Repeated-Measures ANOVA

PSYC 300B - Lecture 2
Dr. J. Nicol

Repeated-Measures ANOVA

- In a repeated-measures design we obtain scores from each participant on more than two occasions
- A repeated-measures ANOVA is like a dependent *t*-test that is used when more than two means are being compared
- Comparisons are made of scores across time within the same subject—so often referred to as a within-subjects design

Advantages and Disadvantages

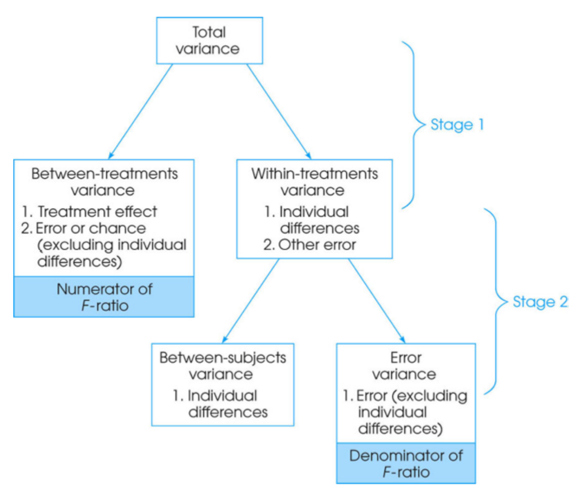
- The primary advantage of a repeated-measure design is that it reduces or eliminates problems caused by individual differences
- Repeated-measures ANOVA allows us to assess individual differences and separate them from the error term (individual differences are a type of systematic variance)
- By removing individual differences, we reduce unsystematic variance (denominator) and in turn increase the chance of finding a significant result
- A repeated-measures design typically requires fewer subjects than a between-groups design

Advantages and Disadvantages

- The repeated-measures design is especially well suited for studying development, or other changes that take place over time (e.g., longitudinal studies)
- Practice effects: participants may gain experience during the first treatment condition, and this may help their performance in subsequent treatments
- One way to deal with time-related factors and order effects is to counterbalance the order of presentation of treatments
- In most cases the advantages outweigh the disadvantages (Howell, 2017)

Assumptions

- Normality within conditions
- Sphericity (homogeneity of variance)—correlations among pairs of conditions of the independent variable are constant
- This is a rather stringent assumption and one that probably is violated at least as often as it is met, and the test is not seriously affected unless the assumption is seriously violated (Howell, 2017)



Calculating Sum of Squares

$$SS_{TOTAL} = \sum (X - M_{GRAND})^2$$

$$SS_{BETWEEN} = n \sum (M_{GROUP} - M_{GRAND})^2$$

$$SS_{SUBJECTS} = k \sum \left(M_{SUBJECT} - M_{GRAND}\right)^2$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{BETWEEN} - SS_{SUBJECTS}$$

Calculating Degrees of Freedom

$$df_{TOTAL} = N - 1$$

$$df_{BETWEEN} = k - 1$$

$$df_{SUBJECTS} = n - 1$$

$$df_{ERROR} = df_{TOTAL} - df_{BETWEEN} - df_{SUBJECTS}$$

Calculating Mean Squares

$$MS_{BETWEEN} = \frac{SS_{BETWEEN}}{df_{BETWEEN}}$$

$$MS_{ERROR} = \frac{SS_{ERROR}}{df_{ERROR}}$$

Calculating the *F*-Ratio

$$F = \frac{MS_{BETWEEN}}{MS_{ERROR}}$$

$$F_{CRITICAL} = (df_{BETWEEN}, df_{ERROR})$$

Post-Hoc Tests

- When the assumption of sphericity is not violated Tukey's *HSD* can be used (Field, 2013)
- For a repeated-measures ANOVA, MS_{ERROR} is a correct estimate of the standard error, so we can run the Fisher's *LSD* as if the means were from independent samples (Howell, 2014)

Post-Hoc Tests

$$Fisher's\ LSD:\ t = \frac{M_1 - M_2}{\sqrt{\left(\frac{MS_{ERROR}}{n_1} + \frac{MS_{ERROR}}{n_2}\right)}}$$

$$Tukey's\ HSD = q\sqrt{\frac{MS_{ERROR}}{n}}$$

Effect Size Measures

$$\eta^2 = \frac{SS_{BETWEEN}}{SS_{TOTAL} - SS_{SUBJECTS}}$$

$$\omega^2 = \frac{\left[\frac{k-1}{nk}(MS_{BETWEEN} - MS_{ERROR})\right]}{MS_{ERROR} + \frac{MS_{SUBJECTS} - MS_{ERROR}}{k} + \left[\frac{k-1}{nk}(MS_{BETWEEN} - MS_{ERROR})\right]}$$

$$d = \frac{M_1 - M_2}{\sqrt{MS_{ERROR}}}$$

Conduct a hypothesis test (α = .05) to determine if there is a significant treatment effect. Perform post-hoc tests and report effect sizes

