

Objective: This lab aims to introduce you to use of spreadsheet for basic data visualisation, summary, and analysis

Introduction to Spreadsheet


Spreadsheets are digital tools that organise, analyse, and display data clearly and effectively. They are versatile and easy to use, making them popular for various tasks, and are very popular in the Industry. Spreadsheets can be used for basic arithmetic operations and more complex data analysis using formulas, functions, visualisation and more. They are used in various applications, including budgeting, financial planning, project management, inventory control, and research analysis. Users can create charts, graphs, and tables to represent data in a way that is easy to understand. This can help to reveal patterns, trends, and connections in the data, which can be helpful for decision-making and communication. Microsoft Excel is the most popular spreadsheet software, but alternatives such as Google Sheets and LibreOffice Calc are available. Learning to effectively use spreadsheet to handling data is a very useful skill.

A spreadsheet is made up rows and columns of cells, with rows numbered as 1,2,3, ... and columns labeled as A, B, C, We refer to a cell by its column-row id. Thus C5 represents the cell in column C and row 5.

Spreadsheets provide many built-in functions that you can access by pressing '=' from within a cell and typing a few alphabets of the possible function you want. Most of the functions will have intuitive names, such as SUM, MIN, AVERAGE, STDEV, etc.

In IE507, we will be using Google Sheets as a tool to analyse and visualise given data.

Instructions:

- Use your credentials to login into the internet.iitb.ac.in via iitb-sso.
- Once the internet is active on your system, login into your Moodle account again from the IITB credential in moodle.iitb.ac.in
- Finally, your task is to log in to google drive (drive.google.com) using institute credentials.
- To open Google Sheets, use your existing location of the IE507 folder on your drive and right-click on the mouse; a new menu will appear from which you choose Google Sheet () , and Google Sheet will appear in another tab and rename it based on your convenience.

Example Using SUM: the SUM function below sums all the values in cells A1 to A10. The 'A1:A10' defines a range of cells.

`=SUM(A1:A10)`

Suppose we want to add up the values in cells A1, A3, A5, and A7 only, then we can say,

`=SUM(A1,A3,A5,A7)`

Example Using IF- condition Below is an IF function in a spreadsheet that analyses data with conditional statements.

```
=IF(A1>10, "Above 10", "Below 10")
```

This formula would return the text "Above 10" if the value in cell A1 is greater than 10 and the text "Below 10" otherwise. With three arguments in IF(x, y, z) as standard, x represents the condition that needs to be tested. If the condition holds true, y is returned otherwise, z is returned as output.

Further, You can also use the IF function to return different values depending on the value of a cell. For example, the following formula would return the value 10 if the value in cell A1 is greater than 5, the value 20 if the value in cell A1 is equal to 5, and the value 30 if the value in cell A1 is less than 5:

```
=IF(A1>5, 10, IF(A1=5, 20, 30))
```

Example Using COUNT: The COUNT function will count all cells in the range that contain numbers. It will ignore cells that contain text, formulas, or blank cells.

```
=COUNT(value1, [value2, ...])
```

value1: (Required) The first item, cell reference, or range within which you want to count numbers.

value2, ...: (Optional) Up to 255 additional items, cell references, or ranges within which you want to count numbers.

Task 0: Spreadsheet Introduction

A link for SpreadsheetIntro on moodle, use this link to open sheet and save this into your drive using File □ Make a Copy. This provided sheet contains around 10 tasks which you need to perform as part of understanding the spreadsheet interface.

Await instructions from your TA and let's proceed in sync for this exercise.

Task 1: Grading : You are given a Google sheet named *scores* (available in Moodle) which contains marks of a full semester course for Quiz 1, Mid Semester exam, Quiz 2, Surprise Quiz and End Semester, for a class of **150+** students. Make a copy of this sheet for your use.

Tasks to do:

You can pair up with your neighbor, talk and discuss.

Each of you have to complete the tasks in your spreadsheet.

