**Sahiti Kovvuri R Homework #2**

**STA 4155**

**R Homework for Chapter 16**

**1. Let’s take Sales as response variable and Price as predictor variable. Fit a linear regression model to each of the four cities. Write down the four fitted models. In which city the pizza sales seem to be more sensitive to price than in others? Explain.**

BALTIMORE

**R code:**

plot(pizza$Baltimore.Price,pizza$Baltimore.Volume,xlab='price',ylab='volume')

#fitted model

imod <- lm(Baltimore.Volume~Baltimore.Price,data=pizza)

summary(imod)

summary(imod)$coefficients

confint(imod, level = 0.95)

**R output:**

**Sales = 126,625 – 34,956(Price)**

Chart, scatter chart

Description automatically generatedText

Description automatically generated

DALLAS

**R code:**

plot(pizza$Dallas.Price,pizza$Dallas.Volume,xlab='price',ylab='volume')

#fitted model

imod2 <- lm(Dallas.Volume~Dallas.Price,data=pizza)

summary(imod2)

summary(imod2)$coefficients

confint(imod2, level = 0.95)

**R output:**

**Sales = 139,457 – 33,527(Price)**

Chart, scatter chart

Description automatically generatedText

Description automatically generated

CHICAGO

**R code:**

plot(pizza$Chicago.Price,pizza$Chicago.Volume,xlab='price',ylab='volume')

#fitted model

imod3 <- lm(Chicago.Volume~Chicago.Price,data=pizza)

summary(imod3)

summary(imod3)$coefficients

confint(imod3, level = 0.95)

**R output:**

**Sales = 1,094,047 - 331,152(Price)**

Chart, scatter chart

Description automatically generatedText

Description automatically generated

DENVER

**R code:**

plot(pizza$Denver.Price,pizza$Denver.Volume,xlab='price',ylab='volume')

#fitted model

imod4<- lm(Denver.Volume~Denver.Price,data=pizza)

summary(imod4)

summary(imod4)$coefficients

confint(imod4, level = 0.95)

**R output:**

**Sales = 181,218 – 52,796(Price)**

Chart, scatter chart

Description automatically generatedText

Description automatically generated

ANSWER: I believe that the city that seems to be the most sensitive to the price is Chicago because it has the largest slope (x hat), meaning that for every increase in price (increments of $1), the volume decreases relatively by $331,152. In comparison to the other cities in the data given, this slope is the highest.

**2. For each of the models fitted above produce a residual plot in the time order, a residual plot against the fitted values, and a Q-Q plot. Is there any regression assumption violated in each model? Explain.**

BALTIMORE

**R code:**

pizza$Week <- as.Date(pizza$Week,format = "%m/%d/%y")

par(mfrow = c(1, 2))

plot(pizza$Week,imod$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod$fitted.values, imod$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#Q-Q plot Baltimore

par(mfrow = c(1, 2))

hist(imod$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod$residuals)

qqline(imod$residuals)

**R output:**

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

ANSWER: In the Baltimore plots, the regression assumption is violated. This is true because there is no symmetric distribution (pictured to the left), the histogram shows no skewness. and the QQ plot shows a slight stray from the regression line (pictured to the right).

DALLAS

**R code:**

par(mfrow = c(1, 2))

plot(pizza$Week,imod2$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod2$fitted.values, imod2$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ plot

par(mfrow = c(1, 2))

hist(imod2$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod2$residuals)

qqline(imod2$residuals)

**R output:**

Chart, diagram, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

ANSWER: In the Dallas plots the regression assumption is violated. This is true because there is no symmetric distribution (pictured to the left), the histogram shows right skewness, and the QQ plot shows a slight stray from the regression line (pictured to the right).

CHICAGO

**R code:**

par(mfrow = c(1, 2))

plot(pizza$Week,imod3$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod3$fitted.values, imod3$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ plot

par(mfrow = c(1, 2))

hist(imod3$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod3$residuals)

qqline(imod3$residuals)

**R output:**

Chart, scatter chart

Description automatically generatedChart

Description automatically generated

ANSWER: In the Chicago plots, the regression assumption is violated. This is true because there is no symmetric distribution (pictured to the left), the histogram shows no skewness, and the QQ plot shows a slight stray from the regression line (pictured to the right).

DENVER

**R code:**

par(mfrow = c(1, 2))

plot(pizza$Week,imod4$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0) > plot(imod4$fitted.values, imod4$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ Plot

par(mfrow = c(1, 2))

hist(imod4$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod4$residuals)

qqline(imod4$residuals)

**R output:**

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

ANSWER: In the Denver plots, the regression assumption is violated. This is true because there is no symmetric distribution (pictured to the left), the histogram shows slight left skewness. and the QQ plot shows a slight stray from the regression line (pictured to the right).

**3. For the remaining questions let’s focus on the model for city Dallas. Show a 90% confidence interval for the slope of Price and interpret it. Based on the interval can we say there is a statistically significant linear relationship between Price and Sales volume? Explain.**

**R code:**

confint(imod2, level = 0.90)

**R output:**

Text

Description automatically generated with medium confidence

ANSWER: The bounds of 90% confidence interval for the slope of the Dallas Price is (-40655.79,

-26398.58),. Based on the output, we are 95% confident that as the Dallas Price increases in increments of 1, the sales on average will decrease between the interval of -40,655.79 and -26,398.58.

**4. Conduct a hypothesis test to see if there is a significant negative correlation between Price and Sales volume in city Dallas, i.e., test H0 : β1 = 0 vs Ha : β1 < 0. State your test conclusion.**

**R code:**

imod2 <- lm(pizza$Dallas.Volume~pizza$Dallas.Price,data=pizza)

summary(imod2)

**R output:**

Text

Description automatically generated with medium confidence

ANSWER:The p-value here is < α, therefore we reject H0 and conclude β1 <0 Price and Sales Volume are negatively correlated. We know the p value is if 9.62e-13 which is less than 0.05.

