

## One-Way Analysis of Variance (ANOVA) and Multiple Comparisons

For this example, we return to the population density of hunter-gatherers in three different forest ecosystems (data taken from Binford 2000). We have three ecosystems ( $s = 3$ ), each with a sample size of ten hunter-gatherer groups ( $n = 10$ ). For this problem we are interested in whether there is a significant difference between the mean group sizes across the three ecosystems, and if so, what is the underlying pattern of variation between the groups. The three ecosystems are boreal, temperate, and tropical forests. Notice that in general the three forest types roughly represent a north to south gradient in latitude.

Let  $\bar{Y}_{BOR}$  = average population density (per km<sup>2</sup>) for boreal forest groups, let  $\bar{Y}_{TEMP}$  = average population density for temperate forest groups, and let  $\bar{Y}_{TROP}$  = average population density (per for tropical forest groups. Formally, we wish to test the hypothesis (at the  $\alpha = 0.05$ , or 95% level):

$$H_O : \bar{Y}_{BOR} = \bar{Y}_{TEMP} = \bar{Y}_{TROP}$$

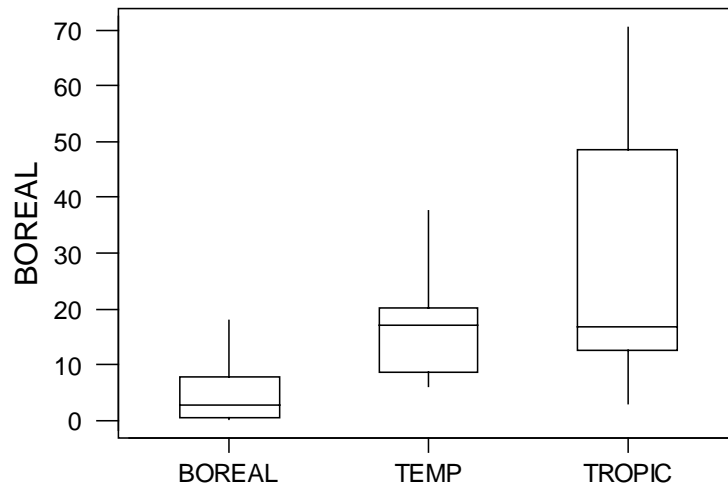
$$H_A : \text{not } H_O$$

The data are:

<b>BOREAL</b>	<b>TEMP</b>	<b>TROPIC</b>
0.64	18.5	15.6
7.5	19.31	2.9
17.93	17.84	14.6
8.69	6	59.8
0.51	8.8	44.65
3.98	11.76	7
0.88	37.5	70.37
3.84	8	17.7
1.8	23.16	33.38
0.33	16.28	15.2

First we look at the descriptive statistics for the three ecosystems and produce some graphics:

<b>Descriptive Statistics</b>						
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
BOREAL	10	4.61	2.82	3.48	5.54	1.75
TEMP	10	16.72	17.06	15.46	9.21	2.91
TROPIC	10	28.12	16.65	25.99	23.06	7.29
Variable	Min	Max	Q1	Q3		
BOREAL	0.33	17.93	0.61	7.80		
TEMP	6.00	37.50	8.60	20.27		
TROPIC	2.90	70.37	12.70	48.44		



We see from the output that the means are not equal, but not enormously different, and the variances are different. An assumption of ANOVA is that these samples are normally distributed, but we will test this formally later and note for now that they look roughly symmetrical and there doesn't seem to be any clear violations of normality, so we can go ahead with our ANOVA.

