

# Fundamentals of statistics for data science

## Tasks

- Import the NumPy and SciPy libraries.
- Load the dataset heights\_and\_weights.csv.
- Calculate the mean, median, and standard deviation of the heights and weights in the dataset using NumPy.
- Calculate the skewness and kurtosis of the heights and weights in the dataset using SciPy.
- Print the results.

```
In [4]: import pandas as pd  
import numpy as np  
import scipy as scp
```



## Read Data

```
In [3]: df = pd.read_csv("height_wieight.csv")  
df
```



Out[3]:

	Index	Height(Inches)	Weight(Pounds)
<b>0</b>	1	65.78331	112.9925
<b>1</b>	2	71.51521	136.4873
<b>2</b>	3	69.39874	153.0269
<b>3</b>	4	68.21660	142.3354
<b>4</b>	5	67.78781	144.2971
...	...	...	...
<b>24995</b>	24996	69.50215	118.0312
<b>24996</b>	24997	64.54826	120.1932
<b>24997</b>	24998	64.69855	118.2655
<b>24998</b>	24999	67.52918	132.2682
<b>24999</b>	25000	68.87761	124.8742

25000 rows × 3 columns

## Mean

```
In [17]: h_mean = np.mean(df['Height(Inches)'])
w_mean = np.mean(df['Weight(Pounds)'])

print("Height Mean: {}\nWidth Mean: {}".format(h_mean,w_mean))
```

Height Mean: 67.99311359679979  
Width Mean: 127.07942116079916

## Median

```
In [18]: h_median = np.median(df['Height(Inches)'])
w_median = np.median(df['Weight(Pounds)'])

print("Height Median: {}\nWidth Median: {}".format(h_median,w_median))
```

Height Median: 67.9957  
Width Median: 127.15775



## Standard Deviation

```
In [19]: h_std = np.std(df['Height(Inches)'])  
w_std = np.std(df['Weight(Pounds)'])  
  
print("Height Standard Deviation: {}\nWidth Standard Deviation: {}".format(h_std,w_std))
```



Height Standard Deviation: 1.9016407372498367  
Width Standard Deviation: 11.66066434332079



## Skewness

It is the degree of distortion from the symmetrical bell curve or the normal distribution. It measures the lack of symmetry in data distribution.

### Positive

Skewness means when the tail on the right side of the distribution is longer or fatter. The mean and median will be greater than the mode.

### Negative

Skewness is when the tail of the left side of the distribution is longer or fatter than the tail on the right side. The mean and median will be less than the mode.

- If the skewness is between -0.5 and 0.5, the data are fairly symmetrical.
- If the skewness is between -1 and -0.5(negatively skewed) or between 0.5 and 1(positively skewed), the data are moderately skewed.
- If the skewness is less than -1(negatively skewed) or greater than 1(positively skewed), the data are highly skewed.

```
In [23]: h_skewness1 = scp.stats.skew(df['Height(Inches)'])  
# OR  
h_skewness2 = (3*(h_mean-h_median))/h_std  
  
print("Height Skewness\n1st way: {}\n2nd way: {}".format(h_skewness1, h_skewness2))
```



Height Skewness  
1st way: -0.005657639882518977  
2nd way: -0.0040802710252500155

```
In [41]: w_skewness1 = scp.stats.skew(df['Weight(Pounds)'])  
# OR  
w_skewness2 = (3*(w_mean-w_median))/w_std  
  
print("Weight Skewness\n1st way: {}\n2nd way: {}".format(w_skewness1, w_skewness2))
```

Weight Skewness  
1st way: -0.026029783883831488  
2nd way: -0.020152069443376747

## Kurtosis

It is all about the tails of the distribution—not the peakedness or flatness. It is used to describe the extreme values in one versus the other tail. It is actually the measure of outliers present in the distribution.

### High kurtosis

in a data set is an indicator that data has heavy tails or outliers. If there is a high kurtosis, then, we need to investigate why do we have so many outliers. It indicates a lot of things, maybe wrong data entry or other things.

### Low kurtosis

in a data set is an indicator that data has light tails or lack of outliers. If we get low kurtosis(too good to be true), then also we need to investigate and trim the dataset of unwanted results.

```
In [45]: h_kurtosis = scp.stats.kurtosis(df['Height(Inches)'])  
print("Height Kurtosis: {}".format(h_kurtosis))
```

Height Kurtosis: -0.03539236835811055

```
In [46]: w_kurtosis = scp.stats.kurtosis(df['Weight(Pounds)'])  
print("Weight Kurtosis: {}".format(w_kurtosis))
```

Weight Kurtosis: 0.044491674304663054

In [ ]:

