Analysis of Survival of Intensive Care Unit Patients

1. Introduction

The Intensive Care Unit (ICU) of a hospital only admits patients who are extremely ill or considered to be at great risk of serious complications requiring the special technology and highly skilled care available in the facility. Patients with care in ICU have a greater mortality rate than any other units of a hospital. A live discharge from ICU should be celebrated and can be regarded as a great gift to patients’ life. This study contains 200 patients sampled from 737 ICU patients. An objective model based on multivariate logistic regression is constructed to ascertain the risk factors attributed to the mortality from the discharge of ICU. We found that the level of consciousness (whether patients have coma or deep stupor), type of admission (elective or emergency), patients’ age and whether patient have cancer or not have a great impact for the mortality rate from the discharge of ICU. Our model fits the data quite well and has a good prediction power. This warrants the use of our model for the future application in estimating the probability of mortality for ICU patients with certain characteristics.

The report is organized as the following: Section 2 summarizes the included variables in the study and displays the results of univariate analysis for the mortality. Section 3 expands the analysis to multivariate logistic regression using stepwise model building procedure. Section 4 summarizes the findings and discusses the issues related to our analysis. The SAS codes that we used to conduct the analysis are included in the Appendix.

2. Explanatory Analysis

Lemeshow, Teres, Avrunin, and Pastides (1988) used 737 ICU patients to develop a predictive model for the outcome of ICU patients. 200 patients are randomly sampled from the original 737 observations. For each patient, 20 variables are recoded. Among the 20, there are 3 demographic variables (AGE, SEX and RACE), 16 descriptive variables (SER, CAN, CRN, INF, CPR, SYS, HRA, PRE, TYP, FRA, PO2, PH, PCO, BIC, CRE and LOC) to characterize the seriousness of patients’ physical condition, and one outcome to indicate the patient’s vital status (STA).

We present the summary of these variables by cross-tabulation with STA in Table 1. The results of univariate analyses using Chi-square test for two-way contingency table on categorical variables and two-sample t-test for continuous variables are also included in the table.
## Table 1: Summary of Descriptive Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>STA=1</th>
<th>STA=0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONTINUOUS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE*</td>
<td>no. (%)</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>40(20)</td>
<td>65.13</td>
</tr>
<tr>
<td></td>
<td>160(80)</td>
<td>55.65</td>
</tr>
<tr>
<td>SYS-systolic blood pressure**</td>
<td>no. (%)</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>40(20)</td>
<td>118.83</td>
</tr>
<tr>
<td></td>
<td>160(80)</td>
<td>135.64</td>
</tr>
<tr>
<td>HRA-heart rate</td>
<td>no. (%)</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>40(20)</td>
<td>100.63</td>
</tr>
<tr>
<td></td>
<td>160(80)</td>
<td>98.50</td>
</tr>
<tr>
<td><strong>CATEGORICAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>Male</td>
<td>24(19.35)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>16(21.05)</td>
</tr>
<tr>
<td>RACE (?)</td>
<td>White</td>
<td>37(21.14)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>1(6.67)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>2(20.00)</td>
</tr>
<tr>
<td>SER-service at ICU Admission*</td>
<td>Medical</td>
<td>26(27.96)</td>
</tr>
<tr>
<td></td>
<td>Surgical</td>
<td>14(13.08)</td>
</tr>
<tr>
<td>CAN-cancer present</td>
<td>No</td>
<td>36(20.00)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>4(20.00)</td>
</tr>
<tr>
<td>CRN-history of chronic renal failure**</td>
<td>No</td>
<td>32(17.68)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>8(42.11)</td>
</tr>
<tr>
<td>INF-infection Probable*</td>
<td>No</td>
<td>16(13.79)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>24(28.57)</td>
</tr>
<tr>
<td>CPR-cpr prior to ICU admission*</td>
<td>No</td>
<td>33(17.65)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>7(53.85)</td>
</tr>
<tr>
<td>PRE-previous admission to ICU</td>
<td>No</td>
<td>33(19.41)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>7(23.33)</td>
</tr>
<tr>
<td>TYP-type of admission*</td>
<td>Elective</td>
<td>2(3.77)</td>
</tr>
<tr>
<td></td>
<td>Emergency</td>
<td>38(25.85)</td>
</tr>
<tr>
<td>FRA-some fracture</td>
<td>No</td>
<td>37(20.00)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>3(20.00)</td>
</tr>
<tr>
<td>PO2-po2 from initial blood gases</td>
<td>&gt;60</td>
<td>35(19.02)</td>
</tr>
<tr>
<td></td>
<td>=60</td>
<td>5(31.25)</td>
</tr>
<tr>
<td>PH-ph from initial blood gases</td>
<td>&gt;7.25</td>
<td>36(19.25)</td>
</tr>
<tr>
<td></td>
<td>&lt;7.25</td>
<td>4(30.77)</td>
</tr>
<tr>
<td>PCO-pco2 from initial blood gases</td>
<td>=45</td>
<td>36(20.00)</td>
</tr>
<tr>
<td></td>
<td>&gt;45</td>
<td>4(20.00)</td>
</tr>
<tr>
<td>BIC-bicarbonate from initial blood gases</td>
<td>=18</td>
<td>35(18.92)</td>
</tr>
<tr>
<td></td>
<td>&lt;18</td>
<td>5(33.33)</td>
</tr>
<tr>
<td>CRE-creatinine from initial blood gases **</td>
<td>=2.0</td>
<td>35(18.42)</td>
</tr>
<tr>
<td></td>
<td>&gt;2.0</td>
<td>5(50.00)</td>
</tr>
<tr>
<td>LOC-level of consciousness at ICU admission *</td>
<td>No Coma</td>
<td>27(14.59)</td>
</tr>
<tr>
<td></td>
<td>Deep Stupor</td>
<td>5(100.00)</td>
</tr>
<tr>
<td></td>
<td>Coma</td>
<td>8(80.00)</td>
</tr>
</tbody>
</table>
The results in Table 1 show that AGE, SYS, SER, CRN, TYP, CRE and LOC are related to the mortality from the ICU at significance level 0.05. Patients who died in the ICU tend to be older than those who survived at average; the survivors apparently had a higher systolic blood pressure than the non-survivors at average; The mortality rate of patients undergone surgical procedure in ICU is less than half of that for patients who were just undergone general medical care in ICU; Patients who experienced chronic renal failure tends to have much higher mortality rate in ICU than those who didn’t (42.11% vs. 17.68%); Patients who received CPR prior to ICU admission have tripled mortality rate in ICU; Patients who admitted ICU in emergency form have much higher mortality rate in ICU than those who are elected to ICU (25.85% vs. 3.77%); Patients whose creatinine level of initial blood gases exceeds 2.0 have much higher mortality rate than those who have normal creatinine level of initial blood gases; Finally, the level of consciousness appears to be a dominant factor to determine the fate of patients admitted in ICU: patients who had coma or deep stupor died in ICU with very higher probability while patients who had no coma can survive from ICU with more than 85% of probability. However, we should be aware that the statistical statement for the effect of LOC is not valid due to the small cell problem.

