\*Midterm Review Problem 1 (Prostate);

\*To import prostate data;

FILENAME REFFILE '/home/sadiet0/prostate.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=prostatedata;

GETNAMES=YES;

RUN;

\*To sort the data in ascending order;

proc sort data = prostatedata;

by gleason;

run;

\*To address ANOVA assumptions on original data with histograms and q-q plots;

proc univariate data = prostatedata;

by gleason;

var psa;

histogram psa;

qqplot psa;

run;

\*To address ANOVA assumptions on original data with a scatter plot;

proc sgplot data = prostatedata;

scatter x= gleason y = psa;

run;

\*To perform a log transformation;

data prostatedata;

set prostatedata;

logpsa = log(psa);

run;

\*To address ANOVA assumptions on logged data with histograms and q-q plots;

proc univariate data = prostatedata;

by gleason;

var logpsa;

histogram logpsa;

qqplot logpsa;

run;

\*To address ANOVA assumptions on logged data with a scatter plot;

proc sgplot data = prostatedata;

scatter x= gleason y = logpsa;

run;

\*To perform a square root transformation;

data prostatedata;

set prostatedata;

sqrtpsa = sqrt(psa);

run;

\*To address ANOVA assumptions on square-rooted data with histograms and q-q plots;

proc univariate data = prostatedata;

by gleason;

var sqrtpsa;

histogram sqrtpsa;

qqplot sqrtpsa;

run;

\*To address ANOVA assumptions on square-rooted data with a scatter plot;

proc sgplot data = prostatedata;

scatter x= gleason y = sqrtpsa;

run;

\*To test if any medians are different (ANOVA on log trasnformed data);

proc glm data = prostatedata;

class gleason;

model logpsa = gleason;

run;

\*To perform bonferroni adjusted lsmeans t-tests;

proc glm data = prostatedata;

class gleason;

model logpsa = gleason;

lsmeans gleason/ adjust = bon pdiff;

run;

\*To examine a linear trend;

proc glm data = prostatedata order = data;

class gleason;

model logpsa = gleason;

contrast "Linear trend in log transformed psa with respect to gleason" gleason -1 0 1;

run;

\*To examine a quadratic trend;

proc glm data = prostatedata order = data;

class gleason;

model logpsa = gleason;

contrast "quadratic trend in log transformed psa with respect to gleason" gleason 1 -2 1;

run;