**N736 Homework 06 – Answer Key**

Logistic Regression

For Homework 06, you will be using the HELP dataset, learn more at:

* <https://melindahiggins2000.github.io/N736Fall2017_HELPdataset/> &
* <https://github.com/melindahiggins2000/N736Fall2017_HELPdataset>

Refer to the logistic regression analysis example and codes we ran during lesson 18 and 19 - see <https://github.com/melindahiggins2000/N736Fall2017_lesson1819>

For the HELP dataset:

* OUTCOME VARIABLE: consider the variable g1b “Experienced serious thoughts of suicide (last 30 days) - Baseline”
* PREDICTOR VARIABLE: consider these variables as potential predictors for g1b:
  + age, female, pss\_fr, homeless, pcs, mcs, cesd, indtot

Complete the following:

1. Consider the continuous variable cesd as a predictor for g1b
   1. run a logistic regression of the probability of suicidal thoughts (g1b) given their depressive symptoms scores (cesd)
   2. make a plot of the the predicted probability of suicidal thoughts (g1b) by the depressive symptoms scores (cesd)
   3. what value of the cesd leads to a probability of suicidal thoughts => 0.5? *(hint: use the plot you just made)*
2. Using variable selection methods, develop a logistic regression model for the probability of suicidal thoughts (g1b) considering all of these variables for possible inclusion: age, female, pss\_fr, homeless, pcs, mcs, cesd, indtot
   1. present the final model results
   2. write a few sentences describing your results including:
      1. model fit
      2. model classification table results - remember to report the threshold used for the classification table - you can change it from 0.5 if you think a different threshold might work better
      3. odds ratios for each significant predictor in the model
3. **Logistic Regression of Suicidal Thoughts (g1b) by Depressive Symptoms (cesd)**

A.

**Model Fit Summary**

Overall the model fit was significant (χ2(1)=56.140, p<.001); -2 Log likelihood (-2LL)=481.38; Cox & Snell R2=0.117; Nagelkerke R2=0.168.

I did not discuss this in class, but the significant p-value for the “Hosmer and Lemeshow Test” (χ2(8)=17.075; p=.029) indicates that this model has a poor fit.

It is worth noting, that the null model (with only the intercept in the model), the classification rate was 72% (127/453) for predicting suicidal thoughts.

**Classification**

For the model with CESD, the classification table is shown below (using the default threshold of 0.5) with an overall classification rate of 71.3% (91.7% correct prediction of subjects NOT having suicidal thoughts but only 18.9% correct prediction of subjects with YES suicidal thoughts). Also, FYI changing the “cut value” threshold to 0.6, increases the overall classification rate to 73.1%, which is only slightly better than the null model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea** | | | | | |
|  | Observed | | Predicted | | |
|  | g1b Experienced serious thoughts  of suicide (last 30 days) - Baseline | | Percentage  Correct |
|  | 0 no | 1 yes |
| Step 1 | g1b Experienced serious thoughts  of suicide (last 30 days) - Baseline | 0 no | 299 | 27 | 91.7 |
| 1 yes | 103 | 24 | 18.9 |
| Overall Percentage | |  |  | 71.3 |
| a. The cut value is .500 | | | | | |

