**Name: Nikhil Lohar R Answer Sheet**

**Regression Analysis in R**

1. Use R to estimate the following four models.

**(a) Model 1:**

**Code:** >lm(data=Airhomes1, Price~Sqft)

>summary(model1)

**Result:**

Call:

lm(formula = Price ~ Sqft, data = Airhomes1)

Residuals:

Min 1Q Median 3Q Max

-214533 -36424 2937 56662 183161

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 45522.27 44555.01 1.022 0.314

Sqft 310.16 22.59 13.731 1.98e-15 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 87160 on 34 degrees of freedom

Multiple R-squared: 0.8472, Adjusted R-squared: 0.8427

F-statistic: 188.5 on 1 and 34 DF, p-value: 1.976e-15

**(b) Model 2:**

**Code: >**model2=lm(data=Airhomes1, formula=Price~Sqft+Baths)

>summary(model2)

**Result:**

Call:

lm(formula = Price ~ Sqft + Baths, data = Airhomes1)

Residuals:

Min 1Q Median 3Q Max

-176783 - 53010 6633 48507 170737

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 50973.29 40417.61 1.261 0.2161

Sqft 212.69 39.36 5.404 5.6e-06 \*\*\*

Baths 81897.32 28245.86 2.899 0.0066 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 78980 on 33 degrees of freedom

Multiple R-squared: 0.8782, Adjusted R-squared: 0.8709

F-statistic: 119 on 2 and 33 DF, p-value: 8.142e-16

**(c) Model 3:**

**Code:** >model3=lm(data=Airhomes1, formula=Price~Beds+Baths)

>summary(model3)

**Result:**

Call:

lm(formula = Price ~ Beds + Baths, data = Airhomes1)

Residuals:

Min 1Q Median 3Q Max

-201475 -62169 -12723 58945 187568

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13115 64258 0.204 0.8395

Beds 73712 23590 3.125 0.0037 \*\*

Baths 172432 21827 7.900 4.14e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 95250 on 33 degrees of freedom

Multiple R-squared: 0.8229, Adjusted R-squared: 0.8122

F-statistic: 76.67 on 2 and 33 DF, p-value: 3.942e-13

**(d) Model 4:**

**Code:** >model3=lm(data=Airhomes1, formula=Price~Sqft+Beds+Baths)

>summary(model3)

**Result:**

Call:

lm(formula = Price ~ Sqft + Beds + Baths, data = Airhomes1)

Residuals:

Min 1Q Median 3Q Max

-162802 -53926 1450 51794 161169

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 29068.6 53939.8 0.539 0.593680

Sqft 193.9 49.9 3.887 0.000481 \*\*\*

Beds 15395.5 24798.6 0.621 0.539115

Baths 85066.1 28965.9 2.937 0.006100 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 79730 on 32 degrees of freedom

Multiple R-squared: 0.8797, Adjusted R-squared: 0.8684

F-statistic: 77.99 on 3 and 32 DF, p-value: 8.379e-15

2. (a) Fill in the table below with the Standard Error of the estimate and Adjusted given by SPSS for each model. (Attach the SPSS output to the answer sheet.) (the significance level is 5%) Each square is 5 points

a)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Joint Significance Test P-value | Individual Significance Test (Only list the significant variables | Standard Error | Adjusted |
| Model 1 | 1.976e-15 | Sqft | 87160 | 0.8427 |
| Model 2 | 8.142e-16 | Sqft, Baths | 78980 | 0.8709 |
| Model 3 | 3.942e-13 | Beds, Baths | 95250 | 0.8122 |
| Model 4 | 8.379e-15 | Sqft, Baths | 79730 | 0.8684 |

(b) Based on you answer in (a) determine the best model for predicting Price.

**Ans:** Model 2 is better for predicting price. As data in above table shows that model 2 has highest Adjusted R2 Value which is 0.8709.