

Probability and Statistics

Prepared by Dr. Israa Lewaa

Lecture 2

جامعة الجميع الذكية

STA 201

Descriptive Statistics



جامعة الجميع الذكية
EVERYONE'S SMART UNIVERSITY

1- Frequency Distributions and Their Graphs

2- More Graphs and Displays

3- Measures of Central Tendency

4- Measures of Variation

5- Measures of Position

1- Frequency Distributions

2- Graphs of Frequency Distributions

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

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

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.

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

Construct a Frequency Distribution from a Data Set

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

Chapter 02

Frequency Distributions and their Graphs

Class Width = 56

Frequency Distribution for
Prices (in dollars) of GPS Navigators

Prices

Number of
GPS navigators

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
		$\Sigma f = 30$

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

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

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

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

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 .

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.

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

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

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	

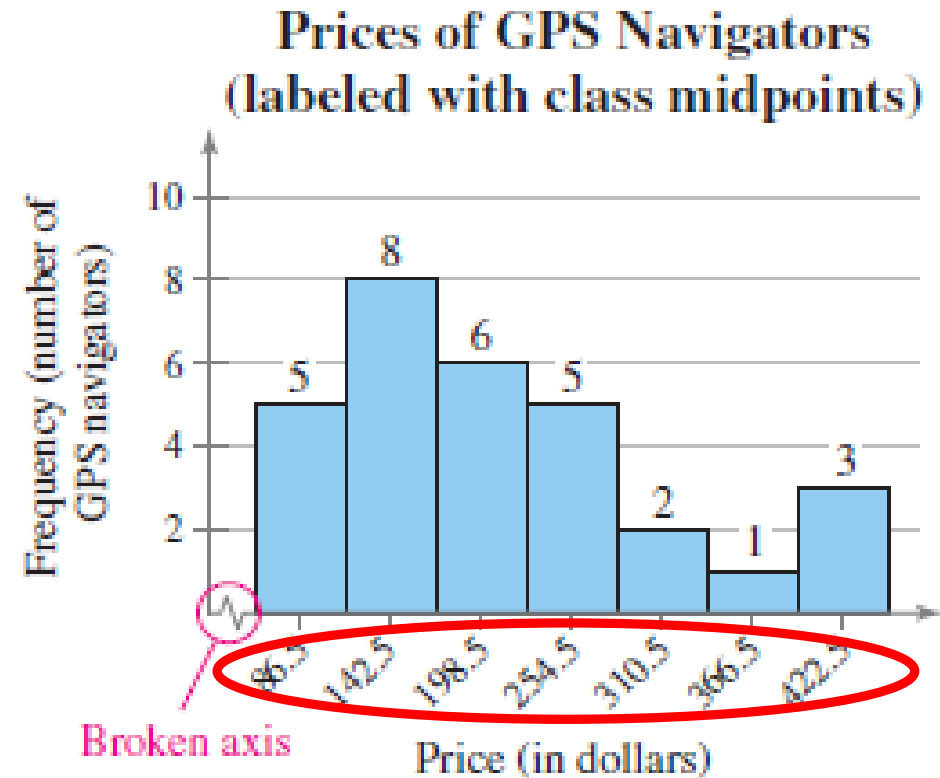
- 1- Frequency Histogram**
- 2- Frequency Polygon**
- 3- Relative Frequency Histogram**
- 4- Ogive**

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.

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

Class	Frequency <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	



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

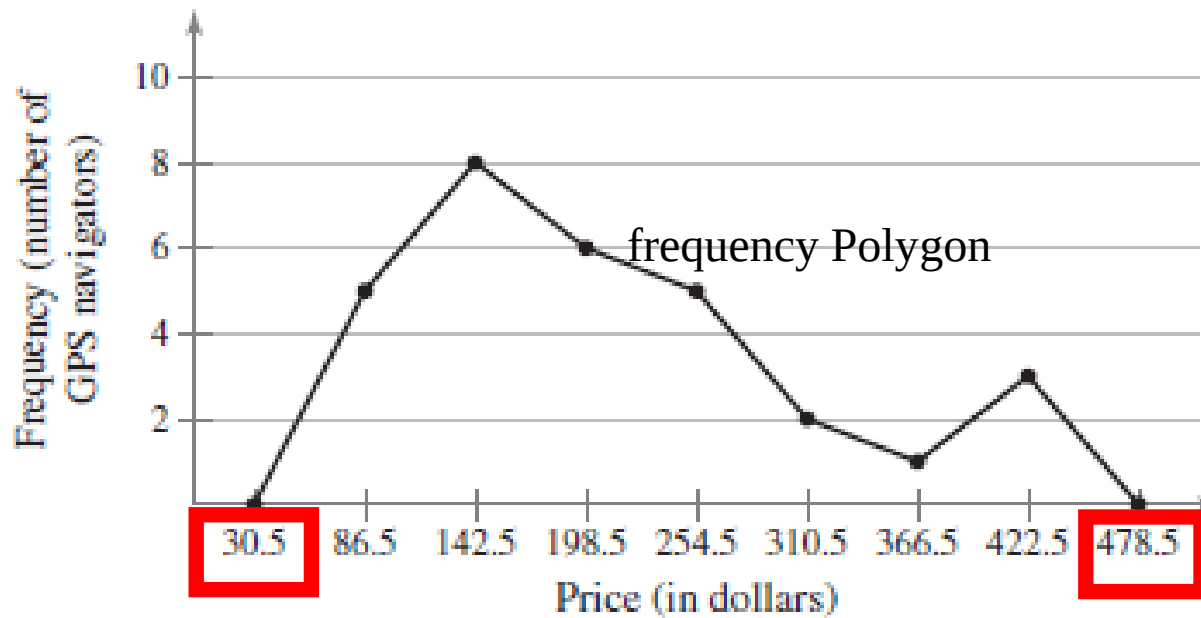
Draw the frequency Polygon based on the data in this table:

Class	Frequency 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

Frequency Polygon Graph

Prices of GPS Navigators



Midpoint

86.5

142.5

198.5

254.5

310.5

366.5

422.5

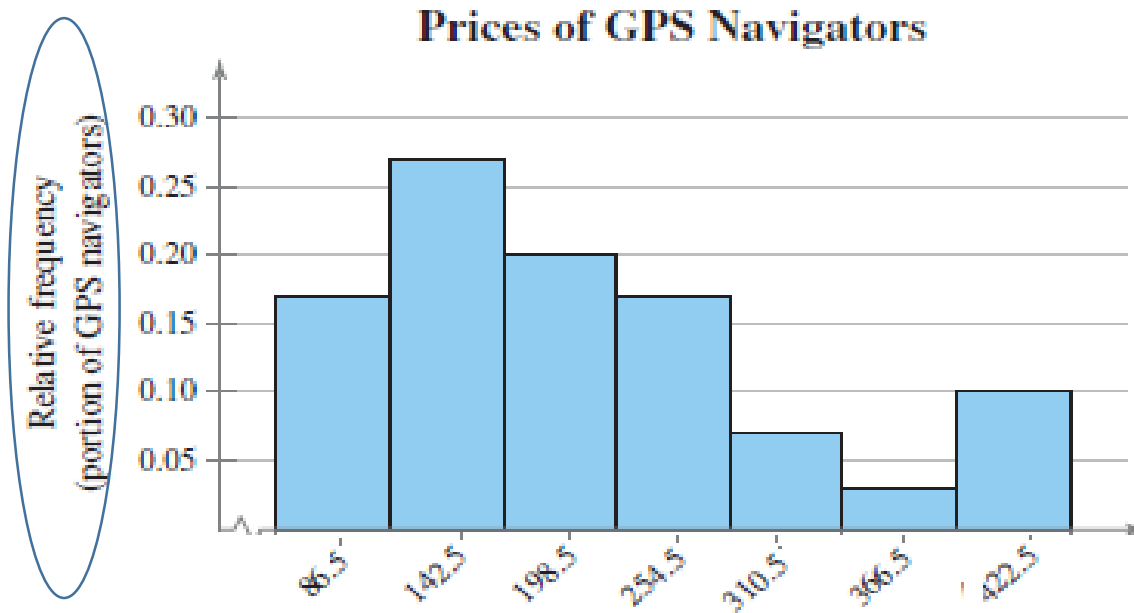


A **relative frequency histogram** has the same shape and the same horizontal scale as the corresponding frequency histogram. The difference is that the vertical scale measures the *relative* frequencies, not frequencies.

