
Part 615

**Analysis of Water
Quality Monitoring
Data**

Issued February 2002

DISCLAIMER: Trade names are used solely to provide specific information. Mention of a trade name does not constitute a guarantee of the products by the U.S. Department of Agriculture nor does it imply endorsement by the Department or the Natural Resources Conservation Service over comparable products that are not named.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326W, Whitten Building, 14th and Independence Avenue, SW, Washington, DC 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.

Preface

Purpose

The purpose of part 615 of the National Water Quality Handbook (NWQH) is to provide guidance in the statistical analysis of water quality data that have been collected according to the designs described in part 614. Part 615 is concerned with the statistical analysis of monitoring results.

Acknowledgments

Part 615 of the National Water Quality Handbook was written by **John C. Clausen**, Ph.D., College of Agriculture and Natural Resources, University of Connecticut. The concept for this project was developed by **James N. Krider**, former national environmental engineer, Natural Resources Conservation Service (NRCS), Washington, DC. Technical leadership was provided by **Bruce J. Newton**, limnologist, NRCS National Water and Climate Center, Portland, Oregon.

The following people contributed to the review of this document:

Richard H. McCuen, University of Maryland, College Park, Maryland

Carl DuPoldt, agricultural engineer, NRCS, Somerset, New Jersey

Barbara M. Vining, pesticide and nutrient specialist, NRCS, Indianapolis, Indiana

Howard Thomas, economist, NRCS, Portland, Oregon

Robert C. Feurer, Philadelphia Suburban Water Company, Bryn Bawr, Pennsylvania

Lynn Sampson, biologist, NRCS, East Lansing, Michigan

William H. Boyd, environmental engineer, NRCS National Water Management Center, Little Rock, Arkansas

James C. Wood, water quality specialist, NRCS, Boise, Idaho

Richard Croft, water quality specialist, NRCS Watershed Science Institute, Burlington, Vermont

Jorge A. Delgado, USDA-Agricultural Research Service

Ron Schierer, NRCS Northern Plains Regional Technical Team, Lakewood, Colorado

Joseph A. Neafsey, water quality specialist, Storrs, Connecticut, was the contract project officer. **Mary R. Mattinson**, editor, and **Wendy R.**

Pierce, illustrator, NRCS National Production Services, Fort Worth, Texas, provided editorial and graphic support.

Part 615

Analysis of Water Quality Monitoring Data

Contents:	Subpart 615.00	Introduction
	Subpart 615.01	Basic Statistics
	Subpart 615.02	Exploratory Data Analysis
	Subpart 615.03	Statistical Assumptions
	Subpart 615.04	Causality
	Subpart 615.05	Hypothesis Testing
	Subpart 615.06	Plot Designs
	Subpart 615.07	Single Watershed
	Subpart 615.08	Above and Below Watersheds
	Subpart 615.09	Paired Watersheds
	Subpart 615.10	Multiple Watersheds
	Subpart 615.11	Trend Analysis
	Subpart 615.12	Statistical Packages
	Appendixes	
	Glossary	
	Index	

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615

National Water Quality Handbook



Subpart 615.00 Introduction

Subpart 615.00 Introduction

Contents:	615.0000	General	00-1
	615.0001	Steps in statistical analysis	00-1
	615.0002	References	00-4

Tables	Table 00-1	Summary of statistical procedures use in Part 615, by subpart	00-2
	Table 00-2	Summary of purpose of statistical procedures used in Part 615	00-3

Figure	Figure 00-1	Steps in data analysis for a water quality monitoring study	00-1
---------------	--------------------	---	------

615.0000 General

In National Water Quality Handbook (NWQH), part 614, the 12 steps for designing a water quality monitoring study were described. The overall purpose of part 615 is to provide assistance in how to analyze water quality data that have been collected according to the designs described in part 614. It is not the intention that part 615 replace a basic course or textbook on statistics; actually the reader would be much better prepared for this part of the handbook having had such a course.

Subparts 615.01 to 615.04 provide background information on statistical analysis; subparts 615.05 to 615.10 provide guidance on how to analyze data obtained from particular monitoring designs; and subpart 615.11 describes information on several available computer packages for statistics. The subparts include several examples that use both hand calculations and computer-generated output. Many computerized statistical packages are available today, and to save time and effort, the user is encouraged to invest in a package. Subpart 615.12 provides guidance on how to select statistical analysis software.

The Statistical Analysis System (SAS) software for a PC is used for illustration purposes throughout part 615 of the NWQH.

Table 00–1 summarizes the statistical procedures used in part 615 and indicates the subpart where that procedure is best described. Table 00–2 summarizes the purpose of the various statistics and statistical tests used in part 615 of the handbook.

615.0001 Steps in statistical analysis

As in part 614, there are several steps in conducting the statistical analysis of water quality data (fig. 00–1). The analysis of data begins with Exploratory data analysis (EDA), which is intended for the analyst to become familiar with the data (Tukey 1977). The next step is to test the appropriate assumptions for the statistical tests to be performed. The assumptions may include randomness, the type of distribution, the homogeneity of variances, and independence. The next step is to determine the appropriate hypotheses to test. This step may have already been completed as part of designing the study. The next step would be to conduct the actual statistical tests. Finally, the conclusions regarding the data are constructed. The following subparts are intended to assist the analyst through these steps of data analysis.

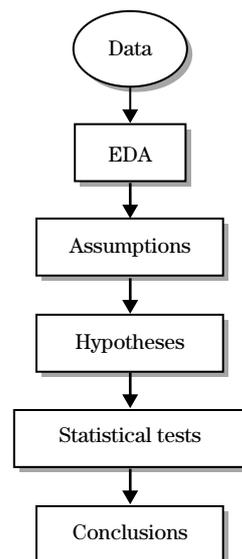
Figure 00–1 Steps in data analysis for a water quality monitoring study

Table 00-1 Summary of statistical procedures used in Part 615, by subpart

Procedure	-----Subpart-----										
	01	02	03	04	05	06	07	08	09	10	11
Basic statistics:											
Mean	X		X								
Median	X		X								
Mode	X										
Variance	X										
Standard deviation	X										
Standard error	X					X					
Coefficient of variation	X										
Coefficient of skewness	X		X								
Kurtosis			X								
Shapiro-Wilk W-statistic			X								
Autocorrelation coefficient			X								
Statistical tests:											
t-test							X	X			
Mann-Whitney U (nonparametric)							X				
Wilcoxon paired sample (nonpar)								X			
F ratio			X							X	
Analysis of variance						X				X	X
one-way						X					
Kruskal-Wallis one-way (nonpar)											
two-way						X					
Tukey's multiple comparisons						X					X
Regression									X		X
Coefficient of determination									X		
Confidence intervals									X		
Analysis of covariance									X		
Kendall tau											X

Table 00-2 Summary of purpose of statistical procedures used in Part 615

Procedure	Purpose
Basic statistics:	
Mean	measure of central tendency
Median	measure of central tendency
Mode	measure of central tendency
Variance	measure of dispersion of a random variable
Standard deviation	measure of dispersion
Standard error	measure of dispersion of a statistic
Coefficient of variation	standardized measure of dispersion
Coefficient of skewness	measure of symmetry
Kurtosis	measure of long tailedness (peakedness) of dispersion
Shapiro-Wilk W-statistic	test for normality
Autocorrelation coefficient	measure of independence of observations on a single random variable
Statistical tests:	
one-sample t-test	comparison of a single mean to a standard
two-sample t-test	comparison of two sample means
Mann-Whitney U (nonparametric)	nonparametric comparison of unpaired two-sample ranks
Wilcoxon paired sample (nonpar)	nonparametric comparison of paired ranks of differences
F ratio	test of homogeneity of variances
Analysis of variance	comparison of several means
one-way	comparison of several means for one factor
Kruskal-Wallis one-way (nonpar)	comparison of several means for one factor, nonparametric
two-way	comparison of several means for two factors
Tukey's multiple comparisons	determine which means are different for a rejected ANOVA test
Regression	relationship between two variables
Coefficient of determination	fraction of variation explained by relationship
Confidence intervals	measure of accuracy of a statistic
Analysis of covariance	comparison of regression slopes and intercepts among groups
Kendall tau	nonparametric measure of correlation for trend detection

615.0002 **References**

Tukey, J.W. 1977. Exploratory data analysis. Addison-Wesley Publ. Co., Reading, MA.

United States Department of Agriculture, Natural Resources Conservation Service. 1996. Design of water quality monitoring systems. National Water Quality Handbook, Part 514.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.01 Basic Statistics

Subpart 615.01 Basic Statistics

Contents:	615.0100	Introduction	01-1
	615.0101	Purpose of statistics	01-1
	615.0102	Statistical terms	01-2
	615.0103	Data types	01-2
	615.0104	Frequencies	01-3
	615.0105	Measures of central tendency	01-6
		(a) Mean	01-6
		(b) Median	01-7
	(c) Mode	01-7	
615.0106	Measures of dispersion	01-8	
615.0107	References	01-11	

Tables	Table 01-1	Raw algae counts from St. Albans Bay, Vermont, 1985	01-4
	Table 01-2	Frequency table of algal counts in St. Albans Bay, Vermont	01-4
	Table 01-3	Measures of central tendency for data types	01-6

Figures	Figure 01-1	Frequency histogram of algal counts in St. Albans Bay, Vermont	01-4
	Figure 01-2	Cumulative frequency of algal counts in St. Albans Bay, Vermont	01-5
	Figure 01-3	Normal distribution	01-8
	Figure 01-4	Distributions showing skewness and kurtosis	01-9
Examples	Example 01-1	Measures of central tendency	01-6
	Example 01-2	Measures of dispersion	01-10

615.0100 Introduction

The understanding of basic statistics is important to the analysis of water quality data. For many, subpart 615.01 is a review of some of the foundations of statistics. Included in this subpart is the purpose of statistics, some statistical terms, definitions of data types, frequencies, measures of central tendency, and measures of dispersion.

615.0101 Purpose of statistics

In water quality monitoring, the use of statistics is important. For example, if our measurement of the quality of water averaged three this year and six next year, has the water quality really doubled in a year? In other words, is the number three different from the six and how confident can I be that they are or are not different?

Almost all water quality data collected are a sample. That is, we sample a certain portion of the entire population of water quality data available. For example, if we sample a well weekly from 2003 to 2008 for nitrate-N, that also means that we are not sampling the well during all other times. Assuming it takes at most 30 minutes to sample a well, we are sampling only 0.3 percent of the time during the week. We also are sampling between 2003 and 2008. We are not sampling before 2003 nor after 2008, which are times that also may be part of the entire population of water quality data. Therefore, the real purpose of statistics is to be able to make conclusions from a sampling of data for the entire population. Because we usually cannot measure the entire population, a sample is necessary. Statistics provide a systematic framework for analysis and summarization of the sample data.

615.0102 Statistical terms

A number of statistical terms used throughout this subpart are defined in this section.

Observation—A record representing a characteristic of a real-world object (EPA 1973). The record is generally a single number; for example, a chemical concentration or the number of macroinvertebrates found in a sample. The observation is the data you collect.

Population—The population is all possible values of a variable and is synonymous with universe (Steel and Torrie 1960).

Sample—A part of the population that should be representative of the population (Steele and Torrie 1960). A sample is a set of observations from the population.

Random sample—A sample that has an equal chance of being selected (Snedecor and Cochran 1980). Usually such a sample is collected to eliminate bias in the data.

615.0103 Data types

The two types of random variables that can be collected in water quality monitoring projects are continuous and discrete. The type of data selected influences the statistics applied and depends on the type of information being collected.

Continuous data means that all values within some range are possible (Steel and Torrie 1960). An example of continuous data would be concentrations. A nearly infinite number of values are possible within some range. More values become possible as detection equipment becomes more precise.

Discrete data means that the possible values can be only a certain set of numbers (Snedecor and Cochran, 1980). Examples include counts, categories, and binary data. The number of fish collected would be discrete data.

In addition to the continuous and discrete data, several scales can be used to measure water quality data. They include nominal, ordinal, interval, and ratio scales.

Nominal data include categories without ranking among the categories. The term nominal means that the category is called a name. Often, nominal data are binary, such as presence or absence. An example of nominal data would be taxa of macroinvertebrates present in a stream.

Ordinal data imply ordering (Ward et al. 1990). Ordinal variables measure the degree of something (Horowitz 1981). Trophic status—oligotrophic, mesotrophic, and eutrophic—is an example of an ordinal scale. However, the differences among the categories do not have to be equal.

Interval data also use ordering, but intervals between the categories are equal. Intervals or categories are used to describe the data. Interval data are used for data sets that do not have a true zero. For example, the intervals for temperature could be <25, 25–50, 50–75, and >75 degrees Fahrenheit. Intervals also are used to describe size classes of fish, such as <10, 10–20, 20–30, and >30.

Ratio data are similar to interval data except that a true zero exists. Therefore, 500 is 5 times greater than 100. Concentration and flow data are ratio data.

615.0104 Frequencies

Water quality data can be presented in many ways. They include tables of raw data or frequencies, seasonal tables, and graphical pie charts or frequency diagrams. A raw data table is given in table 01-1 for algal counts in St. Albans Bay, Lake Champlain, Vermont.

This raw data can be summarized in a frequency table by establishing intervals in the data. For example, the raw algal data in table 01-1 were grouped into intervals of 2,500 organisms per milliliter and are summarized in table 01-2. The *frequency* is the number of observations for that class interval.

The frequency table can also be displayed as a frequency histogram. A histogram graphs frequency as a function of class intervals as rectangles on a graph (fig. 01-1).

Such data may also be presented as a cumulative frequency histogram. The *cumulative frequency* is the summation of all the frequencies up to and including the class interval plotted (fig. 01-2). The points are joined with a line forming a cumulative frequency polygon (Zar 1996).

The frequency histogram and the cumulative frequency polygon can be converted to relative frequency. This is done by changing the Y-axis to either a decimal or percentage scale by dividing the frequencies by the total sample size.

Frequency plots have several values, including:

- help assess the distribution type
- detect characteristics of the data (e.g., central tendency, dispersion)
- identify potential outliers
- assess the range of data

Although these forms of data presentation are useful, there are other ways to describe the data. They include describing a measure of central tendency and a measure of dispersion of the data.

Table 01-1 Raw algal counts (organisms/mL) from St. Albans Bay, Vermont, 1985

Date	Count	Date	Count
1/23	25	8/6	1,564
3/19	125	8/13	6,384
4/23	410	8/20	10,062
5/14	1,883	8/27	6,305
5/30	770	9/4	39,861
6/11	2,229	9/10	6,755
6/18	519	9/17	15,074
6/25	899	9/25	36,823
7/2	882	10/1	29,448
7/9	565	10/8	45,283
7/16	826	10/15	1,336
7/23	299	11/5	1,000
7/30	547	12/4	56

Table 01-2 Frequency table of algal counts in St. Albans Bay, Vermont

Interval	Frequency	Interval	Frequency
0 – 2,500	17	25,000 – 27,500	0
2,500 – 5,000	0	27,500 – 30,000	1
5,000 – 7,500	3	30,000 – 32,500	0
7,500 – 10,000	0	32,500 – 35,000	0
10,000 – 12,500	1	35,000 – 37,500	1
12,500 – 15,000	0	37,500 – 40,000	1
15,000 – 17,500	1	40,000 – 42,500	0
17,500 – 20,000	0	42,500 – 45,000	0
20,000 – 22,500	0	45,000 – 47,500	1
22,500 – 25,000	0		

Figure 01-1 Frequency histogram of algal counts in St. Albans Bay, Vermont

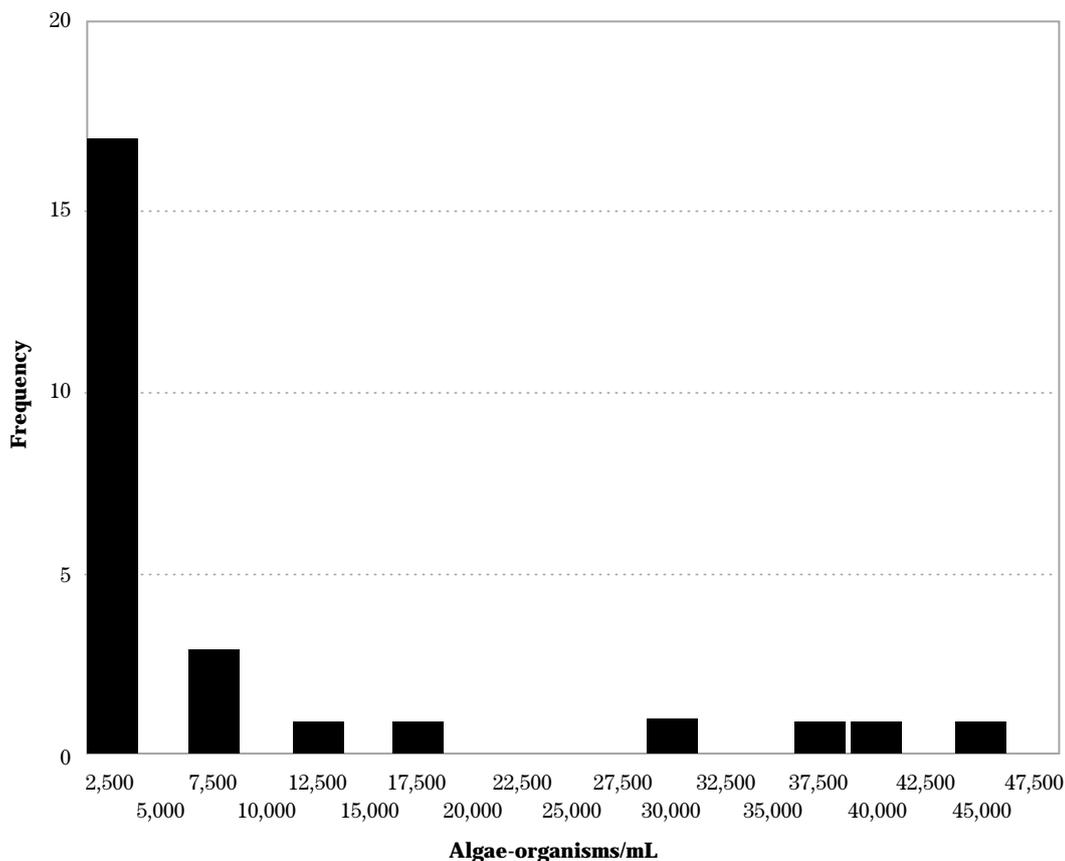
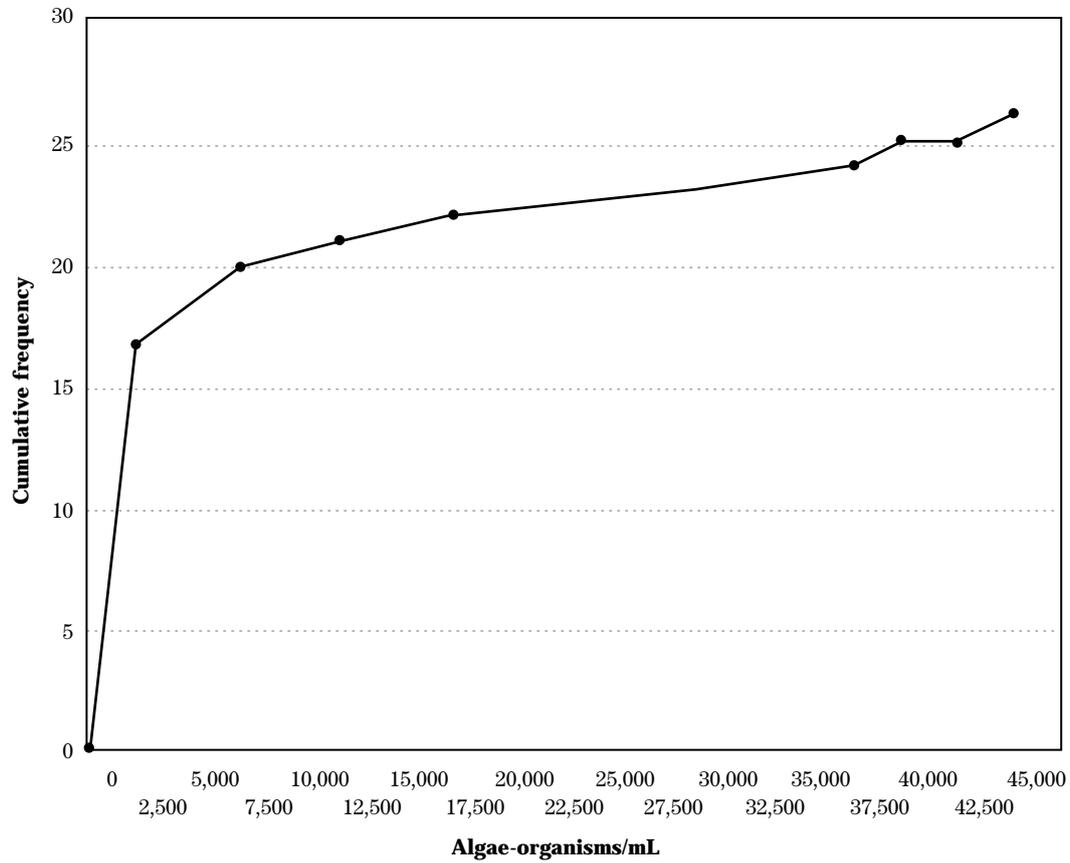


Figure 01-2 Cumulative frequency of algal counts in St. Albans Bay, Vermont



615.0105 Measures of central tendency

Several measures of central tendency for a data set are available. The appropriate measure varies with the type of data (table 01-3). Example 01-1 illustrates the different measures of central tendency.

Table 01-3 Measures of central tendency for data types

Scale	Measure	Example
nominal	mode	taxa
ordinal	median	trophic state
interval	mean	fish age class
ratio	mean	concentrations

(a) Mean

The most commonly used measure is the arithmetic mean or average. The mean (\bar{X}) is the sum of the observations ($\sum X_i$) divided by the number of observations (n):

$$\bar{X} = \frac{\sum X_i}{n} \quad [01-1]$$

The mean is appropriate for interval and ratio data, but not nominal or ordinal types of information. Arithmetic means may not be the best measure of central tendency when distributions are skewed (long tail) left or right. If the data are censored, that is, there are observations below detection limits, the calculation of the mean is more rigorous. The mean for a censored distribution can be calculated from (Newman et al. 1989):

$$\bar{X} = \bar{X} - \sigma \frac{k}{n-k} \frac{f(\epsilon)}{F(\epsilon)} \quad [01-2]$$

Example 01-1 Measures of central tendency

Given: The algal count data from St. Albans Bay in table 01-1.

Determine: The mean, geometric mean, median, and mode.

Solution:

Mean:
$$\bar{X} = \frac{25 + 125 + 410 + \dots + 56}{26} = 8,074 \text{ organisms / mL}$$

Geometric mean:
$$\bar{X}_G = \text{antilog} \frac{\sum_{i=1}^n \log X_i}{n} = \text{antilog} 3.2343 = 1,715 \text{ organisms / mL}$$

Median—Because the data contain an even number of data values ($n=26$), the median is the mean of the two middle values.

$$\text{Median} = \frac{1,336 + 1,000}{2} = 1,168 \text{ organisms / mL}$$

Mode—No value occurred more than once in table 01-1; therefore, the mode does not exist for this data set.

where:

- n = total number of observations
- k = number of observations below the detection limit
- \bar{X} = mean of all the values above the detection limit
- σ = standard deviation
- $f(\varepsilon)$ = distribution function for the normal distribution
- $F(\varepsilon)$ = cumulative distribution function for the normal distribution

ε is obtained from:

$$\varepsilon = \frac{DL - \hat{\mu}}{\sigma} \quad [01-3]$$

where:

- DL = detection limit
- $\hat{\mu}$ = mean

In water quality data a geometric mean is often calculated. The geometric mean \bar{X}_G is the n^{th} root of the product of n values (Landwehr 1978, Zar 1996):

$$\bar{X}_G = \sqrt[n]{X_1 X_2 \dots X_n} \quad [01-4]$$

The geometric mean is also obtained as the antilog of the mean of the log of the values, which is the typical manner of calculating the geometric mean:

$$\bar{X}_G = \text{antilog} \frac{\sum_{i=1}^n \log X_i}{n} \quad [01-5]$$

The geometric mean is only used when all the values are positive and is typically used as the measure of central tendency for log transformed data.

(b) Median

A second measure of central tendency is the median (\bar{X}_m). The median is the value for which 50 percent of observations are greater and 50 percent are lesser. It is the midpoint of a frequency distribution. The median is an appropriate measure of centrality for ordinal data and is often used when the data are highly skewed. If a distribution is symmetrical, then the mean and the median will be the same.

$$\bar{X}_m = \left\{ \begin{array}{ll} \frac{X_{(n+1)}}{2} & \text{if } n \text{ is odd} \\ \left(\frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \right) & \text{if } n \text{ is even} \end{array} \right\} \quad [01-6]$$

(c) Mode

The mode is the final measure of central tendency. It is the value that occurs most frequently. The mode is the only appropriate measure of central tendency for nominal data and quickly describes the most commonly occurring value.

615.0106 Measures of dispersion

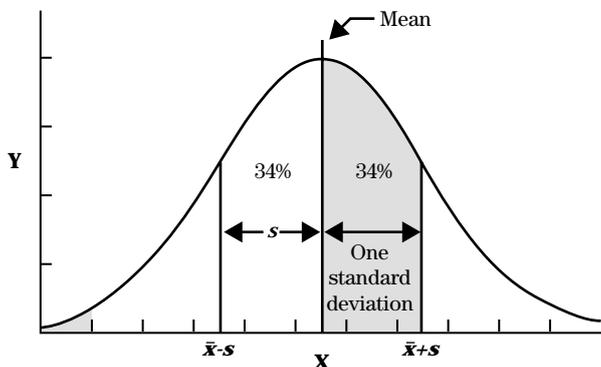
Measures of dispersion are useful to further understand a water quality data set. They indicate how spread out from the central tendency are the observations. The common measures of dispersion include the range, the variance (standard deviation is square root of variance), the standard error (standard deviation of a statistic, such as the mean), and the coefficient of variation.

A normal distribution has a preponderance of values around the mean and fewer observations at the extremes of the range of values. Such a distribution forms the typical bell-shaped curve (fig. 01-3).

The **range** is the distance from the smallest value to the largest value in the data set. It is the most simple of the measures of dispersion, but is subject to extreme values.

The **sample variance** is the sum of the squares of the deviations from the mean divided by the number of observations minus 1. Another term for variance is the mean square, which is the sums of squares divided by the degrees of freedom (n-1). The sample variance is represented by s^2 , and the population variance is represented by σ^2 .

Figure 01-3 Normal distribution



The sample variance is calculated from:

$$s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} \quad [01-7]$$

where:

X_i = value of the observation
 n = number of observations

The population variance is the sum of the squares of the deviations from the mean divided by the number of observations, rather than the number of observations minus one. The population variance is rarely used in water quality studies because sampling is almost always being conducted. Some calculators compute the wrong variance.

The **standard deviation** (s) is the square root of the variance ($\sqrt{s^2}$). The standard deviation carries the same units as the original data. The s is also called the root mean square. For normal distributions, one standard deviation on either side of the mean includes 68 percent of the observations and two standard deviations include 95 percent (fig. 01-3).

The **standard error of the mean** (SE), also termed the standard deviation of the mean, indicates the variability about the estimate of the mean:

$$SE = \frac{s}{\sqrt{n}} \quad [01-8]$$

where:

s = standard deviation
 n = number of observations

The standard error of the mean can be shown as an error bar in graphs summarizing mean values.

The **coefficient of variation** (CV) is a measure of the relative dispersion about the mean. It is defined as the standard deviation expressed as a percent of the mean:

$$CV = \frac{100 \times s}{\bar{X}} \quad [01-9]$$

where:

s = standard deviation
 \bar{X} = mean

The advantage of the coefficient of variation is that it allows direct comparison of variations between variables or among studies.

The *coefficient of skewness* indicates how equally distributed or symmetrical the data are about the mean. It is defined as the cube of the deviations about the mean (SAS 1985):

$$g_1 = \frac{n \sum (X_i - \bar{X})^3}{(n-1)(n-2)s^3} \quad [01-10]$$

The coefficient of skewness is normally distributed with a mean of 0 and a standard deviation of:

$$\left(\frac{6n(n-1)}{(n-2)(n+1)(n+3)} \right)^{0.5} \quad [01-11]$$

If g_1 is greater than four times the standard deviation of the skewness coefficient, then the data are skewed. Snedecor and Cochran (1980) provide a table for determining the significance of the skewness coefficient (appendix B). The sign of the skewness coefficient indicates whether the data are positively skewed (upper tail extended) or negatively skewed (lower tail extended) (fig. 01-4).

Kurtosis is a measure of the long tailedness of the distribution. It is defined as the average of the deviations from the mean raised to the 4th power divided by the standard deviation to the 4th power (SAS 1985):

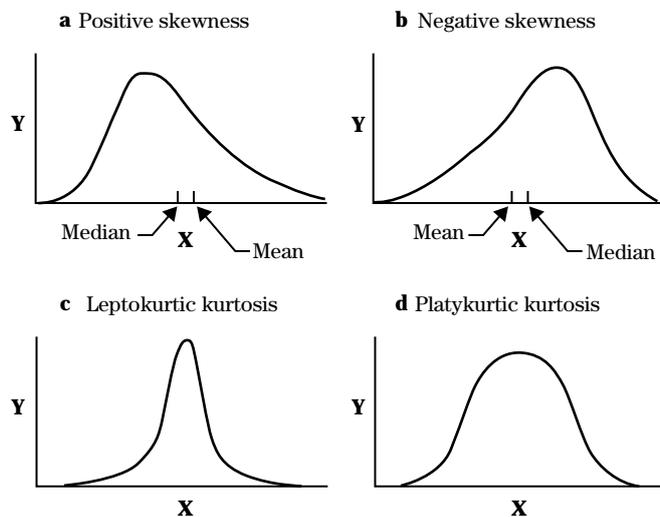
$$g_2 = \frac{n(n+1) \sum (X_i - \bar{X})^4}{(n-1)(n-2)(n-3)s^4} - 3 \frac{(n-1)^2}{(n-2)(n-3)} \quad [01-12]$$

The kurtosis is normally distributed with a mean of -3 and a standard deviation of:

$$\sqrt{\frac{24}{n}}$$

If the ratio of g_2 to standard deviation is less than -2, then the distribution has shorter tails than a normal distribution. If the ratio is more than 2, then the distribution has longer tails than a normal distribution (fig. 01-4). Snedecor and Cochran (1980) provide a table for testing the kurtosis based on the sample size and level of confidence desired.

Figure 01-4 Distributions showing skewness and kurtosis



Example 01-2 Measures of dispersion**Given:** Algal data in table 01-1**Determine:** Range, variance, standard deviation, standard error, coefficient of variation, skewness, and kurtosis values.**Solution:** Range: $45,283 - 25 = 45,258$ organisms/mL

$$\text{Variance: } s^2 = \frac{6,334,495,000 - \frac{(209,930)^2}{26}}{26 - 1} = 185,590,000$$

$$\text{Standard deviation: } s = \sqrt{185,590,000} = 13,623 \text{ organisms / mL}$$

$$\text{Standard error: } SE = \frac{13,623}{\sqrt{26}} = 2,672 \text{ organisms / mL}$$

$$\text{Coefficient of variation: } CV = \frac{100(13,623)}{8,074} = 169\%$$

$$\text{Skewness: } g_1 = \frac{(26)(0.110873E + 15)}{(25)(24)(13,623)^3} = 1.90$$

Since the skewness coefficient is positive, the upper tail is extended. Based upon a table provided by Snedecor and Cochran (1980) (appendix B), this skewness is significant at a probability (p) = 0.01.

$$\begin{aligned} \text{Kurtosis: } g_2 &= \frac{26(27)(0.388106E + 19)}{(25)(24)(23)(13,623)^4} - \frac{3(25)^2}{(24)(23)} \\ &= 2.335 \end{aligned}$$

$$\text{The standard deviation of the kurtosis is: } \sqrt{\frac{24}{n}} = \sqrt{\frac{24}{26}} = 0.9607$$

$$\frac{g_2}{0.9607} = 2.4$$

Since the ratio of g_2 to the standard deviation is greater than 2, the algae data have longer tails than a normal distribution (fig. 01-1).

615.0107 References

- Horowitz, G. 1981. *Sadistic statistics*. Avery Publishing Group, Inc., Wayne, NJ.
- Landwehr, J.M. 1978. Some properties of the geometric mean and its use in water quality standards. *Water Resources Research* 14(3):467-473.
- Newman, M.C., P.M. Dixon, B.B. Looney, and J.E. Pinder, III. 1989. Estimating mean and variance for environmental samples with below detection limit observations. *Water Resources Bulletin* 25(4):905-916.
- SAS Institute, Inc. 1985. *SAS procedures guide for personal computers*. SAS Institute, Inc. Cary, NC 27511.
- Snedecor, G.W., and W.G. Cochran. 1980. *Statistical methods*. 7th Ed. The Iowa State Univ. Press, Ames.
- Steele, R.G.D., and J.H. Torrie. 1960. *Principles and procedures of statistics*. McGraw-Hill Book Co., Inc., New York.
- United States Environmental Protection Agency. 1973. *Biological field and laboratory methods for measuring the quality of surface waters and effluents*. EPA-670/4-73-001. Office of Research and Development. Cincinnati, OH.
- Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. *Design of water quality monitoring systems*. Van Nostrand Reinhold. New York.
- Zar, J.H. 1996. *Biostatistical analysis*. 3rd Ed. Prentice Hall, Upper Saddle River, NJ 07458.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.02 Exploratory Data Analysis

Subpart 615.02 Exploratory Data Analysis

Contents:	615.0200	Introduction	02-1
	615.0201	Writing numbers	02-1
	615.0202	Stem-and-leaf diagrams	02-2
	615.0203	Schematic summaries	02-2
		(a) Box-and-whisker plot	02-2
	615.0204	Transformations	02-4
	615.0205	Comparisons	02-6
	615.0206	Plots of relationship	02-6
	615.0207	Smoothing data	02-7
	615.0208	References	02-8

Figures	Figure 02-1	Stem-and-leaf diagram for algal data from SAS® output	02-2
	Figure 02-2	Box-and-whisker plot	02-3
	Figure 02-3	Boxplot for the algal data from SAS® output	02-3
	Figure 02-4	Boxplot for the algal data from JMP output	02-3
	Figure 02-5	The mean as a function of the size of the constant added in a log transformation	02-4
	Figure 02-6	Boxplots for two annual sets of phosphorus data	02-6
	Figure 02-7	Relationship of observed to predicted runoff (*=p=0.05)	02-7
	Figure 02-8	Time plot of raw data	02-7

Figure 02-9 Smoothed data using medians of three 02-7

Figure 02-10 Smoothed data using the eye 02-8

Figure 02-11 Smoothed data using blurring 02-8

Examples Example 02-1 Log transformations with zero values 02-4

Example 02-2 Transformations 02-5

615.0200 Introduction

The first step in water quality data analysis is exploratory data analysis (EDA). For most data sets, EDA is a necessary step. The basic purpose for EDA is to better become familiar with the data. EDA is "detective work" that examines the data for how it appears (Tukey 1977). EDA, as proposed by Tukey, relies heavily on pictures. It is intended to provide indications rather than confirmations of a specific test. The actual procedure used varies with the type of data being explored, whether univariate, bivariate, or multivariate. Not all techniques are appropriate for all data; however, a number of steps are often examined for routine EDA. They include writing the numbers, stem-and-leaf diagrams, schematic summaries, transformations, comparisons, plots of relationships, and smoothing data.

