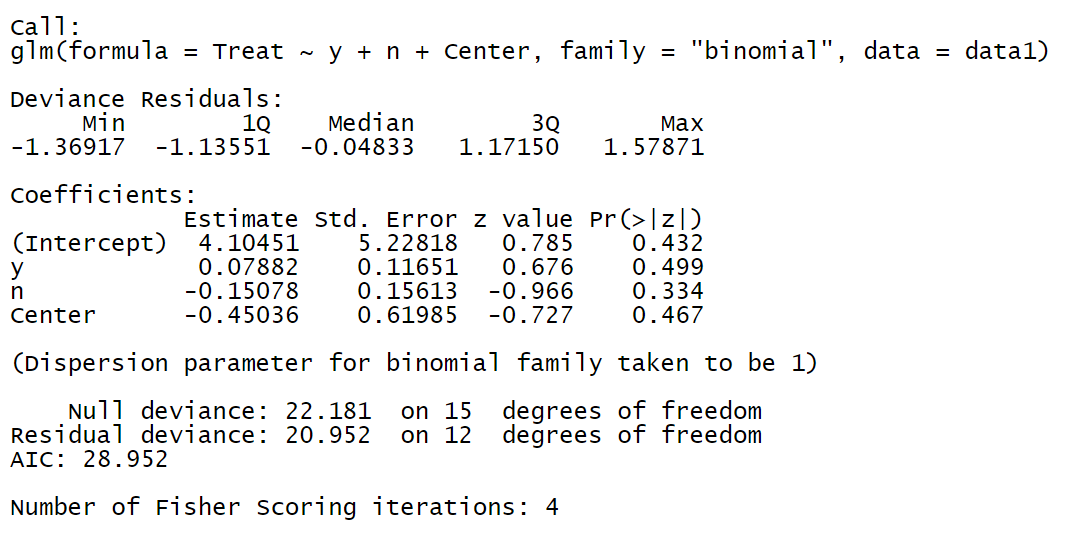
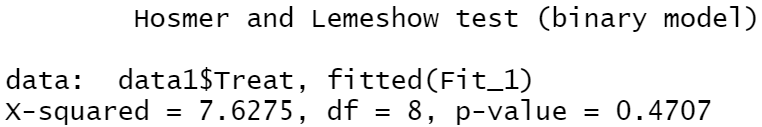
**1.**

**Exercise 4.14**



The results indicate that there is an insignificant difference during success of treatment and placebo effect (β=0.078, p>0.05), same applies to the failure (β=-0.151, p>0.05) and center (β=-0.450, p>0.05).

The following table shows Hosmer and Lemeshow goodness of fit test, the p value is greater than 0.05, which suggests that the model is a good fit and can be relied on.



**Fitting three models, one as an intercept model only, another with everything in(such as the one shown above) and another with centers only, we compare their chi square values to the critical value of chi squared with 8 degrees of freedom.**

**The p-value of the model is much bigger than 0.05 indicating we don’t have sufficient evidence to conclude there is a significant effect of predictors in the full model.**

**Moreover the difference in deviance which makes the Likelihoood ration test doesn’t suggest strong evidence to reject the null hypthesis**

**2.**

**Exercise 4.17:**



The Conditional odds ratio is exp(1.4) = 4.1for blacks

The conditional odds ratio for whites is exp(1.2) = 3.3

The difference between the log odds of 1.2 and 1.4 is the coefficient of cross product term which is 0.22

1. The coefficient of s of 1.2 is the log odds ratio between y and s when r=0 (whites) in which case the interaction does not enter the equation. This means people who smoke at least one pack per day have a log odds of having cancer 1.2 times higher than those in any other category. With a p-value < 0.01 for smoking represents the results of test that log odds ratio between y and s for whites equals 0. The p-value <0.01 also suggests we can reject the null hypothesis that the the coefficients are significant

For R this coefficient is 0.3. Thus, Blacks have log odds of having cancer 0.3 times higher than whites.

