# Microsoft Excel
# Prediction Models & Sensitivity Analysis

**Modeling Background (Correlation & Regression)**

Identifying data relationships is key to modeling behavior of customer, student, and corporate data. First, let's consider two variables and the relationships between them. When comparing two data variables, you can have:

1. No relationship between the variables
2. A positive relationship (when one variable goes up, the other goes up)
3. A negative relationship (when one variable goes up, the other goes down)
4. A curvilinear relationship (a non-linear relationship

Examples of these, from The Research Methods Knowledge Base by Trochim & Donnelly (2007):

**FIGURE 1–1a** No relationship

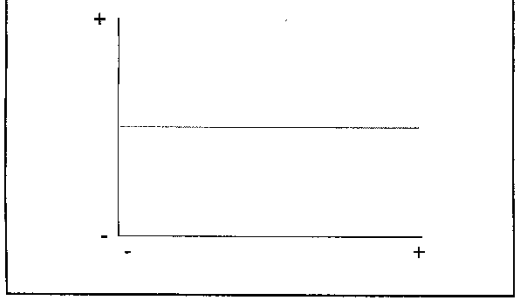



**FIGURE 1–1b** A positive relationship

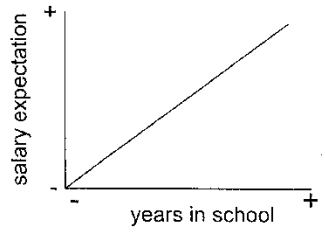



**FIGURE 1–1c** A negative relationship

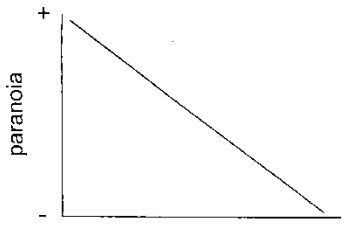
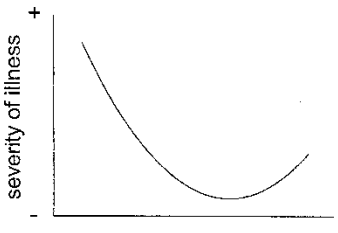
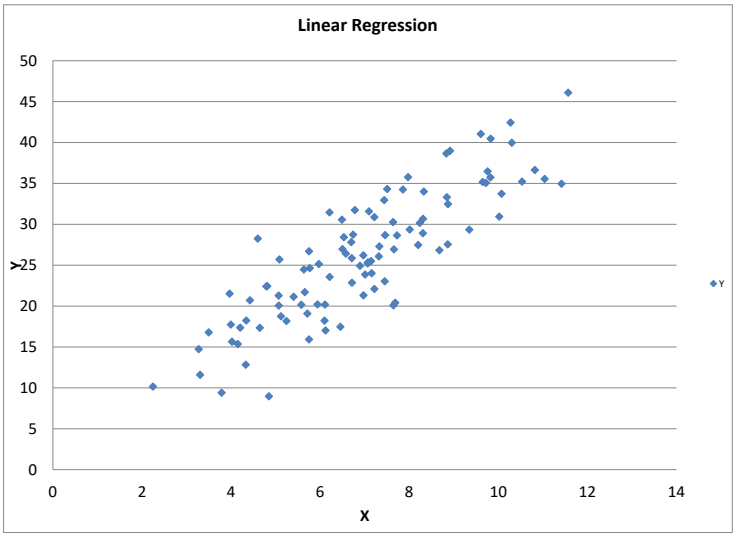



