Simple Statistics and Graphics in Excel

Made possible by the:

Clinical and Translational Science Institute (CTSI)

&

Division of Biostatistics

MEDICAL COLLEGE OF WISCONSIN

ADVANCING HEALTH THROUGH RESEARCH AND DISCOVERY CTSI

Institute *for* Health and Society

Community Engagement | Education | Health Equity | Research

Bridging the sciences BIOSTATISTICS

Simple Statistics and Graphics in Excel

Jessica E. Pruszynski, PhD

Assistant Professor of Biostatistics

Medical College of Wisconsin

CTSI – Advancing Health Through Research and Discovery

Institute for Health and Society
Community Engagement | Education | Health Equity | Research

Bridging the sciences
BIOSTATISTICS

Speaker Disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.

CME Evaluations

- Please help us by fill out an evaluation, even if you are not eligible for CME credit.

Outline

- Data entry
- Descriptive statistics
- Statistical inference
- Types of statistical graphs
- How to create graphs in Excel
- Qualities of good graphs
- Problems with Excel graphics

Data Analysis Tools

- Data Analysis add-in
- To install:
  - Office Button
  - Excel Options
  - Add-Ins
- Once installed, data analysis options will be available under the Data tab

Structure of the Data

- Data should always be entered in the form of a list
  - Color should not contain additional information
  - Each row should contain information on only one subject
- Nothing but data should be on the spreadsheet
- No special codes for missing data

Bad Data Entry

Treatment Group	Control Group	Treatment Group	Control Group
Age	Age	Height	Height
23	68	71	61
28	62	67	71
42	78	63	64
45		74	

Correct Data Entry

Group	Age	Height
1	23	71
1	28	67
1	42	63
1	45	74
2	68	61
2	62	71
2	78	64

- One row for each subject
- Variables clearly labeled
- No additional analysis

Descriptive Statistics

- Can use built in Excel functions to calculate descriptive statistics
  - AVERAGE
  - MEDIAN
  - STDEV
- More are available in the Data Analysis add-in
- Notes
  - Make sure to output results to a separate worksheet
  - Results do not change if data is altered
- Example

Cross Classification Tables

- Pivot tables
- Can display frequencies as well as means for different groups
- Available on the ‘Insert’ tab in Excel
- Example

Correlation

- Available in Data Analysis add-in
- Only gives estimate of the correlation coefficient
- Can also use CORREL function available in Excel
  - CORREL(array1,array2)
- Example

Statistical Tests

- Very limited options in Excel for inference
- Little to no support for categorical inference
- Available tests
  - T-tests (paired and independent)
  - ANOVA (single and two factor)
  - Regression

Independent T Test

- Found in data analysis add-in
- Available for equal and unequal variances
- TTEST function also performs the analysis
- TTEST(group 1, group 2, tails, type)
  - Type=2 → equal variances
  - Type=3 → unequal variances
- Preferable to perform using the add in
  - Function provides only the p-value
- Example

Paired T-Test

- Available in either the data analysis add-in or using the TTEST function
- If using the TTEST function, setting type equal to 1 produces the p-value for the paired t-test
- Example

Regression

- Available in data analysis add in
- Produces several statistics in output
  - R squared
  - ANOVA table
  - Regression coefficients
    - Test statistic
    - P-values
    - Confidence intervals
- Regression line can be added to scatterplot
- Example

Types of Graphs

- Exploratory
  - Designed to help the user visualize the data during the analysis
  - Main purpose is to summarize the data
- Presentation
  - Designed to display the results of a statistical analysis

Pie Charts

- Displays qualitative variable with a small number of categories
- Divides a circle into slices based on the relative frequency of each category
- Insert → Charts → Pie
- Example

Bar Charts

- Also used to display qualitative data
- Easier to visualize data with a bar chart than a pie chart
- Able to display data with more categories
- Insert → Charts → Column → 2D Column
- Example

Line Graphs

- Also referred to as trend plots or frequency polygons
- Often used to display trends in data over time
- Data must be in the form of a frequency distribution
- Select only the data from the frequency column
- Insert → Charts → Line
- Example

Scatterplots

- Graphical view of pairs of measurements on two variables
- Insert → Chart → Scatter
- Regression analysis
- Overlay the fitted regression line on top of the scatterplot
- Procedure in Excel
  - Right click on the points on the scatterplot
  - Select ‘Add Trendline’
  - Select ‘Display Equation’ and ‘Display R-squared’

