Factorial BG ANOVA

Psy 420

Ainsworth

Topics in Factorial Designs

- Factorial?
  - Crossing and Nesting
- Assumptions
- Analysis
  - Traditional and Regression Approaches
  - Main Effects of IVs
  - Interactions among IVs
  - Higher order designs
  - "Dangling control group" factorial designs
- Specific Comparisons
  - Main Effects
  - Simple Effects
  - Interaction Contrasts
- Effect Size estimates
- Power and Sample Size

Factorial?

- Factorial – means that all levels of one IV are completely crossed with all level of the other IV(s).
  - Crossed – all levels of one variable occur in combination with all levels of the other variable(s)
  - Nested – levels of one variable appear at different levels of the other variable(s)

Factorial?

- Crossing example

		Teaching Method		
		Lecture	Media	Lecture/Media
Text book	Tabachnick and Fidell	L & TF	M & TF	LM & TF
	Keppel and Wickens	L & KW	M & KW	LM & KW

  - Every level of teaching method is found together with every level of book
  - You would have a different randomly selected and randomly assigned group of subjects in each cell
    - Technically this means that subjects are nested within cells

Factorial?

- Crossing Example 2 – repeated measures

	Pre - test	Mid - test	Post - test
	s1	s1	s1
	s2	s2	s2
Subjects	s3	s3	s3
	s4	s4	s4
	s5	s5	s5

  - In repeated measures designs subjects cross the levels of the IV

Factorial?

- Nesting Example

		Teaching Method		
		Lecture	Media	Lecture/Media
Text book	T and F	L & TF/ Class 1	M & TF/ Class 3	LM & TF/ Class 5
	K and W	L & KW/ Class 2	M & KW/ Class 4	LM & KW/ Class 6

  - This example shows testing of classes that are pre-existing; no random selection or assignment
  - In this case classes are nested within each cell which means that the interaction is confounded with class

Assumptions

- Normality of Sampling distribution of means
  - Applies to the individual cells
  - 20+ DFs for error and assumption met
- Homogeneity of Variance
  - Same assumption as one-way; applies to cells
  - In order to use ANOVA you need to assume that all cells are from the same population

Assumptions

- Independence of errors
  - Thinking in terms of regression; an error associated with one score is independent of other scores, etc.
- Absence of outliers
  - Relates back to normality and assuming a common population

Equations

- Extension of the GLM to two IVs

$$Y = \mu + \alpha + \beta + \alpha\beta + \varepsilon$$

  - α = deviation of a score, Y, around the grand mean, μ, caused by IV *A* (Main effect of A)
  - β = deviation of scores caused by IV *B* (Main effect of B)
  - $\alpha\beta$ = deviation of scores caused by the interaction of *A and B* (Interaction of AB), above and beyond the main effects

Equations

- Performing a factorial analysis essentially does the job of three analyses in one
  - Two one-way ANOVAs, one for each main effect
  - And a test of the interaction
  - Interaction – the effect of one IV depends on the level of another IV
    - e.g. The T and F book works better with a combo of media and lecture, while the K and W book works better with just lecture

Equations

- The between groups sums of squares from previous 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}

Equations

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

$$\sum_i \sum_a \sum_b (Y_{iab} - GM)^2 = n_a \sum (\bar{Y}_a - GM)^2 + n_b \sum (\bar{Y}_b - GM)^2$$

