Trendline and regression analysis using Excel

Performance Lawn Equipment Analysis

Prepared by

Bukola Williams Orire

April 8th, 2025

Table of Contents

1. Project Overview

In reviewing the data in the Performance Lawn Equipment Database, Elizabeth Burke noticed that defects received from suppliers have decreased (worksheet Defects After Delivery). Upon investigation, she learned that in 2014, PLE experienced some quality problems due to an increasing number of defects in materials received from suppliers. The company instituted an initiative in August 2015 to work with suppliers to reduce these defects, to more closely coordinate deliveries, and to improve materials quality through reengineering supplier production policies. Ms. Burke noted that the program appeared to reverse an increasing trend in defects; she would like to predict what might have happened had the supplier initiative not been implemented and how the number of defects might further be reduced in the near future.

Defects per million items received from suppliers

Month	2014	2015	2016	2017	2018
January	812	828	824	682	571
February	810	832	836	695	575
March	813	847	818	692	547
April	823	839	825	686	542
May	832	832	804	673	532
June	848	840	812	681	496
July	837	849	806	696	472
August	831	857	798	688	460
September	827	839	804	671	441
October	838	842	713	645	445
November	826	828	705	617	438
December	819	816	686	603	436

2. Problem Statement

The objective of this business analytics project is to perform trendline and regression analysis in excel to develop solution to the following key questions:

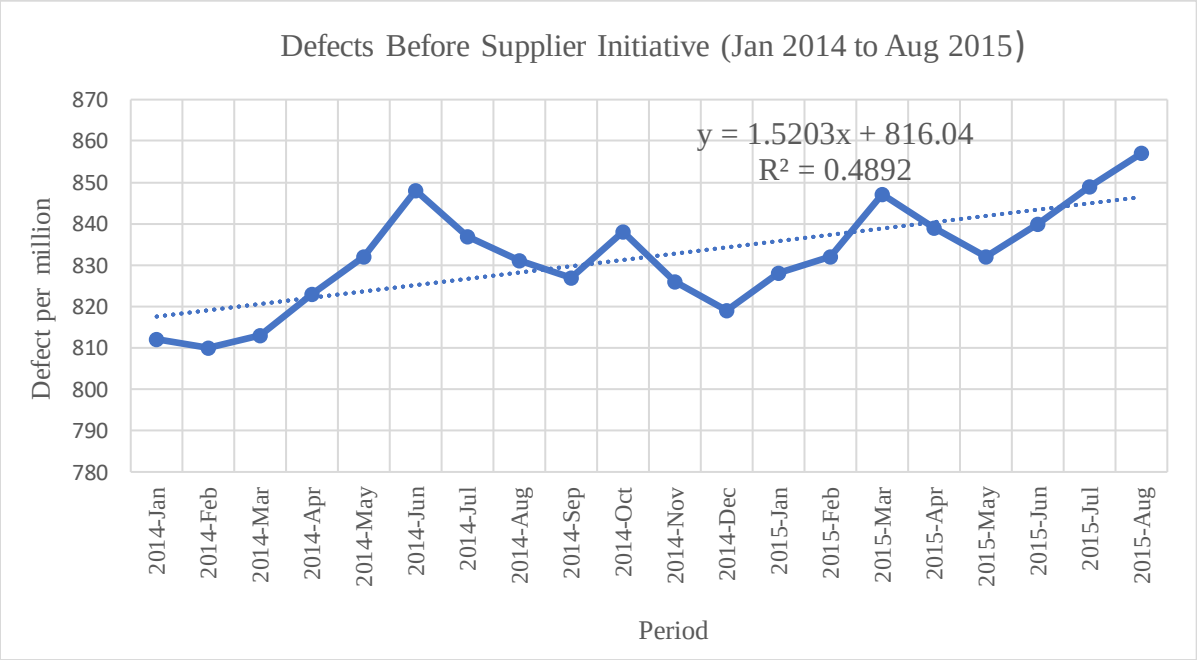
- What might have happened if PLE supplier did not take the initiative as of August 2015?
- How can PLE defects reduce even more in the future?

3. Trendlines and Regression Analysis

Firstly, Let's observe defects overtime. Since we are working on a time series data, it's best to plot a line chart to visualize the difference in defects before and after August 2015.

