201ab Quantitative methods
Repeated Measures

- Measurements of the same thing are correlated.
- Why use ‘repeated measures’ designs?

- 1 within-subject factor, 1 measure per cell per subject
- 1 within-subject factor, >1 measure per cell per subject
- >1 within-subject factors
- Mixed designs: within and between subject effects

- What’s the right error for each effect?

- Blocking as repeated measures.

Correlations from sources of variability

- If we measure the same 'unit' multiple times, those measurements will be correlated. If we treat them as independent samples of the unit's population, we will be very wrong.
  - Goal: put CI on average male height.
    Procedure: I measure my own height 10 times…
    69.3, 68.6, 68.3, 69.1, 68.9, 68.0, 69.4, 69.5, 68.8, 68.4
    Mean = 68.8 SD = 0.5 … sem = 0.16.
    So, CI on male height is 68.5 to 69.2…?
    What's wrong with this?
  - **No matter how many times I measure myself, I am not getting an estimate of the variability of heights across men. I am just getting an estimate of the error in my height measurements (and/or variability of my posture/shoes)**

Correlations from sources of variability

- Sometimes obvious, but hard to track in complex designs
- Example:
  - I measure homework scores
    I have 10 students. 5 assignments. 4 problems/assignment
    So we have 20 measurements per student.
    40 measurements per assignment.
    1 measurement per problem.
    What's the correlation structure / sources of variability?
- Sources of variation:
  - Students (some do better overall).
  - Problems (some are easier than others).
  - Student*Assignment interaction (some students may have had less time on some assignments),
  - We have to respect this correlation structure when doing

Correlations from sources of variability

- When doing repeated-measures or mixed designs, we have to grapple with 'nested' measurements and variability at different scales of our design.
- We now have *conditionally* independent residuals, but collapsing across the nested measurement structure, residuals are correlated.

- This can be very hard.
  - The most general ways to deal with these kinds of data structures are 'hierarchical linear models' or 'linear mixed effects' models. We will talk about those later.
  - Here we will consider the simpler (but still hard!) cases that can be analyzed using mixed-design ANOVAs.

Factor A: Country *(index: i)*

Factor B: Gender *(index: j)*	**North Korea** (i=1)	**USA** (i=2)	**South Korea** (i=3)	**Netherlands** (i=4)	
Male	$y_{1,1,1}$ 67 $y_{1,1,2}$ 66 $y_{1,1,3}$ 64 $y_{1,1,4}$ 64 (i=1, j=1)	$y_{2,1,1}$ 74 $y_{2,1,2}$ 83 $y_{2,1,3}$ 73 $y_{2,1,4}$ 74 $y_{2,1,5}$ 68 (i=2, j=1)	$y_{3,1,1}$ 75 $y_{3,1,2}$ 72 $y_{3,1,3}$ 68 (i=3, j=1)	$y_{4,1,1}$ 71 $y_{4,1,2}$ 77 $y_{4,1,3}$ 70 $y_{4,1,4}$ 80 $y_{4,1,5}$ 73 $y_{4,1,6}$ 79 $y_{4,1,7}$ 75 (i=4, j=1)	j=1
Female	$y_{1,2,1}$ 64 $y_{1,2,2}$ 68 $y_{1,2,3}$ 66 $y_{1,2,4}$ 57 $y_{1,2,5}$ 64 $y_{1,2,6}$ 64 (i=1, j=2)	$y_{2,2,1}$ 59 $y_{2,2,2}$ 63 $y_{2,2,3}$ 68 $y_{2,2,4}$ 60 $y_{2,2,5}$ 67 $y_{2,2,6}$ 64 $y_{2,2,7}$ 59 $y_{2,2,8}$ 68 $y_{2,2,9}$ 72 $y_{2,2,10}$ 57 (i=2, j=2)	$y_{3,2,1}$ 61 $y_{3,2,2}$ 57 $y_{3,2,3}$ 64 $y_{3,2,4}$ 63 $y_{3,2,5}$ 65 $y_{3,2,6}$ 64 (i=3, j=2)	$y_{4,2,1}$ 75 $y_{4,2,2}$ 68 $y_{4,2,3}$ 72 $y_{4,2,4}$ 66 (i=4, j=2)	j=2

So far we have dealt with ANOVA designs/data in which all residuals are presumed to be independent.

Factor A (between subjects): Test *(index: i)*

Midterm 201a	Final 201a	Midterm 201b	Final 201b
$y_{1,1} = 67$	$y_{2,1} = 71$	$y_{3,1} = 55$	$y_{4,1} = 89$
$y_{1,2} = 84$	$y_{2,2} = 67$	$y_{3,2} = 71$	$y_{4,2} = 75$
$y_{1,3} = 73$	$y_{2,3} = 66$	$y_{3,3} = 38$	$y_{4,3} = 56$
$y_{1,4} = 60$	$y_{2,4} = 79$	$y_{3,4} = 69$	$y_{4,4} = 84$
$y_{1,5} = 45$		$y_{3,5} = 63$	
$y_{1,6} = 35$			

So far we have dealt with ANOVA designs/data in which all residuals are presumed to be independent. But this is not always the case, indeed, there is virtue to introducing dependence.

Factor A (within subject): Test *(index: j)*

Factor (random effect): Students *(index: i)*

	Midterm 201a	Final 201a	Midterm 201b	Final 201b
i=1	$y_{1,1} = 67$	$y_{1,2} = 71$	$y_{1,3} = 55$	$y_{1,4} = 89$
i=2	$y_{2,1} = 84$	$y_{2,2} = 67$	$y_{2,3} = 71$	$y_{2,4} = 75$
i=3	$y_{3,1} = 73$	$y_{3,2} = 66$	$y_{3,3} = 38$	$y_{3,4} = 56$
i=4	$y_{4,1} = 60$	$y_{4,2} = 79$	$y_{4,3} = 69$	$y_{4,4} = 84$
i=5	$y_{5,1} = 45$	$y_{5,2} = 90$	$y_{5,3} = 63$	$y_{5,4} = 72$
i=6	$y_{6,1} = 35$	$y_{6,2} = 59$	$y_{6,3} = 65$	$y_{6,4} = 69$
i=7	$y_{7,1} = 25$	$y_{7,2} = 55$	$y_{7,3} = 47$	$y_{7,4} = 65$
i=8	$y_{8,1} = 73$	$y_{8,2} = 66$	$y_{8,3} = 60$	$y_{8,4} = 80$
	j=1	j=2	j=3	j=4

In repeated measures, sampling units (subjects) are measured multiple times; so we can estimate idiosyncratic effects of that unit.

And we can factor them out, to reduce error, and gain power.

Same logic as with a paired t-test, but gets trickier with ANOVA.

Repeated measures

Factor A (within subject): Test *(index: j)*

Factor (random effect): Students *(index: i)*

	Midterm 201a	Final 201a	Midterm 201b	Final 201b
i=1	$y_{1,1}$ = 67	$y_{1,2}$ = 71	$y_{1,3}$ = 55	$y_{1,4}$ = 89
i=2	$y_{2,1}$ = 84	$y_{2,2}$ = 67	$y_{2,3}$ = 71	$y_{2,4}$ = 75
i=3	$y_{3,1}$ = 73	$y_{3,2}$ = 66	$y_{3,3}$ = 38	$y_{3,4}$ = 56
i=4	$y_{4,1}$ = 60	$y_{4,2}$ = 79	$y_{4,3}$ = 69	$y_{4,4}$ = 84
i=5	$y_{5,1}$ = 45	$y_{5,2}$ = 90	$y_{5,3}$ = 63	$y_{5,4}$ = 72
i=6	$y_{6,1}$ = 35	$y_{6,2}$ = 59	$y_{6,3}$ = 65	$y_{6,4}$ = 69
i=7	$y_{7,1}$ = 25	$y_{7,2}$ = 55	$y_{7,3}$ = 47	$y_{7,4}$ = 65
i=8	$y_{8,1}$ = 73	$y_{8,2}$ = 66	$y_{8,3}$ = 60	$y_{8,4}$ = 80
	j=1	j=2	j=3	j=4

- Multiple measurements now share common source of variability: variability of subject.
- In this case, we have a purely within-subject design.

- We want to factor out subject effects (some students do better than others) and measure test effects.
- We are going to do this by saying that we expect different sources of error: some across subjects, some within subject.

- We're gonna need to look at some math to understand.

Different models in cell-math.

Factor A (between subjects): Test *(index: i)*

Midterm 201a	Final 201a	Midterm 201b	Final 201b
$y_{1,1} = 67$	$y_{2,1} = 71$	$y_{3,1} = 55$	$y_{4,1} = 89$
$y_{1,2} = 84$	$y_{2,2} = 67$	$y_{3,2} = 71$	$y_{4,2} = 75$
$y_{1,3} = 73$	$y_{2,3} = 66$	$y_{3,3} = 38$	$y_{4,3} = 56$
$y_{1,4} = 60$	$y_{2,4} = 79$	$y_{3,4} = 69$	$y_{4,4} = 84$
$y_{1,5} = 45$		$y_{3,5} = 63$	
$y_{1,6} = 35$			

In a between subjects design, we have one measurement per subject, and multiple measurements per condition. So we just estimate a single subject error.

Data point j in between-subject cell i

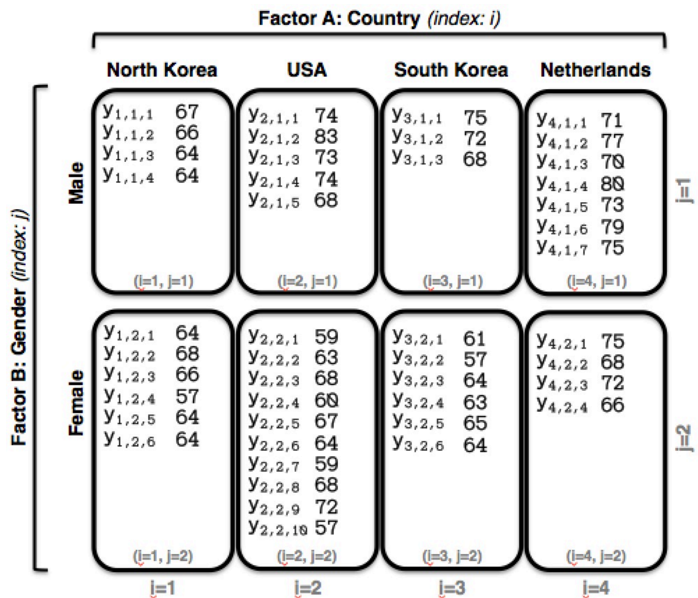
$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_i} + \boxed{\varepsilon_{i,j}}$$

Overall mean

Additive effect of a given treatment (e.g., how much does each test deviate from average performance)

Error of subject j in cell i. (e.g., How much does that subject's score deviate from the average of that cell). Note that this error term includes measurement error as well as subject effects and even subject-cell interactions.

Different models in cell-math.



In a between subjects design, we have one measurement per subject, and multiple measurements per condition. So we just estimate a single subject error.

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_i} + \boxed{\beta_j} + \boxed{\alpha\beta_{i,j}} + \boxed{\varepsilon_{i,j,k}}$$

Data point k in between-subject cell I,j

Overall mean

Additive effect of factor A treatment

Additive effect of factor B treatment

Additive effect of factor AxB interaction for cell i,j

Error of kth observation in cell i,j. Includes measurement error, subject effects, and subject-treatment interactions

Different models in cell-math.

Factor A (between subjects): Test *(index: i)*

Midterm 201a	Final 201a	Midterm 201b	Final 201b
$y_{1,1} = 67$	$y_{2,1} = 71$	$y_{3,1} = 55$	$y_{4,1} = 89$
$y_{1,2} = 84$	$y_{2,2} = 67$	$y_{3,2} = 71$	$y_{4,2} = 75$
$y_{1,3} = 73$	$y_{2,3} = 66$	$y_{3,3} = 38$	$y_{4,3} = 56$
$y_{1,4} = 60$	$y_{2,4} = 79$	$y_{3,4} = 69$	$y_{4,4} = 84$
$y_{1,5} = 45$		$y_{3,5} = 63$	
$y_{1,6} = 35$			

But now let's measure each subject in each test....

Data point j in between-subject cell i

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_i} + \boxed{\varepsilon_{i,j}}$$

Overall mean

Additive effect of a given treatment (e.g., how much does each test deviate from average performance)

Error of subject j in cell i. (e.g., How much does that subject's score deviate from the average of that cell).
Note that this error term includes measurement error as well as subject effects and even subject-cell interactions.

Different models in cell-math.

Factor A (within subject): Test *(index: j)*

Students *(index: i)*	Midterm 201a	Final 201a	Midterm 201b	Final 201b
i=1	$y_{1,1} = 67$	$y_{1,2} = 71$	$y_{1,3} = 55$	$y_{1,4} = 89$
i=2	$y_{2,1} = 84$	$y_{2,2} = 67$	$y_{2,3} = 71$	$y_{2,4} = 75$
i=3	$y_{3,1} = 73$	$y_{3,2} = 66$	$y_{3,3} = 38$	$y_{3,4} = 56$
i=4	$y_{4,1} = 68$	$y_{4,2} = 79$	$y_{4,3} = 69$	$y_{4,4} = 84$

In a within-subject design we can estimate the subject effect, and take it out of our error term!

Data point for subject i in within-subject cell j

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

Overall mean

Additive effect of factor A treatment

Additive effect of subject

Error of measuring subject j in the ith condition.

Note that this includes both measurement error as well as the subject-treatment interaction.

Different models in cell-math.

