

Statistics Cheat Sheet

Statistics basics

Population vs sample

Mean

Median

Range

Percentile

Interquartile range

Box plot

Deviation

Variance

Predicting the population variance from a sample variance

Standard Deviation

Predicting the population standard deviation from a sample standard deviation

Mean to Standard Deviation

Standard Error

Probability

Discrete and Continuous

Degrees of freedom

Central limit theorem

Distribution of the sample means

Distribution

Sample mean

Distribution of sample means

Normal Distribution

Hypothesis Testing

1. Create a null and alternative hypothesis

Null hypothesis

Alternative hypothesis

2. Set a significance level

Significance level

P Values

3. Collect data that is random and independent

4. Calculate the test results

T test

Z score vs T score

T score

[T test](#)
[One tailed and Two tailed t test](#)
[Chi Squared](#)
[5. Interpret the results](#)

It is likely that when you are in a job, and you need to calculate any of these statistics, you will have a program that performs the calculations for you - you can also find them online. To be a good data scientist you need to understand the calculations that these calculators are making.

Twyman's law:

The more unusual or interesting the data, the more probable it is that it was the product of some sort of mistake.

Knowing about statistics helps you dig deeper and find these mistakes

Statistics basics

Population vs sample

μ represents the population mean (average)

\bar{x} represents the sample mean (average)

N represents the population size

n represents the sample size

Population is the entire group that you want to draw conclusions about. If we want to find the average (mean) height of people living in Germany, the population mean would require us to measure every single person living in Germany. As most of the time it is almost impossible to measure an entire group, it is normal to take a sample from the group. If we measured 10,000 people living in Germany and calculated the average (mean) height from this group, then we would have the sample mean.

Population

All the songs of Nirvana

All the job adverts for Data Scientist in Berlin

All the fish in the sea

Sample

Songs on the best of compilation of Nirvana

Top 50 job adverts for Data Scientist in Berlin

Fish seen in a Jacques Cousteau documentary

While looking at statistics, you will see equations with both mu (μ) and x bar (\bar{x}), so please be aware of this difference. Most of the time it won't drastically affect the results if you mix them up as data scientists tend to work with large samples. However, statisticians eat this kind of subtlety for breakfast, so it's important to be familiar with.

Mean

There are different types of mathematical means. Here we are discussing the arithmetic mean, which is **the sum of all of the numbers divided by the number of numbers**

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

If we have 5 people, who's heights are: 142cm, 183cm, 163cm, 161cm, and 172cm

We can calculate the mean by:

$$\frac{142 + 183 + 163 + 161 + 172}{5} = 164.2$$

The average height of our sample of people is 164.2cm

Median

The median is the middle value from a set of values. 50% of the values are greater, and 50% are smaller.

Looking at the heights of the same people as we did in the mean, we need to first arrange them in ascending order, and then find the middle:

142cm, 161cm, **163cm**, 172cm, 183cm

We can see that 163cm is the median value. There are 2 values bigger, and 2 values smaller

How would we find the median if we had 6 people? Again, 1st we arrange the heights in ascending order, then find the middle:

142cm, 161cm, 163cm, **172cm**, 183cm, 189cm

This is a little trickier as there is not one number in the middle. Here the middle is between 163cm and 172cm (there are 3 larger values, and 3 smaller ones). We therefore take the middle point between 163 and 172 as our median. The median is 167.5cm.

Range

The Range is the difference between the lowest and highest values.

If we look at the heights of the five people in the above example, we can see that the shortest person is 142cm, and the tallest 183cm. The range is therefore $183 - 142$, which equals 41.

Percentile

The value below which a percentage of data falls.

