APPLIED ENGINEERING STATISTICS

IE6210

Analysis of the Nutrients from the fertilizers consumed by the various crops

Final Project Report

Submitted by

Chaitanya Dixit (gy2283)

Rajratna Patil (gv2149)

Sudhanshu Authankar (gy2984)

Sharayu Madrewar (gx5538)

Abstract

This report is based on project (Analysis of the nutrients from the fertilizers consumed by the various crops) of Applied Engineering statistics IE 6210.

The survey is taken to estimate plant nutrients of Nitrogen, Phosphate and Potash. National agriculture statistical service has a set of data from 1971-2010. Nutrient contents are considered from different states of USA mainly corn, soybean, cotton, wheat. These nutrients are measured in 1000 nutrient short tons. Data Gathered from economic research service, department of agriculture. Data maintained by Richard Nehring.

In this project we have collected the data of nutrients consumed by the crops, analyzing the data with the different tools we have learned on Minitab. We will develop graphical representation of the data. We will analyze mean of phosphate and potash content are different or not. We will analyze the variance of potash and phosphate nutrients of the various crops like corn, wheat, soybean, cotton.

In summary, we will use following methods to analyze the data like graphs (box-plots, histogram), test of hypothesis, regression (linear, multiple) and ANOVA (one factor, two factor)

Keywords: *Nitrogen, Phosphate, Potash, graphs, hypothesis, regression, ANOVA*

Index

Figures

Abbreviations

Nitrogen Corn- Nitrogen Contain of Corn Nitrogen Cotton- Nitrogen Contain of Cotton Nitrogen Soybean- Nitrogen Contain of Soybean Nitrogen Wheat- Nitrogen Contain of Wheat Nitrogen Other- Nitrogen Contain of Other	Phosphate Corn- Phosphate Contain of Corn Phosphate Cotton- Phosphate Contain of Cotton Phosphate Soybean- Phosphate Contain of Soybean Phosphate Wheat- Phosphate Contain of Wheat Phosphate Other- Phosphate Contain of Other	Potash Corn- Potash Contain of Corn Potash Cotton- Potash Contain of Cotton Potash Soybean- Potash Contain of Soybean Potash Wheat- Potash Contain of Wheat Potash Other- Potash Contain of Other

Problem Statement:

The survey is taken to estimate plant nutrients of Nitrogen ,Phosphate and Potash. National agriculture statistical service has a set of data from 1971-2010. Nutrient contents are considered from different states of USA mainly corn, soybean, cotton, wheat. These nutrients are measured in 1000 nutrient short tons.

1. To check normality of the data of the Nutrients consumed by each crop, by various graphical means like boxplot (to checking symmetry about median) histogram (to check the symmetry about mean) and probability plot
2. To get better quality of crop and to secure neutrality of soil the potash contain which increases alkanity of soil and phosphate contain which increases the acidity of soil should be equal so check the hypothesis of the mean of potash and phosphate contain of the various crops
3. While preparation any fertilizers manufacturer always defines a specific amount of Phosphate and potash after that Nitrogen contain is added to maintain concentration.so find the regression analysis of the data being nitrogen as dependent variable and Phosphate and potash are independent variable.
4. To Find the effect of individual Plants and individual fertilizer content (Factors) on the Responses and to find if there is any effect of interaction between the plant and fertilizer content (Factor interaction) on the responses.

Data

Nutrients consumed by the crops by means of fertilizer

All values in 1000 short ton

Year	*Nitrogen Corn*	Phosphate Corn	Potash Corn	*Nitrogen Cotton*	Phosphate Cotton	Potash Cotton	*Nitrogen Soybeans*	Phosphate Soybeans	Potash Soybeans	*Nitrogen Wheat*	Phosphate Wheat	Potash Wheat	*Nitrogen Other*	Phosphate Other	Potash Other
1971	3,730	2,024	1,946	343	164	140	62	229	282	614	375	136	3,385	2,012	1,728
1972	3,705	1,994	1,992	404	212	175	72	285	370	783	447	157	3,057	1,926	1,633
1973	3,830	1,988	2,052	337	182	151	95	380	498	896	507	181	3,137	2,028	1,767
1974	3,773	2,102	2,361	421	210	173	87	301	404	1,078	621	263	3,798	1,864	1,882
1975	3,885	1,963	2,162	240	102	86	74	273	376	1,085	564	275	3,317	1,605	1,553
1976	5,210	2,550	2,771	353	160	121	70	296	452	1,456	744	312	3,322	1,478	1,553
1977	5,181	2,523	2,835	416	185	110	113	438	602	1,279	647	294	3,658	1,837	1,993
1978	4,888	2,416	2,646	351	163	112	138	510	722	1,047	439	179	3,542	1,569	1,866
1979	5,274	2,499	2,803	352	168	83	149	624	933	1,253	597	276	3,686	1,718	2,149
1980	5,244	2,413	2,927	371	160	100	137	563	881	1,570	677	291	4,085	1,618	2,046
1981	5,588	2,536	3,038	387	171	99	128	513	898	1,791	809	415	4,030	1,406	1,870
1982	5,360	2,341	2,957	330	107	94	102	411	699	1,781	718	318	3,410	1,236	1,564
1983	3,960	1,696	2,124	218	78	62	115	431	714	1,651	715	367	3,183	1,218	1,564
1984	5,391	2,277	2,873	343	128	95	115	468	781	1,866	718	310	3,377	1,310	1,739
1985	5,666	2,153	2,769	325	123	94	80	380	682	1,746	635	218	3,675	1,367	1,790
1986	4,807	1,964	2,331	309	110	98	82	377	665	1,708	623	301	3,518	1,104	1,658
1987	4,194	1,676	2,110	324	108	77	87	342	611	1,633	576	212	3,971	1,307	1,826
1988	4,499	1,856	2,245	390	142	78	104	367	720	1,740	643	307	3,778	1,121	1,623
1989	4,601	1,798	2,196	347	120	68	93	392	720	1,924	751	317	3,628	1,056	1,537
1990	4,748	1,891	2,399	419	133	90	118	326	679	1,800	723	323	3,990	1,271	1,712
1991	4,715	1,868	2,245	477	151	98	118	312	568	1,734	679	301	4,243	1,189	1,789
1992	4,887	1,854	2,256	466	153	140	98	320	584	1,889	688	254	4,106	1,204	1,808
1993	4,369	1,681	2,054	508	171	140	84	304	641	1,986	736	215	4,445	1,543	2,091
1994	4,603	1,740	2,119	649	159	140	100	290	624	2,050	726	227	5,240	1,605	2,158
1995	4,158	1,496	1,800	700	204	173	154	372	665	1,955	719	236	4,752	1,635	2,254
1996	4,829	1,795	2,136	563	193	230	116	393	737	2,208	755	263	4,588	1,391	1,892
1997	4,792	1,783	2,172	525	219	286	175	490	1,016	2,043	719	257	4,816	1,402	1,694
1998	4,846	1,666	2,012	472	212	259	141	415	788	2,017	735	279	4,837	1,586	1,963
1999	4,650	1,580	1,936	544	215	309	139	441	805	1,907	669	246	5,212	1,349	1,657
2000	4,909	1,763	1,920	567	225	304	160	428	762	1,891	636	228	4,808	1,262	1,757
2001	4,249	1,552	1,888	569	236	314	148	448	793	1,764	617	221	4,805	1,404	1,710
2002	4,720	1,701	2,074	508	204	281	155	470	952	1,751	632	227	4,875	1,623	1,447
2003	4,710	1,682	1,963	508	210	280	154	448	827	1,804	651	234	4,916	1,300	1,647
2004	4,792	1,729	2,076	502	206	277	156	464	873	1,957	697	216	5,621	1,729	2,078
2005	5,023	1,781	1,847	517	211	285	151	448	860	1,625	581	198	5,021	1,617	1,983
2006	4,690	1,696	1,901	559	228	304	109	400	755	1,430	538	138	5,256	1,617	1,625
2007	5,714	2,066	2,279	441	156	192	121	367	703	1,689	601	184	5,229	1,382	1,775
2008	5,224	1,888	1,687	421	157	202	120	436	750	1,647	731	175	5,149	1,035	1,846
2009	4,875	1,425	1,457	345	99	110	101	307	511	1,394	479	94	4,746	827	908
2010	5,610	1,933	1,991	380	139	188	111	552	774	1,331	458	90	4,800	1,079	1,417

Box Plot of Data

Boxplot is a way of summarizing data through visualizing the five number summary which consists of the minimum value, first quartile, median, third quartile, and maximum value of a data set.

