Lecture 2

Descriptive Statistics

- Descriptive statistics consists of methods for **organizing, displaying**, and **describing data** by using **tables, graphs**, and **summary measures**.

Organizing, Displaying and Describing Data using Table and Graph

Raw Data

- Data **recorded in the sequence** in which **they are collected** and before they are processed or ranked are called *raw data*.

Qualitative Raw Data

Table 1: Types of smartphones owned by 40 persons

Types of smartphones owned									
I	A	I	A	W	I	A	A	I	W
W	A	I	A	I	F	W	I	A	F
I	A	F	I	I	W	I	A	I	A
A	I	W	F	W	F	I	W	A	I

I = Iphone A = Android W = Window Phone F = Amazon's Fire Phone

Qualitative raw data

Frequency Distribution

how often something occurs

Example : DATA ON THE PREFERENCE OF SHOE BRAND (Sample Size = 50)

Nike	Adidas	Nike	Woodland	Nike	Reebok	Nike
Reebok	Nike	Adidas	Nike	Bata	Adidas	Adidas
Woodland	Adidas	Reebok	Adidas	Reebok	Nike	
Adidas	Nike	Bata	Nike	Nike	Adidas	
Reebok	Woodland	Nike	Woodland	Adidas	Reebok	
Nike	Adidas	Reebok	Nike	Bata	Nike	
Adidas	Nike	Adidas	Woodland	Bata	Adidas	
Nike	Bata	Nike	Nike	Nike	Reebok	

EduSpread|Shubham Kalra

Organizing and Graphing for Qualitative Data

Frequency Distribution for Qualitative Data

A *frequency distribution* for qualitative data

- ✓ lists all the categories of the variable and
- ✓ count the number of elements that belong to each of the categories.

Table

List of the categories of variable	Number of times each category of the variable occurs

Example: Frequency distribution of **Table 1.**

Categories	Tally Marks	Frequency
I		15
A		12
W		8
F		5
Total		40

Example: Frequency distribution of **Table 2.**

Categories	Tally Marks	Frequency
Nike		19
Adidas		13
Woodland		6
Reebok		7
Bata		5
Total		50

Example: A sample of 30 employees from large companies was selected, and these employees were asked how stressful their jobs were. The responses of these employees are recorded below, where *very* represents very stressful, *somewhat* means somewhat stressful, and *none* stands for not stressful at all.

somewhat	none	somewhat	very	very	none

very	somewhat	somewhat	very	somewhat	Somewhat
very	somewhat	none	very	none	Somewhat
somewhat	very	somewhat	somewhat	very	none
somewhat	very	very	somewhat	none	somewhat

Construct a frequency distribution table for these data.

Solution:
Table 3: Frequency distribution of stress on job

Stress on job	Tally	Frequency (f)
Very	/𝍸 𝍸/	10
Somewhat		14
None		6
Total		30

Relative Frequency and Percentage Distribution

- The **relative frequency** of a category is obtained by dividing the frequency of that category by the sum of all frequencies.
- Thus, the relative frequency shows what fractional part or proportion of the total frequency belongs to the corresponding category.

Calculating Relative Frequency
Relative Frequency = $(Frequency\ of\ that\ category)/(Sum\ of\ all\ frequecies)$

Percentage
- The **percentage** for a category is obtained by multiplying the relative frequency of that category by 100.

Calculating Percentage

$$Percentage = Relative\ frequency \times 100$$

EXAMPLE

Table 4 Relative Frequency and Percentage Distributions of Stress on Job

Stress on job	Frequency (f)	Relative Frequency	Percentage
Very	10	10/30=.333	.333(100)=33.3
Somewhat	14	14/30=.467	.467(100)=46.7
None	6	6/30=.200	.200(100)=20.0
Total	30	1.00	100

- Based on the results we can state that 33.3% of the employees said that their jobs were very stressful and 46.7% of the employees said that their jobs were somewhat stressful.
- From the cumulative percent, we can state that 80% of the employees said that their jobs were very or somewhat stressful.
- Notice that the sum of the relative frequencies is always 1.00 (or approximately 1.00 if the relative frequencies are rounded), and the sum of the percentages is always 100 (or approximately 100 if the percentages are rounded).