Subpart 615.02 explains the various approaches to EDA. It presents examples of each of the routine methods used in EDA.

615.0201 Writing numbers

The process of writing numbers may be as simple as the listing of the raw data in a table. Tukey (1977) suggests using colors to highlight differences in the numbers making visual inspection easier.

Table 01–1 in subpart 615.01 is an example of writing numbers. In this case the numbers were written according to date. An alternative presentation would be to write the numbers from lowest to highest.

615.0202 Stem-and-leaf diagrams

Stem-and-leaf diagrams summarize the data visually in a sideways frequency diagram. Each line in the diagram is a stem, and each data point is a leaf on the stem. The stem represents the first digit of an observation in the data set. The leaves indicate the number of observations at that stem and the digits for those observations. Significant figures to the right of the leaves often are dropped. Stem-and-leaf diagrams are presented in many ways. Such a diagram, as presented in output from the Statistical Analysis System software (SAS®) for the algal data in table 01-1, is given in figure 02-1.

In this diagram, each stem is a multiple of 10,000; indicated by a 10**+4 by SAS®. The 4 represents 40,000, 3 represents 30,000 and so on. The leaf of 05 indicates that there are two numbers of 40,000 or greater, after rounding to the nearest 1,000. The data are skewed toward the low values (Stem = 0). There are more values for the leaf column at the stem of 0 than other stem values. SAS® output indicates the number of leaves in each stem by a # column. SAS® output also gives a multiplication factor for the Stem.Leaf data if needed.

Figure 02-1 Stem-and-leaf diagram for algal data from SAS® output

Stem	Leaf	#
4	05	2
3	7	1
2	9	1
1	05	2
0	00000111111111222667	20

-----+-----+-----+-----+

Multiply Stem.Leaf by 10**+4

615.0203 Schematic summaries

The stem-and-leaf diagram can also be summarized using five numbers: the median, maximum, minimum, and upper and lower hinges. The rank of the median can be determined from:

$$\text{median rank} = \frac{1 + \text{count}}{2} \quad [02-1]$$

The hinges are half-way from the extremes to the median and are determined by:

$$\text{hinge} = \frac{1 + \text{median rank}}{2} \quad [02-2]$$

The hinges are so-named because they represent folds in the data between the median and the extremes (Tukey 1977). Another way to characterize the lower and upper hinges is as the 25th and 75th percentile values. The upper hinge is the value that is three-fourths of the way along the values when ranked from lowest to highest.

These five numbers can be provided in a box, as below:

	Med	
H _{low}		H _{high}
Min		Max

For example, using the algae data from table 01-1, the five-number summary would be:

	1,168	
547		6,755
25		45,283

(a) Box-and-whisker plot

Another more common schematic summary is the box-and-whisker plot. This is really a five-number summary in graphical form. The box extends from lower hinge to upper hinge and is crossed with a bar at the median (fig. 02-2). The 75th percentile means that 75 percent of all values are below that value. The whiskers extend from each end of the box to the respective extreme.

In some cases it is desirable to show some data values as farther out than others. H-spread is a term given to the differences between the hinges. A step is 1.5 times the box length (H-spread). An inner fence can be placed at one step outside the hinges; an outer fence is located at two steps outside the hinges.

The values located inside the inner fence, but closest to the inner fence are termed *adjacent*. Values between the inner and outer fences are termed *outside*. And values beyond the outer fences are *far out*.

The box-and-whisker plot is useful in conveying a concept of how even is the data above and below the median. In some cases the whisker may end at the adjacent values.

Boxplots are included in SAS[®] output using PROC UNIVARIATE PLOT (SAS[®] 1985). The boxplot for the algal data is shown in figure 02-3. Another boxplot from the output of JMP (SAS[®] Institute, Inc.) is given in figure 02-4.

The boxplots show that the data are highly skewed to the low values. The bottom and top of the box represent the 25th and 75th percentiles (hinges). The center horizontal line is drawn at the median, and a + is given at the mean (SAS[®] output). In the example, all these lines are so close that they are printed on the same line (fig. 02-3). The whiskers in SAS[®] extend to 1.5 the inter-quartile range (H-spread). Values more extreme, but within three interquartile ranges, are indicated with a zero. Values outside are indicated with an asterisk. For the example in figure 02-3, the three asterisks indicate three extreme values outside three interquartiles. The JMP outlier boxplot uses dots for points beyond the whiskers. The diamond indicates the 95 percent confidence intervals about the mean.

Figure 02-2 Box-and-whisker plot

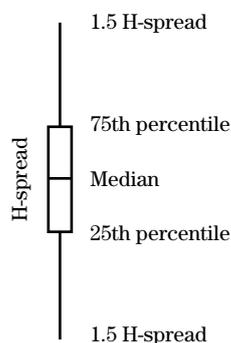


Figure 02-3 Boxplot for the algal data from SAS[®] output

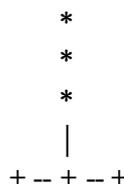


Figure 02-4 Boxplot for the algal data from JMP output



615.0204 Transformations

Transformations of the data are sometimes needed to normalize the data or stabilize the variance. Transformations also change the appearance of the data into a form that may be more readily understandable (Tukey 1977). Some basic rules for different transformations have been described by Tukey (1977):

- Amounts and counts can never be less than zero, but can be large. A transformation may be useful if the ratio of the largest value to the smallest value is large (i.e., 100 or more). If the ratio is small (i.e., 1), the transformation will not modify the appearance of the data.
- Balances, values which can be both positive and negative, are usually not improved by transformations.
- Fractions and percentages may be better expressed with transformations.
- Grades, such as A, B, C, D, also may respond to complex transformations.

A common transformation for water quality data is the use of logarithms. The log distribution for concentration data makes sense because negative values do not exist, many values exist at lower concentrations, and a few values will exist at much higher concentrations (positively skewed). If plotted in a frequency diagram, the typical exponential decay curve results. Logs also are appropriate when the standard deviation in the data is likely to be proportional to the mean or for data that are proportional rather than additive on a linear scale (Snedecor and Cochran 1980, Sokal and Rohlf 1969). Logs tend to squeeze the data together and make it more symmetrical. A log transformation of zero does not exist; zeros can exist in a data set of mass export values. Also, when the data values are less than one, a log transformation gives negative numbers. In such cases the addition of a constant, such as $\log(X+1)$, is recommended (Steel and Torrie 1960, Zar 1984). However, the size of the constant added influences the estimate of the mean for the data set, as shown in example 02-1.

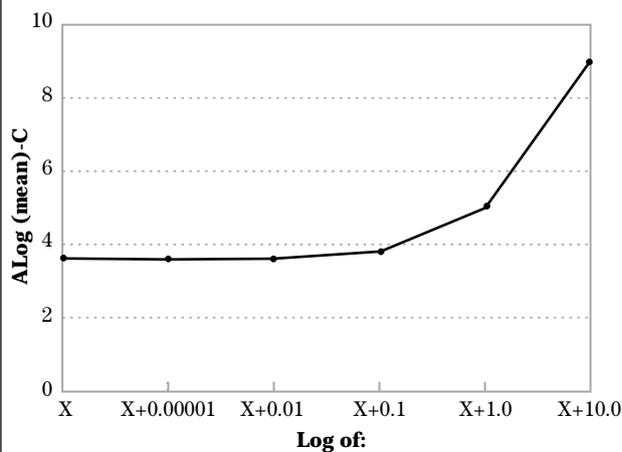
Example 16-1 Log transformations with zero values

A \log_{10} transformation was applied to the following values of X:

0.25	5.0
0.5	8.0
0.8	14.0
1.0	50.0
1.2	100.0

Additional transformations were made by adding the following constants: 10.0, 1.0, 0.1, 0.01, and 0.00001. The mean was obtained for each transformed data set as the antilog of the mean of the logged data minus the added constant. The results from these transformations are plotted in figure 02-5. These transformations indicate that adding smaller constants results in mean values that approach the true mean for the data set.

Figure 02-5 The mean as a function of the size of the constant added in a log transformation



Counts, such as for bacteria data, can be re-expressed with logs and square roots, with root counts more often used (Tukey 1977). When the data numbers are small (<10) the square root transformation is recommended (Steel and Torrie 1960). If the counts are small, Snedecor and Cochran (1980) recommend the square root ($X + 1$) transformation.

Percentage data, based on counts, where the data range from 0 to 20 percent or 80 to 100 percent may be transformed with a square root (Steel and Torrie 1960). Percentage or decimal data based on binomial data can be re-expressed using an arc sine or inverse sine transformation.

For data that are skewed to the left, a value squared transformation has been recommended (Zar 1984).

Generally, when transformations are made, the mean is transformed back to the original scale, but variances or standard deviations should not be transformed back to the original scale (Steel and Torrie 1960).

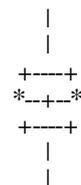
Example 02-2 illustrates the transformation of the St. Albans Bay data from subpart 615.01.

Example 02-2 Transformations

A \log_{10} transformation was made of the St. Albans Bay algal data in table 01-1. The stem-and-leaf diagram for the transformed data indicates that the transformation removed much of the skewness in the algal count data, as compared to figure 02-1.

Stem	Leaf	#
4	5667	4
4	02	2
3	888	3
3	001233	6
2	56778999	8
2	1	1
1	7	1
1	4	1
	-----+	

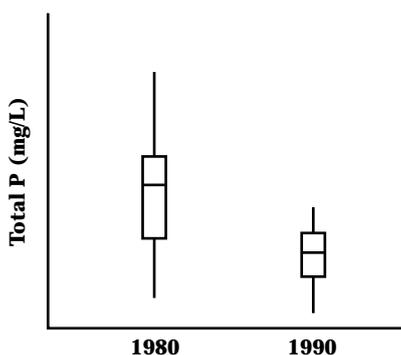
The box-and-whisker plot of the \log_{10} transformed data also shows that the data are now more evenly distributed above and below the median as compared with figure 02-3. The absence of zeros and asterisks in the whiskers indicates that there are no values more extreme than three interquartile ranges. This shape is characteristic of a normal distribution.



615.0205 Comparisons

Different groups of data can be compared in several ways. They include side-by-side stem-and-leaf displays, tables of means or medians, and box-and-whisker plots. Transformations of scale often aid in the comparison among groups. For example, the box plots in figure 02-6 indicate that the phosphorus concentrations for 1990 were lower and less variable than for 1980. The width of the box can be used to reflect the sample size when comparing samples of different sizes (R.H. McCuen 1998, personal communication).

Figure 02-6 Box plots for two annual sets of phosphorus data



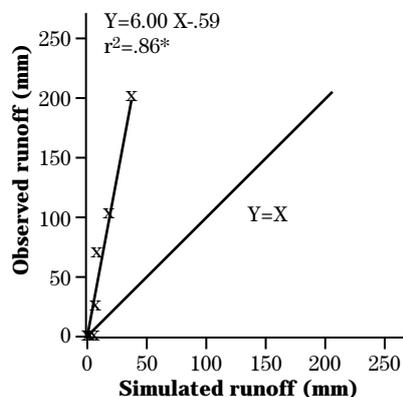
615.0206 Plots of relationship

Plots can be used to describe a relationship between a response variable (dependent) and a factor (independent) (Tukey 1977). The independent variable is usually shown as the abscissa (horizontal X-axis), and the dependent variable is shown as the ordinate (vertical Y-axis). Although default values in computer graphics programs make many decisions for us, there are some general rules that are useful in plotting relationships. These rules include guidance regarding the scale, shape, grid, and labeling of axis.

If comparing different plots with similar information, all plots should be at the same scale even though your graphics program may not default in this manner.

The shape of the plot is another important consideration. Plots can be taller than wide, wider than tall, and of equal dimensions. Taller than wide plots are useful for growth or decay phenomenon. Wider than tall plots facilitate reading from left to right and might be useful for scatter diagrams or time plots. Square plots may be useful in situations where the same units are plotted on each axis and the 45 degree line, representing $Y = X$, has some meaning. Figure 02-7 provides an example of such a graph—the comparison of observed data to data simulated by a model (Jamieson and Clausen 1988).

Figure 02-7 Relationship of observed to predicted runoff (*= $p=0.05$)



The type of grid chosen for the graph influences the interpretation of the graph. Data that are extremely variable, such as suspended solids concentrations, might better be graphed using a log scale rather than a linear scale. Also, exponential relationships are straightened by plotting them on a log-log scale. If the data contain zero values, they cannot be plotted on a log scale unless a constant is added, as described in the previous section on transformations.

The labeling of axis, both in terms of the use of values and tick marks, influences interpretations from the graph. Generally, the number of tick marks and values shown on the graph are minimized because they can be distracting to the eye. An exception would be when the graph is used to pick off points. The origin of the graph, where $Y = X$, is generally zero to show the real magnitude of the values. This guidance is often abused in the media (e.g., stock market) to indicate larger variations than are really occurring.

One of the more common abuses of plots of relationship is termed *spurious correlation* (Kite 1989). This occurs when both axes have a variable in common. For example, a plot of mass export as a function of stream discharge is almost always guaranteed to show a positive relationship. This occurs because the values for the variable discharge are included in both axes. In regression, this would also violate the assumption of independence.

615.0207 Smoothing data

Smoothing data allow definition of general trends without looking at too much detail (Tukey 1977). Generally, the **Y** data are smoothed and the **X** data become intervals. Several techniques are used in smoothing. They include running medians or averages, eye smoothing, blurring, and splitting. An example of a water quality data set where smoothing might be useful is a time plot of concentration data (fig. 02–8).

The data as they appear are quite rough, and general trends are difficult to interpret. To use running medians or averages, take adjacent **Y** values and calculate a new smoothed point. Running implies that a central estimate is made for each point as opposed to creating intervals and deriving a central estimate for each interval. A running 3-day average of daily data is computed by determining the average of the days around Monday, then around Tuesday, and so on. For example, the median of three values running was used to develop the points for figure 02–9. The first three points were: 3, 7, and 10, which would result in a median of 7. New medians are calculated for the second set of three values by skipping the first number. In this case using the medians of a larger number of points may have provided a smoother picture of the data. The mean could be used rather than the median for smoothing. This method is also called the *moving average method*.

Figure 02–8 Time plot of raw data

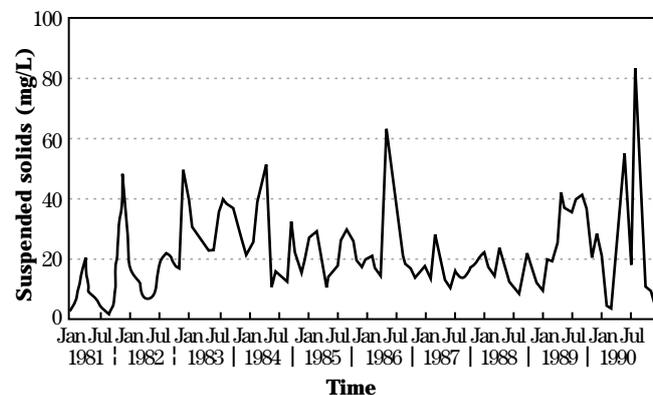
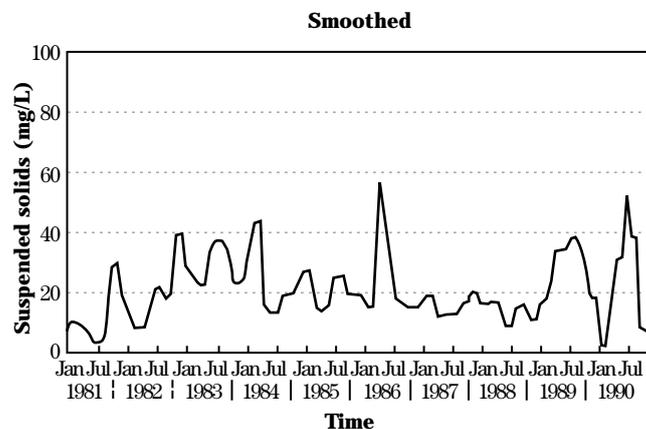


Figure 02–9 Smoothed data using medians of three



Eye smoothing is drawing a smooth curve through the data. However, smoothing by eye allows bias to be used to meet the need or intent of the analyst. For example, figure 02–10 shows two curves fit to the smoothed data that are quite different from each other. Both lines are smoothing of the data. One line attempts to follow the peaks and valleys; the other suggests an even more general trend. This trend is contingent upon when monitoring began. Note that if the sampling began in 1983, a different trend might be suggested.

Blurring is a method of smoothing where the data points are replaced with vertical lines of some length showing their variability. In figure 02–11, the raw data have been blurred, which suggests a band of data rather than a line or a series of points.

Figure 02–10 Smoothed data using the eye

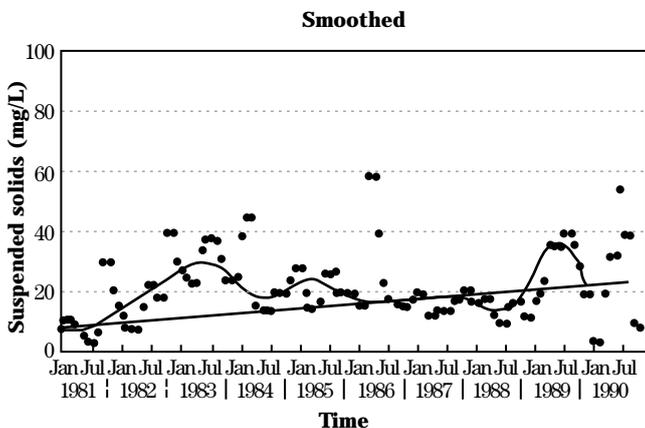
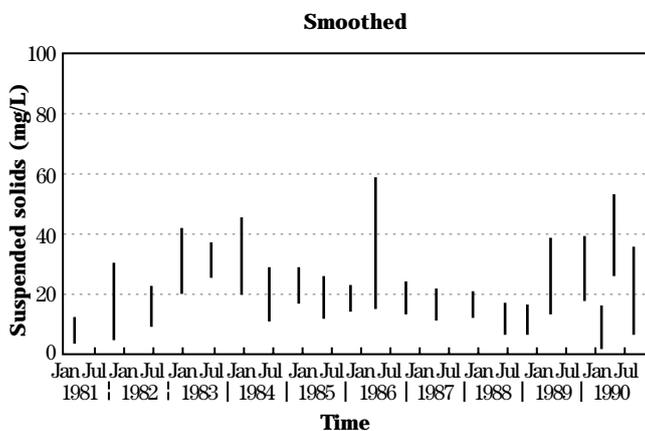


Figure 02–11 Smoothed data using blurring



615.0208 References

- Jamieson, C.A., and J.C. Clausen. 1988. Test of the CREAMS model on agricultural fields in Vermont. *Water Resour. Bul.* 24(6):1219-1226.
- Kite, G. 1989. Some statistical observations. *Water Resour. Bul.* 25(3):483-490.
- SAS Institute, Inc. 1985. SAS procedures guide for personal computers, Version 6 edition, SAS Institute, Inc., Cary, NC.
- Snedecor, G.W., and W.G. Cochran. 1980. *Statistical methods*. 7th Ed., The Iowa State Univ. Press, Ames, IA.
- Sokal, R.R., and F.J. Rohlf. 1969. *Introduction to biostatistics*. W.H. Freeman and Co., San Francisco, CA.
- Steel, R.G.D., and J.H. Torrie. 1960. *Principles and procedures of statistics*. McGraw-Hill Book Co., Inc., New York, NY.
- Tukey, J.W. 1977. *Exploratory data analysis*. Addison-Wesley Publ. Co., Reading, MA.
- Zar, J.H. 1984. *Biostatistical analysis*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.03 Statistical Assumptions

Subpart 615.03 Statistical Assumptions

Contents:	615.0300 Introduction	03-1
	615.0301 Assumptions	03-1
	(a) Randomness	03-1
	(b) Normal distribution	03-2
	(c) Homogeneity of variances	03-3
	(d) Independence	03-4
	(e) Additivity	03-4
	615.0302 Residual plots	03-5
	615.0303 Missing data	03-5
	615.0304 Extreme outliers	03-6
	615.0305 Summary	03-7
	615.0306 References	03-7

Table	Table 03-1 Statistical assumptions and tests	03-7
--------------	---	------

Figures	Figure 03-1 Examples of frequency distributions	03-2
	Figure 03-2 Residual plots for linear regression	03-5

Examples	Example 03-1 Cumulative frequency distributions for St. Albans Bay algae data from SAS output	03-2
	Example 03-2 Test of normality for the St. Albans Bay algae data	03-3

615.0300 Introduction

When applying statistical analyses to water quality data, such as analysis of variance, we must be familiar with several underlying assumptions. It is important to know how to test if these assumptions have been violated and what to do if they are violated. All assumptions are difficult to meet exactly. It is more important to understand whether the violation of an assumption has a serious consequence on the probability statements made based on the assumption (Glass, et al. 1972). The main assumptions are: randomness, normality, homogeneity of variances, independence, and additivity.

Subpart 615.03 describes the various statistical assumptions made when performing statistical tests. The consequences of failing to exactly meet these assumptions are presented for each assumption. The usefulness of residual plots in evaluating assumptions is also detailed as is how to deal with missing data and extreme outliers.

615.0301 Assumptions**(a) Randomness**

The first assumption is that the water quality data are sampled randomly. *Randomness* means that the probability of obtaining a sample remains the same for all possible samples (Steel and Torrie 1960). The purpose of randomization is to design bias out of the study and increase the accuracy of the study (Hurlbert 1984). For example, if a stream was sampled only during stormflow periods, the study would be biased toward higher concentrations than if the stream were sampled mostly during low-flow periods. Water quality data have both random and deterministic components (Moser and Huibregtse 1976). Random components are introduced by precipitation events that are themselves random in most parts of the United States. Nonrandom components are related to trends or seasonality in the data (subpart 614.06, fig. 06-1).

Water quality samples may not be truly random for several reasons. Sampling is not randomized over all possible observations. For example, if sampling were done from 1980 to 1990, the sampling ignores what the water quality may have been for all time before 1980. By sampling within a shorter window than all time, there is a possibility that a nonrandom component is dominating water quality.

The lack of randomness may result in producing a lack of independence, heterogeneous variances, or non-normal distributions (Sokal and Rohlf 1969). No specific test of randomness is available; however, proper design of the sampling program should ensure an appropriate level of randomness.

Sampling methods to maintain randomness are described in part 614 of this handbook in subparts 07 and 08.

(b) Normal distribution

A second assumption is that the data come from a population with a particular frequency distribution of values, usually a normal distribution. Several methods are available for examining the normality of the data. They include graphical and statistical methods. The graphical approach is to plot the data in a cumulative frequency distribution. Normal data plot as a straight line on such a graph (fig. 03-1a). Data that are skewed (long tail) to the left have a cumulative frequency distribution that is concave upward (fig. 03-1b). Data that are skewed right have a cumulative frequency distribution that is concave downward (fig. 03-1c).

Within the Statistical Analysis System, a normal probability plot can be obtained from:

PROC UNIVARIATE PLOT;

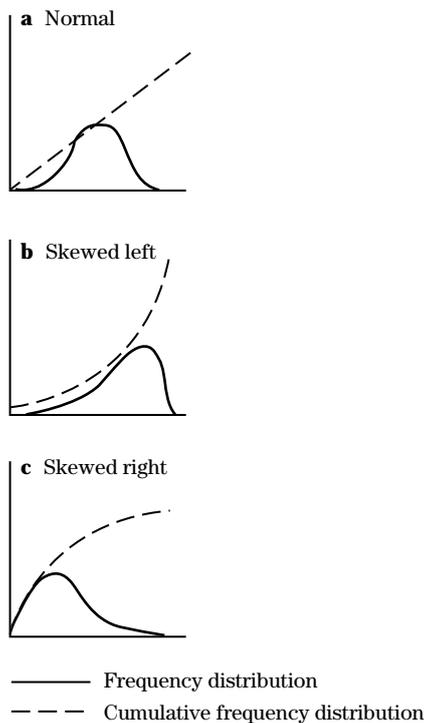
In addition to the normal probability plot, a stem-and-leaf plot and a boxplot are automatically produced.

Example 03-1 illustrates the cumulative frequency distributions for St. Albans Bay algal data in table 01-1 using the SAS® output. The log transformed data produces a straighter line on the normal probability plot than the untransformed data. This finding implies that the data follow a log normal distribution, and a log transformation should be used in subsequent statistical analysis.

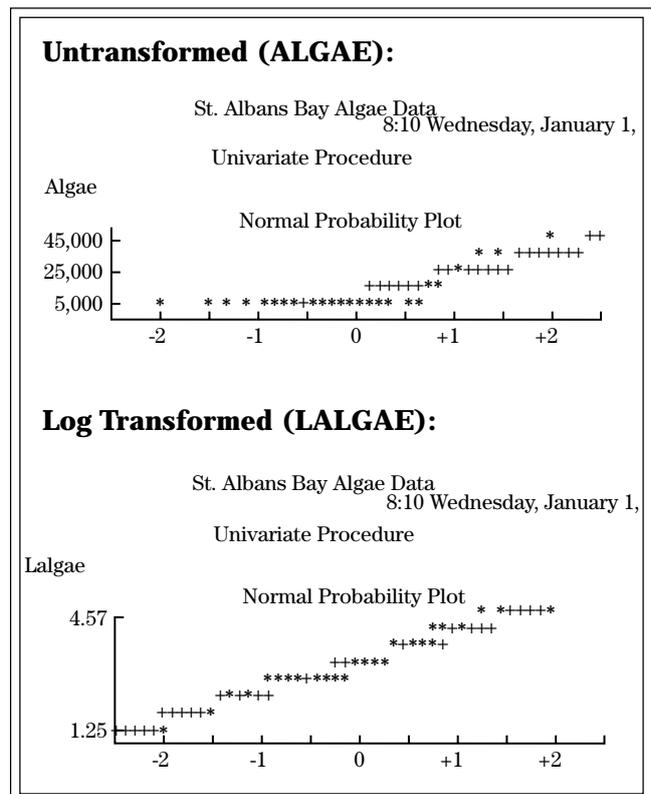
Among the statistical approaches for evaluating the normality of the data is the use of univariate statistics, such as the mean, median, skewness, and kurtosis. Generally, if the median and the mean are very different, the data may not be normally distributed. In addition, tests of either the skewness or the kurtosis will provide information regarding the normality of the distribution (see subpart 615.01).

Several statistical tests have been used for testing normality. One common test is the Chi-square goodness of fit (Snedecor and Cochran 1980; Sokal and

Figure 03-1 Examples of frequency distributions



Example 03-1 Cumulative frequency distributions for St. Albans Bay algal data from SAS® output



Rohlf 1969; Zar 1984). This tests the hypothesis that the sample came from a specific theoretical distribution.

The goodness of fit also may be tested using the Kolmogorov-Smirnov test (Zar 1984). Finally a test for normality can be accomplished by using the Shapiro-Wilk W-statistic. The W statistic has values ranging from 0 to 1; small values for W are significant and indicate nonnormality (Shapiro and Wilk 1965). The decision whether to use the Kolmogorov-Smirnov test is dependent on the sample size. For samples less than 2,000, the Shapiro-Wilk test should be used (SAS 1985). For larger samples, the Kolmogorov-Smirnov test should be used. Example 03-2 illustrates the test of normality using the Shapiro-Wilk W-statistic.

SAS® output provides the W statistic and its probability using the following command:

PROC UNIVARIATE NORMAL;

Example 03-2 Test of normality for the St. Albans Bay algae data

	Untransformed	Log transformed
Mean	8075.	3.234
Median	1168.	3.063
Skewness	1.900	-0.039
Kurtosis	2.334	-0.389
W:Normal	0.626	0.959
Prob<W	0.0001	0.4018

For the St. Albans Bay algal data, the small W for the untransformed data indicates that the W is significant and nonnormal. The log-transformation of this data resulted in a large, nonsignificant W. The hypothesis that the data come from a normal distribution cannot be rejected. Therefore, the log transformed data are assumed to be normally distributed. Note also that the mean and median are closer and the skewness and kurtosis are smaller for the log transformed, as compared to the untransformed data.

Failure to exactly meet the assumption of normality is generally not considered to be a major problem (Glass, et al. 1972, Sokal and Rohlf 1969). The significance levels for t-tests and F-tests do not appear to be affected by nonnormality. That is to say that the probability of the Type I error is not increased significantly by failure to meet the assumption of normality (subpart 615.05). This is especially true for large data sets and when equal numbers of values are being compared. Skewed populations can affect the level of significance for one-tailed tests (Glass, et al. 1972). It is not considered necessary to use nonparametric approaches simply because the assumption of normality has not been exactly met. However, an appropriate transformation to better approximate normality is recommended.

(c) Homogeneity of variances

In uses involving more than one data set, the equality of variances is an important assumption for several statistical tests. If there are two sample data sets that are being compared, the test of the homogeneity of variances is made by computing an F as the ratio between the larger variance divided by the smaller variance (Snedecor and Cochran 1980, Sokal and Rohlf 1969). The computed F is compared to a critical value for F from an F table (appendix C).

If three or more sample data sets are compared, Bartlett's test may be used. The ratio of the test statistic, B, to a correction factor is compared to the chi-square statistic (Snedecor and Cochran 1980, Zar 1984). For nonnormal distributions, some prefer the Levene's test for homogeneity of variances (Snedecor and Cochran 1980). The statistical program BMDP (Biomedical Computer Programs P-Series) computes both the Bartlett's test (BMDP9D) and the Levene's test (BMDP7D) (Dixon and Brown 1979).

A quick test is the F_{\max} test for which an F ratio is computed from S_{\max}^2/S_{\min}^2 and compared to a critical value for F that is given in various tables (appendix C) (Sokal and Rohlf 1969, Peterson and Hartley 1954).

The consequence of failing to meet the assumption of equal variances can be serious, especially when the sample sizes from the two groups are of unequal size (Glass, et al. 1972). When the sample sizes of the groups are equal, there is little effect on the probability

level of committing a Type I error (subpart 615.05). When the sample sizes of the groups being compared are unequal and the variances are heterogeneous, the probability of committing a Type I error may be seriously affected. The probability level may be underestimated when a smaller number of samples come from the more variable population. It may be overestimated when a smaller number of samples come from the less variable population (Glass, et al. 1972).

Transformations often help remove heterogeneous variances. If a transformation does not eliminate the problem, perhaps the data could be aggregated so that the number of samples among groups could be equalized. If this is not possible, a nonparametric approach may be desirable.

(d) Independence

Another assumption is that the experimental errors are independently distributed (Sokal and Rohlf 1969, Steel and Torrie 1960). That is, if the data are arranged in some logical sequence, such as in the order of collection, the errors should follow each other randomly.

Randomization in sampling helps reduce the correlation of observations and their errors over time. This is a special concern in water quality sampling where high values are more likely to follow high values and low values follow low values.

If the errors are not independent, the F-test in analysis of variance (ANOVA) and the *t*-test results can be questioned. With positive serial correlations, the probability level of the Type I error is increased progressively with the size of the correlation. With negative correlations, the probability level of the Type I error is much lower than it really should be (Glass, et al. 1972).

If the sampled data are serially correlated, there is little that can be done. Randomization in the design of the experiment was insufficient. One alternative may be to aggregate the serial data in some logical manner, such as computing means or totals. For example, serially correlated weekly data could be aggregated to monthly data that may not be correlated. Another option would be to use Time Series Analysis (Vandaele 1983). This analysis assumes that the errors are not

independent and are, in fact, correlated according to some time step. Although time series analysis has certain applications in water quality monitoring, such as trend analysis, it is a sophisticated statistical technique requiring special training.

Serial or auto correlation of the residuals can be determined from:

$$r_k = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad [03-1]$$

where:

- r_k = autocorrelation coefficient for any lag k
- y = observation at any time step t
- N = total number of observations

In SAS® the autocorrelation coefficient may be obtained by:

```
PROC REG;
MODEL Y=X / DW;
```

The DW stands for the Durbin-Watson *d* statistic that is a test of the hypothesis that autocorrelation is zero (SAS 1985).

(e) Additivity

The assumption of additivity (also termed linearity) is normally applied to ANOVA and means that the effects of the treatment are additive, not multiplicative (Sokal and Rohlf 1969, Steel and Torrie 1970, Zar 1984). One way of viewing additivity is by writing the model for an ANOVA. A typical one-way ANOVA model would take the form:

$$\chi_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

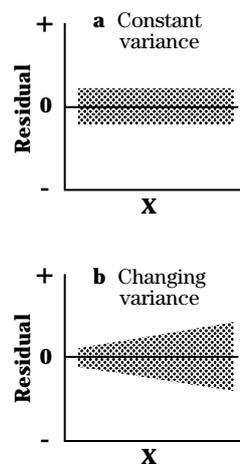
This equation states that an observed value (χ_{ij}) equals the sum of an overall mean (μ), a treatment deviation (α_i), and a random error term (ϵ_{ij}) (Snedecor and Cochran 1980). The three factors in the equation are additive rather than multiplicative. Thus there would be no interaction in this particular model.

A test for nonadditivity has been suggested by Tukey (Snedecor and Cochran 1980). Log transformations of multiplicative effects promote additivity in the data.

615.0302 Residual plots

When using linear regression, an examination of a plot of the residuals, as a function of the independent variable, helps in the assessment of several of the assumptions including equal variances, independence, as well as the adequacy of the linear regression model (Afifi and Azen 1979, Draper and Smith 1981, Ponce 1980, Zar 1984). A residual is the deviation of a datum point from the regression line. For example, if the residuals are independent and of constant variance, then they should be scattered evenly about the horizontal line where the residual is zero (fig. 03-2a). If, on the other hand, the residuals appear to increase or decrease as X increases (fig. 03-2b), the variance may not be constant. A nonconstant variance implies that the regression model is inadequate.

