

# Unit Two: Descriptive Biostatistics

## Unit 2: Lecture 3

Hana Taha, PhD





# Objectives of Tabulation

- **To make complex data simpler:** The main aim of tabulation is to present the classified data in a systematic way. The purpose is to condense the bulk of information (data) under investigation into a simple and meaningful form.
- **To save space:** Tabulation tries to save space by condensing data in a meaningful form while maintaining the quality and quantity of the data.
- **To facilitate comparison:** It also aims to facilitate quick comparison of various observations by providing the data in a tabular form.

# Objectives of Tabulation

- **To facilitate statistical analysis:** Tabulation aims to facilitate statistical analysis because it is the stage between data classification and data presentation. Various statistical measures are easily calculated from data that has been systematically tabulated.
- **To provide a reference:** Since data may be easily identifiable and used when organized in tables with titles and table numbers, tabulation aims to provide a reference for future studies.



# **Classification of Data and Tabular Presentation**

**Spatial  
Classification  
of Data and  
Tabular  
Presentation**

**Quantitative  
Classification of  
Data and Tabular  
Presentation**

**Temporal  
Classification  
of Data and  
Tabular  
Presentation**

**Qualitative  
Classification of  
Data and Tabular  
Presentation**

# Classification of Data

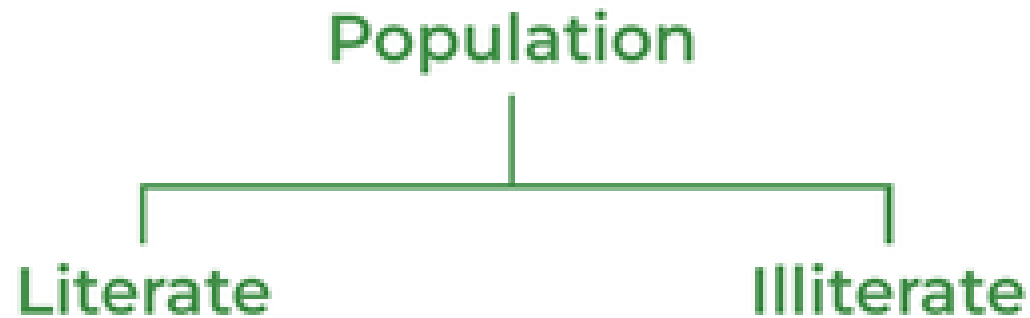
- **Geographical or Spatial Classification.** For example, presenting the population of different governorates of Jordan is done on the basis of geographical location or region.
- **Chronological or Temporal Classification.** For example, the number of students in a school in different years can be presented on the basis of a time period.

# Qualitative Classification

- The classification of data on the basis of descriptive or qualitative characteristics like region, sex, gender, education, etc., is known as **Qualitative Classification**.
- A qualitative classification can be of two types:  
**Simple Classification** and **Manifold Classification**.

# Simple Classification

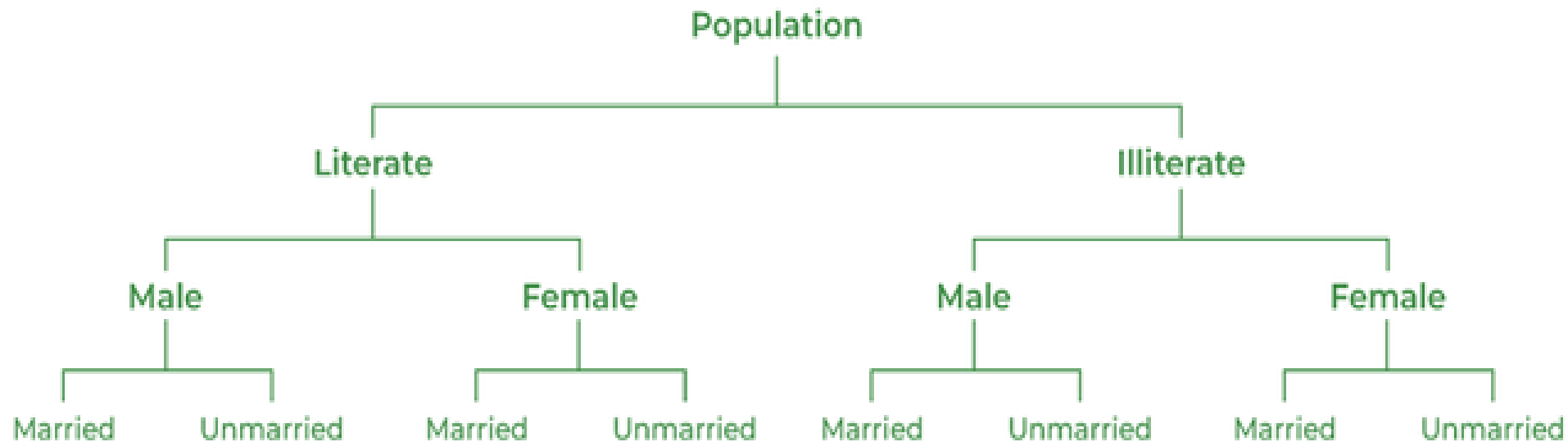
- When based on only one attribute, the given data is classified into two classes, which is known as **Simple Classification**. For **example**, when the population is divided into literate and illiterate, it is a simple classification.





# Manifold Classification

- When based on more than one attribute, the given data is classified into different classes, and then sub-divided into more sub-classes, which is known as **Manifold Classification**.
- **For example**, when the population is divided into literate and illiterate, then sub-divided into male and female, and further sub-divided into married and unmarried, it is a manifold classification.



# Summarizing Qualitative Data

- Frequency Distribution
- Relative Frequency Distribution
- Percent Frequency Distribution
- Bar Graph
- Pie Chart

# Frequency Tables


- ▶ A frequency table is a tabular summary of data showing the frequency (or number) of items in each of several non-overlapping classes.
- ▶ The objective is to provide insights about the data that cannot be quickly obtained by looking only at the original data.

# Example: Assessment of healthcare quality

Patients visiting Ein Al Basha Comprehensive Health Center were asked to rate the quality of care **excellent**, **above average**, **average**, **below average**, or **poor**. The ratings provided by a sample of 20 patients are:

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

## Frequency Table



<u>Rating</u>	<u>Frequency</u>
Poor	2
Below Average	3
Average	5
Above Average	9
Excellent	<u>1</u>
Total	20

## Relative Frequency Table

▶ The relative frequency of a class is the fraction or proportion of the total number of data items belonging to the class.

▶ A relative frequency distribution is a tabular summary of a set of data showing the relative frequency for each class.

# Percent Frequency Table

- ▶ The percent frequency of a class is the relative frequency multiplied by 100.
- ▶ A percent frequency distribution is a tabular summary of a set of data showing the percent frequency for each class.

# Relative Frequency and Percent Frequency Tabulation

<u>Rating</u>	<u>Relative Frequency</u>	<u>Percent Frequency</u>
Poor	.10	10
Below Average	.15	15
Average	.25	25
Above Average	.45	45
Excellent	<u>.05</u>	<u>5</u>
Total	1.00	100

Callout 1:  $.10(100) = 10$  (points to the Percent Frequency for 'Poor')

Callout 2:  $1/20 = .05$  (points to the Relative Frequency for 'Excellent')

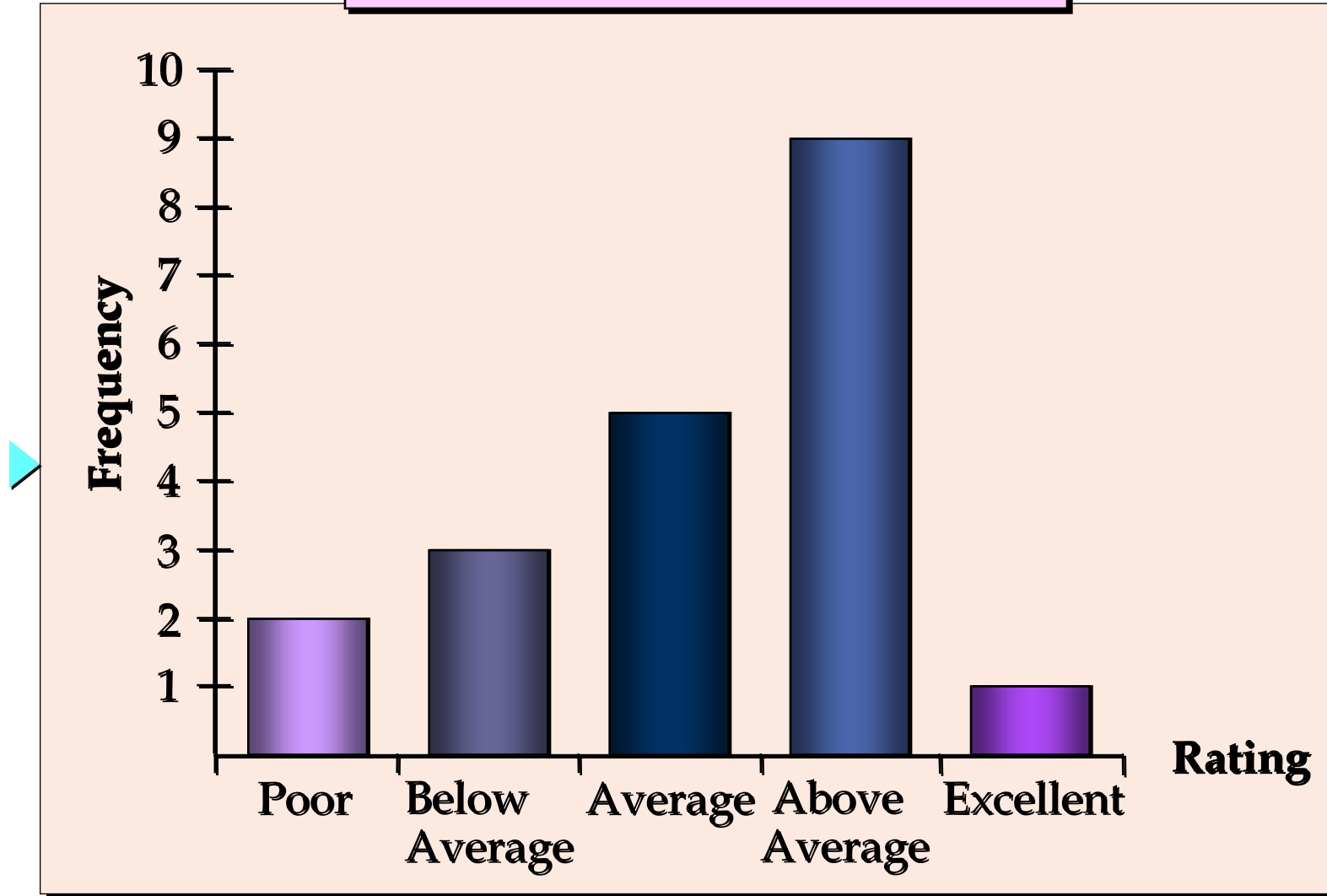


# Bar Graph

- ▶ ■ A bar graph is a graphical device for showing qualitative data.
- ▶ ■ On one axis (usually the horizontal axis), we specify the labels that are used for each of the classes.
- ▶ ■ A frequency, relative frequency, or percent frequency scale can be used for the vertical axis usually
- ▶ ■ Using a bar of fixed width drawn above each class label, we extend the height appropriately.
- ▶ ■ The bars are separated to emphasize the fact that each class is a separate category.

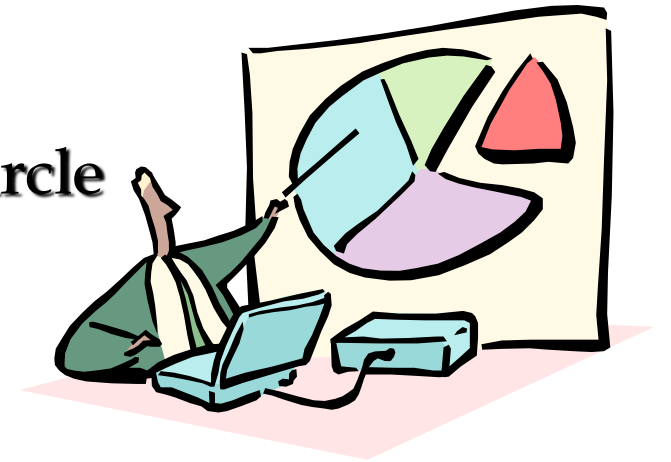
# Bar Chart

Healthcare Quality Ratings



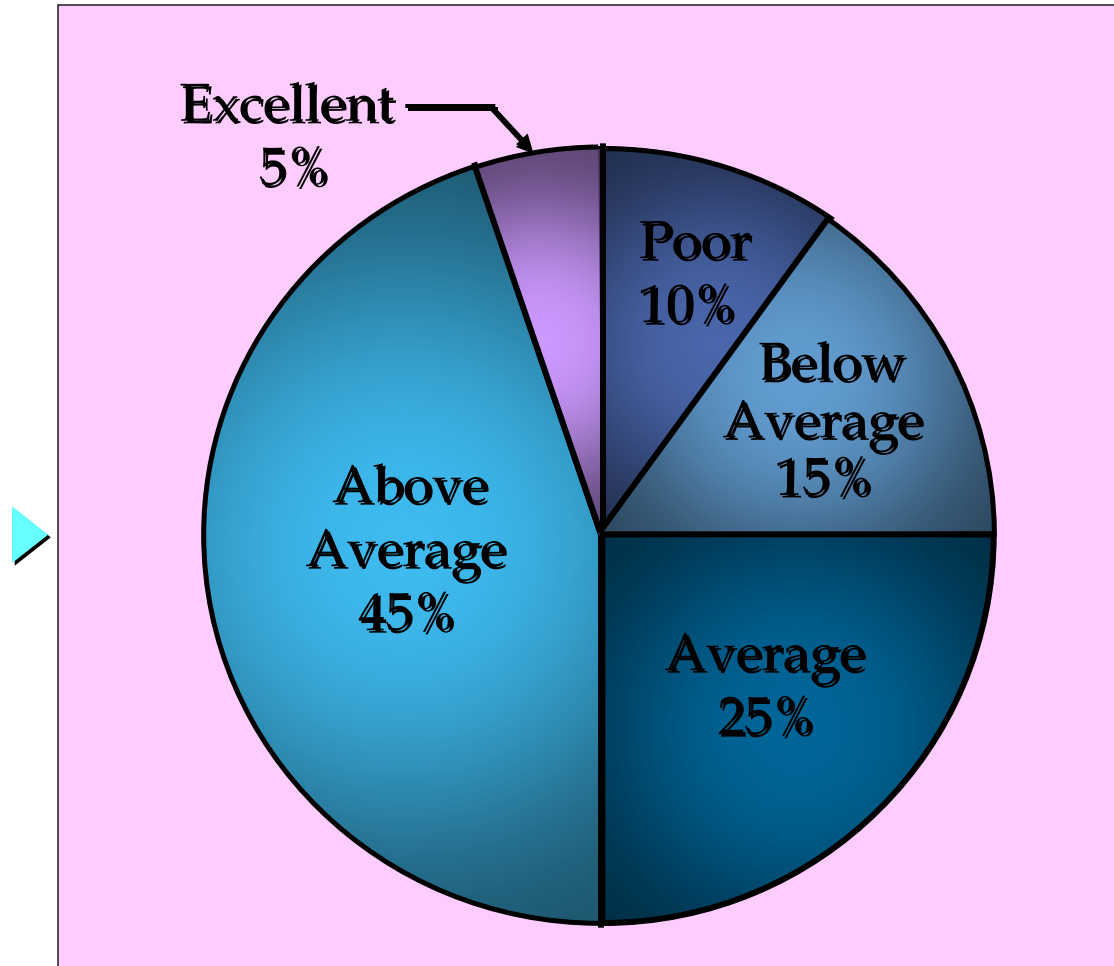
# Pie Chart

- ▶ The pie chart is a commonly used graphical device for presenting relative frequency distributions for qualitative data.
- ▶ We use the relative frequencies to subdivide the circle into sectors that correspond to the relative frequency for each class.

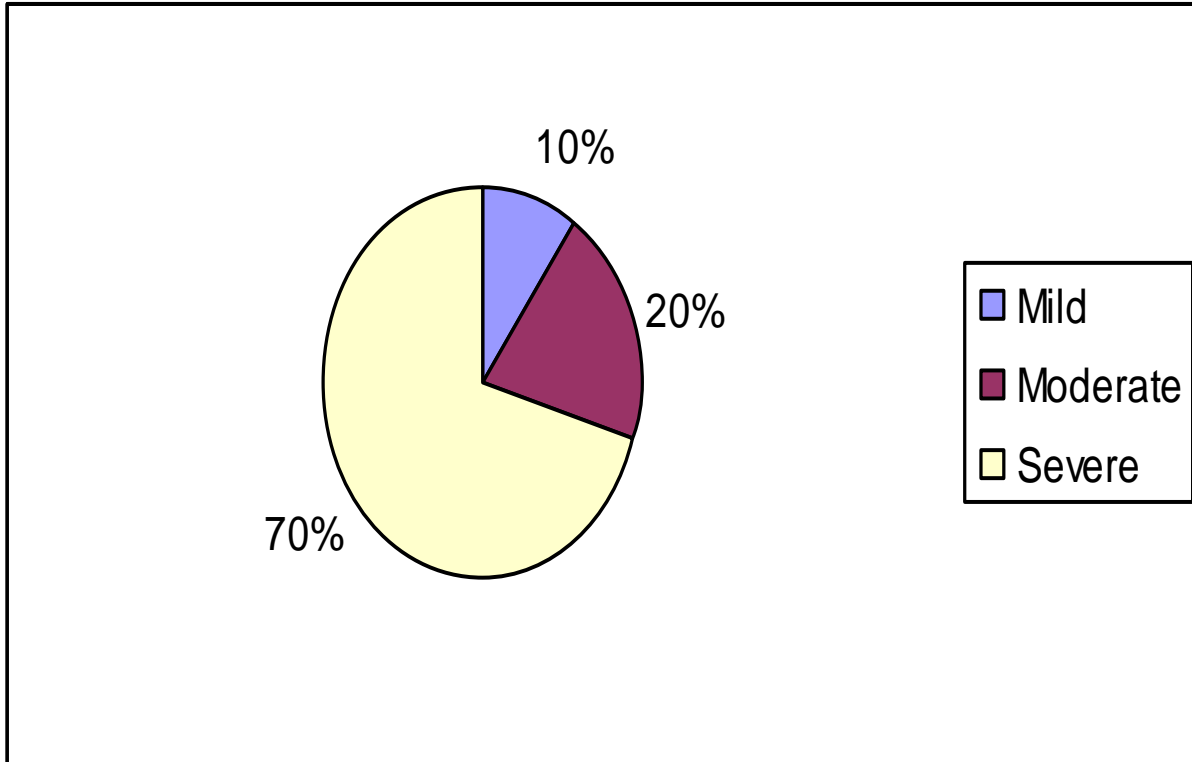


# Pie Chart

Healthcare Quality Ratings



# Pie Chart



- **Circular diagram – total -100%**
- **Divided into segments each representing a category**
- **The amount for each category is proportional to slice of the pie**

**The prevalence of different degree of Hypertension  
in the population**

# Summarizing Quantitative Data

- Recategorization and frequency tabulation
- Relative Frequency and  
Percent Frequency tabulation
- Histogram
- Cumulative Distributions
- Ogive
- Dot Plot

**Example 1: Tabulate the hemoglobin values of 30 adult male patients listed below**

Patient No	Hb (g/dl)	Patient No	Hb (g/dl)	Patient No	Hb (g/dl)
1	12.0	11	11.2	21	14.9
2	11.9	12	13.6	22	12.2
3	11.5	13	10.8	23	12.2
4	14.2	14	12.3	24	11.4
5	12.3	15	12.3	25	10.7
6	13.0	16	15.7	26	12.5
7	10.5	17	12.6	27	11.8
8	12.8	18	9.1	28	15.1
9	13.2	19	12.9	29	13.4
10	11.2	20	14.6	30	13.1

# Steps for making a table

**Step1 Find Minimum (9.1) & Maximum (15.7)**

**Step2 Calculate difference  $15.7 - 9.1 = 6.6$**

**Step3 Decide the number and width of  
the classes or categories (7 c.l) 9.0 -9.9, 10.0-10.9,-----**

**Step4 Prepare dummy table**



## **Table Frequency distribution of 30 adult male patients by Hb**

<b>Hb (g/dl)</b>	<b>No. of patients</b>
<b>9.0 – 9.9</b>	<b>1</b>
<b>10.0 – 10.9</b>	<b>3</b>
<b>11.0 – 11.9</b>	<b>6</b>
<b>12.0 – 12.9</b>	<b>10</b>
<b>13.0 – 13.9</b>	<b>5</b>
<b>14.0 – 14.9</b>	<b>3</b>
<b>15.0 – 15.9</b>	<b>2</b>
<b>Total</b>	<b>30</b>

