A Statistical Report on Cultivation of Paddy in two districts of Tamil Nadu.

Madurai district and Cuddalore district,
For a period of Seventeen years,
From 1997-2013.

By
Abhishek Narayanan & Aditya K R
01FB16ECS016, 01FB16ECS028

TAMIL NADU


TAMIL NADU
DISTRICT MAP
N
ANDHRA PRADESH
KARNATAKA
KERALA
TIRUVALLUR
Tiruvallur
CHENNAI
Vellore
VELLORE
Kanchipuram
KANCHIPURAM
KRISHNAGIRI
Krishnagiri
TIRUVANNAMALAI
Tiruvannamalai
DHARMAPURI
Dharmapuri
Viluppuram
PUDUCHERRY
SALEM
VILUPPURAM
Cuddalore
Bay of Bengal
Salem
ERODE
NILGIRIS
CUDDALORE
Erode
NAMAKKAL
PERAMBALUR
Udagamandalam (Ooty)
Namakkal
Perambalur
ARIYALUR
NAGAPPATTINAM
Coimbatore
Ariyalur
KARAIKAL (PUDUCHERRY)
TIRUPPUR
TIRUCHCHIRAPPALLI
Karur
KARUR
THANJAVUR
COIMBATORE
Nagapattinam
Tiruppur
Tiruchchirappalli
Thanjavur
Thiruvarur
TIRUVARUR
PUDUKKOTTAI
DINDIGUL
NAGAPATTINAM
Dindigul
Pudukkottai
LEGEND
State Boundary
District Boundary
State Capital
District HQ
Theni
Madurai
Sivaganga
THENI
MADURAI
SIVAGANGA
Palk Strait
Virudhunagar
Ramanathapuram
VIRUDHUNAGAR
RAMANATHAPURAM
THOOTHUKUDI
Tirunelveli
Thoothukudi
Gulf of Mannar
SRI LANKA
TIRUNELVELI
KANYAKUMARI
Nagercoil
Map not to Scale
INDIAN OCEAN
Copyright © 2013 www.mapsofworld.com
(Updated on 3rd June, 2013)

Madurai district & Cuddalore district

Madurai District

One of the thirty two districts in the Indian State of Tamil nadu, Madurai is located in the Torrid zone just as most of Tamil nadu. It experiences a largly tropical type of climate with an average temperature of 29.8^{0}C. It is a fertile district with rich nutrients such as Potassium and Phosphorous in larger content.

The district is fed by a seasonal river called Vaigai.

Quick Climate Info

Hottest Month	May (33 °C avg)
Coldest Month	January (26 °C avg)
Wettest Month	October (144.2 mm avg)
Windiest Month	July (22 km/h avg)
Annual Rainfall	697.2 mm (per year)

This climate with adequate rains during the retreating monsoon season is largely favorable for agriculture. Many crops such as cotton, groundnuts, sugarcane and rice are

cultivated in this region. This report is based on statistcal analysis of cultivation of rice for a period of seventeen years (1997-2013).

Cuddalore District

Located in the coastal region of Tamil Nadu, Cuddalore experiences tropical climate with greater rainfall than Madurai. Being a rural district, large areas of land are under cultivation.

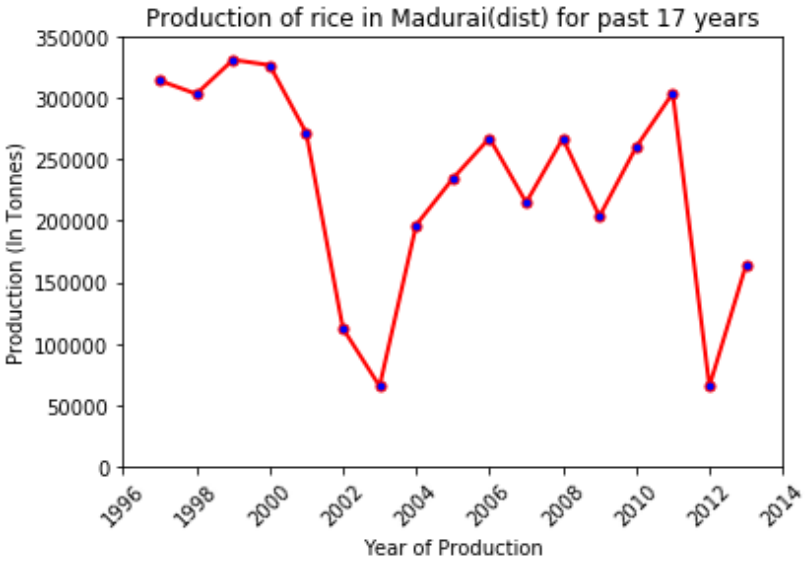
Quick Climate Info

Hottest Month	May (32 °C avg)
Coldest Month	January (26 °C avg)
Wettest Month	November (363.2 mm avg)
Windiest Month	May (4 km/h avg)
Annual Rainfall	1204.1 mm (per year)

Rice is one of the major crop grown in this region. Let us now look into the analy sis of production of paddy in these two districts.

STATISTICAL ANALYSIS

LINE GRAPH SHOWING PRODUCTION OF RICE IN A PERIOD OF SEVENTEEN YEARS
MADURAI DISTRICT



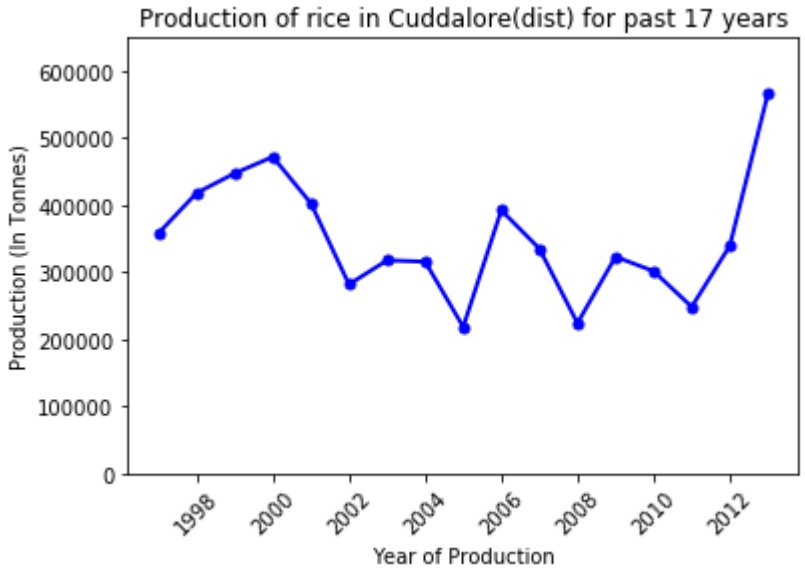
X axis: Defines the year of production
Y axis: Defines the quantitof production in Metric Tonnes.

