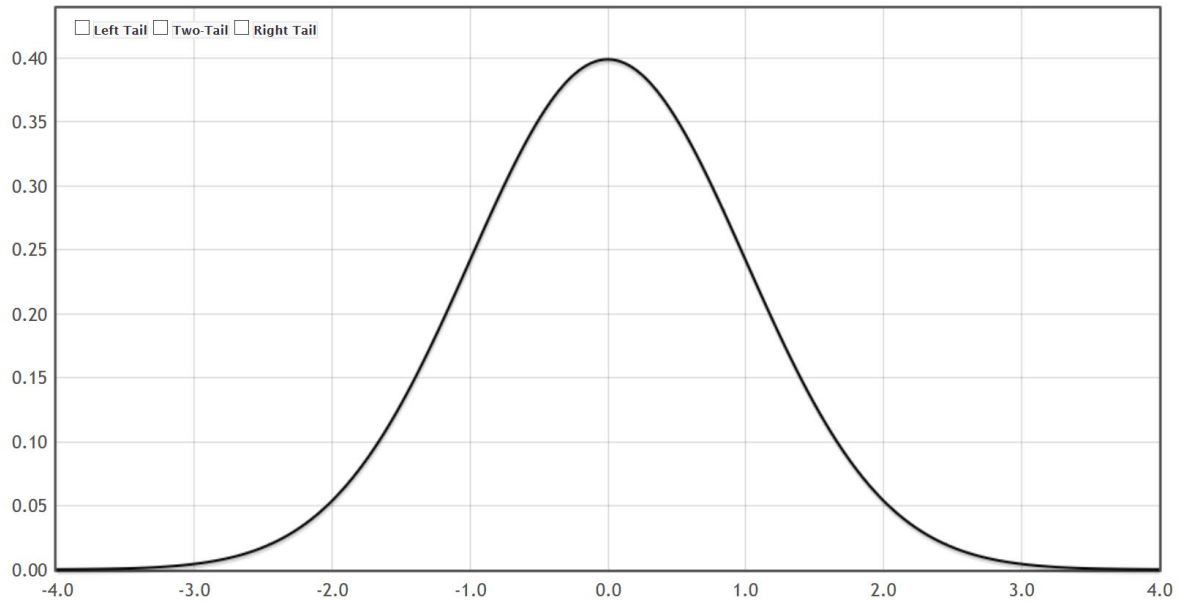
[Edit Parameters](#)



StatKey Theoretical Distribution

Normal Distribution ▾

Reset Plot



Notice that the curve has three buttons on the top left (Left Tail, Two-Tail, and Right Tail).

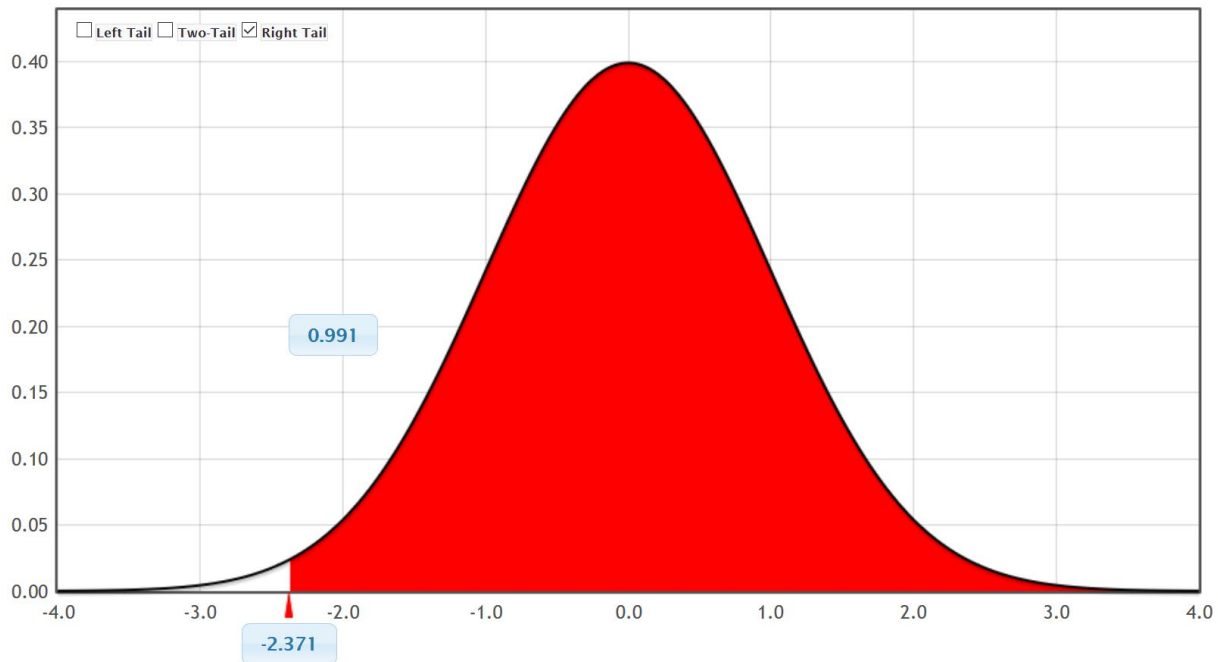
Example: Suppose we want to find what percent of normal data has a Z-score of -2.371 or above. Since we are looking for above, click the right tail button. The upper box is the percentage and the lower box is the Z-score. In this case, we know the Z-score and are looking for the percentage. So in the bottom box type in -2.371 .



