

Association Between Dietary Patterns and Blood Lipid Profiles Among Chinese Women

Stat Project

Contents

1	Introduction	1
2	Data and Methods	2
2.1	Data Loading	2
2.2	Data Filtering and Merging	3
3	Identification of Dietary Patterns	4
3.1	Selection and Standardization of Dietary Variables	4
3.2	Principal Component Analysis	5
3.3	Scree Plot	5
3.4	Factor Loadings	6
4	PCA Scores and Quartile Classification	7
5	Distribution and Normality Assessment	8
6	Simple Linear Regression: Fat Intake and Triglycerides	9
7	Multiple Linear Regression: Dietary Patterns and Lipids	10
8	Diagnostic Tests	12
9	ANOVA and Non-Parametric Tests	13
10	Discussion	15
11	Conclusion	15

1 Introduction

Dietary habits play a central role in lipid metabolism and cardiovascular health. Traditional nutritional analyses often focus on individual nutrients; however, this approach does not reflect real-life dietary behavior where nutrients are consumed in combination.

Dietary pattern analysis, particularly using Principal Component Analysis (PCA), allows identification of dominant eating patterns within a population. These patterns can then be related to biological outcomes such as blood lipid profiles.

The objective of this study is to identify dietary patterns among Chinese women and examine their associations with blood lipid biomarkers, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TG), and total cholesterol (TC). This analysis follows the methodological framework of the article “*Association between dietary patterns and blood lipid profiles among Chinese women*”.

2 Data and Methods

2.1 Data Loading

```
bio <- read.csv("biomarker.csv")
diet <- read.csv("c12diet.csv")

str(bio)

## 'data.frame':   9549 obs. of  49 variables:
## $ IDind      : num  2.11e+11 2.11e+11 2.11e+11 2.11e+11 2.11e+11 ...
## $ UREA       : num  7.24 6.06 7.31 5.92 5.53 NA 5.38 6.46 4.71 6.15 ...
## $ UA         : int  452 200 339 366 279 NA 341 236 197 398 ...
## $ APO_A      : num  1.95 1.79 1.35 1.63 2.39 NA 2.18 1.41 1.35 1.79 ...
## $ LP_A       : int  344 400 10 215 77 NA 187 118 79 34 ...
## $ HS_CRP    : int  1 1 1 3 2 NA 1 1 3 1 ...
## $ CRE        : num  83 71 100 75 70 NA 67 72 76 89 ...
## $ HDL_C      : num  1.15 1.4 0.98 0.84 1.68 NA 1.42 1.19 1.03 1.02 ...
## $ LDL_C      : num  3.71 3.54 2.66 2.59 2.58 NA 3.98 1.83 1.33 2.42 ...
## $ APO_B      : num  1.47 1.22 0.94 1.19 0.89 NA 1.41 0.68 0.58 1.32 ...
## $ MG         : num  0.97 0.84 0.91 0.98 0.93 NA 0.93 0.88 0.85 1.04 ...
## $ FET        : num  154.6 45.9 168.8 106 32.1 ...
## $ INS        : num  12.08 11.21 12.34 18.32 5.57 ...
## $ HGB        : int  142 152 166 134 124 NA 134 120 122 157 ...
## $ WBC        : num  4.2 4 7.5 5.4 4.4 NA 5.8 4.2 5.6 4.5 ...
## $ RBC        : num  4.3 4.9 5.6 4.5 4.1 NA 4.8 4.1 4.1 5.3 ...
## $ PLT        : int  173 221 166 393 253 NA 250 190 22 213 ...
## $ Glu_field  : num  5.02 5.79 4.98 5.99 NA NA 4.83 4.81 5.06 6.01 ...
## $ Y48_2      : int  51 13 21 17 NA NA 16 16 7 29 ...
## $ Y48_3      : num  81.5 76.2 85.5 82 NA NA 72 74.6 75.6 70.5 ...
## $ Y48_4      : num  51 45.4 46 45.6 NA NA 46.5 46.8 46.7 43.4 ...
## $ Y48_5      : num  6.63 6.01 4.35 5.52 NA NA 6.52 3.85 2.69 6.06 ...
## $ Y48_6      : num  2.55 1.32 0.8 3.17 NA NA 1.27 0.99 0.39 4.01 ...
## $ HbA1c     : num  5.2 5.4 5 5.5 NA NA 5.1 5.7 5.4 5.9 ...
## $ TP         : num  79.2 75.6 84.1 79.8 79.3 NA 71 74.9 78.5 67.7 ...
## $ ALB        : num  51.8 45.4 46.5 48.6 47.1 NA 45 45.1 48.5 45.1 ...
## $ GLUCOSE   : num  5.18 5.3 4.49 5.99 4.56 ...
## $ TG         : num  3.13 1.24 0.94 3.16 1.26 NA 1.26 0.97 0.45 4.92 ...
## $ TC         : num  6.4 5.65 4.03 5.27 4.86 NA 6.03 3.28 2.57 5.87 ...
## $ ALT        : num  28 10 15 11 8 NA 10 11 6 21 ...
## $ TRF        : num  231 248 220 266 284 NA 279 221 232 236 ...
## $ TRF_R     : num  0.567 0.879 2.19 NA 1.14 NA 0.965 1.24 0.908 1.19 ...
## $ CRE_MG    : num  0.939 0.803 1.131 0.848 0.792 ...
## $ UA_MG     : num  7.6 3.36 5.7 6.15 4.69 ...
## $ UREA_MG   : num  43.6 36.5 44 35.7 33.3 ...
## $ MG_MG     : num  2.36 2.04 2.21 2.38 2.26 ...
## $ TC_MG     : num  247 218 156 204 188 ...
## $ Y48_5MG   : num  256 232 168 213 NA ...
## $ TG_MG     : num  277.2 109.8 83.3 279.9 111.6 ...
## $ Y48_6MG   : num  225.9 116.9 70.9 280.8 NA ...
## $ HDL_C_MG  : num  44.5 54.1 37.9 32.5 65 ...
## $ LDL_C_MG  : num  143.5 136.9 102.9 100.2 99.8 ...
## $ Glu_field_MG: num  90.4 104.2 89.6 107.8 NA ...
## $ GLUCOSE_MG: num  93.3 95.4 80.8 107.9 82.1 ...
```

```

## $ APO_A_MG : int 195 179 135 163 239 NA 218 141 135 179 ...
## $ APO_B_MG : int 147 122 94 119 89 NA 141 68 58 132 ...
## $ TRF_MG : num 231 248 220 266 284 NA 279 221 232 236 ...
## $ Y46_1DL : num 14.2 15.2 16.6 13.4 12.4 NA 13.4 12 12.2 15.7 ...
## $ wave : int 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...

str(diet)

## 'data.frame': 102575 obs. of 14 variables:
## $ IDind : num 1.11e+11 1.11e+11 1.11e+11 1.11e+11 1.11e+11 ...
## $ wave : int 2011 2011 2011 2011 2011 2011 2011 2011 2011 ...
## $ hhid : int 111101001 111101001 111101002 111101003 111101004 111101004 111101005 111101005 111101005 ...
## $ line : int 1 2 1 1 1 2 1 2 1 2 ...
## $ d3kcal : num 2274 1097 1448 1341 2113 ...
## $ d3carbo: num 270 146 170 162 216 ...
## $ d3fat : num 94.8 35.6 46.7 55.2 105.4 ...
## $ d3protn: num 84 48 86.3 48.6 74.8 ...
## $ t1 : int 11 11 11 11 11 11 11 11 11 ...
## $ t2 : int 1 1 1 1 1 1 1 1 1 ...
## $ t3 : int 1 1 1 1 1 1 1 1 1 ...
## $ t4 : int 1 1 1 1 1 1 1 1 1 ...
## $ t5 : int 1 1 2 3 4 4 5 5 6 6 ...
## $ commid : int 111101 111101 111101 111101 111101 111101 111101 111101 111101 111101 ...

The biomarker dataset contains blood lipid and biochemical measurements, while the dietary dataset contains 3-day nutrient intake data.

