

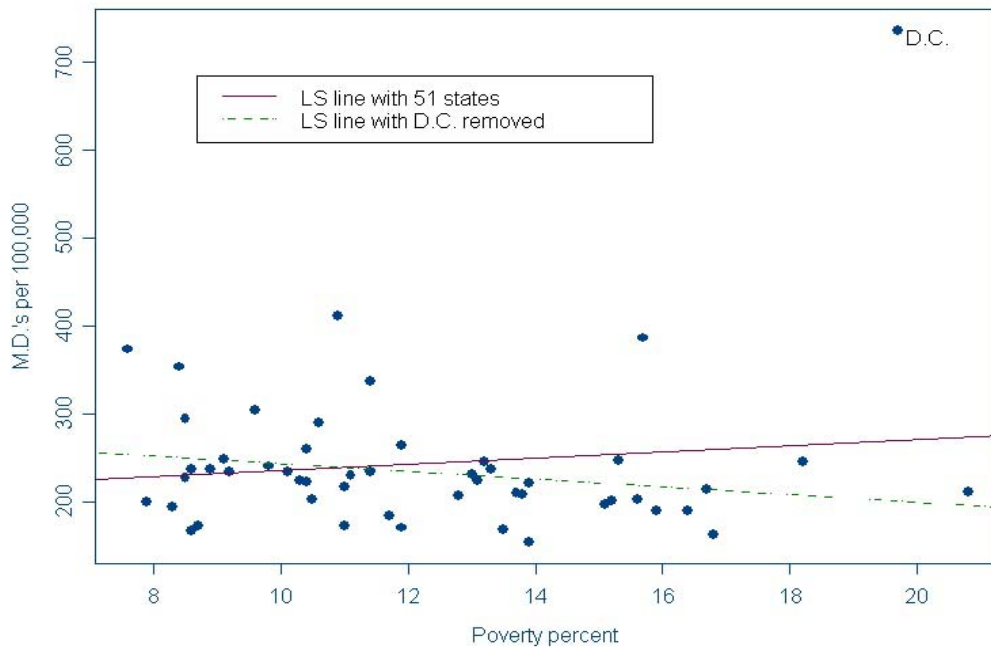
### Question 1

- a) The regression line with “poverty percent” as explanatory variable (X) and “M.D.’s per 100,000” as response variable (Y) using all 51 data points is:

$$\hat{y} = 201.3996 + 3.5096x$$

where slope  $b = 3.5096$  which indicates a positive relationship between poverty percent and M.D.’s per 100,000 which is a bit contradictory with our common sense.

The scatter plot and the line are as below:



- b) D.C. is the second largest horizontally and the largest vertically and is marked on above plot. A point like D.C. this extreme in X, as well as extreme in Y will be influential.
- c) With D.C. removed, we re-calculated the regression line which is:

$$\hat{y} = 287.035 - 4.375x$$

Now, the slope of the line is  $-4.375$  which indicates that the number of doctors now go down with increasing poverty, as we initially expected.

The dramatic change (positive to negative) of regression line slope with and without D.C. shows that D.C. was indeed highly influential.

### Question 2

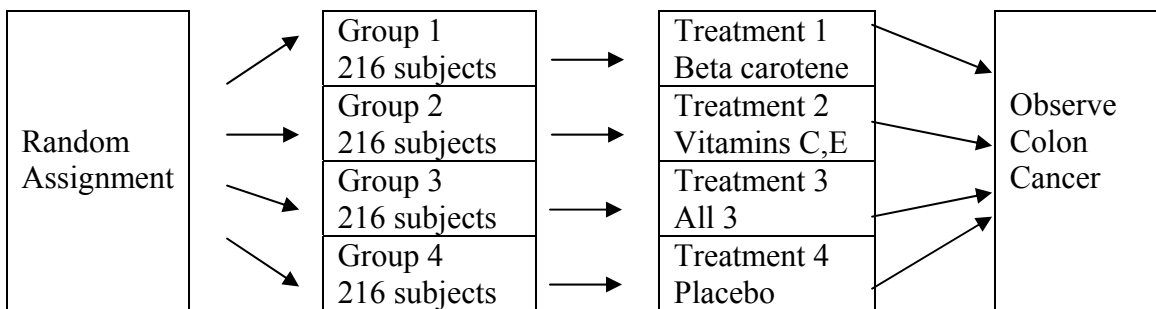
- a) The data was collected after the anesthesia was administered. We did not have any control over whether or not using anesthetics in surgery and neither

randomization for assigning ‘use’ or ‘not use’ anesthetics for each operation. So the National Halothane Study is just an observational study rather than an experiment.

- b) The following are some possible variables that are mixed up with what anesthetic a patient receives: type of surgery, patient allergy to certain anesthetics, patient’s condition like how healthy the patient was before the surgery etc.

**Question 3**

- a) The explanatory variable is diet group and the response variable is whether or not subjects get colon cancer.  
 b) The outline of the experiment is as below:



- c) Assume equal size groups, i.e., 216 in each of the 4 groups. We can first sort the subjects (say in alphabetical order), then label them 001, 002, 003, ... , 864. The following is from Table B starting at line 118. The first 216 groups that are less than 864 will denote subjects for the beta carotene group. 731 **903** 253 304 470 269. Omit 903 because it is greater than 864. Hence subjects labeled 731 253 304 470 269 are chosen for the first group.
- d) Double-blind here means that neither the subject nor the researcher knew which supplement (or placebo) the subject was receiving.
- e) “No significant difference” here means that although the cancer rates may not be exactly the same among the treatment groups, the difference is not larger than might have happened by chance.
- f) Possible lurking variables that could explain why people who eat lots of fruits and vegetables have lower rates of colon cancer are: people who eat more fruits and vegetables may be more health conscious. They may exercise more and see the doctor regularly. They may also eat less “harmful” foods etc.

**Question 4**

The chance of being interviewed is 3/30 for students over age 21 and 2/20 for students under age 21. That is 1/10 in both cases. That means every student has the same chance to be interviewed.

It is not an SRS because not all combinations of students have an equal chance of being interviewed (That is an important characteristic of an SRS). For instance, groups of five students all over age 21 have no chance of being interviewed.