CRWR Online Report 98-4

An Analysis of

Volunteer Water Quality Data

by

Charles Kaough, M.S. EHE.

and

Dr. David R. Maidment, PhD.

Principal Investigator

May 1998

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research • The University of Texas at Austin
J.J. Pickle Research Campus • Austin, TX 78712-4497

This document is available online via World Wide Web at
http://www.ce.utexas.edu/centers/crwr/reports/online.htm

Acknowledgements

I wish to thank Dr. David R. Maidment for supervising and guiding me through this research and Michael Barrett, my reader, who kept me on track with my work. I wish, also to thank Steven Hubbell, Jason Pinchback, John Wedig and Diane Spinney at the Lower Colorado River Authority who furnished me with data and much assistance. Also, I wish to express deep gratitude to my wife, Monica, and children, Maya and Elena, who gave me support and encouragement throughout graduate school.

TABLE OF CONTENTS

1 INTRODUCTION 8

1.1 THE NEED FOR VOLUNTEER WATER QUALITY MONITORING 9
1.2 CREDIBILITY OF VOLUNTEER DATA 12
1.3 THE GOALS OF THE COLORADO RIVER WATCH NETWORK: 14

2 LITERATURE REVIEW 15

3 STATISTICAL ANALYSIS 21

3.1 GOALS OF STATISTICAL ANALYSIS 22
3.2 DESCRIPTION OF THE STATISTICAL ANALYSIS 23
3.3 CHOOSING SITES TO ANALYZE 28
3.4 PARAMETERS STUDIED AND MEASUREMENT METHODS 32
3.5 PARALLEL MEASUREMENTS ANALYSIS 34

4 FOURIER ANALYSIS OF WATER QUALITY DATA 36

4.1 THEORY 36
4.2 SET-UP 37
4.3 RESULTS 41
4.4 CONCLUSION 47

5 STUDY RESULTS 48

5.1 LAKE AUSTIN 48
5.2 BASTROP 61
5.3 SMITHVILLE 75
5.4 LAGRANGE 88

6 SUMMARIES AND CONCLUSIONS 116

7 BIBLIOGRAPHY 122

APPENDIX 124

RAW DATA 124

TABLE OF TABLES

TABLE 3-1 CRITICAL VALUES, PROBABILITY PLOT 25
TABLE 3-2 SPREADSHEET FOR QUANTILE PLOT 25
TABLE 3-3 CRWN SEASONAL DATA DISTRIBUTION 28
TABLE 4-1 RAW DATA EXAMPLE 37
TABLE 4-2 REGRESSION STATISTICS, EXCEL OUTPUT 40
TABLE 4-3 REGRESSION STATISTICS, J=1 41
TABLE 4-4 SUMMARY OF MEANS, DO, J=1 AND J=2 47
TABLE 5-5-1 DESCRIPTIVE STATISTICS, DO- LAKE AUSTIN 48
TABLE 5-5-2 DESCRIPTIVE STATISTICS COMBINED DATA, DO- LAKE AUSTIN 49
TABLE 5-5-3 T-TEST RESULTS, DO- LAKE AUSTIN 52
TABLE 5-5-4 STANDARD ERROR OF THE MEAN, DO- LAKE AUSTIN 53
TABLE 5-5 REGRESSION STATISTICS- LCRA1- LAKE AUSTIN 55
TABLE 5-6 REGRESSION STATISTICS- CRWN1- LAKE AUSTIN 55
TABLE 5-7 SUMMARY OF MEANS- LAKE AUSTIN 56
TABLE 5-8 DESCRIPTIVE STATISTICS, TDS- LAKE AUSTIN 58
TABLE 5-9 RESULT OF THE T-TEST- TDS- LAKE AUSTIN 59
TABLE 5-10 DESCRIPTIVE STATISTICS, DO- BASTROP 61
TABLE 5-11 DESCRIPTIVE STATS. COMBINED DATA- BASTROP 62
TABLE 5-12 T-TEST RESULTS, DO- BASTROP 65
TABLE 5-13 STANDARD ERROR OF THE MEAN, DO-BASTROP 65
TABLE 5-14 REGRESSION STATISTICS, DO-LCRA2 68
TABLE 5-15 REGRESSION STATISTICS, DO-CRWN2 69
TABLE 5-16 SUMMARY OF MEANS, DO- BASTROP 70
TABLE 5-17 DESCRIPTIVE STATISTICS, TDS- BASTROP 71
TABLE 5-18 T-TEST RESULTS, TDS- BASTROP 72
TABLE 5-19 DESCRIPTIVE STATISTICS, DO- SMITHVILLE 75
TABLE 5-20 T-TEST RESULTS, DO- SMITHVILLE 78
TABLE 5-21 STANDARD ERROR OF THE MEAN, DO- SMITHVILLE 78
TABLE 5-22 STANDARD ERROR OF THE MEAN, DO- COMBINED DATA- SMITHVILLE 79
TABLE 5-23 DESCRIPTIVE STATISTICS, DO- COMBINED DATA- SMITHVILLE 80
TABLE 5-24 REGRESSION STATISTICS- DO- LCRA3 80
TABLE 5-25 REGRESSION STATISTICS- DO- CRWN3 81
TABLE 5-26 SUMMARY OF MEANS, DO- SMITHVILLE 82
TABLE 5-27 DESCRIPTIVE STATISTICS, TDS- SMITHVILLE 84
TABLE 5-28 T-TEST RESULTS, TDS- SMITHVILLE 85

Table 5-29 Descriptive Statistics, DO- LaGrange 88
Table 5-30 Descriptive Statistics, Combined Data, DO- LaGrange 89
Table 5-31 t-test Results, DO- LaGrange 92
Table 5-32 Descriptive Statistics w/o Outlier Data, DO- LaGrange 93
Table 5-33 Standard Error of the Mean, DO- LaGrange 95
Table 5-34 Standard Error of the Mean w/o Outlier Data, DO- LaGrange 95
Table 5-35 Regression Statistics, DO- LCRA4 96
Table 5-36 Regression Statistics, DO- CRWN4 96
Table 5-37 Summary of Means, DO- LaGrange 98
Table 5-38 Descriptive Statistics, TDS- LaGrange 100
Table 5-39 t-test Results, TDS- LaGrange 101
Table 5-40 Descriptive Statistics, DO- Columbus 103
Table 5-41 Descriptive Statistics, Combined Data Sets, DO- Columbus 104
Table 5-42 t-test Results, DO- Columbus 107
Table 5-43 Standard Error of the Mean, DO- Columbus 107
Table 5-44 Regression Statistics, DO- LCRA5 109
Table 5-45 Regression Statistics, DO- LCRA5 110
Table 5-46 Summary of Means, DO- Columbus 111
Table 5-47 Descriptive Statistics, TDS- Columbus 112
Table 5-48 t-test Results, TDS- Columbus 113
Table 6-1 Summary of Data, DO 116
Table 6-2 Summary of Data, TDS 117
Table 6-3 Summary of Data, DO, t-tests 117
Table 6-4 Summary of Data, TDS, t-tests 117
Table 6-5 Summary of Data, Means, DO and TDS 118

TABLE OF FIGURES

FIGURE 1-1 FROM TNRCC'S QAAP DOCUMENT 8
FIGURE 1-2 FROM TNRCC'S QAAP DOCUMENT 9
FIGURE 3-1 QUANTILE PLOT 26
FIGURE 3-2 MAP OF MONITORING SITES 29
FIGURE 3-3 PARALLEL TESTING PLOT, DO 35
FIGURE 3-4 PARALLEL TESTING PLOT, TDS 35
FIGURE 4-1 DO SCATTER PLOT 37
FIGURE 4-2 SPREADSHEET LAYOUT FOR FOURIER ANALYSIS 38
FIGURE 4-3 SPREADSHEET FORMULAS FOR FOURIER ANALYSIS 38
FIGURE 4-4 EXCEL REGRESSION TOOL 39
FIGURE 4-5 EXCEL REGRESSION TOOL WINDOW 39
FIGURE 4-6 PLOT OF REGRESSION RESULTS, J=2 42
FIGURE 4-7 PLOT OF REGRESSION RESULTS, J=1 42
FIGURE 4-8 EXCEL SPREADSHEET, REGRESSION PLOT 43
FIGURE 4-9 CRWN DO FOURIER PLOTS, J=1 VS. J=2 44
FIGURE 4-10 LCRA DO FOURIER PLOTS, J=1 VS. J=2 44
FIGURE 4-11 DO FOURIER PLOT, LCRA VS CRWN, J=1 45
FIGURE 4-12 DO FOURIER PLOT, LCRA VS CRWN, J=2 46
FIGURE 5-1 QUANTILE PLOT, DO- CRWN1 50
FIGURE 5-2 QUANTILE PLOT, DO- LCRA1 50
FIGURE 5-3 HISTOGRAMS, DO- LAKE AUSTIN 51
FIGURE 5-4 STANDARD ERROR OF THE MEAN, DO- LCRA1 53
FIGURE 5-5 STANDARD ERROR OF THE MEAN, DO- CRWN1 54
FIGURE 5-6 FOURIER PLOT- LAKE AUSTIN 56
FIGURE 5-7 SCATTER PLOT, TDS- LAKE AUSTIN 57
FIGURE 5-8 HISTOGRAMS, TDS- LAKE AUSTIN 59
FIGURE 5-9 STANDARD ERROR OF THE MEAN- TDS- LCRA1 60
FIGURE 5-10 STANDARD ERROR OF THE MEAN- TDS- CRWN1 60
FIGURE 5-11 QUANTILE PLOT, DO- CRWN2 63
FIGURE 5-12 QUANTILE PLOT, DO- LCRA2 63
FIGURE 5-13 HISTOGRAMS, DO- BASTROP 64
FIGURE 5-14 STANDARD ERROR OF THE MEAN, DO- LCRA2 66
FIGURE 5-15 STANDARD ERROR OF THE MEAN, DO- CRWN2 67
FIGURE 5-16 SCATTER PLOT, TDS- BASTROP 68
FIGURE 5-17 FOURIER PLOT, DO- BASTROP 69
FIGURE 5-18 SCATTER PLOT, TDS- BASTROP 71

FIGURE 5-19 HISTOGRAMS, TDS- BASTROP 72
FIGURE 5-20 STANDARD ERROR OF THE MEAN, TDS- LCRA2 73
FIGURE 5-21 STANDARD ERROR OF THE MEAN TDS, CRWN2........ 73
FIGURE 5-22 QUANTILE PLOT, DO- CRWN3 76
FIGURE 5-23 QUANTILE PLOT DO, LCRA3 76
FIGURE 5-24 HISTOGRAMS, DO- SMITHVILLE 77
FIGURE 5-25 FOURIER PLOT, DO- SMITHVILLE........ 81
FIGURE 5-26 SCATTER PLOT, TDS- SMITHVILLE 83
FIGURE 5-27 HISTOGRAMS, TDS- SMITHVILLE........ 85
FIGURE 5-28 STANDARD ERROR OF THE MEAN, TDS- CRWN3 86
FIGURE 5-29 STANDARD ERROR OF THE MEAN, TDS- LCRA3 86
FIGURE 5-30 HISTOGRAMS, DO- LAGRANGE 90
FIGURE 5-31 QUANTILE PLOT, DO- CRWN4 91
FIGURE 5-32 QUANTILE PLOT, DO- LCRA4........ 91
FIGURE 5-33 STANDARD ERROR OF THE MEAN, DO- CRWN4........ 93
FIGURE 5-34 STANDARD ERROR OF THE MEAN, DO- LCRA4 94
FIGURE 5-35 CRWN4 FOURIER PLOT, DO WITH DATA 97
FIGURE 5-36 FOURIER PLOT, DO- LAGRANGE 98
FIGURE 5-37 SCATTER PLOT, TDS- LAGRANGE 99
FIGURE 5-38 HISTOGRAMS, TDS- LAGRANGE........ 100
FIGURE 5-39 STANDARD ERROR OF THE MEAN, TDS- LCRA4 101
FIGURE 5-40 STANDARD ERROR OF THE MEAN, TDS- CRWN4 102
FIGURE 5-41 QUANTILE PLOT, DO- CRWN5 105
FIGURE 5-42 QUANTILE PLOT, DO- LCRA5........ 105
FIGURE 5-43 HISTOGRAMS, DO- COLUMBUS........ 106
FIGURE 5-44 STANDARD ERROR OF THE MEAN, DO- CRWN5........ 108
FIGURE 5-45 STANDARD ERROR OF THE MEAN, DO- LCRA5 109
FIGURE 5-46 FOURIER PLOT, DO- COLUMBUS........ 110
FIGURE 5-47 SCATTER PLOT, TDS- COLUMBUS 111
FIGURE 5-48 HISTOGRAMS, TDS- COLUMBUS........ 113
FIGURE 5-49 STANDARD ERROR OF THE MEAN, TDS- LCRA5 114
FIGURE 5-50 STANDARD ERROR OF THE MEAN, TDS- CRWN5 114
FIGURE 6-1 DO MEAN VALUES 118
FIGURE 6-2 DO MEAN VALUES, FOURIER METHOD 119
FIGURE 6-3 MEAN VALUES, TDS........ 119
FIGURE 6-4 LCRA MEAN VALUES, DO- STANDARD VS. FOURIER........ 120
FIGURE 6-5 CRWN DO MEAN VALUES, STANDARD VS. FOURIER........ 121

1 INTRODUCTION

The quality of our lives, indeed, our society is related to the quality of our water. Water is an essential element for life. With our population growing and more humans living in urban environments, we are faced with challenges to keep our water quality high and thus our standard of living high. Only recently have we become aware of how our lifestyles effect the quality of our water

The scientific community has done much to quantify water quality and develop methods to preserve and improve it. The government has attempted to use these advancements in science and technology by developing legislation such as the Clean Water Act. Designated uses and water quality standards to support these designations have been set for our streams, lakes, bays and estuaries as mandated by the Clean Water Act. States adopt EPA-approved standards for their waters that define water quality goals for individual waterbodies. Standards consist of designated beneficial uses to be made of the water, criteria to protect those uses, and antidegradation provisions to protect existing water quality[1]. States are required by law to systematically assess and report on the quality of their waters regularly and thoroughly, according to section 305b of the Clean Water Act.

The task of assessing the quality of all of our bodies of water is a huge task and requires a immense amount of data. According to the *EPA National Water Quality Inventory: 1992 Report to Congress* only 643,000 miles of the estimated 3.5 million river miles in the United States have been assessed (Figure 1-1). More than one-fourth of the

[1] 1972 Clean Water Act (Federal Water Pollution Control Act), Section 305b and 1987 , Section 319.

assessed water bodies have been found to be pollution impaired waters as shown in Figure 1-2.

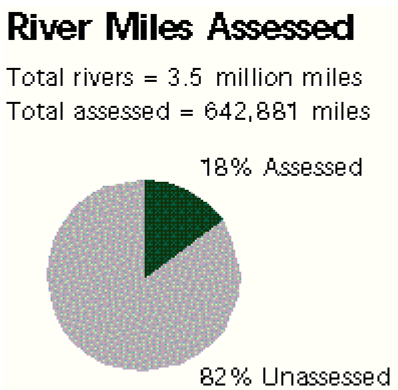


Figure 1-1 **from TNRCC's QAAP document**

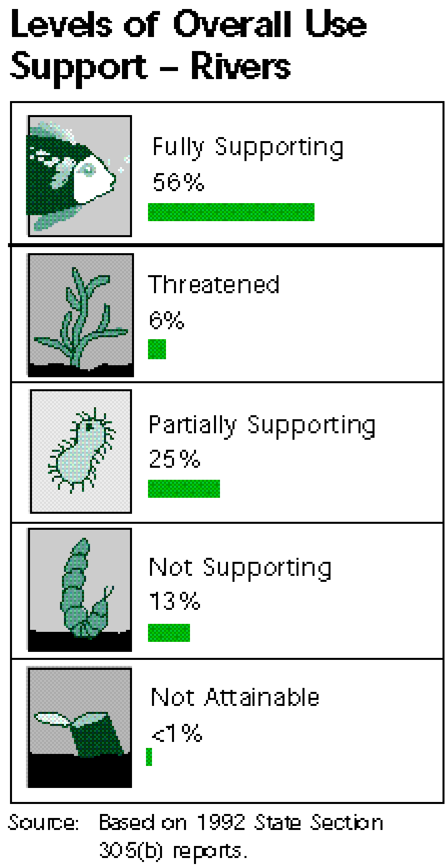


Figure 1-2 **from TNRCC's QAAP document**

1.1 The Need For Volunteer Water Quality Monitoring

States are faced with both the huge task of assessing the quality of the water in the state and limited resources, financial and human, to accomplish the task. Volunteer water quality monitoring programs have the potential to contribute a great deal to both providing the necessary data and contributing to maintaining or improving the quality of our waters.

Also, volunteer monitoring programs provide a means for citizens to become better stewards of the watershed they live in. This quote from the EPA document *The Quality of Our Nations Water* describes this:

> "The EPA encourages each citizen to become a steward of our precious natural resources. Complex environmental threats and diminishing funds for pollution control force us to jointly solve the pollution problems that foul our beaches and lakes or close our favorite fishing sites. We need to understand these problems and become a part of their solution. Once we understand these pollution problems and what is needed to combat them, we will be better able to prioritize our efforts, devise sound solutions, take appropriate action, monitor progress after solutions are implemented, and modify behavior that contributes to the problems.[3]"

In Texas the need for volunteer monitoring is great. According to the Texas Natural Resource Conservation Commission's (TNRCC) Texas Watch, which coordinates state-wide volunteer monitoring programs, four factors create the need for volunteer environmental monitoring in Texas:

1) Texas has a large number of water bodies (about 11,247 rivers and streams large enough to be named) with 191,228 miles of rivers and streams;

2) Texas's population is projected to increase by 59% through the year 2030;

3) Since 1992 TNRCC funding for surface water quality monitoring has dropped from approximately $1,108,000 to $800,000 in 1996;

4) Senate Bill 818, the Clean Rivers Program, states there is a lack of sufficient water quality data needed for state and local governments to make environmentally sound decisions[2].

The scientific community, however, is skeptical about the usefulness of data obtained by citizen monitors. As noted in a master's report by Fatima Paiva*, Volunteer Environmental Monitoring Programs: A Planning Framework*, "Data credibility is the biggest threat to the future of volunteer monitoring programs."

Efforts are being made by citizen monitoring organizations on the national and state levels to develop methods to insure that this data is valid and useful. The goal of this report is to assess the credibility of data collected by citizen monitoring organizations. The data and methods of the state organization, the TNRCC's Texas Watch, and a regional organization, the Lower Colorado River Authority's (LCRA) Colorado River Watch Network will be used as examples.

According to Steven Hubbell, Activity Manager of the Colorado River Watch Network, there is a need for the methodology of volunteer monitoring organizations to be scrutinized by the scientific community. This scrutiny will help improve and legitimize the efforts of these organization. It is the intention of this report to scrutinize the methods of volunteer monitoring, the Texas Watch and the Colorado River Watch Network, and determine if the data collected by these organizations is useful to environmental engineers, policy makers and the scientific community.

[2] Texas Natural Resource Conservation Commission, *Quality Assurance Project Plan*, Texas Watch (Austin, Texas: January 2, 1998)

1.2 Credibility of Volunteer Data

The EPA has issued guidelines through its Office of Water for volunteer organizations to use to insure that its monitoring program is credible. The organization of the TNRCC's Texas Watch and the LCRA's Colorado River Watch Network are based on these guidelines. The following is the general outline of the recommendations[3]:

1. Establishing a Pilot Program
    a) Pick a location
    b) Select sampling equipment
    c) Design a data collection form
    d) Recruit volunteers
    e) Train volunteers
    f) Conduct ongoing quality control
    g) Refine program materials
2. Expand the Program
3. Maintain Volunteer Interest and Motivation
4. Prepare a Quality Assurance Project Plan (QAPP)
    a) Project description
    b) Project organization and responsibilities
    c) Quality Assurance objectives
    d) Sampling procedures

[3] U.S. Environmental Protection Agency, *Volunteer Water Monitoring: A Guide For State Managers*, Office of Water (Washington, D.C.: Government Printing Office, August 1990)

e) Sample custody

f) Calibration procedures and frequency

g) Analytical procedures

h) Data reduction, validation and reporting

i) Internal quality control checks

j) Performance and system audits

k) Preventative maintenance

l) Specific routine procedures used to assess data precision, accuracy and completeness.

m) Corrective action

n) Quality assurance reports

These guidelines illustrate the thoroughness of the efforts of the Texas Watch and the Colorado River Watch Network to produce credible data. These projects must adopt protocols that are straightforward enough for volunteers to master and yet sophisticated enough to generate data of value to resource managers.

The EPA requires that all volunteer organizations that use EPA funds produce a Quality Assurance Project Plan (QAPP). It is recommended that organizations that do not use EPA funds also adopt a QAPP. The first task in this plan is to establish the goals of the organization. Some of the goals could be:

- primarily education oriented
- collect data that can be used in making water quality management decisions

Projects with the second goal might be called primarily data oriented. Data oriented volunteer organizations must show that the measured data is credible. The credibility of the data is indicated by having the following attributes:

- consistent over time and within projects and group members
- collected and analyzed using standardized and acceptable techniques
- comparable to data collected in other assessments using the same methods

1.3 The Goals of the Colorado River Watch Network[4]:

1.Maintain a motivated volunteer monitoring network committed to preserving the integrity of the Colorado River Watershed.

2.Provide educational opportunities about water quality and the environment to the communities in the Lower Colorado River Authority (LCRA) service area.

3.Complement and assist the LCRA with its watershed monitoring effort and act as an early warning system alerting the LCRA to potential water quality threats.

The quality assurance plan:

- details a project's standard operating procedures in the field and lab,
- outlines project organization,
- address issues such as training requirements, instrument calibration, and internal checks on how data are collected, analyzed, and reported.

The quality of the CRWN data will be assessed using statistical techniques.

[4] Colorado River Watch Network, *Technical Instructions*, Fifth Edition(Austin, Texas: May 1996)

2 LITERATURE REVIEW

The focus of this report has been on the reliability of volunteer water quality data. There has been great quantity of material written on water quality monitoring techniques, water quality modeling and data analysis. There is also a large quantity of material written on probability and statistics in general and specifically for hydrologic data. However, a search of the literature did not uncover any work which was specifically concerning the comparison of data sets from volunteer and professional monitors except in parallel studies where the measurements were taken at the same point and time and then compared.

There is wealth of information on quality control and quality assurance of water quality data collected by volunteer monitors in publications by the U.S. Environmental Protection Agency, the Texas Natural Resource Conservation Commission, and by individual monitoring organizations such as the LCRA's Colorado River Watch Network. An example is a detailed discussion of statistical analysis of parallel studies in a series of articles in "The Volunteer Monitor" the national newsletter of volunteer water quality monitoring published by the EPA. This newsletter format is very informative because a variety of techniques and applications from various monitoring groups around the country were compiled and discussed. Using the newsletter format of communicating, the "Volunteer Monitor" was able to first publish a call for articles from its readers on the specific topic of statistical techniques used to quantify how well monitoring techniques were working and the accuracy of monitoring methods. A collection of the short articles was then published in the following publication of the newsletter. Articles detailing specific monitoring techniques, the lab methods, organizational forms, political issues and its application to water quality issues are all discussed in the newsletters.

The EPA has a number of publications aimed specifically at volunteer water monitoring and quality assurance. A particularly useful document used in this study is *Volunteer Water Monitoring: A Guide For State Managers.* This document provides guidelines for statewide monitoring programs with emphasis on quality assurance and quality control. There were some basic statistical techniques discussed and the methods of analyzing and communicating the information gathered by the monitors.

Another particularly useful document from the EPA is *EPA Requirements for Quality Assurance Project Plans (QAPP) for Environmental Data.* The purpose of this document is to provide volunteer monitoring programs with the information they need to develop a quality assurance project plan. Specific statistical techniques for assessing data quality are discussed.

The TNRCC's *Quality Assurance Project Plans* is a document intended to provide guidance for monitoring programs to assure data quality. This document is similar to the EPA's QAPP document except that it has the purpose of standardizing the structure and monitoring techniques of volunteer monitoring programs in Texas in order to integrate the data into the TNRCC's data base.

Many of the statistical analysis techniques used for the study came from a class exercise from Dr. David Maidment's CE397: *Environmental Risk Assessment* Class, Department of Civil Engineering, The University of Texas at Austin, Spring 1998. The exercise covered the use of statistics, using the statistical tools in Microsoft's Excel, to describe the nature of data sets, distribution analysis, and the properties of the mean of the data. The methods for interpreting the descriptive statistics is discussed such as:

1. Determining the best form to use for analysis, original or transformed
2. Use of the frequency histogram to visually assess the nature of the distribution of the data set.
3. Standard Error of the mean to determine if there is a correlation between the number of data points in the data sets and the precision in the result.
4. The difference in the means of two data sets (t-test) to compare two independently obtained data sets and when it is appropriate to use this technique.

The exercise contained detailed discussion and examples of the use of these statistical techniques using environmental data.

Water Quality Monitoring and *Water Quality Assessments* are books published for the United Nation's Environment Program discussing the form of monitoring organizations and the techniques used to accomplish the stated goals of a monitoring organization. The technical and organizational structures needed to assure that quality data is collected is provided.

Water Quality Assessment was particularly useful for this study because it discussed the use of statistics and their application specifically to water quality data. Detailed procedures, examples and interpretations of the results were given. The techniques used were based on the goals of the sampling techniques utilized and the questions that one wants to have answered by the statistical analysis. The following topics relevant to this study were discussed:

1. The use of basic statistics to summarize and assess small or large, simple or complex data sets.
2. The use of descriptive statistics to summarize water quality data sets into simpler and more understandable forms.
3. How to determine the distribution of the data sets.
4. The construction and interpretation of cumulative distribution plots and histograms.
5. The types of methods to use based on the type of distribution of the data.
6. Standard error of the mean analysis.
7. Hypothesis testing, the t-test.
8. Regression analysis and the use of a trigonometric function.

The author also went into some detail about graphical presentation of the results, how to most effectively summarize water quality characteristics.

The *Handbook of Hydrology* has several chapters dedicated to statistical analysis of hydrologic data. There is a very useful discussion on the basic concepts of statistical techniques and their application specifically to hydrologic data. Some of the topics covered included:

1. Histograms and the choice of class intervals utilized for the histogram
2. Quantile plots
3. Hypothesis testing, the t-test and its use for comparing data sets.
4. Multiple regression technique.

Handbook of Hydrology described in detail the use of the sinusoidal function to describe the periodic functions of a time variable. The basic formula used in this study for the Fourier analysis and the various forms of the function to describe the type of cycle that is to be modeled is discussed.

Applied Hydrology, a hydrology textbook, provided the information for constructing the quantile plots used in the study. The section on probability plotting discussed the method of transforming the data to a special probability value for accurately determining whether the data had a normal distribution.

Tips for Statistical Analyses of Parallel Studies, an article by Woodrow Setzer in "The Volunteer Monitor" examined the statistical methods used for Often the main reason for doing parallel testing is to assure government agencies that the data collected by volunteers is reliable enough form the agencies to use. The article points out that parallel testing has been done consistently, resulting in attitude changes toward volunteer data. Another equally important reason for the parallel testing is for the volunteer programs own internal use. The results of these tests can point out the program's strengths and weaknesses. Problems can be spotted with these tests and ways to make improvements can be worked out.

These tests are also good for the moral of the volunteer monitors. They appreciate that their efforts as volunteer monitors is being taken seriously and that the quality of their measurements is being taken into account.

The *Handbook of Hydrology* has a chapter specifically about statistics and hydrologic data. Specific information on basic statistics and more sophisticated statistical analysis can be found there. This chapter examines the difficulty of explaining or predicting hydrologic variables and the sources of uncertainty. It is pointed out how the type of information used for this study is observational data, rather than experimental. This means that the conditions from which the measured samples are taken can not be duplicated like one would do in a laboratory experiment. The sources of uncertainty under with hydrologic systems are:

1) The inherent randomness of the driving variables and the hydrologic system
2) Sampling error due to the fact that the measured sample is a very small part of large population.
3) Incorrect understanding of the processes involved. For instance, understanding the influences on the levels of DO or TDS and how they may vary even over the distance between where the volunteer monitor takes measurements versus the professional monitor.

The concept of using statistics to go beyond simply describing the larger population is examined. The use of statistics to quantify the uncertainty in the knowledge about the population is explained. Knowledge of the magnitude of the uncertainties is essential to identifying those areas where it will be worthwhile to collect additional data. In the case of this study, statistics are used to quantify if the additional data collected by volunteers can be a valuable addition to the data collected by professional monitors.

Very detailed explanation of statistical methods for water resource data can be found in *Statistical Methods in Water Resources*. One particularly useful discussion is in regards to "outlier data", data points whose values are quite different than others in the data set. Outlier data was a concern in this study due to the effects it had on the statistical methods, such as causing larger values of skewness to the data set than if the outlier data value was not included. This reference also contains detailed discussions of distribution plots, regression analysis, and histograms.

3 Statistical Analysis

Statistics is the science that deals with the collection, tabulation, and analysis of numerical data. The properties of a population are assessed based on the properties of a sample from that population. A measure of the uncertainty of the knowledge of this population can be determined by statistics. One can also determine if independent data sets are statistically different.

Intuitively one can see that uncertainty would decrease by increasing the number of samples from the population. Statistics provides a way of quantifying this. For this study, statistics will be used to determine if the increase in the number of water quality samples allowed by volunteer monitoring programs will have a significant effect on decreasing the uncertainty in estimation of population parameters.

This science is particularly important for the task of assessing the quality of a body of water. Water quality assessment relies heavily on statistics since, typically, the overall quality of an entire water body is assessed by a set of grab samples taken at one moment and at one or a few locations on that water body. Natural bodies of water are constantly changing, physically and chemically. There are many processes, natural and those caused by man, effecting the parameters used to assess the quality of a water body. The chemical constituents which are measured to assess water quality are constantly being affected by biological, chemical and physical processes over varying temporal cycles, daily, seasonal and climatic. Random events such as storm events have an effect on water quality.

Federal environmental regulations have forced states to define uses for the main water bodies in the state and the water quality parameters used to determine if the quality of the water is sufficient for that use. Typically, the quality of a river segment or lake is assessed by small samples, only millileters in volume, drawn from a few points on the water body. These samples are drawn infrequently, only several times a year for the typical water quality assessment programs.

In this study basic descriptive statistics and other statistical methods will be used to assess the characteristics of water quality data collected by volunteer monitors. Any statistical analysis begins by stating the goals of the analysis.

3.1 Goals of Statistical Analysis

1) Determine the nature of the data sets using basic statistical methods.

2) Statistically compare the professional and volunteer data sets collected at the same site or sites very close to one another. Evaluate which methods are best to compare them.

3) Estimate the increase in confidence that the sample represents the population with the addition of volunteer data.

4) Compare the water quality data collected at different sites within the same river segment.

5) Determine the increase in data confidence when volunteer data is combined with professional data (at the same site).

6) Over a given time interval, estimate whether there is a substantial increase confidence in the data with the increase in data measurements inherent with volunteer data.

3.2 Description of the Statistical Analysis

a) **Descriptive Statistics**: the mean, median, variance, standard deviation, skewness, and range. These statistics indicate the type of distribution you are dealing with. This information is useful for making decisions on how to handle the data statistically. For instance, highly skewed data with a large range may require using the logarithms of the data for the analysis. These basic statistics were determined using the Descriptive Statistics tool built into Microsoft Excel.

1. What kind of distribution is indicated? It is important to know to what degree the data follows a normal distribution. Most of the statistical procedures used in this analysis require the data to be normally distributed to be fully valid.

2. Are the mean, mode and median values close to each other? A normal distribution is indicated if the mean and median of the data set are nearly equal.

3. How big is the standard deviation? This value gives and indication of the spread of the data value around the mean. Coupled with the mean, the standard deviation gives a good indication of the range of values in the data set.

4. What is the skewness and kurtosis values? The skewness is a measure of how asymmetrically the data are distributed about the mean. The kurtosis measures the extent to which data are more peaked or more flat-topped than in the normal curve. Values close to zero indicate normal distribution.

5. Do the Maximum/Minimum values fall within an expected range? This analysis can be useful for editing bad data out of the data set. They are also useful to describe how much variability there is in the data set.

b) **Histograms:** The histogram is useful visual tool for assessing the distribution of the data. If the histogram resembles the standard, bell shaped curve of a typical normal distribution then it is assumed that the data set has a normal distribution. They cannot be used for more precise judgements such as depicting individual values.

c) **Quantile Plots:** Quantile plots also can be used to determine if a data set is normally distributed.

Quantile plots can be constructed by plotting the data and its probability value on special log normal plotting paper or the data can be transformed into lognormal values and plotted. The value of the standard normal variable, *z*, is used as the horizontal axis to linearize the plot; this is equivalent to using normal probability plotting paper.[5]

The following is the procedure for transforming the data:

i) Arrange the data in descending order

ii) Give each value a rank number, *i*, .

iii) Calculate the probability value *p* (Blom, 1958) for each data point: $p_i = \dfrac{i - 3/8}{n + 1/4}$

iv) An intermediate variable w is calculated: $w = \left[\ln\left(\dfrac{1}{p^2}\right)\right]^{1/2}$

v) The frequency factor for normal distribution, *z*, is calculated using the following formula: $z = w - \dfrac{2.515517 + 0.802853\,w + 0.010328\,w^2}{1 + 1.432788\,w + 0.18969\,w^2 + 0.001308\,w^3}$

vi) Plot the data versus the corresponding value of *z*.

vii) Plot the normal distribution values which correspond to the data set versus the corresponding value of z using the following formula for the predicted value: $y = \bar{y} + \mu z$

where: $\bar{y}$ is the mean of the data set, μ is the standard deviation of the data set

[5] Chow V.T., Maidment D.R. and Mays L.W., "Applied Hydrology", McGraw-Hill, Inc., 1988

From the plot a visual inspection of how close to two plots correspond can give an indication of whether the data is normally distributed and the effects of extreme values. A more accurate method is to calculate the correlation coefficient of the two lines. This value can be compared to the values in Table 3-1, which is specifically for $p_i = (i-0.375)/(n+0.25)$, to determine if the data is normally distributed.

Table 3-1 Critical Values, Probability Plot

Lower Critical Values of the Probability Plot Correlation Test Statistic for the Normal Distibution.			
	Significance Level		
n	**0.10**	**0.05**	**0.01**
10	0.9347	0.9108	0.8804
15	0.9506	0.9383	0.9110
20	0.9600	0.9503	0.9290
30	0.9707	0.9639	0.9490
40	0.9767	0.9715	0.9597
50	0.9807	0.9764	0.9664
60	0.9835	0.9799	0.9710
75	0.9865	0.9835	0.9757
100	0.9893	0.9870	0.9812
300	0.99602	0.99525	0.99354

Just one or a few outlier data points can cause the correlation coefficient to fall beyond the limits specified in the table. If this is the case a judgement should be made as to whether these values should be excluded from the calculation. Table 3-2 shows the spreadsheet used to construct the quantile plot shown in Figure 3-1.

Table 3-2 Spreadsheet For Quantile Plot

Col. Riv. @ Beason's Park- Columbus- Segment 1402
CRWN DATA 1402.0250

I	**DO (mg/l)**	**p**	**w**	**z**	**DO Desc.**	**y**
1	7.1	0.01	3.04	2.27	16.15	12.06
2	7.2	0.03	2.71	1.90	13.40	11.40
3	7.4	0.04	2.52	1.70	10.80	11.04
..	..	..	..	..	..	..
61	9.65	0.96	0.29	-1.51	5.33	5.26
62	8.10	0.97	0.23	-1.68	4.95	4.95
63	8.00	0.99	0.14	-1.95	4.30	4.46

In this example the two extreme data values (16.15 and 13.4 mg/l) caused the correlation factor be lower than the lower limit (0.927 vs 0.971) for normal distribution. By eliminating these values the correlation coefficient value is 0.997.

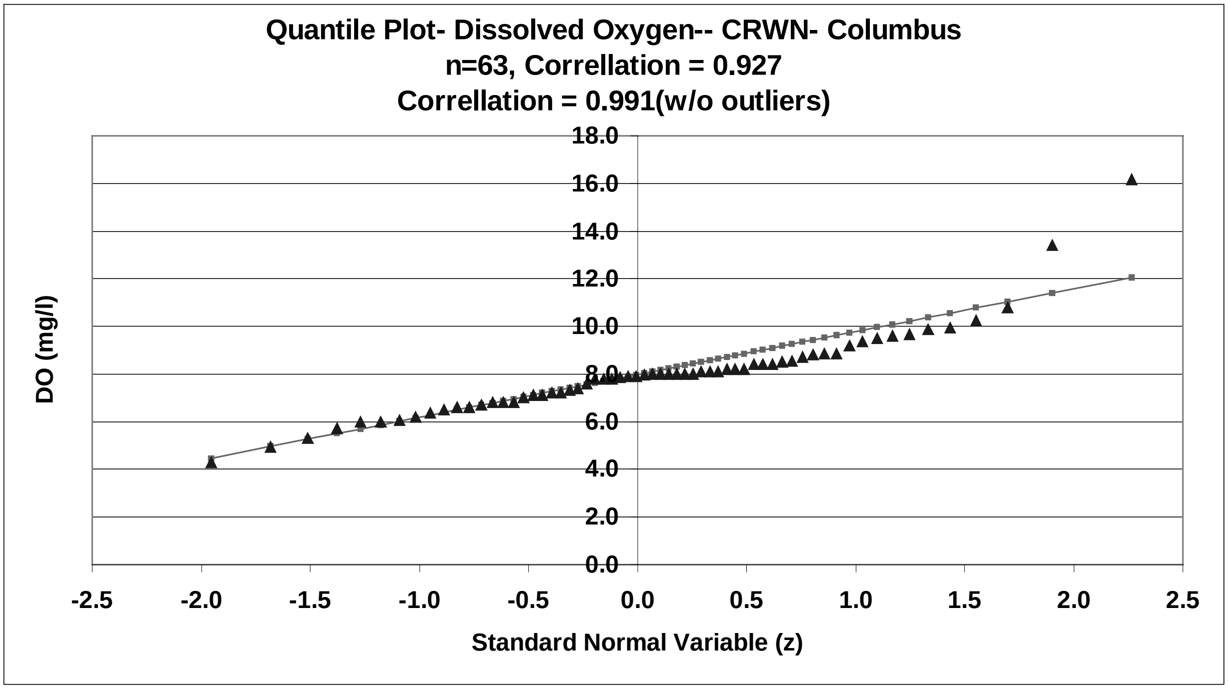


Figure 3-1 Quantile Plot

These results indicate that the data set has a normal distribution.

d) **Standard Error of the Mean analysis**: This process produces the "mean of the means" and an estimate of their spread about that mean. The standard deviation of the mean s_x is known as the standard error of the mean, with which estimates of the reliability of the data mean can be made.

