

Descriptive Statistics

Graphical methods: bar graph, pie chart, box plot, line plot,...

Measures of centre and variability.

Instructor: Qasim Ali

Statistical Methods

Descriptive Statistics

- The methods that are used to ***describe*** the important characteristics of an available set of data.
- ***Graphical techniques*** (bar graph, pie chart, box plot, Histograms,...) present data.
- ***Numerical techniques / measures*** (mean, median, mode, variance,...) to summarize data.

Inferential Statistics

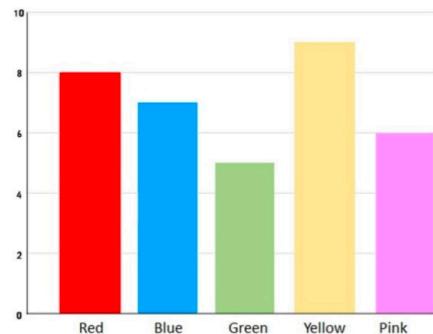
- The methods that use ***sample*** data to make ***inferences*** (or generalizations) about a population.
- ***Numerical techniques*** (regression, correlation, hypothesis testing,...) to estimate, predict or test data.
- ***Measures*** (confidence interval, level of significance, standard error,...) are used to verify the conclusions.

- **Descriptive statistics** deals with methods of organizing, summarizing, and presenting data in a convenient and informative way.
- **Inferential statistics** allow you to estimate, predict or test a hypothesis to assess whether your data is generalizable to the broader population.