Let's look at our original 5 heights to calculate some percentiles: 142cm, 161cm, 163cm, 172cm, 183cm

The 100th percentile is 183, because 100% of the original data is below 183

The 75th percentile is 172, because 75% of the original data is below 172

The 50th percentile is 163, because 50% of the original data is below 163

The 25th percentile is 161, because 25% of the original data is below 161

The 0th percentile is 142, because 0% of the original data is below 142

The 75th percentile is commonly known as the Third quartile (**Q3**)

The 50th percentile is commonly known as the median or the Second quartile (**Q2**)

The 25th percentile is commonly known as the First quartile (**Q1**)

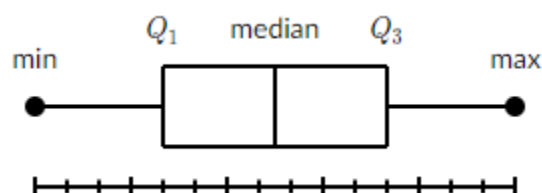
Interquartile range

The range between the 25th percentile and the 75th percentile.

If we look at the heights of the five people in the percentiles example, we can see that the 25th percentile is 161 and the 75th percentile 172. The interquartile range is therefore from 161 to 172. We know that 50% of our data lies in this range.

Box plot

A box plot is a nice way to display a five-number summary of a set of data. The five-number summary is the 0th, 25th, 50th, 75th, and 100th percentiles. 50% of the data lies within the box.



Deviation

Quantifies how spread out our data is.

Deviation is a measure of the difference between a variable's observed value and another value, usually the variable's mean.

Most often we use Mean Absolute Deviation, which tells us how far, on average, our data points are from the mean. The larger the M.A.D., the larger the spread of values

$$\text{Sample mean absolute deviation} = \frac{\sum |x - \bar{x}|}{n}$$

We have a sample of 5 apples that weigh 110g, 130g, 125g, 150g, 125g

First we find out the mean (\bar{x}):

$$\frac{110 + 130 + 125 + 150 + 125}{5} = 128$$

Now we can find out the mean absolute deviation

$$\frac{|110 - 128| + |130 - 128| + |125 - 128| + |150 - 128| + |125 - 128|}{5} = 9.6$$

The mean absolute deviation of our apples is 9.6

Variance

Variance also quantifies how spread out our data is

Variance is the square of the deviations

$$\text{Sample variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n}$$

$$\frac{(110 - 128)^2 + (130 - 128)^2 + (125 - 128)^2 + (150 - 128)^2 + (125 - 128)^2}{5} = 166$$

The variance of our five apples is 166

Predicting the population variance from a sample variance

When we try to estimate variance of a population by simply taking the variance of a sample we tend to underestimate it. This is because the variance measures the spread of the data, and when we take a sample, it is unlikely that this sample will include full spread from minimum to maximum. We compensate for this by minusing one from the denominator to increase the size of the variance ([Bessel's correction](#)).

Unbiased prediction of the population variance from a sample variance:

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

The easiest way to understand this is to consider that we have a population of 1 million people whose ages range from 0-100. If we take a sample of 50 people from this population, it is unlikely that the range of our sample will cover the whole age range, 0-100, of the original population. It is more likely that we will end up with a smaller range, for example from 5 to 88: it's highly unlikely that any sample will include both the highest and lowest value of a population. If we therefore accept the variance of our sample as the variance of the population, we will have underestimated: the sample variance is smaller than the true population variance. We therefore minus 1 from the denominator of our equation to compensate for this.

As we have corrected for our underestimation, this is known as an unbiased estimator. When the denominator is simply left as n, it is known as a biased estimator.

It is not imperative that you totally understand the mathematics of why this happens. But you should be aware that when you see n-1, it represents a correction for the sample not representing the whole population. If you wish to learn more about this, it's pretty interesting, [this video series from Khan academy explains it really well](#).

P.s. This is the same reason why the degrees of freedom minuses 1 in the t-test and chi squared, we are correcting our sample predictions to apply to the whole population

Standard Deviation

Also quantifies how spread out our data is

Standard deviation is the square root of the variance,

Although the variance is useful, it's a huge number because we squared everything. To make it more relatable to our data, we can take the square root of it.

$$\text{Sample standard deviation} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

$$\sqrt{\frac{(110 - 128)^2 + (130 - 128)^2 + (125 - 128)^2 + (150 - 128)^2 + (125 - 128)^2}{5}} \approx 12.88$$

The standard deviation of our sample of 5 apples is approximately 12.88

Predicting the population standard deviation from a sample standard deviation

Biased prediction of the population standard deviation from a sample standard deviation:

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

While the sample variance using Bessel's correction is an unbiased estimator of the population variance, its square root, the sample standard deviation, remains a biased estimate of the population standard deviation. This is because the square root is a concave and nonlinear function. Nevertheless, we still use n-1 for sample standard deviation because it provides us with a better estimation.

