Statistic Tools
Introduction

Franklin Zhou
IPAR, East Carolina University
• Excel
• R
• SAS
• Python
• JMP
• SPSS
• Minitab
• ...

If statistics programs/languages were cars...
• One of the most widely used software
• Up to 1 million rows and 16,000 columns per worksheet.
• Can create and customize charts, graphs, tables, and pivot tables to visualize and summarize data
• Includes many useful built-in functions.
• Limited in functionality and flexibility
What appeals to me

- Easy to create and modify data
- Dynamic updates
- Font and Style
- Directly working on the data makes “WYSIWYG”
- Quick pivot table
- Copy & paste works well with other MS Office apps
- Collaborative work
- Be able to write and draw notes

What troubles me

- Formats
- Customize charts
- Build models
- Not friendly with large dataset
- Undo
What appeals to me

- Easy to create and modify data
- Dynamic updates
- Font and Style
- Directly working on the data makes “WYSIWYG”
- Quick pivot table
- Copy & paste works well with other MS Office apps
- Collaborative work
- Be able to write and draw notes

What troubles me

- Formats
- Customize charts
- Build models
- Not friendly with large dataset
- Undo
- One of the most popular **free** statistics software
- Programming-based
- Rely on third-party contributions
- Lots of online resources
• **Free** tool that provides integrated development environment (IDE) to use R language
• Resource management
• Programming friendly
• More than R script:
  • R Quarto/Markdown notebook and presentations
  • R Shiny Web App
What appeals to me
• Codes can be edited and reused
• More functions than Excel
• Customize plots
• Project management
• Be able to link with GitHub*

What troubles me
• Not as easy as Excel to view and edit data
• Too many ways to solve a problem
• Packages update issue
• Lack of “WYSIWYG”
• One of the most popular commercial data analytics software
• Programming-based
• All functions are built-in
• Lots of online resources
What appeals to me
• Codes can be edited and reused
• Less flexible coding
• More functions than Excel
• Customize plots
• Project management

What troubles me
• Not free to use
• Not as easy as Excel to view and edit data
• Lack of “WYSIWYG”
• Emerging competitor of R in data analysis field
• Has many same features as R:
  • Programming-based
  • Rely on third-party contributions
  • Lots of online resources
• Emerging competitor of R in data analysis field
• Has many same features as R:
  • Programming-based
  • Rely on third-party contributions
  • Lots of online resources
• Markdown for Python
• Free

JupyterLab: A Next-Generation Notebook Interface

Jupyter Notebook: The Classic Notebook Interface

Voilà: Share your results (web applications)
What appeals to me
• Similar as R and R Studio
• Has advantages in Deep Learning, Machine Learning, Data Mining, and AI integration.
• Can do more besides statistical analysis

What troubles me
• Not as easy as Excel to view and edit data
• Too many ways to solve a problem
• Version update issue
• Lack of “WYSIWYG”
• A product of SAS company
• Easy to use with GUI
• Good for plotting and building some advanced models
What appeals to me

• Simple and quick of making figures/plots
• Simple and quick of building certain models:
  o PCA, SEM, Factor Analysis
  o Neural, KNN, SVM
  o Hierarchical Cluster, K Means
  o Nonlinear, Time Series
  o Survival Analysis
  o Text Mining
• JMP offers free webinar every week.

What troubles me

• Not be able to replicate all steps with one button.
• Not be able to cover all statistical methods.
• Not flexible as R / Python
SPSS

- A commercial product from IBM company
- Easy to use with GUI
- Can cover most statistical analysis works
- Supports syntax command
What appeals to me

• Easy to use with GUI
• Simple and quick of building certain models, similar as JMP
• Many social science researchers (including my wife) use it.

What troubles me

• Not good at data processing
• My wife uses it.
- A commercial software
- With GUI
- Satisfy all basic statistical analysis needs
- Supports syntax command
- Less flexible than SPSS
What appeals to me
• Easy to use with GUI
• Simple and quick of building certain models
• Copy & paste the result to Word document

What troubles me
• Not good at data processing
• Does not support some advanced models (extra features costs a lot!)
• Only works within organization’s network (VPN) and one device per account.
<table>
<thead>
<tr>
<th>Software</th>
<th>Platform</th>
<th>GUI</th>
<th>Easy to learn and use</th>
<th>Lots of users</th>
<th>Free</th>
<th>Programming</th>
<th>Best for</th>
</tr>
</thead>
</table>
| Excel    | Windows Mac OS Web | Yes | Yes | Yes | No | No | • Data view  
  • Pivot table  
  • Basic statistical analysis |
| SAS      | Windows Web* | No | No | Yes | No | Yes | • Data processing  
  • Basic and Advanced statistical analysis  
  • Predictive modelling |
| R/Python | Windows Mac OS Web* | No | No | Yes | Yes | Yes | • Data processing  
  • Basic and Advanced statistical analysis  
  • Predictive modelling |
| JMP      | Windows Mac OS | Yes | Yes | No | No | No | • Advanced statistical analysis  
  • Predictive modelling  
  • Graphics |
| SPSS     | Windows Mac OS | Yes | Yes | Yes | No | Some | • Basic and Advanced statistical analysis  
  • Graphics  
  • Predictive modelling |
| Minitab  | Windows | Yes | Yes | No | No | Some | • Basic statistical analysis  
  • Graphics  
  • Basic predictive modelling |
Other tools that help with data analysis

Data visualization tools can help us to find the patterns or trends.
New features of Power BI Premium

Advanced AI capabilities:

- Automated machine learning (AutoML)
  - Binary Prediction
  - Classification
  - Regression Models
  - Time-series forecasting

- Cognitive Services
  - Sentiment Analysis
  - Key Phrase Extraction
  - Language Detection
  - Image Tagging
Share your experience and opinion
Thank you
Thank you for attending the 2024 NCAIR Annual Conference!

There’s a QR code in your program for a conference evaluation form. You’ll also get an e-mail following the conference with a link to the form, which will be available until 4/30.

At your earliest convenience, please take this opportunity to let the planning committee know your thoughts about this year’s conference and where you would like to meet next year.