Neil Yetz

ERHS 642: Logistic regression

02/04/2016

Dr. Bachand

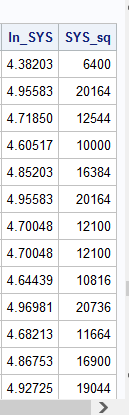
ERHS 642 Logistic Regression Spring 2016

Homework Assignment 1 – 5 points

1. In a temporary ICU data set,

a. Transform the SYS variable using the quadratic and the natural log transformation

Table 1.1: Log transformed (left) and Quadratic transformation of the SYS Variable with the ICU dataset.



b. For the untransformed SYS variable and for the log-transformed SYS variable, obtain

the

• mean

• standard deviation

• minimum, maximum and quartiles

• 5 lowest and 5 highest values

Table 1.2: Univariate procedure for SYS Variable from ICU data set

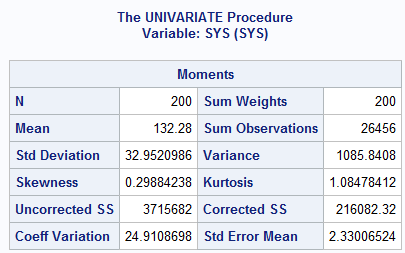


Table 1.3: Minimum, Maximum and quartiles of SYS from ICU data set

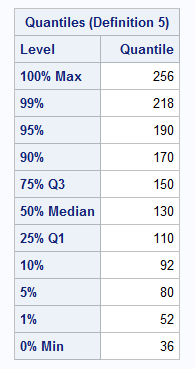


Table 1.4: Extreme observations of SYS from ICU data set

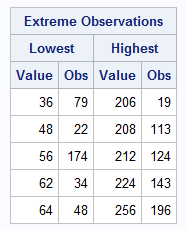


Table 1.5: Univariate function of Log Transformed SYS variable.

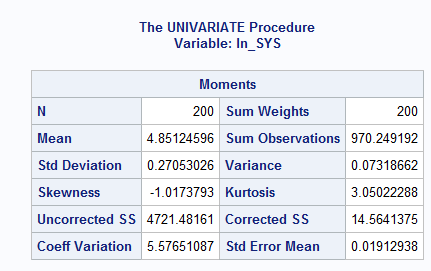


Table 1.6: Minimum, Maximum and quartiles of log transformed SYS variable from ICU data set.

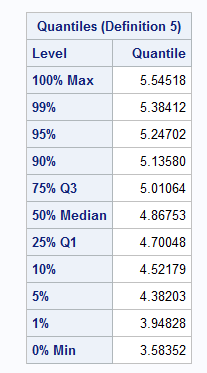
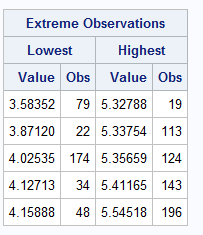


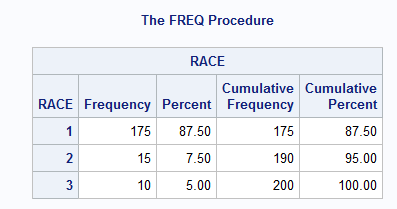
Table 1.7: Extreme observations of log transformed SYS variable form ICU data set



2. Use the temporary ICU data set and

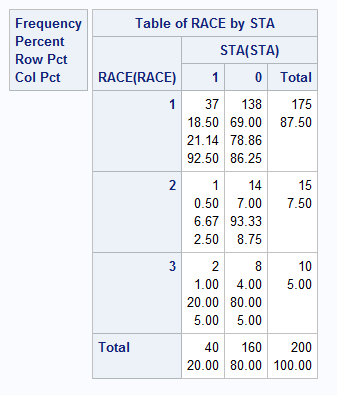
a. Create a frequency table for the categorical variable RACE

Table 2.2: Frequency distribution for RACE from the ICU data set



b. Cross-tabulate STA and RACE

Table 2.2: Cross tabulation of STA and RACE from the ICU data set



3. HL chapter 1,page 32/33, exercise 1

(a)