```

2.2 Data Filtering and Merging

```

diet <- subset(diet, wave == 2009)
data <- merge(diet, bio, by = "IDind")
str(data)

## 'data.frame': 9383 obs. of 62 variables:
## $ IDind : num 2.11e+11 2.11e+11 2.11e+11 2.11e+11 2.11e+11 ...
## $ wave.x : int 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
## $ hhid : int 211101003 211101010 211101012 211101012 211101013 211101015 211101015 211101017 ...
## $ line : int 2 1 1 2 2 1 2 2 4 1 ...
## $ d3kcal : num 2083 1951 2707 2586 2186 ...
## $ d3carbo : num 276 320 324 322 305 ...
## $ d3fat : num 46.2 57.1 116.5 107.3 68.4 ...
## $ d3protn: num 84.1 38.4 90.7 82.8 91.8 ...
## $ t1 : int 21 21 21 21 21 21 21 21 21 ...
## $ t2 : int 1 1 1 1 1 1 1 1 1 ...
## $ t3 : int 1 1 1 1 1 1 1 1 1 ...
## $ t4 : int 1 1 1 1 1 1 1 1 1 ...
## $ t5 : int 3 10 12 12 13 15 15 17 17 62 ...
## $ commid : int 211101 211101 211101 211101 211101 211101 211101 211101 211101 211101 ...
## $ UREA : num 7.24 6.06 7.31 5.92 5.53 NA 5.38 6.46 4.71 6.15 ...
## $ UA : int 452 200 339 366 279 NA 341 236 197 398 ...
## $ APO_A : num 1.95 1.79 1.35 1.63 2.39 NA 2.18 1.41 1.35 1.79 ...
## $ LP_A : int 344 400 10 215 77 NA 187 118 79 34 ...
## $ HS_CRP : int 1 1 1 3 2 NA 1 1 3 1 ...
## $ CRE : num 83 71 100 75 70 NA 67 72 76 89 ...
## $ HDL_C : num 1.15 1.4 0.98 0.84 1.68 NA 1.42 1.19 1.03 1.02 ...

```

```

## $ LDL_C      : num  3.71 3.54 2.66 2.59 2.58 NA 3.98 1.83 1.33 2.42 ...
## $ APO_B      : num  1.47 1.22 0.94 1.19 0.89 NA 1.41 0.68 0.58 1.32 ...
## $ MG         : num  0.97 0.84 0.91 0.98 0.93 NA 0.93 0.88 0.85 1.04 ...
## $ FET        : num  154.6 45.9 168.8 106 32.1 ...
## $ INS        : num  12.08 11.21 12.34 18.32 5.57 ...
## $ HGB        : int   142 152 166 134 124 NA 134 120 122 157 ...
## $ WBC        : num  4.2 4 7.5 5.4 4.4 NA 5.8 4.2 5.6 4.5 ...
## $ RBC        : num  4.3 4.9 5.6 4.5 4.1 NA 4.8 4.1 4.1 5.3 ...
## $ PLT        : int   173 221 166 393 253 NA 250 190 22 213 ...
## $ Glu_field  : num  5.02 5.79 4.98 5.99 NA NA 4.83 4.81 5.06 6.01 ...
## $ Y48_2       : int   51 13 21 17 NA NA 16 16 7 29 ...
## $ Y48_3       : num  81.5 76.2 85.5 82 NA NA 72 74.6 75.6 70.5 ...
## $ Y48_4       : num  51 45.4 46 45.6 NA NA 46.5 46.8 46.7 43.4 ...
## $ Y48_5       : num  6.63 6.01 4.35 5.52 NA NA 6.52 3.85 2.69 6.06 ...
## $ Y48_6       : num  2.55 1.32 0.8 3.17 NA NA 1.27 0.99 0.39 4.01 ...
## $ HbA1c      : num  5.2 5.4 5 5.5 NA NA 5.1 5.7 5.4 5.9 ...
## $ TP          : num  79.2 75.6 84.1 79.8 79.3 NA 71 74.9 78.5 67.7 ...
## $ ALB         : num  51.8 45.4 46.5 48.6 47.1 NA 45 45.1 48.5 45.1 ...
## $ GLUCOSE     : num  5.18 5.3 4.49 5.99 4.56 ...
## $ TG          : num  3.13 1.24 0.94 3.16 1.26 NA 1.26 0.97 0.45 4.92 ...
## $ TC          : num  6.4 5.65 4.03 5.27 4.86 NA 6.03 3.28 2.57 5.87 ...
## $ ALT         : num  28 10 15 11 8 NA 10 11 6 21 ...
## $ TRF         : num  231 248 220 266 284 NA 279 221 232 236 ...
## $ TRF_R       : num  0.567 0.879 2.19 NA 1.14 NA 0.965 1.24 0.908 1.19 ...
## $ CRE_MG      : num  0.939 0.803 1.131 0.848 0.792 ...
## $ UA_MG       : num  7.6 3.36 5.7 6.15 4.69 ...
## $ UREA_MG    : num  43.6 36.5 44 35.7 33.3 ...
## $ MG_MG       : num  2.36 2.04 2.21 2.38 2.26 ...
## $ TC_MG       : num  247 218 156 204 188 ...
## $ Y48_5MG     : num  256 232 168 213 NA ...
## $ TG_MG       : num  277.2 109.8 83.3 279.9 111.6 ...
## $ Y48_6MG     : num  225.9 116.9 70.9 280.8 NA ...
## $ HDL_C_MG   : num  44.5 54.1 37.9 32.5 65 ...
## $ LDL_C_MG   : num  143.5 136.9 102.9 100.2 99.8 ...
## $ Glu_field_MG: num  90.4 104.2 89.6 107.8 NA ...
## $ GLUCOSE_MG  : num  93.3 95.4 80.8 107.9 82.1 ...
## $ APO_A_MG    : int   195 179 135 163 239 NA 218 141 135 179 ...
## $ APO_B_MG    : int   147 122 94 119 89 NA 141 68 58 132 ...
## $ TRF_MG      : num  231 248 220 266 284 NA 279 221 232 236 ...
## $ Y46_1DL     : num  14.2 15.2 16.6 13.4 12.4 NA 13.4 12 12.2 15.7 ...
## $ wave.y      : int   2009 2009 2009 2009 2009 2009 2009 2009 2009 ...

```

3 Identification of Dietary Patterns

3.1 Selection and Standardization of Dietary Variables

Four dietary variables were selected: - Total energy intake (d3kcal) - Carbohydrate intake (d3carbo) - Fat intake (d3fat) - Protein intake (d3protn)

```

diet_vars <- data[, c("d3kcal", "d3carbo", "d3fat", "d3protn")]
diet_scaled <- scale(diet_vars)

```

3.2 Principal Component Analysis

```
library(psych)

pca <- principal(
  diet_scaled,
  nfactors = 3,
  rotate = "varimax",
  scores = TRUE
)

print(pca, digits = 3)

## Principal Components Analysis
## Call: principal(r = diet_scaled, nfactors = 3, rotate = "varimax",
##                 scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##          RC1    RC2    RC3    h2      u2   com
## d3kcal  0.672  0.616  0.403  0.994  0.006491 2.64
## d3carbo 0.958  0.052  0.278  0.998  0.002399 1.17
## d3fat   0.071  0.981  0.177  0.998  0.002042 1.08
## d3protn 0.337  0.235  0.912  1.000  0.000141 1.41
##
##          RC1    RC2    RC3
## SS loadings     1.487  1.400  1.102
## Proportion Var  0.372  0.350  0.276
## Cumulative Var 0.372  0.722  0.997
## Proportion Explained  0.373  0.351  0.276
## Cumulative Proportion 0.373  0.724  1.000
##
## Mean item complexity =  1.6
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.002
## with the empirical chi square  0.662 with prob <  NA
##
## Fit based upon off diagonal values = 1
```

Three dietary patterns were retained based on eigenvalues greater than 1 and interpretability: - Pattern 1: Carbohydrate- and energy-rich pattern - Pattern 2: Fat-rich pattern - Pattern 3: Protein-rich pattern

