Quantitative Techniques

Business Statistics for Decision Making

Dr. Anchal Gupta

Measures of Central Tendency

- One of the important objective of statistical analysis is to get one single value that describes the characteristic of the entire data i.e average.
- For ex- average student in a class, average height, average income etc.
- An average represents the entire data, its value lies somewhere in between two extremes that is , the largest and smallest items. So, average is also known as measures of central tendency.

Mean

- Perhaps the most important measure of location is the mean.
- The mean provides a measure of central location.
- The mean of a data set is the average of all the data values.
- The sample mean $\bar{x}$ is the point estimator of the population mean μ.

Objectives of Averaging

- To get single value that describes the characteristics of the entire group
- To facilitate comparison

Arithmetic Mean

- It's value is obtained by adding together all the items and by dividing the total by the number of items.
- Ungrouped data

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum X}{n}$$

Where n: no. of observations

Numerical for practice

- Davis furniture Company has a revolving credit agreement with the National Bank. The loan showed the following ending monthly balances per year:

- Jan $121,300
- Feb $112,300
- Mar $72,800
- Apr $72,800
- May $72,800
- Jun $57,300

Jul $58,700
Aug $61,100
Sept $50,400
Oct $52,800
Nov. $49,200
Dec. $46,100

- The company is eligible for a reduced rate of interest if its average monthly balance is over $65,000. Does it qualify?

Solution

- Mean=(121300+112300+72800+72800+72800+57300+58700+61100+50400+52800+49200)/12

- Mean= 68966.66 (>65,000)

- Therefore, the company is eligible for getting reduced rate of interest.

For grouped data

- Discrete-

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum f_i}$$

- Continuous-

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i m_i}{\sum f_i}$$

Numerical for discrete

- Calculate mean-

Marks	Frequency
20	8
30	12
40	20
50	10
60	6
70	4

Solution

X	f	fX
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	Sum=Σf= 60	Sum= ΣfX= 2460

Mean=$\Sigma fX/\Sigma f$ = 41

Numerical for Continuous

- Calculate mean-

Marks	Frequency
0-10	5
10-20	10
20-30	25
30-40	30
40-50	20
50-60	10

Solution

X	m	f	fm
0-10	5	5	25
10-20	15	10	150
20-30	25	25	625
30-40	35	30	1050
40-50	45	20	900
50-60	55	10	550
		Sum=Σf=100	Sum=Σfm=3300

Mean=$\Sigma fm/\Sigma f$ = 3300/100=33

Merits/Demerits

Merits

-It is based on all observations.

- It is useful for further algebraic calculations.

-It is simplest and easiest to compute.

Demerits

-Extreme values affect the value of average.

(60+70+10+80)/4 = 55

-It is unable to compute mean for open ended classes.

- difficult to compute when large no. of observations.

Median

- Central / middle value that divides distribution into two equal parts.
- Middle Value In Ordered Sequence
  - If Odd n, Middle Value of Sequence
  - If Even n, Average of 2 Middle Values
- Position of Median in Sequence

$$\text{Positioning Point} = \frac{n+1}{2}$$

Median Example
Odd-Sized Sample

Raw Data: 24.1 22.6 21.5 23.7 22.6

Ordered: 21.5 22.6 22.6 23.7 24.1

Position: 1 2 3 4 5

$$\text{Positioning Point} = \frac{n+1}{2} = \frac{5+1}{2} = 3$$

Median = 22.6

Median Example
Even-Sized Sample

Raw Data: 10.3 4.9 8.9 11.7 6.3 7.7

Ordered: 4.9 6.3 **7.7** **8.9** 10.3 11.7

Position: 1 2 **3** **4** 5 6

$$\text{Positioning Point} = \frac{n+1}{2} = \frac{6+1}{2} = 3.5$$

$$\text{Median} = \frac{7.7 + 8.9}{2} = 8.3$$

Median for grouped data

- $$\text{Median} = L + \frac{N/2 - c.f.}{f} \times i$$

L = lower limit of the median class

c.f. = Cumulative frequency of the class preceding the median class.

f = simple frequency of the median class.

i = The class interval of the median class.

Example

Wages	No. of workers
2000-3000	3
3000-4000	5
4000-5000	20
5000-6000	10
6000-7000	5

Solution

Wages	frequency	Cumulative frequency
2000-3000	3	3
3000-4000	5	8
4000-5000	20	28
5000-6000	10	38
6000-7000	5	43

N/2= 43/2=21.5

Sum=N=Σf=43

L=4000; cf=8; f=20; i=1000

Median=4000+ (21.5-8)/20*1000= 4675

Merits/Demerits

Merits

-It is useful for open ended classes also.