How much is the standard deviation of the mean decreased with the addition of volunteer data? Typical values of n for a calendar year are: professional data, n=6, volunteer data, n>20. Compare the range of the 95% confidence limit (m+2sd) around the mean for each set of data.

e) **<u>The t-test</u>**: With the t-test one can answer the question: Are the means of the professional and volunteer data sets statistically different? This can be done by re-stating in the form of hypotheses:

Null hypothesis: there is no significant difference in the means of the data sets.

Alternate hypothesis: there is a significant difference in the means of the data sets.

As noted above, it is important that the data sets be nearly normally distributed for this test to be reliable. The value of *t* indicates to what degree the means of the two data sets are different. Each unit value of *t*, positive or negative, is one standard error. An absolute value of 2 for the t-stat is an estimate of the 95% confidence limit around the mean, so an absolute value of *t* of less than two indicates that there is no statistical difference between the means of the two data sets.

Knowing that both data sets are normally distributed the t-test is used to determine if the two data sets are statistically different from each other. The t-statistic is determined by the following formula: $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{S_x^2}{n} + \dfrac{S_y^2}{m}}}$ where $\bar{x}$ and $\bar{y}$ are the means of the two data sets, S^2_x and S^2_y are the standard deviations and *n* and *m* are the number of measurements. Microsoft Excel has the t-test as part of it's statistical package. The t-test was done both with the above formula in the spreadsheet and using the t-test in the software package to confirm the accuracy of the software package.

f) **Fourier Analysis**. The Fourier analysis is used to analyze a data set that displays a cyclical nature such a diurnal or seasonal cycle. An analysis of the professional and volunteer dissolved oxygen data sets is done using the Fourier regression technique for several reasons:

1. Dissolved oxygen levels in natural waters vary seasonally with the temperature of the water. The Fourier analysis is used to compare the best-fit lines of the professional and volunteer data and the mean values for the two data sets determined by the analysis can be compared.
2. Volunteer data sets can have the general characteristic of being inconsistent temporally. This is shown numerically in Figure 3-4. Six of the seven volunteer data sets used in this study are from sites being monitored by high school science classes which resulted in a characteristic of a lack of late spring and summer data for these sites. The Fourier analysis is a method with which to predict what the measurements may have been at these sites during these periods.

Table 3-3 CRWN Seasonal Data Distribution

Site	Number of Samples (n)	
	Warm Months (4/1-9/30)	
	Volunteer (n)	Samples %
Lake Austin	91	65
Bastrop	100	21
Smithville	37	8
LaGrange	33	15
Columbus	60	40

A detailed description of the technique used is examined in the chapter for Fourier analysis in this report.

3.3 Choosing sites to analyze

The sites for the study were chosen based on the following criteria:

1) Professional and a volunteer monitoring site are at the same location or within close proximity (about one mile) to each other.

2) The site is on a water body with significant year-round flow.

3) The site is located on river segment identified by the TNRCC

4) At least two years of consistent volunteer monitoring data up to December 1997 are available.

The following is a description of the location of the monitoring sites, the location of the volunteer sites in relation to the professional sites, the site number used for this study, the TNRCC site number and significant features near the sites.

The sites monitored by professional monitors are called LCRA (Lower Colorado River Authority) and the volunteer sites are called CRWN (Colorado River Watch Network). [6]

Figure 3-2 Map of Monitoring Sites

Lake Austin,

LCRA 1: LCRA Site number 1403.0100 at Tom Miller Dam. This station is located over the Colorado River thalweg at the deepest portion, a short distance up the reservoir from the dam outlet.

CRWN 1: CRWN Site number 1403.0100 Colorado River. level 1 site; This site is approximately 0.4 miles upstream of the LCRA site.

[6] Lower Colorado River Authority, *The Texas Clean Rivers Program, Technical Report*, (Austin, Texas: October 1, 1996)

Bastrop

LCRA 2: LCRA Site number 1434.0600 in Bastrop, Texas at Loop 150. This station is located approximately on-half mile upstream of the Bastrop wastewater treatment facility discharge. This provides information for comparison with water quality data from stations upstream and downstream, and will illustrate the water quality of the river before it receives effluents from Bastrop.

CRWN 2: CRWN Site 1428.0600 (level 2) located in Bastrop, Texas at the same location as the LCRA site.

Smithville

LCRA 3: LCRA Site number 1434.0500 (formerly 1402.0505f) in Smithville, Texas at State Highway 95. This station is located approximately one-quarter mile upstream of the Gazley Creek confluence. This station will provide data for upstream of the Smithville wastewater treatment facility discharge. This data is valuable for the analysis of downstream changes in water quality.

CRWN 3: CRWN Site Number 1402.0505 (level 2), in Smithville, Texas at Highway 71, located approximately1 mile downstream of the LCRA site. This location is also downstream of the Gazley Creek confluence and the Smithville wastewater treatment plant, which could be significant when comparing the data from this site with the LCRA site which is upstream of the confluence.

LaGrange

LCRA 4: LCRA Site number 1434.0400 in LaGrange, Texas, located upstream of old HWY 71 bridge and downstream of new HWY 71 bridge. This station is located approximately one-half mile upstream of the LaGrange wastewater discharge facility discharge. Data from this station is useful for comparison with data collected downstream of the outfall and for analysis of downstream water quality changes

CRWN 4: CRWN Site number 1434.0400 (level 2) in LaGrange, Texas located at White Rock Park. This location is three to four miles downstream of the LCRA site. The wastewater treatment plant discharge is between the two sites.

Columbus

LCRA 5: LCRA Site number 1402.0300 in Columbus, Texas. This site is located approximately six miles upstream of the Columbus wastewater treatment facility outfall at Business Hwy. 71. Data from this station is useful for analysis of downstream changes in water quality.

CRWN 5: CRWN Site number 1402.0250(level 2) in Columbus, Texas at Beasons Park. This site is located 4 miles downstream of the LCRA site. This site is also downstream of the Columbus wastewater treatment plant.

3.4 Parameters Studied and Measurement Methods

Dissolved Oxygen (DO) was measured by the volunteer monitors using the Winkler titration method. This is a widely used and accepted method of measuring dissolved oxygen. The accuracy of the volunteer monitor's field test kit is ±0.5 ppm. A judgement must be made, for one part of the test, as to whether the color has changed from blue to clear during a titration. One drop from the titrator is equivalent to 0.3 mg/l on the titrator's scale. This accounts for the rather large range in the accuracy. The professional equipment has an accuracy of ±0.2 mg/l.

This particular parameter is subject to a high degree of quality control and was found to be the most reliable and accurate parameter measured, according to the managers of the River Watch. Duplicate measurements are made and they must be within 0.6 mg/l of each other for the results of the test to be valid.

A potential problem with using this parameter is that many conditions can effect the dissolved oxygen level such as diurnal variations due to vegetation, temperature and sun light, spatial variations and seasonal variations. One factor that may be relevant is that CRWN monitors may sample any time between 7:00 a.m. and 7:00 p.m., while LCRA monitors sample between 9 am and noon.

It is important to note whether there is consistency in the time of day that the measurements are taken and that the samples are taken from the same depth in the water body.

Total Dissolved Solids (TDS) is determined by measuring specific conductance using a conductivity meter. The TDS meter used by the volunteer monitors has an accuracy of ±39.8 mg/l. The accuracy of the professional instrument is ±1%. This method of measuring TDS is widely used and accepted.

The CRWN monitors use meters that directly convert the conductivity measurement into TDS units. The LCRA monitors use a meter that measures conductivity in μmhos/cm. To compare the two data sets the LCRA data was converted into TDS (mg/l) by multiplying the conductance by a conversion factor which varied slightly between the river segments. The conversion factors were obtained from the LCRA.

TDS is a good parameter to study because it is and it is least affected by diurnal variations in other water parameters.

3.5 Parallel Measurements Analysis

The analysis of parallel testing is a way to get some idea as to how well two independent measurements of the water quality parameters used in this study match in an actual "side by side" situation, the same place at the same time with two different instruments. The measurements that are compared in the study were done in a random fashion, at different times and at different points on the river.

The data used in this analysis were compiled from Quality Assurance/Quality Control (QA/QC) meetings the staff of the CRWN has with the volunteer monitors. During part of that meeting the CRWN staff person takes parallel measurements with the volunteer monitor. The volunteer monitors represented in this analysis are not the same monitors in the study.
Figures 3-3 and 3-4 show the results of the parallel tests. The dashed line originating at the (0,0) point represents where the points would lie if the two measurements equaled each other. The solid line is the best-fit line from linear regression. The results show that, when the accuracy of the instruments is taken into account, the measurements match quite well.

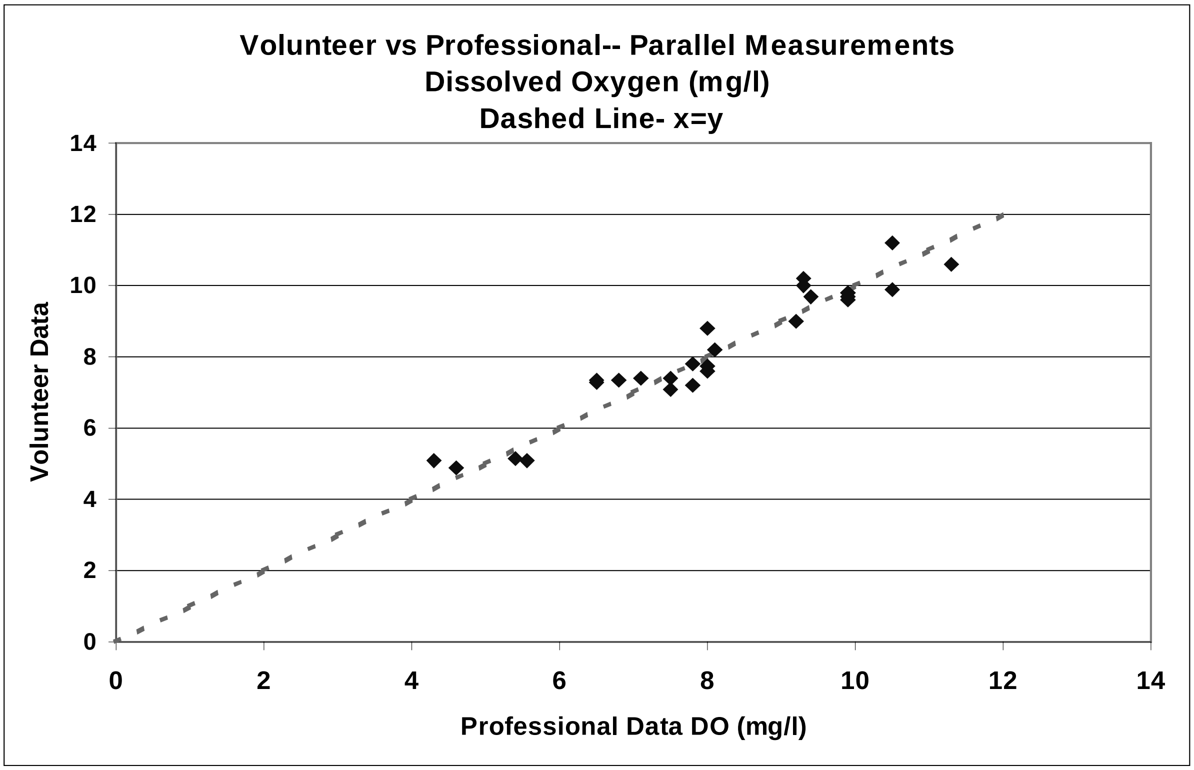


Figure 3-3 Parallel Testing Plot, DO

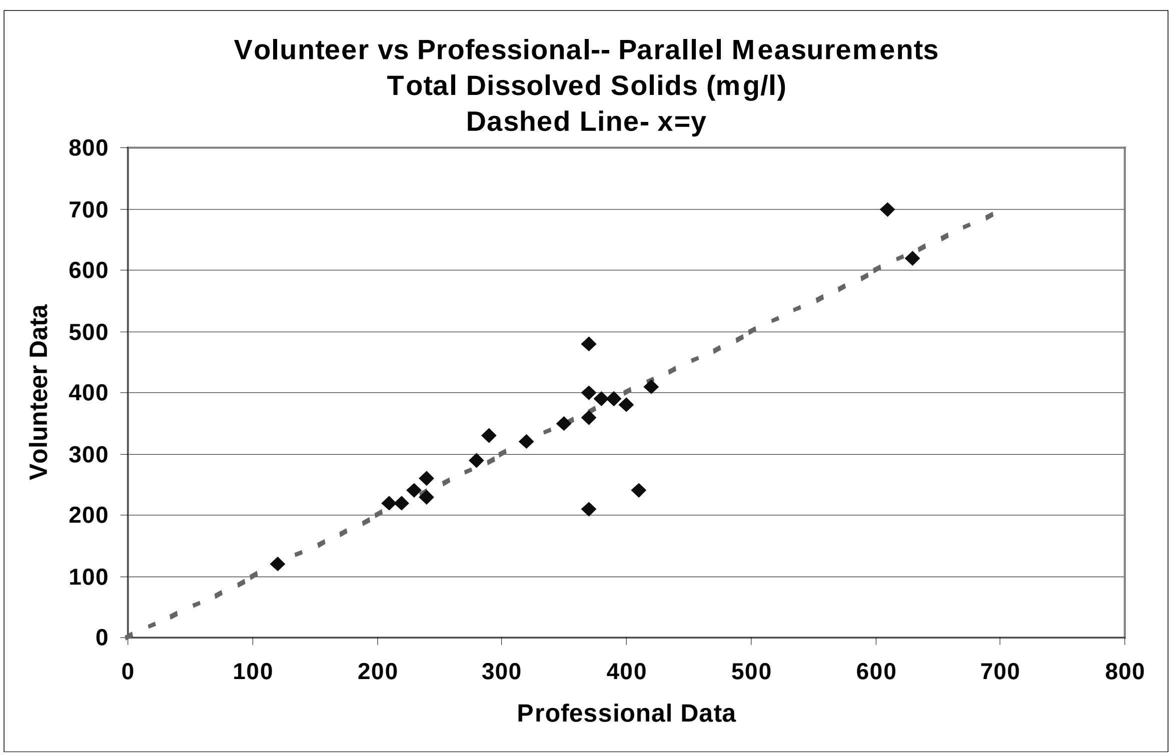


Figure 3-4 Parallel Testing Plot, TDS

4 Fourier Analysis of Water Quality Data

4.1 Theory

If it is desired to study the cyclical behavior over time of a water quality variable, fourier analysis is an effective tool to utilize. The cyclical behavior could be diurnal, seasonal or man-made. The periodic function may be responding to temperature, sunlight, releases from dams, etc. This form of analysis can be an effective tool for predicting the behavior of the variable of interest or it could be used to compare two independent sets of data.

The coefficients of a fourier series can be found by multiple regression analysis. The cycle is described by a sine function with the general form of:

$$y = a_0 + \sum_{j=1}^{J} (a_j \cos(2j\omega t) + b_j \sin(2j\omega t) + \varepsilon)$$

where: $\omega = 2\pi / 365$

t = Julian days

y = water quality parameter

a,b = coefficients

The values of j represents the number of cycles within the given time period. In this case the time period is 365 days over one 2π (circular) cycle for the sinusoidal function. j=1 would represent a 12 month cycle, j=2 a 6month cycle, etc..

This is the form that will be used to analyze the variation of dissolved oxygen with the in the example below. The coefficients for the regression are determined using the regression tool in Microsoft Excel.

4.2 Set-Up

The fourier analysis will now be used to compare two data sets with cyclical behavior and as a tool for predicting the variable where data is missing. The variable is dissolved oxygen measured by volunteer monitors (CRWN) and professional monitors (LCRA) at approximately the same location on the Colorado River. The raw data has the form shown in Table 4-1.

Table 4-1 Raw Data Example

CRWN Data		LCRA Data	
Date	**DO (mg/l)**	**Date**	**DO (mg/l)**
04/19/93	4.8	04/19/93	8.9
06/21/93	6.2	06/01/93	8.9
06/25/93	7.8	08/11/93	8.2
07/09/93	7.8	10/07/93	7.5
07/16/93	8.3	02/14/94	10.4
07/23/93	6.7	04/27/94	9.2
07/30/93	6.7	06/23/94	9
08/06/93	7.7	08/29/94	7.7
08/13/93	6.7	10/18/94	6.4
08/27/93	7	12/22/94	9

A scatter plot (Figure 4-1) of the data shows the cyclical nature of the data:

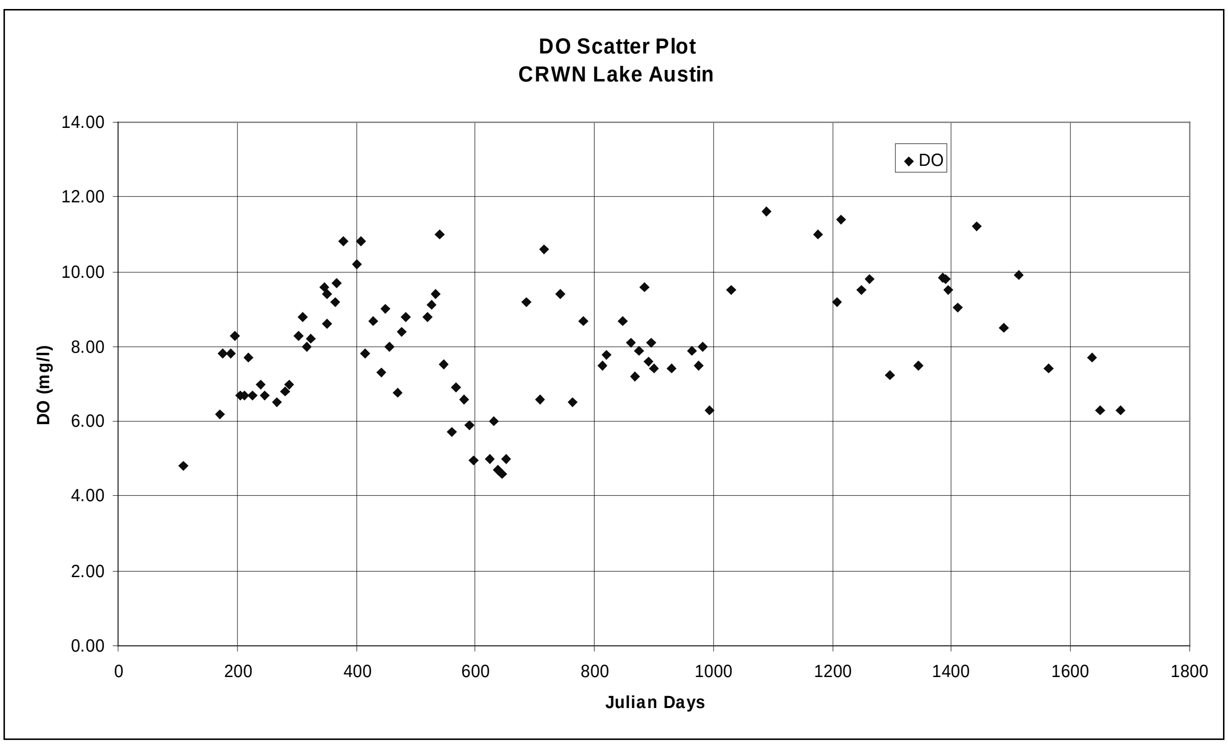


Figure 4-1 DO Scatter Plot

A spreadsheet is developed for each data set. The spreadsheet for j=2 is shown in Figure 4-2 below.

	A	B	C	D	E	F	G	H	I	J
1		CRWN DATA				e				
2	Index	Date	yearfrac	Julian	DO	2pi*t/365	cos (e)	sin (e)	cos(2e)	sin(2e)
3	1	04/19/93	01/01/93	109	4.80	1.88	-0.30	0.95	-0.82	-0.57
4	2	06/21/93	01/01/93	172	6.20	2.96	-0.98	0.18	0.94	-0.35
5	3	06/25/93	01/01/93	176	7.80	3.03	-0.99	0.11	0.98	-0.22
6	4	07/09/93	01/01/93	190	7.80	3.27	-0.99	-0.13	0.97	0.26
7	5	07/16/93	01/01/93	197	8.30	3.39	-0.97	-0.25	0.88	0.48
8	6	07/23/93	01/01/93	204	6.70	3.51	-0.93	-0.36	0.74	0.67
9	7	07/30/93	01/01/93	211	6.70	3.63	-0.88	-0.47	0.56	0.83
10	8	08/06/93	01/01/93	218	7.70	3.75	-0.82	-0.57	0.34	0.94
11	9	08/13/93	01/01/93	225	6.70	3.87	-0.74	-0.67	0.11	0.99
12	10	08/27/93	01/01/93	239	7.00	4.11	-0.56	-0.83	-0.37	0.93
13										

Figure 4-2 Spreadsheet Layout for Fourier Analysis

Column C is used as a reference date used in the formula, *yearfrac*, for determining the Julian day.

The formulas used in colums D through J are shown in Figure 4-3 below.

D	E	F	G	H	I	J
		e				
Julian	DO	2pi*t/365	cos (e)	sin (e)	cos(2e)	sin(2e)
=INT(YEARFRAC(C3,B3,1)*365.25)+1	4.8	=2*PI()*D3/365	=COS(F3)	=SIN(F3)	=COS(2*F3)	=SIN(2*F3)
=INT(YEARFRAC(C4,B4,1)*365.25)+1	6.2	=2*PI()*D4/365	=COS(F4)	=SIN(F4)	=COS(2*F4)	=SIN(2*F4)
=INT(YEARFRAC(C5,B5,1)*365.25)+1	7.8	=2*PI()*D5/365	=COS(F5)	=SIN(F5)	=COS(2*F5)	=SIN(2*F5)
=INT(YEARFRAC(C6,B6,1)*365.25)+1	7.8	=2*PI()*D6/365	=COS(F6)	=SIN(F6)	=COS(2*F6)	=SIN(2*F6)
=INT(YEARFRAC(C7,B7,1)*365.25)+1	8.3	=2*PI()*D7/365	=COS(F7)	=SIN(F7)	=COS(2*F7)	=SIN(2*F7)
=INT(YEARFRAC(C8,B8,1)*365.25)+1	6.7	=2*PI()*D8/365	=COS(F8)	=SIN(F8)	=COS(2*F8)	=SIN(2*F8)
=INT(YEARFRAC(C9,B9,1)*365.25)+1	6.7	=2*PI()*D9/365	=COS(F9)	=SIN(F9)	=COS(2*F9)	=SIN(2*F9)
=INT(YEARFRAC(C10,B10,1)*365.25)+1	7.7	=2*PI()*D10/365	=COS(F10)	=SIN(F10)	=COS(2*F10)	=SIN(2*F10)
=INT(YEARFRAC(C11,B11,1)*365.25)+1	6.7	=2*PI()*D11/365	=COS(F11)	=SIN(F11)	=COS(2*F11)	=SIN(2*F11)
=INT(YEARFRAC(C12,B12,1)*365.25)+1	7	=2*PI()*D12/365	=COS(F12)	=SIN(F12)	=COS(2*F12)	=SIN(2*F12)

Figure 4-3 Spreadsheet Formulas for Fourier Analysis

Now the regression analysis can be performed. The goal is obtain the values for the coefficients a_o, a_j and b_j in the regression formula and to see how well the calculated curve fits the data. Under the Tools Menu go to Data Analysis and choose the Regression tool as shown in Figure 4-4 below.

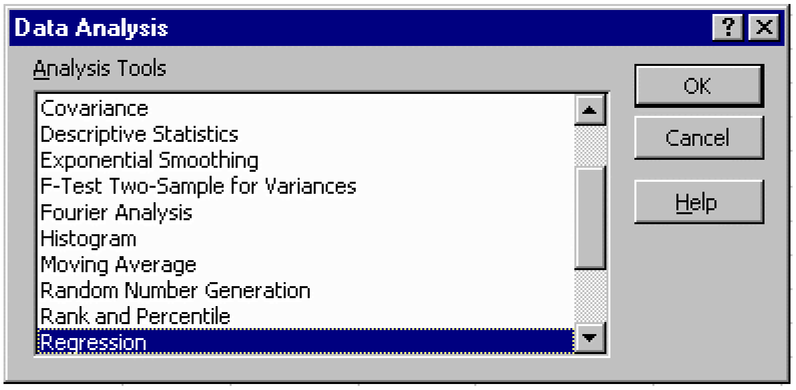


Figure 4-4 Excel Regression Tool

The regression tool window is shown in Figure 4-5 below.

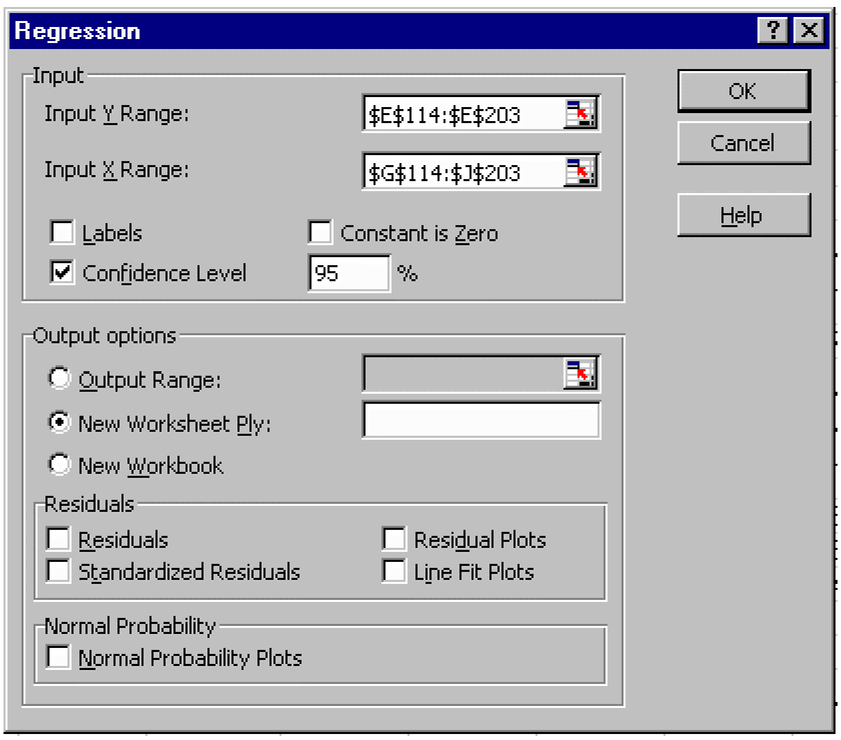


Figure 4-5 Excel Regression Tool Window

The Input Y Range is the column with the dissolved oxygen data and the Input X Range is the field with the values of the sine and cosine functions, for j=1 columns G and H and for j=2 columns G through J. The output of the regression analysis, for j=2, is shown in Table 4-2.

Table 4-2 Regression Statistics, Excel Output

SUMMARY OUTPUT- CRWN Lake Austin

j=2

Regression Statistics	
Multiple R	0.625
R Square	0.391
Adjusted R Square	0.362
Standard Error	1.304
Observations	90

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	92.748	23.187	13.635	0.000
Residual	85	144.549	1.701		
Total	89	237.296			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.187	0.141	58.188	0.000	7.907	8.467
X Variable 1	0.959	0.198	4.852	0.000	0.566	1.352
X Variable 2	0.706	0.200	3.526	0.001	0.308	1.104
X Variable 3	0.671	0.190	3.534	0.001	0.294	1.049
X Variable 4	-0.521	0.204	-2.549	0.013	-0.928	-0.115

For this example the coefficients for the regression formula are: a_o=8.187 mg/l, a_1=0.959 mg/l, b_1=0.706 mg/l, a_2=0.671 mg/l, b_2=-0.5212. The a_o value can be interpreted as being the mean value of DO as determined by the regression. So the formula for the best-fit line for this set of data is:

$$y = 8.1869 + 0.9690\cos\frac{2\pi t}{365} + 0.7056\sin\frac{2\pi t}{365} + 0.6712\cos\frac{4\pi t}{365} + 0.5212\sin\frac{4\pi t}{365}$$

The t-stat values for each variable having an absolute value of greater than 2 indicates that each factor in the equation is contributing significantly to fit of the line. In this case a check into whether the j=2 expansion gives a better model than the j=1 expansion is appropriate. From the shape of the data distribution on the scatter plot it appears that the j=1 expansion would be appropriate to fit the data.

4.3 Results

Values of *j*=1 and j=2 are used in the fourier expansion to determine which one best represents the water quality data's seasonal variability. When this is determined the fourier analysis can be used to compare the volunteer and professional data. The model could also be used to predict values in time periods when real data is lacking.

Table 4-2, below, show the results for j=1.

Table 4-3 Regression Statistics, j=1

CRWN-Lake Austin-SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.505
R Square	0.255
Adjusted R Square	0.238
Standard Error	1.426
Observations	90

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	60.476	30.238	14.878	0.000
Residual	87	176.820	2.032		
Total	89	237.296			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.2163	0.1536	53.4992	0.0000	7.9110	8.5215
X Variable 1	0.8892	0.2126	4.1814	0.0001	0.4665	1.3118
X Variable 2	0.7930	0.2174	3.6482	0.0004	0.3610	1.2251

Graphically the results have the form shown in Figures 4-6 and 4-7.

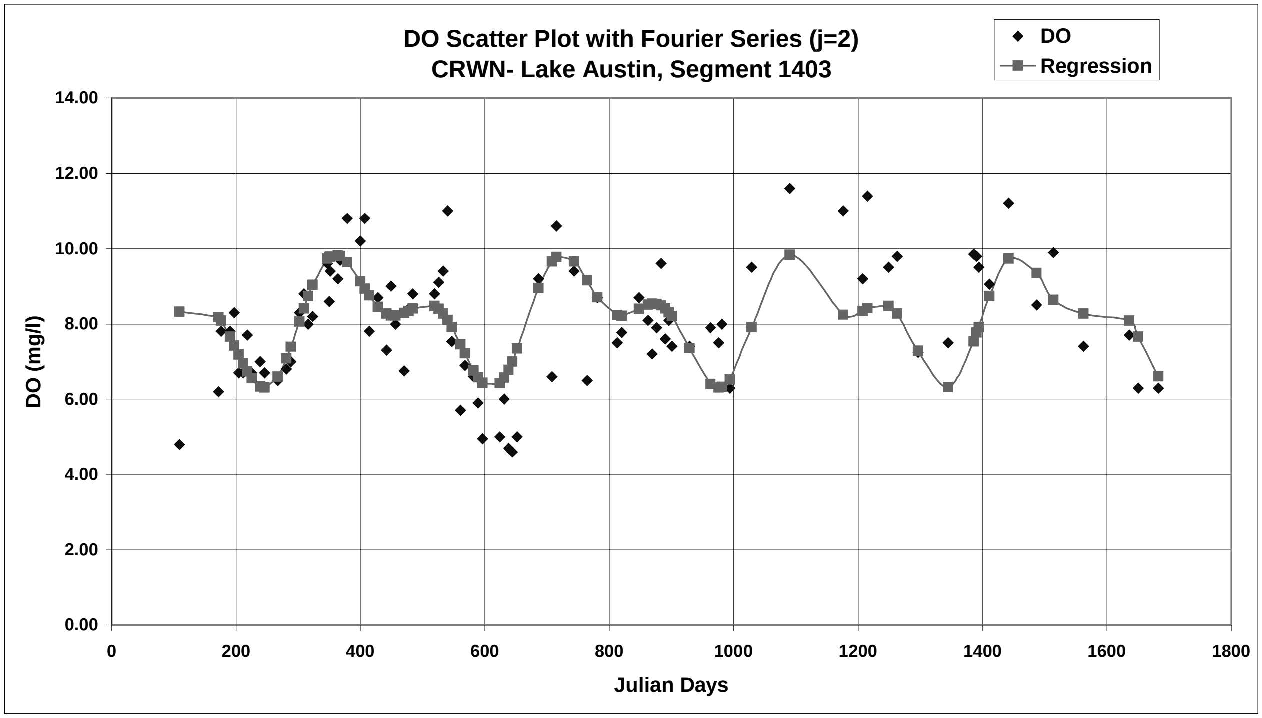


Figure 4-6 Plot of Regression Results, j=2

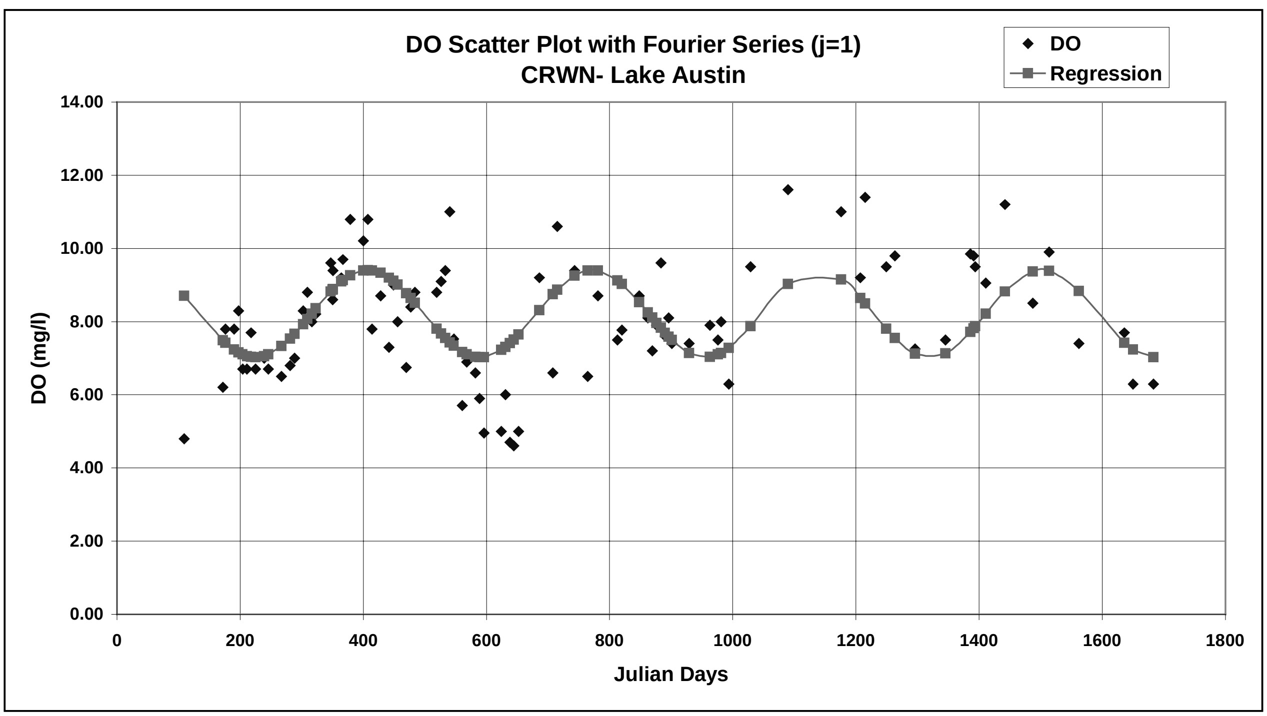


Figure 4-7 Plot of Regression Results, j=1

It can be seen how expanding the equation to j=2 resulted in the best-fit line to have larger extremes, larger maximum and smaller minimum values. A line representing the seasonal variation of the data may best be represented with values of j=1. The same procedure done to the LCRA data had the same conclusion.

A smooth best-fit line can now be produced for each data set using uniform time intervals and the two best-fit lines can be compared graphically. Another Excel spreadsheet was set up with uniform time intervals (7days in this example). The form of the spreadsheet is shown in Figure 4-8.

	A	B	C	D	E	F	G	H	I	J	K
1	**Index**	**Date**	**yearfrac**	**Julian**	**DO**	**2pi*t/365**	**cos (e)**	**sin (e)**	**cos(2e)**	**sin(2e)**	**Regression**
2	1	08/23/93	01/01/93	235	7.10	4.05	-0.62	-0.79	-0.23	0.97	7.33
3	2	08/30/93	01/01/93	242	8.10	4.17	-0.52	-0.85	-0.46	0.89	7.16
4	3	09/06/93	01/01/93	249	9.10	4.29	-0.41	-0.91	-0.66	0.75	7.01
5	4	09/13/93	01/01/93	256	10.10	4.41	-0.30	-0.95	-0.82	0.57	6.89
6	5	09/20/93	01/01/93	263	11.10	4.53	-0.18	-0.98	-0.93	0.36	6.79
7	6	09/27/93	01/01/93	270	12.10	4.65	-0.06	-1.00	-0.99	0.13	6.74
8	7	10/04/93	01/01/93	277	13.10	4.77	0.06	-1.00	-0.99	-0.11	6.73
9	8	10/11/93	01/01/93	284	14.10	4.89	0.18	-0.98	-0.94	-0.35	6.77
10	9	10/18/93	01/01/93	291	15.10	5.01	0.29	-0.96	-0.83	-0.56	6.87

Figure 4-8 Excel Spreadsheet, Regression Plot

Compare the Expansions (j=1 and j=2)

A plot of the two expansions j=1 and j=2 and the data are shown below in Figures 4-9 and 4-10.