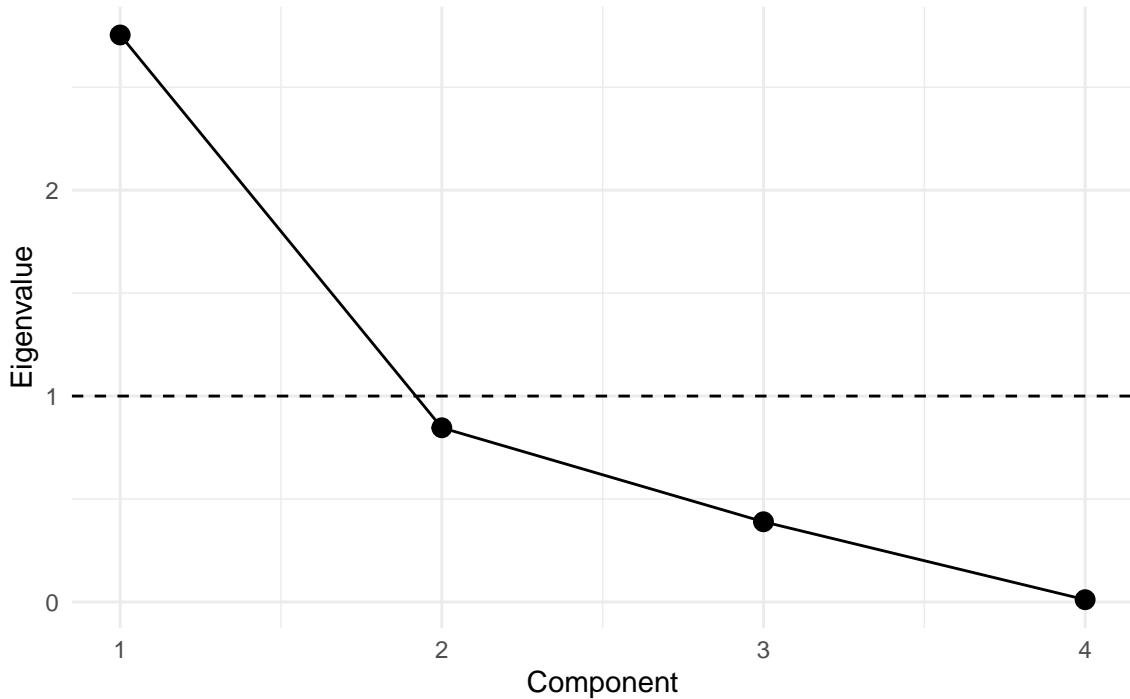
3.3 Scree Plot

```
library(ggplot2)

eig <- pca$values
df <- data.frame(Component = 1:length(eig), Eigenvalue = eig)

ggplot(df, aes(Component, Eigenvalue)) +
  geom_point(size = 3) +
  geom_line() +
  geom_hline(yintercept = 1, linetype = "dashed") +
  theme_minimal() +
  labs(title = "Scree Plot", x = "Component", y = "Eigenvalue")
```

Scree Plot



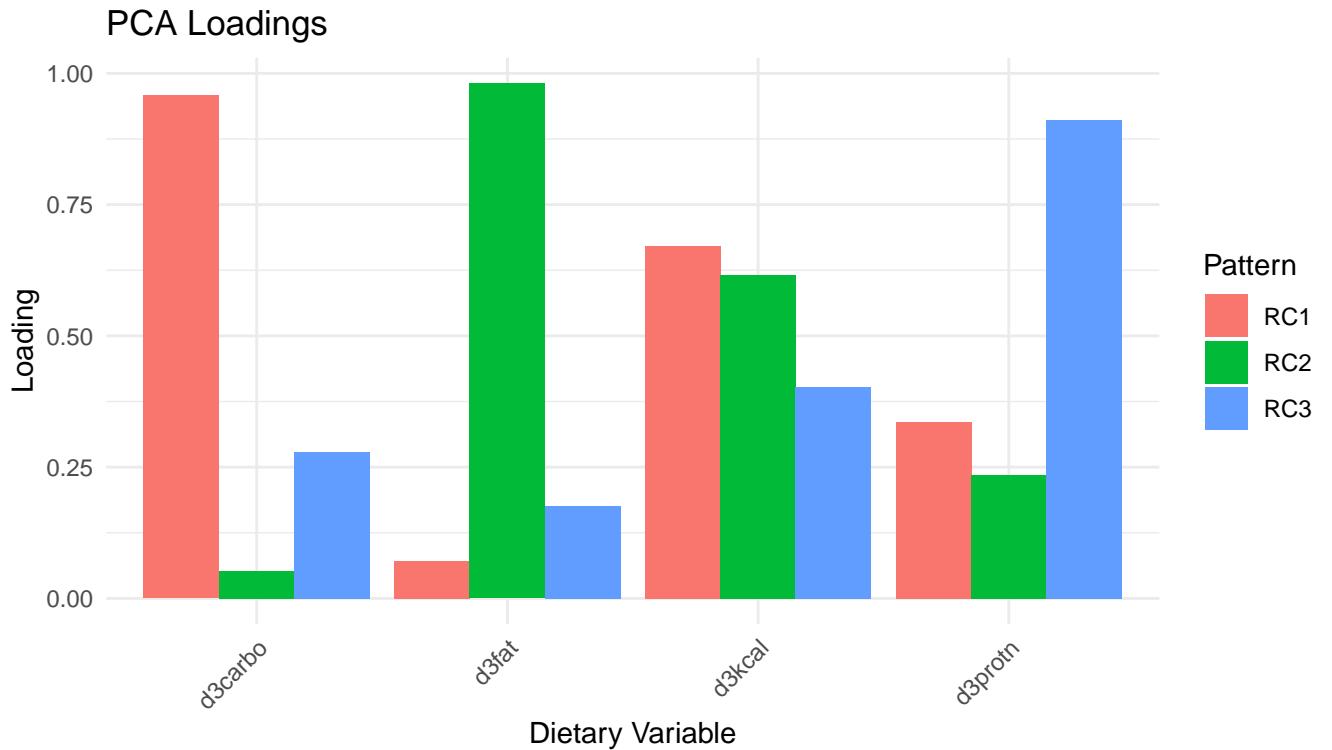
3.4 Factor Loadings

```
library(reshape2)

load_df <- as.data.frame(pca$loadings[,])
load_df$Variable <- rownames(load_df)

melted <- melt(load_df, id.vars = "Variable",
                 variable.name = "Pattern",
                 value.name = "Loading")

ggplot(melted, aes(x = Variable, y = Loading, fill = Pattern)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "PCA Loadings", x = "Dietary Variable", y = "Loading") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



4 PCA Scores and Quartile Classification

