

**Statistical Procedures for Testing  
Homogeneity of Water Quality Parameters**

Xu-Feng Niu  
Professor of Statistics

*Department of Statistics  
Florida State University  
Tallahassee, FL 32306*

May-September 2004

## 1. Nonparametric tests for multiple populations

Median (mean) and variance are the two most important numerical characteristics of a probability population. For a random variable with a symmetric distribution, such as the normal or student- $t$  distribution, the mean and median have the same value. Observations of a water quality parameter from  $k$  different stations (or  $k$  different waterbodies) are samples from  $k$  populations. If either the medians (means) or the variances of the  $k$  populations are different, the populations are not homogenous.

Denote the observations of a water quality parameter collected at  $k$  locations by  $\{X_{1j}, \dots, X_{n_jj}; j = 1, \dots, k\}$ . In this study, we need the following assumptions on the data:

- For a given location  $j$ ,  $\{X_{1j}, \dots, X_{n_jj}\}$  are a random sample from a continuous distribution with distribution function  $F_j$ .
- The  $k$  groups of observations are mutually independent (in other words, observations from the  $k$  different locations are independent).

### a. The Kruskal-Wallis Test for assessing homogeneity of medians.

Consider the following linear model for the observations of a water quality parameter collected from  $k$  locations:

$$X_{ij} = \mu + \alpha_j + e_{ij}, \quad j = 1, \dots, k; \quad i = 1, \dots, n_j, \quad (1)$$

where  $\mu_j = \mu + \alpha_j$  is the median of population  $j$ , the  $e_{ij}$ 's are random errors from a continuous distribution with median 0.

The null hypothesis of interest in this study is that of no differences among the medians  $\{\mu_1, \dots, \mu_k\}$ . In other words, the null hypothesis we are testing is

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k \equiv 0 \quad (\text{or } \mu_1 = \mu_2 = \dots = \mu_k \equiv \mu), \quad (2)$$

Let the total number of observations at the  $k$  locations be  $N = n_1 + n_2 + \dots + n_k$ . To calculate the Kruskal-Wallis statistic (Hollander and Wolfe, 1999, pp190-192), we first combine all  $N$  observations  $\{X_{1j}, \dots, X_{n_jj}; j = 1, \dots, k\}$  from the  $k$  populations from least to greatest. Let  $r_{ij}$  be the rank of  $X_{ij}$  in the joint ranking and let

$$R_j = \sum_{i=1}^{n_j} r_{ij} \quad \text{and} \quad \bar{R}_j = R_j / n_j, \quad j = 1, \dots, k. \quad (3)$$

Then the Kruskal-Wallis statistic  $H$  is given by

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k n_j (\bar{R}_j - (N+1)/2)^2 , \quad (4)$$

where  $\frac{N+1}{2} = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{n_j} r_{ij}$  is the average rank assigned in the joint ranking.

To test the null hypothesis  $H_0$  defined in (2) versus the alternative hypothesis

$$H_1 : \text{ not all medians equal } , \quad (5)$$

at the  $\alpha$  level of significance, we reject  $H_0$  if  $H \geq h_\alpha$  where  $h_\alpha$  is the critical value given in Table A.12 of Hollander and Wolfe (1999).

For large sample sizes (all  $n_j \geq 10$ ), the Kruskal-Wallis statistic  $H$  approximately has a chi-square distribution with  $(k-1)$  degrees of freedom. Then we can reject the null hypothesis if  $H \geq \chi_{k-1, \alpha}^2$ . **In the R language, the command for the Kruskal-Wallis test is `kruskal.test(x, g)` where  $x$  is a numerical data vector of the observations of the water quality parameter and  $g$  is a vector or factor object giving the group for the corresponding elements of  $x$ .**

## b. The Fligner-Killeen test for assessing homogeneity of variances

For testing homogeneity of variances of  $k$  population, Fligner and Killeen (1976) suggested jointly ranking the absolute values  $|X_{ij}|$  and assigning increasing scores such as  $a_{N,i} = i$ ,  $a_{N,i} = i^2$ , or  $a_{N,i} = \Phi^{-1}((1+i/(N+1))/2)$  based on the ranks. Conover, Johnson, and Johnson (1981) modified the Fligner-Killeen test and suggested ranking  $|X_{ij} - \tilde{X}_j|$  instead of  $|X_{ij}|$ , where  $\tilde{X}_j$  is the sample median of the  $j$ th population. This version of the Fligner-Killeen test is called the median-centering Fligner-Killeen test.

