

Lecture No. 3

Introduction to Statistics Statistics and Probability

Dr. Shabbir Ahmad

**Assistant Professor,
Department of Mathematics,
COMSATS University
Islamabad, Wah Campus**

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

Frequency Distribution and Graphs of Frequency Distributions, Cross Tabulation

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

In this lecture

- Frequency, Distribution
- Relative and Percentage Frequency Distribution
- Procedures for Constructing a Frequency Distribution
- Graphs of Frequency Distribution
- Histogram, Frequency Curve, Cumulative Frequency Curve
- Cross Tabulation or Contingency Table

Frequency Distribution

- A **frequency distribution** for qualitative data is a tabular summary of data showing the frequency (or number) of items in each of the categories.
- The **objective** is to provide insights about the data that cannot be quickly obtained by looking only at the original data.

Example: Food Quality

- A New Japanese restaurant offers lunch at the price of \$30. A random sample of 20 customer is selected; and Rate the quality of food as: Excellent, Above Average, Average, Below Average, or Poor. The ratings are listed below:

Above Average	Below Average	Poor
Average	Poor	Above Average
Above Average	Excellent	Above Average
Below Average	Above Average	Below Average
Above Average	Average	Above Average
Above Average	Average	Average
Average	Above Average	

Rating	Frequency
Poor	2
Below Average	3
Average	5
Above Average	9
Excellent	1
TOTAL	20

Relative and Percentage Frequency Distribution

The **relative frequency** of a data category is the fraction while **percentage relative frequency** is percentage of the total number of data items belonging to the category.

$$\text{Relative Frequency of a Category} = \frac{\text{Frequency of that Category}}{\text{Sum of all frequencies}}$$

$$\text{Percent Frequency} = (\text{Relative frequency}) \times 100$$

Frequency, Relative Frequency and Percent Frequency Distributions

Rating	Frequency	Relative Frequency	Percent Frequency
Poor	2	.10	10
Below Average	3	.15	15
Average	5	.25	25
Above Average	9	.45	45
Excellent	1	.05	5
TOTAL	20	1.00	100

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

Frequency Distributions of Quantitative Data

A frequency distribution is a table containing classes (usually) with equal width and the corresponding frequencies with which data fall within each of the classes.

Why Use a Frequency Distribution?

- It is a way to summarize numerical data
- It condenses the raw data into a more useful form...
- It allows for a quick visual interpretation of the data

Procedures for Constructing a Frequency Distribution of Quantitative Data

- Determine the number of non-overlapping classes:
 - Never overlap with one another.
 - Recommend 5 to 20 classes.
 - **GOAL:** Uses enough classes to show the variations in the data.
- Determine the width of each class(Class Width):

$$= \frac{\text{Largest Data Value} - \text{Smallest Data Value}}{\text{Number of Classes}}$$

Procedures for Constructing a Frequency Distribution of Quantitative Data

- Determine the class limits and boundaries:
 - Class Limits:
 - Lower limit: smallest possible class data value;
 - Upper limit: largest possible class data value;
 - Class Boundary:
 - mid-point of the upper limit of class and the lower limit of the next class.
 - Class Mid-point (Class Mark):

$$\text{Class Mark} = \frac{\text{Lower Limit} + \text{Upper Limit}}{2}$$

Example of Constructing a Frequency Distribution of Quantitative Data

- Sample data for the part cost of 50 vehicle check up

Automobile Data

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

- Find the **smallest and largest observations**: 52, 109
- Find the **range**: $109 - 52 = 57$
- Select the **number of non-overlapping classes**: let's use 6 classes.
- Compute class interval **width**: $(109 - 52)/6 = 9.5 \approx 10$
- Determine the class limits.
- Count the number of observations** in each class

Frequency Distribution

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73



Find the lower class limit and upper class limit for the class 60-69.

Class limits are chosen so that each data item belongs to one and only one class.

