Factorial Designs: Programmatic Research, RH: Testing & Causal Inference

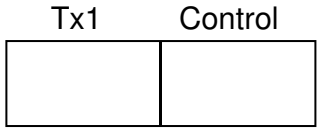
- Applications of Factorial designs in Programmatic Research
- Research Hypotheses for Factorial Designs
- Variable Role Explication in Factorial Designs & Causal Interpretation

Using Factorial Designs in Programmatic Research I

Adding a 2nd IV

Perhaps the most common application of factorial designs it so look at the separate (main) and combined (interaction) effects of two IVs

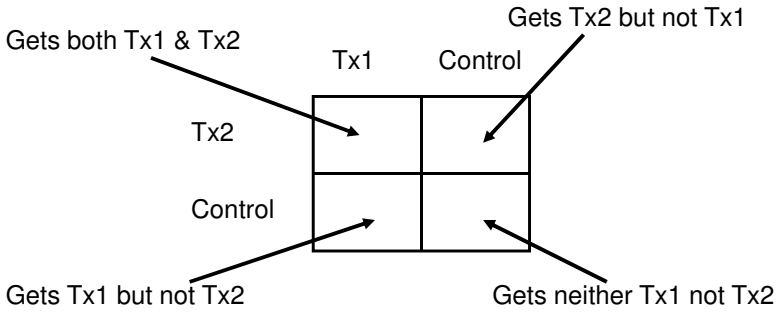
Often our research starts with a simple RH: that requires only a simple 2-group BG research design.



Keep in mind that to run this study, we made sure that none of the participants had any other treatments !

Factorial Designs – Separate (Main) and combined (interaction) effects of two treatments

At some point we are likely use Factorial designs to ask ourselves about how a 2nd Tx/IV also relates to the DV



Using Factorial Designs in Programmatic Research II

"Correcting" Bivariate Studies

Our well sampled, randomly assigned, manipulated, controlled, carefully measured, properly analyzed study showed …

Style1	Style2
40	40

… nothing !

Our well sampled, randomly assigned, manipulated, controlled, carefully measured, properly analyzed study showed …

Context1	Context2
40	40

… nothing !

Looks like neither Style nor Context is related to the DV !!!

However, when we analyzed the same data including both variables as IVs in a Factorial Design …

	Style1	Style2	
Context1	60	20	40
Context2	20	60	40
	40	40	

There are Style effects both for Context1 and Context2 – the marginal Style means are an "aggregation error"

There are Context effects both for those in Style1 & Style2 – the marginal Context means are an "aggregation error"

So, instead of the "neither variable matters" bivariate results, the multivariate result shows that both variables are conditionally related to the DV -- they interact !!!!! BOTH are important!!!

Using Factorial Designs in Programmatic Research III

Generalization across Populations, Settings & Tasks

Often our research starts with a simple RH: that requires only a simple 2-group BG research design.

Computer	Lecture

Keep in mind that to run this study, we had to make some choices/selections:

For example:
population → College Students
setting → Lecture setting
stim/task → teach Psychology

When we've found and replicated an effect, making certain selections, it is important to check whether changing those selections changes the results – by running factorials with the external validity elements as 2nd Ivs and looking for interactions!

Computer	Lecture
60	40

If there are no interactions – if the results "don't depend upon" the population, task/stimulus, setting, etc – we need to know that, so we can apply the results of the study to our theory or practice, confident in their generalizability

If there are an interactions – if the results "depend upon" the population, task/stimulus, setting, etc – we need to know that, so we can apply the "correct version" of the study to our theory or practice

