PH1700 Homework 10

1. This problem uses the ICU data from Hosmer and Lemeshow, Applied logistic regression, 2nd edition, described on page 24-25 "The ICU study data set consists of a sample of 200 subjects who were part of a much larger study on survival of patient following admission to an adult intensive care unit (ICU). The major goal of this study was to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients."

We wish to assess the relationship between Surviving the ICU and age, sex, and CPR prior to entry into the ICU. Therefore, we fit a logistic regression model using vital status (STA) 0 = lived and 1 = died; age in years, sex (0=Male, 1= Female) and CPR prior to ICU entry (CPR, 0=no, 1=yes) The full data can be found in icu.csv and its data dictionary in icu.txt.

Use the following stata output to answer the following questions.

```
. logit sta age sex cpr

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -92.630957
Iteration 2:   log likelihood = -91.884604
Iteration 3:   log likelihood = -91.880181
Iteration 4:   log likelihood = -91.880181

Logistic regression                               Number of obs   =        200
                                                  LR chi2(3)      =      16.40
                                                  Prob > chi2     =     0.0009
Log likelihood = -91.880181                       Pseudo R2       =     0.0819

------------------------------------------------------------------------------
         sta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0302564   .0113226     2.67   0.008     .0080645    .0524482
         sex |    -.16957    .388479    -0.44   0.662    -.9309749    .5918348
         cpr |   1.829403   .6184721     2.96   0.003     .6172204    3.041586
       _cons |  -3.330626   .7509232    -4.44   0.000    -4.802409   -1.858844
------------------------------------------------------------------------------

. logit sta age sex cpr,or

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -92.630957
Iteration 2:   log likelihood = -91.884604
Iteration 3:   log likelihood = -91.880181
Iteration 4:   log likelihood = -91.880181

Logistic regression                               Number of obs   =        200
                                                  LR chi2(3)      =      16.40
                                                  Prob > chi2     =     0.0009
Log likelihood = -91.880181                       Pseudo R2       =     0.0819

------------------------------------------------------------------------------
         sta | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.030719   .0116704     2.67   0.008     1.008097    1.053848
         sex |   .8440276    .327887    -0.44   0.662     .3941693    1.807301
         cpr |   6.230169   3.853186     2.96   0.003     1.853768    20.93844
------------------------------------------------------------------------------
```

2.a) Write the appropriate logistic model for this analysis

2.b) Is the model significant? Report the hypothesis, appropriate test statistics and p-value.

2.c) Is age a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. How would we interpret the coefficient for age? What could we have done to interpret the coefficient for age as an odds ratio?

2.d) Is sex a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. Interpret its odds ratio appropriately, based on the significance.

2.e) Is CPR a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. Interpret its odds ratio appropriately.

Answer 2a

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = 0.03 * Age - 0.17 * Sex + 1.83 * Cpr - 3.33$$

Answer 2b

Yes. The null hypothesis for this test is that removing any of the variables (Age, Sex and Cpr) has no effect. The chi-square statistic equals 16.40 and the p-value is 0.0009 < 0.05. Therefore we can reject the null hypothesis and conclude that the model is significant.

Answer 2c

Yes. The null hypothesis is that the coefficient of Age is 0. The test statistic z equals to 2.67 and the p-value is 0.008 <0.05. Therefore we can reject the null hypothesis and conclude that the coefficient is significantly different from 0. The odds ratio is 1.03. This indicates a one-year increase in Age, the odds of dying in ICU (versus surviving in ICU) increase by a factor of 1.03.

Answer 2d

No. The null hypothesis is that the coefficient of Sex is 0. The test statistic z equals to -0.44 and the p-value is 0.662 >0.05. Therefore we cannot reject the null hypothesis and conclude that the coefficient is not significantly different from 0. The odds ratio equals 0.844. But we cannot interpret the odds ratio since Sex may not be a significant factor.

Answer 2e

Yes. The null hypothesis is that the coefficient of Cpr is 0. The test statistic z equals to 2.96 and the p-value is 0.003 <0.05. Therefore we can reject the null hypothesis and conclude that the coefficient is significantly different from 0. The odds ratio is 6.23. This indicates that the chance of survival for patients who received CPR is 6.23 times greater than those who did not.

