Analysis of Dissolved Oxygen Concentration and Temperature Levels in the upper Chattahoochee River and its Effects on Rainbow Trout

Brock Akerman

University of North Georgia

Author Note

Brock Akerman, Undergraduate, University of North Georgia,

Correspondence concerning this article should be addressed to Gina Reed, Department of Mathematics, University of North Georgia, Gainesville, GA, 30503.

Contact: Gina.Reed@ung.edu

Abstract

The Georgia Department of Natural Resources (GDNR) takes measurements of dissolved oxygen concentrations and temperature from within the Chattahoochee River. Both variables have implications for quality of life for rainbow trout. I acquired the sampled data from the GDNR and used a series of non-parametric tests to examine both variables to determine their likely population median values. Analysis of the data found that although the population median dissolved oxygen concentration near the Buford Dam were close to a level that would adversely affect the quality of life for rainbow trout, the dissolved oxygen levels were within parameters considered acceptable by research and surveying reports by the Georgia Department of Natural Resources (Brown, 1997). Additionally, the population median for dissolved oxygen concentration levels near the trout hatchery and the population median temperatures for both locations were comfortably within acceptable parameters for quality of life for rainbow trout.

Introduction

Two variables, which set the foundation for an ideal habitat for rainbow trout in the Chattahoochee River, are dissolved oxygen concentrations and temperature. These two variables must remain within a particular range in order to provide a reliable quality of life year-round. *Our research question; does dissolved oxygen concentrations or temperature in the Chattahoochee River just below the Buford Dam depart from acceptable conditions required for rainbow trout during the winter months?*

To answer this question, I met with several federal and state governmental offices to acquire data concerning this question. My first step was a visit to the Buford Dam project office where I met with Timothy Rainey, Army Corps of Engineers Operations Project Manager for Lake Lanier. We looked through several data sets and came across a water quality data set that I decided to work with. The data included dissolved oxygen (DO) concentrations and temperature measurements by two submersible devices in the Chattahoochee River—both managed by the Georgia Department of Natural Resources (GDNR). I also met with GDNR biologist Pat Snellings whose office was responsible for both monitoring stations. Mr. Snellings also provided details about the background for the data collected at these stations and gave me some references for acceptable river trout parameters.

The data analyzed was taken from two instruments used to measure temperature and dissolved oxygen concentrations in the Chattahoochee River. One instrument was located in the waters in front of Buford Dam, and the second instrument placed approximately 1.90 miles downstream in the waters just in front of the Georgia Department of Natural Resources Trout Hatchery. The dissolved oxygen and temperature readings are paired together from each device and were taken together in 10-minute intervals. The first instrument reading was performed on

September 25, 2014 at 1200 and ran continuously until terminating on March 31, 2015 at 0650. The frequency of readings has provided a very large sample of data to examine—each with a sample size of 26,884 individual points of data.

The stretch of Chattahoochee River that runs from Buford Dam to Morgan Falls reservoir is mostly perfect for two trout species—brown trout and rainbow trout. Lake Lanier experiences stratification during the summer months; a condition that occurs in warm monomictic lakes that prohibits the establishment of a habitat for trout year-round. The Chattahoochee River south of Morgan Falls Reservoir reaches temperatures that discourage downstream habitats. Although several government bureaus monitor this tract of the river, not much information about this portion of the Chattahoochee River is readily available to the public. Lake Lanier covers an area of approximately 39,000 acres. What makes its way into the lake will eventually make its way to these trout waters. Further water quality studies regarding the upper Chattahoochee River would be helpful in examining long-term trends and implications for all marine biology.

According to a study published by the Georgia Department of Natural Resources, brown trout are more tolerant to changes in the water than their rainbow trout counterparts (Brown, 1997). Brown trout have flourished in the river for quite some time; therefore, there has not been a demand for the Buford Trout Hatchery to stock brown trout in their fishery. In fact, the hatchery has purposefully chosen not to do so for more than a decade because of their ability to propagate naturally in their environment. Instead, the hatchery has shifted its focus on raising and supplying rainbow trout, which are more sensitive to environmental changes.

Given how hardy brown trout are inside the Chattahoochee River and how comfortable the trout hatchery is with their ability of self-sustainment, it would be somewhat trivial to examine waters as they pertain to brown trout. Like the hatchery, I will instead focus on the more sensitive rainbow trout. Higher temperatures and lower dissolved oxygen levels can affect all aspects of quality of life for all fish—rainbow trout included. When temperatures and dissolved oxygen begin approaching levels considered unideal, trout can become stressed. This can affect eating habits, propagation, disease vulnerability, buoyancy, and even mortality (Brown, 1997). It is important that the Chattahoochee River environment remain within the confines of what is considered an acceptable range for trout species to thrive.

The purpose of this analysis is to investigate the dissolved oxygen concentrations and temperature record samples from the Buford Dam and the Trout Hatchery of the Chattahoochee River and determine whether the water retreats into less ideal condition for rainbow trout. I theorize that the median dissolved oxygen concentration and median temperatures fall comfortably within the boundaries for what is considered by the GDNR to be an ideal habitat for a rainbow trout.

Methodology

The data I have acquired was provided to me by the Army Corps of Engineers and authored by biologist Patrick O'Rouke with the Georgia Department of Natural Resources. The data was large, quantitative and continuous, univariate, and sampled from differing populations. The initial steps involved the paring down of choices between all of the various methods of testing. This was accomplished by using Minitab to determine whether the data belonged to a distribution or whether it was distribution free. An examination of skewness determined that the best measure of central tendency was the median. Central tendency and other attributes belonging to the data eliminated methods for testing eventually leading our data to be classified as non-normal and non-parametric.

Distribution and parametric testing having been completed, I worked primarily with the Wilcoxon Signed Rank Test since the data satisfied the Wilcoxon test assumptions best. The Wilcoxon Signed Rank Test was also used to compare both paired groups and their median. To do this required the use of Minitab's calculator formula function to create an expression that took the difference of the paired data and placed the results in a new column. Then, a Wilcoxon Signed Rank Test was performed on this new column. The purpose of using the Wilcoxon testing was to explain as much as possible about the details surrounding the population, at an appropriate significance level, using inferential statistics from the sampled data. I explained in detail the use of Anderson-Darling, Skewness, Kurtosis, and the Wilcoxon Signed Rank test in statistical software and the process of calculating these statistics manually and without the use of software.

Proofs

To hypothesize about the population median, this analysis will require the use of the Wilcoxon Signed Rank test for three different occasions; once to test the population median, once more to test for confidence intervals of the population median, and finally a test for the median differences between both sample data sets. To prove the Wilcoxon Signed Rank Test mean and variance requires the proof of two summations that appear in each proof. See the appendix for both summation proofs.

<u>Proof of Wilcoxon Signed Rank Test mean.</u>

The Central Limit Theorem (CLT) states the probability distribution of a sum of independent and identically distributed (i.i.d.) variables approaches a normal distribution $N(\mu,\sigma)$ given N is significantly large. We can draw from the CLT—and thus a normal distribution—that the mean of the Wilcoxon Signed Rank Test is the first standardize moment of W.

The Wilcoxon Signed Rank Test distribution is $W = \sum_{i=1}^{n} Z_i R_i$. W $= T = \sum_{i=1}^{n} T_i$ where T_i is a Bernoulli Distribution. Under symmetry, $E(T^-) = E(T^+)$.

- $T_i = 0$ with probability $\frac{1}{2}$
- $T_i = i$ with probability of $\frac{1}{2}$

In other words, the distribution is a Bernoulli distribution with the probability of p $= \frac{1}{2}$ and a $(1-p) = q = \frac{1}{2}$.

Proof: We will prove by direct proof that the mean of the Wilcoxon Signed Rank Test is

$$\mu = \frac{n(n+1)}{4}$$

$E[T] =$ (From the CLT, the mean is the first standardized moment)

$E\left[\sum_{i=1}^{n} ix_i\right] =$ $(Since\ T = \sum_{i=1}^{n} Z_i R_i); Z_i = x_i,\ R_i = the\ Index\ i.)$

$\sum_{i=1}^{n} E[ix_i] =$ (Linearity of Expectation property $E\left[\sum_{i=1}^{x} F(x)\right] = \sum_{i=1}^{n} E(F(x))$)

$\sum_{i=1}^{n} iE[x_i] =$ (Property: E(aX) = aE(X))

$E[x_i]\sum_{i=1}^{n} i =$ (Property: $C \cdot \sum_{i=1}^{x} F(x) = \sum_{i=1}^{n} C \cdot F(x)$)

$\frac{1}{2}\sum_{i=1}^{n} i =$ (Expected Value of a Bernoulli Trial is ½)

$\frac{1}{2} \times \frac{n(n+1)}{2} =$ (from Induction Proof (1); see appendix)

$\frac{n(n+1)}{4} = RHS$

Q.E.D.

<u>Proof of Wilcoxon Signed Rank Test mean.</u>

Proof: We will prove by direct proof that the variance of the Wilcoxon Signed Rank Test is

$$\sigma^2 = \frac{n(n+1)(2n+1)}{24}$$

$var[T] =$ From the CLT, the variance is the second standardized moment

$var\left[\sum_{i=1}^{n} i^2 x_i\right] =$ $(Since\ T = \sum_{i=1}^{n} Z_i R_i); Z_i = x_i,\ R_i = the\ Index\ i.)$

$\sum_{i=1}^{n} var\left[i^2 x_i\right] =$ (Linearity of Expectation property/Bienaymé Formula:

$$\text{Var}\left[\sum_{i=1}^{x} F(x)\right] = \sum_{i=1}^{n} Var(F(x)))$$

$\sum_{i=1}^{n} i^2 var[x_i] =$ (Property: var(aX) = $a^2 \cdot$ var(X))

$var[x_i] \sum_{i=1}^{n} i^2 =$ (Property: $C \cdot \sum_{i=1}^{x} F(x) = \sum_{i=1}^{n} C \cdot F(x)$)

$\frac{1}{4} \sum_{i=1}^{n} i^2 =$ (Variance of a Bernoulli Trial is $= p \cdot q = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$)

$\frac{1}{4} \times \frac{n(n+1)(2n+1)}{6} =$ (from Induction Proof (2); see appendix)

$\frac{n(n+1)(2n+1)}{24} = RHS$

$Q.E.D.$

Procedural Calculations

Anderson-Darling normality test process

Anderson-Darling formula:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[lnF(X_i) + ln\left(1 - F\left(X_{n-i+1}\right)\right)\right]$$

The Anderson-Darling normality test is essentially a goodness-of-fit test that formulaically looks for the absence of normality. The test generates a real number called an Anderson-Darling statistic and this statistic is compared to other known critical values associated with various distribution types. The smaller the Anderson-Darling statistic, the likelihood that our sampled data will belong to a particular distribution increases. To demonstrate how the Anderson-Darling test generates its namesake's statistic, I will use the first ten data points drawn from the 'Hatchery DO' dataset and show its application within Microsoft Excel using a manual formula-free input method.

