Krusal-Wallis ANOVA

The Kruskal-Wallis (KW) ANOVA is the non-parametric equivalent of a one-way ANOVA. As it does not assume normality, the KW ANOVA tests the null hypothesis of no difference between three or more group medians, against the alternative hypothesis that a significant difference exists between the medians. The KW ANOVA is basically an extension of the Wilcoxon-Mann-Whitney (WMW) 2 sample test, and so has the same assumptions: 1) the groups have the same spreads; and 2) the data distributions have the same shape.

The workings for the KW ANOVA, in essence, are the same as for the WMW test. That is the KW ANOVA takes the entire data, combines it into one sample, ranks the observations from smallest to largest, and then decomposes the now ranked data into their original groups. The KW ANOVA now runs a straightforward ANOVA on the ranked data. We cannot do Bonferroni Corrections on the groups if we reject the null hypothesis, but as the KW ANOVA is based on the WMW test, we can run WMW tests on each pair, remembering to adjust our *a* level by dividing by the number of comparisons being made.

As our assumption of normality was not entirely met by our data in the ANOVA example, we shall now run the same data through a KW ANOVA.

Formally, let η_{BOREAL} = the median of the population densities of hunter-gatherer groups in boreal forest ecosystems, let η_{TEMP} = the median of the population densities of hunter-gatherer groups in temperate forest ecosystems, and let η_{TROP} = the median of the population densities of hunter-gatherer groups in tropical forest ecosystems. We wish to test the following hypothesis at the $a = 0.05$ (95%) confidence level:

$$H_O : \eta_{BOREAL} = \eta_{TEMP} = \eta_{TROP}$$
$$H_A : \text{not } H_O$$

Here is the same data again:

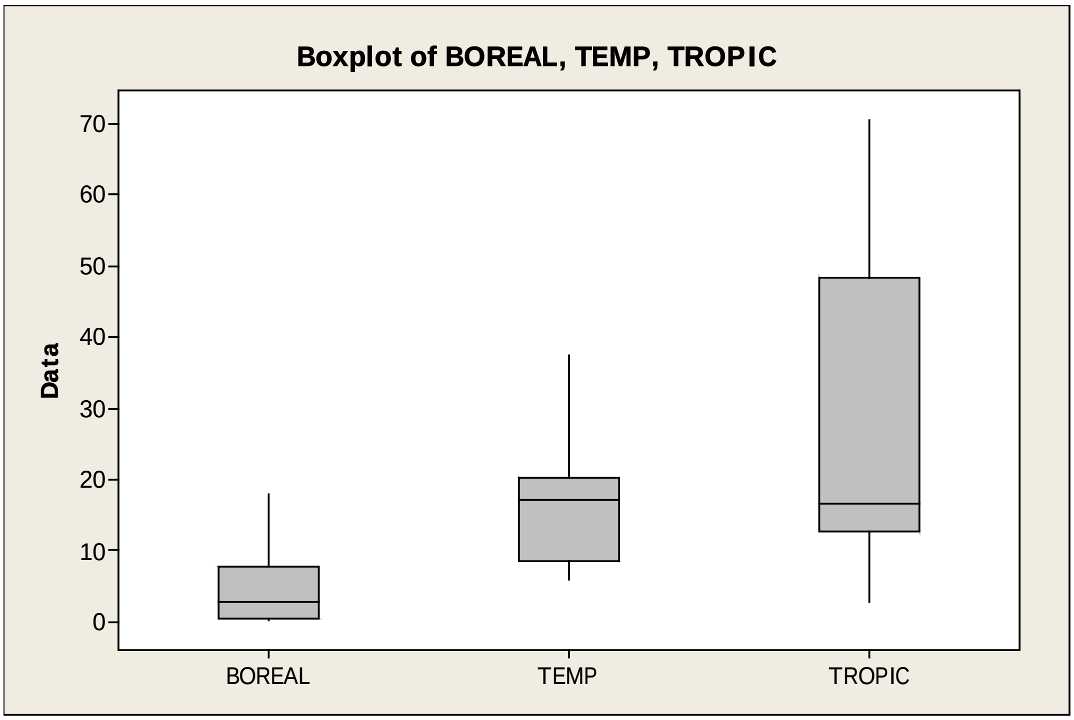
BOREAL	TEMP	TROPIC
0.64	18.50	15.60
7.50	19.31	2.90
17.93	17.84	14.60
8.69	6.00	59.80
0.51	8.80	44.65
3.98	11.76	7.00
0.88	37.50	70.37
3.84	8.00	17.70
1.80	23.16	33.38
0.33	16.28	15.20

Just to be consistent, we will run our descriptive stats again and look at some output.

