Chapter 8

Post-ANOVA Multiple Comparison Tests: SNK

Although ANOVA can be used to determine if three or more means are different, it provides no information concerning where the difference lies. For example, if Ho: mean $_1$ = mean $_2$ = mean $_3$ is rejected, then there are three alternate hypotheses that can be tested:

mean $_1 \neq$ mean $_2 \neq$ mean $_3$, mean $_1 \neq$ mean $_2$ = mean $_3$, or mean $_1$ = mean $_2 \neq$ mean $_3$.

Methods have been constructed to test these possibilities, and they are termed multiple comparison post-tests. There are several of these sorts of tests, and statisticians do not agree on which one is most appropriate in any one situation. Only one of them is covered here – the Student-Newman-Keuls test – but it is one frequently favorably mentioned in statistics texts, and it works in basically the same way as other post-tests that appear in statistics software (e.g. Tukey-Kramer post test in MINITAB).

The Student-Newman-Keuls or SNK test considers a series of null hypotheses of the form Ho: mean $_A$ = mean $_B$ and H_a: mean $_A \neq$ mean $_B$ where the subscripts denote any possible pairs of groups. The first step is to arrange and number all the sample means in order of decreasing magnitude (from highest to lowest). Then pairwise differences between means ($\bar{x}_A - \bar{x}_B$) are used to calculate the test statistic called the "studentized range" which is abbreviated "q." The q value is calculated by dividing the pairwise difference between means by a standard error based upon the average variance of the two samples being considered (s_{AB}), or

$$q = \frac{(\bar{x}_A - \bar{x}_B)}{s_{AB}} \qquad [8.1]$$

The calculation of s_{AB} depends upon the sample size of the two groups being compared (A and B). If their sample size is equal ($n_A = n_B$) then

$$s_{AB} = \sqrt{\frac{\text{MS Error}}{n}} \qquad [8.2]$$

where *MS Error* is that computed in ANOVA of means and the *n* is the sample size of A or B. If the sample sizes are not equal ($n_A \neq n_B$) then

$$s_{AB} = \sqrt{\frac{\text{MS Error}}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} \qquad [8.3]$$

The probability of the q calculated using equation 9.1 is then obtained from a table of critical q values described below. As usual, a 0.05 significance level is used for rejection or acceptance of H_o: mean$_A$ = mean $_B$. Thus, if $P_{(q=0.05,p,df)} > 0.05$, then Ho is accepted, there is no significant difference, and the two means are assumed to come from the same population. Conversely, if $P_{(q=0.05,p,df)} < 0.05$, then Ho is rejected, there is a significant difference between A and B, and the two means are assumed to come from different populations. (Note that the " p" subscript to the q statistic is the number of means across which the comparison is being made, see below).

There are two important rules that must be followed in using SNK.

1) The order of comparison is critical. The largest mean *must* first be tested against the smallest mean, then the largest against the next smallest, and so on, until the largest mean has been compared with the second largest. Then the second largest *must* be compared with the smallest, next smallest, and so on. For example, the sequence of comparisons for four means ranked in descending order from 4 (highest) to 1 (lowest) would be as follows: 4 vs. 1, 4 vs. 2, 4 vs. 3, 3 vs. 1, 3 vs. 2, 2 vs. 1.

2) If there is no difference found between two means, it *must* be concluded that no difference exists between any means enclosed by these two. For example, if in comparing four means it is found that no difference exists between means 4 and 2, then it is also concluded that 4 does not differ from 3, and 3 does not differ from 2.

Finally, the multiple comparison tests including SNK are *not* nearly as powerful as ANOVA. These tests can indicate false significant differences between pairs of mean in cases where no such difference exists according to the more powerful ANOVA procedure. Thus, these tests are never used unless ANOVA indicates there is a significant difference between the means being considered. It is also possible for ANOVA to result in rejecting H_o: means are equal, and then the subsequent SNK procedure fails to detect differences between any pair of means. In which case, one should use another multiple comparison test. However, the most effective next step is to simply compare 95% confidence limits. In any case where the limits overlap it is likely that the samples come from the same population.

q Distribution

The test statistic, the studentized range, is a distribution of the range(s) of a varying number, *p*, of normally distributed items where the range is the difference between the highest and lowest values of the items and the *s* is an independent estimate of the standard deviation of the items.

