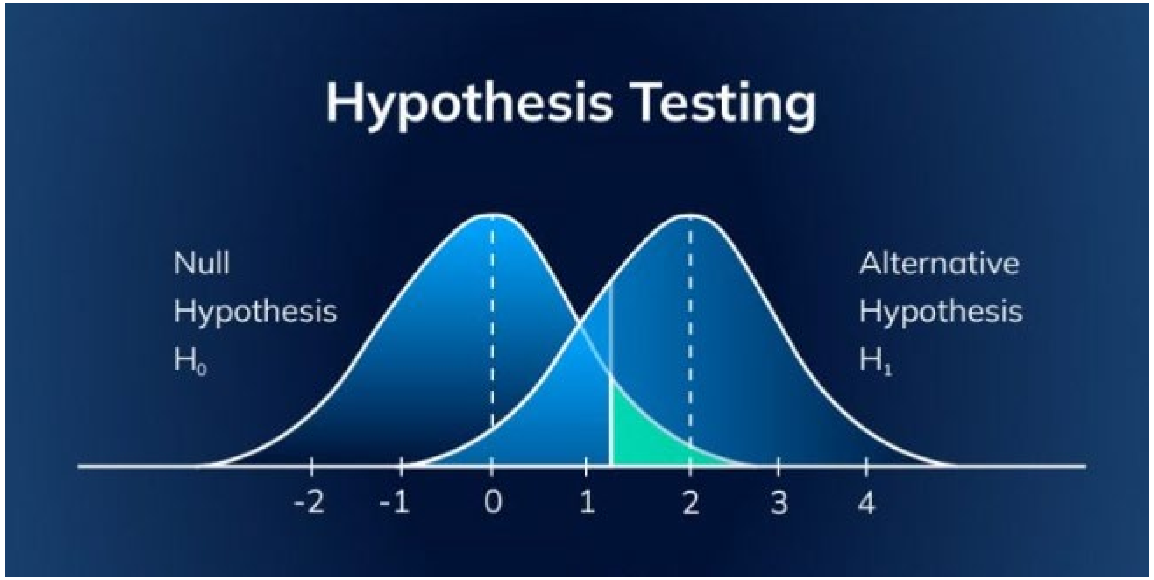
Hypothesis Testing

YORK COLLEGE & UNIVERSITY CENTRE

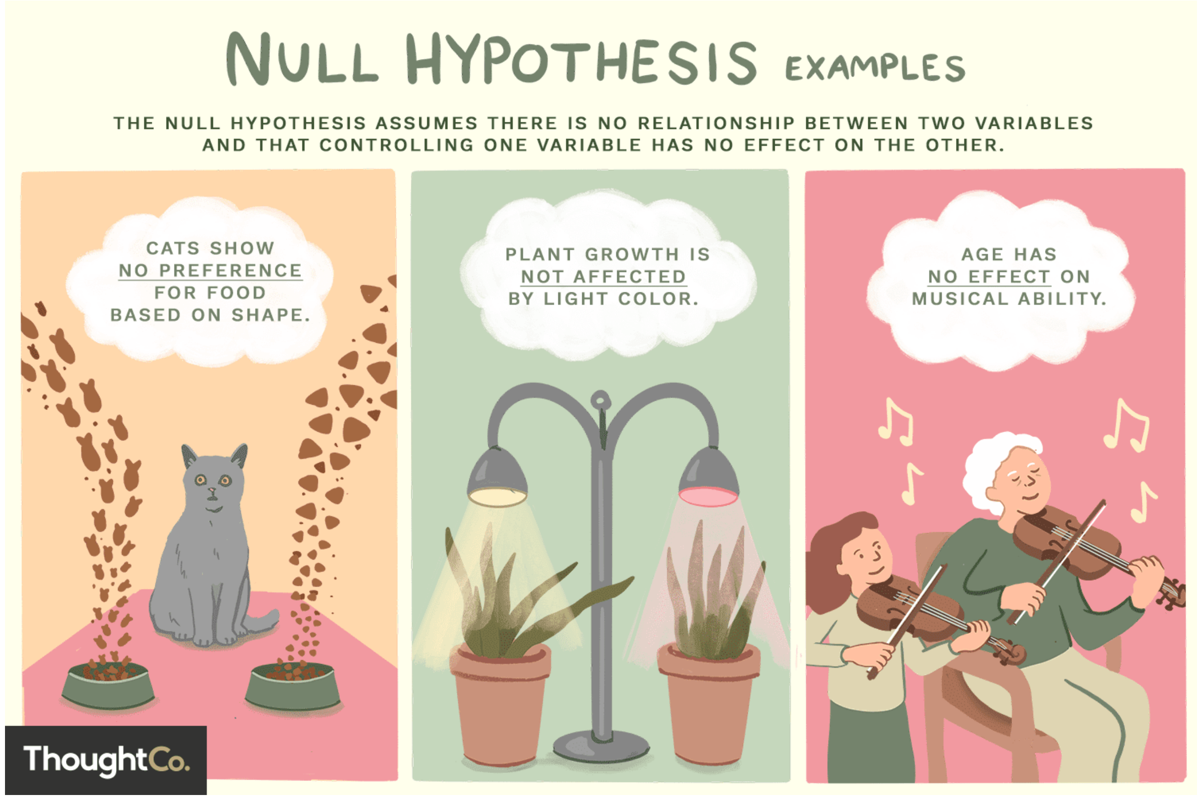
What is hypothesis testing

- Hypothesis testing is the scientific method of checking if a theory is true in a scenario
