Research Methods

Week 3: Experimental Designs (Continued)
Within subjects

Arousal and Working Memory I

- An investigator was interested in the effect of arousal on short term memory. The hypothesis being tested was that caffeine induced arousal helps short term memory. Subjects were given a list of 20 words to study for 2 minutes, and were then asked to count backwards by 3s from 91. They were then asked to recall as many of the words as possible. The average number of words recalled was 10 (sd=3)

- After the recall was completed, subjects were given 200 mg of caffeine and allowed to read for 30 minutes while the caffeine took effect. They were then given the same list to study for 2 minutes, followed by counting backwards again from 91. They were then asked to recall as many words as possible from the list. The average this time was now 12 (sd=3). There were 20 subjects in this within subject experiment and the t-test of the correlated differences was 3.6 (d.f. =19, $p<.01$).

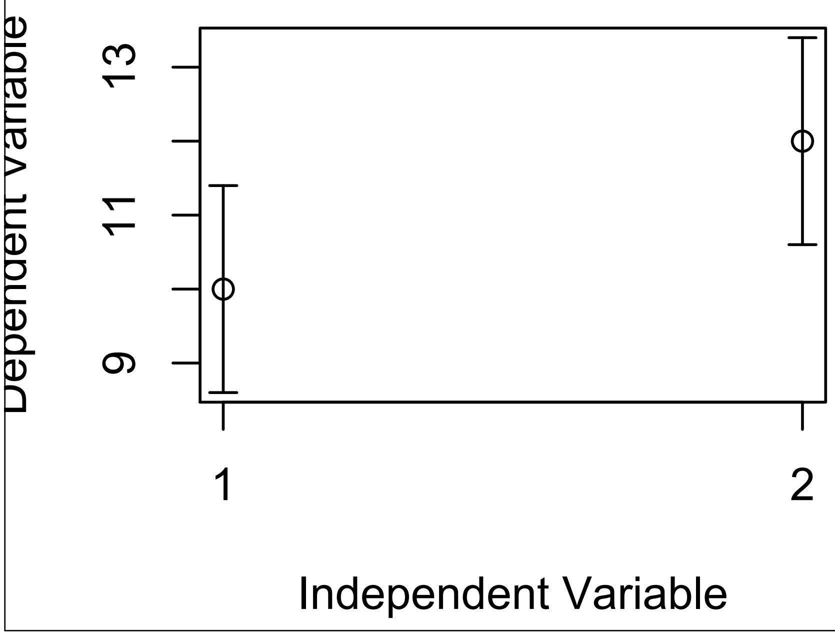
Study 1

```
> print(study1.df,digits=2)
          mean sd  n min max   se
 placebo    10  3 20   4  16 0.67
 caffeine   12  3 20   6  18 0.67
```

Generic Figures

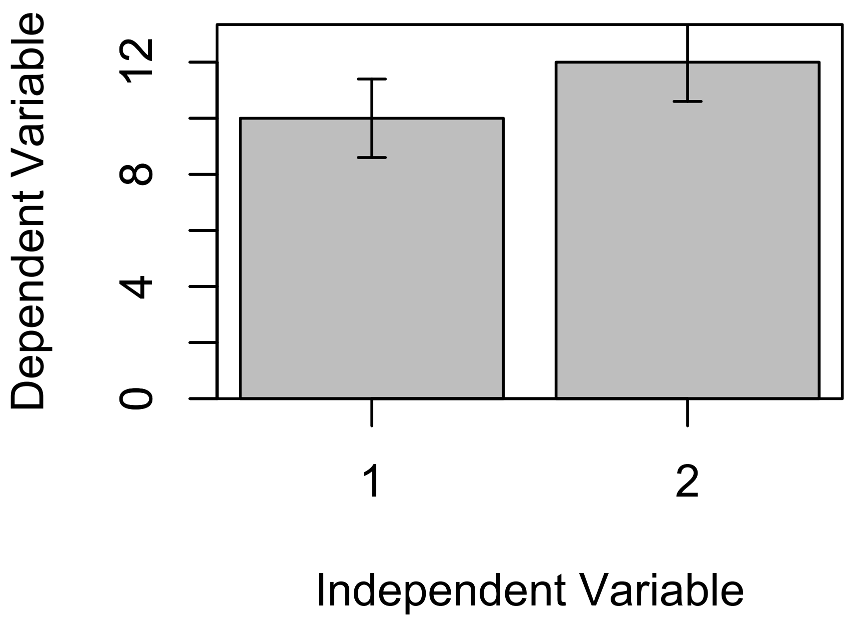
points + error bars

95% confidence limits



bar graph + error bars

95% confidence limits



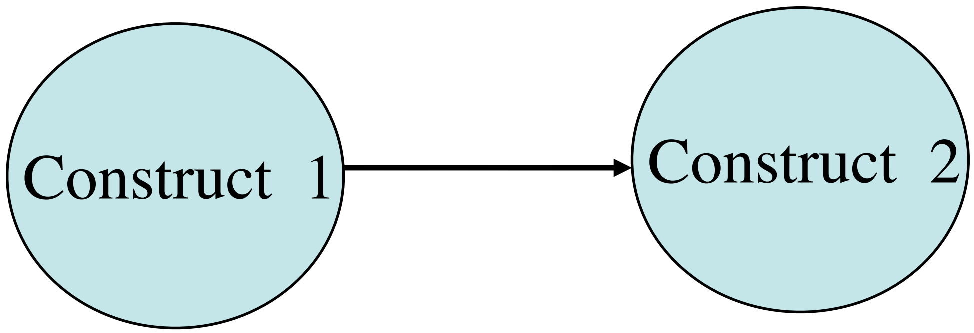
Arousal and Working Memory I

- An investigator was interested in the effect of arousal on short term memory. The hypothesis being tested was that caffeine induced arousal helps short term memory. Subjects were given a list of 20 words to study for 2 minutes, and were then asked to count backwards by 3s from 91. They were then asked to recall as many of the words as possible. The average number of words recalled was 10 (sd=3)
- After the recall was completed, subjects were given 200 mg of caffeine and allowed to read for 30 minutes while the caffeine took effect. They were then given the same list to study for 2 minutes, followed by counting backwards again from 91. They were then asked to recall as many words as possible from the list. The average this time was now 12 (sd=3). There were 20 subjects in this within subject experiment and the t-test of the correlated differences was 3.6 (d.f. =19, $p<.01$).
- From these results, the investigator concluded that the hypothesis that caffeine induced arousal helps working memory was supported.
- Do these results follow?
- Can you think of an alternative explanation for the results?
- How would design a study to control for this alternative explanation?

Questions for evaluating research

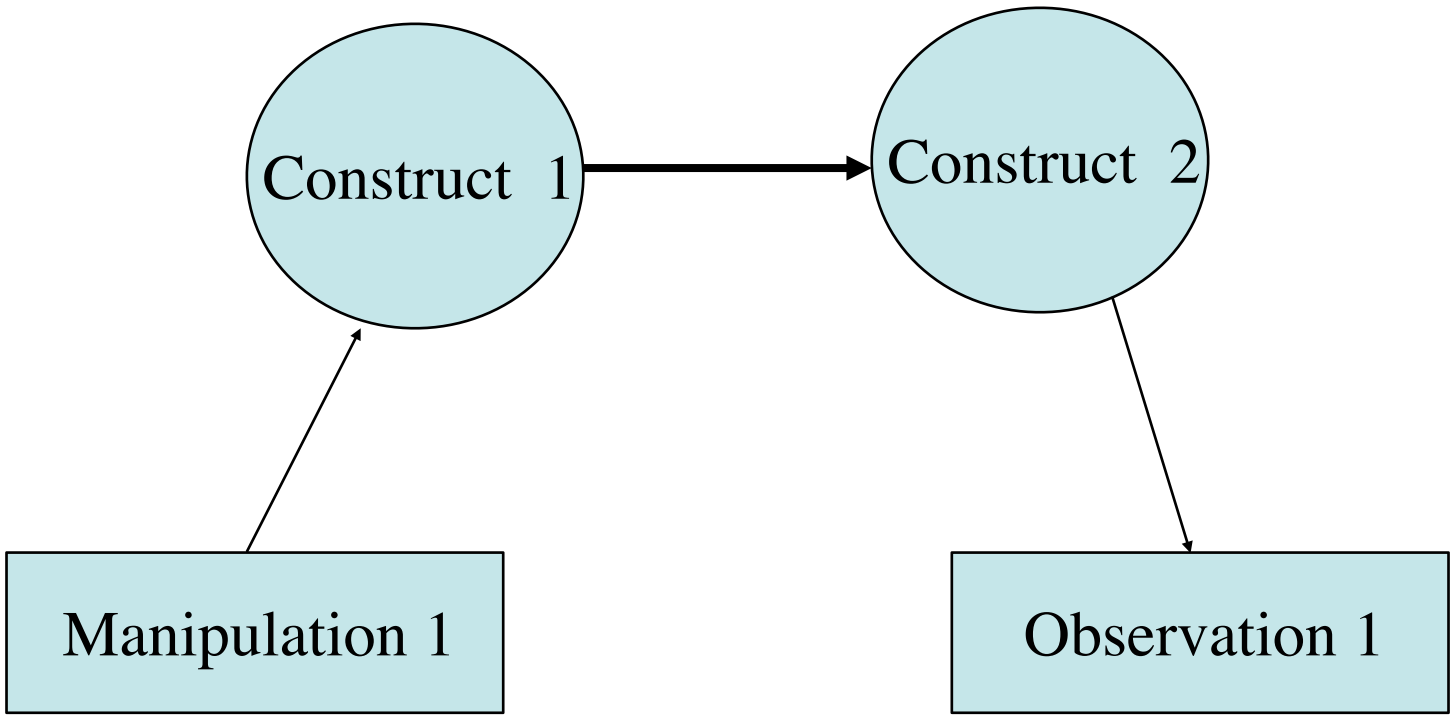
- What are the basic constructs being studied?
- What are the particular operationalizations (observations) associated with the constructs?
- How much of the variability in a construct is due to the (experimental manipulation) independent variable?
- What are possible alternative sources of variation?

Theory and Theory Testing I: Theory



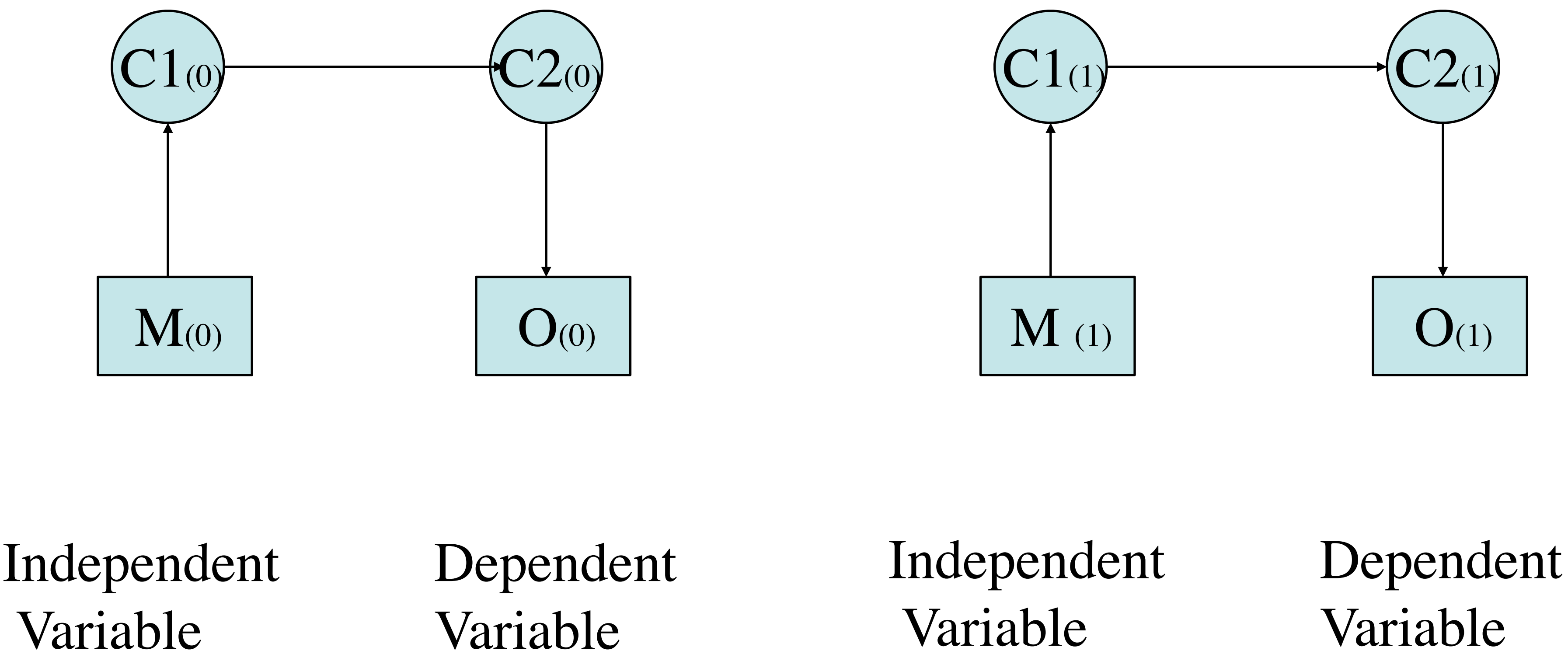
What are the constructs of interest?