**3.**

**a.**

The delta-deviance or test statistics = 173.68 - 170.44 = 3.24

df = 155 -152 = 3 degrees of freedom

chi-squared (0.05, 3) = 7.815

Since 7.815 > 3.24, we fail to reject the null hypothesis that model 1 isn’t enough or that the difference of adding the three term interaction is statistically significant meaning the three factor interaction term can be removed from the model at the 0.05 significance level.

**b.**

In hierarchical modelling the idea is to reduce complexity in each stage, after the model including all the three possible two-way interactions, the following models would be the those which include only two of the possible two-way interactions (models 3a, b and c).

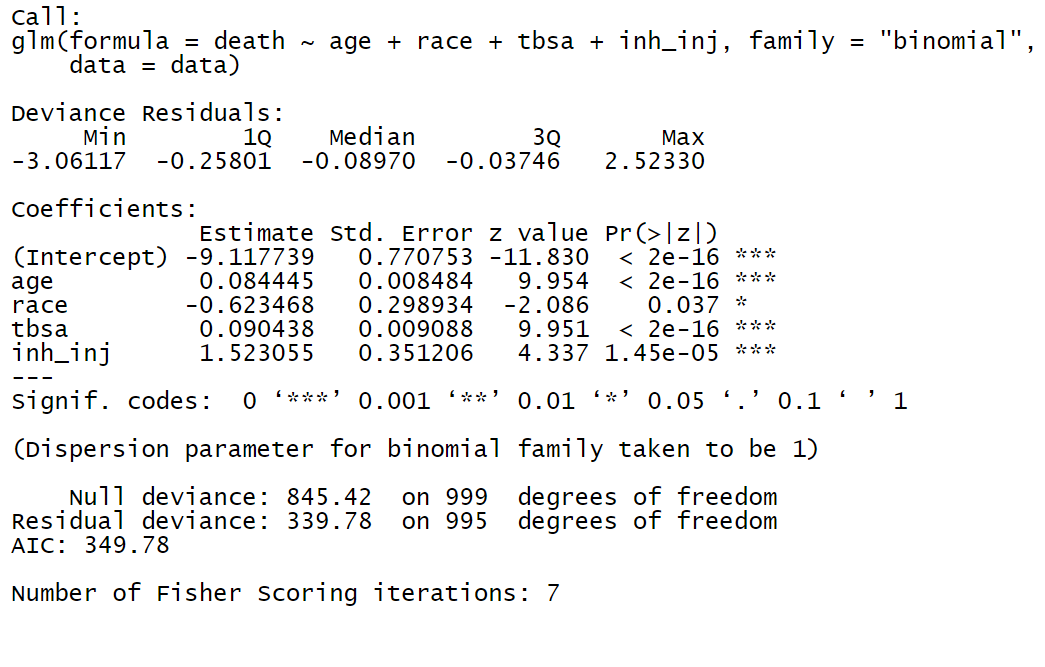
We would drop one term from the model with the smallest change in deviance which will be S\*W since the degrees of freedom go from 155 to 157. Which change in df being 2