- It allows us to make assumptions about a set of data and my predications based on it



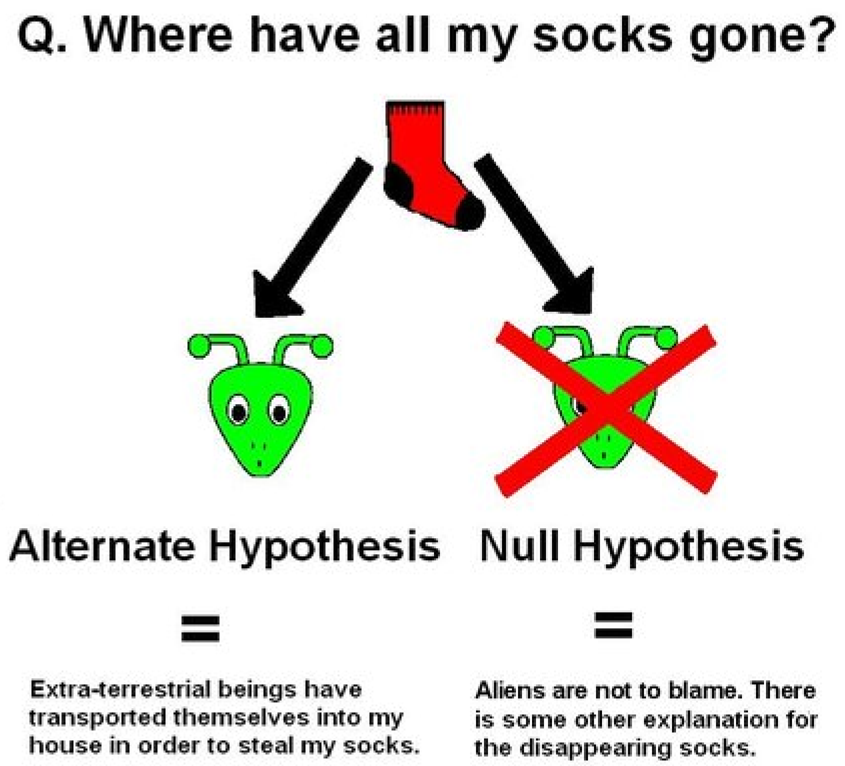
Null Hypothesis

- Symbol $\boldsymbol{H_0}$
- The default assumption that there is no effect or difference
- Example: "There is no difference in test scores between two teaching methods."



