Business Management Toolkit

7. DESCRIPTIVE STATISTICS

7. Descriptive statistics

What are they?

- Descriptive statistics summarizes or describes the characteristics of a data set.

- Descriptive statistics helps facilitate data visualization. It allows for data to be presented in a meaningful and understandable way, which, in turn, allows for a simplified interpretation of the data set in question.

7. Descriptive statistics

These are the ones we are going to learn:

- Mean
- Mode
- Median
- Bar charts
- Pie charts
- Infographics
- Quartiles
- Standard deviation

Central tendency

Frequency distribution

Variability

Watch the video:
https://www.investopedia.com/terms/d/descriptive_statistics.asp

7. Descriptive statistics - 3 basic categories

1. **Frequency Distribution:**

Frequency distribution is basically a presentation or summary of grouped data categorized based on mutually exclusive classes and the number of occurrences in each respective class. It allows for a more structured and organized way to present raw data.

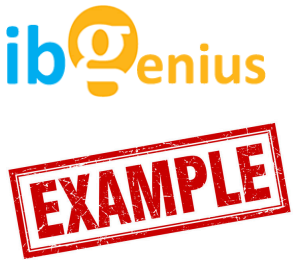
Common charts and graphs used in frequency distribution presentation and visualization include bar charts, histograms, pie charts, and line charts.

7. Frequency Distribution: Bar charts

A bar chart is a graph with rectangular bars. The graph usually compares different categories. Although the graphs can be plotted vertically (bars standing up) or horizontally (bars laying flat from left to right), the most usual type of bar graph is vertical.

A bar graph is useful for looking at a set of data and making comparisons. For example, it's easier to see which items are taking the largest chunk of your budget by glancing at the above chart rather than looking at a string of numbers. They can also shows trends over time, or reveal patterns in periodic sequences.

https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/bar-chart-bar-graph-examples/



7. Frequency Distribution: Bar charts

100 students were asked how they go to school. These are their answers:

Mode of transport	Frequency
Bike	15
Car	55
On foot	18
Bus	12

1. Draw a bar chart based on the above data
2. Comment on the above table

7. Frequency Distribution: Bar charts

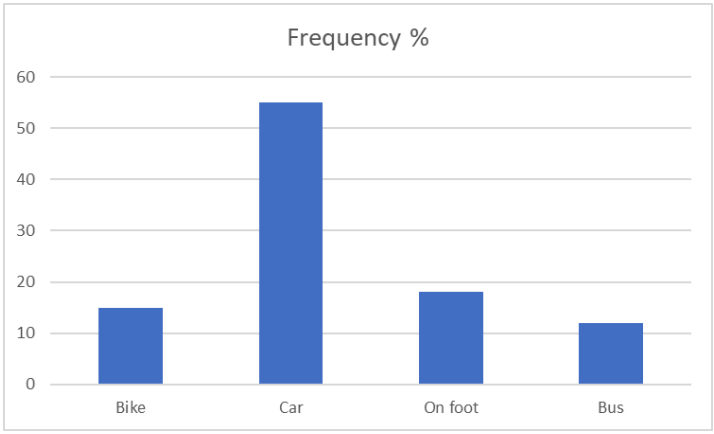
EXAMPLE

1. The bar chart on the right

Tips! Before drawing the bar graph, we have to consider what the requirements of our graph will be from the info in the table:

- There are 4 categories, so it must be wide enough for 4 bars,
- The highest frequency is 55, meaning the scale on the y-axis must go at least at as high as 55.