SPSS Result

Stress

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very	10	33.3	33.3	33.3
	Somewhat	14	46.7	46.7	80.0
	None	6	20.0	20.0	100.0
	Total	30	100.0	100.0	

Graphical Presentation of Qualitative Data

- A graphic display can reveal at a glance the **main characteristics** of a data set.
- The *bar graph* and the *pie chart* are two types of graphs that are commonly used to display qualitative data.

Bar Graphs

- To construct a **bar graph** (also called a *bar chart*), we mark the various categories on the horizontal axis.
- Note that all categories are represented by intervals of the same width.
- We mark the frequencies on the vertical axis.
- Then we draw one bar for each category such that the height of the bar represents the frequency of the corresponding category.
- We leave a small gap between adjacent bars.

Figure 1 Bar graph for the frequency distribution.

Pie Charts

- A **pie chart** is more commonly used to **display percentages**, although it can be used to **display frequencies or relative frequencies**.
- The **whole pie** (or circle) represents the **total sample or population**.
- Then we **divide the pie into different portions** that represent the **different categories.**

- As we know, a circle contains 360 degrees. To construct a pie chart, we **multiply 360 by the relative frequency** of each category to obtain the degree measure or size of the angle for the corresponding category.

Table 5 shows the calculation of angle sizes for the various categories of Table 4.

Table 5 Calculating Angle Sizes for the Pie Chart

Stress on Job	Relative Frequency	Angle Size
Very	.333	360(.333) =119.88
Somewhat	. 467	360(.467) =168.12
None	.200	360(.200) =72.00
Sum	1.000	360

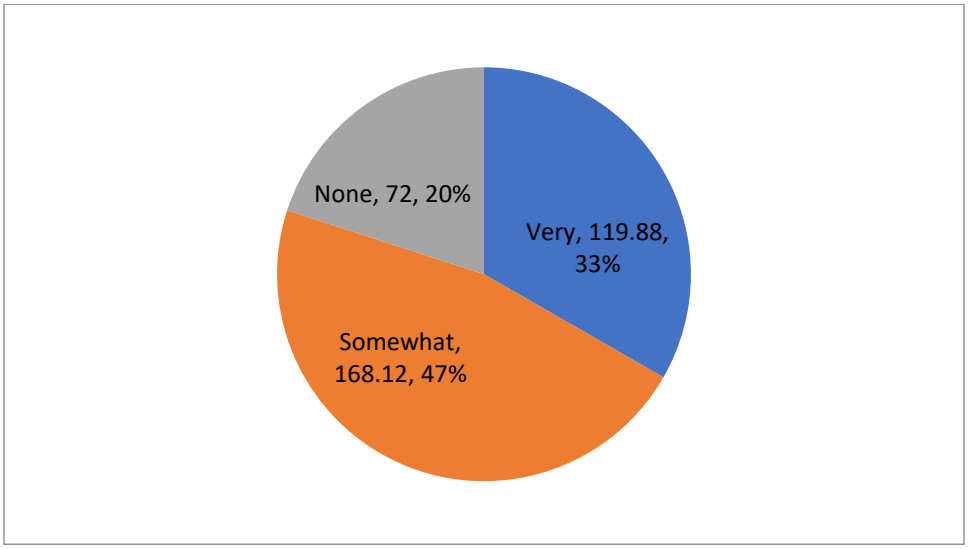

Figure: Pie chart for the percentage distribution of Table 2.5.

Quantitative Raw Data

- Suppose we collect information on the ages (in years) of 50 students selected from a university.
- The data values, in the order they are collected, are recorded in the following Table.

