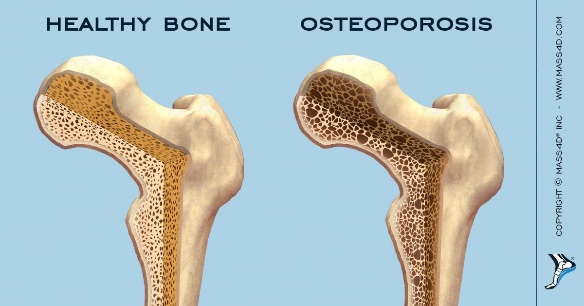
**Module Assignment**

**Module 8**

**QMB-6304 Analytical Methods for Business**



Write a simple R script to execute the following data preprocessing and statistical analysis. Where required show analytical output and interpretations.

**Preprocessing**

1. Load the file “Module 8 Assignment Data Set.xlsx” into R. This file contains information on 500 women age 55 and above and any recent bone fractures they have experienced. The data comes from the Global Longitudinal study of Osteoporosis in Women (GLOW) project. Variables included are:
   1. FRACTURE: A binary variable indicating whether the subject has recently experienced the fracture of any bone. (1=yes)
   2. PRIORFRAC: A binary variable indicating whether the subject has experienced a prior bone fracture. (1=yes)
   3. AGE: A numerical variable showing the age of the subject. Ages range from 55 to 90.
   4. WEIGHT: A numerical variable showing the weight of the subject in pounds.
   5. HEIGHT: A numerical variable showing the height of the subject in inches.
   6. BMI: The Body Mass Index of the subject.
   7. MENOBY45: A binary variable indicating whether the subject experienced early menopause, i.e., by age 45.
   8. MOMFRAC: A binary variable indicating whether the subject’s mother experienced at least one bone fracture.
   9. ARMASSIST: A binary variable indicating whether the subject needs to “push up” with her arms to rise from a chair.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 150 cases from the full data set using the method presented in class. This will be your reduced data set.

**Analysis (Using the Reduced Data Set)**

1. Parameterize a full logistic regression model with FRACTURE as the dependent and all other variables as independent (excluding the CASE variable).
2. Using the *summary()* command report the results of the model from Step 1.
3. State whether you believe the Residual Deviance of your model is markedly different from the Null Deviance.
4. Given your model from Part 1 and ignoring p values, which variable will have the greatest influence in increasing the modeled probability that a subject will have recently experienced a bone fracture?
5. Given your model from Part 1 and ignoring p values, which variable will have the greatest influence in decreasing the modeled probability that a subject will have recently experienced a bone fracture?
6. Using the *expand.grid()* command develop a prediction file with all independent variables in the Step 1 model. For binary independent variables use the *unique()* qualifier. For numerical (continuous) independent variables use the *quantile()* qualifier and set test levels at the 25th, 50th, 75th, and 100th percentiles for the variables as appearing in your reduced data set. Examples of these qualifiers can be found in the video PowerPoints. R will by default calculate predicted probabilities to many decimal places. For convenience in reporting round your stored predictions to only 3 decimal places. Show the predicted probabilities for ONLY the first five cases appearing in your prediction file.
7. Based on your predictions generated in Step 6, find the maximum and minimum predicted probabilities generated. Exclude predictor values for all other predictor variables except AGE, and WEIGHT.

Your deliverable will be a single MS-Word file created using R Markdown. Your file will show 1) the R script which executes the above instructions and 2) the results of those instructions. The first two lines of your deliverable will state this is “Assignment 5” of our course and your name as it appears in Canvas. Your code chunks and analysis results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.