Psychology 205: Research Methods in Psychology
Research Designs
Between Subject Designs

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2021

Outline

Teaching effects

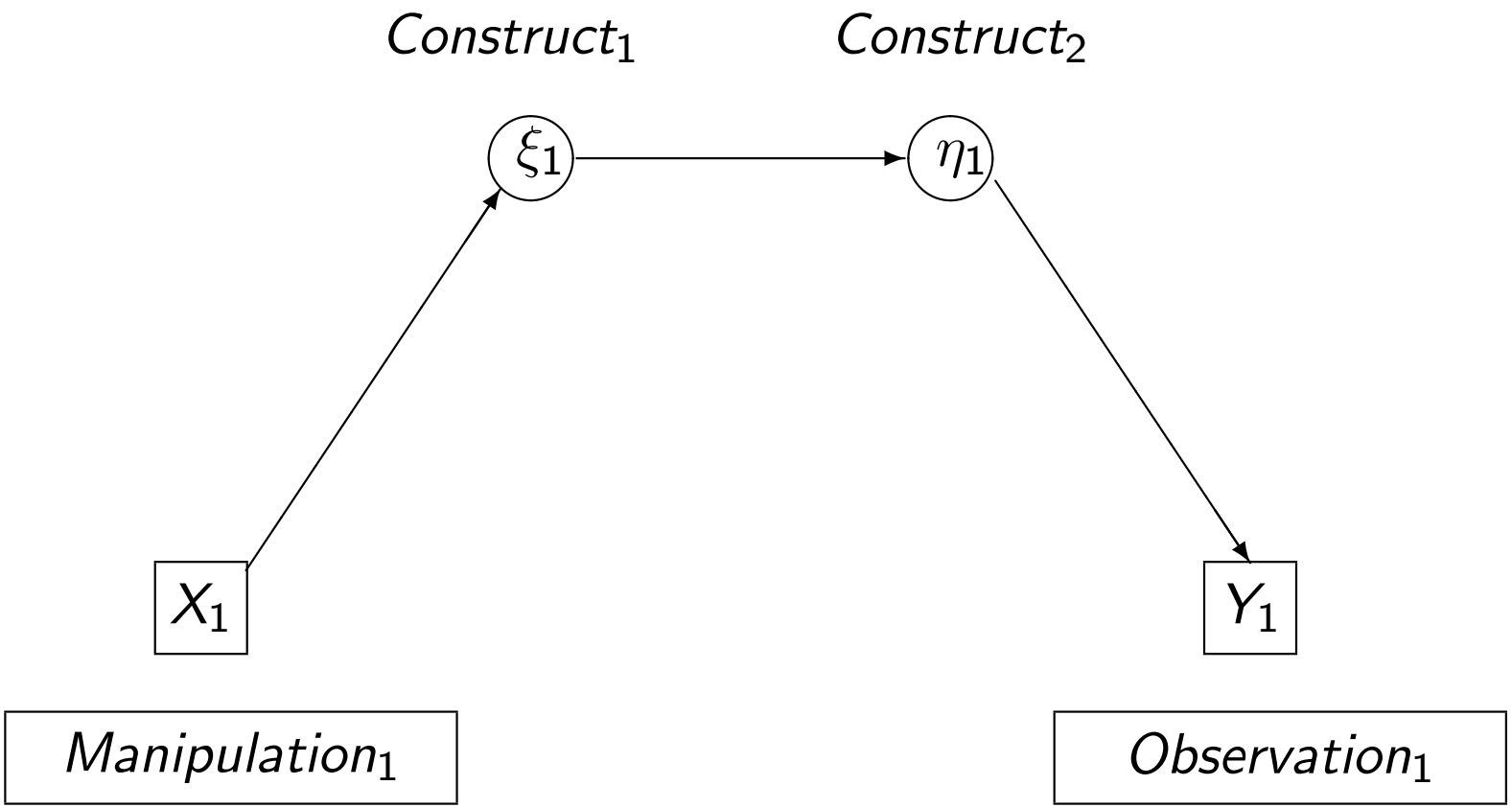
1. A teacher of statistics wanted to compare two methods of teaching introductory statistics. One method relied heavily on the teaching of the theory behind statistics (theory method). The other method was labeled the cookbook method because it consisted of teaching the students various statistical tests and informing them when to use each test.
2. The researcher found that a leading engineering school was using the theory method in all its introductory statistics classes and that a state teachers college was using the cookbook method in all its classes.
3. At the end of each semester he administered a standardized test on the applications of statistics to the statistics classes of both schools. The results of this testing indicated the classes that received the theory method were far superior to the classes that received the cookbook method. The research concluded that the theory method was the superior method and should be adopted by teachers of statistics.

