

# Probability and Statistics

Prepared by Dr. Alfaisal Hasan

STA 201  
Probability and Statistics

Lecture NO 2



# Probability and Statistics

## Lecture NO 2

STA 201

Probability and Statistics

Prepared by  
**Dr. Alfaisal Hasan**

Chapter

## Topics

- Chapter 1: Introduction To Statistics
- **Chapter 2: Descriptive Statistics**
- Chapter 3: Probability
- Chapter 4: Discrete Probability Distributions
- Chapter 5: Normal Probability Distributions
- Chapter 6: Confidence Intervals
- Chapter 7: Hypothesis Testing with One Sample
- Chapter 8: Hypothesis Testing with Two Samples
- Chapter 9: Correlation And Regression
- Chapter 10: Chi-Square Tests and The  $F$ -Distribution

الفصل الأول: مقدمة في الإحصاء

**الفصل الثاني: الإحصاء الوصفي**

الفصل الثالث: الاحتمالات

الفصل الرابع: التوزيعات الاحتمالية المنفصلة

الفصل الخامس: التوزيع الاحتمالي الطبيعي

الفصل السادس: فترات الثقة

الفصل السابع: اختبار الفرضيات لعينة واحدة

الفصل الثامن: اختبار الفرضيات لعينتين

الفصل التاسع: الارتباط والانحدار

الفصل العاشر: اختبارات كاي تربيع وتوزيع  $F$

## Descriptive Statistics

- |                                             |                                     |
|---------------------------------------------|-------------------------------------|
| 1- Frequency Distributions and Their Graphs | التوزيعات التكرارية وتمثيلها بيانيا |
| 2- More Graphs and Displays                 | بعض نماذج التمثيل البياني           |
| 3- Measures of Central Tendency             | مقاييس النزعة المركزية              |
| 4- Measures of Variation                    | مقاييس التشتت                       |
| 5- Measures of Position                     | مقاييس الموقع                       |

# Chap **Frequency Distributions and their Graphs**

**1- Frequency Distributions**

التوزيعات التكرارية

**2- Graphs of Frequency Distributions**

تمثيل التوزيعات التكرارية بيانيا

## Chap Frequency Distributions and their Graphs

A **frequency distribution** is a table that shows **classes** or **intervals** of data entries with a count of the number of entries in each class. The **frequency**  $f$  of a class is the number of data entries in the class.

**To Draw the frequency table, determine the three items:**

- 1- number of classes
- 2- Range
- 3- Class Width

التوزيع التكراري هو جدول يوضح فئات أو فترات إدخلات للفئة  $f$  البيانات مع حساب عدد الإدخالات في كل فئة. التردد هو عدد إدخلات البيانات في الفئة.

لرسم الجدول التكراري حدد العناصر الثلاثة:

- 1- عدد الفئات
- 2- المدى
- 3- عرض الفئة

## Example

The following sample data set lists the prices (in dollars) of 30 portable global positioning system (GPS) navigators. Construct a frequency distribution that has seven classes.

**Construct a Frequency  
Distribution from a Data Set**

| 90  | 130 | 200 | 400 | 350 | 70  | 325 | 250 | 150 | 250 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 275 | 270 | 150 | 130 | 59  | 200 | 160 | 450 | 300 | 130 |
| 220 | 100 | 200 | 400 | 200 | 250 | 95  | 180 | 170 | 150 |

تسرد مجموعة البيانات النموذجية  
التالية الأسعار (بالدولار) لـ 30  
ملاحًا محمولًا يعمل بنظام تحديد  
أنشئ (GPS). المواقع العالمي )  
توزيعًا تكراريًا يحتوي على سبع  
فئات.

**بناء التوزيع التكراري من  
مجموعة البيانات**

## Solution of Example

Number of classes = 7

Range = Max – Min = 450 - 59 = 391

$$\text{Class Width} = \frac{\text{Range}}{\text{Number of classes}} = \frac{391}{7} = 55.86 \cong 56$$

| 90  | 130 | 200 | 400 | 350 | 70  | 325 | 250 | 150 | 250 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 275 | 270 | 150 | 130 | 59  | 200 | 160 | 450 | 300 | 130 |
| 220 | 100 | 200 | 400 | 200 | 250 | 95  | 180 | 170 | 150 |

# Frequency Distributions and their Graphs

Frequency Distribution for  
Prices (in dollars) of GPS Navigators

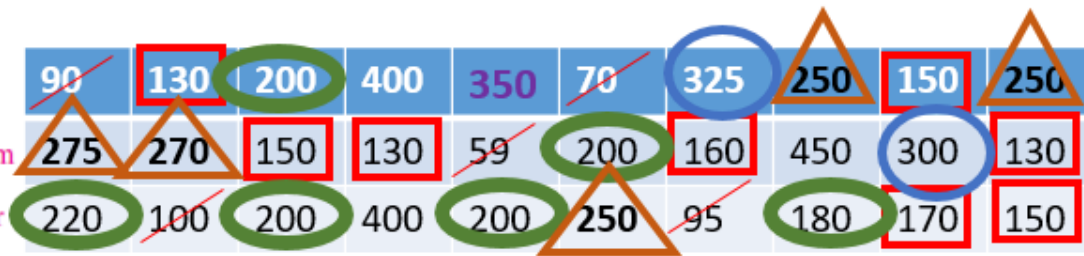
Prices

| Class   | Tally | Frequency, $f$  |
|---------|-------|-----------------|
| 59–114  |       | 5               |
| 115–170 |       | 8               |
| 171–226 |       | 6               |
| 227–282 |       | 5               |
| 283–338 |       | 2               |
| 339–394 |       | 1               |
| 395–450 |       | 3               |
|         |       | $\Sigma f = 30$ |

Number of  
GPS navigators

Check that the sum  
of the frequencies  
equals the number  
in the sample.

Class Width = 56



## Cumulative Frequency Distributions

1- The **midpoint** of a class =  $\frac{\text{Lower class limit} + \text{Upper class Limit}}{2}$

نقطة المنتصف

2- The **relative frequency** of a class =  $\frac{\text{Class Frequency}}{\text{Sample Size}} = \frac{f}{n}$

التكرار النسبي

3- The **cumulative frequency** of a class is the sum of the frequencies of that class and all previous classes. The cumulative frequency of the last class is equal to the sample size  $n$ .

التكرار التراكمي لفئة ما هو مجموع تكرارات تلك الفئة وجميع الفئات السابقة. التكرار التراكمي للفئة الأخيرة يساوي حجم العينة

# Example

Chapter

The following sample data set lists the prices (in dollars) of 30 portable global positioning system (GPS) navigators. Construct a frequency distribution that has seven classes.

**Find the mid point, Relative frequency and a Cumulative Frequency Distribution from a Data Set below**

تسرد مجموعة البيانات النموذجية التالية الأسعار (بالدولار) لـ 30 ملاحًا محمولًا يعمل بنظام تحديد المواقع أنشئ توزيعًا تكراريًا يحتوي على سبع فئات. (GPS العالمي)

ابحث عن نقطة المنتصف والتكرار النسبي والتوزيع التكراري التراكمي من مجموعة البيانات أدناه

| 90  | 130 | 200 | 400 | 350 | 70  | 325 | 250 | 150 | 250 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 275 | 270 | 150 | 130 | 59  | 200 | 160 | 450 | 300 | 130 |
| 220 | 100 | 200 | 400 | 200 | 250 | 95  | 180 | 170 | 150 |

# Solution of Example

Chapter

| Class   | Frequency f | Midpoint                      | Relative Frequency    | Cumulative frequency |
|---------|-------------|-------------------------------|-----------------------|----------------------|
| 59-114  | 5           | $\frac{59 + 114}{2} = 86.5$   | $\frac{5}{30} = 0.17$ | 5                    |
| 115-170 | 8           | $\frac{115 + 170}{2} = 142.5$ | $\frac{8}{30} = 0.27$ | 5+8=13               |
| 171-226 | 6           | 198.5                         | 0.2                   | 5+8+6=19             |
| 227-282 | 5           | 254.5                         | 0.17                  | 5+8+6+5=24           |
| 283-338 | 2           | 310.5                         | 0.07                  | 5+8+6+5+2=26         |
| 339-394 | 1           | 366.5                         | 0.03                  | 5+8+6+5+2+1=27       |
| 395-450 | 3           | 422.5                         | 0.1                   | 5+8+6+5+2+1+3=30     |
| Total   | 30          |                               | 1                     |                      |

# Graphs of Frequency Distributions

**1- Frequency Histogram**

الأعمدة التكرارية

**2- Frequency Polygon**

المضلع التكراري

**3- Relative Frequency Histogram**

الأعمدة التكرارية النسبية

## Frequency Histogram

**Frequency histogram** is a bar graph that represents the frequency distribution of a data set. A histogram has the following properties.

1. The horizontal scale is quantitative and measures the data values.
2. The vertical scale measures the frequencies of the classes.
3. Consecutive bars must touch.

الأعمدة التكرارية هو رسم بياني شريطي يمثل التوزيع التكراري لمجموعة البيانات. يحتوي الرسم البياني على الخصائص التالية.

1. المقياس الأفقي كمي ويقاس قيم البيانات.
2. يقاس المقياس العمودي ترددات الفئات.
3. يجب أن تتلامس الأشرطة المتتالية.

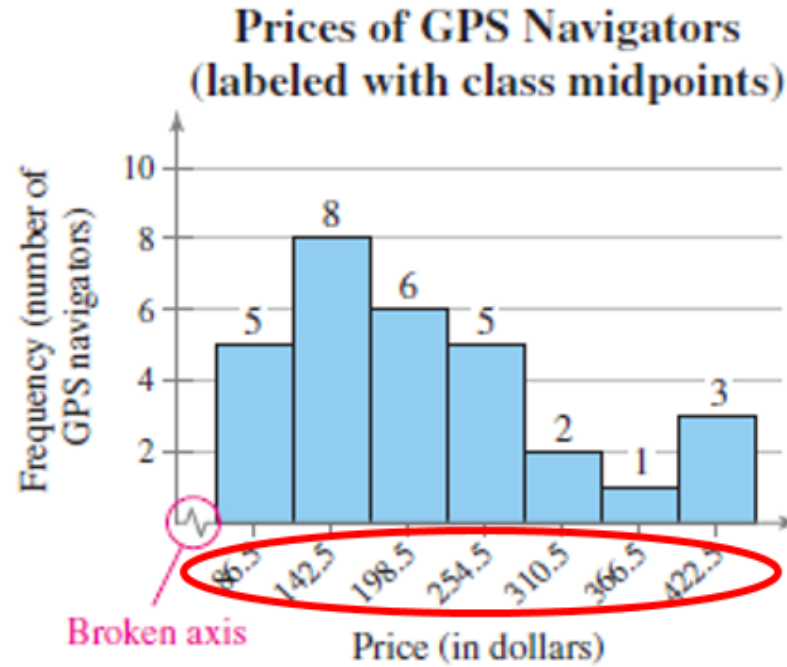
## Example

Draw the frequency Histogram based on the data in the below table:

| Class   | Frequency<br><i>f</i> | Midpoint |
|---------|-----------------------|----------|
| 59-114  | 5                     | 86.5     |
| 115-170 | 8                     | 142.5    |
| 171-226 | 6                     | 198.5    |
| 227-282 | 5                     | 254.5    |
| 283-338 | 2                     | 310.5    |
| 339-394 | 1                     | 366.5    |
| 395-450 | 3                     | 422.5    |
| Total   | 30                    |          |

Chapter

# Frequency Histogram



## Frequency Polygon

A **frequency polygon** is a line graph that emphasizes the continuous change in frequencies.

المضلع التكراري هو رسم بياني خطي يؤكد على التغيير المستمر في التكرارات.

