

Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors

Hae-Young Kim

Department of Dental Laboratory
Science and Engineering, College
of Health Science & Department
of Public Health Science, Graduate
School, Korea University, Seoul,
Korea

When we have a continuous outcome e.g., bonding strength and two categorical explanatory variables such as 4 different resin types and 2 different curing light sources, usually we consider applying the two-way ANOVA for analyzing the relationships. However because implementing the two-way ANOVA is relatively complicated, some clinical researchers prefer to apply the one-way ANOVA for one factor on each level of the other factor, repeatedly. They often insist that they are interested only in one factor (e.g., manipulation methods) and are not interested in the other factor (e.g., brands), claiming that the one-way ANOVA is the more appropriate strategy. Even though the trial with a variety of brands may be considered as a simple way of generalization among various brand types, possible different effects of materials of different brands can never be detected by the one-way ANOVA. Actually materials of different brands may have slightly different ingredient compositions which may elicit different effects on the other factor. Application of the one-way ANOVA cannot detect the possible interaction between two explanatory categories.

Table 1 shows a data of bonding strength of four types of resin (A, B, C and D) on the teeth surface with a simultaneous use of two different curing light sources (Halogen, LED). The highest overall bonding strength is found on resin D followed by resin C and resin B showing insignificant differences (see the superscript a, b, and c). Considering cases using the 'Halogen,' resin D is the strongest among four resin types, while resin C shows the highest value when the 'LED' was used. This explicitly shows that the effects of different resin types are not following a similar trend according to different levels of curing methods, 'Halogen' or 'LED'. Figure 1a shows that trend of (descriptive) mean bonding strengths for resin types are changing according to the levels of curing methods.

Interaction model or main effect (no-interaction) model?

When we have a quantitative continuous outcome and two categorical explanatory variables, we may consider two kinds of relationship between two categorical variables, which could be typically seen in the Figures 1b and 1c. The Figure 1c shows that the relative effect of each level in the material category doesn't change with different levels of curing methods, which means an additive relationship between the two categorical variables, i.e. the second categorical variable takes a role of adding a uniform effect on the relationship between the outcome and the first categorical variable. In this relationship we can distinguish effect of one factor from that of the other factor. This type of model is called a **main effect model** or **no-interaction model**. However, Figure 1b shows that the effect of material depends on the levels of the curing methods and we cannot tell effect of one factor separately, i.e. which light source does produce stronger bonding? This is called an **interaction model** because an interaction relationship is included. We may see that the interaction model could

*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

Table 1. Measurements of bonding strength (Mpa) according to four different types of resin and two curing methods

Curing Lights	Resin types				Mean ± SD
	A	B	C	D	
Halogen	14.5, 15.2, 17.4, 17.5, 19.2, 19.7, 20.1, 21.3, 23.5, 9.3	11.8, 13.3, 19.2, 21.3, 22.2, 23.0, 24.5, 24.6, 27.1, 12	14.5, 15.0, 18.6, 19.6, 21.0, 21.6, 25.5, 25.9, 30.7, 33	35.5, 35.7, 36.3, 37.3, 39.9, 40.9, 41.0, 44.5, 44.7, 47.2	
Mean ± SD	17.8 ± 4.0 ^{α*}	19.9 ± 5.6 ^α	22.5 ± 6.2 ^α	40.3 ± 4.2 ^β	25.1 ± 10.3
LED	27.1, 11.6, 12.2, 15.9, 17.0, 17.2, 18.4, 19.8, 23.4, 28	27.8, 12.8, 16.2, 19.8, 22.4, 23.6, 25.3, 27.9, 34.6, 35.2	16.5, 22.7, 24.2, 26.2, 28.4, 28.5, 30.7, 32.2, 33.8, 34.5	17.3, 19.2, 19.5, 20.5, 20.7, 22.2, 25.8, 29.0, 29.2, 35.1	
Mean ± SD	19.1 ± 5.6 ^A	24.6 ± 7.3 ^{AB}	27.8 ± 5.6 ^B	23.9 ± 5.7 ^{AB}	23.8 ± 6.6
Total Mean ± SD	18.4 ± 4.8 ^a	22.2 ± 6.8 ^{ab}	25.1 ± 6.3 ^b	32.1 ± 9.7 ^c	24.5 ± 8.6

* Different alphabets mean significantly different values at a type one error rate of 0.05.