Inference: From the graph it can be infered that the cultivation of rice in the district of Madurai has been a less stable scenario. The curve is fluctuating, indicating that the produce every year is differing from that of the previous one, by a noticeable range expect for the starting years.

From the year 1997 to 2000 we can see that the variation in production is negligible when compared with the rest of the data. The steep curve starting in 2002 indicates the sudden fall in the production of rice atleast by a quantity of 1.5lakh tonnes. It is also observable that the production recovered with a steep rise in the year 2004. The period from 2004 to 2010 experienced continuous fluctuations which later rose higher and then dropped in the year 2012 to an all time low, just as in the year 2003, only to recover a little in the year 2013.

- Total Production in a period of 17 years is **39,02,567** Metric Tonnes
- Average production of Rice in Madurai per Annum over a period of 17 years is **2, 29,562** Metric Tonnes
- Highest production is in the year 1999, **3,31,342** Metric Tonnes.
- Lowest production is in the year 2012, **65,000** Metric Tonnes.

LINE GRAPH SHOWING PRODUCTION OF RICE IN A PERIOD OF SEVENTEEN YEARS CUDDALORE DISTRICT



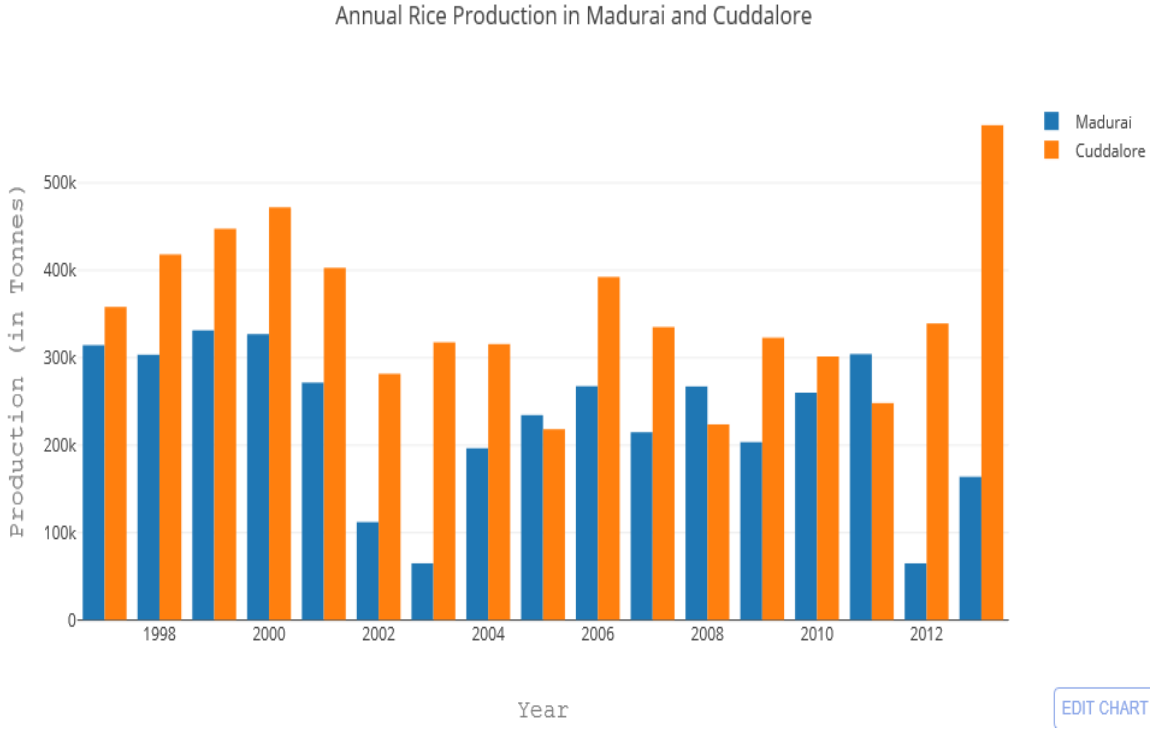
X axis: Defines the year of production
Y axix: Defines the quantitof production in Metric Tonnes.

Inference: This graph describes the production of rice in the district of Cuddalore. It is clearly indicated that most of the distribution in the graph is above the a nnual average of production in Madurai.

The production increased steadily from the year 1997 to 2000 after which there was a drop in the production in the year 2001, which further dropped in the year 2002. With slight increase in the next two consegutive years it further dropped to an all time lo w, slightly more than 2.0lakh tonnes. It later increased steeply in the following year with a slow drop in the next two years, to rise again in 2009. It
rose to an all time high in the year 2013 after a short dipression in the span of two years.

- Total Production in a period of 17 years is **59,61,367** Metric Tonnes
- Average production of Rice in Cuddalore over a period of 17 years is **3,50,668** Metric Tonnes
- Highest production is in the year 2013, **5,65,960** Metric Tonnes.
- Lowest production is in the year 2005, **2,18,345** Metric Tonnes.

