Multiple Group X^2 Designs & Follow-up Analyses

- X^2 for multiple condition designs
- Pairwise comparisons & RH Testing
  - Alpha inflation
  - Effect sizes for k-group X^2
  - Power Analysis for k-group X^2
- gof-X2 & RH Testing
  - Alpha inflation & Bonferroni Correction
- Power Analyses for k-group Chi-square

ANOVA vs. X^2

- Same as before
  - ANOVA – BG design and a quantitative DV
  - X^2 -- BG design and a qualitative/categorical DV

While quantitative outcome variables have long been more common in psychology, there has been an increase in the use of qualitative variables during the last several years.

- improvement vs. no improvement
- diagnostic category
- preference, choice, selection, etc.

For example… I created a new treatment for social anxiety that uses a combination of group therapy (requiring clients to get used to talking with other folks) and cognitive self-appraisal (getting clients to notice when they are and are not socially anxious). Volunteer participants were randomly assigned to the treatment condition or a no-treatment control. I personally conducted all the treatment conditions to assure treatment integrity. Here are my results using a DV that measures whether or not the participants was "socially comfortable" in a large-group situation

$X^2(1) = 9.882, p = .005$

	Group therapy & self-appraisal	Cx
Comfortable	45	25
Not comfortable	10	25

Which of the following statements will these results support?

"Here is evidence that the combination of group therapy & cognitive self-appraisal increases "social comfort." ???

Yep -- treatment comparison causal statement

" You can see that the treatment works because of the cognitive self-appraisal; the group therapy doesn't really contribute anything."

Nope -- identification of causal element statement & we can't separate the role of group therapy & self-appraisal

Same story... I created a new treatment for social anxiety that uses a combination of group therapy (requiring clients to get used to talking with other folks) and cognitive self-appraisal (getting clients to notice when they are and are not socially anxious). Volunteer participants were randomly assigned to the treatment condition or a no-treatment control. I personally conducted all the treatment conditions to assure treatment integrity.

What conditions would we need to add to the design to directly test the second of these causal hypotheses...

The treatment works because of the cognitive self-appraisal; the group therapy doesn't really contribute anything."

Group therapy & self-appraisal	Group therapy	Self-appraisal	No-treatment control

Let's keep going …

Here's the design we decided upon. Assuming the results from the earlier study replicate, we'd expect to get the means shown below.

Group therapy & self-appraisal	Group therapy	Self-appraisal	No-treatment control
45	25	45	25
10	25	10	25

What responses for the other two conditions would provide support for the RH:

The treatment works because of the cognitive self-appraisal; the group therapy doesn't really contribute anything."

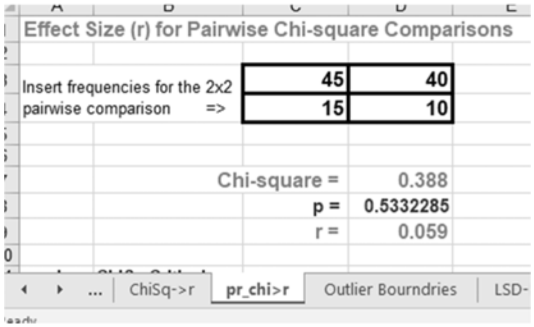
Omnibus X^2 vs. Pairwise Comparisons

- Omnibus X^2
  - overall test of whether there are any response pattern differences among the multiple IV conditions
  - Tests H0: that all the response patterns are equal
- Pairwise Comparison X^2
  - specific tests of whether or not each pair of IV conditions has a response pattern difference
- How many Pairwise comparisons ??
  - Formula, with k = # IV conditions

    # pairwise comparisons = [k * (k-1)] / 2
  - or just remember a few of them that are common
    - 3 groups = 3 pairwise comparisons
    - 4 groups = 6 pairwise comparisons
    - 5 groups = 10 pairwise comparisons

Pairwise Comparisons for X^2

Using the Computator, just plug in the cell frequencies for any 2x2 portion of the k-group design



Effect Size (r) for Pairwise Chi-square Comparisons

Insert frequencies for the 2x2 pairwise comparison =>		45	40
		15	10
	Chi-square =	0.388	
	p =	0.5332285	
	r =	0.059	