By the independent *t*-test: *p*-value (Halogen vs. LED) = 0.498.

By the One-way ANOVA (A vs. B vs. C vs. D): *p*-value (all methods) < 0.001, *p*-value (only Halogen) < 0.001, *p*-value (only LED) = 0.025.

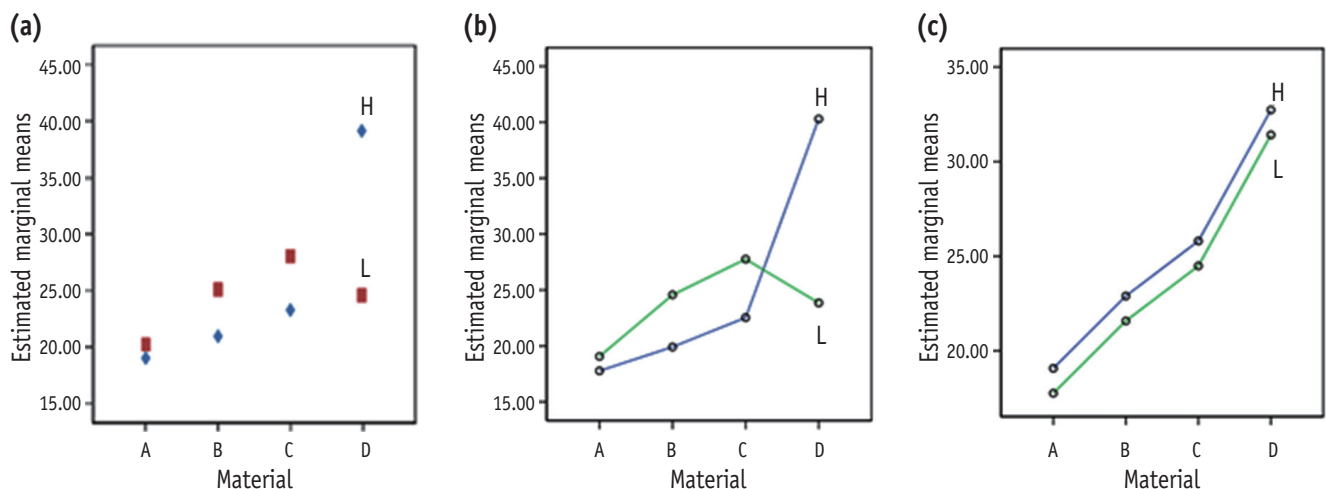


Figure 1. Graphs for bonding strength by resin materials (A, B, C, and D) displayed as separated lines of different curing light sources (Halogen & LED): (a) Descriptive means; (b) Estimated means by the model with the interaction term; (c) Estimated means by the model without the interaction term between two factors.

easily reproduce the actual relationship among descriptive means, as seen in Figure 1a. Therefore generally the first step in application of the two-way ANOVA is fitting the interaction model, specified as the “Full factorial model” (**Part A, d-1**, below) and test the significance of the interaction term. The resulting ANOVA table of two-way ANOVA interaction model is shown in Table 2 and **g-1** (below) and we could find the interaction term (Light*Resin) is statistically significant at an alpha level of 0.05 ($p < 0.001$). As an effect of a level of one variable depends on levels of the other variable, we cannot separate the effects of two variables, neither independent effect of resin types nor independent effect of curing light. The levels of two categorical variables should be combined into a total of eight categories (2 levels of Light * 4 levels of Resin) and the post-hoc multiple comparisons may be implemented among the eight categories as if they consist of one (combined)

Table 2. The ANOVA table from the two-way ANOVA considering two factors with the interaction term (correct)

Source	Sum of squares	df	Mean square	F	p-value
Corrected model	3602.1	7	514.6	16.4	< 0.001
Light	34.7	1	34.7	1.1	0.297
Resin	1997.0	3	665.7	21.2	< 0.001
Light*Resin	1570.4	3	523.5	16.7	< 0.001
Error	2261.6	72	31.4		
Corrected total	5863.7	79			

Table 3. Comparative mean bonding strength under the two-way ANOVA model with the interaction model (correct, g-1)

Curing Light	Resin type, Mean ± SD			
	A	B	C	D
Halogen	17.8 ± 4.0 ^{a*}	19.9 ± 5.6 ^{ab}	22.5 ± 6.2 ^{abc}	40.3 ± 4.2 ^d
LED	19.1 ± 5.6 ^{ab}	24.6 ± 7.3 ^{bc}	27.8 ± 5.6 ^c	23.9 ± 5.7 ^{bc}

* Different alphabets mean significantly different values at a type one error rate of 0.05.
 p-value (model) < 0.001; p-value (light) = 0.297; p-value (resin) < 0.001; p-value (resin*light) < 0.001; R-square = 0.61.

