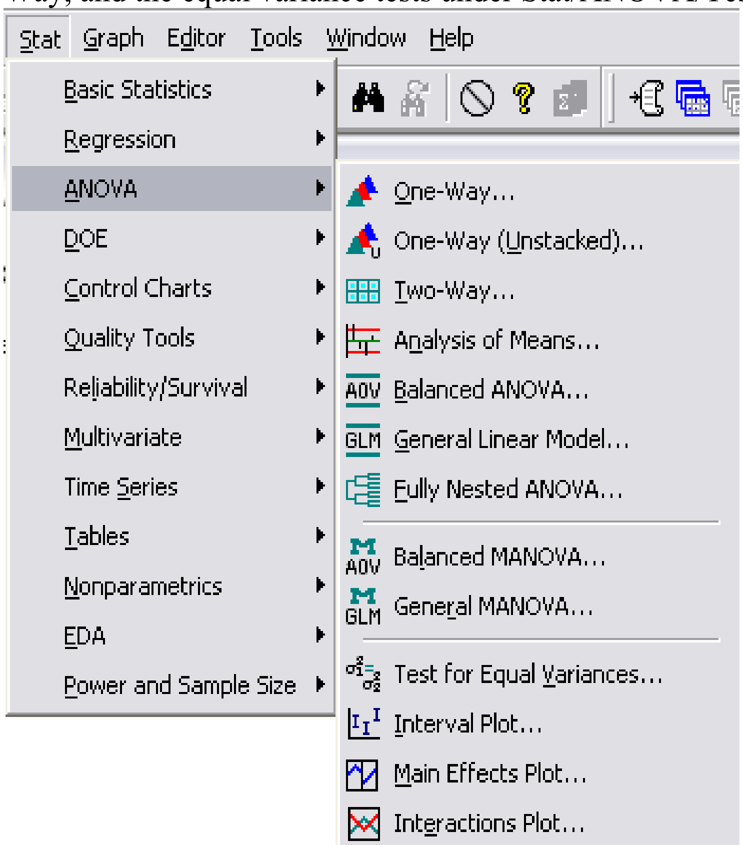
Stat 538 - Biostatistics I - Fall 2005

Lab 8

ANalysis Of VAriance (ANOVA) and Multiple Comparisons –
Hypothesis Testing in the many-sample situation of a population mean

Minitab can perform ANOVA and multiple comparisons under the Stat/ANOVA/One-Way, and the equal variance tests under Stat/ANOVA/Test for Equal Variances.



The example we will use today comes from SW p. 483 exercise 11.13, available on the Labs website in the Chpt 11 directory called spine. However, because the homework has the data in the UNSTACKED row form below, let's use this and transform it into a STACKED form. We need the data stacked for the multiple comparisons.

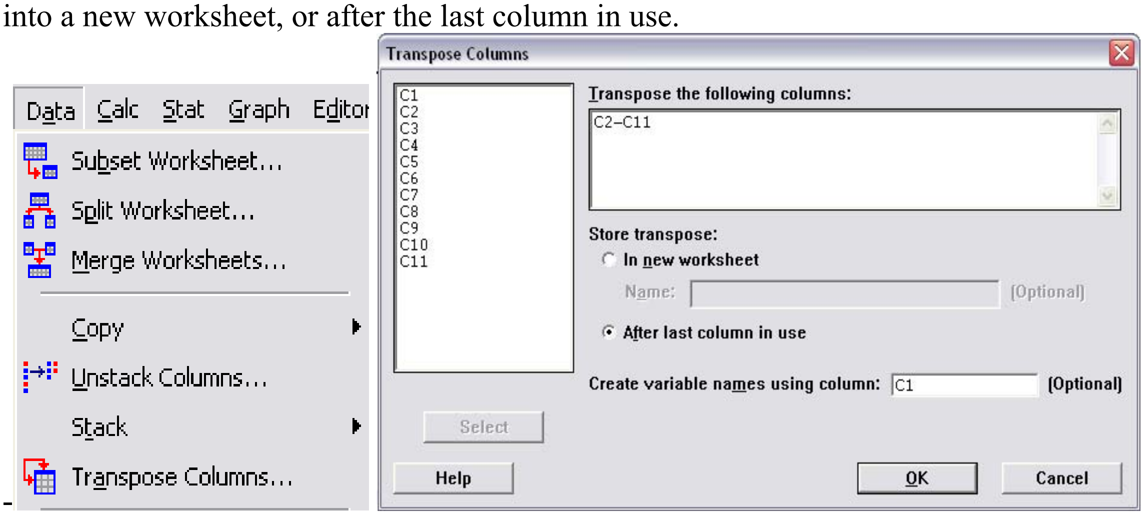
```
aerobics  0.25 -1.50  0.25  1.00 -0.25  0.00  0.25 -1.50  0.25 -0.50
control   0.50 -0.25 -0.25  0.00 -0.75  1.00 -0.25  0.75  0.50
modern    1.00  1.00  0.75  2.00  1.00  0.75  1.00  0.50  2.50 -0.75
```

These are three independent samples, so we will perform a ANOVA hypothesis test and perform a multiple comparison test to identify the difference is one is detected in the ANOVA.

