STA 6166 - Project

Instructions:

- Can work together but the work submitted must be uniquely yours.
- Five (5) pages, with no appendix. Be brief and to the point, each sentence should be 1-2 lines. Five (5) points will be deducted for each page over the limit.
- With each statement/conclusion you make, accompany ONLY the relevant computer output and/or hand calculations.
- Use technically correct methodology, try to avoid "wishy-washy" or "close enough" methods.
- Project needs to be neat, have a flow and easy to follow.
- You are responsible for reading/inputting data into software (from its current format).

Guidelines:

- Use descriptive statistics or overall tests of fit to first determine if it is worthwhile analyzing the dataset (or parts of it) further.
- **Always check if the assumptions of the methodology you plan to use are valid before you proceed**. (If applicable, double check your results with other methods/tools.)
- For linear models, if only the model fit is not good, first try transforming predictors before you transform the response.
- Justify your methodology and conclusions.
- Try to present your findings with an inclusion of a graphic or figure. Remember journal referees love pretty pictures.
- Pay attention to the units of measurement.

1. A university medical center urology group was interested in modeling 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.
   http://www.stat.ufl.edu/~athienit/STA6166/assignment3_1.txt
   Each line of the data set ha an identification number and provides information on 8 other variables

Variable Number	Variable Name	Description
1	ID number	1-97
2	PSA level	Serum prostate-specific antigen level (mg/ml)
3	Cancer volume	Estimate of prostate cancer volume (cc)
4	Weight	Prostate weight (gramms)
5	Age	Age of patient (years)
6	Benign prostatic hyperplasia	Amount of benign prostatic hyperplasia (cm^2)
7	Seminal vesicle invasion	Presence of seminal vesicle invasion: 1 yes; 0 otherwise
8	Capsular penetration	Degree of capsular penetration (cm)
9	Gleason score	Pathologically determined grade of disease. (Scores were either 6, 7, or 8 with higher scores indicating worse prognosis)

Develop a "best" model for predicting PSA and interpret. In addition, create a 90% prediction interval for PSA levels for an individual who has the following values.

Variable Number	Variable Name	Value
3	Cancer volume	4.2633
4	Weight	22.783
5	Age	68
6	Benign prostatic hyperplasia	1.3500
7	Seminal vesicle invasion	0
8	Capsular penetration	0
9	Gleason score	6

2. Three chemical cleaning solvents are a potential hazardous waste. Independent samples of solvents from each type were tested and their sorption rates were recorded as a mole percentage.

Aromatics		Chloroalkanes		Esters		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45	0.91	0.06	0.51	0.09
0.82	1.15	0.57	0.83	0.44	0.10	0.17
0.89	1.12	1.16	0.43	0.55	0.53	0.17
1.05				0.61	0.34	0.60

Analyze the dataset (using appropriate methodology) and determine if there are differences in the mean sorption rate. If so, what are the differences?