Homework 1

Prai 200263486

Q1.

Predictor with highest estimate (in terms of its absolute value) for its regression coefficient is:

X_h = Category_EverythingElse

(a) Probability: $\text{Prob}(Y = \text{Yes} \mid X_h = x) = \frac{1}{1 + e^{-(0.8164 + 2.835 * X_h)}}$

(b) Odds: $\frac{P(Y = \text{Yes})}{P(Y = \text{No})} = \frac{P(Y = \text{Yes})}{1 - P(Y = \text{Yes})}$

$= e^{(0.8164 + 2.835 * X_h)}$

(c) Logit: $\log(\text{odds}) = \log\left(\frac{P(Y = \text{Yes})}{1 - P(Y = \text{Yes})}\right)$

$= 0.8164 + 2.835 * X_h$

Q2. Four predictors: Category_EverythingElse (CEE)
Category_Health/Beauty (CHB)
currency_GBP (GBP)
Category_Coins/Stamps (CCS)

(a) Logit eq:

$$\text{logit} = 0.23831 - 1.89551\,CEE - 1.72639\,CHB + 0.49614\,GBP - 1.16753\,CCS.$$

(b) Odds eq:

$$\text{odds} = e^{(\text{logit})}$$
$$= e^{(0.23831 - 1.89551\,CEE - 1.72639\,CHB + 0.49614\,GBP - 1.16753\,CCS)}$$

(c) Probability:

$$P = \frac{\text{odds}}{1+\text{odds}}.$$
$$= \frac{1}{1 + e^{-(0.23831 - 1.89551\,CEE - 1.72639\,CHB + 0.49614\,GBP - 1.16753\,CCS)}}$$

Q3. X_n = Category_EverythingElse.

Coefficient of X_n = -1.89551 (t)

$$\frac{\text{odds}(X_n+1, X_2 \ldots X_q)}{\text{odds}(X_n, X_2 \ldots X_q)} = \frac{e^{\text{sum(others)} + t*(X_n+1)}}{e^{\text{sum(others)} + t*(X_n)}}$$

$$= e^{t}$$
$$= e^{-1.89551}$$
$$= 0.150242.$$

For logistic regression,

Unit change in variable ⟶ change equal to coefficient of that predictor in the log(odds)

For linear regression,

Unit change in variable ⟶ change equal to coefficient of that predictor in the overall prediction of model ⟶ (−1.89551)

Q4. - Predictors are significant if corresponding p-values < 0.05.

- Upon comparing 2 models using anova-test ⟶

the p-value returned for 'Chisq' test is 0.022792

∴ p-value < 0.05.

∴ it is satistically significant.

∴ null hypothesis is ignored.

Q5: Dispersion of model = $\frac{\text{Residual Deviance}}{\text{Residual degree of freedom}}$

$= \frac{1590.5}{1178} = 1.3502$

Dispersion test was run on data.

$\frac{\text{Obs. variance}}{\text{Theoretical variance}} = 0.45453$

p-value = 1

∴ value 0.45453 is not significantly different from 1.

∴ model is not overdisposed.