

Item analyses vs. subject analyses

- In a typical statistical test, you look at the data across different subjects (collapsing across different items).
 - Ex: you're comparing identification scores for speech spoken quickly vs. carefully.
 - 100 different sentences, each spoken two ways.
 - Subjects identify these in noise.
 - In a typical test, you'd have one score for each subject on all the quick sentences, and one score for each subject on the careful sentence.

Item analyses vs. subject analyses, cont.

- Each "cell" or data point is for one subject, but for many items.
- You would do a t-test comparing normal vs. careful speech using the data from each subject.
- This is called a subject analysis.

	Sentence style	
	normal	careful
subj 1	65%	72%
subj 2	56%	65%
subj 3	85%	88%
subj 4	44%	59%
subj 5	78%	77%
etc		

Item analyses vs. subject analyses, cont.

- You can also do the reverse.
- That is, for each SENTENCE, average across the different subjects in each condition.

	Sentence style	
	normal	careful
sentence 1	65%	75%
sentence 2	55%	65%
sentence 3	85%	90%
sentence 4	45%	60%
sentence 5	80%	75%
etc		

Item analyses vs. subject analyses, cont.

- | | |
|--|---|
| <p>Subject analysis</p> <ul style="list-style-type: none"> • Each data point is an average across all 100 items. • You'd have 20 data points, one for each subject. • What you're really measuring is whether there is a consistent trend across subjects for one condition to be better than the other. • Labeled t_1 or F_1 | <p>Items analysis</p> <ul style="list-style-type: none"> • Each data point is an average across all 20 subjects. • You'd have 100 data points, one for each item. • What you're really measuring is whether there is a consistent trend across items (sentences) for one condition to be better than the other. • Labeled t_2 or F_2 |
|--|---|

Effect size

- A family of indices that measure the magnitude of an effect (not just whether it is present or not) , or the proportion of the variance that can be attributed to the effect
- These used to be rarely reported, but now APA explicitly requests that some measure of effect size be provided for all statistics

Why use them?

- They allow for comparison across studies
- They are not a direct function of sample size, the way p-values are
 - this is one reason why it is not valid to just compare p-values; just because one experimenter ran more subjects doesn't mean their effect is really any bigger

Two classes of effect size measures

- Those based on standardized mean differences (the difference between the means, calculated in terms of standard deviations)
 - Cohen's d
 - Hedges' g
 - Glass' Δ
- Those based on correlational measures of the association between variables
 - Eta squared (η^2)
 - Partial eta-squared (η_p^2)
 - Omega squared (ω^2)

Bonferroni adjustments

- A statistical adjustment for multiple comparisons
 - raises the standard of evidence needed when doing large numbers of comparisons to avoid capitalizing on chance
 - Particularly important for “fishing expeditions”, when the researcher is investigating a wide range of possible associations

Bonferroni adjustments

- Idea: Let's say you want to know whether any of the following might increase behavioral problems in children with ADD
 - Sugar intake; caffeine intake; aspirin intake; salt intake; alcohol intake; chocolate intake; lack of sleep; hours of TV; hours of computer games
- This is 9 factors - if you test all 9 at a .05 level, there's a high likelihood one would be significant just by chance (even if it is really harmless)
 - This is referred to as inflation of the type I error, or the cumulative type I error rate

Bonferroni adjustments

- Another example:
 - You want to see if Hearing Aid X is better than Hearing Aid Y.
 - You have only this one IV (hearing aid), but to maximize your likelihood of finding an effect, you do a variety of outcome measures (SPIN, HINT, etc etc. etc.)
 - If you would accept a significant difference if ANY of these tests gave a significant outcome, that is called the “global null” hypothesis, and you should adjust alpha.
 - If you'd only accept a significant difference if ALL showed a significant outcome, then there's no need.

Bonferroni adjustments

- The adjustment: if you are doing N different tests, divide your alpha level by N .
 - This way, the combined alpha level is still no higher than desired
- Problem: this really assumes the tests are independent; when they're not, this can be too severe of an adjustment, and thus increase type II error!

