36-726: Statistical Practice

Second Progress Reports

Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Outline

- Second Progress Reports (Mar 22 – 31)
  - Data
  - Methods
  - Results, if any
  - Next Steps and Roadblocks

- Third Progress Reports (Apr 12 – 21)
  - Update Data and Methods if needed
  - More complete Results, some Discussion
  - Draft IDMRAD paper in addition to talks

Second Progress Reports

Three Purposes:

- Inform me and your MSP peers about progress on your project
  - Data
  - Methods
  - Next Steps and Road Blocks
  - Turn in your slides for talk, just like First Progress Report
- Outlining and initial material for an IDMRAD paper for your project (you will not turn this outline in)
- Get Feedback

IDMRAD…

- Abstract
  - Summarize I, D, M, R and D of paper
- (I)introduction
  - Why would anyone want to read this paper?
  - What questions will be addressed?
- (D)ata
  - What dataset was used for this study?
  - Typically: Variable definitions, sample size, quick summaries and initial descriptive EDA
- (M)ethods
  - What did you do, to address these questions?
- (R)esults
  - What did you find?

(a)nd

- (D)iscussion
  - What does it all mean?
  - Typically: answer questions, discuss generalizations & limitations
- Technical Appendix
  - Technical details of carrying out the (M)ethods

Second Progress Reports

Section	Approximate Length
Title Page	1 Slide
Introduction	1-3 Slides
Data	3-9 Slides
Methods (preliminary)	1-4 Slides
Results (preliminary)	1-4 Slides
Next Steps & Road Blocks (As of now…)	1-3 Slides

Title Page (1 Slide)

- Project title
- Full name of client point of contact
- Full names of the team members
- Full name of faculty advisor

Introduction (≈ 1-3 Slides)

- Client and Project Overview
  - Briefly remind us of what the project is about
- Conclusions from literature review or other study (if any) about the client and project
- Question(s) to be addressed in the project
  - State as precisely as possible at this time

Data (≈ 3-9 Slides)

- Data Sets to be used
  - What are they?
  - Any pre-processing done?
  - Still waiting for (more) data?
- Data Description
  - What is the “shape” of the data?
    - How many observations?
    - How many variables?
  - Variable names and definitions
- How do the data relate to the project question(s)?
  - Missing or unusual data?
  - Numerical Summaries
  - EDA

Methods (≈ 1-4 Slides)

- What are your current ideas about methods, models, analyses, etc.?
  - Mostly statistical/data-science/quantitative
  - In some cases, methods may also involve summarizing interviews or other types of information
- How does each method help to answer each research question from the Introduction?
- Should be reasonably well thought-out
  - But could still change as the project evolves

Results (≈ 1-4 Slides)

- Examples of analyses you've tried so far
  - Perhaps on small extracts of the data and/or simpler data, especially if
    - Your data is large
    - Your methods take a long time to run
- This can be extensive, or quite preliminary or minimal (perhaps even non-existent!)
  - Depends on how far your team has been able to get

Next Steps, Road Blocks (≈ 1-3 Slides)

- *What are your plans for the rest of the project?*
- For each question in your Introduction
  - What intermediate goals are needed to address them?
  - How will you accomplish these goals
- What products are you expected to produce for the client
  - Are they different from “answering the questions”?
- What roadblocks will prevent progress
  - On the project as a whole
  - On particular goals, milestones or products

Third Progress Reports

- Talks scheduled for Apr 12-21

- Thinking about IDMRAD components
  - Updates of D and M as needed
  - More complete R, and some D

- I will also ask for a draft IDMRAD paper from each team.
  - If a different written paper format would work better for your team, let's set up a meeting to discuss.