Adapted from material by Jamison Fargo, PhD

Cohen Chapter 15

Repeated Measures ANOVA

“The biggest job we have is to teach a newly hired employee how to fail intelligently. We have to train him to experiment over and over and to keep on trying and failing until he learns what will work.”

Charles Kettering, American engineer, 1876 - 1958

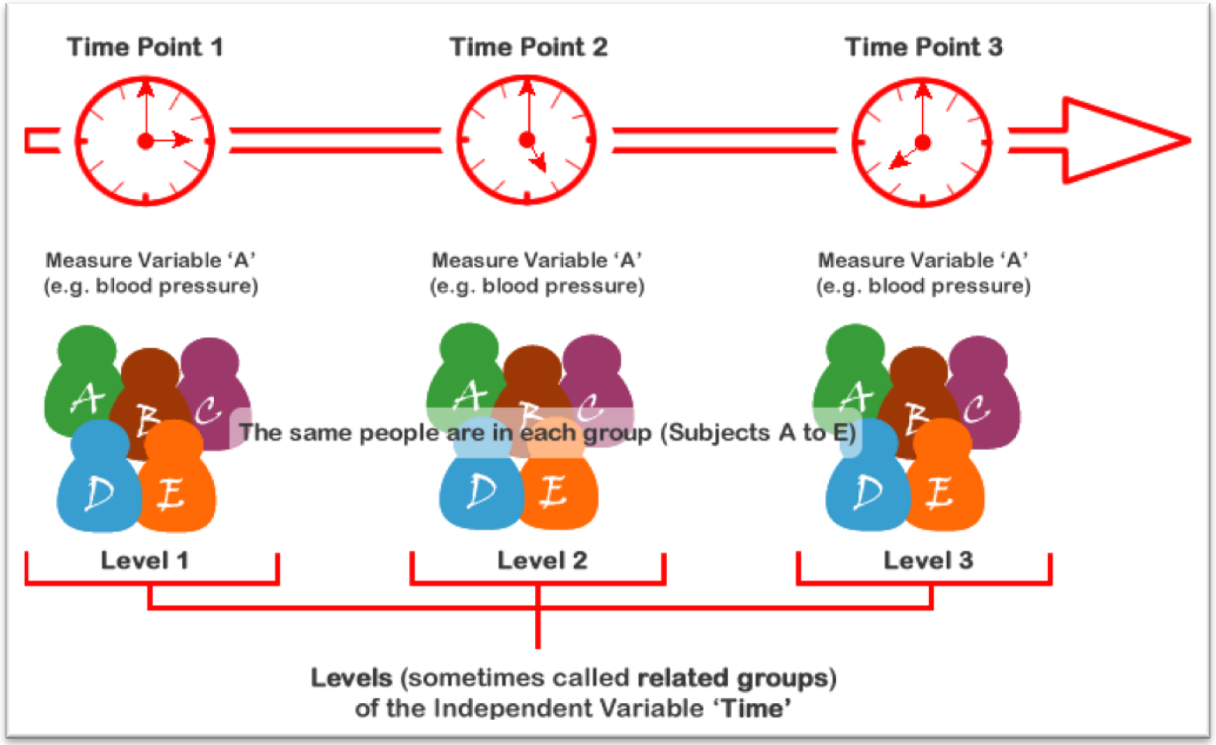
One-Way
Repeated Measures ANOVA

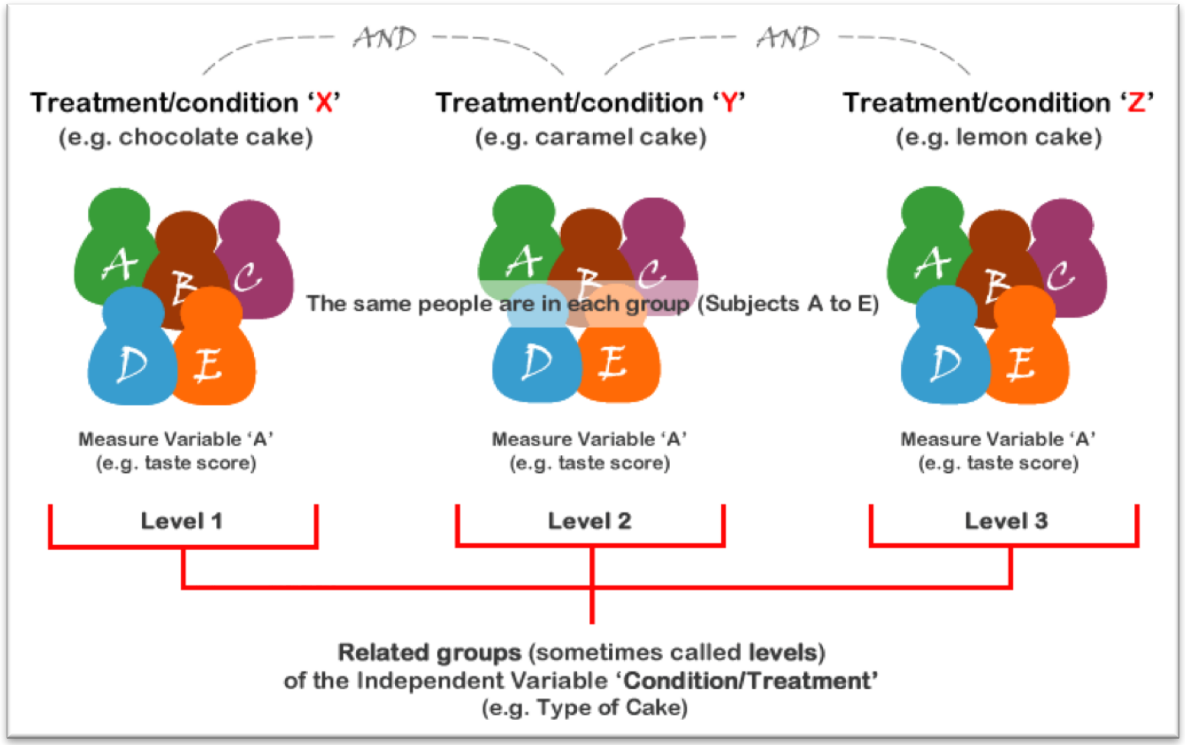
Dr. Pearson is interested in determining whether the average man wants to express his worries to his wife more (or less) the longer they are married. The Desire to Express Worry (DEW) scale is administered to men when they initially get married and then at their 5^{th}, 10^{th}, and 15^{th} wedding anniversaries.

What is the repeated-measures factor and what are its levels?
What is the outcome variable?

Dr. Fairchild wishes to compare reaction time differences for the three subtests of the Stroop Test in patients with Parkinson's Disease: Color, Word, and Color Word.

What is the repeated-measures factor and what are its levels?
What is the outcome variable?