These boxplots shows data of Nitrogen, Phosphate and Potash with Corn, Cotton, Soybean, Wheat and other. Here the median for each crop and element is different. We can see that there are outliers represented by star which is not in Inter Quartile Range (IQR). Nitrogen Wheat, Phosphate Wheat, Potash Corn, Potash Soybeans, and Potash other have extreme outliers

Boxplot of Nitrogen, Phosphate and Potash

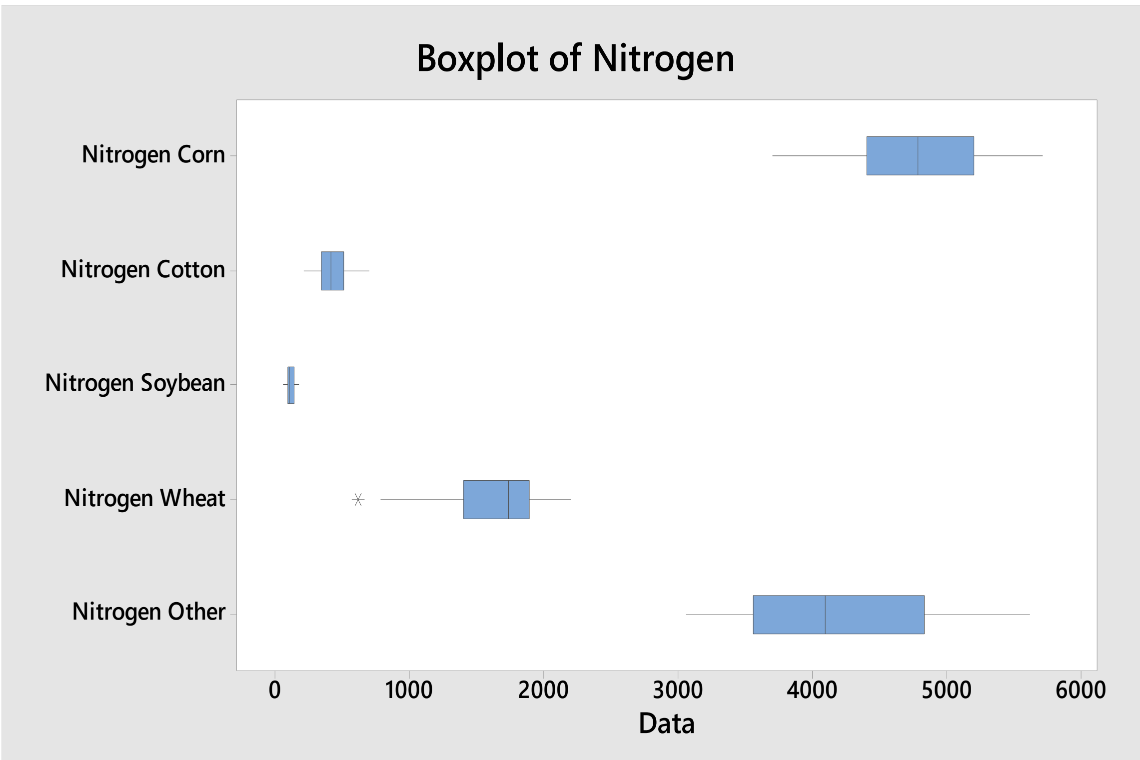

Figure 1 Box Plot of Nitrogen content in crops

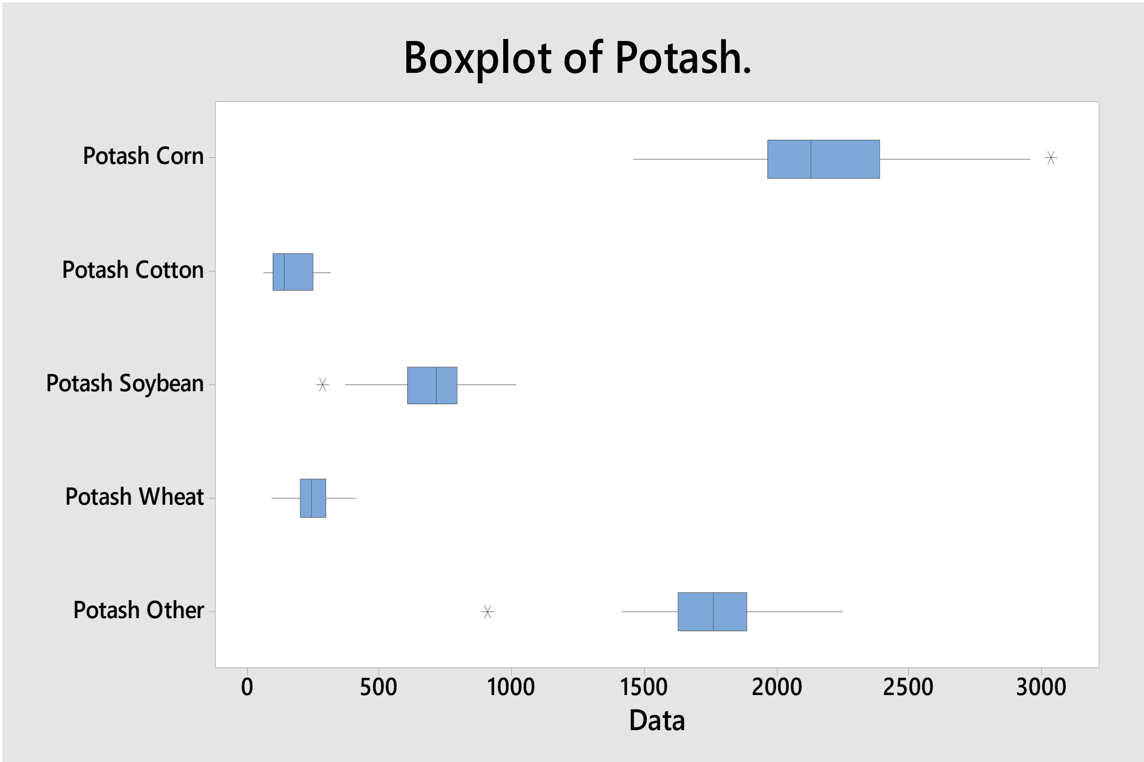

Figure 2 Box Plot of Potash content in crops

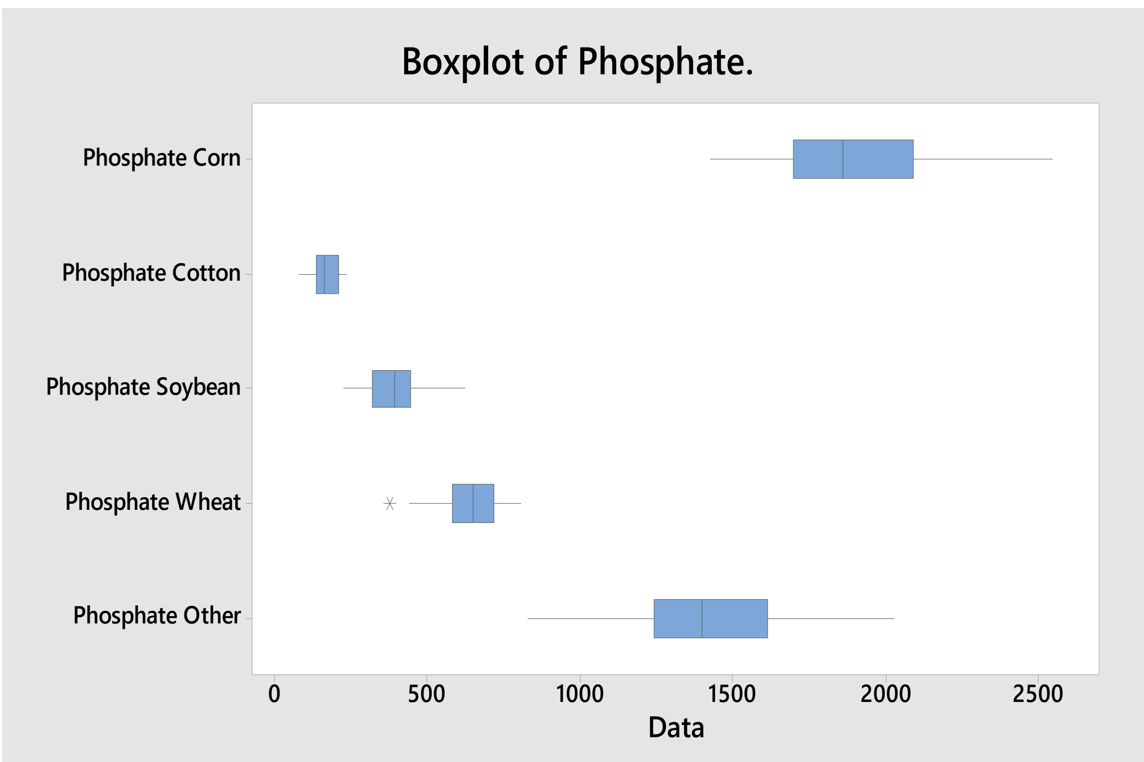

Figure 3 Box Plot of Phosphate contain of Crops

Histogram

A histogram is a plot that lets you understand the underlying frequency distribution of a set of continuous data, this data splits into intervals mainly called bins. Each bin contains the number of usage of fertilizers in the data set that are contained within that bin. The frequencies in each bin have been tabulated along with the fertilizers that contributed to the frequency in each bin. The area of the bar indicates the frequency of occurrences for each bin. This means that the height of the bar does not necessarily indicate how many occurrences fertilizers there were within each bin. One of the reasons is that the height of the bars is often incorrectly assessed as indicating frequency and not the area of the bar.

