PUAF 610
DISCUSSION SECTION 1

Wednesday, September 5, 2012
Mariel Borowitz
mariel@umd.edu
Introductions

- About Me

- About You
  - Name
  - Where you’re from
  - Your concentration area at UMD
  - Why you decided to come to the policy school (in two sentences or less)
  - A time you’ve used excel in the past (if any)
  - How much you love Excel on a scale of 1-10
    - 10 being very in love with Excel, 1 being the minimum amount of love possible for Excel
How I Plan to Run Discussion

• This is not a second lecture
  • Expand on/ reinforce lecture material
  • Talk about readings
  • Work on problem set
  • Answer your questions

• Send me questions in advance whenever possible
  • Lecture, Pset, Readings – anything related to class
  • Email me any time: mariel@umd.edu
Plan for Today

• Microsoft Excel Tutorial
• Lecture 1 Review
• Discuss Readings
• Problem Set
Microsoft Tutorial

• Basics
• Functions
• Graphs
• Pivot Tables
• Installing the data analysis tool pack

• Not covering everything, but we’ll add to it as we go along
Worksheets, Columns, Rows, and Cells
Insert and Delete
# Autofill

![Excel spread sheet with data autofilled](image)

<table>
<thead>
<tr>
<th>First Name</th>
<th>Rank</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Nora</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Oscar</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>Alicia</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>Emily</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>Forrest</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>Lorie</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>Mabel</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>Ian</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>Julie</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>Herman</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>Paul</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>Katie</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Bob</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Corey</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>Dana</td>
<td>16</td>
<td>26</td>
</tr>
</tbody>
</table>
## Sort

A screenshot of a spreadsheet is shown with data sorted by first name and then by rank. The top left of the screen displays the sort options with checkboxes for selecting sort criteria. The top right of the screen displays the sorted data. The bottom left is a pop-up window showing the sort options in detail, including 'My list has headers' and options to sort by first name, then by rank with different sorting orders. The bottom right shows the sorted data in the spreadsheet with columns for 'First Name', 'Rank', and 'Age'. The data includes names like Grant, Nora, Oscar, and more, each with their age.
Formats
# Formulas

![Excel Spreadsheet]

<table>
<thead>
<tr>
<th>First Name</th>
<th>Rank</th>
<th>Age</th>
<th>Pocket Change</th>
<th>Birthdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alicia</td>
<td>4</td>
<td>26</td>
<td>$5.45</td>
<td>1/14/86</td>
</tr>
<tr>
<td>Bob</td>
<td>14</td>
<td>26</td>
<td>$20.00</td>
<td>3/4/86</td>
</tr>
<tr>
<td>Corey</td>
<td>15</td>
<td>26</td>
<td>$150.43</td>
<td>11/1/85</td>
</tr>
<tr>
<td>Dana</td>
<td>16</td>
<td>26</td>
<td>$0.26</td>
<td>4/25/86</td>
</tr>
<tr>
<td>Emily</td>
<td>5</td>
<td>26</td>
<td>$5.97</td>
<td>5/13/86</td>
</tr>
<tr>
<td>Forrest</td>
<td>6</td>
<td>26</td>
<td>$9.08</td>
<td>9/1/86</td>
</tr>
<tr>
<td>Grant</td>
<td>1</td>
<td>26</td>
<td>$112.02</td>
<td>10/11/85</td>
</tr>
<tr>
<td>Herman</td>
<td>11</td>
<td>26</td>
<td>$56.70</td>
<td>9/2/86</td>
</tr>
<tr>
<td>Ian</td>
<td>9</td>
<td>26</td>
<td>$35.94</td>
<td>2/28/86</td>
</tr>
<tr>
<td>Julie</td>
<td>10</td>
<td>26</td>
<td>$1.56</td>
<td>6/19/86</td>
</tr>
<tr>
<td>Katie</td>
<td>13</td>
<td>26</td>
<td>$349.30</td>
<td>5/12/86</td>
</tr>
<tr>
<td>Lorie</td>
<td>7</td>
<td>26</td>
<td>$17.00</td>
<td>4/5/86</td>
</tr>
<tr>
<td>Mabel</td>
<td>8</td>
<td>26</td>
<td>$23.00</td>
<td>8/10/86</td>
</tr>
<tr>
<td>Nora</td>
<td>2</td>
<td>26</td>
<td>$42.42</td>
<td>1/15/86</td>
</tr>
<tr>
<td>Oscar</td>
<td>3</td>
<td>26</td>
<td>$8.96</td>
<td>4/10/86</td>
</tr>
<tr>
<td>Paul</td>
<td>12</td>
<td>26</td>
<td>$25.96</td>
<td>12/15/85</td>
</tr>
</tbody>
</table>

=sum(D2:D17)
## Formulas

A formula in Excel is a mathematical expression that performs calculations on data in a worksheet. Formulas always begin with an equal sign (=) and can be used to perform various calculations, such as addition, subtraction, multiplication, division, and more complex operations. Excel supports a wide range of functions that can be used in formulas to perform specific calculations. These functions include statistical, financial, engineering, and more.

