Data Science Campus

Hypothesis Testing

Overview

- Introduction to hypothesis testing
- Principle ideas of the technique
- Example
- Resources

Introduction

- Scientific claims about reality are called *hypotheses.*

- Any hypothesis should be supported by evidence.

- If they are verified by observations, hypotheses are accepted as true, at least until proved otherwise.

Coin toss

- We are given a coin and asked to determine whether or not it is "fair".

- Therefore we accept that it is fair.
  - We call this the null hypothesis, H_0 or H_N

- We attempt to provide enough evidence to reject this and accept an alternative.
  - We call this the alternative hypothesis, H_1 or H_A

Coin toss

- Let's say we toss the coin 10 times and a total of 9 heads are observed.
- Now, if H_0 is true, getting 9 heads from 10 seems unlikely.
- However, we need to determine if the probability of this occurring by chance is low enough to allow us to reject the null hypothesis and ultimately conclude that the coin is not fair.

Coin toss

- Tossing a coin can be modelled as a binomial variable (set of trials with 2 outcomes)
  - Toss the coin n times
  - Proportion p of heads

- The hypothesis that a coin is fair is to test the hypothesis that $\mathrm{p} = {}^{1}/_{2}$, $H_0\colon \mathrm{p} = {}^{1}/_{2}$

- For a one-tailed test, $H_1\colon p < {}^{1}/_{2}$ or $H_1\colon p > {}^{1}/_{2}$
- For a two tailed test, $H_1\colon p \neq {}^{1}/_{2}$

Type I and Type II errors

- Type I errors – False positives
  - If H_0 is rejected when it is actually true
- Type II errors – False negatives
  - If H_0 is accepted when it was actually false

Decision	H_0 is true	H_0 is false
Reject H_0	Type I error (α)	Correct (1-β)
Do not reject H_0	Correct (1-α)	Type II error (β)

Type I errors

- The probability of committing a Type I error is denoted by α (the significance level).

- If we set the significance level at $\alpha = 0.05$ there is a 5% probability of rejecting H_0, when H_0 is actually correct.

Type II errors

- The probability of committing a Type II error is denoted by β.

$1-\beta$ *is known as the statistical power of the test.*

- The probability of committing a Type II error depends on several factors, including the sample size.

Significance level

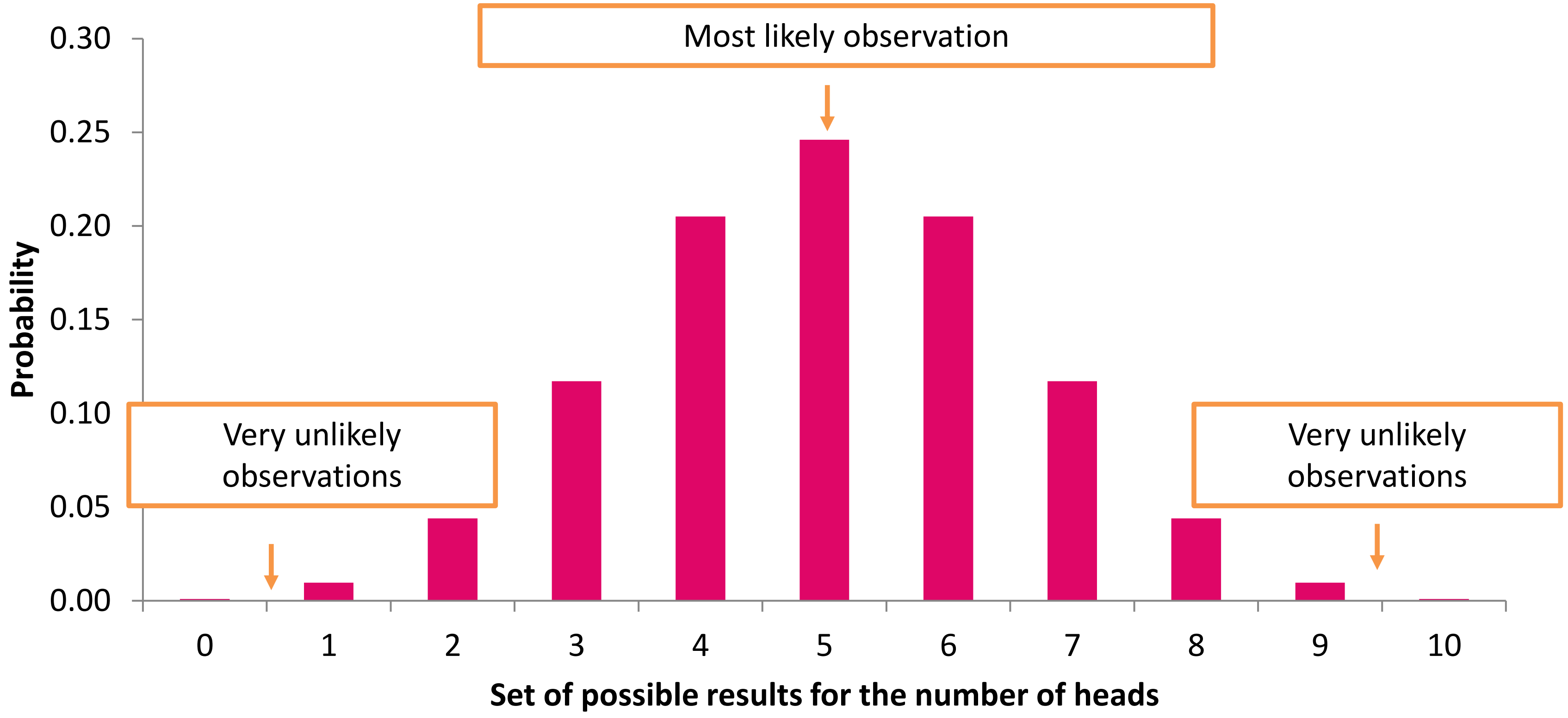
- First we need to decide a significance level for the test. This is denoted by α (alpha).

