

Organized Data

Provides service and convenience

Enables good decision making

Persuades

Saves time and money

Data Table

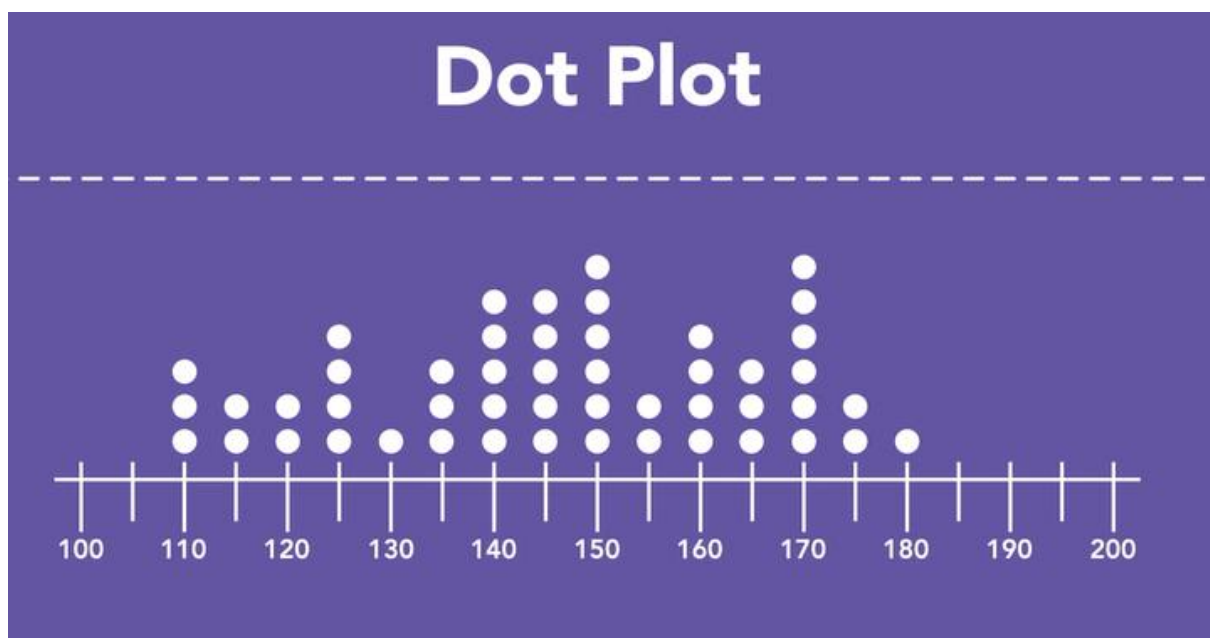
This is a data set of people and their weight in pounds

Data Table			
Adult	Weights	Adult	Weights
1	180	26	145
2	175	27	145
3	175	28	145
4	170	29	145
5	170	30	140
6	170	31	140
7	170	32	140
8	170	33	140
9	170	34	140
10	165	35	135
11	165	36	135
12	165	37	135
13	160	38	135
14	160	39	130
15	160	40	125
16	160	41	125
17	155	42	125
18	155	43	125
19	150	44	120
20	150	45	120
21	150	46	115
22	150	47	115
23	150	48	110
24	150	49	110
25	145	50	110

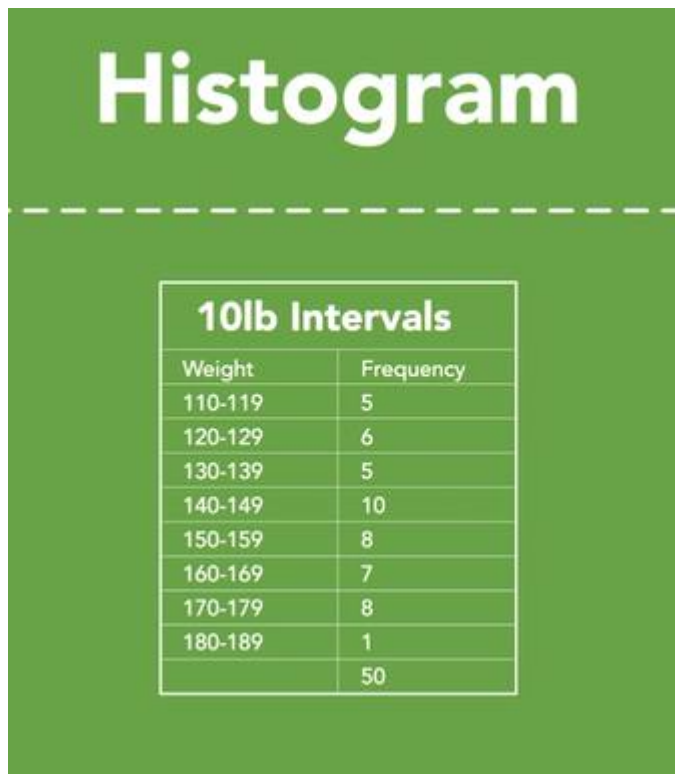
The next table shows the frequency of different weights e.g. how many people weight 130 pounds which is 1 per the table below.

Frequency Table	
Weight	Frequency
180	1
175	2
170	6
165	3
160	4
155	2
150	6
145	5
140	5
135	4
130	1
125	4
120	2
115	2
110	3
	50

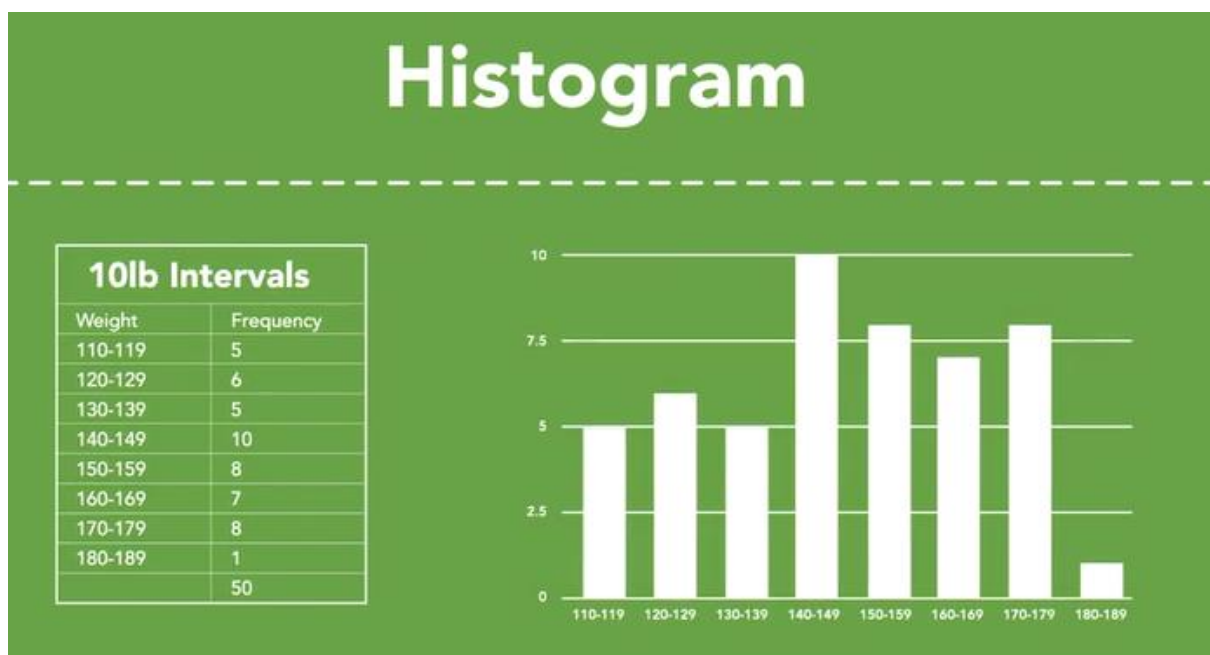
The next figure is another way of presenting the frequency data i.e. by the use of a **dot plot**. This makes the data more visual than the table and is easier for our eyes and minds to visualise.



The next figure we will look at further refines the data by grouping it into intervals e.g. 10 pound intervals in this case. This is called a Histogram



The addition of the chart on the right hand side again makes it more visual for us representing the number of adults who fall into each particular category in a bar chart.



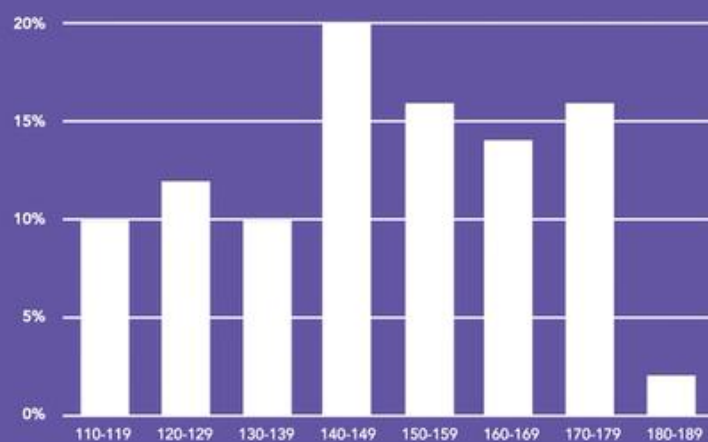
Let's break the data down further by using percentage values e.g. below where 10 adults fall into the 140-149 pounds category which is 20% of the total sample size of 50 adults.

Relative Frequency Table

Relative Frequency Table		
Weight	Frequency	
110-119	5	10%
120-129	6	12%
130-139	5	10%
140-149	10	20%
150-159	8	16%
160-169	7	14%
170-179	8	16%
180-189	1	2%
	50	

Now if we represent the relative frequency as a histogram i.e. bringing in the % values to the Y axis we get the figure below.

Relative Frequency Histogram

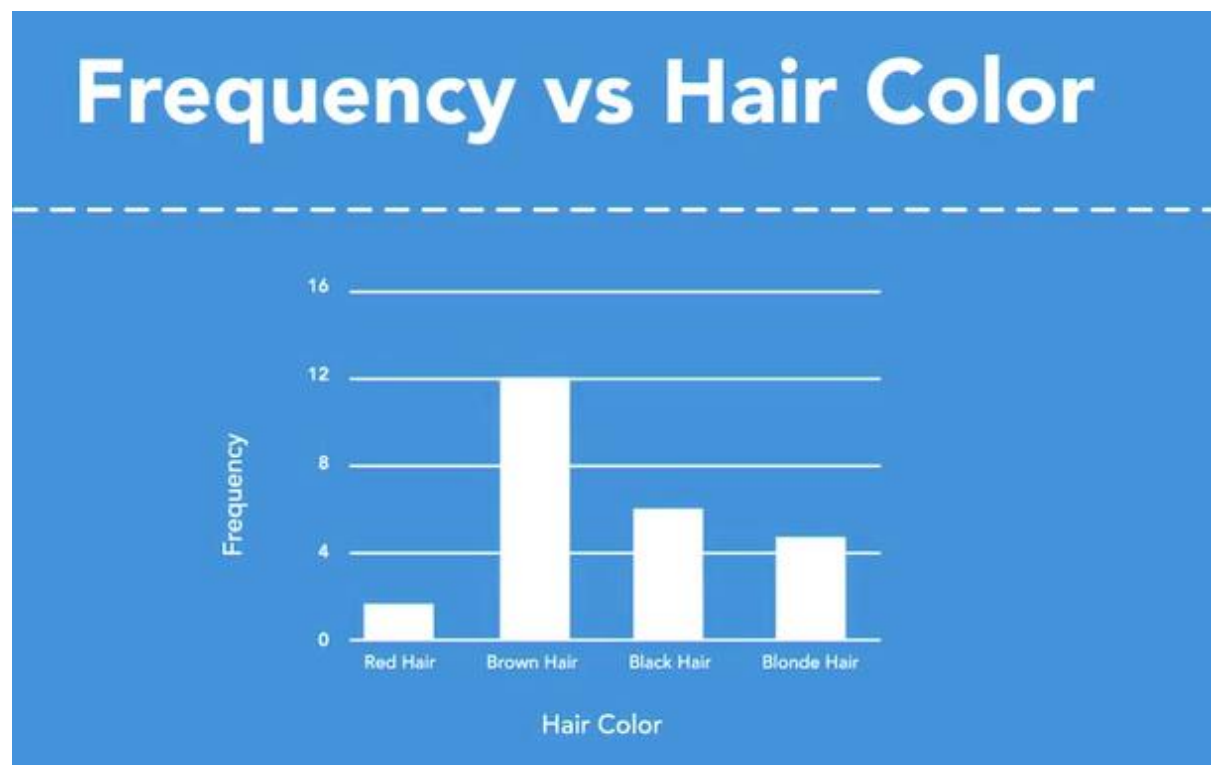


Now let's look at some qualitative (attribute) data for our sample set e.g. hair colour in this case. This data is not quantitative (variable) like the weight values we looked at before.