Let's try another approach. Imagine, taking four random samples out of a normally distributed population, and then computing the means for the four samples, and the average standard deviation by simply obtaining the *MS Error* using ANOVA. The q value is then computed by dividing the range of difference between the highest and lowest mean. This is repeated a billion times using the same sample size for each of the four samples. The data are then plotted with the X-axis being the q value and the Y-axis being the number of times a particular q value is obtained. Obviously, 95% of the q values would fall below some specific q value and 5% would be higher. This specific q value, the q value at the P=0.05 level, is then packaged into a table, and is used to determine whether to accept or reject the null hypothesis in the SNK (and other) tests.

Such a Table is provided in the Tablepack. Some of that Table is reproduced here (next page). This table has two arguments: across the top is *p*, which is the number of means being compared, and along the left margin is *df*, the degrees of freedom of the standard deviation. The latter argument, the *df*, is the degrees of freedom for the Error term in ANOVA. The magnitude of the p value changes as a function of the specific comparison. If four means were being compared, then the p would be 4 if the highest and lowest were being compared while it would be 2 given comparison between two adjacently ranked means. It would also be 3 when comparing the highest to the next to lowest, or the next highest to the lowest.

Table 8.1 Critical values at P=0.05 of q (known also as "studentized range"). The body of the table contains the q value where the top row is p, the number of items (e.g., means) over which the range is computed, and the left hand column is the degrees of freedom (e.g. df of MS Error).

df	p=2	3	4	5	6	7	8	9	10
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254

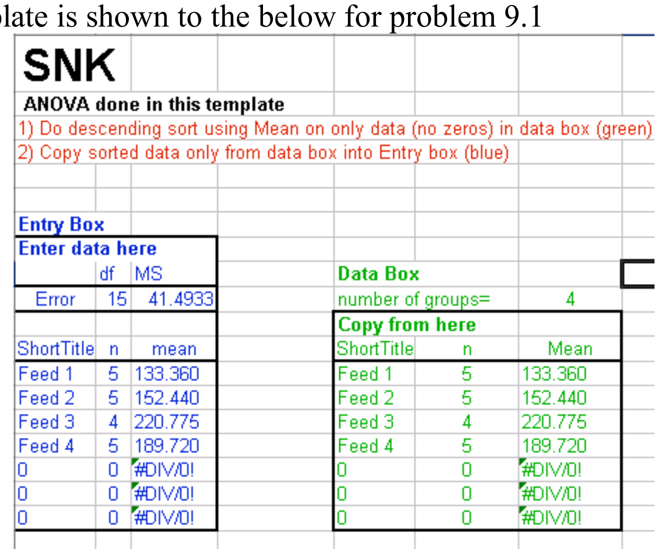
If a calculated value is higher than the appropriate value in the table then the Ho: $mean_A = mean_B$ is rejected. Conversely, if the calculated value is lower than the tabular value then the Ho is accepted. As examples, given *df* = 13 the following statistical decisions based on q test statistics would result.

q	p	decision	q	p	decision
35.23	6	reject	4.233	3	reject
3.523	6	accept	4.233	4	reject
4.695	6	reject	4.233	5	accept

Calculation of SNK Using EXCEL

The SNK Template is the second worksheet of the EXCEL "ANOVA&SNK.XLT" template that was previously used. The SNK template is shown to the below for problem 9.1 before the user had done the required sorting of means. Values from the first worksheet of the "ANOVA&SNK.XLT" template, titled ANOVA, are automatically put into the SNK template. The values are the *MS Error* and *df*, and the short title, sample size, and for each of up to seven data sets from the ANOVA worksheet. These data are put into two boxes – a "Entry Box," which is blue, and a "Data Box," which is green. These may not appear in color here if this document was not printed on a color printer.

As indicted by the instructions, the user has to do a descending sort by the means of only the data in the Data Box.



SNK

ANOVA done in this template

1) Do descending sort using Mean on only data (no zeros) in data box (green)

2) Copy sorted data only from data box into Entry box (blue)

Entry Box

Enter data here

	df	MS			
Error	15	41.4933			

Data Box

number of groups= 4

Copy from here

ShortTitle	n	mean	ShortTitle	n	Mean
Feed 1	5	133.360	Feed 1	5	133.360
Feed 2	5	152.440	Feed 2	5	152.440
Feed 3	4	220.775	Feed 3	4	220.775
Feed 4	5	189.720	Feed 4	5	189.720
0	0	#DIV/0!	0	0	#DIV/0!
0	0	#DIV/0!	0	0	#DIV/0!
0	0	#DIV/0!	0	0	#DIV/0!