First we STACK the data. There are a few ways to do this, but I found the easiest way is to first transpose the columns (making rows into columns), then stack the columns. Copy and paste the data above into Minitab so that it appears this way. Notice that the control group has one fewer value, and this empty space is indicated by a "*". Minitab calls this a missing value; there isn't actually any data missing, we simply have a smaller sample size for this group.

↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	aerobics	0.25	-1.50	0.25	1	-0.25	0.00	0.25	-1.50	0.25	-0.50
2	control	0.50	-0.25	-0.25	0	-0.75	1.00	-0.25	0.75	0.50	*
3	modern	1.00	1.00	0.75	2	1.00	0.75	1.00	0.50	2.50	-0.75

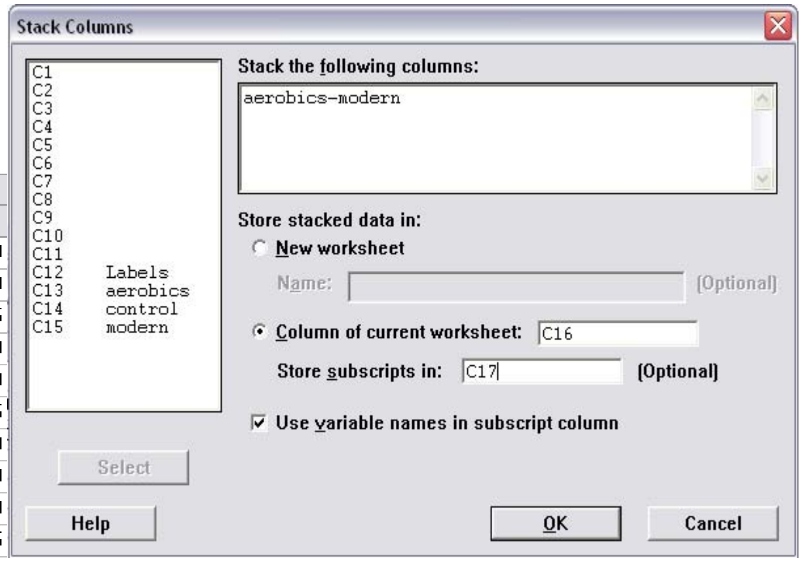
Next Transpose the columns with Data/Transpose Columns. Select the columns with data (C2-C11) for the "Transpose the following columns", put the column with the group names (C1) into "Create variable names using column". You can choose to put the data into a new worksheet, or after the last column in use.



The results looks like the table below, three UNSTACKED columns of data, with a forth column with the original column Labels which we don't care about. At last we stack the columns with Data/Stack/Columns. We want C13-15 (aerobics-modern) in the "Stack the following columns", and check the box "Use variable names in subscript column".

You can store the result in a new worksheet or in a set of columns in the current worksheet as I do here.

C12-T	C13	C14	C15
Labels	aerobics	control	modern
C2	0.25	0.50	1.00
C3	-1.50	-0.25	1.00
C4	0.25	-0.25	0.75
C5	1.00	0.00	2.00
C6	-0.25	-0.75	1.00
C7	0.00	1.00	0.75
C8	0.25	-0.25	1.00
C9	-1.50	0.75	0.50
C10	0.25	0.50	2.50
C11	-0.50	*	-0.75



The result is two columns, one which I've labeled "change_in_flexibility", and the other "class". I show the screenshot to point out that the missing value for the last "control" is still there. The analysis we do will simply ignore this value as though it wasn't there, which is what we want it to do since we really don't have a 10th observation, but only 9 for that group.