**c.**

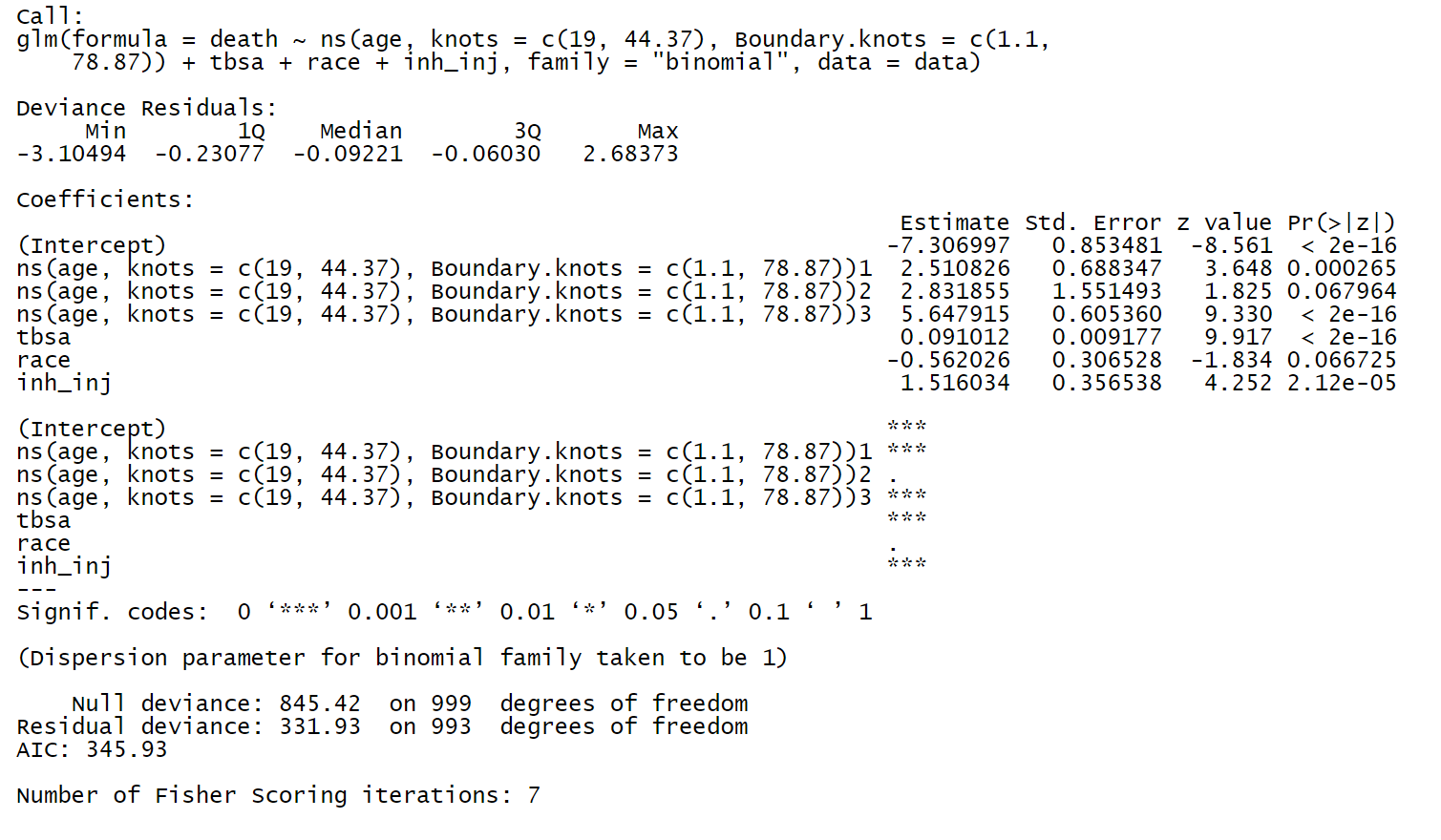
We will take out the C\*W term as model W + C\*S has the largest P-value

4.

**a.**

Age, race, tbsa and inh\_inj are all significant predictors of death at the 0.05 significance level. The difference in null and residual deviance suggest strong evidence that the predictors are significant in predicting the response.