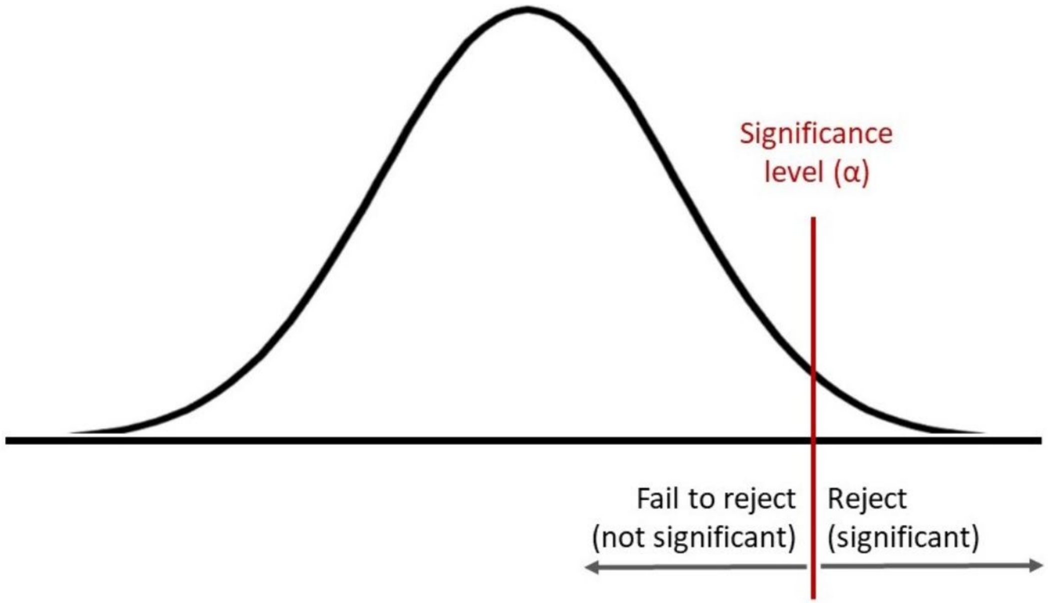
Alternative Hypothesis

- Symbol H_a or H_1
- The assumption we are testing for, ie the effect or difference
- Example: "There is a difference in test scores between two teaching methods."



