

ANALYTICAL REVIEW

Univariate Basics

COMPARING MEANS

- Single samples
- Dependent samples
- Independent samples



WHAT ARE WE TESTING?

○ Single samples

- Single sample mean to a population mean posited by the null hypothesis

○ Paired sample t

- Single sample of difference scores → mean difference score compared to a population mean difference score

○ Independent samples

- Compare the difference between sample means to the population mean difference



GOING FURTHER: EFFECT SIZE D

- Effect size
- Standardized mean difference
 - Cohen's d, Glass's delta and variants
- Applicable to both independent and dependent though caution must be applied to the latter
 - There is choice in standardizer



GOING FURTHER: EQUIVALENCE TESTING

- What if your question regards equivalence rather than a difference between scores?
 - Example: want to know if two counseling approaches are similarly efficacious
- Just do a t-test, and if non-significant conclude they are the same
- Wrong! It would be logically incorrect.
 - Can you think of why?
- Equivalence testing has effect size as a starting point
- Choose a meaningful effect (Δ) such that any difference that size or greater could not be construed as indicating equivalent groups
- However, to show equivalence you must show that the small difference seen did not occur by chance
- Equivalence testing actually tests both difference and equivalence at the same time, places focus on effect size rather than statistical significance, and allow ambiguous outcomes to not be overstated in terms of difference or equivalence if the evidence just simply isn't there



CONFIDENCE INTERVALS

- The mean or difference between means is a point estimate
- We should also desire an interval estimate reflective of the variability and uncertainty of our measurement seen in the data

$$\textit{Limits} = \overline{X} \pm t_{.95} (s_{\overline{X}})$$

Where

\overline{X} = sample mean

$t_{.95}$ = t critical value

$s_{\overline{X}}$ = standard error of the mean



WHAT CONFIDENCE INTERVALS REALLY TELL YOU

- A 95% confidence interval means that:
- 95% of the confidence intervals calculated on repeated sampling of the same population will contain μ
- Note that the population value does not vary i.e. it's not a 95% chance that it falls in that specific interval per se
- There are an infinite number of 95% CIs that are consistent with a fixed population parameter
- In other words, the CI attempts to capture the mean, and the one you have is just one of many that are correct
- http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/index.html



CONFIDENCE INTERVALS

- As suggested by many leading quantitative psych guys, using confidence intervals in lieu of emphasizing p-values may make for a better approach for evaluating group differences etc.
- Confidence intervals provide all the relevant NHST info as well as a means of graphically displaying statically significant differences (with Inferential Confidence Intervals)
- One approach is to simply see if our interval for the statistic of interest contains the H_0 value.
- If not, reject H_0



INITIAL SUMMARY

- Confidence intervals are an important component statistical analysis and should always be reported
- Non-significance on a test of difference does not allow us to assume equivalence
- Methods exist to test the group equivalency, and should be implemented whenever that is the true goal of the research question





ANOVA

- Most popular analysis in psychology
 - Ease of implementation¹
 - Allows for analysis of several groups at once
 - Allows analysis of interactions of multiple independent variables



SOURCES OF VARIABILITY

SS_{Total}

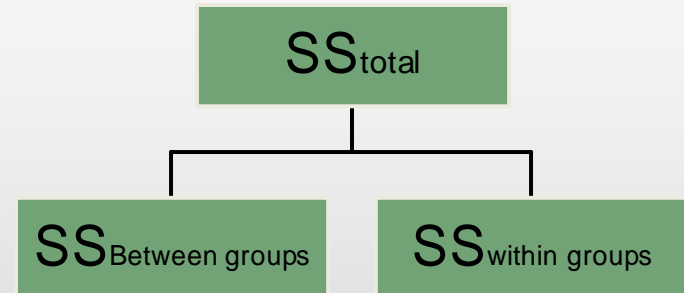
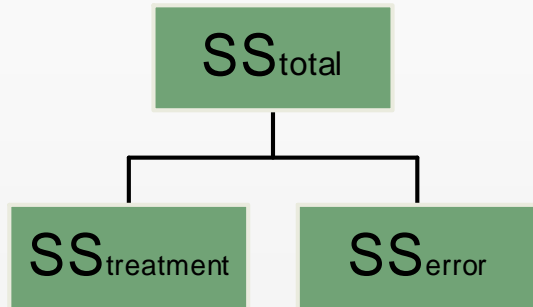
$$\sum (X_{ij} - \bar{X}_{..})^2$$