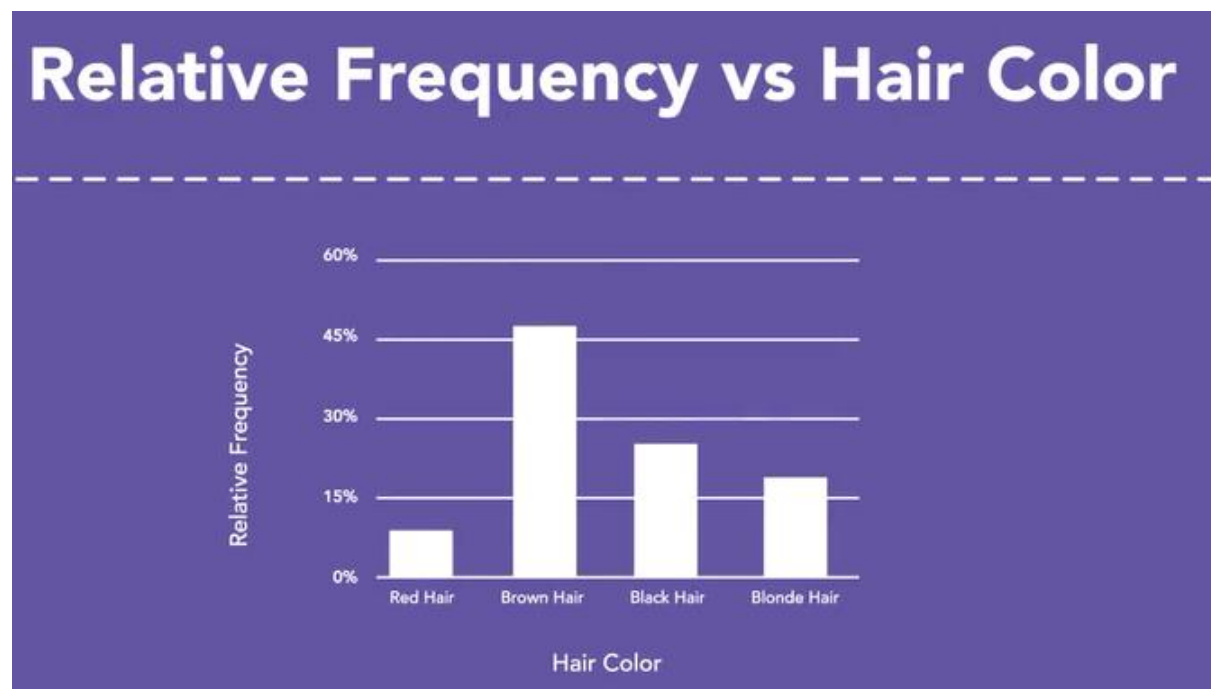
Qualitative Data

Hair Color	Frequency
Red Hair	2
Brown Hair	12
Black Hair	6
Blonde Hair	5
Total	25

Here is the histogram for the hair colour data



Here is the histogram showing the relative frequency of hair colour.



This can also be represented by the use of a Pie Chart which gives a good visual representation of the number of adults who fall into each hair colour category and what % of the relative frequency each hair colour group occupies.

Next let's look at terms like the mean and median values of a dataset.

Mean (Average)

The sum of all data points divided by the total number of observations

We calculate the mean value below for a data set of 25 test scores. The mean value of all the exam results is 65%.

This might indicate that the exam was very difficult as 65% seems quite a low score.

Calculating the Mean Example

DATA SET 1 - 25 test scores				
96	82	78	56	40
95	81	77	51	37
90	80	76	47	34
89	80	69	46	28
83	79	64	42	25

Sum of all test scores 1625

Total # of exams 25

Average 65

Another way to look at the data is through the use of the middle value or median

Median

The midpoint in data with equal number of data points above and below

For this example of 25 test scores the 13th value is the midpoint of the data or median i.e. the 13th value has 12 values lesser than it and 12 values greater than it.

In this case the median value of 76% indicates that the exam was much fairer in that this is a high test score to get.

When we look at the data more closely we see that of the 12 values below 76 only a handful are very low values i.e. the poorest performing 5 or 6 people in the class brought down the class average overall but the median or midpoint value of 76% percent is a fairer reflection of the overall performance of the class.

Finding the Median Example

Student	Score	Student	Score
1	96	14	69
2	95	15	64
3	90	16	56
4	89	17	51
5	83	18	47
6	82	19	46
7	81	20	42
8	80	21	40
9	80	22	37
10	79	23	34
11	78	24	28
12	77	25	25
13	76		

MEDIAN 76

In the example below what do we do when there is not a clear median value on first viewing of the data i.e. 5 has four values below it and 5 values above it and 6 has 5 values below it and 4 values above it.

Data Point	
1	10
2	15
2	15
4	20
5	20
6	30
7	35
8	35
9	40
10	45

Therefore the median value lies somewhere between 20 and 30 and is 25 per the calculation below.

TEN DATA POINTS	
Data Point	
1	10
2	15
2	15
4	20
5	20
6	30
7	35
8	35
9	40
10	45

Actual Mean
 $20 + 30 = 50$
 $50 \div 2 = 25$

Next we will look at the weighted mean which is used when we are combining data sets together and wish to apply a greater relative importance or % value to one data set over the other when we combine them together to give an overall result per the example below.

Calculating Weighted Mean

Calculating Weighted Mean	
Category	Weight
EXAMS 1	30%
EXAMS 2	30%
Quizzes	10%
Homework	10%
Term Paper	20%

In this example the overall course consists of two exams plus quizzes, homework and a term paper which are all assigned % values reflecting their overall importance or weighting in calculating the overall grading for the course.

Calculating Weighted Mean

Calculating Weighted Mean			
Category	Weight	Score	Weight Score
EXAMS 1	30%	90	27.0
EXAMS 2	30%	80	24.0
Quizzes	10%	75	7.5
Homework	10%	100	10.0
Term Paper	20%	85	17.0

$$.30 \times 90 = 27.0$$

$$.30 \times 80 = 24.0$$

$$.10 \times 75 = 7.5$$

$$.10 \times 100 = 10.0$$

$$.20 \times 85 = 17.0$$

Weighted Mean 85.5

In this example multiplying the weight x score gives the weight score for each section of the course on the right hand side and adding these weight scores together gives 85.5% which is the weighted mean or the student's overall result for the course.

We use weighted means to instruct others regards the value we place on different elements of a data set e.g. above the written exams are worth 60% of the overall result but a person can get up to 40% of the overall marks by participating well in class and doing their homework and assignments well.

Creating a Weighted Mean: Considerations

Categories

Weight

You need to carefully consider the categories and weight chosen as the recipients of your data e.g. students in the example above will ask you detailed questions as to why you prioritised one section over another.

Also your categories and weight values should promote the behaviour you want to encourage in your audience. See the example below.

Rate These Rubrics

Example 1

Attendance	40%
Follows directions	30%
Customer service	10%
Sales	20%
	100%

Example 2

Innovative	30%
Technical skills	25%
Professional Certifications	15%
Team evaluations	30%
	100%

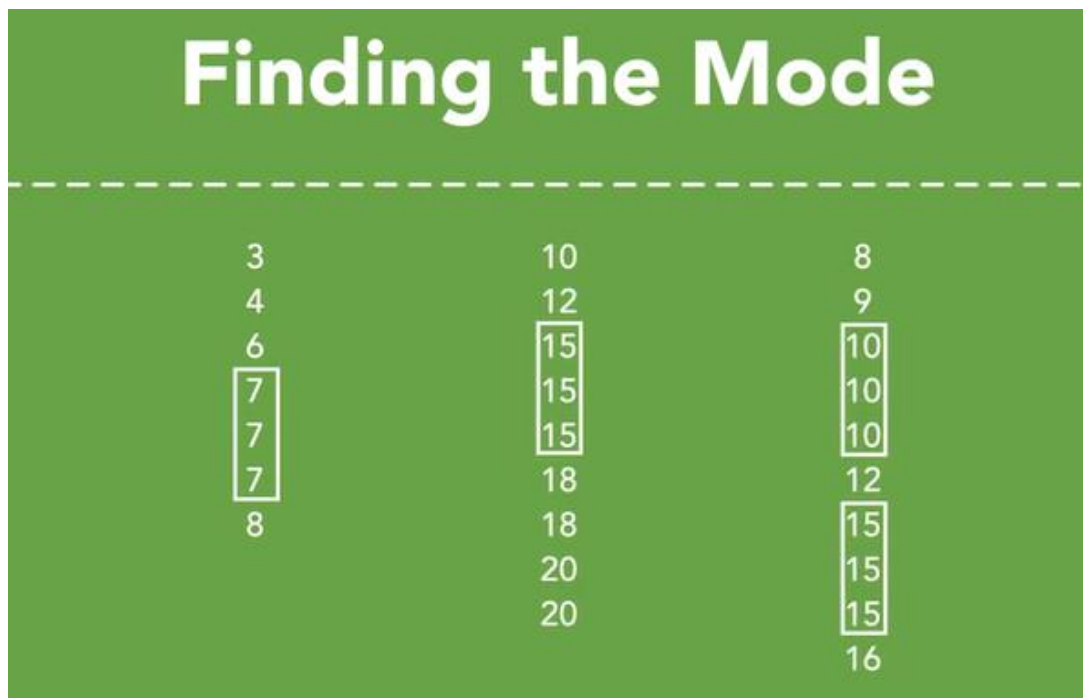
What type of direction do Examples 1 and 2 below give to e.g. Employees in this case.

We will next look at another point in our dataset the Mode. This does not require calculation but is simply;

Mode

The data point that is most prevalent in the data set

The 3rd set of data below has two modes 10 and 15.



In the example below since every data point appears just once it does not have a mode.



Why do we reference the mode? In some cases the mode can represent the most likely outcome in a dataset (see the example below).

Mode: No Minimum Frequency

1	5
3	8
4	8
4	12
4	14
4	14
4	14
4	

1	10
3	11
4	14
4	15
5	17
7	18
8	19
9	

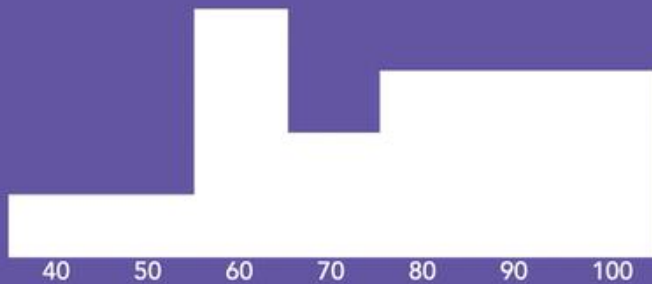
In this example the first data set has 4 as the mode. It occurs 6 times and could be said to represent the most likely outcome in the data set.

In the 2nd dataset 4 is still the mode but it only occurs twice and would not represent the most likely outcome.

The above example shows that the mode is simply the value that occurs most often irrespective of whether it is 6 or 2 times as shown above.

It is therefore important to consider the mean and median in conjunction with the mode value when making decisions regards which value is of greatest importance in a dataset.

Histograms and Mode



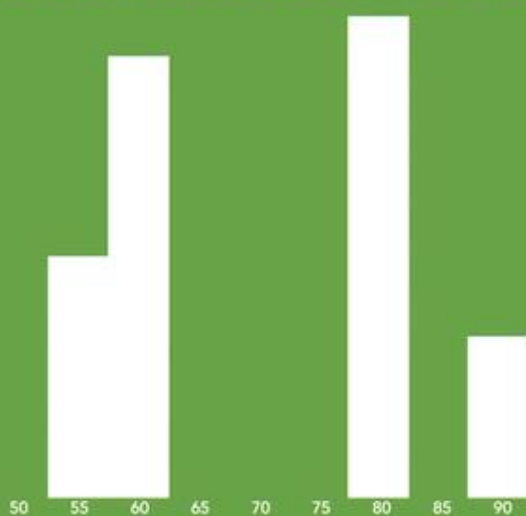
Score	Frequency	Total	
40	1	40	
50	1	50	
60	4	240	
70	2	140	
80	3	240	
90	3	270	
100	3	300	
	17	75.3	Average
		80	Median
		60	Mode

In this example 60 is the mode but the average value of 75.3 and the median value of 80 are more representative of the most frequently occurring values in the dataset.

