Statewide Groundwater Monitoring Project for Oklahoma

Clean Water Act, Ground Water 106 Grant No. I-006400-05

Project 18, Task 200.3 - Fiscal Year 2006/07 Report

Prepared by: Michael S. Houts

Statewide Groundwater Monitoring Project for Oklahoma

Table of Contents

Introduction to Ground Water Monitoring FY' 2006/07

Objective:

The State of Oklahoma has an ambient ground water monitoring program that is administered through the Oklahoma Department of Environmental Quality (DEQ). The primary objective of this project is to assess the quality of groundwater in the State. In the long term, data will be analyzed for trends in order to identify areas where measures should be taken to preserve the beneficial uses of the groundwater. These objectives will be met by sampling a subset of Public Water Supply (PWS) wells to determine current constituent levels. Public Water Supply wells have been selected as the most reasonable choice for this project. Their construction, location, security, and operation should make them a close approximation to dedicated "monitoring wells."

By extended multi-year sampling of a limited subset of wells that are built to similar construction standards and stratified to insure proper randomization, limited assessment and trend statements can be made concerning the State's groundwater quality. Use of sample site stratification will insure proper areal coverage, which will allow for better conclusions to be drawn as a result of the project.

This year approximately 500 water samples were obtained including the necessary QA samples, for laboratory analysis by the State Environmental Laboratory. This resulted in the analysis of an estimated 7,000 parameters.

Selected parameters for measurement.

A total of 16 chemical and physical parameters reflecting secondary drinking water standards and major ions have been selected for analysis.

1. Aluminum
2. Chloride
3. Copper
4. Fluoride
5. Iron
6. Manganese
7. Nitrate as (N)
8. Sulfate
9. Total Dissolved Solids
10. Zinc
11. Calcium
12. Magnesium
13. Sodium
14. Total Alkalinity
15. pH
16. Potassium

Methodology:

1. Selected wells were sampled. Field blanks, duplicates, spikes, etc. were also obtained in order to meet QA/QC requirements.
2. Analyses were completed by the State Environmental Laboratory.
3. Information about measured results were entered into a database.
4. Reports were prepared and made available.

Outputs:

1. Semi-annual reports
2. Final report sent to Oklahoma Secretary of the Environment (OSE)
3. Results made available to the public via the DEQ website

Data Assessment and Summary:

The data assessment phases of this project include an exploratory statistical analysis performed in order to better understand the range, variance and distribution of the data. These exploratory analyses include:

- Calculation of mean
- Calculation of median
- Calculation of quartiles
- Calculation of standard deviation and variance
- Temporal plots and normal probability plots

Results of the statistical analysis effort are contained in the following section. Page one of the data set is a Graphical Summary of the data. It includes all the elements described in the objectives section that are needed to better understand the range, variance and distribution of the data; except Temporal Plots. Temporal plots will be completed when sufficient samples have been collected. However, the first year of samples is not sufficient for that purpose. Three to five sample periods will be necessary in order to give this plot method significant meaning. Page two contains different views of the same data as page one, plus the Normal Probability Plot. The Cumulative Distribution Function plot is on page three, and illustrates how well the data follow a normal distribution pattern.

Explanation of the Graphical Summary

The graphical summary provides graphs to summarize data, as well as an overall statistical summary, all presented within the same window. This information is copied from Minitab software help pages and is paraphrased in places.

Histogram of Data with Normal Curve

The histogram of the data is overlaid with a normal curve to assess the normality of the data. A normal distribution will be symmetric and bell-shaped.

Boxplot

- Boxplots summarize information about the shape, dispersion, and center of the data. They can also help identify outliers.
- The left edge of the box represents the first quartile (Q1), while the right edge represents the third quartile (Q3). Thus the box portion of the plot represents the interquartile range (IQR), or the middle 50% of the observations.
- The line drawn through the box represents the median of the data.
- The lines extending from the box are called whiskers. Whiskers extend outward to indicate the lowest and highest values in the data set (excluding outliers).
- Extreme values, or outliers, are represented by dots. A value is considered an outlier if it is outside of the box (greater than Q3 or less than Q1) by more than 1.5 times the IQR.
- Use the boxplot to assess the symmetry of the data:
    - If the data are fairly symmetrical, the median line will be roughly in the middle of the IQR box and the whiskers will be similar in length.
    - If the data are skewed, the median may not fall in the middle of the IQR box, and one whisker will likely be noticeably longer than the other.

Confidence Intervals for Mean, Standard Deviation, and Median

A confidence interval is an interval used to estimate a population parameter from sample data. The upper and lower bounds of the confidence intervals for m (mu), s (standard deviation), and the median are displayed in the graphical summary. In addition, the confidence intervals for m and the median are displayed graphically.

Confidence intervals are composed of two basic parts:

- Point estimate - a single value computed from the sample data. This value is considered to be an estimate of the parameter of interest. However, it is unlikely that the point estimate is equal to the parameter. Therefore, to account for the possibility of estimation error, the error margin is included in the confidence interval to provide a range of possible parameter values.

- Error margin - determines the width of the confidence interval through the use of probability. To construct the confidence interval, add and subtract the error margin from the point estimate.

If a 95% confidence interval is selected, the method used to construct the interval has a probability of 0.95 of producing an interval containing the parameter of interest. In other words, you can be 95% confident that the true value of the parameter is within the interval. Thus, if one hundred 95% confidence intervals were constructed, you would expect around 95 of the intervals to contain the parameter.

Table of Statistics

Anderson-Darling Normality Test

The Anderson-Darling normality test can help determine whether the data follow a normal distribution. The "A statistic" that the test provides is used to determine the p-value, which ranges from 0 to 1, and indicates how likely it is that the data follows a normal distribution.

First, it must be determined how low the p-value must be to conclude that the data are not normal. A commonly chosen criterion is 0.1. If the p-value is lower than the criterion, it must be concluded that the data do not follow a normal distribution. Otherwise, there is not enough evidence to conclude that the data do not follow a normal distribution.

Mean and N

Mean

The mean, also called the average, is a measure of where the center of the distribution lies. The mean is determined by calculating the sum of all observations divided by the number of observations. The mean is strongly influenced by extreme values.

N

N is the number of non-missing values in the data set.

Standard Deviation (StDev) and Variance

The standard deviation and variance measure dispersion, or how far the observations in a sample deviate from the mean. The standard deviation is analogous to an average distance (independent of direction) from the mean. The variance is the standard deviation squared. Like the mean, the standard deviation (as well as the variance) is very sensitive to extreme values.

Skewness and Kurtosis

Skewness

Skewness refers to a lack of symmetry within a data set. A distribution is skewed if one tail extends farther than the other. A skewness statistic is provided with the graphical summary:

- A value close to 0 indicates symmetrical data.
- Negative values imply negative/left skew.
- Positive values indicate positive/right skew.

Kurtosis

Kurtosis refers to the peak sharpness of a distribution curve. A kurtosis statistic is provided with the graphical summary:

- Values close to 0 indicate normally peaked data.
- Negative values indicate a distribution peak that is flatter than normal.
- Positive values indicate a distribution with a sharper than normal peak.

Minimum and Maximum

One of the easiest ways to assess dispersion within a data set is to compare the minimum and maximum values. The minimum is the smallest value in a data set, and the maximum is the largest value. For the precipitation data the minimum is 1 and the maximum is 10.

Minimum and maximum values are used to calculate the range, which is a statistic often used to describe dispersion within data sets. The range is the (Maximum)-(Minimum).
The range is very sensitive to extreme values.

First and Third Quartiles (Q1 and Q3)

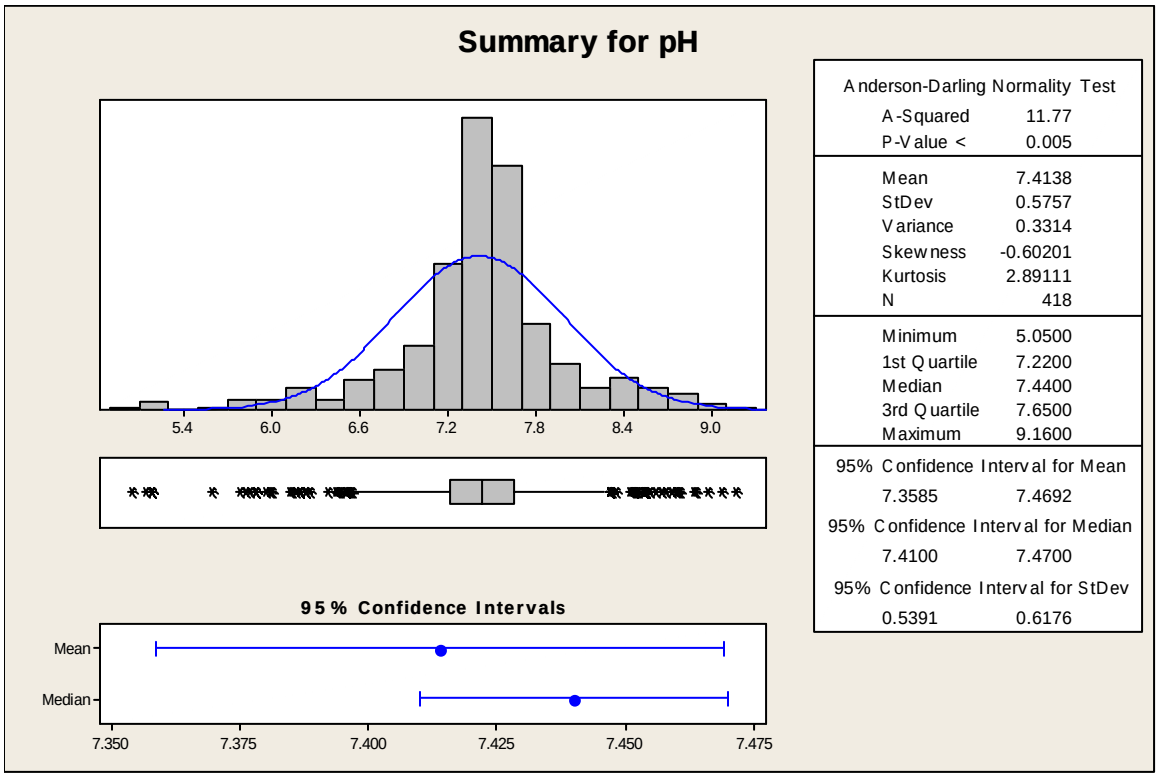
The first quartile (Q1, also called the 25th percentile) is the highest value for the lowest 25% of the observations. The third quartile (Q3, also called the 75th percentile) is the lowest value for the highest 25% of the observations. Q1 and Q3 are often used to calculate the interquartile range (IQR), which is also used to describe dispersion. The IQR is the range of the middle 50% of the values, and is calculated by subtracting Q3 from Q1. The IQR is relatively insensitive to extreme values.

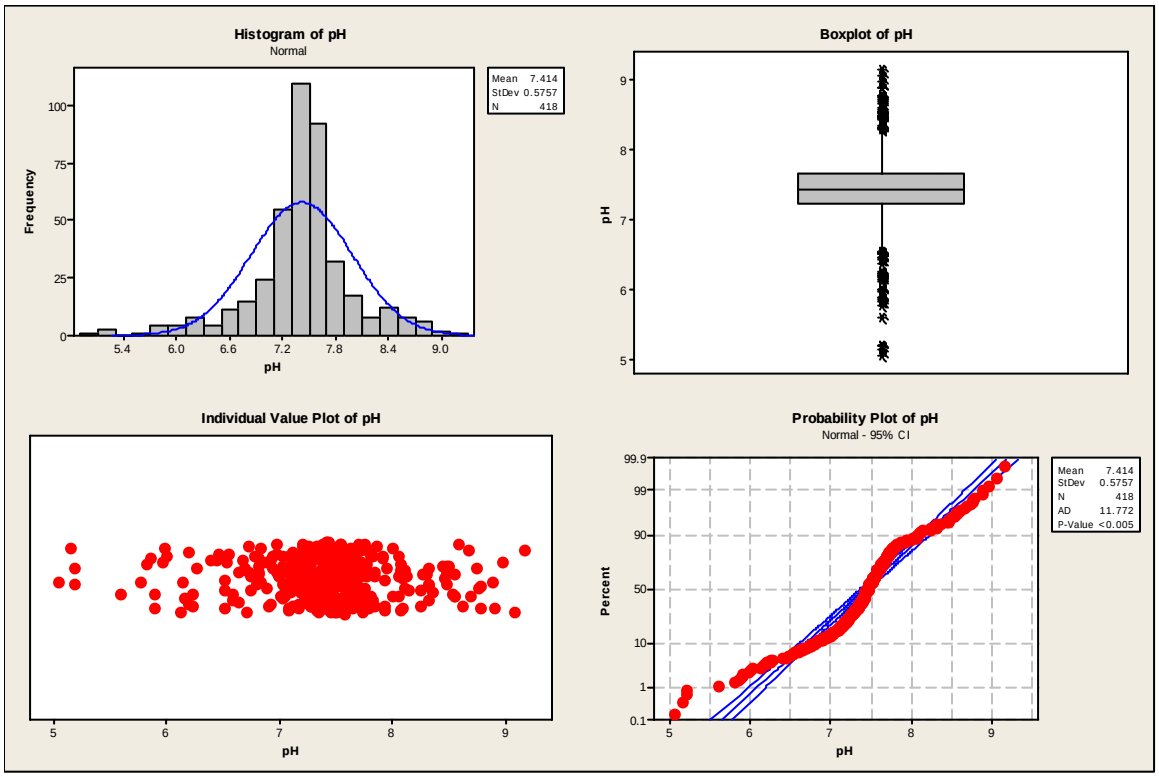
Median

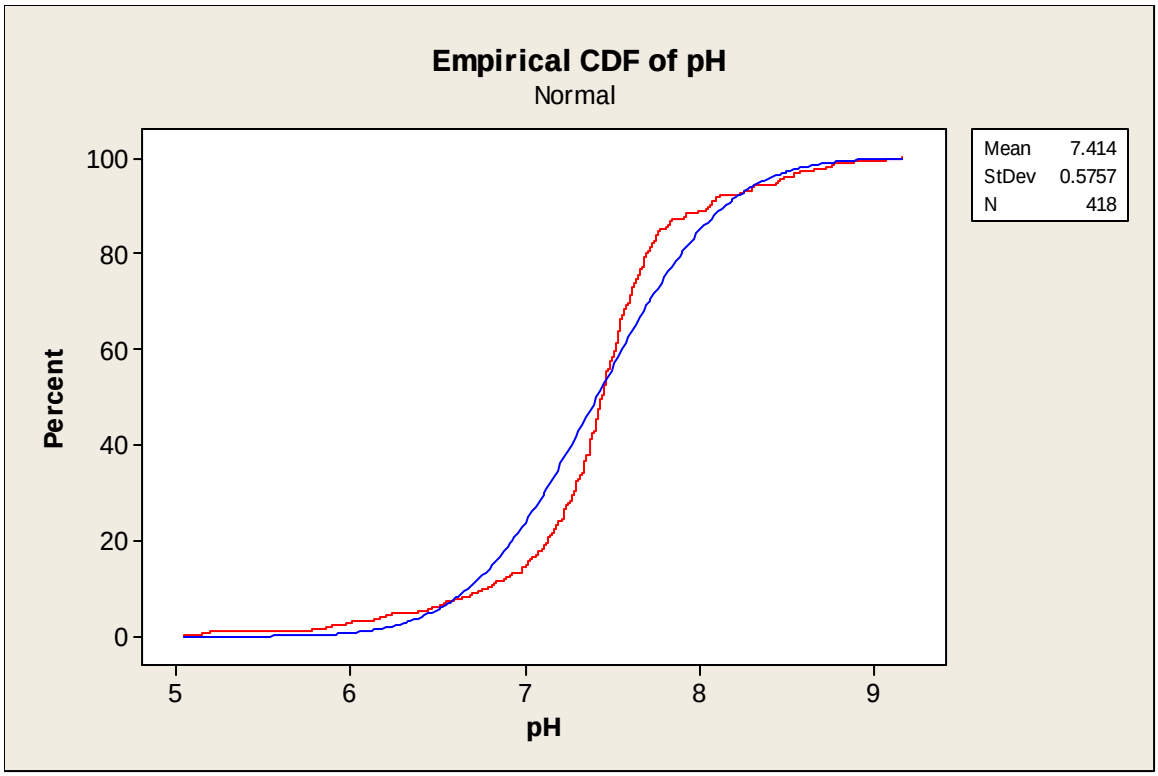
The median, also called the 2nd quartile or 50th percentile, is the middle observation of the data set. The median is determined by ranking the data and locating observation number ([N + 1] / 2). If there are an even number of observations, the median is extrapolated as the value midway between that of observation numbers (N / 2) and ([N / 2] + 1).

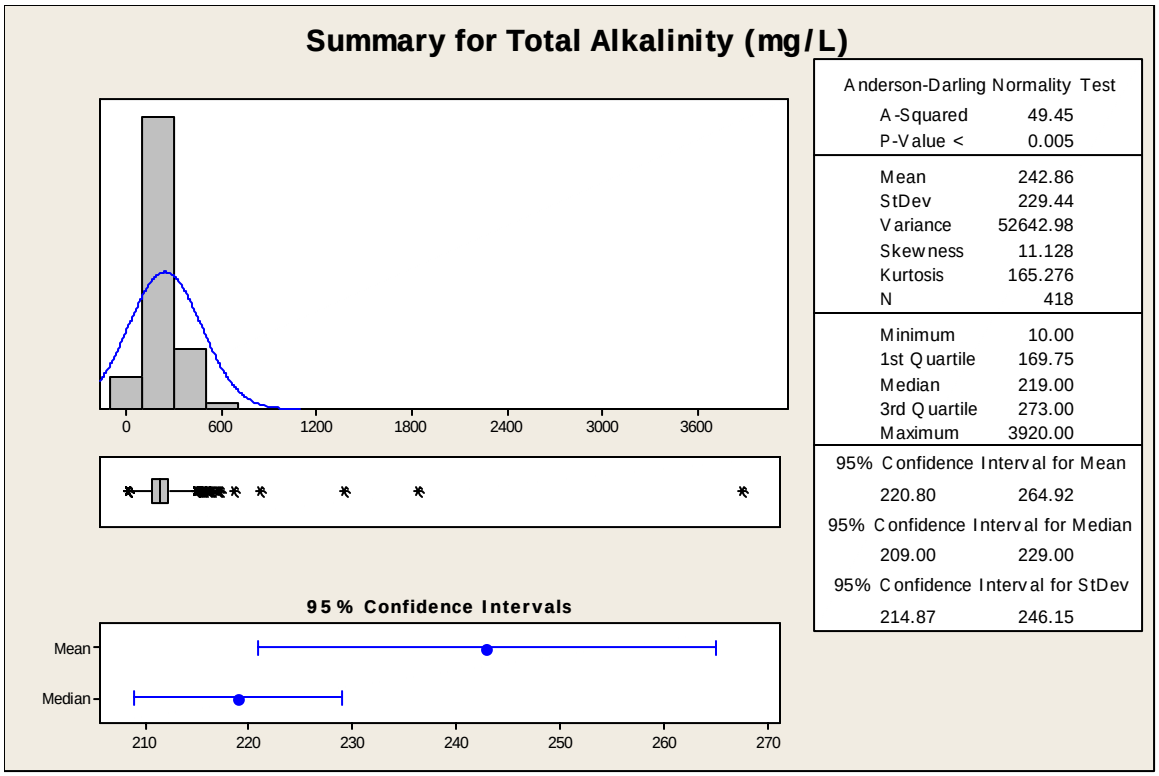
The median is less sensitive to extreme values than the mean. Therefore, the median is often used instead of the mean when data contain outliers, or are skewed.