The most basic of all theory: Construct 1 $->$ Construct 2

ξ η

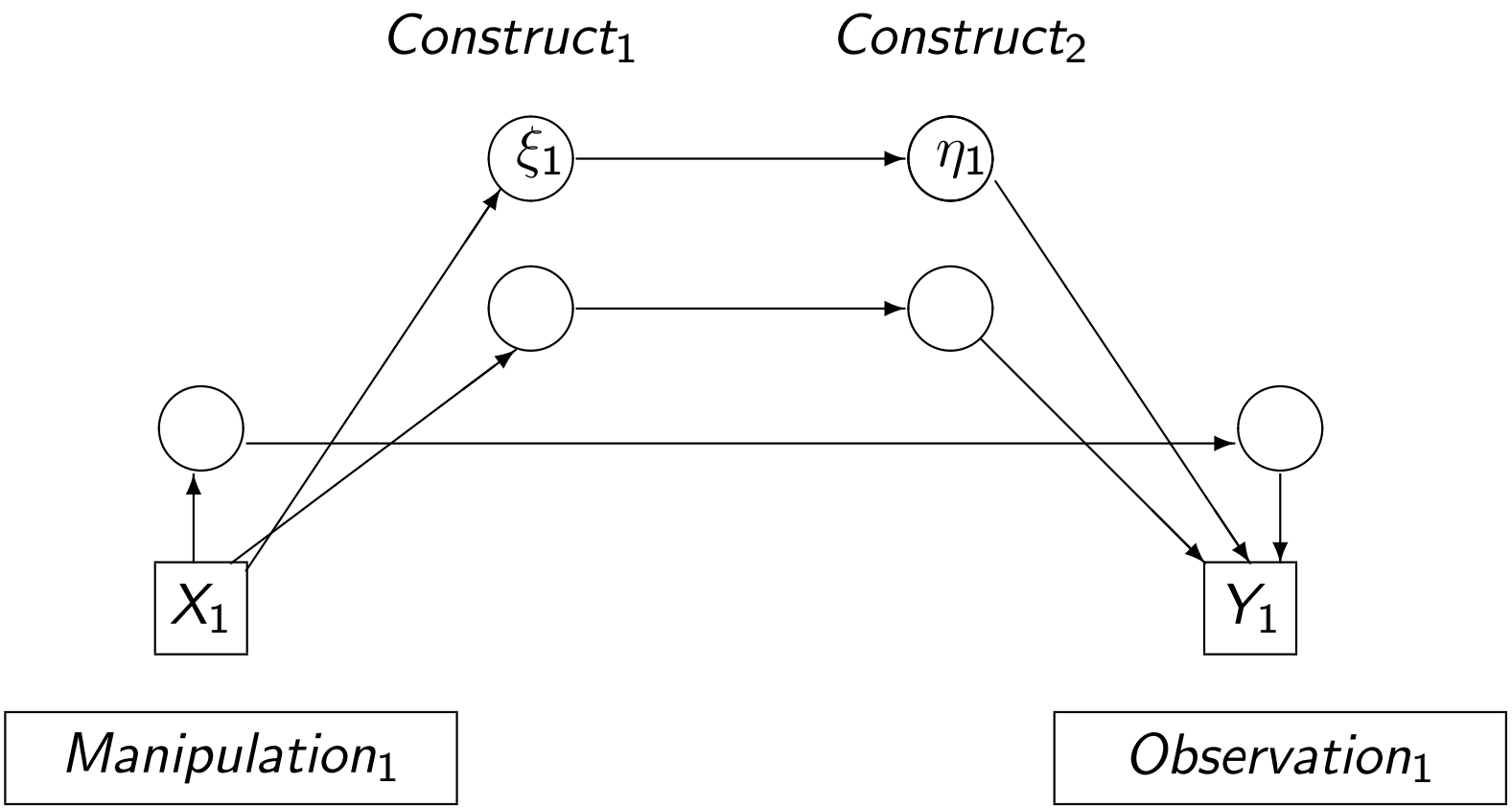
The most basic of all theory: Construct 1 $->$ Construct 2

