Neil Yetz

04/26/2016

ERHS 642

Dr. Bachand

**ERHS 642 Logistic Regression Spring 2016**

**Homework Assignment 9**

Consider the **ICU\_altered** data set.

1.

1. Present a table containing coefficients, standard errors, Wald Chi-Square values and p-values for your final model from homework assignment 6
2. Table 1.1: Table presenting coefficients, standard errors, Wald Chi-Square values, and p-values for Final model from homework assignment 6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Standard Error | Wald Chi-Square | P-Value |
| Intercept | -4.4712 | 1.0104 | 19.5839 | <0.0001 |
| Service at ICU admission: | -1.7275 | 0.4272 | 16.3527 | <0.0001 |
| PO2 from initial blood gases | 0.5553 | 0.5298 | 1.0988 | 0.2945 |
| AGE | 0.0602 | 0.0151 | 15.8519 | <0.0001 |
| Cancer part of the present problem | 1.3949 | 0.6532 | 4.5607 | 0.0327 |

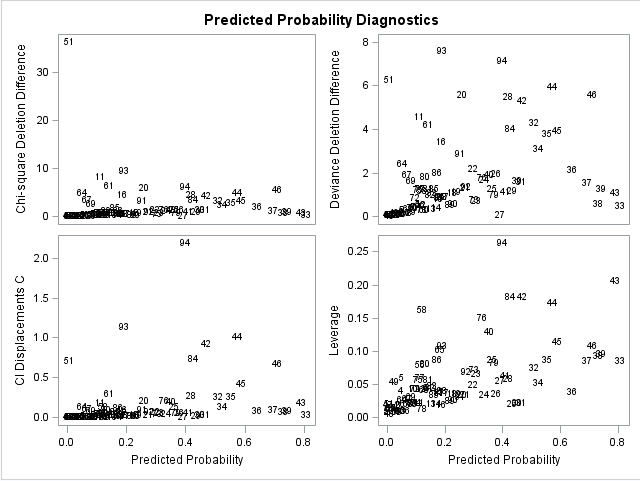
1. Present a table containing odds ratios, 95% confidence intervals and p-values for your final model from homework assignment 6.
2. Table 1.2: Odds Ratios, 95% confidence intervals, and p-values for final Model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Comparison/Unit | OR | 95%CI | | P-Value |
| Service at ICU admission: | Surgical vs Medical | 0.178 | 0.077 | 0.411 | <0.0001 |
| PO2 from initial blood gases | <=60 vs >60 | 1.742 | 0.617 | 4.922 | 0.2945 |
| AGE | 10 | 1.825 | 1.357 | 2.454 | <0.0001 |
| Cancer part of the present problem | Yes vs No | 4.035 | 1.122 | 14.514 | 0.0327 |

2. Use logistic regression diagnostics to identify outliers.

1. Present scatterplots of h, difchisq, difdev and db vs. pihat

Figure 2.1: Scatterplots of h, difichisq, difdev, and db vs their corresponding pihats to identify potential outliers



1. Present a summary table describing the outliers

Table 2.1: Summary table describing the outliers separated by covariate pattern.



1. Explain why the outliers are outliers1

**Covariate pattern outlier 42:** All three individuals fitting this covariate pattern had relatively good factors. Thus, causing them to seem slightly less likely to experience the outcome of death. Yet, all 3 did. I think part of the reason this is an outlier, mainly is because there were 3 of them that fit the exact same profile, and all experienced the outcome, and their pihat was just slightly below the 0.50 threshold, therefore they seem like outliers.

**Covariate pattern outlier 51:** All four individuals in this covariate pattern had really good chances of surviving. Although they were in the emergency setting, they were young, had lower pO2 scores, and no cancer. Yet, 1 individual experienced the outcome of death, therefore causing this to seem like an outlier.

**Covariate pattern outlier 84:** The two individuals in this covariate pattern both experienced death. They were in the emergency setting, making it more likely to experience the outcome. They had cancer as well, again leading them to be more likely. Furthermore, they were old (age 75), yet, the pihat was slightly below 0.50 causing it to seem odd (according to SAS, atheist) that they experienced the outcome of death.

**Covariate pattern outlier 93:** The two individuals in this covariate pattern both experienced the outcome of death. These individuals were in the emergency services and had an old age, thus making them seem like they were at a high risk for death (as I would think), but the there pihats were actually deemed very low, thus making them seem like extreme outliers.

**Covariate pattern outlier 94:** The three individuals in this covariate pattern both experienced the outcome of death. These individuals were in the emergency services and had an old age, thus making them seem like they were at a high risk for death (as I would think), but the there pihats were actually deemed very low, thus making them seem like extreme outliers.

