PSY 3307

Repeated Measures ANOVA

Jonathan A. Pedroza PhD

Cal Poly Pomona

2021-11-04

Agenda

- Within-Subjects Design
  - Repeated Measures ANOVA
  - Differences Between Paired-Samples t-test & Repeated Measures ANOVA
- Longitudinal Designs
- Why They Are Useful
- Assumptions of Repeated Measures ANOVA
- Calculations for One-way Within-Subjects ANOVA
  - Repeated Measures ANOVA
- Repeated Measures in SPSS

Previously With ANOVA

- IV should be two more more groups
  - preferably three groups or more
- no outliers
  - **Cook's distance** is a method to find influential cases/outliers in your model
  - used for both ANOVA and regression models
- **homogeneity of variance** is the assumption that each population has the same variance
- **normality** DV values are normally distributed
- **independence** observations are independent of one another
  - it really is that the residual/error is independent but for now we'll keep it as observations are different from one another

Within-Subjects Design

- Every participant gets each condition
  - the way we will be using this design is with a **repeated measures** design
  - or that every participant will see the same measure multiple times
- An example would be to look at a depression scale before the intervention, after the intervention, and then 6 months after the end of the intervention
- Outside a researcher perspective, we could think of this as a program evaluator
  - you could be evaluating a community on what they think about their safety
  - so you can ask before, during the program promoting safety in the community, and months afterward to see if your program is still working

Comparisons to paired-samples t-test

- Repeated measures is similar to the paired-samples t-test in that we can test multiple time points for every participant
- however, we can test more than two time points with repeated measures ANOVA
- we can also looked at mixed designs (between-subjects x within-subjects)
    - we'll get to this soon

Longitudinal Designs

- this is a major reason why this statistical test is so useful
    - this test accounts for independence of observations/residuals
- **longitudinal designs** are when we are following our sample for a period of time
    - these designs get much more complex than this (same group of people, similar groups over different years)

Repeated Measures ANOVA

- no outliers
  - **Cook's distance** is a method to find influential cases/outliers in your model
  - used for both ANOVA and regression models
- **homogeneity of variance** is the assumption that each population has the same variance
- **normality** DV values are normally distributed
- **independence** is not observed for repeated measures ANOVA because the test accounts for repeating values (same construct different time points)

Repeated Measures ANOVA

- the lack of independence would be a problem but repeated measures ANOVA **partitions** or removes the dependence imposed by multiple measurements on the same participants
- this is due by partialling out or by getting rid of the overlap in the proportion of variance explained
  - we'll get to this when we get to regression
- normal linear regression would actually be worse than repeated measures ANVOA in this case

When won't it work

- missing data in the DV
- unbalanced amount of participants
- time is continuous
- covariates that vary by time
- nested/hierarchical models (students in classrooms in schools)
- non-continuous DV

Example

```r
hungry <- data.frame(time1 = c(9, 7, 8, 6, 10),
                     time2 = c(4, 5, 4, 3, 5),
                     time3 = c(6, 5, 5, 6, 7))
hungry
```

```
##   time1 time2 time3
## 1     9     4     6
## 2     7     5     5
## 3     8     4     5
## 4     6     3     6
## 5    10     5     7
```

Calculations

Let's get the sum for each participant (adding up all of their time points)

```
participant1_sum = 9+4+6
participant2_sum = 7+5+5
participant3_sum = 8+4+5
participant4_sum = 6+3+6
participant5_sum = 10+5+7

participant1_sum
```

```
## [1] 19
```

```
participant2_sum
```

```
## [1] 17
```

```r
participant3_sum
```

```
## [1] 17
```

```r
participant4_sum
```

```
## [1] 15
```

```r
participant5_sum
```

```
## [1] 22
```

Now, let's get the sum of each level of our IV/factor

```
time1_sum = 9+7+8+6+10
time2_sum = 4+5+4+3+5
time3_sum = 6+5+5+6+7

time1_sum
```

```
## [1] 40
```

```
time2_sum
```

```
## [1] 21
```

```
time3_sum
```

```
## [1] 29
```

Let's also get the total sum of values

```
total_sum = time1_sum + time2_sum + time3_sum
total_sum
```

```
## [1] 90
```

Now we can also get the squared sum of each level of our IV/factor

```r
time1_sum_square = 9^2+7^2+8^2+6^2+10^2
time2_sum_square = 4^2+5^2+4^2+3^2+5^2
time3_sum_square = 6^2+5^2+5^2+6^2+7^2

time1_sum_square
```

```
## [1] 330
```

```r
time2_sum_square
```

```
## [1] 91
```

```r
time3_sum_square
```

```
## [1] 171
```

And the total sum of squared values

```
total_sum_square = time1_sum_square + time2_sum_square + time3_squar
total_sum_square
```

```
## [1] 592
```

Let's also get the mean

```
time1_mean = time1_sum/5
time2_mean = time2_sum/5
time3_mean = time3_sum/5

time1_mean
```

```
## [1] 8
```

```
time2_mean
```

```
## [1] 4.2
```

```
time3_mean
```

```
## [1] 5.8
```

Lastly, let's get the n, the N, and the k

```
time1_n = 5
time2_n = 5
time3_n = 5

time1_n
```

```
## [1] 5
```

```
time2_n
```

```
## [1] 5
```

```
time3_n
```

```
## [1] 5
```

```r
k = 3
k
```

```
## [1] 3
```

```r
total_n = time1_n + time2_n + time3_n
total_n
```

```
## [1] 15
```

Let's get the sum of squares total

$$SS_{total} = \Sigma X^2_{total} - (\frac{(\Sigma X_{total})^2}{N})$$

$$SS_{total} = 592 - (\frac{(90)^2}{15})$$

```r
90^2
```

```
## [1] 8100
```

$$SS_{total} = 592 - (\frac{8100}{15})$$

```r
8100/15
```

```
## [1] 540
```

$$SS_{total} = 592 - 540$$

```r
592 - 540
```

```
## [1] 52
```

$$SS_{total} = 52$$

Sum of Squares Between Groups

$$SS_{time} = \Sigma(\frac{(\Sigma X\ in\ each\ column)^2}{n\ in\ each\ column}) - (\frac{(\Sigma X_{total})^2}{N})$$

$$SS_{time} = (\frac{40^2}{5} + \frac{21^2}{5} + \frac{29^2}{5}) - 540$$

```r
40^2
```

```
## [1] 1600
```

```r
21^2
```

```
## [1] 441
```

```r
29^2
```

```
## [1] 841
```

$$SS_{time} = (\frac{1600}{5} + \frac{441}{5} + \frac{841}{5}) - 540$$

```r
1600/5
```

```
## [1] 320
```

```r
441/5
```

```
## [1] 88.2
```

```r
841/5
```

```
## [1] 168.2
```

$$SS_{time} = (320 + 88.2 + 168.2) - 540$$

```r
320 + 88.2 + 168.2
```

```
## [1] 576.4
```

$$SS_{time} = 576.4 - 540$$

```r
576.4 - 540
```

```
## [1] 36.4
```

$$SS_{time} = 36.4$$

Now on to the Sum of Squares for the participants

$$SS_{subj} = \frac{(\Sigma X^2_{subj\ 1} + \Sigma X^2_{subj\ 2} + \Sigma X^2_{subj\ 3} + \Sigma X^2_{subj\ 4} + \Sigma X^2_{subj\ 5})}{k_A} - (\frac{(\Sigma X_{total})^2}{N})$$

$$SS_{subj} = \frac{(19^2 + 17^2 + 17^2 + 15^2 + 22^2)}{3} - 540$$

```r
19^2
```

```
## [1] 361
```

```r
17^2
```

```
## [1] 289
```

```r
17^2
```

```
## [1] 289
```

```r
15^2
```

```
## [1] 225
```

```r
22^2
```

```
## [1] 484
```

$$SS_{subj} = \frac{(361 + 289 + 289 + 225 + 484)}{3} - 540$$

```r
361 + 289 + 289 + 225 + 484
```

```
## [1] 1648
```

$$SS_{subj} = \frac{1648}{3} - 540$$

```r
1648/3
```

```
## [1] 549.3333
```

$$SS_{subj} = 549.33 - 540$$

```r
549.33 - 540
```

```
## [1] 9.33
```

$$SS_{subj} = 9.33$$

Lastly the interaction of our factor by participants

$$SS_{error} = SS_{total} - SS_{time} - SS_{subj}$$

$$SS_{error} = 52 - 36.4 - 9.33$$

```r
52 - 36.4 - 9.33
```

```
## [1] 6.27
```

$$SS_{error} = 6.27$$

Degrees of Freedom Between Groups

$$df_{time} = k_{time} - 1$$

$$df_{time} = 3 - 1$$

```r
3 - 1
```

```
## [1] 2
```

$$df_{time} = 2$$

Degrees of Freedom For the Interaction

$$df_{error} = (k_{time} - 1)(k_{subj} - 1)$$

$$df_{error} = (3 - 1)(5 - 1)$$

```r
(3 - 1)
```

```
## [1] 2
```

```r
(5 - 1)
```

```
## [1] 4
```

$$df_{error} = (2)(4)$$

```r
2*4
```

```
## [1] 8
```

$$df_{error} = 8$$

Mean Square for The Factor/IV

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

$$MS_{time} = \frac{36.4}{2}$$

```r
36.4/2
```

```
## [1] 18.2
```

$$MS_{time} = 18.2$$

Mean Square of the Interaction of Factor by the Participants

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

$$MS_{error} = \frac{6.27}{8}$$

```r
6.27/8
```

```
## [1] 0.78375
```

$$MS_{error} = .78$$

Within-Subjects F-statistic

$$F_{obt} = \frac{MS_{time}}{MS_{error}}$$

$$F_{obt} = \frac{18.2}{.78}$$

```r
18.2/.78
```

```
## [1] 23.33333
```

$$F_{obt} = 23.33$$

```r
tired <- data.frame(time1 = c(1, 2, 4, 3, 2),
                    time2 = c(4, 5, 6, 4, 3),
                    time3 = c(9, 6, 8, 7, 10))
tired
```

```
##   time1 time2 time3
## 1     1     4     9
## 2     2     5     6
## 3     4     6     8
## 4     3     4     7
## 5     2     3    10
```