COMPARITIVE ANALYSIS OF PRODUCTION OF RICE IN MADURAI AND CUDDALORE



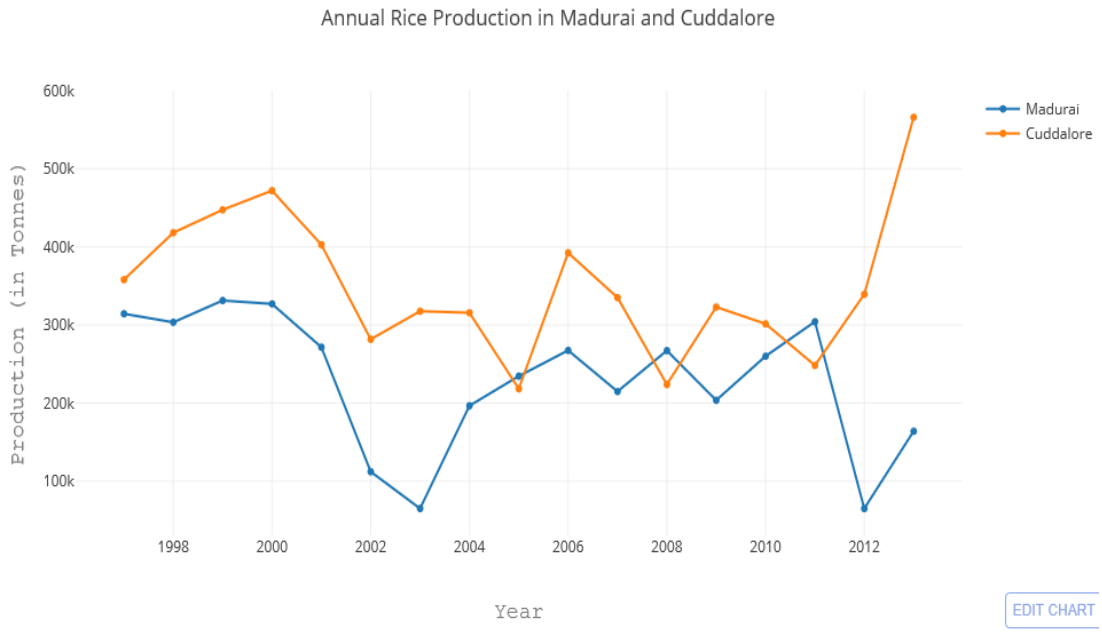
X axis: Defines the year of production

Y axix: Defines the quantitof production in Metric Tonnes.

Inference: The categorical Bar chart compares the production of rice in the two d istricts for a period of 17 years. It is clearly noticeable that the produce in Cuddalore is greater than that of madurai in most cases. Cuddalore's production remains high inspite of the depression in production in Madurai for the year 2013.It has outbeaten Madurai in production by over three times in the same year and in 2003.

The maximum produce in Cuddalore is more than 5.5lakh tonnes, whereas it is only a little more than 3.0lakh tonnes in Madurai, for a period of 17 years.

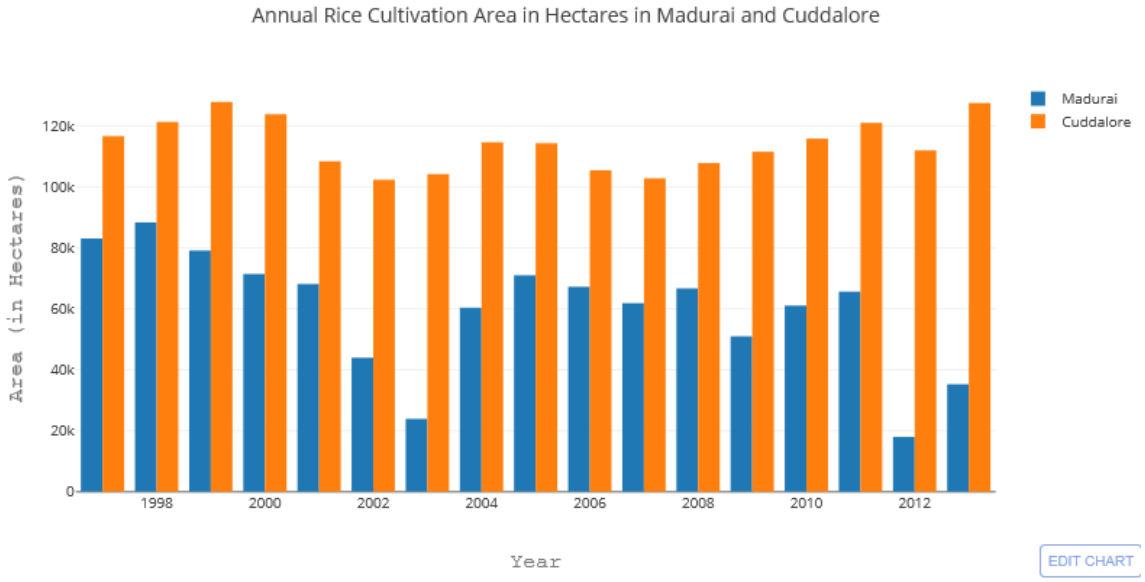
The categorical line graph gives addition information from the data available whi ch otherwise would go un-noticed.



The line plot above distinctly shows three different scenarios where the production of rice in Madurai was greater than that of Cuddalore in the corresponding y ear. Cuddalore's production dipped below that of Madurai in 2005, 2008 and in 2011, which is not easily noticeable in the bar chart.

It is also worthwhile to reacon the fact that the production graph of Cudd-alore l ies entirely above that of Madurai except for these three cases.