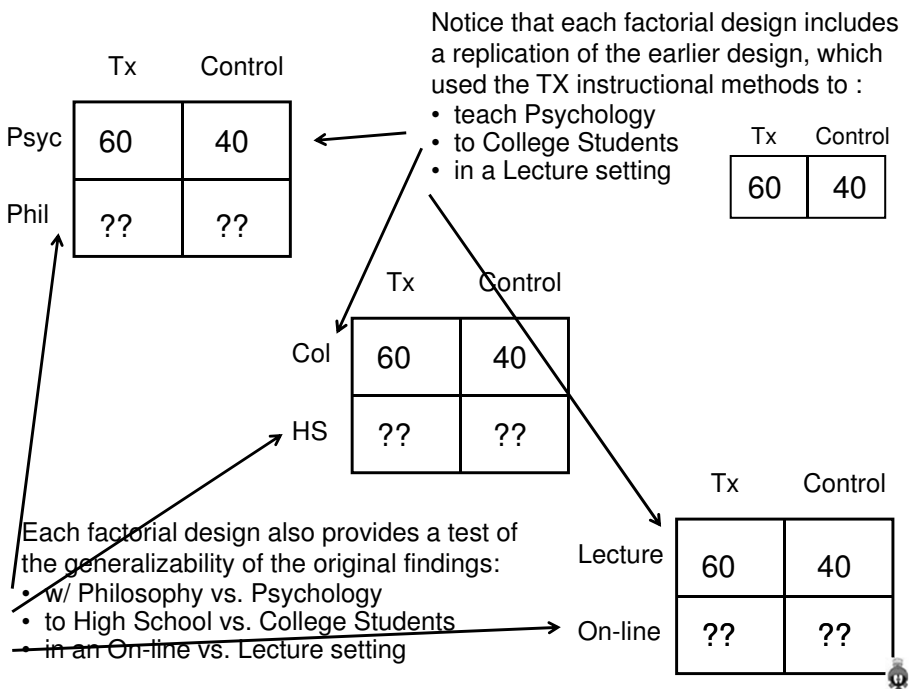
At some point we are likely use BG Factorial designs to ask ourselves how well the results will generalize to:

- other populations – college vs. high school
- other settings – lecture vs. laboratory
- other tasks/stimuli – psyc vs. philosophy

	Tx	Control
Col		
HS		

	Tx	Control
Psyc		
Phil		

	Tx	Control
Lecture		
On-line		



Using Factorial Designs in Programmatic Research IV

Do effects "depend upon" length of treatment ?

As before, often our research starts with a simple RH: that requires only a simple 2-group BG research design.

Tx_1	Tx_2
20	20

Time Course Investigations

In order to run this study we had to select ONE treatment duration (say 16 weeks):

- we assign participants to each condition
- begin treatment of the Tx groups
- treat for 16 weeks and then measured the DV

Using this simple BG design we can "not notice" some important things. A MG Factorial can help explore the time course of the Tx effects.

Tx_1	Tx_2
20	20

By using a MG design, with different lengths of Tx as the 2nd IV, we might find different patterns of data that we would give very different interpretations

	Short	Medium
Tx_1	20	40
Tx_2	20	40

	Short	Medium
Tx_1	20	40
Tx_2	20	20

	Short	Medium
Tx_1	20	40
Tx_2	20	60

	Short	Medium
Tx_1	20	0
Tx_2	20	40

Using Factorial Designs in Programmatic Research V

Evaluating Initial Equivalence when Random assignment is not possible

As before, often our research starts with a simple RH: that requires only a simple 2-group BG research design.

Tx_1	Tx_2

Initial Equivalence Investigations

In order to causally interpret the results of this study, we'd have to have initial equivalence

- but we can't always RA & manipulate the IV
- So what can we do to help interpret the post-treatment differences of the two treatments?
- Answer – compare the groups before treatment too!

By using a MG design, we can compare the groups pre-treatment and use that information to better evaluate post-treatment group differences (but can't really infer cause). For which of these would you be more comfortable conclusing that $Tx_1 > Tx_2$??

	Pre	Post
Tx_1	40	40
Tx_2	20	20

Nah – Post dif = pre dif !

