Neil Yetz

03/24/2016

ERHS 642: Applied Logistic Regression

Dr. Bachand

**MIDTERM EXAM**

Consider Adolescent Placement Study (APS) data set described in the text on pages 26/27.

1. (1 point) Create a permanent SAS data set.

libname sdat 'C:\Users\ndyet\_000\Desktop\Class Folders\Spring 2016\ERHS 642\Data';

/\*data sdat.APS; set APS; run;\*/

**data** APS; set sdat.APS;

**run**;

2. (1 point) Create a new dichotomous outcome variable, PLACE2, as follows:

0 = Outpatient or day treatment

1 = Intermediate residential or residential

**Table 2.1**: Proc Print of newly created PLACE2 variable compared to PLACE variable.



1. (4 points) For the variable LOS (length of hospitalization), obtain separately for those with PLACE2=0 and those with

PLACE2=1

Mean

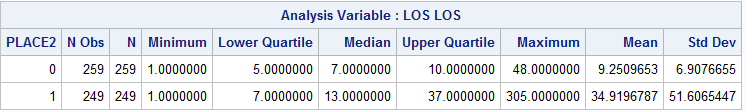
Standard deviation

Quartiles

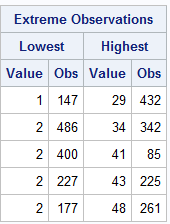
Minimum and maximum

5 lowest and 5 highest values

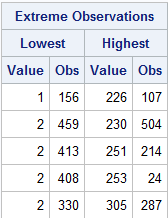
**Table 3.1:** N, minimum, 1st quartile, median, 4th quartile, maximum, mean, and standard deviation separated by PLACE2 variable class using the proc means procedure.



**Table 3.2**: Extreme observations for PLACE2=0.



**Table 3.3**: Extreme observations for PLACE2=1.

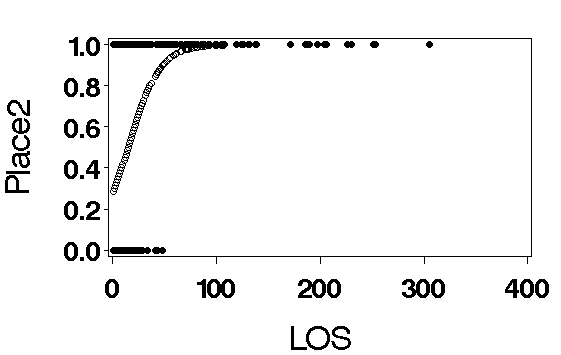


1. (4 points) Obtain the predicted values for LOS. Graph

The predicted values vs PLACE2; and

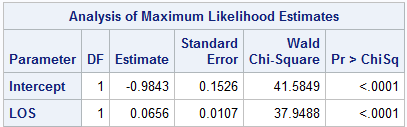
The scatterplot of LOS vs. PLACE2 on the same set of axes.

**Figure 4.1**: scatterplot of LOS vs. PLACE2 with predicted values (pihats) overlaid.



5. (2 points) Show that LOS is a significant predictor of PLACE2.

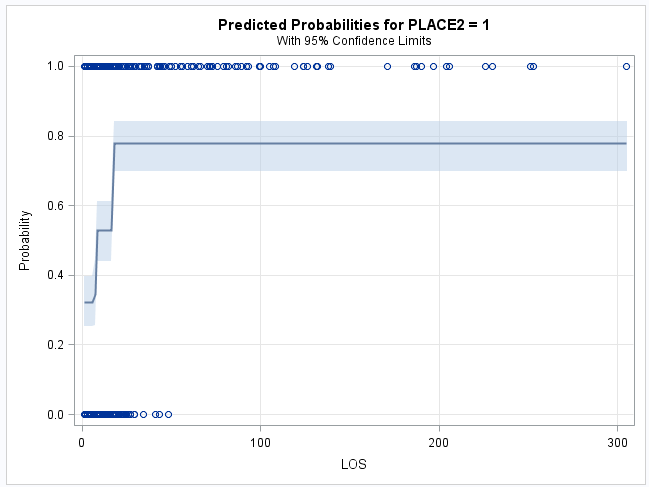
**Table 5.1**: Analysis of Likelihood Estimates for LOS as a predictor of Place 2.



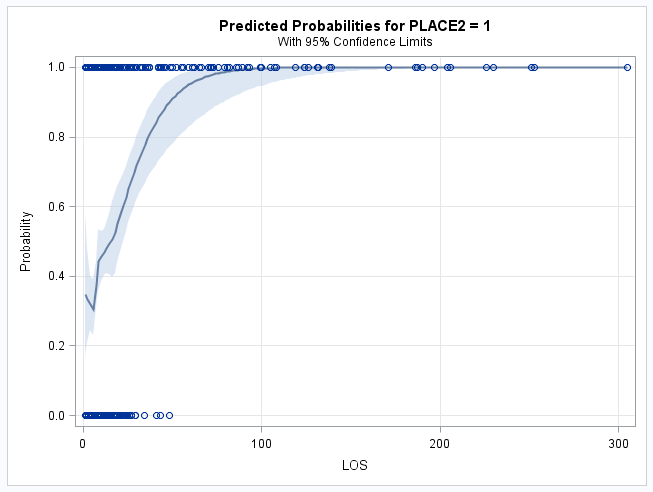
**LOS is a significant predictor of PLACE 2 (p<0.0001).**

6. (4 points) Use spline effect plots to determine if LOS should be transformed. (Select knots and connections you find helpful.)

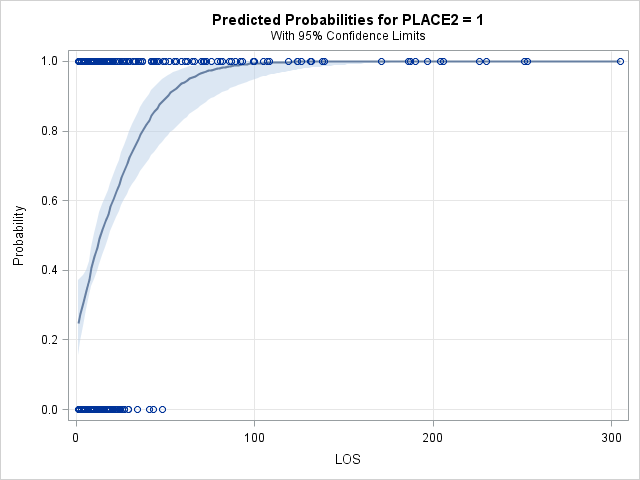
**Figure 6.1**: Spline effect plot using constant connection method showing LOS predicted probabilities of being in PLACE2 using the first, second, and third quartiles as knots.



**Figure 6.2:** Spline effect plot using linear connection method showing LOS predicted probabilities of being in PLACE2 using the first, second, and third quartiles as knots.



**Figure 6.3**: Spline effect plot using cubic connection method showing LOS predicted probabilities of being in PLACE2 using the first, second, and third quartiles as knots.



After viewing all three spline effect plots using quartiles as knots, I believe using transforming LOS is necessary. This is true because none of the methods truly capture all of the points. It is clear that the probability of being in PLACE2 is dependent on how many days your LOS is. Therefore, other transformations need to be considered. Furthermore, it should be noted that other methods of knot placement were assessed, but created such wide confidence intervals that they were deemed not valid due to their arbitrariness.

7. (4 points) Use the fp method to determine the best scale for LOS. Show and describe/explain all results.

**Table 7.1:** Results table for fp transformation macro.



**The results explanations for Table 7.1 are separated by each variable name below:**

**Dev\_Linear**: contains the deviance of the LOS variable from the linear, line. The larger the number, the better it the predicted values fit to the true linear line.

**E\_fp1**: contains the best 1-power transformation. In this case, a “.” Indicates that the linear fitted line is a better fitted line.

**Dev\_fp1**: contains the deviance of the LOS variable from the one-power transformed line. The larger the number, the better it the predicted values fit to the true linear line. In this case, a “.” Indicates that the linear fitted line has a better Deviance.

