psyc3010 lecture 11

One and two-way within participants anova

Last week: Regression topics
next week: mixed anova

announcements

- Assignment 2 due on May 23 (next Monday)
  - submission via Turnitin (on Blackboard)
  - submission deadline = 12 noon
  - If in doubt, or submitting late, e-mail your tutor. You can cc me.

- Quiz 2 marks available after 5 pm Monday the 23rd (or earlier)

- Exam review in Week 13 lecture (last class)
  - exam content + structure
  - how to study for the exam
  - review of course material

- Evaluations in Week 13 lecture
  - course and lecturer
last lecture → this lecture

- last lecture:
  - review of regression topics

- this lecture:
  - back to ANOVA…
  - one-way and two-way within-participants designs

topics for this week

- introduction to within-participants designs
  - research questions
  - power
  - sources of variance

- one-way within-participants ANOVA
  - omnibus tests
  - follow-up tests (main effect comparisons)

- two-way within-participants ANOVA
  - omnibus tests
  - follow-up tests (simple effects and simple comparisons)

- mixed-model: fixed and random effects
- sphericity: problem and solutions
anova – a second look

- **between-participants designs**
  - each person serves in only one treatment/cell
  - we then assume that any difference between them is due to our experimental manipulation (or intrinsic features of the grouping variable, e.g., gender)
  - Within-cell variability is residual error

- **within-participants (repeated-measures) designs**
  - what if each participant served in each treatment?
  - violates the assumption of independence in factorial ANOVA because scores for the participant are correlated across conditions
  - but we can calculate and **remove** any variance due to dependence
  - thus, we can reduce our error term and increase power
### Treatment Means

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td></td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>13</td>
<td></td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>30</td>
<td></td>
<td>27.00</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td></td>
<td></td>
<td>31.67</td>
</tr>
<tr>
<td>Mean</td>
<td>19</td>
<td>21</td>
<td></td>
<td>18.67</td>
</tr>
</tbody>
</table>

Treatment means don’t differ by much – far more variability within each group than between.

---

### An Illustration

In between-participants designs, all within-group variance is error, whereas repeated measures designs remove individual difference variation from the error term.

**People are different**

---

Most of this within-group variance is caused by the fact that some participants learn quickly, and some learn slowly.
an illustration

<table>
<thead>
<tr>
<th>participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>29</td>
<td>30</td>
<td>27.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>31</td>
<td>34</td>
<td>31.67</td>
</tr>
<tr>
<td>mean</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>18.67</td>
</tr>
</tbody>
</table>

solution: firstly remove the between-participants variance (i.e., account for individual differences) and then compare our treatment means

Understanding RM versus BS designs

- In between participants, assign people randomly to j conditions
  - Total Variance = Between group + within group
    - Treatment effect = between group variance
    - Error = within group variance
- No participant variability because each participant has only 1 data point (no variance within individual)
1-way between-participants anova:

- total variation
  - between-groups variance
  - residual/error

any individual differences within groups are considered ‘error’

Understanding RM designs

- In fully within participants design, people are tested in each of j conditions
- “participant” factor is crossed with IV (e.g., factor A)
- End up with A x P (Factor A x Participant) design with only 1 observation per cell
- No within-cell variance – now a cell is one observation (for person i in condition j)
- Weird - So what is error?
  - Interaction of A x P – i.e., the changes (inconsistency) in the effects of A across participants
### A x P design

<table>
<thead>
<tr>
<th>participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>29</td>
<td>30</td>
<td>27.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>31</td>
<td>34</td>
<td>31.67</td>
</tr>
<tr>
<td>mean</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>18.67</td>
</tr>
</tbody>
</table>

Overall, treatment effect for 1 = 16 – 18.67 (-2.67)
treatment effect for 2 = 19-18.67 (+0.33)
treatment effect for 3 = 21 – 18.67 (+2.33)

For P1, treatment effect for 1 = 2 – 4.33 (-2.33)
treatment effect for 2 = 4 – 4.33 (-0.33)
treatment effect for 3 = 7 – 4.33 (+2.67)
### A x P design