2. Most of the students (55%) go to school by car while the bus is used by the least of students (12%)

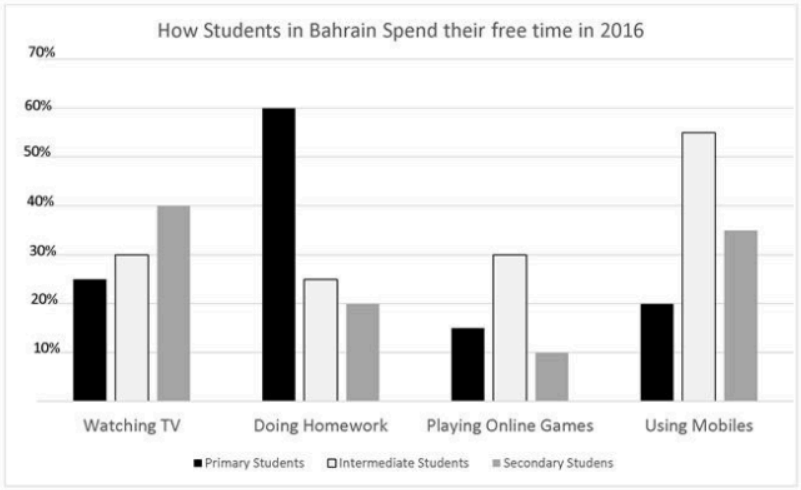


Tip! Remember to use data of the case study on your answers!

ibgenius

ACTIVITY

7. Frequency Distribution:Bar charts - Interpretation



Source: https://en.islcollective.com/english-esl-worksheets/material-type/task-based-learning-tbl-activities/bar-chart-teaching-part-1/97022

Look at the bar chart and answer the following:

1. Which activity do the primary students prefer the most?
2. Which activity the secondary students prefer the least?
3. Compare the activity of doing homework among all students
4. What do you observe about using mobiles among all students?

Tip: In a case study you may be given more information and therefore you will be able to evaluate more in depth.

ibgenius

7. Frequency Distribution: Pie charts

- A Pie Chart is a type of graph that displays data in a circle.
- It uses "pie slices" to show relative sizes of data. The pieces of the graph are proportional to the fraction of the whole in each category. In other words, each slice of the pie is relative to the size of that category in the group as a whole.
- The entire “pie” represents 100 percent of a whole, while the pie “slices” represent portions of the whole.
- Pie charts offers a snapshot of how a group is broken down into smaller pieces.

7. Frequency Distribution: Pie charts

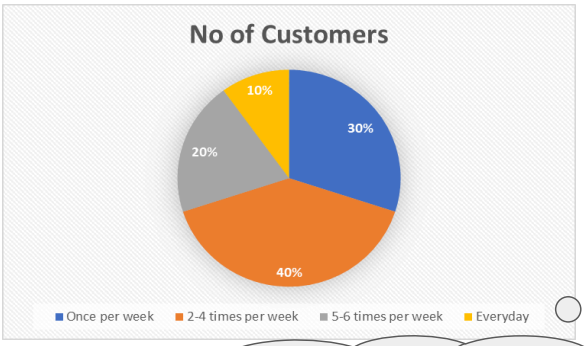
50 customers were asked in a coffee shop how often they buy coffee on a weekly basis. These are the results of the survey:

Frequency of buying coffee	No of Customers	%
Once per week	10	30%
2-4 times per week	25	40%
5-6 times per week	10	20%
Everyday	5	10%

7. Frequency Distribution: Pie charts - inte

EXAMPLE

This is a pie chart representing the results of the survey presented in the previous slide. Comment on the results:



Tip! You can add up percentages in your comments - this shows analysis of results and not only description

Answer:

- 40% of respondents buy coffee 2-4 per week, which is the most frequent answer
- 70% of respondents buy up to 4 times per week coffee
- Only 10% of respondents buy coffee every day

