**ENGINES**

**Engines, which lists 50 samples of the time required to produce a lawn-mower blade using a new technology.**

**Reason**

**Performance Law equipment has produced 50 units on engine production line, therefore with the tests and regression analysis we can note whether PLE is following consistency or not. The sheet is chosen to identify was there an improvement in their production time as the sample increases. So that with the initiative taken will help in producing goods as early as possible. Because as the demand increases, sample will also increase with that it should be necessary to note that whether there is any change in production time.**

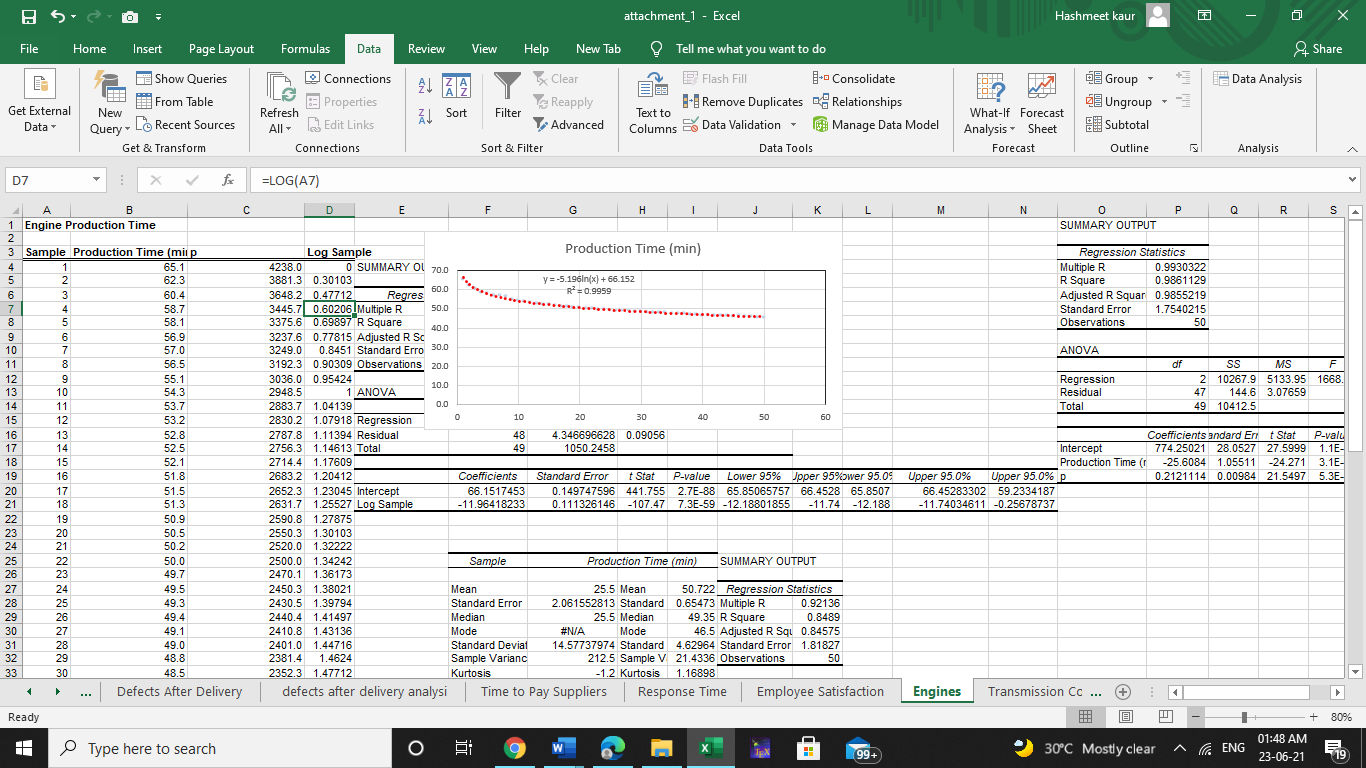
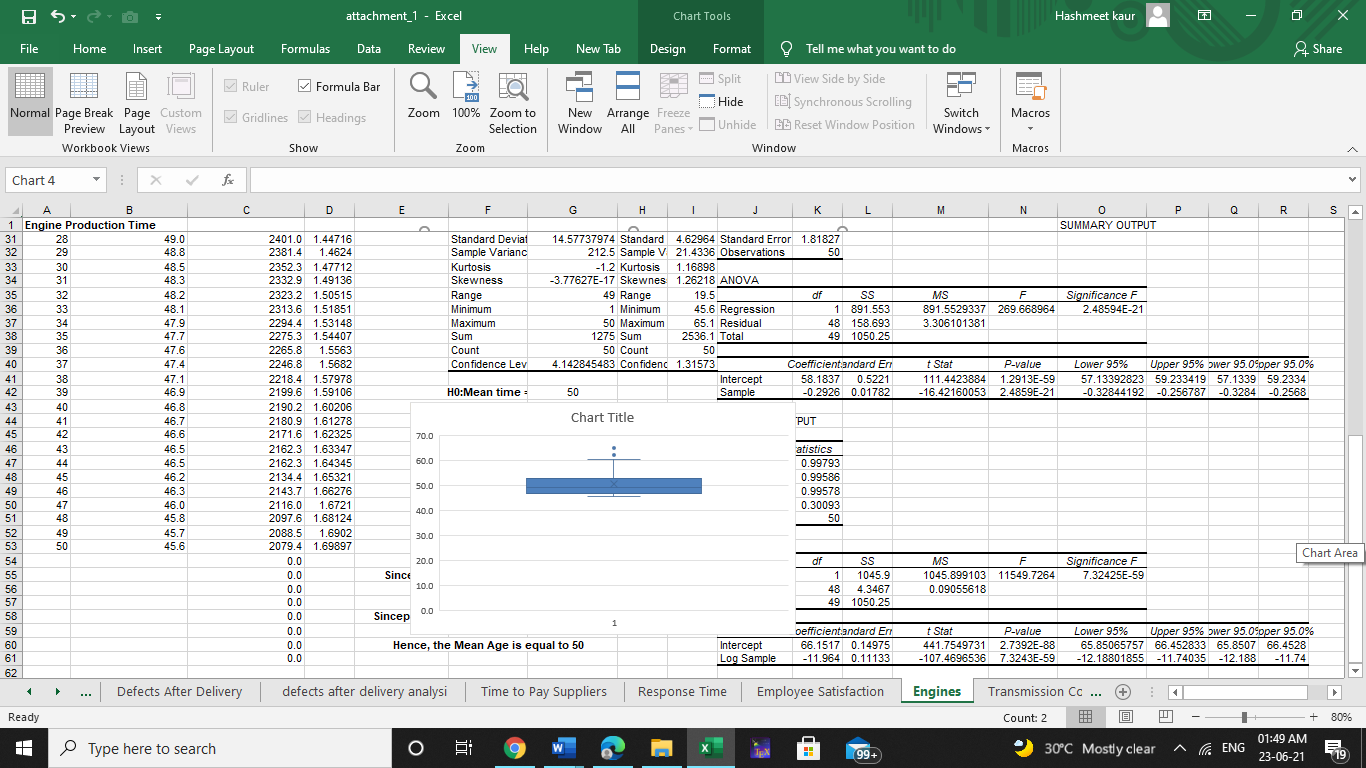
**BUSINESS OBJECTIVES (ANALYTICAL RESEARCH)**

* **As an investigator we would like the production time to decrease with the increase in sample. With the help of regression analysis we’ll see whether there is any change in production time.**
* **We would also like the production time should be 40 min, to achieve that we’ll perform hypothesis test.**
* **Is there any outlier, if yes, then it should be taken care for that to provide a consistency during the process.**
* **What is average production time(min), is there any variation during the process.**
* **What is the maximum and minimum time taken by the engines during the production.**

**HYPOTHESIS TESTING (NULL AND ALTERNATE)**

* **By using logarithmic regression we assume Null Hypothesis(H0): There is no effect on production time with increase in sample size and Alternate Hypothesis(H1): There is some effect on production time with increase in sample size.**
* **By using One Sample T-Test (2 tailed) we assume Null Hypothesis(H0): Average production time is 40min and Alternate Hypothesis(H1): Average production time is not equal to 40min.**

**DATA**

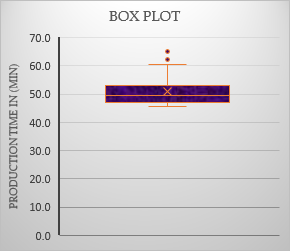
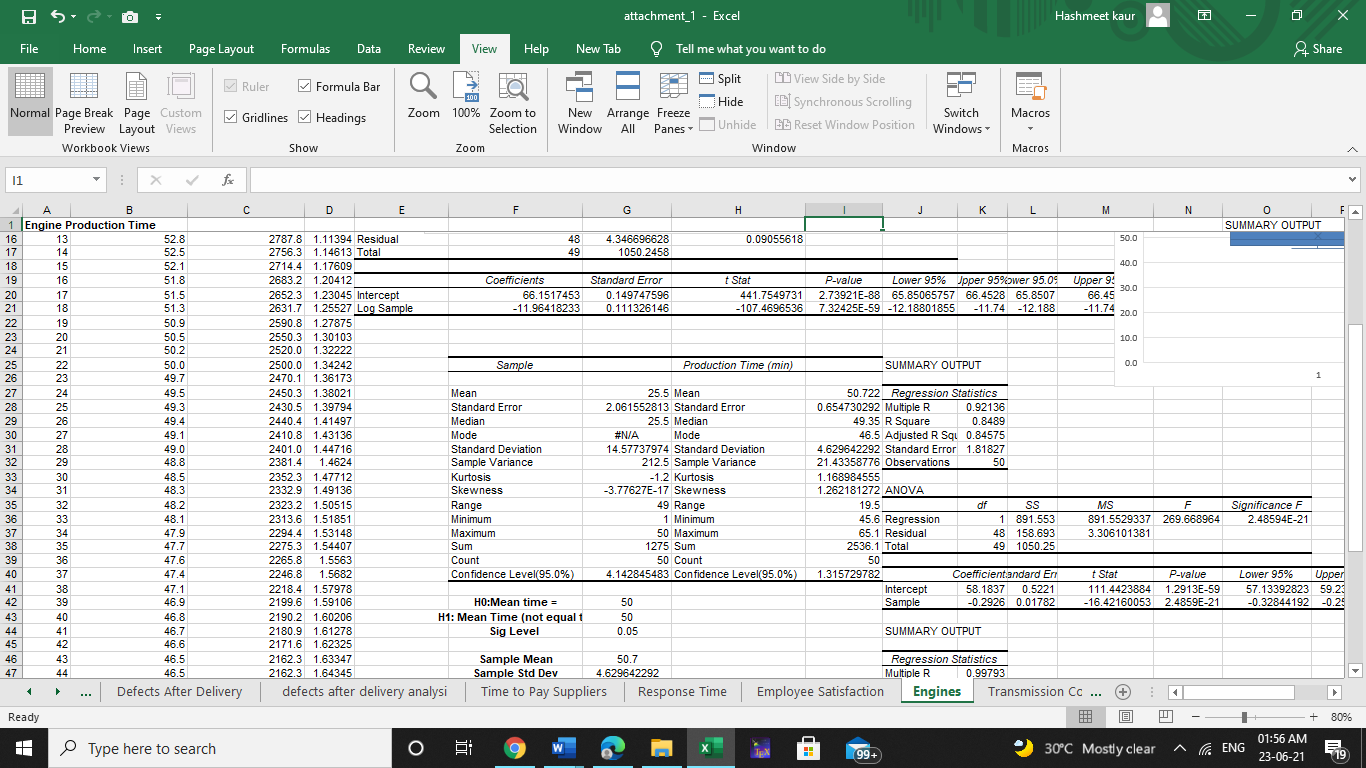
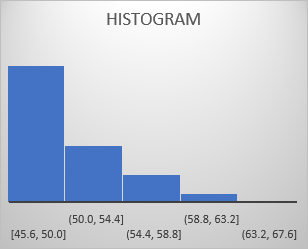
**Here data shows the production time (min) by engines.**