- This is the probability of rejecting the null hypothesis when it is actually true.

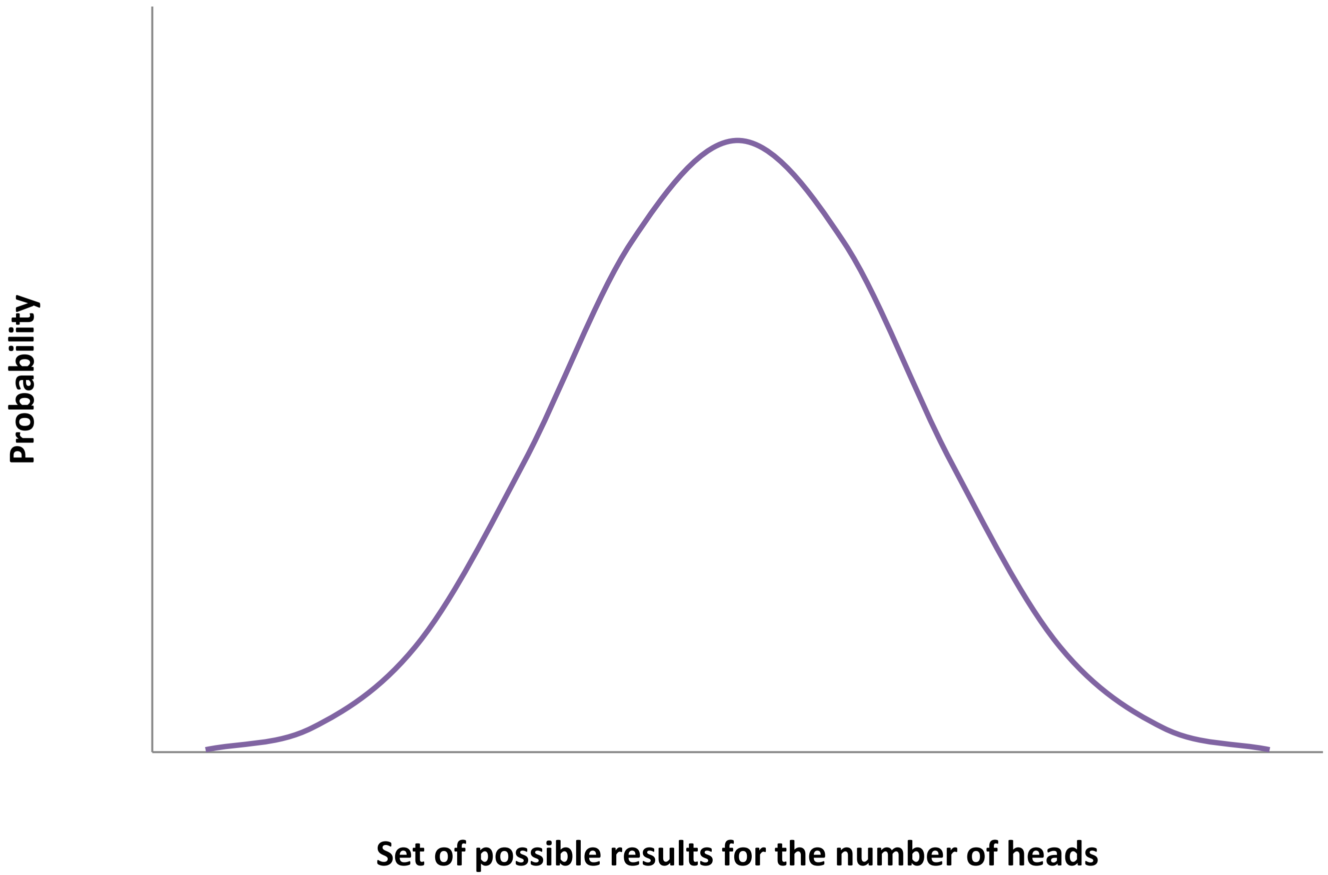
- So for our coin toss that would mean:
  - ➢ Determining the coin was not fair when it actually is.

