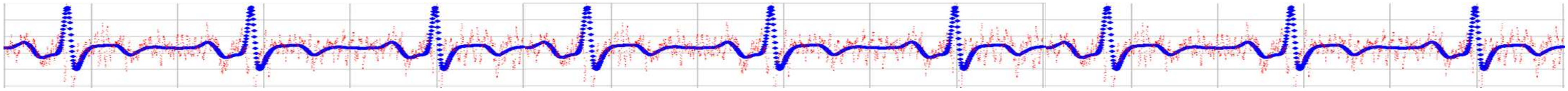
Empirical Research Methods in Information Science

IS 4800 / CS6350



Lecture 22

Outline

- Finish one-way ANOVA
- Two-way (factorial) ANOVA

- Work in teams for T3 – Experimental!
  - Exercise: Checklist

Types of Study Designs

- Qualitative
  - Ethnography
- Quantitative
  - Descriptive
  - Correlational
  - Demonstrative
  - Experimental
    - Between-subjects
      - Single factor, two-level
    - Within-subjects
      - Single factor, two-level

Factor = IV

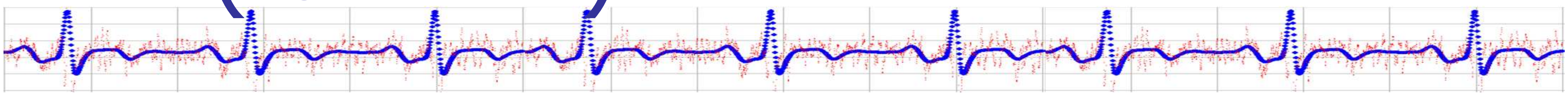
Levels = different values of the factor

Review: Basic Logic of ANOVA

- Null hypothesis
  - Means of all groups are equal.
- Test: do the means differ more than expected given the null hypothesis?

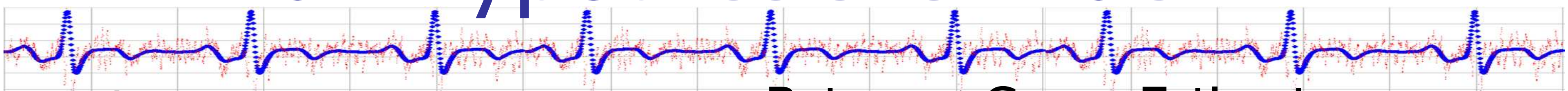
- Terminology
  - Group = Condition = Cell

ANOVA: Single factor, N-level (for N>2)



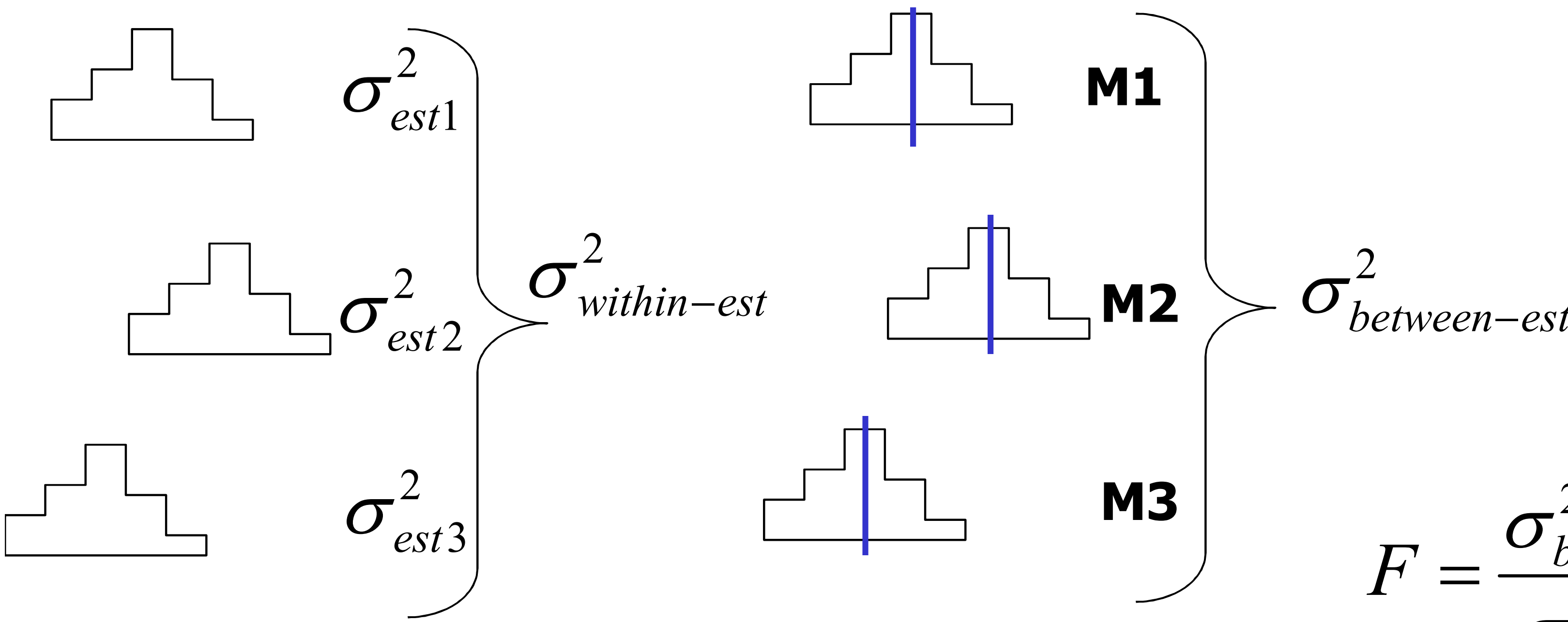
- The *Analysis of Variance* is used when you have more than two groups in an experiment
  - The *F-ratio* is the statistic computed in an Analysis of Variance and is compared to critical values of *F*
  - A significant overall *F* may require further planned or unplanned (*post hoc*) follow-up analyses
  - The analysis of variance may be used with unequal sample size (weighted or unweighted means analysis)

One-Way ANOVA – Assuming Null Hypothesis is True…



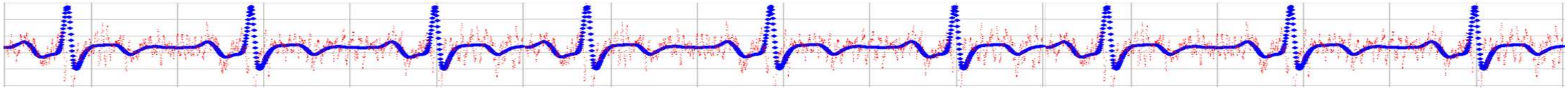
Within-Group Estimate Of Population Variance

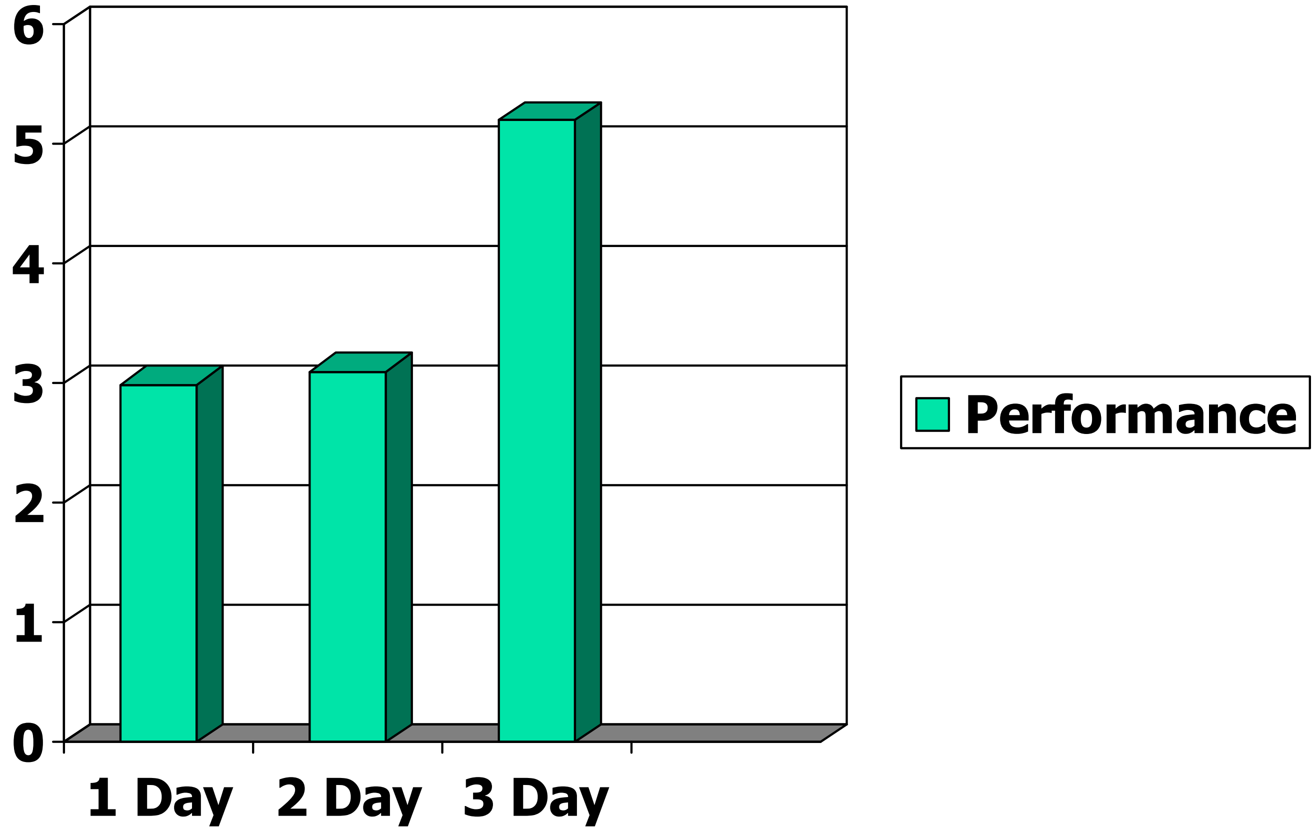
Between-Group Estimate Of Population Variance