Figure 03-2 Residual plots for linear regression



615.0303 Missing data

Missing data are common in water quality sampling. Sometimes samples are missing because they were not collected. Possible reasons for not collecting samples include equipment failure, frozen conditions, or missing an event. Water samples that must be analyzed in a laboratory are subject to accidents or a quality assurance program that may render the sample as in error.

Missing data are important for some water quality monitoring designs, but not for all designs. Missing values may not be important for paired and unbalanced unpaired tests where the number of samples is adequate. The missing value merely eliminates a pair from the analysis and reduces the sample size. However, missing data may have important consequences on trend analysis.

As a cautionary note, the analyst must be aware of how missing data are coded when using computer statistical packages. Some packages read a blank as a zero. If a special value is used, such as -9, the computer may include that in calculations unless specifically informed otherwise. Each statistical package may have different requirements. SAS for example recognizes a '.' as missing. One should also be aware that for some packages a missing value within a line (or case) may result in the elimination of the entire case.

Several techniques are used to estimate missing water quality data. They include linear interpolation, regression with another station or flow, and the use of several stations. In addition, more sophisticated measures are needed for missing blocks in randomized block designs (Snedecor and Cochran 1980, Zar 1984).

Linear interpolation uses the existing values adjacent to the missing value(s) and assumes that the missing value(s) is proportional to the difference between the known values.

For water quality data that are highly correlated to either other water quality data or flow data, missing values could be predicted using a regression equation. For missing flow data, a relationship with precipitation or with flow at a nearby station may provide an adequate predictor of the missing information.

Another approach is that several stations could be used to predict a single missing value if such data are available. For example, the concentration at a fourth station could be determined from the concentrations observed at three other stations and the means at all stations using the equation:

$$C_4 = \frac{1}{3} \left(\frac{\bar{C}_4}{\bar{C}_1} \times C_1 + \frac{\bar{C}_4}{\bar{C}_2} \times C_2 + \frac{\bar{C}_4}{\bar{C}_3} \times C_3 \right) \quad [03-2]$$

where:

C = concentration at stations 1, 2, 3, and 4

\bar{C} = mean for the respective station

615.0304 Extreme outliers

Water quality data sets generally contain values that appear to be extreme outliers. The initial response should be to verify that no mistake has been made in recording the observation. Upon occasion, an error has been made, but the true value cannot be determined. In this case the data could be declared missing.

Several methods are available for determining whether certain observations are outliers (e.g., Dunn and Clark 1987). For example, the maximum normed residual (MNR) can be calculated from:

$$MNR = \frac{\text{Max}|x - \bar{x}|}{\sqrt{\sum (x_1 - \bar{x})^2}} \quad [03-3]$$

where:

x = outlier to be tested (Snedecor and Cochran 1980).

The calculated MNR is compared to a tabular MNR, which varies with the sample size and probability level. If the calculated MNR is less than the tabular MNR, the value is expected to occur more often than the probability level, and thus is not considered an extreme outlier.

615.0305 Summary

Table 03–1 provides a summary of the standard assumptions for parametric statistical tests and the appropriate methods for testing the assumption.

Table 03–1 Statistical assumptions and tests

Assumption	Test
Randomness	Sampling design
Normality	Graphical Shapiro-Wilk Kolmogorov-Smirnov
Equal variances	F ratio Bartlett's Levene's
Independence	Residual plot Autocorrelation
Additivity	Tukey's

615.0306 References

- Afifi, A.A., and S.P. Azen. 1979. *Statistical analysis: a computer oriented approach*. (2nd ed). Academic Press. New York, NY.
- Dixon, W.J., and M.B. Brown (Eds.). 1979. *BMDP biomedical computer programs P-series*. Univ. CA Press, Los Angeles, CA.
- Draper, N.R., and H. Smith. 1981. *Applied regression analysis*. (2nd ed.) John Wiley and Sons. New York, NY.
- Dunn, O.J., and V.A. Clark. 1987. *Applied statistics: Analysis of variance and regression* (2nd ed.) John Wiley and Sons, New York, NY.
- Glass, G.V., P.D. Peckham, and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev. Educ. Research* 42(3):237-288.
- Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2):187-211.
- Moser, J.H., and K.R. Huibregtse. 1976. *Handbook for sampling and sample preservation of water and wastewater*. U.S. EPA 600/4-76-049.
- Peterson, E.S., and H.O. Hartley. 1954. *Biometrika tables for statisticians*. (Vol I.) Cambridge Univ. Press.
- Ponce, S.L. 1980. *Statistical methods used in water quality data analysis*. WSDG Tech. Pap. WSDG-TP-00001. Watershed Systems Development Group, USDA Forest Service, Fort Collins, CO.
- SAS Institute, Inc. 1985. *SAS procedures guide for personal computers*. (Ver. 6 ed.) Cary, NC.
- Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611.

Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods. (7th ed.) The IA State Univ. Press, Ames, IA.

Sokal, R.R., and F.J. Rohlf. 1969. Introduction to biostatistics. W.H. Freeman and Co., San Francisco, CA.

Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., Inc. New York, NY.

Vandaele, W. 1983. Applied time series and Box-Jenkins models. Academic Press, Inc., New York, NY.

Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.04 Causality

Subpart 615.04 Causality

Contents:	615.0400	Introduction	04-1
	615.0401	Association	04-1
	615.0402	Consistency	04-2
	615.0403	Responsiveness	04-4
	615.0404	Mechanism	04-4
	615.0405	Experimental design	04-5
		(a) Time scale	04-5
	(b) System level	04-5	
	(c) Reasonableness of treatment	04-6	
	615.0406	References	04-6

Table	Table 04-1	Correlations between mean weekly discharge concentrations and discharge	04-1
--------------	-------------------	---	------

Figures	Figure 04-1	Jewett Brook phosphorus concentration and manure applied in the watershed	04-2
	Figure 04-2	Mean annual bacteria abundance and the percent of BMP animal units for the Jewett Brook watershed	04-3
	Figure 04-3	Mean annual bacteria abundance and the percent of BMP animal units for the Stevens Brook watershed	04-3
	Figure 04-4	Mean annual bacteria abundance and the percent of BMP animal units for the Rugg Brook watershed	04-3

Figure 04-5	Mean annual bacteria abundance and the percent of BMP animal units for the Mill River watershed	04-3
--------------------	---	------

Figure 04-6	Mechanism for bacteria decreases	04-5
--------------------	----------------------------------	------

Examples	Example 04-1	Correlation	04-1
	Example 04-2	Regression	04-2
	Example 04-3	Consistency	04-2
	Example 04-4	Responsiveness	04-4
	Example 04-5	Mechanism	04-5

615.0400 Introduction

Although the reasons for conducting water quality monitoring are varied (see part 614, subpart 614.00), many involve attempting to develop a cause-and-effect relationship between something that is done on the landscape (cause) and a response in water quality (effect). In statistical terms an experimental design is developed to determine the conclusion desired. An experimental design is a plan of the experimental units, treatments including a control, and the replications to achieve some objective. Four concepts provide a useful framework for designing water quality monitoring studies with causation in mind. These concepts are association, consistency, responsiveness, and mechanism (Mosteller and Tukey 1977).

This subpart describes these four concepts of causality. Examples are used to illustrate each of these requirements. Other features of designing experiments are also described.

615.0401 Association

An association between variables, such as water quality and land treatment, implies that these variables are paired in a related way across the population (Mosteller and Tukey 1977). An association is necessary, but not sufficient to show causality.

An association may be expressed in several ways including correlation and regression (Draper and Smith 1981) or a significance analysis of variance (ANOVA) model. Regression is appropriate when one variable is dependent on the other (Zar 1984). When two variables are associated, but one is not dependent upon the other, correlation analysis is used. For example, the association between runoff and rainfall is best analyzed by regression because runoff is dependent upon rainfall. However, the association between stream order and discharge is best explained by correlation. Discharge would be expected to be greater for higher order streams although there is no mathematical dependence of discharge on stream order.

Examples 04-1 and 04-2 help to illustrate the meaning of association.

Example 04-1 Correlation

Water quality monitoring in the Jewett Brook watershed in Vermont revealed an association between stream discharge and various water quality variables (Hopkins and Clausen 1985). This association is represented by correlation coefficients of log-transformed data (table 04-1).

The correlations in table 04-1 do not necessarily imply dependence. Increased discharge may not cause increased concentrations in streamflow. Rather, other processes, for example snowmelt, can cause increases in both discharge and concentrations. Surely, increased stream concentrations do not cause increased discharge.

Table 04-1 Correlations (r) between mean weekly discharge concentrations (mg/L) and discharge (m³/s) n=52

Variable	Correlation coefficient (r)
Total phosphorus	0.37**
Total kjeldahl nitrogen	0.44**
Total suspended solids	0.61**

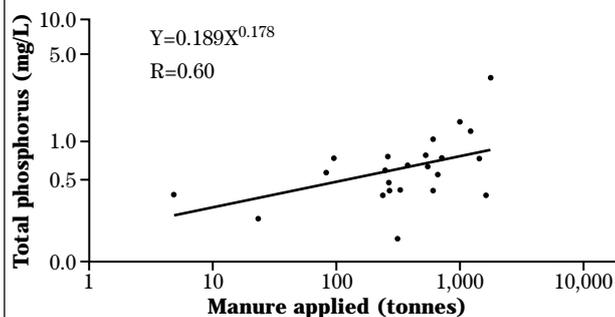
** Indicates p=0.01 (see subpart 615.05).

Example 04-2 Regression

For the watershed described in example 04-1, land treatment data were also collected. These data included the amount of dairy cow manure applied in the watershed between each runoff event. A linear regression was developed between the concentration of total phosphorus in streamflow and the amount of manure applied in the watershed (fig. 04-1). This regression was significant based on analysis of variance for regression.

This association indicates that total phosphorus concentrations in the stream increase with increasing manure applications.

Figure 04-1 Jewett Brook phosphorus concentration and manure applied in the watershed

**615.0402 Consistency**

Another requirement of causation is that the association between the variables is consistent from population to population in both direction and magnitude (Mosteller and Tukey 1977). To assess consistency, different data sets are needed of the same association. Consistency is shown in example 04-3.

Example 04-3 Consistency

Figure 04-2 shows a relationship between either fecal coliform or fecal streptococcus abundance in Jewett Brook as a function of the percentage of the animal units in the watershed that are being managed with best management practices (BMPs). The major BMP used in this case was manure storage during the winter with spring manure spreading followed by rapid incorporation.

The association for fecal streptococcus was statistically significant, but the association for fecal coliform was not. Fecal coliform abundance appeared to be more variable than fecal streptococcus. To show consistency, compare this association to that derived from other data sets. Figures 04-3 through 04-5 show the association between bacteria abundance and the percent of animal units for three other watersheds in the same vicinity.

In all cases illustrated in figures 04-2 through 04-5, the bacteria abundance in the stream declined as the percentage of animal units being managed with BMPs increased. The same general relationship was observed in the LaPlatte River watershed about 50 miles away (Meals 1990). Ideally, this relationship should be tested across the United States to show consistency.

Figure 04-2 Mean annual bacteria abundance and the percent of BMP animal units for the Jewett Brook watershed (n=6)

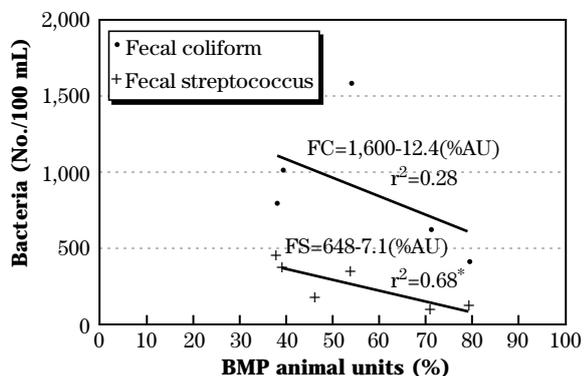


Figure 04-4 Mean annual bacteria abundance and the percent of BMP animal units for the Rugg Brook watershed (n=6)

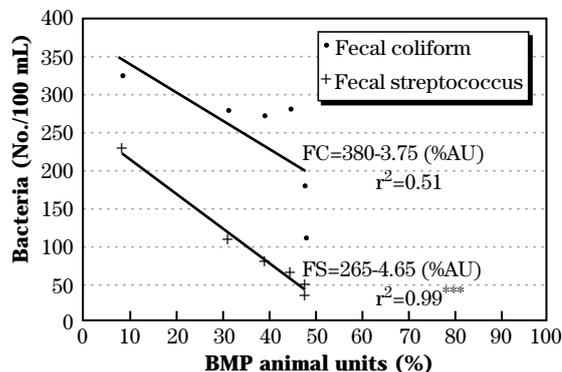


Figure 04-3 Mean annual bacteria abundance and the percent of BMP animal units for the Stevens Brook watershed (n=6)

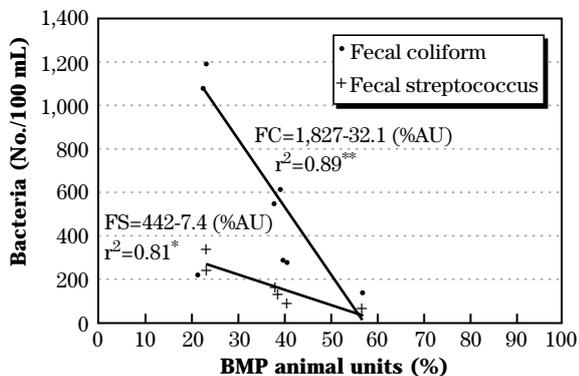
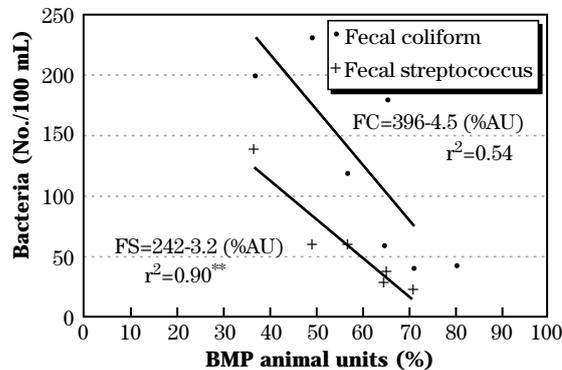


Figure 04-5 Mean annual bacteria abundance and the percent of BMP animal units for the Mill River watershed (n=6)



* Indicates p=0.05
 ** Indicates p=0.01
 *** Indicates p=0.001

615.0403 Responsiveness

Causality is also supported by the concept of responsiveness. By performing an experiment, the dependent variable should respond to manipulation of the independent variables (Mosteller and Tukey 1977). This concept requires that an experiment is performed where we intervene and change the x's and note whether the y's change in a corresponding manner. Example 04-4 illustrates this concept.

Example 04-4 Responsiveness

For the bacteria example, we learned that the bacteria abundance in the streams draining agricultural watersheds was associated to the percent of the BMP animal units. The percent of BMP animal units is actually a surrogate variable for changes that occur in the management of bacteria from animal wastes. Included in these changes are longer storage of manure and incorporation of the manure soon after field application.

At a farm in the St. Albans Bay watershed, a paired watershed study was conducted at a field scale to determine the effect of best manure management on bacteria in runoff. During the calibration period both fields were spread with manure on top of ice and snow during the winter. During the treatment period, the upper field received manure in the winter again, but the lower field was spread with manure in the spring, which was immediately incorporated into the soil. This experiment could determine the change in bacteria abundance in runoff that resulted from the BMP of storage and incorporation. Bacteria abundance in runoff should respond to the application of manure on frozen ground.

615.0404 Mechanism

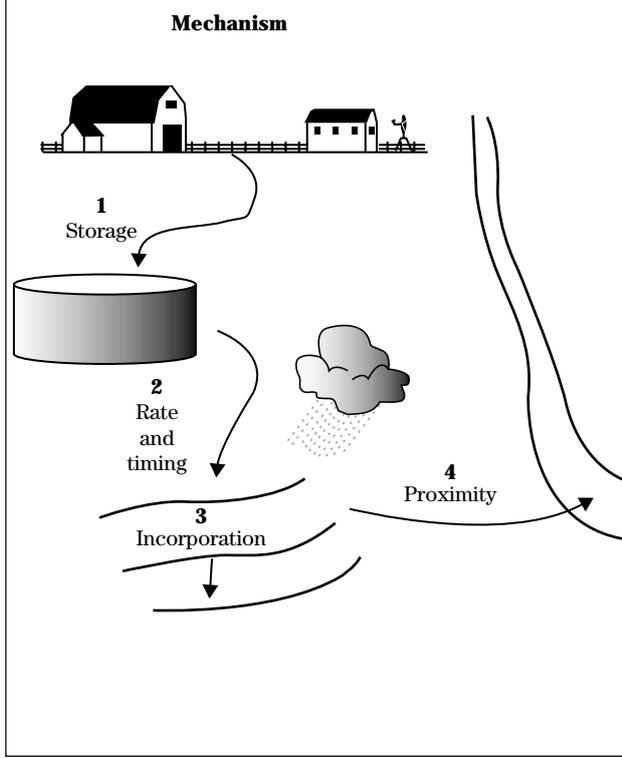
The final requirement for causality is adequate description of a mechanism that provides a step-by-step pathway from the cause to the effect, making the appropriate linkages along the way (Mosteller and Tukey 1977). Example 04-5 illustrates this point.

Example 04-5 Mechanism

Figure 04-6 shows a logical mechanism that explains why bacteria abundance in the example stream may decline after the animal units begin to be managed.

Bacteria would have a tendency to die off, or otherwise decline in abundance, at several points along the pathway. First, bacteria would die off in storage in the manure pit or tank faster than in piled manure (Moore, et al. 1988). Second, the amount and timing of manure applied would be managed based on soil and crop needs. Third, much less manure would be available for runoff if it were incorporated into the soil. Fourth, manure would be applied at a safe distance from the stream off runoff-producing zones. All of these factors should contribute to lower bacteria abundance in streams draining agricultural watersheds that have animal waste BMPs.

Figure 04-6 Mechanism for bacteria decreases



615.0405 Experimental design

Other considerations in analyzing cause and effect depend, in large part, on how the monitoring study is conducted. These factors include the time scale, system level, and reasonableness of treatment.

(a) Time scale

The time scale is important for causality because we all investigate windows within the continuum of time. Numerous temporal cycles, such as diurnal, lunar, seasonal, annual, and astronomical, operate in the natural environment. All these cycles have the potential of influencing our perception of causality. These time scales also influence interpretation of trend data. The timing of flow occurrences during a study can influence our perception of water quality trends. For example, if a wet year occurred early in the study, flow, concentrations, and mass exports would be high during that year. If that year were followed by several years of lower flows, a decreasing trend in flow, concentrations, and mass would be likely.

To avoid or account for problems associated with time scales, the true natural variability must be determined before treatments are imposed. The response observed may be an increase in the variability rather than a change in the mean. Reference watersheds (controls) help account for time scale problems. The experimental design must consider time scale cycles.

(b) System level

Biological systems can be studied at the ecosystem, community, population, individual, cell, and molecular level. Similarly, watersheds (catchments, drainage basins) can be investigated at the watershed, field, and plot level. Because the lower levels of systems are inherently easier to investigate, the tendency is to investigate at a lower level than is needed to answer the question. For example, interest is high in knowing the effect of implementation of BMPs in a watershed on water quality. However, the common approach to investigating cause-and-effect is to look at the effectiveness of an individual BMP on a field or plot basis.

This approach ignores processes that operate on a watershed basis, such as stream transport phenomenon. The project scale should be matched with the objective to avoid misconceptions about the system level being studied.

(c) Reasonableness of treatment

When studying causality, the type of treatment applied should be reasonable and consistent with real world situations. Some treatments may be strong interventions, such as a catastrophe. An example of such treatment is the clearcut and herbicide treatment at Hubbard Brook Experimental Forest (Likens, et al. 1970). Following harvesting, the timber was left on the site and regrowth was prevented with herbicide applications. Stream concentrations increased dramatically in nitrate and cations. By comparison other treatments can be more gradual, such as a change in nutrient management on an agricultural field.

The interaction of the treatment with the environment may be more important than the main effect of the treatment. For example, certain erosion control practices may show no effect during small storms, but may be very effective during the larger, rarer storm events.

A final consideration in causality is understanding the number of variables contributing to a dependent variable. Most water quality issues are multivariate and not univariate. For example, stream phosphorus concentrations may be influenced by precipitation, antecedent moisture, previous stream loading of phosphorus, biological activity, temperature, geologic formation, land activities, and the time available for mineralization. Thus the cause of the level of phosphorus in a stream is potentially the effect of numerous factors that could be considered in the design of the study.

Some causal variables could be unexpected interferences. For example, the midnight dumping of septage, an accidental spill, or routine washing practices at a small point source can create havoc with an experimental design.

615.0406 References

- Draper, N.R., and H. Smith. 1981. Applied regression analysis. (2nd. ed.) John Wiley and Sons, New York, NY.
- Hopkins, R.B., and J.C. Clausen. 1985. Land use monitoring and assessment for nonpoint source pollution control. *In* Perspectives on Nonpoint Source Pollution, U.S. EPA 440/5-85-001, p. 25-29.
- Likens, G.E., F. H. Bormann, N.M. Johnson, D.W. Fisher, R.S. Pierce. 1970. Effects of forest cutting and herbicide treatment on nutrient budgets in the Hubbard Brook watershed-ecosystem. *Ecological Monographs* 40(1):23-47.
- Meals, D.W. 1990. LaPlatte River watershed water quality monitoring and analysis program comprehensive final report. Prog. Rep. No. 12, VT Water Resour. Res. Ctr., Univ. VT, Burlington.
- Moore, J.A., J. Smyth, S. Baker, and J.R. Miner. 1988. Evaluating coliform concentrations in runoff from various animal waste management systems. Agric. Exp. Sta. Spec. Rep. 817, OR State Univ., Corvallis.
- Mosteller, F., and J.W. Tukey. 1977. Data analysis and regression: a second course in statistics. Addison-Wesley Publ. Co., Reading, MA.
- Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615

National Water Quality Handbook



Subpart 615.05 Hypothesis Testing

Subpart 615.05 Hypothesis Testing

Contents:	615.0500	Introduction	05-1
	615.0501	Error types	05-1
	615.0502	One-sample hypotheses	05-2
		(a) One-tailed	05-2
		(b) Two-tailed	05-2
	615.0503	Two-sample hypotheses	05-3
	615.0504	Paired sample hypotheses	05-4
	615.0505	Multisample hypotheses	05-5
	615.0506	Nonparametric hypotheses	05-6
	615.0507	Statistical significance	05-6
	615.0508	Summary	05-7
	615.0509	References	05-7

Tables	Table 05-1	Error types in statistical decisions	05-2
	Table 05-2	Summary of hypotheses by type of data and appropriate test	05-7

Figure	Figure 05-1	Distribution of t showing critical regions	05-2
---------------	--------------------	--	------

Examples	Example 05-1	One-sample hypothesis testing—one tailed	05-3
	Example 05-2	One-sample hypothesis testing—two tailed	05-3
	Example 05-3	Two-sample hypothesis testing—one tailed	05-3
	Example 05-4	Two-sample hypotheses testing—two tailed	05-4
	Example 05-5	Paired sample hypotheses testing—one tailed	05-4
	Example 05-6	Paired sample hypotheses testing—two tailed	05-5
	Example 03-7	Multisample hypotheses testing	05-5

615.0500 Introduction

Developing a hypothesis and testing that hypothesis are fundamental steps in data analysis for water quality monitoring studies. A *hypothesis* is a scientific statement about an assumption regarding the results expected from a study. A *statistical hypothesis* is a statement about a variable describing the distribution of the data, such as the mean (Snedecor and Cochran 1980, Steel and Torrie 1960, Zar 1984). Hypotheses are statements regarding population parameters, not sample statistics. We use hypotheses to draw inferences regarding the assumed population based on sample information. A test of a hypothesis, also termed a test of significance, is a procedure for determining whether a hypothesis should be rejected or accepted (Afifi and Azen 1979).

A *null hypothesis* is the primary hypothesis to be tested and is so termed because it is the hypothesis of no change. The null hypothesis is noted by H_0 . Generally, rejecting the null hypothesis is desirable. An example of a null hypothesis is:

$$H_0: \text{mean (year 1) = mean (year 2)}$$

This seemingly reverse logic exists because data can be collected that can contradict the null hypothesis, but data cannot be obtained to directly accept the hypothesis.

An *alternative hypothesis*, denoted by H_a , is often the hypothesis of interest and is the statement that we may want to assume is true. An example of an alternative hypothesis is:

$$H_a: \text{mean (year 1) } \neq \text{ mean (year 2)}$$

or possibly:

$$H_a: \text{mean (year 1) } < \text{ mean (year 2)}$$

The various types of hypotheses used in water quality studies are described in this subpart. In addition, the consequences of making incorrect hypothesis decisions (error types) and the meaning of statistical significance are described.

615.0501 Error types

When performing a statistical test of a hypothesis, the decision can be wrong because probability, or chance, is involved. Two types of errors can occur. A Type I error can occur when the H_0 is rejected even though it is true (table 05-1). The probability of a Type I error is indicated by α , which is usually a small value that should be decided before the study begins (Steel and Torrie 1960, Zar 1984). This is also termed the statistical significance of the study. Conversely, accepting the null hypothesis when it is true (a correct decision) has the probability of $1-\alpha$, which should be a high value.

A Type II error can occur when the H_0 is not rejected when it should be (table 05-1). The probability of a Type II error is indicated by β . Conversely, the probability of rejecting the null hypothesis when it is false has the probability of $1-\beta$, which is also called the power of the test (Steel and Torrie 1960, Zar 1984).

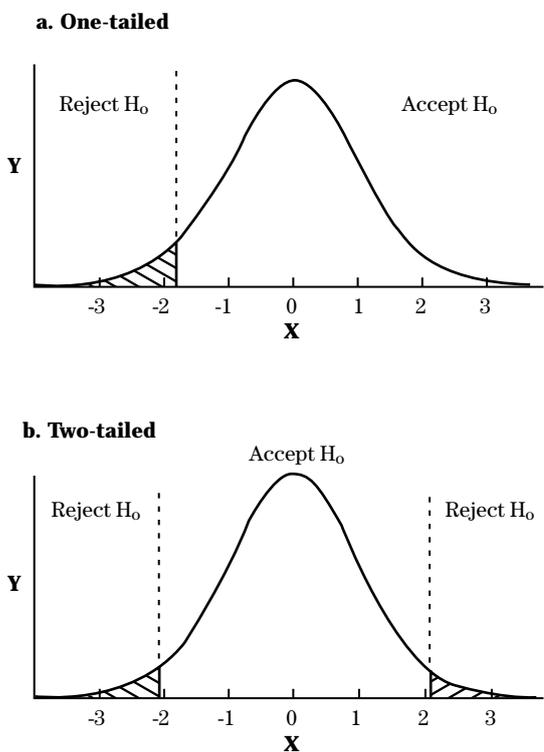
For a given number of samples, α is inversely related to β . This means that if we reduce the probability of rejecting the null hypothesis when it is true (α), we increase the probability of accepting the null hypothesis when it is false (β). Both types of errors can be reduced by larger sample sizes.

Hypotheses will be used throughout the various chapters contained herein. However, some common hypotheses used and their appropriate applications are described further. Hypotheses may be categorized by the number of groups being compared. They are often distinguished as one-sample, two-sample, paired-sample, and multisample (Zar 1984).

Table 05-1 Error types in statistical decisions

Decision	Reality	
	H ₀ is true	H ₀ is false
Reject H ₀	Type I error Prob = α termed <i>significance level</i>	Correct decision Prob = $1 - \beta$ termed <i>power</i>
Accept H ₀	Correct decision Prob = $1 - \alpha$ termed <i>confidence level</i>	Type II error Prob = β

Figure 05-1 Distribution of *t* showing critical regions



615.0502 One-sample hypotheses

A test involving one sample is used when a population parameter (e.g., the mean) is compared to a fixed value that may either be known or hypothesized. Tests can be either one-tailed or two-tailed, depending upon the nature of the problem. These tests are termed one- or two-tailed because they refer to a comparison of a calculated *t* to a critical region of the *t*-distribution at a certain probability. A one-tailed test is used when the mean is to be compared to a fixed value, such as a water quality standard. A two-tailed test is used when the mean could lie on either side of a fixed value.

In figure 05-1a the *t*-distribution is shown for a one-tailed test. If the calculated *t* is greater than the critical *t* (see subpart 615.07 for a definition of *t*), the null hypothesis can be rejected at the probability used. This means that the mean is so different from the fixed value that it lies in the shaded area and has a very small probability of occurring if it were part of the fixed value's population.

The *t* distribution is used rather than the *z* distribution because the population standard deviation (σ) is unknown.

(a) One-tailed

A one-tailed test is appropriate when the mean or some other population parameter is to be compared to some fixed value in a specific direction, such as a water quality standard (Snedecor and Cochran 1980, Zar 1984). We may test that the value is either significantly larger or significantly smaller than the fixed value, but we can only test one direction at a time. See example 05-1 for more information.

(b) Two-tailed

A two-tailed test is appropriate when there is no reason to see whether a value is greater than or less than a fixed value. Therefore, an appropriate null hypothesis would be that the means are identical, and the alternative hypothesis would be that the means are

not equal (Steel and Torrie 1960, Zar 1984). In figure 05-1b, the distribution for t is shown for a two-tailed test. In this case the calculated t can be either positive or negative.

In some cases the appropriate value to compare to the mean might be a zero. This may happen when examining the change in something, such as the change in concentrations before and after some time period. See example 05-2 for more information.

Example 05-1 One-sample hypothesis testing—one-tailed

Implementation of a nutrient management program on cropped fields might be expected to result in reduced ground water $\text{NO}_3\text{-N}$ concentrations below the standard of 10 mg/L. An appropriate null hypothesis would be:

$$H_0: \text{mean NO}_3\text{-N} \geq 10 \text{ mg/L}$$

The alternative hypothesis might be:

$$H_a: \text{mean NO}_3\text{-N} < 10 \text{ mg/L}$$

In this case it is desirable to reject the null hypothesis in favor of the alternative hypothesis.

Example 05-2 One-sample hypothesis testing—two-tailed

When sampling the ground water in a field, we may be uncertain as to whether the $\text{NO}_3\text{-N}$ in the ground water is improving or getting worse over time. An appropriate null hypothesis may be:

$$H_0: \text{mean (year 1)} = \text{mean (year 2)}$$

The alternative hypothesis would be:

$$H_a: \text{mean (year 1)} \neq \text{mean (year 2)}$$

A two-tailed t -test would be appropriate to test these hypotheses. If the calculated t -value was greater than the critical value from a table, then the null hypothesis would be rejected. This t -value could be either positive or negative.

615.0503 Two-sample hypotheses

A two-sample hypothesis is used when testing for the differences between two populations sampled. Often we are testing for the difference between two means; although the two variances could be tested as well. Both one-tailed (example 05-3) and two-tailed (example 05-4) tests are appropriate for two-sample hypothesis; however, the two-tailed test is more commonly used.

Example 05-3 Two-sample hypothesis testing—one-tailed

The nutrient management program described in Example 05-1 could result in a reduced mean concentration of nitrogen in a stream draining the treated watershed. Thus we are interested in detecting a difference in one direction only. The null hypothesis might be:

$$H_0: \text{mean (year 2)} > \text{mean (year 1)}$$

The alternative hypothesis might be:

$$H_a: \text{mean (year 2)} \leq \text{mean (year 1)}$$

If we were less certain about the years, this could be a two-tailed test.

Example 05-4 Two-sample hypotheses testing—two tailed

For long-term trend analysis we may not be certain as to whether the change from year 1 to year 2 might be an increase or a decrease. An appropriate null hypothesis might be:

$$H_0: \text{mean (year 1)} = \text{mean (year 2)}$$

The alternative hypothesis might be:

$$H_a: \text{mean (year 1)} \neq \text{mean (year 2)}$$

These hypotheses could also be stated in terms of their differences. The null hypothesis would be:

$$H_0: \text{mean (year 1)} - \text{mean (year 2)} = 0$$

and the alternative hypothesis would be:

$$H_a: \text{mean (year 1)} - \text{mean (year 2)} \neq 0$$

A *t*-test would be used to test the null hypothesis (subpart 615.07).

615.0504 Paired sample hypotheses

A paired sample hypothesis is appropriate when two samples are associated in some meaningful way. The two-sample hypotheses, described in the previous section, assume that the samples are independent and not associated in some way. For example, comparing the means of monthly observations from one year to the next would be a two-sample test. Months are not paired well from year to year because of climate differences. However, comparing the means of monthly observations from adjacent watersheds for the same year would be a paired sample test. The two adjacent watersheds would be similarly affected by climate from month to month during the year. The paired *t*-test is used to test the null hypothesis (subpart 615.08). Both the one-tailed and two-tailed hypotheses are used with paired comparisons. These tests are illustrated in examples 05-5 and 05-6.

The hypotheses for paired samples are expressed in several ways. One method is to assume that the difference between the means is zero. This is equivalent to stating that the means are equal.

Example 05-5 Paired sample hypotheses testing—one tailed

An erosion control irrigation study was established to determine whether the newer sprinkler irrigation technique results in more than a 1 ton per acre reduction in erosion compared to the older flooded irrigation. To answer the question, paired plots were established with one plot from each pair being irrigated with a sprinkler and the other flooded. An appropriate null hypothesis is:

$$H_0: \text{mean (sprink.)} - \text{mean (flood)} = 1 \text{ ton/acre reduction}$$

The alternative hypothesis is:

$$H_a: \text{mean (sprink.)} - \text{mean (flood)} > 1 \text{ ton/acre reduction}$$

We only wanted to know whether the change in irrigation practice was going to result in less erosion, so a one-tailed test was used.

Example 05-6 Paired sample hypotheses testing—two tailed

For the above-and-below watershed design, samples collected at the above and below stations are associated because of the sampling time; therefore, they should be paired. An appropriate null hypothesis is:

$$H_0: \text{mean (Lower)} - \text{mean (Upper)} = 0$$

The alternative hypothesis would be:

$$H_a: \text{mean (Lower)} - \text{mean (Upper)} \neq 0$$

The paired *t*-test would be used to test the null hypothesis (subpart 615.07).

615.0505 Multisample hypotheses

A multisample hypothesis is used when sampling is from three or more groups. The number of samples taken from each group is not required to be of equal size (unbalanced design). However, equal numbers of samples per group (balanced design) enhance the chance of rejecting the null hypothesis statistically. Example 05-7 illustrates a multisample hypothesis. If two samples are taken, either the *t*-test or ANOVA can be used because they yield identical results.

