Proportional Odds Model On Estimation Obesity Level Based On Eating Habits and Physical Conditions

Introduction

The obesity epidemic has become one of the most serious public health problems globally. According to the research conducted by WHO in 2014, over 1900 million obese adults worldwide, which is double times compared to the number in 1980[1]. Previous research has shown that various social and behavioral factors can explain this rapid growth. Among them, unhealthy eating habits and lifestyle are two crucial risk factors for the development of obesity[3]. As the pandemic of COVID-19 enforced the practice of social-distancing and remote working, the altered eating behaviors and physical conditions such as longer screen time, intake of high caloric food, and reduced activity frequency may contribute to a higher risk of obesity. This analysis aims to conduct a statistical analysis based on previous study data on obesity estimation, using the proportional odds model to assess **whether elements discussed in the study significantly predict obesity levels** and **whether the association between those behaviors or physical conditions obesity differs by gender**.

Data Preprocessing

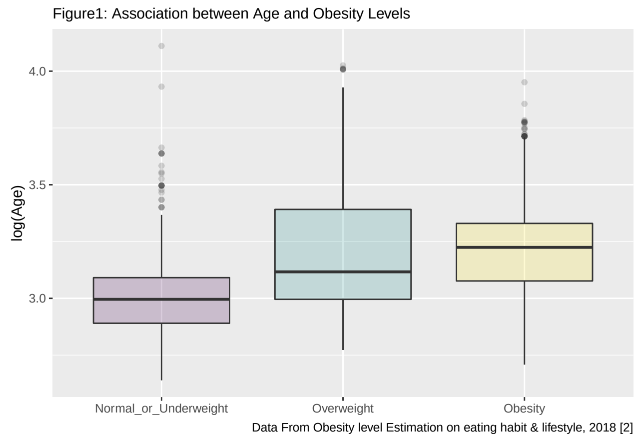
The dataset used in this analysis is collected by a 2018 study on obesity levels estimation based on anonymous survey results collected among individuals from Colombia, Peru, and Mexico [2]. The original data contains 2111 non-empty and 17 variables, including demographic information, eating habits, and physical conditions. Variables 'Time spent on technology devices' and 'water consumption' were deleted because there's a lack of support for matching numeric variables in the dataset to original survey answers. The original dataset contains 4 valid numeric predictors – age, vegetable consumption (in number of meals per day), number of main meals, and daily physical activity (days of a week) – and 8 valid categorical predictors including gender, family obesity history, consumption of high caloric food, between-meal snack eating, smoking, caloric monitoring habit, drinking behavior, and transportation methods.

Since most people adopted a three-meal diet with a combination of meat and vegetables, the distribution of vegetable consumption and the number of main meals are clustered at three. Therefore, two new binary variables, 'main_meal_bin' and 'vege_bin', were created based on levels of 'less_than_three' and 'three_meals_or_more'. Other measurements based on behavior frequency, such as smoking, alcohol consumption, and physical activity, are converted to dummy variables with level "yes" and "no." to ensure the sufficiency of interaction analysis. Individuals' obesity levels were calculated by the BMI function (BMI = Weight / (Height^2)) and categorized using the WHO standard for overweight and obesity [2]. There are 13 categorical predictors, 1 numeric predictor, and a response variable with 3 naturally ordered categories 'normal and underweight', 'overweight', and 'obesity.' in the cleaned dataset.

Exploratory Data Analysis

The cleaned data displays that a baseline probability of becoming overweight is approximately 27.5%, while the likelihood of reaching obesity measurement is 46% in the sample population. It is important to notice the baseline probability does not reveal a true population statistic because the number of individuals in each category is normalized in the original dataset to avoid skewed learning towards the majority class.

Figure 1 identifies the relationship between obesity levels and age for participants, indicating a potential positive association between age and obesity levels. Compared to the normal and overweight category, the increase in the average population age in the obesity category explains obesity as an age-related disease. The increasing risk that corresponds to aging may be explained by hormonal changes, decreases in metabolism, and a less active lifestyle. However, as the dataset's surveyed population is highly skewed towards teenagers and university students, the significance of aging in obesity development needs to be examined in the final model.



Besides aging, family obesity history is another demographic variable besides aging that contributes to greater possibilities of obesity, according to the conditional probability table given whether individuals have family members suffering from exceeded body weights. In the overall sample population, the chance of overweight and obese people with family obesity records is three times higher than those without such histories. Although family history is a potential risk factor in both female and male populations, females have significantly higher chances of obesity, while males tend to become only overweight.

Table 1: Conditional Probability of Obesity Levels Given Family History

	no	yes
Normal_or_Underweight	0.7221	0.1628
Overweight	0.2571	0.2787
Obesity	0.02078	0.5585

Further exploration of eating habits suggests that intake of energy-dense food, between-meal eating, alcohol consumption, and three or more vegetable consumption per day are predictors associated with a higher risk of obesity. While the previous two behaviors' effect appears to be the same regardless of demographic groups, there're potential interaction effects between drinking and gender, and vegetable consumption and gender. Tables 2 and 3 show the difference in the conditional probability of obesity and overweight among females and males given drinking habits, indicating that females are more likely to be influenced by alcohol consumption and become obese. For females with drinking habits, the likelihood of having exceeded body weight increases from 57% to 74%, with a doubled possibility of reaching obesity level, while this statistic remains almost unchanged for males in the same condition. Eating more than three meals per day is negatively correlated to the likelihood of being placed into the obesity end. However, the association is very tricky because people diagnosed as overweight or obese might skip main meals to control weight.

Table 2: Conditional probability of obesity levels among females, based on drinking behavior

	No	Yes
Normal_or_Underweight	0.2896	0.2019
Overweight	0.2418	0.3424
Obesity	0.4687	0.4557

Table 3: Conditional probability of obesity levels among males, based on drinking behavior

	No	Yes
Normal_or_Underweight	0.4178	0.253
Overweight	0.3191	0.2043
Obesity	0.2632	0.5426

Moving to physical conditions, physical exercise, and caloric monitoring habits seem to have a negative correlation with being diagnosed as overweight and obese. However, the chi-square test shows the changes in probability insignificant. The conditional probability table also indicates no significant differences in obesity level in the smoking group than the non-smoking group. Probing into the distribution table, we observed an insufficient number of observations who answer 'yes' on smoking, which may lead to an insignificant predictor in the predictive model. Surprisingly, adopting green transportation methods and getting obese has a positive association. Missing Confounding variables such as income level might contribute to this association. All predictors mentioned in the EDA process will be tested in the model selection process.

Model Selection

The model selection followed two methodologies: AIC Forward Stepwise Selection and ANOVA chi-square test. All main effect was fitted into the initial proportional odds model, and their significances were determined by 95% confidence intervals. Predictors between-meal food consumption, smoking, physical activity, and caloric monitoring were detected as poor predictors for obesity levels and deleted from the model. The interaction term between gender and other variables are added respectively for ANOVA testing using a 0.05 threshold. The results indicate interaction effects between gender and family history, alcohol consumption, and vege_bin significant in obesity level estimation. AIC and ANOVA chi-square test was conducted each time before adding or dropping a predictor from the model.

Final Model

$$log(\frac{Pr[y_i \leq j|x_i]}{Pr[y_i > j|x_i]}) = \beta_{0j} + \beta_1 x_{gender} * (\beta_2 x_{history} + \beta_3 x_{alcohol} + \beta_4 x_{vege\_bin}) + \beta_5 x_{log(age)} + \beta_6 x_{high\ caloric\ food} + \beta_7 x_{main\ meals\_bin} + \beta_8 x_{transportation}$$

Re-fitting to get Hessian

Call: polr(formula = obesity_ordered ~ Gender * (Family_History + Alcohol + vege_bin) + Age + High_Caloric_Food + green_trans + main_meal_bin, data = obesity)

Table 4: Coeficients

	Value	Std. Error	t value
GenderMale	2.866	0.3063	9.357
Family_Historyyes	3.62	0.2172	16.67
AlcoholYes	1.207	0.1649	7.32
vege_binthree_meals_or_more	1.088	0.1565	6.956
Age	0.1308	0.0104	12.58
High_Caloric_Foodyes	1.316	0.1542	8.534
green_transYes	1.328	0.1501	8.844
main_meal_binthree_meals_or_more	-0.4598	0.102	-4.507
GenderMale:Family_Historyyes	-1.723	0.2743	-6.279
GenderMale:AlcoholYes	-1.197	0.2126	-5.63
GenderMale:vege_binthree_meals_or_more	-1.859	0.2427	-7.657

Table 5: Intercepts

	Value	Std. Error	t value
Normal_or_Underweight|Overweight	7.939	0.4342	18.28
Overweight|Obesity	9.796	0.4559	21.49

The summary table suggested the significance of age, gender, family history, drinking, vegetable consumption, number of main meals, and the use of public transportation in predicting obesity levels. Here are some insights extracted from the final model that quantity the effect of each predictor. **Keeping every other variable in the baseline level and constant**:

1) A one-unit increase in age will result in a 14.5% (exp(0.13) = 1.145) increase in the odds of falling in obesity direction rather the normal direction

2) The estimated odds that individuals who self report as having high caloric food eating habit is in the obesity direction rather than the normal or underweight direction is 3.72 (exp(1.316) = 3.72) times the estimated odds for those who don't; while the estimated odds that individuals who eat three or more main meals per day is in the obesity direction rather than the normal or underweight direction decreases by 36.86%, compared to the estimated odds for those who eat less than three.

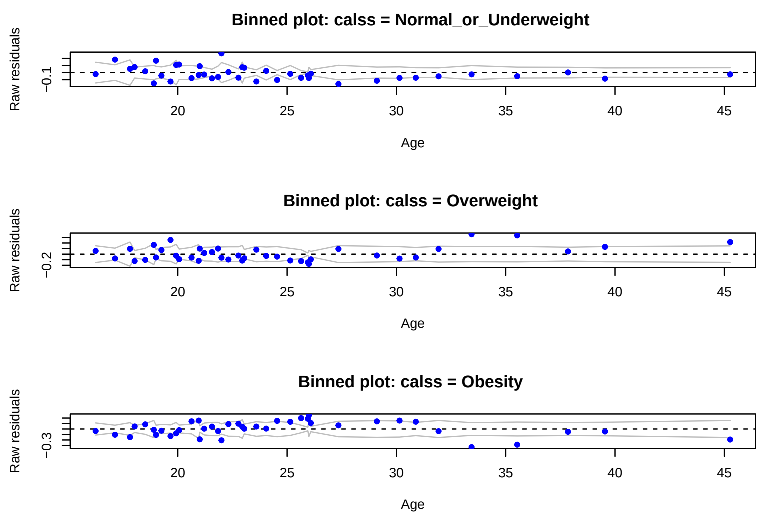
3) The odds of people who adopts green transportation such as bike and public transportation falling into the obesity direction rather than the normal or underweight direction is 3.76 ($\exp(1.328) = 3.76$) times the estimated odds of people who choose automobiles

4) The odds of males with family obesity records falling into the obesity direction rather than the normal or underweight direction is 6.6 ($\exp(3.62 - 1.723) = 6.66$) times the estimated odds of males without such histories.

5) The odds of males with drinking habits is in the obesity direction rather than the normal or underweight direction is increased by 0.3% ($\exp(1.2 - 1.197) = 0.003$) compared to the estimated odds of males without such habits, while the odds of males with drinking habits is in the obesity direction rather than the normal or underweight direction is 10 times ($\exp(1.2 + 1.197) = 10.9$) the estimated odds of females without such habits.

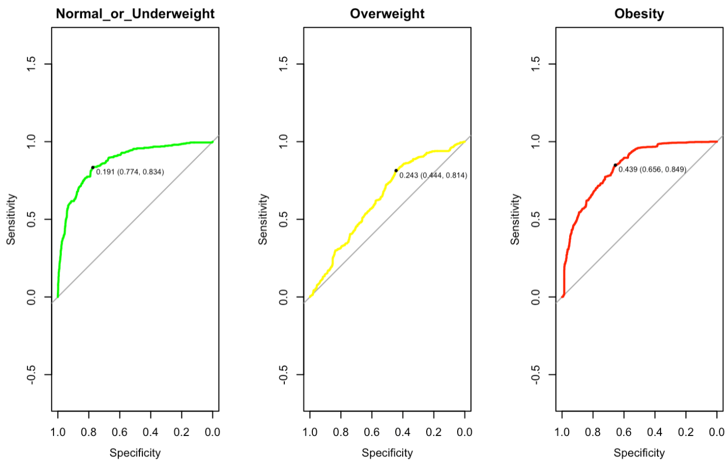
Model Accessment

For the model assessment of the final model, the raw (response) residuals for fitted proportional odds regression in each response variable level are calculated to explore abnormal patterns. In this analysis, age is the only numeric variable. Fitting in age and the raw residuals for each level, respectively, most points fall between the red lines representing a band expected 95% of the observations, regardless of a small group of outliers in the ‘obesity’ class. Therefore, the normality model assumption is not violated in the final model.



The final model’s overall AUC is 64%, indicating the final model has a fair amount of diagnostic ability for predicting obesity levels. Looking into each group respectively, the model is more effectively predicting the ‘normal_or_Underweight’ and ‘Obesity’ class with approximately 75% accuracy while performs much poorly on predicting ‘Overweight’ with only 52.8% accuracy.

According to the ROC curves, at optimal decision threshold, sensitivity for the ‘normal_or_underweight’ class is 0.774, and specificity is 0.834. Sensitivity on the y-axis measures the true positive rate, and among all ‘normal’ cases the final model classified, 77.4% percent of them are correct. Meanwhile, specificity on the x-axis refers that the model returns’ overweight’ or ‘obesity’ for 83.4% non-normal cases the final model reported. The ROC curve for the ‘obesity’ class also shows a fair sensitivity of 0.65 and a high specificity of 0.84. However, the ‘overweight’ class’s sensitivity is only 0.44, indicating a lower predictive power than the previous two.



Conclusion

This analysis on the dataset collected by a 2018 study for obesity level estimation using habits and physical conditions provides some evidence-based conclusions on the impact of age, family history, certain eating habits, and lifestyles on obesity level estimation. In terms of demographic variables, the likelihood of becoming overweight and obese increases as age increases. Family obesity history is also an important predictor as individuals who have a family history of obesity have a much higher chance of being diagnosed as overweight and obese, especially females.

As for eating behavior, the odds of leaning towards the obesity direction rather than the normal direction tripled if individuals have a high-caloric food eating habit while decrease by 36.86% if individuals eat three or more meals per day. As for lifestyle, alcohol consumption and transportation methods are the two most significant predictors. While the likelihood of placing in obesity and overweight increases by only a little for males with drinking habits, the odds of falling into obesity direction rather than normal direction for females who drink are ten times the odds of females who don't. Therefore, females are much more easily influenced by alcohol consumption in terms of obesity risk. Finally, using green transportation methods is positively associated with a higher likelihood of leaning towards the obesity direction. Income level, education, community environment, and other factors linked to the transportation method might contribute to this association.

Limitation

Primarily, the dataset used for analysis is based on anonymous and volunteer survey results. Therefore, skewed response and volunteer-based bias can exist, especially for demographic and behavior related questions such as weight and alcohol consumption. Moreover, the cause of obesity is complex and usually requires a combination of various risk factors. The statistical analysis is conducted without controlling factors related to mental health, economic conditions, and social environment. It is difficult to draw casual conclusions between predictors, such as the number of main meals and transportation method and obesity development because there're many confounding factors such as obesity treatment and income level. Further discussions and observations of those missing elements should provide a clear picture of the impact of eating habits and physical conditions on obesity levels.

Reference

[1]De-La-Hoz-Correa, E., Mendoza Palechor, F., De-La-Hoz-Manotas, A., Morales Ortega, R., & SÃ¡nchez HernÃ¡ndez, A. B. (2019). Obesity level estimation software based on decision trees.

[2]Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in Brief, 104344.

[3]Kuźbicka K, Rachoń D. Bad eating habits as the main cause of obesity among children. Pediatr Endocrinol Diabetes Metab. 2013;19(3):106-10. PMID: 25577898.

Appendix

```r
library(ggplot2)
library(MASS)
library(caret)
library(dplyr)
library(car)
library(arm)
library(pROC)
library(e1071)
library(nnet)
library(knitr)
library(pander)
knitr::opts_chunk$set(echo = F)
knitr::opts_chunk$set(fig.align="center")
obesity <- read.csv('ObesityDataSet_raw_and_data_sinthetic.csv',
                   col.names = c('Gender', "Age", "Height","Weight","Family_History",
                                 "High_Caloric_Food","Vegetable","Main_Meals","Between_Meals",
                                "SMOKE","Water","Monitoring","Physical_Activity","Time_On_Technology_De
                                 "Transportation","Obesity"))
# Categorical variables: gender, family history of obesity, frequent consumption of high caloric food, 

obesity$Gender <- as.factor(obesity$Gender)
obesity$Family_History <- as.factor(obesity$Family_History)
obesity$High_Caloric_Food <- as.factor(obesity$High_Caloric_Food) # frequent consumption of high calori
obesity$Between_Meals <- as.factor(obesity$Between_Meals)
obesity$SMOKE <- as.factor(obesity$SMOKE)
obesity$Monitoring <- as.factor(obesity$Monitoring)
obesity$Alcohol <- as.factor(obesity$Alcohol)
obesity$Transportation <- as.factor(obesity$Transportation)

# Response variables
obesity$Obesity <- as.factor(obesity$Obesity)

# Add BMI for EDA
obesity <- obesity %>% mutate( bmi = Weight / Height^2)

# Combine Levels for Alcohol Consumption, between_meals
obesity <- obesity %>% mutate(
        Alcohol = case_when(
                Alcohol == "Sometimes" ~ 'Yes',
                Alcohol == "Frequently" ~ 'Yes',
                Alcohol == "Always" ~ 'Yes',
                Alcohol == "no" ~ 'No'))

obesity = obesity %>% mutate(
        Between_Meals = case_when(
```

```r
                Between_Meals == "Sometimes" ~ 'Yes',
                Between_Meals == "Frequently" ~ 'Yes',
                Between_Meals == "Always" ~ 'Yes',
                Between_Meals == "no" ~ 'No'))

obesity = obesity %>% mutate(
        green_trans = case_when(
                Transportation == "Automobile" ~ 'No',
                Transportation == "Motorbike" ~ 'Yes',
                Transportation == "Walking" ~ 'Yes',
                Transportation == "Public_Transportation" ~ 'Yes',
                Transportation == "Bike" ~ 'Yes'))

# Combine Levels for obesity
obesity = obesity %>% mutate(
        obesity_new = case_when(
                Obesity == "Insufficient_Weight" ~ 'Normal_or_Underweight',
                Obesity == "Normal_Weight" ~ "Normal_or_Underweight",
                Obesity == "Overweight_Level_I" ~ "Overweight",
                Obesity == "Overweight_Level_II" ~"Overweight",
                Obesity == "Obesity_Type_I"~"Obesity",
                Obesity == "Obesity_Type_II"~"Obesity",
                Obesity == "Obesity_Type_III"~"Obesity"))

# Ordered Level
obesity$obesity_ordered <- ordered(obesity$obesity_new,
                           levels=c("Normal_or_Underweight","Overweight",
                                    "Obesity"))

# Drop Time_on_technology
obesity = obesity[,-14]

# Make categorical variables factors
obesity$Family_History = as.factor(obesity$Family_History)
obesity$High_Caloric_Food = as.factor(obesity$High_Caloric_Food)
obesity$Between_Meals = as.factor(obesity$Between_Meals)
obesity$SMOKE = as.factor(obesity$SMOKE)
obesity$Monitoring = as.factor(obesity$Monitoring)
obesity$Alcohol = as.factor(obesity$Alcohol)
obesity$Transportation = as.factor(obesity$Transportation)
obesity$obesity_ordered = as.factor(obesity$obesity_ordered)

#
obesity = obesity %>% mutate(
        main_meal_bin = case_when(
                Main_Meals < 3 ~ "less_than_three",
                Main_Meals >= 3 ~ "three_meals_or_more"
              ))

obesity = obesity %>% mutate(
        vege_bin= case_when(
                Vegetable < 3 ~ "less_than_three",
                Vegetable == 3 ~ "three_meals_or_more",
```

```
              ))
# Response variable
table(obesity$obesity_ordered)
p = ggplot(obesity, aes(x=obesity_ordered)) +
  geom_bar()
p + theme(axis.text.x = element_text(angle = 90)) + labs(title = 'Obesity Level Distribution')
# The possibility of getting obesity increases as the age increases
p1 = ggplot(data = obesity, aes(x=obesity_ordered, y=log(Age), fill = obesity_ordered)) +
  geom_boxplot(alpha=0.2)
p1 + labs(title = "Figure1: Association between Age and Obesity Levels",
          caption = "Data From Obesity level Estimation on eating habit & lifestyle, 2018 [2]")+
theme(
plot.title = element_text(size=11), legend.position="none", axis.title.x = element_blank(), )
pander(prop.table(table(obesity$obesity_ordered, obesity$Family_History), 2), caption = "Conditional Pro
men = obesity[obesity$Gender == 'Male',]
female = obesity[obesity$Gender == 'Female',]
pander(prop.table(table(men$obesity_ordered, men$Alcohol), 2), caption = "Conditional probability of obe
pander(prop.table(table(female$obesity_ordered, female$Alcohol), 2), caption = "Conditional probability
# Convert to Categorical Variable?
ggplot(data = obesity, aes(x=obesity_ordered, y=Vegetable, fill = obesity_ordered)) +
  geom_boxplot(alpha=0.2)  +  theme(legend.position="none")

# Convert to Categorical Variable? Association not casual effect
ggplot(data = obesity, aes(x=obesity_ordered, y=Main_Meals, fill = obesity_ordered)) +
  geom_boxplot(alpha=0.2)  +  theme(legend.position="none")
hist(obesity$Main_Meals)
summary(obesity$Main_Meals)
# Gender, females are more chance to have Normal_or_Underweight
table(obesity$obesity_ordered, obesity$Gender)
prop.table(table(obesity$obesity_ordered, obesity$Gender), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Gender))

# People with family_history of obesity are more likely to get obesity
table(obesity$obesity_ordered, obesity$Family_History)
prop.table(table(obesity$obesity_ordered, obesity$Family_History), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Family_History))

# Only few observations without high caloric food consumption
# yes means more towards obesity end
table(obesity$obesity_ordered, obesity$High_Caloric_Food)
prop.table(table(obesity$obesity_ordered, obesity$High_Caloric_Food), 2)
chisq.test(table(obesity$obesity_ordered, obesity$High_Caloric_Food))

# few observations without between meal consumption, not significant trend
table(obesity$obesity_ordered, obesity$Between_Meals)
prop.table(table(obesity$obesity_ordered, obesity$Between_Meals), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Between_Meals))

# No significant trend
table(obesity$obesity_ordered, obesity$SMOKE)
prop.table(table(obesity$obesity_ordered, obesity$SMOKE), 2)
chisq.test(table(obesity$obesity_ordered, obesity$SMOKE))
```

```r
# people with caloric monitoring habit has much lower chance for obesity
table(obesity$obesity_ordered, obesity$Monitoring)
prop.table(table(obesity$obesity_ordered, obesity$Monitoring), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Monitoring))

# drinking behavior
table(obesity$obesity_ordered, obesity$Alcohol)
prop.table(table(obesity$obesity_ordered, obesity$Alcohol), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Alcohol))

# transportation
table(obesity$obesity_ordered, obesity$Transportation)
prop.table(table(obesity$obesity_ordered, obesity$Transportation), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Transportation))

# main_meals
table(obesity$obesity_ordered, obesity$main_meal_bin)
prop.table(table(obesity$obesity_ordered, obesity$main_meal_bin), 2)
chisq.test(table(obesity$obesity_ordered, obesity$main_meal_bin))

# vege
table(obesity$obesity_ordered, obesity$vege_bin)
prop.table(table(obesity$obesity_ordered, obesity$vege_bin), 2)
chisq.test(table(obesity$obesity_ordered, obesity$vege_bin))

# physical activity
table(obesity$obesity_ordered, obesity$Physical_Activity)
prop.table(table(obesity$obesity_ordered, obesity$Physical_Activity), 2)
chisq.test(table(obesity$obesity_ordered, obesity$Physical_Activity))

# transportation
table(obesity$green_trans)
prop.table(table(obesity$obesity_ordered, obesity$green_trans), 2)
chisq.test(table(obesity$obesity_ordered, obesity$green_trans))
# Male with alcohol drinking habit has the same probability to have the extreme obesity end
# while alcohol consumption has a huge influence on female
prop.table(table(men$obesity_ordered, men$Family_History), 2)
prop.table(table(female$obesity_ordered, female$Family_History), 2)

prop.table(table(men$obesity_ordered, men$Alcohol), 2)
prop.table(table(female$obesity_ordered, female$Alcohol), 2)
# Interaction with Gender

# no change in association
ggplot(obesity,aes(x=Age, y= obesity_ordered, fill=obesity_ordered)) +
        geom_boxplot(alpha=0.2)  +  theme(legend.position="none") +
        facet_wrap( ~ Gender)

# Female are more easily influenced by family history
prop.table(table(men$obesity_ordered, men$Family_History), 2)
prop.table(table(female$obesity_ordered, female$Family_History), 2)

# men with more vege consumption tend to end up the normal direction
```

```r
# while female with more vege consumption tend to end up the normal direction
prop.table(table(men$obesity_ordered, men$vege_bin), 2)
prop.table(table(female$obesity_ordered, female$vege_bin), 2)

model_full = polr(obesity_ordered ~ Gender + Age + Family_History + High_Caloric_Food + main_meal_bin +
AIC(model_full)
coef(model_full)
confint(model_full) # Between_Meals, Smoke

# Delete Smoke, smoke is not a significant predictor
model1 = polr(obesity_ordered ~ Gender + Age + Family_History + High_Caloric_Food + main_meal_bin + vege
anova(model_full,model1)

# Delete Monitor, Monitor is a significant predictor
model2 = polr(obesity_ordered ~ Gender + Age + Family_History + High_Caloric_Food + main_meal_bin + vege
anova(model1,model2)

# Delete between_meals, between_meals is not a significant predictor
model3 = polr(obesity_ordered ~ Gender + Age + Family_History + High_Caloric_Food + main_meal_bin + vege
anova(model1,model3)
AIC(model3)
# Interaction with Gender and Family History, significant
model4 = polr(obesity_ordered ~ Gender * Family_History + Age + High_Caloric_Food + Physical_Activity +
confint(model4)
AIC(model4)
anova(model4, model3)

# Interaction with Gender and Alcohol, significant
model5 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol) + Age + High_Caloric_Food + Physica
confint(model5)
AIC(model5)
anova(model4, model5)

#interaction between high_caloric and monitor, significant
model7 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol) + Age + High_Caloric_Food * Monitor
AIC(model7)
anova(model5, model7)

#interaction between main_meal and monitor, not significant
model8 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol) + Age + Monitoring * (main_meal_bin
anova(model8, model7)

#interaction between vege and monitor, significant
model9 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol) + Age + Monitoring * (vege_bin + Hi
anova(model7, model9)

#interaction between gender and High_Caloric_Food, not significant
model10 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol + High_Caloric_Food) + Age + High_C
anova(model10, model9)

# interaction between gender and vege_consumption
model11 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol + vege_bin) + Age + High_Caloric_Fo
anova(model5, model11)
```

```r
model12 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol + vege_bin + Physical_Activity) + Ag
anova(model12, model11)

model13 = polr(obesity_ordered ~ Gender * (Family_History + Alcohol + vege_bin) + Age + High_Caloric_Fo
anova(model13, model11)
final_model = polr(obesity_ordered ~ Gender * (Family_History + Alcohol + vege_bin) + Age + High_Calori
pander(summary(final_model))
# confint(final_model)

# Residual and binplot
predprobs <- fitted(final_model)

rawresid1 <- (obesity$obesity_ordered == "Normal_or_Underweight") -  predprobs[,1]
rawresid2 <- (obesity$obesity_ordered == "Overweight") -  predprobs[,2]
rawresid3 <- (obesity$obesity_ordered == "Obesity") -  predprobs[,3]

par(mfrow=c(3,1))
# residual for age
binnedplot(obesity$Age, rawresid1, xlab = "Age", ylab = "Raw residuals", main = "Binned plot: calss = No
binnedplot(obesity$Age, rawresid2, xlab = "Age", ylab = "Raw residuals", main = "Binned plot: calss = Ov
binnedplot(obesity$Age, rawresid3, xlab = "Age", ylab = "Raw residuals", main = "Binned plot: calss = Ob
pred <- predict(final_model)
Conf_mat <- confusionMatrix(as.factor(pred),as.factor(obesity$obesity_ordered))
pander(Conf_mat$byClass[,c("Balanced Accuracy","Sensitivity", "Specificity")])
par(mfrow=c(1,3))
roc((obesity$obesity_ordered == 'Normal_or_Underweight'),predprobs[,1],plot=T,print.thres="best",main="N
roc((obesity$obesity_ordered == 'Overweight'),predprobs[,2],plot=T,print.thres="best",main="Overweight"
roc((obesity$obesity_ordered == 'Obesity'),predprobs[,3],plot=T,print.thres="best",main="Obesity", col =
```