Theory and Theory Testing II: Experimental manipulation



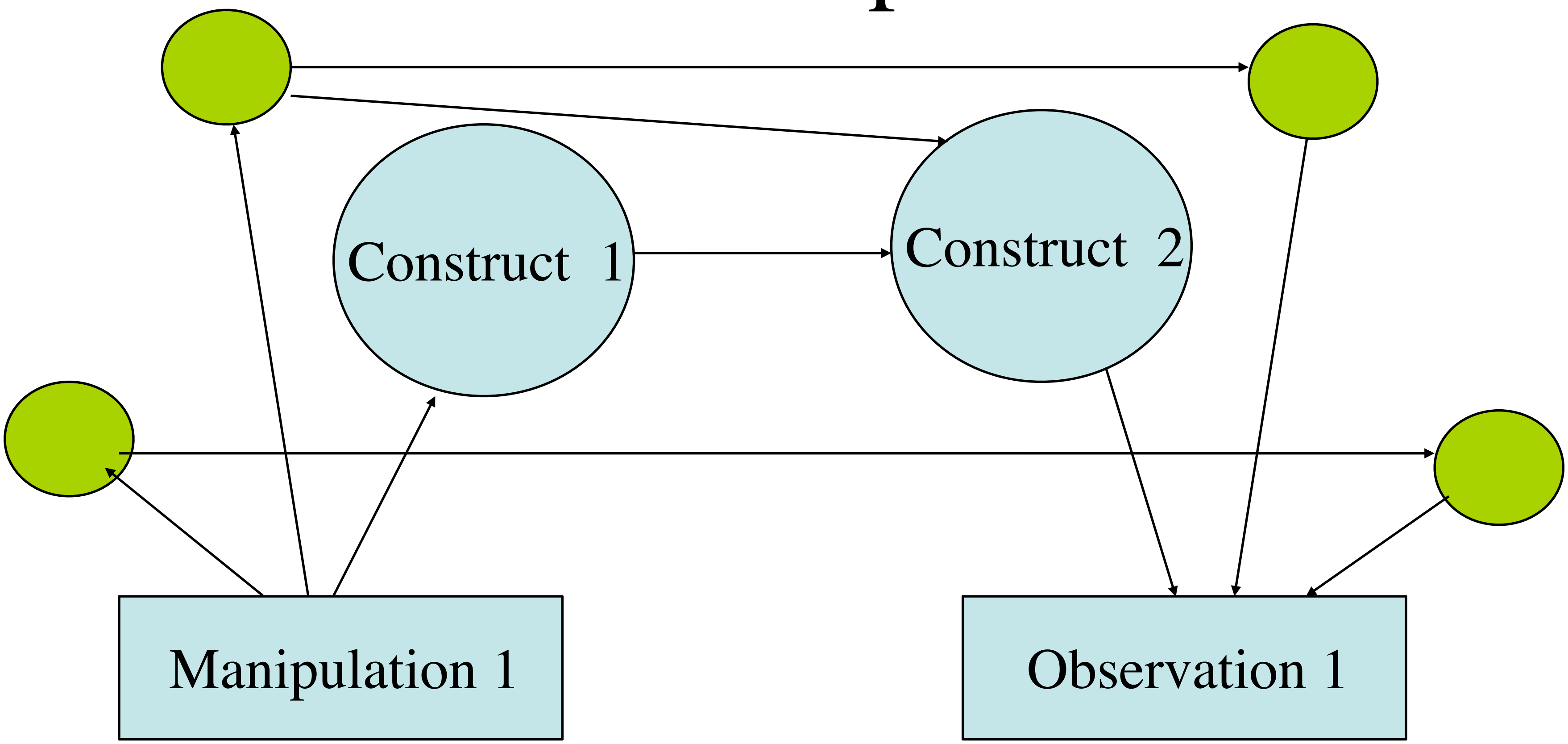
How are the constructs measured/manipulated?

Theory and Theory Testing II: Experimental manipulation



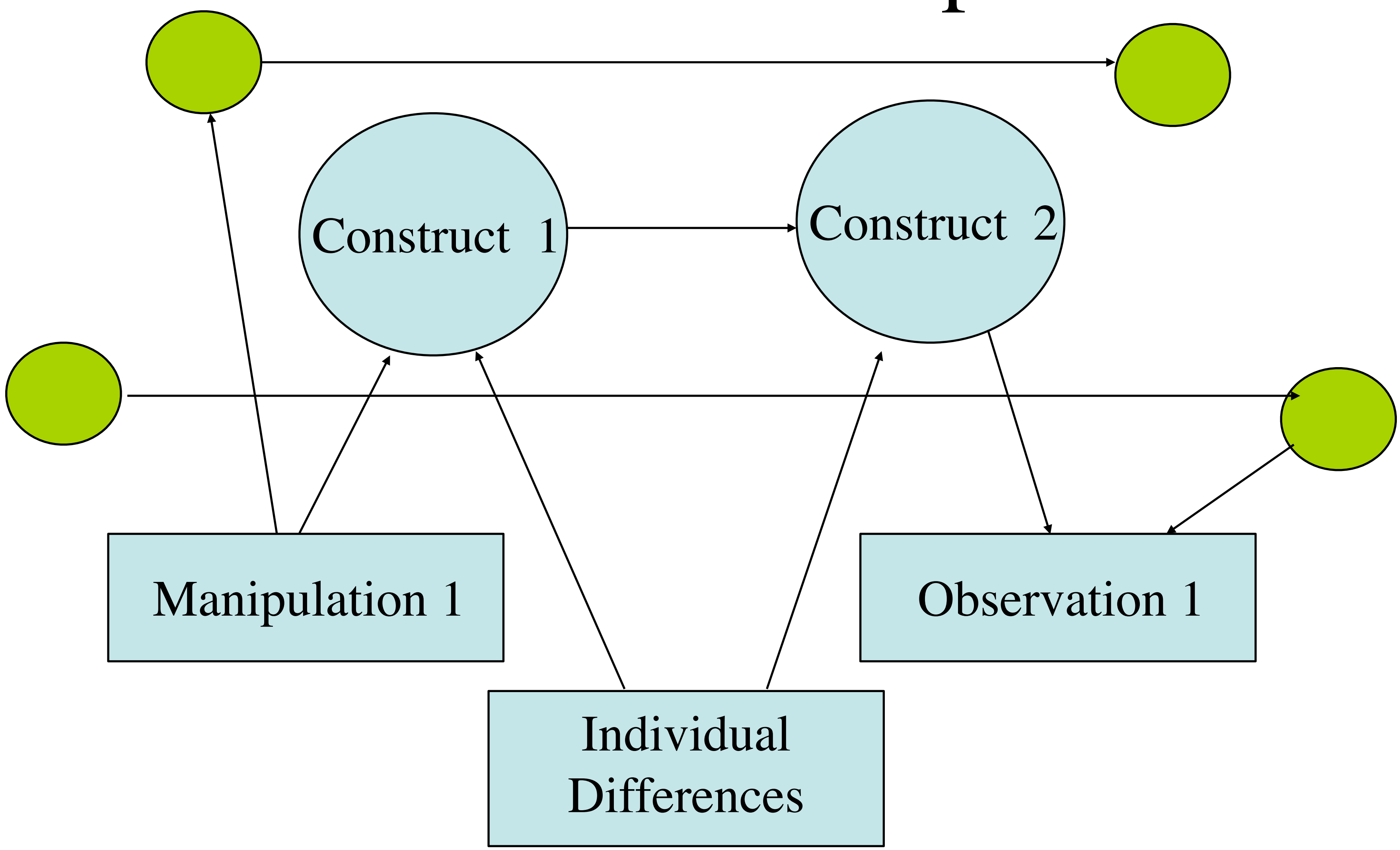
How are the constructs measured/manipulated?

Theory and Theory Testing III: Alternative Explanations



- What are possible alternative sources of variation?

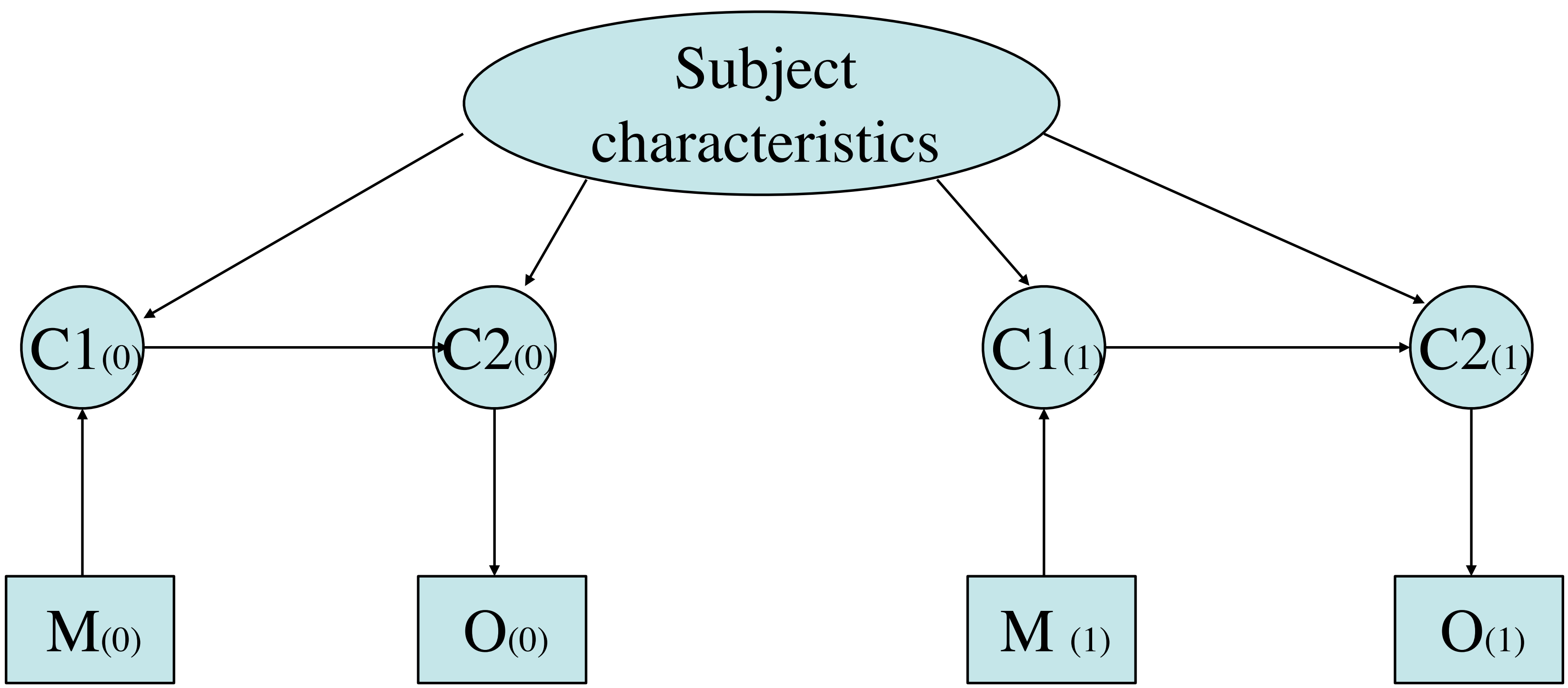
Theory and Theory Testing IV: Eliminate Alternative Explanations



How to control for variability

- Between subject variability
  - People differ because of
    - ability, motivation, practice
  - Use person as their own control
- Within subject variability
  - control for order effects
    - fatigue
    - learning
  - Use counterbalancing

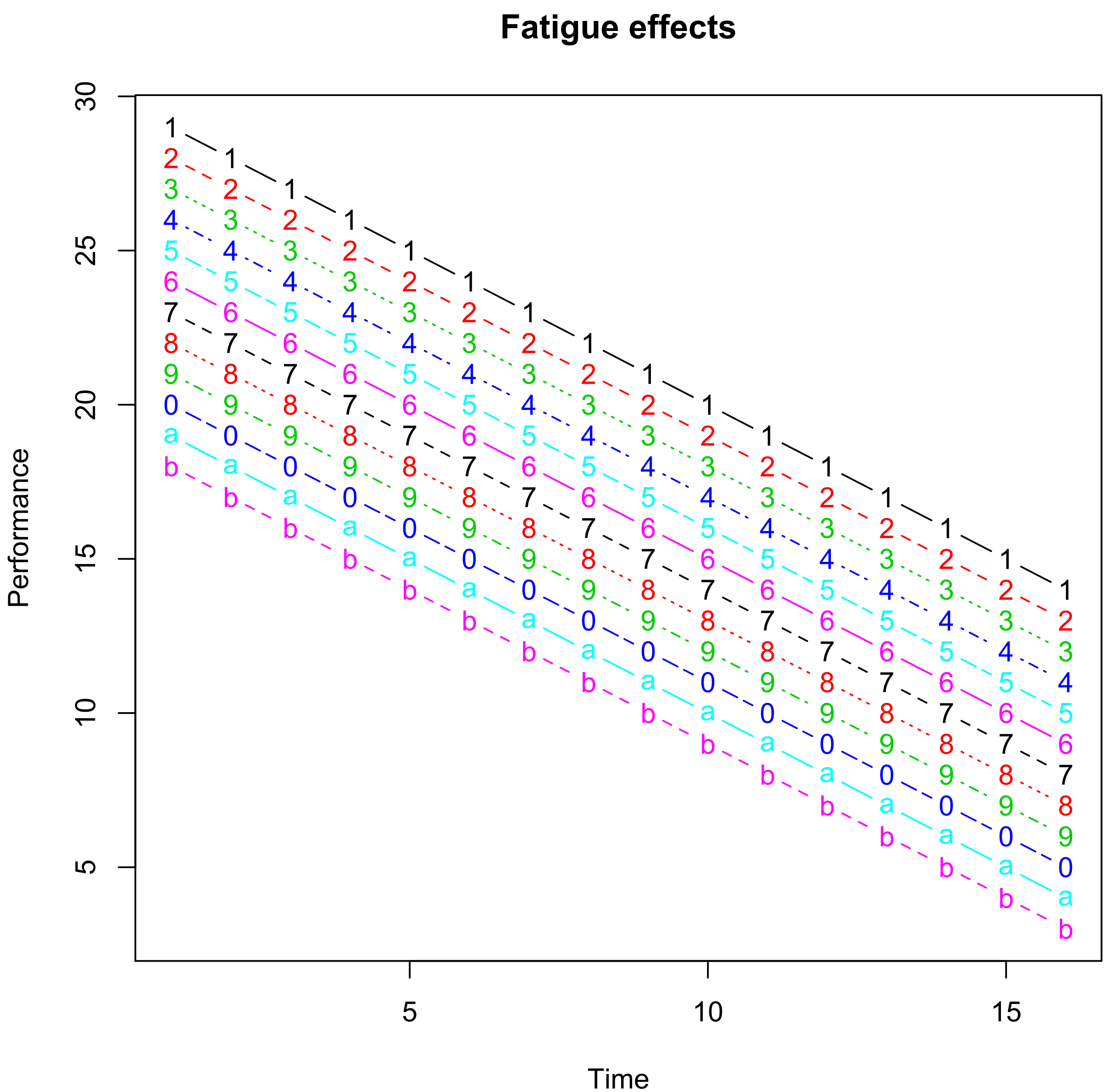
Theory and Theory Testing II: Experimental manipulation- Within Subjects