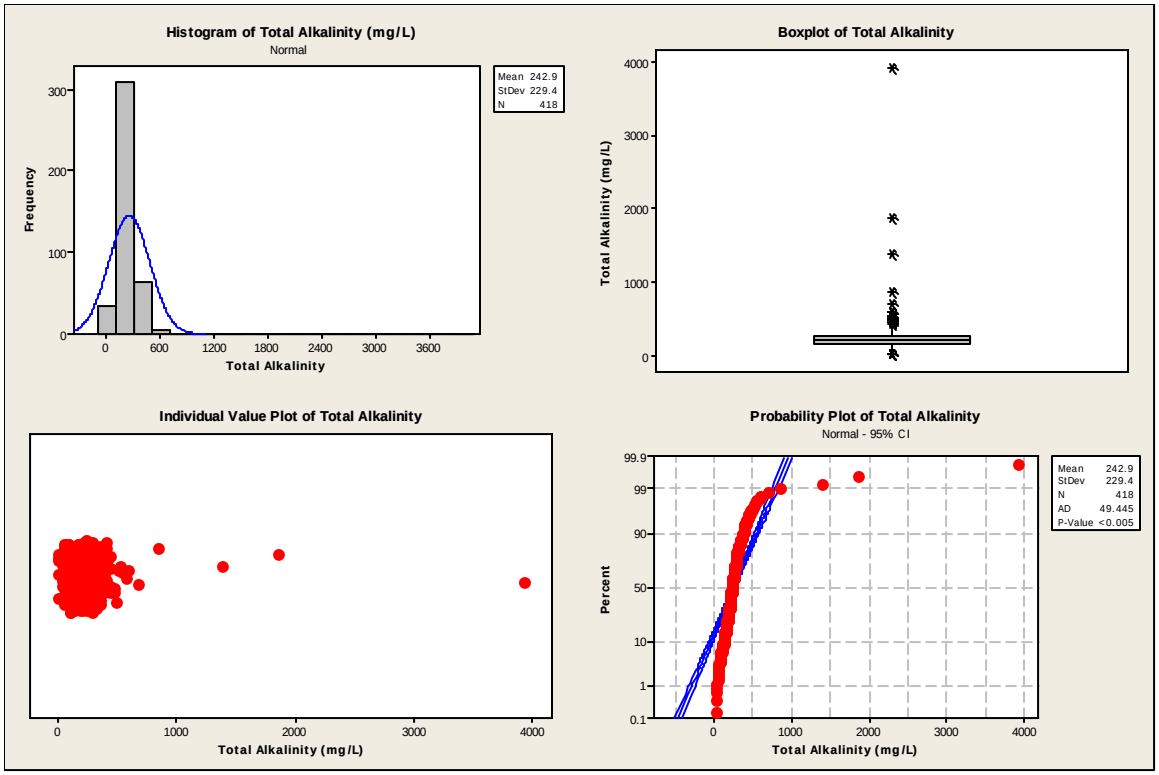
Statistical Tables and Graphs

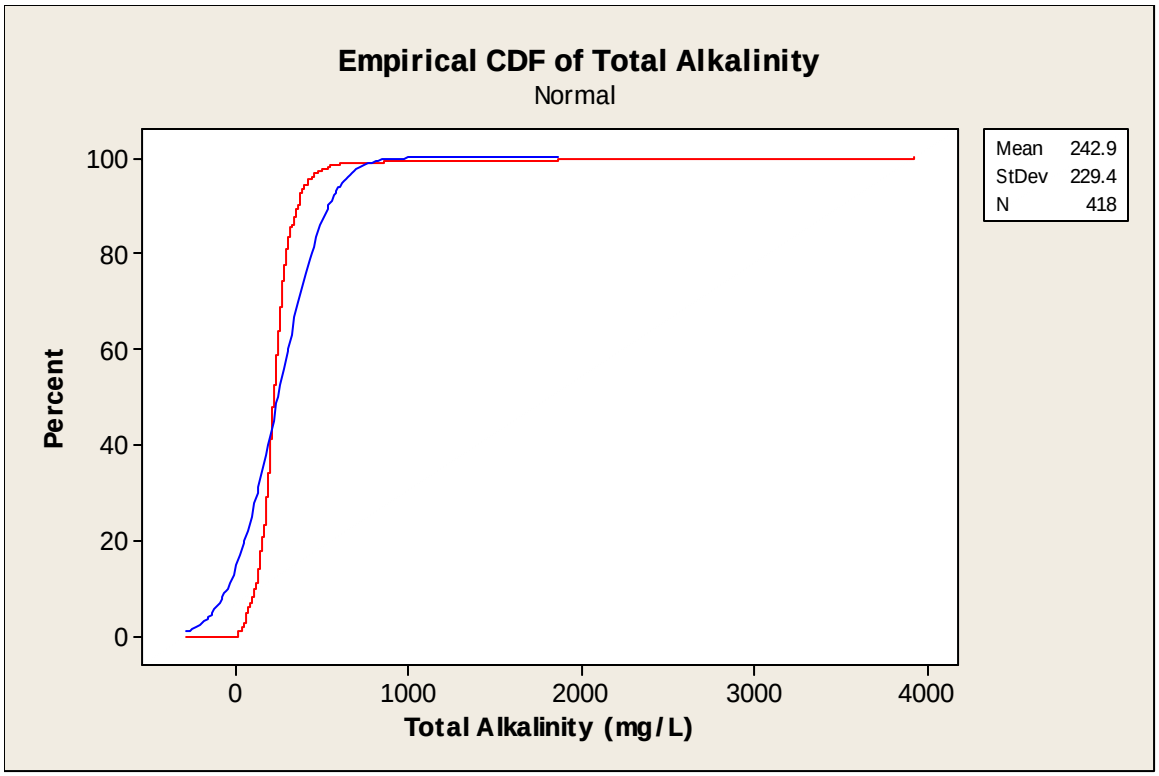

Summary for pH
Anderson-Darling Normality Test
A-Squared 11.77
P-Value < 0.005
Mean 7.4138
StDev 0.5757
Variance 0.3314
Skewness -0.60201
Kurtosis 2.89111
N 418
Minimum 5.0500
1st Quartile 7.2200
Median 7.4400
3rd Quartile 7.6500
Maximum 9.1600
95% Confidence Interval for Mean
7.3585 7.4692
95% Confidence Interval for Median
7.4100 7.4700
95% Confidence Interval for StDev
0.5391 0.6176
5.4 6.0 6.6 7.2 7.8 8.4 9.0
95% Confidence Intervals
Mean
Median
7.350 7.375 7.400 7.425 7.450 7.475


Histogram of pH
Normal
Mean 7.414
StDev 0.5757
N 418
Frequency
pH
Boxplot of pH
pH
Individual Value Plot of pH
pH
Probability Plot of pH
Normal - 95% CI
Mean 7.414
StDev 0.5757
N 418
AD 11.772
P-Value <0.005
Percent
pH

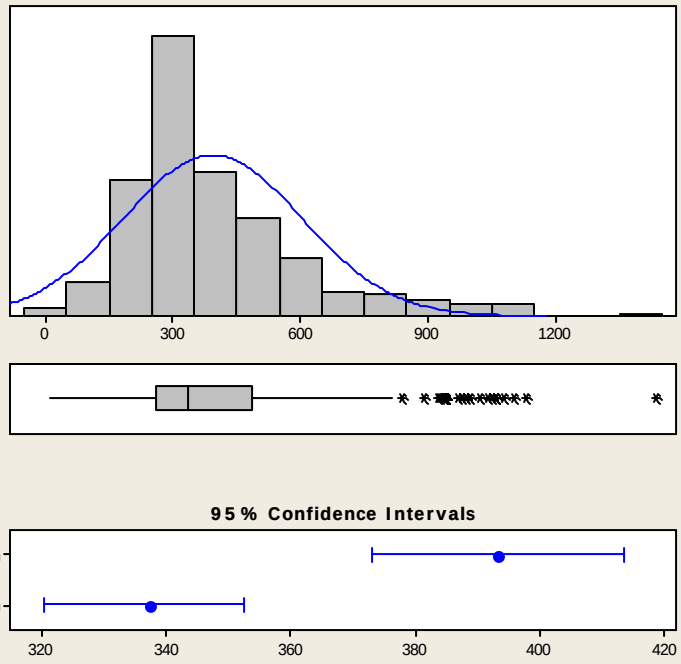

Empirical CDF of pH
Normal
Mean 7.414
StDev 0.5757
N 418
Percent
100
80
60
40
20
0
5
6
7
8
9
pH


Summary for Total Alkalinity (mg/L)
Anderson-Darling Normality Test
A-Squared 49.45
P-Value < 0.005
Mean 242.86
StDev 229.44
Variance 52642.98
Skewness 11.128
Kurtosis 165.276
N 418
Minimum 10.00
1st Quartile 169.75
Median 219.00
3rd Quartile 273.00
Maximum 3920.00
95% Confidence Interval for Mean
220.80 264.92
95% Confidence Interval for Median
209.00 229.00
95% Confidence Interval for StDev
214.87 246.15
0 600 1200 1800 2400 3000 3600
95% Confidence Intervals
Mean
Median
210 220 230 240 250 260 270

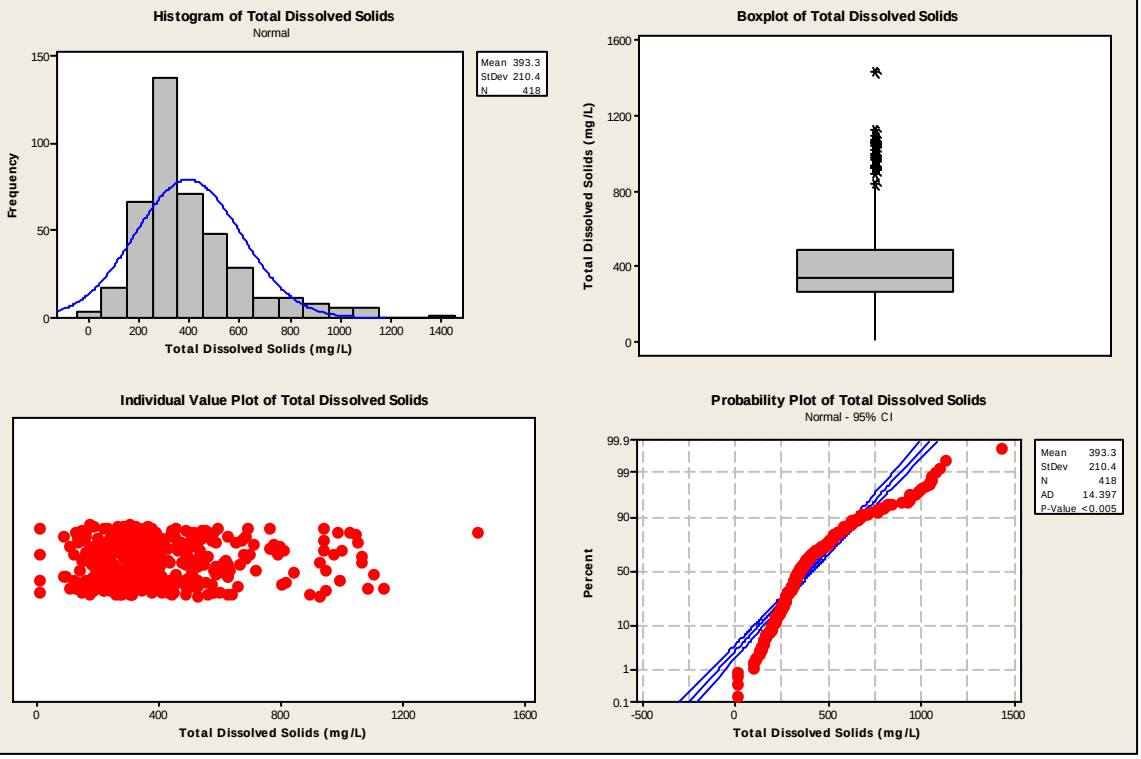

Histogram of Total Alkalinity (mg/L)
Normal
Mean 242.9
StDev 229.4
N 418
Frequency
Total Alkalinity
Boxplot of Total Alkalinity
Total Alkalinity (mg/L)
Individual Value Plot of Total Alkalinity
Total Alkalinity (mg/L)
Probability Plot of Total Alkalinity
Normal - 95% CI
Mean 242.9
StDev 229.4
N 418
AD 49.445
P-Value <0.005
Percent
Total Alkalinity (mg/L)

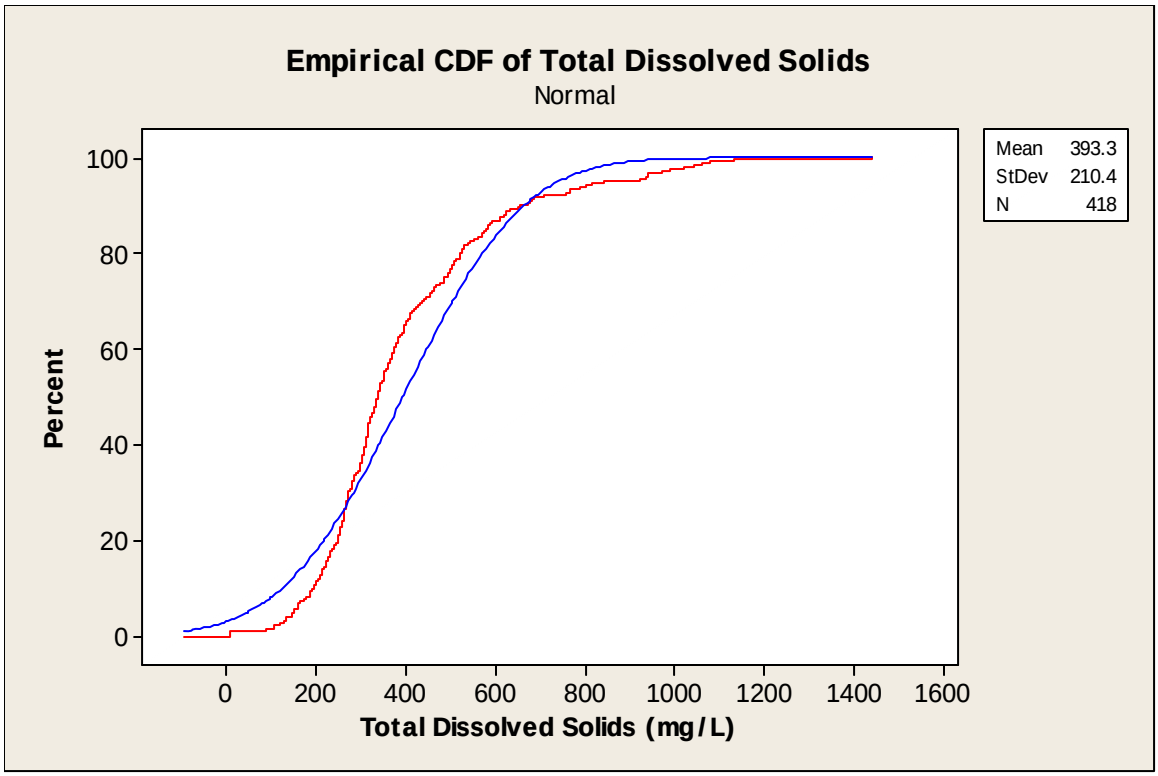

Empirical CDF of Total Alkalinity
Normal
Mean 242.9
StDev 229.4
N 418
Percent
100
80
60
40
20
0
0
1000
2000
3000
4000
Total Alkalinity (mg/L)

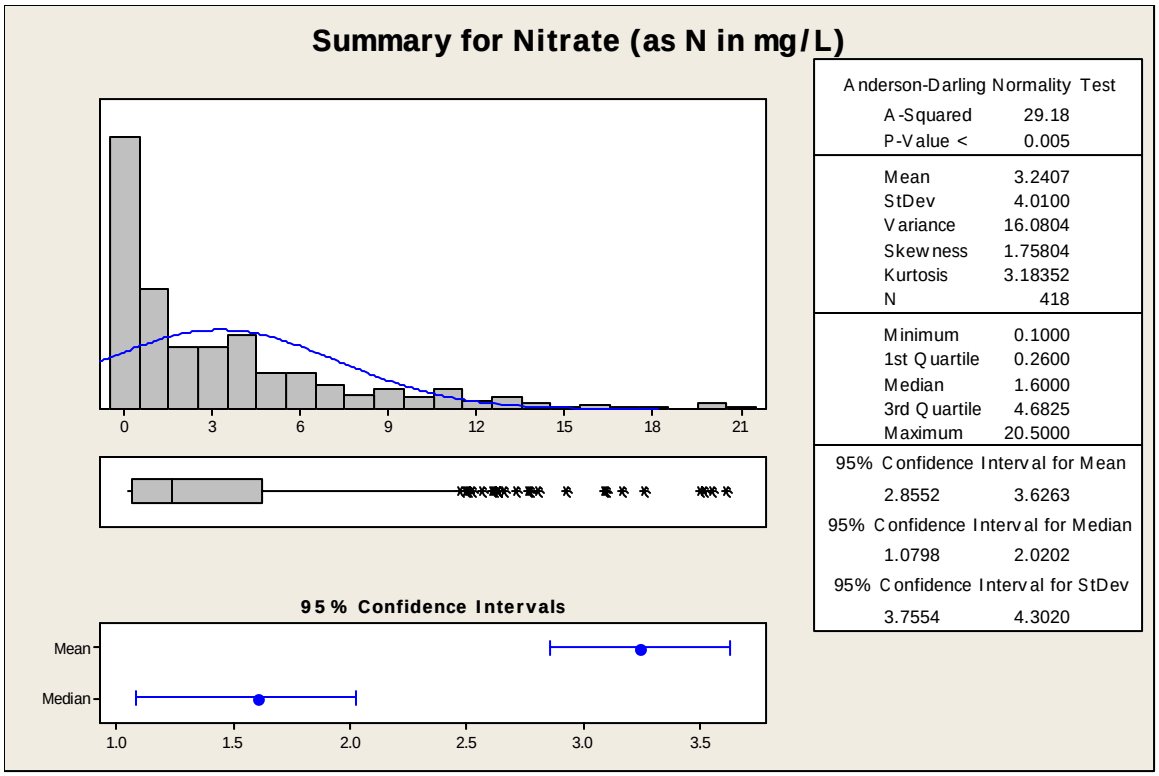
Summary for Total Dissolved Solids (mg/L)

