

## **Homework 7 of STAT 540**

### **Section 001, Fall 2024**

**Due: Wednesday Nov 6, 2024 (before class)**

**Total Points: 50**

Please hand in a hard copy of your homework in class (SAS code and output) and email your SAS code to Kaniz Fatema ([KFATEMA@email.sc.edu](mailto:KFATEMA@email.sc.edu)). Work on the homework independently.

**Problem 1.** Write a SAS program to read in the data file `baseballroster.txt`. The variables for each observation, in order, are: Name, height(feet), height(inches), weight(pounds), position, number of years on team, birthdate, and date the joined the starting lineup (if ever).

- a. Within the data step, do the following:
  - (a1) Create another variable that gives the players' weights in kilograms.
  - (a2) Create another variable that gives the players' heights in meters.
  - (a3) Suppose that the equipment for catchers weighs 10 pounds, the equipment for pitchers and infielders weighs 2 pounds, and the equipment for outfielders weighs 2.5 pounds. Create (as efficiently as possible) another variable that gives the players' weights (in pounds), including their equipment's weight.
  - (a4) Create another variable that gives the players' current ages in years (rounded down, like ages are conventionally stated).
- b. Then complete the following procedures:
  - (b1) Print the entire data set with the players sorted by position (alphabetically), and by years on the team within each position (list most experienced players first). Have SAS print out the dates in

word format including day of the week (e.g., Friday, November 1, 1974).

- (b2) Use PROC MEANS to have SAS show ONLY the mean and standard deviation of the player's weights (in POUNDS, without equipment). Note that in this part have SAS show only the mean and standard deviation and not any other summary statistics, and these statistics are only for the weight and not for other variables.
- (b3) Then have SAS show the mean and standard deviation of weights (again in POUNDS, without equipment), SEPARATELY for each type of position. Create a SAS data set named summary to save the output that contains the mean and standard deviation of weights for each type of position. Print out the data in the data set summary.

Problem 2. Check out the file moviereview.txt, which contains the reviews for 160 movies from 1996 by Gene Siskel and Roger Ebert. The variables in order in the file are movieNumber, reviewSiskel, and reviewEbert.

- (a) Use a Data step to read the data into SAS and generate a SAS data set called moviereviews. Print out the last ten observations of this data set.
- (b) Based on the existing data set moviereviews, create a SAS data set subset10 that contains every 10th observations of moviereviews. Print out all of the observations in subset10 and use movieNumber to replace the obs index.
- (c) Update moviereviews by adding a new variable agree, which takes 1 if the reviews given by Siskel and Ebert are the same and 0 otherwise. Print out observations 20 - 30 of this new data set.
- (d) Use proc means to calculate the proportion of agreement between the reviews by Siskel and Ebert (i.e., mean of variable "agree"). Label this proportion as prop and output it in a sas file called summarydata1. Report this proportion of agreement.
- (e) Calculate the proportions of agreement by reviewSiskel using proc means. That is, for each value of reviewSiskel, find the proportion of agreement by the two reviewers. Label this proportion as prop2 and save

- the output in a sas file called summarydata2. Print out summarydata2.
- (f) Merge the two data sets summarydata2 and moviereviews into a new data set called mergeddata. Print out mergeddata with variables: movieNumber reviewSiskel, reviewEbert, and prop2 for those observations whose movieNumber is a multiplier of 20.
  - (g) Create separate tables to summarize reviewSiskel and reviewEbert using proc Freq. Find the largest category with the most of observations for each of the reviewers.
  - (h) Create a two-way table between reviewSiskel and reviewEbert using proc Freq. Find the category of reviewSiskel and reviewEbert that has the most observations.