

Homework 3: Introduction to Statistical Computing Using SAS

Instructor: H. Peng

STAT 42100: Modern Statistical Modeling Using R and SAS

STAT 52100: Statistical Computing

Instruction: Collaboration on homework assignments is acceptable, but all write-ups must be done independently and clearly indicate the submitter's understanding of the material. Unclear or disorganized homework may have points removed, even if the content is correct. **No hardcopy of the typed report needs to be turned in. The pdf file should be submitted electronically on canvas.** An assignment handed in after the deadline is late, and may or may not be accepted. My solutions to the assignment questions will be available when everyone has handed in their assignment.

Edit the program(s) and output together into a single document, showing the lines of code and relevant output produced by SAS/R. Your answers must be easy for the grader to find. A simple structure is, for each part of each question in order, to put these three things:

- Your code
- Your output
- Your answers and explanation

For SAS, the code and output are naturally separate, so this order is good; for R, the code and output can be intermingled, and that is OK. If your assignment is disorganized or otherwise difficult for the grader to deal with, you can expect to lose marks.

You are reminded that work handed in with your name on it must be entirely your own work. It is as if you have signed your name under it. If it was done wholly or partly by someone else, you have committed an academic offence, and you can expect to be asked to explain yourself. The same applies if you allow someone else to copy your work. The graders will be watching out for assignments that look suspiciously similar to each other (or to my solutions). Besides which, if you do not do your own assignments, you will do badly on the exams.

Problem 1: Brain Activity in Violin and String Players. Studies over the past two decades have shown that activity can effect the reorganization of the human central nervous system. For example, it is known that the part of the brain associated with activity of a finger or limb is taken over for other purposes in individuals whose limb or finger has been lost. In one study, psychologists used magnetic source imaging (MSI) to measure neuronal activity in the brains of nine string players (six violinists, two cellists, and one guitarist) and six controls who had never played a musical instrument, when the thumb and fifth finger of the left hand were exposed to mild stimulation. The researchers felt that stringed instrument players, who use the fingers of their left hand extensively, might show different behavior in the brain-as a result of this extensive physical activity-than individuals who did not play stringed instruments. The attached data contain a neuron activity index from the MSI and the years that the individual had been playing a stringed instrument (zero for the controls). (Data based on a graph in Elbert et al., "Increased Cortical Representation of the Fingers of the Left Hand in String Players," Science 270 (13 October, 1995) 305-7.)

1. Is the neuron activity different in the stringed musicians and the controls?
2. Is the amount of activity associated with the number of years the individual has been playing the instrument?

Create an indicator variable for player vs. non-player. Include appropriate EDA for activity, years, and the new player variable. Perform both ANOVA (for player vs. non-player) and regression to answer the above two questions.

Problem 2: Kentucky Derby. The Kentucky Derby is a 1.25 mile horse race held annually at the Churchill Downs race track in Louisville, Kentucky. The attached data set contain the year of the race, the winning horse, the condition of the track, and the average speed (in feet per second) of the winner, for years 1896-2000. The track conditions have been grouped into three categories: fast, good (which includes the official designations "good" and "dusty"), and slow (which includes the designations "slow", "heavy", "muddy", and "sloppy"). Develop a model for the mean winning speed as a function of year and the track condition factor. Please show diagnostics of your model. The data are from www.kentuckyderby.com. Include appropriate EDA for the outcome and the two explanatory variables.