BIOS 6611 Final Project

Due Friday, December 7, 2018 by midnight to Canvas Assignment Basket

The Overview

This project is intended to give you hands-on experience with a topic that you will choose from the list of two below. These involve analysis of data from actual projects carried out at UCD.

For your chosen topic, you'll prepare a 3-page summary (single spaced; strict maximum) in the form of a brief scientific report. For the data analysis projects, you'll find a general outline for this format on p. 5 of this document. An example is posted on the Final Project page of the Canvas course.

UCD Datasets:

- Parkinson's Exercise Trial
- Mole Development in Colorado Kids

Details for each of the topics are listed below.

Analysis of a UCD Dataset
You'll formulate one research question, develop an analytic plan to address the question, and present the results and interpret them. You can choose one of the two datasets provided for your project. For each dataset, the file names, variable names, and descriptions are on pages 3 and 4, respectively, of this document. Each of the data analysis projects has its own subsection on the Canvas page for the Final Project. On Canvas, you will find the data and the resources available for each project. You can use either R or SAS, or some of each, to conduct your analysis.

Develop one question and analyze data to answer it
The resource material for each project will give you an idea of the potential interrelationships among the variables included. Use those materials to develop one question that could be answered with the data you have.

One study is a completed clinical trial with longitudinal data in a "wide" format, i.e. the repeated measures are all one record for a patient. The other is an ongoing observational study, also longitudinal, that grew out of a clinical trial. These data are also in wide format.

In your analysis you should involve data from *at least two* time points for your outcome of interest, e.g. the difference between the outcome measured at the last time point and the baseline time point, the difference between baseline and an average of the post-baseline measures. (**Note:** You will learn to work with the longitudinal data in detail next semester. For now, you will use methods appropriate for uncorrelated data, e.g. by taking differences and treating the difference as the outcome of interest.) For the observational study it will be important to identify *potential* confounders of the relationships between the variables you choose to focus on.

Your project must include the following aspects:

Apply Resampling to understand Sampling Variability of Parameters of Interest
You should investigate and summarize the bootstrap sampling distribution of at least one parameter estimate of interest related to your study question.

Apply Reproducible Research (RR) principles for reporting
In addition to the 3-page summary, you should include a brief Appendix with the R or SAS code that you used to perform the analysis. This will ensure that anyone looking at your analysis and summary will be able to know exactly what you did.

***Be sure to*:** Comment your program, attach useful names to any variables you create and analyze, and attach labels to variable values using formats.

See the last page of this document (p. 6) on Reproducible Research for resources you can use and apply.

<u>Data Analysis Project 1</u> – A Randomized Clinical Trial of Exercise in Early to Mid-Stage Parkinson's Disease

Lead Investigator: Margaret Schenkman PhD

Data are in a .csv file: PD Exercise RCT Selected Secondary Outcomes - Wide.csv

Summary:

121 participants randomized into one of 3 exercise groups
33 variables
Repeated outcomes: Baseline (0 months), 4 Months, 10 Months, 16 Months; some missing data; suffix for outcome measures below denotes when the measure was taken

Data Dictionary:

Variable/Field Name	Label/Attributes
Participant	id number
Group	4 = Home Exercise 5 = Flexibility, Balance, and Functional Training 6 = Aerobic Conditioning
Gender	1 = Male 2 = Female
Age	years
YearsDx	Years with PD
HYStage0, HYStage4, HYStage10, HYStage16	Hoehn and Yahr stage of PD – scale from 1 (lowest) to 4 (worst) in increments of 0.5 at baseline, 4, 10, and 16 months
FiveM_Wk0, FiveM_Wk4, FiveM_Wk10, FiveM_Wk16	Five meter walk in number of steps at baseline, 4, 10, and 16 months
FiveM_Tm0, FiveM_Tm4, FiveM_Tm10, FiveM_Tm16	Five meter walk in seconds at baseline, 4, 10, and 16 months
TUG0, TUG4, TUG10, TUG16	Timed Up and Go in seconds at baseline, 4, 10, and 16 months
UPDRS0, UPDRS4, UPDRS10, UPDRS16	Total score on UPDRS (see scale information in resource document) at baseline, 4, 10, and 16 months
SixMn_Wk0, SixMn_Wk4, SixMn_Wk10, SixMn_Wk16	Six Minute Walk in meters at baseline, 4, 10, and 16 months
LEDD0, LEDD4, LEDD10, LEDD16	Levodopa equivalents (mg/day) (commonly prescribed medication for PD) at baseline, 4, 10, and 16 months