Descriptive Statistics: BOREAL, TEMP, TROPIC

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
BOREAL	10	0	4.61	1.75	5.54	0.330	0.608	2.82	7.80	17.93
TEMP	10	0	16.72	2.91	9.21	6.00	8.60	17.06	20.27	37.50
TROPIC	10	0	28.12	7.29	23.06	2.90	12.70	16.65	48.44	70.37

And the boxplot:



So we see again that the data seems slightly skewed to the right and the means do not equal the medians. Of course, a normality test on the residuals in the last example suggested we these data are *nearly* normal, but not quite.

As the mechanics are the same as the WMW test we shall not run through them by hand but go straight into the MINITAB output.

First, if we have inputted our data into three columns, we need to stack the data using the same procedure as in the last example. Then we can go ahead with our KW ANOVA.

>STAT

>NON-PARAMETRIC

>KRUSKAL-WALLIS

>in RESPONSE, select your single data column

>in FACTOR, select your factor column

>OK

We get the output:

Kruskal-Wallis Test: TOTAL versus TYPE

```
Kruskal-Wallis Test on TOTAL

TYPE      N  Median  Ave Rank      Z
BOREAL   10   2.820       7.6  -3.48
TEMP     10  17.060      18.9   1.50
TROPIC   10  16.650      20.0   1.98
Overall  30              15.5

H = 12.16  DF = 2  P = 0.002
```

Here we see the medians for the groups, the average ranks and the associated Z scores for each group. The *H* statistic is the actual value of the test, and we see that the $p = 0.002$. Notice that this value is smaller than our ANOVA result, but the result is the same: reject the null hypothesis in favor of the alternative. This means we now have to run WMW multiple comparisons.

First, the total number of comparisons is given by $0.5s(s\text{-}1) = 3$.

Our adjusted $a = (0.05/3)*100 = 1.667$

So, we run three WMW tests at the $a = 0.01667$ (98.3%) confidence level between η_{BOREAL} and η_{TEMP}, η_{BOREAL} and η_{TROP}, and η_{TEMP} and η_{TROP}.

In the MINITAB dialog box for the WMW test remember to change the significance level of the test to 98.3% from the default level of 95%.

The output from the three tests is as follows:

Mann-Whitney Test and CI: BOREAL, TEMP

```
         N  Median
BOREAL  10    2.82
TEMP    10   17.06


Point estimate for ETA1-ETA2 is -11.19
98.6 Percent CI for ETA1-ETA2 is (-18.98,-4.02)
W = 64.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0022
```

Mann-Whitney Test and CI: BOREAL, TROPIC

```
         N  Median
BOREAL  10    2.82
TROPIC  10   16.65


Point estimate for ETA1-ETA2 is -14.70
98.6 Percent CI for ETA1-ETA2 is (-52.29,-3.02)
W = 67.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0046
```

Mann-Whitney Test and CI: TEMP, TROPIC

```
         N  Median
TEMP    10   17.06
TROPIC  10   16.65


Point estimate for ETA1-ETA2 is -6.17
98.6 Percent CI for ETA1-ETA2 is (-38.64,8.56)
W = 98.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.6232
```

We are primarily interested in the significance levels of this output.

Test 1 suggests there is a significant difference between the Boreal groups and the Temperate groups as $p < a$. Test 2 indicates there is also a significant difference between Boreal and Tropical groups (not suprising), but Test 3 indicates there is no significant difference between Temperate and Tropical groups. This result can be summarized:

BOREAL TROP TEMP

This is a different result to the Bonferroni adjustments, where there was no significant difference between the Boreal and Temperate groups. This result suggests that the Boreal population density is significantly lower than either of the other two forest ecosystems and we can see why by looking at the initial test output: the median population density for Boreal groups is 2.82 people per km^2, whereas for Temperate groups it is 17.06, and for Tropical groups, it is 16.65. This result tells us that we were right to question the assumption of normality of our data from the residuals of the one-way ANOVA, not in terms of the effect non-normality was having on the initial ANOVA result (they were both highly significant), but for our multiple comparisons, we should have been testing the medians as they are the more accurate measure of central tendency in this case.

Now, we still have the problem that not all the variances are equal, and so one additional thing we could do is to log transform the data, and see if we can standardize those variances. But I will let you do that.