Final Projects (Fall 2019)- Statistics 601. *Due on December 18, 2019 at 5 pm central time.*

*Work on this exam by yourself and be sure to reference any material you use on your exam. Only discuss the exam with me, Ms. Fuglsby, or Ms. Hajebi and not the other students in the class.*

***Do two of the following three problems. Turn in one pdf and RMD file for each problem clearly labeling which problems you have chosen to do.***

**1:** The data shown in schizophrenia.csv were collected in a follow-up study of women patients with schizophrenia and summarized in ﻿Davis (2002), *Statistical Methods for the Analysis of Repeated Measurements*, Springer, New York. The binary response recorded at 0, 2, 6, 8 and 10 months after hospitalization was thought disorder (absent or present). The single covariate is the factor indicating whether a patient had suffered early or late onset of her condition (age of onset less than 20 years or age of onset 20 years or above). ***The question of interest is whether the course of the illness differs between patients with early and late onset?*** Investigate the question of interest.

* + 1. *Provide a two to three-page write-up (including graphs) explaining your analysis of the experiment and the conclusions you can draw from it.*
    2. *As a secondary component provide annotated code that replicates your analysis.*

*Make sure to discuss any concerns about the modeling assumptions used in your analysis.*

*The .csv file has the following variables.*

*subject*

* + the patient ID, a factor with levels 1 to 44.

*onset*

* + the time of onset of the disease, a factor with levels < 20 yrs and > 20 yrs.

*disorder*

* + whether thought disorder was absent or present, the response variable.

*month*

* + month after hospitalization.

*Please note that you may have already explored this dataset in the class. Even so, please do a complete and extended analysis answering the questions, with the focus of writing and explaining the what you have found in your analysis.*

**2:** (Vole Data)- Consider the *“microtus"* dataset in the “Flury" library in R.

*Background from Airoldi\_Flury\_Salvioni\_JTheorBiol\_1995: Discrimination Between Two Species of Microtus using both Classified and Unclassified Observations.*

*“1. Introduction*

*Microtus subterraneus and M. multiplex are now considered to be two distinct species (Niethammer, 1982; Krapp, 1982), contrary to the older view of Ellerman & Morrison-Scott (1951). The two species differ in the number of chromosomes: 2n=52 or 54 for M. subterraneus, and 2n=46 or 48 for M. multiplex. Hybrids from the laboratory have reduced fertility (Meylan, 1972), and hybrids from the field, whose karyotypes would be clearly recognizable, have never been found (Krapp, 1982).*

*The geographic ranges of distribution of M. subterraneus and M. multiplex overlap to some extent in the Alps of southern Switzerland and northern Italy (Niethammer, 1982; Krapp, 1982). M. subterraneus is smaller than M. multiplex in most measurements, and occurs at elevations from 1000 m to over 2000 m, except in the western part of its range (for example, Belgium and Brittany), where it is found in lower elevations. M. multiplex is found at similar elevations, but also at altitudes from 200–300 m south of the Alps (Ticino, Toscana).*

*The two chromosomal types of M. subterraneus can be crossed in the laboratory (Meylan, 1970, 1972), but no hybrids have so far been found in the field. In M. multiplex, the two chromosomal types show a distinct distribution range, but they are morphologically indistinguishable, and a hybrid has been found in the field (Storch & Winking, 1977).*

*No reliable criteria based on cranial morphology have been found to distinguish the two species. Saint Girons (1971) pointed out a difference in the sutures of the posterior parts of the premaxillary and nasal bones compared to the frontal one, but this criterion does not work well in many cases. For both paleontological and biogeographical research it would be useful to have a good rule for discriminating between the two species, because much of the data available are in form of skull remains, either fossilized or from owl pellets.*

*The present study was initiated by a data collection consisting of eight morphometric variables measured by one of the authors (Salvioni) using a Nikon measure-scope (accuracy 1/1000 mm) and dial calipers (accuracy 1/100 mm). The sample consists of 288 specimens collected mostly in Central Europe (Alps and Jura mountains) and in Toscana. One peculiar aspect of this data set is that the chromosomes of 89 specimens were analyzed to identify the species. Only the morphometric characteristics are available for the remaining 199 specimens…”*

Develop a model from the 89 specimens that you can use to predict the group membership of the remaining 199 specimens’.

* 1. *Explain your GLM and assess the quality of the fit with the classified observations.*

*You may want to use Cross Validation to predict the accuracy of your model.*

* 1. *Provide a two to three-page write-up (including graphs) explaining your analysis of the dataset and your recommendations on the usefulness of your predictions.*
  2. *Provide predictions for the unclassified observations.*
  3. *As a secondary component provide annotated code that replicates your analysis.*

**3.** In Claflinet al. (2011) *Effects of high- and low-velocity resistance training on the contractile properties of skeletal muscle fibers from young and older humans.* Journal of Applied Physiology, 111, 1021-1030. The authors are interested in the effect of a 12-week program of Progressive Resistance Training (PRT) on:

* + - * 1. the power output of the overall musculature associated with movements of the ankles, knees, and hips;
        2. the cross-sectional area and the force and power of permeabilized single fibers obtained from the vastus lateralis muscle; and
        3. the ability of young and elderly men and women to safely arrest standardized falls.

The training consisted of repeated leg extensions by shortening contractions of the leg extensor muscles against a resistance that was increased as the subject trained using a specially designed apparatus.

There are two files in that contain the data associated with this study.

The first file is “prog.res.trn.csv” it has four columns containing the experimental results.

id

* + experimental subject ID (Index of the 63 subject in the experiment)

iso.fo

* + numeric values of isometric force ranging from 0.16 to 2.565

spec.fo

* + numeric values of specific force ranging from 80.5 to 290

occ.f

* + 2 levels Pre, Pos, i.e. pre- and post- intervention measurement

fiber.f

* + 2 levels Type 1, Type 2, i.e. Type 1 and Type 2 muscle fiber being measured

The second file is “prog.res.trn.subs.csv” it has four columns containing information on the subjects that participated in the study.

id

* + experimental subject ID (Index of the 63 subjects in the experiment). This will allow you to link information between the two files.

prt.f

* + Whether or not the subject was randomized between a “high” and “low” intensity of a 12-week PRT.

age.f

* + Whether the subject was Young (21-30 years of age), Old (65 to 80 years of age)

sex.f

* + Whether the subject was Female, Male

bmi

* + The Body Mass Index of the subject, ranging from 18.4 to 32.3

***The question of interest is whether or not*** a 12-week program of Progressive Resistance Training (PRT) on effects:

* + - * 1. the power output of the overall musculature associated with movements of the ankles, knees, and hips;
        2. the cross-sectional area and the force and power of permeabilized single fibers obtained from the vastus lateralis muscle; and
        3. the ability of young and elderly men and women to safely arrest standardized falls.

***Investigate the question of interest by analyzing the isometric force and specific force responses individually.***

* + 1. *Provide a two to three-page write-up (including graphs) explaining your analysis of the experiment and the conclusions you can draw from it.*
    2. *As a secondary component provide annotated code that replicates your analysis.*

*Make sure to discuss any concerns about the modeling assumptions used in your analysis.*