**Neil Yetz**

**04/12/2016**

**ERHS 642: Aplied Logistic regression**

**ERHS 642 Logistic Regression Spring 2016**

**Homework Assignment 7 – New Version**

Using the ICU\_altered data set with STA as the outcome variable,

1. Perform best subsets selection of the main effects model
   * You can use collapsed variables and scale assessment results from HW 6
   * If the main effects model is unstable (huge standard errors or 95% confidence intervals for some variables) or if it contains statistically non-significant variables, consider removing those variables from the model before proceeding to question 2.

* Viewing these models, first off I will say that they include variables such as CPR and CRN, which I purposely left out due to wide confidence intervals due to sparse cells.
* Therefore, I am including the best model not including variables without wide confidence intervals included.

Table 1.01: Step procedure best 3 main effect model

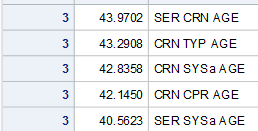


Table 1.02: Step procedure best 4 main effect models

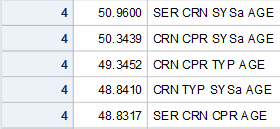


Table 1.03: Step procedure best 5 main effect models

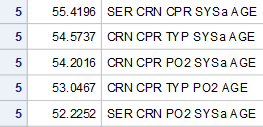


Table 1.04: Step procedure best 6 main effect models

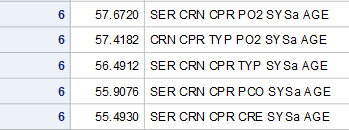


Table 1.05: Step procedure best 7 main effect models

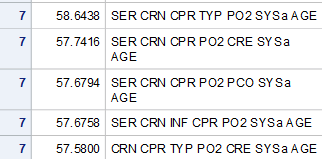


Table 1.06: Step procedure best 8 main effect models

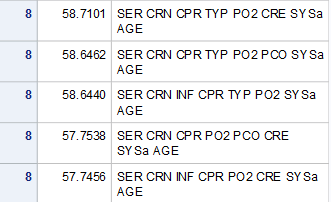


Table 1.1: Maximum Likelihood estimates for best 4 main effect model (minus CRN due to wide confidence intervals).

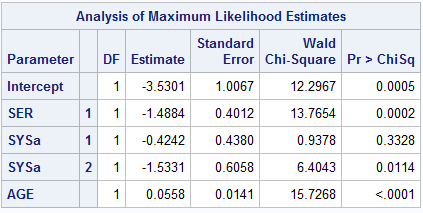
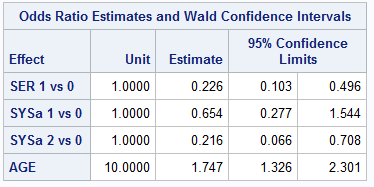


Table 1.2: Odds ratio estimates for best 4 main effect model (minus CRN due to wide confidence intervals).



1. Compare your main effects model from HW 6 to the model obtained in question 1; explain any differences. Did you miss any important variables in HW 6?

Table 2.1: Maximum likelihood estimates of final model from HW6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

Table 2.2: Odds Ratios of final model from HW6.

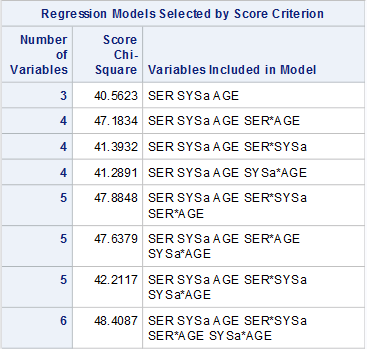
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Comparison/Unit | OR | 95%CI | | P-Value |
| Service at ICU admission: | Surgical vs Medical | 0.178 | 0.077 | 0.411 | <0.0001 |
| PO2 from initial blood gases | <=60 vs >60 | 1.742 | 0.617 | 4.922 | 0.2945 |
| AGE | 10 | 1.825 |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

* Comparing the automated stepwise table as compared to my final model from HW6, it is clear to see that there are differences in our models. Honestly, looking back, I do NOT think I should have included Cancer as part of the present problem variable into my model, but, regardless there were deinfitely some difference when looking at the best subsets analyses.
* The subsets analyses suggested including my categorized SYS variable. Which I did not include due to the insignificance of one of the levels of the variable.
* Otherwise, it appears that My model has more differences than the suggested models. Systolic blood pressure seems to be the only difference in the variable.

3.

* 1. Perform best subsets selection of the interaction terms added to the main effects model
  2. If the final model is unstable or if it contains statistically non-significant variables, consider removing those variables from the model before proceeding to question 4.

Table 3.1 Results from subset analysis including interactions.



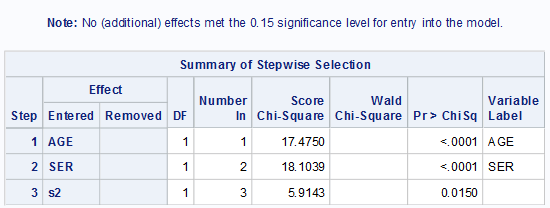
1. Compare your final model from HW 6 to the model obtained in question 3b; explain any differences.

* My final model from Homework 6 did not contain any interactions. The main reason for the difference in the best subsets selection is that is forcing for an interaction to be in the model because of the coding to include certain variables. Therefore, I consciously made the decision not to include any itercations based n the p-values whereas in the the subset analyses it was forced to include the interactions, regardless of significance. It clearly just chose the most significant (not necessarily significant) interactions to include in the model.