1. Present one or more tables showing effect of deleting outliers on model coefficients&p-values, GOF statistics&p-values and model stability

Table 2.2: Odds rations and corresponding p-values if outlier covariate pattern is are removed.



Table 2.3: Table showing changes in GOF statistics for Hosmer-Lemshow test with outlier covariate patterns removed compared to the original model statistics.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Hosmer-Lemshow Test** | | |
| Outlier | Chi-square | DF | p-value |
| Original | 17.4168 | 8 | 0.0261 |
| 42 | 7.756 | 7 | 0.3546 |
| 51 | 8.3 | 8 | 0.4047 |
| 84 | 8.546 | 8 | 0.382 |
| 93 | 8.834 | 8 | 0.3564 |
| 94 | 9.688 | 8 | 0.2876 |

Table 2.4: Table showing changes in GOF statistics for Pearson and Deviance GOF tests with outlier covariate patterns removed compared to the original model statistics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Pearson** | | | **Deviance** | | |
| Outlier | Value | DF | p-value | Value | DF | p-value |
| Original | 148.614 | 89 | <0.0001 | 132.871 | 89 | 0.0018 |
| 42 | 152.868 | 88 | <0.0001 | 127.565 | 88 | 0.0038 |
| 51 | 123.917 | 88 | 0.007 | 126.57 | 88 | 0.0045 |
| 84 | 141.993 | 88 | 0.0002 | 128.853 | 88 | 0.003 |
| 93 | 156.619 | 88 | <0.0001 | 125.246 | 88 | 0.0056 |
| 94 | 149.507 | 88 | <0.0001 | 125.714 | 88 | 0.0052 |

Table 2.5: Table showing changes in GOF statistics for Osius-Rojek GOF test and Stukel p-value test for the tails with outlier covariate patterns removed compared to the original model statistics.

|  |  |  |
| --- | --- | --- |
|  | **Osius-Rojek** | **Stukel** |
| Outlier | p-value | p-Value |
| Original | 0.005 | 0.285 |
| 42 | 0.0083 | 0.1662 |
| 51 | 0.292 | 0.933 |
| 84 | 0.0112 | 0.154 |
| 93 | 0.0105 | 0.165 |
| 94 | 0.0064 | 0.116 |

1. Propose how to deal with the outliers (a hypothetical answer is sufficient; don’t redo the analysis)

Personally, I do not like the idea of removing outliers at all. Removing the outliers honestly just seems like you are trying to sway the data in the direction you want. But, there are definitely instances where I am sure that the outliers are stopping you from seeing the true effect that is there.

In this case of this, I would **NOT** remove any outlier covariates. Even with most of them being removed, it does not seem to change the GOF tests much. The one I would most consider removing would be covariate pattern 51 because it does definitely seem like an odd one when I look at the pattern with such a young age definitely makes it seem off. The removal of case 51 definitely changes some of the results of the Osius-Rojek GOF test (0.005 to 0.292) and the Hosmer-Lemishow GOF test (0.0261 to 0.404) which definitely seems to be a case that the outlier is definitely effecting the model and they both make it seem much more fit. But again, I do not think it is wise to remove any of the outlier covariates that is the only one I would somewhat consider though.

SAS Code

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

**data** ICU\_altered; set sdat.ICU\_altered;

if race=**1** then do; r1=**0**; r2=**0**; end;

else if race=**2** then do; r1=**1**; r2=**0**; end;

else if race=**3** then do; r1=**0**; r2=**1**; end;

if **16**<= SYS <**110** then SYSa=**0**;

else if **110**<= SYS <**150** then SYSa=**1**;

else if SYS >= **150** then SYSa=**2**;

**run**;

\*\* Final model from chapter 4\*\*;

**proc** **logistic** descending data=ICU\_altered;

model STA=SER PO2 age CAN;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Pearson chi-square, deviance and Hosmer-Lemeshow test \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **logistic** descending data=ICU\_altered;

model STA=SER PO2 age CAN

/scale=n aggregate lackfit;

**run**;

\*"/scale=n aggregate lackfit" gives you chi square, H-L and deviance, and remember H-L High p-value=good;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_altered; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_altered; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_altered; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Diagnostics \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **sort** data=ICU\_altered; by SER PO2 age CAN; **run**;

**proc** **means** n sum noprint data=ICU\_altered;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j;

**run**;

**proc** **logistic** descending data=jdat plots(only label)=(phat);

model y\_j/m\_j=SER PO2 age CAN;

output out=diag h=h difchisq=difchisq difdev=difdev c=db p=pihat;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\* Evaluate outliers \*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** diag; set diag; j=\_N\_; **run**;

