

22 August 2023

Tuesday

3rd qtr 14th

Mam Kannan

Quiz Weeks - 3

$$\text{Assignments} = 8\% = 8\%$$

$$\text{Quizzes} = 4 \times 3\% = 12$$

$$\text{Mids} = 15 + 15\% = 30\%$$

Statistics :-

Technique and methods of data collection, analysis and (integration, interpretation and manipulation)

Sample and Population

Sampling :- partially covering a population by taking non-probability samples that is random.

→ Results conducted from population is the parameter

→ The samples produce quantities that are called statistic.

Sample mean : \bar{x} (x -bar)

Population mean : μ (mean)

$$\bar{x} = \frac{\Sigma x}{n} = 3 \quad \left. \begin{array}{l} \text{Sample mean} \\ \text{diff.} \end{array} \right\} \text{er}$$

$$\mu = \frac{\Sigma x}{N} = 3.5 \quad \left. \begin{array}{l} \text{Population mean} \\ \text{er} \end{array} \right\}$$

Branches of Statistics :-

- 1) Descriptive Statistics
- 2) Inferential Statistics

Descriptive Statistics :-

Summary :-

- 1) mean (central)
- 2) Dispersion
- 3) graphical visualization
- 4) classification

Inferential Statistics

To draw inferences from data-

1) Hypothesis testing

2) Parameter estimation

Hypothesis

Variables

1) Quantitative (Quantitative)

2) Non-Happy Stage (Qualitative)

1) Qualitative

Variables

1) Qualitative }

2) Quantitative } discrete + continuous

Data Types

→ Primary & Secondary.

↓

generated

data manually

(Raw data)

22/08/23

Population :-

A collection of set of individuals, objects or events whose properties are to be analyzed

Sample :-

A subset of population

Statistics

descriptive (used to
statistics summarize data
graphically)

Inferential
statistics
(measure
variability)

Measure of
variability

22/08/23

Population :- Set of individuals, objects or events as a collection whose properties is to be analyzed

Sample :-

A subset of population. (Statistics)

Types of Statistics

Descriptive Stats (summarize data) Inferential stats

Measure of Tendency

- a) mean
- b) median
- c) mode

Measure of variability

- a) Range
- b) variance
- c) Dispersion

Qualitative

Quantitative

discrete

continuous

Data types → primary data (raw data)

↓
secondary data (already built/collected/refined)

24/08/23

Exercise 1.1

3.0, 4.4, 4.0, 5.2, 3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7,

a)

Sample size = 15

b)

$$\frac{56.8}{15} = 3.786$$

15

$$\text{mean} = 3.8$$

c) median

Sort the samples:-

2.5, 2.8, 2.8, 2.9, 3.0, 3.3, 3.4, 3.6, 3.7,
4.0, 4.4, 4.8, 5.2, 5.6, 4.8, 5.2, 5.6

$$\frac{15}{2} = 7.5 \rightarrow 8^{\text{th}} \text{ term} = 3.6$$

(d) time ↑

6

5

4

3

2

1 2 3 4 5 6 7 8 9 10 → brand

1.01, 1.02, 1.03, 1.04

c) 20). trimmed mean

2.9, 3.0, 3.3, 3.4, 3.6, 3.7, 4.0, 4.4, 4.8

$$\text{mean} = \frac{33.4}{9} = 3.67 = 3.7$$

(f) 0.1. It is almost the same as descriptive as sample mean.

Exercise 1.28-

$$\text{mean} = \frac{21.86}{20} =$$

$$\text{mean} = 20.76$$

$$\star 50\% = \text{median} = 20.6$$

18.71, 21.41, 20.72, 21.81, 19.29, 22.43, 20.17,
23.71, 19.44, 20.50, 18.92, 20.33, 23.00,
22.85, 29.28, 21.77, 19.77, 18.04, 21.02

Sort :- 18.04, 18.71, 18.92, 19.25, 19.29,
19.44, 19.77, 20.17, 20.33, 20.55,
20.72, 21.02, 21.41, 21.77, 21.81,
22.01, 22.43, 22.85, 23, 23.71

$$\text{mean} = 415.36 / 20 = 20.768 = 20.77$$

$$\star \text{median} = 10^{\text{th}} \text{ value} = 20.55 = 20.6$$

$$\text{trimmed 10% mean} = 19.496 = 19.50$$

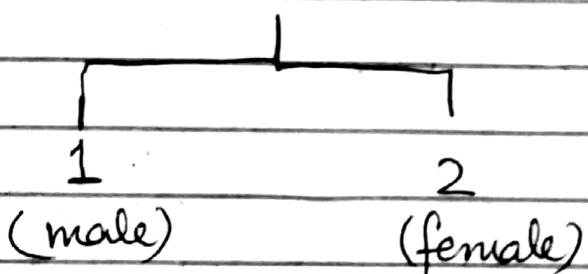
24/08/23

Nominal Scale } Qualitative
Ordinal Scale } Data

Interval Scale } Quantitative Data
Ratio Scale

Nominal Scale :-

Gender



- Nominate a number to each gender without preference
- No order of grouping

Ordinal Scale :-

- greater than or less than
- ordering of groups e.g.

