Neil Yetz

05/03/2016

ERHS 642: Applied Logistic Regression

Dr. Bachand

**ERHS 642 Logistic Regression Spring 2016**

**Homework Assignment 10**

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 |

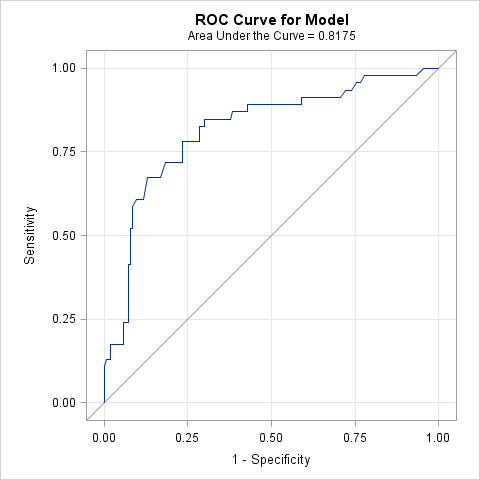
1. Present a table containing odds ratios, 95% confidence intervals and p-values for your final model from homework assignment 6.
2. Table 1.2: Odds Ratios, 95% confidence intervals, and p-values for final Model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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. Determine how well your model predicts the outcome,

1. Plot the ROC curve, determine the area under the ROC curve and draw conclusions

Table 2.1: ROC curve for the for the main effects model.

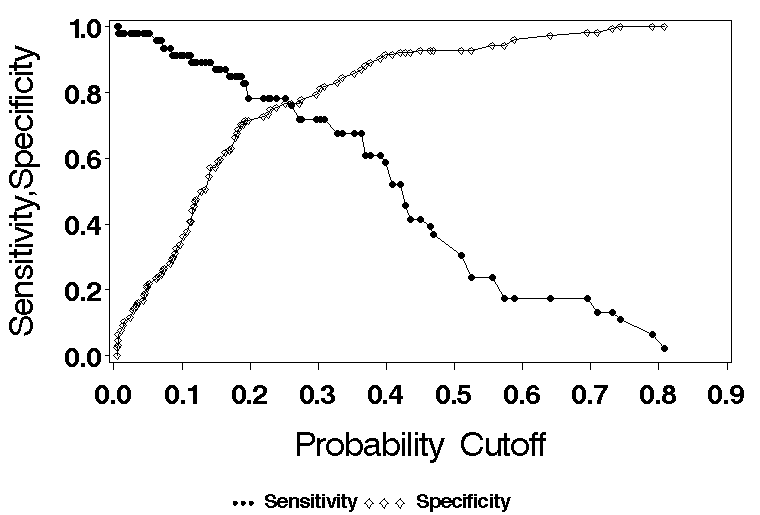


Area under the ROC curve = 0.8175

Our area under the ROC curve indicates that there is excellent discrimination in our model. Furthermore, this means that our model is doing an excellent job in correctly discriminating between those that were actually predicted, and not predicted, to have the outcome of death.

1. Plot sensitivity and specificity vs. possible cutpoints and select the “best” cutpoint

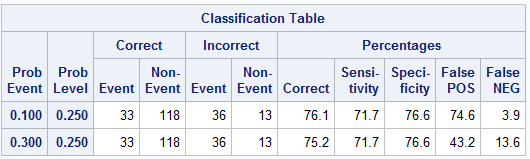
Figure 2.2: Graph representing sensitivity & specificity and corresponding pihats with the best indicated cut point drawn in.



Best cutpoint = **0.25**

1. Based on this cutpoint, calculate sensitivity and specificity

Table 2.1: Classification table of model with 0.25 cutpoint and 0.1 & 0.3 event probability



Sensitivity at cutpoint 0.25 = **71.7**

Specificity at cutpoint 0.25 = **76.6**

1. Draw conclusions based on your results in part c. Include definitions of sensitivity and specificity and specifically address limitations of sensitivity and specificity in assessing model fit; how may the limitations have affected sensitivity and specificity in your model (show evidence).

**Sensitivity:**

* Sensitivity is the proportion of individuals that have an outcome were correctly identified as having the outcome.
* In my model, the sensitivity is indicating that we correctly predicted 71.7% of the individuals to have the outcome of death, based on the predictor variables we had in our model.