**FIGURE 1–1d** A curvilinear relationship

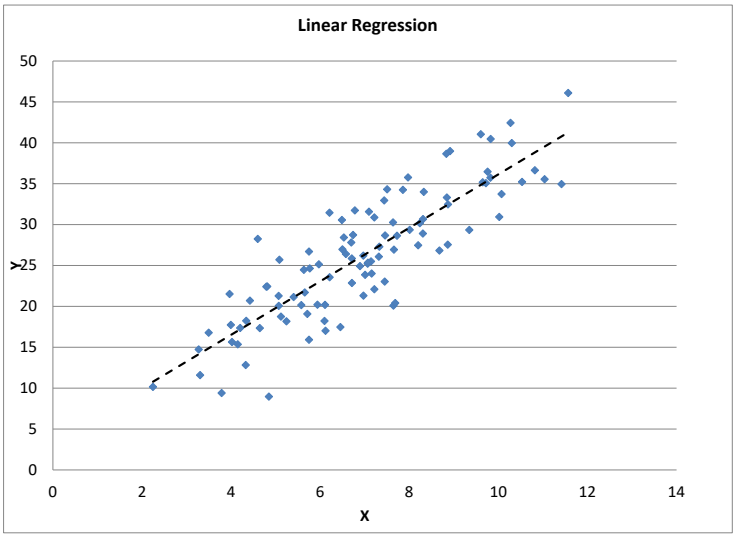


## Regression

Linear regression is a technique that calculates the relationship between a dependent variable Y and one or more independent variables, or X's. Assume that you have data similar to the picture below.




You can calculate a regression trend line based on the data. This dashed line represents $\hat{Y}$ which is the estimate of the Y equation.

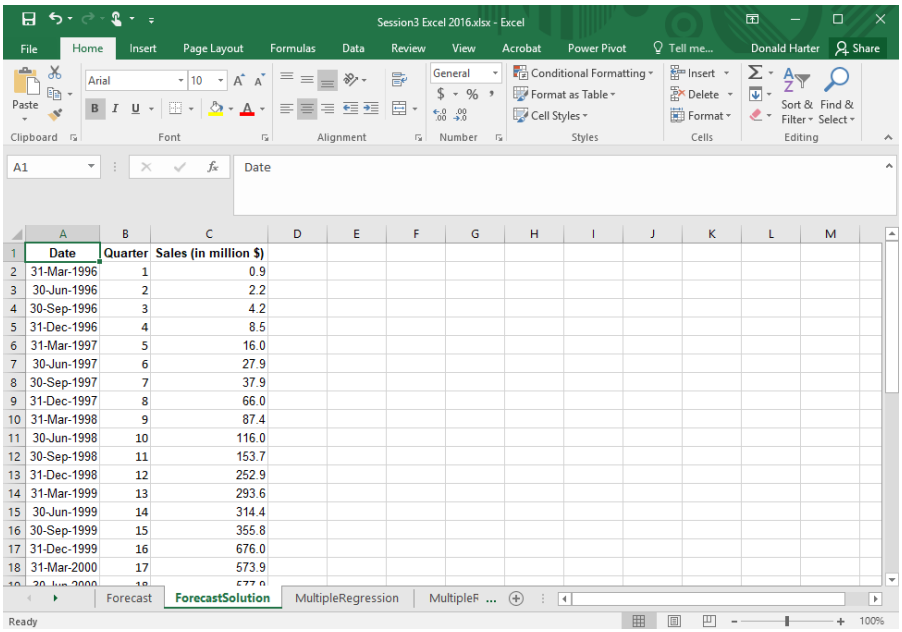



The vertical distance between the line and the data point is called the residual or error term.