We should also be aware that many variables in the study are highly correlated in common sense in medical knowledge, so the statement of effect without proper justification does not necessarily reflect the truth. To better assess the effect for the mortality in ICU, multivariate analysis is indispensable.

3. Analysis Using Multivariate Logistic Regression

We will build a multivariate logistic regression model to ascertain the risk factors and predict the mortality rate in ICU. Among these variables in the study, there are two categorical variables (RACE and LOC) which have more than two levels. Close looking at these two variables, who find that there is either a zero cell or close to a zero cell which may trouble the model fitting procedure in logistic regression. To avoid the problem, we have collapsed two levels for these two variables. The new RACE variable has only two levels: 0=white and 1=other and so does the new variable LOC: 0=no coma and 1=coma or deep stupor.

The stepwise procedure is a very efficient way to screen all the included variables for selecting potential significant risk factors to the outcome. We use p=0.15 for entering variable and p=0.20 for retaining the variables that have been entered from the previous steps. This procedure yields the following potential risk factors (in the order of importance) for the further analysis: LOC, TYP, AGE, CAN, SYS, PCO and PH. The likelihood ratio test as a very powerful test is applied to assess the statistical
significance of the selected variables. Table 2 presents the procedures of performing a sequence of the likelihood ratio test. Method 1 is referred to as a growing method which starts with the simplest model then grows the model until no more significant addition. As a counterpart, Method 2 is referred to as a pruning method which starts with the most complicated model then trims the model until no more variables can be removed in terms of statistical significance.