SS_{Treat}

$$\sum n(\bar{X}_j - \bar{X}_{..})^2$$

SS_{error}

$$\sum \sum (X_{ij} - \bar{X}_j)^2$$



ASSUMPTIONS

- If the assumptions of the test are not met we may have problems in its interpretation
- Assumptions
 - Independence of observations
 - Normally distributed variables of measure
 - Homogeneity of Variance
 - Sphericity (repeated measures)
 - Homogeneity of covariance matrices among groups (mixed design)



GIST

- Approach One-way ANOVA much as you would a t-test
 - Same assumptions and interpretation taken to 3 or more groups
 - One would report similar info: effect sizes, confidence intervals, graphs of means etc.
- With ANOVA one must run planned comparisons or post hoc analyses to get to the good stuff as far as interpretation
- Turn to nonparametric robust options in the face of yucky data and/or violations of assumptions



THE ONE-WAY DESIGN

- One-way ANOVA
- What does it tell us?
 - Means are different
 - How?
 - Don't know
- What if I want to know the specifics?
 - Multiple comparisons



OMNIBUS F TEST

- We can do multiple comparisons with a correction for type I error or planned contrasts to test the differences we expect based on theory
- A priori vs. Post hoc
 - Before or after the fact
- A priori (planned)
 - Do you have an expectation of the results based on theory?
 - A priori
 - Few comparisons
 - More statistically powerful
- Post hoc
 - Look at all comparisons while maintaining type I error rate



OMNIBUS F TEST

- Do we need an omnibus F test first?
 - Current thinking is no
 - Most multiple comparison procedures maintain type I error rates without regard to omnibus F results
 - Multivariate analogy
 - We do multivariate and then interpret in terms of uni anovas
 - Begg the question of why the multivariate approach was taken in the first place



WHICH POST HOC?

- Some are better than others
- However which are better may depend on situation
- Try alternatives, but if one is suited specifically for your situation use it.
- Some suggestions
 - Assumptions met: Tukey's or REWQ of the traditional options, FDR for more power
 - Unequal n: Gabriel's or Hochberg (latter if large differences)
 - Unequal variances: Games-Howell
- More later



PLANNED CONTRASTS

- The point of these type of analyses is that you had some particular comparison in mind before even collecting data.
- Why wouldn't one do a priori all the time?
 - Though we have some idea, it might not be all that strong theoretically
 - Might miss out on other interesting comparisons
 - Too lazy to think more about the research problem



SUMMARY FOR MULTIPLE COMPARISONS

- Let theory guide which comparisons you look at
 - Have a priori contrasts in mind whenever possible
- Test only comparisons truly of interest
- Use more recent methods for post hocs for more statistical power



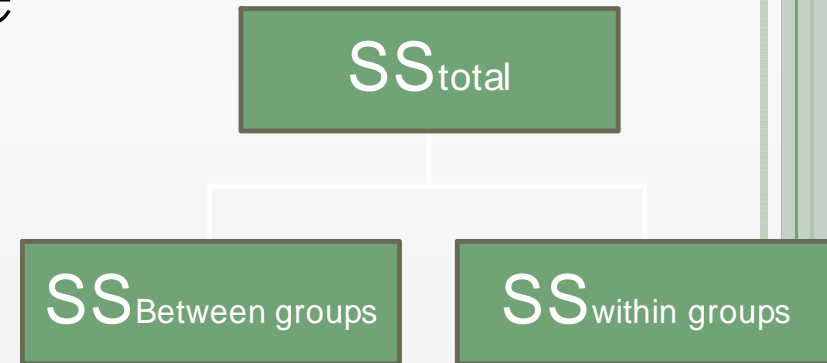
MORE COMPLEX DESIGNS

- Factorial between groups
- Repeated Measures
- Mixed



THE ONE-WAY DESIGN AS A STARTING POINT

- Total variability in the outcome is a source of:
 - Differences between groups
 - Differences within groups

