Lecture 24: Factorial Designs

Dewayne E Perry
ENS 623
perry@ece.utexas.edu

Comparing Multiple Conditions

- 4 treatment conditions
  - Means: PD = 8, P = 4, D = 4, O = 2
  - P vs O: $t(8) = 1.55$, $p = .08$, one-tailed
  - PD vs D: $t(8) = 3.10$, $p = .01$
- Instead do one simultaneous *t* test
  - (PD + P)/2 vs (D + O)/2
    - Compare conditions with P against those without
    - Advantage: greater power (increases n1 and n2)
  - (PD + D)/2 vs (P + O)/2
    - Compare benefits of drug therapy
    - Again, greater power

Comparing Multiple Conditions

$$t = \frac{\left[(8+4)/2 - (4+2)/2\right]}{\sqrt{\left(\frac{1}{6}+\frac{1}{6}\right)2.5}} = \frac{6-3}{0.913} = 3.29$$

- Fisher noticed that a one-way analysis of variance could be rearrayed to form a two dimensional design of much greater power:
  - Factorial design
  - Two or more levels of each factor in combination with two or more levels of every other factor

Comparing Multiple Conditions

Drug	Psychotherapy *present*	*absent*	*mean*
present	8[3]	4[3]	6[6]
absent	4[3]	2[3]	3[6]
mean	6[6]	3[6]	4.5[12]

- → Comparisons
  - Column means: effect of psychotherapy
  - Row means: effect of drug therapy
  - Number of observations for mean has doubled
  - Greater economy:
    - Each condition or group contributes data to several comparisons

Analysis of Variance

- Can decomposed 4 basis means into
  - Grand mean
    - *(8 + 4 + 4 + 2)/4 = 4.5*
  - Residual/interaction effects of group membership

group mean	-	*grand mean*	=	*residual effect*
8	-	4.5	=	3.5
4	-	4.5	=	-0.5
4	-	4.5	=	-0.5
2	-	4.5	=	-2.5
18		18		0.0

  - Sum of residual effects is always zero

Analysis of Variance

- In 2 way factorial design get
  - Row effects
    - Row means – grand mean = row effects
    - *DP = 1.5, DA = -1.5*
  - Columns effects
    - column means – grand mean = column effects
    - *PP = 1.5, PA = -1.5*
  - Residual/interaction effects
    - *Group mean – grand mean – row effect – column effect = interaction effect*

GrpM		GrdM		RowE		ColE		InterE
8	-	4.5	-	1.5	-	1.5	=	0.5
4	-	4.5	-	1.5	-	-1.5	=	-0.5
4	-	4.5	-	-1.5	-	1.5	=	-0.5
2	-	4.5	-	-1.5	-	-1.5	=	0.5

What Do We Learn

- Group Mean tells us general level of measurements
  - Usually not of great interest
- Row Effects
  - Better to receive drug therapy than not
- Column Effects
  - Better to receive psychotherapy than not
- Interaction effects
  - Better to receive both than either
  - Indication that it is better to receive neither than either is more than offset by row/column effects

Individual Differences

- Analysis so far does not tell quite the whole story
  - Does not take into account in various scores
  - Variability from mean – deviations
  - Call these deviations errors
    - *Error = score – group mean*
    - Large error: falls far from mean
    - Small error: falls close to the mean
  - *Score = grand mean + row effect + column effect + interaction effect + error*

Individual Differences

- PD: S=9, E=1; S=8, E=0; S=7, E=-1
- D: S=5, E=1; S=4, E=0; S=3, E=-1
- P: S=6, E=2; S=4, E=0; S=2, E=-2
- O: S=4, E=2; S=2, E=0; S=0, E=-2

- Sum of scores 54
- Sum of squared scores 320
- Sum of squared grand means 243
- Sum of squared row effects 27
- Sum of squared column effects 27
- Sum of squared interaction effects 3
- Sum of squared errors 20
- Total SS 77

- Between-conditions SS 57
  - *BC SS = RE SS + CE SS + IE SS*
- Within conditions SS 20
  - Notice it is the same as sum of squared errors

Individual Differences

- Variance
  - Drug therapy and psychotherapy
    - Large *eta* (.76), and significant (*p* = .012)
  - Interaction effect
    - Not trivial *eta* (.36), not close to statistically significant (*p* = .30)
    - Important in two way and higher order analyses of variance
    - Often misinterpreted

