

Math 401 Notes

Lecture 1&2 (Statistics)

Ahmed Nasser

May 16, 2021

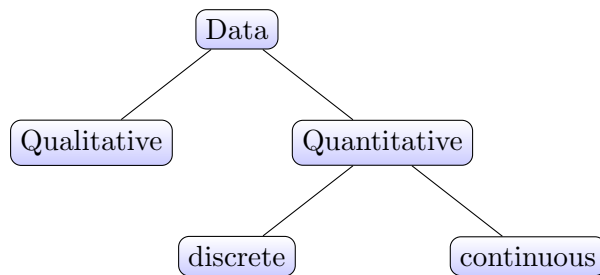
Contents

1	Statistics Steps	2
2	Types of Data	2
3	Types of statistical studies	2
4	Present (Graphs)	2
4.1	Histogram	2
4.2	Ogive	2
5	Range & Width	2
6	grouped frequency distribution	3
7	The Median	3
7.1	Discrete Data	3
7.2	Grouped data	3
8	The Mode	3
8.1	Discrete Data	3
8.2	Grouped data	3
9	The Mean	3
9.1	Discrete Data	3
9.2	Grouped data	4
9.3	Weighted data	4
10	The Standard Deviation	4
11	The Population Variance	4
11.1	Discrete Data	4
11.2	Grouped data	4
12	Chebyshev's Theorem	4
13	Percentiles	5
14	Quartiles	5

1 Statistics Steps

1. Collect
2. Organize
3. Present
4. Summarize
5. Draw Conclusion

2 Types of Data



3 Types of statistical studies

1. Descriptive statistics

collecting, organizing, presenting, and summarizing of data, in an informative way.

2. Inferential statistics

drawing conclusions and making decisions using the given data

4 Present (Graphs)

4.1 Histogram

- X-axis – Class Limits
- Y-axis – “ f ” or “ $r.f.$ ”

4.2 Ogive

- X-axis – Upper limits of each class limit
- Y-axis – “ $c.f.$ ” or “ $c.r.f.$ ”

5 Range & Width

1. Range = Max Val - Min Val

2. Class Width = $\frac{\text{Range} + 1}{\text{No. of classes}}$

If this value is not an integer ROUND IT UP

6 grouped frequency distribution

- If the **number of classes is given** and the given data values are integers: we proceed using the **direct method**
- If the **number of classes is NOT given** and the given data values are integers or decimals

If an **initial value is NOT given**, we proceed using the **Stem-leaf method**.

If an **initial value is given**, we proceed using the (\sqrt{n}) **method**. (Width = $\frac{\text{Range}}{\text{No. of classes}}$)

7 The Median

7.1 Discrete Data

- Sort the Data value
- MD = middle point (in case of odd num)
- MD = (sum of the two middle points) / 2 (in case of odd num)

7.2 Grouped data

$$\text{MD} = \frac{\frac{n}{2} - \text{pre."c.f."}}{f}(W) + L_m$$

where W is the class width and L_m is the lower bound

ex: if the median is 80 then about 50% of student achieved more than 80 and 50% achieved less than 80

8 The Mode

8.1 Discrete Data

- The mode is a data value that has the highest frequency in a data set
- If all of data values have the same frequency then there is no mode

8.2 Grouped data

- We use Modal class instead of mode in grouped data
- The modal class is the class corresponds to the highest frequency

9 The Mean

9.1 Discrete Data

- For a population

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

- For a Sample

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

9.2 Grouped data

$$\mu = \frac{\sum f \cdot X_m}{\sum f} = \frac{\sum f \cdot X_m}{N}$$

Where X_m is the class midpoint

9.3 Weighted data

$$\mu = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i}$$

10 The Standard Deviation

- For a population

$$\sigma := \sqrt{\sigma^2}$$

- For a Sample

$$S := \sqrt{S^2}$$

11 The Population Variance

11.1 Discrete Data

- For a population

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

- For a Sample

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$
$$S^2 = \frac{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}{n - 1}$$

11.2 Grouped data

- For a population

$$\sigma^2 = \frac{\sum f \cdot (X_m - \mu)^2}{\sum f}$$

- For a Sample

$$S^2 = \frac{\sum f \cdot X_m^2 - \frac{(\sum f \cdot X_m)^2}{\sum f}}{(\sum f) - 1}$$

12 Chebyshev's Theorem

$[\bar{x} - ks, \bar{x} + ks]$ is the smallest interval that contains at least $(1 - \frac{1}{k^2}) * 100\%$ of the data

13 Percentiles

- Sort your data values

$$\text{Percentile} = \frac{0.5 + \text{No. of values below } X}{n} * 100\%$$

- n is number of data and p is the percentile then evaluate $m = n.p$
 - if m is not integer round it up and then p corresponds **data value**[m]
 - if m is integer then p corresponds the average of **data value**[m] and **data value**[$m+1$]

14 Quartiles

- $Q_1 = 25^{\text{th}}$ percentile
- $Q_2 = 50^{\text{th}}$ percentile —or— median
- $Q_3 = 75^{\text{th}}$ percentile