StatKey Theoretical Distribution

Normal Distribution ▾

Reset Plot

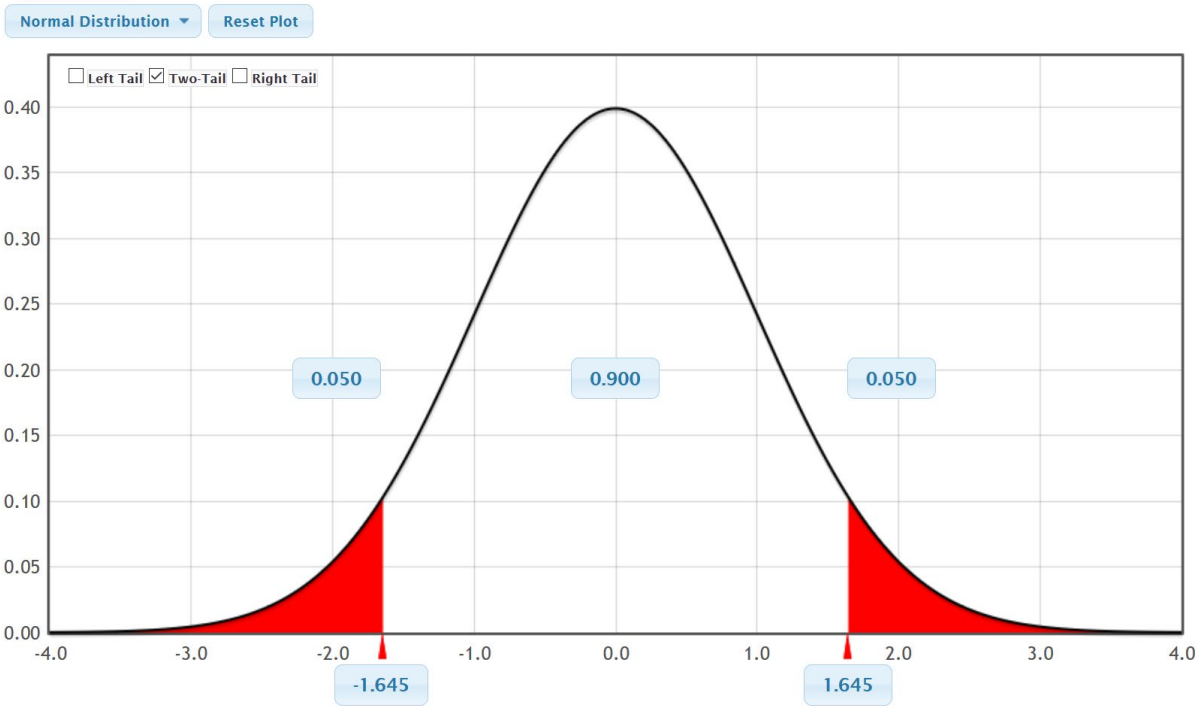


Notice the top box is the answer, 99.1% of normal data values will have a Z-score of -2.371 or higher.

Example: Push the “reset plot” button. Suppose we want to find the two Z-scores that 90% (0.9) of normal data values are in between. Since we are looking for “in between”, click the two-tail button. The upper boxes are the percentages and the lower boxes are the Z-scores. In this case, we know the percentage in between, but need to find the Z-scores. So in the upper middle box type in the decimal proportion equivalent of 90% (0.9).



StatKey Theoretical Distribution



Notice the Z-score answers we are looking for are at the bottom. Therefore, the middle 90% of normal data values have a Z-score between -1.645 and $+1.645$. These are the famous Z-scores for 90% confidence intervals that we will study in later chapters.

Percentages for any normal data

We often want to calculate percentages for normal quantitative data without calculating Z-scores first. StatKey can do that as well. Push the “reset plot” button. Right now, the mean is set at zero and the standard deviation is at one.

Normal Distribution

Mean	Standard Deviation
0	1



Example: Suppose we want to calculate percentages associated with the women's height data we studied earlier. We found that the women's heights were normally distributed with a mean of 63.195 inches and a standard deviation of 2.741 inches. Click on the button that says, "edit parameters" and put those numbers into StatKey.

Normal Distribution

Mean	Standard Deviation
63.195	2.741

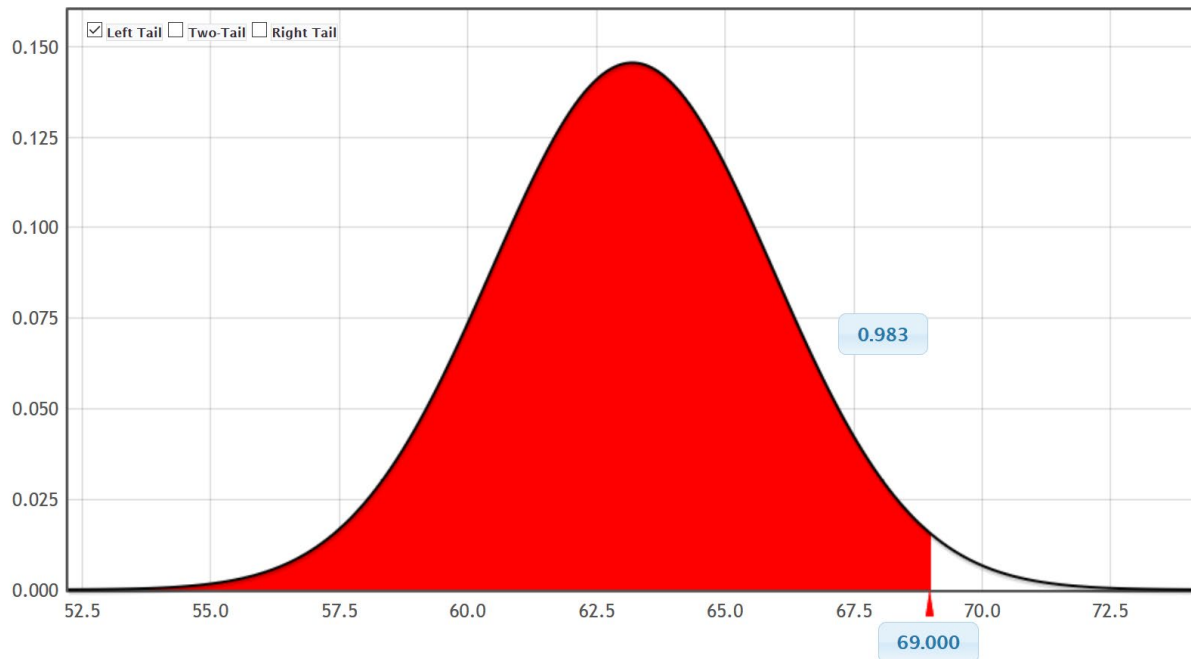
[Edit Parameters](#)

Suppose we want to know what percentage of women in the data have a height of 69 inches or less. Since we are looking for "less than", click on left tail. Remember the top box is the percentage (proportion). The bottom box is now the height. Since we know the height is 69, type in 69 into the bottom box. The proportion in the top box is our answer. So about 98.3% of the women in the sample data have a height below 69 inches.

StatKey Theoretical Distribution

Normal Distribution ▾

[Reset Plot](#)



Note: Be careful about generalizing results of sample data to the population. This does not mean that 98.3% of all women have a height of 69 inches or below. As we learned in chapter one, samples may have bias and not represent the population.

Example: Suppose we wish to find the two heights corresponding to the middle 35%. That is the two heights that 35% of women are in between. Just push the "two-tail" button and put 0.35 in the upper middle box. The answer will be in the two lower boxes.

