library(UsingR)

library("ggpubr")

setwd("/Users/estherji")

#Part 1

fish <- read.csv("ass3fish.csv", header=TRUE)

fish <- na.omit(fish)

head(fish)

summary(fish)

A picture containing text, receipt, screenshot

Description automatically generated

#(1)

plot(fish$Number.of.meals.with.fish, fish$Total.Mercury.in.mg.g, xlab = 'Number of meals with fish', ylab = 'Total mercury', main = 'scatterplot between meals with fish and Mercury')

Chart, scatter chart

Description automatically generated

#From the image. There might be a positive correlation between Number of meals with fish and Total Mercury. When the fisherman eats more fish, they have more mercury in their body.

#(2)

cf<- cor(fish$Number.of.meals.with.fish, fish$Total.Mercury.in.mg.g, use="complete.obs")

#0.6991094

#Since the correlation coefficient is 0.6991094. It is not 0. Therefore, there is a positive correlation between number of meals with fish and total mercury in body.

#(3)

xbar <- mean(fish$Number.of.meals.with.fish)

sx <- sd(fish$Number.of.meals.with.fish)

ybar <- mean(fish$Total.Mercury.in.mg.g)

sy<-sd(fish$Total.Mercury.in.mg.g)

r <- cf

b1 <- r\*sy/sx

b0 <- ybar - (b1\*xbar)

b0

#1.687643

b1

#0.2759503

#y = 1.69+ 0.276x

flm <- lm( fish$Total.Mercury.in.mg.g ~ fish$Number.of.meals.with.fish)

abline( a = b0, b=b1)

ggscatter(fish,"Number.of.meals.with.fish", "Total.Mercury.in.mg.g",

add = "reg.line", conf.int = TRUE,

cor.coef = TRUE, cor.method = "pearson",

xlab = "Number of meals with fish", ylab = "Total mercury")

Graphical user interface, text, application, email

Description automatically generated

Chart, scatter chart

Description automatically generated

#(4)B1 is 0.276, this means the slope of this line is 0.276. Since 0.276 is a positive number, therefore, as Number.of.meals.with.fish increases, Total.Mercury.in.mg.g is also increase

#B0 is 1.69, this means the when x equals to 0, the intersection of this line and the y-axis. this also means when Number.of.meals.with.fish is 0, the Total.Mercury is 1.69.

#(5)

#h0: there is no linear association.

#h1: there is a linear association.

#df = 100-2 = 98

qf(0.9, df1 = 1, df2 = 98)

#2.75743 with a = 0.1

anova(flm)

#F = 93.689Graphical user interface, text, application

Description automatically generated

#Since 93.7 > 2.76, reject null hypothesis and conclude that there is linear relationship between Number.of.meals.with.fish and Total.Mercury.in.mg.g

summary(flm))Text, letter

Description automatically generated

#r\_squared is 0.4835. Since r-squared is between 0 to 1, and 1 means this model that explains all the variation in the response variable around its mean.

# Then, 0.485 means 48.8% of the data fit in the model

#Calculate and interpret the 90% confidence interval for B1.

#The 90% confidence interval means there are 90% sure that the value should between the upper and lower limits. Therefore, there are 90% chance that as x increased by 1, y will increase 0.229 to 0.323.

confint(flm, level = 0.9)

Text

Description automatically generated

R code:

library(UsingR)

library("ggpubr")

setwd("/Users/estherji")

#Part 1

fish <- read.csv("ass3fish.csv", header=TRUE)

fish <- na.omit(fish)

head(fish)

summary(fish)

#(1)

plot(fish$Number.of.meals.with.fish, fish$Total.Mercury.in.mg.g, xlab = 'Number of meals with fish', ylab = 'Total mercury', main = 'scatterplot between meals with fish and Mercury')

#From the images,there might be a possitive correlation between Number of meals with fish and Total Mercury. When the fisherman eat more fish, they have more mercury in their body.

#(2)

cf<- cor(fish$Number.of.meals.with.fish, fish$Total.Mercury.in.mg.g, use="complete.obs")

#0.6991094

#Since the correlation coefficient is 0.6991094. It is not 0. Therefore, there is a positive correlation bewteen number of meals with fish and total mercury in body.

#(3)

xbar <- mean(fish$Number.of.meals.with.fish)

sx <- sd(fish$Number.of.meals.with.fish)

ybar <- mean(fish$Total.Mercury.in.mg.g)

sy<-sd(fish$Total.Mercury.in.mg.g)

r <- cf

b1 <- r\*sy/sx

b0 <- ybar - (b1\*xbar)

b0

#1.687643

b1

#0.2759503

#y = 1.69+ 0.276x

flm <- lm( fish$Total.Mercury.in.mg.g ~ fish$Number.of.meals.with.fish)

abline( a = b0, b=b1)

ggscatter(fish,"Number.of.meals.with.fish", "Total.Mercury.in.mg.g",

add = "reg.line", conf.int = TRUE,

cor.coef = TRUE, cor.method = "pearson",

xlab = "Number of meals with fish", ylab = "Total mercury")

#(4)B1 is 0.276, this means the slope of this line is 0.276. Since 0.276 is a positive number, therefore, as Number.of.meals.with.fish increases, Total.Mercury.in.mg.g is also increase

#B0 is 1.69, this means the when x equals to 0, the intersection of this line and the y-axis. this also means when Number.of.meals.with.fish is 0, the Total.Mercury is 1.69.

#(5)

#h0: there is no linear association.

#h1: there is a linear association.

#df = 100-2 = 98

qf(0.9, df1 = 1, df2 = 98)

#2.75743 with a = 0.1

anova(flm)

#F = 93.689

#Since 93.7 > 2.76, reject null hypothesis and concloud that there is linear relationship between Number.of.meals.with.fish and Total.Mercury.in.mg.g

summary(flm)

#r\_squared is 0.4835. Since r-squared is between 0 to 1, and 1 means this model that explains all the variation in the response variable around its mean.

# Then, 0.485 means 48.8% of the data fit in the model

#calculate and interpret the 90% confidence interval for B1.

#The 90% confidence interval means there are 90% sure that the value should between the upper and lower limits. Therefore, there are 90% chance that as x increased by 1, y will increase 0.229 to 0.323

confint(flm, level = 0.9)