Factorial ANOVA

Cal State Northridge
Ψ320
Andrew Ainsworth PhD

Topics in Factorial Designs

- What is Factorial?
- Assumptions
- Analysis
- Multiple Comparisons
  - Main Effects
  - Simple Effects
  - Simple Comparisons
- Effect Size estimates
- Higher Order Analyses

Factorial?

- Factorial – means that:
  1. You have at least 2 IVs
  2. And all levels of one variable occur in combination with all levels of the other variable(s).
- Assumptions
  - Same as one-way ANOVA but they are tested within each cell
  - i.e. Normality, Homogeneity and Independence

Simplest Form: 2 x 2 ANOVA

		B	
		b_1	b_2
A	a_1		
	a_2		

		Video Game	
		GTA	NBA 2K7
Gender	Men		
	Women		

Analysis

- Performing a factorial analysis does the job of three analyses in one
  - Two one-way ANOVAs, one for each IV (called a main effect)
  - And a test of the interaction between the IVs
  - Interaction? – the effect of one IV depends on the level of another IV
    - The variability that is left over after you assess each IV
    - The 2 IVs together work to affect scores over and above either of them independently

Analysis

- The between groups sums of squares from 1-way ANOVA is further broken down:
  - Before $SS_{bg} = SS_{effect}$
  - Now $SS_{bg} = SS_A + SS_B + SS_{AB}$
  - In a two IV factorial design A, B and AxB all differentiate between groups, therefore they all add to the SS_{bg}

Analysis

- Total variability = (variability of A around GM) + (variability of B around GM) + (variability of each group mean {AB} around GM) + (variability of each person's score around their group mean)
- $SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{error}$

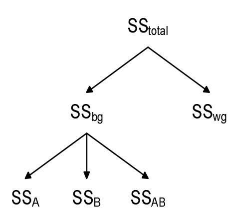
$$\sum (Y_i - \bar{Y}_{GM})^2 = \sum n_a(\bar{Y}_a - Y_{GM})^2 + \sum n_b(\bar{Y}_b - Y_{GM})^2$$
$$+\left[\sum n_{ab}(\bar{Y}_{ab} - \bar{Y}_{GM})^2 - \sum n_a(\bar{Y}_a - Y_{GM})^2 - \sum n_b(\bar{Y}_b - Y_{GM})^2\right]$$
$$+\sum (Y_i - \bar{Y}_{ab})^2$$

Analysis

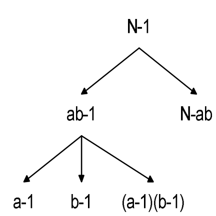
- Degrees of Freedom
  - df_A = #groups$_A$ – 1
  - df_B = #groups$_B$ – 1
  - $df_{AB} = (a – 1)(b – 1)$
  - $df_{error} = ab(n – 1) = abn – ab = N – ab$
  - $df_{total} = N – 1 = a – 1 + b – 1 + (a – 1)(b – 1) + N – ab$

Analysis

- Breakdown of sums of squares



- Breakdown of degrees of freedom



Analysis

- Mean square
  - The mean squares are calculated the same
  - SS/df = MS
  - You just have more of them, MS_A, MS_B, MS_{AB}, and MS_{WG}
  - This expands when you have more IVs
    - One for each main effect, one for each interaction (two-way, three-way, etc.)

Analysis

- F-test
  - Each effect and interaction is a separate F-test
  - Calculated the same way: MS_{effect}/MS_{WG} since MS_{WG} is our error variance estimate
  - You look up a separate F_{crit} for each test using the df_{effect}, df_{WG} and tabled values

Example

	B: Vacation Length		
A: Profession	b1: 1 week	b2: 2 weeks	b3: 3 weeks
	0	4	5
a1: Administrators	1	7	8
	0	6	6
	5	5	9
a2: Belly Dancers	7	6	8
	6	7	8
	5	9	3
a3: Politicians	6	9	3
	8	9	2