COMPARITIVE ANALYSIS OF AREA UNDER CULTIVATION IN MADURAI & CUDDALORE



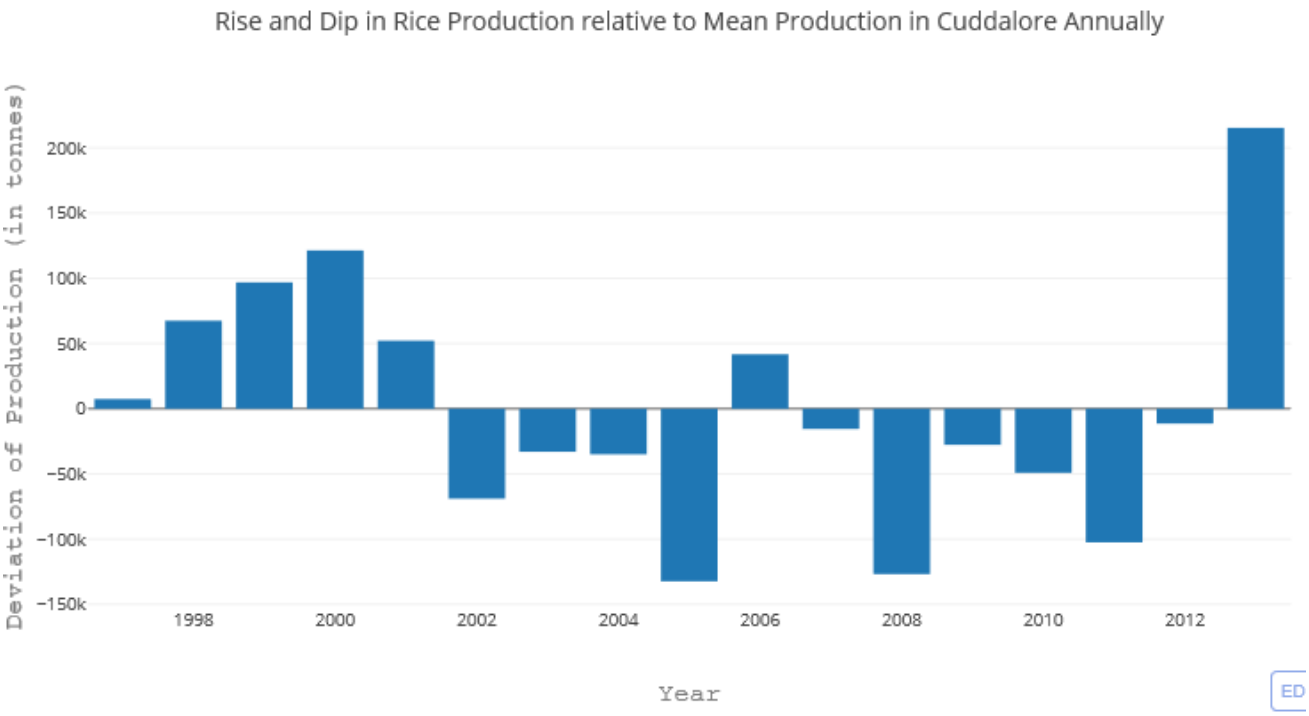
It is interesting to note that the greater yeild in Cuddalore is of no wonder considering the fact that it had larger areas under cultivation than Madurai, which in all cases is greater than 1.0lakh hectares. Whereas Madurai being a semi Urban area had 80,000 hectares under cultivation in only one case (1998) and the rest lower.

- Largest area under cultivation in Cuddalore is in the year 1999, **1,27,843** Hectares.

- Lowest area under cultivation in Cuddalore is in the year 2002, **1,02,336** Metric Tonnes.
- Largest area under cultivation in Madurai is in the year 1998, **88,338** Hectares.
- Lowest area under cultivation in Cuddalore is in the year 2012, **18,024** Metric Tonnes.

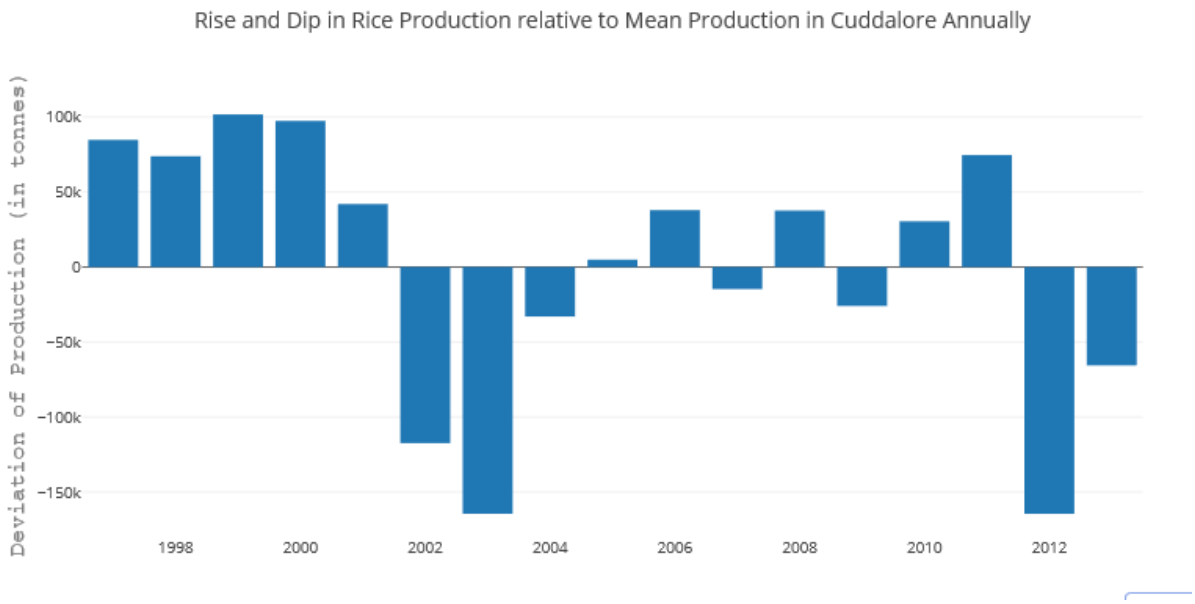
By comparing the production graph and the area graph it is justified that the low production of rice in Madurai in years (2003 and 2012) is primarily due to the fact that the area under cultivation was extremely low when compared to that of its elf in other years and to that of Cuddalore.

Analysis of Deviation in Production from Annual Average for Cuddalore



The graph above shows the deviation of production from Annual average f-or C uddalore. The highest positive deviation is in the year 2013, with deviation ov-er 2lakh Metric Tonnes, and the highest negative deviation in the year 2005 with Production deviating about 1.3lakh Metric Tonnes from Mean Annual Average. It can also be observed that during the periods from 2002-2005 and from 2007-2011, the production was much lesser than the Annual Production. The span 1 of crop failure was due to labour protests and the span 2 of crop failure was due to excessive rainfall. (correlation coefficient being negative).