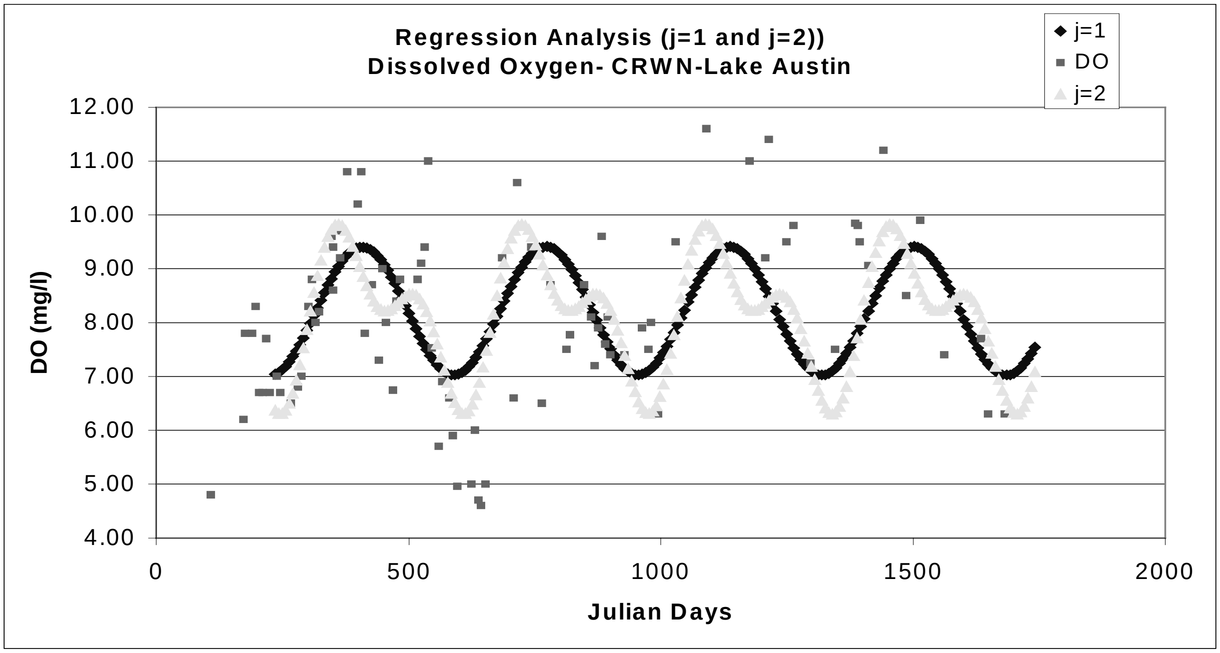


Figure 4-9 CRWN DO Fourier Plots, j=1 vs. j=2

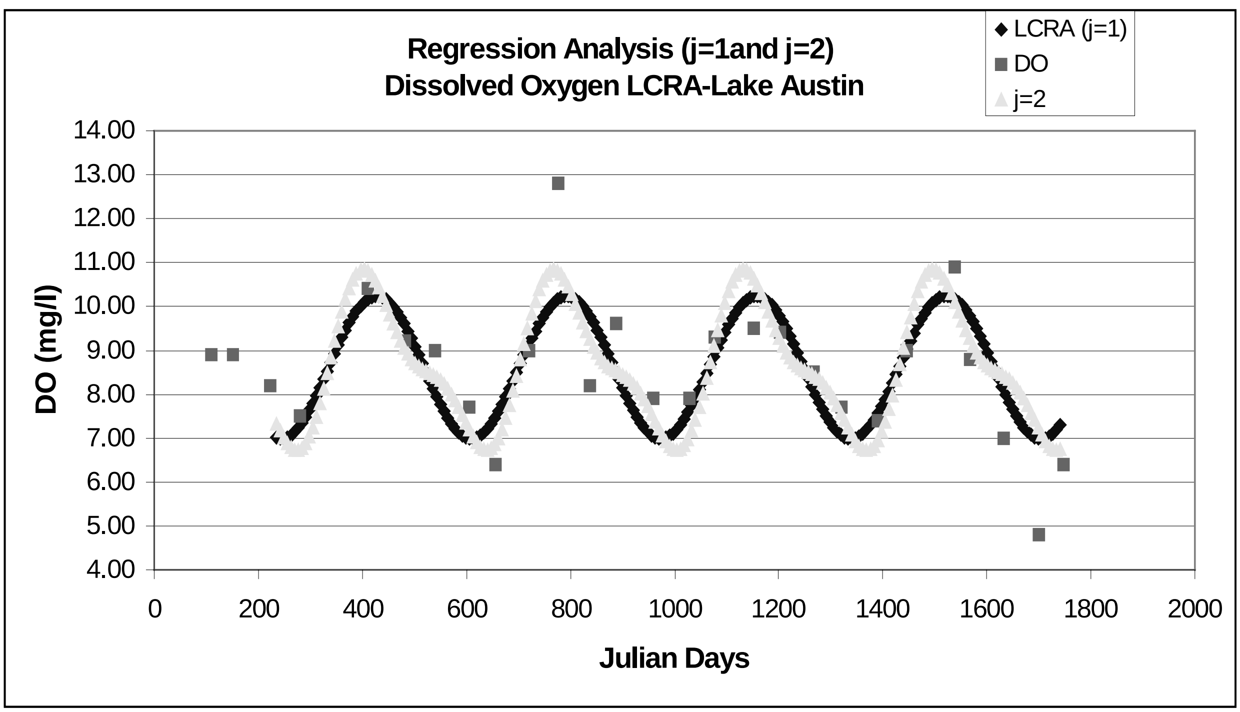


Figure 4-10 LCRA DO Fourier Plots, j=1 vs. j=2

Compare the two data sets with j=1 and j=2

The plot of the two best-fit lines (j=1) for the two data sets is shown in Figure 4-11.

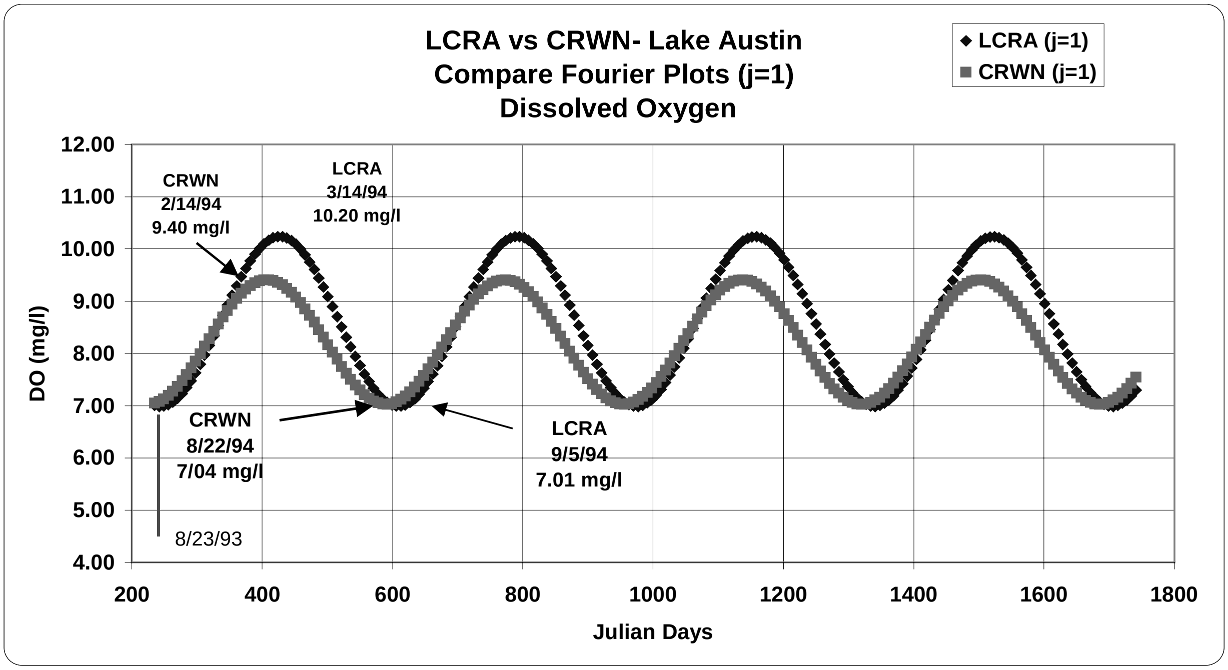


Figure 4-11 DO Fourier Plot, LCRA vs CRWN, j=1

The plot of the two best-fit lines (j=2) for the two data sets is shown in Figure 4-12.

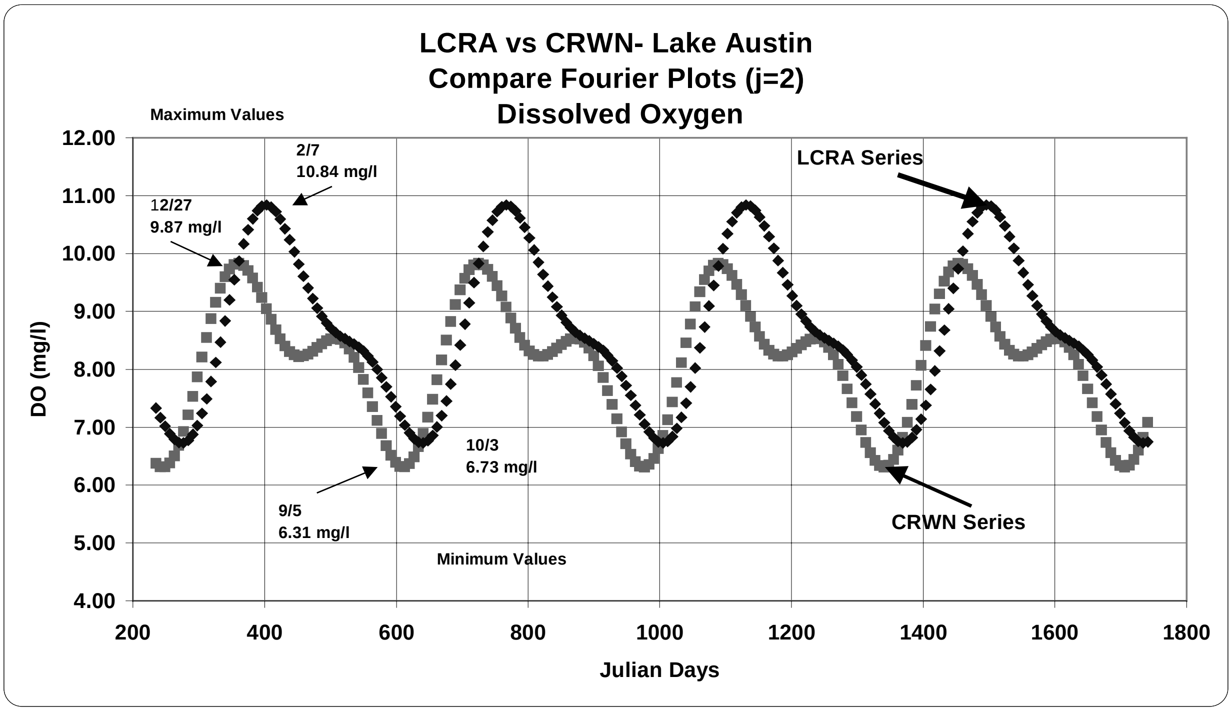


Figure 4-12 DO Fourier Plot, LCRA vs CRWN, j=2

It can be seen that each best-fit line has the same general form but there is a significant phase shift between the curves and the values of the predicted maximum and minimums. These differences are not as pronounced in the j=1 plots. The predicted time and magnitudes of the maximum and minimums are indicated on the graph.

For these particular data sets there is a significantly larger number of data points in the CRWN data set (n=90 vs. n=27). The t-test for the two data sets indicated that their means are not significantly different.

Table 4-4 shows the means of the two data sets determined by standard statistics and the fourier analysis.

Table 4-4 Summary of Means, DO, j=1 and j=2

Method	LCRA (n=27) Mean ± Standard Error	CRWN (n=90) Mean ± Standard Error
Standard Statistics	8.53 ± 0.30 mg/l	8.05 ± 0.17 mg/l
Fourier Analysis (Intercept)	(j=1) 8.70 ± 0.19 mg/l (j=2) 8.61 ± 0.20 mg/l	(j=1) 8.19 ± 0.14 mg/l (j=2) 8.21 ± 0.15 mg/l

The mean values for the two methods are in good agreement. The standard error range for both data sets overlap and with j=1 and j=2 in the fourier analysis.

4.4 Conclusion

‘There does not appear to be any advantage to expanding the regression more than j=1 for this particular type of data set. The best-fit line begins to respond to the daily variability at higher expansions and distorts the curve from seasonal variations. Also the mean values determined by the j=1 expansion are within the 95% error range of the j=2 expansion as shown above.

For this study the j=1 expansion will be used.

5 STUDY RESULTS

5.1 Lake Austin

Site Analysis- Dissolved Oxygen-- Sites LCRA 1 and CRWN 1

These sites are located on the downstream end of segment 1403 above Tom Miller Dam. The volunteer (CRWN) site is located about 0.4 miles upstream from the professional (LCRA) site. The volunteer site is a level 1 site monitored by an employee of the LCRA, who is not a professional monitor. It is worthwhile to note that the CRWN data is consistent throughout the year for the entire monitoring period being considered.

A time span of approximately four and a half years was used for the analysis:

CRWN data- 4/19/93 to 9/5/97, n=90

LCRA data- 4/19/93 to 10/13/97, n=27

Descriptive Statistics

Table 5-1 Descriptive Statistics, DO- Lake Austin

CRWN DATA- Lk. Austin		***LCRA DATA- Lk. Austin***	
Mean	8.05	Mean	8.53
Standard Error	0.17	Standard Error	0.30
Median	7.95	Median	8.80
Mode	6.70	Mode	9.00
Standard Deviation	1.63	Standard Deviation	1.55
Sample Variance	2.67	Sample Variance	2.40
Kurtosis	-0.39	Kurtosis	1.89
Skewness	0.02	Skewness	0.25
Range	7.00	Range	8.00
Minimum	4.60	Minimum	4.80
Maximum	11.60	Maximum	12.80
Sum	724.46	Sum	230.30
Count	90.00	Count	27.00
Confidence Level(95.0%)	0.34	Confidence Level(95.0%)	0.61

The mean and median values are very close for both sets of data. There are about three times more data values for the CRWN site. The skewness and kurtosis values are very small and both sets of data have about the same range. Essentially, the two data sets are very similar. The descriptive statistics (Table 5-1) indicate normal distribution of the data.

Combining the Data Sets

Table 5-2 is the descriptive statistics of the combined data sets.

Table 5-2 Descriptive Statistics Combined Data, DO- Lake Austin

Combined Data Sets	
Lake Austin	
Mean	8.16
Standard Error	0.15
Median	8.1
Mode	6.7
Standard Deviation	1.62
Sample Variance	2.62
Kurtosis	-0.01
Skewness	0.04
Range	8.2
Minimum	4.6
Maximum	12.8
Sum	955
Count	117

The standard error for the combined data sets is lower that either individual data sets which indicates an improvement in the estimation of the mean.

Quantile Plots

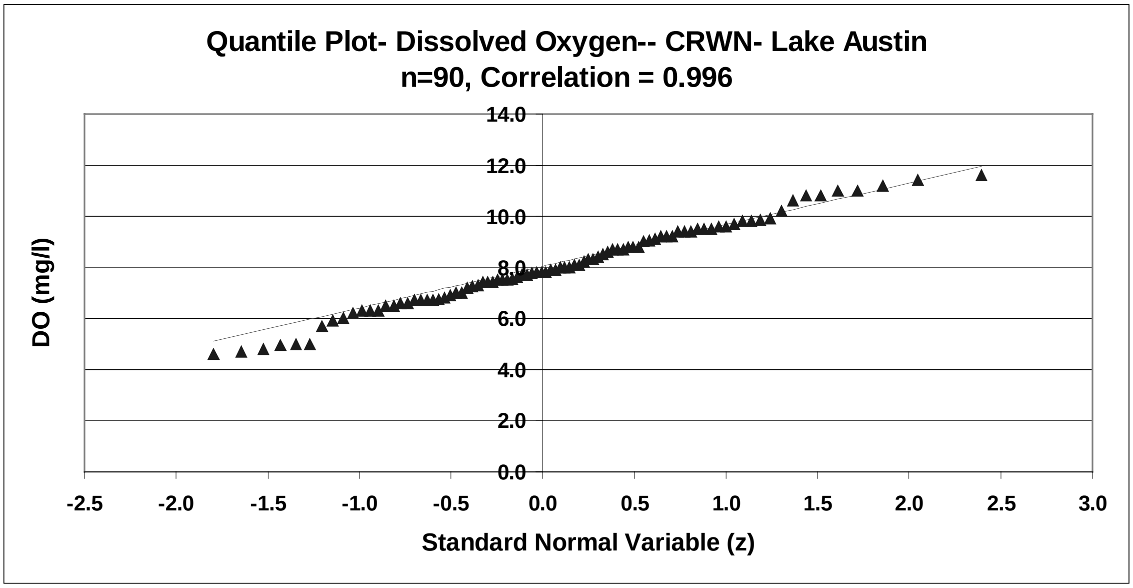


Figure 5-1 Quantile Plot, DO- CRWN1

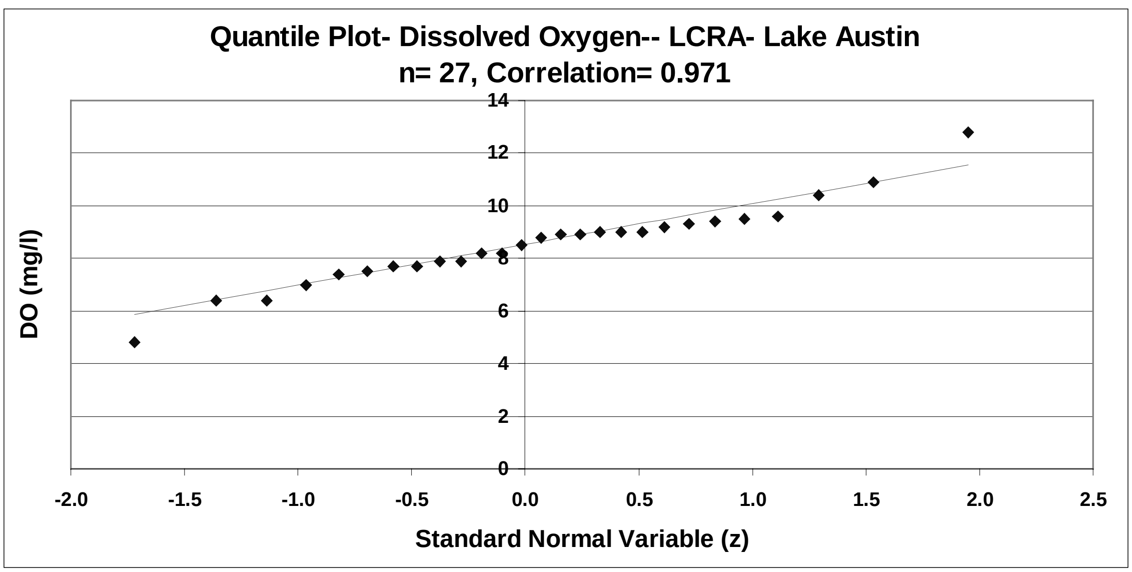
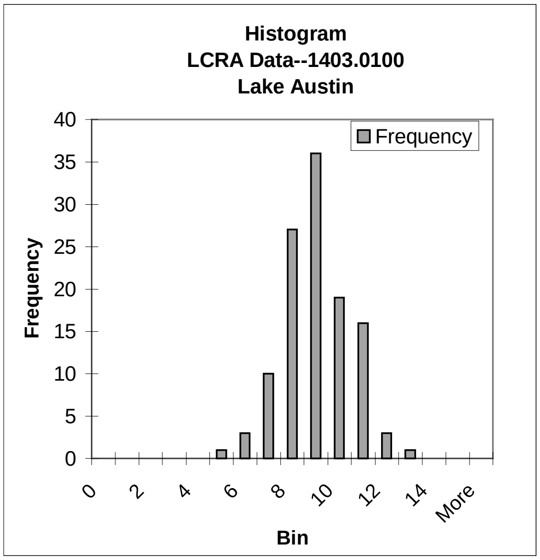


Figure 5-2 Quantile Plot, DO- LCRA1

The quantile plots (Figures 5-1 and 5-2) show that both data sets are normally distributed. Both correlation coefficients are greater than the lower critical value of the probability plot correlation test statistic for the normal distribution.

Histograms



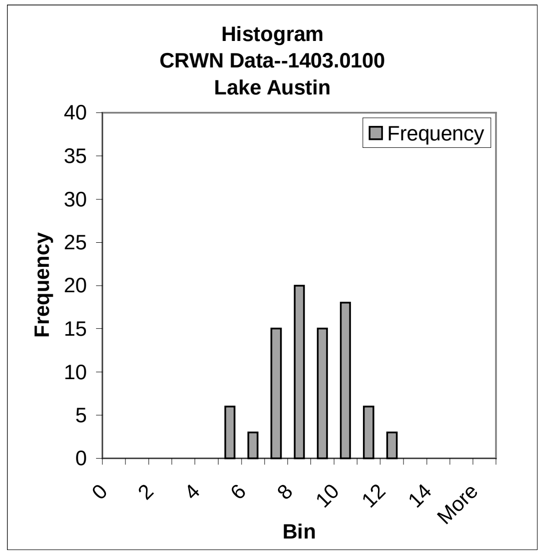


Figure 5-3 Histograms, DO- Lake Austin

The histograms in Figure 5-3 show a normal distribution also.

t-test

Because the data is normally distributed the t-test is appropriate to compare the means of the two data sets.

Table 5-3 t-test Results, DO- Lake Austin

	Calculated t= -1.395	
t-Test: Two-Sample Assuming Unequal Variances		
	CRWN	***LCRA***
Mean	8.050	8.53
Variance	2.666	2.40
Observations	90	27
df	45	
t Stat	-1.395	
P(T<=t) one-tail	0.085	
t Critical one-tail	1.679	
P(T<=t) two-tail	0.170	
t Critical two-tail	2.014	

The t-value, shown in Table 5-3, of -1.395 indicates that statistically the means data sets are not significantly different.

<u>Standard Error of the Mean</u>

Table 5-4 shows that the LCRA data has a much lower standard error than the CRWN data even though there are fewer measurements. The range for the volunteer data around the mean for n=90 is 0.69 or the mean DO with 95% confidence is 8.05 ± 0.35 mg/l. The range for the professional data around the mean for n=27 is 0.11 or the mean DO with 95% confidence is 8.53 ± 0.055 mg/l.

Table 5-4 Standard Error of the Mean, DO- Lake Austin

Standard Error of the Mean CRWN Data- Lake Austin						
index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	4.80	4.80	0.00	4.80	4.80	0.00
2	6.20	5.50	0.70	4.10	6.90	2.80
3	7.80	6.27	0.87	4.53	8.00	3.47
4	7.80	6.65	0.72	5.20	8.10	2.89
5	8.30	6.98	0.65	5.68	8.28	2.60
6	6.70	6.93	0.53	5.87	8.00	2.13
..	..	..	..	..	..	..
87	7.40	8.09	0.18	7.74	8.45	0.70
88	7.70	8.09	0.17	7.74	8.44	0.69
89	6.30	8.07	0.17	7.72	8.42	0.69
90	6.30	8.05	0.17	7.71	8.39	0.69

Standard Error of the Mean LCRA Data- Lake Austin						
index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	8.90	8.90	0.00	8.90	8.90	0.00
2	8.90	8.90	0.00	8.90	8.90	0.00
3	8.20	8.67	0.08	8.51	8.82	0.31
4	7.50	8.38	0.12	8.13	8.62	0.50
5	10.40	8.78	0.10	8.58	8.98	0.39
6	9.20	8.85	0.08	8.69	9.01	0.33
..	..	..	..	..	..	..
24	8.80	8.84	0.03	8.78	8.90	0.12
25	7.00	8.76	0.03	8.71	8.82	0.11
26	4.80	8.61	0.03	8.56	8.67	0.11
27	6.40	8.53	0.03	8.47	8.59	0.11

Figures 5-4 and 5-5 are the graphical representation of the standard error of the mean.

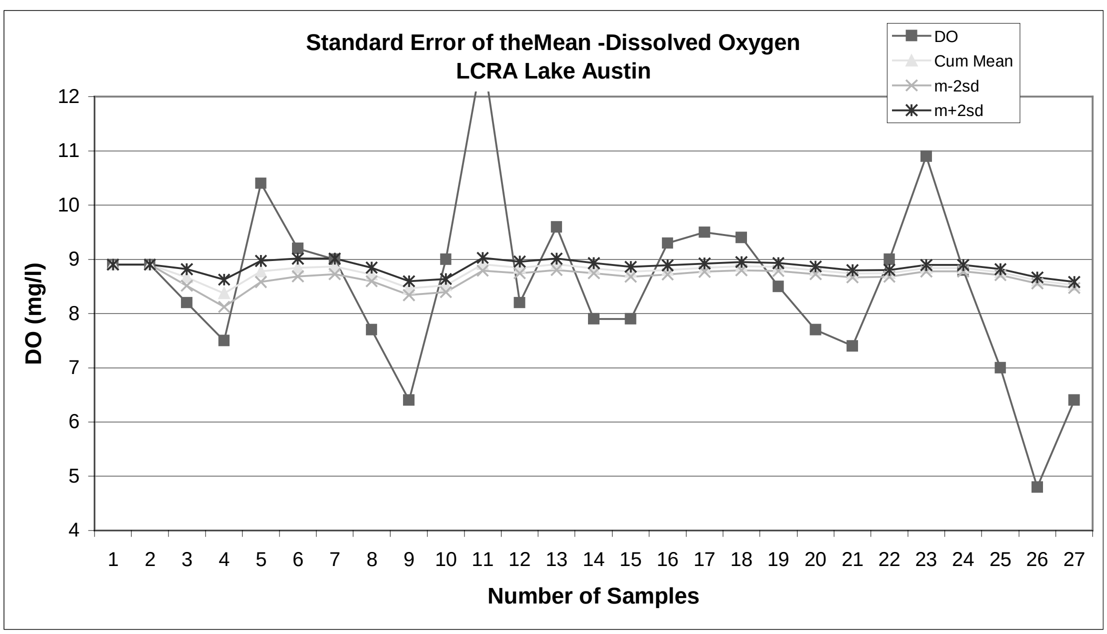


Figure 5-4 Standard Error of the Mean, DO- LCRA1

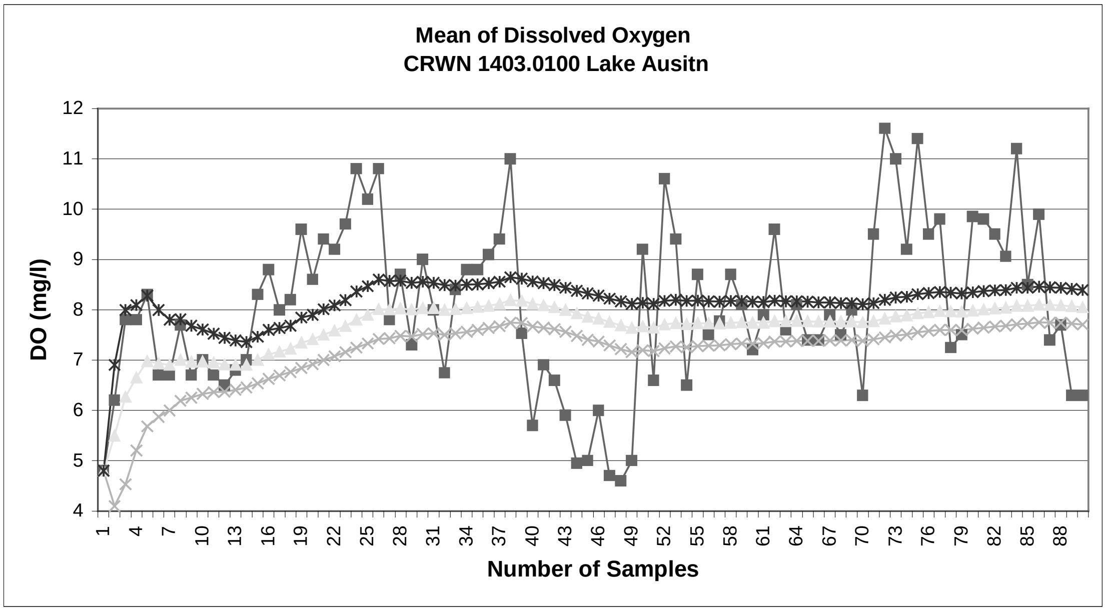


Figure 5-5 Standard Error of the Mean, DO- CRWN1

Fourier Analysis--Dissolved Oxygen

The statistics for the regression analysis of the two data sets are shown in Tables 5-5 and 5-6.

Table 5-5 Regression Statistics- LCRA1- Lake Austin

LCRA- Lake Austin SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.763
R Square	0.583
Adjusted R Square	0.548
Standard Error	1.041
Observations	27

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	36.305	18.152	16.761	0.000
Residual	24	25.992	1.083		
Total	26	62.296			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.609	0.203	42.343	0.000	8.189	9.029
X Variable 1	0.784	0.308	2.548	0.018	0.149	1.419
X Variable 2	1.423	0.267	5.322	0.000	0.871	1.975

Table 5-6 Regression Statistics- CRWN1- Lake Austin

CRWN-Lake Austin-SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.505
R Square	0.255
Adjusted R Square	0.238
Standard Error	1.426
Observations	90

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	60.476	30.238	14.878	0.000
Residual	87	176.820	2.032		
Total	89	237.296			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.2163	0.1536	53.4992	0.0000	7.9110	8.5215
X Variable 1	0.8892	0.2126	4.1814	0.0001	0.4665	1.3118
X Variable 2	0.7930	0.2174	3.6482	0.0004	0.3610	1.2251

The plot of the best fit lines for the two data sets is shown in Figure 5-6.

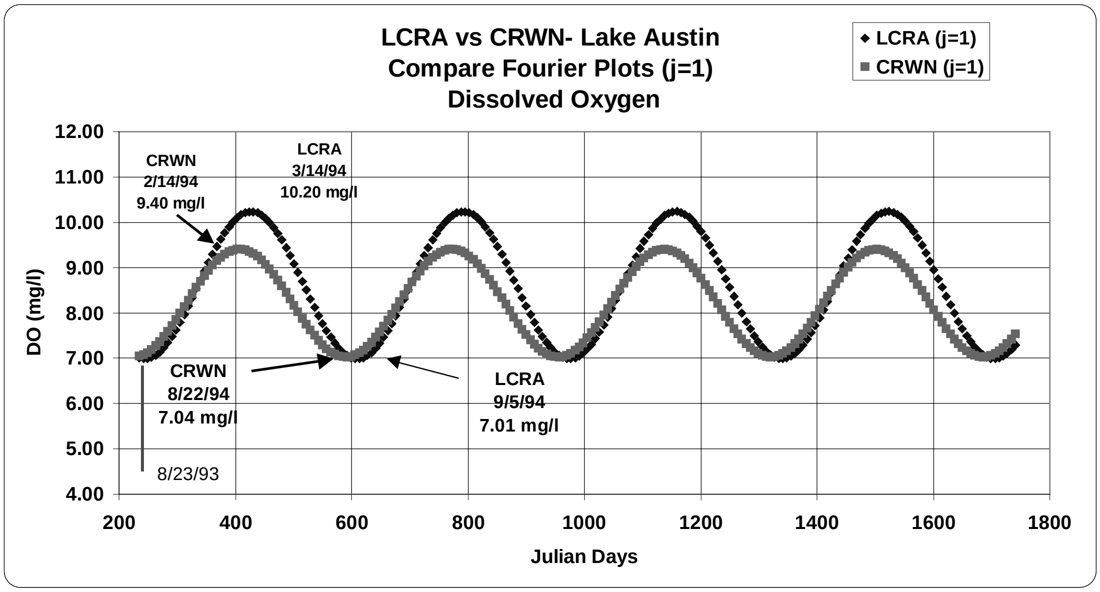


Figure 5-6 Fourier Plot- Lake Austin

Table 5-7 shows the comparison of the means of the two data sets as determined by the two types of analysis:

Table 5-7 Summary of Means- Lake Austin

Method	LCRA (n=27) Mean ± Standard Error	CRWN (n=90) Mean ± Standard Error
Standard Statistics	8.53 ± 0.30 mg/l	8.05 ± 0.17 mg/l
Fourier Analysis (Intercept)	8.61 ± 0.20 mg/l	8.22 ± 0.15 mg/l

There is good agreement between the two methods as illustrated above. The mean values determined by standard statistics and fourier analysis are within the margins of error. The CRWN data is well distributed temporally with at least one measurement per month throughout the time period analyzed.

Total Dissolved Solids

The same analysis technique will now be used for Total Dissolved Solids. A plot of the raw data is shown in Figure 5-7.

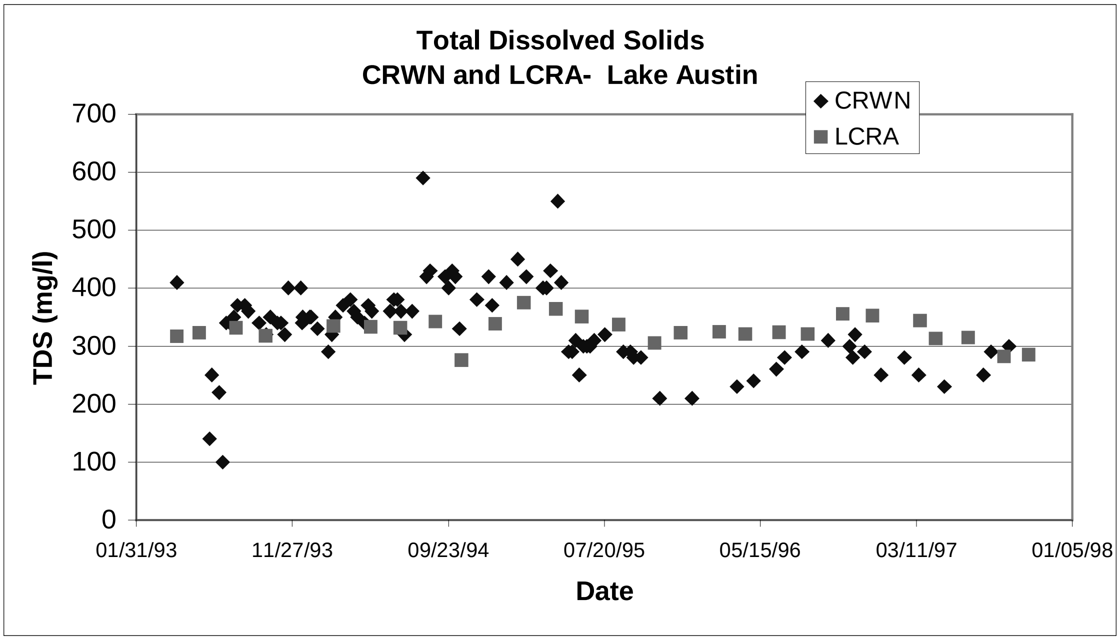


Figure 5-7 Scatter Plot, TDS- Lake Austin

This plot shows a much wider range and variability in the volunteer data.

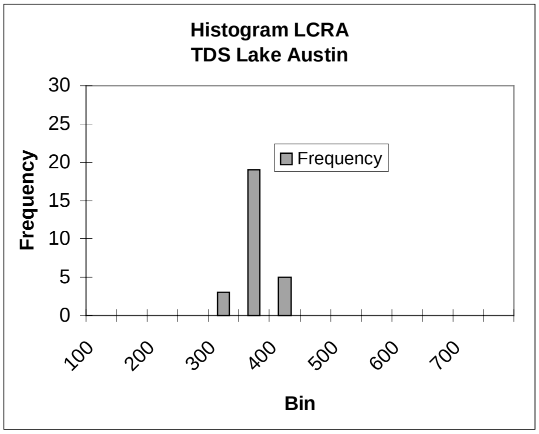
Descriptive Statistics

Table 5-8 Descriptive Statistics, TDS- Lake Austin

TDS- CRWN		***TDS- LCRA***	
Mean	333.96	Mean	327.07
Standard Error	7.79	Standard Error	4.48
Median	340	Median	324.6
Mode	290	Mode	322.8
Standard Deviation	74.28	Standard Deviation	23.26
Sample Variance	5517.51	Sample Variance	540.84
Kurtosis	2.02	Kurtosis	0.43
Skewness	0.12	Skewness	-0.35
Range	490	Range	99
Minimum	100	Minimum	275.4
Maximum	590	Maximum	374.4
Sum	30390	Sum	8831
Count	91	Count	27
Confidence Level(95.0%)	15.47	Confidence Level(95.0%)	9.20

The descriptive statistics shown in Figure 5-7, indicate that both data sets show the signs of normal distribution. The means and medians of the two data sets are relatively close together and they both have small skewness values. The means of the two data sets are very close also, 334mg/l for CRWN versus 327 mg/l for LCRA. In other aspects they are quite different. The higher degree of variability of the volunteer data is shown. The standard deviation is three times larger than the professional data and the range is almost five times as large. The confidence levels for each data sets are very reasonable for this water quality parameter.

Histograms



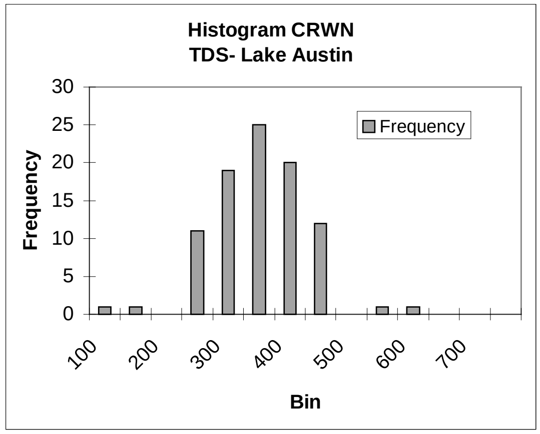


Figure 5-8 Histograms, TDS- Lake Austin

The histograms (Figure 5-8) show a normal distribution for each data set.

t-test

The result of the t-test are shown in Table 5-9.

Table 5-9 Result of the t-test- TDS- Lake Austin

t-Test: Two-Sample Assuming Unequal Variances

	CRWN	*LCRA*
Mean	333.96	327.07
Variance	5517.51	540.84
Observatio	91	27
Hypothesiz	0	
df	116	
t Stat	0.767	
P(T<=t) one	0.222	
t Critical on	1.658	
P(T<=t) two	0.445	
t Critical tw	1.981	

The t-test indicate that there is no significat difference in the means of the two data sets.

Standard Error of the Mean

The standard error of the mean plots for the two data sets are shown in Figures 5-9 and 5-10.