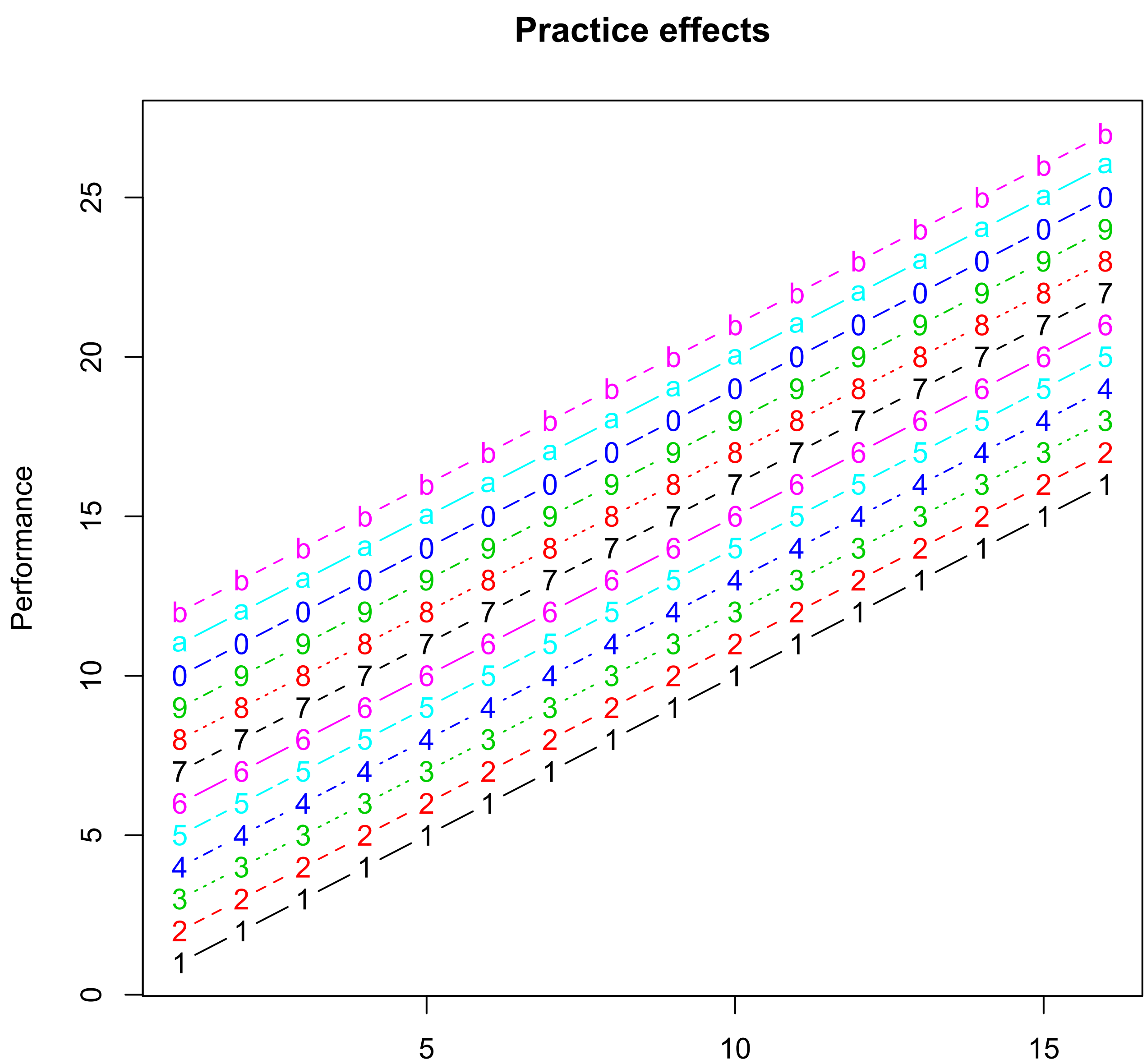
Threats to validity of within subject designs

- If we have two or more conditions, then we need to worry about order effects.
  - Practice
  - Fatigue
  - Carryover
- Can use complete within subject design and counter balance
- Mixed within-between to control for materials

Fatigue -> decrease over time



Practice -> Improvement over time



Hypothetical data

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
S1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
S2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
S3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
S4	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
S5	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
S6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
S7	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
S8	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
S9	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
S10	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
S11	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
S12	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Consider alternative designs

- First half versus 2nd half
  - mean(my.data[,1:8]) 10
  - mean(my.data[,9:16]) 18
- Odds vs. Even
  - mean(my.data[,c(1,3,5,7,9,11,13,15)]) 13.5
  - mean(my.data[,-c(1,3,5,7,9,11,13,15)]) 14.5
- ABBA counter balance
  - mean(my.data[,c(1,4,5,8,9,12,13,16)]) 14
  - mean(my.data[,-c(1,4,5,8,9,12,13,16)]) 14

Experimental Designs

- Within Subjects
  - Controls for subject variability
  - Sensitive to within subject changes such as fatigue, learning, differential transfer
- Between subjects
  - Controls for within subject changes
  - Sensitive to between subject variability
    - Effects due to subject selection, attrition, randomization
- Mixed designs
  - Controls for materials effect (i.e., are some word lists easier to learn)

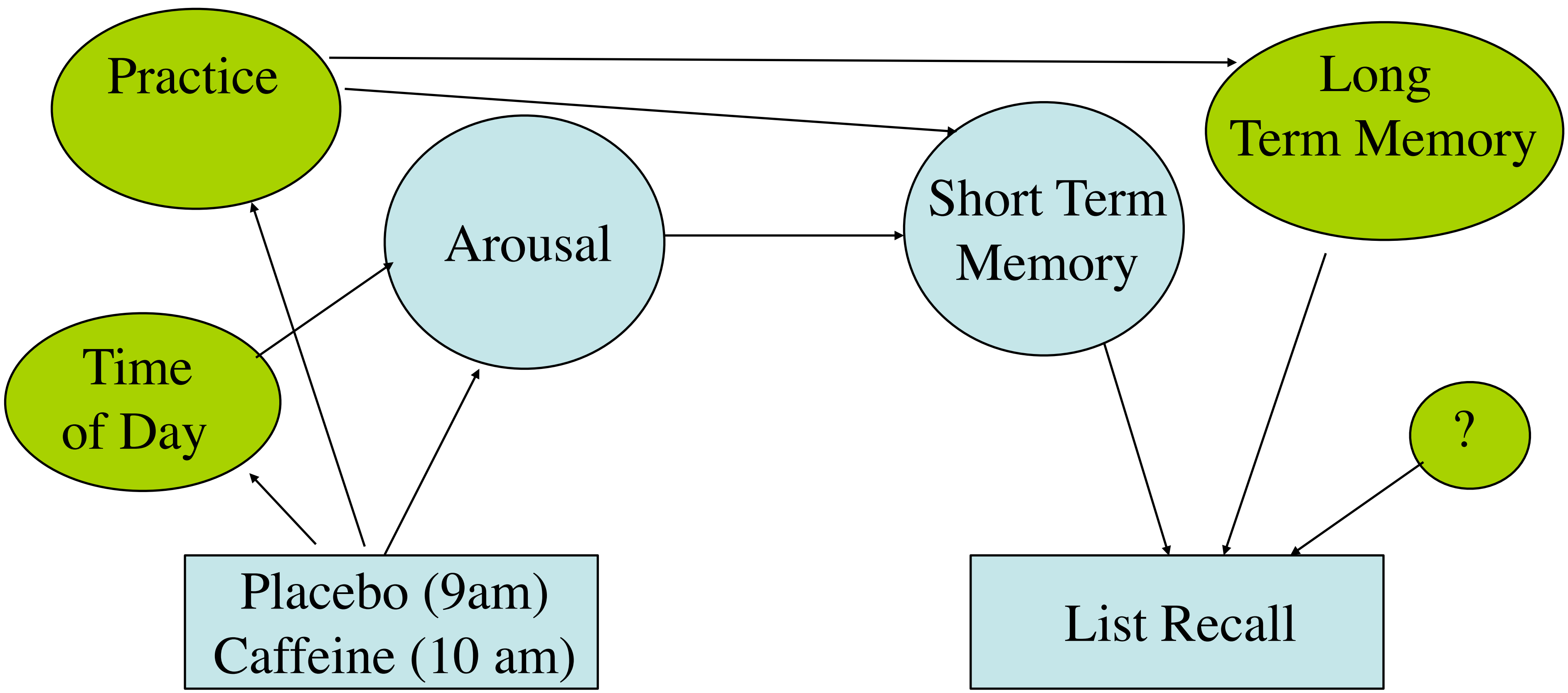
Analysis of any study

- What are the constructs of interest?
- How are they measured/manipulated?
- What are possible alternative sources of variation?
  - Within subjects threats
  - Between subject threats
- How strong is the relationship between the manipulation/observation of the IV and the measurement of the DV?

Arousal and Working Memory

- Hypothesis
  - Alertness (arousal) facilitates short term memory
- Constructs
  - Arousal
  - Short Term Memory (memory for very recent events)
- Manipulations/Observables
  - Caffeine increases arousal
  - Study list - Filled Delay interval (why)
  - Immediate List recall
- Alternative Explanations

Confounded Within Subject design



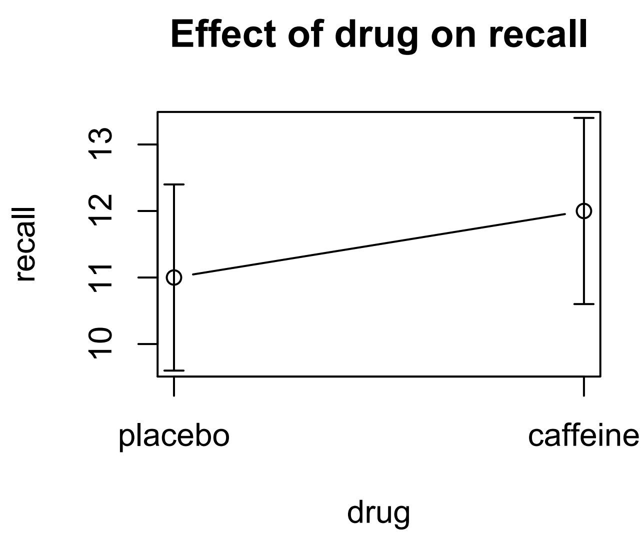
- What are possible alternative sources of variation?

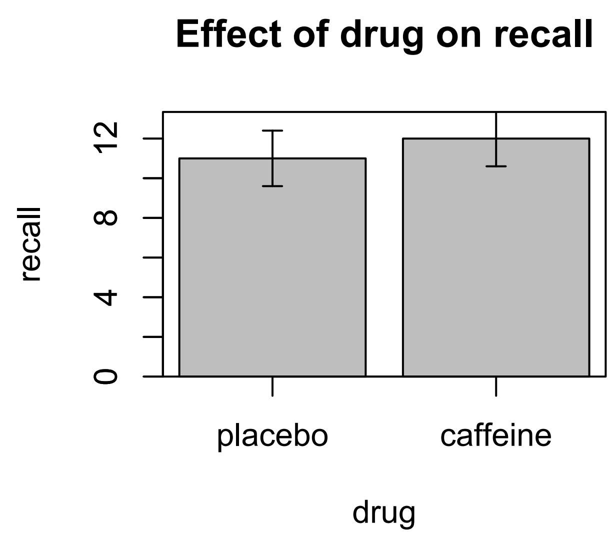
Arousal and Working Memory II

- Another investigator was interested in the effect of caffeine induced arousal on short term memory. The hypothesis being tested was that caffeine induced arousal helps short term memory. To control for time of day effects, all subjects were run at 8 am. Subjects were given a list of 20 words to study for 2 minutes and were then asked to count backwards from 91 by 3s. They were then asked to recall as many of the word as possible. The average number of words recalled was 11 (sd=3.)
- After the recall was completed, subjects were allowed to read quietly for an hour in order to minimize any possible carry over from the previous trial. Then the participants were given 200 mg. of caffeine and then allowed to read for 30 minutes while the caffeine took effect. They were then given a new list of words to study for 2 minutes, followed by counting forwards by 7s from 17. they were then asked to recall as many words as possible from the list. The average this time was now 12 (sd=2.5). With 30 subjects, this difference had a t-test of correlated differences of 2.8, df=29, $p<.01$.
- From the results of this within subject study, the investigator concluded that the hypothesis that caffeine induced arousal helps working memory as supported.

Study 2

```
> print(study2.df,digits=2)
         mean  sd  n min max   se
placebo    11 3.0 20   4  16 0.67
caffeine   12 2.5 20   6  18 0.67
```





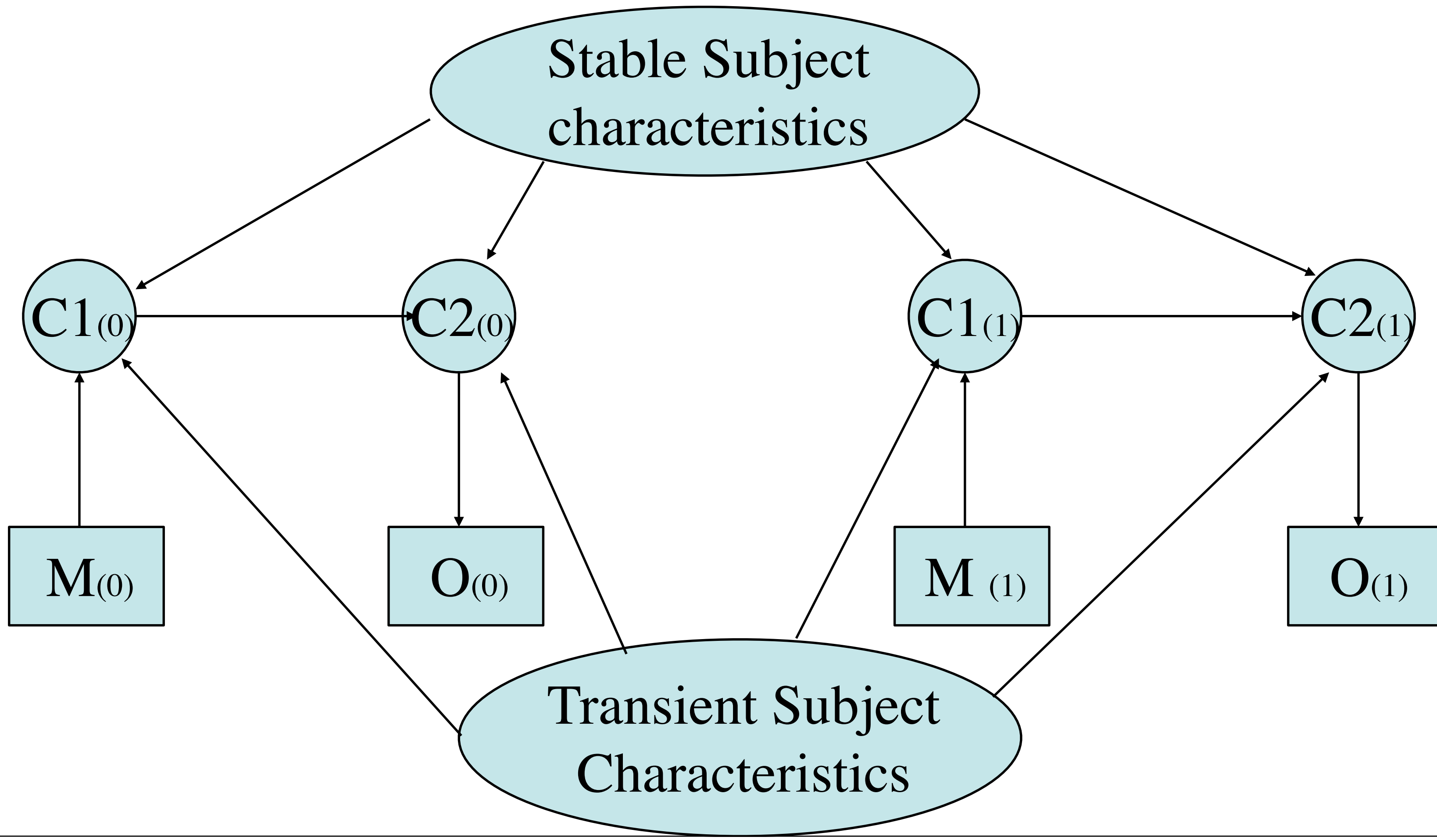

```
error.bars(stats=study2.df,ylab="recall",xlab="drug",main="Effect of drug on recall",typ="b")
error.bars(stats=study2.df,ylab="recall",xlab="drug",main="Effect of drug on recall",bars=TRUE)
```

