Marcus Crowder
R Homework #2
Statistics 3155

Write down the four fitted models. In which city the pizza sales seem to be more sensitive to price than in others? Explain.

```
> pizza <- read.table('~/Downloads/Frozen_Pizza.txt', sep = '\t', header = TRUE)
> head(pizza)
       Week Baltimore.Volume Baltimore.Price Dallas.Volume Dallas.Price Chicago.Volume Chicago.Price Denver.Volume Denver.Price
1  1/8/1994            27982            2.76         58224         2.55         353412          2.34         58171         2.45
2 1/15/1994            26951            2.98         47699         2.74         264862          2.61         59348         2.40
3 1/22/1994            28782            2.78         59578         2.39         204975          2.77         63137         2.41
4 1/29/1994            32074            2.62         61595         2.49         208763          2.70         61271         2.29
5  2/5/1994            19765            2.81         64889         2.21         326558          2.45         70480         2.22
6 2/12/1994            22393            3.02         46388         2.75         176891          2.78         53496         2.48
> imod <- lm(Baltimore.Volume ~ Baltimore.Price, data = pizza)
> summary(imod)
```

Linear Model of Baltimore Sales Volume to Sales Price.

```
Call:
lm(formula = Baltimore.Volume ~ Baltimore.Price, data = pizza)

Residuals:
   Min     1Q Median     3Q    Max 
-16093  -5883  -1148   3663  58135 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       126625       8070   15.69   <2e-16 ***
Baltimore.Price   -34956       2821  -12.39   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9678 on 154 degrees of freedom
Multiple R-squared:  0.4992,	Adjusted R-squared:  0.4959 
F-statistic: 153.5 on 1 and 154 DF,  p-value: < 2.2e-16
```

```
> imod <- lm(Dallas.Volume ~ Dallas.Price, data = pizza)
> summary(imod)
```

Linear Model of Dallas Sales Volume to Dallas Sales Price

```
Call:
lm(formula = Dallas.Volume ~ Dallas.Price, data = pizza)

Residuals:
   Min     1Q Median     3Q    Max 
-14235  -6345  -1116   3553  28988 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    139547      11302  12.347  < 2e-16 ***
Dallas.Price   -33527       4308  -7.783 9.62e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8365 on 154 degrees of freedom
Multiple R-squared:  0.2823,	Adjusted R-squared:  0.2776 
F-statistic: 60.57 on 1 and 154 DF,  p-value: 9.618e-13
```

```
> imod <- lm(Chicago.Volume ~ Chicago.Price, data = pizza)
> summary(imod)
```

Linear Model of Chicago Sales Volume to Sales Price.

```
Call:
lm(formula = Chicago.Volume ~ Chicago.Price, data = pizza)

Residuals:
    Min      1Q  Median      3Q     Max 
-114091  -28590     636   22693  207755 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1094047      39746   27.53   <2e-16 ***
Chicago.Price  -331152      15140  -21.87   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40000 on 154 degrees of freedom
Multiple R-squared:  0.7565,	Adjusted R-squared:  0.7549 
F-statistic: 478.4 on 1 and 154 DF,  p-value: < 2.2e-16
```

