Empirical Research Methods in Information Science

IS 4800 / CS 6350

Lecture 23

Multi-factor Between-Subjects Designs

1

Experiment

A = Read article about exercise.
B = Read article about F&V.
C = Motivational F&V intervention.

Outcomes:
F&V Change
PA Change
Meditation Change

Hypotheses?

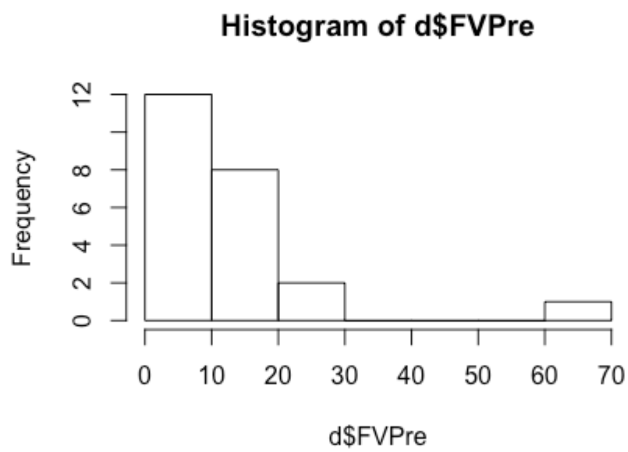
2

```
> library(readxl)
> d <- read_excel("Documents/courses/2018-Spring-
IS4800-methods/Lectures/L19/In-Class-Experiment/
anondata.xlsx")
> View(d)

> table(d$Condition)

A B C
7 8 8

> table(d$Sex,d$Condition)

   A B C
 F 3 3 5
 M 3 3 3
```

3

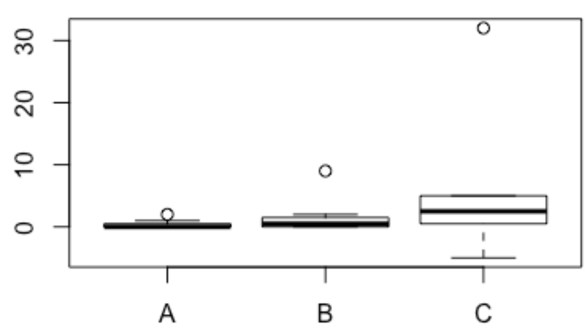
```
> hist(d$FVPre)
> hist(d$FVPost)
> hist(d$ExercisePre)
> hist(d$ExercisePost)
> hist(d$MeditationPost)
> hist(d$MeditationPre)
> hist(d$FVChange)
> hist(d$ExerciseChange)
> hist(d$MeditationChange)

> d$Condition<-factor(d$Condition)
> boxplot(d$FVChange~d$Condition)
> boxplot(d$ExerciseChange~d$Condition)
> boxplot(d$MeditationChange~d$Condition)
```

4



5

boxplot(d$FVChange~d$Condition)



6

```
> # BASELINE ANALYSIS

> summary(d$FVPre[d$Condition=='A'])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.00    6.00    8.00   11.43   14.50   28.00

> IQR(d$FVPre[d$Condition=='A'])
[1] 8.5

> # ETC
```

7

```
> # BASELINE ANALYSIS

> kruskal.test(d$FVPre~d$Condition)
      Kruskal-Wallis rank sum test
data:  d$FVPre by d$Condition
Kruskal-Wallis chi-squared = 1.5819, df = 2, p-
value
= 0.4534

> chisq.test(table(d$Condition,d$Sex))
      Pearson's Chi-squared test
data:  table(d$Condition, d$Sex)
X-squared = 0.30303, df = 2, p-value = 0.8594

> # ETC
```

8

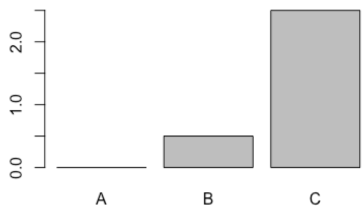
```
> # OUTCOME DESCRIPTIVES

> summary(d$FVChange[d$Condition=='A'])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.4286  0.5000  2.0000

> IQR(d$FVChange[d$Condition=='A'])
[1] 0.5

> # ETC
```

9

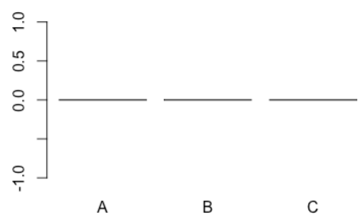
```
> # OUTCOME INFERENTIALS

> kruskal.test(d$FVChange~d$Condition)
      …
      Kruskal-Wallis … p-value = 0.1813

> kruskal.test(d$ExerciseChange~d$Condition)
       …
      Kruskal-Wallis … p-value = 0.2933

> kruskal.test(d$MeditationChange~d$Condition)
     …
     Kruskal-Wallis … p-value = 1
```

10

FVChange
(Medians)



ExerciseChange
MeditationChange
(Medians)



All zero!

