proAdeotiE_555

2024-09-17

Obesity

Abstract

Discussions about obesity can be an emotional topic, depending on the audience, but some of the reasons why it happens vary and sometimes could be dependent on the individual. The data-set that was selected focused more on the estimation of obesity levels that might stem from eating habits and variables like physical conditions.
The data-set includes 17 attributes and 2111 records, some of the variable names were changed to be self-explanatory, variable names like NObesity (Obesity Level), to Obes_level, which allows data classifications from 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 gathered synthetically from countries in South America (Mexico, Peru, and Colombia) and 23% of the data was collected from web users' input. Some variable observations would be altered to be binary to help apply classical inferences and classification methods. The data-set picked was used for a different purpose like generating computational tools to identify and recommend systems that help individuals with obesity. This paper will discuss the role genetics and lifestyle plays in obesity or its lack thereof.

```
##  [1] "Gender"                         "Age"
##  [3] "Height"                         "Weight"
##  [5] "family_history_with_overweight" "FAVC"
##  [7] "FCVC"                           "NCP"
##  [9] "CAEC"                           "SMOKE"
## [11] "CH2O"                           "SCC"
## [13] "FAF"                            "TUE"
## [15] "CALC"                           "MTRANS"
## [17] "NObeyesdad"
```

```
##     Gender               Age           Height         Weight      
##  Length:2111        Min.   :14.00   Min.   :1.450   Min.   : 39.00  
##  Class :character   1st Qu.:19.95   1st Qu.:1.630   1st Qu.: 65.47  
##  Mode  :character   Median :22.78   Median :1.700   Median : 83.00  
##                     Mean   :24.31   Mean   :1.702   Mean   : 86.59  
##                     3rd Qu.:26.00   3rd Qu.:1.768   3rd Qu.:107.43  
##                     Max.   :61.00   Max.   :1.980   Max.   :173.00  
##                                                                     
##       CH2O            FAF              NCP             FCVC      
##  Min.   :1.000   Min.   :0.0000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:1.585   1st Qu.:0.1245   1st Qu.:2.659   1st Qu.:2.000  
##  Median :2.000   Median :1.0000   Median :3.000   Median :2.386  
##  Mean   :2.008   Mean   :1.0103   Mean   :2.686   Mean   :2.419  
##  3rd Qu.:2.477   3rd Qu.:1.6667   3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :3.000   Max.   :3.0000   Max.   :4.000   Max.   :3.000  
##                                                                  
##       TUE             FAVC              SMOKE               SCC           
##  Min.   :0.0000   Length:2111        Length:2111        Length:2111       
##  1st Qu.:0.0000   Class :character   Class :character   Class :character  
##  Median :0.6253   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :0.6579                                                           
##  3rd Qu.:1.0000                                                           
##  Max.   :2.0000                                                           
##                                                                           
##  family_history_with_overweight     CAEC               CALC          
##  no : 385                       Length:2111        Length:2111       
##  yes:1726                       Class :character   Class :character  
##                                 Mode  :character   Mode  :character  
##                                                                      
##                                                                      
##                                                                      
##                                                                      
##     MTRANS                          NObeyesdad       obes.3class 
##  Length:2111        Insufficient_Weight:272   Norm_weight:559  
##  Class :character   Normal_Weight      :287   Obesity    :972  
##  Mode  :character   Obesity_Type_I     :351   Overweight :580  
##                     Obesity_Type_II    :297                    
##                     Obesity_Type_III   :324                    
##                     Overweight_Level_I :290                    
##                     Overweight_Level_II:290                    
```

```
##   Gender Age Height Weight CH2O FAF NCP FCVC TUE FAVC SMOKE SCC
## 1 Female  21   1.62   64.0    2   0   3    2   1   no    no  no
## 2 Female  21   1.52   56.0    3   3   3    3   0   no   yes yes
## 3   Male  23   1.80   77.0    2   2   3    2   1   no    no  no
## 4   Male  27   1.80   87.0    2   2   3    3   0   no    no  no
## 5   Male  22   1.78   89.8    2   0   1    2   0   no    no  no
## 6   Male  29   1.62   53.0    2   0   3    2   0  yes    no  no
##   family_history_with_overweight      CAEC        CALC                MTRANS
## 1                            yes Sometimes          no Public_Transportation
## 2                            yes Sometimes   Sometimes Public_Transportation
## 3                            yes Sometimes Frequently Public_Transportation
## 4                             no Sometimes Frequently               Walking
## 5                             no Sometimes   Sometimes Public_Transportation
## 6                             no Sometimes   Sometimes            Automobile
##            NObeyesdad obes.3class
## 1       Normal_Weight Norm_weight
## 2       Normal_Weight Norm_weight
## 3       Normal_Weight Norm_weight
## 4  Overweight_Level_I  Overweight
## 5 Overweight_Level_II  Overweight
## 6       Normal_Weight Norm_weight
```

Introduction

The role of genetics in obesity might go under the radar most of the time, with food habits or unhealthy lifestyle taking precedence of why people might be obese.
My questions centers on the Part genetics have to play in Obesity, regardless of eating habits.? And why certain eating habits do not always result in obesity. Certain types of people tend to have unhealthy eating habits and their propensity to gain weight appears to be on the low side, this is in comparison to some people that have average eating habits, but they are more susceptible to be obese if they do not pay enough attention to their calorie intake often. Coming across multiple scenarios like this has always piqued my interest in the why and the how. Finding a dataset that covers some part of my questions to an extent was why I decided to work on this topic. Having some sought of answer to why this might happen would be personable, considering the high rise in obesity level, especially in the U.S and around central America.

Background

The truth is that there are multiple factors that contribute to obesity, factors like pervasive overeating, which is one of the accepted explanations for the primary cause of obesity. Nonetheless, contributing factors also include higher calorie intake from foods containing processed carbohydrate, which is widely affordable, but if estimated energy expenditure is on the lower side per day, our body has retained massive calorie intake with minimum output.
Another contributing factor is the type of diet that is consumed, the previously mentioned factor is somewhat associated with this. The type of diet that is consumed on a daily basis is important; “First, it is increasingly clear that the types and quality of foods consumed interact with the composition and health of our gut microbiota to influence digestive efficiency and flux (including the location, rate, and/or completeness of nutrient digestion), relative (host compared with microbiome) nutrient utilization, host metabolic expenditure, and host adipocyte function”. Generally, summarizes the need for a quality consumed diet for the gut microbiome and the host metabolic interactions to increase energy expenditure and could potentially reduce obesity risks.
The final factor this paper would focus on is the genetic role in obesity. “The obesity epidemic may be driven by

intergenerational influences. Potential pathways include maternal-to-infant transmission of microbiome species (and thereby health risk); in utero epigenetic changes caused by maternal stress, obesity, and poor diet; and inter- or transgenerational transmission of sperm or oocyte noncoding RNAs (ncRNAs). As successive generations become more obese, risk may be transmitted to the next generation that increases their susceptibility independently of energy intake. The composition of the microbiome, clearly linked to risk of obesity, is transmitted from one generation to the next. Both dietary changes and diet-induced microbial metabolites can also induce epigenetic changes that influence risk of weight gain and obesity." The cellular biological specifications that contribute to obesity were spoken of to an extent. Obesity can be transgenerational, due to the contributions of our DNA, RNA synthesis in cohorts with unregulated poor diet quality from parents which have physiologically influenced how the gut microbiome interacts with food.

Data Description

The Data was generated from countries of Mexico, Peru, and Colombia, based on their eating habits and physical condition. The Introductory paper and data formation was done by By Fabio Mendoza Palechor, Alexis De la Hoz Manotas in 2019 (https://doi.org/10.24432/C5H31Z (https://doi.org/10.24432/C5H31Z)). The data was collected through a web platform which accounted for 23% of the dataset, then an estimation, which was synthetically done accounted for the other 77% of the dataset.
Variable names, types, and descriptions

- Gender, categorical, sexuality at birth
- Age, Continuous
- Height, Continuous (km)
- Weight, Continuous (Kg)
- Family_history_with_overweight, Binary, has a family member suffered or suffers from overweight?
- FAVC, Binary, do you eat high caloric food frequently?
- FCVC, Integer, do you usually eat vegetables in your meals?
- NCP, continuous, How many main meals do you have daily?
- CAEC Categorical, do you eat any food between meals?
- Smoke, Binary, do you smoke?
- CH2O, Continuous, how much water do you drink daily?
- SCC, Binary, do you monitor the calories you eat daily?
- FAF, Continuous, how often do you have physical activity?
- TUE, Integer, how much time do you use technological devices such as cell phone, video games, television, computer, and others
- CALC, Categorical, how often do you drink alcohol?
- Mtrans, Categorical, Which transportation do you usually use?
- NObeyesdad, Categorical, Obesity level.

In all the Variables listed, there were no missing values accounted for. There would not be a need to omit values, or cleaning data. We define BMI categories as following.

BMI < 18.5 is underweight. BMI ranges from 18.5 -> 24.99 is normal. BMI ranges from 23 -> 24.99 is overweight. BMI > 25 is obese.

Numerical and visualization techniques

```
##       Age    Height    Weight      CH2O       FAF       NCP      FCVC       TUE 
## 20.787192  1.683656 56.195030  1.860444  1.248722  2.824182  2.405679  0.755515
```

```
##         Age      Height      Weight        CH2O         FAF         NCP 
##  25.8061810   1.7155531 109.0823438   2.0726387   0.8748878   2.7167863 
##        FCVC         TUE 
##   2.5200769   0.6033382
```

```
##        Age     Height     Weight       CH2O        FAF        NCP       FCVC 
## 25.2073275  1.6957921 78.1760495  2.0419291  1.0074341  2.4998740  2.2626047 
##        TUE 
##  0.6551335
```

```
##                Age        Height     Weight       CH2O         FAF         NCP
## Age    17.73227170 -0.0435444754  7.2308397 0.13483255 -0.08597280 -0.16036528
## Height -0.04354448  0.0094587704  0.5961974 0.01425466  0.03160026  0.03715147
## Weight  7.23083970  0.5961973905 99.3864590 1.10084473  1.86564871  1.45313798
## CH2O    0.13483255  0.0142546628  1.1008447 0.38517458  0.14560947  0.07731381
## FAF    -0.08597280  0.0316002589  1.8656487 0.14560947  0.88583880  0.26396494
## NCP    -0.16036528  0.0371514733  1.4531380 0.07731381  0.26396494  0.79183154
## FCVC    0.32461245 -0.0004948708 -0.2340257 0.05484371  0.01035260 -0.05168979
## TUE    -0.53110831  0.0081392284 -0.3067033 0.02133192  0.01308718  0.04232907
##                 FCVC          TUE
## Age     0.3246124456 -0.531108314
## Height -0.0004948708  0.008139228
## Weight -0.2340256711 -0.306703266
## CH2O    0.0548437123  0.021331915
## FAF     0.0103526016  0.013087175
## NCP    -0.0516897944  0.042329069
## FCVC    0.3504583830 -0.103282623
## TUE    -0.1032826228  0.449323574
```

```
##                Age       Height      Weight        CH2O          FAF
## Age     35.0343244 -0.100912973 -29.3493621 -0.57752448 -0.479348690
## Height  -0.1009130  0.007920136   0.9882766  0.01317496  0.027507574
## Weight -29.3493621  0.988276591 301.2133565  2.20695446  2.905828984
## CH2O    -0.5775245  0.013174964   2.2069545  0.37352647  0.052410628
## FAF     -0.4793487  0.027507574   2.9058290  0.05241063  0.592294264
## NCP     -0.2413621  0.012854347   4.3082203  0.02619084  0.004365296
## FCVC    -0.3861985 -0.004908411   3.5126357  0.01464070 -0.019113841
## TUE     -1.1443386  0.006748749   0.8946143  0.02989340  0.055948586
##                 NCP         FCVC          TUE
## Age    -0.241362074 -0.386198532 -1.144338597
## Height  0.012854347 -0.004908411  0.006748749
## Weight  4.308220323  3.512635735  0.894614324
## CH2O    0.026190842  0.014640703  0.029893401
## FAF     0.004365296 -0.019113841  0.055948586
## NCP     0.383116051  0.034122206 -0.043878857
## FCVC    0.034122206  0.262663670  0.017017515
## TUE    -0.043878857  0.017017515  0.298239088
```

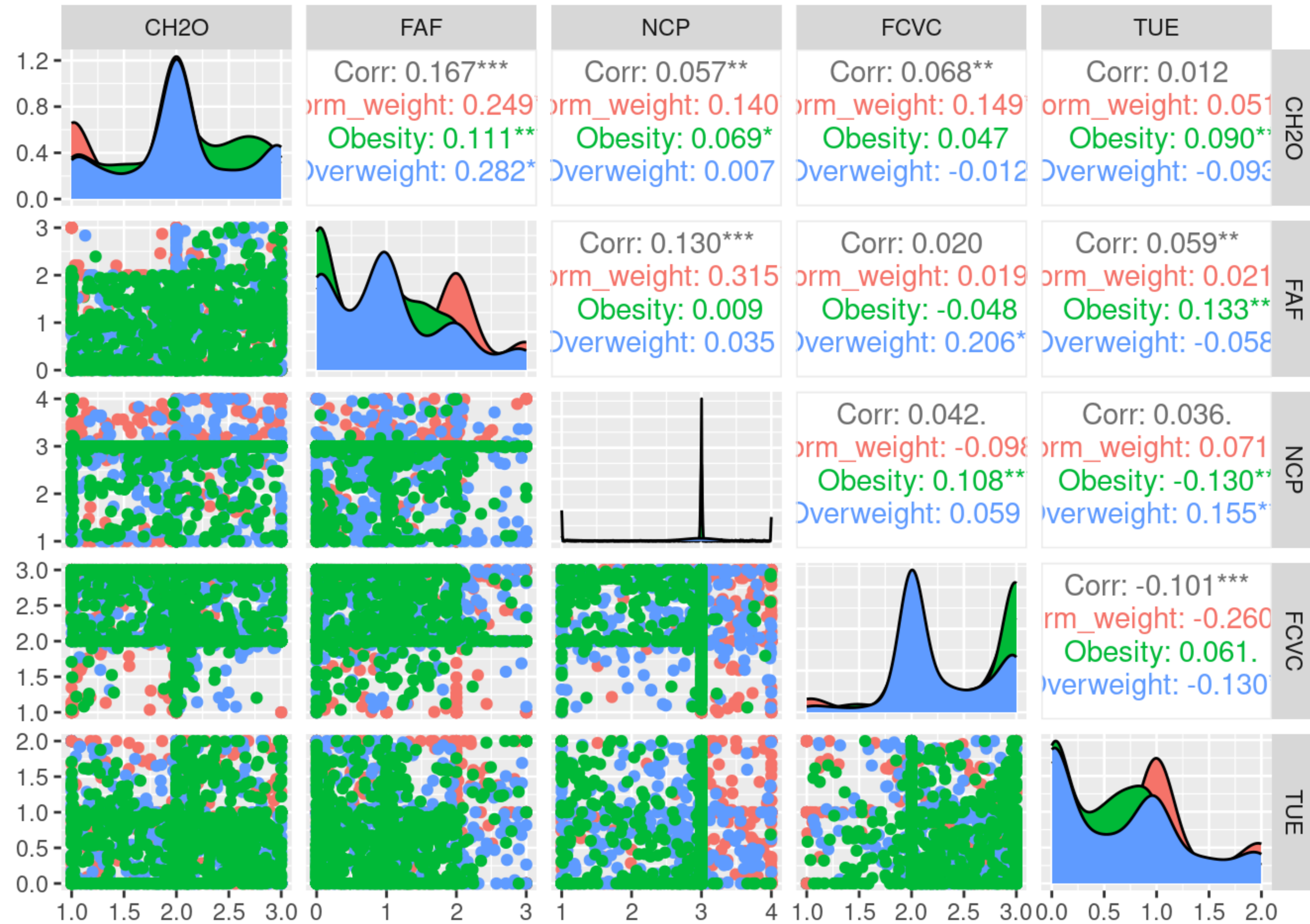
```
##               Age       Height     Weight         CH2O         FAF          NCP
## Age    54.3685139  0.064333505 12.6331032 -0.498315013 -0.81072766 0.328589747
## Height  0.0643335  0.008656293  0.7901878  0.004766021  0.01593644 0.007692166
## Weight 12.6331032  0.790187769 86.7723400  0.487679392  1.23837154 0.679032161
## CH2O   -0.4983150  0.004766021  0.4876794  0.342367536  0.13828301 0.003453002
## FAF    -0.8107277  0.015936438  1.2383715  0.138283006  0.70393913 0.025272605
## NCP     0.3285897  0.007692166  0.6790322  0.003453002  0.02527260 0.745666535
## FCVC    0.3773773 -0.001720584 -0.1769305 -0.003307214  0.08067234 0.023894462
## TUE    -1.2789384 -0.005460588 -0.4624396 -0.034494531 -0.03078387 0.084790412
##                FCVC          TUE
## Age     0.377377321 -1.278938360
## Height -0.001720584 -0.005460588
## Weight -0.176930514 -0.462439573
## CH2O   -0.003307214 -0.034494531
## FAF     0.080672343 -0.030783867
## NCP     0.023894462  0.084790412
## FCVC    0.218817515 -0.038748170
## TUE    -0.038748170  0.403859523
```

```
##
## Norm_weight     Obesity  Overweight
##         559         559         580
```



```
##
## Norm_weight     Obesity  Overweight
##         559         559         580
```

```
##                          Age Height Weight  CH2O   FAF   NCP  FCVC   TUE
## Age                     1.00   0.00   0.25 -0.02 -0.15 -0.03  0.03 -0.28
## Height                  0.00   1.00   0.46  0.20  0.29  0.24 -0.03  0.04
## Weight                  0.25   0.46   1.00  0.19 -0.05  0.06  0.15 -0.08
## CH2O                   -0.02   0.20   0.19  1.00  0.19  0.05  0.06 -0.01
## FAF                    -0.15   0.29  -0.05  0.19  1.00  0.15  0.05  0.05
## NCP                    -0.03   0.24   0.06  0.05  0.15  1.00  0.03  0.06
## FCVC                    0.03  -0.03   0.15  0.06  0.05  0.03  1.00 -0.14
## TUE                    -0.28   0.04  -0.08 -0.01  0.05  0.06 -0.14  1.00
## obes.3classNorm_weight -0.35  -0.11  -0.70 -0.15  0.16  0.13  0.03  0.09
## obes.3classObesity      0.20   0.14   0.78  0.10 -0.13  0.02  0.14 -0.07
## obes.3classOverweight   0.14  -0.02  -0.08  0.06 -0.03 -0.15 -0.16 -0.02
##                        obes.3classNorm_weight obes.3classObesity
## Age                                     -0.35               0.20
## Height                                  -0.11               0.14
## Weight                                  -0.70               0.78
## CH2O                                    -0.15               0.10
## FAF                                      0.16              -0.13
## NCP                                      0.13               0.02
## FCVC                                     0.03               0.14
## TUE                                      0.09              -0.07
## obes.3classNorm_weight                   1.00              -0.49
## obes.3classObesity                      -0.49               1.00
## obes.3classOverweight                   -0.50              -0.50
##                        obes.3classOverweight
## Age                                     0.14
## Height                                 -0.02
## Weight                                 -0.08
## CH2O                                    0.06
## FAF                                    -0.03
## NCP                                    -0.15
## FCVC                                   -0.16
## TUE                                    -0.02
## obes.3classNorm_weight                 -0.50
## obes.3classObesity                     -0.50
## obes.3classOverweight                   1.00
```



Methods & Research plan

The data set picked has multivariate characteristics, it was highlighted as part of the data description. We will try using classical interference methods, like MANOVA and Hotelling's to test if multiple response variables differ significantly across groups and compare the means generated for the normal distributions, respectively. The sample size might help with avoiding some limitations, to interpret the data accurately, calculated covariance matrices must not be unreliable.

This data provides multiple response variables and a classification variable that allows for classification of this data. We would test the relationships between the variables using logistic regression. We would try multiple models to help with linearity assumption of the response and predictors and compare the best performing models using ANOVA between both (regular non-interactive relationships and variables with interactions). The model's limitations would also be visualized, this would help us determine the outliers and interpretation issues. The Mass body index values would be recalculated to fit the weight level and used to classify obesity levels which would be used as the response for some binary and continuous variables.

The obesity levels would be useful for the classification methods; we would apply and evaluate the performance of the methods using cross-validation Ordered Logistic Regression, classification Tree and random forest This CV would provide accuracy and compare the limitations of the classifications mentioned. With all the methods that would be used already mentioned, each method has specific contexts where it excels and situations where its limitations require alternative approaches. The computations of this method would determine which of them helps us describe and better explains research directions.

```
##   Gender      Age   Height    Weight     CH2O      FAF      NCP     FCVC
## 1   Male 22.27697 1.849950 121.78648 1.555534 0.348839 2.272214 3.000000
## 2   Male 30.00203 1.759324 112.00038 2.003563 0.000000 3.000000 1.572036
## 3   Male 24.18489 1.768834  97.44974 2.973729 2.491642 3.000000 2.000000
## 4   Male 26.94779 1.647807  99.59222 1.000000 1.089891 1.845858 2.935157
## 5 Female 22.87522 1.624367  82.00000 2.000000 0.000000 1.000000 1.826885
## 6 Female 19.52894 1.817917 142.55916 2.562002 1.976427 3.000000 3.000000
##        TUE FAVC SMOKE SCC family_history_with_overweight      CAEC      CALC
## 1 0.000000    1     0   0                              1 Sometimes Sometimes
## 2 0.340196    1     0   0                              1 Sometimes Sometimes
## 3 1.365950    1     0   0                              1 Sometimes        no
## 4 0.715993    1     0   0                              1 Sometimes        no
## 5 0.459274    1     0   0                              1 Sometimes        no
## 6 0.740331    1     0   0                              1 Sometimes Sometimes
##                  MTRANS       NObeyesdad obes.3class  bmi
## 1 Public_Transportation  Obesity_Type_II     Obesity 35.6
## 2            Automobile  Obesity_Type_II     Obesity 36.2
## 3 Public_Transportation   Obesity_Type_I     Obesity 31.1
## 4 Public_Transportation  Obesity_Type_II     Obesity 36.7
## 5 Public_Transportation   Obesity_Type_I     Obesity 31.1
## 6 Public_Transportation Obesity_Type_III     Obesity 43.1
```

```
## Call:
## polr(formula = obes.3class ~ family_history_with_overweight +
##     Age + SCC + FAF + CH2O + FAVC + FCVC + NCP + SMOKE, data = b.obeslevel,
##     subset = train, Hess = TRUE)
##
## Coefficients:
##                                   Value Std. Error t value
## family_history_with_overweight  1.75010    0.17887  9.7845
## Age                             0.07019    0.01008  6.9639
## SCC                             0.37566    0.32610  1.1520
## FAF                            -0.18106    0.07289 -2.4838
## CH2O                            0.42769    0.09840  4.3465
## FAVC                           -0.06336    0.19302 -0.3282
## FCVC                           -0.36306    0.11017 -3.2954
## NCP                            -0.40192    0.07773 -5.1705
## SMOKE                          -0.55589    0.38873 -1.4300
##
## Intercepts:
##                     Value   Std. Error t value
## Norm_weight|Obesity  0.9913  0.4671     2.1222
## Obesity|Overweight   2.7387  0.4719     5.8035
##
## Residual Deviance: 2207.357
## AIC: 2229.357
```

```
##                                    Value Std. Error    t value      p value
## family_history_with_overweight  1.75010401 0.17886558  9.7844652 1.312875e-22
## Age                             0.07019023 0.01007922  6.9638563 3.310824e-12
## SCC                             0.37566163 0.32609733  1.1519923 2.493242e-01
## FAF                            -0.18105938 0.07289472 -2.4838477 1.299714e-02
## CH2O                            0.42769029 0.09839822  4.3465245 1.383116e-05
## FAVC                           -0.06335523 0.19301545 -0.3282392 7.427308e-01
## FCVC                           -0.36306341 0.11017235 -3.2954131 9.827709e-04
## NCP                            -0.40192148 0.07773351 -5.1705049 2.334623e-07
## SMOKE                          -0.55588935 0.38873381 -1.4300000 1.527170e-01
## Norm_weight|Obesity             0.99127504 0.46709397  2.1222176 3.381947e-02
## Obesity|Overweight              2.73874037 0.47190992  5.8035237 6.493558e-09
```

The Pvalues listed are all statistical significant against the null hypothesis. `The Alternative Hypothesis` states that the listed predictors in the formula used, does affect the probability of a person being in different BMI classes.

we would convert the coefficient values into odd ratios for easier interpretation to get the OR interpretations together with the confidence intervals through exponentiation of their earlier scaled log values. CI would be calculated using the standard errors, also assuming a normal distribution. If the 95% CI does not cross 0, the parameter estimate is statistically significant.

```
##                                      2.5 %      97.5 %
## family_history_with_overweight  1.39953393  2.10067410
## Age                             0.05043532  0.08994513
## SCC                            -0.26347740  1.01480065
## FAF                            -0.32393041 -0.03818835
## CH2O                            0.23483332  0.62054727
## FAVC                           -0.44165855  0.31494809
## FCVC                           -0.57899726 -0.14712957
## NCP                            -0.55427635 -0.24956660
## SMOKE                          -1.31779361  0.20601491
```

```
## family_history_with_overweight                            Age
##                      5.7552013                      1.0727122
##                            SCC                            FAF
##                      1.4559544                      0.8343858
##                           CH2O                           FAVC
##                      1.5337110                      0.9386100
##                           FCVC                            NCP
##                      0.6955423                      0.6690333
##                          SMOKE
##                      0.5735619
```

```
##                                      OR     2.5 %    97.5 %
## family_history_with_overweight 5.7552013 4.0533104 8.1716766
## Age                            1.0727122 1.0517288 1.0941143
## SCC                            1.4559544 0.7683750 2.7588134
## FAF                            0.8343858 0.7233006 0.9625316
## CH2O                           1.5337110 1.2646979 1.8599456
## FAVC                           0.9386100 0.6429691 1.3701882
## FCVC                           0.6955423 0.5604601 0.8631821
## NCP                            0.6690333 0.5744878 0.7791384
## SMOKE                          0.5735619 0.2677254 1.2287715
```

Interpreting odds ratio

For people with positive obesity in their family history, the odds of having BMI ranges from 23 - 24.99 (overweight)and BMI > 25 (obese) is 5.24 times likely to occur than people with negative obesity in family history, holding all other variables constant.

(*) For students in private school, the odds of being more likely to apply is 1.06 times [i.e., 1/0.943] that of public school students, holding constant all other variables (positive odds ratio).

Proportional odds

We can use the values in this table to help us assess whether the proportional odds assumption is reasonable for our model. The null hypothesis here is that (to determine if)the sets of coefficients are the same; in other words that (to determine if)the predictor variables predictions is the same from “normal weight” to “over-weight” and over-weight” to “obesity”.

```
##
## Tests for Proportional Odds
## polr(formula = obes.3class ~ family_history_with_overweight +
##     Age + SCC + FAF + CH2O + FAVC + FCVC + NCP + SMOKE, data = b.obeslevel,
##     subset = train, Hess = TRUE)
##
##                                  b[polr] b[>Norm_weight] b[>Obesity] Chisquare
## Overall                                                                150.84
## family_history_with_overweight    1.7501          2.2324      0.5630    287.02
## Age                               0.0702          0.1883      0.0447     47.85
## SCC                               0.3757         -0.0544      0.7904   -109.51
## FAF                              -0.1811         -0.3590     -0.0610     11.48
## CH2O                              0.4277          0.7626      0.2921     14.98
## FAVC                             -0.0634          0.6299     -0.5228     90.81
## FCVC                             -0.3631         -0.0280     -0.6575     25.33
## NCP                              -0.4019         -0.4428     -0.3411      1.36
## SMOKE                            -0.5559         -0.8244     -0.6400      0.11
##                                df Pr(>Chisq)
## Overall                         9    < 2e-16 ***
## family_history_with_overweight  1    < 2e-16 ***
## Age                             1    4.6e-12 ***
## SCC                             1    1.00000
## FAF                             1    0.00070 ***
## CH2O                            1    0.00011 ***
## FAVC                            1    < 2e-16 ***
## FCVC                            1    4.8e-07 ***
## NCP                             1    0.24288
## SMOKE                           1    0.74047
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  Length   Class    Mode
##       3 formula    call
```

```
## Error in `[.default`(logodds.obes, , 4): incorrect number of dimensions
```

```
## Error in `[.default`(logodds.obes, , 3): incorrect number of dimensions
```

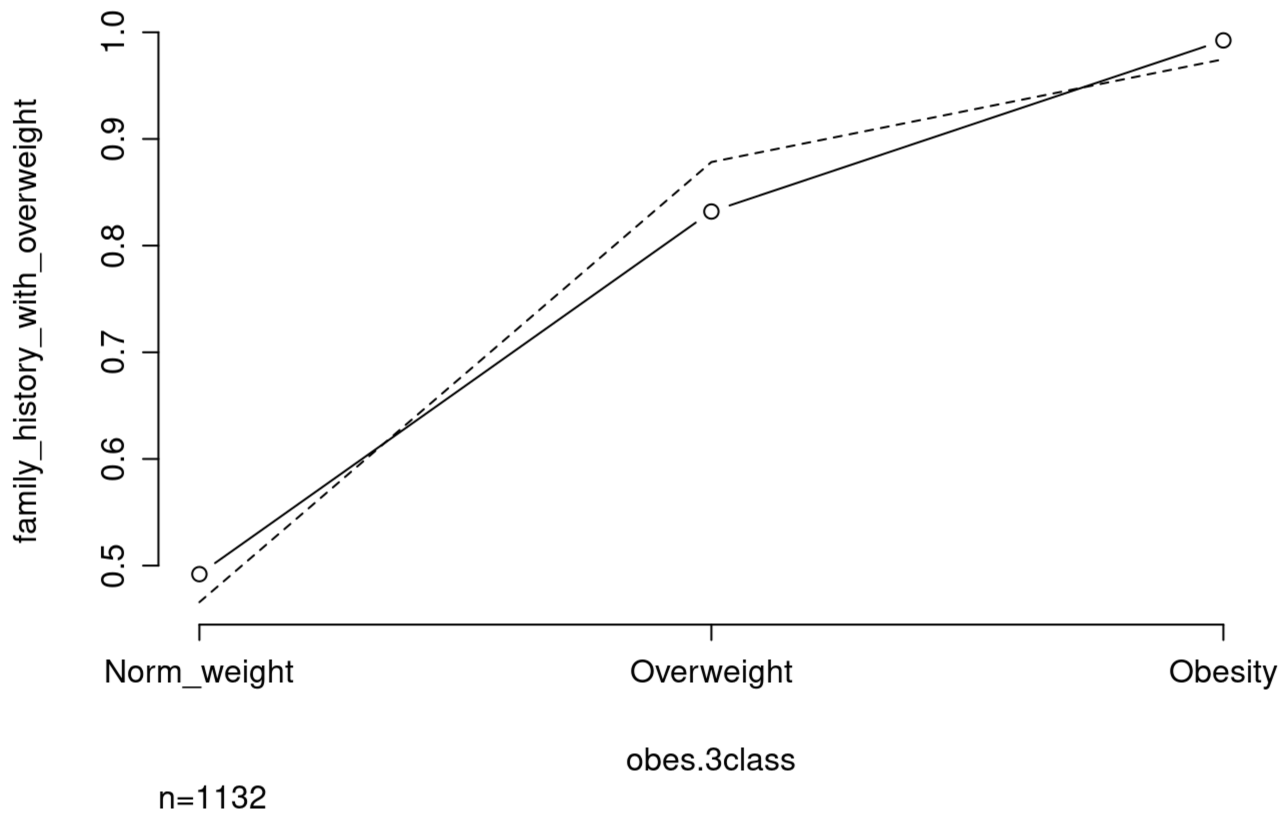
```
##  Length   Class    Mode
##       3 formula    call
```

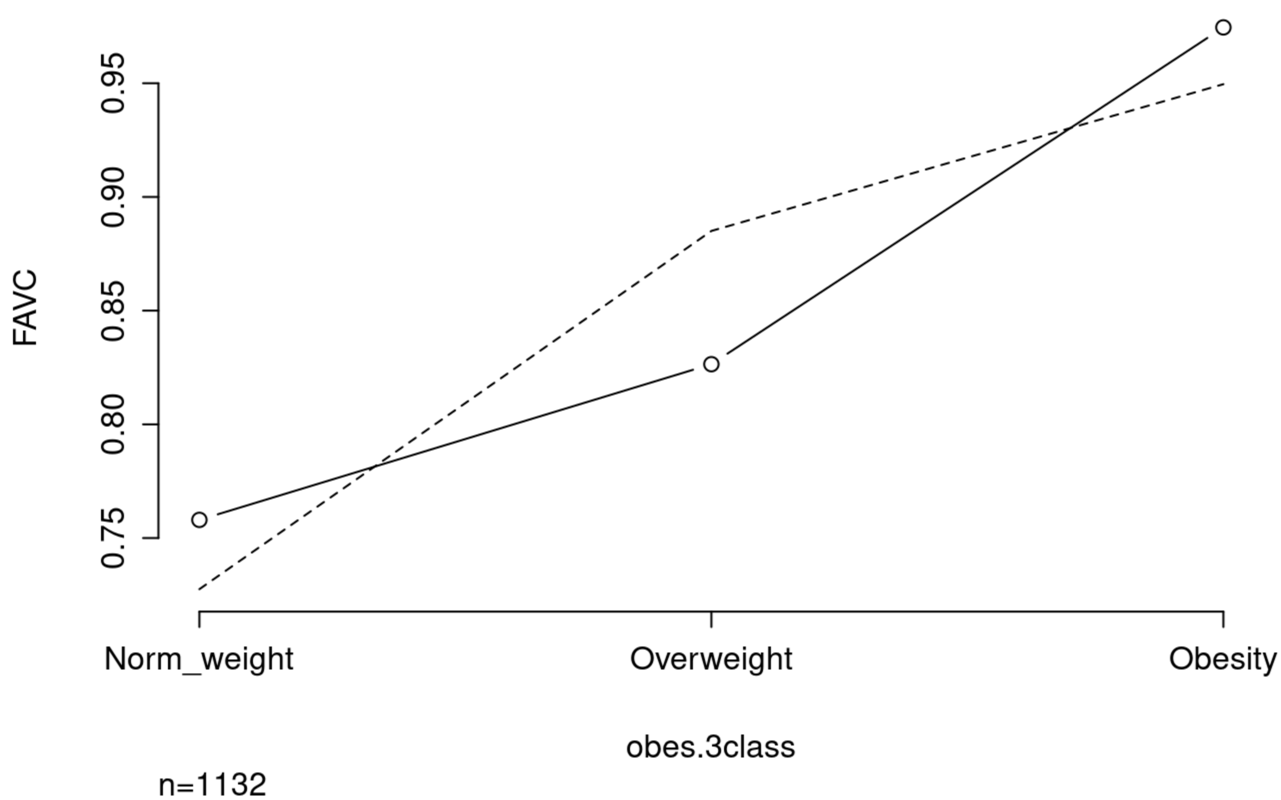
The t values are the ratios of values to standard errors. t values larger than 2 or 3 are good evidence that the sign on the coefficient is correct. In the output above, the t value of 3.53 for the coefficient is strong evidence that the effect of treatment is positive. The differences in the distance between the sets used to calculate the values for the predictor may suggest that the parallel slopes assumption does not hold for the model

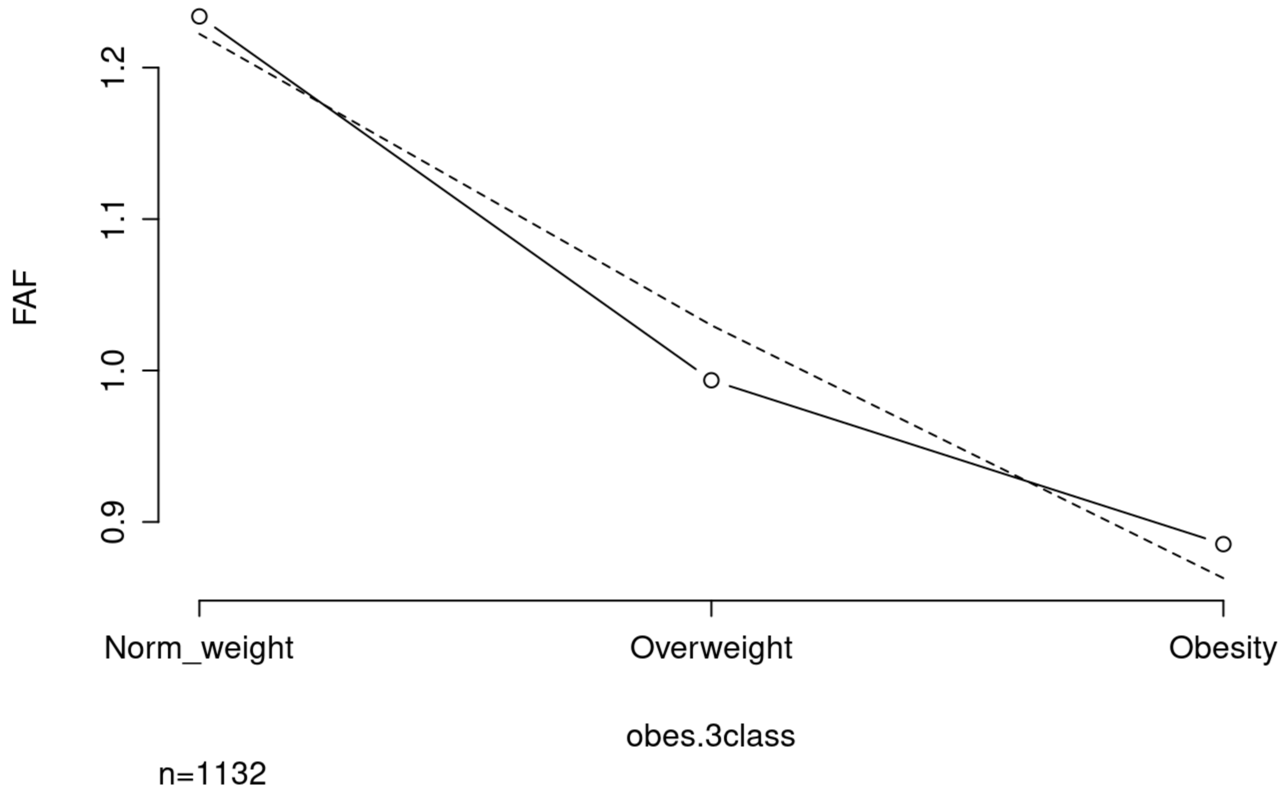
$X^2 = 201.23$, df = 9, p-value = 0 The Omnibus test checks whether the proportional odds assumption holds for the entire model. Since the p-value is 0 (less than 0.05), the null hypothesis is rejected, meaning the Proportional Odds assumption is violated for the model as a whole. the larger the sample size, the stronger that tendency will be; i tried to avoid this, which is the reason for the undersampling in my dataset from ~ 2000- ~1600

A graphical method for assessing the proportional odds assumption is available in the rms package. The function plot.xmean.ordinaly plots both When the solid and dashed lines `roughly` follow the same trajectory, we have good evidence that the PO assumption is safe. This looks okay!

```
# rms:: allows us to use the function without loading the rms package
rms::plot.xmean.ordinaly(obes.3class ~ family_history_with_overweight+FAVC+FAF,data= b.obesleve
l, subset = train)
```


family_history_with_overweight
0.5
0.6
0.7
0.8
0.9
1.0
Norm_weight
Overweight
Obesity
obes.3class
n=1132


FAVC
0.75
0.80
0.85
0.90
0.95
Norm_weight
Overweight
Obesity
obes.3class
n=1132



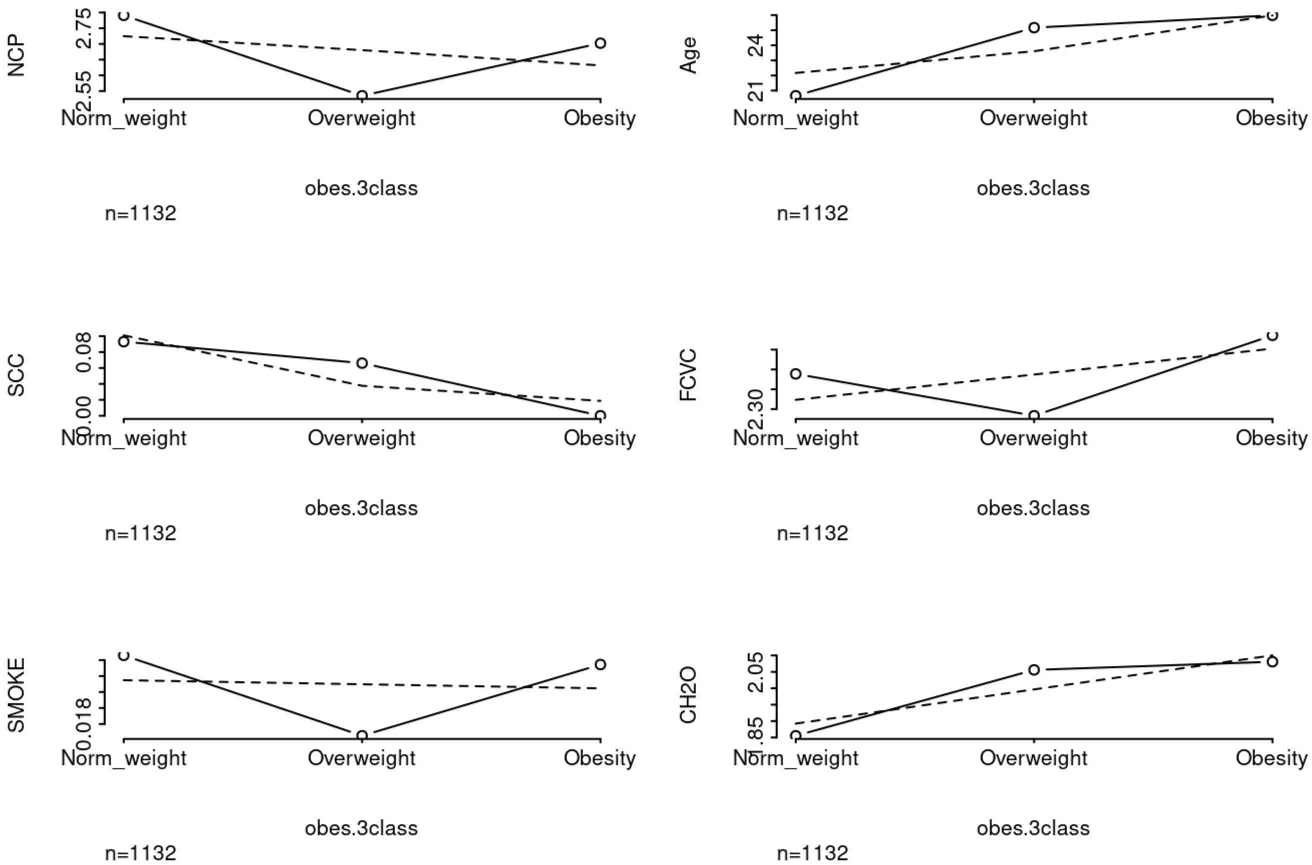

```
# Load the rms package
#library(rms)

# Set up the plotting layout to display plots side by side
# 'mfrow = c(1, 2)' means 1 row and 2 columns (adjust as needed)
#par(mfrow = c(1, 2))  # For 2 plots side by side, change numbers accordingly

# Create individual plots for the predictors
#plot.xmean.ordinaly(obes.3class ~ family_history_with_overweight, data = b.obeslevel)
#plot.xmean.ordinaly(obes.3class ~ Age, data = b.obeslevel)

# If you want to add more plots, continue adding them like this:
#plot.xmean.ordinaly(obes.3class ~ SCC, data = b.obeslevel)
# Set up for 4 plots in 2 rows and 2 columns


# Generate the plots
par(mfrow = c(3, 2))
rms::plot.xmean.ordinaly(obes.3class ~ NCP, data = b.obeslevel, subset = train)
rms::plot.xmean.ordinaly(obes.3class ~ Age, data = b.obeslevel, subset = train)
rms::plot.xmean.ordinaly(obes.3class ~ SCC, data = b.obeslevel, subset = train)
rms::plot.xmean.ordinaly(obes.3class ~ FCVC, data = b.obeslevel, subset = train)
rms::plot.xmean.ordinaly(obes.3class ~ SMOKE, data = b.obeslevel, subset = train)
rms::plot.xmean.ordinaly(obes.3class ~ CH2O, data = b.obeslevel, subset = train)
```




```
# Reset to default
#par(mfrow = c(1, 1))
#plot(logodds.obes, which=1:3, pch=1:3, xlab='logit', main=' ', xlim=range(logodds.obes[,3:4]))
```

The assumption of our model does not hold according to the plots provided.also, According to Frank Harrell, Violation of Proportional Odds is Not Fatal. See https://www.fharrell.com/post/po/ (https://www.fharrell.com/post/po/) Their are other model methods we could use that might help prove the null hypothesis, like The partial proportional odds model, which allows some predictors to have multiple coefficients that vary with the response level, while keeping the assumption for others. But the visualization of the provided plot helps prove the null hypothesis on each variable, excluding the model as a whole.

Predicting the fitted Model

```
knitr::opts_chunk$set(echo=FALSE, error = TRUE, message = FALSE,  warning = FALSE)
orpred= predict(obes.logr, newdata = b.obeslevel[-train,],
        type = "class") #Auto[-train,], type = "class"
with(b.obeslevel[-train,], table(orpred, as.ordered(obes.3class)))
```

```
##
## orpred        Norm_weight Overweight Obesity
##   Norm_weight         116         45      17
##   Obesity              48         83      77
##   Overweight           19         89      72
```

```
# Calculate and print accuracy
(sum(diag(with(b.obeslevel[-train,], table(orpred, as.ordered(obes.3class))))) / length(orpred))
```

```
## [1] 0.4787986
```

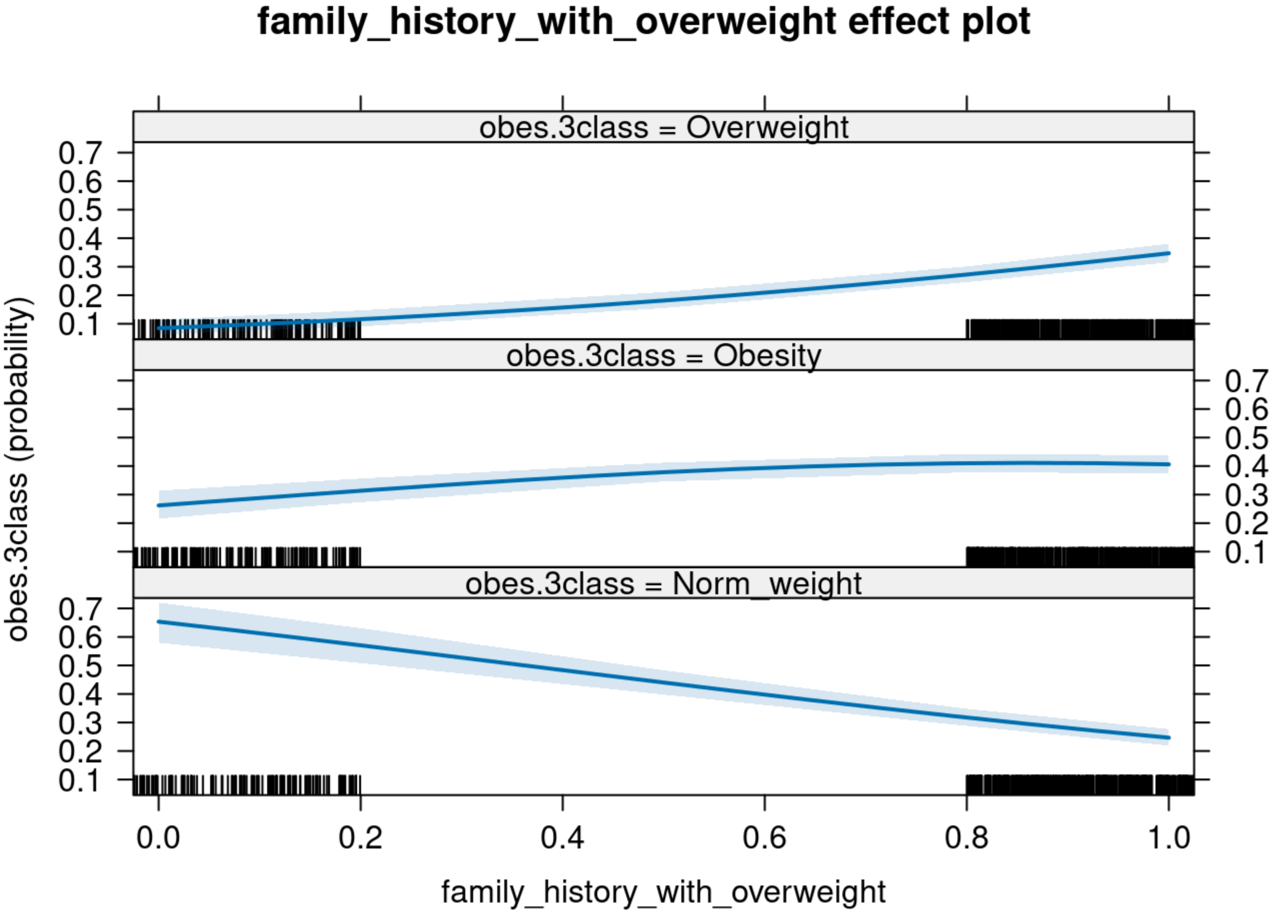
```
print(paste("Model accuracy:", (sum(diag(with(b.obeslevel[-train,], table(orpred, as.ordered(obe
s.3class))))) / length(orpred))))
```

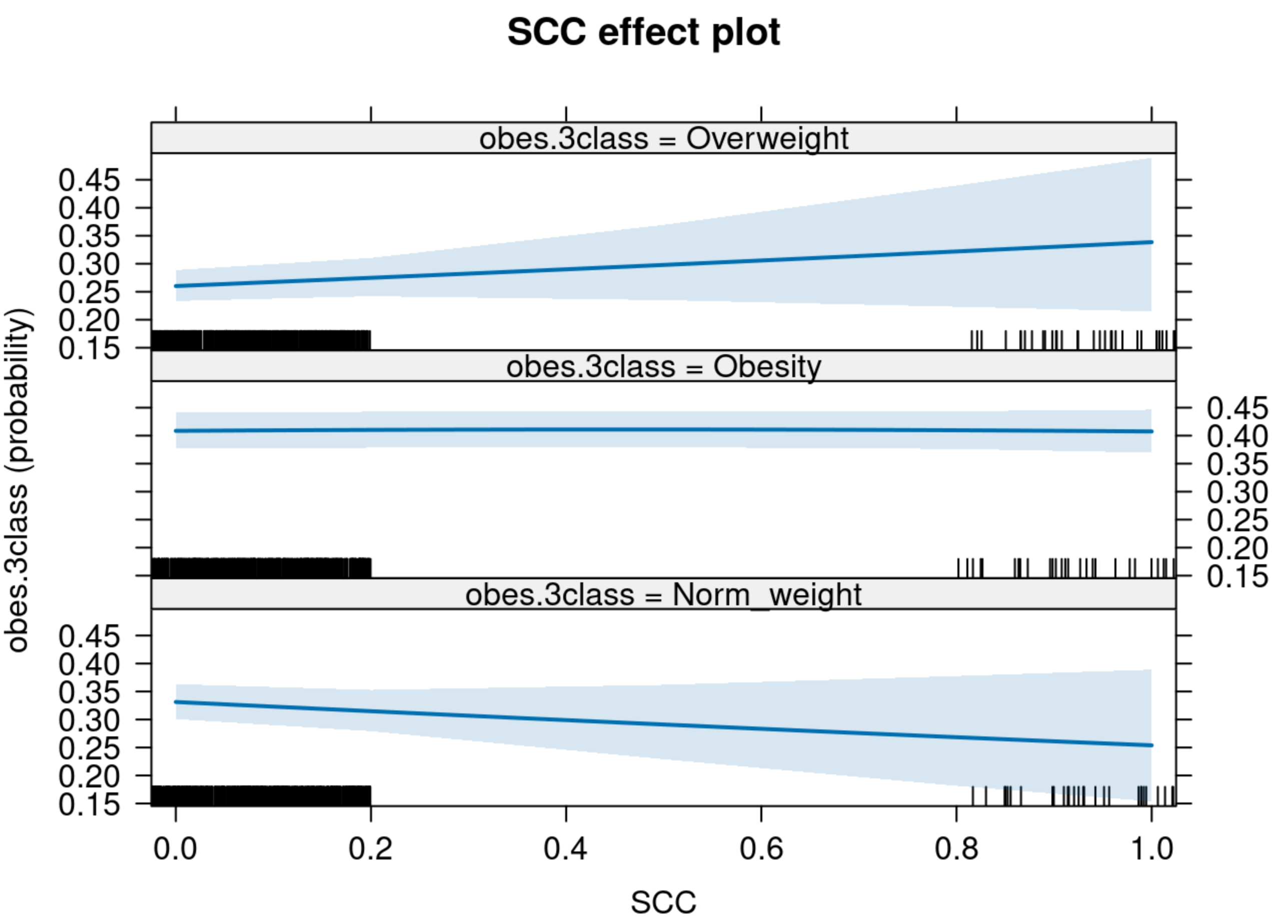
```
## [1] "Model accuracy: 0.478798586572438"
```

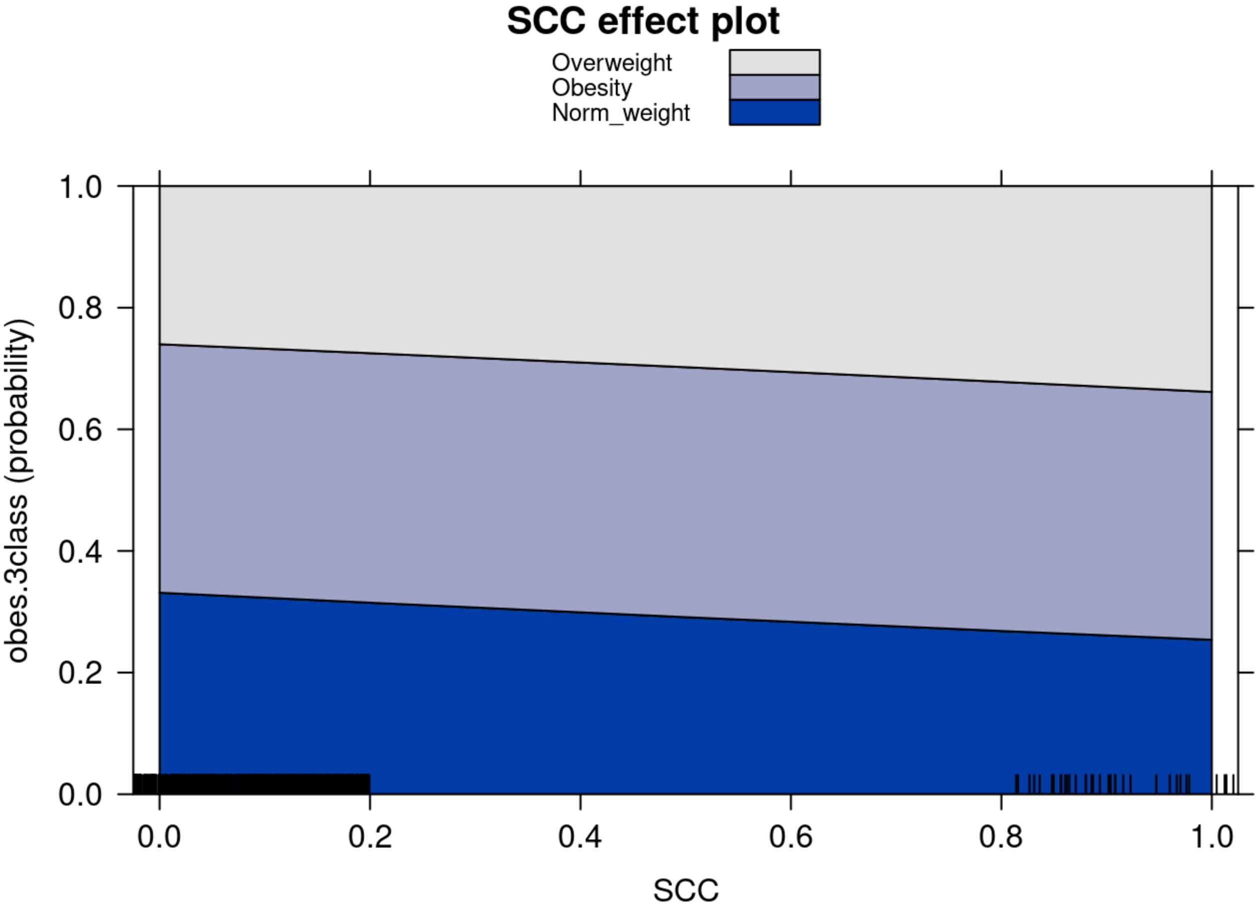
```
orprob= predict(obes.logr, newdata = b.obeslevel[-train,],
        type = "prob")
# Summarize the probabilities for each class
(colMeans(orprob))  # Get mean probabilities for each class. This gives concise class probabilit
ies
```

```
## Norm_weight     Obesity  Overweight
##   0.3495037   0.3426916   0.3078047
```

The following graph results in a plot for each outcome. We can see the probability of Marked improvement is much higher for the Treated group than the Placebo group. Likewise the probability of no improvement is much lower in the Treated group than in the Placebo group.


family_history_with_overweight effect plot
obes.3class = Overweight
obes.3class = Obesity
obes.3class = Norm_weight
obes.3class (probability)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.2
0.4
0.6
0.8
1.0
family_history_with_overweight


SCC effect plot
obes.3class = Overweight
obes.3class = Obesity
obes.3class = Norm_weight
obes.3class (probability)
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.0
0.2
0.4
0.6
0.8
1.0
SCC



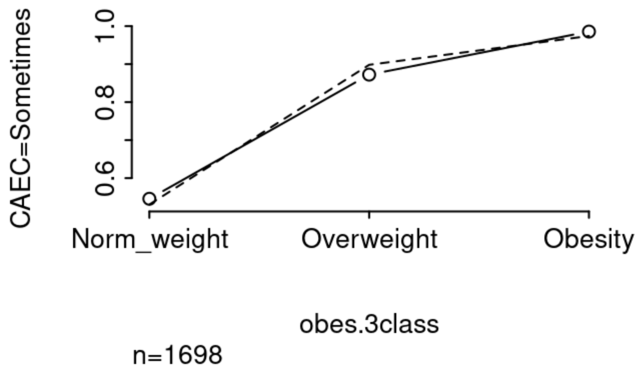
Simple Ordinal Model; M2

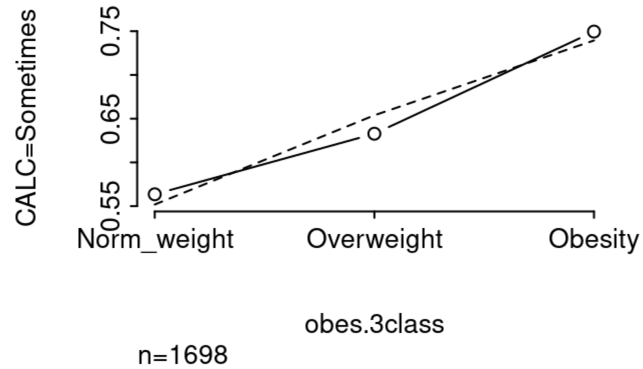
After the completion of the complex model, we would compute a simpler model using the variables that were excluded from the previous models predictor variables, while including family_history_with_overweight.

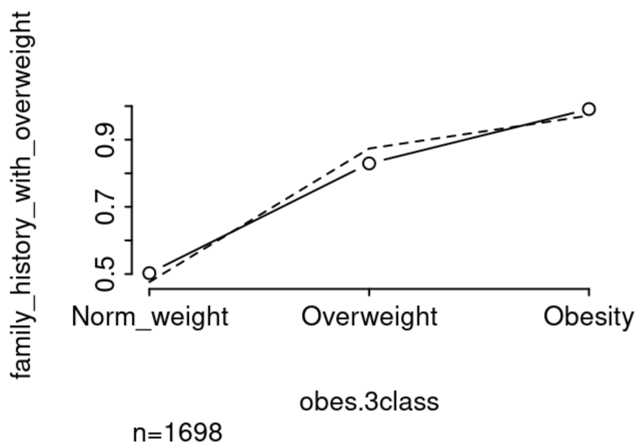
```
## Call:
## polr(formula = obes.3class ~ family_history_with_overweight +
##     CAEC + CALC, data = b.obeslevel, subset = train, Hess = TRUE)
##
## Coefficients:
##                                 Value Std. Error t value
## family_history_with_overweight  2.631     0.1857  14.168
## CAECFrequently                 -1.047     0.4688  -2.234
## CAECno                          2.199     0.5177   4.248
## CAECSometimes                   1.874     0.4079   4.595
## CALCFrequently                 13.055     0.2567  50.860
## CALCno                         12.541     0.1647  76.135
## CALCSometimes                  13.141     0.1532  85.784
##
## Intercepts:
##                         Value   Std. Error t value
## Norm_weight|Overweight 15.5685  0.3241    48.0317
## Overweight|Obesity     17.5261  0.3460    50.6477
##
## Residual Deviance: 1948.162
## AIC: 1966.162
```

```
##                                   Value Std. Error   t value      p value
## family_history_with_overweight  2.630736  0.1856836 14.167838 1.448940e-45
## CAECFrequently                 -1.047457  0.4688360 -2.234166 2.547216e-02
## CAECno                          2.199165  0.5177307  4.247700 2.159768e-05
## CAECSometimes                   1.874338  0.4079467  4.594566 4.336505e-06
## CALCFrequently                 13.055410  0.2566944 50.859729 0.000000e+00
## CALCno                         12.541157  0.1647225 76.135064 0.000000e+00
## CALCSometimes                  13.140707  0.1531830 85.784350 0.000000e+00
## Norm_weight|Overweight         15.568523  0.3241300 48.031730 0.000000e+00
## Overweight|Obesity             17.526099  0.3460397 50.647650 0.000000e+00
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: obes.3class
##                                                                          Model
## 1                                        family_history_with_overweight + CAEC + CALC
## 2 family_history_with_overweight + Age + SCC + FAF + CH2O + FAVC + FCVC + NCP + SMOKE
##   Resid. df Resid. Dev   Test    Df  LR stat. Pr(Chi)
## 1      1123   1948.162
## 2      1121   2207.357 1 vs 2     2 -259.1949       1
```







Comparison

The Two models were compared together using `ANOVA`. The simpler training model (M2) produced a lower residual deviance and the resulting LR stat suggest the complex model had a lower performance among the two models, suggesting a significant difference between both models, with Pr(chi) also supporting this assertion.

References

- “Estimation of Obesity Levels Based On Eating Habits and Physical Condition .” UCI Machine Learning Repository, 2019, https://doi.org/10.24432/C5H31Z (https://doi.org/10.24432/C5H31Z).
- Mozaffarian, Dariush. “Perspective: Obesity-an unexplained epidemic.” The American journal of clinical nutrition vol. 115,6 (2022): 1445-1450. doi:10.1093/ajcn/nqac075 (doi:10.1093/ajcn/nqac075)