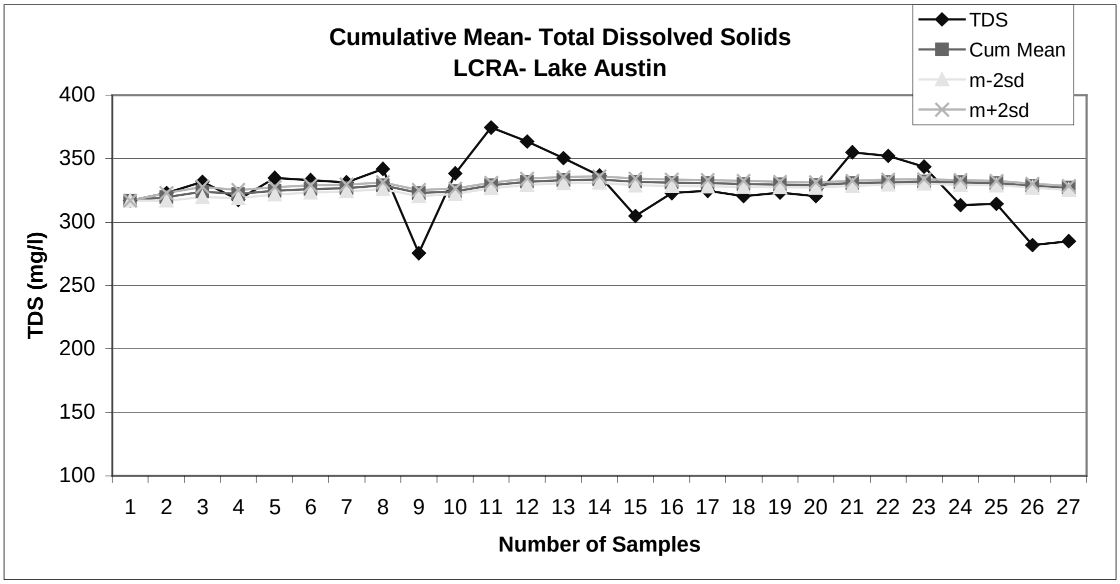


Figure 5-9 Standard Error of the Mean- TDS- LCRA1

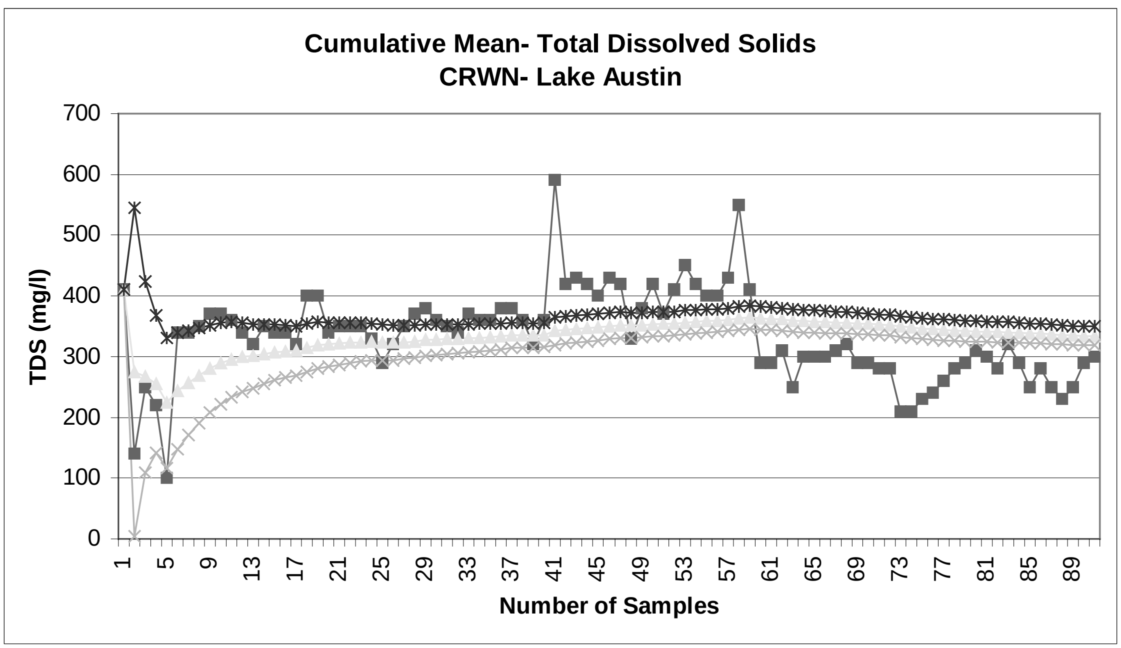


Figure 5-10 Standard Error of the Mean- TDS- CRWN1

The plots show that each data set is stable and have a small error interval.

5.2 Bastrop

<u>Site Analysis- Dissolved Oxygen-- Sites LCRA 2 and CRWN 2</u>

This site is located in the upstream part of river segment 1434. The LCRA monitors at this site six times a year and the CRWN has a volunteer site at the same location. The volunteer data lacks summer month data. Only 21% of the measurements occurred between April1 and October 1.

For the following analysis a time span of approximately four years is used:

LCRA data- 8/23/93 to 10/1/97, n=26

CRWN data- 9/28/93 to 12/16/97, n=104

<u>Descriptive Statistics</u>

The following tables in Figure 5-10 are the descriptive statistics for each site:

Table 5-10 Descriptive Statistics, DO- Bastrop

CRWN Data- Bastrop		*LCRA DATA- Bastrop*	
Mean	9.35	**Mean**	8.96
Standard Error	0.19	**Standard Error**	0.34
Median	9.45	**Median**	8.65
Mode	10.00	**Mode**	7.10
Standard Deviation	1.94	**Standard Deviation**	1.72
Sample Variance	3.78	**Sample Variance**	2.96
Kurtosis	-0.09	**Kurtosis**	0.11
Skewness	0.09	**Skewness**	0.90
Range	10.70	**Range**	6.30
Minimum	4.10	**Minimum**	7.00
Maximum	14.80	**Maximum**	13.30
Sum	972.90	**Sum**	232.90
Count	104.00	**Count**	26.00

Note that there are almost four times more CRWN data values than LCRA data values. Both of the data sets have mean and median values that are very close to each other and relatively small skewness values. The CRWN skewness value is more than 50% lower than the LCRA values, probably because of the larger number of data points. The range of the CRWN is substantially higher than the LCRA range, this being due to the larger number of CRWN measurements. The descriptive statistics indicate that both sets of data are normally distributed.

Combining the Data Sets

Table 5-11 is the descriptive statistics of the combined data sets.

Table 5-11 Descriptive Statistics Combined Data- Bastrop

Combined Data	
Bastrop	
Mean	9.28
Standard Error	0.17
Median	9.2
Mode	10
Standard Deviation	1.90
Sample Variance	3.62
Kurtosis	-0.14
Skewness	0.23
Range	10.70
Minimum	4.1
Maximum	14.8
Sum	1205.8
Count	130

The standard error for the combined data sets is lower than either individual data sets which indicates an improvement in the estimation of the mean.

Quantile Plot

Figures 5-11 and 5-12 are the cumulative frequency plots and histograms for the two data sets:

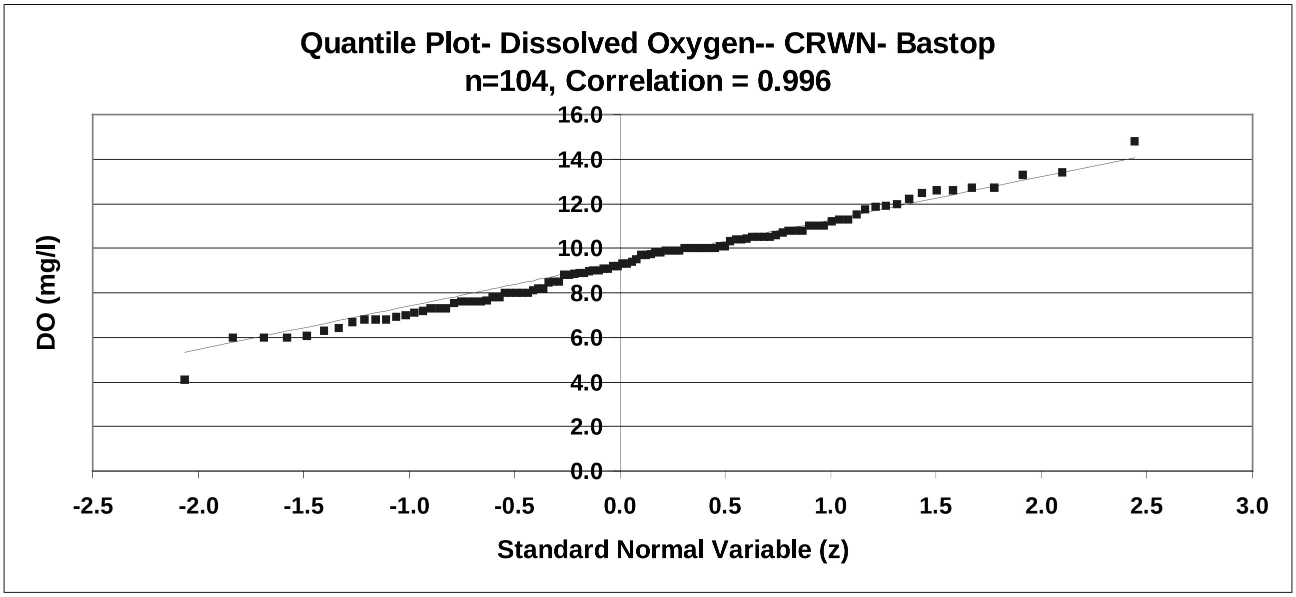


Figure 5-11 Quantile Plot, DO- CRWN2

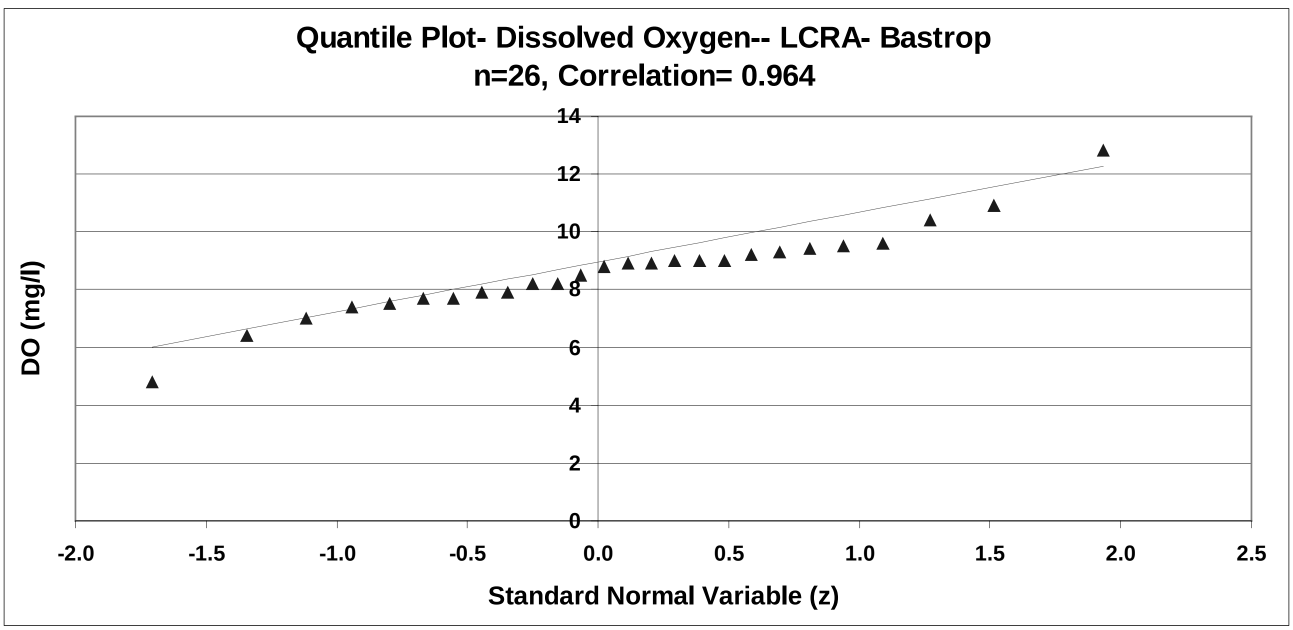
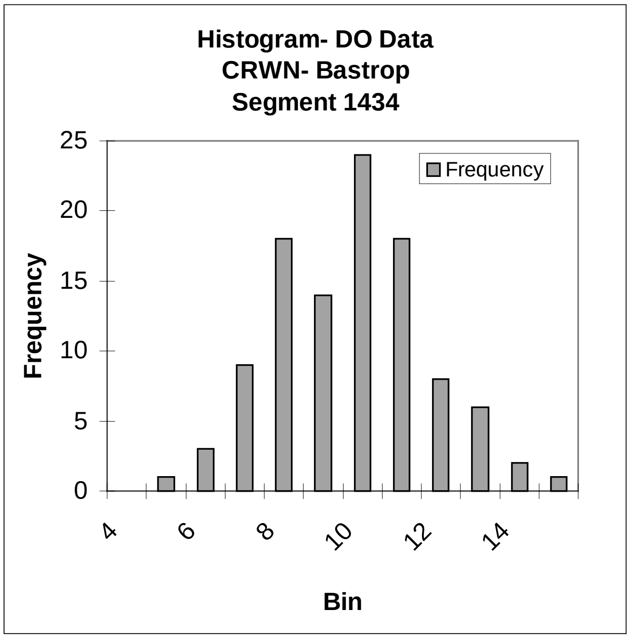


Figure 5-12 Quantile Plot, DO- LCRA2

The quantile plots show that both data sets are normally distributed. Both correlation coefficients are greater than the lower critical value of the probability plot correlation test statistic for the normal distribution.

Histograms



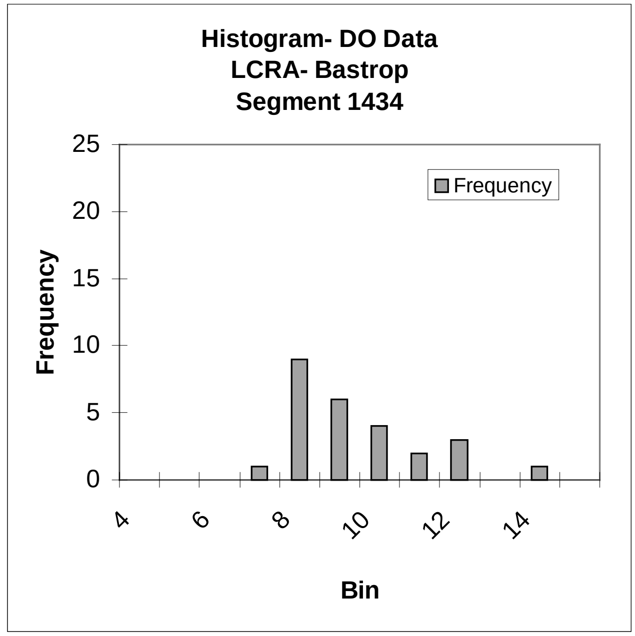


Figure 5-13 Histograms, DO- Bastrop

The histogram for the CRWN data has the appearance of a normal distribution, slightly skewed in the positive direction. The LCRA data appears to be slightly more skewed but there is only a difference of three data point between the intervals with the highest number of data points.

t-test

The t-test should work well with these data sets since they are normally distributed and the standard deviations are approximately the same. The t-test was done two ways for this analysis: using Excel's built in test and using the above formula in the spreadsheet. The results are in Table 5-12, below.

Table 5-12 t-test Results, DO- Bastrop

Calculated t=	1.108	
t-Test: Two-Sample Assuming Unequal Variances		
	CRWN	**LCRA**
Mean	9.401	8.958
Variance	4.371	2.959
Observations	95.000	26
Hypothesized Mean Difference	0.000	
df	47.000	
t Stat	1.108	
P(T<=t) one-tail	0.137	
t Critical one-tail	1.678	
P(T<=t) two-tail	0.274	
t Critical two-tail	2.012	

The results using Excel and the formula were the same. The value of t = 1.108 indicates that the hypothesis there is not a significant difference between the two means cannot be rejected.

Standard Error of the Mean

Table 5-13 Standard Error of the Mean, DO-Bastrop

Standard Error of the Mean CRWN Data- Bastrop

index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	7.00	7.00	0	7.000	7.000	0.000
2	9.00	8.00	0.500	7.000	9.000	2.000
3	8.00	8.00	0.333	7.333	8.667	1.333
4	7.60	7.90	0.243	7.414	8.386	0.971
..	..	..	..	..	..	..
100	10.8	9.29	0.045	9.201	9.381	0.179
101	11.75	9.32	0.044	9.227	9.404	0.178
102	10.45	9.33	0.044	9.239	9.415	0.176
103	9.8	9.33	0.044	9.244	9.418	0.174
104	10.3	9.34	0.043	9.255	9.427	0.172
105	10.8	9.35	0.043	9.269	9.440	0.171

Standard Error of the Mean LCRA Data- Bastrop

index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	7.10	7.10	0	7.100	7.100	0.000
2	7.20	7.15	0.025	7.100	7.200	0.100
3	10.70	8.33	0.403	7.527	9.139	1.612
4	11.70	9.18	0.501	8.173	10.177	2.003
..	..	..	..	..	..	..
21	9.80	9.01	0.150	8.715	9.314	0.599
22	10.20	9.07	0.143	8.783	9.354	0.571
23	9.10	9.07	0.136	8.797	9.343	0.546
24	8.40	9.04	0.131	8.780	9.303	0.523
25	7.60	8.98	0.125	8.733	9.235	0.501
26	8.30	8.96	0.120	8.717	9.199	0.482

The results in Figure 5-26, above, indicate that in 95 out of 100 similar measurements the mean would lie in the approximate range around the mean indicated in the last column. The range for the volunteer data around the mean for n=105 is 0.171 or the mean DO with 95% confidence is9.35 ± 0.085 mg/l. The range for the professional data around the mean for n=26 is 0.482 or the mean DO level with 95% confidence is 8.96 ± 0.241 mg/l.

The graphical representation of the above analysis is shown in Figures 5-14 and 5-15:

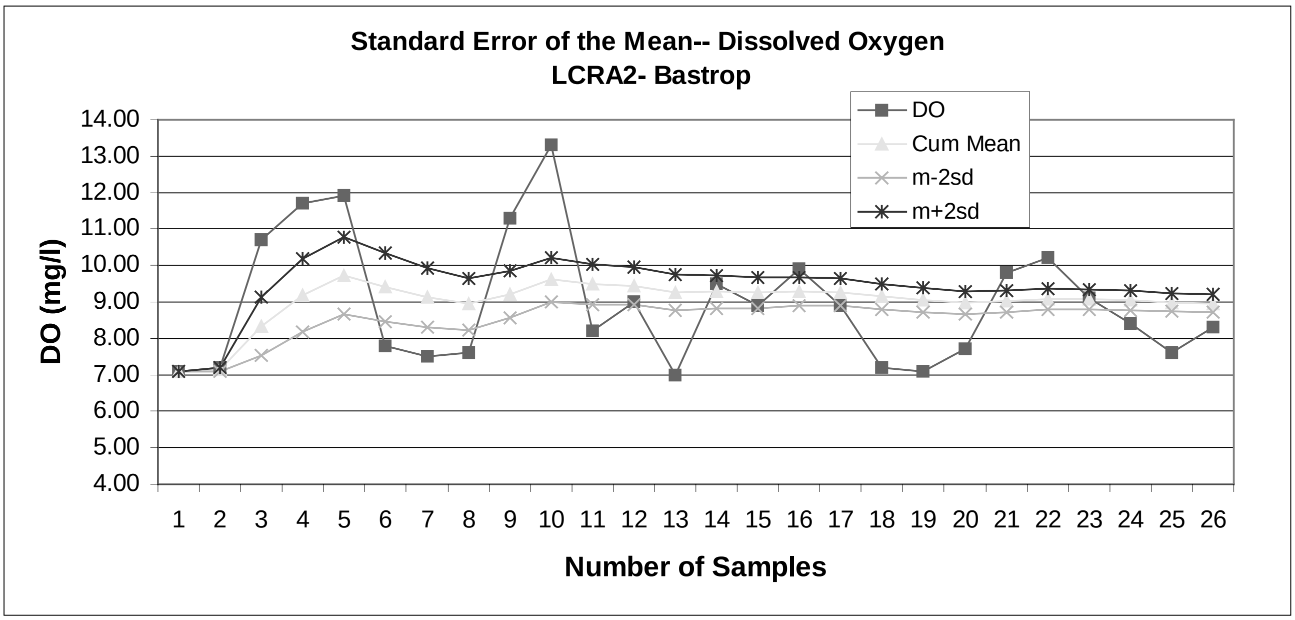


Figure 5-14 Standard Error of the Mean, DO- LCRA2

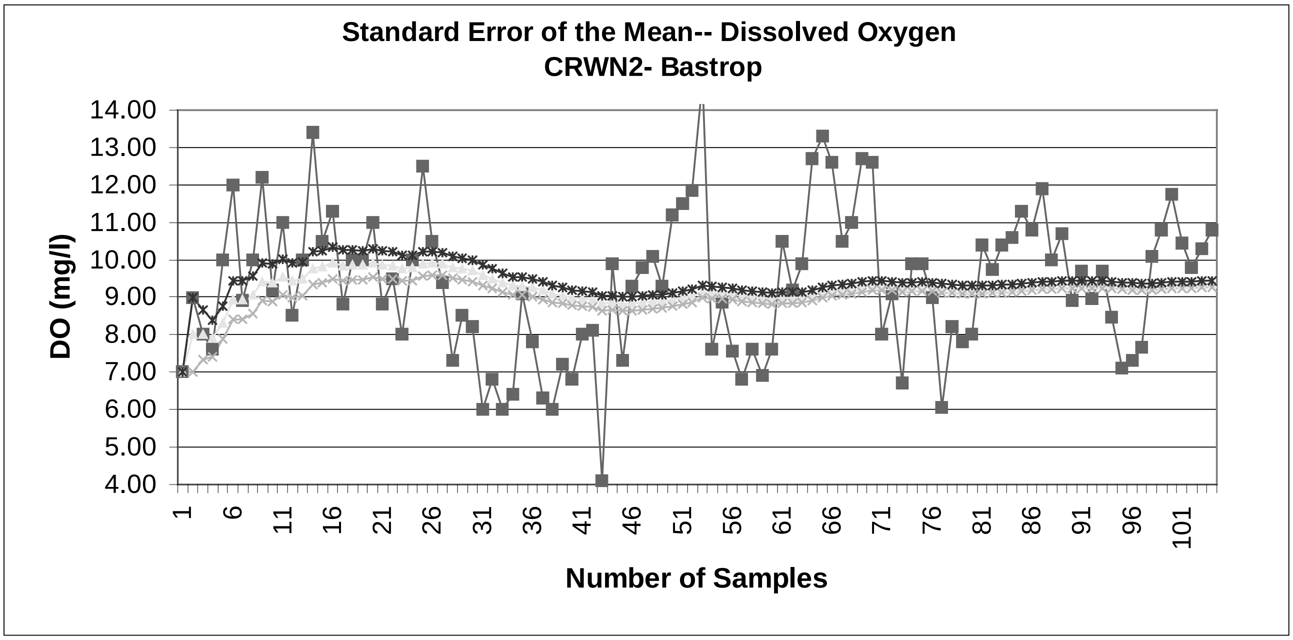


Figure 5-15 Standard Error of the Mean, DO- CRWN2

Fourier Analysis-- Dissolved Oxygen

A scatter plot of the data shown in Figure 5-16 illustrates the seasonal variations in the dissolved oxygen levels:

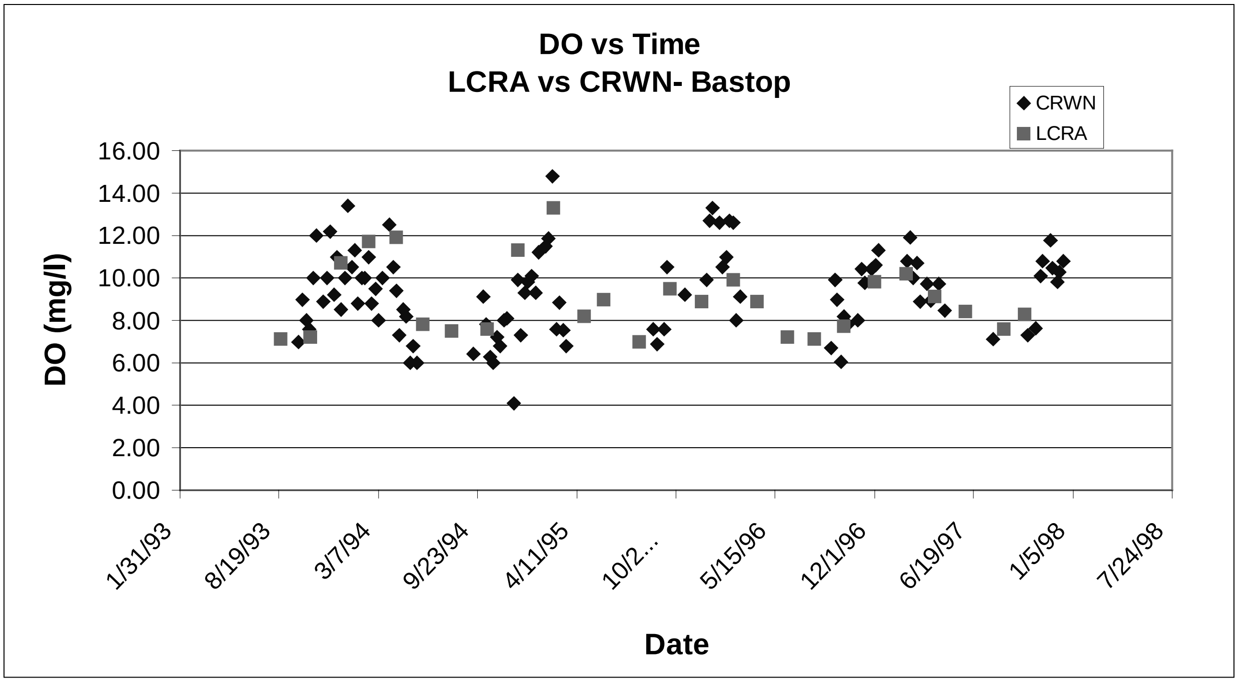


Figure 5-16 Scatter Plot, TDS- Bastrop

The statistics for the regression analysis of the two data sets are shown in Tables 5-14 and 5-15.

Table 5-14 Regression Statistics, DO-LCRA2

LCRA- Bastrop- SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.82
R Square	0.66
Adjusted R Square	0.64
Standard Error	1.04
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	49.18	24.59	22.81	0.00
Residual	23	24.80	1.08		
Total	25	73.98			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	9.04	0.20	44.30	0.00	8.62	9.46
X Variable 1	1.58	0.30	5.27	0.00	0.96	2.19
X Variable 2	1.16	0.28	4.16	0.00	0.58	1.73

Table 5-15 Regression Statistics, DO-CRWN2

CRWN- Bastrop-SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.57
R Square	0.33
Adjusted R Square	0.32
Standard Error	1.61
Observations	104

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	128.27	64.13	24.83	0.00
Residual	101	260.90	2.58		
Total	103	389.17			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.47	0.21	39.46	0.00	8.05	8.90
X Variable 1	2.04	0.32	6.30	0.00	1.40	2.69
X Variable 2	0.64	0.21	3.03	0.00	0.22	1.05

The plot of the best fit lines for the two data sets is shown in Figure 5-17.

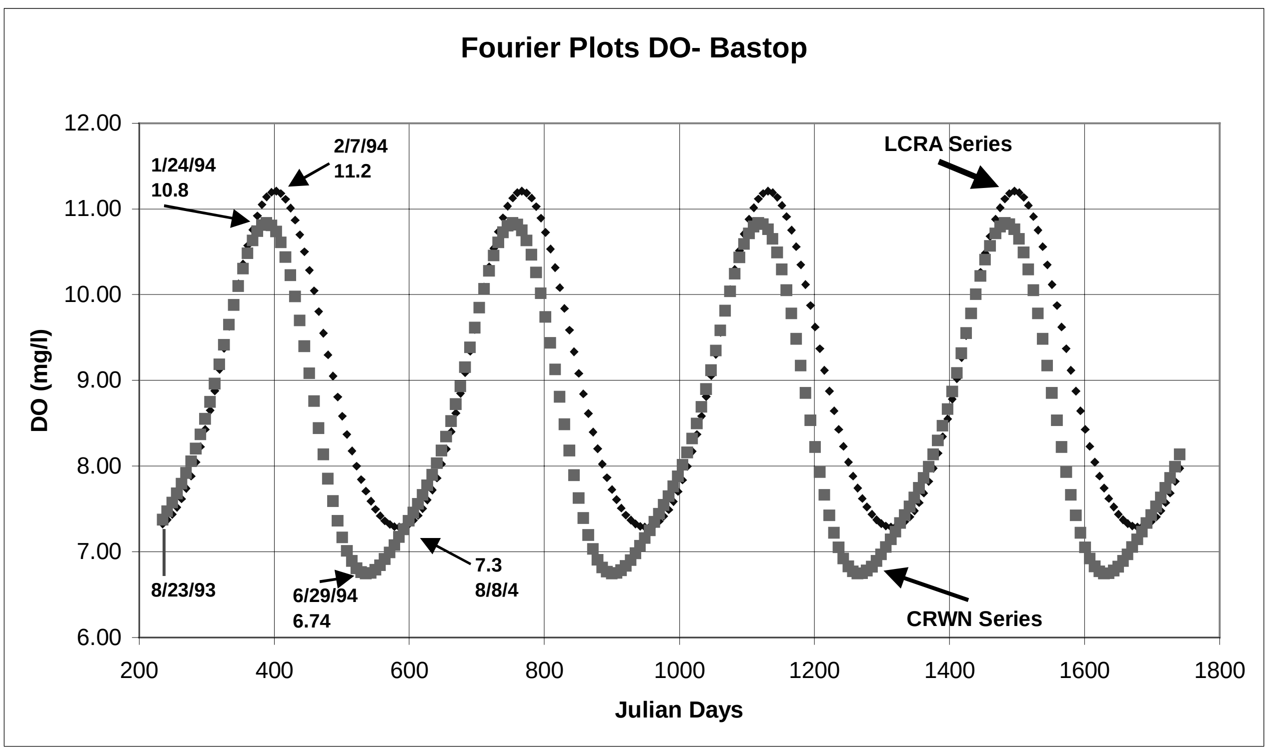


Figure 5-17 Fourier Plot, DO- Bastrop

The fourier plots show a phase shift in the cycles of the two data sets of about 14 days for the maximums and 40 days for the minimums. It should be kept in mind that the CRWN data set lacked data points for the summer months, which may account for the phase shift being larger in the summer period when the minimums occur.

Comparing the mean values determined by standard statistics and fourier analysis is shown in Table 5-16.

Table 5-16 Summary of Means, DO- Bastrop

Method	LCRA (n=27) Mean ± Standard Error	CRWN (n=90) Mean ± Standard Error
Standard Statistics	8.96 ± 0.34 mg/l	9.35 ± 0.19 mg/l
Fourier Analysis (Intercept)	9.04 ± 0.21 mg/l	8.55 ± 0.29 mg/l

The means of the LCRA data sets as determined by the two methods are in good agreement, differing by only 0.08 mg/l. The means CRWN data are considerably different, though, differing by 0.80 mg/l. The mean of the CRWN data set calculated by the fourier method is the lower value and is more realistic. The CRWN data set lacks summer data. The t-stat for the two data sets is 1.11. So there is not a significant difference in the means of the data sets.

Total Dissolved Solids Analysis-- Bastrop

The scatter plot of TDS data is shown in Figure 5-18.

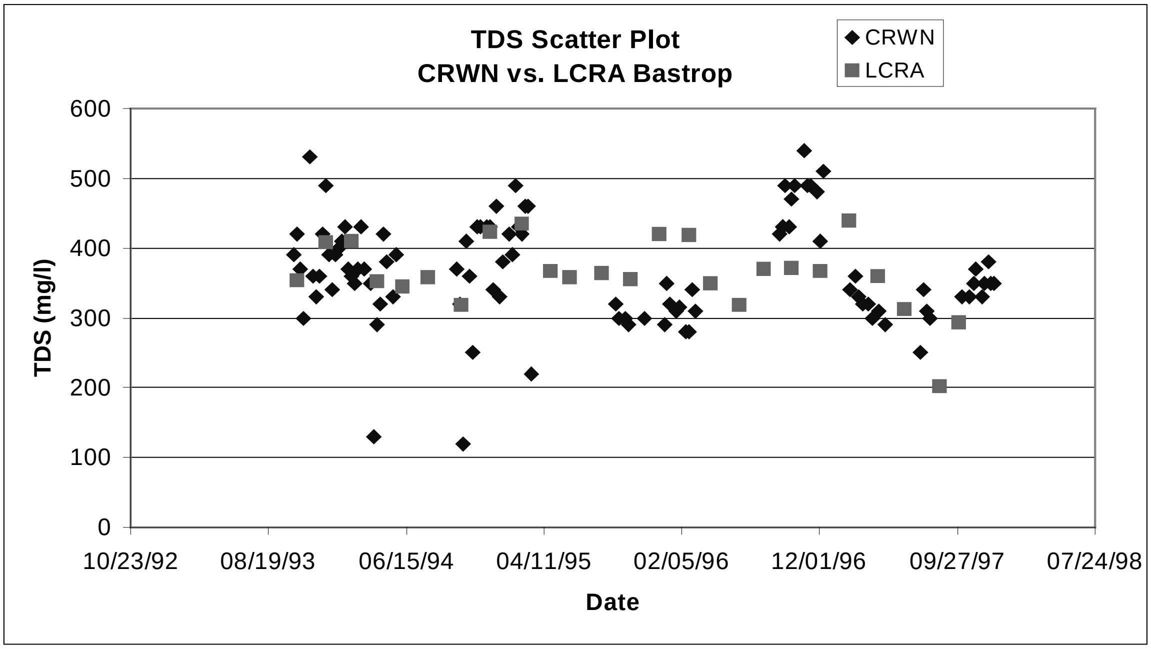


Figure 5-18 Scatter Plot, TDS- Bastrop

Descriptive Statistics:

Table 5-17 Descriptive Statistics, TDS- Bastrop

CRWN- Bastrop		*LCRA- Bastop*	
TDS		**TDS**	
Mean	368.34	**Mean**	362.74
Standard Error	7.52	**Standard Error**	10.25
Median	360	**Median**	360
Mode	430	**Mode**	358.2
Standard Deviation	75.17	**Standard Deviation**	51.25
Sample Variance	5650.61	**Sample Variance**	2626.35
Kurtosis	1.06	**Kurtosis**	2.81
Skewness	-0.28	**Skewness**	-1.08
Range	420	**Range**	236.4
Minimum	120	**Minimum**	202.2
Maximum	540	**Maximum**	438.6
Sum	36834	**Sum**	9068.4
Count	100	**Count**	25

The descriptive statistics, shown in Table 5-17, indicate that the two data sets are very similar. The median values are exactly the same and the means are close to equal. The main difference is the greater range and variability of the CRWN data.