Pre	Post	Follow-up
6	8	10
5	5	5
1	2	3
0	1	2

Pre	Post	Follow-up	
6	8	10	$M_{S1} = 8$
5	5	5	$M_{S2} = 5$
1	2	3	$M_{S3} = 2$
0	1	2	$M_{S4} = 1$
$M_{PRE} = 3$	$M_{POST} = 4$	$M_{FOLLOW} = 5$	$M_{GRAND} = 4$

- H_0: $\mu_{TIME1} = \mu_{TIME2} = \mu_{TIME3}$
- H_1: not all the μ's are equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 102$
- $SS_{BETWEEN} = n\sum(M_{GROUP} - M_{GRAND})^2 = 4[(3\text{-}4)^2 + (4\text{-}4)^2 + (5\text{-}4)^2] = 8$
- $SS_{SUBJECTS} = k\sum(M_{SUBJECT} - M_{GRAND})^2$
  - $= 3[(8\text{-}4)^2 + (5\text{-}4)^2 + (2\text{-}4)^2 + (1\text{-}4)^2] = 90$
- $SS_{ERROR} = 102 - 8 - 90 = 4$
- $df_{TOTAL} = N - 1 = 12 - 1 = 11$
- $df_{BETWEEN} = k - 1 = 3 - 1 = 2$
- $df_{SUBJECTS} = n - 1 = 3$
- $df_{ERROR} = 11 - 2 - 3 = 6$

- $MS_{BETWEEN} = SS_{BETWEEN}/df_{BETWEEN} = 8/2 = 4$
- $MS_{ERROR} = SS_{ERROR}/df_{ERROR} = 4/6 = 0.67$
- $F = MS_{BETWEEN}/MS_{ERROR} = 4/0.67 = 6$
- $F\,(2, 6)_{CRITICAL} = 5.14$
- Reject H_0 and conclude there is a significant treatment effect across time, $F\,(2, 6) = 6.0$, $p < .05$, $\eta^2 = 0.67$

Source	*SS*	*df*	*MS*	*F*
Between	8	2	4	6
Within				
Subjects	90	3		
Error	4	6	0.67	
Total	102	11		

- k = 3 and df_{ERROR} = 6, so q = 4.34
- $HSD = 4.34\sqrt{(0.67/4)} = 1.78$
- $M_{PRE} - M_{POST} = 3 - 4 = 1$ ($p > .05$)
- $M_{PRE} - M_{FOLLOW\text{-}UP} = 3 - 5 = 2$ ($p < .05$); $d = 2/\sqrt{0.67} = 2.44$
- $M_{POST} - M_{FOLLOW\text{-}UP} = 4 - 5 = 1$ ($p > .05$)
- Scores were significantly higher at follow-up compared to pre-treatment

Conduct a hypothesis test (α = .05) to determine if spelling errors decrease as students progress through grades. Perform post-hoc tests and report effect sizes

4th Grade	5th Grade	6th Grade
4	3	1
8	6	4
5	3	3
7	4	2
6	4	0

4th Grade	5th Grade	6th Grade	
4	3	1	$M_{S1} = 2.67$
8	6	4	$M_{S2} = 6.0$
5	3	3	$M_{S3} = 3.67$
7	4	2	$M_{S4} = 4.33$
6	4	0	$M_{S5} = 3.33$
$M_{4TH} = 6.0$	$M_{5TH} = 4.0$	$M_{6TH} = 2.0$	$M_{GRAND} = 4.0$

- H_0: $\mu_{4TH} = \mu_{5TH} = \mu_{6TH}$
- H_1: not all the μs are equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 66$
- $SS_{BETWEEN} = n\sum(M_{GROUP} - M_{GRAND})^2 = 5[(6\text{-}4)^2 + (4\text{-}4)^2 + (2\text{-}4)^2] = 40$
- $SS_{SUBJECTS} = k\sum(M_{SUBJECT} - M_{GRAND})^2 = 3[(2.67\text{-}4)^2 + (6\text{-}4)^2 + (3.67\text{-}4)^2 + (4.33\text{-}4)^2 + (3.33\text{-}4)^2] = 19.33$
- $SS_{ERROR} = 66 - 40 - 19.33 = 6.67$
- $df_{TOTAL} = N - 1 = 15 - 1 = 14$
- $df_{BETWEEN} = k - 1 = 3 - 1 = 2$
- $df_{SUBJECTS} = n - 1 = 4$
- $df_{ERROR} = 14 - 2 - 4 = 8$

- $MS_{BETWEEN} = SS_{BETWEEN}/df_{BETWEEN} = 40/2 = 20$
- $MS_{ERROR} = SS_{ERROR}/df_{ERROR} = 6.67/8 = 0.83$
- $F = MS_{BETWEEN}/MS_{ERROR} = 20/0.83 = 24.1$
- $F\,(2, 8)_{CRITICAL} = 4.46$
- Reject H_0 and conclude there is a significant effect of grade on spelling performance, $F\,(2, 8) = 24.1$, $p < .05$, $\eta^2 = 0.86$

Source	*SS*	*df*	*MS*	*F*
Between	40	2	20	24.1
Within				
Subjects	19.33	4		
Error	6.67	8	0.83	
Total	66.0	14		

- t (8) critical = ±2.306
- 4th vs. 5th:
  - $t = 6.0 - 4.0/\sqrt{(0.83/5+0.83/5)} = 3.45$, $p < .05$
  - $d = 2/\sqrt{0.83} = 2.19$
- 4th vs. 6th:
  - $t = 6.0 - 2.0/\sqrt{(0.83/5+0.83/5)} = 6.89$, $p < .05$
  - $d = 4/\sqrt{0.83} = 4.38$
- 5th vs. 6th:
  - $t = 4.0 - 2.0/\sqrt{(0.83/5+0.83/5)} = 3.45$, $p < .05$
  - $d = 2/\sqrt{0.83} = 2.19$
- All pairwise comparisons were significant

- $k = 3$ and $df_{ERROR} = 8$, so $q = 4.04$
- $HSD = 4.04\sqrt{(0.83/5)} = 1.65$
- $M_{4TH} - M_{5TH} = 6 - 4 = 2$ ($p < .05$); $d = 2/\sqrt{0.83} = 2.19$
- $M_{4TH} - M_{6TH} = 6 - 2 = 4$ ($p < .05$); $d = 4/\sqrt{0.83} = 4.38$
- $M_{5TH} - M_{6TH} = 4 - 2 = 2$ ($p < .05$); $d = 2/\sqrt{0.83} = 2.19$

Conduct a hypothesis test (α = .05) to determine if manner of dress affects how comfortable a person is in a social setting. Perform post-hoc tests and report effect sizes

Casual	Semiformal	Formal
4	9	1
6	12	3
8	4	4
2	8	5
10	12	2

Casual	Semiformal	Formal	
4	9	1	$M_{S1} = 4.67$
6	12	3	$M_{S2} = 7.0$
8	4	4	$M_{S3} = 5.33$
2	8	5	$M_{S4} = 5.0$
10	12	2	$M_{S5} = 8.0$
$M_C = 6$	$M_S = 9$	$M_F = 3$	$M_{GRAND} = 6.0$

- H_0: $\mu_{CASUAL} = \mu_{SEMI} = \mu_{FORMAL}$
- H_1: not all the μs are equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 184$
- $SS_{BETWEEN} = n\sum(M_{GROUP} - M_{GRAND})^2 = 5[(6-6)^2 + (9-6)^2 + (3-6)^2] = 90$
- $SS_{SUBJECTS} = k\sum(M_{SUBJECT} - M_{GRAND})^2$
    - $= 3[(4.67-6)^2 + (7-6)^2 + (5.33-6)^2 + (5-6)^2 + (8-6)^2] = 24.67$
- $SS_{ERROR} = 184 - 90 - 24.67 = 69.33$
- $df_{TOTAL} = N - 1 = 15 - 1 = 14$
- $df_{BETWEEN} = k - 1 = 3 - 1 = 2$
- $df_{SUBJECTS} = n - 1 = 4$
- $df_{ERROR} = 14 - 2 - 4 = 8$

- $MS_{BETWEEN} = SS_{BETWEEN}/df_{BETWEEN} = 90/2 = 45$
- $MS_{ERROR} = SS_{ERROR}/df_{ERROR} = 69.33/8 = 8.67$
- $F = MS_{BETWEEN}/MS_{ERROR} = 45/8.67 = 5.19$
- $F\ (2, 8)_{CRITICAL} = 4.46$
- Reject H_0 and conclude that manner of dress has a significant effect on social comfort, $F\ (2, 8) = 5.19$, $p < .05$, $\omega^2 = 0.48$

Source	*SS*	*df*	*MS*	*F*
Between	90.0	2	45.0	5.19
Within				
Subjects	24.67	4		
Error	69.33	8	8.67	
Total	184.0	14		

- $k = 3$ and $df_{ERROR} = 8$, so $q = 4.04$
- $HSD = 4.04\surd(8.67/5) = 5.31$
- $M_{CASUAL} - M_{SEMI} = 6 - 9 = 3$ ($p > .05$)
- $M_{CASUAL} - M_{FORMAL} = 6 - 3 = 3$ ($p > .05$)
- $M_{SEMI} - M_{FORMAL} = 9 - 3 = 6$ ($p < .05$); $d = 6/\surd 8.67 = 2.04$
- Participants were significantly more comfortable attending the party in semi-formal clothing than formal clothing

Conduct a hypothesis test (α = .01) to determine if there is an effect of study strategy on test performance. Perform post-hoc tests and report effect sizes

Reread	Prep. Questions	Create Questions
2	5	8
3	9	6
8	10	12
6	13	11
5	8	11
6	9	12

Reread	Prep. Quests	Create Quests	
2	5	8	$M_{S1} = 5.0$
3	9	6	$M_{S2} = 6.0$
8	10	12	$M_{S3} = 10.0$
6	13	11	$M_{S4} = 10.0$
5	8	11	$M_{S5} = 8.0$
6	9	12	$M_{S6} = 9.0$
M_{READ}=5.0	M_{PREP}=9.0	M_{CREATE}=10.0	M_{GRAND} = 8.0

- H_0: $\mu_{REREAD} = \mu_{PREPARED} = \mu_{CREATED}$
- H_1: not all the μs are equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 172.0$
- $SS_{BETWEEN} = n\sum(M_{GROUP} - M_{GRAND})^2 = 6[(5\text{-}8)^2 + (9\text{-}8)^2 + (10\text{-}8)^2] = 84$
- $SS_{SUBJECTS} = k\sum(M_{SUBJECT} - M_{GRAND})^2$
  - $= 3[(5\text{-}8)^2 + (6\text{-}8)^2 + (10\text{-}8)^2 + (10\text{-}8)^2 + (8–8)^2 + (9\text{-}8)^2] = 66.0$
- SS_{ERROR} = 172.0 - 84.0 - 66.0 = 22.0
- $df_{TOTAL} = N - 1 = 18 - 1 = 17$
- $df_{BETWEEN} = k - 1 = 3 - 1 = 2$
- $df_{SUBJECTS} = n - 1 = 5$
- df_{ERROR} = 17 - 2 - 5 = 10

- $MS_{BETWEEN} = SS_{BETWEEN}/df_{BETWEEN} = 84/2 = 42.0$
- $MS_{ERROR} = SS_{ERROR}/df_{ERROR} = 22/10 = 2.20$
- $F = MS_{BETWEEN}/MS_{ERROR} = 42/2.20 = 19.09$
- $F\ (2, 10)_{CRITICAL} = 7.56$
- Reject H_0 and that study strategy has a significant effect on test performance, $F\ (2, 10) = 19.09$, $p < .01$, $\eta^2 = 0.79$

Source	*SS*	*df*	*MS*	*F*
Between	84.0	2	42.0	19.09
Within				
Subjects	66.0	5		
Error	22.0	10	2.20	
Total	172.0	14		

- $k = 3$ and $df_{ERROR} = 10$, so $q = 5.27$
- $HSD = 5.27\sqrt{(2.20/6)} = 3.19$
- $M_{READ} - M_{PREP} = 5-9 = 4$ ($p < .01$); $d = 4/\sqrt{2.2} = 2.70$
- $M_{READ} - M_{CREATE} = 5-10 = 5$ ($p < .01$); $d = 5/\sqrt{2.2} = 3.37$
- $M_{PREP} - M_{CREATE} = 9 - 10 = 1$ ($p > .01$)
- Performance was significantly better using the prepare questions strategy than the rereading strategy, and significantly better using the create question strategy than the rereading strategy

End of Lecture