Research 201

Using Advanced Excel Tools for Managing Your Data

Sean Kolhoff, PhD
Senior Research Analyst
American Academy of PAs

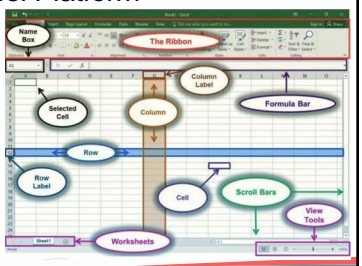
1

Excel Functions

2

Navigating the Excel Platform

- This presentation assumes you have a basic understanding of the Excel platform.
- If you are a true Excel novice, I recommend going to YouTube, Coursera, or LinkedIn Learning for intro information.

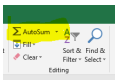


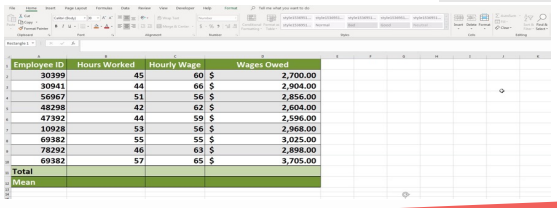
3

Excel Tips and Tricks

Using built in tools for quick summary stats

- Functions auto-update when numbers change
- Need correct cell references (be careful when adding/removing columns)



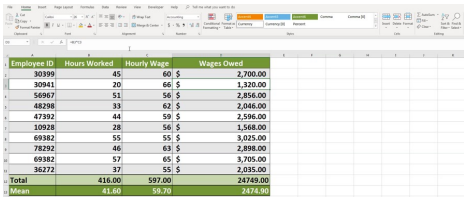


Employee ID	Hours Worked	Hourly Wage	Wages Owed
30399	45	60	$ 2,700.00
30941	44	66	$ 2,904.00
56967	51	56	$ 2,856.00
48298	42	62	$ 2,604.00
47392	44	59	$ 2,596.00
10928	53	56	$ 2,968.00
69382	55	55	$ 3,025.00
78292	46	63	$ 2,898.00
69382	57	65	$ 3,705.00
Total			
Mean			

4

Excel Tips and Tricks

Add conditional formatting to make cells stand out



Employee ID	Hours Worked	Hourly Wage	Wages Owed
30399	45	60	$ 2,700.00
30941	20	66	$ 1,320.00
56967	51	56	$ 2,856.00
48298	33	62	$ 2,046.00
47392	44	59	$ 2,596.00
10928	28	56	$ 1,568.00
69382	55	55	$ 3,025.00
78292	46	63	$ 2,898.00
69382	57	65	$ 3,705.00
36272	37	55	$ 2,035.00
Total	416.00	597.00	24749.00
Mean	41.60	59.70	2474.90

5

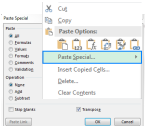
Manipulating/Rotating Data

The "Transpose" special paste option can let you turn rows into columns (and vice versa)!

Copy -> Right Click -> Paste Special -> Transpose

Be aware if you are copying cells that have formulas/reference other cells

NOTE: you cannot *cut* and transpose



Employee ID	Hours Worked	Hourly Wage	Bi Weekly Wages Owed
30399	45	60	$ 2,700.00
30941	44	66	$ 2,904.00
56967	51	56	$ 2,856.00

Employee ID	30399	30941	56967
Hours Worked	45	44	51
Hourly Wage	60	66	56
Bi Weekly Wages Owed	$ 2,700.00	$ 2,904.00	$ 2,856.00

6

Relative, Absolute, and Mixed Cell References

Relative Cell Reference (ex. A1)

- **Changes** if the formula is moved or copied to a new cell
- *Useful when*: carrying out formula when the values do not matter
  - i.e. Hours Worked (B1) x Hourly Wage (B2) = Compensation (B1 * B2)

Absolute Cell Reference (ex. A1)

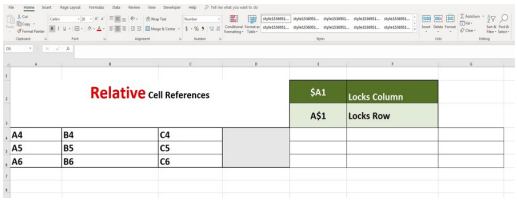
- If you move or copy the formula to a new cell, it will **always** reference the absolute cell reference
- *Useful when:* carrying out a formula that includes a value that is fixed
  - *i.e.* Hours Worked x Hourly Wage x Overtime = Compensation

Mixed Cell Reference

- **A$1** relative column, absolute row
  - If moved or copied, the **column changes**, but the row does not
- **$A1** absolute column, relative row
  - If moved or copied, the **row changes**, but the column does not

7

So, what does this look like?