<table>
<thead>
<tr>
<th>participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>29</td>
<td>30</td>
<td>27.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>31</td>
<td>34</td>
<td>31.67</td>
</tr>
<tr>
<td>mean</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>18.67</td>
</tr>
</tbody>
</table>

For P2, treatment effect for 1 = 10 – 11.67 (-1.67)
- treatment effect for 2 = 12-11.67 (+0.33)
- treatment effect for 3 = 13-11.67 (+1.33)

### A x P design

<table>
<thead>
<tr>
<th>participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4.33</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>29</td>
<td>30</td>
<td>27.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>31</td>
<td>34</td>
<td>31.67</td>
</tr>
<tr>
<td>mean</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>18.67</td>
</tr>
</tbody>
</table>

For P3, treatment effect for 1 = 22 – 27 (-5)
- treatment effect for 2 = 29-27 (+2)
- treatment effect for 3 = 30 – 27 (+3)
1-way within-participants anova:

**total variation**

- **between-participants variance**
  - any individual differences are removed first
- **within-participants variance**
- **between-treatments**
- **error/residual [interaction P x tr]**
Within-participants design

- **Total Variance = Between participants + within participants**

Between participants variance due to individual differences is partitioned out of error (and treatment)!
  - **Within participants = between treatment [treatment effect] + treatment x participant interaction**
    - [residual error – i.e., inconsistencies in the treatment effect]
  - **F test = TR / TR x P**

- Acknowledges reality that variability within conditions/groups and between conditions/groups are both influenced by participant factor [people doing study]

---

the conceptual model

\[ X_{ij} = \mu + \pi_i + \tau_j + e_{ij} \]

for \( i \) cases and \( j \) treatments:

\( X_{ij} \), any DV score is a combination of:

- \( \mu \rightarrow \) the grand mean,
- \( \pi_i \rightarrow \) variation due to the \( i \)-th person \( (\mu_i - \mu) \)
- \( \tau_j \rightarrow \) variation due to the \( j \)-th treatment \( (\mu_j - \mu) \)
- \( e_{ij} \rightarrow \) error - variation associated with the \( i \)-th cases in the \( j \)-th treatment – error = \( \pi \tau_{ij} \) (plus chance)
partitioning the variance

![Pie chart diagram](image)

IV: block
- 40 trials through whole experiment
- want to compare over 4 blocks of 10 to see if learning has occurred

DV = number of correct responses per block

worked example

basic learning study
- 1-way within-participants design (n=5)
- IV: block
  - 40 trials through whole experiment
  - want to compare over 4 blocks of 10 to see if learning has occurred
- DV = number of correct responses per block
correct trials over 4 blocks of 10

<table>
<thead>
<tr>
<th>block 1</th>
<th>block2</th>
<th>block3</th>
<th>block 4</th>
<th>P total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Participant 2</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Participant 3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Participant 4</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Participant 5</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

| block total | 15 | 20 | 25 | 30 | 90 |
| block mean  | 3  | 4  | 5  | 6  |    |

Definitional formulae

- Total variability – deviation of each observation from the grand mean:
  \[ SS_T = \sum (Y_i - \bar{Y})^2 \]

- Variability due to factor – deviation of factor group means from grand mean:
  \[ SS_A = n \sum (\bar{Y}_j - \bar{Y})^2 \]

- Variability due to participants – deviation of each participant’s mean from the grand mean:
  \[ SS_P = a \sum \sum (\bar{Y}_i - \bar{Y})^2 \]

- Error – changes (inconsistencies) in the effect of factor across participants (TR x P interaction):
  \[ SS_{A\times P} = \sum \sum (Y - \bar{Y}_i - \bar{Y}_j + \bar{Y})^2 \] or \[ SS_{A\times P} = SS_T - SS_A - SS_P \]
degrees of freedom

number of participants * number of conditions

\[ df_{total} = nj-1 = N-1 = 19 \]

\[ df_P = n-1 = 4 \]

\[ df_{tr} = j-1 = 3 \]

\[ df_{error} = (n-1)(j-1) = 12 \]

