## Homework for Lecture 16 Multivariate Regression Analysis Section 9.5

Consider the following subset of the dataset of death rate (B) with a number of explanatory variables:

- A1, the average annual precipitation
- A6, the number of years of schooling for persons over 22
- A9, the size of the nonwhite population
- A10, the number of office workers
- A11, the number of families with an income less than \$3000
- A12, the hydrocarbon pollution index

## References:

R F Gunst and R L Mason, Regression Analysis and Its Applications, Dekker, 1980, pages 370-371.

Helmut Spaeth, Mathematical Algorithms for Linear Regression, Academic Press, 1991, ISBN 0-12-656460-4.

Use the JMP output below to answer the following questions



- 1. Perform the 6-step hypothesis test for the linear model described in JMP.
  - 1)  $H_0$ : \_\_\_\_\_\_ vs.  $H_1$ : \_\_\_\_\_\_

where [fill in]

2) [fill in]

3) test statistic:  $F = \frac{SSR}{SSE}$  follows a F-distribution with df: \_\_\_\_\_ and \_\_\_\_\_

4) [fill in]

- 5) [fill in]
- 6) [fill in]
- 2. Interpret the meaning of  $r^2$  in the context of the problem.

3. Provide the linear regression equation for predicting death rate from the given explanatory variables in the JMP output. Predict the death rate for A1 = 40, A6 = 10, A9 = 20, A10 = 40, A11 = 20, and A12 = 20.

4. Which explanatory variable is the least significant?

5. Do the residuals appear to be healthy? Explain.

Use the JMP output below to answer the following question



6. Are any of the independent variables highly correlated? If so, which ones?