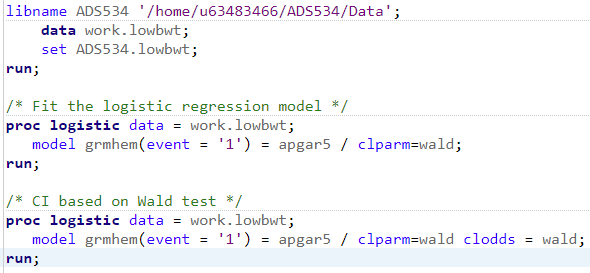
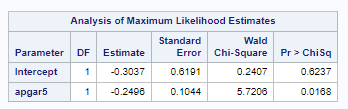
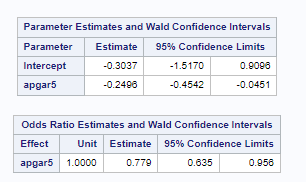
1. The dataset lowbwt.sas7bdat contains information for the sample of 100 low birth weight infants born in Boston, Massachusetts. The variable grmhem is a dichotomous random variable indicating whether an infant experienced a germinal matrix hemorrhage. The value 1 indicates that a hemorrhage occurred and 0 that it did not. The infants' five-minute apgar scores are saved under the name apgar5, and indicators of toxemia - where 1 represents a diagnosis of toxemia during pregnancy for the child's mother and 0 no such diagnosis - under the variable name tox.

First, we fit a logistic regression model where germinal matrix hemorrhage is the response and five-minute apgar score is the predictor, using 1 to represent the regression coefficient of apgar score.

* 1. At the 0.05 significance level, test the null hypothesis that β1 is equal to 0 using Wald test in SAS. What is the value of the test statistic? What's the distribution of the test statistic (including degrees of freedom)? What do you conclude?



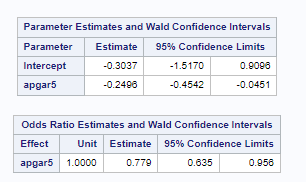




In a logistic regression, we find the test statistic W using W = {(βˆ1 - 0)/SE(βˆ1)}2. Using this output, we have W = {(-0.2496 - 0) / (0.1044)}2, or -2.39082, or 5.7159. With n = 100, we follow a chi square distribution, with 1 degree of freedom since we are testing for 1 parameter. Based on the chi-square distribution, the probability is 0.017, below the acceptance level of 0.05. This gives us evidence to reject the null hypothesis and conclude that apgar5 is a statistically significant predictor variable in determining the occurrence of grmhem.

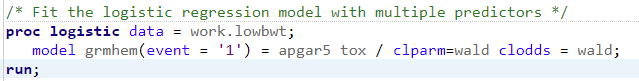
* 1. Construct a 95% Wald test based confidence interval for the population odds ratio of suffering a germinal matrix hemorrhage associated with 1 unit increase in five-minute apgar score. Does this interval contain the value 1? What does this tell you?

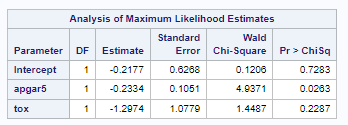
Next, fit a logistic regression model where germinal matrix hemorrhage is the response and toxemia status is the predictor, using β2 to represent the regression coefficient of toxemia status.



A 95% confidence interval for β1 can be easily calculated as βˆ1 ± 1.96SE(βˆ1). -0.2496 ± 1.96(0.1044) comes out to (-0.4542, -0.0450), which aligns with the calculated confidence interval from part a.

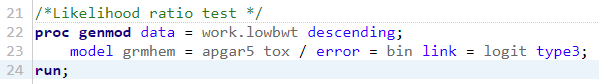
Based on a 1-unit increase in apgar5, the 95% Wald test based confidence interval is (-0.4542, -0.0451). This interval does not contain the value 1, which indicates that there is a statistically significant association between apgar5 and grmhem.

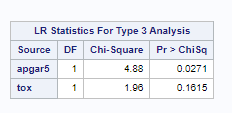




β2, based on the logistic regression model with multiple predictors, has a regression coefficient of -1.2974.

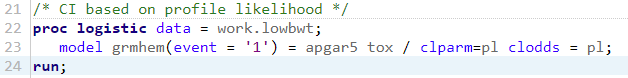
* 1. At the 0.05 significance level, test the null hypothesis that β2 is equal to 0 using likelihood ratio test in SAS. What is the value of the test statistic? What's the distribution of the test statistic (including degrees of freedom)? What do you conclude?