Macro Data		Hatchery_DO	Sequence	Sort	$F(X_i)$	$1-F(X_i)$	$F(X_{n-i+1})$	S_i
		6.14	1	6.14	0.05573	0.9443	0.08484	-5.354
Mean	6.242	6.19	2	6.17	0.1306	0.8694	0.0848	-13.51
Standard Dev	0.06408328	6.17	3	6.19	0.20856	0.7914	0.2269	-15.25
N	10	6.22	4	6.22	0.36569	0.6343	0.3894	-13.64
		6.24	5	6.24	0.48755	0.5124	0.4503	-13.65
		6.25	6	6.25	0.54967	0.4503	0.5124	-13.94
		6.26	7	6.26	0.6106	0.3894	0.6343	-12.33
		6.29	8	6.29	0.77308	0.2269	0.7914	-7.369
		6.33	9	6.33	0.91516	0.0848	0.8694	-3.886
		6.33	10	6.33	0.91516	0.0848	0.9443	-2.774
							$Sum(S_i)$ =	-101.7
			$A^2 = -n-(1/n)*S_i = -10-(1/10)*(-101.7) =$					0.17

The Anderson-Darling statistic has many calculable parts. The mean and standard deviation of our data is required so I calculated both and placed them off to the side for use later. Next, I worked on the summation portion of the formula. Since we will be working with ordered -ith terms, I took our ten data points and reorganized them in the 'sort' column by their ascending order. The 'sequence' column is the -ith value assigned to each ascending ordered data point. From here, we will need to calculate the cumulative distribution function of a normal distribution for each value and then take its compliment. The CDF is denoted by:

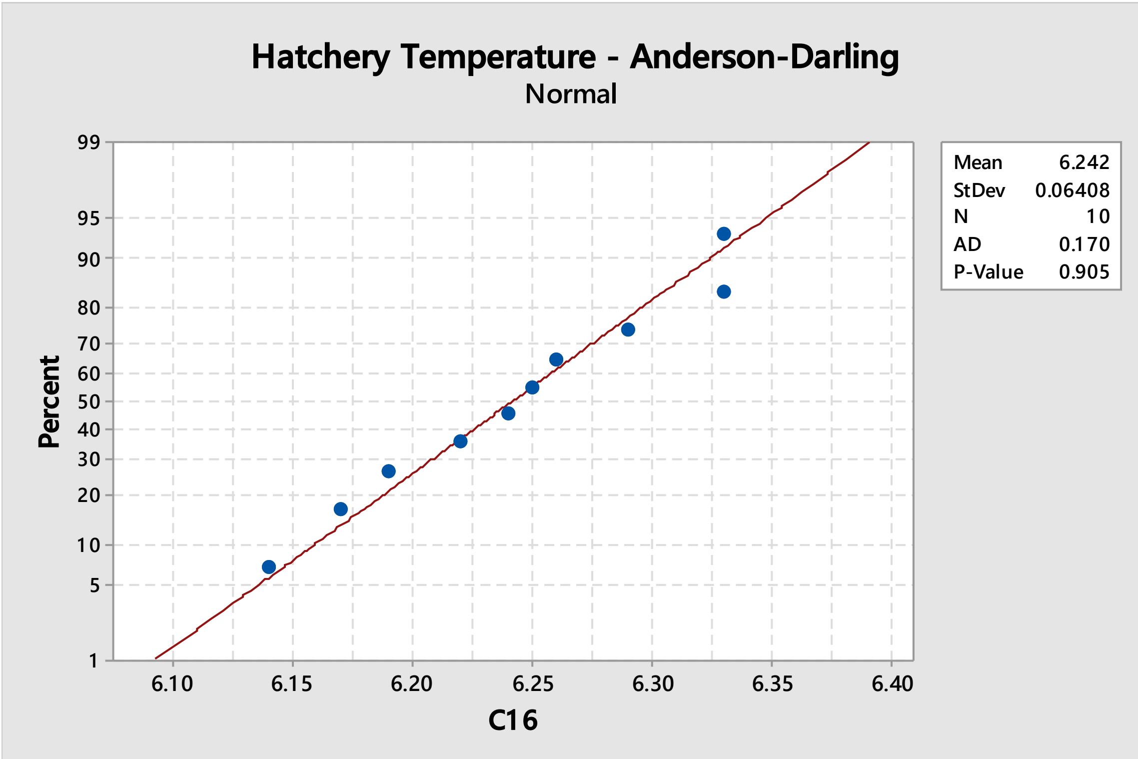
$$F(X_i) = \frac{1}{\sqrt{2\pi}\,(\sigma)}\int_{-\infty}^{x} e^{-\frac{(t-\bar{x})^2}{2\sigma^2}}\,dt$$

Since the output is a cumulative distribution function (CDF), each value collects the probability of the values before it and adds to it its own. We can use this property of CDF's to check the CDF column for errors as our data should be on an interval $0 \leq F(x_i) \leq 1$ and increasing—which from observation, it is. The next column is the compliment of the probabilities of $F(x_i)$ and so it will be decreasing and on the same interval. We will use the compliment to calculate the other function of the Anderson-Darling formula; $F(x_{n-i+1})$. The function $F(x_{n-i+1})$ is the cumulative distribution function of the compliment of the cumulative distribution function of $F(x_i)$. What $F(x_{n-i+1})$ does is reorders the column $1-F(x_i)$ in ascending order.

Now that each relevant term of our summation has been calculated, we only need to enter them into the summation portion of the Anderson-Darling formula and sum them. The 'S_i' column collects and sums the -ith terms and we are left with an integer that we use to calculate the Anderson-Darling statistic.

$$AD = -n - \frac{1}{n} * S_i = -10 - \frac{1}{10} * (-101.7) = .170$$

To check our work, I placed the same ten values into Minitab and performed an Anderson-Darling normality test. From the Minitab output, the AD value is the same value we found by manually calculating it by hand.



<u>Skewness</u>

There are several methods for checking skew and of them; Minitab v.17 uses the adjusted Fisher-Pearson standardized moment coefficient to determine skew.

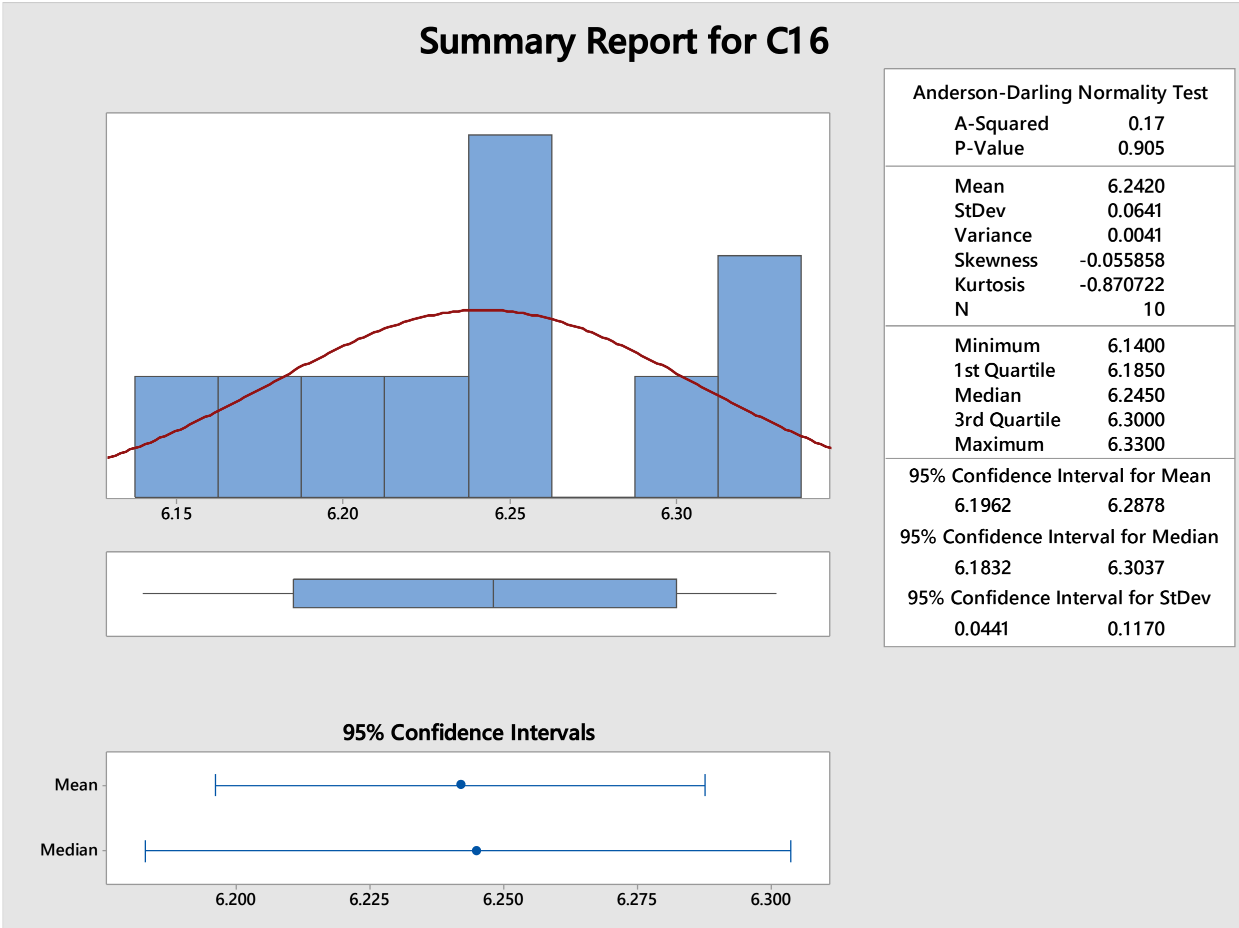
Adjusted Fisher-Pearson standardized moment coefficient

$$G_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

To demonstrate how to apply this formula, I will use the same ten data points we used to demonstrate the calculation of the Anderson-Darling statistic. Our sample size of 10 is much smaller than the 26,884 so we should expect the skew value to be significantly different from the sample skew.

Hatchery_DO	Sequence	Sort	$(X_i\text{-}X)$	$(X_i\text{-}X)^3$	$S_i = (X_i\text{-}X)^3/S^3$
6.14	1	6.14	-0.102	-0.001061208	-4.032425344
6.19	2	6.17	-0.072	-0.000373248	-1.418284346
6.17	3	6.19	-0.052	-0.000140608	-0.53428853
6.22	4	6.22	-0.022	-0.000010648	-0.040460744
6.24	5	6.24	-0.002	-0.000000008	-3.03988E-05
6.25	6	6.25	0.008	0.000000512	0.00194552
6.26	7	6.26	0.018	0.000005832	0.022160693
6.29	8	6.29	0.048	0.000110592	0.420232399
6.33	9	6.33	0.088	0.000681472	2.589487606
6.33	10	6.33	0.088	0.000681472	2.589487606
				$\text{Sum}(S_i) =$	-0.402175538
Mean	6.242				
St.Dev	0.0640833				
N	10	$\text{Skew} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^3 \approx -0.0558577$			

Skew is simpler to calculate than an Anderson-Darling statistic and requires only some algebra. Like Anderson-Darling test, we will need the mean and standard deviation to compute the skew statistic. The column labeled $(X_i\text{-}X)$ takes the difference of each of the observations and the mean. The column to the right cubes the result and the final column divides the difference of each observation minus the mean cubed by the standard deviation cubed. Finally, sum all of the parts and then multiply it by the polynomial coefficient in our formula and you we will have obtained the skew. To check the skew, I ran the ten points of data through Minitab and obtained the same result with the output pictured below.