**E1\_fp2:** contains the best first part of the two power transformation. In this case a “0.5” indicates that the best first part of a 2-power transformation is the

**E2\_fp2:** contains the best 2nd part of the two power transformation. In this case a “3” indicates that the best 2nd part of a 2-power transformation is the . Therefore, the combination of these 2 for the two power transformation results is .

**Dev\_fp2:** contains the deviance of the LOS variable from the two-power transformed line. The larger the number, the better it the predicted values fit to the true linear line.

**P\_lin\_fp1:** indicates how significantly different the linear transformation is from the one-power transformation. In this case, a “.” Indicates that there is no difference between the two or the original linear is a better fitted line.

**P\_lin\_fp2:** indicates how significantly different the linear transformation is from the two-power transformation. In this case, p=0.775, meaning that the two power transformation is similar to the linear line, therefore, you should keep it linear.

**P\_fp1\_fp2:** indicates how significantly different the one-power transformation is from the two-power transformation. In this case, a “.” Indicates that there is no difference between the two or the original two-power transformation is a better fitted line.

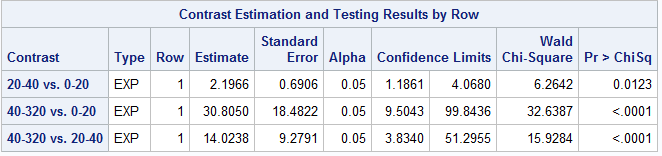
8. (3 points) Based on your results in questions 6 and 7, how would you model LOS? Explain!

Hint: Try out your idea(s) to see if they actually work.

Based on my results from questions 6 & 7, I do not think I would use splines or an fp transformation. Looking at the results, it is clear that the data for LOS with the outcome of PLACE2 do not fit constant connection, linear, or cubic. Furthermore, after assessing the fp transformation method, keeping with the linear method was recommended over any 1-power or 2-power transformation; which was already previously determined to not be the best fitting line for LOS with an outcome of PLACE2.

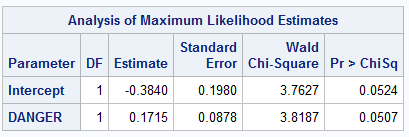
Based on these results I would recommend modeling LOS using a *categorization method.* Initially, I started out categorizing LOS evenly, but that did not work because there is such an odd distribution at the tail end of LOS, leading to many 0 cells in the proc frequency table. Eventually, I found out that using the categories of 0-20, 20-40, and 40-320 (This last category, mainly to show the “rest” really had a better shot of seeing what was going on. As you can tell from table 7.2, initially, the OR is for LOS being 0-20 days in LOS is around 2. Then, as the LOS becomes longer, you can see the OR’s become quite larger. Thus indicating a need for categorization and presentation of multiple OR’s. It should be noted that the confidence intervals are quite wide from 20-40 days, and 40-320 days and should be interpreted with caution.

**Table 7.2:** Contrast statements indicating different categories for LOS as a predictor for outcome PLACE2



9. (3 points) In a model containing only DANGER as the independent variable, is DANGER a significant predictor of PLACE2? Show a table containing the maximum likelihood estimates of the model coefficients and their p-values.

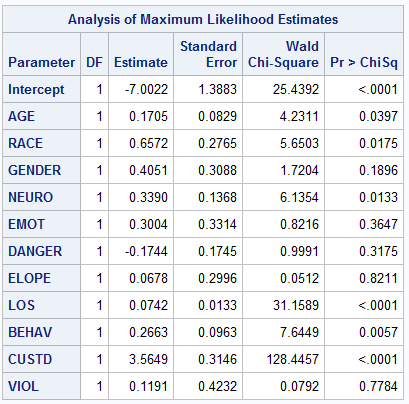
**Table 9.1**: Maximum Likelihood Estimates of DANGER as a predictor for PLACE2.



**Danger is not a significant predictor of PLACE2 at the P<.05 level, but it is relatively close (p=0.0507)**

10. (3 points) In a model containing all study variables as independent variables (except ID and the outcome variables), is DANGER a significant predictor of PLACE2? (For simplicity, keep all continuous variables linear.) Show a table containing the maximum likelihood estimates of the model coefficients and their p-values. Explain why your results for DANGER are different from the results in question 9.