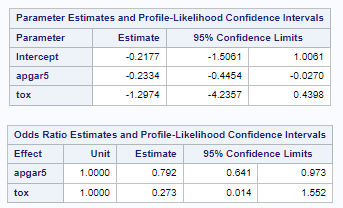




In a logistic regression, we find the test statistic using LR = -2log(L(βˆ0|β1 = 0)/L(βˆ0,βˆ1)) = -2log(maxL at H0 / maxL at MLE). Based on the LR statistics for type 3 analysis table, our test statistic calculates out to 1.96. With n = 100, we follow a chi square distribution, with 1 degrees of freedom since we are testing for 1 parameter. Based on the chi-square distribution, the probability is 0.162, which is above the acceptance level of 0.05. This does not give us evidence to reject the null hypothesis and conclude that tox is not a statistically significant predictor variable in determining the occurrence of grmhem.

* 1. Construct a 95% profile likelihood based confidence interval for the population odds ratio of suffering a germinal matrix hemorrhage for children whose mothers were diagnosed with toxemia relative to children whose mothers were not. Does this interval contain the value 1? What does this tell you?





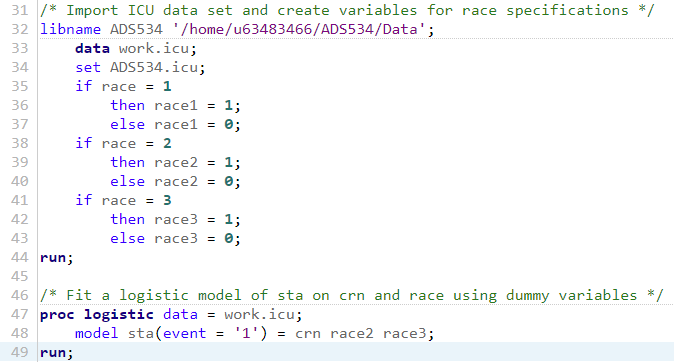
A 95% confidence interval for β2 can be calculated using 2log(L(βˆ0,βˆ1)/L(βˆ0|β1 = 0)) ≤ χ21,1-ɑ.

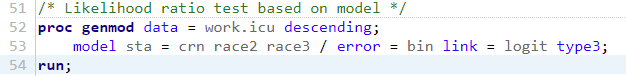
The 95% Wald test based confidence interval is (0.014, 1.552). This interval does contain the value 1, which indicates that there is not a statistically significant association between tox and grmhem.

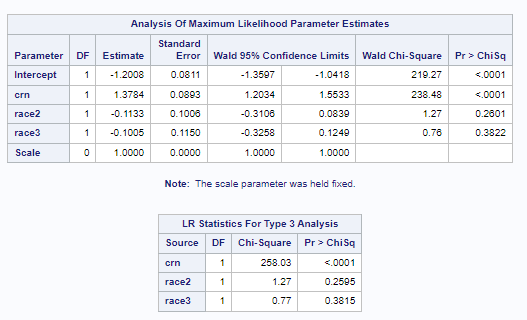
1. A sample of 2500 subjects was involved in a study on the association of patient death rate following admission to an adult intensive care unit (ICU) and hospital discharge of these patients. The data file icu.sas7bdat contains six columns correspond to id, sta (vital status; 0=alive, 1=dead), age (in years), sex (1=female, 0=male), race (1=white, 2=black, 3=other), crn (history of chronic renal failure; 0=no, 1=yes). Let π be the probability of death following admission to ICU.
   1. Fit a logistic model of sta on crn and race using dummy variables

log {π / 1 - π} = β0 + β1crn + β2race2 + β3race3 (1)

where race2 = 1 if race = black and 0 otherwise, and race3 = 1 if race = other and 0 otherwise. Test for no association between crn and sta (i.e., H0 : β1 = 0) adjusting for race using likelihood ratio test based on the logistic model (1).





