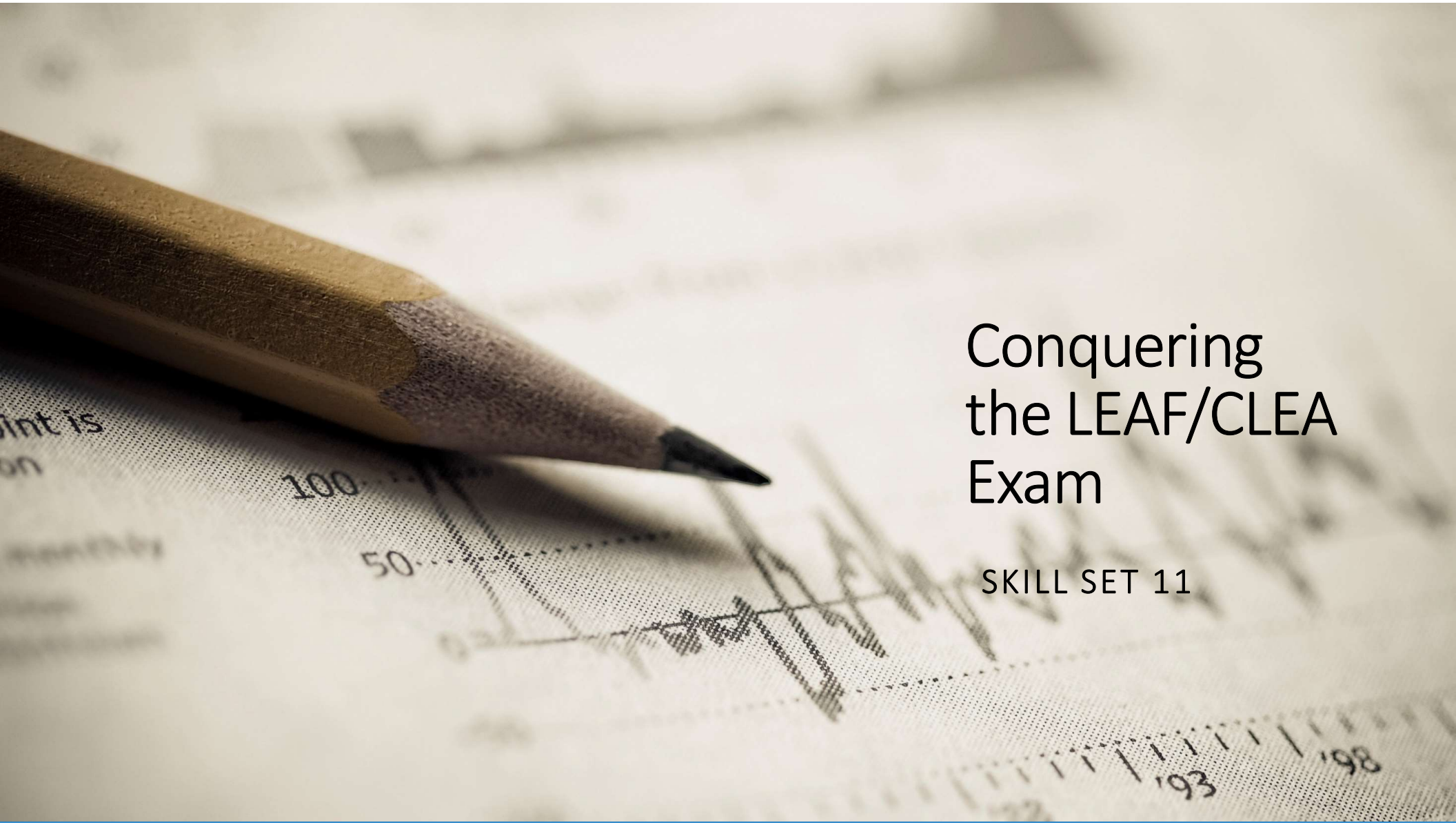

# Conquering the LEAF/CLEA Exam

SKILL SET 11

# About the Instructor/Course

- Instructor – Jenny Zawitz Jennifer.Zawitz@gmail.com
- CLEA Study Guide: https://iaca.net/wp-content/uploads/2021/06/CLEA-Skill-Sets_Study-Resources-051821.pdf
- LEAF Study Guide: https://iaca.net/wp-content/uploads/2021/06/en_LEAF-Core-Competencies_Study-Resources.pdf
- Exploring Crime Analysis: Readings on Essential Skills (3rd Edition) - IACA
- Each month will cover a different section of the study guide
- Intended as a supplement NOT a substitute for the texts and the Essential Skills classes
  - This course will help you focus your studying, but the courses and text will provide the actual understanding you need to pass the tests

