**Analytic Proposal Review #1**

For this assignment you are to write a proposal describing the analytic data product that you intend to develop. The purpose of writing this proposal is to help ensure that you develop an analytic that delivers the features desired by an end-user in the allotted time (by the end of the quarter).

The proposal will have two sections. For this review you are to complete the first section: Provide some basic information about the analytic you plan to develop. The information you provide in this section will be used both within your analytic and as part of the GitHub repository to aid collaboration.

This proposal will be included as a file in the `inst` directory of your package and will serve as the README file for your GitHub repository. The proposal must be pushed no later than midnight on 16 January 2018. Your grade on this proposal will be based on how closely it aligns with the actual final product (consideration will be given to the level of difficulty to complete the project). The reasoning behind this grading scheme is that it forces you prioritize what needs to be done in the allotted time and avoids the situation in which you ‘over-promise and under-deliver’.

**Section 1: Basic information about the analytic data product**

1. Provide a short name (either a single word or an acronym) for the analytic you plan to develop.

logreg

1. Provide a brief title (1-2 sentences) describing – at the 50,000-foot level – what your analytic does. Your title should be short and to the point, but should also be clear to an end user.

LOGREG performs logistic regression on a set of data stored in Excel.

1. Provide a description (2-3 paragraphs) of why this analytic data product will be useful for an end-user. This description should address the following points (where applicable).
   1. Describe each of the features that your analytic will perform when complete

LOGREG at its core will perform logistic regression on a set of data stored in Excel and output the beta coefficients for the modeled variables. Full functionality will be a bidirectional step-wise logistic regression selection model. Sub-functions will include the calculation of the G statistic to assist the user with selecting the next best variable to add to the model. Additionally, sub-functions will exist for discriminating the goodness of fit in the form of the chi-square p-value, the area under the receiver operating characteristic (ROC) curve value, model accuracy in the form of sensitivity and specificity, and the Hosmer-Lemeshow c-hat statistic.

* 1. Describe the typical end-user for whom this analytic is being developed

This tool is designed for an analyst already familiar with logistical regression who wishes to step-wise build models rather than accept a block box output from tools such as JMP.

* 1. Describe any specific knowledge/skills/abilities an end-user must have to use your analytic

Although this tool provides many statistics to help guild the user along to build a model, the analyst is solely responsible for understanding the outputs to interpret model integrity. A foundation in logistic regression is highly desirable.

* 1. If your analytic implements known statistical methods, specified them

In order to build the logistic model, a series of known statistics will be performed to include the logit transform, the conditional mean of the dependent variable, coding of the dependent variable, the likelihood function, deviance, and the G statistic.

* 1. If your analytic builds on existing statistical methods or R packages, specify them

It is intended that the statistical methods will be built from the ground up.

1. How will end-users access your analytic data product?

Initially, access will be conducted via command prompt in R. Ultimately, a GUI will assist the end-user through model building.

1. Are there any security concerns that need to be addressed?

No.

1. Are there any appearance/design constraints that your analytic must adhere to?

Data for the model must be formatted in Excel with the logit binary variable in column A. Row 1 will be variables names with data following below. There must not be any missing data.