Electronic Spreadsheets in Data Processing

Zhang Kun Qi
Informatics, Kyoto University
Kunqizhang24@gmail.com

Abstract

Spreadsheets are easy and useful to work with data. This paper is going to talk about how to use excel to help with fixing errors, cleaning data, and checking data. Example is given to show how to let people use excel to work with data. The paper also talks about some limitations of excel when the data is very large.

1 Introduction

Nowadays, spreadsheets like excel are very popular among people. They help us organize and check data. This paper is going to talk about some excel tools that make data better and how to use them in real life.

1.1 What is Data Cleaning?

When we work with data, it is not always good because we will make mistakes and miss parts. In that way cleaning data can fix these problems.

1.2 Why Is It Important?

If the data has mistaken, the result of all the actions and progress will be wasted and wrong. Cleaning data makes sure the result is correct to save time.

1.3 What Will This Paper Talk About?

This paper will show how Excel helps fix mistakes, remove bad data, and check data. It will also talk about some limitations of excel.

2 Excel Tools

Excels can find cells with mistake. Ex. You can press keyboards, choose “Special”, and find all the wrong cells. This makes it easy to fix them.

2.1 Finding Mistakes

Sometimes, mistakes in data are not easy to find. Excel makes this easy with its "Go to Special" feature:

Press Ctrl + G.

Choose "Special" and then select "Formulas."

Select "Errors" to find cells with mistakes.

This tool helps highlight problems so they can be fixed quickly.

2.2 Removing Duplicate Data

Data can sometimes have duplicate entries, which can affect the results. There is a “Conditional Formatting” feature that highlights duplicate data. Users can then remove and correct these duplicates.

2.3 Checking Data

Data Validation in Excel helps make sure the data is correct. For example:

A cell can be set to accept only numbers between 1 and 100.

A cell can be set to accept only dates within a certain range.

To use Data Validation:

Select the cells you want to check.
Go to the "Data" tab and click "Data Validation."
Set the rules, like number ranges or text formats.
Click "OK" to apply the rules.

2.4Automating Tasks

Excel is not only for manual data work; it also helps save time by automating tasks. The "Macro" feature in Excel allows users to record repetitive steps and play them back with one click. Ex, if you need to clean a similar dataset every week, you can:

Click "View" → "Macros" → "Record Macro."

Perform the actions you want Excel to remember, like removing duplicates or applying formatting.

Stop recording and save the macro.

Later, you can run this macro to do the sane steps automatically. This can make workflows faster and reduce human errors.

2.5 Visualizing Data

Visualizing data helps people understand it better. Excel provides different types of charts, such as bar charts, line graphs, and pie charts. For instance:

A bar chart is great for comparing categories, like sales across different months.

A line chart shows trends over time, like tracking temperature changes.

A pie chart displays proportions, like how much of your budget is spent on food.

To create a chart:

Highlight your data.

Click "Insert" → "Chart" and choose the type you want.

Customize the chart with labels and colors.

Using charts makes data easier to explain, especially in reports and presentations.

2.6 Using Conditional Formatting

Conditional formatting is a powerful feature in Excel that changes the appearance of cells based on their values. This is especially useful when you want to highlight specific data, like numbers above a certain threshold or duplicate entries.

Example:

Highlight high values: If sales exceed $10000, excel can automatically change the cell color to green.

Low stock alerts: Inventory items with stock below 10 can be highlighted in red.

Steps to Use Conditional Formatting:

Select the cells you want to format.

Go to the "Home" tab → "Conditional Formatting."

Choose a rule, such as "Greater Than" or "Duplicate Values."

Set the condition and choose the formatting

This feature helps users quickly spot important trends or problems in their data

2.7 Data Filtering and Sorting

Excels filter and sorting tools make it easy and conventional to find specific information in large datasets. For examples:

Filter: if we just want to see the sales make by a specific person or during a certain time or month of a list of 1000 sales transactions. We can filter for that.

Sorting: we can organize data in alphabetical order, numerically or by date.

How to Sort:

Select your data range.

Go to "Data" tab → "Sort."

Choose the column and order (e.g., ascending or descending).

These tools make large datasets more manageable and ensure that key information is easy to find

2.8 Handling Missing Data

The set of the data may not be complete, with some cells left blank. This can create problems during analysis, Excel provide tools to handles missing data:

Fill with Default Values: Replace blanks with a default value, like "0" or "Not Available."