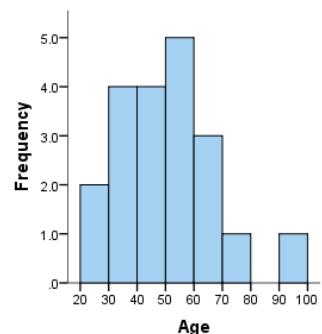
Descriptive Statistics

Graphical Methods

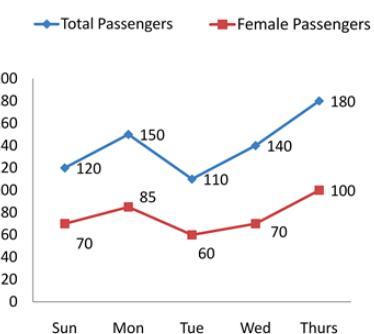
Bar Chart



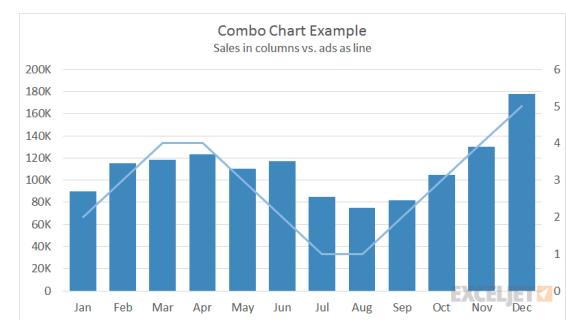
Histograms



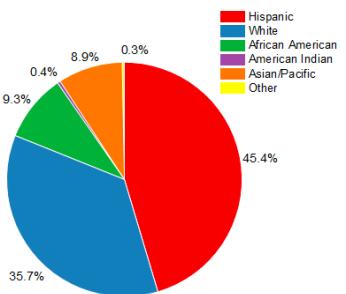
Line plot



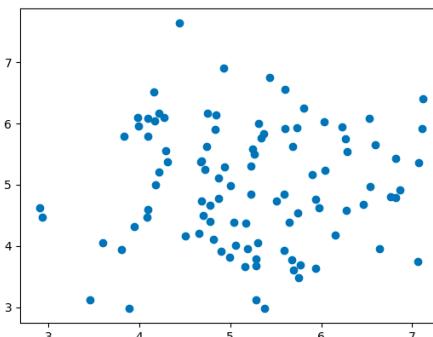
Combo Charts



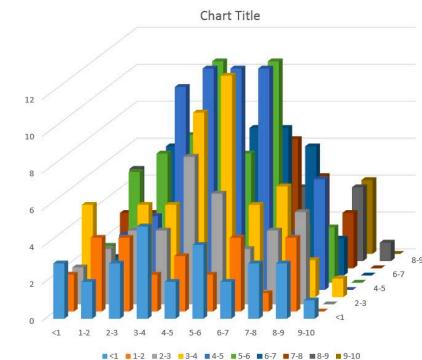
Pie Chart



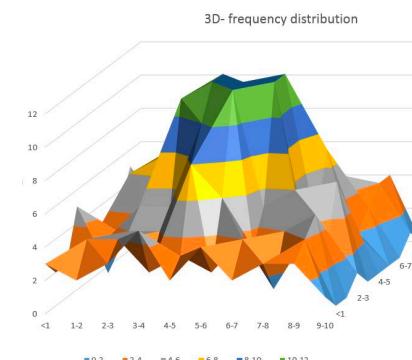
Scatter and bubble plot



3D Plot



Surface plot



Graphical Methods

Craft engaging charts and graphs in Excel

Learning Objectives

- Identifying chart elements and chart types
- Creating a basic chart
- Choosing a chart layout and style
- Working with column, bar, and line charts

Keyboard Shortcuts

<https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f>

Graphical Methods

Craft engaging charts and graphs in Excel

Learning Objectives

- Identifying chart elements and chart types
- Creating a basic chart
- Choosing a chart layout and style
- Working with column, bar, and line charts

Graphical Methods

Craft engaging charts and graphs in Excel

Identifying chart elements and chart types

- Identifying chart elements
- Select the right chart type
- Understand Chart subtypes
- Understand the Chart Tools, Ribbon, Design and Format Tab

Graphical Methods

Practice Problems

Answer the following questions:

- What is the difference between Bar Chart and Histogram?
- Do we have same options in Chart Elements for each type of chart?
- What are the most common charts in Excel?

Answer: Histogram is used to plot quantitative data in bar charts plots qualitative data

Answer: No, charts can have less or more options in chart elements list.

Answer: Column chart, Line chart, Pie chart and Bar chart

Graphical Methods

Craft engaging charts and graphs in Excel

Learning Objectives

- Identifying chart elements and chart types
- **Creating a basic chart**
- Choosing a chart layout and style
- Working with column, bar, and line charts

Graphical Methods

Craft engaging charts and graphs in Excel

Creating basic Charts

- Select data to display as a chart
- Create charts and use basic tools
- Develop charts with the Recommended Charts feature
- Build charts with standard menu commands

Graphical Methods

Practice Problems

Answer the following questions:

- Where can we find the charts in Excel? How many different charts and graphs are present?
- How can we see the previews of charts before inserting them in our excel sheet?
- How the excel show us the recommended charts, either based on our dataset or the most commonly used charts by other people?

Answer: Charts are placed under Insert Tab and around 58 charts are available.

Answer: Recommended charts option allow us to see the previews.

Answer: Based on our dataset

Graphical Methods

Craft engaging charts and graphs in Excel

Learning Objectives

- Identifying chart elements and chart types
- Creating a basic chart
- **Choosing a chart layout and style**
- Working with column, bar, and line charts

Graphical Methods

Craft engaging charts and graphs in Excel

Choosing a chart appearances

- Choose a chart layout
- Select a chart style from a gallery of colorful variations
- Switch rows and columns for a different view of the data
- Change a chart's type
- Move and resize a chart within a worksheet

Graphical Methods

Practice Problems

Answer the following questions:

- What is the purpose of quick chart layout and where can we find it?
- Can we change chart type?
- Can we move the chart to new worksheet?

Answer: Quick Chart layout are predefined layouts of the chart type that is plotted. You can find it in Design menu.

Answer: Yes, it is available in Chart Design tab.

Answer: Yes, right click on the chart and click on Move Chart option.

Graphical Methods

Craft engaging charts and graphs in Excel

Learning Objectives

- Identifying chart elements and chart types
- Creating a basic chart
- Choosing a chart layout and style
- **Working with column, bar, and line charts**

Graphical Methods

Craft engaging charts and graphs in Excel

Choosing a chart layout and style

- Use column and bar charts
- Work with line charts

Measures of Central Location

Descriptive Statistics

Measures of Central Location

Which Group is Smarter?

Class A – IQs of 13 Students

102	115
128	109
131	89
98	106
140	119
93	97
110	

Class B – IQs of 13 Students

127	162
131	103
96	111
80	109
93	87
120	105
109	

Each individual may be different. If you try to understand a group by remembering the qualities of each member, you become overwhelmed and fail to understand the group.

