Syed Thahirr Shaa Shahid Syed Yacoob

Minitab 5

Correlation and Regression

**OBJECTIVE:**

1) Create Scatter Plots in Minitab  
2) Calculate Correlation in Minitab  
3) Create Fitted Line Plots and Residual Plots  
4) Perform Regression Analysis in Minitab

**INSTRUCTIONS:**

1) Refer to the Homes.mtw Minitab worksheet file. This data set contains real estate data on 75 randomly selected homes.

2) Calculate the Correlation Coefficient for Price and Area and Append to your Report. Create a Scatter plot of Price vs. Area. Be sure to put Price on the Y-axis as it is the Response Variable; also be sure that your graph is well labeled. Append to Report.

3) Create a Fitted Line Plot of Price vs. Area. Append to Report.

4) Predict the Price of a house that is 1800 sq. ft. Append Regression Analysis to Report.

5) Create a Residual Plot for the residuals vs. the explanatory variable Area. Append only the graph to your Report.

**6) Answer the following:**

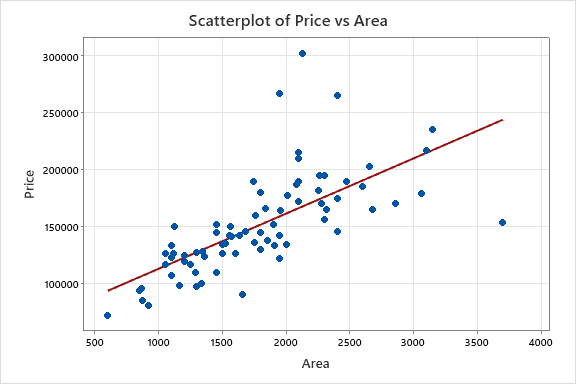
a. What is the association between Price vs. Area?  
b. How does Price correlate as the Area of the House increase?  
c. What is the predicted price of a house with 1800 sq. ft. of Area?  
d. What is the most extreme point that you have in the X direction (Area)?  
e. What is the point with the highest residual? Estimate both the Area and the Residual value from your graph.

1) Refer to the Homes.mtw Minitab worksheet file. This data set contains real estate data on 75 randomly selected homes.