## Forecasting

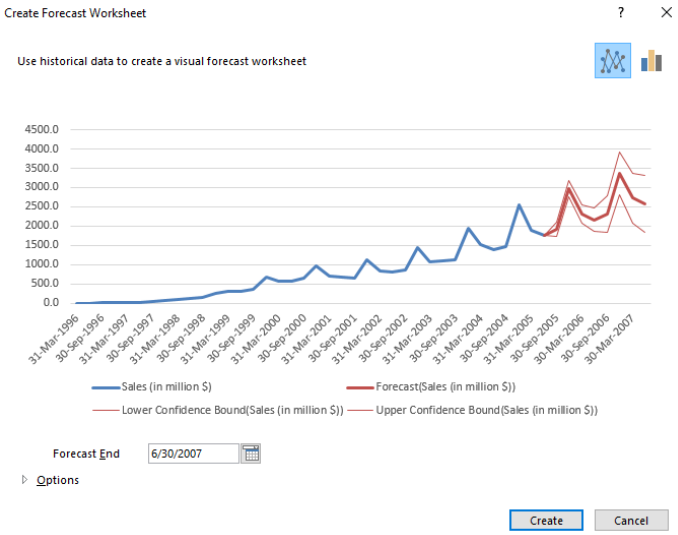
Regression models describe how changes in explanatory variables affect the outcome over the period of the data. If you want to forecast a trend into the future, then the new forecast feature initially offered in Excel 2016 is available. Forecasts use one date variable and one outcome variable. Note that forecasting outcomes over time can be risky because there is no guarantee that the trend will continue.

1. Use the Forecast spreadsheet for this exercise. The data below is the quarterly sales data for Amazon.com. Recall that the data displayed seasonality.

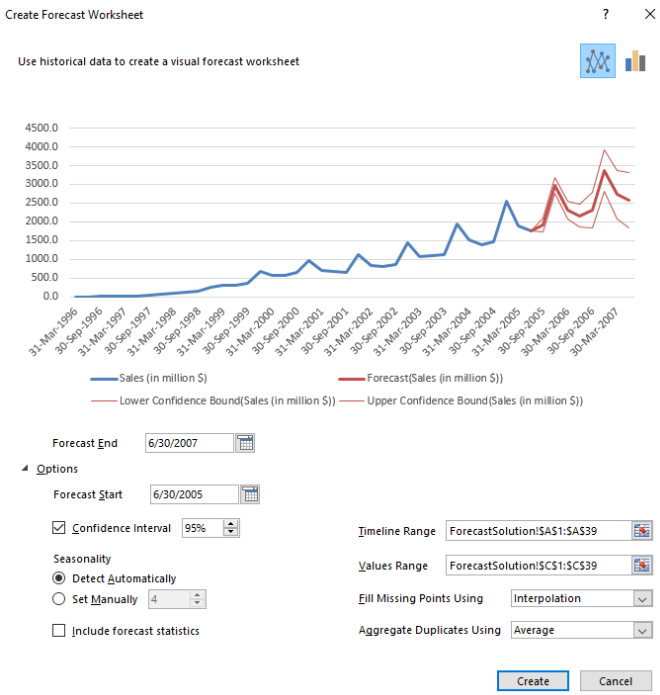


| | Date | Quarter | Sales (in million $) |
|---|---|---|---|
| 2 | 31-Mar-1996 | 1 | 0.9 |
| 3 | 30-Jun-1996 | 2 | 2.2 |
| 4 | 30-Sep-1996 | 3 | 4.2 |
| 5 | 31-Dec-1996 | 4 | 8.5 |
| 6 | 31-Mar-1997 | 5 | 16.0 |
| 7 | 30-Jun-1997 | 6 | 27.9 |
| 8 | 30-Sep-1997 | 7 | 37.9 |
| 9 | 31-Dec-1997 | 8 | 66.0 |
| 10 | 31-Mar-1998 | 9 | 87.4 |
| 11 | 30-Jun-1998 | 10 | 116.0 |
| 12 | 30-Sep-1998 | 11 | 153.7 |
| 13 | 31-Dec-1998 | 12 | 252.9 |
| 14 | 31-Mar-1999 | 13 | 293.6 |
| 15 | 30-Jun-1999 | 14 | 314.4 |
| 16 | 30-Sep-1999 | 15 | 355.8 |
| 17 | 31-Dec-1999 | 16 | 676.0 |
| 18 | 31-Mar-2000 | 17 | 573.9 |

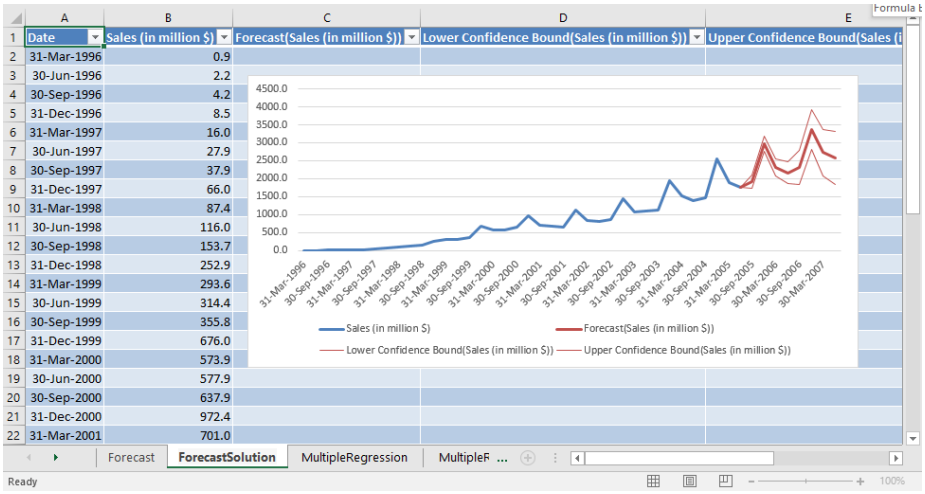
2. Click on the Data tab, then Forecast Sheet. In this example, Excel automatically identifies the relevant date and trend data.



3. Click on the Options drop down arrow in the lower left corner.

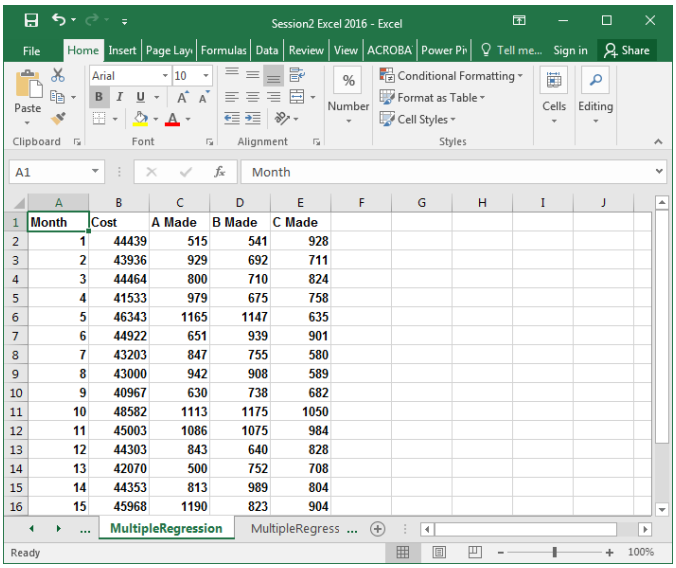



4. You can set the Forecast Start and Forecast End
5. Uncheck the box on Confidence Interval. This removes the confidence interval.
6. Change the Confidence Interval to 50%. What happens?
7. Excel Forecast is usually good at identifying Seasonality. If it has difficulty, you can click on Set Manually and set the seasonality parameter (4 for quarters, 12 for months, 52 for weeks, etc.)
8. If your data has missing data points, you can select Fill Missing Points Using: Interpolation or Zeros. Interpolation is usually better.
9. If there are duplicates in the data, set Aggregate Duplicates Using: Average.
10. In the upper right corner is the option for line versus bar chart. Click each.
11. Click on Create to generate the forecast. New columns with forecasted data are created.



**Multiple regression Review**

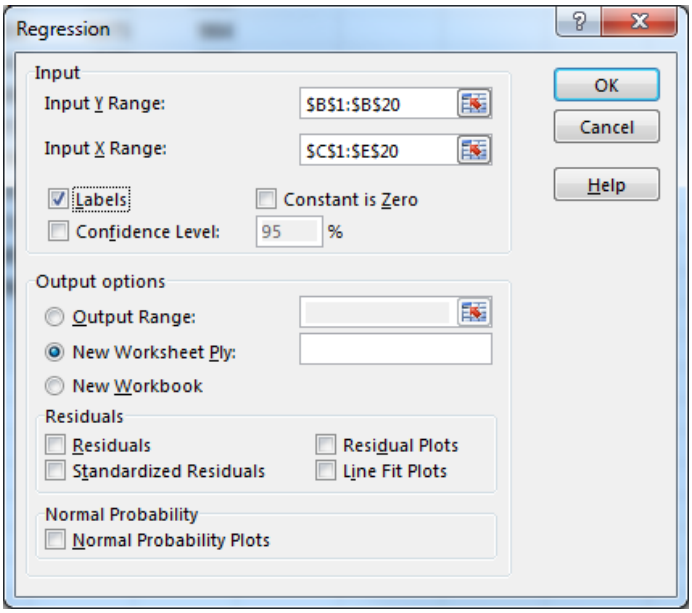
Multiple regression includes several independent variables. Use the Multiple Regression spreadsheet.