# Spreadsheets….but why?

- Spreadsheets allow analyst to evaluate new datasets, calculate crime rate, produce graphs/charts, and quantify conclusions.
- Can be data repositories or used to pass data from one database to another
- Can filter, sort, clean, and aggregate data; perform statistical calculations, and create graphs/charts.
- New versions every 3-4 years. IACA recommends versions 2010 or later.

# When to Use vs. When Not to Use

## USE EXCEL WHEN:

- Calculating crime rate
- Calculating time series analysis for crime series and patterns
- Producing charts and graphs
- Calculating simple/summary statistics
- Automating routine calculations
- Developing simple models (forecasting)
- Answering one-time data requests

## USE SOMETHING ELSE WHEN:

- Complex networking of cells can get too convoluted and cause errors since cells can simply to told to refer to each other.
- Can multiply the impact of “off by one” errors. Errors can quickly add up.
- Can be difficult to verify or audit.
- Not great for managing data with regards to data entry or searching for detail.
- Not as robust a repository for information as many believe. Should not be a sole data storage or repository source (RMS).

# Limitations of Excel

- Not a relational database or a report writer
- Better to transfer post-calculation results to a different program (ex: PowerPoint for chart presentations)
- Designed for data calculation and analysis vs. data management (database/servers)
- Recall relational databases store and join multiple databases and keep categorized data in unique tables and relate the tables to show information. (ex: Access)
- Excel cells and tabs are independent of one another and thus less effective in as a relational database.
- With expansion of row/column capabilities, lag time in opening/saving files may occur.
- Not as efficient at processing large amounts of data. Ex: can't store a full year of CFS.

# Key Spreadsheet Commands

- Formatting: currency, date, time, percentage, fraction, or decimal
- Style: color and size of cell's background border and font
- Conditional formatting: allows values meeting a set of criteria to have a certain format/style
  - Common  when displaying percentage change
- Sorting: can sort by value, color, type
- Filtering: target specific data by removing unwanted records. Can be filtered by value or cell background color.  Can be done on multiple columns and then sorted.  Resetting will return data to the original form
- Major command ordered at the top of the screen under headings.  Can be customized.

# Common Simple Commands

| Command | What it does |
|---|---|
| Ctrl+a | selects all cells in the spreadsheet |
| Ctrl+c | Copies all selected cells |
| Ctrl+v | Pastes all copied cells |
| Ctrl+ | Enters today's data |
| Shift+Ctrl+ | Enters current time |

Can search online for Microsoft Excel shortcut for additional lists

Add-Ins allow you to add additional functionality – can add Analyst ToolPak to conduct regression, correlation, and other advanced statistical functions.

# Crime Analysis Formulas

- All formulas are calculations to be performed by referencing other cells in the spreadsheet.
- All formulas begin with an equal (=) sign.
- After the equal sign is the name of the formula and an open parenthesis. Specific components for the calculation are required in a specific order. If you aren't sure, click on the Formulas tab and then Insert Function (all the way to the left). This will walk you through the requirements for the formula and show you what it will look like before you insert it (including any errors).
- Each component separated by a comma.
- Formula must end in a closed parenthesis to indicate that it is complete.
- Ex: =IF(A1>B1,"YES","NO") (note no spaces). Read like a sentence. If cell A1 is greater than B1, then show the word YES in the cell. Otherwise, show the word NO.