```

scores <- as.data.frame(pca$scores)
colnames(scores) <- c("Pattern1", "Pattern2", "Pattern3")
data <- cbind(data, scores)

data$Pattern1_Q <- cut(
  data$Pattern1,
  quantile(data$Pattern1, probs = seq(0,1,0.25), na.rm = TRUE),
  include.lowest = TRUE
)

data$Pattern2_Q <- cut(
  data$Pattern2,
  quantile(data$Pattern2, probs = seq(0,1,0.25), na.rm = TRUE),
  include.lowest = TRUE
)

data$Pattern3_Q <- cut(
  data$Pattern3,
  quantile(data$Pattern3, probs = seq(0,1,0.25), na.rm = TRUE),
  include.lowest = TRUE
)

table(data$Pattern1_Q)

##
##  [-2.79,-0.692] (-0.692,-0.115]  (-0.115,0.59]  (0.59,4.95]

```

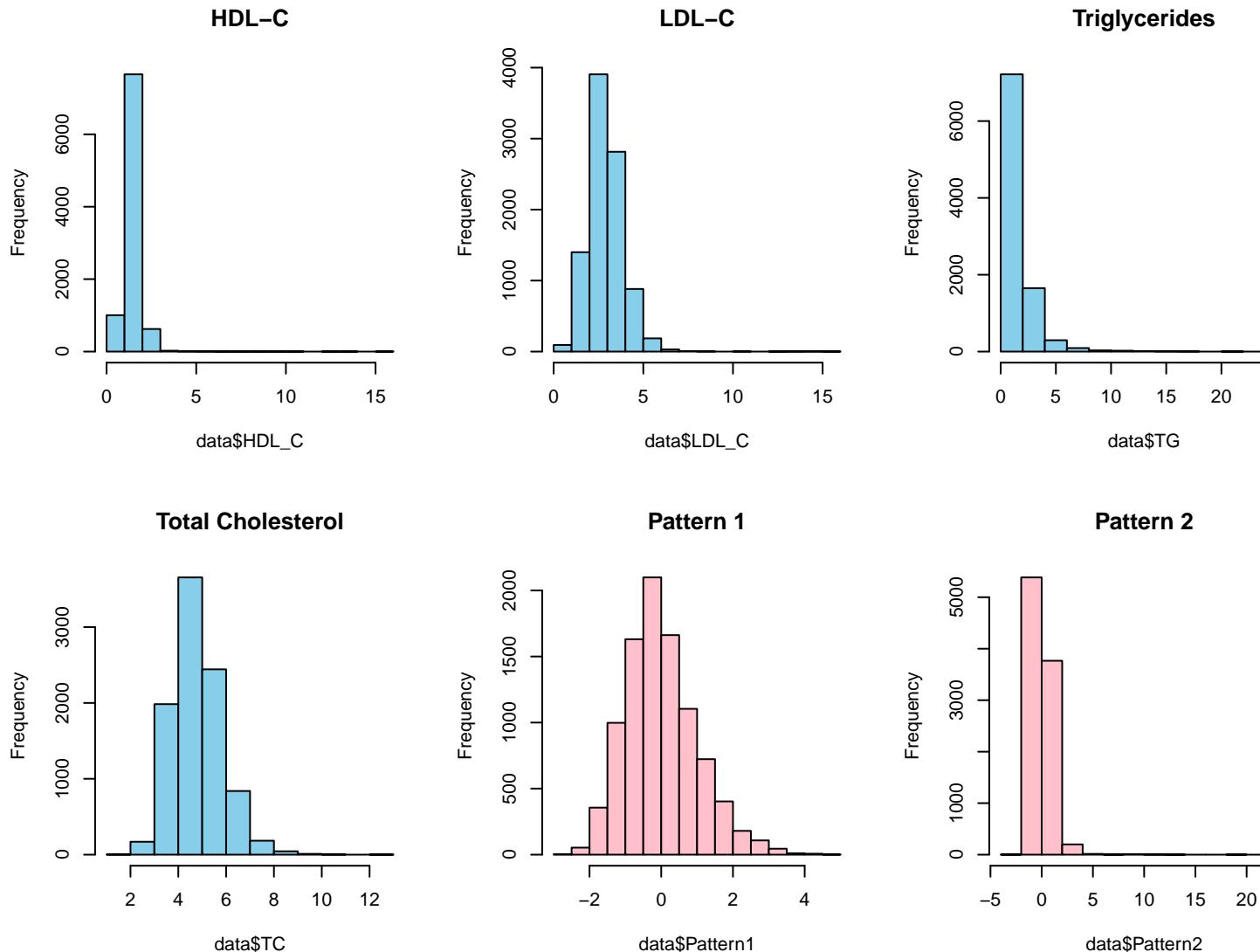
```
##          2346      2346      2345      2346
```

5 Distribution and Normality Assessment

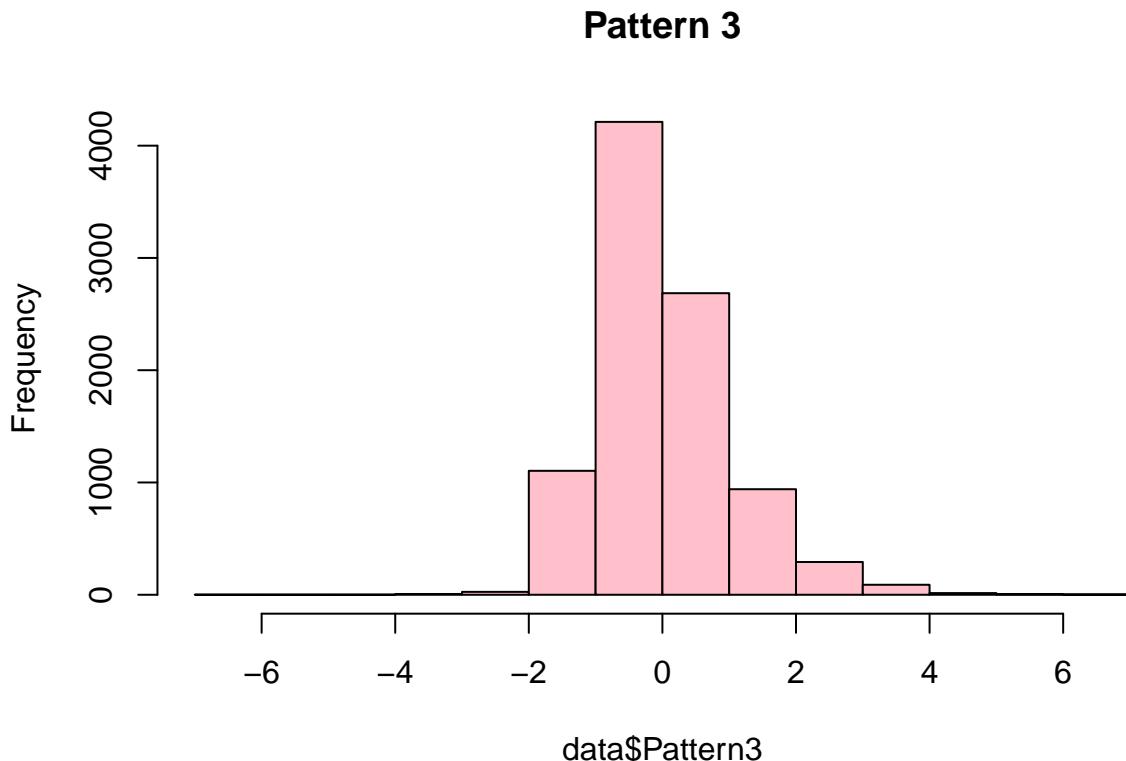
```
par(mfrow=c(2,3))

