DATA SCIENCE INFINITY
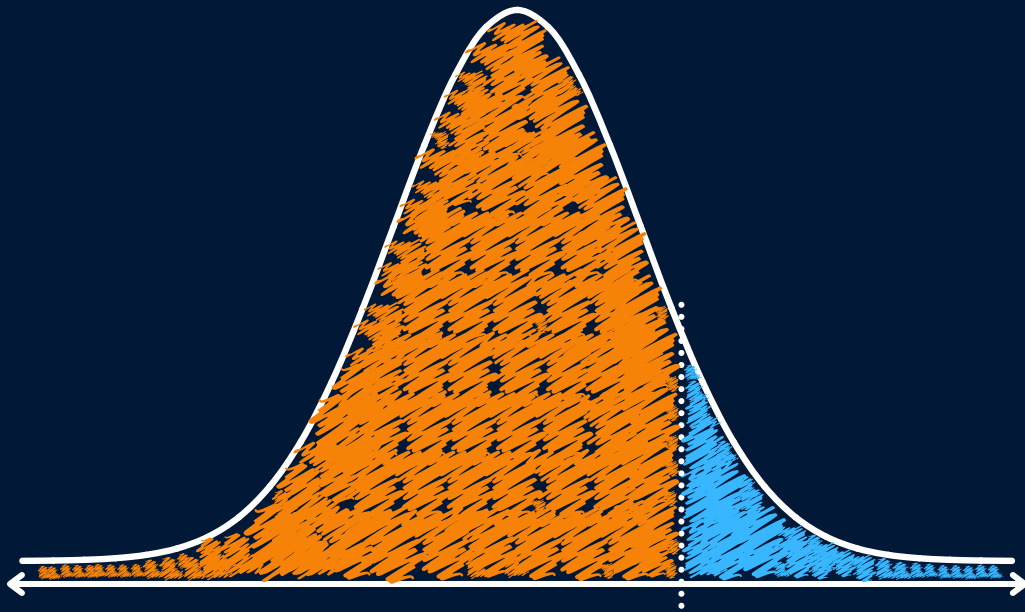
COMMON

HYPOTHESIS

TESTS

AN OVERVIEW

What?

In Statistics, a **Hypothesis Test** is used to assess & understand the plausibility, or likelihood of some assumed viewpoint (*a hypothesis*) - based upon data. In other words, we are using Statistics to test or investigate ideas.

Let's say we're the coach of an NBA Basketball team - we might find ourselves with the below dilemma...

The New Sensation >>

Games Played: 2
Shooting Rate: 60%

The Current Star >>

Games Played: 102
Shooting Rate: 49%

Coach

My new player seems to be a better shooter...but she's only played 2 games. I need more confidence before making a big change!
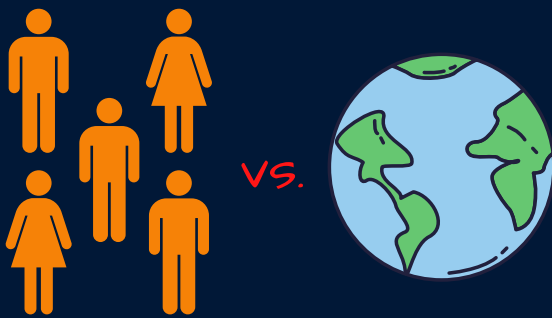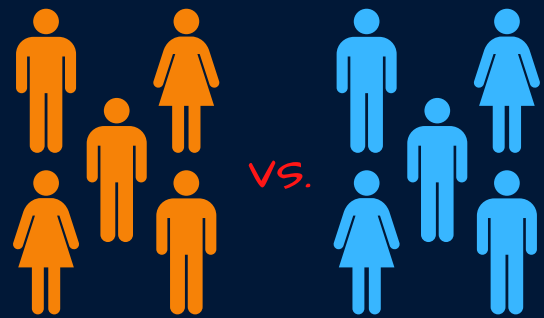
Let's test it!

Common Tests...

There are many, many different types of hypothesis test - each of which are appropriate for different types of data, and/or dealing with different scenarios & comparisons.

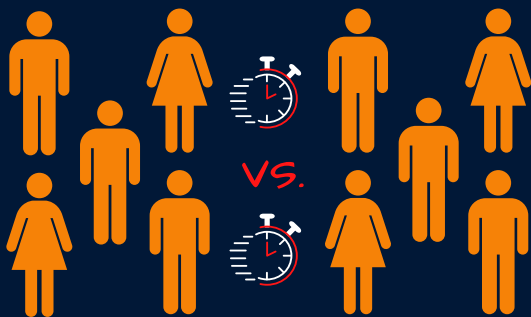Here we will cover several commonly used tests...

