A university medical center urology group was interested in the association between prostate-specific antigen (PSA) and a number of prognostic clinical measurements in men with advanced prostate cancer. Data were collected on 97 men who were about to undergo radical prostectomies. Each line of the data has an identification number and provides information on 8 other variables for each person. The 9 variables are:

|  |  |  |
| --- | --- | --- |
| **Variable Number** | **Variable Name** | **Description** |
| 1 | Identification Number | 1-97 |
| 2 | PSA level | Logarithm of Serum prostate-specific antigen level (mg/ml) |
| 3 | Cancer Volume | Estimate of prostate cancer volume (cc) |
| 4 | Weight | Prostate weight (gm) |
| 5 | Age | Age of patient (years) |
| 6 | Benign prostatic hyperplasia | Amount of benign prostatic hyperlesia ( |
| 7 | Seminal vesicle invasion | Presence or absence of seminal vesicle invasion:1 if yes;0 otherwise |
| 8 | Capsular penetration | Degree of capsular penetration (cm) |
| 9 | Gleason score | Pathologically determined grade of disease using total score of two patterns (summed scores were either 6, 7, or with higher scores indicating worse prognosis) |

The data are given in the prostate.txt file. You are expected to conduct Regression Analysis for relationship between prostate-specific antigen (PSA) and a number of prognostic clinical measurements in men with advanced prostate cancer. For this analysis, please answer the following question.

*(25 pts)* **a)** Find the best model without interaction terms, interpret the results

*(25 pts)* **b)** Is there any multicollinearity problem in the best model ? If yes, please handle it

*(25 pts****)* c)** Is there any confounding factor in the best model ? If yes, please write which variable is confounding variable. Interpret your findings.

*(25 pts)* **d)** Find the best model with interaction terms. Compare the results with part a result.