Example 05-7 Multisample hypotheses testing

For a trend study being conducted over several years, we may be interested in comparing annual means. An appropriate null hypothesis might be:

$$H_0: \text{mean (year 1)} = \text{mean (year 2)} = \dots \text{mean (year k)}$$

The alternative hypothesis might be:

$$H_a: \text{mean (year 1)} \neq \text{mean (year 2)} \neq \dots \text{mean (year k)}$$

Analysis of variance (ANOVA) is used to test the null hypothesis (subpart 615.10) using the *F*-statistic. The test indicates whether all of the population means are different, but not which of those means are different. To answer this question, a multiple comparison test is needed (subpart 615.10).

615.0506 Nonparametric hypotheses

Nonparametric or distribution-free tests have the advantage that they do not assume that the populations are normal or have equal variances (Zar 1984). Nonparametric tests could be used in most cases where a parametric test may be used. A parametric test is better to use than a nonparametric test because it has greater power; that is, the probability of rejecting the null hypothesis is higher when it is false. A greater probability of a Type II error occurs when using nonparametric approaches. Nonparametric tests are described in detail in subsequent subparts.

Most nonparametric approaches require that the data be ranked from either lowest to highest or highest to lowest, and values are assigned the rank of 1, 2, and so forth. The rank, rather than the actual value, becomes the basis of comparison. Ranking eliminates the impact of outliers in the tail regions of distributions.

The actual hypotheses stated will be the same as previously described; however, a nonparametric statistic is used to test the null hypothesis.

615.0507 Statistical significance

The significance level is the probability of committing a Type I error and is denoted as α . By convention, an α of 0.05 is used because it is considered to be a small chance of committing a Type I error. However, in some cases an α of 0.01 is used. The selection of the significance level is somewhat arbitrary. Reporting the level of significance helps the reader in making their own conclusions regarding significance (Zar 1984). The significance level should be decided when the null hypothesis is constructed. Because the significance level is affected by the sample size, a smaller α might be used for a smaller experiment (Steel and Torrie 1960).

The concept of biological significance has two meanings. The first meaning is that a much higher α is acceptable in biological systems because we simply cannot get any better. An α of 0.2 is sometimes acceptable. The second meaning of biological significance is related to the interpretation of results. For example, just because the negative correlation is significant between elevation and abundance of macroinvertebrates, does it mean that high elevation causes lower abundance? This relationship may not have biological significance even though it may have statistical significance.

615.0508 Summary

Table 05-2 provides a summary of the appropriate null hypotheses and statistical test for various data types. In most cases we are interested in comparing means. However, in some cases a comparison of variances may be of greater interest. For example, we may want to know if a particular water quality constituent has become less variable over time.

Table 05-2 Summary of hypotheses by type of data and appropriate test

Data type	Rejection region	Null hypothesis	Test
One-sample	one-tailed	$\bar{x} > x_0$	<i>t</i>
	two-tailed	$\bar{x} > x_0$	<i>t</i>
Two-sample	one-tailed	$\bar{x}_1 > \bar{x}_2$	<i>t</i>
	two-tailed	$\bar{x}_1 = \bar{x}_2$	<i>t</i>
		$\sigma_1 - \sigma_2 = 0$	F ratio
Paired-sample	one-tailed	$\bar{x}_1 - \bar{x}_2 \leq x_0$	<i>t</i>
	two-tailed	$\bar{x}_1 - \bar{x}_2 = 0$	<i>t</i>
		$\sigma_1 - \sigma_2 = 0$	F ratio
Multisample		$\bar{x}_1 = \bar{x}_2 = \bar{x}_k$	F
		$\sigma_1 = \sigma_2 = \sigma_k$	Bartlett's

615.0509 References

- Afifi, A.A., and S.P. Azen. 1979. Statistical analysis: a computer oriented approach (2nd ed). Academic Press. New York, NY.
- Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods (7th ed). The Iowa State Univ. Press. Ames, IA.
- Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., Inc., New York.
- Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.06 Plot Designs

Subpart 615.06 Plot Designs

Contents:	615.0600	Introduction	06-1
	615.0601	Replications	06-1
		(a) Replications per treatment	06-1
		(b) Observations per plot	06-2
	615.0602	Assumptions	06-3
	615.0603	One-way ANOVA	06-3
	615.0604	Two-way ANOVA	06-6
	615.0605	Factorial	06-8
	615.0606	Nonparametric ANOVA	06-8
	615.0607	Multiple mean comparisons	06-9
	615.0608	References	06-9

Tables	Table 06-1	Annual NO ₃ -N export from plots receiving different methods of N fertilizer applications	06-3
	Table 06-2	One-way ANOVA	06-4
	Table 06-3	One-way ANOVA of fertilizer data	06-5
	Table 06-4	Two-way ANOVA	06-6
	Table 06-5	Two-way ANOVA of fertilizer data	06-7
	Table 06-6	Annual NO ₃ -N export from plots receiving different methods of N fertilizer applications (ranks)	06-8

	Table 06-7	Tukey's multiple comparison test of the N fertilizer data	06-10
	Table 06-8	Nonparametric multiple comparison test of the N fertilizer data	06-10
Figure	Figure 06-1	Layout of plot design for fertilizer study	06-3
Examples	Example 06-1	Replications per treatment	06-2
	Example 06-2	One-way ANOVA	06-3
	Example 06-3	Two-way ANOVA	06-6
	Example 06-4	Nonparametric ANOVA	06-8
	Example 06-5	Tukey multiple comparison test	06-9
	Example 06-6	Multiple comparison using Kruskal-Wallis nonparametric ANOVA	06-10

615.0600 Introduction

Plots are generally small areas that are replicated on the land or water. In a plot design, all plots are treated alike except for the factors under study. Data from a plot design are usually organized into multiple data sets corresponding to control plots and treatment plots. A further description of the plot design is in subpart 614.03 of the National Water Quality Handbook (NWQH).

The principal tool for the analysis of plot data is the analysis of variance (ANOVA) (Snedecor and Cochran 1980, Sokal and Rohlf 1969, Steel and Torrie 1960, Zar 1984). Normally, more than two plots are used for plot studies because the treatment applied is replicated. The ANOVA procedure is needed to test multisample hypotheses, such as whether the means of several treatments are different.

When designing a plot study, two of the important decisions are selecting the treatment(s) to be tested and the number of replications for each treatment. Also, the number of observations per plot and whether blocking will be used need to be determined.

This subpart describes the methods used to analyze plot data. Hand calculations and SAS® programs are used to illustrate the statistical methods. Examples of parametric and nonparametric statistics are provided. All possible plot designs are not covered in this chapter. A statistical textbook should be consulted for more complicated designs. These other designs are mentioned in this subpart.

615.0601 Replications

Replications in plot studies can be of two kinds:

- number of replications (plots) per treatment
- number of observations (samples or subsamples) per plot

(a) Replications per treatment

One of the most important initial decisions in a plot study is to determine the number of replications of each treatment to use. Often this decision is based upon economic considerations, such as not enough funding to have more than two replications per treatment. However, such judgments often result in studies with insignificant findings. It is far better to simplify the number of treatments tested rather than sacrifice the number of replications. The number of replications per treatment that would be desired is a function of the variability in the data, the precision desired, and the type of sampling used, as further described in NWQH, part 614, subpart 614.08. Example 06-1 illustrates the selection of the number of replications per treatment.

(b) Observations per plot

A second decision is to determine the number of observations per plot. This decision is partly controlled by the objective of the study. For example, only one annual export value can be obtained from a plot per year, but sampling the soil generally requires that several soil samples be obtained per plot. Having more than one replicate per plot modifies the ANOVA used. In a randomized block design (for an example see figure 03-1, NWQH part 614, subpart 614.03) an interaction term between treatment and block is added, which represents the experimental error. A within-plot sampling error is also determined. The ANOVA for a randomized block design is discussed in several introductory statistics textbooks. Blocking is used when the blocks are believed to be significant. For example, soil type changes across the experimental area could be blocked if soil types contribute to the variability observed in the data being measured. All treatments would be assigned to each block. Blocking is further described in section 615.0604, Two-way ANOVA.

Example 06-1 Replications per treatment

A plot study is being planned to assess the effect of different N fertilizer treatments on the export of $\text{NO}_3\text{-N}$ in water. For this example the export in surface water and ground water are combined into one number. The methods described in NWQH, part 614, subpart 614.08 are used for this calculation, especially equation 08-1, which is repeated here:

$$n = \frac{t^2 s^2}{d^2} \quad [08-1]$$

A published study, similar to the one planned resulted in the following:

mean $\text{NO}_3\text{-N}$ export	=	59 kg/ha
standard deviation	=	7.05 kg/ha
n	=	5

The difference (d) for 10 percent from the mean would be:

$$d = 0.1 \times 59 \text{ kg/ha} = 5.9 \text{ kg/ha}$$

To determine the number of samples needed to estimate the mean value within 10 percent of the true mean, two iterations of equation 08-1 are needed. The *t*-value would be 2.776 for *n*-1 degrees of freedom, where *n*=5 from the published study (appendix A). Using equation 08-1, the following number of replications needed is calculated:

First iteration

$$n = \frac{(2.776)^2 (7.05)^2}{(5.9)^2} = 11$$

For the second iteration the *t*-value is 2.228 (appendix A) at 11-1 = 10 degrees of freedom.

Second iteration

$$n = \frac{(2.228)^2 (7.05)^2}{(5.9)^2} = 8$$

Based on this previous study, eight replications of each treatment would be recommended to estimate a mean value within 10 percent of its true value. If only a 20 percent difference were used, *n* would equal two replications per treatment.

615.0602 Assumptions

Because ANOVA is being used to analyze the plot data, the assumptions associated with ANOVA must be considered. First, it is assumed that the treatments have been assigned randomly to the plots. ANOVA also assumes that the errors are normally distributed, are independent, and have a common variance. Tests to determine if the plot data meet these assumptions are described in detail in subpart 615.03. In cases where the data do not meet these assumptions, you should first try a transformation of the data (subparts 615.02 and 615.03). For example, a log transformation may convert a non-normal distribution to an approximate normal distribution. If the transformation still does not result in meeting the assumptions of the test, then you should consider the use of nonparametric statistics.

615.0603 One-way ANOVA

In a one-way classification we are interested in only the effect of one factor on the water quality variable. To design this type of study, each plot is assigned one of the treatments at random with approximately the same number of plots receiving each treatment. This type of design is also termed a *completely randomized design*. Example 06-2 provides the calculations used to perform a one-way ANOVA of data from this design. This example has one observation per plot.

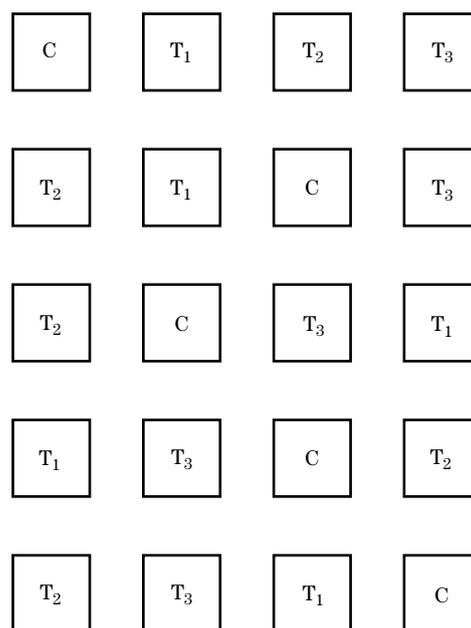
Example 06-2 One-way ANOVA

A plot study was conducted to assess the effect of different N fertilizer treatments on the overall mass export of $\text{NO}_3\text{-N}$ in water. The treatments included spring, split, and spring slow-release applications and a control plot with no fertilizer. There were five replications of each treatment. Figure 06-1 displays the plot layout and assignment of treatments. The data are summarized in table 06-1.

Table 06-1 Annual $\text{NO}_3\text{-N}$ export (kg/ha) from plots receiving different methods of N fertilizer applications

	Control	Spring	Split	Slow release	$\Sigma X_j =$	$\Sigma X_j^2 =$
Block 1	55	64	78	62	259	17,049
Block 2	62	72	91	70	295	22,209
Block 3	49	68	97	67	281	20,923
Block 4	64	77	82	76	299	22,525
Block 5	66	56	85	55	262	17,742
$\Sigma X_i =$	296	337	433	330	1,396	100,448
$\bar{X} =$	59	67	87	66		
$\Sigma X_i^2 =$	17,722	22,969	37,723	22,034	100,448	

Figure 06-1 Layout of plot design for fertilizer study



Example 06-2 One-way ANOVA—Continued

The calculations for a one-way ANOVA are shown in table 06-2. The null hypothesis for this experiment would be:

$$H_0: \bar{X}_1 = \bar{X}_2 = \bar{X}_3 = \bar{X}_4$$

The alternative hypothesis is:

$$H_a: \bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3 \neq \bar{X}_4$$

For the calculations in table 06-2:

X = observation from table 06-1

i = ith treatment

j = jth replication

t = number of treatments

r = number of replicates per treatment

Hand calculations are shown in most beginning statistical books (e.g., Snedecor and Cochran 1980, Sokal and Rohlf 1969, Steel and Torrie 1960, Zar 1984). To perform these calculations by hand, initially determine $\sum X_i$, $\sum X_i^2$, and \bar{X} for each treatment and overall (table 06-1). The additional calculations follow:

Sums of squares

Between treatments:

$$\begin{aligned} SS_{\text{Bet}} &= \frac{\sum X_{ij}^2}{r} - \frac{(\sum X_{ij})^2}{rt} \\ &= \frac{296^2 + 337^2 + 433^2 + 330^2}{5} - \frac{(1,396)^2}{(5)(4)} \\ &= 99,514.8 - 97,440.8 = 2,074 \end{aligned}$$

Total

$$\begin{aligned} SS_{\text{Total}} &= \sum X_{ij}^2 - 97,440.8 \\ &= 100,448 - 97,440.8 = 3,007.2 \end{aligned}$$

Within treatment

$$\begin{aligned} SS_{\text{Within}} &= SS_{\text{Total}} - SS_{\text{Bet}} \\ &= 3,007.2 - 2,074 = 933.2 \end{aligned}$$

Mean squares

$$MS_{\text{Bet}} = \frac{SS_{\text{Bet}}}{df} = \frac{2,074}{4-1} = 691.333$$

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{df} = \frac{933.2}{4(5-1)} = 58.325$$

Table 06-2 One-way ANOVA

Source of variation	Degrees of freedom	Sum of squares (SS)	Mean squares (MS)	F
Between treatments	t-1	$\frac{\sum X_{ij}^2}{r} - \frac{(\sum X_{ij})^2}{rt}$	SS/df	$\frac{MS_{\text{between}}}{MS_{\text{within}}}$
Within treatments	t(r-1)	by subtraction	SS/df	
Total	rt-1	$\sum X_{ij}^2 - \frac{(\sum X_{ij})^2}{rt}$		

Example 06-2 One-way ANOVA—Continued**F-ratio**

$$F = \frac{MS_{\text{Bet}}}{MS_{\text{Within}}} = \frac{691.333}{58.325} = 11.853$$

These calculations are summarized in table 06-3.

To determine whether to reject the null hypothesis of no difference between treatment means, the calculated F ratio is compared to the table F ratio for 3 and 16 degrees of freedom (appendix C). The table F is 3.24 and 5.29 for the 0.05 and 0.01 probability levels, respectively. Because the calculated F exceeds the table F, we can reject the null hypothesis with a 99 percent level of confidence. Therefore, the different fertilizer application methods most likely resulted in a difference in nitrate export. However, which treatments were different are not yet known. To determine which treatment means are different, the methods described in section 615.0607, Multiple mean comparisons, should be consulted.

Using SAS®, the appropriate program would be:
SAS PC Program

```
data nitrate;
    title 'ANOVA of Plot Data';
    infile 'a:nitrate.dat';
    input treat nitrate;
Proc ANOVA;
    class treat;
    model nitrate=treat;
run;
```

Table 06-3 One-way ANOVA of fertilizer data

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Between	3	2074.0	691.333	11.853
Within	16	933.2	58.325	
Total	19	3,007.2		

615.0604 Two-way ANOVA

A two-way classification is useful when we are interested in the effect of two factors on the water quality variable. In plot studies, for example, plots that are adjacent to one another will have a tendency to give more similar results than plots located further away from each other. This may be because of some physical factor, such as soil heterogeneity. Another example would be if up slope plots have the potential to impact downslope plots. To account for this variability, the land can be subdivided into blocks of similar conditions. Blocks are sometimes referred to as replications.

When assigning treatments to the plots, they are assigned randomly within each block with a new randomization for each block. This type of design is termed a *randomized complete-block design*. The primary advantage of this design is that the variability contributed by field differences can be accounted for and eliminated from the treatment effect. Example 06-3 illustrates a two-way ANOVA.

Example 06-3 Two-way ANOVA

For the N fertilizer experiment described in example 06-2, the plots were laid out in the field by placing them across four elevation transects (fig. 06-1). Treatments were randomly assigned to plots across each of the transects. In table 06-1, blocks are represented by rows. The calculations for a two-way ANOVA are shown in table 06-4

Sums of squares

$$\begin{aligned} SS_{\text{Blocks}} &= \frac{\sum X_{ij}^2}{t} - \frac{(\sum X_{ij})^2}{rt} \\ &= \frac{259^2 + 295^2 + 281^2 + 299^2 + 262^2}{4} - \frac{(1,396)^2}{(5)(4)} \\ &= 97,778 - 97,440.8 = 337.2 \end{aligned}$$

Table 06-4 Two-way ANOVA

Source of variation	Degrees of freedom	Sum of squares (SS)	Mean squares (MS)	F
Blocks	r-1	$\frac{\sum_j X_{ij}^2}{t} - \frac{(\sum X_{ij})^2}{rt}$	SS/df	$\frac{MS_{\text{block}}}{MS_{\text{error}}}$
Treatments	t-1	$\frac{\sum_j X_{ij}^2}{t} - \frac{(\sum X_{ij})^2}{rt}$	SS/df	$\frac{MS_{\text{treatment}}}{MS_{\text{error}}}$
Error	(r-1)(t-1)	by subtraction		
Total	rt-1	$\sum X_{ij}^2 - \frac{(\sum X_{ij})^2}{rt}$		

Example 06-3 Two-way ANOVA—Continued

Treatments

$$\begin{aligned} SS_{\text{Bet}} &= \frac{\sum X_{ij}^2}{r} - \frac{(\sum X_{ij})^2}{rt} \\ &= \frac{296^2 + 337^2 + 433^2 + 330^2}{5} - \frac{(1,396)^2}{(5)(4)} \\ &= 99,514.8 - 97,440.8 = 2,074 \end{aligned}$$

Total

$$\begin{aligned} SS_{\text{Total}} &= \sum X_{ij}^2 - 97,440.8 \\ &= 100,448 - 97,440.8 = 3,007.2 \end{aligned}$$

Error

$$\begin{aligned} SS_{\text{Error}} &= SS_{\text{Total}} - SS_{\text{Block}} - SS_{\text{Bet}} \\ &= 3,007.2 - 337.2 - 2,074 = 596 \end{aligned}$$

Mean squares

$$\begin{aligned} MS_{\text{Block}} &= \frac{SS_{\text{Block}}}{df} = \frac{337.2}{5-1} = 84.3 \\ MS_{\text{Bet}} &= \frac{SS_{\text{Bet}}}{df} = \frac{2,074}{4-1} = 691.333 \\ MS_{\text{Error}} &= \frac{SS_{\text{Error}}}{df} = \frac{596}{(5-1)(4-1)} = 49.667 \end{aligned}$$

F-ratio

$$\begin{aligned} F_{\text{Block}} &= \frac{MS_{\text{Block}}}{MS_{\text{Error}}} = \frac{84.3}{49.667} = 1.697 \\ F_{\text{treatment}} &= \frac{MS_{\text{Bet}}}{MS_{\text{within}}} = \frac{691.333}{49.667} = 13.919 \end{aligned}$$

These calculations are summarized in table 06-5.

Table 06-5 Two-way ANOVA of fertilizer data

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Blocks	4	337.2	84.3	1.697
Treatment	3	2074.0	691.333	13.919
Error	12	596.0	49.667	
Total	19	3007.2		

Based upon the ANOVA of the N fertilizer data, the block effect is not significant while the treatment effect was significant as before. A significant block effect would indicate that the design has been made more precise by blocking (Steel and Torrie 1960). Note that in the two-way ANOVA the error mean square has been reduced by apportioning some of the sums of squares to the block effect. This results in an overall higher treatment effect. If blocks are not different, they can be pooled into the error term, which results in an increase in the error degrees of freedom. However, in this example a higher significance was obtained with blocking than without it. Introductory statistical textbooks describe the calculation of the efficiency added by blocking.

Using SAS®, the appropriate program would be:

SAS PC Program

```
data nitrate;
    title 'ANOVA of Plot Data with Blocking';
    infile 'a:nitrate.dat';
    input block treat nitrate;
Proc ANOVA;
    class block treat;
    model nitrate = block treat;
run;
```

615.0605 Factorial

More complicated factorial design, split plot designs, and Latin squares are rare in water quality studies, but common in agronomic and soil investigations. An introductory statistics text should be consulted before planning one of these designs.

615.0606 Nonparametric ANOVA

If data are found to violate the assumptions of normality and especially homogeneous variances, a nonparametric approach may be used (Zar 1984). The Kruskal-Wallis test can be used for a one-way ANOVA. Other similar nonparametric tests, such as Friedman's, exist for a two-way ANOVA and more complicated designs. Because this test is based on rank rather than variance, the test statistic is determined from:

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1) \quad [06-1]$$

where:

- n_i = number of observations in treatment i
- N = total number of observations
- R_i = sum of the ranks for each observation in treatment i

Observations are ranked from low (1) to high (N). The Kruskal-Wallis nonparametric ANOVA for the N fertilizer data is demonstrated in example 06-4.

Example 06-4 Nonparametric ANOVA

For the N fertilizer data described in previous examples, determine the effects of the different fertilizer treatments on $\text{NO}_3\text{-N}$ export using a nonparametric approach (see table 06-6).

If the calculated H is greater than the table H (appendix D, or χ^2 for more than 5 groups) then the null hypothesis is rejected. In this case the table H is 5.78 at $p = 0.05$, and the null hypothesis that the nitrate exports are the same for each treatment is rejected.

$$H = \frac{12}{20(20+1)} \left[\frac{24^2}{5} + \frac{51^2}{5} + \frac{90^2}{5} + \frac{45^2}{5} \right] - 3(20+1)$$

$H = 13.011$

Table 06-6 Annual $\text{NO}_3\text{-N}$ export (kg/ha) from plots receiving different methods of N fertilizer applications (ranks are in parentheses)

	Control	Spring	Split	Slow release
	55 (2)	64 (8)	78 (16)	62 (6)
	62 (5)	72 (13)	91 (19)	70 (12)
	49 (1)	68 (11)	97 (20)	67 (10)
	64 (7)	77 (15)	82 (17)	76 (14)
	66 (9)	56 (4)	85 (18)	55 (3)
R	(24)	(51)	(90)	(45)

615.0607 Multiple mean comparisons

From ANOVA we may have determined that the means are different; however, we do not know which of the means are statistically different from one another. Multiple comparison tests may be used to determine which of the means are different (Zar 1984). Although many such tests exist (e.g.; Duncan, LSD), the Tukey test is recommended for most cases and will be described further in example 06-5. The multiple comparison using the rank sums from the Kruskal-Wallis nonparametric ANOVA is described further in example 06-6.

Example 06-5 Tukey multiple comparison test

For the N fertilizer data, it was determined that the mean $\text{NO}_3\text{-N}$ exports were not equal. Using the Tukey multiple comparison test, determine for which groups the means are different.

First, the standard error is calculated from:

$$SE = \sqrt{\frac{S^2}{n}}$$

where:

- SE = standard error
- S^2 = variance (mean square error from the ANOVA)
- n = number of observations per group

For the example without blocking, the standard error would be:

$$SE = \sqrt{\frac{58.325}{5}} = 3.415$$

615.0608 References

- Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods (7th ed.). The IA State Univ. Press, Ames.
- Sokal, R.R., and F.J. Rohlf. 1969. Introduction to biostatistics. W.H. Freeman and Co., San Francisco, CA.
- Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., Inc., New York.
- Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Second, the means from table 05-1 should be arranged in increasing order and coded with a name or number, such as:

1	2	3	4
59	66	67	87

The statistic q for each possible pair combination is calculated from:

$$q_{\alpha} = \frac{\overline{X}_B - \overline{X}_A}{SE}$$

If the calculated q is greater than the tabular q, the null hypothesis that the means are equal is rejected. The order of comparisons affects the conclusions. Therefore, the largest should be compared with the smallest first, then the second smallest and so on. The calculations are summarized in table 06-7.

The tabular q at $rt-1 = 16$ and $k = 4$ means degrees of freedom and $p = 0.05$ is 4.05 (appendix E). Therefore, group 4 is different from groups 1, 2, and 3, but no other groups are different. These results can be displayed by drawing a line under the groups that are not different, as shown above. More often the

Example 06-5 Tukey multiple comparison test—Continued

means are listed in a table with letters following them and a notation that the means followed by the same letter are not different at $p = 0.05$, as follows:

Treatment	NO ₃ -N Export (kg/ha)
Control	59 a
Spring	67 a
Split	87 b
Slow release	66 a

The conclusion for this study would be that the split application resulted in significantly higher NO₃-N export from the plots than all other treatments, including the control.

Using SAS®, the following statement could be added below the Proc ANOVA statement. The mean values will also be printed.

means tukey;

Table 06-7 Tukey's multiple comparison test of the N fertilizer data

Comparison	Difference	q
4 vs 1	87 - 59 = 28	8.20
4 vs 2	87 - 66 = 21	6.15
4 vs 3	87 - 67 = 20	5.86
3 vs 1	67 - 59 = 8	2.34
3 vs 2	67 - 66 = 1	0.29
2 vs 1	66 - 59 = 7	2.05

Example 06-6 Multiple comparison using Kruskal-Wallis nonparametric ANOVA

A multiple comparison can also be made for a nonparametric ANOVA. The method is similar to that described for Tukey's in example 06-5, but uses the rank sums from the Kruskal-Wallis nonparametric ANOVA (Zar 1984). The standard error is determined from:

$$SE = \sqrt{\frac{n(nk)(nk+1)}{12}} = 13.23$$

where:

n = number of observations per k groups
(Zar 1984)

The rank sums from the table 06-6, rather than the means, are used for arranging the data:

1	2	3	4
24	45	51	90

The q statistic is determined as before. Table 06-8 displays nonparametric multiple comparison test of the N fertilizer data. In this case only the split treatment was higher than the control. There was no difference among all other treatments.

Table 06-8 Nonparametric multiple comparison test of the N fertilizer data

Comparison	Difference	q
4 vs 1	90 - 24 = 66	4.99
4 vs 2	90 - 45 = 45	3.40
4 vs 3	90 - 51 = 39	2.95
3 vs 1	51 - 24 = 27	2.04
3 vs 2	51 - 45 = 6	0.45
2 vs 1	45 - 24 = 21	1.59

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.07 Single Watershed

Subpart 615.07 Single Watershed

Contents:	615.0700	Introduction	07-1
	615.0701	Unpaired comparison of means	07-1
	615.0702	Nonparametric two-sample test	07-5
	615.0703	Presentation of results	07-6
	615.0704	References	07-7

Table	Table 07-1	Mean daily total phosphorus concentrations in watershed runoff from a period before and after implementation of best manure management	07-2
	Table 07-2	Test of normality for the total phosphorus data	07-3
	Table 07-3	Summary of calculations for \log_{10} transformed phosphorus data	07-3
	Table 07-4	Ranks of total phosphorus concentrations for the before and after study of manure management	07-6

Figure	Figure 07-1	Boxplots of phosphorus data	07-6
---------------	--------------------	-----------------------------	------

Examples	Example 21-1	Single watershed analysis	07-2
	Example 21-2	Nonparametric single watershed analysis	07-6

615.0700 Introduction

The single watershed design is used when a single station is monitored both before and after a watershed treatment occurs. As indicated in the National Water Quality Handbook (NWQH), part 614, subpart 614.03, the single watershed design is not recommended because any difference observed is difficult to attribute to the treatment rather than other influences that change over time, such as climate. However, the appropriate statistical approach when such a comparison is made is the unpaired *t*-test of pre and post data. This test actually determines the difference between the effects (Snedecor and Cochran 1980).

Comparisons between groups can be either paired or unpaired (independent). Paired comparisons occur when two samples can be paired in some meaningful way. For example, one pair could constitute an individual watershed measured before and after a treatment. However, in this case there is only one comparison and to make the test meaningful and valid, many watersheds (degrees of freedom) must be compared. It is generally not appropriate to pair observations, such as weekly or monthly data, from a single watershed across years. Because of climate variability, there is no reason to believe that the water quality of the 13th week or for July should be a valid pair across years. Therefore, the unpaired comparison is more common and is presented here.

615.0701 Unpaired comparison of means

The appropriate null hypothesis for the comparison of means is:

$$H_0: \bar{X}_1 - \bar{X}_2 = 0 \text{ or } \bar{X}_1 = \bar{X}_2$$

The appropriate alternative hypothesis would be:

$$H_a: \bar{X}_1 - \bar{X}_2 \neq 0 \text{ or } \bar{X}_1 \neq \bar{X}_2$$

The test of the significance of the difference between the means is based on the *t* distribution where *t* is defined as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_d} \quad [07-1]$$

where:

\bar{X} = the mean for either group 1 or 2

S_d = standard deviation of the difference between the means, which is determined from:

$$S_d = \sqrt{S_p^2 \frac{(n_1 + n_2)}{n_1 n_2}} \quad [07-2]$$

for the case where $n_1 \neq n_2$ and is determined from:

$$S_d = \sqrt{2 \frac{S_p^2}{n}} \quad [07-3]$$

for the case $n_1 = n_2$.

S_p^2 is the pooled sample variance determined from:

$$S_p^2 = \frac{\left[\sum X_1^2 - \frac{(\sum X_1)^2}{n_1} \right] + \left[\sum X_2^2 - \frac{(\sum X_2)^2}{n_2} \right]}{(n_1 - 1) + (n_2 - 1)} \quad [07-4]$$

where:

S_p = pooled standard deviation

S_p is calculated by pooling the individual standard deviations as calculated from equation 01-6 (Steel and Torrie 1960, Zar 1984). The *t*-test is appropriate when

the distributions are normally distributed and have equal population variances. Example 07-1 illustrates the analysis of a single watershed.

Example 07-1 Single watershed analysis

Table 07-1 presents a summary of total phosphorus concentrations in watershed runoff for a before and after study of manure applications. The before period (X_1) occurred during the period when manure was applied to the watershed during the winter on ice and snow. The after period (X_2) represents samples that were taken during the period when manure was applied during the spring and incorporated into the soil. Each value listed in the table is the daily mean of eight 4-hour composite samples.

To determine whether the difference in phosphorus concentrations is significant between the two periods, the appropriate null hypothesis is:

$$H_0: \bar{X}_1 - \bar{X}_2 = 0 \text{ or } \bar{X}_1 = \bar{X}_2$$

The appropriate alternative hypothesis would be:

$$H_a: \bar{X}_1 > \bar{X}_2$$

The t -test assumes that the data are normally distributed and the groups have equal variances, so the data should first be tested for these assumptions.

Table 07-1 Mean daily total phosphorus concentrations (mg/L) in watershed runoff from a period before and after implementation of best manure management

---- Total phosphorus ----
Before (X_1) After (X_2)

(mg/L)	(mg/L)
6.330	0.185
2.166	0.049
0.642	0.040
0.754	0.087
0.728	0.142
0.478	0.060
0.464	0.187
0.444	0.068
0.375	0.043
0.120	0.039
0.086	0.404
0.064	0.110
0.099	0.085
0.054	0.082
0.063	0.138
0.197	1.617
0.088	0.798
0.089	0.104
0.110	0.341
0.105	0.055
0.081	0.295
	0.090
	0.211
	0.151
	0.158
	0.047
	0.029
	0.027
	0.065
	0.152
	0.087
	0.041
	0.544
	0.296

Example 07-1 Single watershed analysis—Continued

Using a statistical package, such as SAS, a test for normality is made as described in subpart 615.03. Table 07-2 shows the test results for the total phosphorus data.

Table 07-2 Test of normality for the total phosphorus data

	Untransformed		Log ₁₀ transformed	
	X ₁	X ₂	X ₁	X ₂
Mean	0.645	0.201	-0.631	-0.934
Median	0.120	0.097	-0.921	1.014
Skewness	3.840	3.690	0.998	0.764
Kurtosis	15.70	15.78	0.531	0.404
W:Normal	0.445	0.559	0.884	0.954
Prob < W	<0.001	<0.001	0.015	0.204

Based on the nonsignificant Shapiro-Wilk W statistic, the data appear to be log-normally distributed. Therefore, the log transformation is used prior to the *t*-test. The next step is to calculate S_d, the standard deviation of the difference between means. Since n₁ does not equal n₂, equation 07-2 is used to calculate S_d. Table 07-3 provides a summary of calculations needed to determine S_d.

Table 07-3 Summary of calculations for log₁₀ transformed phosphorus data

	X ₁	X ₂
n	21	34
ΣX	-13.242	-31.764
ΣX ²	14.595	35.465
Log \bar{X}	-0.631	-0.934
\bar{X}	0.234	0.116

First S_p is calculated from equation 07-4:

$$S_p^2 = \frac{\left[14.595 - \frac{(-13.242)^2}{21} \right] + \left[35.465 - \frac{(-31.764)^2}{34} \right]}{(21-1) + (34-1)}$$

$$= \frac{6.245 + 5.790}{660} = 0.018235$$

S_d is calculated from equation 07-2:

$$S_d = \sqrt{0.018235 \frac{(21+34)}{(21)(34)}} = 0.037479$$

Student's *t* is calculated from equation 07-1:

$$t = \frac{0.631 - (-0.934)}{0.037479} = 8.085$$

From appendix A the table *t*-value is 2.006 for df = (n₁-1)+(n₂-1) = 53 and p = 0.05. Therefore, since the calculated *t* is greater than the table *t*, the H₀ is rejected. The mean is determined on the log-transformed values. Therefore, to transform the mean back to original units, the antilog of the log mean is taken by taking the value 10 and raising it to the power of the log mean.

Based upon the *t*-test, this before and after study determined that the mean phosphorus concentration was significantly reduced by 50 percent after the implementation of the practice as compared to before the practice. Confidence limits can be added to this estimate of differences between means from:

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha} S_{\bar{X}_1 - \bar{X}_2} \quad [07-5]$$

Where the standard error is calculated from:

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \quad [07-6]$$

or for log normal distributions when n is not large, consult page 170 of Gilbert (1987).