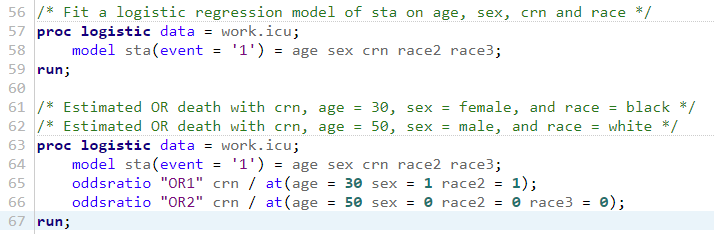


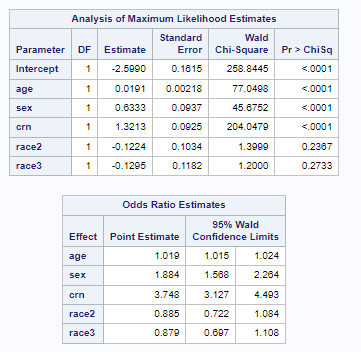
Based on the outcome of the likelihood ratio test, crn has an intercept of 1.3784 with standard error of 0.0893 and p-value of <0.0001. If we use a significance level of 0.05, this gives us evidence to reject the null hypothesis β1 = 0 and conclude that there is a statistically significant relationship where crn is a predictor of sta.

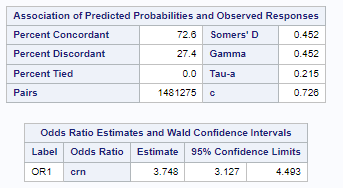
* 1. Fit a logistic regression model of sta on age, sex, crn and race.

log {π / 1 - π} = β0 + β1age + β2sex + β3crn + β4race2 + β5race3 (2)

What is the estimated odds ratio of death associated with crn for those with age = 30, sex = female and race = black? What is the odds ratio of death associated with crn for those with age = 50, sex = male and race = white?







The estimated odds ratio of death associated with crn for those with age = 30, sex = female and race = black is ORcrn = exp(β^3) = exp(1.3213) = 3.7483.

The odds ratio of death associated with crn for those with age = 50, sex = male and race = white is still ORcrn = exp(β^3) = exp(1.3213) = 3.7483.

* 1. Interpret β^1 in model (2). Estimate the odds ratio of death for every 10 years increase in age, adjusting for other predictors.

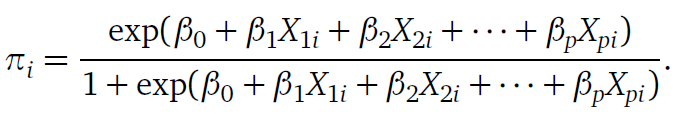
ORcrn = exp(β^1) = exp(0.0191) = 1.0193

ORcrn = exp(10β^1) = exp(10\*0.0191) = exp(0.191) = 1.2105

For every 10 years increase in age, there is an estimated odds ratio of death increase of 1.2105.

* 1. Based on model (2), calculate the estimated probability of death for an ICU patient with age = 50, sex = female, crn = yes and race = black?

Pr(sta=1|age=50 sex=1 crn=1 race2=1) =

 =

exp(-2.5990 + (50\*0.0191) + (1\*0.6333) + (1\*1.3213) + (1\*-0.1224) + (0\*-0.1295)

/ (1 + exp(-2.5990 + (50\*0.0191) + (1\*0.6333) + (1\*1.3213) + (1\*-0.1224) +

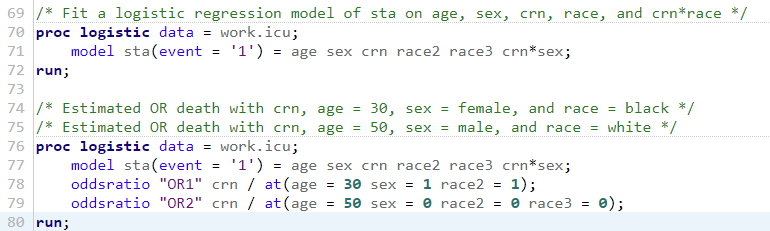
(0\*-0.1295)) = 0.5469

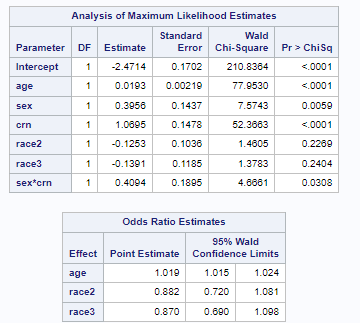
The estimated probability of death for a 50 year-old, female, black ICU patient is 0.5469.