Kurtosis

The kurtosis formula:

$$Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

This formula—also known as the “sample excess kurtosis formula”—is used to calculate kurtosis in Minitab and it takes into consideration the adjustment for sampled data against the population by subtracting a scaling fraction toward the end of the equation. Kurtosis of a normal distribution is equal to zero. We are interested in values differing from zero, or the excess. After calculating the kurtosis of our data, a numerical value is given that is relative to zero. A number closer to zero shows a distribution whose tailed-ness is close to representing that of a normal distribution. Departures from zero in a positive direction describes a leptokurtic distribution where data is found primarily in the center of the distribution and less in the tails. The shape of a leptokurtic distribution is taller and narrower than a normal distribution with longer tails and higher variability within the two deviations from the mean. A departure from zero in a negative direction describes a platykurtic distribution that has nearly as much data in the center of the distribution as is found in the tails. A platykurtic distribution has an arch that falls below a normal distribution with shorter tails and variability found further from central tendency than compared to a normal distribution.

Hatchery_DO	Sequence	Sort	$(X_i\text{-}X)$	$(X_i\text{-}X)^4$	$S_i = (X_i\text{-}X)^4/S^4$
6.14	1	6.14	-0.102	0.000108243	6.41832613
6.19	2	6.17	-0.072	2.68739E-05	1.593496374
6.17	3	6.19	-0.052	7.31162E-06	0.433545286
6.22	4	6.22	-0.022	2.34256E-07	0.013890306
6.24	5	6.24	-0.002	1.6E-11	9.48727E-07
6.25	6	6.25	0.008	4.096E-09	0.000242874
6.26	7	6.26	0.018	1.04976E-07	0.006224595
6.29	8	6.29	0.048	5.30842E-06	0.314764716
6.33	9	6.33	0.088	5.99695E-05	3.555918367
6.33	10	6.33	0.088	5.99695E-05	3.555918367
				$\text{Sum}(S_i) =$	15.89232796

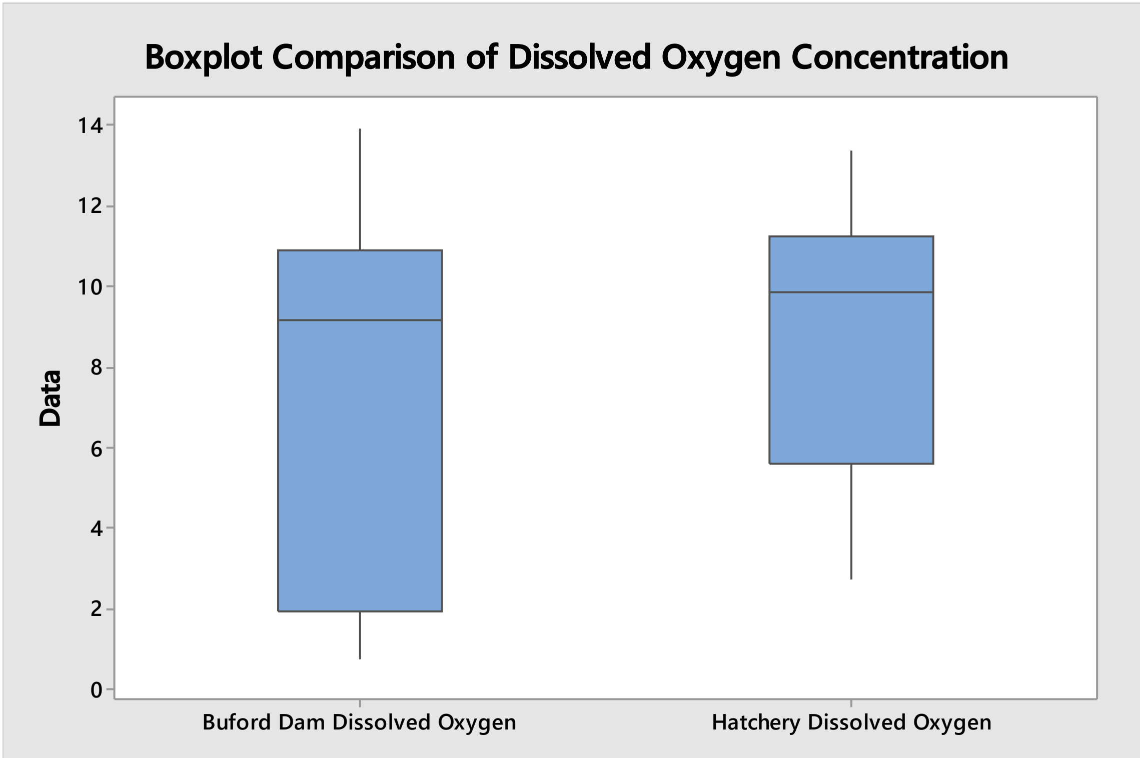
Mean	6.242
St.Dev	0.0640833
N	10

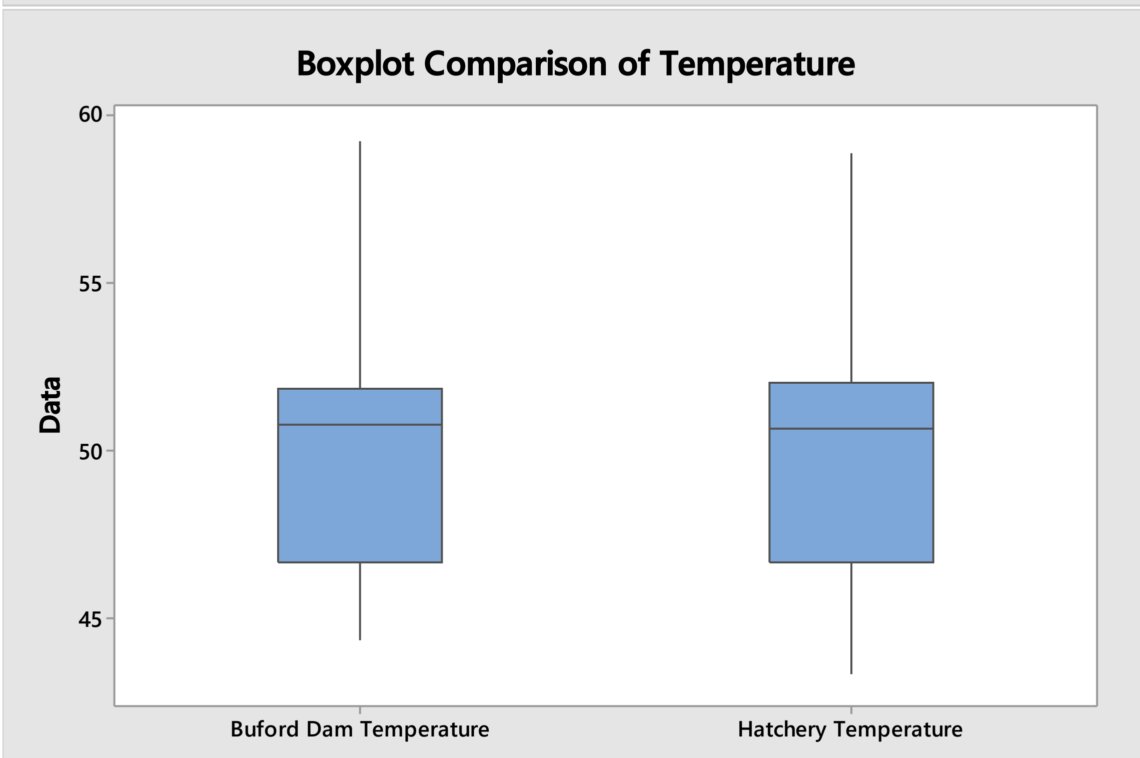
$$\text{Kurt} = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i-1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \approx -0.8707243$$

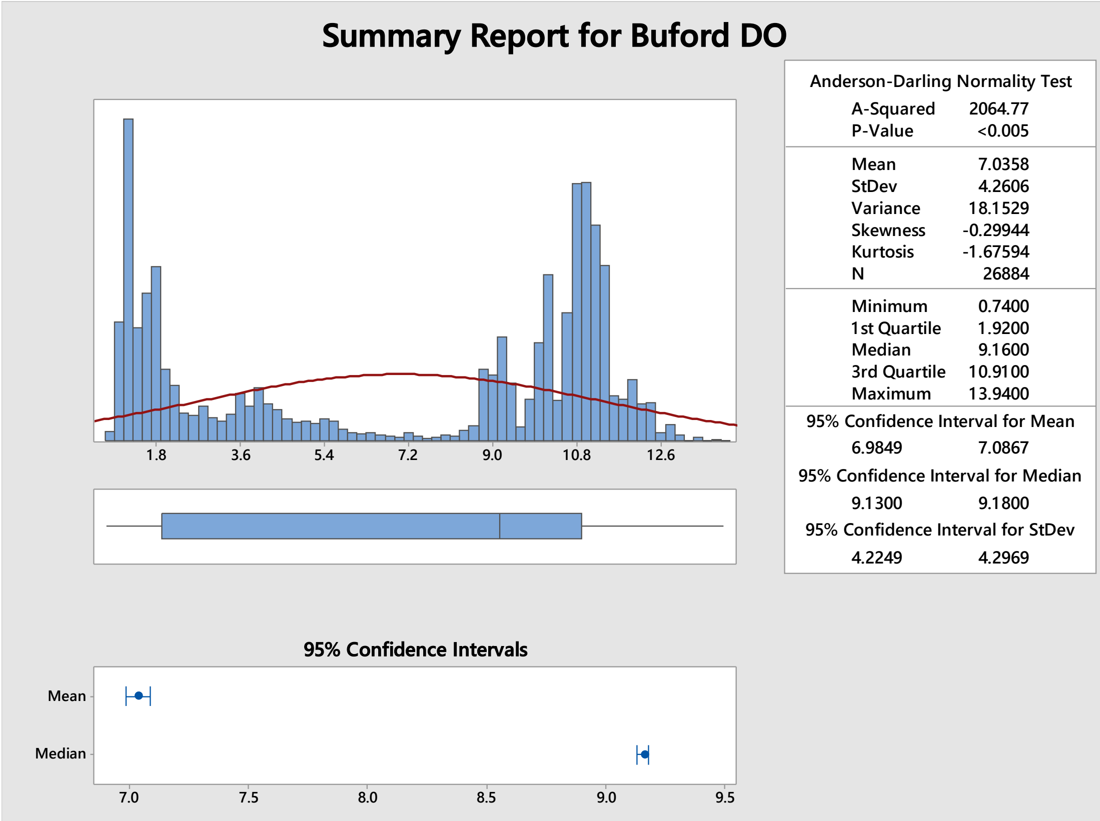
Calculating the kurtosis follows the same concept as calculating skewness. Find the fourth power of the difference between each observation and the mean; then divide that result by the square root of the sample standard deviation to the fourth power. The fractional components of the formula are simple algebra. Replace 'n' with the sample size. After calculating the different components of the kurtosis formula and enter them into Microsoft Excel, we can see that the same conclusion from Microsoft Excel as that found in the Minitab output from our earlier calculation of skewness has been reached.

