Final Project

EPIB607 - Inferential Statistics[a]

[a]Fall 2021, McGill University

This version was compiled on September 1, 2021

Final project instructions. Due date December 22, 2021.

Final project

1. Final Project

The objective of the group project is to construct an exercise and solutions suitable for testing or demonstrating understanding of basic principles of biostatistics as discussed in this course.

Exercises must be based on (i) one to two articles in a scientific journal or perhaps in the lay press or (ii) a *publicly available* dataset. The data must not be taken from an RA project, but must be freely available on the web or another public source. The article or data should concern some health problem amenable to statistical investigation. The narrative of the exercise should be clear and concise. The exercise should comprise 5-7 questions requiring altogether about three hours for completion. The questions may cover any part of this course. You must also produce a separate set of model answers; these should be equally short and to the point.

The group project will be evaluated using the following criteria (for a total of 10 points):

1. The choice of subject and ingenuity (2.0 points)

2. Testing of important biostatistical principles (2.5 points)

3. Exercises that are clear, concise, and creative. It's better to have one question that tests several concepts together, vs. several questions that have no link with each other (2.5 points)

4. Quality of solutions (2.0 points)

5. Is the report reproducible (1.0 points)

Projects should be done in groups of 2 to 4 people. Examples final projects prepared by students in previous years have been posted on MyCourses. All projects must be uploaded to myCourses. One submission per group.

The upload should consist of the following:

1. One `.Rmd` file containing the questions and solutions. This must be fully reproducible using the techniques discussed in this class, i.e., I should be able to download your submission, open the `.Rmd` file, and compile it without error. Be aware of file paths and hard coded solutions.
2. One compiled `.pdf` or `.html` file of the Rmarkdown document
3. Any article(s) on which the questions are based
4. Any data-sets used in the questions, in text or CSV format. If the dataset is publicly available then a link to the dataset or the R package is sufficient.