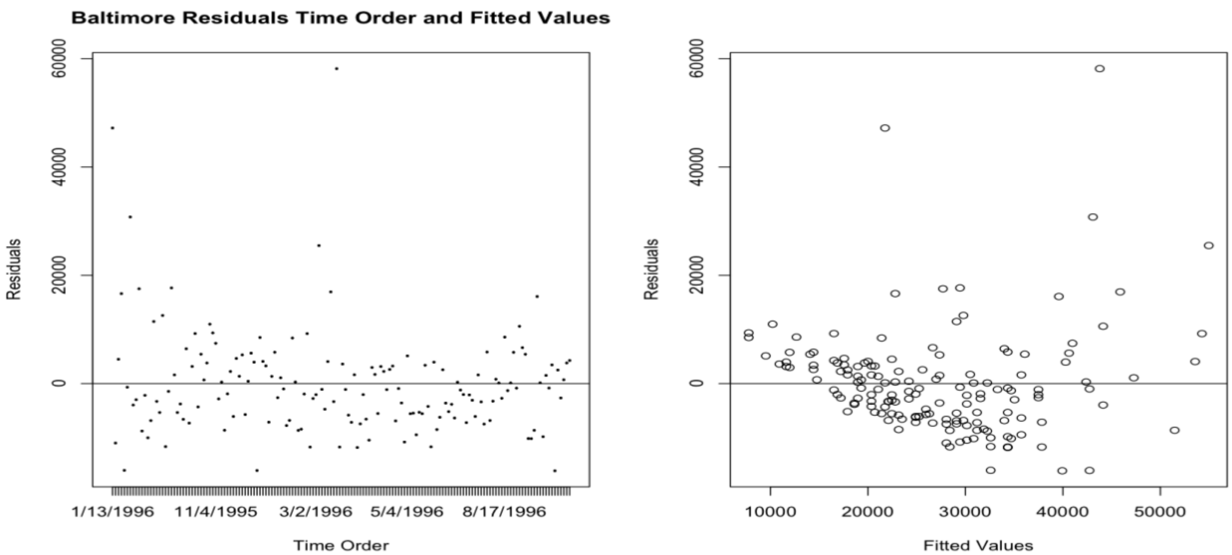
Answer: With The highest negative slope Chicago seems to be the most price sensitive city than others. Its price is in the -300000 thousand range while other slopes don’t even crack the negative hundred thousand range.

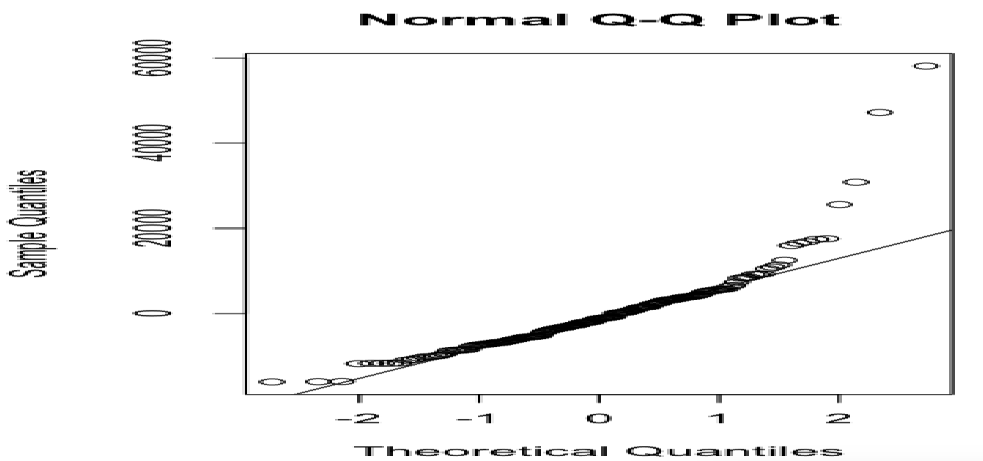
Linear Model of Denver Sales Volume to Denver Sales Price.

```
> imod <- lm(Denver.Volume ~ Denver.Price, data = pizza)
> summary(imod)

Call:
lm(formula = Denver.Volume ~ Denver.Price, data = pizza)

Residuals:
     Min       1Q   Median       3Q      Max
-15890.1  -5173.3   -352.9   3570.1  28762.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    181218       9077   19.96   <2e-16 ***
Denver.Price   -52796       3529  -14.96   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7859 on 154 degrees of freedom
Multiple R-squared:  0.5924,	Adjusted R-squared:  0.5898
F-statistic: 223.8 on 1 and 154 DF,  p-value: < 2.2e-16
```

