**Part B. Extra Credit**

**Methods and Results**

To assess reliability of the prediction model for infant birthweight developed from the training dataset in the previous section, a validation dataset containing the same variables but with different observations is imported and used. The validation dataset is processed in the same way as the training dataset by deleting missing values, centering numeric values on their mean, and creating higher order terms such as quadratic terms and relevant interaction terms. The prediction model is then used to predict birthweight in the validation dataset. A subsequent correlation analysis shows that the correlation between predicted birthweight and observed birthweight is 0.341 (p<.001), resulting in the R-squared value of 0.1163. This value is lower than the R-squared value of 0.1331 from the training dataset because the prediction model may have been overfit to a certain extent to the training dataset. Further analysis could include k-fold cross-validation for reduced bias and increased model reliability.

**Code**

**data** validate;

set library.nmihs88\_hold\_278;

if smoking = **0**;

junkid= **1**;

**run**;

**proc** **means** data = validate NMISS N;

var momage educ prenatalClass smoking parity wtprepreg BMIprepregCat wtGain gestWeeks male bwt;

**run**;

**data** validate;

set validate;

if parity ne **.** AND BMIprepregCat ne **.** AND bwt ne **.**;

**run**;

**proc** **means** data = validate NMISS N;

var momage educ prenatalClass smoking parity wtprepreg BMIprepregCat wtGain gestWeeks male bwt;

**run**;

**proc** **means** data = validate;

by junkid;

var momage wtprepreg wtgain gestweeks;

output out=sumstats mean=meanmomage meanwtprepreg meanwtgain meangestweeks;

**run**;

**data** validate;

merge sumstats validate;

by junkid;

momage = momage-meanmomage;

wtprepreg = wtprepreg - meanwtprepreg;

wtgain = wtgain - meanwtgain;

gestweeks = gestweeks - meangestweeks;

momage2 = momage\*\***2**;

wtprepreg2 = wtprepreg\*\***2**;

wtgain2 = wtgain\*\***2**;

gestweeks2 = gestweeks\*\***2**;

wtgain\_male = wtgain\*male;

momage\_bmiprepregcat = momage\*bmiprepregcat;

**run**;

**proc** **reg** data = b outest = regress;

model bwt = momAge wtGain gestWeeks momage2 gestweeks2 prenatalClass BMIprepregCat male wtgain\_male momage\_bmiprepregcat;

title "reg bwt";

**run**;

**proc** **score** data = validate score = regress out= new type=parms nostd predict;

var momAge wtGain gestWeeks momage2 gestweeks2 prenatalClass BMIprepregCat male wtgain\_male momage\_bmiprepregcat;

**run**;

**proc** **print** data=new;

var bwt model1;

label bwt = "Observed Birthweight" model = "Prediced Birthweight";

title "Predicted Scores for Regression";

**run**;

**proc** **corr** data = new;

var model1 bwt;

**run**;

**Output**

**Table

Description automatically generatedTable

Description automatically generated**