Histograms

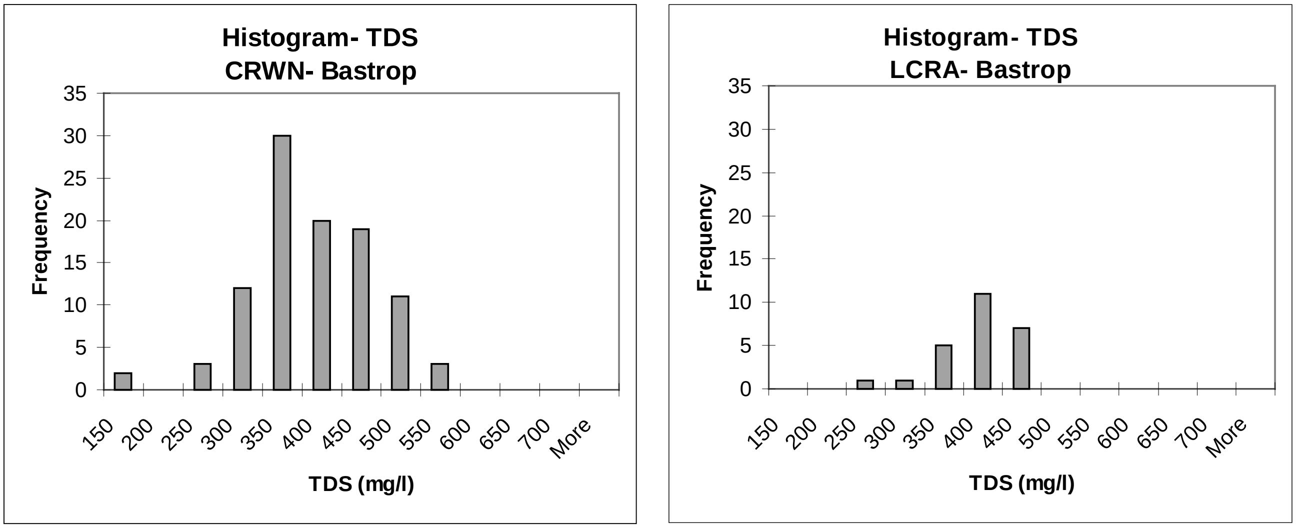


Figure 5-19 Histograms, TDS- Bastrop

t-test

Table 5-18 t-test Results, TDS- Bastrop

t-Test: Two-Sample Assuming Unequal Variances

	CRWN	*LCRA*
Mean	368.34	362.74
Variance	5650.61	2626.35
Observations	100	25
Hypothesized Mean Difference	0	
df	53	
t Stat	0.44	
P(T<=t) one-tail	0.33	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.66	
t Critical two-tail	2.01	

Standard Error of the Mean Plots

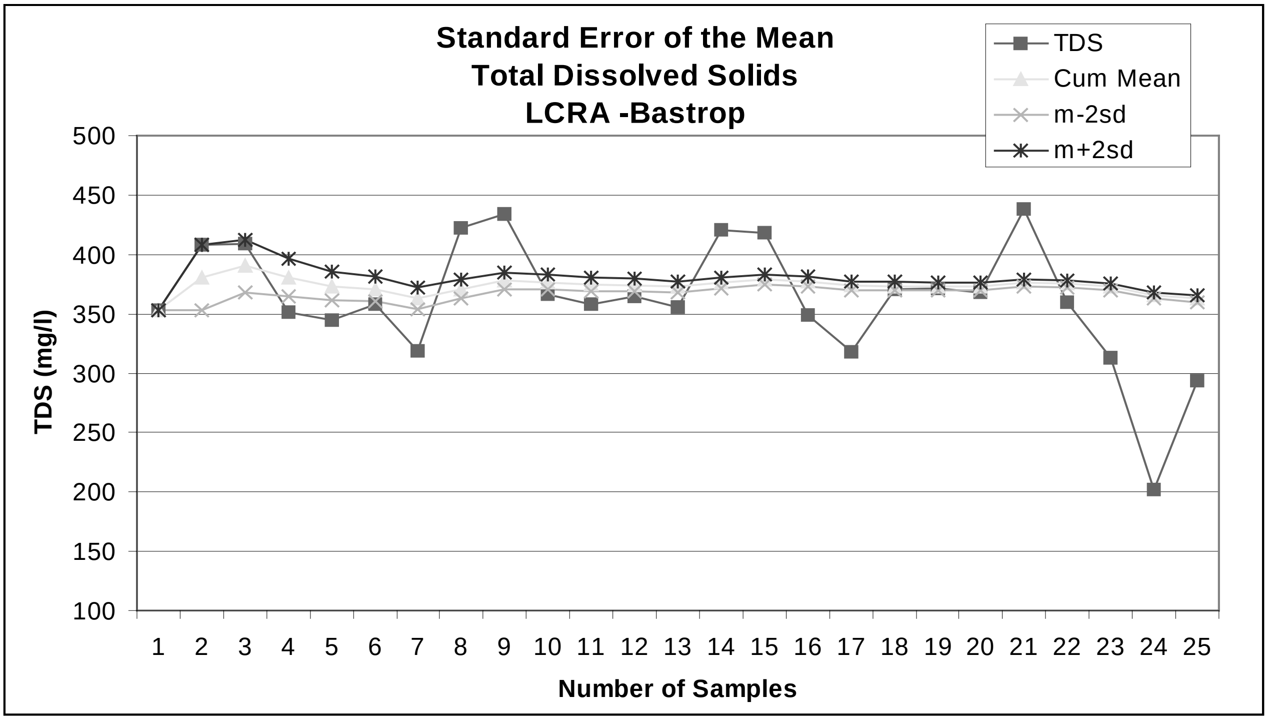


Figure 5-20 Standard Error of the Mean, TDS- LCRA2

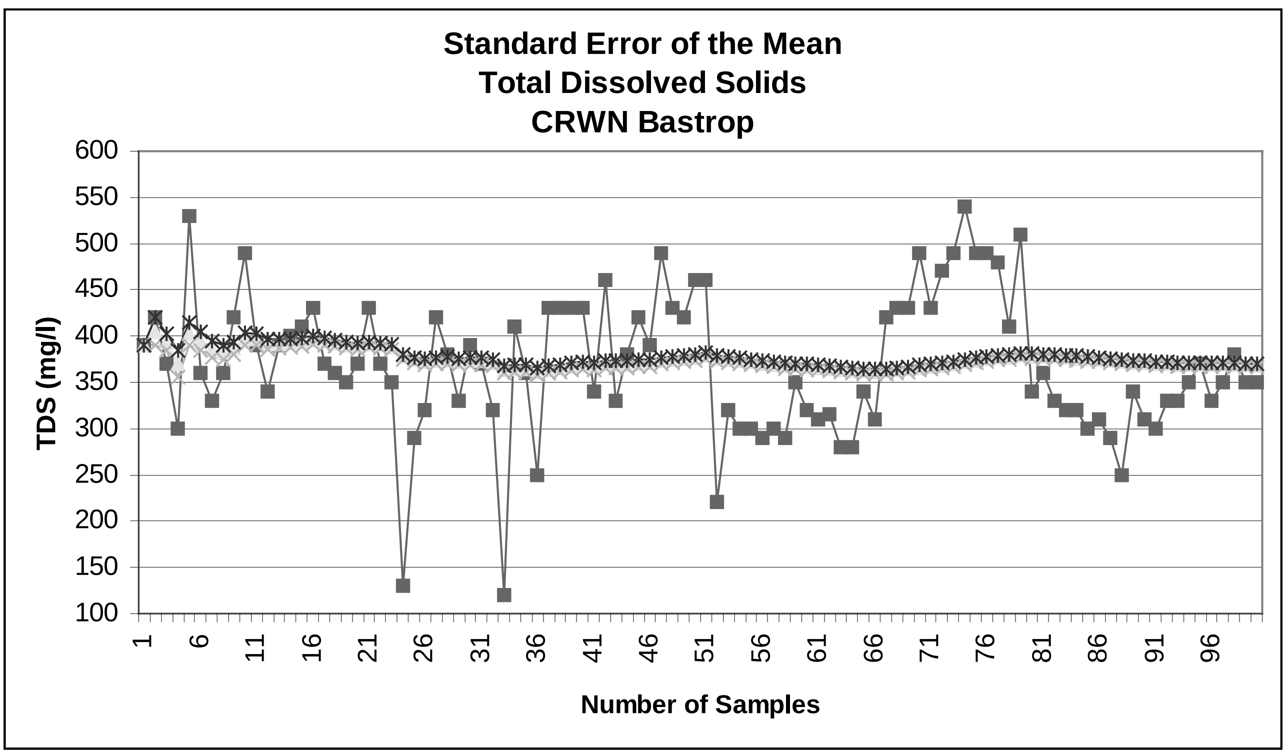


Figure 5-21 Standard Error of the Mean TDS, CRWN2

Both plots (Figures 5-38 and 5-39) indicate a great degree of stability in the TDS mean values for both data sets. The 95% confidence interval is very small after only a few measurements.

5.3 Smithville

Site Analysis- Dissolved Oxygen-- Sites LCRA 3 and CRWN 3

This site is located in the middle of river segment 1434. The LCRA monitors at this site six times a year and the CRWN has a volunteer site one mile downstream of the LCRA site. The volunteer data lacks summer month data. Only 8% of the measurements occurred between April1 and October 1.

For the following analysis a time span of approximately three and a half years is used:

LCRA data- 8/23/93 to 3/20/97, n=23

CRWN data- 8/23/93 to 4/1/97, n=39

Descriptive Statistics

Table 5-19 Descriptive Statistics, DO- Smithville

CRWN Smithville		***LCRA Smithville***	
Mean	7.73	Mean	8.59
Standard Error	0.36	Standard Error	0.32
Median	8.15	Median	8.40
Mode	7.80	Mode	6.90
Standard Deviation	2.24	Standard Deviation	1.55
Sample Variance	5.01	Sample Variance	2.40
Kurtosis	0.86	Kurtosis	-0.19
Skewness	-1.03	Skewness	0.47
Range	9.50	Range	6.2
Minimum	2	Minimum	5.9
Maximum	11.5	Maximum	12.1
Sum	301	Sum	198
Count	39	Count	23
Confidence Level(95.0%)	0.725	Confidence Level(95.0%)	0.670

The mean of the volunteer data is lower than the professional data even though volunteer data is lacking during the period when the DO would be the lowest. The mean and median values are close to one another for both data sets, but the professional data is closer.

Quantile Plots

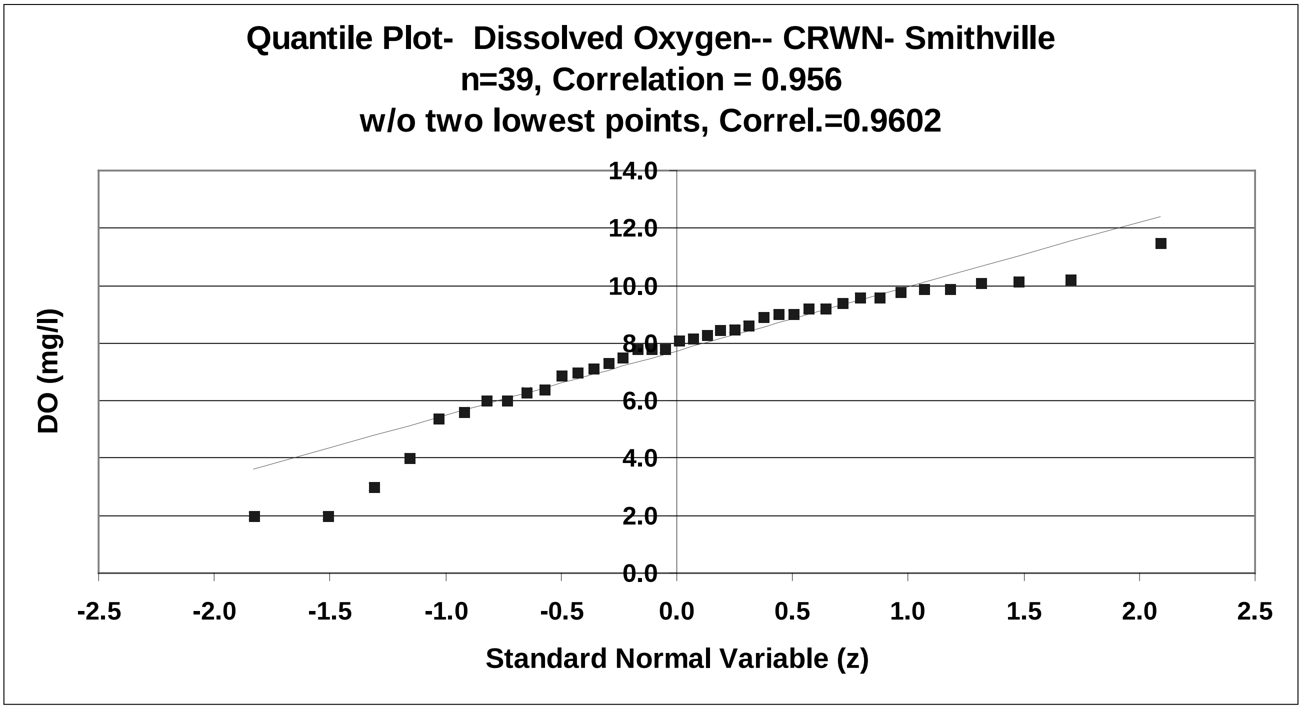


Figure 5-22 Quantile Plot, DO- CRWN3

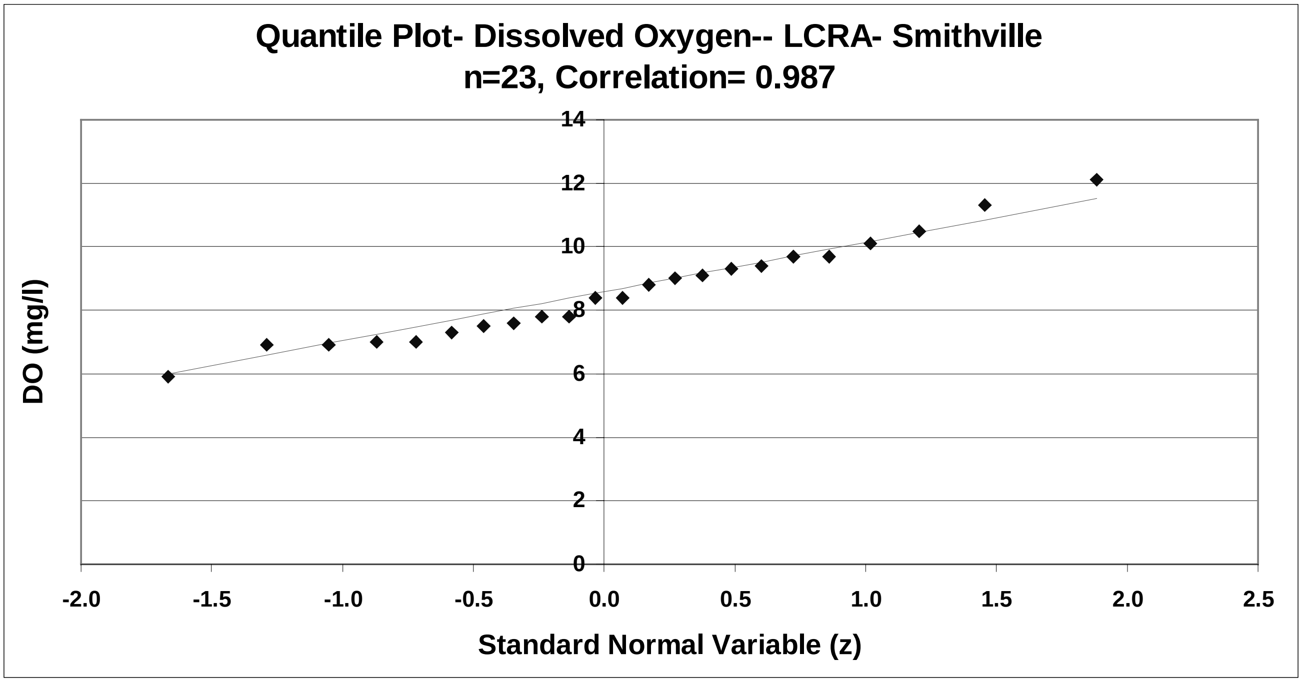
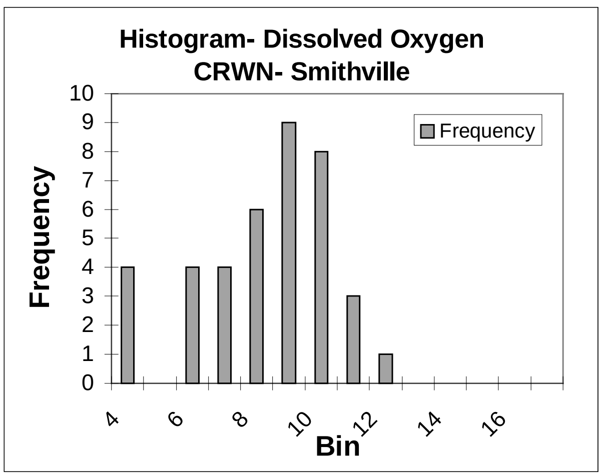


Figure 5-23 Quantile Plot DO, LCRA3

The quantile plots in Figures 5-22 and 5-23 show that both data sets are normally distributed. The LCRA correlation coefficient is greater than the lower critical value of the probability plot correlation test statistic for the normal distribution.

The CRWN data set had "outlier" data which caused the correlation coefficients to be less than the lower critical value of the probability plot correlation test statistic for the normal distribution. Eliminating just the two lowest values caused the coefficient to be greater than the lower critical value.

Histograms



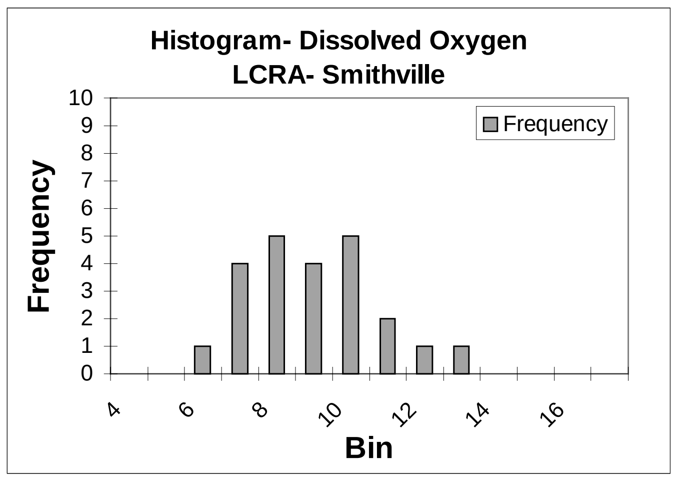


Figure 5-24 Histograms, DO- Smithville

The histogram of the volunteer data (Figure 5-24) shows a strong skewness in the negative direction (skewness=-1.03) and professional data has a good normal distribution.

t-test:

Table 5-20 t-test Results, DO- Smithville

Calculated t=	-1.777	
Smithville		
	CRWN	***LCRA***
Mean	7.73	8.59
Variance	5.01	2.40
Observations	39	23
df	58	
t Stat	-1.777	
P(T<=t) one-tail	0.040	
t Critical one-tail	1.672	
P(T<=t) two-tail	0.081	
t Critical two-tail	2.002	

The absolute value of the t Stat is less than two but still relatively high, reflecting the differences in the data sets indicated by the descriptive statistics.

Standard Error of the Mean

Table 5-21 Standard Error of the Mean, DO- Smithville

Standard Error of the Mean CRWN Data- Smithville						
index	**DO**	**Cum Mean**	**Std Dev/n^0.5**	**m-2sd**	**m+2sd**	**range**
1	6.00	6.00	0.00	6.00	6.00	0.00
2	7.80	6.90	0.45	6.00	7.80	1.80
3	2.00	5.27	0.47	4.32	6.21	1.89
4	7.50	5.83	0.34	5.15	6.50	1.36
5	2.00	5.06	0.32	4.41	5.71	1.29
6	4.00	4.88	0.31	4.27	5.49	1.22
..	..	..	..	..	..	..
36	11.50	7.60	0.14	7.32	7.88	0.56
37	9.60	7.65	0.14	7.37	7.93	0.55
38	9.20	7.69	0.14	7.42	7.97	0.55
39	9.20	7.73	0.14	7.46	8.00	0.55

Standard Error of the Mean LCRA Data- Smithville						
index	**DO**	**Cum Mean**	**Std Dev/n^0.5**	**m-2sd**	**m+2sd**	**range**
1	6.90	6.90	0.00	6.90	6.90	0.00
2	7.50	7.20	0.15	6.90	7.50	0.60
3	10.10	8.17	0.38	7.40	8.93	1.53
4	12.10	9.15	0.51	8.13	10.17	2.04
5	9.70	9.26	0.48	8.29	10.23	1.94
6	7.00	8.88	0.41	8.05	9.71	1.66
..	..	..	..	..	..	..
20	7.80	8.49	0.13	8.23	8.74	0.51
21	9.30	8.52	0.12	8.28	8.77	0.49
22	9.70	8.58	0.12	8.34	8.81	0.47
23	8.80	8.59	0.11	8.36	8.81	0.45

The results in Table 5-21, above, indicate that in 95 out of 100 similar measurements the mean would lie in the approximate range around the mean indicated in the last column. The range for the volunteer data around the mean for n=39 is 0.55 or the mean DO with 95% confidence is 7.73 ± 0.275 mg/l. The range for the professional data around the mean for n=23 is 0.45 or the mean DO level with 95% confidence is 8.59 ± 0.225 mg/l.

Combining the Professional and Volunteer Data

Table 5-22 Standard Error of the Mean, DO- Combined Data- Smithville

Combining Volunteer and Professional Data

Standard Error of the Mean CRWN Data- Smithville							
Date	**index**	**DO**	**Cum Mean**	**Std Dev/n^0.5**	**m-2sd**	**m+2sd**	**range**
08/23/93	1	6.90	6.90	0.00	6.90	6.90	0.000
09/09/93	2	6.00	6.45	0.22	6.00	6.90	0.900
10/01/93	3	7.80	6.90	0.15	6.60	7.20	0.600
10/07/93	4	2.00	5.68	0.29	5.10	6.25	1.156
10/15/93	5	7.50	6.04	0.24	5.56	6.52	0.962
10/20/93	6	7.50	6.28	0.20	5.89	6.68	0.789
..	..	..	..	..	..	..	..
12/05/96	55	5.40	7.83	0.10	7.63	8.04	0.409
01/21/97	56	10.20	7.87	0.10	7.67	8.08	0.405
01/31/97	57	11.50	7.94	0.10	7.74	8.14	0.401
02/03/97	58	9.70	7.97	0.10	7.77	8.17	0.397
02/14/97	59	9.60	8.00	0.10	7.80	8.19	0.393
02/27/97	60	9.20	8.02	0.10	7.82	8.21	0.390
03/20/97	61	9.20	8.04	0.10	7.84	8.23	0.386
04/01/97	62	8.80	8.05	0.10	7.86	8.24	0.383

By combining the two data sets the 95% confidence interval around the mean was reduced to a value which is less than either data set individually. The standard error of the mean analysis (Table 5-22) shows that for these particular data sets there is improvement in confidence in the accuracy of the data.

The descriptive statistics for the combined data sets are shown in Table 5-23 below.

Table 5-23 Descriptive Statistics, DO- Combined Data- Smithville

Descriptive Statistics	
Combined Data-Smithville	
Mean	8.048
Standard Error	0.259
Median	8.350
Mode	7.8
Standard Deviation	2.040
Sample Variance	4.160
Kurtosis	1.523
Skewness	-0.947
Range	10.1
Minimum	2
Maximum	12.1
Sum	498.95
Count	62
Confidence Level(95.0%)	0.518

The standard error for the combined data sets is lower than either individual data sets which indicates an improvement in the estimation of the mean. The data is strongly skewed in the negative direction.

Fourier Analysis-- Dissolved Oxygen-- Smithville

The statistics for the regression analysis of the two data sets in shown in Tables 5-24 and 5-25.

Table 5-24 Regression Statistics- DO- LCRA3

LCRA- Smithville-SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.83
R Square	0.68
Adjusted R Square	0.65
Standard Error	0.92
Observations	23

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	36.04	18.02	21.44	0.00
Residual	20	16.81	0.84		
Total	22	52.85			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.52	0.19	44.48	0.00	8.12	8.92
X Variable 1	1.57	0.28	5.62	0.00	0.98	2.15
X Variable 2	0.87	0.26	3.30	0.00	0.32	1.42

Table 5-25 Regression Statistics- DO- CRWN3

CRWN- Smithville-SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.64
R Square	0.41
Adjusted R Square	0.38
Standard Error	1.76
Observations	39

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	78.74	39.37	12.71	0.00
Residual	36	111.55	3.10		
Total	38	190.28			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	7.69	0.44	17.41	0.00	6.80	8.59
X Variable 1	0.53	0.68	0.78	0.44	-0.85	1.91
X Variable 2	1.84	0.37	4.93	0.00	1.08	2.59

The plot of the best-fit lines for the two data sets is shown in Figure 5-25 below.

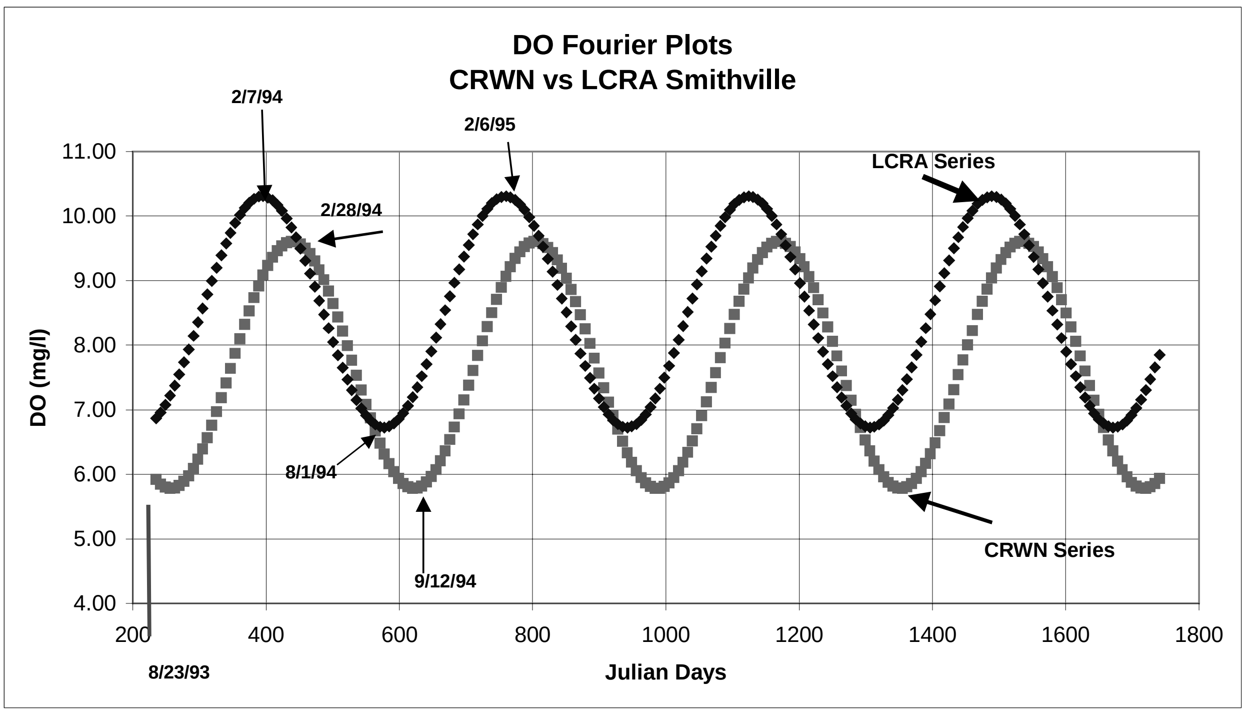


Figure 5-25 Fourier Plot, DO- Smithville

The plots of the best-fit lines for the two data sets indicate that the two data sets are very dissimilar. The CRWN data set had only 8% of the measurements occurred between April1 and October 1. Also, these data sets had the smallest number of measurement of all the data sets analyzed.

Table 5-26 below shows the comparison of the means of the two data sets as determined by the two types of analysis:

Table 5-26 Summary of Means, DO- Smithville

Method	LCRA (n=23) Mean ± Standard Error	CRWN (n=39) Mean ± Standard Error
Standard Statistics	8.59± 0.32 mg/l	7.73 ± 0.36 mg/l
Fourier Analysis (Intercept)	8.52 ± 0.20 mg/l	7.69 ± 0.44 mg/l

The mean values determined by the two methods for the data sets are very close. The means for the LCRA data differ by only 0.07 mg/l and the CRWN data by 0.16 mg/l. The means of the two data sets also differ from one another by almost 1.0 mg/l using both methods to determine the mean. It is interesting that the mean of volunteer data set is substantially lower than that of the professional data set when there is such a lack of summer data in the volunteer data.

Total Dissolved Solids Analysis-- Smithville

The scatter plot of the TDS data for the two data sets is shown in Figure 5-26 below.

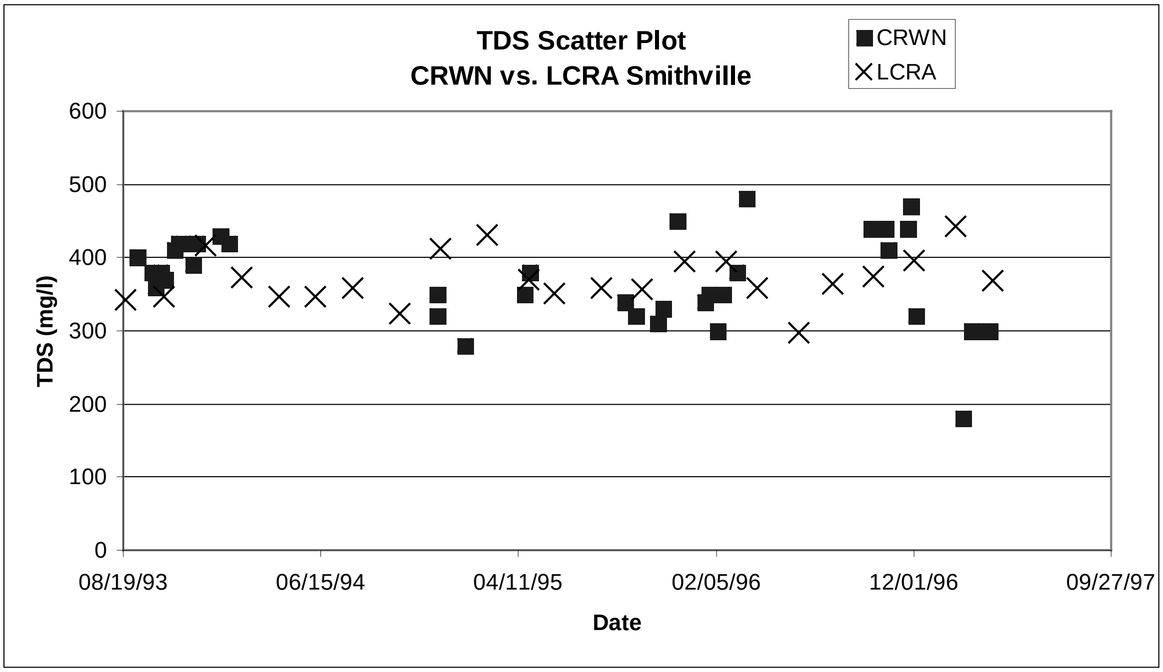


Figure 5-26 Scatter Plot, TDS- Smithville

The scatter plot indicates good agreement between the two data sets with the range and variability being about the same. The CRWN data set has been manipulated though. In March of 1995 two extremely high values were indicated, 910 and 1130 mg/l. The above plot does not include these data points and below, the descriptive statistics were done both with and without these values in the volunteer data set.

Descriptive Statistics

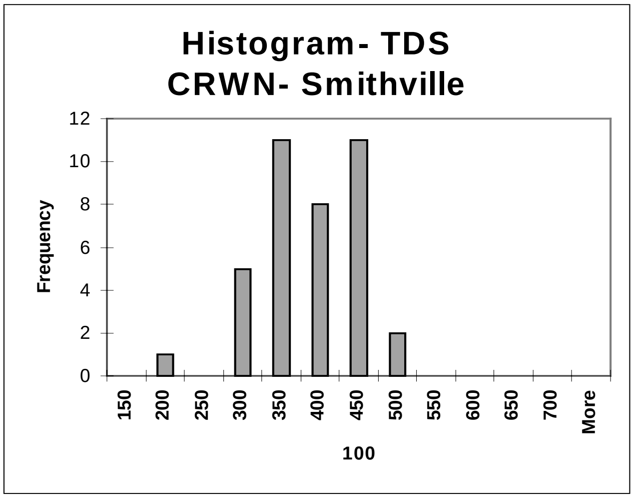
Table 5-27 Descriptive Statistics, TDS- Smithville

CRWN- Smithville		*LCRA- Smithville*	
TDS		**TDS**	
Mean	401.75	**Mean**	370.64
Standard Error	24.94	**Standard Error**	7.15
Median	380	**Median**	364.8
Mode	380	**Mode**	346.8
Standard Deviation	157.75	**Standard Deviation**	34.28
Sample Variance	24886.60	**Sample Variance**	1175.15
Kurtosis	13.73	**Kurtosis**	0.25
Skewness	3.42	**Skewness**	0.30
Range	950	**Range**	145.8
Minimum	180	**Minimum**	297.6
Maximum	1130	**Maximum**	443.4
Sum	16070	**Sum**	8524.8
Count	40	**Count**	23

CRWN- Smithville	
TDS w/ outliers deleted	
Mean	368.38
Standard Error	10.28
Median	370
Mode	380
Standard Deviation	62.52
Sample Variance	3908.41
Kurtosis	0.76
Skewness	-0.52
Range	300
Minimum	180
Maximum	480
Sum	13630
Count	37

With the extreme data values eliminated the two data sets are very similar as shown in Table 5-27. The range of the volunteer data set is almost double that of the professional data.

Histograms



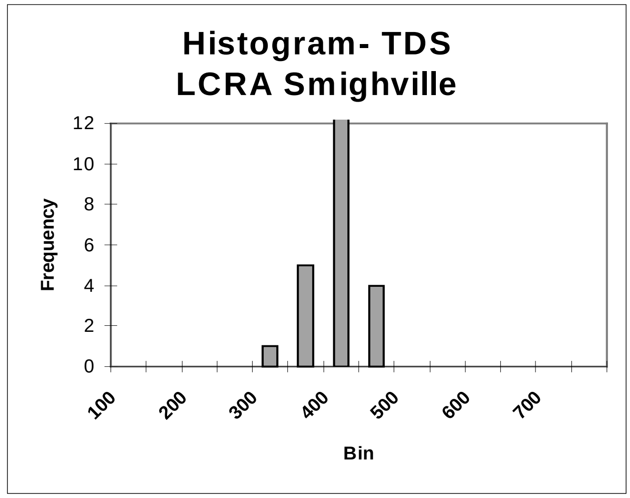


Figure 5-27 Histograms, TDS- Smithville

The histograms in Figure 5-27 indicate that the two data sets are normally distributed.

t-test:

Table 5-28 t-test Results, TDS- Smithville

t-Test: Two-Sample Assuming Unequal Variances
Smithville

	CRWN	*LCRA*
Mean	369.21	370.64
Variance	3829.09	1175.15
Observations	38	23
df	59	
t Stat	-0.12	
P(T<=t) one-tail	0.45	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.91	
t Critical two-tail	2.00	

The absolute value of the t-stat is very low, indicating that there is no statistical difference in the means of the two data sets.

The Standard Error of the Mean

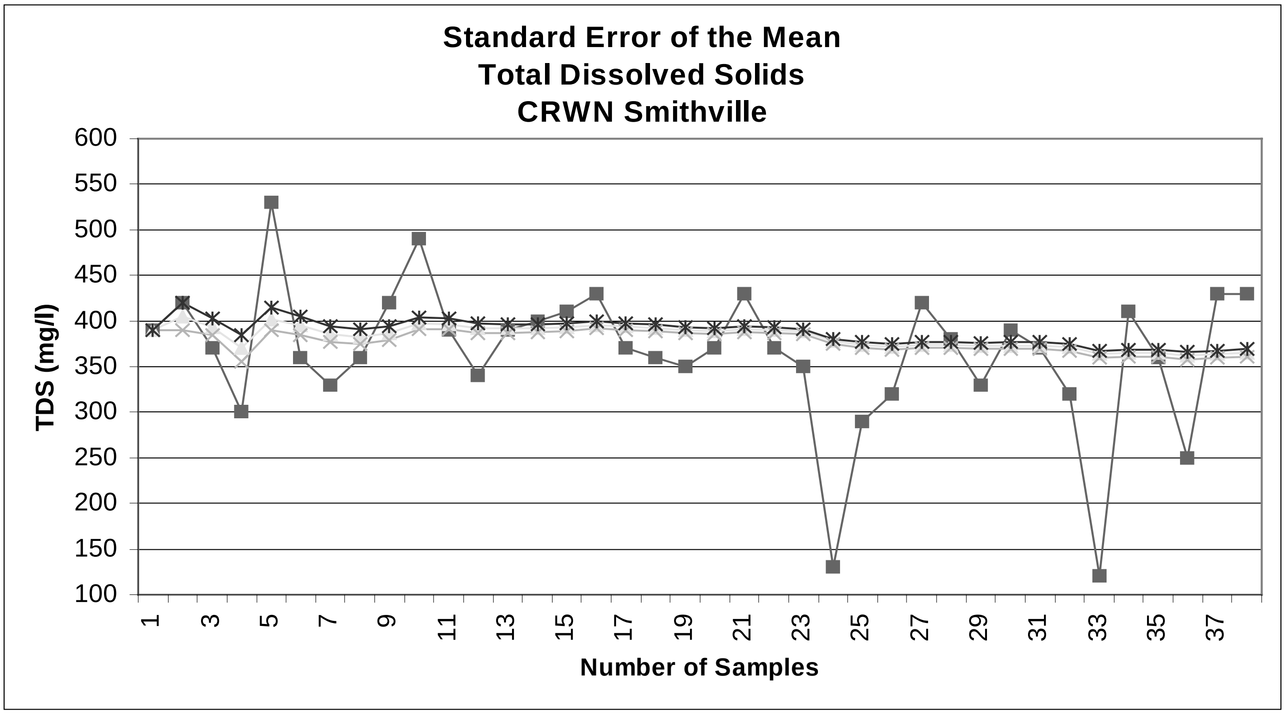


Figure 5-28 Standard Error of the Mean, TDS- CRWN3

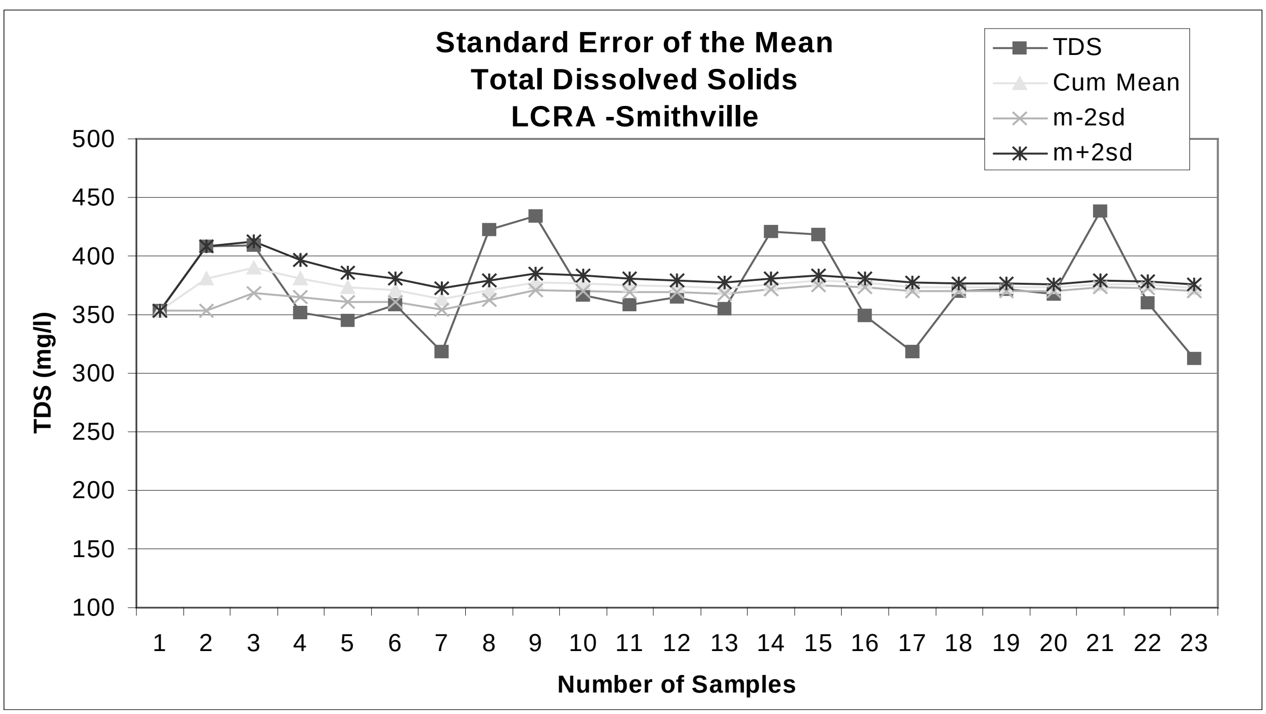


Figure 5-29 Standard Error of the Mean, TDS- LCRA3

The plots in Figures 5-28 and 5-29 show how both data sets converge at almost the same value with a very small range in the error around the means.

5.4 LaGrange

<u>Site Analysis- Dissolved Oxygen-- Sites LCRA 4 and CRWN 4</u>

These sites are located at the upstream end of river segment 1402. The CRWN site is a level 2 site monitored by a high school science class downstream of the LCRA site. The CRWN site lacked data during the period from April to August. 85% of the measurements occurred between October 1 and March 31. For the following analysis a time span of approximately two years was used.

CRWN data- 10/19/95 to 10/14/97, n=34.

