Nutrition studies

Racheal\_wanyoike

2022-06-20

## Nutrition study R Markdown

This dataset contains information on the eating,general habits and health of 54 respondents who completed a nutrition survey similar to a food frequency questionnaire.This data will be used to find associations between food and different outcomes using P-values.The data is in a csv file, flat file database ,structured (rows and column), delimited format.Delimiters used commas and quotes.

## To read the data, I used the read\_csv() in the readr package.

nutrition\_data <- read.csv ("https://raw.githubusercontent.com/fivethirtyeight/data/master/nutrition-studies/raw\_anonymized\_data.csv")

## Next, I cleaned the data by selecting 12 variables that may have an association to cancer.

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

cancerlinktofoods <- select(nutrition\_data,cancer,BEEFPORKDISHFREQ,DRIEDFRUITFREQ,OTHERFRESHFRUITFREQ,CORNFREQ,SUGARINCOFFEE,SALTFREQ,HOTTEAFREQ,RAWTOMATOESFREQ,MILKFREQ,BUTTERFREQ,OTHERBREADSFREQ, BROCCOLIFREQ)

The dataframe cancerlinktofoods has 54 rows and 13 columns. The names of the columns and a brief description of each are in the table below:

description\_table <- data.frame(  
 Names = c("cancer","BEEFPORKDISHFREQ","DRIEDFRUITFREQ","OTHERFRESHFRUITFREQ","CORNFREQ","SUGARINCOFFEE","SALTFREQ","HOTTEAFREQ","RAWTOMATOESFREQ","MILKFREQ","BUTTERFREQ","OTHERBREADSFREQ","BROCCOLIFREQ"),  
Descriptions = c("Respondents have/ever have cancer?Yes/No","Frequency of eating beefporkdishes","frequency of eating dried fruits","frequency of eating other fresh fruits","frequency of eating corn","frequency of putting sugar in coffee","frequency of salt added","frequency of drinking hot tea","frequncy of eating raw tomatoes","frequency of drinking milk","frequncy of using/adding butter to food","frequency of eating bread other than cornbread","frequency of eating broccoli")   
)  
description\_table

## Names Descriptions  
## 1 cancer Respondents have/ever have cancer?Yes/No  
## 2 BEEFPORKDISHFREQ Frequency of eating beefporkdishes  
## 3 DRIEDFRUITFREQ frequency of eating dried fruits  
## 4 OTHERFRESHFRUITFREQ frequency of eating other fresh fruits  
## 5 CORNFREQ frequency of eating corn  
## 6 SUGARINCOFFEE frequency of putting sugar in coffee  
## 7 SALTFREQ frequency of salt added  
## 8 HOTTEAFREQ frequency of drinking hot tea  
## 9 RAWTOMATOESFREQ frequncy of eating raw tomatoes  
## 10 MILKFREQ frequency of drinking milk  
## 11 BUTTERFREQ frequncy of using/adding butter to food  
## 12 OTHERBREADSFREQ frequency of eating bread other than cornbread  
## 13 BROCCOLIFREQ frequency of eating broccoli

# Next, I picked three columns to obtain summary:

library(dplyr)  
sub\_cancerlinktofoods <- select(cancerlinktofoods,BEEFPORKDISHFREQ,RAWTOMATOESFREQ,BROCCOLIFREQ)

#Then, saved summary of the three columns in object summary table

summarytable <- summary(sub\_cancerlinktofoods)