Histogram of Nitrogen, Phosphate and Potash

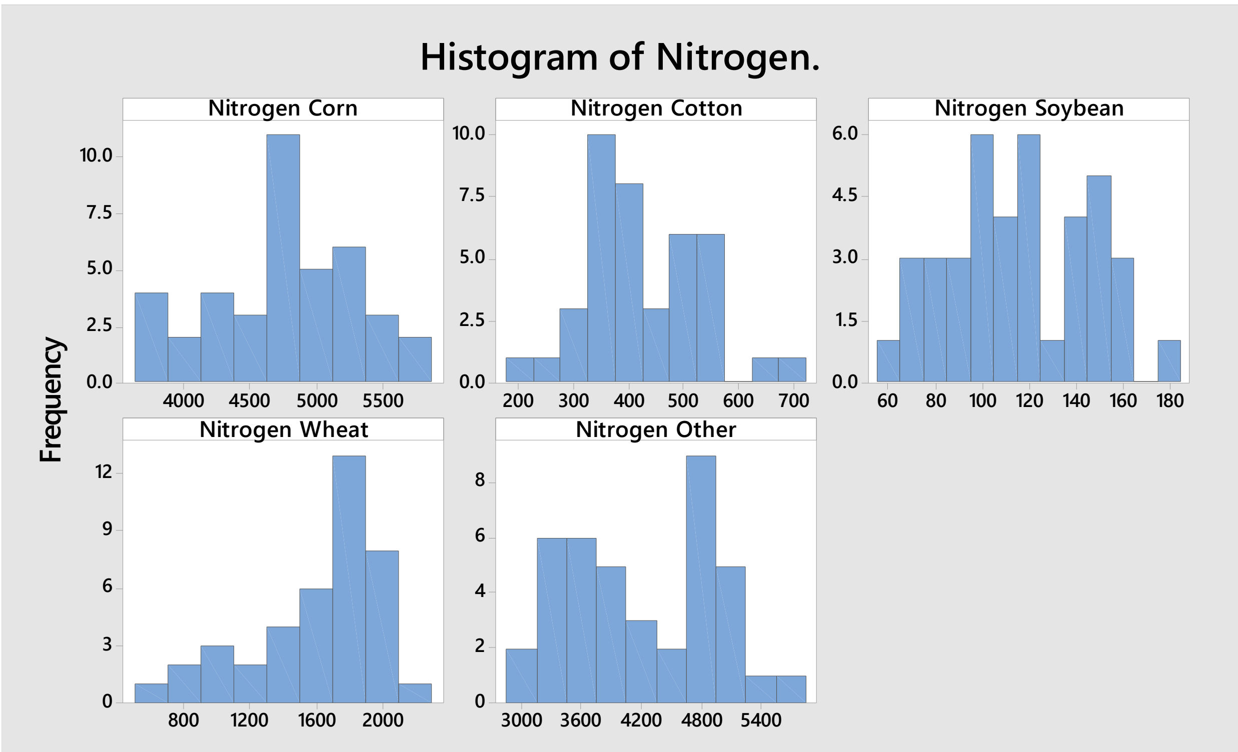

Figure 4 Histogram of Nitrogen Content in crops

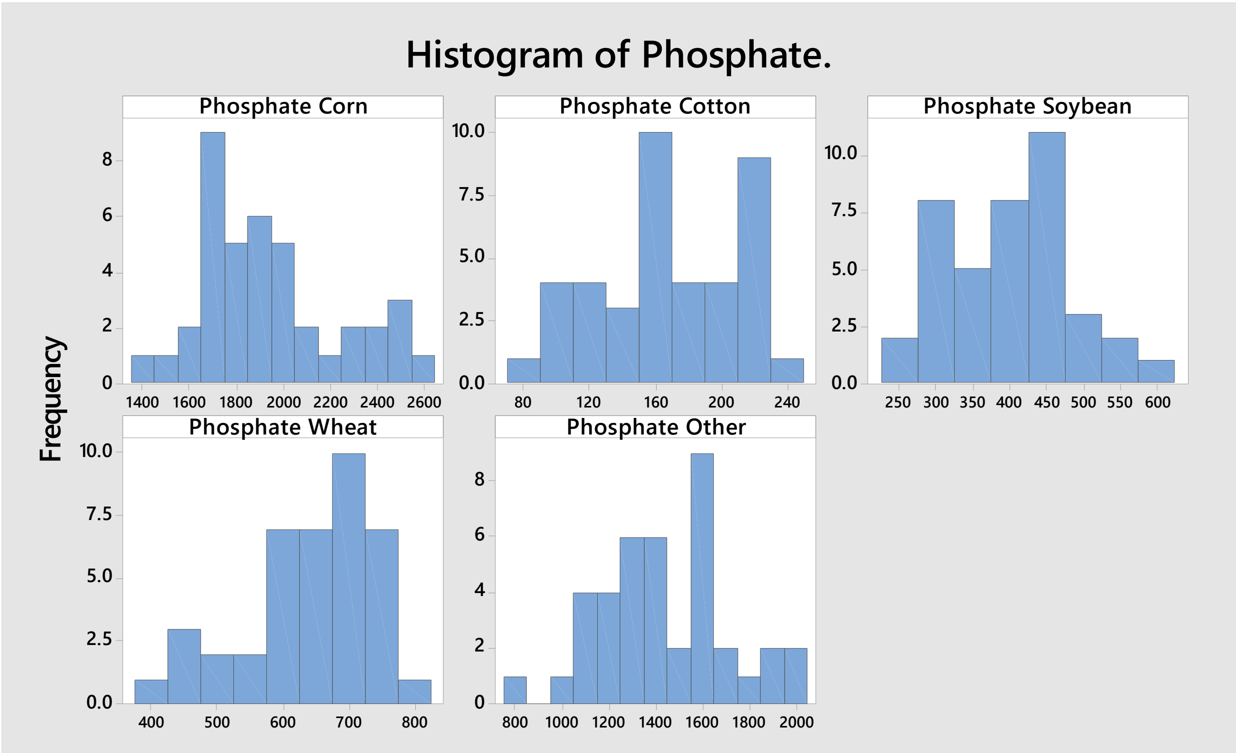

Figure 5 Histogram of Phophate contain in Crops

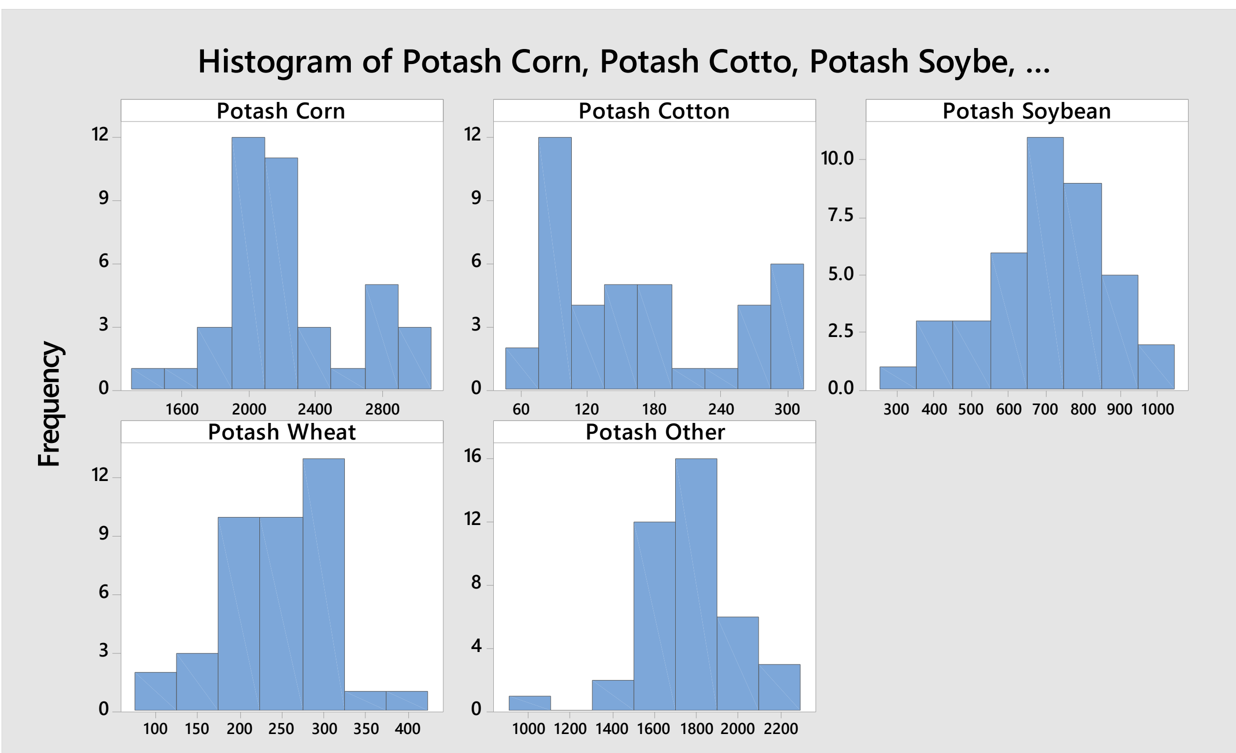

Figure 6 Histogram of Potash Contain in Crops

Probability Plot

It is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data. The general procedure is very simple and can be performed quickly. It is also more reliable than histogram for small-to moderate-size samples. Probability plotting typically uses special axes that have been scaled for the hypothesized distribution.

