Data Analysis Problem

A medical center urology group is studying the association between prostate-specific antigen (PSA) levels and a number of clinical measurements on 97 men with prostate cancer. We focus here on the relationship between cancer volume and PSA. The data are available as prostate.csv on the course website. There are 9 variables for each man:

id = identification number

PSA = prostate-specific antigen level (mg/ml)

cancvol = estimate of prostate cancer volume (cc)

weight = prostate weight (gm)

age = age of patient (yrs)

bph = amount of benign prostatic hyperplasia (cm2)

sem = presence/absence of seminal vesicle invasion

capspen = degree of capsular penetration (cm)

gleason = grade of disease (6, 7, or 8 for these men)

Here is the code to enter the data into SAS:

**data** psa;

infile 'c:Desktop/prostate.csv' dlm = ',' firstobs=**2**;

input id psa cancvol weight age bph sem capspen gleason;

logpsa = log(psa);

logvol = log(cancvol);

run;

Code for a different data set. This data set gathered information on hospital stays in order to determine main causes of hospital infections. In this set of code, we input the data, print it, and obtain a regression model with residuals and influence statistics.

**data** senic;

infile 'c:Desktop/Courses/senic.csv' firstobs=**2** dlm=',';

input id stay age inf cult xray bds m r pat nur facil;

run;

**proc** **print** data=senic; **run**;

ODS rtf;

ODS graphics on;

**proc** **reg** data=senic;

model inf = stay /P CLM CLI CLB influence;

**run**;

ODS Graphics off;

ODS rtf close;