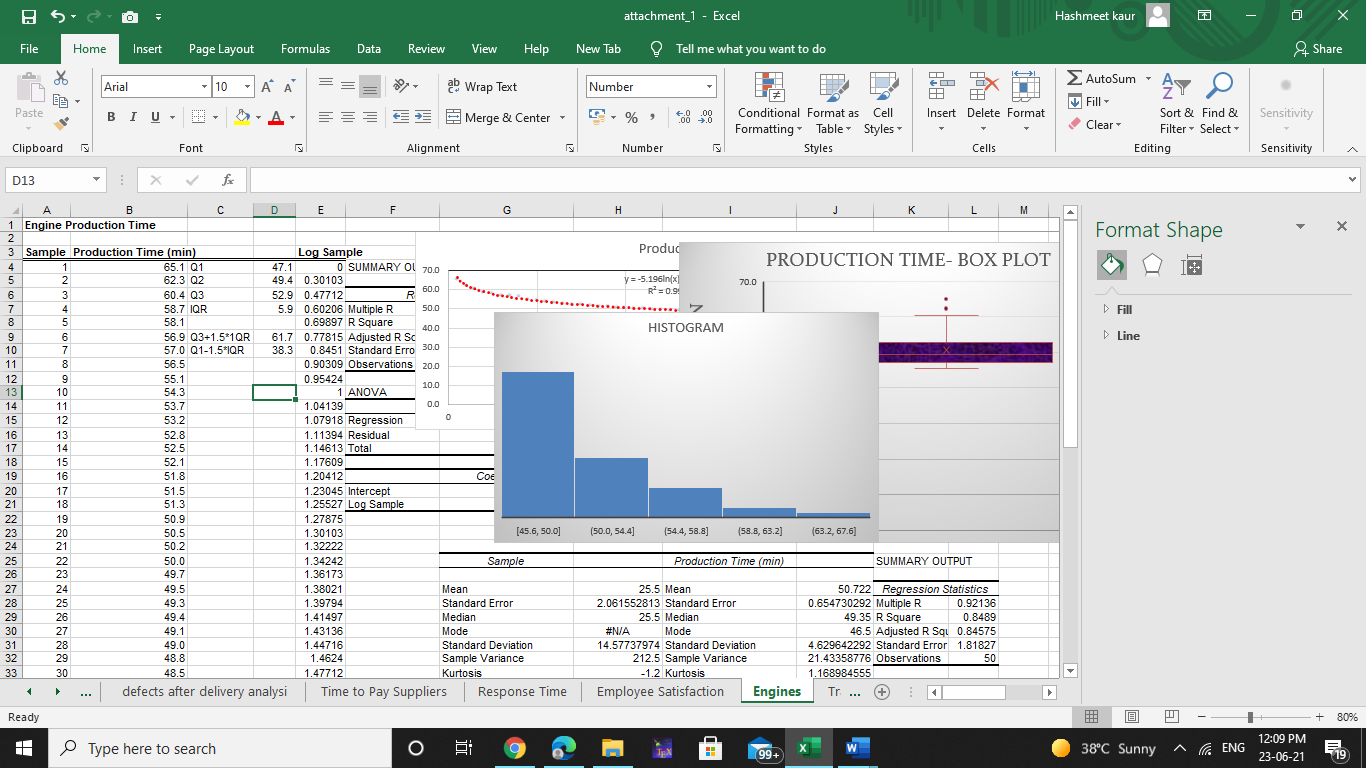
**Sample size: 50**

**Time: in minutes**

**STATISTICAL TECHNIQUES**

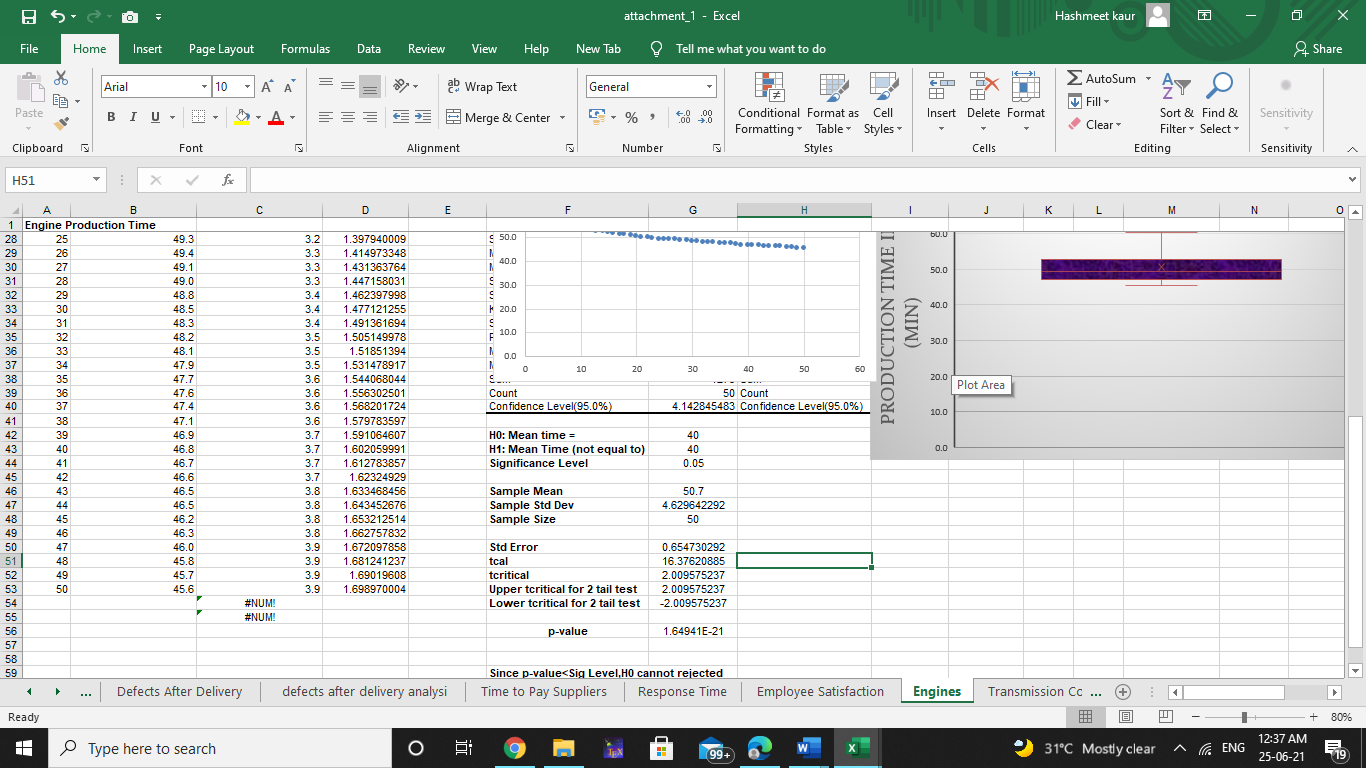
* **DESCRIPTIVE STATISTICS**

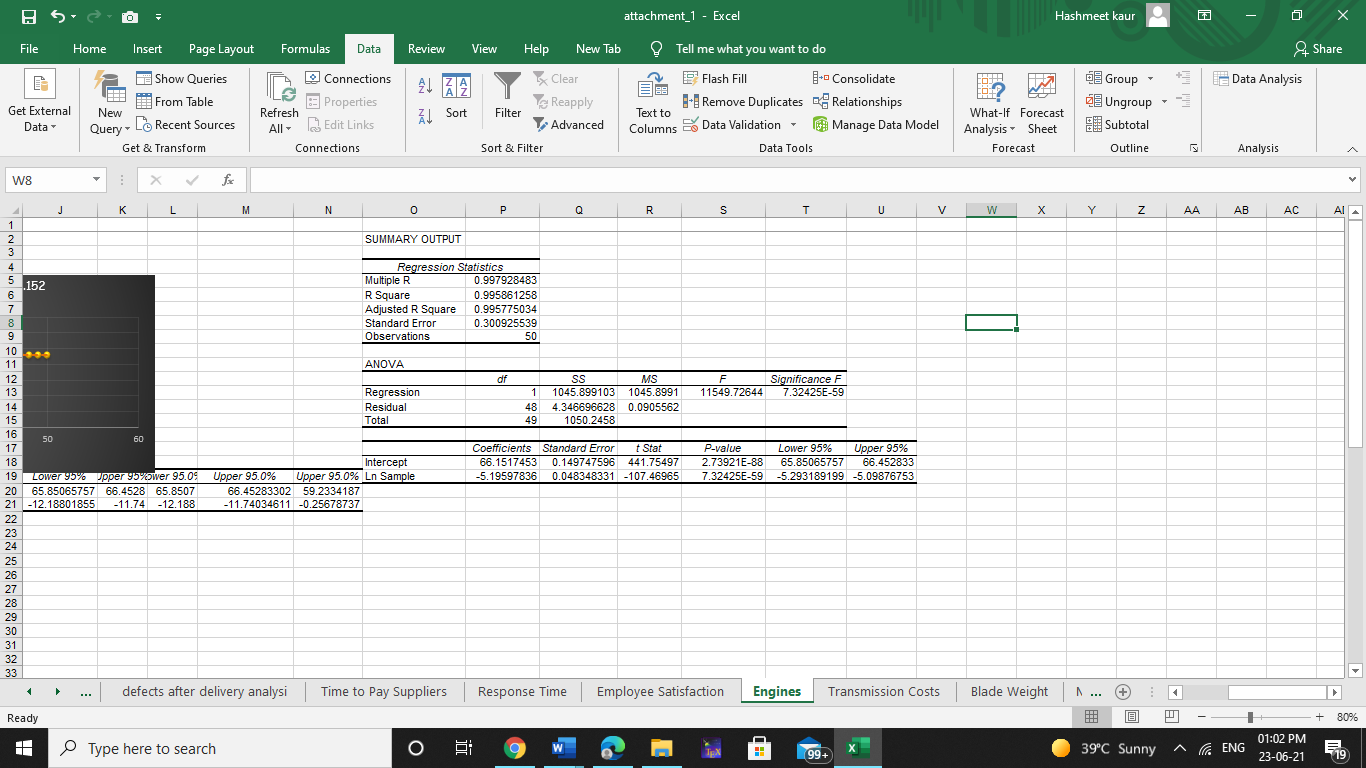


* **INFERENTIAL STATISTIS**

1. **ONE-SAMPLE TEST ( 2 TAILED )**

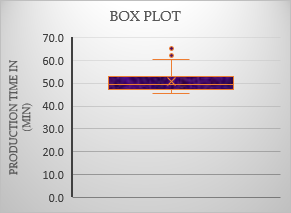
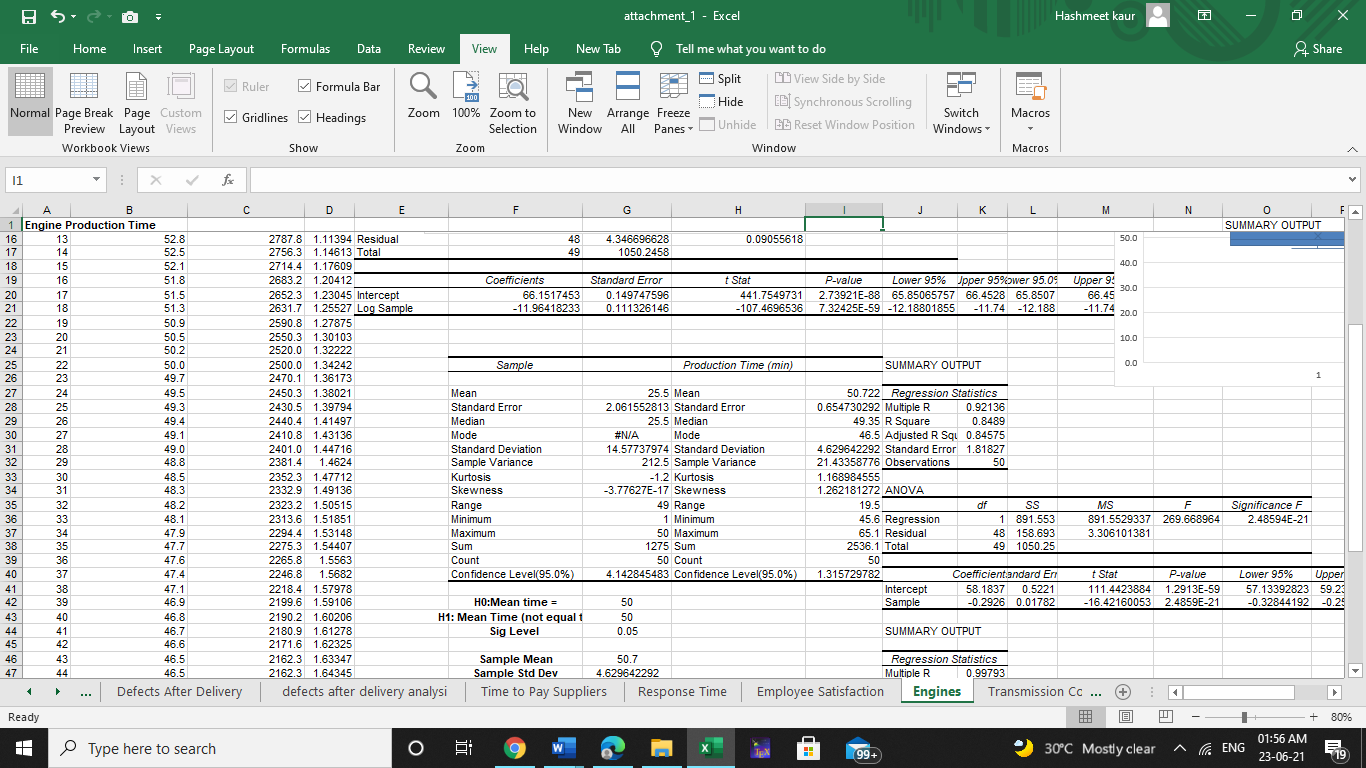
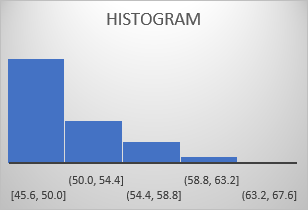


1. **OVERALL REGRESSION ANALYSIS (LOGRAITHMIC)**

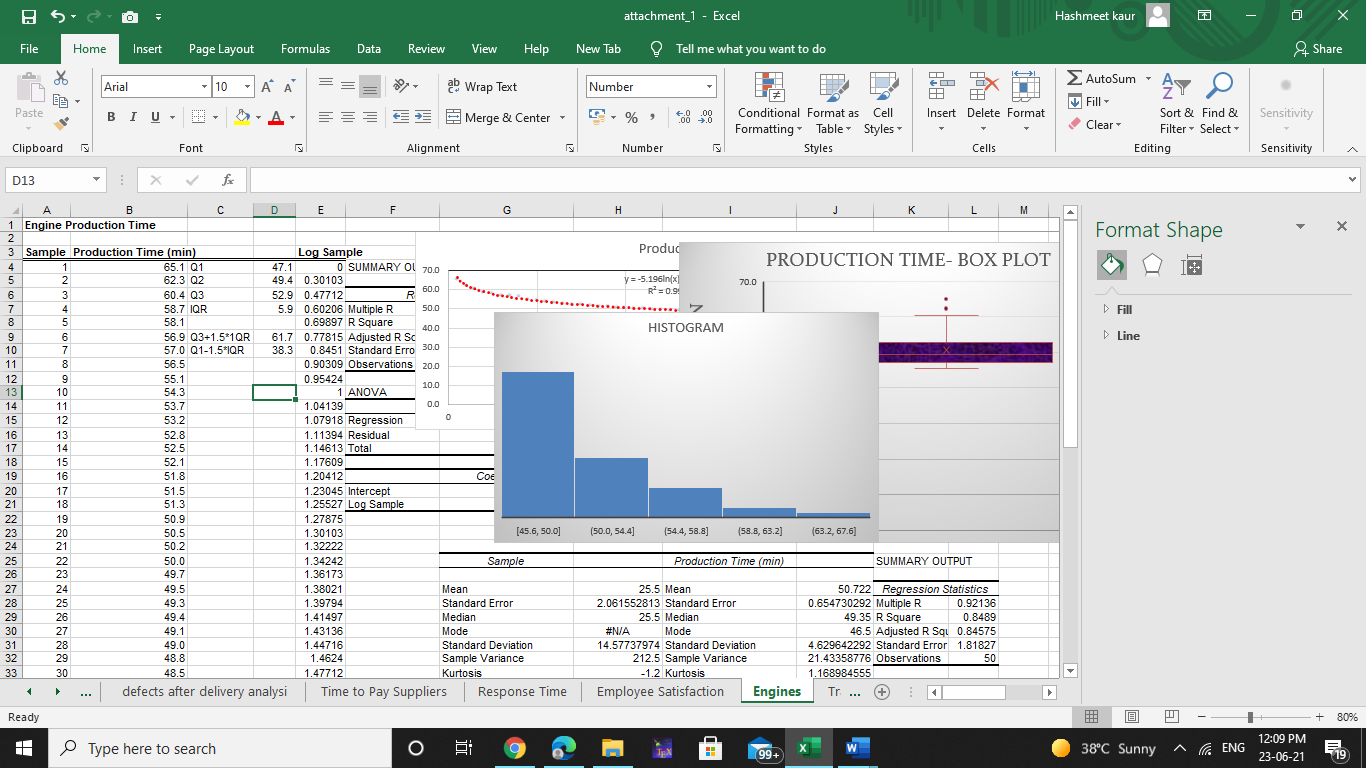


**RESULTS AND DISCUSSION**

* **DESCRIPTIVE STATISTICS**

Overall analysis shows that average production time is 50.7min. Least time taken is 45.6 min and maximum time taken is 65.1min with a range 19.5min in a given sample of size 50.



MEASURES OF DISPERSION

Box-Plot tells there are some outliers above Q3+1.5IQR, that is 61.7min. From graphs it’s clear that the data is not Normally distributed and it’s right skewed, with kurtosis: 1.168 which is a high degree of skewness.

* **INFERENTIAL STATISTIS**

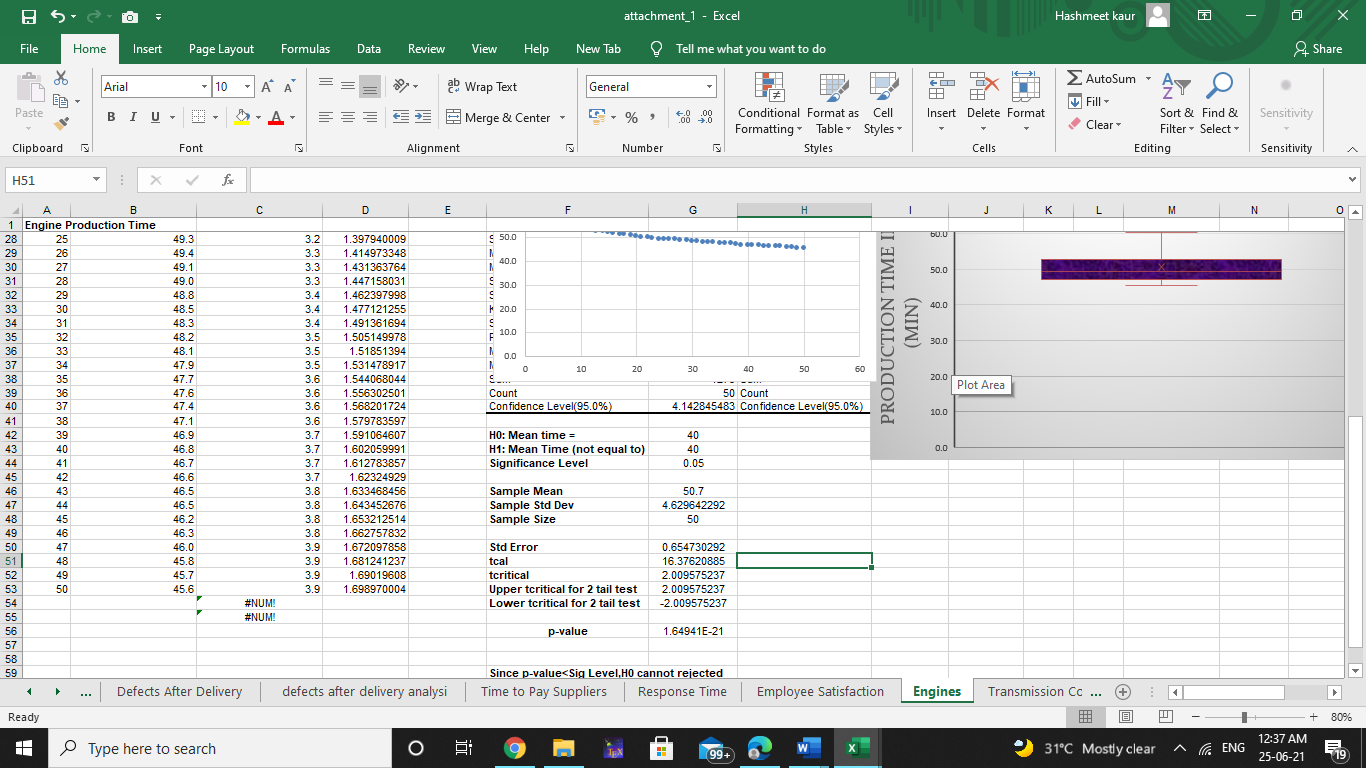
ONE-SAMPLE T-TEST (2-TAILED)

Null Hypothesis(H0): Average production time is 40min

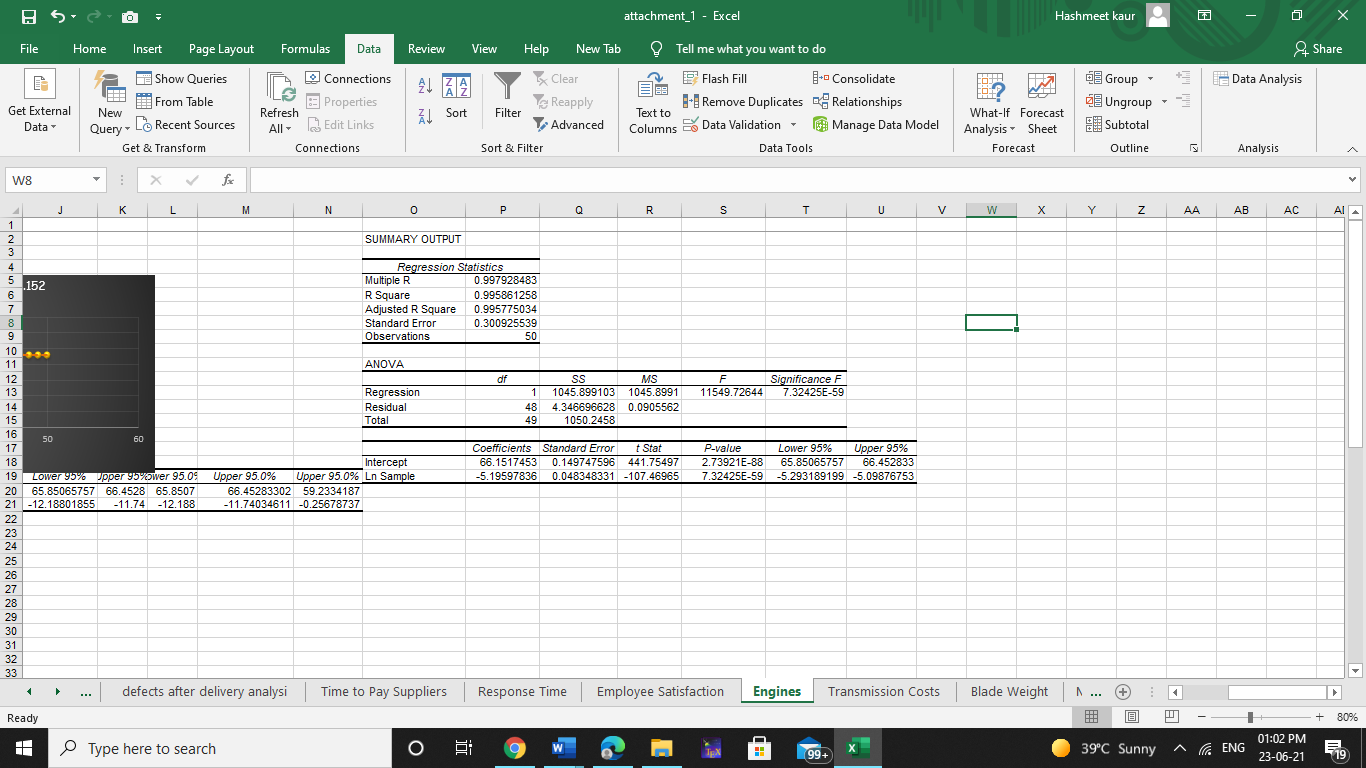
Alternate Hypothesis(H1): Average production time is not equal to 40min

Through the analysis its clear that p-value < 0.05 (p-value=1.64E-21). Therefore, we reject the Null Hypothesis, as Average production time of Null hypothesis is approximately less than sample average production time.