Types of Study Designs

- Qualitative
  - Ethnography
- Quantitative
  - Descriptive
  - Correlational
  - Demonstrative
  - Experimental
    - Between-subjects
      - Single factor, two-level
      - Single factor, N-level (for N>2)
      - Two factor, two-level
    - Within-subjects
      - Single factor, two-level

Accompanying Statistics

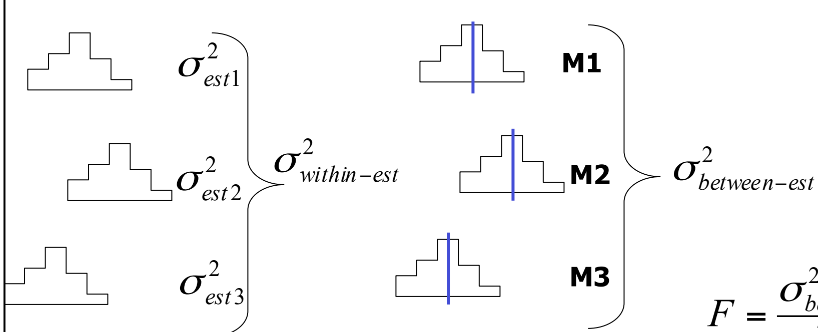
- Between-subjects
  - Single factor, 2-level
    - t-test for independent means
  - Single factor, N-level (for N>2)
    - One-way Analysis of Variance (ANOVA)
  - Two factor, two-level (or more!)
    - Factorial Analysis of Variance
    - AKA N-way Analysis of Variance (for N IVs)
    - AKA N-factor ANOVA
- Within-subjects
  - Single factor, two level
    - Paired sample t-test
  - Repeated-measures ANOVA (not discussed)
    - AKA within-subjects ANOVA

13

One-Way ANOVA – Assuming Null Hypothesis is True…

Within-Group Estimate Of Population Variance

Between-Group Estimate Of Population Variance



$$F = \frac{\sigma^2_{between-est}}{\sigma^2_{within-est}}$$

16

Post hoc analysis

- Once the ANOVA indicates there is a significant difference ("omnibus" test), you do either
  - Planned comparisons, or
  - Post hoc tests
- to determine which pairwise comparisons are significantly different

- There are many post hoc tests (B&A 446)
  - Sheffe, Dunnett, Tukey, etc.
    - Very conservative

18

Factorial ANOVA Designs

- Two or more nominal independent variables, each with two or more levels, and a interval or ratio dependent variable.
- Factorial ANOVA teases apart the contribution of each IV separately, as well as *every combination* of IVs.
- Terminology
  - For N IVs, aka "N-way" ANOVA
  - For L_i levels per factor, "L_1 by L_2 by L_3… ANOVA"
- Most common: 2 by 2 ANOVA

21

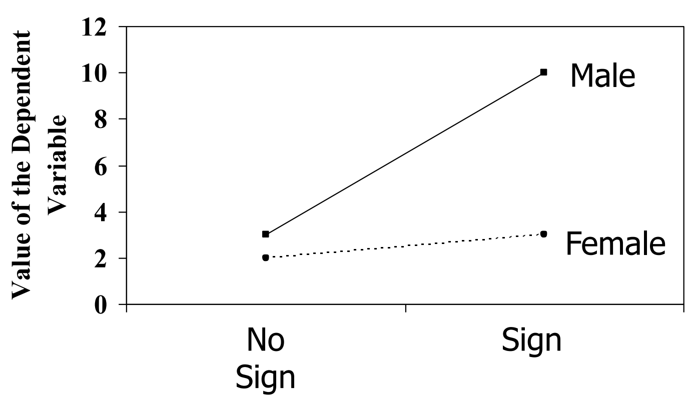
Factorial Designs

- Two effects of IVs on DV can be assessed
  - A *MAIN EFFECT* of each independent variable
    - The separate effect of each independent variable
    - Analogous to separate experiments involving those variables
  - An *INTERACTION* between independent variables
    - When the effect of one independent variable changes over levels of a second
    - Also – when the effect of one variable depends on the level of the other variable.

