Homework

Unity ID -> hagrawa2
Name -> Harsh Agrawal

1)
Predictor with highest estimate (in terms of absolute value) is : Currency_EUR_US (currency usd and EUR combined) so $X_h = Currency\_EUR\_USD$

Equation for fit.single using this predictor is:

a)

$$Prob(Y = yes \mid X_h = Currency\_EUR\_US) = \frac{1}{(1 + e^{-(-0.22 - 0.215 * currencyEURUS)})}$$

b) Odds = P(Y = yes) / P (Y = no)

$$\frac{Prob(Y = yes)}{1 - Prob(Y = yes\ )} = e^{(-0.22 - 0.215 * currencyEURUS)}$$

c) Logit

$$\log\left(\frac{Prob(Y = yes)}{1 - Prob(Y = yes\ )}\right) = (-0.22\ -\ 0.215 * currencyEURUS)$$

2) Four most significant predictors are with their coefficients are

duration_5 = D5	-0.12367
Closeprice = CP	0.135496990839086
Openprice = OP	- 0.151161171550078
currency_eur_us = CED	- 0.215341437564903

a) Logit equation

$$logit = -0.22 - 0.21534 * CED -\ 0.15116 * OP + 0.13549 * CP - 0.12367 * D5$$

b) Equation for odds = e^{logit}

$$odds = e^{-0.22 - 0.21534 * CED - 0.15116 * OP + 0.13549 * CP - 0.12367 * D5}$$

c) Equation for probability

$$probabilty = \frac{odds}{1 + odds}$$

$$= \frac{1}{1 + e^{-(-0.22 - 0.21534 * CED - 0.15116 * OP + 0.13549 * CP - 0.12367 * D5)}}$$

3) For our fit.all model $X_h = Currency_{EUR}$, Let coefficient be c, then have

$$\frac{odds(X_h + 1, X_2, X_3 \dots)}{odds(X_h, X_2, X_3 \dots)} = \frac{e^{((X_2 + X_3 \dots) + c(X_h + 1))}}{e^{((X_2 + X_3 \dots) + c.X_h)}} = e^c = e^{-0.21534} = 0.80623$$

For Logistics regression, Unit change in a variable will be equal to $e^{coefficient}$ of that variable.
If it was a linear regression, the change would be directly equal to the coefficient of that variable.

4) We can identify the significant predictors from the p-value in the summary table of the fit.all. They are also marked by * in R. We only have 2 significant predictors in our case.

1) OpenPrice - OP
2) ClosePrice – CP

p-value for both of these is < 0.05, further confirming our analysis.
Building a model using these 2 and using the anova() function to perform the chi-square test, we get a chi square value(p-value) of 0.267, which is insignificant. According to slide 27, we can conclude that the new model with fewer variables fits as well as the more complex model. Our null hypothesis is proved and hence we can say that the models are equivalent.

5) Overdispersion

$$\phi = \frac{Residual\ Deviance}{Residual\ df} = \frac{1258}{1175} = 1.07$$

We can say that the model is not over-dispersed. Running the test on the data (qcc.overdispersion.test), we get a p value of 1, and observed variance/ theoretical variance = 0.47. The value of 0.47 is not statically significantly different from 1, HJence we can conclude that the model is not overdispersed.