1. **ONE-SAMPLE TEST ( 2 TAILED )**



1. **OVERALL REGRESSION ANALYSIS (LOGRAITHMIC)**



OVERALL REGRESSION (LOGRATHMIC)

We take the natural log of the independent variable, Sample and then use this data with the Production Time using Regression analysis.

Null Hypothesis(H0): There is no change in Production time(min) with increase in samples.

Alternate Hypothesis(H1): There is a change in Production time(min) with increase in samples.

Through the analysis it’s clear that p-value of sample (Ln) < 0.05 (p-value=7.324E-59) Therefore, we reject the Null Hypothesis, which conclude that there was a change in production time as the sample increase.

In the ANNOVA analysis, we see that because the Significance F is very slow, we can reject the null hypothesis and conclude the slope is significant. In, examining the trend lines for the chart of engine production time, we see that the trend line has highest R squared value which is a Logarithmic trend line. With both these tests it tells us that the independent variable or Sample (Ln) is significant and does affect the dependent variable the engine production time. The further down the line engines are sampled, smaller we expect the production time.

**IMPLICATIONS AND MANGERIAL RECOMMENDATIONS**

1. **First the average production time is around 50min. There is a huge variation of about 21.43min.**
2. **Considering the outliers, there are 2 outliers which are about Q3+1.5IQR above, that is 61.7min, this can be the one of the reasons like defects in materials or can be engines, which has to be taken care.**
3. **Although through regression its clear that the sample variable is significant meaning increase in samples would likely cause production time to decrease making the slope negative.**
4. **Our assumption of having 40min as a production time was failed as the average production time is around 50min.**
5. **As its clear that production time decreases as sample increases therefore the data is positively skewed.**
6. **Its recommended that may be by increasing the sample size may decrease the production time to around our assumption but it might not makes the data normally distributed.**

**MOWER BLADE WEIGHTS**

**PLE has recorded Mower Blade Weights of 350 samples to evaluate the consistency of the production process.**

**Reason**

**As an investigator its compulsory to have a check on the production process so that there should have a smoothing service, for that its necessary to have a check that the tools which are used in that process are they providing consistency during the process. With the help of Regression and Hypothesis testing we’ll see whether the Mower Blade will continue to manufacture at a steady average or not.**

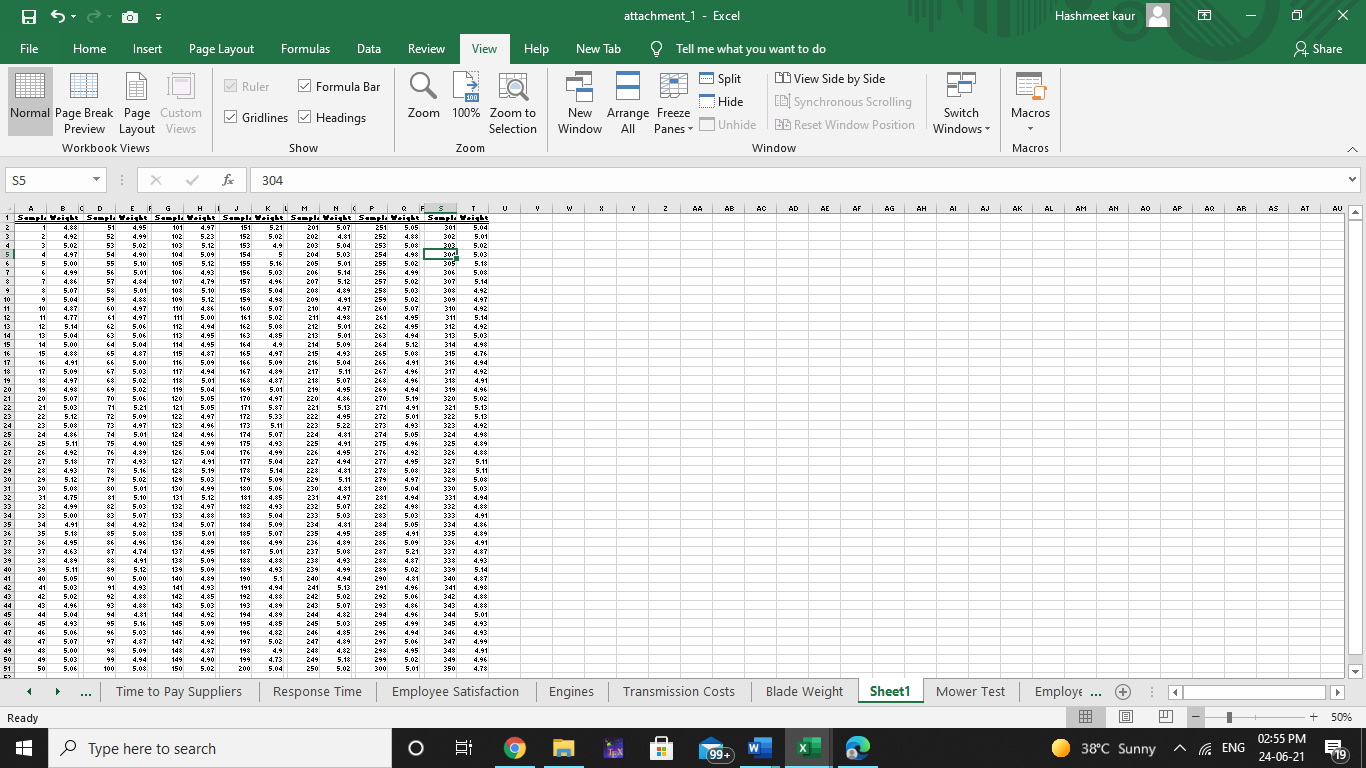
**BUSINESS OBJECTIVES (ANALYTICAL RESEARCH)**

* **Upon investigation, PLE has performed an experiment to check whether the Mower Blade will continue to manufacture at a steady average of 4.9,will there be any change in Mower Blade Weights with the increase in sample size.**
* **As an investigator we want blade weights should be less than 5.01, will it achieve.**
* **How much variability is occurring in the measurements of blade weights.**
* **Assuming data is normally distributed what is the probability that blade weights during this process will exceed 5.20?**
* **What is the probability that weights will be less than 4.8?**
* **Is the process that makes the blades stable over time, were there any apparent changes in the pattern.**
* **Is there any outliers which might indicate a problem during the manufacturing process.**
* **What is the sampling distribution, with overall mean and standard error.**
* **How many blade weights must be measured to find a 95% confidence interval for the mean blade weight a sampling error at most 0.05.**

**HYPOTHESIS TESTING (NULL AND ALTERNATE)**

1. **Using Regression Analysis(Linear) we assume Null Hypothesis(H0): Sample is insignificant variable, there is no change in blade weights and Alternate Hypothesis(H1): Sample is significant variable, there is a change in blade weights.**
2. **With the help of One Sample T-Test (1 Tailed) we assume Null Hypothesis(H0): Blade weights should be greater than or equal to 5.1 and Alternate Hypothesis(H1): Blade weights should be less than 5.1.**