After the scores  $a_{N,i}$ 's are assigned to  $|X_{ij} - \tilde{X}_j|$ , the modified Fligner-Killeen test can be calculated by

$$X^2 = \sum_{j=1}^k n_j (\bar{A}_j - \bar{a})^2 / V^2 , \quad (6)$$

where  $\bar{A}_j$  is the mean score for the  $j$ th sample,  $\bar{a}$  is the overall mean score, i.e.,

$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_{N,i}$ , and  $V^2 = \frac{1}{N-1} \sum_{i=1}^N (a_{N,i} - \bar{a})^2$ . For large sample sizes, the Fligner-Killeen statistic  $X^2$  has an asymptotic chi-square distribution with  $(k-1)$  degrees of freedom.

Conover, Johnson, and Johnson (1981) conducted a simulation study to compare tests for homogeneity of variances and found that the modified Fligner-Killeen test is most robust against departures from normality. **The version of the Fligner-Killeen test implemented in the R language uses median centering in each of the samples. The command for this test is `fligner.test(x, g)` where  $x$  is a numerical data vector of the observations of the water quality parameter and  $g$  is a vector or factor object giving the group for the corresponding elements of  $x$ .**

## 2. Nonparametric tests for two populations

Nonparametric test methods for homogeneity of two populations are well developed in statistical literature. In this study, three most popular tests are presented: the Wilcoxon Rank Sum Test for location shift, the Ansari-Bradley test for dispersion, and the Kolmogorov-Smirnov test for general differences in two populations.

### a. The Wilcoxon Rank Sum test for location shift

The Wilcoxon Rank Sum test is also called the Wilcoxon-Mann-Whitney Rank Sum test (Hollander and Wolfe, 1999, pp106-125), which is used widely in location-shift (mean or median) tests.

Consider observations of a water quality parameter collected from two locations (populations A and B) in a watershed. Denote the observations by  $\{X_{1j}, \dots, X_{n_j}; j = 1, 2\}$ . The assumptions on the samples, specified in Section 2, are assumed to be valid in this section.

The null hypothesis tested by the Wilcoxon Rank Sum test is

$$H_0: \text{The two sampled populations have identical probability distributions.} \quad (7)$$

The one-tailed alternative hypothesis is

$$H_1: \text{The probability distribution for population A is shifted to the right of that for B} \quad (8)$$

and the two-tailed alternative hypothesis is

$$H_1: \text{The probability distribution for population A is shifted to the left or to the right of that for B.} \quad (9)$$

Let the total number of observations from the two locations be  $N = n_1 + n_2$ . To calculate the Wilcoxon Rank Sum statistic (Hollander and Wolfe, 1999, pp106-125), we first combine all  $N$  observations  $\{X_{1j}, \dots, X_{n_j}; j = 1, 2\}$  of the two populations from least to greatest. Let  $r_{ij}$  be the rank of  $X_{ij}$  in the joint ranking and let

$$R_j = \sum_{i=1}^{n_j} r_{ij}, \quad j = 1, 2. \quad (10)$$

To test the null hypothesis against the two-tailed alternative hypothesis at the  $\alpha$  level of significance, reject the null hypothesis if  $R_2 \geq w_{\alpha/2}$  or if  $R_2 \leq n_2(n_1 + n_2 + 1) - w_{\alpha/2}$  where  $w_{\alpha/2}$  can be found in Table A.6 of Hollander and Wolfe (1999).

