# Food Preferences and Consumption Patterns

Vidya Sri L

2024-11-18

## Introduction

Dietary preferences and food consumption patterns are key aspects of human behavior that have significant implications for public health, nutrition, and the food industry. These preferences are shaped by a complex interplay of demographic factors, health consciousness, cultural influences, and individual lifestyle choices. Understanding these patterns is crucial for designing effective public health interventions, developing targeted marketing strategies, and addressing the diverse needs of consumers.

This study aims to explore the intricate relationships between dietary preferences and various influencing factors such as gender, health consciousness, meal preferences, and eating habits. By employing categorical data analysis techniques, the research seeks to uncover the underlying associations that drive food choices. Specifically, the study will analyze how gender affects dietary preferences, assess the impact of health consciousness on food choices, examine the relationship between meal preferences and eating out frequency, and develop predictive models to forecast dietary preferences. Additionally, advanced statistical methods, including multinomial logistic regression and log-linear modeling, will be used to delve into more complex relationships among multiple categorical variables related to food consumption.

## Objectives

1. Analyze Dietary Preferences: Examine the association between gender and dietary preferences, using odds ratio and relative risk to quantify differences.
2. Assess Health Consciousness Impact: Determine how health consciousness influences food preferences through Chi-Square tests and visualizations.
3. Evaluate Meal Preferences: Explore the link between meal preferences and frequency of eating out, applying likelihood ratio tests and heatmaps.
4. Predict Dietary Preferences: Fit and evaluate logistic regression models to predict dietary preferences based on various predictors.
5. Apply Multinomial Logistic Regression: Investigate shopping behaviors related to cooking frequency using multinomial logistic regression.
6. Fit Log-Linear Model: Test the adequacy of the log-linear model in representing relationships among categorical variables related to food preferences.

## Formulation of Problem

The study aims to explore and analyze dietary preferences and food consumption patterns using categorical data analysis. The focus is on understanding how various factors such as gender, health consciousness, meal preferences, and eating habits influence dietary choices among respondents. The key issues include:

1. Dietary Preferences by Gender: Assessing how gender affects preferences for vegetarian versus non-vegetarian diets.

2. Health Consciousness and Food Choices: Investigating the relationship between health consciousness levels and preferred food types.
3. Meal Preferences and Eating Out Frequency: Evaluating whether meal preferences are associated with how often individuals eat out.
4. Dietary Preferences Prediction: Developing models to predict dietary preferences based on demographic and behavioral factors.
5. Multinomial and Log-Linear Modeling: Employing advanced statistical techniques to understand the complex relationships among multiple categorical variables related to food consumption.

## Questionnaire

The following questionnaire was designed to collect comprehensive data on respondents' food preferences, eating habits, and spending patterns. It includes various types of questions to categorize responses and assess key aspects of food consumption.

1. What is your age group?
   - Under 18
   - 18-28
   - 29-38
   - 39-48
   - 49 and above
2. What is your gender?
   - Male
   - Female
3. What is your dietary preference?
   - Vegetarian
   - Non-Vegetarian
4. What is your favorite type of cuisine?
   - Italian
   - Chinese
   - North Indian
   - South Indian
5. How often do you eat out?
   - Daily
   - Weekly
   - Monthly
   - Occasionally
6. Which meal of the day do you prefer the most?
   - Breakfast
   - Lunch
   - Dinner
   - Snacks

7. How important is health to you when choosing food?
   - Important
   - Neutral
   - Does not care
8. What is your preferred type of food?
   - Fast Food
   - Home-Cooked
   - Organic
   - Street Food
9. How much do you typically spend on food per week?
   - Under 200
   - 200-500
   - 600-1000
   - 1000 and above
10. Where do you primarily shop for food?
    - Supermarket
    - Local Market
    - Organic Store
    - Online
11. How often do you cook meals at home?
    - Daily
    - Several times a week
    - Weekly
    - Occasionally
12. How often do you try new foods or cuisines?
    - Very Often
    - Often
    - Occasionally
    - Never

## Data Collection

### Description

Data was collected using an online survey tool, distributed among participants from various age groups. The survey automatically recorded respondents' email addresses to ensure authenticity.

### Sample Size

A total of 131 respondents participated in the survey, providing diverse insights into their food preferences and consumption habits.

**Methodology**

The study utilizes a structured questionnaire to collect data on respondents' food preferences, eating habits, and spending patterns. The survey was distributed online, targeting a diverse sample of participants from various age groups, resulting in a total of 131 responses. The dataset was pre-processed to remove unnecessary columns, and various statistical analyses were performed to meet the study's objectives.

- **Contingency Tables and Tests:** We created contingency tables to explore the relationships between categorical variables, such as gender and dietary preferences. The odds ratio and relative risk were calculated to quantify differences, while Chi-Square tests were employed to assess the significance of associations, such as between health consciousness and preferred food types.

- **Visualization:** Heatmaps were generated to visually represent the relationships between variables, providing a clearer understanding of how different factors, such as health consciousness, influence food preferences.

- **Logistic Regression:** A logistic regression model was fitted to predict dietary preferences based on selected predictors, such as gender and favorite cuisine. Model selection techniques, including forward selection and backward elimination, were used to identify the most important predictors.

- **Multinomial Logistic Regression:** This analysis was used to investigate shopping behaviors related to cooking frequency, providing insights into how often people shop at different places based on how frequently they cook meals at home.

- **Log-Linear Modeling:** A log-linear model was fitted to test the adequacy of representing relationships among categorical variables related to food preferences, allowing us to understand the interactions between multiple factors.