Interaction Effects

- $eta = \sqrt{SS_{between} / SS_{between} + SS_{within}}$
- *eta*, like *r*, represents square root of proportion of variance accounted for
  - But, *eta* is a very non-specific index of effect size when it is based on a source of variance with *df* > 1
  - Eg, *eta* = .86 based on *df* = 3 is very large, but cannot say why it is large
  - When *df* = 1, *eta* is identical with *r*
    - Drug/psychotherapy: *eta = r = .76*
    - Get all the ways of interpreting *r*
- While not significant (p = .30), *eta* is of promising magnitude (*eta* = *r* = .36)
- We regard each effect size estimate as though it were the only one in the study
- Remember that when r^2 or eta^2 exceeds 1.00
  - *.574 + .574 + .130 = 1.278*

Variance

- Proportions of SS
  - Proportion of total SS =
    - SS effect of interest, divided by
    - SS eoi + SS within + SS all other between effects
  - SS: D=27, P=27, I=3, WC=20. T=77
  - D=.35, P=.35, I=.04, WC=.26, T=1.00
  - r^2 or eta^2 =
    - SS effect of interest, divided by
    - SS effect of interest + SS within

Testing Grand Mean

- Lack of interest in the magnitude of grand mean
  - In part due to arbitrary units of measurement often employed
- Sometimes, the constant of measurement may be of interest
  - When we failed to replicate a relationship obtained in an earlier experiment; compare our sample of subjects with an earlier sample
  - When dependent variable might estimate some skill that might or might not be better than chance
  - When our dependent variable might already be a difference score - eg, the difference between pre and post test
    - GM is then a equivalent to a *matched pair t test*

Testing Grand Mean

- $t = \overline{M} - C \Big/ \sqrt{\left(1/N\right) MS_{error}}$
  - C is the comparison score established on theoretical grounds
  - MS error is the estimate of the variation of scores within their experimental conditions, ie, MS within
  - Eg, is the grand mean significantly greater than 0
    - $t = 4.5 - 0 \Big/ \sqrt{\left(1/12\right) 2.5} = 9.86$
    - We use 0 for *C*
    - which with 8 *df* (the *df* for the MS error, ie MS within) is significant at *p* < .000005

Testing Grand Mean

- Eg, comparing our grand mean with that of a large norm group whose grand mean is 5
  - As above, use *C* is 5
  - $$t = 4.5 - 5 \Big/ \sqrt{\left(1/12\right)2.5} = -1.10$$
  - At 8 *df* has an associated *p* values of about .30 two-tailed – not a very significant difference
  - Also, the power on basis of 12 patients to the reject the null hypotheses is quite low unless the true effect size is quite large

Testing Grand Mean

- Eg, suppose compare to another grand mean of 3
  - $t = 4.5 - 3.0 / \sqrt{(1/12)2.5} = 3.29$
  - With 8 *df* is significant at about .01, two-tailed
  - Assumed that comparison scores are theoretical scores know exactly, not estimates
  - Suppose comparison mean is an estimate based on 6 patients (8 *df*, *p* = .10, two tailed)

$$t = \overline{M} - C / \sqrt{(1/N_{\overline{M}} + 1/N_C) MS_{error}}$$

$$t = 4.5 - 3.0 / \sqrt{(1/12 + 1/6)2.5} = 1.90$$

Testing Grand Mean

- Assume MS error is equivalent
- Assume degrees of freedom is the same
- Compute MS error pooled and then *t*

$$MS_{errorpooled} = \frac{df_1 MS_{error_1} + df_2 MS_{error_2}}{df_1 + df_2}$$

- *df* for *t* is the sum of the two *df* 's used

$$t = \overline{M} - C \Big/ \sqrt{\left(1/N_{\overline{M}} + 1/N_C\right) MS_{errorpooled}}$$

F Test

- $t^2 = F$ when a single *df* for numerator of *F*
  - *F* tests on the grand mean of any one study involve only a single *df* for the numerator
  - *t* test procedures can be employed as *F* test procedures
- If totals rather than means then
  - With 1 *df* for numerator, 8 for denominator, $p < .00001$
  - The square root of *F* is 9.86, the value of *t*

$$F = \left(\left(\sum X\right)^2 / N\right) \Big/ MS_{error}$$

$$F = (54)^2 / 12 \Big/ 2.5 = 243 / 2.5 = 97.20$$

Unequal Sizes

- For one-way or omnibus analysis of variance it does not matter if we have the same number of units per condition or not
- For two-way or higher order analysis must take special care when number varies from condition to condition
  - One possible approach:
    - discard units til all conditions are equal
    - Almost never justified