For large samples ( $n_1 \geq 10$  and  $n_2 \geq 10$ ), the Wilcoxon Rank Sum statistic  $W=R_2$  can be approximated by a normal distribution. Specifically, let

$$W^* = \frac{R_2 - [n_2(n_1 + n_2 + 1)/2]}{[n_1 n_2 (n_1 + n_2 + 1)/12]^{1/2}}. \quad (11)$$

Then the statistic  $W^*$  has a standard normal distribution approximately. To test the null hypothesis against the two-tailed alternative hypothesis at the  $\alpha$  level of significance, reject the null hypothesis if  $|W^*| \geq z_{\alpha/2}$  where the critical value  $z_{\alpha/2}$  can be found in a standard normal distribution table.

**In the R language, the command for the Wilcoxon Rank Sum test is `wilcox.test(x, y)` where  $x$  is a numerical data vector of the observations of the water quality parameter from the first location and  $y$  is a numerical data vector of the observations of the water quality parameter from the second location.**

## b. The Ansari-Bradley test for dispersion.

Consider the two populations considered in part (a). Suppose that samples from the two populations passed the Wilcoxon Rank Sum test, i.e., there is no location shift in the two populations (or in other words, the means and medians of the two populations are identical). Let the variances of the two populations be  $\sigma_1^2$  and  $\sigma_2^2$ , respectively. We would like to test the null hypothesis of

$$H_0 : \gamma^2 = \sigma_1^2 / \sigma_2^2 = 1. \quad (12)$$

The Ansari-Bradley test was designed to test this null hypothesis (Hollander and Wolfe, 1999, 142-158). To calculate the Ansari-Bradley statistic, we first order the combined sample from least to greatest. Assign the score 1 to both the smallest and the largest observations in the combined sample, assign score 2 to the second smallest and the second largest, and continue in this manner. If  $N = n_1 + n_2$  is an even number, the array of assigned scores is

$$\{1, 2, 3, \dots, N/2, N/2, \dots, \dots, 3, 2, 1\}.$$

If  $N$  is an odd number, the array of assigned scores is

$$\{1, 2, 3, \dots, (N-1)/2, (N+1)/2, (N-1)/2, \dots, \dots, 3, 2, 1\}.$$

Let  $\{S_j, j = 1, \dots, n_2\}$  be the scores assigned to the observations in the second population and let

$$C = \sum_{j=1}^{n_2} S_j . \quad (13)$$

To test the null hypothesis against the two-sided alternative hypothesis  $H_1 : \gamma^2 \neq 1$  at the  $\alpha$  level of significance, we reject the null hypothesis if  $C \geq c_\alpha$  or  $C \leq (c_{1-\alpha} - 1)$  where the critical values  $c_\alpha$  and  $c_{1-\alpha}$  can be found in Table A.8 of Hollander and Wolfe (1999).

For large samples ( $n_1 \geq 10$  and  $n_2 \geq 10$ ), the Ansari-Bradley  $W=R_2$  can be approximated by a normal distribution with mean

$$E_0(C) = \frac{n_2(N + 2)}{4} \quad (14)$$

and variance

$$Var_0(C) = \frac{n_1 n_2 (N + 2)(N - 2)}{4N} . \quad (15)$$

Then the standardized statistic

$$C^* = \frac{C - E_0(C)}{[Var_0(C)]^{1/2}} \quad (16)$$

has a standard normal distribution. To test the null hypothesis against the two-tailed alternative hypothesis at the  $\alpha$  level of significance, reject the null hypothesis if  $|C^*| \geq z_{\alpha/2}$  where the critical value  $z_{\alpha/2}$  can be found in a standard normal distribution table.

**In the R language, the command for the Ansari-Bradley test is `ansari.test(x, y)` where  $x$  is a numerical data vector of the observations of the water quality parameter from the first location and  $y$  is a numerical data vector of the observations of the water quality parameter from the second location.**

### c. The Kolmogorov-Smirnov test for general differences.

Consider a water quality parameter observed at two locations. Assume observations from the two locations are from continuous probability distributions  $F$  and  $G$ , respectively. The Kolmogorov-Smirnov test (Hollander and Wolfe, 1999, pp178-185) can be used to test whether there are any differences whatsoever between the two probability distributions. Specifically, we may test the null hypothesis

$$H_0 : \text{The two sampled populations have identical probability distributions } (F(t) \equiv G(t)) \quad (17)$$

versus the alternative hypothesis

$$H_1 : \text{The two sampled populations have different probability distributions} \\ (F(t) \neq G(t), \text{ for some } t). \quad (18)$$

Details of the test can be found in Hollander and Wolfe (1999, pp178-185).