# Common Crime Analysis Formulas

- Percent Change =(A2-B2)/B2 where A2 = new value, B2 = old value so (New-Old)/Old. To get the percentage, you can either format your cells to show the value as a percent or you can add *100 to your formula =((A2-B2)/B2)*100)
  - Common error for this formula is the divide by zero error (#DIV/0!) because you cannot divide any number by 0. If you had 0 homicides last year and 2 this year, you'll get this error. You can create an IF statement to correct this, or you can note this as N/A for percent change.
- Mean or Average =AVERAGE(A1:A22)
- Median =MEDIAN(A1:A22)
- Mode =MODE.SNGL(A1:A22) or =MODE.MULT(A1:A22)
  - Note these can be influenced by outliers. Mode formula depends on your version of Excel.
- Rounding (note for display purposes only – original number will still be stored)
  - Can round =ROUND(A1) round up =ROUNDUP(A1) round down =ROUNDDOWN(A1)

# VLOOKUP

- Far more complicated than a single slide for practical purposes, but for the exam here are the highlights.
- Function allows datasets to be joined. V stands for “vertical” so the datasets are joined by columns (as opposed to HLOOKUP which is for horizontal or row lookups)
- Four components to VLOOKUP
  - Cell in the data that is joined to the lookup table
  - Array or set of data in the lookup table (note that the field to be joined must be the left most column in the array)
  - The column number of the array to be returned
  - “True” returns the closest match and “False” returns an exact match
- =VLOOKUP(A3,Sheet1!A:B,2,FALSE)

# VLOOKUP Practical

| VLOOKUP Formula | |
|---|---|
| Number | Month |
| 1 | Jan |
| 2 | Feb |
| 1 | Jan |
| 1 | Jan |
| 3 | Mar |
| 6 | Jun |
| 9 | Sep |
| 12 | Dec |

- ‘=VLOOKUP(A3,Sheet1!A:B,2,FALSE)

# Formulas – Text Manipulation

- Concatenate: merging values in multiple cells into one cell.
  - =CONCAT(B1," ",B2," ",B3)  ex: entering an address.  If leave out spaces, everything put together.
  - If you have extra spaces in your data, use the =TRIM function to get rid of them with the cell in parenthesis being the cell where your concatenation was placed.
- Parsing: process of separating out a value in one cell into new values in multiple cells.
  - Text To Columns command
    - Select the data you want to separate or select the entire column as a whole
    - Find the Text to Column Command (Data tab → Text to Column)
    - In the Convert Text, choose Delimited if you have columns or spaces –or- Fixed Width if you want to set the columns yourself
    - If choose Delimited, select your delimiter (check the preview before you commit).  If choose Fixed Width, click on the location you want the breaks to create the solid line.  If you want to remove a solid line, double click it.
    - Click finish
    - NOTE: ALWAYS check to make sure you have enough blank cells to transfer your data.

# Excel Dates and Times

- Dates and times are stored as a number of days since January 1, 1900 plus a fractional portion of the 24 hour day : ddddd.tttttt. This is called the serial date or serial date-time.
- Ex: January 1, 2014 is 41640
- One second is 1/86400 so 0600 hrs is .25 or 25% of a 24 hour day. 1800 hrs would be .75.
- This makes performing calculations with dates/times easier, but determining the interval of date and time to calculate may be difficult. Can use formulas for this
- To add a year to a date, use formula =EDATE(A1,12) where A1 is the date and 12 represents the number of months.
- Similarly, =A1+TIME(1,2,3) where A1 is the original time and adds 1 hour, 2 minutes, and 3 seconds.
- Excel will not display negative time and will display ### as the error.

# Copying Formulas

- Formulas can be copies and pasted to other cells.
- Cells default to Relative References in column row format (A1, B1, C34). If you copy the Excel formula to a different location, it will change the relative reference cell to the new cell.
- Can change this to Absolute Reference by adding a $. This means the formula will always look for that value in a specified cell even if you move it to a different location. Absolute references are written as $A$1.
- $A1 means you should always refer to column A but the row number will change when you copy and paste.
- A$1 means you should always refer to row 1 when the formula is copied, but the column will change.
- Giant list of formulas on page 227-228 in book.