X ξ η Y



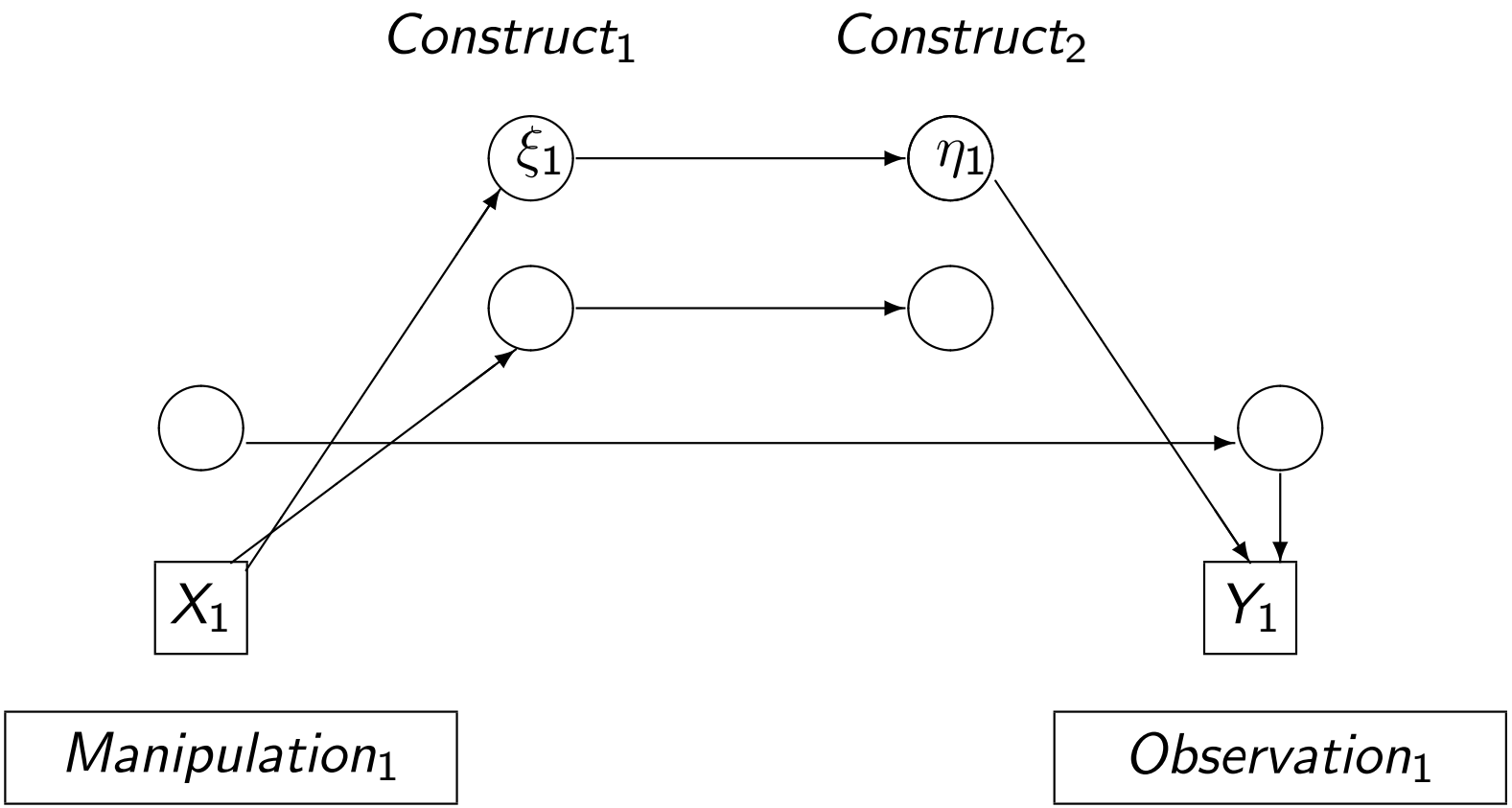
But all manipulations and observations have multiple effects and causes

X ξ η Y



Design eliminates extraneous paths

X ξ η Y



Overview of the problem

1. Theoretical problem: understanding the relationship between latent variables (constructs)
   - Relationships among latent variables
   - Relationships between latent variables and observed variables
2. Generalization of results and threats to external validity (how to choose subjects, how general are the results)
3. Proper design maximizes internal validity

Generalization of results and threats to external validity-I

1. Limitations of generalization for subjects
2. Limits of generalization for conditions – interactions with other variables

A hypothetical problem

Dependent Variable = f(Independent Variable + Confounding Variables 1 ... N)

Table: DV $\sim$ IV1 +CV_1 CV_2 + CV_3 + ...+ CV_n

A table from the psych package in R

Subject	DV	IV_1	CV_1	CV_2	CV_3	...	CV_n
1	2	1	1	0	0	...	0
2	4	0	2	0	1	...	1
3	6	0	3	1	0	...	2
4	9	1	4	1	1	...	2
5	9	1	5	2	0	...	1
6	9	0	6	2	1	...	0

Unfortunately, although we can observe the IV and the DV, we do not observe the confounding variables. We want to control for their effects by proper design.
Design tries to reduce the correlation between confounding variables and the Independent Variable

Are our subjects from WEIRD cultures?

Do our results generalize?

1. Western
2. Educated
3. Industrialized
4. Rich
5. Democratic

The weirdest people in the world (Henrich, Heine & Norenzayan, 2010)

Most subjects are American Undergrads

1. 68% of subjects are from US, 96% from North America, Europe, Australia, Israel
2. In JPSP 67% of American studies are undergraduates, 80% of studies from other countries
3. "In other words, a randomly selected American undergraduate is more than 4,000 times more likely to be a research participant than is a randomly selected person from outside of the West." (p 63)
4. Should it be the Journal of Social and Personality Psychology of Western subjects?

The Muller-Lyer illusion

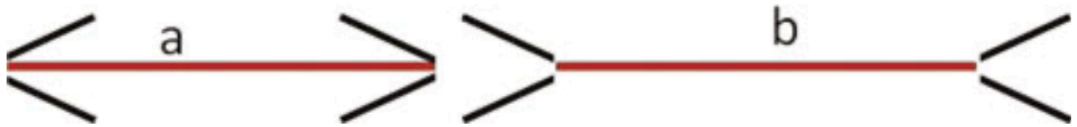


Figure 1. The Müller-Lyer illusion. The lines labeled "a" and "b" are the same length. Many subjects perceive line "b" as longer than line "a".

