Math 4581 **Additional Notes II: ANOVA**

ANOVA is short for analysis of variation. It is used to test to see if the means of two or more samples are the same.

Example: Suppose we have the following three groups and want to test at the 5% level of signficance to see if the population means are the same:
1. 270, 245, 190, 215, 250, 230
2. 240, 135, 160, 230, 250, 200, 200, 210
3. 180, 155, 200, 120, 140, 180, 140, 130
We get $\bar{x}_1 = 233.3$, $s_1 = 28.23$; $\bar{x}_2 = 203.125$, $s_2 = 39.36$; $\bar{x}_3 = 155.625$, $s_3 = 26.39$

We have our usual question: is the difference between the means real or due to random chance?

Idea: If the variance between groups (find the variance of the numbers $\bar{x}_1, \bar{x}_2, \bar{x}_3$) is much bigger than the variance within groups (find the pooled variance of s_1, s_2, s_3), then population means for the three groups are probably not all the same.

Notes:
(i) The null and alternate hypotheses for ANOVA are always:
$H_0 : \mu_1 = \mu_2 = \ldots = \mu_g$, H_a :the means are not all the same. Therefore, like the χ^2-test, there is no 1-sided version.

(ii) It is assumed that each sample is a random sample from a normal population. If the sample sizes are large, this can be assumed. If the sample sizes are small, this should be checked (one method is to get the normal quintile plots).

(iii) It is assumed the variance of each population is the same, σ^2. A rough guideline is that if the largest standard deviation is less than double the smallest, this assumption is appropriate. The estimator for σ^2 is the pooled variance s_p^2.

Details:

We will mostly use a calculator or software for ANOVA, but should know how it works to decide if it's appropriate.

We will assume that $x_{ij} = \mu_i + \epsilon_{ij}$, where x_{ij} is the jth point in the ith group, μ_i is the population mean of the ith group, and ϵ_{ij} is the 'error' of x_{ij}. Given our assumptions, we can assume the ϵ_{ij} are from a normal population with a mean of 0 and variance of σ^2.

The estimator for μ_i is $\bar{x}_i$. Under the null hypothesis it's assumed that the μ_i are all the same μ, the estimator for this is $\bar{x}$ (the mean of all the points).

It will be convenient at this point to look at sums of squares instead of variances. Define (let g=number of groups, N=total number of points):

SSG=$\displaystyle\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{x})^2$= sum of squares between groups

It can be shown that SSG=$\displaystyle\sum_{i=1}^{g} n_i(\bar{x}_i - \mu)^2 - n(\bar{x} - \mu)^2$ and this can be used to show

E(SSG)=$(g-1)\sigma^2 + \displaystyle\sum_{i=1}^{g} n_i(\mu_i - \mu)^2$

Notice that if H_0 is true then E(SSG)=$(g-1)\sigma^2$, so we will reject H_0 if we get something much larger than this. Of course we will need to know the distribution, but you should be able to guess:

Theorem: If H_0 is true then SSG/σ^2 has a χ^2 distribution with $g-1$ degrees of freedom.

The problem is this assumes that we know σ and we usually don't. We will estimate this using the pooled standard deviation (really the sum):

SSE=$\sum_{i=1}^{g}(n_i-1)s_i^2$= sum of squares of the error (or within groups)

SSE/σ^2 is also a χ^2 distribution but with $n-g$ degrees of freedom.

Note: SST=$\sum\sum(x_{ij}-\bar{x})^2$= sum of squares of the total= SSG+SSE.

and we have the corresponding means:

MSG=SSG/(g-1) MSE=SSE/(N-g) MST=SST/(N-1).

Theorem: If H_0 is true then $\frac{MSG}{MSE}$ has an F distribution with $g-1$ and $n-g$ degrees of freedom.

Note: If U is a χ^2 distribution with m degrees of freedom and V is a χ^2 distribution with n degrees of freedom, then $\frac{U/m}{V/n}$ is an F distribution with m and n degrees of freedom:

$$f_X(x)=\frac{\Gamma(\frac{m+n}{2})m^{m/2}n^{n/2}x^{(m/2)-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(n+mx)^{(m+n)/2}} \text{ for } x\geq 0$$

Aside: you might notice that the square of the student distribution, T^2, is $\frac{Z^2}{s^2}$ which makes it an F distribution with 1 and n-1 degrees of freedom.

Now let's go back to the example and use a calculator to get these numbers.

The ANOVA command gives:

(Press STAT, choose EDIT, and put the points into columns L1, L2, and L3;
press STAT, move over to Tests, and choose ANOVA;
in the main screen fill in ANOVA(L1,L2,L3) and press enter).

F=10.118, p=.001

Factor (this is our between groups information, so SSG, MSG)
df=2, SS=21729.73, MS=10864.87

Error (this is our error or within groups information)
df=19, SS=20402.08, MS=1073.79

s_{xp}=32. 77 (this is the pooled standard deviation=$\sqrt{MSE}$)

Since the p-value is less than .05, we reject the null hypothesis and believe the means of the three groups are not all the same. Notice that we really only need the p-value from the calculator's output, although we will use s_p for the follow-up tests.

If we were doing this by hand, we would get the test statistic (which we found to be F=10.118 here) and find the critical value from the F-table. The degrees of freedom are 2 and 19 (in the denominator) which gives F=3.52 as the critical value.

Homework:

Test each of the following at the 5% level to see if the groups are the same.