5.

1. Perform stepwise selection of the main effects model
   * You can use collapsed variables and scale assessment results from HW 6
   * If the main effects model is unstable (huge standard errors or 95% confidence intervals for some variables) or if it contains statistically non-significant variables, consider removing those variables from the model before proceeding to question 6.

Table 5.1: Best main effects model with stepwise selection.



1. What is the problem with removing variables from the stepwise selected main effects model? How could you resolve the problem?

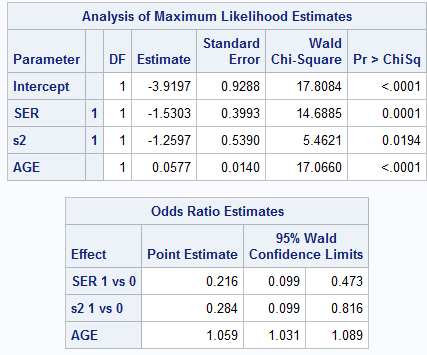
6. Compare your main effects model from HW 6 to the model obtained in question 5; explain any differences.

* For one, the main effects model proposes much less variables than my final model. Compare 4 variables to 3 in the stepwise selection model.
* Furthermore, the stepwise model suggests leaving only one portion of the dummy coded SYS variable. Therefore, some extra though needs to be put into this model because it does not make sense for only one part of the variable to be included.

7.

1. Perform stepwise selection of the interaction terms added to the main effects model

Table 7.1: Maximum likelihood estimates and Odds ratios for variables after the addition of interaction items. Note: No interaction terms made it into the model.



1. If the final model is unstable or if it contains statistically non-significant variables, consider removing those variables from the model before proceeding to question 8.

8. Compare your final model from HW 6 to the model obtained in question 7b; explain any differences.

Comparing the final model from my stepwise selection, it is clear that the SER and AGE are definitely significant predictors. Where my model differs is in the addition of PO2 and CAN, Which I included CAN because I believed it would be a big part of a predictor of life or death. Furthermore, the stepwise selection model only includes one part of the SYS dummy coded variable, which does not necessarily make sense. Therefore, it is quite difference from what I would have chosen in part of the final model.

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**;

if SYSa=**0** then do; s1=**0**; s2=**0**; end;

else if SYSa=**1** then do; s1=**1**; s2=**0**; end;

else if SYSa=**2** then do; s1=**0**; s2=**1**; end;

**run**;

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

model STA= SER CRN INF CPR TYP PO2 PCO CRE SYSa age

/ selection=score start=**3** stop=**8** best=**5**;

**run**;

\*best 3 main effects model;

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

class SER CRN /param=ref ref=first;

model STA=SER CRN AGE/clodds=wald;

units age=**10**;

**run**;

\*best 4 vmain effects model;

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

class SER CRN SYSa/param=ref ref=first;

model STA=SER CRN SYSa AGE/clodds=wald;

units age=**10**;

**run**;

\*Best 4 main effects model without CRN;

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

class SER SYSa/param=ref ref=first;

model STA=SER SYSa AGE/clodds=wald;

units age=**10**;

**run**;

\*best 5 main effects model;

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

class SER CRN CPR /param=ref ref=first;

model STA=SER CRN CPR AGE/clodds=wald;

units age=**10**;

**run**;

\*best 6 main effects model;

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

class SER CRN CPR PO2/param=ref ref=first;

model STA=SER CRN CPR PO2 AGE/clodds=wald;

units age=**10**;

**run**;

\*best 7 main effects model;

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

class SER CRN CPR PO2 SYSa TYP/param=ref ref=first;

model STA=SER CRN CPR PO2 TYP SYSa AGE/clodds=wald;

units age=**10**;

**run**;

\*best 8 main effects model;

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

class SER CRN CPR PO2 SYSa CRE TYP/param=ref ref=first;

model STA=SER CRN CPR PO2 TYP CRE SYSa AGE/clodds=wald;

units age=**10**;

**run**;

\*Best subset model with interaction terms;

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

model STA= SER SYSa age

SER\*SYSa SER\*AGE SYSa\*age

/ selection=score start=**4** stop=**6** best=**4** include=**3**;

**run**;

\*Best 4 main effect model w/ interactins

proc logistic descending data=ICU\_altered;

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

class SER SYSa /param=ref ref=first;

model STA=SER SYSa AGE SER\*AGE/clodds=wald;

units age=**10**;

**run**;

\*Stepwise selection;

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

class SER CRN INF CPR TYP PO2 PCO CRE s1 s2/param=ref ref=first;

model STA= SER CRN INF CPR TYP PO2 PCO CRE s1 s2 age

/stepwise sle=**0.15** sls=**0.20** details;

**run**;

\*Reduced stepwise selection;

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

class SER INF TYP PO2 PCO s1 s2/param=ref ref=first;

model STA= SER INF TYP PO2 PCO s1 s2 age

/stepwise sle=**0.15** sls=**0.20** details;

**run**;

\*Stepwise with interactions;

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

class SER s2/param=ref ref=first;

model STA= SER s2 age

SER\*s2 SER\*age s2\*AGE

/stepwise sle=**0.15** sls=**0.20** include=**3** details;

**run**;