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

Data





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

- Many post hoc tests (B&A 446) (generally, making testing more conservative)

Cohen's conventions: .01 small, .07 medium, and .14 large

Effect size

- In *t* test, took difference between two means and divide by standard deviation
- But now >2 means
- Instead, use proportion of variance accounted for (R^2)

$$R^2 = \frac{(S^2{}_{Between})(df_{Between})}{(S^2{}_{Between})(df_{Between}) + (S^2{}_{Within})(df_{Within})}$$

$$R^2 = \frac{(F)(df_{Between})}{(F)(df_{Between}) + df_{Within}}$$

Power table

Table 9-9 Approximate Power for Studies Using the Analysis of Variance Testing Hypotheses at the .05 Significance Level

Participants per Group (n)	Effect Size: Small (R^2 = .01)	Medium	Large
40	.17	.81	*
50	.21	.90	*
100	.40	*	*

*Nearly 1.

Table 9-10 Approximate Number of Participants Needed in Each Group (Assuming Equal Sample Sizes) for 80% Power for the One-Way Analysis of Variance Testing Hypotheses at the .05 Significance Level

	Effect Size: Small ($R^2 = .01$)	Effect Size: Medium ($R^2 = .06$)	Effect Size: Large ($R^2 = .14$)
Three groups ($df_{Between} = 2$)	322	52	21
Four groups ($df_{Between} = 3$)	274	45	18
Five groups ($df_{Between} = 4$)	240	39	16

Tension

- Why not ONLY do planned contrast (comparison) tests vs. overall (omnibus) F test?
  - Some argue diffuse test is not useful and should be abandoned
  - If so, what is lost?

Example Paper

- Thank you – I did not see that: In-car, speech-based information systems for older adults.
- Critique?

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

What if you have two IVs?

Measure income:

- Education: HighSchool / College
- Age: Younger / Older

Options?

- Two experiments
- Factorial: All three together (save resources, but also, see if interactions!)

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

Terminology

- Two-factor design with
  - Two levels Factor A
  - Three levels Factor B

  2x3 ("two by three") factorial design

- Three-factor design w/ 3 levels each factor

  3x3x3 ("three by three by three") factorial design

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.

Fear appeal example

Want to separately assess impact of each IV on the DV (recall of details): the **main effect**

Also want to asses whether effect of on IV changes across the levels of the other IV: **interaction**

		Fear type	
		Physical	Social
Fear intensity	Low	2.19	2.41
	High	3.85	3.02

Fear appeal example

Main effect

Difference (1.11) is main effect of fear intensity

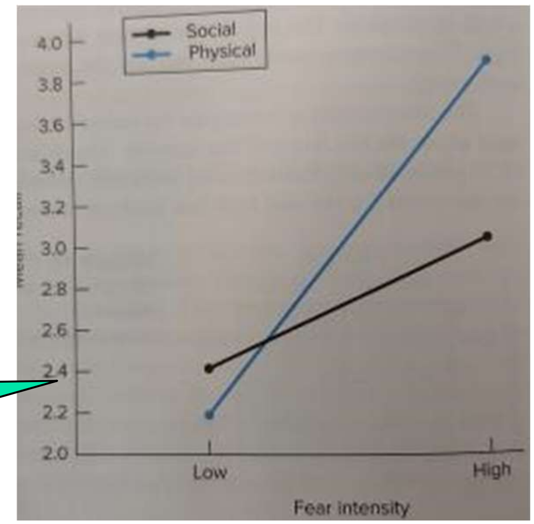
		Fear type		
		Physical	Social	
Fear intensity	Low	2.19	2.41	2.30
	High	3.85	3.02	3.41
		3.02	2.72	

Marginal means

Difference (.30) is main effect of fear type

Fear appeal example

Interaction: Type of fear impacts recall when intensity high, but not when intensity low



Simple main effect of fear intensity

Look for non-parallel lines to indicate a *possible* interation

		Fear type	
		Physical	Social
Fear intensity	Low	2.19	2.41
	High	3.85	3.02

Understanding interactions

Result A — Interaction + 2 main effects

Age	High School	College	Overall
Younger	40	40	40
Older	40	60	50
Overall	40	50	

Result B — Interaction + no main effects

Age	High School	College	Overall
Younger	60	40	50
Older	40	60	50
Overall	50	50	

Result C — No interaction but two main effects

Age	High School	College	Overall
Younger	20	60	40
Older	40	80	60
Overall	30	70	

Result D — No interaction + 1 main effect (age)

Age	High School	College	Overall
Younger	20	20	20
Older	120	120	120
Overall	70	70	

Result E — Interaction + 2 main effects

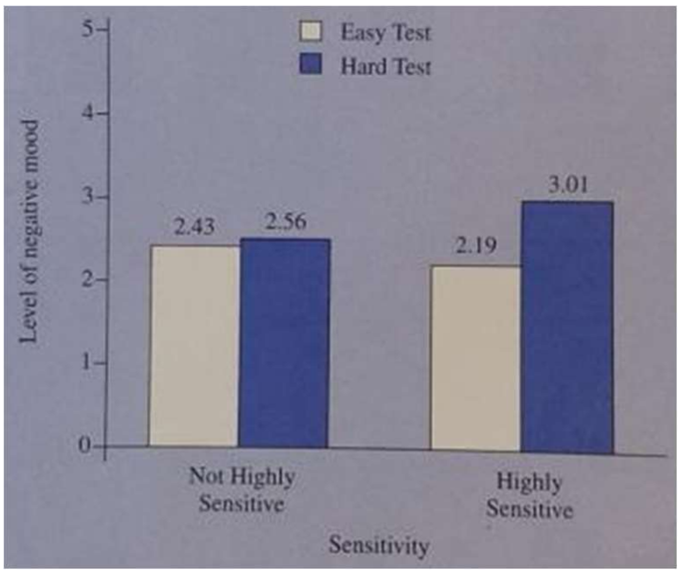
Age	High School	College	Overall
Younger	40	60	50
Older	40	80	60
Overall	40	70	

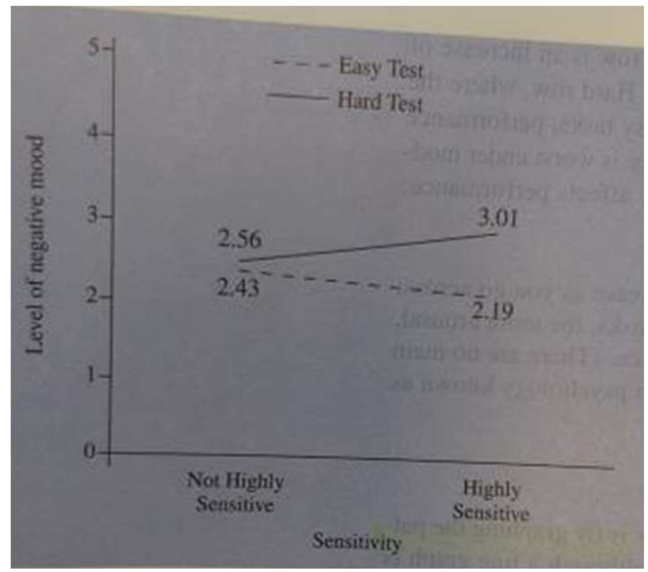
Result F — Interaction + 2 main effects

Age	High School	College	Overall
Younger	30	45	37.5
Older	35	60	47.5
Overall	32.5	52.5	

