Prediction of Obesity Levels Based on Various Factors

Background on Obesity:

Obesity is a complex, multifactorial condition characterized by an excessive accumulation of body fat. It is typically assessed using body mass index (BMI), which is calculated as a person's weight in kilograms divided by the square of their height. Obesity has become a major global health concern due to its association with a range of serious health problems, including cardiovascular disease, type 2 diabetes, cancer, respiratory problems, joint disorders, and mental health. Preventing and managing obesity requires a comprehensive approach that includes promoting healthy eating habits and encouraging regular physical activity.

The Dataset Background:

The dataset "Estimation of Obesity Levels Based on Eating Habits and Physical Condition" is available on the public domain, UCI Machine Learning Repository. It contains data related to the estimation of obesity levels in individuals based on their eating habits and physical condition. The dataset was created by analyzing individuals from the National Health and Nutrition Examination Survey (NHANES) conducted between 2013 and 2014 in the United States. The dataset consists of 17 attributes, including demographic information (age, gender), measurements (height, weight), and lifestyle choices (eating habits, physical activity). The target variable is the obesity level, which is categorized into four classes based on the Body Mass Index (BMI) of the individuals: Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III. 77% of the data was generated through ML tools such as the Weka tool and the SMOTE filter, while 23% of the data was collected through a web platform. The dataset is structured and suitable for classification tasks, particularly for predicting the obesity level of individuals based on their characteristics and habits.

Previous Research Using Dataset:

The article from the MDPI (Multidisciplinary Digital Publishing Institute), "Estimation of Obesity Levels through the Proposed Predictive Approach Based on Physical Activity and Nutritional Habits," discusses the application of various machine learning algorithms (using statistical methods), including logistic regression, XG boosting, and random forests to predict obesity levels using this dataset. The study compares the performance of these algorithms in predicting obesity and provides insights into the factors influencing obesity levels. The researchers concluded that logistic regression provided the most accurate and precise results compared to the other methods and was the best way to predict obesity levels based on physical activity and eating habits.

My goal is to conduct a comprehensive evaluation of the logistic regression model using the dataset, delving deeper into its performance and characteristics. While the research paper primarily focused on identifying the most accurate statistical method for predicting obesity levels, my aim is to provide a detailed analysis of the logistic regression model specifically as it pertains to the actual prediction of obesity levels. By thoroughly examining the model's predictions, accuracy, and other relevant metrics, I intend to uncover insights that may not have been fully explored in the previous study as well as compare some of my findings such as the AUC performance.

Research Questions:

-How do eating habits and physical condition relate to the likelihood of being obese?

-Which specific eating habits or physical conditions are most strongly associated with different levels of obesity?

-Can we predict obesity levels based on a person's eating habits and physical condition?

-What are the significant factors for having higher odds of being at a higher obesity level?

-How well can a logistic regression model predict obesity levels based on the provided features?

-Can the logistic regression model be used to inform interventions or policies aimed at reducing obesity rates based on eating habits and physical condition?

Summary of the Data (Survey Questions Used):

Variable Name	Variable Type	Survey Question	Survey Answers
Gender	Categorical	What is your gender?	Male, Female
Age	Continuous	What is your age?	Numeric value
Height	Continuous	What is your height?	Numeric value in meters
Weight	Continuous	What is your weight?	Numeric value in kilograms
family_history_with_overweight	Binary	Has a family member suffered or suffers from being overweight?	Yes, No
FAVC	Binary	Do you eat high-caloric food frequently?	Yes, No
FCVC	Integer	Do you usually eat vegetables in your meals?	Never, Sometimes, Always
NCP	Continuous	How many main meals do you have daily?	Between 1 and 2, Three, More than three
CAEC	Categorical	Do you eat any food between meals?	No, Sometimes, Frequently, Always
SMOKE	Binary	Do you smoke?	Yes, No
CH2O	Continuous	How much water do you drink daily?	Less than a liter, Between 1 and 2 L, More than 2 L
SCC	Binary	Do you monitor the calories you eat daily?	Yes, No
FAF	Continuous	How often do you have physical activity?	I do not have, 1 or 2 days, 2 or 4 days, 4 or 5 days
TUE	Integer	How much time do you use technological devices such as cell phones, video games, television, computer, and others?	0–2 hours, 3–5 hours, More than 5 hour

CALC	Categorical	How often do you drink alcohol?	I do not drink, Sometimes, Frequently, Always
MTRANS	Categorical	Which transportation do you usually use?	Automobile, Motorbike, Bike, Public Transportation, Walking

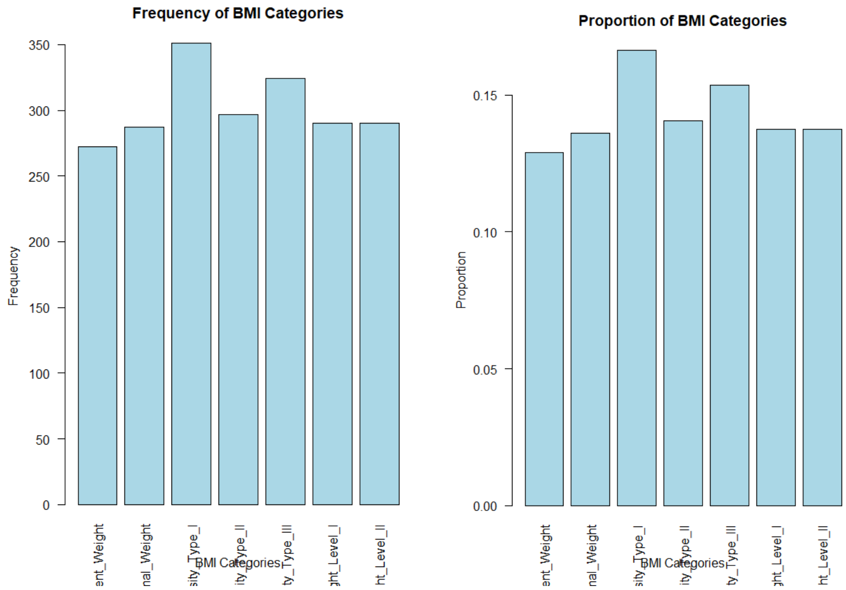
Target Variable (NObeyesdad, Categorical): Body Mass Index (BMI) of the individuals: Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III (Underweight Less than 18.5, Normal 18.5 to 24.9, Overweight 25.0 to 29.9, Obesity I 30.0 to 34.9, Obesity II 35.0 to 39.9, Obesity III Higher than 40)

Statistical Method:

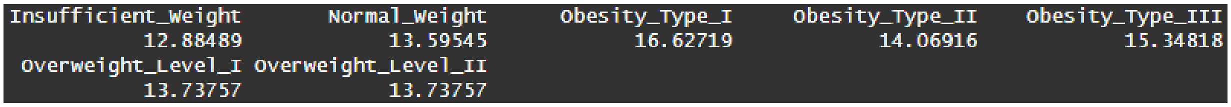
Multinomial Logistic Regression: This type of regression is used to predict the probabilities of the different possible outcomes of a categorical dependent variable, based on one or more independent variables. Multinomial LR helps understand how the independent variables influence the likelihood of each category of the dependent variable, providing insights into the relationships between the variables and aiding in prediction and classification tasks.

Exploratory Analyses:

Frequency and Proportion histograms of BMI Categories:



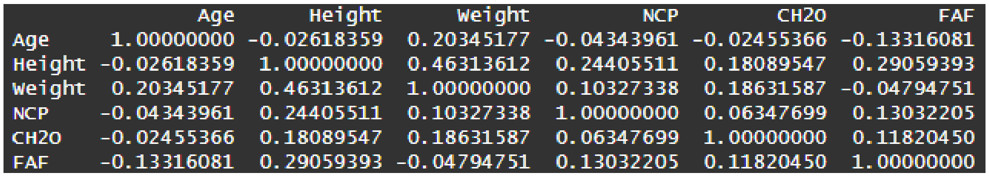
Percentages of each BMI category:



```
Insufficient_Weight       Normal_Weight       Obesity_Type_I      Obesity_Type_II     Obesity_Type_III
           12.88489            13.59545             16.62719             14.06916             15.34818
 Overweight_Level_I Overweight_Level_II
           13.73757            13.73757
```

BMI categories all look relatively equal based on the histograms and percentages. This suggests that the distribution of BMI categories is roughly uniform. In other words, each BMI category occurs with similar frequencies or proportions in the dataset. This balance in levels helps in achieving a more stable and reliable model, as each level of the target variable has a sufficient number of observations for the model to learn from.

Correlation Matrix of Continuous Variables:



```
               Age      Height      Weight         NCP        CH2O         FAF
Age     1.00000000 -0.02618359  0.20345177 -0.04343961 -0.02455366 -0.13316081
Height -0.02618359  1.00000000  0.46313612  0.24405511  0.18089547  0.29059393
Weight  0.20345177  0.46313612  1.00000000  0.10327338  0.18631587 -0.04794751
NCP    -0.04343961  0.24405511  0.10327338  1.00000000  0.06347699  0.13032205
CH2O   -0.02455366  0.18089547  0.18631587  0.06347699  1.00000000  0.11820450
FAF    -0.13316081  0.29059393 -0.04794751  0.13032205  0.11820450  1.00000000
```

The continuous variables do not seem to have much positive or negative correlation with each other. The most prevalent correlation would be the height and weight variables with a 0.463 positive correlation. Correlations can guide variable selection by identifying which variables are most strongly related to the outcome variable. Variables with low correlations may be less relevant and could be candidates for removal to simplify the mode. High correlations between independent variables (multicollinearity) can cause issues in the model, such as unstable coefficients and inflated standard errors. In this case, there is no real concern. It is expected that height and weight would be correlated as well.

Assumptions:

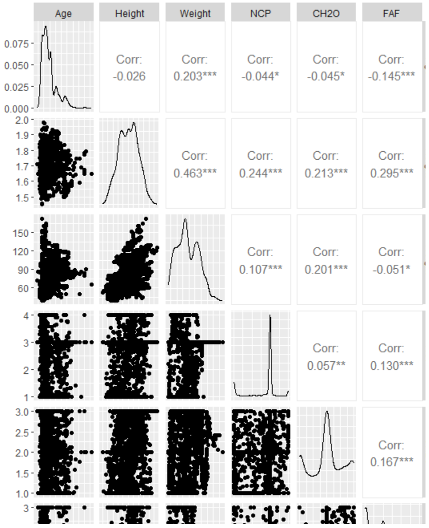
Independence of Observations: Correlation matrix didn't show strong correlation (only moderate between height and weight) among the continuous independent variables, which is a sign that the observations are independent of each other. The scatter plot also doesn't depict any strong relationships between the continuous independent variables.

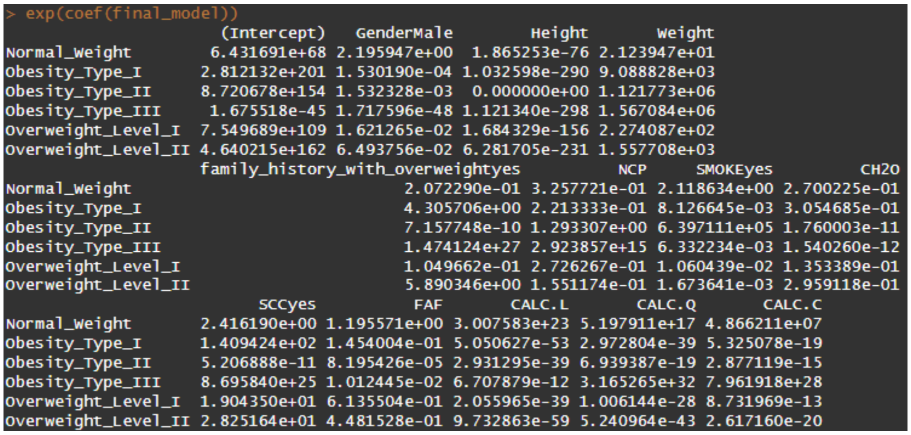
Multicollinearity: Correlation matrix also shows no multicollinearity because there are no pairwise interaction values higher than 0.70.

Linearity and Outliers: There also doesn't appear to be any outliers, linearity will be checked after modeling

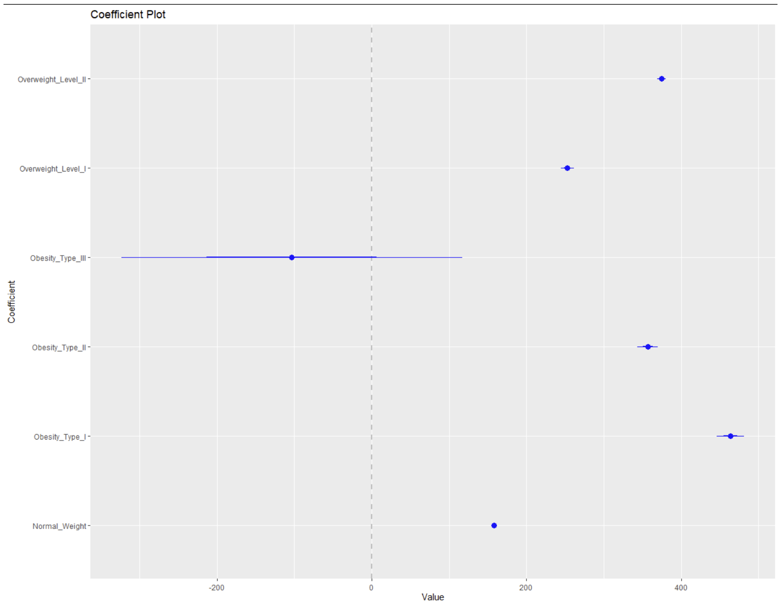
Multinomial Logistic Regression Model for Estimation of Obesity Level:

Original model had a higher AIC value than when using Stepwise AIC variable selection. A lower AIC value after using the stepwise AIC variable selection method suggests that the selected model (with fewer variables) provides a better balance between goodness of fit and model complexity compared to the original model. The variable selection chose 10 variables and excluded FAVC, TUE, CAEC, MTRANS, FCVC, and Age.





```
> exp(coef(final_model))
                      (Intercept)   GenderMale        Height       Weight
Normal_Weight        6.431691e+68 2.195947e+00  1.865253e-76 2.123947e+01
Obesity_Type_I      2.812132e+201 1.530190e-04 1.032598e-290 9.088828e+03
Obesity_Type_II     8.720678e+154 1.532328e-03  0.000000e+00 1.121773e+06
Obesity_Type_III     1.675518e-45 1.717596e-48 1.121340e-298 1.567084e+06
Overweight_Level_I  7.549689e+109 1.621265e-02 1.684329e-156 2.274087e+02
Overweight_Level_II 4.640215e+162 6.493756e-02 6.281705e-231 1.557708e+03
                    family_history_with_overweightyes          NCP     SMOKEyes         CH2O
Normal_Weight                            2.072290e-01 3.257721e-01 2.118634e+00 2.700225e-01
Obesity_Type_I                           4.305706e+00 2.213333e-01 8.126645e-03 3.054685e-01
Obesity_Type_II                          7.157748e-10 1.293307e+00 6.397111e+05 1.760003e-11
Obesity_Type_III                         1.474124e+27 2.923857e+15 6.332234e-03 1.540260e-12
Overweight_Level_I                       1.049662e-01 2.726267e-01 1.060439e-02 1.353389e-01
Overweight_Level_II                      5.890346e+00 1.551174e-01 1.673641e-03 2.959118e-01
                          SCCyes          FAF       CALC.L       CALC.Q       CALC.C
Normal_Weight       2.416190e+00 1.195571e+00 3.007583e+23 5.197911e+17 4.866211e+07
Obesity_Type_I      1.409424e+02 1.454004e-01 5.050627e-53 2.972804e-39 5.325078e-19
Obesity_Type_II     5.206888e-11 8.195426e-05 2.931295e-39 6.939387e-19 2.877119e-15
Obesity_Type_III    8.695840e+25 1.012445e-02 6.707879e-12 3.165265e+32 7.961918e+28
Overweight_Level_I  1.904350e+01 6.135504e-01 2.055965e-39 1.006144e-28 8.731969e-13
Overweight_Level_II 2.825164e+01 4.481528e-01 9.732863e-59 5.240964e-43 2.617160e-20
```



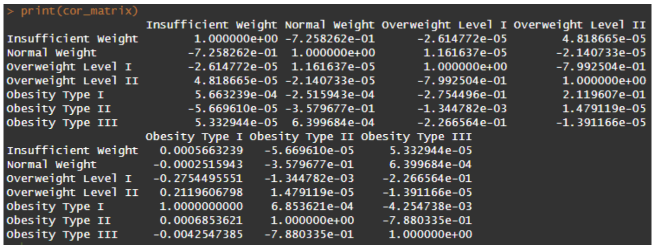
All the P-values for each coefficient (alpha = 0.05):

	(Intercept)	GenderMale	Height	Weight	family_history_with_overweightyes	NCP
Normal_Weight	0	0.5663352134	**0**	**0**	2.014771e-01	0.07196950
Obesity_Type_I	0	**0.0004203993**	**0**	**0**	5.362563e-01	0.13903357
Obesity_Type_II	0	0.1605553441	**0**	**0**	**2.758451e-05**	0.86269134
Obesity_Type_III	0	**0.0000000000**	**0**	**0**	**0.000000e+00**	**0.00000000**
Overweight_Level_I	0	**0.0194004537**	**0**	**0**	1.195885e-01	0.09632238
Overweight_Level_II	0	0.1467404529	**0**	**0**	2.813189e-01	0.03103997

	SMOKEyes	CH2O	SCCyes	FAF	CALC.L	CALC.Q	CALC.C
Normal_Weight	0.88128686	2.010947e-01	7.584732e-01	7.618779e-01	**0**	**0**	**1.758109e-05**
Obesity_Type_I	0.47986711	4.186315e-01	2.555628e-01	**4.172055e-02**	**0**	**0**	**0.000000e+00**
Obesity_Type_II	**0.01828608**	**0.000000e+00**	**3.205578e-10**	**3.002564e-09**	**0**	**0**	**2.662223e-07**
Obesity_Type_III	0.23637281	**8.622815e-08**	**0.000000e+00**	5.754031e-01	**0**	**0**	**0.000000e+00**
Overweight_Level_I	0.48828794	9.917454e-02	3.333415e-01	4.801353e-01	**0**	**0**	**7.118328e-11**
Overweight_Level_II	0.33764982	3.538720e-01	2.983486e-01	3.081822e-01	**0**	**0**	**0.000000e+00**

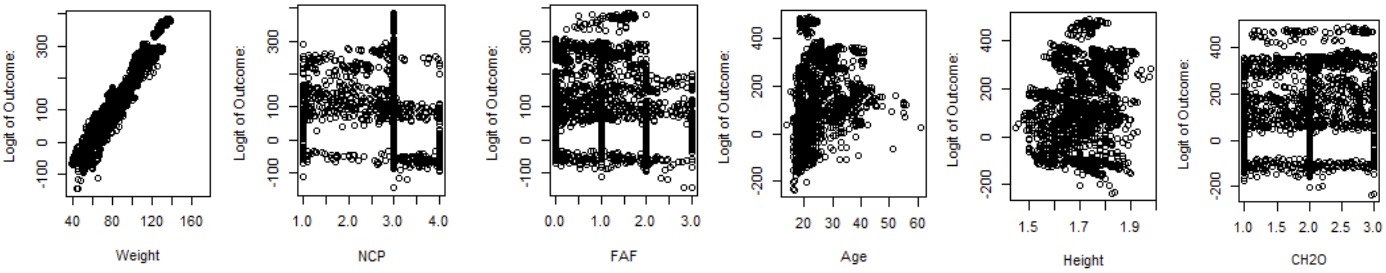
Assumptions re-check:

Independence: This correlation matrix shows the correlations between the residuals of the model for different levels of obesity. Most correlations are close to 0, indicating that the residuals for different levels of the dependent variable are largely independent. However, there are a few notable exceptions, such as the correlation of -0.358 between "Normal Weight" and "Obesity Type II," which suggests a moderate negative linear relationship between the residuals for these two levels. This might be of concern to some underlying relationship in the data, but not enough to violate the independence assumption.



```
> print(cor_matrix)
                    Insufficient Weight Normal Weight Overweight Level I Overweight Level II
Insufficient Weight        1.000000e+00 -7.258262e-01      -2.614772e-05        4.818665e-05
Normal Weight             -7.258262e-01  1.000000e+00       1.161637e-05       -2.140733e-05
Overweight Level I        -2.614772e-05  1.161637e-05       1.000000e+00       -7.992504e-01
Overweight Level II        4.818665e-05 -2.140733e-05      -7.992504e-01        1.000000e+00
Obesity Type I             5.663239e-04 -2.515943e-04      -2.754496e-01        2.119607e-01
Obesity Type II           -5.669610e-05 -3.579677e-01      -1.344782e-03        1.479119e-05
Obesity Type III           5.332944e-05  6.399684e-04      -2.266564e-01       -1.391166e-05
                    Obesity Type I Obesity Type II Obesity Type III
Insufficient Weight   0.0005663239   -5.669610e-05     5.332944e-05
Normal Weight        -0.0002515943   -3.579677e-01     6.399684e-04
Overweight Level I   -0.2754495551   -1.344782e-03    -2.266564e-01
Overweight Level II   0.2119606798    1.479119e-05    -1.391166e-05
Obesity Type I        1.0000000000    6.853621e-04    -4.254738e-03
Obesity Type II       0.0006853621    1.000000e+00    -7.880335e-01
Obesity Type III     -0.0042547385   -7.880335e-01     1.000000e+00
```

Linearity: Plots of logit outcomes for continuous variables

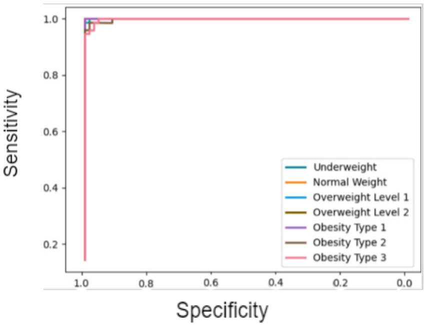


There does not seem to be linear relationships between the continuous independent variables and the logit of the outcome variable for each level of the outcome variable, except for weight and maybe age. However, due to the nature of the way the surveys were set up for the continuous independent variables, it was expected and is not very concerning considering the AUC curve has great predictive power (indicating a multinomial regression model is good for prediction). Non-linearity is not very severe for multinomial regression. While the linearity assumption is important, logistic regression can still perform reasonably well if the departure from linearity. Machine learning tools that generated the data also could have been a contributor to this nonlinearity as well and it could have chosen these values with this in mind, so this could be intended. This might conversely lead to overfitting and bias however.

Outliers and Multicollinearity: No outliers or multicollinearity issues as checked before.

AUC Curve:

AUC=0.98



Model seems to have very good predictive power.

Example Interpretation of Significant Coefficients:

Variables (Categorical)	The odds of being in the category of: (relative to the other categories)	are X times higher
Gendermale	Obesity Type I	1.717596e-48 (for men than for women, holding all other variables constant.)
Gendermale	Obesity Type III	6.493756e-02
Gendermale	Overweight Level II	1.532328e-03
family_history_with_overweightyes	Obesity Type II	1.049662e-01(for yes than for no)
family_history_with_overweightyes	Obesity Type III	5.890346e+00
SMOKEyes	Obesity Type II	1.060439e-02
SCCyes	Obesity Type II	1.904350e+01 (for yes monitoring cal than for no)
SCCyes	Obesity Type III	2.825164e+01
CALC.L	Each category	Check coefficients in exp(coef(final_model))
CALC.Q	Each category	Check coefficients in exp(coef(final_model))
CALC.C	Each category	Check coefficients in exp(coef(final_model))

For continuous variables (Height, Weight, NCP, CH2O, FAF), a one-unit increase in the variable is associated with the given increase/decrease in the odds of being in each category compared to the reference category. For example, for each unit increase in Height, the odds of being in the category of Obesity_Type_I vs. Normal_Weight decreases by a factor of 1.87 x 10^-76.

Conclusion:

-How do eating habits and physical condition relate to the likelihood of being obese?

Significant eating/drink habits such as NCP and CALC did contribute to the likelihood of being obese. For NCP (Number of Main Meals), having more main meals daily is associated with lower odds of being in the "Overweight Level I" category, but has very higher odds of being "Obesity Type I" category. For CALC (alcohol consumption), that higher alcohol consumption is associated with increased likelihood of being in higher obesity categories. For significant physical conditions such as FAF (Frequency of Physical Activity), having physical activity 1-2 days a week (compared to none) is associated with lower odds of being in the "Overweight Level I" category. Also, higher levels of physical activity are generally associated with lower odds of obesity across all categories. Overall, having more meals/more alcohol and having lower physical activity gives an individual higher odds to be in the obese categories than normal.

-Which specific eating habits or physical conditions are most strongly associated with different levels of obesity?

These findings suggest that specific eating habits (such as the number of main meals and consumption of food between meals), alcohol consumption, and physical activity are strongly associated with different levels of obesity. These factors can play a significant role in determining an individual's likelihood of being in a certain obesity category.

-Can we predict obesity levels based on a person's eating habits and physical condition?

We can predict obesity levels based on a person's eating habits and physical condition using the multinomial regression model. The model provides coefficients for each predictor variable, which can be used to calculate the probability of belonging to each obesity category based on the individual's eating habits and physical condition. By inputting the values of these variables into the model, we can estimate the probabilities of being in each obesity category and determine the most likely category for a given individual.

-What are the significant factors for having higher odds of being at a higher obesity level?

The significant factors are gender, family history with being overweight, smoking, monitoring calories, alcohol consumption, height, weight, daily meals, daily water intake, and physical activities.

-How well can a logistic regression model predict obesity levels based on the provided features?

Based on the provided multinomial regression model with an AUC of 0.98 and a great ROC curve, the model appears to have excellent predictive power within the dataset it was trained on. This is similar to the findings from the article "Estimation of Obesity Levels through the Proposed Predictive Approach Based on Physical Activity and Nutritional Habits" that was mentioned earlier.

-Can the logistic regression model be used to inform interventions or policies aimed at reducing obesity rates based on eating habits and physical condition?

The model shows that certain eating habits, such as frequent consumption of high-caloric food or eating between meals, are strongly associated with odds of being in higher levels of obesity. Policymakers could use this information to develop interventions that promote healthier eating habits, such as providing education on nutrition or implementing policies to increase access to healthy food options.

Future questions/ways to further analyze the problem:

While the model shows great predictive power within the dataset it was trained on, its generalizability to real-world data may be limited. It's crucial to validate the model on independent datasets and consider ways to mitigate overfitting, such as using regularization techniques or reducing the complexity of the model, to ensure its reliability in practical applications. Linearity could also be improved in the model through transformation which might fix the overfitting of the ML data and the way the survey was set up for continuous variables.

References:

Gozukara Bag, H.G., Yağın, F.H., Gormez, Y., González, P.P., Çolak, C., Gülü, M., Badicu, G., & Ardigò, L.P. (2023). Estimation of Obesity Levels through the Proposed Predictive Approach Based on Physical Activity and Nutritional Habits. Diagnostics, 13. https://doi.org/10.3390/diagnostics13182949

Palechor, F. M., & Manotas, A. H. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in brief, 25, 104344. https://doi.org/10.1016/j.dib.2019.104344

R Appendix:

```
##FINAL PROJECT STAT654 OBESITY LOGISTIC REGRESSION##
##Aaron PONGSUGREE##

#obtaining data set
setwd("C:/Users/Aaron/Documents/R Scripts/datasets")
obesity.data <- read.csv("ObesityDataSet_raw_and_data_sinthetic.csv")
View(obesity.data)

##data cleaning and adjusting
#changing categorical and integer variables to factors
str(obesity.data)
obesity.data$NObeyesdad = factor(obesity.data$NObeyesdad, ordered = TRUE)
obesity.data$Gender = factor(obesity.data$Gender)
obesity.data$family_history_with_overweight = factor(obesity.data$family_history_with_overweight)
obesity.data$FAVC = factor(obesity.data$FAVC)
obesity.data$SMOKE = factor(obesity.data$SMOKE)
obesity.data$SCC = factor(obesity.data$SCC)
obesity.data$MTRANS = factor(obesity.data$MTRANS)


#have to round the columns first because the ML algorithm gave outputs in between levels
#they are ordinal integers as well
obesity.data$FCVC = round(obesity.data$FCVC)
obesity.data$FCVC = factor(obesity.data$FCVC, levels = 1:3, labels = c("Never", "Sometimes", "Always"),
                           ordered = TRUE)
obesity.data$TUE = round(obesity.data$TUE)
obesity.data$TUE = factor(obesity.data$TUE, levels = 0:2, labels = c("0-2 hours", "3-5 hours", "More than 5
        hours"),
                          ordered = TRUE)
#rounding these based off the answers in the survey
obesity.data$Age = round(obesity.data$Age)
obesity.data$NCP = round(obesity.data$NCP)
obesity.data$CH2O = round(obesity.data$CH2O)
obesity.data$FAF = round(obesity.data$FAF)
#change some categorical factors to ordinal based on the answers in the survey
obesity.data$CAEC = factor(obesity.data$CAEC, levels = c("no", "Sometimes", "Frequently", "Always"),
                         ordered = TRUE)
obesity.data$CALC = factor(obesity.data$CALC, levels = c("no", "Sometimes", "Frequently", "Always"),
                         ordered = TRUE)

#summary statistics on obesity level outcomes
freq_table <- table(obesity.data$NObeyesdad)
print(freq_table)
options(repr.plot.width=8, repr.plot.height=4)
barplot(freq_table,
        main = "Frequency of BMI Categories",
        xlab = "BMI Categories",
        ylab = "Frequency",
        col = "lightblue",
        border = "black",
        las = 2)
prop_table <- prop.table(freq_table)
print(prop_table)
barplot(prop_table,
        main = "Proportion of BMI Categories",
        xlab = "BMI Categories",
        ylab = "Proportion",
        col = "lightblue",
        border = "black",
        las = 2)

summary(obesity.data$NObeyesdad)
table(obesity.data$NObeyesdad) / nrow(obesity.data) * 100

#correlation matrix of continuous variables
continuous_vars <- c("Age", "Height", "Weight", "NCP", "CH2O", "FAF")
correlation_matrix <- cor(obesity.data[, continuous_vars])
print(correlation_matrix)
correlation_matrix <- cor(obesity.data[, -which(names(obesity.data) == "NObeyesdad")])
print(correlation_matrix)

#checking assumptions
#independence
#scatterplot matrix for continuous variables
library(GGally)
ggpairs(obesity.data, columns = c("Age", "Height", "Weight", "NCP", "CH2O", "FAF"))
#multicollinearity




#multinomial logistic regression model
library(nnet)

mlog <- multinom(NObeyesdad ~ ., data = obesity.data)
summary(mlog)
library(car)
```

```
vif(mlog)


#AIC stepwise selection
library(MASS)
final_model <- stepAIC(mlog, direction = "both")
summary(final_model)
final_model
vif(final_model)
exp(coef(final_model))

# install.packages("nnet")
library(nnet)
confint(final_model)

str(summary(final_model))
summary(final_model)$standard.errors

#linearity
fitted_values <- predict(final_model, type = "probs")
par(mfrow=c(3,3))  # Set up a 3x3 grid of plots
for (i in 1:ncol(fitted_values)) {
  for (var in c("Age", "Height", "Weight", "NCP", "CH2O", "FAF")) {
    x <- obesity.data[[var]]
    y <- log(fitted_values[, i] / fitted_values[, 1])
    if (length(x) != length(y)) {
      cat("Lengths differ for", var, "and outcome level", levels(fitted_values)[i], "\n")
      cat("Length of x:", length(x), "\n")
      cat("Length of y:", length(y), "\n")
    } else {
      plot(x, y, xlab = var, ylab = paste("Logit of Outcome:", levels(fitted_values)[i]))
    }
  }
}

#independence
residuals <- residuals(final_model, type = "pearson")
cor_matrix <- cor(residuals)
print(cor_matrix)

# Create a coefficient plot
install.packages("coefplot")
library(coefplot)
coefplot(final_model, intercept = FALSE)p

#p-values for coefficients
z <- summary(final_model)$coefficients/summary(final_model)$standard.errors
 # 2-tailed Wald z tests to test significance of coefficients
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
```

```
#AUC
install.packages("pROC")
library(pROC)
pred <- predict(final_model, type = "probs")
roc_obj <- multiclass.roc(response = obesity.data$NObeyesdad, predictor = as.numeric(pred), levels =
        levels(obesity.data$NObeyesdad))
plot(roc_obj, print.auc = TRUE, print.auc.x = 0.5, print.auc.y = 0.3)
```