22S: Biostatistics 101 Project (October 22, 2003)

Due December 5 (Friday), 2003

You will be analyzing the data set bed stored on the disk that accompanied the text. I will also put the data set on my web side (www.stat.uiowa.edu/∼jian/S101/Datasets/datasets.html), so that you can download the dataset and read it into R, as we did in our computer lab sessions with other datasets. The statistical package R is recommended for the analysis. However, you may also use other softwares of your choice.

The data set describes the number of community hospital beds per 1,000 population that are available in each state in the United States and the District of Columbia in the years 1980 and 1986. The values for 1980 are stored under the variable named bed80; those for 1986 are saved under bed86. For each state, the number of beds per 1,000 population can be considered an observation of a random variable whose distribution depends on the year. Suppose that you are working for an agency that would like to know if the mean of the 1986 distribution is different from the mean of the 1980 distribution. You are to prepare a report which answers the questions below and supports those answers with the tests and graphs. The paper should be neatly typed (laser printing is not necessary) and address all the issues below bearing in mind that brief reports (about 2 pages not including graphs and computer outputs) will be appreciated. You may place all graphs and R output in an appendix at the end of your report. Make certain that all items are clearly labeled so that the report is easy to read.

1. Graphic examination of the data and summary statistics

(i) Generate a box plot for each of the variables bed80 and bed86. Discuss these plots. For example, you could mention which variable has the largest median or which has the largest or smallest observation.

(ii) Construct a two-way scatter plot with bed86 on the vertical axis. Comment on the appearance of this graph and note any points that vary from the observable trend. Give the number of any state or states with unusual observations. Explain what makes these observations unusual.

(iii) Compute the sample mean and standard deviation for bed80 and bed86.

2. Formal statistical test: Question #1

(i) Conduct a paired t-test to determine whether or not the mean of the 1986 distribution is different from the mean of the 1980 distribution. Give a p-value for

the test and draw a conclusion based on a 0.05 level of significance. Provide the corresponding 95% confidence interval for the difference in means and interpret.

(ii) Conduct a two-sample t-test to determine whether or not the mean of the 1986 distribution is different from the mean of the 1980 distribution. Give a p-value for the test and draw a conclusion based on a 0.05 level of significance. Does this conclusion depend on whether or not common population variance is assumed? Provide the corresponding 95% confidence interval for the difference in means and interpret.

(iii) Discuss which of the tests in part 4 and 5 is appropriate for this data.

(iv) Generate histograms for bed80 and bed86. If the distributions for 1980 and 1986 are not approximately normal, the t-tests in parts 4 and 5 may be invalid. Comment on the validity of t-tests in this case.

3. Formal statistical test: Question # 2

Now suppose that the states numbered 1 to 24 are the states west of the Mississippi River; and suppose that your agency is interested in determining how the distribution of hospital beds per 1,000 population differs between the East and West for the year 1986.

(i) Compute means and standard deviations for both the eastern and western states based on the 1986 data.

(ii) Conduct the appropriate test at the 0.05 significance level to answer your agency's question. Explain the results of the test, and give reasons why you chose to use this test. Back your argument with data and/or graphs if necessary.