**In the R language, the command for the Kolmogorov-Smirnov test is `ks.test(x, y)` where  $x$  is a numerical data vector of the observations of the water quality parameter from the first location and  $y$  is a numerical data vector of the observations of the water quality parameter from the second location.**

#### 4. A step-up procedure Testing procedures for testing homogeneity of samples.

The step-up procedure starts at any one of the  $k$  samples. The chosen sample is compared with each of the remaining  $k-1$  samples using the nonparametric tests for two populations given in Section 3. The procedure consists of the following five steps.

- Step 1. List the  $k$  samples in order based on region or median values or some other criteria. In this study, we simply list the samples from 1 to  $k$ .
- Step 2. Start from sample 1. Compare this sample with each of the remaining  $k-1$  samples using the nonparametric tests listed in Section 3. For example, if samples 1 and 2 pass the three tests, these two samples will be combined. The combined sample is then compared to each of the remaining  $k-2$  samples. On the other hand, if samples 1 and 2 cannot pass one of the three tests, sample 1 is then compared to sample 3. Continue this process until the last sample is compared.
- Step 3. Perform the same procedure in Step 2 for sample 2. Compare sample 2 with samples 1 to  $k$  (not sample 2). Continue this procedure for the remaining samples until all sample pairs are compared.
- Step 4. Based on the comparison results in Steps 2 and 3, list all sub groups that can be combined.
- Step 5. Perform homogeneity tests on the sub groups to decide whether further merging of the samples is possible or not.

