E1

a.

The tabulate below shows that

1. In the sample, the number of patients with malignant tumor is almost twice as large as that of patients with benign tumor.
2. The mean proportion uniform shape of cells for patients with malignant tumor is less than 50% but slightly higher than 30%, which is much smaller than the proportion for patients with benign tumor, which is higher than 90%.
3. Uniformity of the cell’s shape has a higher standard deviation for patients with cancer, which is 2.56, compared with those who do not have cancer (s.d. =2.56).



b. Interpretation:

1. As shown in the tables below, the models fit much better when covariate “ushape” is included, because the AIC drops from 902.53 (intercept only) to 288.25 (intercept and covariate).
2. Both intercept and ushape are significant (p-values <.0001).
3. When ushape is evaluated at 0, the log-odds of this patient having cancer is -5.06. However, as a matter of fact, evaluating ushape at zero is out of the range of plausible ushape scores (1-10).
4. More specifically, with 1-unit increase in ushape, the difference in log-odds for having cancer is expected to increase by 1.4068, and the odds of having cancer (malignant tumor) is expected to increase by 308% (e^1.4068=4.08).



c.

1. The plot below shows that as the proportion of cells uniformity increases, the estimated probability of having cancer (malignant tumor) increases. However, the speed of probability increase is relatively slow when ushape is below 2 or is above 6. When ushape is rising from approximately 3 to 4, the speed is the largest of increase of the probability of having cancer (the slope is the largest).
2. According to the plot, when ushape is approximately 3.8, the probability of having malignant tumor exceeds 0.5.



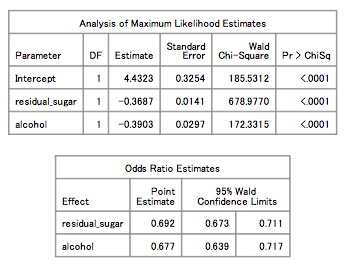
E2

a.

1. The procedure that is used here for model selection is stepwise selection. As is shown below, the model includes 10 out of the 11 predictor, with only the variable “ fixed\_aciditys” being left out.
2. The model tells us that quality does not predict whether the wine is red or white (p-values >.05). Given other variables in the model are held constant, with 1-unit increase in volatile\_acidity, chlorides, free\_sulfur\_dioxide, density, sulphates, and alcohol, respectively, the odds that the wine is red is expected to increase by more than 99899.9%, more than 99899.9%, 7%, more than 99899.9%, 1850.7%, and 415%.
3. For 1-unit increase in citri\_acid, residual\_sugar, and total\_sulfur\_dioxide, the odds that the wine is red is expected to decrease by 90%, 58%, and 5.3%.
4. The predictor “quality” does not have a significant impact when it comes to predicting the odds of the wine being red.
5. Difficulties: quality is not a statistically significant predictor, but it was included in the model, which makes it hard to explain how quality influences the odds of the wine being red.

b.

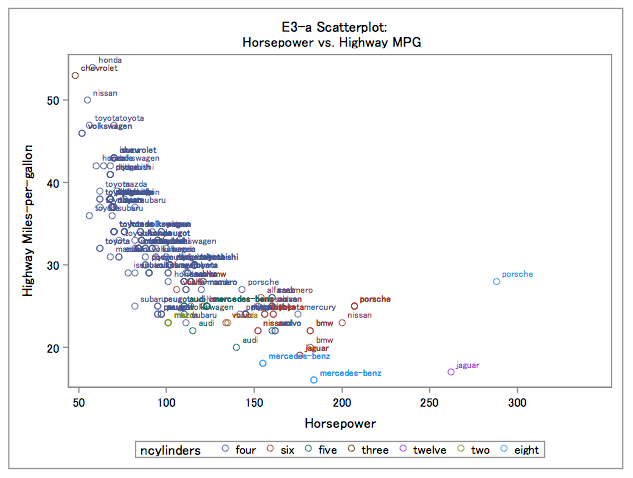
1. When alcohol is held fixed, for 1-unit increase in sugar, we can expect approximately 30.8% decrease in the odds of the wine being red; when sugar is held fixed, for 1-unit increase in alcohol, we can expect approximately 32.3% decrease in the odds of the wine being red.
2. The main effects of both predictors are statistically significant, in that p-values are less than 0.0001, and 95% Wald CL do not include 1.
3. According to the model, we can conclude that compared with white wine, red wine tends to have lower levels of residual sugar as well alcohol.



E3

a.

1. According to the scatter plot, most cars that fall within the low horsepower high mpg area have four cylinders.
2. Compared with four-cylinder cars, six-cylinder cars tend to have higher horsepower and lower mpg.
3. Eight-cylinder cars tend to perform well on either horsepower or mpg, but with high horsepower, the mpg is relatively high (slightly higher than six-cylinder cars), and with low mpg, the horsepower falls in the middle of the continuum.
4. Two-cylinder cars have low horsepower and relatively low mpg.
5. Three-cylinder cars have both the lowest horsepower and highest mpg.
6. Twelve-cylinder cars have high horsepower and low mpg.
7. Overall, there’s a negative relationship between horsepower and mpg; cars with higher power tend to have lower mpg.
8. Cars with more cylinders tend to have higher horsepower, and lower mpg, except that two-cylinder cars actually have similar horsepower as four-cylinder cars but with relatively low mpg.