- In general, this is usually around 5%.

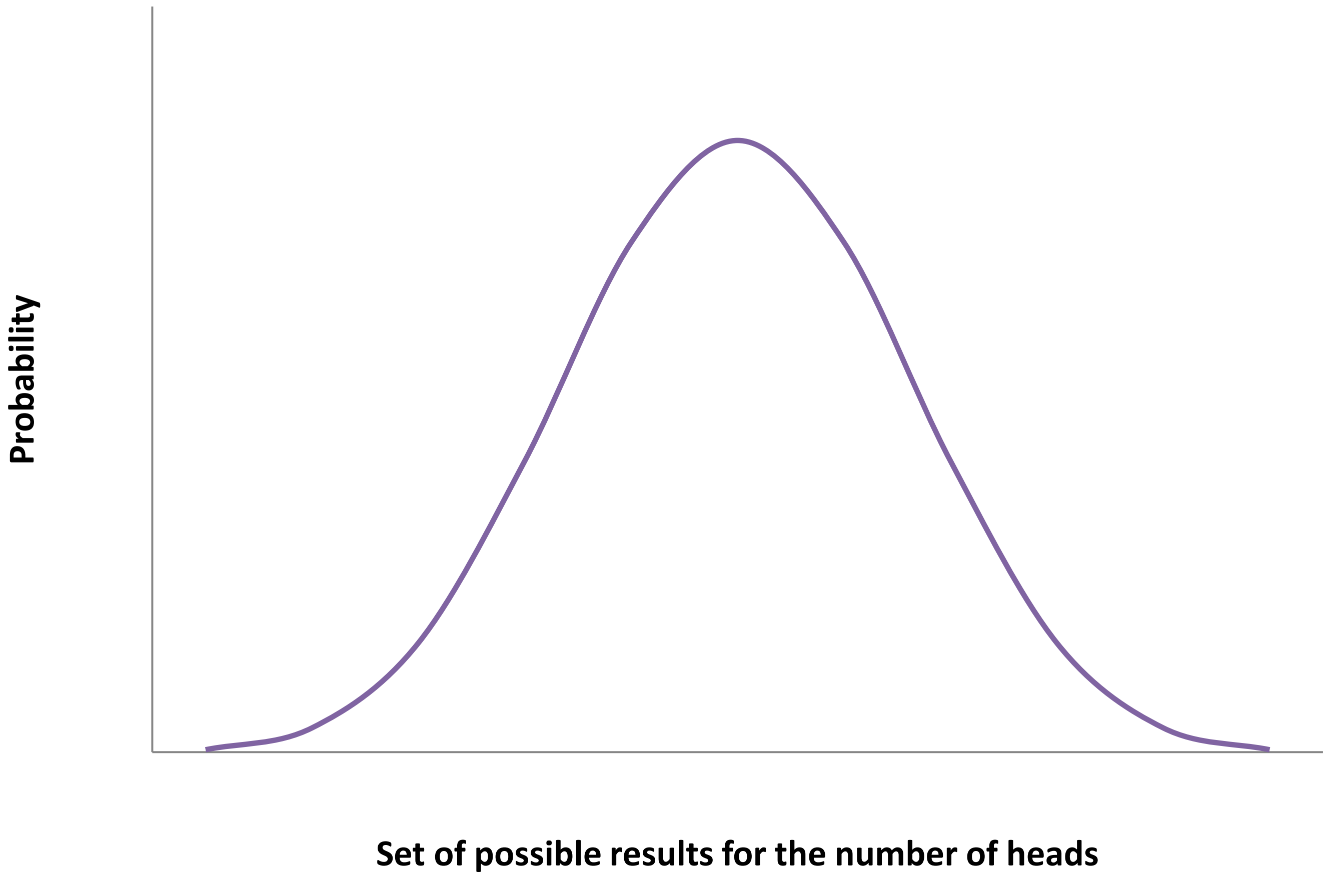
Probability mass function



$H_1\text{: p} > {}^{1}/_{2}\ ,\ \alpha = 0.05$