First we shall go through calculating the quantities:

Quantity 1: *The Grand Total*

$$\sum^s \sum^n Y = 46.1 + 196.15 + 270.2 = \mathbf{494.45}$$

Quantity 2: *The sum of squared individual observations*

$$\sum^s \sum^n Y^2 = 488.63 + 3556.80 + 12426.69 = \mathbf{16739.52}$$

Quantity 3: *The sum of squared group totals, each divided by its sample size*

$$\sum \frac{(\sum^n Y)^2}{n_i} = \frac{(46.1)^2}{10} + \frac{(196.15)^2}{10} + \frac{(270.2)^2}{10} = \mathbf{10913.78}$$

Quantity 4: *The Correction Term*

$$CT = \frac{(\sum^s \sum^n Y)^2}{\sum^s n_i} = \frac{(483.45)^2}{30} = \mathbf{8149.36}$$

Quantity 5: *The total sum of squares*

$$SS_{TOTAL} = \sum^s \sum^n Y^2 - CT = 16472.12 - 7790.797 = \mathbf{8590.157}$$

Quantity 6: *The group sum of squares*

$$SS_{GROUP} = \sum^s \frac{(\sum^n Y)^2}{n_i} - CT = 10307.24 - 7790.797 = \mathbf{2764.417}$$

Quantity 7: *The sum of squares within*

$$SS_{WITHIN} = SS_{TOTAL} - SS_{GROUPS} = 8681.31 - 2516.441 = \mathbf{5825.74}$$

Then we calculate the ANOVA table:

Source of Variation	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F<sub>s</sub></i>
<i>Among groups</i>	2	2764.417	1382.2085	6.41
<i>Within groups</i>	27	5825.74	215.768	
<i>Total</i>	29	8681.321		

$F_{CRIT}(df_{AMONG} = 2, df_{WITHIN} = 27, \alpha = 0.05) = 3.35$ . To establish a  $p$  value, we look for the values that bracket our  $F_{STAT}$  in the table at  $df(2, 27)$ , and that is 6.49 and 5.49. The corresponding  $p$  value is  $0.005 > p < 0.01$ , and as 6.41 is closer to 6.49 than 5.49, our  $p$  value will be slightly larger than 0.005. Anyway, at this stage we reject the null hypothesis of no difference in favor of the alternative, as at the  $\alpha = 0.05$  level  $p < \alpha$  (and  $F_{STAT} > F_{CRIT}$ ): there is a significant difference between the average population densities across ecosystems. Are they all

different from each other, or is one ecosystem type different from the others? Now we move to the multiple comparisons. But first, let's do the above test through MINITAB.

The procedure is:

### 1. DATA PREPARATION

**>Enter your data in three columns first so you can run the descriptive stats**

### 2. DATA STACKING

**>MANIP**

**>STACK/UNSTACK**

**>STACK**

**>Choose your three data columns**

**>STORE DATA in an empty column**

**>STORE SUBSCRIPTS IN the next empty column**

**>OKAY**

You will now have two more columns to your spread sheet; a stacked data column and a factor column. Keep track of the factors; 1 = BOREAL, 2 = TEMPERATE, 3 = TROPICAL.

### 3. THE ANOVA

**>STAT**

**>ANOVA**

**>ONE WAY (NOT UNSTACKED)**

**>RESPONSE is your stacked data column**

**>FACTOR is your factor column**

**>STORE RESIDUALS**

**>Click COMPARISONS**

**>Select FISHER'S I.E.R.**

**>Input 1.67\***

**>OKAY**

\* This value is the correction term for the *Fisher's Individual Error Rate* that tricks MINITAB into doing a *Bonferroni Correction*.