It is not affected by extreme values.

-It is most appropriate average in dealing with qualitative data.

Demerits

-For this, we need to arrange the data other averages do not need it.

-It is not based on all the observations.

Mode

- In french means fashion
-
- 'most fashionable' value
- - customer's preference
- - garment industry
- Most common wage, most common income, most common size of shoe or garment
- Category (or value) with the highest frequency

Mode Calculation

- Value which occurs most frequently in a distribution
- E.g.
- Size of shoes No. of persons

Size of shoes	No. of persons
5	10
6	20
7	25
8	40
9	22
10	15
11	6

What is the modal size? - 8

Which is the Mode?

- Distribution A:

  7 8 4 2 2 1 9 8 2 7 6

  Distribution B:

  8 0 2 3 6 7 0 8 1 4 7

  Distribution C

  3 5 7 8 1 9 2 4 11

Which is the Mode?

- When there are two or more values having same maximum frequency (mode is ill defined). This series is known as bimodal or multimodal.
- Income (in Rs.)

  110 120 130 120 110 140 130 120 130 140

Size of item	No. of times it occurs
110	2
130	3
130	3
140	2

Since 120 & 130 have maximum frequency- mode is bimodal

MODE(Grouped data)

- $\text{Mode} = L + \dfrac{f1 - f0}{2f1 - f0 - f2}$

L= lower limit of modal class
f1=frequency of modal class
f0=frequency of class preceding the modal class
f2=frequency of class succeeding the modal class

- Relationship between Mean-Median-Mode
- Mode = 3 Median – 2 Mean

Example

Marks	Frequency	
30-40	4	
40-50	6	
50-60	8	f0
60-70	12	f1
70-80	9	f2
80-90	7	
90-100	4	

Solution

L=60, f1=12, f0=8, f2=9

Mode= 60 + [(12-8)/(2*12-8-9)]*10

=60+(4/7)*10

= 65.7 approx

Merits/Demerits

Merits

-It is easy to calculate.

It is not affected by extreme values.

-

Demerits

-Sometimes it is ill defined. Like cases of no mode, bimodal and multimodal.

-It is not based on all the observations.

Which Measure Do I Use?

- Which measure of central tendency is most appropriate?
  - In general, the mean is preferred, since it has nice mathematical properties (in particular, see chapter 7)
  - The median and quartiles, are <u>resistant</u> to outliers
- Consider the following three datasets
  - 1, 2, 3 (median=2, mean=2)
  - 1, 2, 6 (median=2, mean=3)
  - 1, 2, 30 (median=2, mean=11)
  - All have median=2, but the mean is <u>sensitive</u> to the outliers
- In general, if there are outliers, the median is preferred to the mean

Percentiles

- Percentile - measures of central tendency that divide a group of data into 100 parts
- At least n% of the data lie at or below the n^{th} percentile, and at most (100 - n)% of the data lie above the n^{th} percentile
- Example: 90th percentile indicates that at 90% of the data are equal to or less than it, and 10% of the data lie above it

Calculating Percentiles

- To calculate the p^{th} percentile,
  - Order the data
  - Calculate $i = N (p/100)$
  - Determine the percentile
    - If i is a whole number, then use the average of the i^{th} and $(i+1)^{th}$ ordered observation
    - Otherwise, round i up to the next highest whole number

80^{th} Percentile

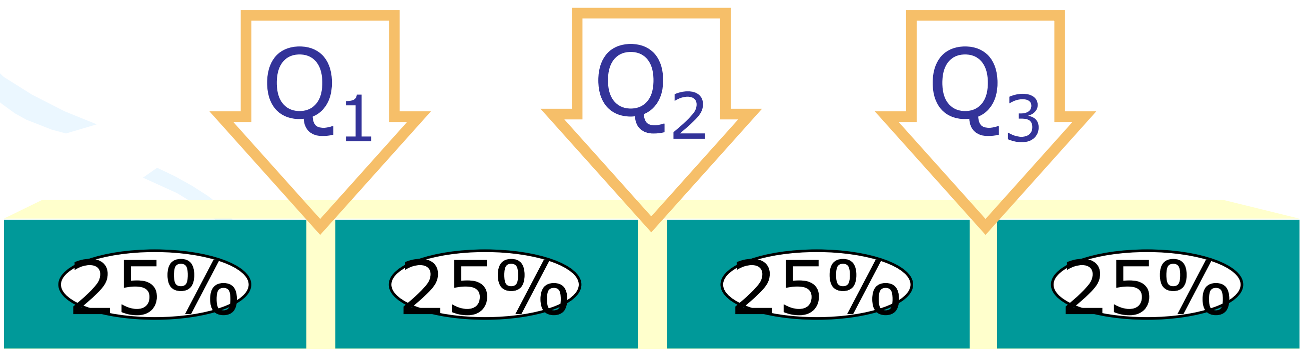
- Example: Apartment Rents
  - $i = (p/100)n = (80/100)70 = 56$
  - Averaging the 56^{th} and 57^{th} data values:
  - 80th Percentile = (535 + 549)/2 = 542