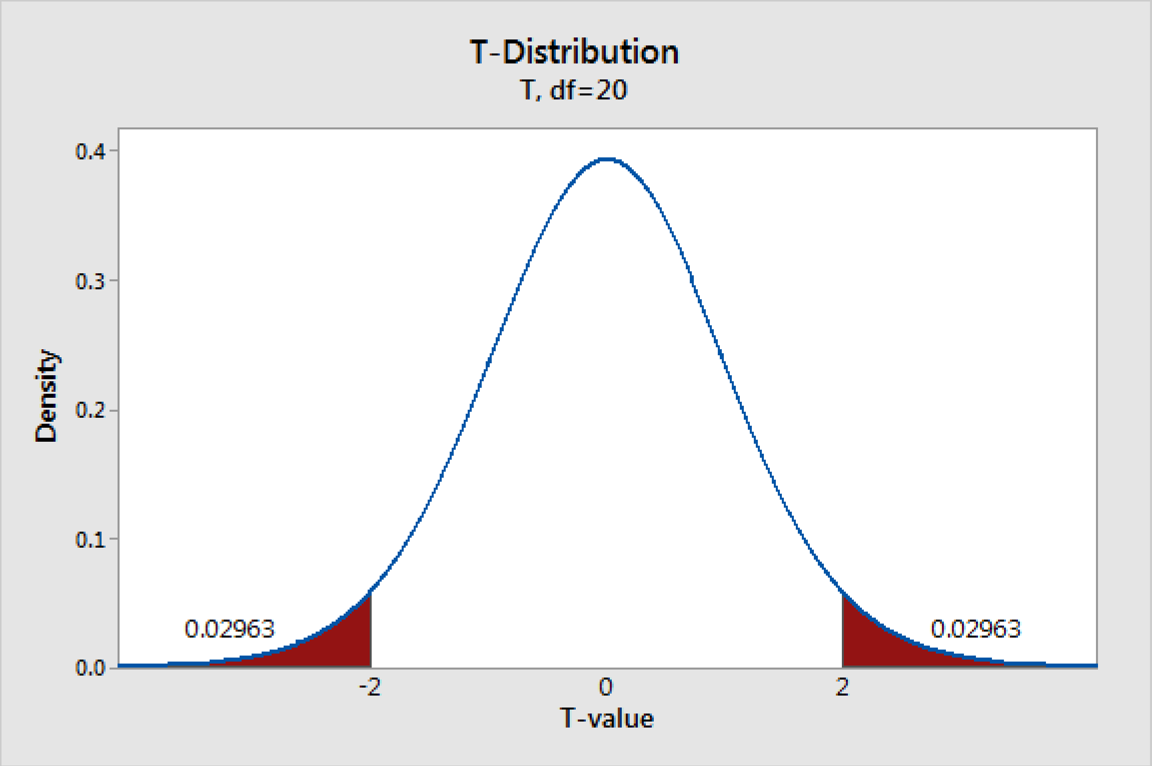
Significance level

- Symbol $\boldsymbol{\alpha}$
- The probability of rejecting H_0 when it is true
- Typically, 0.05 (5%) or 0.01 (1%)



Test Statistic

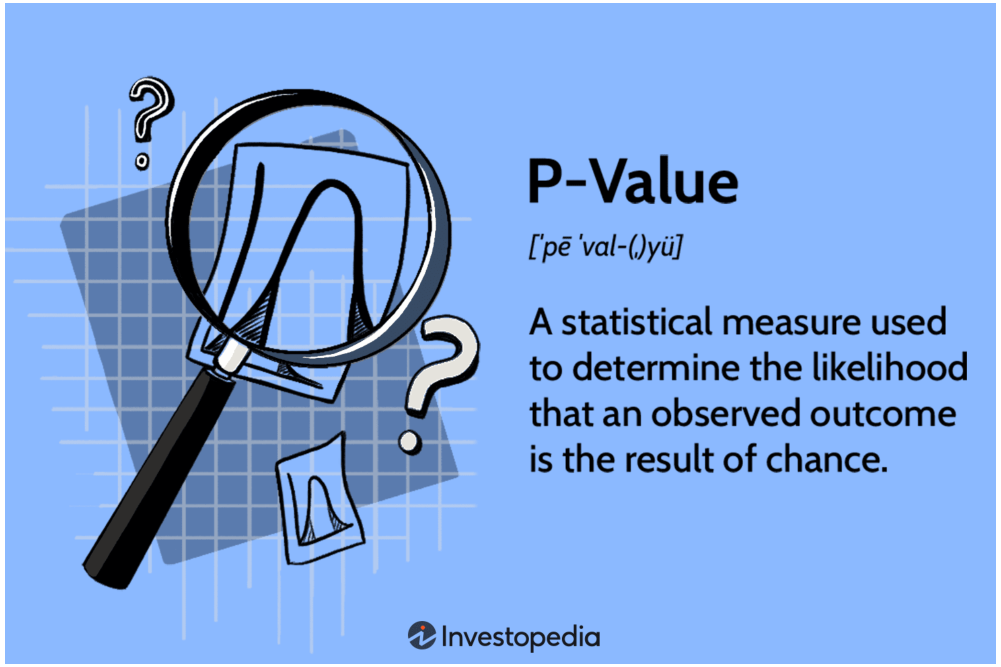
- A value computed from the sample data used to decide whether to reject H_0
- For our use case will typically be a one or two tailed t-test
- Though a z-test can also be written



P Value

- The probability of obtaining test results at least as extreme as the observed data, assuming H_0 is true.

