**Generalized Linear Models**

**Name: Rachit Shah (rshah25)**

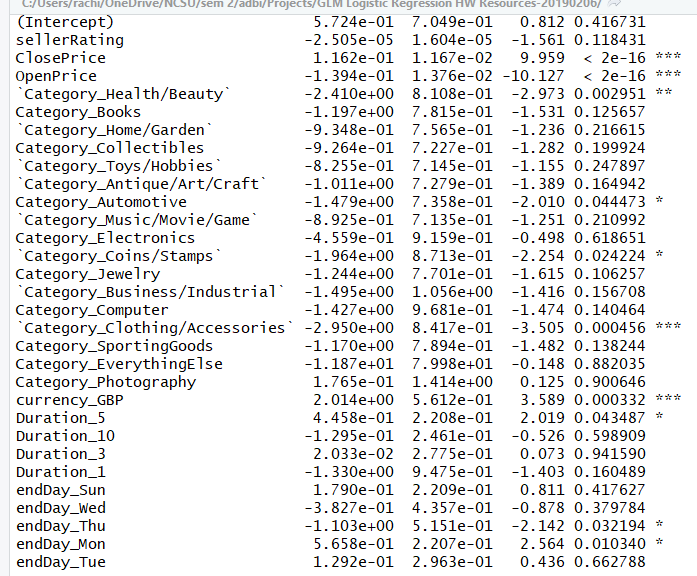
**Problem 1 (25 points: 5 points each question): Building and analyzing the logistic regression model**

For the problem below, build the logistic regression model (fit.all) using all the predictors and answer the following questions by including the corresponding R code and showing all the required mathematical derivations used to answer these questions:

1. **Let be the predictor with the highest estimate (in terms of its absolute value) for its regression coefficient. Build a single predictor logistic regression model (*fit.single*) using as the predictor. Write the equations relating the dependent variable (Response) to the explanatory variable in terms of:**

ANS:

From the summary of fit.all, we can see that the predictor with highest estimate is with an estimate of 2.014



1. Probabilities:
2. Odds:

1. Logit
2. **Write the estimated equation for the *fit.all* model in all three formats (if the number of predictors is more than four, then include only those four predictors whose absolute value estimates are the highest):**

**ANS:**

The 4 predictors with the highest estimates are: currencyGBP, endDayMon, Duration5 and endDaySun.

* 1. The logit as a function of the predictors.
  2. The odds as a function of the predictors.
  3. The probability as a function of the predictors

1. **Let be the predictor with the highest estimate (in terms of its absolute value) for its regression coefficient in the *fit.all*. Compute the odds ratio that estimated a single unit increase in , holding the other predictors constant. For example, if :**

**Provide the interpretation for this regression coefficient. If it were a linear regression model, how would the interpretation change for a single unit increase in .**

**ANS:**

Here, and the rest of the predictors are constant. Hence,

Since the estimate for currencyGBP is 2.014,

This means that for a unit increase of currencyGBP, the response variable will change 7.493 times for logistic regression. For 10 times increase in currencyGBP will cause 7.493^10 increase in response variable.

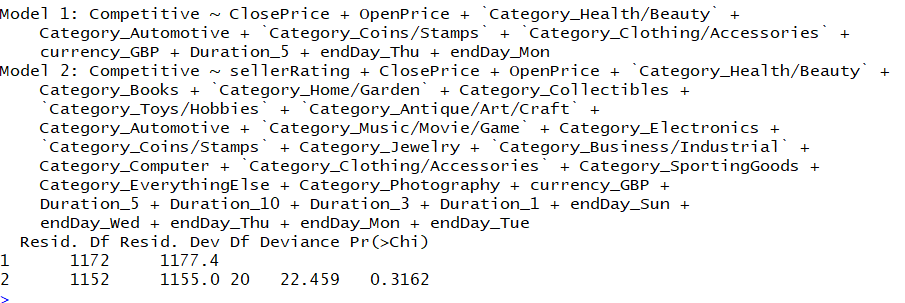
However, for linear regression, the change would be proportional to 2.014 and not its exponential.