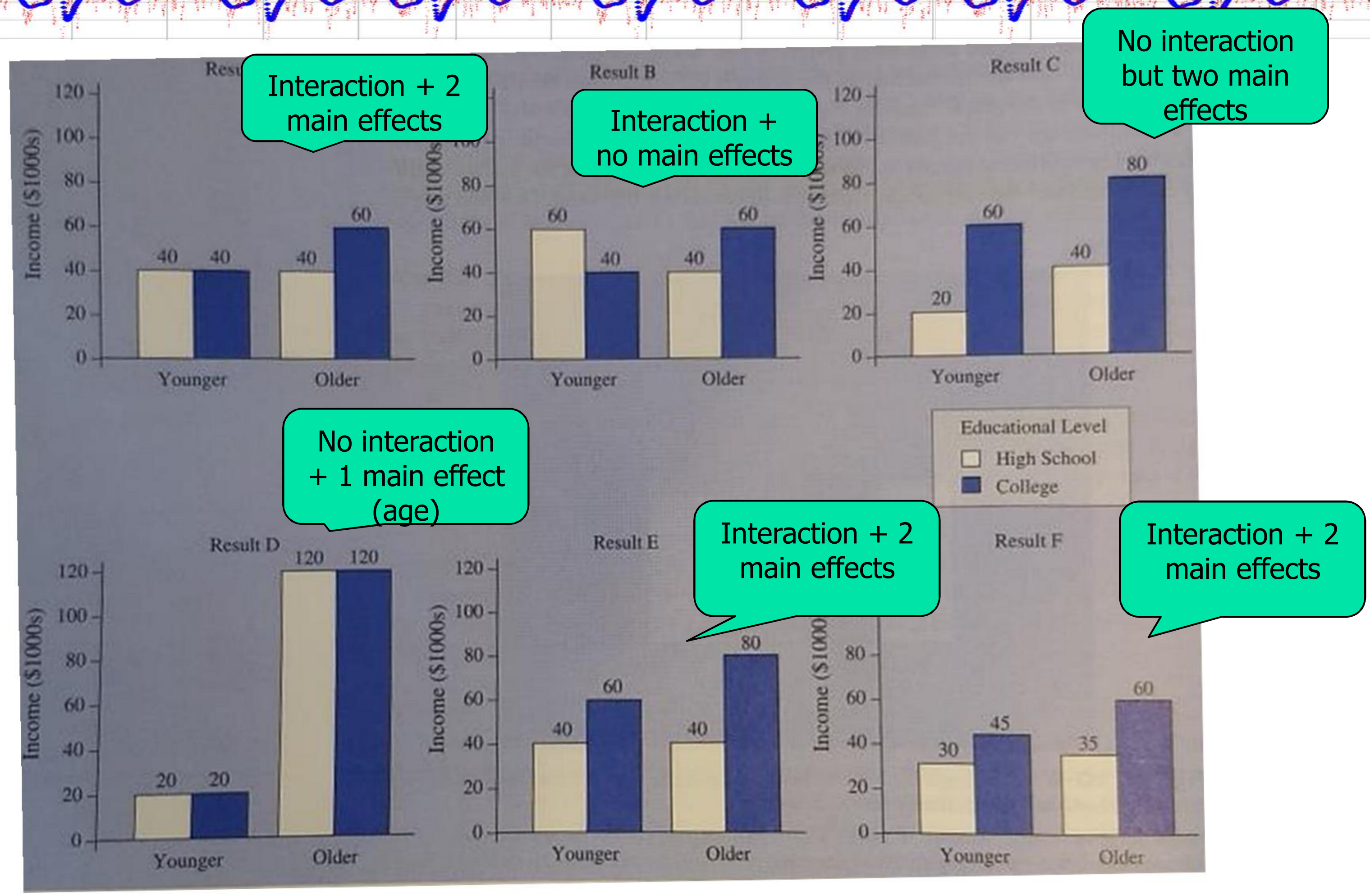
Note: Assuming statistical significance

Graphing interaction





Graphing interaction



Understanding interactions

Result A

Task	Low	Moderate	High	Overall
	Manipulated Arousal			
Easy	10	10	10	10
Hard	20	20	20	20
Overall	15	15	15	

Result B

Task	Low	Moderate	High	Overall
	Manipulated Arousal			
Easy	10	20	30	20
Hard	10	20	30	20
Overall	10	20	30	

Result C

Task	Low	Moderate	High	Overall
	Manipulated Arousal			
Easy	10	20	30	20
Hard	20	30	40	30
Overall	15	25	35	

Result D

Task	Low	Moderate	High	Overall
Easy	10	20	30	20
Hard	10	20	60	30
Overall	10	20	45	

Result E

Task	Low	Moderate	High	Overall
Easy	10	20	10	13.3
Hard	20	10	20	16.7
Overall	15	15	15	

Result F

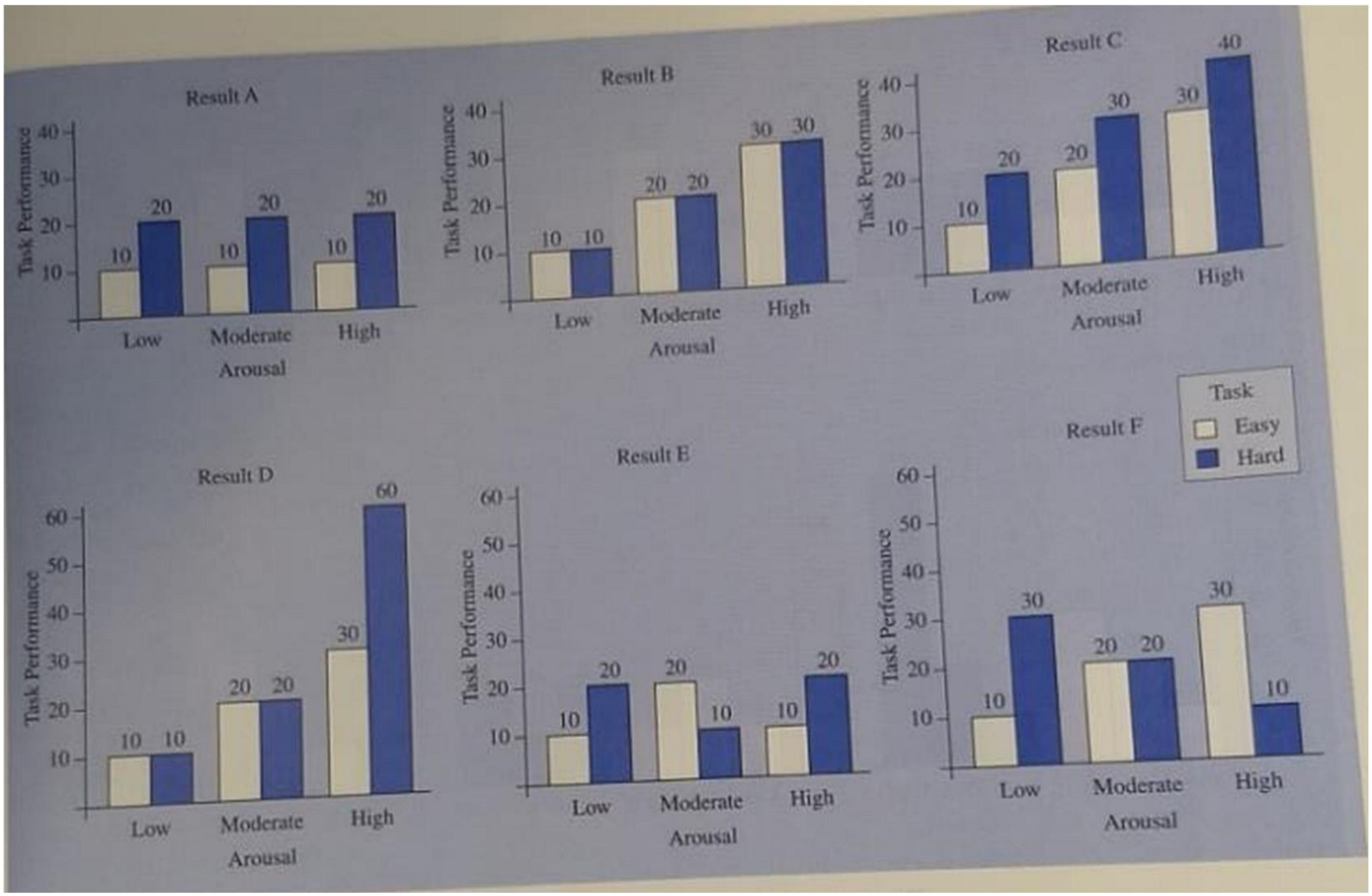
Task	Low	Moderate	High	Overall
Easy	10	20	30	20
Hard	30	20	10	20
Overall	20	20	20	

Careful! Meaning of main effect is tied up in interaction. Here, what drives the effect in difficulty?