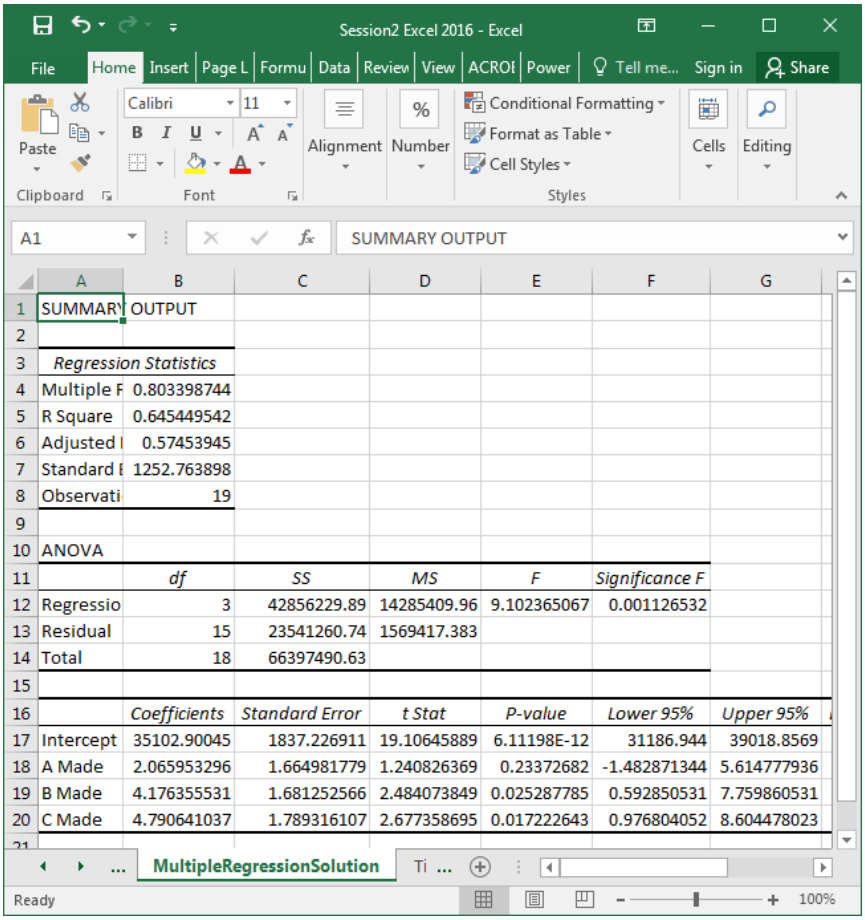



| Month | Cost | A Made | B Made | C Made |
|---|---|---|---|---|
| 1 | 44439 | 515 | 541 | 928 |
| 2 | 43936 | 929 | 692 | 711 |
| 3 | 44464 | 800 | 710 | 824 |
| 4 | 41533 | 979 | 675 | 758 |
| 5 | 46343 | 1165 | 1147 | 635 |
| 6 | 44922 | 651 | 939 | 901 |
| 7 | 43203 | 847 | 755 | 580 |
| 8 | 43000 | 942 | 908 | 589 |
| 9 | 40967 | 630 | 738 | 682 |
| 10 | 48582 | 1113 | 1175 | 1050 |
| 11 | 45003 | 1086 | 1075 | 984 |
| 12 | 44303 | 843 | 640 | 828 |
| 13 | 42070 | 500 | 752 | 708 |
| 14 | 44353 | 813 | 989 | 804 |
| 15 | 45968 | 1190 | 823 | 904 |

To run a multiple regression:

1. Click on the data tab, data analysis, regression, then OK.
2. For the Y-range, highlight the values in the B column for cost
3. For the X-range, highlight the values in the C, D, and E columns.
4. If you included the headings at the top of the columns, click labels.
5. Click OK.






| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | SUMMARY | OUTPUT | | | | | |
| 2 | | | | | | | |
| 3 | *Regression Statistics* | | | | | | |
| 4 | Multiple F | 0.803398744 | | | | | |
| 5 | R Square | 0.645449542 | | | | | |
| 6 | Adjusted I | 0.57453945 | | | | | |
| 7 | Standard E | 1252.763898 | | | | | |
