4810\_Starbucks Project – final report

Group 7

2024-03-19

# Read and Introduce our Dataset

star = read.csv('..//data//starbucks\_drinkMenu\_expanded.csv')  
head(star)

## Beverage\_category Beverage Beverage\_prep Calories  
## 1 Coffee Brewed Coffee Short 3  
## 2 Coffee Brewed Coffee Tall 4  
## 3 Coffee Brewed Coffee Grande 5  
## 4 Coffee Brewed Coffee Venti 5  
## 5 Classic Espresso Drinks Caffè Latte Short Nonfat Milk 70  
## 6 Classic Espresso Drinks Caffè Latte 2% Milk 100  
## Total.Fat..g. Trans.Fat..g. Saturated.Fat..g. Sodium..mg.  
## 1 0.1 0.0 0.0 0  
## 2 0.1 0.0 0.0 0  
## 3 0.1 0.0 0.0 0  
## 4 0.1 0.0 0.0 0  
## 5 0.1 0.1 0.0 5  
## 6 3.5 2.0 0.1 15  
## Total.Carbohydrates..g. Cholesterol..mg. Dietary.Fibre..g. Sugars..g.  
## 1 5 0 0 0  
## 2 10 0 0 0  
## 3 10 0 0 0  
## 4 10 0 0 0  
## 5 75 10 0 9  
## 6 85 10 0 9  
## Protein..g. Vitamin.A....DV. Vitamin.C....DV. Calcium....DV. Iron....DV.  
## 1 0.3 0% 0% 0% 0%  
## 2 0.5 0% 0% 0% 0%  
## 3 1.0 0% 0% 0% 0%  
## 4 1.0 0% 0% 2% 0%  
## 5 6.0 10% 0% 20% 0%  
## 6 6.0 10% 0% 20% 0%  
## Caffeine..mg.  
## 1 175  
## 2 260  
## 3 330  
## 4 410  
## 5 75  
## 6 75

# Data Structure

str(star)

## 'data.frame': 242 obs. of 18 variables:  
## $ Beverage\_category : chr "Coffee" "Coffee" "Coffee" "Coffee" ...  
## $ Beverage : chr "Brewed Coffee" "Brewed Coffee" "Brewed Coffee" "Brewed Coffee" ...  
## $ Beverage\_prep : chr "Short" "Tall" "Grande" "Venti" ...  
## $ Calories : int 3 4 5 5 70 100 70 100 150 110 ...  
## $ Total.Fat..g. : chr "0.1" "0.1" "0.1" "0.1" ...  
## $ Trans.Fat..g. : num 0 0 0 0 0.1 2 0.4 0.2 3 0.5 ...  
## $ Saturated.Fat..g. : num 0 0 0 0 0 0.1 0 0 0.2 0 ...  
## $ Sodium..mg. : int 0 0 0 0 5 15 0 5 25 0 ...  
## $ Total.Carbohydrates..g.: int 5 10 10 10 75 85 65 120 135 105 ...  
## $ Cholesterol..mg. : int 0 0 0 0 10 10 6 15 15 10 ...  
## $ Dietary.Fibre..g. : int 0 0 0 0 0 0 1 0 0 1 ...  
## $ Sugars..g. : int 0 0 0 0 9 9 4 14 14 6 ...  
## $ Protein..g. : num 0.3 0.5 1 1 6 6 5 10 10 8 ...  
## $ Vitamin.A....DV. : chr "0%" "0%" "0%" "0%" ...  
## $ Vitamin.C....DV. : chr "0%" "0%" "0%" "0%" ...  
## $ Calcium....DV. : chr "0%" "0%" "0%" "2%" ...  
## $ Iron....DV. : chr "0%" "0%" "0%" "0%" ...  
## $ Caffeine..mg. : chr "175" "260" "330" "410" ...