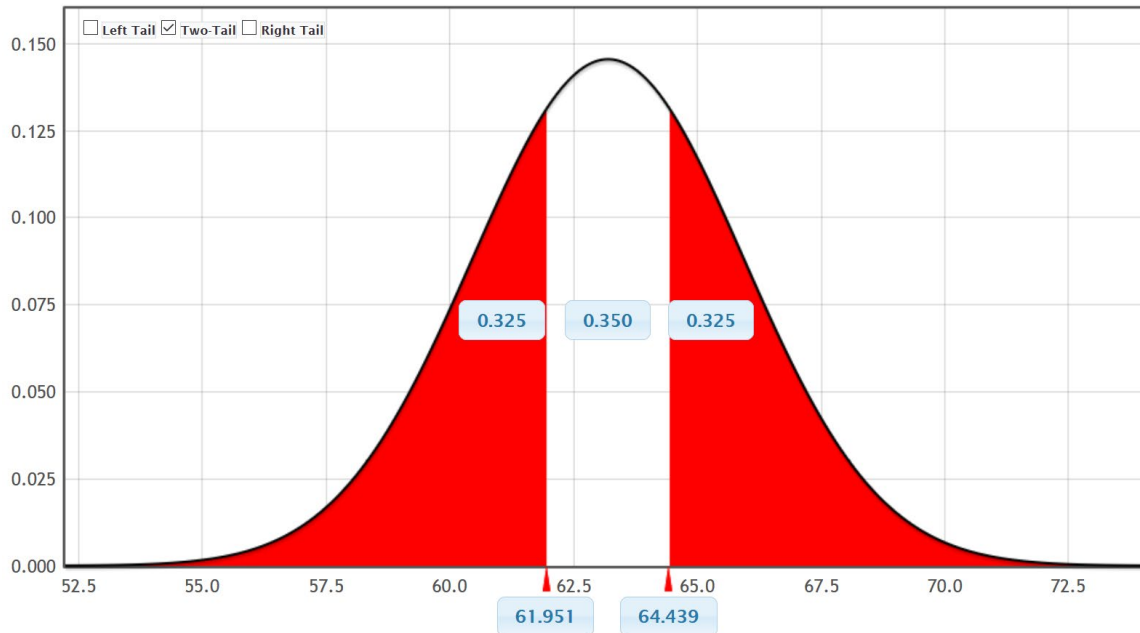


This chapter is from [Introduction to Statistics for Community College Students](#), 1st Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a "CC-By" [Creative Commons Attribution 4.0 International license](#) – 10/1/18

StatKey Theoretical Distribution

Normal Distribution ▾

Reset Plot



So about 35% of the women in the data have a height between 61.951 and 64.439 inches.

Note: These percentages are based on perfectly normal curves, yet real data is rarely perfectly normal. There are actually 15 women in the data had a height between 61.951 and 64.439. This was actually 37.5%. This is off from the theoretical percentage because the data was not perfectly normal. It is important to realize that theoretical distributions rarely match up exactly with real data.

Calculating Percentages for Normal Curves with Statcato

Z-scores, X-values and percentages for normal curves can also be calculated with Statcato. Go to the “Calculate” menu, click on “Probability Distributions” and then “Normal”.



Distribution

Distribution Parameters:

Mean:

Standard deviation:

Compute:

Probability density

Cumulative probability

Inverse cumulative probability

Inputs and Outputs

Input(s):

Column:

Constant:

Store Results in: (optional)

(e.g. C1 for column label, or variable name)

If you leave the mean at zero and the standard deviation at one, then Statcato is set up to calculate Z-scores or percentages from Z-scores. To calculate a Z-score from a percentage less than the Z-score, put in the proportion (decimal equivalent of the percentage) into the box that says “constant”. Then click “inverse cumulative” and “compute”. For example, what is the Z-score that 85% of values in a normal data set are less than? The answer is under “X”. The Z-score is 1.0365.

Normal Distribution: mean = 0.0 stdev = 1.0

Input: 0.85

Type: Inverse cumulative probability

$P(<=X)$ X

0.85 1.0365



Normal Probability Distribution ×

Help F1

Distribution

Distribution Parameters:

Mean:

Standard deviation:

Compute:

Probability density
 Cumulative probability
 Inverse cumulative probability

Inputs and Outputs

Input(s):

Column:
 Constant:

Store Results in: (optional)

(e.g. C1 for column label, or variable name)

Suppose we want to find the percentage less than a Z-score of 2.36. Put 2.36 in the constant box and press "Cumulative Probability". The answer is under "P(<= X)". So the answer 0.990863 or about 99.1%.

