The ANOVA for 2x2 Independent Groups Factorial Design

Please Note: In the analyses above I have tried to avoid using the terms "Independent Variable" and "Dependent Variable" (IV and DV) in order to emphasize that statistical analyses are chosen based on the type of variables involved (i.e., qualitative vs. quantitative) and the type of statistical question (i.e., linear relationship vs. mean difference between two quantitative variables). As emphasized in class and lab, the issue of causal interpretation, when the terms IV and DV are most appropriate, is an issue of the type of research design that is used, not an issue of the statistic that is employed to determine if there is a relationship between the variables. Factorial designs, however are most commonly used in experimental settings, and so the terms IV and DV are used in the following presentation.

Application: This analysis is applied to a design that has two between groups IVs, both with two conditions (groups, samples). There are three separate "effects" tested as part of the 2x2 ANOVA, one corresponding to each main effect and the third involving the interaction (joint effect) of the two IVs.

H0: There are three, each corresponding to one of the effects of the study. Two of these involve the two main effects. For each main effect the corresponding H0: is that the populations represented by the conditions (groups, samples) of that IV have the same mean on the quantitative DV. The third involves the interaction, for which the H0: is that the effect of each IV is the same for both conditions (groups, samples) of the other IV.

To reject H0: is to say that the populations represented by the conditions of that IV (for each main effect), or combination of IVs (for the interaction) have different means on the quantitative DV.

The data: The data for this example are taken from a study that looked at the joint influence upon vocabulary test performance of the method used to study and the familiarity of the words. Five subjects were randomly assigned to each of four conditions of this factorial design (familiar words studied using definitions, familiar words studied using literature passages, unfamiliar words studied using definitions, and unfamiliar words studied using literature passages).

Here are the data from 20 subjects. For each, the Type of Study condition (D = Definition and P = passage) and the Word Type condition (F = familiar and U = unfamiliar) are given along with their score on the 20-word vocabulary test. The number of subjects in each IV conditions is represented as n (n = 5 in each of these conditions), the number of Type of Study conditions is represented as a (a = 2 for these data), the number of Word Type conditions is represented as b (b = 2 for these data) and the total number of subjects is represented as N (N = a * b * n = 2 * 2 * 5 = 20, for the example data).

```
F,D,16; U,P,8; U,D,11; F,P,14; F,P,14; U,P,6; U,D,10; F,D,16; U,P,7; F,D,13;
U,D,12; F,P,15; F,D,12; U,P,9; F,D,15; U,D,11; U,D,14; F,P,12; U,P,6; F,P,12
```

Research Hypothesis: The researcher's three hypotheses were: 1) there would be a main effect for Word Type, students would have better overall vocabulary scores with familiar than with unfamiliar words, 2) there would be a main effect for Type of Study, students would have better overall vocabulary scores following practice using definitions than using literature passages, 3) there would be an interaction of Word Type and Type of Study; for familiar words the two types of study would work equally well, but for unfamiliar words study of definitions would lead to better vocabulary scores than would study using literature passages.

H0: for this analysis: There are three: 1) there is no main effect for Word Type, 2) there is no main effect for Type of Study, and there is no interaction of Word Type and Type of Study.

Step 1 Assemble the data for analysis. Rearrange the data so that the subjects in each IV condition are in the same condition. There should be as many columns as there conditions in the "A" IV and as many rows as there are conditions in the "B" IV. Be sure each subject is in the cell that corresponds to their combination of IV conditions. Assign subject numbers consecutively within each IV condition (if your subjects are already assigned numbers in your data you may use them - the values of the subjects numbers doesn't influence the computation of F.