$H_1: \text{p} \neq {}^{1}/_{2}$, $\alpha = 0.05$



p-value

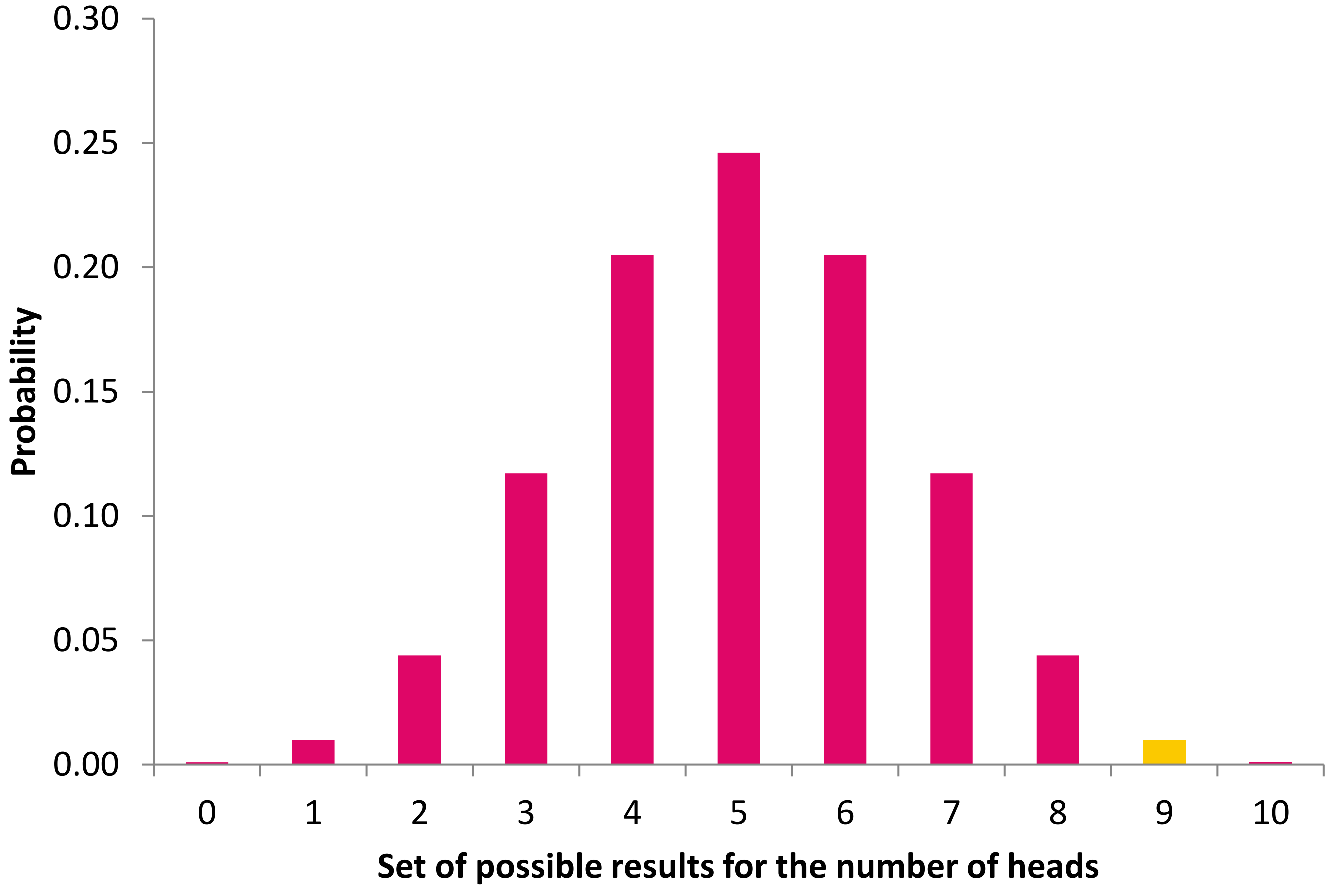
- The probability of obtaining a result at least as extreme, assuming H_0 is true.