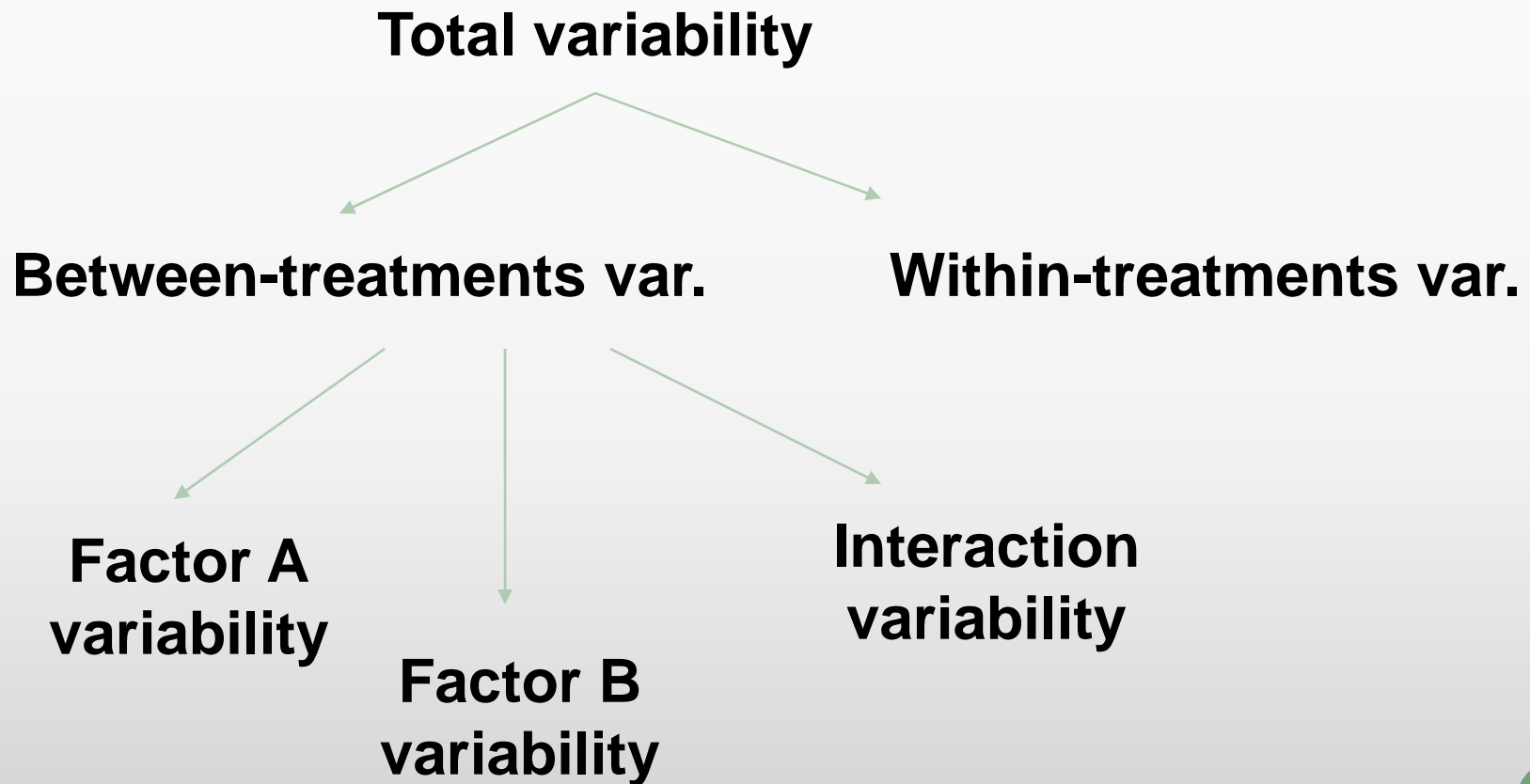


FACTORIAL ANOVA

- With factorial designs we have additional sources of variability to consider
- Main effects
 - Mean differences among the levels of a particular factor
- Interaction: what it means
 - Differences among cell means not attributable to main effects
 - Interactions in this sense are residual effects (i.e. whatever is left over after main effects are accounted for)
 - Note the GLM for a factorial ANOVA
 - When the effect of one factor on a DV is influenced by the levels of another

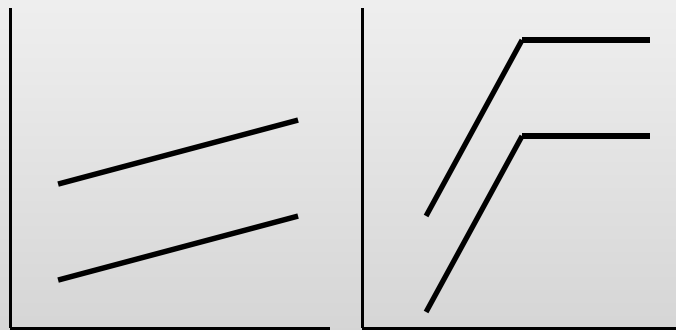


PARTITIONING THE DV VARIANCE

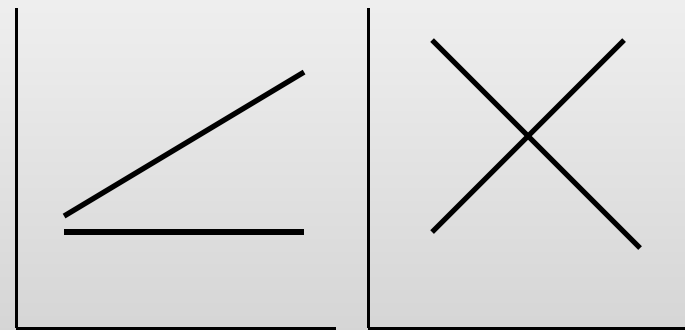


GRAPHICAL DISPLAY OF INTERACTIONS

- What are we looking for?
- Do the lines behave similarly (are parallel) or not?
- Does the effect of one factor depend on the level of the other factor?



No interaction



Interaction



INTERPRETATION OF INTERACTIONS

- Note that with a significant interaction, the main effects are understood only in terms of that interaction
- In other words, they cannot stand alone as an explanation and must be qualified by the interaction's interpretation



INTERPRETATION OF INTERACTIONS

- However, interpretation depends on common sense, and should adhere to theoretical considerations
 - Plot your results in different ways
- If main effects are meaningful, then it makes sense to at least talk about them, whether or not an interaction is statistically significant or not
- To help you interpret results, test simple effects
 - Is simple effect of A significant within specific levels of B?
 - Is simple effect of B significant within specific levels of A?