The Muller-Lyer illusion is a Western effect

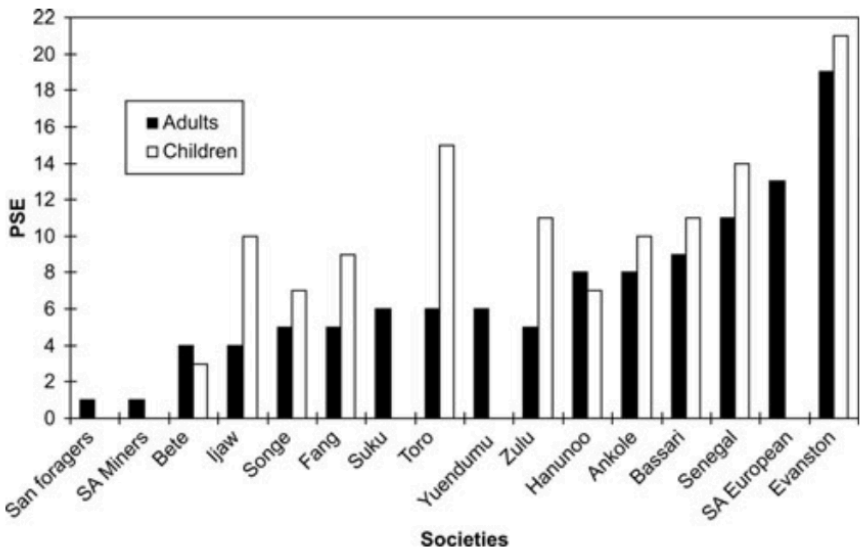


Figure 2. Müller-Lyer results for Segall et al.'s (1966) cross-cultural project. PSE (point of subjective equality) is the percentage that segment a must be longer than b before subjects perceived the segments as equal in length. Children were sampled in the 5-to-11 age range.

Generalization of results and threats to external validity-I

1. Limitations of generalization for subjects (Are our subjects WEIRD?)
   - (1) freshman psych students at NU
   - (2) students at NU
   - (3) college students at selective research universities
   - (4) college students
   - (5) 18-24 year olds
   - (6) North Americans
   - (7) Humans
2. Limits of generalization due to confounding variables.
3. Limits of generalization for conditions – interactions with other variables.

Generalization of results and threats to external validity-II

1. Lmits of generalization for conditions – interactions with other variables
2. Problems and benefits of interactions
3. (a) xy relationship depends upon z
4. (b) Example from (Revelle, Humphreys, Simon & Gilliland, 1980)
   - i) In the morning, caffeine facilitates working memory performance
   - ii) In the evening, caffeine hinders working memory performance
5. (2) Interactions limit generalization
6. (3) Interactions test theoretical limits

Practical problems and threats to internal validity

1. Manipulations affect more than the construct of interest
   - Examples
   - (1) caffeine induces alertness and motor tremor
   - (2) failure induces anxiety, depression, anger
   - (3) practice leads to motivational changes as well as changes in skill
2. Observable variables reflect more than the construct of interest
   - Self report of alertness reflects base line differences in mood.
   - Cognitive performance reflects ability, motivation, training, and practice.
   - Slowness of responding reflects caution as well as processing speed

Types of designs – None are perfect

1. Within subject designs
   - Controls for subject variability
   - Confounds practice/order effects with manipulation
   - Two or more conditions – repeated many, many times
2. Between subject designs
   - Subject variables as an alternative explanation of results – threats to validity
   - Randomization as a control
3. Mixed – Within/Between

Between Subject designs

Subject variables as threat to external validity

1. Ability
2. Practice
3. Motivation
4. Interest
5. Gender
6. Age
7. Culture

Between Subject designs

Confounded effects that can lead to subject variability

1. Time of day
   - Naturally occurring rhythms of alertness
   - Classroom effects
   - Fatigue
2. Time of week, month, season, year
   - Class schedules
   - Mid terms
   - Papers
   - Weather
3. Volunteer effects
4. Experimenter-Subject interactions

Between Subject designs

1. Subject variables as threat to external validity
2. Confounded effects that can lead to subject variability
3. Randomization as a control

Leadership training

1. A YMCA official in a small town wanted some evidence to prove that his program was valuable in training future leaders. He went back to the membership records and got the names of those boys who were active members in his program 20 years before. He also took school records and got the names of boys who were not YMCA members. He compared the two groups as to present occupations, salaries, and so on, and found that the YMCA group was doing much better. He concluded that this result was due to the influence of his program.
2. What were the constructs of interest?
3. What are possible threats to the validity of this study?

A hypothetical problem

Dependent Variable = f(Independent Variable + Confounding Variables 1 ... N)

Table: DV $\sim$ IV1 +CV_1 CV_2 + CV_3 + ...+ CV_n

A table from the psych package in R

Subject	DV	IV_1	CV_1	CV_2	CV_3	...	CV_n
1	2	1	1	0	0	...	0
2	4	0	2	0	1	...	1
3	6	0	3	1	0	...	2
4	9	1	4	1	1	...	2
5	9	1	5	2	0	...	1
6	9	0	6	2	1	...	0

Unfortunately, although we can observe the IV and the DV, we do not observe the confounding variables. We want to control for their effects by proper design.
Design tries to reduce the correlation between confounding variables and the Independent Variable

Randomization

1. Dependent variable is still a function of the confounding variables, but the correlation between the experimental Independent Variable with the Confounding Variables is reduced. (The expected correlation is 0, although the observed correlation will vary from sample to sample.)
2. We have a purer measure of the influence of the IV upon the DV independent of the effects of the Confounding Variables.

