ISyE 6414 - HW 2

Manvitha Kalicheti
gtID: 903838438

September 9, 2022

Part 1

1

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

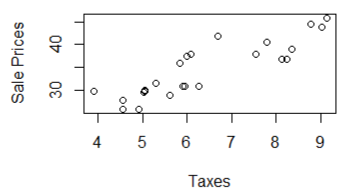

Figure 1: Scatter Plot (Taxes vs. Sale Prices)

2

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

$$SalePrice = 3.3244 * Tax + 13.3202 \tag{1}$$

```
Call:
lm(formula = sale_price ~ taxes)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8343 -2.3157 -0.3669  1.9787  6.3168 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  13.3202     2.5717   5.179 3.42e-05 ***
taxes         3.3244     0.3903   8.518 2.05e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.961 on 22 degrees of freedom
Multiple R-squared:  0.7673,	Adjusted R-squared:  0.7568 
F-statistic: 72.56 on 1 and 22 DF,  p-value: 2.051e-08
```

Figure 2: Linear Model Summary from R

3

β_1 is the slope. It tells us how much the mean value of y changes for a one-unit change in x. Here, we see that $\hat{\beta_1} = 3.3244$. This is a positive value, which means that as the tax increases, sale price also increases. More specifically, the sale price increases by \$33,244 for every \$10,000 increase in tax.

4

β_0 is the y intercept. It tells us the mean value of y when x is 0. Here, it says that the mean value of Sale Price when Tax is 0 is \$13,3202. It does not have a practical meaning here as tax cannot be 0.

5

$\hat{\sigma} = 2.961$

$\hat{\sigma}^2 = 8.768$

$df = n - 2 = 24 - 2 = 22$

$SSE = \hat{\sigma}^2 * df = 192.885$

R code for Part 1

```
data = read.csv(file = "6414-HW2-taxes.csv", sep=",", head = FALSE)
data = data[-1,]
sale_price = data[-1,1]
sale_price = as.numeric(sale_price)
taxes = data[-1,2]
taxes = as.numeric(taxes)
model = lm(sale_price ~ taxes)
summary(model)
plot(taxes, sale_price, xlab = "Taxes", ylab = "Sale Prices")
```

Part 2

6

Given :

$n = 14$

$\sum_{i=1}^{n} y_i = 572$

$\sum_{i=1}^{n} y_i^2 = 23530$

$\sum_{i=1}^{n} x_i = 43$

$\sum_{i=1}^{n} x_i^2 = 157.42$

$\sum_{i=1}^{n} x_i y_i = 1697.80$

We know that:

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{43}{14} = 3.0714$

and

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{572}{14} = 40.8571$

Also,

$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \boxed{\text{-2.3289}}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \boxed{48.0099}$

7

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = \sum_{i=1}^{n} yi^2 - n\bar{y}^2 - \hat{\beta}_1(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y})$

$= 23530 - 23370.2367 - (-2.3289) * (-59.0389) = 22.2676$

$\hat{\sigma} = \frac{SSE}{n-2} = \frac{22.2676}{12} = \boxed{1.8556}$

8

Euqation of the fitted line:

$\hat{y}_i = 48.0099 - 2.3289 x_i$

For $x = 3.7$, $\hat{y} = \boxed{39.3929}$

Given $y = 46.1$, residual $y - \hat{y} = \boxed{\text{r = 6.7071}}$