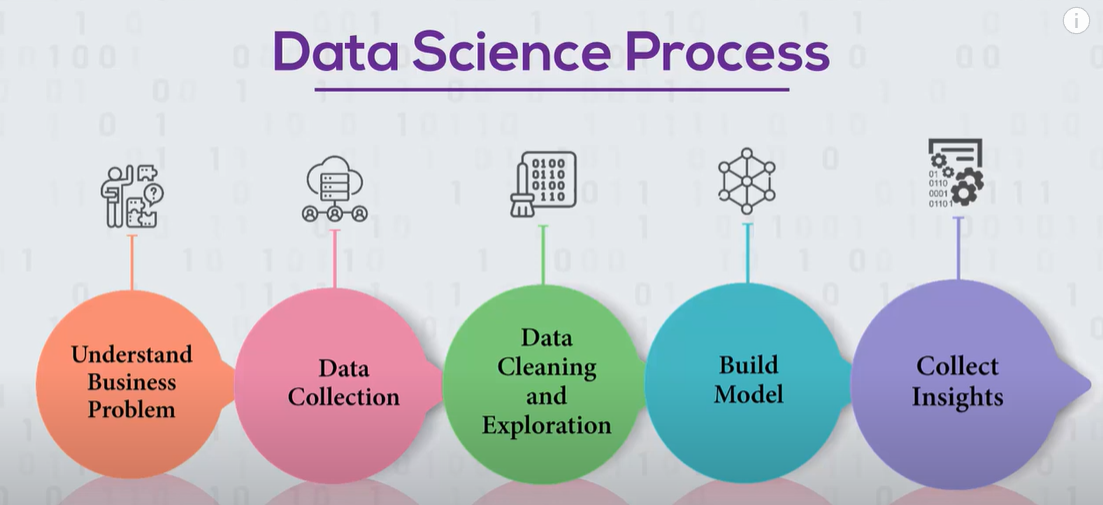
# Data Science



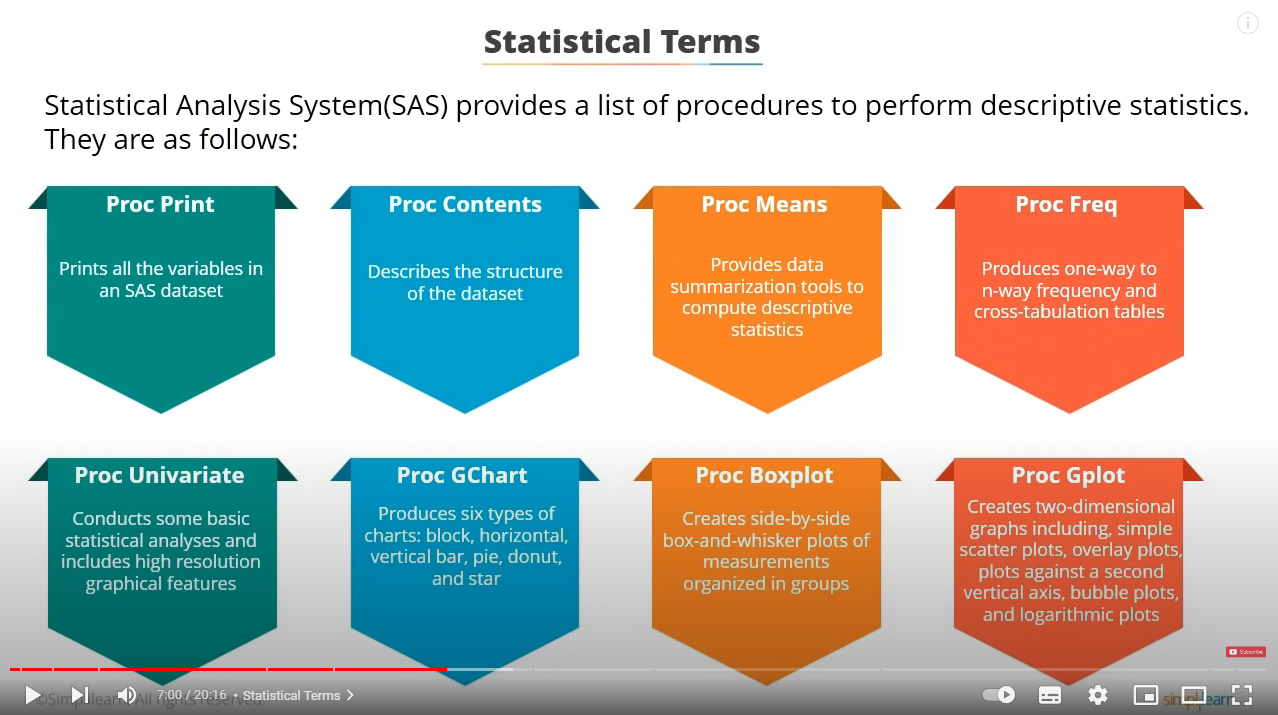
# applications

## Health care

Predict disease based on measuring wearable divide. Monitoring sleeping patters to predict heart disease.