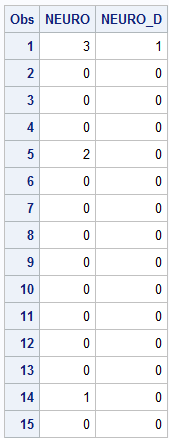
**Table 10.1**: Maximum Likelihood Estimates of all independent variables from the study with PLACE2 as the outcome.



Looking at the DANGER variable Table 10.1, it is clear to see that DANGER is not as significant in the presence of all other variables (p=0.3175). This is because in the presence of other variables, it is hard for DANGER to have as much as an effect. For example, the model is *predicting* for the *outcome* of place 2, the more predictors there are, the more it takes away from the predictive capabilities of the other variables because they are controlling for each other. Therefore, it makes sense that in the presence of less variables, DANGER is a better predictor of PLACE2.

11. (3 points) Create a new variable as follows: NEURO\_D=0 if NEURO=0, 1 or 2 NEURO\_D=1 if NEURO=3 Based on the results in question 10 and other considerations, why does it make sense to combine NEURO categories 0, 1 and 2?

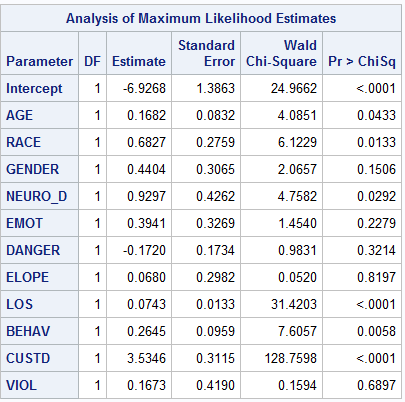
**Table 11.1**: Proc Print of the newly created NEURO\_D variable compared to the NEURO variable



It makes sense to combine Neuro categories 0, 1, and 2 because these are NONE, MILD, and MODERATE. Compared to the code 3, which means SEVERE. Categories, 0, 1, and 2 may not have the same implications as 3 has, therefore it makes sense to combine all three of these categories into one as it likely to give a better representation of the outcome.

12. (2 points) In the model you created in question 10, replace the 4-category NEURO variable with the dichotomous variable NEURO\_D you created in question 11. Show a table containing the maximum likelihood estimates of the model coefficients and their p-values. Is NEURO\_D statistically significant?

**Table 12.1**: Maximum Likelihood Estimates with NEURO\_D replacing NEURO and all other independent variables from the study with PLACE2 as the outcome.

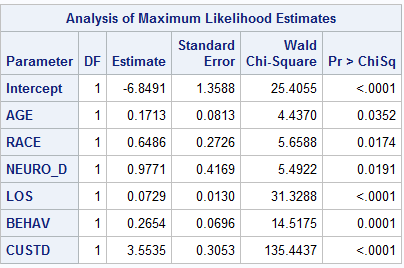


**Neuro\_D is a significant predictor of the PLACE2 outcome (p=0.0292).**

13. Start with the model you created in question 12. Remove all variables that are non-significant at the 0.05 level.

1. (2 points) Show a table of the reduced model containing the maximum likelihood estimates of the model coefficients and their p-values.

**Table** 13.1: Maximum Likelihood Estimates of the reduced model from item 12 with the outcome of PLACE2.

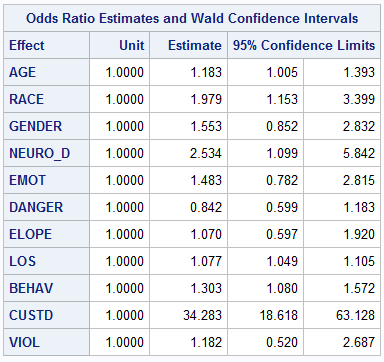


1. (4 points) Determine if the full model from question 12 is significantly better than the reduced model.

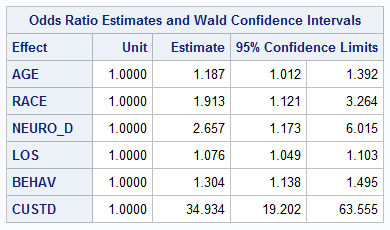
All of the variables in the reduced model remained significant; therefore, the reduced model is significantly better than the full model in question 12.