22

Example of An Interaction - Student Center Sign - 2 Genders x 2 Sign Conditions

Is a Sign better than no Sign?



Two-way ANOVA in R

Book	Instructor	Knowledge
1	1	1.5
2	2	2
1	1	2
1	1	0.5
2	2	2
2	2	2
1	1	2
2	2	2
1	1	1
1	2	6.5

24

Two-way ANOVA in R

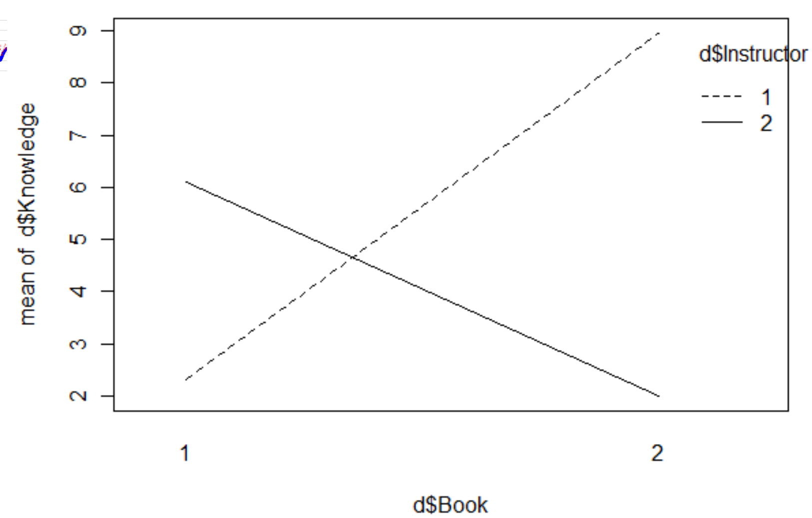
```
> out <- aov(Knowledge ~ Book * Instructor,  data=d)
>
> summary(out)
                Df  Sum Sq Mean Sq F value    Pr(>F)
Book             1   1.477   1.477   1.161    0.2941
Instructor       1   0.022   0.022   0.017    0.8975
Book:Instructor  1 123.450 123.450  97.032 4.073e-09 ***
Residuals       20  25.445   1.272
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
```

25

Interaction Plot

```
# x.factor = X-axis
# trace.factor = separate plot lines factor
# response = response variable

> interaction.plot(x.factor=d$Book,
         trace.factor=d$Instructor,
         response=d$Knowledge)
```

26



27

Degrees of Freedom

- df for between-group variance estimates for main effects
  - Number of levels – 1
- df for between-group variance estimates for interaction effect
  - Total num cells – df for both main effects – 1
  - e.g. For 2x2, it is 4 – (1+1) – 1 = 1
- df for within-group variance estimate
  - Sum of df for each cell = N – num cells
- Report: "F(*bet-group, within-group*)=*F, Sig.*"

