Lecture 26: Repeated Measures

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

Use of Repeated Measures

- Between subjects designs
  - Sampling units only observed once
  - Variation based on individual differences *between* subjects
  - subjects *nested* with in their treatment conditions
- Within Subjects Designs
  - very efficient to administer two or more treatments to same sampling units
  - sampling units serve as their *own control*
  - subjects *crossed* by treatment conditions
  - the more correlated, the more advantageous this approach

Use of Repeated Measures

- Intrinsic nature of experiment might call for repeated measures type of design
  - effect of practice on learning a task
  - effects in a longitudinal study of development
  - series of tests or subtests for a variety of reasons
- Simplest type: subjects measured twice
  - compare scores under each condition
  - use non-independent *t* test to compare correlation

Computations

- Consider performance scores on repeated practice sessions

S1	0	7	3	mean 3.33
S2	1	7	4	mean 4.00
S3	3	8	5	mean 5.33
S4	4	8	6	mean 6.00
mean	2.0	7.5	4.5	GM 4.67

  - $$SS_{total} = \sum (X - \overline{M})^2$$
  - $$SS_{total} = (0 - 4.67)^2 + \ldots + (6 - 4.67)^2 = 76.67$$

Computations

- $SS_{row} = \sum[c(M_r - \overline{M})^2]$
- $SS_{row} = 3(3.33 - 4.67)^2 + \ldots + 3(6 - 4.67)^2 = 13.35$
- $SS_{column} = \sum[r(M_c - \overline{M})^2]$
- $SS_{column} = 4(2 - 4.67)^2 + \ldots + 4(4.5 - 4.67)^2 = 60.67$
- $SS_{interaction} = SS_{total} - SS_{row} - SS_{column}$
- $SS_{interaction} = 76.67 - 13.35 - 60.67 = 2.65$

Computations

- Sources of variation
  - Between subjects
    - *Ss* = 13.35, *df* = 3, *MS* = 4.45
  - Within subjects
    - Sessions: *SS* = 60.67, *df* = 2, *F* = 68.93, *eta* = .98 and $p < .001$
    - Sessions x subjects (error term for sessions effect): *SS* = 2.65, *df* = 6, *MS* = 0.44

Fixed and Random Effects

- Distinction to help us employ the appropriate error term
  - Fixed
    - Selected particular levels of the factor
    - Cannot generalize to other levels
    - Includes most factors involving experimental manipulations, various organismic variables and repeated measures factors
  - Random
    - Randomly sampled from population of levels
      - ✓ Most common is that of sampling units, especially people
- In previous example
  - If we regard between subjects as random
    - We can test its significance only very conservatively
  - If we regard it as fixed
    - Restrict inferences to this four subjects
    - Can test subject factors against sessions x subjects interactions
  - Will consider all combinations
    - Fixed and random
    - For between and within subjects factors

Fixed and Random Effects

- Examples
  - 4 countries as our between sampling units factor
    - If only interested in these 4, fixed
    - If view as a sample from which we want to generalize, random
  - Longitudinal design with a summary score for each country for each of 3 decades
    - Scores are repeated measures, or within sampling units factors
    - Regard as fixed if we have chosen them specifically
    - Regard as random if we view as samples from which to generalize
- General principle that helps in determining the appropriateness of the error term
  - The effect (fixed or random) we to test are properly tested by dividing *MS* for that effect by the *MS* for a random source of variation

Error Terms, 4 Designs

- Type A (between and within fixed)
  - Interaction *MS* can be used as the error term for *MS* between and *MS* within
  - Likely to lead to *F* 's that are two conservative
  - *F* accurate only when a zero interaction effect
  - KEYS used subsequently
    - B = between subjects *MS*,
    - W = within subjects *MS*,
    - BW = between x within subjects interaction *MS*,
    - O = ordinary error
  - BW used when O is not available

Error Terms, 4 Designs

- ***Between countries:***
  - computed as the row effect of any two way factorial design
- ***Within countries:***
  - *decades* – computed as the column effect
- ***Within countries:***
  - *decades x countries* – computed as the interaction effect
- ***Within countries:***
  - *years within decade x country combinations* – computed as the within cell error
- **Illustration of type A design**

	abbr	error	err/cons
Between countries:	B	O	BW
Within countries			
Decades	W	O	BW
Decades x countriesBW	O		
years within d x o	O		