This is probably the best way to use the mode i.e. in conjunction with the mean and median values to give a clearer overall picture of the spread of values in a dataset.

In the example below the mode is 80 the mean is 70 and the median is 60 with all three values contributing to our understanding of the overall dataset.

Data Set 2



Data	Frequency	Total	
50	0	0	
55	6	330	
60	11	660	
65	0	0	
70	0	0	
75	0	0	
80	12	960	
85	0	0	
90	4	360	
	33	70.00	Average
		60	Median
		80	Mode

The Range

Range

The difference between the largest and smallest values

In the small dataset below the range is 50.

Range

20	60	51	62	57
45	55	70	48	51

$$70 - 20 = 50$$

Below the range is 150

Range

75	35	66	127	96
60	105	63	185	66

$$185 - 35 = 150$$

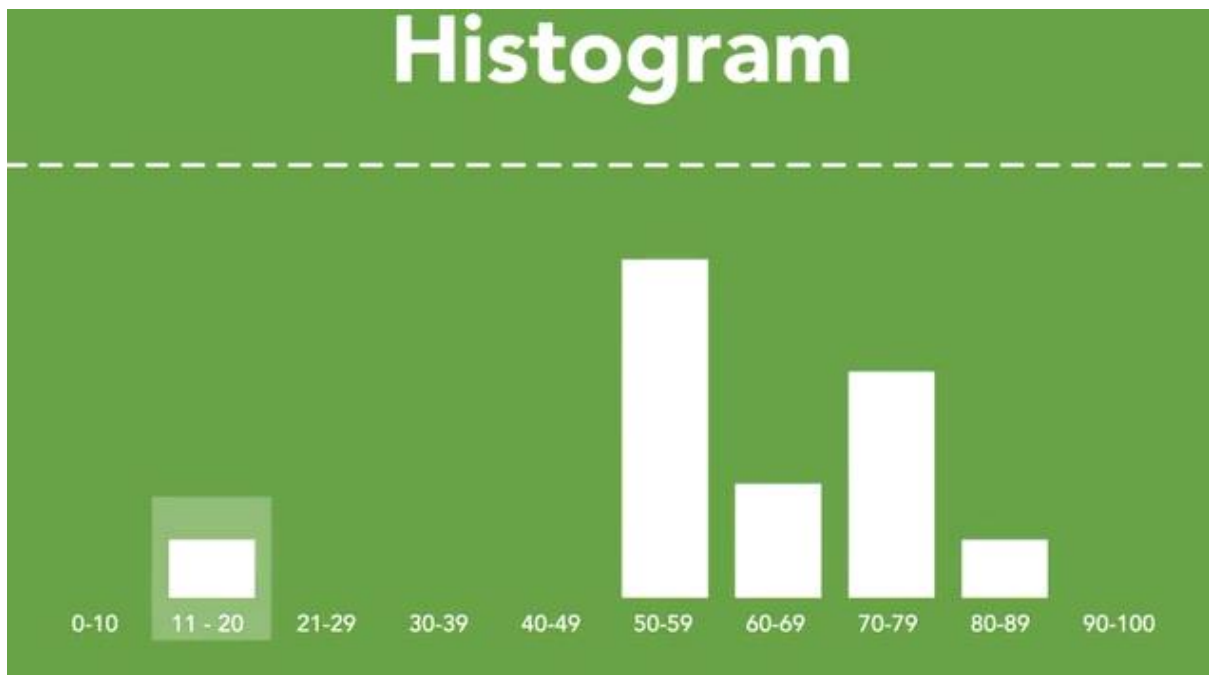
Be careful about making assumptions. A mean of 60 with a range of 70 could make you assume that the statement below is true but the data behind it could tell a different story.

Mean 60
Range 70

$$60 + 35 = \text{Highest range } 95?$$

$$60 - 35 = \text{Lowest range } 25?$$

A Histogram can better explain the make up of a range i.e. below there is one student who got a very low score which really opened up the range i.e. one rogue data point can really exaggerate what the range seems to be telling you if the range is viewed in isolation.



Standard Deviation

Standard Deviation

Sort of the average distance from the mean

The actual definition is below

Standard Deviation

The average squared distance from the mean

In the example below the standard deviation helps us to deduce that although the mean value is the same for the three data sets and the range is the same value for data sets 2 and 3, data set 2 has more outlier values in it which gives it the highest standard deviation value of the three data sets.

	Data Set 1	Data Set 2	Data Set 3
	15	2	3
	18	3	18
	25	3	16
	16	23	15
	14	38	16
	20	39	40
AVERAGE	18	18	18
StDev	4.05	17.73	12.05
Max	25	39	40
Min	14	2	3
RANGE	11	37	37

The values below when all considered together give us a fuller picture of what the data is really telling us.

Statistics Tools

Mean

Range

Median

Standard Deviation

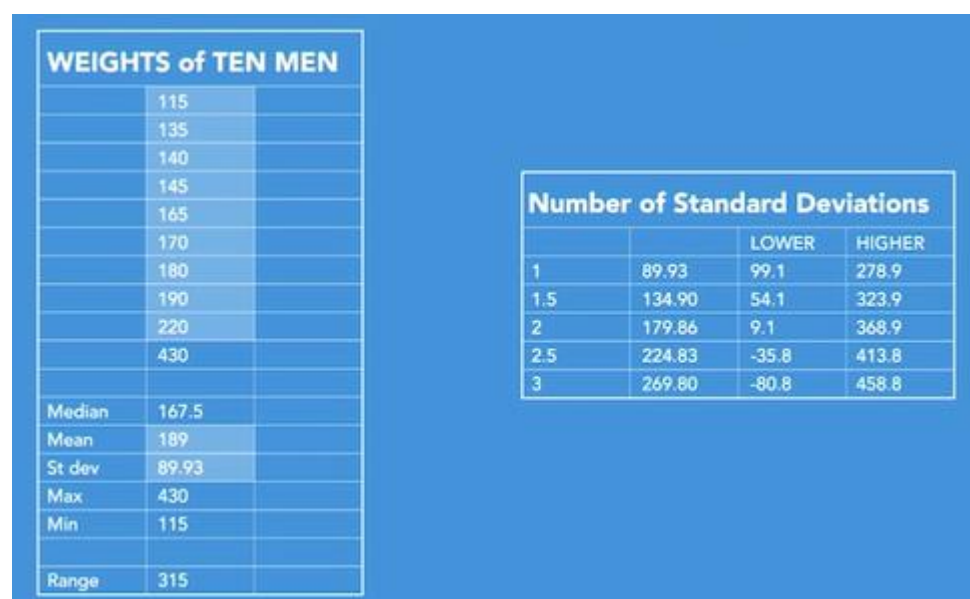
Mode

What can Standard Deviation tell us about individual values in a dataset? Consider the example below.

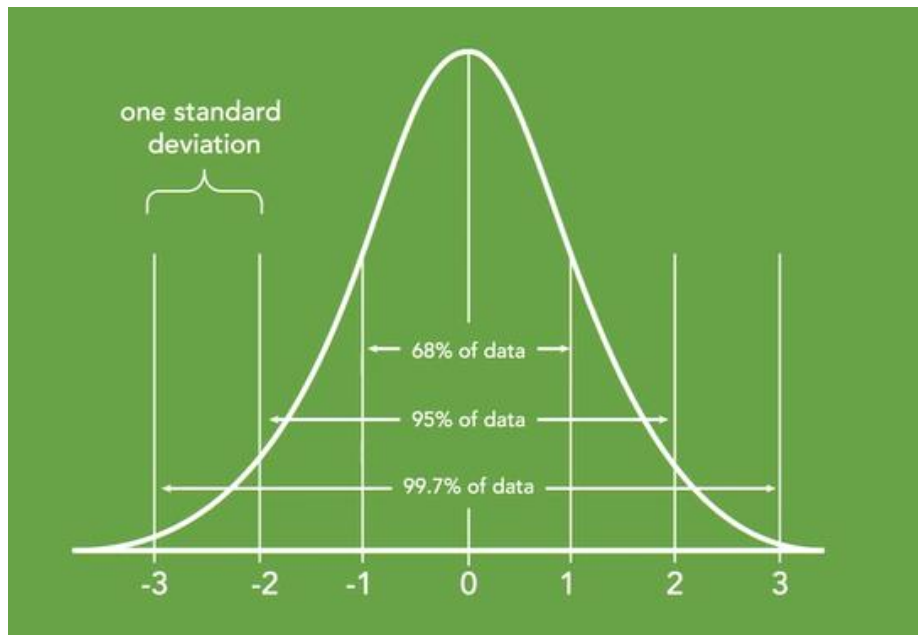


The first nine values are within one standard deviation of the mean per the box on the right above.

For the 10th value i.e. 430 pounds that does not fall within our statistical range until we get to 3 Standard Deviations from the mean.

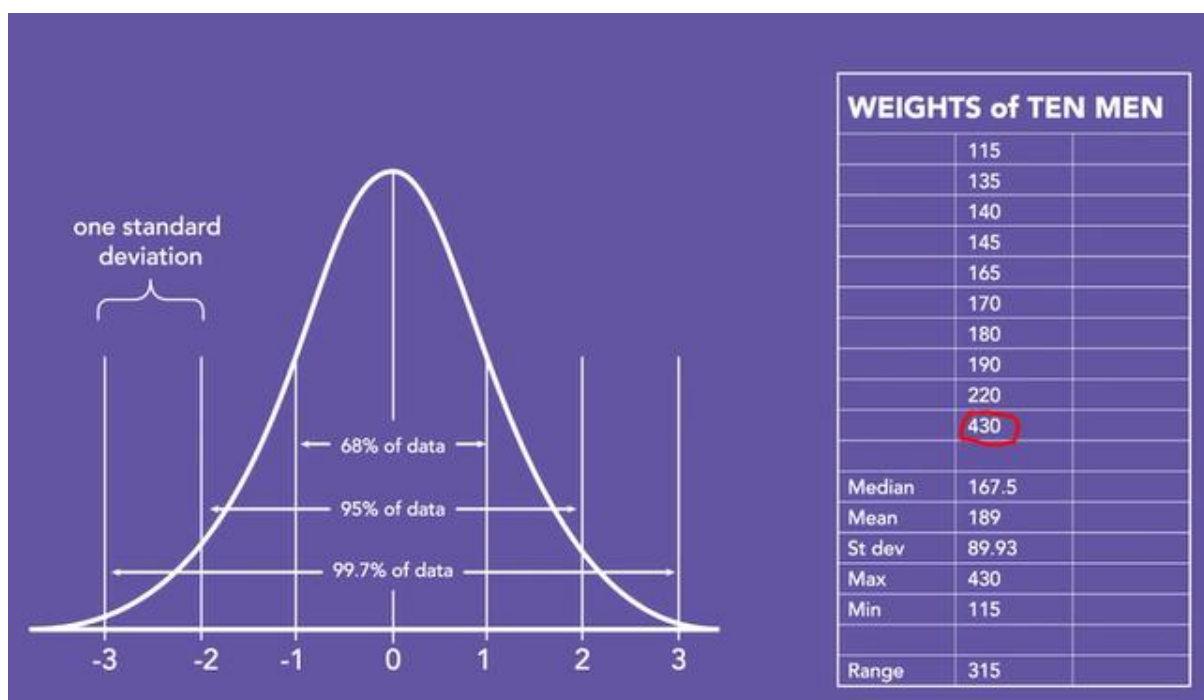


Normal (Bell Shaped) Distribution Curve



The Normal Distribution tells us that 68% of our data should fall within 1 standard deviation of the mean as shown above. In our example 90% of our data points fell within 1 standard deviation of the mean which gives us a good level of confidence about the consistency of our data.

Our last data point falls within the 3rd standard deviation from the mean as shown in the graphic below. Therefore it can be considered an “outlier value” in relation to the other data points which all lie between 1 standard deviation from the mean.

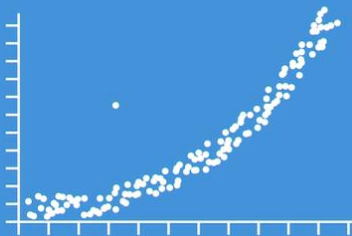


What is an Outlier?

Outlier

A data point that is an abnormal distance from other values in a data set

Outliers Are Opportunities



An outlier value could possibly be the beginning of a new trend and should always be given some consideration rather than being simply ignored.

It is good to ask some searching questions about outliers e.g. if a student has performed way above the average in an exam did they e.g. receive extra tuition, study other subjects that really complimented and improved their performance in this subject. This could allow us to make improvements to the course offering for other students to help improve their exam result in this subject as well.