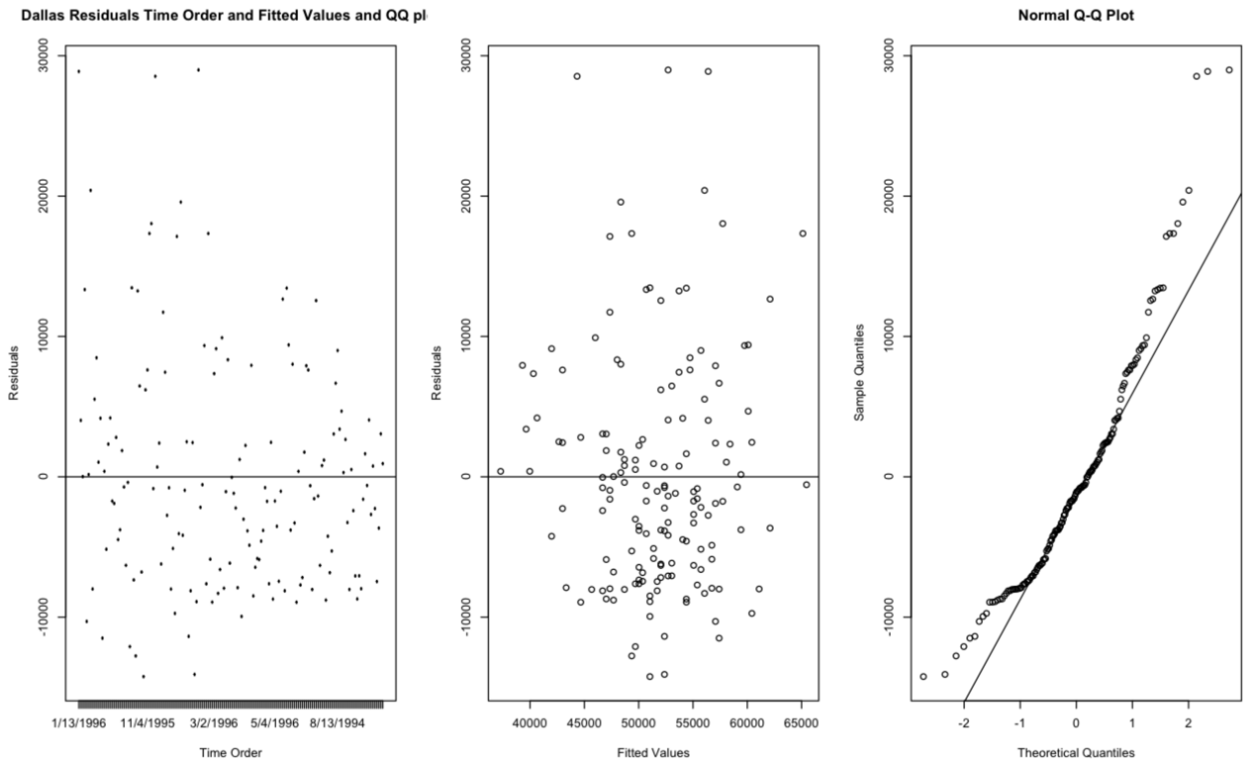
2. For each of the models fitted above produce a residual plot in the time order, a residual plot against the fitted values, and a Q-Q plot. Is there any regression assumption violated in each model? Explain.

```
> par(mfrow=c(1,2))

> plot(pizza$Week, imod1$residuals, ylab = "Residuals", xlab = "Time Order", main =
"Baltimore Residuals Time Order and Fitted Values")

> abline(a=0, b=0)

> plot(imod1$fitted.values, imod1$residuals, ylab = "Residuals", xlab = "Fitted Values")

> abline(a=0, b=0)

> qqnorm(imod1$residuals)

> qqline(imod1$residuals)
```

