MATLAB Introduction
MATLAB stands for MATrix LABoratory

It provides an easy-to-use interface to perform mathematical computations, create plots, analyze data etc.
Variable definition: \( x = 5 \)

Matrix definition:

\[
A = \begin{bmatrix}
16 & 3 & 2 & 13 \\
5 & 10 & 11 & 8 \\
9 & 6 & 7 & 12 \\
4 & 15 & 14 & 1 \\
\end{bmatrix}
\]

In general: matrix = [line 1; line 2... ; line N]

Useful note: indices start from 1 and not from 0!
Lot of useful functions:

- sum
- mean
- median
- abs
- ...

You can also create your own functions as in every other programming language

```
function [ return variables ] = function_name( arguments )
    ...
    ...
    ...
end
```
And for loops...

```
for i = starting value : step : ending value
  ...
  ...
  ...
  ...
  ...
end
```
And if statements...

- if condition1==TRUE ...
  - ...
  - ...
- elseif condition2==TRUE
  - ...
  - ...
- else
  - ...
  - ...
- end
How to plot data:

- `figure`
- `plot(x,y)`
- `xlabel('x axis name')`
- `ylabel('y axis name')`
- `title('title of plot')`
- `legend('line 1','line 2',...) % if you plot more than two lines`
Standard deviation

std function

Gaussian distribution

99.7% of the data are within 3 standard deviations of the mean
95% within 2 standard deviations
68% within 1 standard deviation
Percentiles

A percentile is a number where a certain percentage of scores fall below that number.

\[ x = \text{prctile}(\text{vector}, \text{desired percentile}) \]

**Question:** What is the value of the 50th percentile on the Gaussian?
You can also see interesting things by creating a histogram:

\[
\text{histogram}(x)
\]

You can also specify the bins that are created by setting the number of bins:

\[
\text{histogram}(x, \text{bin\_number})
\]
Cumulative distribution function

a function whose value is the probability that a corresponding continuous random variable has a value less than or equal to the argument of the function

- \([y, x] = \text{ecdf}(\text{vector});\)
- \(\text{plot}(x, y)\)

Can you spot the 50th percentile here?
K-S test
How to run and interpret

\[ \text{[hypothesis_result, pvalue]} = \text{kstest2(variable 1, variable 2)} \]

This test examines if the two input distributions originate from the same distribution.

It calculates the maximum distance between the ECDFs of the two distributions.

It returns a \( p \)-value that expresses the possibility of two random selected samples (one of each distribution) having the max distance.

**Null hypothesis**: input distributions originate from the same distribution

*p-value > 0.05*: if \( p \)-value is not in the 5% of the most extreme values, then the result is not statistical significant and the null hypothesis is not rejected.

*p-value < 0.05*: if \( p \)-value is in the 5% of the most extreme values, we reject the null hypothesis and consider the result statistically significant.
You can plot two variables together to discover how they are related!

```
scatter(var1, var2)
```

Note that here we cannot see any linear relation between the variables.

use `corr()` function to check if two variables are linearly correlated
**corr(x,y) - Pearson Correlation**

**How to interpret it?**

Pearson Correlation is a measure of the linear correlation between two variables.

It takes values in $[-1, 1]$

- *Negative value* (negative correlation) means that when $x$ is incrementing, $y$ is decreasing.
- *Positive value* (positive correlation) means that when $x$ is incrementing, $y$ is incrementing too.
- *Zero* correlation means that data are randomly distributed.
For more information about MATLAB

https://www.mathworks.com/help/

or Google it :)