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

Relative frequency Histogram Graph

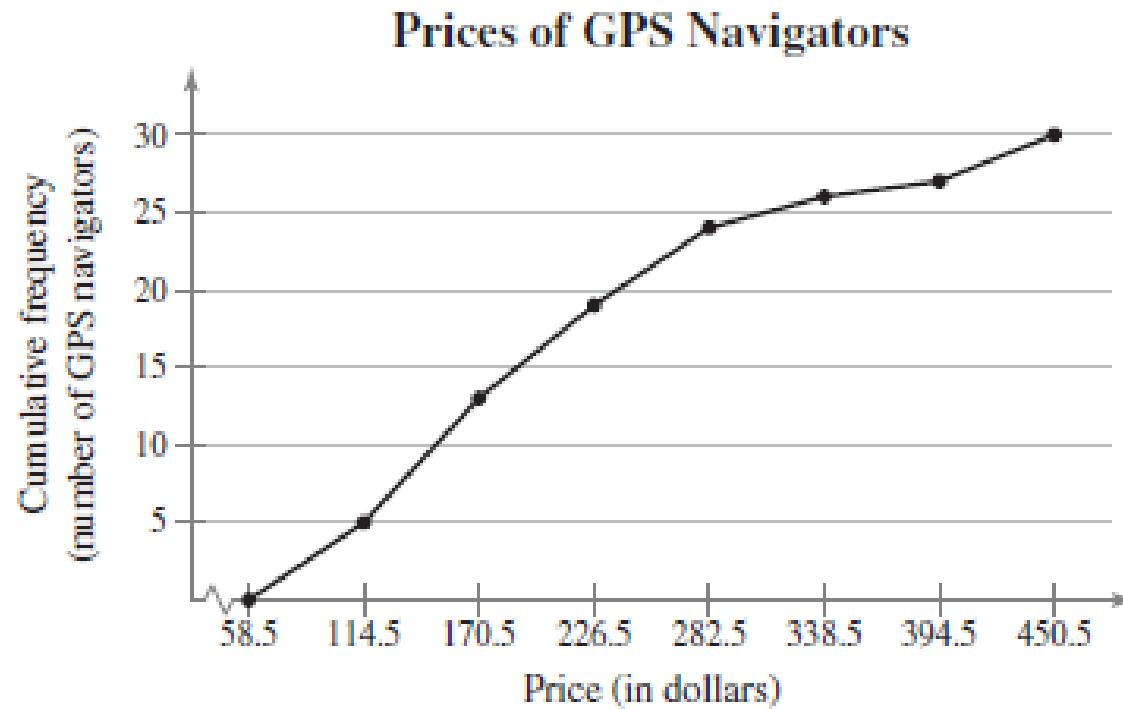


1. Construct a cumulative frequency distribution.
2. Specify the horizontal and vertical scales. The horizontal scale consists of upper class boundaries, and the vertical scale measures cumulative frequencies.
3. Plot points then the graph.
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).

Draw Ogive based on the data in the below table:

Class	Frequency f	Upper class boundaries	Cumulative frequency
59-114	5	$59 - 0.5 = 58.5$	Zero
115-170	8	$115 - 0.5 = 114.5$	5
171-226	6	$171 - 0.5 = 170.5$	13
227-282	5	$227 - 0.5 = 226.5$	19
283-338	2	$283 - 0.5 = 282.5$	24
339-394	1	$339 - 0.5 = 338.5$	26
395-450	3	$395 - 0.5 = 394.5$	27
450 - 506		$450 - 0.5 = 449.5$	30

Ogive Graph



1- Frequency Distributions and Their Graphs

2- More Graphs and Displays

3- Measures of Central Tendency

4- Measures of Variation

5- Measures of Position



We have three measures of Central Tendency:

1- Mean

2- Median

3- Mode

- 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 = $\mu = \frac{\sum X}{N}$

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

81 75 66 16 90 90

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

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.

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,**

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

3- Mode:

The most repeated number in the data. So,

Mode = 90

Since there is an outlier which is (16). Therefore, the best central measure is the median NOT the mean.

We have three measures of Variations:

1- Range

2- Variance

3- Standard deviation

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

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

81

75

66

84

96

Calculate the dispersion measures for the above data

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
Σ		Zero	493.2

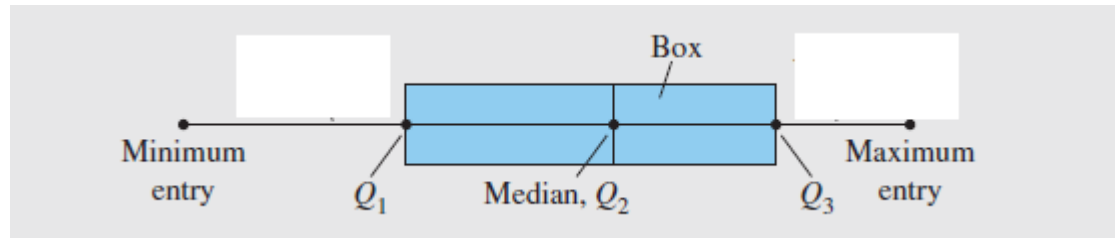
1- Range = Max – Min = 96- 66 = 30

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
Σ		Zero	493.2

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

- Q1 is the first Quartile
- Q2 is the Second Quartile (Median)
- Q3 is the third Quartile
- Interquartile range $IQR = Q3 - Q1$
- Box Plot



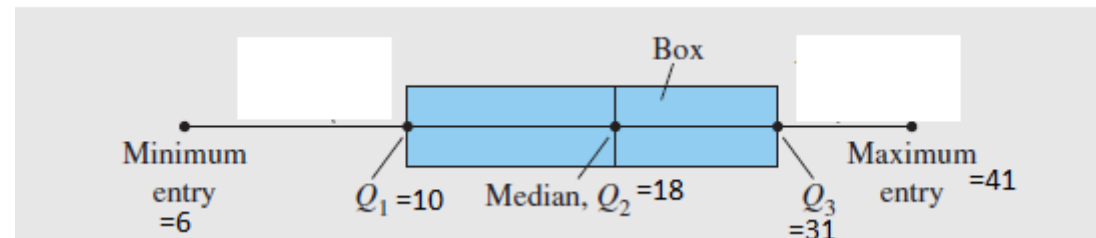
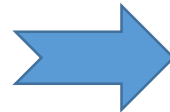
The number of nuclear power plants in the top 15 nuclear power-producing countries in the world are listed.

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

- [illegible]

- **Box Plot**



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
- 4- Ogive

Measures of Variation:

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