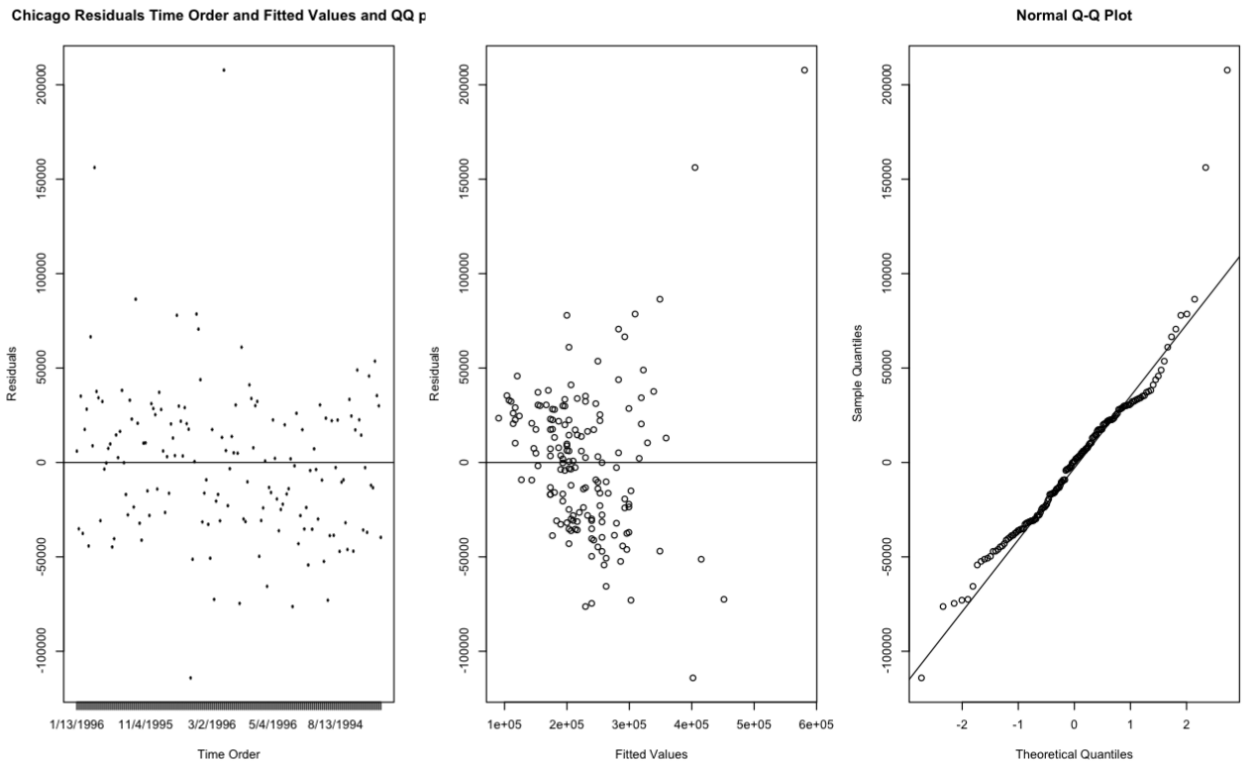

Baltimore Residuals Time Order and Fitted Values
Residuals
Time Order
1/13/1996
11/4/1995
3/2/1996
5/4/1996
8/17/1996
Fitted Values
10000
20000
30000
40000
50000
60000
0


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
0
20000
40000
60000
-2
-1
0
1
2