The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Deviation from the straight line indicates deviation from normality.

Probability Plot of Nitrogen, Phosphate and Potash

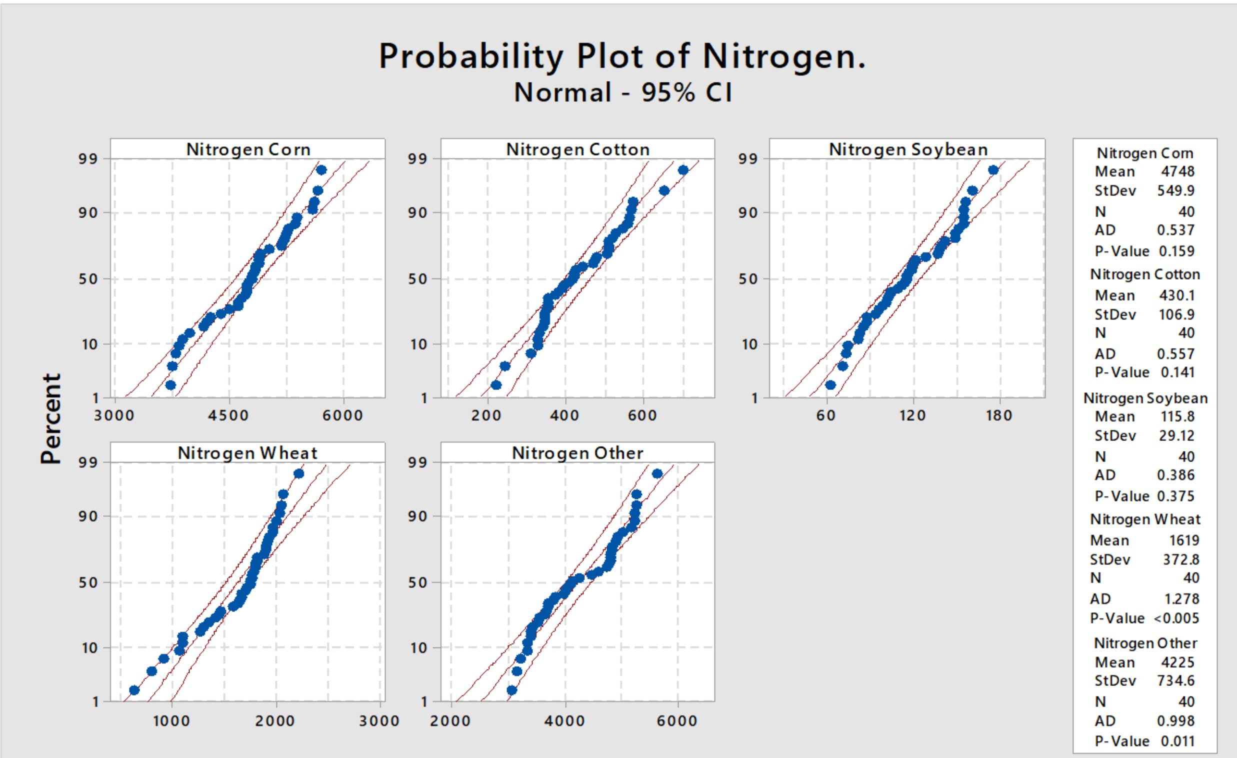

Figure 7 Probability Plot of Nitrogen contain in crops

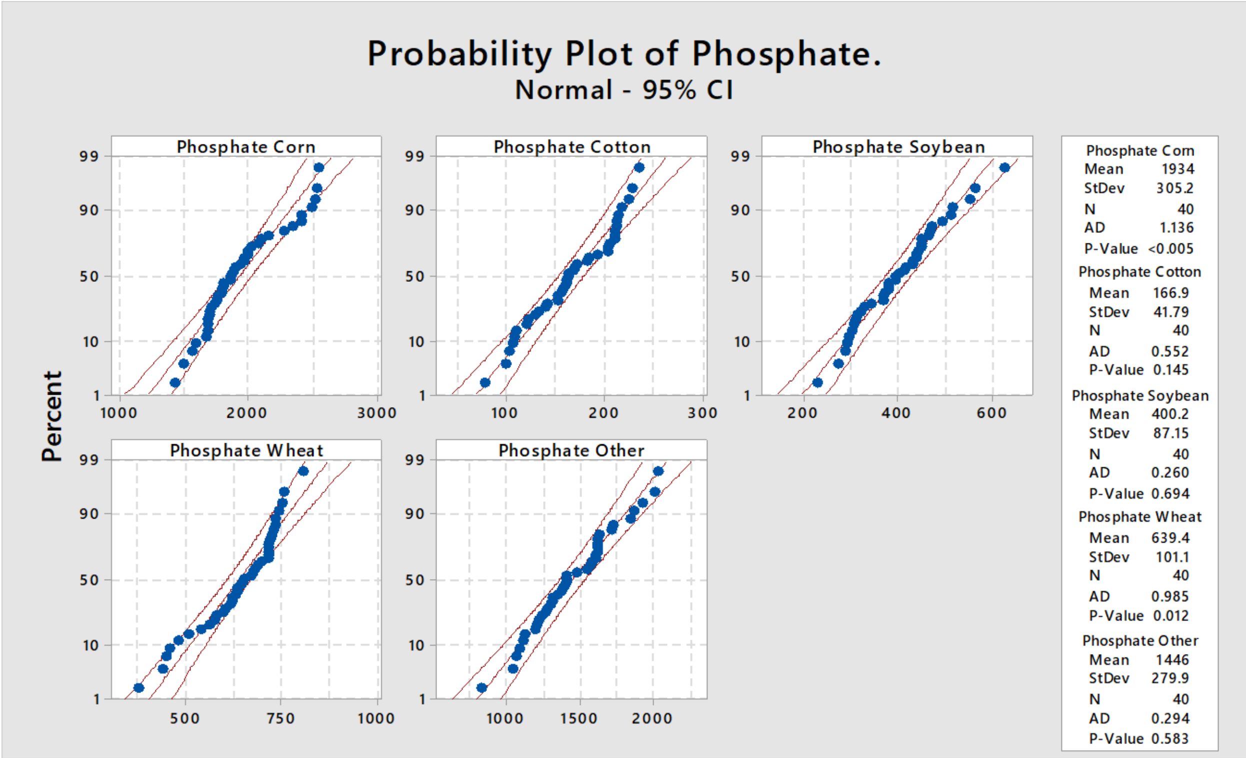

Figure 8 Probability Plot of Phosphate contain in crops

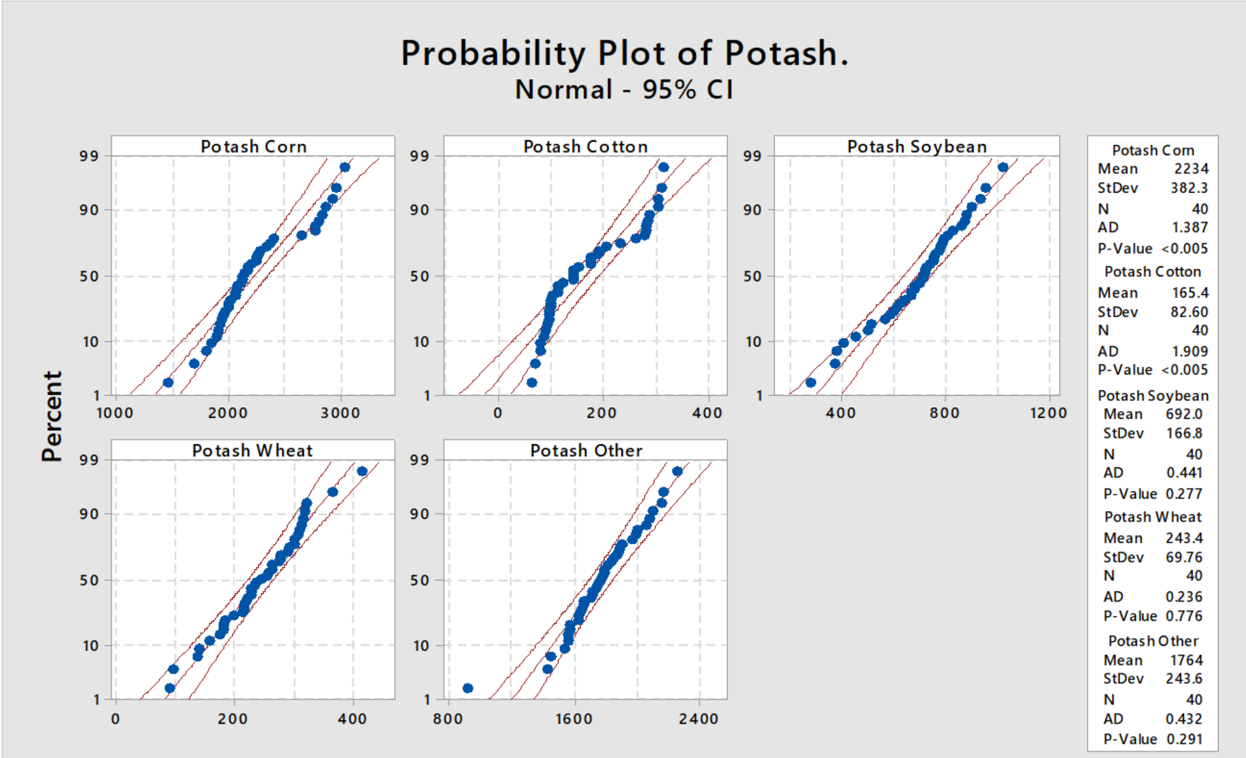

Figure 9 Probability plot of Potash Contain in crops

Hypothesis Testing

Hypothesis is confirmatory data analysis that is testable on the basis of observing the process that is modeled via set of variables. Hypothesis tests are used when determining what outcome of the study would results rejection of null hypothesis.

