STAT 6021 Project2 (Group 2) Proposal: Characteristics of nursing homes in New Mexico.

Gregory Madden, Christina Kuang, Chi Do, Trey Hamilton

Part 2: Proposal

ideas in bold.

Each group will submit a project proposal for approval by the instructor, to ensure that an appropriate data set is being used and an appropriate amount of analysis will be done to produce a successful 4 week project. Each proposal should provide:

- The associated data set for the project should be provided, in one of the following ways:

1) providing the names of the R package and the R dataframe;

R Package: "data("Nursing")"

R dataframe: "Nursing"

```
# from data('Nursing') package
`?`(Nursing)
Data <- data("Nursing")
```

2) a link to the data set;

Data can be found at: http://lib.stat.cmu.edu/DASL/Datafiles/Nursingdat.html

3) a file containing the data, as an Excel spreadsheet, a .csv file, or a .txt file. Please note that a list of some publicly available sources of data sets is provided on Collab.

```
write.csv(Data, "Nursing.csv")
```

• Project objectives/goals. What questions is the group trying to answer, as well as potential practical implications (the more interesting and/or practical, the better) of the results. Your project should involve both linear regression and logistic regression, so clearly state the response variables involved.

Q1. What characteristics of nursing homes in New Mexico dictate annual nurse salaries at those institutions?

Practical implications of a linear model for predicting nurse salaries could be used by policymakers to rationally distribute subsidy funds to institutions that are expected to have the lowest salaries.

Q2. What factors are associated with rural vs. non-rural nursing homes.

Rural patients suffer from a lack of locally available nursing home beds. Understanding the relationships of these characteristics and how they define rural vs. non-rural nursing homes is helpful to know how to make rural nursing homes financially viable.

• Some data visualizations and commentary related to the project objectives/goals. At least one visualization related to linear regression, and at least one visualization related to logistic regression.

Objective 1: Fit a multiple linear regression model with annual nurse salary for individual nursing homes using the available financial characteristics for each institution. The goal is to develop a model using these available data to reliably predict annual nursing salaries among the larger group of nursing homes across the state.

Objective 2: Fit a logistic regression model with rural vs. non-rural status as the response using the available financial characteristics.

The proposal should be no longer than 6 pages. The proposal will be graded on a pass / fail basis. The proposal will be evaluated on the following:

• The appropriateness of the data set. – The instructor should be able to access the data set easily. – The data set should be formatted in a manner where each row represents an observation, and each column represents a variable. – Please note that regression methods assume the observations are independent. One way of assessing whether your observations are independent is to ask if the order of the rows in your dataframe matters. If you can scramble the order of the rows without affecting any structure in your data, then your observations are likely to be independent. If scrambling the rows upends the structure of the dataframe, then your observations are not independent, and regression methods are not meant to handle dependent data. – For the categorical response variable, be sure it is binary. We have not covered enough material to tackle categorical response variables with more than 2 classes.

• The objectives are appropriate for a project that spans 4 weeks. The instructor will assess if you have at least one question involving linear regression and at least one question involving logistic regression.

Feedback will be provided for group proposals that are rejected. Groups will have the option of submitting one revised proposal (but given the relatively short time you have to complete the project, this should be avoided).

Submission Please submit your group's proposal (.pdf file) via Assignments (1 upload per group).