## Data Import and Initial Processing

```r
library(readxl)
#Load data
library(readxl)
data<- read_excel("/Users/apple/Downloads/CDA_Project (Responses).xlsx")
data <- data[,-c(1, 2)]
# Print first few rows of data
head(data)
```

```
## # A tibble: 6 x 12
##   age      gender diet_pref  fav_cuisine eat_out_freq meal_pref health_conscious
##   <chr>    <chr>  <chr>      <chr>       <chr>        <chr>     <chr>
## 1 18-28    Male   Non - Veg~ South Indi~ Daily        Lunch     Neutral
## 2 Under 18 Female Non - Veg~ South Indi~ Weekly       Lunch     Important
## 3 18-28    Female Non - Veg~ Italian     Weekly       Lunch     Important
## 4 18-28    Female Non - Veg~ South Indi~ Weekly       Dinner    Important
## 5 29-38    Female Vegetarian South Indi~ Occasionally Dinner    Neutral
## 6 18-28    Female Non - Veg~ South Indi~ Monthly      Lunch     Important
## # i 5 more variables: pref_typefood <chr>, spend_week <chr>, shop_place <chr>,
## #   cook_freq <chr>, new_food <chr>
```

## Analysis

```r
# Summary statistics for numerical variables
summary(data)
```

```
##      age               gender            diet_pref         fav_cuisine
##  Length:133         Length:133         Length:133         Length:133
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##  eat_out_freq        meal_pref         health_conscious   pref_typefood
##  Length:133         Length:133         Length:133         Length:133
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##   spend_week         shop_place          cook_freq           new_food
##  Length:133         Length:133         Length:133         Length:133
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```r
# Frequencies of categorical variables
table(data$age)
```

```
##
##       18-28       29-38       39-48 49 and above     Under 18
##         110           7           7            4            5
```

```r
table(data$gender)
```

```
##
## Female   Male
##     70     63
```

```r
table(data$diet_pref)
```

```
##
## Non - Vegetarian       Vegetarian
##               96               37
```

```r
table(data$fav_cuisine)
```

```
##
##      Chinese      Italian North Indian South Indian
##            7            6           17          103
```

```r
table(data$eat_out_freq)
```

```
##
##        Daily      Monthly Occasionally       Weekly
##           27           15           27           64
```

```r
table(data$meal_pref)
```

```
##
## Breakfast    Dinner     Lunch    Snacks
##        23        51        50         9
```

```r
table(data$health_conscious)
```

```
##
## Does not care     Important       Neutral
##            10            75            48
```

```r
table(data$pref_typefood)
```

```
##
##   Fast Food Home-Cooked     Organic Street Food
##          16          94          12          11
```

```r
table(data$spend_week)
```

```
##
## 1000  above     200-500    600-1000   Under 200
##          16          61          27          29
```

```r
table(data$shop_place)
```

```
##
##  Local Market        Online Organic Store   Supermarket
##            56            24             7            46
```

```r
table(data$cook_freq)
```

```
##
##                Daily         Occasionally Several times a week
##                   43                   47                   29
##               Weekly
##                   14
```

```r
table(data$new_food)
```

```
##
##        Never Occasionally        Often   Very Often
##           12           70           33           18
```

**Demographics:** 69 females and 62 males. Majority are non-vegetarian (94).

**Cuisine Preferences:** South Indian is the most popular cuisine (101). Other favorites include North Indian (17), Italian (6), and Chinese (7).

**Meal Preferences:** Most preferred meals are dinner (51) and lunch (49). Fewer respondents prefer breakfast (22) and snacks (9).

**Eating Out Frequency:** Weekly (63) is the most common frequency. Other frequencies: Occasionally (26), Daily (27), and Monthly (15).

**Food Preferences:** Home-cooked food is highly preferred (92). Less preference for fast food (16) and street food (11). Organic food has a moderate preference (12).

**Shopping Habits:** Local Market (55) is the most popular shopping place. Followed by Supermarket (45), Online (24), and Organic Store (7).

**Food Importance:** Most consider food important (73). A smaller number are indifferent (10).

**Analysis of Gender and Dietary preferences:**

```r
# Create contingency table with gender as the response variable
contingency_table_gen <- table(data$diet_pref, data$gender)

# Display the contingency table
print(contingency_table_gen)
```

```
##
##                      Female Male
##   Non - Vegetarian       43   53
##   Vegetarian             27   10
```

The contingency table reveals the following dietary preferences: • Females: 42 prefer non-vegetarian, and 27 prefer vegetarian. • Males: 52 prefer non-vegetarian, and 10 prefer vegetarian.

```r
# Define the values from the contingency table
a <- 42 # Females preferring non-vegetarian
b <- 27 # Females preferring vegetarian
c <- 52 # Males preferring non-vegetarian
d <- 10 # Males preferring vegetarian
# Calculate Odds Ratio
odds_ratio <- (a / b) / (c / d)
print(paste("Odds Ratio:", odds_ratio))
```

```
## [1] "Odds Ratio: 0.299145299145299"
```

The odds ratio of approximately 0.299 indicates that the odds of females preferring non-vegetarian food compared to vegetarian food are about 30% of the odds for males. This suggests that females are significantly less likely to prefer non-vegetarian food compared to males.

```r
# Calculate Relative Risk
risk_female_non_veg <- a / (a + b)
risk_male_non_veg <- c / (c + d)
relative_risk <- risk_female_non_veg / risk_male_non_veg
print(paste("Relative Risk:", relative_risk))
```

```
## [1] "Relative Risk: 0.725752508361204"
```

The relative risk of approximately 0.727 further supports this, showing that females are about 73% as likely as males to prefer non-vegetarian food.