It was mentioned that Potash and Phosphate content of Fertilizers should be same to avoid alkalinity or acidity of soil and to get better quality yield of the crop. So we did Hypothesis testing on mean of content of potash and phosphate and standard deviation of potash and phosphate.

1. Two-Sample T-Test and CI: Phosphate Corn, Potash Corn

Method

μ_1: mean of Phosphate Corn

μ_2: mean of Potash Corn

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Phosphate Corn	40	1934	305	48
Potash Corn	40	2234	382	60

Estimation for Difference

Difference	95% CI for Difference
-300.3	(-354.4, +146.2)

Test

Null Hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-1.66	74	0.68

Here, value from T table

$T_{\frac{\alpha}{2},n-1}$=$T_{\frac{0.05}{2},74-1}$=1.99

Here the magnitude T stat value is less than $T_{\frac{\alpha}{2},n-1}$ so null hypothesis $\mu_1 = \mu_2$ is fail to reject so there is no difference between these 2 means for Corn.

2. Two-Sample T-Test and CI: PhosphateCotton, PotashCotton

Method

μ_1: mean of Phosphate Cotton

μ_2: mean of Potash Cotton

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Phosphate Cotton	40	166.9	41.8	6.6
Potash Cotton	40	165.4	82.6	13

Estimation for Difference

Difference	95% CI for Difference
1.4	(-27.9, 30.7)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.10	57	0.722

Here, value from T table

$T_{\frac{\alpha}{2},n-1}$=$T_{\frac{0.05}{2},57-1}$=2.02

Here the magnitude T stat value is less than $T_{\frac{\alpha}{2},n-1}$ so null hypothesis μ_1= μ_2 is fail to reject so there is no difference between these 2 means for cotton.

3. Two-Sample T-Test and CI: PhosphateSoybeans, PotashSoybeans

Method

μ_1: mean of Phosphate Soybeans

μ_2: mean of Potash Soybeans

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Phosphate Soybeans	40	400.2	87.2	14
Potash Soybeans	40	692	167	26

Estimation for Difference

Difference	95% CI for Difference
-291.8	(-351.3, -232.2)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$

Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-9.80	58	0.000

Here, value from T table

$T_{\frac{\alpha}{2},n-1}$=$T_{\frac{0.05}{2},58-1}$=2.018

Here the magnitude T stat value is more than $T_{\frac{\alpha}{2},n-1}$ so null hypothesis μ_1= μ_2 is rejected so there is difference between these 2 means for Soybean

We can see there is more potash contain added to Soybean to get more quantity of Grains.

4. Two-Sample T-Test and CI: Phosphate Wheat, Potash Wheat

Method

μ_1: mean of Phosphate Wheat

μ_2: mean of Potash Wheat

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Phosphate Wheat	40	639	101	16
Potash Wheat	40	243.4	69.8	11

Estimation for Difference

Difference	95% CI for Difference
396.0	(357.3, 434.8)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$

Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
20.39	69	0.000

Here, value from T table

$T_{\frac{\alpha}{2},n-1}=T_{\frac{0.05}{2},69-1}=1.999$

Here the magnitude T stat value is more than $T_{\frac{\alpha}{2},n-1}$ so null hypothesis $\mu_1 = \mu_2$ is rejected so there is no difference between these 2 means for wheat.

We can see there is more phosphate contain added to Wheat to get more quantity of Grains.

5. Two-Sample T-Test and CI: Phosphate Other, Potash Other

Method

μ_1: mean of Phosphate Other

μ_2: mean of Potash Other

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Phosphate Other	40	1446	280	44
Potash Other	40	1764	244	39

Estimation for Difference

Difference	95% CI for Difference
-318.0	(-434.9, -201.2)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-5.42	76	0.000

Here, value from T table

$T_{\frac{\alpha}{2},n-1}$=$T_{\frac{0.05}{2},76-1}$=1.988

Here the magnitude T stat value is more than $T_{\frac{\alpha}{2},n-1}$ so null hypothesis μ_1= μ_2 is rejected so there is difference between these 2 means for other crops

We can see there is more potash contain added to other crops to get more quantity of Production.

We are Checking the Variance are equal or not for the Corn and Cotton for which mean of potash and phosphate are equal
for Corn and Cotton we are checking the spread of data or variance having equal mean.

1. Test and CI for Two Variances: Phosphate Corn, Potash Corn

Method

σ_1: standard deviation of Phosphate Corn

σ_2: standard deviation of Potash Corn

Ratio: σ_1/σ_2

F method was used. This method is accurate for normal data only.

Descriptive Statistics

Variable	N	StDev	Variance	95% CI for σ
Phosphate Corn	40	305.207	93151.481	(250.014, 391.897)
Potash Corn	40	382.305	146157.024	(313.169, 490.893)

Ratio of Standard Deviations

Estimated Ratio	95% CI for Ratio using F
0.798335	(0.581, 1.098)

Test

Null hypothesis H_0: $\sigma_1 / \sigma_2 = 1$

Alternative hypothesis H_1: $\sigma_1 / \sigma_2 \neq 1$

Significance level $\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
F	0.64	39	39	0.164

$F_{\alpha/2,n1-1,n2-1} = F_{0.05/2,39-1,39-1}=1.88$

Here the magnitude of Test stats is less than F value for $\alpha/2$ and degree of freedom n_1-1,n_2-1 so null hypothesis is accepted so there is no difference between these 2 Variance. So for Corn mean as well as variance of potash and Phosphate contain is same.

2. Test and CI for Two Variances: Phosphate Cotton, Potash Cotton

Method

σ_1: standard deviation of Phosphate Cotton

σ_2: standard deviation of Potash Cotton

Ratio: σ_1/σ_2

F method was used. This method is accurate for normal data only.

Descriptive Statistics

Variable	N	StDev	Variance	95% CI for σ
Phosphate Cotton	40	41.788	1746.244	(34.231, 53.657)
Potash Cotton	40	82.597	6822.338	(67.661, 106.058)

Ratio of Standard Deviations

Estimated Ratio	95% CI for Ratio using F
0.505925	(0.368, 0.696)

Test

Null hypothesis	H_0: σ_1 / σ_2 = 1
Alternative hypothesis	H_1: σ_1 / σ_2 ≠ 1
Significance level	α = 0.05

Method	Test Statistic	DF1	DF2	P-Value
F	2.26	39	39	0.000

$F_{\alpha/2,n1-1,n2-1} = F_{0.05/2,39-1,39-1}$=1.88

Here the magnitude of Test stats is more than F value for $\alpha/2$ and degree of freedom n_1-1,n_2-1 so null hypothesis is accepted so there is difference between these 2 Variance.

So for Cotton although mean of potash and phosphate contain is same but their spread over the mean is different, i.e. their variance is different.

Regression Analysis

Regression analysis is used to infer casual relationships between the independent and dependent variables. In regression it is mandatory to justify why existing relationships have predictive power or why a relationship between two variables has a casual interpretation.

In this regression analysis, phosphate and potash are regressor variable or predictor variable and nitrogen is a dependent or response variable.

