# Home Work # 3. AMS 597

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# The quiz is due on Monday, April 6, 2020, at 12noon. Solutions will be emailed to you by 1pm. The homework will not be graded and all complete submission will warrant full points. Quiz 5 will be given on Tuesday, April 7, 2020, in class, covering the entire lecture; and will be graded. So please study hard from now on. Please email your completed homework to your TA at: [song.jiecheng@stonybrook.edu](mailto:song.jiecheng@stonybrook.edu)

# Please include (1) R code

# (2) Output from R

# (3) Answers to all the questions asked

#### Logistic Regression (\* A type of Generalized Linear Model) with the bwt Data

The bwt data is a built in R dataset containing 9 variables and 189 cases – please see the following link for detailed description:

<http://garthtarr.github.io/mplot/articles/birthweight.html>

Your task is to use the other 8 variables to predict the binary dependent variable ‘low’ that indicates whether a given infant was born with low birth weight or not.

You can follow the detailed description of related methods in the following paper:

<http://atm.amegroups.com/article/view/9706/pdf>

Please note that for the logistic regression model, one does not use the coefficient of determination for model selection, but rather, one should use the information criterion, typically either the AIC, or the BIC.

They are both functions of the likelihood albeit with the opposite signs from the likelihood:

<https://en.wikipedia.org/wiki/Bayesian_information_criterion>

<https://en.wikipedia.org/wiki/Akaike_information_criterion>

Since a larger likelihood corresponds to a better model, a smaller AIC or BIC will correspond to a better model. The BIC will usually results in a smaller model, hence more robust and preferred than the AIC.

1. Please find a model that best predicts whether the infant would be born with low birth weight or not using the stepwise variable selection method and the BIC. Please only use the original variables and no need to include any other variables such as interactions. Please report the final model and the associated BIC value.
2. Please find a model that best predicts whether the infant would be born with low birth weight or not using the stepwise variable selection method and the BIC. Please use the original variables plus all the two-way interactions. Please report the final model and the associated BIC value. (\*Please always remember that if an interaction term is found significant, then one must include both original variables even if they are not significant.)
3. Please find a model that best predicts whether the infant would be born with low birth weight or not using the best subset variable selection method and the BIC. Please only use the original variables and no need to include any other variables such as interactions. Please report the final model and the associated BIC value.
4. Please find a model that best predicts whether the infant would be born with low birth weight or not using the best subset variable selection method and the BIC. Please use the original variables plus all the two-way interactions. Please report the final model and the associated BIC value. (\*Please always remember that if an interaction term is found significant, then one must include both original variables even if one/both is/are not significant.)
5. Among all the models selected in steps 1, 2, 3, 4 above, which one is the best and why?