Example 07-1 Single watershed analysis—Continued

For this example

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{0.0182}{21} + \frac{0.0182}{34}} = 0.037$$

The confidence limit is:

$$0.234 - 0.116 \pm 2.004(0.037) =$$

$$0.118 \pm 0.074$$

However, because of the limitations of this experimental design, it is possible that the differences are actually the result of some climate difference from the first year to the second. The design does not provide a way to correct for any deterministic features in the data, such as cyclic patterns or rainfall. For example, the change in concentrations might also be caused by a dry year following a wet year.

The SAS® program for the *t*-test in example 07-1 would be:

SAS PC Program

```
Data phos;
  title 'TTest of Phos Data';
  infile 'a:phos.dat';
  input trt phos;
logphos = log10(phos);
Proc TTEST;
  class trt;
run;
```

615.0702 Nonparametric two-sample test

If data violate the assumptions of normal distributions or equal variances, nonparametric or distribution-free approaches may be used (Zar 1984). The Mann-Whitney test is the nonparametric equivalent to the *t*-test for two-samples. As previously described for other nonparametric approaches, the ranks of the values are used rather than the values themselves. Ranking is done from highest to lowest, with the largest value in both groups given a value of 1 and so on.

The Mann-Whitney U statistic is calculated from:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad [07-7]$$

and

$$U' = n_1 n_2 - U \quad [07-8]$$

where:

n = number of samples in each group

R = sum of the ranks for that group (Zar 1984)

If either *U* or *U'* is equal to or greater than the table *U*, the *H*₀ is rejected at the appropriate α .

The data in table 07-1 is used in example 07-2, which is a nonparametric approach for single watershed analysis.

Example 07-2 Nonparametric single watershed analysis

Table 07-4 provides the ranks for the data in table 07-1.

Table 07-4 Ranks of total phosphorus concentrations for the before (X_1) and after (X_2) study of manure management

	X_1	X_2	X_2
	1	20	32
	2	48	17
	7	52	23
	5	36	21
	6	24	49
	9	45	54
	10	19	55
	11	41	42
	13	50	22
	26	53	35
	37	12	51
	43	27	8
	31	38	15
	47	39	
	44	25	
	18	3	
	34	4	
	33	30	
	28	14	
	29	46	
	40	16	
n	21	34	
R	474	1066	

$$U = (21)(34) + \frac{21(21+1)}{2} - 474 = 471$$

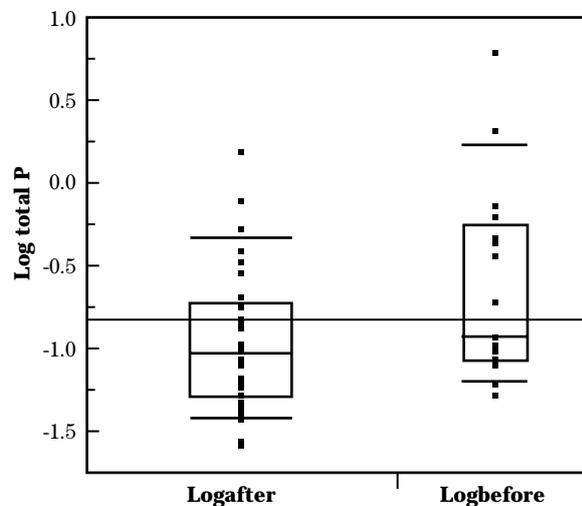
$$U' = (21)(34) - 471 = 243$$

The table value for U is 450 ($\alpha=0.05$) (Zar, 1984). Since the calculated U is greater than the table U, the H_0 of equal concentrations is rejected. Using either parametric approaches with a transformation or nonparametric approaches, the conclusion was that there was a significant difference in the mean concentrations of total phosphorus in runoff.

615.0703 Presentation of results

The presentation of results from a before and after study is generally a presentation of means. Box plots (fig. 07-1) are also an appropriate presentation of the data. The bottom and top of the box represent the 25th and 75th percentiles, the center horizontal line is the medium, and the outer lines are the 10th and 90th percentiles. In some cases time plots of the data can be used; however, since the data are not paired in a meaningful manner, the time plot could result in a misleading interpretation.

Figure 07-1 Box plots of phosphorus data



615.0704 References

Gilbert, R.O. 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York, NY.

Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods (7th ed.). The IA State Univ. Press, Ames, Iowa.

Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., Inc., New York, NY.

Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.08 Above and Below Watersheds

Subpart 615.08 Above and Below Watersheds

Contents:	615.0800	Introduction	08-1
	615.0801	Assumptions	08-1
	615.0802	Paired comparison of means	08-2
	615.0803	Nonparametric paired-sample test	08-4
	615.0804	Presentation of results	08-5
	615.0805	References	08-5

Tables	Table 08-1	Mean daily total phosphorus concentrations in watershed runoff above and below an area receiving manure applications in the winter	08-3
	Table 08-2	Test of normality for the difference in total phosphorus data	08-3
	Table 08-3	Summary calculation for determining the value of t	08-3

Figure	Figure 08-1	Time plot of above-and-below phosphorus concentration data	08-5
---------------	--------------------	--	------

Examples	Example 08-1	Above-and-below watershed analysis	08-2
	Example 08-2	Nonparametric above-and-below watershed analysis	08-4

615.0800 Introduction

The above-and-below design is often thought of as a way to isolate the effect of a treatment. Theoretically, if we sample the water before it flows into an area and then again after it leaves an area, the difference in water quality will be a result of the treatment in the area. In some cases this may be true; however, the difference may be caused by watershed differences as well. An alternative is to conduct an above-and-below study before and after the treatment. Such a study becomes a paired watershed study as described in subpart 615.09.

The above-and-below design is actually one watershed physically nested within another. This design is applicable to streams as well as ground water systems. The appropriate statistical approach is the paired *t*-test of above-and-below data.

This subpart describes the assumptions used for the above-and-below design, provides examples of how to analyze the data using both parametric and nonparametric approaches, and gives examples of how to present the results from the study.

615.0801 Assumptions

The *t*-test assumes that the data are normally distributed and the two groups being compared are of equal variances (Snedecor and Cochran 1980, Steel and Torrie 1960, Zar 1984). If the data fail these assumptions, a transformation or nonparametric approach should be used. One of the conditions of the paired *t*-test is that pairs actually exist. Thus if data are collected at one station, but not the other, no pair exists. Flow occurring at one station, but not at the other still constitutes a pair since one of the values is a zero and the other is above zero. However, when there is no water to measure, a concentration value does not exist and, therefore, a concentrated pair does not exist.

615.0802 Paired comparison of means

The paired comparison of means assumes that the paired values are correlated in some way (Steel and Torrie 1960). Therefore, when one value of the pair was large, we would expect the other value to also be large. The variance is then computed on the difference between paired values rather than on the individual observations as for the unpaired example.

The appropriate null hypothesis of the paired comparison of means is the same as for the unpaired comparison described in subpart 615.07:

$$H_0: \bar{X}_1 - \bar{X}_2 = 0$$

The appropriate alternative hypothesis would be:

$$H_a: \bar{X}_1 - \bar{X}_2 \neq 0$$

The test of the significance of the difference between the means is based on the t distribution (Steel and Torrie 1960, Zar 1984) where t is defined as:

$$t = \frac{\bar{d}}{S_d} \quad [08-1]$$

where:

\bar{d} = the mean of the differences between the paired observations

S_d = standard deviation of the difference between the means, which is determined from:

$$S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n(n-1)} \quad [08-2]$$

where:

d_i = difference between the paired observation

n = number of observation pairs

Example 08-1 illustrates the statistical analysis using the above-and-below process.

Example 08-1 Above-and-below watershed analysis

Table 08-1 presents a summary of total phosphorus concentrations in watershed runoff above and below an area that received winter manure applications on ice and snow. Each value listed in the table is the daily means of eight 4-hour samples. The below data are the same as those listed as before data in table 07-1 in subpart 615.07. This example allows a direct comparison of the single watershed analysis to the above-and-below analysis since the data are real observations from a watershed in Vermont.

Determine whether a significant difference in phosphorus concentrations occurs between the above and below stations. The appropriate null hypothesis is:

$$H_0: \bar{X}_1 - \bar{X}_2 = 0$$

The appropriate alternative hypothesis would be:

$$H_a: \bar{X}_1 - \bar{X}_2 \neq 0$$

Because the t -test assumes that the data are normally distributed and the groups have equal variances, the data should first be tested for these assumptions.

Using a statistical package, such as SAS[®], the data should be examined for normality. As shown in table 08-2, the data appear to be log-normally distributed. Therefore, the log transformation is used before the t -test. To calculate S_d and t , the values in table 08-3 are calculated.

From appendix A, the table t -value is 2.101 for $df = n-1 = 18$ and $p = 0.05$. Therefore, since the calculated t is greater than the table t , the H_0 is rejected. The mean is determined on the log transformed values. To transform the mean back to original units, the antilog of the log mean is obtained by taking the value 10 and raising it to the power of the log mean. If a negative value had been obtained for the difference, a constant would need to be added

Example 08-1 Above-and-below watershed analysis—Continued

to all difference values before a log transformation could be used because the log of a negative number does not exist.

Based upon the *t*-test, this above-and-below study determined that the phosphorus concentration was significantly increased in runoff by 0.173 mg/L as a result of the winter application of manure. However, because of the limitations of this experimental design, it may be possible that the differences may actually be the result of an inherent watershed difference between the upstream and downstream stations.

Table 08-1 Mean daily total phosphorus concentrations (mg/L) in watershed runoff above and below an area receiving manure applications in the winter

----- Total phosphorus (mg/L) -----			
Above	Below	Difference	Rank
0.060	6.330	6.270	19
0.095	2.166	2.071	18
0.117	0.642	0.525	15
0.073	0.754	0.681	17
0.050	0.728	0.678	16
0.034	0.478	0.444	14
0.250	0.464	0.214	11
0.211	0.444	0.233	12
0.090	0.375	0.285	13
0.032	0.120	0.088	10
0.027	0.086	0.059	7
0.076	0.064	-0.012	1
0.058	0.099	0.041	2
0.012	0.054	0.042	3
0.011	0.063	0.052	5
0.056	0.088	0.032	1
0.029	0.089	0.060	8
0.040	0.110	0.070	9
0.049	0.105	0.056	6
0.036	0.081	0.045	4

Table 08-2 Test of normality for the difference in total phosphorus data

	Untransformed	Log transformed
	d	d
Mean	0.629	-0.7634
Median	0.088	-1.0555
Skewness	3.696	0.922
Kurtosis	14.442	0.158
W:Normal	0.449	0.888
Prob < W	<0.001	0.029

Table 08-3 Summary calculation for determining the value of *t*

log (d)	
n	19
$\sum d_i$	-14.5048
$\sum d_i^2$	18.6759
$\log \bar{X}$	-0.763
\bar{X}	0.173 (mean difference in mg/L)

$$S_d^2 = \frac{18.6759 - \frac{(-14.5048)^2}{19}}{19(19-1)} = 0.0222$$

$$t = \frac{-0.763}{0.0222} = -34.369$$

The SAS® program for the *t*-test in example 08-1 would be:

SAS PC Program

```
Data phos;
  title 'TTest of Phos Data';
  infile 'a:phos.dat';
  input phos1 phos2;
diff=phos2-phos1;
logdiff = log10(diff);
Proc MEANS Mean Stderr T PRT;
  Var diff;
run;
```

615.0803 Nonparametric paired-sample test

If the data violate the assumptions of normal distributions or equal variances, nonparametric or distribution-free approaches may be used (Zar 1984) as was used for the unpaired comparison of means in subpart 615.07. The Wilcoxon paired sample test is the nonparametric equivalent to the *t*-test for paired samples. As previously described for other nonparametric approaches, the ranks *of the differences* between the values are used rather than the differences themselves. Ranking is done from lowest to highest with the smallest difference given a value of 1 and so on. The sign of the difference is also carried with the rank. Ranks are summed for both positive (T+) and negative (T-) ranks. The T values are compared to a tabular T value; if either value is less than or equal to the table T value, the H_0 of equal values is rejected.

The data in table 08-1 are used in example 08-2, which illustrates the nonparametric approach to analysis of the above-and-below design data.

Example 08-2 Nonparametric above-and-below watershed analysis

$$T+ = 19 + 18 + \dots + 1 = 209$$

$$T- = 1$$

From appendix G, T at $n = 20$ df and $p = 0.05 = 52$. Since T- is less than the table T, the null hypothesis of equal concentrations above and below is rejected. Using either the \log_{10} transformation or nonparametric approaches, the conclusion was that there was a significant difference in the mean total phosphorus concentrations in runoff at the below station as compared to those at the above station.

615.0804 Presentation of results

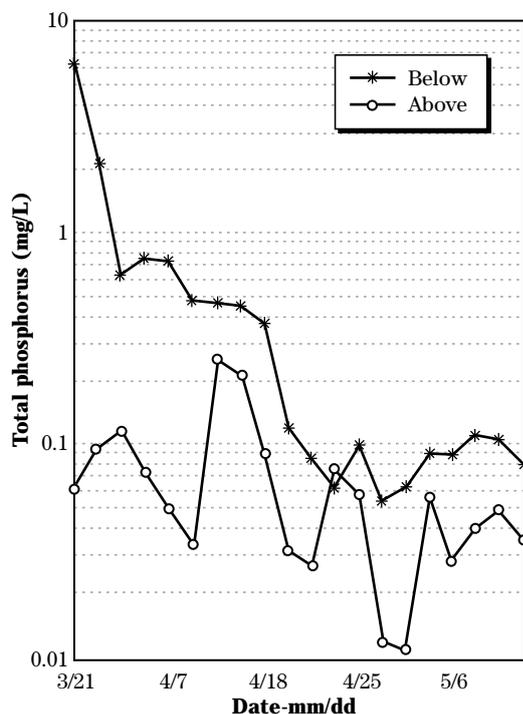
The presentation of results from an above-and-below study is usually a presentation of means. In this case the mean total phosphorus concentration was increased from 0.052 to 0.234 mg/L or 4.5 times. Box plots would be an informative graphic approach to presenting the comparison between above and below data.

In some cases time plots are useful in presenting the results. For example 08-1, the time plot in figure 08-1 reveals that the below station was consistently higher in phosphorus concentration than the above station. The plot also reveals that the difference was greater during the early part of the snowmelt season and became progressively less as time went on.

615.0805 References

- Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods. 7th Ed. The IA State Univ. Press. Ames, IA.
- Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., Inc., New York, NY.
- Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Figure 08-1 Time plot of above-and-below phosphorus concentration data



United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.09 Paired Watersheds

Subpart 615.09 Paired Watersheds

Contents:	615.0900	Introduction	09-1
	615.0901	Calibration	09-2
		(a) Regression significance	09-2
		(b) Calibration duration	09-3
		(c) Residual errors	09-3
	615.0902	Treatment	09-4
	615.0903	Nonlinear/multiple regression	09-5
	615.0904	Displaying results	09-5
	615.0905	References	09-6

Tables	Table 09-1	Schedule of BMP implementation	09-1
	Table 09-2	Analysis of variance for linear regression	09-2
	Table 09-3	Analysis of covariance for comparing regression lines	09-4
	Table 09-4	Analysis of variance for regression of treatment watershed runoff on control watershed runoff	09-7
	Table 09-5	ANOVA for regression of treatment watershed runoff on control watershed runoff for the treatment period	09-8
	Table 09-6	ANCOVA for comparing calibration and treatment regressions	09-8
	Table 09-7	Mean values by period and watershed	09-9
	Table 09-8	Example analysis of covariance for comparing regression lines	09-9

Figures	Figure 09-1	Calibration period regression	09-2
	Figure 09-2	Treatment and calibration period regressions	09-9
	Figure 09-3	Observed deviations from predicted discharge	09-9
Example	Example 09-1	Paired watershed analysis	09-7

615.0900 Introduction

The purpose of this subpart is to describe data analysis for the paired watershed design for conducting nonpoint source (NPS) water quality studies. The monitoring system design requires a minimum of two watersheds—control and treatment—and two periods of study—calibration and treatment (Green 1979, Hewlett 1971, Hewlett and Pienaar 1973, Ponce 1980, Reinhart 1967).

The control watershed accounts for year-to-year or seasonal climate variations. The management practices within the control watershed remain the same during the study. The treatment watershed has a change in management at some point during the study. During the calibration period, the two watersheds are treated identically, and paired water quality data are collected (table 09–1). Such paired data could be annual means or totals, or for shorter studies (<5 yr), the observations could be seasonal, monthly, weekly, or event-based (Reinhart 1967). During the treatment period, one watershed is treated with a best management practice (BMP) while the control watershed remains in the original management (table 09–1). The treated watershed should be selected randomly by such means as a coin toss.

The reverse of this schedule is possible for certain BMPs; the treatment period could precede the calibration period (Reinhart 1967). For example, the study could begin with two watersheds in two different treatments, such as **BMP** and **no BMP**. Later both watersheds could be managed identically to calibrate

them. Since no calibration exists before the treatment occurs, this reversed design is considered risky because you will not find out if the watersheds are properly calibrated until the end of the study.

The basis of the paired watershed approach is that

- The relationship between paired water quality data for the two watersheds is quantifiable.
- This relationship is valid until a major change is made in one of the watersheds (Hewlett 1971). At that time, a new relationship will exist.

This basis does not require that the quality of runoff be statistically the same for the two watersheds. It does require that the relationship between paired observations of water quality remains the same over time except for the influence of the BMP. Often, in fact, the analysis of paired observations indicates that the water quality is different between the paired watersheds. This difference further substantiates the need to use a paired watershed approach. This is because the technique does not assume that the two watersheds are the same; it does assume that the two watersheds respond in a predictable manner together. Example 09–1 illustrates a paired watershed analysis.

Table 09–1 Schedule of BMP implementation

Period	----- Watershed -----	
	control	treated
Calibration	no BMP	no BMP
Treatment	no BMP	BMP

615.0901 Calibration

The relationship between watersheds during the calibration period is described by a simple linear regression equation (fig. 09-1) between the paired observations, taking the form:

$$\text{treated} = b_o + b_1(\text{control}_i) + e \quad [09-1]$$

where:

treated and control = flow, water quality concentration, or mass values for the appropriate watershed

b_o and b_1 = regression coefficients representing the regression intercept and slope, respectively

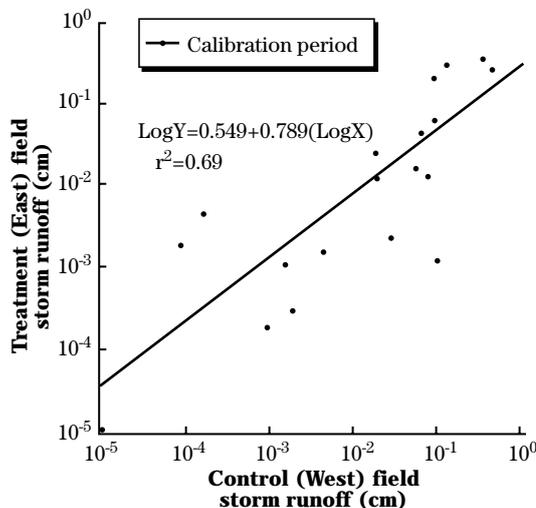
e = residual error

Three important questions must be answered before shifting from the calibration period to the treatment period:

- Is there a significant relationship between the paired watersheds for all parameters of interest?
- Has the calibration period continued for a sufficient length of time?
- Are the residual errors about the regression smaller than the expected BMP effect?

In addition, the observations should cover the full range of observations expected during treatment.

Figure 09-1 Calibration period regression



(a) Regression significance

The significance of the relationship between paired observations is tested using analysis of variance (ANOVA). The test assumes that the regression residuals are normally distributed, have equal variances between treatments, and are independent.

Hand calculations to test for the significance of the relationship are shown in Snedecor and Cochran (1980, p. 157) and in table 09-2. The values for the table are calculated from:

$$S_y^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \quad [09-2]$$

$$S_x^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n} \quad [09-3]$$

$$S_{xy} = \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} \quad [09-4]$$

$$S_{yx}^2 = \frac{S_y^2 - \frac{(S_{xy})^2}{n}}{n - 2} \quad [09-5]$$

Table 09-2 Analysis of variance for linear regression

Source	Degrees of freedom	Sum of squares	Mean squares	F
regression	1	$\frac{(S_{xy})^2}{S_x^2}$	$\frac{(S_{xy})^2}{S_x^2}$	$\left[\frac{(S_{xy}^2)}{S_x^2} \right] / \frac{S_{yx}^2}{S_{yx}^2}$
residual	n-2	$S_y^2 - \frac{(S_{xy})^2}{S_x^2}$	S_{yx}^2	
total	n-1	S_y^2		

Also, the regression coefficients and coefficient of determination are determined from:

$$b_1 = \frac{S_{xy}}{S_x^2} \quad [09-6]$$

$$b_o = \bar{Y} - b_1 \bar{X} \quad [09-7]$$

$$r^2 = \frac{(S_{xy})^2}{S_x^2 S_y^2} \quad [09-8]$$

To perform the calculations by hand, initially calculate:

$$\sum X_i, \sum Y_i, \sum X_i Y_i, \sum X_i^2, \sum Y_i^2, \bar{X}, \bar{Y}$$

The mean squares (MS) are determined by dividing the sum of squares by the degrees of freedom (df).

Using SAS[®], the appropriate program is shown as:

SAS PC Program

```
data flow;
  title 'Total Flow (cm)';
  infile 'fname.dat';
  input flow1 flow2;
  logflow1=log10(flow1);
  logflow2=log10(flow2);
  Proc reg;
    Model logflow2=logflow1
      /P CLM;
run;
```

This program was used to generate table 09-4 in example 09-1.

(b) Calibration duration

Methods for determining whether the length of the calibration period has been sufficient have been described by Wilm (1949), Kovner and Evans (1954), and Reinhart (1967). The ratio between the residual variance (mean squares, S_{yx}^2) for the regression and the smallest worthwhile difference (d) for the treatment watershed is used to determine if a sufficient sample has been taken to detect that difference, from (Kovner and Evans 1954):

$$\frac{S_{yx}^2}{d^2} = \left(\frac{n_1 n_2}{n_1 + n_2} \right) \left[\frac{1}{F(1 + F_{n_1 + n_2 - 2})} \right] \quad [09-9]$$

where:

- S_{yx}^2 = estimated residual variance about the regression
- d^2 = square of the smallest worthwhile difference
- n_1 and n_2 = numbers of observations in the calibration and treatment periods ($n_1 = n_2$ for this calculation because n_2 is not known yet)
- F = table value ($p = 0.05$) for the variance ratio at 1 and $n_1 + n_2 - 3$ df (appendix C)

The difference (d) is selected based on experience and would vary with project expectations. If the left side of the equation is greater than the right side, then the number of samples taken was not sufficient to detect the difference.

(c) Residual errors

The confidence bands for the regression equation allow determining the level of change needed to have a significant treatment effect. In other words, how far away from the calibration regression must the treatment data be to be significantly different? Confidence bands for the regression are determined from:

$$CI = \pm(t)(S_{yx}) \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_x^2}} \quad [09-10]$$

where:

- CI = confidence interval
- S_{yx} = square root of S_{yx}^2
- n and S_x^2 = factors have been previously defined
- t = Student's t
- X_i = value at the point of comparison to compare to the mean on the regression line

Confidence limits can be generated in SAS[®] by adding / P CLM to the MODEL statement.

615.0902 Treatment

At the end of the treatment period the significance of the effect of the BMP is determined using analysis of covariance (ANCOVA). The analysis is actually a series of steps determining:

- significance of the treatment regression equation
- significance of the overall regression that combines the calibration and treatment period data
- difference between the slopes of the calibration and treatment regressions
- difference between the intercepts of the calibration and treatment regressions

The analysis can be computed by hand as shown in table 09-3 (Snedecor and Cochran 1980, p. 386). The summation's symbol (Σ) in table 09-3 is used to signify the addition of the column entries above it.

An example program using SAS® is shown below. This program contains both a test of the treatment regression in the PROC REG statement and a test comparing the regression lines in the PROC GLM statement.

SAS PC Program

```
Proc reg;
    model logflow2=logflow1;
run;
Proc glm;
    class period;
    model logflow2=logflow1 period
    logflow1 * period;
run;
```

Table 09-3 Analysis of covariance for comparing regression lines

Source	df	S_x^2	S_{xy}	S_y^2	b_1	df	SS	MS	F
Within calibration	n_1-1	Eq 09-3	Eq 09-4	Eq 09-2	Eq 09-6	n_1-2	$S_y^2 - \frac{(S_{xy})^2}{S_x^2}$	Eq 09-5	
Within treatment	n_2-1	Eq 09-3	Eq 09-4	Eq 09-2	Eq 09-6	n_2-2	$S_y^2 - \frac{(S_{xy})^2}{S_x^2}$	Eq 09-5	
				Pooled	Error	Σ	Σ	SS/df	
Slopes	$n_1 + n_2 - 2$	Σ	Σ	Σ	Eq 09-6	$n_1 + n_2 - 3$	$S_y^2 - \frac{(S_{xy})^2}{S_x^2}$	Eq 09-5	
			Slope difference			1	Slope SS - Error SS		MS/ Error MS
						1	Combined SS - Slope SS		MS/ Slope MS
Intercepts	$n_1 + n_2 - 1$	combined data				$n_1 + n_2 - 2$	$S_y^2 - \frac{(S_{xy})^2}{S_x^2}$		

615.0903 Nonlinear/ multiple regression

At times the effect of the treatment may be nonlinear. Examples of nonlinear treatment effects include different responses to storm size or gradual vegetation changes. Swindel and Douglass (1984) described approaches for testing nonlinear treatment effects including quadratic approaches and fitting to a gamma distribution. Multiple regression may also be used for paired watershed studies (Hibbert 1969, Snyder 1962).

Regression through the origin can be used where zero flow is expected to occur from both watersheds at approximately the same time. This would occur for adjacent, equally sized watersheds, but not for watersheds of different sizes.

615.0904 Displaying results

The most common methods for displaying the results include a bivariate plot of paired observations together with the calibration and treatment regression equations (fig. 09–2). Another useful graph is a plot of deviations ($y_{\text{observed}} - y_{\text{predicted}}$) as a function of time during the treatment. The predicted values are obtained from the calibration regression equation.

Results should be provided of mean values for each period and each watershed. The overall results caused by the treatment can be expressed as the percent change based on the mean predicted and observed values.

615.0905 References

- Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, New York, NY.
- Hewlett, J.D. 1971. Comments on the catchment experiment to determine vegetal effects on water yield. *Water Res. Bul.* 7(2):376-381.
- Hewlett, J.D., and L. Pienaar. 1973. Design and analysis of the catchment experiment. *In Proc. Symp., Use of Small Watersheds*, E.H. White (ed.), Univ. KY.
- Hibbert, A.R. 1969. Water yield changes after converting a forested catchment to grass. *Water Resour. Res.* 5(3):634-640.
- Kovner, J.L., and T.C. Evans. 1954. A method for determining the minimum duration of watershed experiments. *Trans. AGU* 35(4):608-612.
- Ponce, S.L. 1980. Statistical methods commonly used in water quality data analysis. WSDG Tech. Pap. WSDG-TP-00001, USDA Forest Service, Fort Collins, CO.
- Reinhart, K.G. 1967. Watershed calibration methods. *In Proc. Intern. Symp. on Forest Hydrology*, W.E. Sopper and H.W. Lull. (eds.), Pergamon Press. Oxford. pp. 715-723.
- SAS Institute, Inc. 1986. SAS system for linear models. Cary, NC
- Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods (7th ed.). IA State Univ. Press, Ames.
- Snyder, W.M. 1962. Some possibilities for multivariate analysis in hydrologic studies. *J. Geophysical Res.* 67(2):721-729.
- Swindel, B.F., and J.E. Douglass. 1984. Describing and testing nonlinear treatment effects in paired watershed experiments. *Forest Science* 30(2):305-313.
- Wilm, H.G. 1949. How long should experimental watersheds be calibrated? *Amer. Geophys. Union Trans, Part II*, 618-622.

Example 09-1 Paired watershed analysis

Data from a study in Vermont is used to illustrate the paired watershed approach. The purpose of the study was to compare changes in field runoff as a result of conversion of conventional tillage to conservation tillage. Information included:

- West watershed was the control and was 1.46 hectares (ha) in area.
- East watershed was the treatment field and was 1.10 ha.
- Conventional tillage was moldboard plow whereas conservation tillage was a single disk harrow.
- The calibration period was 1 year during which 49 paired observations of storm runoff were made.
- The treatment period was 3 years during which 114 paired observations of runoff were made.

The assumptions were tested for ANOVA. Data were log-transformed to approach normality based upon the Shapiro-Wilks (W) statistic. The equality of variances between periods was tested using the F-test. Residual plots were examined to check for independence of errors. The statistical package SAS[®] was used for all analyses (SAS 1986).

The regression coefficients of paired observations are calculated by hand as follows:

$$\sum X_i = -123.403$$

$$\sum Y_i = -180.704$$

$$\sum X_i Y_i = 533.553$$

$$\sum X_i^2 = 381.713$$

$$\sum Y_i^2 = 814.847$$

$$\bar{X} = -2.518 \left(10^X = 0.003041 \text{ cm} \right)$$

$$\bar{Y} = -3.688 \left(10^Y = 0.000205 \text{ cm} \right)$$

Therefore,

$$S_y^2 = 148.441$$

$$S_{xy} = 78.463$$

$$S_x^2 = 70.933$$

$$S_{yx}^2 = 1.312$$

The resulting F statistic for this example would indicate that the regression adequately explains a significant amount ($p < 0.001$) of the variation in paired data.

For the example, S_{yx}^2 was 1.312 (from table 09-4), $n_1 = n_2$ was 49, and F was 3.94. A 10 percent change from the mean was considered a worthwhile difference; therefore,

$$d = 0.10 \times \bar{X} = 0.10 \times \log 0.003041 \text{ cm}$$

$$\frac{S_{yx}^2}{d^2} = 20.7$$

The right side of equation 09-9 equals 6. Because 20.7 is greater than 6, the number of observations was not sufficient to detect a 10 percent change in discharge. Enough samples were taken to detect a 20 percent change in discharge:

$$\frac{S^2}{d^2} = 5.2$$

Table 09-4 Analysis of variance for regression of treatment watershed runoff on control watershed runoff

Source	df	MS	F	p
model	1	86.79	66.17	0.0001
error	47	1.31		
total	48			

Example 09-1 Paired watershed analysis—Continued

To perform the calculations for determining analysis of covariance (ANCOVA) by hand, determine the following for the example treatment data:

$$\begin{aligned}\sum X_i &= -358.14 \\ \sum Y_i &= -416.05 \\ \sum X_i Y_i &= 1,408.37 \\ \sum X_i^2 &= 1,352.54 \\ \sum Y_i^2 &= 1,653.43 \\ \bar{X} &= -3.1416 \\ \bar{Y} &= -3.650 \\ n &= 114\end{aligned}$$

Therefore,

$$\begin{aligned}S_y^2 &= 135.00 \\ S_{xy} &= 101.32 \\ S_x^2 &= 227.43\end{aligned}$$

The treatment period regression was found to be significant based on the analysis of variance for regression (table 09-5).

The analysis of covariance obtained in SAS[®] output summarizes the significance of the overall model, compares the two regression equations, the regression intercepts, and the slopes (table 09-6). The ANCOVA indicates that the overall treatment and calibration regressions were significantly different and that the slopes and intercepts of the equations also were different. The difference in slopes is evident in figure 09-2. The slight differences in F values between the hand calculation method and the SAS[®] output are caused by rounding errors.

For the example, the plot of deviations indicates that for most paired observations, the observed value was less than that predicted by the calibration regression equation (fig. 09-3).

In the example, a 64 percent reduction in mean runoff was attributed to the treatment (table 09-7).

The ANCOVA is completed for the example in table 09-8.

Since the slopes were found to be different, the differences in intercepts do not have any real meaning and do not need to be calculated. That is, if slopes are different, intercepts generally are different. However, the calculation for the test of intercepts is presented to show the method. The combined data are determined by summing the $\sum X_i$, $\sum Y_i$, $\sum X_i Y_i$, $\sum X_i^2$, and $\sum Y_i^2$ values for both the calibration and treatment periods and calculating new values for S_y^2 , S_{xy} , and S_x^2 . The calculation of F for the intercept uses the slope MS in the denominator. The F for the slope test uses the error MS in the denominator. A significant difference in intercepts, but not slopes indicates an overall parallel shift in the regression equation.