- If $p \leq \alpha$, reject H_0 otherwise fail to reject H_0



Critical Value

- Found using a statistics software or a t-table
- Uses our significance level and degrees of freedom (n-1) to find a value to compare our test-statistic to

Critical values of *t* for two-tailed tests

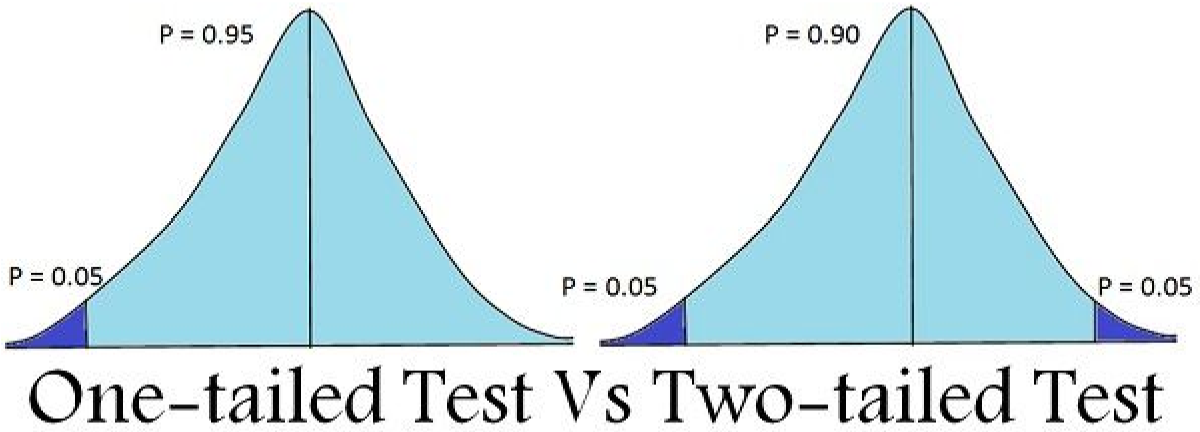
Significance level (α)

Degrees of freedom *(df)*	.2	.15	.1	.05	.025	.01	.005	.001
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	636.619
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.599
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.924
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.610
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.869
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.959
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.408
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.041
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.781
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.587
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.437
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.318
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.221
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.140
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.073
16	1.337	1.512	1.746	2.120	2.473	2.921	3.252	4.015
17	1.333	1.508	1.740	2.110	2.458	2.898	3.222	3.965
18	1.330	1.504	1.734	2.101	2.445	2.878	3.197	3.922
19	1.328	1.500	1.729	2.093	2.433	2.861	3.174	3.883
20	1.325	1.497	1.725	2.086	2.423	2.845	3.153	3.850
21	1.323	1.494	1.721	2.080	2.414	2.831	3.135	3.819
22	1.321	1.492	1.717	2.074	2.405	2.819	3.119	3.792
23	1.319	1.489	1.714	2.069	2.398	2.807	3.104	3.768
24	1.318	1.487	1.711	2.064	2.391	2.797	3.091	3.745
25	1.316	1.485	1.708	2.060	2.385	2.787	3.078	3.725
26	1.315	1.483	1.706	2.056	2.379	2.779	3.067	3.707
27	1.314	1.482	1.703	2.052	2.373	2.771	3.057	3.690
28	1.313	1.480	1.701	2.048	2.368	2.763	3.047	3.674
29	1.311	1.479	1.699	2.045	2.364	2.756	3.038	3.659
30	1.310	1.477	1.697	2.042	2.360	2.750	3.030	3.646
40	1.303	1.468	1.684	2.021	2.329	2.704	2.971	3.551
50	1.299	1.462	1.676	2.009	2.311	2.678	2.937	3.496
60	1.296	1.458	1.671	2.000	2.299	2.660	2.915	3.460
70	1.294	1.456	1.667	1.994	2.291	2.648	2.899	3.435
80	1.292	1.453	1.664	1.990	2.284	2.639	2.887	3.416
100	1.290	1.451	1.660	1.984	2.276	2.626	2.871	3.390
1000	1.282	1.441	1.646	1.962	2.245	2.581	2.813	3.300
Infinite	1.282	1.440	1.645	1.960	2.241	2.576	2.807	3.291