Randomization as a control

1. Only the expected values of groups are equal – not the observed values
   - In any particular experiment, groups are not equivalent.
   - Expected value of the (signed) group difference=0.
   - Randomization does not introduce systematic bias.
2. Confounding variables are (on the average) not correlated with our experimental variables

Types of Randomization

1. Subjects matched on variable of interest and then assigned to condition,
2. Blocking to control for order effects
   - Controls for stable subject effects.
   - Eliminates subject effects associated with time of appearance.
3. Complete randomization
   - "Failures" of randomization
   - Problems at the end of the experiment
   - Power is maximized with equal cell sizes
   - Randomization will tend not to produce equal size groups
4. Block Randomization
   - Randomize within blocks of subjects
   - Will lead to equal cell sizes, reduces chance of end effects

four random orders

ord1 <- sample(2,12,replace=TRUE)

Table: four random orders

A table from the psych package in R

Variable	ord1	ord2	ord3	ord4
1	2	1	1	2
2	2	2	2	2
3	1	2	2	1
4	2	1	2	2
5	2	2	2	1
6	1	2	1	2
7	2	2	2	1
8	1	1	2	1
9	2	2	2	1
10	2	2	1	1
11	2	2	2	2
12	2	2	2	2
# 1s	3	3	3	6

Number of 1s per column is not the expected value of 6

Randomization with an end effect

If we want equal cell sizes (to maximize power) some randomize but then when the cells are filled, fill with the remaining orders. This produces an "end effect".

Table: Simple randomization will lead to end effects if we balance cell sizes at end of experiment

A table from the psych package in R

Variable	ord1	ord1 "adjusted"
1	2	2
2	2	2
3	1	1
4	2	2
5	2	2
6	1	1
7	2	2
8	1	2
9	2	1
10	2	1
11	2	1
12	2	1

Block randomization avoids end effects

R code

```
b1 <- block.random(12,2) #12 subjects, 2 subjects per random block
bl2 <- block.random(12,2)  #do another 12 with 2 per block
df <- data.frame(b1,b12) # put them together to show them
```

Table: Block randomization will avoid end effects

Randomize within **blocks** of subjects.

A table from the psych package in R

Variable	blcks	IV1	blc.1	IV1.1
S1	1	1	1	2
S2	1	2	1	1
S3	2	2	2	2
S4	2	1	2	1
S5	3	1	3	1
S6	3	2	3	2
S7	4	2	4	2
S8	4	1	4	1
S9	5	2	5	1
S10	5	1	5	2
S11	6	1	6	2
S12	6	2	6	1

Proof Reading and error rate

1. An investigator does a study on proofreading for typographical errors. There are three levels of the independent variable (three conditions), distinguished by the number of errors per line. In one case there is one error per every 5 lines, in another one error per every 10 lines, and in the third one error per every 20 lines. The three typescripts are exactly the same except for the typographical errors inserted. Random groups are used and the subjects are instructed to go through the script as rapidly as possible, identifying all errors with a slash. For the dependent variable (response measure), the investigator uses the mean number of failures to detect an error. For the three levels, these values are 5.0, 4.9, and 5.1. These do not differ reliably.
2. The investigator concludes that the error rate is independent of the number of to-be-detected errors.
3. This is probably an inappropriate conclusion.
4. Why?

Learning to learn

1. An investigator set about to get a definitive answer on progressive changes in learning as a function of practice (learning-to learn) for free-recall lists. He decided to study learning-to learn as a function of 2, 4, 6, 8, and 10 successive lists. Five different random groups were used, the subjects being assigned to the five conditions (2, 4,6, 8, or 10 lists) by a block-randomized schedule. In terms of method, procedure, balancing of lists, and so on, the experiment was immaculate.
2. However, we would have to say it was a very inefficient way to obtain the information he sought.
3. Why?
4. What is a better way of doing this study?

Infant Nutrition

1. An investigator developed the idea that an excess of a certain chemical in the brain during infancy produced permanent mental retardation. To gather evidence germane to this notion, he used two groups of 15 newborn monkeys each. The 30 babies were assigned to the groups by a block-randomized schedule. One group, the C Group, was nursed by the mother monkeys, and it is assumed that the investigator could measure the amount of milk consumed. The other group, the E Group, was fed by bottle, but the milk was of exactly the same kind and of the same amount as that received by the naturally-fed monkeys. Of course, the baby monkeys in the E Group were separated from the mothers so that they would not nurse from them and thereby get more milk than those in the C Group. The independent variable X was given to the E Group by including the chemical in the nursing bottle. Tests of mental development were made on both groups at various points in time, even far beyond the nursing period. At every point of testing the monkeys in the E Group were found to be inferior to those the C Group. Such a finding would support the idea prompting the experiment.
2. The independent variable is confounded. How?
3. Should it be concluded that X is not responsible for the observed differences?