Error Terms, 4 Designs

- Type B (Between fixed, within random)
  - Interaction *MS* is appropriate error term for between units effect
  - Only for within or repeated measure effects if interaction is zero
  - Appropriate error term for within subjects effect is the variation of the multiple observations made for each combination of row and column
  - Illustration of type B design

	abbr	error	err/cons
Between countries:	B	BW	
Within countries			
Decades	W	O	BW
Decades x countriesBW	O		
years within BW	O		

Error Terms, 4 Designs

- Type C (between random, within fixed)
  - Interaction MS is the appropriate error term for within sampling units effect
  - Appropriate for between units effect only if interaction effect is really zero
  - Appropriate error term is the variation of the multiple observations for each combination of row and column
  - Illustration of type C design

	abbr	error	err/cons
Between countries:	B	O	BW
Within countries			
Decades	W	BW	
Decades x countriesBW	O		
years within BW	O		

Error Terms, 4 Designs

- Type D (between and within random)
  - Interaction MS is the appropriate error term for both between and within subject effects
  - Interaction effect could be tested against the variation of multiple observations made for each combination of row and column
  - Illustration of type D design

	abbr	error	err/cons
Between countries:	B	BW	
Within countries			
Decades	W	BW	
Decades x countriesBW	O		
years within BW	O		

Latin Squares

- Consider three drugs and 4 patients
  - Suppose each subject given three drugs in the same sequence
    - Confound drug and order
    - Suppose A is best
    - Rival hypothesis is the first is best
  - Use *counterbalancing* to avoid confounding
    - Sequence is systematically varied
    - Essential in organization and sequencing presentation
      - ✓ *Primacy*: opinions influenced by arguments presented 1st
      - ✓ *Recency*: opinions influenced by what is presented last

Latin Squares

- Latin squares has counterbalancing built in
  - Nr of rows equals the nr of columns
  - The letter presenting treatments appears in each column and row only once
  - Effects of treatment, order and sequence are isolated – systematic counterbalancing

Order	1	2	3
seq 1	A	B	C
seq 2	B	C	A
seq 3	C	A	B

Latin Squares

- Analysis
  - Sequence effects tell how sequencings differ
  - Order effects tell how orders differ
  - Where is the treatment effect in latin squares?
    - In a 2 x 2 latin square, treatment effect is the sequence x order interaction effect
    - Compares the two diagonals
    - Sources of variation:
      - ✓ Sequences: 1 *df*
      - ✓ Orders: 1 *df*
      - ✓ Treatments (s x o): 1 *df*
  - Significance testing is a problem
    - No *df* available for error terms for the 3 sources of variation
    - Can use the mean of the *MS* 's of the other two in computing very conservative *F* 's
  - As the size of the latin square increases, the df for an error term increases
    - Get less conservative *F* 's
    - More accurate in the sense of less Type II errors
    - On average *F* 's will be too small

Latin Squares

- Larger squares
  - Sources of variation
    - Sequences: $a - 1$ *df*
    - Orders: $a - 1$ *df*
    - (sequences x orders): $((a - 1)^2)$ *df*
      - ✓ Treatments: $a - 1$ *df*
      - ✓ Residual sequences x orders: $(a - 1)(a - 2)$ *df*
- Computations
  - M_s M_o M_t for mean sequences, orders and treatments
- Sources of variation – *df* :

	3x3	4x4	5x5
sequences	2	3	4
orders	2	3	4
(s x o)	(4)	(9)	(16)
treatments	2	3	4
residual s x o	2	6	12

Latin Squares

- Compute *SS* 's as follows
  - $$SS_{total} = \sum(X - \overline{M})^2$$
  - $$SS_{sequences} = \sum[a(M_s - \overline{M})^2]$$
  - $$SS_{orders} = \sum[a(M_o - \overline{M})^2]$$
  - $$SS_{s \times o} = SS_{total} - SS_{sequences} - SS_{orders}$$
  - $$SS_{treatment} = \sum[a(M_t - \overline{M})^2]$$
  - $$SS_{residual} = SS_{s \times o} - SS_{treatment}$$

