Mean can be ========= > statistic or parameter

If you identify mean on population data , then the mean is called parameter

If you identify mean on sample data, then the mean is called statistic

Mean

Median

Mode ========== all will deal about centre of the data

Flow of the data

Range

Mean deviation

Absolute mean deviation

Variance

Standard deviation

Percentile and quartile

Percentage is different from percentile

No percentage concept

10 11 12 13 14 15 16 17 18 19 20

How many data points are there 11 data points are there

Median = 15 ====== > 50 percentile

Out of the total data (11 observations) 50 percentage of the data has less than value 15

Half the data less than median value, that’s why median = 50 percentile

Suppose GATE exam ========= > 100max marks

Number applicants are 1000

You got 25marks out of 100

Your percentile is 70

Out of 1000 members 70 perentage , 700 pepole has marks less than 25

PASHa CAT ========= > max marks = 100

PASHA GOT 99 perentile ============ it is not 99 marks

His marks only 50

Out of the total applicants 99 percentage of students has marks less than 50

CAT/GMAT/GATE ======= percentile

Mechanical ======= top 30 /100

**$2,038 $1,758 $1,721 $1,637 $2,097 $2,047 $2,205 $1,787 $2,287 $1,940 $2,311 $2,054 $2,406 $1,471 $1,460**

Step-1: make all observations in an order

1. $1,460 6) 1787 11) 2097
2. $1,471 7) 1940 12) 2205
3. $1,637 8) 2038 13) 2287
4. $1,721 9) 2047 14) 2311
5. $1,758 10) 2057 15) 2406

Step-2: how many observations are there =15

Step-3: what is the median value

8th observation is my 50 percentile data

N=15 and P=50 ================ > 8th

Half of observations less than 50perecntile value

Half of observations less than 2038

1. $1,460 6) 1787 11) 2097
2. $1,471 7) 1940 12) 2205
3. $1,637 8) 2038 13) 2287
4. $1,721 9) 2047 14) 2311
5. $1,758 10) 2057 15) 2406

8th observation is my 50 percentile data : 2038

Step-4: I want 25perentile value : 4th observation is ==== > 1721 (15+1)\*25/100= 16/4=4th

Out of total observation(15), 25perecentage of data less than 1721

15\*25/100=15/4=3.75 <4th

Step-5: I want 75 percentile value : 12th observation====== > 2205 (15+1)\*75/100= 16\*3/4=12th

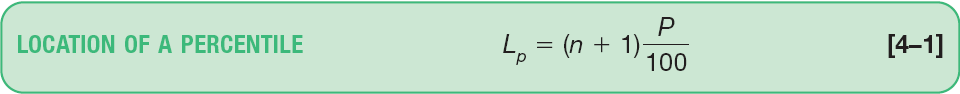
Out of total observation(15), 75perecentage of data less than 2205

15\*3/4=45/4=11.

25p ============ 50p ============ 75p

1721 2038 2205

N\*50/100



P=50 N= 15

(15+1)\*50/100 = 16\*50/100= 16/2= 8th location

Quartiles:

Quarter ============ > 25 percentile

Percent =========== > reference to 100

1p 2p 3p ======= 25p ====== 50p ==== 75p ====== 100p

100/25 ========= 4

Data divided into 4 parts

0 ========== 25 ============ 50 ========= 75 ==========100

Min\_value ========== Q1============== Q2=============Q3============max value

Min\_value========== 25p ============ 50p ============ 75p===========max value

Quartile 1 : Q1 ====== > 25p

Quartile 2: Q2 ======= > 50p

Quartile 3: Q3 ======= > 75p



**Outlier analysis**:

Outlier is an observation it skews your entire data

Outlier might be a huge max value or a huge min value

You want to identify below poverty line people

You want to identify middle class people

You want to identify rich people

Outliers will be available in rich people data

Outliers will be available in poor people data

Min value Q1 Q2 Q3 Max value

Where can be poor people data available ========= > min to Q1

Middle class ========= Q1 to Q3

Rich class ===== Q3 to max

Outlier cut off between min to q1

Q3 to max

Box plot:

Outlier > Q3

Outlier < Q1

IQR========= > inter quartile range

IQR= Q3-Q1

Outlier = Q3 + 1.5\* IQR

Outlier = Q1 - 1.5\* IQR

Mild outlier : 1.5\* IQR

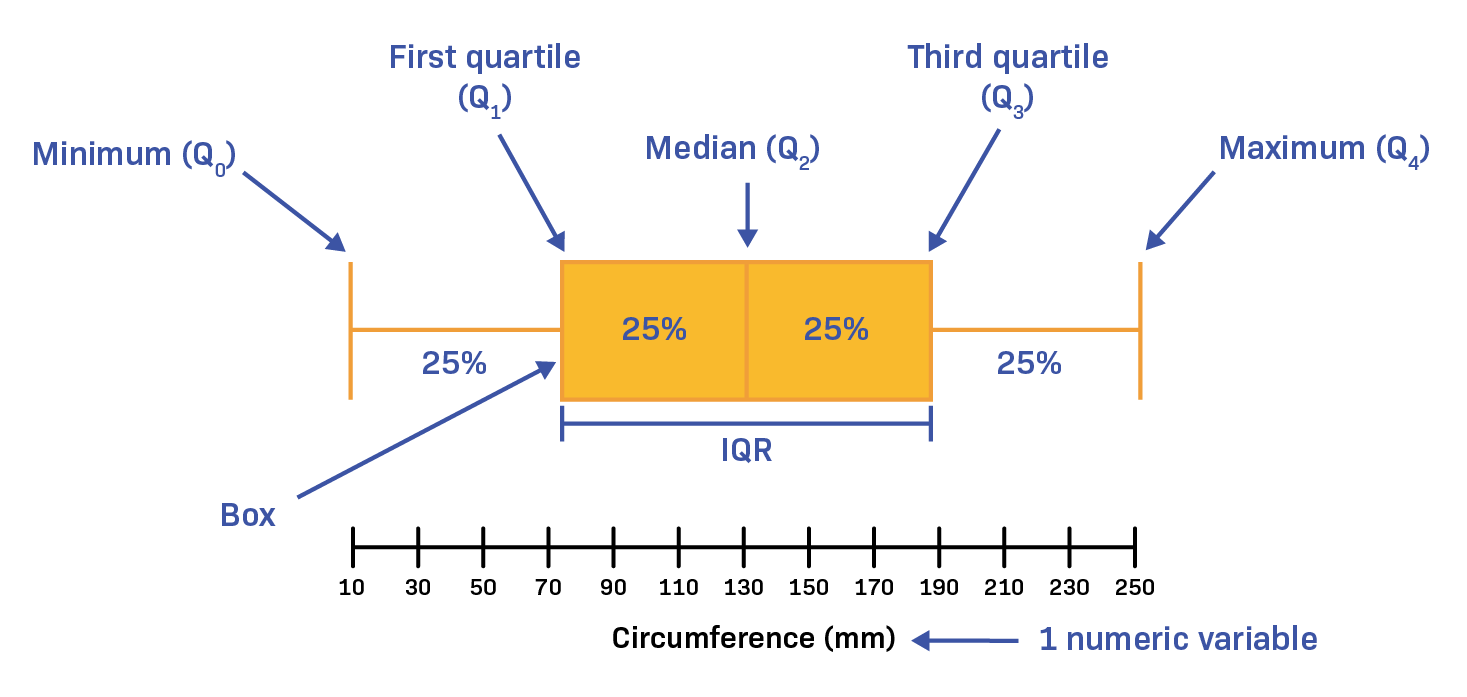
Huge outlier : 3\*IQR

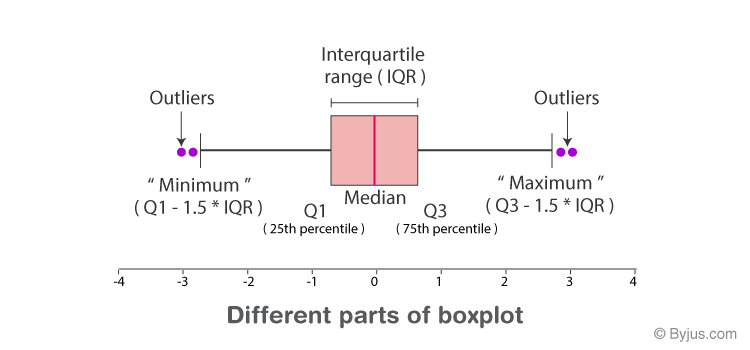
Outlier : Q3 + 1.5\*IQR or Q3+3\*IQR

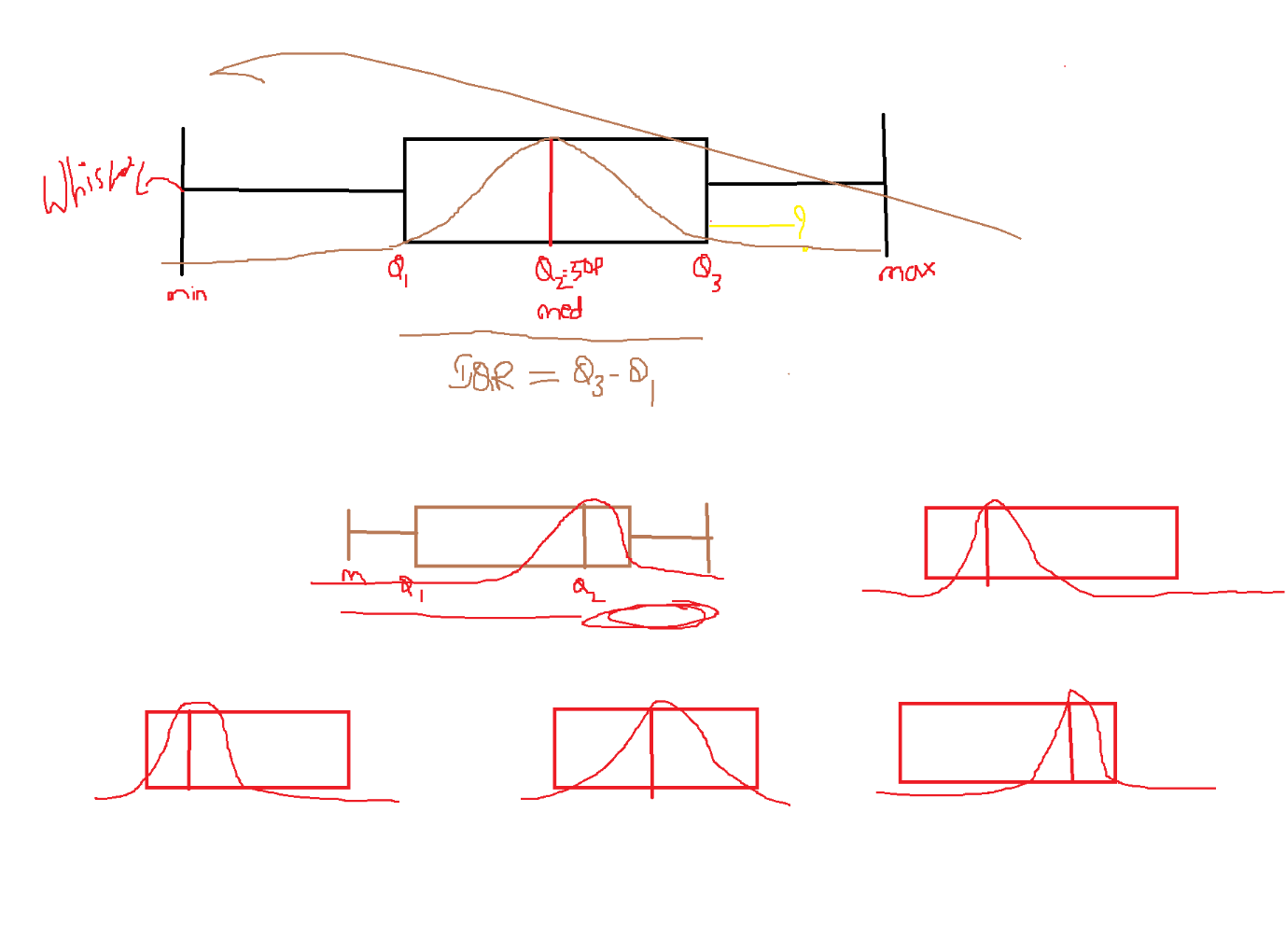
Q1-1.5\*IQR or Q1-3\*IQR

Hi sir, why only 1.5 and 3

Q3> ============= > Q3+ 3\*IQR







3 types of analysis

Age income gender

Age ======== numerical

Income ========== numerical

Gender =========== categorical

Variables /features/columns/

Univariate analysis : one

We are analyse only age column

Only income

Only gender

Bi variate analysis: two

Analyse age and income together

Multi variate analysis:

Analyse age --- gender --- income together

Univariate : bar/pie/histogram/boxplot

Bivariate : scatter plots

Age income

30 50k

35 55k

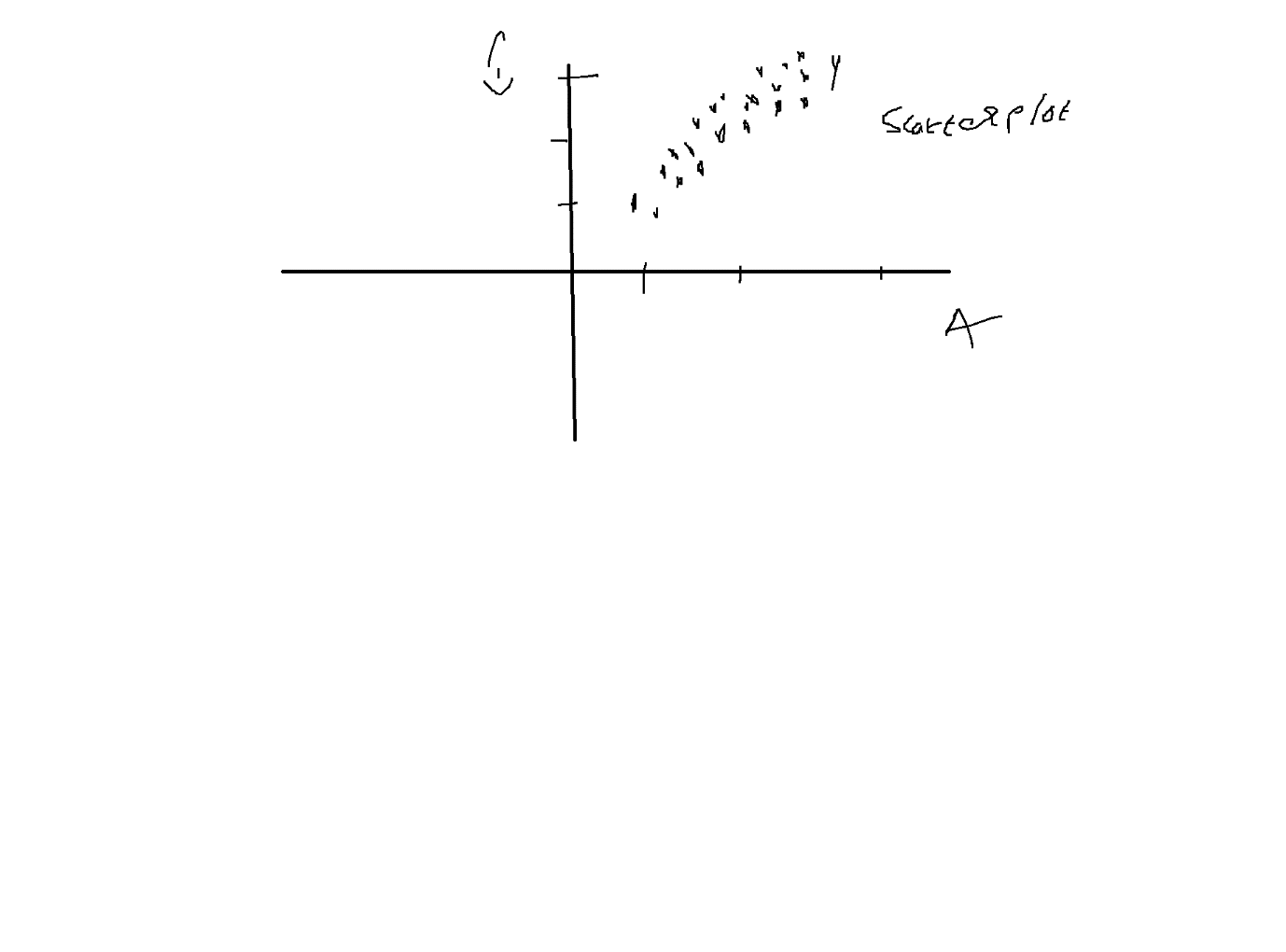
40 6ok

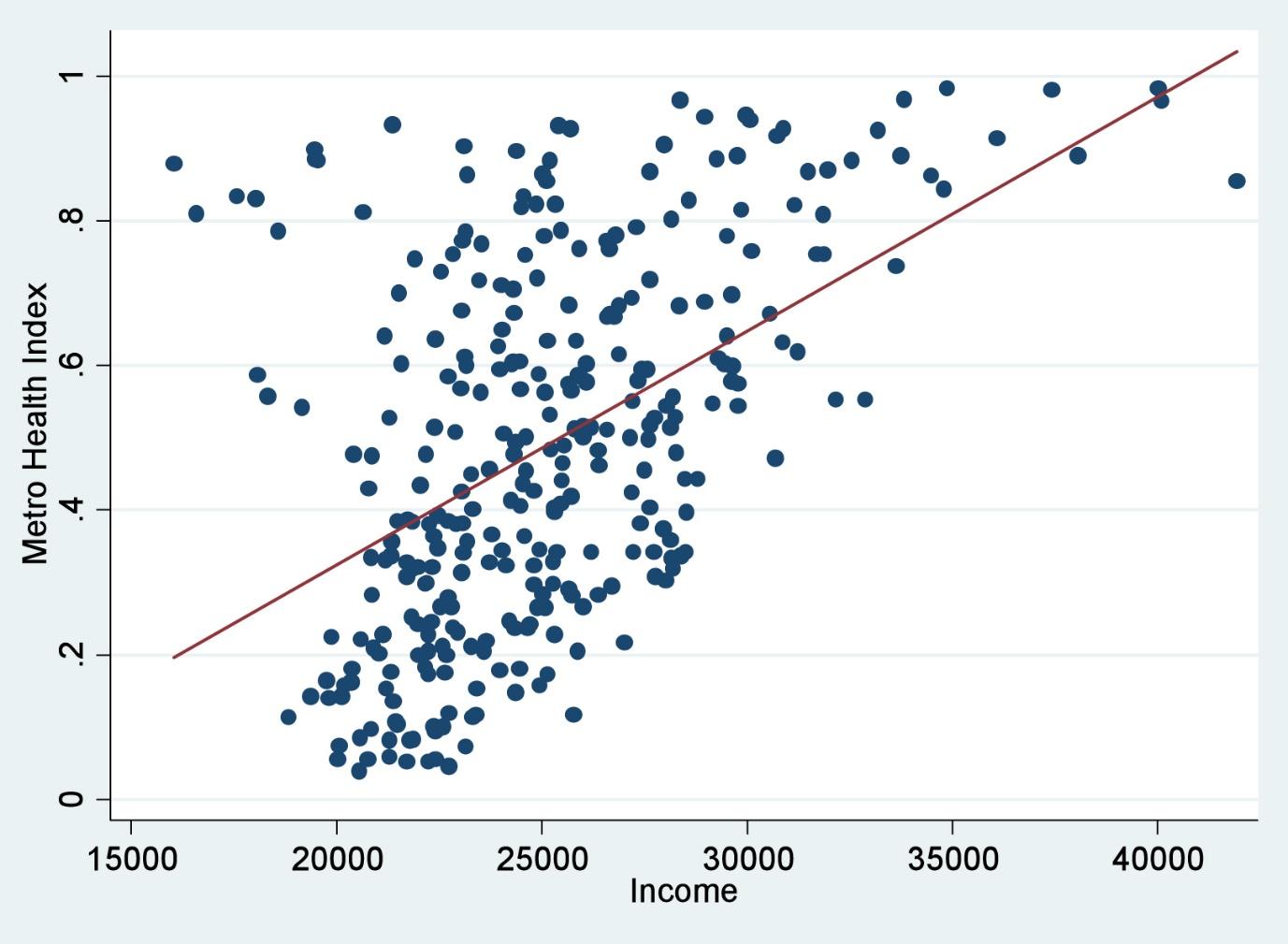
If you have a data has 100 columns ====100dimensions data

Age and gender 2 colmns ====== 2D data

(30,50k) (35,55k) (40,60k)

**Scatter plot:**





If age increases income also increasing ======== positively correlated

If age increase income is decreasing ========== negatively correlated

If age increase your income is constant ======= no correlation

Costart

Colleage

Co ============ two

Diff between univarite and bivariate

Univariate ===== uni + variate

Single variable analysis

4D data =======