# Convert column Total.Fat..g. / Vitamin.A….DV./ Vitamin.C….DV. / Calcium….DV./ Iron….DV./Caffeine..mg.

# from chr to num

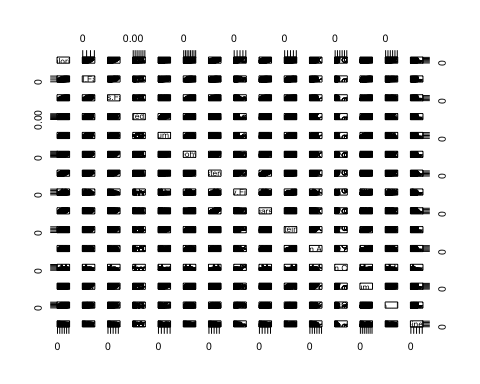
# Replace all NA values with 0 in the entire dataframe  
star[is.na(star)] <- 0  
star$Total.Fat..g.[238] = 3.2  
star$Total.Fat..g. = as.numeric(star$Total.Fat..g.)  
star$Vitamin.A....DV. = as.numeric(sub("%", "", star$Vitamin.A....DV.))  
star$Vitamin.C....DV. = as.numeric(sub("%", "", star$Vitamin.C....DV.))  
star$Calcium....DV. = as.numeric(sub("%", "", star$Calcium....DV.))  
star$Iron....DV. = as.numeric(sub("%", "", star$Iron....DV.))  
star$Caffeine..mg. = as.numeric(sub("%", "", star$Caffeine..mg.))

## Warning: NAs introduced by coercion

star[is.na(star)] <- 0

# get all the scatter plot of numerical variables

numeric\_vars <- sapply(star, is.numeric)  
pairs(star[, numeric\_vars])



# according to drink type, combine some levels of the Beverage category

Frappuccino = c('Frappuccino® Blended Coffee','Frappuccino® Blended Crème','Frappuccino® Light Blended Coffee')  
Espresso = c('Classic Espresso Drinks','Coffee','Signature Espresso Drinks')  
# others: Shaken Iced Beverages, Tazo® Tea Drinks, Smoothies   
star$Drink.Type = ifelse(star$Beverage\_category %in% Frappuccino,'Frappuccino',ifelse(star$Beverage\_category %in% Espresso,'Espresso','non-coffee'))

# from literature review, we omit sugar , Dietary.Fibre..g., trans fat, saturated fat, because Total.Fat..g. and Total.Carbohydrates..g. can represent them

# Model selection

library(MASS)  
MASS::stepAIC(lm(Calories~Total.Fat..g. +   
 Sodium..mg. + Total.Carbohydrates..g. +   
 Cholesterol..mg. + Protein..g. +   
 Vitamin.A....DV. + Vitamin.C....DV. + Calcium....DV. + Iron....DV. +   
 Caffeine..mg. + Drink.Type,data=star),direction="both")

## Start: AIC=807.98  
## Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Protein..g. + Vitamin.A....DV. + Vitamin.C....DV. +   
## Calcium....DV. + Iron....DV. + Caffeine..mg. + Drink.Type  
##   
## Df Sum of Sq RSS AIC  
## <none> 6126 807.98  
## - Vitamin.A....DV. 1 61 6187 808.38  
## - Caffeine..mg. 1 116 6242 810.52  
## - Total.Carbohydrates..g. 1 266 6392 816.27  
## - Drink.Type 2 388 6514 818.86  
## - Protein..g. 1 627 6753 829.56  
## - Sodium..mg. 1 765 6891 834.47  
## - Vitamin.C....DV. 1 788 6914 835.26  
## - Calcium....DV. 1 3020 9146 902.97  
## - Iron....DV. 1 3568 9694 917.05  
## - Total.Fat..g. 1 38157 44283 1284.68  
## - Cholesterol..mg. 1 460067 466193 1854.35