# **Table Frequency distribution of adult patients by Hb and gender:**

<b>Hb (g/dl)</b>	<b>Gender</b>		<b>Total</b>
	<b>Male</b>	<b>Female</b>	
<b>&lt;9.0</b>	<b>0</b>	<b>2</b>	<b>2</b>
<b>9.0 – 9.9</b>	<b>1</b>	<b>3</b>	<b>4</b>
<b>10.0 – 10.9</b>	<b>3</b>	<b>5</b>	<b>8</b>
<b>11.0 – 11.9</b>	<b>6</b>	<b>8</b>	<b>14</b>
<b>12.0 – 12.9</b>	<b>10</b>	<b>6</b>	<b>16</b>
<b>13.0 – 13.9</b>	<b>5</b>	<b>4</b>	<b>9</b>
<b>14.0 – 14.9</b>	<b>3</b>	<b>2</b>	<b>5</b>
<b>15.0 – 15.9</b>	<b>2</b>	<b>0</b>	<b>2</b>
<b>Total</b>	<b>30</b>	<b>30</b>	<b>60</b>

# Elements of a Table

Ideal table should have

- Number
- Title
- Column headings
- Foot-notes

Number – Table number for identification in a report

Title, place  
Time period - Describe the body of the table, variables,  
(What, how classified, where and when)

Column - Variable name, No. , Percentages (%), etc.,  
Heading

Foot-note(s) - to describe some column/row headings,  
special cells, source, etc.,

# Example 2: Al Rahmah Hospital

- ▶ The manager of the hospital aims to tabulate the daily price of patient stay in the hospital. She examines 50 invoices, and the prices are listed below.

91	78	93	57	75	50	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

# Frequency Distribution

- Guidelines for Selecting Number of Classes
  - ▶ • Use between 5 and 20 classes.
  - ▶ • Data sets with a larger number of elements usually require a larger number of classes.
  - ▶ • Smaller data sets usually require fewer classes.

# Frequency Distribution

- Guidelines for Selecting Width of Classes

- ▶ • Use classes of equal width.

- ▶ • Approximate Class Width =

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

# Frequency Distribution

▶ Approximate Class Width =  $(109 - 50)/6 = 9.83 \cong 10$

<u>Daily price (\$)</u>	<u>Frequency</u>
50-59.9	2
60-69.9	13
70-79.9	16
80-89.9	7
90-99.9	7
100-110	<u>5</u>
Total	50

# Relative Frequency and Percent Frequency Distributions

The table displays the relationship between relative and percent frequencies. Two callouts illustrate the conversion:  $2/50$  is shown as a callout for the relative frequency of 0.04, and  $.04(100)$  is shown as a callout for the percent frequency of 4.

<u>Daily price (\$)</u>	<u>Relative Frequency</u>	<u>Percent Frequency</u>
50-59.9	.04	4
60-69.9	.26	26
70-79.9	.32	32
80-89.9	.14	14
90-99.9	.14	14
100-110	<u>.10</u>	<u>10</u>
Total	1.00	100