Factor A (within subject): Test *(index: j)*

j): Students *(index: i)*	Midterm 201a	Final 201a	Midterm 201b	Final 201b
i=1	$y_{1,1} = 67$	$y_{1,2} = 71$	$y_{1,3} = 55$	$y_{1,4} = 89$
i=2	$y_{2,1} = 84$	$y_{2,2} = 67$	$y_{2,3} = 71$	$y_{2,4} = 75$
i=3	$y_{3,1} = 73$	$y_{3,2} = 66$	$y_{3,3} = 38$	$y_{3,4} = 56$
i=4	$y_{4,1} = 68$	$y_{4,2} = 79$	$y_{4,3} = 69$	$y_{4,4} = 84$

In a within-subject design we can estimate the subject effect, and take it out of our error term!

Data point for subject i in within-subject cell j

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

Overall mean

Additive effect of factor A treatment

Additive effect of subject

Error of measuring subject j in the ith condition.

Note that this includes both measurement error as well as the subject-treatment interaction.

In a between subject design, the subject effect would be lumped with the error.

But, look: here, because we have multiple measurements per subject, we can estimate the "subject effect" and remove it from the error! This gives us power!

What kind of design would we need to estimate the subject-treatment interaction?

Simple 1-way repeated measure (1/cell)

```
D1.data
     student            test score
1          1  201-A.midterm    58
1.1        1    201-A.final    38
1.2        1 201-A.homework    26
1.3        1  201-B.midterm    32
1.4        1    201-B.final    38
1.5        1 201-B.homework    53
2          2  201-A.midterm    74
2.1        2    201-A.final    58
2.2        2 201-A.homework    50
2.3        2  201-B.midterm    68
2.4        2    201-B.final    64
2.5        2 201-B.homework   101
3          3  201-A.midterm    73
3.1        3    201-A.final    44
3.2        3 201-A.homework    29
3.3        3  201-B.midterm    55
...... ...... ...... ...... ......
10        10  201-A.midterm    81
10.1      10    201-A.final    58
10.2      10 201-A.homework    49
10.3      10  201-B.midterm    86
10.4      10    201-B.final    52
10.5      10 201-B.homework    86
```

The simplest possible repeated measures design is what we just saw: we have one within-subject factor (test), and one observation per subject per factor level.

Here we have 10 students, each being assessed on 6 different ‘tests’, with one score for each test.

Total measurements: 60

Measurements/subject: 6

Simple 1-way repeated measure (1/cell)

```
D1.data
      student              test  score
1           1   201-A.midterm     58
1.1         1     201-A.final     38
...... ...... ...... ...... ......
10.5       10 201-B.homework     86
```

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

Here we have 10 students, each being assessed on 6 different 'tests', with one score for each test.

Factor A (within subject): Test *(index: j)*

Factor (random effect): Students *(index: i)*

	Midterm 201a	Final 201a	Midterm 201b	Final 201b
i=1	$y_{1,1}$ = 67	$y_{1,2}$ = 71	$y_{1,3}$ = 55	$y_{1,4}$ = 89
i=2	$y_{2,1}$ = 84	$y_{2,2}$ = 67	$y_{2,3}$ = 71	$y_{2,4}$ = 75
i=3	$y_{3,1}$ = 73	$y_{3,2}$ = 66	$y_{3,3}$ = 38	$y_{3,4}$ = 56
i=4	$y_{4,1}$ = 60	$y_{4,2}$ = 79	$y_{4,3}$ = 69	$y_{4,4}$ = 84
i=5	$y_{5,1}$ = 45	$y_{5,2}$ = 90	$y_{5,3}$ = 63	$y_{5,4}$ = 72
i=6	$y_{6,1}$ = 35	$y_{6,2}$ = 59	$y_{6,3}$ = 65	$y_{6,4}$ = 69
i=7	$y_{7,1}$ = 25	$y_{7,2}$ = 55	$y_{7,3}$ = 47	$y_{7,4}$ = 65
i=8	$y_{8,1}$ = 73	$y_{8,2}$ = 66	$y_{8,3}$ = 60	$y_{8,4}$ = 80
	j=1	j=2	j=3	j=4

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

So we want to adopt this sort of model: one that factors out the subject effect from the error.

Like this, but with 10 subjects (rather than 8, as pictured) and including two more 'test' levels: 201a-homework and 201b-homework.

Simple 1-way repeated measure (1/cell)

```
D1.data
      student             test  score
1           1   201-A.midterm     58
1.1         1     201-A.final     38
...... ...... ...... ...... ......
10.5       10 201-B.homework     86
```

10 students, each assessed on 6 'tests'; with 1 score per student per test

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

We could just **ignore the subject effect**, and then all the subject effects get lumped in with the error.

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_i} + \boxed{\varepsilon_{i,j}}$$

```
summary(aov(score~test))
              Df Sum Sq Mean Sq F value   Pr(>F)
test           5  11251  2250.2    9.83 1.01e-06 ***
Residuals     54  12361   228.9
```

But that would be silly: why lose power by failing to factor out subject effects?

Simple 1-way repeated measure (1/cell)

```
D1.data
       student              test  score
1            1  201-A.midterm     58
1.1          1     201-A.final     38
...... ...... ...... ...... ......
10.5        10 201-B.homework     86
```

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

To factor out subject effects, we have to add them to the model. In this simple case, we can add subject as a factor.

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

```
summary(aov(score~test + student))
            Df Sum Sq Mean Sq F value   Pr(>F)
test         5  11251  2250.2   15.17 9.99e-09 ***
student      9   5686   631.8    4.26 0.000487 ***
Residuals   45   6675   148.3
```

Note: our SS and df error dropped because that variability was rightly attributed to a main effect of subject.

Simple 1-way repeated measure (1/cell)

```
D1.data
      student             test  score
1           1   201-A.midterm     58
1.1         1     201-A.final     38
...... ...... ...... ...... ......
10.5       10 201-B.homework     86
```

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

To factor out subject effects, we have to add them to the model. For consistency with other models, we should add them not as a factor, but as an 'error' / 'random effect' term.

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

```
summary(aov(score~test + Error(student)))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

Notes: (1) this analysis doesn't explicitly test if there is a significant subject effect, but we usually don't care about it anyway. (2) We see that we are 'splitting' the error into two strata: error between subjects, and error 'within' subjects.

Simple 1-way repeated measure (1/cell)

```
D1.data
     student              test score
1          1  201-A.midterm     58
1.1        1    201-A.final     38
...... ...... ...... ...... ......
10.5      10 201-B.homework     86
```

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

Something we can't do: Add a student:test interaction

```
summary(aov(score~test*student))
             Df Sum Sq Mean Sq
test          5  11251  2250.2
student       9   5686   631.8
test:student 45   6675   148.3
```

Because we only have 1 measurement per student-test combination, if we estimate a student:test interaction, there is no error left over. Indeed, our previous error term *was* the student:test interaction!

Simple 1-way repeated measure (1/cell)

```
D1.data
      student            test score
1           1   201-A.midterm    58
1.1         1     201-A.final    38
...... ...... ...... ...... ......
10.5       10 201-B.homework    86
```

The simplest possible repeated measures design: one within-subject factor, one observation per subject per factor level.

If we write the model in the complete way: specifying which factors are nested within students, the fact that the student:test interaction is the within-subject error term is made explicit for us.

$$y_{i,j} = \boxed{\mu} + \boxed{\alpha_j} + \boxed{\rho_i} + \boxed{\varepsilon_{i,j}}$$

```
summary(aov(score~test + Error(student/test)))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: student:test
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

Simple 1-way repeated measure (1/cell)

For the simplest possible repeated measures design: one within-subject factor and one observation per subject per factor level, we have three equivalent ways to specify the model.

```
summary(aov(score~test + student))
          Df Sum Sq Mean Sq F value    Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
student    9   5686   631.8    4.26 0.000487 ***
Residuals 45   6675   148.3
```

```
summary(aov(score~test + Error(student)))
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

```
summary(aov(score~test + Error(student/test)))
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: student:test
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

They all get the correct SS and F value for the effect of test.

They all get the correct within-subject error.

And they all factor out subject effects.

Simple 1-way repeated measure (1/cell)

For the simplest possible repeated measures design: one within-subject factor and one observation per subject per factor level, there are two **wrong ways** to specify the model

```
summary(aov(score~test))
            Df Sum Sq Mean Sq F value   Pr(>F)
test         5  11251  2250.2    9.83 1.01e-06 ***
Residuals   54  12361   228.9
```

WRONG: Don't add students to the model… Subject variability is now lumped in with the within-subject error. This is inefficient. Moreover, it will yield wrong answers when introducing more factors.

```
summary(aov(score~test*student))
              Df Sum Sq Mean Sq
test           5  11251  2250.2
student        9   5686   631.8
test:student  45   6675   148.3
```

WRONG: Adding a test:student interaction doesn't work because the test:student interaction *is* our within-subject error term. Adding the interaction means there is no error left over!

Simple 1-way repeated measure (1/cell)

For the simplest possible repeated measures design: one within-subject factor and one observation per subject per factor level, we have three equivalent ways to specify the model.

```
summary(aov(score~test + student))
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
student    9   5686   631.8    4.26 0.000487 ***
Residuals 45   6675   148.3
```

(1) Add student as a factor. Works here, but **will break if we have any between-subject factors.**

```
summary(aov(score~test + Error(student)))
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

(2) Add student as a general random effect. Works here, but **will break if we have more than 1 within-subject factor.**

```
summary(aov(score~test + Error(student/test)))
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: student:test
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

(3) Add student as a random effect, specifying the nested within-subject factors. Works here, and will work for all balanced mixed designs with one random effect (aov can't handle crossed random effects – see lmer)

Simple 1-way repeated measure (1/cell)

For the simplest possible repeated measures design: one within-subject factor and one observation per subject per factor level, we have three equivalent ways to specify the model.

```
summary(aov(score~test + Error(student/test)))
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   5686   631.8

Error: student:test
          Df Sum Sq Mean Sq F value   Pr(>F)
test       5  11251  2250.2   15.17 9.99e-09 ***
Residuals 45   6675   148.3
```

We will stick with the most general method, so we don't have to adopt a new syntax every time we change models.

1-way repeated measure (>1/cell)

We have one within-subject factor (test), and *more than one* observation per subject per factor level.

Factor A (within subject): Test *(index: j)*

Factor (random effect): Students *(index: i)*

	Midterm 201a	**Final 201a**	**Midterm 201b**	**Final 201b**
i=1	$y_{1,1,1} = 67$ $y_{1,1,2} = 69$ $y_{1,1,3} = 50$	$y_{1,2,1} = 76$ $y_{1,2,2} = 71$ $y_{1,2,3} = 75$	$y_{1,3,1} = 55$ $y_{1,3,2} = 50$ $y_{1,3,3} = 59$	$y_{1,4,1} = 89$ $y_{1,4,2} = 85$ $y_{1,4,3} = 93$
i=3	$y_{3,1,1} = 70$ $y_{3,1,2} = 73$ $y_{3,1,3} = 76$	$y_{3,2,1} = 69$ $y_{3,2,2} = 66$ $y_{3,2,3} = 62$	$y_{3,3,1} = 38$ $y_{3,3,2} = 41$ $y_{3,3,3} = 35$	$y_{3,4,1} = 53$ $y_{3,4,2} = 51$ $y_{3,4,3} = 56$
i=5	$y_{5,1,1} = 41$ $y_{5,1,2} = 45$ $y_{5,1,3} = 47$	$y_{5,2,1} = 87$ $y_{5,2,1} = 90$ $y_{5,2,1} = 92$	$y_{5,3,1} = 60$ $y_{5,3,2} = 64$ $y_{5,3,3} = 61$	$y_{5,4,1} = 70$ $y_{5,4,2} = 74$ $y_{5,4,3} = 72$
i=7	$y_{7,1,1} = 22$ $y_{7,1,2} = 25$ $y_{7,1,3} = 27$	$y_{7,2,1} = 58$ $y_{7,2,2} = 55$ $y_{7,2,3} = 53$	$y_{7,3,1} = 50$ $y_{7,3,2} = 47$ $y_{7,3,3} = 45$	$y_{7,4,1} = 62$ $y_{7,4,2} = 67$ $y_{7,4,3} = 65$

1-way repeated measure (>1/cell)

D2.data

```
   student            test rep score
1        1   201-A.midterm   1    67
1.6      1   201-A.midterm   2    79
1.7      1   201-A.midterm   3    63
1.1      1     201-A.final   1    79
1.1.1    1     201-A.final   2   101
1.1.2    1     201-A.final   3    89
1.2      1  201-A.homework   1    37
1.2.1    1  201-A.homework   2    28
1.2.2    1  201-A.homework   3    53
1.3      1   201-B.midterm   1    41
1.3.1    1   201-B.midterm   2    38
1.3.2    1   201-B.midterm   3    16
1.4      1     201-B.final   1    41
1.4.1    1     201-B.final   2    39
1.4.2    1     201-B.final   3    12
1.5      1  201-B.homework   1    45
1.5.1    1  201-B.homework   2    56
1.5.2    1  201-B.homework   3    61
...... ...... ...... ...... ......
10.4.2 10     201-B.final   3    55
10.5   10  201-B.homework   1    77
10.5.1 10  201-B.homework   2    81
10.5.2 10  201-B.homework   3    58
```

We have one within-subject factor (test), and *more than one* observation per subject per factor level.

Here we have 10 students, each being assessed on 6 different ‘tests’, with 3 scores for each test.

Total measurements: 180
Measurements/subject: 18
Measurements/sub-cell: 3

1-way repeated measure (>1/cell)

We have one within-subject factor (test), and *more than one* observation per subject per factor level.

D2.data

```
  student            test rep score
1        1  201-A.midterm   1    67
1.6      1  201-A.midterm   2    79
1.7      1  201-A.midterm   3    63
1.1      1    201-A.final   1    79
1.1.1    1    201-A.final   2   101
1.1.2    1    201-A.final   3    89
...... ...... ...... ...... ......
10.4.2 10    201-B.final   3    55
10.5   10 201-B.homework   1    77
10.5.1 10 201-B.homework   2    81
10.5.2 10 201-B.homework   3    58
```

What do we do with multiple observations per subject per level?

Option 1: Meh? Ignore it.