Table 2 Model selection with Likelihood Ratio Test

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable Entered</th>
<th>Log Like.</th>
<th>G</th>
<th>df</th>
<th>p-value</th>
<th>G</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>-100.081</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>LOC</td>
<td>-82.779</td>
<td>34.604</td>
<td>1</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>TYP</td>
<td>-77.257</td>
<td>11.043</td>
<td>1</td>
<td>0.0008</td>
<td>37.118</td>
<td>6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3</td>
<td>AGE</td>
<td>-72.656</td>
<td>9.203</td>
<td>1</td>
<td>0.0024</td>
<td>16.872</td>
<td>4</td>
<td>0.0024</td>
</tr>
<tr>
<td>4</td>
<td>CAN</td>
<td>-69.568</td>
<td>6.176</td>
<td>1</td>
<td>0.0129</td>
<td>10.695</td>
<td>3</td>
<td>0.0135</td>
</tr>
<tr>
<td>5</td>
<td>SYS</td>
<td>-67.806</td>
<td>3.523</td>
<td>1</td>
<td>0.0605</td>
<td>7.172</td>
<td>2</td>
<td>0.0277</td>
</tr>
<tr>
<td>6</td>
<td>PCO</td>
<td>-66.441</td>
<td>2.731</td>
<td>1</td>
<td>0.0984</td>
<td>4.441</td>
<td>1</td>
<td>0.0351</td>
</tr>
<tr>
<td>7</td>
<td>PH</td>
<td>-64.220</td>
<td>4.441</td>
<td>1</td>
<td>0.0351</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on Method 2, all 7 variables selected should be considered as significant risk factors attributed to the mortality in ICU.

We have two continuous variables, AGE and SYS, selected as risk factors for the mortality. In the stepwise selection we just performed above, the linear form for any continuous variables are assumed. We want to evaluate if there is other simple mathematical form for the continuous variables that would improve the model fitting. The fractional polynomial method is applied. A selected number of possible simple polynomial forms with power from \{-2,-1,-0.5, 0, 0.5, 1, 2, 3\} is considered. The best \(J=1\) model is referred to the model with the polynomial form for the continuous variable selected from the above set that yields the largest likelihood function. The best \(J=2\) model is referred to the model with two combined polynomial terms each selected from the above that yields the largest likelihood function.

The best \(J=2\) model for AGE is the one with \(\log(\text{AGE}) \& \text{AGE}^{1/2}\), the likelihood ratio test for this model against the original model does not support the use of this form for AGE \((p=0.669)\). The best \(J=2\) model for SYS is the one with \(\text{SYS}^{-2} \& \text{SYS}^{3}\), again the likelihood test for this model against the original model does not support the use of this form for SYS \((p=0.889)\).

Two demographic variables, SEX and RACE, are not selected by the stepwise procedure. Though the two variables are commonly used as confounding variables in many biomedical studies, there is little logic to argue that the two variables would confound other risk factors in this particular study. Moreover, inclusion of these two variables does not change the regression parameters drastically. Unless there is
special interest to consider these demographic effects on the mortality in ICU, we would not consider these two variables in our model thereafter.

Interaction effect known as effect modifier is always a great interest in any biomedical study. Based on the main effect model that we just built, we adopt the same stepwise model fitting procedure to ascertain possible interaction effects among these seven variables. We are aware that the zero-cell problem once again occurs when interactions between LOC and other binary variables are considered. This is because there are only 15 cases of either coma or deep stupor and among them 13 did not survive. Further grouping these cases according other variables would cause empty cells in some subgroups. Thus, the interactions between LOC and other variables are removed from stepwise procedure.

Two interaction terms associated with AGE stand out from the procedure as possible important factors to be considered based on entry probability 0.15 and removing probability 0.20. The result is displayed in Table 3.

Table 3 Analysis of Maximum Likelihood Estimator

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\beta}$</th>
<th>$\text{s.e.}(\hat{\beta})$</th>
<th>Wald-$\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-16.815</td>
<td>4.820</td>
<td>12.172</td>
<td>0.0005</td>
</tr>
<tr>
<td>LOC</td>
<td>5.119</td>
<td>1.288</td>
<td>15.790</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TYP</td>
<td>3.926</td>
<td>1.365</td>
<td>8.272</td>
<td>0.0040</td>
</tr>
<tr>
<td>AGE</td>
<td>0.207</td>
<td>0.067</td>
<td>9.431</td>
<td>0.0021</td>
</tr>
<tr>
<td>CAN</td>
<td>-5.128</td>
<td>4.941</td>
<td>1.077</td>
<td>0.2994</td>
</tr>
<tr>
<td>SYS</td>
<td>0.069</td>
<td>0.032</td>
<td>4.615</td>
<td>0.0317</td>
</tr>
<tr>
<td>PCO</td>
<td>-2.513</td>
<td>1.039</td>
<td>5.850</td>
<td>0.0156</td>
</tr>
<tr>
<td>PH</td>
<td>1.951</td>
<td>0.872</td>
<td>5.007</td>
<td>0.0252</td>
</tr>
<tr>
<td>AGE*CAN</td>
<td>0.134</td>
<td>0.083</td>
<td>2.644</td>
<td>0.1040</td>
</tr>
<tr>
<td>AGE*SYS</td>
<td>-0.0013</td>
<td>0.0005</td>
<td>7.036</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

Based on Wald-test, we would select a relative simple model with only one interaction term AGE*SYS included. This will be considered as a candidate for the final model in this study.