Is this really an outlier?

How did this happen?

What can be learned?

What needs to change?

Z Score i.e. how many standard deviations from the mean does our data point lie? This is a simple calculation based on our three values below.

Z-Score

$$z\text{-score} = \frac{x_i - \bar{x}}{s}$$

x_i = data point

$\bar{x}$ = mean

s = standard deviation

For example go back to the weight of ten men dataset on p. 21. Performing the calculation above for the highest and lowest value gives us

Highest value

$$Z \text{ score} = \frac{430 - 189}{89.93} = 2.67 \text{ standard deviations from the mean on the positive side}$$

Lowest value

$$Z \text{ score} = \frac{115 - 189}{89.93} = -0.82 \text{ standard deviations from the mean on the negative side}$$

The Empirical Rule

When you have the bell shaped curve i.e. the normal distribution 99.7% of your data falls within three standard deviations or 3 Sigma of the mean.

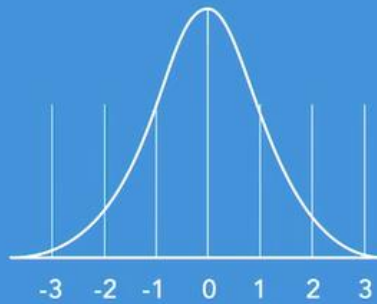
Empirical Rule

Works for
symmetrically
distributed data



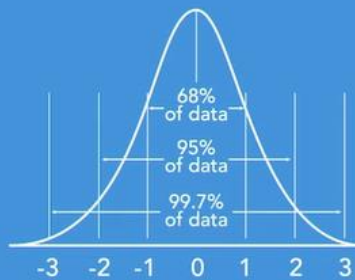
Empirical Rule

Symmetrically distributed data follows a pattern whereby most data points fall within 3 standard deviations of the mean



Empirical Rule

99.7% of data points are within 3 StDev



The Empirical or 3 Sigma rule works best for data that falls into the symmetrical or bell shaped distribution. It is very useful to combine your z values with the Empirical rule to evaluate where your data point lies in the overall % distribution of data e.g. a z value of 1.8 falls within 2 sigma of the mean i.e. lies within 95% of all the data in the distribution.

However a z value of 3.0 lies beyond 99.7% of the data in your distribution and can be considered to be an outlier.

Percentile Rank

This term is used to describe where the data point lies within the overall data set i.e. if a test score is within the top 4% in the class it is in the 96% percentile. However if a student scores 85% in an exam it does not mean that they are in the top 15% in the class as if all the other students scored above 85% it would place the student at a much lower percentile level.

Percentile Rank

$$\% \text{ rank} = \left[\frac{(\# \text{ of values below } x) + 0.5}{\text{Total \# of values}} \right] 100$$

Percentile Rank

Student	Score
1	97
2	96
3	96
4	94
5	91
6	90
7	90
8	85
9	83
10	82

$$\% \text{ rank} = \left[\frac{(\# \text{ of values below } x) + 0.5}{\text{Total \# of values}} \right] 100$$

$$\% \text{ rank} = \left[\frac{2 + 0.5}{10} \right] 100$$

= 25th percentile

i.e. there are two values lower than 85% in the class.

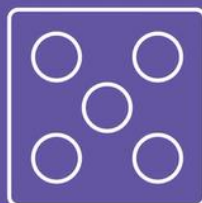
Probability

Look at the ratio below i.e. the desired outcome divided by all possible outcomes e.g. the probability of rolling a 5 from a dice is 1 in 6.

Probability

The likelihood that some event will occur

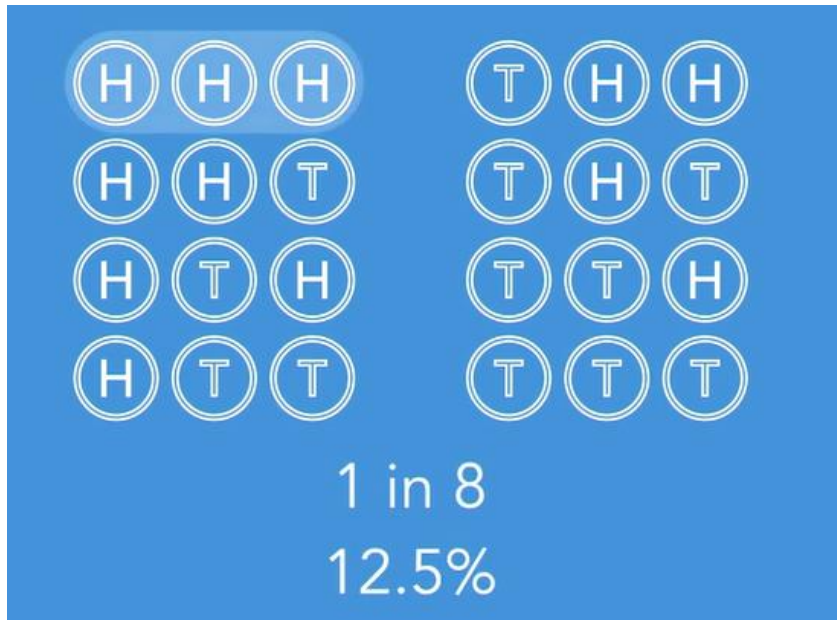
$\frac{\text{Desired outcome}}{\text{All possible outcomes}}$



Random experiment: Tossing a coin

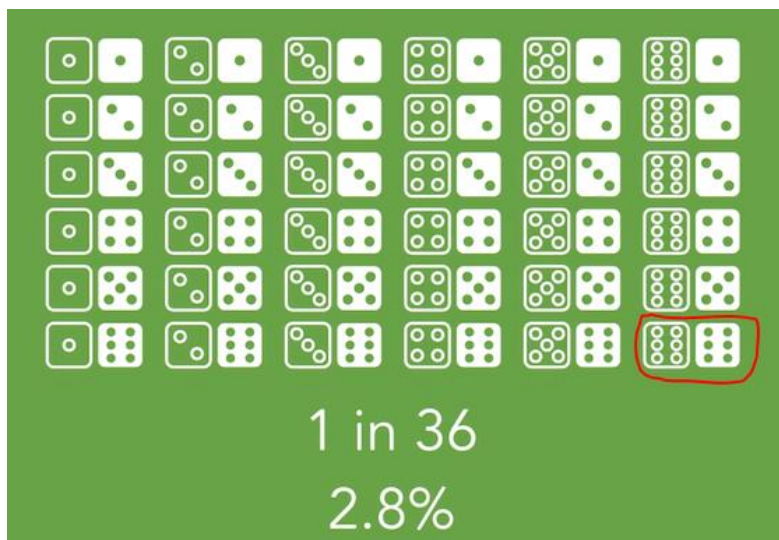
What are the chances of being able to toss 3 heads in a row?

First we look at how many possible outcomes there are when we toss a coin three times. The number is 8 as shown in the graphic below.



Therefore the chance of throwing 3 heads in row is 1 in 8 or 12.5%.

If we try this with rolling a pair of dice e.g. what is the chance of rolling a double 6? This is a 1 in 36 or 2.8% chance per the graphic below.



NB: Probability **does not guarantee an outcome** it simply informs you of the **possibility that a particular outcome** will occur e.g. the chances of winning the Lotto are extremely low but you could be the lucky person who does a quick pick at the local garage and gets the winning numbers.

The aim of using probability is to **understand all the different factors** that could affect an outcome and if these are well understood to calculate a probability of the event occurring which allows an **informed decision to be made** of the action to be taken e.g. the chance of getting 4 numbers in the Lotto is much greater than getting 6 for the overall jackpot. However getting 4 numbers might only produce a prize of €20 so my decision could be that the potential reward I might get is not worth the amount of money spent each week in playing the Lotto.

The probability of all possible outcomes must sum to 100% i.e. for our coin tossing or dice rolling examples below

Total Probability of a Sample Space

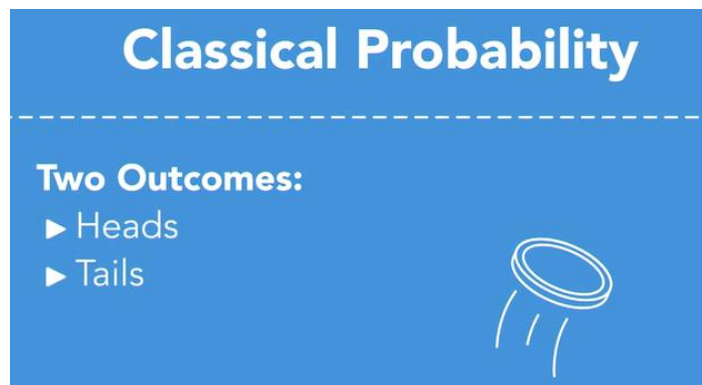
The sum of probabilities of all possible outcomes must add up to 100%.



The probability of an event occurring can never be less than 0% or greater than 100%.

The highest probability for any scenario is 100%.

The lowest probability for any scenario is 0%.



Classical Probability works well when you know all possible outcomes and each outcome is equally likely to occur e.g. rolling a 6 sided dice;

The % chance of rolling e.g. a 1 is $1/6 = 16.6\%$

The % chance of rolling e.g. a 1 or a 2 is $2/6 = 33\%$

But what happens when things are not all fair and equal. For this we turn to empirical probability. Since every situation is different we can only rely on the observations we have made up to this point. The greater the number of observations made the more confident we feel about the validity of the predictions we make.

An example would be the number of penalties scored by a soccer player in the Champions League. If we look at last season's data we could calculate i.e.

$$\frac{\text{the number of penalties scored}}{\text{the number of penalties attempted}} = \frac{6}{10} = 60\%$$

However the quality of the number of penalties attempted can vary based on a number of factors e.g. player health, fatigue, time of the match when penalty was won (a penalty in the first few mins could carry less pressure than one in the last minute). Also if the penalty was won in the group stage which is less pressurised than the knockout stage.

For this reason you might want to break down your analysis into chance of scoring a penalty in the group stage v the chance of scoring in the knockout stage when you are trying to predict the chance of scoring at a particular stage of this season's competition.

Consider the types of probabilities you will encounter

Classical

Empirical

Subjective

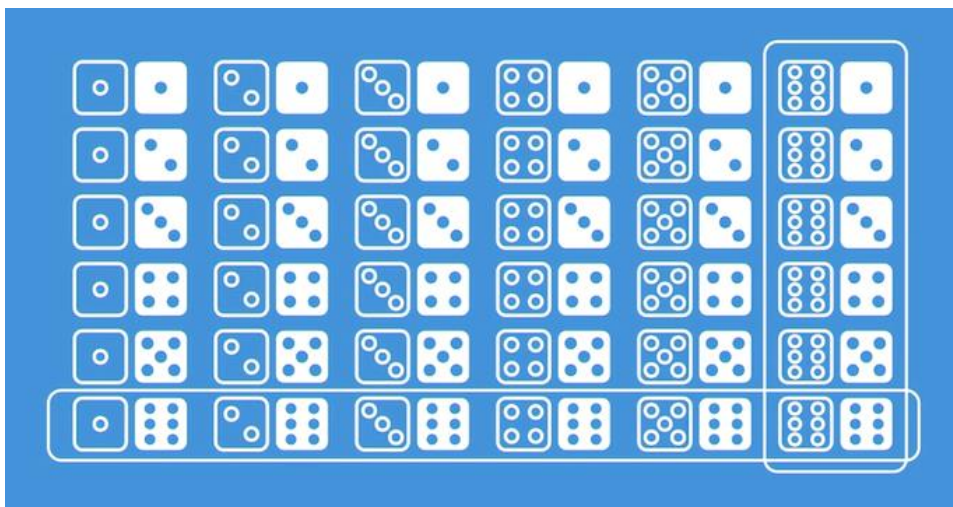
Classical is based on well-defined data where the number of possible outcomes is known.

Empirical is based on past performance but needs to consider the situation in which the information was gathered before making a prediction about future performance

Subjective data is more difficult to predict than the above two and relies on opinions and judgement calls of experienced people when examining past data and using their own knowledge to try and make future predictions. This can border on guesswork e.g. an economist trying to predict the unemployment rate in the United States for the coming year when the previous year's data or 5 years data may vary a lot and not be a good indicator of future performance in this area.

Probability of two events: Either event? Both events?

Roll two 6 sided dice. What is the probability of rolling a 6 on the 1st or 2nd throw.