##   
## Call:  
## lm(formula = Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Protein..g. + Vitamin.A....DV. + Vitamin.C....DV. +   
## Calcium....DV. + Iron....DV. + Caffeine..mg. + Drink.Type,   
## data = star)  
##   
## Coefficients:  
## (Intercept) Total.Fat..g. Sodium..mg.   
## -0.33750 9.68492 -0.38500   
## Total.Carbohydrates..g. Cholesterol..mg. Protein..g.   
## 0.04972 3.88229 1.16408   
## Vitamin.A....DV. Vitamin.C....DV. Calcium....DV.   
## 0.12051 0.21517 0.74777   
## Iron....DV. Caffeine..mg. Drink.TypeFrappuccino   
## -0.63385 0.01218 -2.20099   
## Drink.Typenon-coffee   
## 3.24672

# from the stepAIC, we get model fit1, Caffeine and Vitamin A were removed

fit1 = lm(formula = Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
 Cholesterol..mg. + Protein..g. + Vitamin.C....DV. + Calcium....DV. +   
 Iron....DV. + Drink.Type, data = star)  
  
summary(fit1)

##   
## Call:  
## lm(formula = Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Protein..g. + Vitamin.C....DV. + Calcium....DV. +   
## Iron....DV. + Drink.Type, data = star)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.4136 -3.3026 -0.3737 3.1825 23.9667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.06969 0.91941 1.163 0.24584   
## Total.Fat..g. 9.49258 0.23347 40.659 < 2e-16 \*\*\*  
## Sodium..mg. -0.34581 0.06755 -5.119 6.46e-07 \*\*\*  
## Total.Carbohydrates..g. 0.06257 0.01500 4.173 4.26e-05 \*\*\*  
## Cholesterol..mg. 3.87007 0.02947 131.304 < 2e-16 \*\*\*  
## Protein..g. 1.14992 0.24084 4.775 3.20e-06 \*\*\*  
## Vitamin.C....DV. 0.22633 0.03824 5.918 1.17e-08 \*\*\*  
## Calcium....DV. 0.76986 0.06626 11.619 < 2e-16 \*\*\*  
## Iron....DV. -0.60842 0.05399 -11.270 < 2e-16 \*\*\*  
## Drink.TypeFrappuccino -4.25453 2.41608 -1.761 0.07957 .   
## Drink.Typenon-coffee 2.58841 0.92413 2.801 0.00553 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.221 on 231 degrees of freedom  
## Multiple R-squared: 0.9975, Adjusted R-squared: 0.9974   
## F-statistic: 9331 on 10 and 231 DF, p-value: < 2.2e-16

library(car)

## Loading required package: carData

car::vif(fit1)

## GVIF Df GVIF^(1/(2\*Df))  
## Total.Fat..g. 4.177547 1 2.043905  
## Sodium..mg. 3.004762 1 1.733425  
## Total.Carbohydrates..g. 13.466100 1 3.669619  
## Cholesterol..mg. 3.321110 1 1.822391  
## Protein..g. 12.170173 1 3.488578  
## Vitamin.C....DV. 2.689424 1 1.639946  
## Calcium....DV. 8.208210 1 2.864997  
## Iron....DV. 2.833330 1 1.683250  
## Drink.Type 13.266348 2 1.908481

# Detected multicollinearity between several variables​

# Total Fat and Total Carbohydrates are essential to our model​

# => Removed Protein and Calcium from the model

# fit2 , model without Calcium….DV. , Protein..g.

fit2 = lm(formula = Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
 Cholesterol..mg. + Vitamin.C....DV. +   
 Iron....DV. + Drink.Type, data = star)  
  
summary(fit2)