2. The following output is from an analysis of the Smoke.dat.dta data with an additional variable that splits the sample into two groups: high log co adjusted, and low log co adjusted, using a median cutpoint. The analysis compared recidivism of high versus low log CO adjusted scores. Looking at the output, what do you conclude about the recidivism of the two groups? [write 1-2

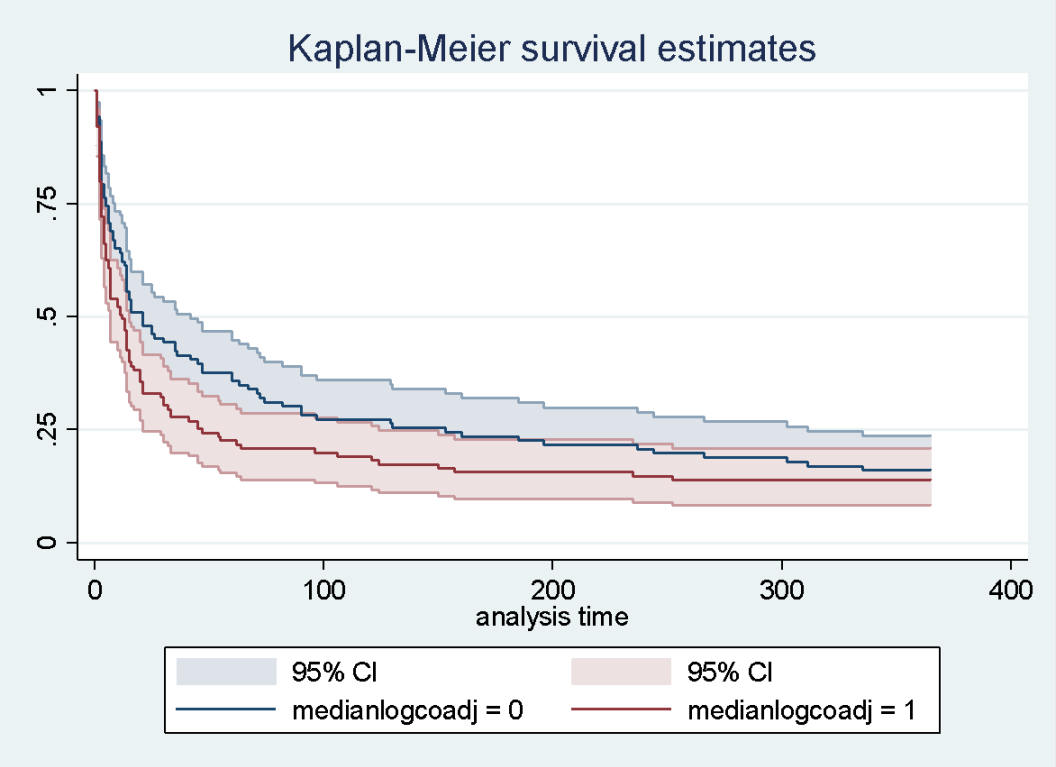
sentences interpreting the KM Plot. For the log rank test, state the null and alternative hypotheses, write the p-value and 1-2 sentences interpreting the log rank test.]

Stata output:

```
     failure event:  smoked == 1
obs. time interval:  (0, day_abs]
 exit on or before:  failure
```

```
        234  total observations
         13  observations end on or before enter()
```

```
        221  observations remaining, representing
        188  failures in single-record/single-failure data
     18,712  total analysis time at risk and under observation
                                              at risk from t =         0
                                   earliest observed entry t =         0
                                        last observed exit t =       365
```



```
        failure _d:  smoked == 1
  analysis time _t:  day_abs
```

Log-rank test for equality of survivor functions

medianlogc~j	Events observed	Events expected
0	89	100.46
1	99	87.54
Total	188	188.00

chi2(1) = 2.96
Pr>chi2 = 0.0854

Answer

The KM plot shows that the confidence band of medianlogcoadj=0 is above the confidence band of medianlogcoadj=1. But there are also overlapping area between two groups. It is difficult to determine from the plot whether there are differences between two curves.

The null hypothesis is that there is no association between recidivism and high versus low log CO adjusted scores. The test statistic equals 2.96 and p-values is 0.0854 > 0.05. Therefore we cannot reject the null hypothesis and conclude that there is no difference in recidivism between high and low log CO adjusted groups.