Big N = Number of observations

error df is different from between-participants anova – because error is now interaction of participant factor x treatment factor

the summary table

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects (P)</td>
<td>59</td>
<td>4</td>
<td>14.75</td>
<td></td>
</tr>
<tr>
<td>Treatment (TR)</td>
<td>25</td>
<td>3</td>
<td>8.33</td>
<td>6.66*</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>12</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ * p < .05 \quad F_{crit}(3,12) = 3.49 \]

\[ MS_P = \text{estimate of variance in DV attributable to INDIVIDUAL DIFFERENCES} \]
\[ \text{(averaged over treatment levels)} \quad \text{-- but ignore this & don't report in write-up} \]

\[ MS_{TR} = \text{estimate of variance in DV attributable to TREATMENT} \]
\[ \text{(averaged over participants)} \]

\[ MS_{Error} = \text{RESIDUAL: estimate of variance in DV not attributable to S or TR} \]
\[ \text{(interaction - the change in the treatment effect across participants = error)} \]
assuming the data was obtained from a between-participants design . . .

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (TR)</td>
<td>25</td>
<td>3</td>
<td>8.33</td>
<td>1.80</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>16</td>
<td>4.63</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$F_{crit}(3,16) = 3.24$

in between-participants designs, individual differences are inseparable from error, hence contribute to the error term.

in within-participants designs it is possible to partial out (i.e., remove) individual differences from the error term.

smaller error term $\rightarrow$ greater POWER ☺

a note on error terms...

- hand calculations in within-participants anova are no different to those in between-participants anova
  - only the error term (and df) changes

- in 1-way within-participants the error term (and df) is the treatment x participants interaction
  - $MS_{error} = MS_{TRxP}$
Nobody made a greater mistake than he who did nothing because he could do only a little.

--Edmund Burke, statesman and writer (1729-1797)

following up the main effect of treatment . . .

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (TR)</td>
<td>25</td>
<td>3</td>
<td>8.33</td>
<td>6.66*</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>12</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$  \( F_{crit}(3,12) = 3.49 \)

in between-participants anova, \( MS_{error} \) is the term we would use to test any effect, including simple comparisons [error = differences between participants – expect within-cell variance is the same across conditions]

but within-participants ANOVA we partition out and ignore the main effect of participants and compute an error term estimating inconsistency as participants change over WS levels
separate error terms: following-up main effects

- We expect inconsistency in TR effect x participants so in simple comparisons use only data for conditions involved in comparison & calculate separate error terms each time

<table>
<thead>
<tr>
<th>block 1</th>
<th>block2</th>
<th>block3</th>
<th>block 4</th>
<th>P total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Participant 2</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Participant 3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Participant 4</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Participant 5</td>
<td>7</td>
<td>6</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>block total</td>
<td>20</td>
<td>25</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>block mean</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B2 vs B3

separate error terms: following-up main effects

• We expect inconsistency in TR effect x participants so in simple comparisons use only data for conditions involved in comparison & calculate separate error terms each time

<table>
<thead>
<tr>
<th>block 1</th>
<th>block2</th>
<th>block3</th>
<th>block 4</th>
<th>P total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Participant 2</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Participant 3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Participant 4</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Participant 5</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>block total</td>
<td>15</td>
<td>30</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>block mean</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B1 vs B4
Simple comparisons in between-participants anova:

- **Total Variation**
  - **Between-groups variance**
    - Contrast 1
    - Contrast 2
    - Contrast 3
  - **Residual/error**

Partition treatment variance to follow-up, but use same error term (within-cell variance) for main effect (treatment) test and for all follow-ups.