**Data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Row** | **Price** | **Area** | **Acres** | **Rooms** | **Baths** |
| 1 | 123500 | 1362 | 0.4000 | 7 | 2.0 |
| 2 | 190000 | 2100 | 1.3000 | 8 | 1.5 |
| 3 | 96000 | 864 | 0.3200 | 4 | 1.0 |
| 4 | 126500 | 1500 | 0.5000 | 7 | 1.5 |
| 5 | 135500 | 1526 | 0.3000 | 7 | 1.5 |
| 6 | 134000 | 1502 | 0.3500 | 7 | 1.5 |
| 7 | 180000 | 1800 | 1.5200 | 8 | 2.5 |
| 8 | 152000 | 1450 | 0.3000 | 6 | 1.5 |
| 9 | 128500 | 1344 | 0.9360 | 6 | 2.0 |
| 10 | 107000 | 1100 | 0.1700 | 5 | 1.0 |
| 11 | 122000 | 1950 | 0.5000 | 7 | 1.5 |
| 12 | 302000 | 2130 | 11.9100 | 8 | 1.5 |
| 13 | 133000 | 1908 | 0.4600 | 7 | 2.0 |
| 14 | 150000 | 1122 | 3.0900 | 5 | 2.0 |
| 15 | 94000 | 850 | 0.1100 | 4 | 1.0 |
| 16 | 141900 | 1950 | 0.7500 | 8 | 2.5 |
| 17 | 125000 | 1200 | 0.3300 | 5 | 1.0 |
| 18 | 175000 | 2400 | 0.7000 | 8 | 3.0 |
| 19 | 267000 | 1950 | 18.7000 | 7 | 2.5 |
| 20 | 187000 | 2080 | 1.2300 | 8 | 2.5 |
| 21 | 215000 | 2100 | 0.5000 | 8 | 2.5 |
| 22 | 185000 | 2600 | 0.7500 | 8 | 2.0 |
| 23 | 165000 | 2680 | 0.5000 | 9 | 3.0 |
| 24 | 156000 | 2300 | 0.6500 | 7 | 3.0 |
| 25 | 122500 | 1100 | 0.3700 | 7 | 1.0 |
| 26 | 117000 | 1248 | 0.3000 | 6 | 1.0 |
| 27 | 136000 | 1750 | 0.5000 | 7 | 2.0 |
| 28 | 165000 | 2320 | 0.4000 | 8 | 2.5 |
| 29 | 141000 | 1575 | 0.2500 | 7 | 1.5 |
| 30 | 80500 | 922 | 0.3000 | 5 | 1.0 |
| 31 | 203000 | 2653 | 1.8000 | 9 | 3.0 |
| 32 | 110000 | 1289 | 0.2500 | 6 | 1.0 |
| 33 | 164000 | 1956 | 0.5000 | 8 | 2.5 |
| 34 | 194500 | 2300 | 0.9100 | 8 | 2.5 |
| 35 | 141900 | 1632 | 3.0000 | 6 | 3.0 |
| 36 | 169900 | 2858 | 0.7900 | 9 | 3.0 |
| 37 | 137900 | 1856 | 0.3300 | 7 | 1.5 |
| 38 | 98000 | 1165 | 0.1200 | 6 | 1.0 |
| 39 | 190000 | 2473 | 1.2500 | 9 | 2.5 |
| 40 | 100000 | 1338 | 0.1200 | 6 | 1.0 |
| 41 | 145000 | 1800 | 0.6580 | 8 | 2.5 |
| 42 | 170000 | 2277 | 0.8000 | 8 | 3.0 |
| 43 | 153500 | 3700 | 1.1000 | 10 | 3.0 |
| 44 | 210000 | 2100 | 0.5000 | 8 | 2.5 |
| 45 | 172000 | 2100 | 1.0000 | 8 | 2.5 |
| 46 | 235000 | 3150 | 0.3000 | 11 | 4.0 |
| 47 | 119000 | 1200 | 0.2500 | 7 | 1.0 |
| 48 | 116300 | 1050 | 0.4300 | 5 | 1.5 |
| 49 | 146000 | 2400 | 0.4000 | 7 | 2.5 |
| 50 | 181500 | 2250 | 0.3300 | 9 | 2.5 |
| 51 | 142500 | 1552 | 0.4600 | 6 | 1.5 |
| 52 | 72000 | 600 | 0.5000 | 3 | 1.0 |
| 53 | 160000 | 1760 | 0.0500 | 7 | 2.0 |
| 54 | 179000 | 3060 | 0.7500 | 8 | 2.0 |
| 55 | 150000 | 1564 | 0.3328 | 6 | 2.0 |
| 56 | 89900 | 1660 | 0.2100 | 7 | 1.0 |
| 57 | 217000 | 3100 | 0.5400 | 10 | 3.5 |
| 58 | 110000 | 1450 | 0.3000 | 6 | 2.0 |
| 59 | 125900 | 1118 | 0.5600 | 7 | 1.5 |
| 60 | 195000 | 2265 | 0.8500 | 8 | 2.5 |
| 61 | 85000 | 875 | 0.2600 | 5 | 1.0 |
| 62 | 265000 | 2400 | 2.0000 | 7 | 2.0 |
| 63 | 126500 | 1050 | 1.0000 | 5 | 1.0 |
| 64 | 190000 | 1745 | 0.5800 | 7 | 2.5 |
| 65 | 177000 | 2010 | 0.6800 | 8 | 1.5 |
| 66 | 126500 | 1600 | 0.2600 | 8 | 1.5 |
| 67 | 134500 | 2000 | 0.7000 | 8 | 1.0 |
| 68 | 144900 | 1450 | 0.3000 | 7 | 1.0 |
| 69 | 97000 | 1300 | 0.3700 | 5 | 1.0 |
| 70 | 133000 | 1100 | 0.3300 | 6 | 1.0 |
| 71 | 145900 | 1680 | 0.5000 | 6 | 1.5 |
| 72 | 130000 | 1800 | 0.3000 | 7 | 1.5 |
| 73 | 127500 | 1296 | 0.5000 | 9 | 1.0 |
| 74 | 165900 | 1840 | 1.1620 | 8 | 2.0 |
| 75 | 151500 | 1900 | 0.7500 | 7 | 2.0 |

2) Calculate the Correlation Coefficient for Price and Area and Append it to your Report. Create a Scatter plot of Price vs. Area. Be sure to put Price on the Y-axis as it is the Response Variable; also, be sure that your graph is well labeled. Append to Report.

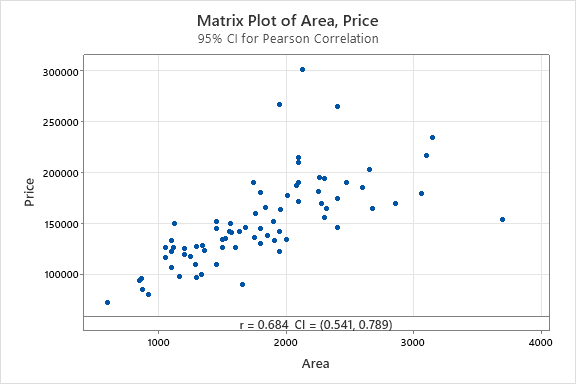


**Model Summary**

|  |  |  |
| --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** |
| 31891.6 | 46.80% | 46.07% |