When to use vs. not use

- Global null hypothesis vs. more restrictive
- Exploratory analysis: if you are clear that it's just exploratory, that may be fine - but if you want the results to stand on their own without follow-ups, you'll need to correct
- When you have a large number of tests, and can't identify any as primary (if there are just a few primary analyses, and others are clearly secondary, not as necessary)
- Post-hoc vs. pre-planned tests (when the whole study was designed around a few tests, don't adjust)

MANOVA (Multivariate Analysis of Variance)

- Used when there is more than one DV at a time.
- Ex: We try two different textbooks, and are interested in the students' improvements in math and physics.
 - Two dependent variables, and our hypothesis is that both together are affected by the difference in textbooks.
- Instead of a univariate F value, we would obtain a multivariate F value (Wilks' lambda).
- MANOVA takes into account that the two measures are probably correlated in some way.

Multiple regression

- Allows you to analyze the interaction of several variables at once.
- Finds an equation that predicts one variable (the DV, Y) from two or more independent (X) variables.
- Allows you to predict scores on one test from scores on other tests
 - Ex: predicting blood pressure from age, weight & gender.
 - Ex: predicting how well a kid will do on a speech perception task after getting a CI, where there are a number of variables that might play a role (memory skill, age at implantation, etc.).

Multiple regression, cont.

- In some situations, your goal may be to examine several variables at once.
 - Ex: Which variable has the largest influence on blood pressure: age, weight or gender?
 - Or your goal may be to find an equation that best predicts blood pressure from those three variables.
- In other situations, you really only care about one of the IVs, but your analysis needs to adjust for differences in other variables.
 - Ex: Does blood pressure vary with age, after correcting for differences in weight and differences between the sexes?

Multiple regression, cont.

- Result is a formula out that looks like
 - $y = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$.
- Looks for linear relationships only.
- Can do a multiple regression if you have multiple IVs and a single DV, or multiple DVs and a single IV.

Multiple Correlation Coefficient

- A part of multiple regression – this tells you how much your IV is related to the composite of your DVs.
- Ex: You want to identify attributes of speaking behavior that influence judgments of stuttering severity.
 - First, speech samples rated for degree of severity.
 - These are analyzed acoustically to see what attributes they had (freq. of stuttering, average duration of moments of stuttering, etc.) that might influence people's ratings.
 - Do a multiple regression.
 - If the multiple correlation coefficient is high, it would suggest that the group of attributes (as a whole) tended to influence people's ratings of stuttering severity.

Order of factors

- Different factors can be entered in different orders.
 - Ex: predicting grad school GPA from undergrad GPA & GRE scores.
 - Computer could determine relationship between undergrad & grad GPA, and then add in the GRE scores.
 - Or it could look for GRE & GPA, and then add in the undergrad GPA...
 - The order in which the DVs are entered into the equation will result in slight differences.

Types of multiple regression

- Hierarchical multiple regres.
 - You have a theoretical reason why one thing should be more important than another, and you use this to set the order of entry of the variables.
 - The computer calculates at each stage whether that additional variable adds anything meaningful to the equation.
 - If you have a theoretical reason to believe one thing should be more important than another, then finding that relationship is less likely to be a fluke.
- Stepwise multiple regression
 - The computer calculates the regression in stages.
 - The IV that best correlates with the DV is included first; then the one that is next best once the first one is controlled for, etc.
 - Continue until the addition of a remaining IV does not increase R^2 significantly (or until none left)
 - Maximizes the likelihood of getting a significant formula by looking for all relationships, but thus may be less likely to generalize (capitalizes on chance)

Coefficients of partial correlation

- Partial out the amount of relationship due to a 3rd factor, and tells you how much of a relationship is “left over.” (Factors out the 3rd variable statistically.)
- Used when there is a factor that you couldn’t control experimentally, but might play a role.
 - Ex: is there a relationship between auditory discrimination ability & articulation proficiency for kids in kindergarten?
 - There could be a relationship between the two because both are related to mental age. The higher the mental age of the child, the likely the child is to do well on both tasks.
 - Really want to know if there would there be any relationship if the effects of mental age were eliminated? (i.e., held constant)

MANCOVA & ANCOVA

- Tests for differences between groups once some other factor is controlled for statistically.
 - If there is a variable you don’t care about that could cause a confound, you can avoid problems experimentally by holding that factor constant or match it across groups.
 - ANCOVA is a way of doing the same thing statistically, after the fact, rather than by some selection criteria before the data are gathered.
 - Ex: You want to compare 2 intervention programs, but the two groups of kids available differ slightly in age. Using ANCOVA, you basically eliminate differences between pre- & post-intervention scores that might have been caused by age, and then run an ANOVA on these adjusted scores.