	Type of Study (A)			
	Definition (a1)		Passage (a2)	
Word Type (B)	Subject #	a1,b1	Subject #	a2,b1
Familiar (b1)	1	16	6	14
	2	16	7	14
	3	13	8	15
	4	12	9	12
	5	15	10	12
	Subject #	a1,b2	Subject #	a2,b2
Unfamiliar (b2)	11	11	16	8
	12	10	17	6
	13	12	18	7
	14	11	19	9
	15	14	20	6

Step 2 Compute the squared score for each subject and place in a column next to their scores.

	Type of Study (A)					
	Definition (a1)			Passage (a2)		
Word Type (B)	Subject #	a1,b1	X^2	Subject #	a2,b1	X^2
Familiar (b1)	1	16	256	6	14	196
	2	16	256	7	14	196
	3	13	169	8	15	225
	4	12	144	9	12	144
	5	15	225	10	12	144

	Type of Study (A)					
	Definition (a1)			Passage (a2)		
	Subject #	a1,b2	X^2	Subject #	a2,b2	X^2
Unfamiliar (b2)	11	11	121	16	8	64
	12	10	100	17	6	35
	13	12	144	18	7	49
	14	11	121	19	9	81
	15	14	196	20	6	36

Step 3 Compute ΣX, ΣX^2, the mean and sum of squares for each condition (following the formulas shown in the Descriptive Statistics section earlier) and complete the following table.

Word Type (B)		Type of Study (A): Definition (a1)		Passages (a2)
Familiar (b1)	ΣX	72.00	ΣX	67.00
	ΣX^2	1050.00	ΣX^2	905.00
	$\bar{X}$	14.40	$\bar{X}$	13.40
	SS	13.20	SS	7.20
Unfamiliar (b2)	ΣX	58.00	ΣX	36.00
	ΣX^2	682.00	ΣX^2	265.00
	$\bar{X}$	11.60	$\bar{X}$	7.20
	SS	9.20	SS	6.80

Step 4 Compute SS_{Total}

$$SS_{Total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = (1050 + 905 + 682 + 265) - \frac{(72 + 67 + 58 + 36)^2}{20}$$

$$= 2902 - \frac{233^2}{20} = 2902 - 2714.45 = 187.55$$

Step 5 Compute SS_{Error} (also called SS_{Within} or SS_{WG})

$$SS_{Error} = (SS_{a1,b1} + SS_{a2,b1} + SS_{a1,b2} + SS_{a2,b2}) = 13.20 + 7.20 + 9.20 + 6.80 = 36.40$$

Step 6 Compute df_{Error} (also called df_{Within} or df_{WG})

$$df_{Error} = a * b * (n-1) = 2 * 2 * (5-1) = 2 * 2 * 4 = 16$$

Step 7 Compute MS_{Error} (also called MS_{Within} or $MS_{(WG)}$)

$$MS_{Error} = \frac{SS_{Error}}{df_{Error}} = \frac{36.40}{16} = 2.28$$

Step 8 Compute SS_A (Sum of squares to test for a main effect of the A independent variable)
(Sum of squares for the main effect of studying using Definitions versus Passages)

$$SS_A = \frac{(\Sigma X_{a1,b1} + \Sigma X_{a1,b2})^2 + (\Sigma X_{a2,b1} + \Sigma X_{a2,b2})^2}{b * n} - \frac{(\Sigma X)^2}{N} =$$

$$= \frac{(72 + 58)^2 + (67 + 36)^2}{2 * 5} - 2714.45 = \frac{16900 + 10609}{10} - 2714.45 = 36.45$$

Step 9 Compute df_A (degrees of freedom for the main effect of the A independent variable)

$$df_A = a - 1 = 2 - 1 = 1$$

Step 10 Compute MS_A (Mean square to test for a main effect of the A independent variable)

$$MS_A = \frac{SS_A}{df_A} = \frac{36.45}{1} = 36.45$$

Step 11 Compute F_A (F-test for the main effect of the A independent variable)

$$F_A = \frac{MS_A}{MS_{Error}} = \frac{36.45}{2.28} = 15.97$$

Step 12 Use F Table to determine the critical value of F for $\alpha = .05$ and the appropriate degrees of freedom

Numerator degrees of freedom = $df_A = 1$ and
Denominator degrees of freedom = $df_{Error} = 16$,

$$F(1, 16, \alpha = .05) = 4.49$$

Step 13 Compare the obtained F_A and critical F, and determine whether to reject or retain the null hypothesis, following the same procedures we used with the other F-tests.