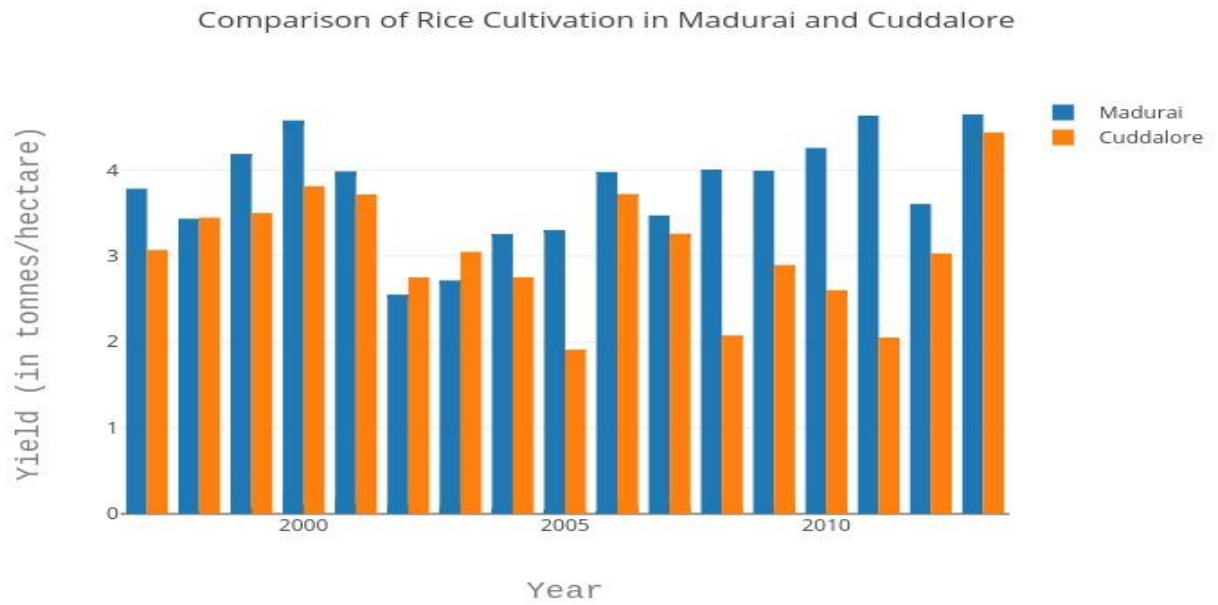
Analysis of Deviation in Production from Annual Average for Madurai



The graph above shows the deviation of production from Annual average f-or M adurai. The highest positive deviation is in the year 1999, with deviation of ov-er 75 tho usand Metric Tonnes from the Mean Annual Average, and the highest negative deviatio n is in the year 2003 and in 2012 with 165 thousand Metric Tonnes of negative deviating from Annual average.

It can also be observed that during the period from 2002-2004 the production w as much lesser than the Annual Production. This failure of crop was due to the statewid e labour protests leading to low production.

Comparitive Analysis of Productivity (Yield in Tonnes per Hectare) in Madurai and Cuddalore



Interestingly, unlike what was expected, this graph shows us a trend which is worth noting. Inspite of Cuddalore having greater production in the period of 17 years except for a few hitches, the graph tells us a different story. Previously we have discovered that the area under cultivation in Cuddalore is far greater than that of Madurai. Which means there is enough justification for Madurai's produce bei-ng lesser than Cuddalore's. However this might make us overlook the chance of Madurai's productivity being higher, which this graph highlights.

In all the years, except for 2002 & 2003 we see a clear trend of productivity in madurai being higher. Which means, the Yield in Metric Tonnes per Hectare in Madurai is larger by a considerable margin than that of cuddalore. There might be various reasons for madurai's productivity being higher than Cuddalore's. The prime factor here is the fertility of soil.

These are the key nutrients required for satisfactory yeilds of rice Ph and K.

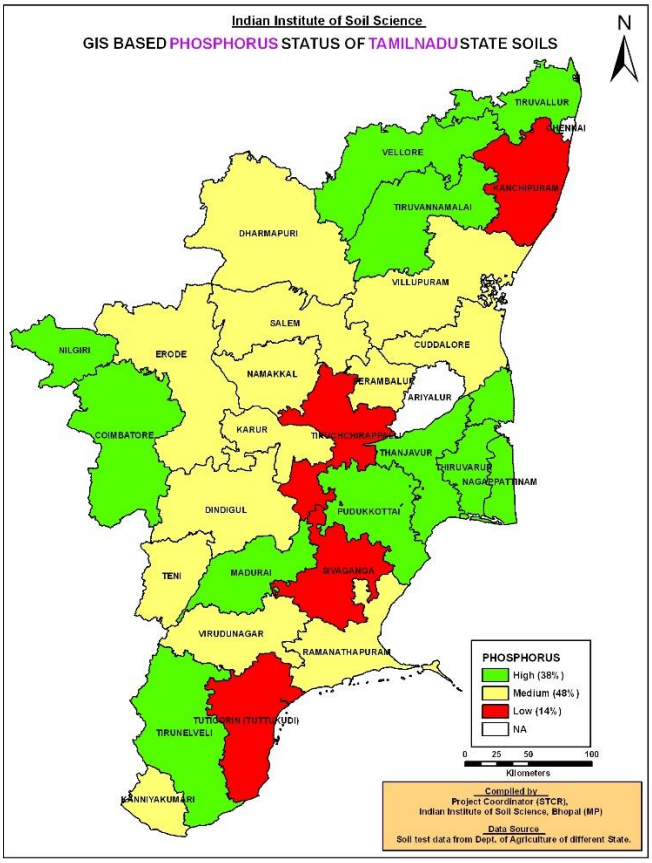
Table 6.1: Minimum soil levels of key nutrients necessary for satisfactory rice yields

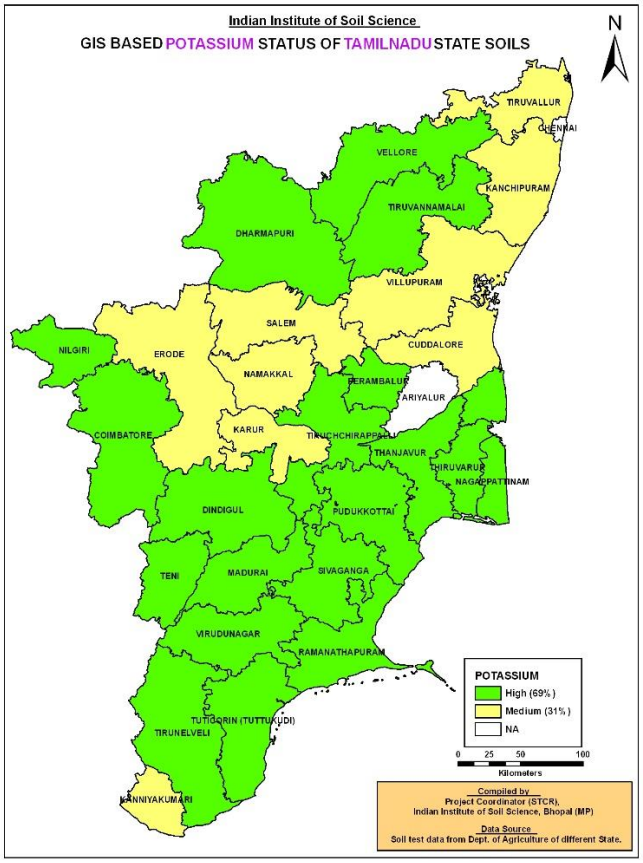
Element	Soil test method	Critical value
Phosphorus	$NaHCO_3$	6 ppm PO_4-P(In cold years, may be as high as 9 ppm)
Potassium	NH_4Ac	60 ppm K
Zinc	DTPA	0.5 ppm Zn

Field research has established the critical concentrations of P, K, and Zn in the soil by correlating soil test values of these mineral nutrients and rice plant performance. The presence of critical or subcritical values of a given nutrient will generally result in a reduction of 10% or more of growth and grain yield. When soil test values are near or below the critical level, application of the deficient nutrient will usually prevent this reduction. Therefore, proper soil sampling, chemical analysis and interpretation help avoid plant nutrient deficiencies and are valuable guides in soil fertility management.