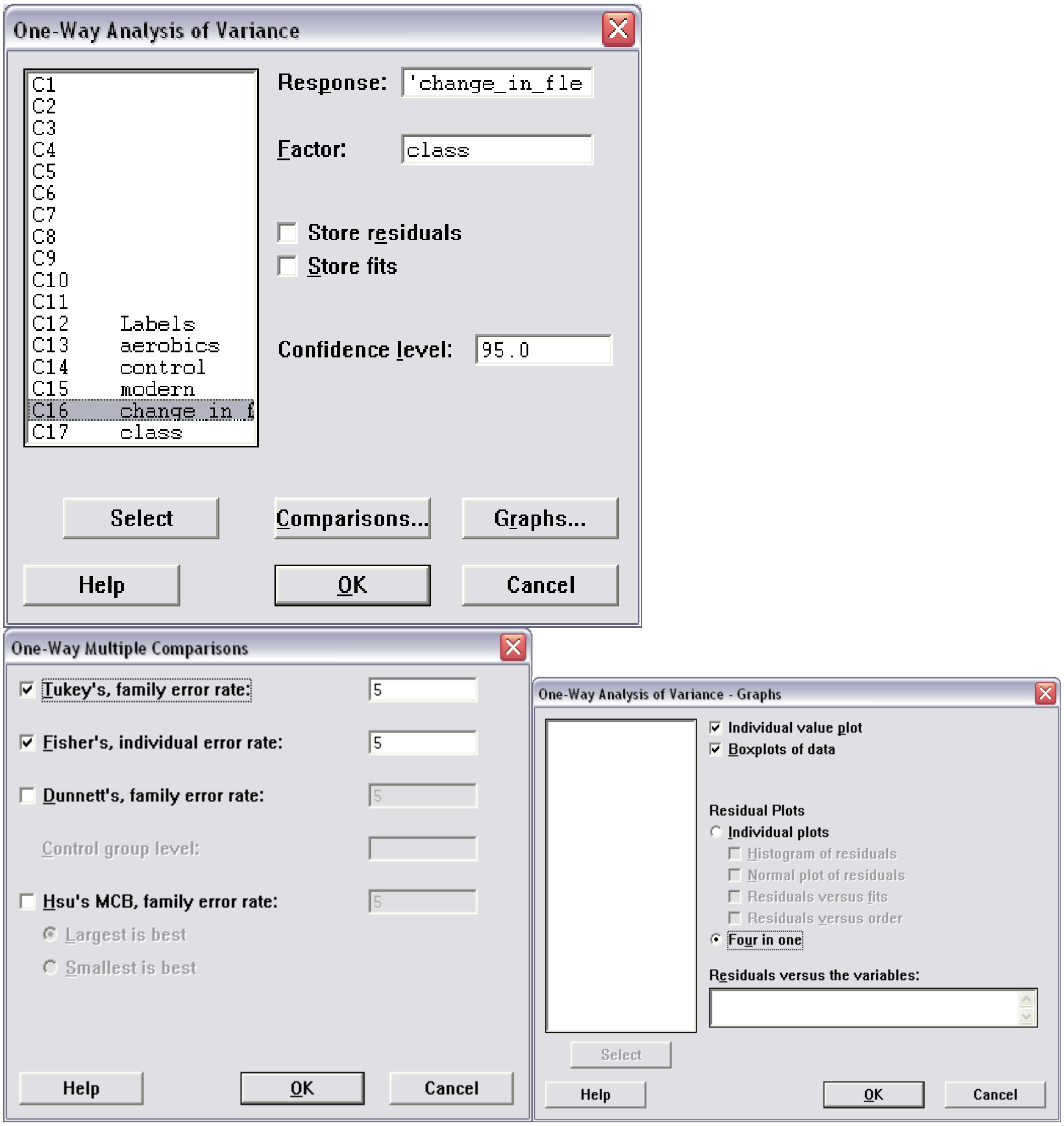
C16	C17-T
change_in_flexibility	class
0.25	aerobics
-1.50	aerobics
0.25	aerobics
1.00	aerobics
-0.25	aerobics
0.00	aerobics
0.25	aerobics
-1.50	aerobics
0.25	aerobics
-0.50	aerobics
0.50	control
-0.25	control
-0.25	control
0.00	control
-0.75	control
1.00	control
-0.25	control
0.75	control
0.50	control
*	control
1.00	modern
1.00	modern

Let's perform an Analysis of Variance, ANOVA, in Minitab. Minitab allows you to do almost everything you'd want to do in one step. This lab will help you understand what the different pieces are, what they tell us, and how to keep them organized.

1. First we will look at boxplots and individual value plots to visually see how the groups compare. We will describe what we see.
2. Second, we will perform an ANOVA to test whether these three populations have roughly equal means, or whether at least one mean is different from the others. We will check the assumptions of normality and equal standard deviations at this point.
3. Third, if we reject H_0 and detect that there is a difference in means, then we will perform a (set of) multiple comparison tests to identify what the differences between means are. We will discuss these differences.