Unequal Sizes

- Multiple regression procedures available
  - Yield identical results when sample sizes equal
  - Vary substantially when samples sizes become increasingly unequal
  - procedure here represents yields closer to the "fully simultaneous multiple regression method" (FSMR) recommended by Overall et al 75
  - For factorial designs of any size, always having 2 levels per factor, yields results identical to FSMR

Unweighted Means Analysis

- Can be used for equal or unequal sample sizes
- Three simple steps
  - Compute a one-way analysis of variance on the *k* groups or conditions
  - Compute a two-way (or higher) analysis of variance on the *means* of all conditions
  - Compute error term required for the analysis in step 2 by multiplying the *MS error* from step 1 by 1 over the *harmonic mean* of the various sample sizes of the different conditions to scale down the *MS error*
- Step 1
  - PD: 9, 8, 7 mean 8
  - P: 6, 4, 2 mean 4
  - D: 5, 4, 3 mean 4
  - O: 4, 2, 0 mean 2
  - Between: *SS=57, df=3, MS=19.0, F=7.60, eta=.86, p=.01*
  - Within: *SS=20, df=8, MS=2.5*

Unweighted Means Analysis

- Step 2
  - $SS_{total} = \sum(M - \overline{M})^2$
  - $SS_{total} = (8-4.5)^2 + (4-4.5)^2 + (4-4.5)^2 + (2-4.5)^2 = 19$
  - $SS_{row} = \sum[c(M_r - \overline{M})^2]$ where c is # of columns
  - $SS_{row} = 2(6-4.5)^2 + 2(3-4.5)_2 = 9$
  - $SS_{column} = \sum[r(M_c - \overline{M})^2]$ where r is # of rows
  - $SS_{column} = 2(6-4.5)^2 + 2(3-4.5)_2 = 9$
  - $SS_{interaction} = SS_{total} - SS_{row} - SS_{column}$
  - $SS_{interaction} = 19 - 9 - 9 = 1$

Unweighted Means Analysis

- Step 3
  - $$\frac{1}{\bar{n}_h} = \frac{1}{k}\left(\frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_k}\right)$$
  - $$\frac{1}{\bar{n}_h} = \frac{1}{4}\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) = \frac{1}{3}$$
  - **New error term: 1/3 x 2.5 = 0.833**

Unweighted Means Analysis

- Table of variance

SS :	D=9	P=9	I=1
df :	D=1	P=1	I=1
MS :	D=9	P=9	I=1
F :	D=10.80	P=10.8	I=1.20
Eta :	D=.76	P=.76	I=.36
P :	D=.012	P=.012	I=.30

- Results of the analysis
  - Results of the *F* test and magnitude of *eta* and *p* are identical
  - Magnitude of SS and MS are smaller by the reciprocal of the harmonic mean
  - Ie, shrinks SS and MS but has no effect on effect size or significance

Unweighted Means Analysis

- Effects on *F* of unequal sample sizes
  - *F* increases (as did *t*) as the sample sizes become more nearly equal
  - 3,3,3,3 *MSerror=.833, F(1,8)=10.60, eta=.76, p=.012*
  - 2,2,4,4 *MSerror=.938, F(1,8)=9.59, eta=.74, p=.015*
  - 1,1,5,5 *MSerror=1.500, F(1,8)=6.00, eta=.65, p=.040*
  - 1,1,1,9 *MSerror=1.944, F(1,8)=4.63, eta=.61, p=.064*
  - *F* and *eta* both become smaller
    - *F* as much as 57%
    - *eta* as much as 20%
  - *p* becomes larger, ie less significant
    - From quite significant to "non-significance" ;-)
  - Can be even more extreme
    - 25,25,25,25 *F(1,96)=131.58, eta=.76*
    - 1,1,1,97 *F(1,96)=7.00, eta=.26*
    - *95% 66%*

Higher Order Factorial Designs

- So far dealt only with 2 way
- Suppose current example had been done twice, once for females, once for males
  - Benefits: more subjects, more comparisons
  - 2x2x2 factorial design = 2^3 factorial
  - 3 factors: drug, psychotherapy, gender
  - N = 2x2x2x3 = 24 if the same twice
  - MSerror = 2.5, adjustment factor 1/3 = .833

