

Homework 2: 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 In studies of the placebo effect, it has been established that nausea can arise after a medicine is taken by mouth even though there is no physical cause for the distress. As a result, a placebo (a tablet consisting of inert material and free of the drug) is given to half of the patients in a drug trial. The patients have no idea if they are getting a placebo or not, and their response (nausea or no nausea) is recorded after the dose. The results are given in the following table:

	Nauseated	Not Nauseated
Drug Given	15	35
Placebo Given	4	46

Is there evidence of an association between nausea and the taking of the drug? Explain which statistic(s) you used, and give the associated p-value(s).

Problem 2 The data set *home.txt* (download from canvas please) contains data on Albuquerque housing prices based on a random sample of over 100 homes sold Feb 15 to Apr 30, 1993. The data were obtained from the Albuquerque Board of Realtors.

- Create scatterplots of TAXES versus FEATS separately for the four different classes of houses defined by combinations (0,0), (1,0), (0,1), (1,1) of (COR,NE) (Use WHERE statement). Does these plots tell you anything about whether the relationship between TAXES and FEATS is different in the four different classes of houses defined by (COR,NE) ?
- Create side-by-side boxplots of TAXES for the 4 (COR,NE) groups, and of FEATS for the 4 (COR,NE) groups, to see whether these groups differ from each other in their Taxes and their numbers of Features. What do you conclude from these pictures ?
- Break the set of houses into three groups according to whether they have LOW, MEDIUM, or HIGH taxes. Then use a Chi-square test of Row-Column independence, calculated through PROC FREQ, to determine whether there is any relationship between the tax group and the COR status of the houses. Explain your conclusions.