Wacaha

Pricing and Retail Analytics

Natalie Kvochak, Maryclare Leonard, Sean O'Grady

3/29/31

1 Technical section

This is where your technical material should go. You might start by reading in the data.

1.1 Data setup

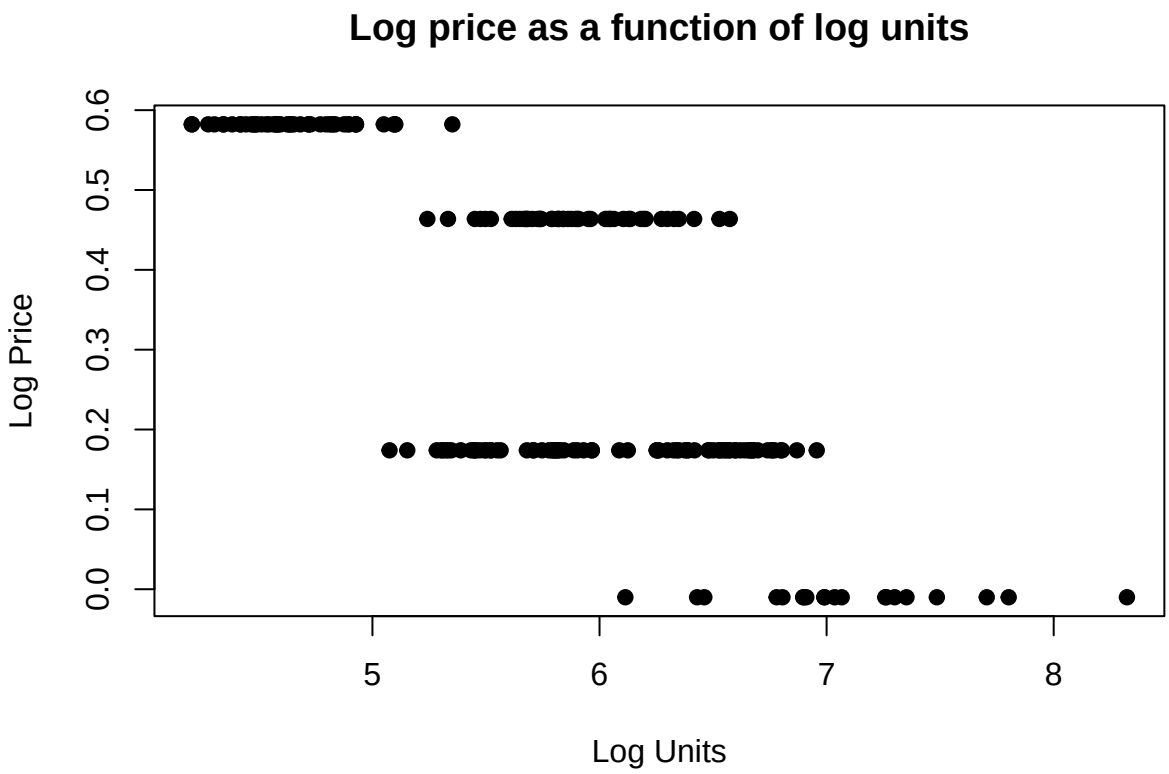
```
# Read in the data
df <- read_excel("small_salsa.xlsx")

# Create logged prices and quantities
df <- df %>%
  mutate(ln_p = log(price),
         ln_q = log(units),
         Dholiday = factor(holiday),
         Dstore = factor(store),
         Dzone= factor(zone))
```

1.2 Data Exploration

```
#plot price versus units
plot(df$price, df$units,main="Units as a function of price",xlab="Price",ylab="Units",pc
```


```
plot(df$ln_q,df$ln_p,main="Log price as a function of log units",xlab="Log Units",ylab="
```