14. (4 points) Compare the ORs from the models in questions 12 and 13 and determine if there is evidence of confounding.

**Table 14.1**: Crude ORs from full model



**Table 14.2**: Adjusted ORs from reduced model



Confounding =

**AGE**= = 0.003 = 0.3% **NO EVIDENCE OF CONFOUNDING**

**RACE** = = 0.035 = 3.5% **NO EVIDENCE OF CONFOUNDING**

**Neuro\_D** = = 0.046 = 4.6% **NO EVIDENCE OF CONFOUNDING**

**LOS** = = 0.0009 = 0.09% **NO EVIDENCE OF CONFOUNDING**

**BEHAV** = = 0.0008 = 0.08% **NO EVIDENCE OF CONFOUNDING**

**CUSTD** = = 0.019 =1.9% **NO EVIDENCE OF CONFOUNDING**

**There is NO evidence of confounding.**

15. Start with the reduced model you created in question 13.

a. (3 points) Determine if NEURO\_D is a multiplicative effect modifier of any of the other model covariates (use the 0.1 level of significance)

**Table 15.1**: Maximum Likelihood Estimates of the interaction of each variable from the reduced model with the NEURO\_D variable. NOTE: These were conducted through 5 separate Proc logistic procedures and only the interaction section is shown to save space.







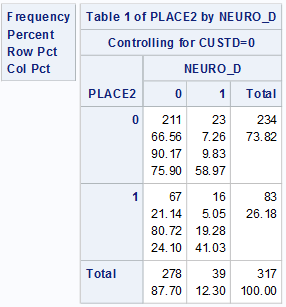




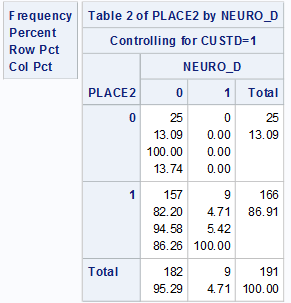
**Neuro\_D is a multiplicative effect modifier of RACE (p=0.0709) and LOS (p=0.0797).**

1. (3 points) Notice that inclusion of the interaction between NEURO\_D and CUSTD results in quasicomplete separation. Explain why this happens and show results to back up your conclusion.

**Table 15.2**: Proc Freq of PLACE2 and Nuero\_D controlling for CUSTD=0



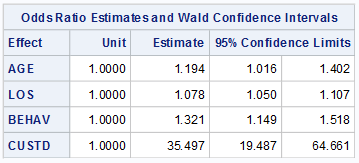
**Table 15.3**: Proc Freq of PLACE2 and Nuero\_D controlling for CUSTD=1



One very likely reason that the interaction from Neuro\_D and CUSTD is “Quasicomplete” is because there were **no** participants in this study that fulfilled the role of having a severe neuropsychiatric disorder, in state custody and in Intermediate residential or residential treatment. As is shown in Table 15.3. Having 0 in any frequency cell does not allow for a true logistic regression procedure, therefore not al conditions were satisfied due to the 0 participants in one cell creating a quasicomplete separation.

16. (4 points) Start with the reduced model you created in question 13 and add the interaction between NEURO\_D and RACE. Present meaningful ORs and 95% CIs for the model covariates.

**Table 16.1**: Odds ratio estimated and 95% Confidence Intervals for variables in the reduced model from question 13.



17. (3 points) Interpret the ORs and 95% CIs from question 16.

Controlling for LOS, BEHAV, CUSTD & the interaction between RACE & NEURO\_D for every year older someone is, 1.194 times as likely to end up in Intermediate Residential or Residential treatment. We can say with 95% confidence that the true odds ratio lies somewhere between 1.016 and 1.402.

Controlling for AGE, BEHAV, CUSTD & the interaction between RACE & NEURO\_D for every day longer their length of stay is they are 1.078 times as likely to end up in Intermediate Residential or Residential treatment. We can say with 95% confidence that the true odds ratio lies somewhere between 1.050 and 1.107.