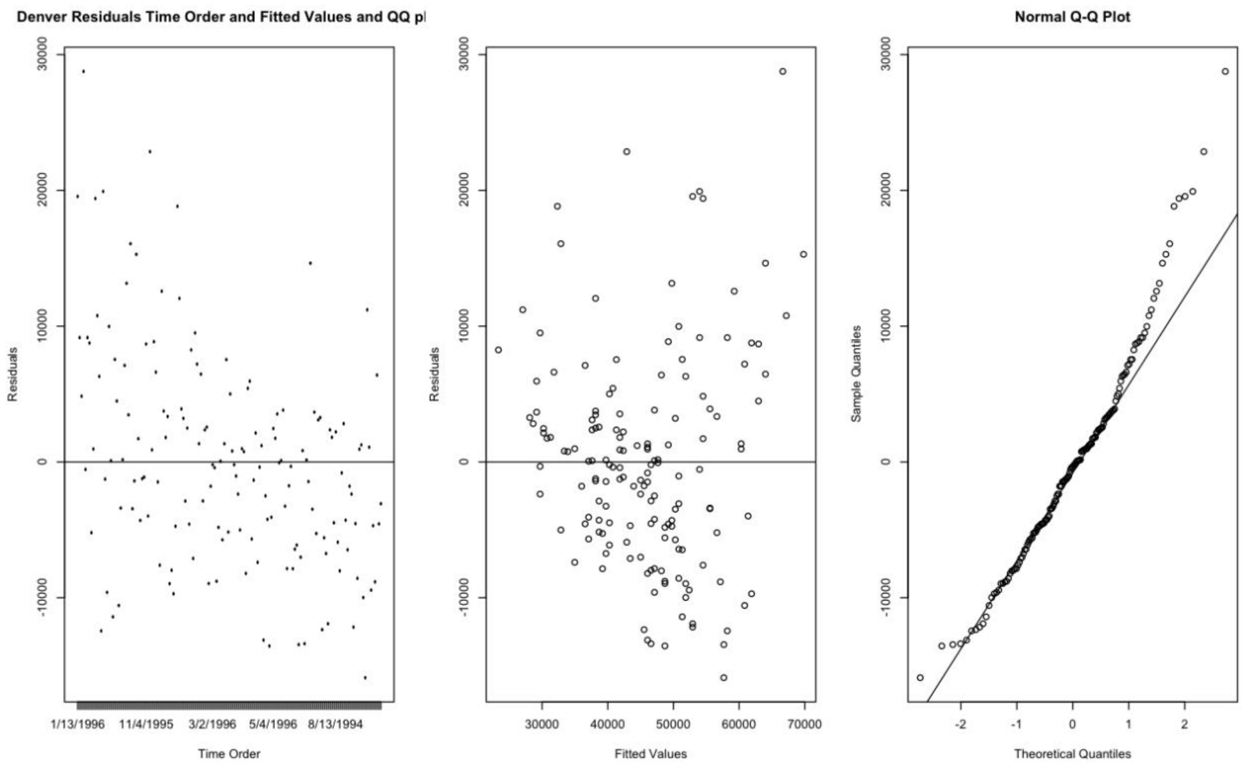
```
> par(mfrow=c(1,3))
> plot(pizza$Week, imod2$residuals, ylab = "Residuals", xlab = "Time Order", main = "Dallas Residuals Time Order and Fitted Values and QQ plot")
> abline(a=0, b=0)
> plot(imod2$fitted.values, imod2$residuals, ylab = "Residuals", xlab = "Fitted Values")
> abline(a=0, b=0)
> qqnorm(imod2$residuals)
> qqline(imod2$residuals)
```



```
> par(mfrow=c(1,3))
> plot(pizza$Week, imod3$residuals, ylab = "Residuals", xlab = "Time Order", main = "Chicago Residuals Time Order and Fitted Values and QQ plot")
> abline(a=0, b=0)
> plot(imod3$fitted.values, imod3$residuals, ylab = "Residuals", xlab = "Fitted Values")
> abline(a=0, b=0)
> qqnorm(imod3$residuals)
> qqline(imod3$residuals)
```



```
> par(mfrow=c(1,3))
> plot(pizza$Week, imod4$residuals, ylab = "Residuals", xlab = "Time Order", main = "Denver Residuals Time Order and Fitted Values and QQ plot")
> abline(a=0, b=0)
> plot(imod4$fitted.values, imod4$residuals, ylab = "Residuals", xlab = "Fitted Values")
> abline(a=0, b=0)
> qqnorm(imod4$residuals)
> qqline(imod4$residuals)
```



Question2 Answer: Besides some outliers the graphs seem to not violate any regression assumptions. The equal variance, Independence linearity and normality Assumptions seem to be satisfied on the graphs.

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.

```
> confint(imod2, level = .90)
                   5 %      95 %
(Intercept)  120844.48 158250.38
Dallas.Price -40655.79 -26398.58
```

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 $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 < 0$. State your test conclusion.

Answer: I would reject H0 because the p-value for the slope is close to zero which is a low p-value. With a p-value that low you should reject H0. By rejecting H0 we can conclude the slope is significant. The slope is also negative which brings us to the conclusion that the correlation between Price and Sales volume in Dallas has a significant negative correlation.

```
> summary(imod2)

Call:
lm(formula = Dallas.Volume ~ Dallas.Price, data = pizza)

Residuals:
   Min     1Q Median     3Q    Max 
-14235  -6345  -1116   3553  28988 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    139547      11302  12.347  < 2e-16 ***
Dallas.Price   -33527       4308  -7.783 9.62e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8365 on 154 degrees of freedom
Multiple R-squared:  0.2823,	Adjusted R-squared:  0.2776 
F-statistic: 60.57 on 1 and 154 DF,  p-value: 9.618e-13
```

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.