ibgenius

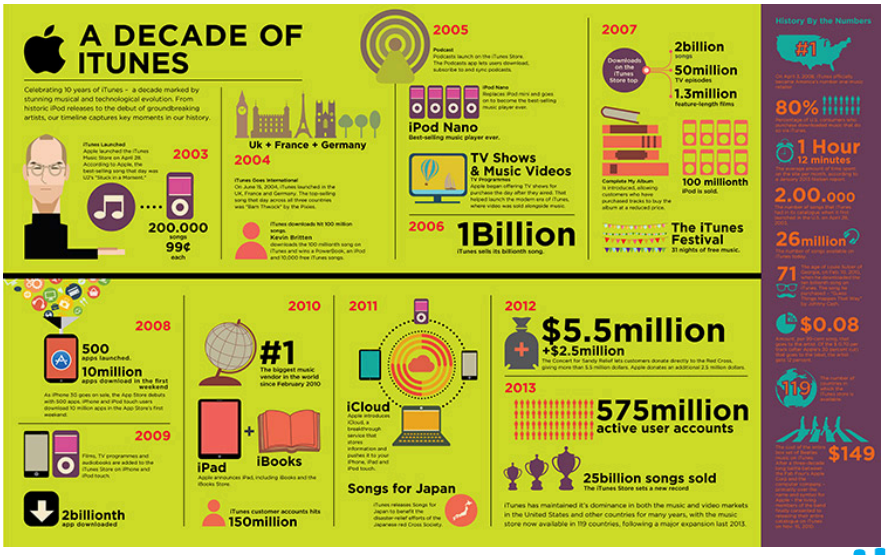
7. Frequency Distribution: Infographics

Infographics (comes from "information" and "graphics") are graphic visual representations of information, data, or knowledge intended to present information quickly and clearly.

- The three parts of all infographics are the visual, the content, and the knowledge.
- Statistics and facts usually serve as the content for infographics
- Infographics are effective because of their visual element.

ibgenius

7. Frequency Distribution: Infographics



Source: https://www.vectordiary.com/design/best-timeline-infographic-templates/

7. Frequency Distribution:Infographics

Source: https://www.worldbank.org/en/news/feature/2013/09/09/world-bank-group-goals

7. Infographics

EXAMPLE

Source: https://www.un.org/en/sustainable-development-goals

ibgenius

7. Infographics - Evaluation

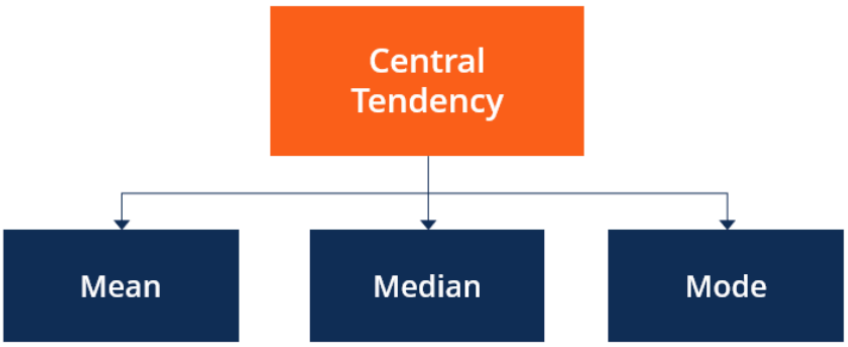
Advantages	Disadvantages
• It turns complicated information into visual graphics that are easy to understand. • The information is easier to retain. • Eye-catching • Easy to share	• The information on an infographic can still be misinterpreted. • Limited information is provided

ibgenius

7. Descriptive statistics - 3 basic categories

2. Measures of Central Tendency

Central tendency refers to a dataset's descriptive summary using a single value reflecting the center of the data distribution.



7. Central tendency: Mean

The mean, or M, considered the most popular measure of central tendency, is the average or most common value in a data set.

To find the mean, simply add up all response values and divide the sum by the total number of responses.

$$Mean = \frac{sum\ of\ all\ data\ values}{the\ number\ of\ data\ values}$$

7. Central tendency: Mean

Example:

The number of visits in the museum over a period of 7 days is:

150 30 45 60 35 90 200

The mean is= (150+30+45+60+35+90+200)/7= 87.14

7. Central tendency: Median

The median is the middle value of an ordered data set. **In other words, the median refers to the middle score for a data set in ascending order.**

The median splits the data in halves. Half of the data values are less than or equal to the median and half are greater or equal to it.

E.g. if the median mark for a revision test is 56% then you know that half of the class scored less than or equal to 56% and half scored greater or equal to 56%.

7. Central tendency: Median

- **For an odd number of data values**, the median is one of the original data values.
  e.g. if the values are 2, 15, 18, 22, 25 the median is 18

- **For an even number of data values**, the median is the average of the two middle values, and hence may not be in the original data set.
  e.g. if the values are 2, 15, 18, 22 the median is (15+18)/2= 16.5

Tip! If there are n values listed in order from the smallest to the largest the median is the $(\frac{n+1}{2})$ data value.

e.g. n= 13, the median is (13+1)/2=7
n= 14, the median is (14+1)/2=7.5

7. Central tendency: Mode

The mode refers to the score or value that is most frequent in a data set.

