CSC 591, Homework Generalized Linear Models

Student Id: 200261202

Mkhan8@ncsu.edu

**Solution 1:**

Let be the predictor with the highest estimate (in terms of its absolute value) for its regression coefficient.

Regression coefficient will highest coefficient is dummy\_EverythingElse with a coefficient value of -2.127e+00. So, we can say that .

Using as a single predictor for , the value of response variable



**Solution 2:**

The top four coefficients of fit.all model are:

* dummy\_EverythingElse (2.126912)
* dummy\_GBP (1.739977)
* dummy\_Health/Beauty (1.608681)
* dummy\_Coins/Stamps (1.359195)

**Solution 3:**

The highest coefficient of the fit.all model is dummy\_EverythingElse with a value of -2.126912. If we increase the value of by 1 keeping all the other co-efficient constant, the value of log odd will increase by .

We know the odds is given by where

If we increase the value of by 1 🡪 ,

Where

We know that

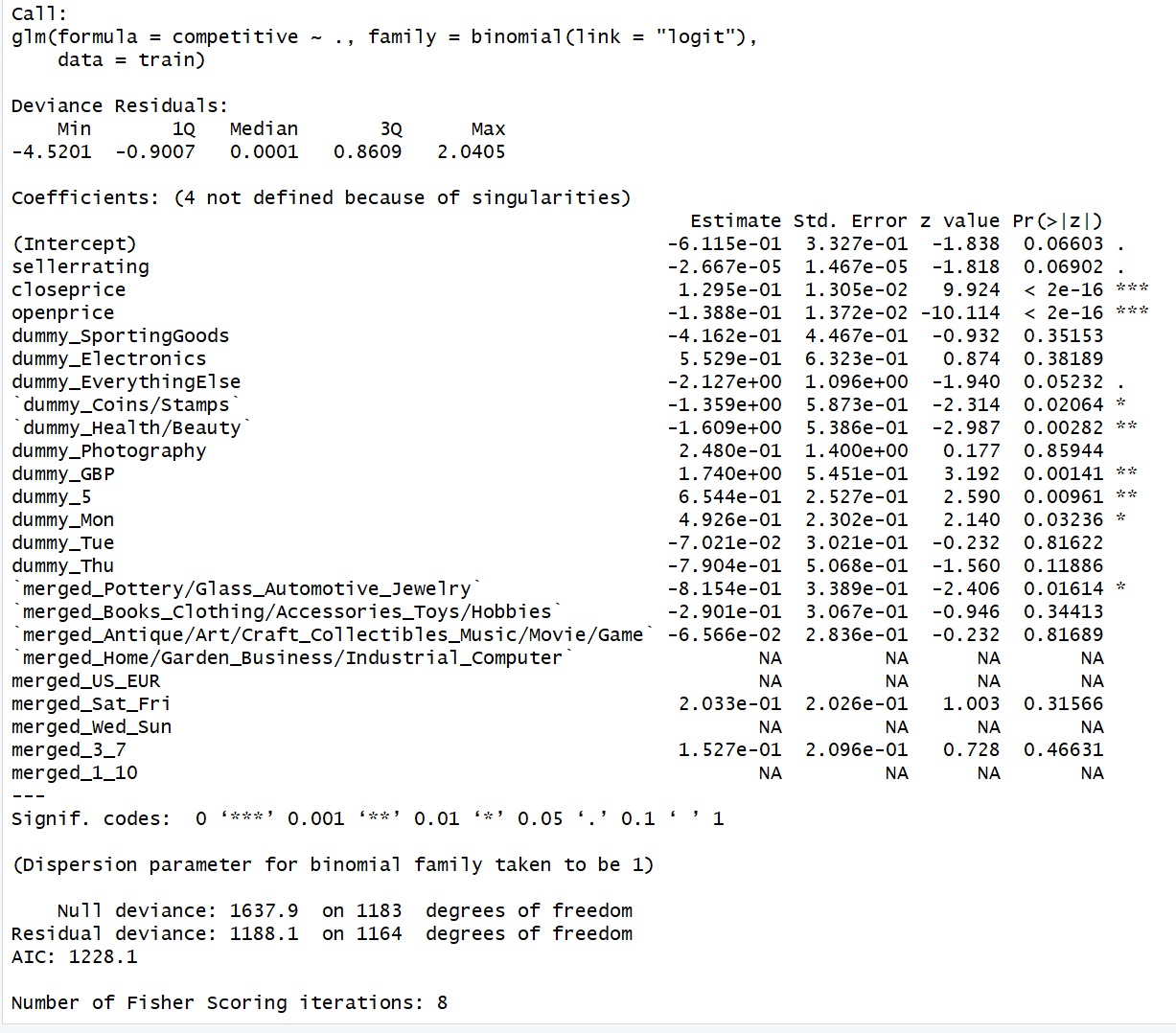
In case of linear regression:

And

In this case, the value will decrease by the value of the coefficient i.e. 2.126912

**Solution 4:**

The output of summary statistics of fit.all model:



The significant coefficients from the summary statistics of fit.all model are:

* Openprice
* Closeprice
* dummy\_Health/Beauty
* dummy\_GBP
* dummy\_5
* dummy\_Mon
* merged\_Pottery/Glass\_Automotive\_Jewelry
* dummy\_Coins/Stamps

**Model Comparison:**

Accuracy:

Fit.all model: 80.5%

Fit.reduced: 80.2%

Chisquare test:

anova(fit.reduced, fit.all, test='Chisq') 🡪 0.04615

* at and 0.001: the value 0.046 is non-significant. The null hypothesis is accepted, and the alternate hypothesis is ignored. Therefore, we can say that the models are equivalent.
* at 05: the value 0.046 is significant. The null hypothesis is ignored. Therefore, we can say that the models are not equivalent.

**Solution 5:**

We know that,

For a well-fitting model: Residual Deviance ≈ Residual d.f. or we can say that

In case of fit.reduced model , we know the from summary statistic of the model that :

Residual deviance: 1188.1 on 1164 degrees of freedom

As Residual deviance ≈ residual degree of freedom, so we can say that the model is not over-dispersed and is a well fitted model.