STA108 Project I

Jasper Dong, Allison Peng

2023-05-01

R Appendix

```r
knitr::opts_chunk$set(echo = FALSE, comment = NA)
options(scipen = 999) #Remove the scientific notation
library(readr)
library(MASS)
library(ggplot2)
SENIC <- read_csv("SENIC (1).csv")
SENIC
# plot exploratory data analysis scatterplots
plot(SENIC$infection, SENIC$length, xlab = "Probability of Aquiring Infection in Hospital (Percent)", yl
plot(SENIC$facility, SENIC$length, xlab = "Percent of 35 Potential Facilities and Services Provided by I
plot(SENIC$Xray, SENIC$length, xlab = "Ratio of Number of X-rays Performed to Patients without Pneumonia

# find mean for each variable
mean(SENIC$length)
mean(SENIC$infection)
mean(SENIC$facility)
mean(SENIC$Xray)

# find SD for each variable
sd(SENIC$length)
sd(SENIC$infection)
sd(SENIC$facility)
sd(SENIC$Xray)

# removing outliers
outliers = c(47, 112)
SENIC1 = SENIC[-outliers,]
SENIC1
##### MODELS #####
infect.model = lm(length ~ infection, data = SENIC1)
facility.model = lm(length ~ facility, data = SENIC1)
xray.model = lm(length ~ Xray, data = SENIC1)
SENIC1$ei1 = infect.model$residuals
SENIC1$yhat1 = infect.model$fitted.values
SENIC1

# plotting the estimated regression line
plot(SENIC1$infection, SENIC1$length, xlab = "Probability of Aquiring Infection in Hospital (Percent)",
abline(new.model1,col = "purple",lwd = 2)
```

```r
plot(SENIC1$facility, SENIC1$length, xlab = "Percent of 35 Potential Facilities and Services Provided b
abline(new.model2,col = "purple",lwd = 2)

plot(SENIC1$Xray, SENIC1$length, xlab = "Ratio of Number of X-rays Performed to Patients without Pneumo
abline(new.model3,col = "purple",lwd = 2)

# finding the r^2 value for each model
cor(SENIC1$infection, SENIC1$length)^2
cor(SENIC1$facility, SENIC1$length)^2
cor(SENIC1$Xray, SENIC1$length)^2

 # ANOVA tables for each model
anova(infect.model)
anova(facility.model)
anova(xray.model)
  # Assessing Infection Linearity
options(scipen = 8)
reduced.model = lm(length ~ 1, data = SENIC1)
anova.table = anova(reduced.model, infect.model)
anova.table
  # Assessing Infection Independence
plot(infect.model$residuals, main = "Residual Index plot", xlab = "Index", ylab = "residuals", pch = 19
abline(h = 0, lty = 2)
# Confidence Interval for Parameters
alpha = 0.05
infect.CIs = confint(infect.model, level = 1-alpha)
infect.CIs
 # Shapiro-Wilks Test for Infection Normality
ei = infect.model$residuals
infect.SWtest = shapiro.test(ei)
infect.SWtest
 # Fligner-Killeen test for Infection Homoscedasticity
SENIC$ei1 = infect.model$residuals
Group = rep("Lower",nrow(SENIC1))
Group[SENIC1$length < median(SENIC1$length)] = "Upper"
Group = as.factor(Group)
SENIC1$Group = Group
infect.FKtest= fligner.test(SENIC1$ei1, SENIC1$Group)
infect.FKtest
```