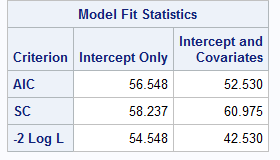
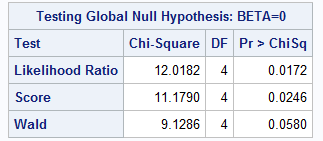
N = 40 observations

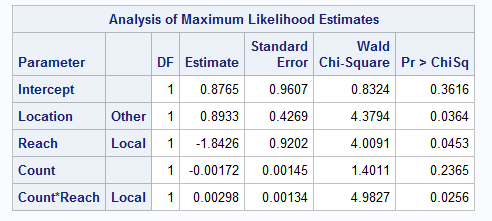
AIC and -2 Log Likelihood indicate covariates are important in model:



Beta = 0 test for model significance



analysis of effects using wald tests indicate significance for location and reach in the logistic model for bias (1). Count is not significant, however, we find a significant interaction between article word count and reach at p = 0.0256



We must exponentiate SAS MLE coefficients with exp(2\*coefficient) in order to obtain odds ratios. Location shows that bias in non-south news outlets are 5.97 (6) times the odds of south news outlets covering the Kim Davis story. Local outlets are 0.025 as likely to be biased as regional news outlets, indicating regional and national news outlets have greater (almost 40 times) odds of bias than do local news outlets in the presence of covariates location, article word count and the interaction of word count and reach. The interaction is effect is significant, however slight, in the presence of covariates previously mentioned. Indicating longer local articles have greater odds to be biased.

Under the proposed model, an article on Kim Davis before July 3rd 2015, written in a non-southern state by a local newspaper with 250 words would have a:

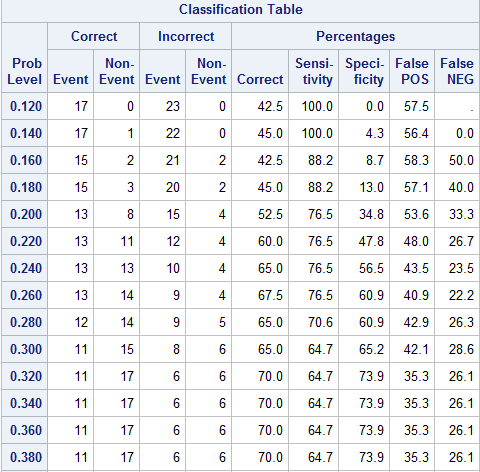
exp(2\*0.8765+2\*0.8933+(2\*-1.8426)+(2\*-0.00172)\*250+(2\*0.00298)\*250)/(1+exp(2\*0.8765+2\*0.8933+(2\*-1.8426)+(2\*-0.00172)\*250+(2\*0.00298)\*250)) = 62% probability of bias in the article

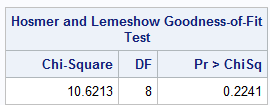
If the article has regional reach and is written in a non-southern state there is a:

exp(2\*0.8765+2\*0.8933)/(1+exp(2\*0.8765+2\*0.8933) = 97% probability of bias in the article

Classification Table

If we were to use this model to predict bias in an article, a probability threshold of 36% would be optimal to maximize correct classification. This data is based on the training data set itself and is used to classify the training set directly, therefore appropriate train / test procedures would need to be employed to ensure model fit for prediction is optimal.





Hosmer and Lemeshow goodness of fit test indicates the maximum likelihood estimate model is appropriate for our analysis ( p value of 0.22 on chi-squared distribution) and no quasi-MLE model is necessary. However, care should be taken in interpreting the HL test as it is primarily used for continuous predictors. We happen to have many categorical predictors.

Effect plot for probability of bias for location based on word count for reach = regional: