Analysis

220225605

2023-09-01

# MSc Dissertation

## Does ethnicity and income moderate the relationship between the consumption of saturated fats and free sugars and BMI?

### Loading the required libraries

library(here)

## here() starts at C:/Users/solom/Desktop/Final Project MSc/MSc Dissertation

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

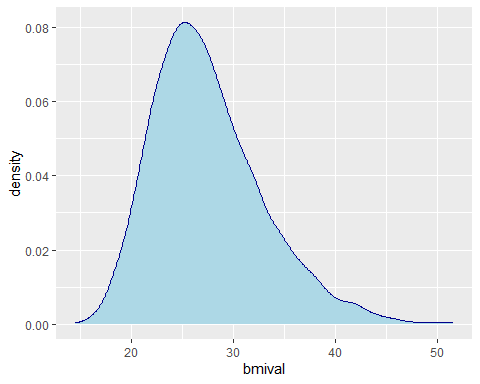
library(ggfortify)

### Importing cleaned and processed data

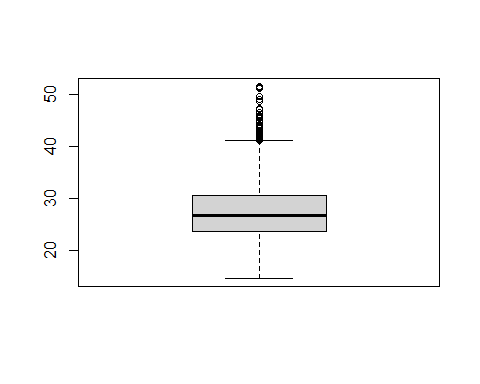
load(here("processed data/ProcessedData.RData"))

### Visualising the outcome variable

# density plots for the outcome variable   
p <- ggplot(FinalData, aes(x = bmival))  
p + geom\_density(color = "darkblue", fill = "lightblue")



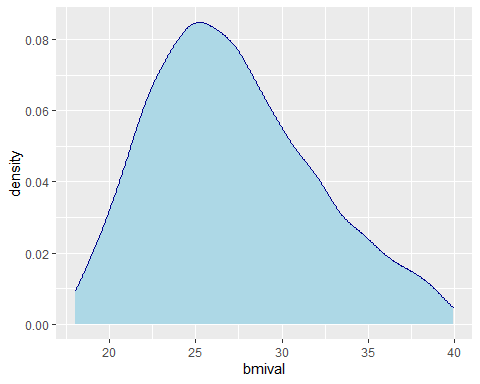
# checking for outliers  
boxplot(FinalData$bmival)



# following the exclusion criteria (removal of participants with a BMI less than 18 and BMI above 45)  
FinalDataTemp <- FinalData[FinalData$bmival > 18,]  
FinalDataTemp <- FinalDataTemp[FinalDataTemp$bmival < 40,]  
  
# checking for the range   
range(FinalDataTemp$bmival)

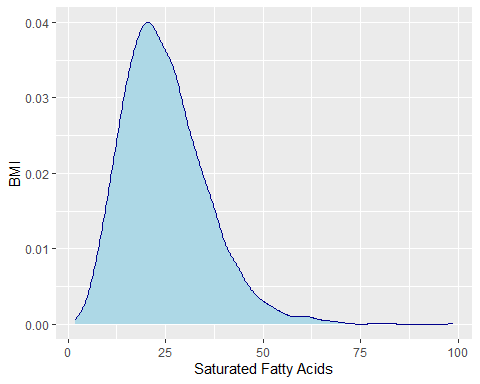
## [1] 18.02171 39.95838

# density plot of the outcome variable after the removal of underweight and morbidly obese participants   
p1 <- ggplot(FinalDataTemp, aes(x = bmival))  
p1 + geom\_density(color = "darkblue", fill = "lightblue")

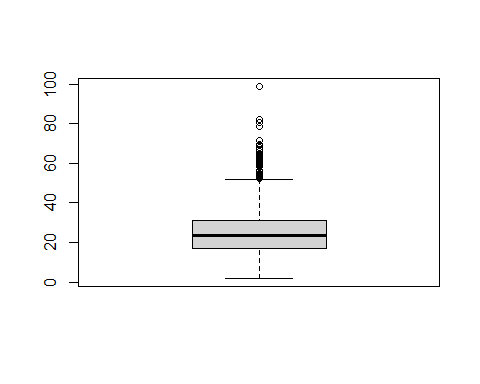


### Checking the relationship between predictors and the outcome variable

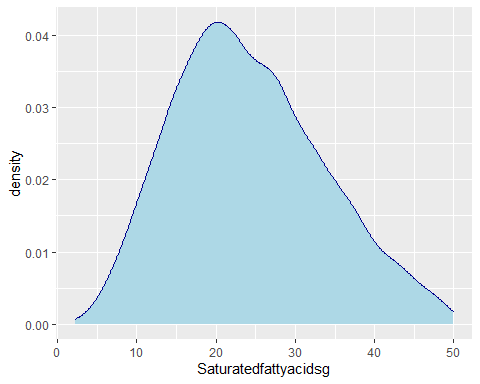
# Predictor 1 - Saturated Fatty Acids ----------------------------------------  
ggplot(FinalData, aes(x = Saturatedfattyacidsg)) + geom\_density(color = "darkblue", fill = "lightblue")+ labs(x = "Saturated Fatty Acids", y = "BMI")



# checking for outliers   
boxplot(FinalData$Saturatedfattyacidsg)

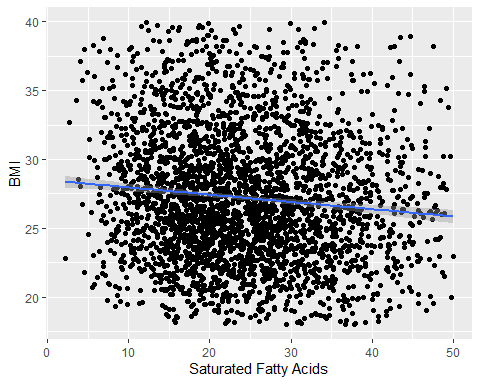


# removing outliers   
FinalDataTemp <- FinalDataTemp[FinalDataTemp$Saturatedfattyacidsg < 50,]  
ggplot(FinalDataTemp, aes(x = Saturatedfattyacidsg)) + geom\_density(color = "darkblue", fill = "lightblue")

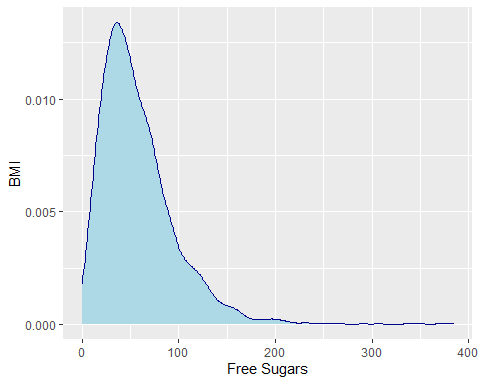


# scatter plot with saturated fat intake on the x axis and BMI on the y axis   
p1 <- ggplot(FinalDataTemp, aes(x = Saturatedfattyacidsg, y = bmival))  
p1 + geom\_point(color = "grey0")+ labs(x = "Saturated Fatty Acids", y = "BMI") + geom\_smooth(method = "lm")

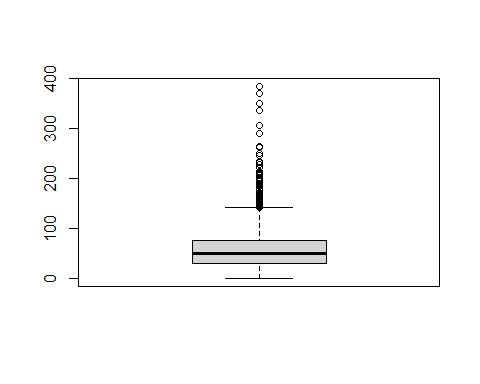
## `geom\_smooth()` using formula = 'y ~ x'



# Predictor 2 - Free Sugar Consumption-----------------------  
ggplot(FinalData, aes(x = FreeSugarsg)) + geom\_density(color = "darkblue", fill = "lightblue") + labs(x = "Free Sugars", y = "BMI")



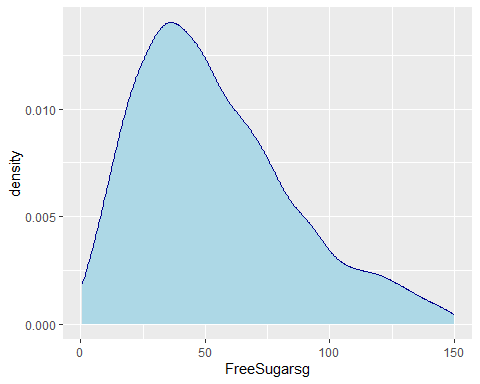
# checking for outliers   
boxplot(FinalData$FreeSugarsg)



# removal of outliers   
FinalDataTemp <- FinalDataTemp[FinalDataTemp$FreeSugarsg < 150,]  
range(FinalDataTemp$FreeSugarsg)

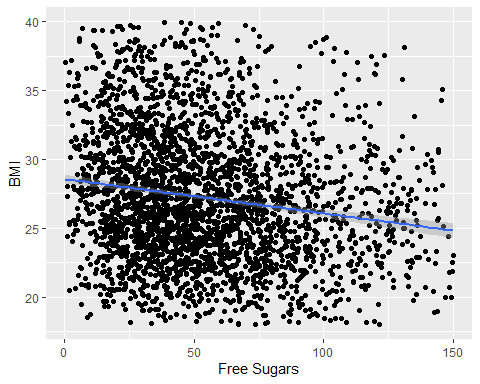
## [1] 0.480125 149.904900

ggplot(FinalDataTemp, aes(x = FreeSugarsg)) + geom\_density(color = "darkblue", fill = "lightblue")



# scatter plot with free sugar intake on the x axis and BMI on the y axis  
p2 <- ggplot(FinalDataTemp, aes(x = FreeSugarsg, y = bmival))  
p2 + geom\_point(color = "grey0")+ labs(x = "Free Sugars", y = "BMI") + geom\_smooth(method = "lm")

## `geom\_smooth()` using formula = 'y ~ x'



### Descriptive Statistics

# frequency table - weight statuses   
table(FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese   
## 2225 806

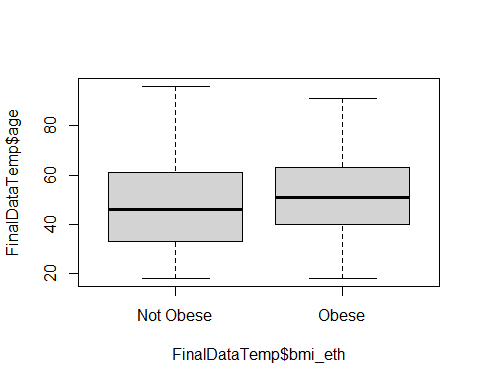
# mean and standard deviation of age across weight statuses   
FinalDataTemp %>% group\_by(bmi\_eth) %>% summarise\_at(vars(age, eqvinc), list(mean = mean, sd = sd))

## # A tibble: 2 × 5  
## bmi\_eth age\_mean eqvinc\_mean age\_sd eqvinc\_sd  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Not Obese 47.0 33514. 18.1 24472.  
## 2 Obese 51.2 28762. 15.8 20536.

sd(FinalDataTemp$age) # standard deviation of age

## [1] 17.6184

# performing a t-test to compare the mean age across weight statuses   
boxplot(FinalDataTemp$age ~ FinalDataTemp$bmi\_eth)



t.test (age ~ bmi\_eth, var.equal=TRUE, data = FinalDataTemp)

##   
## Two Sample t-test  
##   
## data: age by bmi\_eth  
## t = -5.8317, df = 3029, p-value = 6.065e-09  
## alternative hypothesis: true difference in means between group Not Obese and group Obese is not equal to 0  
## 95 percent confidence interval:  
## -5.613693 -2.788646  
## sample estimates:  
## mean in group Not Obese mean in group Obese   
## 46.98742 51.18859

# ethnicity across weight status   
table(FinalDataTemp$EthGrpGB, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## 1 2137 781  
## 2 40 14  
## 3 43 10  
## 4 5 1

# converting equivalised household income into three categories (tertiles)  
FinalDataTemp <- FinalDataTemp %>%  
 mutate(tertiles = ntile(eqvinc, 3)) %>%  
 mutate(tertiles = if\_else(tertiles == 1, 'Low', if\_else(tertiles == 2, 'Medium', 'High')))  
# equivalised income across weight status   
table(FinalDataTemp$tertiles, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## High 784 226  
## Low 701 310  
## Medium 740 270

# chi-square test (Categorical variables - income tertiles and weight status)  
chisq.test(FinalDataTemp$tertiles, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$tertiles and FinalDataTemp$bmi\_eth  
## X-squared = 17.788, df = 2, p-value = 0.0001372

# gender across weight statuses   
table(FinalDataTemp$Sex, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## Male 963 317  
## Female 1262 489

# chi-square test (Categorical variables - gender and weight status)  
chisq.test(FinalDataTemp$Sex, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$Sex and FinalDataTemp$bmi\_eth  
## X-squared = 3.7857, df = 1, p-value = 0.05169

# qualifications across weight statuses   
table(FinalDataTemp$qual7, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## A level or equivalent 356 102  
## Foreign or other qualifications 74 39  
## GCSE or equivalent 455 182  
## Higher education and above 824 239  
## No qualifications 392 220  
## Still in FT education 124 24

# chi-square test (Categorical variables - educational qualifications and weight status)  
chisq.test(FinalDataTemp$qual7, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$qual7 and FinalDataTemp$bmi\_eth  
## X-squared = 54.088, df = 5, p-value = 2.011e-10

# presence of children across weight statuses   
table(FinalDataTemp$NumCh118, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## Do not have Children 1371 524  
## Have Children 854 282

# chi-square test (Categorical variables - presence of children and weight status)  
chisq.test(FinalDataTemp$NumCh118, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$NumCh118 and FinalDataTemp$bmi\_eth  
## X-squared = 2.9094, df = 1, p-value = 0.08807

# current cigarette smoking status across weight statuses   
table(FinalDataTemp$cigst2, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## Non-smoker 1657 632  
## Smoker 568 174

# chi-square test (Categorical variables - cig. smoking status and weight status)  
chisq.test(FinalDataTemp$cigst2, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$cigst2 and FinalDataTemp$bmi\_eth  
## X-squared = 4.9681, df = 1, p-value = 0.02582

# saturated fat intake status across weight status   
# calculating daily caloric intake from saturated fats alone   
FinalDataTemp <- FinalDataTemp %>% mutate(SF\_calories = Saturatedfattyacidsg\*9)  
# creating a new column to determine if a calories from saturated fat exceed 10% of total caloric intake  
FinalDataTemp <- FinalDataTemp %>% mutate(SF\_Status = case\_when(SF\_calories >= 0.10\*Energykcal ~ "Yes", SF\_calories < 0.10\*Energykcal ~ "No"))  
# frequency table (Saturated Fat Intake and Weight Status)  
table(FinalDataTemp$SF\_Status, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## No 531 220  
## Yes 1694 586

# chi-square test (Categorical variables - saturated fat intake status and weight status)  
chisq.test(FinalDataTemp$SF\_Status, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$SF\_Status and FinalDataTemp$bmi\_eth  
## X-squared = 3.735, df = 1, p-value = 0.05328

# free sugar intake status across weight status   
# calculating daily caloric intake from free sugars alone   
FinalDataTemp <- FinalDataTemp %>% mutate(FS\_calories = FreeSugarsg\*4)  
# creating a new column to determine if a calories from free sugars exceed 10% of total caloric intake  
FinalDataTemp <- FinalDataTemp %>% mutate(FS\_Status = case\_when(FS\_calories >= 0.10\*Energykcal ~ "Yes", FS\_calories < 0.10\*Energykcal ~ "No"))  
# frequency table (Free Sugar Intake and Weight Status)  
table(FinalDataTemp$FS\_Status, FinalDataTemp$bmi\_eth)

##   
## Not Obese Obese  
## No 826 378  
## Yes 1399 428

# chi-square test (Categorical variables - free sugar intake status and weight status)  
chisq.test(FinalDataTemp$FS\_Status, FinalDataTemp$bmi\_eth, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: FinalDataTemp$FS\_Status and FinalDataTemp$bmi\_eth  
## X-squared = 23.61, df = 1, p-value = 1.18e-06

### Multiple Linear Regression

#### Adding covariates to the model

Including age, gender, educational attainments, presence of children and cigarette smoking status as predictors and BMI as the outcome variable

FinalDataTemp$qual7 <- relevel(FinalDataTemp$qual7, ref = "No qualifications")  
lm1 <- lm(bmival ~ age + Sex + qual7 + NumCh118 +cigst2, data = FinalDataTemp)  
summary(lm1)

##   
## Call:  
## lm(formula = bmival ~ age + Sex + qual7 + NumCh118 + cigst2,   
## data = FinalDataTemp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.4913 -3.3938 -0.6419 2.8842 14.5497   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 26.350838 0.463385 56.866 < 2e-16 \*\*\*  
## age 0.038730 0.006234 6.212 5.93e-10 \*\*\*  
## SexFemale -0.169685 0.170539 -0.995 0.319821   
## qual7A level or equivalent -1.028942 0.301285 -3.415 0.000646 \*\*\*  
## qual7Foreign or other qualifications -0.565660 0.472238 -1.198 0.231078   
## qual7GCSE or equivalent -0.434492 0.272585 -1.594 0.111049   
## qual7Higher education and above -1.161457 0.250950 -4.628 3.84e-06 \*\*\*  
## qual7Still in FT education -2.250400 0.476978 -4.718 2.49e-06 \*\*\*  
## NumCh118Have Children 0.238630 0.199044 1.199 0.230669   
## cigst2Smoker -0.804314 0.202637 -3.969 7.38e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.612 on 3021 degrees of freedom  
## Multiple R-squared: 0.05288, Adjusted R-squared: 0.05006   
## F-statistic: 18.74 on 9 and 3021 DF, p-value: < 2.2e-16

# confidence intervals   
confint(lm1,level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 25.44225667 27.25942019  
## age 0.02650643 0.05095455  
## SexFemale -0.50406850 0.16469948  
## qual7A level or equivalent -1.61968673 -0.43819798  
## qual7Foreign or other qualifications -1.49160087 0.36028108  
## qual7GCSE or equivalent -0.96896334 0.09998018  
## qual7Higher education and above -1.65350731 -0.66940703  
## qual7Still in FT education -3.18563350 -1.31516634  
## NumCh118Have Children -0.15164587 0.62890567  
## cigst2Smoker -1.20163400 -0.40699433

#### Adding the primary predictors to the model

Including saturated fatty acids intake and free sugars intake to the model controlling for the covariates

FinalDataTemp$qual7 <- relevel(FinalDataTemp$qual7, ref = "No qualifications")  
lm2 <- lm(bmival ~ age + Sex + qual7 + NumCh118 +cigst2 +Saturatedfattyacidsg + FreeSugarsg, data = FinalDataTemp)  
summary(lm2)

##   
## Call:  
## lm(formula = bmival ~ age + Sex + qual7 + NumCh118 + cigst2 +   
## Saturatedfattyacidsg + FreeSugarsg, data = FinalDataTemp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.669 -3.359 -0.615 2.793 14.538   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 28.566982 0.528604 54.042 < 2e-16 \*\*\*  
## age 0.034106 0.006247 5.460 5.15e-08 \*\*\*  
## SexFemale -0.628080 0.177251 -3.543 0.000401 \*\*\*  
## qual7A level or equivalent -0.896971 0.298503 -3.005 0.002679 \*\*   
## qual7Foreign or other qualifications -0.482105 0.466921 -1.033 0.301911   
## qual7GCSE or equivalent -0.342009 0.269872 -1.267 0.205146   
## qual7Higher education and above -1.040086 0.249055 -4.176 3.05e-05 \*\*\*  
## qual7Still in FT education -2.101673 0.471873 -4.454 8.74e-06 \*\*\*  
## NumCh118Have Children 0.180302 0.196897 0.916 0.359889   
## cigst2Smoker -0.751621 0.200598 -3.747 0.000182 \*\*\*  
## Saturatedfattyacidsg -0.036297 0.009951 -3.648 0.000269 \*\*\*  
## FreeSugarsg -0.017490 0.003009 -5.813 6.77e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.558 on 3019 degrees of freedom  
## Multiple R-squared: 0.07534, Adjusted R-squared: 0.07197   
## F-statistic: 22.36 on 11 and 3019 DF, p-value: < 2.2e-16

confint(lm2, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 27.53052211 29.60344142  
## age 0.02185759 0.04635410  
## SexFemale -0.97562541 -0.28053410  
## qual7A level or equivalent -1.48226058 -0.31168061  
## qual7Foreign or other qualifications -1.39762010 0.43341015  
## qual7GCSE or equivalent -0.87116116 0.18714274  
## qual7Higher education and above -1.52841980 -0.55175125  
## qual7Still in FT education -3.02689812 -1.17644715  
## NumCh118Have Children -0.20576389 0.56636718  
## cigst2Smoker -1.14494362 -0.35829836  
## Saturatedfattyacidsg -0.05580795 -0.01678701  
## FreeSugarsg -0.02338891 -0.01159059