Table 09-5 ANOVA for regression of treatment watershed runoff on control watershed runoff for the treatment period

Source	df	MS	F	p
model	1	45.13	56.25	0.0001
error	112	0.80		
total	113			

Table 09-6 ANCOVA for comparing calibration and treatment regressions

Source	df	MS	F	p
model	3	43.99	46.17	0.001
error	159	0.95		
overall	1	103.09	108.18	0.0001
intercept	1	5.47	5.74	0.0178
slope	1	23.42	24.58	0.0001

Example 09-1 Paired watershed analysis—Continued

Figure 09-2 Treatment and calibration period regressions

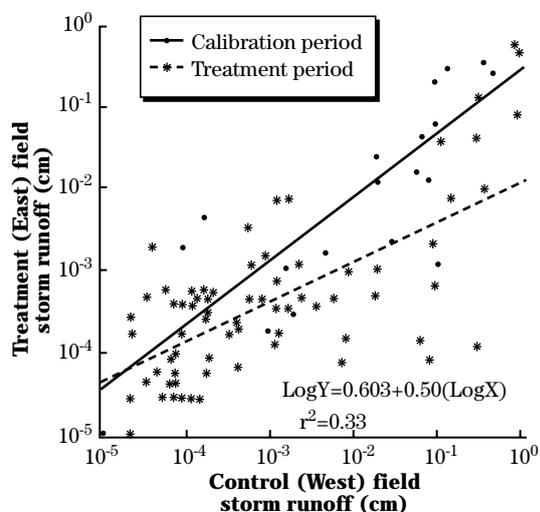


Table 09-7 Mean values by period and watershed

		Runoff (cm) x 10 ⁻²	
Calibration			
Control		0.30	
Treatment		1.63	
Treatment			
Control		0.08	
Treatment		0.04	
Predicted		0.11	-64%

Figure 09-3 Observed deviations from predicted discharge

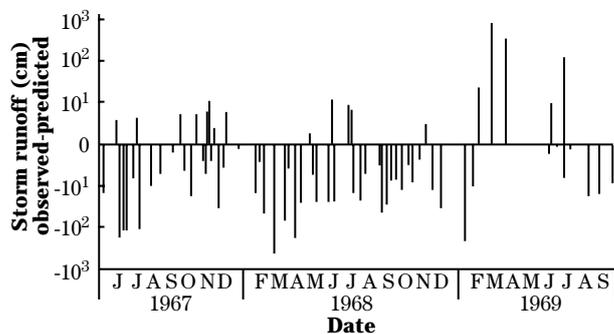


Table 09-8 Example analysis of covariance for comparing regression lines

Source	df	S _x ²	S _{xy}	S _y ²	b ₁	df	SS	MS	F
Within calibration	48	70.933	78.463	148.441	1.106	47	61.650	1.3117	
Within treatment	113	227.430	101.315	135.000	0.445	112	89.866	0.8024	
Error						159	151.516	0.9529	
Slopes	161	298.363	179.778	283.441	0.603	160	175.116	1.0945	
Slope difference						1	23.600	23.600	24.77***
						1	5.8453	5.8453	5.34*
Intercepts	162	311.671	178.762	283.492		161	180.961		

*** indicates significance at p=0.001
 * indicates significance at p=0.05

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.10 Multiple Watersheds

Subpart 615.10 Multiple Watersheds

Contents:	615.1000	Introduction	10-1
	615.1001	Assumptions	10-2
	615.1002	Number of watersheds	10-2
	615.1003	Comparison of groups	10-3
	615.1004	Nonparametric approaches	10-5
	615.1005	Presentation of results	10-6
	615.1006	References	10-6

Tables	Table 10-1	Phosphorus concentration of runoff from paved and unpaved barnyards	10-4
	Table 10-2	Univariate statistics for barnyard phosphorus concentration data	10-4
	Table 10-3	ANOVA for barnyard phosphorus concentration data	10-4
	Table 10-4	Mean total phosphorus concentrations of runoff from the paved and unpaved barnyards	10-4
	Table 10-5	Wilcoxon rank-sum test for the barnyard phosphorus data using JMP	10-5

Figures	Figure 10-1	Box plots of the paved and unpaved barnyard phosphorus concentration data	10-5
----------------	--------------------	---	------

Examples	Example 10-1	Parametric multiple watershed analysis—chemical	10-3
	Example 10-2	Nonparametric multiple watershed analysis using the phosphorus barnyard data	10-5

615.1000 Introduction

The purpose of this subpart is to describe data analysis for the multiple watershed approach for conducting nonpoint source water quality studies. The multiple watershed approach is a study involving more than two watersheds in the design. Wicht (1967) described this approach that was intended to overcome some of the disadvantages of the paired watershed approach. These disadvantages included:

- Inability to always find a stable control watershed
- Uncertainty in predicting the length of the calibration period
- Risk that meteorological conditions may change at the same time as when treatment begins
- Progressive long-term response, such as during major land use changes

In addition, extrapolation of the results from paired watershed studies to broader areas or regions can be questioned, and there is no true replicate in paired watershed investigations.

For the multiple watershed approach, the treatments are intended to be applied to a series of watersheds that have comparable geology, topography, and initial vegetative cover, and are subject to the same or related uncontrolled climate influences (Wicht 1967).

Striffler (1965) also described a multiwatershed method that used multiple regression analysis to assess the relationship between a dependent variable, such as sediment yield, and several independent variables, such as watershed area, soil or vegetative types, and precipitation. Many watersheds selected represent different levels for the independent variables. A major advantage of such an approach is that a large range of watershed conditions is being sampled. Sampled watersheds also can vary in size and other characteristics, such as varying levels of a disturbance.

However, a different approach is more appropriate for nonpoint source pollution studies. Watersheds that have the treatment already in place could be selected across a region of interest. The size of the region would be dictated by the objectives of the study, but could be as large as a state or perhaps limited to an

ecoregion or smaller unit. Once the watersheds were selected, sampling of the appropriate water quality variables would be conducted over a period of time. Clausen and Brooks (1983a, b) used such an approach when comparing the water quality associated with different types of wetlands and when comparing mined to unmined bogs.

This subpart describes the assumptions made in a multiple watershed experiment, presents examples of how to analyze the data from such designs using both parametric and nonparametric approaches, and gives examples of how to present the results from the study.

615.1001 Assumptions

The primary statistical approach for comparing groups of watersheds is the analysis of variance. Therefore, the assumptions made are the same as those previously described for analysis of variance (ANOVA). The major assumptions are:

- Water quality data are sampled randomly.
- Data come from a normal distribution.
- Variances are homogeneous across groups.
- Experimental errors are independently distributed.
- Treatment effects are additive.

The approaches used to test these assumptions are described in subpart 615.03. When using nonparametric approaches, the assumption of normality is no longer appropriate.

615.1002 Number of watersheds

One of the first decisions to make when designing a multiple watershed monitoring study is determining the number of watersheds in each group to monitor. Part 614, subpart 614.08, National Water Quality Handbook describes procedures for estimating the number of sampling units for water quality monitoring. The basic requirements are knowledge of the variance among watersheds and a desired precision to achieve in the study. Clausen and Brooks (1983a) found that 15 watersheds of each type were sufficient to determine differences in the water quality of different peatland types.

615.1003 Comparison of groups

The original analytical approach suggested was a series of paired comparisons between different pairs of watersheds using covariance analysis as for the paired watershed technique (Wicht 1967). Both parametric and nonparametric approaches can be used to compare the results from several groups. The methods of analysis are quite similar to those used for plot studies (part 615, subpart 615.06). Example 10-1 illustrates parametric method of data analysis.

Example 10-1 Parametric multiple watershed analysis

The multiple watershed approach was used to assess the water quality effects associated with paving dairy barnyards in Vermont. The objective of the study was to determine the effect of paving on runoff water quality within a 26,000-acre watershed. Five paved and five unpaved barnyards were sampled for runoff on an event basis nine times over 1 year. During each rainfall event, one or two grab samples were collected from each barnyard. Samples were analyzed for phosphorus (table 10-1), nitrogen, and suspended solids; however, only the total phosphorus concentration data are used in this example. Missing concentration data occurred during the study when either there was no runoff or the sample was destroyed during the analysis process.

Using PROC UNIVARIATE in SAS® (SAS 1995) the phosphorus concentration data were found to be log normally distributed (table 10-2). A P-value <0.05 for the unlogged data (i.e., the data prior to log transformation) indicated that the distribution may be significantly different from a normal distribution based on the Shapiro-Wilk W-statistic.

PROC ANOVA was used to test the null hypothesis that the mean phosphorus concentrations were the same in runoff from the paved and unpaved barnyards. The resulting ANOVA (table 10-3) indicated that there was a significant difference between barnyard types, and the null hypothesis is rejected.

The log mean and antilog mean phosphorus concentrations for the barnyard data are reported in table 10-4. The antilog was obtained by taking 10 to the power of the log value. These results indicate that the paved barnyards in this watershed had runoff phosphorus concentrations that were about two times greater than that in runoff from the unpaved barnyards.

Example 10-1 Parametric multiple watershed analysis—Continued

Table 10-1 Phosphorus concentration of runoff from paved and unpaved barnyards

Date	Paved	Unpaved	Date	Paved	Unpaved
----- mg/L -----			----- mg/L -----		
6/12	20.20	1.90	3/20	12.40	90.00
	67.80	16.10		50.00	22.30
	3.40	4.90		13.30	---
	38.20	5.10		192.50	17.70
	25.70	23.30		132.50	29.00
6/13	36.70	7.20	6/2	13.40	13.40
	132.7	14.70		52.00	13.80
	12.20	25.50		17.00	36.70
	80.70	7.20		134.30	8.60
	32.70	20.30		7.40	15.03
9/27	22.00	53.00	6/8	10.30	17.70
	---	18.20		105.50	27.60
	19.80	40.85		47.30	18.80
	59.20	23.30		68.80	6.40
	---	35.90		17.20	19.10
10/5	22.90	13.30	8/7	---	19.00
	54.15	19.00		63.35	26.00
	38.30	44.40		93.02	---
	73.70	14.60		86.68	9.80
	96.60	43.10		83.02	22.20
11/5	82.27	27.78			
	50.75	25.11			
	47.01	22.10			
	44.34	7.48			
	35.79	18.16			

Table 10-2 Univariate statistics for barnyard phosphorus concentration data

	Unlogged	Log10
Skewness	2.04	-0.17
Kurtosis	4.68	-0.02
W-statistic	0.777	0.986
P-value	<0.001	0.877

Table 10-3 ANOVA for barnyard phosphorus concentration data

Source of variation	Degrees freedom	Sum of squares	Mean squares	F	P > F
Between	1	2.627	2.627	21.53	<0.0001
Within	83	10.127	0.122		
Total	84	12.754			

Table 10-4 Mean total phosphorus concentrations of runoff from the paved and unpaved barnyards

	Log mean	Mean
----- mg/L -----		
Paved	1.4099	25.70
Unpaved	1.0927	12.38

615.1004 Nonparametric approaches

The nonparametric approaches described in subparts 615.06 to 615.08 are also appropriate for multiple watersheds data analysis. For the comparison of two groups, the Mann-Whitney or Wilcoxon rank-sum test may be used. For the comparison of more than two groups, the Kruskal-Wallis nonparametric analysis of variance may be appropriate.

Example 10–2 uses the Wilcoxon rank-sum test for the barnyard phosphorus data analyzed in example 10–1. From the previous example it was determined that the data were not normally distributed, which serves as justification for performing nonparametric analysis.

Example 10–2 Nonparametric multiple watershed analysis using the phosphorus barnyard data

For the data in example 10–1, test the null hypothesis that the median phosphorus concentrations are the same for the paved and unpaved barnyards. The alternative hypothesis would be that the median concentrations are different.

Using JMP (SAS 1995), the box-and-whisker plots in figure 10–1 were obtained. This boxplot shows the median, the 25th and 75th quartiles framing the box, and two lines indicating the 10th and 90th percentiles.

Output for the Wilcoxon rank-sums test is given in table 10–5. This analysis indicates that the medians are significantly different and the null hypothesis is rejected. The median phosphorus concentration for the paved barnyard runoff of 47.2 mg/L was 2.5 times greater than the median of 19.0 mg/L for the unpaved barnyard. These results are similar to the parametric results presented in example 10–1.

Figure 10–1 Boxplots of the paved and unpaved barnyard phosphorus concentration data

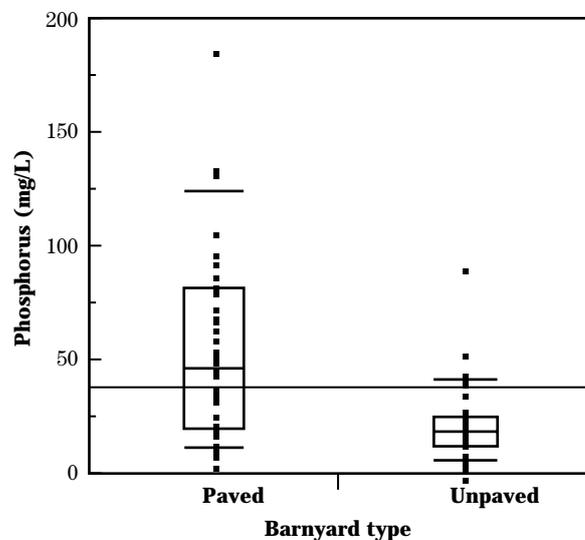


Table 10–5 Wilcoxon rank-sum test for the barnyard phosphorus data using JMP

Level	Count	Score sum	Score mean	Mean-mean0 Std0
Paved	42	2273.5	54.1310	4.105
Unpaved	43	1381.5	32.1279	-4.105

2-sample test, normal approximation

S	Z	Prb> Z
2273.5	4.10501	0.0000

1-way test, chi-square approximation

ChiSquare	DF	Prob>ChiSq
16.8872	1	0.0000

615.1005 Presentation of results

The presentation of results depends in part on the number of groups being compared. However, side-by-side boxplots, as shown in figure 10–1, are a favored method of presenting results because they display graphically the data distributions. When viewing boxplots, if the boxes do not overlap each other, the groups are usually different.

615.1006 References

- Clausen, J.C., and K. N. Brooks. 1983a. Quality of runoff from Minnesota peatlands: I. a characterization. *Water Resour. Bul.* 19(5):763-767.
- Clausen, J.C., and K.N. Brooks. 1983b. Quality of runoff from Minnesota peatlands: II. a method for assessing mining impacts. *Water Resources Bul.* 19(5):769-772.
- SAS Institute, Inc. 1995. *JMP statistics and graphics guide*. Ver. 3.1, SAS Institute, Inc. Cary, NC.
- Striffler, W.D. 1965. The selection of experimental watersheds and methods in disturbed forest areas. *Publ. No. 66, I.A.S.H. Symp. of Budapest*, pp. 464-473.
- Wicht, C.L. 1967. The validity of conclusions from South African multiple watershed experiments. *In Forest Hydrology*, W.E. Sopper and H.W. Lull (eds.), *Proc. Intern. Symp. on Forest Hydrol.*, Pergamon Press, Oxford.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.11 Trend Analysis

Subpart 615.11 Trend Analysis

Contents:	615.1100	Introduction	11-1
	615.1101	Missing data	11-1
	615.1102	Time plots	11-2
	615.1103	Least squares regression	11-2
	615.1104	Comparison of annual means	11-5
	615.1105	Cumulative distribution curves	11-5
	615.1106	Q-Q plots	11-6
	615.1107	Double mass analysis	11-6
	615.1108	Paired regression analysis	11-7
	615.1109	Nonparametric approaches	11-7
		(a) Kendall tau	11-7
615.1110	Summary	11-9	
615.1111	References	11-9	

Tables	Table 11-1	Univariate statistics for fecal coliform data	11-3
	Table 11-2	Analysis of variance for regression of log fecal coliform over time	11-4
	Table 11-3	T-test of slope different from zero for fecal coliform trend data	11-4
	Table 11-4	Analysis of variance across years for fecal coliform data	11-4

Table 11-5	Univariate statistics for fecal coliform data for 1984 and 1989	11-6
Table 11-6	Fecal streptococcus data for August from Jewett Brook, St. Albans Bay Watershed, VT	11-8
Table 11-7	Summary of pluses and minuses for calculation of Kendall's tau for the August fecal streptococcus data for Jewett Brook, St. Albans Bay watershed, VT	11-8
Table 11-8	Summary of trend detection techniques	11-9

Figures

Figure 11-1	Fecal Coliform bacteria in Jewett Brook in the St. Albans Bay watershed, Vermont	11-2
Figure 11-2	Frequency histograms and box plots for fecal coliform data from JMP	11-3
Figure 11-3	Regression of log fecal coliform data over time	11-4
Figure 11-4	Annual boxplots for fecal coliform data	11-4
Figure 11-5	Cumulative frequency curves for the fecal coliform data	11-5
Figure 11-6	Q-Q plot of \log_{10} fecal coliform data for 1984 and 1989	11-6
Figure 11-7	Double-mass analysis of fecal coliform data	11-7

Examples

Example 11-1	Determination of least squares regression of the fecal coliform data over time for Jewett Brook	11-3
Example 11-2	Kendall's tau for August fecal streptococcus data	11-8

615.1100 Introduction

Several techniques have been applied to detect trends in water quality data. A trend as used in this subpart is intended to mean a persistent increase or decrease in a hydrologic or water quality variable over time (Erlebach 1978). Trend analysis methods range from the simple to the complex. Different techniques can be used to select different types of trends, such as monotonic and step trends. *Monotonic trends* are continuous increases or decreases over time (Helsel and Hirsch 1992). *Step trends* are comparisons of two non-overlapping periods of data, perhaps caused by some intervention or time gap between the two periods. Trends may also be persistent or not persistent, and some trends may exhibit seasonality.

Some trend detection techniques require a continuous time-series of data. Thus, interruptions in the temporal data set must be eliminated for these detection techniques. Several methods are available for replacing missing data.

The true first step in trend analysis is actually exploratory data analysis (EDA) as described in subpart 615.02. Thus, the data should be examined using such techniques as stem-and-leaf diagrams and box-and-whisker plots. Transformations, such as the log transformation, of the data may be needed to bring out the trend as well as to meet certain statistical assumptions. Finally, some smoothing approaches may be useful in detecting trends.

In this subpart several techniques for trend detection are presented along with examples. Both parametric and nonparametric approaches are used. Generally, more than one trend method should be used when evaluating water quality data. The different techniques show trends in different ways. The existence of a trend does not mean causality. In fact, a major weakness of relying on trend analysis for an experimental design is that no causality can be inferred from a trend alone. The trend must be explained by other data in conjunction with the trend data. Hipel and McLeod (1994) present methods for testing causality between two time series.

615.1101 Missing data

Several techniques are used for dealing with missing data in a water quality data set. They include linear interpolation, regression analysis, and seasonal adjustment modeling. Linear interpolation may be appropriate if only one or two adjacent data points are missing. The missing data could be estimated by a linear interpolation between the known values before and after the gap. Regressions between water quality observations at different stations or between a water quality variable and flow may also be used to fill in missing data (Dunne and Leopold 1978). The gap in the missing data can be filled using the regression and the known independent values.

In seasonal adjustment modeling, the data are broken up into long-term, seasonal, and nonseasonal components (McLeod, et al. 1983). A missing data point is calculated from an equation representing the summation of influences related to long-term (median), stable seasonal (e.g., monthly), and irregular nonseasonal components.

615.1102 Time plots

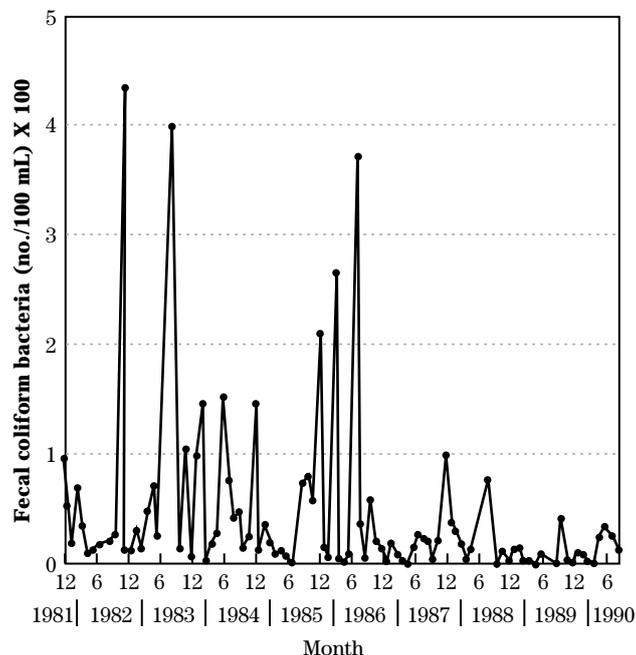
Perhaps the first step in trend analysis is to plot the data versus time. Figure 11-1 is a time plot of fecal coliform bacteria counts obtained from Jewett Brook in the St. Albans Bay watershed from 1981 to 1990.

The fecal coliform data suggest that there is a monotonic decrease in fecal coliform abundance over time. The data also indicate that the variance of the data about this trend is also decreasing. A log transformation could be used to decrease the differences in the variance over time. The data may also show seasonality, but such variations are not obvious. Using just a time plot, the rate of the decrease cannot be obtained.

615.1103 Least squares regression

A parametric regression analysis can be used to test the null hypothesis that the slope of the regression is 0 (i.e., no trend). This test requires the assumptions of normality, constant variance, and independence of errors. In example 11-1 the fecal coliform data previously described are tested using regression.

Figure 11-1 Fecal coliform bacteria in Jewett Brook in the St. Albans Bay watershed, Vermont



Example 11-1 Determination of least squares regression of the fecal coliform data over time for Jewett Brook

Figure 11-2 contains frequency histograms and box plots for the fecal coliform and \log_{10} fecal coliform data. The distribution and boxplots suggest that the untransformed data are not normally distributed and that the \log_{10} transformed data are normally distributed. The univariate statistics are shown in table 11-1. Since the P-value for the untransformed data is less than 0.05, the data are not normally distributed. With a \log_{10} transformation, the data appear to be normally distributed and the \log_{10} transformed will be used for further analysis. Tests for normality are described in detail in subpart 615.03.

Figure 11-3 is a plot of the \log_{10} transformed fecal coliform data as a function of month including a regression line.

The following linear regression equation was obtained using the statistical package JMP (SAS 1995):

$$\text{Log fecal} = 2.673 (\text{month}) - 0.0074$$

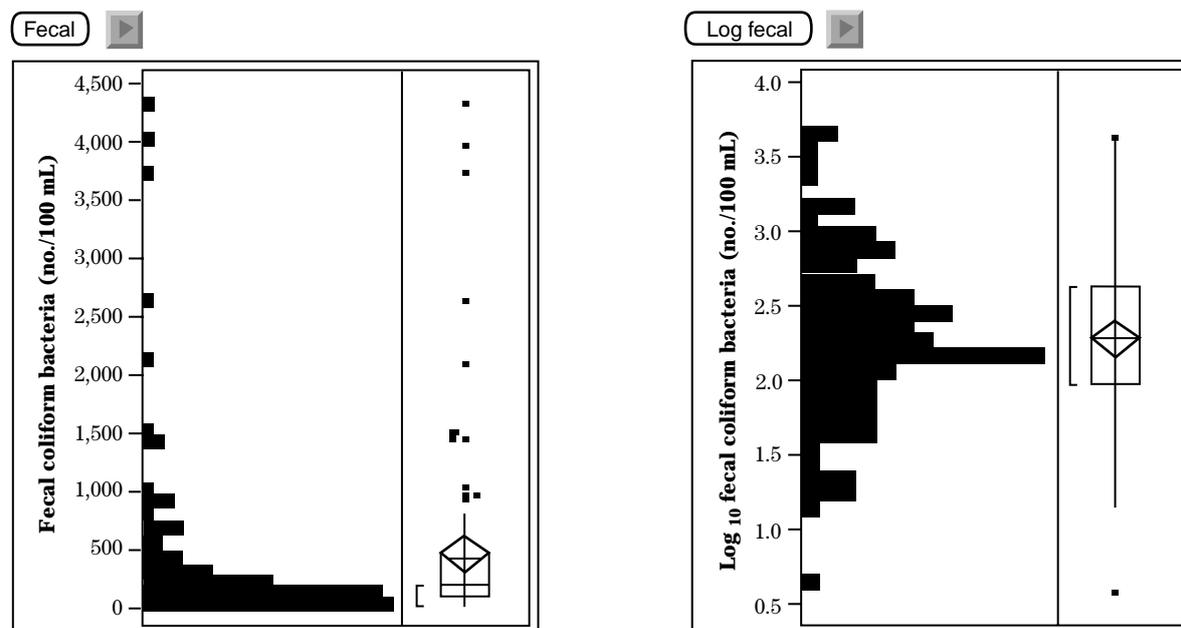
The analysis of variance for the regression is shown in table 11-2. The ANOVA indicates that the regression is significant. The H_0 : slope = 0 is rejected. Also, based on the *t*-statistic, the slope of the regression is significantly different from zero. The results of the *t*-test are shown in table 11-3.

Using the slope of -0.0074 , the fecal coliform bacteria are decreasing at a rate of 0.98 colonies per month (antilog of -0.0074).

Table 11-1 Univariate statistics for fecal coliform data

	Untransformed	\log_{10} transformed
Mean (No./100 mL)	458	2.285
Median (No./100 mL)	190	2.280
Shapiro-Wilk W	0.550	0.985
P<W	0.000	0.786

Figure 11-2 Frequency histograms and box plots for fecal coliform data from JMP



Example 11-1 Determination of least squares regression of the fecal coliform data over time for Jewett Brook—Continued

Annual means were used for the fecal coliform data. Boxplots for each year are shown in figure 11-4. Generally, years might be expected to be different when the boxes do not overlap each other, as for 1984 versus 1989. An analysis of variance indicates that the means are different at $p=0.05$ (table 11-4).

To determine which means were different, annual means were compared using the Tukey-Kramer honestly significant difference (HSD) test (SAS 1995). Only the means for 1984 and 1989 were different. In this example the comparison of annual means does not show a definite trend, but rather a high year early and a low year later. Additional methods of trend analysis are recommended to further analyze the data.

Table 11-2 Analysis of variance for regression of log fecal coliform over time

Source	DF	Sums of squares	Mean squares	F
Model	1	4.750	4.750	16.378
Error	94	27.265	0.290	
			P>F	0.0001

Table 11-3 T-test of slope different from zero for fecal coliform trend data

Term	Estimate	Std error	t ratio	Prob>t
Intercept	2.673	0.111	24.18	0.0000
Month	-0.0074	0.002	-4.05	0.0001

Table 11-4 Analysis of variance across years for fecal coliform data

Source	DF	Sums of squares	Mean squares	F
Model	9	6.626	0.762	2.494
Error	86	25.390	0.295	
			P>F	0.0139

Figure 11-3 Regression of log fecal coliform data over time

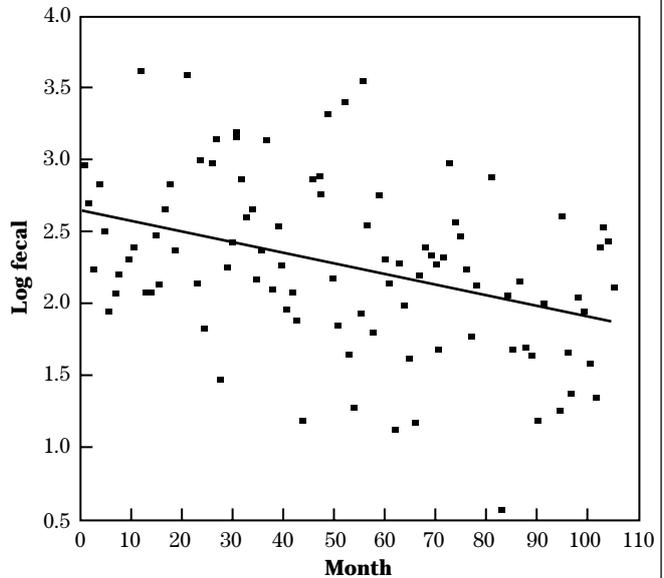
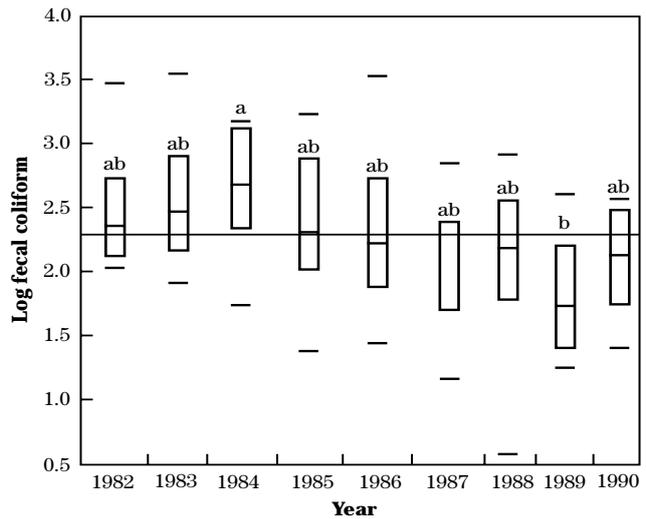


Figure 11-4 Annual boxplots for fecal coliform data ^{1/}



^{1/} Boxes with the same letter are not significantly different at $p=0.05$.

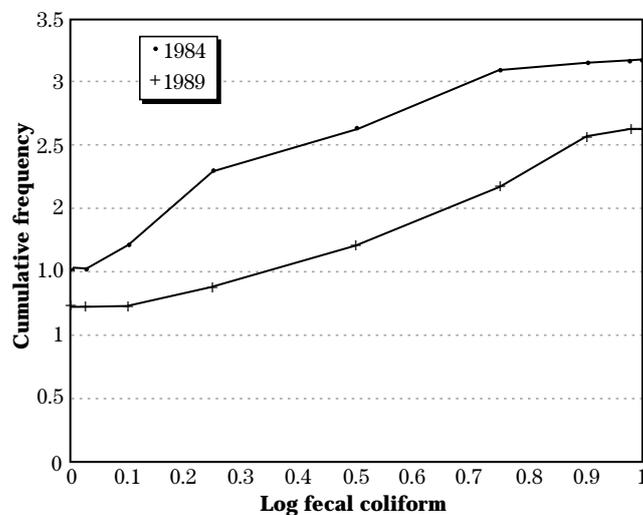
615.1104 Comparison of annual means

A comparison of means may be used to infer trends. Means across years, or some other unit of time such as every 2 or 3 years, may be compared for the analysis. The decision of what time unit to use is based partly on the degrees of freedom for the time unit as well as some scientific reasoning for dividing the time series. It is important that the units of time be equal for the analysis (UNESCO 1978).

615.1105 Cumulative distribution curves

The comparison of cumulative distribution curves may also be used to determine trends. Using the fecal coliform data, cumulative distribution curves were created for each year. By comparing the various curves, such as the 1984 curve to the 1989 curve, the decrease in fecal coliform bacteria is evident from 1984 to 1989 (fig. 11-5). These data could be tested using the Kolmogorov-Smirnov Goodness of fit (Zar 1996). The differences between these two curves is partly because of their individual means.

Figure 11-5 Cumulative frequency curves for the fecal coliform data



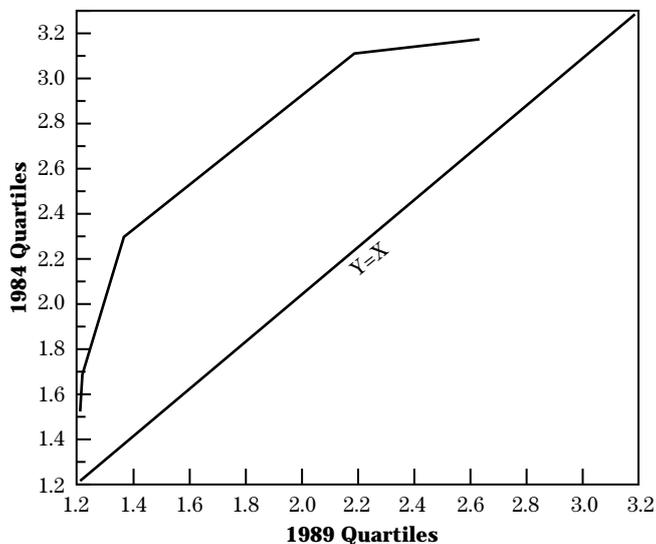
615.1106 Q-Q plots

For a Q-Q plot, the percentile (quartile) of one data set is plotted against another. For distributions that are similar, the points should follow along a line defined by $Y = X$ (UNESCO 1978). The 1984 fecal coliform data in table 11-5 are plotted against the 1989 data in figure 11-6. The 1984 quartiles are clearly higher than the 1989 quartiles, indicating a trend toward decreasing fecal coliform in the stream.

Table 11-5 Univariate statistics for fecal coliform data for 1984 and 1989

Quartile	Log ₁₀ fecal coliform (No./100mL)	
	1984	1989
0% (min)	1.53	1.23
10%	1.73	1.24
25%	2.32	1.39
50%	2.66	1.72
75%	3.12	2.19
100% (max)	3.18	2.63

Figure 11-6 Q-Q plot of log₁₀ fecal coliform data for 1984 and 1989



615.1107 Double mass analysis

Double-mass curves are plots of accumulated values for a water quality station of interest as a function of an average from a number of stations or a control or reference station. This type of trend analysis requires data from several different locations, preferably in close proximity to each other.

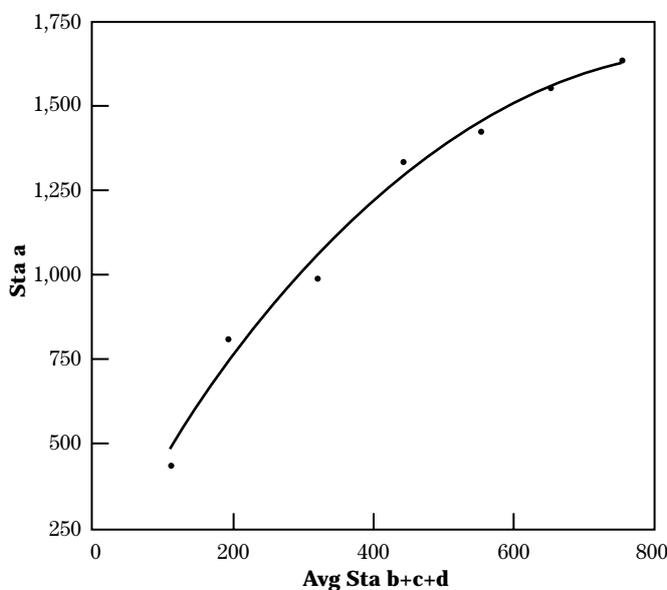
Double mass analysis is commonly used to assess changes in precipitation stations (Dunne and Leopold 1978). It is a visual tool that can be used to describe changes in one station in reference to a control station(s). A break in the slope of the line may indicate a trend or intervention. A comparison of slopes can be evaluated statistically (subpart 615.09) by analysis of covariance as pointed out by Dingman (1994).

A double mass curve of the fecal coliform data is shown in figure 11-7. In this case the cumulative coliform counts from a watershed receiving animal waste treatment (Sta a) are plotted as a function of the average among several stations (Sta b+c+d) that did not have watershed treatments. From this example the double mass analysis indicates that fecal coliform levels have fallen off gradually as compared to the average at the other three stations. A series of plots could be developed to check the other stations for trends (e.g., b vs. a+c+d, etc.)

615.1108 Paired regression analysis

Paired regressions can be used to infer trends if the data from two stations, one a control and one a treatment, are grouped into before and after time periods. Such data analysis was described in detail in subpart 615.09. A significant change in the paired regressions could signify a trend.

Figure 11-7 Double-mass analysis of fecal coliform data



615.1109 Nonparametric approaches

Several nonparametric approaches are used in trend detection. The primary advantages of nonparametric approaches are that there are no assumptions regarding the distribution, censored data, outliers, and missing data (Hirsch, et al. 1982). However, both parametric and nonparametric approaches assume that the data are not autocorrelated (i.e., that one observation is not related to the next observation).

(a) Kendall's tau

Kendall's tau is a measure of correlation between a water quality variable and time for monotonic trends (Helsel and Hirsch 1992). Like most nonparametric approaches the procedure is based on rank, rather than the actual values. Although the calculation of tau is on many statistical packages (subpart 615.12), in example 11-2 a hand calculation is performed.

When seasonality or flow effects are removed from the trend, Spearman's rho test may be superior to the Kendall test (Hipel and McLeod 1994)

Example 11-2 Kendall's tau for August fecal streptococcus data

The fecal streptococcus data from Jewett Brook used in the previous example was used for this example. To simplify the calculations, only the data for August is used (table 11-6).

The null hypothesis is that there is no correlation (trend) between bacteria level and time. The alternative hypothesis is that they are correlated.