### Example Formulas

1. **Summation**
   - Formula: `=SUM(D2:D17)`
   - This formula calculates the sum of the values in cells D2 through D17.

2. **Average**
   - Formula: `=AVERAGE(D2:D17)`
   - This formula calculates the average of the values in cells D2 through D17.

3. **VLOOKUP**
   - Formula: `=VLOOKUP(D3, A1:B10, 2, FALSE)`
   - This formula looks up the value in column A for the given condition in column D and returns the value in the second column of the lookup table.

### Excel Functions

Excel offers a vast library of functions that can be used in formulas. Some commonly used functions include:

- **SUM**
- **AVERAGE**
- **VLOOKUP**
- **IF**
- **MAX**
- **MIN**
- **STDEV**
- **COUNT**
- **TEXT**

These functions help in performing various data analysis tasks efficiently.
Formulas

• Basic Functions
  • Example of use: Step B is due 10 days after Step A
  • Equal: =B2
  • Addition: =A1+10
  • Subtraction: =A1-10
  • Multiplication, Division, etc.

• Minimum
  • Example of use: Step A must be done before either Event B or Event C occurs (whichever comes first)
  • Minimum: =Min(B2,C2)

• Maximum
  • Example of use: Step A is due after Event B and Event C are complete
  • Maximum: =Max(B1,C1)
Formulas

• SUM
  • Example of use: Want to add all weights to show they add to 100
  • SUM: =SUM(A2:A5) – Adds numbers in A2 through A5

![Excel sheet with formula]

A6

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>34%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100%</td>
<td></td>
<td>=SUM(A2:A5)</td>
</tr>
</tbody>
</table>
Autocomplete Formulas

• Relative References
  • This is the norm for copy and paste/autocomplete
  • It copies the formula using cells in the same relative location
    • C1=A1+B1
    • C2=A2+B2
    • C3=A3+C3

• Absolute References
  • If you want to continue referencing one specific cell (or row or column), you can add a $ to the formula
    • C1=A1+B$1
    • C2=A1+B$1
    • C3=A1+B$1
Formulas

Relative References in autocomplete

<table>
<thead>
<tr>
<th>First Name</th>
<th>Number of Apples</th>
<th>Number of Oranges</th>
<th>Number of Fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Bob</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Corey</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Dona</td>
<td>7</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Emily</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Forrest</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Grant</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Herman</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Ian</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Julie</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Katie</td>
<td>8</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Lorie</td>
<td>3</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Mabel</td>
<td>8</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Nera</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Oscar</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Paul</td>
<td>8</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Formulas

- **AND**
  - Example of use: Step C complete if Step A and Step B complete
  - AND: =AND(A1,B1)

- **OR**
  - Example of use: Step C complete if Step A or Step B complete
  - OR: =OR(A1,B1)

- **IF**
  - Example of use: If organization A and organization B are the same, then due date C is Jan. 1, 2009, otherwise due date C is February 1, 2009
  - IF: 
    - IF(logical_test, value_if_true, value_if_false)
Example: Describe this function in words:

\[
\text{IF(AND(A1>0, A2>0), 100, 0)}
\]

Translation

- If A1 and A2 are both greater than zero, the result is 100
- Otherwise the result is zero

Given the following values, give the result of the function:

- 1) A1=5, A2=10
- 2) A1=0, A2=5
- 3) A1=0, A2=0

Answers: 1)100; 2)0; 3)0
Graphs/ Charts

- Each column is a variable
- Each row is an observation
- Highlight data and click a type of chart
Graphs/Charts

• Ask five or more people near you:
  • Their first name
  • Their eye color
  • Number of pennies they have with them
  • Number of uncles they have
  • Whether one of those uncles is named Mike

• Working on your own or in a group, use this data to make:
  • A bar or column chart
  • A pie chart
  • A scatter chart
  • One other type of chart
Pivot Tables

- Pivot tables allow you to quickly view the data in different ways.
- We’ll introduce this now and look at it in more detail in the future.
- Select data and click on “Pivot Table”
Pivot Tables

