Power Analyses, Effect Sizes, Confidence Intervals & Data Cleaning

- Power analyses – n, S & N
- Effect Size Estimates
- Confidence Intervals
- Combining the information
- Data Cleaning

k-group Power Analyses

Power analysis is about the interrelationships among four things...

1. the size of the effect - r
2. the sample size of the study -- N
3. Type I Error Rate (False Alarm) -- α, usually .05
4. Type II Error rate (Miss) -- β , usually .20
   - remember that power is 1 - β, usually set at .80

A priori power analyses

- conducted before the study is begun
- start with r, β & α and determine the needed N

Post hoc power analysis

- conducted after retaining H0:
- start with r, N & α and determine β (1-power) of the study

Power analyses are dependent upon having accurate effect size estimates !!!!

Determining the power you need ..

For a 2-condition design...

- the omnibus-F is sufficient -- retain or reject, you're done !
- you can easily determine the sample size needed to test any expected effect size with a given amount of power

For a k-condition design …

- the power of the omnibus-F - isn't what matters !
- a significant omnibus-F only tells you that the "most different" means are significantly different
- follow-up (pairwise) analyses will be needed to test if the pattern of the mean differences matches the RH:
- you don't want to have a "pattern of results" that is really just a "pattern of differential statistical power"
- you need to assure that you have sufficient power for the smallest pairwise effect needed to test your specific RH:

Effect Size & Power Analyses for 2-BG designs

Effect Size computation

$$r = \sqrt{[F / (F + df_{error})]}$$

Friedman's table was designed for 2-BG designs …

- S as the total number of participants and ...
- n = S / 2
  - n = the number of participants in each condition

Effect Size & Power Analyses for k-BG designs

Effect Size Computation

- you won't have F-values for the pairwise comparisons
- Use a 2-step computation
  - $d = (M_1 - M_2) / \sqrt{MSerror}$
  - $r = \sqrt{\dfrac{d^2}{d^2 + 4}}$ (This is an "approximation formula")

Using Friedman's table for k-BG designs …

- when we base power analyses on pairwise comparisons, S tells us the number of participants *in those two conditions*
- n (the number in each IV condition) is still S/2, but …
- for a k-group design, N = k * n (or N = S/2 * k)
  - so, if you S = 120 for the smallest pairwise comp, and you have k=4 conditions, you need n = 60 for each or N=240

Effect Size & Power Analyses for 2-WG designs

Effect Size computation

$$r = \sqrt{[F / (F + df_{error})]}$$

Using Friedman's table for 2-WG designs …

- S as the total number of participants and …
- Each participant completes both IV conditions

Power Analyses for k-WG designs

Effect Size Computation

- you won't have F-values for the pairwise comparisons
- Use a 3-step computation
  - $d = (M_1 - M_2) / \sqrt{(MSerror * 2)}$
  - $\mathbf{d_w = d * 2}$
  - $r = \sqrt{\dfrac{d_w^2}{d_w^2 + 4}}$ (This is an "approximation formula")

Using Friedman's table for k-WG designs …

- S as the total number of participants and …
- Each participant completes all "k" IV conditions

Remember, Pairwise NHST isn't the only approach to describing your data and testing your research hypotheses

Confidence Intervals around single means and mean differences are both available, the latter can be used to augment pairwise significance tests.

Using the combination of these techniques – effect sizes, power analyses & confidence intervals -- you will also get useful information about what effects need further study (perhaps with a modified research design), because the results of this analysis are inconclusive.

Confidence Intervals for Pairwise Mean Differences

Note: MS_{Error} and df_{Error} taken from the omnibus-F - for both below

CI = pairwise mean difference +/- $t * \sqrt{(2 * MS_{error} / n)}$

For CI formula:

t = t-critical for full model df_{Error} and degree of confidence desired
n = number of datapoints in each IV condition (or the average)

You can "adjust" for experiment wise error by changing the p-value used to look up the t in this formula.

You can compute an "HSD CI" by substituting Q for t in this formula.

Note: the +/- value of the 95% CI is the same as the LSD (p=.05) minimum mean difference.

Note: The same formula applies for BG and WG designs

Combing these different types of information …

		Cx			Tx1		
	mean	M dif	95% CI	η	M dif	95% CI	η
Cx	20.3						
Tx1	24.6	4.3	(-1.8 to 10.4)	.22			
Tx2	32.1	11.8*	(5.7 to 17.9)	.54	7.5*	(1.4 to 13.6)	.41

* indicates mean difference is significant based on LSD criterion (min dif = 6.1)

Examining these results…

• The effect size of Cx vs. Tx1 is substantial (Cohen calls .30 "medium and .10 small"), but is not significant (note LSD and CI agree), suggesting we should check the power of the study for testing an effect of this size.

• Related to this, the size of the Cx vs. Tx1 mean difference could be as large as 10.4 (the large interval also suggests the sample size is small)

So, what do you get out of all these analyses ???

effect size estimates

- mean difference means -- most basic description/inference but…
  - DV scale can be difficult to generalize
  - does not account for variability around the or sample size
- F-value -- integrates effect size, variability and sample size, but (without practice) is most useful to obtain p-value
- d, r, etc. -- tells "how big" is the effect considering variability, but without considering sample size/power - easy to interpret metrics (r & d), but tells nothing about the likelihood of α or β

assessing statistical conclusion error

- CI -- expresses mean difference taking variability and sample size (α) into account -- allows testing of non-nil H0: ("practical significance"
- p-value -- probability that a rejected H0: is a Type I error
- post-hoc power analysis - prob that a retained H0: is a Type II error

Data Cleaning for k-group ANOVA

BG Designs

- Perform outlier analyses separately for each IV condition
- Data from all IV conditions must be transformed the same way
- Consider results from trimming and Winsorizing
- Consider results from outlier treatments &/or symmetrizing

WG Designs

- Remember each IV condition is a separate variable
- Data from all IV conditions must be transformed the same way
- Trimming means you lose all the data for that case
- Consider results from outlier treatments &/or symmetrizing