Latin Squares

→ Example

	O1	O2	O3	O4	R mean
S1 (ABCD)	5	3	8	5	5
S2 (BCDA)	0	6	7	7	5
S3 (CDAB)	2	2	10	10	6
S4 (DABC)	6	5	7	14	8
C mean	3	4	8	9	6

Latin Squares

- Compute *SS* 's
  - $SS_{total} = (4-6)^2 + (3-6)^2 + \ldots + (14-6)^2 = 186$
  - $SS_{sequences} = 4(5-6)^2 + \ldots + 4(8-6)^2 = 24$
  - $SS_{orders} = 4(3-6)^2 + \ldots + 4(9-6)^2 = 104$
  - $SS_{s \times o} = 186 - 24 - 104 = 58$

- Treatments

	TA	TB	TC	TD	R mean
S1	4	3	8	5	5
S2	7	0	6	7	5
S3	10	10	2	2	6
S4	5	7	14	6	8
C mean	6.5	5.0	7.5	5.0	6

Latin Squares

- Remaining *SS* 's
  - $$SS_{treatment} = 4(6.5-6)^2 + \ldots + 4(5.0-6)^2 = 18$$
  - $$SS_{residual} = 58 - 18 = 40$$
- Table of variance

	SS	*df*	*MS*	*F*	*eta*	*p*
Seq's	24	3	8.00	1.2	.61	.39
Orders	104	3	34.67	5.2	.85	.042
SxO	(58)	(9)				
T's	18	3	6.00	0.90	.56	.49
R sxo	40	6	6.67			

Other Counterbalancing Designs

- What if we have an unequal number of subjects and treaments?
- Useful strategies
  - Multiple squares
  - Rectangular arrays
- Rectangular arrays
  - Eg, 3 treatments and 6 subjects
    - Could do two 3x3 squares
    - Could assign each subject a unique sequence of treatments since 3! = 6
  - If 4 treatments, then would need 4! = 24 subjects
  - Call such designs *t* x *t*! designs (*t* = 2 is latin squares)
  - Fewer than *t*! sample units?
    - Multiple latin squares
    - Random assignment from *t*! Sequences
      - ✓ Constraint: ensure maximum degree of balancing in the resulting samples
  - More than *t*! Sample units?
    - Multiple rectangular arrays
    - Subjects-within sequences designs

Other Counterbalancing Designs

- Subjects-within sequences Designs
  - Suppose 2 x *t* ! Subjects
    - Could randomly assign half the subjects to each of two rectangular arrays
      - ✓ Treat each array as a different and replicated experiment
    - Could randomly assign from *t* ! Sequences
      - ✓ Constraint: ensure maximum counterbalancing
  - Suppose 18 subjects, 3 treatments
    - 3! = 6, assign 3 subjects at random to each sequence
    - Subjects are *not* confounded with sequence as in latin squares
    - Subjects are *nested* within sequences so sequences can be tested

Other Counterbalancing Designs

- Sources of variance: subject within sequences

	df	test against
Between subjects:	(17)	
sequences	5	subjects
subjects within sequences	12	
Within subjects:	(36)	[*N x df* r'd measures]
orders	2	o x subjects
orders x sequences	10	usually not tested
treatments	2	o x subjects
residual orders x seq's	8	o x subjects
orders x subj's within seq's	24	

Other Counterbalancing Designs

- Noteworthy features of this design
  - More than a single error term in the design
    - ✓ Previously only one error term
    - ✓ Usually associate with differences among subjects
    - ✓ Have that: subjects within sequences
      - Used to test whether sequences differ from each other
      - Note: error is within *conditions* but *between* subjects
    - ✓ Other error term: orders x subjects within-sequences interaction
      - Test all within-subjects sources of variation
      - Is itself a within-subjects source of variation
      - Formed by crossing repeated measures factors by the random factor of sampling units