Simple comparisons in RM designs:

- **Total Variation**
  - **Between-participants**
  - **Within-participants**
    - **Between-treatments**
    - **Residuals**

Partition treatment variance and residual variance for follow-ups. Each contrast effect is tested against error term = C x P interaction.
### Summary Table

These are the SS\textsubscript{contrasts} we can calculate in the same way as in between-participants ANOVA:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2 vs B3</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>1.25</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B1 vs B4</td>
<td>22.5</td>
<td>1</td>
<td>22.5</td>
<td>22.5*</td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

For comparison, the df for each within-participants effect is the same as usual (i.e., 1).

The error df is calculated as follows:

\[
df\text{error} = (n-1)(j-1) = (5-1)(2-1) = 4
\]

But SS\textsubscript{P\times contrast} terms are calculated separately for each within-participants effect.
2-way within-participants designs

- calculations are similar to a 2-way between-participants ANOVA
  - main effects for A and B are tested, as well as a AxB interaction
  - with a within-participants design, each effect tested has a separate error term
  - this error term simply corresponds to an interaction between the effect due to participants, and the treatment effect

- main effect of A → error term is $MS_{AxP}$
- main effect of B → error term is $MS_{BxP}$
- AxB interaction → error term is $MS_{ABxP}$

2-way between-participants anova:

Partition between-groups variance into A, B and AxB, but use same error term (within-cell variance) for each test (and all follow-ups)
2-way within-participants anova:

Partition treatment variance and residual variance for each effect. Each effect is tested against error term = effect x P interaction.

2-way within-participants example

another learning study:
- 2 x 4 repeated-measures factorial design (n=4)
- first factor: phase
  - phase 1: no reinforcement (100 trials)
  - phase 2: reward for correct response (100 trials)
- second factor: block
  - each phase split into four blocks of 25
  - enables us to compare performance for trials later in each phase with trials early in each phase – thereby assessing learning
- DV = number of correct responses per block
Phase x Block repeated measures design
[phase x block x participants]

<table>
<thead>
<tr>
<th>Particip</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
<th>Particip</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>P1</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>P2</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>P2</td>
<td>10</td>
<td>12</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>P3</td>
<td>7</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>P3</td>
<td>10</td>
<td>15</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>P4</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>P4</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

Data

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between participants</td>
<td>272.6</td>
<td>3</td>
<td>90.867</td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>116.28</td>
<td>1</td>
<td>116.28</td>
<td>59.63*</td>
</tr>
<tr>
<td>Phase x P</td>
<td>5.84</td>
<td>3</td>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>Block</td>
<td>129.6</td>
<td>3</td>
<td>43.20</td>
<td>12.24*</td>
</tr>
<tr>
<td>Block x P</td>
<td>31.77</td>
<td>9</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>Phase x Block</td>
<td>3.34</td>
<td>3</td>
<td>1.11</td>
<td>3.26</td>
</tr>
<tr>
<td>Ph x B x P</td>
<td>3.04</td>
<td>9</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

Critical $F(1,3) = 10.13$
Critical $F(3,9) = 3.86$

summary table . . .
following up main effects

- as with one-way repeated measures designs, use of error term for effect (e.g., $MS_{Block \times P}$) is not appropriate for follow-up comparisons
- a separate error term must be calculated for each comparison undertaken ($MS_{Block_{COMPA_{COMP}} \times P}$)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{COMP}$</td>
<td>18.06</td>
<td>1</td>
<td>18.06</td>
<td>6.62</td>
</tr>
<tr>
<td>$B_{COMP \times P}$</td>
<td>8.19</td>
<td>3</td>
<td>2.73</td>
<td></td>
</tr>
</tbody>
</table>

Critical $F(1,3) = 10.13$

following up interactions...