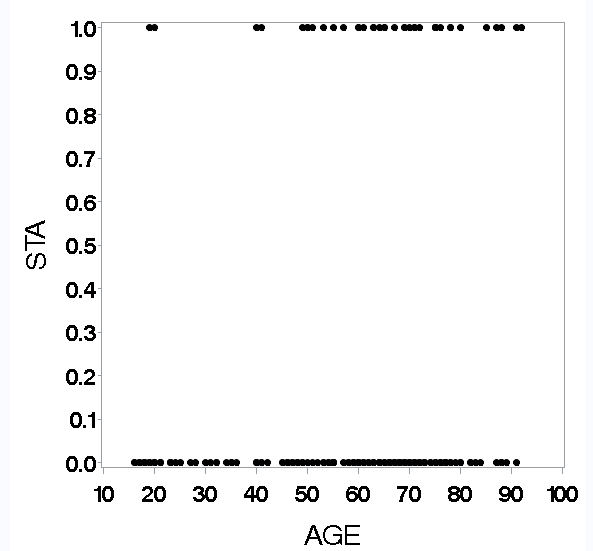
Logistic regression Model: π(AGE) =

Logit: g(age) =

STA is dichotomous

(b)

Table 2.3: Scatter plot of Age by STA.



(c)

Table 2.4: The mean procedure of the STA variable set to the mid points of each age variable.



Table 2.5: Scatterplot with age midpoint probabilities.

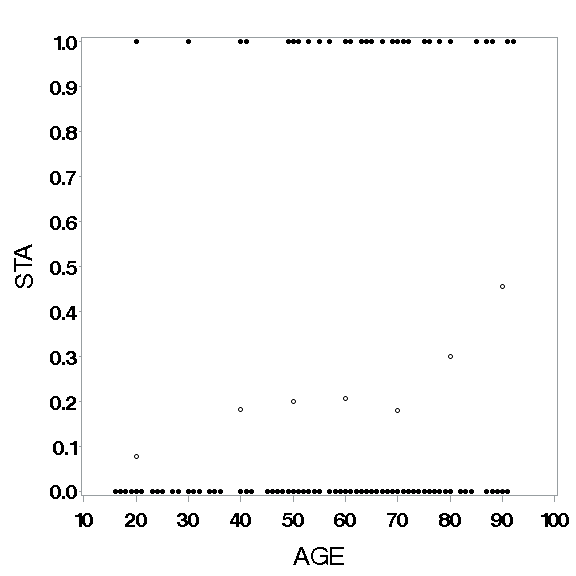
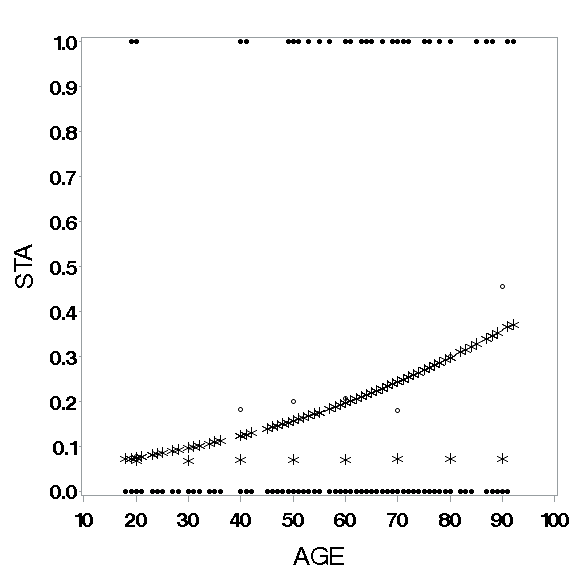


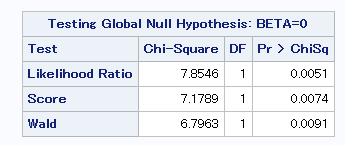
Table 2.6: Scatter plot with age mid point probabilities and pihats

(e) 

π(AGE) =

(f)

Table 3.2: Results from significance tests for age



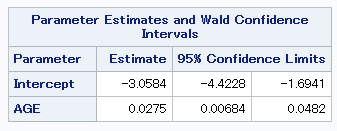
All Significant. Reject Null. Age has a significant effect on Status.

Assumptions = assuming we have a large sample size and enough in both the 0 and 1 categories.

We meet the assumption of enough subjects. We do NOT have enough subjects in each category though (1=40, 0=160). Might have not overly selected. Could be fixed by stratification.

