Open lecture on statistics

ISSN 2234-7658 (print) / ISSN 2234-7666 (online)
http://dx.doi.org/10.5395/rde.2014.39.1.74

RDE Restorative Dentistry & Endodontics

Analysis of variance (ANOVA) comparing means of more than two groups

Hae-Young Kim

Department of Dental Laboratory Science and Engineering, College of Health Science & Department of Public Health Science, Graduate School & BK21+ Program in Public Health Sciences, Korea University, Seoul, Korea

***Correspondence to**
Hae-Young Kim, DDS, PhD.
Associate Professor,
Department of Dental Laboratory Science & Engineering, Korea University College of Health Science, San 1 Jeongneung 3-dong, Seongbuk-gu, Seoul, Korea 136-703
TEL, +82-2-940-2845; FAX, +82-2-909-3502, E-mail, kimhaey@korea.ac.kr

Mean values obtained from different groups with different conditions are frequently compared in clinical studies. For example, two mean bond strengths between tooth surface and resin cement may be compared using the parametric Student's *t* test when independent groups are subjected to the comparison under the assumptions of normal distribution and equal variances (or standard deviation). In a condition of unequal variances we may apply the Welch's *t* test as an adaptation of the *t* test. As the nature and specific shape of distributions are predetermined by the assumption, the *t* test compares only the locations of the distribution represented by means, which is simple and intuitive. The *t* statistic is the ratio of mean difference and standard errors of the mean difference.

Even when more than two groups are compared, some researchers erroneously apply the *t* test by implementing multiple *t* tests on multiple pairs of means. It is inappropriate because the repetition of the multiple tests may repeatedly add multiple chances of error, which may result in a larger α error level than the pre-set α level. When we try to compare means of three groups, A, B, and C, using the *t* test, we need to implement 3 pairwise tests, i.e., A vs B, A vs C, and B vs C. Similarly if comparisons are repeated k times in an experiment and the α level 0.05 was set for each comparison, an unacceptably increased total error rate of $1-(0.95)^k$ may be expected for the total comparison procedure in the experiment. For a comparison of more than two group means the one-way analysis of variance (ANOVA) is the appropriate method instead of the *t* test. As the ANOVA is based on the same assumption with the *t* test, the interest of ANOVA is on the locations of the distributions represented by means too. Then why is the method comparing several means the 'analysis of variance', rather than 'analysis of means' themselves? It is because that the relative location of the several group means can be more conveniently identified by variance among the group means than comparing many group means directly when number of means are large.

The ANOVA method assesses the relative size of variance among group means (between group variance) compared to the average variance within groups (within group variance). Figure 1 shows two comparative cases which have similar 'between group variances' (the same distance among three group means) but have different 'within group variances'. When the between group variances are the same, mean differences among groups seem more distinct in the distributions with smaller within group variances (a) compared to those with larger within group variances (b). Therefore the ratio of between group variance to within group variance is of the main interest in the ANOVA.

Table 1 displays an artificial data of bond strength according to different resin types and Table 2 shows the result of the one-way ANOVA. The 'SSB' represents the sum of squares between groups which is the variation of group means from the total grand mean, and the mean of squares between groups (MSB) is subsequently obtained by dividing SSB with degrees of freedom. The 'SSW' represents sum of squares within

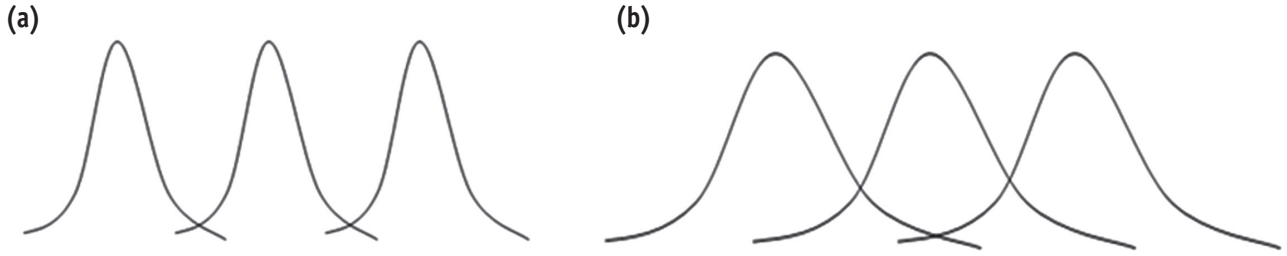


Figure 1. Distributions with the same between group variance. (a) smaller variance within groups; (b) larger variance within groups.

