Data Analysis and Presentation

Carey Williamson

Department of Computer Science

University of Calgary

UNIVERSITY OF CALGARY

Data Analysis and Presentation

- There are many “tricks of the trade” used in data analysis and results presentation
- A few will be mentioned here:
  - statistical analysis
  - multi-variate analysis
  - ANOVA
  - tabular presentation of results
  - graphical presentation of results

UNIVERSITY OF CALGARY

Statistical Analysis

- "Math and stats are your friends!!!" CW
- There are lots of "standard" techniques from mathematics, probability, and statistics that are of immense value in performance evaluation work:
  - confidence intervals, null hypotheses, F-tests, T-tests, linear regression, non-linear regression, least-squares fit, maximum likelihood estimation (MLE), correlation, time series analysis, transforms, Q-Q plots, EM...
  - working knowledge of commonly-observed statistical distributions

UNIVERSITY OF CALGARY

Data Analysis Checklist (meta-level)

- Where did this data come from?
- How was it collected?
- What can it tell me?

- Do some exploratory data analysis (see next slide)

- Does this data make sense?
- Is it representative?
- What are the key properties?
- Does it resemble anything I've seen before?
- How best to model it?

UNIVERSITY OF CALGARY

Data Analysis Checklist (detailed-level)

- How much data do I have? (N)
- Is it discrete or continuous?
- What is the range for the data? (min, max)
- What is the central tendency? (mean, median, mode)
- How variable is it? (mean, variance, std dev, CV)
- What is the shape of the distribution? (histogram)
- Are there gaps, outliers, or anomalies? (tails)
- Is it time series data? (time series analysis)
- Is there correlation structure and/or periodicity?
- Other interesting phenomena? (scatter plot)

UNIVERSITY OF CALGARY

Multi-Variate Analysis

- For in-depth and really messy data analysis, there are multi-variate techniques that can be immensely helpful
- In many cases, good data visualization tools will tell you a lot (e.g., plotting graphs), but in other cases you might try things like:
  - multi-variate regression: find out which parameters are relevant or not for curve fitting
  - ANOVA: analysis of variance can show the parameters with greatest impact on results

UNIVERSITY OF CALGARY

Presentation of Results

- Graphs and tables are the two most common ways of illustrating and/or summarizing data
  - graphs can show you the trends
  - tables provide the details
- There are good ways and bad ways to do each of these
- Again, it is a bit of an “art”, but there are lots of good tips and guidelines as well

UNIVERSITY OF CALGARY

Table Tips

- Decide if a table is really needed; if so, should it be part of main paper, or just an appendix?
- Choose formatting software with which you are familiar; easy to import data, export tables
- Table caption goes at the top
- Clearly delineate rows and columns (lines)
- Logically organize rows and columns
- Report results to several significant digits (consistently)
- Be consistent in formatting wherever possible

UNIVERSITY OF CALGARY

Graphing Tips (1 of 2)

- Choose a good software package, preferably one with which you are familiar, and one for which it is easy to import data, export graphs
- Title at top; caption below (informative)
- Labels on each axis, including units
- Logical step sizes along axes (1's, 10's, 100's…)
- Make sure choice of scale is clear for each axis (linear, log-linear, log-log)
- Graph should start from origin (zero) unless there is a compelling reason not to do so

UNIVERSITY OF CALGARY

Graphing Tips (2 of 2)

- Make judicious choice of type of plot
  - scatter plot, line graph, bar chart, histogram
- Make judicious choice of line types
  - solid, dashed, dotted, lines and points, colours
- If multiple lines on a plot, then use a key, which should be well-placed and informative
- If graph is “well-behaved”, then organize the key to match the order of lines on the graph (try it!)
- Be consistent from one graph to the next wherever possible (size, scale, key, colours)

UNIVERSITY OF CALGARY

Summary

- There are many “tricks of the trade” used in data analysis and presentation
- A few have been mentioned here
- Effective data analysis and presentation is important in an effective performance evaluation study
- Not always easy to do, but it is worth it!