Kendall's S is calculated from:

$$S = P - M \quad [11-1]$$

where:

P = number of pluses or the number of times the y's increase as the x's increase

M = number of minuses or the number of times the y's decrease as the x's increase (Helsel and Hirsch 1982)

Table 11-6 Fecal streptococcus data for August from Jewett Brook, St. Albans Bay Watershed, VT

Date	No./100 mL
8/82	200
8/83	4,000
8/84	430
8/85	390
8/86	370
8/87	237
8/88	790
8/89	60
8/90	140

Table 11-7 Summary of pluses and minuses for calculation of Kendall's tau for the August fecal streptococcus data for Jewett Brook, St. Albans Bay watershed, VT

200	4,000	430	390	370	237	790	60	140
+	-	-	-	-	+	-	+	
+	-	-	-	+	-	-		
+	-	-	+	-	-			
+	-	+	-	-				
+	-	-	-					
+	-	-						
-	-							
-								

To calculate the values, first compare 200/100 mL to all other values. For example, since 4,000 is greater than 200, a + is recorded, then 430 is greater than 200, a + is recorded, and so on. This can be summarized in a matrix format (table 11-7). Summing the pluses and minuses yields 11 P's and 25 M's.

$$S = 11 - 25 = -14.$$

Tau is calculated from:

$$\tau = \frac{S}{n \frac{(n-1)}{2}} \quad [11-2]$$

$$\tau = \frac{-14}{9 \frac{(9-1)}{2}} = -0.389$$

From appendix H, for S = (x)=-14 and n=9, p = 2 x 0.090 = 0.180. Because the calculated tau is greater than the table tau, the null hypothesis of no change is rejected because tau is significantly different from zero. The alternative hypothesis that there is a significant trend is accepted.

For a data set with seasonality (for example, months across years are different), the seasonal Kendall test may be used (Hirsch et al. 1982). For each season a separate S is calculated. They then are summed across seasons.

A seasonal slope estimator (B) can be calculated as the median of all the slopes between all possible data pairs within the same season (Helsel and Hirsch 1992). The individual slopes are calculated using equation 11-3 (Hirsch, et al. 1982):

$$d_{ijk} = \frac{(x_{ij} - x_{ik})}{j - k} \quad [11-3]$$

where:

I = 1, 2, ..., 12 months

j = k+1, 2, ..., n years

k = 1, 2, ..., n-1 years

The slope estimator is determined in subpart 615.12 in the WQStat II package.

615.1110 Summary

Table 11–8 summarizes the trend methods described in this subpart and whether they are suitable for missing data, censored data, or seasonality.

Table 11–8 Summary of trend detection techniques

Trend method	Missing data	Censored data	Seasonality	Comments
Time plot	ok	ok	ok	
Least squares regression	ok	no	no	
Annual means	ok	no	no	
Cumulative distribution	ok	no	no	
Q-Q plots	ok	no	no	
Double mass analysis	ok	no	no	
Paired regressions	ok	no	ok	
Nonparametric seasonal Kendall	ok	ok	ok	Distribution free

615.1111 References

- Dingman, S.L. 1994. Physical hydrology. Prentice Hall. Englewood Cliffs, NJ.
- Dunne, T., and L.B. Leopold. 1978. Water in environmental planning. W. H. Freeman and Co., San Francisco, CA.
- Erlebach, W.E. 1978. A systematic approach to monitoring trends in the quality of surface waters. *In* Establishment of water quality monitoring programs, American Water Resources Association, pp. 7-19.
- Helsel, D.R., and R.M. Hirsch. 1992. Chapter 12. Trend analysis. *In* Statistical methods in water resources, Studies in Environmental Science 49, Elsevier, New York, NY.
- Hipel, K.W., and A.I. McLeod. 1994. Time series modeling of water resources and environmental systems. Elsevier Science BV. Amsterdam, The Neatherlands.
- Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18(1):107-121.
- McLeod, A.I., K.W. Hipel, and F. Comancho. 1983. Trend assessment of water quality time series. *Water Resources Bulletin* 19(4):537-547.
- SAS Institute, Inc. 1995. Statistics and graphical guide. Ver. 3.1, SAS Institute, Inc., Cary, NC.
- UNESCO/WHO. 1978. Water quality surveys - a guide for the collection and interpretation of water quality data. IHD-WHO Working Group on Quality of Water, UNESCO/WHO, Geneva, Switzerland.
- Zar, J.H. 1996. Biostatistical analysis. Prentice Hall. Upper Saddle River, NJ.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Subpart 615.12 Statistical Packages

Subpart 615.12 Statistical Packages

Contents:	615.1200	Introduction	12-1
	615.1201	Sample size and sampling frequency estimator	12-1
	615.1202	Water quality statistical software	12-2
		(a) WQStat II	12-2
		(b) DETECT	12-6
		(c) SDS	12-9
		(d) ESTREND	12-9
	615.1203	General statistics	12-10
	615.1204	References	12-12

Tables	Table 12-1	WQStat II Mean / Skew values for the fecal streptococcus data	12-3
	Table 12-2	WQStat II seasonal box and whiskers for the fecal streptococcus data	12-4
	Table 12-3	WQStat II annual box and whiskers for the fecal streptococcus data	12-4
	Table 12-4	WQStat II autocorrelations for the fecal streptococcus data	12-5
	Table 12-5	WQStat II Kruskal-Wallis test for seasonality for the fecal streptococcus data (test statistic = 11.62)	12-5
	Table 12-6	WQStat II seasonal Kendal test for the fecal streptococcus data (test statistic = -3.987)	12-5
	Table 12-7	ANOVA table for equality of means for the fecal streptococcus data	12-8

	Table 12-8	Autocorrelation coefficients for the fecal streptococcus data	12-8
	Table 12-9	DETECT Kendall's test for trend for the fecal streptococcus data	12-8
	Table 12-10	Summary of characteristics and capabilities of water quality statistical packages	12-9
	Table 12-11	Summary of cost (1997) and capabilities of statistical software packages	12-10
	Table 12-12	Summary of statistical methods included in software packages	12-11
<hr/>			
Figures	Figure 12-1	Time series of fecal coliform data from DETECT	12-7
	Figure 12-2	Double-mass curve for the fecal streptococcus data	12-8
	Figure 12-3	CUSUM function for the fecal streptococcus data	12-8
<hr/>			
Examples	Example 12-1	WQStat II using the fecal streptococcus data from the St. Albans Bay RCWP	12-3
	Example 12-2	DETECT using the fecal streptococcus data from the St. Albans Bay RCWP	12-7

615.1200 Introduction

Several statistical software packages were developed specifically to aid in water quality data analysis. These packages include WQStat II (Loftis 1989), DETECT (Cluis 1989), SDS (Gaugush 1993), and ESTREND (Shertz, et al. 1991). In addition, numerous statistical software packages are available to assist in data analysis of most any type data. This subpart describes the packages available for water quality data analysis so that their usefulness for your particular situation can be determined. Statistical packages generally available for personal computers are described as well.

615.1201 Sample size and sampling frequency estimator

A sample size estimator has been developed by Region 6 of the United States Environmental Protection Agency (USEPA). This Windows program is downloadable free of charge from:

www.epa.gov/earth1r6/6wq/

This program estimates sample sizes for linear and step trends, estimation of means and differences between means. One major advantage of the software is that it performs iterative procedures.

615.1202 Water quality statistical software

(a) WQStat Plus

WQStat II was developed at Colorado State University (Loftis 1989, Ward, et al. 1990). The most recent version is WQStat Plus. The package is IBM-PC compatible and includes both data management and data analysis capabilities. Although data for any frequency of time series can be used in this program, WQStat creates either a monthly or quarterly data file for analysis. Data can be either manually entered, or the program can read various files.

The following summary statistics are provided by WQStat Plus as part of an exploratory data analysis (EDA):

- mean
- median
- standard deviation
- number of data points
- skewness and significance
- kurtosis and significance
- frequency histogram
- correlogram (autocorrelation)

A time series plot also can be obtained as well as indicators of seasonality:

- seasonal box-and-whisker plot
- annual box-and-whisker plot
- Kruskal-Wallis test for seasonality

For trend detection the program determines:

- Kendall tau
- seasonal Kendall test
- seasonal Kendall slope estimator
- analysis of covariance

An analysis is also provided across groups using medians. This analysis allows comparison of sites or time periods within a single site. The following non-parametric approaches are used:

- Wilcoxon Signed Rank test
- Mann-Whitney test
- Kruskal-Wallis test

The package also provides an analysis of extreme values, such as the proportion of values exceeding a standard.

Example 26–1 gives an application of WQStat Plus using the fecal streptococcus time series data for Jewett Brook in the St. Albans Bay watershed in Vermont, used in subpart 615.11.

WQStat Plus is available from Intelligent Decision Technologies, 203 South Main Street, Longmont, Colorado 80501, www.idt-ltd.com.

Example 12-1 WQStat using the fecal streptococcus data from the St. Albans Bay RCWP

Monthly mean fecal coliform values for Jewett Brook for the period December 1981 through August 1990 were entered into WQStat using a Lotus 1-2-3 file. A plot of the time series is shown in figure 11-1 in subpart 615.11.

Following the WQStat main menu, the summary statistics shown in table 12-1 were obtained.

The package produces a time plot and a seasonal Box-and-Whisker Plot. The Box-and-Whisker data indicate that seasons (months) are not greatly different (table 12-2).

An annual Box-and-Whisker Plot is provided. For the fecal streptococcus data, the annual Box-and-Whisker Plot indicates that the median and quartiles appear to decrease with time (table 12-3).

The program produces histograms and correlogram plots. The autocorrelations output is presented in table 12-4.

The autocorrelations shown in table 12-4 indicate that no significant serial correlation exists within the data. The highest autocorrelation was for the lag 9-month period, but it was not significant.

The Kruskal-Wallis test for seasonality using medians indicated that seasonality was not significant in the fecal streptococcus data (table 12-5).

The Seasonal Kendall test for trend was used since there were more than 5 years of data (table 12-6).

For this example, WQStat indicated that there was a declining trend in fecal streptococcus in Jewett Brook of 30 organisms per 100 mL per year, which is significant.

Table 12-1 WQStat Mean / Skew values for the fecal streptococcus data

Mean	Skew values (No / 100 mL)
Mean	458.010
Median	158.000
Standard deviation	783.943
Number of data points	96

Skew test for normality
(skew value = 3.400)

Confidence level	Test	Significance
98%	3.400 > 0.579	significant
90%	3.400 > 0.397	significant
80%	3.400 > 0.306	significant

Kurtosis test for normality
(Kurtosis value = 15.11)

Confidence level	Test	Significance
98%	15.11 > 4.42	significant
90%	15.11 > 3.79	significant
80%	15.11 > 3.53	significant

Example 12-1 WQStat using the fecal streptococcus data from the St. Albans Bay RCWP—Continued**Table 12-2** WQStat seasonal box and whiskers for the fecal streptococcus data

Season	Minimum	Interquartile	Median	Interquartile	Maximum
1/1-2/1	1.4E+01	1.3E+02	1.5E+02	5.3E+02	9.8E+02
2/1-3/1	7.9E+01	1.6E+02	2.1E+02	3.6E+02	1.5E+02
3/1-4/1	3.4E+01	5.5E+01	1.5E+02	6.9E+02	2.7E+03
4/1-5/1	2.5E+01	4.8E+01	6.5E+01	3.4E+02	4.8E+02
5/1-6/1	1.6E+01	2.1E+01	1.3E+02	2.9E+02	7.1E+02
6/1-7/1	8.4E+01	9.9E+01	1.5E+02	2.1E+02	1.5E+03
7/1-8/1	1.7E+01	1.8E+02	2.8E+02	7.6E+02	3.7E+03
8/1-9/1	1.4E+02	2.4E+02	4.0E+02	7.9E+02	4.0E+03
9/1-10/1	2.0E+01	6.9E+01	2.2E+02	4.8E+02	7.6E+02
10/1-11/1	4.0E+00	1.0E+02	2.1E+02	3.5E+02	8.0E+02
11/1-12/1	5.1E+01	1.7E+02	2.4E+02	4.2E+02	4.3E+03
12/1-1/1	2.6E+01	7.6E+01	1.5E+02	1.5E+03	2.1E+03

Table 12-3 WQStat annual box and whiskers for the fecal streptococcus data

Season	Minimum	Interquartile	Median	Interquartile	Maximum
1981	9.6E+02	9.6E+02	9.6E+02	9.6E+02	9.6E+02
1982	9.8E+01	1.3E+02	2.2E+02	6.1E+02	4.3E+03
1983	7.6E+01	1.5E+02	2.9E+02	7.1E+02	4.0E+03
1984	3.4E+01	1.8E+02	4.6E+02	1.2E+03	1.5E+03
1985	1.7E+01	9.2E+01	2.0E+02	7.8E+02	2.1E+03
1986	2.1E+01	5.8E+01	1.6E+02	4.9E+02	3.7E+03
1987	1.4E+01	3.1E+01	1.9E+02	2.3E+02	1.0E+03
1988	4.0E+00	6.5E+01	1.5E+02	3.9E+02	7.9E+02
1989	1.7E+01	2.6E+01	5.3E+01	1.5E+02	4.3E+02
1990	2.5E+01	3.4E+01	1.3E+02	2.8E+02	3.6E+02

Example 12-1 WQStat using the fecal streptococcus data from the St. Albans Bay RCWP—Continued

Table 12-4 WQStat autocorrelations for the fecal streptococcus data

Rho 1	:	-0.0232
Rho 2	:	-0.0996
Rho 3	:	0.1379
Rho 4	:	0.0825
Rho 5	:	-0.0045
Rho 6	:	0.0403
Rho 7	:	0.0529
Rho 8	:	-0.0347
Rho 9	:	0.1988
Rho10	:	0.0816
Rho 11	:	0.0048
Rho 12	:	0.0235
Rho 13	:	-0.0651
Rho 14	:	-0.0469
Rho 15	:	0.0434
Rho 16	:	-0.0051
Rho 17	:	-0.0099
Rho 18	:	-0.0379
Rho 19	:	0.1106
Rho 20	:	0.0232
Rho 21	:	0.0004
Rho 22	:	0.0219
Rho 23	:	-0.0296
Rho 24	:	-0.0175

Boundary value = 0.2041

Table 12-5 WQStat Kruskal-Wallis test for seasonality for the fecal streptococcus data (test statistic = 11.62)

Confidence level	Test	Significance
95%	11.62<19.68	not significant
90%	11.62<17.28	not significant
75%	11.62<13.70	not significant

Table 12-6 WQStat seasonal Kendall test for the fecal streptococcus data (test statistic = -3.987)

Confidence level	Test	Significance
95%	-3.987<-1.960	significant
90%	-3.987<-1.645	significant
80%	-3.987<-1.282	significant

Seasonal Kendall slope estimate:
Slope = -30.00000 units/year

(b) DETECT

The program DETECT was developed in Quebec, Canada, to utilize nonparametric approaches to detect trends in water quality data (Cluis 1989). This package is IBM-PC compatible and is somewhat directed toward Canada's national water quality data collection program (NAQUADAT). A typical input file contains the date (YY MM DD), the concentration, and the discharge (optional). Mass loading information may be input as well. The input file must be in columns with a row in a strict FORTRAN format:

(12X, I2, 1X, I2, 1X, I2, 16X, F12.6, F12.6)

This format is designed to read as: 12 spaces, YY, one space, MM, one space, DD, 16 spaces, concentration in 12 spaces with 6 following decimal, and discharge in 12 spaces with 6 following decimal (optional). Concentration data should be in milligrams per liter and discharge in cubic meters per second.

Graphic analysis includes a time plot, double-mass curves, and the CUSUM function. Double-mass curves show the accumulated sum of the concentration or discharge as a function of accumulated time (days from first observation). The CUSUM function, or cumulative sum, is the summation of the deviations of the observations from the mean plotted as a function of time.

$$\text{CUSUM}(X_t) = \sum_{j=1}^t X_j - j(\bar{X}) \quad [12-1]$$

where:

t = time (Cluis 1989, Hipel and McLeod 1994)

DETECT allows elimination of high and low outliers. Among the tests in DETECT is one for seasonality based on ANOVA. Missing values may be replaced using three different options:

- Temporal interpolation
- Seasonal mean
- Concentration-discharge relationship

Persistence in the trend is examined using autocorrelation coefficients. The appropriate test for trend recommended in the user's manual is suggested based on:

- Type of trend—monotonic or stepwise
- Persistence—Markovian or none
- Seasonality

The following trend tests are available:

- Lettenmaier/Spearman (Lettenmaier 1976)
- Hirsch and Slack (Hirsch and Slack 1984)
- Spearman/Kendal (Helsel and Hirsch 1992)
- Kendall seasonality (Helsel and Hirsch 1992)
- Lettenmaier/Mann-Whitney (Lettenmaier 1976)
- Mann-Whitney (Lettenmaier 1976)

Example 12-2 shows an application of DETECT using the fecal streptococcus time series data for Jewett Brook in the St. Albans Bay watershed in Vermont, used in subpart 615.11.

Example 12-2 DETECT using the fecal streptococcus data from the St. Albans Bay RCWP

Monthly mean fecal coliform values for Jewett Brook for the period December 1981 through August 1990 were prepared for entry into DETECT by editing a file in a DOS editor to put it in the proper format. A plot of the time series generated by DETECT is shown in figure 12-1.

The outliers were not eliminated from the data set for this example. As indicated in the manual, non-parametric tests yield stable results even with outliers present.

The double mass curve generated by DETECT is shown in figure 12-2. This plot shows the accumulated sum of fecal coliform abundance as a function of the accumulated time in months. The plot contains several lines. A general mean line is drawn from the origin to the upper right hand corner of the graph. The individual points are shown as X's. The general mean slope can be compared to groups of points in the double-mass curve. Slope of a group of points less than the general mean line indicates that the mean of these points would be less than the general mean. The lines above and below the mean line are termed *rails* and are two standard deviations from the mean line based on deviations from only its side of the line. Rails located far from the

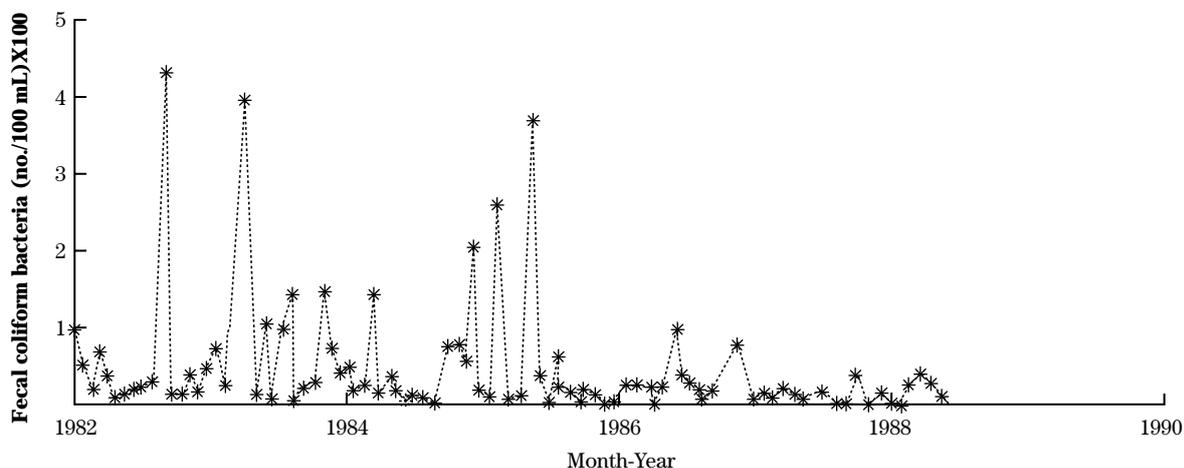
mean line indicate large variability in the data. If no trend is present, points on the double-mass curve are located on both sides of the mean line randomly. This is not the case in this example, indicating a trend is most likely present.

The CUSUM function is shown in figure 12-3. This plots the summation of the deviations from the general mean line in the previous figure.

Departures on one side of the line at $Y=0$ indicate a likely trend, as in this case. If the curve is parabolic, a monotonic linear trend is suggested. If the curve includes discontinuous lines, a stepwise trend is suggested. In this case a monotonic trend is suspected. The analysis of variance in table 12-7 tests whether monthly means are different as a test of seasonality.

The ANOVA in table 12-7 indicates no difference among months. Also, a Bartlett's test of the equality of variances is performed, which indicates in this case that the variances may not be equal across groups. Some data was missing in the fecal streptococcus data set, and the interpolation option was selected to fill missing data.

Figure 12-1 Time series of fecal coliform data from DETECT



Example 12-2 DETECT using the fecal streptococcus data from the St. Albans Bay RCWP—Continued

Autocorrelation correlation coefficients were used for an analysis of persistence (table 12-8). The autocorrelation coefficient is significant if the value is at least two times the standard deviation. In this case there was no significant persistence since no autocorrelations were significant. If the lag 1 r was significant and the lag 2 r was not, this would be termed Markovian persistence.

Using the decision tree in the program, the data displayed a monotonic trend without persistence or seasonality. Therefore, the Kendall test was used for analysis of the trend. Table 12-9 shows the results from Kendall's test as displayed by DETECT.

Table 12-7 ANOVA table for equality of means for the fecal streptococcus data

Source	df	MS	F
Month	11	0.61241E-06	0.995
Error	84	0.61568E-06	
Total	95	0.61530E-06	

Equality of means accepted
No seasonality
Equality of variances is rejected!

Table 12-8 Autocorrelation coefficients for the fecal streptococcus data

	Lag			
	1	2	3	4
coeff.	0.16	-0.02	0.10	0.12
std. dev	0.10	0.10	0.10	0.10

Table 12-9 DETECT Kendall's test for trend for the fecal streptococcus data

statistic	-1375.63
test value	-3.86
signif. level	0.00

Comment: Decreasing monotonic trend detected.

Figure 12-2 Double-mass curve for the fecal streptococcus data

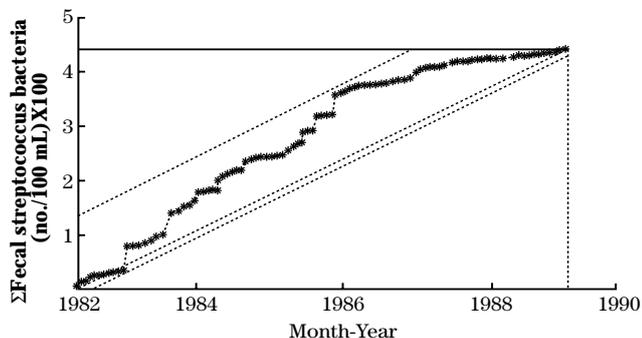
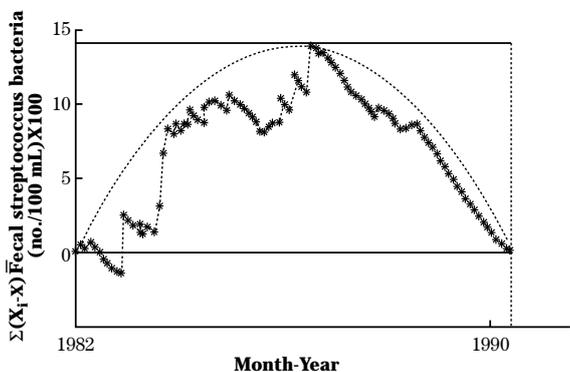


Figure 12-3 CUSUM function for the fecal streptococcus data



(c) SDS

The Sampling Design Software (SDS) was developed by the U.S. Army Corps of Engineers. This software is used to determine sample sizes, variance components, optimization of stratified samples among strata, and clustering of groups to increase efficiency of sampling (Gaugush 1993).

The sample size determination can be based on multi-variable sampling using either simple random or stratified sampling. The decision analysis is based on the mean, coefficient of variation, precision level, acceptable error, and the costs of sampling. The program displays the sample sizes and costs for each variable at different precision and error levels.

The variance component program determines the contribution to the variability in a water quality variable from different factors, such as station, date, and depth. The analysis attempts to determine which factors are most important in sampling and, therefore, which factors should dominate the design. For example, if most of the variance was explained by date, the station and depth subsampling could be reduced.

The number of samples applied to different strata can be optimized using error analysis in the program. The percent variance for each strata is compared to the percent of the number of samples and a percent optimum number of samples. Generally, more samples are allocated to strata with the higher variability.

Cluster analysis can be used to identify redundancy in the sampling program. For example, if a number of water quality stations are producing the same type of information, one or more could be dropped.

(d) ESTREND

ESTREND (Shertz, et al. 1991) is used by the U.S. Geological Survey for nonparametric trend analysis at its various water quality stations. The program is written for UNIX and has been commonly used on Prime™ computers.

Table 12-10 summarizes the characteristics and capabilities of the various water quality statistical packages.

Table 12-10 Summary of characteristics and capabilities of water quality statistical packages

	WQStat II	DETECT	SDS
Data manager			
Data type	monthly, seasonal	monthly	summary
ASCII import	X	X	X
Lotus 1-2-3 import	X		
Manual entry	X		
Missing data		X	
Data analyses			
EDA	X		
Trends	X	X	
Group comparisons	X		
Extreme values	X		
Autocorrelation		X	
Sample sizes			X
Graphics			
time plot	X	X	
double-mass		X	
CUSUM		X	

615.1203 General statistics

Many statistical packages are commercially available that will perform the statistical analyses described in NWQH Part 615. Table 12–11 provides a summary of some of the capabilities and features of these packages, and table 12–12 summarizes the statistical methods included in each package.

Table 12–11 Summary of cost (2001) and capabilities of general statistical software packages

Statistical package	Cost (\$)	Win/ Mac	Graphics	Documentation	Comments
Analyse-it	125	W	X	on-line	Plug-in for MS Excel, www.analyse-it.com (note British spelling)
DataDesk	£399 ^{1/}	W/M	X	manual	www.longman.net/datadesk-activstats
Instat	79	W/M	X		www.graphpad.com
JMP	895	W/M	X	manual	www.jmpdiscovery.com
Quick Statistica	495	W			www.statsoft.com
SAS		W	X	manual	Primarily for mainframe computers, www.sas.com
SPSS	858 ^{2/}	W/M			www.spss.com
Statistica	1095	W			www.statsoft.com
Statistix	495	W			www.statistix.com
SYSTAT	1299	W			www.spss.com
WINKS Basic	99	W	X	manual	www.texasoft.com/homepage

^{1/} Price in British Pounds.

^{2/} GSA Schedule.

Table 12-12 Summary of statistical methods included in software packages

Package	Descriptive/ univariate	Boxplot	Test of normality	Regression/ correlation	t-test	ANOVA	Multiple comparisons	ANCOVA	Nonparametric
Analyse-it	X	X	X	X	X	X	O	O	X
DataDesk	X	X	X	X	X	X	X	X	X
Instat	X	O	X	X	X	X	X	O	X
JMP	X	X	X	X	X	X	X	X	X
Quick Statistica	X	X	X	X	X	X	O	X	X
SAS	X	X	X	X	X	X	X	X	X
SPSS	X	X	X	X	X	X	X	X	X
Statistica	X	X	X	X	X	X	X	X	X
Statistix	X	X	X	X	X	X	O	X	O
SYSTAT	X	X	X	X	X	X	O	X	X
WINKS	X	X	X	X	X	X	X	X	X

O = blank

615.1204 References

- Cluis, D. 1989. Detection of trend and norm violation. User's manual—DETECT and EXCEED, Ver. 2, INRS-Eau, Quebec, Canada.
- Gaugush, R.F. 1993. Sample design software user's manual. Instructions report W-93-1, Vicksburg, MS, U.S. Army Engineer Waterways Experiment Station.
- Helsel, D.R., and R.M. Hirsch. 1992. Trend analysis, chapter 12. *In* Statistical Methods in Water Resources, Studies in Environmental Science 49, Elsevier, New York, NY.
- Hipel, K.W., and A.I. McLeod. 1994. Time series modeling of water resources and environmental systems. Elsevier Science. The Neatherlands.
- Hirsch, R.M., and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20(6):727-732.
- Lettenmaier, D.P. 1976. Detection of trend in water quality data from records with dependent observations. *Water Resources Research* 12(5):1037-1046.
- Loftis, J.C. 1989. User's manual—WQStat II. A water quality statistics program, CO State Univ.
- Shertz, T.L., R.B. Alexander, and D.J. Ohe. 1991. Computer program estimate trend (ESTREND). *Water Resources Investigations WRI 91-4040*.
- Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. Design of water quality monitoring systems. Van Nostrand Reinhold, New York.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615

National Water Quality Handbook



Appendixes

Appendixes

Contents:	Appendix A	Distribution of t	A-2
	Appendix B	Skewness	A-3
	Appendix C	Values of F	A-4
	Appendix D	Kruskal-Wallis H	A-14
	Appendix E	Tukey's q	A-17
	Appendix F	Mann-Whitney U distribution	A-19
	Appendix G	Wilcoxon T distribution	A-20
	Appendix H	Quantities for Kendall's Tau	A-21
	Appendix I	Conversion factors	A-22

Appendix A Distribution of *t* (two-tailed) ^{1/}

Degrees of Freedom	0.500	0.400	0.20	Probability of a Larger Value, Sign Ignored				0.010	0.005	0.001
			0.10	0.050	0.025					
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657			
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598	
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941	
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610	
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859	
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959	
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405	
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041	
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781	
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587	
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437	
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318	
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221	
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140	
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073	
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015	
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965	
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922	
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883	
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850	
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819	
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792	
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767	
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745	
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725	
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707	
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690	
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674	
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659	
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646	
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591	
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551	
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520	
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496	
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476	
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460	
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435	
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416	
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402	
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390	
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373	
∞	.6745	.8416	1.2816	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905	

1/ Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods, 7th ed. Iowa State Univ. Press, Ames. (No part of this appendix may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher.)

Appendix B Table for testing skewness (one-tailed) ^{1/}

Size of sample <i>n</i>	-- Percentage points --		Standard deviation
	5%	1%	
25	0.711	1.061	0.4354
30	.661	.982	.4052
35	.621	.921	.3804
40	.587	.869	.3596
45	.558	.825	.3418
50	.533	.787	.3264
60	.492	.723	.3009
70	.459	.673	.2806
80	.432	.631	.2638
90	.409	.596	.2498
100	.389	.567	.2377
125	.350	.508	.2139
150	.321	.464	.1961
175	.298	.430	.1820
200	.280	.403	.1706
250	.251	.360	.1531
300	.230	.329	.1400
350	.213	.305	.1298
400	.200	.285	.1216
450	.188	.269	.1147
500	0.179	0.255	0.1089

^{1/} Snedecor, G.W., and W.G. Cochran. 1980. Statistical methods, 7th ed. Iowa State Univ. Press, Ames. (No part of this appendix may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher.)

Appendix C Values of F_{α}

Denominator df	Probability of a larger F	----- Numerator df -----								
		1	2	3	4	5	6	7	8	9
1	0.050	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50
	0.010	4052.00	4999.50	5403.00	5625.00	5764.00	5859.00	5928.00	5982.00	6022.00
2	0.050	18.51	19.00	19.16	19.25	19.30	19.33	9.35	19.37	19.38
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	0.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	0.010	8.88	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03

See footnote at end of table.

Appendix C Values of F_{α} —Continued

----- Numerator df -----										
10	12	15	20	24	30	40	60	120	∞	P
241.90	243.90	245.90	248.00	249.10	250.10	251.10	252.20	253.30	254.30	0.050
6056.00	6106.00	6157.00	6209.00	6235.00	6261.00	6287.00	6313.00	6339.00	6366.00	0.010
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	0.050
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	0.010
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	0.050
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	0.010
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	0.050
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.63	13.56	13.46	0.010
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	0.050
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	0.010
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	0.050
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	0.010
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	0.050
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	0.010
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	0.050
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	0.010
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	0.050
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	0.010
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	0.050
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	0.010
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	0.050
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	0.010
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	0.050
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	0.010
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	0.050
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	0.010
2.54	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	0.050
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	0.010

Appendix C Values of F_{α} —Continued

Denom- inator df	Probability of a larger F	----- Numerator df -----								
		1	2	3	4	5	6	7	8	9
15	0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	0.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	0.050	4.41	3.35	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	0.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	0.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	0.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	0.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	0.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	0.010	7.95	5.72	4.62	4.31	3.99	3.76	3.59	3.45	3.35
23	0.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	0.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	0.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	0.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	0.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	0.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	0.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	0.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	0.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	0.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	0.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	0.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12

See footnote at end of table.

Appendix C Values of F_{α} —Continued

----- Numerator df -----										
10	12	15	20	24	30	40	60	120	∞	P
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	0.050
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	0.010
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	0.050
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	0.010
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	0.050
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	0.010
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	0.050
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	0.010
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	0.050
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	0.010
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	0.050
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	0.010
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	0.030
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	0.010
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	0.050
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	0.010
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	0.050
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	0.010
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	0.050
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	0.010
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	0.050
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	0.010
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	0.050
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	0.010
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	0.050
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	0.010
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	0.050
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	0.010

Appendix C Values of F_{α} —Continued

Denom- inator <i>df</i>	Probability of a larger <i>F</i>	----- Numerator <i>df</i> -----								
		1	2	3	4	5	6	7	8	9
29	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	.050	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
	.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

1/ Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill, Inc., New York, NY. (Reproduced with permission of the McCraw-Hill Companies.)

Appendix C Values of F_{α} —Continued

----- Numerator df -----										
10	12	15	20	24	30	40	60	120	∞	P
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	.050
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	.010
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	.050
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	.010
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	.050
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	.010
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	.050
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	.010
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	.050
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	.030
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	.050
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	.010

Appendix D Critical Values of the Kruskal-Wallis H Distribution ^{1/}

n_1	n_2	n_3	$\infty =$	0.10	0.05	0.02	0.01	0.005	0.002	0.001
2	2	2		4.571						
3	2	1		4.286						
3	2	2		4.500	4.714					
3	3	1		4.571	5.143					
3	3	2		4.556	5.361	6.250				
3	3	3		4.622	5.600	6.489	(7.200)	7.200		
4	2	1		4.500						
4	2	2		4.458	5.333	6.000				
4	3	1		4.056	5.208					
4	3	2		4.511	5.444	6.144	6.444	7.000		
4	3	3		4.709	5.791	6.564	6.745	7.318	8.018	
4	4	1		4.167	4.967	(6.667)	6.667			
4	4	2		4.555	5.455	6.600	7.036	7.282	7.855	
4	4	3		4.545	5.598	6.712	7.144	7.598	8.227	8.909
4	4	4		4.654	5.692	6.962	7.654	8.000	8.654	9.269
5	2	1		4.200	5.000					
5	2	2		4.373	5.160	6.000	6.533			
5	3	1		4.018	4.960	6.044				
5	3	2		4.651	5.251	6.124	6.909	7.182		
5	3	3		4.533	5.648	6.533	7.079	7.636	8.048	8.727
5	4	1		3.987	4.985	6.431	6.955	7.364		
5	4	2		4.541	5.273	6.505	7.205	7.573	8.114	8.591
5	4	3		4.549	5.656	6.676	7.445	7.927	8.481	8.795
5	4	4		4.619	5.657	6.953	7.760	8.189	8.868	9.168
5	5	1		4.109	5.127	6.145	7.309	8.182		
5	5	2		4.623	5.338	6.446	7.338	8.131	6.446	7.338
5	5	3		4.545	5.705	6.866	7.578	8.316	8.809	9.521
5	5	4		4.523	5.666	7.000	7.823	8.523	9.163	9.606
5	5	5		4.940	5.780	7.220	8.000	8.780	9.620	9.920
6	1	1		-----						
5	2	1		4.200	4.822					
6	2	2		4.545	5.345	6.182	6.982			
5	3	1		3.909	4.855	6.236				
5	3	2		4.682	5.348	6.227	6.970	7.515	8.182	
6	3	3		4.538	5.615	6.590	7.410	7.872	8.628	9.346

See footnote at end of table.

