Madison Pribble

Dr. Poplin

Report 2

Monthly Apartment Rental Cost Based on Size

An agent for a real estate company in a small city wants to predict the monthly apartment rental cost, based on the size of the apartment in square footage. The agent selected 25 apartments in a specific residential neighborhood and gathered the data which is stored in the file Rent.csv (or as seen in the appendix). This report will discuss the regression of this data. (The output from R Studio is in the Appendix section).

1. First, we will look at a scatter plot of the data, along with the Least Squares Regression Line (LSRL).



 The scatterplot shows us that this is a strong, positive, linear relationship. In order to test the strength of the relationship we can look at the r value. The r value is .9035, which ensures that there is a strong relationship.

1. Next, we will look at the LSRL (Least Squares Regression Line) equation. The LSRL equation would be ŷ = 1.088133x + 141.945668.
2. The slope would be 1.088 dollars/square footage. For each additional square foot in size, we expect the price to increase by approximately 1.088 dollars on average. Furthermore, if we change the units to hundreds of dollars, for each additional 100 square feet, we expect the price to increase by approximately $108.
3. Next, we will look at $r^{2}$. Calculations is R tells us that R-squared= 0.8164. Approximately 81.64% of the variation in the rent is explained by changes in the size of the apartment.
4. Now we want to run a hypothesis test to answer the following question: “Is there a significant relationship between the variables?” We also want to check our conditions. First, the information regarding collection does not say whether it was a simple random sample, so use caution when analyzing the data. When looking at the scatterplot, we are able to ensure that the data is linear, so this condition is satisfied. The scatterplot of residuals shows random scatter, so this condition is also satisfied. However, the histogram of residuals seems to be right skewed so use caution when determining if this is a good model to use. (The scatter plot and histogram of residuals is in the Appendix and in Part 5).

β1= the slope of the regression line. Next, we will look at the null and alternative hypothesis that are being used to run this hypothesis test. The null hypothesis is Ho: β1=0 and the alternative hypothesis is Ha: β1≠ 0. After running the test we see the t-value is 10.113. The p-value would be 6.18e-10. Since the p-value is less than 0.05 we reject the null hypothesis. We have strong evidence that this is a linear model and that there is a relationship between the variables because β1 does not equal 0.

1. In order to determine how useful the regression model is, we will look at the residual scatterplot and the residual histogram.



 

 The scatterplot of residuals is random with uniform variation, which explains that this model is useful. However, the histogram appears to be right skewed which is a deviation from normality. I think this regression model is useful, however, use caution since the histogram deviates from normality.

1. Two variables that may contribute to the rent price that are not mentioned include the age of the apartment and what is included with the apartment (electricity, furnished, etc.). Other variables include the number of bedrooms, parking availability, yard size, and if pets are allowed.
2. Now, we will calculate a 95% confidence interval for the slope of the line. We are 95% confident that the slope of the line is between 0.8655612 and 1.310705. We are 95% confident that each additional square foot will have an increase of $0.865 to $1.31 in the price, on average. Or in other words, we are 95% confident that for each additional 100 square feet, the price will have an increase between $86.50 and $131 in the price, on average.
3. Next, we want to calculate a 95% confidence interval for the mean rent for apartments of 1100 square feet. We are 95% confidence that the mean rent for apartments of 1100 square feet is between $1,285.78 and $1,392 a month.
4. Now we will calculate a 95% confidence interval for the rent of a particular apartment of 1100 square feet. We are 95% confident that the rent of a particular apartment of 1100 square feet is between $1,069.56 and $1,608.22 a month.
5. My friends, Alice and Bob, are considering signing a lease for an apartment in this residential neighborhood. They are trying to decide between two apartments, one with 1,000 square feet for a monthly rent of $1,275 and the other with 1,200 square feet for a monthly rent of $1,425. We would first run a 95% confidence interval for an apartment of 1,200 square feet. We are 95% confident that the rent of a particular apartment of 1200 square feet is between $1,177.92 and $1,717.49. I think that the better option would be the apartment with 1,000 square feet with a monthly rent of $1,275 because we determined earlier that with each additional 100 square feet, we would expect the price to increase by approximately $108. However, this price increased by $150 for the extra 100 square feet. This is not a major increase than what we expected, but it is more than what was expected from the slope. The price for each apartment is within the price range determined by the confidence interval, so either choice would be in the expected range, however, I think the first option is a better option.

Appendix

Below is the data entered into R Studio.

rents = data.frame(
  Rent = c(950, 1600, 1200, 1500, 950, 1700,
           1600, 935, 875, 1150, 1400, 1650,
           1900, 1750, 1400, 1450, 1250, 1700,
           1200, 1150, 1450, 1500, 1200, 1050,
           1650),
  Size = c(850, 1450, 1085, 1232, 725, 1485,
           1225, 700, 700, 960, 1100, 1285,
           1575, 1375, 1175, 1225, 1200, 1260,
           1150, 900, 1360, 1040, 875, 1000, 1200) )

Below is the scatterplot of the data.



Histogram of Residuals



Scatterplot of Residuals



Output from R:

#load data

> r= read.csv("rent.csv", header=T)

>

> head(r)

 Rent Size

1 950 850

2 1600 1450

3 1200 1085

4 1500 1232

5 950 725

6 1700 1485

>

> #plot data/examine data

> plot(Rent~Size, data=r, main="Rent vs. Size")

>

> fit=lm(Rent~Size, data=r)

> abline(fit)

>

> coefficients(fit)

(Intercept) Size

 141.945668 1.088133

> summary(fit)

Call:

lm(formula = Rent ~ Size, data = r)

Residuals:

 Min 1Q Median 3Q Max

-197.71 -116.86 17.47 105.94 226.40

Coefficients:

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

(Intercept) 141.9457 123.7337 1.147 0.263

Size 1.0881 0.1076 10.113 6.18e-10 \*\*\*

---

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

Residual standard error: 127.6 on 23 degrees of freedom

Multiple R-squared: 0.8164, Adjusted R-squared: 0.8084

F-statistic: 102.3 on 1 and 23 DF, p-value: 6.176e-10

>

> #examine residuals

> plot(fit$residuals~r$Size, main="Scatterplot of Residuals")

> abline(0,0)

> hist(fit$residuals, main="Residuals")

>

> #confidence interval for the slope

> confint(fit, level = 0.95)

 2.5 % 97.5 %

(Intercept) -114.0169108 397.908248

Size 0.8655612 1.310705

> #confidence interval for mean response

> predict(fit, data.frame(Size=1100), interval="confidence")

 fit lwr upr

1 1338.892 1285.784 1392

> #confidence interval for mean response

> predict(fit, data.frame(Size=1100), interval="prediction")

 fit lwr upr

1 1338.892 1069.559 1608.225

> predict(fit, data.frame(Size=1200), interval="prediction")

 fit lwr upr

1 1447.705 1177.918 1717.493