1. **Build a reduced logistic regression model (*fit.reduced*) using only the predictors that are statistically significant. Assess if the reduced model is equivalent to the full model. Justify your answer.**

**ANS:**

The statistically significant predictors which we can ascertain from fit.all are: ClosePrice, OpenPrice, `Category\_Health/Beauty`, Category\_Automotive, `Category\_Coins/Stamps`, `Category\_Clothing/Accessories`, currency\_GBP, Duration\_5, endDay\_Thu and endDay\_Mon.

After fitting this reduced model and performing chi-square anova test we can find whether they are equivalent or not. From the result of the test, the p-value is 0.3162 which states that the difference is not significant and hence they are equivalent. Hence, we should choose the simpler model.



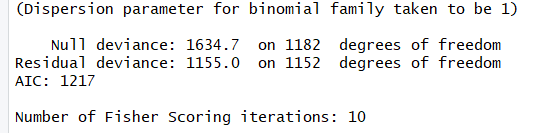
1. **Compute the dispersion of your model and run the dispersion diagnostic test. If the constructed model is overdispersed, then discuss the ways to deal with the issue.**

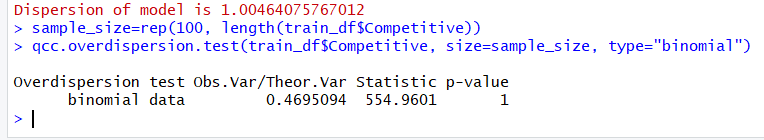
**ANS:**

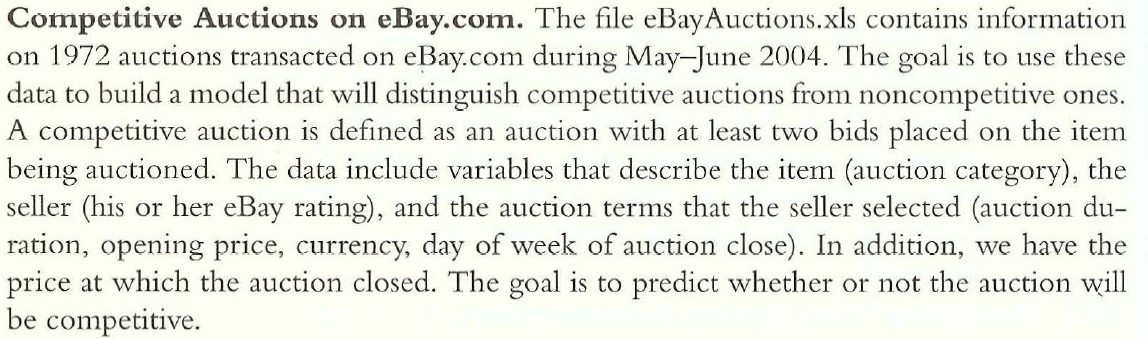
The dispersion of the model can be calculated by the formula:

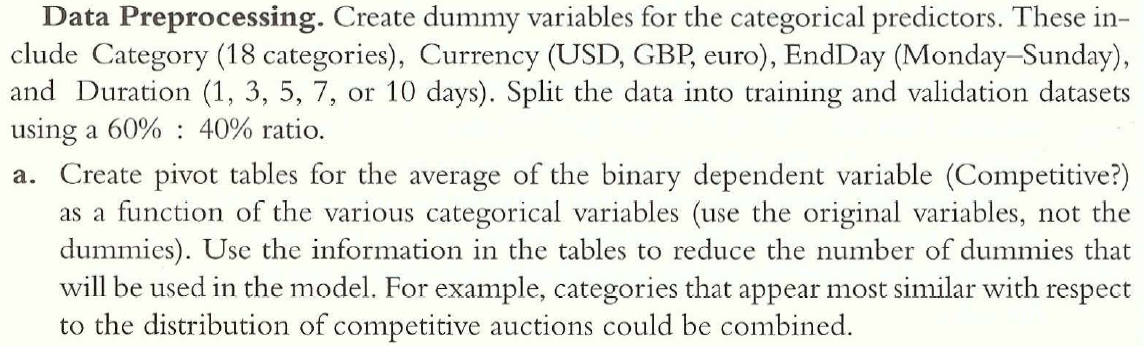
Hence, the dispersion is not too great then 1. Also, the dispersion diagnostic test in qcc package returns a p-value of 1 signifying that the model is not overdispersed.

If the test had resulted positive and there was overdispersion in our model, then we would have to refit our model with quasi-binomial distribution instead of binomial.

****

****





See R code