Antiallergenic drug

1. A prominent pharmaceutical company wishes to test the effects of a new allergy drug, Drug X. They gather a sample of people with a certain minimal allergy level and randomly assign them to one of two conditions: 100 mg of X and a control group that was given no drug. They discovered that the subjects in the drug condition had a significant reduction in allergy symptoms but those in the control group did not. They concluded that Drug X is effective in fighting allergies and should be sold to the public.
2. Do the conclusions follow?
3. How could this study be improved?

Vitamin C and health

1. An experimenter wanted to test whether rats treated with Vitamin C were healthier than rats that received no Vitamin C. In order to test his hypothesis that Vitamin C would be beneficial to rats' health, he bought 40 random rats and placed them in a box. There were two conditions, a Vitamin C condition in which rats were given Vitamin C for several weeks and a control condition in which rats were not given Vitamin C. The researcher assigned rats to each condition by reaching into the box and randomly picking out rats. The first 20 rats picked from the box were assigned to the control condition, and the remaining 20 were assigned to the Vitamin C condition. After several weeks, the researcher discovered that the rats that received Vitamin C were on average healthier than the control rats. He concluded that Vitamin C was beneficial to rats' health.
2. Does the conclusion follow?
3. How could this study be improved?

Interactions and Main Effects

1. Anova framework of "Rows", "Columns", "Interactions"
2. Regression framework of DV $\sim$ IV_1 + IV_2 + IV_1*IV_2
3. The original ANOVA design comes from agricultural studies where one was crossing (e.g.,) seeds in rows with fertilizer in columns. Each plot of land was given a different condition.
4. The use of the interaction was to see if different fertilizers had different effects upon different seeds.

Interactions and Main Effects: a table

Table: A two by two ANOVA table

	IV_{2-low}	IV_{2-high}	Row Mean
IV_{1-high}	Aa	AA	A.
IV_{1-low}	aa	aA	a.
Column mean	.a	.A	Grand mean

- Row effect = A. - a. = Aa + AA - aa - aA
  = (Aa+AA) - (aa + aA)
- Column effect = .A - .a = AA + aA - Aa - aa
  = (AA + aA) - (Aa+aa)
- Interaction effect (do the slopes differ)
  = (AA + aa) - (Aa + aA)

Anova as multiple t-tests

	IV_{2-low}	IV_{2-high}	Row Mean
IV_{1-high}	Aa	AA	A.
IV_{1-low}	aa	aA	a.
Column mean	.a	.A	Grand mean

Effects can be thought of as combinations of cells, then a t-test on the weighted sum.

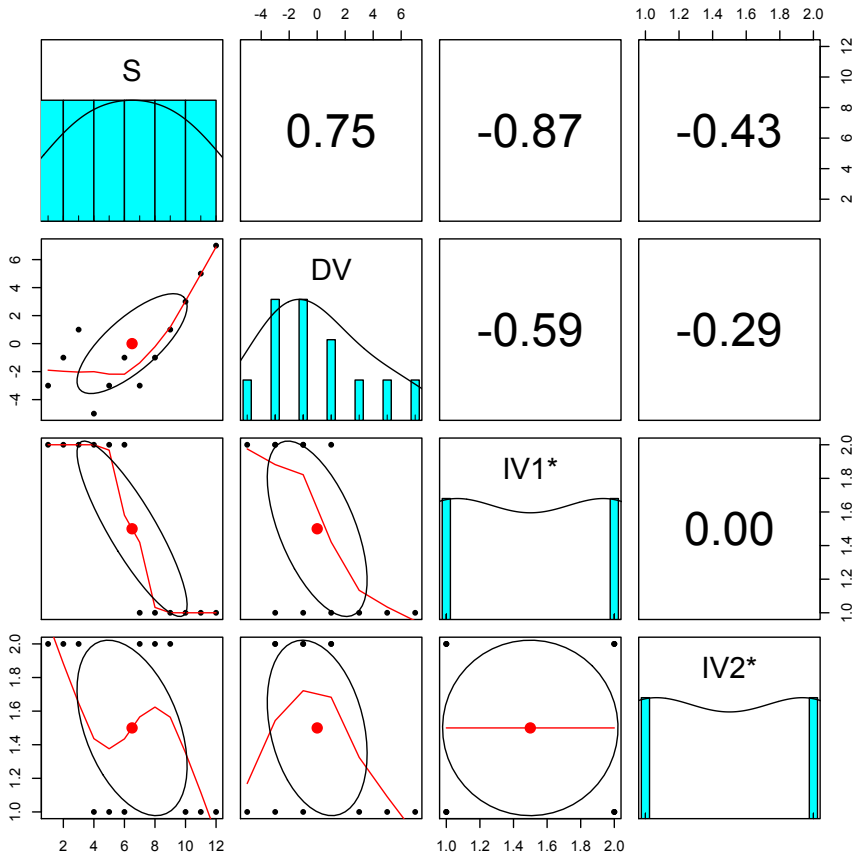
Cells	Aa	AA	aa	aA
Columns	-1	1	-1	1
Rows	-1	-1	1	1
Interactions	1	-1	-1	1

Some hypothetical data

R code

```
my.data <- read.clipboard()
```

	S	DV	IV1	IV2
1	-3		L	L
2	-1		L	L
3	1		L	L
4	-5		L	H
5	-3		L	H
6	-1		L	H
7	-3		H	L
8	-1		H	L
9	1		H	L
10	3		H	H
11	5		H	H
12	7		H	H



Note the confounding between Subject order and the IVs. Is the effect of the IVs on the DV due to this?