Other Counterbalancing Designs

- To test for treatments, we must reach into the order x sequences interactions and pull out the variation of the treatment means around the grand mean
- Analysis – pretty much same as already discussed
  - ✓ Between-subjects SS is broken down into a sequences SS and a subjects-within-sequences SS
  - ✓ Later is the difference between the between-subjects SS and the sequences SS
  - ✓ etc

3 or more Factors

- So far considered only two factors
  - Between-subjects factor
  - Within-subjects or repeated measures factor
  - Often have two or more of each
- 2 or more between-subjects factors
  - Does not increase complexity of the design as much increasing the number of within-subjects factors
  - Eg, 4 subtests of personality test of three age levels and two genders

3 or more Factors

- Assume two subjects for each of 3 x 2 = 6 between-subjects conditions

Source	*df*
Between subjects:	(11)
age	2
gender	1
age x gender	2
subjects (within conditions)	6

Source	*df*
Within subjects:	(36)
subtests	3
subtests x age	6
subtests x gender	3
subtests x age x gender	6
subtests x subjects (within)	18

Subtests x between subjects = s x a, s x g, s x a x g, and s x s

3 or more Factors

- Computation straight forward
  - Think of it as a 12 x 4 measurement array
  - Compute all between subject *SS* 's
  - The age, gender, age x gender *SS* 's and subtract from the total between subjects *SS* to give us the subjects within conditions *SS*
  - *df* for each source becomes more useful as designs become more complex
    - ✓ A check on whether have left out any source of variance
    - ✓ Check to make sure all components of betweena nd within are accounted for
      - Between subjects: 2 + 1 + 2 + 6 = 11

3 or more Factors

Age	gender	subj	st1	st2	st3	st4	mean
12	f	1	2	3	7	8	5.0
12	f	2	1	2	3	6	3.0
12	m	3	1	2	1	4	2.0
12	m	4	1	2	1	4	2.0
14	f	5	5	4	7	8	6.0
14	f	6	4	5	8	7	6.0
14	m	7	1	2	4	5	3.0
14	m	8	1	4	6	9	5.0
16	f	9	5	9	9	9	8.0
16	f	10	6	5	8	9	7.0
16	m	11	5	6	9	8	7.0
16	m	12	4	5	7	8	6.0
Mean `		3.0	4.0	6.0	7.0	5.0	

3 or more Factors

	Gender		
Age	female	male	mean
12	4.0[8]	2.0[8]	3.0[16]
14	6.0[8]	4.0[8]	5.0[16]
16	7.5[8]	6.5[8]	7.0[16]
Mean	5.83[24]	4.17[24]	5.0[48]

3 or more Factors

- $SS_{total} = \sum(X - \overline{M})^2 = (2-5)^2 + \ldots + (8-5)^2 = 340$
- $SS_{row(subject)} = \sum[c(M_r - \overline{M})^2] = 4(5.0-5)^2 + \ldots + 4(6.0-5)^2 = 184$
- $SS_{column(r-ms)} = \sum[r(M_c - \overline{M})^2] = 12(3.0-5)^2 + \ldots + 12(7.0-5)^2 = 120$
- $SS_{r \times c\ interaction} = SS_{total} - SS_{row} - SS_{column} = 340 - 184 - 120 = 36$
- $SS_{age} = \sum[ngt(M_a - \overline{M})^2] = [2x2x4(3.0-5)^2 + \ldots = 128$
- $SS_{gender} = \sum[nat(M_g - \overline{M})^2] = [2x3x4(5.83-5)^2 + \ldots = 33$
- $SS_{age \times gender} = \sum[nt(M_{ag} - \overline{M})^2] - SS_{age} - SS_{gender}$
  $= [2 \times 4(4.0-5)^2 + \ldots = 164 - 128 - 33 = 3$
- $SS_{subjects-within} = SS_{row} - SS_{age} - SS_{gender} - SS_{age \times gender}$
  $= 184 - 128 - 33 - 3 = 20$

3 or more Factors

Means: subtests x age

	Subtests				
Age	1	2	3	4	Mean
12	1.25[4]	2.00	3.50	5.25	3.00[14]
14	2.75	3.75	6.25	7.25	5.00
16	5.00	6.25	8.25	8.50	7.00
Mean	3.00[12]	4.00	6.00	7.00	5.00[48]