	Pre	Post
Tx_1	60	40
Tx_2	20	20

Nah – Tx_1 lowered score

	Pre	Post
Tx_1	30	60
Tx_2	20	40

Maybe – more in-crease by Tx_1

	Pre	Post
Tx_1	20	40
Tx_2	20	20

As good as it gets!

Replication & Generalization in Factorial Designs

Identifying the "replication" in a factorial design

Most factorial designs are an "expansion" or an extension of an earlier, simpler design, often by adding a second IV that "makes a variable out of an earlier constant". This second IV may related to the population, setting or task/stimulus involved.

Study #1 – Graphical software

Mean failures PC = 5.7, std = 2.1

Mean failures Mac = 3.6, std = 2.1

What gives us the most direct replication? The main effect of PC vs. Mac or one of the SEs of PC vs. Mac?

Study #2

	PC	Mac
Graphical	5.9	3.6
Computing	3.1	3.8
	4.5	3.7

Did Study #2 replicate Study #1?

Replication & Generalization in Factorial Designs, cont…

Identifying the "replication" in a factorial design

Most factorial designs are an "expansion" or an extension of an earlier, simpler design, often by adding a second IV that "makes a variable out of an earlier constant". This second IV may related to the population, setting or task/stimulus involved.

Study #1 – Mix of Networked & Stand-alone computers

Mean failures PC = 5.7, std = 2.1

Mean failures Mac = 3.6, std = 2.1

What gives us the most direct replication? The main effect of PC vs. Mac or one of the SEs of PC vs. Mac?

Study #2

	PC	Mac
Networked	8.9	1.6
Stand-alone	3.1	5.8
	6.0	3.7

Did Study #2 replicate Study #1?

RH: for Factorial Designs

Research hypotheses for factorial designs may include

- RH: for main effects
  - involve the effects of one IV, while ignoring the other IV
  - tested by comparing the appropriate marginal means
- RH: for interactions
  - usually expressed as differences between hypothesized results for a set of simple effects
  - tested by comparing the results of the appropriate set of simple effects
  - That's the hard part -- determining which set of simple effects gives the most direct test of the interaction RH:

#1 Sometimes the Interaction RH: is explicitly stated

- when that happens, one set of SEs will provide a direct test of the RH: (the other won't)

Here's an example:

Easy tasks will be performed equally well using paper or computer presentation, however, hard tasks will be performed better using computer presentation than paper.

Task Diff.	Presentation: Comp	Presentation: Paper
Easy	=	
Hard	>	

This is most directly tested by inspecting the simple effect of paper vs. computer presentation for easy tasks, and comparing it to the simple effect of paper vs. computer for hard tasks.

Your Turn...

Snapping turtles will prefer Crickets, while Painted turtles will have no preference?

Species	Food offered: Crickets	Food offered: Carrion
Snapping	>	
Painted	=	

SE Food @ Species

Judges will rate confessions as more convincing than do Lawyers, however, Lawyers will rate witnesses as more convincing than do Judges.

Rater	Type of Evidence: Confession	Type of Evidence: Witness
Judge	V	Λ
Lawyer		

SE Rater @ Evidence

#2 Sometimes the set of SEs to use is "inferred" ...

Often one of the IVs in the study was used in previous research, and the other is "new".

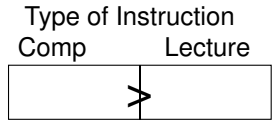
- In this case, we will usually examine the simple effect of the "old" variable, at each level of the "new" variable
- this approach gives us a clear picture of the replication and generalization of the "old" IV's effect.

e.g., Previously I demonstrated that computer presentations lead to better learning of statistical designs than does using a conventional lecture. I would like to know if the same is true for teaching writing.

Let's take this "apart" to determine which set of SEs to use to examine the pattern of the interaction...