<u>Parts Cost (\$)</u>	<u>Frequency</u>
50-59	2
60-69	13
70-79	16
80-89	7
90-99	7
100-109	<u>5</u>
Total	50

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

Relative Frequency and Percent Frequency Distributions

Parts Cost (\$)	Relative Frequency	Percent Frequency
50-59	.04	4
60-69	.26	26
70-79	.32	32
80-89	.14	14
90-99	.14	14
100-109	.10	10
TOTAL	1.00	100

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

Relative Frequency and Relative and Percent Frequency Distributions

- What do you find from the frequency distribution on page 28?
- Insights Gained from the Percent Frequency Distribution
- Only 4% of the parts costs are in the \$50-59 class.
- 30% of the parts costs are under \$70.
- The greatest percentage (32% or almost one-third) of the parts costs are in the \$70-79 class.
- 10% of the parts costs are \$100 or more.

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

More About Frequency Distribution

Intervals (Marks)		Freq	Mid Points X	Class Boundries		Cumulative Freq
1	9	6	5	0.5	9.5	6
10	18	12	14	9.5	18.5	18
19	27	8	23	18.5	27.5	26
28	36	13	32	27.5	36.5	39
37	45	8	41	36.5	45.5	47
46	54	3	50	45.5	54.5	50

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:57 PM

Graphs of Frequency Distribution

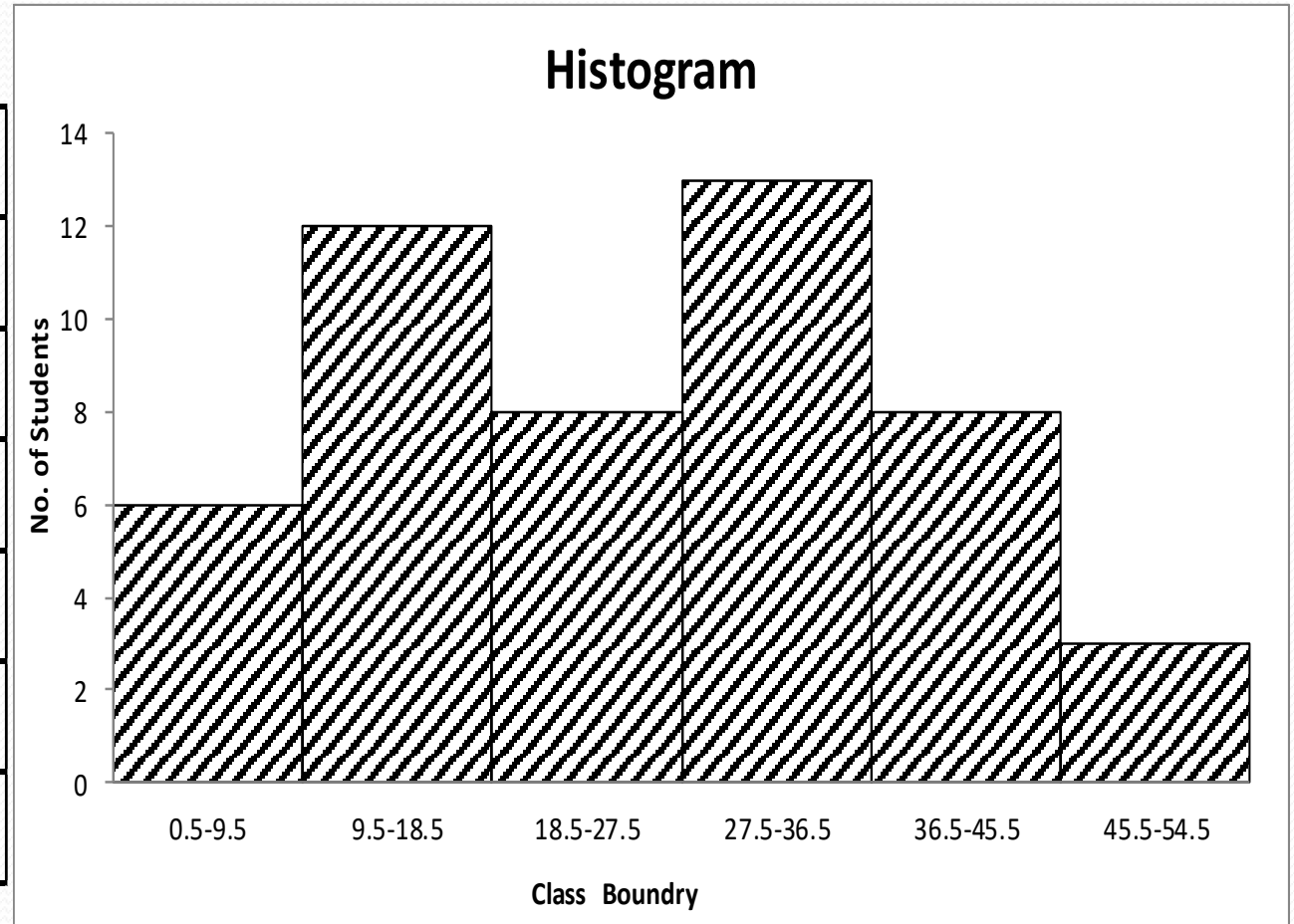
- Histogram
- Frequency Curve
- Cumulative Frequency Curve