Let's look at Minitab's ANOVA procedure and look at the options:

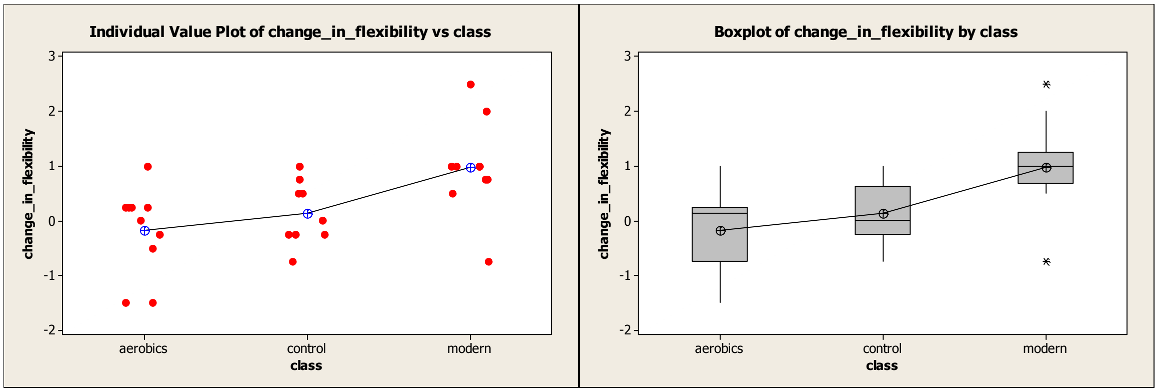
First, input the data values "change_in_flexibility" in the "response", and the grouping variable "class" in "factor". Note the Confidence level is set at 95%, which will give us an α=0.05 level test. Just this alone will give us our ANOVA table with test of the hypothesis that all the means are equal versus at least one mean is different.
Under the **comparisons** button is where the multiple comparisons are performed. We have discussed **Fisher's LSD**, **Tukey's HSD**, and **Bonferroni**. Note: Bonferroni is simply Fisher with α divided by the number of comparisons, where the number of pair-wise comparisons among *k* groups is *k(k-1)/2*. For *k*=3 groups, we have 3*2/2=3 pair-wise comparisons. For *k=4* groups, we have 4*3/2=6 pair-wise comparisons, and so on. In our example with three groups, and with α=0.05, the Bonferroni method uses Fisher's with .05/3=1.66666. Dunnett's and Hsu's methods are in similar spirit but different.
Under the **graphs** button, we can get boxplots and individual value plots. Additionally, diagnostic plots of the residuals are available. Since these were not discussed in lecture, I will simply show you what they are, but won't go into any detail. I choose the four-in-one to get all 4 available plots together.



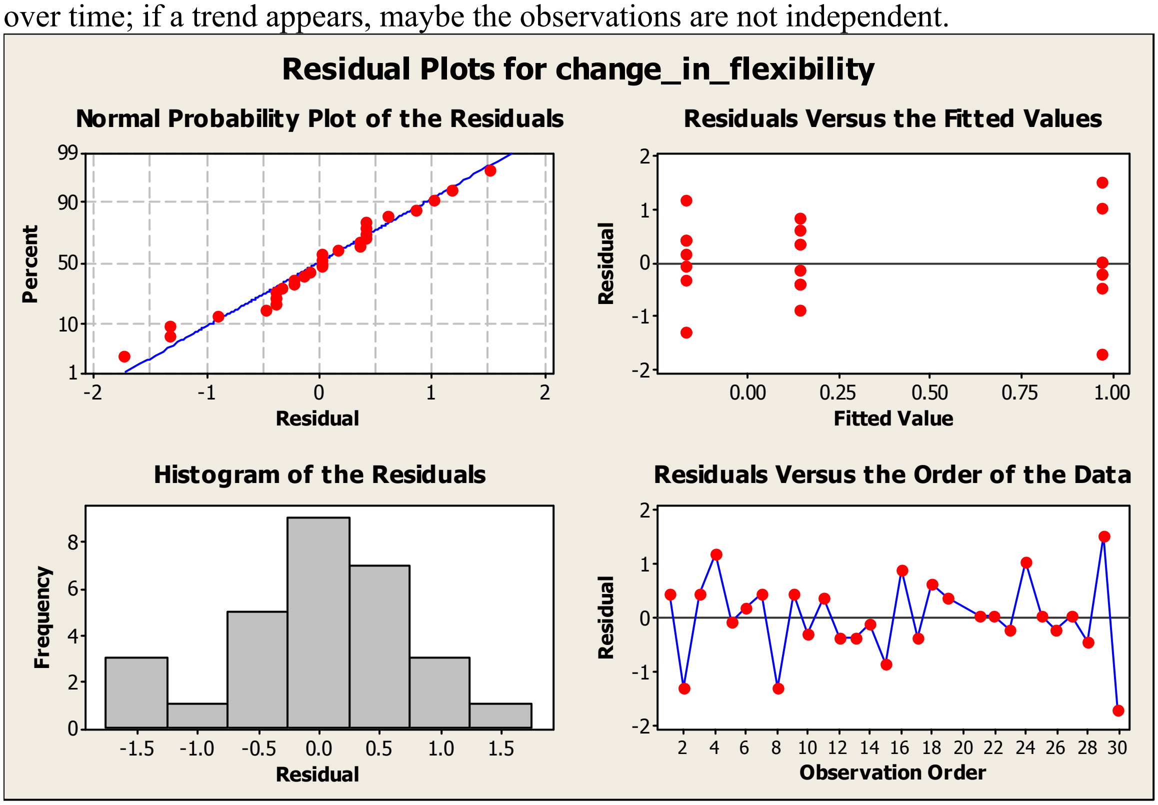
When you click OK, lots of output and graphs appear. We'll look at the graphs, at the ANOVA table, then at the multiple comparisons.