The first map here describes the phosphorous content in each of the districts of Tamil Nadu. It is to be noted that Madurai has greater percentage of phoshorous in its soil than cuddalore. Ph being an importnt nutrient for rice cultivation is found in lesser percentage in Cuddalore which has directlt affected the yeild of rice

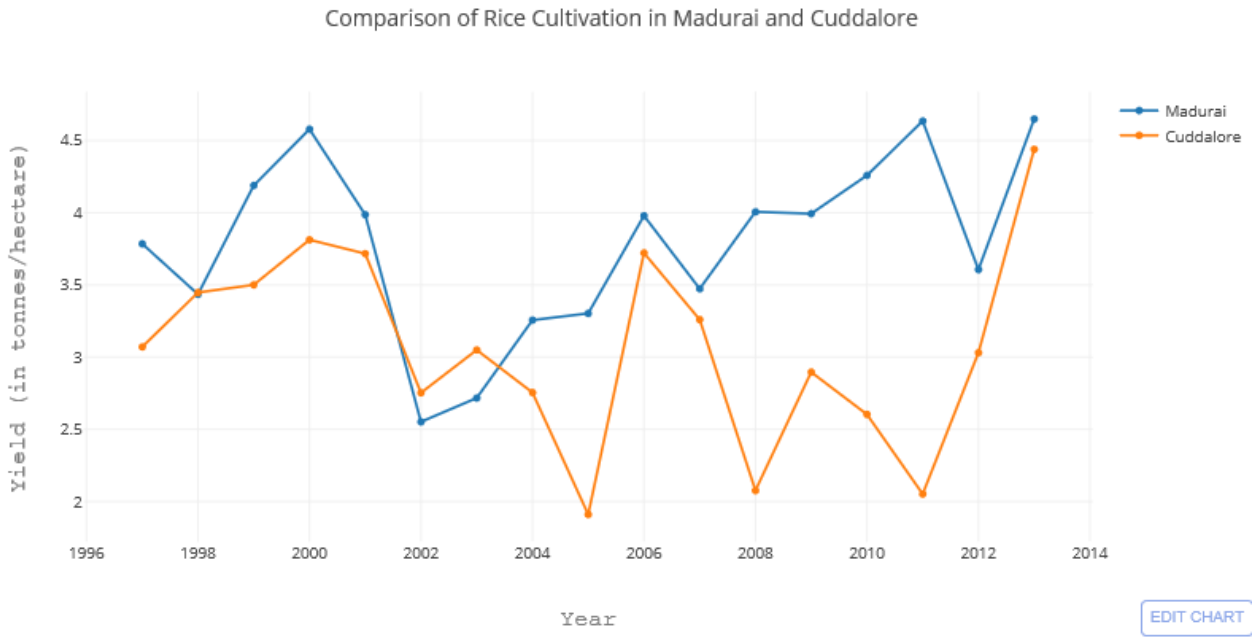
The graph on the left also justifies the greater productivity in Madurai with the argument that it has larger potassium content in its soil than Cuddalore.



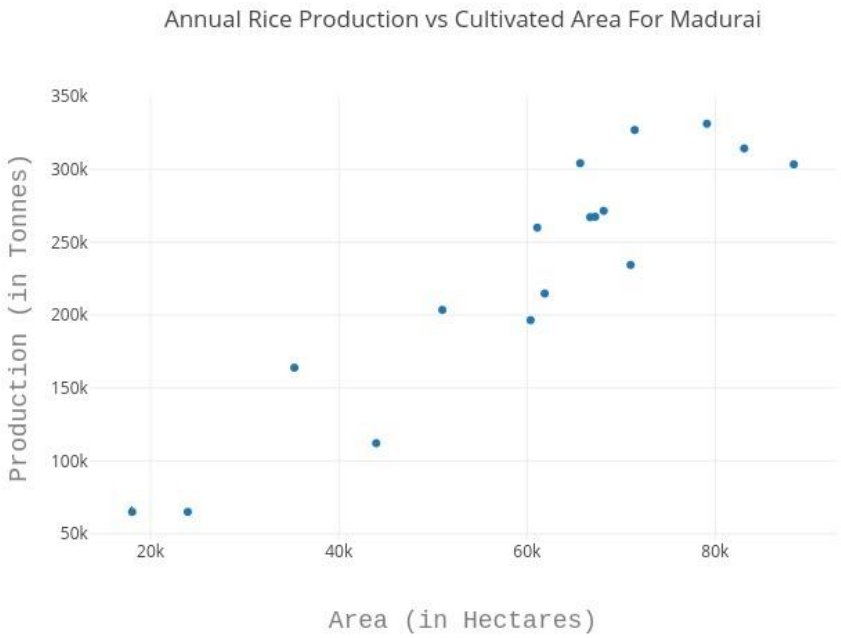


Thus we can arrive at a conclusion that higher fertility of soil is the main cause fo r greater productivity in Madiurai. Greater potential of Madurai's fertail land can be unlo cked if larger areas of land are brought under cultivation. In such a scenario Madurai mig ht outbeat cuddalore in its production.

The line plot at the bottom highlights the dip in Madurai's rice roductivity for a s pan of two years.