**Analysis of Health consciousness and Preferred foodtype**

```r
# Create a contingency table with health_conscious and pref_typefood
contingency_table_hf<- table(data$health_conscious, data$pref_typefood)

# Display the contingency table
print(contingency_table_hf)
```

```
##
##                  Fast Food Home-Cooked Organic Street Food
##   Does not care          3           5       1           1
##   Important              6          58       9           2
##   Neutral                7          31       2           8
```

```r
# Perform the Chi-Square test
chi_square_test <- chisq.test(contingency_table_hf)
```

```
## Warning in chisq.test(contingency_table_hf): Chi-squared approximation may be
## incorrect
```

```r
# Display the results
print(chi_square_test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table_hf
## X-squared = 14.261, df = 6, p-value = 0.02685
```

```r
# Critical value for alpha = 0.05 and df = 6
critical_value <- qchisq(0.95, df = 6)
print(critical_value)
```

```
## [1] 12.59159
```

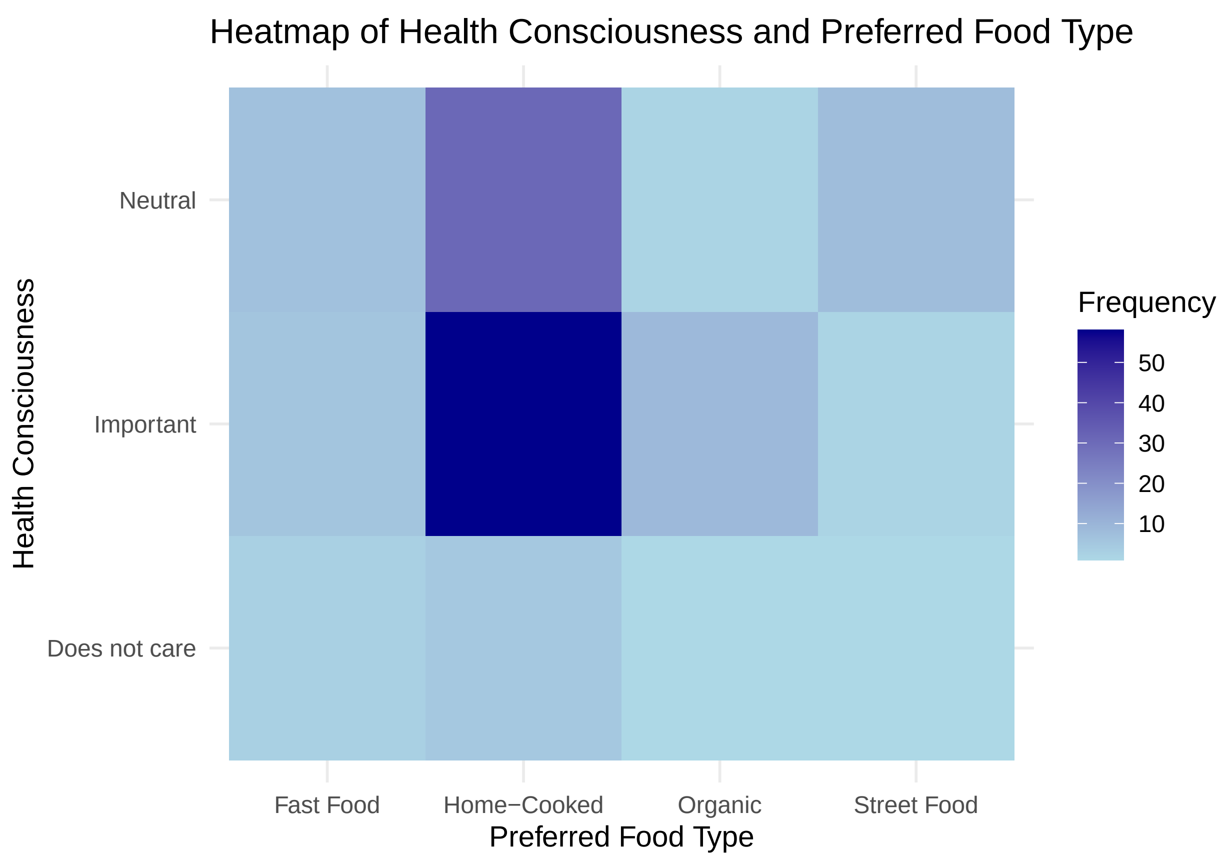
The Chi-Square test results for the contingency table with health_conscious and pref_typefood show a test statistic of 13.839 with 6 degrees of freedom, and a p-value of 0.03149. This p-value is below the typical alpha level of 0.05, indicating that there is a statistically significant association between health consciousness and preferred food type.The critical value for the Chi-Square distribution with 6 degrees of freedom at a 0.05 significance level is approximately 12.592. Since the test statistic 13.839 exceeds this critical value, we reject the null hypothesis.

In other words, the differences observed in the frequency of preferred food types across different levels of health consciousness are unlikely to be due to chance, suggesting a meaningful relationship between health consciousness and food preference.

```r
# Convert the contingency table to a data frame
contingency_df <- as.data.frame(contingency_table_hf)

# Load ggplot2 package
library(ggplot2)

# Create a heatmap with a color gradient from light to dark
ggplot(contingency_df, aes(x = Var2, y = Var1, fill = Freq)) +
geom_tile() +
scale_fill_gradient(low = "lightblue", high = "darkblue") +
labs(title = "Heatmap of Health Consciousness and Preferred Food Type",
x = "Preferred Food Type",
y = "Health Consciousness",
fill = "Frequency") +
theme_minimal()
```




**Visualisation**

The heatmap visually supports this finding by showing how the frequency of each food preference varies with health consciousness, with some categories (like "Important" in "Home-Cooked") being more pronounced than others.

**Analysis of Meal preference and eating outside frequently**

```r
#Create contingency table
table = table(data$meal_pref, data$eat_out_freq)
```

```r
table
```

```
##
##             Daily Monthly Occasionally Weekly
##   Breakfast     4       1            3     15
##   Dinner       14       5           11     21
##   Lunch         6       6           11     27
##   Snacks        3       3            2      1
```

```r
#Likelihood Ratio Test
library(DescTools)

likelihood_ratio_test <- GTest(table)
print(likelihood_ratio_test)
```