Limitations

- Several options for graphs in the Chart Wizard
- Many common statistical graphs that are not available in Excel
  - Dot plots
  - Box plots
  - Stem and leaf plots
- Many of these graphs can be produced by manipulating the available charts

Qualities of Good Graphs

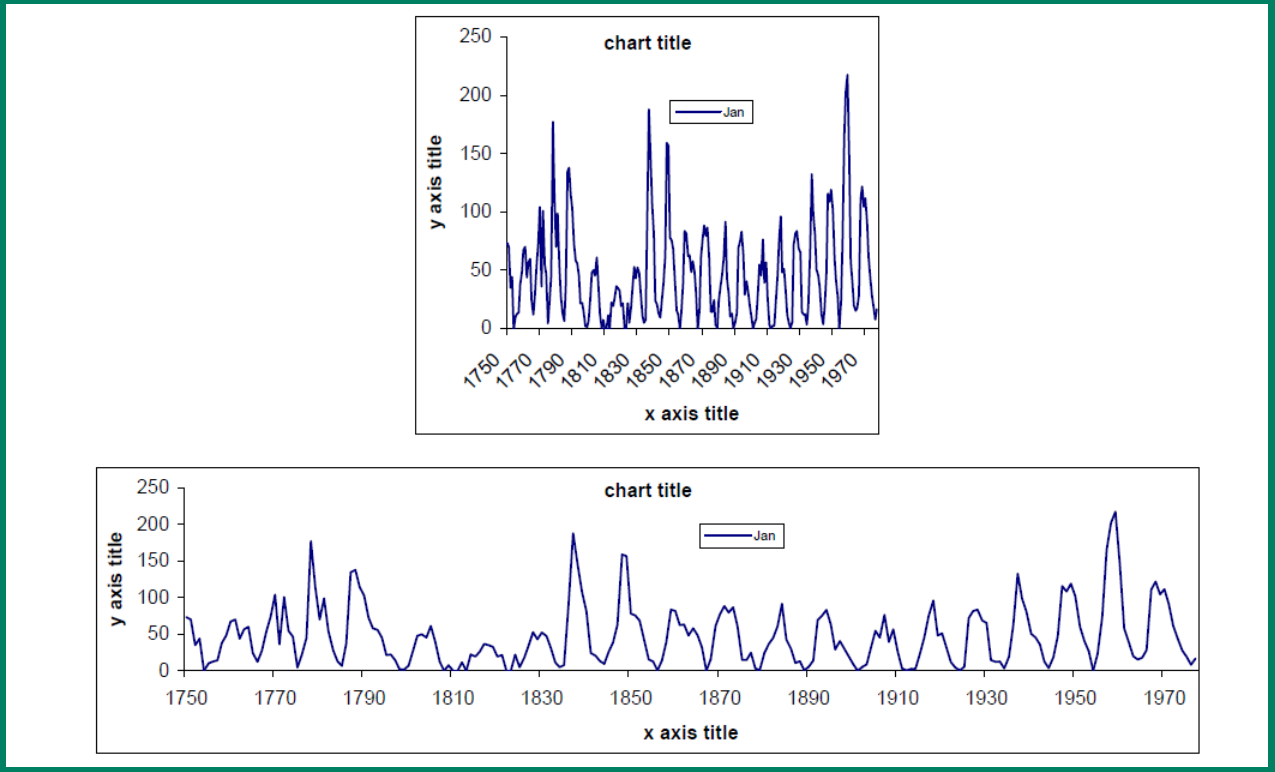
- Portray information without distortion
- No distracting elements
- Axes should contain the correct range
- Axes should be labeled appropriately
- Descriptive titles, captions, and legends
- The default graphs in Excel meet very few of these criteria

Third Dimension

- There is an option in Excel for a 3D line graph
- Data display is not optimal
- Third dimension is meaningless