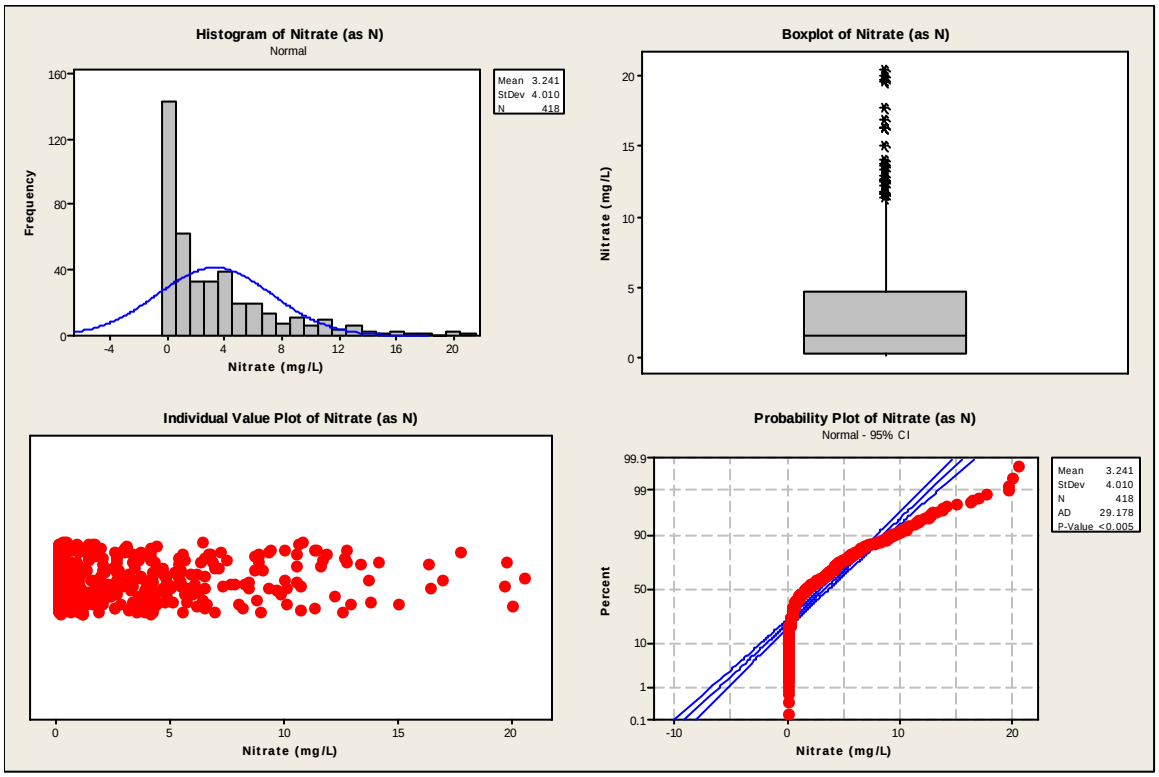


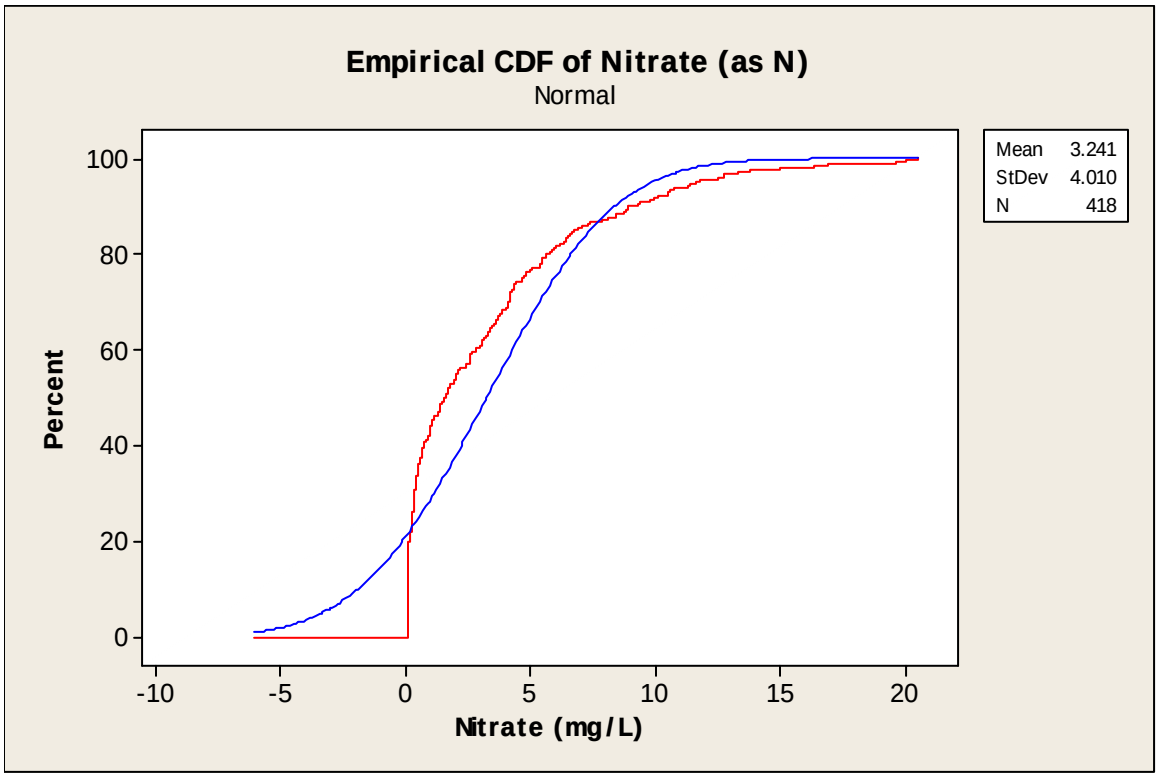
Anderson-Darling Normality Test	
A-Squared	14.40
P-Value <	0.005
Mean	393.32
StDev	210.39
Variance	44265.37
Skewness	1.49435
Kurtosis	2.84044
N	418
Minimum	10.00
1st Quartile	261.75
Median	337.50
3rd Quartile	488.25
Maximum	1438.00
95% Confidence Interval for Mean	
373.09	413.55
95% Confidence Interval for Median	
320.49	352.51
95% Confidence Interval for StDev	
197.03	225.71

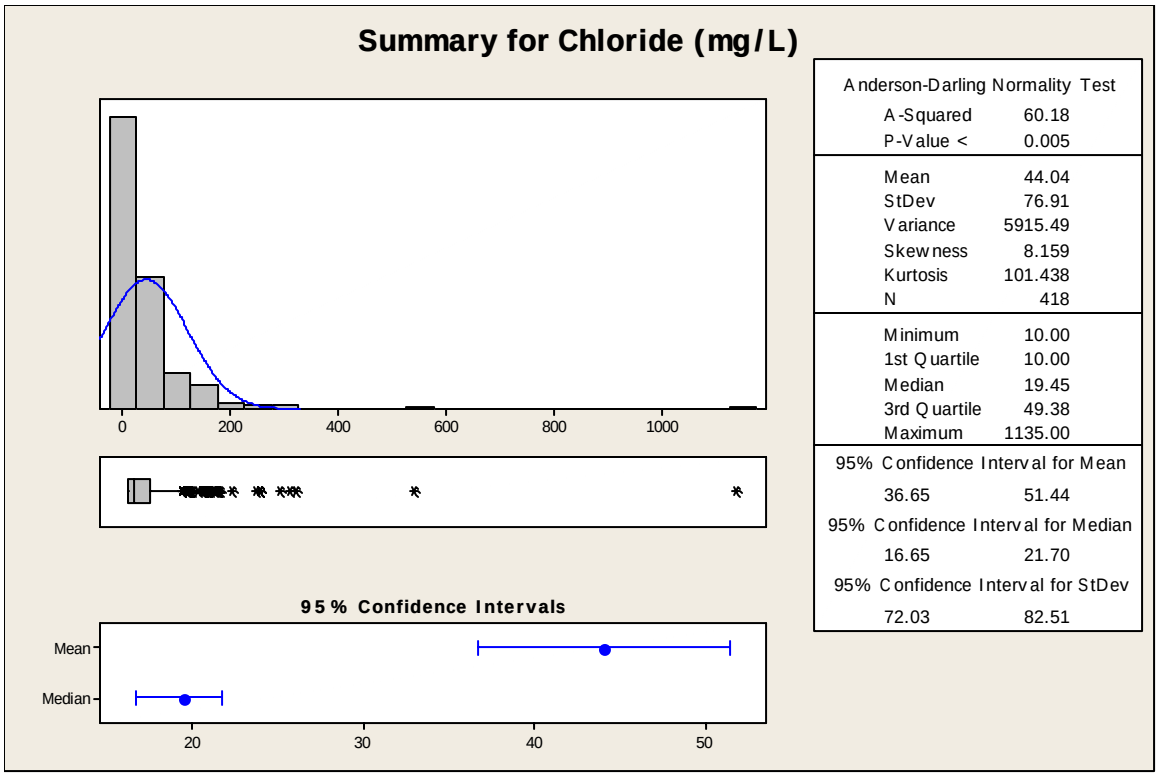


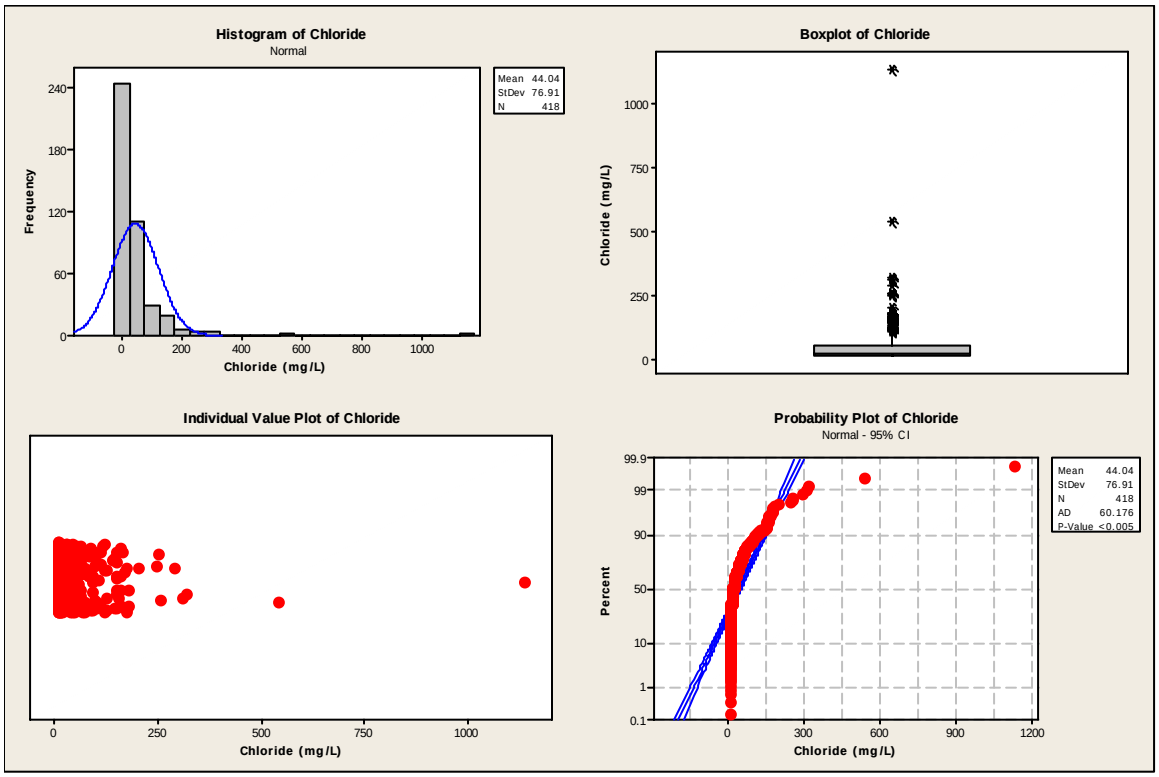

Empirical CDF of Total Dissolved Solids
Normal
Mean 393.3
StDev 210.4
N 418
Percent
100
80
60
40
20
0
0
200
400
600
800
1000
1200
1400
1600
Total Dissolved Solids (mg/L)

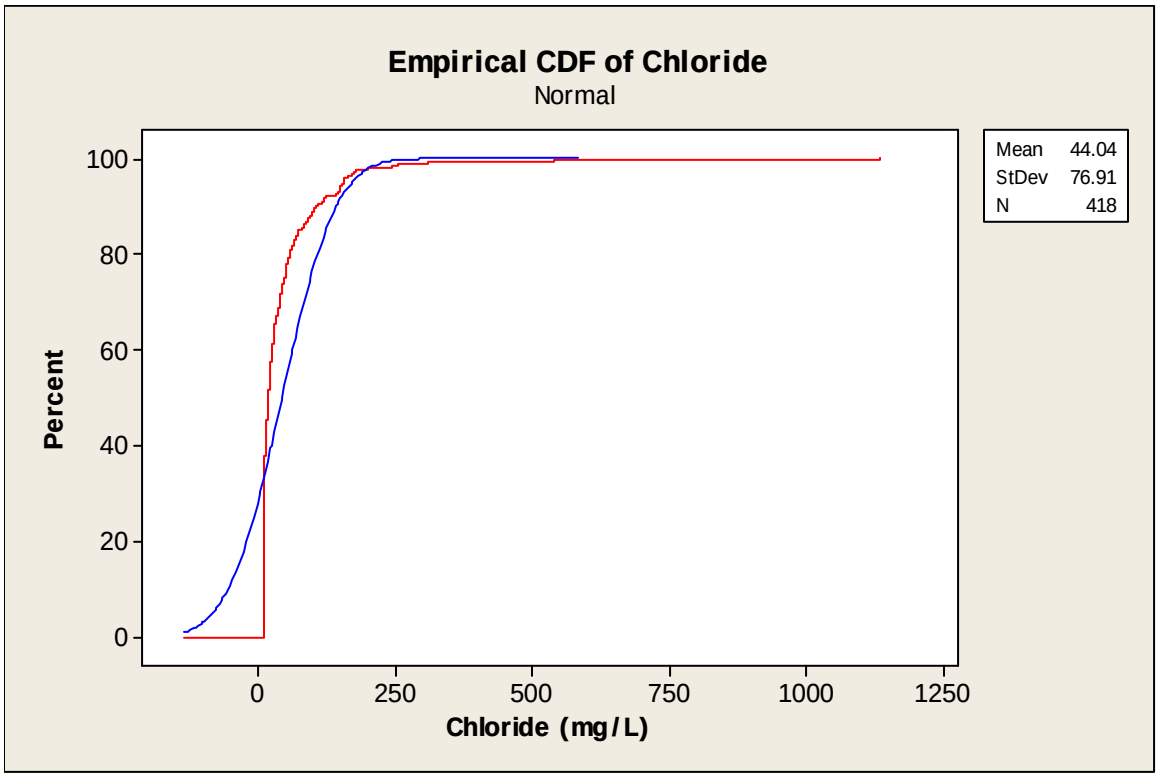

Summary for Nitrate (as N in mg/L)
Anderson-Darling Normality Test
A-Squared 29.18
P-Value < 0.005
Mean 3.2407
StDev 4.0100
Variance 16.0804
Skewness 1.75804
Kurtosis 3.18352
N 418
Minimum 0.1000
1st Quartile 0.2600
Median 1.6000
3rd Quartile 4.6825
Maximum 20.5000
95% Confidence Interval for Mean
2.8552 3.6263
95% Confidence Interval for Median
1.0798 2.0202
95% Confidence Interval for StDev
3.7554 4.3020
95% Confidence Intervals
Mean
Median
0 3 6 9 12 15 18 21
1.0 1.5 2.0 2.5 3.0 3.5


Histogram of Nitrate (as N)
Normal
Mean 3.241
StDev 4.010
N 418
Frequency
Nitrate (mg/L)
Boxplot of Nitrate (as N)
Nitrate (mg/L)
Individual Value Plot of Nitrate (as N)
Nitrate (mg/L)
Probability Plot of Nitrate (as N)
Normal - 95% CI
Mean 3.241
StDev 4.010
N 418
AD 29.178
P-Value <0.005
Percent
Nitrate (mg/L)

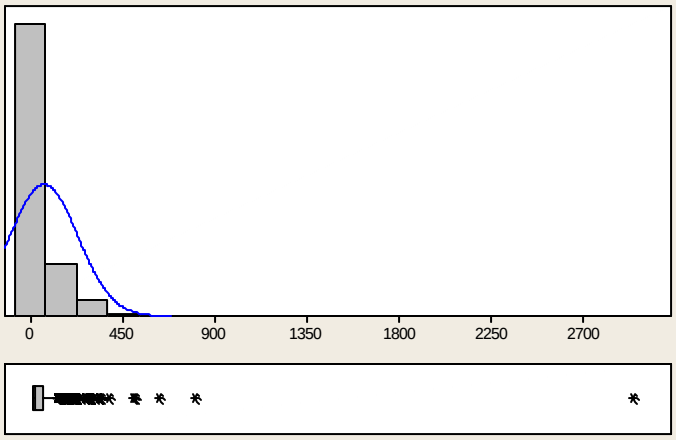

Empirical CDF of Nitrate (as N)
Normal
Mean 3.241
StDev 4.010
N 418
Percent
100
80
60
40
20
0
-10
-5
0
5
10
15
20
Nitrate (mg/L)

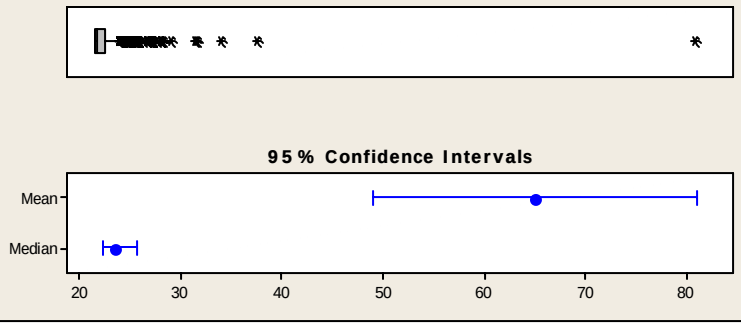

Summary for Chloride (mg/L)
Anderson-Darling Normality Test
A-Squared 60.18
P-Value < 0.005
Mean 44.04
StDev 76.91
Variance 5915.49
Skewness 8.159
Kurtosis 101.438
N 418
Minimum 10.00
1st Quartile 10.00
Median 19.45
3rd Quartile 49.38
Maximum 1135.00
95% Confidence Interval for Mean
36.65 51.44
95% Confidence Interval for Median
16.65 21.70
95% Confidence Interval for StDev
72.03 82.51
0 200 400 600 800 1000
95% Confidence Intervals
Mean
Median
20 30 40 50


Histogram of Chloride
Normal
Mean 44.04
StDev 76.91
N 418
Frequency
Chloride (mg/L)
Boxplot of Chloride
Chloride (mg/L)
Individual Value Plot of Chloride
Chloride (mg/L)
Probability Plot of Chloride
Normal - 95% CI
Mean 44.04
StDev 76.91
N 418
AD 60.176
P-Value < 0.005
Percent
Chloride (mg/L)

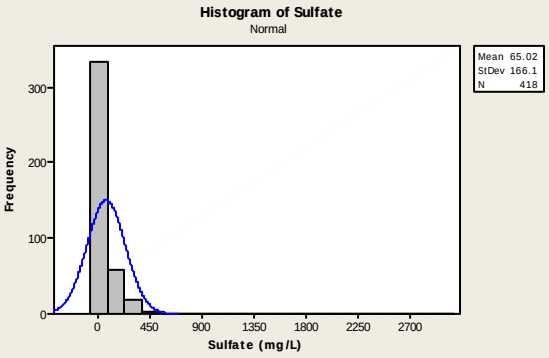

Empirical CDF of Chloride
Normal
Mean 44.04
StDev 76.91
N 418
Percent
100
80
60
40
20
0
0
250
500
750
1000
1250
Chloride (mg/L)