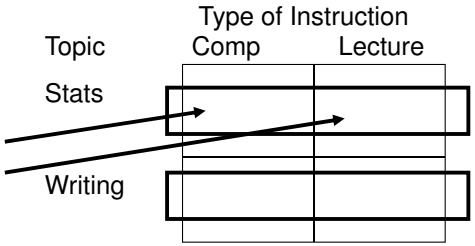
Previously I demonstrated that computer presentations lead to better learning of statistical designs than does using a conventional lecture. I would like to know if the same is true for teaching writing.

Here's the design and result of the earlier study about learning stats.



Here's the design of the study being planned.

What cells are a replication of the earlier study ?

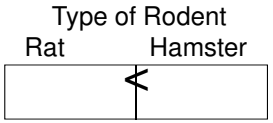


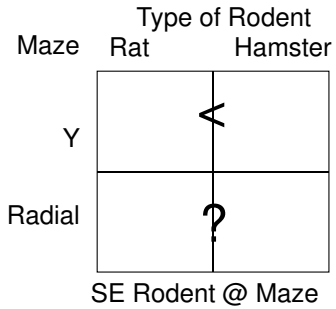
So, which set of SEs will allow us to check if we got the replication, and then go on to see of we get the same results with the new topic ?

Yep, SE of Type of Instruction, for each Topic ...

Your turn ..

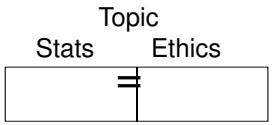
I have previously demonstrated that rats learn Y-mazes faster than do hamsters. I wonder if the same is true for radial mazes ?



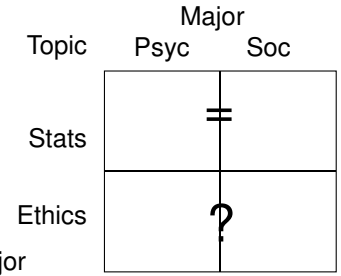


SE Rodent @ Maze

I've discovered that Psyc majors learn statistics & Ethics about equally well. My next research project will also look at how well Sociology majors learn these topics.



SE Topic @ Major

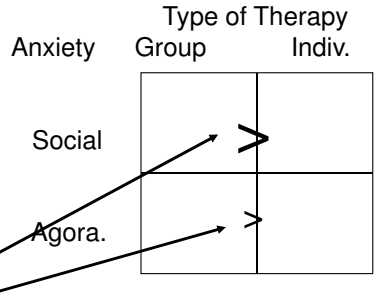


#3 Sometimes the RH: about the interaction and one about the main effects are "combined"

- this is particularly likely when the expected interaction pattern is of the > vs. > type (the most common pattern)

Here's an example…

Group therapy tends to work better than individual therapy, although this effect is larger for patients with social anxiety than with agoraphobia.



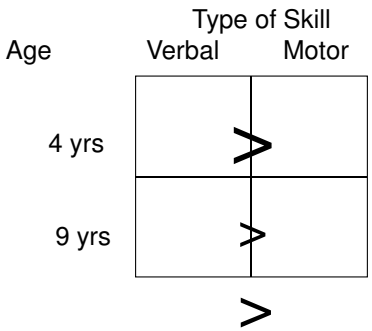
Int. RH:

Main effect RH:

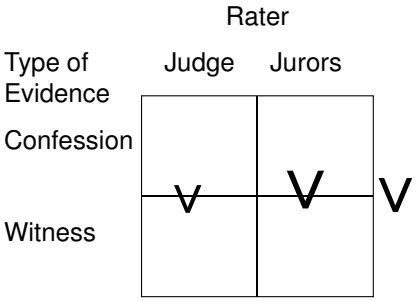
So, we would examine the interaction by looking at the SEs of Type of Therapy for each type of Anxiety.

Your Turn…

Young children have better verbal skills than motor skills, however the difference gets smaller with age (DV = skill score)



SE Skill @ Age



SE Evidence @ Rater

Confession is considered more convincing than eyewitness testimony. This preference is stronger for jurors than judges.

(DV = convincingness rating)