Descriptive Statistics

In the previous video

- Why we cannot always use mode to summarize our continuous data?
- What is mean?
- How to calculate mean of the data?
- Mean is not robust

In this video

- What do we really mean by robust?
- What is an outlier?
- Reasons for having outliers in data

Analytics Vidhya
Learn everything about analytics

Can I now trust the mean value ?

Mean with 9700= 164.24

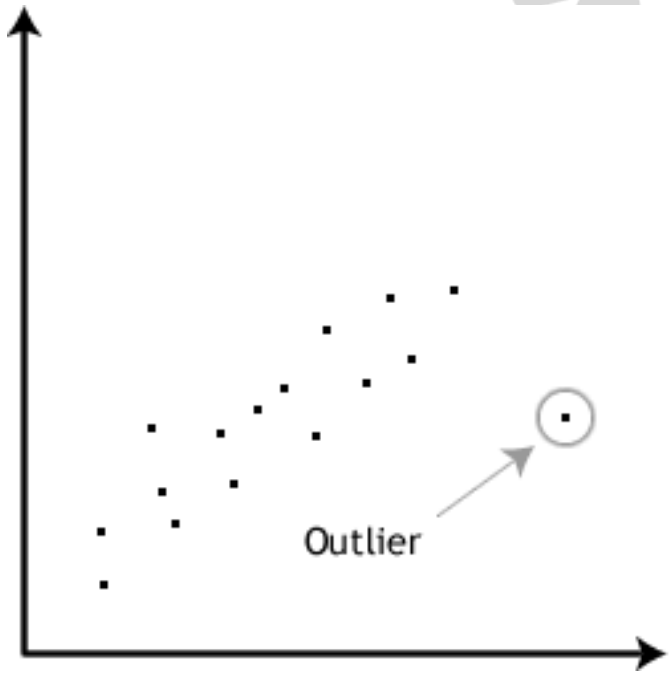
Mean with 97 = 68.21

Mode = 88

Student	Overall Marks	Gender
Anish	9700	Male
Rishabh	83	Male
Ajay	78	Male
Abhinav	86	Male
Anurag	91	Male
Ajeet	15	Male
Varun	83	Male
Rajeev	64	Male
sanjay	66	Male
NISHA	58	Female
SURESH	62	Male
Aniket	62	Male
manu	20	Male
ALOK	85	Male
Pranav	49	Male

Outliers

- Any values will fall really outside the range of the data is termed as a **outlier**

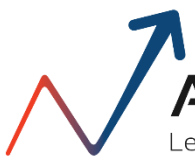


Student	Overall Marks	Gender
Anish	9700	Male
Rishabh	83	Male
Ajay	78	Male
Abhinav	86	Male
Anurag	91	Male
Ajeet	15	Male
Varun	83	Male
Rajeev	64	Male
sanjay	66	Male
NISHA	58	Female
SURESH	62	Male
Aniket	62	Male
manu	20	Male
ALOK	85	Male
Pranav	49	Male

Reasons for Outliers in the data

- **Typos** – Outliers during data collection. For example adding an extra zero by mistake
- **Measurement error** - outliers in data due to measurement operator being faulty. For example weights of people measured on a faulty weighing machine.
- **Intentional error** – these are errors which are induced by people intentionally. For example teens might claim lesser alcohol than they actually claim.
- **Legit outliers** – these are values which are not actually errors but are in the data due to legitimate reasons. For example a CEOs salary might actually be very high as compared to other employees.

Analytics Vidhya
Learn everything about analytics



Mean and Mode

Without Outlier	With Outlier
4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7	4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7,300
Mean = 5.45	Mean = 30.00
Mode = 5.00	Mode = 5.00