**proc** **print** data=diag noobs; where j in (**51**, **94**, **93**, **84**, **42**); \*<-- you change these numbers based on the predicted probability dignostics. You are looking for outliers;

var j SER PO2 age CAN y\_j m\_j pihat h difchisq difdev db; \*<-- Change to your variables do NOT include interactions keep everything else that is not a variable;

format pihat h difchisq difdev db **7.3**; \*<-- keep this the same, this is just making your table look nice.;

**run**;

\*You don't need to use the below macro if it confuses you... but it will make things way easier;

**%macro** outliers(j);

ODS output Logistic.ContrastEstimate=ors;

proc logistic descending data=diag; where j ne &j;

model y\_j/m\_j=SER PO2 age CAN;

contrast 'SER 1 vs 0' SER **1**/estimate=exp; \*<- insert one contrast per effect;

contrast 'PO2 1 vs 0' PO2 **1**/estimate=exp; \*<- insert one contrast per effect;

contrast 'AGE increase of 10 years' AGE **10**/estimate=exp; \*<- insert one contrast per effecct;

contrast 'CAN 1 vs 0' CAN **1**/estimate=exp; \*<- insert one contrast per effect;

run;

data ors&j; set ors;

drop type row stderr alpha lowerlimit upperlimit waldchisq;

rename estimate=OR&j probchisq=p&j; run;

proc print data=ors&j; run;

**%mend** outliers;

%***outliers***(**0**);

%***outliers***(**42**);

%***outliers***(**51**);

%***outliers***(**84**);

%***outliers***(**93**);

%***outliers***(**94**);

**data** ors; merge ors0 ors42 ors51 ors84 ors93 ors94; **run**;

**proc** **print** data=ors noobs;

var contrast or0 or42 or51 or84 or93 or94 p0 p42 p51 p84 p93 p94;

format ors0 ors42 ors51 ors84 ors93 ors94 **6.2**;

**run**;

**proc** **datasets**; delete ors; **run**; **quit**;

\*\*\*\*\*\*\*\*\*deleting outlier covariate patterns for HL, Person & deviance\*\*\*\*\*\*\*\*;

\*Take out 42\*;

**proc** **logistic** descending data=diag;

where j ne (**42**);

model y\_j/m\_j = SER PO2 age CAN/scale=n aggregate lackfit;

**run**;

\*Take out 51\*;

**proc** **logistic** descending data=diag;

where j not in (**51**);

model y\_j/m\_j = SER PO2 age CAN/scale=n aggregate lackfit;

**run**;

\*Take out 84\*;

**proc** **logistic** descending data=diag;

where j not in (**84**);

model y\_j/m\_j = SER PO2 age CAN/scale=n aggregate lackfit;

**run**;

\*Take out 93\*;

**proc** **logistic** descending data=diag;

where j not in (**93**);

model y\_j/m\_j = SER PO2 age CAN/scale=n aggregate lackfit;

**run**;

\*Take out 94\*;

**proc** **logistic** descending data=diag;

where j not in (**94**);

model y\_j/m\_j = SER PO2 age CAN/scale=n aggregate lackfit;

**run**;

\*Finding outlier covariate patterns;

\*42;

**Proc** **print** data=ICU\_altered;

where SER=**0** and PO2=**1** and AGE=**63** and CAN=**0**;

**run**;

\*51;

**Proc** **print** data=ICU\_altered;

where SER=**1** and PO2=**0** and AGE=**20** and CAN=**0**;

**run**;

\*84;

**Proc** **print** data=ICU\_altered;

where SER=**1** and PO2=**0** and AGE=**75** and CAN=**1**;

**run**;

\*93;

**Proc** **print** data=ICU\_altered;

where SER=**1** and PO2=**1** and AGE=**70** and CAN=**0**;

**run**;

\*94;

**Proc** **print** data=ICU\_altered;

where SER=**1** and PO2=**1** and AGE=**87** and CAN=**0**;

**run**;

\*creating new datasets deleting covariate patterns;

\*data w/o covariate pattern 42;

**data** ICU\_42;

set ICU\_altered;

where ID not in (**16**, **17**, **18**);

**run**;

**proc** **print** data=ICU\_42; **run**;

\*data w/o covariate pattern 51;

**data** ICU\_51;

set ICU\_altered;

where ID not in (**96**, **194**, **195**, **200**);

**run**;

\*data w/o covariate pattern 84;

**data** ICU\_84;

set ICU\_altered;

where ID not in (**52**, **53**);