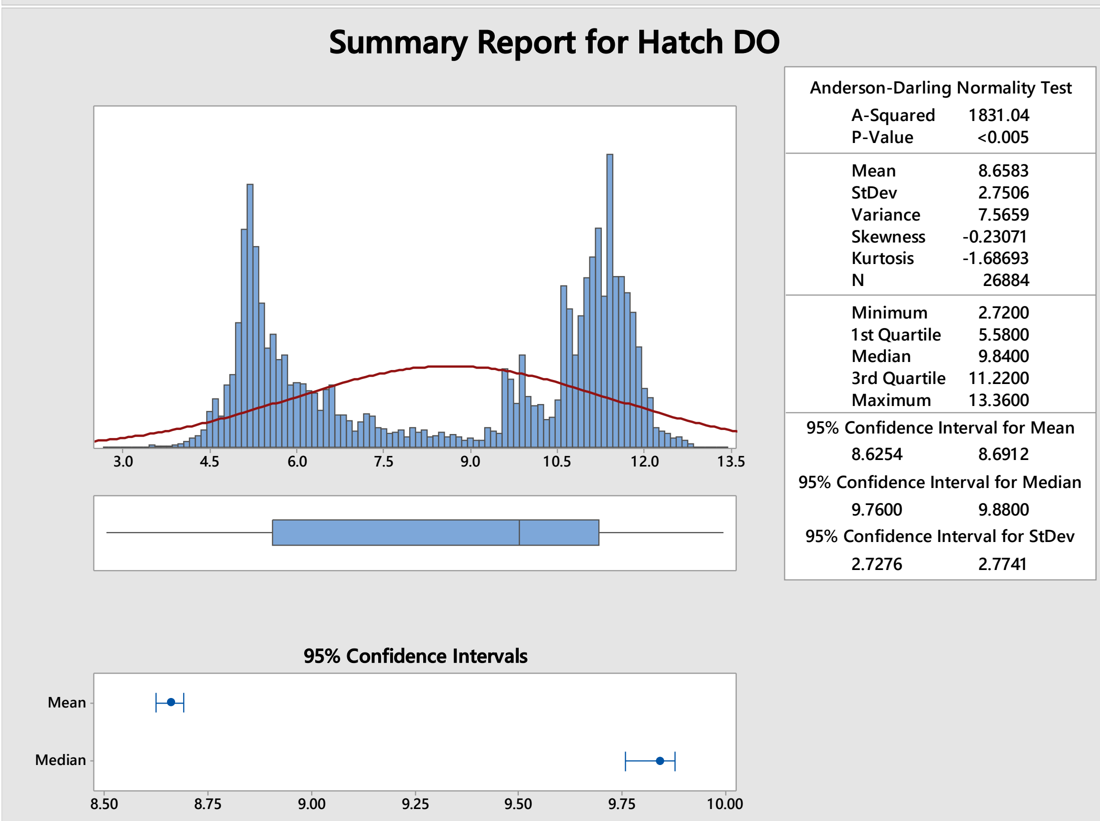
Results

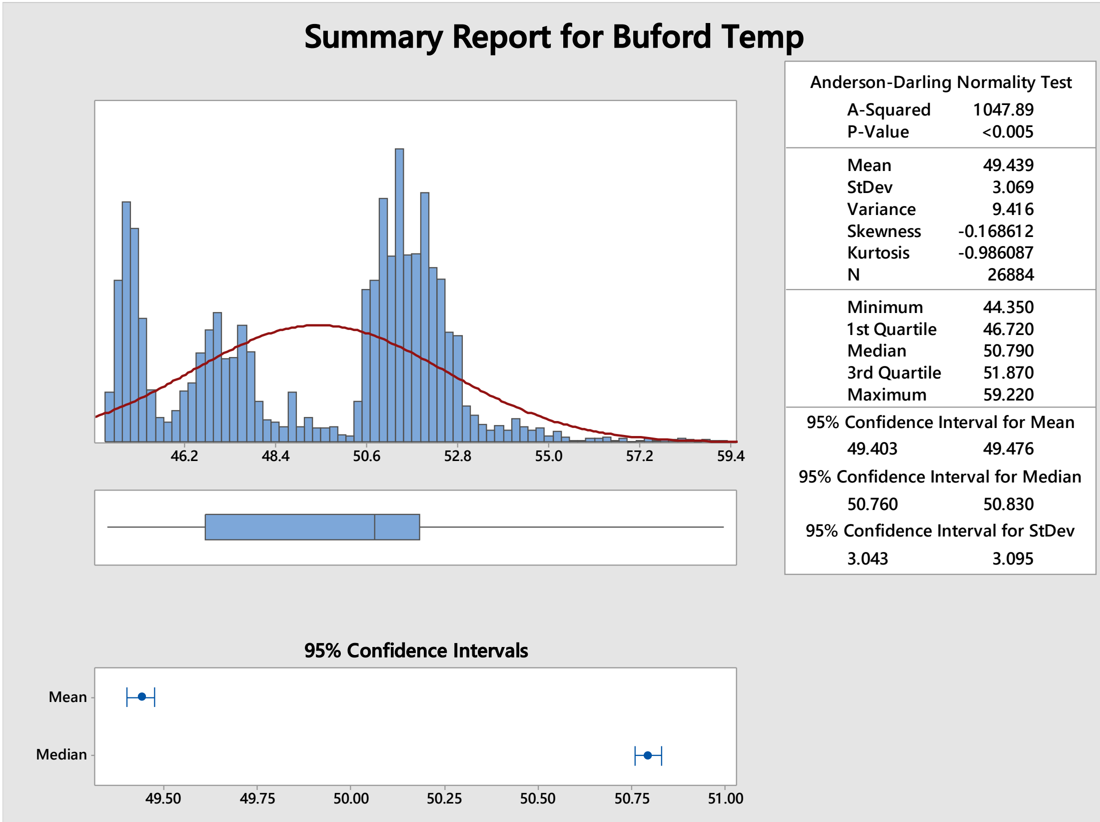
Box-plot output:

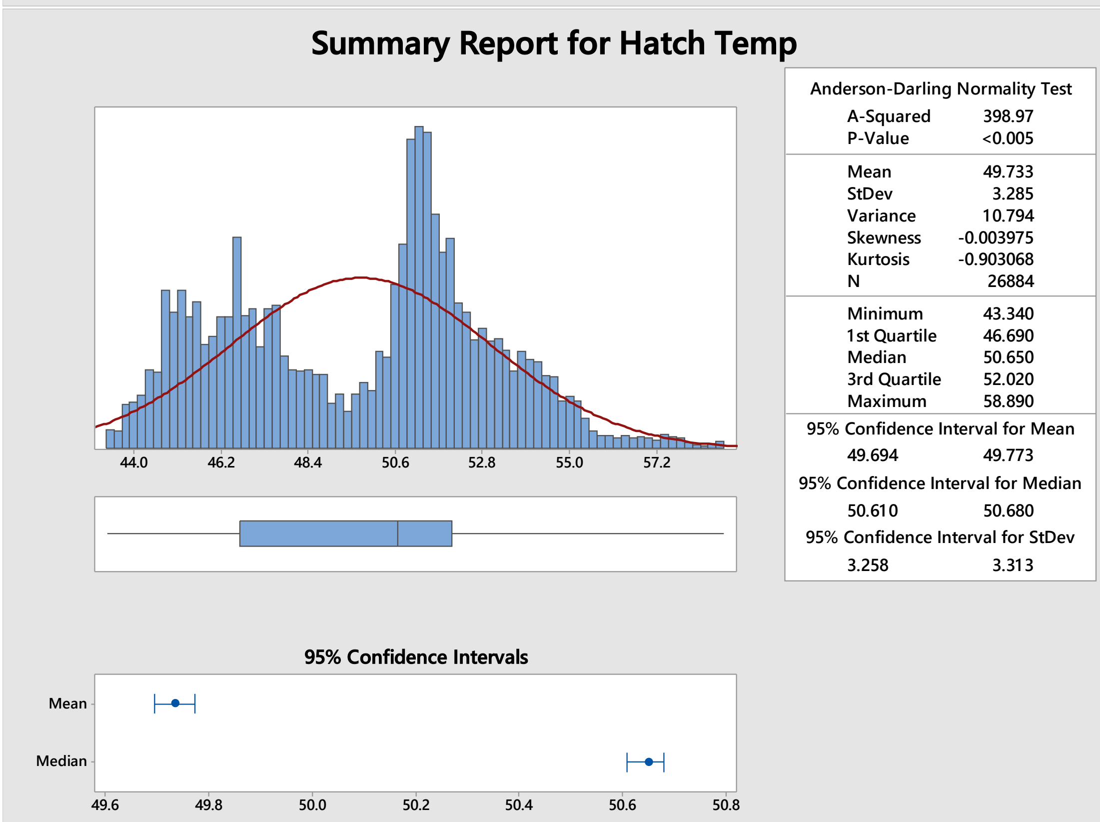




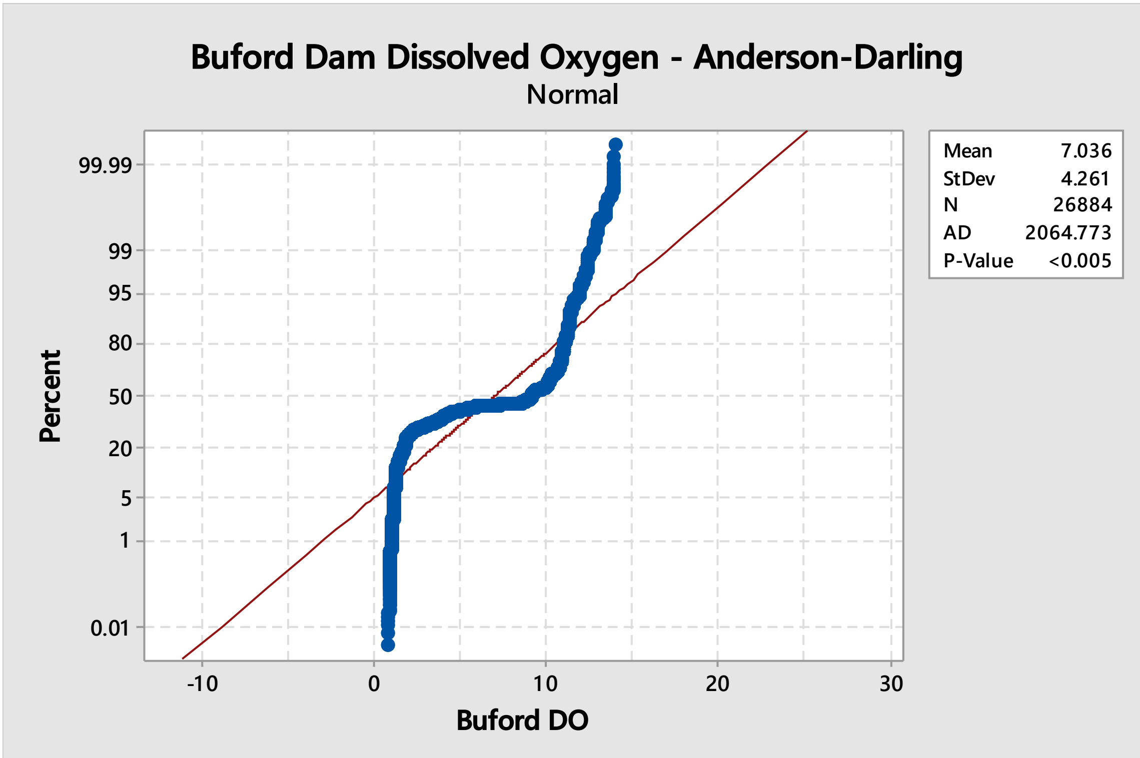

Summary Report for Buford DO
Anderson-Darling Normality Test
A-Squared 2064.77
P-Value <0.005
Mean 7.0358
StDev 4.2606
Variance 18.1529
Skewness -0.29944
Kurtosis -1.67594
N 26884
Minimum 0.7400
1st Quartile 1.9200
Median 9.1600
3rd Quartile 10.9100
Maximum 13.9400
95% Confidence Interval for Mean
6.9849 7.0867
95% Confidence Interval for Median
9.1300 9.1800
95% Confidence Interval for StDev
4.2249 4.2969
1.8 3.6 5.4 7.2 9.0 10.8 12.6
95% Confidence Intervals
Mean
Median
7.0 7.5 8.0 8.5 9.0 9.5

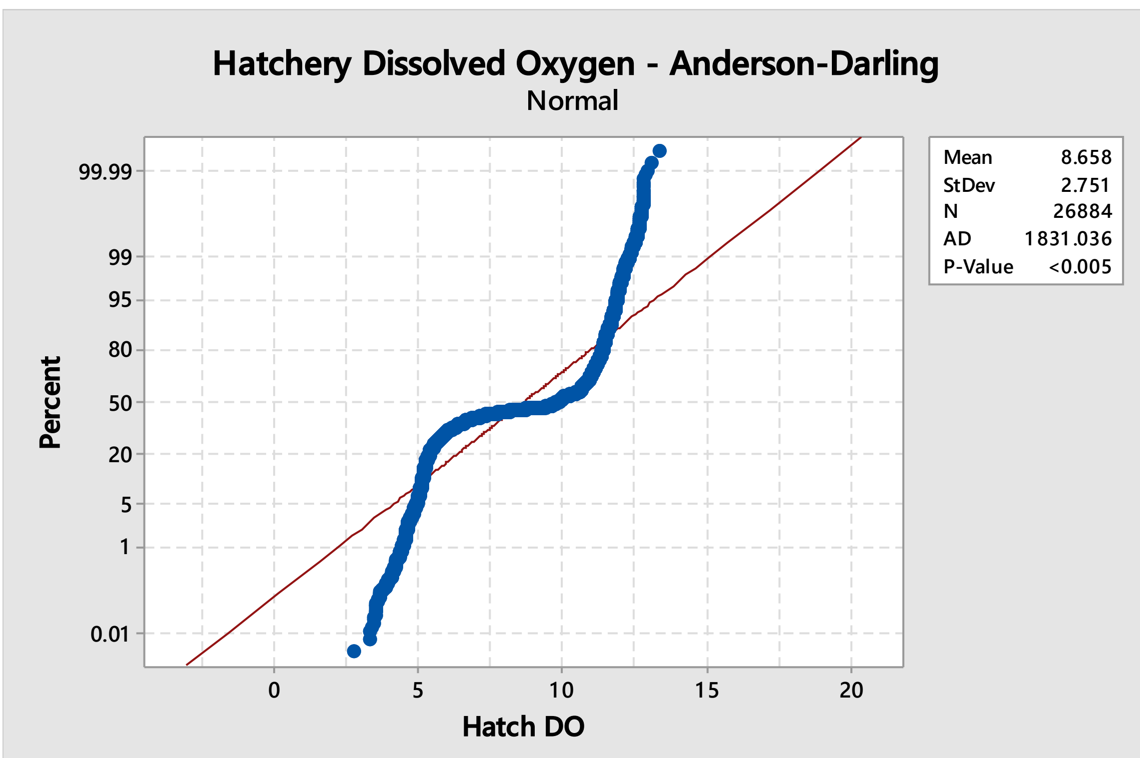

Summary Report for Hatch DO
Anderson-Darling Normality Test
A-Squared 1831.04
P-Value <0.005
Mean 8.6583
StDev 2.7506
Variance 7.5659
Skewness -0.23071
Kurtosis -1.68693
N 26884
Minimum 2.7200
1st Quartile 5.5800
Median 9.8400
3rd Quartile 11.2200
Maximum 13.3600
95% Confidence Interval for Mean
8.6254 8.6912
95% Confidence Interval for Median
9.7600 9.8800
95% Confidence Interval for StDev
2.7276 2.7741
3.0 4.5 6.0 7.5 9.0 10.5 12.0 13.5
95% Confidence Intervals
Mean
Median
8.50 8.75 9.00 9.25 9.50 9.75 10.00

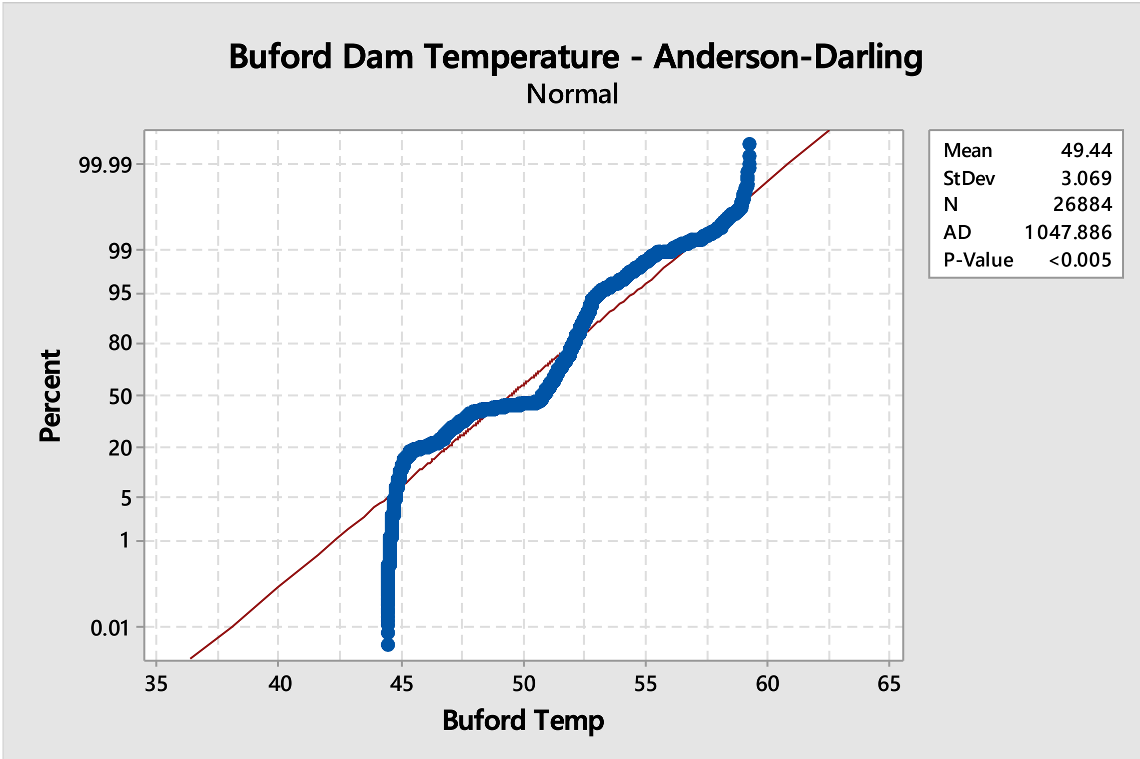

Summary Report for Buford Temp
Anderson-Darling Normality Test
A-Squared 1047.89
P-Value <0.005
Mean 49.439
StDev 3.069
Variance 9.416
Skewness -0.168612
Kurtosis -0.986087
N 26884
Minimum 44.350
1st Quartile 46.720
Median 50.790
3rd Quartile 51.870
Maximum 59.220
95% Confidence Interval for Mean
49.403 49.476
95% Confidence Interval for Median
50.760 50.830
95% Confidence Interval for StDev
3.043 3.095
46.2 48.4 50.6 52.8 55.0 57.2 59.4
95% Confidence Intervals
Mean
Median
49.50 49.75 50.00 50.25 50.50 50.75 51.00

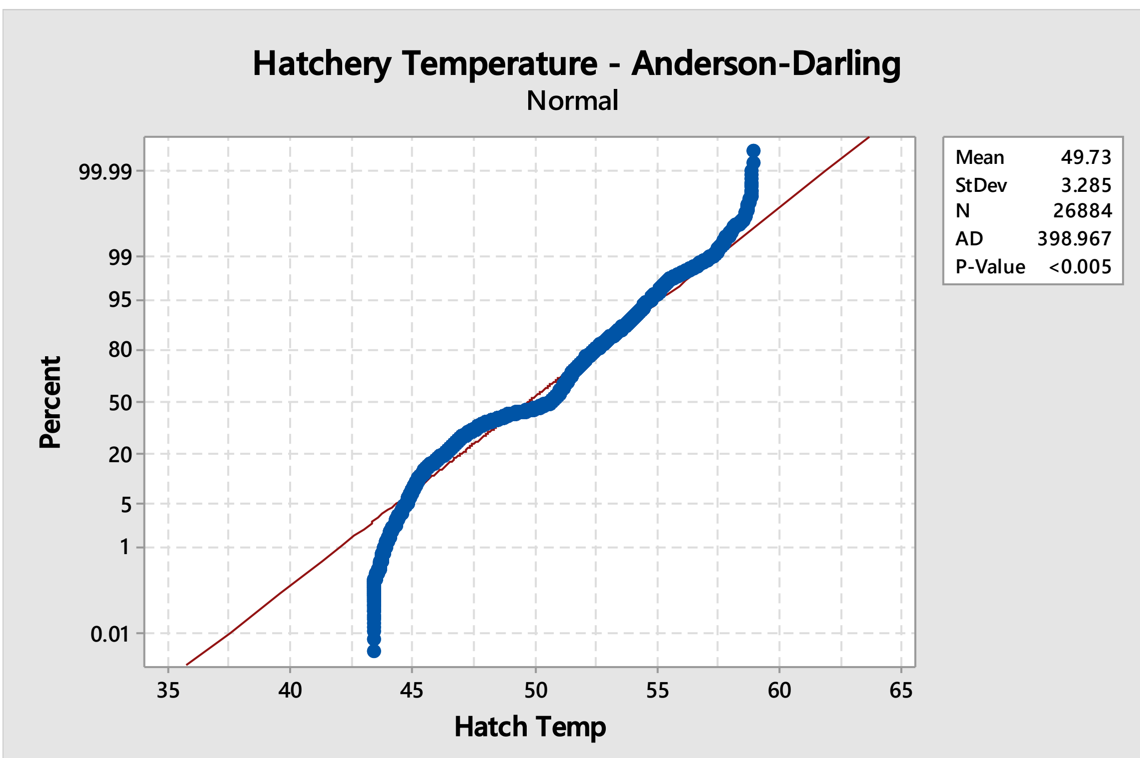

Summary Report for Hatch Temp
Anderson-Darling Normality Test
A-Squared 398.97
P-Value <0.005
Mean 49.733
StDev 3.285
Variance 10.794
Skewness -0.003975
Kurtosis -0.903068
N 26884
Minimum 43.340
1st Quartile 46.690
Median 50.650
3rd Quartile 52.020
Maximum 58.890
95% Confidence Interval for Mean
49.694 49.773
95% Confidence Interval for Median
50.610 50.680
95% Confidence Interval for StDev
3.258 3.313
44.0 46.2 48.4 50.6 52.8 55.0 57.2
95% Confidence Intervals
Mean
Median
49.6 49.8 50.0 50.2 50.4 50.6 50.8