Mean to Standard Deviation

Mean

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

Deviation

$$\frac{\sum (x - \bar{x})}{n}$$

Variance

$$\frac{\sum (x - \bar{x})^2}{n}$$

Standard Deviation

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

Standard Error

Quantifies how much the spread of the means of multiple samples

The standard error is the standard deviation of the means of multiple samples

Look at the table below. It represents four boxes of apples on a table at a farmers market. Each box contains five apples. We've weighed all of the apples and calculated the mean for each box.

	Apple 1	Apple 2	Apple 3	Apple 4	Apple 5	Mean
Box 1	110g	160g	130g	145g	110g	131g
Box 2	145g	130g	150g	165g	170g	152g
Box 3	130g	155g	100g	120g	110g	123g
Box 4	170g	140g	165g	160g	125g	152g

Mean of the means:

$$\frac{131 + 152 + 123 + 152}{4} = 139.5$$

Mathematically speaking, each box of apples is a sample as it contains a sample of apples from the total number of apples on a table.

In our example, the standard error is the standard deviation of the means of the boxes of apples.

$$\sqrt{\frac{(131 - 139.5)^2 + (152 - 139.5)^2 + (123 - 139.5)^2 + (152 - 139.5)^2}{4}} \approx 12.82$$

The standard error of our boxes of apples is 12.82

Probability

The proportion of times the outcome would occur if we observed the random process an infinite number of times.

Discrete and Continuous

A **continuous variable** is one whose value is derived from measurement and can take on an **uncountable** range of values. For example, time can be infinitely broken down into years, months, days, hours, seconds, milliseconds, etc....

Discrete variables are countable. For example a survey of favourite Simpsons characters (there's a lot of characters, but still a finite amount that are independent and separate from one another)

Degrees of freedom

Consider a data sample of five positive integers. The average of the total data sample is 6, and four of the numbers in the sample are 3, 4, 5, and 8. This means that the fifth number must be a

ten. It can't possibly be anything else. It does not have the ability to change. As a result, this data sample has four degrees of freedom.

The Degrees of freedom is always the size of our sample minus one: $n - 1$

Central limit theorem

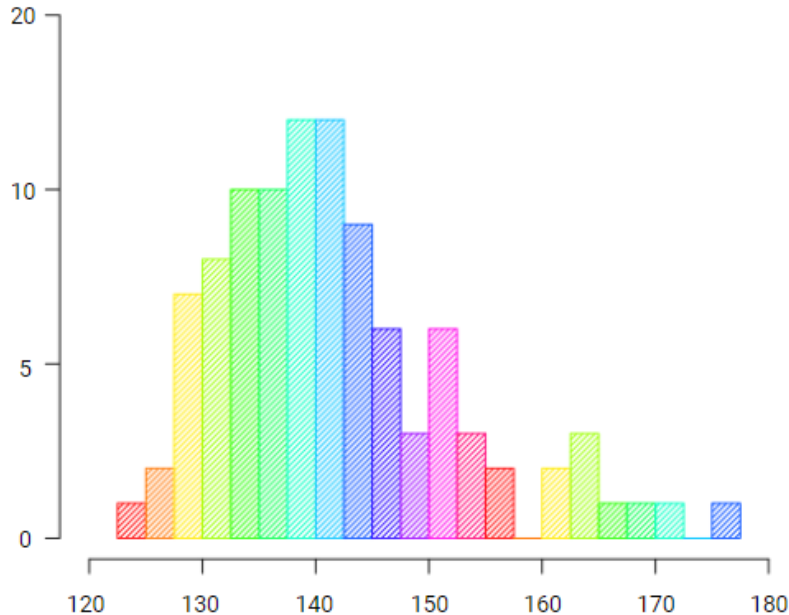
The Central Limit Theorem tells us that, no matter the original distribution of the population, as we take more and more samples, the distribution of sample means will approach a normal distribution.