**run**;

\*data w/o covariate pattern 93;

**data** ICU\_93;

set ICU\_altered;

where ID not in (**56**, **57**);

**run**;

\*data w/o covariate pattern 94;

**data** ICU\_94;

set ICU\_altered;

where ID not in (**1**, **2**, **3**);

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test -cov 42 \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_42; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_42; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test -51 \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_51; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_51; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test -84 \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_84; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_84; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test -93 \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_93; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_93; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\* Osius-Rojek test -94 \*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* If interaction terms are in the model, create the interaction terms ;

\* If categorical variables with more than 2 categories are in the model;

\* create design variables ;

\* Sort new data set by model covariates;

**proc** **sort** data=ICU\_94; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

**run**;

\* For each covariate pattern, j, save m\_j= # with covariate pattern j and;

\* y\_j = # with outcome=1 in covariate pattern j ;

**proc** **means** n sum noprint data=ICU\_94; \*\*\*\* for a different data set change independent variable names in by statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\* and outcome variable name in var statement \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

by SER PO2 age CAN;

var STA; output out=jdat n=m\_j sum=y\_j; \*<--- indicate outcome;

**run**;

\*LOOK ABOVE! after "var" is where your outcome should be;

\* Run proc logistic for covariate patterns rather than individuals ;

\* outcome=y\_j / m\_j (not 0 or 1) , save fitted values ;

**proc** **logistic** noprint descending data=jdat; \*\*\*\* for a different data set change independent variable names in model statement. keep jdat as jdat. do NOT change this\*\*\*\*\*;

model y\_j/m\_j= SER PO2 age CAN;

output out=pdat p=p\_j;

**run**;

\* Create v\_j, c\_j, the chi-square terms and the terms in the sum in A;

**data** pdat; set pdat;

v\_j=m\_j\*p\_j\*(**1**-p\_j); c\_j=(**1**-**2**\*p\_j)/v\_j; chisq\_j=(y\_j-m\_j\*p\_j)\*\***2**/v\_j;

m\_j\_inv=**1**/m\_j;

**run**;

\* Create and save chi-square & sum for A ;

\* Perform weighted linear regression, save SS ;

\* Calculate RSS,A,z & p-val for z ;

**proc** **means** sum noprint data=pdat;

var chisq\_j m\_j\_inv; output out=cdat sum=chisq m\_inv; **run**;

**proc** **reg** noprint data=pdat outest=ss; \*\*\*\* for a different data set change independent variable names in model statement \*\*\*\*\*;

model c\_j=SER PO2 age CAN;

weight v\_j; **run**;

**data** zdat; merge cdat (keep=\_freq\_ chisq m\_inv) ss (keep=\_rmse\_);

rss=(\_freq\_-**4**-**1**)\*\_rmse\_\*\***2**; A=**2**\*(\_freq\_-m\_inv); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

z=(chisq-(\_freq\_-**4**-**1**))/sqrt(A+rss); z=abs(z); \*\*\*\* for a different data set change 8 to number of variables in the model \*\*\*\*\*;

pval=(**1**-probnorm(z))\***2**;

**run**;

**proc** **print** noobs data=zdat; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*42\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_42; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*51\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_51; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*84\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_84; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*93\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_93; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*94\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\* Stukel test of logistic regression model assumption \*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

ODS trace on;

ODS output GlobalTests=gt1;

**proc** **logistic** descending data=ICU\_94; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement \*\*\*\*\*;

model STA=SER PO2 age CAN;

output out=pdat2 xbeta=g\_j p=p\_j;

**run**;

**data** pdat2;

set pdat2;

if p\_j>=**0.5** then ind1=**1**; else ind1=**0**;

if p\_j< **0.5** then ind2=**1**; else ind2=**0**;

z1\_j=**0.5**\*g\_j\*\***2**\*ind1;

z2\_j=-**0.5**\*g\_j\*\***2**\*ind2;

**run**;

ODS output GlobalTests=gt2;

**proc** **logistic** descending data=pdat2 ; \*\*\*\* for a different data set change outcome and independent variable names \*\*\*\*\*;

\*\*\*\* in model statement but keep z1\_j and z2\_j \*\*\*\*\*;

model STA=SER PO2 age CAN z1\_j z2\_j;

**run**;

**data** pval;

merge gt1(rename=(ChiSq=ChiSq1))

gt2(rename=(ChiSq=ChiSq2));

if \_N\_=**1**;

drop Test df ProbChisq;

lr=ChiSq2-ChiSq1;

pval=(**1**-probchi(lr,**2**));

**run**;

**proc** **print** noobs data=pval; var pval; **run**;