##   
## Call:  
## lm(formula = Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Vitamin.C....DV. + Iron....DV. + Drink.Type,   
## data = star)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -20.6639 -5.9908 -0.4518 5.3581 24.5926   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.42696 1.34204 6.279 1.65e-09 \*\*\*  
## Total.Fat..g. 9.07731 0.34866 26.035 < 2e-16 \*\*\*  
## Sodium..mg. -0.17925 0.09895 -1.811 0.0714 .   
## Total.Carbohydrates..g. 0.27799 0.01513 18.375 < 2e-16 \*\*\*  
## Cholesterol..mg. 3.77759 0.04693 80.496 < 2e-16 \*\*\*  
## Vitamin.C....DV. 0.20838 0.04194 4.968 1.31e-06 \*\*\*  
## Iron....DV. -0.32015 0.07691 -4.163 4.42e-05 \*\*\*  
## Drink.TypeFrappuccino -41.74122 2.08733 -19.997 < 2e-16 \*\*\*  
## Drink.Typenon-coffee 1.74793 1.48235 1.179 0.2395   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.428 on 233 degrees of freedom  
## Multiple R-squared: 0.9935, Adjusted R-squared: 0.9933   
## F-statistic: 4458 on 8 and 233 DF, p-value: < 2.2e-16

library(car)  
car::vif(fit2)

## GVIF Df GVIF^(1/(2\*Df))  
## Total.Fat..g. 3.575761 1 1.890968  
## Sodium..mg. 2.474530 1 1.573064  
## Total.Carbohydrates..g. 5.260239 1 2.293521  
## Cholesterol..mg. 3.231318 1 1.797587  
## Vitamin.C....DV. 1.241304 1 1.114138  
## Iron....DV. 2.206756 1 1.485516  
## Drink.Type 3.725340 2 1.389285

# compare fit1, fit2

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

lmtest::lrtest(fit2,fit1)

## Likelihood ratio test  
##   
## Model 1: Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Vitamin.C....DV. + Iron....DV. + Drink.Type  
## Model 2: Calories ~ Total.Fat..g. + Sodium..mg. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Protein..g. + Vitamin.C....DV. + Calcium....DV. +   
## Iron....DV. + Drink.Type  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 -854.63   
## 2 12 -737.71 2 233.84 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Anova(fit2)

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include  
## arithmetic operators in their names;  
## the printed representation of the hypothesis will be omitted

## Anova Table (Type II tests)  
##   
## Response: Calories  
## Sum Sq Df F value Pr(>F)   
## Total.Fat..g. 48145 1 677.8134 < 2.2e-16 \*\*\*  
## Sodium..mg. 233 1 3.2814 0.07136 .   
## Total.Carbohydrates..g. 23984 1 337.6589 < 2.2e-16 \*\*\*  
## Cholesterol..mg. 460248 1 6479.6668 < 2.2e-16 \*\*\*  
## Vitamin.C....DV. 1753 1 24.6849 1.307e-06 \*\*\*  
## Iron....DV. 1231 1 17.3297 4.425e-05 \*\*\*  
## Drink.Type 30124 2 212.0516 < 2.2e-16 \*\*\*  
## Residuals 16550 233   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

fit3 , model without Sodium..mg.

fit3 = lm(formula = Calories ~ Total.Fat..g. + Total.Carbohydrates..g. +   
 Cholesterol..mg. + Vitamin.C....DV. +   
 Iron....DV. + Drink.Type, data = star)  
  
summary(fit3)

##   
## Call:  
## lm(formula = Calories ~ Total.Fat..g. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Vitamin.C....DV. + Iron....DV. + Drink.Type,   
## data = star)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.7475 -6.0205 -0.4253 5.0416 25.1299   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.28547 1.34628 6.154 3.24e-09 \*\*\*  
## Total.Fat..g. 8.69984 0.28090 30.971 < 2e-16 \*\*\*  
## Total.Carbohydrates..g. 0.27310 0.01496 18.258 < 2e-16 \*\*\*  
## Cholesterol..mg. 3.77604 0.04715 80.087 < 2e-16 \*\*\*  
## Vitamin.C....DV. 0.20382 0.04207 4.845 2.31e-06 \*\*\*  
## Iron....DV. -0.24055 0.06342 -3.793 0.00019 \*\*\*  
## Drink.TypeFrappuccino -40.93062 2.04872 -19.979 < 2e-16 \*\*\*  
## Drink.Typenon-coffee 1.75659 1.48955 1.179 0.23949   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.469 on 234 degrees of freedom  
## Multiple R-squared: 0.9934, Adjusted R-squared: 0.9932   
## F-statistic: 5046 on 7 and 234 DF, p-value: < 2.2e-16