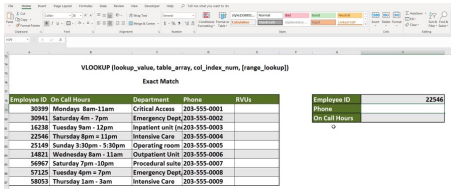
Where do we go from here?

8

Excel Tips and Tricks

Using VLOOKUP to lookup a value in a table array and retrieve a matched value

- **VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])**

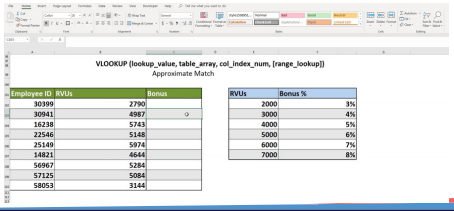


9

Excel Tips and Tricks

Using VLOOKUP approximate match to lookup a value in a table array and retrieve a matched value

- **[range_lookup]) =TRUE (Data must be sorted ascending)**



VLOOKUP (lookup_value, table_array, col_index_num, [range_lookup])

Approximate Match

Employee ID	RVUs	Bonus
30399	2790	
30941	4987	
16238	5743	
22546	5148	
25149	5974	
14821	4644	
56967	5284	
57125	5084	
58053	3144	

RVUs	Bonus %
2000	3%
3000	4%
4000	5%
5000	6%
6000	7%
7000	8%

10

VLOOKUP: Pros and Cons

Pros

- Easy to use
- Effective on small data sets

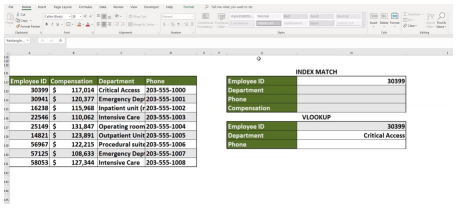
Cons

- Easy to make mistakes
  - *Requires mixed and absolute references*
  - *Adding/removing columns leads to invalid references*
- Limited functionality and flexibility
  - *You can only look up values on the "right" of the reference value*
  - *Only works when the data is vertically arranged ("**V**" lookup)*
- Requires manual assignment, which can be cumbersome.
  - *Manual work is slow, other functions (like INDEX) have more automation*

11

Excel Tips and Tricks

Using INDEX vs. VLOOKUP



Employee ID	Compensation	Department	Phone
30399	$ 117,014	Critical Access	203-555-1000
30941	$ 120,377	Emergency Dep	203-555-1001
16238	$ 115,968	Inpatient unit (r	203-555-1002
22546	$ 110,062	Intensive Care	203-555-1003
25149	$ 131,847	Operating room	203-555-1004
14821	$ 123,891	Outpatient Unit	203-555-1005
56967	$ 122,215	Procedural suite	203-555-1006
57125	$ 108,633	Emergency Dep	203-555-1007
58053	$ 127,344	Intensive Care	203-555-1008

INDEX MATCH

Employee ID	30399
Department	
Phone	
Compensation	

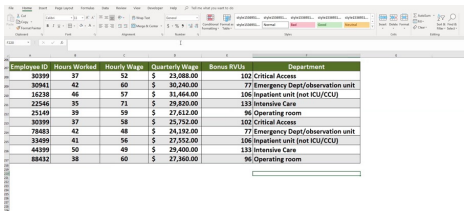
VLOOKUP

Employee ID	30399
Department	Critical Access
Phone	

12

Excel Tips and Tricks

Pivot tables and charts



Employee ID	Hours Worked	Hourly Wage	Quarterly Wage	Bonus RVUs	Department
30399	37	52	$ 23,088.00	102	Critical Access
30941	42	60	$ 30,240.00	77	Emergency Dept/observation unit
16238	46	57	$ 31,464.00	106	Inpatient unit (not ICU/CCU)
22546	35	71	$ 29,820.00	133	Intensive Care
25149	39	59	$ 27,612.00	96	Operating room
30399	37	58	$ 25,752.00	102	Critical Access
78483	42	48	$ 24,192.00	77	Emergency Dept/observation unit
33499	41	56	$ 27,552.00	106	Inpatient unit (not ICU/CCU)
44399	50	49	$ 29,400.00	133	Intensive Care
88432	38	60	$ 27,360.00	96	Operating room

13

Data Visualization

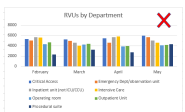
AAPA
American Academy of PAs

14

Visualization on Excel: What charts to use

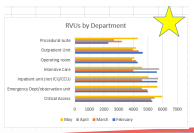
Clustered Column

- 2-4 data groups (series)
- Compared on the same data *type*
- Compensation differences between men and women