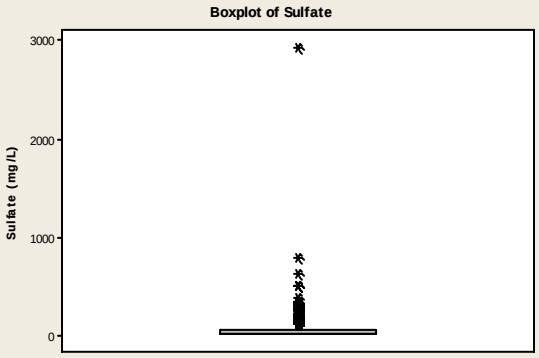
Summary for Sulfate (mg/L)

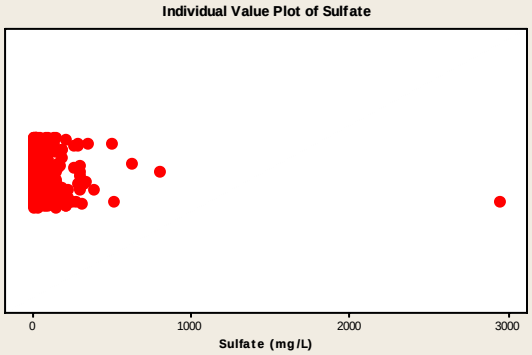


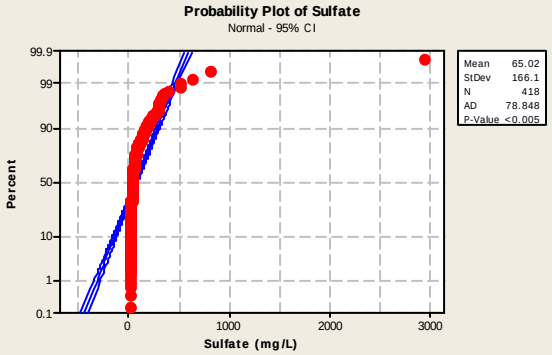


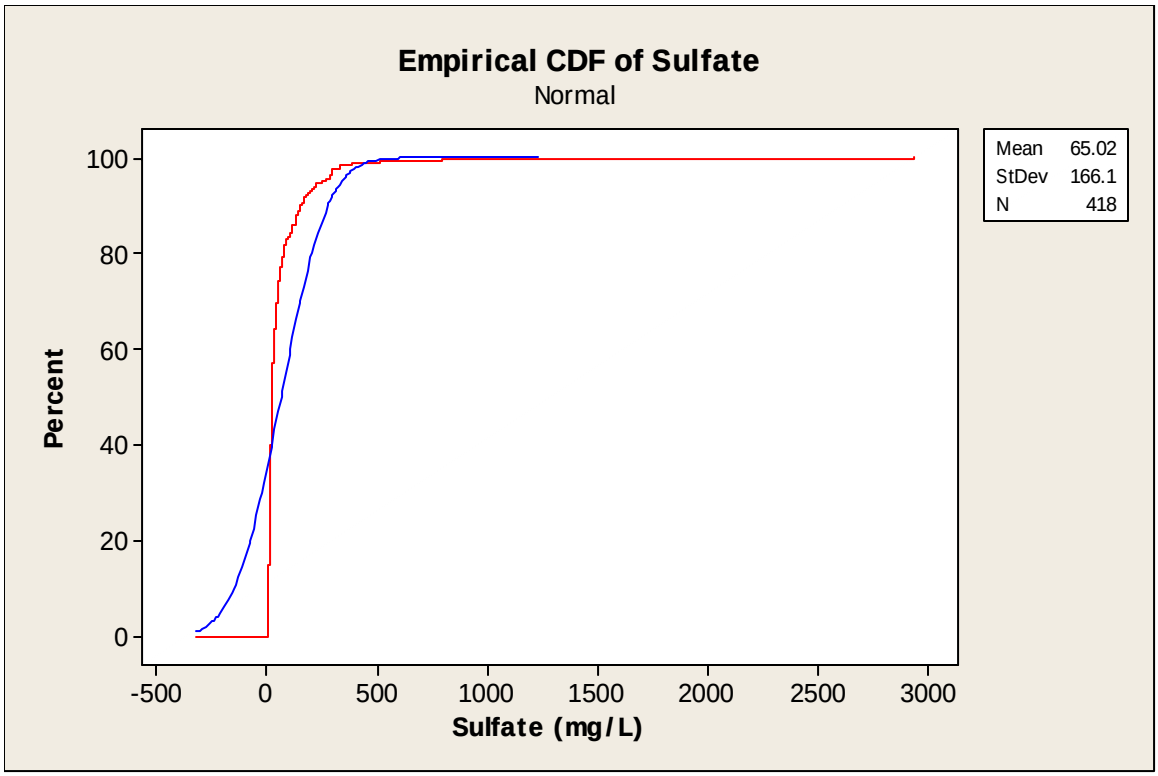
Anderson-Darling Normality Test	
A-Squared	78.85
P-Value <	0.005
Mean	65.02
StDev	166.10
Variance	27587.97
Skewness	12.967
Kurtosis	215.840
N	418
Minimum	10.00
1st Quartile	13.60
Median	23.60
3rd Quartile	58.25
Maximum	2936.00
95% Confidence Interval for Mean	
49.05	80.99
95% Confidence Interval for Median	
22.30	25.70
95% Confidence Interval for StDev	
155.55	178.19



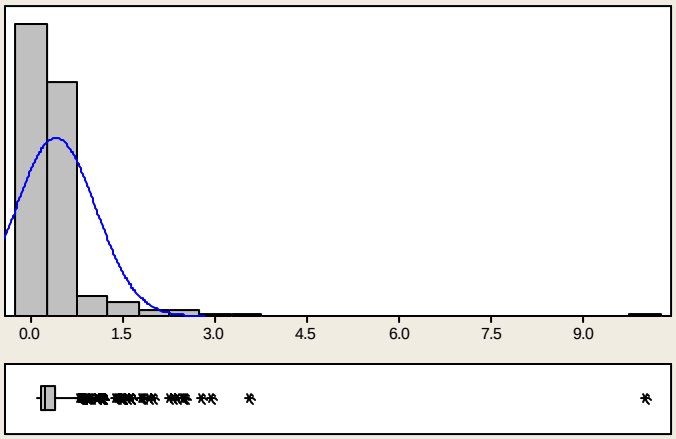


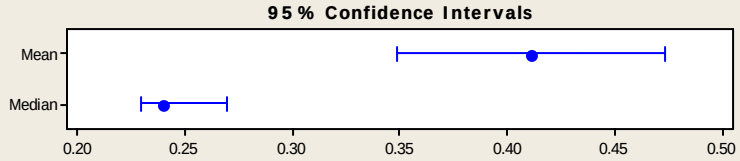




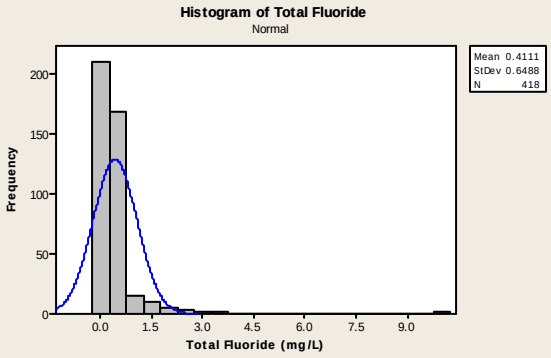

Empirical CDF of Sulfate
Normal
Mean 65.02
StDev 166.1
N 418
Percent
100
80
60
40
20
0
-500
0
500
1000
1500
2000
2500
3000
Sulfate (mg/L)

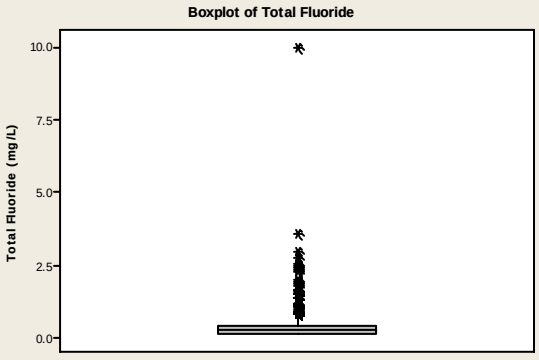
Summary for Total Fluoride (mg/L)

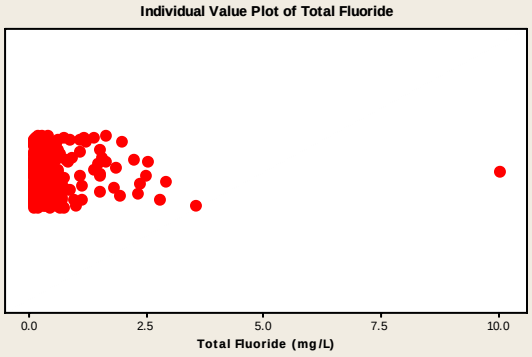


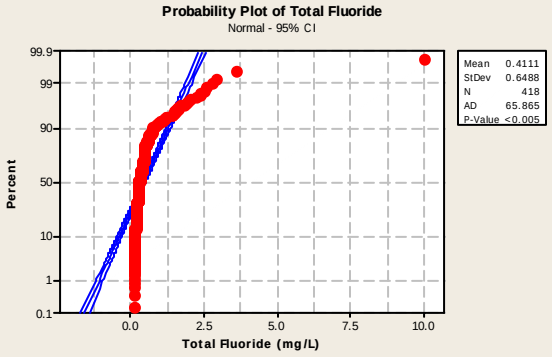


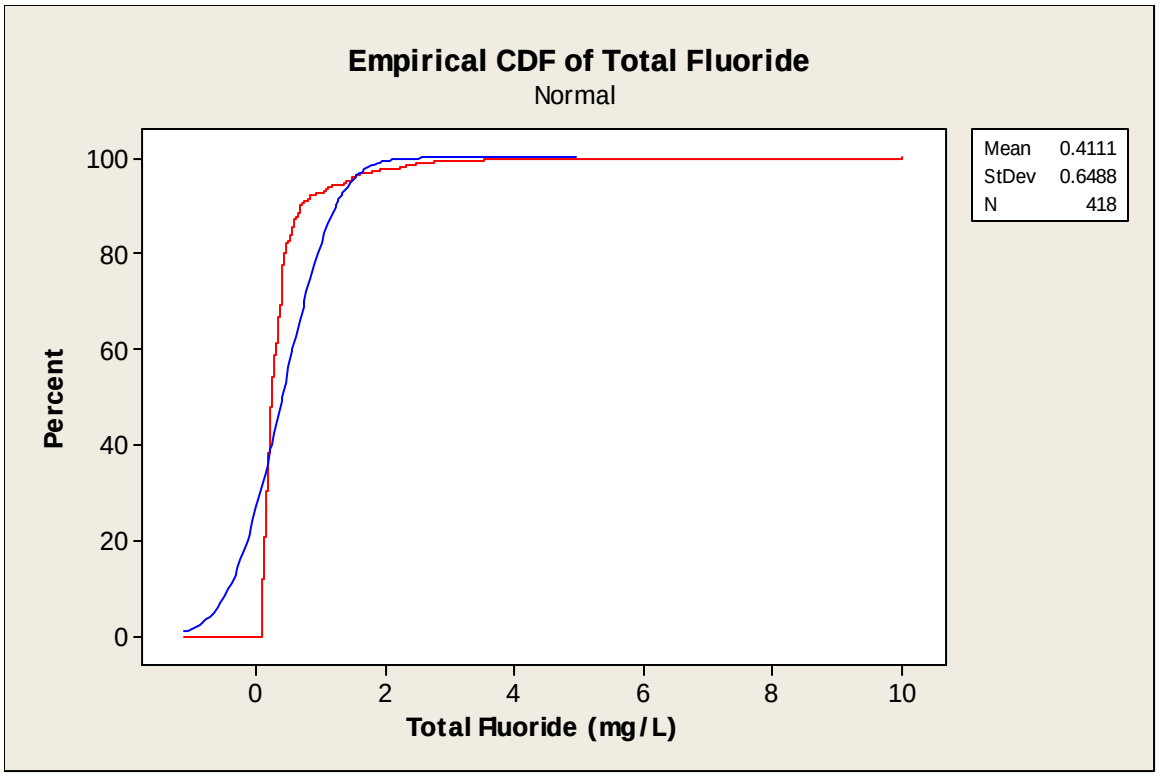
Anderson-Darling Normality Test	
A-Squared	65.86
P-Value <	0.005
Mean	0.4111
StDev	0.6488
Variance	0.4209
Skewness	8.883
Kurtosis	116.939
N	418
Minimum	0.1000
1st Quartile	0.1600
Median	0.2400
3rd Quartile	0.4000
Maximum	10.0000
95% Confidence Interval for Mean	
0.3487	0.4735
95% Confidence Interval for Median	
0.2300	0.2700
95% Confidence Interval for StDev	
0.6076	0.6960



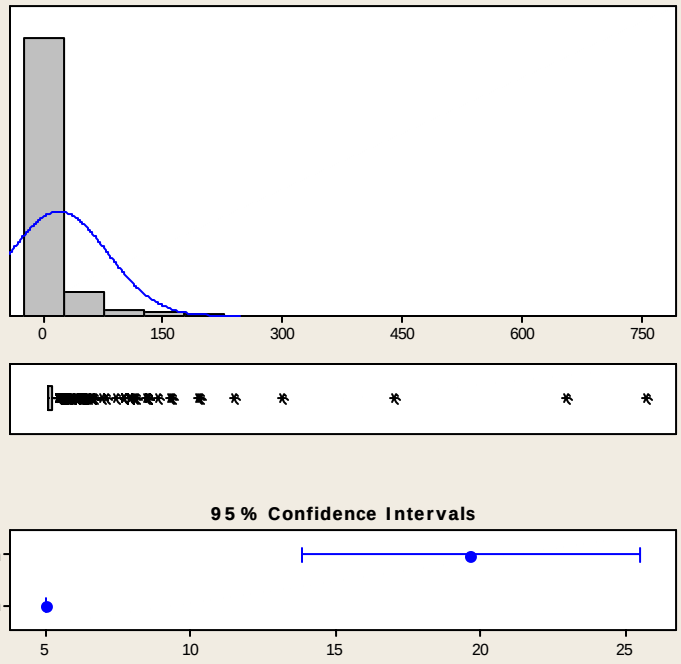




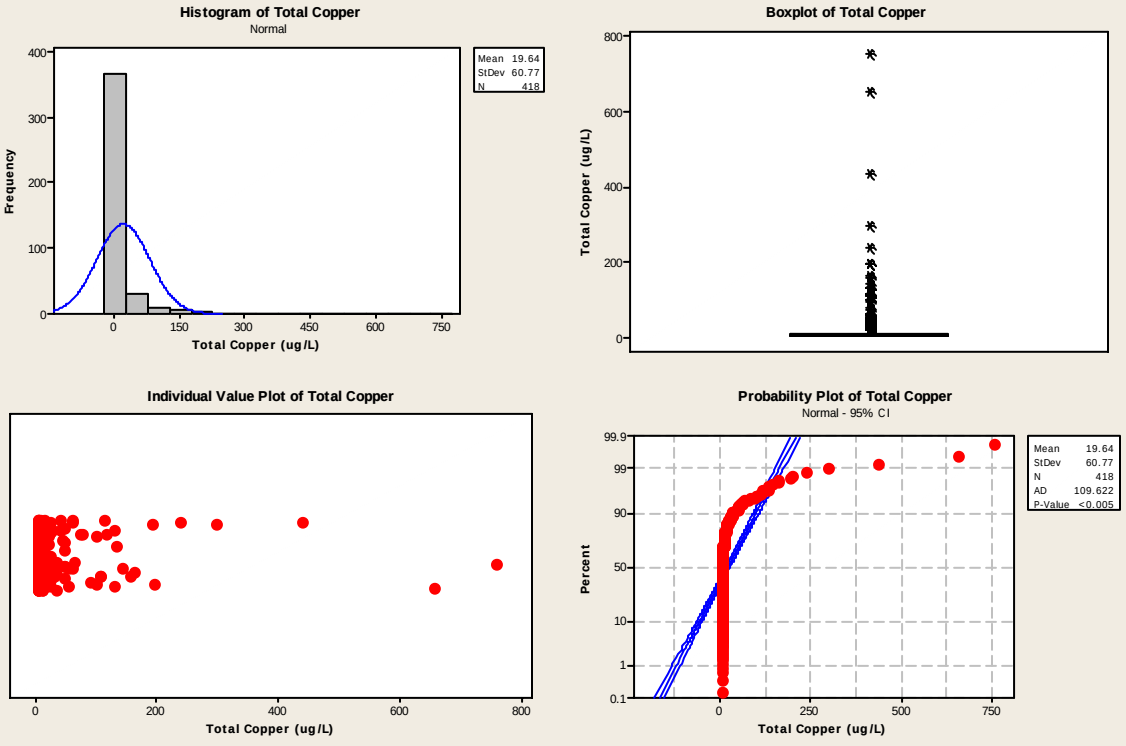


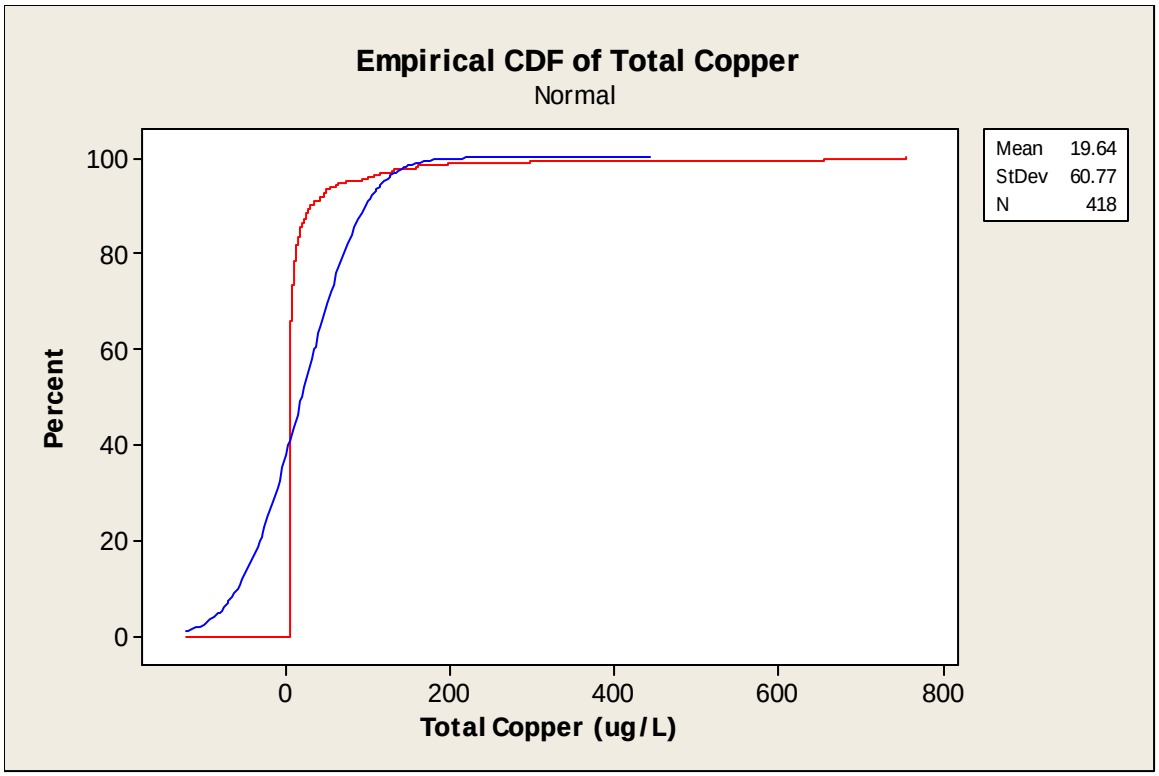

Empirical CDF of Total Fluoride
Normal
Mean 0.4111
StDev 0.6488
N 418
Percent
100
80
60
40
20
0
0
2
4
6
8
10
Total Fluoride (mg/L)