LCRA data- 10/17/95 to 10/2/97, n=13

<u>Descriptive Statistics</u>

Table 5-29 is the descriptive statistics from each site for dissolved oxygen:

Table 5-29 Descriptive Statistics, DO- LaGrange

CRWN DATA- LaGrange		***LCRA DATA- LaGrange***	
Mean	9.44	Mean	8.38
Standard Error	0.40	Standard Error	0.25
Median	9.15	Median	8.40
Mode	9.20	Mode	9.40
Standard Deviation	2.31	Standard Deviation	0.92
Sample Variance	5.34	Sample Variance	0.84
Kurtosis	3.16	Kurtosis	-0.78
Skewness	1.36	Skewness	-0.09
Range	12.15	Range	3.1
Minimum	5.15	Minimum	6.8
Maximum	17.3	Maximum	9.9
Sum	320.9	Sum	109
Count	34	Count	13
Confidence Level(95.0%)	0.81	Confidence Level(95.0%)	0.56

The mean and median for both data sets are very close to each other. The CRWN data has a much higher range (2.5 times more data), a much higher standard deviation and the data is more skewed. The means of the two data sets differ by 1.0 mg/l. The higher mean dissolved oxygen level for the CRWN data is at least partially due to the fact that the CRWN data set lacks summer data when, generally, the dissolved oxygen level would normally be at its lowest. This would also account for the positive skewness.

Combining the Data Sets

Table 5-30 is the descriptive statistics of the combined data sets.

Table 5-30 Descriptive Statistics, Combined Data, DO- LaGrange

Combined Data	
LaGrange	
Mean	8.97
Standard Error	0.249
Median	8.8
Mode	7.8
Standard Deviation	1.692
Sample Variance	2.863
Kurtosis	1.432
Skewness	0.83
Range	8.95
Minimum	5.15
Maximum	14.1
Sum	412.6
Count	46

The standard error for the combined data sets is lower than either individual data sets which indicates an improvement in the estimation of the mean.

Histograms

The histograms, Figure 5-30, indicate a normal distribution for both data sets:

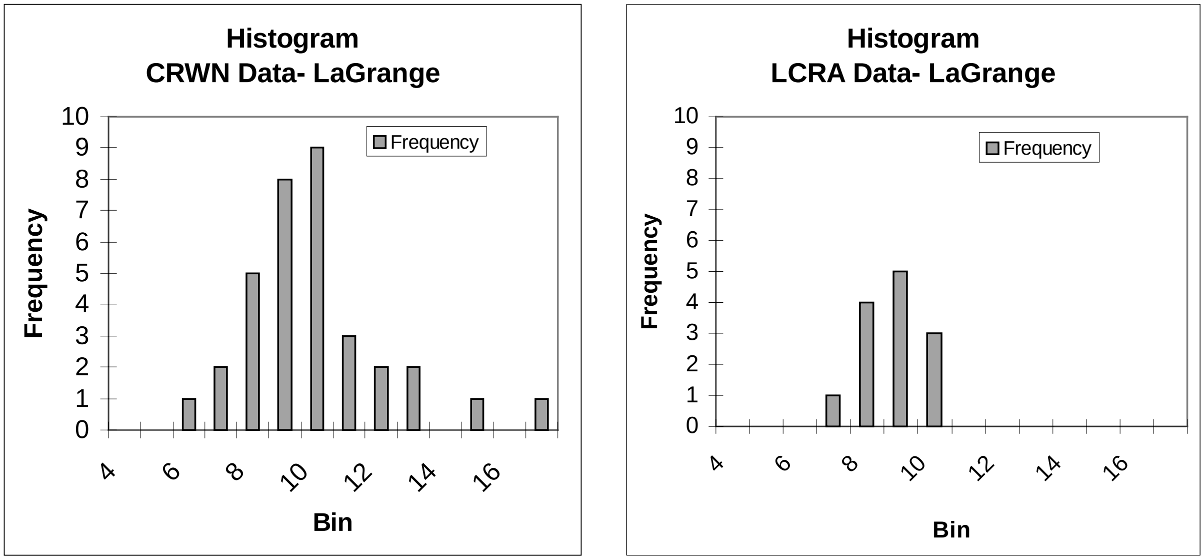


Figure 5-30 Histograms, DO- LaGrange

Quantile Plots

The quantile plots, Figures 5-31 and 5-32, show that both data sets are normally distributed. Both correlation coefficients are greater than the lower critical value of the probability plot correlation test statistic for the normal distribution.

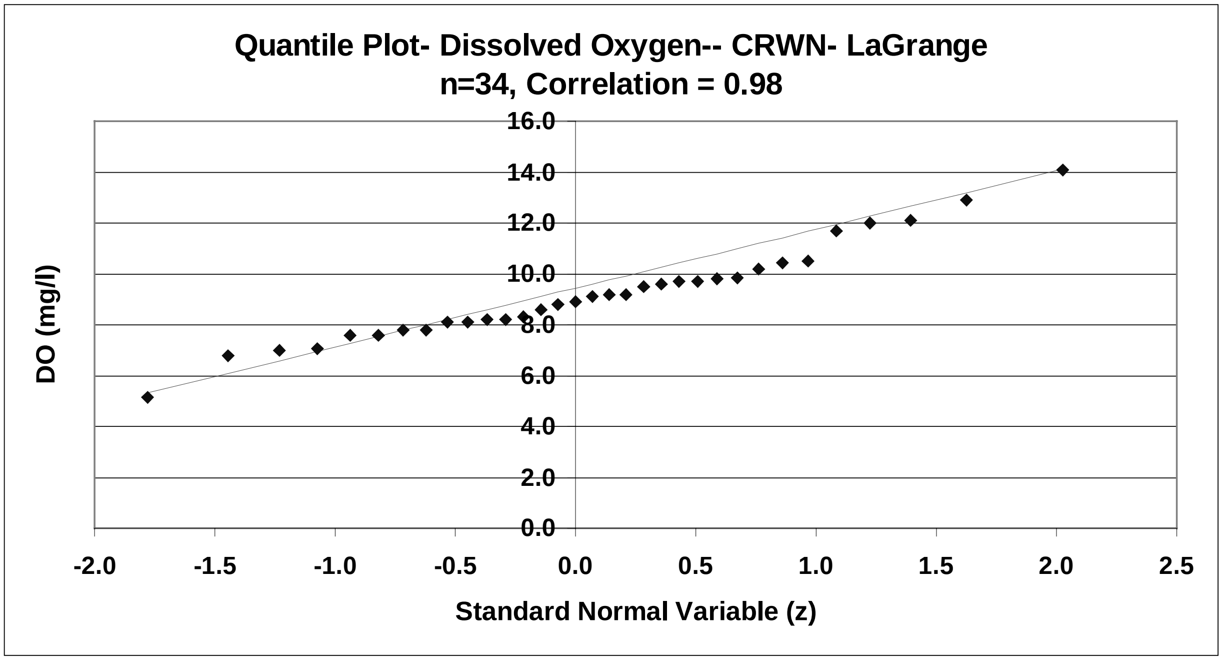


Figure 5-31 Quantile Plot, DO- CRWN4

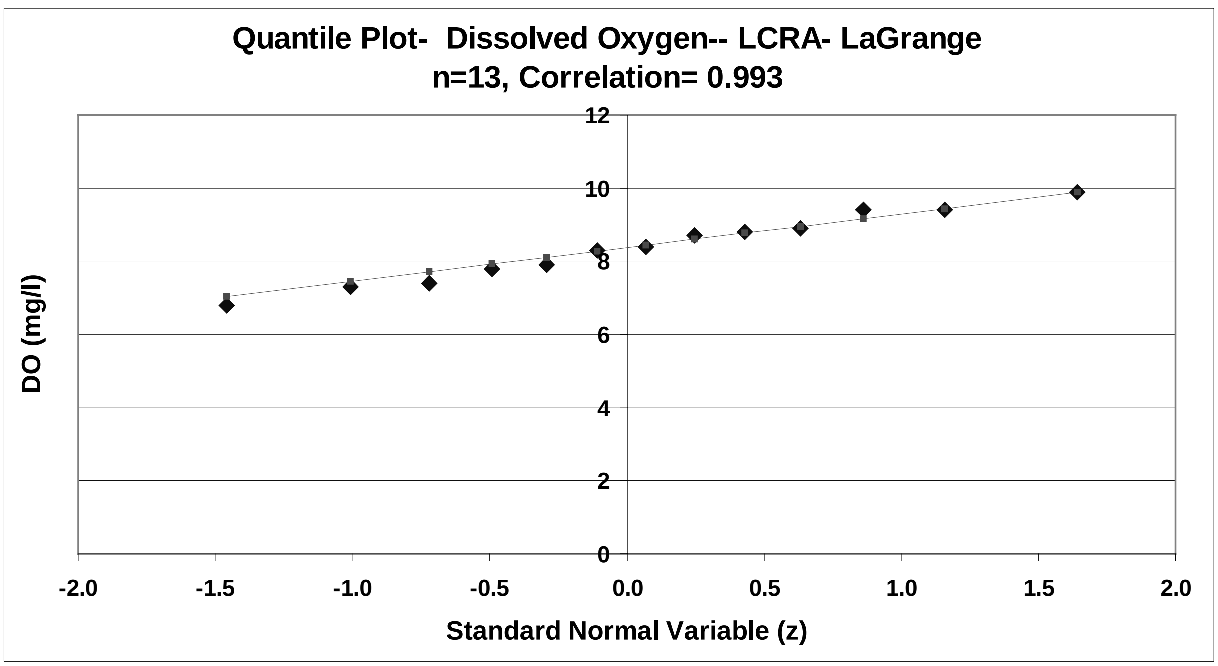


Figure 5-32 Quantile Plot, DO- LCRA4

t-test:

Table 5-31 t-test Results, DO- LaGrange

	calculated t=	2.236
	CRWN	*LCRA*
Mean	9.44	8.38
Variance	5.34	0.84
Observations	34	13
Hypothesized Mean Difference	0	
df	45	
t Stat	2.236	
P(T<=t) one-tail	0.02	
t Critical one-tail	1.68	
P(T<=t) two-tail	0.03	
t Critical two-tail	2.01	

The t value for these data sets (Table 5-31) is greater than 2. This is due mainly to the difference in the mean values of 1.06 mg/l, the largest difference of all of the sites analyzed. This indicates that the hypothesis that there is not a significant difference between the data sets is not valid. The CRWN data contained one measurement that was considerably higher than any other measurement, 17.3 mg/l. Eliminating that data point resulted in the calculated t-value being 1.969, or slightly less than 2.0. This illustrates the sensitivity of this analysis to "outlier" data. Also, the gross lack of volunteer data for one half of the year would contribute to the higher value of t. Table 5-32, below, shows the descriptive statistics of the CRWN site with the high value eliminated:

Table 5-32 Descriptive Statistics w/o Outlier Data, DO- LaGrange

CRWN DATA with highest value eliminated	
Mean	9.20
Standard Error	0.33
Median	9.1
Mode	9.2
Standard Deviation	1.87
Sample Variance	3.52
Kurtosis	0.71
Skewness	0.59
Range	8.95
Minimum	5.15
Maximum	14.1
Sum	303.6
Count	33
Calculated t Stat	1.969

Standard Error of the Mean

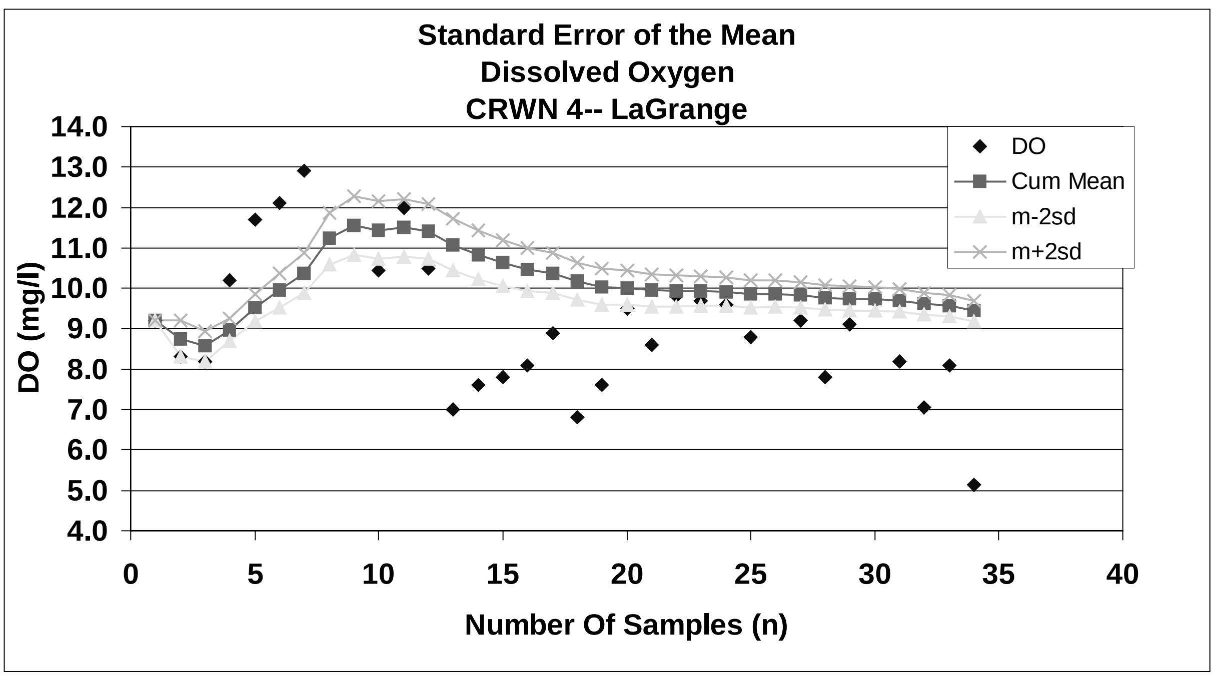


Figure 5-33 Standard Error of the Mean, DO- CRWN4

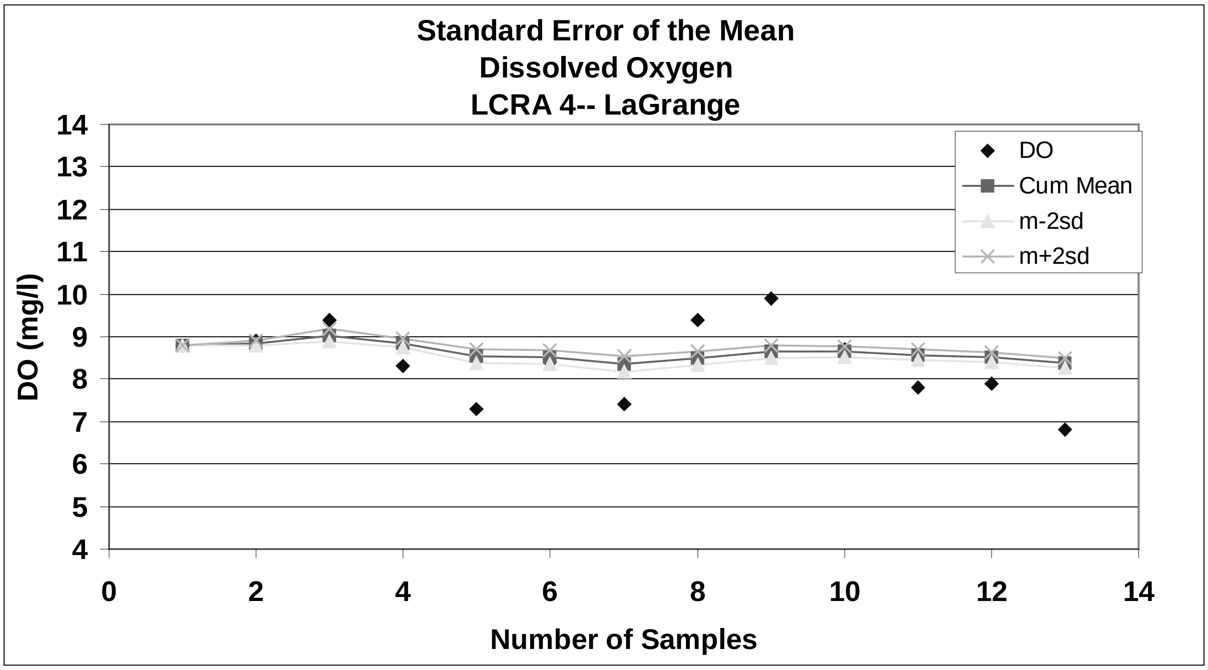


Figure 5-34 Standard Error of the Mean, DO- LCRA4

The Standard Error of the Mean plots, Figures 5-33 and 5-34, indicate the variability in the CRWN data set, while the LCRA data sets appears very stable with a very small 95% confidence interval. Below, Table 5-33, is the tabular results of the standard error of the mean evaluation of the two data sets:

Table 5-33 Standard Error of the Mean, DO- LaGrange

	Standard Error of the Mean CRWN Data- LaGrange					
index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	9.2	9.20	0	9.200	9.200	0.000
2	8.3	8.75	0.225	8.300	9.200	0.900
3	8.2	8.57	0.188	8.190	8.943	0.753
4	10.2	8.98	0.137	8.700	9.250	0.549
5	11.7	9.52	0.168	9.185	9.855	0.670
6	12.1	9.95	0.209	9.532	10.368	0.836
..	..	..	..	..	..	..
31	8.2	9.70	0.140	9.416	9.977	0.561
32	7.05	9.61	0.137	9.340	9.888	0.547
33	8.1	9.57	0.134	9.301	9.836	0.535
34	5.15	9.44	0.131	9.176	9.701	0.525

	Standard Error of the Mean LCRA Data- LaGrange					
index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	8.8	8.80	0	8.800	8.800	0.000
2	8.9	8.85	0.025	8.800	8.900	0.100
3	9.4	9.03	0.071	8.891	9.175	0.284
4	8.3	8.85	0.051	8.747	8.953	0.205
5	7.3	8.54	0.079	8.381	8.699	0.317
6	8.4	8.52	0.082	8.353	8.680	0.327
..	..	..	..	..	..	..
10	8.7	8.65	0.066	8.519	8.781	0.262
11	7.8	8.57	0.060	8.453	8.693	0.240
12	7.9	8.52	0.056	8.404	8.629	0.224
13	6.8	8.38	0.056	8.274	8.496	0.222

The variability around the mean is significantly higher for the volunteer site despite the fact that there is more than twice as much data. There is a much more variability in the volunteer measurements which may warrant taking a closer look at characteristics of the measurements, such as the consistency in time of day of the measurements and the time span that is missing. The standard error of the mean for the CRWN data with the high value of 17.3 eliminated is shown in Table 5-34 below

Table 5-34 Standard Error of the Mean w/o Outlier Data, DO- LaGrange

	Standard Error of the Mean CRWN Data- LaGrange					
index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
33	5.2	9.20	0.102	8.995	9.405	0.410

The range around the mean has been reduced by 0.115 mg/l by eliminating the one high value. This, again, illustrates the sensitivity of the analysis to "outlier" data. This also illustrates the need to have a strong quality assurance system in place to prevent erroneous data from skewing the data.

Fouier Analysis--Dissolved Oxygen

The statistics for the regression analysis of the two data sets are shown in Tables 5-35 and 5-36:

Table 5-35 Regression Statistics, DO- LCRA4

LCRA- LaGrange- SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.78
R Square	0.60
Adjusted R Square	0.53
Standard Error	0.63
Observations	13

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	6.13	3.06	7.63	0.01
Residual	10	4.01	0.40		
Total	12	10.14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.41	0.18	47.62	0.00	8.02	8.81
X Variable 1	0.87	0.26	3.30	0.01	0.28	1.45
X Variable 2	0.46	0.24	1.93	0.08	-0.07	0.99

Table 5-36 Regression Statistics, DO- CRWN4

CRWN- LaGrange- SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.65
R Square	0.43
Adjusted R Square	0.39
Standard Error	1.47
Observations	33

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	47.86	23.93	11.10	0.00
Residual	30	64.64	2.15		
Total	32	112.50			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.16	0.34	23.76	0.00	7.46	8.86
X Variable 1	2.13	0.49	4.36	0.00	1.13	3.13
X Variable 2	1.19	0.38	3.10	0.00	0.41	1.98

The plot of the CRWR-LaGrange with the raw data (Figure 5-35):

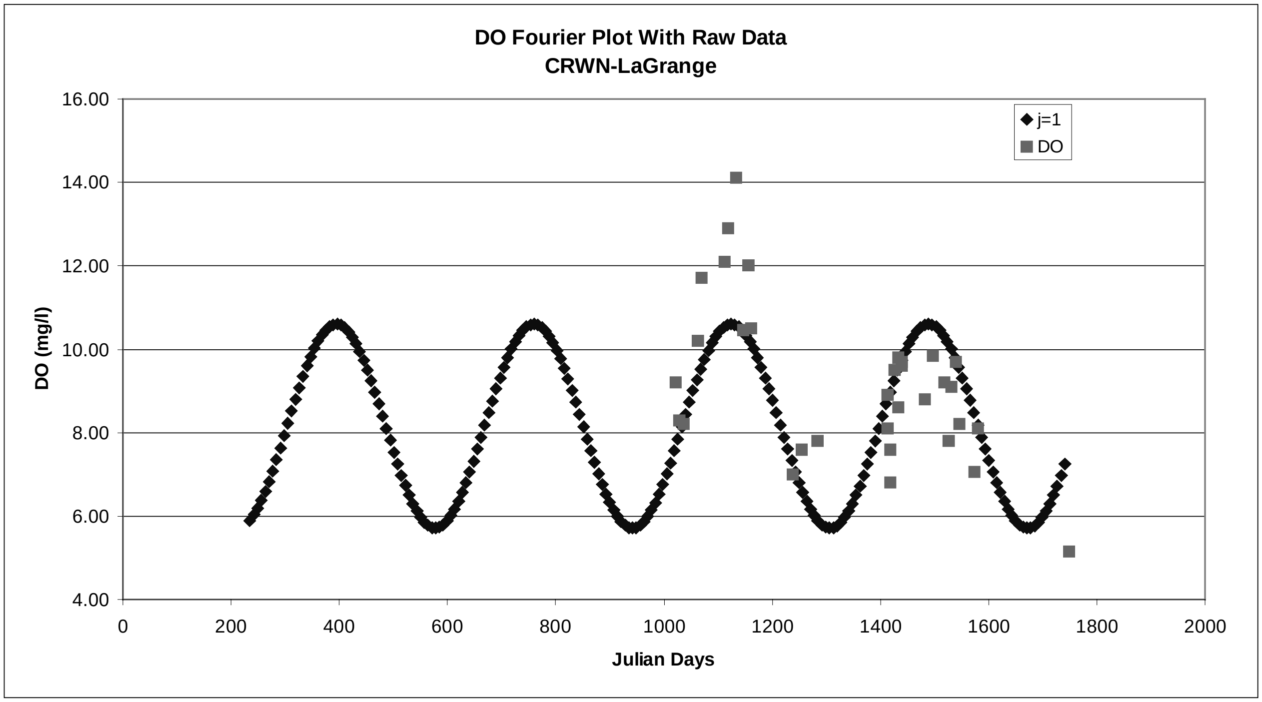


Figure 5-35 CRWN4 Fourier Plot, DO with Data

It is interesting to note that the minimums of the fourier curves are well below all but one data point. As shown below the two curves match well in time but the minimums differ by nearly 1.0 mg/l.

The plot of the best fit lines for the two data sets is shown in Figure 5-36:

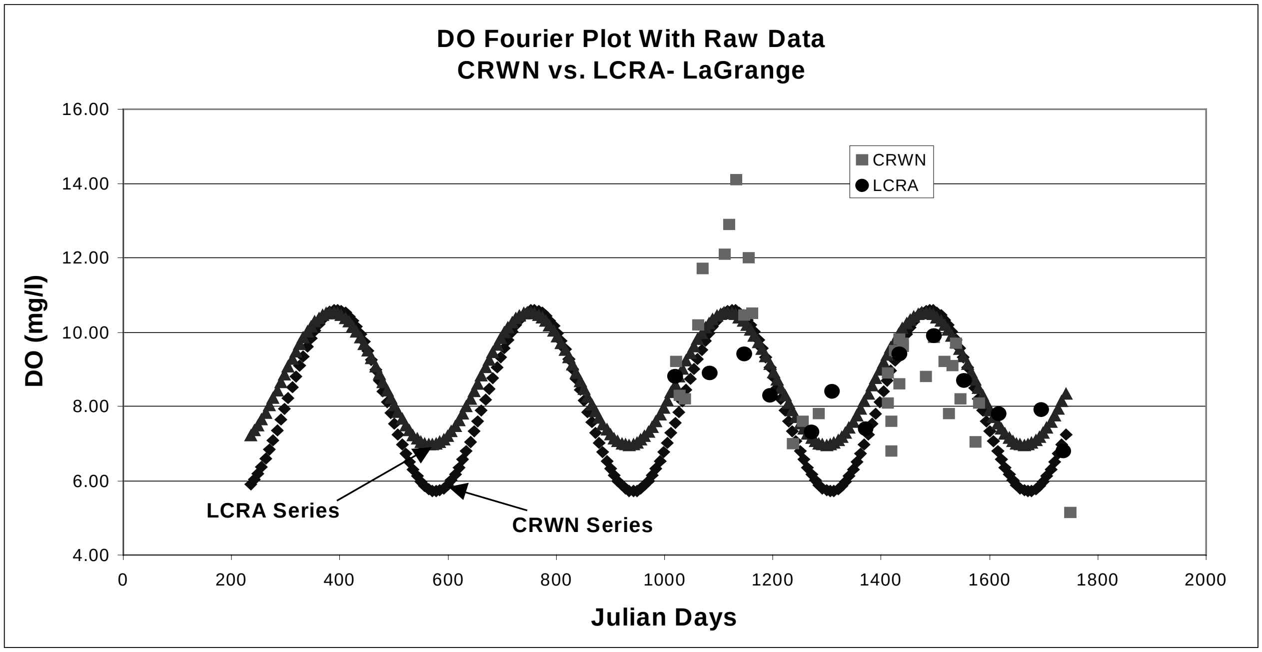


Figure 5-36 Fourier Plot, DO- LaGrange

The results of comparing the means of the two data sets as determined by the two types of analysis are shown in Table 5-37, below.

Table 5-37 Summary of Means, DO- LaGrange

Method	LCRA (n=13) Mean ± Standard Error	CRWN (n=34) Mean ± Standard Error
Standard Statistics	8.38 ± 0.25 mg/l	9.44 ± 0.40 mg/l
Fourier Analysis (Intercept)	8.41 ± 0.18 mg/l	8.16 ± 0.34 mg/l

The difference between the means of the two data sets is much lower with the fourier analysis, 0.25 mg/l, and 1.06 mg/l with standard statistics.

Total Dissolved Solids Analysis

The scatter plot for the TDS data, LCRA vs. CRWNis shown in Figure 5-37:

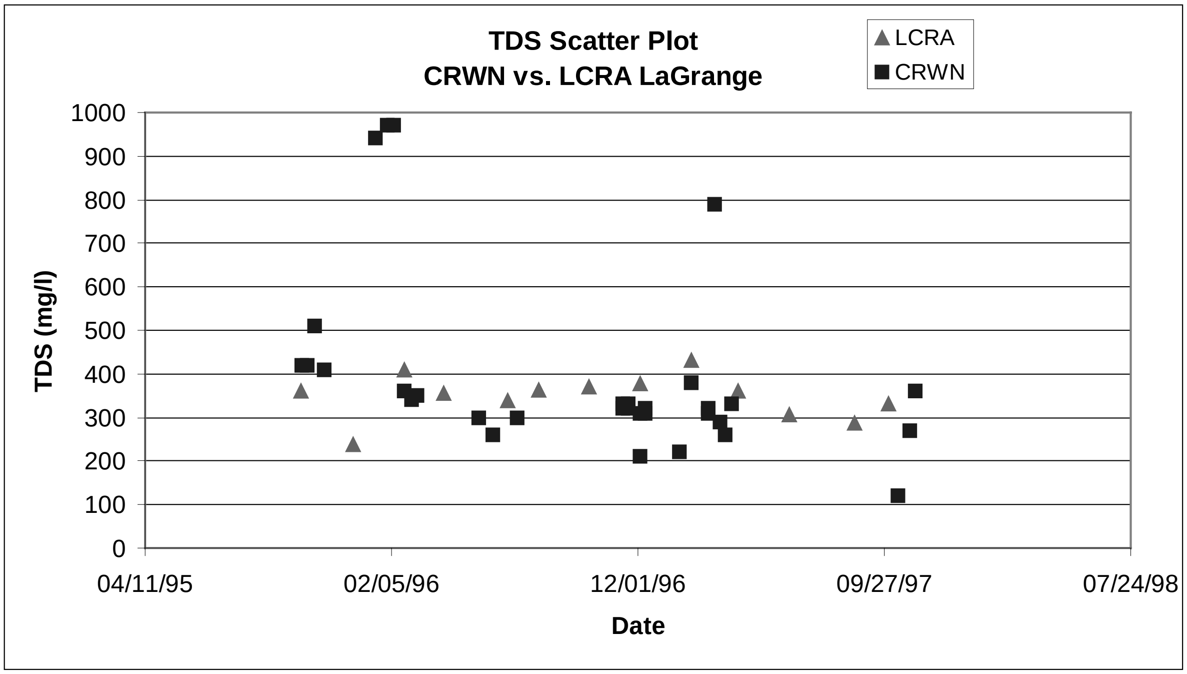


Figure 5-37 Scatter Plot, TDS- LaGrange

The two data sets are in good agreement except for 4 "outlier" points in the CRWN data with values of between 800 and 1000 mg/l.   On 2/21/96 both the professional and the volunteer monitors measured and were in within 50 mg/l of each other that day.

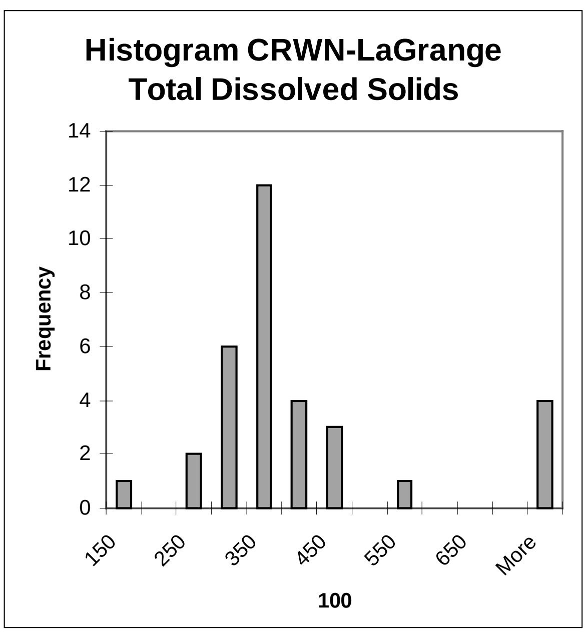
Descriptive Statistics for TDS:

Table 5-38 Descriptive Statistics, TDS- LaGrange

CRWN- LaGrange		*LCRA- LaGrange*	
TDS		**TDS**	
Mean	394.85	**Mean**	347.95
Standard Error	36.61	**Standard Error**	14.08
Median	330	**Median**	359.4
Mode	320	**Mode**	#N/A
Standard Deviation	210.31	**Standard Deviation**	50.77
Sample Variance	44232.01	**Sample Variance**	2577.09
Kurtosis	3.27	**Kurtosis**	0.83
Skewness	2.00	**Skewness**	-0.58
Range	850	**Range**	193.8
Minimum	120	**Minimum**	238.2
Maximum	970	**Maximum**	432
Sum	13030	**Sum**	4523.4
Count	33	**Count**	13

The high values in the CRWN data, shown in Figure 5-38, has caused a skewness in the positive direction. The range in the CRWN data is more than four time the LCRA data.

Histograms



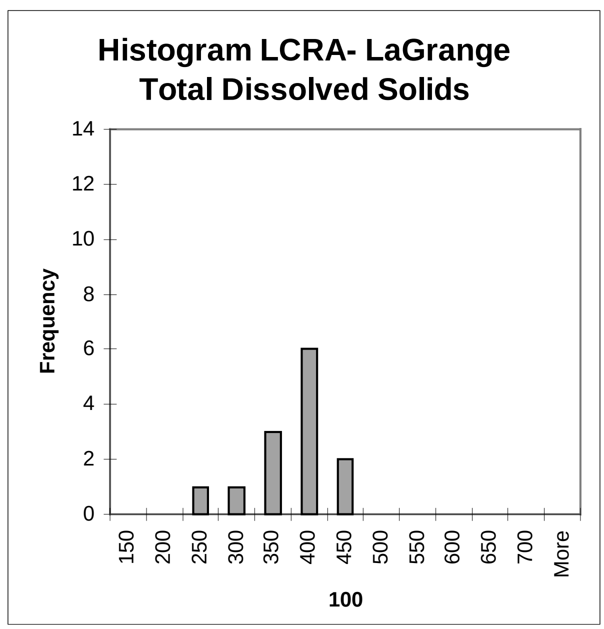


Figure 5-38 Histograms, TDS- LaGrange

The CRWN data show a strong normal distribution except for a few outlier points.

t-test

Table 5-39 t-test Results, TDS- LaGrange

Calculated t-stat.		1.196
t-Test: Two-Sample Assuming Unequal Variances LaGrange		
	CRWN	***LCRA***
Mean	394.85	347.95
Variance	44232.01	2577.09
Observations	33	13
Hypothesized Mean Difference	0	
df	40	
t Stat	1.20	
P(T<=t) one-tail	0.12	
t Critical one-tail	1.68	
P(T<=t) two-tail	0.24	
t Critical two-tail	2.02	

Despite the large differences due to the outlier data points the t-stat remains well below 2 (Table 5-39) which indicates that there is not a significant difference in the means of the data sets.

The Standard Error of the Mean Plots

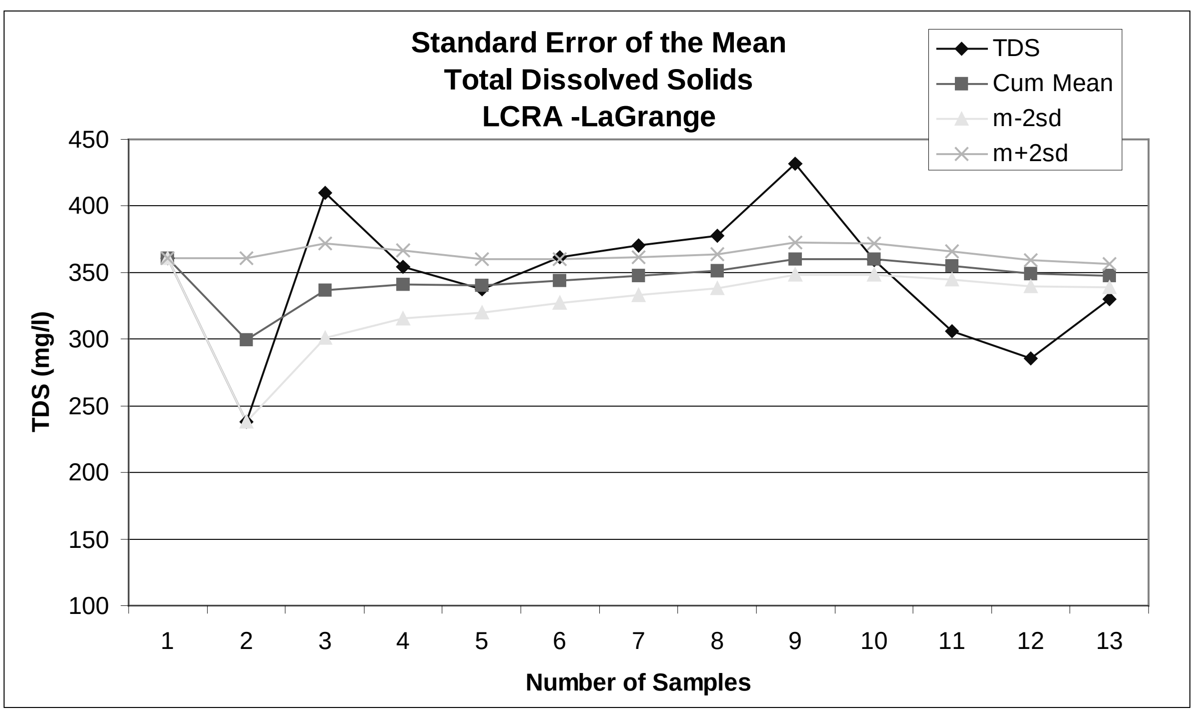


Figure 5-39 Standard Error of the Mean, TDS- LCRA4

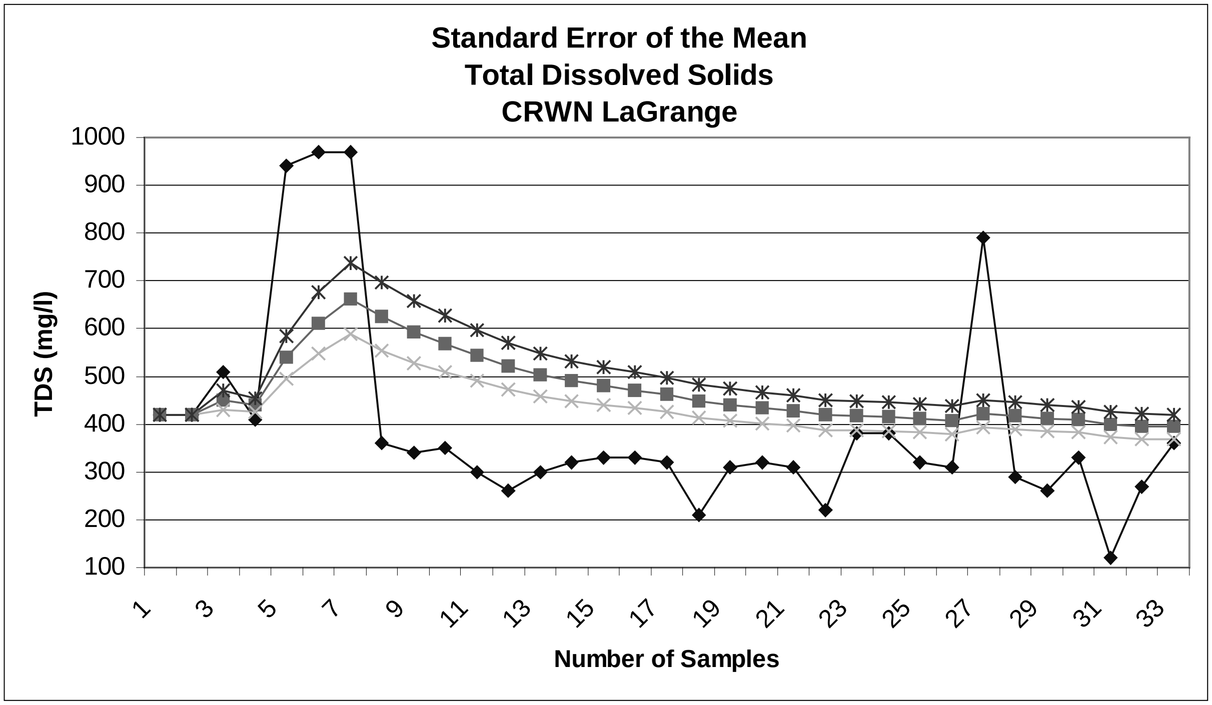


Figure 5-40 Standard Error of the Mean, TDS- CRWN4

The standard error of the mean plots, Figures 5-39 and 5-40, show how the two data sets are converging within 50 mg/l of each other. Without the high values the CRWN data set would be converging much closer with the professional data set. The two data sets are very similar to each other except for the few extremely high values in the volunteer data. Both have relatively small confidence intervals around the mean.