Option 2: Average to collapse them to 1 observation per subject per level. (not always possible)

Option 3: Specify the correct model to respect nesting structure.

1-way repeated measure (>1/cell)

We have 10 subjects. one within-subject factor (test: 6-levels), and *3* observation per subject per factor level.

What do we do with multiple observations per subject per level?

Option 1: Meh? Ignore it.

```
summary(aov(score~test + Error(student)))
```

```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: Within
           Df Sum Sq Mean Sq F value   Pr(>F)
test        5  13014  2602.7   10.56 8.14e-09 ***
Residuals 165  40649   246.4
```

BIG PROBLEM: This analysis assumes that every measurement is independent, but we may (and should!) expect that there may be some sort of interaction between test and student (e.g., some students are hung over for some tests, but not others). Thus, all measurements of that student-test will be correlated, because of this test:student interaction, and are not independent! This is like using multiple measurements of my height as independent samples of the population of male heights. WRONG!

1-way repeated measure (>1/cell)

We have 10 subjects. one within-subject factor (test: 6-levels), and *3* observation per subject per factor level.

What do we do with multiple observations per subject per level?

Option 2: Aggregate to get 1 measure/cell

```
D1.data.agg = D1.data %>%
    group_by(student,test) %>%
    summarize(score=mean(score))
```

```
   student          test     score
1        1    201-A.final  89.66667
2        2    201-A.final  90.66667
3        3    201-A.final  58.00000
4        4    201-A.final  82.66667
5        5    201-A.final  75.33333
6        6    201-A.final  42.00000
7        7    201-A.final  44.00000
8        8    201-A.final  65.33333
9        9    201-A.final 105.00000
10      10    201-A.final  46.33333
11       1 201-A.homework  39.33333
. . . . . . . . . . . . . . . . . . .
60      10  201-B.midterm  68.33333
```

So now, instead of having 180 measurements (with 3 per subject per test) we have 60 measurements with 1 per subject per cell. With that 1 corresponding to the average of the 3 we had before.

1-way repeated measure (>1/cell)

We have 10 subjects. one within-subject factor (test: 6-levels), and *3* observation per subject per factor level.

What do we do with multiple observations per subject per level?
Option 2: Aggregate to get 1 measure/cell

```
summary(aov(score~test + Error(student), data=D1.data.agg))
```

```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   4965   551.7

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
test       5   4338   867.6   4.648 0.00167 **
Residuals 45   8399   186.6
```

Everything looks peachy, and this is the correct answer.

But… this strategy will not work if we have multiple within-subject factors!!

1-way repeated measure (>1/cell)

We have 10 subjects. one within-subject factor (test: 6-levels), and *3* observation per subject per factor level.

What do we do with multiple observations per subject per level?

Option 3: Specify the correct nesting structure

```
summary(aov(score~test + Error(student/test)))
```

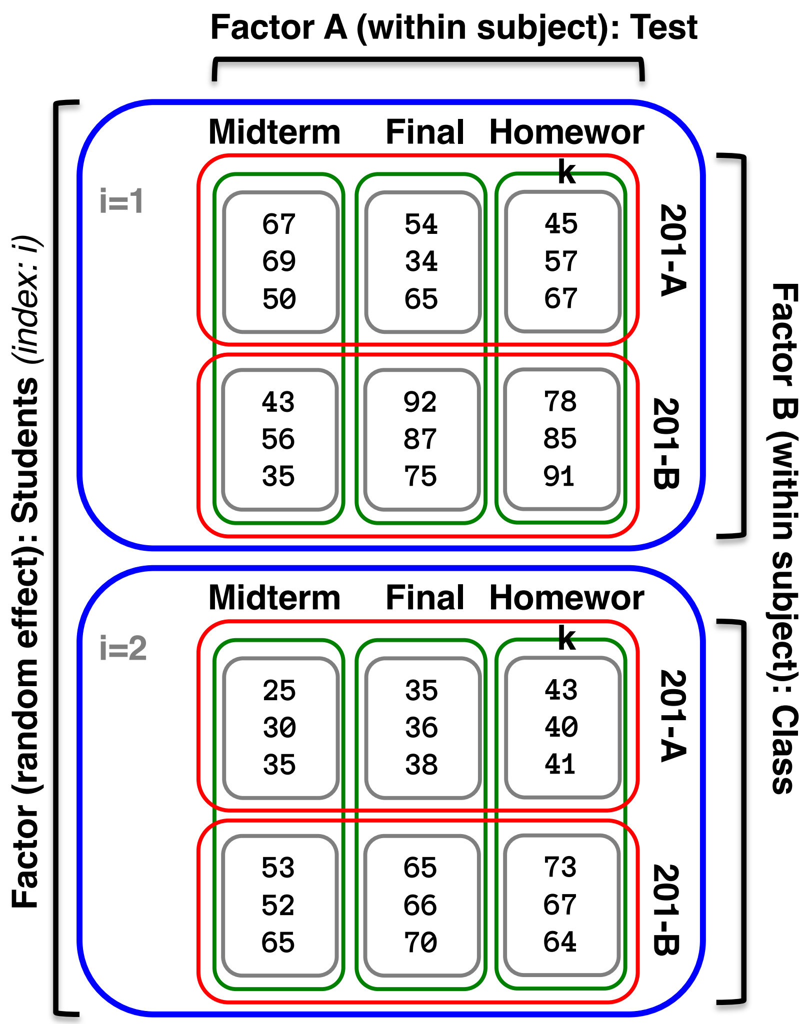
```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:test
          Df Sum Sq Mean Sq F value  Pr(>F)
test       5  13014  2602.7   4.648 0.00167 **
Residuals 45  25197   559.9

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

Note: We get the same answer as option 2 for the effect of test. But critically, we've clarified that the relevant error for the effect of test is the student:test interaction. The 'within' error, is the variability of multiple measurements per subject per test.

This is the general strategy we need to use if we have multiple within-subject factors.

n-way repeated measures

We have multiple within-subject factors (class and test), and potentially, >1 measurement per subject per cell.



n-way repeated measures

D3.data

```
student class      test rep score
      1 201-A  midterm   1    67
      1 201-A  midterm   2    79
      1 201-A  midterm   3    63
      1 201-A    final   1    79
      1 201-A    final   2   101
      1 201-A    final   3    89
      1 201-A homework   1    37
      1 201-A homework   2    28
      1 201-A homework   3    53
      1 201-B  midterm   1    41
      1 201-B  midterm   2    38
      1 201-B  midterm   3    16
      1 201-B    final   1    41
      1 201-B    final   2    39
      1 201-B    final   3    12
      1 201-B homework   1    45
      1 201-B homework   2    56
      1 201-B homework   3    61
      2 201-A  midterm   1    51
      2 201-A  midterm   2    67
      2 201-A  midterm   3    85
. . . . . . . . . . . . . . . . . . 
     10 201-B homework   3    58
```

We have two within-subject factors (class and test).

Here we have 10 students, each being assessed on 3 'tests' in 2 classes, with 3 scores for each test.

Total measurements: 180
Measurements/subject: 18
Measurements/sub-cell: 3

But now we have 2 within-subject factors!

n-way repeated measures

```
D3.data
student class      test rep score
      1 201-A  midterm   1    67
      1 201-A  midterm   2    79
. . . . . . . . . . . . . . . . . .
     10 201-B homework   3    58
```

We have two within-subject factors (class and test).

What do we do with multiple within-subject factors?

We can't ignore it, and we can't average to reduce to just one measurement.

Option 3: Specify the correct model to respect nesting structure.

n-way repeated measures

Option 3: Specify the correct nesting structure

```
summary(aov(score~class*test + Error(student/(class*test))
```

```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

Different student:factor interactions yield the appropriate error terms which we compare to various within-subject factor effects.

Now the general formulation
Y~A*..*K + Error(Sub/(A*B))
Makes sense: we have to specify which factors are nested within subjects, so we can get all the right error terms out.

n-way repeated measures

NOT SPECIFYING NESTING STRUCTURE GIVES WRONG ERROR TERMS!

```
summary(aov(score~class*test + Error(student)))
```

```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: Within
            Df Sum Sq Mean Sq F value   Pr(>F)
class        1   1755    1755   7.123  0.00837 **
test         2   1957     978   3.972  0.02067 *
class:test   2   9302    4651  18.879 4.14e-08 ***
Residuals  165  40649     246
```

Because there is some correlation among multiple measurements of the same subject in the same condition, if we do not appropriately specify which conditions are nested in subjects, we do not account for these correlations. Thus, these correlation will yield spurious effects and Type 1 errors.

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))
```

```
Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

We partition into various error "Strata"

- sums of squares
- Degrees of freedom

Within each error stratum we partition the sums of squares and degrees of freedom into

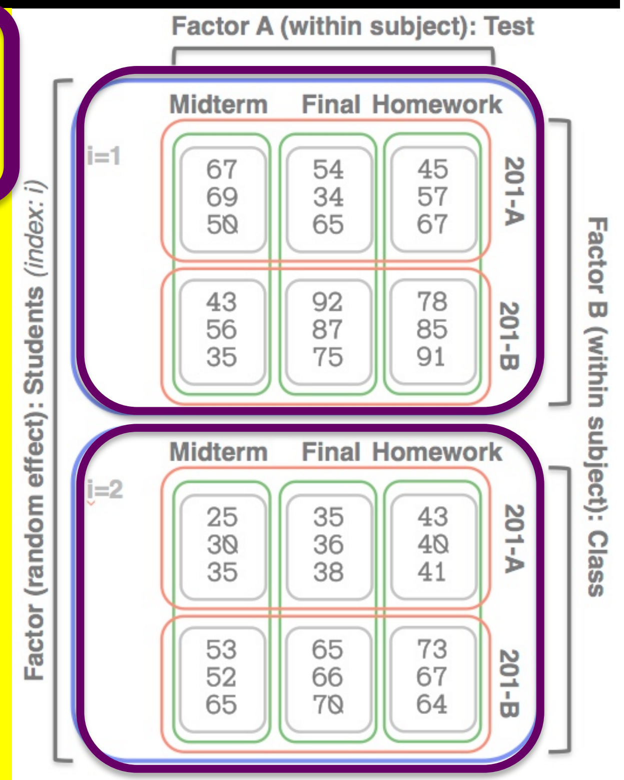
- Explanatory variables
- Residuals

We then compute MS[effect]/MS[residual] within the error strata to get F ratios.

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

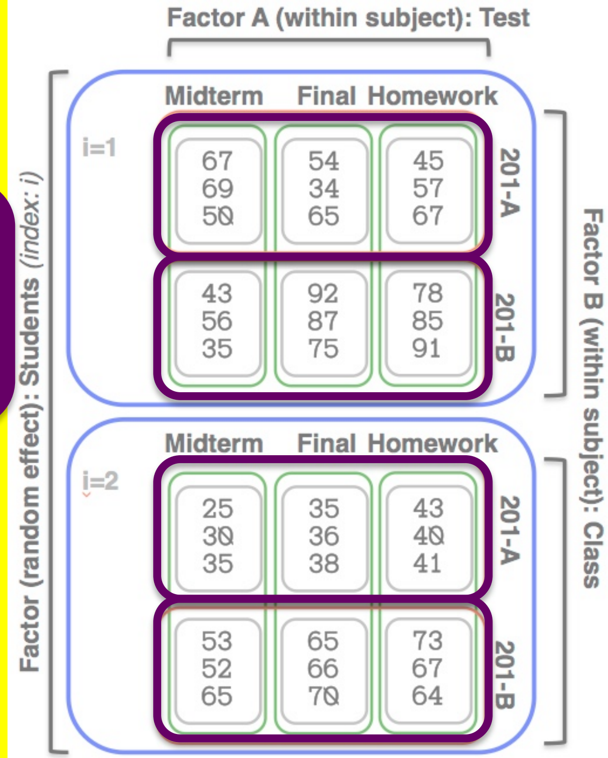


How much variability is there in student scores averaging over all tests for each student?

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

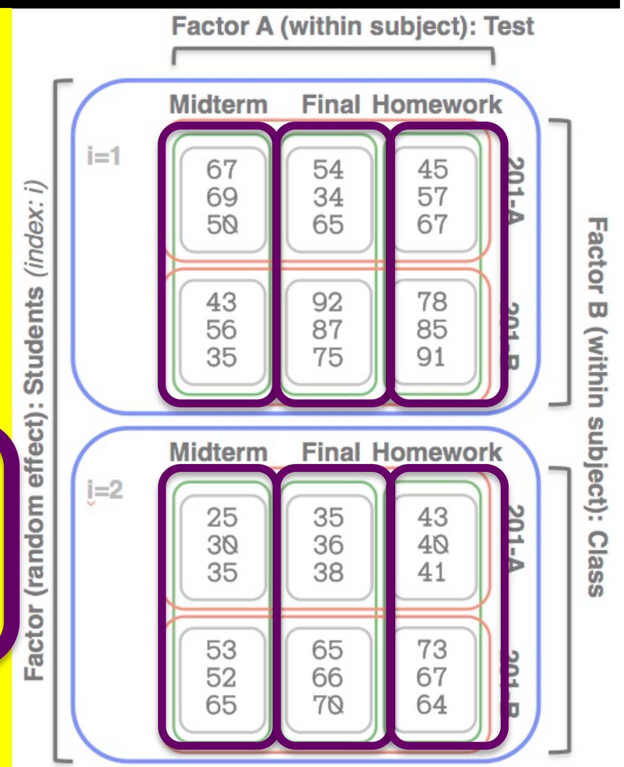


How much variability is there in average student-class scores (after factoring out student average), and how much of that can we explain with

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

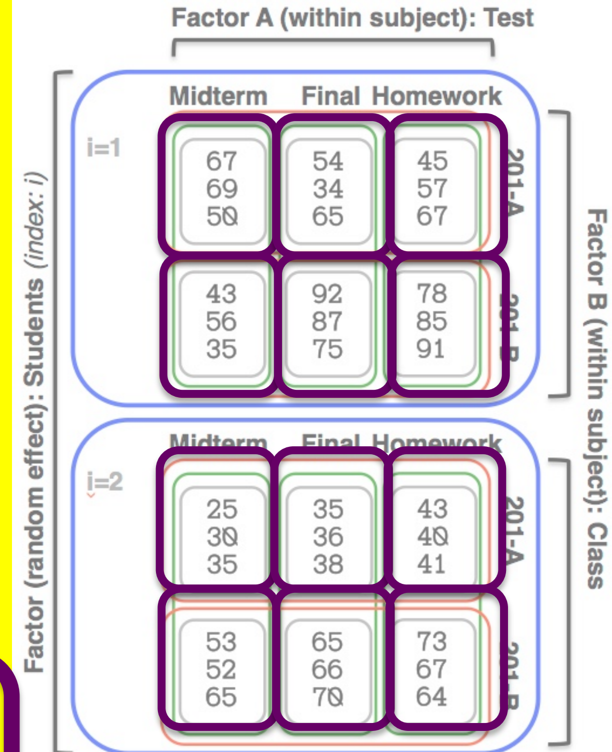


How much variability is there in average student-test scores (after factoring out student average), and how much of that can we explain with test effects?