Table 1. Measurements of bonding strength according to three different types of resin (artificial data)

	A	B	C
	19.7, 20.1, 21.3, 23.5, 9.3 27.1, 11.6, 12.2, 15.9, 17.0 17.2, 18.4, 19.8, 23.4, 28.0	23.0, 24.5, 24.6, 27.1, 12,0 27.8, 12.8, 16.2, 19.8, 22.4 23.6, 25.3, 27.9, 34.6, 35.2	21.6, 25.5, 25.9, 30.7, 33.0 16.5, 22.7, 24.2, 26.2, 28.4 28.5, 30.7, 32.2, 33.8, 34.5
Mean	18.97	23.80	27.62
SD	5.38	6.72	5.09

Table 2. One-way ANOVA table

Source of variation	Sum of squares	df	Mean square	F	*p* value
Between groups	563.7 (SSB)	2 (p-1)	281.8 [SSB/(p-1) = MSB]	8.4 (MSB/MSW)	0.001
Within groups	1399.7 (SSW)	42 (N-p)	33.3 [SSW/(N-p) = MSW]		
Total	1963.4 (SST)	44 (N-1)	SST/(N-1) = s^2		

groups which is the sum of squared deviations from the group means and individual observations because the equal variances in all the groups were already assumed. The mean of square within groups (MSW) is subsequently obtained by dividing SSW with degrees of freedom, in the same way. The ratio of MSB and MSW determines the degree of how relatively greater the difference is between group means (between group variance) compared to within group variance. If the ratio is greater than expected by chance we may think not all the group means are the same which means that at least one mean is substantially different. As the result is interpreted about the whole set of groups, it is called as a global or overall test. The ratio of MSB and MSW is known to follow the F distribution. Therefore, to get a statistical conclusion we may compare the F value calculated from the observed data with the critical value at an α error level of 0.05 in the F table.

$$F\ (\text{observed}) = \frac{MSB}{MSW} = \frac{\text{variance between groups}}{\text{variance within groups}}$$

Larger F value implies that means of the groups are greatly different from each other compared to the variation of the individual observations in each groups. Larger F value than the critical value supports that the differences between group means are larger than what would be expected by chance. In this example the critical F value is 3.23 in the F table when the degrees of freedom of numerator and denominator are 2 and 42 respectively at the α error level 0.05. As the observed F value 8.4 is larger than the critical value, the result in Table 2 may be interpreted as statistically significant difference among the means of the groups at the α error level 0.05. The result suggests to rejection of the null hypothesis that all the group means are the same, and coincidently supports that at least one group mean differs from other group means.

If any significant difference is detected by the 'overall F test' above, we need to examine what specific pair of group means shows difference and what pairs do not. While many different kinds of *post-hoc* multiple comparison procedures have been proposed, the choice needs to be made according to the specific research question. One basic method is implementing multiple pairwise *t* tests using the common variance as MSW and appropriately adjusting α error level to get the optimal α error level for the whole experiment. For example, the Bonferroni correction is a simple method that adjusts comparison-wise type α error level as the usual experiment-wise α error level divided by the number of comparisons, e.g., 0.05/k. However, caution is needed because in some situations the Bonferroni correction may be substantially conservative that actual experiment-wise α error level applied may be lower than 0.05. Tukey's HSD, Schaffe method, and Duncan multiple range test are more frequently preferred methods for the multiple comparison procedures. Table 3 displays the analysis results by both the ANOVA and multiple comparison procedure. We usually need to report the *p*-value of overall F test and the result of the *post-hoc* multiple comparison. Table 3 shows that 'C' resin has the highest bond strength and 'A' resin shows the lowest.

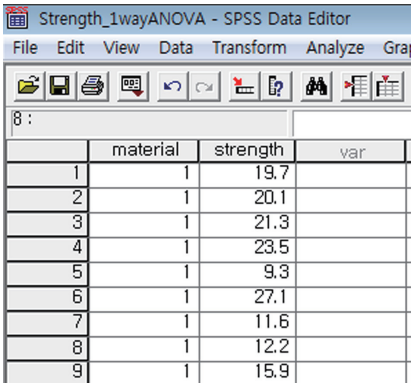
Table 3. Comparative mean bond strength according to different types of resin (display of ANOVA results)

	Resin types			*p* value
	A	B	C	
Bonding strength, mean ± SD	18.97 ± 5.38[a]*	23.80 ± 6.72[ab]	27.62 ± 5.09[b]	0.001

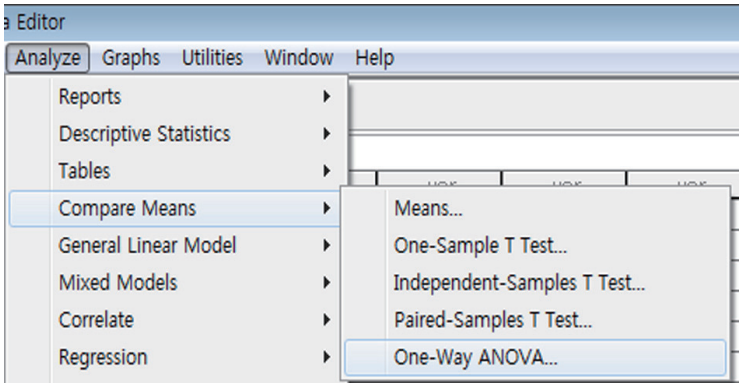
*Different superscripts mean statistically different.

