Role of Statistics in EDA

What is Statistics?

What is Statistics?

Statistics is set of rules and concepts for analysis and interpretation of the data

Role of Statistics in EDA: Univariate

Statistics is set of rules and concepts for analysis and interpretation of the data

- **Univariate** : central tendencies, distributions and spread of data

Role of Statistics in EDA: Univariate

1458.71
5390.37
3913.16
2291.91
927.72
15202.20
7006.93
10096.58
1355.86
4957.95
1928.76
6886.40
554.38
4548.48
5431.42
2010.98

Role of Statistics in EDA: Univariate

- Central tendency

1458.71
5390.37
3913.16
2291.91
927.72
15202.20
7006.93
10096.58
1355.86
4957.95
1928.76
6886.40
554.38
4548.48
5431.42
2010.98

Role of Statistics in EDA: Univariate

- Central tendency
- Range

1458.71
5390.37
3913.16
2291.91
927.72
15202.20
7006.93
10096.58
1355.86
4957.95
1928.76
6886.40
554.38
4548.48
5431.42
2010.98

Role of Statistics in EDA: Univariate

- Central tendency
- Range
- Average

1458.71
5390.37
3913.16
2291.91
927.72
15202.20
7006.93
10096.58
1355.86
4957.95
1928.76
6886.40
554.38
4548.48
5431.42
2010.98

Role of Statistics in EDA: Bivariate

Statistics is set of rules and concepts for analysis and interpretation of the data

- **Univariate** : central tendencies, distributions and spread of data
- **Bivariate** : relationship b/w variables, similarity/hypothesis testing

Role of Statistics in EDA: Bivariate

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Role of Statistics in EDA: Bivariate

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

All Male Data

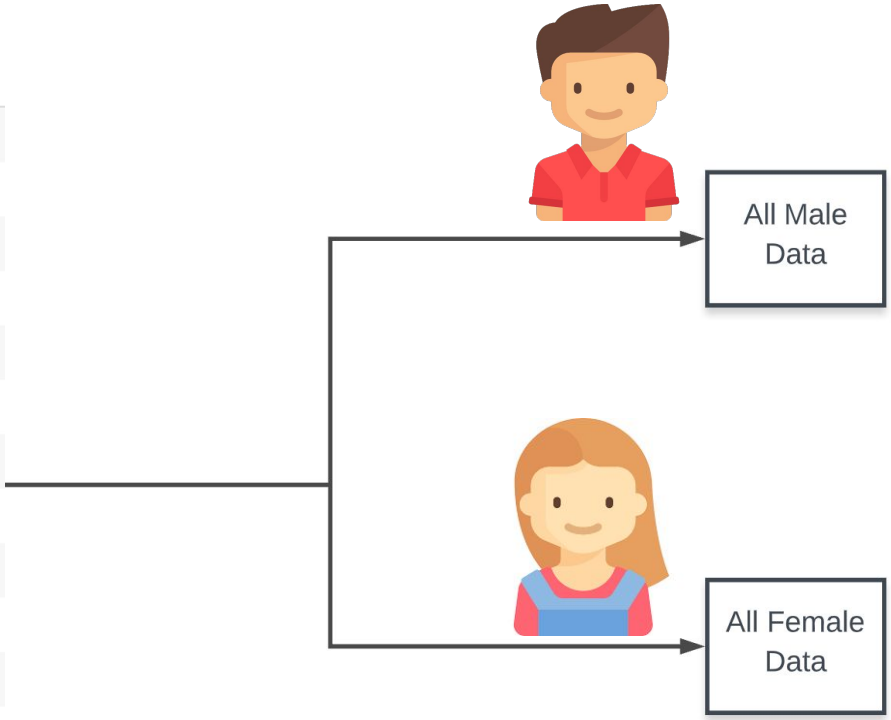
All Female Data

Analytics Vidhya
Learn everything about analytics

Role of Statistics in EDA: Bivariate

- Are these groups different?