Table 4. Comparative mean bonding strength under the two-way ANOVA using the main effect model (no-interaction model, incorrect[†], g-2)

Curing Light	Resin type, Mean ± SD				p-value (resin)
	A	B	C	D	
Halogen	17.8 ± 4.0 ^{a*}	19.9 ± 5.6 ^{ab}	22.5 ± 6.2 ^b	40.3 ± 4.2 ^c	0.412
LED	19.1 ± 5.6 ^a	24.6 ± 7.3 ^{ab}	27.8 ± 5.6 ^b	23.9 ± 5.7 ^c	
p-value (light)	< 0.001				

[†] This table simply shows how to report results of the main effect model, only for the purpose of illustration. Actually the Table 3 displays the correct results which reflect given data well.

* Different alphabets mean significantly different values at a type one error rate of 0.05.
 p-value (model) < 0.001; R-square = 0.35.

factor (shown in **h-1**①; variable name = 'light_mat'). Table 3 provides the complete report of analyses results from the interaction model as well as post-hoc multiple comparisons. The plot in Figure 1 may be displayed by requesting plots (**e**, below). The underlying assumptions of the two-way ANOVA model are the same with those of the one-way ANOVA, normal distribution of outcomes and equal variances. The assumption of normality should be checked in an exploratory procedure and the assumption of equal variances may be tested as the homogeneity test for the null hypothesis of equal variances for all groups, as shown in the procedure **f**, below.

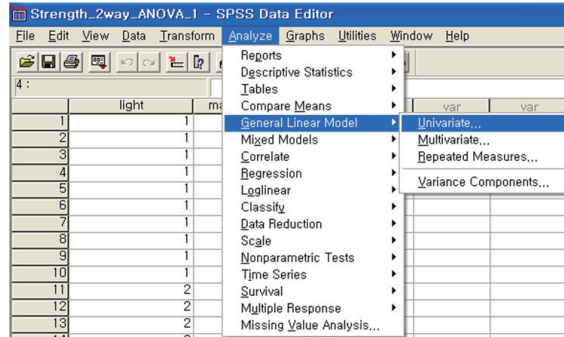
On the other hand, if we have an insignificant interaction term, different from the results above, we consider a main effect (no-interaction) model as shown in **Part B** below, which may be actually incorrect in modeling this data. Table 4 shows the comprehensive results of the analyses based on the main effect model, although the model is actually inadequate because it doesn't fit the data well. The superscripts represent statistical differences among levels of Resin types only because the effect of Light was insignificant ($p = 0.412$). You may add superscripts of upper cases to represent statistical differences in the levels of the Light variable if the Light is significant.

The two-way ANOVA with interaction term using the SPSS statistical package (SPSS Inc., Chicago, IL, USA) as following procedures:

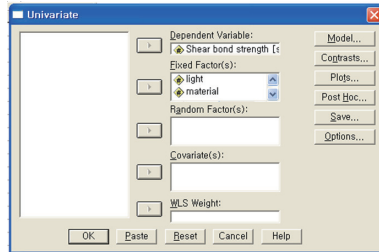
(a) Input data

	light	material	strength	var
1	1	1	14.49	
2	1	1	15.20	
3	1	1	17.43	
4	1	1	17.50	
5	1	1	19.20	
6	1	1	19.70	
7	1	1	20.10	
8	1	1	21.30	
9	1	1	23.50	
10	1	1	9.32	
11	2	1	27.10	
12	2	1	11.60	

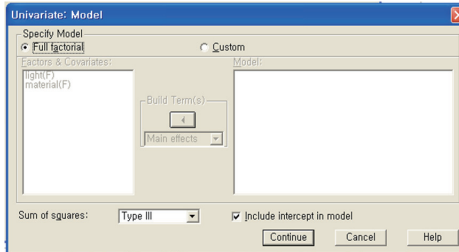
(b) Analysis – General Linear Model – Univariate