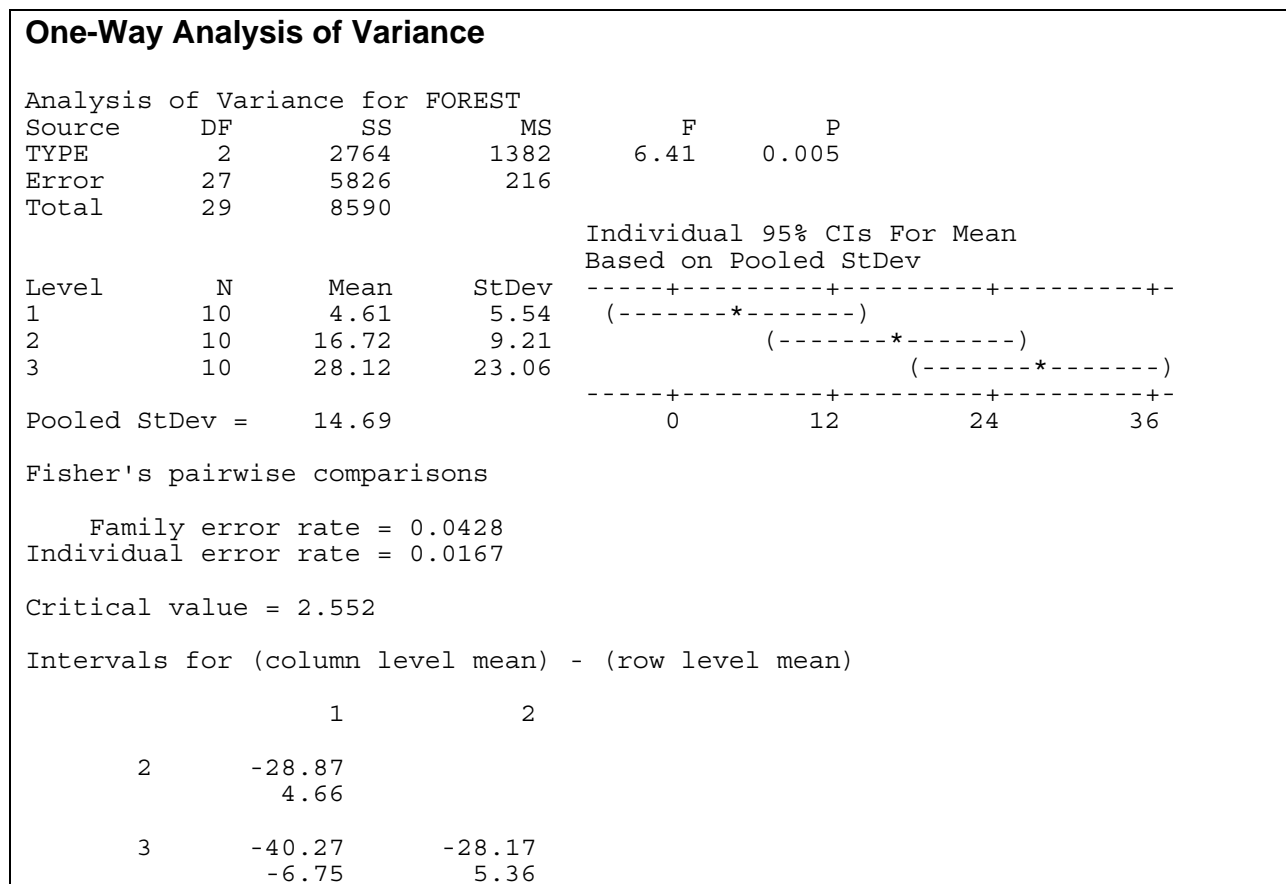
$$\text{The Bonferroni Correction} = \left( \frac{\alpha}{[s(s-1)]0.5} \right) 100 = \left( \frac{0.05}{3} \right) 100 = 1.66\bar{6} \%$$

The denominator of the first term in the Bonferroni Correction the number of possible comparisons in a given sample.

$$\# \text{ of Comparisons} = \frac{s(s-1)}{2}$$

The Bonferroni Comparison corrects for accumulating  $\alpha$  errors by dividing by the number of comparisons being made.

The MINITAB output should look like:



Let's go through the output. First, notice the computed ANOVA table is the same as our hand-calculated ANOVA table, including the  $p$  value, leading us to the same conclusion; reject the null hypothesis in favor of the alternative.

As we have rejected our null hypothesis, we look at the Multiple Comparisons to find out the underlying structure of the relationships between the ecosystems.

First we see that there are means and standard deviations given for each ecosystem, 1, 2, and 3 (remember that 1 = BOREAL, 2 = TEMPERATE, and 3 = TROPICAL). Next to that output we see a simple graphic of confidence intervals. This is given because, in reality, multiple comparisons are simply 2-tailed 2 Sample  $t$ -tests between all groups, adjusting for  $\alpha$ . Next, the Fisher's *Family Error Rate* is really the Bonferroni Comparison with the *IER* set at the adjusted  $\alpha$  level.