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```

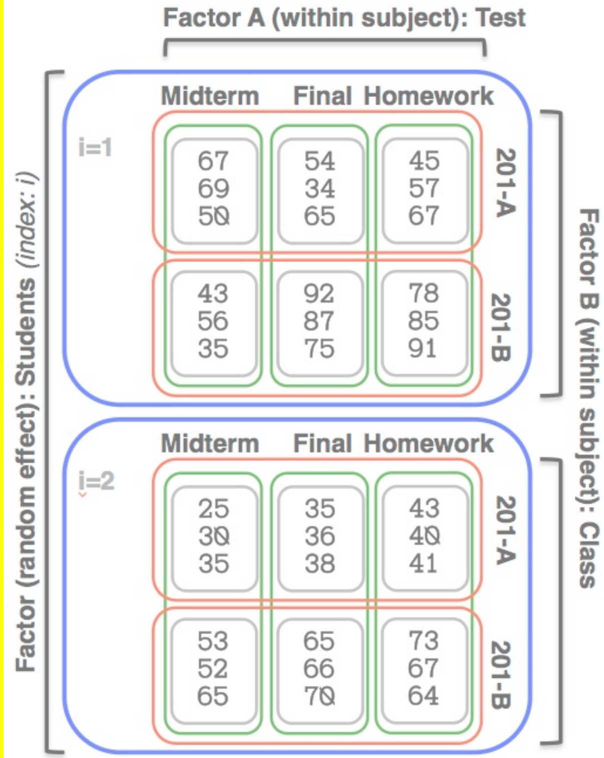


How much variability is there in average student-class-test scores (after factoring out student average, student-class average and student-test average), and how much of that can we explain with the class:test interaction?

n-way repeated measures

What's happening here?

```
summary(aov(score~class*test + Error(student/(class*test))

Error: student
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  14896    1655

Error: student:class
          Df Sum Sq Mean Sq F value Pr(>F)
class      1   1755  1754.7   2.378  0.157
Residuals  9   6641   737.8

Error: student:test
          Df Sum Sq Mean Sq F value Pr(>F)
test       2   1957   978.4   2.392   0.12
Residuals 18   7363   409.0

Error: student:class:test
           Df Sum Sq Mean Sq F value  Pr(>F)
class:test  2   9302    4651   7.479 0.00432 **
Residuals  18  11194     622

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 120  15452   128.8
```



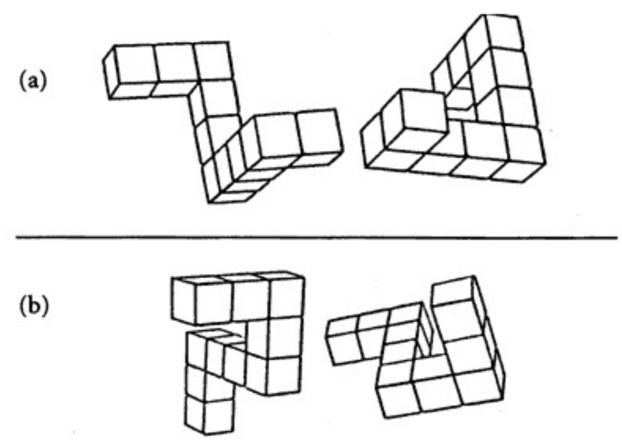
How much variability is there within student-class-test cells (after factoring out the student-class-test average)?

Blocking

- We want to factor out subject effects via repeated-measures analysis, but we may not be able to do a within-subject design.
  - We are comparing autistic to typical kids.
  - We are assigning students to different classrooms.
  - We are performing different kinds of surgery on patients.
  - We give different sorts of drugs to patients.
  - Etc.
- One approach: create fake ‘pseudo-subjects’ or ‘blocks’ that we think might share common sources of variability
  - Block kids based on IQ
  - Block students based on SES
  - Block patients based on severity of ailment

Complete Block Designs

- You test kids:
  - w/ Autism
  - w/ Williams
  - w/ Downs
  - Controls
- Do they differ in spatial reasoning tasks controlling for SES?
  - Create SES blocks



Mental Rotation Test—Are these two figures the same except for their orientation?

Complete Block Designs

SES measure (income percentile)

Controls	Autistics	Downs	Williams
93	76	74	38
15	82	59	12
35	24	16	95
47	48	25	67
54	6	41	52
68	35	39	86
3	15	96	6
85	99	86	44
79	66	2	29
27	55	66	77

SES measure (income percentile)

	Autistics	Downs	Williams	Controls
block 1	6	2	6	3
block 2	15	16	12	15
block 3	24	25	29	27
block 4	35	39	38	35
block 5	48	41	44	47
block 6	55	59	52	54
block 7	66	66	67	68
block 8	76	74	77	79
block 9	82	86	86	85
block 10	99	96	95	93

- Create SES blocks
- Effectively: you've created a new factor.
  - Each level of that factor corresponds to some approximate value of SES.
  - Each kid group has 1 member in each block
  - Kids are very closely matched within block
  - This is a 'complete block' design.
- The blocks serve as repeated measures, and you can factor out variability due to SES block!

Complete Block Designs

SES measure (income percentile)

	Autistics	Downs	Williams	Controls
block 1	6	2	6	3
block 2	15	16	12	15
block 3	24	25	29	27
block 4	35	39	38	35
block 5	48	41	44	47
block 6	55	59	52	54
block 7	66	66	67	68
block 8	76	74	77	79
block 9	82	86	86	85
block 10	99	96	95	93

Spatial reasoning measure…

	Autistics	Downs	Williams	Controls
block 1	41	74	49	70
block 2	28	40	36	24
block 3	14	35	14	88
block 4	56	42	36	19
block 5	44	80	85	17
block 6	18	83	80	45
block 7	44	50	44	83
block 8	19	86	66	67
block 9	89	70	36	79
block 10	62	79	75	76

- Create complete SES blocks
- To analyze this, you would include the blocking measure as a repeated measure

```
aov(spatial.reasoning ~ kid.type + Error(ses.block / kid.type))
```

  – Since this is now just a 1 factor repeated measures design, with one observation per unit-level, you could just add ses.block as a factor, but let’s stick with being explicit.

Complete Block Designs

SES measure (income percentile)

	Autistics	Downs	Williams	Controls
block 1	6	2	6	3
block 2	15	16	12	15
block 3	24	25	29	27
block 4	35	39	38	35
block 5	48	41	44	47
block 6	55	59	52	54
block 7	66	66	67	68
block 8	76	74	77	79
block 9	82	86	86	85
block 10	99	96	95	93

Spatial reasoning measure...

	Autistics	Downs	Williams	Controls
block 1	41	74	49	70
block 2	28	40	36	24
block 3	14	35	14	88
block 4	56	42	36	19
block 5	44	80	85	17
block 6	18	83	80	45
block 7	44	50	44	83
block 8	19	86	66	67
block 9	89	70	36	79
block 10	62	79	75	76

- Create complete SES blocks
- This is the ideal world where you can create complete blocks because magically, we had one kid from every group in every income range. This is unlikely to happen.
  - Often: do a *yoked* experiment – pick control kids to match a special population kid. Feasible for one special population, not three. (often used to equate subject-driven

Complete Block Designs

SES measure (income percentile)

Autistics	Downs	Williams	Controls
2	5	4	19
15	20	17	39
60	23	18	45
72	25	22	45
75	49	30	50
80	76	36	54
82	76	44	56
83	78	44	66
93	95	78	70

SES measure (income percentile)

	Autistics	Downs	Williams	Controls
? Block 1	2	5	4	19
? Block 2	15	20	17	39
? Block 3	60	23	18	45
? Block 4	72	25	22	45
? Block 5	75	49	30	50
? Block 6	80	76	36	54
? Block 7	82	76	44	56
? Block 8	83	78	44	66
? Block 9	93	95	78	70

- Real world: force blocking?
- This doesn’t work. Our income measures are not distributed in a matched way across special populations, so our blocks don’t have a useful meaning.

Blocking in the real world

- Real world blocking across populations is often hard, and requires some forethought
  - E.g. yoked designs.
- If you are doing random assignment to conditions, then blocking becomes much easier.
  - Someone comes in, you get a measure of their blocking variable, then assign them to a condition to ensure a complete block.
  - This also yields a 'randomized complete block' design, where the blocking factor has no relationship to the treatment factor(s).
    (if we have two factors outside of our control [SES and diagnosis], we don't get to randomize).

Factoring out extraneous variability

- There are sources of variability besides our factors of interest. We want to to account for this variability, to reduce error and gain power.
  - **Repeated measures**: we measure each experimental unit (subject / family / school) in every treatment, this way factoring out all the variability due to the experimental unit.
    Very powerful – always do this if you can.
  - **Blocks**: Analysis/design constructs where we pool individuals matched on some variable into blocks. This is a good idea, but not always easy to do.
  - **Covariates**: we measure continuous variables that contribute linearly to our measure of interest, and factor them out via regression (ANCOVA, later).

Factoring out extraneous variability

- We aim to gain power by factoring out variability
  - We decrease SS[error] and df[error].
  - We'd like the decrease in SS[error] to be a bigger proportion than the decrease in df[error] so that MS[error] also drops.
  - Repeated measures
    - df[error] loss: # of subjects (-1)
  - Blocks:
    - df[error] loss: # of blocks (-1)
  - Covariates:
    - df[error] loss: # of covariates.

- Measurements of the same thing are correlated.
- Why use ‘repeated measures’ designs?

- 1 within-subject factor, 1 measure per cell per subject
- 1 within-subject factor, >1 measure per cell per subject
- >1 within-subject factors
- Mixed designs: within and between subject effects

- What’s the right error for each effect?

- Blocking as repeated measures.

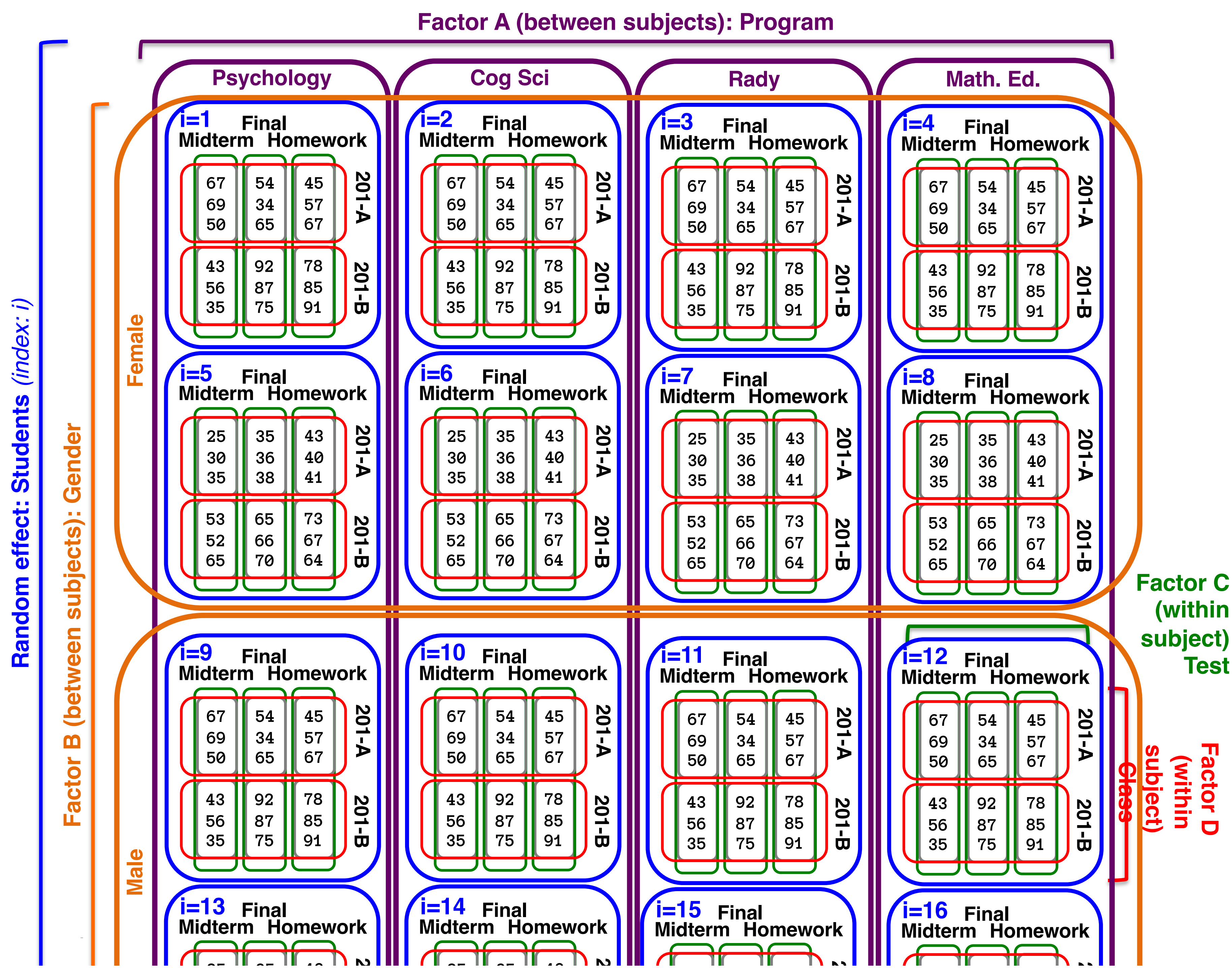
We have 10 teachers each teach 12 classes in a 2-way factorial (within-teacher) design: topic [physics, algebra, history, art] * method [lecture, "discovery", "flipped"]. In each class we measure the pre and post-class score on a standardized test for that topic for each of 15 students, and then record their pre-to-post percentile improvement (so we get one improvement number per student).