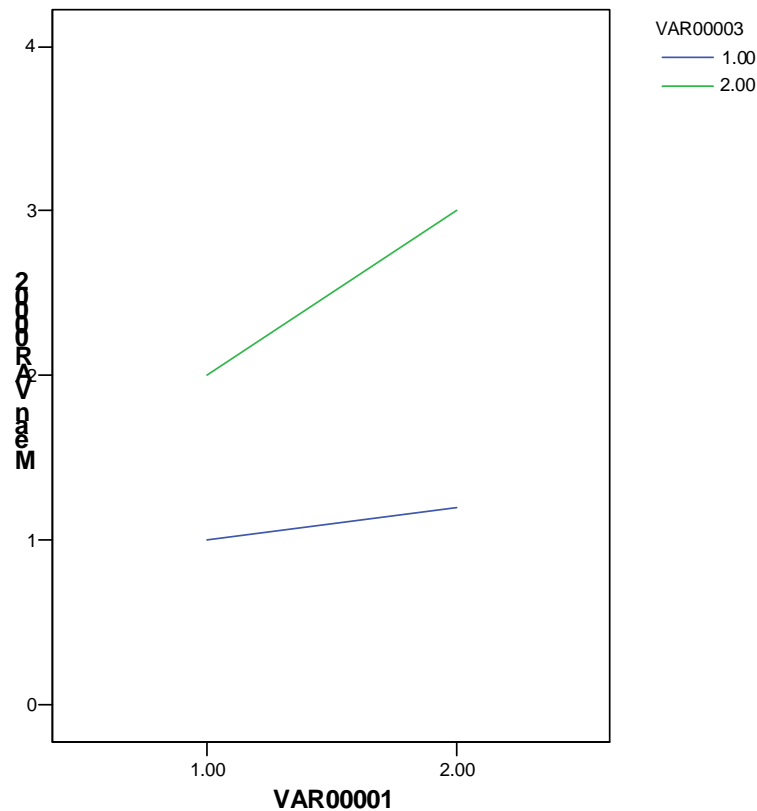


SIMPLE EFFECTS

- Analysis of the effects of one factor at one level of the other factor
- Example:
 - Levels of A_{1-3} at B_1 and A_{1-3} at B_2
- Note that the simple effect represents a partitioning of $SS_{\text{main effect}}$ and $SS_{\text{interaction}}$
 - It is not just a breakdown of the interaction variance!



NONSIGNIFICANT INTERACTION



- What if there was no significant interaction, can I still test for simple effects?
- Technically yes¹
- A significant simple effect suggests that at least one of the slopes across levels is significantly different than zero
- However, one would not conclude that the interaction is 'close enough' just because there was a significant simple effect
- The nonsig interaction suggests that the slope seen is not significantly different from the other(s) under consideration.



REPEATED MEASURES

- Instead of having one score per subject, studies are frequently conducted in which multiple scores are gathered for each case
- Repeated Measures or Within-subjects design
- Advantages
 - Design – nonsystematic variance (i.e. error, that not under experimental control) is reduced
 - Take out variance due to individual differences
 - Efficiency – fewer subjects are required
 - More sensitivity/power all else being equal



WHEN TO USE?

- Measuring performance on the same variable over time
 - Example: looking at changes in performance during training or before and after a specific treatment
- The same subject is measured multiple times under different conditions
 - Example: performance when taking Drug A and performance when taking Drug B
- The same subjects provide measures/ratings on different characteristics
 - Example: the desirability of red cars, green cars and blue cars
- Note how we could do some RM as regular between subjects designs
 - Example: Randomly assign to drug A or B

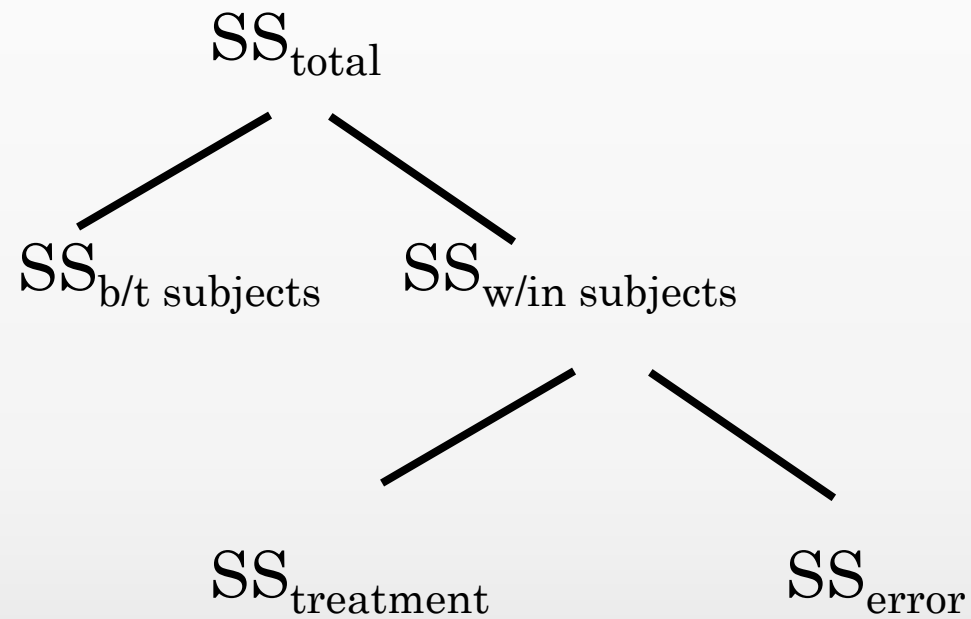


INDEPENDENCE

- Analysis of variance as discussed previously assumes cells are independent
- But here we have a case in which that is unlikely
 - For example, those subjects who perform best in one condition are likely to perform best in the other conditions



PARTITIONING THE VARIANCE



INTERPRETATION

- As with a regular one-way Anova, the omnibus RM analysis tells us that there is some difference among the treatments (drugs)
- Often this is not a very interesting outcome, or at least, not where we want to stop in our analysis



CONTRASTS

Deviation	Compares the mean of one level to the mean of all levels (grand mean)
Simple	Compares each mean to some reference mean (either the first or last category e.g. a control group)
Difference (reverse Helmert)	Compares level 1 to 2, level 3 with the mean of the previous two etc.
Helmert	Compares level 1 with all later, level 2 with the mean of all later, level 3 etc.
Repeated	Compares level 1 to level 2, level 2 to level 3, 3 to 4 and so on
Polynomial	Tests for trends (e.g. linear) across levels

- If you had some particular relationship in mind¹ you want to test due to theoretical reasons (e.g. a linear trend over time) one could test that by doing contrast analyses (e.g. available in the contrast option in SPSS before clicking ok).
- This table compares the different contrasts available in SPSS
 - Deviation, simple, difference, Helmert, repeated, and polynomial.



SPHERICITY

- Observations may covary, but the degree of covariance must remain the same
- If covariances are heterogeneous, the error term will generally be an underestimate and F tests will be positively biased
- Such circumstances may arise due to carry-over effects, practice effects, fatigue and sensitization
- Suppose the factor TIME had 3 levels – before, after and follow-up
- RM ANOVA assumes that the 3 correlations
 - r (Before-After)
 - r (Before-Follow up)
 - r (After-Follow up)
- are all about the same in size



MIXED DESIGN

- $A \times (B \times S)$
- At least one between, one within subjects factor
- Each level of factor A contains a different group of randomly assigned subjects.
- On the other hand, each level of factor B at any given level of factor A contains the same subjects

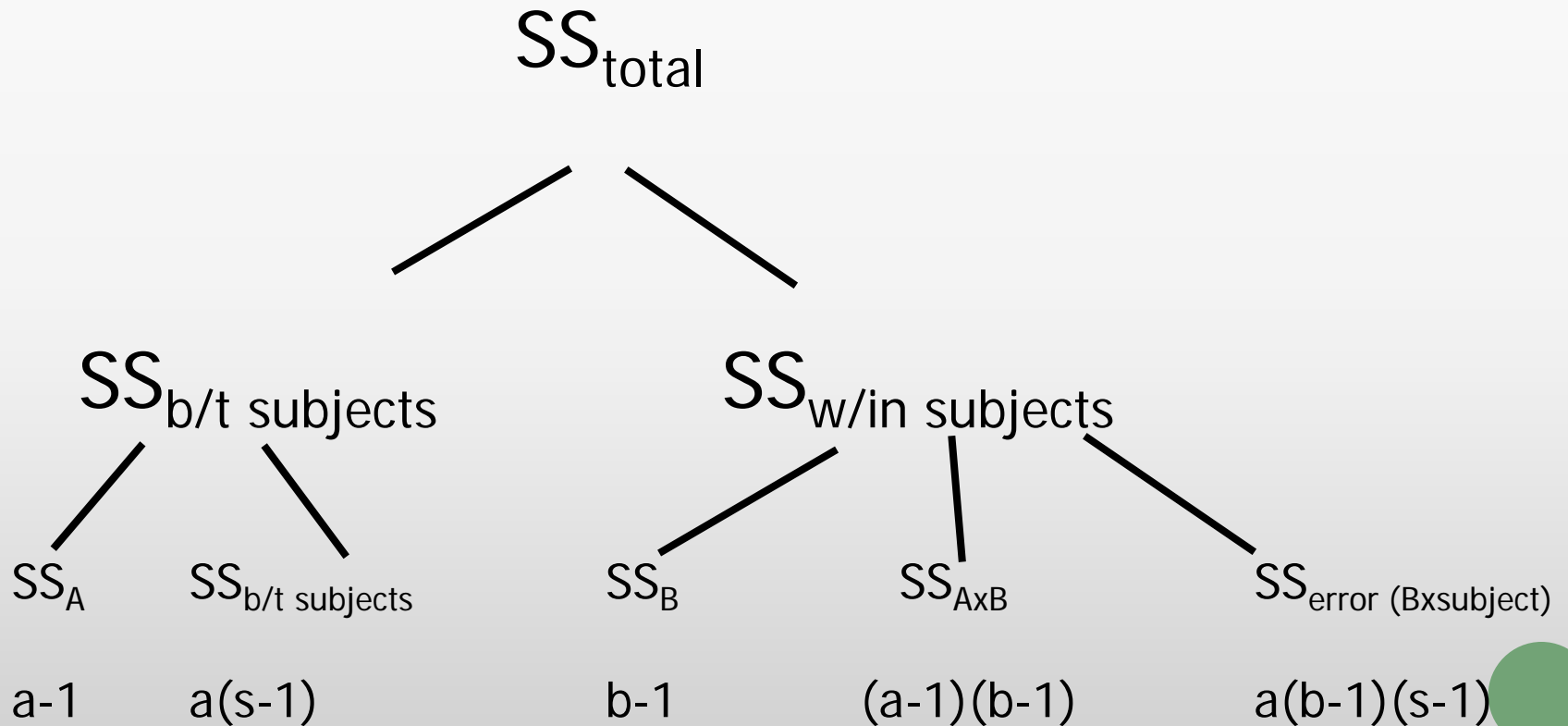


PARTITIONING THE VARIANCE

- Again we adopt the basic principle we have followed previously in looking for effects. We want to separate between treatment effects and error
 - A part due to the manipulation of a variable, the treatment part (treatment effects)
 - A second part due to all other uncontrolled sources of variability (error)
- The deviation associated with the error can be divided into two different components:
 - Between Subjects Error
 - Estimates the extent to which chance factors are responsible for any differences among the different levels of the between subjects factor.
 - Within Subjects Error
 - Estimates the extent to which chance factors are responsible for any differences observed within the same subject



HOW IT BREAKS DOWN¹



COMPARING THE DIFFERENT DESIGNS

B/t groups

SS_A
 $SS_{A/S}$

W/in groups

SS_S
 SS_B
 $SS_{B \times S}$

Mixed

SS_A
 $SS_{A/S}$
 SS_B
 $SS_{A \times B}$
 $SS_{B \times S}$

- Note that the between groups outcome is the same in the mixed and b/t groups design
- The same is true for the within groups design, except in the mixed the 'subjects' are nested within the factor of A



COMPARING THE DIFFERENT DESIGNS

- The $SS_{b/t \text{ subjects}}$ reflects the deviation of subjects from the grand mean while the $SS_{w/in}$ reflects their deviation from their own mean
- The mixed design is the conjunction of a randomized single factor experiment and a single factor experiment with repeated measures
 - Though this can be extended to have more than one b/t groups or repeated factor