1.3 Correlation

```
corr.test(df %>% select(price, pop, units, holiday))
```

```
Call:corr.test(x = df %>% select(price, pop, units, holiday))
Correlation matrix
        price   pop units holiday
price    1.00 -0.17 -0.61   -0.50
pop     -0.17  1.00  0.45    0.00
units   -0.61  0.45  1.00    0.67
holiday -0.50  0.00  0.67    1.00
Sample Size
[1] 200
Probability values (Entries above the diagonal are adjusted for multiple tests.)
        price  pop units holiday
price    0.00 0.03     0       0
pop      0.02 0.00     0       1
units    0.00 0.00     0       0
holiday  0.00 1.00     0       0
```

```
 To see confidence intervals of the correlations, print with the short=FALSE option
```

1.4 Regression

```
reg1 <- lm(ln_q ~ ln_p + Dzone + Dholiday, data=df)
summary(reg1)
```

```

Call:
lm(formula = ln_q ~ ln_p + Dzone + Dholiday, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-0.63263 -0.19428 -0.02146  0.18465  0.76698

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.96265    0.04703 148.058  < 2e-16 ***
ln_p        -2.34677    0.11293 -20.780  < 2e-16 ***
Dzone2      -0.92446    0.03835 -24.108  < 2e-16 ***
Dholiday1    0.56942    0.07604   7.489 2.32e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.267 on 196 degrees of freedom
Multiple R-squared:  0.8998,    Adjusted R-squared:  0.8983
F-statistic: 586.9 on 3 and 196 DF,  p-value: < 2.2e-16
```

The regression indicates that the price elasticity is -2.347.

1.5 Regression Calculation for Zone 1

```
reg2 <- lm(ln_q ~ ln_p + Dholiday, data=df %>% filter(Dzone==1))
summary(reg2)
```

```

Call:
lm(formula = ln_q ~ ln_p + Dholiday, data = df %>% filter(Dzone ==
    1))

Residuals:
```

```
     Min       1Q   Median       3Q      Max
-0.65367 -0.18896 -0.01899  0.15606  0.87123

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.94964    0.07245  95.926  < 2e-16 ***
ln_p        -2.27334    0.19871 -11.440  < 2e-16 ***
Dholiday1    0.47892    0.11352   4.219 5.53e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2715 on 97 degrees of freedom
Multiple R-squared:  0.7706,    Adjusted R-squared:  0.7658
F-statistic: 162.9 on 2 and 97 DF,  p-value: < 2.2e-16
```

The regression indicates that the price elasticity for Zone 1 is -2.273.

1.6 Regression Calculation for Zone 2

```
reg3 <- lm(ln_q ~ ln_p + Dholiday, data=df %>% filter(Dzone==2))
summary(reg3)
```

```
Call:
lm(formula = ln_q ~ ln_p + Dholiday, data = df %>% filter(Dzone ==
    2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.62176 -0.17960 -0.01737  0.18030  0.69492

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.03401    0.06075  99.325  < 2e-16 ***
ln_p        -2.36523    0.13525 -17.488  < 2e-16 ***
Dholiday1    0.67766    0.10303   6.577 2.43e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2603 on 97 degrees of freedom
Multiple R-squared:  0.8731,    Adjusted R-squared:  0.8705
F-statistic: 333.6 on 2 and 97 DF,  p-value: < 2.2e-16
```

The regression indicates that the price elasticity for Zone 2 is -2.365.

1.7 Population Regression Calculation for Zone 1

```
reg2 <- lm(ln_q ~ ln_p + pop, data=df %>% filter(Dzone==1))
summary(reg2)
```

```
Call:
lm(formula = ln_q ~ ln_p + pop, data = df %>% filter(Dzone ==
    1))

Residuals:
    Min      1Q  Median      3Q     Max
-0.6605 -0.1797 -0.0380  0.1585  1.0998

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.5930     0.8907   6.279 9.61e-09 ***
ln_p         -2.7796     0.1696 -16.385  < 2e-16 ***
pop           0.2224     0.1270   1.751    0.083 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2907 on 97 degrees of freedom
Multiple R-squared:  0.7368,    Adjusted R-squared:  0.7314
F-statistic: 135.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

The regression indicates that the price elasticity for Zone 1 is -2.78.

2 Managerial Discussion

Managerial discussion goes here.