Descriptive Statistics

Measures of Central Location

Test Scores

96	69
95	64
90	56
89	51
83	47
82	46
81	42
80	49
80	37
79	34
78	28
77	25
76	

Another dataset of test scores of 25 students

Mean:

Median:

Mode:

Descriptive Statistics

Measures of Central Location

Mean: Simple average = the sum of all the data points divided by the total number of observations

1. Sum all the data points
2. Divide the data points by the total number of data points

Median: The middle value of the data.

1. Place data in sequence
2. Find the value that is at the middle of the data

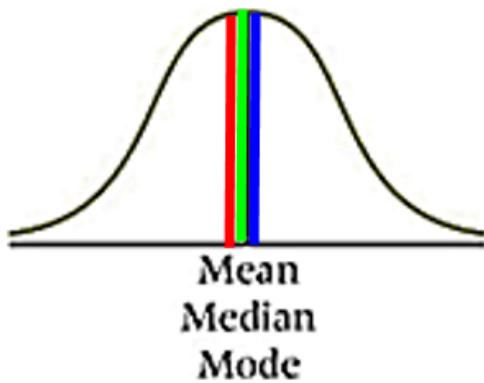
Mode: Most frequent value in the data (Data must have repeated values)

1. Count the occurrence of each value in the data
2. Mode is the value that has maximum number of occurrence in the data

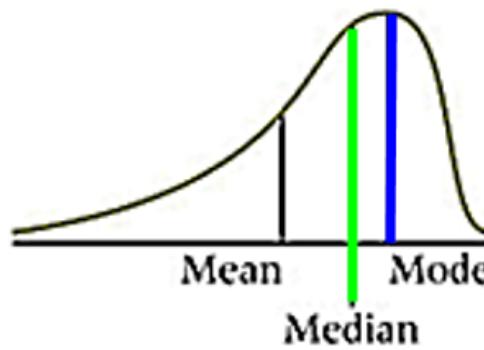
Descriptive Statistics

Measures of Central Location

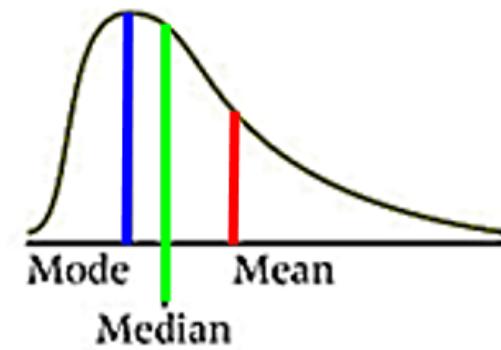
The following picture shows how relation of mean, median, and mode affects skewness of the distribution.



(a)
Symmetric distribution
(no skewness)



(b)
Negatively Skewed
Or
Skewed Left



(c)
Positively Skewed
Or
Skewed Right

Descriptive Statistics

Measures of Central Location

- When analyzing a graphical display, you can draw conclusions based on several characteristics of the graph.
- **You may ask questions such as:**
 - Where is the approximate middle, or center, of the graph?
 - How spread out are the data values on the graph?
 - What is the overall shape of the graph?
 - Does it have any interesting patterns?

Descriptive Statistics

Measures of Central Location

The following data represent duration of hospital stay for a sample 16 patients from the Crossways Hospital.

2	11	5	12	15	28	7	12
10	3	7	7	5	15	3	10

Find the sample mean, median, and mode. Round to one decimal place.

Solution: The sample mean is equal to

$$x = (2+11+5+12+\dots+10)/16 = 152/16 = 9.5 \text{ days.}$$

To find the median, we first arrange the days in order (from smallest to largest):

2	3	3	5	5	7	7	7	10	10	11	12	12	15	15	28
								↑ $x_8 = 7$	↑ $x_9 = 10$						

$$\text{median} = x = (x_8 + x_9)/2 = (7+10)/2 = 8.5$$

$$\text{mode} = M = 7 \text{ (occurs three times)}$$

Measures of Variability

Descriptive Statistics

Measures of Variability

Dataset 1:
110
70 30
90 50

Dataset 2:
70
70 65
70 75

Mean: 70

Median: 70

Mode: 70

A **measure of variability** is a summary statistic that represents the amount of dispersion in a dataset.