3 or more Factors

Means: subtests x gender

	Subtests				
Gender	1	2	3	4	Means
female	3.83[6]	4.67	7.00	7.83	5.83[24]
male	2.17	3.33	5.00	6.17	4.17
Mean	3.00[12]	4.00	6.00	7.00	5.00[48]

3 or more Factors

Means: subtests x age x gender

		subtests				
Age	gender	1	2	3	4	Mean
12	female	1.5[2]	2.5	5.0	7.0	4.0[8]
12	male	1.0	1.5	2.0	3.5	2.0
14	female	4.5	4.5	7.5	7.5	6.0
14	male	1.0	3.0	5.0	7.0	4.0
16	female	5.5	7.0	8.5	9.0	7.5
16	male	4.5	5.5	8.0	8.0	6.5
Mean		3.0[12]	4.0	6.0	7.0	5.0[48]

3 or more Factors

- $$SS_{subtest\times age} = \sum[ng(M_{ta} - \overline{M})^2] - SS_{subtest} - SS_{age}$$
  $$= [2\times 2(1.25-5)^2 + \ldots = 252 - 120 - 128 = 4$$
- $$SS_{subtest\times gender} = \sum[na(M_{tg} - \overline{M})^2] - SS_{subtest} - SS_{gender}$$
  $$= [2\times 3(3.83-5)^2 + \ldots = 154 - 120 - 33 = 1$$
- $$SS_{subtest\times age\times gender} = \sum[n(M_{tag} - \overline{M})^2] - SS_{subtest} - SS_{age} - SS_{gender}$$
  $$- SS_{subtest\times age} - SS_{subtest\times gender} - SS_{age\times gender}$$
  $$= 2(1.5-5)^2 + \ldots = 300 - 120 - 128 - 33 - 4 - 1 - 3 = 11$$
- $$SS_{subtest\times subjects-within} = SS_{row\times column} - SS_{subtest-age} - SS_{subtest\times gender}$$
  $$- SS_{subtest\times age\times gender}$$
  $$= 36 - 4 - 1 - 11 = 20$$

3 or more Factors

Source	*SS*	*df*	*MS*	*F*	*eta*	*p*
Between subj's	(184)	(11)				
age	128	2	64.00	19.22	.93	.003
gender	33	1	33.00	9.91	.79	.02
a x g	3	2	1.5	0.45	.36	.66
subj's (w)	20	6	3.33			
Within subj's	(156)	(36)				
subtests	120	3	40.00	36.04	.93	.0001
s x a	4	6	.67	0.60	.41	.73
s x g	1	3	.33	0.30	.22	.82
s x a x g	11	6	1.83	1.65	.60	.19
s x subj's (w)	20	18	1.11			

3 or more Factors

- **Summary**
  - **All three main effects were significant**
  - **Interactions**
    - ✓ None were significant
    - ✓ Though several effect sizes (*eta*) were substantial
    - ✓ Suggests replications with larger sample sizes might reach statistical significance
- **Use equal sample sizes – with unequal**
  - **Use the harmonic mean instead**
  - **Ie, replace *n* with**

$$\bar{n}_h = 1/[1/k(1/n_1 + \ldots + 1/n_k)]$$

Fixed vs Random Factors

- Examples so far have assumed that all factors other than subjects within conditions have been fixed
  - The most common situation
- Example:
  - 5 f and 5 m teachers, 4 schools, teach lesson to 3 pupils, one designated as bright
  - If fixed, 2 error terms
  - If school is random factor, 5 error terms

Do repeated measures help?

- Basic utility: subjects are their “own control”
- High correlation yields advantages
- If low correlations, little advantage

Note on Assumptions

- For *F* and *t* tests
  - Independence of errors
  - Homogeneity of variance
  - Normality
- Additional needed for repeated measures
  - Homogeneity of correlation coefficients among the various levels of the repeated measures factors
  - Patterns of inter-correlations is consistent among various levels is consistent from level to level