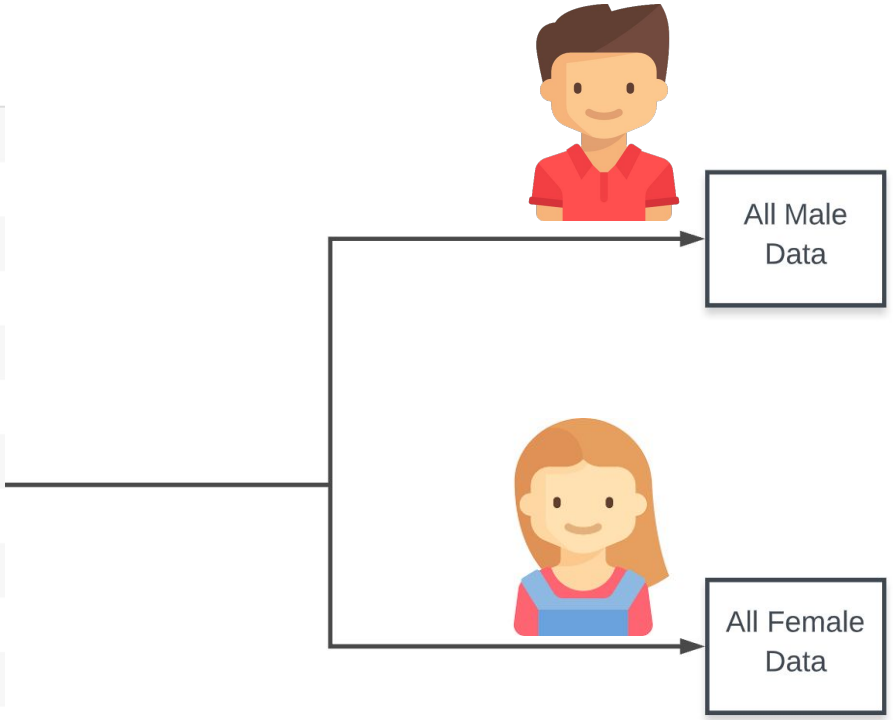
Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51



Role of Statistics in EDA: Bivariate

- Are these groups different?
- Is the difference significant enough?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51



Role of Statistics in EDA: Experiments

Statistics is set of rules and concepts for analysis and interpretation of the data

- **Univariate** : central tendencies, distributions and spread of data
- **Bivariate** : relationship b/w variables, similarity/hypothesis testing
- **Experiments** : T-tests (one-sample, paired, two-sample)

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Before Experiment

Male	72
Male	46
Male	31
Male	40
Male	68
Male	32

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Before Experiment

Male	72
Male	46
Male	31
Male	40
Male	68
Male	32

After Experiment

Male	38
Male	47
Male	75
Male	39
Male	80
Male	57

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Before Experiment

Male	72
Male	46
Male	31
Male	40
Male	68
Male	32

?

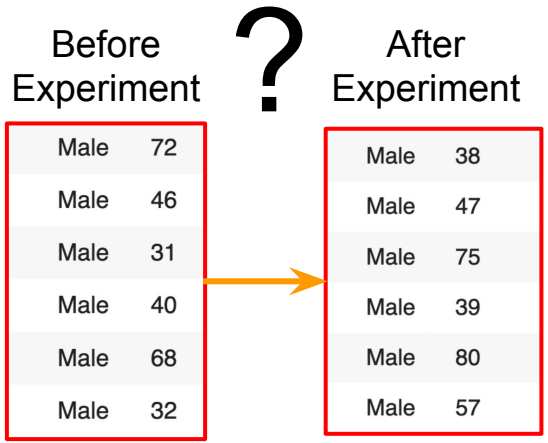
After Experiment

Male	38
Male	47
Male	75
Male	39
Male	80
Male	57

Analytics Vidhya
Learn everything about analytics

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51



Analytics Vidhya
Learn everything about analytics

Why is Statistics Important?

Male	66
Male	35
Male	31
Female	42
Male	72
Male	46
Male	31
Male	40
Male	68
Male	32
Male	73
Male	50
Female	48
Female	51

Before Experiment

Male	72
Male	46
Male	31
Male	40
Male	68
Male	32

?

After Experiment

Male	38
Male	47
Male	75
Male	39
Male	80
Male	57

Role of Statistics in EDA: Population Estimation

Statistics is set of rules and concepts for analysis and interpretation of the data

- **Univariate** : central tendencies, distributions and spread of data
- **Bivariate** : relationship b/w variables, similarity/hypothesis testing
- **Experiments** : T-tests (one-sample, paired, two-sample)
- **Population Estimation** : central limit theorem

Thank You!