#### Including ethnicity and income as possible predictors (main effects)

FinalDataTemp$qual7 <- relevel(FinalDataTemp$qual7, ref = "No qualifications")  
FinalDataTemp$EthGrpGB <- relevel(FinalDataTemp$EthGrpGB, ref = 1)  
lm3 <- lm(bmival ~ age + Sex + qual7 + NumCh118 +cigst2 +Saturatedfattyacidsg + FreeSugarsg + EthGrpGB + eqvinc, data = FinalDataTemp)  
summary(lm3)

##   
## Call:  
## lm(formula = bmival ~ age + Sex + qual7 + NumCh118 + cigst2 +   
## Saturatedfattyacidsg + FreeSugarsg + EthGrpGB + eqvinc, data = FinalDataTemp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.6623 -3.3216 -0.5903 2.8027 14.2968   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.908e+01 5.490e-01 52.970 < 2e-16 \*\*\*  
## age 3.143e-02 6.285e-03 5.001 6.02e-07 \*\*\*  
## SexFemale -6.500e-01 1.773e-01 -3.666 0.000250 \*\*\*  
## qual7A level or equivalent -8.040e-01 2.993e-01 -2.686 0.007270 \*\*   
## qual7Foreign or other qualifications -3.860e-01 4.670e-01 -0.826 0.408591   
## qual7GCSE or equivalent -2.974e-01 2.698e-01 -1.102 0.270521   
## qual7Higher education and above -7.898e-01 2.594e-01 -3.045 0.002345 \*\*   
## qual7Still in FT education -2.095e+00 4.723e-01 -4.437 9.47e-06 \*\*\*  
## NumCh118Have Children 4.616e-02 2.007e-01 0.230 0.818068   
## cigst2Smoker -8.313e-01 2.017e-01 -4.121 3.87e-05 \*\*\*  
## Saturatedfattyacidsg -3.497e-02 9.963e-03 -3.510 0.000454 \*\*\*  
## FreeSugarsg -1.803e-02 3.009e-03 -5.993 2.31e-09 \*\*\*  
## EthGrpGB2 3.084e-01 6.291e-01 0.490 0.623972   
## EthGrpGB3 -5.497e-01 6.362e-01 -0.864 0.387647   
## EthGrpGB4 -2.924e+00 1.869e+00 -1.565 0.117795   
## eqvinc -1.270e-05 3.870e-06 -3.282 0.001042 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.551 on 3015 degrees of freedom  
## Multiple R-squared: 0.0796, Adjusted R-squared: 0.07502   
## F-statistic: 17.38 on 15 and 3015 DF, p-value: < 2.2e-16

confint(lm3, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 2.800359e+01 3.015647e+01  
## age 1.910866e-02 4.375418e-02  
## SexFemale -9.976871e-01 -3.023794e-01  
## qual7A level or equivalent -1.390950e+00 -2.171087e-01  
## qual7Foreign or other qualifications -1.301746e+00 5.297422e-01  
## qual7GCSE or equivalent -8.265009e-01 2.317158e-01  
## qual7Higher education and above -1.298364e+00 -2.812820e-01  
## qual7Still in FT education -3.021589e+00 -1.169381e+00  
## NumCh118Have Children -3.473014e-01 4.396305e-01  
## cigst2Smoker -1.226792e+00 -4.358073e-01  
## Saturatedfattyacidsg -5.450452e-02 -1.543609e-02  
## FreeSugarsg -2.393526e-02 -1.213403e-02  
## EthGrpGB2 -9.250970e-01 1.541982e+00  
## EthGrpGB3 -1.797068e+00 6.977221e-01  
## EthGrpGB4 -6.588798e+00 7.405162e-01  
## eqvinc -2.029045e-05 -5.113639e-06

#### Interaction Effects

lm4 <- lm(bmival ~ age + Sex + qual7 + NumCh118 +cigst2 +Saturatedfattyacidsg + FreeSugarsg + EthGrpGB + eqvinc + Saturatedfattyacidsg\*EthGrpGB +FreeSugarsg \*EthGrpGB + Saturatedfattyacidsg\*eqvinc + FreeSugarsg \* eqvinc + Saturatedfattyacidsg\*EthGrpGB\*eqvinc + FreeSugarsg\*EthGrpGB\*eqvinc, data = FinalDataTemp)  
summary(lm4)