# Statistics

<https://www.youtube.com/watch?v=Lv0xcdeXaGU>



## scipy

### Describe

Ytb [link](https://www.youtube.com/watch?v=REiAGmXwZNw)

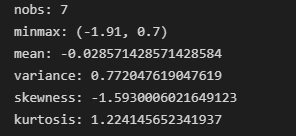
from scipy.stats import describe

data = [0.2, -1.91, 0.41, 0.7, -0.03, 0.53, -0.1]

desciption = describe(data)

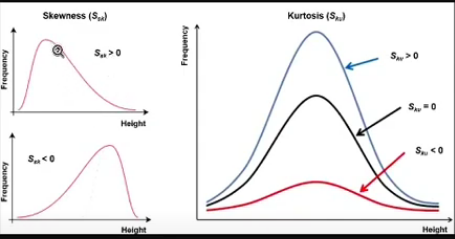
for key, value in desciption.\_asdict().items():

    print(f"{key}: {value}")



Mean – average

Variance – how spread the dataset

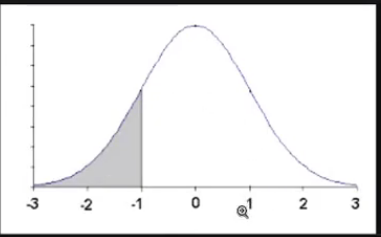


Skewness – symmetry of dataset

F1 shows large outlier on right side

F2 shows outlier on left side

### normal distribution



from scipy.stats import norm

import numpy as np

# mean, standard deviation

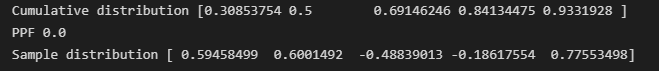
distribution = norm(0, 2)

array = np.array([-1, 0.0, 1, 2, 3])

print("Cumulative distribution", distribution.cdf(array))

print("PPF", distribution.ppf(0.5))

print("Sample distribution", distribution.rvs(size = 5))



cdf for -1 is 0.3085… it is the area under curve

Ppf is reverse of cdf for eg(0.5 is cdf of 0)

Sample distribution – sample some values

### ttest

from scipy import stats

# Manufacture claims that the volume of a drink is 255 ml

# is this true?

# volumes look like they are lower than 255 ml

# level of significance = 0.05

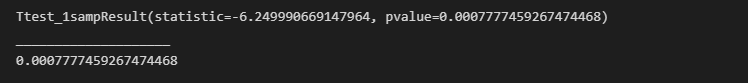
volumes = [220.1, 220.5, 221.2, 221.8, 222.5, 223.1, 223.7]

result = stats.ttest\_1samp(volumes, 225)

print(result)

print("\_"\*20)

print(result[1])



If value less than 0.5 then the hypothesis is false

## Statistics tutorial

### Probability Density Function

<https://www.simplilearn.com/tutorials/statistics-tutorial/probability-density-function>

from matplotlib import pyplot

from numpy.random import normal

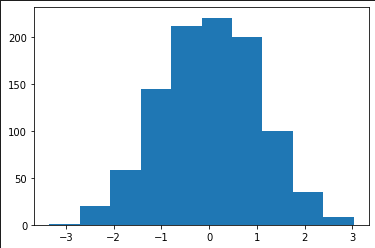
from numpy import mean, std

from scipy.stats import norm

sample = normal(size=1000)

pyplot.hist(sample, bins=10)

pyplot.show()



sample = normal(loc=50, scale=5, size=1000)

sample\_mean = mean(sample)

sample\_std = std(sample)

print("Mean", sample\_mean, "STD", sample\_std)

dist = norm(sample\_mean, sample\_std)

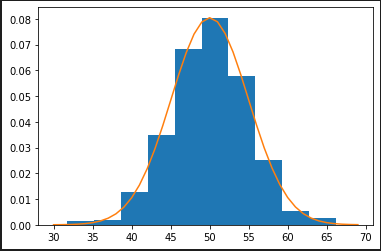
values = [value for value in range(30, 70)]

probabilities = [dist.pdf(value) for value in values]

pyplot.hist(sample, bins=10, density=True)

pyplot.plot(values, probabilities)

pyplot.show()



### central-limit-theorem

<https://www.simplilearn.com/tutorials/statistics-tutorial/central-limit-theorem>

### measures-of-central-tendency

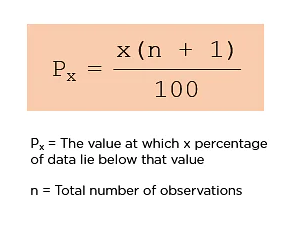
<https://www.simplilearn.com/tutorials/data-analytics-tutorial/measures-of-central-tendency>