Arousal and Working Memory II

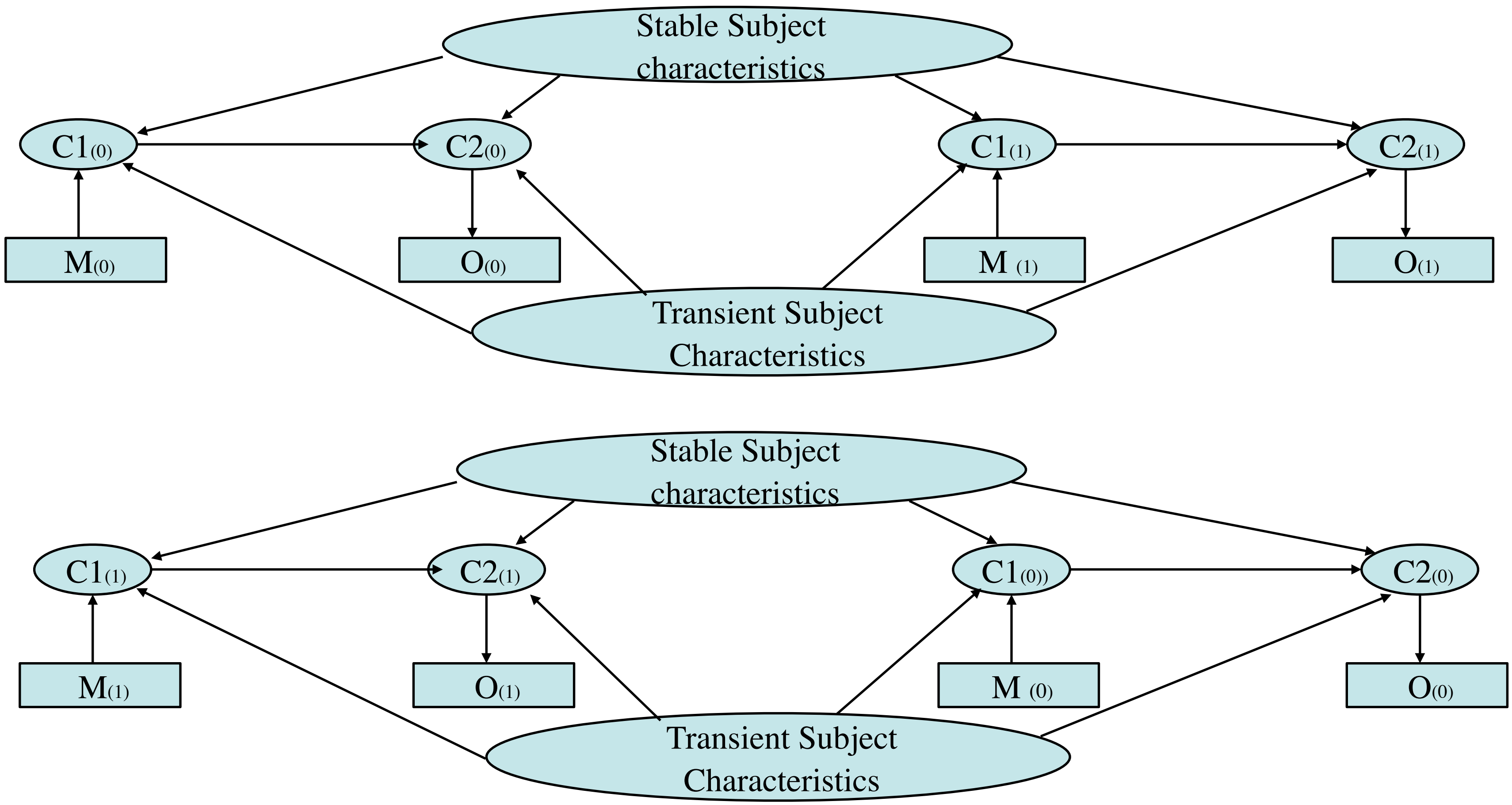
- Another investigator was interested in the effect of caffeine induced arousal on short term memory. The hypothesis being tested was that caffeine induced arousal helps short term memory. To control for time of day effects, all subjects were run at 8 am. Subjects were given a list of 20 words to study for 2 minutes and were then asked to count backwards from 91 by 3s. They were then asked to recall as many of the word as possible. The average number of words recalled was 11 (sd=3.)
- After the recall was completed, subjects were allowed to read quietly for an hour in order to minimize any possible carry over from the previous trial. Then the participants were given 200 mg. of caffeine and then allowed to read for 30 minutes while the caffeine took effect. They were then given a new list of words to study for 2 minutes, followed by counting forwards by 7s from 17. they were then asked to recall as many words as possible from the list. The average this time was now 12 (sd=2.5). With 30 subjects, this difference had a t-test of correlated differences of 2.8, df=29, p<.01.
- From the results of this within subject study, the investigator concluded that the hypothesis that caffeine induced arousal helps working memory as supported.

- Do these results follow?
- Can you think of an alternative explanation for the effects?
- How would design a study to control for this alternative explanation?

Theory and Theory Testing II: Experimental manipulation- Within Subjects



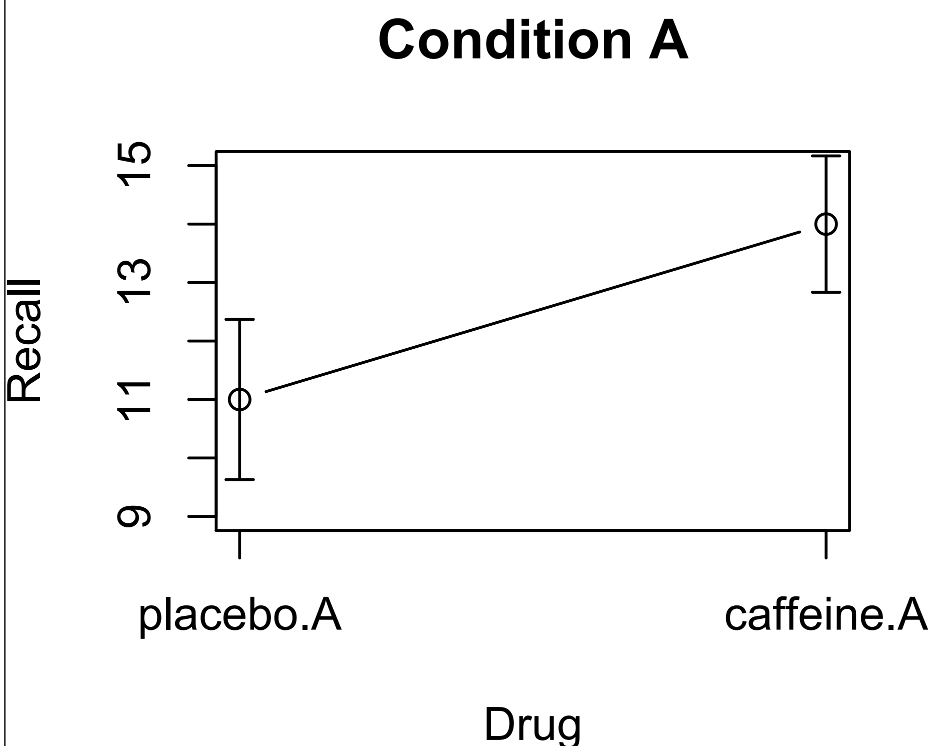
Theory and Theory Testing II: Experimental manipulation-Within Subjects-Counter balancing

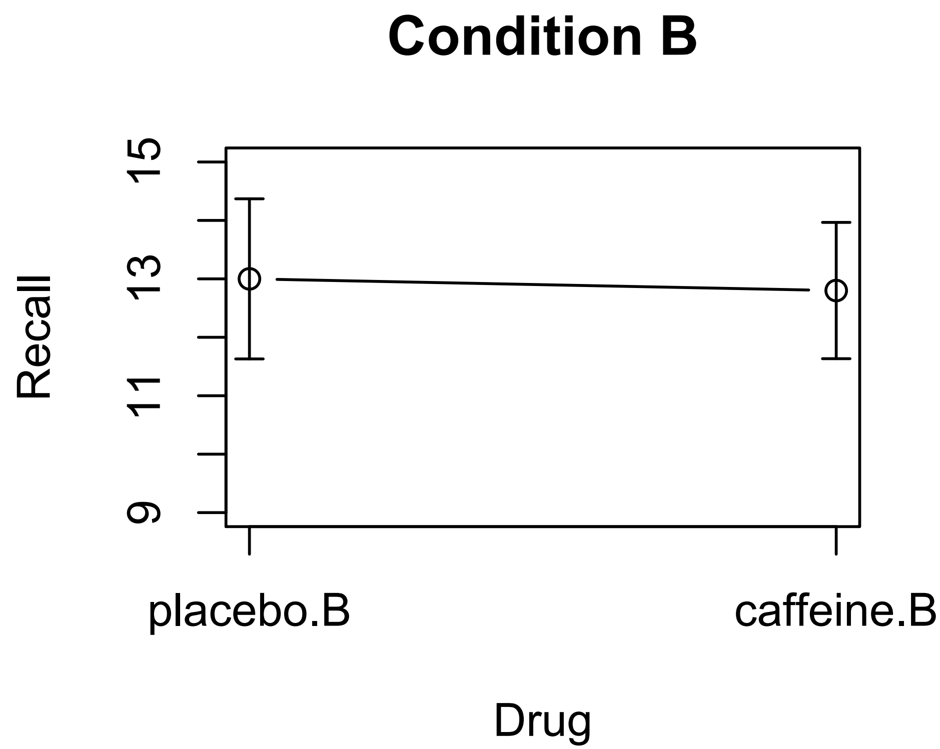


Arousal and Working Memory III

- Yet another investigator was interested in the effect of caffeine induced arousal on short term memory The hypothesis being tested was that caffeine induced arousal helps short term memory. To control for time of day effects, all subjects were run at 8 am.
- However, to control for possible order effects, 1/2 of the participants were run in one within subject condition, the other half in the other condition.
- That is, half were given a list of 20 words to study for 2 minutes and were then asked to count backwards from 91 by 3s. They were then asked to recall as many of the word as possible. The average number of words recalled for this group was 11 (sd=3.) Then the participants were given 200 mg. of caffeine and then allowed to read for 30 minutes while the caffeine took effect. They were then given a new list of words to study for 2 minutes, followed by counting forwards by 7s from 17. they were then asked to recall as many words as possible from the list. The average this time was now 14 (sd=2.5). With 30 subjects, this difference had a t-test of correlated differences of 2.8, df=29, p<.01.
- The other half of the participants were given the caffeine on trial one and not given anything on trial 2. Their performance on trial 1 was 13 (sd=2) and on trial 2 was 12.8 (sd=2). This difference was not reliably different from a chance difference (t=.4 ns.)

	order	mean	sd	n	min	max	se
placebo.A	A	11	3	30	4	16	0.55
caffeine.A	A	14	3	30	6	18	0.55
placebo.B	B	13	2	30	4	16	0.37
caffeine.B	B	13	3	30	6	18	0.55

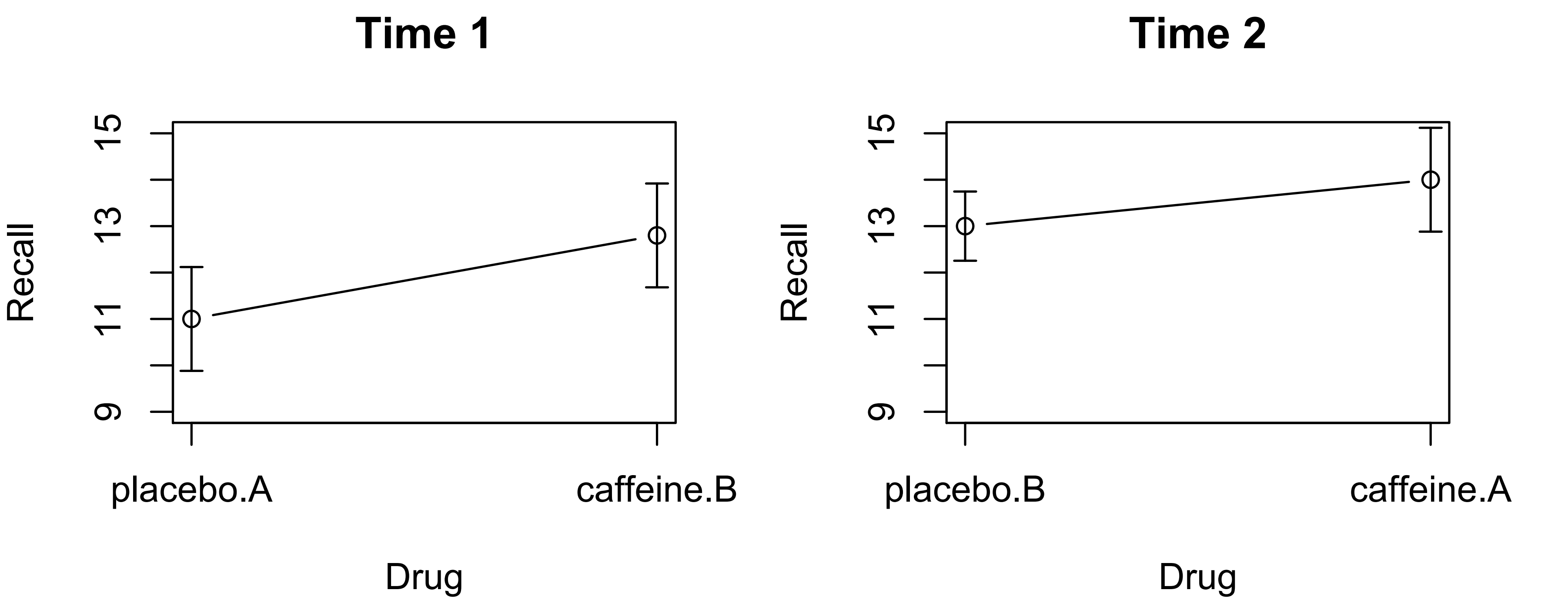




Arousal and Working Memory III

- Yet another investigator was interested in the effect of caffeine induced arousal on short term memory The hypothesis being tested was that caffeine induced arousal helps short term memory. To control for time of day effects, all subjects were run at 8 am.
- However, to control for possible order effects, 1/2 of the participants were run in one within subject condition, the other half in the other condition.
- That is, half were given a list of 20 words to study for 2 minutes and were then asked to count backwards from 91 by 3s. They were then asked to recall as many of the word as possible. The average number of words recalled for this group was 11 (sd=3.) Then the participants were given 200 mg. of caffeine and then allowed to read for 30 minutes while the caffeine took effect. They were then given a new list of words to study for 2 minutes, followed by counting forwards by 7s from 17. they were then asked to recall as many words as possible from the list. The average this time was now 14 (sd=2.5). With 30 subjects, this difference had a t-test of correlated differences of 2.8, df=29, p<.01.
- The other half of the participants were given the caffeine on trial one and not given anything on trial 2. Their performance on trial 1 was 13 (sd=2) and on trial 2 was 12.8 (sd=2). This difference was not reliably different from a chance difference (t=.4 ns.)
- Although the one order showed the effect and the other did not, the investigator then pooled the data from the two orders and found that the caffeine condition in general led to better performance. (mean caffeine = 13, mean control = 11.9). From these results the investigator concluded that the hypothesis that caffeine induced arousal helps working memory as supported.
- Do these results follow?
- Can you think of an alternative explanation for the effects?
- Can you think of an explanation for the difference between the two orders?