- again, separate error terms must be used for each effect tested
  - simple effects
    - error term is $MS_{A \text{ at } B1 \times P}$
    - the interaction between the $A$ treatment and participants, at $B1$
  - simple comparisons
    - error term is $MS_{A_{COMP} \text{ at } B1 \times P}$
    - interaction between the $A$ treatment (only the data contributing to the comparison, $A_{COMP}$), and participants, at $B1$
2 approaches to within-participants designs

- **mixed-model approach**
  - what we have been doing with examples so far calculations
  - treatment is a **fixed** factor, participants is a **random** factor
    - **Fixed factor**: You chose the levels of the IV.
      - You have sampled all the levels of the IV or
      - You have selected particular levels based on a theoretical reason
    - **Random factor**: The levels of the IV are chosen at random
      - Random factors have different error terms: all ANOVA we have done to date has assumed the IVs are fixed. For most of you, the participant factor is the only random factor you will ever meet (be grateful). 😊 You can read up on random factor ANOVA models in advanced textbooks if you need to (e.g., as a postgrad).
  - powerful when assumptions are met
  - mathematically user-friendly
    - just like a factorial anova
  - restrictive assumptions, but adjustments available if they are violated

- **multivariate approach**...which we will discuss briefly later
assumptions of mixed-model approach

- not dissimilar to between-participants assumptions:
  1. sample is *randomly drawn* from population
  2. DV scores are *normally distributed* in the population
  3. **compound symmetry**
     - homogeneity of variances in levels of repeated-measures factor
     - homogeneity of covariances (equal correlations/covariances between pairs of levels)

- compound symmetry
- the variance-covariance matrix:

\[
\Sigma = \begin{bmatrix}
  T1 & T2 & T3 \\
  T1 & 158.92 & 163.33 & 163.00 \\
  T2 & 163.33 & 172.67 & 170.67 \\
  T3 & 163.00 & 170.67 & 170.00 \\
\end{bmatrix}
\]
**compound symmetry**

- the variance-covariance matrix:

\[
\Sigma = \begin{bmatrix}
T1 & T2 & T3 \\
T1 & 158.92 & 163.33 & 163.00 \\
T2 & 163.33 & 172.67 & 170.67 \\
T3 & 163.00 & 170.67 & 170.00 \\
\end{bmatrix}
\]

**compound symmetry** requires that variances are roughly equal (homogeneity of variance)

---

**compound symmetry**

- the variance-covariance matrix:

\[
\Sigma = \begin{bmatrix}
T1 & T2 & T3 \\
T1 & 158.92 & 163.33 & 163.00 \\
T2 & 163.33 & 172.67 & 170.67 \\
T3 & 163.00 & 170.67 & 170.00 \\
\end{bmatrix}
\]

**compound symmetry** requires that covariances are roughly equal (homogeneity of covariance)
Mauchly’s test of sphericity

- compound symmetry is a very restrictive assumption – often violated
- sphericity is a more broad and less restrictive assumption
- SPSS – Mauchley’s test of sphericity
  - examines overall structure of covariance matrix
  - determines whether values in the main diagonal (variances) are roughly equal, and if values in the off-diagonal are roughly equal (covariances)
  - evaluated as $\chi^2$ – if significant, sphericity assumption is violated
  - not a robust test AT ALL – very commonly fail to find Mauchley’s sphericity is sig even when violations of sphericity are present in the data

violations of sphericity

when sphericity doesn’t matter
- in between-participants designs, because treatments are unrelated (different participants in different treatments)
  - the assumption of homogeneity of variance still matters though
- when within-participant factors have two levels, because only one estimate of covariance can be computed

when it does matter
- in all other within-participants designs (3 + levels)
- when the sphericity assumption is violated, F-ratios are positively biased
  - critical values of $F$ [based on df $a – 1$, $(a – 1)(n – 1)$] are too small
  - therefore, probability of type-1 error increases
adjustments to degrees of freedom

- Best to assume that have a problem and make adjustment proactively – change critical $F$ by adjusting degrees of freedom

- **epsilon ($\varepsilon$) adjustments**
  - epsilon is simply a value by which the degrees of freedom for the test of $F$-ratio is multiplied
  - equal to 1 when sphericity assumption is met (hence no adjustment), and < 1 when assumption is violated
  - the lower the epsilon value (further from 1), the more conservative the test becomes

different types of epsilon

- **Lower-bound epsilon**
  - Act as if have only 2 treatment levels with maximal heterogeneity
  - used for conditions of maximal heterogeneity, or worst-case violation of sphericity → often too conservative

