One-way Analysis of Variance (ANOVA) followed by a stepwise multiple comparisons procedure called Student–Newman–Keuls (SNK).

Peter A Noble PhD, November 2020

Overview: To run ANOVA and SNK tests on 2000 samples, I developed a program in C++ called 'ANOVA.cpp'. One problem I encountered in calculating ANOVA happened when all replicates in a group have zero values. Zeros could mean that no sample was taken or that the value of the sample was below the minimum detection level. To solve this problem, the program looks for samples that were all zeros and essentially removes them from the ANOVA.

To ensure that the ANOVA works properly, I used an example on an internet site (https://www.youtube.com/watch?v=ZQEmSzrmBJc) and got identical results. There are additional web sites I used to check my code: https://goodcalculators.com/one-way-anova-calculator/ and https://www.socscistatistics.com/tests/anova/default2.aspx.

Assumption: I assumed that the number of groups and replicates were fixed, the data was normally distributed, and the critical value was determined beforehand and incorporated in the program.

Unix line command to create application:
g++ ANOVA.cpp -o ANOVA

Unix line command to run application:
./ANOVA 3 8 test.txt anova_out.txt

The './ANOVA' invokes the application, '3' is the number of groups, '8' is the number of replicates, 'test.txt' is the input data file and 'anova_out.txt' is the output file containing the results.

Output:
Here is the output to the terminal screen (which is exactly the same as that presented in the first youtube video I mentioned above).

```
6       36      8       64      9       81
6       36      7       49      8       64
5       25      6       36      9       81
8       64      5       25      9       81
7       49      6       36      7       49
6       36      7       49      8       64
3       9       8       64      6       36
4       16      6       36      9       81
```

```
sums[]45      53      65      163
means[]       5.625   6.625   8.125
sum2[]        271     359     537     1167
sum_sum_x_squared_div_nt=1107.04
dfb=2
nt=24
dfw=21
sum_square_x_div_n[s]       253.125       351.125       528.125
sum_sum_x_squared_div_nt2       1132.38
SSb=   25.3333
SSw=   34.625
MSb=   12.6667
MSw=   1.64881
F=     7.68231
SNK tests
1       2       1       0.453984        2.20272
1       3       2.5     0.453984        5.50681
2       3       1.5     0.453984        3.30408
```

The output file (anova_out.txt) yields the following information:

```
7.68231       1       3       5.625   8.125
```

The first number is the F-statistic, the second and third number shows that group 1 and 3 are significantly different by SNK. The last two numbers are the respective means of groups 1 and 3. The critical value for the ANOVA and SNK tests is 3.47 (for this example with 2, 21 degrees of freedom). Therefore, any F-statistic or SNK test having a value greater than the critical value is significant to $P<0.05$. The terminal output shows that groups 1 and 3 had an SNK value of 5.50681, meaning the groups were significantly different.

Tables for critical values are located here: https://www.statology.org/wp-content/uploads/2018/09/f.05.png.