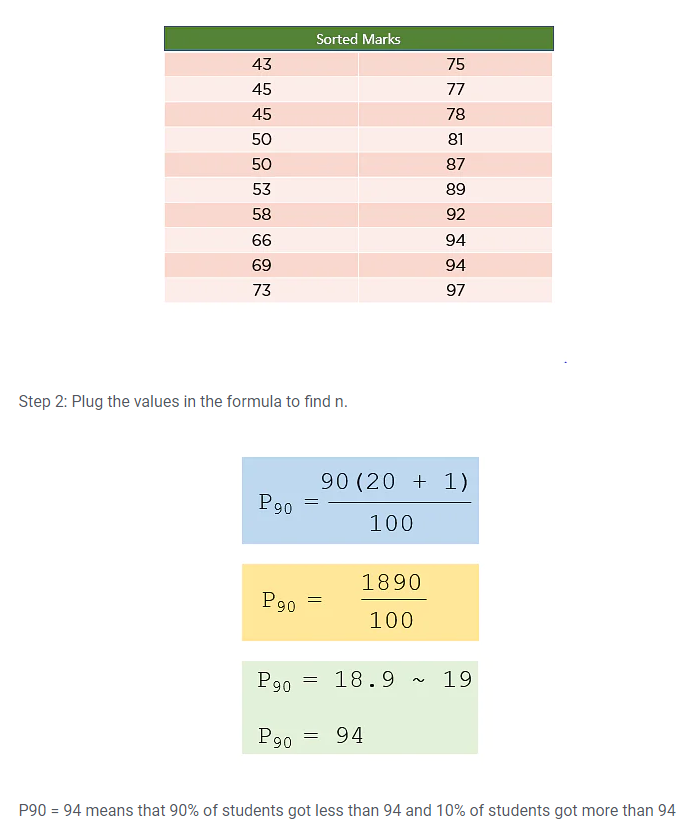
### conditional-probability

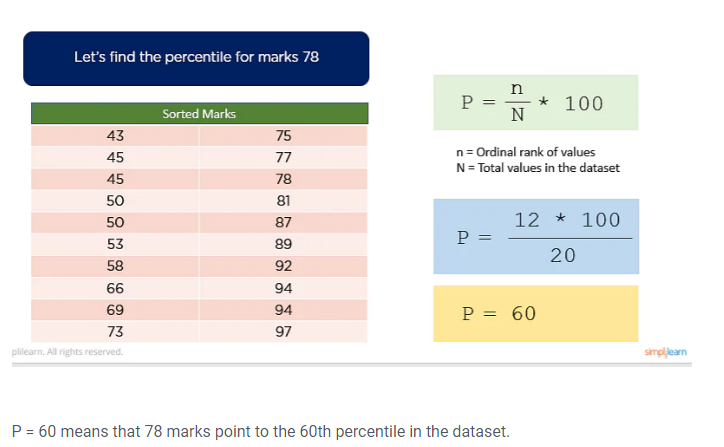
<https://www.simplilearn.com/tutorials/statistics-tutorial/conditional-probability>

### percentile-in-statistics

<https://www.simplilearn.com/tutorials/data-analytics-tutorial/percentile-in-statistics>







### bayes-theorem

<https://www.simplilearn.com/tutorials/statistics-tutorial/bayes-theorem>

An insurance company insured 2000 scooter drivers, 4000 car drivers, and 6000 truck drivers. The probability of an accident involving a scooter driver, car driver, and a truck is 0.01, 0.03, and 0.015 respectively. One of the insured persons meets with an accident. What is the probability that he is a scooter driver?

Let E1, E2, E3, and A be the events defined as follows:

E1 = person chosen is a scooter driver

E2 = person chosen is a car driver

E3 = person chosen is a truck driver and

A = person meets with an accident

Since there are 12000 people, therefore:

P(E1) = 2000/12000 = ⅙

P(E2) = 4000/12000 = ⅓

P(E3) = 6000/12000 = ½

It is given that P(A / E1) = Probability that a person meets with an accident given that he is a scooter driver = 0.01

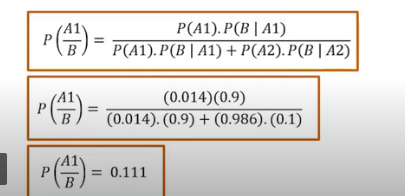
Similarly, you have P(A / E2) = 0.03 and P(A / E3) = 0.15

You are required to find P(E1 / A), i.e. given that the person meets with an accident, what is the probability that he was a scooter driver?

P(E1/A) = P(E1) P(A/E1)P(E1) P(A/E1) + P(E2) P(A/E2) + P(E3) P(A/E3)

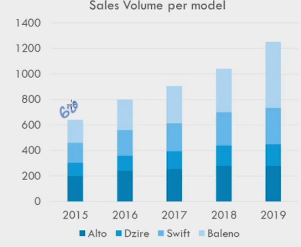
= (1/6 \* 0.01) / ((1/6 \* 0.01) + (1/3 \* 0.03) + (1/2 \* 0.15))

= 1/52



# Data Representation

## Stacked bar chart



Representing the total and year wise growth.

Eg: selling of product over the year