Controlling for AGE, LOS, CUSTD & the interaction between RACE & NEURO\_D for every point higher their Behavioral score is 1.321 times as likely to end up in Intermediate Residential or Residential treatment. We can say with 95% confidence that the true odds ratio lies somewhere between 1.149 and 1.518.

Controlling for AGE, LOS, BEHAV & the interaction between RACE & NEURO\_D people also in state custody are 35.497 times as likely to end up in Intermediate Residential or Residential treatment. We can say with 95% confidence that the true odds ratio lies somewhere between 19.487 and 64.661. Note: Because the Confidence intervals for CUSTD is so wide, these results should be interpreted with caution.

18. (3 points) Based on earlier analyses in this exam, is your conclusion for LOS correct? Explain!

**NO**, our conclusion for LOS was NOT correct because based on the spline analyses and fp analyses, LOS needs to be somehow categorized. Length of stay in LOS is contingent on categorization and will need multiple odds ratios to come up with proper conclusions. I conclude it is necessary to talk to the experts and look into reasons LOS would differ in PLACE2 and present differing OR’s for individuals based on their LOS.

**YOU MUST SHOW ALL RELEVANT MODEL OUTPUT**

**SAS CODE**

libname sdat 'C:\Users\ndyet\_000\Desktop\Class Folders\Spring 2016\ERHS 642\Data';

\*Question 1;

/\*data sdat.APS; set APS; run;\*/

**data** APS; set sdat.APS;

\*Question 2;

if PLACE=**0** then PLACE2=**0**;

else if PLACE=**1** then PLACE2=**0**;

else if PLACE=**2** then PLACE2=**1**;

else if PLACE=**3** then PLACE2=**1**;

\*question 11;

if **0**=<Neuro<**3** then NEURO\_D=**0**;

else if Neuro=**3** then Neuro\_D=**1**;

/\* Question 12

This makes sense because now we can make a proper

assessmnet, as mild and moderate are not nearly as bad as severe\*/

/\*

if 0<=LOS<80 then x\_c=1;

else if 80<=LOS<160 then x\_c=2;

else if 160<=LOS<240 then x\_c=3;

else if 240<=LOS<320 then x\_c=4;

\*/

if **0**<=LOS<**20** then x\_c=**1**;

else if **20**<=LOS<**40** then x\_c=**2**;

else if **40**<=LOS<**320** then x\_c=**3**;

/\*

if 0<=LOS<20 then x\_c=1;

else if 20<=LOS then x\_c=2;

\*/

**run**;

if **11**=<AGE<**12** then AGE2=**11**;

else if **12**=<AGE<**13** then AGE2=**12**;

else if **13**=<AGE<**14** then AGE2=**13**;

else if **14**=<AGE<**15** then AGE2=**14**;

else if **15**=<AGE<**16** then AGE2=**15**;

else if **16**=<AGE<**17** then AGE2=**16**;

else if **17**=<AGE<**18** then AGE2=**17**;

**run**;

**proc** **print** data=APS;

var Neuro Neuro\_D AGE AGE2;

**run**;

**proc** **freq** data=APS;

tables PLACE2\*LOS;

**run**;

**proc** **freq** data=APS;

tables PLACE2\*X\_C;

**run**;

**proc** **print** data=APS;

var PLACE PLACE2 los;

**run**;

**proc** **sort** data=APS;

by descending PLACE2;

**run**;

\*Question 3;

**proc** **univariate** data=APS;

var los;

class PLACE2;

**run**;

**proc** **means** data=APS

mean std min q1 median q3 max;

var los AGE;

class PLACE2;

**run**;

\*Question 4;

**proc** **logistic** descending data=APS;

model PLACE2=LOS;

output out=pdat p=pihat;

**run**;

**proc** **print** data=pdat;

var PLACE2 LOS pihat; **run**;

axis1 minor=none label=(f=swiss h=**2.5** 'LOS');

axis2 minor=none label=(f=swiss h=**2.5** a=**90** 'Place2');

goptions FTEXT=swissb HTEXT=**2.0** HSIZE=**6** in

VSIZE=**6** in;

symbol1 c=black v=dot;

symbol2 c=black v=circle;

symbol3 c=black v=star h=**2**;

**proc** **gplot** data=pdat;

plot (PLACE2 pihat)\*LOS/overlay haxis=axis1 vaxis=axis2;

**run**; **quit**;

\*question 5;

**proc** **logistic** descending data=APS;

model Place2=LOS;

**run**;

**Proc** **univariate** data=APS;

var LOS;

**run**;

/\*question 6; quartiles used as cutpoints.

