

- **Note:**

The take home should be turned in on Monday at 1:00pm. You may not talk to other students about the exam problems, and you may not look at other students' exams. Violations of this policy may result in a 0 on the exam.

1. You are working as a statistical assistant for the ABC baseball team at Austin of Texas and plan to do analysis to players' performance. You has gathered data on 100 players, measuring 11 variables on each, which represent various characteristics of the players: Salary (in thousands of dollars), batting average, On-base percentage, Number of runs, Number of hits, Number of doubles, Number of home runs, Number of walks, Number of strike-outs, Number of errors, and Player's name (in quotation marks). The data set is named baseball.csv.

- Players' batting averages are calculated as the ratio of number of hits to the number of hits plus the number of outs.
- On-base percentage is the ratio of number of hits plus the number of walks to the number of hits plus the number of walks plus the number of outs.
- A batting average above .300 is very good; OBP above .400 is excellent.

The questions that the team boss want to hear from you are:

- (3 pts) Are there particular players that are highly unusual in terms of the measured characteristics? If so, identify them.
- (3 pts) Are there notable associations/relationships between some of the variables? If so, describe them.
- (3 pts) Is there a way to graphically represent the raw data for the 100 players and draw conclusions about the data set from such a graph?
- (5 pts) Baseball provides a rare opportunity to judge the value of an employee. In this case, a player -- by standardized measures of performance. The question is, can we find a less number of items to describe at least 90% of variations of the original data set or the original number of characteristics?
- (10 pts) If so, what are those items? Is there convenient interpretation of any of the items?
- (5 pts) Are there any players that are similar or different from each other in any aspects?
- (6 pts) What are those characteristics that is worth to be consider the best predictors of the players' salary of the ABC baseball team?

You will type a report detailing your analysis of the data and your conclusions. You may include graphs that illustrate and/or support your findings. Do NOT include computer code within the main body of your report. You may include such code in an appendix if you wish.

2. The researchers who study of mammals collected a data set on 52 mammals. The variables measured include weight measurements of mammals (body weight and brain weight) and characteristics measurements of mammal (total daily sleep, maximum lifespan, and gestation time). The data file contains the observed values of the 6 variables (plus a labeling column with the names of the species) for 52 mammals. The data set is named mammals.csv. The units of measurements are given below

- species of animal
- body weight in kg
- brain weight in g
- total sleep (hrs/day)
- maximum life span (years)
- gestation time (days)

NOTE: It is recommended by the ecologists that you use a natural log transformation of the body weight and brain weight variables before doing the analysis.

The questions that the ecologists would like answered include:

- (1) (5 pts) Are there notable associations/relationships between some of the variables? (if so, describe them)
- (2) Considering the data set as two groups of data sets: Weight measurement (log body weight and log brain weight) and characteristics measurement (total sleep, life span, and gestation time).
 - (5 pts) Test for independence between two sets of variables.
 - (5 pts) Determine the number of significant canonical variate pairs, if null hypothesis is rejected in part a.
 - (5 pts) Compute the correlation between the canonical variate pairs, if null hypothesis is rejected in part a.
 - (5 pts) Compute the canonical variates from the data.
 - (5 pts) Interpret each member of a canonical variate pair using its correlations with the member variables.

You will type a report detailing your analysis of the data and your conclusions. You may include graphs that illustrate and/or support your findings. Do NOT include computer code within the main body of your report. You may include such code in an appendix if you wish.