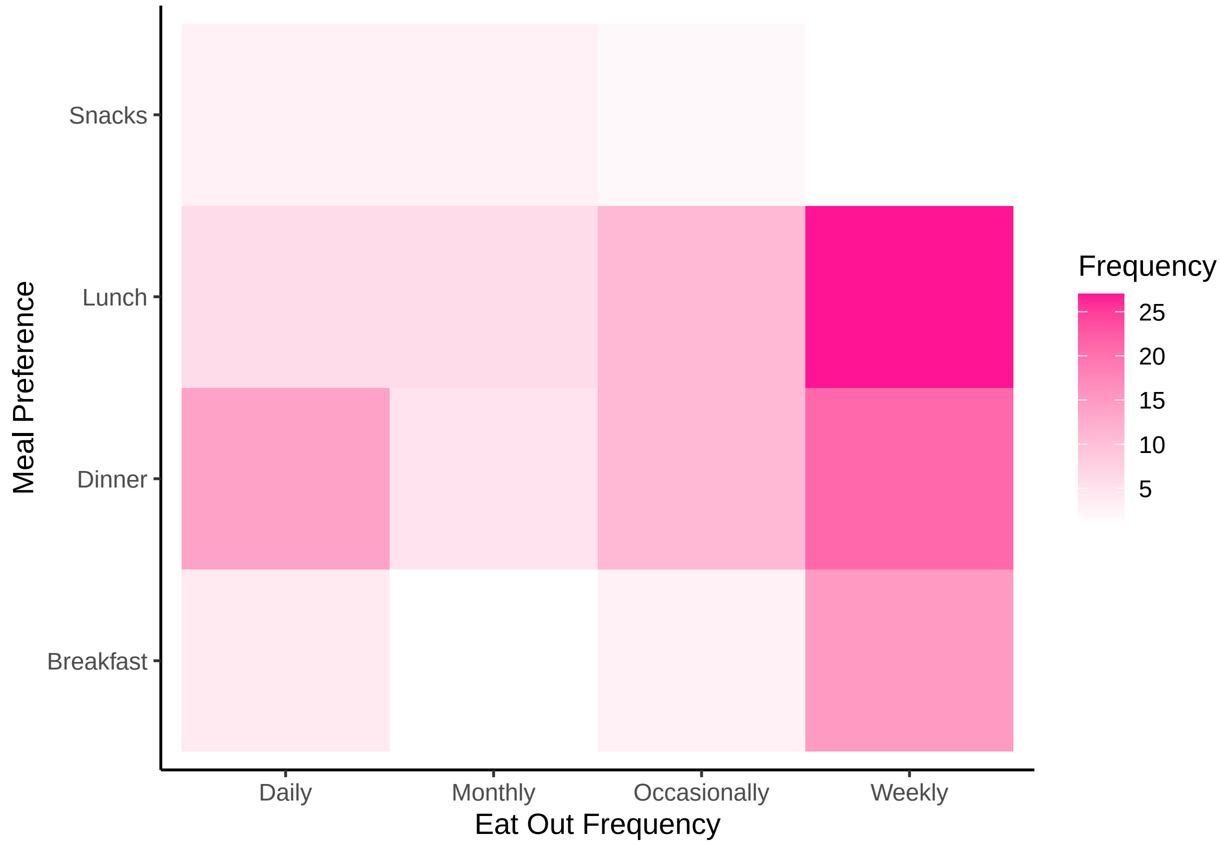
```
##
##  Log likelihood ratio (G-test) test of independence without correction
##
## data:  table
## G = 14.495, X-squared df = 9, p-value = 0.1058
```

```r
# Calculate the degrees of freedom
df <- (nrow(table)- 1) * (ncol(table)- 1)
# Calculate the chi square tabulated value
p_value <- qchisq(0.95, df)
print(paste("Chi square tabulated value ", p_value))
```

```
## [1] "Chi square tabulated value  16.9189776046204"
```

Here, the likelihood ratio test (G-test), is used to determine whether there is a significant association between meal preference and frequency of eating out. The p-value is greater than the typical significance level of 0.05, and Calculate G is less than table value, therefore we fail to reject the null hypothesis which suggests that there is no significant association between meal preference and frequency of eating out at the 5% level.

```r
# Convert the contingency table to a data frame
table_df <- as.data.frame(table)
# Rename the columns
colnames(table_df) <- c("meal_pref", "eat_out_freq", "freq")
# Create a heatmap using ggplot2
ggplot(table_df, aes(x = eat_out_freq, y = meal_pref, fill = freq)) +
geom_tile() +
scale_fill_gradient(low = "white", high = "deeppink") +
labs(x = "Eat Out Frequency", y = "Meal Preference", fill = "Frequency") +
theme_classic()
```




**Visualisation**

The heatmap shows that “Lunch” is the most common meal for eating out, especially on a weekly basis, as indicated by the darkest pink shade. “Dinner” is also popular, particularly on an occasional and weekly basis. Eating out daily is less common across all meal types. “Snacks” and “Breakfast” have the lowest frequencies, with lighter shades across all eating-out frequencies. Overall, weekly dining out, especially for lunch, is the most frequent behavior.

**Logistic Regression and Model Evaluation on Dietary Preferences**

```r
data$diet_pref <- ifelse(data$diet_pref== "Vegetarian", 1, 0)
model=lm(data$diet_pref~.,data=data)

print(data)
```

**Data Transformation and Initial Model Fitting**

```
## # A tibble: 133 x 12
##    age      gender diet_pref fav_cuisine eat_out_freq meal_pref health_conscious
##    <chr>    <chr>      <dbl> <chr>       <chr>        <chr>     <chr>
##  1 18-28    Male           0 South Indi~ Daily        Lunch     Neutral
##  2 Under 18 Female         0 South Indi~ Weekly       Lunch     Important
##  3 18-28    Female         0 Italian     Weekly       Lunch     Important
##  4 18-28    Female         0 South Indi~ Weekly       Dinner    Important
##  5 29-38    Female         1 South Indi~ Occasionally Dinner    Neutral
```

```
##  6 18-28    Female         0 South Indi~ Monthly      Lunch     Important
##  7 39-48    Female         1 Chinese     Weekly       Snacks    Neutral
##  8 18-28    Female         0 South Indi~ Monthly      Lunch     Important
##  9 18-28    Female         1 South Indi~ Weekly       Breakfast Important
## 10 18-28    Male           0 South Indi~ Weekly       Lunch     Important
## # i 123 more rows
## # i 5 more variables: pref_typefood <chr>, spend_week <chr>, shop_place <chr>,
## #   cook_freq <chr>, new_food <chr>
```

The diet_pref column is converted into a binary variable where “Vegetarian” is coded as 1 and others as 0. A linear model is fitted with diet_pref as the response variable and all other variables as predictors.