Anova(fit3)

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include  
## arithmetic operators in their names;  
## the printed representation of the hypothesis will be omitted

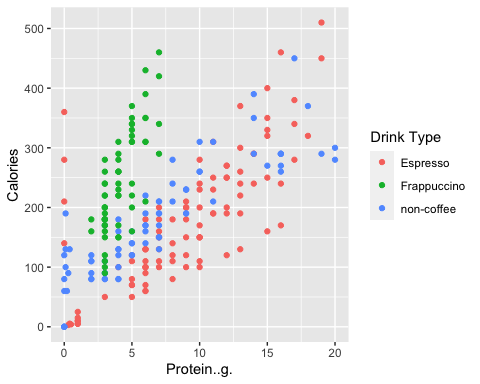
## Anova Table (Type II tests)  
##   
## Response: Calories  
## Sum Sq Df F value Pr(>F)   
## Total.Fat..g. 68797 1 959.214 < 2.2e-16 \*\*\*  
## Total.Carbohydrates..g. 23909 1 333.355 < 2.2e-16 \*\*\*  
## Cholesterol..mg. 460024 1 6413.976 < 2.2e-16 \*\*\*  
## Vitamin.C....DV. 1684 1 23.473 2.306e-06 \*\*\*  
## Iron....DV. 1032 1 14.386 0.0001896 \*\*\*  
## Drink.Type 30417 2 212.045 < 2.2e-16 \*\*\*  
## Residuals 16783 234   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# we know that different drink type may have different ingredients. For example, smoothies may have more sugar than classic expresso coffee.

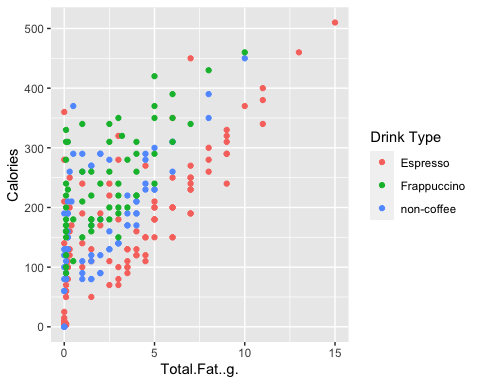
# so we want to know is there interaction between Drink.Type and other explanatory variables

library(ggplot2)  
var\_names = c('Protein..g.', 'Total.Fat..g.' , 'Total.Carbohydrates..g.' , 'Cholesterol..mg.', 'Vitamin.C....DV.' , 'Iron....DV.')  
for (variable\_name in var\_names) {  
 p = ggplot(star, aes(x = star[[variable\_name]], y = Calories, color = Drink.Type)) +  
 geom\_point(aes(color = Drink.Type)) + # Color points by Drink.Type  
 # geom\_smooth(method = "lm", se = FALSE) +   
 labs(x = variable\_name, y = "Calories", color = "Drink Type")  
 print(p)  
  
}

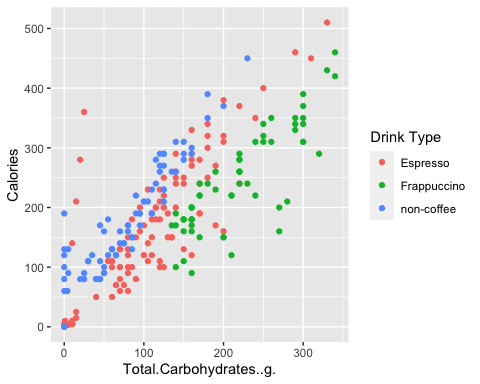
## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



