Assignment 3

Aishat Olatunji

2022-09-09

#file.choose()  
income\_data <- read.csv("/Users/aishatolatunji/Documents/Data Analytics/Income Data.csv")  
summary(income\_data)

## job type income education   
## Length:45 Length:45 Min. : 7.00 Min. : 7.00   
## Class :character Class :character 1st Qu.:21.00 1st Qu.: 26.00   
## Mode :character Mode :character Median :42.00 Median : 45.00   
## Mean :41.87 Mean : 52.56   
## 3rd Qu.:64.00 3rd Qu.: 84.00   
## Max. :81.00 Max. :100.00   
## prestige   
## Min. : 3.00   
## 1st Qu.:16.00   
## Median :41.00   
## Mean :47.69   
## 3rd Qu.:81.00   
## Max. :97.00

head(income\_data)

## job type income education prestige  
## 1 accountant prof 62 86 82  
## 2 pilot prof 72 76 83  
## 3 architect prof 75 92 90  
## 4 author prof 55 90 76  
## 5 chemist prof 64 86 90  
## 6 minister prof 21 84 87

attach(income\_data) # The purpose of using this attach is for r to know all the variables present in my data.

#Number one  
summary(income\_data['income'])

## income   
## Min. : 7.00   
## 1st Qu.:21.00   
## Median :42.00   
## Mean :41.87   
## 3rd Qu.:64.00   
## Max. :81.00

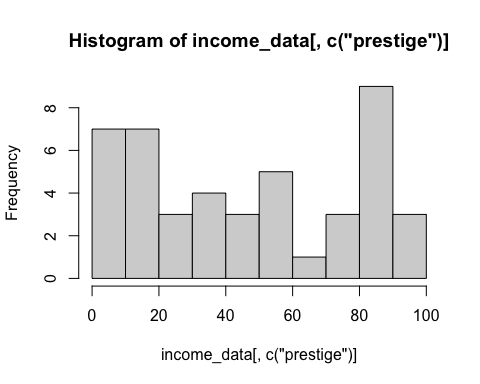
summary(income\_data['education'])

## education   
## Min. : 7.00   
## 1st Qu.: 26.00   
## Median : 45.00   
## Mean : 52.56   
## 3rd Qu.: 84.00   
## Max. :100.00

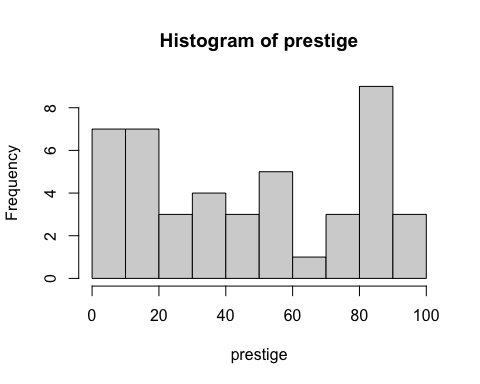
summary(income\_data['prestige'])

## prestige   
## Min. : 3.00   
## 1st Qu.:16.00   
## Median :41.00   
## Mean :47.69   
## 3rd Qu.:81.00   
## Max. :97.00

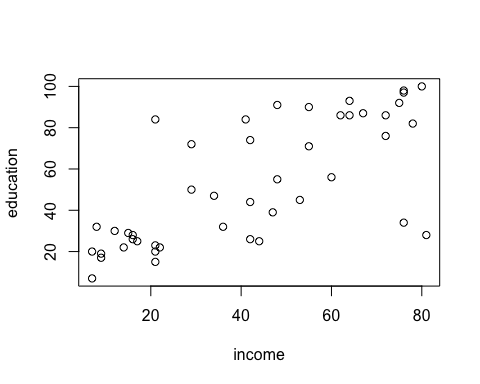
#Number two  
hist(income\_data[,c('prestige')])



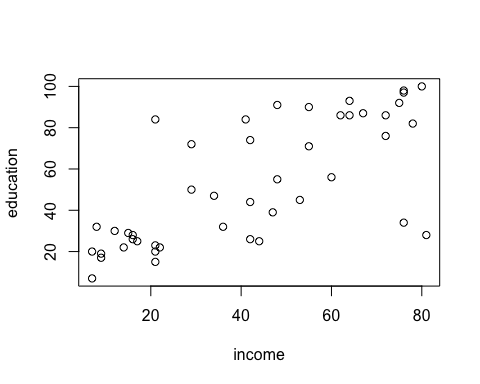
# or  
hist(prestige)



#Number three  
plot(income,education)



# or  
plot(income\_data[,c('income','education')])



#Number 4  
var(income)

## [1] 597.0727

## ii  
sqrt(var(income))

## [1] 24.43507

#or  
var(income\_data[,c("income")])

## [1] 597.0727

sqrt(var(income\_data[,c("income")])  
)

## [1] 24.43507

# number 5  
cor(income,education, method = c("pearson"))

## [1] 0.7245124

# we have quite a strong positive relationship between the two variables i.e the income and education.  
  
# Finding the pvalue between the income and education  
library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

rcorr(income,education, type="pearson")

## x y  
## x 1.00 0.72  
## y 0.72 1.00  
##   
## n= 45   
##   
##   
## P  
## x y   
## x 0  
## y 0

# The pvalue is 0 (zero) which is signicantly small,   
  
# setting my hypothesis  
#Ho: There is no relationship  
#H1: There is a significant relationship  
  
# Decision rule: reject Ho(null hypothesis) if the pvalue is less than alpha(0.05) otherwise do not reject Ho  
#Decision: since,i tried to compare it at 95% significant level which is also less than my alpha(0.05). therefore i reject the null hypothesis and conclude that this result shows that there is a significant and positive relationship between the two variables.