For the example data, we would decide to reject the null hypothesis, because the obtained F value of 15.97 is larger than the critical F value of 4.49. This tells us that, overall, participants who studied synonyms performed better than did those who studied passages (as hypothesized).

Step 14 Compute SS_B (Sum of squares to test for a main effect of the B independent variable)
(Sum of squares for the main effect of studying Familiar versus Unfamiliar words)

$$SS_B = \frac{(\Sigma X_{b1,a1} + \Sigma X_{b1,a2})^2 + (\Sigma X_{b2,a1} + \Sigma X_{b2,a2})^2}{a * n} - \frac{(\Sigma X)^2}{N} =$$

$$= \frac{(72 + 67)^2 + (58 + 36)^2}{2 * 5} - 2714.45 = \frac{19321 + 8836}{10} - 2714.45 = 101.25$$

Step 15 Compute df_B (degrees of freedom for the main effect of the B independent variable)

$$df_B = b - 1 = 2 - 1 = 1$$

Step 16 Compute MS_A (Mean square to test for a main effect of the B independent variable)

$$MS_B = \frac{SS_B}{Df_B} = \frac{101.25}{1} = 101.25$$

Step 17 Compute F_B (F-test for the main effect of the B independent variable)

$$F_B = \frac{MS_B}{MS_{Error}} = \frac{101.25}{2.28} = 44.41$$

Step 18 Use Table F to determine the critical value of F for $\alpha = .05$ and the appropriate degrees of freedom

numerator degrees of freedom = df_B = 1 and denominator degrees of freedom = df_{Error} = 16,

$F(1, 16, \alpha = .05) = 4.49$

Step 19 Compare the obtained F_B and critical F, and determine whether to reject or retain the null hypothesis.

For the example data, we would decide to reject the null hypothesis, because the obtained F value of 4.41 is larger than the critical F value of 4.49. This tell us that, overall, participants performed better with familiar than with unfamiliar words.

Step 20 Compute SS_{Effect} (A combination of the SS due to the main effects and the interaction)

$$SS_{Effect} = \frac{(\Sigma X_{b1,a1})^2 + (\Sigma X_{b1,a2})^2 + (\Sigma X_{b2,a1})^2 + (\Sigma X_{b2,a2})^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{72^2 + 67^2 + 58^2 + 36^2}{5} - 2714.45 = \frac{5184 + 4489 + 3364 + 1296}{5} - 2714.45 = 152.15$$

Step 21 Computational check

$$SS_{Total} = SS_{Effect} + SS_{Error} \qquad 188.55 = 152.15 + 36.40$$

Step 22 Compute SS_{AB} (Sum of squares to test for an interaction of the A and B independent variables)

$SS_{AB} = SS_{Effect} - SS_A - SS_B = 152.15 - 36.45 - 101.25 = 14.45$

Step 23 Compute df_{AB} (degrees of freedom for an AB interaction)

$df_{AB} = (a - 1) * (b - 1) = (2 - 1) * (2 - 1) = 1$

Step 24 Compute MS_{AB} (Mean square to test for an AB interaction)

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{14.45}{1} = 14.45$$

Step 25 Compute F_{AB} (F-test for the AB interaction)

$$F_{AB} = \frac{MS_{AB}}{} = \frac{14.45}{} = 6.34$$

MS_{Error} 2.28

Step 26 Use Table F to determine the critical value of F for $\alpha = .05$ and the appropriate degrees of freedom

numerator degrees of freedom = $df_{AB} = 1$ and denominator degrees of freedom = $df_{Error} = 16$,
$F(1, 16, \alpha = .05) = 4.49$

Step 27 Compare the obtained F_{AB} and critical F, and determine whether to reject or retain the null hypothesis.