Analysis

- Sample data reconfigured into cell and marginal means (with variances)

	B: Vacation Length			
A:Profession	b_1: 1 week	b_2: 2 weeks	b_3: 3 weeks	Marginal A means
a_1: Administrators	$\bar{Y}_{a_1b_1} = 0.333$ $s^2_{a_1b_1} = 0.333$	$\bar{Y}_{a_1b_2} = 5.667$ $s^2_{a_1b_2} = 2.333$	$\bar{Y}_{a_1b_3} = 6.333$ $s^2_{a_1b_3} = 2.333$	$\bar{Y}_{a_1} = 4.111$
a_2: Belly Dancers	$\bar{Y}_{a_2b_1} = 6$ $s^2_{a_2b_1} = 1$	$\bar{Y}_{a_2b_2} = 6$ $s^2_{a_2b_2} = 1$	$\bar{Y}_{a_2b_3} = 8.333$ $s^2_{a_2b_3} = 0.333$	$\bar{Y}_{a_2} = 6.778$
a_3: Politicians	$\bar{Y}_{a_3b_1} = 6.333$ $s^2_{a_3b_1} = 2.333$	$\bar{Y}_{a_3b_2} = 9$ $s^2_{a_3b_2} = 0$	$\bar{Y}_{a_3b_3} = 2.667$ $s^2_{a_3b_3} = 0.333$	$\bar{Y}_{a_3} = 6$
Marginal B Means	$\bar{Y}_{b_1} = 4.222$	$\bar{Y}_{b_2} = 6.889$	$\bar{Y}_{b_3} = 5.778$	$\bar{Y}_{...} = 5.630$

$\sum Y^2 = 1046$

Example – Sums of Squares

$$SS_{total} = \sum (Y_i - \bar{Y}_{GM})^2 =$$
$$= (\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 +$$
$$(5 - 5.630)^2 + (7 - 5.630)^2 + (6 - 5.630)^2 +$$
$$(5 - 5.630)^2 + (6 - 5.630)^2 + (8 - 5.630)^2 +$$
$$(4 - 5.630)^2 + (7 - 5.630)^2 + (6 - 5.630)^2 +$$
$$+ \ldots + (3 - 5.630)^2 + (2 - 5.630)^2 = 190.296$$

Example – Sums of Squares

$$SS_A = \sum n_a (\bar{Y}_a - Y_{GM})^2 =$$
$$= [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] + [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] +$$
$$[\_\_\_ * (6 - 5.630)^2] = 33.852$$

Example – Sums of Squares

$$SS_B = \sum n_b(\bar{Y}_b - Y_{GM})^2 =$$
$$= [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] + [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] +$$
$$[\_\_\_ * (5.778 - 5.630)^2] = 32.296$$

Psy 320 - Cal State Northridge 16

Example – Sums of Squares

$$SS_{AB} = \sum n_{ab}(\bar{Y}_{ab} - \bar{Y}_{GM})^2 - \sum n_a(\bar{Y}_a - Y_{GM})^2 - \sum n_b(\bar{Y}_b - Y_{GM})^2 =$$
$$= [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] + [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] +$$
$$+ [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] + [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] +$$
$$+ [\_\_\_ * (\_\_\_\_ - \_\_\_\_)^2] + [\_\_\_ * (9 - 5.630)^2] +$$
$$[\_\_\_ * (6.333 - 5.630)^2] + [\_\_\_ * (8.333 - 5.630)^2] +$$
$$+ [\_\_\_ * (2.667 - 5.630)^2] = 170.296$$
$$= 170.296 - 33.825 - 32.296 = 104.148$$