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

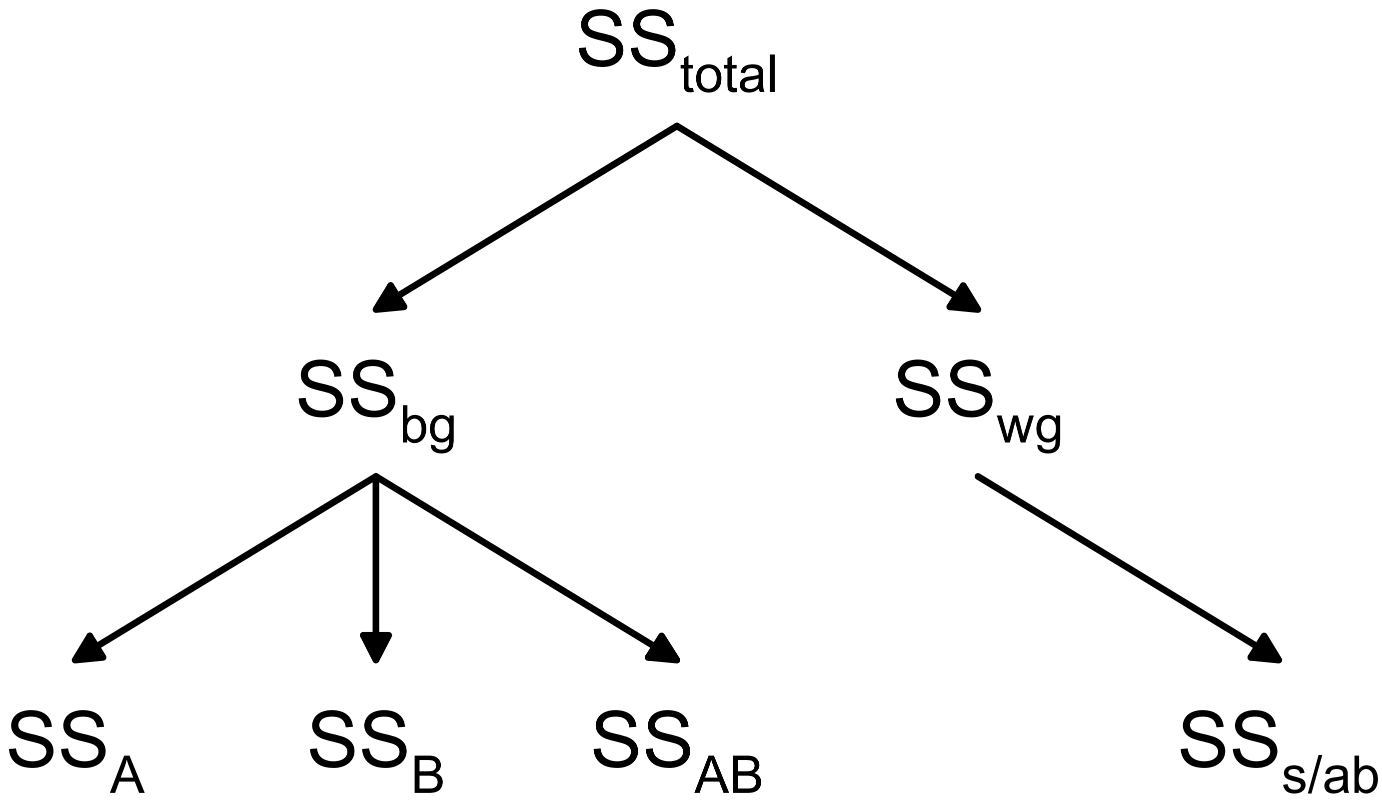
$$+ \sum_i \sum_a \sum_b (Y_{iab} - \bar{Y}_{ab})^2$$

Equations

- Degrees of Freedom
  - $df_{effect} = \#groups_{effect} - 1$
  - $df_{AB} = (a - 1)(b - 1)$
  - $df_{s/AB} = ab(s - 1) = abs - ab = abn - ab$
    $= N - ab$
  - $df_{total} = N - 1 = a - 1 + b - 1 + (a - 1)(b - 1)$
    $+ N - ab$