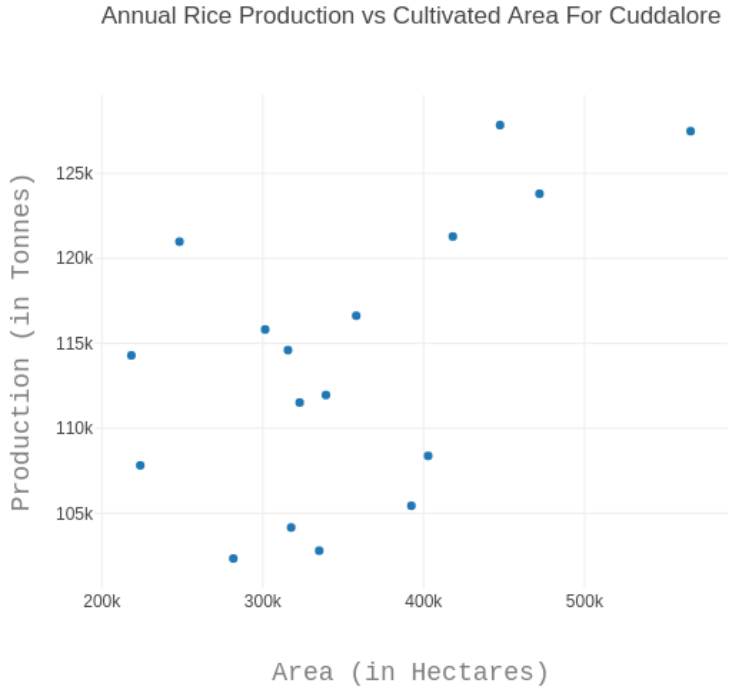


Analysis of Optimum area of cultivation for Maximum Yield for Madurai



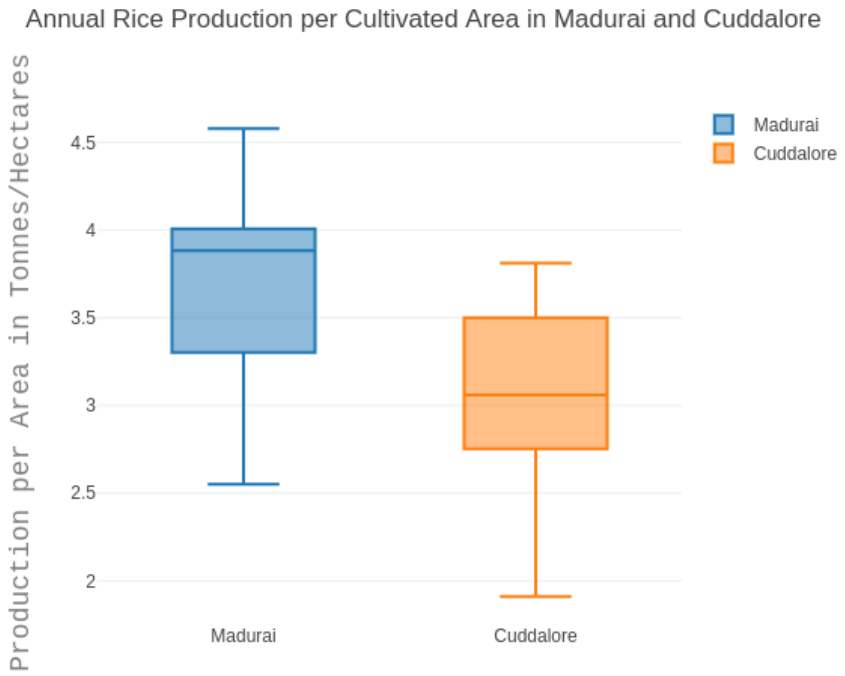
The figure portrays the scatter plot of production vs area in Madurai. It is clearly visualized that the points almost follow a linear trend, and therefore a best fit straight line may be plotted throgh the points, whose slope (=production/area) would indicate the yield of rice in the district for a particular year. It is observed that if such a line were to be fit through the application of linear regression, then the yield (=slope of the line of best fit) increases with area annually. From this plot, we in fer the values of areas of cultivated land for which production is optimum.
Such an optimum range for area is 60,000 to 80,000 hectares.

Analysis of Optimum area of cultivation for Maximum Yield for Cuddalore



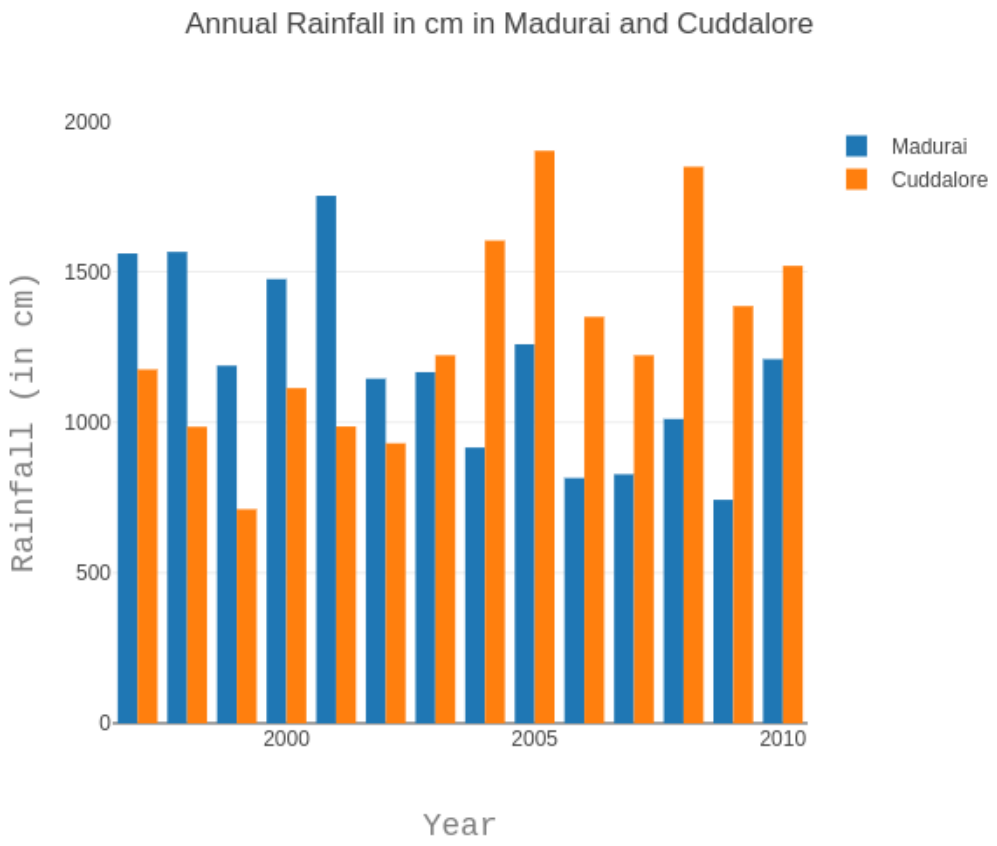
The figure portrays the scatter plot of production vs area in Cuddalore. It c-an be clearly visualized that the points follow no linear trend, and therefore perhaps a polyno mial curve may be plotted through the points, whose slope (=production/area) would in dicate the yield of rice in the district for a particular year. It is difficult to observe a clear trend in the yield values without the application of polynomial regression. From this plot however, we infer the values of areas of cultivated land for which production is optimum . Such an optimum range for area is 4,00,000 to
500,000 hectares. Thus a clear trend in the graph is not observable.