The Hosmer and Lemeshow test, as an ad-hoc method, is used to assess the goodness-of-fit of this model since our model included continuous covariates. This test does not provide evidence of lack of fit to the data using this model as compared to the saturated model based on the selected variables. ($p=0.564$).

Conventionally, one may want to classify a patient as a survivor from ICU if the predicted probability of mortality in ICU is less than 50% and non-survivor otherwise. The quantity of the area under the ROC curve is a good measure of the discriminating power for a model: a good model should discriminate the survivor from non-survivor
via the predicted probability of the mortality in ICU, namely, a survivor should have a smaller predicted probability of the mortality than a non-survivor. The ROC curve for this model is presented in Figure 1 below.

The quantity for this model is 87.7% which strongly suggests the model have a satisfactory discriminating power.

Before we put this model in force for the future prediction of the probability of the mortality in ICU, we need to conduct model diagnostic analysis to find the possible influential observations in the data. The decision on either retaining or removing these possible influential observations for our final model will be made through close examination on these individual observations. Figures 2-4 plot the one step difference of Pearson Chi-square, deviance statistics and corresponding overall regression coefficient change when individual observations are deleted from the model fitting one at a time.

Both Figures 2 and 3 indicate that the four observations appear to have relative larger Pearson Chi-square and Deviance Statistics that measure the discrepancy between the actual observation and predicted value. In Figure 4 we pick up observations which yield overall regression coefficient change by more than 0.5 if deleted for the further investigation. The characteristics of these observations are summarized in Table 4.
We note that two patients with ID #84 and #881, respectively, were the only two patients in this study who experienced coma but survived from ICU. Although these two patients may have tremendous impact on the model fitting, we cannot remove them from the study if the effect of level of consciousness on survival is one of the primary interests. There are three young patients (#127, #380, #285) who died in ICU but had very low estimated probability of dying. Close looking at their characteristics used in our model, we found that they did not seem to have a strong medical reason for death. A reasonable guess is that they might die due to accident such as gun shot, car crash etc. If we want to study the mortality in ICU for a normal death, these three cases may be treated as “outliers” for the study and are suggested for removing from the study. Patient #154 who is 53 years old has a low estimated probability of death, but actually did not survive from the discharge of ICU. We would suggest for the
further investigation on this patient for the cause of death. Other observations appear not to be strong influential cases. We refit the model without the four influential observations: #127, #154, #285 and #380. The comparison of the regression coefficients between the two models is shown in Table 5.

Table 5: The Comparison of Estimates of Regression Coefficients between Two Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model with Full Data Set</th>
<th>Model without 4 cases</th>
<th>Change %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-15.947(4.759)</td>
<td>-24.280(6.592)</td>
<td>-52.25</td>
</tr>
<tr>
<td>LOC</td>
<td>4.932(1.187)</td>
<td>6.603(1.554)</td>
<td>33.88</td>
</tr>
<tr>
<td>TYP</td>
<td>3.158(0.983)</td>
<td>3.825(1.169)</td>
<td>21.12</td>
</tr>
<tr>
<td>AGE</td>
<td>0.203(0.068)</td>
<td>0.317(0.091)</td>
<td>56.16</td>
</tr>
<tr>
<td>CAN</td>
<td>2.605(0.901)</td>
<td>3.836(1.113)</td>
<td>47.26</td>
</tr>
<tr>
<td>SYS</td>
<td>0.065(0.032)</td>
<td>0.096(0.042)</td>
<td>47.69</td>
</tr>
<tr>
<td>PCO</td>
<td>-2.631(1.027)</td>
<td>-4.283(1.377)</td>
<td>-62.79</td>
</tr>
<tr>
<td>PH</td>
<td>1.830(0.867)</td>
<td>2.310(1.033)</td>
<td>26.23</td>
</tr>
<tr>
<td>AGE*SYS</td>
<td>-0.0012(0.0005)</td>
<td>-0.0017(0.0006)</td>
<td>-41.67</td>
</tr>
</tbody>
</table>

Clearly, two models are quite different. With the four observations deleted, the estimated regression parameters have changed a lot, the change ranges from 21.12% to 62.79%. If our conjecture about the young patients who died in ICU is justified, the modified model is encouraged to use in practice.
Finally, we apply our modified final model to predict the mortality in ICU for patients ranged from 40 to 90 who have the same clinical characteristics considered by the model: LOC=0, TYP=1, CAN=1, SYS=160, PCO=1 and PH=1. The estimated probability versus age along with the 95% confidence band is plotted in Figure 5. From Figure 5, we can roughly say that the probability of death in ICU ranges from 17% to 55% for patients age from 40 to 90 correspondingly given the above clinical characteristics.