- Pivot tables have four areas:
  - Report Filter
    - This area contains the fields that enable you to page through the data summaries shown in the actual pivot table by filtering out sets of data – they act as the filters for the report. For example, you may choose to filter by “eye color” allowing you to show all eye colors, or just one, like “brown.”
  - Column Labels
    - This area contains the fields that determine the arrangement of data shown in the columns of the pivot table.
  - Row Labels
    - This area contains the fields that determine the arrangement of data shown in the rows of the pivot table.
  - Values
    - This area contains the fields that determine which data are presented in the cells of the pivot table – they are the values that are summarized in the last column (totaled by default)
- The information button on each dataset allows you to change the type of info shown
Pivot Tables

• Using the data you collected, create a pivot table
• Try different combinations of filters, rows, columns, and values
• After five minutes, we will share with the class any interesting combinations – be prepared to say what you put in each category
Install the Data Analysis Add-in

Excel 2010 – PC
1. Click the Microsoft Office Button, and then click Excel Options
2. Click Add-ins, and then in the Manage box, select Excel Add-ins.
3. Click Go.
4. In the Add-ins available box, select the Analysis ToolPak check box, and then click OK.
   • If Analysis ToolPak is not currently listed in the Add-ins available box, click Browse to locate it.
   • If you get prompted that the Analysis Toolpak is not currently installed on your computer, click Yes to install it.
5. After you load the Analysis ToolPak, the Data Analysis command is available in the Analysis group on the Data tab.

Excel 2010 – Mac
• Search for “StatPlus” and follow the instructions on the website to download
Interested in more about Excel?

- Multiple useful tutorials online at all levels
When all else fails...

Dear various parents, grandparents, co-workers, and other "not computer people."

We don't magically know how to do everything in every program. When we help you, we're usually just doing this:

1. Start
2. Find a menu item or button which looks related to what you want to do.
   - If you can't find one, pick one at random.
   - If you've tried them all, use Google. The name of the program plus a few words related to what you want to do. Follow any instructions.
3. Have you been trying this for over half an hour?
   - Yes: ask someone for help or give up.
   - No: click it.
4. Did it work?
   - Yes: you're done!
   - No: ask someone for help or give up.

Please print this flowchart out and tape it near your screen. Congratulations; you're now the local computer expert!
Lecture 1 Review

- Limits to Human Rationality
- Definition of Statistics
- Types of Variables
- Types of Data
- Tables and Graphs – Common Errors
- Significant Digits
Limits to Human Rationality

• In small groups, discuss and list some of the limits to human rationality that were discussed in lecture (try to get to three - think back to the example questions!)
  • Likelihood of rare events
  • Use extraneous information
  • See cause and effect relations that aren’t there
Limits to Human Rationality

- Put these in order of likelihood of occurrence (most likely to least likely):
  - Getting hit by lightning in your lifetime = 1/10,000
  - Winning the Powerball Jackpot = 1/175,223,510
  - An individual will die by skydiving (per jump) = 1/100,000
  - Becoming the President of the United States (assuming you’re a U.S. citizen) = 1/20,000,000
  - Odds of winning an Oscar = 1/7,000,000

Sources:
http://www.lightningsafety.noaa.gov/medical.htm
http://www.powerball.com/powerball/pb_prizes.asp
http://adventure.howstuffworks.com/skydiving8.htm
Definition of Statistics

• What is statistics?
  • Collecting, analyzing, and presenting numerical facts or data

• What is the focus of descriptive statistics?
  • Organizing and summarizing data

• What is the focus of inferential statistics
  • Drawing general conclusions from sample data
Types of Variables

• What are the two most broad types of variables/data? Name the two types and an example of each
  • Numerical and Categorical

• What are the two types of numerical variables? Examples?
  • Discrete vs. Continuous

• What are the two types of Continuous Variables? Examples?
  • Interval vs. Proportion

• What are two types of categorical variables?
  • Ordinal vs. Nominal

• What are dichotomous variables? They are a subset of which other category?
  • Nominal variables with only two values
Types of Variables

- Classify each of these variables:
  - Gender
    - Categorical, Nominal, Dichotomous
  - Age
    - Numerical, Continuous, Interval
  - Salary
    - Numerical, Continuous, Interval
  - Number of Children
    - Numerical, Discrete
  - Mode of transportation used to get to class
    - Categorical, Nominal
  - Level of agreement (strongly agree, agree, disagree, strongly disagree)
    - Categorical, Ordinal
  - Ratio of men to women in each undergrad class
    - Numerical, Continuous, Proportion
Types of Data