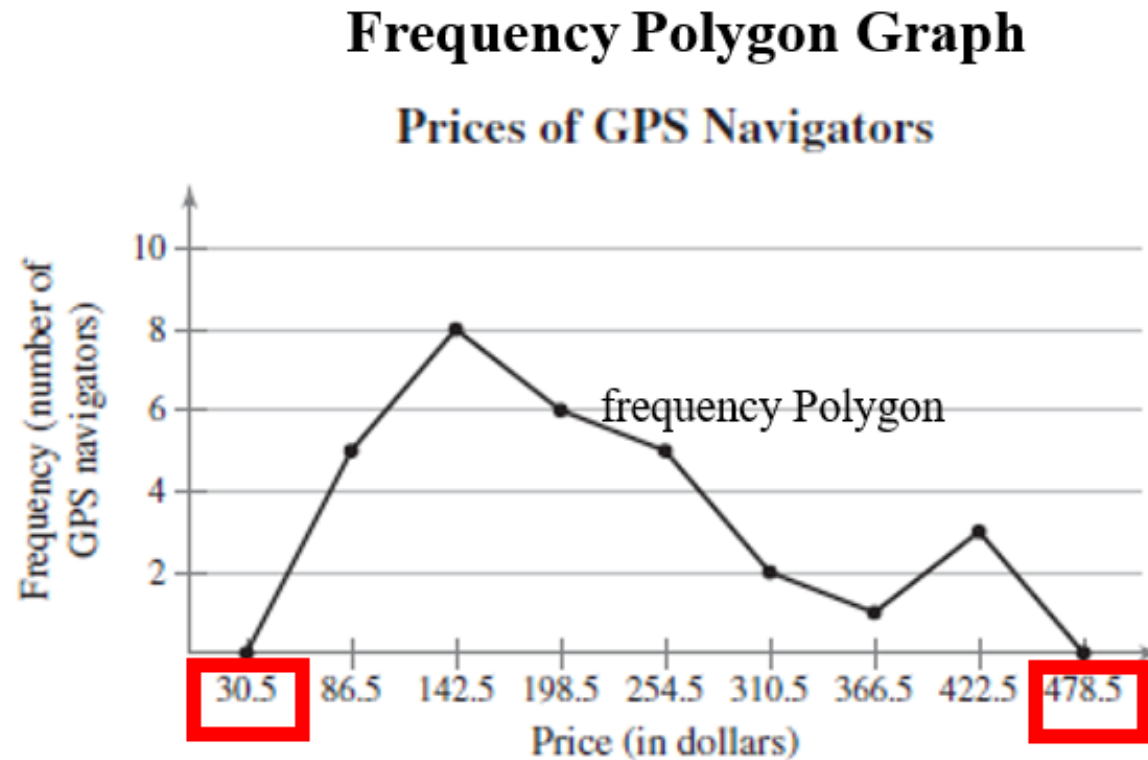
## Example

Draw the frequency Polygon based on the data in this table:

| Class   | Frequency<br>$f$ | Midpoint |
|---------|------------------|----------|
| 59-114  | 5                | 86.5     |
| 115-170 | 8                | 142.5    |
| 171-226 | 6                | 198.5    |
| 227-282 | 5                | 254.5    |
| 283-338 | 2                | 310.5    |
| 339-394 | 1                | 366.5    |
| 395-450 | 3                | 422.5    |
| Total   | 30               |          |

An increase of midpoints by 56 and the same for other points

## Solution of Example



### Midpoint

86.5

142.5

198.5

254.5

310.5

366.5

422.5

## Relative Frequency Histogram

A **relative frequency histogram** has the same shape and the same horizontal scale as the corresponding frequency histogram. The difference is that the vertical scale measures the *relative* frequencies, not frequencies.

الأعمدة البيانية التكرارية النسبية له نفس الشكل ونفس المقياس الأفقي مثل الرسم البياني للتردد المقابل. والفرق هو أن المقياس الرأسي يقيس التكرارات النسبية وليس التكرارات.