For the example data, we would decide to reject the null hypothesis, because the obtained F value of 6.34 is larger than the critical F value of 4.49. (We'll look at the pattern of this interaction below).

Step 28 Preparing the Summary Table (Also called a Source Table)

The Summary Table for a factorial design looks very much like the one for an independent groups ANOVA, except that there are two main effects and an interaction presented. The values come from the ANOVA example. The table on the left shows which steps in the analysis produced each value, while the one on the right shows the actual values. df = degrees of freedom, SS = sum of squares, and MS = mean squares

Computational steps

Source	df	SS	MS	F
A	9	8	10	11
B	15	14	16	17
AB	25	24	26	27
Error	6	5	7	
Total	4			

Source	df	SS	MS	F	p.
A	1	36.45	36.45	15.97	< .05
B	1	101.25	101.25	44.41	< .05
AB	1	14.45	14.45	6.34	< .05
Error	16	36.40	2.28		
Total	19	187.55			

Computing and using LSD to Describe the Pattern of the Interaction and "Check-up" on the Main Effects

Step 1 Obtain the df_{Error} from analysis

From the example data, $df_{Error} = 16$

Step 2 Use Table t to determine the critical value of t for $\alpha = .05$ with df_{Error}. (Sometimes, with larger df, the table doesn't include the df you are looking for. When this happens, just use the next smaller df that is included on the table. For example, if you had df = 33, you would use the t for df = 30.)

$t\ (16, \alpha = .05) = 2.12$

Step 3 Obtain the MS_{Error} from the analysis

From the example data, $MS_{Error} = 2.28$

Step 4 Obtain n, the number of subjects in each "condition" of the design. If there is an unequal number of subjects in the different IV conditions, use the average number of subjects in the conditions.

From the example data, n = 5

Step 5 Compute the specific d_{LSD} -- the minimum significant pairwise mean difference, based on the LSD procedure.

$$d_{LSD} = \frac{t * \sqrt{(2 * MS_{Error})}}{\sqrt{n}} = \frac{2.12 * \sqrt{(2 * 2.28)}}{\sqrt{5}} = 2.02$$

Step 6 Build the table of cell means and marginal means and use <, > and = to describe each set of simple effects.

	Type of Study (A)		
Word Type (B)	Definition (a1)	Passages (a2)	
Familiar (b1)	14.40	13.40	13.90
Unfamiliar (b2)	11.60	7.20	9.40
	13.00	10.30	

Step 7 Describe the pattern of each set of simple effects. Note: Only one set of simple effects will be used to describe the pattern of the interaction, but each is necessary to determine whether the corresponding main effect is descriptive or misleading.

For the simple effect of Type of Study for each Word Type: Definition = Passages for Familiar words (the difference between 14.40 and 13.40 is smaller than the LSD value of 2.02, whereas Definition > Passages for Unfamiliar words (the difference between 11.50 and 7.20 is larger than the LSD value of 2.02).

For the simple effect of Word Type for each Type of Study: Familiar > Unfamiliar when Definitions were studied (the difference between 14.40 and 11.60 is larger than the LSD value of 2.02), and Familiar > Unfamiliar when Passages were studied (the difference between 13.40 and 7.20 is larger than the LSD value of 2.02), however the effect is larger for Passages (13.4 - 7.20 = 6.20) than for Definitions (14.4 - 11.6 = 2.8).

Describing the Results of the Factorial Analysis

There is a lot of information to organize and present when describing the analysis of a factorial design! Here is a summary of the information and the order of presentation. You will notice that although we computed the F-tests for the main effects first, we report the results of the interaction first. There are two reasons for this: 1) Usually in a factorial design the primary research question is about the presence and pattern of the interaction -- present first what you want the reader to focus their attention upon, 2) if the interaction is significant, then we know that one or both of the main effects might be potentially misleading.

