HGDS 200 - Foundations of Data-Driven Analysis 2018/2019 - Handout 3

1/7/2019 - 1/14/2019

Name: Connor Voglewede

1. The coefficient for *IsMale* being larger than the coefficient for *LDL* is expected when *LDL* can take on values between 0 and infinity theoretically. *IsMale* can only take on values [0,1]. I’m not sure we can deduce which variable is more important just from the equation without seeing the data or model outputs. If we wanted the equation to tell us more between the *IsMale* and *LDL* variables, we could make *LDL* a binary classification where *LDL*=1 if LDL>160 mg/dL and 0 otherwise.
2. The female is the greater risk. The minimum predicted value is f(-1), or ~.27, is greater than the maximum predicted value the male could have, which is f(-3) or .05.
3. Again, I would create the dummy variable for high risk LDL where *LDL* could take on the value of 1 when LDL>160 mg/dL and 0 otherwise.