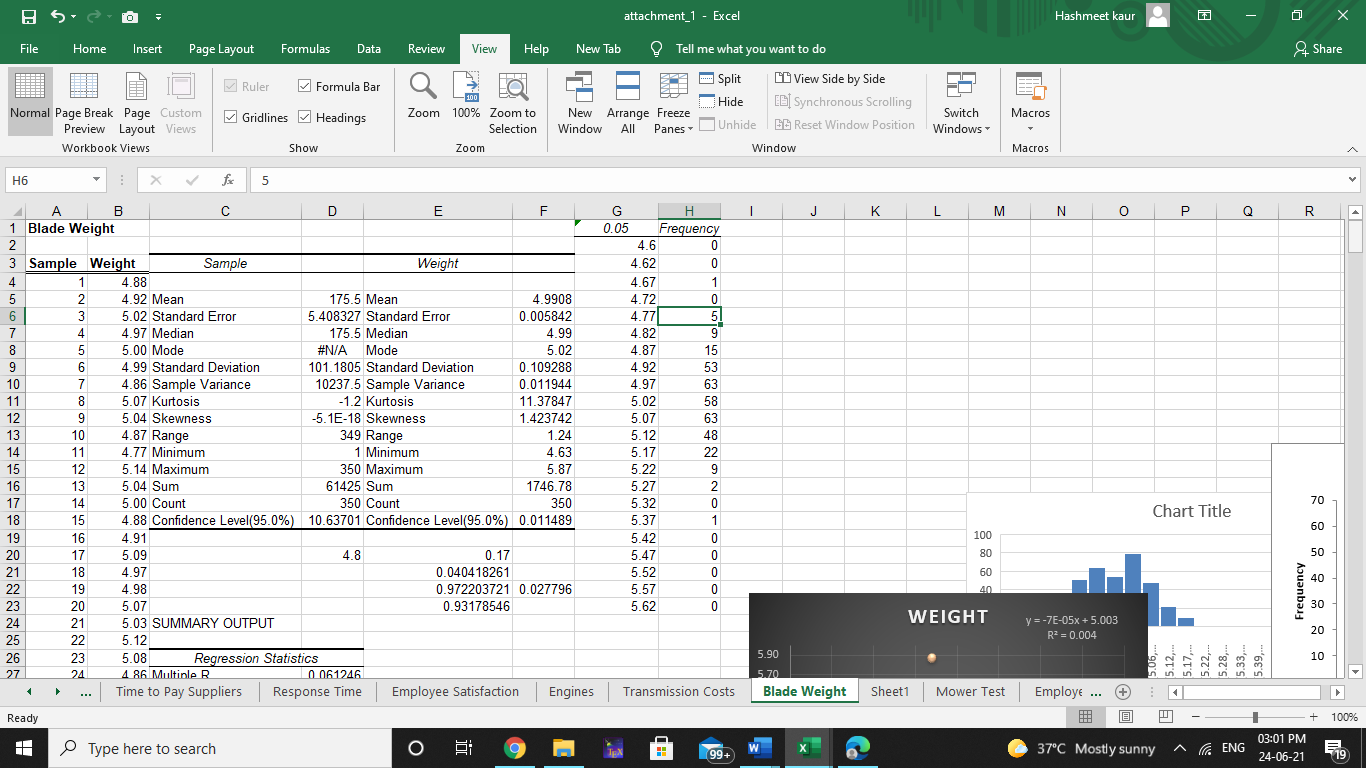
**DATA**

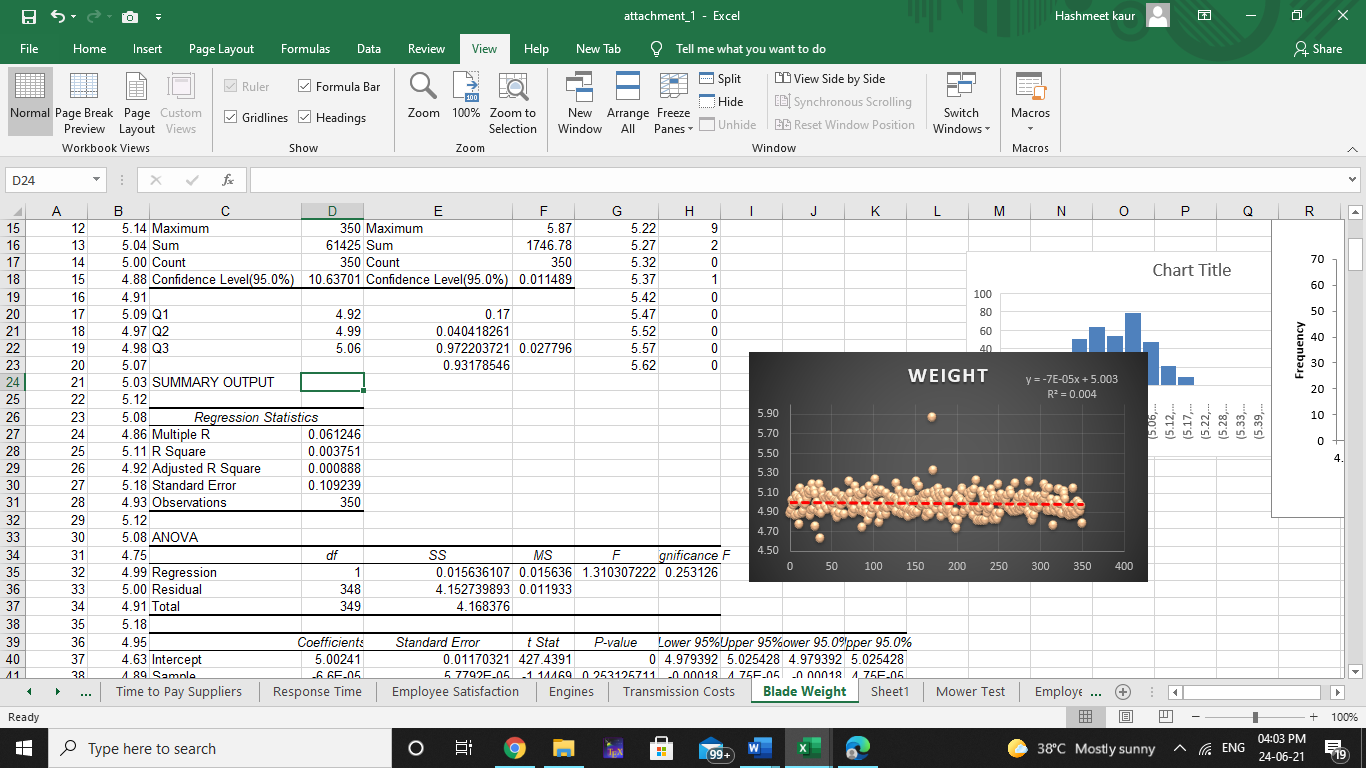


**Sample Size: 350 samples**

**STATISTICAL TECHNIQUES**

* **DESCRIPTIVE STATISTICS**

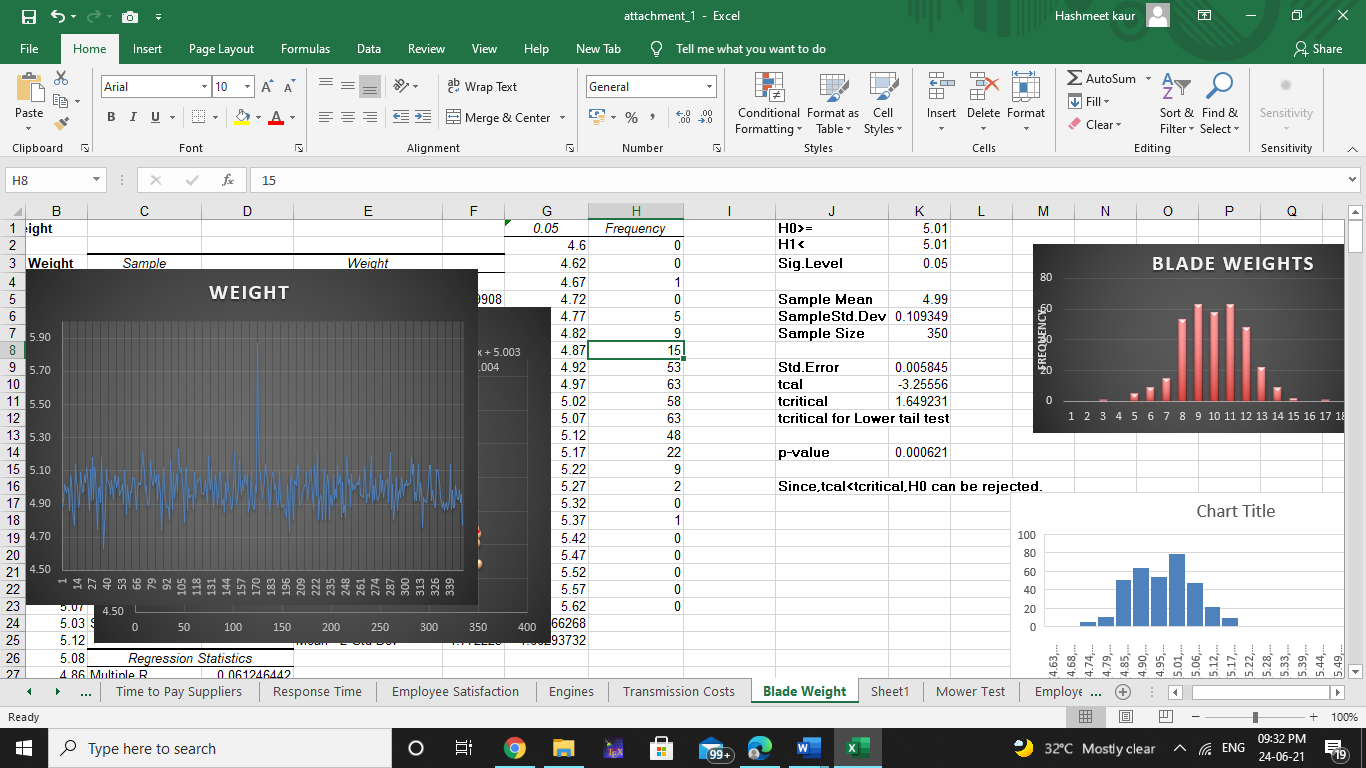




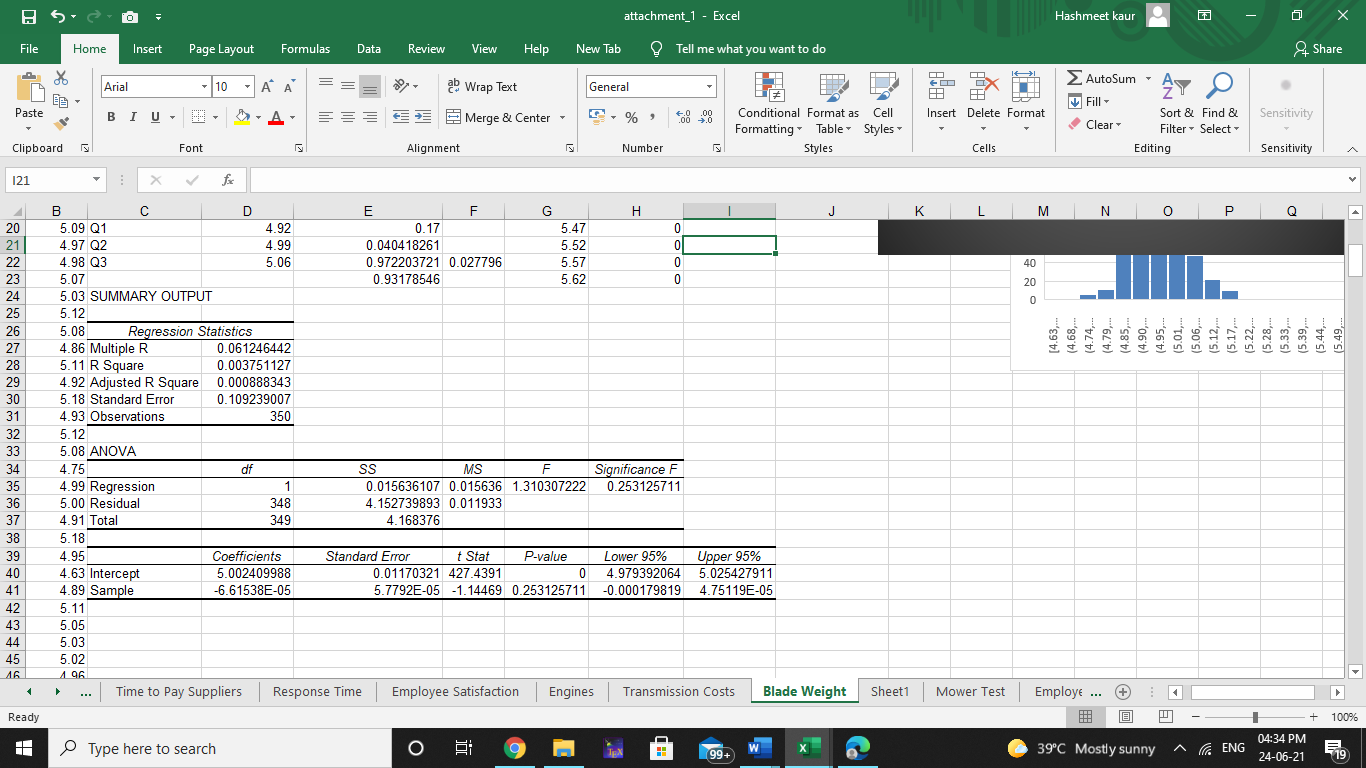
Overall analysis shows that average is 4.99, with sample variance approximate 0.011. Minimum and maximum weight of blades are 4.63 and 5.87 respectively with a range of1.2.

* **INFERENTIAL STATISTICS**

1. **ONE – SAMPLE T-TEST ( 1 TAILED )**

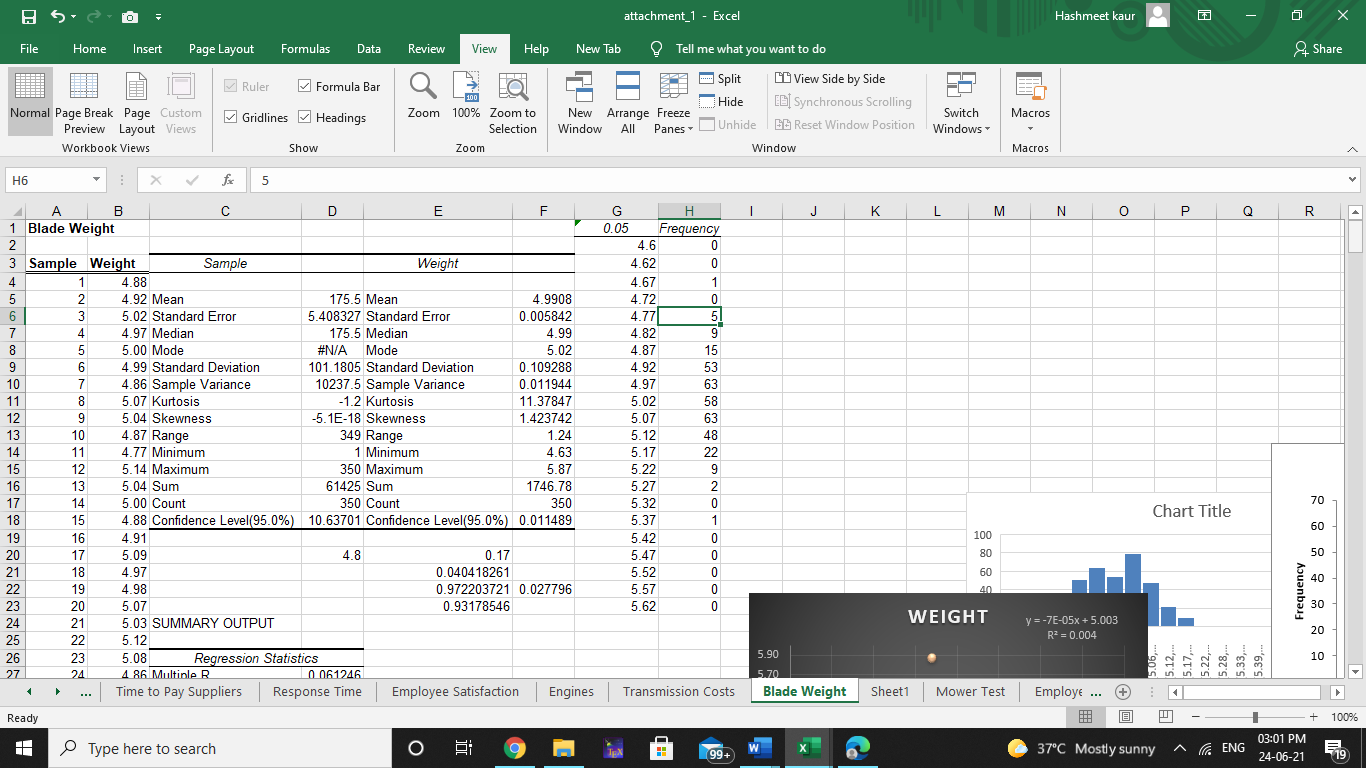


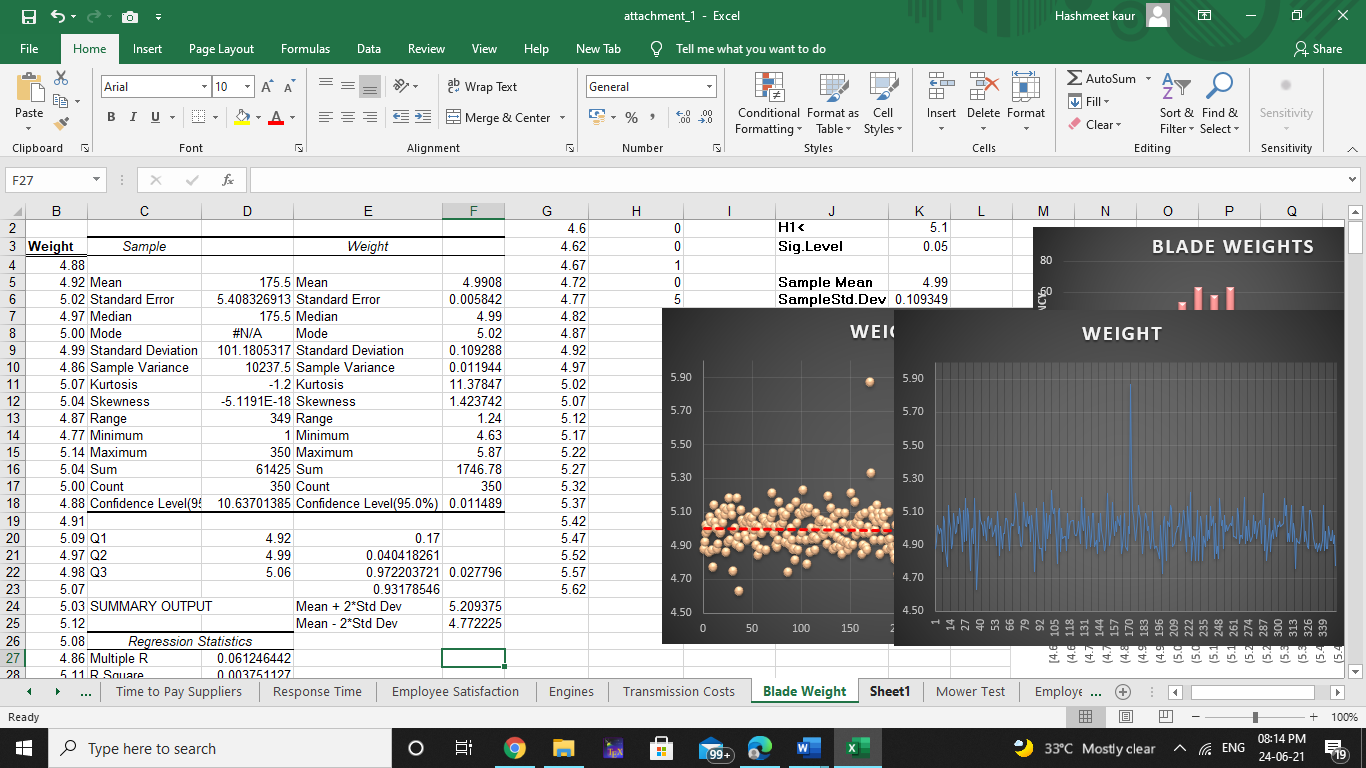
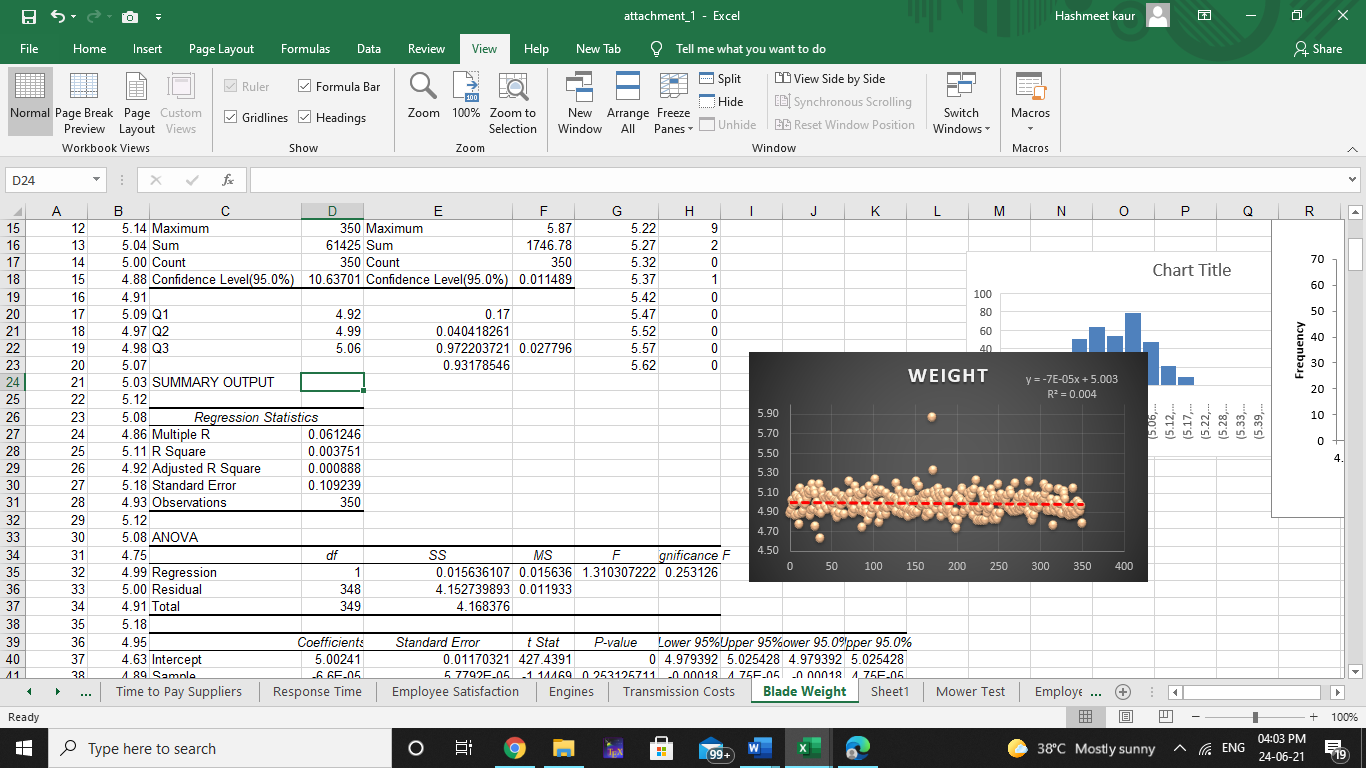
1. **REGRESSION ANALYSIS**



**RESULTS AND DISCUSSION**

1. **Descriptive analysis**





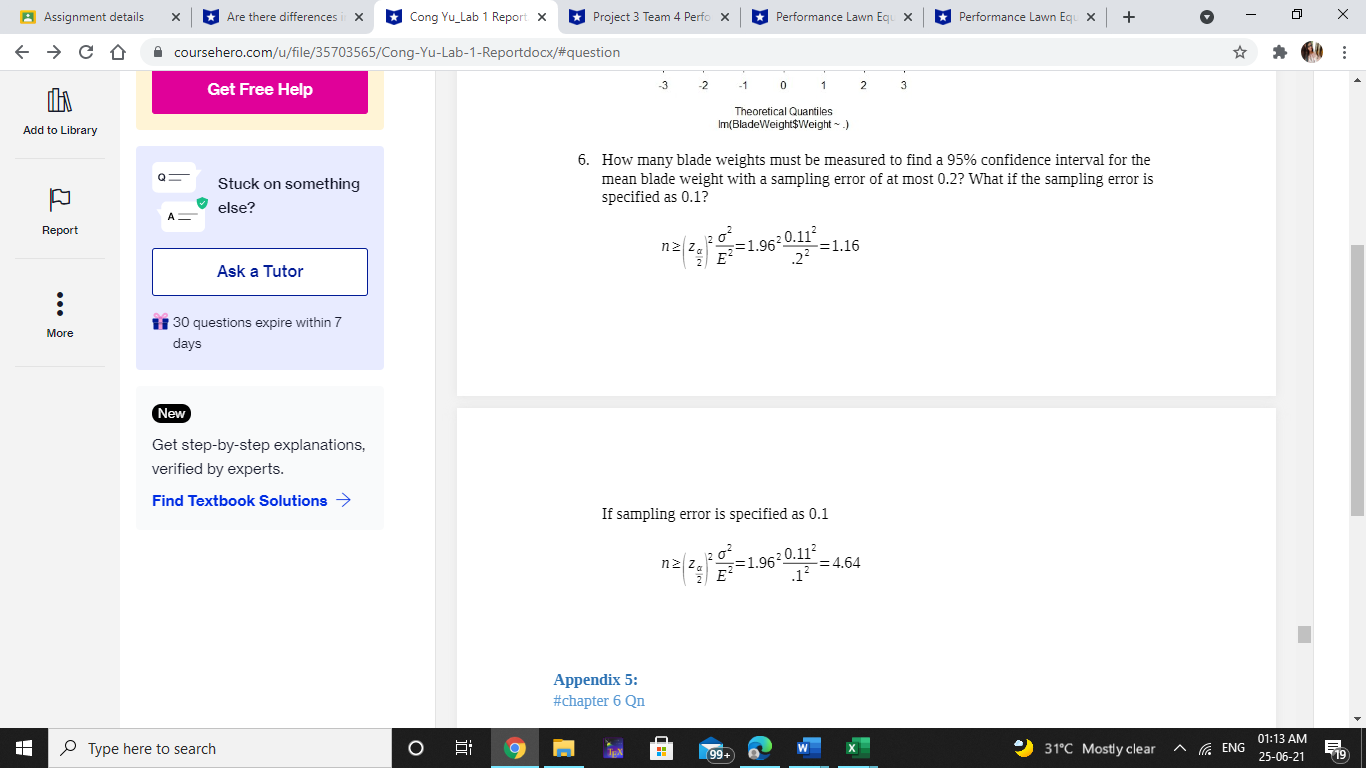
The average weight of 350 samples of blades is approximately 4.99, with 0.10 standard deviation. There is some variability in the data which is 0.011. Most of the weights of blades are 5.02. With the help of Quartiles, 75% of samples have less weight than 5.06.

