Data Analysis using Excel

Advanced Information Systems and Business Analytics for Air Transportation

M.Sc. Air Transport Management

June 1-6, 2015

Slides prepared by Benny Mantin

Background

- Excel is a spreadsheet application
- It has great functionality:
  - Calculation
  - Graphing
  - Pivot tables
  - VB Programming (macros)
- Main advantage: wide spread
- Useful Add-ins
  - Make sure you have Data Analysis and Solver installed

Some useful shortcuts

Key	Description
CTRL+PgUp (or PgDn)	Switches between worksheet tabs
F9	Calculates all worksheets in all open workbooks.
Ctrl+Shift+Arrow Key	Extend the selection to the last cell with content in row or column.
Shift+Arrow Keys	Extend the selection by one cell.
Ctrl+Shift+Home	Extend the selection to the beginning of the worksheet.
Ctrl+Shift+End	Extend the selection to the last used cell on the worksheet (lower-right corner).

Complete list of shortcuts:
http://www.shortcutworld.com/en/win/Excel_2010.html

Large data sets

- Excel is not best choice for working with large data sets.

- We will work with DOT's datasets.

- Specifically, we will experiment with On-Time Performance
http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time

- The file available contains the entire data for the month of January, 2014.

On-Time Performance

- The (slightly cleaned) data contains information on:
    - Time period
    - Airline
    - Origin
    - Destination
    - Departure Performance
    - Arrival Performance
    - Cancellations and diversions
    - Flights summaries
    - Cause of Delay
    - (Gate return information)
    - (diverted airport information)

On-Time Performance

- Given the data available at hand, we may be interested in exploring some characteristics of flight delays.
- Here are some ideas:
  - What is the average delay (in minutes)?
  - Does it change by the day of the week?
  - Are some airlines (airports) better than others?
  - Does a departure delay result with an arrival delay?
  - What is the distribution of delays?
    - Is there evidence that airline game the system? (i.e., slightly more 15 min than 16 min)
  - Does it depend on the distance? Other factors?
  - What is the major cause for delays?

On-Time Performance

- What other data sources might be of relevance?
  - Weather (NOAA)
    - Extracting weather data can help understand underlying patterns
  - Transactions (DB1B)
    - Can lead to other interesting exploration questions:
      - Do airline price markets differently based on expected delays? Based on delay uncertainty? Cancelled/diverted flights?

Excel activity

- In this demonstration we will:
  - Work with large a dataset
  - Clean the data
  - Insert table
  - Execute pivot tables
  - Draw histograms
  - Work with functions
  - Run regressions
  - Construct a data model