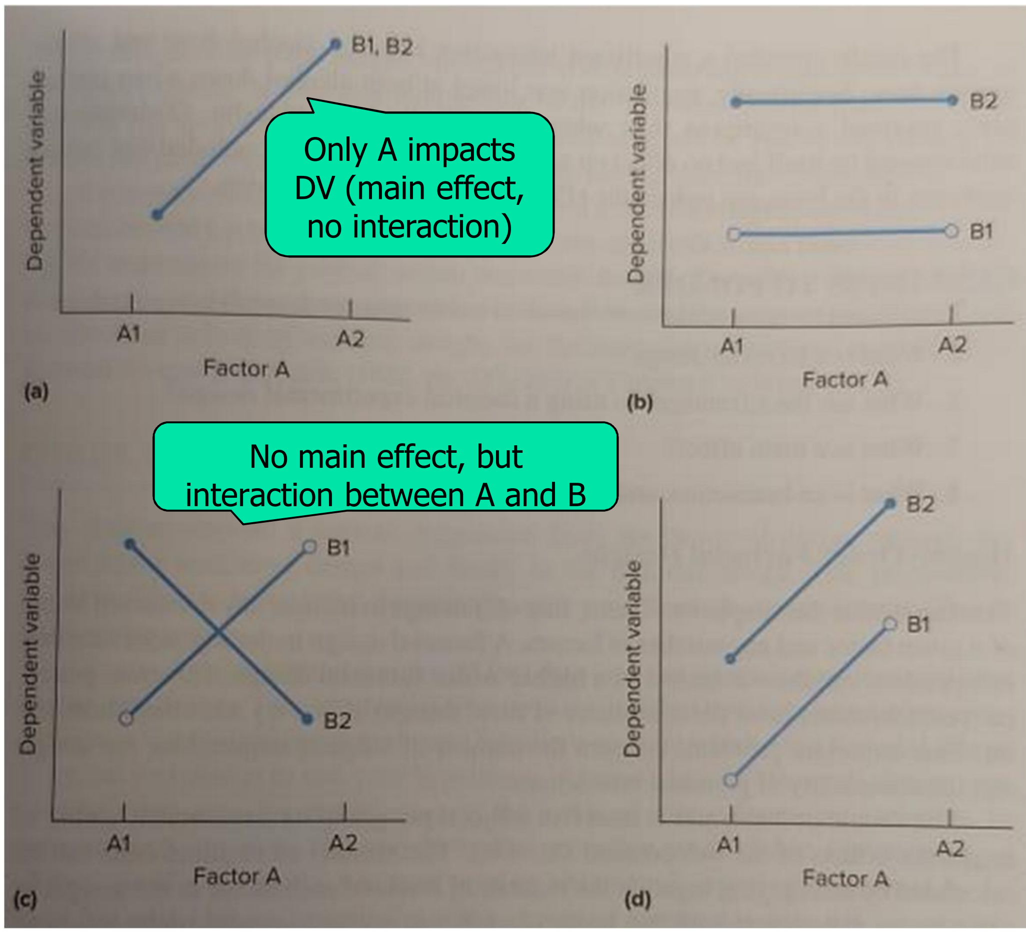
But main effect of arousal holds up across difficulty

Misleading to talk about hard vs. easy task without talking about impact of arousal.

Graphing interactions



Interpreting ANOVA graphs



Only A impacts DV (main effect, no interaction)

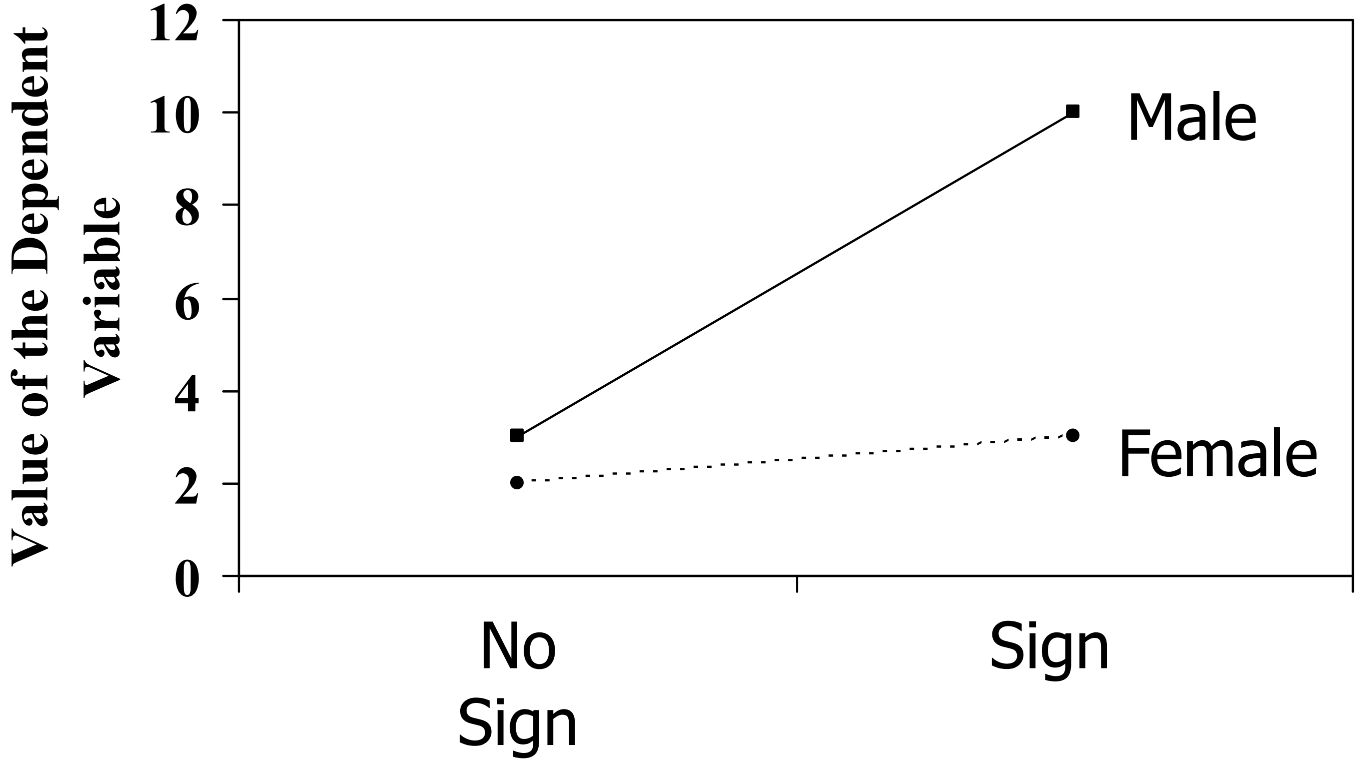
Only B impacts DV (main effect, no interaction)

No main effect, but interaction between A and B

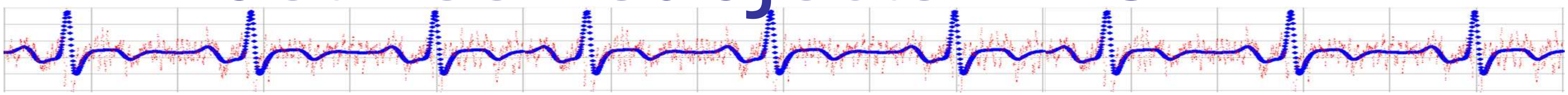
Both A and B main effect, but no interaction

Example of An Interaction - Student Center Sign - 2 Genders x 2 Sign Conditions (on satisfaction)

Is a Sign better than no Sign?



Statistical analysis: two-factor, between subjects ANOVA



- More complex than one-way because must assess statistical significance of main affect + interaction
- In two-way, three F ratios:
  - Grouping variable across columns (column main effect)
  - Grouping variable across rows (rows main effect)
  - Interaction effect

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."*

N is total scores

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.

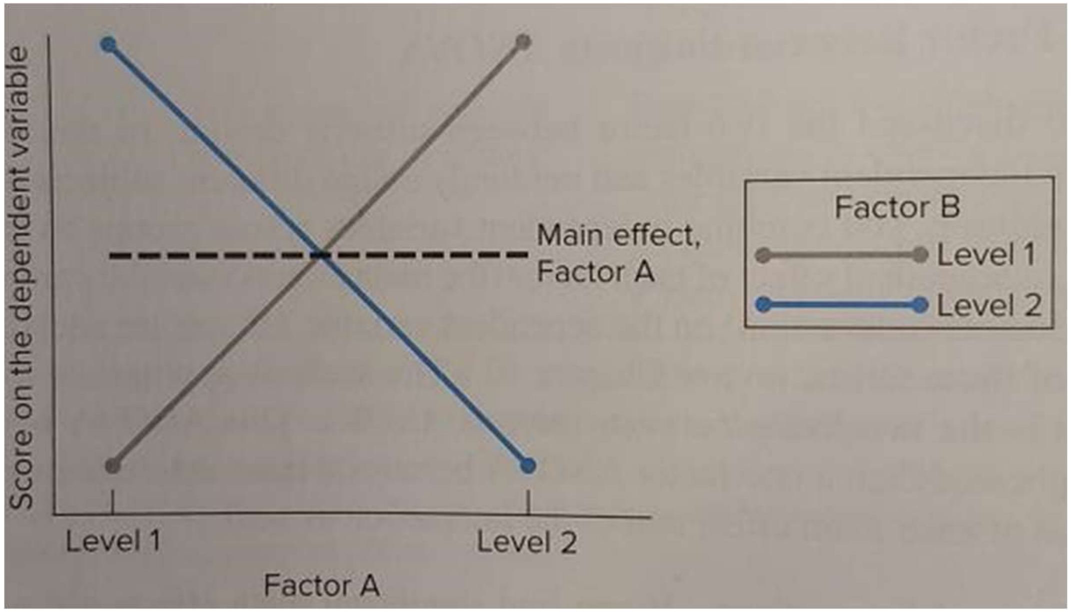
There is a significant interaction effect of Book and Instructor on Knowledge gain.