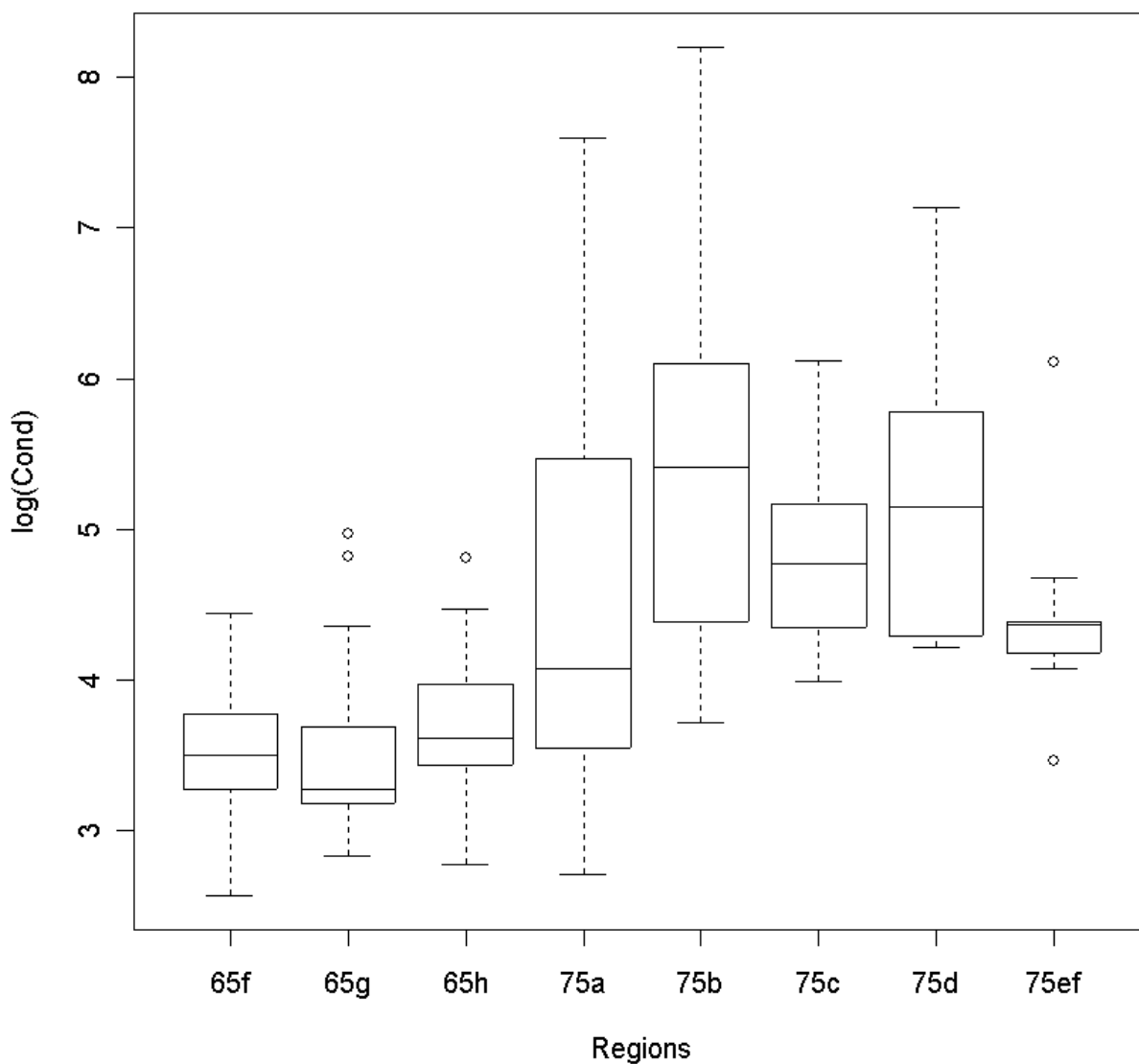
## 5. Illustration examples

In this section, we utilize the step-up procedure to test the homogeneity of water quality variables in eight ecoregions. All nonparametric tests proposed in this study are available in the R language that is a **free** statistical computing package and can be downloaded from

<http://cran.us.r-project.org/>

### 1). Example 1 (Specific Conductivity Field Medians in the Eight Ecoregions; after the natural logarithm transformation)

Figure 1. Specific Conductivity Field Medians in the Eight Ecoregions





**Table 1. Initial merging groups based on the step-up procedure and the specific conductivity field medians in the eight ecoregions ( $\alpha = 0.05$  is used).**

Start Ecoregion	Merging groups
<b>65f (n=11)</b>	"65f" "65g" "65h"
65g (n=33)	"65f" "65g" "65h"
65h (n=25)	"65f" "65h"
<b>75a (n=20)</b>	"75a" "75b" "75d" "75ef"
75b (n=18)	"75a" "75b" "75d" "75ef"
75c (n=14)	"75c" "75d" "75ef"
<b>75d (n=9)</b>	"75b" "75c" "75d"
<b>75ef (n=9)</b>	"75b" "75c" "75d" "75ef"