As the data is normally distributed, we can find in what range of weights will fall under 95% confidence level:

Mean+2\*std dev=5.209375

Mean-2\*std dev=4.772225

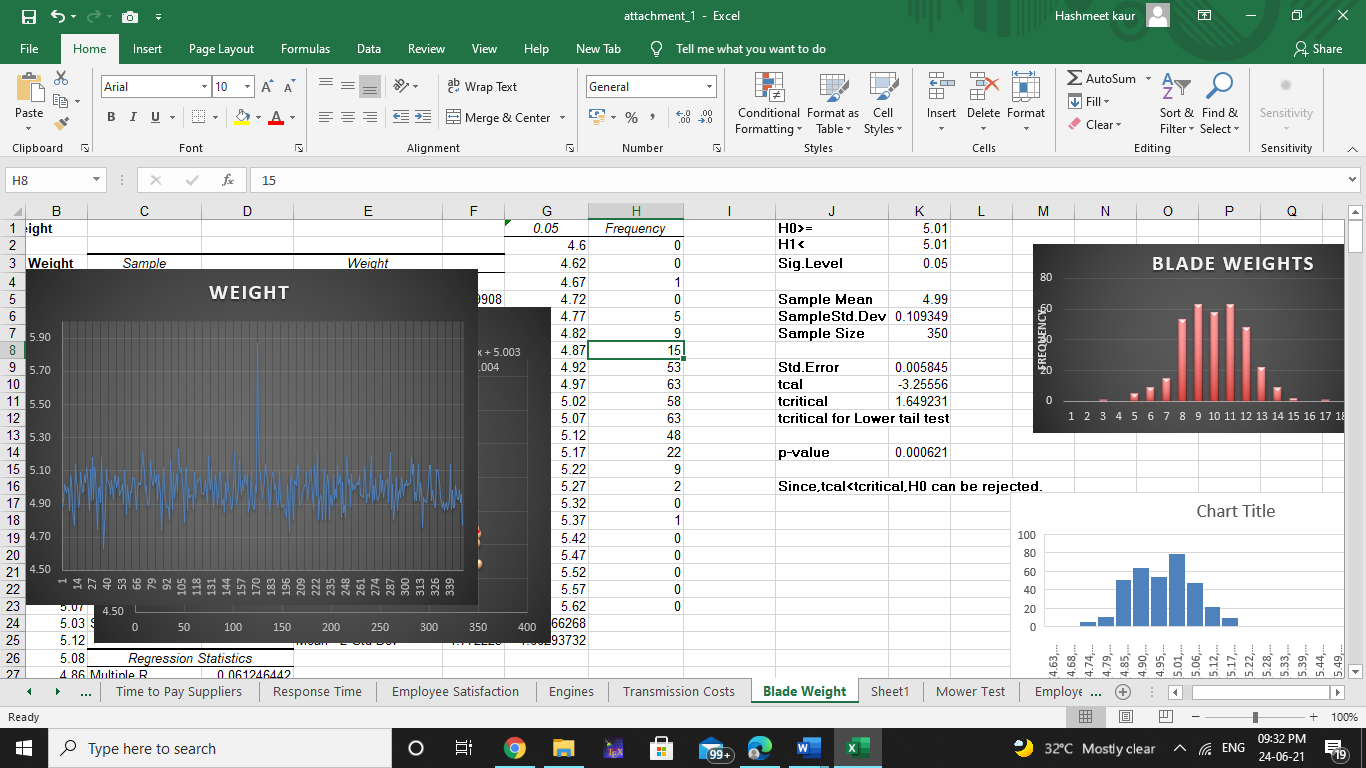
Therefore 95% of sample lie between 4.77 and 5.20. With 95% confidence interval and standard error of 0.05, will results :

 =(1.96)^2\*[(0.10)^2/(0.05)^2] = 15.366 samples, aprroximately 15 samples.Assuming data is normally distributed probability of weights measured above 5.2 will be 2.77% similarly probability of weight measured less than 4.8 will be 4.0%.

**INFERENTIAL analysis**

There are, however, 2 outliers in the data set. Blade number 37 whose weight is 4.63 which is actually 3 standard deviation away from mean. Sample 171 was recorded as 5.87, which is outside the expected range of values. Also the mean, median, and mode are all similar values and the values are evenly distributed on either side of the mean, a normal distribution is the best-fitting model for this data, as shown in the histogram below.

* **ONE – SAMPLE T-TEST ( 1 TAILED )**



One sample T-Test (1 tailed)

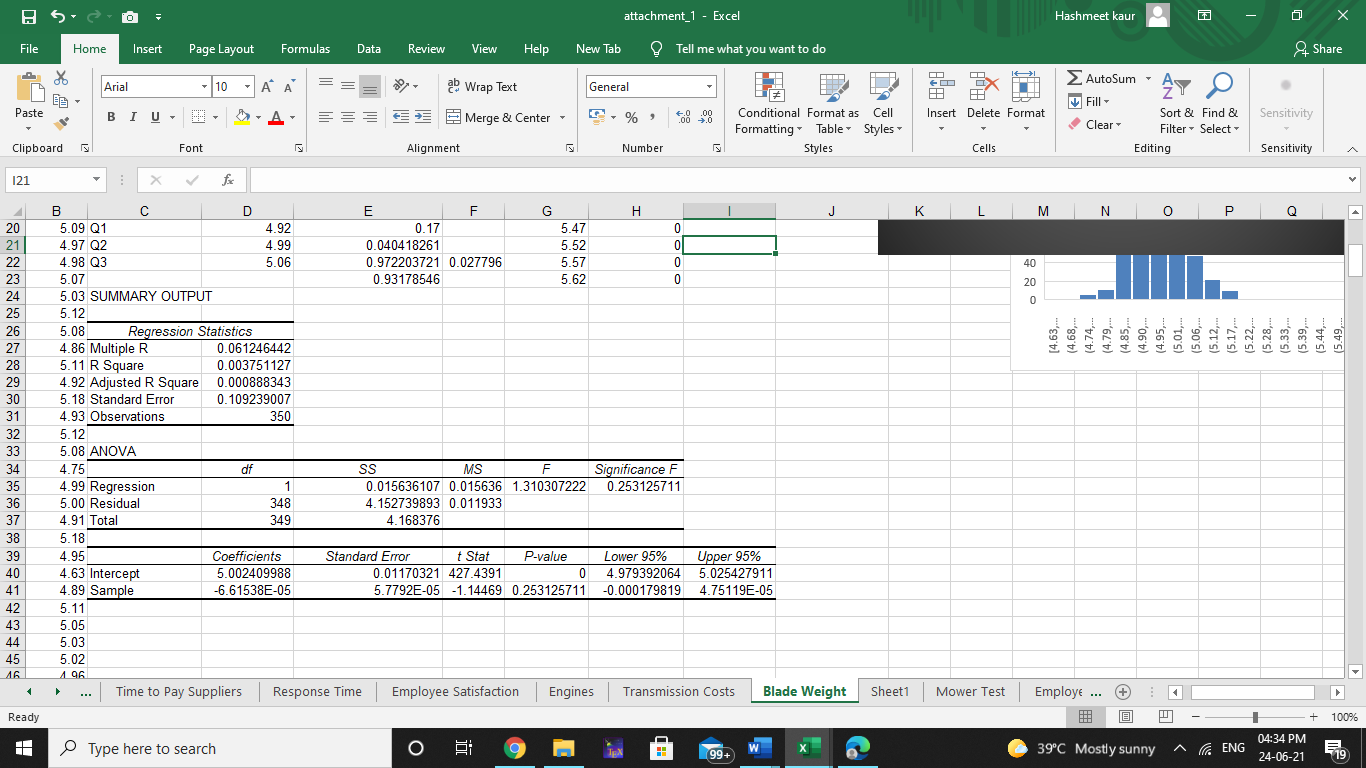
Null Hypothesis(H0): Average blade weights is greater than or equal to 5.01

Alternate Hypothesis(H1): Average blade weights is less than 5.01.

As tcal < tcritical and p-value < 0.05, so we’ll reject the Null Hypothesis.

Therefore, its clear to say that weights are reduced to 4.99 which is average mean weights of 350 sample blades.

* **REGRESSION ANALYSIS**



Regression Analysis (Linear)

Null hypothesis(H0): Sample is insignificant variable, there is no change in blade weights

Alternate Hypothesis(H1): Sample is significant variable, there is a change in blade weights

As p-value of Sample is greater than 0.05 (p-value: 0.25), therefore we accept the Null Hypothesis. Plotting a trend line of the average weight gives us a slope of -7E-5

Which declines at a very slight rate of -0.00007. This indicated a negative slope, which is enough to be considered that sample is insignificant variable as blades continue to be manufactured at a steady average of approx. 4.99.

**IMPLICATIONS AND MANGERIAL RECOMMENDATIONS**

1. **First of its clear that through regression analysis that sample is insignificant so we can safely assume that blades will continue to manufactured at a steady average weight of approximate 4.99.**
2. **Overall, we can assume Blades manufacturing process is stable and predictable .**
3. **Also from the Histogram graph its clear that there were 2 outliers out of which one was 3 standard deviation away from mean and other was out of range of values.**
4. **Considering the outliers, there could result number of issues such as individual defects, problems with the material or error in measuring or recording the value.**
5. **Its also clear as the data is evenly distributed on the either side of the mean, normal distribution is the best fitting model for this data.**
6. **Through hypothesis testing its clear that our assumption of having less weight than 5.01 was correct.**
7. **Also considering the sample error of 0.05 with 95% confidence interval, will have 15 samples.**

**DEFECTS AFTER DELIVERY**

**Defects after delivery, shows the number of defects in supplier-provided material found in all shipments received from suppliers. The data shows the defects per million year and month wise.**

**REASON**

**Upon investigation we can find that how Performance Law Equipment experienced some quality problems due to an increasing number of defects from 2010 to 2014. After performing some analysis like regression and hypothesis testing we can find what might have happened which made the supplier initiative not been implemented and how the number of defects might further be reduced in the near future. Defect tracking has long been used to measure the quality problems, with an emphasis on finding as many bugs as possible early in the development cycle.**

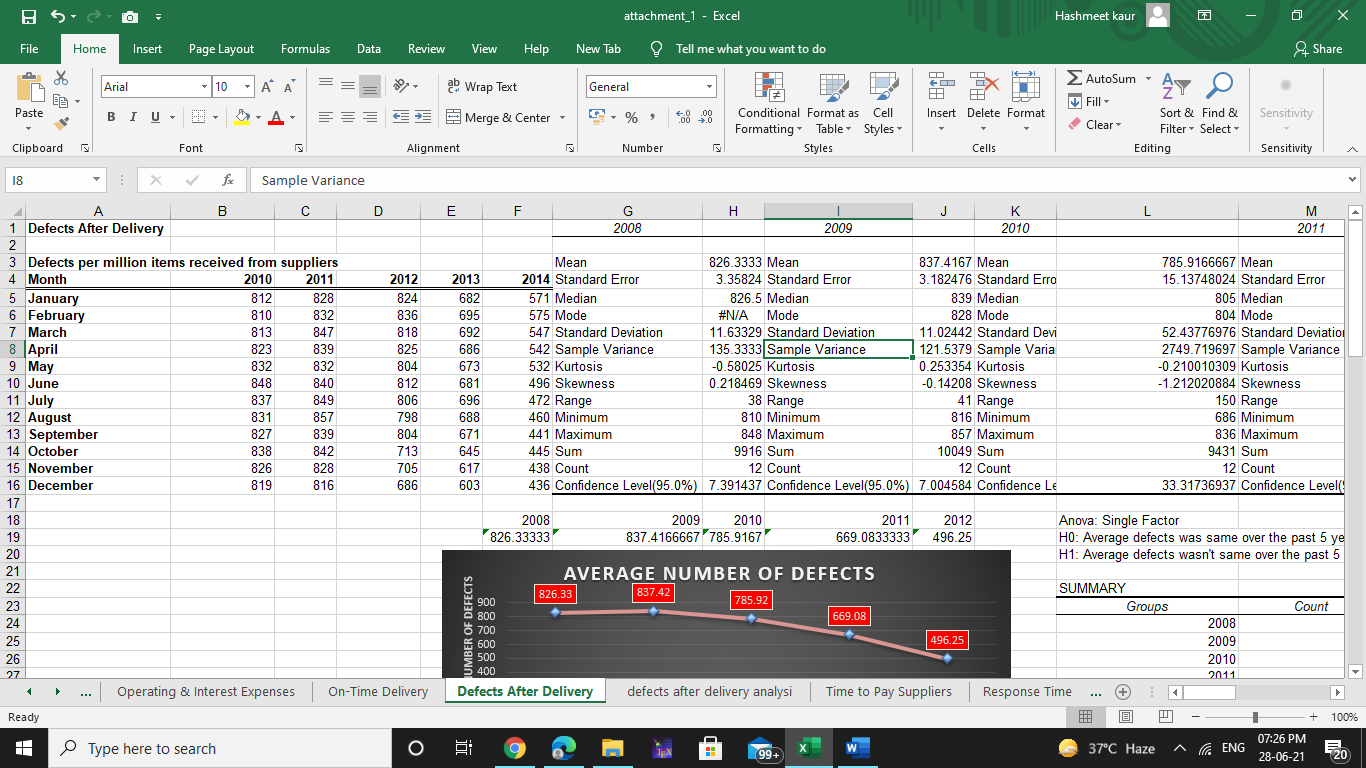
**BUSINESS OBJECTIVES (ANALYTICAL RESEARCH)**

* **Upon investigation, PLE has experienced some quality problems due to an increasing number of defects in materials received from suppliers.** **The company instituted an initiative in August 2011 to work with suppliers to reduce these defects, to more closely coordinate deliveries, and to improve materials quality through reengineering supplier production policies. So we’ll predict what might have happened which made supplier initiative not been implemented and how the number of defects might further be reduced in the near future.**
* **Have the defects in the data has changed significantly over the past 5 years.**
* **Have the number of defects reduced to 600.**

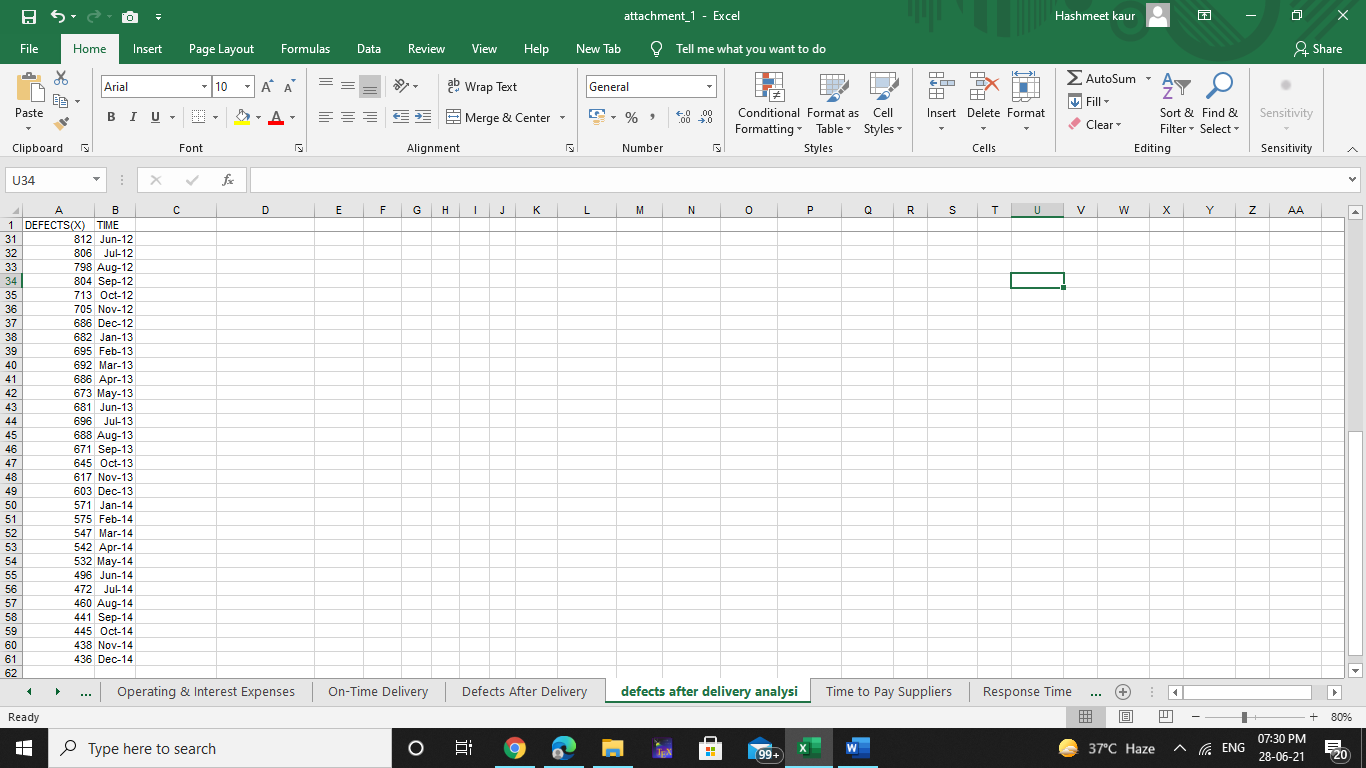
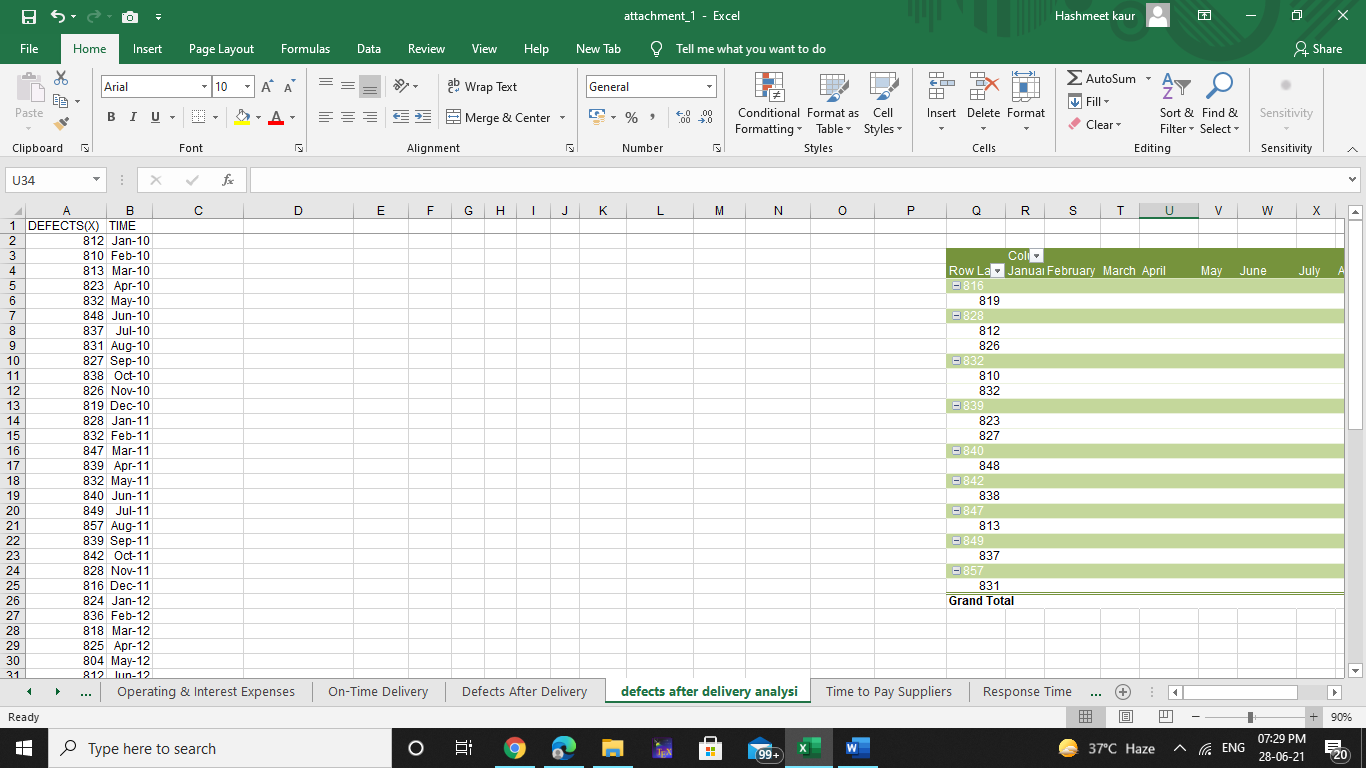
**HYPOTHESIS TESTING (NULL AND ALTERNATE)**

