**Assignment #4**

**20175404 Rocku Oh**

***Problem 1.***

**This is a multinomial logistic regression problem. Quality is the dependent variable. Treat it as a categorical variable. Regress predictors on quality. Report the results. Interpret the output statistics including coefficients, those that are significant and non-signficant.**

The number of observation whose quality is 3 (very bad) and 9 (very good) is too small to compare the normal quality(6). We compared the normal quality of wine with bad(quality=4) and good(quality=8) rather. The multinomial logistic regression result is summarized in Table

*Low quality wine (quality = 4)*

In the case of the residual sugar, the relative risk of being in the bad quality wine group would be 14.1% less likely when the other variables in the model are held constant with the confidence level of 0.95. The less free sulfur dioxide contains in the low quality wine by the amount of 4.6% compared to the normal quality wine (significance level of 0.05). In the case of the proportion of the alcohol, 42% decrease in the low quality wine relative to the normal wine.

The relative risk ratio compared with red wine and white wine is 4400% for being in low quality wine vs. normal wine. In other words, the expected risk of being a normal wine is much lower for wine which is red in wine type. The other effect sizes are less substantial.

*Good quality wine (quality = 8)*

If one-unit increase in fixed acidity, the relative risk for good quality wine(8) relative to the normal quality wine(6) would be expected to increase by 59% given the other variables in the model are held constant (confidence level of 0.95). In the case of the residual sugar, the good wine contains 30% more as compared to the normal wine. We can say that if a wine were to increase its residual sugar, it would be expected to be a good wine as compared to the normal wine. For every unit increase in free sulfur dioxide, sulphates, and alcohol, the wine is more likely to be classified as a good wine by the factor of 2%, 391%, and 51% respectively with the significance level of 0.05. The other effect sizes are less substantial.

Table 1

*Multinomial Logistic Regression Result for Wine quality*

|  |  |  |
| --- | --- | --- |
| Predictors | Relative Risk Ratio  Exp (95% CI) | Relative Risk Ratio  Exp (95% CI) |
| Fixed acidity | 1.100 (0.880, 1.376) | **1.59 (1.159, 2.169)** |
| Volatile acidity | **1958.512 (660.026, 5811.543)** | 0.20 (0.037, 1.131) |
| Citric acid | 0.996 (0.283, 3.502) | 1.46 (0.357, 5.938) |
| Residual sugar | **0.859 (0.789,0.935)** | **1.30 (1.151, 1.467)** |
| Chlorides | 4.491 (0.030, 670.244) | 0.01 (0.000, 1161.514) |
| Free sulfur dioxide | **0.954 (0.939, 0.969)** | **1.02 (1.007, 1.031)** |
| Total sulfur dioxide | 1.000 (0.995, 1.005) | 1.00 (0.991, 1.003) |
| Density |  |  |
| pH | 1.604 (0.396, 6.506) | **15.77 (3.189, 77.956)** |
| Sulphates | 0.262( 0.065, 1.055) | **4.91 (1.433, 16.797)** |
| Alcohol | **0.580 (0.460, 0.732)** | **1.51 (1.031, 2.213)** |
| Type | **44.513 (19.205, 103.171)** | 0.66 (0.217, 2.034) |
|  |  |  |
| *Category* | *4 (Bad quality)* | *8 (Good quality)* |
| *Base outcome* | *6 (normal quality)* | *6 (normal quality)* |
| *Observations* | *6,497* | *6,497* |

*Note.* Multinomial Logistic regression was used to predict the quality of wine.