Psy 320 - Cal State Northridge 17

Example – Sums of Squares

$$SS_{Error} = \sum (Y_i - \bar{Y}_{ab})^2 =$$
$$= (\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 +$$
$$(\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 + (\_\_\_\_ - \_\_\_\_)^2 +$$
$$(5 - 6.333)^2 + (6 - 6.333)^2 + (8 - 6.333)^2 +$$
$$(4 - 5.667)^2 + (7 - 5.667)^2 + (6 - 5.667)^2 +$$
$$+ \ldots + (3 - 2.667)^2 + (2 - 2.667)^2 = 20$$

Psy 320 - Cal State Northridge 18

Analysis – Computational

- Marginal Totals – we look in the margins of a data set when computing main effects
- Cell totals – we look at the cell totals when computing interactions
- In order to use the computational formulas we need to compute both marginal and cell totals

Analysis – Computational

- Sample data reconfigured into cell and marginal totals

	B: Vacation Length			
A: Profession	b_1: 1 week	b_2: 2 weeks	b_3: 3 weeks	Marginal Sums for A
a_1: Administrators	1	17	19	$a_1 = 37$
a_2: Belly Dancers	18	18	25	$a_2 = 61$
a_3: Politicians	19	27	8	$a_3 = 54$
Marginal Sums for B	$b_1 = 38$	$b_2 = 62$	$b_3 = 52$	T = 152

Analysis – Computational

- Formulas for SS

$$SS_A = \frac{\sum\left(\sum a\right)^2}{bn} - \frac{T^2}{abn}$$

$$SS_B = \frac{\sum\left(\sum b\right)^2}{an} - \frac{T^2}{abn}$$

$$SS_{AB} = \frac{\sum\left(\sum ab\right)^2}{n} - \frac{\sum\left(\sum a\right)^2}{bn} - \frac{\sum\left(\sum b\right)^2}{an} + \frac{T^2}{abn}$$

$$SS_{error} = \sum Y^2 - \frac{\sum\left(\sum ab\right)^2}{n}$$

$$SS_T = \sum Y^2 - \frac{T^2}{abn}$$

Analysis – Computational

- Example

$$SS_A = \frac{\sum\left(\sum a\right)^2}{bn} - \frac{T^2}{abn}$$

$$SS_A = \frac{\_\_\_^2 + \_\_\_^2 + 54^2}{3(3)} - \frac{\_\_\_^2}{3(3)(3)} = \_\_\_\_ - \_\_\_\_ = 33.85$$

$$SS_B = \frac{\sum\left(\sum b\right)^2}{an} - \frac{T^2}{abn} =$$

$$SS_B = \frac{38^2 + \_\_\_^2 + \_\_\_^2}{3(3)} - \frac{\_\_\_^2}{3(3)(3)} = 888 - 855.7 = 32.30$$

Analysis – Computational

- Example

$$SS_{AB} = \frac{\sum\left(\sum ab\right)^2}{n} - \frac{\sum\left(\sum a\right)^2}{bn} - \frac{\sum\left(\sum b\right)^2}{an} + \frac{T^2}{abn}$$

$$SS_{AB} = \frac{\_\_\_^2 + \_\_\_^2 + \_\_\_^2 + 18^2 + 18^2 + 25^2 + 19^2 + 27^2 + 8^2}{3}$$

$$- \frac{37^2 + 61^2 + 54^2}{3(3)} - \frac{38^2 + 62^2 + 52^2}{3(3)} + \frac{152^2}{3(3)(3)}$$

$$= \_\_\_\_ - 889.55 - 888 + 855.7 = 104.15$$

Analysis – Computational

- Example

$$SS_{error} = \sum Y^2 - \frac{\sum\left(\sum ab\right)^2}{n}$$

$$SS_{error} = \_\_\_\_\_ - \frac{1^2 + 17^2 + 19^2 + 18^2 + 18^2 + 25^2 + 19^2 + 27^2 + 8^2}{3}$$

$$= \_\_\_\_ - 1026 = 20$$

$$SS_T = \sum Y^2 - \frac{T^2}{abn}$$

$$SS_T = 1046 - \frac{152^2}{3(3)(3)} = 1046 - 855.7 = 190.30$$

Analysis – Computational

- Example

$$df_A = a-1 = 3-1 = 2$$
$$df_B = b-1 = 3-1 = 2$$
$$df_{AB} = (a-1)(b-1) = (3-1)(3-1) = 2(2) = 4$$
$$df_{Error} = abn - ab = 27-9 = 18$$
$$df_{total} = abn - 1 = 27 - 1 = 26$$