- **Greenhouse-Geisser epsilon**
  - size of $\varepsilon$ depends on degree to which sphericity is violated
  - $1 \geq \varepsilon \geq 1/(k-1)$: varies between 1 (sphericity intact) and lower-bound epsilon (worst-case violation)
  - *generally recommended – not too stringent, not too lax*
different types of epsilon

- **Huynh-Feldt epsilon**
  - an adjustment applied to the GG-epsilon
  - often results in epsilon exceeding 1, in which case it is set to 1
  - used when "true value" of epsilon is believed to be \( \geq .75 \)

---

**spss output from our previous example**

Mauchly's Test of Sphericity

<table>
<thead>
<tr>
<th>Within Subjects Effect</th>
<th>Mauchly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Greenhouse-Geisser</th>
<th>Huynh-Feldt</th>
<th>Lower-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHASE</td>
<td>1.000</td>
<td>.000</td>
<td>0</td>
<td>.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>BLOCK</td>
<td>.111</td>
<td>3.785</td>
<td>5</td>
<td>.634</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PHASE * BLOCK</td>
<td>.000</td>
<td>.874</td>
<td>5</td>
<td>.704</td>
<td>.587</td>
<td>1.000</td>
<td>.333</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept  
   Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

No sphericity test for effects involving phase – only 2 levels

test for block is not significant (sphericity not violated) but **we aren’t going to trust it!**
spss output from our previous example

Mauchly's Test of Sphericity

<table>
<thead>
<tr>
<th>Within Subjects Effect</th>
<th>Mauchly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Greenhouse-Geisser</th>
<th>Huynh-Feldt</th>
<th>Lower-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHASE</td>
<td>1.000</td>
<td>.000</td>
<td>0</td>
<td>.</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>BLOCK</td>
<td>.111</td>
<td>3.785</td>
<td>5</td>
<td>.634</td>
<td>.367</td>
<td>1.000</td>
<td>.333</td>
</tr>
<tr>
<td>PHASE * BLOCK</td>
<td>.000</td>
<td>.000</td>
<td>5</td>
<td>.</td>
<td>348</td>
<td>370</td>
<td>.333</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

- May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
- Design: Intercept
  Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

compare the epsilon values…

*The lower the epsilon, the greater the adjustment, the more conservative the test*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHASE</td>
<td>116.281</td>
<td>1</td>
<td>116.281</td>
<td>59.695</td>
<td>.005</td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>116.281</td>
<td>1.000</td>
<td>116.281</td>
<td>59.695</td>
<td>.005</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>116.281</td>
<td>1.000</td>
<td>116.281</td>
<td>59.695</td>
<td>.005</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>116.281</td>
<td>1.000</td>
<td>116.281</td>
<td>59.695</td>
<td>.005</td>
</tr>
<tr>
<td>Error(PHASE)</td>
<td>5.844</td>
<td>3</td>
<td>1.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>5.844</td>
<td>3.000</td>
<td>1.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>5.844</td>
<td>3.000</td>
<td>1.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>5.844</td>
<td>3.000</td>
<td>1.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-bound</td>
<td>5.844</td>
<td>3.000</td>
<td>1.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOCK</td>
<td>129.594</td>
<td>3</td>
<td>43.198</td>
<td>12.233</td>
<td>.002</td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>129.594</td>
<td>1.760</td>
<td>73.621</td>
<td>12.233</td>
<td>.011</td>
</tr>
<tr>
<td>Greenhouse-Geiser</td>
<td>129.594</td>
<td>3.000</td>
<td>43.198</td>
<td>12.233</td>
<td>.002</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>129.594</td>
<td>3.000</td>
<td>43.198</td>
<td>12.233</td>
<td>.002</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>129.594</td>
<td>1.000</td>
<td>129.594</td>
<td>12.233</td>
<td>.040</td>
</tr>
<tr>
<td>Error(BLOCK)</td>
<td>3.031</td>
<td>3.000</td>
<td>1.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>3.031</td>
<td>3.000</td>
<td>1.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geiser</td>
<td>3.031</td>
<td>3.000</td>
<td>1.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>3.031</td>
<td>3.000</td>
<td>1.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-bound</td>
<td>3.031</td>
<td>3.000</td>
<td>1.010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p values are different for Block despite same F of 12.233 – why?*
*because each F value is associated with a different df*
*thus, each calculated F value is compared to a different critical F value to determine whether it meets the criteria for statistical significance*
**spss output from our previous example**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOCK</td>
<td>Sphericity Assumed</td>
<td>129.594</td>
<td>3</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>129.594</td>
<td>1.760</td>
<td>73.621</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>129.594</td>
<td>3.000</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>129.594</td>
<td>1.000</td>
<td>129.594</td>
<td>12.233</td>
</tr>
<tr>
<td>Error(BLOCK)</td>
<td>Sphericity Assumed</td>
<td>31.781</td>
<td>9</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>31.781</td>
<td>5.281</td>
<td>6.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>31.781</td>
<td>9.000</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>31.781</td>
<td>3.000</td>
<td>10.594</td>
<td></td>
</tr>
</tbody>
</table>