hist(data$HDL_C, main="HDL-C", col="skyblue")
hist(data$LDL_C, main="LDL-C", col="skyblue")
hist(data$TG, main="Triglycerides", col="skyblue")
hist(data$TC, main="Total Cholesterol", col="skyblue")

hist(data$Pattern1, main="Pattern 1", col="pink")
hist(data$Pattern2, main="Pattern 2", col="pink")
```



```
hist(data$Pattern3, main="Pattern 3", col="pink")
```



Triglycerides and several lipid variables exhibit right-skewed distributions, suggesting deviations from normality.

6 Simple Linear Regression: Fat Intake and Triglycerides

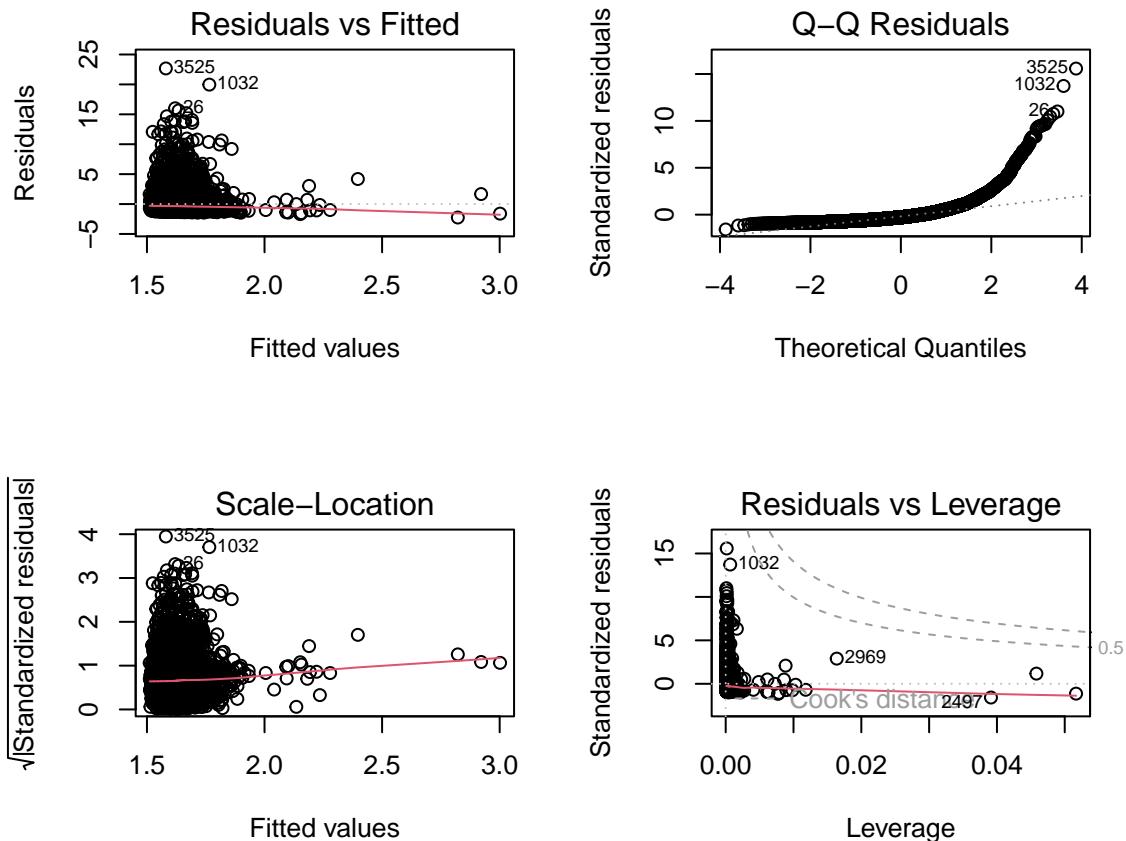
```
model_simple <- lm(TG ~ d3fat, data = data)
summary(model_simple)

