ISyE 6414 - HW 3

Manvitha Kalicheti
gtID: 903838438

September 13, 2022

1

From the scatter plot figure below, it does seem like a simple linear regression model is appropriate here.

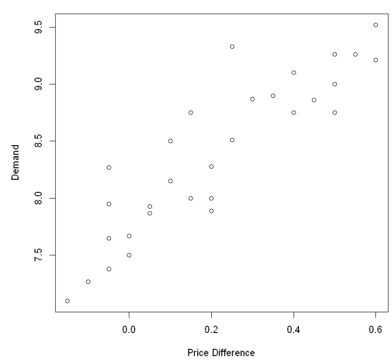


Figure 1: Scatter Plot (Price Difference vs. Demand)

2

I used R to create a linear regression model on the given data. The regression equation is

$$Demand = 2.665 * PriceDif + 7.814$$

From the summary of the model:

$$\hat{\beta}_0 = 7.814$$
$$\hat{\beta}_1 = 2.665$$
$$\hat{\sigma} = 0.3166$$
$$se(\hat{\beta}_0) = 0.0799$$
$$se(\hat{\beta}_1) = 0.2585$$

```
Call:
lm(formula = Demand ~ PriceDif)

Residuals:
     Min       1Q   Median       3Q      Max
-0.45713 -0.21121 -0.04898  0.14314  0.84961

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.81409    0.07988   97.82  < 2e-16 ***
PriceDif     2.66521    0.25850   10.31 4.88e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3166 on 28 degrees of freedom
Multiple R-squared:  0.7915,    Adjusted R-squared:  0.7841
F-statistic: 106.3 on 1 and 28 DF,  p-value: 4.881e-11
```

Figure 2: Linear Model Summary from R

3

$\alpha = 0.05$
$n - 2 = 28$
From the tables, $t_{0.025,28} = 2.048$

$$
\begin{aligned}
\text{Confidence interval for } \beta_1 &= \left(\hat{\beta}_1 - \left(t_{\frac{\alpha}{2},n-2}\sqrt{\frac{MSE}{S_{xx}}}\right), \hat{\beta}_1 + \left(t_{\frac{\alpha}{2},n-2}\sqrt{\frac{MSE}{S_{xx}}}\right)\right) \\
&= (\hat{\beta}_1 - t_{\frac{\alpha}{2},n-2}se(\hat{\beta}_1), \hat{\beta}_1 + t_{\frac{\alpha}{2},n-2}se(\hat{\beta}_1)) \\
&= (2.665 - 2.048 * 0.2585, 2.665 + 2.048 * 0.2585) \\
&= \boxed{(2.136, 3.195)}
\end{aligned}
$$

This means that we can state with 95% confidence, (or probability of 0.95) that the acual value of β_1 will lie within the interval $(2.136, 3.195)$.

4

Let's say our null hypothesis is $H_0 : \beta_1 = 0$ and our alternative hypothesis is $H_a : \beta_1 \neq 0$.

$$
\begin{aligned}
t - value &= \frac{\hat{\beta}_1 - 0}{\sqrt{\hat{\sigma}^2/S_{xx}}} \\
&= \frac{\hat{\beta}_1\sqrt{S_{xx}}}{\hat{\sigma}} \\
&= \frac{2.665 * \sqrt{1.49967}}{0.3166} \\
&= 10.30822
\end{aligned}
$$

From the tables, the critical t value $= t_{\frac{alpha}{2},n-2} = t_{0.025,28} = 2.048$.

$|t| = 10.308$ which is very large compared to the critical t value. Thus, we can reject the null hypothesis. This means that β_1 is statistically significant.

By setting β_1 value (or slope of x) to 0, we are saying that the value of x does not play a role in determining the value of y. We were proven wrong by the hypothesis test, and hence, we now know that the predictor x is statistically significant.

5

From the model summary, the p value ($Pr(>|t|)$) of the intercept is $< 2e-16$, which is very low. This implies that the intercept is statistically significant.