4. Conclusion and Discussion

We build a multivariate logistic regression model to ascertain the possible risk factors attributed to the mortality in ICU. Seven factors: level of consciousness (LOC), type of admission (TYP), patient’s age (AGE), cancer as a present problem (CAN), systolic blood pressure (SYS), PCO2 from initial blood gases (PCO) and PH value from initial blood gases are found statistically significantly related to the mortality in ICU. The model we constructed does not lack of fit and has a satisfactory discriminating power. Detailed diagnostic analysis identifies four possible influential observations which may be considered as unusual cases compared to most of other cases. Decision on whether or not including these four observations for analysis needs further identification on the cause of death for these patients.

The model with the four patients removed shows that the odds of death in ICU for patients under coma or deep stupor is more than 700 times of patients who isn’t in coma nor deep stupor adjusting for other variables; The odds of the death for patients admitted to ICU in emergency form is 46 times of those in the elected form adjusting for the other variables; As cancer is one of the leading cause of death, it is expected to see that cancer plays an important role in explaining the death in ICU: controlling for other variables, the odds of death for patients with some forms of cancer is also more than 46 times of those without; Age is not surprised to be an important factor attributing to death. As in nature, the elder people have higher odds of death than their younger counterpart. However, the age effect has reduced as patients systolic blood pressure increases based on our model.

The extremely high odds of death for patients in coma or deep stupor is reflected by the factor that 13 out of 15 who had this incidence died in ICU in this data set. At this moment, we are not sure if this is basically the scenario in ICU, otherwise, the estimate of odds ratio is greatly biased since our data are not the representative of the ICU population. One way to remove this possible bias is to only consider patients who did not have coma or deep stupor in ICU if we think people in coma or deep stupor is the “sign” of death.

Though the probability curve over age is useful, the confidence band, however, is very uninformative, since the width of the confidence band is too wide. We have to note that the data set used to build the model does not have many observations in the neighborhood at any points of our test data set. This will cause a large variance in
estimation which in turn gives a wide confidence band. Using more data to build a model will ultimately provide a reasonable confidence band for the probability curve.

ACKNOWLEDGMENT

I own thanks to Ernie Martin for pointing out a useful source to obtain a SAS macro for annotating texts to the plots. This greatly improves the presentation of diagnostic plots in this analysis.

REFERENCE


Appendix: SAS codes

/* Descriptive Analysis*/
var age sys hra;
by sta;
run;

proc ttest data=logistic.icu;
   class sta;
   var age sys hra;
run;

proc freq data=logistic.icu;
   tables sex*sta race*sta ser*sta can*sta crn*sta inf*sta cpr*sta pre*sta typ*sta fra*sta po2*sta ph*sta pco*sta bic*sta cre*sta loc*sta/chisq;
run;

/* Data recoding*/
Data newicu;
   set logistic.icu;
   if RACE=1 then RACE=0;
   else RACE=1;
   if LOC>0 then LOC=1;
run;

/* The stepwise procedure*/
proc logistic data=newicu Descending;
   model STA=AGE SEX RACE SER CAN CRN INF CPR SYS HRA PRE TYP FRA PO2 PH PCO BIC CRE LOC /selection=stepwise slentry=0.15 slstay=0.2;
run;

data pvalue;
input G df;
p=1-probchi(G,df);
cards;
34.601 1
11.044 1
9.514 1
5.864 1
3.524 1
2.730 1
4.442 1
71.560 7
37.118 6
26.074 5
16.560 4
10.696 3
7.172 2;
run;

/* Scale for continuous variables AGE and SYS */
set newicu;
agen2=1/(age*age);
agen1=1/age;
agen05=1/sqrt(age);
agelog=log(age);
age05=sqrt(age);
age2=age*age;
age3=age*age*age;
sysn2=1/(sys*sys);
sysn1=1/sys;
sysn05=1/sqrt(sys);
syslog=log(sys);
sys05=sqrt(sys);
sys2=sys*sys;
sys3=sys*sys*sys;

/* Check for AGE*/
model STA=LOC TYP AGEN2 CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
model STA=LOC TYP AGEN1 CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
model STA=LOC TYP AGEN05 CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
model STA=LOC TYP AGELOG CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
model STA=LOC TYP AGE05 CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
model STA=LOC TYP AGE CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGEN2*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGEN1 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGEN05 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGE CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN2 AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGEN1*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGEN05 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGE05 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGE CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN1 AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGEN05*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGELOG CAN SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE05 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGELOG AGE05 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGElogs CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGEN05 AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGELOG AGE05*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGELOG AGE CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGELOG AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGELOG AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE AGE*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE AGE2 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE2 AGE2*AGELOG CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE2 AGE3 CAN SYS PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE3 AGE3*AGELOG CAN SYS PCO PH;
run;
data P_value;
  input G df;
  P_value=1-probchi(G,df);
cards;
  1.558 3
;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSN2*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSN1 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSN05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSPCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYS3 PCO PH;
run;

proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSN2*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSN1 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYSN2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYSPCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN2 SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYSN1*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYSN05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYS05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN1 SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYS05*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYS05 SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYS05 SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYSLOG SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSN05 SYSLOG SYS05 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSLOG SYSL0G SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSLOG SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
  model STA=LOC TYP AGE CAN SYSLOG SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS05 SYS05*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS05 SYS PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS05 SYS2 PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS05 SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS SYS2 SYS2*SYSLOG PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS SYS2 SYS3 PCO PH;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS SYS3 SYS3*SYSLOG PCO PH;
run;
/* A e SEX an RACE confounding variables ?*/
    proc logistic data=NEWICU Descending;
        model STA=LOC TYP AGE CAN SYS PCO PH SEX RACE;
run;
proc logistic data=NEWICU Descending;
    model STA=LOC TYP AGE CAN SYS PCO PH;
/* Interaction effects?*/
    proc logistic data=NEWICU Descending;
        model STA=LOC TYP AGE CAN SYS PCO PH LOC|TYP|AGE|CAN|SYS|PCO|PH
            /selection=stepwise slentry=0.15 slstay=0.2 include=7;
run;
/* LOC appears to be a questionable variable to consider for the interaction effect */

```sas
proc logistic data=NEWICU Descending;
   model STA=LOC TYP AGE CAN SYS PCO PH TYP|AGE|CAN|SYS|PCO|PH /selection=stepwise slentry=0.15 slstay=0.2 include=7;
run;
```

/* Final elected model */

```sas
proc logistic data=NEWICU Descending;
   model STA=LOC TYP AGE CAN SYS PCO PH AGE*SYS/lackfit rsq outlier=rocdata;
run;
```

/* ROC curve */
```sas
goptions cback=white
colors=(black)
border;
axis1 length=2.5in;
axis2 order=(0 to 1 by 0.1) length=2.5in;
proc gplot data=rocdata;
symbol1 i=join v=none;
title1;
title2 'Figure 1. ROC CURVE of the Mortality Model';
plot _sensit_*_1mspec_ /haxis=axis1 vaxis=axis2;
```

/* Model diagnostic */
```sas
set newicu;
agesys=age*sys;
proc logistic data=newicu2 Descending;
   model STA=LOC TYP AGE CAN SYS PCO PH AGESYS;
   output out=diagnostic (keep=id pred dloc dtyp dage dcan dsys dpco dph
dagesys r dchi ddev c cbar HMD) p=pred dfbetas=int dloc
dtyp dage dcan dsys dpco dph
dagesys RESCHI=r DIFCHISQ=dchi DIFDEV=ddev C=c CBAR=cbar H=HMD;
run;
data influence;
   set diagnostic;
dbeta=r*r*hmd/((1-hmd)*(1-hmd));
```

```sas
/* ------------------------------------------------------------------*
* Name: label.sas                                                  *
* Title: Create an Annotate dataset to label observations           *
* in a scatterplot                                                 *
* Doc: http://www.math.yorku.ca/SCS/vcd/label.html                  *
* ------------------------------------------------------------------*
* Author: Michael Friendly <friendly@yorku.ca>                      *
*------------------------------------------------------------------*/
```
The LABEL macro creates an Annotate data set used to label observations in a 2D (PROC GPLOT) or 3D (PROC G3D) scatterplot. The points which are labeled may be selected by an arbitrary logical expression from those in the input dataset. The macro offers flexible ways to position the text label relative to either the data point or the center of the plot. The resulting Annotate data set would then be used with the ANNO= option of PROC GPLOT or PROC G3D.

**Usage:**

Values must be supplied for the X=, Y= and TEXT= parameters. For a PROC G3D plot, supply a value for the Z= parameter as well. The label macro is called with keyword parameters. The arguments may be listed within parentheses in any order, separated by commas. For example:

```plaintext
%label(x=age, y=response, text=name);
```

**Parameters:**

- **DATA=** The name of the input data set [Default: DATA=_LAST_]
- **X=** The name of the (numeric) X variable for the scatterplot, a numeric constant, or data step expression, in SYS= coordinates.
- **Y=** The name of the (numeric) Y variable for the scatterplot or a numeric constant, in SYS= coordinates.
- **Z=** The name of the (numeric) Z variable for a 3D scatterplot or a numeric constant, in SYS= coordinates.
- **BY=** The name(s) of any BY variable(s) to be used for multiple plots.
- **XOFF=** An X-offset for the text label. You may specify a numeric constant (XOFF=-1) in data units, or the name of a variable in the input data set. Positive values move the label up, negative values move it down.
to the right relative to the point; negative values move it
to the left.