**5. For city Dallas estimate the mean Sales if the Price is $2.50 and $3.00 using 95% confidence intervals. Interpret both intervals. Can we also estimate the mean Sales if the Price is $3.50? Explain.**

**R code:**

# if the price is $2.5

range(pizza$Dallas.Price)

new<-data.frame(Dallas.Price=c(2.50))

predict(imod2,newdata = new,interval='confidence',level=0.95)

**R output:**

Text

Description automatically generated

**R code:**

#if the price is $3

new1<-data.frame(Dallas.Price=c(3.00))

predict(imod2,newdata = new1,interval='confidence',level=0.95)

**R output:**

A picture containing text

Description automatically generated

ANSWER: The predicted mean sales for Dallas is $54,053.11 (look at the fit output) and its confidence interval is (52,604.49, 55,500.73) 🡪 lower and upper bounds respectfully.

We can say that we are 95% confident that the mean sales volume for Dallas of all years with the price of $2.50 is between $52,604.49 and $55,500.73.

However, based on this estimate, you cannot estimate the mean Sales of Dallas if the Price is $3.50 because the maximum range for the price is 3.05.

6. For city Dallas we know the pizza price was $2.77 in the last week of 1996. Suppose the price would increase to $2.99 in the following week. Can you predict the sales for that week and account for the uncertainty of your prediction? Do you think the resulting prediction is useful? Explain.

**R code:**

dallasnew <-lm(pizza$Dallas.Volume~pizza$Dallas.Price,data=pizza)

new2<-data.frame(Dallas.Price=c(2.99))

predict(dallasnew,newdata = new2,interval='prediction',level=0.95)

**R output:**

Text

Description automatically generated

ANSWER: Based on the output, the predicted sales for the following week in Dallas is $54,053.11 and its prediction interval is (37,465.11, 70,641.11) 🡨 lower and upper bounds respectfully. With the confidence of 95%, we can say the sales for the following weeks is between the bounds of $37,465.11 and $70,641.11 if the Price is $2.99.

**Appendix for R code:**

pizza <- read.table('Frozen\_Pizza.txt', sep = '\t', header = TRUE)

#Baltimore

plot(pizza$Baltimore.Price,pizza$Baltimore.Volume,xlab='price',ylab='volume')

#fitted model

imod <- lm(Baltimore.Volume~Baltimore.Price,data=pizza)

summary(imod)

summary(imod)$coefficients

confint(imod, level = 0.95)

#Dallas

plot(pizza$Dallas.Price,pizza$Dallas.Volume,xlab='price',ylab='volume')

#fitted model

imod2 <- lm(Dallas.Volume~Dallas.Price,data=pizza)

summary(imod2)

summary(imod2)$coefficients

confint(imod2, level = 0.95)

#chicago

plot(pizza$Chicago.Price,pizza$Chicago.Volume,xlab='price',ylab='volume')

#fitted model

imod3 <- lm(Chicago.Volume~Chicago.Price,data=pizza)

summary(imod3)

summary(imod3)$coefficients

confint(imod3, level = 0.95)

#Denver

plot(pizza$Denver.Price,pizza$Denver.Volume,xlab='price',ylab='volume')

#fitted model

imod4<- lm(Denver.Volume~Denver.Price,data=pizza)

summary(imod4)

summary(imod4)$coefficients

confint(imod4, level = 0.95)

#2

#Baltimore

pizza$Week <- as.Date(pizza$Week,format = "%m/%d/%y")

par(mfrow = c(1, 2))

plot(pizza$Week,imod$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod$fitted.values, imod$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#Q-Q plot Baltimore

par(mfrow = c(1, 2))

hist(imod$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod$residuals)

qqline(imod$residuals)

#Dallas

par(mfrow = c(1, 2))

plot(pizza$Week,imod2$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod2$fitted.values, imod2$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ plot

par(mfrow = c(1, 2))

hist(imod2$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod2$residuals)

qqline(imod2$residuals)

#Chigago

par(mfrow = c(1, 2))

plot(pizza$Week,imod3$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0)

plot(imod3$fitted.values, imod3$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ plot

par(mfrow = c(1, 2))

hist(imod3$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod3$residuals)

qqline(imod3$residuals)

#Denver

par(mfrow = c(1, 2))

plot(pizza$Week,imod4$residuals,xlab = 'Time', ylab = 'Residual')

abline(a = 0, b = 0) > plot(imod4$fitted.values, imod4$residuals, xlab = 'Fitted value', ylab = 'Residual')

abline(a = 0, b = 0)

#QQ Plot

par(mfrow = c(1, 2))

hist(imod4$residuals, xlab = 'Residual', main = 'Histogram of Residuals')

qqnorm(imod4$residuals)

qqline(imod4$residuals)

#3

confint(imod2, level = 0.90)

#4

imod2 <- lm(pizza$Dallas.Volume~pizza$Dallas.Price,data=pizza)

summary(imod2)

#5

# if the price is $2.5

range(pizza$Dallas.Price)

new<-data.frame(Dallas.Price=c(2.50))

predict(imod2,newdata = new,interval='confidence',level=0.95)

#if the price is $3

new1<-data.frame(Dallas.Price=c(3.00))

predict(imod2,newdata = new1,interval='confidence',level=0.95)

#6

dallasnew <-lm(pizza$Dallas.Volume~pizza$Dallas.Price,data=pizza)

new2<-data.frame(Dallas.Price=c(2.99))

predict(dallasnew,newdata = new2,interval='prediction',level=0.95)