Sphericity assumed – i.e., no adjustment

This is what we based our degrees of freedom on before, i.e., $b-1 = 4-1 = 3$, $(n-1)(b-1) = 3 \times 3 = 9 \rightarrow 3,9$

**Lower-bound – for worst case heterogeneity**

i.e., $df = 1$, $b-1$ – here we come close to concluding non-significance (which could be a type-2 error)
### spss output from our previous example

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOCK</td>
<td>Sphericity Assumed</td>
<td>129.594</td>
<td>3</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>129.594</td>
<td>1.760</td>
<td>73.621</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>129.594</td>
<td>3.000</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>129.594</td>
<td>1.000</td>
<td>129.594</td>
<td>12.233</td>
</tr>
<tr>
<td>Error(BLOCK)</td>
<td>Sphericity Assumed</td>
<td>31.781</td>
<td>9</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>31.781</td>
<td>5.281</td>
<td>6.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>31.781</td>
<td>9.000</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>31.781</td>
<td>3.000</td>
<td>10.594</td>
<td></td>
</tr>
</tbody>
</table>

Greenhouse-Geisser adjustment does not change significance of result

### spss output from our previous example

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOCK</td>
<td>Sphericity Assumed</td>
<td>129.594</td>
<td>3</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>129.594</td>
<td>1.760</td>
<td>73.621</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>129.594</td>
<td>3.000</td>
<td>43.198</td>
<td>12.233</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>129.594</td>
<td>1.000</td>
<td>129.594</td>
<td>12.233</td>
</tr>
<tr>
<td>Error(BLOCK)</td>
<td>Sphericity Assumed</td>
<td>31.781</td>
<td>9</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>31.781</td>
<td>5.281</td>
<td>6.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>31.781</td>
<td>9.000</td>
<td>3.531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>31.781</td>
<td>3.000</td>
<td>10.594</td>
<td></td>
</tr>
</tbody>
</table>

Huynh-Feldt – adjusts GG

no different to ‘sphericity assumed’ – indicates that $\varepsilon > 1$
Writing up…

Changes in participants’ learning with practice and with or without reinforcement were explored in a 2 [phase] x 4 [Block] repeated measures ANOVA. In these analyses, the Huynh-Feldt correction was applied to the degrees of freedom, however the full degrees of freedom are reported here. Contrary to predictions, the interaction was not significant, $F(3,9) = 3.309, p = .159, \eta^2 = \text{??}$. However, as hypothesised, participants learned more in the phase with reinforcement ($M = 42.5; SD = \text{??}$) than in the phase without ($M = 27.25; SD = \text{??}$), $F(1, 3) = 59.70, p = .005, \eta^2 = \text{??}$. A main effect of Block, $F(3,9) = 12.23, p = .002, \eta^2 = \text{??}$, was followed up with a series of contrasts. These revealed that …
multivariate approach