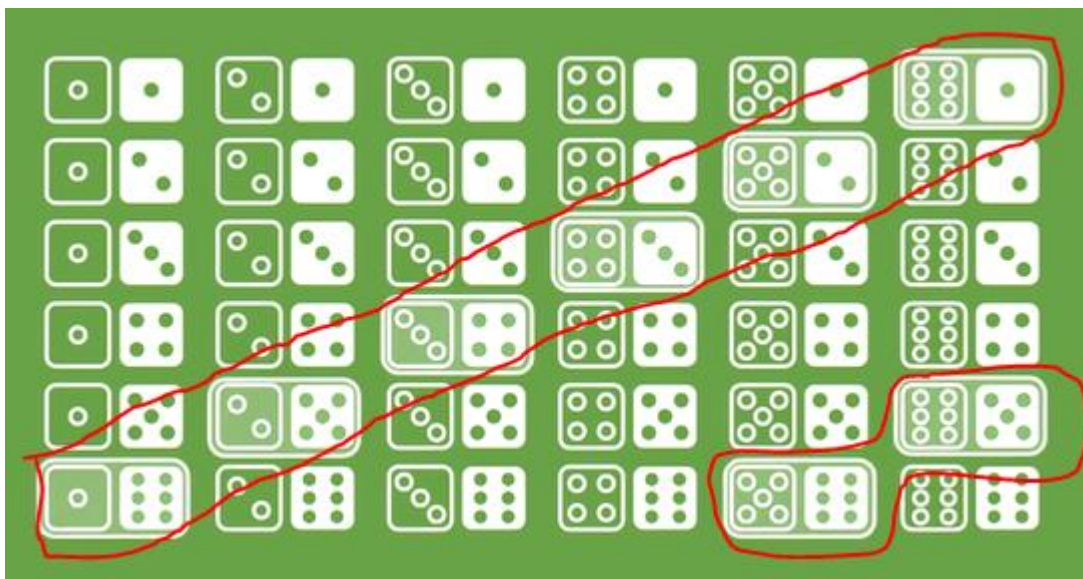
There are two sets of 6 events where a 6 is rolled on the 1st or 2nd throw. However the scenario in the bottom right hand corner is an overlap of the two sets. Therefore the calculation is as per below

$$\frac{6}{36} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}$$

$$= 30.56\%$$

i.e. add the probability of each set and subtract the overlapping value to give the final % outcome.

Now let's take a scenario where the values do not overlap i.e. throw 2 dice and count the outcomes where the sum of the two dice = 7 or = 11



$$\frac{6}{36} + \frac{2}{36} = \frac{8}{36}$$

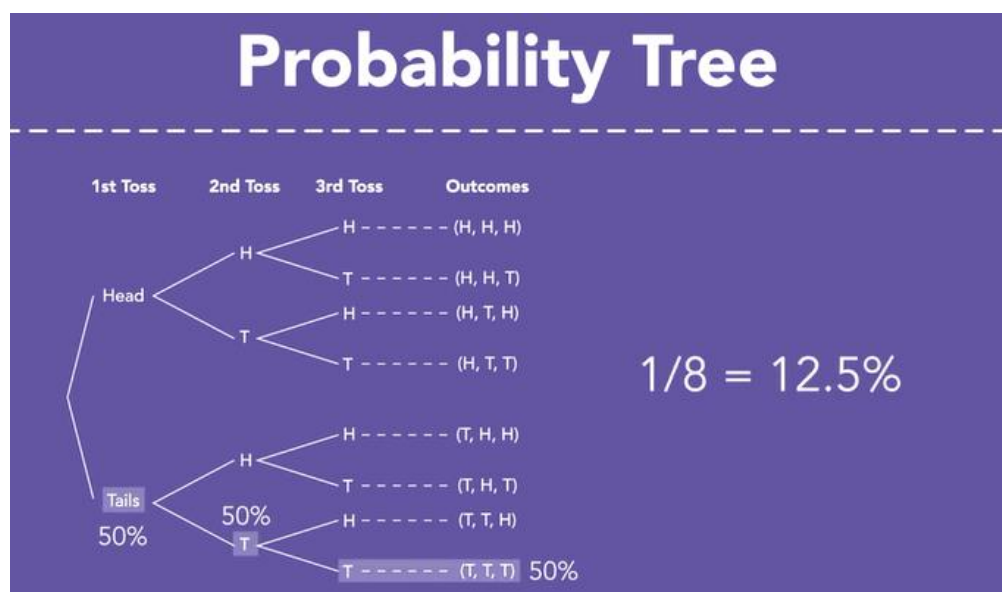
$$22.2\%$$

You can add probabilities and you can also subtract probabilities per the examples above.

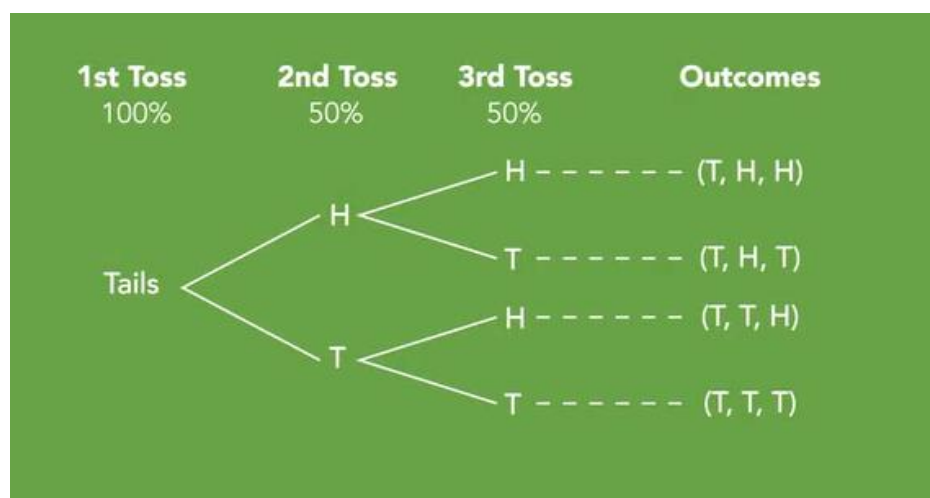
Conditional Probability

The probability of an event, given that another has already occurred

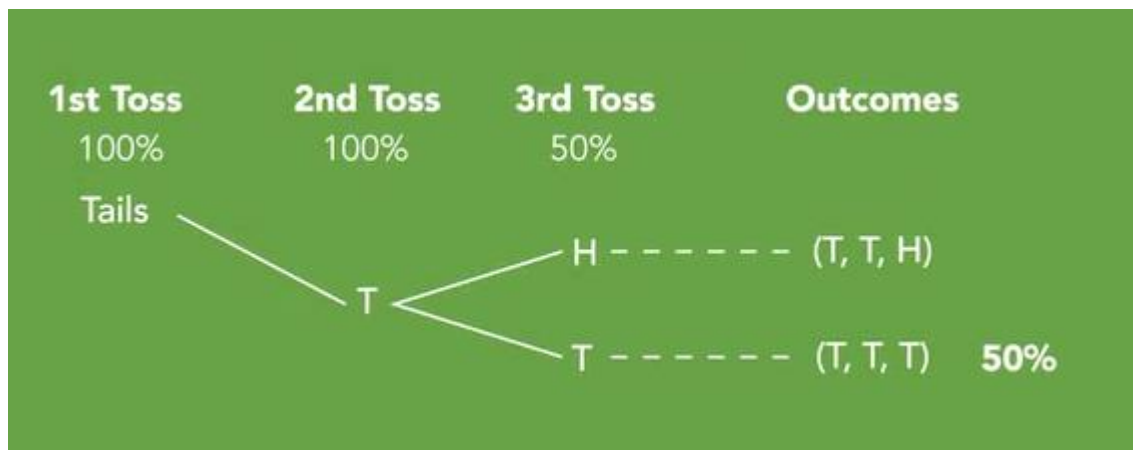
It often helps to draw a Probability Tree of all the possible outcomes as shown below.



The odds that tails will come up three time in row are 1 in 8 = 12.5%



Once we flip the first tails the odds of three in a row is now increased to 1 in 4 = 25%.



After the 2nd tail is flipped then the odds jump to 50%.

Think of a more nuanced real life example e.g. in a Golf Major tournament a player starts out with bookies odds of 70:1 of winning at the end of the 4 day tournament. After day 1 if he is in the top 10 scorers his odds of winning the tournament will be reduced.

As an exercise can we draw a probability tree to factor in how we would reduce his odds after day 1 of play. We will do this as a small class assignment.

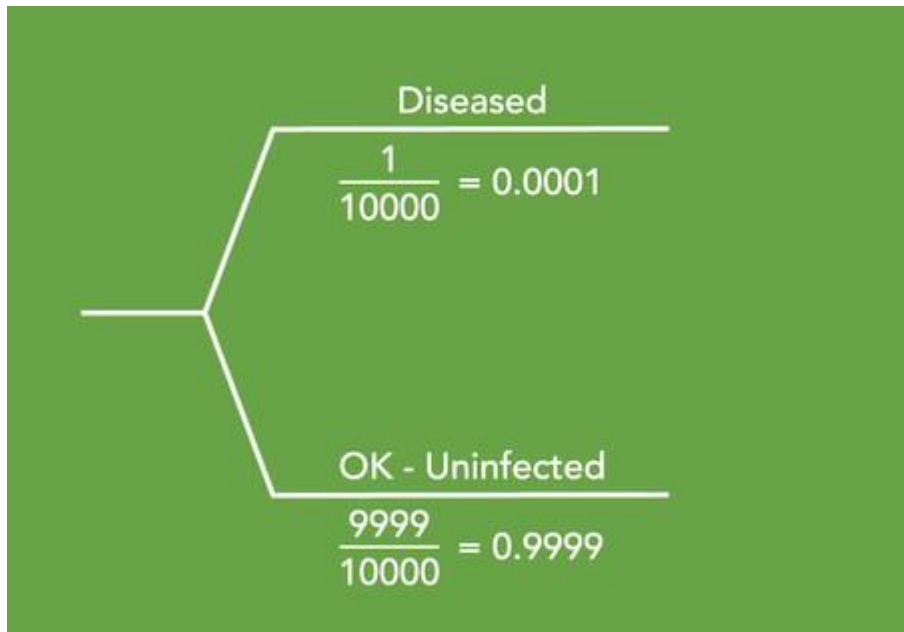
Relationship between two events: Independence vs. dependence

I need to get a better example of this i.e. the mathematical proof that shows that the probable outcome of two events are separate from each other.

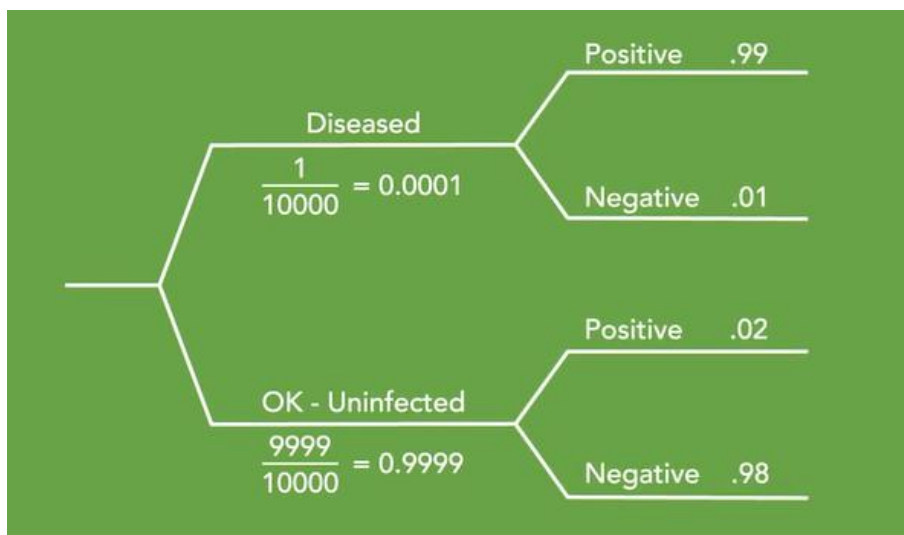
Bayes's theorem and false positives

Two scenarios presented below. One in Ten thousand people have a probability of contracting a rare disease as seen in the top branch below = 0.0001

In the bottom branch are the 9999 uninfected people i.e. 0.9999 is the probability of not having contracted the disease.



However if you do test positive for the disease what is the probability that you have received a false positive result i.e. that you do not actually have the disease.



If you do have the infection there is a 99% chance that you will test positive for the disease as seen in the top branch above.

The bottom branch tells us that of the negative test results there is a 98% chance of this being the correct diagnosis and a 2% chance that the person will test positive of the disease even though they are uninfected. This is called a false positive result.