Reporting rule

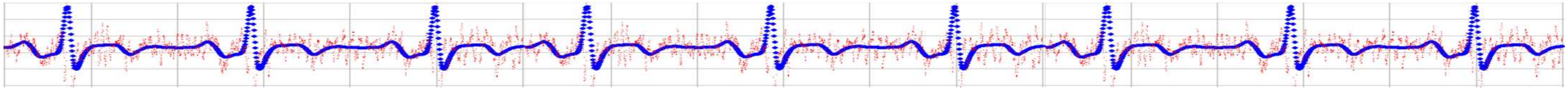
- IF you have a significant interaction
- THEN
  - In general, only report interaction, not any main effects, even if significant.
  - However, you must inspect the means to determine if main effects make sense to report
- Interaction => you cannot interpret the effect of one factor without the other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

Reporting

Why you must be careful with main effects when there are interactions:



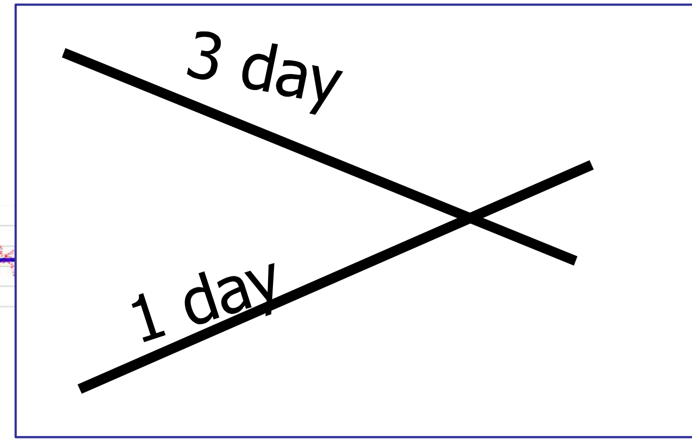
Results?



	Sig.
TrainingDays	0.34
Trainer	0.12
TrainingDays * Trainer	0.41

n.s.

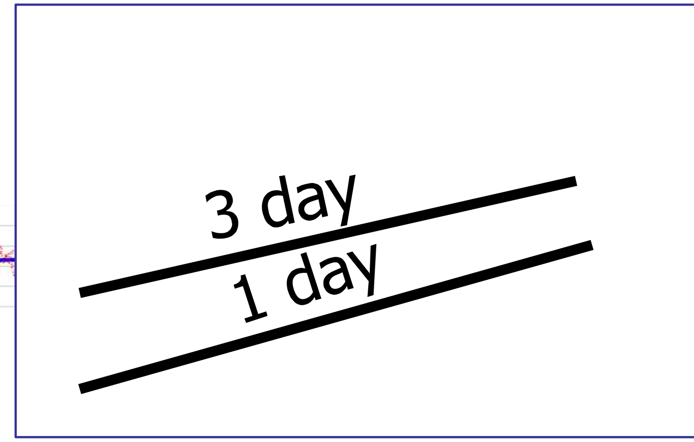
Results?



	Sig.
TrainingDays	0.34
Trainer	0.12
TrainingDays * Trainer	0.02

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

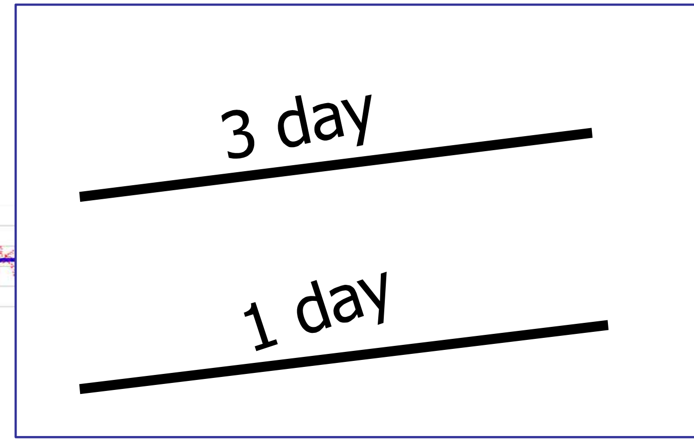
Results?



	Sig.
TrainingDays	0.34
Trainer	0.02
TrainingDays * Trainer	0.41

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

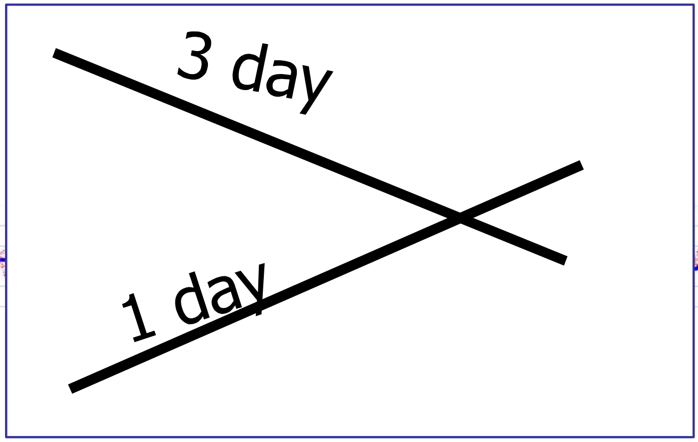
Results?



	Sig.
TrainingDays	0.02
Trainer	0.34
TrainingDays * Trainer	0.41

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

Results?

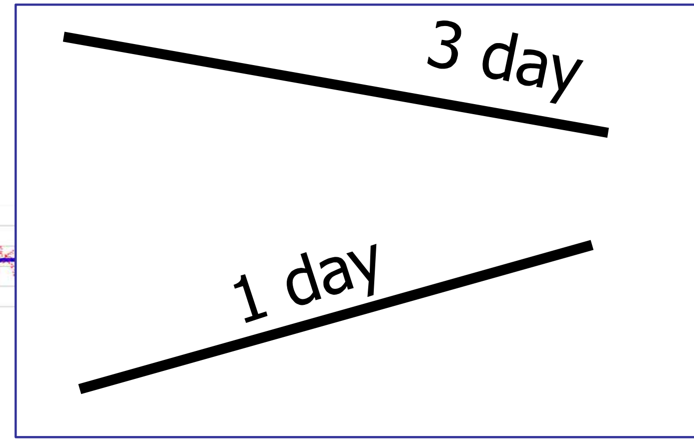


	Sig.
TrainingDays	0.04
Trainer	0.12
TrainingDays * Trainer	0.01

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

Do not report TrainingDays as significant

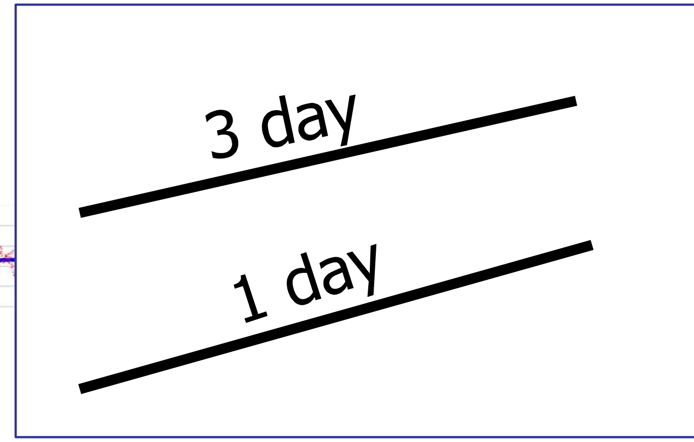
Results?



	Sig.
TrainingDays	0.04
Trainer	0.12
TrainingDays * Trainer	0.01

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

Results?