425	430	430	435	435	435	435	435	440	440
440	440	440	445	445	445	445	445	450	450
450	450	450	450	450	460	460	460	465	465
465	470	470	472	475	475	475	480	480	480
480	485	490	490	490	500	500	500	500	510
510	515	525	525	525	535	549	550	570	570
575	575	580	590	600	600	600	600	615	615

Note: Data is in ascending order.

Quartiles

- Quartile - measures of central tendency that divide a group of data into four subgroups
- Q1: 25% of the data set is below the first quartile
- Q2: 50% of the data set is below the second quartile
- Q3: 75% of the data set is below the third quartile



Quartiles

- Quartiles are specific percentiles.
- First Quartile = 25th Percentile

Second Quartile = 50th Percentile = Median

- Third Quartile = 75th Percentile

Third Quartile

- Example: Apartment Rents
  - Third quartile = 75th percentile
  - $i = (p/100)n = (75/100)70 = 52.5 = 53$

Third quartile = 525

425	430	430	435	435	435	435	435	440	440
440	440	440	445	445	445	445	445	450	450
450	450	450	450	450	460	460	460	465	465
465	470	470	472	475	475	475	480	480	480
480	485	490	490	490	500	500	500	500	510
510	515	525	525	525	535	549	550	570	570
575	575	580	590	600	600	600	600	615	615

Note: Data is in ascending order.

More Examples for Practice

Airthmetic Mean (Ungrouped Data)

Find the mean of the following observations:

32,35,36,37,39,41.43,47,48

Soln:

Mean=(32+35+36+37+39+41+43+47+48)/9

= 358/9 = 39.77

More Examples for Practice

Arithmetic Mean (Grouped Data-Discrete)

Find the mean of the following data:

No. of students	5	9	13	21	20	15	8	3

Solution

Marks (x)	No. of students (f)	fx
5	5	25
10	9	90
15	13	195
20	21	420
25	20	500
30	15	450
35	8	280
	$\Sigma f = 94$	$\Sigma fx = 2{,}080$

Mean= $\Sigma fx / \Sigma f = 2080/94 = 22.127$

More Examples for Practice

Arithmetic Mean (Grouped Data-Continuous)

Class	Frequency
20-25	10
25-30	12
30-35	8
35-40	20
40-45	11
45-50	4
50-55	5

Solution

Class	Midpoint (m)	Frequency (f)	fm
20-25	22.5	10	225
25-30	27.5	12	330
30-35	32.5	8	260
35-40	37.5	20	750
40-45	42.5	11	467.5
45-50	47.5	4	190
		$\Sigma f = 70$	$\Sigma fm = 2485$

Mean= $\Sigma fx / \Sigma f = 2485/70 = 35.5$

More Examples for Practice

Median (Grouped)

Marks	No. of students
Below 10	8
10-20	10
20-40	22
40-60	25
60-80	10
Above 80	5

Solution

Marks (x)	No. of students (f)	Cumulative Frequency
Below 10	8	8
10-20	10	18
20-40	22	40
40-60	25	65
60-80	10	75
Above 80	5	80
	N= Σf = 80	

N/2= 80/2= 40

L=20; cf=18; f=22; i=40-20=20

Median=20+ (40-18)/22*20= 40

More Examples for Practice

Mode (Grouped)

Class Interval	Frequency
10-20	10
20-30	12
30-40	18 f0
40-50	30 f1
50-60	16 f2
60-70	6
70-80	8

Solution

L=40, f1=30, f0=18, f2=16

Mode= 40 + [(30-18)/(2*30-18-16)]*10

=40+(12/26)*10

= 44.61 approx