A mixed design can be analyzed as a between design

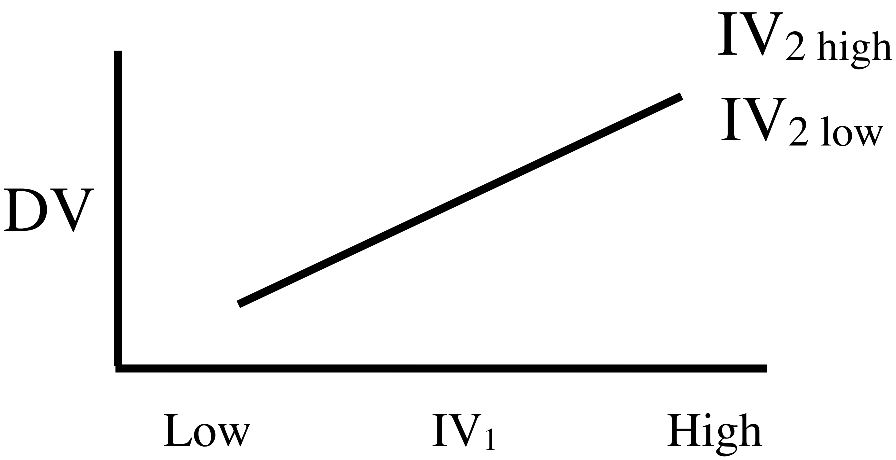


Two variables - 3 analyses

- When we study two variables at the same time, we can ask three different questions:
  - Is there an effect of Variable 1?
  - Is there an effect of Variable 2?
  - Does the effect of Variable 1 depend upon Variable 2 (do they interact)?
- Typically discussed in terms of analysis of variances, but can also be done in terms of regressions --
  - The question is do the slopes differ from 0 and from each other?

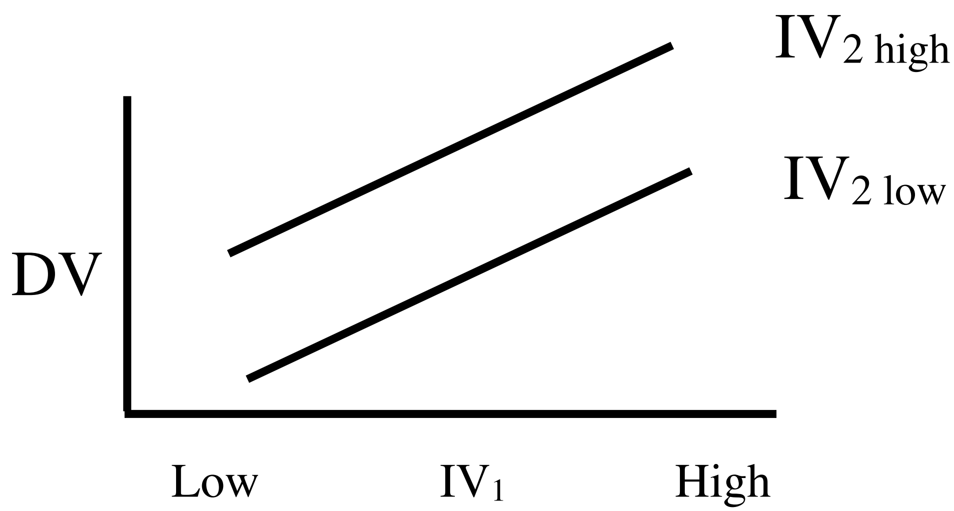
Types of results

Main effect of IV_1

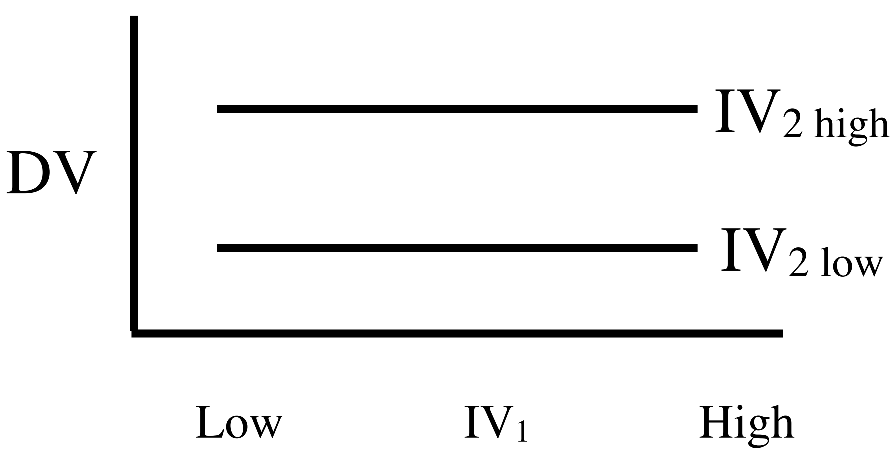


Main effect of IV_1

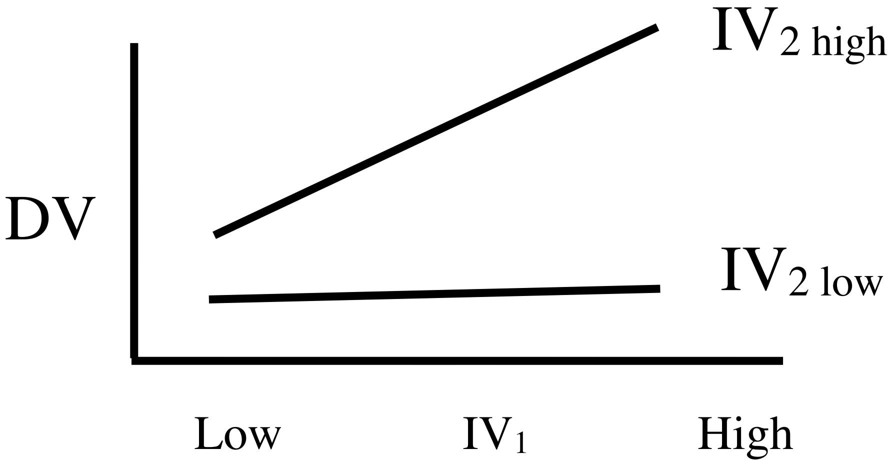
Main effect of IV_2



Main effect of IV_2

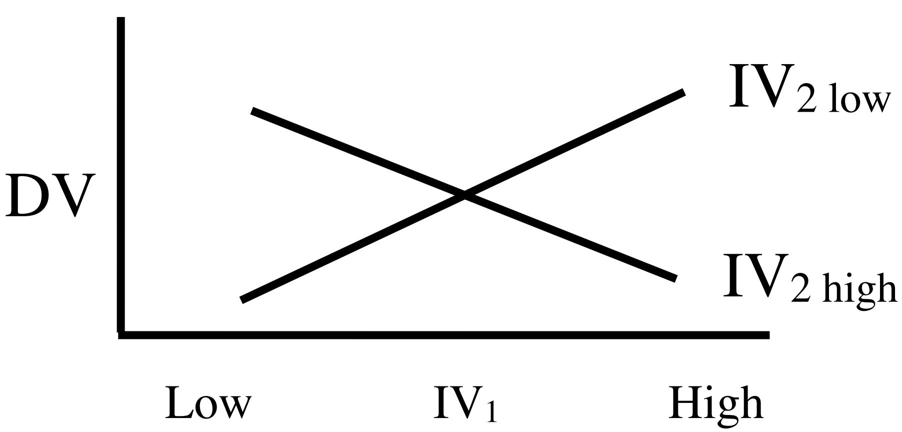


Types of interactions



Effect of IV_2 depends upon that of IV_1

Fan Fold



Cross over

Inferential power of an interaction

- Main effect of a variable shows that there is a relationship between IV and DV.
- Interaction of two IVs with DV means that the effect of one IV depends upon the other IV.
- By having an interaction, we are able to specify the limits of our effects.
- Interactions allow more powerful inference, for they can exclude more alternative hypotheses

Earliest known example of a within subject study with a cross over interaction (double dissociation)

- Gideon was an early methodologist who understood principles of good design (Judges 6:36-40)
- Day 1: Make the wool wet, keep the floor dry
  - alternative explanations for effect
- Day 2: reverse conditions: keep the wool dry, make the floor wet
  - by having a reversal, it is harder to explain effect

Gideon's tests for God are an early example of a double dissociation and probably the first published example of a cross over interaction.   On the first night, the wool was wet but the floor was dry. On the second night, the floor was wet but the wool was dry (Judges 6:36-40)

Experimental Designs

- Within Subjects -- Every subject is own control
  - Every subject is a complete experiment
  - Controls for subject variability
    - Ability
    - Motivation
  - Sensitive to within subject changes
    - Fatigue
    - Learning
  - Counterbalancing controls for some transient effects but is open to threats of
    - Differential transfer

Varieties of Counterbalancing

- Within subject counterbalancing
  - ABBA and BAAB controls for linear order effects but not transfer
  - Within subject randomization if many trials
    - possible to do block randomization
- Complete counterbalancing across subjects
  - One order for each subject, all orders appear
    - Two conditions: two Orders AB BA
    - Three conditions: six orders
      - ABC, ACB, BAC, BCA, CAB, CBA
    - Four conditions, 24 orders! N of orders =C!

Example of within subject counterbalancing

- Class replication of Roediger and McDermott
- How to examine presentation modality and recall vs. math within subjects
- Why not do between subjects?
  - consider subject "cost"
  - also consider sources of between subject error
- how to study several variables at a time within subjects
  - need to manipulate IV_1 and IV_2 independently

How to study several within subject variables at the same time

- Counterbalancing to avoid confounding
  - IV_1 and IV_2 are experimentally independent
- Conditions crossed with conditions
  - All conditions for IV_1 occur with all conditions of IV_2
  - no systematic relationship between IV_1 and IV_2
- Conditions balanced across orders of presentation

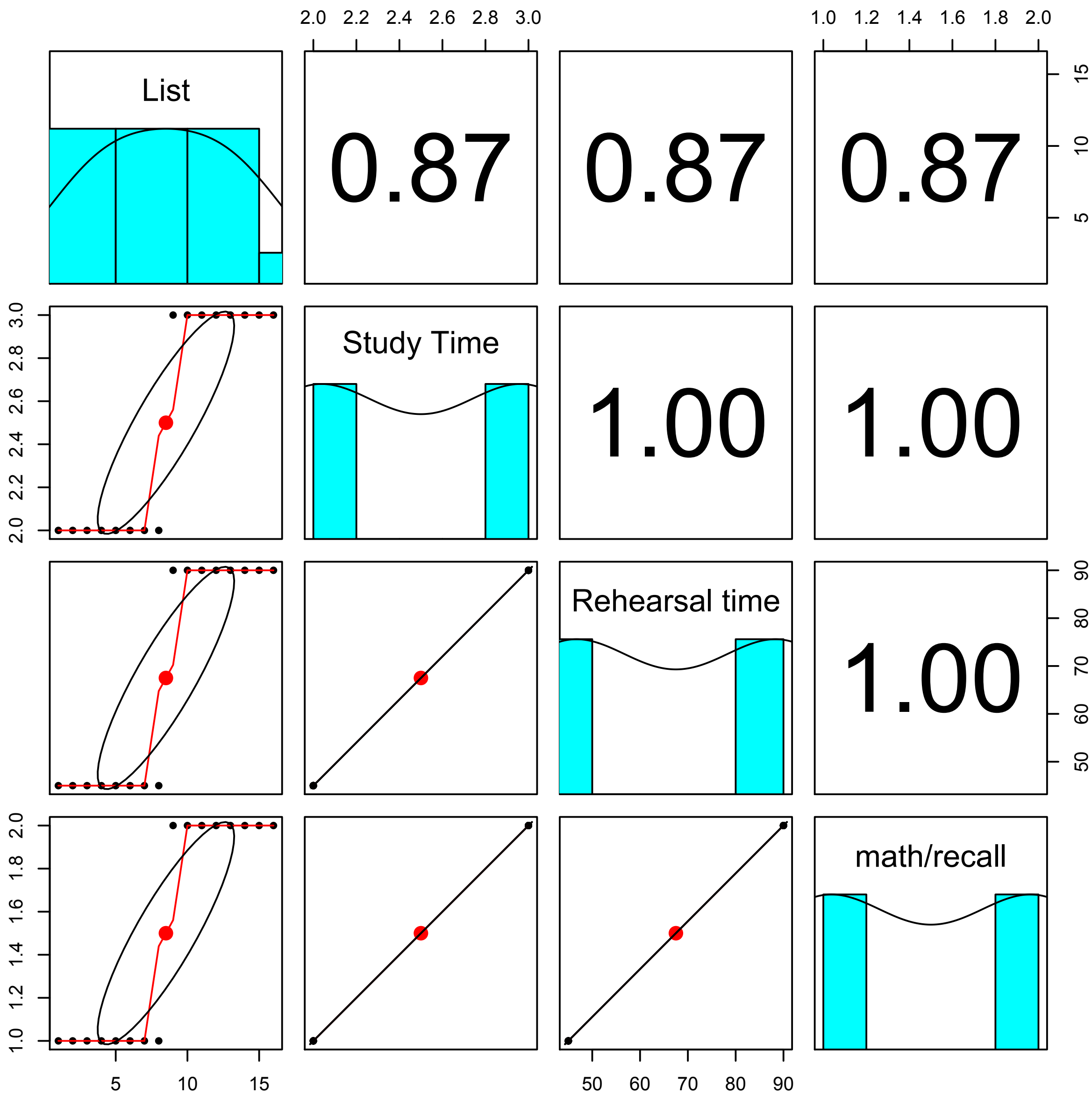
Purpose of counterbalancing

- Conditions are independent of order and of each other
- This allows us to determine effect of each variable independently of the other variables.
- If conditions are related to order or to each other, we are unable to determine which variable is having an effect

Complete Confounding! Math and study and recall time from another design

	List	Study Time	Rehearsal time	math/recall
1	1	2	45	1
2	2	2	45	1
3	3	2	45	1
4	4	2	45	1
5	5	2	45	1
6	6	2	45	1
7	7	2	45	1
8	8	2	45	1
9	9	3	90	2
10	10	3	90	2
11	11	3	90	2
12	12	3	90	2
13	13	3	90	2
14	14	3	90	2
15	15	3	90	2
16	16	3	90	2