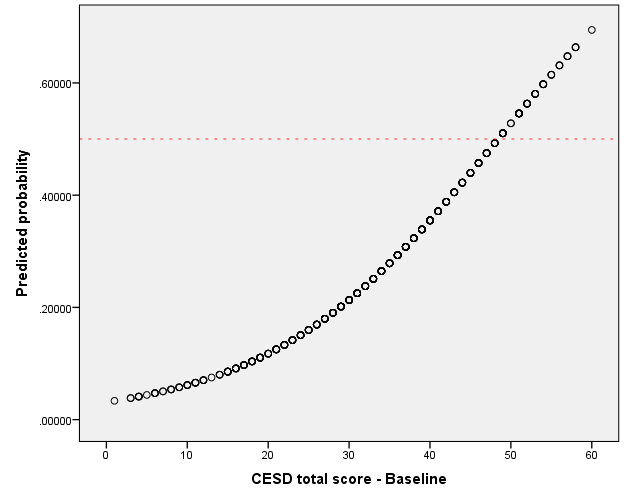
**Logistic Regression Equation**

Depressive symptoms were significant predictive of suicidal thoughts (p<.001); for every one point higher someone scored on the CESD, their odds of having suicidal thoughts increased by OR=1.073 times (95% Confidence Interval 1.052, 1.096).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | | | |
|  | | B | S.E. | Wald | df | p-value | Exp(B)  Odds Ratio (OR) | 95% C.I.for EXP(B) | |
| Lower | Upper |
| Step 1a | cesd | .071 | .010 | 45.723 | 1 | .000 | 1.073 | 1.052 | 1.096 |
| Constant | -3.433 | .404 | 72.097 | 1 | .000 | .032 |  |  |
| a. Variable(s) entered on step 1: cesd. | | | | | | | | | |

B.

**Plot of predicted probability (from logistic regression model) of suicidal thoughts by CESD scores**



C.

CESD scores of about 48-49 or higher lead to a probability of suicidal thoughts > 0.5.

**2. Fit a model for suicidal thoughts (g1b) using age, female, pss\_fr, homeless, pcs, mcs, cesd, indtot**

Depending on which software package you used and whether or not you used forward or backward variable selection, your set of variables retained in the final model may be different.

**Model Fit**

Overall the model fit was significant (χ2(4)=75.408, p<.001); -2 Log likelihood (-2LL)=462.110; Cox & Snell R2=0.153; Nagelkerke R2=0.221.