6

From the model summary, the p value ($Pr(>|t|)$) of β_1 is $4.88e-11$, which is very low. Our p value is much smaller than all 3 values of $alpha = 0.10, 0.05, 0.005$. Hence, H_0 can be rejected. This suggests that y and x are strongly related, and we cannot disregard x's role while trying to estimate y.

7

Using R, we get that the point estimate is 8.0806 and the 95% confidence interval is $[7.9479, 8.2133]$.

8

Using R, we get that the point prediction is 8.0806 and the 95% prediction interval is $[7.4187, 8.7425]$.

The confidence interval half length is 0.1327 and the prediction interval half length is 0.6619. Yes, the prediction interval is wider than the confidence interval as we see from their half lengths.

9

Using R, we get that the point estimate is 8.4803 and the 95% confidence interval is $[8.3604, 8.6003]$. The half length of this interval is 0.1199.

No, the half length of this interval is not too similar to the half length of the confidence interval for $PriceDif = 0.10$. The half length of confidence interval for $PriceDif = 0.25$ is smaller than the half length of the confidence interval for $PriceDif = 0.10$ as 0.25 is closer to the mean of $PriceDif$ (0.2133). As we move away from the mean, the interval becomes wider.

R codes attached at the end

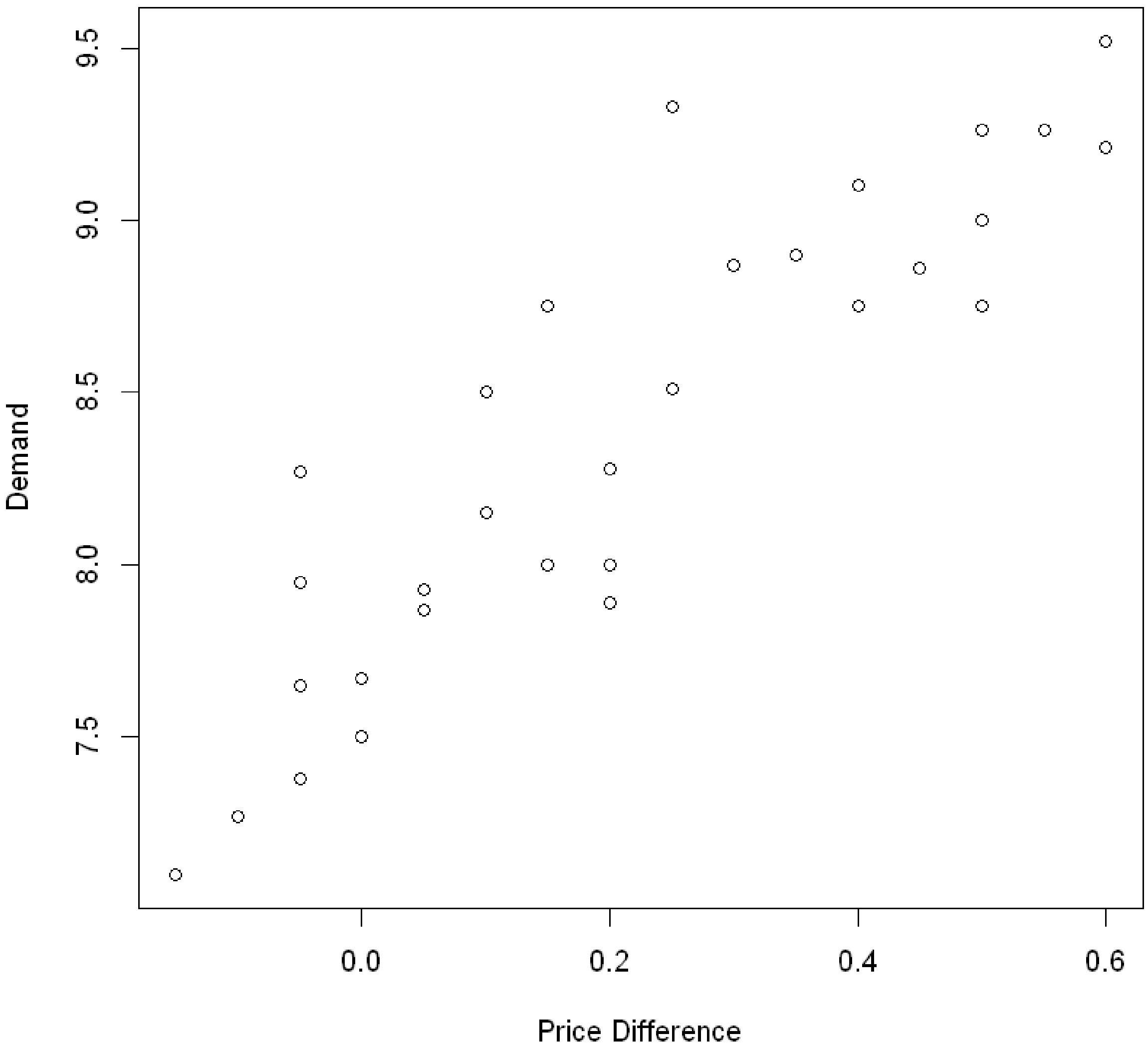
10

No-intercept model: $y = \beta_1 x + \epsilon$.
To minimise $SSE = \sum_{i=1} n(y_i - \hat{y_i})^2$, we have to set $\frac{\partial}{\partial \beta_1}(SSE)$ to 0.

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1}(SSE) &= 0 \\
\frac{\partial}{\partial \beta_1}\sum_{i=1} n(y_i - \hat{y}_i)^2 &= 0 \\
\frac{\partial}{\partial \beta_1}\sum_{i=1} n(y_i - \hat{\beta}_1 x)^2 &= 0 \\
\sum_{i=1}^{n} 2(y_i - \hat{\beta}_1 x_i)(-x_i) &= 0 \\