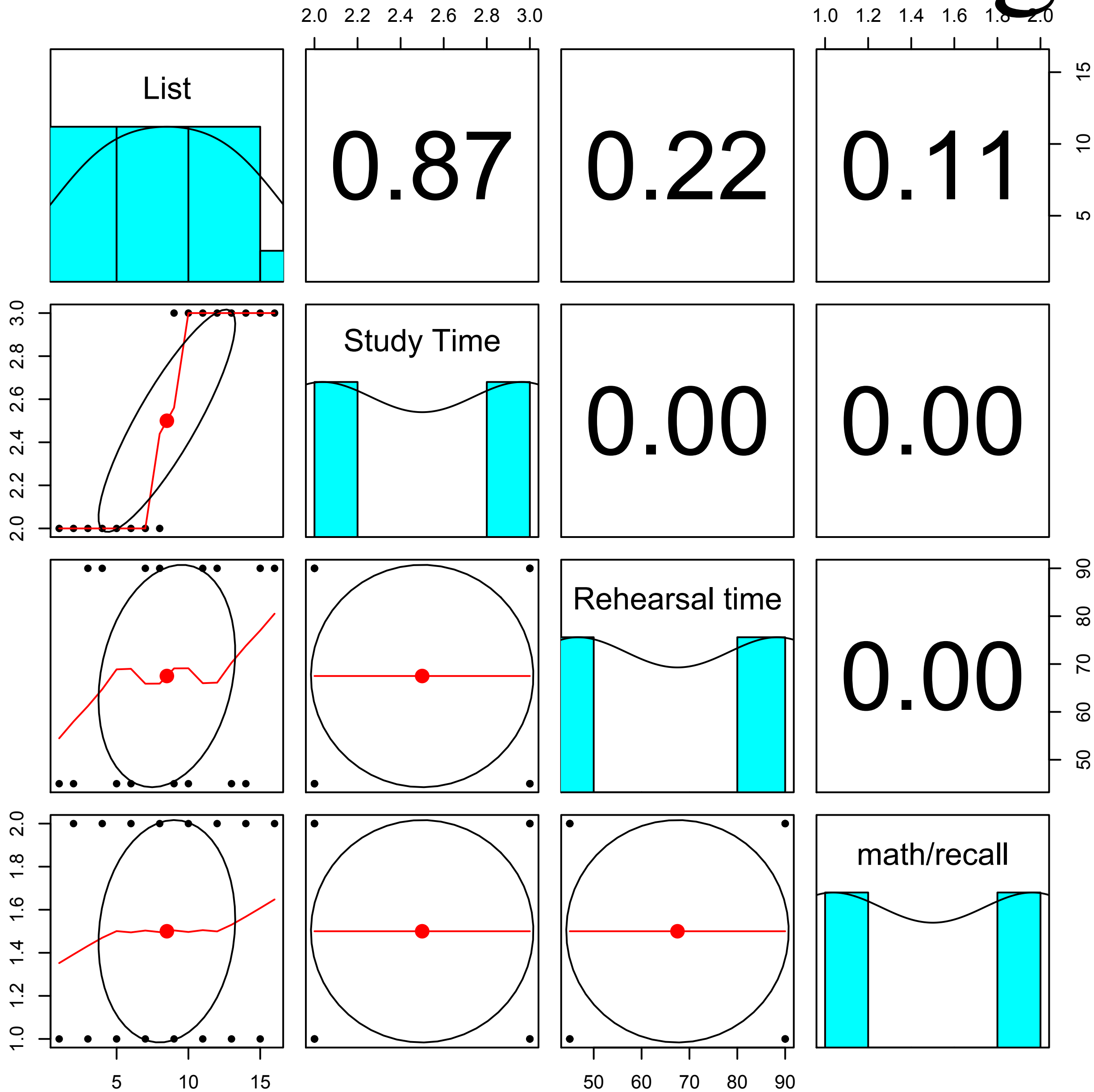
Complete confounding



Partial Confounding
variables confounded with order

	List	Study Time	Rehearsal time	math/recall
1	1	2	45	1
2	2	2	45	2
3	3	2	90	1
4	4	2	90	2
5	5	2	45	1
6	6	2	45	2
7	7	2	90	1
8	8	2	90	2
9	9	3	45	1
10	10	3	45	2
11	11	3	90	1
12	12	3	90	2
13	13	3	45	1
14	14	3	45	2
15	15	3	90	1
16	16	3	90	2

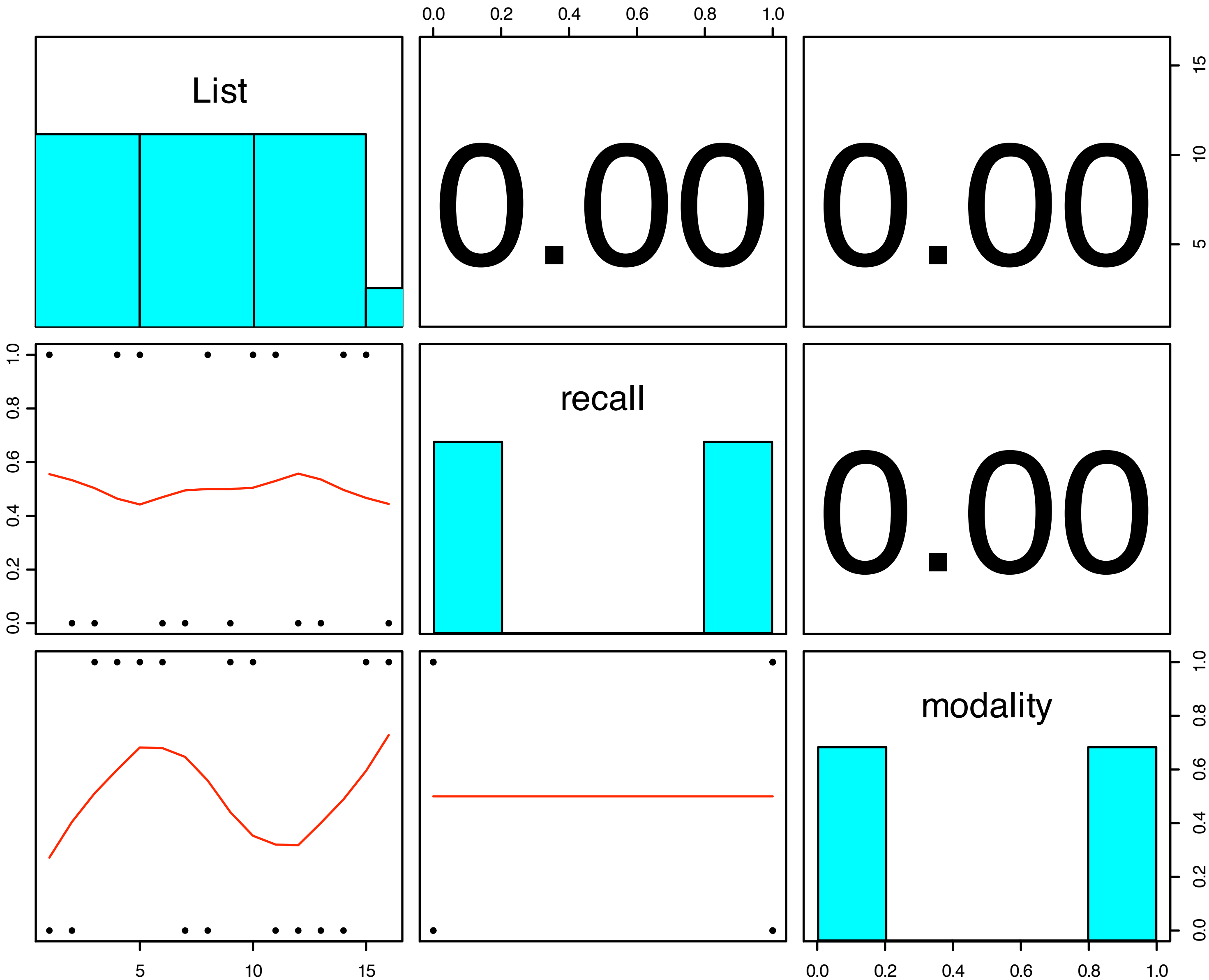
Partial confounding



Class Design- counterbalancing

List	Modality (within)	A/B (between)
1	Visual	Recall
2	Visual	Math
3	Aural	Math
4	Aural	Recall
5	Aural	Recall
6	Aural	Math
7	Visual	Math
8	Visual	Recall
9	Aural	Math
10	Aural	Recall
11	Visual	Recall
12	Visual	Math
13	Visual	Math
14	Visual	Recall
15	Aural	Recall
16	Aural	Math

Design matrix shows no correlations



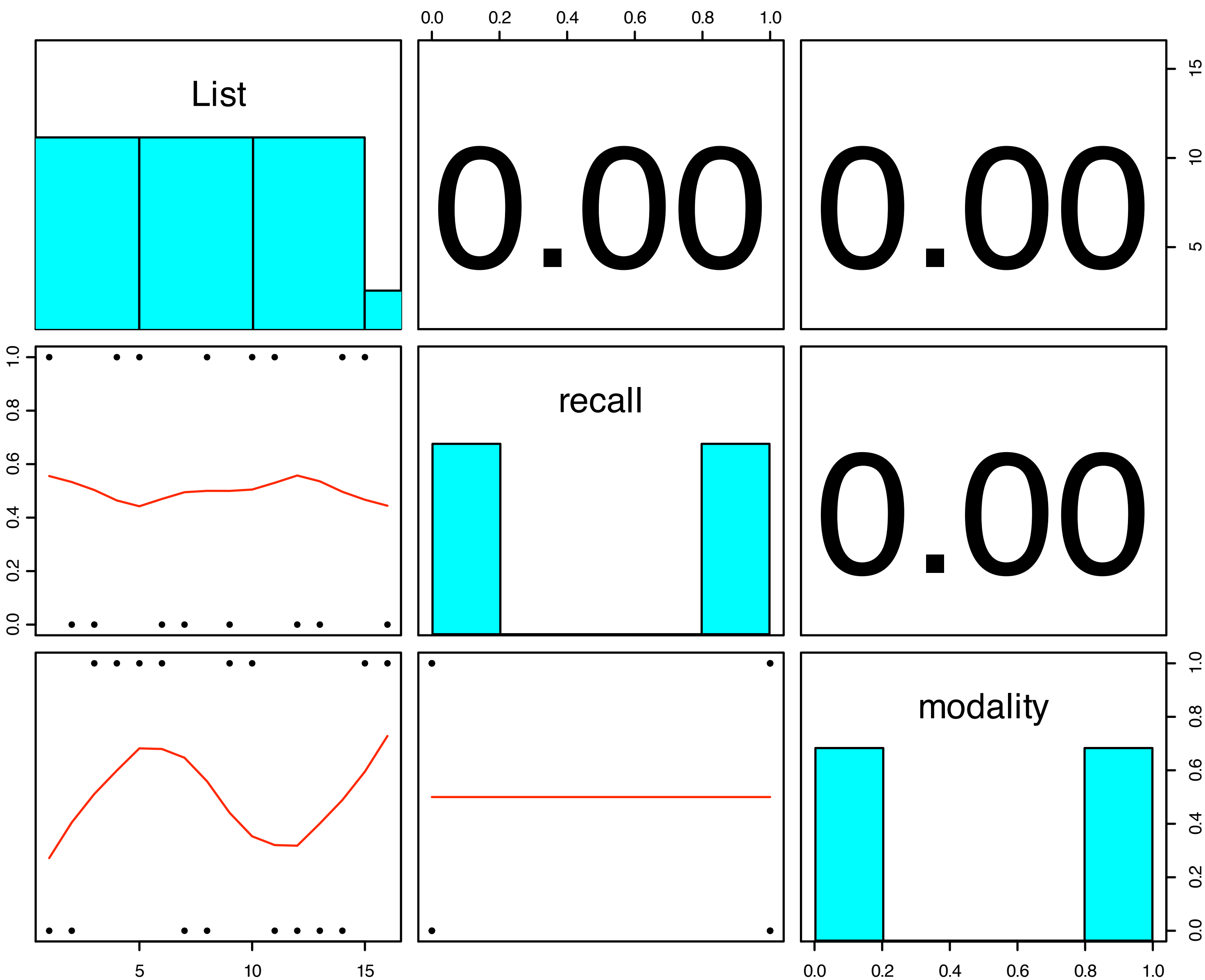
Class Design- counterbalancing

List	Modality (within)	A/B (between)
1	Visual	Recall
2	Visual	Math
3	Aural	Math
4	Aural	Recall
5	Aural	Recall
6	Aural	Math
7	Visual	Math
8	Visual	Recall
9	Aural	Math
10	Aural	Recall
11	Visual	Recall
12	Visual	Math
13	Visual	Math
14	Visual	Recall
15	Aural	Recall
16	Aural	Math

Purpose of counterbalancing

- Conditions are independent of order and of each other
- This allows us to determine effect of each variable independently of the other variables.
- If conditions are related to order or to each other, we are unable to determine which variable is having an effect

Design matrix shows no correlations



Results - Descriptive

- Descriptive statistics vs. Inferential stats
- Describe the results --
  - Say it in words

    Say it in pictures (figures)

    Say it in numbers
- Inferential: What is the likelihood that the results could happen by chance?

    Estimate a parameter and give confidence intervals for that parameter

Results - selective summary

- No need to report every analysis, just the ones that tell the important story
- Think about how to aggregate the data to best summarize it
- Transforms of data to make more understandable
  - e.g., percent correct rather than raw number
- Story must be truth
  - don’t hide “inconvenient data”
  - assume someone else will want to analyze your data

Data= Model + Residual

- The process of science is improve the model and reduce the error
- Models are progressively more complicated
- Consider the recall data:
  - Model 0: Data
  - Model 1: Data = Mean + Residual
  - Model 2: Data = $Position_i$ + Residual
  - Model 3: Data = Type of presentation + Residual
  - ...