Summary for Total Copper (ug/L)

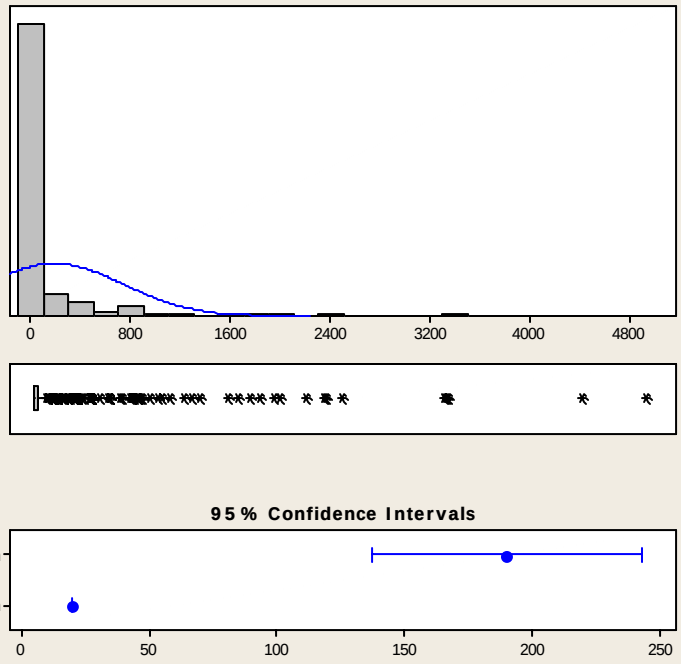


Anderson-Darling Normality Test	
A-Squared	109.62
P-Value <	0.005
Mean	19.642
StDev	60.774
Variance	3693.485
Skewness	8.4408
Kurtosis	85.8100
N	418
Minimum	5.000
1st Quartile	5.000
Median	5.000
3rd Quartile	10.000
Maximum	755.000
95% Confidence Interval for Mean	
13.799	25.485
95% Confidence Interval for Median	
5.000	5.000
95% Confidence Interval for StDev	
56.915	65.199

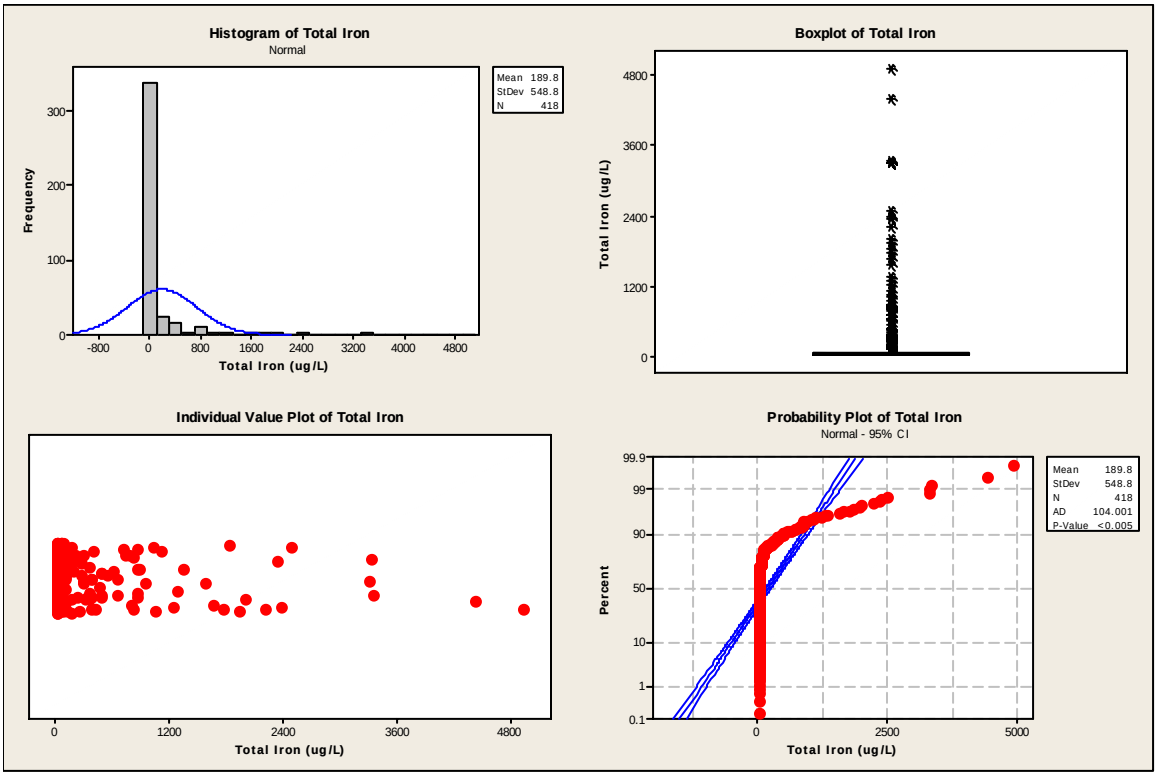


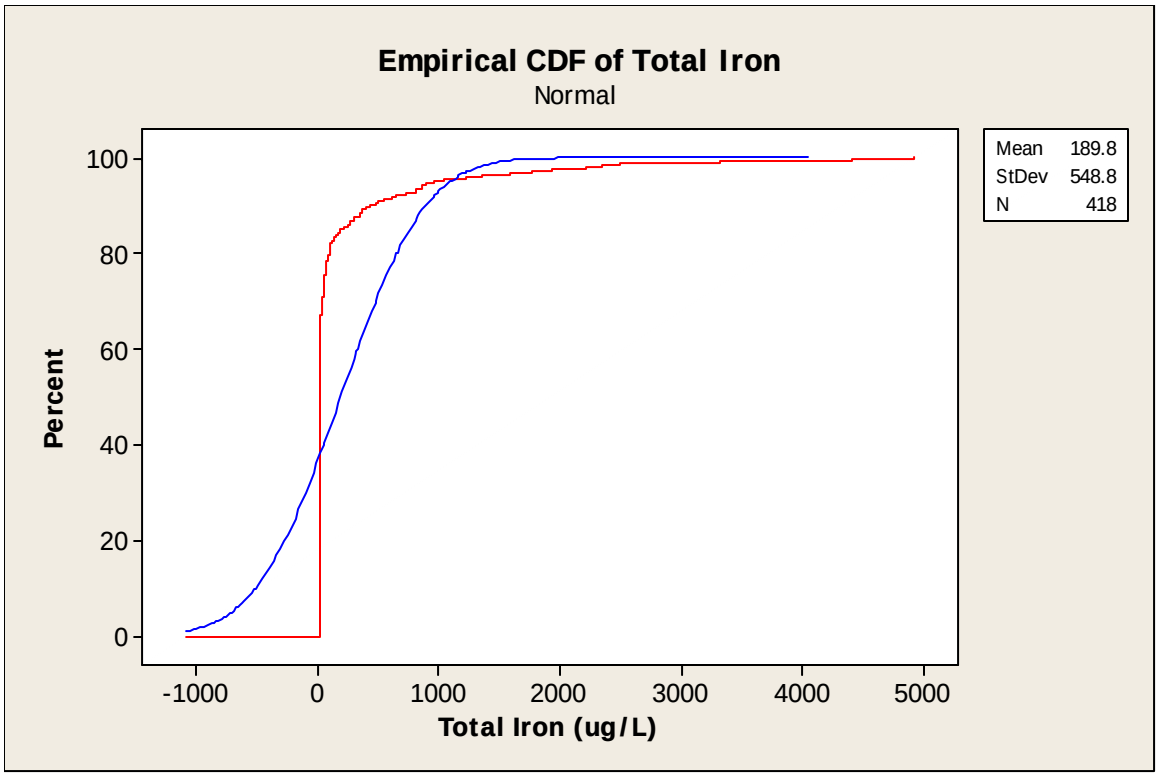

Empirical CDF of Total Copper
Normal
Mean 19.64
StDev 60.77
N 418
Percent
100
80
60
40
20
0
0
200
400
600
800
Total Copper (ug/L)

Summary for Total Iron (ug/L)

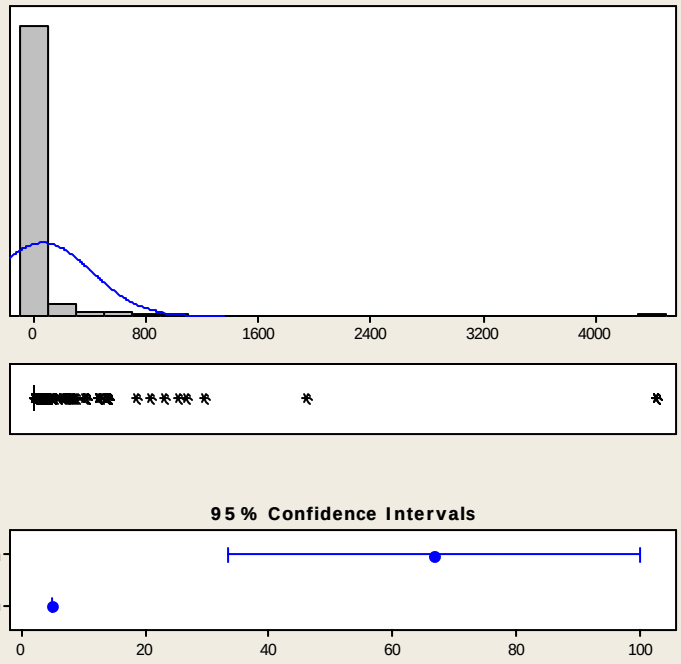


Anderson-Darling Normality Test	
A-Squared	104.00
P-Value <	0.005
Mean	189.82
StDev	548.79
Variance	301170.64
Skewness	5.0697
Kurtosis	30.4069
N	418
Minimum	20.00
1st Quartile	20.00
Median	20.00
3rd Quartile	60.18
Maximum	4920.00
95% Confidence Interval for Mean	
137.06	242.59
95% Confidence Interval for Median	
20.00	20.00
95% Confidence Interval for StDev	
513.94	588.75

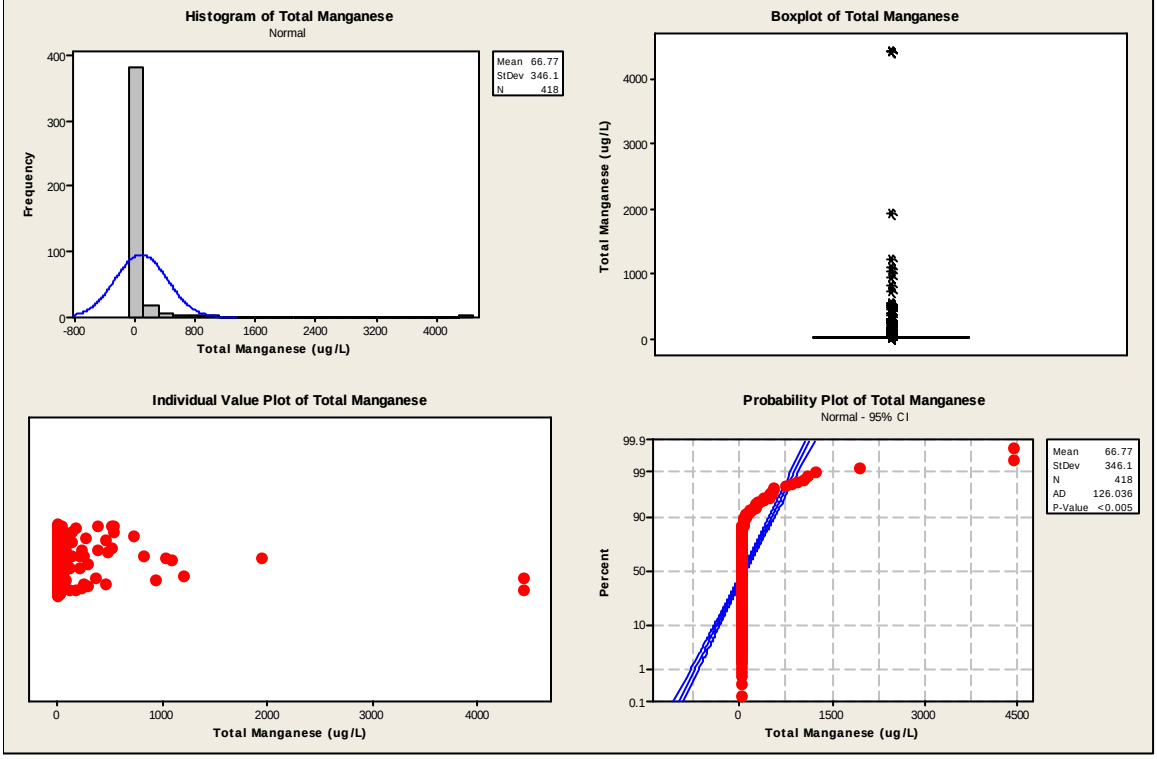


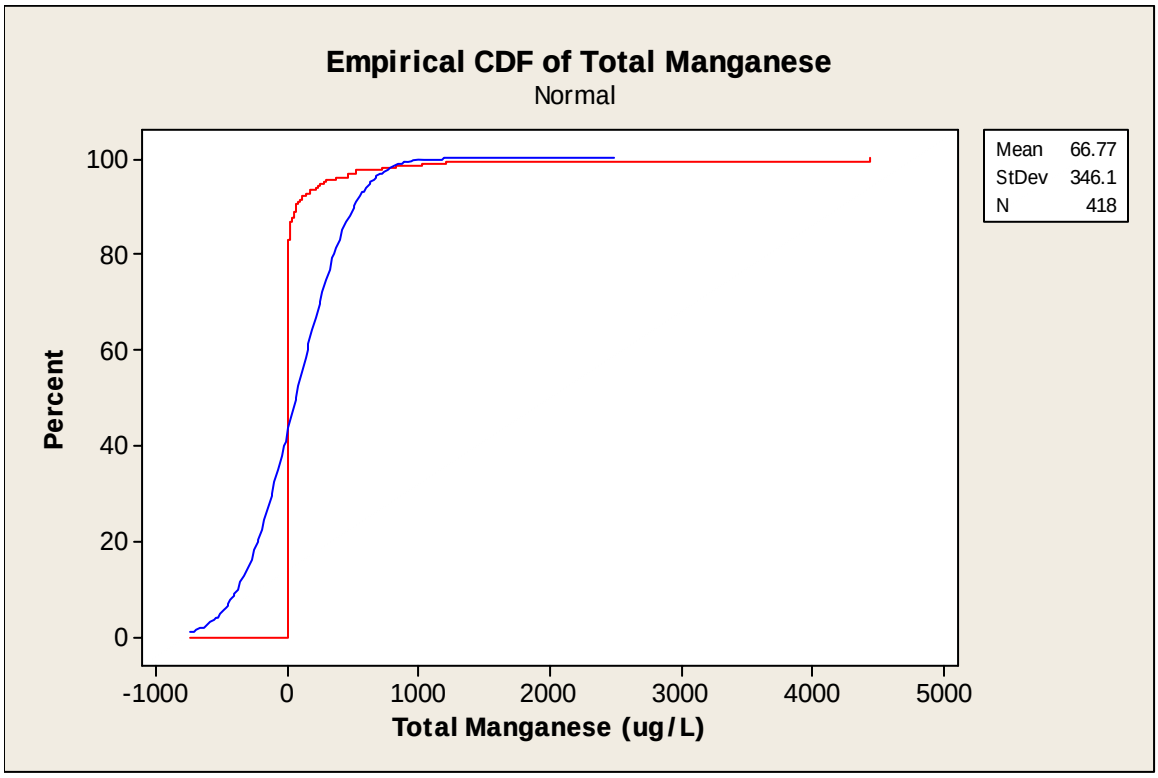

Empirical CDF of Total Iron
Normal
Mean 189.8
StDev 548.8
N 418
Percent
100
80
60
40
20
0
-1000
0
1000
2000
3000
4000
5000
Total Iron (ug/L)

Summary for Total Manganese (ug/L)

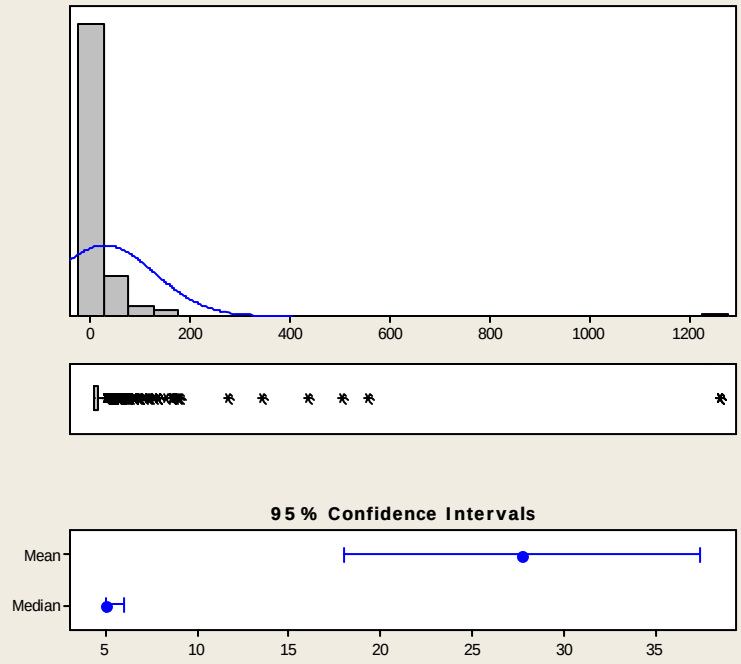


Anderson-Darling Normality Test	
A-Squared	126.04
P-Value <	0.005
Mean	66.77
StDev	346.10
Variance	119787.09
Skewness	10.357
Kurtosis	122.825
N	418
Minimum	5.00
1st Quartile	5.00
Median	5.00
3rd Quartile	5.00
Maximum	4430.00
95% Confidence Interval for Mean	
33.49	100.04
95% Confidence Interval for Median	
5.00	5.00
95% Confidence Interval for StDev	
324.12	371.30

