For both problems, include your R code and its output at the end of the problem (in a fixed-width font such as Courier).

1. Use the badhealth data set from the COUNT package in R for this problem. The only information available about this data set is that it is a German survey from 1998 (or at least regarding the year 1998). The variables are: numvisits, the number of visits to the doctor in 1998; badh, which is an indicator variable of whether the patient self-classified as being in bad health, and age, the age (in years) of the patient.

Analyze how the mean number of doctor’s visits in 1998 varies with age for those who are in bad health, and for those who are not. In particular, test whether or not there is a difference in the relationship between the mean number of doctor’s visits and age for those two types of people (those who classify themselves as being in bad health, and those who don’t). Also, estimate this difference (whether or not you find statistically significant evidence of such a difference).

2. Use the loomis data set from the COUNT package in R for this problem. The only information available about this data set is that it was collected in 2002 as a survey about use of the Snake River in Jackson Hole, Wyoming. Surveys were mailed out to a random sample of Wyoming residents (purchased from a company that specializes in such things) with a personalized cover letter and a dollar enclosed. Response rate was 55%.

The column anvisits gives the number of annual visits reported; gender gives the subject’s reported gender (0 being female, 1 being male); income is a categorical variable giving the subject’s reported income level, higher income being a higher number; and travel gives the travel time for the subject to go to Snake River. You will need to convert these last two to categorical variables for your analysis.

Use this data set to analyze the question of whether or not men and women (in the population to which these results extend, which you should discuss) in 2002 have different average visit frequencies to Snake River, after adjusting for income level and travel distance to Snake River. This should include both a hypothesis test and a confidence interval computation.