Types of hypothesis tests

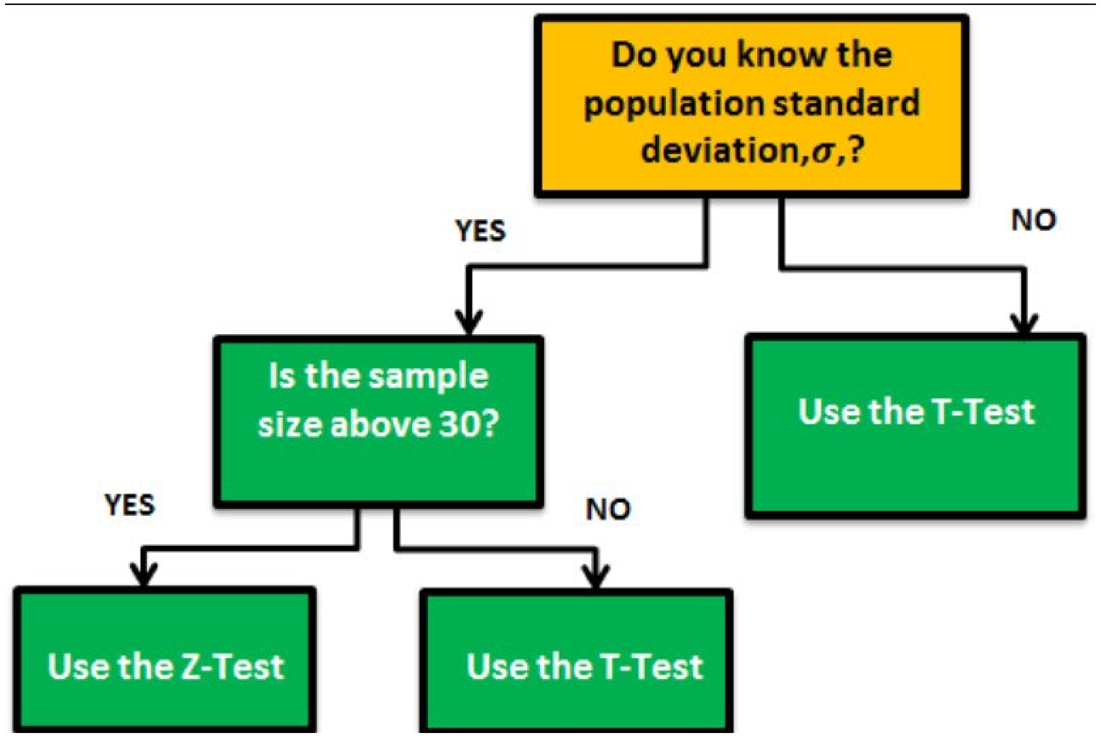
- One-tailed test: Tests for an effect in one direction (e.g., greater than or less than).

- Two-tailed test: Tests for an effect in either direction (e.g., different but unspecified direction).



Common hypothesis tests

- Z-test: Used when population variance is known, and sample size is large ($n \geq 30$).
- T-test: Used when population variance is unknown, and sample size is small.
- Chi-square test: Used for categorical data to test independence or goodness-of-fit.
- ANOVA (Analysis of Variance): Compares means across multiple groups.



Different samples

The equation we use changes depending on how many sets of data (columns) we have, for both these equations the data must be independent (not linked)

One sample example:
A factory produces metal rods, and the average rod length is 50 cm. A manager takes a random sample of 30 rods from a new production batch and wants to test if their average length is different from 50 cm.

Two sample example:
A nutritionist is testing two different diets. She randomly selects 30 people for Diet A and 30 people for Diet B and records their weight loss over 3 months.

One-Sample T-Test

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ = obersved mean of the sample
μ = assumed mean
s = standard deviation
n = sample size

Two-Sample T-Test

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = observed mean of 1st sample
$\bar{x}_2$ = observed mean of 2nd sample
s_1 = standard deviation of 1st sample
s_2 = standard deviation of 2nd sample
n_1 = sample size of 1st sample
n_2 = sample size of 2nd sample

Paired T-test

- When one sample relates to another sample we use a paired t-test.
- So, if we have sample a and then we do something to sample a to get sample b we use a paired t-test
- Paired t-test is super important and most used
- **Example:**

A researcher wants to test whether a new sleep aid increases sleep duration. They conduct an experiment where 10 participants record their sleep duration (in hours) for one week without the sleep aid and then for another week with the sleep aid.