**Test for Normality** **Shapiro-Wilk Test** • H0: The residuals are normally distributed • H1: The residuals are not normally distributed.

```
r=rstudent(model)
shapiro.test(r)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  r
## W = 0.94307, p-value = 2.812e-05
```

Here p-value 6.01e-05 <0.05,we reject ho.Hence the residuals are not normally distributed.

### Model Selection

**Forward Selection** Forward selection helps in identifying the most important predictors for diet_pref by incrementally adding variables that improve model performance.

```
fitstart=lm(data$diet_pref~1,data = data)
model <- lm(data$diet_pref~., data =data)
fwd=step(fitstart,direction = 'forward',scope = formula(model))
```

```
## Start:  AIC=-211.52
## data$diet_pref ~ 1
##
##                    Df Sum of Sq    RSS     AIC
## + gender            1   1.70835 24.998 -218.31
## + fav_cuisine       3   2.24053 24.466 -217.18
## + spend_week        3   1.57874 25.128 -213.63
## + eat_out_freq      3   1.29936 25.407 -212.16
## <none>                          26.707 -211.52
## + new_food          3   0.99371 25.713 -210.57
## + health_conscious  2   0.44760 26.259 -209.77
## + meal_pref         3   0.72794 25.979 -209.20
## + pref_typefood     3   0.45608 26.251 -207.81
## + shop_place        3   0.31287 26.394 -207.09
## + cook_freq         3   0.06526 26.642 -205.85
## + age               4   0.29508 26.412 -205.00
```

```
##
## Step:  AIC=-218.31
## data$diet_pref ~ gender
##
##                    Df Sum of Sq    RSS     AIC
## + fav_cuisine       3   1.88845 23.110 -222.76
## <none>                          24.998 -218.31
## + spend_week        3   1.06216 23.936 -218.09
## + health_conscious  2   0.53229 24.466 -217.18
## + new_food          3   0.72614 24.272 -216.24
## + eat_out_freq      3   0.70651 24.292 -216.13
## + meal_pref         3   0.49096 24.508 -214.95
## + pref_typefood     3   0.33250 24.666 -214.09
## + shop_place        3   0.20724 24.791 -213.42
## + cook_freq         3   0.06060 24.938 -212.64
## + age               4   0.37681 24.622 -212.33
##
## Step:  AIC=-222.76
## data$diet_pref ~ gender + fav_cuisine
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          23.110 -222.76
## + spend_week        3   0.83110 22.279 -221.63
## + health_conscious  2   0.40421 22.706 -221.11
## + eat_out_freq      3   0.68714 22.423 -220.78
## + new_food          3   0.54497 22.565 -219.94
## + meal_pref         3   0.49140 22.619 -219.62
## + pref_typefood     3   0.46162 22.648 -219.44
## + shop_place        3   0.23422 22.876 -218.12
## + cook_freq         3   0.06675 23.043 -217.15
## + age               4   0.29545 22.814 -216.47
```

Gender and favourite cuisine are the most important predictors for diet_pref.