Results

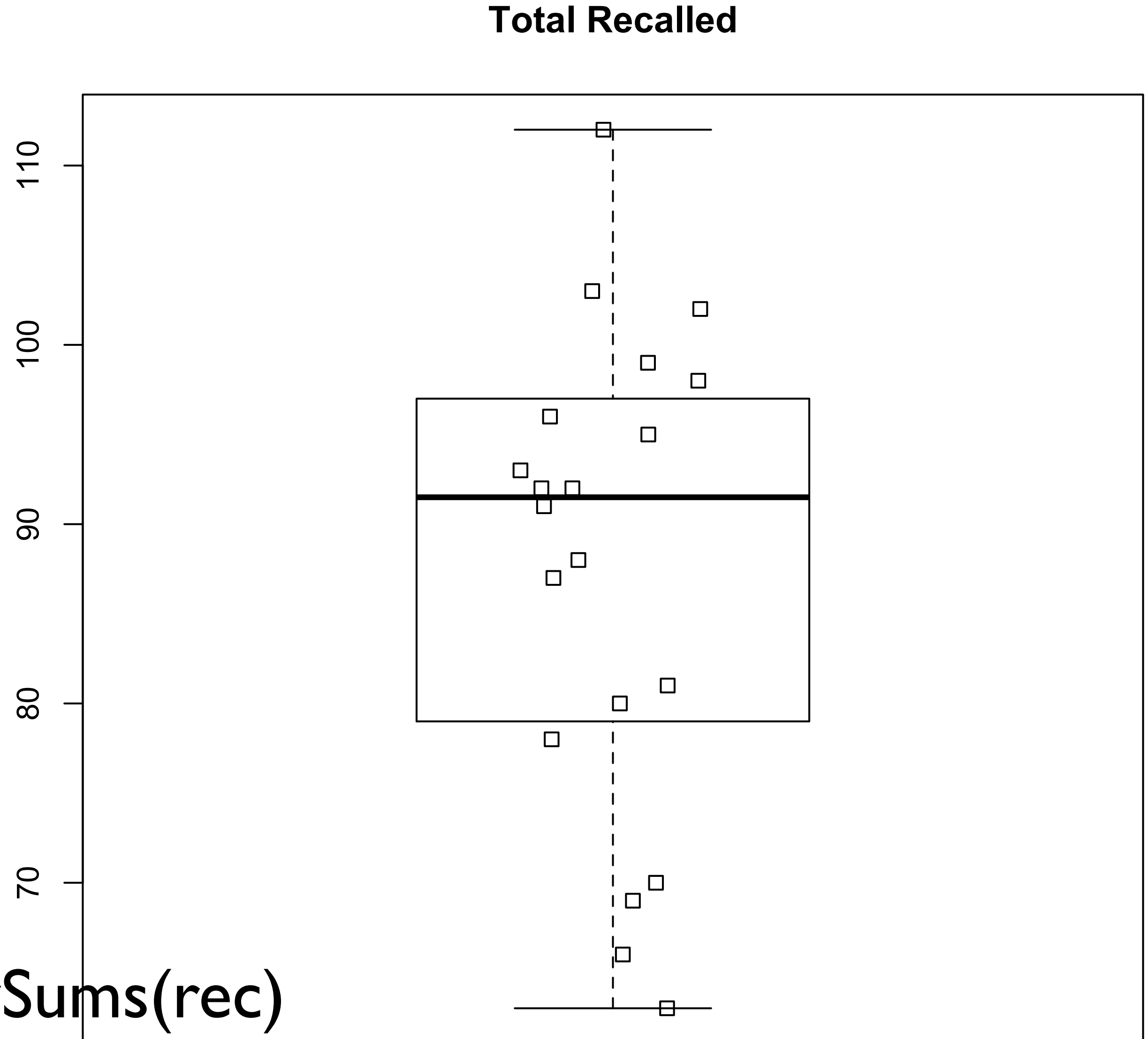
- Recall (manipulation check)
  - Is there a serial position effect?
    - Primacy
    - Recency (particularly given the instructions)
  - Did people just recall on recall tasks?
- Recognition
  - Is there a false memory effect?
  - What manipulations affect it?
  - Are these the same manipulations that affect real recognition?

The raw data by position

```
> rec
```

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
S1	7	7	7	7	6	4	5	8	7	7	7	7	8	7	8
S2	7	7	5	7	6	6	4	6	5	6	5	5	7	8	8
S3	6	6	5	7	8	8	7	5	5	4	5	5	6	6	8
S4	5	5	6	4	7	4	4	4	5	3	4	4	6	3	6
S5	8	7	7	7	6	7	5	4	4	6	3	5	5	7	7
S6	6	8	5	5	5	7	6	7	6	3	6	5	2	5	5
S7	8	8	8	8	7	7	6	6	7	7	6	6	7	5	7
S8	8	7	7	7	8	8	7	6	6	4	6	5	7	7	6
S9	7	6	4	5	5	4	4	3	1	5	3	2	3	5	6
S10	7	6	6	6	4	2	6	5	3	6	4	3	6	6	8
S11	8	4	7	7	5	5	5	5	7	7	7	4	8	6	8
S12	7	7	7	6	6	8	4	7	6	4	5	7	5	6	7
S13	8	8	8	7	5	6	7	4	6	8	7	5	5	6	5
S14	8	6	7	6	7	7	7	3	6	6	6	7	8	6	8
S15	7	5	7	7	6	4	2	3	3	2	2	5	4	5	4
S16	8	6	5	3	7	6	4	1	4	3	1	4	4	5	8
S17	7	8	6	6	7	5	6	5	2	3	6	7	6	5	8
S18	7	8	7	6	5	3	4	2	3	5	7	6	6	6	5
S19	7	6	6	8	6	8	6	8	5	7	5	4	6	6	8
S20	8	8	7	7	7	8	7	7	8	7	7	8	8	7	8

Model 1: Median + Residual




```
total <- rowSums(rec)
boxplot(total,main="Total Recalled")
stripchart(total,method="jitter",vertical=TRUE,add=TRUE)
```

Summary statistics

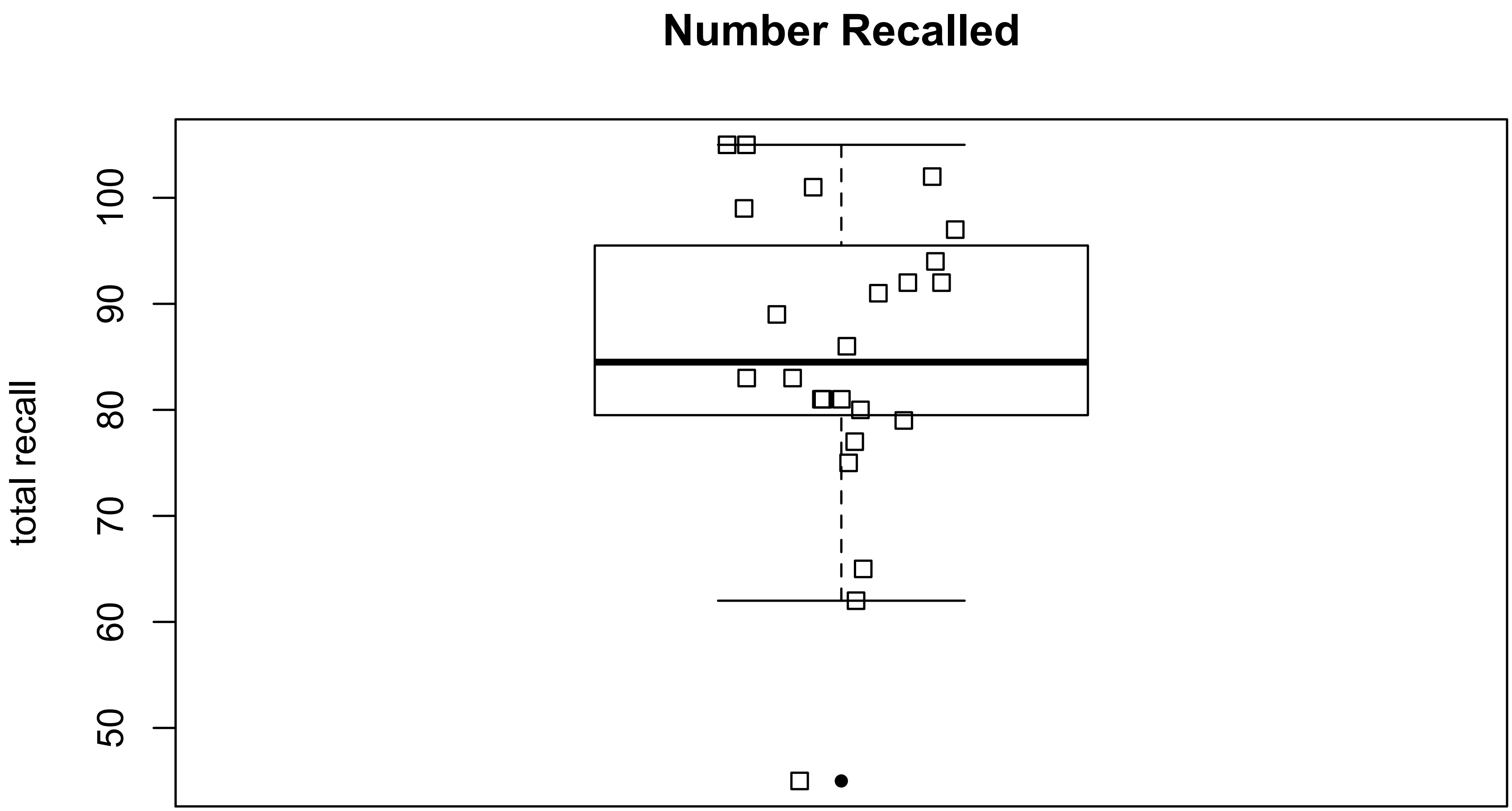
summary(total)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
63.00	79.50	91.50	87.75	96.50	112.00

describe(total)

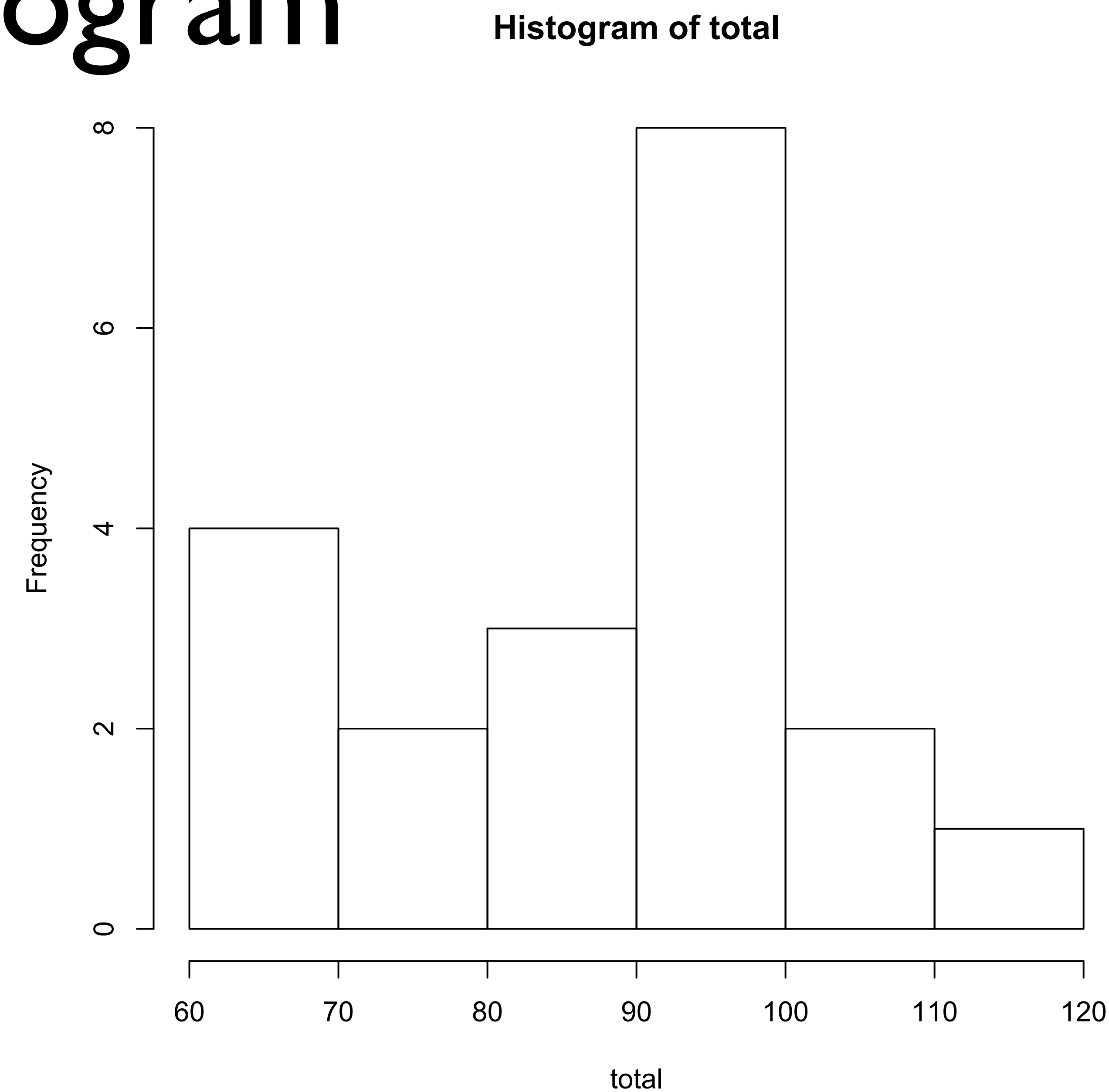
	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	1	20	87.75	13.38	91.5	88.19	13.34	63	112	49	-0.31	-0.95	2.99

Compare to last year



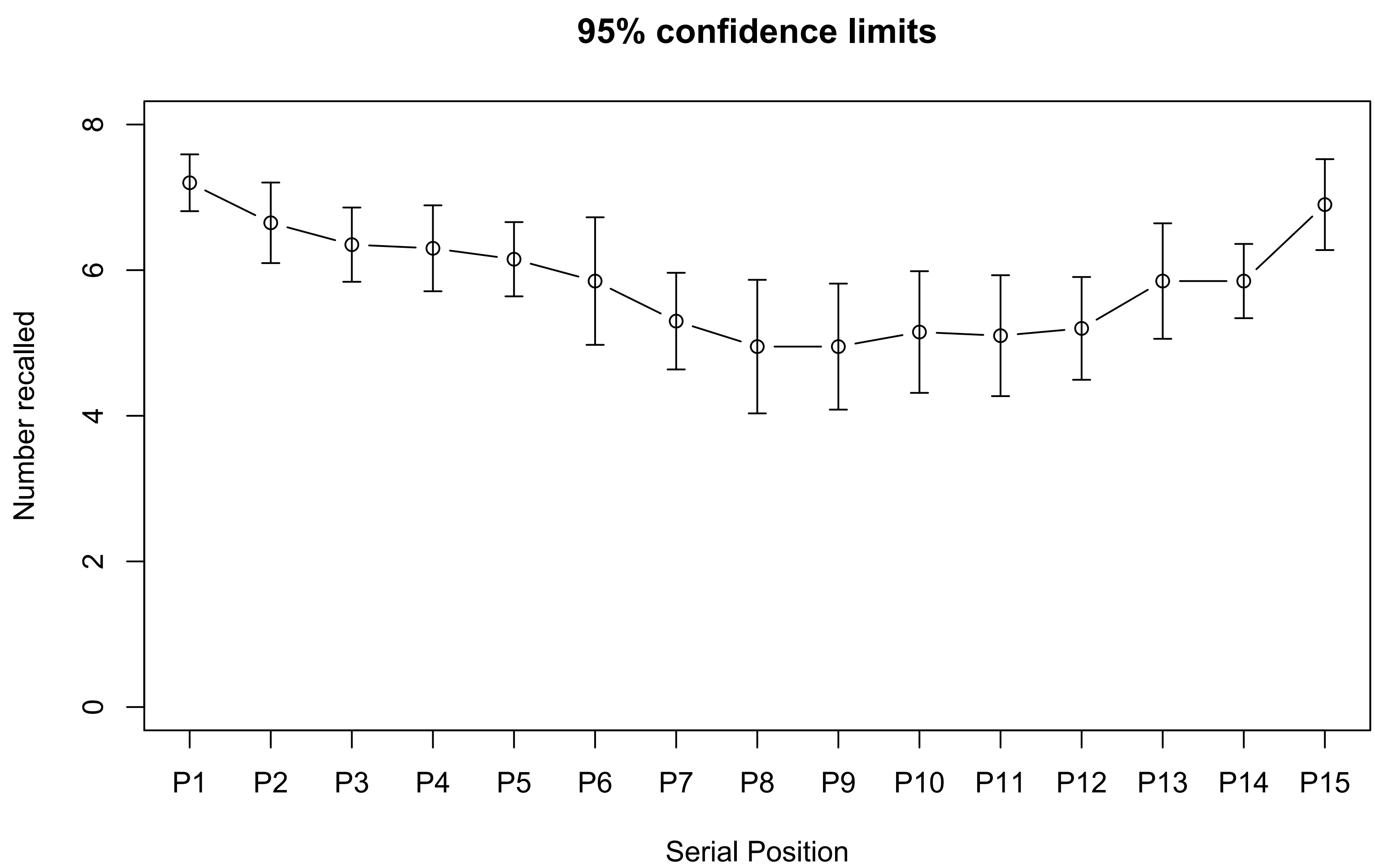
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
45.00	79.75	84.50	85.21	94.75	105.00