- We toss the coin 10 times and a total of 9 heads are observed.

- Let X be the number of heads obtained, $X \sim B(10{,}0.5)$.

- $\mathrm{P}(\mathrm{X} \geq 9) = P(X = 9) + P(X = 10) = 0.01074.$

p-value



0.30
0.25
0.20
0.15
0.10
0.05
0.00
Probability
0
1
2
3
4
5
6
7
8
9
10
Set of possible results for the number of heads

Result

- p-value $= 0.01074$
- Two tailed test, $H_1 : p \neq 1/2$, $\alpha = 0.05$
- p-value < 0.025 we reject H_0
- Can conclude, at a significance level of 5% that the coin is not fair.

Example - Correlation

- One way to quantify the relationship between two variables is to use the Pearson correlation coefficient, which is a measure of the linear association between two variables.

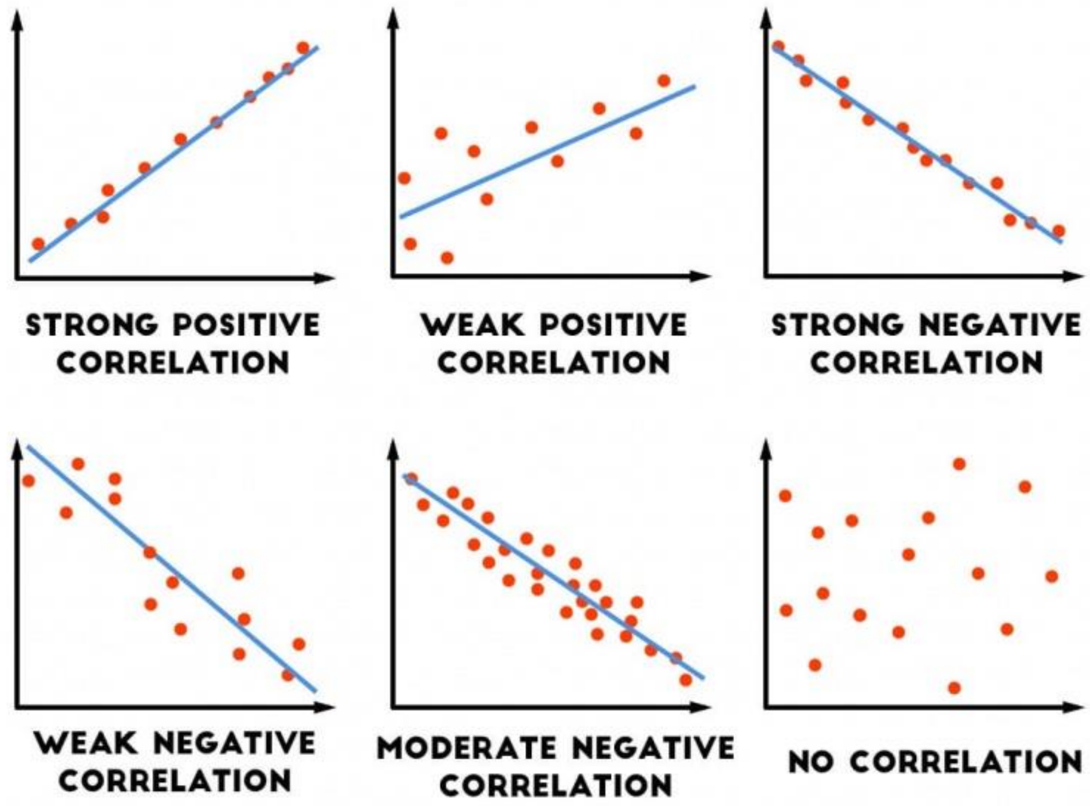
- To determine if a correlation coefficient is statistically significant, you can calculate the corresponding test statistic and p-value.