Use google / help to search & learn various commands, syntax, examples, etc.

After completion of each task, please show it to your TA.

- A. For each exam, compute the scores'
 - maximum,
 - minimum,
 - median,
 - mode,
 - average
 - standard deviation.
- B. For each student, compute Insem (excl. End Semester) score on 60%. The weightage for Quiz1, midsem, Quiz2 and SurpriseTest are also given in the spreadsheet. Compute the basic statistics for data series.
- C. For each student, compute TOTAL (**includes**. End Semester) score on 100%. Compute the basic statistics for data series.
- D. Create another new data column, called 'Rounded Score', which contains the weighted total score of the student, rounded to the *nearest* integer. Use appropriate functions. Compute the basic statistics for data series.
- E. Create another new data column, where the raw top score is taken as 100%, and the rest of the scores are scaled accordingly. Use the appropriate formula. Compute the basic statistics for data series
- F. Create correlation matrix, showing the correlation between the different tests.
- G. GRAPH: Create a SCATTER chart between MIDSEM and Endsem scores. Use raw scores as the independent values (X-axis values) and Endsem raw score as the dependent values (Y-axis values).
 - Can we say that a student who performed well in midsem also performs well in Endsem? Write your observation below the chart.
 - Add a linear trend line & display the equation on the chart.
 - Change the type of fit to get the lowest R^2 value.

- H. GRAPH: Create a SCATTER chart between INSEM and ENDSEM scores. Use raw scores as the independent values (X-axis values) and Endsem raw score as the dependent values (Y-axis values).
- Can we say that a student who performed well in insem also performs well in Endsem? Write your observation below the chart.
 - Add a trend line & display the equation on the chart.
- I. HISTOGRAM: Create a histogram of the total scores, using the charts feature in Google sheet. Try out different numbers of bins to see which histogram seems more useful.

Duplicate your Sheet1. Do all the below tasks in the copy of Sheet1.

- J. Sort the data set in descending order of the total score.
- You can either select the dataset, and use the menu Data>Sort Range, or
 - select the header row, click Data>Create a Filter, and then use the filter menu to sort.
 - *Note: DO NOT sort the statistics part of the spreadsheet.*
- K. Identify how many students are there in the following three intervals: (0, mean – sd), (mean – sd, mean + sd), (mean + sd, 100)?
- Hint:** You can do this manually.
- L. Assign appropriate letter grades to each student. The grades can be either AA, AB, BB, BC, CC, CD, DD, or FR. Ensure that:
- Only about 5-10% of students get AA
 - no more than 25% of students get AA or AB
 - If a student's score is less than 30-35% of the class highest, then that student should get FR grade.
 - The range of marks within grades should be consistent and reasonable
 - Class CPI to be around 7.5 (say, between 7 and 8).

$$\text{Class CPI} = \frac{\sum_{x \in \text{students}} \text{gradePoint}}{\text{TotalClassSize}}$$

- Class CPI = $\frac{\sum_{x \in \text{students}} \text{gradePoint}}{\text{TotalClassSize}}$
- Grade points associated with the letter grades are as follows:
AA-10; AB-9; BB-8; BC-7; CC-6; CD-5; DD-4; or FR-0.

Write a few lines in the spreadsheet on your grading scheme. Prepare a summary of the number of students in each grade.

- M. Compute the average CPI of BTechs, MTechs, PhDs and MScs and DDs separately.

Duplicate your Sheet1.

- N. Identify the fraction and list of students whose
- (a) performance improved over the course of semester
 - (b) performance worsened over the course of semester
 - (c) performance remind more or less the same over the course of semester
 - (d) performance improved in Endsem over the insemester performance
 - (e) performance worsened in Endsem over the insemester performance

For this task, you may need to think about what constitutes a performance improvement. You may need to do some data transformations, some more basic computations etc as needed. Also, present the analysis in a nice manner. Note that there may be more than one way to do this.