1. Below are the overall ratings of instructors of four calculus 1 classes:
A. 4.9, 4.2, 2.7, 4.4, 3.9, 4
B. 4.8, 3.5, 4, 3.9, 4.2
C. 4.6, 3.5, 2.9, 4.2, 4.4, 3.7, 2.6, 4.1, 4.4, 4.5, 4.5, 4.3
D. 4.2, 4.2, 4.1, 4.1, 3.8, 4.1, 4.8, 4.8, 4.4, 3.5, 3.1

2. Below are the grades on a final from 3 different sections of the same course:
A. 364, 379, 358, 226, 319, 380, 213, 309, 362, 338, 295, 370, 379, 384, 294
B. 346, 395, 347, 284, 375, 385, 371, 243, 331, 384, 374, 386, 396
C. 354, 346, 367, 386, 364, 357, 375, 393, 353, 395, 381, 374, 339, 397

3. Below are the size of classes given in the Spring, Summer, and Fall:

Spring. 40, 41, 36, 33, 15, 32, 13, 41, 31
Summer. 21, 47, 42, 35, 25, 30
Fall. 66, 36, 27, 36, 46, 22, 28, 32, 19, 43, 17, 35

4. Below are the weights of a type of tree, in kg, one year after they are planted under different treatments:

None: .15, .02, .16, .37, .22
Fertilizer: 1.34, .14, .02, .08, .08
Irrigation: .23, .04, .34, .16, .05
Fert. and Irrig.: 2.03, .27, .92, 1.07, 2.38

Now let's go back to the example. We found that they do not all have the same mean, but this should be unsatisfying. We don't know which pairs, if any, are different. Given $\bar{x}_1 = 233.3$, $\bar{x}_2 = 203.125$, $\bar{x}_3 = 155.625$, it's not a bad guess that the third group is different from the other two but we want to be able to validate that guess.
There are several ways to do this, we'll look at two.

Contrasts

Looking back at the example, possible questions are: 'are the first two groups different from the third?'; 'are the first and third groups different?'.
Definition: A contrast is a combination of the population means, $\sum a_i\mu_i$, so that:
$\sum a_i = 0$ for some a_i

Look at the example using the question: are the first two groups different from the third? This can be represented with the hypotheses:

$H_0 : \frac{1}{2}(\mu_1 + \mu_2) = \mu_3$, $H_a : \frac{1}{2}(\mu_1 + \mu_2) \neq \mu_3$

This gives: $\frac{1}{2}(\mu_1 + \mu_2) - \mu_3 = 0$, so $a_1 = a_2 = \frac{1}{2}, a_3 = -1$

We now want to know the decision rule.

Theorem: The contrast $C = \sum a_i\mu_i$ has the estimator $c = \sum a_i\bar{x}_i$ where $Var(c) = MSE \sum \frac{a_i^2}{n_i}$.
If we're testing $H_0 : C = 0$,
$t = \dfrac{c}{\sqrt{Var(c)}}$ is a t-statistic with N-g degrees of freedom
The confidence interval for C is: $c \pm t\sqrt{Var(c)}$
Note: Contrasts should be designated before the data is found.

Bonferroni Method

A contrast is suitable if we want to look at one comparison. A more systematic method would instead look at all pairs. Basically we'll be doing two sample t-tests with a couple variations:

(i) When we compare μ_i to μ_j, we look at: $t = \dfrac{\bar{x}_i - \bar{x}_j}{s_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$ (note the s_p here is the one from all the samples)

(ii) If we are testing at the α level of significance then we will test each pair at the α/p level where p is the number of pairs.

Note: As usual we can also look at the confidence interval for the differences:

$$(\bar{x}_i - \bar{x}_j) \pm t \cdot s_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Let's go back to the example:

For the contrast $H_0 : \frac{1}{2}(\mu_1 + \mu_2) = \mu_3$, $H_a : \frac{1}{2}(\mu_1 + \mu_2) \neq \mu_3$, if we're testing at the 5% level we would get:

$c = \frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \bar{x}_3 = 62.59$, $Var(c) = 1073.79\left(\frac{(1/2)^2}{6} + \frac{(1/2)^2}{8} + \frac{(-1)^2}{8}\right) = 477.7$

$\rightarrow t = \frac{62.59}{\sqrt{477.7}} = 2.86$=test statistic

t(df=19, .025)=2.093=critical value.

Thus we would reject H_0 and believe the third group is different from the other two.

For the pairs (Bonferroni), we would get:

$t_{1,2}$ (for the first two groups)= $\dfrac{233.3 - 203.125}{32.77\sqrt{\frac{1}{6} + \frac{1}{8}}} = 1.705.$

In a similar way $t_{1,3} = 4.39$ and $t_{2,3} = 2.9$

t(df=19, .025/3=.0083)=2.861=critical value (I use .005 from the table, this is more conservative than using .001).

We conclude that the first and third pairs and the second and third pairs are different, while the first and second could have the same population mean.

Note: Since we have made it harder to reject the null hypothesis for the pairs, it's possible that we reject the null hypothesis for the ANOVA test and yet reject the null hypothesis for none of the pairs.

Homework:

Test at the 5% level.

5. Looking at problem **1**, are the first two groups different from the second two?

6. Looking at problem **2**, test all the pairs to see which, if any, are different.

7. Looking at problem **3**,
a) test to see if the summer is different from the other two semesters.
b) test all the pairs.

8. Looking at problem **4**,
a) test to see if the None and Irrigation groups are different from the other two.
b) Test at the 6% level of significance to see if the second and third groups are equal using the Bonferroni method.