➢

Table: Ages of 50 Students

21	19	24	25	29	34	26	27	37	33
18	20	19	22	19	19	25	22	25	23
25	19	31	19	23	18	23	18	23	19
22	28	21	20	22	21	20	19	21	22
36	25	18	37	27	23	21	25	21	24

Organizing and Graphing Quantitative Data

Frequency Distribution for Quantitative Data :

Types Frequency Distribution for Quantitative Data

1. Ungrouped and
2. Grouped

Ungrouped Frequency Distribution:

- An ungrouped *frequency distribution* for quantitative data lists all the values and the number of values that belong to each value.
- It is constructed when the list of the values of the variable is short.

EXAMPLE: The following data represents the number of family members of 50 families

7	3	5	6	2	4	5	6	3	4
2	5	4	5	5	6	2	3	4	5
5	6	3	5	7	5	8	6	5	7
4	8	5	2	6	3	4	5	3	8
7	5	6	5	6	5	4	7	8	5

Construct an ungrouped frequency distribution.

Solution:

Number of family members (x_i)	Tally	Frequency (f_i)
2		4
3		6
4		7
5		16
6		8
7		5
8		4
Total		50

SPSS RESULT

Familymem

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	4	8.0	8.0	8.0
	3	6	12.0	12.0	20.0
	4	7	14.0	14.0	34.0
	5	16	32.0	32.0	66.0
	6	8	16.0	16.0	82.0
	7	5	10.0	10.0	92.0
	8	4	8.0	8.0	100.0
	Total	50	100.0	100.0	

CONSTRUCTION OF A GROUPED FREQUENCY DISTRIBUTION

A grouped *frequency distribution* for quantitative data lists

- all the classes/groups/class intervals and
- the number of values that belong to each class.

This type of frequency distribution is constructed when the list of the values of the variable is long.

- ❖ However, the classes or intervals for a frequency distribution of numerical data are not as easy identifiable.
- ❖ To determine the intervals of a frequency distribution for numerical data requires answers to 4 questions:
  - (i) How many intervals/classes/groups should be used?
  - (ii) How wide should each interval be?
  - (iii) Where does the first interval begin?
  - (iv) Where does the last interval end?

Answers to the above questions

Number of Intervals (answer to question #(i))

Determine k, the number of intervals (classes). k may be determined by the Sturges's rule $k = 1 + 3.322(\log_{10} n)$, where n is the total number of observations.

Alternative Formula

$$2^k \leq n$$

For n = 40

$$2^k \leq 40$$

$$2^5 = 32 \leq 40$$

So, k = 5.

Class Width (answer to question #(ii))

Intervals should be of the same width, w; the width is determined by the following:

$$w = Interval\ \ Width = \frac{Largest\ Number - Smallest\ Number}{Number\ of\ \ Intervals(k)}$$

- ➢ The value of w calculated by this formula use as guideline for appropriate class width.
- ➢ The value of ***w*** should be rounded up to a closest convenient whole number (usually 5, 10, 15, 20, 25, 30 etc.) to provide for easy interpretation and easy to write and for a nice presentation of data as a grouped frequency distribution.

Example:

Table 6: Days to maturity 40 short-term investments.

71	64	99	55	64	89	87	65	62	38
67	70	60	69	78	39	75	56	71	51
99	68	95	86	57	53	47	50	55	81
80	98	51	36	63	66	85	79	83	70

Construct a frequency distribution.

Solution:

Number of Classes

$$k = 1 + 3.322(\log_{10} 40)$$

$$= 1 + 3.322(1.60206)$$

$$= 6.32$$

Class interval width = $\frac{R}{k} = \frac{99-36}{6.32} = \frac{63}{6.32} = 9.97$

Now we round up this approximate width to the closest convenient number, say 10.

Lower Limit of the First Class or the Starting Point

- **Any convenient number** that is equal to or less than the smallest value in the data set can be used as the lower limit of the first class.
- As the lowest value is 36, we use 30 as the lower limit of the first class.
- If the last digit of the selected class interval is 0, the last digit of the lower limit should be 0. If the last digit of the selected class interval is 5, the last digit of the lower limit should be either 0 or 5.

Table: Frequency distribution of the data in Table 2.6.

Class interval	Tally marks	Frequency
30—39	𝍸	3
40—49	𝍸	1
50—59	𝍸 𝍸	8
60—69	𝍸 𝍸	10
70—79	𝍸	7
80—89	𝍸	7
90—99	𝍸	4
Total		40

SPSS Result

```
RECODE Data (30 thru 39=1) (40 thru 49=2) (50 thru 59=3) (60 thru 69=4) (70
thru 79=5) (80 thru 89=6) (90 thru 99=7)
```

Frequency distribution in recode

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	3	7.5	7.5	7.5
	2.00	1	2.5	2.5	10.0
	3.00	8	20.0	20.0	30.0
	4.00	10	25.0	25.0	55.0
	5.00	7	17.5	17.5	72.5
	6.00	7	17.5	17.5	90.0
	7.00	4	10.0	10.0	100.0
	Total	40	100.0	100.0	

Assignment: Construct a frequency distribution for following data.

table 3.1 Scores from statistics exam ($N = 70$)

95	57	76	93	86	80	89
76	76	63	74	94	96	77
65	79	60	56	72	82	70
67	79	71	77	52	76	68
72	88	84	70	83	93	76
82	96	87	69	89	77	81
87	65	77	72	56	78	78
58	54	82	82	66	73	79
86	81	63	46	62	99	93
82	92	75	76	90	74	67

Raw Scores

Class Boundary: The *class boundary* is given by the midpoint of the upper limit of one class and the lower limit of the next class.

- The upper boundary of the preceding class and the lower boundary of the succeeding class are the same.

Finding Class Width:
Class width = Upper boundary - Lower boundary

Calculating Class Midpoint or Mark
Class midpoint or mark= (Lower limit + Upper limit)/2

Table 8: Class Boundaries, Class Widths, and Class Midpoints for Table 2.7

Class Limits	Class Boundaries	Frequency	Class Width	Class Midpoint
30-39	29.5 to less than 39.5	**3**	**10**	34.5
40-49	39.5 to less than 49.5	1	10	44.5
50-59	49.5 to less than 59.5	8	10	54.5
60-69	59.5 to less than 69.5	10	10	64.5
70-79	69.5 to less than 79.5	7	10	74.5
80-89	79.5 to less than 89.5	7	10	84.5
90-99	89.5 to less than 99.5	4	10	94.5

- Note that when we write classes using class boundaries, we write *to less than* to ensure that each value belongs to one and only one class.
- As we can see, the upper boundary of the preceding class and the lower boundary of the succeeding class are the same.

Relative Frequency and Percentage Distributions
Calculating Relative Frequency and Percentage

Relative frequency of a class = $\frac{Frequency\ of\ that\ class}{Sum\ of\ all\ frequencies} = \frac{f_i}{\sum f_i}$

Percentage = Relative frequency X 100

Table 9: Class Boundaries, Class Widths, and Class Midpoints for Table 2.7

Class Limits	Class Boundaries	Relative frequency	Percentage
30-39	29.5 to less than 39.5	3/40=0.075	7.5
40-49	39.5 to less than 49.5	1/40=0.025	2.5
50-59	49.5 to less than 59.5	8/40=0.200	20.0
60-69	59.5 to less than 69.5	10/40=0.250	25.0
70-79	69.5 to less than 79.5	7/40=0.175	17.5
80-89	79.5 to less than 89.5	7/40=0.175	17.5
90-99	89.5 to less than 99.5	4/40=0.100	10.0
Total		1.000	100