ods rtf;  
  
/\* Project: 2009 Community health status Indicators report  
that was compiled by the CHSI Project Working Group as part of the CDC \*/  
/\* Purpose: The motivation for this study is to determine whether certain characteristics  
in a community can predict whether the area will have a health professional shortage.  
I will be analyzing poverty level, population size, population density, and percent populations  
of certain ethnicities to see if any of these variables can be used to predict the likelihood  
of a health Professional shortage in a specific community. \*/  
/\* Programmer: Patrick Lutz \*/  
/\* Date: May 28, 2017 \*/  
/\*Using macro variables to upload data and create the permanent library,CHSI\*/  
%let path=C:\Users\ptl09\Downloads;  
%put &path;  
libname CHSI "&path";  
  
/\*checking to make sure everything's there\*/  
proc contents data=chsi.demographics;  
run;  
  
proc contents data=chsi.riskfactors;  
run;  
  
/\*Creating smaller data sets with only the needed variables\*/  
data demographics1;  
 set chsi.demographics;  
 keep 'State\_FIPS\_Code'n 'Population\_Size'n 'Population\_Density'n Poverty White   
 Black 'Native\_American'n Asian Hispanic;  
run;  
  
data riskfactors1;  
 set chsi.riskfactors;  
 keep 'State\_FIPS\_Code'n 'Community\_Health\_Center\_Ind'n 'HPSA\_Ind'n;  
run;  
  
/\*Merging the two data sets by State\_FIPS\_Code\*/  
data health;  
 merge demographics1 riskfactors1;  
 by 'State\_FIPS\_Code'n;

run;  
  
/\*Setting up an indicator variable for my binary response and deleting all missing observations\*/  
Data health1;  
 set health;  
  
 if 'HPSA\_Ind'n=2 then  
 shortage=1;  
 Else  
 shortage=0;  
  
 If poverty < 0 then  
 delete;  
  
 If 'Population\_Density'n < 0 then  
 delete;  
Run;  
  
/\*Using a panel display of scatterplots with a loess curve to see which  
variables might be a good fit for logistic regression\*/  
proc sgscatter data=health1;  
 compare y=shortage x=('State\_FIPS\_Code'n 'Population\_Size'n   
 'Population\_Density'n poverty)/loess;  
run;  
  
proc sgscatter data=health1;  
 compare y=shortage x=(White Black 'Native\_American'n Asian Hispanic)/loess;  
run;  
  
/\*'State\_FIPS\_Code'n:as expected the state\_fips loess curve is all over the place depending on which state your in\*/  
/\*'Population\_Size'n: It looks like as the population increases the probability of a shortage increases\*/  
/\*'Population\_Density'n: Except for some weird outlier, as density increases, shortage increases\*/  
/\* poverty: Interesting one. As poverty increases, shortage decreases. Maybe poverty demands more health care professionals\*/  
/\* White: As white population increases so does shortage until about 90% white. Then the shortage drops significantly.  
Maybe high white communities have an increased demand for healthcare professionals, until the community is almost all white. Then demand is met\*/  
/\*Black: For small black percent there are increased shortages, but decreasing shortages with higher black percentages\*/  
/\*Native American: same as black more or less\*/  
/\*Asian: All over the place\*/  
/\*Hispanic; same as black\*/  
/\*Running a logistical regression on health data  
with health care profession shortage as the response and poverty as the predictor\*/  
proc logistic data=health1;  
 model shortage(event='1')=poverty;  
run;  
  
/\*Intercept=2.76 Poverty=-0.1154 or Shortage=2.76-0.1154Poverty\*/  
/\* Produce the psuedo r^2 to see how much of the variation in health care shortages  
is accounted for by poverty level \*/  
proc logistic data=health1;  
 model shortage(event='1')=poverty/expb rsquare;  
run;  
  
/\*Create a empirical Logit plot to determine linearity\*/  
proc rank data=health1 groups=4 out=out;  
 /\*4 groups in the  
 output file out\*/  
 var Poverty;  
 ranks bin;  
  
 /\*Create the variable bin for grouping\*/  
run;  
  
/\*Summarize the data by each bin\*/  
proc means data=out nway noprint;  
 class bin;  
 var shortage Poverty;  
 output out=bins sum(shortage)=Shortage mean(Poverty)=Poverty;  
Run;  
  
/\* Empirical Logit Plot\*/  
data bins1;  
 set bins;  
 logit=log((shortage+.5)/(\_FREQ\_ - shortage + .5));  
run;  
  
Title 'Empirical Logit Plot';  
  
proc sgplot data=bins1;  
 reg x=Poverty y=Logit;  
run;  
  
Title;  
  
/\* The model appears to be negatively linear\*/  
/\*Performing a residual analysis to determine if there are any outliers\*/  
proc logistic data=health1;  
 model shortage(event='1')=Poverty/iplots;  
run;  
  
/\*no significant outliers\*/  
/\*Use a correlation analysis to determine whether there are any  
multicollinearity problems with all the predictors included\*/  
proc corr data=health1;  
 var 'Population\_Size'n 'Population\_Density'n Poverty White Black   
 'Native\_American'n Asian Hispanic;  
run;  
  
/\*Black and White are negatively correlated as expected but nothing else indicates a multicollinearity problem\*/  
/\* Variable selection in Logistic Regression using backwards elimination\*/  
proc logistic data=health1;  
 model shortage (event='1')='Population\_Size'n 'Population\_Density'n Poverty   
 White Black 'Native\_American'n Asian Hispanic/selection=backward;  
  
 /\*all sebsets selection\*/  
run;  
  
/\* The best model includes population size, population density, Poverty, Black, Native\_American, Hispanic\*/  
/\* Produce a nested F test to compare the full model with the reduced model\*/  
proc logistic data=health1;  
 model shortage(event='1')='Population\_Size'n 'Population\_Density'n Poverty   
 White Black 'Native\_American'n Asian Hispanic;  
 test Asian, white;  
run;  
  
/\*'Asian' and 'White' had no effect on the model\*/  
/\*Compare ROC curves for the ful model, 'reduced' best model, and the single 'Poverty' model\*/  
proc logistic data=health1 plots=ROC;  
 model shortage (event='1')='Population\_Size'n 'Population\_Density'n Poverty   
 White Black 'Native\_American'n Asian Hispanic;  
 ROC 'best' 'Population\_Size'n 'Population\_Density'n Poverty Black   
 'Native\_American'n Hispanic;  
 ROC 'single' Poverty;  
 ROCcontrast reference('best')/estimate e;  
run;  
  
/\*In Conclusion it appears that poverty level, population size, population density,  
and the percent of Black, Hispanic, and Native American populations in a county  
all have an effect on there being a health care professional shortage in that same community.  
The fact that pop. size and density are the only two variables that increase the likelihood  
of there being a shortage makes sense because with more people and a higher concentration of people in an area,  
there is going to be a higher demand on health care.  
The interesting effect of this study is that an increased poverty level  
and the increased minority populations of Blacks, Hispanics, and Native Americans  
actually seem to decrease the likelihood of there being a shortage.  
This may be because lower income people are actually using health care services less  
or that health care professionals are drawn to lower-income areas because there’s more work. \*/  
ods rtf close;