# Quiz 5. AMS 597

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

# The quiz is due at the end of the lecture by 11:20am – please submit no later than 11:30am. Please email your completed quiz 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

# Please keep yourself on Zoom video until you have emailed your solutions.

# Please plug your computer in power source to avoid running low on battery.

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

The banknote data (see attached) were extracted from images that were taken from genuine and forged banknote-like specimens. Yes, this is a ***Catch Me if You Can*** story. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Wavelet Transform tool were used to extract features from images. **There are 1372 banknotes, and 5 variables:**

1. variance of Wavelet Transformed image (continuous)   
2. skewness of Wavelet Transformed image (continuous)   
3. curtosis of Wavelet Transformed image (continuous)   
4. entropy of image (continuous)   
5. class (binary) – this is the response variable of interest, 0 (forged) or 1 (genuine).

1. Please find a model that best predicts whether the banknote is forged or genuine 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 banknote is forged or genuine 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 banknote is forged or genuine 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 banknote is forged or genuine 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?