TCEC Webinar Guide: Data Analysis in Excel

Learning Objectives:

Perform basic Excel functions (format, filter, sort, hide, conditional formatting, etc.)

Calculate descriptive statistics in Excel

Recognize materials required before starting data analysis

Classify which variables to prioritize

Identify why/when/how to calculate confidence intervals for categorical data

Compare confidence intervals and provide an interpretation

Operate pivot tables

Excel Tips:

- Save! Save! Save!
  - **Always save raw data, and don't touch it!**
  - "Save As" another file so that the raw data are left untouched

- Label! Label! Label!
  - Label variables, label sheets, label files, label tables, label titles, label everything!

- Document! Document! Document!
  - Create a data analysis plan and code book if not already provided with the data
  - Record steps via macros or on a Word document so you can retrace your steps
  - List file locations and define naming conventions if necessary

- Get organized – it's more efficient!
  - Gather all materials – data files, codebook, data analysis plan, shell tables, etc.
  - Do NOT do it by hand! Use the resources including TCEC for tips and options
  - Use shortcuts:
    - Ctrl + down/up/left/right arrow to go to the end of in that direction
    - Ctrl + Shift + down/up/left/right arrow to select all cells in that direction
    - Ctrl + A = select all
    - Ctrl + Z = undo last change
    - Ctrl + Y = redo last change
    - Shift + Space = select row
    - Ctrl + Spacebar = select column
    - Ctrl + Home = go to cell A1
    - Ctrl + F = open Find/Replace window
    - Tab to complete a formula
    - =VLOOKUP is amaaaaaaziiiing!
    - F2 to check formulas
    - F4 in a formula = relative cell reference

Basic Excel Functions

Note: These steps do not change the data; they merely change the view of the data.

1. **Turn on filters** = allows for easy sorting and filtering
    a. Select the row of variable names → Home → Sort & Filter (toward the left side) → Filter

2. **Highlight variable labels** = make the names stand out somehow e.g. bold, change size or color of font or cell, and definitely freeze panes.
    a. BONUS Use "Format as Table" feature under the Home ribbon
    b. **Freeze variable labels** = keeps the variables in view even when scrolling through rows
        i. Select the cell to the RIGHT and BELOW the column and row you want to freeze → View → Freeze Panes → Freeze Panes

3. **Resize and/or wrap cells** = allows all of a cell's contents to be seen
    a. Resize: Select row or column to be resized (can select multiple rows or columns at once) → move cursor over the line to the right of the column or below the row → click and hold → drag to left to make smaller or right to make bigger
        AND/OR
    b. Wrap: Select cell(s) to be resized → Home → Wrap test
    c. BONUS Shortcut to auto size: Ctrl + A → Double click line between rows and/or columns

4. **Hide variables** = show less on the screen at once e.g. modules not chosen
    a. Select columns to be hidden → right click in selected area → Hide

5. **Unhide variables** = show hidden columns
    a. Select columns before and after hidden columns or select all columns → right click in selected area → Unhide

6. **Change view** = zoom in or out of the screen
    a. Select the – or + sign in the bottom right of the screen

7. **Change number format** = tell Excel how to interpret numbers (type of data)
    a. Select cells to reformat → Home → under the "Number" section of the ribbon, there are different options: $ sign, % sign, Increase or Decrease decimal, etc.

Basic Excel Functions Continued:

Note: These steps DO change the data

8. **Check for duplicates** = endeavor to start with the most accurate and valid data
    a. Select column with unique ID → Home → Conditional Formatting → Highlight Cell Rules → Duplicate Values → Sort or Filter by cell color to easily locate and fix
    b. Can also sort/filter a few variables. Filter by zip, sort by city, etc.

9. **Copy AND label sheets** = for segmenting data, exploring data safely, other uses
    a. Right click on a sheet tab (bottom left of the screen) → Move or Copy → Select the sheet to copy → Select the box next to Create a cope → OK
        AND THEN
    b. Right click the sheet that was just created → Rename → Give it a name
    c. To **reorder** sheets, drag and drop to the desired location

PART 2:

Calculate Descriptive Statistics

Descriptive statistics for nominal data

Label	Formula
sample size aka "n"	=COUNTA (S2:S62)
missing	=COUNTBLANK (S2:S62)
sum (frequency for binary variables)	=SUM(S2:S62)
count (frequency for other nominal variables	=COUNTIF(S2:S62
percentage	=frequency/n

Descriptive statistics for ratio data:

Label	Formula
min	=MIN (CK2:62)
max	=MAX (CK2:62)
range	=max cell – min cell
mean	=AVERAGE(CK2:62)
median	=MEDIAN(CK2:62)
mode	=MODE(CK2:62)

Prioritize Variables:

1. First, **Find** the variable(s) of interest → Ctrl + F
2. Then, **Hide** other variables → Select columns to be hidden → right click in selected area → Hide
3. Calculate descriptive statistics for those variables
    a. **Sample size** aka **"n"** =COUNTA(E2:E62)
    b. **Missing data** =COUNTBLANK(E2:E62)
    c. For binary variables (i.e. 0 or 1 values) can use **Sum** =SUM(E2:E62)
    d. **Count** aka **frequency** =COUNTIF(E2:E62, 1)
    e. **Percentage** =frequency/n

Calculate Confidence Intervals in Excel:

Label	Formula	Formula
n [same as the n from Descriptive stats]	=E64	Sample size
count yes	=COUNTIF(E2:E61,1)	number of 1s in data
count no	=COUNTIF(E2:E61,".")	Number of decimal points
p	=E92/E91	Number of 1s over sample
1-p	=E93/E91	Number of decimal points over sample
C.L. [confidence level]	0.95 [could also be .99]	0.95 [could also be .99]
alpha [this is my margin of error, set at (1-C.I), in this case 1-.95=.05]	0.05 [could also be .01]	0.05 [could also be .01]
alpha/2 [the total of (1-C.I)/2 in this case .05/2=.025]	=E97/2	The alpha over 2
z score	1.96 [could also be 2.576]	1.96 [could also be 2.576]
standard error [statistical formula: (P*(1-P))/(n=)]	= SQRT((E94*E95)/(E91))	Square root of (Number of 1s over sample) times

		(Number of decimal points over sample) over the sample
margin of error [statistical formula: (Standard Error)*(Z score)]	=E100*E99	(Standard error) times Z score
lower limit [proportion - margin of error]	=E94-E101	(margin of error) minus (Number of 1s over sample)
upper limit [proportion + margin of error]	=E101+E94	(margin of error) Plus (Number of 1s over sample)

PivotTables in Excel:

First, there are some requirements that Excel demands in order to create a pivot table:

- Every column needs a variable name
- Cannot have a space between columns
- Cannot have a space between rows
- If the data change, you need to refresh the pivot table
- Clicking off of the pivot table makes the window disappear; click it to make it appear

1. **Insert a Pivot Table**
    a. Select a cell in the dataset → Insert → Pivot Table → OK = pivot table in a new sheet
    OR
    b. Select a cell in the dataset → Insert → Pivot Table → Select Existing Worksheet → Select where you want the pivot table to go → OK = pivot table in a specific location

2. **Rename Sheet**
    a. Right click sheet tab to rename → Rename → Enter a proper label for this sheet

3. **Drag/Drop Variables**
    a. Drag "StoreCity" to Row Labels
    b. Drag C7_1, C7_2, C8_1, C8_2 to Values
    c. For all variables in Values → Select arrow → Value Field Settings → Sum → OK
        i. Since we're working with binary variables, we can/need to use the sum function
    d. Rename labels (store type, variable names, etc.)

4. **Calculate Percentages**
    a. Select all cells in the PivotTable → Copy → Paste as Values in 2 locations
    b. Label table
    c. Enter equation and copy into all cells

5. **Compare city of Apple to the rest**
    a. Select all cities except for Apple → right click in the selected area → Group
    b. Rename group to "County" → Select cell → type County
    c. Roll up groups → Select the minus sign to collapse group
    d. Repeat Step 4. Calculate Percentages

6. **Compare Peaches to the rest**
    a. Select group → right click in selected area → Ungroup
    b. Select all cities except for Peaches → right click in the selected area → Group
    c. Rename group to "County" → Select cell → type County
    d. Roll up groups → Select the minus sign to collapse group
    e. Repeat Step 4. Calculate Percentages

7. BONUS:
    a. Double click in a PivotTable to create a segment of the data in a separate sheet

Resources

- Anne Emery workshops, blog, and other evaluation resources: http://annkemery.com/workshops/
- Mike "excelisfun" Girvin's YouTube video tutorials: https://www.youtube.com/user/ExcelIsFun
- Mr. Nystrom, YouTube videos on Stats https://www.youtube.com/user/MrNystrom
- American Evaluation Association courses: http://comm.eval.org/coffee_break_webinars/estudy
- Tobacco Control Evaluation Center: www.tobaccoeval.ucdavis.edu or 530-752-9951