3.1 Defects Pre-Initiative

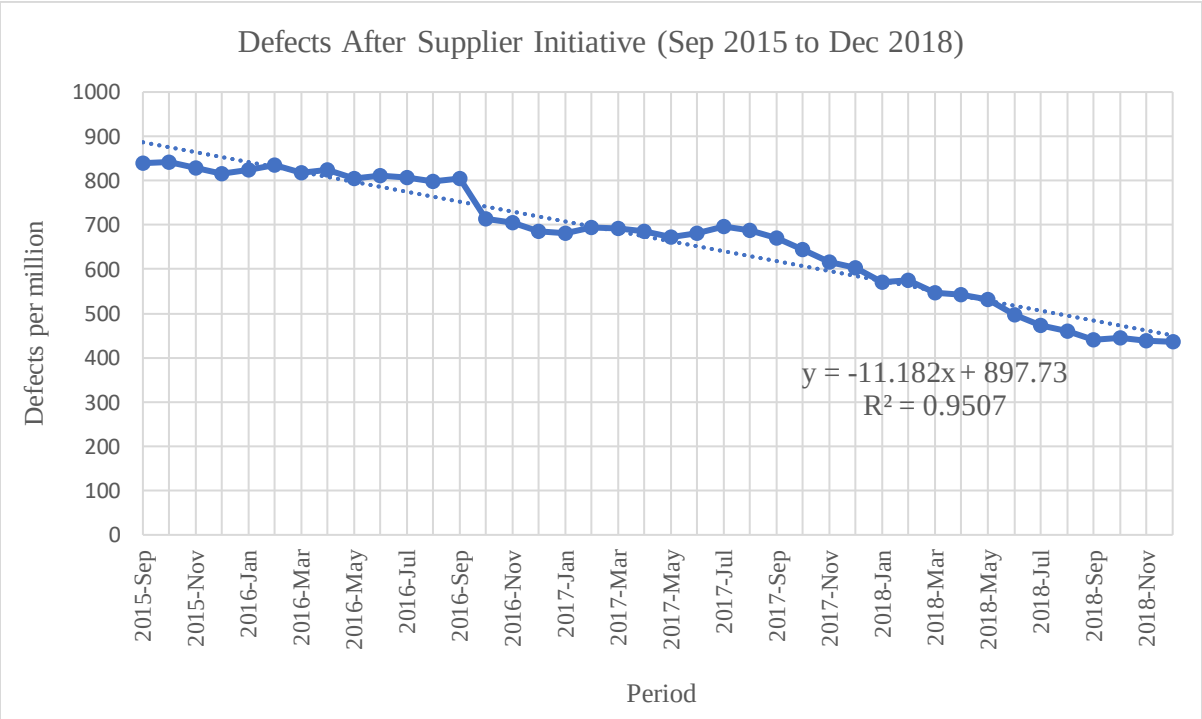
Figure 1



The line Chart above indicates fluctuations but mostly has an upward trend in the number of defects overtime. The linear equation $Y = 1.5203x + 816.04$ shows that defects were rising at a rate of 1.52 per month during the pre-initiative period. The $R^2 = 0.4892$ suggests that our model explains 48.92% variability of the data reflecting a moderate fit.

3.2 Defects Post-Initiative

Figure 2



The linear equation Y = -11.182 + 897.73 indicates that defects were decreasing at an average rate of 11.18 per month during the post-initiative period. The $R^2 = 0.9507$ suggests that 95% to the data fits our model which proves the supplier's initiative was effective in improving the quality of raw materials visualized on the chart consistent decline in defects overtime.

4. Prediction Based on Model

To predict what might have happened without initiative, we will use the trendline from pre-initiative to project likely increase. Let's implement the linear function from Figure 1 Y = 1.5203x + 816.04. The linear function shows steady increase or decrease overtime. Correlation does not imply causation. In this case, time/month does not directly have an impact on defects, it only acts as method to observe trends. To do this, we assign a variable x starting from 1 to 60 representing month for each period.

Alternatively, we can use the data analysis tool in excel to interpret our model illustrated below:

Figure 3- Summary Output

Regression Statistics	
Multiple R	0.699418705
R Square	0.489186525
Adjusted R Square	0.460807998
Standard Error	9.442739538
Observations	20

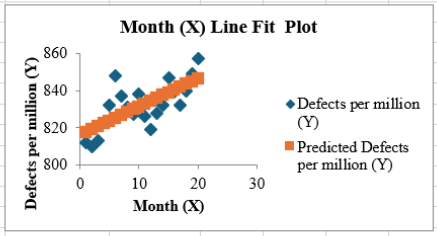
ANOVA

	df	SS	MS	F	Significance F
Regression	1	1537.02406	1537.02	17.2379	0.000598997
Residual	18	1604.97594	89.1653		
Total	19	3142			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	816.0368421	4.386449547	186.036	5.14E-31	806.8212536	825.252431	806.8212536	825.252431
Month (X)	1.520300752	0.366173733	4.15186	0.0006	0.750998285	2.28960322	0.750998285	2.28960322