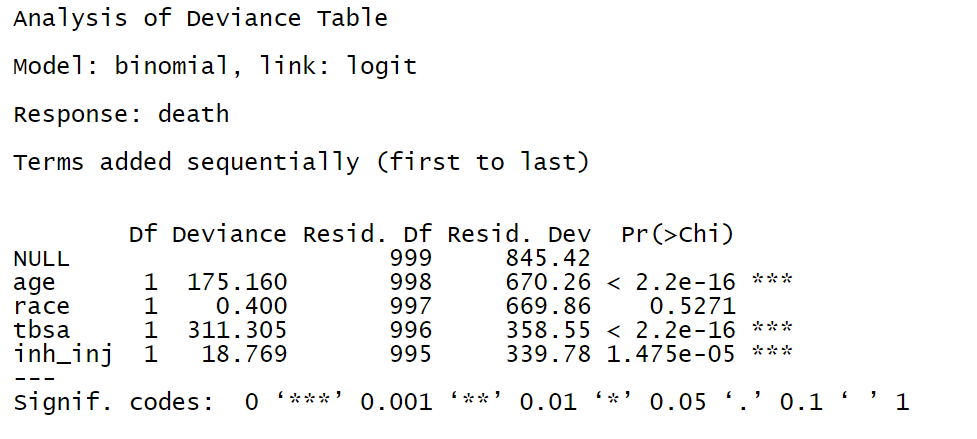
**b.**

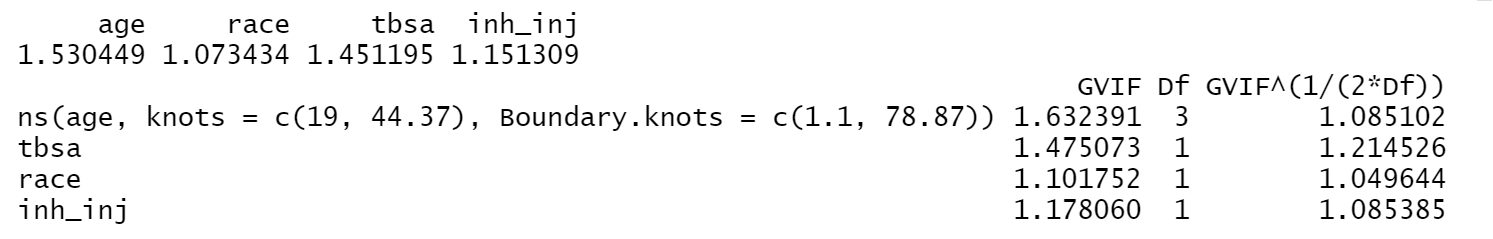


Once again the results suggests that smoothing the splines have significant effect on death looking at the high difference in deviance and the decrease in AIC. The results match with the desired outcome as required.

5.

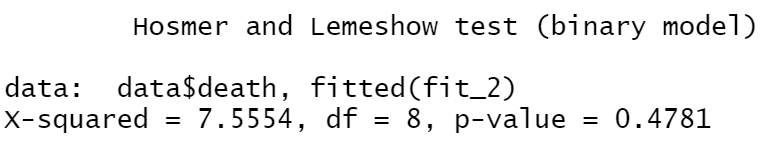
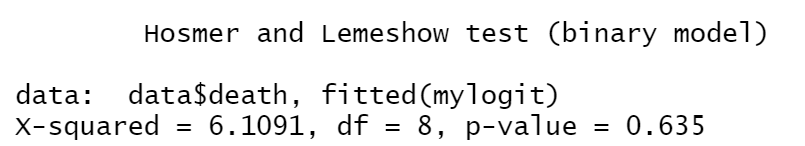
Diagnostics:





For the diagnostics of the model, two analyses have been employed, the ANOVA that provides the deviance of coefficients and the variance inflation factor that suggests whether or not we a multicollinearity problem. The results suggest that the race is not significant, thus removing it would provide a better model. Further, based on VIF, it appears that the model is free of multicollinearity and the values for VIF for both models are less than 10 which isn’t concerning.