Histogram

- A common graphical presentation of quantitative data.
 - X-axis: Variable of Interest;
 - Recommend 5 to 20 equal subintervals;
 - Y-axis: frequency, relative frequency, or percent frequency;
 - A rectangle with appropriate heights is drawn above each class;
- Unlike a bar graph, a histogram has no natural separation between rectangles of adjacent classes.

Histogram

CB	F
0.5-9.5	6
9.5-18.5	12
18.5-27.5	8
27.5-36.5	13
36.5-45.5	8
45.5-54.5	3

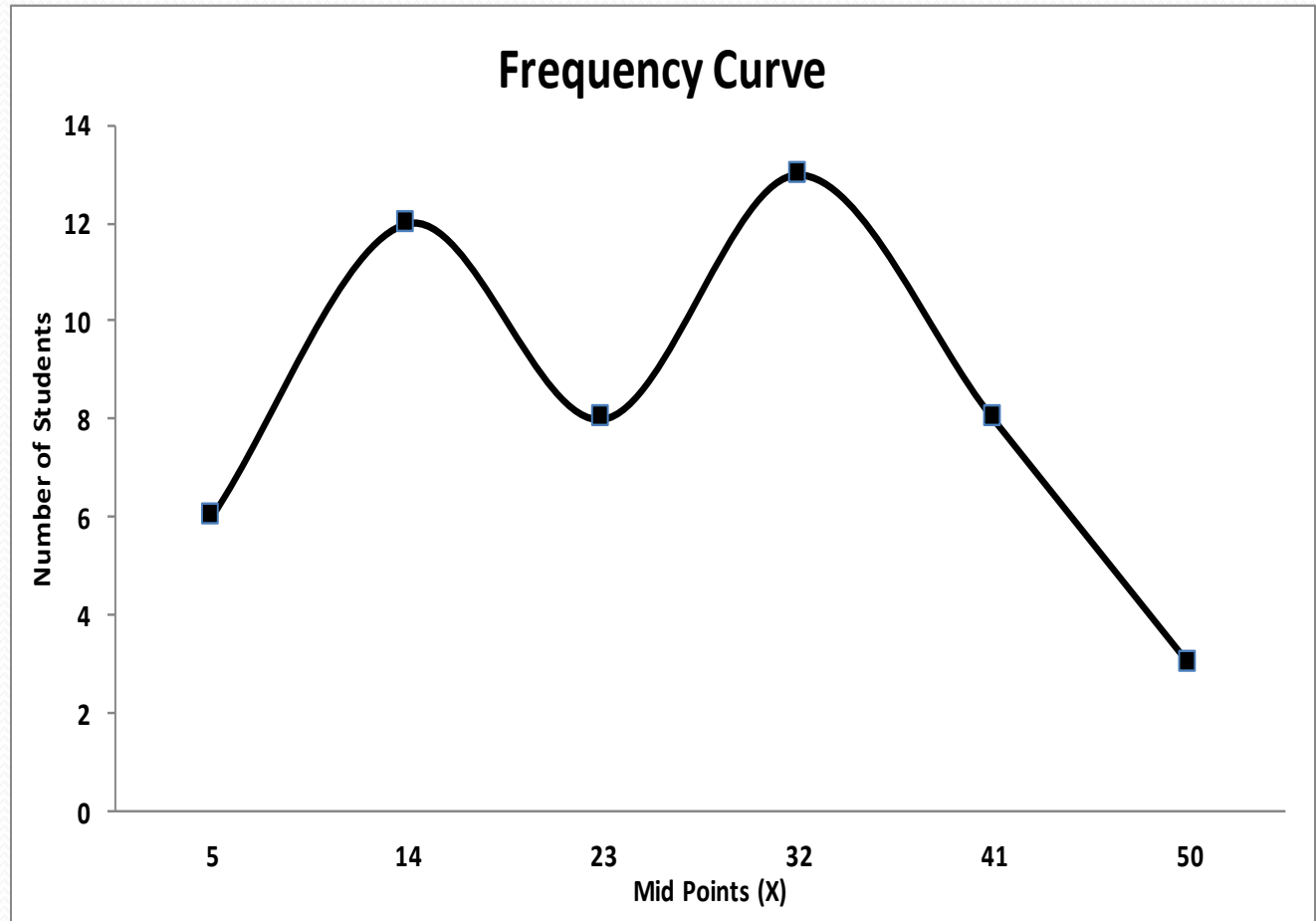


Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:58 PM

Frequency Curve

Mid Points	Freq
5	6
14	12
23	8
32	13
41	8
50	3



Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:58 PM

Cumulative Frequency Curve/Ogive

- An ogive is a graph of a cumulative distribution.
- The data values are shown on the X-axis.
- Shown on the Y-axis are the:
 - cumulative frequencies, or
 - cumulative relative frequencies, or
 - cumulative percent frequencies
- The frequency (one of the above) of each class is plotted as a point above the upper boundaries of each class.
- The plotted points are connected by straight lines.

Cumulative Distributions

Cost (\$)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
≤ 59	2	.04	4
≤ 69	15	.30	30
≤ 79	31	.62	62
≤ 89	38	.76	76
≤ 99	45	.90	90
≤ 109	50	1.00	100

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:58 PM

Cumulative Distributions

- Cumulative frequency distribution

shows the number of items with values less than or equal to the upper limit of each class.