28

Publication format

```
> summary(out)
                   Df  Sum Sq Mean Sq F value    Pr(>F)    
Book                1   1.477   1.477   1.161    0.2941    
Instructor          1   0.022   0.022   0.017    0.8975    
Book:Instructor     1 123.450 123.450  97.032 4.073e-09 ***
Residuals          20  25.445   1.272                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F(1,20)=97.0, p<.05.

There is a significant interaction effect of Book and Instructor on Knowledge gain.

29

Reporting rule

- IF you have a significant interaction
- THEN
  - In general, only report interaction, not any main effects, even if significant.
  - However, you must inspect the means to determine if main effects make sense to report
- Interaction => you cannot interpret the effect of one factor without the other (in general)

30

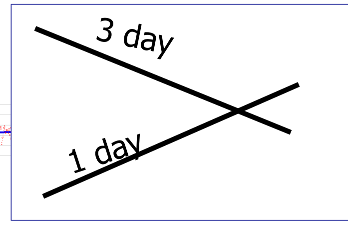
Results?

	Sig.
TrainingDays	0.34
Trainer	0.12
TrainingDays * Trainer	0.41

→ n.s.

31

Results?

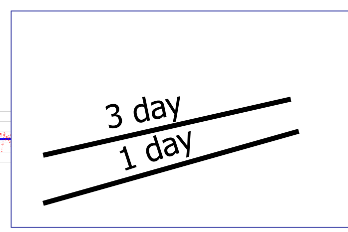


	Sig.
TrainingDays	0.34
Trainer	0.12
TrainingDays * Trainer	0.02

Significant interaction between TrainingDays And Trainer, $F(1,22)=.584$, $p<.05$

32

Results?

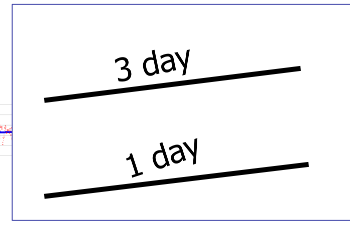


	Sig.
TrainingDays	0.34
Trainer	0.02
TrainingDays * Trainer	0.41

Main effect of Trainer, $F(1,22)=3.9$, $p<.05$

33

Results?

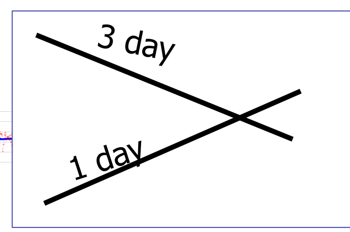


	Sig.
TrainingDays	0.02
Trainer	0.34
TrainingDays * Trainer	0.41

Main effect of TrainingDays, $F(1,22)=7.20$, $p<.05$

34

Results?



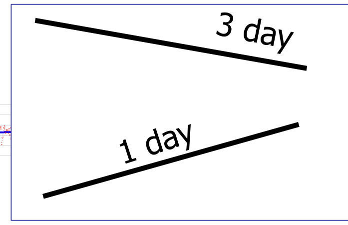
	Sig.
TrainingDays	0.04
Trainer	0.12
TrainingDays * Trainer	0.01

Significant interaction between TrainingDays and Trainer, $F(1,22)=.584$, $p<.05$

Do not report TrainingDays as significant

35

Results?

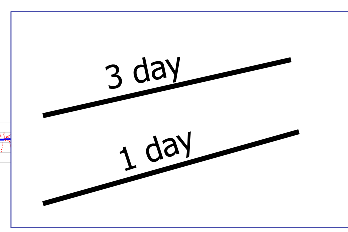


	Sig.
TrainingDays	0.04
Trainer	0.12
TrainingDays * Trainer	0.01

Significant interaction between TrainingDays and Trainer, $F(2,22)=.584$, $p<.05$
Also a main effect of TrainingDays, $F(2,22)=.684$, $p<.05$, since learning is always greater for 3 days vs. 1 day, regardless of who the trainer is

36

Results?

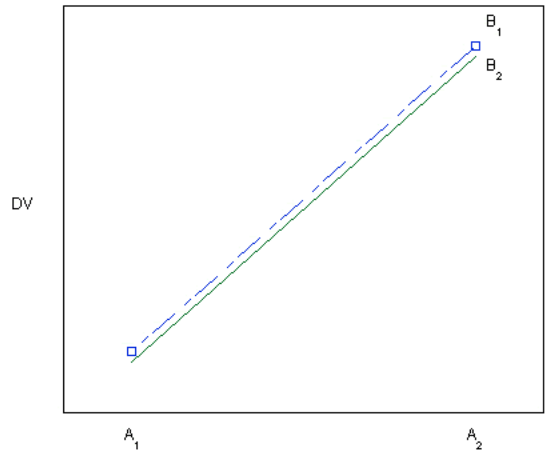


	Sig.
TrainingDays	0.04
Trainer	0.02