-For a mean price of \$2.50 we are 95% confident that sales volume will fall between 54063 and 57396

-For a mean price of \$3.00 we are 95% confident that sales volume will fall between 35464 and 42467

We can estimate the mean sales price if it is 3.50 but we shouldn't because it is not in the range of observed means. The range for Dallas is 2.21 to 3.05

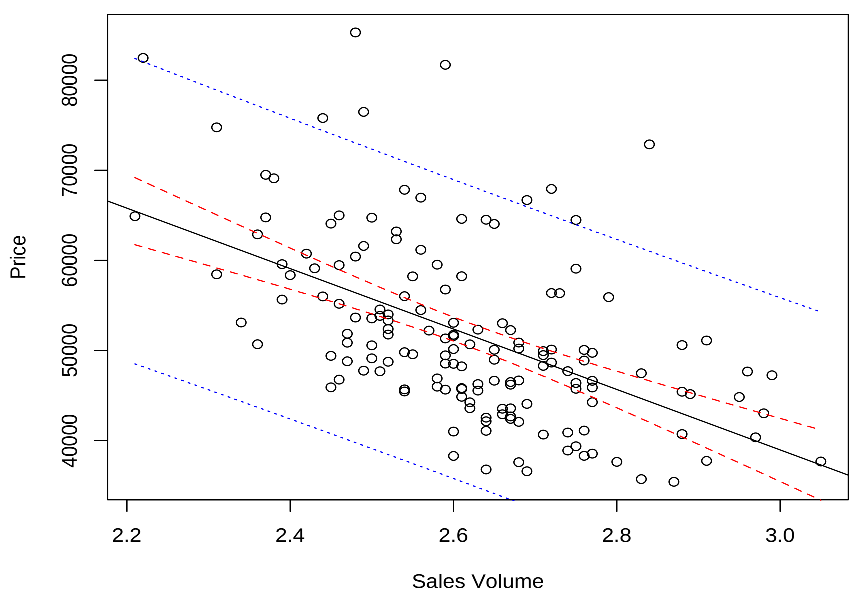
```
> new <- data.frame(Dallas.Price = c(2.50))
> predict(imod2, newdata = new, interval = 'confidence', level = .95)
       fit      lwr      upr
1 55729.47 54063.14 57395.79
> new <- data.frame(Dallas.Price = c(3.00))
> predict(imod2, newdata = new, interval = 'confidence', level = .95)
       fit      lwr      upr
1 38965.87 35464.31 42467.44
> new <- data.frame(Dallas.Price = c(2.50))
> predict(imod2, newdata = new, interval = 'confidence', level = .95)
       fit      lwr      upr
1 55729.47 54063.14 57395.79
> range(pizza$Dallas.Price)
[1] 2.21 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.

> xx <- seq(min(pizza$Dallas.Price), max(pizza$Dallas.Price), length.out = 100)

> conf <- predict(imod2, newdata = new.band, interval = 'confidence', level = .95)

> pred <- predict(imod2, newdata = new.band, interval = 'prediction', level = .95)



```
> plot(pizza$Dallas.Volume, pizza$Dallas.Price, xlab = 'Price', ylab = 'Sales Volume')
> plot(pizza$Dallas.Volume, pizza$Dallas.Price, ylab = 'Price', xlab = 'Sales Volume')
> plot(pizza$Dallas.Price, pizza$Dallas.Volume, ylab = 'Price', xlab = 'Sales Volume')
> abline(imod2)
> lines(xx, conf[, 'lwr'], lty = 2, col = 'red')
> lines(xx, conf[, 'upr'], lty = 2, col = 'red')
> lines(xx, pred[, 'lwr'], lty = 3, col = 'blue')
> lines(xx, pred[, 'upr'], lty = 3, col = 'blue')
>
```

Answer: By looking at the graph we can see the prediction interval compared to the Confidence Interval. The prediction interval however does not account for all values within the bands. We can predict the interval but I would not consider it that useful. Even though the prediction interval is wider I would have a hard time accounting for the outliers.