**Table 2. Further test results for the subgroups of the eight ecoregions.**

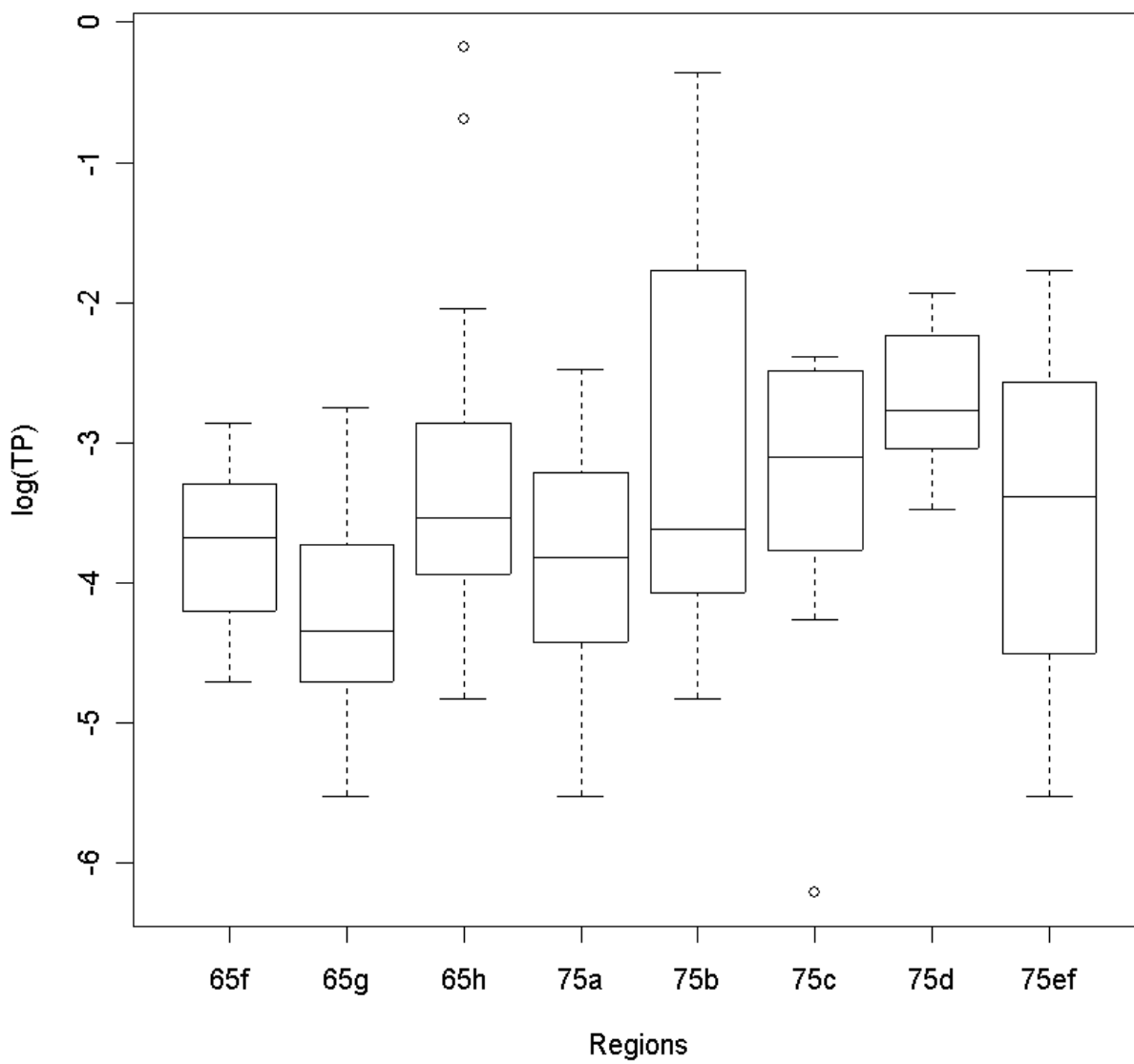
	Wilcoxon Rank Sum Test	Ansari Bradley Test	Kolmogorov Smirnov Test
Ecoregion group ("65f", "65g", "65h") vs. Ecoregion group ("75a" "75b" "75d" "75ef")	<b>W = 570.5</b> <b>p = <math>1.46 \times 10^{-11}</math></b>	<b>AB = 2376.5</b> <b>p = 0.0653</b>	<b>D = 0.6162</b> <b>p = <math>1.3 \times 10^{-10}</math></b>
Ecoregion 75c vs. Ecoregion group ("75a" "75b" "75d" "75ef")	<b>W = 405.5</b> <b>p = 0.8486</b>	<b>AB = 340.5</b> <b>p = 0.0093</b>	<b>D = 0.25</b> <b>p = 0.4858</b>

### Conclusions:

- Ecoregion 75c may be merged with Ecoregion group ("75a" "75b" "75d" "75ef") if the significance level  $\alpha = 0.005$  is used.
- The final merging groups based on specific conductivity field medians are ("65f", "65g", "65h"), 75c, and ("75a" "75b" "75d" "75ef").