##
## Call:
## lm(formula = TG ~ d3fat, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.2521 -0.7972 -0.4028  0.2655 22.6697 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.5088631  0.0309354 48.775 < 2e-16 ***
## d3fat       0.0015324  0.0003669   4.176 2.99e-05 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## Residual standard error: 1.455 on 9332 degrees of freedom
##   (49 observations deleted due to missingness)
## Multiple R-squared:  0.001866, Adjusted R-squared:  0.001759
## F-statistic: 17.44 on 1 and 9332 DF, p-value: 2.989e-05
par(mfrow=c(2,2))
plot(model_simple)

```



```
par(mfrow=c(1,1))
```

A statistically significant but weak positive association was observed between fat intake and triglyceride levels.

7 Multiple Linear Regression: Dietary Patterns and Lipids

```

model_hdl <- lm(HDL_C ~ Pattern1 + Pattern2 + Pattern3, data=data)
model_ldl <- lm(LDL_C ~ Pattern1 + Pattern2 + Pattern3, data=data)
model_tg <- lm(TG ~ Pattern1 + Pattern2 + Pattern3, data=data)
model_tc <- lm(TC ~ Pattern1 + Pattern2 + Pattern3, data=data)

summary(model_hdl)
##
```

```

## Call:
## lm(formula = HDL_C ~ Pattern1 + Pattern2 + Pattern3, data = data)
##
## Residuals:
##      Min      1Q Median      3Q     Max
## -1.1468 -0.2743 -0.0557  0.1931 14.1752
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.438748  0.005187 277.356 <2e-16 ***
## Pattern1    0.001640  0.005192  0.316   0.752
## Pattern2   -0.002925  0.005179 -0.565   0.572
## Pattern3   -0.005437  0.005188 -1.048   0.295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5011 on 9329 degrees of freedom
## (50 observations deleted due to missingness)
## Multiple R-squared:  0.0001625, Adjusted R-squared:  -0.000159
## F-statistic: 0.5054 on 3 and 9329 DF, p-value: 0.6785
summary(model_ldl)

##
## Call:
## lm(formula = LDL_C ~ Pattern1 + Pattern2 + Pattern3, data = data)
##
## Residuals:
##      Min      1Q Median      3Q     Max
## -2.8867 -0.6555 -0.0842  0.5470 12.4764
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.91424   0.01032 282.464 < 2e-16 ***
## Pattern1   -0.04365   0.01033 -4.227 2.39e-05 ***
## Pattern2    0.01975   0.01030  1.918  0.0552 .
## Pattern3    0.01609   0.01032  1.559  0.1190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9966 on 9327 degrees of freedom
## (52 observations deleted due to missingness)
## Multiple R-squared:  0.002565, Adjusted R-squared:  0.002244
## F-statistic: 7.994 on 3 and 9327 DF, p-value: 2.555e-05
summary(model_tg)

##
## Call:
## lm(formula = TG ~ Pattern1 + Pattern2 + Pattern3, data = data)
##
## Residuals:
##      Min      1Q Median      3Q     Max
## -1.7547 -0.7919 -0.4017  0.2666 22.6694
##

```

```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.62164   0.01504 107.814 < 2e-16 ***
## Pattern1    -0.02614   0.01505  -1.737 0.082473 .
## Pattern2     0.05433   0.01502   3.617 0.000299 ***
## Pattern3     0.08084   0.01504   5.375 7.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.453 on 9330 degrees of freedom
##   (49 observations deleted due to missingness)
## Multiple R-squared:  0.004798, Adjusted R-squared:  0.004478
## F-statistic: 14.99 on 3 and 9330 DF, p-value: 9.849e-10
summary(model_tc)