Factor analysis

- (principal factor anal. & principal component anal.)
- A technique for identifying general factors or abilities that influence performance on a series of tasks.
 - Ex: Imagine you told someone with aphasia to name objects, then point to objects, then do some type of motor task. One general factor that underlies all three is the ability to understand instructions.
- Factor analysis says how many underlying factors there are, and what particular tasks can be used to evaluate each one.

Factor analysis, cont.

- Allows you to
 - reduce the number of variables
 - detect structure in the relationships between variables (classify variables)
 - Ex: You want to measure people’s satisfaction with their lives.
 - Design a satisfaction questionnaire; among the questions are how satisfied people are with their hobbies (item 1) and how intensely they are pursuing a hobby (item 2).
 - Most likely, the responses to these two items are highly correlated with each other (and thus redundant). To make a more efficient test you’d want to only have the question that better gets at that issue.

Factor analysis, cont.

- Ex: You want to develop a comprehensive test for aphasia. You test many aphasics on a large number of possible tasks, and want to know which are important to include and which are redundant.
- Can use factor analysis to tell you:
 - the number of general abilities that presumably influenced the aphasics’ performance
 - specific tasks that had relatively high loadings and low loadings for each ability.
 - A task with a high loading for a particular factor would be good for assessing that factor; one with a low loading would not.
- Then you select tasks with a high loading for each factor.

Factor analysis, cont.

- The computer can not label what the general factors are – it can just say there's one factor that seems to be related to performance on these particular tasks, to these particular degrees, etc.
- You then need to infer what ability that factor represents.
 - Ex: If a particular factor was related to how people did on sentence speaking tasks, but not to how they did on motor skills, speech repetition skills, or comprehension tasks, you might label that factor as "sentence planning".

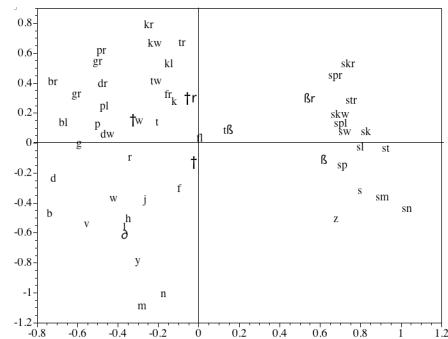
Discriminant function analysis

- Statistical approach to either:
 - Assign people to categories
 - Ex: Is a given 4-year old "normally disfluent" or "beginning to stutter"
 - Determine which variables discriminate between two or more naturally occurring groups.
 - Ex1: What variables best differentiate "normally disfluent speech" vs. "beginning stuttering" for 4-year-olds?
 - Ex2: Researcher wants to investigate which variables discriminate between high school graduates who go to college vs. trade school vs. no further education. Can collect data on numerous IVs prior to students' graduation, then see which categories people fall into. Discriminant analysis could be used to determine which IVs are the best predictors of students' subsequent educational choice.

Multidimensional scaling (MDS)

- Technique by which the computer seeks to identify attributes of stimuli or events to which people attend.
 - Ex: You're interested in what phonemes are confusable with one another in the context of noise. You have data on how often each of 18 consonants was confused with each other consonant.
 - MDS will pull out the different dimensions to which listeners were sensitive.
 - Plots a multidimensional space, in which the smaller the distance between two objects, the more similar they are.

MDS for cluster similarity ratings



Cluster analysis

- An exploratory procedure for identifying possible subgroups of subjects.
 - Groupings consist of subjects who performed similarly to one another and differently from the others.
 - Groupings are based on calculations of the distances between measures of the subjects' behavior during the performance of one or more tasks.
- The researcher can use this to make hypotheses about how subgroups differ from one another, which then could be tested in later experiments.
 - Ex: Are there different types of phonological disorders?