

The Statistical Analysis of Experimental Data

Kenneth D. Hughes, Ph.D., Associate Professor
Kennesaw State University

A crucial aspect of owning or caring for a pool or spa involves collecting chemical water quality data (making measurements) and adjusting the concentration of active chemical species in the water. There are many important reasons for this activity including: 1) maintaining the health and safety of bathers, 2) reducing the degradation of pool structural components, 3) creating an aesthetically pleasing body of water, and 4) reducing the magnitude and cost of chemical treatment programs.

Although the phrase "Statistical analysis of data" has never incited much enthusiasm for most individuals and probably never will, it is an important aspect of any chemical measurement program that attempts to understand, analyze, and control water quality in pools and spas. It is unlikely that a chemical treatment program will be efficient in obtaining the goals listed earlier if the errors associated with chemical measurements are not understood.

Understanding the basic concepts of statistical analysis and the science behind making chemical measurements of water quality requires a concerted effort over time and includes as much "book-work" or studying as it does practical experience gained in the field. The technical skills required for successfully obtaining accurate water quality data and generating correct decisions from this data include becoming facile in 1) using chemical measurement tools (testing methods), 2) making observations of pool or spa bulk water characteristics, 3) understanding the impact of physical location on the pool or spa, 4) recording and tracking water quality chemical data over long periods of time, and 5) understanding the impact of decisions based upon this recorded data. The impact of decision making requires frequent testing of water parameters after chemical additions.

Many texts are required for the complete treatment of these pertinent topics, and reading in the following subject areas should be conducted: water chemistry theory, water chemistry monitoring/testing methods, statistical calculations, and the mathematical analysis of data. The primary goal of this article is to provide an overview and "jumping-in-point" for understanding factors that impact chemical measurements and basic statistical analysis of pool and spa water quality data. The topics covered in this article will aid in beginning a water quality monitoring program.

Understanding Water Chemistry in Pools and Spas

The water contained in pools and spas is a complex chemical system. Complete understanding of the water quality and the interactions of chemical species in this water system imply that the water quality characteristics can be predicted and controlled over extended periods of time. In everyday practice this capability is rarely achieved since many chemical tests are required and each water system is situated in a unique environment and experiences unique chemi-

- | |
|---|
| <ol style="list-style-type: none">1. Pool or spa fill water composition2. Chemicals intentionally introduced into the water for treatment purposes3. Chemicals unintentionally introduced into the water by bathers or external factors4. Chemical species that are generated by chemical and biological reactions occurring between chemical species listed in items 1, 2, and 3. |
|---|

Table 1 – Factors Controlling Water Quality

cal influences. These chemical influences are broadly categorized in Table 1.

Understanding pool and spa water quality is further complicated by physical changes that include lighting levels, temperature changes, evaporation, and dilution as a result of precipitation. Although understanding all the variables associated with water quality is a seemingly impossible task, significant progress can be made and a healthy, aesthetically pleasing water systems can be achieved. The key is to understand the basic chemical factors that control water quality in the pools or spas, and the desired water quality criteria. The water quality criteria for pools and spas has been outlined by the National Spa and Pool Institute (NSPI) and the National Sanitation Foundation (NSF).

The Need for Measurements

There are no inexpensive rapid methods that can provide a complete "picture" of water quality in a time period of several minutes. Complete understanding of complex water systems such as those present in pools and spas can only be obtained through frequent careful observation and the use of water quality chemical tests.

Observations of bulk water characteristics can be very useful and should be a component of any chemical measurement program. As an example, the lack of water clarity, the presence of color, and the presence of debris is indicative of low quality water. Unfortunately, the presence of uncolored water with good clarity and the lack of debris is not always associated with good water quality. Even the most seasoned professional can not visually determine the presence or absence of chemical species that generate the observed bulk water properties, control bacteria and organics, and minimize structural component degradation. As a result, accurate chemical measurements must be completed on a routine basis.

Chemical measurements are categorized as either qualitative or quantitative. Qualitative measurements provide information related to the presence or absence of a chemical species or a chemical concentration that is relative (greater or lessor than) to a solution of known concentration. A quantitative chemical measurement, by definition, provides a numerical value that corresponds to the concentration of the chemical species investigated.