Distribution of the sample means

Let's start by breaking this convoluted title down into its two need to know components - distribution and sample mean.

Distribution

Let's continue with the theme of apples. Say we took all apples from an orchard and weighed them all. Then we can look at the distribution of weights. We can do this numerically, but it's much easier to comprehend visually. Below is how a distribution looks.



As you can see from the distribution, most of the apples are around 140g and they range from about 120g to 180g. The distribution helps us see the probability of getting an apple with a particular weight. From the distribution we can see that it's highly probable a random apple will weigh between 130g and 150g, and highly improbable that it will weigh more than 180g.

Sample mean

If we take a sample from the above distribution, say 5 apples, and calculate the mean of the 5 apples, then we have the sample mean. It's literally that, the mean of a sample.

Distribution of sample means

The distribution of sample means is therefore a distribution consisting of the means of multiple samples. We take the mean of the 1st sample of five apples and plot it. Then we take the mean of a 2nd sample of apples and plot it on the same distribution as the 1st. We then take a 3rd sample, find the mean, plot it, and so on and so forth. This way we end up with a distribution of multiple sample means, which is also known as the distribution of sample means.

This is significant because the bigger the sample size and the higher the number of samples, the closer this distribution resembles the normal distribution. No matter how the distribution of our original data looks, when we plot a distribution of the sample means, and we do this with enough samples of large enough size, we will always end up with a chart that's a very close approximation of the normal distribution.

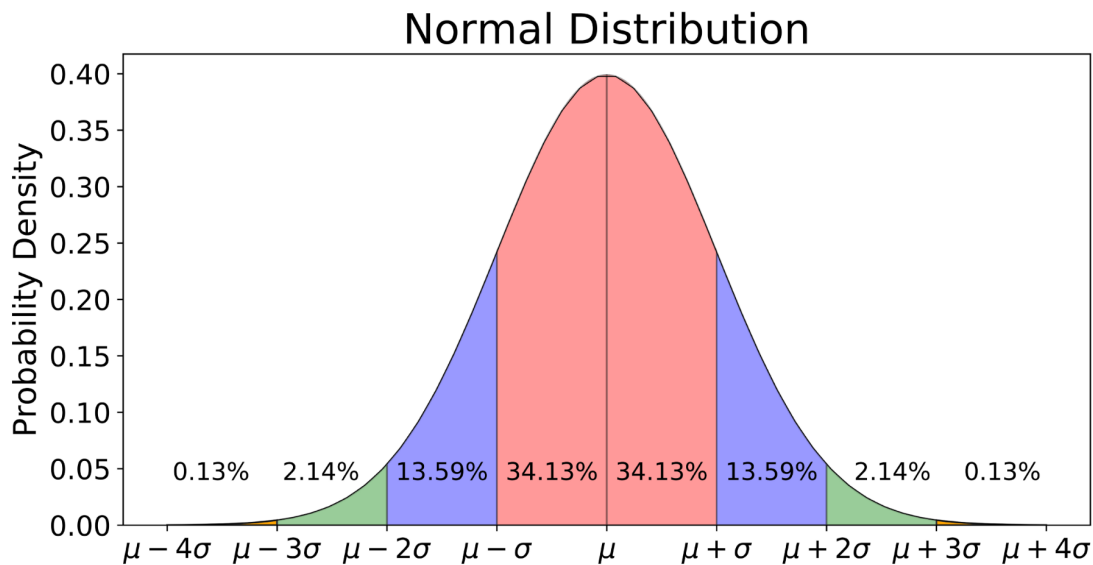
Normal Distribution

Normal distribution is a mathematical ideal that helps calculate probability. The normal distribution is normally explained with a lot of technical terms. All you really need to know is that it's a symmetrical bell shaped curve, where the peak of the curve is the mean, and as you can see from the graph below:

64.26% of the data falls within one standard deviation (σ) of the mean (μ) - the red area.

