PUBHLTH 460

Fall 2019

## Project Assignment

For the project you will create, in small groups (3-4 students), a thorough analysis of a dataset chosen by the team (see below). The project will have components completed by the group and by each individual student.

Each team is required to submit their group project to the Undergraduate Statistics Project Competition (USPROC) by Wednesday, December 11, 2019. See https://www.causeweb.org/usproc/usclap

## Guidelines for the group presentation

- A ten-slide presentation summarizing the key findings in the group write-up (see below).
- 7-8 minutes maximum per group.

## Guidelines for the group write-up

As a team, you are encouraged to pay close attention to the submission guidelines for the USPROC (see https://www.causeweb.org/usproc/usclap).

From the USPROC:

A 3-page (or shorter) paper reporting the results of your project that includes the following:

- The research question(s)
- Background/significance of the research
- The methods used to obtain and analyze the data
- The results of the analysis (tables, charts, graphs, significance, confidence intervals, descriptive text)
- A discussion of the research, the limitations of the current research, reasonableness of any assumptions made, possibilities of future work/studies that should be conducted, etc.
- In addition to the 3 page paper, you may have (in the same file) a blinded title page which includes: The title of the project and a one-paragraph abstract of the project (recommended length of no more than 150 words).

- References should be listed at the end of the paper and do not count against the 3 page limit.

The entire written summary must be **no more than 3 pages** (single spaced, 11pt or 12pt font with standard 1 inch margins).

Your group should assemble an outline of the key elements of the story that you want to tell and which team member will be responsible for each element. The general idea is that each element should focus on one key observation or insight about the dataset. Tell a short, compelling story with a small number of elements. The elements should complement each other and together tell a coherent story about your dataset. Elements could be data visualizations, regression analyses, integrations from other datasets, or some other quantitative piece of the story.

### Guidelines for the individual write-up

Each member of the group will serve as the lead on one or more elements that tell a story about the assigned dataset. In addition to incorporating these elements into the final group deliverables, each individual student will produce a **2-page write-up** (including tables and figures) about their individual analyses. These individual analyses should provide more technical detail about the analyses performed as well as giving a brief introduction and conclusion to the analysis. Each write-up should stand on its own, providing tables and figures as necessary.

The individual data analysis write-ups will be due Thursday December 19th at 5pm, to be handed in via Moodle. Individual analyses should be handed in as PDFs knitted using RMarkdown. No code should be displayed in these reports, however, figures resulting from your analyses should be dynamically created.

### Project grading

Your project grade makes up 25% of your final grade for the class. The grading rubric for the group portion of the project will be evaluated based on the juding rubric provided by the USPROC with the following adjustment to the point scales:

- The final product produced by the group: **60 points**
    - Description of the data source (12 points)
    - Accuracy of data analysis (12 points)
    - Accuracy of conclusions and discussion (12 points)
    - Overall clarity and presentation (12 points)
    - Originality and significance of the study (12 points)

- Individually prepared data analysis: **40 points**
    - 30 points: overall quality of analysis (correct implementation and interpretation of method(s) used, appropriate use of equations to show what methods/models have been used, appropriate use of graphics/tables to support central results, succinct summary of key results)
    - 10 points: clarity and presentation (clear statement/summary of goals and central results, clear and accurate description of methods/models used, use of figures rather than text to illustrate central ideas, figures dynamically generated within the RMarkdown file, page limit adhered to)

### Milestones and deadlines

- **Tuesday Oct 22**: Project announced
- **Tuesday Oct 29**: Teams + dataset (submit on Moodle, 1 submission per team).
- **Tuesday Nov 5:** Background paper discussions/Research Question (1 page write up, submit on Moodle, 1 submission per team)
- **Tuesday Nov 19**; **Thursday Nov 21; Tuesday Dec 3**: In-class project work (Attendance mandatory).
- **Thursday Dec 5**: Group presentations
- **Tuesday Dec 10**: Group presentations
- **Wednesday Dec 11**: Group write up submitted by 11:59pm on Moodle and to USPROC (1 submission per team).
- **Thursday Dec 19**: Individual project write-up submitted by 5pm on Moodle.

### Datasets

1. **ASA Public Health Data Challenge (Fall 2019) [Homelessness in 3 cities]** https://thisisstatistics.org/falldatachallenge/

2. **Breast cancer proteogenomics.** Proteogenomics connects somatic mutations to signaling in breast cancer, Merlins, P. et al., *Nature* volume 534, pages 55–62 (02 June 2016). See Moodle for datasets and README file with information on data formats.

**Example successful projects:**

ASA Public Health Data Challenge (Fall 2018) [Opioid Crisis]

https://thisisstatistics.org/public-health-data-challenge-winners/

**High dimensional data – visualization and analysis:** We will hold an *optional* session during one of the TA hours on tips for analyzing/visualizing high dimensional data. This may be useful if you choose the breast cancer proteogenomics data for your project.