Descriptive Statistics

Measures of Variability

Range:

difference between the largest and smallest observations in a sample

$$\begin{aligned}\text{Range} &= \text{Max value} - \text{Min value} \\ &= 70 - 20 = 50\end{aligned}$$

20	60	51	62	57
45	55	70	48	51

Mean: 51.9, Median: 53, Mode: 51,

$$\begin{aligned}\text{Range} &= \text{Max value} - \text{Min value} \\ &= 185 - 35 = 150\end{aligned}$$

75	35	66	127	96
60	105	63	185	66

Mean: 87.8, Median: 70.5, Mode: 66,

Descriptive Statistics

Measures of Variability

Range:

difference between the largest and smallest observations in a sample

Mean	60
Median	58
Range	70

Descriptive Statistics

Measures of Variability

Range:

difference between the largest and smallest observations in a sample

Mean	60
Median	58
Range	70

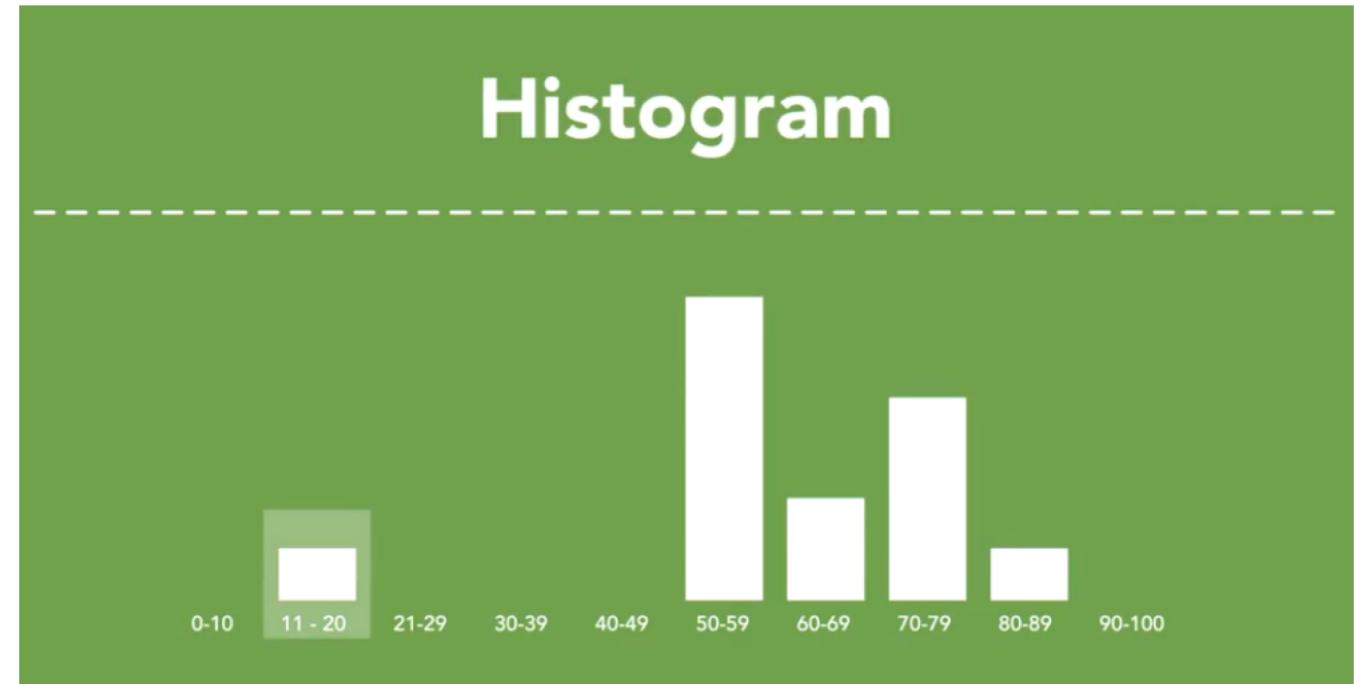
Descriptive Statistics

Measures of Variability

Range: difference between the largest and smallest observations in a sample

15	58	85	70	65
65	55	55	55	75
50	70	55	70	55

Mean	
Median	
Range	70



Descriptive Statistics

Measures of Variability

Standard Deviation:

The average distance of the data points from their own mean

A low standard deviation indicates that the data points are clustered around the mean.