Regression Analysis: Nitrogen Corn
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	2681478	1340739	5.44	0.008
Phosphate Corn	1	12276	12276	0.05	0.825
Potash Corn	1	454134	454134	1.84	0.183
Error	37	9111309	246252		
Total	39	11792787			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
496.237	22.74%	18.56%	7.13%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3174	511	6.21	0.000	
Phosphate Corn	0.123	0.551	0.22	0.825	4.48
Potash Corn	0.598	0.440	1.36	0.183	4.48

Regression Equation

Nitrogen Corn = 3174 + 0.123 Phosphate Corn + 0.598 Potash Corn

Fits and Diagnostics for Unusual Observations

Obs	Nitrogen Corn	Fit	Resid	Std Resid	
4	3773	4844	-1071	-2.20	R
40	5610	4602	1008	2.11	R

R Large residual

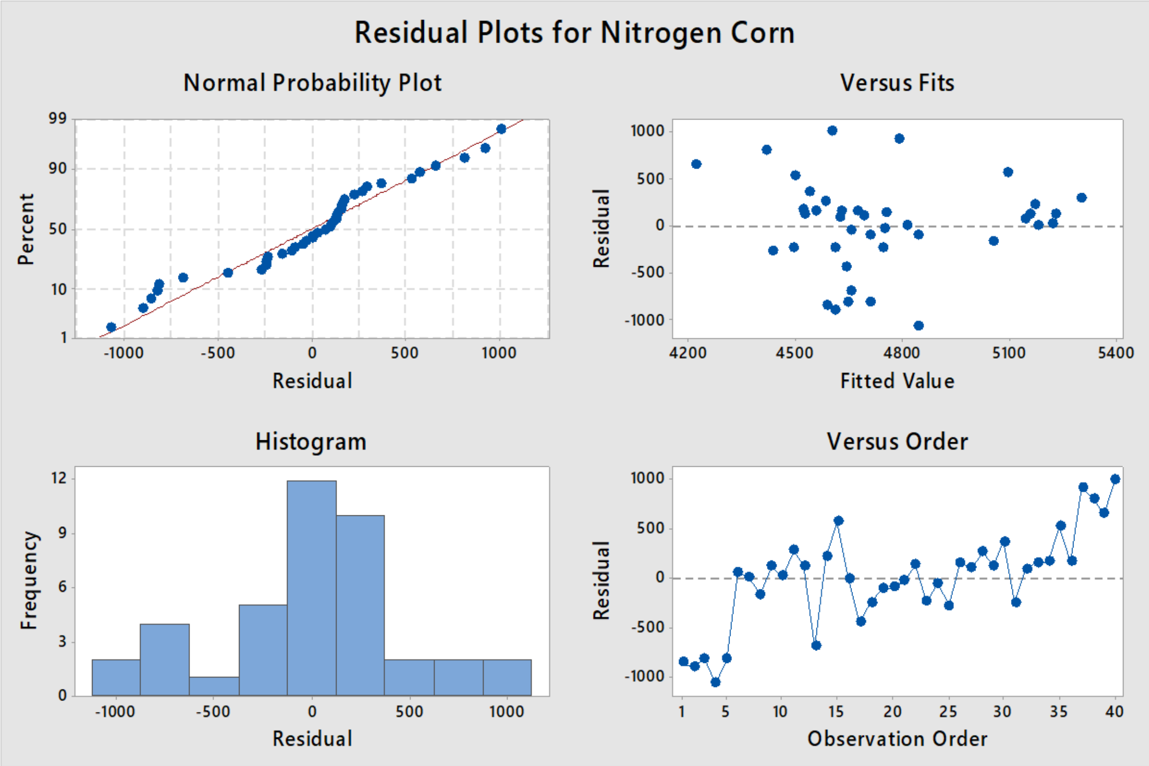

Figure 10 Residual plot for Nitrogen contain in Corn

First table denotes the 'Analysis of Variance'. In this table, the contribution of an individual regressor variable to the model is tested. In this approach, the increase in the regression sum of squares obtained by adding a variable xj to model is determined. This test is called the general regression significance test, or the extra sum of squares method. From this procedure we can investigate the contribution of a subset of the regressor variables to the model.

From F-value we can determine whether the regressor variable is associated with the response. If the p-value for regression is less than significance level α, we can conclude that the model defines a linear relationship. In second table, hypothesis on the individual regression coefficients is tested. This test is used for determining the potential value of each regressor variable in the regression model. From this table, we can find out that model is more effective with the inclusion of additional variables or with the deletion of one or more of the regressor variables in the model

For t, if p-value is less than α, we can conclude that the model specifies a linear relationship between the given two variables. If p-value is greater than significance level α of regressor xj, then we cannot reject null hypothesis. Therefore, regressor xj can be deleted from the model.

Regression Analysis: Nitrogen Cotton

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	256945	128473	25.17	0.000
Phosphate Cotton	1	39499	39499	7.74	0.008
Potash Cotton	1	8587	8587	1.68	0.203
Error	37	188884	5105		
Total	39	445829			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
71.4490	57.63%	55.34%	52.10%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	146.9	54.4	2.70	0.010	
Phosphate Cotton	1.376	0.495	2.78	0.008	3.26
Potash Cotton	0.324	0.250	1.30	0.203	3.26

Regression Equation

Nitrogen Cotton = 146.9 + 1.376 Phosphate Cotton + 0.324 Potash Cotton

Fits and Diagnostics for Unusual Observations

Obs	Nitrogen Cotton	Fit	Resid	Std Resid	
24	649.0	411.3	237.6	3.37	R
25	699.7	483.4	216.3	3.16	R

R Large residual

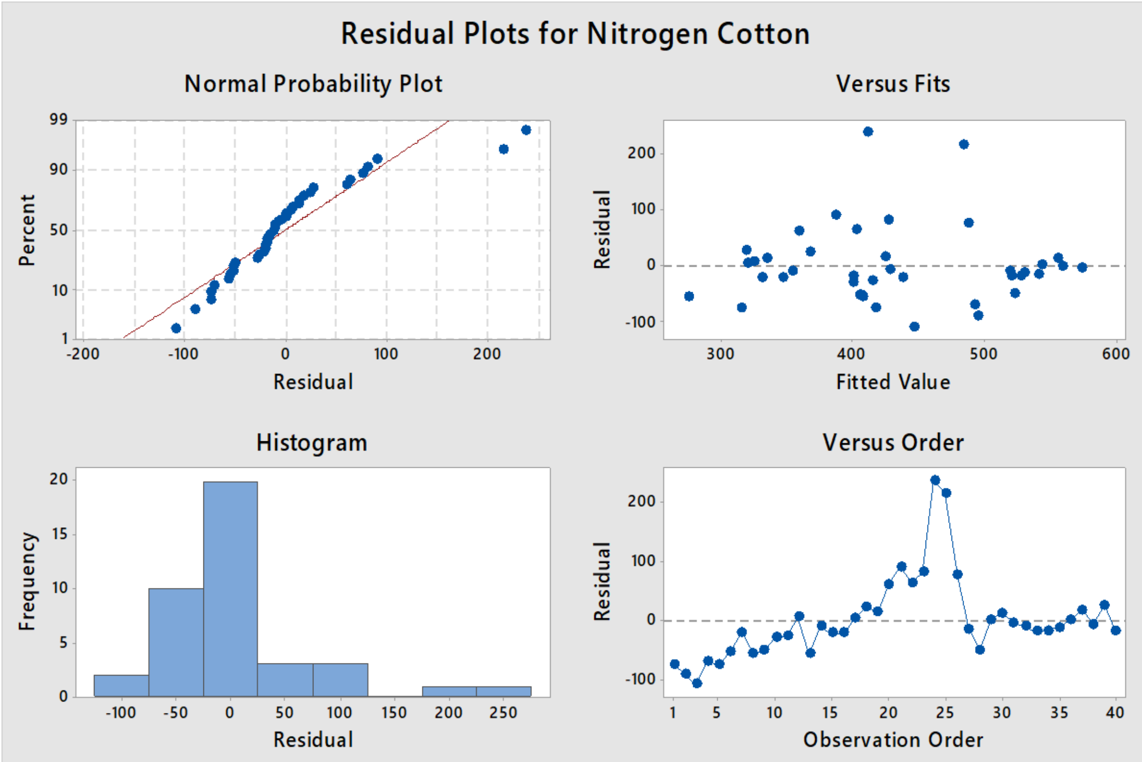

Figure 11 Residual plot for Nitrogen contain in Cotton

Regression Analysis: Nitrogen Soybean

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	22138.4	11069.2	37.44	0.000
Phosphate Soybean	1	25.8	25.8	0.09	0.769
Potash Soybean	1	5971.3	5971.3	20.20	0.000
Error	37	10938.1	295.6		
Total	39	33076.5			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
17.1937	66.93%	65.14%	63.18%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	15.4	13.0	1.18	0.245	
Phosphate Soybean	0.0170	0.0576	0.30	0.769	3.33
Potash Soybean	0.1353	0.0301	4.49	0.000	3.33

Regression Equation

Nitrogen Soybean = 15.4 + 0.0170 Phosphate Soybean + 0.1353 Potash Soybean

Fits and Diagnostics for Unusual Observations

Obs	Nitrogen Soybean	Fit	Resid	Std Resid		
9	148.53	152.22	-3.69	-0.25		X
25	154.25	111.65	42.60	2.51	R	
30	160.41	125.74	34.67	2.05	R	