Columbus

Site Analysis- Dissolved Oxygen-- Sites LCRA 5 and CRWN 5

These sites are located in the middle of river segment 1402. The LCRA monitors at this site six times a year upstream from the CRWN site. The site is monitored by a high school science class, which results in a lack of data in the summer months. There is a higher percentage of measurements from April to September, 40%.

A time span of approximately four years was used for this analysis:

CRWN data- 9/27/93 to 12/10/97, n=63.

LCRA data- 8/24/93 to 10/2/97, n=26.

Descriptive Statistics

Table 5-40 is the descriptive statistics from each site for dissolved oxygen:

Table 5-40 Descriptive Statistics, DO- Columbus

CRWN DATA- Columbus		*LCRA DATA- Columbus*	
Mean	7.98	Mean	7.70
Standard Error	0.23	Standard Error	0.21
Median	8.00	Median	7.55
Mode	8.00	Mode	7.00
Standard Deviation	1.80	Standard Deviation	1.08
Sample Variance	3.24	Sample Variance	1.17
Kurtosis	6.84	Kurtosis	-0.14
Skewness	1.75	Skewness	0.08
Range	11.85	Range	4.70
Minimum	4.30	Minimum	5.30
Maximum	16.15	Maximum	10.00
Sum	502.65	Sum	200.10
Count	63.00	Count	26.00
Confidence Level(95.0%)	0.45	Confidence Level(95.0%)	0.44

Note that there are about two times more CRWN data values than LCRA data values. Both data sets have mean and median values that are very close to each other and small skewness values. The mean of the CRWN data is slightly higher than the LCRA site, probably due to the lack of summer monitoring data. The LCRA data is less skewed (0.08) than the CRWN data (1.75). The range of the CRWN data is much higher than the LCRA data (11.85 vs. 4.70). Also the means of each data set are different by almost 0.5 mg/l. The descriptive statistics indicate that the two data sets are normally distributed.

Combining the Data Sets

Table 5-41 is the descriptive statistics of the combined data sets.

Table 5-41 Descriptive Statistics, Combined Data Sets, DO- Columbus

Combined Data Sets	
Columbus	
Mean	7.90
Standard Error	0.17
Median	7.9
Mode	8
Standard Deviation	1.62
Sample Variance	2.63
Kurtosis	7.80
Skewness	1.76
Range	11.85
Minimum	4.3
Maximum	16.15
Sum	702.745
Count	89
Confidence Level(95.0%)	0.34166695

The standard error for the combined data sets is lower than either individual data sets which indicates an improvement in the estimation of the mean.

Quantile Plots

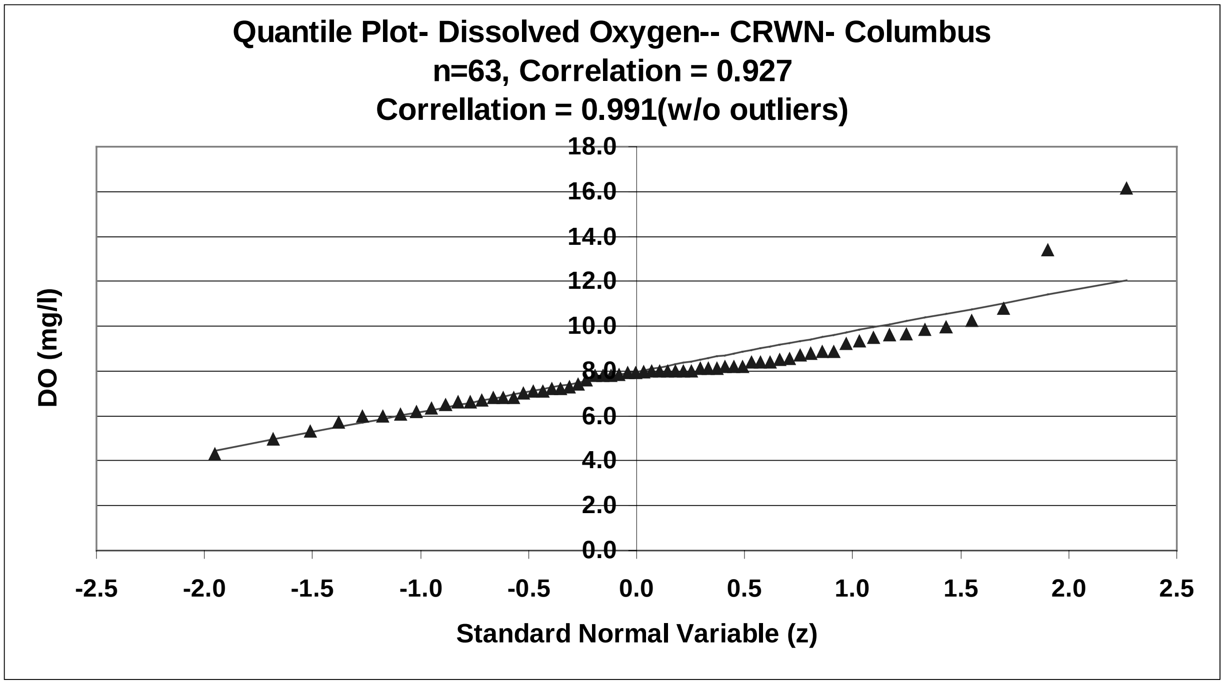


Figure 5-41 Quantile Plot, DO- CRWN5

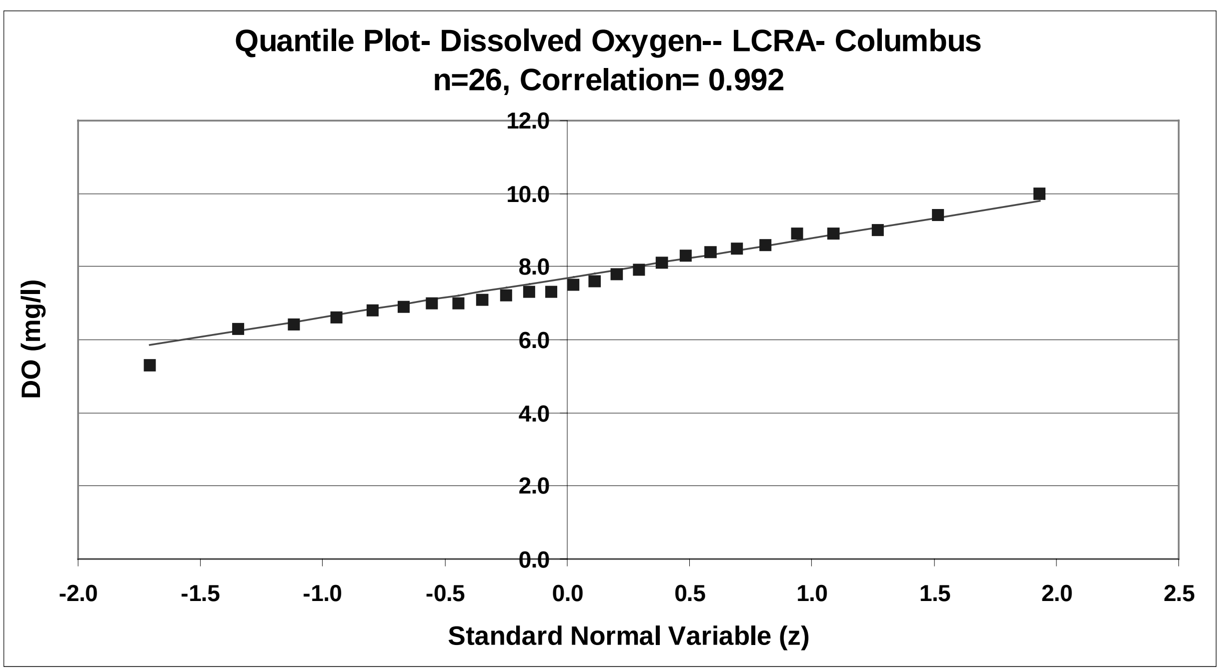
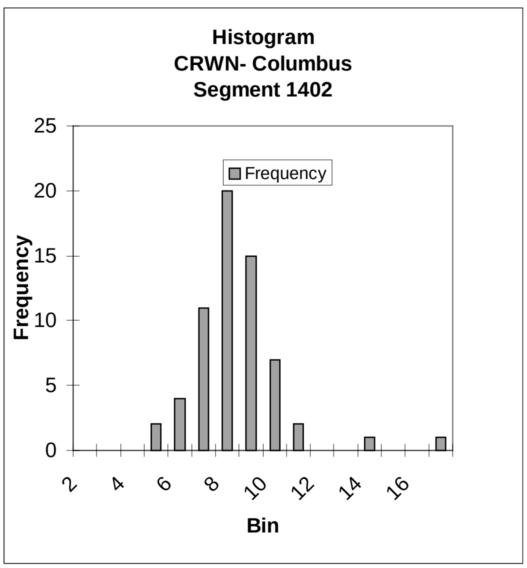


Figure 5-42 Quantile Plot, DO- LCRA5

The quantile plots in Figures 5-41 and 5-42 show that both data sets are normally distributed. The LCRA correlation coefficient is greater than the lower critical value of the probability plot correlation test statistic for the normal distribution. The CRWN data set had "outlier" data which caused the correlation coefficients to be less than the lower critical value of the probability plot correlation test statistic for the normal distribution. Eliminating just the two highest values caused the coefficient to be greater than the lower critical value.

<u>Histograms</u>

The histograms in Figure 5-43 show a good visual indication of the normality of the distributions.



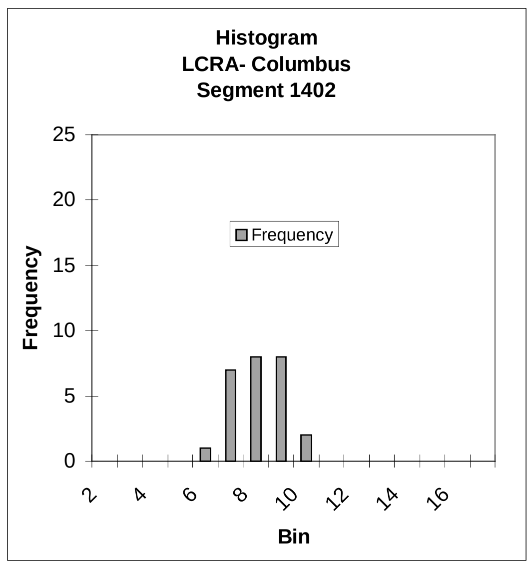


Figure 5-43 Histograms, DO- Columbus

t-test

Table 5-42 t-test Results, DO- Columbus

Calculated t= 0.909

	CRWN	*LCRA*
Mean	7.98	7.70
Variance	3.24	1.17
Observations	63	26
Hypothesized Mean Difference	0.00	
df	75.00	
t Stat	0.909	
P(T<=t) one-tail	0.18	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.37	
t Critical two-tail	1.99	

The t value of 0.909, shown in Table 5-42, indicate that there is no significant difference between the means of the two data sets.

Standard Error of the Mean

Table 5-43 Standard Error of the Mean, DO- Columbus

Standard Error of the Mean CRWN Data- Columbus

index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	7.10	7.10	0.000	7.100	7.100	0.000
2	7.20	7.15	0.025	7.100	7.200	0.100
3	7.40	7.23	0.039	7.156	7.311	0.156
4	8.00	7.43	0.071	7.282	7.568	0.286
..	..	..	..	..	..	..
58	6.60	7.94	0.043	7.858	8.031	0.173
59	8.00	7.95	0.043	7.860	8.031	0.171
60	8.10	7.95	0.042	7.864	8.032	0.168
61	9.65	7.98	0.041	7.893	8.059	0.166
62	8.10	7.98	0.041	7.896	8.060	0.163
63	8.00	7.98	0.040	7.898	8.059	0.161

Standard Error of the Mean LCRA Data- Columbus

index	DO	Cum Mean	Std Dev/n^0.5	m-2sd	m+2sd	range
1	7.00	7.00	0.000	7.000	7.000	0.000
2	6.90	6.95	0.025	6.900	7.000	0.100
3	8.90	7.60	0.209	7.182	8.018	0.835
4	10.00	8.20	0.294	7.612	8.788	1.176
..	..	..	..	..	..	..
21	8.90	7.72	0.066	7.591	7.856	0.265
22	8.50	7.76	0.063	7.633	7.886	0.253
23	8.60	7.80	0.060	7.675	7.917	0.242
24	7.60	7.79	0.058	7.672	7.903	0.231
25	6.80	7.75	0.056	7.637	7.859	0.222
26	6.40	7.70	0.053	7.589	7.803	0.214

Table 5-43, above, show that the results indicate that in 95 out of 100 similar measurements the mean would lie in the approximate range around the mean indicated in the last column. The range around the mean for the volunteer data, n=63, is 0.161 or the mean DO with 95% confidence is 7.98 ±0.08 mg/l. The range around the mean for the professional data, n=26, is 0.214 or the mean DO with 95% confidence is 7.70 ± 0.107 mg/l.

The graphical representation of the above analysis is shown below in Figures 5-44 and 5-45. Both data sets have very small confidence intervals around the mean.

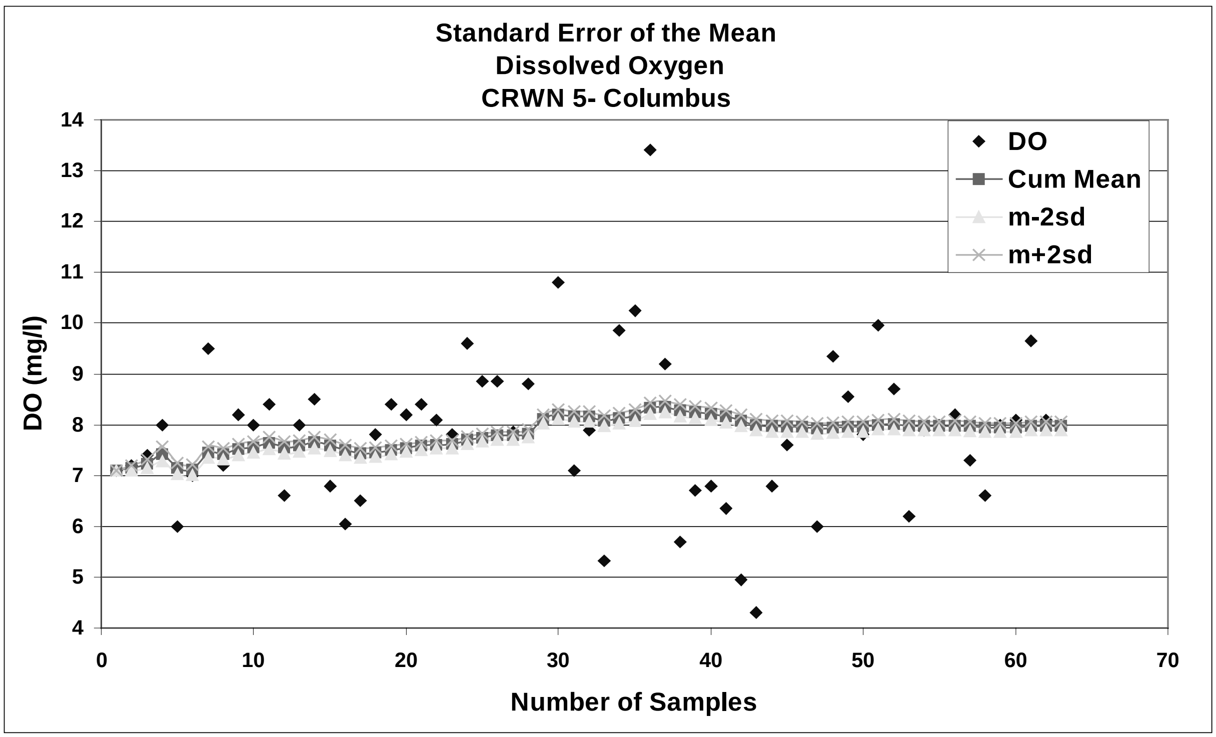


Figure 5-44 Standard Error of the Mean, DO- CRWN5

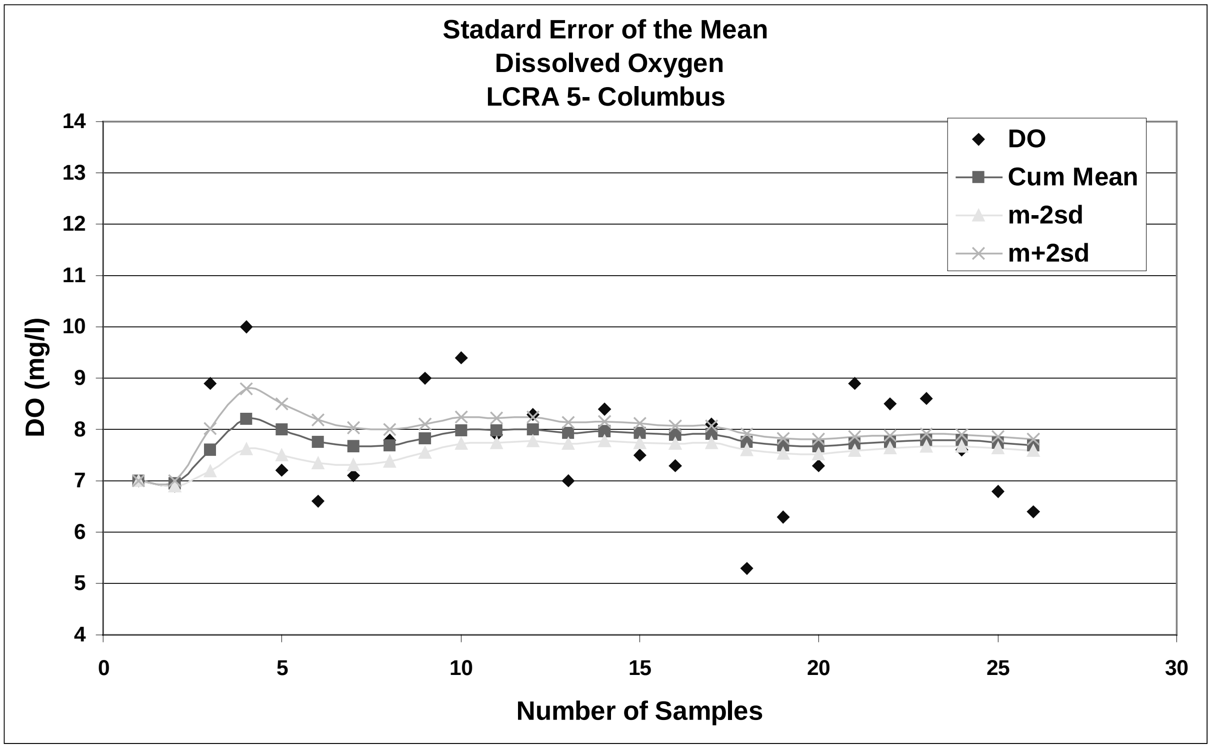


Figure 5-45 Standard Error of the Mean, DO- LCRA5

Fourier Analysis-- Dissolved Oxygen

The statistics for the regression analysis of the two data sets are shown in Tables 5-44 and 5-45.

Table 5-44 Regression Statistics, DO- LCRA5

LCRA Columbus SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.73
R Square	0.53
Adjusted R Square	0.49
Standard Error	0.78
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	15.46	7.73	12.86	0.00
Residual	23	13.83	0.60		
Total	25	29.29			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	7.75	0.15	50.80	0.00	7.44	8.07
X Variable 1	1.00	0.22	4.48	0.00	0.54	1.46
X Variable 2	0.48	0.21	2.33	0.03	0.05	0.91

Table 5-45 Regression Statistics, DO- LCRA5

CRWN Columbus SUMMARY OUTPUT
j=1

Regression Statistics	
Multiple R	0.41
R Square	0.17
Adjusted R Square	0.14
Standard Error	1.67
Observations	63

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	33.20	16.60	5.95	0.00
Residual	60	167.54	2.79		
Total	62	200.75			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	7.73	0.23	33.35	0.00	7.26	8.19
X Variable 1	1.16	0.37	3.16	0.00	0.43	1.90
X Variable 2	0.45	0.27	1.66	0.10	-0.09	1.00

The plot of the best fit lines for the two data sets are shown in Figure 5-46.

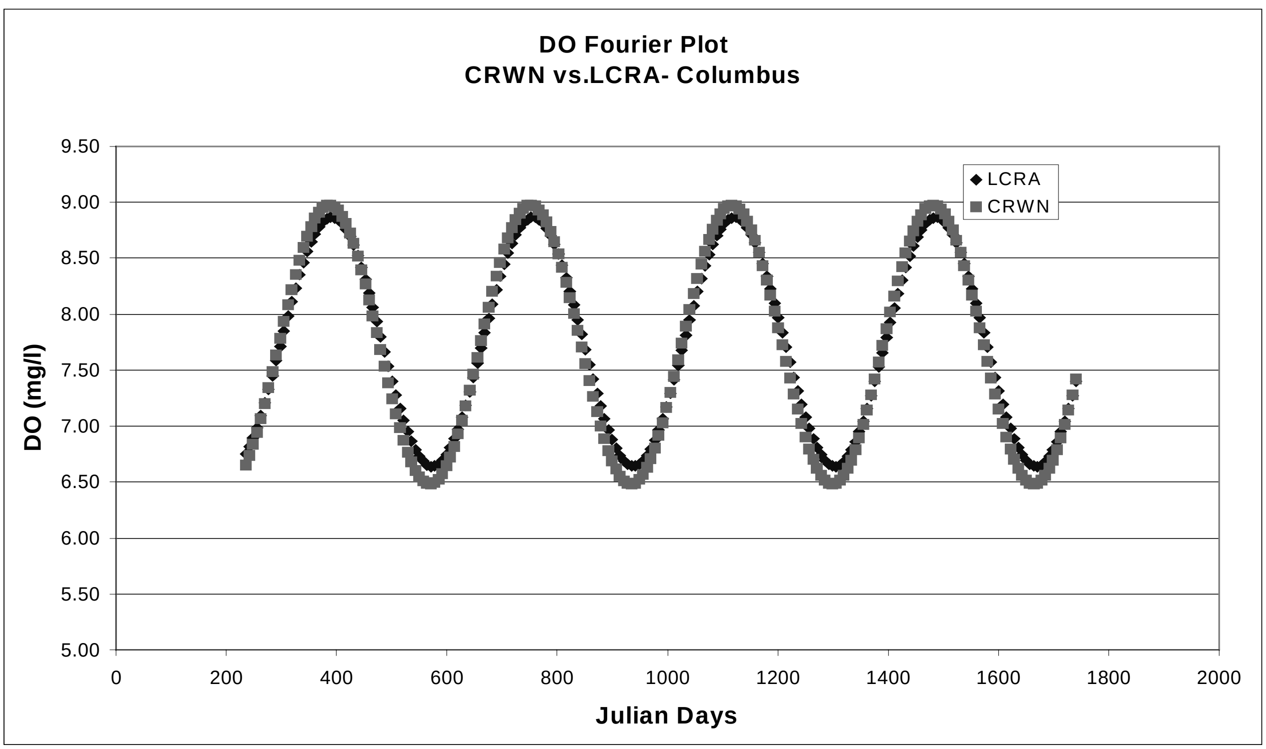


Figure 5-46 Fourier Plot, DO- Columbus

The best-fit lines for these two data sets match up very well in time and in magnitudes. Table 5-46, below, shows the comparison of the mean values determined by the two methods.

Table 5-46 Summary of Means, DO- Columbus

Method	**LCRA (n=26)** **Mean ± Standard Error**	**CRWN (n=63)** **Mean ± Standard Error**
Standard Statistics	7.70 ± 0..21 mg/l	7.98 ± 0.23 mg/l
Fourier Analysis (Intercept)	7.75 ± 0.15 mg/l	7.73 ± 0.23 mg/l

The mean values of the two data sets as determined by the two types of analysis are in good agreement. The two means for each data set are within the error range. The mean for the CRWN data is lower when calculated by the fourier series method. Again, this is probably due to the lack of summer data for CRWN data set.

Total Dissolved Solids Analysis- Columbus

The TDS scatter plot for the two data sets is shown in Figure 5-47.

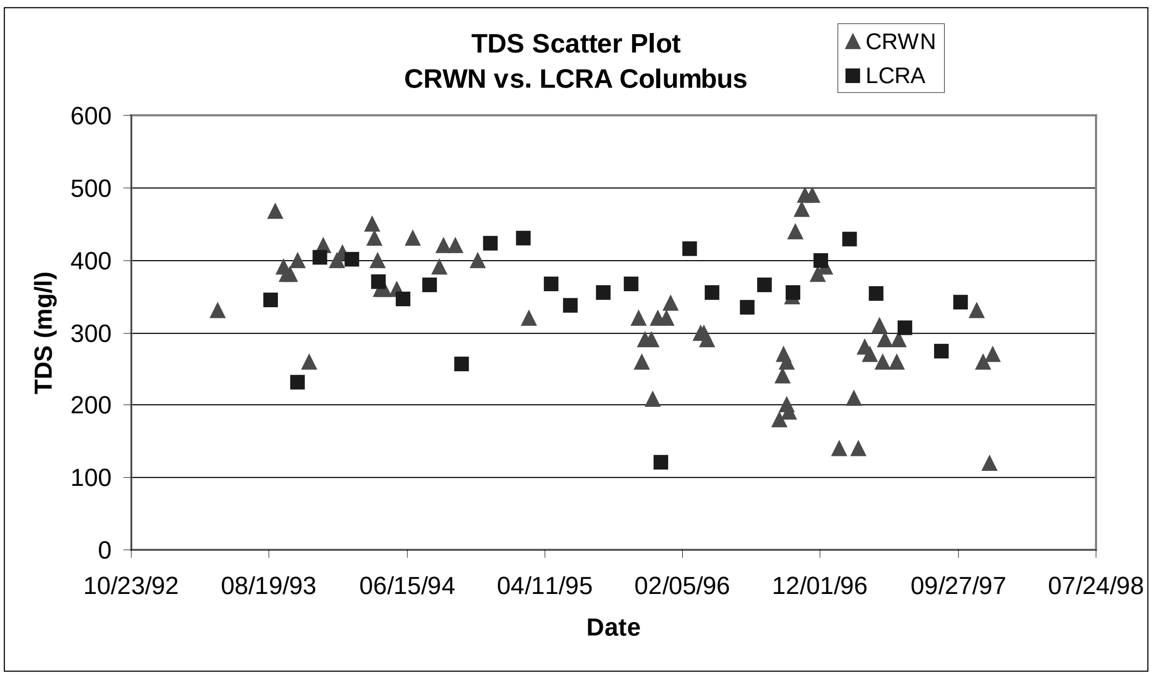


Figure 5-47 Scatter Plot, TDS- Columbus

The scatter plot indicates good agreement between the two data sets in the range and variability.

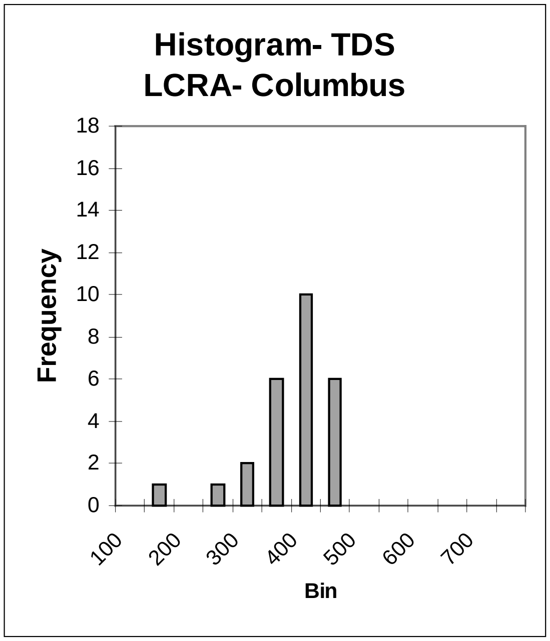
<u>Descriptive Statistics:</u>

Table 5-47 Descriptive Statistics, TDS- Columbus

CRWN- Columbus			***LCRA- Columbus***	
TDS			**TDS**	
Mean	326.60		**Mean**	348.14
Standard Error	11.56		**Standard Error**	13.31
Median	320		**Median**	355
Mode	260		**Mode**	355
Standard Deviation	89.58		**Standard Deviation**	67.85
Sample Variance	8024.31		**Sample Variance**	4603.37
Kurtosis	-0.46		**Kurtosis**	4.06
Skewness	-0.25		**Skewness**	-1.69
Range	370		**Range**	310
Minimum	120		**Minimum**	121
Maximum	490		**Maximum**	431
Sum	19596		**Sum**	9052
Count	60		**Count**	26

Table 5-47 shows that both of the data set's statistics indicate normal distribution with the mean and median values close to one another and low skewness values, though the LCRA data is slightly skewed in the negative direction. The range and maximum/minimum values are in good agreement.

Histograms



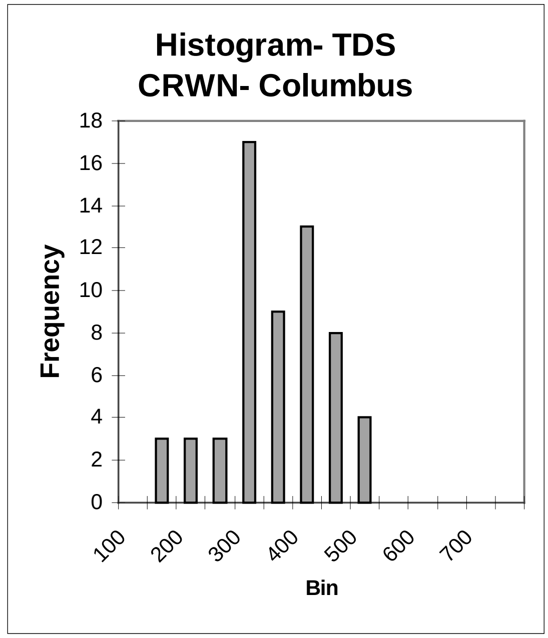


Figure 5-48 Histograms, TDS- Columbus

The histograms, Figure 5-48, show both data sets being skewed in the negative direction.

t-test

Table 5-48 t-test Results, TDS- Columbus

t-Test: Two-Sample Assuming Unequal Variances
Columbus

	CRWN	*LCRA*
Mean	326.60	348.14
Variance	8024.31	4603.37
Observations	60	26
df	62	
t Stat	-1.22	
P(T<=t) one-tail	0.11	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.23	
t Critical two-tail	2.00	

The absolute value of the t-stat is less than 2 indicating that there is not a significant difference in the means of the two data sets.

__**The Standard Error of the Means Plots**__

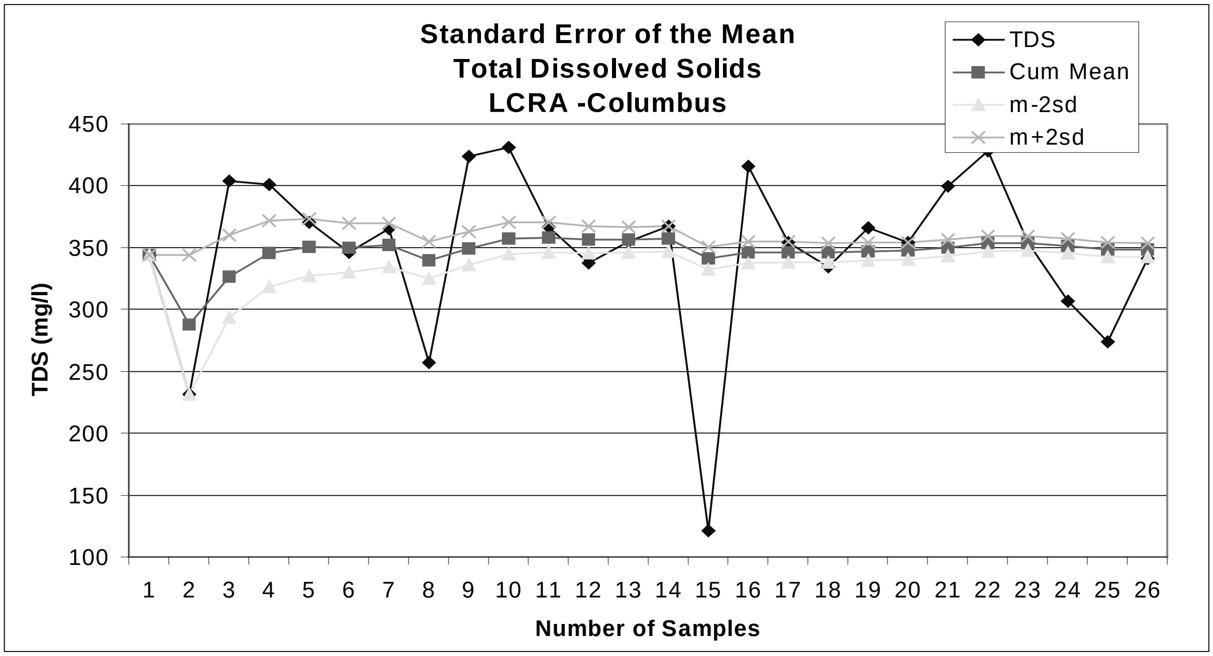


Figure 5-49 Standard Error of the Mean, TDS- LCRA5

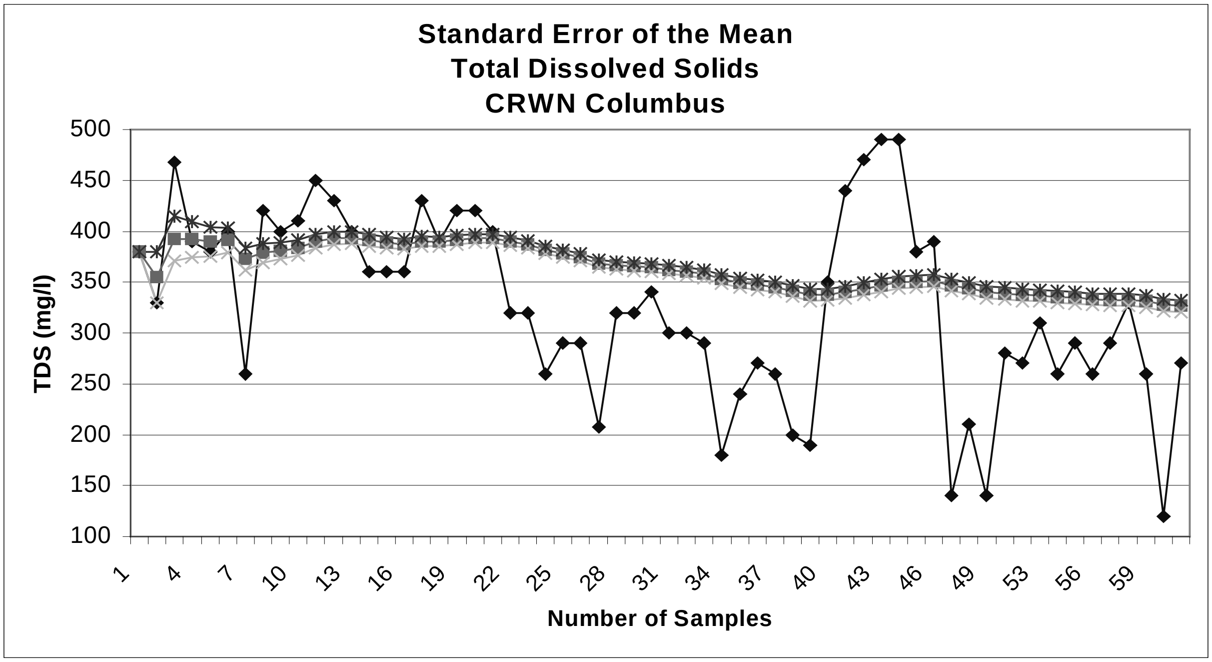


Figure 5-50 Standard Error of the Mean, TDS- CRWN5

Both data sets converge close to 350 mg/l with a very small error range around the mean. The mean of CRWN data appears to be in a steady decline. This trend does not appear in the LCRA data. It would be interesting to note what happens with trend with future measurements.

6 SUMMARIES AND CONCLUSIONS

All of the data sets used in this study were shown to be normally distributed using descriptive statistics, histograms and quantile plots. The quantile plot is the most precise tool to determine if the data is normally distributed.

The results of the quantile plot analysis for two data sets were distorted by a few "outlier" data values in the set. The elimination of these outlier data points caused the quantile analysis correlation value to strongly indicate normal distribution.

Once it was established that all of the data sets were normally distributed the statistical methods to determine the confidence interval of the mean and to compare the two independent data sets could be chosen.

The standard error of the mean analysis consistently showed an increase in confidence in the mean with an increase in the number of measurements in the data set. The volunteer data consistently had a larger number of data points, a larger range and more variability than the professional data sets resulting in a larger standard error of the mean for the volunteer data in some instances. This may be inherent to simply having a larger number of measurements over a given time period. The accuracy of the instruments and the measurement methodology can also contribute to this. Tables 6-1 and 6-2 summarize the standard error of the mean analysis of the data sets for dissolved oxygen and total dissolved solids respectively.