Higher Order Factorial Designs

- Improvement scores in 8 conditions
  - D: FP=10, FNP=5, MP=6, MNP=3
  - ND: FP=4, FNP=1, MP=4, MNP=3
- Gender x drug
  - D: F=7.5[2], M=4.5[2], Mean=6.0[4]
  - ND: F=2.5[2], M=3.5[2], Mean=3.0[4]
  - Mean: F=5.0[4], M=4.0[4], Mean=4.5[8]
- Gender x psychotherapy
  - P: F=7.0[2], M=5.0[2], Mean=6.0[4]
  - PD: F=3.0[2], M=3.0[2], Mean=3.0[4]
  - Mean: F=5.0[4], M=4.0[4], Mean=4.5[8]
- Drug x psychotherapy
  - D: P=8.0[2], NP=4.0[2], Mean=6.0[4]
  - ND: P=4.0[2], NP=2.0[2], Mean=3.0[4]
  - Mean: P=6.0[4], NP=3.0[4], Mean=4.5[8]

Higher Order Factorial Designs

- General strategy
  - Compute main effects first
  - Then two way interactions
    - Ie, residuals when two contributing main effects are subtracted from the variation in the two tables
  - Then the three way interactions
    - Ie, the residuals when the three main effects and the three two way interactions are subtracted from the total variation from the total variation among the 8 conditions

Higher Order Factorial Designs

- $$SS_{total} = \sum(M - \overline{M})^2$$
- $$SS_{total} = (10 - 4.5)^2 + \ldots + (3 - 4.5)^2 = 50$$
- $$SS_{gender} = \sum[dp(M_g - M^2)]$$
- $$SS_{gender} = [2 \times 2(5 - 4.5)^2] + [2 \times 2(4 - 4.5)^2] = 2$$
- $$SS_{drug} = \sum[gp(M_d - \overline{M})^2]$$
- $$SS_{drug} = [2 \times 2(6 - 4.5)^2] + [2 \times 2(3 - 4.5)^2] = 18$$
- $$SS_{psychotherapy} = \sum[gd(M_p - \overline{M})^2]$$
- $$SS_{psychotherapy} = [2 \times 2(6 - 4.5)^2] + [2 \times 2(3 - 4.5)^2] = 18$$

Higher Order Factorial Designs

- $$SS_{g\times d} = \sum[p(Mg_d - \overline{M})^2] - SS_g - SS_d$$
- $$SS_{g\times d} = 2(7.5-4.5)^2 + \ldots + 2(3.5-4.5)^2 - 2 - 18 = 8$$
- $$SS_{g\times p} = \sum[d(Mg_{dp} - \overline{M})^2] - SS_g - SS_p$$
- $$SS_{g\times p} = 2(7-4.5)^2 + \ldots + 2(3-4.5)^2 - 2 - 18 = 2$$
- $$SS_{d\times p} = \sum[g(Mg_{dp} - \overline{M})^2] - SS_d - SS_p$$
- $$SS_{g\times d} = 2(8-4.5)^2 + \ldots + 2(2-4.5)^2 - 18 - 18 = 2$$
- $$SS_{g\times d\times p} = SS_{total} - SS_g - SS_d - SS_p - SS_{g\times d} - SS_{g\times p} - SS_{d\times p}$$ $$SS_{g\times d\times p} = 50 - 2 - 18 - 18 - 8 - 2 - 2 = 0$$

Higher Order Factorial Designs

Unweighted means analysis

- G: SS=2, df=1, MS=2, F(1,16)=2.4, eta-.36, p=.14
- D: SS=18, df=1, MS=18, F(1,16)=21.61, eta-.76, p=.0003
- P: SS=18, df=1, MS=18, F(1,16)=21.61, eta-.76, p=.0003
- GxD: SS=8, df=1, MS=8, F(1,16)=9.60, eta-.61, p=.007
- GxP: SS=2, df=1, MS=2, F(1,16)=2.40, eta-.36, p=.14
- DxP: SS=2, df=1, MS=2, F(1,16)=2.40, eta-.36, p=.14
- GxDxP:SS=0, df=1, MS=8, F(1,16)=0.00, eta-.00, p=1.00
- Error Term df=16, MS=.833

Higher Order Factorial Designs

- Summary
  - Effect sizes of .76, .76, and .36 are identical in earlier two way analysis
  - *F* scores have all increases
  - *p* values are much smaller
    - As we would expect: study size increased
  - Tendency of gender to make some difference
    - gender and drug interactions significant
- Generalized strategy
  - Eg four way factorial
    - Construct all possible 2 and 3 way tables
    - Compute 4 main effects, 6 two way interactions, 4 three way interactions, and on four way interaction