Anderson-Darling Normality output:






Buford Dam Temperature - Anderson-Darling
Normal
Percent
99.99
99
95
80
50
20
5
1
0.01
35
40
45
50
55
60
65
Buford Temp
Mean 49.44
StDev 3.069
N 26884
AD 1047.886
P-Value <0.005


Hatchery Temperature - Anderson-Darling
Normal
Percent
99.99
99
95
80
50
20
5
1
0.01
35
40
45
50
55
60
65
Hatch Temp
Mean 49.73
StDev 3.285
N 26884
AD 398.967
P-Value <0.005

Anderson-Darling Hypothesis Testing:

H_o: The Buford Dam Dissolved Oxygen sample data set follows a specified distribution.

H_a: The Buford Dam Dissolved Oxygen sample data set does not follow a specified distribution.

Sample Mean = 7.036

Sample Standard Deviation = 4.261

Sample Size = 26884

Anderson-Darling Statistic = 2064.773

P-Value < 0.005

Reject H_o. There is sufficient evidence at the alpha level of 0.05 to reject the claim that the Buford Dam Dissolved Oxygen sample data set follows a specified distribution.

H_o: The Hatchery Dissolved Oxygen sample data set follows a specified distribution.

H_a: The Hatchery Dissolved Oxygen sample data set does not follow a specified distribution.

Sample Mean = 8.658

Sample Standard Deviation = 2.751

Sample Size = 26884

Anderson-Darling Statistic = 1831.036

P-Value < 0.005

Reject H_o. There is sufficient evidence at the alpha level of 0.05 to reject the claim that the Hatchery Dissolved Oxygen sample data set follows a specified distribution.

H_o: The Buford Dam Temperature sample data set follows a specified distribution.

H_a: The Buford Dam Temperature sample data set does not follow a specified distribution.

Sample Mean = 49.440

Sample Standard Deviation = 3.069

Sample Size = 26884

Anderson-Darling Statistic = 1047.886

P-Value < 0.005

Reject H_o. There is sufficient evidence at the alpha level of 0.05 to reject the claim that the Buford Dam Temperature sample data set follows a specified distribution.

H_o: The Hatchery Temperature sample data set follows a specified distribution.

H_a: The Hatchery Temperature sample data set does not follow a specified distribution.

Sample Mean = 49.73

Sample Standard Deviation = 3.285

Sample Size = 26884

Anderson-Darling Statistic = 398.967

P-Value < 0.005

Reject H_o. There is sufficient evidence at the alpha level of 0.05 to reject the claim that the Hatchery Temperature data set follows a specified distribution.

Skewness:

Variable	Skewness
Buford Dam DO	-0.29944
Hatchery DO	-0.23071
Buford Dam Temperature	-0.16861
Hatchery Temperature	-0.00398

Wilcoxon Signed Rank Test to compare one group to a hypothetical value:

Biologists and surveyors with the Georgia Department of Natural Resources have performed stress-test research on rainbow trout. Their findings concluded that rainbow trout require that their environment have particular temperatures and dissolved oxygen concentration levels. Rainbow trout require water temperatures that are at or below 70 degrees Fahrenheit. Rainbow trout also begin exhibiting signs of stress when dissolved oxygen concentrations are less than 6 mg/L. We will want to use the Wilcoxon Signed Rank Test to examine the data collected to see if Chattahoochee River water at both locations supports habitable conditions for Rainbow Trout.

Is the population median of dissolved oxygen concentrations greater than or equal to 6 mg/L at the Buford Dam?

H_o: The population median of Buford Dam dissolved oxygen concentration is greater than or equal to 6 mg/L.

H_a: The population median of Buford Dam dissolved oxygen concentration is less than 6 mg/L.

Wilcoxon Signed Rank Test: Buford Dam DO

Test of median ≥ 6.000 versus median < 6.000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Buford Dam DO	26884	26879	243180342.0	1.000	6.615

Conclusion: Fail to Reject H_o. There is insufficient evidence at the alpha level of 0.05 significance to reject the claim that the population median of Buford Dam dissolved oxygen concentration is greater than or equal to 6 mg/L. Data supports the claim that Chattahoochee River dissolved oxygen concentrations at the Buford Dam supports a healthy environment for rainbow trout.

Is the population median of dissolved oxygen concentrations greater than or equal to 6 mg/L at the Hatchery?

H_o: The population median of Hatchery dissolved oxygen concentration is greater than or equal to 6 mg/L.

H_a: The population median of Hatchery dissolved oxygen concentration is less than 6 mg/L.

Wilcoxon Signed Rank Test: Hatchery DO

Test of median ≥ 6.000 versus median < 6.000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Hatchery DO	26884	26861	310920493.0	1.000	8.475

Conclusion: Fail to Reject H_o. There is insufficient evidence at the alpha level of 0.05 significance to reject the claim that the population median of Hatchery dissolved oxygen concentration is greater than or equal to 6 mg/L. Data supports the claim that Chattahoochee River dissolved oxygen concentrations at the Hatchery supports a healthy environment for rainbow trout.

Is the population median for temperature less than or equal to 70ºF at the Buford Dam?

H_o: The population median of Buford Dam temperature is less than or equal to 70ºF.

H_a: The population median of Buford Dam temperature is greater than 70ºF.

Wilcoxon Signed Rank Test: Buford Dam Temperature

Test of median ≤ 70.00 versus median > 70.00

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Buford Dam Temperature	26884	26884	0.0	1.000	49.33

Conclusion: Fail to Reject H_o. There is insufficient evidence at the alpha level of 0.05 significance to reject the claim that the population median of Buford Dam temperature is less than or equal to 70ºF. Data supports the claim that Chattahoochee River temperature at the Buford Dam supports a healthy environment for rainbow trout.

Is the population median for temperature less than or equal to 70ºF at the Hatchery?

H_o: The population median of Hatchery temperature is less than or equal to 70ºF.

H_a: The population median of Hatchery temperature is greater than 70ºF.

Wilcoxon Signed Rank Test: Hatchery Temperature

Test of median ≤ 70.00 versus median > 70.00

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Hatchery Temperature	26884	26884	0.0	1.000	49.64

Conclusion: Fail to Reject H_o. There is insufficient evidence at the alpha level of 0.05 significance to reject the claim that the population median of Hatchery temperature is less than or equal to 70ºF. Data supports the claim that Chattahoochee River temperature at the Hatchery supports a healthy environment for rainbow trout.

The utility of the Wilcoxon Signed Rank Test also extends to test for likely confidence intervals pertaining to the population median. We would like to know the confidence interval for the population median of each data set to see how close the interval comes to the critical value the GDNR considers unsafe for rainbow trout habitat.

<u>Wilcoxon Signed Rank Confidence Interval for likely population median based on sampling data</u>*:*

What is a confidence interval for the population median of dissolved oxygen concentration at the Buford Dam?

Wilcoxon Signed Rank CI: Buford Dam DO

	N	Estimated Median	Achieved Confidence	Confidence Interval Lower	Confidence Interval Upper
Buford Dam DO	26884	6.615	95.0	6.580	6.655

A 95% confidence interval for the Buford Dam dissolved oxygen concentration median is between 6.580 mg/L and 6.655 mg/L. The confidence interval result supports the hypothesis test conclusion that our population median of dissolved oxygen concentrations at the Buford Dam is greater than or equal to 6 mg/L.

What is a confidence interval for the population median of dissolved oxygen concentration at the Hatchery?

Wilcoxon Signed Rank CI: Hatchery DO

	N	Estimated Median	Achieved Confidence	Confidence Interval Lower	Confidence Interval Upper
Hatchery DO	26884	8.475	95.0	8.460	8.495

A 95% confidence interval for the Hatchery dissolved oxygen concentration median is between 8.460 mg/L and 8.495 mg/L. The confidence interval result supports the hypothesis test conclusion that our population median of dissolved oxygen concentrations at the Hatchery is greater than or equal to 6 mg/L.

What is a confidence interval for the population median of temperature at the Buford Dam?

Wilcoxon Signed Rank CI: Buford Dam Temperature

	N	Estimated Median	Achieved Confidence	Confidence Interval Lower	Confidence Interval Upper
Buford Dam Temperature	26884	49.33	95.0	49.30	49.37

A 95% confidence interval for the Buford Dam Temperature median is between 49.30ºF and 49.37ºF. The confidence interval result supports the hypothesis test conclusion that our population median of temperature at the Buford Dam is less than or equal to 70ºF.

What is a confidence interval for the population median of temperature at the Hatchery?

Wilcoxon Signed Rank CI: Hatchery Temperature

	N	Estimated Median	Achieved Confidence	Confidence Interval Lower	Confidence Interval Upper
Hatchery Temperature	26884	49.64	95.0	49.61	49.70

A 95% confidence interval for the Hatchery Temperature median is between 49.61ºF and 49.70ºF. The confidence interval result supports the hypothesis test conclusion that our population median of temperature at the Hatchery is less than or equal to 70ºF.

Wilcoxon Signed Rank Test to compare population median of two groups:

Are the dissolved oxygen concentration medians between the Buford Dam and the Hatchery equal?

H_o: The median of Buford Dam dissolved oxygen concentration are equal to the median of Hatchery dissolved oxygen concentration. [$\Theta_{DAM} = \Theta_{HATCHERY\,=}\,0$]

H_a: The median of Buford Dam dissolved oxygen concentration are not equal to the median of Hatchery dissolved oxygen concentration. [$\Theta_{DAM} - \Theta_{HATCHERY\,\neq}\,0$]

Wilcoxon Signed Rank Test: DO_Difference

```
Test of median = 0.000000 versus median ≠ 0.000000

                      N for   Wilcoxon                    Estimated
                N     Test    Statistic        P          Median
DO_Difference   26884  26807  24063613.0       0.000      -1.715
```

Conclusion: Reject H_o. There is sufficient evidence at the alpha level of 0.05 significance to reject the claim that the median for Buford Dam dissolved oxygen concentration is equal to the median for hatchery dissolved oxygen concentration.