* **By using regression analysis we assume Null Hypothesis (H0): Defects don’t change with the time and Alternate Hypothesis (H1): Defects changed with the time**
* **By using ANNOVA Single Factor we assume Null Hypothesis (H0): Average defects was same over the past 5 years and Alternate Hypothesis (H1): Average Defects wasn’t same over the past 5 years similarly with T-Test: Two sample assuming unequal variance we assume Null Hypothesis (H0): Average Defects wasn’t same in 2010 and 2014 and Alternate Hypothesis (H1): Average Defects was same in 2010 and 2014 respectively.**
* **As an investigator we would like number of defects to be reduced so we’ll perform ONE SAMPLE T-Test where we Null Hypothesis(H0): Average number of defects are greater than or equal to 600 and Alternate Hypothesis(H1): Average number of defects are less than 600**

**DATA**



**This data shows the number of defects initiated from suppliers over the time period from 2010 to 2014.**

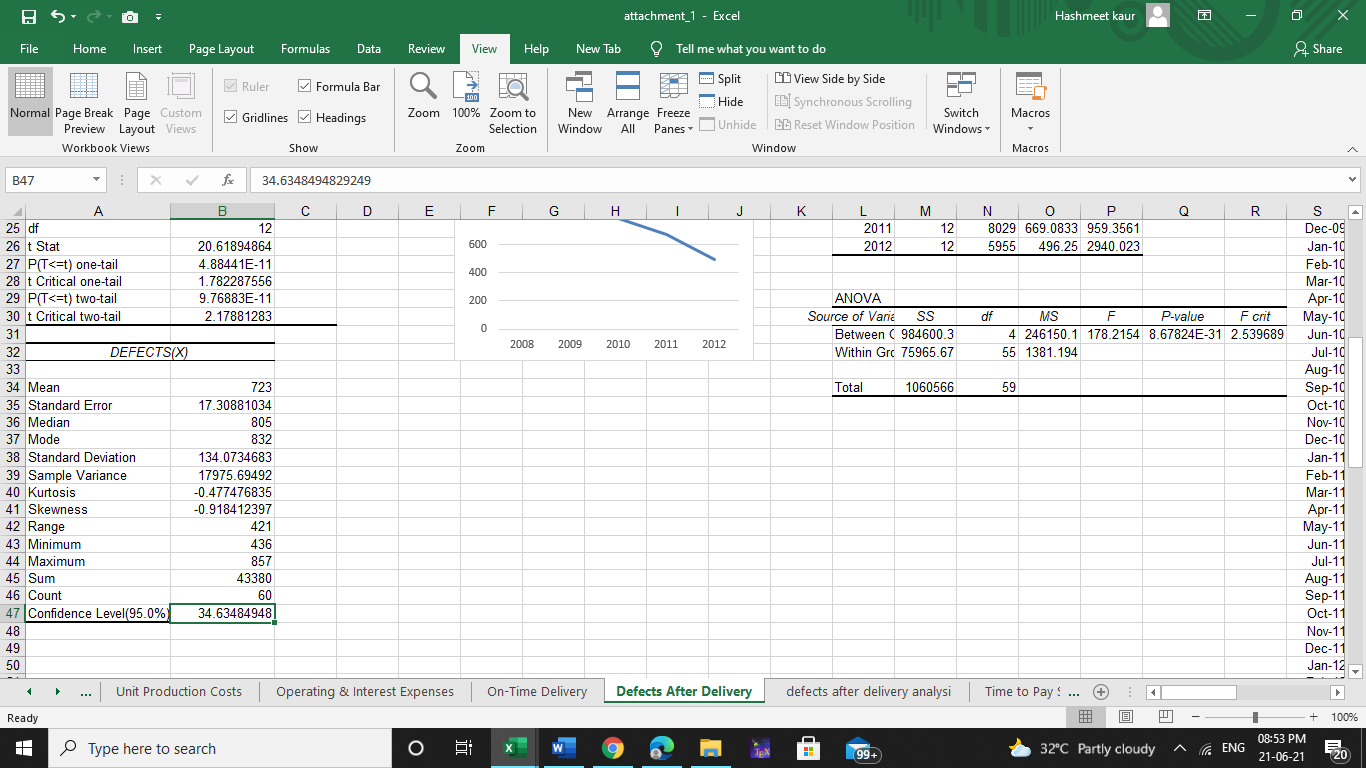


**Sample size: 60 samples (if the data is transformed into month and year wise)**

**Time Period: From January 2010 to December 2014**

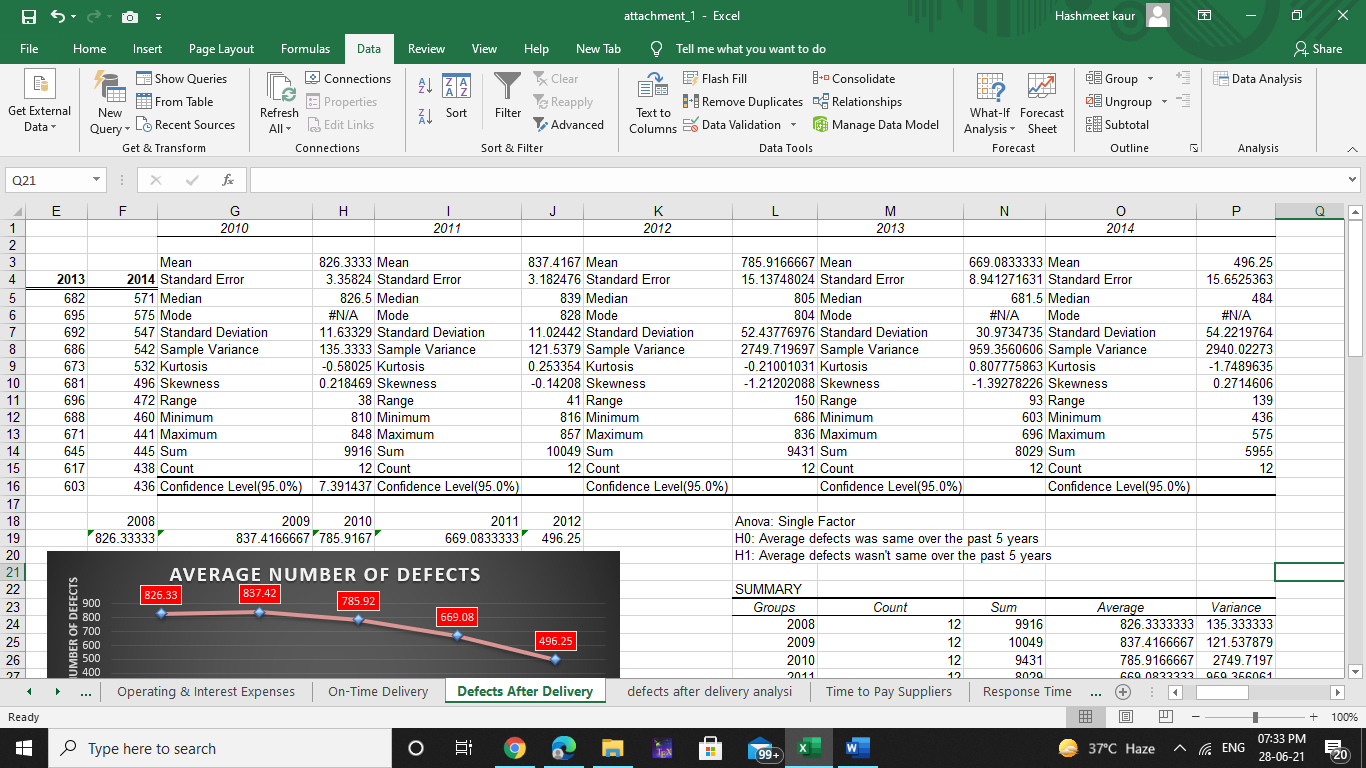
**STATISTICAL TECHNIQUES**

* **DESCRIPTIVE STATISTICS**
* **OVERALL ANALYSIS**



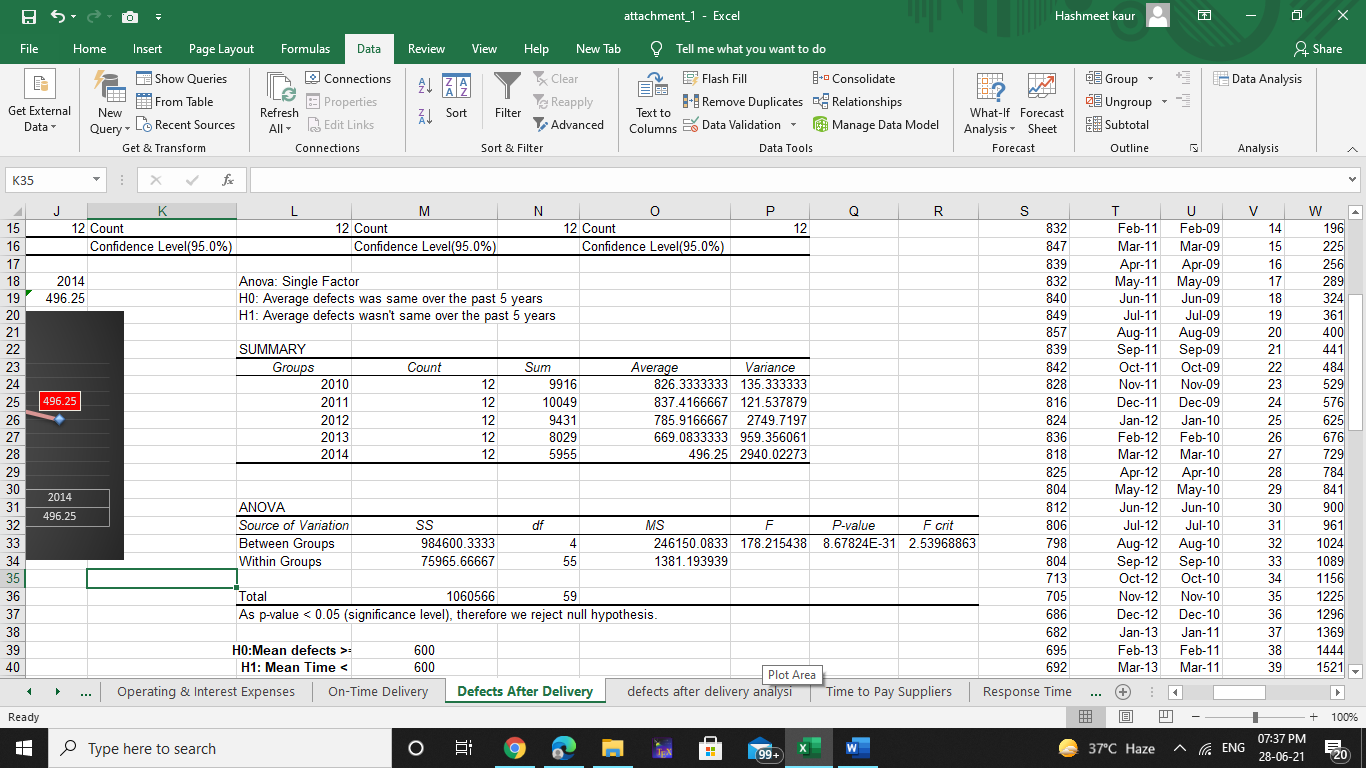
Overall analysis shows that average number of defects over the past 5 years is 723, with sample variance approximate 17,975. Least number of defects were 436 which was increased to 857 with a range of 421.

* **YEAR WISE ANALYSIS**

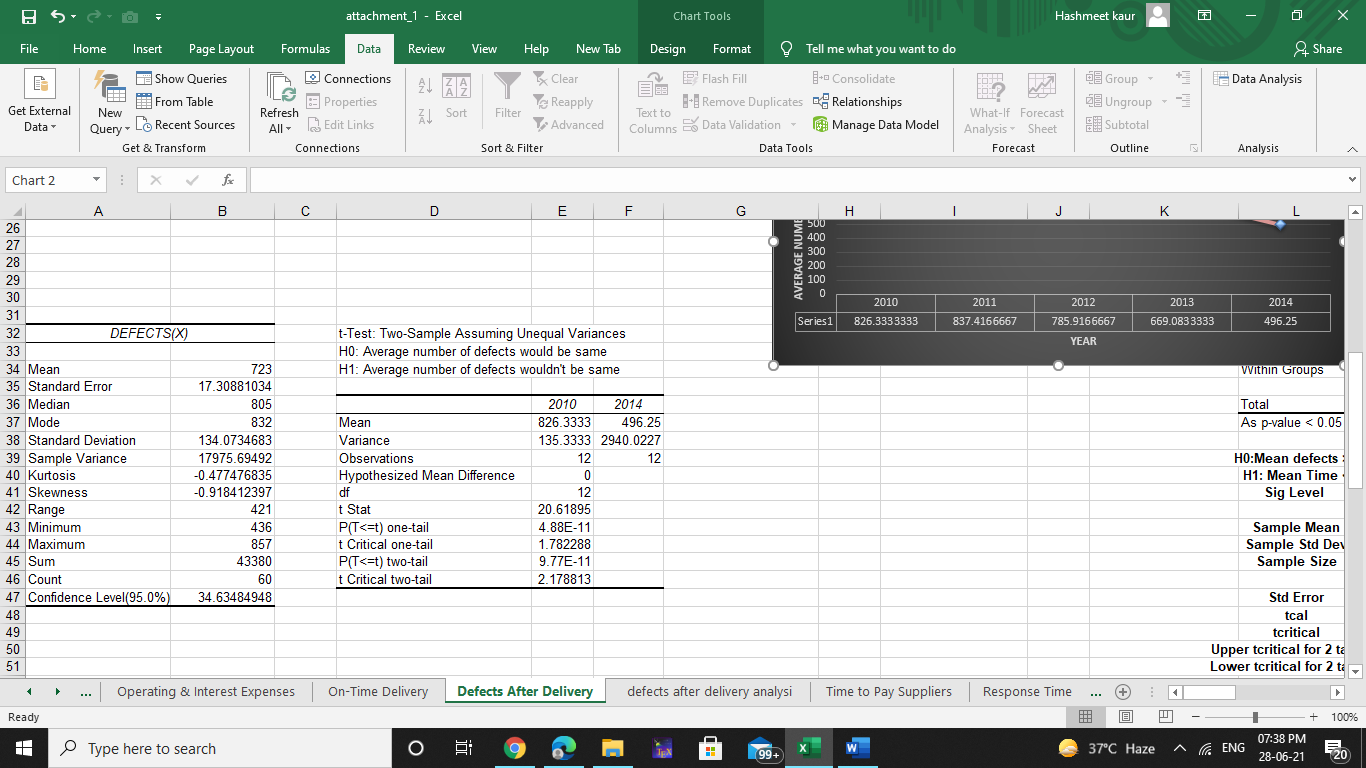


Year wise analysis shows how average number of defects has decreased between 2010 to 2014.