Use Formulas: Use formulas like =IF(ISBLANK(A1), "Missing", A1) to flag missing data.

Interpolate Missing Values: For numeric data, you can use interpolation to estimate missing values. For example, if two adjacent cells have values of 10 and 20, the missing value between them can be calculated as 15.

By knowing the address of the missing data, users can improve the accuracy of their results and make sure the database is ready for further processing.

3 Applications of Excel

Excel is used in various fields, including academics, business, and daily life. Here are some detailed examples:

3.1 Academic Use

Excels can helps students and researchers in schools and universities to analyze data. For example , finding “outliers” is a common task in research. Outliers are data points that are very different from others in the dataset. To detect them, students can use the formula:

$$\text{Standardized Value} = (X - \mu)/\sigma$$

where X is the data point, μ is the mean, and σ is the standard deviation. This formula helps identify unusual data points that may need further review or correction

3.2 Business Use

Companies use Excel for sales analysis, inventory management, and financial planning. For instance:

Month	Sales
January	10000
February	15000
March	20000

With simple formulas like =SUM (B2:B4), businesses can quickly calculate total sales or analyze monthly trends. Conditional formatting can be used to highlight months where sales exceed $15,000, helping managers make informed decisions

3.3 Personal Use

Excel is a valuable tool for managing personal budgets. People can track their expenses and savings. For example:

date	item	Cost

2024-01-01	groceries	50
2024-01-05	electricity	30
2024-01-10	transport	20

Using a formula like =SUM (C2:C4), individuals can calculate their total monthly expenses. Creating a pie chart from this data provides a visual representation of spending habits, making it easier to plan finances.

3.4 Data Visualization

Visualization is crucial for understanding data. Excel's chart options, like bar charts, line graphs, and pie charts, make data easier to interpret. For example:

A line chart can show sales trends over time.

A pie chart can display budget distribution.

To create a chart:

Select the data.

Click "Insert" → "Chart."

Choose the desired chart type.

This feature is commonly used in presentations and reports to communicate data findings effectively.

3.5 Advanced Features in Research

Excel's Pivot Tables allow researchers to summarize large datasets. For example:

Group data by categories like age groups or regions.

Calculate metrics like averages or totals.

Creating a Pivot Table is simple:

Select the dataset.

Click "Insert" → "Pivot Table."

Drag fields to the "Rows" and "Values" sections.

This feature is particularly useful for researchers working with surveys or experimental data.

4. Limits of Excel

While Excel is powerful, it has some limitations

4.1 Data Size

Excel can handle only up to 1,048,576 rows and 16,384 columns. For larger datasets, tools like SQL or Python are recommended. For example, analyzing datasets with millions of entries would be inefficient in Excel

4.2 Performance Issues

As files grow larger, Excel becomes slower. Complex calculations and large visualizations can take longer to process, affecting efficiency

4.3 Lack of Advanced Automation

Excel's macros are helpful but limited compared to full programming languages. For tasks requiring advanced automation, such as dynamic data analysis or machine learning, languages like Python offer better solutions

5. Conclusion

Excel is a versatile tool for organizing, analyzing, and visualizing data. It is widely used in education, business, and personal tasks due to its simplicity and accessibility. Despite its limitations with large datasets and automation, Excel remains one of the best tools for small to medium-sized tasks.

Acknowledgments

Preparation of the Microsoft Word file was supported by the International Joint Conference on Artificial Intelligence (IJCAI). An early version of this document was created by Shirley Jowell and Peter F. Patel-Schneider. It was subsequently modified by Jennifer Ballentine and Thomas Dean, Bernhard Nebel, Daniel Pagen Stecher, Toby Walsh, and Carles Sierra before arriving at its current form.

References

Sight Seti Yanto & Ismail Setiawan (2022). *Data Science with Excel*. International Journal of Computer and Information System, 3(3), 104-110.

Pandita, R., Parnin, C., Hermans, F., & Murphy-Hill, E. (2018). *No half-measures: A study of manual and tool-assisted end-user programming tasks in Excel*. IEEE Symposium on Visual Languages and Human-Centric Computing, 95–103.

Huang, Z., & He, Y. (2018). *Auto-detect: Data-driven error detection in tables*. Proceedings of the International Conference on Management of Data, 1377–1392.