Basic question: What should I do when I've measured the same e.u. multiple times?
Subquestion 1: should I assume equal variances at all times?

- I think in terms of the correlation model, but computing is based on the covariance matrix.
- Most repeated measures models allow unequal variances for each time point.
- Different same sort of unequal variance than we've previously seen
  - Where unequal variances depended on the mean response
  - so may vary between treatments
- Modern computing provides lots of options to model unequal variances
  - Transformation is the simplest, often appropriate
  - but often is not always
- Two general strategies when you want to worry about unequal variances,
  - 1) think about what might "cause" inequality. Construct a model to evaluate that cause
  - 2) blindly use the data: fit lots of models and see which fits best
  - best when lots of replicates, so lots of information about variability
  - often misled when only a few replicates, e.g., 3 per treatment/day

Subquestion 2: What if the time intervals are not equal?

What if you measured something on days 1, 2, 3, 5, 7, 14, 21?

- ar(1) assumes correlation between observations, ρ, is same for all adjacent observations
- e.g., Cor(day 1, day 2) = ρ and Cor(day 3, day 5) = ρ and Cor(day 14, day 21) = ρ.
- might be appropriate (if something "slows" down over time)
- but probably not

Correlation models where the correlation between observations depends on the time separation

- The spatial exponential model, not shown in the table of correlation matrices

$$\text{Cor day } i, \text{ day } j = \exp\left(-|j-i|/\gamma\right)$$

  - Available JMP, R, and SAS

  - R: need to use lme() or gls() with correlation=corExp, not in mmrm()
  - Cor depends on the number of days, $j - i$, between two observations
  - Cor day 3, day 5 = Cor day 1, day 3
  - Only positive correlation
  - γ quantifies strength of correlation, $\gamma \geq 0$
  - $\gamma \approx 0$: no correlation for any time lag
  - γ large: strong correlation, declines slowly with time lag
  - When obs are equally spaced, sp exp is same as ar(1), $\rho = \exp(-1/\gamma)$
  - JMP: Spatial power also same as ar(1), often fewer computational issues, $\rho = \gamma$

- The antedependence model, ANTE(1).
  - SAS assumes different variances (more general than you may want)
  - R/mmrm and JMP provide both equal and different variance versions
  - more general than spatial exponential
    * spatial exponential uses one parameter for times 1-2 and 7-14
    * but does account for different length of gap between obs times
    * antedependence uses one parameter per sequential pair of times
    * so correlation time 1 - time 2 can be a completely different value than times 7 - 14

2nd type of repeated measurement: different responses, each measured once
Examples:

- Psychological study: at end of study administer 3 different assessments
- Animal nutrition: at end of study record wt gain, condition score, backfat thickness
- Ecological study: record species richness, biomass, % grass

In each, there are 3 different responses. How to analyze and report?

Raises issues of multiple testing: same concept as multiple comparisons
p-value = P[reject H0 when H0 true], i.e., P[claim an effect when there isn't one].

Common issue:

- if you do 1 test where H0 is true, you're not likely to claim an effect (only 5% of the time)
- if you do 20 tests where H0 is true in all, you'll claim an effect in 1 test, on average

- if you do 200 tests where H0 is true in all, you'll claim to find 10 effects, on average
- If you do 20 tests, you might not want to be excited about that 1 claimed effect
- Called the multiple comparisons problem when comparing groups
- Called the multiple testing problem in general

Multiple options:

- Analyze each response separately.
  - probably the most common outside of clinical and related (e.g. nutrition) work
  - ignores the multiple testing issue.
  - researcher responsibility to interpret results carefully. e.g., when they have 30 tests
  - find only one "significant" result for a response not expected to show an effect
    - is this is a false positive?
    - Or, a suggestion for a completely new line of research?
- Adjust all p-values (and confidence intervals) for multiple testing
  - Bonferroni is most common: if k tests, adjusted p-value is $k \times p$
    Simultaneous 95% Confidence intervals are individual $1 - 0.05/k$ intervals, e.g. 99.5% intervals when $k = 10$ tests
  - There are improved versions of Bonferroni: e.g., Holm
  - and alternatives to Bonferroni: e.g., false discovery rate
- First, use MANOVA, multivariate ANOVA, to test whether any response is significantly different
  - If yes, then test each response separately without adjustment
  - Same concept as Fisher's protected LSD for pairwise comparisons of means
  - But MANOVA requires a much stronger assumption than do tests of any individual response

The facts: Adjusting for multiple testing:

- Reduces the number of false positive results
- Also reduces power, increases CI width
- Penalizes researches for being efficient.
  - Consider two researchers, both primarily interested in response A

  - Researcher 1: measures A and one other response. Declares A has an effect if $p < 0.025$
  - Researcher 2: measures A and 9 other responses. Declares A has an effect if $p < 0.005$

Many opinions! Here are mine

- 1) Declare, before seeing the data, one (perhaps 2) primary outcomes
  - Key is that these are pre-specified (e.g., response A in above scenario)
  - Based on common practice in clinical studies
  - Clinical trial registry (clinicaltrials.gov) includes the primary outcome(s)
  - Test primary outcome without adjustment
  - Test all other outcomes with multiple testing adjustment
- 2) If you're really interested in an unexpected result, run a second experiment
  - was that unexpected result something repeatable (at least once), or a fluke?

3rd type of repeated analysis: subgroup analyses

Example:

- Smoking cessation study, ca 1000 individuals
  - 1st analysis: all subjects, you find no evidence of an effect
  - 2nd analysis: men and women separately
  - 3rd analysis: men 16-25, men 26-70, women 16-25, women 26-70
- Same multiplicity issues as multiple responses
- My opinion
  - 1st analysis is the primary analysis, no adjustment
  - all subsequent subgroup analyses should be adjusted