R Large residual
X Unusual X

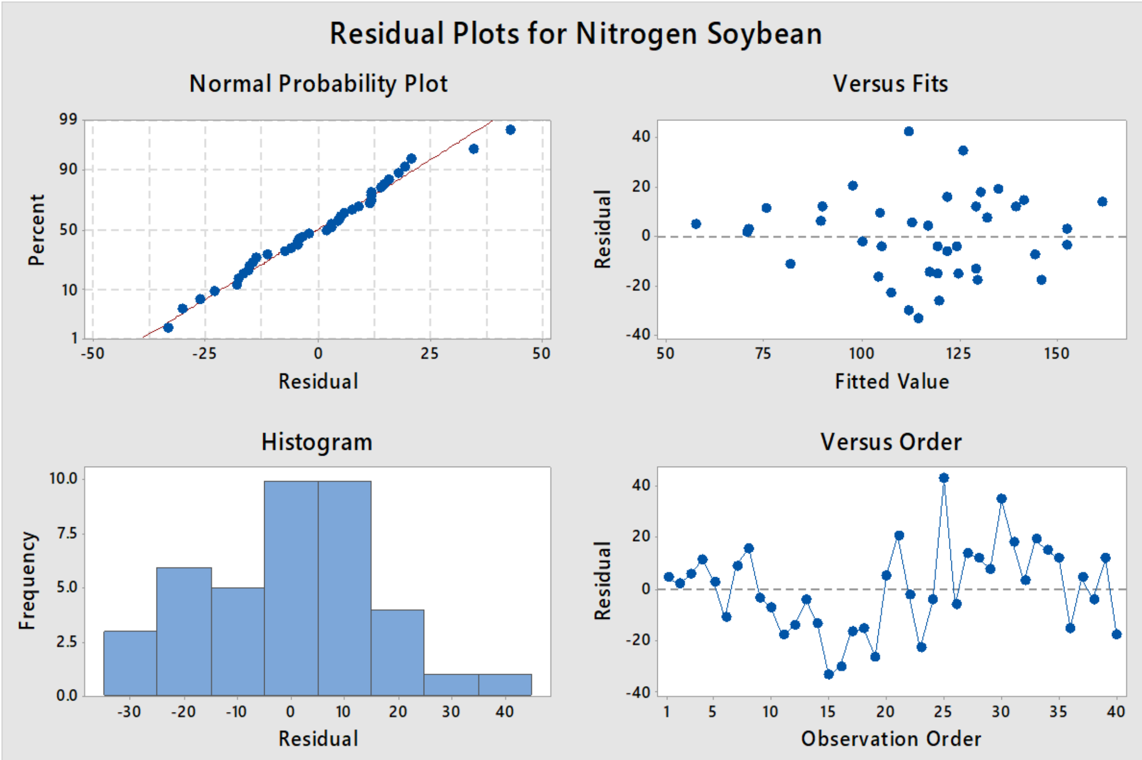

Figure 12 Residual plot for Nitrogen contain in Soybean

Regression Analysis: Nitrogen Wheat

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	3989052	1994526	51.53	0.000
Phosphate Wheat	1	3177782	3177782	82.09	0.000
Potash Wheat	1	488135	488135	12.61	0.001
Error	37	1432250	38709		
Total	39	5421302			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
196.747	73.58%	72.15%	67.78%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-466	209	-2.23	0.032	
Phosphate Wheat	4.161	0.459	9.06	0.000	2.17
Potash Wheat	-2.364	0.666	-3.55	0.001	2.17

Regression Equation

Nitrogen Wheat = -466 + 4.161 Phosphate Wheat - 2.364 Potash Wheat

Fits and Diagnostics for Unusual Observations

Obs	Nitrogen Wheat	Fit	Resid	Std Resid	
4	1078.4	1496.4	-417.9	-2.16	R
6	1455.6	1890.0	-434.5	-2.27	R
38	1647.0	2162.0	-515.0	-2.92	R

R Large residual

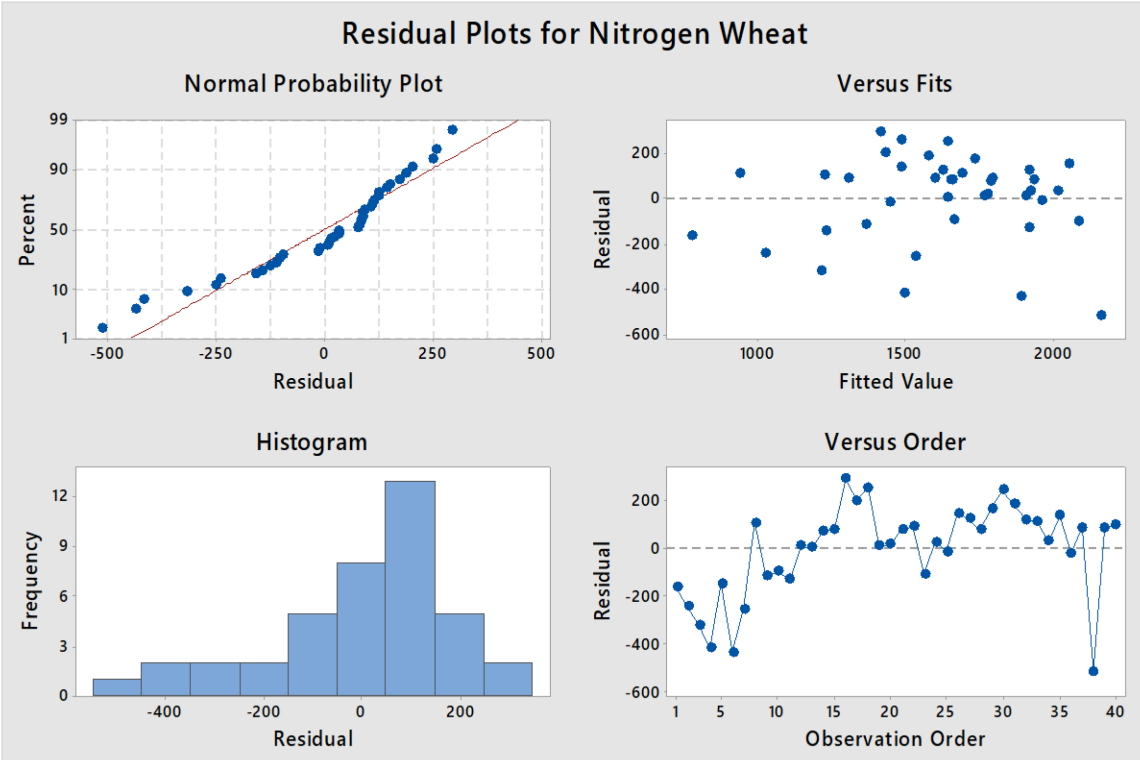

Figure 13 Residual plot for Nitrogen contain in Wheat

Regression Analysis: Nitrogen Other

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	2128919	1064460	2.08	0.139
Phosphate Other	1	1673179	1673179	3.27	0.079
Potash Other	1	1519319	1519319	2.97	0.093
Error	37	18916245	511250		
Total	39	21045164			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
715.017	10.12%	5.26%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3810	861	4.42	0.000	
Phosphate Other	-0.855	0.473	-1.81	0.079	1.34
Potash Other	0.937	0.543	1.72	0.093	1.34

Regression Equation

Nitrogen Other = 3810 - 0.855 Phosphate Other + 0.937 Potash Other

Fits and Diagnostics for Unusual Observations

Obs	Nitrogen Other	Fit	Resid	Std Resid	
39	4746	3953	794	1.37	X

X Unusual X

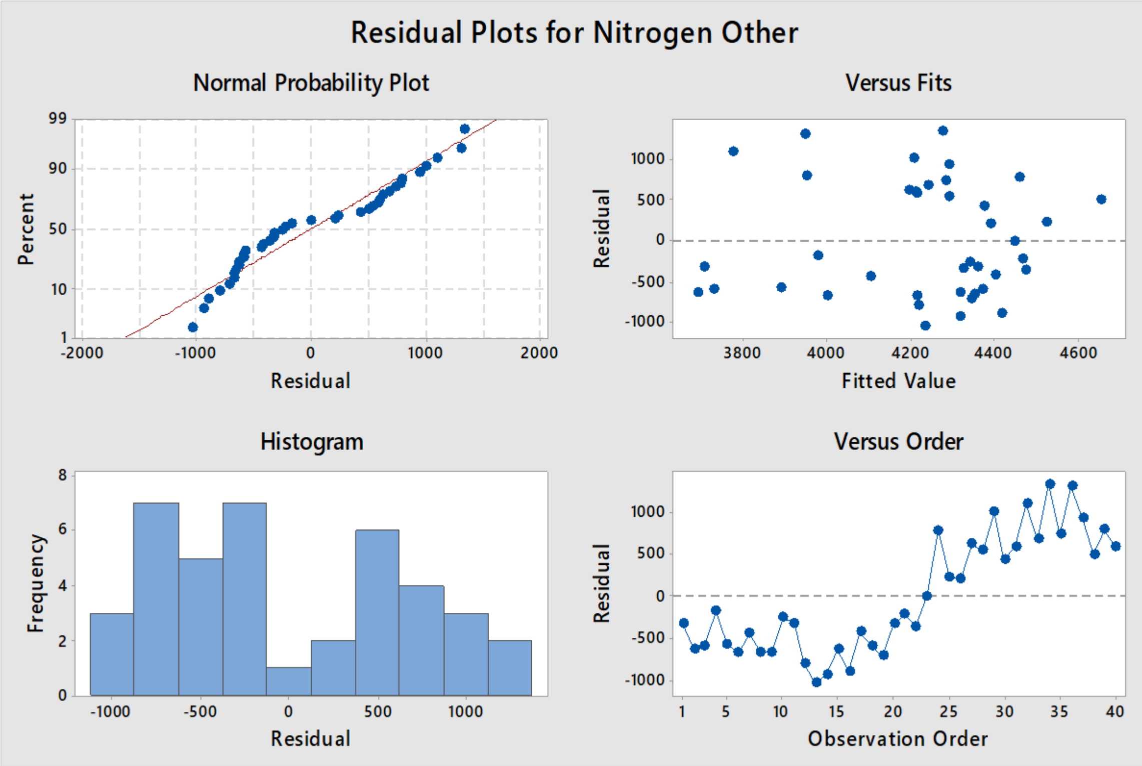

Figure 14 Residual plot for Nitrogen contain in Other

ANOVA of the data

We are analyzing the data to identify if there is any effect of individual Plants and individual fertilizer content (Factors) on the Responses. Also if there is any effect of interaction between the Plant and fertilizer content (Factor interaction) on the Responses.

