Homework #4

The Medical Expenditure Panel Survey is an annual survey which collects information about medical expenditures, income, employment, demographics, health information, &c for a representative sample of Americans.

I have prepared an extract of these data for 1996, and it is available on the course website. The following are the columns in the data, in order:

Variable	Meaning
age	age of person in years
sex	sex of person, 1=male & 0=female
income	income in 1996 $
employed	1=employed, 0=not employed
insured	1=had health insurance, 0=not
health	perceived health status, higher is sicker
spending	spending on health care, 1996 $

To begin with, let's consider a model like the one we used on the midterm:

$$
\begin{aligned}
\text{spending}_i \;=\; & \beta_1 + \beta_2 \text{income} + \beta_3 \text{age} + \beta_4 \text{sex} \\
+\; & \beta_4 \text{employed} + \beta_5 \text{insured} + \beta_6 \text{health} \qquad (1)
\end{aligned}
$$

1. What do you think of the claim that income and sex do not belong in this model?

2. Consider the health status variable. Respondents were asked to rate their health status; their choices were excellent, very good, good, fair, or poor. These were assigned the numerical values 1-5. Does it make sense to enter health status as a single continuous variable as in equation 1?

   Enter health status into the model as a set of dummies, and then test whether they belong.

3. How much more do people in very good health status spend than do people in excellent health status (estimate and CI).

4. Test whether it was correct to enter health status linearly.

5. Test whether insurance affects spending for people of different health statuses differently and discuss.