Rating in Uber :-

1 2 3 4 5
verybad bad fine good v. good

Player ratings

Interval Scale :-

→ value of zero (0) is not considered as truly absolute or nothing

→ pH, temperature (0°C , $\text{Opt pH} \Rightarrow \text{acidic}$)

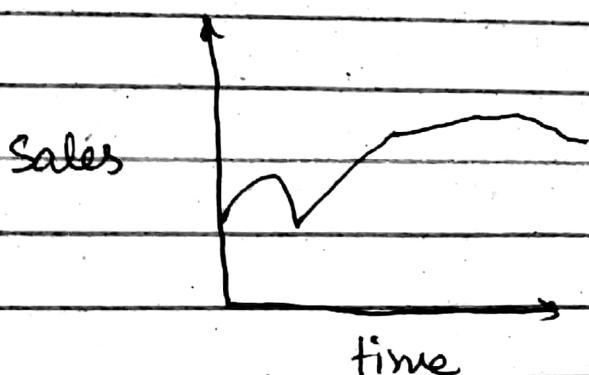
Ratio Scale :-

→ Zero implies nothing - null.

→ distance, GPA, student's age.

Time Series :-

→ data associated w time



Cross Sectional :-

fixed time

Ungrouped data :-

→ Measure of central tendency

→ mean, median, mode

mean

$$\bar{x} = \frac{\sum x}{n}$$

mode

$$\mu = \frac{\sum x}{N}$$

sample size

population size

→ If data has outliers, this gives error

Median :- → divides data into equal halves
→ if $\frac{n}{2}$ is an integer then

$$\text{median} = \frac{1}{2} \left[\left(\frac{n}{2} \right)^{\text{th}} \text{ value} + \left(\frac{n+1}{2} \right)^{\text{th}} \text{ value} \right]$$

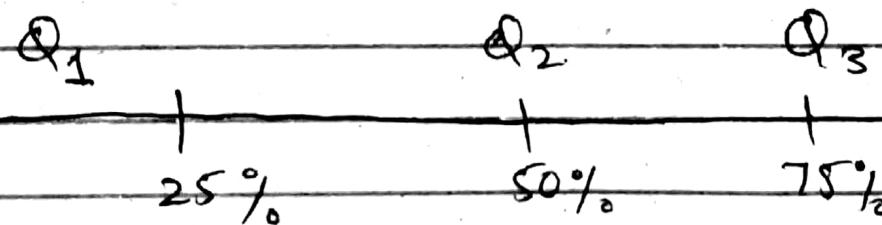
→ If $\frac{n}{2}$ is not an integer (odd case)

$$\text{median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ value}$$

29 / 08 / 23

⇒ Percentile and Quartile

- There are 4 parts of data - therefore Quartile. There are 3 values of Quartile
- There are 100 parts of data therefore percentile has 99 values.



50% percentile is Q_2 (mode)
25% percentile is Q_1
75% percentile is Q_3

$Q_1 \rightarrow$ lower Quartile
 $Q_2 \rightarrow$ upper Quartile

Steps to calculate percentile / Quantile

- (1) Sort the data in ascending order
- (2) calculate index

$$i = \left(\frac{P}{100} \right) n \quad (\text{Position of each sample})$$

- (3) P can range from 1 to 99 in index

(3)a) If i is an integer then p th percentile is average of the positions i and $i+1$.

b) If i is not an integer (odd case) then we pick the value which is present at the rounded value of i

Example

Monthly starting salaries for a sample of 12 business school graduates :-

graduate	salary (\$)	graduate
1	3450	
2	3550	
3	3650	
4	3480	
5	3355	
6	3310	
7	3490	
8	3730	
9	3540	
10	3425	

graduate	Salary (\$)
1	3450 ✓
2	3556 ✓
3	3650 ✓
4	3480 ✓
5	3355 ✓
6	3310 ✓
7	3490 ✓
8	3730 ✓
9	3540 ✓
10	3925
11	3520 ✓
12	3480 ✓

85th percentile = 3650

50th percentile = 3490

Upper and lower quartile.



3556



3480

3550

Saydawng, Wallpole, Anderson

3310, 3355, 3450, 3480, 3480
3490, 3520, 3540, 3550,
3650, 3730, 3925

8's

P_{85} 85th Percentile = 3660 3730

P_{50} 50th Percentile = 3480 3505

P_{25} 25th (lower) = 3465

P_{75} 75th (upper) = 3600

Q_1 = lower, Q_3 = upper

$P_{50} = \frac{1}{2} (6^{\text{th}} \text{ position value} + 7^{\text{th}} \text{ value})$

$Q_1 = P_{25} = \frac{1}{2} (3^{\text{rd}} \text{ value} + 4^{\text{th}} \text{ value})$