# Example

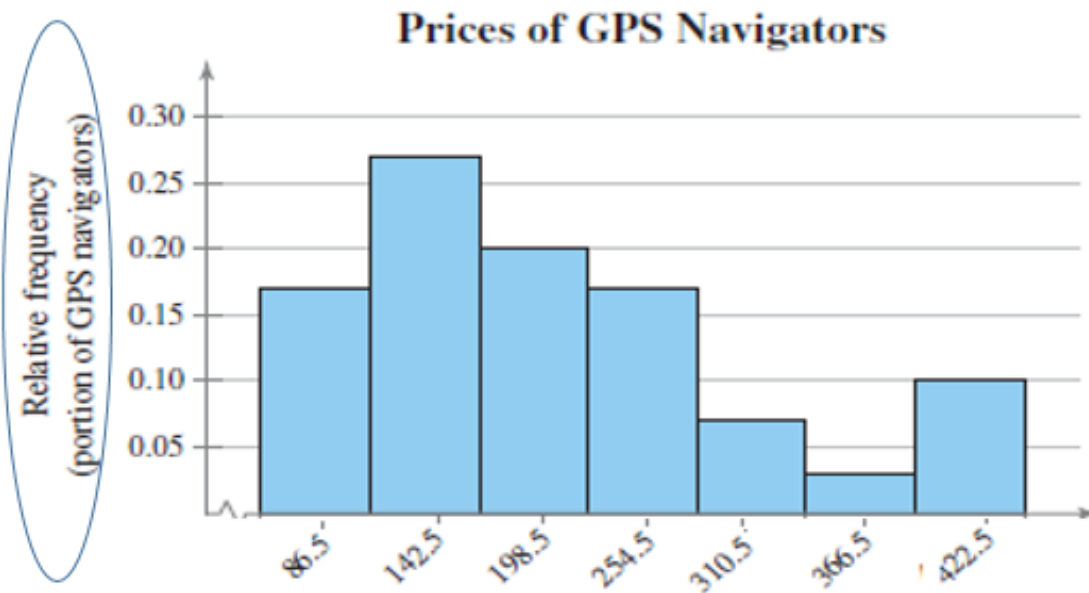
Chapter

Draw the relative frequency Histogram based on the data in this table:

| Class   | Frequency $f$ | Midpoint | Relative Frequency |
|---------|---------------|----------|--------------------|
| 59-114  | 5             | 86.5     | 0.17               |
| 115-170 | 8             | 142.5    | 0.27               |
| 171-226 | 6             | 198.5    | 0.2                |
| 227-282 | 5             | 254.5    | 0.17               |
| 283-338 | 2             | 310.5    | 0.07               |
| 339-394 | 1             | 366.5    | 0.03               |
| 395-450 | 3             | 422.5    | 0.1                |
| Total   | 30            |          | 1                  |

## Solution of Example

### Relative frequency Histogram Graph



# Frequency Distributions and their Graphs

- 1- Frequency Distributions and Their Graphs    التوزيعات التكرارية وتمثيلها بيانيا
- 2- More Graphs and Displays    بعض نماذج التمثيل البياني
- 3- Measures of Central Tendency    مقاييس النزعة المركزية
- 4- Measures of Variation    مقاييس التشتت
- 5- Measures of Position    مقاييس الموقع

## Measures of Central Tendency

We have three measures of Central Tendency:      مقياس النزعة المركزية

1- Mean      المتوسط الحسابي

2- Median      الوسيط

3- Mode      المنوال

## Important Notes

- Mean is not preferred to use as a central measure for the data contained an outlier.
- An **outlier** is a data entry that is far removed from the other entries in the data set.
- **Sample Mean** =  $\bar{X} = \frac{\sum X}{n}$
- **Population Mean** =  $\mu = \frac{\sum X}{N}$

## Example

**Given the following Macroeconomics exam grades for a sample of 6 students:**

|           |           |           |           |           |           |
|-----------|-----------|-----------|-----------|-----------|-----------|
| <b>81</b> | <b>75</b> | <b>66</b> | <b>16</b> | <b>90</b> | <b>90</b> |
|-----------|-----------|-----------|-----------|-----------|-----------|

**Calculate the central tendency measures and state the best central measure for this data.**

## Solution of Example

### 1- Mean:

$$\text{Sample Mean} = \bar{X} = \frac{\sum X}{n} = \frac{81+75+66+16+90+90}{6} = 69.67$$

The mean mark is equal to 69.67 which represents the value at the Centre of the data where the majority of observations are around.

المتوسط الحسابي يساوي 69.67 والتي تمثل القيمة في مركز البيانات حيث توجد غالبية الملاحظات.

## Solution of Example

### 2- Median:

- put data in order: (smallest Largest)

16, 66, 75, 81, 90, 90

- **The value of the median is the value between 75 and 81. So,**

$$\text{value of median: } \frac{75+81}{2} = 78$$

## **Solution of Example**

**3- Mode:**

**The most repeated number in the data. So,**

**Mode = 90**

## Example

**Given the following Macroeconomics exam grades for a sample of 7 students:**

**5      6      9      10      14      11      12**

**Calculate the central tendency measures and state the best central measure for this data.**

## Solution of Example

### 1- Mean:

$$\text{Sample Mean} = \bar{X} = \frac{\sum X}{n} = \frac{5+6+9+10+14+11+12}{7} = 9.57$$

The mean mark is equal to 9.57 which represents the value at the Centre of the data where the majority of observations are around.