Table 6-1 Summary of Data, DO

Summary of Data-- Standard Error of the Mean Results-- Dissolved Oxygen								
Site	Number of Samples (n)		Professional			Volunteer		
	Professional (n)	Volunteer (n)	Range	Mean	95%	Range	Mean	95%
Lake Austin	27	90	8	8.53	0.06	7	8.05	0.35
Bastrop	26	105	6.3	8.96	0.24	10.7	9.35	0.09
Smithville	23	37	6.2	8.59	0.22	9.5	7.73	0.27
LaGrange	13	34	3.1	8.38	0.11	12.15	9.44	0.26
Columbus	26	63	4.7	7.70	0.17	11.85	7.98	0.08

Table 6-2 Summary of Data, TDS

Summary of Data-- Standard Error of the Mean Results-- Total Dissolved Solids								
Site	Number of Samples (n) Professional (n)	Volunteer (n)	Professional Range	Professional Mean	Professional 95%	Volunteer Range	Volunteer Mean	Volunteer 95%
Lake Austin	27	91	99	327	1.6	490	334	15.5
Bastrop	25	101	236	363	2.8	420	368	2.0
Smithville	23	38	146	371	3.1	300	369	3.0
LaGrange	13	33	194	348	8.9	850	395	26.0
Columbus	26	61	310	348	5.4	370	327	5.8

The t-values shown in Table 6-3 for dissolved oxygen and Table 6-4 for total dissolved solids are consistently below the value of two which indicates that the means of all of the data sets compared were not significantly different.

Table 6-3 Summary of Data, DO, t-tests

Summary of Data --Descriptive Statistics, t-value									
Sites	Professional (n)	Volunteer (n)	Professional Range	Professional Max.	Professional Min.	Volunteer Range	Volunteer Max.	Volunteer Min.	t
Lake Austin	27	90	8	12.8	4.8	7	11.6	4.6	1.39
Bastrop	27	90	6.3	13.3	7	10.7	14.8	4.1	1.11
Smithville	23	40	6.2	12.1	5.9	9.5	11.5	2	1.78
LaGrange	13	33	3.1	9.9	6.8	12.15	17.3	5.15	1.97
Columbus	26	63	4.7	10	5.3	11.85	16.15	4.3	0.91

Table 6-4 Summary of Data, TDS, t-tests

Summary of Data --Descriptive Statistics-- Total Dissolved Solids									
Sites	Professional (n)	Volunteer (n)	Professional Range	Professional Max.	Professional Min.	Volunteer Range	Volunteer Max.	Volunteer Min.	t
Lake Austin	27	91	99	374	275	490	590	100	0.77
Bastrop	25	100	236	439	202	420	540	120	0.44
Smithville	23	37	146	443	298	*300	480	180	1.78
LaGrange	13	33	194	432	238	850	970	120	1.97
Columbus	26	60	310	431	121	370	490	120	1.22

* Outlier Data Removed

It should be noted that in some of the cases the professional and volunteer sites were separated by several river miles and had discharges from creeks and wastewater treatment plants between them. For instance both Smithville and LaGrange have wastewater treatment discharge points between the two monitoring sites. This type of situation could be the focus of further study of the data.

The table in Table 6-5 is a summary of the mean values of dissolved oxygen and total dissolved solids. The sites are in downstream order.

Table 6-5 Summary of Data, Means, DO and TDS

Table 1. Summary of Data- Mean Values-- Fourier vs. Standard								
Site	Number of Samples (n)		Mean Values Dissolved Oxygen (mg/l)				Mean Values Total Dissolved Solids (mg/l)	
	Professional (n)	Volunteer (n)	Professional Standard	Professional Fourier	Volunteer Standard	Volunteer Fourier	Professional	Volunteer
Lake Austin	27	90	8.53	8.61	8.05	8.22	327	333
Bastrop	27	90	8.96	9.04	9.35	8.55	363	368
Smithville	23	40	8.59	8.52	7.73	7.69	370	368
LaGrange	13	33	8.38	8.41	9.44	8.16	347	394
Columbus	26	63	7.70	7.75	7.98	7.73	348	326

Three of the four volunteer sites with data lacking during the summer months, when DO levels would be the lowest, had mean values which were higher than the professional data set. The exception, Smithville, had several extremely low values which pulled its mean down below the mean of the corresponding professional data set. The fourier method for determining the means resulted in these same three volunteer data sets having lower mean values than the professional data sets. This would be due to the correction for the lack of measurements in the summer by using the fourier method to determine the mean. Figures 6-1 and 6-2 are the graphical representations of the data in Table 6-5.

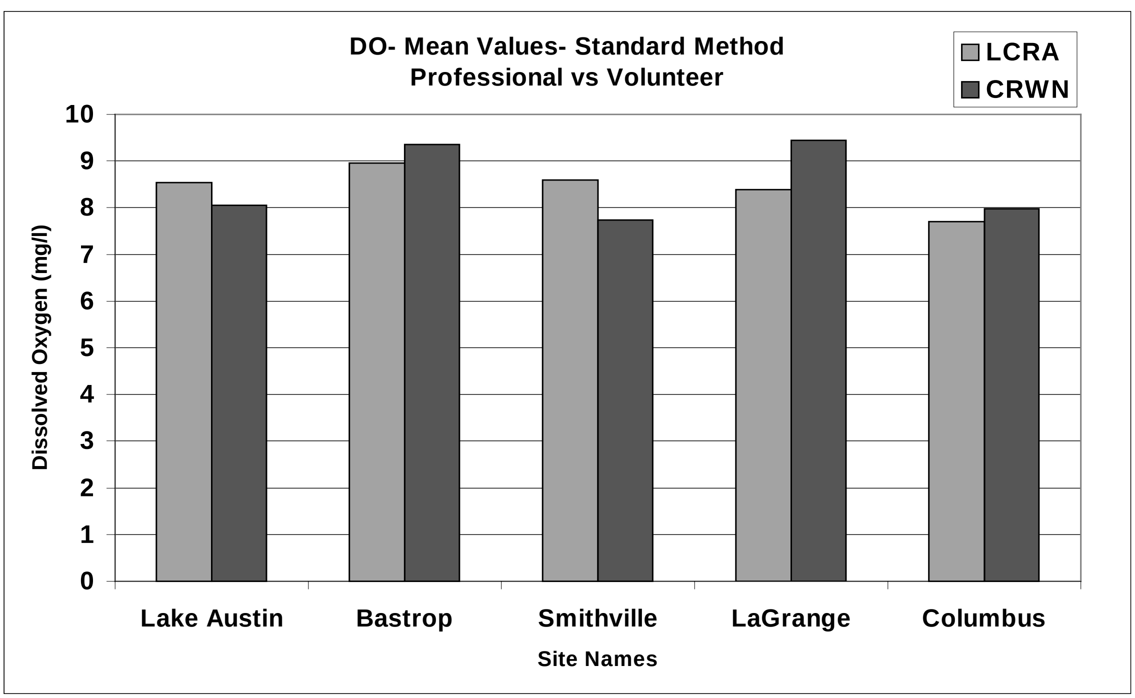


Figure 6-1 DO Mean Value

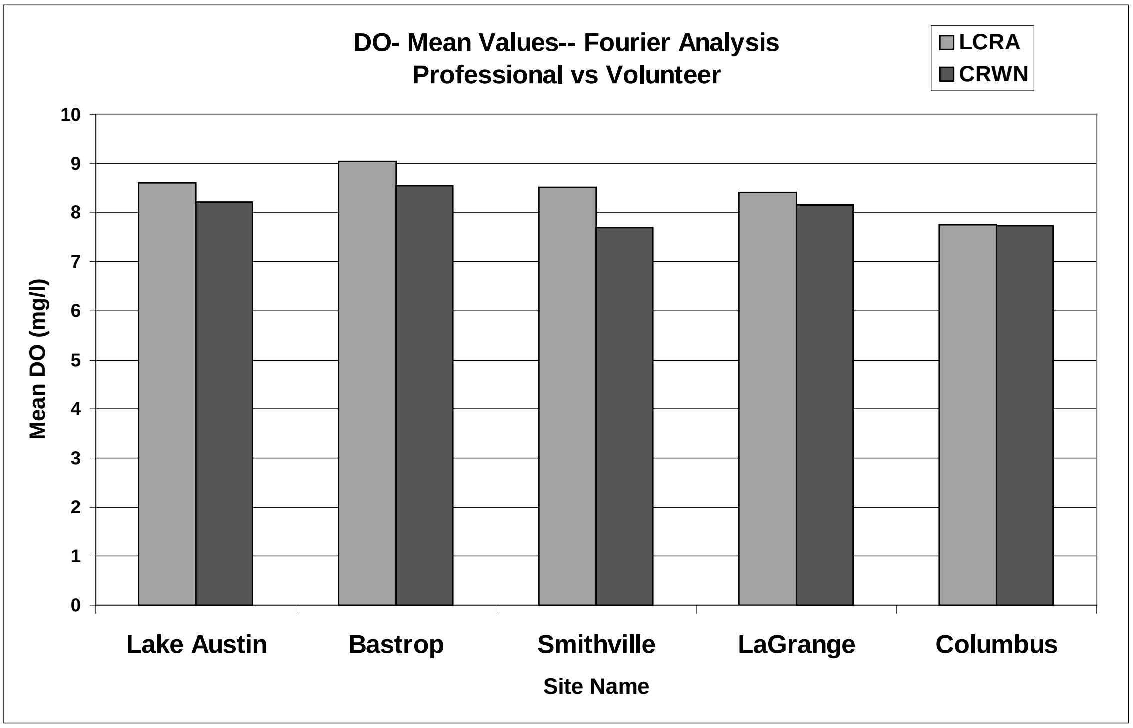


Figure 6-2 DO Mean Values, Fourier Method

There is no real difference between the professional and volunteer TDS mean values. This is illustrated in the graph, Figure 6-3, below.

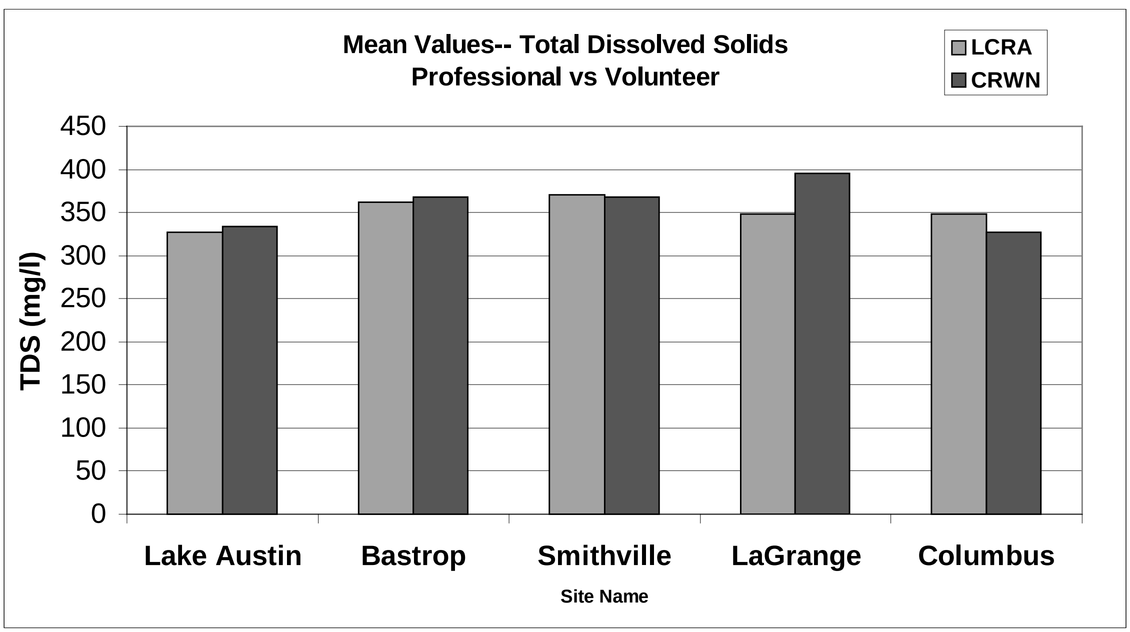


Figure 6-3 Mean Values, TDS

Figures 6-4 and 6-5 illustrate the differences in the mean values of the dissolved oxygen data sets using the standard method of calculating the mean versus the fourier method. The professional (LCRA) mean values, Figure 6-4, were essentially the same using both methods. The professional data set had a very regular temporal distribution over the time periods studied. This illustrates how the fourier method is a legitimate way to determine the mean of a data set.

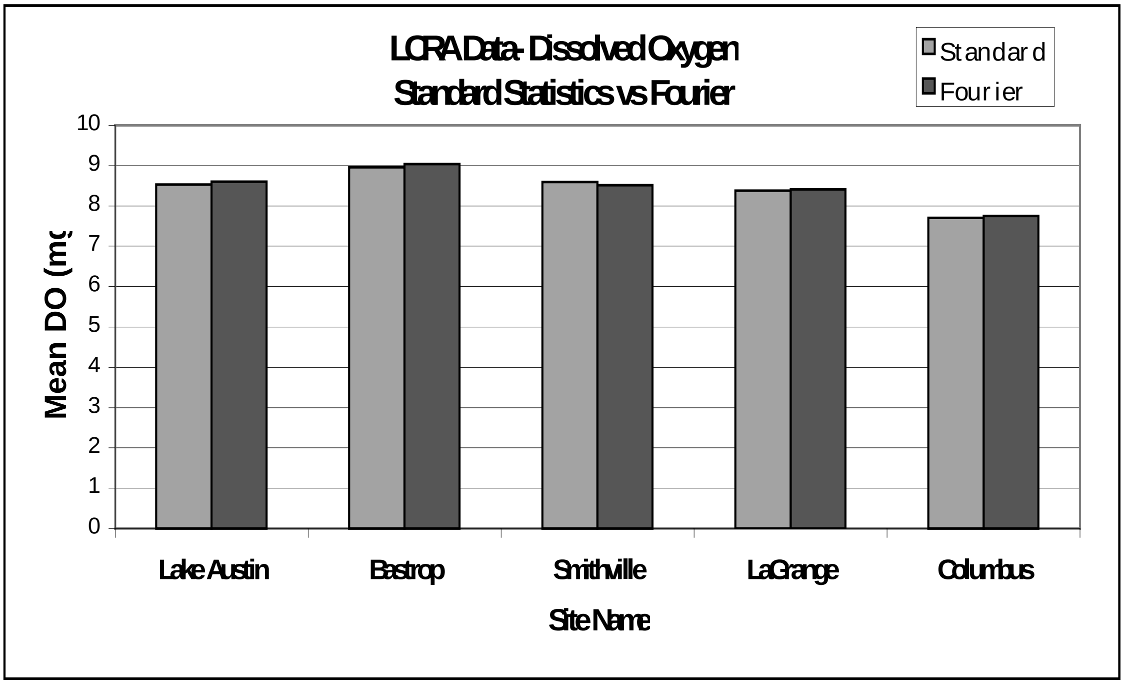


Figure 6-4 LCRA Mean Values, DO- Standard vs. Fourier

In contrast the volunteer dissolved oxygen mean values, Figure 6-5, had larger differences between the means calculated using the standard method and the means determined by the fourier method. The means determined by the fourier method was consistently lower for those sites which lacked data during the warmer time of the year when DO values would be lowest. This illustrates how the fourier method corrected for that lack of data.

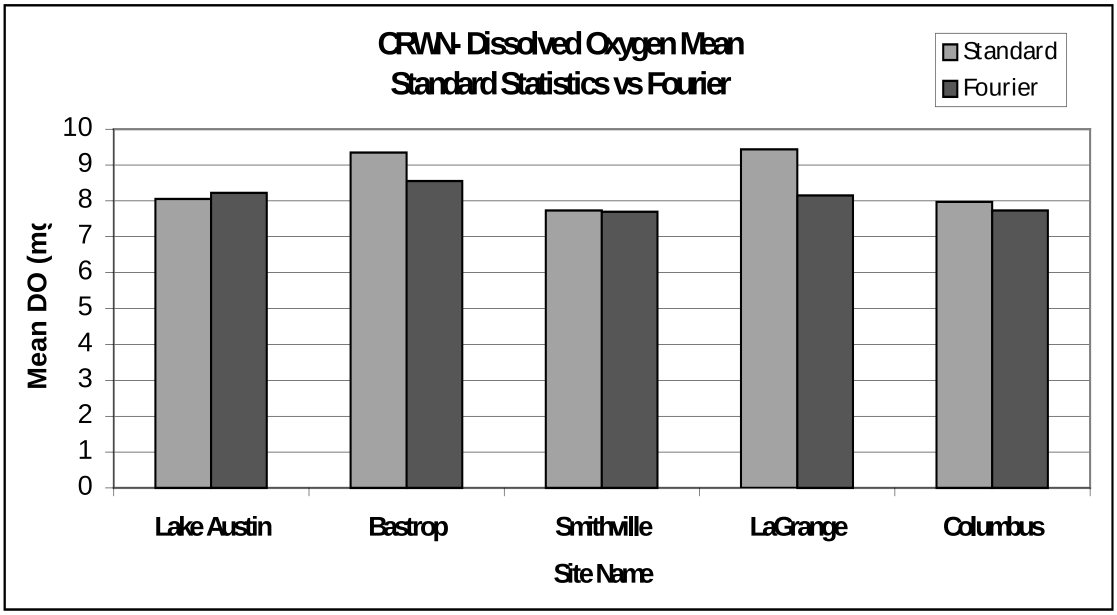


Figure 6-5 CRWN DO Mean Values, Standard vs. Fourier

This study indicates that there is good statistical agreement between the data collected by volunteer monitors and data collected by professional monitors. The main weakness of the volunteer monitoring is the lack of consistency temporally.

7 BIBLIOGRAPHY

Bartram, Jamie and Balance, Richard. *Water Quality Monitoring: A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*. E & FN Spon, London,1996.

Chapman, D, *Water Quality Assessments,* E & FN Spon, London, 1996.

Chow V.T., Maidment D.R. and Mays L.W., "Applied Hydrology", McGraw-Hill, Inc., 1988.

Colorado River Watch Network, *Technical Instructions*, Fifth Edition(Austin, Texas: May 1996)

Ely, Eleanor, ed. *Building Credibility* The Volunteer Monitor, Volume 9, No.1, Spring 1997, pp 19-20.

Ely, Eleanor, ed. *Managing and Presenting Your Data*. The Volunteer Monitor, Volume 7, number 1, Spring 1995. San Francisco, CA.

Helsel, D.R. and R.M. Hirsch, "Statistical Methods in Water Resources", Elsevier Science B.V. 1995

Hubbell, Steven. “Eyes on the Water: A Framework for Evaluating Volunteer Water Quality Monitoring Programs.” Master’s Report, The University of Texas at Austin, May, 1993.

Lower Colorado River Authority, *The Texas Clean Rivers Program, Technical Report*, (Austin, Texas: October 1, 1996)

Maidment, David, *Assignment #1: Statistical Analysis of Environmental Data*, CE 397: Environmental Risk Assessment, Department of Civil Engineering, The University of Texas at Austin

Maidment D., *Handbook of Hydrology* McGraw-Hill, Inc., 1993

Setzer,Woodrow. “Tips for Statistical Analyses of Parallel Studies", The Volunteer Monitor, Volume 9, No.1, Spring 1997, pp 19-20.

Setzer, Woodrow, Denise Stoeckel “Parallel Testing—Volunteers Versus Professionals",

Texas Natural Resource Conservation Commission, *Quality Assurance Project Plan,* Texas Watch (Austin, Texas: January 2, 1998)

US Environmental Protection Agency, Office of Water, "The Quality of Our Nation's Water" Website: http://www.epa.gov/305b/sum1.html#SEC3

U.S. Environmental Protection Agency, *National Water Quality Inventory: 1992 Report to Congress*, Office of Water (Washington, D.C.: Government Printing Office, April 1994)

U.S. Environmental Protection Agency, *Volunteer Water Monitoring: A Guide For State Managers*, Office of Water (Washington, D.C.: Government Printing Office, August 1990)

U.S. Environmental Protection Agency, Office of Water, Washington, DC. *EPA Requirements for Quality Assurance Project Plans (QAPP) for Environmental Data Operations.*, EPA/QA/R-5. August 1994. U.S.

U.S Environmental Protection Agency, Office of Research and Development, Washington, DC., *Guidance for Data Quality Assessment*. EPA QA/G-9, March 1995. U.S.

8 APPENDIX

Raw Data

Lake Austin Note: LCRA TDS values are converted from conductivity values

Dissolved Oxygen (mg/L)				Total Dissolved Solids (mg/L)			
Date	CRWN	Date	LCRA	Date	CRWN	Date	LCRA
04/19/93	4.8	04/19/93	8.9	04/19/93	410	4/19/93	317
06/21/93	6.2	06/01/93	8.9	06/21/93	140	6/1/93	323
06/25/93	7.8	08/11/93	8.2	06/25/93	250	8/11/93	332
07/09/93	7.8	10/07/93	7.5	07/09/93	220	10/7/93	317
07/16/93	8.3	02/14/94	10.4	07/16/93	100	2/14/94	335
07/23/93	6.7	04/27/94	9.2	07/23/93	340	4/27/94	333
07/30/93	6.7	06/23/94	9	07/30/93	340	6/23/94	331
08/06/93	7.7	08/29/94	7.7	08/06/93	350	8/29/94	342
08/13/93	6.7	10/18/94	6.4	08/13/93	370	10/18/94	275
08/27/93	7	12/22/94	9	08/27/93	370	12/22/94	338
09/03/93	6.7	02/15/95	12.8	09/03/93	360	2/15/95	374
09/24/93	6.5	04/18/95	8.2	09/24/93	340	4/18/95	364
10/08/93	6.8	06/07/95	9.6	10/08/93	320	6/7/95	350
10/15/93	7	08/17/95	7.9	10/15/93	350	8/17/95	337
10/29/93	8.3	10/25/95	7.9	10/29/93	340	10/25/95	305
11/05/93	8.8	12/14/95	9.3	11/05/93	340	12/14/95	323
11/12/93	8	02/26/96	9.5	11/12/93	320	2/26/96	325
11/19/93	8.2	04/16/96	9.4	11/19/93	400	4/16/96	320
12/13/93	9.6	06/20/96	8.5	12/13/93	400	6/20/96	323
12/16/93	8.6	08/14/96	7.7	12/16/93	340	8/14/96	320
12/17/93	9.4	10/21/96	7.4	12/17/93	350	10/21/96	355
12/30/93	9.2	12/17/96	9	12/30/93	350	12/17/96	352
01/02/94	9.7	03/18/97	10.9	01/02/94	350	3/18/97	344
01/14/94	10.8	04/17/97	8.8	01/14/94	330	4/17/97	313
02/04/94	10.2	06/19/97	7	02/04/94	290	6/19/97	314
02/11/94	10.8	08/27/97	4.8	02/11/94	320	8/27/97	282
02/18/94	7.8	10/13/97	6.4	02/18/94	350	10/13/97	285
03/04/94	8.7			03/04/94	370		
03/18/94	7.3			03/18/94	380		
03/25/94	9			03/25/94	360		
04/01/94	8			04/01/94	350		
04/15/94	6.75			04/15/94	340		
04/22/94	8.4			04/22/94	370		
04/29/94	8.8			04/29/94	360		
06/03/94	8.8			06/03/94	360		
06/10/94	9.1			06/10/94	380		
06/17/94	9.4			06/17/94	380		
06/24/94	11			06/24/94	360		
07/01/94	7.53			07/01/94	320		
07/15/94	5.7			07/15/94	360		
07/22/94	6.9			08/05/94	590		
08/05/94	6.6			08/12/94	420		
08/12/94	5.90			08/19/94	430		
08/19/94	4.95			09/16/94	420		
09/16/94	5.00			09/23/94	400		
09/23/94	6.00			09/30/94	430		
09/30/94	4.70			10/06/94	420		
10/06/94	4.60			10/14/94	330		
10/14/94	5			11/17/94	380		
11/17/94	9.20			12/09/94	420		
12/09/94	6.6			12/16/94	370		
12/16/94	10.60			01/13/95	410		
01/13/95	9.4			02/03/95	450		
02/03/95	6.5			02/20/95	420		
02/20/95	8.7			03/24/95	400		
03/24/95	7.5			03/31/95	400		
03/31/95	7.77			04/07/95	430		
04/28/95	8.7			04/21/95	550		
05/12/95	8.1			04/28/95	410		
05/19/95	7.2			05/12/95	290		
05/26/95	7.9			05/19/95	290		
06/02/95	9.6			05/26/95	310		
06/09/95	7.6			06/02/95	250		
06/16/95	8.1			06/09/95	300		
06/22/95	7.4			06/16/95	300		
07/21/95	7.4			06/22/95	300		
08/25/95	7.9			06/30/95	310		
09/08/95	7.5			07/21/95	320		
09/14/95	8			08/25/95	290		
09/28/95	6.3			09/08/95	290		

Bastrop **Note: LCRA TDS values are converted from conductivity values**

Dissolved Oxygen (mg/L)				Total Dissolved Solids (mg/L)			
Date	CRWN	Date	LCRA	Date	CRWN	Date	LCRA
09/28/93	7.00	08/23/93	7.10	10/12/93	390	10/20/93	353
10/05/93	9.00	10/20/93	7.20	10/19/93	420	12/21/93	408
10/12/93	8.00	12/21/93	10.70	10/27/93	370	02/15/94	409
10/19/93	7.60	02/15/94	11.70	11/02/93	300	04/13/94	352
10/27/93	10.00	04/13/94	11.90	11/16/93	530	06/06/94	345
11/02/93	12.00	06/06/94	7.80	11/23/93	360	08/02/94	358
11/16/93	8.90	08/02/94	7.50	11/30/93	330	10/13/94	319
11/23/93	10.00	10/13/94	7.60	12/07/93	360	12/13/94	422
11/30/93	12.20	12/13/94	11.30	12/14/93	420	02/22/95	434
12/07/93	9.20	02/22/95	13.30	12/22/93	490	04/25/95	367
12/14/93	11.00	04/25/95	8.20	12/29/93	390	06/05/95	358
12/22/93	8.50	06/05/95	9.00	01/04/94	340	08/15/95	365
12/29/93	10.00	08/15/95	7.00	01/12/94	390	10/16/95	355
01/04/94	13.40	10/16/95	9.50	01/19/94	400	12/19/95	421
01/12/94	10.50	12/19/95	8.90	01/25/94	410	02/20/96	418
01/19/94	11.30	02/20/96	9.90	02/01/94	430	04/08/96	349
01/25/94	8.80	04/08/96	8.90	02/08/94	370	06/10/96	318
02/01/94	10.00	06/10/96	7.20	02/15/94	360	08/01/96	370
02/08/94	10.00	08/01/96	7.10	02/22/94	350	10/01/96	371
02/15/94	11.00	10/01/96	7.70	03/01/94	370	12/02/96	368
02/22/94	8.80	12/02/96	9.80	03/08/94	430	02/03/97	439
03/01/94	9.50	02/03/97	10.20	03/15/94	370	04/07/97	360
03/08/94	8.00	04/01/97	9.10	03/29/94	350	06/03/97	313
03/15/94	10.00	06/03/97	8.40	04/06/94	130	08/20/97	202
03/29/94	12.50	08/20/97	7.60	04/12/94	290	10/01/97	294
04/06/94	10.50	10/01/97	8.30	04/19/94	320		
04/12/94	9.40			04/26/94	420		
04/19/94	7.30			05/03/94	380		
04/26/94	8.50			05/17/94	330		
05/03/94	8.20			05/24/94	390		
05/10/94	6.00			10/04/94	370		
05/17/94	6.80			10/11/94	320		
05/24/94	6.00			10/18/94	120		
09/14/94	6.40			10/25/94	410		
10/04/94	9.10			11/01/94	360		
10/11/94	7.80			11/08/94	250		
10/18/94	6.30			11/15/94	430		
10/25/94	6.00			11/22/94	430		
11/01/94	7.20			12/06/94	430		
11/08/94	6.80			12/13/94	430		
11/15/94	8.00			12/20/94	340		
11/22/94	8.10			12/27/94	460		
12/06/94	4.10			01/03/95	330		
12/13/94	9.90			01/10/95	380		
12/20/94	7.30			01/24/95	420		
12/27/94	9.30			01/31/95	390		
01/03/95	9.80			02/07/95	489		
01/10/95	10.10			02/14/95	430		
01/18/95	9.30			02/21/95	420		
01/24/95	11.20			02/28/95	460		
02/07/95	11.50			03/07/95	460		
02/14/95	11.85			03/14/95	220		
02/21/95	14.80			09/13/95	320		
02/28/95	7.60			09/20/95	300		
03/07/95	8.85			10/04/95	300		
03/14/95	7.55			10/11/95	290		
03/21/95	6.80			11/15/95	300		
09/13/95	7.60			12/29/95	290		
09/20/95	6.90			01/03/96	350		
10/04/95	7.60			01/10/96	320		
10/11/95	10.50			01/24/96	310		
11/15/95	9.20			01/30/96	315		
12/29/95	9.90			02/13/96	280		
01/03/96	12.70			02/20/96	280		
01/10/96	13.30			02/27/96	340		
01/24/96	12.60			03/05/96	310		
01/30/96	10.50			09/05/96	420		
02/06/96	11.00			09/13/96	430		
02/13/96	12.70			09/13/96	430		
02/20/96	12.60			09/17/96	490		

Smithville Note: LCRA TDS values are converted from conductivity values

Dissolved Oxygen (mg/L)				Total Dissolved Solids (mg/L)			
Date	CRWN	Date	LCRA	Date	CRWN	Date	LCRA
09/09/93	6.00	8/23/93	6.90	09/09/93	400	8/23/93	343
10/01/93	7.80	10/20/93	7.50	10/01/93	380	10/20/93	347
10/07/93	2.00	12/21/93	10.10	10/07/93	360	12/21/93	416
10/15/93	7.50	2/15/94	12.10	10/15/93	380	2/15/94	373
10/21/93	2.00	4/13/94	9.70	10/21/93	370	4/13/94	346
11/04/93	4.00	6/6/94	7.00	11/04/93	410	6/6/94	347
11/12/93	9.00	8/2/94	6.90	11/12/93	420	8/2/94	359
11/18/93	3.00	10/13/94	7.60	11/18/93	420	10/13/94	323
12/09/93	7.80	12/13/94	10.50	12/03/93	390	12/13/94	412
01/13/94	9.00	2/22/95	11.30	12/09/93	420	2/22/95	431
01/26/94	6.90	4/26/95	7.80	01/13/94	430	4/26/95	370
12/08/94	7.80	6/5/95	9.00	01/26/94	420	6/5/95	350
12/08/94	7.30	8/15/95	7.00	12/08/94	320	8/15/95	358
01/20/95	8.60	10/16/95	9.40	12/08/94	350	10/16/95	357
03/03/95	9.90	12/19/95	8.40	01/20/95	280	12/19/95	394
03/30/95	9.60	2/20/96	9.10	03/03/95	910	2/20/96	394
04/20/95	8.15	4/8/96	8.40	03/30/95	1130	4/8/96	358
04/28/95	8.10	6/10/96	7.30	04/20/95	350	6/10/96	298
09/20/95	6.30	8/1/96	5.90	04/28/95	380	8/1/96	365
10/05/95	7.00	10/1/96	7.80	09/20/95	340	10/1/96	374
11/08/95	8.90	12/2/96	9.30	10/05/95	320	12/2/96	397
11/17/95	8.50	2/3/97	9.70	11/08/95	310	2/3/97	443
12/07/95	5.60	4/1/97	8.80	11/17/95	330	4/1/97	369
01/19/96	9.90			12/07/95	450		
02/08/96	10.15			01/19/96	340		
02/15/96	10.10			01/26/96	350		
03/08/96	9.80			02/08/96	300		
03/22/96	9.40			02/15/96	350		
09/28/96	8.45			03/08/96	380		
10/18/96	8.30			03/22/96	480		
10/24/96	6.00			09/28/96	440		
11/22/96	7.10			10/18/96	440		
11/26/96	6.40			10/24/96	410		
12/05/96	5.40			11/22/96	440		
01/21/97	10.20			11/26/96	470		
01/31/97	11.50			12/05/96	320		
02/14/97	9.60			02/14/97	180		
02/27/97	9.20			02/27/97	300		
03/20/97	9.20			03/20/97	300		
				03/26/97	300		

LaGrange Note: LCRA TDS values are converted from conductivity values

Dissolved Oxygen (mg/L)				Total Dissolved Solids (mg/L)			
Date	CRWN	Date	LCRA	Date	CRWN	Date	LCRA
10/19/95	9.2	10/17/95	8.8	10/19/95	420	10/17/95	361
10/25/95	8.3	12/20/95	8.9	10/25/95	420	12/20/95	238
11/03/95	8.2	02/21/96	9.4	11/03/95	510	02/21/96	410
11/29/95	10.2	04/09/96	8.3	11/15/95	410	04/09/96	355
12/06/95	11.7	06/25/96	7.3	01/17/96	940	06/25/96	337
01/17/96	12.1	08/02/96	8.4	01/31/96	970	08/02/96	362
01/24/96	12.9	10/02/96	7.4	02/07/96	970	10/02/96	370
02/07/96	14.1	12/03/96	9.4	02/21/96	360	12/03/96	377
02/21/96	10.5	02/04/97	9.9	03/01/96	340	02/04/97	432
03/01/96	12.0	04/02/97	8.7	03/07/96	350	04/02/97	359
03/07/96	10.5	06/04/97	7.8	05/22/96	300	06/04/97	306
05/22/96	7.0	08/21/97	7.9	06/08/96	260	08/21/97	286
06/08/96	7.6	10/02/97	6.8	07/07/96	300	10/02/97	330
07/07/96	7.8			11/13/96	320		
11/13/96	8.1			11/13/96	330		
11/13/96	8.9			11/19/96	330		
11/19/96	6.8			11/19/96	320		
11/19/96	7.6			12/03/96	210		
11/26/96	9.5			12/03/96	310		
12/03/96	8.6			12/10/96	320		
12/03/96	9.8			12/10/96	310		
12/10/96	9.7			01/21/97	220		
12/10/96	9.6			02/04/97	380		
01/21/97	8.8			02/04/97	380		
02/04/97	9.9			02/25/97	320		
02/25/97	9.2			02/25/97	310		
03/05/97	7.8			03/05/97	790		
03/11/97	9.1			03/11/97	290		
03/18/97	9.7			03/18/97	260		
03/25/97	8.2			03/25/97	330		
04/22/97	7.1			10/14/97	120		
04/29/97	8.1			10/28/97	270		
10/14/97	5.2			11/04/97	360		

Columbus Note: LCRA TDS values are converted from conductivity values

Dissolved Oxygen (mg/L)				Total Dissolved Solids (mg/L)			
Date	CRWN	Date	LCRA	Date	CRWN	Date	LCRA
04/28/93	7.10	08/24/93	7.00	09/27/93	380	08/24/93	344
09/01/93	7.20	10/21/93	6.90	04/28/93	330	10/21/93	232
09/20/93	7.40	12/08/93	8.90	09/01/93	468	12/08/93	404
09/27/93	8.00	02/16/94	10.00	09/20/93	390	02/16/94	401
10/04/93	6.00	04/14/94	7.20	10/04/93	380	04/14/94	370
10/19/93	7.00	06/07/94	6.60	10/19/93	400	06/07/94	346
11/15/93	9.50	08/03/94	7.10	11/15/93	260	08/03/94	365
12/14/93	7.20	10/13/94	7.80	12/14/93	420	10/13/94	257
01/13/94	8.20	12/14/94	9.00	01/13/94	400	12/14/94	424
01/25/94	8.00	02/23/95	9.40	01/25/94	410	02/23/95	431
03/31/94	8.40	04/26/95	7.90	03/31/94	450	04/26/95	367
04/05/94	6.60	06/06/95	8.30	04/05/94	430	06/06/95	337
04/12/94	8.00	08/16/95	7.00	04/12/94	400	08/16/95	355
04/19/94	8.50	10/17/95	8.40	04/19/94	360	10/17/95	367
04/26/94	6.80	12/20/95	7.50	04/26/94	360	12/20/95	121
05/23/94	6.05	02/21/96	7.30	05/23/94	360	02/21/96	416
06/28/94	6.50	04/09/96	8.10	06/28/94	430	04/09/96	355
08/24/94	7.80	06/25/96	5.30	08/24/94	390	06/25/96	335
09/02/94	8.40	08/02/96	6.30	09/02/94	420	08/02/96	366
09/29/94	8.20	10/02/96	7.30	09/29/94	420	10/02/96	355
11/17/94	8.40	12/03/96	8.90	11/17/94	400	12/03/96	400
01/05/95	8.10	02/04/97	8.50	03/08/95	320	02/04/97	428
01/19/95	7.80	04/02/97	8.60	11/01/95	320	04/02/97	354
01/26/95	9.60	06/04/97	7.60	11/09/95	260	06/04/97	307
02/02/95	8.85	08/21/97	6.80	11/15/95	290	08/21/97	274
03/08/95	8.86	10/02/97	6.40	11/30/95	290	10/02/97	341
11/01/95	7.85			12/01/95	208		
11/09/95	8.80			12/13/95	320		
11/15/95	16.15			01/01/96	320		
11/30/95	10.80			01/11/96	340		
12/01/95	7.10			03/14/96	300		
12/13/95	7.90			03/21/96	300		
01/01/96	5.33			03/28/96	290		
01/11/96	9.86			09/03/96	180		
03/14/96	10.25			09/09/96	240		
03/21/96	13.40			09/12/96	270		
03/28/96	9.20			09/18/96	260		
09/03/96	5.70			09/20/96	200		
09/09/96	6.70			09/24/96	190		
09/12/96	6.80			09/30/96	350		
09/18/96	6.35			10/08/96	440		
09/20/96	4.95			10/21/96	470		
09/24/96	4.30			10/29/96	490		
09/30/96	6.80			11/14/96	490		
10/08/96	7.60			11/26/96	380		
10/21/96	7.95			12/11/96	390		
10/29/96	6.00			01/11/97	140		
11/14/96	9.35			02/13/97	210		
11/26/96	8.55			02/21/97	140		
12/11/96	7.80			03/07/97	280		
01/11/97	9.95			03/20/97	270		
02/13/97	8.70			04/10/97	310		
02/21/97	6.20			04/17/97	260		
03/07/97	7.90			04/21/97	290		
04/10/97	8.00			05/15/97	260		
04/17/97	8.20			05/20/97	290		
04/21/97	7.30			11/05/97	330		
05/15/97	6.60			11/19/97	260		
05/20/97	8.00			12/03/97	120		
11/05/97	8.10			12/10/97	270		
11/19/97	9.65						
12/03/97	8.10						
12/10/97	8.00						