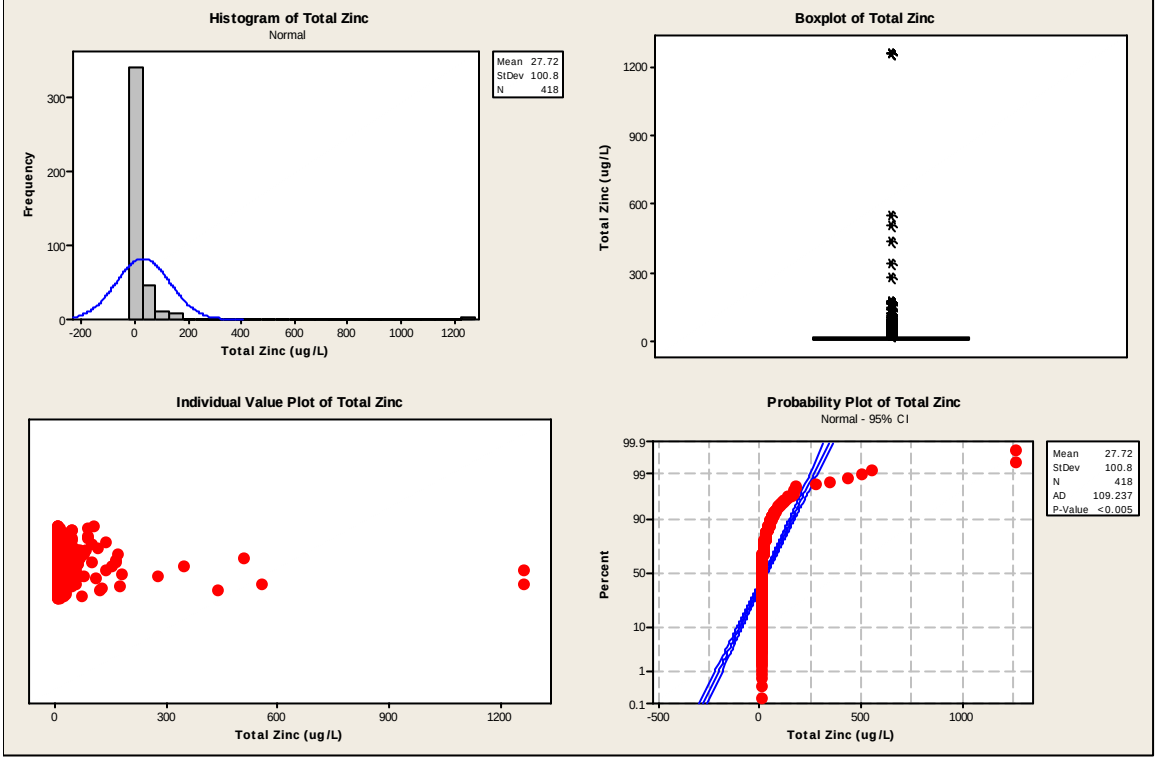


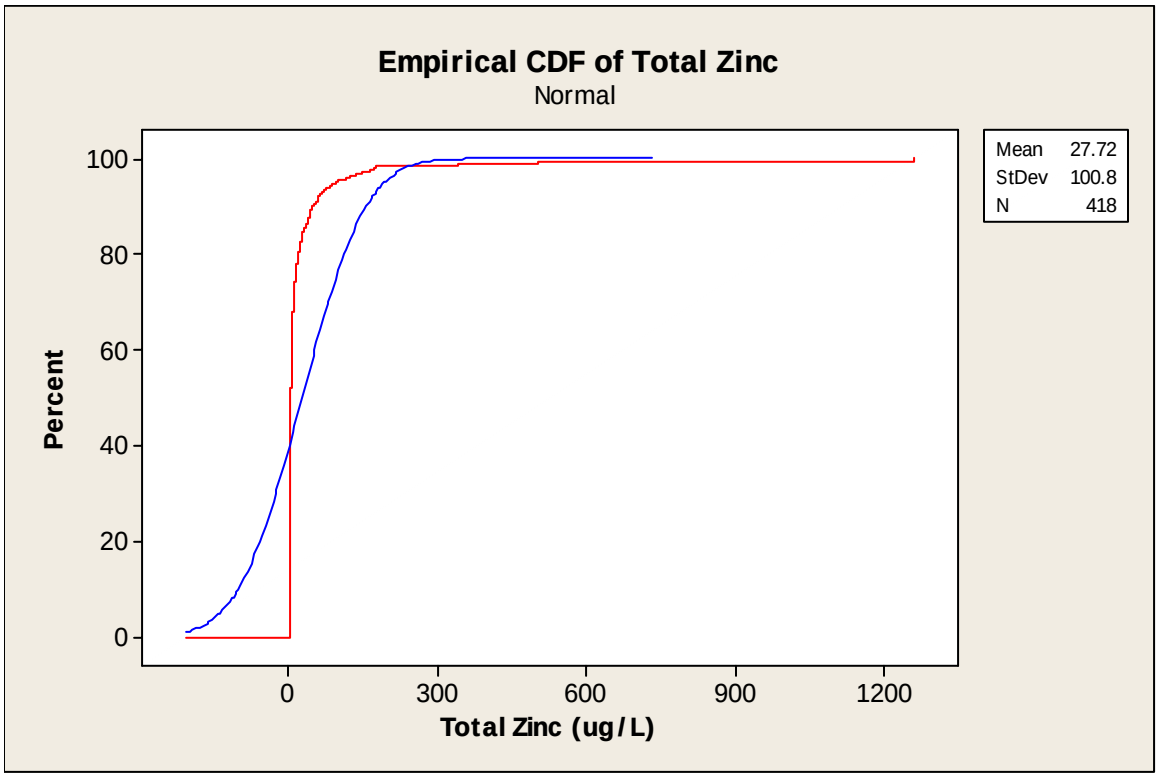

Empirical CDF of Total Manganese
Normal
Mean 66.77
StDev 346.1
N 418
Percent
100
80
60
40
20
0
-1000
0
1000
2000
3000
4000
5000
Total Manganese (ug/L)

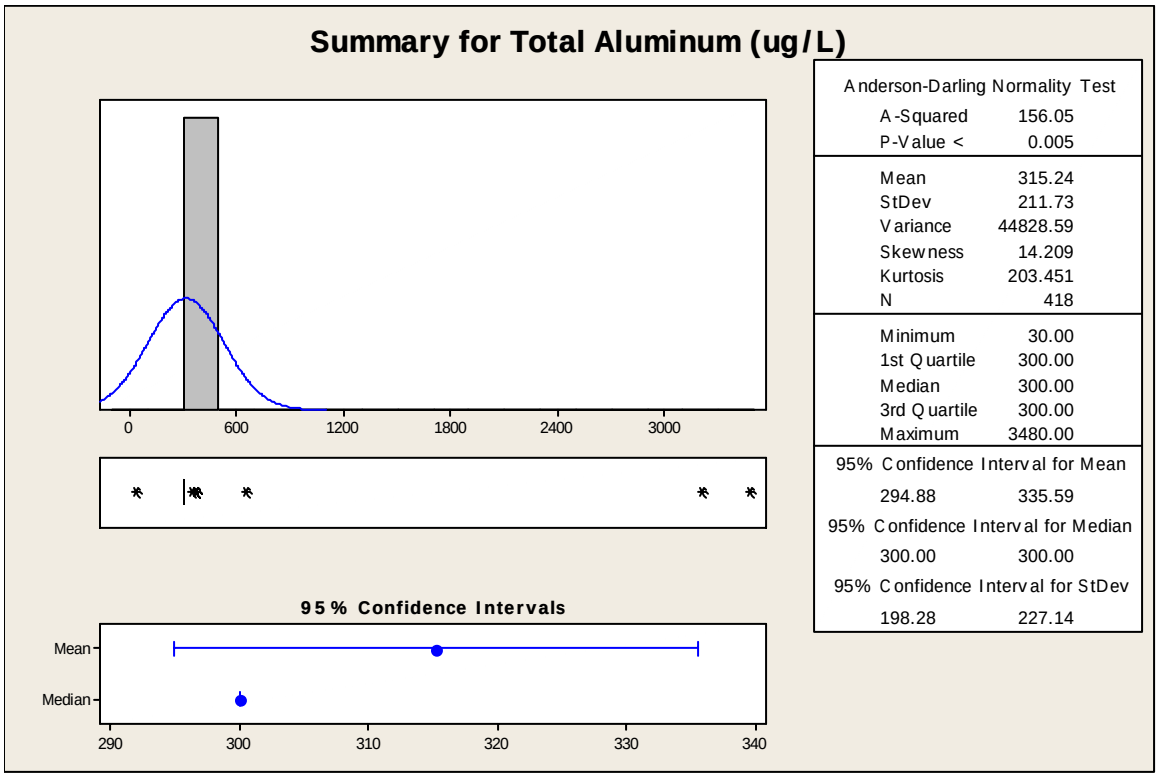
Summary for Total Zinc (ug/L)

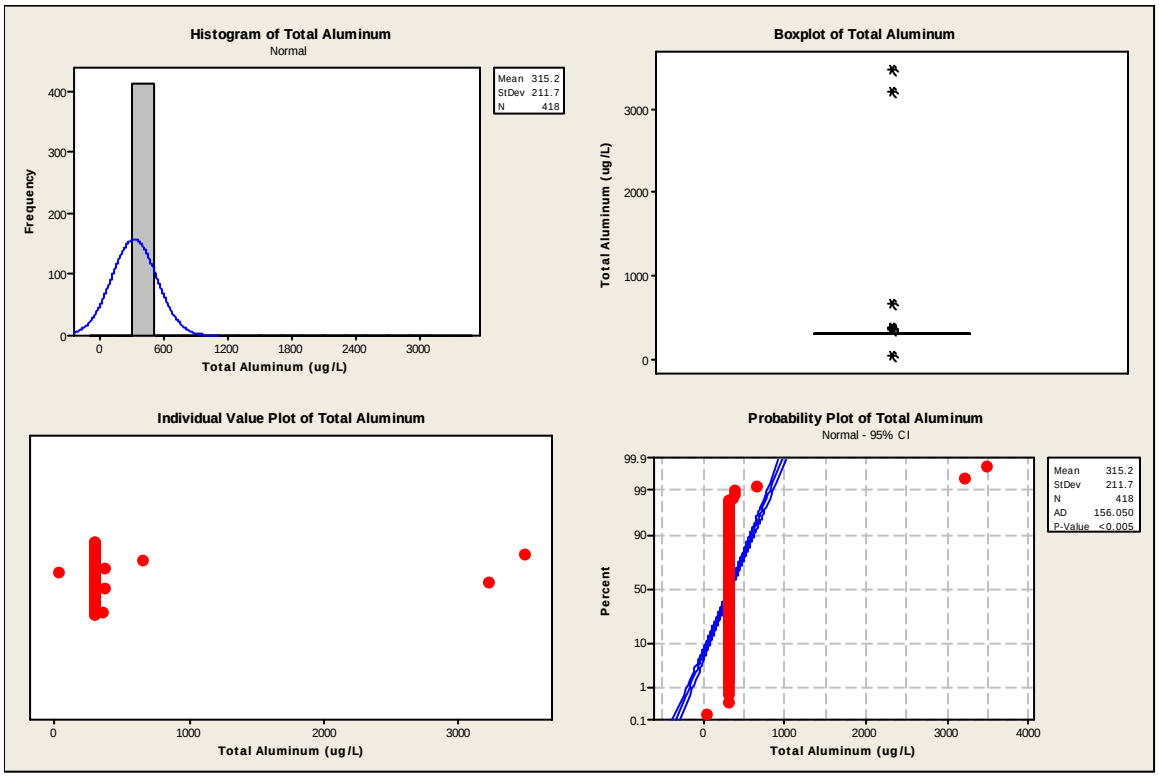


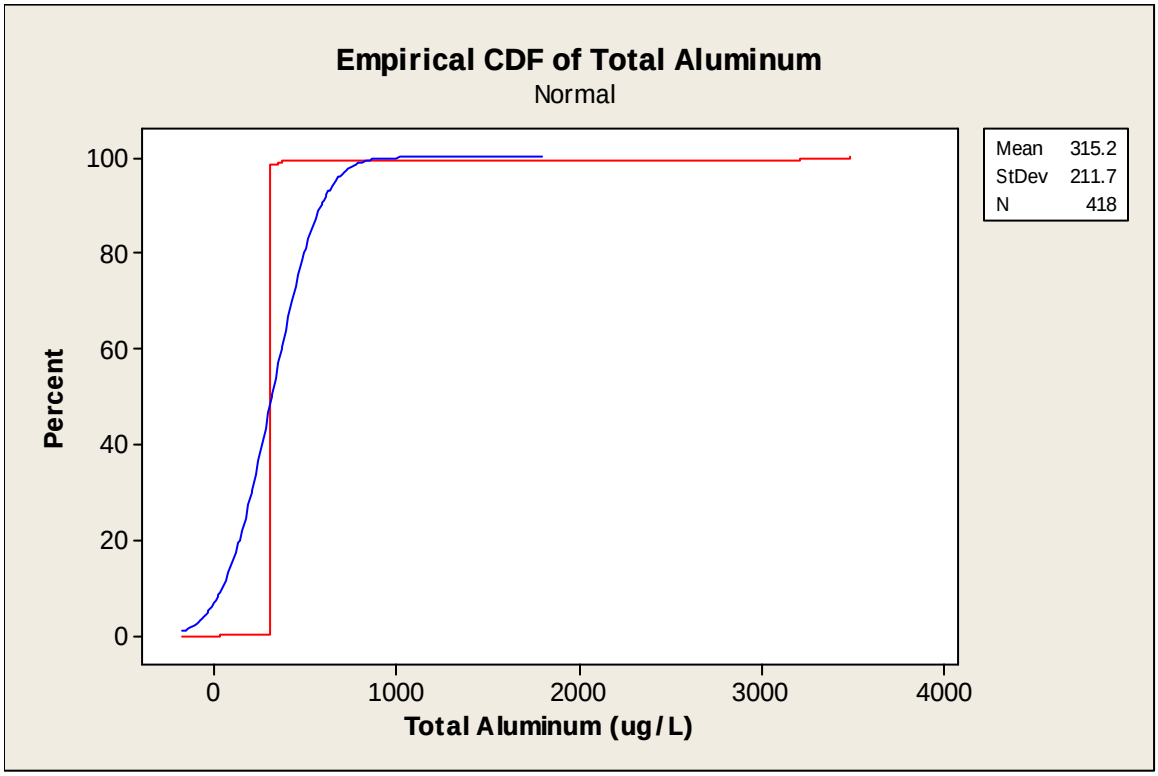
Anderson-Darling Normality Test	
A-Squared	109.24
P-Value <	0.005
Mean	27.72
StDev	100.77
Variance	10155.46
Skewness	9.723
Kurtosis	109.811
N	418
Minimum	5.00
1st Quartile	5.00
Median	5.00
3rd Quartile	14.00
Maximum	1260.00
95% Confidence Interval for Mean	
18.03	37.41
95% Confidence Interval for Median	
5.00	6.00
95% Confidence Interval for StDev	
94.37	108.11

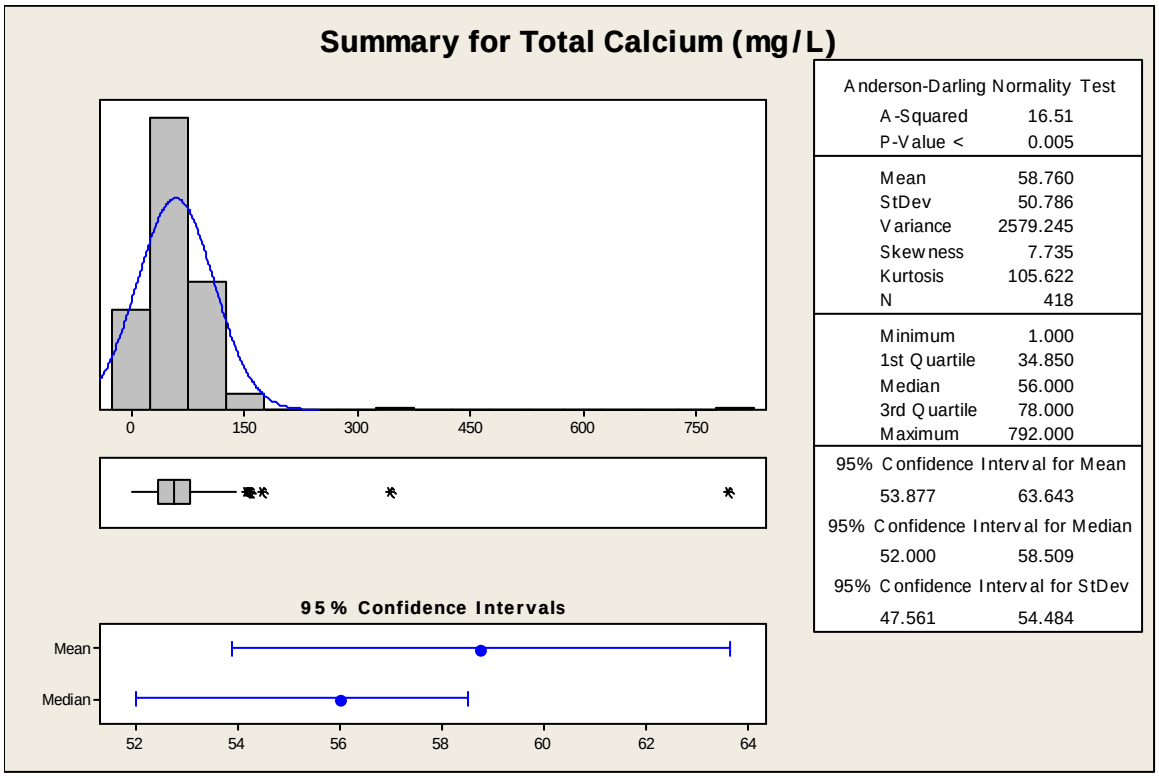


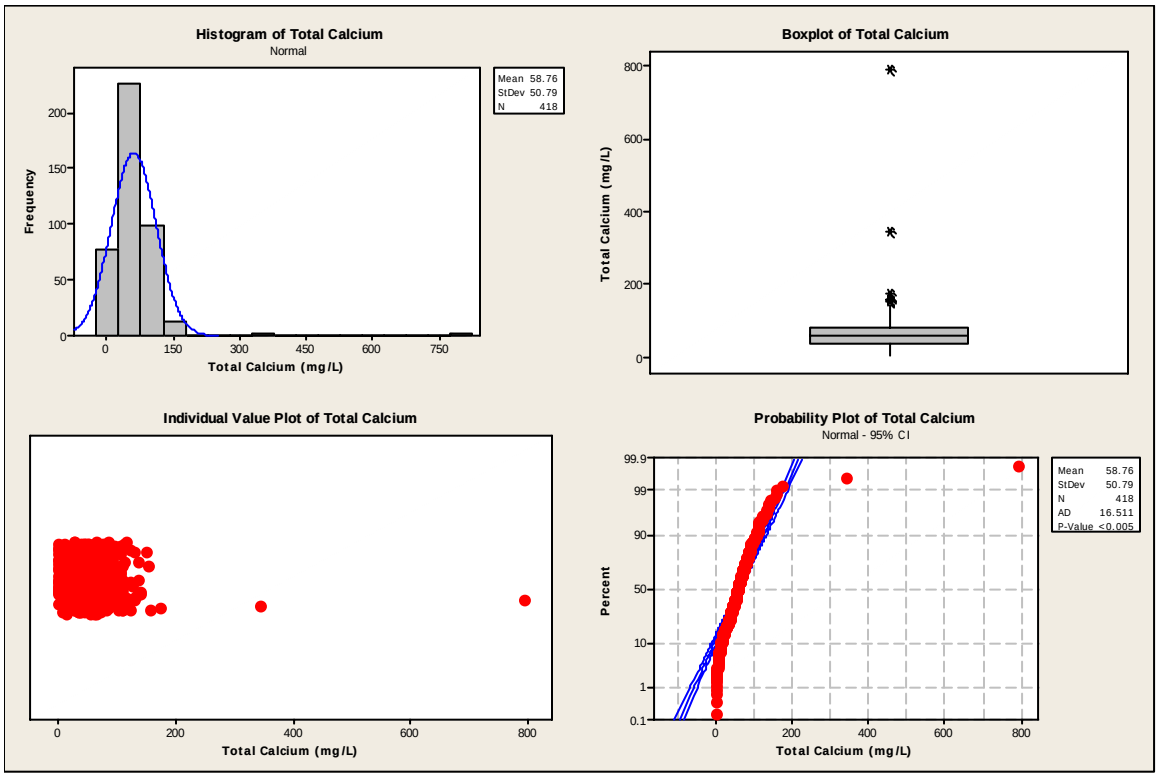

Empirical CDF of Total Zinc
Normal
Mean 27.72
StDev 100.8
N 418
Percent
100
80
60
40
20
0
0
300
600
900
1200
Total Zinc (ug/L)