- Write:
  - The R command you would use to do a repeated measures analysis on the effects of topic, teaching method, and their interaction on student improvement.
  - Write out the structure of the ANOVA table(s) we would expect to get from this analysis, including the degrees of freedom for each entry.

Source	Df	SS	MS	F	P
Err: teacher	*9*	*5000*			
Residuals	9	5000	555.6		
Err: teacher:topic	*30*	*1500*			
topic	3			4	
Residuals	27				
Err: teacher:method	*20*				
method	2		163		0.01
Residuals	18				
Err: teacher:topic:method	*60*				
topic:method	6			3	
Residuals	54				
Err: within	*1680*	*10000*			
Residuals	1680	10000			
Total	1799	20000			

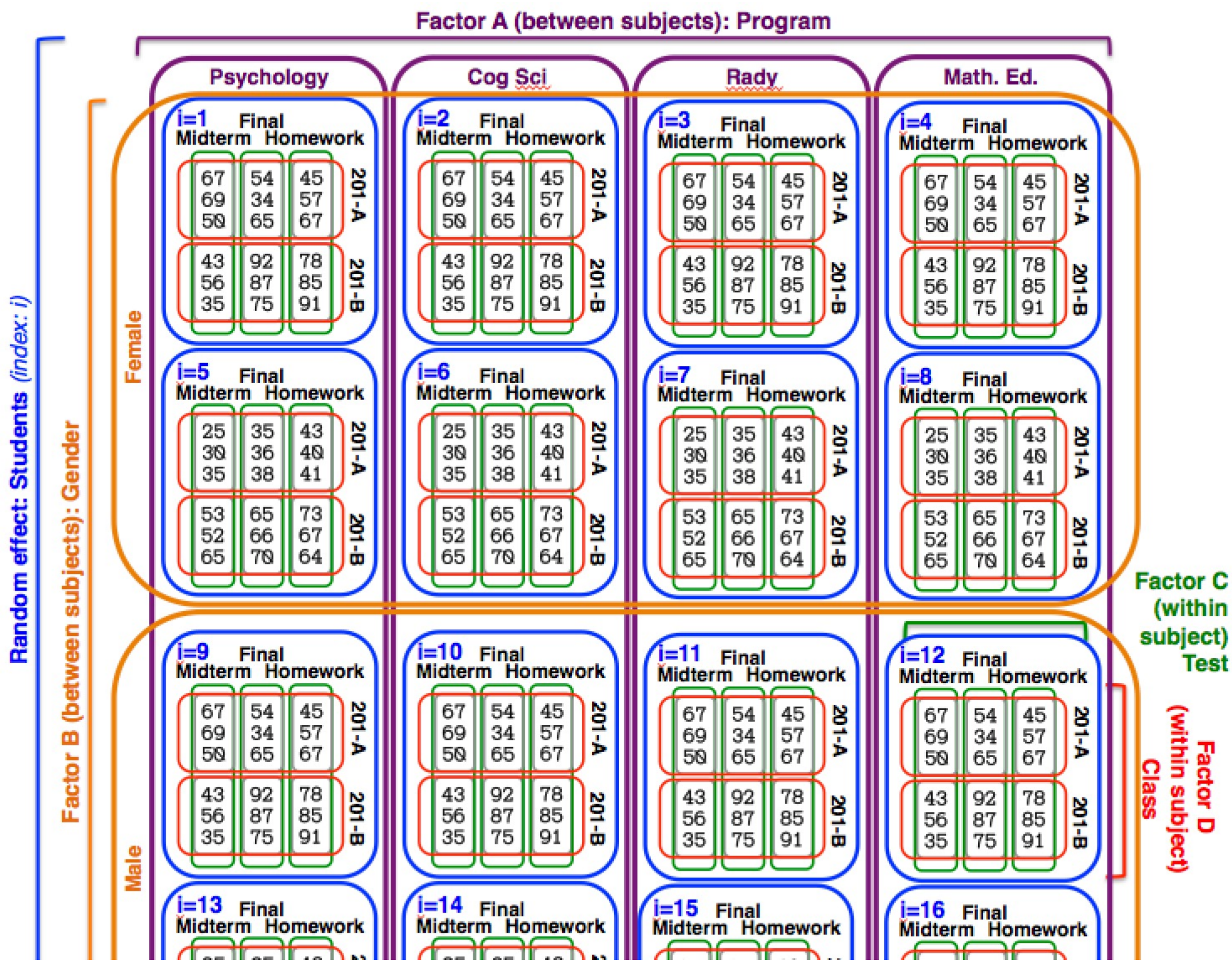
Source	Df	SS	MS	F	P
Err: teacher	*9*	***5000***			
Residuals	9	5000	556		
Err: teacher:topic	*30*	*1500*			
topic	3	462	154	4	0.018
Residuals	27	1038	38		
Err: teacher:method	*20*	*814*			
method	2	326	163	6	0.01
Residuals	18	488	27		
Err: teacher:topic:method	*60*	*2686*			
topic:method	6	672	112	3	0.013
Residuals	54	2015	37		
Err: within	*1680*	*10000*			
Residuals	1680	10000	6		
Total	1799	20000			

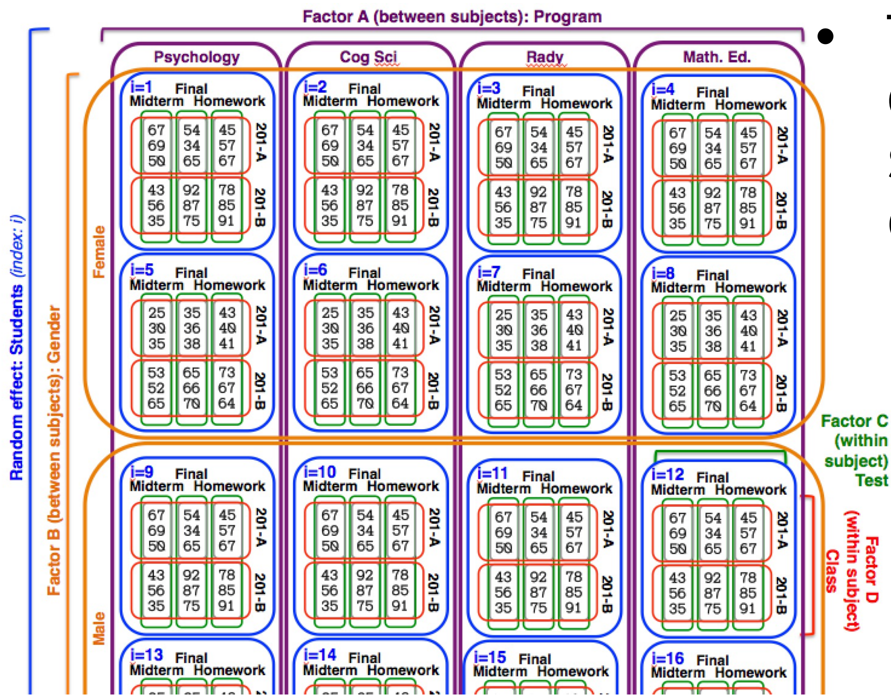
What (roughly) do the different MS/SS terms measure?


Factor A (between subjects): Program
Psychology
Cog Sci
Rady
Math. Ed.
Random effect: Students (index: i)
Factor B (between subjects): Gender
Female
Male
i=1
i=2
i=3
i=4
i=5
i=6
i=7
i=8
i=9
i=10
i=11
i=12
i=13
i=14
i=15
i=16
Final
Midterm Homework
201-A
201-B
Factor C (within subject) Test
Factor D (within subject) Class

Mixed design (factors within,between)

We have multiple within-subject factors (class and test), and potentially, ›1 measurement per subject per cell.
Subjects, are in various between subject conditions.





- The complicated nesting structure of our variables means that different sources of variability will yield correlations.
  - Some students do better than others, this will influence *all their scores.*
  - Some students suffer under time pressure (this will influence *all* their midterm/final scores)
  - Some students get complacent in one class, or another. This will influence all their scores in a class.
  - Some students were hung over on some day. This will influence all their scores on that day.
- Our task is to factor out these different independent sources of variability, and then see if they can be explained by our factors.

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

All the factors, crossed.

Random effect of student, with nesting of within-student conditions specified.

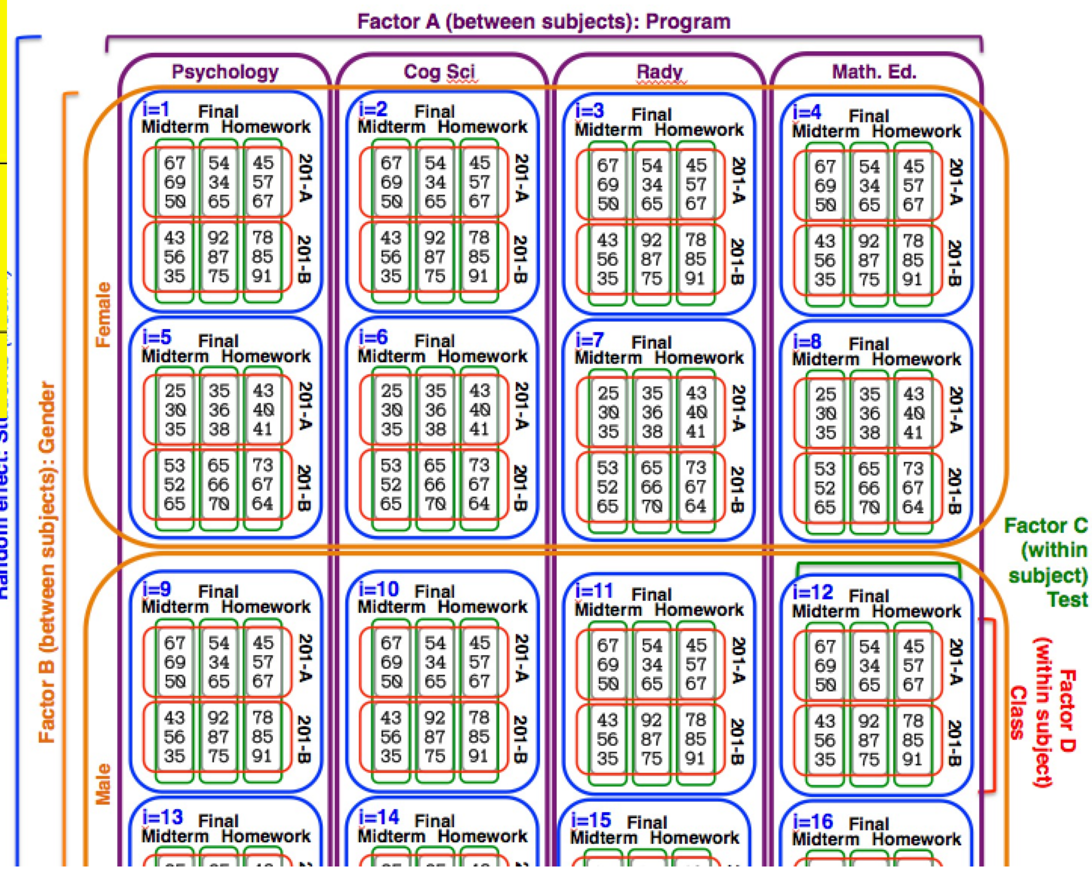
```
Error: student
           Df Sum Sq Mean Sq F value   Pr(>F)
sex         1  40524   40524  20.409 5.43e-05 ***
program     4  21430    5357   2.698   0.0442 *
sex:program 4   3712     928   0.467   0.7593
Residuals  40  79422    1986

Error: student:class
                 Df Sum Sq Mean Sq F value  Pr(>F)
class             1   1933  1933.2   9.072 0.00448 **
sex:class         1     81    80.6   0.378 0.54204
program:class     4    417   104.3   0.490 0.74329
sex:program:class 4    208    52.1   0.245 0.91127
Residuals        40   8524   213.1

Error: student:test
                Df Sum Sq Mean Sq F value   Pr(>F)
test             2   7826    3913  16.213 1.23e-06 ***
sex:test         2    321     160   0.665    0.517
program:test     8   1253     157   0.649    0.734
sex:program:test 8   2365     296   1.225    0.295
Residuals       80  19306     241

Error: student:class:test
                      Df Sum Sq Mean Sq F value   Pr(>F)
class:test             2   6310  3155.1  13.579 8.37e-06 ***
sex:class:test         2    387   193.5   0.833    0.439
program:class:test     8   1342   167.8   0.722    0.671
sex:program:class:test 8   1775   221.9   0.955    0.477
Residuals             80  18588   232.4

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Mixed design (factors within,between)

```
score ~ sex*program*class*test +
                    Error(student/(class*test))

Error: student
             Df Sum Sq Mean Sq F value   Pr(>F)
sex           1  40524   40524  20.409 5.43e-05
program       4  21430    5357   2.698   0.0442
sex:program   4   3712     928   0.467   0.7593
Residuals    40  79422    1986
Error: student:class
                   Df Sum Sq Mean Sq F value  Pr(>F)
class               1   1933  1933.2   9.072 0.00448
sex:class           1     81    80.6   0.378 0.54204
program:class       4    417   104.3   0.490 0.74329
sex:program:class   4    208    52.1   0.245 0.91127
Residuals          40   8524   213.1
Error: student:test
                  Df Sum Sq Mean Sq F value   Pr(>F)
test               2   7826    3913  16.213 1.23e-06
sex:test           2    321     160   0.665    0.517
program:test       8   1253     157   0.649    0.734
sex:program:test   8   2365     296   1.225    0.295
Residuals         80  19306     241
Error: student:class:test
                        Df Sum Sq Mean Sq F value   Pr(>F)
class:test               2   6310  3155.1  13.579 8.37e-06
sex:class:test           2    387   193.5   0.833    0.439
program:class:test       8   1342   167.8   0.722    0.671
sex:program:class:test   8   1775   221.9   0.955    0.477
Residuals               80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

What's all this craziness?