One Sample T-Test

VS.

Independent Samples T-Test

VS.

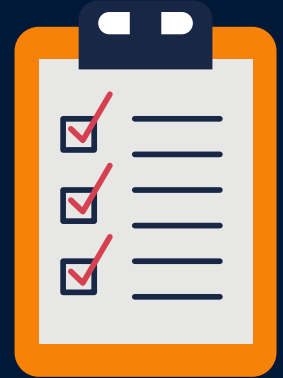### Paired Samples T-Test

VS.

Chi-Square Test Of Independence

% VS. %

The Setup...

It is important to lay down our hypotheses/assumptions and a required level of confidence before running the test itself.

Generally, we specify 3 things; the **Null Hypothesis**, the **Alternate Hypothesis**, and our **Acceptance Criteria**...

The **Null Hypothesis** is where we state our initial viewpoint. In statistics, and specifically hypothesis testing, our initial viewpoint is always that the **result is purely by chance** or that there is **no relationship or association between two outcomes or groups**

The **Alternate Hypothesis** is is essentially the opposite viewpoint to the Null Hypothesis - that **the result is not by chance**, or that **there is a relationship between two outcomes or groups**

The **Acceptance Criteria** is the threshold we set that will determine if there is **enough evidence** to support the Null Hypothesis. This is often set to a p-value of 0.05 but it **does not have to be**. If we want more certainty, we can set this to a lower value

One Sample T-Test

A **One Sample T-Test** looks to assess differences between a sample, and the entire population from which that sample resides.

VS.

If we were the coach of an NBA Basketball team, this might help us with the following question…

Is the average height (cm) of my team higher than that of the entire NBA?

Independent Samples T-Test

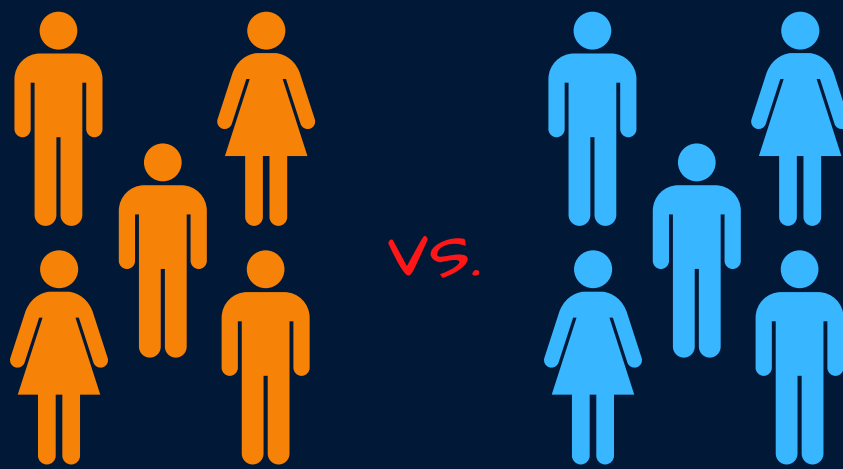A **Independent Samples T-Test** looks to assess differences between a sample, and another sample.

If we were the coach of an NBA Basketball team, this might help us with the following question…

Is the average height (cm) of my team higher than that of our rival team?

Paired Samples T-Test

A **Paired Samples T-Test** looks to assess differences between a sample, and that same sample, at another point in time.
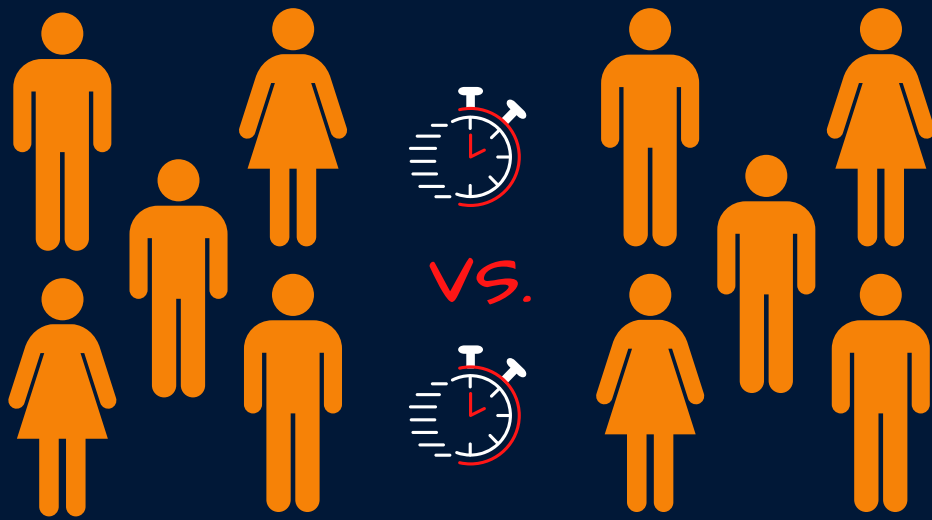
VS.

