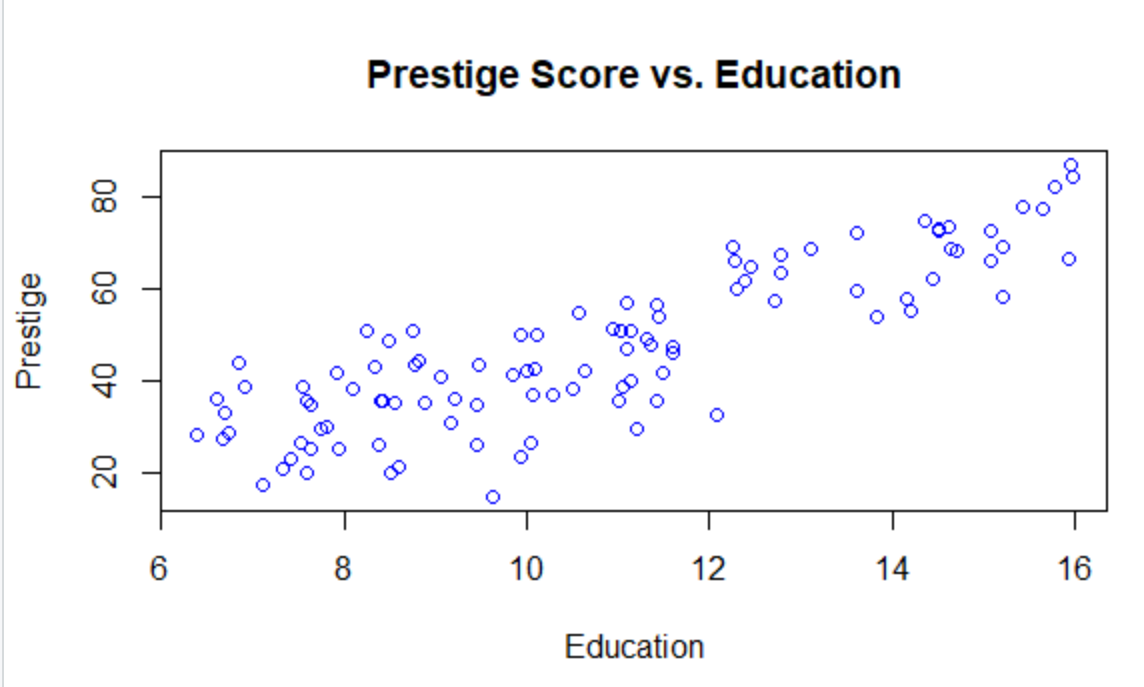
1. 

The correlation between year of education and prestige score is .8501769

The form of the scatterplot is linear.

The direction of the scatterplot is positive.

The strength of the association of the variables is strong.

2.)

Chart, scatter chart

Description automatically generated

Assumptions:

Linear: Yes, there is no distinct pattern

Independent - Yes, the observations are independent

Constant Variation - variability decreases slightly when X is high

Normally Distributed - yes, it’s normally distributed according to the below histogram

I think that education, prestige point (9.62, 14.8) is an outlier aka(3.2, 14.8) on the residual plot. Since it is an outlier on the y axis, it is not an influence point

Chart, histogram

Description automatically generated

3.)

Text

Description automatically generated

Text

Description automatically generated

Step 1

* H0: BEducation = BIncome = BWoman = 0
* H1: BEducation ≠0 and/or BIncome ≠0 and/or BWoman ≠0
* α=0.05

Step 2

Test Statistic:

K = 3, N – K – 1 = 102 – 3 -1 = 98

* F=MS Reg/MS Res with 3 and 98 degrees of freedom

Step 3

F Critical Value

3, 98 degrees of freedom

α = .05

Decision Rule: Reject H0 if F is greater than or equal to 2.697423

Step4

F Statistic:

* F= 129.2

Step 5

Since the F statistic is greater than the f critical value, we reject the null hypothesis that there is no linear association

R Squared = 0.78 which signifies that 78% of the variability in prestige can be attributed to the combination of the other 3 variables

4.)

T critical value = 2.697423

Text

Description automatically generated

Both Income and Education are significant in determining Prestige because the t value is greater than 2.697

Text, letter

Description automatically generated

The estimate coefficients imply that for every year of education, the prestige score will be 4.18 higher, and for every dollar per month, the prestige score will be .000762 higher. In other words, for every extra 1000 per month, the prestige score will be 7.6 higher.

The 95% confidence interval for education is between (3.415,4.958)

The 95% confidence interval for income is between (.000762, .00186)

Diagram, schematic

Description automatically generated

5.) Yes the fit of the model is reasonable, R squared is .7982

Chart, scatter chart

Description automatically generated

rm(list=ls()); cat("\014")

#Set directory

setwd("C:/Users/HP/Documents/555")

getwd()

#1

#Import spreadsheet

census = read.csv("Census.csv", header = TRUE)

as.data.frame(census)

plot(census$Education, census$Prestige, col = "blue",

xlab= "Education", ylab = "Prestige",

main = "Prestige Score vs. Education")

cor(census$Education, census$Prestige)

# 2

# Perform Linear Regression.

m <- lm(census$Education ~ census$Prestige)

m

resid(m)

plot(census$Prestige, resid(m),

ylab="Residuals", xlab="Prestige",

main="Residual Plot", col = "purple")

abline(0,0)

#Check residual for normal distribution

hist(resid(m))

fitted(m)

#3.)

m2 <- lm(census$Prestige ~ census$Education + census$Income + census$Women)

m2

#find N

length(census$Education)

#f critical value

qf(.95,3,98)

summary(m2)

#4.)

confint(m2, level = .95)

anova(m2)

summary(m2)

#find t critical value

qt(.025,98)

#Find correlation between two significant variables < .8 ok

cor(census$Education, census$Income)

#plot

par(mfrow = c(2 ,2))

plot(census$Education , resid(m2) , axes = TRUE , frame.plot = TRUE , xlab = "Education" , ylab = "Residual")

plot(census$Income , resid(m2) , axes = TRUE , frame.plot = TRUE , xlab = "Income" , ylab = "Residual")

plot(census$Women , resid(m2) , axes = TRUE , frame.plot = TRUE , xlab = "Women" , ylab = "Residual")

#5.)

resid(m2)

fitted(m2)

plot(fitted(m2), resid(m2), axes = TRUE , frame.plot = TRUE,

xlab = "Fitted values", ylab = "Residual", col = "green")

abline(0,0)