						1 MILLION PEOPLE	
DISEASED BRANCHES	0.0001	0.99	Pos	0.000099		99	POSITIVE
	0.0001	0.01	Neg	0.000001		1	
UNINFECTED BRANCHES	0.9999	0.02	Pos	0.019998		19998	POSITIVE
	0.9999	0.98	Neg	0.979902		979902	
			POS	20097	0.0049261084	0.49%	

So what does this mean if we take a population for example of 1 million people.

If we multiply out the probabilities above we see that for the uninfected population i.e. 999,000 if we multiply our 0.99999 probability of not being infected with the 2% i.e. 0.02 probability of it being a false positive result we get 19,998 people out of a 1 million population who will get a false positive result.

In other words 20,097 people out of 1 million get a test result that says they have the disease but only 99 actually have the disease based on the figures above.

Baye's Theorem continued

Millionaires		Inherited their wealth
HS Drop-out	0.05	0.50
HS Diploma	0.15	0.20
Undergraduate Degree	0.60	0.20
Graduate Degree	0.20	0.15

What is the probability that a millionaire did not complete high school given that they earned their wealth?

Let's draw our decision tree again as shown below. To simplify things a little let's say that our total number of millionaires is 1000.



Add text



To calculate the above decision tree mathematically

Bayes Theorem

$$0.031 = \frac{(.05) (.50)}{(.05 \times .50) + (.15 \times .80) + (.60 \times .80) + (.20 \times .85)}$$

Bayes theorem helps us to understand both relationships and probabilities.

Permutations

The number of ways in which objects can be arranged.

Permutations

The number of ways in which objects can be arranged

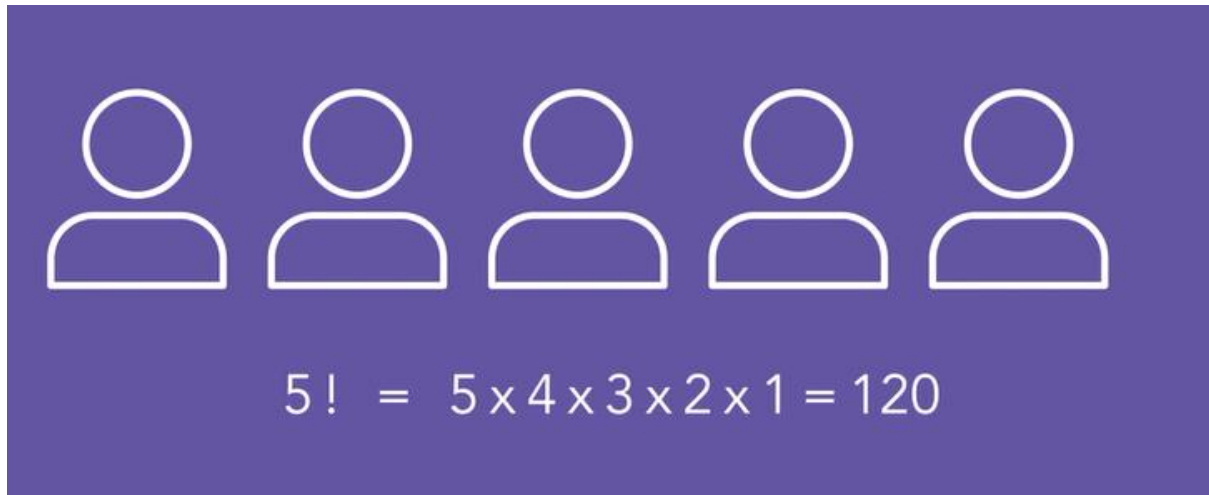
The numbers 123 have six permutations

- 123
- 132
- 213
- 231
- 312
- 321

Permutation calculation (Factorial formula)

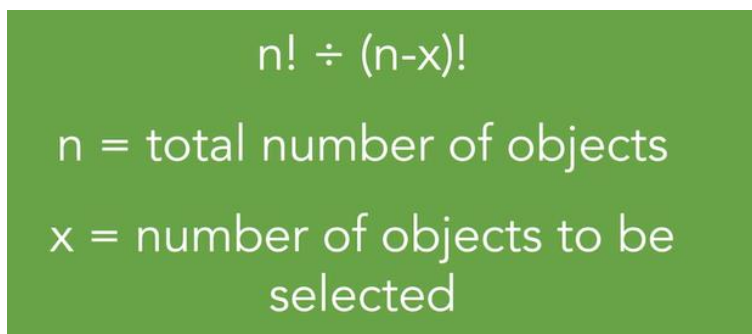
How many different combinations if we have 5 objects?

5! is called 5 factorial and is calculated as described below

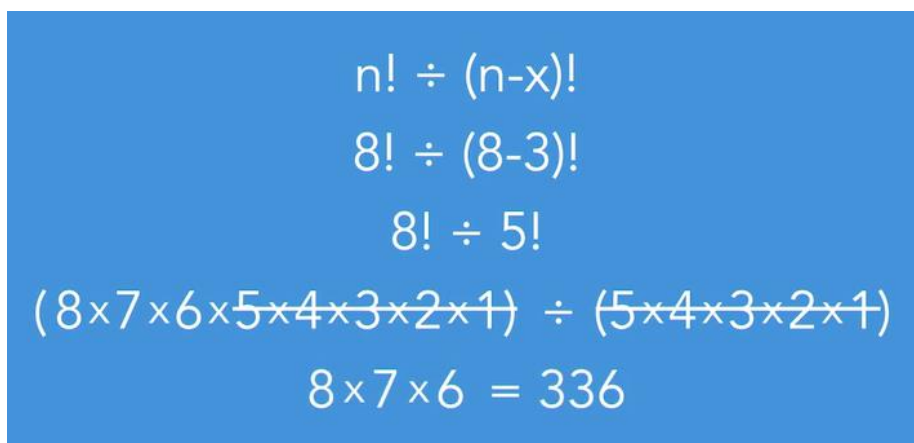


Now let's take a different scenario ;

If we have 8 people and we are awarding 1st, 2nd and 3rd prizes how many combinations of 1st, 2nd, 3rd are there among the 8 people?



This gives us



Combinations vs Permutations

Combinations: When we select a team the order in which the names are picked does not affect the nature of the team.

Combinations: Formula

$$n! \div [(n-x)! x!]$$

n = Total number of objects

x = Number of objects chosen
at one time

In the above example we have e.g. a panel of 10 students and are picking teams of 4 to represent the school in a competition i.e. n = 10 and x = 4.

Combinations: Formula

$$n! \div [(n-x)! x!]$$

$$10! \div [(10-4)! 4!]$$

$$10! / [(6)! 4!]$$

$$\frac{10 \times 9 \times 8 \times 7 \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{(\cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1})(4 \times 3 \times 2 \times 1)}$$
$$(10 \times 9 \times 8 \times 7) / (4 \times 3 \times 2 \times 1) = 210$$

If we cancel out the same calculation values on top and bottom we get a result of 210 possible combinations of teams of 4 from a panel of 10 where the order in which the names are picked doesn't matter.

Now let's look at another scenario. We want to find out what the chances are of two people from the total of 10 being picked on the same team.

Therefore we can look at it as 2 people have already been picked so there are 8 people remaining and what is the probability our two people e.g. John and Anne will be picked together on the team?

So n is then = 8 to represent the 8 remaining people after the first 2 have been picked.

The 2 in the calculation below represents our two people John and Anne.

$$\begin{aligned} & n! \div [(n-x)! x!] \\ & 8! / [(8-2)! 2!] \\ & 8! / [(6)! 2!] \\ & \frac{8 \times 7 \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{(\cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1})(2 \times 1)} \\ & (8 \times 7) / (2 \times 1) = 28 \end{aligned}$$

Therefore the result above means there are 28 possible teams of 4 that could be chosen using the first 2 picked plus 2 from remaining 8 students with both John and Anne being on the team.

$$\begin{aligned} & 210 \text{ Total combinations} \\ & 28 \text{ Desired outcomes} \\ & 28 \div 210 \\ & 13.3\% \end{aligned}$$

Therefore of the 210 possible teams of 4 there is a $28/210 = 13.3\%$ chance that both John and Anne will be picked together on the team.

Discrete Random Variables

Continuous Random Variables

Random Variable

Value of the outcome is unknown

A random variable e.g. the sum of two dice rolled. What are the possible results here?

Discrete Random Variables: Dice Example

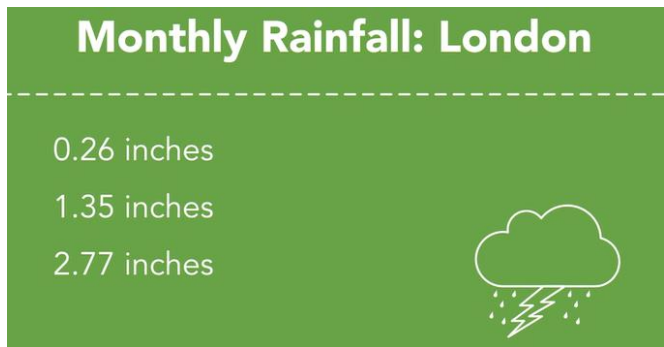
2 3 4 5 6 7 8 9 10 11 12



The possible outcomes are a minimum value of 2 (1+1) to a maximum value of 12.

Note: It is not possible to have values that occur between the above values e.g. 3.5, the values have to be whole numbers or as we would call it **discrete** values.

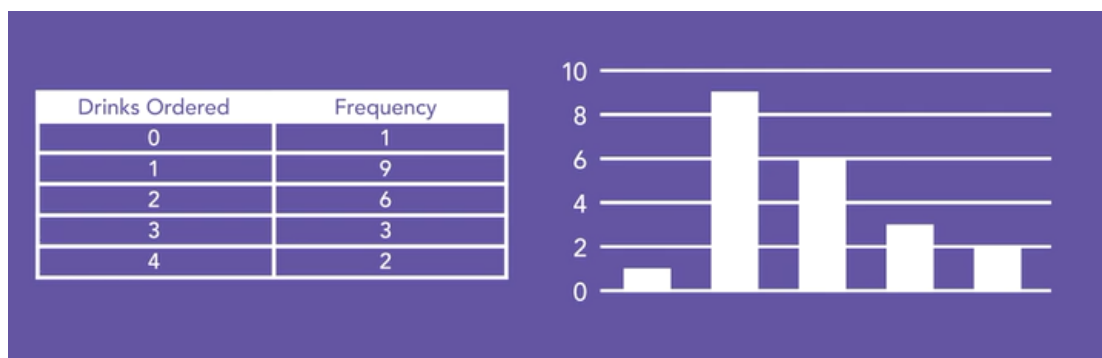
However if we want to measure e.g. the monthly rainfall in London this is a random variable i.e. includes decimal values as well as whole numbers.



Why is this distinction between discrete and continuous important in determining probabilities?

Consider the two examples below

- 1) Coffee shop example: This is the number of drinks ordered by an individual customer on the left and on the right the number of customers who order this amount of drinks in a one hour sample period. This gives the distribution graph on the right showing the spread of values across this time period.



- 2) Rainfall example: Here the same value does not repeat even once across the total number of 20 values. Therefore the distribution table does not present a clear picture like the one above for discrete data.



Since we have an endless number of possibilities for the continuous random variables, we need to consider alternative ways to calculate those types of probabilities.

Mean and standard deviation of discrete probability distributions

Here's the data for the number of drinks ordered per customer at a particular Starbucks between 8am and 9am. This table is called the probability distribution.

Probability Distribution

Drinks Ordered	Frequency
0	1
1	35
2	8
3	3
4	2
5	1
TOTAL	50

We can see that we had 50 different paying customers. Note, we're only counting the number of drinks in their order. 35 customers ordered only one drink. Eight customers placed an order for two drinks in their purchases. We even had one customer that ordered zero drinks during their purchase, perhaps they bought a pastry or some fruit. As you can see, the customers ordered between zero drinks and five drinks during their purchases. Now that we know how many of the 50 customers ordered each of those discrete drink totals, we can find the relative frequencies.

Mean of a Discrete Probability Distribution

Drinks Ordered	Frequency	Relative Frequency
0	1	$(1/50) = 0.02$
1	35	$(35/50) = 0.70$
2	8	$(8/50) = 0.16$
3	3	$(3/50) = 0.06$
4	2	$(2/50) = 0.04$
5	1	$(1/50) = 0.02$
TOTAL	50	

So what was the mean number of drinks ordered per customer during this single hour? For this we will be calculating a weighted mean, in other words, I will multiply zero drinks times the relative frequency 0.02. I will multiply one drink by 0.70. I will multiply two drinks by the relative frequency of 0.16.