In the majority of cases those individuals charged with the responsibility of maintaining water quality must perform quantitative chemical measurements (or analysis) using analytical tools such as test kits, test strips, and electronic probes and instrumentation. The numerical values are then compared to the chemical threshold values established by the NSPI and the NSF.

Errors Involved in Chemical Measurements

It is unfortunate that all quantitative measurements targeting chemical and physical properties of matter contain some level of error or uncertainty, but it is a fact. The uncertainty in measured numerical values complicates the analysis of the collected data and ultimately the decisions that are generated in the data analysis process. Therefore, it is important in any chemical measurement program that the source and magnitude of potential error in the water quality measurement be understood, controlled, and minimized.

In the end, the total error in any chemical measurement is comprised of two separate factors, the error associated with obtaining the sample for measurement and the error associated with performing the measurement. This relationship is described mathematically in the following equation:

$$(\text{Total Error})^2 = (\text{Sampling Error})^2 + (\text{Measurement Error})^2$$

It should be noted that the errors associated with both the sampling step and the measurement step in this equation are "squared" (mathematical function). The impact of this relationship is demonstrated in Table 2. It is clear by comparing these two calculations that the step with the largest error controls the total error in the final result. Specifically, the decrease in measurement error by a factor of five produces only a small decrease in the total error, from 27% to 25%. In order to reduce the total error in this case, the error associated with the sampling step must be addressed.

Controlling Sampling Errors

Practical reality requires that only a small portion of pool or spa water be taken (sampled) in order to perform a chemical analysis. This sample is then used to calculate the mass of chemicals that must be added to the water system. It is interesting to think about this process and the volume relationship between the sample taken and the total volume of the pool or spa. Clearly, the sample collected is a very small fraction of the total water volume and any er-

	Case A:	Case B:
Sampling error	25 %	25 %
Measurement error	10 %	2 %
Total Error	27 %	25 %

Table 2 – Examples of Total Error Calculations

ror associated with the sample collection and measurement is magnified when extrapolated to the larger water system.

The primary means of controlling sampling errors involves obtaining a representative sample from the larger water system. If the sample collected for chemical measurement is not representative of the larger body of water (pool or spa) then the chemical measurement will have little value. A representative sample is defined as a sample that is indicative or accurately reflects the composition of the larger chemical system. Unfortunately, obtaining a representative sample is not as easy as it initially seems. Three major requirements for obtaining a representative sample include, 1) understanding the variations in physical locations in the water system 2) understanding the effect of time on sampling (early morning versus late afternoon), and 3) clearly understanding the size of the sample that must be obtained.

In a small pool or spa a representative sample can be obtained after the pump and filtration system have been running for a time period that provides complete mixing of the total water volume. Unfortunately, large pools require many hours for even one complete turnover of the total water volume. It is in this case, that the water collected for chemical analysis may not be representative of the total pool volume. To address this problem several samples may be collected from different locations in the pool and then combined (mixed) to generate a more representative sample.

Additional sampling errors that are possible include the use of storage containers for the sample, transport characteristics (time and temperature) of samples from a site to a testing laboratory, and the time span between sampling the water system and performing the chemical analysis. An example of a poor sampling strategy is to collect a pool water sample for free chlorine analysis in an unclean clear glass jar, and then allowing this sample to sit at an elevated temperature in full sunlight, for several hours before making the measurement.

Controlling Measurement Errors:

There are three fundamental types of measurement error. These are classified as 1) gross errors, 2) determinant errors, and 3) random errors. The first two types of errors may be eliminated while the third, random error, may be controlled and minimized but not eliminated. These errors are addressed through the use of statistical analysis.

Gross Errors – Gross errors involve, as the name implies, large errors in the numerical value for a measurement. These errors are obvious when a series of individual data points of similar measurements,

such as pH, are compared. As example, a pH reading of 1.0 included in a series of pH readings ranging from 6.5 through 7.5 indicates something distinctly different or unique about this sample measurement. In many cases the cause of this error can be determined. Possibilities include using the wrong reagent, using an incorrect sample volume, or performing a calculation incorrectly. When using an instrument to perform a chemical measurement, a gross error may occur when an electrical component fails or fluctuations in electrical power occur. The procedure for handling this situation is simply the removal of the measurement result in question from further data analysis. These errors are easily identified if a good notebook is kept, and even more noticeable when the results collected are graphed with a simple spreadsheet program.