	Sig.
TrainingDays	0.04
Trainer	0.02
TrainingDays * Trainer	0.41

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

Reporting example

η^2, or eta squared, (effect size) is same as R^2 from last lecture

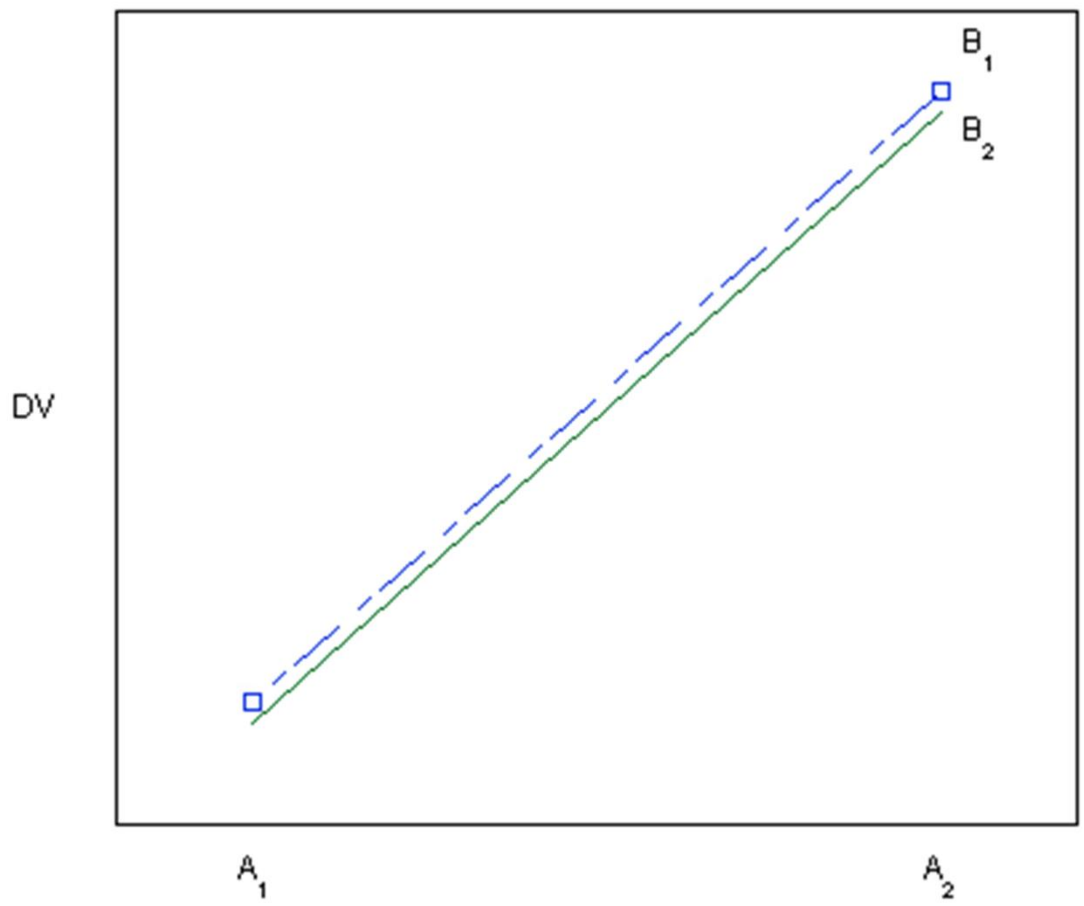
- "A 3x2 analysis of variance (ANOVA) on the procedure satisfaction scale showed main effects of both procedure, $F(1,136) = 94.28, p<.01, \eta^2=.41$, and group belongingness, $F(2,136)=3.7\text{-}0, p<.03, \eta^2=.05$. More important … was that this analysis also yielded the predicted interaction effect, $F(2,136)=3.46, p<.04, \eta^2=.05$. The cell means and standard deviations are shown in Table [next slide]. Findings showed that inclusion in a group leads to stronger effects of voice as opposed to no-voice procedures on participants' ratings of procedureal satisfaction than exclusion from a group."

Table example

Procedure	Group Belongingness					
	Inclusion		Exclusion		Not yet known	
	M	*SD*	*M*	*SD*	*M*	*SD*
Voice	4.86_a	1.82	3.31_b	1.83	4.46_a	1.95
No voice	1.89_c	1.14	1.81_c	0.84	1.69_c	0.83

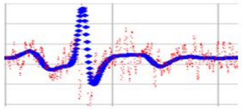
Note: Means are on 7-point scales, with higher values indicating more positive ratings of procedural satisfaction. Means with no subscript in common differ as indicated by a least significant difference test for multiple comparisons between means ($p < .05$).
Source: van Prooijen, J. W., van den Bos, K., & Wilke, H. A. M. (2004). Group belongingness and procedural justice: Social inclusion and exclusion by peers affects the psychology of voice. *Journal of Personality and Social Psychology*, *87*, 66–79. Published by the American Psychological Association. Reprinted with permission.

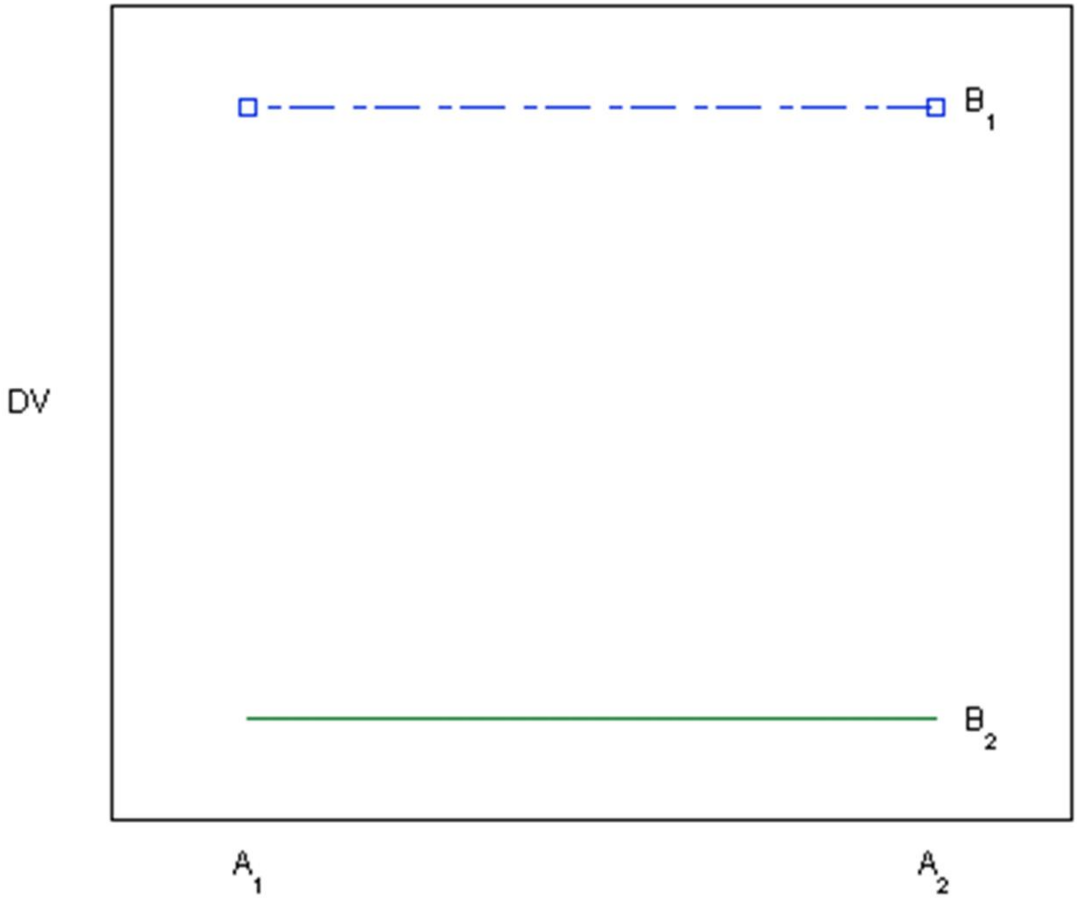
Possible interpretation?

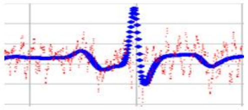


http://courses.washington.edu/smartpsy/interactions.htm#

Possible interpretation?