I will then add up all of these products and this tells me that my mean for this probability distribution is 1.46 drinks. Our average customer between 8am and 9am ordered 1.46 drinks.

Mean of a Discrete Probability Distribution

Drinks Ordered	Frequency	Relative Frequency	Weight
0	1	0.02	$(0 \times 0.02) = 0.00$
1	35	0.70	$(1 \times 0.70) = 0.70$
2	8	0.16	$(2 \times 0.16) = 0.32$
3	3	0.06	$(3 \times 0.06) = 0.18$
4	2	0.04	$(4 \times 0.04) = 0.16$
5	1	0.02	$(5 \times 0.02) = 0.10$
TOTAL	50		
		MEAN	1.46

Standard Deviation of a Discrete Probability Distribution

Drinks Ordered	Drinks Squared	Frequency	Relative Frequency	Weight
0	0	1	0.02	0.00
1	1	35	0.70	0.70
2	4	8	0.16	0.32
3	9	3	0.06	0.18
4	16	2	0.04	0.16
5	25	1	0.02	0.10
TOTAL		50		
			TOTAL	3.02
	WEIGHT TOTAL 3.02	Mean Squared 2.13	VARIANCE 0.89	Standard Deviation 0.94

$$\begin{aligned}
 &0 \times 0.02 \\
 &1 \times 0.70 \\
 &4 \times 0.16 \\
 &9 \times 0.06 \\
 &16 \times 0.04 \\
 &25 \times 0.02
 \end{aligned}$$

$$3.02 - 2.13 = 0.89$$

Next, let's figure out the standard deviation for this probability distribution. We have the same data set here, but I've sandwiched in a new column between drinks ordered and frequency.

This column is simply the number of drinks ordered squared. Just like with the mean, my next step is to multiply. This time I multiply drinks squared by the relative frequency. So for five drinks, 25 times the relative frequency 0.02. I do this for each discrete drink order zero to five. I then add up the products. Here I get a total of 3.02. Remember this number, 3.02, use drinks squared.

So, to get the variance, we subtract 3.02 minus our means squared. Our mean was 1.46. 1.46 squared is 2.13. So, our variance is 0.89. If we take the square root of our variance, 0.89, we get our standard deviation, which is often referred to as sigma. The standard deviation is therefore 0.94.

Range for 1 Standard Deviation

$$\text{Lower Limit} = 1.46 - 0.94 = 0.52$$

Mean of 1.46

$$\text{Upper Limit} = 1.46 + 0.94 = 2.40$$

Range for 1 Standard Deviation

$$0.52 \leftarrow 1.46 \rightarrow 2.40 \text{ drinks}$$

Drinks Ordered	Frequency	Relative Frequency
0	1	0.02
1	35	0.70
2	8	0.16
3	3	0.06
4	2	0.04
5	1	0.02

86% of customers

The standard deviation for this probability distribution is 0.94 drinks. So, one standard deviation from the mean would be a range from 1.46 drinks minus .94 to 1.46 drinks plus .94, a range of 0.52 drinks to 2.4 drinks.

Since drink orders are discrete, though, this range really just accounts for drink orders of one or two drinks. According to our table, this accounts for 86% of our customers.

Range for 2 Standard Deviations

$$\text{Lower Limit} = 1.46 - 2(0.94) = -0.43$$

Mean of 1.46

$$\text{Upper Limit} = 1.46 + 2(0.94) = 3.35$$

$$-0.43 \leftarrow 1.46 \rightarrow 3.35 \text{ drinks}$$

Drinks Ordered	Frequency	Relative Frequency
0	1	0.02
1	35	0.70
2	8	0.16
3	3	0.06
4	2	0.04
5	1	0.02

94% of customers

Two standard deviations from the mean ranges from negative .43 drinks to 3.35 drinks which accounts for drink orders between zero and three drinks. This accounts for 94% of our customers.

Range for 3 Standard Deviations

$$\text{Lower Limit} = 1.46 - 3(0.94) = -1.37$$

Mean of 1.46

$$\text{Upper Limit} = 1.46 + 3(0.94) = 4.29$$

$$-1.37 \leftarrow 1.46 \rightarrow 4.29 \text{ drinks}$$

Drinks Ordered	Frequency	Relative Frequency
0	1	0.02
1	35	0.70
2	8	0.16
3	3	0.06
4	2	0.04
5	1	0.02

98% of customers

And three standard deviations from the mean ranges from negative 1.37 drinks to 4.29 drinks which accounts for drink orders between zero and four drinks. This would include 98% of our customers. And you might remember that three standard deviations should capture 99.7% of our data points. So, according to our calculations, that single order of five drinks was a true outlier. And just like that, you're becoming a real statistician.

Expected Monetary Value

Expected Monetary Value (EMV)

Total of the weighted payoffs
associated with a decision

In this scenario we are entering our team in a 16 team tournament. The entry fee is \$750 but four rounds will be played in the knockout competition and a team will win \$700 for each round won. However once you lose a round you are out of the tournament. For this exercise the assumption is that all teams are at the same ability level and all have an equal chance of winning.

Wins	Frequency
4	1
3	1
2	2
1	4
0	8
TOTAL	16

Above we see the frequency of wins for the 16 teams that take part i.e. at the top the winner will play 4 games and win each of them.

Next down is the beaten finalist who played 4 games winning 3.

Next down are the two teams who were beaten in the 3rd round i.e. winning 2 games.

Wins	Frequency	Relative Frequency	Winnings	WEIGHT
4	1	0.0625	\$2,800	175
3	1	0.0625	\$2,100	131.25
2	2	0.1250	\$1,400	175
1	4	0.2500	\$700	175
0	8	0.5000	\$0	0
TOTAL	16	1.00		656.25

Next we calculate the relative frequency for the teams from winners downward as shown in the table above. Multiplying the actual \$ winnings at each stage by the weighted value gives us a figure of \$656.25 as we work our way towards the Expected Monetary Value (EMV).

\$656.25	-	\$750	= -\$93.75
Expected winnings		Entry fees	EMV

However each team has to pay the entry fee of \$750 so for the average team the EMV is a loss of \$93.75.

Wins	Probability	Winnings		WEIGHT
4	0.05	2800		140
3	0.15	2100		315
2	0.25	1400		350
1	0.25	700		175
0	0.30	0		0
TOTAL	1.00			980

$$\begin{array}{rclcl}
 \$980 & - & \$750 & = & \$230 \\
 \text{Expected winnings} & & \text{Entry fees} & & \text{EMV}
 \end{array}$$

However what if our team is a bit better than average? In the scenario above an independent consultant has assessed our winnings stats for the last three seasons and estimated that we have a 5% chance of winning the tournament, a 15% chance of getting to the final, a 25% chance of winning either 1 or 2 games and a 30% chance of being knocked out in the first round.

When we multiply out these scenarios above we get an EMV of \$230 which is monetary gain overall from our position after entering the competition.

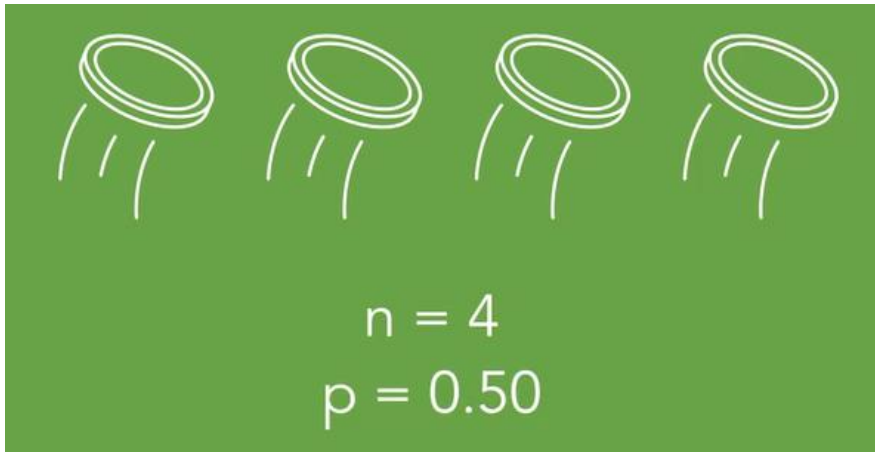
So, as you're trying to figure out what your commissions for next year might be, what the cost of a big corporate project might be, or perhaps as you consider the sales probabilities for next quarter, you can use expected monetary value to help you understand the possibilities, consider your options, or you might even use expected monetary value to persuade others about a money related decision.

Binomial Random Variable

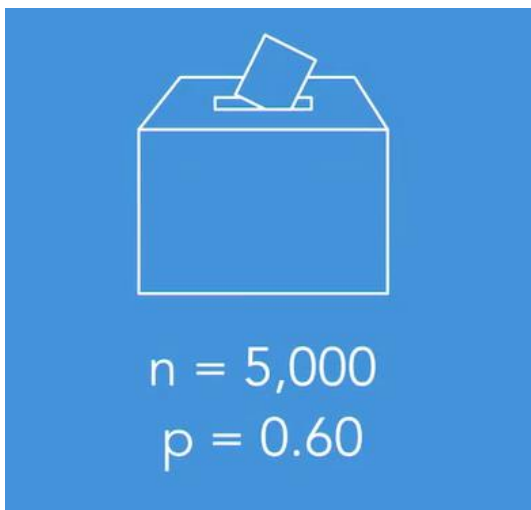
Binomial Random Variable

An experiment that has only two possible outcomes

An example here would be a coin toss i.e. only two outcomes, heads or tails



We also take into account in the coin number of coin flips and $p = 0.5$ the chance that the coin will show the heads or tails face.



Now let's take voter turn out as an example. In the example above there are 5000 people registered to vote with a 60% turn out rate so $n = 5000$ and $p = 60\%$.



Now let's take a company example i.e. the one above where a number of people are introduced to the company each month e.g. through an open company day and 20% of the new people who attend the event will join the company each month.

Let's be clear about the question above. "Only one" means not less than and not greater than one person joins the organisation out of the three people who attend the event.

Tom	Joins	0.2
Lori	Does NOT Join	0.8
Fred	Does NOT Join	0.8
		0.128

In the scenario above Tom joins (and is part of the 20%) and Lori and Fred do not (and are part of the 80%).

Therefore multiplying $0.2 \times 0.8 \times 0.8$ gives us 0.128 or 12.8%. However the same scenario applies if Lori or Fred are the one person who joins the Company.

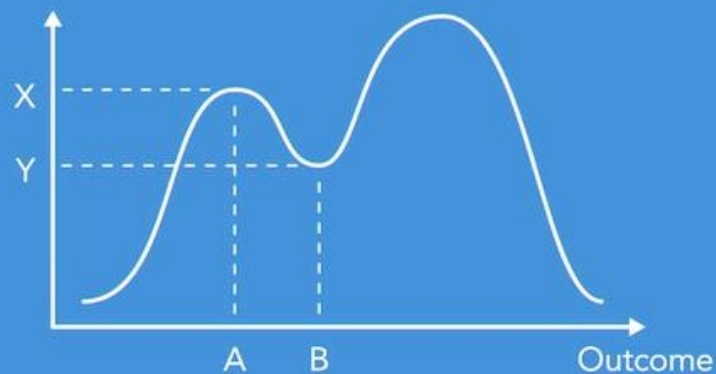
This gives us a 38.4% chance that "only one" of our three people will join the Company by adding the 12.8% together for the three people as shown below.

$$12.8\% + 12.8\% + 12.8\% = 38.4\%$$

Using the same method, we could figure out the probability that exactly zero, one, two, or all three of our new meeting attendees would join our organization as shown below.