Not a good representation, need to do fp procedure;\*/

**proc** **logistic** descending data=APS;

effect LOSs=spline(LOS/knotmethod=list(**6** **8** **17**)

basis=tpf(noint) degree=**0**);

model PLACE2=LOSs;

effectplot;

**run**;

**proc** **logistic** descending data=APS;

effect LOSs=spline(LOS/knotmethod=list(**6** **8** **17**)

basis=tpf(noint) degree=**1**);

model PLACE2=LOSs;

effectplot;

**run**;

**proc** **logistic** descending data=APS;

effect LOSs=spline(LOS/knotmethod=list(**6** **8** **17**)

basis=tpf(noint) naturalcubic);

model PLACE2=LOSs;

effectplot;

**run**;

\*\* Macro for fp assessment \*\*;

**%macro** fp1(dset,y,var,lb,p1);

%do %until(&p1=**7**);

%put \*\*\*\*\* &p1 \*\*\*\*\*;

ODS output FitStatistics = mfs;

data fpdat; set &dset; if &var>&lb; pc=&p1/**2**;

if pc ne **0** then F1=&var\*\*pc; else if pc = **0** then F1=log(&var);

run;

proc logistic descending data=fpdat;

model &y=F1; \*-------------------F1 represents the variable being tested

for scale;

run;

data mfs; set mfs; if criterion='-2 Log L'; drop Criterion InterceptOnly;

run;

proc append data=mfs base=tres; run;

proc datasets; delete fpdat mfs; run;

quit;

%let p1=%eval(&p1+1);

%end;

**%mend** fp1;

%***fp1***(APS,PLACE2,LOS,**0**,-**4**); \*-----------Enter data set name, outcome

variable name and name of variable being tested for scale;

**data** pvals; do p1=-**4** to **6**; output; end; **run**;

**data** pvals; set pvals; p1=p1/**2**; **run**;

**data** tres; merge pvals tres; if p1 in (-**1.5**, **1.5**, **2.5**) then delete; **run**;

**proc** **sort** data=tres; by InterceptAndCovariates; **run**;

**data** tres; set tres; if \_N\_=**1** or p1=**1**; **run**;

**%macro** fp2(dset,y,var,lb,p1,p2);

%do %until(&p1=**7**);

%do %until(&p2=**7**);

%put \*\*\*\*\* &p1 &p2 \*\*\*\*\*;

ODS output FitStatistics = mfs;

data fpdat; set &dset; if &var>&lb; pc1=&p1/**2**; pc2=&p2/**2**;

if pc1 ne **0** then F1=&var\*\*pc1; else if pc1 = **0** then F1=log(&var);

if pc1 ne pc2 then do; if pc2 ne **0** then F2=&var\*\*pc2;

else if pc2 = **0** then F2=log(&var); end;

if pc1=pc2 then F2=F1\*log(&var);

run;

proc logistic descending data=fpdat;

model &y=F1 F2; \*------------F1 and F2 represent the variable being tested

for scale;

run;

data mfs; set mfs; if criterion='-2 Log L'; drop Criterion InterceptOnly;

run;

proc append data=mfs base=tres2; run;

proc datasets; delete fpdat mfs; run;

quit;

%let p2=%eval(&p2+1);

%end;

%let p2=%eval(-4);

%let p1=%eval(&p1+1);

%end;

**%mend** fp2;

%***fp2***(APS,PLACE2,LOS,**0**,-**4**,-**4**); \*-----------Enter data set name, outcome

variable name and name of variable being tested for scale;