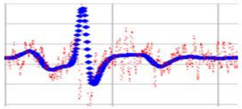


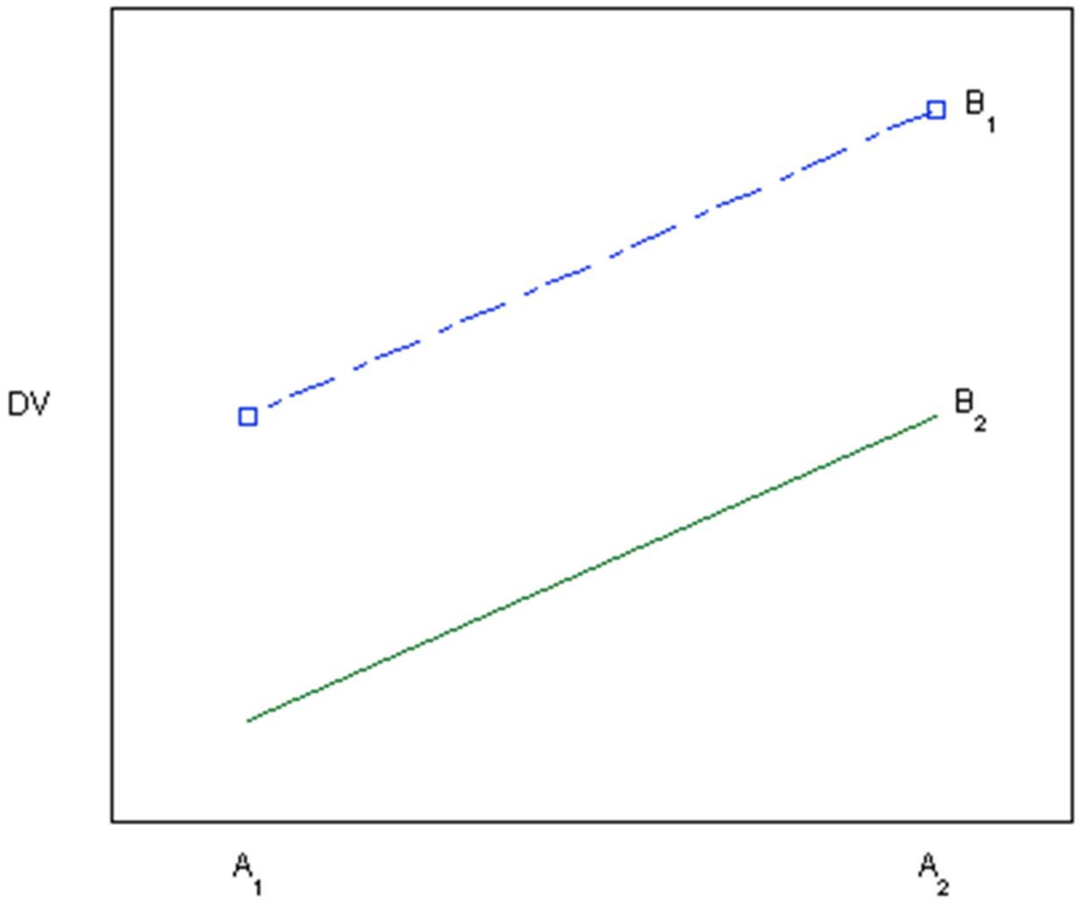


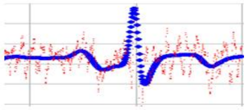


http://courses.washington.edu/smartpsy/interactions.htm#

Possible interpretation?

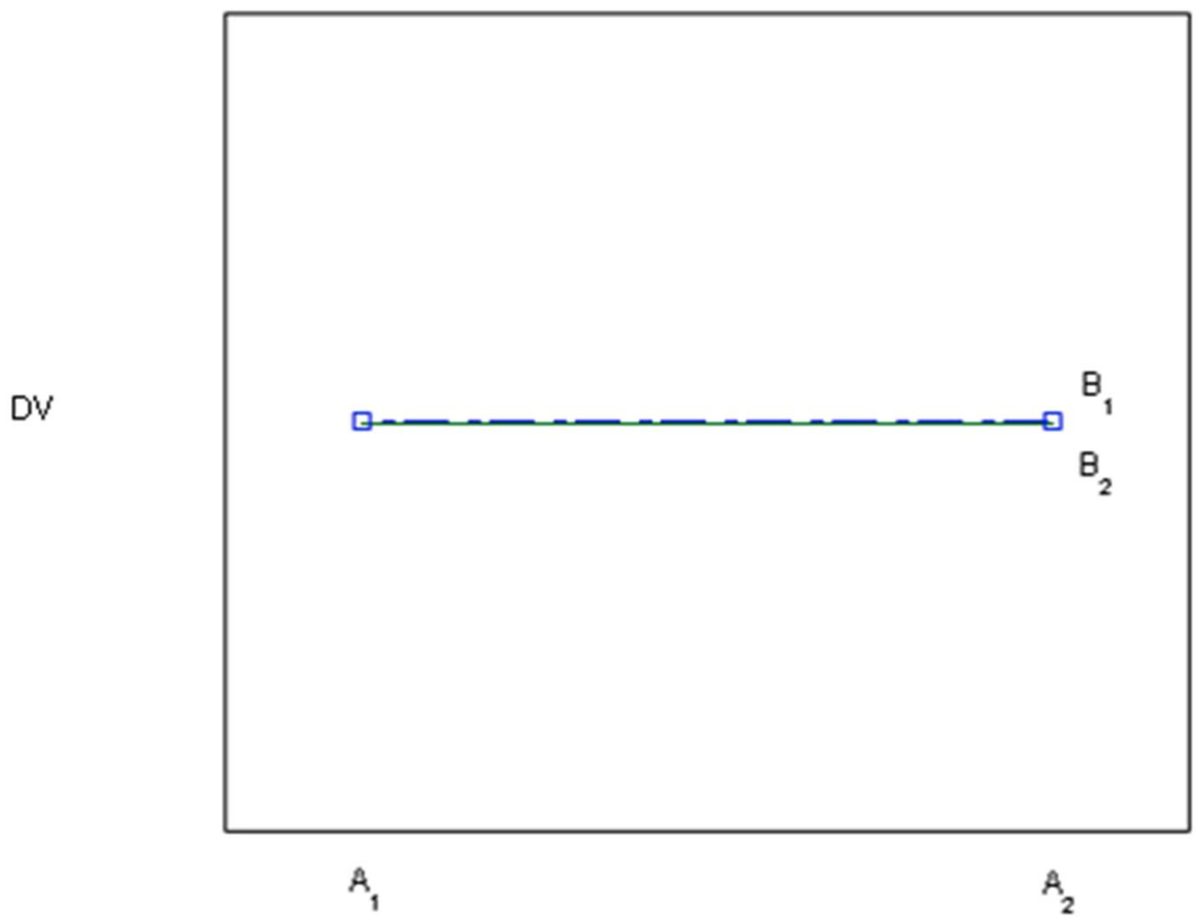






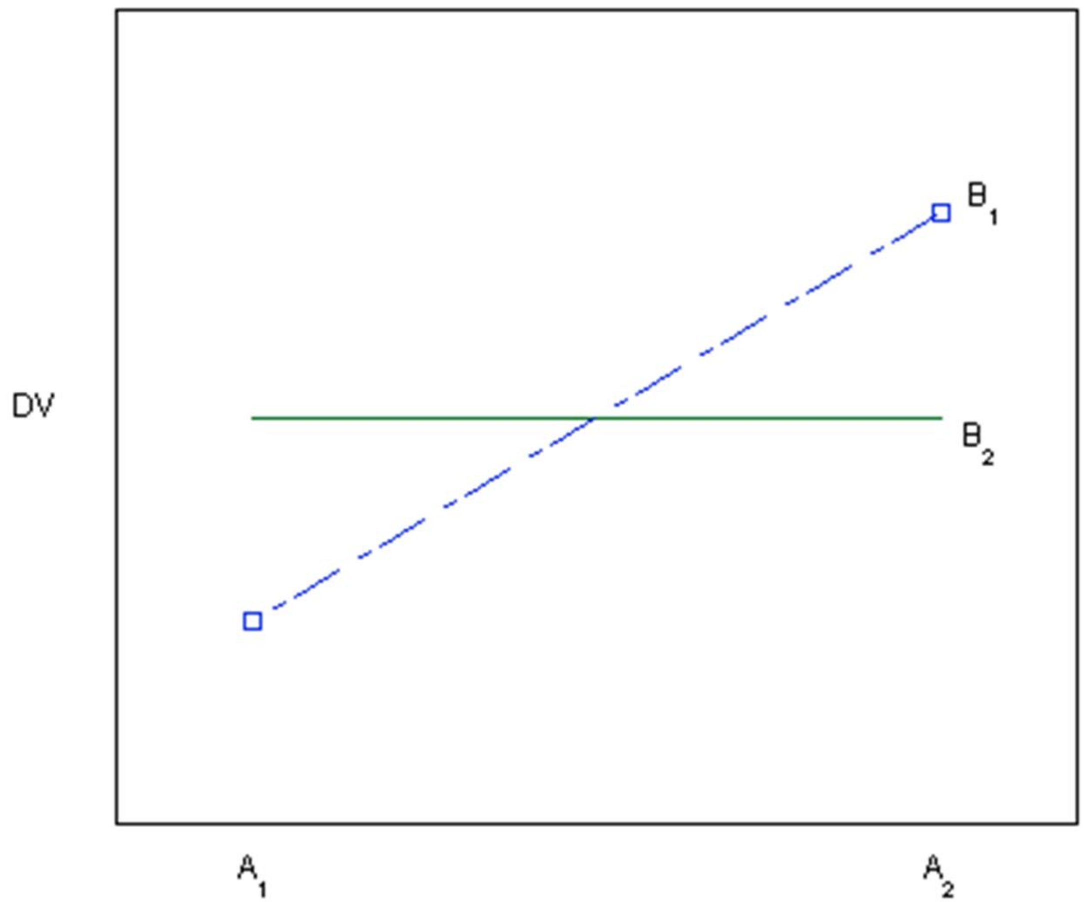
http://courses.washington.edu/smartpsy/interactions.htm#

Possible interpretation?