## Solution of Example

### 2- Median:

- put data in order: (smallest Largest)

5, 6, 9, 10, 11, 12, 14

- **The value of the median is the value 10**

## Solution of Example

**3- Mode:**            5    6    9    10    14    11    12

**The most repeated number in the data. So,**

**The data has **No mode** عديمة المنوال**

## Example

**Find Mean – Median – Mode for the following data**

**أوجد المتوسط الحسابي – الوسيط – المنوال للبيانات التالية**

**18    29    43    24    31    9    12    19    16    14**

## Solution of Example

**Mean** المتوسط الحسابي

$$= \frac{18 + 29 + 43 + 24 + 31 + 9 + 12 + 19 + 16 + 14}{10} = \frac{215}{10} = 21.5$$

## Solution of Example

**Median**      الوسيط

9   12   14   16   18   19   24   29   31   43

$$\frac{18+19}{2} = 18.5$$

## Solution of Example

**Mode**    المنوال

The data has **No mode**    عديمة المنوال

## Example

**Find Mean – Median – Mode for the following data**

**أوجد المتوسط الحسابي – الوسيط – المنوال للبيانات التالية**

**14 16 9 8 12 15 20 9 10 24 22 10 10**

## Solution of Example

Mean المتوسط الحسابي

14 16 9 8 12 15 20 9 10 24 22 10 10

$$= \frac{14+16+9+8+12+15+20+9+10+24+22+10+10}{13} = \frac{179}{13} = 13.769 \approx 13.77$$

## Solution of Example

**Median**      الوسيط

14   16   9   8   12   15   20   9   10   24   22   10   10

8   9   9   10   10   10   12   14   15   16   20   22   24

Median = 12

## Solution of Example

**Mode**      المنوال

8    9    9    10   10   10   12   14   15   16   20   22   24

Mode = 10

## Example

In the following three groups, it is required to calculate the mean, median, and mode:

Group 1: 2 – 4 – 4 – 6

Group 2: -2 – 4 – 4 - 10

Group 3: -70 – 4 – 4 - 78

## Solution of Example

| Group | Mean | Median | Mode |
|-------|------|--------|------|
| 1     | 4    | 4      | 4    |
| 2     | 4    | 4      | 4    |
| 3     | 4    | 4      | 4    |

## Measures of Variation

**We have three measures of Variations:**

**1- Range**

المدى

**2- Variance**

التباين

**3- Standard deviation**

الانحراف المعياري

## Chapter 10 Important Notes

- **Population Variance** =  $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$
- **Sample Variance** =  $S^2 = \frac{\sum(X-\bar{X})^2}{n-1}$
- **Population Standard deviation** =  $\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$
- **Sample Standard deviation** =  $S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

**Example**

Given the following Mathematics exam grades for a sample of 5 students.

8      12      20      7      3

Calculate the dispersion measures for the above data

## Solution of Example

1- Range = Max – Min = 20- 3 = 17

2- Sample Variance =  $S^2 = \frac{\sum(X-\bar{X})^2}{n-1}$

| $X$      | $\bar{X}$ | $(X - \bar{X})$ | $(X - \bar{X})^2$ |
|----------|-----------|-----------------|-------------------|
| 8        | 10        | -2              | 4                 |
| 12       | 10        | 2               | 4                 |
| 20       | 10        | 10              | 100               |
| 7        | 10        | -3              | 9                 |
| 3        | 10        | -7              | 49                |
| $\Sigma$ |           | zero            | 166               |

## Solution of Example

1- Range = Max – Min = 20- 3 = 17

2- Sample Variance =  $S^2 = \frac{\sum(X-\bar{X})^2}{n-1}$

| $X$      | $\bar{X}$ | $(X - \bar{X})$ | $(X - \bar{X})^2$ |
|----------|-----------|-----------------|-------------------|
| 8        | 10        | -2              | 4                 |
| 12       | 10        | 2               | 4                 |
| 20       | 10        | 10              | 100               |
| 7        | 10        | -3              | 9                 |
| 3        | 10        | -7              | 49                |
| $\Sigma$ |           | zero            | 166               |

3- Sample Standard deviation =  $S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{166}{4}} = 6.44$

**Example**

Given the following Macroeconomics exam grades for a sample of 5 students.