2). Example 2 (Total Phosphorus Medians in the Eight Ecoregions; after the natural logarithm transformation)

Figure 2. Total Phosphorus Medians in the Eight Ecoregions



**Table 3. Initial merging groups based on the step-up procedure and the total phosphorus medians in the eight ecoregions ( $\alpha = 0.05$  is used).**

<b>Start Ecoregion</b>	<b>Merging groups</b>
<b>65f (n=10)</b>	<b>"65f" "65h" "75a" "75b" "75c" "75ef"</b>
65g (n=29)	<b>"65g" "75a"</b>
65h (n=20)	"65f" "65h" "75a" "75b" "75c" "75ef"
75a (n=17)	"65f" "65h" "75a" "75b" "75c" "75ef"
75b (n=18)	"65f" "65h" "75a" "75b" "75c" "75ef"
75c (n=10)	"65f" "65h" "75a" "75b" "75c" "75ef"
75d (n=8)	<b>"75c" "75d" "75ef"</b>
75ef (n=9)	"65f" "65h" "75a" "75b" "75c" "75ef"

**Table 4. Further test results for the subgroups of the eight ecoregions.**

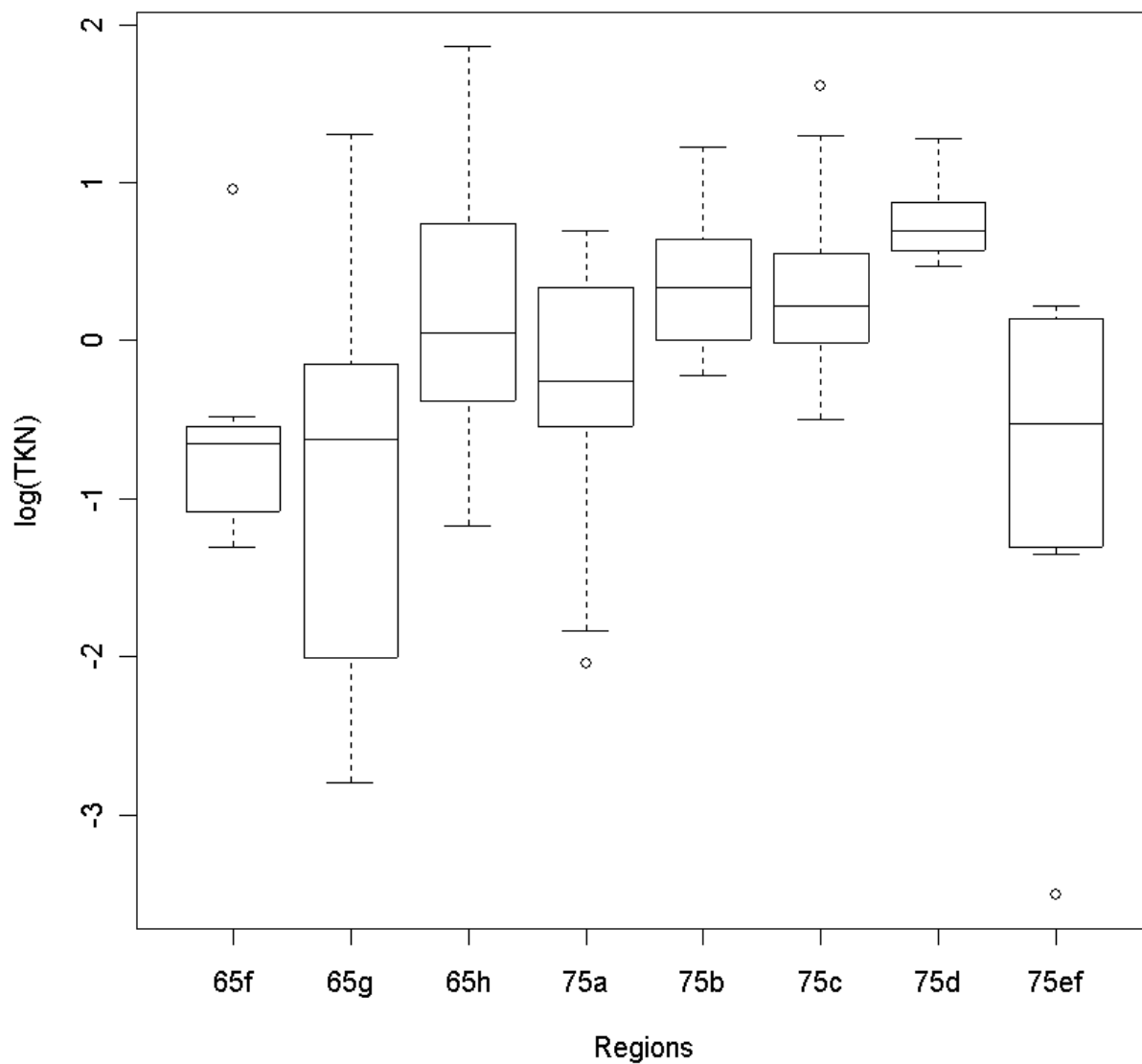
	<b>Wilcoxon Rank Sum Test</b>	<b>Ansari Bradley Test</b>	<b>Kolmogorov Smirnov Test</b>
Ecoregion 65g vs. Ecoregion group ("65f" "65h" "75a" "75b" "75c" "75ef")	<b>W = 1810.5</b> <b>p = <math>9.89 \times 10^{-5}</math></b>	<b>AB = 2409.5</b> <b>p = 0.9405</b>	<b>D = 0.3633</b> <b>p = 0.0068</b>
Ecoregion 75d vs. Ecoregion group ("65f" "65h" "75a" "75b" "75c" "75ef")	<b>W = 158</b> <b>p = 0.01389</b>	<b>AB = 1968</b> <b>p = 0.8678</b>	<b>D = 0.5774</b> <b>p = 0.01535</b>