# Threshold Analysis

- When reporting crime numbers and percent changes, need to provide context. Is a 200% increase in homicide cause for alarm or is it simply to be expected? Ex: go from 1 homicide last year to 3 homicides this year as opposed to 50 last year to 150 this year. What if the previous 5 year homicide totals had been 400, 300, 250, and 275? Would that 150 number look a little better/50 number look odd?
- Threshold analysis can help analyst determine what is normal/to be expected and prioritize tasks. This type of analysis helps to determine where the spikes occur.
- Common standard for threshold analysis is 3 to 7 years of data or data sets. Any longer may not be as indicative of what is "typical" anymore.
- Let's create a threshold analysis

# Creating a Threshold Analysis

- To calculate the threshold analysis you need:
  - Mean =Average(B2:F2)
  - Standard Deviation (how spread out the data points are) =STDEVP(B2:F2)
  - Low range and high range (floor and ceiling of the data points using one standard deviation)
    =CONCAT(ROUND((H2-I2),0),"-",ROUND((H2+I2),0))
  - z-score squared (how many standard deviations the most recent value is from the average, either above or below) =ROUND(((G2-H2)/I2),2)
- Z-score of -2 or less means the value is abnormally low whereas a z-score of 2 or above indicates that the individual score is abnormally high. Z-score from -1 to 1 is normal.
- Using this, we see that 150 homicides is still outside the normal range and abnormally low for this area.

# Threshold Analysis

| Offense | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Percent Ch | Mean | St Dev | Range | Zscore | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Homicide | 400 | 300 | 250 | 275 | 50 | 150 | 200 | 255 | 114.4552 | -55-455 | -0.2 | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

# Pivot Tables

- Allows data to be aggregated or cross-tabulated.  Again, too extensive to map out here.
- Categorizes data into different dimensions.
- Must be at least one measure and one dimension.  They are defined by the data's field (column headers).  The measure is a type of aggregate (count, sum, etc.).  Dimension is the category and shows how the data will be divided and grouped.  These can be rows or columns or both.
- Pivot tables can be filtered by adding fields to the Report Filter.  Can also filter using drop downs.
- Can also subtotal different categories within your original table.  Very useful for quick summaries in reports.

# Pivot Tables

| Offense | Off. Age | Off. Sex | Sector |
|---|---|---|---|
| Homicide | 15 | M | 1 |
| Homicide | 42 | M | 1 |
| Robbery | 34 | M | 2 |
| Burglary | 53 | F | 1 |
| Robbery | 36 | F | 4 |
| Burglary | 17 | M | 2 |
| MV Theft | 15 | M | 1 |
| MV Theft | 16 | M | 4 |
| Burglary | 25 | M | 3 |
| Rape | 23 | M | 2 |
| Robbery | 36 | F | 4 |

# Charts

- Line and Bar Charts
  - Display a single category or multiple categories over time.
  - Multiple options for showing trend lines
- Time of Day and Day of Week Charts
  - Frequently misused. Tendency to summarize volume of incidents per hour or day and then use that to show which days are more likely or less likely to have a crime occur.
  - Can perform a tabulation by hour of day and day of week and show those results using a surface chart. Can use this to visualize density of incidents over a 168 hour week.
  - Can put into Word, Outlook, Publisher, or PowerPoint.
  - If paste the chart, must have access to the program where it is originally stored or it won't open – or – store the chart in the program that you copied it into.
- Dashboards
  - Viewer can make a series of selections to display the data that they wish. Can customize results.
  - Slicers introduced in 2010 for Pivot Tables to help this. Can substitute for report filters.

# Macros and Visual Basic

- Used for automation.
- Macros record keystrokes and mouse operations so they can be applied to similar situations later. Save time by automating repetitive tasks. Can also standardize data by applying formatting changes.
- Macros don't require writing programming language, but are very specific and apply only to the sheet you're working on when you record them. When you apply it to another worksheet, the screen will blink and it will repeat the task you recorded.
- Visual Basic for Applications (VBA) is a Microsoft Programming Language that allows users to write scripts to create new functions or commands in Excel. Complicated and requires knowledge of programming language.
- If your automation won't require user interaction, use macros. If your automation must provide feedback to the user or make decisions based on user input, then use VBA.