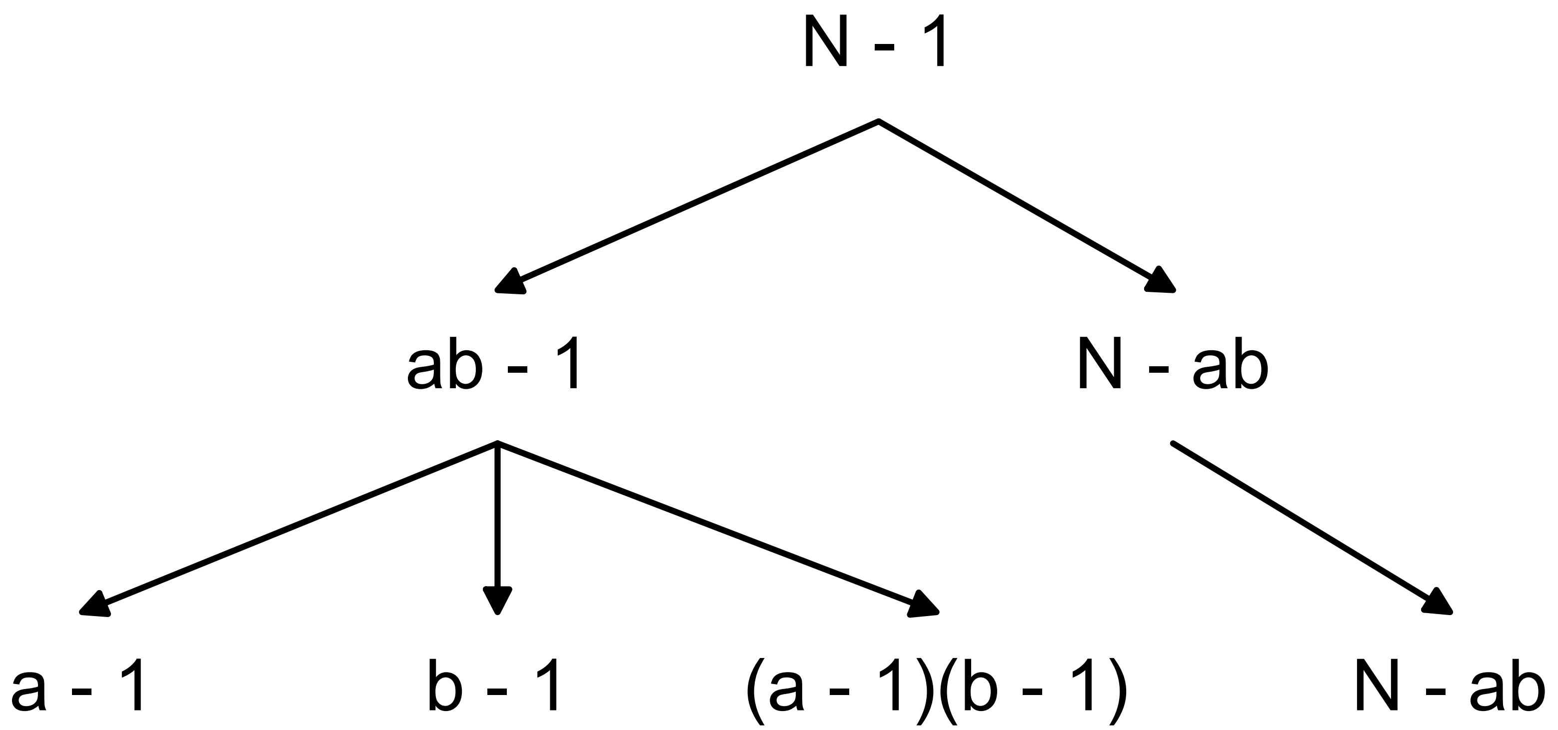
Equations

- Breakdown of sums of squares



Equations

- Breakdown of degrees of freedom



Equations

- 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 $MS_{S/AB}$
  - This expands when you have more IVs
    - One for each main effect, one for each interaction (two-way, three-way, etc.)

Equations

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

Sample data

	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

$$\sum Y^2 = 0^2 + 1^2 + \cdots + 2^2 = 1046$$

Sample data

- Sample info
  - So we have 3 subjects per cell
  - A has 3 levels, B has 3 levels
  - So this is a 3 x 3 design

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

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

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

$$SS_{AB} = \frac{\sum\left(\sum ab_{jk}\right)^2}{n} - \frac{\sum\left(\sum a_j\right)^2}{bn} - \frac{\sum\left(\sum b_k\right)^2}{an} + \frac{T^2}{abn}$$

$$SS_{S/AB} = \sum Y^2 - \frac{\sum\left(\sum ab_{jk}\right)^2}{n}$$

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

Analysis – Computational

- Example

$$SS_A = \frac{37^2 + 61^2 + 54^2}{3(3)} - \frac{152^2}{3(3)(3)} = 889.55 - 855.7 = 33.85$$

$$SS_B = \frac{38^2 + 62^2 + 52^2}{3(3)} - \frac{152^2}{3(3)(3)} = 888 - 855.7 = 32.30$$

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

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

$$= 1026 - 889.55 - 888 + 855.7 = 104.15$$

Analysis – Computational

- Example

$$SS_{S/AB} = 1046 - \frac{1^2 + 17^2 + 19^2 + 18^2 + 18^2 + 25^2 + 19^2 + 27^2 + 8^2}{3}$$

$$= 1046 - 1026 = 20$$

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

Analysis – Computational

- Example

$$df_A = a - 1 = 3 - 1 = 2$$

$$df_B = b - 1 = 3 - 1 = 2$$

$$df_{AB} = (a-1)(b-1) = (3-1)(3-1) = 2(2) = 4$$

$$df_{S/AB} = abn - ab = 27 - 9 = 18$$

$$df_{total} = abn - 1 = 27 - 1 = 26$$

Analysis – Computational

- Example

Source	SS	df	MS	F
Profession	33.85	2	16.93	15.25
Length	32.3	2	16.15	14.55
Profession x Length	104.15	4	26.04	23.46
Subjects/Profession x Length	20	18	1.11	
Total	190.3	26		

Analysis – Computational

- $F_{crit}(2,18)=3.55$
- $F_{crit}(4,18)=2.93$
- 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