Analysis

- Example

Tests of Between-Subjects Effects

Dependent Variable: ENJOY

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
PROFESSION	33.852	2	16.926	15.233	.000
LENGTH OF STAY	32.296	2	16.148	14.533	.000
PROFESSION * LENGTH	104.148	4	26.037	23.433	.000
WITHIN GROUPS	20.000	18	1.111		
TOTAL	190.296	26			

- The MS_{WG} is also the pooled (average) variance across the cells, since all n are equal:

(.333+2.333+2.333+1+1+.333+2.333+0+.333)/9 = 1.111

Analysis

- $F_{crit}(2,18)=3.55$
- $F_{crit}(4,18)=2.93$
- Since 15.25 > 3.55, the effect for profession is significant
- Since 14.55 > 3.55, the effect for length is significant
- Since 23.46 > 2.93, the effect for profession * length is significant

Effect Size Revisited

- Eta Squared is calculated for each effect

$$\eta^2_{effect} = \frac{SS_{effect}}{SS_{total}}$$

- Omega Squared also for each effect

$$\omega^2_{Effect} = \frac{SS_{Effect} - (k_{Effect} - 1)MS_{WG}}{SS_T + MS_{WG}}$$

Effect Size Example

- Effect Size for Profession

$$\eta^2_{\text{Profession}} = \frac{SS_{\text{Profession}}}{SS_{\text{total}}} = \frac{33.852}{190.296} = .178$$

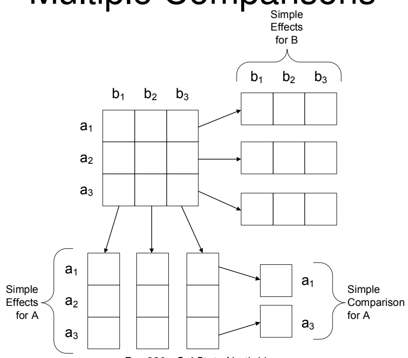
$$\omega^2_{\text{Profession}} = \frac{SS_{\text{Profession}} - (k_{\text{Profession}} - 1)MS_{WG}}{SS_T + MS_{WG}} =$$

$$\omega^2_{\text{Profession}} = \frac{33.853 - [(3-1)*1.111]}{190.296 + 1.111} = .165$$

Multiple Comparisons

- If a main effect is significant and has more than 2 levels, than you need to do marginal comparisons
- If the interaction is significant
  - You should break the interaction down by performing a **simple effect analysis** of A at each level of B (The effect of A at B_1, A at B_2, A at B_3, etc.) and vice versa
  - If any of them are significant and if A has more than 2 levels, follow up with simple comparisons

Multiple Comparisons



Specific Comparisons

- If the comparisons were planned than analyze them without any adjustment to the critical value
- If they were post-hoc than the values needs to be adjusted (e.g. Tukey, Bonferroni, etc.)
  - This is the same as previously covered

Multiple Comparisons Example Main Effect: Profession

Multiple Comparisons

Dependent Variable: ENJOY

	(I) PROFESS	(J) PROFESS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
LSD	1 Administrators	2 Belly Dancers	-2.67*	.497	.000	-3.71	-1.62
		3 Politicians	-1.89*	.497	.001	-2.93	-.84
	2 Belly Dancers	1 Administrators	2.67*	.497	.000	1.62	3.71
		3 Politicians	.78	.497	.135	-.27	1.82
	3 Politicians	1 Administrators	1.89*	.497	.001	.84	2.93
		2 Belly Dancers	-.78	.497	.135	-1.82	.27
Bonferroni	1 Administrators	2 Belly Dancers	-2.67*	.497	.000	-3.98	-1.36
		3 Politicians	-1.89*	.497	.004	-3.20	-.58
	2 Belly Dancers	1 Administrators	2.67*	.497	.000	1.36	3.98
		3 Politicians	.78	.497	.405	-.53	2.09
	3 Politicians	1 Administrators	1.89*	.497	.004	.58	3.20
		2 Belly Dancers	-.78	.497	.405	-2.09	.53

Based on observed means.