• What is cross-sectional data?
  • Looks at one set (cross-section) of observations (members of the sample) at one point in time

• What is longitudinal/ time-series data?
  • Follows one or more observations (members of the sample) through time

• What is panel data?
  • Follows a cross-section of observations through time (cross-section and time-series data combined)
Tables and Graphs

• In small groups, discuss: what are some common errors made on tables and graphs? (Think of at least 5)
  • Errors in significant digits
  • Chart junk/ Unnecessary 3D
  • Non-linear axes
  • Axes that don’t start at zero
  • Data not in context
  • Data not adjusted for inflation
  • Data not adjusted for population growth
  • Confusing interpretation due to use of area or volume rather than height
Tables and Graphs
Tables and Graphs

**My Hobby: Extrapolating**

As you can see, by late next month you’ll have over four dozen husbands.

Better get a bulk rate on wedding cake.
Tables and Graphs

Number of living humans who have walked on another world.
LAKES AND OCEANS

DEPTHS AND ANIMAL DISTRIBUTION
ARE TO SCALE. HORIZONTAL DISTANCE IS NOT.

1,000 m
2,000 m
3,000 m
4,000 m
5,000 m
6,000 m
7,000 m
8,000 m
9,000 m
10,000 m
11,000 m
12,000 m

PUERTO RICO TRENCH
MARIANAS TRENCH
MAUNA KEA, HAWAII
MARIANAS TRENCH (INCREASING HORIZONTAL SCALE)
MAUNA KEA, HAWAII (INCREASING HORIZONTAL SCALE)

OIL

10,000 ft
20,000 ft
30,000 ft
40,000 ft
50,000 ft
60,000 ft
70,000 ft
80,000 ft
90,000 ft
100,000 ft

Significant Digits

• 4.4 + 5.16 = 9.6
• 100 - 1 = 100
• 43 - 6 = 37
• 65,000,000 + 12 = 65,000,000
• 890 - 10 = 880
• 6,400 + 205 = 6,600

• 5 \times 10 = 50
• 4.32 \times 100 = 400
• 4.32 \times 101 = 436
• 8 / 2 = 4
• 900 / 50 = 20
• 4040 / 40 = 100
• 404 / 4.00 = 101
Significant Digits

• What is the rule for addition/subtraction?
  • For addition/subtraction, the answer has significant digits only in the decimal places where all of the original numbers have significant digits [Look at where the digits are]

• What is the rule for multiplication/division?
  • For multiplication and division, the number of significant digits in the answer is the same as in the original number with the fewest significant digits [Count the number of digits]

• When should you round the digits you’re working with?
  • Rounding off is the last step
Readings

- Carlberg – Introduction
- Kahneman - Introduction
“To shoehorn statistics and Excel into 400 pages or so takes some picking and choosing.”

No background in statistics assumed

Watch out for vocabulary
- Excel: Cell = intersection of a row and a column
- Statistics: Cell = specific sub-group (placebo vs. test group)

Excel lets you focus on concepts, not calculations

Excel is not perfect, but it is accurate to a reasonably high level

Book will deal with descriptive and inferential statistics
Carlberg - Introduction

• Discuss: What are some of the positive and negative aspects to using Excel for statistics?
Kahneman - Introduction

• “I hope to enrich the vocabulary people use when they talk about the judgments and choices of others, the company’s new policies, or a colleague’s investment decisions.

• Importance of intuition and impressions

• Worked with colleague to document systematic errors in thinking of normal people

• Fast and slow thinking described as two characters: System 1 and System 2
Kahneman - Introduction

• Consider the letter K.
• Is K more likely to appear as the first letter in a word OR as the third letter?

• How or why might an individual’s answer to this question be biased?
Problem Set

1. In the student survey distributed during our first class meeting, list the variable type for questions 1, 2, 5, 6, 7, 8, and 10.

2. Find an example of a graph in a newspaper or magazine and comment on its design, noting features that you find deficient or commendable. Bonus points if you can redraw the graph and significantly improve it.

3. Find an example of a table or other use of numerical data in a newspaper or magazine and comment on the use of significant digits. Bonus points if you find an egregious abuse of significant digits (e.g., the meteorite example in the class notes).
Questions?

"You've stumped me with that question, I think that's something you need to Google."

"I know nothing about the subject, but I'm happy to give you my expert opinion."

Arthur answers the eternal questions.

Arthur, when will we know all the answers?

Sometimes after we've asked all the questions.