Time Point 1
Time Point 2
Time Point 3
Measure Variable 'A'
(e.g. blood pressure)
Measure Variable 'A'
(e.g. blood pressure)
Measure Variable 'A'
(e.g. blood pressure)
A
B
C
D
E
The same people are in each group (Subjects A to E)
Level 1
Level 2
Level 3
Levels (sometimes called related groups)
of the Independent Variable 'Time'


AND
AND
Treatment/condition 'X'
(e.g. chocolate cake)
Treatment/condition 'Y'
(e.g. caramel cake)
Treatment/condition 'Z'
(e.g. lemon cake)
A
B
C
D
E
The same people are in each group (Subjects A to E)
Measure Variable 'A'
(e.g. taste score)
Measure Variable 'A'
(e.g. taste score)
Measure Variable 'A'
(e.g. taste score)
Level 1
Level 2
Level 3
Related groups (sometimes called levels)
of the Independent Variable 'Condition/Treatment'
(e.g. Type of Cake)

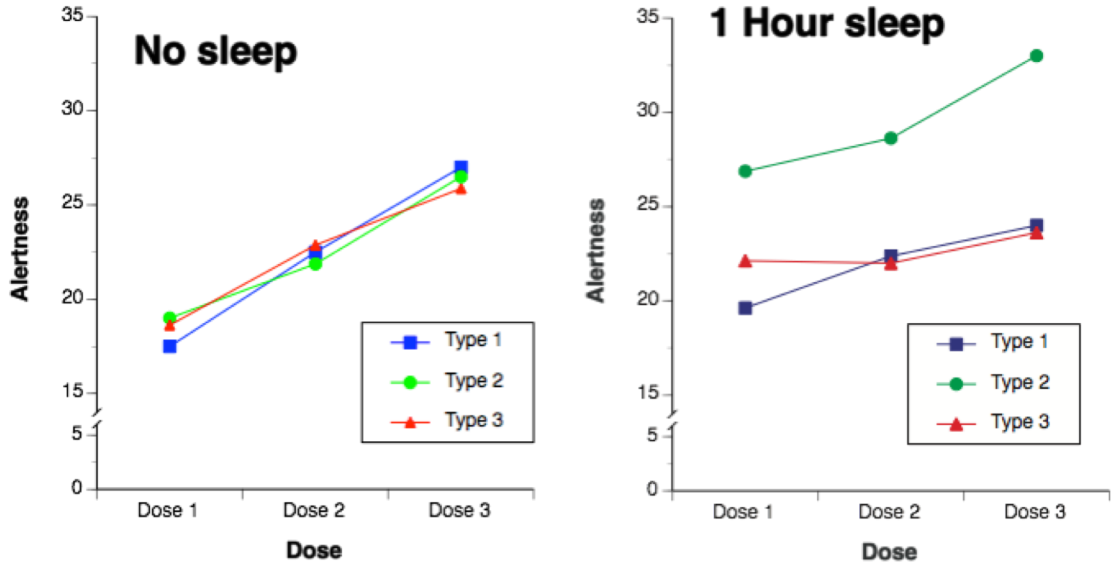
Design Types

1. Same outcome, same cases, **different occasions**
   Time points are levels of factor

2. **Different outcomes** (all on same metric) on same cases
   Different **outcomes** are levels of factor

3. Same outcome, different condition/exposure, on cases that are **matched into sets** prior to random assignment
   Different **conditions** are levels of factor

- Experimental
- Quasi-experimental
- Field/Naturalistic studies
- Longitudinal/Developmental studies



More powerful:

- Each case serves as their **own control**, less between-subject variation
- Error term (denominator) of *F*-test for RM ANOVA is often **less** than in Independent Groups ANOVA

More economical:

- **Fewer cases** required
- Independent Groups ANOVA:
  - 3 conditions,
  - 10 cases per condition
  - = 30 cases
- RM ANOVA:
  - 3 conditions,
  - same 10 cases used in all conditions
  - = 10 cases

Repeated-Measures (RM) factor often referred to as:

'Within-Subjects' factor

- Time 1, Time 2, Time 3, etc…
- Condition1, Condition2, Condition3, etc…

May have…

- Multiple RM factors → Factorial RM ANOVA
- A combination of RM and independent groups factors → Mixed Design ANOVA

- *Lack of independence of observations → must be accounted for in analysis*

Time as a RM Factor

Can answer questions such as:

Do measurements on outcome change over time or conditions?
Is change linear? Quadratic?
Is change positive or negative?
Does change 1st increase, then decrease (or vice versa)?
How long does change last?
Is change permanent over duration of study?
Is outcome same at beginning and end of study?

- Researcher chooses when and how frequently to observe outcome, time is not traditionally considered experimental variable
  - Not a manipulated factor, cannot counterbalance time, or randomize participants to have different times or orders of observation
  - Although many experiments are longitudinal, they include an additional treatment variable that is experimentally manipulated
- Time intervals must be equally spaced
  - If spacing is unequal, ANOVA with random-effects must be used instead

Time as a RM Factor

	Month			
	Month 1	**Month 2**	**Month 3**	**Row Means**
s1	1	3	6	3.33
s2	1	4	8	4.33
s3	3	3	6	4.00
s4	5	5	7	5.67
s5	2	4	5	3.67
Column Means	2.40	3.80	6.40	**4.20**

	Treatment			
	A1	***A2***	***A3***	**Row Means**
	s1	*s1*	*s1*	.
	s2	*s2*	*s2*	.
	s3	*s3*	*s3*	.
	s4	*s4*	*s4*	
	s5	s5	s5	.
Column Means	.	.	.	**GM**

Condition as the RM Factor

Simultaneous RM Factors

- Sometimes levels of RM factors are administered:

simultaneously or inter-mixed

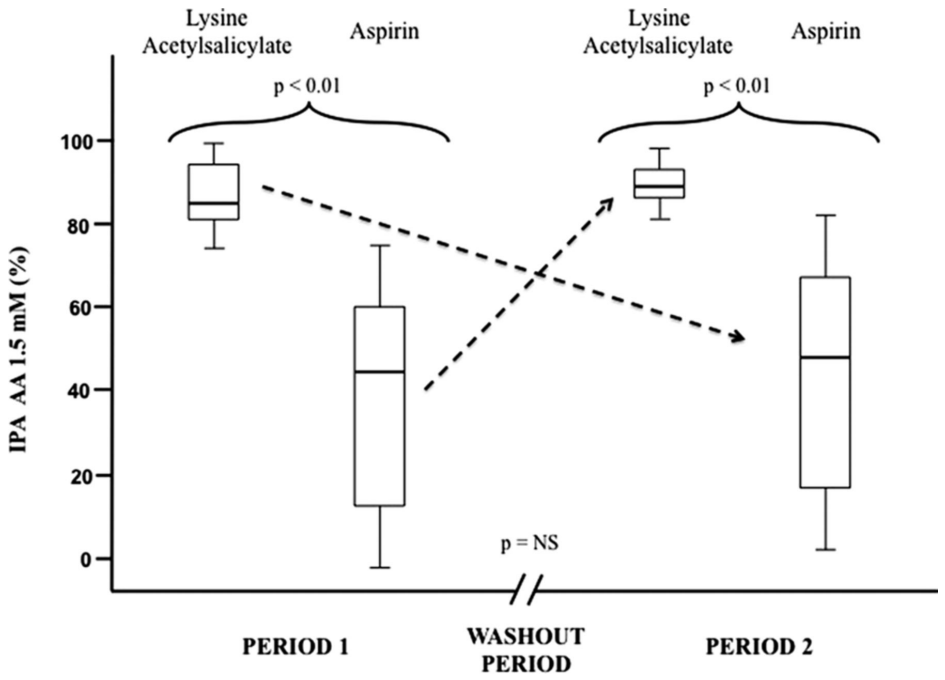
within one experimental or observational study

For example…

- Levels of RM factor might be verbs, nouns, and adjectives, which appear randomly within a passage to be memorized
- # of words of each type recalled by participants are recorded

