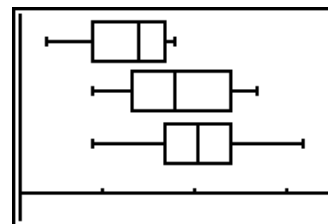# Chapter 14 – Analysis of Variance (ANOVA)

We have already seen two-sample tests for equality of the means in chapter 10. What if there are more than two groups? Answering the question relies on comparing variation among the groups to variation within the groups, hence the name. The null hypothesis for ANOVA is always that all groups have the same mean and the alternate is that at least one group has a mean different from the others.

Wild irises are beautiful flowers found throughout North America and northern Europe. Sir R. A. Fisher collected data on the sepal lengths in centimeters from random samples of three species. The data are below. Do these data indicate the mean sepal lengths are similar or different?

Iris setosa	Iris versicolor	Iris virginica
5.4	5.5	6.3
4.9	6.5	5.8
5.0	6.3	4.9
5.4	4.9	7.2
5.8	6.7	6.4
5.7	5.5	5.7
4.4	6.1	
	5.2	

TI-83/84 Procedure

I have entered the data into lists L1, L2, and L3. We will first construct side-by-side boxplots of the data for visual comparison. Visually, the medians are somewhat different with Iris virginica being the largest.

From the STAT TESTS menu select choice F:ANOVA(. This is the last test on the menu, so it is easiest to find by pressing the up arrow. The command shell is transferred to the home screen.

To complete the command, enter the list names separated by commas, then press ENTER.

This is the first portion of the output. The value of the F statistic is 2.95 and the p-value for the test is 0.0779 which indicates at $\alpha = 5\%$ there is not a significant difference in the mean sepal lengths for the three species, based on this sample. The Factor degrees of freedom are k-1 where k is the number of groups, so with three groups, this is 2. MS is SS/df.

105

Pressing the down arrow several times gives the remainder of the output. Degrees of freedom for Error are $n - k$, where n is the total number of observations in all groups (21 here) and k is the number of groups (3). MS is again SS/df. Sxp is the estimate of the common standard deviation and is the square root of MSE. The F statistic is MSTR/MSE.

```
One-way ANOVA
↑ MS=1.22074405
 Error
  df=18
  SS=7.44803571
  MS=.413779762
 Sxp=.643257151
■
```

TI-89 Procedure

From the STAT Tests menu select choice C:ANOVA. This is the next-to-last test on the menu, so it is easiest to find by pressing the up arrow.

The first input screen asks if the data are in lists or if there are merely summary statistics. We also need to identify the number of groups (3 in this example). In each case, press the right arrow and make the appropriate selection. Press ENTER when finished to proceed.

The second input screen asks for the list names. Use 2nd – (VAR-LINK) to access the list of list names and select those you have used.

This is the first portion of the output. The value of the F statistic is 2.95 and the p-value for the test is 0.0779 which indicates at α = 5% there is not a significant difference in the mean sepal lengths for the three species, based on this sample. The Factor degrees of freedom are k-1 where k is the number of groups, so with three groups, this is 2. MS is SS/df.

Pressing the down arrow several times gives the remainder of the output. Degrees of freedom for Error are n-k, where n is the total number of observations in all groups (21 here) and k is the number of groups (3). MS is again SS/df. Sxp is the estimate of the common standard deviation and is the square root of MSE. The F statistic is MSFactor/MSError.

After pressing ENTER to clear the output screen, we find something has been added into the Statistics editor. The first column contains the means for each group. The next two lists, lowlist and uplist, contain the lower and upper limits of individual 95% *t*-intervals for the mean, where Sxp has been used as the common standard deviation. Looking at these is a crude method of determining whether group means are equal or not. If the intervals have a large amount of overlap it is an indication that group means are not different. Since all of these intervals overlap, it is a confirmation of the decision that the mean sepal lengths of the three species are similar.
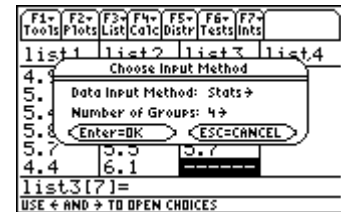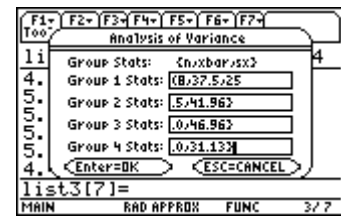
Another TI-89 example

The TI-89 can also perform the analysis using the summary statistics for each group. Your text uses an example about the number of bacteria left on hands after washing with different types of cleanser. The statistics are presented in the table below.

Level	n	Mean	St. Dev.
Alcohol Spray	8	37.5	25.56
Antibacterial soap	8	92.5	41.96
Soap	8	106.0	46.96
Water	8	117.0	31.13

On the first Anova screen, I have indicated that I am using the Stats method of input, and that there are four groups.

Input the summary statistics for each group in the order specified, and enclosed by curly braces. These are [2nd][(] and [2nd][)]. Failure to include the curly braces will result in an error message that (at least) one is missing.

The output indicates gives the F statistic 7.13, with a p-value of 0.0011. We reject the null hypothesis and conclude that the method of hand washing does affect the number of bacteria left on the skin. Note that there is some rounding error present in my calculations. This is probably due to the fact that I used the (rounded) summary statistics, rather than the actual data.
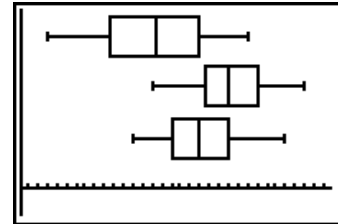
ANOTHER EXAMPLE

The following data represent yield (in bushels) for plots of a given size under three different fertilizer treatments. Does it appear the type of fertilizer makes a difference in mean yield?