Normal Distribution: mean = 0.0 stdev = 1.0

Input: 2.36

Type: Cumulative probability

X P(<=X)

2.36 0.990863



Distribution

Distribution Parameters:

Mean:

Standard deviation:

Compute:

Probability density

Cumulative probability

Inverse cumulative probability

Inputs and Outputs

Input(s):

Column:

Constant:

Store Results in: (optional)

(e.g. C1 for column label, or variable name)

We can also calculate X-values and percentage for those X-values for normally distributed data. We need to input the mean and standard deviation into Statcato. For example, earlier we saw some random sample data for women's heights was normally distributed with a mean of 63.195 and a standard deviation of 2.741. Suppose we want to find the percentage of women in the data that have a height below 64 inches. We see that the answer is 0.615502 or about 61.6%. Note that Statcato can only calculate for less than. If we want to know what percent of women in the data have a height above 64 inches, we first calculate less than and then subtract the answer from 100%. In this case, $100\% - 61.6\% = 38.4\%$.

Normal Distribution: mean = 63.195 stdev = 2.741

Input: 64.0

Type: Cumulative probability

X P(\leq X)

64.0 0.615502



Normal Probability Distribution ×

Help F1

Distribution

Distribution Parameters:

Mean:

Standard deviation:

Compute:

Probability density

Cumulative probability

Inverse cumulative probability

Inputs and Outputs

Input(s):

Column:

Constant:

Store Results in: (optional)

(e.g. C1 for column label, or variable name)

You can also use the “Inverse Cumulative Probability” function to calculate the height that 15% of women are taller than. Remember, Statcato only works with less than, so if 15% of women are greater than this height, then 85% of women are less than this same height. So we will enter 85% (0.85) into the constant box. We see the answer under “X”. Therefore, 85% of women have a height less than 66.0358 inches. This also means that 15% of women have a height above 66.0358 inches.

Normal Distribution: mean = 63.195 stdev = 2.741

Input: 0.85

Type: Inverse cumulative probability

$P(\leq X)$ X

0.85 66.0358



Normal Probability Distribution ×

Help F1

Distribution

Distribution Parameters:

Mean:

Standard deviation:

Compute:

Probability density

Cumulative probability

Inverse cumulative probability

Inputs and Outputs

Input(s):

Column:

Constant:

Store Results in: (optional)

(e.g. C1 for column label, or variable name)

Calculating between is challenging with Statcato. Statcato does not have a between button, so we must use the percentages less than an X-value. If we want to find the two values that the middle 40% are in between, we have to think about the percentages less than each X-value. If 40% is in the middle, then the remaining 60% is divided into the two tails. Therefore, each tail must be 30%. So the X-value on the left will have 30% (0.3) less than. The X-value on the right will have 70% (0.7) less than. Put 0.3 into the "Constant" box and press inverse cumulative. Then put 0.7 into the "Constant" box and press inverse cumulative. For women's heights, we would get that the middle 40% of women's heights are between 61.7576 inches and 64.6324 inches.

Normal Distribution: mean = 63.195 stdev = 2.741

Input: 0.3

Type: Inverse cumulative probability

$P(<=X)$ X

0.3 61.7576

Normal Distribution: mean = 63.195 stdev = 2.741

Input: 0.7

Type: Inverse cumulative probability

$P(<=X)$ X

0.7 64.6324



Note on Rounding Statistics for Quantitative Data

It is often best to not round if you are unsure. Data analysts usually prefer better accuracy and can round to their own specifications. Rounding too much interferes with accuracy. If you must round, here are some general guidelines.

Percentages and proportions are usually rounded to three significant figures. Proportions are rounded to the thousandths place and percentages are rounded to the tenths place.

Quantitative statistics like the mean or standard deviation are usually rounded to one more decimal place to the right than the original data has. Notice the women's heights data is rounded to the tenths place (one number to the right of the decimal). So statistics calculated from this data would usually be rounded to the hundredths place (two numbers to the right of the decimal).

Mean (women's height) = $63.195 \approx 63.20$ inches

Standard Deviation (women's height) = $2.741 \approx 2.74$ inches

B
Women Ht (in)
64.3
66.4
62.3
62.3
59.6
63.6