Analysis of Rice Production per cultivated area for Madurai and Cuddalore



The above graph refects the Box Plot for yield (=production/area). It is obviously inferred from the plot that quite a lot of yield values are higher for madurai
than for cuddalore since the plot of madurai extends above that of cuddalore. Since the l ower whiskers of these plots are longer than upper whiskers, the yield data for both of t he districts is inferred to be left or negatively skewed, i.e. mean<median<mode. The ske wed nature of the yield data is further confirmed by the cumulative
distribution function plot as shown which has a longer left tail.

The inference obtained from this skewness is that over all of the years, there hav e been more years with less rice yield as compared to those years with prosperous yied a nd production.



The initial years analysed show a good production in Madurai. This is shown to b e weakly correlated to the high rainfall during the period and the correlation coefficient i s found to be 0.36. The later years are seen to obtain moderate or decent amount of rai nfall which may be one of the fators determining the production but does not clearly ex plain the reason for certain drastic fluctuations in production as no such trend is seen in the rainfall graph.
On the other hand, Cuddalore has been observed to recieve moderate rainfall initially, s howing a steep increase in the later years. However excessive rainfall has been found to have an adverse effect on the production in the recent years as suggested by the extrem ely negative coefficient of correlation which has a value of -0.76.

Abstract :

Madurai and Cuddalore are two semi-urban districts in Tamil Nadu which have b een prospering in rice production and are one of the chief contributers to
the net production as obtained from the state of Tamil Nadu. In this report, we have pre sented a detaied analysis of rice cultivation and production in the two districts from the year 2002 to 2014. We conducted a detailed analysis of the area cultivated and producti on of rice in these areas and have made an attempt to predict the factors or parameters which have deep infuence on the trends of rice yield and production observed through v

arious vizualization means. The detaied analysis report along with the vizualizations and inferences concluded from the trends observed have been explained in detail in further s ections of this report.

Description of the Dataset Used :

The dataset for the purpose of this analysis has been collected from the Indian G overnment databases via the data.gov.in website. The dataset possesses annual details a bout cultivated area and production for a variety of food and cash crops for all districs in each State across India cultivated both in rabi and kharif seasons.
The data obtained has been arranged in .csv (comma separated values) format and only the fields required in our analysis, i.e., the data about the two districts, Madurai and Cud dalore in Tamil Nadu have been extracted. Subsequently, we perform vizualization and a nalysis of this data for a period of 17 years from 2002-2014. The details about the results and inferences have been summarized as follows.

INTRODUCTION TO DATA SCIENCE ASSIGNMENT -II

HYPOTHESIS TESTING

We assume two populations, consisting of yield of rice (in tonnes/hectare) of Madurai and Cuddalore in the state of Tamil Nadu, India.We obtained a sample comprising of yield data for 17 years starting from 1997 to 2014.

We wish to determine whether yield of rice in the district of Madurai is greater than that in Cuddalore at 5% significance level and therefore set the Hypothesis as follows :

$$H_0 : \mu X - \mu Y <= 0 \quad \text{versus} \quad H_1 : \mu X - \mu Y > 0$$

Where μX and μY are the mean yield of rice from the population containeing the data for all years for Madurai and Cuddalore Districts respectively.

From the sample data we obtain the following results :
For Madurai :

sample mean for Madurai ($\bar{X}$) =3.7885 tonnes/hectare

sample standard deviation sx =0.61658

number of samples n=17

For Cuddalore :

sample mean for Cuddalore ($\bar{Y}$) =3.06442 tonnes/hectare

sample standard deviation sy =0.68243

number of samples n=17

Null distribution for the test considered is $\bar{X} - \bar{Y} \sim N(0, 0.04976)$

Since sampe size is small (n=17) and standard deviation of population is unknown, we use T-Test.

T statistic for our hypothesis has been calculated to be 3.24598

From the student's t-table, we find that $0.005 < P(t>3.24598) < 0.001$

Hence p-value which is $P(t>3.24598)$ is less than 0.05 which is the α-value at 5% significance level.

Hence we can reject H0, i.e., our Alternate Hypothesis is true and we conclude that the yield of rice in tonnes/hectare is greater in Madurai than in Cuddalore.

WILCOXON RANK SUM TEST OR MANN WHITNEY TEST :

For conducting the Wilcoxon Rank Sum Test, our sample is a large one as size is greater than 8.

Therefore m=n=17

We used the Wilcoxon rank-sum test to test:

$H_0 : \mu_X <= \mu_Y$ versus $H_1 : \mu_X > \mu_Y$

Where μX and μY are the mean yield of rice from the population containing the data for all years for Madurai and Cuddalore Districts respectively.

When both sample sizes m and n are greater than 8, it can be shown by advanced methods that the null distribution of the test statistic W is approximately normal:

$W \sim N(m(m+n+1)/2 , mn(m+n+1)/12)$

Test is performed by computing the z-score of W , and then using the normal table to find the
P-value.

Combining the data of the two samples, assigning ranks to them and adding the ranks corresponding to Madurai, we get the value of W as 379.

We computed the z-score for the above W value as 2.807 or approximately 2.81.

$P(z>2.81) = 0.0025$ from the standard normal distribution table, which is less than our α-value at 5% significance level (0.05). Hence we reject our null Hypothesis and our Alternate Hypothesis is concluded to be True.

Therefore we arrive at the conclusion that the yield of rice in Madurai in tonnes/hectare is gretaer for the distrct of Madurai than that of Cuddalore, both in the state of Tamil Nadu.

Conclusion:

Having taken great pleasure in completing this Assignment, we would like to work on many such interesting tasks in the future.

Abhishek Narayanan & Aditya K R.