

Indian Institute of Management Tiruchirappalli
PGP Term IV
BSF (2019) - Assignment 4 (Logistic Regression)

Submission Date: 12th August 2019 11:59 PM
Maximum Marks: 12

This is a group assignment. You need to use R programming for all data cleaning, manipulation, and analysis. Submit the soft copy of your assignment, R codes, and other supporting documents in the Google Classroom (No hard copy/email). Show adequate reasoning, including calculations, if any, in support of your answers. State any assumption you make (your assumptions need to be reasonable).

Mention your Group No (1 or 2) and team No (1 to 10) in your submission file-name (Assignment1_GR1_TEAM5.R). Also, mention your team details in one of your uploading files.

1. **[3 Marks]** *Compare Logit and Probit* Refer the article, 'Chapter 24 Logistic Regression'. You can find the data in the google classroom (graded individual assignment). Develop the model discussed in page 582 ($Y = \text{resp}$, $X_s = \text{Mortgage}$, $Famsize$) using probit and logit link. Compare the probabilities of 'take up the offer'? What can you find the similarity or difference in the model coefficients.
Hint: go through the help files ?glm ?family
2. **[3 Marks]** *Switching the response variable* Refer the article, 'Chapter 24 Logistic Regression'. You can find the data in the google classroom (graded individual assignment). Develop the model discussed in page 582 ($Y = \text{resp}$, $X_s = \text{Mortgage}$, $Famsize$) using response as 'take offer'. Develop the second model using response as 'decline offer'. Can I get the probability of 'take up the offer' from the second model? Compare the probabilities of 'take up the offer'? What can you find the similarity or difference in the model coefficients.
3. **[6 Marks]** Consider the German Credit Data from the following link.
[https://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))
Use the *first 5 attributes* and response (1 = Good, 2 = Bad) for this exercise.
 - (a) Perform univariate and bivariate analysis.
 - (b) Divide the data randomly into training (70%) and validation (30%) partitions, and develop a logistic regression model. Interpret the result.
 - (c) Report model validation output (lift and gain chart).