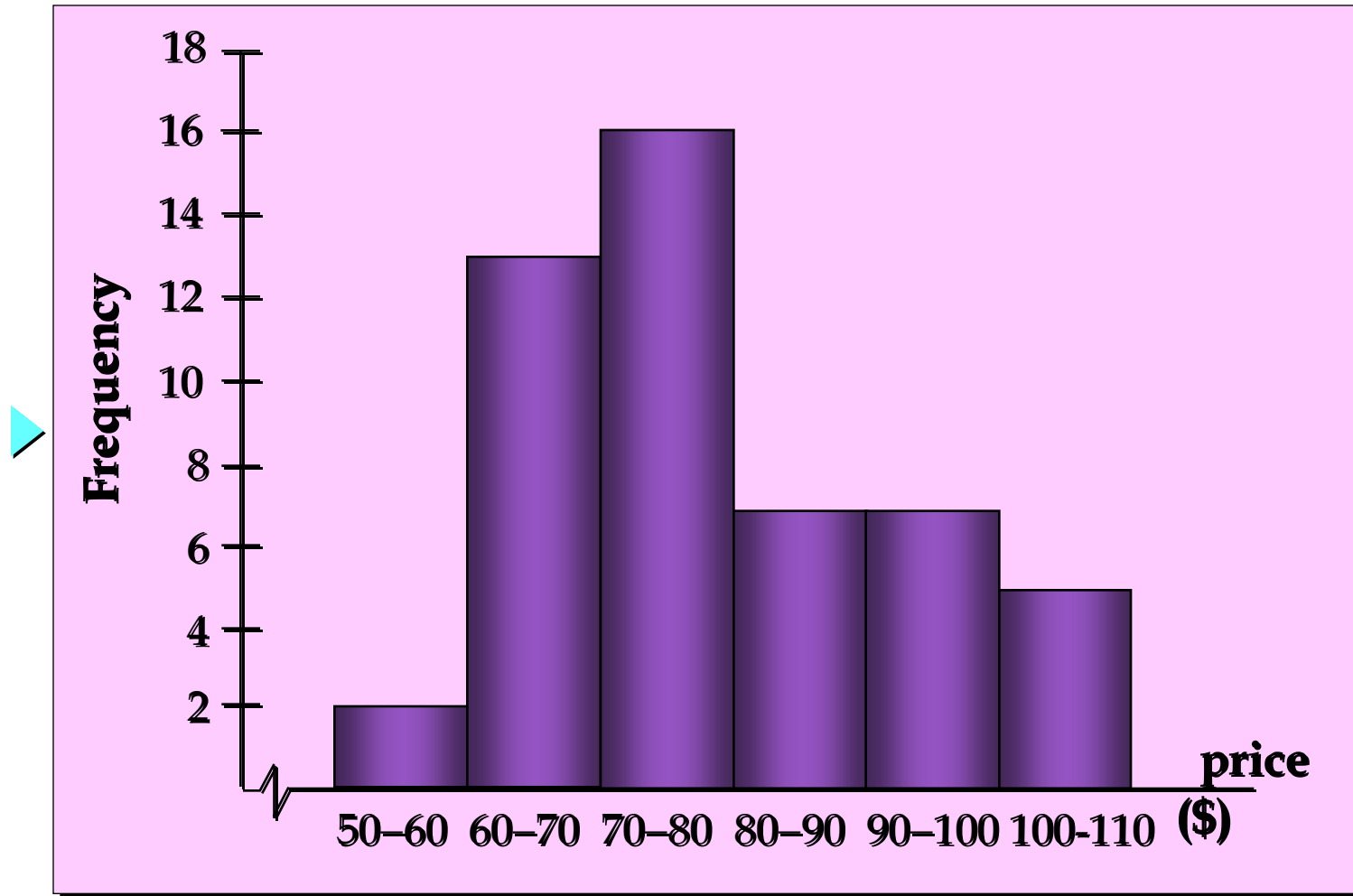


# Histogram

- ▶ ■ A common graphical presentation of quantitative data is a histogram.
- ▶ ■ The variable of interest is placed on the horizontal axis.
- ▶ ■ A rectangle is drawn above each class interval with its height corresponding to the interval's frequency, relative frequency, or percent frequency.
- ▶ ■ Unlike a bar graph, a histogram has no natural separation between rectangles of adjacent classes.

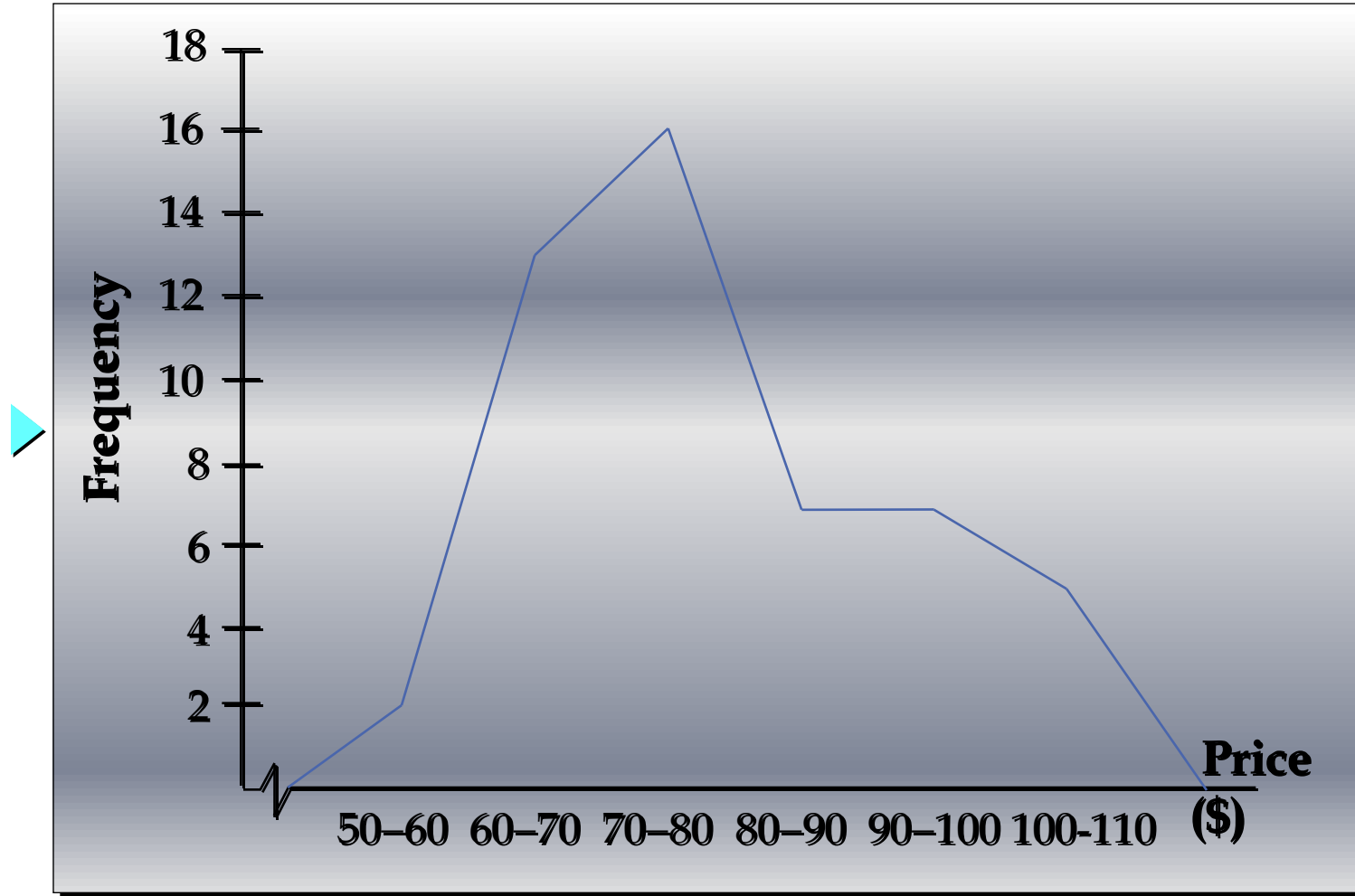
# Histogram

Daily price of staying in Al Rahmah Hospital



# Polygon

Daily price of staying in Al Rahmah Hospital



# Cumulative frequency tables

▶ Cumulative frequency – shows the *number* of items with values less than or equal to the upper limit of each class..

