Generalized Linear Models

Ans 1- Using fit.all, the predictor (X_h) with the highest estimate of the regression coefficient is "Category - Coins. Stamps.
Hence, using 'Category - Coins. Stamps' as the single predictor to build the regression model.

a) Probabilities : Prob ($Y =$ Yes $| X_h = x$)

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} = \frac{1}{1+e^{-(0.43011-1.07387x)}}$$

$$= \frac{1}{1+e^{-(0.43-1.07 \times \text{Category\_Coins.Stamps})}}$$

b) Odds : Prob ($Y =$ Yes)

$$\text{odds} = e^{\beta_0+\beta_1 x}$$

$$= e^{-(0.43-1.07 \times \text{Category\_Coins\_Stamps})}$$

c) Logit : $\log_e\left(\frac{p}{1-p}\right)$

$$= \log_e\left(\frac{\frac{1}{1+e^{-z}}}{1-\frac{1}{1+e^{-z}}}\right) \quad \text{where } z = \beta_0+\beta_1 x$$

$$= \log_e\left(\frac{\cancel{\frac{1}{1+e^{-z}}}}{\frac{\cancel{1+e^{-z}}-1}{\cancel{1+e^{-z}}}}\right) = \log_e \frac{1}{e^{-z}} = \log_e e^{z}$$

$$= z = \beta_0+\beta_1 x$$

$$= 0.43 - 1.07 \times \text{Category\_Coins.Stamps}$$

Ans2. Equation for fit.all model (with 4 predictors)

a) $\text{logit} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$

$= 0.22 - 0.76x_1 - 0.55x_2 + 0.31x_3 + 0.29x_4$

where x_1 = Category_Coins. Stamps

x_2 = endDay_Sat

x_3 = currency_GBP

x_4 = Duration_1

b) $\text{odds} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4}$

$= e^{0.22 - 0.76x_1 - 0.55x_2 + 0.31x_3 + 0.29x_4}$

where x_1 = Category_Coins. Stamps

x_2 = endDay_Sat

x_3 = currency_GBP

x_4 = Duration_1

c) Probability : $P(Y = \text{Yes} \mid x_1, x_2, x_3, x_4)$

$$= \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)}}$$

$$= \frac{1}{1+e^{-(0.22 - 0.76x_1 - 0.55x_2 + 0.31x_3 + 0.29x_4)}}$$

where x_1 = Category_Coins. Stamps

x_2 = endDay_Sat

x_3 = currency_GBP

x_4 = Duration_1

Ans3-

$$\text{Odds Ratio} = \frac{odds(x_1+1, x_2, x_3, x_4)}{odds(x_1, x_2, x_3, x_4)}$$

$$= \frac{e^{\beta_0+\beta_1(x_1+1)+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4}}{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4}}$$

$$= \frac{e^{\beta_1} \times \cancel{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_4+\beta_4 x_4}}}{\cancel{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4}}}$$

$$= e^{\beta_1} = e^{-0.76} = 0.47$$

Interpretation : Odds Ratio of 0.47 (< 1) indicates a negative relationship between x_1 and y ie between Category_Coins.Stamps and Competitive. An increase in the value of Category_Coins.Stamps by 1 indicates a decrease in the odds of getting Competitive = 1 by 0.47.

If it were a linear regression, then an increase of the predictor variable Category_Coins.Stamp by 1 would have implied a decrease in the value of Response variable by 0.76.

Ans 4- Reduced logistic regression model. (fit.reduced).

$$p(x) = \frac{1}{1+e^{-\beta_0+\beta_1 x_1+\beta_2 x_2+ \ldots +\beta_5 x_5}}$$

$$p(x) = \frac{1}{1+e^{-(0.46-0.79x_1-0.68x_2-0.9x_3+0.76x_4-3.21\times10^{-5}x_5)}}$$

where x_1 = Category_Coins.Stamps
x_2 = endDay_Sat
x_3 = OpenPrice
x_4 = ClosePrice
x_5 = sellerRating

In order to determine equivalence of this model to the full model, using the ANOVA test. The p-value obtained from the test is 0.2559 which is > 0.05. Hence, we cannot reject null hypothesis and this means that this model is not significantly different from the ~~fit~~ fit.all model.

Ans5- ~~Over~~dispersion, $\phi = \frac{\text{Residual deviance}}{\text{Residual df}}$

$$= \frac{1282.8}{1184} = 1.083 \approx 1$$

Hence, there is no overdispersion.
Also, p-value of 1 obtained by running the dispersion diagnostic test suggests that there is no overdispersion in the model.