It is important that *only* data be selected before doing the sort. For example, if only three data sample have been analyzed then it is important to select only the appropriate three rows and not include any additional rows from the data box. Sort can be found under Tools|Sort. After sorting, these data are then copied and pasted in the appropriate rows in the Entry Box. As soon as this is done, the analysis is complete.

The output of the SNK template should be self-explanatory. The Description or short titles from ANOVA and the two entry boxes are used to indicate what comparisons are being made (Columns A-C). The difference in means and the standard error are displayed to orient the user (Columns D, E). Both of these values should be positive; if not, the typical reason is that the sort was not done or not done properly. The calculated test statistic is the crucial one to observe (Column G). It is compared with the tabular q value that is shown in the adjacent column (Column H). Remember this tabular q value is the P=0.05 value and is imported from the 3rd worksheet in the ANOVA&SNK.XLT template. The statistical decision of acceptance or rejection of the null hypothesis is obtained by simply comparing the calculated q to the tabular value (Columns G vs. H). If the calculated value, Column G, is greater than the tabular value (H), then the probability that sample A and B come from the same population is less than 0.05. Thus, the Ho is rejected. Conversely, a calculated q lower than tabular q results in a probability greater than 0.05 that the two samples come from the same population; thus, the Ho is accepted.

Entry Box				Data Box		
Enter data here						
	df	MS				
Error	15	41.4933		number of groups=		4
				Copy from here		
ShortTitle	n	mean		ShortTitle	n	Mean
Feed 3	4	220.775		Feed 3	4	220.775
Feed 4	5	189.720		Feed 4	5	189.720
Feed 2	5	152.440		Feed 2	5	152.440
Feed 1	5	133.360		Feed 1	5	133.360
0	0	#DIV/0!		0	0	#DIV/0!
0	0	#DIV/0!		0	0	#DIV/0!
0	0	#DIV/0!		0	0	#DIV/0!

Comparison					# of	Calculated	Critical	
A		B	diff of means	SE	means	q	q	Concl.
***	vs	***	***	***	7	***	***	***
***		***	***	***	6	***	***	***
***		***	***	***	5	***	***	***
Feed 3		Feed 1	87.415	3.055	4	28.609	4.076	reject
Feed 3		Feed 2	68.335	3.055	3	22.365	3.674	reject
Feed 3		Feed 4	31.055	3.055	2	10.164	3.014	reject

Example Problem

(Problem 8.1)

I. STATISTICAL STEPS

A. Statement of Ho and Ha

Ho: mean 1 = mean 2 = mean 3 = mean 4

Ha: not precisely known since differences could exist anywhere

B. Statistical test – ANOVA

F value computed to test Ho: MS factor ≤ MS error   Ha: MS factor > MS error

C. Computation of test statistic

The template "ANOVA&SNK.XLT" was used to determine if there are significant differences between the means. The F value calculated was $F_{[\alpha(1)=0.05,3,15]}=164.38$

	A	B	C	D	E	F
1	ANOVA		Must enter: data into b20:h100			
2		SSX	SSX2	S(SX)^2/n	CT	
3	Sums	3260.700	580671.410	580049.011	559587.605	
4						
5	Source	SS	df	MS	F	P
6	Factor	20461.406	3	6820.469	164.375	.000
7	Error	622.400	15	41.493		
8	Total	21083.805	18			
9	Description	Feed 1:	Feed 2:	Feed 3:	Feed 4:	
10	n	5	5	4	5	
11	mean	133.360	152.440	220.775	189.720	
12	SE	3.045	2.862	3.053	2.840	
13	Low 95% CL	124.907	144.493	211.059	181.835	
14	High 95% CL	141.813	160.387	230.491	197.605	
15	Low	125.300	145.800	212.300	182.800	
16	High	143.100	162.700	224.600	198.600	
17						
18	S(SX)^2/n	88924.448	116189.768	194966.403	179968.392	

D. Determination of the P of the test statistic

$P(F_{[\alpha(1)=0.05,3,15]}=164.38)<<0.05$

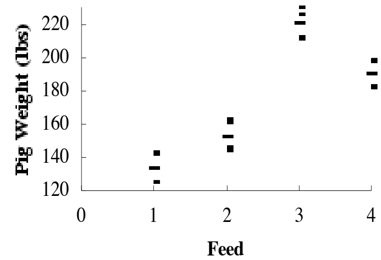
E. Statistical Inference

Reject Ho that MS factor ≤ MS. MS factor is significantly greater than MS error. Therefore, conclude that means are significantly different.