- **multivariate analysis of variance (manova)**
  - creates a *linear composite* of multiple DVs
  - In MANOVA approach to repeated measures designs, our repeated measures variable is treated as multiple DVs and combined / weighted to maximise the difference between levels of other variables (similar to the approach regression uses to combined multiple predictors)
    - multivariate tests – *Pillai’s Trace, Hotelling’s Trace, Wilk’s Lambda, Roy’s Largest Root*
    - does not require restrictive assumptions that mixed model within participants design does
  - more complex underlying math

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai’s Trace</th>
<th>Wilks’ Lambda</th>
<th>Hotelling’s Trace</th>
<th>Roy’s Largest Root</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHASE</strong></td>
<td>.952</td>
<td>.048</td>
<td>19.898</td>
<td>19.898</td>
</tr>
<tr>
<td></td>
<td>59.695*</td>
<td>59.695*</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>BLOCK</strong></td>
<td>.992</td>
<td>.008</td>
<td>129.050</td>
<td>129.050</td>
</tr>
<tr>
<td></td>
<td>43.017*</td>
<td>43.017*</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td><strong>PHASE * BLOCK</strong></td>
<td>.993</td>
<td>.010</td>
<td>102.333</td>
<td>102.333</td>
</tr>
<tr>
<td></td>
<td>102.333*</td>
<td>102.333*</td>
<td>2.000</td>
<td>2.000</td>
</tr>
</tbody>
</table>

a. Exact statistic
b. Design: Intercept
   Within Subjects Design: PHASE+BLOCK+PHASE*BLOCK

---

66
Take home message

- What is MANOVA doing?
  - Weighting the DV for each level of the repeated measures IV with coefficients (like what happens to scores for each IV in multiple regression) to create a predicted DV score that maximises differences across the levels of the IV
  - Problem: Instead of adapting model to observed DVs, selectively weight or discount DVs based on how they fit the model.
    - A-theoretical, over-capitalises on chance
- Therefore, don’t use MANOVA approach to repeated measures
- With repeated measures designs, report the mixed model Fs not the MANOVA statistics
- Usually report GG Fs to ensure adjustment for sphericity violations which are common (regardless of Mauchley’s test, which is too conservative and may not be sig. even when there are large violations)
- Personally I always use the GG or HF adjustment (HF can be more liberal) but report full df – this is common

pros and cons

advantages of within-participants designs:

- more efficient
  - n Ps in j treatments generate nj data points
  - simplifies procedure
- more sensitive
  - estimate individual differences (SSparticipants) and remove from error term
pros and cons

disadvantages of within-participants designs:
- restrictive statistical assumptions
- sequencing effects:
  - learning, practice — improved later regardless of manipulation
  - Fatigue — deteriorating later regardless of manipulation
  - Habituation — insensitivity to later manipulations
  - Sensitisation — become more responsive to later manipulations
  - Contrast — previous treatment sets standard to which react
  - Adaptation — adjustment to previous manipulations changes reaction to later
  - Direct carry-over — learn something in previous that alters later
  - Etc!
- An essential methodological practice in RM designs is to counterbalance to reduce sequencing effects
  - i.e., half participants receive order A1 then A2; half receive A2 then A1
  - But can still get treatment x order interactions

most important points

- in within participants anova, the error term used for ANY effect is equal to the interaction between that effect and the effect of participants (a random factor)
  - this applies to:
    - main effects
      - follow-up (main) comparisons
    - interactions
      - simple effects
      - follow-up (simple) comparisons
- due to problems causes by lack of compound symmetry/sphericity, adjustments (such as Greenhouse-Geisser adjustment) to our degrees of freedom are needed
- unless we used the manova approach, which we shouldn’t, because it is inferior
In class next week:
- Mixed ANOVA

In the tutes:
- This week: Consult for A2
- Next week: Within-participants and mixed designs

Readings:
- Howell
  - chapter 14
- Field
  - Chapter 11