81

75

66

84

96

Calculate the dispersion measures for the above data

## Solution of Example

1- Range = Max – Min = 96- 66 = 30

2- Sample Variance =  $S^2 = \frac{\sum(X-\bar{X})^2}{n-1}$

| x                          | $\bar{x}$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ |
|----------------------------|-----------|---------------|-----------------|
| 81                         | 80.4      | 0.6           | 0.36            |
| 75                         | 80.4      | -5.4          | 29.16           |
| 66                         | 80.4      | -14.4         | 207.36          |
| 84                         | 80.4      | 3.6           | 12.96           |
| 96                         | 80.4      | 15.6          | 243.36          |
| <b><math>\Sigma</math></b> |           | <b>Zero</b>   | <b>493.2</b>    |

## Solution of Example

1- Range = Max – Min = 96- 66 = 30

2- Sample Variance =  $S^2 = \frac{\sum(X-\bar{X})^2}{n-1} = \frac{493.2}{5-1} = 123.3$

| x        | $\bar{x}$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ |
|----------|-----------|---------------|-----------------|
| 81       | 80.4      | 0.6           | 0.36            |
| 75       | 80.4      | -5.4          | 29.16           |
| 66       | 80.4      | -14.4         | 207.36          |
| 84       | 80.4      | 3.6           | 12.96           |
| 96       | 80.4      | 15.6          | 243.36          |
| $\Sigma$ |           | <b>Zero</b>   | <b>493.2</b>    |

3- Sample Standard deviation =  $S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{123.3} = 11.1$

## Example

**Find Mean – Median – Mode and Range for the following data**

**أوجد المتوسط الحسابي – الوسيط – المنوال والمدى للبيانات التالية**

**14 16 9 8 12 15 20 9 10 24 22 10 10 9 13**

## Solution of Example

Mean المتوسط الحسابي

14 16 9 8 12 15 20 9 10 24 22 10 10 9 13

$$= \frac{14+16+9+8+12+15+20+9+10+24+22+10+10+9+13}{15} = \frac{201}{15} = 13.4$$

## Solution of Example

Median الوسيط

14 16 9 8 12 15 20 9 10 24 22 10 10 9 13

8 9 9 9 10 10 10 12 13 14 15 16 20 22 24

Median = 12

## Solution of Example

Mode      المنوال

8   9   9   9   10   10   10   12   13   14   15   16   20   22   24

Mode = 9 ,10

## Solution of Example

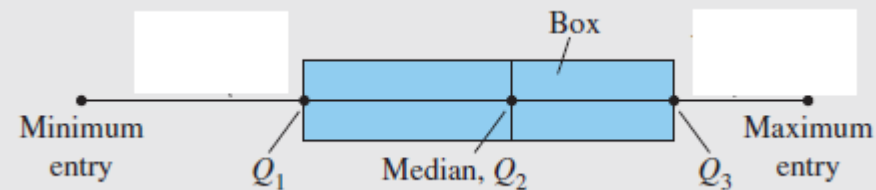
**Range**      المدى

8   9   9   9   10   10   10   12   13   14   15   16   20   22   24

$$\text{Range} = 24 - 8 = 16$$

## Measures of Position

- Q1 is the first Quartile الربع الأول
- Q2 is the Second Quartile ( Median ) الربع الثاني (الوسيط)
- Q3 is the third Quartile الربع الثالث
- Interquartile range  $IQR = Q3 - Q1$  المدى الربيعي
- Box Plot

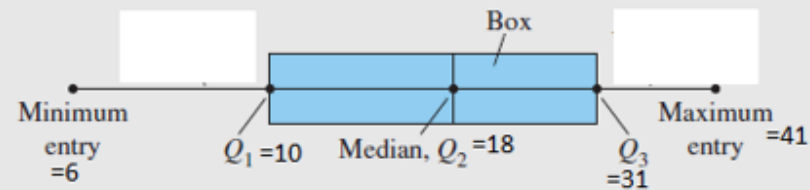


## Example

The number of nuclear power plants in the top 15 nuclear power-producing countries in the world are listed.