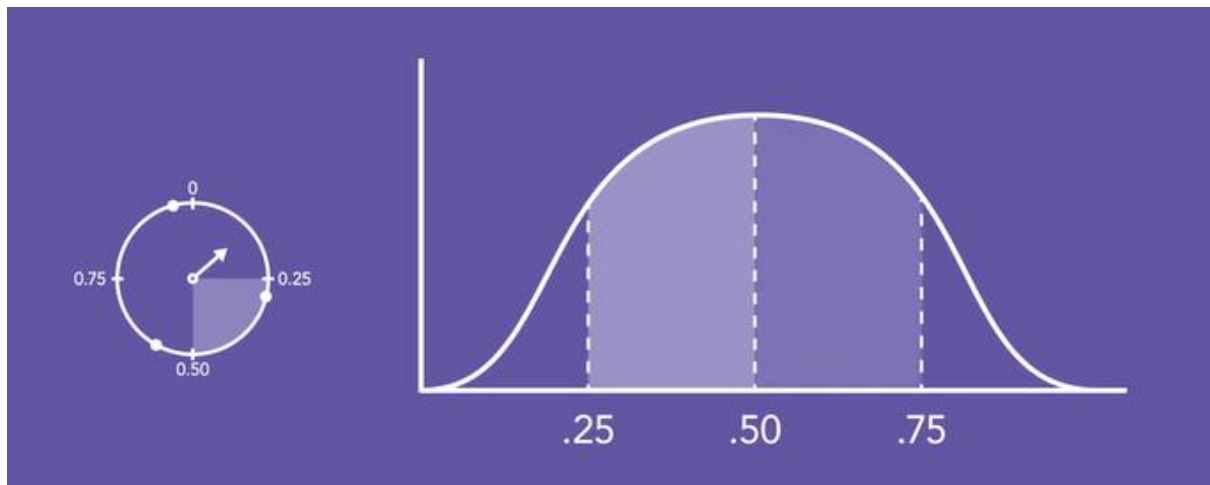
Number that join	
0	0.512
1	0.384
2	0.096
3	0.008

Curves: Probability Densities



The area under the curve represents the probability of each and every outcome. For this probability density, the probability of outcome A is X. The probability of outcome B is only Y.

Since the area under the curve represents all possible outcomes, then the area under the curve is equal to 1.0 or 100%. Because the probability of any exact single scenario like A or B is so low, most of the time, people will calculate the probability of a group of outcomes.

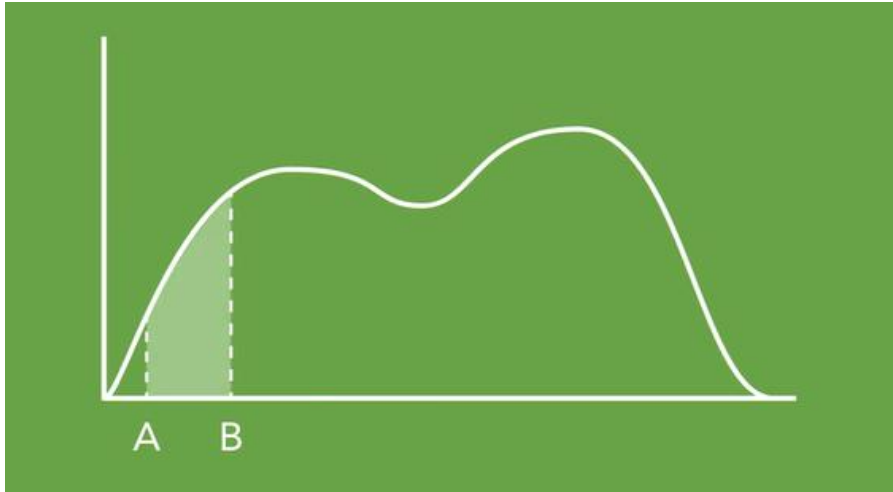


The example above looks at a spinning wheel on the left. We might ask, what is the probability the arrow will land on or between 0.25 and 0.50? If this was the probability density for the spinning wheel, we could see that the area under the curve between points 0.25 and 0.50 would represent the probability that the arrow would land between those two points.

If we wanted to know the probability that the arrow would fall between 0.25 and 0.75, we can now see that the probability is easily over 50% since the shaded area under the curve is a significant

percentage of all the area under the curve. Of course, if we ask, what are the odds that the arrow will fall between zero and 0.9999, it would be very close to 100%.

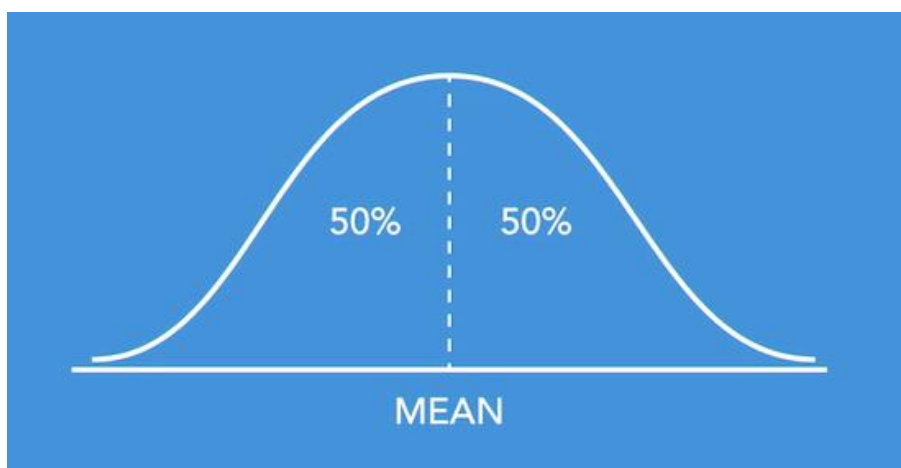
Some simple take aways from the probability density curve



For example, for this probability density, the odds of outcome A are very low, and the probability of outcome B is higher than the probability of outcome A.

The probability of having an outcome between points A and B is certainly less than 25%. A large percentage of the area under the curve is not shaded in.

The famous bell-shaped curve: Introduction

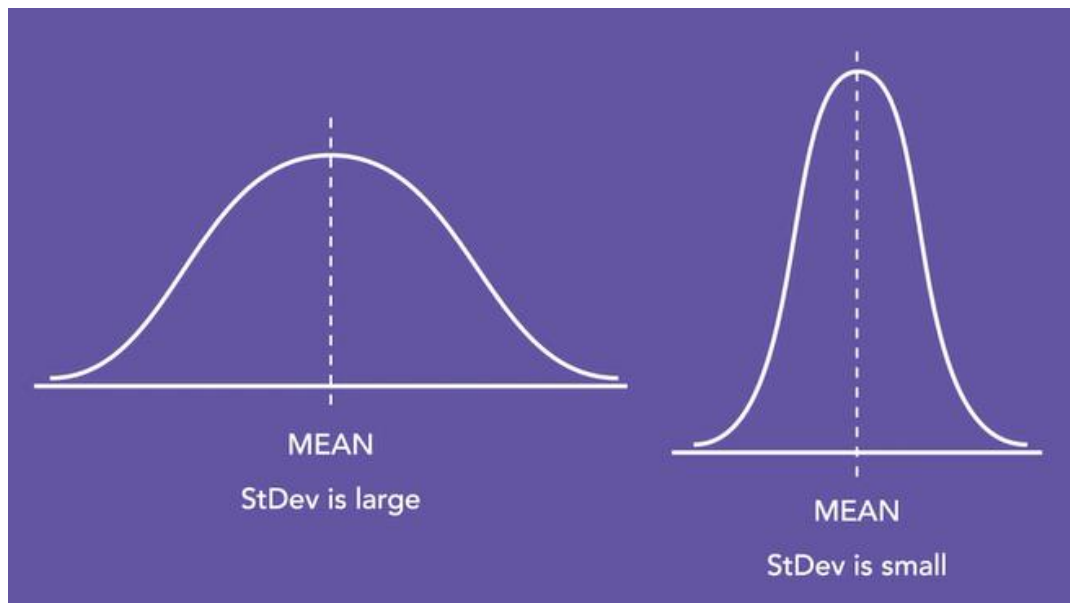


Data will often take on this bell shape for its probability distribution. What are we looking at? Often, the mean of the data is centered at the highest point of the curve. The data around the mean is symmetrical. In other words, 50% of the data is above the mean, 50% is below the mean. The farther we get from the mean, the lower the probability of those outcomes. And notice the curve never touches the axis.

It just keeps on going to infinity in either direction. Also remember that the area under the curve is equal to one, which means that the area under the curve accounts for 100% of all the possible outcomes. This is what we call the classic normal curve.

It is vital to understanding probabilities, especially when you're told that the data is normally distributed. Essentially, they're saying that the data is taking on the shape of a normal curve.

Differently shaped normal distribution curves



Not all normal curves are created equal though. These are both normal curves, but one of them is taller and more narrow. The more narrow curve has a smaller standard deviation. As you can see, the flatter curve is more accommodating. It has more room under the curve at distances farther away from the mean. Still, you might be asking, why is this so important? Very few data sets are symmetrical, so normal distributions must be very rare.

Well, believe it or not, the normal distribution is more widespread than most would ever imagine.

The Fuzzy Central Limit Theorem

The Fuzzy Central Limit Theorem

Data influenced by many small and unrelated random effects are normally distributed.

The Fuzzy Central Limit Theorem essentially says that most things we observe in nature and in our day-to-day life abide by the normal distribution. What it is saying is that when data is influenced by dozens or even hundreds of small and often unrelated random effects, the results end up being normally distributed.

Think about things like global temperatures, the stock market, and the average weight of men in the United States. In each of these cases, the resulting temperatures, stock prices, and weights were influenced by thousands of different things but their probability curves when plotted follow the normal distribution.

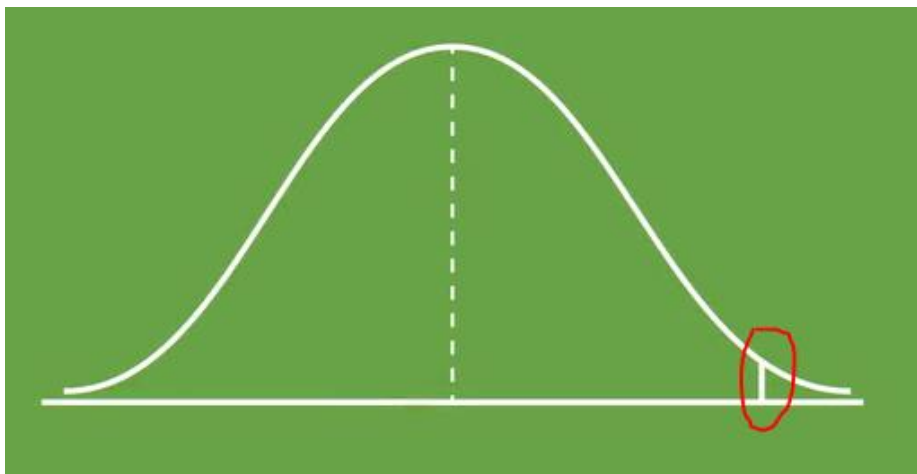
Using the Z transformation to find probabilities

$$Z\text{-score} = (\text{data point} - \text{mean}) / \text{Std Dev}$$

$$Z\text{-score} = (211 - 150) / 25 = 2.44$$

In a previous section, we discussed the Z-score. We were able to figure out how far from the mean, our data point was, in terms of standard deviations. To recap, if I gave you a mean of 150, and a standard deviation of 25. And then I gave you a data point of 211, you could find the Z-score of 2.44.

On our normal curve, that would be over here, far to the right of the mean.



Let's use the Z transformation to answer a question. What is the probability a man weighs between 140 pounds, and 170 pound. Now we need two Z-scores. First, the Z-score for 170 pounds, plugging our numbers into the formula, we get a Z-score of 0.80. The chart tells us the value for this is 0.7881, or 78.81% Now, the Z-score for 140 pounds.

170 pounds

$$Z\text{-score} = (\text{data point} - \text{mean}) / \text{Std Dev} = 2.40$$

$$Z\text{-score} = (170 - 150) / 25 = 0.80$$

$$\text{Chart value} = .7881 = 78.81\%$$

Now let's calculate the z score for 140lb

$$\begin{aligned} &140 \text{ pounds} \\ \text{Z-score} &= (\text{data point} - \text{mean}) / \text{Std Dev} = 2.40 \\ \text{Z-score} &= (140-150) / 25 = -0.40 \end{aligned}$$

Here we get a Z-score of negative 0.40. How do we find that on the chart below?

Standard Normal Distributions Table					
	0.00	0.01	0.02	0.03	0.04
0.0	0.500	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495
1.7	0.9554	0.9564	0.9573	0.9582	0.9591
1.8	0.9641	0.9649	0.9656	0.9664	0.9671
1.9	0.9713	0.9719	0.9726	0.9732	0.9738

Well, for this negative value, you find 0.40. The chart tells us the value for this is 0.6554. Because the Bell Curve is symmetrical, we can subtract this number from 1.0, to find where we are on the left-hand side of the mean.

So, one minus 0.6554, gives us 0.3446, or 34.46%. Remember, we are trying to find the probability a man weighs between 140 pounds and 170 pounds. Well, the probability a man weighs 170 pounds or less, is 78.81%.

The probability a man weighs 140 pounds or less, is 34.46%. So, by subtracting those two percentages, we now know that the probability that a man weighs between 140 pounds and 170 pounds, is 44.35%. Means, standard deviations, normal curves, Z-scores, and Z-score tables. Not only do you know what they mean, you're now able to calculate interesting probabilities all on your own.