First, we will report whether or not there is a significant interaction and the F-information
- -- if there is an interaction, we then report one set of simple effects to describe the interaction pattern
- -- if there is a research hypothesis about the pattern of the interaction, we should select the set of simple effects which directly addresses that hypothesis.

Second, we will report whether or not there is a main effect for one of the IVs, and the F-information
- -- if there is a research hypothesis about that main effect, we should describe whether or not the results support that researchhypothesis
- -- if there is an interaction, we would then report whether or not that main effect is descriptive or is potentially misleading, by comparing it with the pattern of the corresponding simple effects

Third, we will report whether or not there is a main effect for the other IVs, and that F-information
- -- if there is a research hypothesis about that main effect, we should describe whether or not the results support that research hypothesis

-- if there is an interaction, we would then report whether or not that main effect is descriptive or is potentially misleading, by comparing it with the pattern of the corresponding simple effects

Here are the researcher's hypotheses and how we might report the results of these analyses.
The researcher's hypotheses were: 1) there would be a main effect for Word Type, students would have better overall vocabulary scores with familiar than with unfamiliar words, 2) there would be a main effect for Type of Study, students would have better overall vocabulary scores following practice using definitions than using literature passages, 3) there would be an interaction of Word Type and Type of Study; for familiar words the two types of study would work equally well, but for unfamiliar words study of definitions would lead to better vocabulary scores than would study using literature passages.

A between groups factorial ANOVA was designed to examine how Type of Study and Word Type relate to performance on a vocabulary test. Table 1 summarizes the data form the analysis.
There is an interaction of Word Type and Type of Study as they vocabulary, $\underline{F}(1,16) = 6.34$, $\underline{p}<.05$, $\underline{MSe} = 2.28$. As hypothesized, the pattern of that interaction was that with Familiar Words there was no simple effect of Type of Study, however with Unfamiliar Words study using definitions led to better vocabulary performance than did study using Passages.[1]
There was a main effect of Type of Study ($\underline{F}(1,16) = 15.97$, $\underline{p}<.05$), with better overall performance following study using Definitions than using Passages. However, analysis of the simple effects of Type of Study revealed that this pattern of data was true only for Unfamiliar words. There was no Type of Study simple effect for Familiar words.
There was a main effect of Word Type, $\underline{F}(1,16) = 44.41$, $\underline{p}<.05$) with better overall performance with Familiar than with Unfamiliar words. Analysis of the simple effects of Word Type revealed that this pattern resulted with both study using Definitions and using Passages.

[1]Cell mean comparisons were based on the LSD minimum mean difference value of 2.02).

Table 1.
Vocabulary performance data.

	Type of Study		
	Definition	Passages	
Word Type			
Familiar	14.40	13.40	13.90
Unfamiliar	11.60	7.20	9.40
	13.00	10.30	