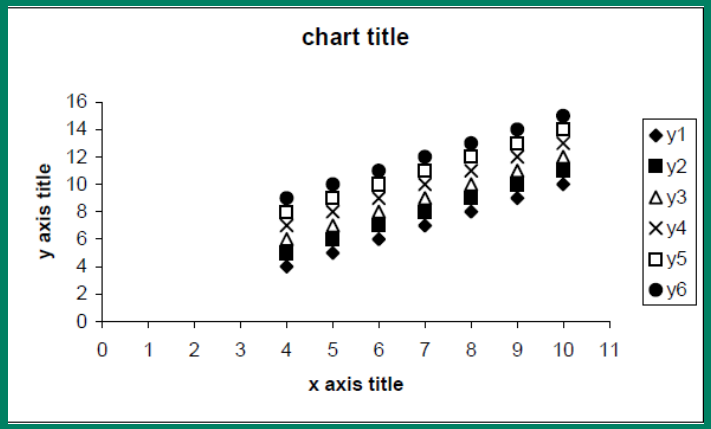
Aspect Ratio

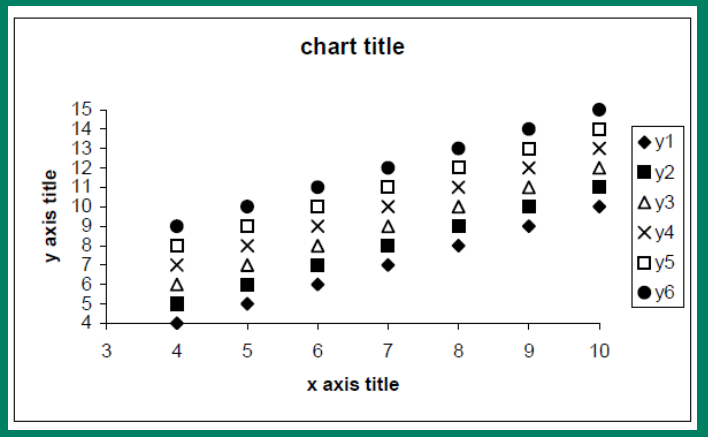
- Height/width ratio is extremely important
- Aspect ratio is easily changed in Excel



Axis Ranges

- Default in Excel can leave blank space
- Variation is often concealed in the default plot





Modifying the Graphs

- Visual aspects of the default graphs are often not optimal
- Not appropriate for publication or presentation
- Aspects that can be modified
  - Background
  - Gridlines
  - Colors
  - Legends
  - Titles and captions

Modifying the Graphs

- Background
  - Default background color is dark grey
  - Distracting and can often hide features of the graph
  - Solution: Right click on graph and select ‘Clear’

- Gridlines
  - Default setting is solid gridlines on the graph
  - Can overwhelm the graph
  - Not necessary
  - Solution: Double click the gridlines and make the appropriate changes in the dialog box

Modifying the Graphs

- Colors
  - Default colors are often difficult to distinguish
  - Think carefully about what colors can add to the graph
  - Different colors may not be necessary
  - Solution: Tools → Options → Color → Chart lines

- Legends
  - Often necessary to identify different series in the plot
  - Default locations are all outside the plot area
  - If there is available space, legend should be placed inside of the plot area
  - Can easily be moved if needed

Modifying the Graphs

- Titles
  - Graphs should always have titles and axes labels
  - Excel does not automatically add these labels
  - Solution
    - Double click on the graph
    - Layout → Axis Title
    - Layout → Chart Title

Conclusions

- Statistical inference
  - Can be a useful tool
  - Caution should be used
    - Numerical instability
    - Handling of missing data
- Statistical graphics
  - Excel does have several good graphing utilities
  - Easy to use for those with no training in statistical software
  - The graphs produced by Excel should be used with caution

Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support www.mcw.edu/biostatistics.htm

Free Drop-In Consulting

- MCW/ Froedtert: 1 – 3 PM
  - Monday, Wednesday, Friday @ CTSI Administrative offices (LL772A- TRU offices)
  - Tuesday, Thursday 1 – 3 PM @ Health Research Center, H2400
- VA: Every Monday, 9:30-10:30 am
  - VA Medical Center, Room 70-A 314-A
- Marquette: Every Tuesday, 8:30-10:30 am
  - School of Nursing, Clark Hall, Office of Research & Scholarship

References

- Pace, L.A. (2007). *The Excel 2007 Data & Statistics Cookbook.* Anderson, SC: TwoPaces LLC.
- Su, Y-S. (2008). *It's easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs.*
- University of Reading Section of Applied Statistics (2006). *Guidelines for Good Statistical Graphics in Excel.*