##   
## Call:  
## lm(formula = bmival ~ age + Sex + qual7 + NumCh118 + cigst2 +   
## Saturatedfattyacidsg + FreeSugarsg + EthGrpGB + eqvinc +   
## Saturatedfattyacidsg \* EthGrpGB + FreeSugarsg \* EthGrpGB +   
## Saturatedfattyacidsg \* eqvinc + FreeSugarsg \* eqvinc + Saturatedfattyacidsg \*   
## EthGrpGB \* eqvinc + FreeSugarsg \* EthGrpGB \* eqvinc, data = FinalDataTemp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.6854 -3.2887 -0.6032 2.8093 14.4333   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 2.962e+01 6.348e-01 46.657 < 2e-16  
## age 3.109e-02 6.295e-03 4.938 8.31e-07  
## SexFemale -6.662e-01 1.777e-01 -3.749 0.000181  
## qual7A level or equivalent -7.966e-01 2.997e-01 -2.658 0.007892  
## qual7Foreign or other qualifications -3.665e-01 4.672e-01 -0.784 0.432848  
## qual7GCSE or equivalent -2.964e-01 2.699e-01 -1.098 0.272167  
## qual7Higher education and above -7.678e-01 2.599e-01 -2.954 0.003162  
## qual7Still in FT education -2.091e+00 4.755e-01 -4.397 1.13e-05  
## NumCh118Have Children 4.388e-02 2.015e-01 0.218 0.827616  
## cigst2Smoker -8.238e-01 2.020e-01 -4.079 4.64e-05  
## Saturatedfattyacidsg -3.421e-02 1.664e-02 -2.055 0.039941  
## FreeSugarsg -2.762e-02 4.899e-03 -5.639 1.87e-08  
## EthGrpGB2 -2.648e+00 3.052e+00 -0.868 0.385698  
## EthGrpGB3 2.436e+00 2.682e+00 0.908 0.363916  
## EthGrpGB4 2.372e+02 2.171e+02 1.093 0.274678  
## eqvinc -2.919e-05 1.066e-05 -2.738 0.006210  
## Saturatedfattyacidsg:EthGrpGB2 9.573e-02 1.564e-01 0.612 0.540607  
## Saturatedfattyacidsg:EthGrpGB3 -2.269e-01 1.772e-01 -1.280 0.200603  
## Saturatedfattyacidsg:EthGrpGB4 -4.944e+00 4.134e+00 -1.196 0.231821  
## FreeSugarsg:EthGrpGB2 -1.585e-02 3.470e-02 -0.457 0.647960  
## FreeSugarsg:EthGrpGB3 4.699e-02 5.023e-02 0.936 0.349577  
## FreeSugarsg:EthGrpGB4 -4.137e+00 3.912e+00 -1.057 0.290388  
## Saturatedfattyacidsg:eqvinc 3.988e-08 4.028e-07 0.099 0.921140  
## FreeSugarsg:eqvinc 2.730e-07 1.190e-07 2.295 0.021830  
## EthGrpGB2:eqvinc 1.417e-04 1.114e-04 1.272 0.203462  
## EthGrpGB3:eqvinc -5.725e-05 7.804e-05 -0.734 0.463289  
## EthGrpGB4:eqvinc -7.372e-03 6.188e-03 -1.191 0.233597  
## Saturatedfattyacidsg:EthGrpGB2:eqvinc -7.263e-06 5.994e-06 -1.212 0.225699  
## Saturatedfattyacidsg:EthGrpGB3:eqvinc 3.659e-06 5.069e-06 0.722 0.470530  
## Saturatedfattyacidsg:EthGrpGB4:eqvinc -1.618e-04 2.528e-04 -0.640 0.522202  
## FreeSugarsg:EthGrpGB2:eqvinc 1.778e-06 1.202e-06 1.478 0.139397  
## FreeSugarsg:EthGrpGB3:eqvinc -7.542e-07 1.388e-06 -0.543 0.587066  
## FreeSugarsg:EthGrpGB4:eqvinc 3.017e-04 2.878e-04 1.048 0.294699  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## SexFemale \*\*\*  
## qual7A level or equivalent \*\*   
## qual7Foreign or other qualifications   
## qual7GCSE or equivalent   
## qual7Higher education and above \*\*   
## qual7Still in FT education \*\*\*  
## NumCh118Have Children   
## cigst2Smoker \*\*\*  
## Saturatedfattyacidsg \*   
## FreeSugarsg \*\*\*  
## EthGrpGB2   
## EthGrpGB3   
## EthGrpGB4   
## eqvinc \*\*   
## Saturatedfattyacidsg:EthGrpGB2   
## Saturatedfattyacidsg:EthGrpGB3   
## Saturatedfattyacidsg:EthGrpGB4   
## FreeSugarsg:EthGrpGB2   
## FreeSugarsg:EthGrpGB3   
## FreeSugarsg:EthGrpGB4   
## Saturatedfattyacidsg:eqvinc   
## FreeSugarsg:eqvinc \*   
## EthGrpGB2:eqvinc   
## EthGrpGB3:eqvinc   
## EthGrpGB4:eqvinc   
## Saturatedfattyacidsg:EthGrpGB2:eqvinc   
## Saturatedfattyacidsg:EthGrpGB3:eqvinc   
## Saturatedfattyacidsg:EthGrpGB4:eqvinc   
## FreeSugarsg:EthGrpGB2:eqvinc   
## FreeSugarsg:EthGrpGB3:eqvinc   
## FreeSugarsg:EthGrpGB4:eqvinc   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.548 on 2998 degrees of freedom  
## Multiple R-squared: 0.0858, Adjusted R-squared: 0.07604   
## F-statistic: 8.792 on 32 and 2998 DF, p-value: < 2.2e-16