If we were the coach of an NBA Basketball team, this might help us with the following question...

Has the average jumping height (cm) for my team increased after our 4 week fitness programme?

Chi-Square Test Of Independence

The Chi-Square Test For Independence is used to determine if there is a signficant relationship between two **categorical** variables. It examines the **actual and expected frequencies** to determine if there is a **dependence** between the two variables.

% VS. %

If we were the coach of an NBA Basketball team, this might help us with the following question...

> Is the 3-point shooting % of my newly signed player, higher than that of my current star player?

Let's talk about p-values

The outcome of a hypothesis test (in other words whether or not we support our initial assumption) is influenced by our **Acceptance Criteria**.

This Acceptance Criteria is often based upon a **p-value...**

...So let's talk about what a p-value **is**, and what a p-value **is not!**

*A **p-value** essentially helps us assess whether the results of some finding or test we have conducted are either, **likely to be ordinary, or likely to be strange.***

*In other words - these results we've got - do we think we'd get similar results if we ran **many more tests**, or do we think our results might be something of an rarity?*

*With an Acceptance Criteria of 0.05, we are essentially saying that we want a likelihood of 5% (or lower) that our result happened **by chance**.*

***A p-value is not a probability of an event occuring**, it is a probability or likelihood of seeing a different result if we were to sample many times*

***A p-value does not tell us how different two samples are**. Two samples with the same difference in means, but larger/smaller samples sizes will get different p-values. A p-value is instead telling us how likely it is that they are different (or in other words how confident we can be that they are different)*

Do you want to **learn more** about this topic - and how to **apply it** in the real world?

Do you want to land an incredible role in the exciting, future-proof, and lucrative field of **Data Science**?

DATA SCIENCE INFINITY

LEARN THE RIGHT SKILLS

A curriculum based on input from **hundreds** of leaders, hiring managers, and recruiters

DATA SCIENCE INFINITY

BUILD YOUR PORTFOLIO

✅

Create your professionally made portfolio site that includes **10 pre-built** projects

⌄

DATA SCIENCE INFINITY

EARN THE CERTIFICATION

Prove your skills with the DSI **Data Science Professional Certification**

DATA SCIENCE INFINITY

LAND AN AMAZING ROLE

✓

Get guidance & support based upon **hundreds of interviews** at top tech companies

∨

DATA SCIENCE INFINITY

Taught by former Amazon & Sony PlayStation Data Scientist **Andrew Jones**

What do **DSI** students say?

"I had over 40 interviews without an offer. After **DSI** I quickly got **7 offers** including one at **KPMG** and my amazing new role at **Deloitte**!"

- Ritesh

"The **best program** I've been a part of, hands down"

- Christian

"DSI is **incredible** - everything is taught in such a clear and simple way, even the more complex concepts!"

- Arianna

"**I got it!** Thank you so much for all your advice & help with preparation - it truly gave me the **confidence to go in and land the job!**"

- Marta

"I've taken a number of Data Science courses, **and without doubt, DSI is the best**"

- William

"One of the **best purchases towards learning** I have ever made"

- Scott

"I learned more than on **any other course**, or reading entire books!"

- Erick

"I started a bootcamp last summer through a well respected University, **but I didn't learn half as much from them!**"

- GA

"100% worth it, it is amazing. I have never seen such a good course and I have done plenty of them!"

- Khatuna

"This is a world-class Data Science experience. I would recommend this course to every aspiring or professional Data Scientist"

- David

"Andrew's guidance with my Resume & throughout the interview process helped me land my amazing new role (and at a much higher salary than I expected!)"

- Barun

"DSI is a fantastic community & Andrew is one of the best instructors!"

- Keith

"I'm now at University, and my Data Science related subjects are a **piece of cake** after completing this course!

I'm so glad I enrolled!"

- Jose

"In addition to the great content, Andrew's dedication to the growing DSI community is **amazing**"

- Sophie

"The course has such high quality content – you get your ROI even from the first module"

- Donabel

"The Statistics 101 section was awesome! I have now started to get confidence in Statistics!"

- Shrikant

"I can't emphasise how good this programme is...**well worth the investment!**"

- Dejan

Come and join the hundreds & hundreds of other students getting the **results** they want!