TrainingDays * Trainer	0.41

Main effects for both TrainingDays, $F(2,22)=7.20$, $p<.05$, and Trainer, $F(1,22)=.001$, $p<.05$

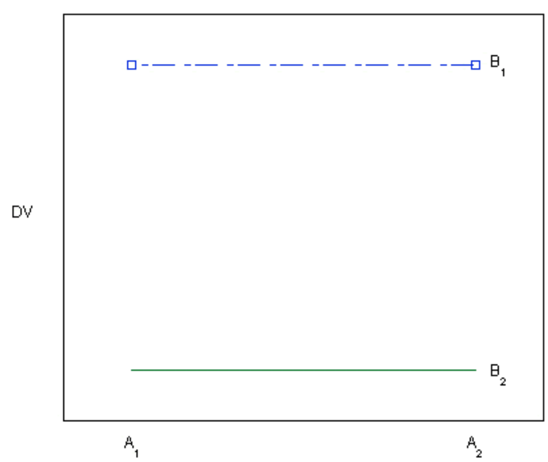
37

Possible interpretation?



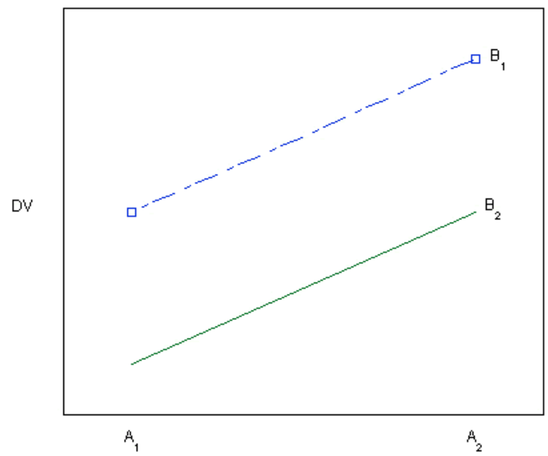
http://courses.washington.edu/smartpsy/interactions.htm# 38

Possible interpretation?



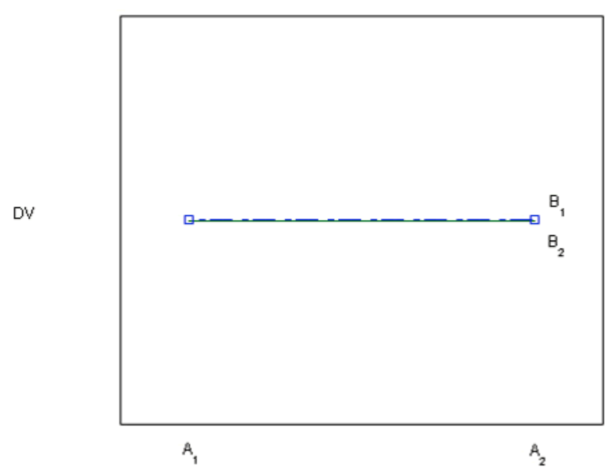
http://courses.washington.edu/smartpsy/interactions.htm# 39

Possible interpretation?



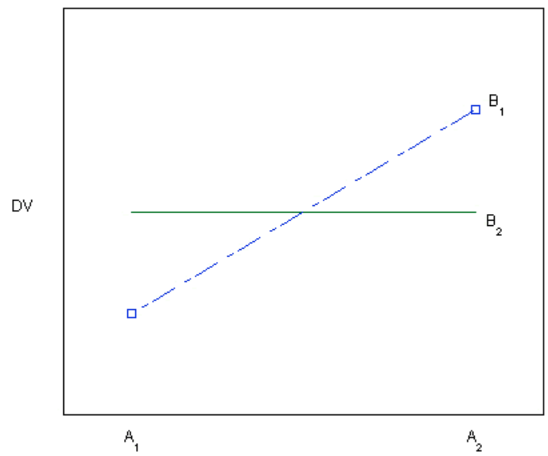
http://courses.washington.edu/smartpsy/interactions.htm# 40

Possible interpretation?



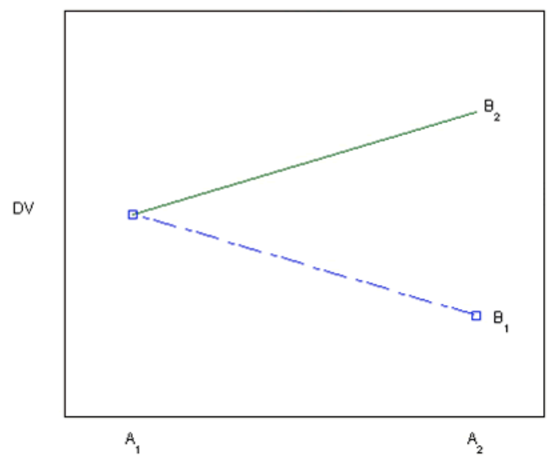
http://courses.washington.edu/smartpsy/interactions.htm# 41

Possible interpretation?



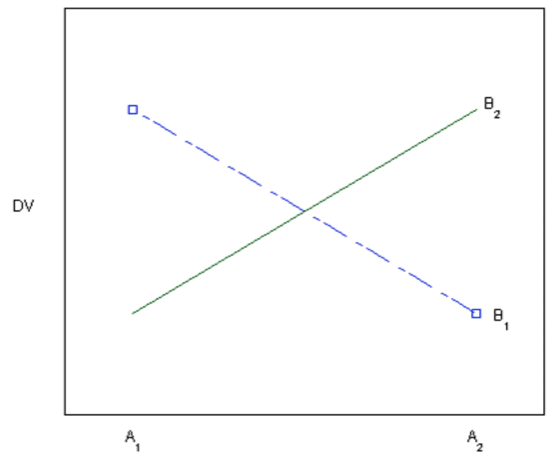
http://courses.washington.edu/smartpsy/interactions.htm# 42

Possible interpretation?



http://courses.washington.edu/smartpsy/interactions.htm# 43

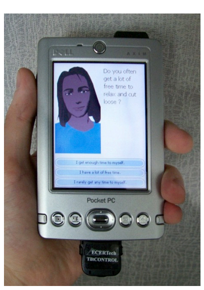
Possible interpretation?



http://courses.washington.edu/smartpsy/interactions.htm# 44

"Factorial Design"

- Not all cells in your design need to be tested
  - But if they are, it is a "full factorial design", and you do a "full factorial ANOVA"



	Real-Time	Retrospective
Agent	√	√
Text	√	X

45

Higher-Order Factorial Designs

- More than two independent variables are included in a *higher-order factorial design*
  - As factors are added, the complexity of the experimental design increases
    - The number of possible main effects and interactions increases
    - The number of subjects required increases exponentially
    - The volume of materials and amount of time needed to complete the experiment increases exponentially
    - The difficulty of interpreting the results can also greatly increase.

46

ANOVA Assumptions

- Population in each cell is normal
- Populations have equal variances across cells

47

Group Exercise

- For each problem, write
  1. Kind of study design
  2. Kind of analysis
  3. Research & ~~Null hypotheses~~ (~~Means~~ & English)
  4. Test criteria
  5. Plot results
  6. Test results
     - English & Publication format (requires df)
  7. Implications

ANOVA effect size

- There are several.
- Most common: Eta squared (η^2)
  - In R:

```
> library(lsr)
> etaSquared(aovResult)
```

  - The variance explained by one IV after excluding variance explained by other IVs
  - Cohen: 0.01 = small, 0.06 = medium, 0.14 = large
  - Roughly: the % variance explained by one IV

Power Analysis & Multi-factorial designs

- 'N' computed for your criteria for a between-subjects design is for *each cell* of your experimental design
- A two-factor x two-level design has *four cells*
- B&A: Need at least 5 Ss per cell
- But usually need much more.

50

Power Analysis for Multi-Factor ANOVA

Table 10-16 Approximate Number of Participants Needed in Each Cell (Assuming Equal Sample Sizes) for 80% Power for Studies Using a 2 × 2 or 2 × 3 Analysis of Variance, Testing Hypotheses at the .05 Significance Level

	Effect Size		
	Small ($R^2 = .01$)	Medium ($R^2 = .06$)	Large ($R^2 = .14$)
2 × 2: All effects	197	33	14
2 × 3: Two-level main effect	132	22	9
Three-level main effect and interaction	162	27	11

- Example: medium effect size, 2x2, for all effects, requires 33x4 = 132 Ss!

51

Quasi-Independent Variable

- A *quasi-independent variable* is a correlational variable (e.g., gender) that is treated like an experimental variable
- Resulting design looks like a factorial experimental design
- The quasi-independent variable must not be interpreted as <u>causing</u> changes in the dependent variable

- If one or more of your IVs a quasi-IV, then your design is a quasi-experimental design.

52

Rickenberg & Reeves

53

Design Examples

- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 3 games leads to greatest engagement, and whether there are gender differences. You randomly assign participants to play one of the three games for 30 minutes and record their gender. After this you let them continue playing as long as they want (noting the time), then send them home.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 2 games leads to greatest satisfaction. You randomly assign participants to play RockBand or GuitarHero, ask them to rate satisfaction on a scale from 1 to 10, then send them home.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of two games, played on two different consoles, leads to greatest satisfaction. You randomly assign participants to play RockBand or GuitarHero on either Wii or Xbox and then ask them to fill out a SUS questionnaire and send them home.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 2 games leads to greatest engagement. You randomly assign participants to play RockBand or GuitarHero and keep track of how long they play. When they are done you let them play the other game for as long as they want and keep track of the time.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 2 games leads to greatest satisfaction. You go to a community center during an after school program, where you know they have PacMan and DonkeyKong on the computers. You wait until a kid plays one of these games, then ask them to fill out a 12-item composite measure of satisfaction, before scaring them away.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 3 games leads to greatest satisfaction. You randomly give participants BioShock, StarCraftII or TombRaider, ask them to go home and play for a week, then fill out a 12-item composite measure of satisfaction (then they are done).
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Homework

- Project T2!
- Next next class:
  - Project T2 presentations
  - Non-parametric statistics (B&A Ch 14, 458-466).

61