**data** pvals2; do p1=-**4** to **6**; do p2=-**4** to **6**; output;end; end; **run**;

**data** pvals2; set pvals2; p1=p1/**2**; p2=p2/**2**; **run**;

**data** tres2; merge pvals2 tres2;

if p1 in (-**1.5**, **1.5**, **2.5**) or p2 in (-**1.5**, **1.5**, **2.5**) then delete; **run**;

**proc** **sort** data=tres2; by InterceptAndCovariates; **run**;

**data** tres2; set tres2; if \_N\_=**1**; **run**;

**data** comb; set tres tres2; **run**;

**data** c1; set comb; if p1=**1** and p2=**.**; rename

InterceptAndCovariates=Dev\_linear;

drop p1 p2; **run**;

**data** c2; set comb; if p1 ne **1** and p2=**.**; rename

InterceptAndCovariates=Dev\_fp1;

rename p1=e\_fp1; drop p2; **run**;

**data** c3; set comb; if p2 ne **.**; rename InterceptAndCovariates=Dev\_fp2;

rename p1=e1\_fp2; rename p2=e2\_fp2; **run**;

**data** c;

merge c1 c2 c3;

diff\_lin\_fp1=Dev\_linear-Dev\_fp1;

diff\_lin\_fp2=Dev\_linear-Dev\_fp2;

diff\_fp1\_fp2=Dev\_fp1-Dev\_fp2;

p\_lin\_fp1=**1**-probchi(diff\_lin\_fp1,**1**);

p\_lin\_fp2=**1**-probchi(diff\_lin\_fp2,**3**);

p\_fp1\_fp2=**1**-probchi(diff\_fp1\_fp2,**2**);

**run**;

**proc** **print** noobs data=c;

var Dev\_linear e\_fp1 Dev\_fp1 e1\_fp2 e2\_fp2 Dev\_fp2 p\_lin\_fp1 p\_lin\_fp2

p\_fp1\_fp2;

format p\_lin\_fp1 p\_lin\_fp2 p\_fp1\_fp2 **6.4**;

**run**;

**proc** **datasets**; delete tres tres2 pvals pvals2 comb c c1 c2 c3; **run**; **quit**;

\* End macro for fp assessment \*;

**proc** **freq** data=APS;

tables AGE;

**run**;

**proc** **logistic** descending data=APS;

class x\_c/param=ref ref=first;

model PLACE2=x\_c;

contrast '20-40 vs. 0-20' x\_c **1** **0**/estimate=exp;

contrast '40-320 vs. 0-20' x\_c **0** **1**/estimate=exp;

contrast '40-320 vs. 20-40' x\_c -**1** **1**/estimate=exp;

**run**;

**proc** **freq** dat=APS;

tables DANGER Neuro;

**run**;

\*question 9 Danger NOT a sig predictor;

**proc** **logistic** descending data=APS;

model PLACE2=DANGER;

**run**;

/\*question 10: DANGER still not sig predictor

BUt by much les than w/o all other variables\*/

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE GENDER NEURO EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL;

**run**;

/\*Question 12

Neuro\_D statistically significant\*/

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE GENDER NEURO\_D EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL;

**run**;

\*Question 13 reduced model is better;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD;

**run**;

\*question 14;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE GENDER NEURO\_D EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL/clodds=Wald;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD/clodds=Wald;

**run**;

/\*Question 15a;

Neuro\_D has Sigificant interaction on:RACE (.07)LOS(.079) )\*/

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD AGE\*Neuro\_D;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD RACE\*Neuro\_D;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD LOS\*Neuro\_D;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD BEHAV\*Neuro\_D;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD CUSTD\*Neuro\_D;

**run**;

/\*Question 15b;

Cell has 0 for frequency\*/

**proc** **freq** data=APS;

table CUSTD\*PLACE2\*Neuro\_D;

**run**;

\*Question 16;

**proc** **freq** data=APS;

table Neuro\_D;

**run**;

**proc** **logistic** descending data=APS;

model PLACE2= AGE RACE NEURO\_D LOS BEHAV CUSTD NEURO\_D\*RACE/clodds=wald;

**run**;

/\*Question 17; CUSTDY VERY High with wide confidence intervals\*/