Final Projects (Fall 2020)- Statistics 601. *Due on December 5, 2020 at 12 pm central time. This will give me enough to grade the exams and ask questions if something is unclear concerning your submission.*

*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 Mr. Claussen and not the other students in the class. All coding should be done in R.*

*Each question is worth 50% of the final grade.*

***Turn in one pdf and RMD file for each problem.***

**1:** (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 specimen’s.

* 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 (****excluding graphs****) explaining your analysis of the dataset and your recommendations on the usefulness of your predictions. This can be done in a Word document, in RMarkdown, or another text editor, but submit a PDF.*
  2. *Provide predictions for the unclassified observations. Upload as a separate excel file.*
  3. *As a secondary component provide annotated code that replicates your analysis. Upload as an .R or .RMD file.*

**2.** In Martyna et al. (2013) (*The evidential value of microspectrophotometry measurements made for pen inks.* Analytical Methods) the authors are focused on developing a type of similarity score, between two ink samples, known as a forensic likelihood ratio. We are interested in three of the univariate LR’s (corresponding to the “X”, “Y”, and “Z” measures of color) that were constructed and their relationship to an omnibus LR that I (Dr. Saunders) constructed. Please describe and characterize the relationship between the 4 forensic likelihood ratios, with the ultimate aim of predicting the omnibus Likelihood ratio in terms of the three univariate likelihood ratios. Note that all of the likelihood ratios are given on a “log” scale.

There is one csv file that contains the data for this study.

The ~~first~~ file is “dat.LLR.int.csv” it has six columns containing:

Comp

* + Comparison of interest (from 1 to 820).

Omni.LLR.int

* + numeric values of for the Log of the Omnibus likelihood ratio, this is the response variable we are interested in.

Type

* + Type of comparison, either “wi” for within-source comparison or “bw” for between-source comparison

LLR.x

* + The Log Likelihood ratio for the X color variable

LLR.y

* + The Log Likelihood ratio for the Y color variable

LLR.z

* + The Log Likelihood ratio for the Z color variable

***The question of interest is whether or not*** there is a relationship between the marginal LR’s (i.e. LLR.x, LLR.y, and LLR.z) and the Omnibus LR (i.e. Omni.LLR.int). If there is a relationship, please characterize said relationship.

* + 1. *Provide a two to three-page write-up (excluding graphs) explaining your analysis of the experiment and the conclusions you can draw from it. This can be done in a Word document, in RMarkdown, or another text editor, but submit a PDF.*
    2. *As a secondary component provide annotated code that replicates your analysis. Upload as an .R or .RMD file.*

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