* YOFF=       A Y-offset for the text label. Positive values move the
label
towards larger Y values.

* ZOFF=       A Z-offset for the text label, for a 3D plot.

* TEXT=       The text used to label each point. TEXT= may be
specified as a variable in the data set or a SAS
expression involving dataset variables (e.g.,
TEXT=SCAN(MODEL,1)) and/or string
constants. If you supply an expression, use
the
C<%str()> macro function, e.g.,
C<TEXT=%str(trim(name || '-' || place))> to
protect special
characters.

* LEN=        Length of the TEXT variable [Default: LEN=16]

* POS=        Specifies the position of the label relative to the
data point. The POS= value can be a character
constant
(one of the characters in
"123456789ABCDEF<+>", as used
by the Annotate POSITION variable), an
expression involving
dataset variables which evaluates to one of
the special
characters, "/", "|", or "-". The special
position values
(moved outward
toward the edges of the plot relative to the
data point.)
by comparing the coordinates of the point
to the
mean of
X and Y (/), or to the mean of X only (|), or
mean of Y only (-).

* SYS=        Specifies the Annotate XSYS, YSYS (and ZSYS) values,
as a list of up to 3 numeric values [Default: SYS=2]

* COLOR=      Label color (the name of a dataset character variable or
a
string constant enclosed in quotes. [Default:
COLOR='BLACK']

* SIZE=       The size of label (in whatever units are given by the
GUNIT goption). There is no
default, which means that the labels inherit the
global HTEXT setting.

* FONT=
The name of the font used for the label. There is no
default, which means that the labels inherit the
global FTEXT setting.

* ANGLE=
Baseline angle for label.

* ROTATE=
Character rotate for label

* SUBSET=
An expression (which may involve any dataset variables)
to
select points. A point will be labeled if the expression
evaluates to non-zero for the current
observation.
[Default: SUBSET=1]

* COPY=
The names of any variables to be copied to output dataset

* IN=
The name of an optional input annotate data set. If
specified, the IN= data set is concatenated with the
OUT= data set.

* OUT=
The name of the annotate data set produced. [Default:
OUT=_LABEL_] =Example:

This example plots Weight against Price for American cars in the Auto
data, labeling the most expensive cars.

%label(data=auto, x=price, y=weight,  
color='red', size=1.2,  
subset=origin='A' and price>10000,  
pos=1, text=scan(model,1));

proc gplot data=auto(where=(origin='A'));  
plot weight * price / frame anno=_label_;  
symbol1 v='+' i=none color=black h=1.5;
=*

%macro label(
  data=_LAST_,  
x=,             /* X variable for scatterplot */  
y=,             /* Y variable for scatterplot */  
z=,             /* Z variable for G3D (optional) */  
cvar=,         /* name of a curve variable */  
by=,            /* BY variable(s) (mult plots) */  
xoff=0,         /* X-offset for label (constant */  
yoff=0,         /* Y-offset for label or */  
zoff=0,         /* Z-offset for label variable */  
text=,          /* text variable or expression */  
len=16,         /* length of text variable */
);
%*-- Reset required global options;
%if &sysver >= 7 %then %do;
  %local o1 o2;
  %let o1 = %sysfunc(getoption(notes));
  %let o2 = %sysfunc(getoption(validvarname,keyword));
  options nonotes validvarname=upcase;
%end;
%else %do;
  options nonotes;
%end;
%* -- pos can be a constant, an expression, or / or -;
%* if a character constant, put "" around it;
%if %index(//-,&pos) %then %do;
  %*-- Out-justify wrt means of x,y;
  proc summary data=&data;
    var &x &y;
    output out=_means_ mean=mx my;
%end;
%else %if "&pos" ^= "" %then %do;
  %if %verify(&pos,%str(123456789ABCDEF<+>)) = 0
     %then %let pos="&pos" ;
%end;
%else %let pos = "5";
%if %length(&by) %then %do;
  proc sort data=&data;
    by &by &cvar
      %if %datatyp(&x) = CHAR %then &x;
    ;
%end;
run;
%local xsys ysys zsys;
%let xsys = %scan(&sys &sys &sys,1,%str( ));
%let ysys = %scan(&sys &sys &sys,2,%str( ));
%let zsys = %scan(&sys &sys &sys,3,%str( ));
options notes;
data &out;
  set &data;
  %if %length(&by) %then %do;
    by &by &cvar
      %if %datatyp(&x) = CHAR %then &x;
  ;
};
keep x y xsys ysys position function
%if %length(&size) %then size ;
%if %length(&angle) %then angle ;
%if %length(&rotate) %then rotate ;
color text &by ©
length function color $8 text $ &len position $1;
xsys = &"xsys"; ysys = &"ysys"; function='LABEL';
x = &x + &xoff ;
y = &y + &yoff ;
%if &z ^= %str() %then %do;
retain zsys &"zsys"; keep z zsys;
z = &z + &zoff;
%end;
%if "&text" ^= ""
%then %do; text=&text; %end;
%else %do; text=left(put(_n_,5.)); %end;
%if %length(&size) %then %str(size=&size;);
%if %length(&angle) %then %str(angle=∠);
%if %length(&rotate) %then %str(rotatee=&rotate;);
color=&color;
%if &font ^= %str() %then %do;
keep style;
style = &"font";
%end;
%if "&pos" = "/" %then
%do;
retain mx my;
if _n_ = 1 then set _means_(keep=mx my);
if x > mx then
  if y > my then position = '3';
  else position = '9';
else
  if y > my then position = '1';
  else position = '7';
%end;
%else %if "&pos" = "-" %then
%do;
retain mx my;
if _n_ = 1 then set _means_(keep=mx my);
if y > my then position = '2';
else position = '8';
%end;
%else %if "&pos" = "|" %then
%do;
retain mx my;
if _n_ = 1 then set _means_(keep=mx my);
if x > mx then position = '6';
else position = '4';
%end;
/* if pos has more than one character, use them cyclically */
%else %if %qsubstr(&pos,1,1) eq %str("") %then
  %str(position=substr(&pos,1+mod(_n_,length(&pos)),1));
%else %str(position = &pos);
if (&subset);
run;
%if %length(&in) %then %do;
  data &out;
    set &in &out;
    %if %length(&by) %then %do;
      by &by;
    %end;
  %end;
%end;
%done:
  %*-- Restore global options;
  %if &sysver >= 7 %then %do;
    options &o1 &o2;
  %end;
  %else %do;
    options notes;
  %end;
%mend label;