General Linear Model: US Plant Nutrient

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Factor 1_Plant Name	Fixed	5	CORN, COTTON, OTHER, SOYABEAN, WHEAT
Factor 2_Fertilizer Content	Fixed	2	PHOSPHATE, POTASH

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor 1_Plant Name	4	103787672	25946918	143.44	0.000
Factor 2_Fertilizer Content	1	9681432	9681432	53.52	0.000
Error	394	71270040	180888		
Lack-of-Fit/**Interaction**	4	59697881	14924470	502.98	0.000
Pure Error	390	11572159	29672		
Total	399	184739144			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
425.310	61.42%	60.93%	60.24%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	761.6	21.3	35.81	0.000	
Factor 1_Plant Name					
CORN	287.9	42.5	6.77	0.000	1.60
COTTON	-595.4	42.5	-14.00	0.000	1.60
OTHER	843.2	42.5	19.83	0.000	1.60
SOYABEAN	-215.5	42.5	-5.07	0.000	1.60
Factor 2_Fertilizer Content					
PHOSPHATE	155.6	21.3	7.32	0.000	1.00

Regression Equation

US Plant Nutrient(1,000 nutrien = 761.6 + 287.9 Factor 1_Plant Name_CORN

- 595.4 Factor 1_Plant Name_COTTON
+ 843.2 Factor 1_Plant Name_OTHER
- 215.5 Factor 1_Plant Name_SOYABEAN
- 320.2 Factor 1_Plant Name_WHEAT
+ 155.6 Factor 2_Fertilizer Content_PHOSPHATE
- 155.6 Factor 2_Fertilizer Content_POTASH

Fits and Diagnostics for Unusual Observations

Obs	US Plant Nutrient(1,000 nutrient)	Fit	Resid	Std Resid	
4	2102.0	1205.0	897.0	2.12	R
6	2550.0	1205.0	1345.0	3.19	R
7	2523.0	1205.0	1318.0	3.12	R
8	2416.0	1205.0	1211.0	2.87	R
9	2499.0	1205.0	1294.0	3.07	R
10	2413.0	1205.0	1208.0	2.86	R
11	2536.0	1205.0	1331.0	3.15	R
12	2341.0	1205.0	1136.0	2.69	R
14	2277.0	1205.0	1072.0	2.54	R
15	2153.0	1205.0	948.0	2.25	R
37	2066.0	1205.0	861.0	2.04	R
359	827.0	1760.3	-933.3	-2.21	R

R Large residual

Factor 1 – $F_{0.05,4,390} = 2.37$ $F_0 = 143.44$ ---> $F_0 > F_{0.05,4,390}$

So **Factor 1** has significant effect on our Responses.

Factor 2 – $F_{0.05,1,390} = 3.84$ **$F_0 = 53.52$** ---> $F_0 > F_{0.05,1,390}$

So **Factor 2** has significant effect on our Responses.

Interaction -- $F_{0.05,4,390} = 2.37$ **$F_0 = 502.98$** ---> $F_0 > F_{0.05,4,390}$

So **Interaction** has significant effect on our Responses.

We analyzed that there is each factor is affecting the responses and the interaction of two factors is significant and affects the responses. This is based on the p-values in ANOVA table. All P-values<0.05 as Confidence interval assumed 95%

Also MINITAB gave us some interesting data on unusual readings in the data set.

Standardized residuals are used to detect outliers. Standardized residuals less than -2 and greater than 2 are usually considered large. The Fits and Diagnostics for Unusual Observations identify these observations with an 'R' tag. Such observations do not follow the regression equation well. However, we may have some unusual observations.

But, based on the criteria for large standardized residuals, one would expect roughly 5% of your observations to be tagged as having a large standardized residual.
In our case out of 400 observations analyzed we found 12 unusual observations.
So that makes 3 % of data to be unusual. This is quite under the 5 % criteria.
Standardized residuals are useful as raw residuals might not be good indicators of outliers.
The variance of each raw residual can differ by the x-values associated with it. This unequal variation causes it to be difficult to assess the magnitudes of the raw residuals. Standardizing the residuals solves this problem by converting the different variances to a common scale.

Residual

A residual (e_i) is the difference between an observed value (y) and the corresponding fitted value, ($\hat{y}$), which is the value predicted by the model.

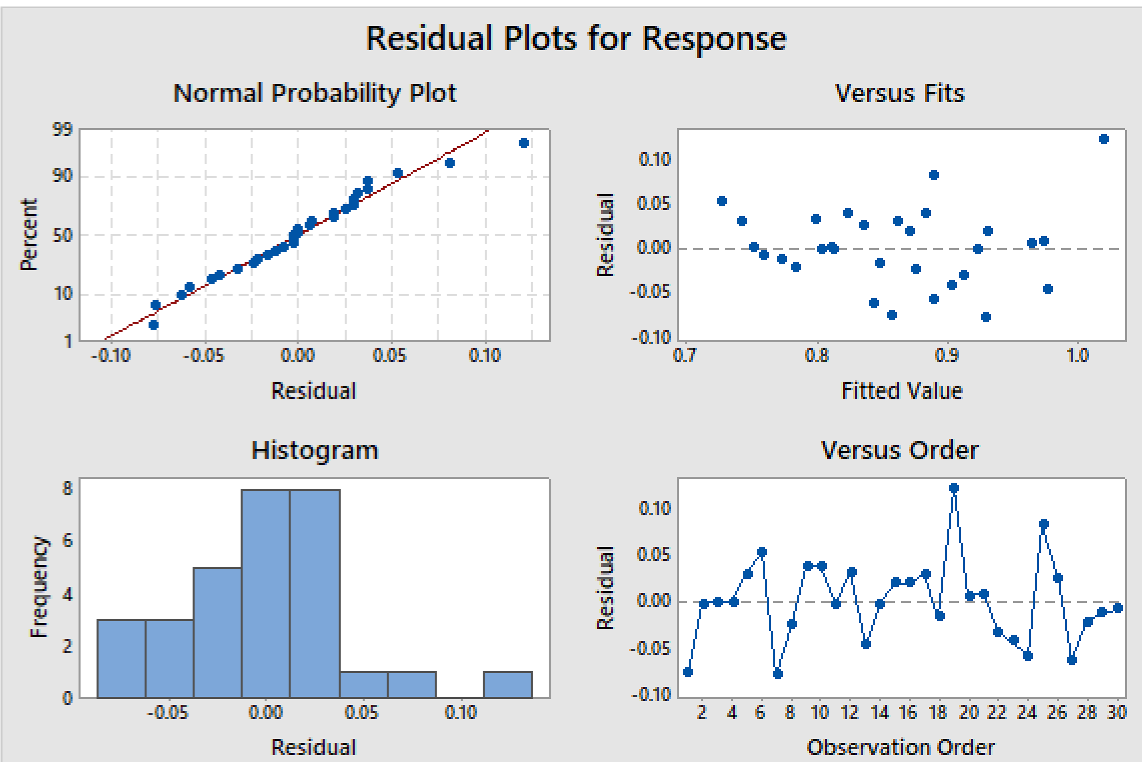

Figure 15 Residual plots of response

Conclusion

- Boxpolts, Histogram and probability plot gave us the spread of the data about median, outliers, mean, Normality.
- From the hypotheses we conclude that mean contain of potash and phosphate is equal for Corn and cotton but their spread about mean i.e variance is different.
- From the regression we conclude that there is a relation between Nitrogen (dependent) and Potash, Phosphate (independent) contain in the fertilizer.
- We observed that each factor is affecting the responses and the interaction of two factors (Crops and Fertilizers) is significant and affects the responses.

References

https://www.data.gov

https://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx

https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/unusual-observations/

https://support.minitab.com/en-us/minitab/19/help-and-how-to/modeling-statistics/regression/how-to/fit-regression-model/interpret-the-results/all-statistics-and-graphs/fits-and-diagnostics-table/

https://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm#:~:targetText=The%20normal%20probability%20plot%20

https://statistics.laerd.com/statistical-guides/understanding-histograms.php