Is the Buford Dam dissolved oxygen concentration median greater than or equal to the Hatchery dissolved oxygen concentration median?

H_o: The Buford Dam has a dissolved oxygen concentration median that is greater than or equal to the Hatchery dissolved oxygen concentration median. [$\Theta_{DAM} - \Theta_{HATCHERY} \geq 0$]

H_a: The Buford Dam has a dissolved oxygen concentration median that is less than the Hatchery dissolved oxygen concentration median. [$\Theta_{DAM} - \Theta_{HATCHERY} < 0$]

Wilcoxon Signed Rank Test: DO_Paired

Test of median ≥ 0.000000 versus median < 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
DO_Paired	26884	26807	24063613.0	0.000	-1.715

Conclusion: Reject H_o. There is sufficient evidence at the alpha level of 0.05 significance to reject the claim that the Buford Dam has a dissolved oxygen concentration median that is greater than or equal to the Hatchery dissolved oxygen concentration median. This result shows a likely increase in dissolved oxygen concentrations as the Chattahoochee River water flows from the Buford Dam to the Hatchery.

Are the temperature medians between the Buford Dam and the Hatchery equal?

H_o: The median for Buford Dam temperature is equal to the median for Hatchery temperature. $[\Theta_{DAM} = \Theta_{HATCHERY} = 0]$

H_a: The median for Buford Dam temperature is not equal to the median for Hatchery temperature. $[\Theta_{DAM} \neq \Theta_{HATCHERY} \neq 0]$

Wilcoxon Signed Rank Test: Temp_Difference

Test of median = 0.000000 versus median ≠ 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Temp_Difference	26884	26395	150403089.0	0.000	-0.1450

Conclusion: Reject H_o. There is sufficient evidence at the alpha level of 0.05 significance to reject the claim that the median for Buford Dam temperature is equal to the median for Hatchery temperature.

Is the Buford Dam temperature median greater than or equal to the Hatchery temperature median?

H_o: The Buford Dam has a temperature median that is greater than or equal to the Hatchery temperature median. $[\Theta_{DAM} - \Theta_{HATCHERY} \geq 0]$

H_a: The Buford Dam has a temperature median that is less than the Hatchery temperature median. $[\Theta_{DAM} - \Theta_{HATCHERY} < 0]$

Wilcoxon Signed Rank Test: Temp_Paired

```
Test of median ≥ 0.000000 versus median < 0.000000

                    N for  Wilcoxon             Estimated
             N      Test   Statistic     P      Median
Temp_Paired  26884  26395  150403089.0  0.000   -0.1450
```

Conclusion: Reject H_o. There is sufficient evidence at the alpha level of 0.05 significance to reject the claim that the Buford Dam has a temperature median that is greater than or equal to the Hatchery temperature median. This result shows a likely increase in temperature as the Chattahoochee River water flows from the Buford Dam to the Hatchery.

Discussion

That data supports the hypothesis; that dissolved oxygen concentration and temperatures likely fall within what is considered by the GDNR as an ideal habitat for a rainbow trout.

Early analysis of our data requires that we limit the variety of statistical testing possibilities at our disposal to only the strongest and most appropriate ones. To accomplish this, we first need to determine whether our data belongs to a particular distribution and whether the data is parametric or non-parametric. It was an important first step for me to begin by examining the distribution's shape and symmetry so I could create a long-term plan on how to approach our four data sets. I did this by examining the box plot Minitab output included at the beginning of the results section. A visual comparison between the Buford Dam and Hatchery temperatures illustrates a similar shape and distribution. Comparing the dissolved oxygen boxplots shows them to have pronounced differences. Especially noticeable are the differences in the interquartile range and the overall range between data sets. Many statistical tests have assumptions about them that must be satisfied. An asymmetric data set might eliminate some tests so we will need verify symmetry—or lack thereof—by examining skewness and kurtosis.

Normality and Parametric Testing

The first appropriate measure would be to apply a normality test. Minitab has three different types of normality tests; the Anderson-Darling, the Ryan-Joiner and the Kolmogorov-Smirnov test. Each test has justification for its use and I will explain why using Anderson-Darling (A-D) test for this analysis makes the most sense.

Of the three choices available, Anderson-Darling is the test to use when there is a need emphasize the weight in the tails of a distribution—the tails being the most important portion of the distribution for testing. Kolmogorov-Smirnov weighs the distribution differently, placing more emphasis on the center of the distribution. Ryan-Joiner is useful in cases where you would be interested in working with correlation between sample data and data that would be expected from a normal distribution. The strength of Anderson-Darling is particularly evident as the sample size tested increases. Anderson-Darling is also extremely sensitive to small changes within the tails of the distribution. Given the strength in larger sample sizes and sensitivity found in Anderson-Darling, plus the Kolmogorov-Smirnov and Ryan-Joiner's lack of robustness and strength, make the choice of using Anderson-Darling a cautious and conservative one.

All four data sets were treated to the Anderson-Darling normality test in Minitab and all four data sets failed hypothesis testing of belonging to any particular distribution. We can move forward with confidence that the data is likely to be distribution-free.

In continuing the need to put limitations on the number of testing choices, the data collected needs to be tested to determine whether it is parametric or not. Defining parameters, or the absence of them, will help to partition the data set into one of two groups—each group having specific tests.

The skew of a distribution is a shape statistic of that distribution, which measures how far the departure from horizontal symmetry is. Skew explains the best measure of central tendency and central tendency is the metric that determines whether our data is parametric. Measuring skew will produce an output in the form of a real number that will describe the direction of the skew and how severe the skew is when compared to that of a normal and symmetric distribution. The closer to zero our skew value is, the closer to symmetry the distribution is. From the Minitab output we can see that the data is approximately symmetric but still contains skew. The mean would not be an appropriate measure of central tendency. Since mode is used exclusively on dichotomous nominal data only, the median will be the best measure for central tendency.

The preliminary results indicate that after performing an Anderson-Darling normality test and examining the skewness of each of the data sets, we can confidently conclude that all four samples are distribution-free non-parametric samples.

Non-Parametric Testing

Now that the parametric and normality testing is completed, the next logical step is to test the likely confidence interval for which a population median falls within. The same test can also compare the sample median against a particular value. The proximity of the likely population median and its confidence interval in relation to the values the GDNR considers unsafe is important in determining the placement of the rest of our data in relation to that value. A median closer to the critical value can be interpreted as more of the distribution being closer to that critical value as well. The converse is likely true as well when the median is further away.

There are two choices available to perform non-parametric interval testing on the median; the 1-Sample Sign Test and the 1-Sample Wilcoxon Signed Rank Test. The 1-Sample Sign Test is the weaker of the two tests so ideally we would like to use the stronger of the two tests available. I need to make sure my data fulfills assumptions regarding my data. .

For the 1-Sampled Sign Test, there are four assumptions:

- Any dependent variables should be continuous and non-nominal.
- Independent variables should be comprised of two categorical “related groups” or “matched pairs”.
- Paired observations for each participant needs to be independent.

- The differences between the paired observations are from a continuous distribution.

For the 1-Sample Wilcoxon Signed Ranks Test,

- Data is comprised of dependent variables that are either ordinal, interval, or ratio.
- Independent variables are two categorical groups that are matched pairs.
- The distribution of the differences between the two groups are symmetrical in shape.

The Chattahoochee River water data is composed entirely of dependent variables. River water temperature and dissolved oxygen concentration levels will depend on other independent variables such as Lake Lanier water temperature and dissolved oxygen concentration profiles. The data is also continuous and non-nominal; with temperature and dissolved oxygen being an interval and ratio data type respectively. The four samples meet all but one of the assumptions for both tests—that will be examined. The last assumption requires that both samples are symmetric to each other. Paired symmetry is decided by examining two “shape statistics”; skew and kurtosis. Similarity between their values explains why we will use the 1-Sample Wilcoxon Signed Ranks Test over the weaker 1-Sample Sign Test.

The first statistic we want to look at is the skew. We have already examined—and explored the process of calculating—the skew of our samples earlier while determining central tendency. At this point, we want to compare the two samples to each other instead of a normal distribution. This is different to comparing skew to zero. The magnitude of the skew between both variables is most interesting in that we can determine approximate symmetry by examining the magnitude of the difference between skew values of each variable. The ‘Buford DO’ dataset has a skewness of -0.29944 and the 'Hatchery DO' dataset has a skewness of -0.23071. The 'Buford Temp' dataset has a skewness of -0.168612 and the 'Hatchery Temp' has a skewness of -0.003975. According to Bulmer, our skew values are close to pairwise symmetry that we can proceed as though they are symmetric (1979).

Kurtosis is another shape statistic that is relevant to symmetry. This statistic measures the distributions "tailed-ness" and gives insight into where the data is located within the distribution. Observe that the comparison of the paired-data kurtosis values is approximately the same as well. The 'Buford DO' data kurtosis is -1.67594 and 'Hatchery DO' is -1.6869; 'Buford Temp' kurtosis is -0.986087 and 'Hatchery Temp' is -0.903068. We can establish a pattern of symmetry in the kurtosis for each pair of data. With skew and kurtosis having like-paired values, we can use the stronger of the two tests for our analysis; the 1-sample Wilcoxon signed ranks test.

Wilcoxon Signed Rank Test for the population median

Research conducted by John Biagi and Richard Brown of the Georgia Department of Natural Resources determined that rainbow trout waters are recommended to have a dissolved oxygen concentration greater than or equal to 6 mg/L for sustainability (1997). The same research also found that the recommended water temperature for rainbow trout should be less than or equal to 70ºF. I used the 1-Sample Wilcoxon ranked signed test to determine if our population median for dissolved oxygen concentrations and temperatures were at values unideal for a rainbow trout habitat.

From the results of dissolved oxygen concentration data, we can say with a significance level of 95% that there is insufficient evidence to refute the claim that the population median for both locations of dissolved oxygen concentrations are equal to or greater than 6mg/L. This is good news for the rainbow trout's quality-of-life as we would not want the median of the distribution to be found below 6 mg/L.

Assessing the water temperature output, we are 95% confident that there is insufficient evidence to refute the claim that the population median temperature at the Buford Dam and the Hatchery locations are equal to or below 70ºF. We would not want the median of our temperature distributions to exceed 70ºF.

Wilcoxon Signed Rank Test for population median confidence interval

The Wilcoxon Signed Rank test also has the capability to provide a confidence interval for the values the population median is likely to fall within. Knowing this information is important in that it will give us an idea of where our population medians confidence intervals are in proximity to our critical values. The closer the confidence interval approaches the critical value, the more probable parts of the distribution are to be equal to or beyond the acceptable habitable limits.