Appendix D Critical Values of the Kruskal-Wallis H Distribution $1/$ —Continued

n_1	n_2	n_3	$\infty =$	0.10	0.05	0.02	0.01	0.005	0.002	0.001
6	4	1		4.038	4.947	6.174	7.106	7.614		
6	4	2		4.494	5.340	6.571	7.340	7.846	8.494	8.827
5	4	3		4.604	5.610	6.725	7.500	8.033	8.918	9.170
5	4	4		4.595	5.681	6.900	7.795	8.381	9.167	9.861
6	5	1		4.128	4.990	6.138	7.182	8.077	8.515	
6	5	2		4.596	5.338	6.585	7.376	8.196	8.967	9.189
5	5	3		4.535	5.602	6.829	7.590	8.314	9.150	9.669
5	5	4		4.522	5.661	7.018	7.936	8.643	9.458	9.960
6	5	5		4.547	5.729	7.110	8.028	8.859	9.771	10.271
5	6	1		4.000	4.945	6.286	7.121	8.165	9.077	9.692
6	6	2		4.438	5.410	6.667	7.467	8.210	9.219	9.752
6	6	3		4.558	5.625	6.900	7.725	8.458	9.458	10.150
5	6	4		4.548	5.724	7.107	8.000	8.754	9.662	10.342
5	5	5		4.542	5.765	7.152	8.124	8.967	9.948	10.524
6	6	6		4.643	5.801	7.240	8.222	9.170	10.187	10.889
7	7	7		4.594	5.819	7.332	8.378	9.373	10.516	11.310
8	8	8		4.595	5.805	7.355	8.465	9.495	10.805	11.705
2	2	1	1	-----						
2	2	2	1	5.357	5.679					
2	2	2	2	5.667	6.167	(6.667)	6.667			
3	1	1	1	-----						
3	2	1	1	5.143						
3	2	2	1	5.556	5.833	6.500				
3	2	2	2	5.544	6.333	6.978	7.133	7.533		
3	3	1	1	5.333	6.333					
3	3	2	1	5.689	6.244	6.689	7.200	7.400		
3	3	2	2	5.745	6.527	7.182	7.636	7.873	8.018	8.455
3	3	3	1	5.655	6.600	7.109	7.400	8.055	8.345	
3	3	3	2	5.879	6.727	7.636	8.105	8.379	8.803	9.030
3	3	3	3	6.026	7.000	7.872	8.538	8.897	9.462	9.513
4	1	1	1	-----						
4	2	1	1	5.250	5.833					
4	2	2	1	5.533	6.133	6.667	7.000			
4	2	2	2	5.755	6.545	7.091	7.391	7.964	8.291	
4	3	1	1	5.067	6.178	6.711	7.067			

See footnote at end of table.

Appendix D Critical Values of the Kruskal-Wallis H Distribution ^{1/}—Continued

n_1	n_2	n_3		$\alpha =$	0.10	0.05	0.02	0.01	0.005	0.002	0.001
4	3	2	1		5.591	6.309	7.018	7.455	7.773	8.182	
4	3	2	2		5.750	6.621	7.530	7.871	8.273	8.689	8.909
4	3	3	1		5.589	6.545	7.485	7.758	8.212	8.697	9.182
4	3	3	2		5.872	6.795	7.763	8.333	8.718	9.167	8.455
4	3	3	3		6.016	6.984	7.995	8.659	9.253	9.709	10.016
4	4	1	1		5.182	5.945	7.091	7.909	7.909		
4	4	2	1		5.568	6.386	7.364	7.886	8.341	8.591	8.909
4	4	2	2		5.808	6.731	7.750	8.346	8.692	9.269	9.462
4	4	3	1		5.692	6.635	7.660	8.231	8.583	9.038	9.327
4	4	3	2		5.901	6.874	7.951	8.621	9.165	9.615	9.945
4	4	3	3		6.019	7.038	8.181	8.876	9.495	10.105	10.467
4	4	4	1		5.564	6.725	7.879	8.588	9.000	9.478	9.758
4	4	4	2		5.914	6.957	8.157	8.871	9.486	10.043	10.429
4	4	4	3		6.042	7.142	8.350	9.075	9.742	10.542	10.929
4	4	4	4		6.088	7.235	8.515	9.287	9.971	10.809	11.338
2	1	1	1	1	-----						
2	2	1	1	1	5.785						
2	2	2	1	1	6.250	6.750					
2	2	2	2	1	6.600	7.133	(7.533)	7.533			
2	2	2	2	2	6.982	7.418	8.073	8.291	(8.727)	8.727	
3	1	1	1	1	-----						
3	2	1	1	1	6.139	6.583					
3	2	2	1	1	6.511	6.800	7.400	7.600			
3	2	2	2	1	6.709	7.309	7.836	8.127	8.327	8.618	
3	2	2	2	2	6.955	7.682	8.303	8.682	8.985	9.273	9.364
3	3	1	1	1	6.311	7.111	7.467				
3	3	2	1	1	6.600	7.200	7.892	8.073	8.345		
3	3	2	2	1	6.788	7.591	8.258	8.576	8.924	9.167	9.303
3	3	2	2	2	7.026	7.910	8.667	9.115	9.474	9.769	10.026
3	3	3	1	1	6.788	7.576	8.242	8.424	8.848	(9.455)	9.455
3	3	3	2	1	6.910	7.769	8.590	9.051	9.410	9.769	9.974
3	3	3	2	2	7.121	8.044	9.011	9.505	9.890	10.330	10.637
3	3	3	3	1	7.077	8.000	8.879	9.451	9.846	10.286	10.549
3	3	3	3	2	7.210	8.200	9.267	9.876	10.333	10.838	11.171
3	3	3	3	3	7.333	8.333	9.467	10.200	10.733	10.267	11.667

^{1/} Zar, J.H. 1996. Biostatistical analysis. 3rd ed., Prentice Hall, Upper Saddle River, NJ 07458.

Appendix E Upper Percentage Points of the Studentized Range, $q_\alpha = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{S_x} \sqrt{p}$

Error df	α	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
	.01	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	.05	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.09	7.17
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
	.01	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	.05	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.20	6.27	6.34	6.40	6.47
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.99	6.06	6.14	6.20	6.26	6.33
	.01	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.62	5.71	5.80	5.88	5.95	6.03	6.09	6.15	6.21
	.01	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.05	6.11
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.72	5.79	5.85	5.92	5.97	6.03
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	.05	3.01	3.67	4.08	4.37	4.60	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.58	5.65	5.72	5.79	5.85	5.90	5.96
	.01	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.72	5.79	5.84	5.90
	.01	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	.05	2.98	3.63	4.02	4.30	4.52	4.71	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.54	5.59
	.01	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61

Appendix E Upper Percentage Points of the Studentized Range, $q_\alpha = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{S_x} / \text{--- Continued}$

Error df	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
30	.05	2.89	3.49	3.84	4.10	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.91	4.98	5.05	5.11	5.16	5.22	5.27	5.31	5.36
	.01	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.16	5.20	5.24
	.01	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	.05	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.48	4.56	4.64	4.72	4.78	4.84	4.90	4.95	5.00	5.05	5.09	5.13
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

1/ Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill, Inc., New York, NY. (Reproduced with permission of the McGraw-Hill Companies.)

Appendix F Wilcoxon two-sample rank test (Mann-Whitney test) ^{1/}

$n_2 =$ larger n	P	$n_1 =$ smaller n													
		2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	.05			10											
	.01			—											
5	.05		6	11	17										
	.01		—	—	15										
6	.05		7	12	18	26									
	.01		—	10	16	23									
7	.05		7	13	20	27	36								
	.01		—	10	17	24	32								
8	.05	3	8	14	21	29	38	49							
	.01	—	—	11	17	25	34	43							
9	.05	3	8	15	22	31	40	51	63						
	.01	—	6	11	18	26	35	45	56						
10	.05	3	9	15	23	32	42	53	65	78					
	.01	—	6	12	19	27	37	47	58	71					
11	.05	4	9	16	24	34	44	55	68	81	96				
	.01	—	6	12	20	28	38	49	61	74	87				
12	.05	4	10	17	26	35	46	58	71	85	99	115			
	.01	—	7	13	21	30	40	51	63	76	90	106			
13	.05	4	10	18	27	37	48	60	73	88	103	119	137		
	.01	—	7	14	22	31	41	53	65	79	93	109	125		
14	.05	4	11	19	28	38	50	63	76	91	106	123	141	160	
	.01	—	7	14	22	32	43	54	67	81	96	112	129	147	
15	.05	4	11	20	29	40	52	65	79	94	110	127	145	164	185
	.01	—	8	15	23	33	44	56	70	84	99	115	133	151	171
16	.05	4	12	21	31	42	54	67	82	97	114	131	150	169	
	.01	—	8	15	24	34	46	58	72	86	102	119	137	155	
17	.05	5	12	21	32	43	56	70	84	100	117	135	154		
	.01	—	8	16	25	36	47	60	74	89	105	122	140		
18	.05	5	13	22	33	45	58	72	87	103	121	139			
	.01	—	8	16	26	37	49	62	76	92	108	125			
19	.05	5	13	23	34	46	60	74	90	107	124				
	.01	3	9	17	27	38	50	64	78	94	111				
20	.05	5	14	24	35	48	62	77	93	110					
	.01	3	9	18	28	39	52	66	81	97					
21	.05	6	14	25	37	50	64	79	95						
	.01	3	9	18	29	40	53	68	83						
22	.05	6	15	26	38	51	66	82							
	.01	3	10	19	29	42	55	70							
23	.05	6	15	27	39	53	68								
	.01	3	10	19	30	43	57								
24	.05	6	16	28	40	55									
	.01	3	10	20	31	44									
25	.05	6	16	28	42										
	.01	3	11	20	32										
26	.05	7	17	29											
	.01	3	11	21											
27	.05	7	17												
	.01	4	11												
28	.05	7													
	.01	4													

^{1/} Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill, Inc., New York, NY. (Reproduced with permission of the McGraw-Hill Companies.)

Appendix G Wilcoxon's signed rank test (tabulated values of T are such that smaller values, regardless of sign, occur by chance with stated probability) ^{1/}

Pairs n	-- Probability --		
	.05	.02	.01
6	0	—	—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

^{1/} Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill, Inc., New York, NY. (Reproduced with permission of the McCraw-Hill Companies.)

Appendix H Quantiles (p-values) for Kendall's tau correlation coefficient ($p = \text{Prob}[S \geq x] = \text{Prob}[S \leq -x]$) ^{1/}

x	----- Number of data pairs = n -----				x	---- Number of data pairs = n ----		
	4	5	8	9		6	7	10
0	0.625	0.592	0.548	0.540	1	0.500	0.500	0.500
2	0.375	0.408	0.452	0.460	3	0.360	0.386	0.431
4	0.167	0.242	0.360	0.381	5	0.235	0.281	0.364
6	0.042	0.117	0.274	0.306	7	0.136	0.191	0.300
8		0.042	0.199	0.238	9	0.068	0.119	0.242
10		0.0083	0.138	0.179	11	0.028	0.068	0.190
12			0.089	0.130	13	0.0083	0.035	0.146
14			0.054	0.090	15	0.0014	0.015	0.108
16			0.031	0.060	17		0.0054	0.078
18			0.016	0.038	19		0.0014	0.054
20			0.0071	0.022	21		0.0002	0.036
22			0.0028	0.012	23			0.023
24			0.0009	0.0063	25			0.014
26			0.0002	0.0029	27			0.0083
28			<0.0001	0.0012	29			0.0046
30				0.0004	31			0.0023
32				0.0001	33			0.0011
					35			0.0005
					37			0.0002

1/ Helsel, D.R., and R.M. Hirsch. 1992. Chapter 12, Trend analysis. *In* Statistical methods in water resources, Studies in Environmental Science 49, Elsevier, New York, NY.

Appendix I Conversion Factors**Length**

From:	To:	Multiply by:
foot	inch	12
foot	meter	.3048
inch	centimeter	2.54
kilometer	mile	0.621
meter	yard	1.094
mile	kilometer	1.6093
yard	inch	36

Area

From:	To:	Multiply by:
acre	ft ²	43,560
acre	hectare	0.405
ft ²	m ²	0.0929
hectare	acre	2.471
hectare	m ²	10 ⁴
mile ²	kilometer ²	2.59

Volume

From:	To:	Multiply by:
ft ³	liter	28.317
ft ³	gallon	7.481
gallon	liter	3.785
m ³	ft ³	35.314
m ³	liter	1,000

Discharge

From:	To:	Multiply by:
ft ³ /s	gpm	448.83
ft ³ /s	m ³ /s	.0283
m ³ /s	liter/s	1,000
m ³ /s	gpm	15,850

Mass

From:	To:	Multiply by:
pound	kilogram	0.4536
ton	pound	2,000
tonnes	pound	2,205
pound/ac	kg/ha	1.1208
ft ³ - water	pound	62.4

Temperature

$$^{\circ}\text{F} = \frac{9}{5} (^{\circ}\text{C}) + 32$$

$$^{\circ}\text{C} = \frac{5}{9} (^{\circ}\text{F} - 32)$$

Concentration

From:	To:	Multiply by:
mg/L	ppm	1.0
ppm	ppb	1,000
mg/L	mg/kg	1.0
ug/L	mg/m ³	1.0
g/m ³	mg/L	1.0
lb/ac	kg/ha	1.120851
% solution	mg/L	1 x 10 ⁴

Metric

To convert SI prefixes

From:	To:	Multiply by:
Suffix	mega (M)	1 x 10 ⁶
Suffix	kilo (k)	1,000
Suffix	hecto (c)	100
Suffix	deca	10
Suffix	Suffix	1
Suffix	deci	.1
Suffix	centi	.01
Suffix	milli	.001
Suffix	micro	.000001

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615
National Water Quality Handbook



Glossary

Glossary

Aerobic	Containing oxygen. Used to describe organisms living, active, or occurring only in the presence of oxygen.
Alternate hypothesis	Any hypothesis alternative to the one under a test.
Anaerobic	Containing no oxygen. Used to describe organisms living, active, or occurring in the absence of oxygen.
Analysis of variance	An analysis of the total variation displayed by a set of observations, measured by the sums of squares of deviations from the mean. The variation is usually separated into components associated with sources of interest.
Aquifer	A geologic formation containing water, usually able to yield appreciable water.
Baseflow	A part of stream discharge not attributed to direct runoff from precipitation or snowmelt and usually contributed by subsurface flow.
Baseline	Initial or background water quality conditions. Also a surveyed line.
Bedload	Sediment, not in suspension, moving along the streambed by rolling or bouncing.
Benthos	The assemblage of organisms living on or at the bottom of a body of water.
Best Management Practice	A practice or combination of practices found to be the most effective, practicable (including economic and institutional considerations) means of preventing or reducing the amount of pollution generated by nonpoint sources to a level compatible with water quality goals.
Blurring	An exploratory data analysis technique of smoothing by replacing data points with short vertical lines of appropriate length beginning with the median of the residuals.
Calibration	The beginning period of time for a paired watershed design somewhat synonymous with a baseline period.
Catchment	The area providing runoff to a lake, stream, or well (drainage area, drainage basin, watershed).
Coefficient of determination	The square of the correlation coefficient. Decimal fraction of percent of variance explained.
Coefficient of variation	The standard deviation of a distribution divided by the mean.
Coliform bacteria	A group of bacteria predominantly found in the intestines of animals, but also occasionally found elsewhere.
Composite sample	A combination of individual samples taken at selected intervals or volumes to minimize variability.

Concentration	The amount of a substance dissolved or suspended in a unit volume of water.
Conductance	The measure of the ability of a solution to conduct electricity that is equal to the reciprocal of the resistance.
Confidence level	The measure of probability (α) of the truth of a statement.
Confidence limits	The values of an upper and lower t of a confidence interval. The interval has a probability (α) that the value will lie between the upper and lower limits.
Confined aquifer	An aquifer that is surrounded by formations of less permeable or impermeable material that is isolated from the atmosphere. (Artesian aquifer)
Conservation practice	An engineered structure or management activity that eliminates or reduces an adverse environmental effect of a pollutant and conserves soil, water, plant, or animal resources.
Contamination	An introduction of a substance into water in a sufficient concentration to make the water unfit for its intended use.
Continuous data	Data for which all values in some range are possible, such as height and weight.
Control	In a study, a standard for comparison against which other treatments are compared, but is either untreated or receives a standard treatment. Also, a stable cross section in a stream that controls flow upstream.
Critical area	An area within a watershed determined to be an important source of a pollutant.
Current meter	A device for measuring the velocity of flowing water.
Discharge	The rate or volume of water flowing at a specific cross section within a specified time.
Discharge rating curve	A curve showing the relationship between the stage at a cross section and the discharge at that cross section.
Discrete data	Data for which the possible values are fixed, such as counts.
Dispersion	The mixing of the concentration of a substance in the water with another body of water due to the flow of water.
Dissolved oxygen	The oxygen dissolved in water, expressed in milligrams per liter or percentage saturation.
Drainage basin	See Catchment.

Drainage density	The density of natural drainage channels in a given area, expressed as length per unit area.
Effluent stream	A stream that receives water from saturated ground water.
Epilimnion	The upper waters of a thermally stratified lake.
Equipotential line	A contour line that connects points of equal head for the water table or equipotential surface.
Error	The difference between an occurring value and its true or expected value.
Eye smoothing	Drawing a smooth curve through points of data on a graph.
Field	A small agricultural unit implying a management area.
Filter strip	A conservation practice that is a strip of vegetated land established downslope of a nonpoint source of pollution with the purpose of reducing the pollutant.
Flow line	A line indicating the direction of ground water flow toward the point of discharge. Flow lines are perpendicular to equipotential lines and together they form a flow net.
Flume	An open conduit for flow.
Frequency distribution	A listing of the way the frequencies of members of a population are distributed according to the values of the variable. The distribution is usually shown in a table.
Gage	A device for determining the water level.
Grab sample	A single sample taken at a certain time and place.
Ground water	Subsurface water in the saturated zone below the water table.
Hydrograph	A graph showing discharge as a function of time for a given location on a stream.
Hypolimnion	The bottom water of a thermally stratified lake.
Hypothesis	A hypothesis concerning the parameters or form of the probability distribution for a designated population.
Intermittent stream	A stream or portion that flows only in direct response to precipitation.
Interval scale	A measurement with a constant interval size, but no true zero, such as temperature (arbitrary zero) and time.
Kurtosis	The extent to which a unimodal frequency curve is peaked.

Least squares regression	Estimation of regression parameters by minimizing a quadratic form.
Linnocorral	A device used in lakes that isolates the water column from surrounding water.
Load	The quantity of material entering a receiving body of water.
Lysimeter	A device used to measure the water quantity or quality draining through the soil.
Macroinvertebrate	A large animal without a backbone that can be observed without the aid of magnification.
Macrophyton	A large plant that can be observed without the aid of magnification.
Mean	The arithmetic average of the values for a variate.
Median	That value of the variate which divides the total frequency into two halves.
Mesocosm	A medium-sized experimental unit with boundaries.
Metalimnion	The middle layer of a thermally stratified lake.
Mode	The value of the variate that has the greatest number of members of the population.
Model	A description of a system; often mathematical.
Nonparametric statistics	Better termed distribution-free statistics. Testing a hypothesis that does not depend on the form of the underlying distribution.
Nonpoint source	A diffuse location with no particular point of origin.
Null hypothesis	A hypothesis under test that determines the probability of the Type I error. Also a hypothesis under a test of no difference.
Objective	A statement describing what is to be accomplished that contains an infinitive verb and an object.
Observation	Data that are collected or analyzed.
Ordinal scale	Data that consist of an ordering or ranking of measurements, such as A is bigger than B.
Parametric statistics	A statistical test that assumes the distribution type is known.
Perennial stream	A stream that flows continuously all seasons of a year and during both wet and dry years.
Periphyton	Small or microscopic aquatic plants attached to submerged objects.

Phytoplankton	Small or microscopic aquatic plants.
Piezometer	An instrument for measuring pressure head in the soil.
Plankton	Small or microscopic aquatic organisms that are floating, or weakly motile and generally considered to be at the mercy of the currents.
Plot	A small experimental unit with boundaries.
Pollutant	An undesirable substance in water, soil, or air at sufficient concentrations to impair the intended use of the resource.
Pollution	A condition caused by the presence of harmful or objectionable substances in water.
Population	A collection of individuals.
Random sample	A sample collected from a population where every sample has an equal probability of being selected.
Rating	A relation between stage and discharge of a stream.
Ratio scale	Measurements having a constant interval size and a true zero point, such as lengths, weights, volumes, and rates.
Reconnaissance survey	A survey to obtain a general view of water quality; may imply samples collected at approximately the same time (synoptic survey).
Regression	A statistical method to investigate relationships between two components.
Replication	The execution of an experiment more than once.
Resource management system	A combination of conservation practices and management identified by the primary use of land or water.
Responsiveness	In establishing cause-and-effect, the evidence that the dependent variable is related to the independent variable.
Runoff	That portion of precipitation or irrigation found in surface channels and streams.
Runoff coefficient	The ratio of the depth of runoff from a watershed to the depth of precipitation.
Sample	A part of all the possible measurements in some larger group, such as the population.
Sampler	A device used to obtain an aliquot of water.
Significance	The probability of committing a Type I error (α). Biological significance refers to an underlying assumption about relationships.

Skewness	A measure of asymmetry in a frequency distribution.
Smoothing	The process of removing fluctuations in a series of data.
Specific conductance	The ability of water to conduct electricity across a specific length at a specified temperature.
Stage	The elevation of the water surface above some datum.
Stage-discharge relation	The relationship between stream stage and discharge at a gaging station.
Standard deviation	A measure of dispersion of a frequency distribution that is the square root of the variance.
Statistic	A summary value calculated from a sample of observations.
Statistical error	See Error.
Statistics	The science of collecting, analyzing, and interpreting data.
Steady-state	Conditions that are averaging constant over time.
Stilling well	A chamber with small inlets connected to a water body used for measuring the water level.
Streamflow	Water flowing in a stream channel. (Stream discharge)
Surface runoff	The portion of runoff that reaches a stream by traveling over the surface of the land. (Overland flow)
Suspended solids	Solids in suspension in water.
Synoptic survey	See reconnaissance survey.
Tensiometer	An instrument filled with water with a porous cup used for measuring the soil water potential.
Turbidity	A condition in water caused by suspended matter that causes the scattering and absorption of light.
Unconfined aquifer	An aquifer where the water table is exposed to the atmosphere. (Water table aquifer)
Vadose zone	Zone of soil between the surface and the water table that is not saturated.
Variance	The mean of the squares of the deviations from the mean.
Velocity meter	A meter used to measure stream velocity.
Water quality	The physical, chemical, and biological properties of water with respect to its suitability for an intended use.

Water quality management	The management of the physical, chemical, and biological characteristics of water.
Water quality monitoring	The collection of information on the characteristics of water.
Water quality standards	A rule established by an agency or units of government; often numerical.
Water table	The upper surface of the saturated zone in a soil that is at atmospheric pressure.
Water-level recorder	A device used for recording the water elevation over time.
Watershed	The area contributing water to a stream, lake, or well.
Weir	A device used in a stream with a damming crest and an opening of some known geometric shape, such as a V-notch.
Zooplankton	Small or microscopic aquatic animals.

United States
Department of
Agriculture

**Natural
Resources
Conservation
Service**

Part 615

National Water Quality Handbook



Index

Index

A

Above-and-below watershed 05-5,
08-2 to 08-4
Analysis of covariance 00-2, 00-3,
09-4, 09-8, 09-9, 11-6, 12-2
ANOVA (Analysis of variance)
00-2 to 00-4, 03-1, 03-4,
03-7, 04-1, 04-2, 05-5, 06-1
to 06-10, 09-2, 09-7, 09-8,
10-2 to 10-5, 11-3, 11-4, 12-6
to 12-8, 12-11
one-way ANOVA 03-4
Association 04-1, 04-2, 11-9
Assumption 00-1, 05-1
additivity 03-1, 03-4, 03-7
independence 00-1, 00-3, 02-7,
03-1, 03-4, 03-5, 03-7, 09-7,
11-2
normality 00-3, 03-1, 03-2,
03-3, 03-7, 06-8, 07-3, 08-2,
08-3, 09-7, 10-2, 11-2, 11-3,
12-3, 12-11
randomness 00-1, 03-1, 03-7
Autocorrelation 00-2, 00-3, 03-4,
03-7, 12-2, 12-3, 12-5, 12-8,
12-9

B

Box-and-whisker plot 02-2, 02-3,
02-5, 02-6, 10-5, 11-1, 12-2,
12-3
Boxplot 02-3, 03-2, 10-5, 10-6,
11-3, 11-4, 12-11

C

Calibration 04-4, 09-1, 09-2,
09-3, 09-4, 09-5, 09-6, 09-7,
09-8, 09-9, 10-1
Causality 04-1, 04-4, 04-5, 04-6,
11-1
Coefficient of determination 00-2,
00-3, 09-3
Coefficient of skewness 00-2,
00-3, 01-9
Coefficient of variation 00-2,
00-3, 01-8, 01-9, 01-10, 12-9

Comparison of mean 07-1, 08-2,
08-4, 11-5
Computer 02-6, 02-8
Computer package 00-1
Confidence intervals 00-2, 00-3,
02-3, 09-3
Consistency 04-1, 04-2
Cumulative distribution 01-7,
11-5

D

Data analysis 02-1, 02-8
DETECT 12-1, 12-6 to 12-9,
12-12
Double mass analysis 11-6

E

EDA 00-1, 02-1, 11-1, 12-2, 12-
9
Error 07-3
standard error 00-2, 00-3,
01-8, 01-10, 06-9, 06-10, 07-3
type I 03-3, 03-4, 05-1, 05-2,
05-6
type II 05-1, 05-2, 05-6
Experimental design 04-1, 04-5,
04-6, 07-4, 08-3, 11-1
Exploratory data analysis 00-1,
00-4, 02-1, 02-8, 12-2

F

F ratio 00-2, 00-3, 03-3, 03-
7, 05-7, 06-5
Frequency(ies) 01-1, 01-3 to
01-5, 01-7, 02-2, 02-4, 03-2,
11-3, 11-5, 12-1, 12-2

G

Goodness of fit 03-2, 03-3, 11-5

H

Homogeneity of variance 00-
1, 00-3, 03-1, 03-3

Hypothesis 03-3, 03-4, 05-1,
05-2, 05-3, 05-4, 05-5, 05-6,
05-7, 06-4, 06-5, 06-8, 06-9,
07-1, 07-2, 08-2, 08-4, 10-3,
10-5, 11-2, 11-8
one-sample 05-3
two-sample 05-3
paired-sample 05-4, 08-4

I

Independence 00-1, 00-3, 02-7,
03-1, 03-4, 03-5, 03-7, 09-7,
11-2

K

Kendall's tau 00-2, 00-3, 11-
7, 11-8, 12-2, A-19
Kruskal-Wallis 00-2, 00-3, 06-8,
06-9, 06-10, 10-5, 12-2, 12-3,
12-5, A-11, A-12, A-13
Kurtosis 00-2, 00-3, 01-9, 01-10,
03-2, 03-3, 07-3, 08-3, 10-4,
12-2, 12-3

M

Mann-Whitney U 00-2, 00-3,
07-5, 10-5, 12-2, 12-6
Mean 00-2, 00-3, 01-1, 01-2,
01-6 to 01-9, 01-11, 02-2 to
02-7, 03-1 to 03-4, 03-6, 04-1,
04-3, 04-5, 05-1 to 05-7, 06-1,
06-2, 06-4 to 06-7, 06-9,
06-10, 07-1 to 07-3, 07-6,
08-2 to 08-5, 09-3, 09-5, 10-3,
10-4, 10-5, 11-1, 11-3, 11-4,
12-1 to 12-3, 12-6 to 12-9
Mean comparison 06-5, 06-9
Median 00-2, 00-3, 01-6, 01-7,
02-2, 02-3, 02-5 to 02-7,
03-2, 03-3, 07-3, 08-3, 10-5,
11-1, 11-3, 11-8, 12-2 to 12-4
Missing value 03-5, 03-6, 12-6
Mode 00-2, 00-3, 01-6, 01-7
Multiple comparison 00-2, 00-3,
05-5, 06-9, 06-10
Multiple watershed 10-1 to 10-6

N

Nonparametric 00-2, 00-3, 03-3, 03-4, 05-6, 06-1, 06-3, 06-8 to 06-10, 07-5, 07-6, 08-1, 08-4, 10-1 to 10-3, 10-5, 11-1, 11-7, 11-9, 12-2, 12-6, 12-7, 12-9, 12-11, 12-12

Normal distribution 01-7 to 01-10, 02-5, 03-1 to 03-3, 06-3, 07-3, 07-5, 08-4, 10-2, 10-3

Normality 00-3, 03-1 to 03-3, 03-7, 06-8, 07-3, 08-2, 08-3, 09-7, 10-2, 11-2, 11-3, 12-3, 12-11

O

Observation 00-3, 01-1 to 01-3, 01-6 to 01-8, 01-11, 02-2, 02-8, 03-1, 03-4, 03-6, 05-4, 06-1 to 06-4, 06-8 to 06-10, 07-1, 08-2, 09-1 to 09-3, 09-5, 09-7, 09-8, 11-1, 11-7, 12-6, 12-12

One-sample t-test 00-3

Outliers 01-3, 02-3, 03-1, 03-6, 12-6, 12-7

P

Paired comparison 05-4, 07-1, 08-2, 08-4, 10-3

Paired watershed 04-4, 08-1, 09-1, 09-2, 09-5 to 09-9, 10-1, 10-3

Plot 02-1 to 02-7

Plot design 06-1, 06-3, 06-8

Population 01-1, 01-2, 01-8, 03-2 to 03-4, 04-1, 04-2, 04-5, 05-1 to 05-3, 05-5, 05-6, 07-2

Q

Q-Q plot 11-6, 11-9

R

Random sample 01-2

Randomness 00-1, 03-1, 03-7

Regression 00-2, 00-3, 02-7, 03-5, 03-7, 04-1, 04-2, 04-6, 09-2 to 09-5, 09-7 to 09-9, 10-1, 11-1 to 11-4, 11-7, 11-9, 12-11

Replications 03-7, 04-1, 06-1 to 06-3, 06-6

Residual plot 03-1, 03-5, 03-7, 09-7

Residuals 03-4, 03-5, 09-2, 09-3, 09-7

Responsiveness 04-1, 04-4

Runoff 02-7

S

Sample 00-2, 00-3, 01-1 to 01-3, 01-8, 01-9, 01-11, 02-6, 03-1, 03-3 to 03-7, 05-1 to 05-7, 06-1, 06-2, 07-1, 07-2, 07-5, 08-1, 08-2, 08-4, 09-3, 09-7, 10-1 to 10-3, 10-5, 12-1, 12-9, 12-12

Sample size 01-3, 01-9, 02-6, 03-3 to 03-6, 05-1, 05-6, 12-1, 12-9

Shapiro-Wilk W-statistic 00-2, 00-3, 03-3, 09-7, 10-3, 11-3

Single watershed 07-1 to 07-6, 08-2

Skewness 00-2, 00-3, 01-9, 01-10, 02-5, 03-2, 03-3, 07-3, 08-3, 10-4, 12-2, A-3

Smoothing 02-1, 02-7, 02-8, 11-1

Standard deviation 00-2, 00-3, 01-7 to 01-10, 02-4, 02-5, 05-2, 06-2, 07-1, 07-3, 08-2, 12-2, 12-3, 12-7, 12-8

Standard error 00-2, 00-3, 01-8, 01-10, 06-9, 06-10, 07-3

Statistical significance 05-1, 05-6

Statistics, purpose of 00-1, 00-2, 00-3, 01-1

Stem-and-leaf diagram 02-1, 02-2, 02-5, 11-1

T

T-test 00-2, 00-3, 03-3, 03-4, 05-3 to 05-5, 07-1 to 07-3, 07-5, 08-1 to 08-4, 11-3, 11-4, 12-11

Time plot 02-6, 02-7, 07-6, 08-5, 11-2, 11-9, 12-3, 12-6, 12-9

Transformation 02-1, 02-4 to 02-7, 03-2 to 03-4, 06-3, 07-3, 07-6, 08-1 to 08-4, 10-3, 11-1 to 11-3

Treatment period 04-4, 09-1 to 09-4, 09-7, 09-8

Trend analysis 03-4, 03-5, 05-4, 11-1, 11-2, 11-4, 11-6, 11-9, 12-9, 12-12, A-19

Tukey's multiple comparison 00-2, 00-3, 06-10

Two-sample hypothesis 05-3

Two-sample t-test 00-3

Type I error 03-3, 03-4, 05-1, 05-2, 05-6

Type II error 05-1, 05-2, 05-6

V

Variance 00-1 to 00-3, 01-8, 01-10, 01-11, 02-4, 02-5, 03-1, 03-3 to 03-5, 03-7, 04-1, 04-2, 05-5, 06-1, 06-3, 06-8, 06-9, 07-1, 08-1 to 08-4, 09-2 to 09-4, 10-2, 10-3, 10-5, 11-2 to 11-4, 11-6, 12-2, 12-7 to 12-9

W

Watersheds
 above-and-below 05-5, 08-2 to 08-4
 multiple 10-1 to 10-6
 paired 04-4, 08-1, 09-1, 09-2, 09-5 to 09-9, 10-1, 10-3
 single 07-1 to 07-6, 08-2

Wilcoxon 08-4,
 Wilcoxon paired sample 00-2, 00-3, 08-4, 10-5, 12-2, A-17, A-18

WQStat II 11-8, 12-1, 12-2, 12-9, 12-12