Table F: ANOVA (F-tests) Critical values of F for α = .05 & α = .01

Denominator df	α	Numerator df 1	2	3	4	5	6
1	.05	161	200	216	225	230	234
2	.05	18.5	19.0	19.2	19.2	19.3	19.3
	.01	98.5	99.0	99.2	99.2	99.3	99.3
3	.05	10.1	9.55	9.28	9.12	9.01	8.94
	.01	34.1	30.8	29.5	28.7	28.2	27.9
4	.05	7.71	6.94	6.59	6.39	6.26	6.16
	.01	21.2	18.0	16.7	16.0	15.5	15.2
5	.05	6.61	5.79	5.41	5.19	5.05	4.95
	.01	16.3	13.3	12.1	11.4	11.0	10.7
6	.05	5.99	5.14	4.76	4.53	4.39	4.28
	.01	13.7	10.9	9.78	9.15	8.75	8.47
7	.05	5.59	4.74	4.35	4.12	3.97	3.87
	.01	12.2	9.55	8.45	7.85	7.46	7.19
8	.05	5.32	4.46	4.07	3.84	3.69	3.58
	.01	11.3	8.65	7.59	7.01	6.63	6.37
9	.05	5.12	4.26	3.86	3.63	3.48	3.37
	.01	10.6	8.02	6.99	6.42	6.06	5.80
10	.05	4.96	4.10	3.71	3.48	3.33	3.22
	.01	10.0	7.56	6.55	5.99	5.64	5.39
11	.05	4.84	3.98	3.59	3.36	3.20	3.09
	.01	9.65	7.21	6.22	5.67	5.32	5.07
12	.05	4.75	3.89	3.49	3.26	3.11	3.00
	.01	9.33	6.93	5.95	5.41	5.06	4.82
13	.05	4.67	3.81	3.41	3.18	3.03	2.92
	.01	9.07	6.70	5.74	5.21	4.86	4.62
14	.05	4.60	3.74	3.34	3.11	2.96	2.85
	.01	8.86	6.51	5.56	5.04	4.69	4.46
15	.05	4.54	3.68	3.29	3.06	2.90	2.79
	.01	8.68	6.36	5.42	4.89	4.56	4.32
16	.05	4.49	3.63	3.24	3.01	2.85	2.74
	.01	8.53	6.23	5.29	4.77	4.44	4.20
17	.05	4.45	3.59	3.20	2.96	2.81	2.70
	.01	8.40	6.11	5.18	4.67	4.34	4.10
18	.05	4.41	3.55	3.16	2.93	2.77	2.66
	.01	8.29	6.01	5.09	4.58	4.25	4.01

Numerator df Denominator df	α	1	2	3	4	5	6
19	.05	4.38	3.52	3.13	2.90	2.74	2.63
	.01	8.18	5.93	5.01	4.50	4.17	3 94
20	.05	4.35	3.49	3.10	2.87	2.71	2.60
	.01	8.10	5.85	4.94	4.43	4.10	3.87
22	.05	4.30	3.44	3.05	2.82	2.66	2.55
	.01	7.95	5.72	4.82	4.31	3.99	3.76
24	.05	4.26	3.40	3.01	2.78	2.62	2.51
	.01	7.82	5.61	4.72	4.22	3.90	3.67
26	.05	4.23	3.37	2.98	2.74	2.59	2.47
	.01	7.72	5.53	4.64	4.14	3.82	3.59
28	.05	4.20	3.34	2.95	2.71	2.56	2.45
	.01	7.64	5.45	4.57	4.07	3.75	3.53
30	.05	4.17	3.32	2.92	2.69	2.53	2.42
	.01	7.56	5.39	4.51	4.02	3.70	3.47
40	.05	4.08	3.23	2.84	2.61	2.45	2.34
	.01	7.31	5.18	4.31	3.83	3.51	3.29
60	.05	4.00	3.15	2.76	2.53	2.37	2.25
	.01	7.08	4.98	4.13	3.65	3.34	3.12
120	.05	3.92	3.07	2.68	2.45	2.29	2.17
	.01	6.85	4.79	3.95	3.48	3.17	2.96
200	.05	3.89	3.04	2.65	2.42	2.26	2.14
	.01	6.76	4.71	3.88	3.41	3.11	2.89
∞	.05	3.84	3.00	2.60	2.37	2.21	2.10
	.01	6.65	4.61	3.78	3.32	3.02	2.S0

Critical values of t for a = .05 & a = .01

df	α = .05	α = .01
1	12.71	63.66
2	4.30	9.92
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.36	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3 17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.14	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
20	2.09	2.84
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77
28	2.05	2.76
29	2.04	2.76
30	2.04	2.75
40	2.02	2.70
60	2.00	2.66
120	1.98	2.62
∞	1.96	2.58