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained variance.



Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Each of these different sources of variability gets its own ANOVA:
Each one has a total sum of squares (e.g., SS[students], SS[student:class], etc.) and these get divided up into SS[factors] and SS[error].
The F value is computed within each ANOVA in the standard way
F = MS[factor]/MS[error]
With df[factor] and df[error]

So, after we understand what has been factored, how, and why, the rest is reasonably straight forward.

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
```

Between-subject error: variation of average performance across students. This is the relevant variability against which to compare between-subject effects.

```
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
```

Interaction of students and classes: this is the relevant variability against which to compare all class effects (main effect of class, and class:between-factor interactions.

```
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
```

Interaction of students and test: relevant variability against which to compare all test effects (main effect of test, and test:between-Ss interactions.

```
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06 ***
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
```

Interaction of students and class-test combination: relevant variability against which to compare all class:test effects

```
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Variability across multiple measurements of a student in a class on a test: not relevant.

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Student variability (overall student performance)

This can be explained by Their sex (e.g., women do better than men)

Their program (e.g., Rady students do better)

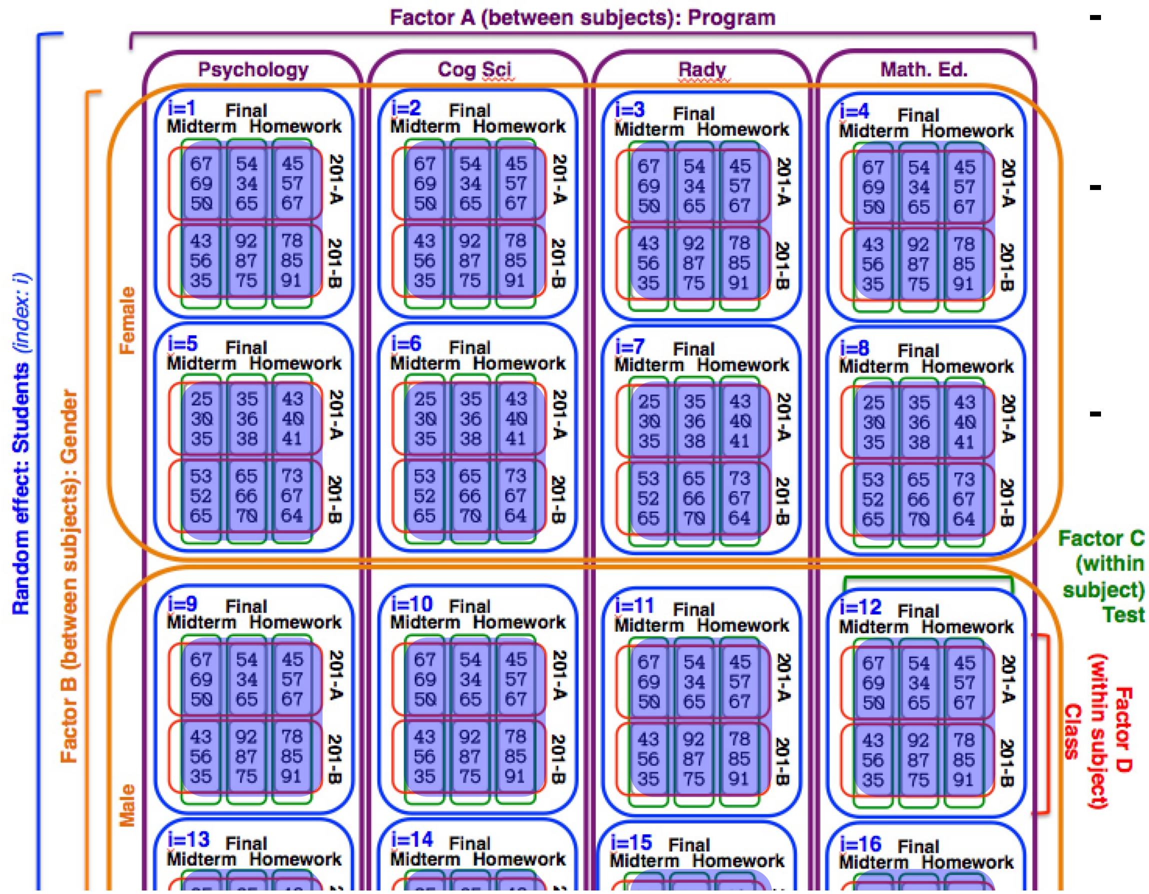
The sex:program interaction (e.g., discrepancy between sex is different in the different programs)

The remaining idiosyncratic variability of each subject (error/residuals)

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained variance

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
```



Between subject variability.

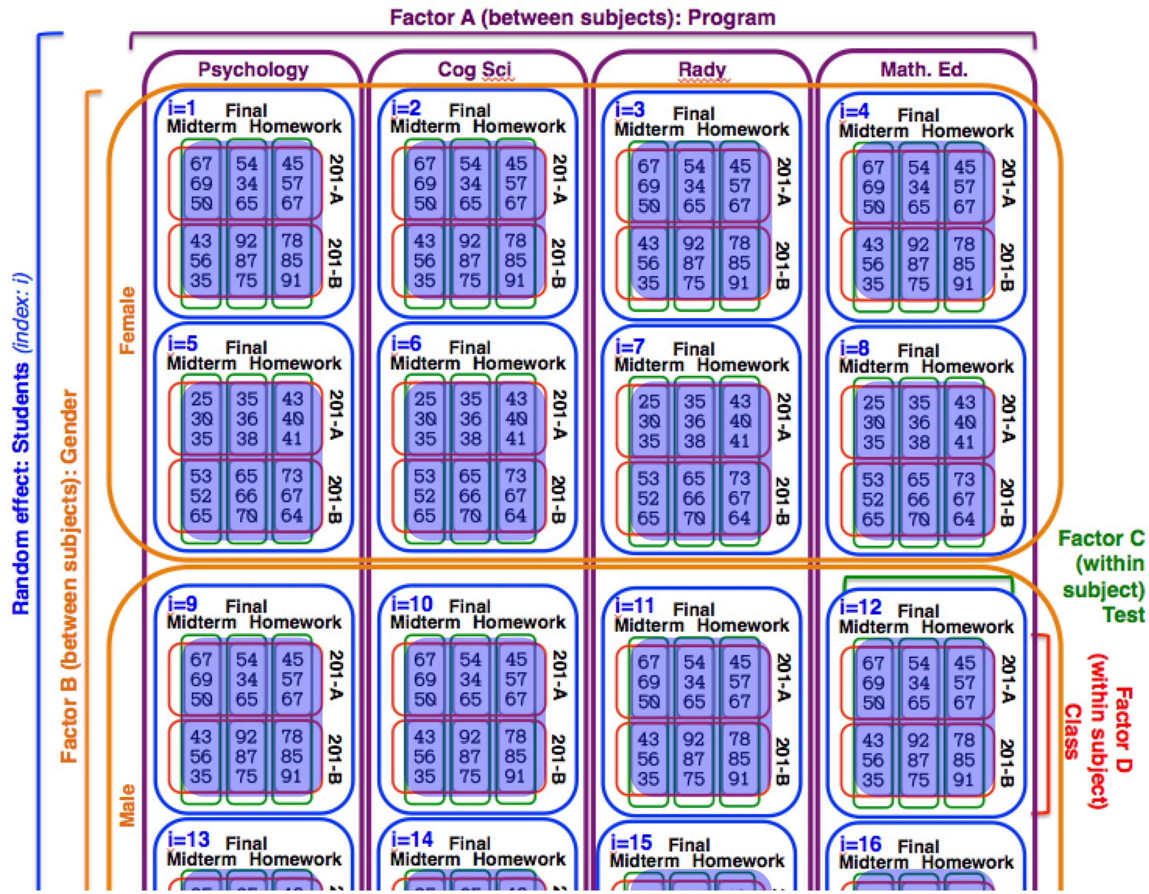
Roughly:

- Average all data within a subject.
- Sum of squares of those subject averages is the total subject variability.
- Subject variability is divided into variability attributable to sex, program, sex:program interaction, and the residual error.
- Tests compare subject variability explained by these between-subject factors, and the remaining subject error.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained variance

```
Error: student
             Df Sum Sq Mean Sq F value    Pr(>F)
sex           1  40524   40524  20.409 5.43e-05 ***
program       4  21430    5357   2.698   0.0442 *
sex:program   4   3712     928   0.467   0.7593
Residuals    40  79422    1986
```



Between subject variability.

How many degrees of freedom should there be in the between-subject variability, and how does it get divided?

- Number of subjects (here 5*5*2) -1: 49
  These get divided into
- K-1 for each between subject factor (2-1, and 5-1 here)
- (Ka-1)*(Kb-1) for the interaction (2-1)*(5-1)
- Remainder into error (40)

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Student:class variability (factoring out overall student performance, did they do better in one class or another on average)

This can be explained by

- The class (e.g., everyone does bettern in 201a)
- Sex:Class interaction (e.g., men do better in 201b, women in 201a)
- The program:class interaction (e.g., psych students do better in 201a, rady in 201b)
- Program:Sex:class interaction (e.g., gender difference in 201a-201b performance differs by program)

Remaining error.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:class
                    Df Sum Sq Mean Sq F value  Pr(>F)
class                1   1933  1933.2   9.072 0.00448
sex:class            1     81    80.6   0.378 0.54204
program:class        4    417   104.3   0.490 0.74329
sex:program:class    4    208    52.1   0.245 0.91127
Residuals           40   8524   213.1
```

Student:class variability.

Roughly:

- Subtract subject effect from all data.
- Average result within each subject:class group.
- Compute SS of these averages.
- Assess whether this variability can be explained by class, or by class interacting with the between-subject factors.
- Compare explained and to unexplained variability.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:class
                   Df Sum Sq Mean Sq F value  Pr(>F)
class               1   1933  1933.2   9.072 0.00448
sex:class           1     81    80.6   0.378 0.54204
program:class       4    417   104.3   0.490 0.74329
sex:program:class   4    208    52.1   0.245 0.91127
Residuals          40   8524   213.1
```

How many degrees of freedom should there be in the student:class variability, and how does it get divided?

- # subjects * # classes - # subjects (because we factor out the subject effects): 50*2-50 = 50

These get divided into

- 2-1 for class factor
- (2-1)*(2-1) for sex:class interaction
- (2-1)*(5-1) for program:class interaction
- (2-1)*(2-1)*(5-1) for sex:program:class interaction

Remainder into error (40)

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Student:test variability (factoring out overall student performance, did they do better on midterms/finals/homework)

This can be explained by

- The test (e.g., everyone does better on hw)
- Sex:test interaction (e.g., men do well on midterm, women on finals)
- The program:tes interaction (e.g., psych students do better in hw, rady on finals)
- Program:Sex:test interaction (e.g., gender difference in midterm-final performance differs by program)

Remaining error.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
```

Student:test variability.

Roughly:

- Subtract subject effect from all data.
- Average result within each subject:test group.
- Compute variability of these averages.
- Assess whether this variability can be explained by test, or by testinteracting with the between-subject factors.
- Compare explained and to unexplained variability.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:test
                  Df Sum Sq Mean Sq F value   Pr(>F)
test               2   7826    3913  16.213 1.23e-06
sex:test           2    321     160   0.665    0.517
program:test       8   1253     157   0.649    0.734
sex:program:test   8   2365     296   1.225    0.295
Residuals         80  19306     241
```

How many degrees of freedom should there be in the student:test variability, and how does it get divided?

- # subjects * # tests - # subjects (because we factor out the subject effects):
  50*3-50 = 100

These get divided into

- 3-1 for test factor
- (3-1)*(2-1) for sex:test interaction
- (3-1)*(5-1) for program:test interaction
  (3-1)*(2-1)*(5-1) for sex:program:test interaction
  Remainder into error (80)

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
             Df Sum Sq Mean Sq F value   Pr(>F)
sex           1  40524   40524  20.409 5.43e-05 ***
program       4  21430    5357   2.698   0.0442 *
sex:program   4   3712     928   0.467   0.7593
Residuals    40  79422    1986
Error: student:class
                   Df Sum Sq Mean Sq F value  Pr(>F)
class               1   1933  1933.2   9.072 0.00448 **
sex:class           1     81    80.6   0.378 0.54204
program:class       4    417   104.3   0.490 0.74329
sex:program:class   4    208    52.1   0.245 0.91127
Residuals          40   8524   213.1
Error: student:test
                  Df Sum Sq Mean Sq F value   Pr(>F)
test               2   7826    3913  16.213 1.23e-06 ***
sex:test           2    321     160   0.665    0.517
program:test       8   1253     157   0.649    0.734
sex:program:test   8   2365     296   1.225    0.295
Residuals         80  19306     241
Error: student:class:test
                        Df Sum Sq Mean Sq F value   Pr(>F)
class:test               2   6310  3155.1  13.579 8.37e-06
sex:class:test           2    387   193.5   0.833    0.439
program:class:test       8   1342   167.8   0.722    0.671
sex:program:class:test   8   1775   221.9   0.955    0.477
Residuals               80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Student:class:test variability (factoring out overall student performance, class performance, and test performance, did they do better on particular combinations of 201a/201b and midterm/final/hw?)

This can be explained by

- Class:test (e.g., everyone does better on 201a-final)
- Sex:class:test interaction (e.g., men do well on 201a midterm, women on 201b final)
- The program:class:test interaction (e.g., psych students do better in 201a-hw, rady on 201b-finals)
- Program:Sex:class:test (e.g., gender difference in 201a-midterm-201b-final differs by program)

Remaining error.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:class:test
                       Df Sum Sq Mean Sq F value    Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833     0.439
program:class:test      8   1342   167.8   0.722     0.671
sex:program:class:test  8   1775   221.9   0.955     0.477
Residuals              80  18588   232.4
```

Student:class:test variability.