▶ Cumulative relative frequency– shows the *proportion* of items with values less than or equal to the upper limit of each class.

▶ Cumulative percent frequency– shows the *percentage* of items with values less than or equal to the upper limit of each class.

# Cumulative Frequency Table

Daily price of staying in Al Rahmah Hospital

<u>Price (\$)</u>	<u>Cumulative Frequency</u>	<u>Cumulative Relative Frequency</u>	<u>Cumulative Percent Frequency</u>
$\leq 50$	0	0	0
$\leq 60$	2	.04	4
$\leq 70$	15	.30	30
$\leq 80$	31	.62	62
$\leq 90$	38	.76	76
$\leq 100$	45	.90	90
$\leq 110$	50	1.00	100

Callouts for the row  $\leq 80$ :

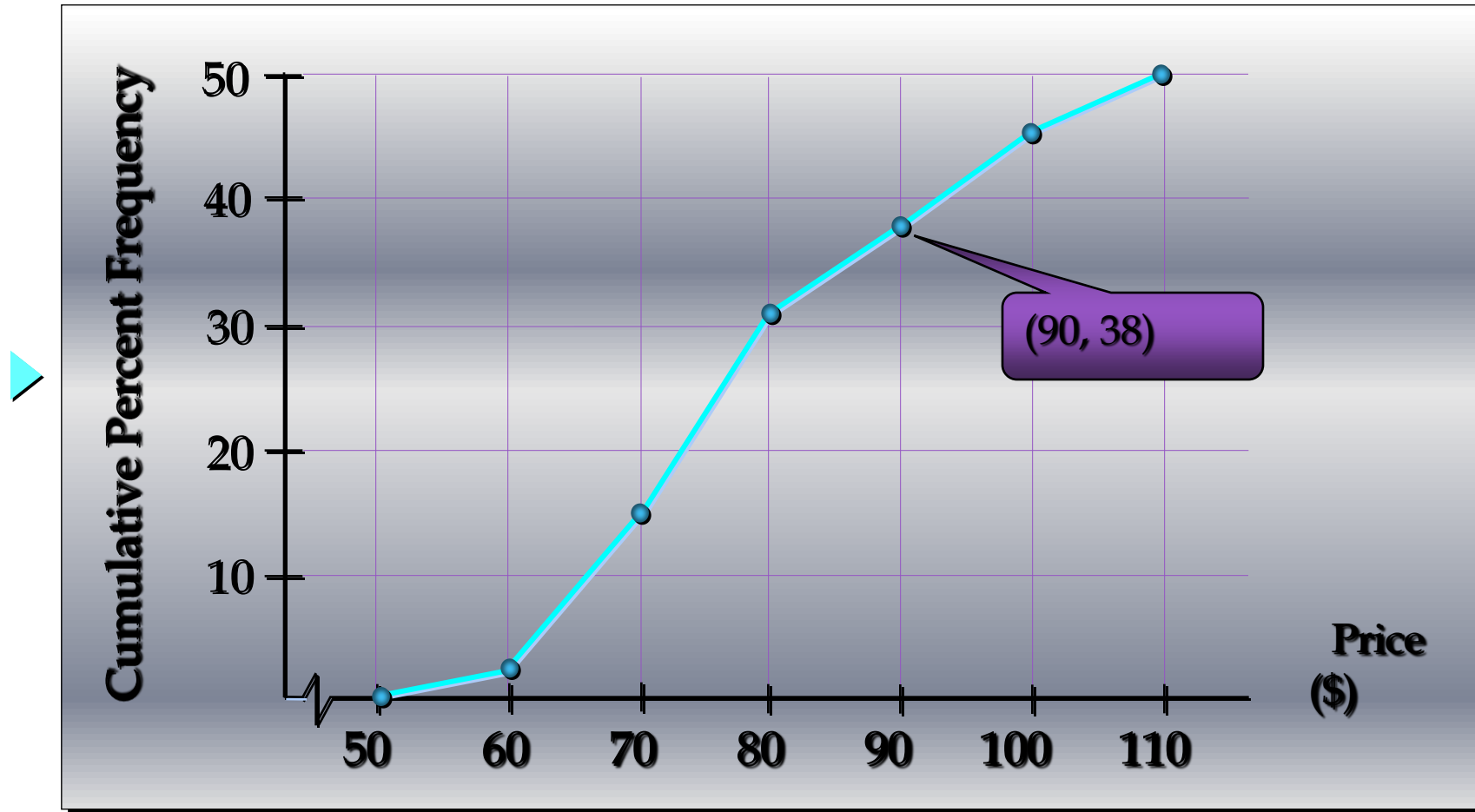
- Cumulative Frequency:  $2 + 13$
- Cumulative Relative Frequency:  $15/50$
- Cumulative Percent Frequency:  $.30(100)$

# Ogive

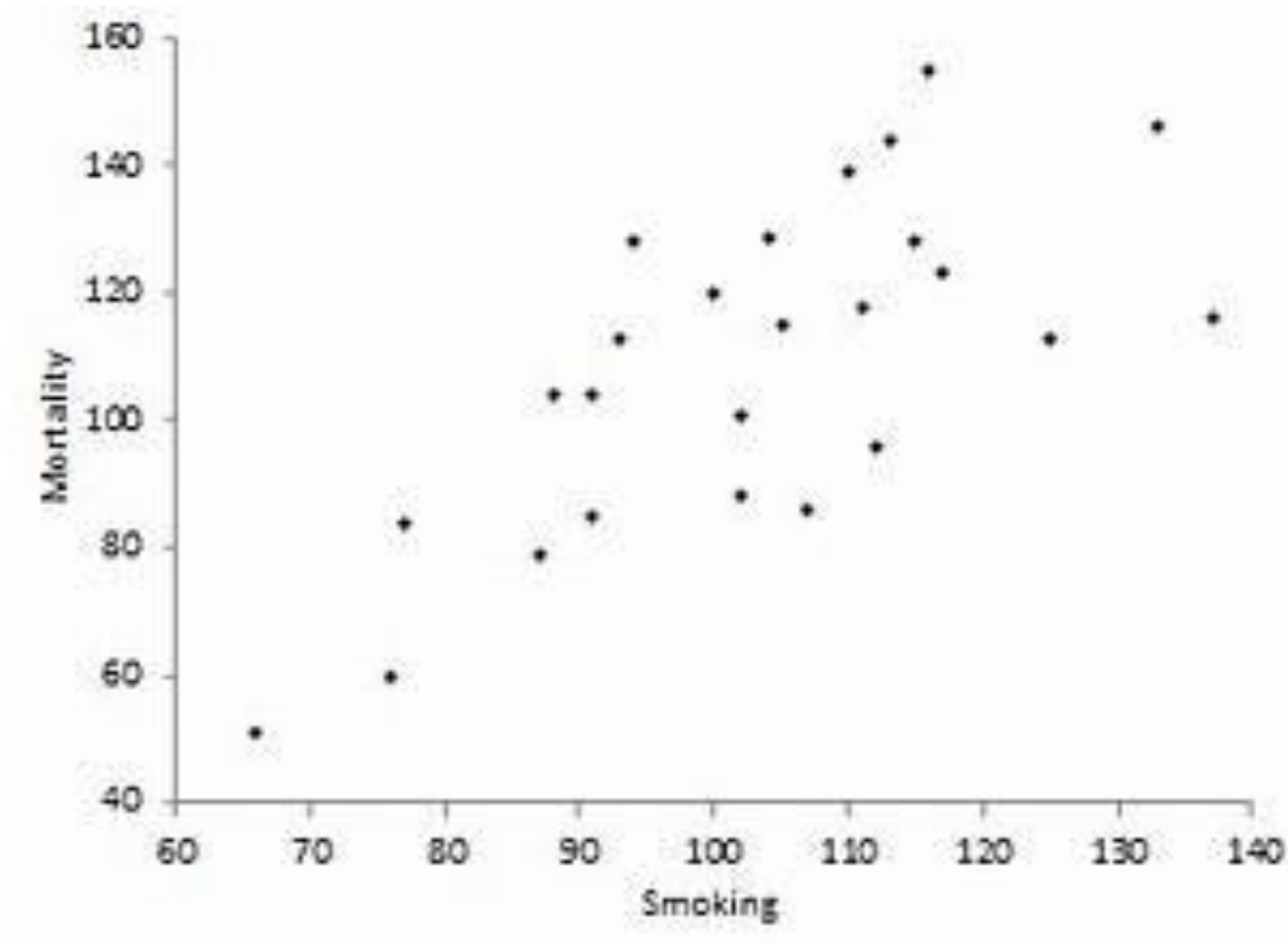
- ▶ □ An ogive is a graph of a cumulative distribution.
- ▶ □ The data values are shown on the horizontal axis.
- ▶ □ Shown on the vertical axis are the:
  - cumulative frequencies, or
  - cumulative relative frequencies, or
  - cumulative percent frequencies
- ▶ □ The frequency (one of the above) of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

# Ogive with Cumulative Frequencies

Daily price of staying in Al Rahmah Hospital



# Dot Plot



Dot chart or dot plot:  
Consists of data points plotted on a simple scale. It is very similar to the frequency table, but instead of using numbers to show frequency, it uses dots. Each dot represents a data point.



# General rules for designing graphs

- A graph should have a clear title and a self-explanatory legend
- A graph should help reader to understand data
- Scales are important; axis must be labeled and units of measurement indicated
- Avoid graphs with three-dimensional impression, it may be misleading (reader visualize less easily)

# Why Tables and Graphs?

- Summarize relevant data and results in a concise and easy to understand manner
- Save readers time and energy and reduce word count in the results section of a research manuscript
- If poorly constructed, can lead to incorrect interpretation or misinterpretation of results
- Readers who go beyond the abstract of a paper are likely to examine the graphs and tables next

# General Principles

- What is your study question? What points would you like to convey through tables and graphs?
- Know your data. What type of variables do you have? (nominal, ordinal, discrete, continuous)
- Look at distributions of all study variables, create the tables and graphs as needed.

## Community medicine: Biostatistics Attendance Barcode

- Register your attendance with your university number
- Make sure that the settings of your phone allow tracking location. The location should be turned on to complete this process. Go to settings > privacy> location> services> make sure that location services is ON
- If the student's name doesn't pop up, he should stop by the examination office to add it manually.



**Thank you**