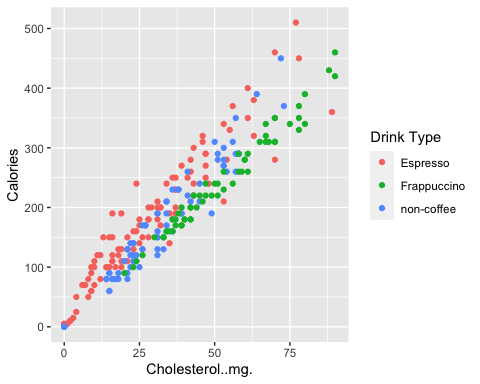
## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



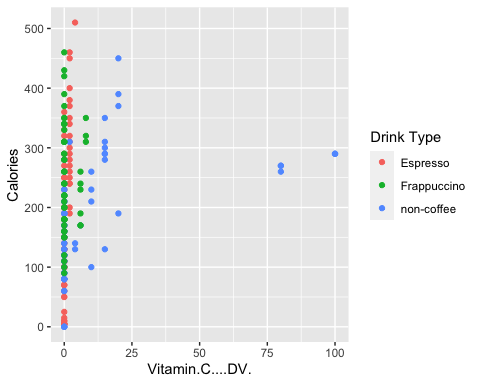
## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



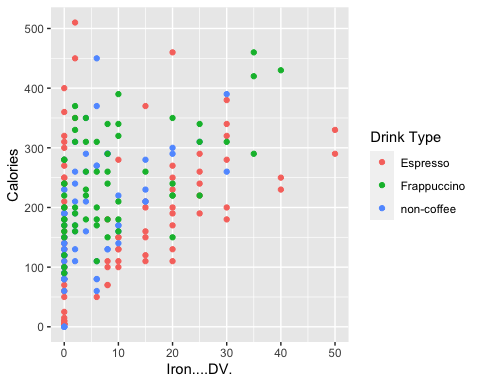
## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



## Warning: Use of `star[[variable\_name]]` is discouraged.  
## ℹ Use `.data[[variable\_name]]` instead.



# from Scatter plot above, we can find that there seems interaction between Drink type and Total.carbonhydrates;

#fit4 ,model with interaction between Drink type and Total.carbonhydrates;

fit4 = lm(formula = Calories ~ Total.Fat..g. + Total.Carbohydrates..g. +   
 Cholesterol..mg. + Vitamin.C....DV. +   
 Iron....DV. + Drink.Type +Drink.Type:Total.Carbohydrates..g., data = star)  
  
summary(fit4)

##   
## Call:  
## lm(formula = Calories ~ Total.Fat..g. + Total.Carbohydrates..g. +   
## Cholesterol..mg. + Vitamin.C....DV. + Iron....DV. + Drink.Type +   
## Drink.Type:Total.Carbohydrates..g., data = star)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.8154 -4.2324 -0.7367 3.4898 24.1859   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 4.26391 1.22764 3.473  
## Total.Fat..g. 7.91726 0.21268 37.226  
## Total.Carbohydrates..g. 0.31821 0.01279 24.880  
## Cholesterol..mg. 3.80784 0.03415 111.499  
## Vitamin.C....DV. 0.08612 0.03174 2.713  
## Iron....DV. -0.12194 0.04674 -2.609  
## Drink.TypeFrappuccino -3.31367 3.24319 -1.022  
## Drink.Typenon-coffee -2.14766 1.72986 -1.242  
## Total.Carbohydrates..g.:Drink.TypeFrappuccino -0.20184 0.01685 -11.978  
## Total.Carbohydrates..g.:Drink.Typenon-coffee 0.07084 0.01631 4.343  
## Pr(>|t|)   
## (Intercept) 0.000613 \*\*\*  
## Total.Fat..g. < 2e-16 \*\*\*  
## Total.Carbohydrates..g. < 2e-16 \*\*\*  
## Cholesterol..mg. < 2e-16 \*\*\*  
## Vitamin.C....DV. 0.007163 \*\*   
## Iron....DV. 0.009668 \*\*   
## Drink.TypeFrappuccino 0.307973   
## Drink.Typenon-coffee 0.215667   
## Total.Carbohydrates..g.:Drink.TypeFrappuccino < 2e-16 \*\*\*  
## Total.Carbohydrates..g.:Drink.Typenon-coffee 2.1e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.093 on 232 degrees of freedom  
## Multiple R-squared: 0.9966, Adjusted R-squared: 0.9965   
## F-statistic: 7606 on 9 and 232 DF, p-value: < 2.2e-16