Analysis of Variance as way of estimating the size of an effect

1. We expect samples to differ due to random sampling.
2. The expected variance of sample means is $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$
3. If the observed variance of the sample means is bigger than the expected value, we conclude that this is unusual given the hypothesis that the samples came from the same population.
4. We find an estimate for the population variance from the within groups variance and from the between group variance (times n).
5. These variance estimates are called Mean squares.

Analysis of Variance using the aov function

R code

```
model1 <- aov(DV ~ IV1 + IV2 + IV1*IV2,data=my.data)
summary(model1)  #show the result
print(model.tables(model1,"means"))
```

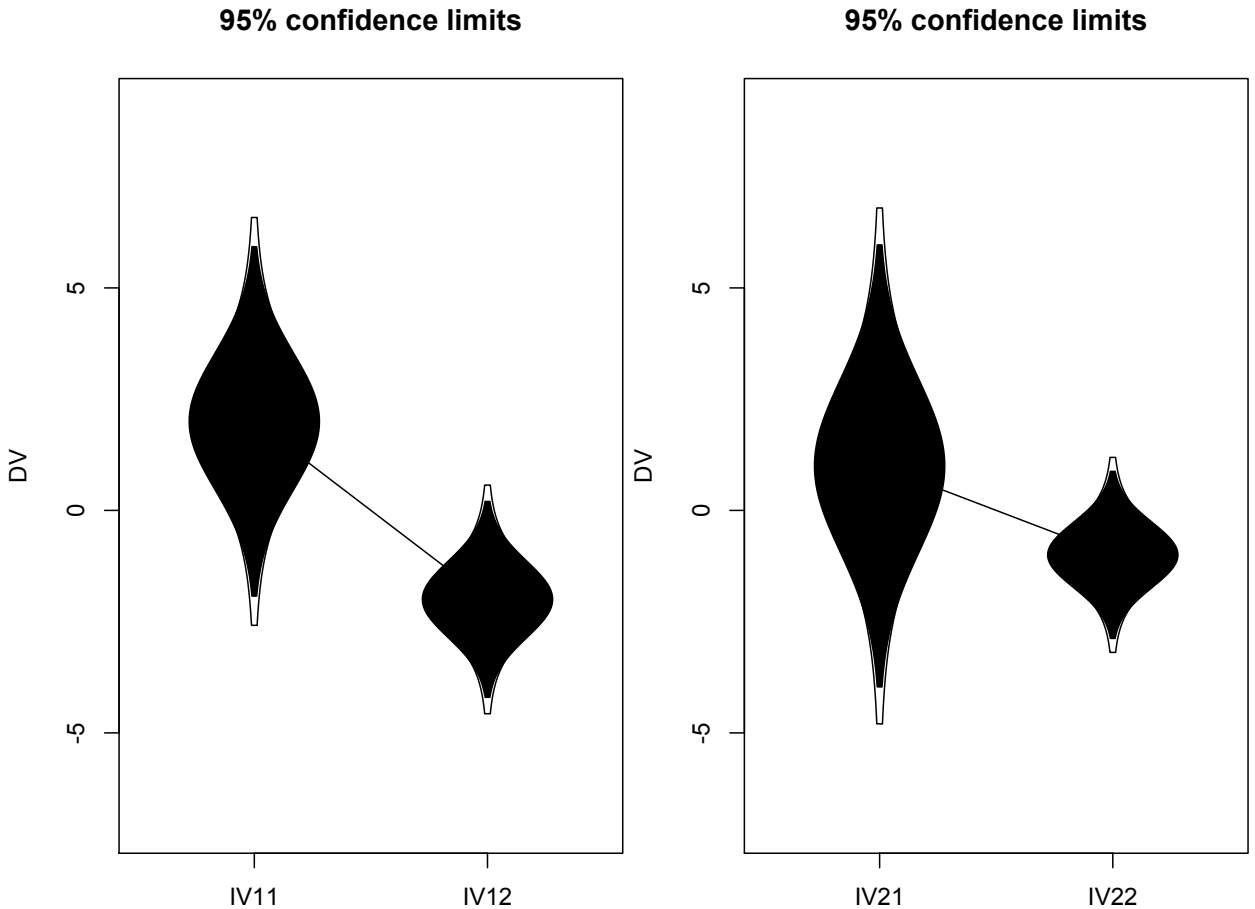
```
            Df Sum Sq Mean Sq F value  Pr(>F)
IV1          1     48      48      12 0.00852 **
IV2          1     12      12       3 0.12150
IV1:IV2      1     48      48      12 0.00852 **
Residuals    8     32       4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Tables of means
Grand mean
-2.56395e-16

IV1
 H  L
 2 -2
IV2
 H  L
 1 -1

 IV1:IV2
   IV2
IV1 H  L
  H  5 -1
  L -3 -1
```

1. Mean Squares are the variance estimates associated with particular effects.
2. Fs are comparisons of variance between groups to the residual variance within groups.
3. If the groups do not differ, these should be the same.