### Backward Elimination

Backward elimination provides another method to refine the model by removing predictors that do not contribute meaningfully to the prediction of diet_pref.

```
bwd=step(model,direction = "backward")
```

```
## Start:  AIC=-194.35
## data$diet_pref ~ age + gender + fav_cuisine + eat_out_freq +
##     meal_pref + health_conscious + pref_typefood + spend_week +
##     shop_place + cook_freq + new_food
##
##                    Df Sum of Sq    RSS     AIC
## - age               4   0.11121 19.177 -201.57
## - cook_freq         3   0.13296 19.198 -199.42
## - eat_out_freq      3   0.31941 19.385 -198.14
## - new_food          3   0.42065 19.486 -197.45
## - shop_place        3   0.56037 19.626 -196.50
## - meal_pref         3   0.66857 19.734 -195.76
```

```
## - pref_typefood     3   0.68788 19.753 -195.63
## - health_conscious  2   0.56842 19.634 -194.44
## - spend_week        3   0.87219 19.938 -194.40
## <none>                          19.065 -194.35
## - gender            1   0.44839 19.514 -193.26
## - fav_cuisine       3   1.40179 20.467 -190.91
##
## Step:  AIC=-201.58
## data$diet_pref ~ gender + fav_cuisine + eat_out_freq + meal_pref +
##     health_conscious + pref_typefood + spend_week + shop_place +
##     cook_freq + new_food
##
##                    Df Sum of Sq    RSS     AIC
## - cook_freq         3   0.14253 19.319 -206.59
## - eat_out_freq      3   0.33495 19.512 -205.27
## - new_food          3   0.45667 19.633 -204.44
## - shop_place        3   0.59088 19.767 -203.54
## - meal_pref         3   0.68741 19.864 -202.89
## - pref_typefood     3   0.71795 19.895 -202.69
## - spend_week        3   0.85350 20.030 -201.78
## <none>                          19.177 -201.57
## - health_conscious  2   0.59599 19.773 -201.50
## - gender            1   0.40954 19.586 -200.76
## - fav_cuisine       3   1.50165 20.678 -197.55
##
## Step:  AIC=-206.59
## data$diet_pref ~ gender + fav_cuisine + eat_out_freq + meal_pref +
##     health_conscious + pref_typefood + spend_week + shop_place +
##     new_food
##
##                    Df Sum of Sq    RSS     AIC
## - eat_out_freq      3   0.36315 19.682 -210.11
## - new_food          3   0.38233 19.701 -209.98
## - shop_place        3   0.56591 19.885 -208.75
## - pref_typefood     3   0.61239 19.932 -208.44
## - meal_pref         3   0.75185 20.071 -207.51
## - spend_week        3   0.78626 20.105 -207.28
## - health_conscious  2   0.57561 19.895 -206.69
## <none>                          19.319 -206.59
## - gender            1   0.44291 19.762 -205.58
## - fav_cuisine       3   1.49030 20.809 -202.71
##
## Step:  AIC=-210.11
## data$diet_pref ~ gender + fav_cuisine + meal_pref + health_conscious +
##     pref_typefood + spend_week + shop_place + new_food
##
##                    Df Sum of Sq    RSS     AIC
## - new_food          3   0.38947 20.072 -213.51
## - shop_place        3   0.47748 20.160 -212.93
## - pref_typefood     3   0.63300 20.315 -211.90
## - meal_pref         3   0.72231 20.405 -211.32
## - spend_week        3   0.84263 20.525 -210.54
## <none>                          19.682 -210.11
## - health_conscious  2   0.72084 20.403 -209.33
```

```
## - gender            1   0.45190 20.134 -209.09
## - fav_cuisine       3   1.34165 21.024 -207.34
##
## Step:  AIC=-213.51
## data$diet_pref ~ gender + fav_cuisine + meal_pref + health_conscious +
##     pref_typefood + spend_week + shop_place
##
##                    Df Sum of Sq    RSS     AIC
## - shop_place        3   0.43838 20.510 -216.63
## - pref_typefood     3   0.61260 20.684 -215.51
## - meal_pref         3   0.86806 20.940 -213.88
## - health_conscious  2   0.57273 20.645 -213.76
## <none>                          20.072 -213.51
## - spend_week        3   1.12597 21.198 -212.25
## - gender            1   0.52914 20.601 -212.05
## - fav_cuisine       3   1.54799 21.620 -209.63
##
## Step:  AIC=-216.63
## data$diet_pref ~ gender + fav_cuisine + meal_pref + health_conscious +
##     pref_typefood + spend_week
##
##                    Df Sum of Sq    RSS     AIC
## - pref_typefood     3   0.74211 21.252 -217.91
## - meal_pref         3   0.76812 21.278 -217.74
## - health_conscious  2   0.55203 21.062 -217.10
## - spend_week        3   0.94426 21.454 -216.65
## <none>                          20.510 -216.63
## - gender            1   0.64000 21.150 -214.55
## - fav_cuisine       3   1.56068 22.071 -212.88
##
## Step:  AIC=-217.91
## data$diet_pref ~ gender + fav_cuisine + meal_pref + health_conscious +
##     spend_week
##
##                    Df Sum of Sq    RSS     AIC
## - meal_pref         3   0.64594 21.898 -219.92
## - health_conscious  2   0.47496 21.727 -218.97
## - spend_week        3   0.85431 22.107 -218.66
## <none>                          21.252 -217.91
## - fav_cuisine       3   1.43092 22.683 -215.24
## - gender            1   0.85277 22.105 -214.67
##
## Step:  AIC=-219.92
## data$diet_pref ~ gender + fav_cuisine + health_conscious + spend_week
##
##                    Df Sum of Sq    RSS     AIC
## - health_conscious  2   0.38062 22.279 -221.63
## - spend_week        3   0.80751 22.706 -221.11
## <none>                          21.898 -219.92
## - fav_cuisine       3   1.45420 23.352 -217.37
## - gender            1   0.99517 22.893 -216.01
##
## Step:  AIC=-221.63
## data$diet_pref ~ gender + fav_cuisine + spend_week
```

```
##
##               Df Sum of Sq    RSS     AIC
## - spend_week   3   0.83110 23.110 -222.76
## <none>                     22.279 -221.63
## - fav_cuisine  3   1.65740 23.936 -218.09
## - gender       1   0.95259 23.231 -218.06
##
## Step:  AIC=-222.76
## data$diet_pref ~ gender + fav_cuisine
##
##               Df Sum of Sq    RSS     AIC
## <none>                     23.110 -222.76
## - fav_cuisine  3    1.8884 24.998 -218.31
## - gender       1    1.3563 24.466 -217.18
```

Gender and favourite cuisine are the most important predictors for diet_pref selected through backward selection.

### Logistic Regression Model

```r
glm=glm(data$diet_pref~ gender + fav_cuisine,data=data,family = binomial)
summary(glm)
```

```
##
## Call:
## glm(formula = data$diet_pref ~ gender + fav_cuisine, family = binomial,
##     data = data)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)              0.02204    0.79434   0.028   0.9779
## genderMale              -1.19554    0.44931  -2.661   0.0078 **
## fav_cuisineItalian      -0.71519    1.17515  -0.609   0.5428
## fav_cuisineNorth Indian  0.85252    0.94637   0.901   0.3677
## fav_cuisineSouth Indian -0.80073    0.82790  -0.967   0.3334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 157.27  on 132  degrees of freedom
## Residual deviance: 139.47  on 128  degrees of freedom
## AIC: 149.47
##
## Number of Fisher Scoring iterations: 4
```

A generalized linear model (GLM) is fitted using gender and fav_cuisine as predictors, with diet_pref as the binary outcome.The logistic regression analysis shows that gender is a significant predictor of dietary preference. Specifically, males are significantly less likely to be vegetarian compared to females (Estimate = -1.20226, p = 0.00757). The decrease from 155.95 (null deviance) to 138.44 (residual deviance) suggests that the model with gender and fav_cuisine improves the fit compared to a model with no predictors