Anova(fit4)

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include  
## arithmetic operators in their names;  
## the printed representation of the hypothesis will be omitted  
  
## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include  
## arithmetic operators in their names;  
## the printed representation of the hypothesis will be omitted  
  
## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include  
## arithmetic operators in their names;  
## the printed representation of the hypothesis will be omitted

## Anova Table (Type II tests)  
##   
## Response: Calories  
## Sum Sq Df F value Pr(>F)   
## Total.Fat..g. 51451 1 1385.7858 < 2.2e-16 \*\*\*  
## Total.Carbohydrates..g. 23909 1 643.9662 < 2.2e-16 \*\*\*  
## Cholesterol..mg. 461571 1 12432.0257 < 2.2e-16 \*\*\*  
## Vitamin.C....DV. 273 1 7.3617 0.007163 \*\*   
## Iron....DV. 253 1 6.8074 0.009668 \*\*   
## Drink.Type 30417 2 409.6227 < 2.2e-16 \*\*\*  
## Total.Carbohydrates..g.:Drink.Type 8169 2 110.0173 < 2.2e-16 \*\*\*  
## Residuals 8614 232   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

H0: fit3 better, beta\_interaction = 0 Ha: fit4 better, beta\_interaction != 0

lrtest(fit3,fit4)

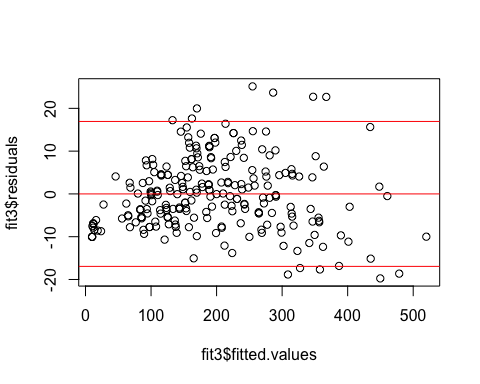
## Likelihood ratio test  
##   
## Model 1: Calories ~ Total.Fat..g. + Total.Carbohydrates..g. + Cholesterol..mg. +   
## Vitamin.C....DV. + Iron....DV. + Drink.Type  
## Model 2: Calories ~ Total.Fat..g. + Total.Carbohydrates..g. + Cholesterol..mg. +   
## Vitamin.C....DV. + Iron....DV. + Drink.Type + Drink.Type:Total.Carbohydrates..g.  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 9 -856.32   
## 2 11 -775.61 2 161.42 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## since lrtest p-value < 0.05,we reject H0, and conclude that model with interaction between Drink type and Total.carbonhydrates is better.

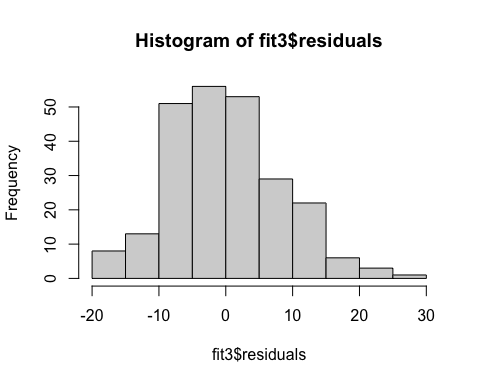
## However, R2 adjusted only increased by 0.32%​ Since simple model is preferred. We stay with Model 3 as final model.

### Residuals

s = summary(fit3)$sigma  
plot(fit3$fitted.values, fit3$residuals)  
abline(h = 0, col = "red")  
abline(h = c(-2, 2)\*s, col = "red")



hist(fit3$residuals)



qqnorm(fit3$residuals)  
qqline(fit3$residuals,col='red')