| 8 | Observati | 19 | | | | | |
| 9 | | | | | | | |
| 10 | ANOVA | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Significance F* | |
| 12 | Regressio | 3 | 42856229.89 | 14285409.96 | 9.102365067 | 0.001126532 | |
| 13 | Residual | 15 | 23541260.74 | 1569417.383 | | | |
| 14 | Total | 18 | 66397490.63 | | | | |
| 15 | | | | | | | |
| 16 | | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| 17 | Intercept | 35102.90045 | 1837.226911 | 19.10645889 | 6.11198E-12 | 31186.944 | 39018.8569 |
| 18 | A Made | 2.065953296 | 1.664981779 | 1.240826369 | 0.23372682 | -1.482871344 | 5.614777936 |
| 19 | B Made | 4.176355531 | 1.681252566 | 2.484073849 | 0.025287785 | 0.592850531 | 7.759860531 |
| 20 | C Made | 4.790641037 | 1.789316107 | 2.677358695 | 0.017222643 | 0.976804052 | 8.604478023 |

**Prediction Models**

A prediction model allows you to enter values for each of the inputs (independent variables or X variables) and make a prediction of the outcome (Y variable or dependent variable).

The general form of the equation is:

$$Y = \beta_0 + \beta_1*X_1 + \beta_2*X_2 + \beta_3*X_3 + \ldots.$$

For the regression results above, we have:

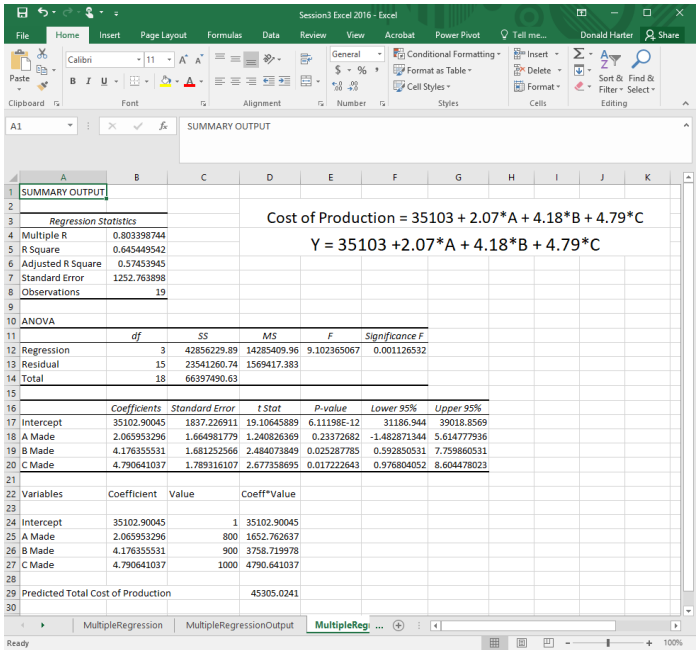
$$Y = 35103 + 2.07*X_1 + 4.18*X_2 + 4.79*X_3$$