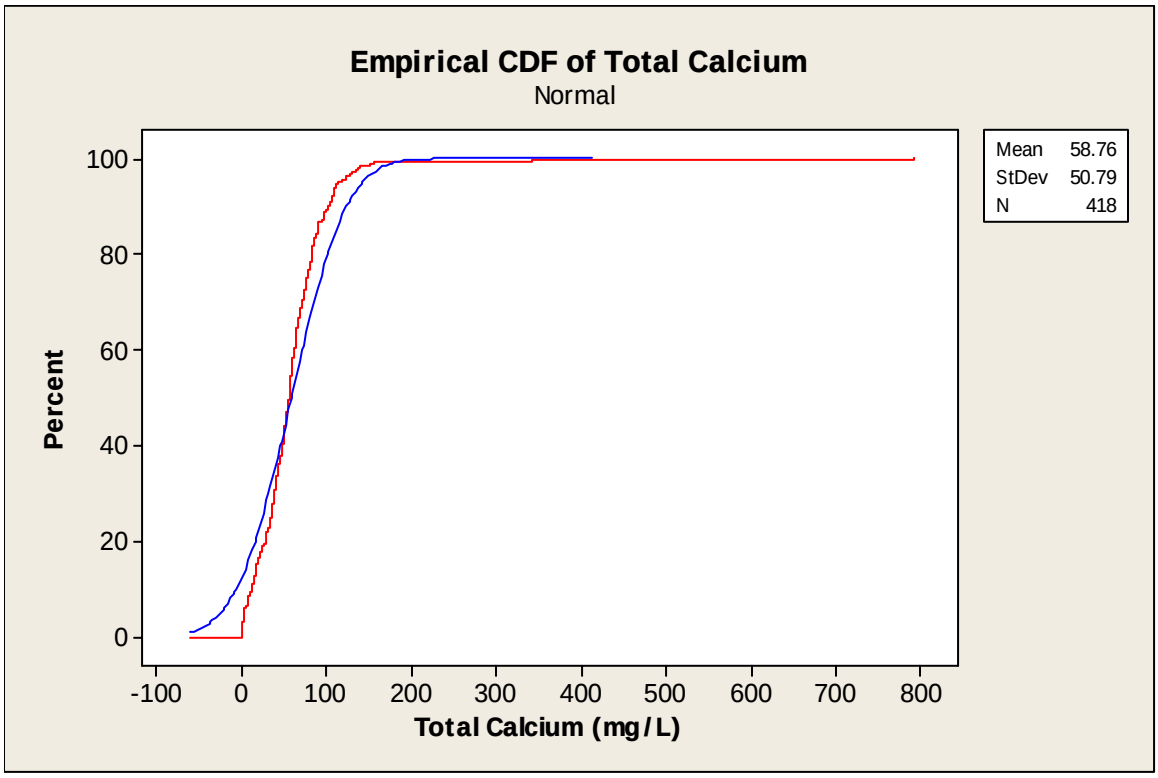

Summary for Total Aluminum (ug/L)
Anderson-Darling Normality Test
A-Squared 156.05
P-Value < 0.005
Mean 315.24
StDev 211.73
Variance 44828.59
Skewness 14.209
Kurtosis 203.451
N 418
Minimum 30.00
1st Quartile 300.00
Median 300.00
3rd Quartile 300.00
Maximum 3480.00
95% Confidence Interval for Mean
294.88 335.59
95% Confidence Interval for Median
300.00 300.00
95% Confidence Interval for StDev
198.28 227.14
0 600 1200 1800 2400 3000
95% Confidence Intervals
Mean
Median
290 300 310 320 330 340

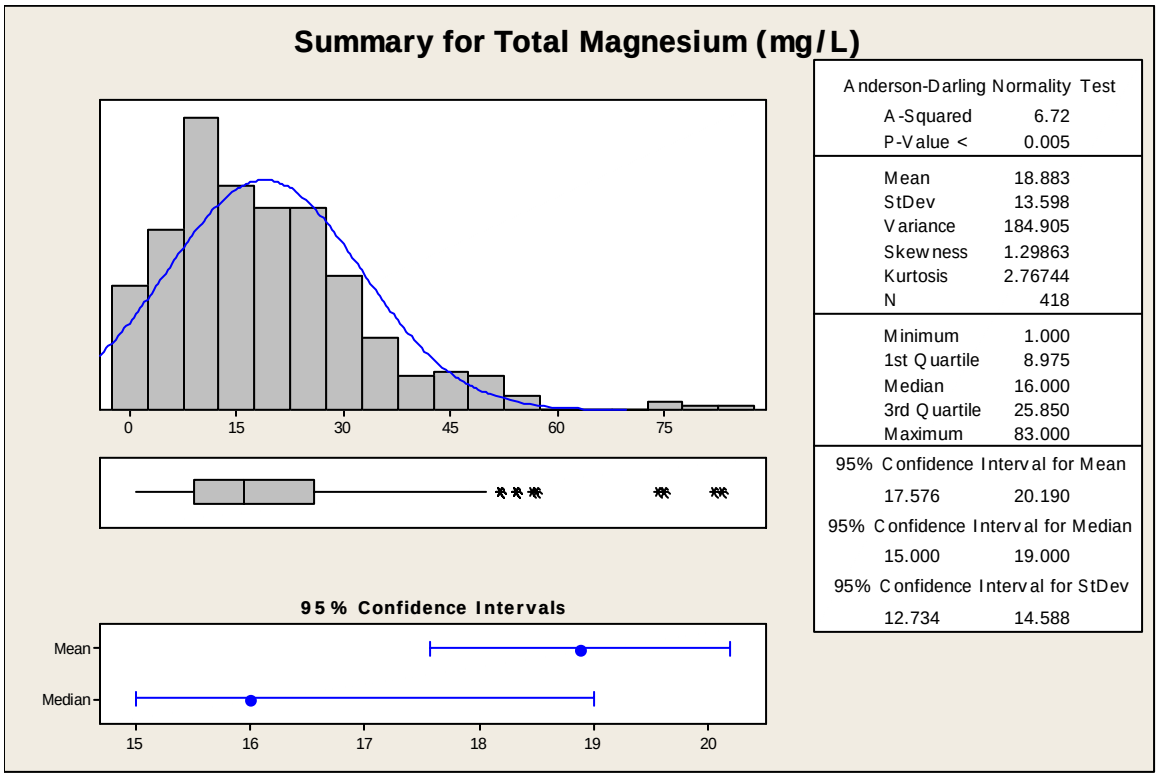

Histogram of Total Aluminum
Normal
Mean 315.2
StDev 211.7
N 418
Frequency
Total Aluminum (ug/L)
Boxplot of Total Aluminum
Total Aluminum (ug/L)
Individual Value Plot of Total Aluminum
Total Aluminum (ug/L)
Probability Plot of Total Aluminum
Normal - 95% CI
Mean 315.2
StDev 211.7
N 418
AD 156.050
P-Value <0.005
Percent
Total Aluminum (ug/L)

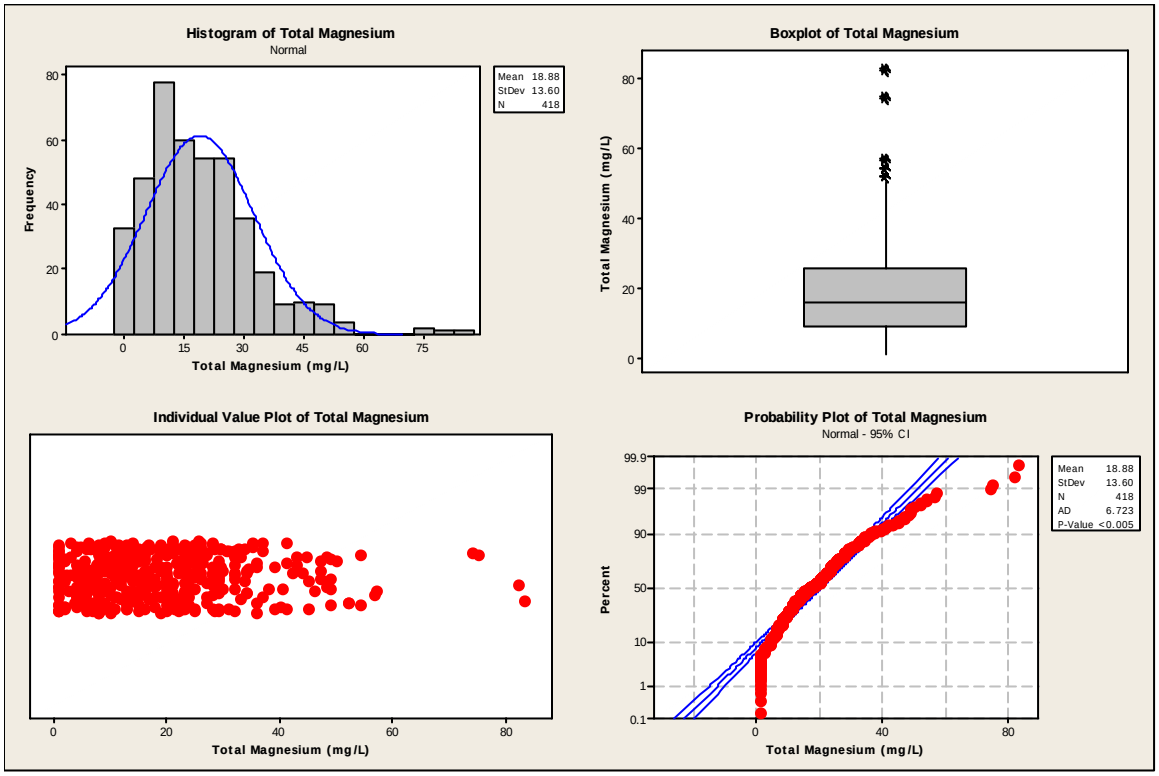

Empirical CDF of Total Aluminum
Normal
Mean 315.2
StDev 211.7
N 418
Percent
100
80
60
40
20
0
0
1000
2000
3000
4000
Total Aluminum (ug/L)

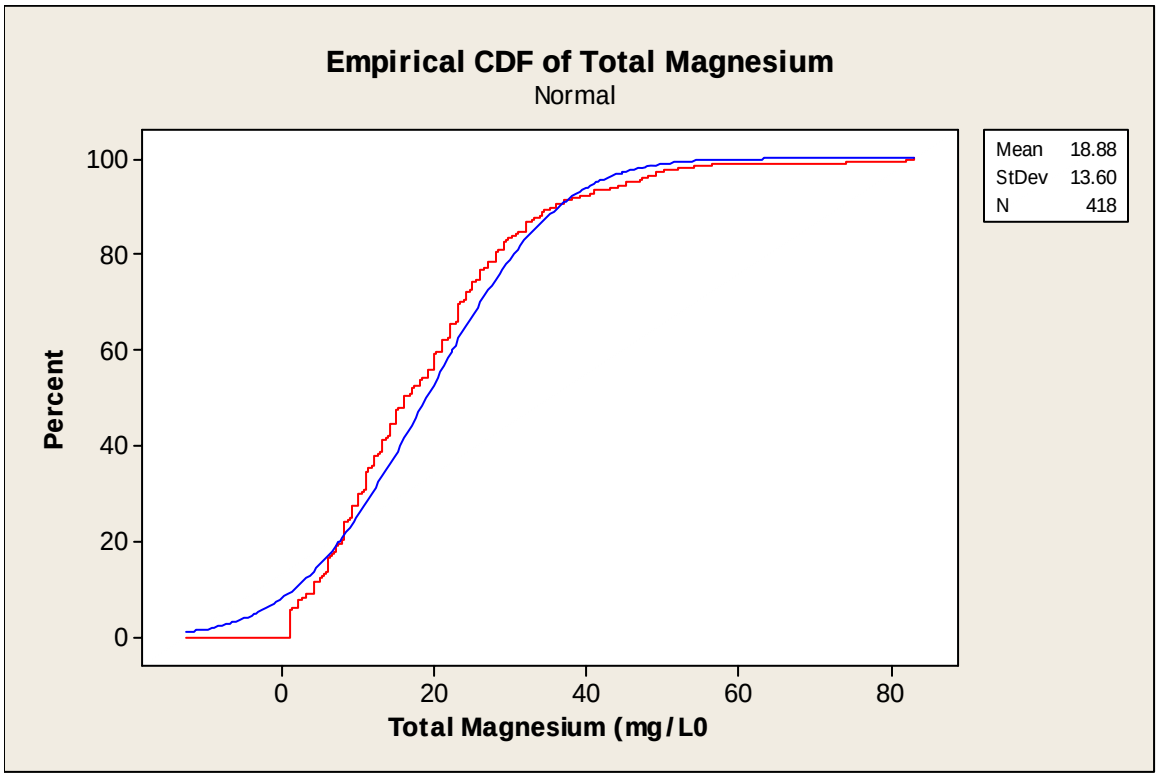

Summary for Total Calcium (mg/L)
Anderson-Darling Normality Test
A-Squared 16.51
P-Value < 0.005
Mean 58.760
StDev 50.786
Variance 2579.245
Skewness 7.735
Kurtosis 105.622
N 418
Minimum 1.000
1st Quartile 34.850
Median 56.000
3rd Quartile 78.000
Maximum 792.000
95% Confidence Interval for Mean
53.877 63.643
95% Confidence Interval for Median
52.000 58.509
95% Confidence Interval for StDev
47.561 54.484
0 150 300 450 600 750
95% Confidence Intervals
Mean
Median
52 54 56 58 60 62 64

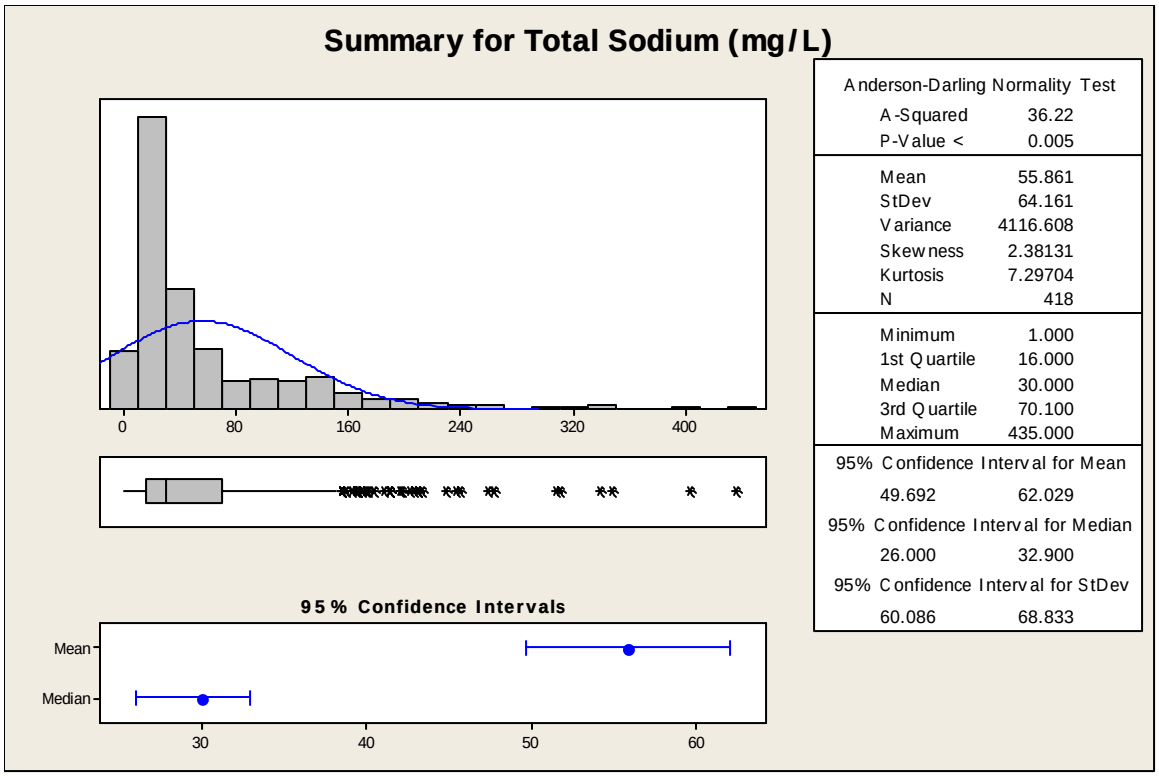

Histogram of Total Calcium
Normal
Mean 58.76
StDev 50.79
N 418
Frequency
Total Calcium (mg/L)
Boxplot of Total Calcium
Total Calcium (mg/L)
Individual Value Plot of Total Calcium
Total Calcium (mg/L)
Probability Plot of Total Calcium
Normal - 95% CI
Mean 58.76
StDev 50.79
N 418
AD 16.511
P-Value <0.005
Percent
Total Calcium (mg/L)

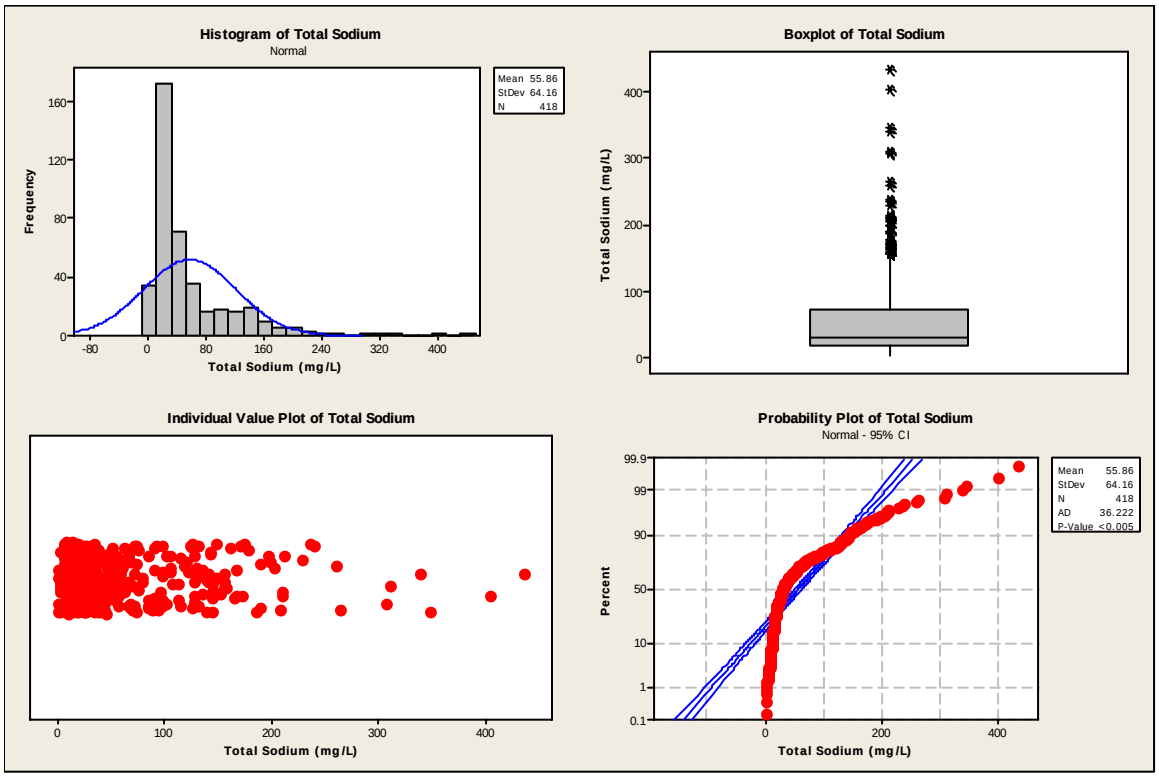

Empirical CDF of Total Calcium
Normal
Mean 58.76
StDev 50.79
N 418
Percent
100
80
60
40
20
0
-100
0
100
200
300
400
500
600
700
800
Total Calcium (mg/L)