title1;
%label(data=influence,x=pred,y=dchi,text=id,subset=dchi>10, pos=1, color='red');
  plot dchi*pred /vaxis=axis1 haxis=axis1 annotate=_label_;
  symbol v=dot i=none height=0.5;
  axis minor=none width=2 major=(width=2);
  title2 'Figure 2. Difference of Preason Chi-Square Statistics';
run;

title2;
%label(data=influence,x=pred,y=ddev,text=id,subset=ddev>5, pos=1, color='red');
  plot ddev*pred /vaxis=axis1 haxis=axis1 annotate=_label_;
  symbol v=dot i=none height=0.5;
  axis minor=none width=2 major=(width=2);
  title3 'Figure 3. Difference of Deviance Statistics';
run;

title3;
%label(data=influence,x=pred,y=dbeta,text=id,subset=dbeta>0.5, pos=1, color='red');
proc gplot data=influence;
  plot dbeta*pred /vaxis=axis1 haxis=axis1 annotate=_label_; 
  symbol v=dot i=none height=0.5;
  axis minor=none width=2 major=(width=2);
  title4 'Figure 4. Difference of Overall Regression Coeffcients';
data influencel;
  merge newicu2 influence;
  where id =84 or id=208 or id=204 or id=202 or id=645 or id=881 or 
    id=285 or id=127 or id=154 or id=380;
  keep id loc typ age can sys pco ph sta pred;
run;
/* refit the model after deleting observations #127, #154, #285 and #380. */
data newicu3;
  set newicu2;
  if id=127 then delete;
  if id=154 then delete;
  if id=285 then delete;
  if id=380 then delete;
run;

proc logistic data=newicu3 Descending;
  model STA=LOC TYP AGE CAN SYS PCO PH AGESYS;
run;

/* make a test data set*/
data testdat;
  do age=40 to 90 by 1;
    id=999; sex=0; race=1; ser=1; can=1; crn=1; inf=1; cpr=0;
    sys=160; hra=85; pre=0; fra=0; po2=0; ph=1; pco=1; bic=1; cre=1;
    loc=0; typ=1; sta=.
    output;
  end;
run;

data newdata;
  set newicu3 testdat;
run;

/* Draw confidence band */
proc logistic data=newdata descending;
  model STA=LOC TYP AGE CAN SYS PCO PH AGE*SYS;
  output out=death p=pred lower=lci upper=uci;
run;

data death;
  set death;
  where id=999;
run;
goption reset=all;
proc gplot data=death;
  symbol1 c=bl v=none i=join;
  symbol2 c=b v=none i=join;
  title5 " Figure 5 The 95% Confidence Band for the Probability of Death";
  plot pred*age=1 lci*age=2 uci*age=2 /frame overlay;
run;