### Model Evaluation

```
# Load necessary libraries
library(caret) # For confusionMatrix and metrics
```

```
## Loading required package: lattice

##
## Attaching package: 'caret'

## The following objects are masked from 'package:DescTools':
##
##     MAE, RMSE
```

```
library(pROC) # For ROC curve and AUC
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
# Predict probabilities
pred_probs <- predict(glm, data, type = "response")
# Convert probabilities to binary predictions
pred <- ifelse(pred_probs > 0.5, 1, 0)
# Calculate confusion matrix
conf_matrix <- confusionMatrix(as.factor(pred), as.factor(data$diet_pref))
# Extract accuracy, sensitivity, specificity
accuracy <- conf_matrix$overall['Accuracy']
sensitivity <- conf_matrix$byClass['Sensitivity']
specificity <- conf_matrix$byClass['Specificity']
# Print accuracy, sensitivity, and specificity
print(paste("Accuracy:", round(accuracy, 4)))
```

```
## [1] "Accuracy: 0.7744"
```

```
print(paste("Sensitivity:", round(sensitivity, 4)))
```

```
## [1] "Sensitivity: 0.9583"
```

```
print(paste("Specificity:", round(specificity, 4)))
```

```
## [1] "Specificity: 0.2973"
```

```r
# Calculate F1 Score
precision <- conf_matrix$byClass['Precision']
recall <- sensitivity
f1_score <- 2 * (precision * recall) / (precision + recall)
print(paste("F1 Score:", round(f1_score, 4)))
```

```
## [1] "F1 Score: 0.8598"
```

```r
# Print AUC value
roc_curve <- roc(data$diet_pref, pred_probs)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```r
auc_value <- auc(roc_curve)
print(paste("AUC:", round(auc_value, 4)))
```

```
## [1] "AUC: 0.7007"
```

The model shows good accuracy (77.1%) and is excellent at identifying vegetarians (sensitivity: 95.74%). However, it struggles to correctly identify non-vegetarians (specificity: 29.73%). The F1 Score (0.8571) reflects a good balance between precision and recall, while the AUC (0.6998) indicates moderate overall discrimination ability.

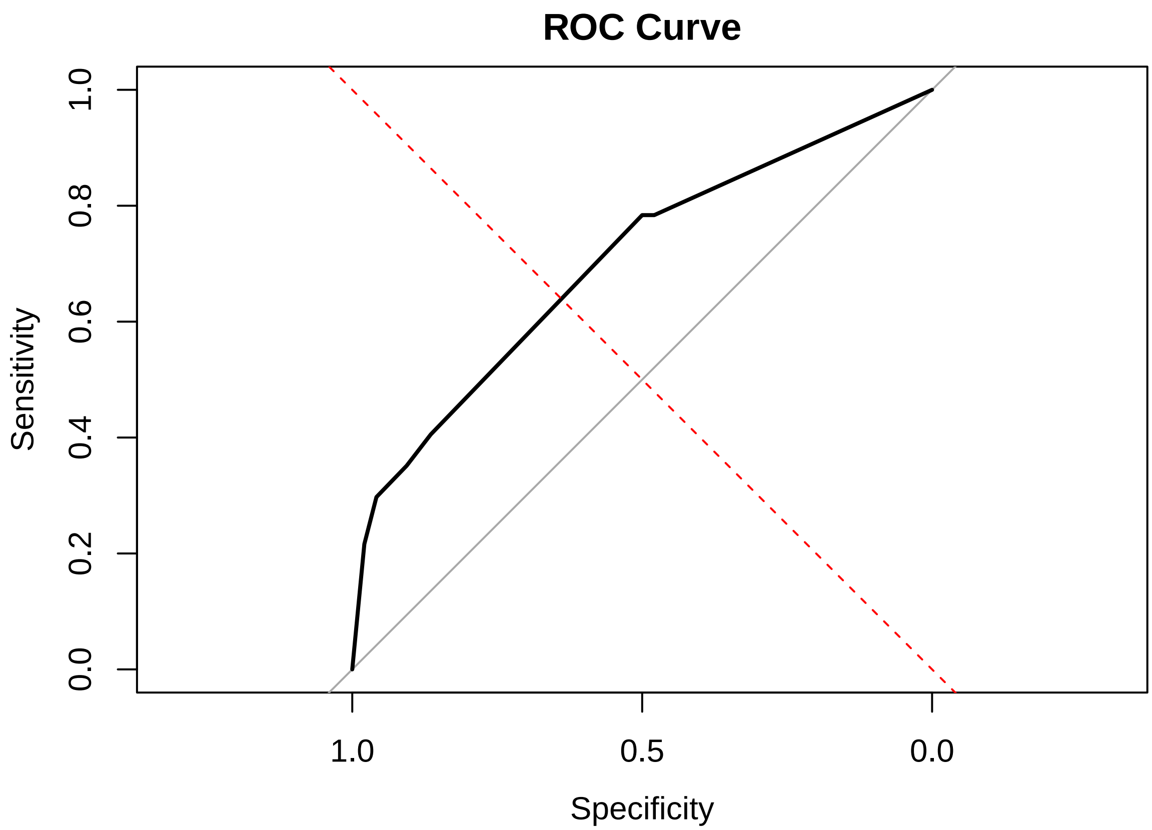
### ROC curve

```r
# Plot ROC curve
roc_curve <- roc(data$diet_pref, pred_probs)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```r
auc_value <- auc(roc_curve)
plot(roc_curve, main = "ROC Curve")
abline(a = 0, b = 1, lty = 2, col = "red") # Add a diagonal line
```




The ROC curve shows that the model has moderate performance. The curve is close to the diagonal, indicating that the model's ability to distinguish between classes is not strong. This suggests the model is only somewhat better than random guessing, with room for improvement.