ChiSq->r | pr_chi>r | Outlier Bourndries | LSD-

It also calculates the effect size of the pairwise comparison, more later…

Example of pairwise analysis of a multiple IV condition design

	Tx1	Tx2	Cx
Comfortable	45	40	25
Not comfortable	15	10	20

$X^2(2) = 7.641$, $p = .034$

	Tx1	Tx2
C	45	40
~C	15	10

$X^2(1) = .388$, $p > .05$

Retain H0:

Tx1 = Tx2

	Tx1	Cx
C	45	25
~C	15	20

$X^2(1) = 4.375$, $p < .05$

Reject H0:

Tx1 > Cx

	Tx2	Cx
C	40	25
~C	10	20

$X^2(1) = 6.549$, $p < .05$

Reject H0:

Tx2 > Cx

What to do when you have a RH:

The RH: was, “In terms of the % who show improvement, immediate feedback (IF) is the best, with delayed feedback (DF) doing no better than the no feedback (NF) control.”

Determine the pairwise comparisons, how the RH applied to each …

IF > DF IF > NF DF = NF

Run the omnibus X^2 -- is there a relationship ?

	IF	DF	NF
Improve	78	40	65
Not improve	10	32	18

$X^2(2) = 23.917$, $p < .001$

Perform the pairwise X^2 analyses

	IF	DF
i	78	40
~i	10	32

$X^2(1)=22.384$, $p<.001$

Reject H0: IF > DF

	IF	NF
i	78	65
~i	10	18

$X^2(1)=3.324$, $p>.05$

Retain H0: IF = NF

	DF	NF
i	40	65
~i	32	18

$X^2(1)=9.137$, $p<.005$

Reject H0: DF < NF

Determine what part(s) of the RH were supported by the pairwise comparisons …

RH:	IF > DF	IF > NF	DF = NF
well ?	supported	not supported	not supported

We would conclude that the RH: was partially supported !

Alpha Inflation

- Increasing chance of making a Type I error the more pairwise comparisons that are conducted

Alpha correction

- adjusting the set of tests of pairwise differences to “correct for” alpha inflation
- so that the overall chance of committing a Type I error is held at 5%, no matter how many pairwise comparisons are made

There is no equivalent to HSD for X^2 follow-ups

- one approach is to use p=.01 for each pairwise comparison, reducing the alpha inflation
- Another is to “Bonferronize” p = .05 / #comps to hold the experiment-wise Type I error rate to 5%
- As with ANOVA → when you use a more conservative approach you can find a significant omnibus effect but not find anything to be significant when doing the follow-ups!

Alpha Corrected pairwise comparisons for Chi-square

Effect Size (r) for Pairwise Chi-square Comparison

Insert frequencies for the 2x2 pairwise comparison =>		45	35
		15	10
		Chi-square =	0.109
		p =	0.7408568
		r =	0.032

Bonferroni-corrected p-values for various numbers of pairwise comparisons

Number of comditions	Number of comparisons	critical p-value
2	1	0.05
	2	0.025
3	3	0.0167
	4	0.0125
	5	0.01
4	6	0.0083
5	10	0.005
6	15	0.0033

The computator also shows the critical Chi-square value for different p-values for “corrected” comparisons.

First: Compute the pariwise chi-square.

Second: Determine the p-value to use for NHST of that pairwise comparison.

If this were the comparison of two conditions from a 3-condition design…

3 conditions requires 3 comparisons, so we would use the p-value of .0167

Based on this Bonferroni-corrected p-value of .0167, we would conclude that this pariwise comparison (with a p-value of .032) is no significant!

Power Analyses for k-group designs

Important Symbols

S is the total # of participants in that pairwise comp

n = S / 2 is the # of participants in each condition of that pairwise comparison

N = n * k is the total number or participants in the study

Example

- the smallest pairwise X^2 effect size for a 3-BG study was .25
- with r = .25 and 80% power S = 120
- for each of the 2 conditions n = S / 2 = 120 / 2 = 60
- for the whole study N = n * k = 60 *3 = 180