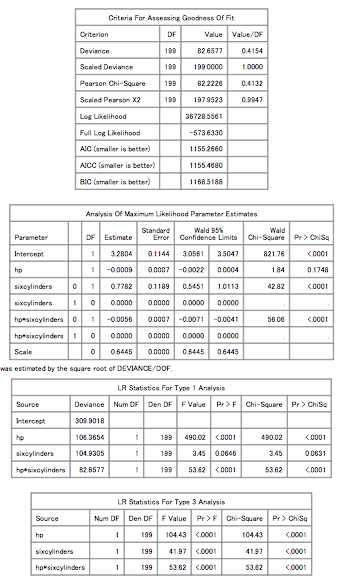


b.

AIC: 1155.27

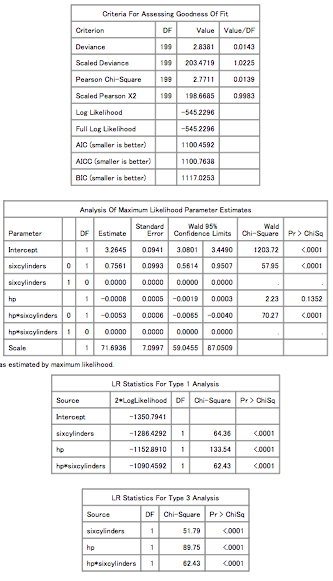
BIC: 1168.52

The interaction effect is statistically significant, in that the p-values obtained from maximum likelihood parameter estimates, type 1 and type 3 analysis are all less than 0.0001.



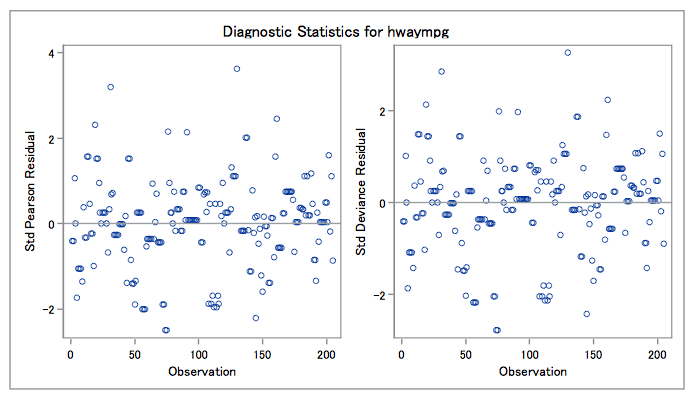
c.

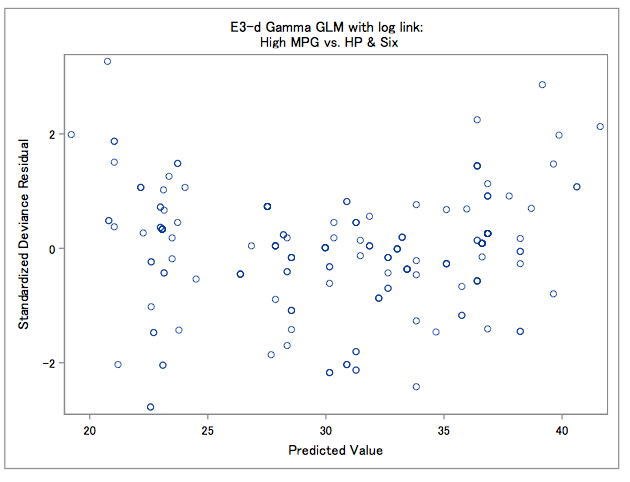
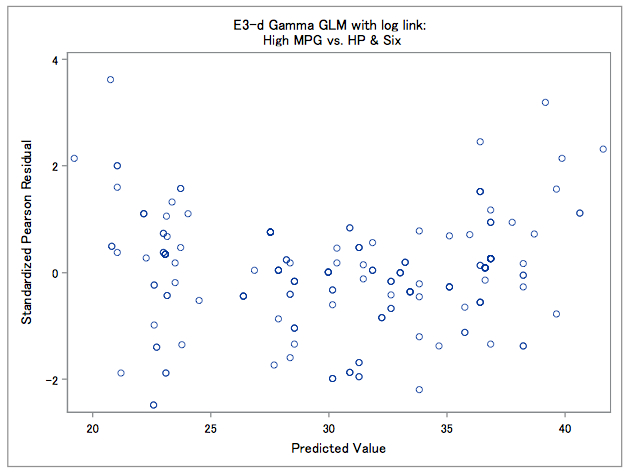
According to the tables below, the Poisson model and gamma model roughly agree with each other, in maximum likelihood parameter estimates in both models, hp is not significant, but according to type 1 and type 3 analysis, hp is significant. Sixcylinders is not significant in type 1 analysis in the Poisson model, but is significant in other analysis in the Poisson model and all three analyses in the gamma model. The interaction effect is significant in all analysis in both models. Basically, the Poisson and gamma models do not give very different results regarding predictors of mpg, which is that hp, sixcylinders and their interaction effect are all reasonable to be included in the model to predict mpg.



d.

The gamma model:





According to the graphs above, the residuals appear to be relatively evenly distributed around 0, which indicates that there are no problems with the model assumptions. Also, there does not appear to be a trend in the residuals against predicted values.

For cars with fewer than 6 cylinders, for a 10-unit increase in horsepower, we can expect an increase in highway MPG as large as e^0.7561=2.13. And for cars with more than 6 cylinders, for a 10-unit increase in the horsepower, we can expect a decrease in highway MPG as large as e^0.7561=2.13.