Determinant Errors – Determinant errors are characterized by measurement error that is directional with respect to the true or accepted value. In many cases the true value of the measurement is not known, but where it is known a determinant error will manifest itself in a numerical value that is always greater than or less than the true value. The source of these errors can be chemical, physical, or mechanical when an instrument is involved. An example of this type of measurement error would be a balance that always yielded a weight for an object that was greater than the object's true weight. These errors may be identified by using a proper calibration solution or analyzing a solution where the numerical value of the measurement is known. Elimination of determinant errors is completed by recalculating the concentration of reagent solutions and adjusting instrumental parameters. In the case of the balance, the balance is calibrated by using an object of known mass.

Random Errors – Random errors in the numerical value of a measurement occur as a result of a multitude of unavoidable small errors. These small errors will tend to cancel each other out due to the fact that some errors lead to increases in the numerical value while other errors lead to decreases in the numerical value. The outcome of this phenomena is measurement values that fall symmetrically around the true or accepted numerical value for the chemical parameter. Examples of phenomena that can generate small errors include temperature and pressure fluctuations, which affect chemical reactions. The means for handling and reporting these small random errors is known as statistical analysis.

Several concepts should be noted. The term accuracy is defined as the difference between the measurement value and the true or accepted value. The term precision is defined as the difference in numeri-

cal value between data points in a data set of measurements. Therefore, determinant errors will affect accuracy while random errors affect precision.

Experimental Data Calculations

In many situations only one measurement of each chemical parameter, such as free chlorine, alkalinity, or pH, is completed and thus there are few calculations that are required. As discussed earlier, if this measurement is not representative of the pool or spa then multiple measurements from different locations need to be performed or a combined sample generated.

When multiple measurements are completed and a set of values (known as a data set) is generated there are several calculations that can and should be performed. In this specific situation, calculation of the average (sample mean) and the standard deviation, which is a measure of the error in the measurements, should be completed. If multiple measurements are collected from different locations in a large pool, the greater the standard deviation the larger the difference in pool water conditions in the pool. Table 3 provides the equations for calculating the average and standard deviation. Many inexpensive calculators have these equations programmed for easy calculation and all spreadsheets have single functions that perform these tasks. Although it may make sense to generate an average from measurements collected from a single pool over a short time period, it does not make sense to average values collected over several days or weeks or with data collected from different pools or spas.

Record Keeping

Record keeping is a must in a chemical measurement program. Good record keeping can aid in understanding these water systems, assist in chemical ordering and purchasing, and serve as a reference data in legal situations. Each pool or spa should have a separate logbook or section in a single notebook where qualitative observations of bulk water characteristics and quantitative data from chemical tests are recorded and graphed. By observing the trends in the chemical data and matching this information with the recorded observations a greater understanding of the water quality dynamics for a particular pool and or pools in a specific geographical location will be obtained. Appropriate discussions concerning the practice of keeping a notebook for experimental data can be found in introductory chemistry textbooks and laboratory books. Several copies from different authors can usually be found in public libraries.

Plotting Data

Whether making a single measurement or calculating an average from multiple measurements, it is advantageous to keep a running list of the values for each pool or spa analyzed. Unfortunately, most individuals are not able to see the trends and features in a data set unless a chart or plot of the values is generated. Once data are plotted in a simple graph with values plotted according to magnitude and date/time, many of the trends in the data are observable. An example plot is provided in Figure 1. It can clearly be seen that the chlorine demand is increasing in the

Sample Average (X_{avg})	=	$\frac{(X_1 + X_2 + X_3 + X_4 + X_5 + \dots X_N)}{N}$
where:	X_1	= data point 1
	X_n	= last data point
	N	= total number of data points
Sample Standard Deviation (s)	=	$\{[(X_1 - X_{avg})^2 + \dots (X_n - X_{avg})^2] \div (N - 1)\}^{-1/2}$
where:	X_1	= data point 1
	X_n	= last data point
	N	= total number of data points