1. Plots

Below are the individual value plot, the box plot, and the residual plots. We observe from both the individual value and box plots that the mean change in flexibility increases from aerobics, to control, to modern. The class modern is a bit higher than the other two, which are close to the same.



Here is what the residual plots tell us. A residual is the difference between the individual observations and the predicted mean, in this case the actual mean of each group. The first plot is testing whether the data appears normal, which in this case it does. The second plot (going to the right) shows the spread of each of the groups from their mean, used to assess whether the variance is the same. The third plot is a histogram of the residuals, which appear unimodal, symmetric, no outliers, so roughly normal. The last plot shows the observations in the order they appear in the data set, which is a visual check for trends over time; if a trend appears, maybe the observations are not independent.



2. The ANOVA table below tests:

H_0: $\mu_a=\mu_c=\mu_m$ that all means are equal, versus

H_A: At least one mean is different from the others.

There are three required assumptions: (1) Independent random samples from each population, (2) the population frequency curves are normal, and (3) the populations have equal standard deviations, $\sigma_a= \sigma_c=\sigma_m$.

Because the *p*-value in the table is less than α=0.05, we reject H_0 in favor of $H_{A,}$ concluding at least one mean is different from the others. Below the table are the group's sample sizes N, means, and StDevs, with a visual representation of their confidence intervals. Notice that the aerobics and modern intervals do not overlap – but this does not mean they are different in this case, we need to perform a multiple comparison test to test this difference.

One-way ANOVA: change_in_flexibility versus class

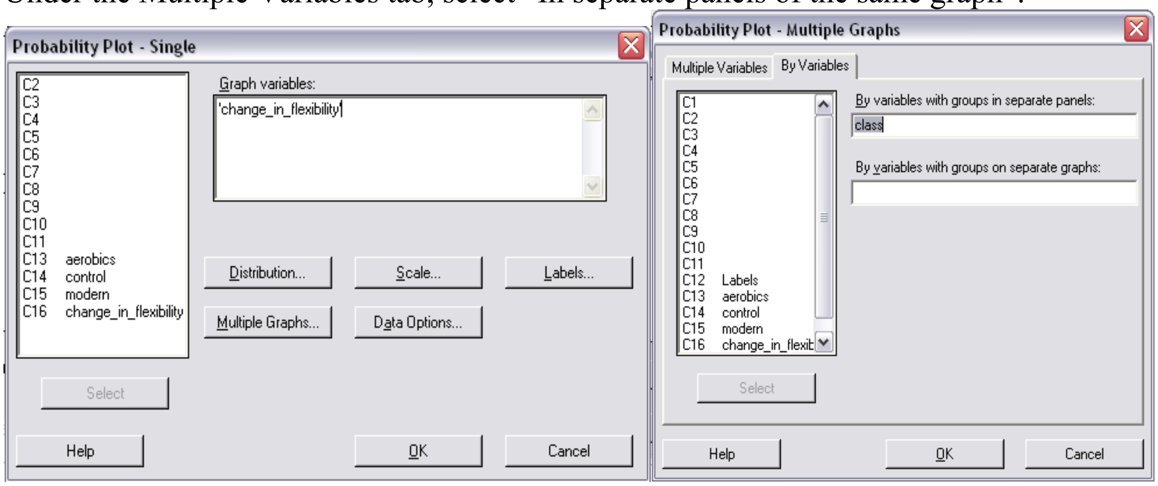
```
Source  DF      SS     MS     F      P
class    2   7.036  3.518  6.07  0.007
Error   26  15.076  0.580
Total   28  22.112

S = 0.7615   R-Sq = 31.82%   R-Sq(adj) = 26.57%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level      N     Mean   StDev    -+---------+---------+---------+--------
aerobics  10  -0.1750  0.7997    (-------*--------)
control    9   0.1389  0.5743         (-------*---------)
modern    10   0.9750  0.8616                        (-------*-------)
                                 -+---------+---------+---------+--------
                               -0.60      0.00      0.60      1.20
Pooled StDev = 0.7615
```