Or:

Cost of Production = 35103 + 2.07*A + 4.18*B + 4.79*C

Let's now build the prediction model:

1. In cell A22, enter Variables
2. In cells A24:A27 enter Intercept, A Made, B Made, C Made
3. In cells B22, enter Coefficient
4. In cells B24:B27, copy the coefficients from the regression
5. In cell C22, enter Values
6. In cells C24:C27, enter 1, 800, 900, 1000. Note that the value for the intercept should be one; the values for the X variables must be in the range of the original data
7. In cell D22, enter Coeff*Value
8. In cell D24, enter the formula =B24*C24
9. Copy the formula from D24 to D25:D27
10. In cell A29, enter Predicted Total Cost of Production
11. In cell D29, enter the formula =sum(D24:D27)

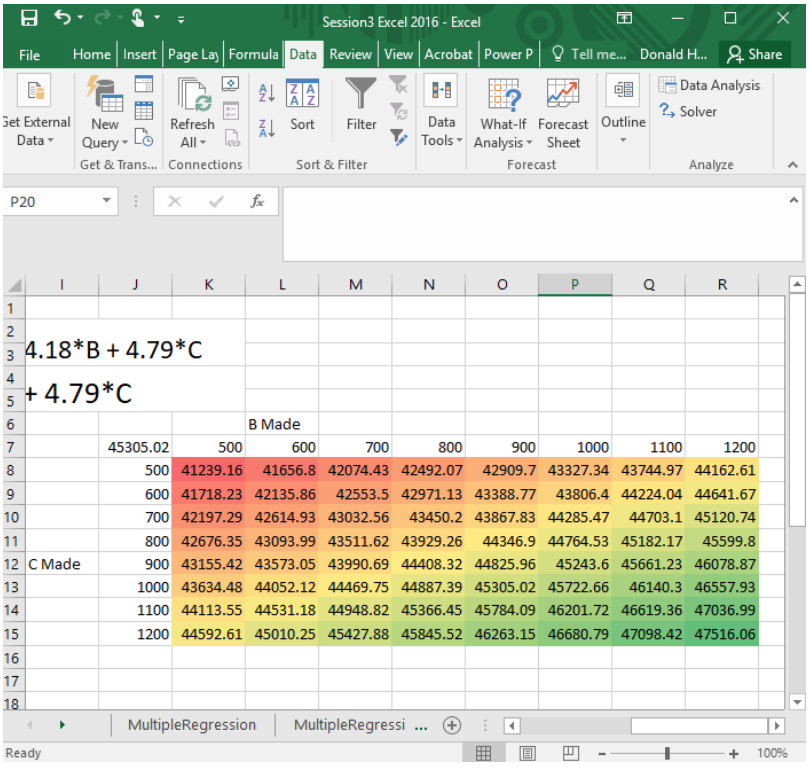



| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | SUMMARY OUTPUT | | | | | | |
| 2 | | | | | | | |
| 3 | *Regression Statistics* | | | | | | |
| 4 | Multiple R | 0.803398744 | | | | | |
| 5 | R Square | 0.645449542 | | | | | |
| 6 | Adjusted R Square | 0.57453945 | | | | | |
| 7 | Standard Error | 1252.763898 | | | | | |
| 8 | Observations | 19 | | | | | |
| 9 | | | | | | | |
| 10 | ANOVA | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Significance F* | |
| 12 | Regression | 3 | 42856229.89 | 14285409.96 | 9.102365067 | 0.001126532 | |
| 13 | Residual | 15 | 23541260.74 | 1569417.383 | | | |
| 14 | Total | 18 | 66397490.63 | | | | |
| 15 | | | | | | | |
| 16 | | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| 17 | Intercept | 35102.90045 | 1837.226911 | 19.10645889 | 6.11198E-12 | 31186.944 | 39018.8569 |
| 18 | A Made | 2.065953296 | 1.664981779 | 1.240826369 | 0.23372682 | -1.482871344 | 5.614777936 |
| 19 | B Made | 4.176355531 | 1.681252566 | 2.484073849 | 0.025287785 | 0.592850531 | 7.759860531 |
| 20 | C Made | 4.790641037 | 1.789316107 | 2.677358695 | 0.017222643 | 0.976804052 | 8.604478023 |
| 21 | | | | | | | |
| 22 | Variables | Coefficient | Value | Coeff*Value | | | |
| 23 | | | | | | | |
| 24 | Intercept | 35102.90045 | 1 | 35102.90045 | | | |
| 25 | A Made | 2.065953296 | 800 | 1652.762637 | | | |
| 26 | B Made | 4.176355531 | 900 | 3758.719978 | | | |
| 27 | C Made | 4.790641037 | 1000 | 4790.641037 | | | |
| 28 | | | | | | | |
| 29 | Predicted Total Cost of Production | | | 45305.0241 | | | |
| 30 | | | | | | | |

**Sensitivity Analysis of Regression Results**

After you build the prediction model, you can create a sensitivity analysis of the regression results.

1. In cell L6, enter B Made
2. In cells K7:R7, enter 500, 600, …, 1200
3. In cell I12, enter C Made
4. In cells J8:J15, enter 500, 600, …, 1200
5. In cell J7, enter a formula that points to Predicted Cost, =D29
6. Highlight the data, cells J7:R15
7. Click on the Data tab, What if analysis, Data table
8. For Row input cell, enter C26, the value for B Made
9. For Column input cell, enter C27, the value for C Made
10. Click OK
11. To add conditional formatting, highlight the production costs, cells K8:R15
12. Click on the Home tab
13. Click on Conditional Formatting, Color Scales, your choice of color




| | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | | | | B Made | | | | | | |
| 7 | | 45305.02 | 500 | 600 | 700 | 800 | 900 | 1000 | 1100 | 1200 |
| 8 | | 500 | 41239.16 | 41656.8 | 42074.43 | 42492.07 | 42909.7 | 43327.34 | 43744.97 | 44162.61 |
| 9 | | 600 | 41718.23 | 42135.86 | 42553.5 | 42971.13 | 43388.77 | 43806.4 | 44224.04 | 44641.67 |
| 10 | | 700 | 42197.29 | 42614.93 | 43032.56 | 43450.2 | 43867.83 | 44285.47 | 44703.1 | 45120.74 |
| 11 | | 800 | 42676.35 | 43093.99 | 43511.62 | 43929.26 | 44346.9 | 44764.53 | 45182.17 | 45599.8 |
| 12 | C Made | 900 | 43155.42 | 43573.05 | 43990.69 | 44408.32 | 44825.96 | 45243.6 | 45661.23 | 46078.87 |
| 13 | | 1000 | 43634.48 | 44052.12 | 44469.75 | 44887.39 | 45305.02 | 45722.66 | 46140.3 | 46557.93 |
| 14 | | 1100 | 44113.55 | 44531.18 | 44948.82 | 45366.45 | 45784.09 | 46201.72 | 46619.36 | 47036.99 |
| 15 | | 1200 | 44592.61 | 45010.25 | 45427.88 | 45845.52 | 46263.15 | 46680.79 | 47098.42 | 47516.06 |