				Confidence Interval	
	N	Estimated Median	Achieved Confidence	Lower	Upper
Buford Dam DO	26899	6.620	95.0	6.580	6.660
Hatchery DO	26884	8.475	95.0	8.460	8.495
Buford Temp	26899	49.33	95.0	49.30	49.37
Hatchery Temp	26884	49.64	95.0	49.61	49.70

From the Minitab output, we can see that the median population confidence intervals are within comfortable margins. An interesting observation is how close the confidence interval the dissolved oxygen concentrations near the Buford Dam is. The GDNR does not want to see the dissolved oxygen concentration levels to be less than 6 mg/L. The confidence interval at the Buford Dam is within 0.58-0.66 mg/L from that value.

To explain why this confidence interval is so much closer to the critical GDNR value than the other data sets, we turn to Pat Snellings, Biologist with the Georgia Department of Natural Resources. Lake Lanier is a thermally stratified warm water monomictic lake. During the hot months of summer, water temperatures and densities are vertically separated and layered within the lake. The top layer—called the epilimnion layer—extends approximately 15 meters below the lake surface. It contains warm low-density water (approximately 70-89ºF) and high dissolved oxygen concentrations (6-12 mg/L). The next layer—called the thermocline—is a thin layer of stratification that is marked by rapidly declining water temperature and dissolved oxygen concentration. This layer is about 5 meters in depth and separates the epilimnion from the bottom layer called the hypolimnion. The hypolimnion layer of the lake contains water that is absent of wildlife, cold (approximately 46-49ºF), dense and low in dissolved oxygen concentration (0.6-1.0 mg/L).

Timothy Rainey, Army Corps of Engineers Project Manager says the lake water that is drawn into the dam is done so at a depth ranging between 39.92-46.33 meters. That would place the water intake of the dam well within the depths of the hypolimnion layer of Lake Lanier. Timothy Rainey also mentions that the lake water that is drawn into the dam mechanically aerates itself by passing through the hydro-electrical components of the Buford Dam. Once the lake water enters the origins of the upper Chattahoochee River, it continues aeration as the river water breaks and tumbles over rocks and other natural features within the river. This is congruent with the data that I have showing lower dissolved oxygen concentrations just below the Buford dam and an increased margin between the population median confidence interval and critical value two miles downstream at the hatchery.

The population median temperature confidence interval at both locations is significantly below the 70 degree Fahrenheit mark—by more than 20 degrees. Rainbow trout are likely to enjoy temperatures well within their comfort level.

Wilcoxon Signed Rank Test for the difference between medians

The Wilcoxon Signed Rank Test also can test to see if there a departure between the median temperatures and dissolved oxygen concentrations of both locations exists. To do this requires a little more work. First, we need to take the difference between the two groups and put them in an entirely new column. In Minitab, I used the calculator function, created a formula taking the difference between Buford Dam DO and Hatchery DO, and created a new column to accept the results. After the differences were recorded, I ran a Wilcoxon Signed Rank Test on the newly columned data and asked the question, "are the median temperatures and dissolved oxygen concentrations different between the Buford Dam and the Hatchery locations?"

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
DO_Differ.	26884	26807	24063613.0	0.000	-1.715
Temp_Differ.	26884	26395	150403089.0	0.000	-0.1450

The conclusion we can draw from both Wilcoxon test results is that the median between temperature and dissolved oxygen concentrations are not equal to each other, and likely change on the 2-mile span of the Chattahoochee River. There is an inequality in temperature and dissolved oxygen concentration between both observation points; however, the change is not significant with respect to living conditions for rainbow trout. Both median temperature and dissolved oxygen concentrations and their confidence interval are within acceptable parameters. What we have learned from this test is that the medians likely differ. The median temperature is likely to have increased by 0.145ºF from Buford Dam to Hatchery. The median dissolved oxygen concentrations is also likely to increase 1.715 mg/L in that same 2-mile downstream stretch of river.

Conclusion

From inferential statistics, we can note with a high confidence level that the Chattahoochee River, between the Buford Dam and the Trout Hatchery, can safely harbor rainbow trout during the winter months. The temperature at both stations are likely to remain comfortably below the 70ºF benchmark. Rainbow will enjoy nearly uniform temperatures throughout this stretch of river water.

Far more interesting is the measurements taken near the Buford Dam that shows a depressed—but still acceptable—level of dissolved oxygen concentrations. The dissolved oxygen levels nearer the dam are explained by the hypolimnion layered water from Lake Lanier entering and passing through the Buford Dam and into the Chattahoochee River. The water mechanically aerates while passing through the dam and reaches the river at just above the acceptable limit for Rainbow Trout. The aeration continues while traveling downstream with the disruption of surface water and the breaking of water around natural features found around the river. Downstream, the Chattahoochee River waters are likely to experience a 1.7 mg/L increase in dissolved oxygen concentrations along the first 2-miles. This would bring this dissolved oxygen concentrations well within a comfortable margin we would like to see in a thriving habitat for rainbow trout.

The lower dissolved oxygen concentrations near the dam are not nearly as alarming when considering the length of habitable river water available to Georgia's trout. The rainbow trout habitat is approximately 33 miles in length; extending from Buford Dam down to Morgan falls reservoir. The tributaries that also supply the Chattahoochee River with water have tendrils that extend into highly commercialized and population dense areas. This might have the effect of raising water temperatures in the Chattahoochee River. Considering the length of habitable river rainbow trout have at their disposal and the safe level of dissolved oxygen concentration and temperature readings of the river water, it is recommended that dissolved oxygen concentrations and temperature readings be collected further downstream—from the hatchery down to Morgan Falls Reservoir.

That is but one possible future research endeavor. Another possibility is regarding the Buford Dam. The dam has two methods for releasing water; through the hydropower units and an alternative sluice channel, that bypasses the hydropower components. Measurements and statistical analysis concerning dissolved oxygen concentration readings in the river after a hydropower water release versus a sluice bypass release would be interesting to those with vested interests in marine wildlife just below the dam. Consider also the temperature increase, if any, caused by water passing through the Buford Dam versus the sluice.

Appendix

<u>Mathematical Induction Proof Summation Component of the Wilcoxon Signed Rank Test Mean</u>:

Proof: We will prove by induction that, for all $n \in \mathbb{Z}_+$,

(1) $\sum_{i=1}^{n}(i) = \frac{n(n+1)}{2}$

Verify: When $n = 1$, the left side of (1) is = 1, and the right side is (1(1+1))/2 = 2/2 = 1, so both sides are equal and (1) is true for $n = 1$.

Assume: Let $k \in \mathbb{Z}_+$ be given and suppose (1) is true for $n = k$ such that

$$1 + 2 + \dots + (k-1) + k = \frac{k(k+1)}{2}$$

Prove: $P(k+1)$ is true for all $k \in \mathbb{Z}_+$

$$\sum_{i=1}^{k+1} i = \frac{((k+1)(k+1+1))}{2}$$

$$1 + 2 + \dots + (k-1) + k + (k+1) = \frac{(k+1)(k+2)}{2}$$

$$\frac{k(k+1)}{2} + (k+1) = \text{ (by induction hypothesis)}$$

$$\frac{k(k+1)}{2} + \frac{2(k+1)}{2} =$$

$$\frac{k(k+1) + 2(k+1)}{2} =$$

$$\frac{(k+1)(k+2)}{2} = RHS$$

Thus, (1) holds for $n = k + 1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, (1) is true for all $n \in \mathbb{Z}_+$.

Q.E.D.

Mathematical Induction Proof Summation Component of the Wilcoxon Signed Rank Test Variance:

Proof: We will prove by induction that, for all $n \in \mathbb{Z}_+$,

(2) $\sum_{i=1}^{n}(i^2) = \frac{n(n+1)(2n+1)}{6}$

Verify: When $n = 1$, the left side of (2) is = 1, and the right side is (1(1+1)(2(1)+1))/6 = 6/6 = 1, so both sides are equal and (2) is true for $n = 1$.

Assume: Let $k \in \mathbb{Z}_+$ be given and suppose (2) is true for $n = k$ such that

$$1^2 + 2^2 + \dots + (k-1)^2 + k^2 = \frac{k(k+1)(2k+1)}{6}$$

Prove: $P(k+1)$ is true for all $k \in \mathbb{Z}_+$

$$\sum_{i=1}^{k+1} i^2 = \frac{(k+1)(k+1+1)(2(k+1)+1)}{6}$$

$$1^2 + 2^2 + \dots + (k-1)^2 + k^2 + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6}$$

$$\frac{k(k+1)(2k+1)}{6} + (k+1)^2 = \text{ (by induction hypothesis)}$$

$$\frac{k(k+1)(2k+1)}{6} + \frac{6(k+1)^2}{6} =$$

$$\frac{k(k+1)(2k+1) + 6(k+1)^2}{6} =$$

$$\frac{(k+1)(k(2k+1) + 6(k+1))}{6} =$$

$$\frac{(k+1)(2k^2 + k + 6k + 1)}{6} =$$

$$\frac{(k+1)(2k^2 + 7k + 1)}{6} =$$

$$\frac{(k+1)(k+2)(2k+3)}{6} = RHS$$

Thus, (2) holds for $n = k + 1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, (2) is true for all $n \in \mathbb{Z}_+$.

$Q.E.D.$

Works Cited

Adam Lund, Mark Lund. (2013). *Laerd Statistics*. Retrieved from Laerd Statistics: https://statistics.laerd.com/spss-tutorials/wilcoxon-signed-rank-test-using-spss-statistics.php

Brown, J. B. (1997, September). *Environmental Protection Division*. Retrieved from Environmental Protection Division: https://epd.georgia.gov/sites/epd.georgia.gov/files/related_files/site_page/WQS_Biagi_Temperature_Tolerance_Trout_1997.pdf

Bulmer, M. (1979). *Principals of Statistics.* New York: Dover Publications Inc.

Fondriest Environmental, Inc. (2013, November 19). *Dissolved Oxygen, Fundamentals of Environmental Measurements*. Retrieved from http://www.fondriest.com/environmental-measurements/parameters/water-quality/dissolved-oxygen/

Graham K. Taylor, R. L. (2010, February 26). *The Royal Society Publishing*. Retrieved from The Royal Society Publishing: http://rsif.royalsocietypublishing.org/content/early/2010/02/17/rsif.2009.0522.full

McNeese, D. B. (2011, June). *SPC For Excel - BPI Consulting, LLC*. Retrieved from SPC For Excel - BPI Consulting, LLC: https://www.spcforexcel.com/knowledge/basic-statistics/anderson-darling-test-for-normality

Scholz, S. (1987, September 1). K-Sample Anderson-Darling Tests. *Journal of the American Statistical Association*, pp. Volume 82, Number 399, 918-924. Retrieved from http://www.hep.caltech.edu/~fcp/statistics/hypothesisTest/PoissonConsistency/ScholzStephens1987.pdf

Shmukler, C. (2012, July 5). *Hatch Magazine*. Retrieved from Hatch Magazine: http://www.hatchmag.com/articles/trout-and-water-temperature-how-hot-too-hot/771553