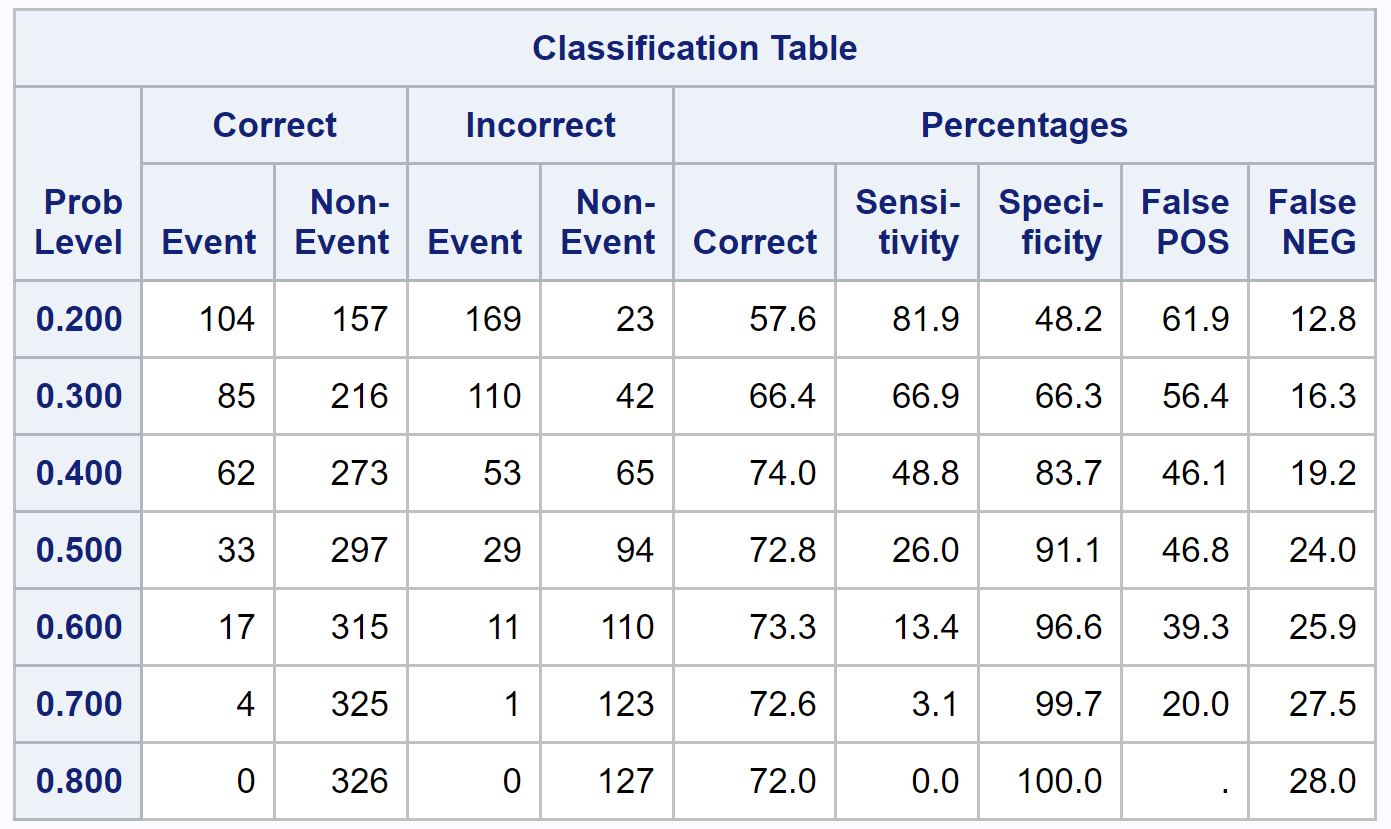
**Classification**

The final model (using a cut value or threshold of 0.5) yielded an overall classification rate of 73.1% (with 91.4% correct classification of subjects with NO suicidal thoughts, but only 26.0% correct classification of subjects with YES suicidal thoughts). You may have chosen a different threshold value.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea** | | | | | |
|  | Observed | | Predicted | | |
|  | g1b Experienced serious  thoughts of suicide  (last 30 days) - Baseline | | Percentage  Correct |
|  | 0 no | 1 yes |
| Step 4 | g1b Experienced serious  thoughts of suicide (last  30 days) - Baseline | 0 no | 298 | 28 | 91.4 |
| 1 yes | 94 | 33 | 26.0 |
| Overall Percentage | |  |  | 73.1 |
| a. The cut value is .500 | | | | | |

**Classification Cutoff/Threshold trade off table**

Best trade off looks to be around the probability level of 0.3.



**Logistic Regression Equation**

In the final fitted model, gender (female), being homeless, the mental component score of the SF36, and depressive symptoms (CESD) were retained. Being female increased the odds of having suicidal thoughts by 1.78 times and being homeless increased the odds by 1.85 times. For every 1 point higher someone scored on the CESD the odds of having suicidal thoughts increased by 1.043 times, and for every 1 point higher scored on the mental component score the odds of having suicidal thoughts was lower (OR=0.937) – thus the odds of NOT having suicidal thoughts was higher by 1/0.937 = 1.07 times (i.e. take the inverse of the odds ratio for mcs).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| Lower | Upper |
| Step 4d | female | .578 | .260 | 4.969 | 1 | .026 | 1.783 | 1.072 | 2.966 |
| homeless | .613 | .232 | 6.990 | 1 | .008 | 1.846 | 1.172 | 2.909 |
| mcs | -.039 | .013 | 8.856 | 1 | .003 | .962 | .937 | .987 |
| cesd | .042 | .013 | 10.391 | 1 | .001 | 1.043 | 1.017 | 1.070 |
| Constant | -1.749 | .757 | 5.332 | 1 | .021 | .174 |  |  |
| a. Variable(s) entered on step 1: cesd. | | | | | | | | | |
| b. Variable(s) entered on step 2: mcs. | | | | | | | | | |
| c. Variable(s) entered on step 3: homeless. | | | | | | | | | |
| d. Variable(s) entered on step 4: female. | | | | | | | | | |