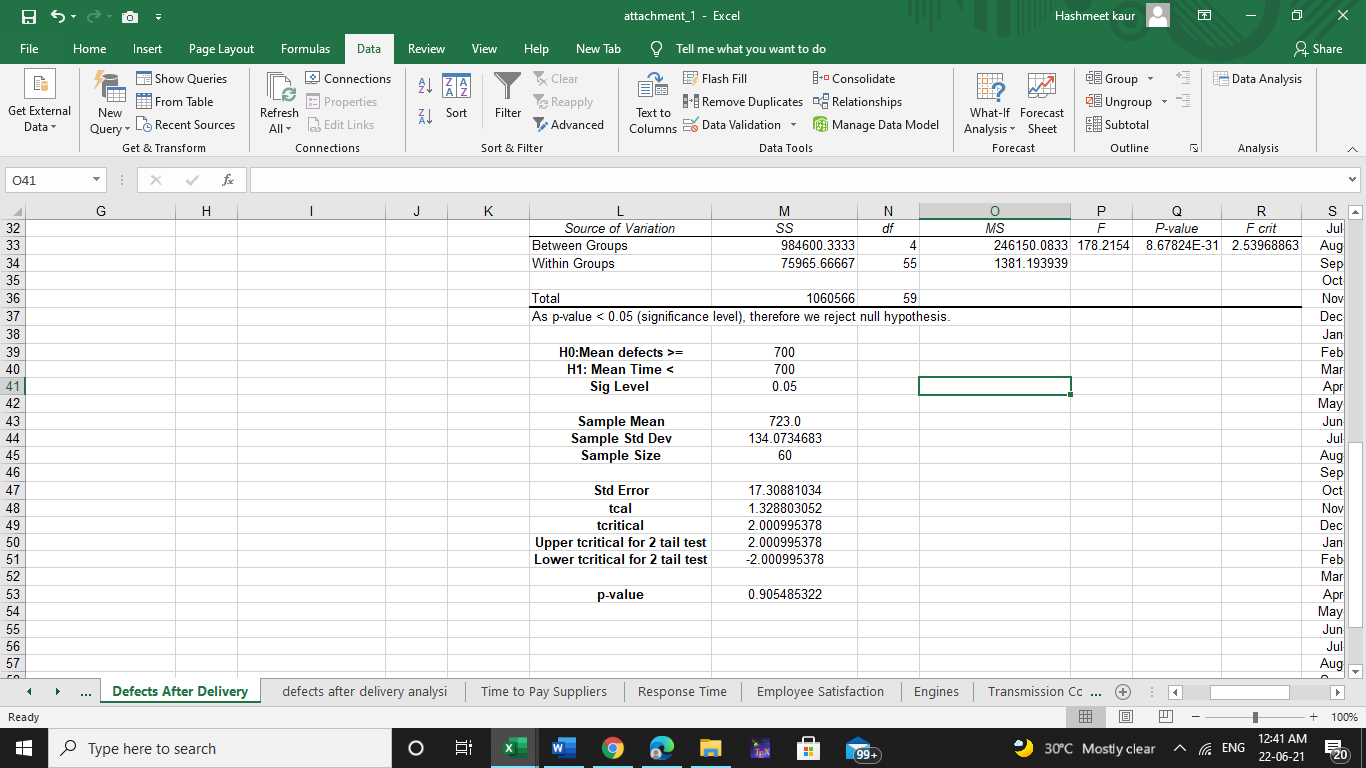
* **INFERENTIAL STATISTICS**
* **ANNOVA-SINGLE FACTOR**



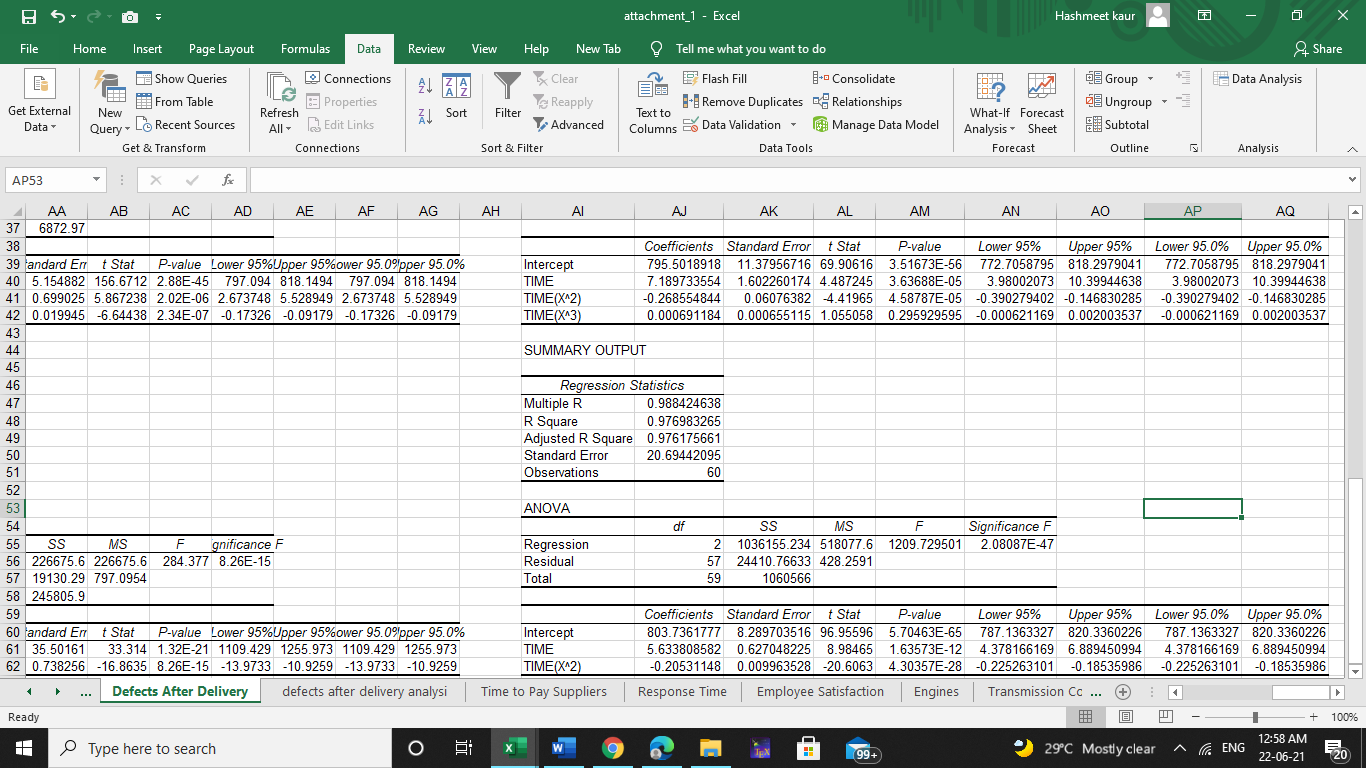
* **T-Test: Two sample assuming unequal variance**



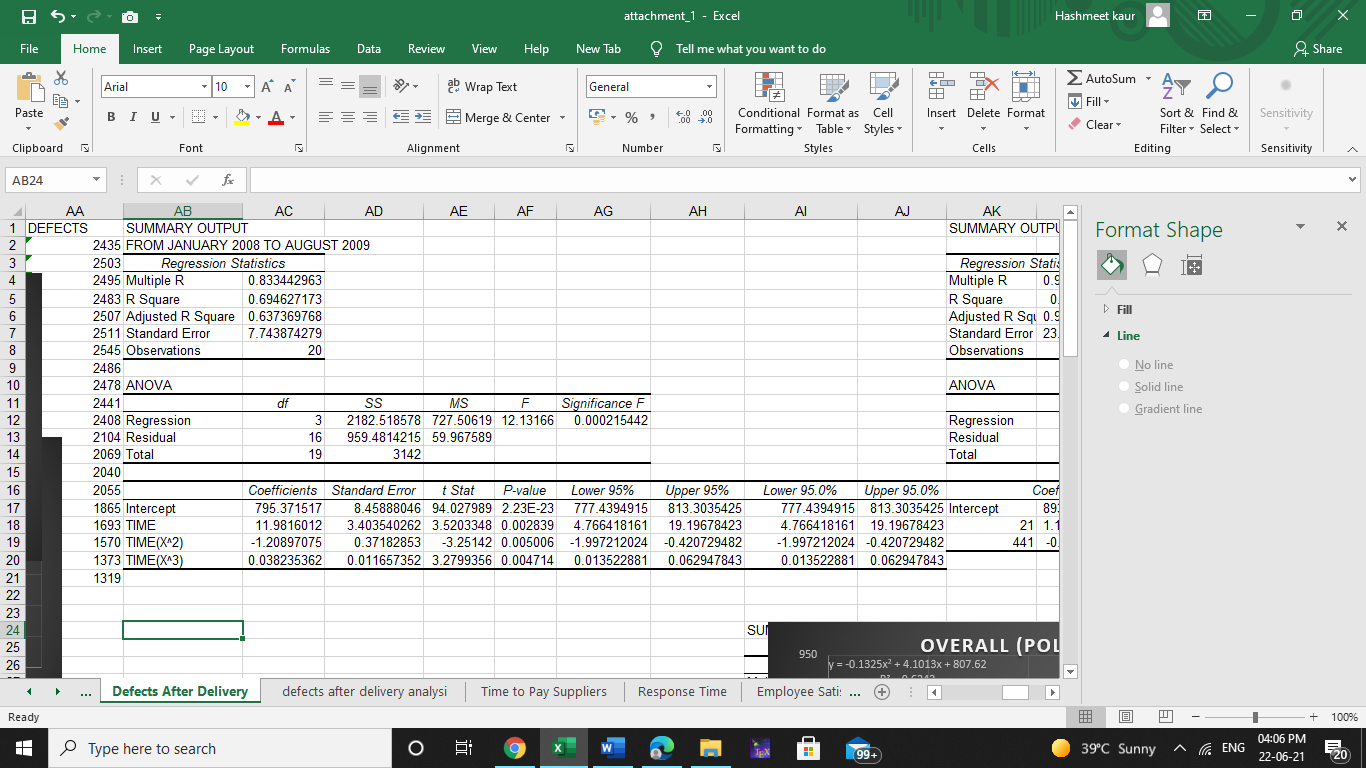
* **One sample T- test**



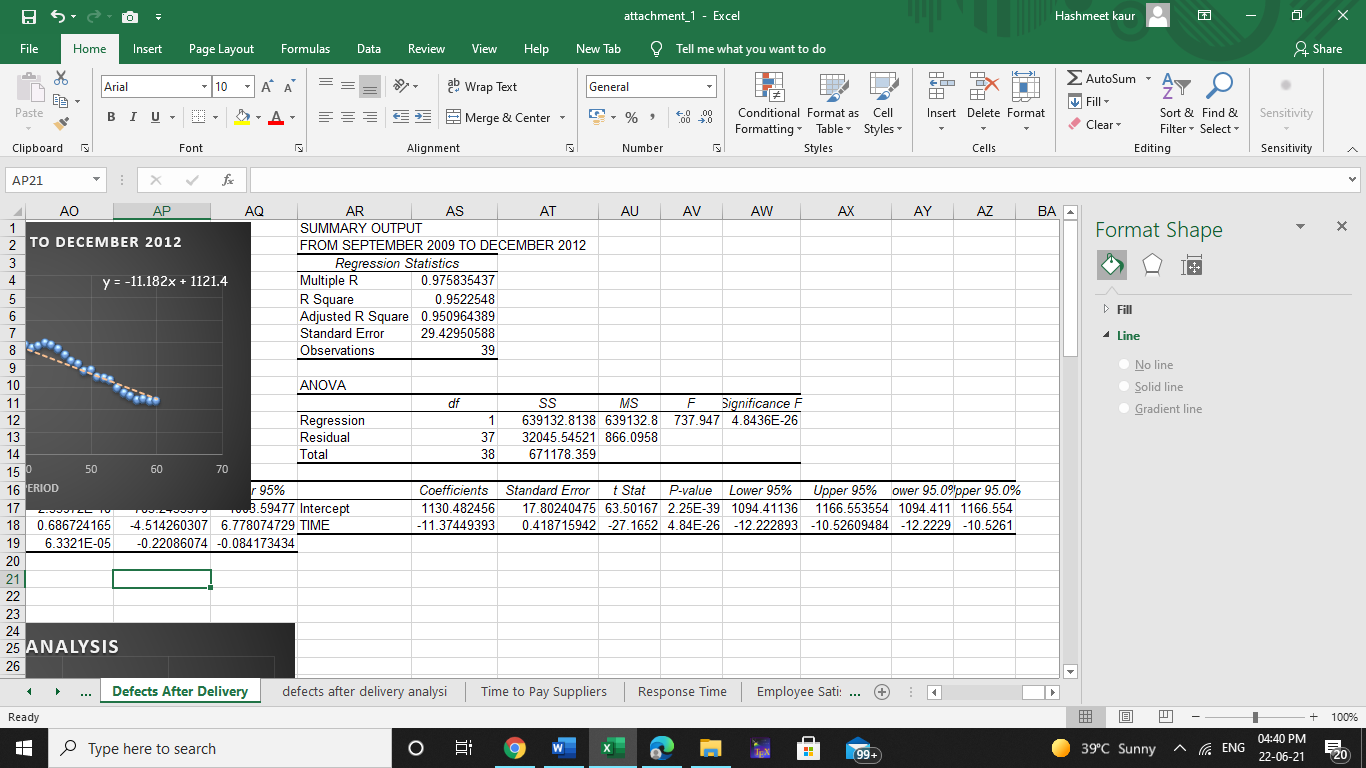
* **REGRESSION ANALYSIS**
* **OVERALL ANALYSIS**



* **FROM JANUARY 2010 TO AUGUST 2011**



* **FROM SEPTEMBER 2011 TO DECEMBER 2014**



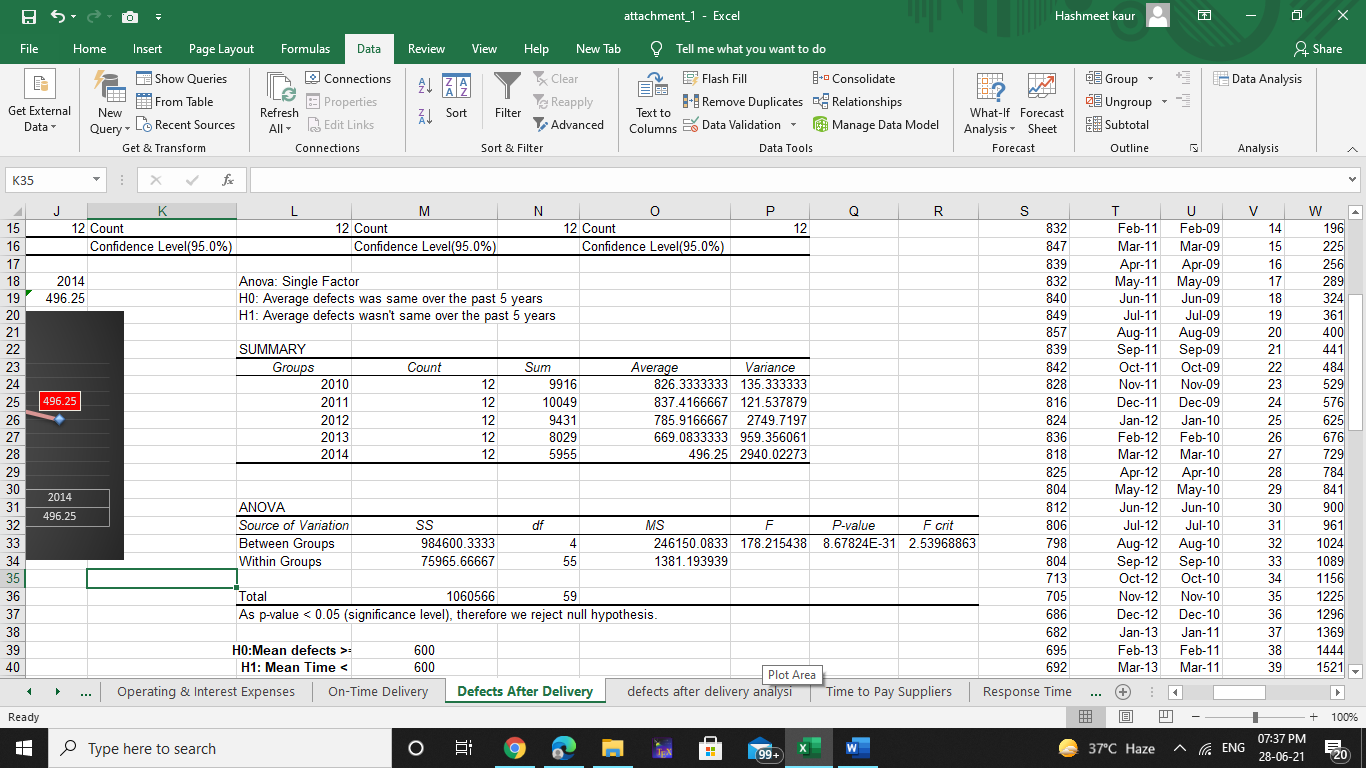
**RESULTS AND DISCUSSION**

1. **Average number of defects**

As p-value < 0.05 (significance level), therefore we reject Null Hypothesis.

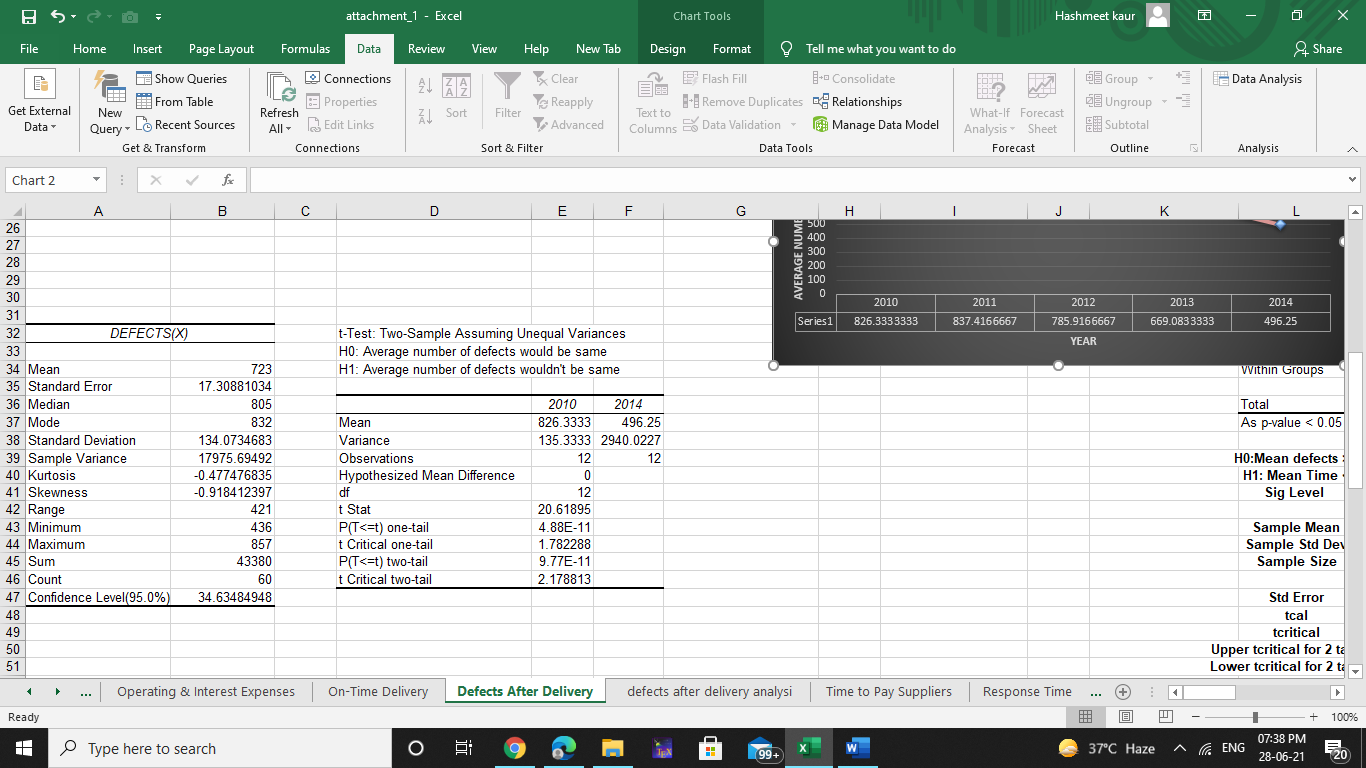
Therefore, average defects wasn’t same over the past 5 years.