# Additional Functions

- Help can be found by pressing F1 or selecting the question mark icon in the upper right hand corner of the Excel window.
- You can access forums, blogs, videos, and websites
- IACA listservs are helpful for complex questions.
- Can also find webinars on IACA website
- Books dedicated to Microsoft Excel – Better Policing with Microsoft Office 2007

# Conclusions

- Read the books and take the classes to strengthen understanding.
- Check out Statistics for People Who Hate Statistics.
- Try to apply the things learned to your every day work to “make them stick”.
- Use the study guides.
  - https://iaca.net/about-clea/ (links for program outline and study guides here)
  - https://iaca.net/about-leaf/ (links for program outline and study guides here)
- Next month: Temporal Analysis (Skill Set 12)

## Any questions?

# Good luck to everyone taking the test next week!

Don’t worry – you’ll do a great job!

# Finding Formulas (Extras)

A list of formulas can be found under the Formulas tab

Hundreds of available formulas to do just about anything you need

Most common formulas are found under the tabs within the formula tab.

- AutoSum: Basic functions (Sum, Avg, Count, Min, Max)
- Recently Used: Your recently used formulas/functions
- Financial: More useful in a business setting
- Logical: AND, IF, FALSE, TRUE, NOT – will discuss in depth in November
- Text: Converts Data to Text
- Date & Time: Formulas for…date and time. :-/
- Lookup & Reference: Formulas to direct you another location on your sheet – October class
- Math & Trig: Mathematical functions
- More Functions: Stats/Informational/Other Functions

# FORMULA Etiquette (Extras)

Start all formulas with “=“

Text has to be in quotes.  This includes spaces “ “

Hover over formula if you are unsure, and it will tell you what it does and walk you through the steps for completion.

Don’t forget PEMDAS

- Parentheses
- Exponents
- Multiplication
- Division
- Addition
- Subtraction

# Common Mathematical Formulas (Extra)

=SUM(A1:A4) – Adds cells selected

=COUNT(A1:A4) – Provides a count of the cells that contain numbers

=COUNTA(A1:A4) – Provides a count of all cells that aren't empty

=COUNTBLANK(A1:A4) – Provides a count of the number of blank cells

=COUNTIF(A1:A4) – Provides a count of the number of cells that meet a certain condition

=COUNTIFS(A1:A4) – Provides a count of the number of cells that meet a set of conditions

=((B1-A1)/A1)*100 – Percent change (new-old)/old

=ROUND(B1/A1,2) – rounds to the nearest hundredth

=A3-A2 – Days Between Incidents

# Statistical Functions (Extra)

=AVERAGE(A1:A4) – returns mean

=MEDIAN(A1:A4) – returns median

=MODE(A1:A4) – returns mode

- =MODE.MULT(A1:A4) (Ctrl, Shift, Enter) and =MODE.SNGL(A1:A4) in 2010+

=MAX(A1:A4) – returns maximum value in those cells

=MIN(A1:A4) – returns minimum value in those cells

=STDEV.P(A1:A4) – Standard deviation of entire population

=STDEV.S(A1:A4) – Standard deviation of sample

=STDEV(A1:A4) – if using 2007

# Common Text Formulas (Extra)

=CONCATENATE(A1," ",B1) – combines cells as text.  Good for addresses

=PROPER(A1) – removes all capital letters

=TRIM(A1) – removes extra spacing from cells

=TEXT – converts value to text in specific format (see next slide)

=LEFT(A1,2) – grabs the left two most characters in the reference cell

=RIGHT(A1,2) – grabs the right two most characters in the reference cell

=SUBSTITUTE(A1,”Street”,”St”)

Text to Column Function – separates text in a cell

# Common Dates/Times (Extra)

Day of Week

- =TEXT(A1,”ddd”) =TEXT(A1,”dddd”)

Month

- =TEXT(A1,”mmm”) =TEXT(A1,”mmmm”)

Year

- =TEXT(A1,”yy”) =TEXT(A1,”yyyy”)

Time of Day in Hours

- =TEXT(A1,”HH”)