**Generalized Linear Models**

Ｗen-Han Hu (whu24)

**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:
   1. Probabilities:
   2. Odds:
   3. Logita
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):
   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
3. 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 :

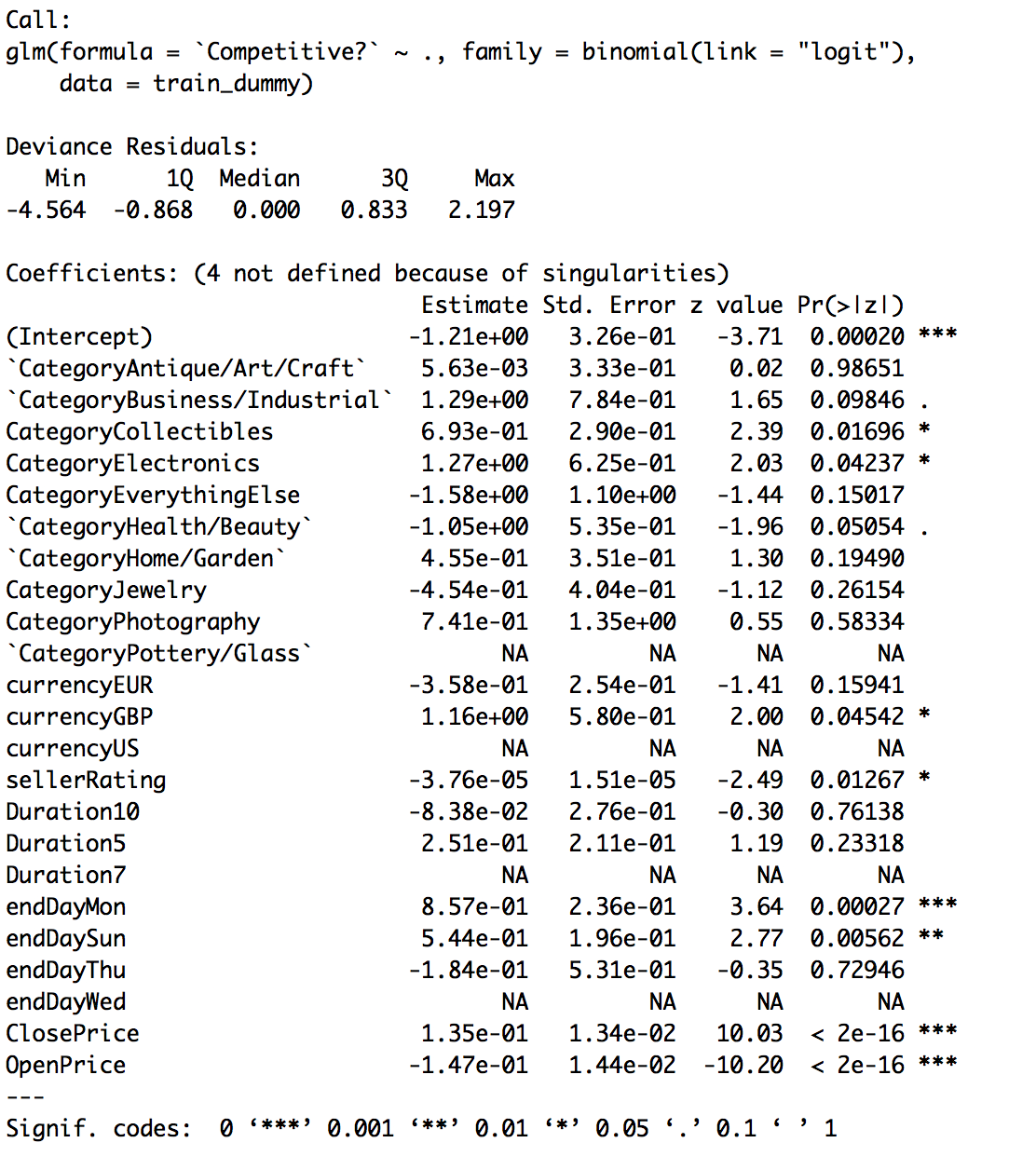
Provide the interpretation for this regression coefficient. If it were a linear regression model, how would the interpretation change for a single unit increase in .

Thus, odds ratio increases when a single unit increases in .

If it were a linear regression model, the increase of single unit will depend on its coefficient and directly reflect to Y. So, Y will increase .

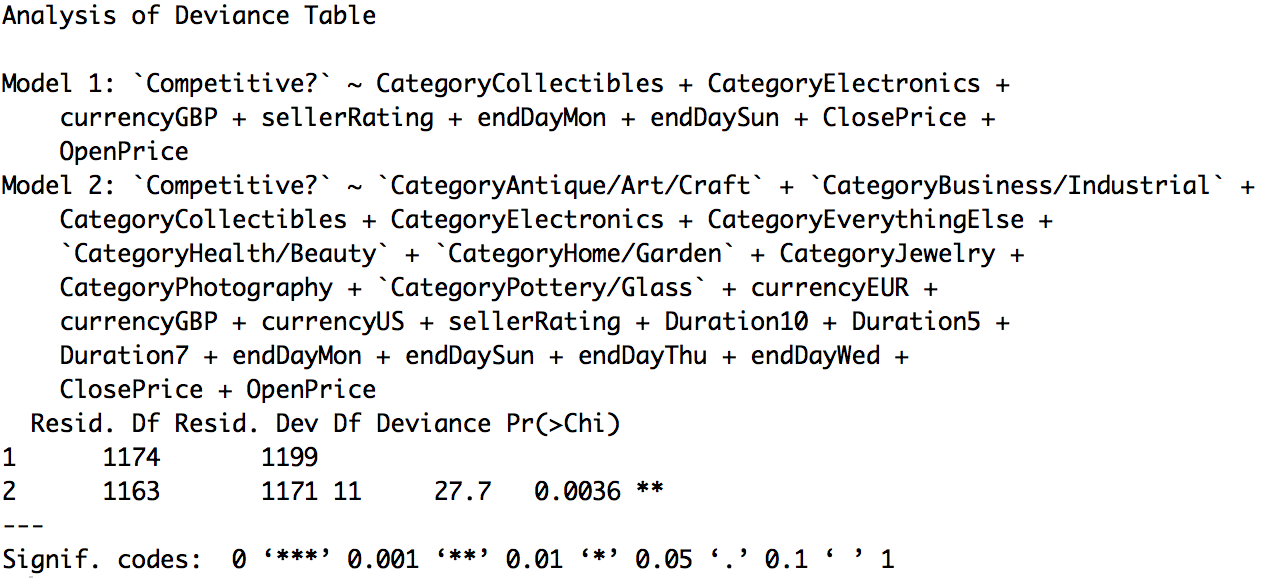
1. Build a reduced logistic regression model (*fit.reduced*) using only the predictors that are statistically significant. Assess if the reduced model is equivalent to the full model. Justify your answer.

The following picture is the screenshot of *fit.all:*



Those variables that the p-value <0.05 are significant predictors.

Then, do the *fit.reduced* with significant predictors. And the screenshot of comparing *fit.reduced to fit.all* as following:



Since the P-value ≤ α(0.05), we conclude that there is a statistically significant association between the variables. Thus, the reduced model is equivalent to the full 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.

Thus, no overdispersion on data.