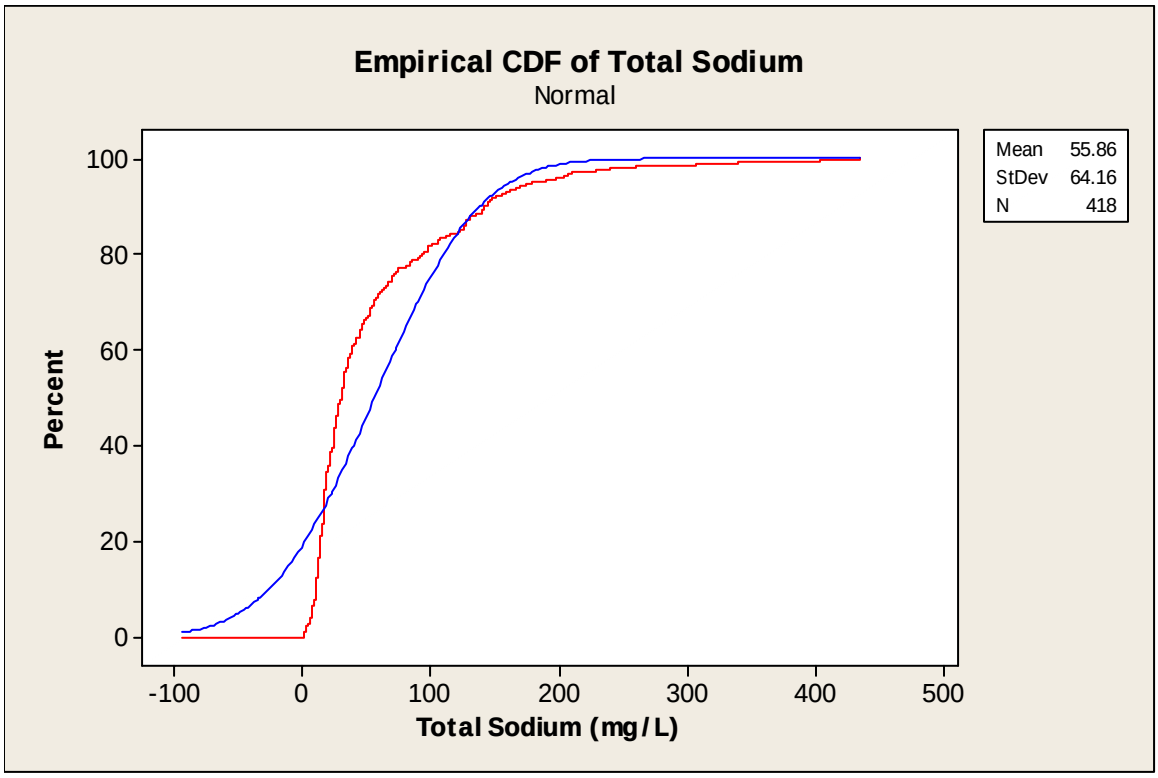

Summary for Total Magnesium (mg/L)
0
15
30
45
60
75
95% Confidence Intervals
Mean
Median
15
16
17
18
19
20
Anderson-Darling Normality Test
A-Squared 6.72
P-Value < 0.005
Mean 18.883
StDev 13.598
Variance 184.905
Skewness 1.29863
Kurtosis 2.76744
N 418
Minimum 1.000
1st Quartile 8.975
Median 16.000
3rd Quartile 25.850
Maximum 83.000
95% Confidence Interval for Mean
17.576 20.190
95% Confidence Interval for Median
15.000 19.000
95% Confidence Interval for StDev
12.734 14.588

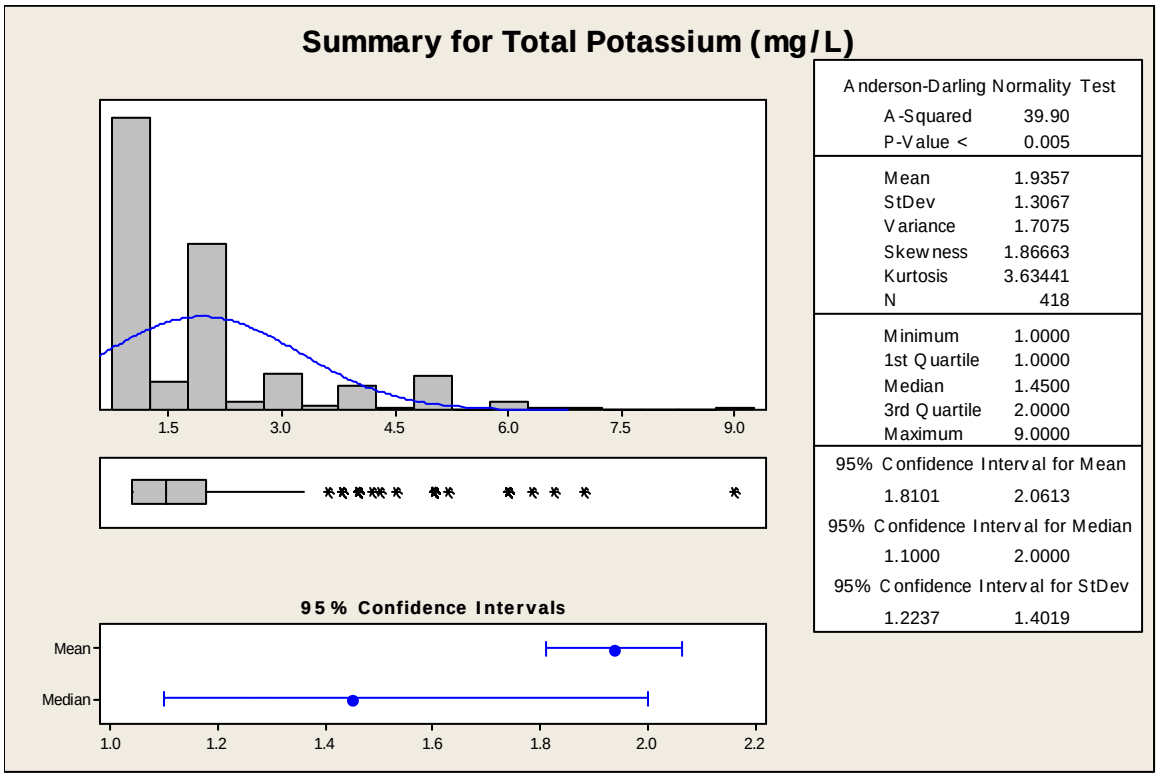

Histogram of Total Magnesium
Normal
Mean 18.88
StDev 13.60
N 418
Frequency
Total Magnesium (mg/L)
Boxplot of Total Magnesium
Total Magnesium (mg/L)
Individual Value Plot of Total Magnesium
Total Magnesium (mg/L)
Probability Plot of Total Magnesium
Normal - 95% CI
Mean 18.88
StDev 13.60
N 418
AD 6.723
P-Value <0.005
Percent
Total Magnesium (mg/L)

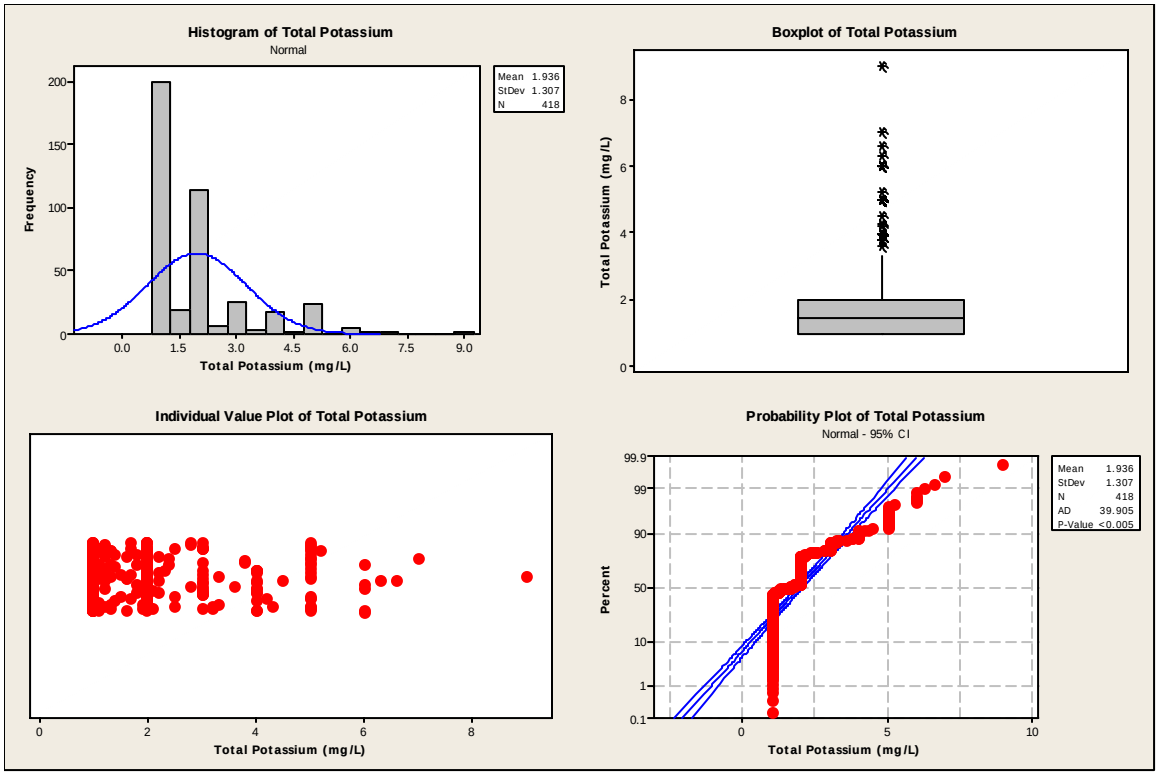

Empirical CDF of Total Magnesium
Normal
Mean 18.88
StDev 13.60
N 418
Percent
100
80
60
40
20
0
0
20
40
60
80
Total Magnesium (mg/L0

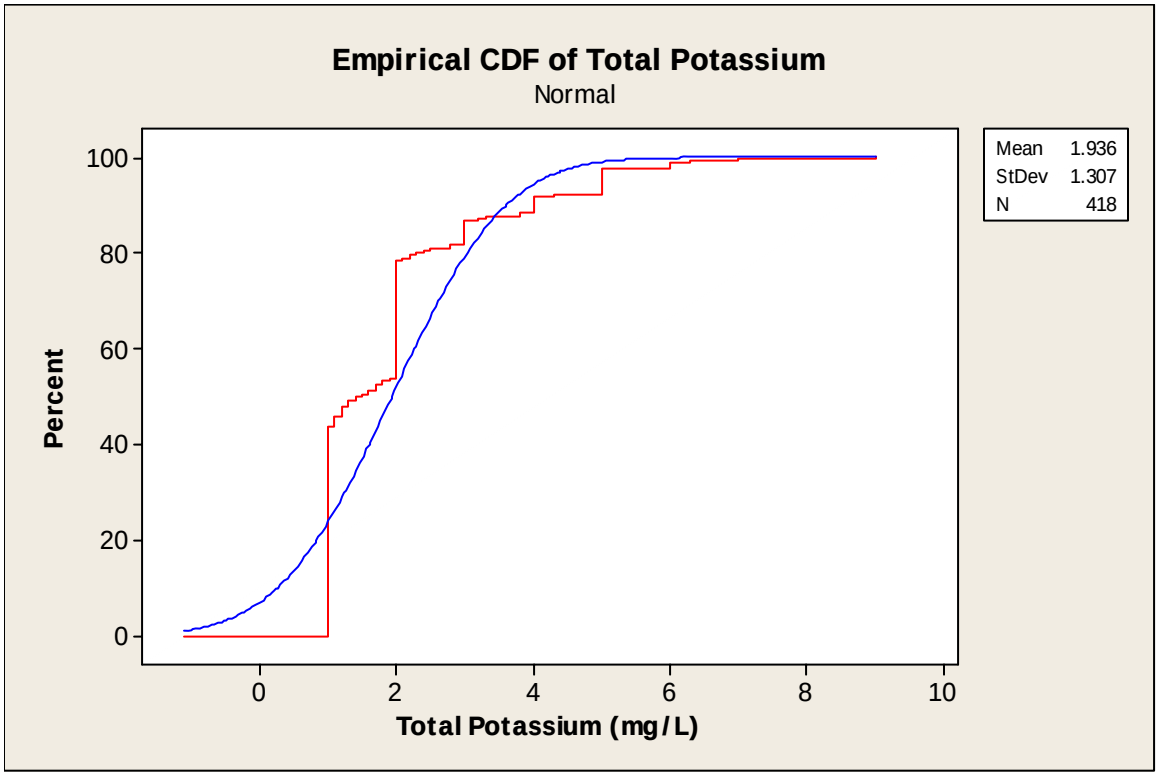

Summary for Total Sodium (mg/L)
Anderson-Darling Normality Test
A-Squared 36.22
P-Value < 0.005
Mean 55.861
StDev 64.161
Variance 4116.608
Skewness 2.38131
Kurtosis 7.29704
N 418
Minimum 1.000
1st Quartile 16.000
Median 30.000
3rd Quartile 70.100
Maximum 435.000
95% Confidence Interval for Mean
49.692 62.029
95% Confidence Interval for Median
26.000 32.900
95% Confidence Interval for StDev
60.086 68.833
0 80 160 240 320 400
95% Confidence Intervals
Mean
Median
30 40 50 60


Histogram of Total Sodium
Normal
Mean 55.86
StDev 64.16
N 418
Frequency
Total Sodium (mg/L)
Boxplot of Total Sodium
Total Sodium (mg/L)
Individual Value Plot of Total Sodium
Total Sodium (mg/L)
Probability Plot of Total Sodium
Normal - 95% CI
Mean 55.86
StDev 64.16
N 418
AD 36.222
P-Value < 0.005
Percent
Total Sodium (mg/L)


Empirical CDF of Total Sodium
Normal
Mean 55.86
StDev 64.16
N 418
Percent
100
80
60
40
20
0
-100
0
100
200
300
400
500
Total Sodium (mg/L)


Summary for Total Potassium (mg/L)
Anderson-Darling Normality Test
A-Squared 39.90
P-Value < 0.005
Mean 1.9357
StDev 1.3067
Variance 1.7075
Skewness 1.86663
Kurtosis 3.63441
N 418
Minimum 1.0000
1st Quartile 1.0000
Median 1.4500
3rd Quartile 2.0000
Maximum 9.0000
95% Confidence Interval for Mean
1.8101 2.0613
95% Confidence Interval for Median
1.1000 2.0000
95% Confidence Interval for StDev
1.2237 1.4019
1.5 3.0 4.5 6.0 7.5 9.0
95% Confidence Intervals
Mean
Median
1.0 1.2 1.4 1.6 1.8 2.0 2.2


Histogram of Total Potassium
Normal
Mean 1.936
StDev 1.307
N 418
Frequency
Total Potassium (mg/L)
Boxplot of Total Potassium
Total Potassium (mg/L)
Individual Value Plot of Total Potassium
Total Potassium (mg/L)
Probability Plot of Total Potassium
Normal - 95% CI
Mean 1.936
StDev 1.307
N 418
AD 39.905
P-Value < 0.005
Percent
Total Potassium (mg/L)


Empirical CDF of Total Potassium
Normal
Mean 1.936
StDev 1.307
N 418
Percent
100
80
60
40
20
0
0
2
4
6
8
10
Total Potassium (mg/L)

Results and Conclusions

The data analyses indicate that none of the parameters are normally distributed. Most of these data are skewed because of data outliers from wells that have a problem with one or more parameters. Because of the skewed data, the mean is not representative of the true middle of the data distribution. Some skewness is due to lower detection limits imposed by instrument capability. A more appropriate measure for central tendency in this set of water quality data is the median.

Page one of the data set is a Graphical Summary of the data and additional clarifying plots including the Normal Probability Plot. Page two has the Cumulative Distribution Function Graph. These pages include all the elements described in the objectives section needed to better understand the range, variance and distribution of the data except Temporal Plots.

The null hypothesis for this project was that the mean of water quality samples collected in Oklahoma would fall below the Primary Drinking Water Standard. Aluminum was the only parameter that did not fail to reject the null hypothesis however; the laboratory detection limit for aluminum was 300 ug/L, which is above the secondary standard of 0.2 mg/L.

Mann-Whitney Test were run on all 16 parameters to compare year two and year three data results. The null hypothesis was that all the medians of the data would be equal. Results indicate that we would fail to reject the null hypothesis on all pairs except for pH. pH results for Mann-Whitney were .02 we have to therefore reject the null hypothesis and say there is a difference between year two and year three values.

Chart of Means

Chemical Parameter	Mean yr. 1	Mean yr. 2	Mean yr. 3	DW Standard
pH	7.5125	7.557	7.4138	6.5-8.5
Total Alkalinity	213.76	230.15	242.86	no std.
Total Dissolved Solids	383.89mg/L	407.09	393.32	500mg/L
Nitrate	2.6358mg/L	3.5023	3.2407	10mg/L
Chloride	38.621mg/L	41.137	44.04	250mg/L
Sulfate	55.29mg/L	55.473	65.02	250mg/L
Fluoride	0.35846mg/L	0.375	0.4111	2mg/L
Copper	16.618ug/L	23.58	19.642	1mg/L
Iron	141.41ug/L	224.07	189.82	0.3mg/L
Manganese	33.79ug/L	73.59	66.77	0.05mg/L
Zinc	18.05ug/L	25.11	27.72	5mg/L
Aluminum	299.92ug/L	319.79	315.24	0.05-2mg/L
Calcium	49.095mg/L	57.079	58.76	no std.
Magnesium	16.45mg/L	18.8	18.883	no std.
Sodium	64.86mg/L	59.191	55.861	no std.
Potassium	1.87mg/L	1.9618	1.9357	no std.

% exceed MCL Yr. 2		**% exceed MCL Yr. 3**	
pH	4.7	pH	9.8
TDS	21.5	TDS	23.4
Nitrate	9.1	Nitrate	8.4
Chloride	0.0	Chloride	1.6
Fluoride	1.9	Sulfate	5.0
Copper	0.47	Copper	0.0
Iron	12.9	Iron	1.2
Manganese	11.0	Manganese	3.1
Zinc	0.0	Zinc	0.0
Aluminum	0.47	Aluminum	0.47