Biological Inference

There was a significant difference in pig body weights when raised on different diets.

Now we need to determine where the differences lie. Inspection of the data at this point using a modified Dice-Lerass graph generated in the mean template is quite helpful. Based on the 95% confidence intervals one would expect the feeds to be significantly different from one another with Feed 3 being the highest, followed by Feed 4, then Feed 2 with Feed 1 giving the lowest pig weight. However, it could be that Feeds 1 and 2 are not significantly different.



F. 2nd Statement of H_0

Ho_2: $mean_A = mean_B$

Ha_2: $mean_A \neq mean_B$

G. 2nd Statistical test - SNK

Use SNK procedure where A first becomes the largest mean (Feed 3), the first B is the smallest mean (Feed 1), then the second smallest mean (Feed 2), then the third smallest (Feed 4). Then A becomes the next largest (Feed 4) and the whole sequence is done as before.

H. Computation of test statistic q

The template "ANOVA&SNK.XLT" was used to make these calculations. The calculated q's are in column G as indicted by its heading. For all the comparisons the calculated q's are greater than the tabular q's.

	A	B	C	D	E	F	G	H	I
1	SNK								
2	2) Copy sorted data only from data box into Entry box (blue)						ANOVA done elsewhere		
3	1) Do descending sort using Mean on only data (no zeros) in data box (						1) Import Error df & MS into en		
4	ANOVA done in this template						2) In green data box, place numb		
5							3) Do descending sort using Mear		
6							4) Copy sorted data only from dat		
7	Entry Box								
8	Enter data here								
9		df	MS		Data Box				
10	Error	15	41.4933		number of groups=		4		
11					Copy from here				
12	ShortTitle	n	mean		ShortTitle	n	Mean		
13	Feed 3:	4	220.775		Feed 3:	4	220.775		
14	Feed 4:	5	189.720		Feed 4:	5	189.720		
15	Feed 2:	5	152.440		Feed 2:	5	152.440		
16	Feed 1:	5	133.360		Feed 1:	5	133.360		
17									
18									
19	Comparison					# of	Calculated	Critical	
20	A		B	diff of means	SE	means	q	q	Concl.
21	Feed 3:		Feed 1:	87.415	3.055	4	28.609	4.076	reject
22	Feed 3:		Feed 2:	68.335	3.055	3	22.365	3.674	reject
23	Feed 3:		Feed 4:	31.055	3.055	2	10.164	3.014	reject
24									
25	Feed 4:		Feed 1:	56.360	2.881	3	19.564	3.674	reject
26	Feed 4:		Feed 2:	37.280	2.881	2	12.941	3.014	reject
27									
28	Feed 2:		Feed 1:	19.080	2.881	2	6.623	3.014	reject
29									

I. 2nd Determination of the P of the test statistics

a) feed 3 vs feed 1: $P(q_{[4,15]} = 28.609) < 0.05$

b) feed 3 vs feed 2: $P(q_{[3,15]} = 22.365) < 0.05$

c) feed 3 vs feed 4: $P(q_{[2,15]} = 10.164) < 0.05$

d) feed 4 vs feed 1: $P(q_{[3,15]} = 19.564) < 0.05$

e) feed 4 vs feed 2: $P(q_{[2,15]} = 12.941) < 0.05$

f) feed 2 vs feed 1: $P(q_{[2,15]} = 6.623) < 0.05$

J. 2nd Statistical Inference

Reject Ho and accept Ha in all cases. Thus, feed 3 > feed 4 > feed 2 > feed 1

II. BIOLOGICAL INTERPRETATION

1. Each feed resulted in significantly different body weights.
2. Feed 3 produced pigs with the greatest body weight.

3. The remaining feeds produced significantly lower body weights ranking from feed 1 (lowest), feed 2 and feed 4.

Results

There was a significant difference in pig body weight based on different feeds ($F_{0.05,3,15} = 164.4$, $P < 0.001$) with feed 3 producing the heaviest body weight (Table 1). All feeds produced significantly different body weights of pigs (q's = 6.62 – 28.6, all $P < 0.05$). The remaining feeds produced significantly lower body weights ranking from feed 1 (lowest), to feed 2 to feed 4 (Fig. 1). The differences in body weight between each feed average about 18% with the top feed resulting in 65% greater weight gain compared to the poorest diet (Feed 3 vs. 1).