Paired T-test Formula

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2) - (\sum d)^2}{n-1}}}$$

where d: difference per paired value

n: number of samples

Paired t-test example

- This is example data for the example in the previous slide
- We know its paired as it's the same group just with and without aid
- Our first step is to declare our hypotheses
- H_0 sleep aid does not increase sleep duration
- H_a sleep aid increases sleep duration

Participant	Sleep without aid (hrs)	Sleep with aid (hrs)
1	6.5	7.2
2	5.8	6.5
3	6.0	6.8
4	6.2	7.1
5	5.5	6.3
6	6.3	7.0
7	6.1	7.4
8	5.9	6.8
9	5.7	6.6
10	6.4	7.2

Paired t-test example

- Our next step is to add a difference column onto our data and then work out the difference between the two values

Participant	Sleep without aid (hrs)	Sleep with aid (hrs)	Difference
1	6.5	7.2	+0.7
2	5.8	6.5	+0.7
3	6.0	6.8	+0.8
4	6.2	7.1	+0.9
5	5.5	6.3	+0.8
6	6.3	7.0	+0.7
7	6.1	7.4	+1.3
8	5.9	6.8	+0.9
9	5.7	6.6	+0.9
10	6.4	7.2	+0.8

Paired t-test example

- Next let's add a d^2 column to the table

Participant	Sleep without aid (hrs)	Sleep with aid (hrs)	Difference	d^2
1	6.5	7.2	+0.7	0.49
2	5.8	6.5	+0.7	0.49
3	6.0	6.8	+0.8	0.64
4	6.2	7.1	+0.9	0.81
5	5.5	6.3	+0.8	0.64
6	6.3	7.0	+0.7	0.49
7	6.1	7.4	+1.3	1.69
8	5.9	6.8	+0.9	0.81
9	5.7	6.6	+0.9	0.81
10	6.4	7.2	+0.8	0.64

Paired t-test example

- Finally lets add a sum row to the bottom of our table
- This means:
- $\sum d = 8.5$
- $\sum d^2 = 7.51$

Participant	Sleep without aid (hrs)	Sleep with aid (hrs)	Difference	d^2
1	6.5	7.2	+0.7	0.49
2	5.8	6.5	+0.7	0.49
3	6.0	6.8	+0.8	0.64
4	6.2	7.1	+0.9	0.81
5	5.5	6.3	+0.8	0.64
6	6.3	7.0	+0.7	0.49
7	6.1	7.4	+1.3	1.69
8	5.9	6.8	+0.9	0.81
9	5.7	6.6	+0.9	0.81
10	6.4	7.2	+0.8	0.64
			8.5	7.51

Paired t-test example

- Now we can put our values into our equation
- $\sum d = 8.5$
- $\sum d^2 = 7.51$
- n = 10
- $t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2)-(\sum d)^2}{n-1}}} = \frac{8.5}{\sqrt{\frac{10*(7.51)-8.5^2}{10-1}}} = 15.105$

Paired T-test Formula

cuemath
THE MATH EXPERT

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2)-(\sum d)^2}{n-1}}}$$

where d: difference per paired value
n: number of samples

Paired t-test example

- Next we need to work out our critical value
- We know our df is n-1 so 10-1=9
- If we have a significance of 0.05 we have a critical value of 2.262

Critical values of *t* for two-tailed tests

Significance level (α)

Degrees of freedom (*df*)	.2	.15	.1	.05	.025	.01	.005	.0
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	63
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.

Paired t-test example

- Finally, we compare our test statistic and critical value:
- $t = 15.105$
- $crit = 2.262$
- As t is much greater than the critical, we can reject the null
- This means there is significant chance that the hypothesis of the sleep aid increasing sleep duration