Goodness of Fit Test is done using the HL test



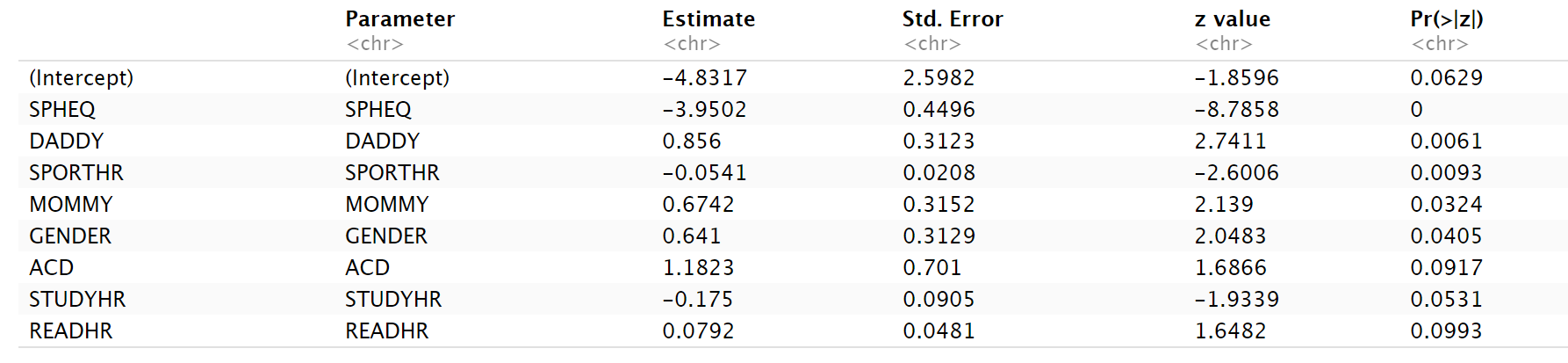
For Hosmer and Lemeshow test suggest p-values are pretty large which supports the evidence the fit was not great. For both models under analyses the p-values are greater than 0.05 and are significantly large. Thus, we conclude that the models are good fit.

6 The curve looks linear as expected.

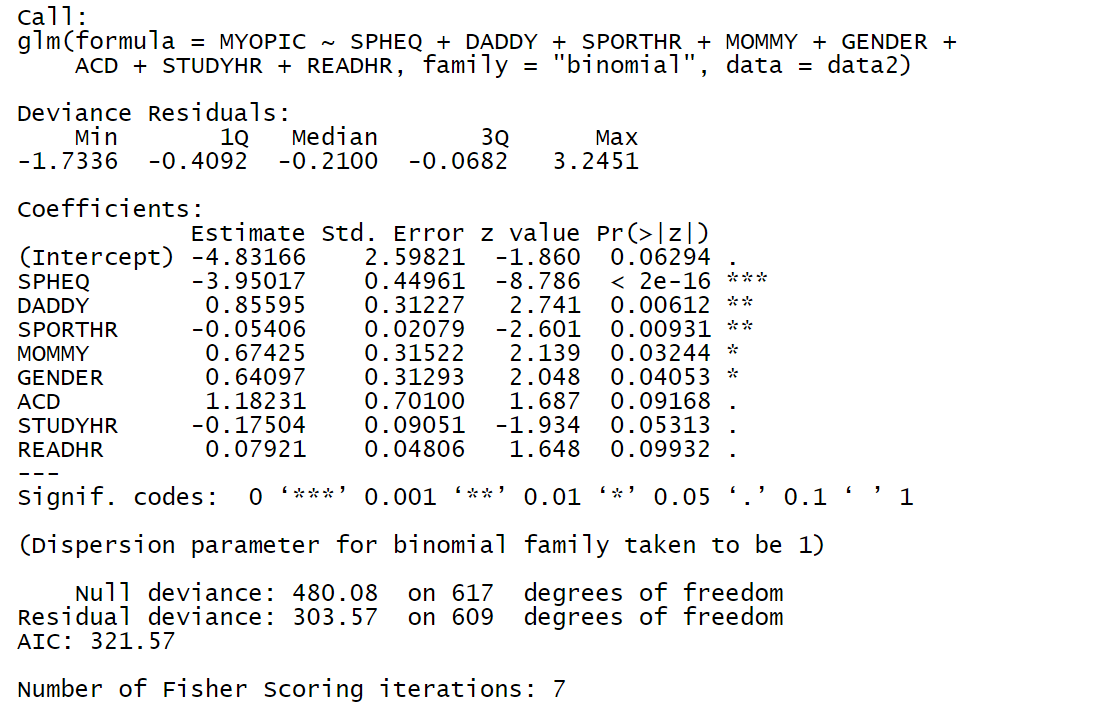
7.