table(FinalDataTemp$EthGrpGB)

##   
## 1 2 3 4   
## 2918 54 53 6

confint(lm4, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 2.837234e+01 3.086162e+01  
## age 1.874358e-02 4.342900e-02  
## SexFemale -1.014617e+00 -3.177332e-01  
## qual7A level or equivalent -1.384151e+00 -2.090611e-01  
## qual7Foreign or other qualifications -1.282552e+00 5.495784e-01  
## qual7GCSE or equivalent -8.256426e-01 2.327818e-01  
## qual7Higher education and above -1.277406e+00 -2.581458e-01  
## qual7Still in FT education -3.023329e+00 -1.158624e+00  
## NumCh118Have Children -3.511700e-01 4.389253e-01  
## cigst2Smoker -1.219792e+00 -4.277764e-01  
## Saturatedfattyacidsg -6.683955e-02 -1.572793e-03  
## FreeSugarsg -3.722900e-02 -1.801894e-02  
## EthGrpGB2 -8.632790e+00 3.336676e+00  
## EthGrpGB3 -2.823473e+00 7.694626e+00  
## EthGrpGB4 -1.884589e+02 6.627748e+02  
## eqvinc -5.008443e-05 -8.288300e-06  
## Saturatedfattyacidsg:EthGrpGB2 -2.109920e-01 4.024525e-01  
## Saturatedfattyacidsg:EthGrpGB3 -5.743971e-01 1.206330e-01  
## Saturatedfattyacidsg:EthGrpGB4 -1.305104e+01 3.162144e+00  
## FreeSugarsg:EthGrpGB2 -8.388812e-02 5.219536e-02  
## FreeSugarsg:EthGrpGB3 -5.149533e-02 1.454806e-01  
## FreeSugarsg:EthGrpGB4 -1.180799e+01 3.533929e+00  
## Saturatedfattyacidsg:eqvinc -7.499050e-07 8.296636e-07  
## FreeSugarsg:eqvinc 3.970698e-08 5.062434e-07  
## EthGrpGB2:eqvinc -7.669774e-05 3.600073e-04  
## EthGrpGB3:eqvinc -2.102678e-04 9.577465e-05  
## EthGrpGB4:eqvinc -1.950457e-02 4.760658e-03  
## Saturatedfattyacidsg:EthGrpGB2:eqvinc -1.901619e-05 4.489500e-06  
## Saturatedfattyacidsg:EthGrpGB3:eqvinc -6.281257e-06 1.359855e-05  
## Saturatedfattyacidsg:EthGrpGB4:eqvinc -6.575708e-04 3.339262e-04  
## FreeSugarsg:EthGrpGB2:eqvinc -5.799085e-07 4.135035e-06  
## FreeSugarsg:EthGrpGB3:eqvinc -3.476584e-06 1.968291e-06  
## FreeSugarsg:EthGrpGB4:eqvinc -2.627198e-04 8.660782e-04

# finding the mean income of each tertile   
data\_mean <- aggregate(x = eqvinc ~ tertiles,   
 data = FinalDataTemp,  
 FUN = mean)

#### Model diagnostics

autoplot(lm4)