Underneath that we see a matrix of comparisons:

Intervals for (column level mean) - (row level mean)		
	1	2
2	-28.87 4.66	
3	-40.27 -6.75	-28.17 5.36

This matrix is comparison of all possible combinations of our data, given in terms of the confidence limits of a 2 Sample  $t$ -test. In that case, the first entry is a test of sample 1 vs. sample 2, and the confidence limits are set around the difference between the two sample means. If one bound is negative, and one is positive they bound a hypothetical value of zero, meaning that there is no statistical difference between the two. This matrix indicates that there is no significant difference between samples 1 and 2, or 2 and 3, but there is a difference between 1 and 3 as both bound are negative. This can be summarized in the following way:

Align the groups from the smallest mean to the largest and link the groups that are non-significant with a line:

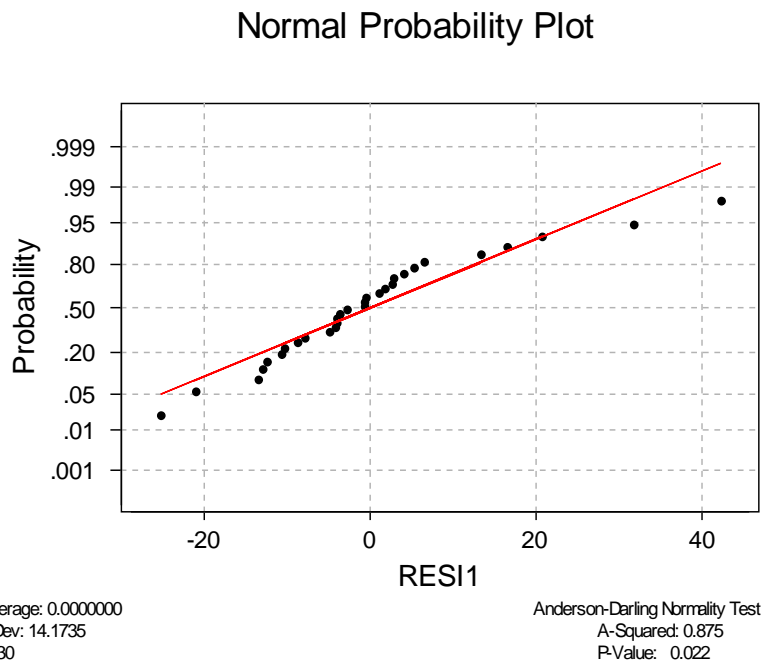
1      2      3  
 \_\_\_\_\_

This gives us the same information but in an easier way. Now we have to be careful about how we interpret it: the Bonferonni Correction suggests that that there is no significant difference between samples 1 and 2, and 2 and 3, but we do not have enough evidence to suggest there is no significant difference between samples 1 and 3 (we state it in these terms as we have not proved that there is no difference between these two samples, our data simply doesn't detect a difference).

What does this mean? Well our ANOVA suggested that there is a significant difference between the mean population densities of hunter-gatherers across three different ecosystem types. The Bonferonni Corrections (or comparisons, or statistics) suggest that hunter-gatherer population densities in boreal forests do not differ from those in temperate forests, and those in temperate forests do not differ from tropical forests, however, there is a significant difference between

population densities in tropical forests and boreal forests. This makes intuitive sense as biogeographically boreal forests are the most northern, followed by temperate forests, and the tropical forests are closest to the equator. We know that most measures of environmental production decrease from north to south, and if we assume the productivity of the environment affects population density, especially for hunter-gatherer populations, we have found an underlying macroecological pattern; hunter-gatherer population densities follow basic patterns of global biogeography to the extent that while latitudinal-bordering forest types do not differ statistically, population densities in the tropics are statistically greater than at northern latitudes.

However, we are not finished quite yet. In order to check our assumptions we should run a normality test on the stored residuals. These should be normally distributed as they are due to random error (in the mathematical, not the colloquial sense).



Here we see our residuals do not follow the straight line exactly, but it is close. The Anderson-Darling  $p = 0.022$ , which is less than the usual  $\alpha = 0.05$ , but it is relatively close, so I can be relatively confident in these results. I could also run a non-parametric ANOVA to see if there was any difference in the outcome. So I did, and there isn't.