The model is very effective at detecting vegetarians but struggles with accurately identifying non-vegetarians. The high sensitivity and F1 score suggest it's well-tuned for situations where identifying vegetarians is more critical, but the low specificity indicates it may produce many false positives for vegetarians.

```r
table1 = table(data$shop_place, data$cook_freq)
table1
```

**Fitting of Multinomial Logistic Regression:**

```
##
##                  Daily Occasionally Several times a week Weekly
##   Local Market      22           24                    8      2
##   Online             8            7                    7      2
##   Organic Store      2            2                    1      2
##   Supermarket       11           14                   13      8
```

```r
# Load the necessary libraries
library(nnet)
# Convert the shop_place and cook_freq variables to factors
data$shop_place <- factor(data$shop_place)
```

```r
data$cook_freq <- factor(data$cook_freq)

# Perform multinomial logistic regression analysis
multinom_shop_place <- multinom(shop_place~ cook_freq, data = data)
```

```
## # weights:  20 (12 variable)
## initial  value 184.377150
## iter  10 value 152.309885
## final  value 152.261895
## converged
```

```r
# Summarize the results
summary(multinom_shop_place)
```

```
## Call:
## multinom(formula = shop_place ~ cook_freq, data = data)
##
## Coefficients:
##                (Intercept) cook_freqOccasionally cook_freqSeveral times a week
## Online          -1.0115922           -0.22055150                     0.8780091
## Organic Store   -2.3980399           -0.08700827                     0.3185245
## Supermarket     -0.6931117            0.15413969                     1.1785746
##                cook_freqWeekly
## Online                1.011715
## Organic Store         2.398164
## Supermarket           2.079489
##
## Std. Errors:
##                (Intercept) cook_freqOccasionally cook_freqSeveral times a week
## Online           0.4128609             0.5958004                     0.6620493
## Organic Store    0.7385994             1.0427210                     1.2925027
## Supermarket      0.3692708             0.4994543                     0.5816188
##                cook_freqWeekly
## Online               1.0818849
## Organic Store        1.2432013
## Supermarket          0.8725836
##
## Residual Deviance: 304.5238
## AIC: 328.5238
```

```r
# Extract the coefficients for each category
coef_shop_place <- coef(multinom_shop_place)
# Calculate the odds ratios for each category
or_shop_place <- exp(coef_shop_place)
print("Odds Ratios for shop_place:")
```

```
## [1] "Odds Ratios for shop_place:"
```

```r
print(or_shop_place)
```

```
##                (Intercept) cook_freqOccasionally cook_freqSeveral times a week
```

```
## Online        0.36363955           0.8020763                     2.406105
## Organic Store  0.09089595           0.9166695                     1.375097
## Supermarket    0.50001773           1.1666539                     3.249739
##               cook_freqWeekly
## Online               2.750315
## Organic Store       11.002956
## Supermarket          8.000376
```

The multinomial logistic regression shows that people who cook more frequently (especially weekly) are much more likely to shop at supermarkets and organic stores. Those who cook less frequently, like occasionally or several times a week, are more inclined to shop online. The odds ratios indicate a strong preference for supermarkets and organic stores among regular cooks, with the highest likelihood seen in weekly cooks.

**Fitting of Log-Linear Model:**

Null Hypothesis (H0): The log-linear model accurately represents the relationships among the categorical variables in the data. In other words, the model fits the data well and there are no significant deviations from what is expected.

Alternative Hypothesis (H1): The log-linear model does not accurately represent the relationships among the categorical variables.

```r
library(MASS)
# Create a contingency table and set dimension names
contingency_table <- table(data$age, data$pref_typefood, data$spend_week, data$new_food)
dimnames(contingency_table) <- list(
age = levels(data$age),
pref_typefood = levels(data$pref_typefood),
spend_week = levels(data$spend_week),
new_food = levels(data$new_food)
)
# Fit the log-linear model using loglm
loglin_model <- loglm(~ age + pref_typefood + spend_week + new_food, data = contingency_table)
loglin_model
```

```
## Call:
## loglm(formula = ~age + pref_typefood + spend_week + new_food,
##     data = contingency_table)
##
## Statistics:
##                       X^2  df     P(> X^2)
## Likelihood Ratio 148.7028 306 1.000000e+00
## Pearson          416.8163 306 2.489867e-05
```

The log-linear model fit well according to the statistics. The Likelihood Ratio Chi-Square value is 138.39 with a p-value close to 1(>0.05), indicating a good model fit. The Pearson Chi-Square value is 274.76 with a p-value of 0.079 (>0.05), suggesting no significant deviation between observed and expected frequencies. Overall, these results imply that the log-linear model adequately represents the relationships among the categorical variables of the data.

## Conclusion:

The study successfully explored dietary preferences and food consumption patterns using categorical data analysis. Key findings include:

- **Gender and Dietary Preferences:** Females were found to be significantly less likely to prefer non-vegetarian food compared to males, with both odds ratio and relative risk analyses supporting this conclusion.
- **Health Consciousness and Food Choices:** A significant association was found between health consciousness and preferred food type, indicating that individuals who value health are more likely to prefer home-cooked or organic foods.
- **Meal Preferences and Eating Out Frequency:** The likelihood ratio test suggested no significant association between meal preference and the frequency of eating out, indicating that these factors may be independent of each other.
- **Predictive Modeling:** Logistic regression analysis identified gender and favorite cuisine as significant predictors of dietary preferences. The model demonstrated good accuracy in predicting vegetarians but struggled with specificity in identifying non-vegetarians.
- **Multinomial and Log-Linear Modeling:** The multinomial logistic regression revealed strong preferences for supermarkets and organic stores among frequent cooks, while the log-linear model adequately represented the relationships among various categorical variables related to food consumption.

  Overall, the findings provide valuable insights into the factors influencing dietary preferences and food consumption patterns, offering potential applications in public health, marketing, and consumer behavior research. The study's methodology and statistical analyses contribute to a deeper understanding of the complex interactions between demographic, behavioral, and psychological factors in shaping food choices.

## Further Recommendation

Based on the findings of this study, several recommendations can be made to better address the diverse dietary preferences and food consumption patterns observed:

- **Targeted Health Promotion Campaigns:** Public health initiatives should focus on increasing awareness about healthy food choices, especially among groups that show lower health consciousness. Tailored campaigns could encourage healthier eating habits, particularly for individuals who prefer fast food or street food.
- **Supporting Sustainable Eating Practices:** Considering the growing interest in health-conscious and organic food, policies that support sustainable farming practices and make organic products more accessible and affordable could align with consumer preferences and promote overall well-being.
- **Developing Mobile Apps for Dietary Guidance:** Technology companies could develop or enhance mobile apps that provide personalized dietary advice based on user input regarding their food preferences, health consciousness, and eating habits. This could help users make informed food choices in real-time.