WITHYOU WITHME

LESSON HANDOUT

Applying Basic Descriptive Statistics

Why?

Descriptive statistics will help you:

- Understand the nature of your dataset
- Direct your analysis
- Describe your results

Using formulas in Excel

A formula is an expression which calculates the value of a cell or cells.

- Formulas begin with an =
- Then formula name
- Then conditions within brackets ()

Formula	**Definition**	Expression which calculates the value of a cell/ cells					
	Components	**=**	**Name**	**(**	**Values or Cells**	**, Conditions**	**)**

Basic descriptive statistics functions

- **Average** - the average or mean of a selection
  =average(select cells to average)

- **Median** - the middle value from a selection
  =median(select cells from which we want the median)

- **Mode** - the most often occurring value of a selection
  =mode(select cells from which we want the mode)

- **Max** - the highest cell value of a selection
  =max(select cells from which we want the max)

- **Min** - the lowest cell value of a selection
  =min(select cells from which we want the min)

- **Range** - the difference between the Max and Min value
  = (max value – min value)

Advantages of Databases:

Scalability: Databases are able to store very large volumes of data. If needed, additional storage capacity can be added.

Security: Access rights can be established and monitored to protect data from unauthorized users. This ensures security and privacy of information.

Multi-access: Enables multiples users to access the same data at the same time.

Single Source: Consolidates data from different sources into a single accessible location.

Databases:

There are 2 main types of Databases:

- Relational (Structured) Databases
- Unstructured Databases

Relational (Structured) Databases: The most commonly used database.

Relational databases hold data which can be stored in tabular formats. and are structured to recognize relations amongst stored items of information. Data is stored in a series of **interconnected (related) tables.**

They are accessed using a programming language called Structured Query Language (SQL).

Oracle and pgAdmin are examples of relational databases

Unstructured Databases: store various types of data in various forms. They are used to store what is called **Big Data** which is data that is not stored in a structured manner and requires special treatment due to its unconventional nature

Big data is defined by the 3 Vs:

- Volume : The scale or number of data involved
- Variety : The different varieties of data to be stored such as video, audio or text.
- Velocity: The fast speed at which the data must be processed.

Hadoop and Apache Spark are some examples of unstructured databases.