### **Conclusions:**

- **Ecoregion 75d may be merged with the major ecoregion group ("65f" "65h" "75a" "75b" "75c" "75ef") if  $\alpha = 0.01$  is used.**
- **Ecoregion 65g could not be merged with the major ecoregion group ("65f" "65h" "75a" "75b" "75c" "75ef").**

**3). Example 3 (Total Kjeldahl Nitrogen (TKN) Medians in the Eight Ecoregions; after the natural logarithm transformation)**

**Figure 3. Total Kjeldahl Nitrogen Medians in the Eight Ecoregions**



**Table 5. Initial merging groups based on the step-up procedure and Total Kjeldahl Nitrogen medians in the eight ecoregions ( $\alpha = 0.05$  is used).**

<b>Start Ecoregion</b>	<b>Merging groups</b>
65f (n=9)	"65f" "75ef"
65g (n=26)	"65g" "75a" "75ef"
65h (n=18)	"65h" "75a" "75c"
75a (n=18)	"65g" "75a" "75ef"
75b (n=18)	"65h" "75b" "75c"
75c (n=12)	"65h" "75a" "75b" "75c"
75d (n=7)	"75d"
75ef (n=8)	"65f" "65g" "75ef"

**Table 6. Further test results for the subgroups of the eight ecoregions.**

	<b>Wilcoxon Rank Sum Test</b>	<b>Ansari Bradley Test</b>	<b>Kolmogorov Smirnov Test</b>
Ecoregion 75d vs. Ecoregion group ("65h" "75a" "75b" "75c")	<b>W = 87.5</b> <b>p = 0.0073</b>	<b>AB = 1253.5</b> <b>p = 0.5336</b>	<b>D = 0.697</b> <b>p = 0.0043</b>
Ecoregion group ("65f" "65g" "75ef") vs. Ecoregion group ("65h" "75a" "75b" "75c")	<b>W = 2249.5</b> <b>p = <math>2.6 \times 10^{-7}</math></b>	<b>AB = 1957.5</b> <b>p = 0.1181</b>	<b>D = 0.4843</b> <b>p = <math>1.04 \times 10^{-5}</math></b>

### **Conclusions:**

**The final merging ecoregion groups based on TKN are ("65f" "65g" "75ef"), ("65h" "75a" "75b" "75c"), and 75d.**

**References:**

Conover, W. J., Johnson, M.E., and Johnson M. M. (1981), "A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data."

*Technometrics* \_ **23**, 351-361.

Hollander, M., and D. Wolfe, (1999), *Nonparametric Statistical Methods, Second Edition*, John Wiley and Sons, New York, N.Y. 787p.