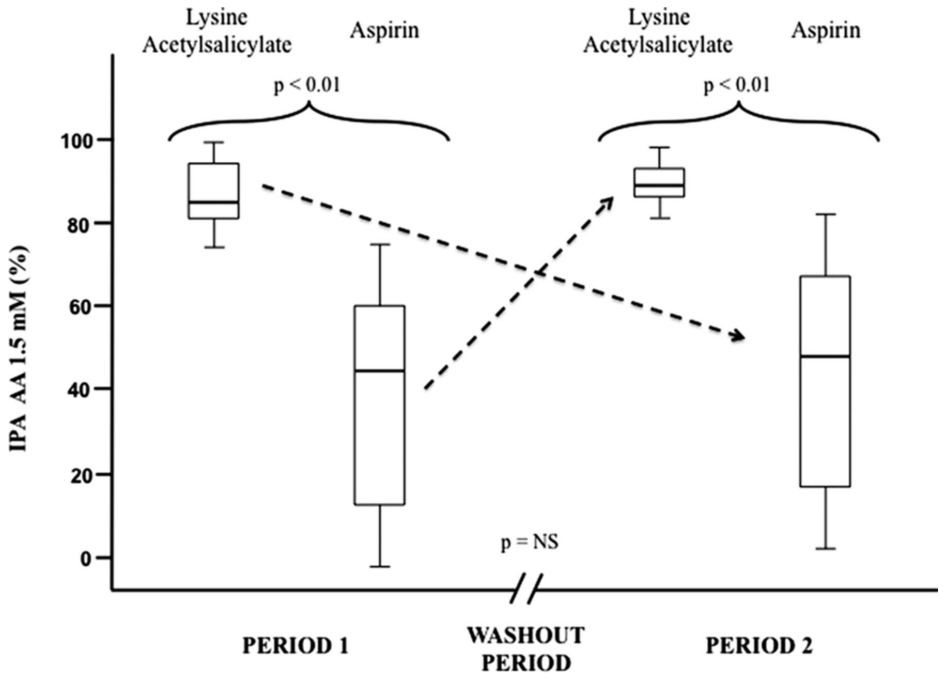
Carryover Effects: *The Problem…*

- Exposure to treatment or participation in study/outcome at one time influences responses at another
  - Biases related to practice, fatigue, etc.
- *When time is RM factor, carryover effects are the focus of study*
  - *Learning, change over time*
- When CONDITION is RM factor and participants rotate through conditions, carryover effects are not of interest and may lead to spurious results
  - Magnitude of carryover effects will vary across treatment order
  - Differential carryover effects are very problematic
    - Effect of some levels of RM factor are more long-lasting than others



Carryover Effects: *Possible Solutions*

- Counterbalancing: Varying RM condition order across subjects
  - 3-level RM factor: ABC, ACB, BCA, BAC, CAB, CBA
- Partial counterbalancing (Latin Squares): Too many possible orders of RM conditions so a representative set is used
- Each subject receives a random order of RM conditions
- Each subject receives a 'run-in' period (a series of practice trials) at beginning of study to 'stabilize' performance
- Intervening (distractor, neutral) trials between conditions
- Larger time interval, washout period, between conditions
- *Note: Effects may not be eliminated by any of these methods*

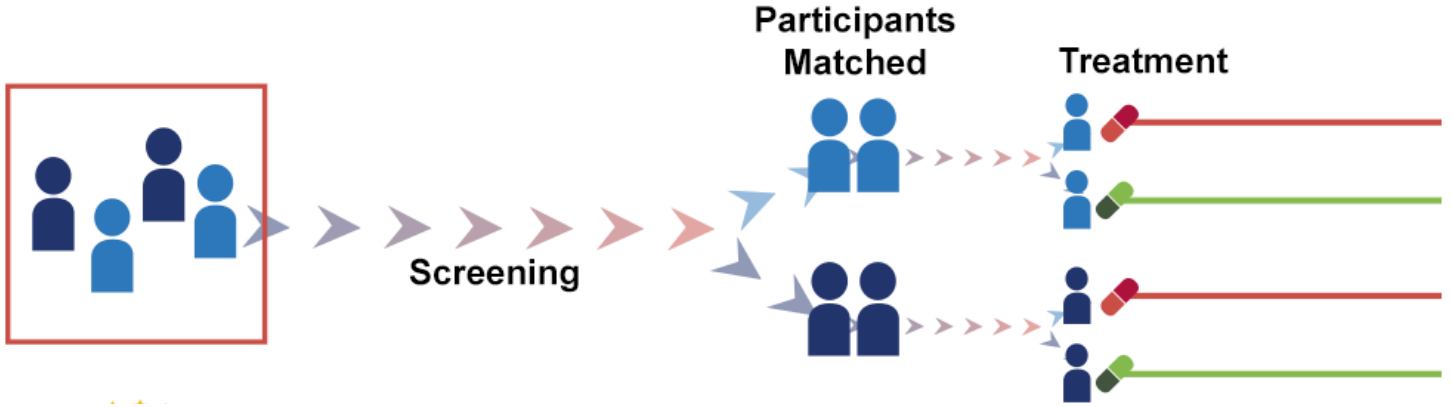


Matched Designs

- Alternative to having same cases engage in all RM conditions
  - Used to limit problems associated with…
    - Confounding variables (e.g., age, sex, education)
    - Other threats to internal validity associated with RM studies, such as carryover effects or ordering

- Each member of a **set** of unique, but similar or matched, participants is **randomly assigned** to one condition

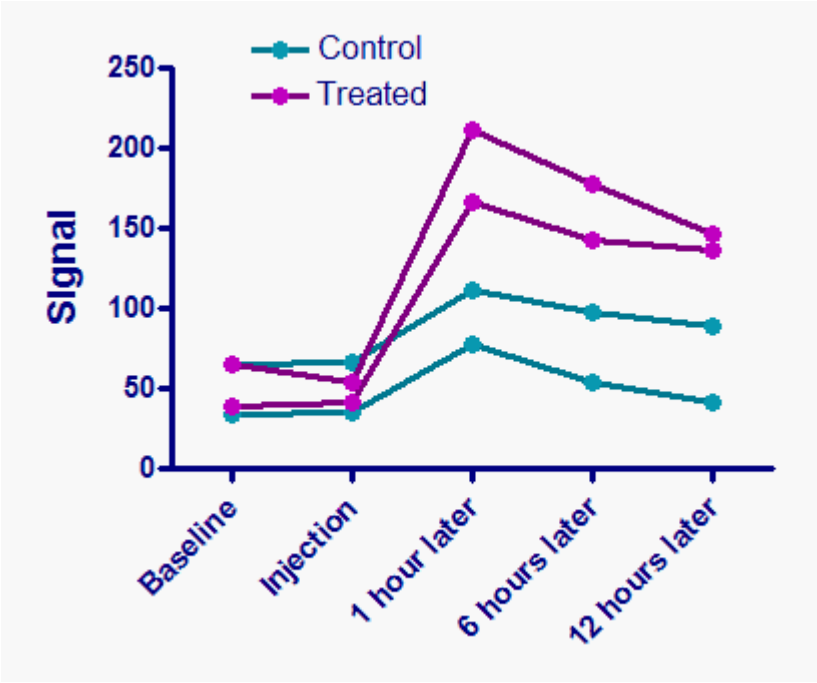
- In analysis, each **set of participants** treated **as if** they are the **same** participant
- Participants matched into sets on **potentially confounding variables** (e.g., pretest scores, other characteristics) prior to random assignment
  - *Researcher may have too much faith in matching*
  - *Need to report on process used for matching*
  - *Usually only match (if at all) on 1 or 2 variables*



May match and conduct 1-Way Independent Groups ANOVA to be more conservative in statistical results

1-Way RM ANOVA
is actually a
2-Way Independent Groups ANOVA
in disguise!!

- Factor 1: RM or Within-Subjects factor: Time, Condition
- Factor 2: Subject factor: 8 participants = 8 levels



Hypothesis:

- Only made with respect to marginal means of RM factor
- Same form as 1-Way Independent Groups ANOVA
  - H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
  - H_1: *H_o is not true*

Partitioning Variance

- RM factor: Same or similar outcome is measured more than once (each level) by multiple participants
- Subject factor: Same or similar outcome is measured more than once (each level) by same participants or sets of matched participants
- RM x Subject factor interaction

Total variation partitioned into 3 parts…but no SS_W or error term!

$$SS_{Total} = SS_{RM} + SS_{Subj} + SS_{RMxSubj}$$

Note: only 1 score per cell (n = 1) in previous 1-Way RM ANOVA cross-classification, thus, no variability within cells; $SS_W = 0$

- $SS_{RMxSubj}$ is used as error term and represents variation in outcome explained by…
    1. Interaction of participants with levels of RM factor
    2. Random (i.e., left-over) variation (error)

$SS_{Repeated\ Measure}$

In computing column or marginal means of RM factor all scores in a given level are averaged regardless of row

- n_k = # participants per RM level

$$SS_{RM} = n_k[(\bar{X}_{RM1} - \bar{X}_{GM})^2 + (\bar{X}_{RM2} - \bar{X}_{GM})^2 + \ldots + (\bar{X}_{RMk} - \bar{X}_{GM})^2]$$

$$SS_{RM} = \frac{\left(\sum_{i=1}^{n} X_{RM1}\right)^2 + \left(\sum_{i=1}^{n} X_{RM2}\right)^2 + \ldots + \left(\sum_{i=1}^{n} X_{RMk}\right)^2}{n_k} - \frac{\left(\sum_{i=1}^{n} X\right)^2}{N}$$

$SS_{Subject}$

- In computing individual subject means, all scores in a given row are averaged, regardless of level of RM factor
  - n_{row} = # repeated measurements of outcome from same participant, since n = 1 per cell

$$SS_{Subj} = n_{row}[(\bar{X}_{Subject1} - \bar{X}_{GM})^2 + (\bar{X}_{Subj2} - \bar{X}_{GM})^2 + \ldots + (\bar{X}_{N} - \bar{X}_{GM})^2]$$

$$SS_{Subj} = \frac{\left(\sum_{i=1}^{n} X_{Subj1}\right)^2 + \left(\sum_{i=1}^{n} X_{Subj2}\right)^2 + \ldots + \left(\sum_{i=1}^{n} X_{N}\right)^2}{n_{row}} - \frac{\left(\sum_{i=1}^{n} X\right)^2}{N}$$

$SS_{interaction}$

- Variability among cell means when variability due to individual Subject and RM effects have been **removed**

$$SS_{RMxS} = [(\bar{X}_{cell11} - \bar{X}_{GM})^2 + (\bar{X}_{cell12} - \bar{X}_{GM})^2 + \ldots + (\bar{X}_{cell\ rc} - \bar{X}_{GM})^2] - SS_{RM} - SS_{Subj}$$

$$SS_{RMxS} = \left(\sum_{i=1}^{n} X_{cell11}\right)^2 + \left(\sum_{i=1}^{n} X_{cell12}\right)^2 + \ldots + \left(\sum_{i=1}^{n} X_{cell\ rc}\right)^2 - \frac{\left(\sum_{i=1}^{n} X\right)^2}{N} - SS_{RM} - SS_{Subj}$$

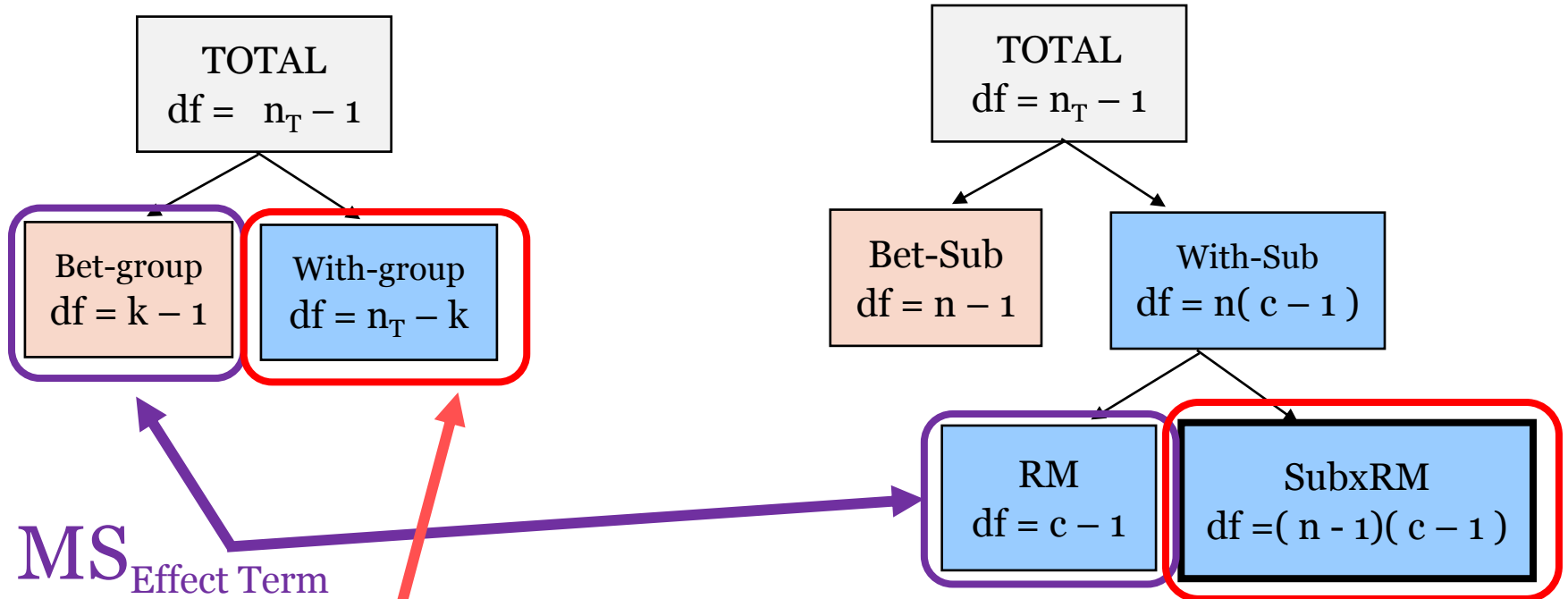
SS & DEGREE OF FREEDOM

Independent Groups ANOVA

$SS_{Total} = SS_{Row} + SS_{Within}$

Repeated Measures ANOVA

$SS_{Total} = SS_{RM} + SS_{Subj} + SS_{RMxS}$



$$F = \frac{MS_{Effect\ Term}}{MS_{Error\ Term}}$$

$$MS_{Subj} = SS_{Subj} \ / \ df_{Subj}$$

- Generally **ignored**, considered nuisance variable

- However, may be of interest to know if participants vary significantly on outcome:
  - Considered 'random effect'
    - assumed participants (which serve as levels) are a random sample

  - Correct analysis is random- or mixed-effects ANOVA
    - Mixed-effects ANOVA: Includes both fixed and random effects (which can either be independent or repeated)
    - Mixed-design ANOVA: Includes both independent (between-subjects) and repeated-measures (within-subjects) factors

$$MS_{RM*S} = SS_{RM*S} \; / \, df_{RM*S}$$

$$SS_{Within} = SS_{Subj} + SS_{RMxS}$$

- Not always of inferential interest
- Useful for **testing assumptions** (later)
- Indicates whether RM effect is **similar for all participants**
  - When MS_{RMxS} = **0**, effect of RM factor is consistent across participants → **desirable**
  - When MS_{RMxS} is **large**, effect of RM factor likely differs across participants → **undesirable**
  - **Line plot** of individual participant means across conditions/time can shed light

- Variation due to participants (MS_{Subj}) is not included in error term for *F*-test of RM factor, MS_{RMxS}
- Thus, error term is generally smaller in RM ANOVA than Independent Groups ANOVA

- However, when matching leads to no variation across subjects ($SS_{Subj} \approx 0$) and $MS_{RMxS} = MS_{Within}$
  - Results of RM ANOVA same as Independent Groups ANOVA
  - Increased effect of matching or repeating participants
    - SS_{RMxS} decreases, SS_{Subj} increases
  - Decreased effect of matching or repeating participants
    - SS_{RMxS} increases, SS_{Subj} decreases

1-Way RM ANOVA: Summary Table

Source	*SS*	*df*	*MS*	*F*	*p*
RM					
Subj			**X**	**X**	**X**
Error(RM x Subj)				**X**	**X**
Total			**X**	**X**	**X**

Assumptions

- Participants are a **random sample** from population and are **independent** of one another (*Although participant observations are dependent, participants themselves are independent)*
- DV **normally** distributed in the population

  Less concerned: equal *n* per level and $df_{Intrx} \approx 20$ (CLT) ← investigate via plotting
- **Homogeneity** of variance

  Variance of DV is similar for all levels of RM factor ← Leven's or visual inspection
- If *Time* is RM factor, data are measured at (near) **equal intervals**

- ****Sphericity**** and **Compound symmetry**

  CS is a special case of sphericity
  - If CS is satisfied, sphericity is satisfied
  - However, if CS is not satisfied, sphericity may still be satisfied

Sphericity

- Informally, it is the degree of violation of **independence same** for all levels of RM factor?
- Taking DV, difference scores can be calculated for each participant between all possible pairs of levels of RM factor
  - *A variance can be calculated for each set of difference scores*
  - *When assumption of sphericity is met, difference score variances will be equal*
- **Mauchly's test of sphericity**
  - Based on χ^2 distribution
  - H_o: Variances of difference scores between all pairs of levels of RM factor are equal (sphericity)
  - Test not extremely useful as most "tests of other tests" tend to be…misleading*
    - Small N = ↑ Type II error
    - Large N, non-normality, +heterogeneity of covariances = ↑ Type I error
- When using this test, assess all RM main effect(s)

- **Rule of thumb:** cause for concern may exist when the **largest variance is 4x greater than smallest**

*Kesselman, Rogan, Mendoza, & Breen, 1980

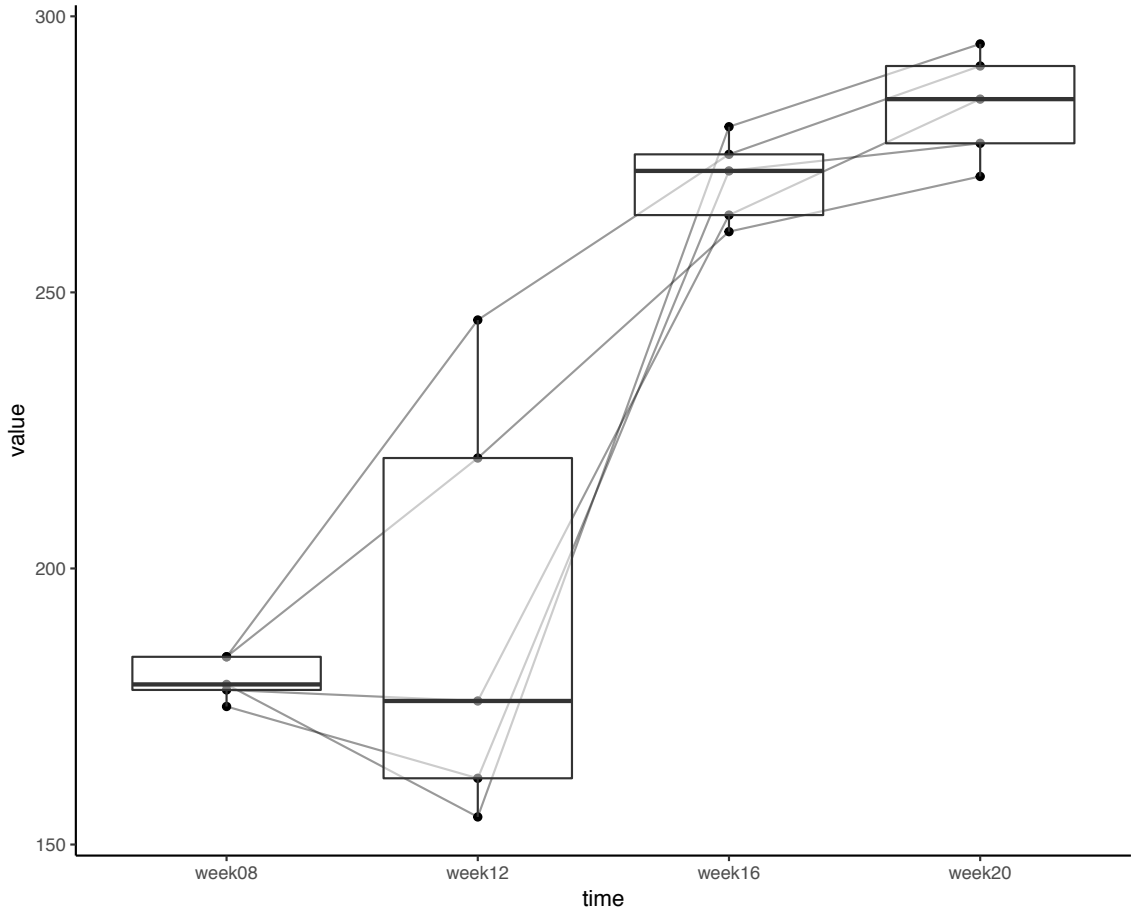
Sphericity: Mauchly's test

Only applies to RM factors with > 2 levels

- Cannot compare variances of difference scores when there is only 1 set of differences
- Sphericity always met when $k = 2$ (RM factor)

When violated, ↑ risk of Type I error

- Critical *F*-statistics will be too small
- *F*-test is + *biased* when sphericity is violated
- Several "alternatives", discussed later



Compound Symmetry

A bit stricter than sphericity, which is a special case, and is subsumed by CS

- ❑ Homogeneity of **variances** of difference scores
  - Variance of difference scores assumed to be equal
  - Same as previously mentioned for sphericity
- ❑ Homogeneity of **covariances** of difference scores
  - Covariances of difference scores (between all possible pairs of levels of the RM factor) assumed to be equal
  - Most software does not assess this assumption
- ❑ **Additivity** (discussed in later slides)

Independence

	A	B	C	D
A	s_A^2	0	0	0
B	0	s_B^2	0	0
C	0	0	s_C^2	0
D	0	0	0	s_D^2

Groups or levels are independent of one another as there are different participants in each level; variances are non-0 and assumed equal, covariances are 0

Compound Symmetry

	A	B	C	D
A	s_A^2	s_{AB}	s_{AC}	s_{AD}
B	s_{BA}	s_B^2	s_{BC}	s_{AB}
C	s_{CA}	s_{CB}	s_C^2	s_{AC}
D	s_{DA}	s_{DB}	s_{DC}	s_D^2

Groups or levels are dependent or correlated. Variances are non-0 and assumed equal as are covariances (assumption met)

Additivity

- **Error term for RM ANOVA is *RMxS* interaction**
  - Should only represent random error, not error plus variation of subjects over time or across conditions
  - Possible that effect of level A of RM factor is different for different subjects, and thus an interaction between RM and S truly exists
  - Then, some of what we consider to be error when we calculate RMxS, is really an interaction effect, and not just random error
- **Thus, Additivity = absence of RMxS interaction**
  - Presence of such an interaction indicates a multiplicative or nonadditive effect where different participants have different patterns of response to RM factor
  - Error term is thus distorted by inclusion of a systematic (non-random) source of variation (due to *Subjects*)
  - Must determine what extraneous (between-subjects) factor (e.g., Gender) is causing interaction and test it explicitly (e.g., Gender X RM Factor interaction)
  - Inclusion removes effects from error term (MS_{Intrx}) -> **Mixed-Design ANOVA** (*discussed next lecture*)
- Since nonadditivity implies heterogeneous variances for difference scores, sphericity assumption will be violated if this assumption is not met
- A test exists for this assumption, called the "Tukey test for nonadditivity", available in `additivityTests::tukey.test()`

Assessing Assumptions

If we want to assess these assumptions, we rely on results of the following approaches in practice:

- Homogeneity of variances
  - **Levene's** (or Bartlett's) test
- Sphericity/Compound Symmetry
  - **Mauchly test**
  - Examination of variance-covariance matrix
  - Examination of variances among pairs of difference scores
- Additivity
  - Small MS_{Intrx}
  - Individual Subject lines in a means **plot are mostly parallel**

Violations of Assumptions

Mostly concerned with **sphericity** -- > If violated, should pursue some alternative

- If sphericity is met, 5 options:
  - *Use **standard univariate F-tests (recommended)***
  - *Use **trend analysis** (recommended, **IF** this is the goal)*
  - *Use a multivariate test (not recommended as findings should be same as standard univariate F-tests)*
  - *Use a **maximum likelihood** procedure (highly recommended)*
  - *Use a nonparametric test (not recommended, less power)*
    - *Friedman test (1-way only)*

- If sphericity is NOT met, 5 options:
  - *Use an **adjusted or alternative F-test** (recommended)*
  - *Use **trend analysis** (recommended, if this is the goal)*
  - *Use a multivariate test (less recommended in most cases)*
  - *Use a **maximum likelihood procedure** (highly recommended)*
  - *Use a **nonparametric test** (recommended, as a last resort)*
    Friedman test (1-way only)

Alternatives

- Standard univariate *F*-tests are not recommended when sphericity is violated
  - As mentioned before, will be too liberal and inaccurate (increased risk for Type I error)

Trend analysis

- Sphericity assumption irrelevant
- Series of smaller pairwise comparisons across levels of the RM factor
- Preferred for questions regarding the shape of the pattern in the DV over time

Adjusted or alternative univariate *F*-tests *(Useful for **"smaller" N**)*

- *DEGREES OF FREEDOM* (*numerator and denominator*) are REDUCED by multiplying by EPSILON
  - Epsilon = an adjustment factor describing the magnitude of the departure from sphericity
  - If sphericity assumption is perfectly met, epsilon = 1
  - Epsilon < 1 indicates departure from sphericity
    - Lower-bound depends on *k* levels of RM factor
      - 1 / (*k* – 1), thus when *k* = 3, epsilon can be as small as .50

- MORE conservative *F*-critical value
  - *df* correction approaches have been **criticized as too conservative**,
  - increasing risk of Type II error, as they assume maximal heterogeneity among cells

Several approaches (most-to-least conservative)

- Lower-bound: Uses the lower bound estimate of epsilon in the *df* correction
- **Greenhouse-Geisser: Considered conservative and tends to underestimate epsilon when epsilon is close to 1 (danger for over-correction)**
- Huynh-Feldt: Considered less conservative when true value of epsilon is ≥ .75; but also overestimates sphericity

Multivariate *F*-tests

- DV is treated as a set of variables, ignores (does not assume) sphericity;
- Assumes general covariance structure
- Cost: Less powerful than RM ANOVA and should be avoided UNLESS…
  - k is low (< 5) and N is > (15 + k) (or k is high (5 to 8) and N is > (30 + k)) , epsilon is low (< .70), and correlations among levels of RM factor are high
- Computed on differences among means
- Most often used in context of **non-experimental research**
- Different forms exist:
  - Pillai's trace, +Wilk's λ, Hotelling's trace, Roy's largest root
    - +Preferred and most commonly used
    - All yield same result for 1-Way RM ANOVA

- **Additional assumptions** for multivariate *F*-tests
  - Difference scores are multivariately normally distributed in population
    - Difference scores on outcome for each pair of levels are normally distributed
    - Difference scores on outcome for each pair of levels are normally distributed at every combination of the values of other factors
  - Difference scores from any one participant are independent from those of any other participant

- Use multivariate η^2 for main effect or interaction when using multivariate F-tests
  - Multivariate η^2 = 1 – Wilk's Lambda (Λ)

Maximum likelihood procedures

- Mixed-effects, multilevel, or hierarchical linear models
  - Wave of the (present and) future
  - Structure of **variance-covariance matrix** is modeled explicitly
  - not assumed to follow compound symmetry (can be tested empirically)
    - Autoregressive, exchangeable, or unstructured correlational structures are but a few examples

Effect of N on results of the Mauchly test of sphericity

- Could have large N, reject H_0, apply corrections, which are only minimal and unlikely to affect outcome of results
- Could have small N, fail to reject H_0, not apply corrections and obtain spurious results
- If epsilon is near 1, a correction is probably not necessary; however, if epsilon is near the lower bound, a correction is likely necessary
  - Could run both RM ANOVA (with corrections for sphericity) and Multivariate analyses and report analysis that is statistically significant as that analysis has the greater power given the circumstances

Effect Size: η^2

- Little evidence for a RM factor X Subject interaction (additivity met) (Keppel & Wickens, 2004)

$$\text{Partial } \eta^2 = \frac{SS_{RM}}{SS_{RM} + SS_{Intrx}}$$

- Evidence for a RM factor X Subject interaction (non-additivity) (Myers & Well, 1991)
  - Conservative or 'lower bound' estimate

$$\eta^2 = \frac{SS_{RM}}{SS_{RM} + SS_{Subj} + SS_{Intrx}} = \frac{SS_{RM}}{SS_{Total}}$$

Effect Size: ω^2

- Little evidence for a RM factor X Subject interaction

$$\text{Partial } \omega^2 = \frac{df_{RM}(MS_{RM} - MS_{Intrx})}{df_{RM}(MS_{RM} - MS_{Intrx}) + k_{RM}N(MS_{Intrx})}$$

- Evidence for a RM factor X Subject interaction
  - Conservative or 'lower bound' estimate

$$\omega^2 = \frac{df_{RM}(MS_{RM} - MS_{Intrx})}{df_{RM}(MS_{RM} - MS_{Intrx}) + k_{RM}N(MS_{Intrx}) + N(MS_{Subj})}$$

In both equations, N = # independent participants or sets of participants

FACTORIAL

REPEATED MEASURES ANOVA

Dr. Evans wishes to evaluate various coping strategies for pain.

He obtains 8 volunteers to come to the lab on 2 consecutive days. On both days, the volunteers plunge their hands into freezing cold water for 90 seconds.

They rate how painful the experience is on a scale from 1 to 50 (not painful) after 30 seconds, then 60 seconds, and then 90 seconds.

On one day they are given pain avoidance instructions and on the other day they are given concentration on pain instructions.

In order to counterbalance the design, 4 students are given the avoidance and 4 students are given the concentration strategy the 1st day, then switched the 2nd day.

What are the RM factors? What are their levels?
What is the outcome variable?

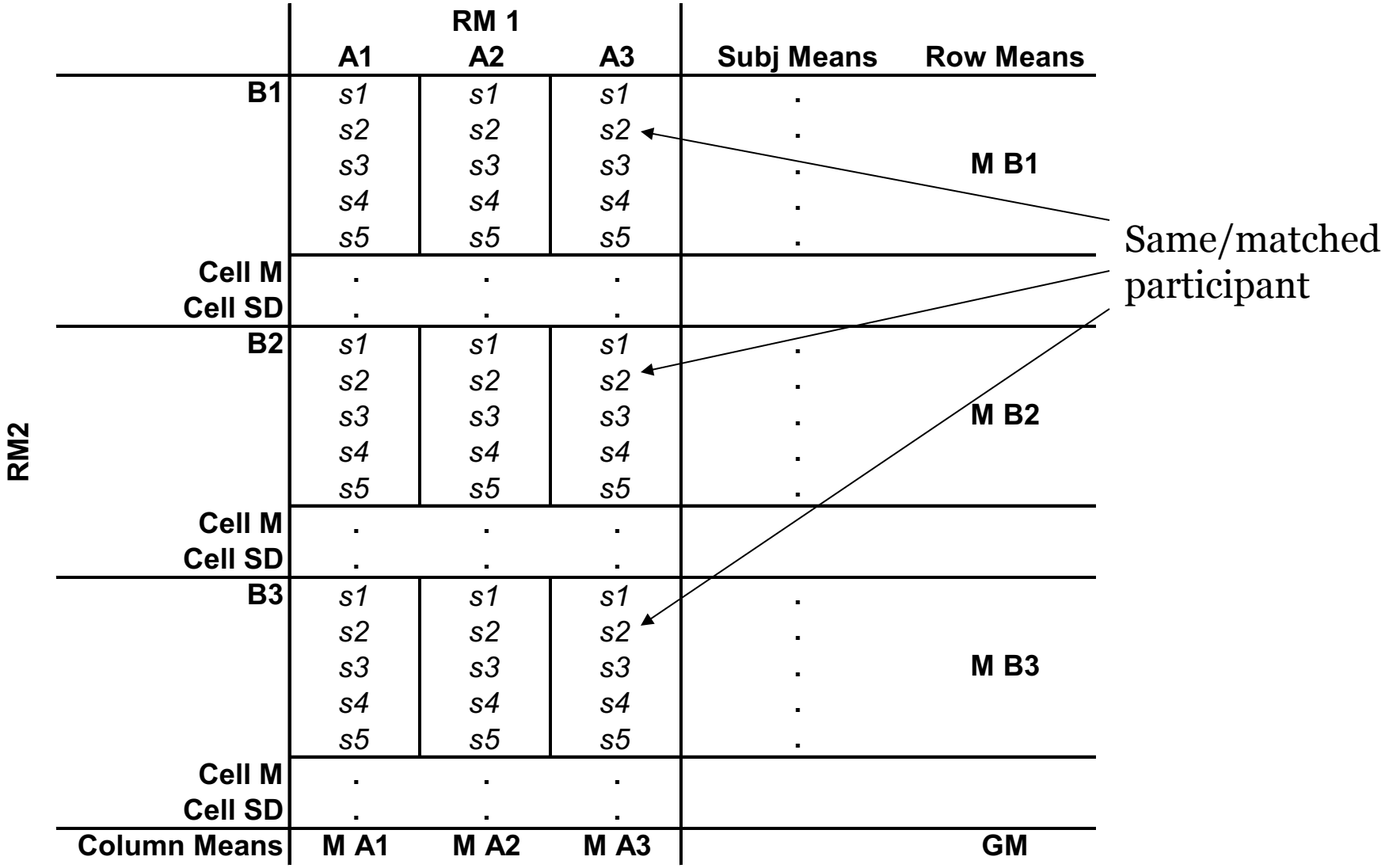
Generally, 'Order' would be another factor (not RM) that would need to be included in the ANOVA. For our purposes, we will say that this factor had no effect.

Dr. Chapman wishes to examine the effect of drugs A and B as well as their interaction on blood flow. Each drug has two possible formulations (levels). Each participant received each of the 4 possible combinations of the 2 drugs over several days (A1B1, A1B2, A2B1, A2B2). The half-life of each drug was such that there were no carry-over effects.

What are the RM factors? What are their levels?

What is the outcome variable?

Factorial RM ANOVA



Factorial RM ANOVA

2 or more RM factors (no independent factors)

Separate error term for each RM main effect and for interaction(s) among RM factors

Error terms = RM effect being tested (main effect or interaction) x Subjects interaction

- 1st RM main effect error term = RM_1 x Subjects intrx
- 2nd RM main effect error term = RM_2 x Subjects intrx
- RM_1 x RM_2 interaction error term = RM_1 x RM_2 x Subjects intrx

Factorial RM ANOVA: Summary Table

Source	*SS*	*df*	*MS*	*F*	*p*
Subj			X	X	X
RM1					
Error(RM1 x Subj)				X	X
RM2					
Error(RM2 x Subj)				X	X
RM1 x RM2					
Error(RM1 x RM2 x Subj)				X	X
Total			X	X	X

Effect Size: η^2

- Little evidence for a RM factor X Subject interaction (additivity met) (Keppel & Wickens, 2004)
  - Compute depending on effect of interest

$$\text{Partial } \eta^2 = \frac{SS_{RM1}}{SS_{RM1} + SS_{RM1xS}} \text{ or } \frac{SS_{RM2}}{SS_{RM2} + SS_{RM2xS}} \text{ or } \frac{SS_{RM1xRM2}}{SS_{RM1xRM2} + SS_{RM1xRM2xS}}$$

- Evidence for interaction (non-additivity)
  - Conservative or 'lower bound' estimate
  - Compute depending on effect of interest

$$\eta^2 = \frac{SS_{RM1}}{SS_{Total}} \text{ or } \frac{SS_{RM2}}{SS_{Total}} \text{ or } \frac{SS_{RM1xRM2}}{SS_{Total}}$$

  - Present the range

Effect Size: ω^2

- Little evidence for a RM factor X Subject interaction
  - Compute depending on effect of interest

Main RM effect: Partial $\omega^2 =$

$$\frac{df_{RM1}(MS_{RM1} - MS_{RM1xRM2xSubj})}{df_{RM1}(MS_{RM1} - MS_{RM1xSubj}) + k_{RM1}N(MS_{RM1xSubj})}$$

Interaction between RM factors: Partial $\omega^2 =$

$$\frac{df_{RM1xRM2}(MS_{RM1xRM2} - MS_{RM1xRM2xSubj})}{df_{RM1xRM2}(MS_{RM1xRM2} - MS_{RM1xRM2xSubj}) + k_{RM1xRM2}N(MS_{RM1xRM2xSubj})}$$

Where $k_{RM1xRM2}$ = Number of cells in RM ANOVA factorial design; RM factors only, not including levels due to participants.

Example: 2x3 RM ANOVA, $k = 6$

In both equations, N = # independent participants or sets of participants

Multiple Comparisons

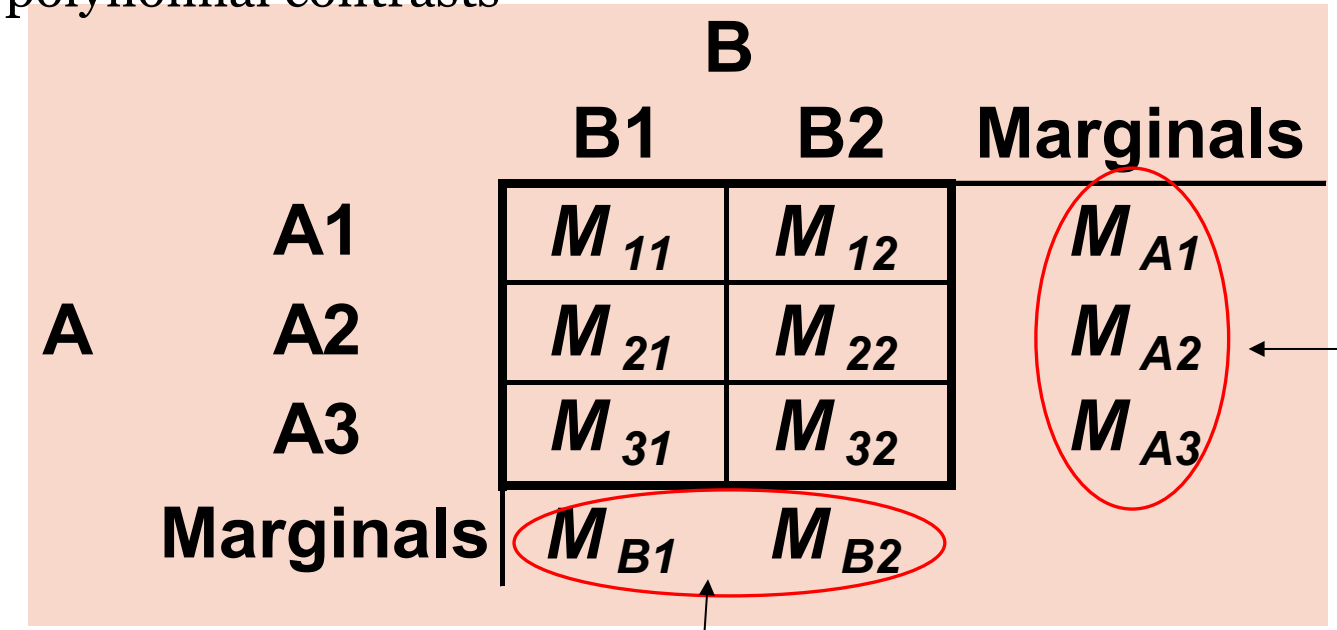
- Similar procedures as other ANOVA designs

- Different error term **technically required for each RM** comparison
  - Error represents differences among participants across levels of RM factor + random error
  - When a contrast omits one or more levels of the RM factor, how do we know whether omnibus error term represented by RM x Subjects factors still applies to remaining levels? Hard to say…
- However, use of $\boldsymbol{MS_{Intrx}}$ **as error term in omnibus multiple comparisons** is usually justified
  - i.e., Follow-up 1-Way RM ANOVAs for simple main effects following interaction
  - Similar to follow-up 1-Way Independent Groups ANOVAs following significant Factorial ANOVA

- **Simple or pairwise comparisons** avoid this problem by use of paired-samples *t*-tests or trend analysis procedures (*recommended*)

Non-Significant Interaction(s)

- Only significant RM main effects
  - Reduces to two 1-Way RM ANOVAs
- Marginal means are contrasted
  - Paired-samples *t*-tests; α_{PC} adjustment
  - Trend analysis or polynomial contrasts



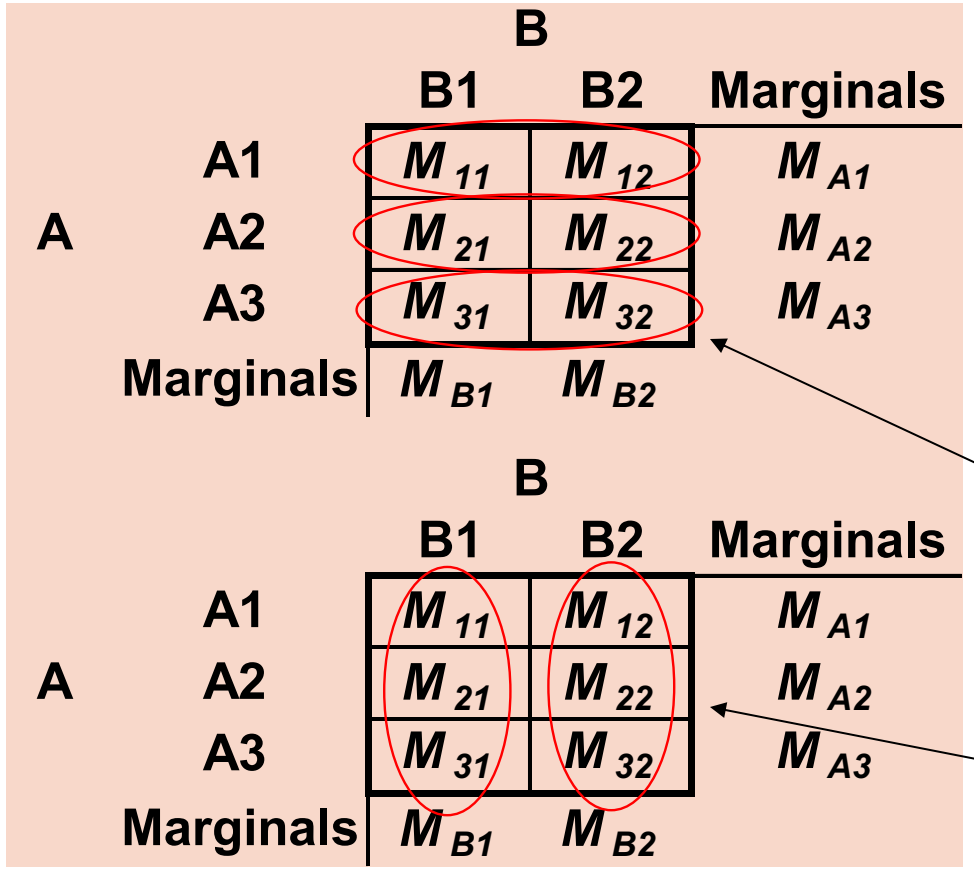
Simple or complex comparisons among marginal means (levels) if *F*-test significant

No further tests if ***F***-test of main-effect indicates difference

Significant Interaction(s)

- **Visualize: Plot means**
- Tests of simple (main) effects
  - Contrast means from levels of one RM factor within levels of another RM factor using 1-way RM ANOVA, paired-samples *t*-tests, or polynomial contrasts
- **Avoid interpretation of main effects**
- Alternative: Tests of interaction contrasts
  - Create difference scores between levels of one factor within each level of another factor and compare with paired-samples *t*-tests
    - Order dictates valence of difference scores
  - Results will indicate whether mean differences across one condition vary across levels of other condition

Significant Interaction(s)



- Direction of 'simple effect' testing determined by researcher
- Simple effects generally tested for each level of stratifying factor
  - Simple comparisons
    - Paired-samples *t*-tests
  - 1-way RM ANOVA followed by simple or complex comparisons (e.g., Paired-samples *t*-tests)

Reporting Results

- **Summary** information: sample means and either *SD*s, *SE*s, *CI*s
- **Effect size** measures for main effects or interactions (even if non-significant)
- Results of **post hoc** comparisons
- Mean differences and interactions can be **graphically** depicted

Problems

- **Extraneous factors (internal validity)**
  - Passage of time in longitudinal studies
    - Do conditions, equipment, experimenters, participants change (interest, practice, skills) over the course of the study in ways that may invalidate results?
  - Need methodological control
- **Generalizability (external validity)**
  - Using fewer participants, so sample is less representative of population
- **Poor matching, small *n*, violated assumptions** may lead to deflated power in RM ANOVA so that its power is same as Independent Groups ANOVA
- If a participant is **missing data** on outcome from any level of any RM factor, all data from that participant is removed from analysis
  - Decreased *N* → less power
  - However, easier to impute missing data in RM ANOVA than in randomized- or independent-groups designs
    - Other outcome scores are available from participants with missing values
    - Imputation results in several data sets on which the same analysis is conducted and results are compared

Supplemental

MS_{RM*S} Can use to calculate the ICC

$$ICC = \frac{MS_{Sub} - MS_{RMxS}}{MS_{Sub} + (c-1)MS_{RMxS} + \frac{c}{n}(MS_{RM} - MS_{RMxS})}$$