- Cumulative relative frequency distribution

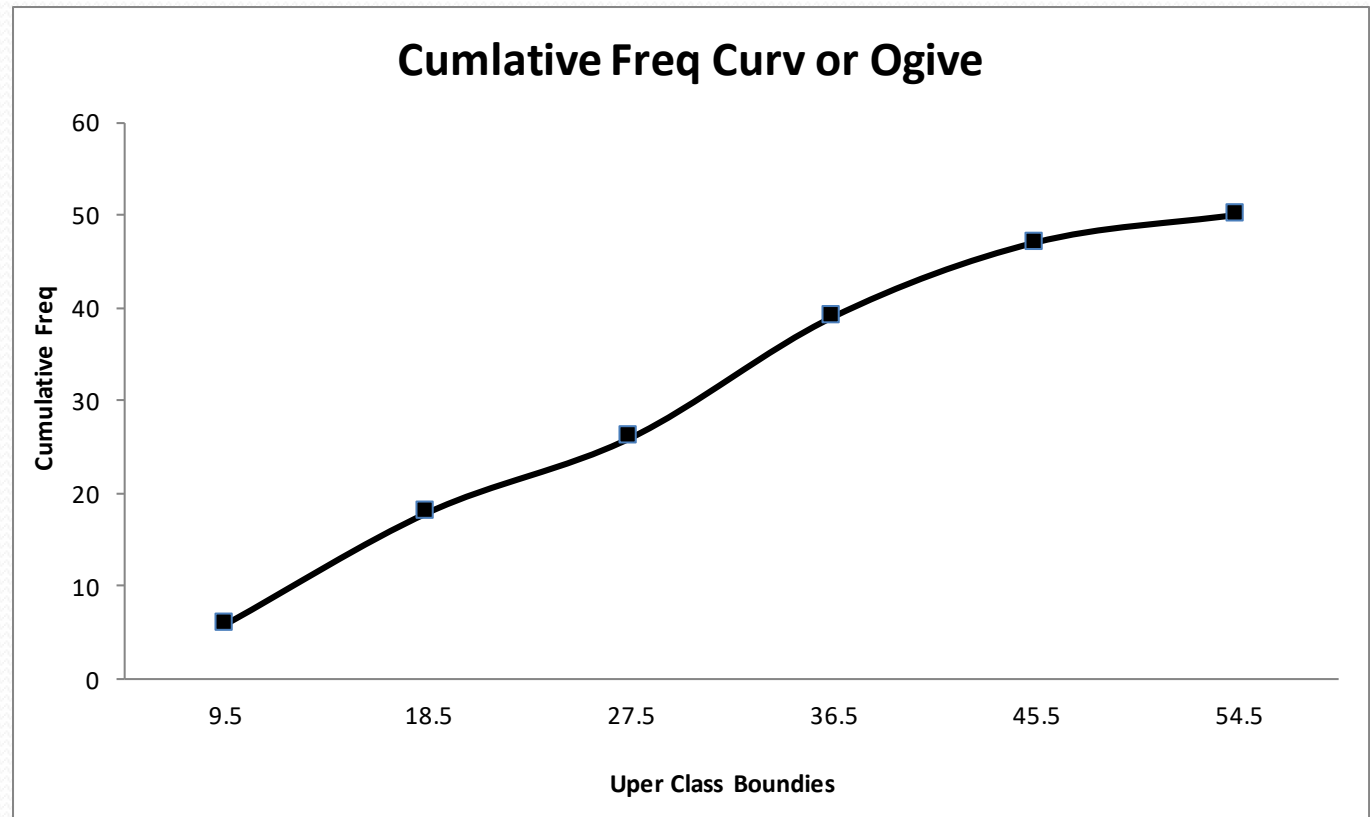
shows the proportion of items with values less than or equal to the upper limit of each class.

- Cumulative percent frequency distribution

shows the percentage of items with values less than or equal to the upper limit of each class.

Cumulative Frequency Curve/Ogive

UCB	CF
9.5	6
18.5	18
27.5	26
36.5	39
45.5	47
54.5	50



Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:58 PM

Cross Tabulation or Contingency Table

- A cross tabulation or contingency table is a tabular summary of data for two variables.
- Cross tabulation can be used when:
 - Qualitative vs Quantitative;
 - Qualitative vs Qualitative;or
 - Quantitative vs Quantitative.
- The left and top margin labels define the classes for the two variables.

Cross Tabulation or Contingency Table

- Example: Automobile insurance




An automobile company sells three types of cars: Full-size, pick-up-truck and convertible. To better understand the purchasing behavior, and to improve the advertising plan, an initial study was performed based on the recent 200 sales. In particular, we studied the following two variables.

Variable 1: Type of car

Variable 2: Sex of car owner

Both variables are qualitative.

Cross Tabulation or Contingency Table

Sex	<u>Type of car</u>			Total
	 Full-size	 Pick-up-truck	 Convertible	
Male	57	38	20	115
Female	63	12	10	85
Total	120	50	30	200

Dr. Shabbir Ahmad

Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 11-Sep-21 4:17:58 PM