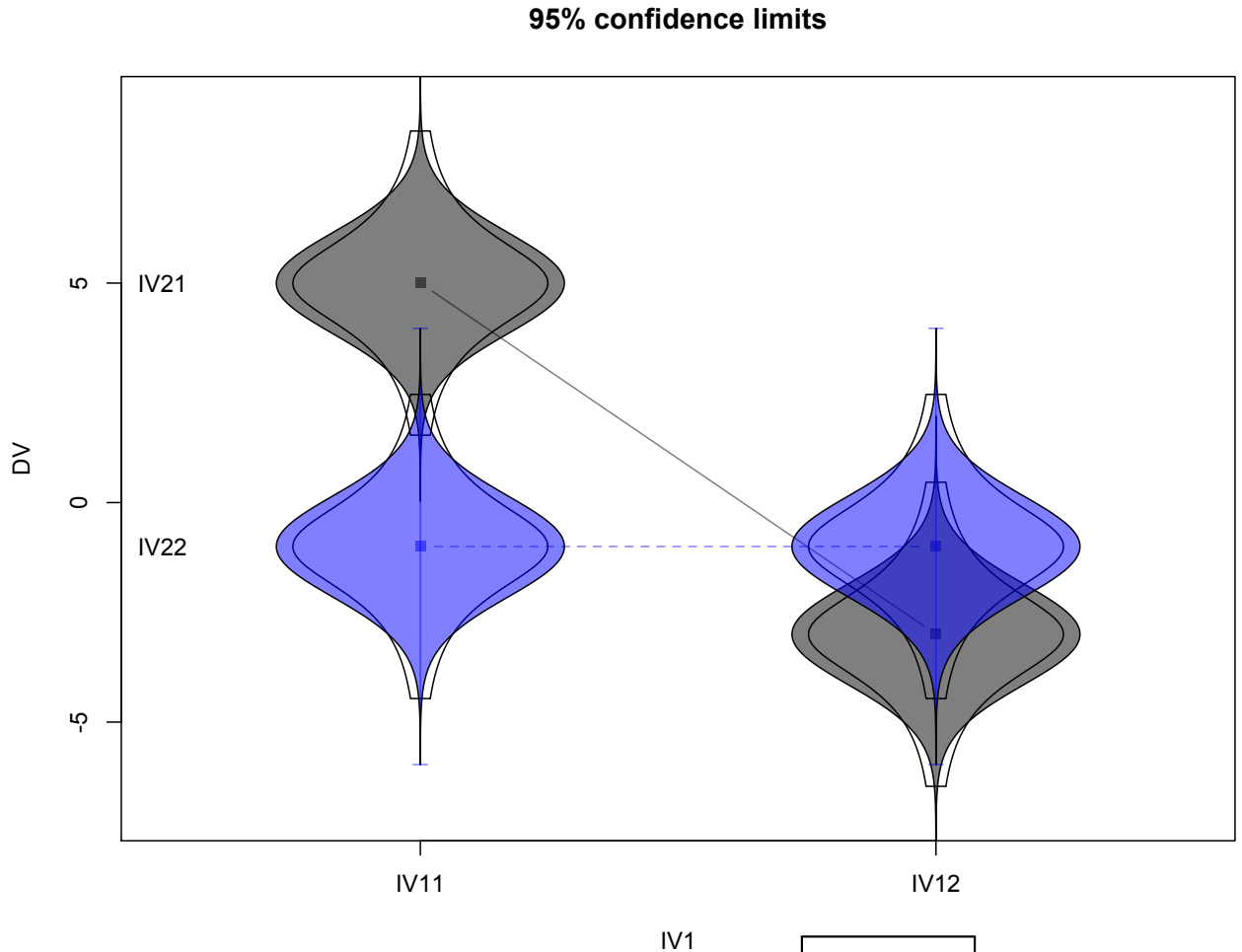
Graphic displays show this more clearly



R code

```
op <- par(mfrow=c(1,2))  #two panel graph
 error.bars.by(DV ~ IV1,data=my.data)
 error.bars.by(DV ~ IV2,data=my.data)
op <- par(mfrow=c(1,1))  #back to one panel
```

Graphics are particularly helpful for interactions



Main effects are hard (impossible?) to interpret in the presence of an interaction.

R code

```
error.bars.by(DV ~ IV1 + IV2,data=my.data,density=-30,add.labels="left")
 #negative density controls transparency
```

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Revelle, W., Humphreys, M. S., Simon, L., & Gilliland, K. (1980). Interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology General*, *109*(1), 1–31.