Homework Solutions
Applied Logistic Regression

WEEK 8

Exercise 1:

Consider a multivariable model for the ICU study with the variables: **LOC2 TYP AGE CAN SYS** and **AGExSYS**. Assess the fit of the model.

To include the interaction term in the model, we use the command, "gen agesys = AGE*SYS". Type either "**logistic STA LOC2 TYP AGE CAN SYS agesys**" or "**logit STA LOC2 TYP AGE CAN SYS agesys**" to obtain a logistic regression output.

```
. gen agesys=AGE*SYS

. logistic STA LOC2 TYP AGE CAN SYS agesys
Logit Estimates                                   Number of obs =    200
                                                  chi2(6)       =  69.32
                                                  Prob > chi2   = 0.0000
Log Likelihood = -65.420861                       Pseudo R2     = 0.3463

------------------------------------------------------------------------------
         STA | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LOC2 |   45.39315   42.68579      4.057   0.000       7.187146    286.6977
         TYP |     18.886   17.34661      3.199   0.001       3.121185    114.2774
         AGE |   1.187805   .0787806      2.595   0.009       1.043013    1.352698
         CAN |   9.275526   7.801173      2.648   0.008        1.78417    48.22151
         sys |   1.055576   .0329379      1.733   0.083       .9929533    1.122148
      agesys |   .9989684    .000478     -2.157   0.031       .9980319    .9999056
------------------------------------------------------------------------------
```

Use the Hosmer-Lemeshow test to evaluate the overall fit of the model.

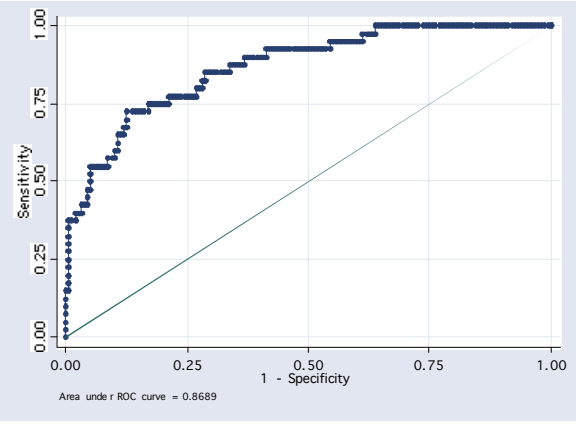
Type “lfit, group(10) table” in the command window to obtain both the goodness of fit result as well as the table of deciles.

```
. lfit, group (10) table

Logistic model for STA, goodness-of-fit test

  (Table collapsed on quantiles of estimated probabilities)
  +--------------------------------------------------------+
  | Group |    Prob | Obs_1 | Exp_1 | Obs_0 | Exp_0 | Total |
  |-------+---------+-------+-------+-------+-------+-------|
  |     1 |  0.0091 |     0 |   0.1 |    20 |  19.9 |    20 |
  |     2 |  0.0234 |     0 |   0.3 |    20 |  19.7 |    20 |
  |     3 |  0.0513 |     1 |   0.8 |    19 |  19.2 |    20 |
  |     4 |  0.0793 |     2 |   1.3 |    18 |  18.7 |    20 |
  |     5 |  0.1187 |     1 |   2.0 |    19 |  18.0 |    20 |
  |-------+---------+-------+-------+-------+-------+-------|
  |     6 |  0.1411 |     2 |   2.6 |    18 |  17.4 |    20 |
  |     7 |  0.2019 |     4 |   3.5 |    16 |  16.5 |    20 |
  |     8 |  0.2788 |     6 |   4.9 |    14 |  15.1 |    20 |
  |     9 |  0.5229 |     8 |   7.7 |    12 |  12.3 |    20 |
  |    10 |  0.9771 |    16 |  16.8 |     4 |   3.2 |    20 |
  +--------------------------------------------------------+

      number of observations =       200
            number of groups =        10
     Hosmer-Lemeshow chi2(8) =      2.38
                Prob > chi2 =      0.9669
```

The test indicates that the fit of the model is adequate.

Use the area under the ROC Curve to assess the model’s ability to discriminate between those subjects with the outcome versus those without the outcome.

```
. lroc

Logistic model for STAnumber of observations =      200
area under ROC curve   =   0.8689
```



The area under the ROC curve, 0.8689, indicates excellent discrimination.

Please note that the lfit and lroc command uses the latest model that was fit.