**You don’t need to answer the question about the deviance.**

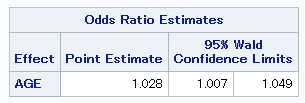
(g) Table 3.3: Wald Confidence intervals for Beta-zero and Beta 1 of STA by AGE logistic regression model.



B1 = .0275(x) 95%CI= [.00684, .0482]

Essentially, this confidence interval tells us that change in probability of having STA from 1 age interval to the next could vary from .00684 and as high as .0482 with 95% confidence.

Table 3.4: Wald Confidence intervals for odds ratio of STA by AGE logistic regression model.



Essentially, this confidence interval tells us that The Odds ratio for Having STA given age is 1.028 but can vary between 1.007 and 1.049. We have 95% confidence that the true parameter is thin this confidence interval. In other words, if we rant this test 100 times this would be one of our confidence intervals, and 95 out of 100 confidence intervals contain the true parameter.

Note: Don’t do parts (d) or (h)

Note: The ICU data set is described in HL Chapter 1

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

**data** ICU; set sdat.ICU;

**data** ICU; set sdat.ICU;

if **15**<=age<=**24** then a=**1**;

else if **25**<= age<=**34** then a=**2**;

else if **35**<= age<=**44** then a=**3**;

else if **45**<= age<=**54** then a=**4**;

else if **55**<= age<=**64** then a=**5**;

else if **65**<= age<=**74** then a=**6**;

else if **75**<= age<=**84** then a=**7**;

else if **85**<= age<=**94** then a=**8**;

\* Create the natural log of SYS \*;

ln\_SYS = log(SYS);

\* Create SYS squared \*;

SYS\_sq=SYS\*\***2**;

**run**;

**proc** **print** data = ICU;

var ln\_SYS SYS\_sq;

**run**;

**proc** **univariate** data=ICU; var SYS; **run**;

**proc** **univariate** data=ICU; var ln\_SYS; **run**;

**proc** **freq** data=ICU order=data;

tables RACE RACE\*STA;

**run**;

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

model STA=age;

output out=pdat p=pihat;

**run**;

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

var AGE STA pihat;

**run**;

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

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

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=ICU;

plot STA\*AGE/haxis=axis1 vaxis=axis2;

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

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

model STA=AGE;

output out=pdat p=pihat;

**run**;

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

var AGE STA pihat;

**run**;

**proc** **sort** data=pdat;

by age;

**run**;

**proc** **means** noprint data=icu;

by a;

var sta;

output out=sta\_means mean=stamean;

**run**;

**proc** **print** data=sta\_means;

**run**;

**data** sta\_means;

set sta\_means;

\*now assigning actual ages that equal the midpoint of the categories to the new data set;

if a=**1** then age=**20**;

else if a=**2** then age=**30**;

else if a=**3** then age=**40**;

else if a=**4** then age=**50**;

else if a=**5** then age=**60**;

else if a=**6** then age=**70**;

else if a=**7** then age=**80**;

else if a=**8** then age=**90**;

drop \_type\_ \_freq\_ a;

**run**;

**proc** **print** data=sta\_means;

**run**;

**proc** **gplot** data=sta\_means;

plot (stamean)\*age

run;

**quit**;

**proc** **logistic** descending data=icu;

model sta=age;

\*sta=response or independent var, age=dependent var;

output out=lr\_data p=pihat;

**run**;

**proc** **print** data=lr\_data; **run**;

**proc** **print** data=sta\_means; **run**;

**data** plotdat;

merge pdat sta\_means;

\*merge 2 new datasets;

**run**;

**proc** **print** data= plotdat;

**run**;

**proc** **gplot** data=plotdat;

plot (sta stamean)\*age /overlay haxis=axis1 vaxis=axis2;

**run**;

**quit**;

**proc** **gplot** data=plotdat;

\*plot variables from the new merged dataset, along with sta;

plot (sta stamean pihat)\*age /overlay haxis=axis1 vaxis=axis2;

**run**;

**quit**;

**proc** **logistic** descending data=icu;

model sta=age/cl;

\*sta=response or independent var, age=dependent var;

output out=lr\_data p=pihat;

**run**;

**proc** **logistic** descending data=icu;

model sta=age/clodds=both;

\*model confidence intervals;

**run**;