**Population :**

Entire group of data or whole data ,that you want to draw conclusion about .

**Sample :**

Subset of population

**Types of Statistics**

i) **Descriptive Statistics :**

- Refer to methods for summarizing raw observations into information that we can understand & share

which includes calculation of

a) 'Central Tendency': Mean,Median,Mode

b) 'Spread of data': variance , Standard Deviation

c) 'Visualize data': plot central tendency & spread of data

ii) **Inferential Statistics :**

- Refers to quantifying properties of data(sample|population) .

which includes

'hypothesis testing'

**Sample & Population**

**Population** : All possible observations from a group.

**Sample** : A subset of observations from a group

1. Central Tendency : Single value that tells about data distribution.
2. **Mean :**

**Arithmetic mean** :

Averaged of group of numbers

Applicable only for ‘interval’ & ‘ratio’ data-types

Affected by each data-points, including outlier

**Weighted mean (mean of grouped data) :**

**Median** :

- Middle value of ordered series of data-points

- Applicable only for ‘ordinal’, ‘interval’ , ‘ratio’ data-types

- Not affected by extreme large values(outliers)

**Weighted Median :**

**Mode** :

- Most frequent occurred value in data

- Applicable for ‘nominal’ , ‘ordinal’ , ‘interval’ , ‘ratio’

**Weighted Mode :**

**Percentile :**

**-** Measure Central tendency by dividing data into 100 parts

- Applicable for ‘ordinal’ , ‘interval’ , ‘ratio’ data-types

ex: 90 percentile of data: 90% of data under that

Formula

**Quartile :**

Measure Central tendency by dividing data into 4 sub-groups

Q1 = 25% below 1st quartile

Q2 = 50% below 2nd quartile(median)

Q3 = 75% below 3rd quartile

1. Dispersion | Spread of Data :

- Measure of variability describe the ‘spread of data’

- Used to calculate reliable centre of data distribution(when ‘central tendency’ misleads)

**Range :** Measured by difference between ‘smallest’ & ‘largest’ value of dataset

Formula range = largest - smallest

**IQR(Inter-Quartile Range) :**

- Measured as data ranges between 1st Quartile & 3rd Quartile

- Less influenced by Outliers

Formula : IQR = Q3 - Q1

**Deviation From Mean:**

How deviate the data-points from mean

**Mean Absolute Deviation(MAD) :**

Average of absolute deviation of data-points from mean

**Variance :**

- Average of Squared Deviation of data-points from Mean

- low variance: values grouped around mean (narrow bell-shape)

- high variance:values spread out from mean (wide bell-shape)

**Standard Deviation :**

- Square-root of Average of Squared Deviation of data-points from Mean .

- Square-root of Variance of data

**Skewness :**

- Measure of absence of Symmetry in data-distribution

- Shows presence of Extreme Values in one-side of data-distribution

- It shows the shape of data-distribution

**Kurtosis :**

- Measure of Peakedness of data-distribution

- Shows degree to given distribution is more | less “Peaked” relative to “Normal Distribution”

**c) Data Distributions**

- Refers to shape 'graph',when it plotted .

a) ***Gaussian Distribution :***

- Many data fits in Gaussian / Normal distribution

- 'Data graph'forms a "bell curve" during data plotting

- It is helpful when data is ‘ in Gaussian / assume as Gaussian distribution’ for calculating statistics cuz have many ‘statistical methogs’

e:g People’s incomes , Population of cities ,Sales of books

# generate and plot an idealized gaussian

from numpy import arange

from matplotlib import pyplot

from scipy.stats import norm

# x-axis for the plot

x\_axis = arange(-3, 3, 0.001)

# y-axis as the gaussian

y\_axis = norm.pdf(x\_axis, 0, 1)

# plot data

pyplot.plot(x\_axis, y\_axis)

pyplot.show()

b) ***Uniform Distribution :***

- It is another common distribution, often seen when each item or value has an equal value for being selected. The shape of a graph of the uniform distribution is a at line.

Exception :

|  |  |
| --- | --- |
| [StopIteration](https://www.python-ds.com/python-3-built-in-exceptions#StopIteration) |  |

|  |  |
| --- | --- |
| * [StopAsyncIteration](https://www.python-ds.com/python-3-built-in-exceptions#StopAsyncIteration) |  |

|  |  |
| --- | --- |
| [ArithmeticError](https://www.python-ds.com/python-3-built-in-exceptions#ArithmeticError) |  |
| * [FloatingPointError](https://www.python-ds.com/python-3-built-in-exceptions#FloatingPointError) |  |
| * [OverflowError](https://www.python-ds.com/python-3-built-in-exceptions#OverflowError) |  |
| * [ZeroDivisionError](https://www.python-ds.com/python-3-built-in-exceptions#ZeroDivisionError) |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| datetime  'astimezone',  'combine',  'ctime',  'date',  'day',  'dst',  'fold',  'fromordinal',  'fromtimestamp',  'hour',  'isocalendar',  'isoformat',  'isoweekday',  'max',  'microsecond',  'min',  'minute',  'month',  'now',  'replace',  'resolution',  'second',  'strftime',  'strptime',  'time',  'timestamp',  'timetuple',  'timetz',  'today',  'toordinal',  'tzinfo',  'tzname',  'utcfromtimestamp',  'utcnow',  'utcoffset',  'utctimetuple',  'weekday',  'year' | datetime.date  **ctime',**  **'day',**  **'fromordinal',**  **'fromtimestamp',**  **'isocalendar',**  **'isoformat',**  **'isoweekday',**  **'max',**  **'min',**  **'month',**  **'replace',**  **'resolution',**  **'strftime',**  **'timetuple',**  **'today',**  **'toordinal',**  **'weekday',**  **'year'** | time  **'dst',**  **'fold',**  **'hour',**  **'isoformat',**  **'max',**  **'microsecond',**  **'min',**  **'minute',**  **'replace',**  **'resolution',**  **'second',**  **'strftime',**  **'tzinfo',**  **'tzname',**  **'utcoffset'** | **Time delta**  **'days',**  **'max',**  **'microseconds',**  **'min',**  **'resolution',**  **'seconds',**  **'total\_seconds'** | tzinfo | Timezone  'dst',  'fromutc',  'max',  'min',  'tzname',  'utc',  'utcoffset' |

|  |  |
| --- | --- |
|  | Date |
| “%y” | Year short(yy) (Ex: 21) |
| “%Y” | Year full (yyyy) (Ex: 2021) |
| “%a” | Week-day short (E.x: wed ) |
| “%A” | week-day full (E.x: wednesday ) |
| “%w” | week-day as no (E.x: 0:sunday ) |
| “%d” | Day of Month (1-31) |
| “%b” | Month Name short (E.x: Dec ) |
| “%B” | Month Name short (E.x: December ) |
| “%m” | Month no. (1 - 12) |
| “%j” | Day No. of year (001 - 366) |
| “%U” | Week No. of year (00 - 53) (Sunday: day 1) |
| “%W” | Week No. of year (00 - 53) (Monday: day 1) |
| “%c” | Local date & time (D M d hr:min:sec Y) |
| “%x” | Local Version Of Date (mm/dd/yy) |
| “%X” | Local Version Of time (hr : min : sec) |
|  | Time & Time Zone |
| “%H” | Hour (00-23) |
| “%I” | Hour (00-12) |
| “%p” | AM | PM |
| “%M” | Minute (00-59) |
| “%S” | Second (00-59) |
| “%f” | Micro-Second (000000-999999) |
| “%z” | UTC Offset |
| “%Z” | Time Zone (CST) |
| “%%” | '%' char |