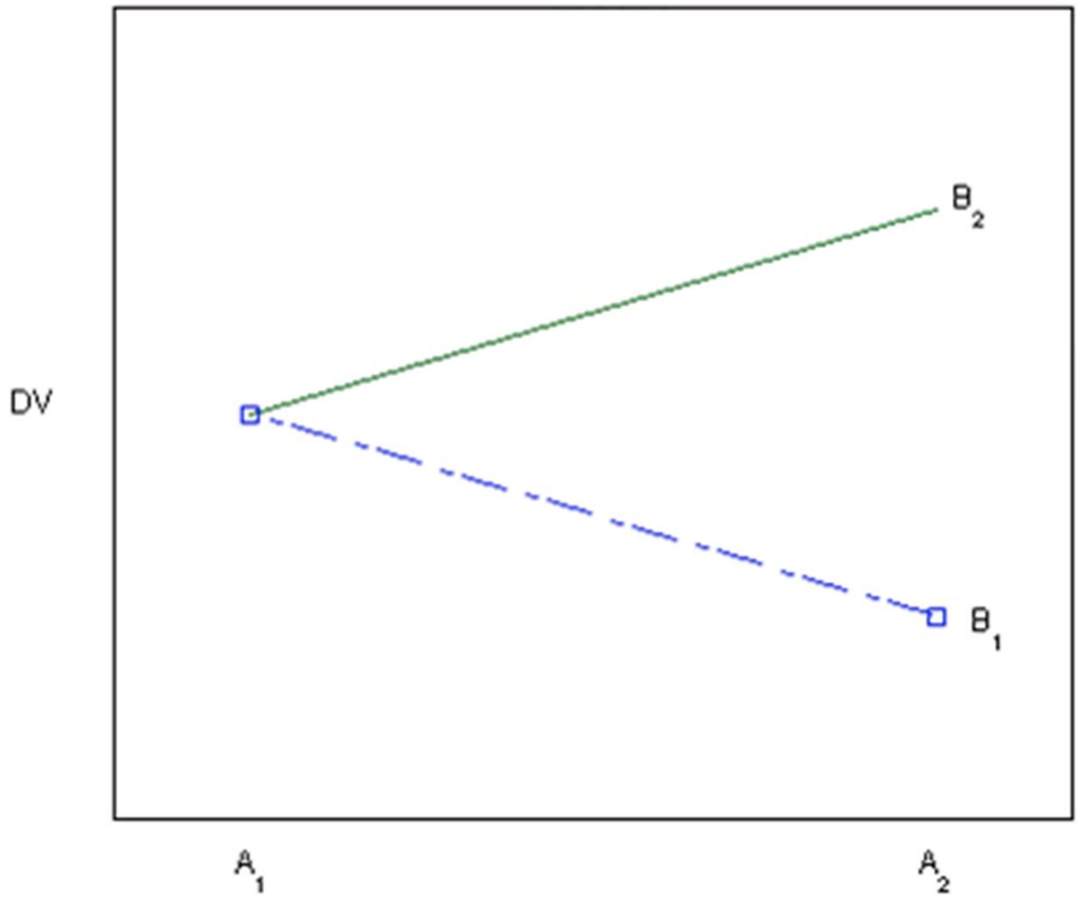
http://courses.washington.edu/smartpsy/interactions.htm#

Possible interpretation?



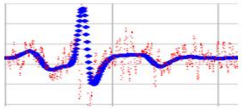
http://courses.washington.edu/smartpsy/interactions.htm#

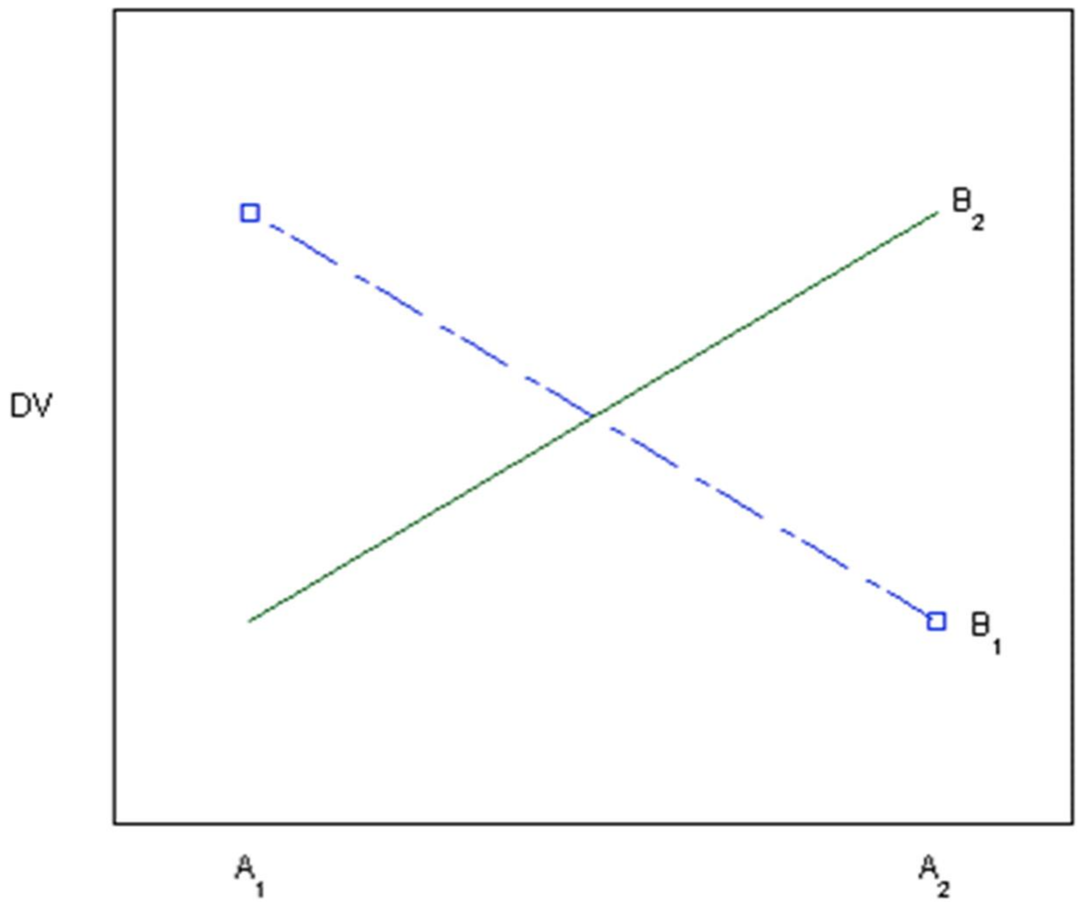
Possible interpretation?

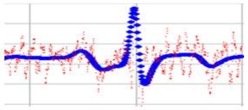


http://courses.washington.edu/smartpsy/interactions.htm#

Possible interpretation?



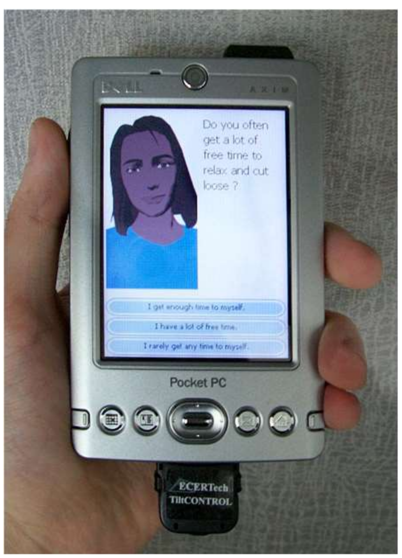




http://courses.washington.edu/smartpsy/interactions.htm#

“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

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.
  - Text: *at least* 5 participants per group

Higher-order designs: Often unrealistic

- 2 x 2: *At least* 20 people
- 2 x 2 x 2:
  - *At least* 40 people
  - Three main effects plus three two-way interactions (AxB, AxC, BxC) and one three-way interaction (AxBxC)
- Three-factor, three level (i.e., 3 x 3 x 3): *At least* 135 people More complex interpretation

ANOVA Assumptions

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

\+ *Minimally*, have 5 participants per cell (usually need much more)

Dichotomization

- Suppose you have two continuous IVs
  - Aptitude
  - Age
- Can you break at mean (i.e., median split, or dichotomize) to get groups?

Dichotomization: Caveats

- Lose information – increase chance of Type II error
- Sometimes a more “conservative” approach (does not increase chance of Type I), but not always so

Alternative is to use techniques based on multiple regression

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)
  - 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 <u>much</u> more.

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!

Factorial within-subjects design

- Each subject exposed to every combination of levels of all factors (IVs)
- Counterbalance order to deal with carryover effects

		Fear type	
		Physical	Social
Fear intensity	Low	2.19	2.41
	High	3.85	3.02

Group time: T3