The comparison of more than two group means by ANOVA using the SPSS statistical package (SPSS Inc., Chicago, Il) according to the following procedures:

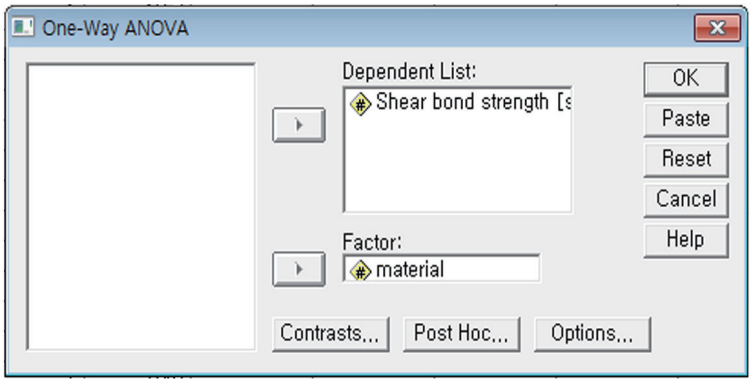
(a) Input data



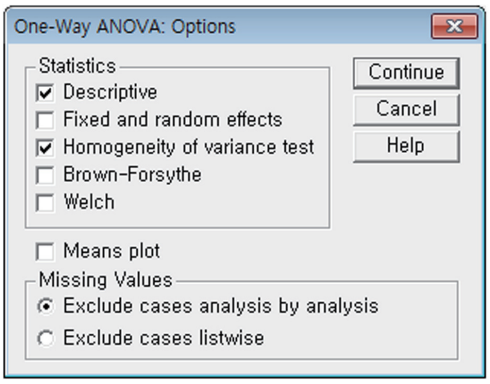
(b) Analysis – Compare means –One-way ANOVA



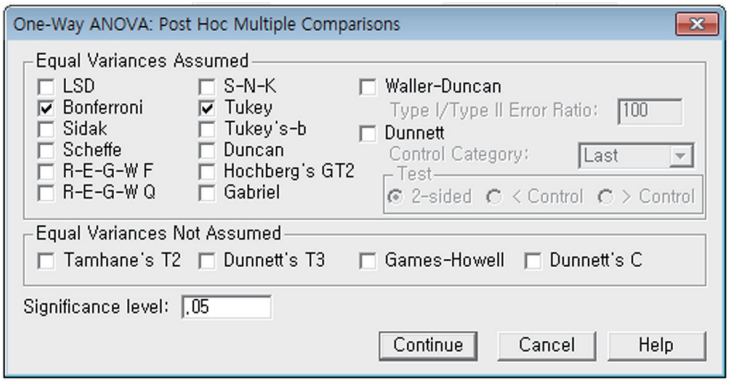
(c) Write variables on the windows



(d) Select options



(e) *Post-hoc* multiple comparison



(f) Homogeniety of variances

Test of Homogeneity of Variances

Shear bond strength

Levene Statistic	df1	df2	Sig.
.234	2	42	.792

(g) ANOVA table

ANOVA

Shear bond strength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	563.692	2	281.846	8.457	.001
Within Groups	1399.717	42	33.327		
Total	1963.409	44			

(h) Results of *post-hoc* multiple comparison

Multiple Comparisons

Dependent Variable: Shear bond strength

	(I) material	(J) material	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1	2	-4.8273	2.1080	.068	-9.949	.294
		3	-8.6500*	2.1080	.001	-13.771	-3.529
	2	1	4.8273	2.1080	.068	-.294	9.949
		3	-3.8227	2.1080	.178	-8.944	1.299
	3	1	8.6500*	2.1080	.001	3.529	13.771
		2	3.8227	2.1080	.178	-1.299	8.944
Bonferroni	1	2	-4.8273	2.1080	.081	-10.084	.429
		3	-8.6500*	2.1080	.001	-13.907	-3.393
	2	1	4.8273	2.1080	.081	-.429	10.084
		3	-3.8227	2.1080	.231	-9.079	1.434
	3	1	8.6500*	2.1080	.001	3.393	13.907
		2	3.8227	2.1080	.231	-1.434	9.079

*. The mean difference is significant at the .05 level.

(i) Homogeneous subsets

Homogeneous Subsets

Shear bond strength

	material	N	Subset for alpha = .05: 1	2
Tukey HSD[a]	1	15	18.968	
	2	15	23.795	23.795
	3	15		27.618
	Sig.		.068	.178

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 15.000.

Reference

1. Rosner B. Fundamentals of biostatistics. Belmont: Duxbury Press; 2006. p557-581.