Additional Resources:

UPDRS,H&Y, S & E_MedEl_tool.doc
UPDRS Background Paper.pdf
Schenkman at al 2012 PTJ Exercise Early MidStage PD 16 Month RCT.pdf

Data Analysis Project 2 - Mole Count Study in Colorado Children

Lead Investigator: Lori Crane PhD

Data are in an Excel spreadsheet: Mole Count Data 2004-2008.xls

Summary:

472 children age 6 followed from baseline to age 10
15 variables
Longitudinal study of mole development over five years

Data Dictionary:

Variable/Field Name	Label/Attributes
Respondent Code Number	id number
oca2 status	0 = gg 1 = ga 2 = aa 9 = missing
gender	1 = Female 2 = Male
Hispanic	0 = No 1 = Yes
molecount2004	Number of moles in 2004
molecount2005	Number of moles in 2005
molecount2006	Number of moles in 2006
molecount2007	Number of moles in 2007
molecount2008	Number of moles in 2008
eyecolor	1 = blue, green or combo 2 = light/dark brown 3 = hazel
baseskincolor	Skin color based on a continuous score, higher is darker
haircolor	1 = blonde 2 = red 3 = brown 4 = black
number vacs birth thru 2005	Total number of waterside vacations from birth through 2005
number vacs birth thru 2006	Total number of waterside vacations from birth through 2006
number vacs birth thru 2007	Total number of waterside vacations from birth through 2007

Additional Resources:

Mole Study R01 July 2009 Draft.doc
Crane et al. 2009 – Nevus development in children (pdf)
Pettijohn et al. 2009 – Waterside vacations and nevus development (pdf)
Crane et al 2012 AJPM Mailed Intervention Sun Prot Children RCT (pdf)

Outline of Brief Scientific Report

Introduction with brief background – 0.5 page

Question or hypothesis – 0.125 page

Materials and Methods – 0.25 page

Analysis Plan – 0.25 page

Results – 1.5 page
- "Table 1"
- Descriptive or other plot(s)
- Analytic results table

Discussion and Conclusion – 0.375 page
- Include limitations

References (not included in 3-page limit)

Appendix for Reproducible Research Goal (not included in 3-page limit):
- SAS or R code

Refer to example: AJPH Example Brief Report.pdf, located on the Final Project Canvas page

General Writing Resource: Tips on Writing Results for a Scientific Paper Amstat News - Cummins - Sept 2009.pdf, located on the Final Project Canvas page

Reproducible Research

You will need to follow Reproducible Research principles in carrying out your project. Adherence to the principles of reproducible research is essential to the reporting of scientific results.

Be sure to familiarize yourself with some of the resources on this topic listed below.

An outstanding example of reproducible research by Keith Baggerly and Kevin Coombes of the MD Anderson Cancer Center Bioinformatics and Computational Biology department is posted in the Paper Repository ("Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology" in the *Annals of Applied Statistics* by Baggerly and Coombes, and *Cancer Letter* 1, 2, and 3 by Paul Goldberg).

RR Resources: (Canvas Paper Repository)

Gentleman and Temple Lang paper – Statistical Analysis and Reproducible Research; from the Bioconductor Project

Roger Peng 2009 Editorial in *Biostatistics*

Roger Peng 2011 article in *Science*

Frank Harrell presentation on the use of R in clinical trials research

IOM Report on Evolution of Translational Omics

Lehrer 2010 New Yorker The Truth Wears Off

Dynamic/computable documents using R
http://yihui.name/knitr/
http://rmarkdown.rstudio.com/

Dynamic/computable documents using SAS
http://sites.northwestern.edu/stattag/

Recommended, but not required:

Use one of these (or another) computable document applications.

Git and GitHub – version control, collaboration tools
https://git-scm.com/
https://github.com/