* 1. Fit a logistic regression model

log {π / 1 - π} = β0 + β1age + β2sex + β3crn + β4race2 + β5race3 + β6crn\*sex (3)

What is the estimated odds ratio of death associated with crn for those with age = 30, sex = female and race = black? What is the odds ratio of death associated with crn for those with age = 50, sex = male and race = white?



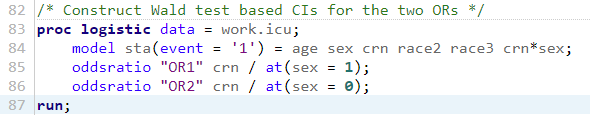


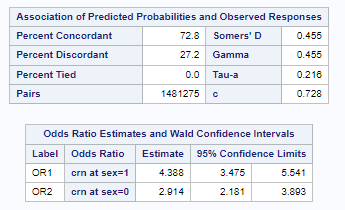
crn is located in β3 and β6, so the result will be a combination of those values.

The estimated odds ratio of death associated with crn for those with age = 30, sex = female and race = black is exp(β^3 + β^6) = exp(1.0695 + 0.4094) = exp(1.4789) = 4.3881.

The odds ratio of death associated with crn for those with age = 50, sex = male and race = white is exp(β^3) since β^6 will be equal to 0 based on the value for male. exp(β^3) = exp(1.0695) = 2.9139.

* 1. Construct 95% Wald test based confidence intervals for the two odds ratios in part (e). Do these intervals contain the value 1? What do they tell you?





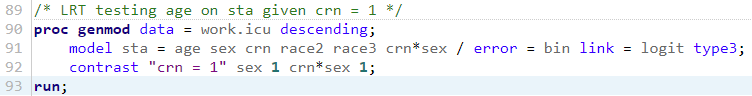
The odds ratio code was adjusted based on the major variable that held power in the assessment - sex. Based on this, the odds ratio for OR1 is (3.475, 5.541) and the odds ratio for OR2 is (2.181, 3.893). Neither the ratio for OR1 nor OR2 contains 1, indicating that there is a statistically significant association between crn and sta.

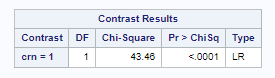
* 1. Based on model (3), we want to test whether sex is significantly associated with sta among those people with history of chronic renal failure (crn = 1). Write out the null and alternative hypotheses.

H0 : β^2 + β^6 = 0

HA : β^2 + β^6 ≠ 0

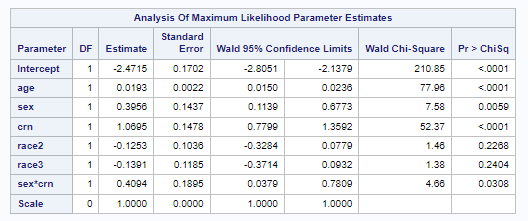
* 1. Conduct a likelihood ratio test for part (g) using SAS proc genmod procedure. Draw a conclusion under significance level 0.05.





The likelihood ratio test has a test statistic of 43.46 with a p-value of <0.0001, which is below the significance level of 0.05. This gives us evidence to reject the null hypothesis and conclude that sex is significantly associated with sta.

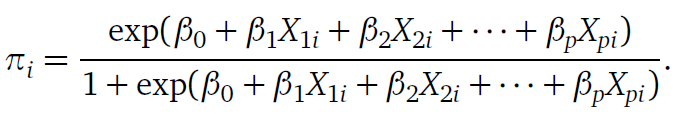
* 1. Based on model (3), obtain exp(β^3). Does it have a simple interpretation? Explain your answer.



exp(β^3) = exp(1.0695) = 2.9139. This is only part of the story of the model, as crn is also involved with sex in β^6, so on its own, it cannot give us any conclusions to draw.

* 1. Based on model (3), calculate the estimated probability of death for an ICU patient with age = 50, sex = female, crn = yes and race = black?

Pr(sta=1|age=50 sex=1 crn=1 race2=1) =

 =

exp(-2.4715 + (50\*0.0193) + (1\*0.3956) + (1\*1.0695) - (1\*0.1253) - (0\*0.1391) + (1\*1\*0.4094) / (1 + exp(-2.4715 + (50\*0.0193) + (1\*0.3956) + (1\*1.0695) - (1\*0.1253) - (0\*0.1391) + (1\*1\*0.4094)) = 0.5604

The estimated probability of death for an ICU patient with age = 50, sex = female, crn = yes and race = black is 0.5604.