95.44% of the data falls within 2 standard deviations of the mean - the red plus purple

99.72% of the data falls within 3 standard deviations of the mean - red, purple, green



We use these measurements in standard deviations from the mean to help predict probability. We define any value in a normal distribution by “how many standard deviations away from the mean” it falls. For example, in a normal distribution with a $\mu=0$ and $\sigma=2$, the value 6 is “three standard deviations away from the mean”. The proper term to express this is the z-score. So, in our example, the z-score of 6 is 3.

Hypothesis Testing

Hypothesis testing tries to determine whether what we are testing are the same or different.

Is drug A the same as, or different from, drug B? Does our new anti dandruff shampoo affect dandruff the same as, or different from, our regular shampoo? Does the new red button on our home page have the same or different effect on sales as the original white button design?

To perform a hypothesis test you need to:

1. Create a null and alternative hypothesis
2. Set a Significance Level
3. Collect data that is random and independent
4. Calculate the test results
5. Interpret the results

1. Create a null and alternative hypothesis

Null hypothesis

Often represented as H_0

The most common null hypothesis states that there is no difference between two groups or samples.

For example, a researcher performs an experiment for a new hair growth drug. The control group received a placebo and the test group received the new drug. The null hypothesis in this case would be: the new drug has no effect on hair growth.

We then look at the results of the experiment. If it's possible that the difference in outcomes between the two groups could have happened by chance, we accept the null hypothesis. However, if it is likely that the difference in results did not happen by chance, then we can reject the null hypothesis.

Please keep in mind that the null hypothesis can be accepted or rejected, but it can't be proven. There is always a degree of certainty about the outcome.

Alternative hypothesis

Often represented as H_1

The alternative hypothesis is the opposite of the null hypothesis. It is what's accepted if we reject the null hypothesis. In the case stated above about a new hair growth drug, the alternative hypothesis would state that: the new drug does affect hair growth.

2. Set a significance level

Significance level

Often represented as alpha (α).

If the null hypothesis is rejected, there is strong evidence that the test groups are truly different. This strong evidence must pass the significance level. There is a trend of 95% being the significance level needed to reject the null hypothesis. This means there's a 95% probability that the null hypothesis is wrong. This level can change, it depends on what you're testing and the severity of consequences if you get it wrong. Even though 95% is a standard, always consider the effects of the outcome of your experiment and adjust the significance level accordingly.

P Values

P values are calculated from the results of our experiment. When the p value is lower than the equivalent significance level, we reject the null hypothesis. When the p value is greater than the significance level, we accept the null hypothesis.

If your significance level is 95%, then you need a p value of 0.05 or lower, to be able to consider rejecting your null hypothesis.

If your significance level is 99%, then you need a p value of 0.01 or lower, to be able to consider rejecting your null hypothesis.

P values are related to probability, but are not the same as probability.

The probability of something happening is exactly that, the probability of this one thing occurring.

A p value is the probability of this one thing occurring plus the probability of anything as equally or more improbable.

3. Collect data that is random and independent

This is not something we can take care of on the course (as we are not collecting data). However, please be aware of biases that creep into data. Probability relies on randomness, if you are influencing or guiding your data, then you are only fooling yourself as you will get a false result, and your test will be neither repeatable nor reliable.

A double blind test is a good example of one measure that can be taken to ensure your data is random and independent. It is quite normal to hear of studies where participants do not know if they are in a control group, who receive a placebo, or if they are in the experiment group that's receiving a new drug; this is a blind test, as the participants are blind to which group they are in. In a double blind experiment, neither the participants nor the researcher know who is in which group. A double blind study is conducted so that the researcher can't treat the two groups differently based on their knowledge of the experiment's structure. It's an extra measure to ensure the test is completely fair and random.