Table 3 – Equations for Calculating the Average and Standard Deviation

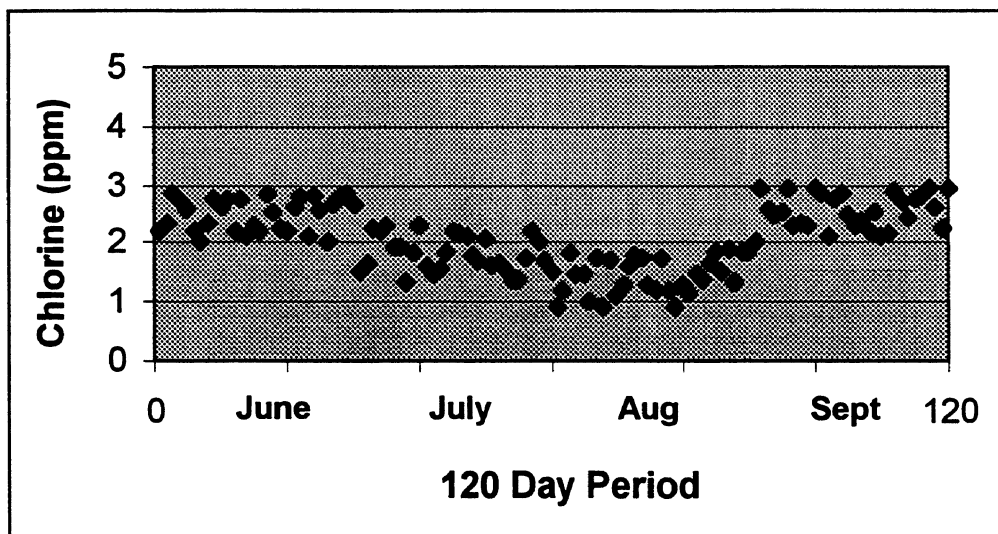


Figure 1 – Graph of hypothetical chlorine concentration measured over a four month summer season. Notice the trends in the data.

months of July and August when compared to June and September. The reason(s) or cause effect relationships for these trends are not indicated by the data values. The trends could be due to low stabilizer level, greater bather load over certain periods, a change in chlorine source or greater solar radiation levels. Comparison of these trends with nearby pools and with pools that lie further away can provide valuable information, which can be used in the next season to understand chlorine demands and the need for modifications in the measurement of chemical parameters and chemical application.

Summary

Maintaining water quality in a pool or spa requires continual measurement and chemical addition. In order to minimize the cost and time required with this maintenance task the variables that control water chemistry must be understood and the error associated with chemical measurements must be minimized. A list of important concepts that should be included in any chemical measurement plan is provided in Table 4.

1. Read the manufacturer's instructions for each test kit. If the instructions are not clear call the manufacturer of the test kit and ask for clarification.
2. Make sure the water sample taken is representative of the pool or spa, and if it is not take multiple measurements in different locations and calculate the average, or combine the samples before measurement.
3. If there are any questions concerning the validity (age, contamination, and decomposition) of the reagents used in the test kit, a solution of known concentration should be evaluated or the reagent replaced.
4. Chemical measurements must be conducted frequently for maintaining water quality in pools and spas and chemical levels adjusted.
5. Good record keeping and the plotting of data aids tremendously in understanding water systems and minimizing the cost and application of treatment chemicals.

Table 4 – Fundamental Concepts in All Chemical Measurement Programs

About the Author

Kenneth D. Hughes, Ph.D. received a BS degree from Muhlenberg College in Pennsylvania and a Ph.D. from Purdue University in Analytical Chemistry. He performed research at the University of Illinois – Urbana. Dr. Hughes has recently moved from Georgia Tech to Kennesaw State University, both in Atlanta. His research interests include water chemistry measurement and the use of lasers in environmental measurements.

General References

1. *Standard Methods for the Examination of Water and Wastewater*, 19th edition, American Public Health Association, 1995.
2. *Chemistry*, 4th edition, Steven S. Zumdahl, Houghton Mifflin Company, New York, 1997.