To find the best model, we need to run the stepwise logistic regression. I showed bidirectional step wise logistic regression to accommodate for the both.

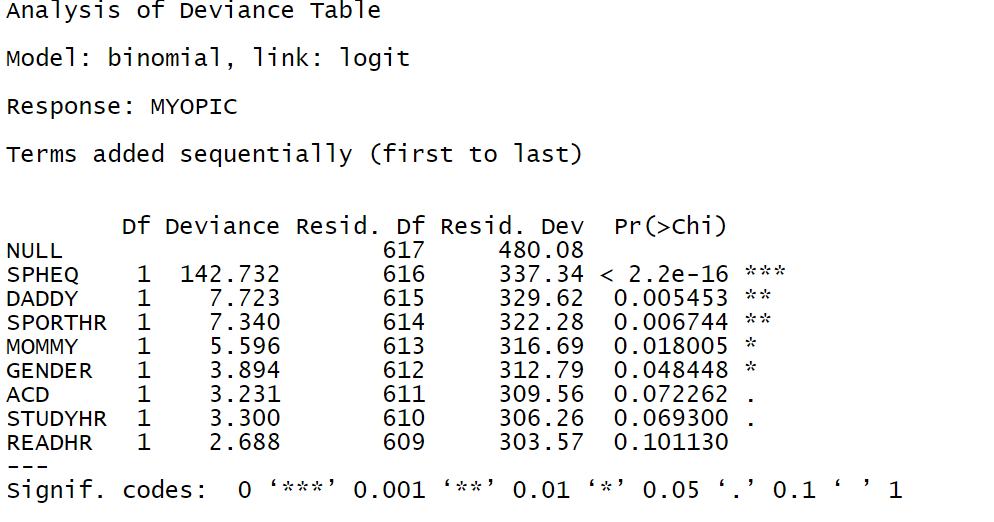
The following table shows that the significant predictors of Myopic are SPEHEQ, DADDY, SPORTHR, MOMMY, GENDER, ACD, STUDYHR and READHR as the p-values are less than the significance level of 10% (0.10).



Following the detection of key variables, we rerun the model in glm function to conduct subsequent test. From the results tabulated, GENDER, MOMMY, SPORTHR, DADDY and SPHEQ appear significant. The following table shows the logistic regression outcome for the best predictors.

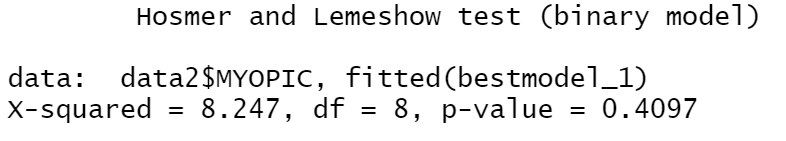


Now we run the ANOVA test to diagnose the data, the results show that the filtered variables are significant at 10% level of significance.



We further evaluate the model for the goodness of fit using the Hosmer-Lemeston GOF test.

The following table shows the results for the goodness of fit.



The results for Hosmer and Lemeshow goodness of fit show that the model is a good fit as the p-value of 0.409 which is greater than the alpha valiue of 0.05. Moreover, with a chi squared value of 8.25 with df= 8