7 18 11 6 59 17 18 54 104 20 31 8 10 15 19

- Find the first, second, and third quartiles of the data set.
- Calculate IQR
- Draw the Box plot



## Example

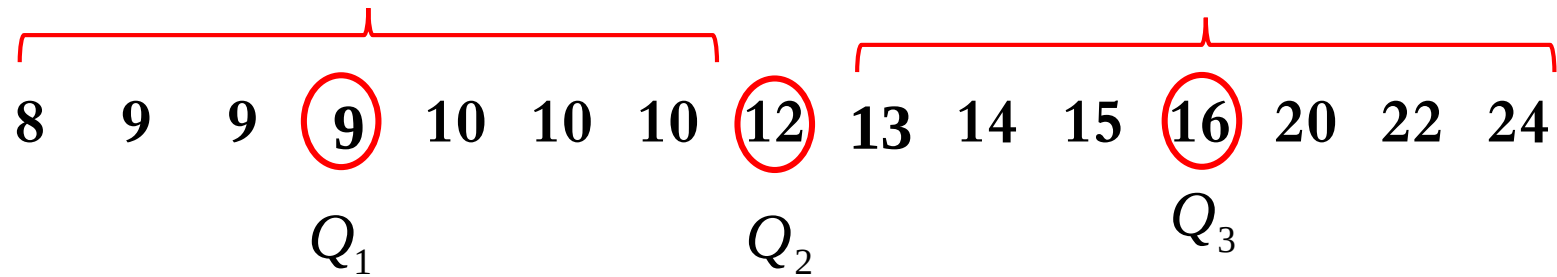
For the following data.

**14 16 9 8 12 15 20 9 10 24 22 10 10 9 13**

- Find the first, second, and third quartiles of the data set.
- Calculate IQR

## Solution of Example

- Put data in order: (smallest Largest)



$$\text{IQR} = Q_3 - Q_1 = 16 - 9 = 7$$

# Chapter Summary

## Frequency Distributions

### Measures of Central Tendency:

- 1- Mean
- 2- Median
- 3- Mode

### Measures of Position:

- 1- First Quartile
- 2- Second Quartile
- 3- Third Quartile
- 4- IQR
- 5- Box Plot

### Graphs and Displays:

- 1- Frequency Histogram
- 2- Frequency Polygon
- 3- Relative Frequency Histogram

### Measures of Variation:

- 1- Range
- 2- Variance
- 3- Standard deviation

# Un Solved Examples

Chapter

**Given the following data**

**11                      25                      30                      24                      22                      10                      12**

**Find Mean – Median – Mode - Range**

# Un Solved Examples

**Given the following data**

**11                      25                      30                      24                      22      10                      12**

**Find Mean – Median – Mode - Range**

$$\text{Mean} = \frac{11+25+30+24+22+10+12}{7}$$

$$\text{Mean} = 19.14$$

# Un Solved Examples

**Given the following data**

**11                      25                      30                      24                      22      10                      12**

**Find Mean – Median – Mode - Range**

Median = 10 – 11 – 12 – **22** – 24 – 25 – 30

Median = 22

# Un Solved Examples

Chapter

**Given the following data**

**11                      25                      30                      24                      22                      10                      12**

**Find Mean – Median – Mode - Range**

**Mode = Has No Mode**

# Un Solved Examples

**Given the following data**

**11                      25                      30                      24                      22      10                      12**

**Find Mean – Median – Mode - Range**

$$\text{Range} = 30 - 10$$

$$\text{Range} = 20$$

# Un Solved Examples

Chapter

The value of Q1 for the values [12,10,33,49,18,90,100]

# Un Solved Examples

Chapter

The value of Q1 for the values [12,10,33,49,18,90,100]

Lower Half                      Upper Half

$\overbrace{10-12-18} - 33 - \overbrace{49-90-100}$

$$Q_1 = 12$$

# Un Solved Examples

Chapter

The value of Q3 for the values [12,10,33,49,18,90,100]

# Un Solved Examples

Chapter

The value of Q3 for the values [12,10,33,49,18,90,100]

Lower Half                      Upper Half

$\overbrace{10-12-18} \quad - \quad 33 - \quad \overbrace{49-90-100}$

$$Q_3 = 90$$

# Un Solved Examples

Chapter

The value of Q2 for the values [12,10,33,49,18,90,100]

# Un Solved Examples

Chapter

The value of Q2 for the values [12,10,33,49,18,90,100]

Lower Half                      Upper Half

$\overbrace{10-12-18} - 33 - \overbrace{49-90-100}$

$$Q_2 = 33$$

# Un Solved Examples

Chapter

The value of IQR for the values [12,10,33,49,18,90,100]

# Un Solved Examples

Chapter

The value of IQR for the values [12,10,33,49,18,90,100]

$$\begin{aligned}
 IQR &= Q_3 - Q_1 \\
 &= 90 - 12 \\
 &= 78
 \end{aligned}$$



# Thank You