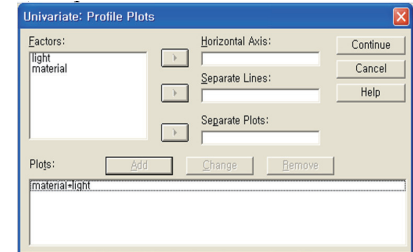
(c) Write variables on the windows



(d-1) interaction Model

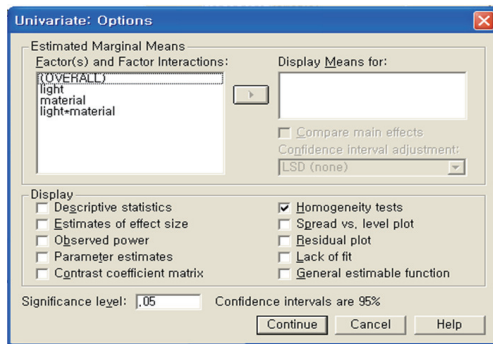


(e) Request Plot



(f) Assess assumption: equal variances

① Request homogeneity test



② Result of homogeneity test

Levene's Test of Equality of Error Variances^a

Dependent Variable: Shear bond strength

F	df1	df2	Sig.
.775	7	72	.610

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+light+material+light * material

Part A: Interaction model

(g-1) ANOVA Table: interaction model

Tests of Between-Subjects Effects

Dependent Variable: Shear bond strength

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3602,103 ^a	7	514,586	16,382	,000
Intercept	47894,642	1	47894,642	1524,759	,000
light	34,716	1	34,716	1,105	,297
material	1996,998	3	665,666	21,192	,000
light * material	1570,389	3	523,463	16,665	,000
Error	2261,612	72	31,411		
Total	53758,357	80			
Corrected Total	5863,715	79			

a. R Squared = .614 (Adjusted R Squared = .577)

(h-1) Post-hoc multiple comparisons: with significant interaction

- ① Re-arrangement of data
- ② Multiple comparison
- ③ Homogeneous subsets

	light	material	strength	light_mat
1	1	1	14.49	11
2	1	1	15.20	11
3	1	1	17.43	11
4	1	1	17.50	11
5	1	1	19.20	11
6	1	1	19.70	11
7	1	1	20.10	11
8	1	1	21.30	11
9	1	1	23.50	11
10	1	1	9.32	11
11	2	1	27.10	21
12	2	1	11.60	21

Homogeneous Subsets

Shear bond strength

light_mat	N	Subset for alpha = .05			
		1	2	3	4
Halogen & A	10	17,740			
LED & A	10	19,0600			
Halogen & B	10	19,9090			
Halogen & C	10	22,5320			
LED & D	10	22,5320	22,5320		
LED & B	10	23,8500	23,8500		
LED & C	10	24,5680	24,5680		
Halogen & D	10			27,7590	
Sig.		,087	,052	,059	1,000

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 10,000.

Part B: Main effect (no-interaction) model

d-2) Model: Main-effect model

(g-2) ANOVA Table

Tests of Between-Subjects Effects

Dependent Variable: Shear bond strength

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2031,714 ^a	4	507,928	9,941	,000
Intercept	47894,642	1	47894,642	937,395	,000
light	34,716	1	34,716	,679	,412
material	1996,998	3	665,666	13,028	,000
Error	3832,001	75	51,093		
Total	53758,357	80			
Corrected Total	5863,715	79			

a. R Squared = ,346 (Adjusted R Squared = ,312)

h-2) Post-hoc multiple comparison

- ① Post-hoc multiple comparison
- ② Homogeneous subsets

Shear bond strength

Scheffe^{a,b}

material	N	Subset		
		1	2	3
A	20	18,4170		
B	20	22,2385	22,2385	
C	20		25,1455	
D	20			32,0710
Sig.		,420	,649	1,000

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 51,093.
a. Uses Harmonic Mean Sample Size = 20,000.
b. Alpha = ,05.

Reference

1. Daniel WW. Biostatistics - Basic concepts and methodology for the health sciences, 9th ed. Singapore: John Wiley & Sons (Asia) Pte Ltd; 2010. p334-368.