**Specificity:**

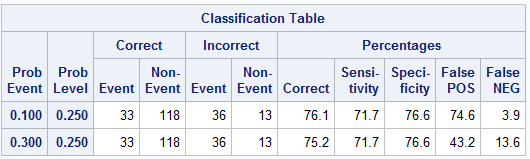
* Specificity is the proportion of individuals that were correctly identified as *NOT* having the outcome.
* In my model, the specificity is indicating that we correctly predicted 76.6% of the individuals to *NOT* have the outcome of death, based on the predictor variables we had in our model.

**Limitations:**

* One extreme limitation of using sensitivity and specificity in logistic regression is that it treats observations that are nothing alike completely different as similar and observations that are very similar very differently based on the cutpoints designated.
  + For example, an observation with a pihat of 0.26 will be treated exactly the same as an individual with a pihat as 0.67 based on our cutpoint (0.25), even though obviously have very different characteristics. Furthermore, observations with pihats of 0.26 are treated completely different than observations with pihat of 0.24.
* Based on Figure 2, it is clear that these limitations have affected our specificity and sensitivity. As stated in the limitations, there are clearly observations that fall very close to the cutpoint and as well far away. Thus causing both over-discrimination and under discrimination.

1. Calculate the positive and negative predictive value assuming the prevalence of the outcome in the population of interest is 30% and assuming the prevalence of the outcome in the population of interest is 10%.

Table 2.2: Classification table of model with 0.25 cutpoint and 0.1 & 0.3 event probability



**Positive Predictive Value:**

* **0.10 Prevalence:** 100%-FalsePOS 🡪 100%-74.6% = **23.4%**
* **0.30 Prevalence:** 100%-FalsePOS 🡪 100%-43.2% = **56.8%**

**Negative Predictive Value:**

* **0.10 Prevalence:** 100%-FalseNEG 🡪 100%-3.90% = **96.1%**
* **0.30 Prevalence:** 100%-FalseNEG 🡪 100%-13.6% = **86.4%**

1. Draw conclusions based on your results in part e. Include definitions of positive and negative predictive values.

* **Positive Predicted Value (PPV):** The probability that those indicated with having the outcome truly have the outcome.
  + **0.10 Prevalence:** As indicated from the above equation, the results indicate that our PPV is 23.4% at 0.10 prevalence. Therefore, based on the data and the population prevalence, 23.4% of people with the indicated outcomes would correctly be predicted to have the outcome of death.
  + **0.30 Prevalence:** As indicated from the above equation, the results indicate that our PPV is 56.8% at 0.30 prevalence. Therefore, based on the data and the population prevalence, 56.8% of people with the indicated outcomes would correctly be predicted to have the outcome of death.
* **Negative Predictive Value (NPV):** The probability that those indicated with having the *NOT* having the outcome truly do *NOT* have the outcome.
  + **0.10 Prevalence:** As indicated from the above equation, the results indicate that our NPV is 96.1% at 0.10 prevalence. Therefore, based on the data and the population prevalence, 96.1% of people without the indicated outcomes would correctly be predicted to NOT have the outcome of death.
  + **0.30 Prevalence:** As indicated from the above equation, the results indicate that our NPV is 86.4% at 0.30 prevalence. Therefore, based on the data and the population prevalence, 86.4% of people without the indicated outcomes would correctly be predicted to NOT have the outcome of death.
* Based on these results, our model does a great job at predicting those that do not have the outcome of death. Unfortunately, it is not as good at predicting who has the outcome of death, therefore it is not that good of a predictive model when basing it of PPV.

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**;

\*Specificity/Sentiivity graph;

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

model STA=SER PO2 age CAN / outroc=rocdat;

**run**;

**proc** **print** data=rocdat;

**run**;

\*Calculate specificity;

**data** rocdat;

set rocdat;

spec=**1**-\_1mspec\_;

**run**;

\*Plotting\*;

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

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

goptions FTEXT=swissb HTEXT=**2.0** HSIZE=**8** in VSIZE=**6** in;

symbol1 v=dot i=join c=black h=**1**;

symbol2 v=diamond i=join c=black h=**1**;

footnote1 c=black f=special h=**1** 'J J J' f=swissb h=**1.5** ' Sensitivity'

c=black f=special h=**1** ' D D D' f=swissb h=**1.5** ' Specificity';

**proc** **gplot** data=rocdat;

plot (\_sensit\_ spec)\*\_prob\_

/overlay haxis=axis1 vaxis=axis2;

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

footnote;

\*Creating a classification table\*;

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

model STA=SER PO2 age CAN

/ctable pprob=(**0.25**) pevent=**0.10** **0.30**;

**run**;