- If a data set has two values which both occur most frequently, we say it is bimodal.
- If a data set has three or more values which all occur most frequently, the mode is not an appropriate measure of center to use.
- To find the mode, order your data set from lowest to highest and find the response that occurs most frequently.

***Note**: For discrete data, the mode is the most frequently occurring value in the data set. For continuous data, we cannot talk about a mode in this way, but about a modal class, which is the class or group that has the highest frequency (e.g. age group)*

7. Central tendency: Mode

Example: The data collected are: 1, 5, 13, 14, 5, 21, 67, 32, 5, 7

1. **We need to put them in order: 1, 5, 5, 5, 7, 13, 14, 21, 32, 67**
2. **The mode is 5 as it occurs three times.**

7. Activity

The number of orders of birthday cakes at a local pastry shop each day over a 15-day period are:

2 4 4 9 8 8 6 4 7 9 1 3 5 3 5

a. Calculate the mean average orders (2 marks)
b. Calculate the median (2 marks)
c. Calculate the mode (2 marks)

7. Activity - ANSWER

The number of orders of birthday cakes at a local pastry shop each day over a 15-day period are:

2 4 4 9 8 8 6 4 7 9 1 3 5 3 5

a. Mean= (2+4+4+9+8+8+6+4+7+9+2+3+5+3+5)/15=5.2 orders
b. Median: as n=15 then (n+1)/2=8, the 8th value is the median after the data are put in order

1 2 3 3 4 4 4 5 5 6 7 8 8 9 9

The median is 5

c. The mode is 4 as this is the data value that occurs more frequently

7. Interpretation of data

The mean, mode and median can all be used **to indicate the centre of a set of numbers.** The most appropriate measure will depend on the type of data to be analysed. The following table will help you **for the interpretation of data**:

Statistic	**Properties**
Mode	- Gives the most usual value - Only takes common values into account - Not affected by extreme values
Mean	- Takes all values into account - Affected by extreme values
Median	- Gives the halfway point of the data - Only takes middle values into account - Not affected by extreme values

7. Descriptive statistics - 3 basic categories

3. **Measures of Variability**

A measure of variability is a summary statistic reflecting the degree of dispersion in a sample. The measures of variability determine how far apart the data points appear to fall from the center. In other words, it describes the dispersion of data within the set.

The range, standard deviation, and variance are used, respectively, to depict different components and aspects of the spread. Here, we will work only on standard deviation.

7. Measures of Variability: Quartiles

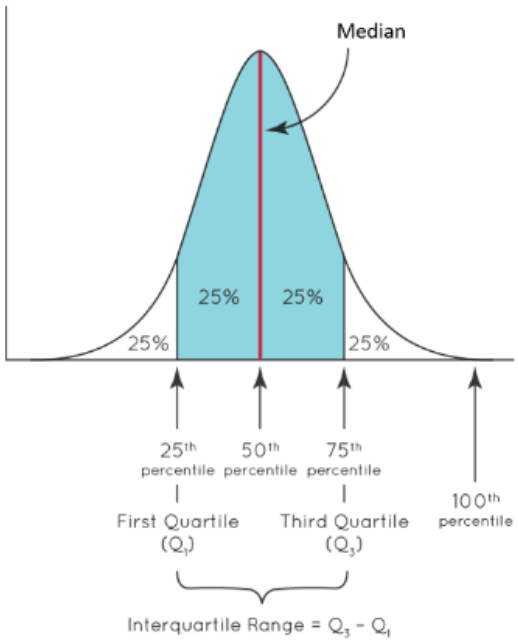
"**A quartile is a statistical term that describes a division of observations into four defined intervals based on the values of the data and how they compare to the entire set of observations**". (investopedia.com)

In other words, quartiles are values that divide the data into quarters. However, quartiles aren't shaped like pizza slices (i.e. pie charts); Instead they divide the data into four segments according to where the numbers fall on the number line.

