Name: - Shiva Kumar Peddapuram

Registration Number: - 811235874

college mail id: - [speddapu@kent.edu](mailto:speddapu@kent.edu)

submitted to: - Jun Li

subject: - Applied Statistics (MATH 50015-001)

* + 1. Identify the predictor and the response.

Answer: -

**Predictor: - region, group, Fertility, lifeExpF, pctUrban,**

Response: - Per capita gross domestic product in US dollars (ppgdp)

Draw the scatterplot of fertility on the vertical axis versus ppgdp on the horizontal axis and summarize the information in this graph. Does a straight-line mean function seem to be plausible for a summary of this graph?

Answer:-

getwd() ## getting the present working directory

install.packages("alr4") #installling all the required packages

library(alr4)

a <- UN11 #assigning the data package

install.packages("ggplot2") #instaling ggplot2 package for

library(ggplot2)

ggplot(data=a, aes(x=fertility,y=ppgdp))+geom\_point(pch=16,col=3)+ggtitle("scatter plot")

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

* + 1. Draw the scatterplot of log(fertility) versus log(ppgdp) using natural logarithms. Does the simple linear regression model seem plausible for a summary of this graph? If you use a different base of logarithms, the shape of the graph won’t change, but the values on the axes will change

Code:-

##1.1.3

install.packages("alr4")

library(alr4)

install.packages("ggplot2")

library(ggplot2)

b <- UN11

b

p2 <- ggplot(b, aes(x=log(ppgdp), y=log(fertility))) +geom\_point() +labs(title= "log comparison",x= "GNP",y = "Fertility") +theme\_minimal()

p2

Graphical user interface, text, application, email

Description automatically generated

Chart, scatter chart

Description automatically generated

2.1.1 Draw a scatterplot of wt on the vertical axis versus ht on the horizontal axis. On the basis of this plot, does a simple linear regression model make sense for these data? Why or why not?

Solution:-

data<-read.csv("htwt.csv",header = T, sep=",")

#2.1.1

#scatter plot

height=data$ht

weight=data$wt

par(cex=.8)

plot(height, weight, main="Height-Weight Scatterplot", xlab="Height", ylab="Weight", pch=19)

Graphical user interface, application

Description automatically generated

2.1.2 Show that x = 165.52, y = 59.47, SXX = 472.08, SYY = 731.96, and SXY = 274.79. Compute estimates of the slope and the intercept for the regression of Y on X. Draw the fitted line on your scatterplot.

Code:-

#2.1.2

library(alr4)

data(Htwt)

head(Htwt)

input <- Htwt[,c('ht','wt')]

x = input$ht

y = input$wt

mean (x)

sxx <- sum((x - mean(x))^2)

sxx

syy <- sum((x - mean(x))^2)

syy

sxy <- sum((x-mean(x)) \* (y - mean(y)))

sxy

bta\_1 <- sxy/sxx

bta\_1

bta\_0 <- mean(y) - bta\_1\*mean(x)

bta\_0

plot(x = input$ht, y = input$wt)

abline(lm(y~x, data = Htwt),col = "green",lwd=3)

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

2.1.3 Obtain the estimate of σ2 and find the estimated standard errors of ˆ β0 and ˆ β1. Also find the estimated covariance between ˆ β0 and ˆ β1. Compute the t-tests for the hypotheses that β0 = 0 and that β1 = 0 and find the appropriate p-values using two-sided tests.

Solution:-

summary(fit)  
# Beta-0 std. error = 64.4728  
# Beta-1 std. error = 0.3892

#t-value Beta-0 = -.572  
#t-value Beta-1= 1.496

#p-value Beta-0=.583  
#p-value Beta-1=.173

var(data$ht)  
#estimated covarience  
cov(data$ht, data$wt)

Graphical user interface, text, application

Description automatically generated

Reference :-

<https://rpubs.com/>

r blogs

r tutorial