Roughly:

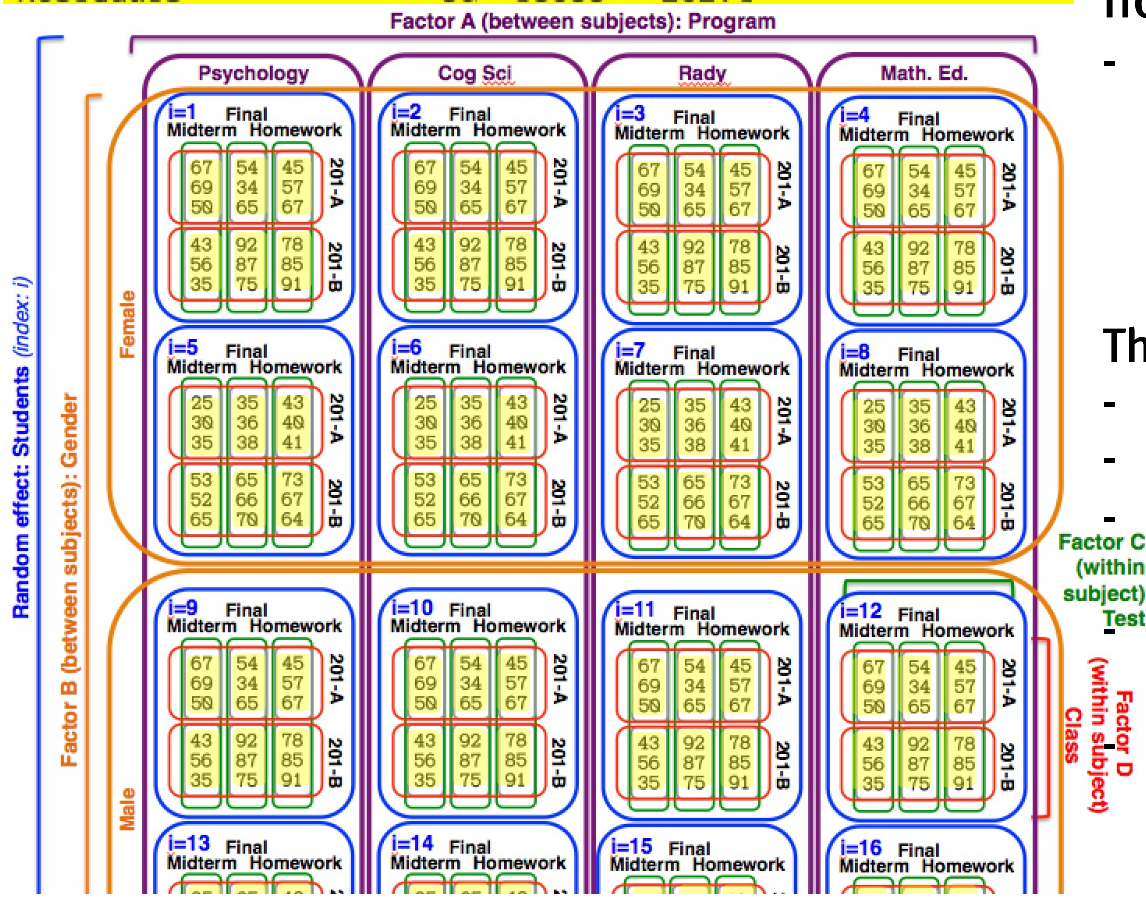
- Subtract subject effect, subject:class interaction and subject:test interaction from all data.
- Average result within each subject:class:test group.
- Compute variability of these averages.
- Assess whether this variability can be explained by class:test, or by class:test interacting with the between-subject factors.

Compare explained and to unexplained variability.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained

```
Error: student:class:test
                         Df Sum Sq Mean Sq F value   Pr(>F)
class:test                2   6310  3155.1  13.579 8.37e-06
sex:class:test            2    387   193.5   0.833    0.439
program:class:test        8   1342   167.8   0.722    0.671
sex:program:class:test    8   1775   221.9   0.955    0.477
Residuals                80  18588   232.4
```



How many degrees of freedom should there be in the student:class:test variability, and how does it get divided?

- # subjects * # classes*# tests - # subjects – df. Student:class – df student:test
  50*2*3-50-50-100 =
  50*(2-1)*(3-1) = 100

These get divided into

- (2-1)(3-1) for class:test
- (2-1)(3-1)(2-1) for sex:class:test
- (2-1) (3-1)(5-1) for program:class:test
- (2-1)(3-1)(2-1)(5-1) for sex:program:class:test
- Remainder into error (80)

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)    
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *  
sex:program  4   3712     928   0.467   0.7593    
Residuals   40  79422    1986                     

Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)   
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204   
program:class      4    417   104.3   0.490 0.74329   
sex:program:class  4    208    52.1   0.245 0.91127   
Residuals         40   8524   213.1                   

Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)    
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517    
program:test      8   1253     157   0.649    0.734    
sex:program:test  8   2365     296   1.225    0.295    
Residuals        80  19306     241                     

Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4                 

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75               
```

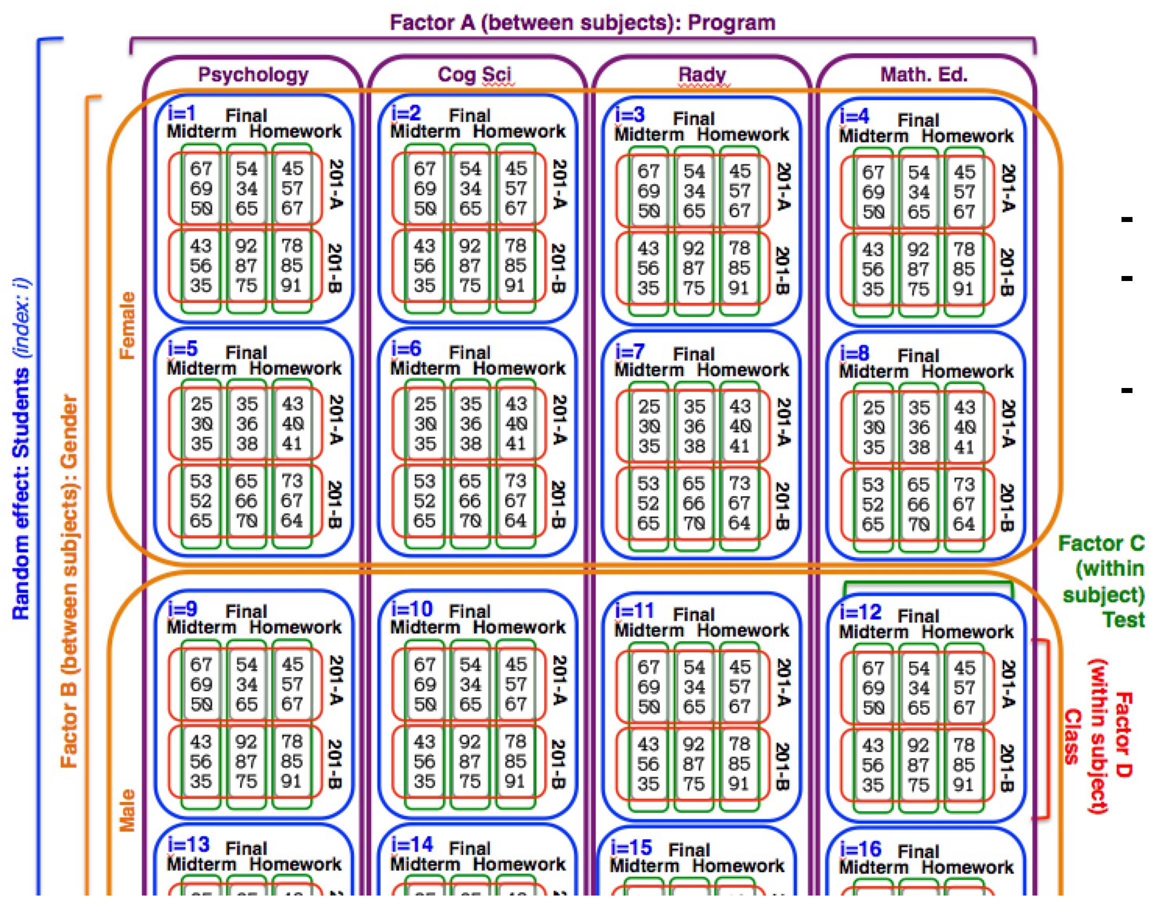
Within-Student:class:test variability (factoring out student:class:test average, what was the variability across problems?)

We have no explanatory variables for this, so it's just residuals.

Mixed design (factors within,between)

We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained variance

```
Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```



'Within' cell variability.

Roughly:

- Subtract subject effect, subject:class interaction, subject:test interaction, and subject:class:test interaction from all data.
- Compute variability of these data.
- This is just the extra measurement variability isolated.
- Since we have multiple measurements per subject-within-subject-cell, this is something we have, but don't use.

Total d.f.:

subjects * (# test levels-1) * (# class levels - 1) * (# reps)

Mixed design (factors within,between)

```
summary( aov(score ~ sex*program*class*test + Error(student/(class*test)) )
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05 ***
program      4  21430    5357   2.698   0.0442 *
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
```

Between-subject error: variation of average performance across students. This is the relevant variability against which to compare between-subject effects.

```
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448 **
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
```

Interaction of students and classes: this is the relevant variability against which to compare all class effects (main effect of class, and class:between-factor interactions.

```
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06 ***
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
```

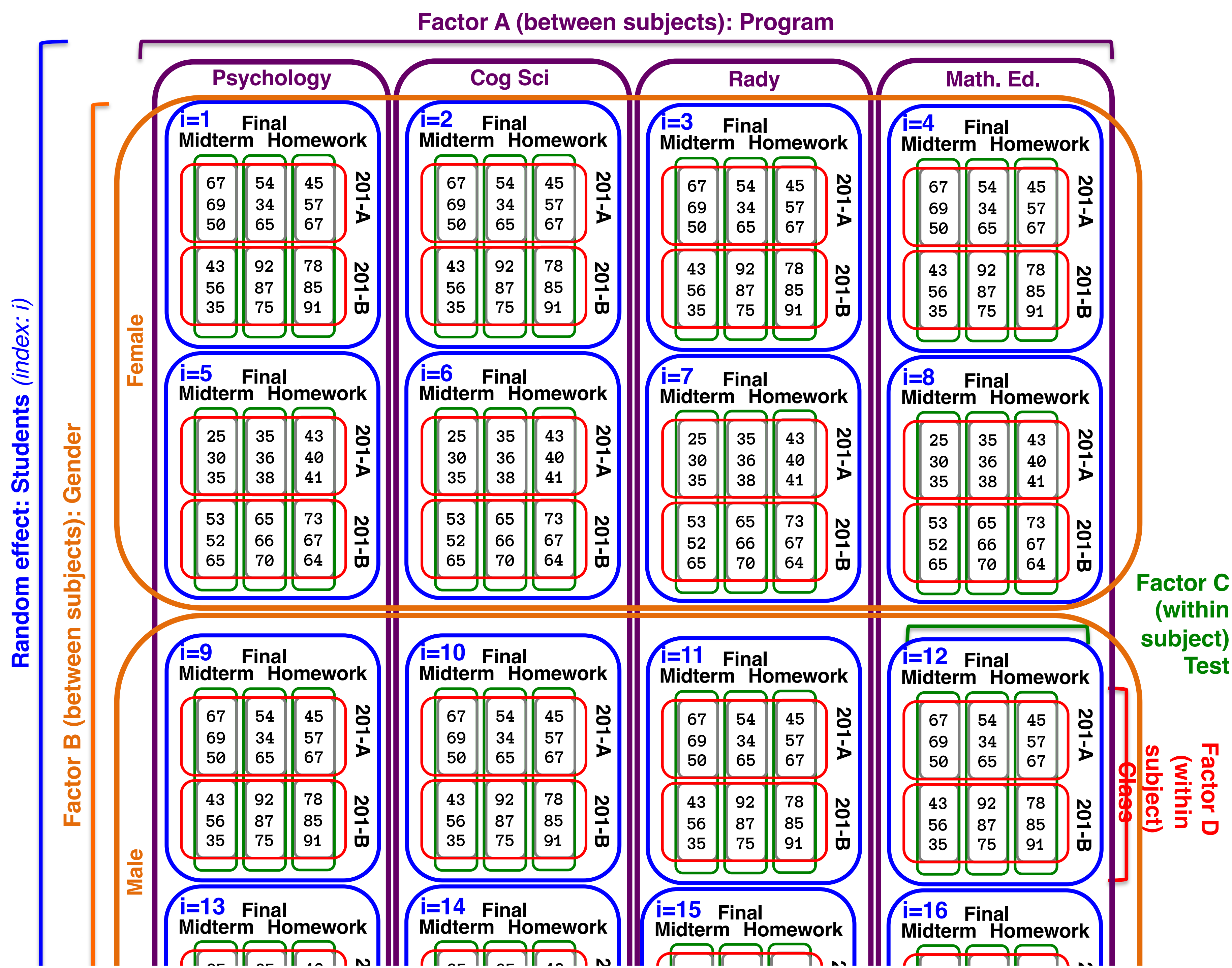
Interaction of students and test: relevant variability against which to compare all test effects (main effect of test, and test:between-factor interactions.

```
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06 ***
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
```