##
## Call:
## lm(formula = TC ~ Pattern1 + Pattern2 + Pattern3, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1052 -0.7201 -0.0993  0.6076  7.6668
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.78238   0.01054 453.683 < 2e-16 ***
## Pattern1    -0.06526   0.01055  -6.187 6.38e-10 ***
## Pattern2     0.04470   0.01053   4.247 2.19e-05 ***
## Pattern3     0.04379   0.01054   4.154 3.30e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.018 on 9330 degrees of freedom
##   (49 observations deleted due to missingness)
## Multiple R-squared:  0.007827, Adjusted R-squared:  0.007508
## F-statistic: 24.53 on 3 and 9330 DF, p-value: 8.321e-16

```

8 Diagnostic Tests

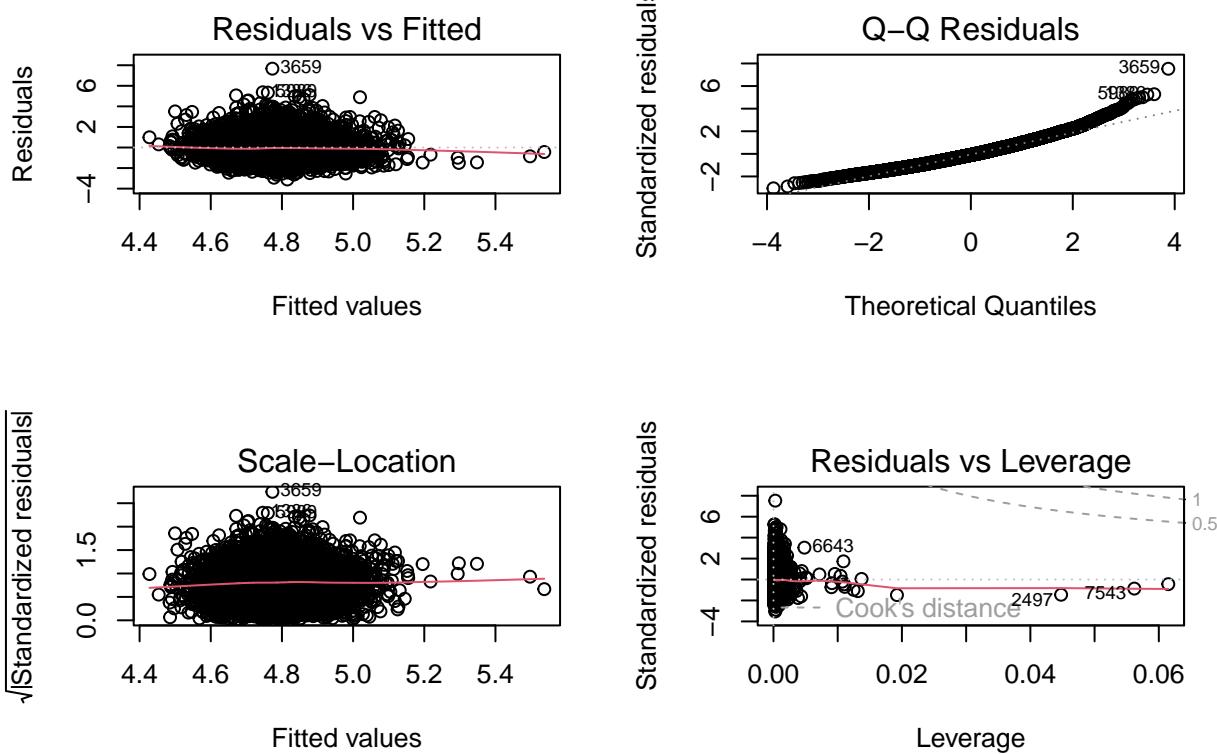
```

library(lmtest)

check_assumptions <- function(model) {
  par(mfrow=c(2,2))
  plot(model)
  par(mfrow=c(1,1))
  print(bptest(model))
  print(dwtest(model))
}

check_assumptions(model_tc)

```



```
##
## studentized Breusch-Pagan test
##
## data: model
## BP = 13.602, df = 3, p-value = 0.003501
##
##
## Durbin-Watson test
##
## data: model
## DW = 1.6819, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

Several models violate classical linear regression assumptions, motivating complementary non-parametric analyses.

9 ANOVA and Non-Parametric Tests

```
summary(aov(TC ~ Pattern1_Q, data = data))

##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pattern1_Q     3     49   16.46   15.83 2.91e-10 ***
## Residuals  9330  9704     1.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 49 observations deleted due to missingness
```

```

summary(aov(LDL_C ~ Pattern1_Q, data = data))

##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pattern1_Q     3    30  10.076   10.15 1.13e-06 ***
## Residuals  9327  9257   0.993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 52 observations deleted due to missingness
summary(aov(HDL_C ~ Pattern1_Q, data = data))

##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pattern1_Q     3      3  0.9881   3.939 0.00807 **
## Residuals  9329  2340   0.2509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 50 observations deleted due to missingness
summary(aov(TG ~ Pattern1_Q, data = data))

##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pattern1_Q     3     28   9.495   4.481 0.00379 **
## Residuals  9330  19768   2.119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 49 observations deleted due to missingness
bartlett.test(TC ~ Pattern1_Q, data = data)

##
##  Bartlett test of homogeneity of variances
##
## data: TC by Pattern1_Q
## Bartlett's K-squared = 19.231, df = 3, p-value = 0.0002449
kruskal.test(TC ~ Pattern1_Q, data = data)

##
##  Kruskal-Wallis rank sum test
##
## data: TC by Pattern1_Q
## Kruskal-Wallis chi-squared = 47.53, df = 3, p-value = 2.681e-10
kruskal.test(LDL_C ~ Pattern1_Q, data = data)

##
##  Kruskal-Wallis rank sum test
##
## data: LDL_C by Pattern1_Q
## Kruskal-Wallis chi-squared = 32.3, df = 3, p-value = 4.525e-07
kruskal.test(HDL_C ~ Pattern1_Q, data = data)

##
##  Kruskal-Wallis rank sum test
##
## data: HDL_C by Pattern1_Q
## Kruskal-Wallis chi-squared = 19.025, df = 3, p-value = 0.0002702

```

```
kruskal.test(TG ~ Pattern1_Q, data = data)

##
##  Kruskal-Wallis rank sum test
##
## data: TG by Pattern1_Q
## Kruskal-Wallis chi-squared = 22.78, df = 3, p-value = 4.488e-05
```

10 Discussion

Dietary patterns derived from PCA were significantly associated with blood lipid profiles. Carbohydrate-rich patterns were associated with lower LDL-C and TC levels, whereas fat- and protein-rich patterns were associated with higher triglyceride and cholesterol concentrations. HDL-C showed no strong association with dietary patterns.

11 Conclusion

This study demonstrates that dietary patterns, rather than isolated nutrients, are meaningfully associated with lipid metabolism. PCA provided a useful framework for summarizing dietary intake, and appropriate statistical methods were applied after careful assessment of model assumptions. These findings highlight the relevance of dietary pattern analysis in nutritional epidemiology.