1. **ANNOVA SINGLE FACTOR**



From the graph its clear that average number of defects has decreased over the past 5 years starting from 826 to 496 defects.

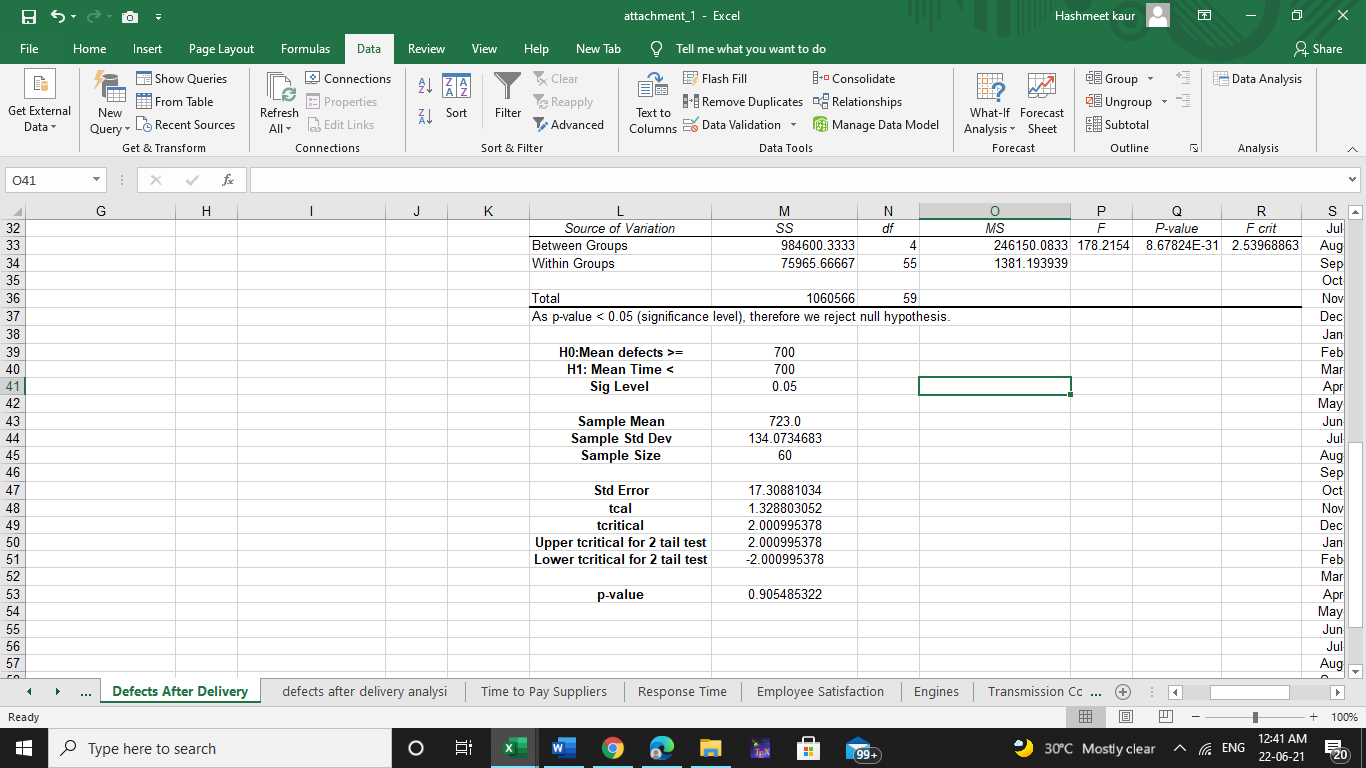
1. **TWO SAMPLE T-TEST ASSUMING UNEQUAL VARIANCE**



1. **ONE SAMPLE T-TEST**

As p-value < 0.05 (significance level), therefore we reject Null Hypothesis.

Therefore, average defects wasn’t same in 2008 and 2012.



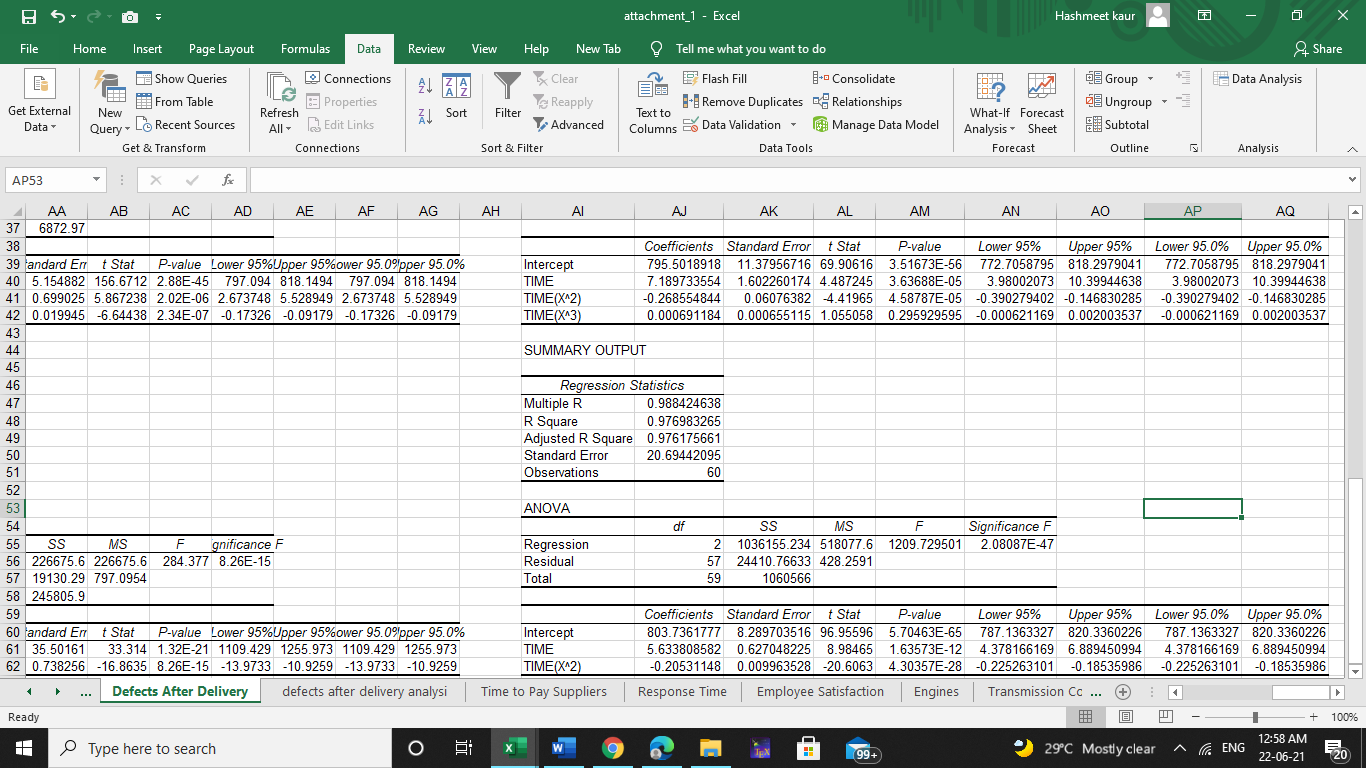
As p-value > 0.05 (significance level), therefore we accept the Null Hypothesis.

Therefore, its clear that number of defects has

Increased.

1. **REGRESSION ANALYSIS**

* **OVERALL ANALYSIS (POLYNOMIAL REGRESSION)**



**If we try to analyse the graph we can split the time period from January 2010 to August 2011 and September 2011 to December 2014 by using polynomial regression analysis.**

Trend line equations shows the number of defects is related to time as follows.

Number of defects = 803.73 + 5.633\*Time – 0.2053\*(Time^2)

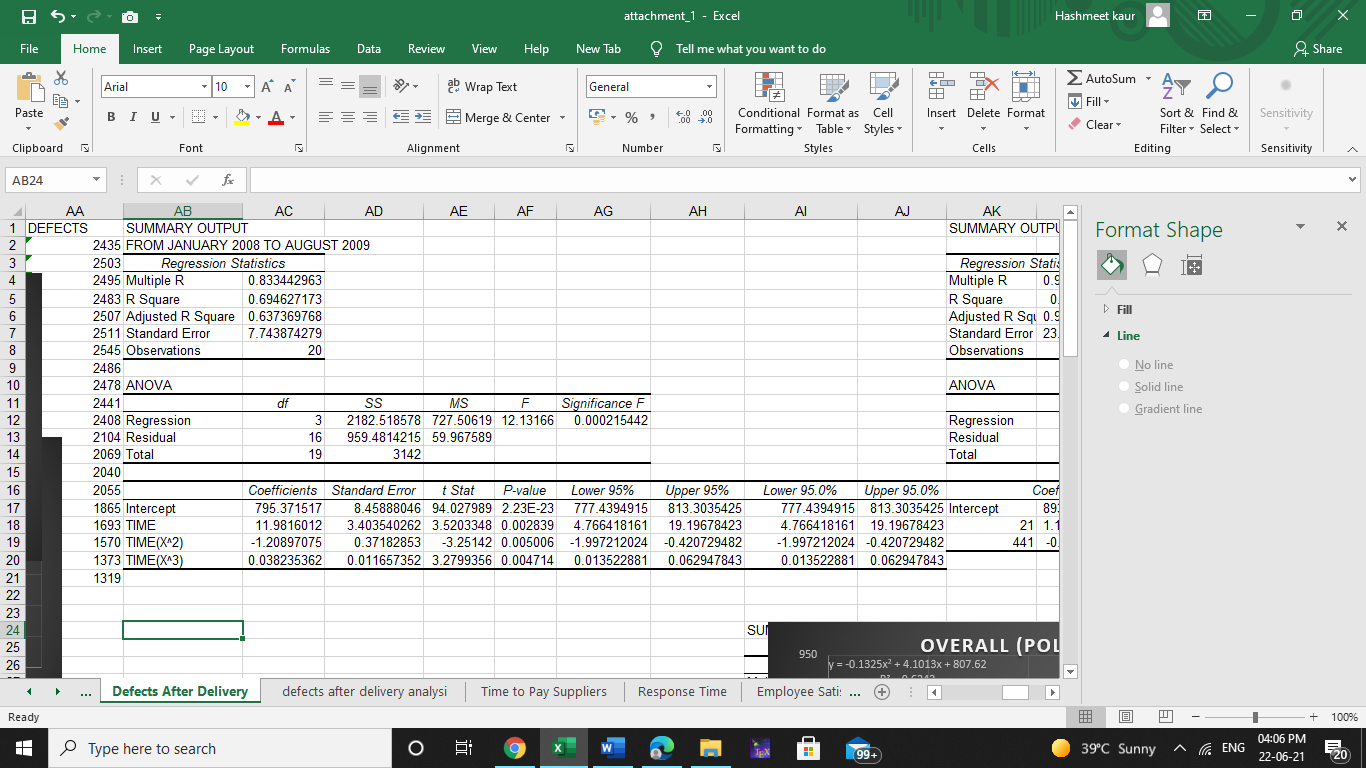
NOTE: Time has been incremented by months starting from January 2010.

H0: Time and Time^2 are not significant variables, therefore defects doesn’t changes (decrease) with time.

H1: Time and Time^2 are significant variables, therefore defects changes (decrease) with time.

From the regression analysis its concluded that p-value for Time :1.63E-12 and Time^2 :4.30E-28 are very small than 0.05 (alpha). So we reject the Null Hypothesis and conclude that defects got decreased with the time.

* **FROM JANUARY 2010 TO AUGUST 2011**



Trend line equations shows the number of defects is related to time as follows.

Number of defects = 795.37 + 11.98\*Time – 1.20\*(Time^2) + 0.038\*(Time^2)

NOTE: Time has been incremented by months starting from January 2010.

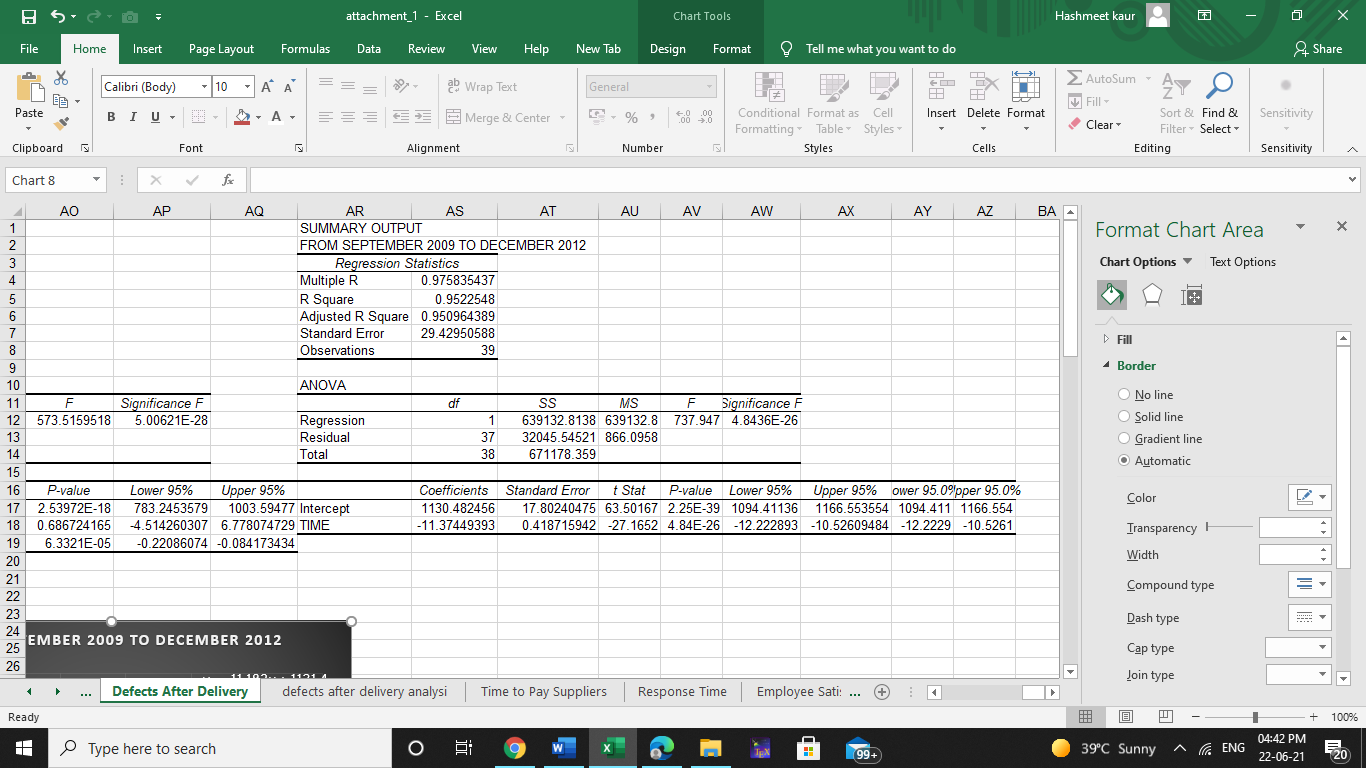
H0: Time, Time^2 and Time^3 are not significant variables, therefore defects doesn’t changes (increase) with time.

H1: Time, Time^2 and Time^3 are significant variables, therefore defects changes (increase) with time.

From the polynomial regression analysis its concluded that p-value for Time: 0.00283,

Time^2: 0.00500 and Time^3: 0.00471 are very small than 0.05 (alpha). So we reject the Null Hypothesis and conclude that defects got increase with the time. On the otherhand R square =0.62, which means 62% of the increase in defects received with the time, (which indicates quality problem increasing over the time), meaning if no initiative is taken then the number of defects will increase over the time

* **FROM SEPTEMBER 2011 TO DECEMBER 2014**



Trend line equations shows the number of defects is related to time as follows.

Number of defects = 1130.48 - 11.37\*Time

NOTE: Time has been incremented by months starting from January 2010.

H0: Time are not significant variables, therefore defects doesn’t changes (decrease) with time.

H1: Time are significant variables, therefore defects changes (decrease) with time.

From the linear regression analysis its concluded that p-value for Time: 4.84E-26 is very small than 0.05 (alpha). So we reject the Null Hypothesis and conclude that defects got decreased with the time. On the other hand R square =0.95, which means 95% of the decrease in defects received with the time, (which indicates that the quality initiative taken during August 2009 helped in reducing the defects.

**IMPLICATIONS AND MANGERIAL RECOMMENDATIONS**

1. **As PLE was experienced some quality problems due to which some initiatives were taken in August 2011, through regression analysis its concluded that defects got increased but after the initiatives there can be seen some decrease in defects with increase in samples.**
2. **Taking time(year) as a consideration there is a huge variation in average number of defects, this shows that their initiative was helpful.**
3. **But this doesn’t guarantee us that in the coming year number of defects will not increase until an unless we come to know which factor actually caused the increase in defects, incomplete data can also be the reason.**
4. **So, its recommended that company should bring some changes in the coming year so that we can come to know through regression analysis that which factor actually cause increase in defects, as increase in samples doesn’t give a clear idea, otherwise predicting defects in future won’t be possible.**
5. **After the initiative, company’s assumption having reduction in defects might decrease to 700, which was rejected, even though there is not a huge difference in average number of defects which is approximately 723.**