Type A	Type B	Type C
21	41	35
24	44	37
31	38	33
42	37	46
38	42	42
31	48	38
36	39	37
34	32	30

Here are side-by-side boxplots of the data. All three distributions appear symmetric and there is not a large difference in spread, so ANOVA is appropriate. Using `1-Var Stats` we find the mean for Type A is 32.125 bushels, Type B has a mean of 40.125 bushels, and Type C has a mean of 37.25. They're different, but are they different enough to believe the three types of fertilizer have different effects on crop yield?

The ANOVA output of interest is at right. We see the F-statistic is 4.05 and the p-value for the test is 0.0325. At α = 0.05, we will reject the null hypothesis and conclude that at least one fertilizer has a different mean yield than the others. Scrolling down, we further find `Sxp` = 5.696.

WHICH MEAN(S) IS (ARE) DIFFERENT?

Having rejected the null hypothesis, we would like to know which mean (or means) are different from the rest. For a rough idea of the difference one can do confidence intervals for each mean (using the pooled standard deviation found in the ANOVA) and look for intervals which do not overlap, but this method is flawed. Similarly, testing each pair of means (doing three tests here) has the same problem: the problem of multiple comparisons. If we constructed three individual confidence intervals, the probability they all contain the true value is $0.95^3 = 0.857$, using the fact that the samples are independent of each other.

The Tukey Method

The Tukey method is one way to solve this question. From tables, we find the critical value q^* for k groups and $n - k$ degrees of freedom for error. My table gives $q^*_{3,21,0.05} = 3.58$. The "honestly significant difference" is computed as $HSD = \dfrac{q^*}{\sqrt{2}} s_p \sqrt{\dfrac{2}{n_i}}$ where each group has n_i observations. (NOTE: most tables of this distribution require the division of q^*, but not all. Check your particular table.) Here, $HSD = \dfrac{3.58}{\sqrt{2}} * 5.695 \sqrt{\dfrac{2}{8}} = 7.208$. Groups with means that differ by more than 7.208 are significantly different. The mean for Type A was 32.125; for Type B 40.125 and for Type C 37.25. We therefore declare types A and C have similar means; Types B and C also have similar means. The means which are different from each other are those for types A and B.

The Bonferroni Method

The Bonferroni method is based on t distributions, unlike the Tukey method. One disadvantage is that you will require a calculator with the inverse t function, or a computer to find exact t^* multipliers. Another disadvantage with this method is that you must know ahead of time how many comparisons you will be doing.

The Bonferroni method for making j comparisons, involves computing a "confidence interval" for each comparison using a confidence level of $1 - \alpha/j$ which entails finding a critical value t^{**} with $n - k$ degrees of freedom, and $1 - \alpha/2j$ tail probability. For our fertilizer data we want three comparisons, so for 95% overall confidence we need a 98.33% confidence interval, or t^{**} for 99.16% area below the critical value. My calculator indicates the critical value is 2.601.

Our data had 8 observations in each group, and a pooled standard deviation of 5.696, so the least significant difference with this method (the margin of error for a confidence interval for the difference) becomes

$$LSD = t** \times s_{xp} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.601 * 5.696 \sqrt{\frac{1}{8} + \frac{1}{8}} = 7.408.$$ Any two group means

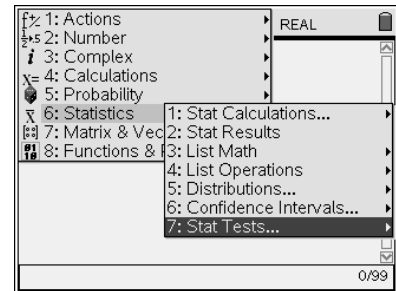that are farther apart than 7.408 will be called significantly different. Again, Types A and B are different.

Notice the Bonferroni least significant difference is slightly larger than the Tukey quantity. If you have many groups, and want comparisons for all possible pairs, the Bonferroni method will yield much larger LSDs than the Tukey method.

Commands for the TI-Nspire™ Handheld Calculator

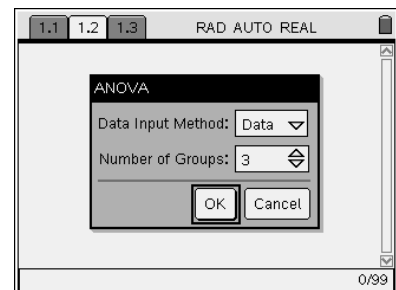To run an ANOVA, first place the data into lists. Values for the flowers example are shown.

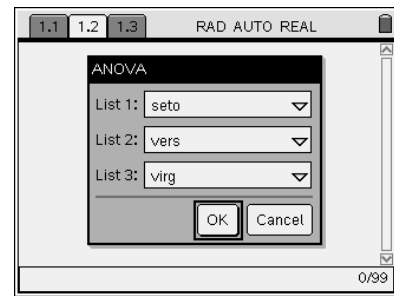Then open a Calculator page, press (menu) , select Statistics and Stat Tests.

Select ANOVA.

Select Data and the number of groups, three for this example.

109

Select the names of the lists for the groups.