RESIDUAL OUTPUT

Observation	Predicted Defects per million (Y)	Residuals
1	817.5571429	-5.557142857
2	819.0774436	-9.077443609
3	820.5977444	-7.597744361
4	822.1180451	0.881954887
5	823.6383459	8.361654135
6	825.1586466	22.84135338
7	826.6789474	10.32105263
8	828.1992481	2.80075188
9	829.7195489	-2.719548872
10	831.2398496	6.760150376
11	832.7601504	-6.760150376
12	834.2804511	-15.28045113



Linear function: y= a+bx

Where y = dependent variable (defects)

a = intercept

b = slope

x = independent variable (month)

Hence, to predict for September 2015, y = 816.04 + 1.50203 (21) = 848

4.1 Using Linear Model.

Figure 4

Date	Month (X)	Actual Value (Y)	Predicted Value (Y)
2014-Jan	1	812	818

2014-Feb	2	810	819
2014-Mar	3	813	821
2014-Apr	4	823	822
2014-May	5	832	824
2014-Jun	6	848	825
2014-Jul	7	837	827
2014-Aug	8	831	828
2014-Sep	9	827	830
2014-Oct	10	838	831
2014-Nov	11	826	833
2014-Dec	12	819	834
2015-Jan	13	828	836
2015-Feb	14	832	837
2015-Mar	15	847	839
2015-Apr	16	839	840
2015-May	17	832	842
2015-Jun	18	840	843
2015-Jul	19	849	845
2015-Aug	20	857	846
2015-Sep	21	839	848
2015-Oct	22	842	849
2015-Nov	23	828	851
2015-Dec	24	816	853
2016-Jan	25	824	854
2016-Feb	26	836	856
2016-Mar	27	818	857
2016-Apr	28	825	859
2016-May	29	804	860
2016-Jun	30	812	862
2016-Jul	31	806	863

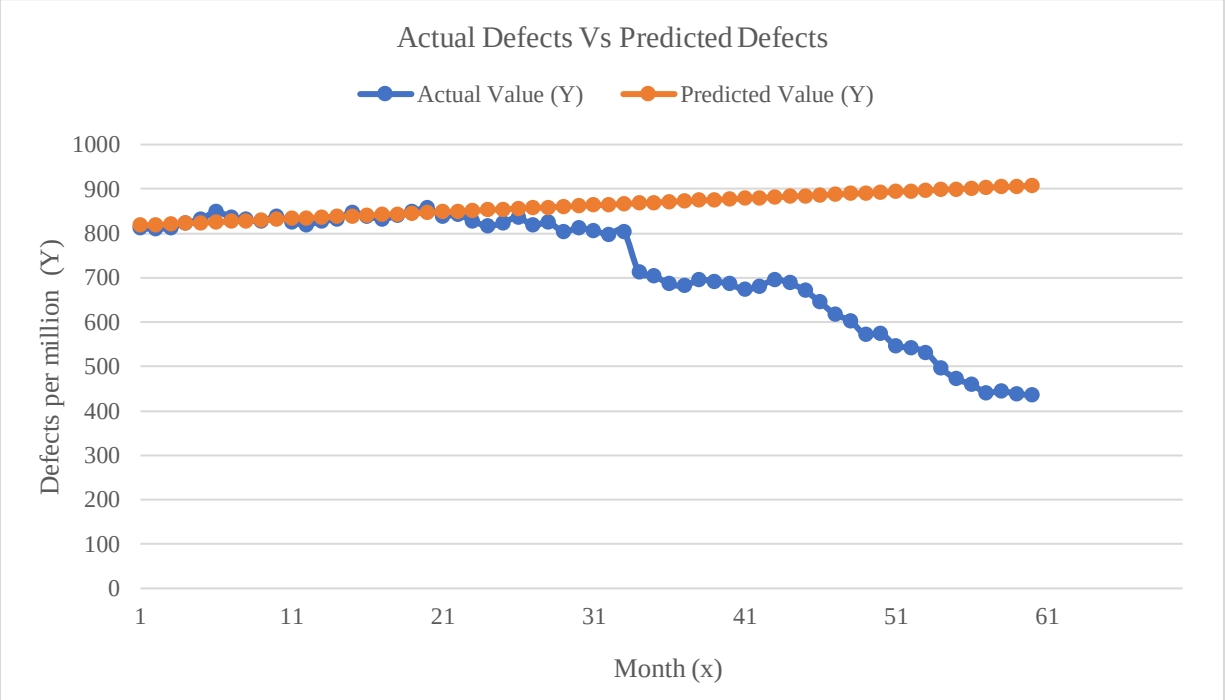
2016-Aug	32	798	865
2016-Sep	33	804	866
2016-Oct	34	713	868
2016-Nov	35	705	869
2016-Dec	36	686	871
2017-Jan	37	682	872
2017-Feb	38	695	874
2017-Mar	39	692	875
2017-Apr	40	686	877
2017-May	41	673	878
2017-Jun	42	681	880
2017-Jul	43	696	881
2017-Aug	44	688	883
2017-Sep	45	671	884
2017-Oct	46	645	886
2017-Nov	47	617	887
2017-Dec	48	603	889
2018-Jan	49	571	891
2018-Feb	50	575	892
2018-Mar	51	547	894
2018-Apr	52	542	895
2018-May	53	532	897
2018-Jun	54	496	898
2018-Jul	55	472	900
2018-Aug	56	460	901
2018-Sep	57	441	903
2018-Oct	58	445	904
2018-Nov	59	438	906
2018-Dec	60	436	907