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

Multiple Comparisons Example Main Effect: Length of Stay

Multiple Comparisons

Dependent Variable: ENJOY

	(I) LENGTH	(J) LENGTH	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1 1 week	2 2 weeks	-2.67*	.497	.000	-3.71	-1.62
		3 3 weeks	-1.56*	.497	.006	-2.60	-.51
	2 2 weeks	1 1 week	2.67*	.497	.000	1.62	3.71
		3 3 weeks	1.11*	.497	.038	.07	2.16
	3 3 weeks	1 1 week	1.56*	.497	.006	.51	2.60
		2 2 weeks	-1.11*	.497	.038	-2.16	-.07
Bonferroni	1 1 week	2 2 weeks	-2.67*	.497	.000	-3.98	-1.36
		3 3 weeks	-1.56*	.497	.017	-2.87	-.24
	2 2 weeks	1 1 week	2.67*	.497	.000	1.36	3.98
		3 3 weeks	1.11	.497	.115	-.20	2.42
	3 3 weeks	1 1 week	1.56*	.497	.017	.24	2.87
		2 2 weeks	-1.11	.497	.115	-2.42	.20

Based on observed means.

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

Simple Effect and Simple Comp. Profession at 1 week

ANOVA

ENJOY

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	68.222	2	34.111	27.909	.001
Within Groups	7.333	6	1.222		
Total	75.556	8			

Multiple Comparisons

Dependent Variable: ENJOY

	(I) PROFESS	(J) PROFESS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	1 Administrators	2 Belly Dancers	-5.67*	.903	.001	-7.88	-3.46
		3 Politicians	-6.00*	.903	.001	-8.21	-3.79
	2 Belly Dancers	1 Administrators	5.67*	.903	.001	3.46	7.88
		3 Politicians	-.33	.903	.725	-2.54	1.88
	3 Politicians	1 Administrators	6.00*	.903	.001	3.79	8.21
		2 Belly Dancers	.33	.903	.725	-1.88	2.54
Bonferroni	1 Administrators	2 Belly Dancers	-5.67*	.903	.002	-8.63	-2.70
		3 Politicians	-6.00*	.903	.002	-8.97	-3.03
	2 Belly Dancers	1 Administrators	5.67*	.903	.002	2.70	8.63
		3 Politicians	-.33	.903	1.000	-3.30	2.63
	3 Politicians	1 Administrators	6.00*	.903	.002	3.03	8.97
		2 Belly Dancers	.33	.903	1.000	-2.63	3.30

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

Higher-Order Designs

- Higher-order – meaning more than 2 IVs
  - With 3 IVs; each with 2 levels you have a 2 x 2 x 2 design
  - If we have even 5 subjects per cell we are talking about a minimum of 40 subjects
  - We are also talking about:
    - $SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_{WG}$

Higher-Order Designs

- Higher-order – meaning more than 2 IVs
  - With 4 IVs; each with 2 levels you have a 2 x 2 x 2 x 2 design
  - If we have even 5 subjects per cell we are talking about a minimum of 80 subjects
  - We are also talking about:
    - $SS_T = SS_A + SS_B + SS_C + SS_D + SS_{AB} + SS_{AC} + SS_{AD} + SS_{BC} + SS_{BD} + SS_{CD} + SS_{ABC} + SS_{ABD} + SS_{ACD} + SS_{BCD} + SS_{ABCD} + SS_{WG}$