WORKSHEET 1

**Correlation: Area, Price**



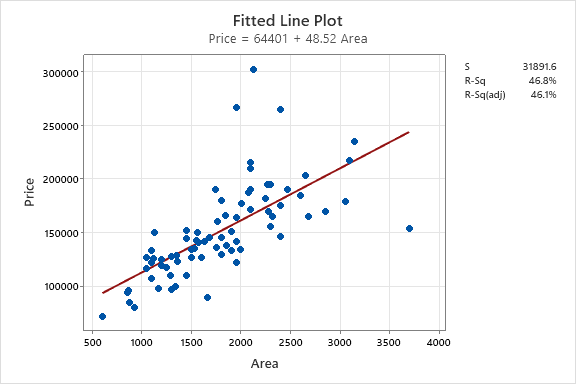
**Method**

|  |  |
| --- | --- |
| Correlation type | Pearson |
| Number of rows used | 75 |

**Correlations**

|  |  |
| --- | --- |
|  | **Area** |
| Price | 0.684 |

3) Create a Fitted Line Plot of Price vs. Area. Append to Report.



4) Predict the Price of a house that is 1800 sq. ft. Append Regression Analysis to Report.

WORKSHEET 1

**Prediction for Price**

**Regression Equation**

|  |  |  |
| --- | --- | --- |
| Price | = | 64401 + 48.52 Area |

**Settings**

|  |  |
| --- | --- |
| **Variable** | **Setting** |
| Area | 1800 |

**Prediction**

|  |  |  |  |
| --- | --- | --- | --- |
| **Fit** | **SE Fit** | **95% CI** | **95% PI** |
| 151744 | 3682.84 | (144404, 159084) | (87761.3, 215726) |

WORKSHEET 1

**Regression Analysis: Price versus Area**

**Regression Equation**

|  |  |  |
| --- | --- | --- |
| Price | = | 64401 + 48.52 Area |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | 64401 | 11460 | 5.62 | 0.000 |  |
| Area | 48.52 | 6.06 | 8.01 | 0.000 | 1.00 |

**Model Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** | **R-sq(pred)** |
| 31891.6 | 46.80% | 46.07% | 42.56% |

**Analysis of Variance**

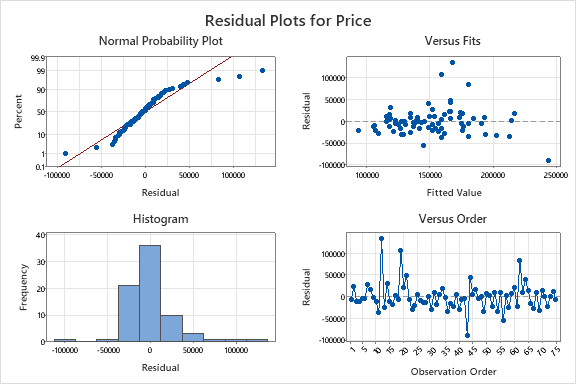
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Regression | 1 | 65304546178 | 65304546178 | 64.21 | 0.000 |
| Area | 1 | 65304546178 | 65304546178 | 64.21 | 0.000 |
| Error | 73 | 74246535688 | 1017075831 |  |  |
| Lack-of-Fit | 57 | 49541327355 | 869146094 | 0.56 | 0.942 |
| Pure Error | 16 | 24705208333 | 1544075521 |  |  |
| Total | 74 | 1.39551E+11 |  |  |  |

**Fits and Diagnostics for Unusual Observations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Price** | **Fit** | **Resid** | **Std Resid** |  |  |
| 12 | 302000 | 167756 | 134244 | 4.25 | R |  |
| 19 | 267000 | 159022 | 107978 | 3.41 | R |  |
| 43 | 153500 | 243938 | -90438 | -3.07 | R | X |
| 62 | 265000 | 180858 | 84142 | 2.67 | R |  |

*R  Large residual  
X  Unusual X*

5) Create a Residual Plot for the residuals vs. the explanatory variable Area. Append only the graph to your Report.



**6) Answer the following:**

a. What is the association between Price vs. Area?

Direct Relationship between Price and Area.   
b. How does Price correlate as the Area of the House increase?

Since it is a direct relationship, as Area increases the Prices also Increase.   
c. What is the predicted price of a house with 1800 sq. ft. of Area?

The Regression Equation is : 64401 + 48.52(Area)

Predicted Price (1800) = 64401 + 48.52(1800)

Predicted Price = 151737  
d. What is the most extreme point that you have in the X direction (Area)?

Row 43 – Price = 153500 Area = 3700  
e. What is the point with the highest residual? Estimate both the Area and the Residual value from your graph.

Row 12 – Price = 302000 Area = 2130