The table above calculates the predicted value for defects indicating that if PLE supplier did not take action to improve the quality of raw materials, defects would have kept increasing at a consistent rate over time. However, it's essential to consider that they are major factors that directly affect defect rates such as environmental conditions.

Below is a visual representation of what might have happened if the supplier did not take the initiative.

4.2 Actual Defects vs Predicted Defects

Figure 5



5. Limitations

When working on this project, I faced challenges in model selection. Which predictive model would be the best fit for our data? To decide, I had to consider linear, exponential, polynomial, and moving average trendline using excel. The linear model has a moderate variability with a R^2 of 48.9% indicating a good fit. However, Evans, J. R. (2021) discussed in his text that *"The R^2 is a measure of the "fit" of the line to the data. The larger the value of R^2, the better the fit.* (p.314)

Thus, exploring other models illustrated below:

Figure 6- Comparing Models R squared

Model	Function	R squared
Linear	$y = 1.5203x + 816.04$	$R^2 = 0.4892$
Polynomial order-3	$y = 0.0382x^3 - 1.209x^2 + 11.982x + 795.37$	$R^2 = 0.6946$
Polynomial order-2	$y = -0.0046x^2 + 1.616x + 815.69$	$R^2 = 0.4893$
Power	$y = 808.25x^{0.0136}$	$R^2 = 0.5097$

I decided to choose between linear or polynomial function since the other models R^2 are slightly similar to linear trendline. Furthermore, I determined the better model using Mean Percentage Error which states how reliable a model's predictions will be, based on the difference between the actual values and predicted values.

Figure 7 – Determining Model Selection

			Y=a+bx	Y=ax^+bx^+cx+d		
			Linear	Polynomial order =3	Absolute Percentage Error	
Period	Month (X)	Actual Value (Y)	Predicted Value (Y)	Predicted Value (Y)	Linear	Polynomial
2014-Jan	1	812	818	806	0.68%	0.72%
2014-Feb	2	810	819	815	1.12%	0.59%
2014-Mar	3	813	821	821	0.93%	1.04%
2014-Apr	4	823	822	826	0.11%	0.41%
2014-May	5	832	824	830	1.00%	0.26%
2014-Jun	6	848	825	832	2.69%	1.89%
2018-Sep	57	441	903	4625	104.69%	948.68%
2018-Oct	58	445	904	4877	103.19%	995.85%
2018-Nov	59	438	906	5139	106.79%	1073.35%
2018-Dec	60	436	907	5413	108.09%	1141.53%
				MAPE	25.53%	203.72%

To avoid the risk of overfitting, I used ABS function to determine the Absolute percentage Error (APE). The model with a lower MAPE will be more accurate in a better explanatory power. Hence, linear model is accurate for when data has consistent slope.

6. Conclusion

PLE supplier initiative was effective in minimizing defect rate of raw materials overtime. The defect would have continued to increase if no action was implemented.

Moreover, it is crucial to consider other factors that directly affect defect rate such as change in temperature. An additional variable may have higher explanatory power in predicting what might have happened without initiative and even forecast future outcomes.

7. Reference

Evans, J. R. (2021). *"Business analytics: Methods, models, and decisions (3rd ed.)"*. Boston, MA: Pearson.