Correlation coefficient

The correlation coefficient always takes on a value between -1 and 1 where:

- -1 indicates a perfectly negative linear correlation between two variables
- 0 indicates no linear correlation between two variables
- 1 indicates a perfectly positive linear correlation between two variables

Correlation coefficent



Test statistic

- A test statistic assesses how consistent your sample data are with the null hypothesis in a hypothesis test.

- Test statistic calculations take your sample data and boil them down to a single number that quantifies how much your sample diverges from the null hypothesis.

- As a test statistic value becomes more extreme, it indicates larger differences between your sample data and the null hypothesis.

Test statistic

- When your test statistic indicates a sufficiently large incompatibility with the null hypothesis, you can reject the null and state that your results are statistically significant.

- To use a test statistic to evaluate statistical significance, you either compare it to a critical value or use it to calculate the p-value.

Test statistic

- When a t-value equals 0, it indicates that your sample data match the null hypothesis exactly.

- The significance level uses critical values to define how far the test statistic must be from the null value to reject the null hypothesis.

- When the test statistic exceeds a critical value, the results are statistically significant.

Test statistic

- The formula to calculate the t-score of a correlation coefficient (r) is:

$$t = r * \sqrt{n-2} / \sqrt{1-r2}$$

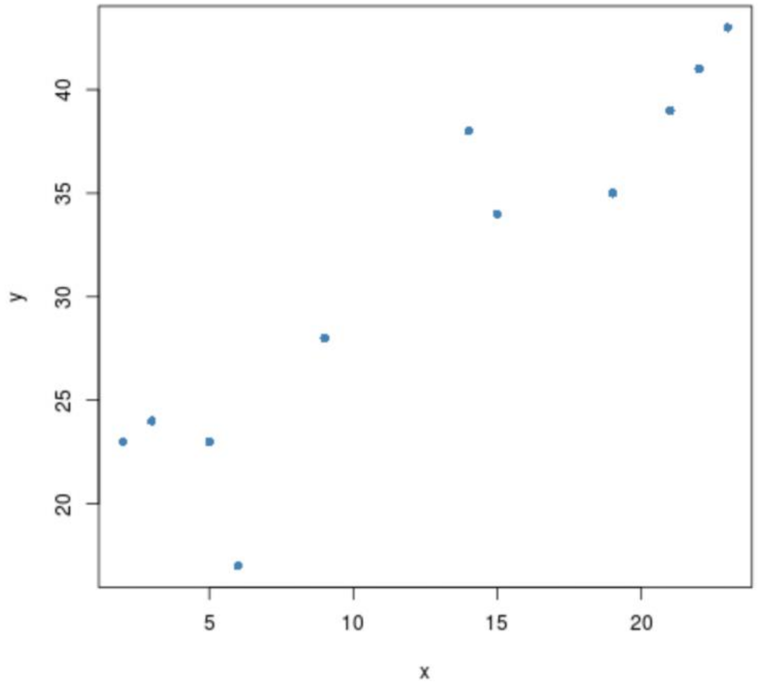
- The p-value is calculated as the corresponding two-sided p-value for the t-distribution with n-2 degrees of freedom.

Example

Suppose we have the following 2 vectors:

x = {2, 3, 3, 5, 6, 9, 14, 15, 19, 21, 22, 23}
y = {23, 24, 24, 23, 17, 28, 38, 34, 35, 39, 41, 43}

There appears to be a positive correlation between the two variables. That is, as one increases the other tends to increase as well.



Example

- The correlation coefficient between the two vectors turns out to be 0.9279869.

- The test statistic turns out to be 7.8756 and the corresponding p-value is 1.35e-05.

- Since this value is less than .05, we have sufficient evidence to say that the correlation between the two variables is statistically significant.

Example

Pearson's product-moment correlation
t = 7.8756, df = 10, p-value = 1.35e-05

df\ α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

Next steps

- Python – independent study of Hypothesis Testing in Python course.
- R - independent study of Chapter 3 of Statistics in R course.

Questions?