Table 1. Mean, sample size (n), standard error (SE), 95% confidence interval (CL), and range of pig body weights (pounds) after being raised on four different feeds.

Feed	Mean	n	SE	95% CL	Range
1	133.36	5	3.04	124.91 - 141.81	125.3 - 143.1
2	152.44	5	2.86	144.49 - 160.39	145.8 - 162.7
3	220.78	4	3.05	211.06 - 230.29	212.3 - 225.8
4	189.72	5	2.84	181.84 - 197.61	182.8 - 198.6

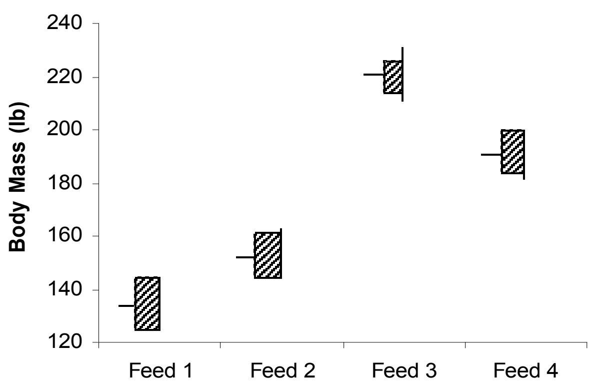


Figure 1. Pig body weights (pounds) after being raised on four different diets. Error bars represent 95% confidence intervals.

Problem Set – ANOVA and SNK

8.1) Pigs were assigned at random to one of four experimental groups with each group fed a different diet. Analyze and interpret the pigs' body weights (pounds) after being raised on these diets.

Feed 1:	133.8	125.3	143.1	128.9	135.7
Feed 2:	151.2	149.0	162.7	145.8	153.5
Feed 3:	225.8	224.6	220.4	212.3	
Feed 4:	193.4	185.3	182.8	188.5	198.6

8.2) The pH's of water samples from four different ponds were determined. Analyze and interpret.

Pond 1:	7.68	7.70	7.72	7.73	7.73	7.76	7.78	7.71
Pond 2:	7.69	7.70	7.71	7.73	7.74	7.74	7.78	7.81
Pond 3:	7.74	7.75	7.77	7.78	7.80	7.81	7.81	
Pond 4:	7.71	7.71	7.74	7.79	7.84	7.85	7.87	7.91

8.3) Body temperatures (^{o}C) of individuals from four species of mammals were recorded to determine if there were any differences. Analyze and interpret.

Species A:	37.9	37.6	38.1	38.2	37.8	37.6	37.8	37.6
Species B:	37.7	37.6	37.9	37.8	37.5	37.6	37.8	37.4
Species C:	37.7	37.8	38.0	38.0	37.8	37.5	37.9	37.5
Species D:	37.9	38.1	38.4	38.2	37.6	37.7	37.5	37.3

8.4) Weights (kg) of food consumed per day by adult deer were collected at different times of the year. Analyze and interpret.

February:	4.7	4.9	5.0	4.8	4.7	
May:	4.6	4.4	4.3	4.4	4.1	4.2
August:	4.8	4.7	4.6	4.4	4.7	4.8
November:	4.9	5.2	5.4	5.1	5.6	

8.5) Amounts of free amino acids [μmoles·(g dry weight)$^{-1}$] were determined for four species of marine arthropods. Analyze and interpret.

Species 1:	431.1	440.2	443.2	445.5	448.6	451.2	
Species 2:	477.1	479.0	481.3	487.8	489.6		
Species 3:	385.5	387.9	389.4	399.1	403.7		
Species 4:	366.8	269.8	371.4	373.2	377.2	379.4	381.3

8.6) The bird *Singus sweetus* is divided into three populations by geographic barriers. An investigator wishes to know if evolution has caused any significant morphological differences in the three populations and measured the wing length (mm) of samples of birds.

A: 69, 72, 71, 72, 72, 73, 71, 68, 69, 75
B: 71, 75, 72, 72, 74, 76, 73, 70, 75, 76, 73, 74
C: 71, 74, 72, 70, 75, 74, 73, 68, 73, 75, 74, 74, 72, 70, 74