7. Quartiles

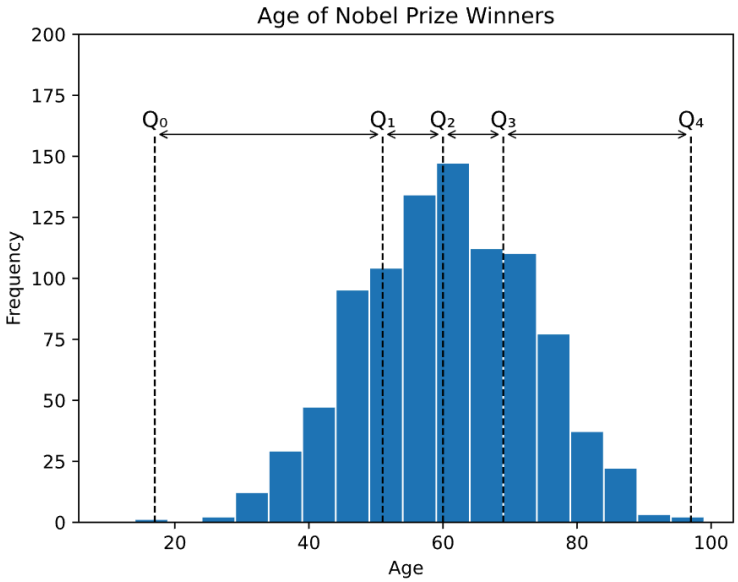
- The quartile measures the spread of values above and below the mean by dividing the distribution into four groups.
- A quartile divides data into three points—a lower quartile, median, and upper quartile—to form four groups of the dataset.
- The four quarters that divide a data set into quartiles are:
  - The lowest 25% of numbers.
  - The next lowest 25% of numbers (up to the median).
  - The second highest 25% of numbers (above the median).
  - The highest 25% of numbers.
- Quartiles are used to calculate the interquartile range, which is a measure of variability around the median.

7. Quartiles



Source: https://www.cuemath.com/quartile-formula/

7. Quartiles



Source: Source: https://www.cuemath.com/quartile-formula/

The quartiles (Q0,Q1,Q2,Q3,Q4) are the values that separate each quarter.

Between Q0 and Q1 are the 25% lowest values in the data. Between Q1 and Q2 are the next 25%. And so on.

- Q0 is the smallest value in the data.
- Q1 is the value separating the first quarter from the second quarter of the data.
- Q2 is the middle value (median), separating the bottom from the top half.
- Q3 is the value separating the third quarter from the fourth quarter
- Q4 is the largest value in the data.

7. Standard deviation

Standard deviation (σ) measures how far a 'typical' observation is from the average of the data (μ).

If the data points are further from the mean, there is a **high deviation** within the data set; thus, the more spread out the data, the higher the standard deviation.

If the data points are near the mean, there is a **low deviation** within the data set; thus, the fewer spread out of the data, the lower the standard deviation.

7. Standard deviation

Tip! Standard deviation is useful in comparing sets of data which may have the same mean but a different range.

For example, the mean of the following two sets of data is the same:

a. 14, 14, 15, 16, 16 (mean=15, standard deviation σ=0.89)
b. 2, 7, 14, 22, 30 (mean=15, standard deviation σ=10.08)

What do you observe?

The second is clearly more spread out than the first one, and the second standard deviation is higher than the first one (10.08 vs 0.89). If a set has a low standard deviation, the values are not spread out too much.

7. Standard deviation in Business

1. **Standard Deviation in Real Estate**

Standard deviation is a metric that is used often by real estate agents. Real estate agents calculate the standard deviation of house prices in an area to inform their customers of the type of variation in house prices they can expect.

2. **Standard Deviation in Human Resources**

Standard deviation is often used by Human Resource departments at companies. The Human Resource manager often calculates the standard deviation of salaries in a certain field so that they can know what type of variation in salaries to offer to new employees.

7. Standard deviation

3. **Standard Deviation in Marketing**
Standard deviation is often used by marketing departments to gain an understanding of competitors' advertisements. They calculate the standard deviation of the number of ads used by competitors to understand whether or not competitors are using more or less ads than normal during a given period.

4. **Standard Deviation in Operations Management**
Standard deviation is often used by operations management departments. They calculate the standard deviation of raw materials' prices in order to an understanding of price variations they may get when purchasing raw materials.

7. Standard deviation explained

7. Standard deviation explained

7. Standard deviation explained

Are you ready for the quizz?

- **Do the quiz in the ibGenius platform**
- **You need 70% to pass!**
- **Then you are ready to move on to the next Business Management Tool!**