4. Calculate the test results

There are various tests that can be used for hypothesis testing. On the course we will focus on the t test and chi squared test: they are most often used. Once you understand them you should have no problems understanding the binomial or anova tests, should you wish to progress with a little bit of personal study.

Which test to use?		What type of data do you have?	
		Numerical	Categorical
What type of comparison are you making?	Sample vs number	1 Sample T-Test	Binomial Test
	2 Samples	2 Sample T-Test	Chi square
	More than 2 samples	ANOVA	

T test

Z score vs T score

Both are a statistical way of testing a hypothesis

As a rule of thumb:

We use a z score when:

We know the standard deviation and the sample size is above 30

We use a t score when:

We don't know the standard deviation, or the sample size is below 30

However, in real life it is more often than not that we don't know the standard deviation. Also, when the sample size is large enough, z score and t score become roughly the same. We will therefore only explain the t score here, as that's what you'll be using the vast majority of the time. If you wish to learn more about the z score, please use some of the further reading suggestions at the end of this cheat sheet.

T score

T score can be converted to p values through an online calculator. Or we can use tables like the one below.

The higher the t score the more likely we are to reject the null hypothesis

The lower the p value the more likely we are to reject the null hypothesis

.... going in opposite directions is confusing, but that's the way it is

Looking at the table below, you can see that across the top are the confidence levels for both one-sided and two-sided t tests. Down the left hand side we have degrees of freedom, which is the size of your sample minus one.

For example, if we performed a two tailed t test on a sample of ten apples with a 95% significance level, we can trace the rows and columns for 9 degrees of freedom and two sided 95%. This gives us a t score of 2.262. We therefore know that if our t score is higher than 2.262 it is likely that we will reject our null hypothesis.

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437

T test

Don't worry too much if you don't follow the formula below, you will have a calculator that performs the maths when passed your sample data. It's included here, so you can get a better understanding of what's happening if you'd like to..

When comparing two samples, the t value is calculated by dividing the difference between the means of the two samples by the square root of the sum of the variance divided by the number of samples. This can be written as:

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

To compare two samples let's enter some data into the formula for our t score, and compare that against the t table. This will allow us to decide whether we should accept our null hypothesis or reject it.

For our example we have two samples of apples.

Null hypothesis: They are the same variety of apple (there is no difference)

Alternative hypothesis: The samples are different varieties of apple

We will use a significance level of 95%

Sample 1: 114g, 165g, 144g, 134g, 184g, 143g, 111g, 155g, 145g, 144g

Sample 2: 164g, 187g, 176g, 155g, 145g, 177g, 136g, 183g, 155g, 165g

From looking at our samples, it appears that Sample 2 is heavier on average. We will now run the t test to check how likely this is to happen by chance.

From the equation above, we can see that we need to know three things about each sample: mean, variance, sample size

Sample 1:

Mean: 143.9

Variance: 472.54444

Sample size: 10

Sample 2:

Mean: 164.3

Variance: 278.9

Sample size: 10

$$\frac{|143.9 - 164.3|}{\sqrt{\frac{472.54444}{10} + \frac{278.9}{10}}} = \frac{20.4}{8.6686} = 2.3533$$

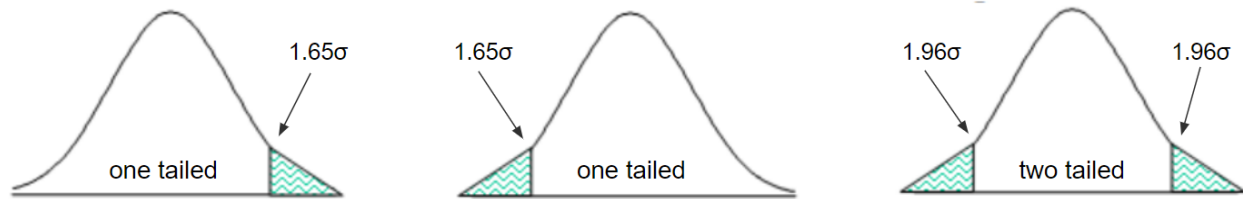
Our t value is 2.3533

We can see from the t table that a two sided t test with a 95% significance level and nine degrees of freedom gives us a t score of 2.262

As our t value is higher than the t score from the table we are able to say that there is something statistically significant between these two datasets. We therefore reject our null hypothesis and accept our alternative hypothesis.