\sum_{i=1}^{n} (\hat{\beta}_1 x_i^2 - x_i y_i) &= 0 \\
\hat{\beta}_1 \sum_{i=1}^{n}(x_i^2) - \sum_{i=1}^{n}(x_i y_i) &= 0 \\
\boxed{\implies \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1} n x_i^2}}
\end{aligned}
$$

In [1]:

```
data = read.csv('C:/Users/manvi/Documents/GT Acads/Fall 2022/Regression/HW3/6414_HW3
Demand = data[,1]
PriceDif = data[,2]
plot(PriceDif, Demand, xlab = "Price Difference", ylab = "Demand")
```



In [2]:

```
model = lm(Demand ~ PriceDif)
summary(model)
```

```
Call:
lm(formula = Demand ~ PriceDif)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.45713 -0.21121 -0.04898  0.14314  0.84961 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.81409    0.07988   97.82  < 2e-16 ***
PriceDif     2.66521    0.25850   10.31 4.88e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3166 on 28 degrees of freedom
Multiple R-squared:  0.7915,    Adjusted R-squared:  0.7841
F-statistic: 106.3 on 1 and 28 DF,  p-value: 4.881e-11
```

In [20]:

```
new = data.frame(PriceDif = 0.10)
q7 = predict.lm(model, new, interval = "confidence", level = 0.95)
q7
```

fit	lwr	upr
8.080609	7.947878	8.21334

In [21]:

```
new = data.frame(PriceDif = 0.10)
q8 = predict.lm(model, new, interval = "predict", level = 0.95)
q8
```

fit	lwr	upr
8.080609	7.418719	8.7425

In [22]:

```
new = data.frame(PriceDif = 0.25)
q9 = predict.lm(model, new, interval = "confidence", level = 0.95)
q9
```

fit	lwr	upr
8.480391	8.36042	8.600362

In [49]:

```
q7hl = (q7[,'upr'] - q7[,'lwr']) /2
q7hl
```

0.132730642733149

In [50]:

```
q8hl = (q8[,'upr'] - q8[,'lwr']) /2
q8hl
```

0.661890519414294

In [51]:

```
q9hl = (q9[,'upr'] - q9[,'lwr']) /2
q9hl
```

0.119970889973496

In [48]:

```
mean(PriceDif)
```

0.213333333333333