Checking Assumptions

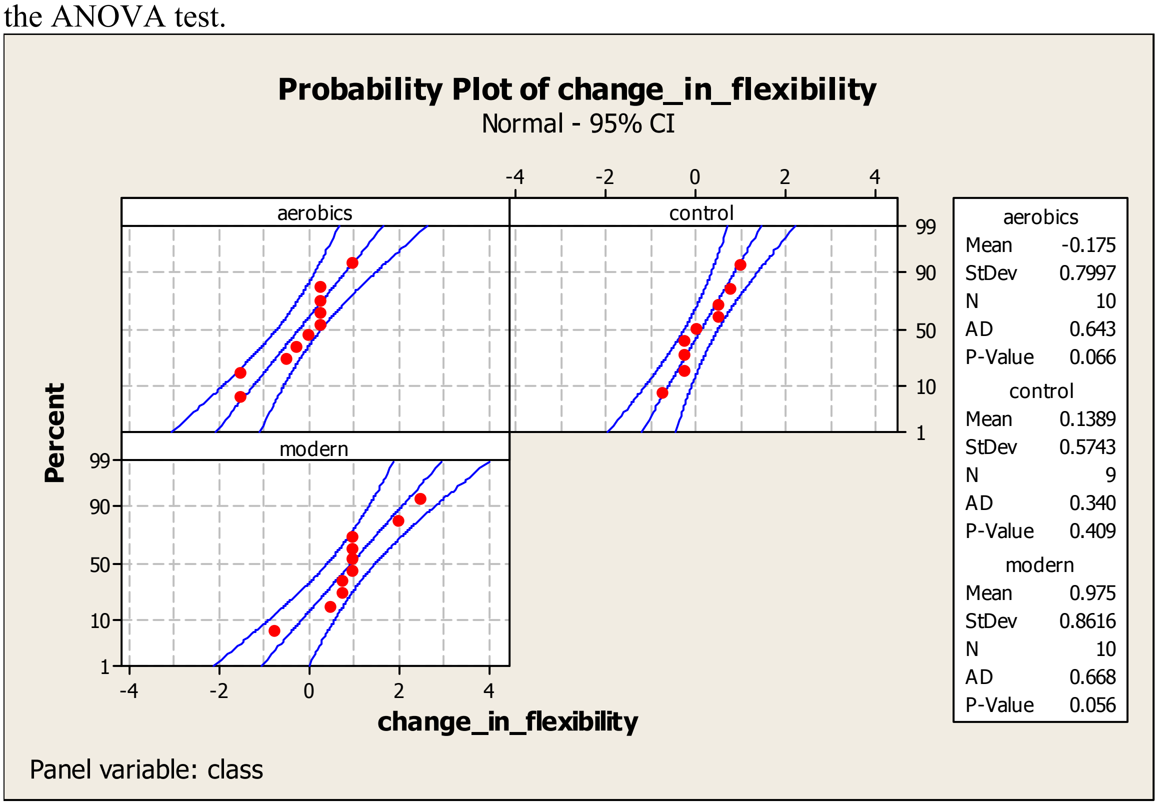
First we'll check normality, then equal variances (standard deviations).

To check normality, create a rack of normal probability plots, which will include their Anderson-Darling normality test statistics and *p*-values.

Graph/Probability Plot
Under the Multiple Variables tab, select "In separate panels of the same graph".



These plots indicate each of the populations are reasonably normal to use the results of the ANOVA test.



Stat/ANOVA/Test for Equal Variances.

Bartlett's and Levene's test is of

H_0: $\sigma_a = \sigma_c = \sigma_m$ all standard deviations are equal, versus

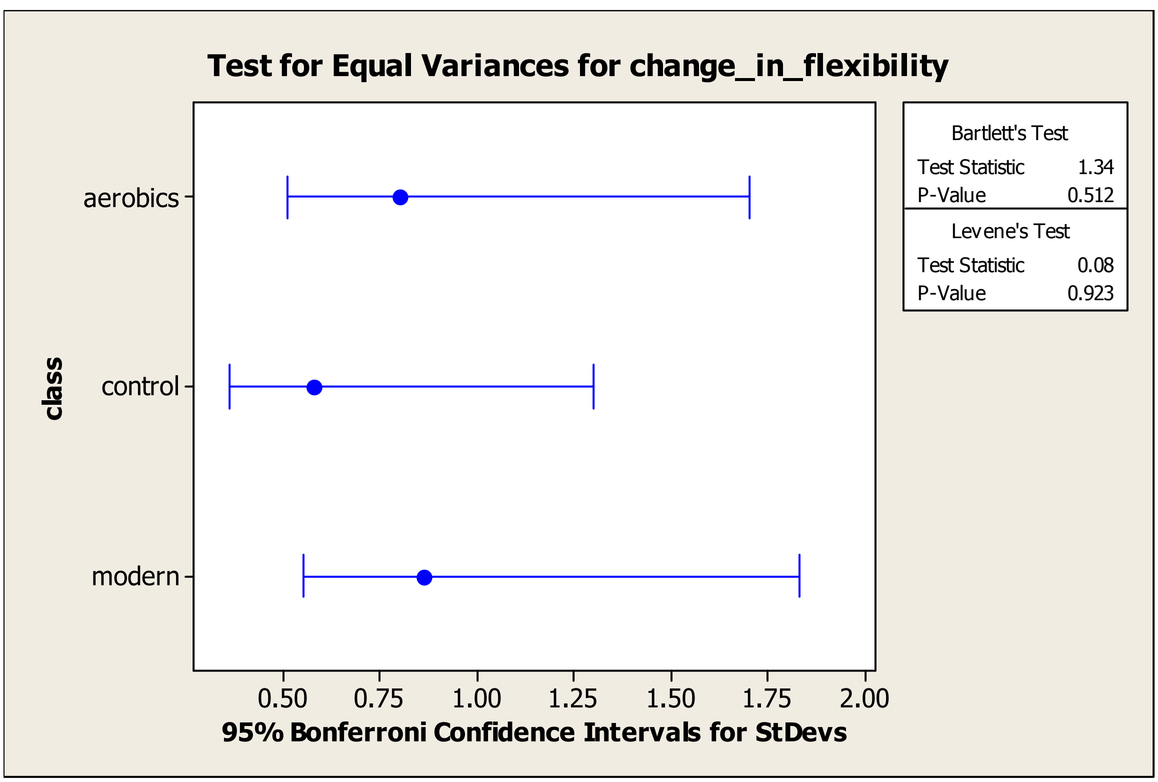
H_A: At least one standard deviation is different from the others.

Bartlett's assumes the populations are normal, while Levene's just assumes the populations are continuous (not discrete data).

Below both Bartlett's test and Levene's test have large p-values, also all the confidence intervals overlap, so we fail to reject H_0 concluding there is insufficient evidence that these standard deviations are different.

Test for Equal Variances: change_in_flexibility versus class

```
95% Bonferroni confidence intervals for standard deviations

   class   N     Lower     StDev    Upper
aerobics  10  0.509465  0.799740  1.70191
 control   9  0.358037  0.574335  1.30063
  modern  10  0.548875  0.861604  1.83356

Bartlett's Test (normal distribution)
Test statistic = 1.34, p-value = 0.512

Levene's Test (any continuous distribution)
Test statistic = 0.08, p-value = 0.923
```



Because both assumptions are satisfied, we can be comfortable with the conclusion of the ANOVA test, and continue with our multiple comparisons to determine which difference exists between the groups to cause us to reject H_0 in the ANOVA.

Multiple comparison tests:

I will proceed in the order from most liberal (most likely to detect a difference) to most conservative (least likely to detect a difference). That is, Fisher's LSD, Tukey's HSD, and Bonferroni's method.

How to use these. Consider this comparison graph taken from the first Fisher's LSD, which compares the mean of classes control and modern to aerobics. Aerobics is the class mean we are "comparing to", control and modern are the class means we're "comparing". Because the confidence interval for control contains 0, control population mean is not different from aerobics population mean (which is, by the way, the same as saying that aerobics population mean is not different from control population mean). The modern confidence interval does not contain 0, so modern population mean is different from aerobics population mean.

```
class = aerobics subtracted from:

class       Lower  Center   Upper  ------+---------+---------+---------+---
control   -0.4053  0.3139  1.0331              (------*------)
modern     0.4500  1.1500  1.8500                      (------*-------)
                                   ------+---------+---------+---------+---
                                      -1.0       0.0       1.0       2.0
```