Interaction of students and class-test combination: relevant variability against which to compare all class:test effects

```
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```

Variability across multiple measurements of a student in a class on a test: not relevant.



Factor A (between subjects): Program
Psychology
Cog Sci
Rady
Math. Ed.
Random effect: Students (index: i)
Factor B (between subjects): Gender
Female
Male
Factor C (within subject) Test
Factor D (within subject) Class
i=1 Final Midterm Homework
i=2 Final Midterm Homework
i=3 Final Midterm Homework
i=4 Final Midterm Homework
67 69 50 | 54 34 65 | 45 57 67 201-A
43 56 35 | 92 87 75 | 78 85 91 201-B
i=5 Final Midterm Homework
i=6 Final Midterm Homework
i=7 Final Midterm Homework
i=8 Final Midterm Homework
25 30 35 | 35 36 38 | 43 40 41 201-A
53 52 65 | 65 66 70 | 73 67 64 201-B
i=9 Final Midterm Homework
i=10 Final Midterm Homework
i=11 Final Midterm Homework
i=12 Final Midterm Homework
67 69 50 | 54 34 65 | 45 57 67 201-A
43 56 35 | 92 87 75 | 78 85 91 201-B
i=13 Final Midterm Homework
i=14 Final Midterm Homework
i=15 Final Midterm Homework
i=16 Final Midterm Homework

Maybe helpful?

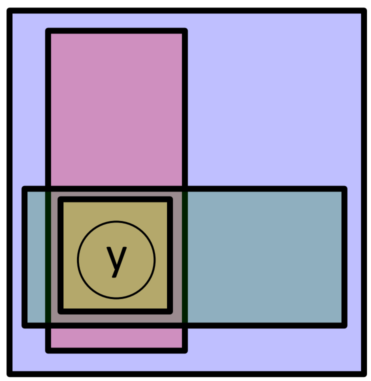
Each data point is nested inside a number of different 'scopes' of variability.

Different students have different additive effects.

Different student:test combinations have different effects.

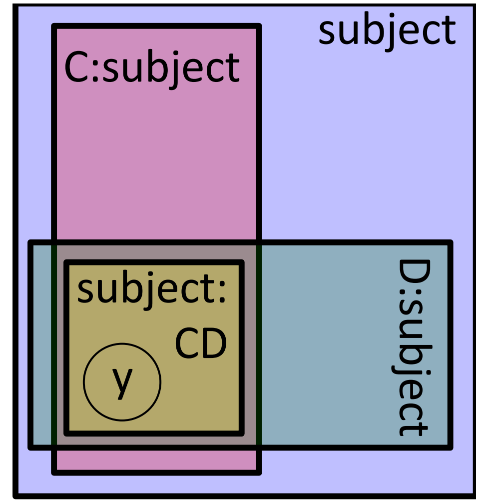
Different student:class combinations differ in effects.

Different student:class:test combinations differ in effects.



Some of each of these variance sources may be explained by various factors. That's what we aim to find out.

Maybe helpful?



Total sum of squares, df.

Between subjects | **Subject:C** | **Subject:D** | **Subject:CD** | **Within Subject:CD**

Then, within each of these, the SS and df are split among explanatory factors and residuals.
Those can be compared via F tests.

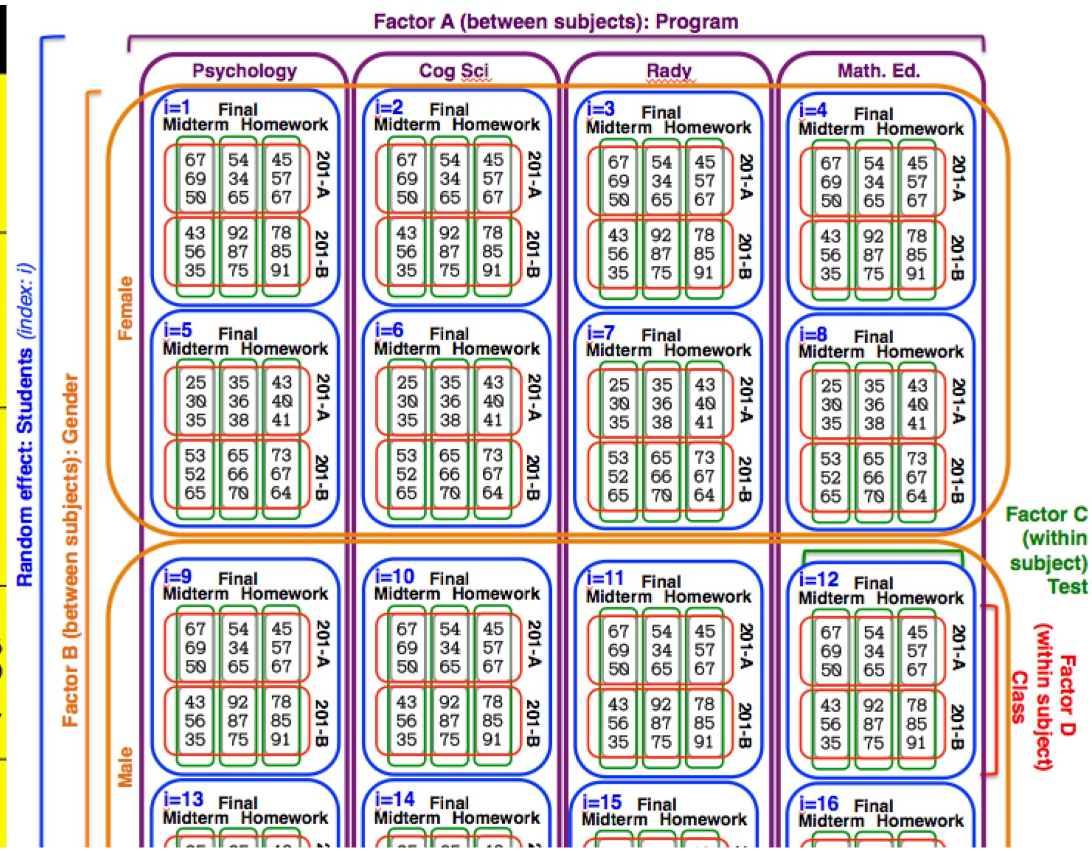
Mixed design (factors within,between)

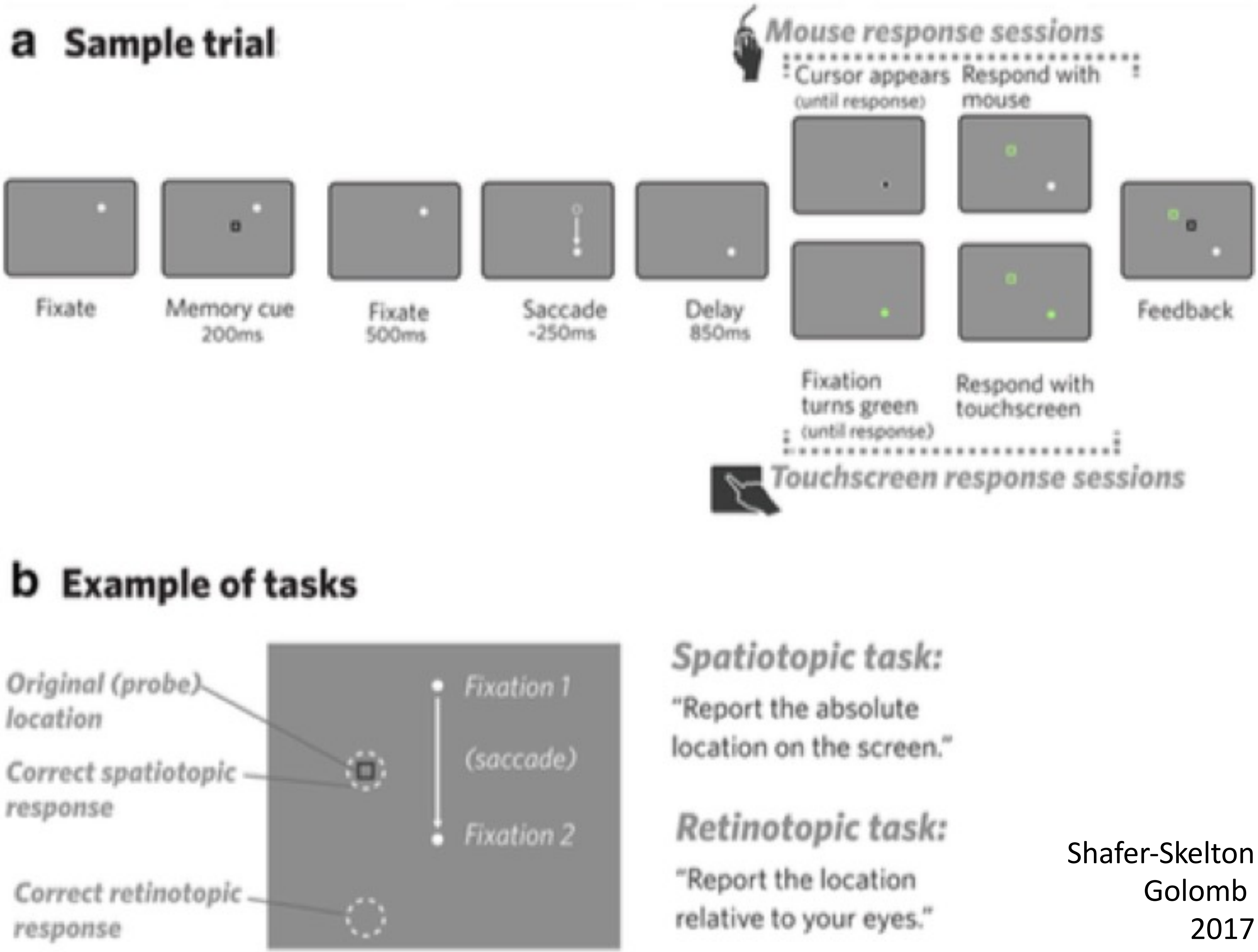
We partition all the variability into different independent sources. Then we partition those sources of variability into explained and unexplained variance

Note: our strategy of pooling by averaging, or really any sum-of-squares strategy, won't work with unbalanced designs. As usual, unbalanced designs give us a credit-assignment problem, and in this case, we get 'leakage' of variability across error strata. Unbalanced designs will thus give us nonsensical ANOVA tables. Beware! Let's avoid those, and ignore this complication until we start using likelihood-based methods later.

```
score ~ sex*program*class*test +
                       Error(student/(class*test))
```

```
Error: student
            Df Sum Sq Mean Sq F value   Pr(>F)
sex          1  40524   40524  20.409 5.43e-05
program      4  21430    5357   2.698   0.0442
sex:program  4   3712     928   0.467   0.7593
Residuals   40  79422    1986
Error: student:class
                  Df Sum Sq Mean Sq F value  Pr(>F)
class              1   1933  1933.2   9.072 0.00448
sex:class          1     81    80.6   0.378 0.54204
program:class      4    417   104.3   0.490 0.74329
sex:program:class  4    208    52.1   0.245 0.91127
Residuals         40   8524   213.1
Error: student:test
                 Df Sum Sq Mean Sq F value   Pr(>F)
test              2   7826    3913  16.213 1.23e-06
sex:test          2    321     160   0.665    0.517
program:test      8   1253     157   0.649    0.734
sex:program:test  8   2365     296   1.225    0.295
Residuals        80  19306     241
Error: student:class:test
                       Df Sum Sq Mean Sq F value   Pr(>F)
class:test              2   6310  3155.1  13.579 8.37e-06
sex:class:test          2    387   193.5   0.833    0.439
program:class:test      8   1342   167.8   0.722    0.671
sex:program:class:test  8   1775   221.9   0.955    0.477
Residuals              80  18588   232.4
Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 300  20625   68.75
```





Shafer-Skelton
Golomb
2017

```
load(url('http://vulstats.ucsd.edu/data/shaferskelton.rdata'))
```

Subject identifier:
- Subject_Initials

Within subject conditions:
- Task
- Mouse/Touchscreen
- saccade_condition

Response variable:
- Error_Dist

Run appropriate aov()
Figure out why it didn't work right
Fix the data
Re-run appropriate aov()
Make a graph.

Correlations from sources of variability

- If we measure the same 'unit' multiple times, those measurements will be correlated. If we treat them as independent samples of the unit's population, we will be very wrong.
- Our task with mixed designs is
  (a) identifying the 'units' being measured at different scales of the analysis.
  (b) Factoring out different independent sources of variability arising from multiple measurements of the same 'unit'.
  (c) Matching up variability of some units, to factors that might explain variability of those units, and then doing an Analysis of variance for each source of variability separately.

Repeated/Mixed ANOVAs

If we have between-subject factors b.A, b.B, b.C, … and within subject factors w.A, w.B, w.C, … we can analyze the data by specifying the model as follows to aov()

```
Y~b.A*b.B*b.C*w.A*w.B*w.C + Error(subject / (w.A*w.B*w.C))
```

Aov() will then split up the overall variability into different independent sources that apply at different scales. (these are sometimes called 'error strata')

And will then do separate ANOVAs for each independent source of variability to figure out how much of that variability can be explained by relevant factors.

If we have unbalanced designs, this process goes awry

What can we not analyze in this way?

- Crossed random effects (subjects and items)
  - Need linear mixed models and different least squares/likelihood calculation.
- Nested/hierarchical designs.
  - Nested ANOVA to partition variance, or hierarchical models.
- Multivariate ANOVA (MANOVA)
  - Relaxes assumptions about residual covariance structure, but loses some degrees of freedom.