A large standard deviation indicates that they are widely scattered around the mean

The standard deviation of a sample is denoted by 's' and a population is denoted by 'u'

$$\text{Sample variance } s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$\text{Population Variance } \sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

$$\text{Sample standard deviation } s = \sqrt{s^2}$$

$$\text{Population standard deviation } \sigma = \sqrt{\sigma^2}$$

The **sample coefficient of variation (CV)** is defined by:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

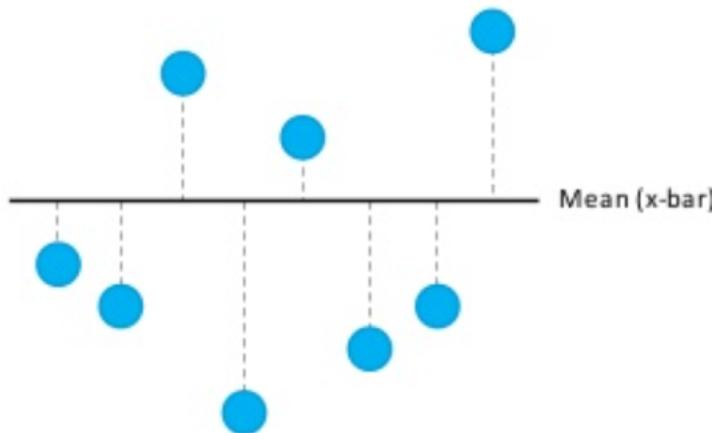
Descriptive Statistics

Measures of Variability

Standard Deviation:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

s = standard deviation
 \bar{x} = mean
 x = values of the data set
 n = size of the data set



It is a more robust measure of variability than finding the range

Descriptive Statistics

Measures of Variability

Test Scores

96	69
95	64
90	56
89	51
83	47
82	46
81	42
80	49
80	37
79	34
78	28
77	25
76	

Another dataset of test scores of 25 students

Mean:

Median:

Mode:

Standard deviation:

Descriptive Statistics

Standard Deviation

Example: Consider a sample of ten technical measurements ranked in order from the lowest to the highest:

6.8 7.4 7.9 8.2 8.3 8.3 8.4 8.8 9.1 9.8

Find the sample range, variance, standard deviation, and coefficient of variation (CV). Round to two decimal places.

Solution: Range = $x_{max} - x_{min} = 9.8 - 6.8 = 3.00$

To calculate the sample variance/standard deviation, we first calculate the sample mean:

$$x = (6.8 + 7.4 + \dots + 9.8) / 10 = 83.0 / 10 = 8.30$$

Then we have to use the formula:

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

Note that in this case $n=10$.

Thus, for each x value we calculate the difference between x value and the sample mean, then square difference (see table below). In the final step, we calculate the sum of all squared differences.

Descriptive Statistics

Standard Deviation

Thus, for each x value we calculate the difference between x value and the sample mean, then square difference (see table below). In the final step, we calculate the sum of all squared differences.

Sample variance:

$$s^2 = \frac{\sum(x - \bar{x})^2}{n-1} = \frac{6.38}{10-1} = \frac{6.38}{9} = 0.708889 \approx 0.71$$

Sample standard deviation

$$s = \sqrt{s^2} = \sqrt{0.71} = 0.842615 \approx 0.84$$

$$CV = \frac{s}{\bar{x}} \times 100\% = \frac{0.84}{8.30} \times 100\% = 0.101205 \times 100\% \\ = 10.12\%$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
6.8	-1.5	2.25
7.4	-0.9	0.81
7.9	-0.4	0.16
8.2	-0.1	0.01
8.3	0	0
8.3	0	0
8.4	0.1	0.01
8.8	0.5	0.25
9.1	0.8	0.64
9.8	1.5	2.25
Total		$\sum(x - \bar{x})^2 = 6.38$

Descriptive Statistics

Summary and Conclusion

- ✓ Graphical Methods
 - ✓ Identifying chart elements and chart types
 - ✓ Creating a basic chart
 - ✓ Choosing a chart layout and style
 - ✓ Working with column, bar, and line charts
- ✓ Measure of Central Tendency (or Groups' "Middle Values")
 - ✓ Mean
 - ✓ Median
 - ✓ Mode
- ✓ Variation (or Measure of Variability)
 - ✓ Range
 - ✓ Variance
 - ✓ Standard Deviation