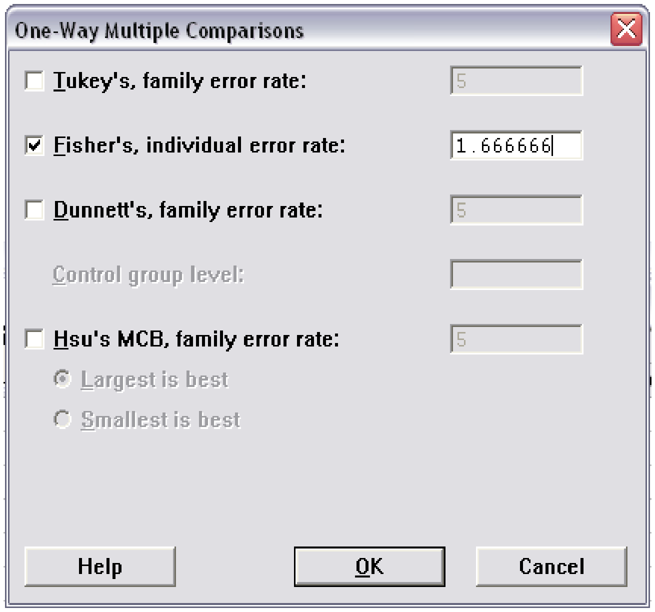
Fisher's LSD does not control the family-wise error rate. These are the standard individual confidence intervals. Below we first see that control is not different from aerobics, and that modern is different from aerobics. Next we see that modern is different from control. By "different", I mean that the population means for the populations are significantly different.

```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of class

Simultaneous confidence level = 88.09%

class = aerobics subtracted from:
class      Lower  Center   Upper  ------+---------+---------+---------+---
control  -0.4053  0.3139  1.0331             (------*------)
modern    0.4500  1.1500  1.8500                     (------*-------)
                                  ------+---------+---------+---------+---
                                     -1.0       0.0       1.0       2.0

class = control subtracted from:
class    Lower  Center   Upper  ------+---------+---------+---------+---
modern  0.1169  0.8361  1.5553                   (------*-------)
                                ------+---------+---------+---------+---
                                   -1.0       0.0       1.0       2.0
```

Tukey's HSD controls the family-wise error rate. These are slightly wider than the individual confidence intervals, notice below the individual confidence level is increased to 98.01. Below we first see that control is not different from aerobics, and that modern is different from aerobics. Next we see that modern is **not** different from control, which is different from the conclusion using Fisher's LSD. Again, by "different", I mean that the population means for the populations are significantly different.

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of class

Individual confidence level = 98.01%

class = aerobics subtracted from:
class      Lower  Center   Upper  -------+---------+---------+---------+--
control  -0.5545  0.3139  1.1823            (--------*--------)
modern    0.3048  1.1500  1.9952                     (--------*-------)
                                  -------+---------+---------+---------+--
                                      -1.0       0.0       1.0       2.0

class = control subtracted from:
class     Lower  Center   Upper  -------+---------+---------+---------+--
modern  -0.0323  0.8361  1.7045                  (-------*--------)
                                 -------+---------+---------+---------+--
                                     -1.0       0.0       1.0       2.0
```

Bonferroni's method controls extremely conservatively for the family-wise error rate. These are much wider than the individual confidence intervals, by setting the individual confidence level to 98.333. [Again, Bonferroni is simply Fisher with α divided by the number of comparisons, where the number of pair-wise comparisons among *k* groups is *k(k-1)/2*. For *k*=3 groups, we have 3*2/2=3 pair-wise comparisons. For *k=4* groups, we have 4*3/2=6 pair-wise comparisons, and so on. In our example with three groups, and with α=0.05, the Bonferroni method uses Fisher's with .05/3=1.66666.] Below is a screen shot for how to do this in the ANOVA/Comparisons procedure.



Below we first see that control is not different from aerobics, and that modern is different from aerobics. Next we see that modern is not different from control, which is different from the conclusion using Fisher's LSD, but the same as Tukey's HSD. Again, by "different", I mean that the population means for the populations are significantly different.

```
Fisher 98.3333% Individual Confidence Intervals
All Pairwise Comparisons among Levels of class

Simultaneous confidence level = 95.74%

class = aerobics subtracted from:
class      Lower  Center   Upper  -------+---------+---------+---------+--
control  -0.5814  0.3139  1.2092            (--------*--------)
modern    0.2786  1.1500  2.0214                     (---------*-------)
                                  -------+---------+---------+---------+--
                                      -1.0       0.0       1.0       2.0

class = control subtracted from:
class     Lower  Center   Upper  -------+---------+---------+---------+--
modern  -0.0592  0.8361  1.7314                (--------*--------)
                                 -------+---------+---------+---------+--
                                     -1.0       0.0       1.0       2.0
```

Wasn't that fun?!