A histogram



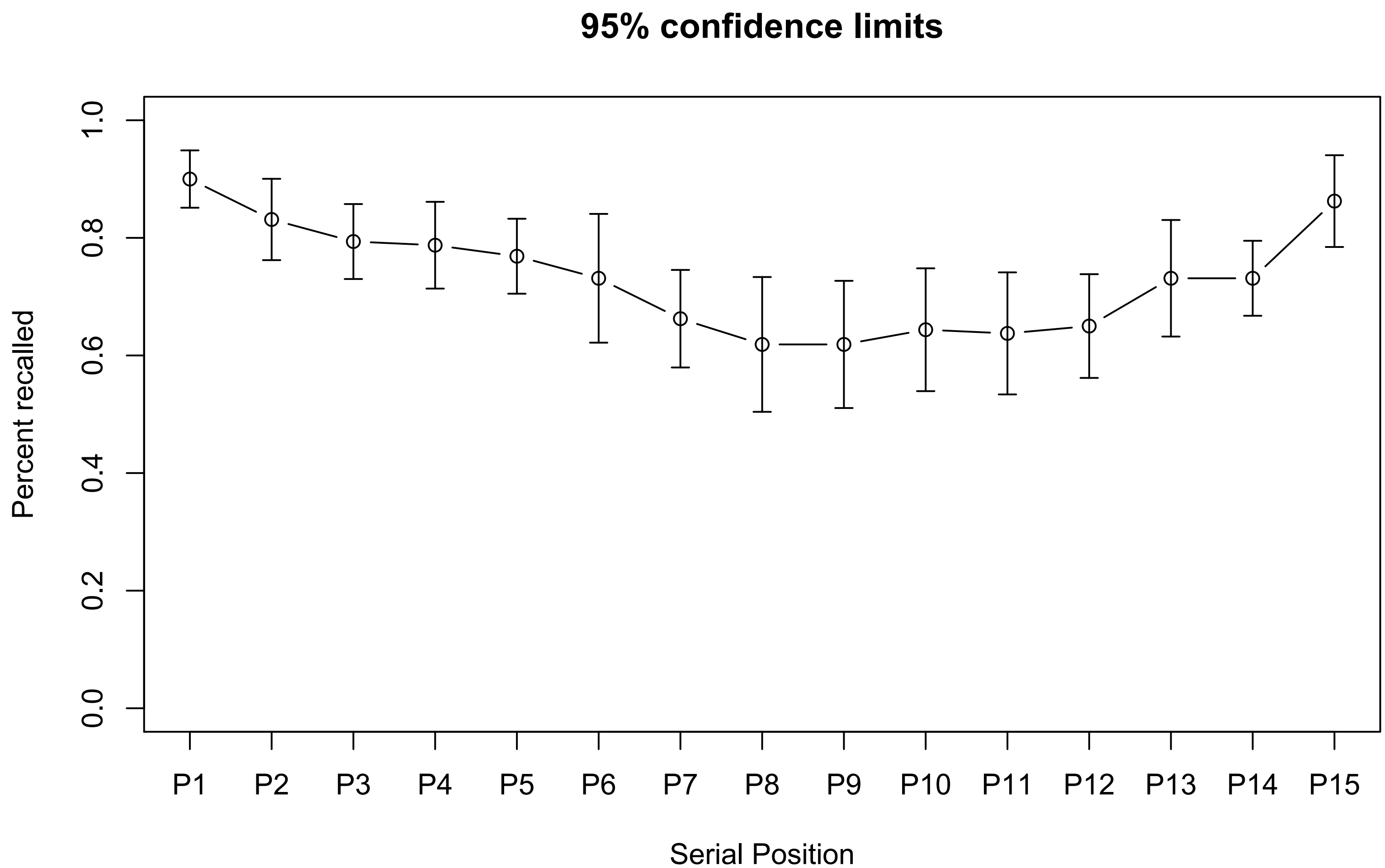
hist(total)

Total recalled by position




```
error.bars(rec,ylim=c(0,8),ylab="Number recalled",xlab="Serial
Position",typ="b")
```

Percent recalled by position




```
error.bars(rec/8,ylim=c(0,1),ylab="Percent recalled",xlab="Serial
Position",typ="b")
```

The recall data organized by list

```
> words
    W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15 W16
S1  NA 12 14 NA NA 11 13 NA 13  NA  NA  14  15  NA  NA  10
S2  NA 10 12 NA NA 11 13 NA 11  NA  NA  11  10  NA  NA  14
S3  NA  8 12 NA NA 13 13 NA 12  NA  NA  12  13  NA  NA   8
S4  10 NA NA 10 10 NA NA 10 NA   4   9  NA  NA   9   8  NA
S5  12 NA NA 11  8 NA NA 13 NA   9  12  NA  NA  11  12  NA
S6  NA 10 11 NA NA 10 10 NA 12  NA  NA   9  11  NA  NA   8
S7  NA 12 12 NA NA 15 15 NA 14  NA  NA  13  12  NA  NA  10
S8  13 NA NA 15  9 NA NA 14 NA  11  13  NA  NA  12  12  NA
S9   7 NA NA 10 11 NA NA  7 NA  11   5  NA  NA   6   6  NA
S10  9 NA NA  8 10 NA NA 11  2  14  14  NA  NA   5   5  NA
S11 11 NA NA 10 11 NA NA 13 NA  13  12  NA  NA  11  12  NA
S12 NA 11 11 NA NA 12 13 NA 13  NA  NA  12   9  NA  NA  11
S13 NA  9 13 NA NA 11 13 NA 15  NA  NA  11  11  NA  NA  12
S14 10 NA NA 10 15 NA NA 11 NA  11  14  NA  NA  14  13  NA
S15 NA 11 11 NA NA  5  6 NA  5  NA  NA   8  12  NA  NA   8
S16 NA  9  9 NA NA  5 10 NA 10  NA  NA  10   9  NA  NA   7
S17 NA  9  9 NA NA  9 12 NA 12  NA  NA  11  13  NA  NA  12
S18 12 NA NA 11  8 NA NA 12 NA   9   9  NA  NA  11   8  NA
S19 12 NA NA 12 12 NA NA 12 NA  12  12  NA  NA  13  11  NA
S20 NA 15 13 NA NA 15 14 NA 15  NA  NA  13  14  NA  NA  13
```

Score Visual and Oral and find descriptive statistics

```
VO <- c(1,2,7,8,11:14)  #specify the columns to score
Visual <- rowSums(words[,VO],na.rm=TRUE)#find the sum
Oral <- rowSums(words[,-VO],na.rm=TRUE) #score the others
VisOral <- data.frame(Visual=Visual,Oral=Oral) #organize them
describe(VisOral) #descriptive stats
```

```
VisOral
    Visual Oral
S1      54   48
S2      44   48
S3      46   45
S4      38   32
S5      48   40
S6      40   41
S7      52   51
S8      52   47
S9      25   38
S10     39   39
S11     47   46
S12     45   47
S13     44   51
S14     49   49
S15     37   29
S16     38   31
S17     45   42
S18     44   36
S19     49   47
S20     56   56
```

```
       var  n  mean   sd median trimmed  mad min max range  skew kurtosis   se
Visual   1 20 44.60 7.14   45.0    45.0 6.67  25  56    31 -0.75     0.61 1.60
Oral     2 20 43.15 7.29   45.5    43.5 7.41  29  56    27 -0.38    -0.89 1.63
```

Modality effects on recall

Two conditions: visual and oral, do they differ?

```
> with(VisOral,t.test(Visual,Oral,paired=TRUE))

	Paired t-test

data:  Visual and Oral
t = 1.2042, df = 19, p-value = 0.2433
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.07022  3.97022
sample estimates:
mean of the differences
                   1.45
```

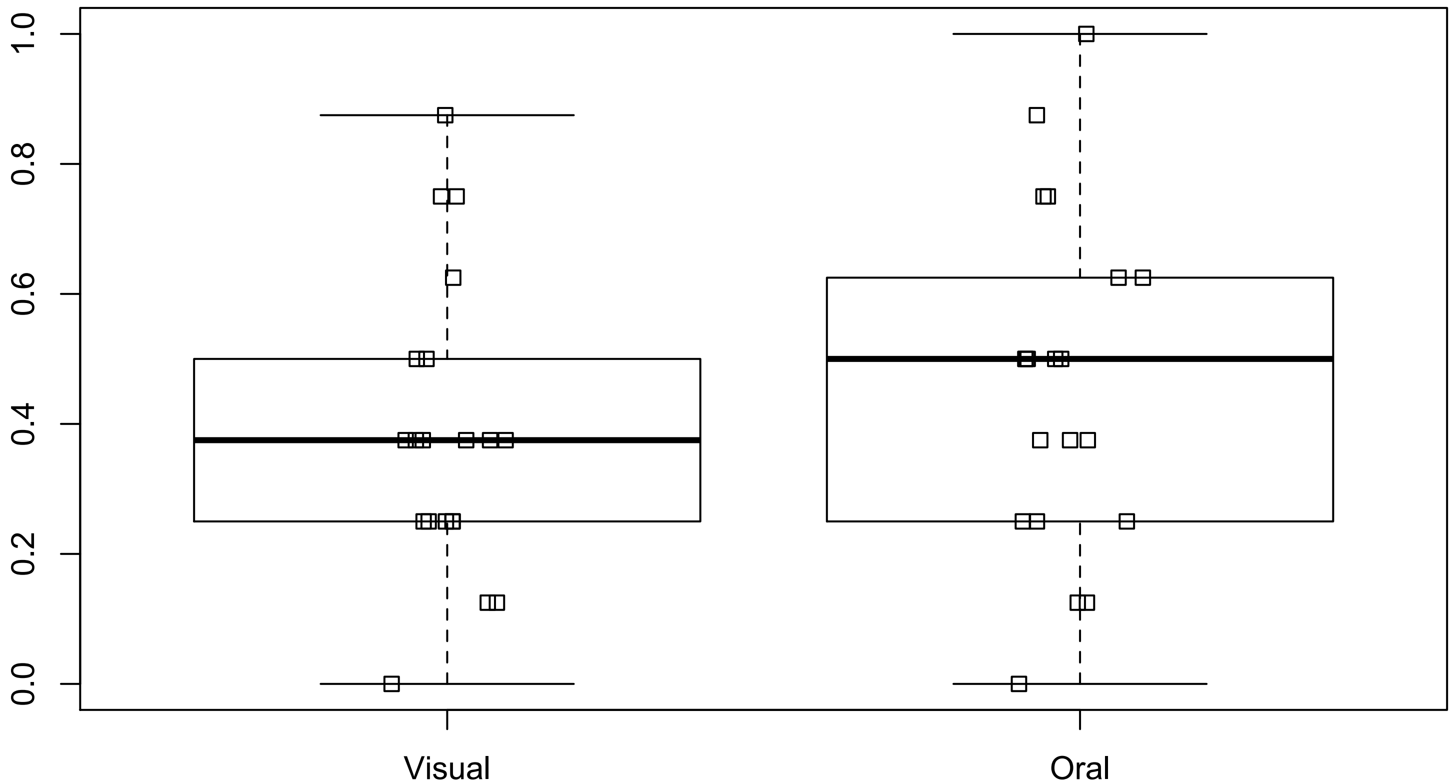
Probability is .24 to get an effect this big by chance

Is there a false recognition effect?

- Are there any false recognitions?
- If so, do they differ as function of our conditions?
- describe(recog.df)

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Visual	1	20	3.1	1.80	3	3.00	1.48	0	7	7	0.52	-0.52	0.40
Oral	2	20	3.7	2.08	4	3.62	2.22	0	8	8	0.21	-0.75	0.47

40-48% false recognition!



Descriptive and Inferential Statistics

- Describe the data
  - Central Tendencies and Dispersion
  - Means, standard deviations
- Inferential -- the Null Hypothesis model
- How likely are the data given a model of no difference
  - consider the t-test

Multiple ways to model variance

- t test compares the difference of two groups
- F-test (ANOVA) is a generalization of t to compare multiple groups
- If the independent variable is categorical, then it can be thought of in terms of groups and we can use ANOVA
- If the independent variable is continuous, then we use the linear model.
- ANOVA is a special case of linear model

Recall and Recognition Hypothesis testing

- How likely would differences of this magnitude be observed if in fact there were no effect in the population.
- Null Hypothesis Test
  - H_0 The groups do not differ in the population
  - H_1 The groups come from different populations
  - How likely are the results if H_0?
  - What is the probability of data given H_0?
  - Reject H_0 if $p <$ critical value

Significance testing using Analysis of Variance

- ANOVA as a generalization of t-test.
  - t-test compares the difference between two means in terms of the expected standard deviation of the mean = observed standard deviation/sqrt(N-1)
- ANOVA compares the variance of the sample means to the variance within groups
- Possible to do ANOVA for multiple comparisons (combinations of variables)

Interpretation of ANOVA

- Each anova is a comparison of two estimates of the population variance:
  - an estimate from the variance between groups and an estimate from the variance within groups.
- F is the ratio of these estimates. If the two groups are random samples from the same population, we would expect the F ratio to be 1. The more the F deviates from 1, the less likely is the hypothesis that the samples came from the same population.

Alternative to hypothesis testing

- Effect size and confidence interval.
- How big is the effect and what is the expected range of the effect?

Central Tendencies and error

- Sample means reflect population values +/- error variability
- standard deviation of a mean (the standard error) $= s.d/\sqrt{N}$
- observed mean +/- 1 standard error includes the population value 68% of the time
- means that differ by 2.8 standard errors are unlikely to be from same population
- errors of within subject designs are more complicated to show

Results

- Recall (manipulation check)
  - Is there a serial position effect?
    - Primacy
    - Recency (particularly given the instructions)
  - Did people just recall on recall tasks?
  - Do the lists differ in recall ease?
- Recognition
  - Is there a false memory effect?
  - What manipulations affect it?

Conclusions

- Big picture
  - Possible to show false memory, particularly in a recognition task
- Smaller picture
  - variables that affect false recognition
- Take home message:
  - What does this all mean