Bar vs. Column Chart

- Can better display multiple data series
- Bar charts allow you to have longer labels

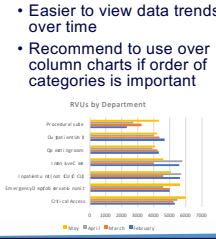


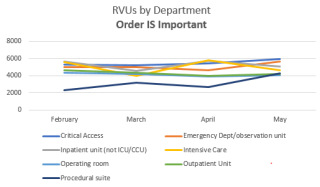
15

Visualization on Excel: What charts to use

Line Charts

- Easier to view data trends over time
- Recommend to use over column charts if order of categories is important



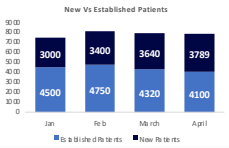


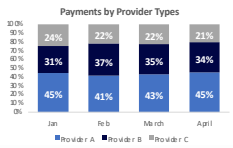
16

Visualization on Excel: What charts to use

Stacked Column (values) vs. Stacked Column (%)

- Both compare the proportion of the series within each category
- Stacked values make the trend in the absolute difference easier to interpret
- Stacked % emphasizes proportion of categories



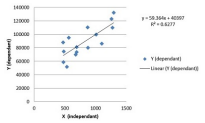


17

Visualization on Excel: What charts to use

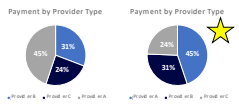
Scatter Chart

- Charts relationship between two variables
- More data points, stronger chart
- Illustrates "why" factors are related



Pie Chart

- Composition of a single variable
- Works best with less than 5 categories
- Pie slices should be ordered from largest to smallest

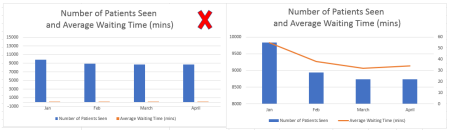


18

Visualization on Excel: What charts to use

Combination Charts

- Best used when your data series are two different *types* of data
  - For example, measuring % class attendance and average test score
- Data would not be interpreted well on the same scale/axis



19

Communicating your Findings

AAPA
American Academy of PAs

20

Communicating your findings

The language you use to share your results is related to how well you can leverage your data.

You are not trying to sell the *numbers*; you are trying to tell a story.

- Communicate the big picture of the results through your tables and figures
- Why does this matter? What should be done?

21

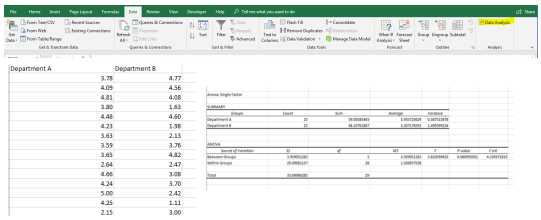
Stress the importance of the data (numerically)!

- Excel can be used to calculate statistical tests (which lets us say if the interactions between numbers matter)
- The Data Analysis button can be used to calculate:
  - ANOVA
  - Correlation
  - Moving averages
  - Regression
  - T-Test
  - Z-Test

Excel (especially scatter and line charts) can be great for forecasting and trend analysis

22

Finding the Data Analysis Tool



Department A	Department B
3.78	4.77
4.09	4.56
4.81	4.08
3.80	1.63
4.48	4.60
4.23	1.98
3.63	2.13
3.59	3.76
3.65	4.82
2.64	2.47
4.66	3.08
4.24	3.70
5.00	2.42
4.25	1.11
2.15	3.00

23

Communicating your findings

Here are some general tips for communicating your research

- Avoid jargon
  Do not assume the audience all understands an industry-specific (or workplace specific) term/acronym. "Plain" language is easier to follow.
- Be concise
  Keep your points short and include only what is necessary. Your audience can easily lose their "train of thought".
- Know your audience
  This is key. Relate your message/data to what your AUDIENCE cares about.
- Fact check yourself
  Understand your data. Be prepared for questions. Some people in the audience may try to poke holes in your logic.
- Prepare for anything
  You might lose power, the technology might fail, or you may run into someone interested in the hall. Be ready to adapt your message to the situation.

24

Communicating your findings

Looking for more? Try Grice's Maxims of Conversation

Maxim of Quantity: Be informative but say no more than is needed.
Maxim of Quality: Be truthful and say nothing that is not supported by your data/evidence.
Maxim of Relation: Be relevant to the discussion at hand.
Maxim of Manner: Be clear, brief, and orderly.

25