If we use an online t test to p value calculator, we can see that our p value is 0.03. Implying that there is only a 3% probability that the difference between these two samples is due to chance.

One tailed and Two tailed t test



1.65σ and 1.96σ represent the number of standard deviations from the mean a sample needs to be, to be considered statistically significant with a 95% significance level.

Most often you will perform a two tailed t test. We use a two-sided test to examine whether an effect is either better or worse

We use a one-sided test when we only wish to perceive an effect in one direction, either better or worse.

You may be tempted to perform a one-tailed test anytime you have a hypothesis about the direction of an impact because it provides more power to detect an effect. Consider the repercussions of missing an effect in the opposite direction before going ahead. Assume you've created a new medicine that you believe is superior to one that already exists. You choose a one-tailed test to increase your ability to identify the improvement. You are negating the potential that the new drug is less effective than the existing drug by doing so and should use a two-tailed test instead.

Chi Squared

The chi squared test is used when the data is categorical and we are testing two or more samples.

Look at the example below. We can see how many people had jury duty and their race. By comparing how many people were in a jury from each racial group with the expected total based on the makeup of the population, we can determine whether the jury selection is biased, or whether this constellation could have happened by chance.

	White	Black	Hispanic	Other	Total
Representation in juries	180	22	24	14	240
Percentage of population	0.74	0.06	0.08	0.12	1
Expected number on juries	177.6	14.4	19.2	28.8	240

To calculate this we use the Chi Squared test

$$\sum \frac{(x_1 - x_2)^2}{x_2}$$

$$\frac{(180 - 177.6)^2}{177.6} + \frac{(22 - 14.4)^2}{14.4} + \frac{(24 - 19.2)^2}{19.2} + \frac{(14 - 28.8)^2}{28.8} = 12.85$$

We can then use an online calculator to work out our p value. Or we can look at the Chi Squared table to interpret the result

Degrees of freedom (df)	χ^2 value ^[9]										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P value (Probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
	Nonsignificant								Significant		

The degrees of freedom is the number of categories minus one, which in our case means 3. Let's stick with the most often used significance level of 95%. We therefore follow the row for 3 degrees of freedom to the p value column 0.05 and we get 7.82.

The higher the chi squared score the more likely we are to reject the null hypothesis

The lower the p value the more likely we are to reject the null hypothesis

.... going in opposite directions is confusing, but that's the way it is

The results of our chi squared test was 12.85. This is higher than 7.82, which means we should reject the null hypothesis.

If you look closely at the table you can see that our chi squared score is also higher than the chi statistic for 3 degrees of freedom and a significance level of 99%. This implies that there is a less than one percent chance that our spread of races amongst jurors happened by chance.

Using an online chi squared calculator we can see that our p value for this data is 0.005, implying that there is a 0.05% probability that our sample of jurors was randomly sampled from the population

5. Interpret the results

If you read through the sections on t test and chi squared test you will see a lot of language such as implying and suggesting and probability. None of these words are certainties, they all imply an element of chance and arbitrariness. This is because we can't be 100% sure that our results mean what we believe they do. Think about the standard 95% confidence level, which means that if we performed the test 100 times on samples that aren't statistically significant, we should get 95 results that suggest the samples are statistically insignificant and 5 results that suggest there is a statistical significance. What if the time we performed this test it happened to be one of those 5 times where we prove statistical significance? This is why tests should be repeatable, so that we can verify the results. Not only that, if your experiment gives you a p value of exactly 0.05 you may wish to think about whether you accept or reject the null hypothesis, as this value is right on the border. By interpreting the results, we mean that you should think about your test and the concept of probability and make an informed decision as to whether your results truly imply what you believe they do.