# Designing an Experiment (part 1)

# 13 Probability and Statistics

COMS10011

Dr. Anne Roudaut
csxar@bristol.ac.uk

# let’s do an experiment!

# memorization game

group 1

memorize as much
as you can

group 2

if you beat group 1 =
chocolate!

take a piece of paper and a pen

I will tell a list of numbers

 “1,2,3,6,write”

**only when “write”** -> write the list on paper

I will show the list

1, 2, 3, 6

if you are **correct** continue the game

if you wrong stop the game, remember *best score*

practice trials

1, 4, 9 (size=3)

practice trials

8, 7, 3, 5, 6, 1 ,2 (size=7)

let’s start the real experiment!

trial

3, 2, 8 (size=3)

trial

4, 2, 5, 1 (size=4)

trial

7, 2, 5, 3, 1 (size=5)

trial

6, 2, 9, 8, 5, 1 (size=6)

trial

7, 4, 1, 8, 6, 3, 2 (size=7)

trial

2, 7, 4, 9, 3, 1, 5, 9 (size=8)

trial

1, 6, 7, 8, 5, 3, 1, 4, 6 (size=9)

trial

6, 4, 1, 9, 3, 8, 2, 1, 7, 9 (size=10)

trial

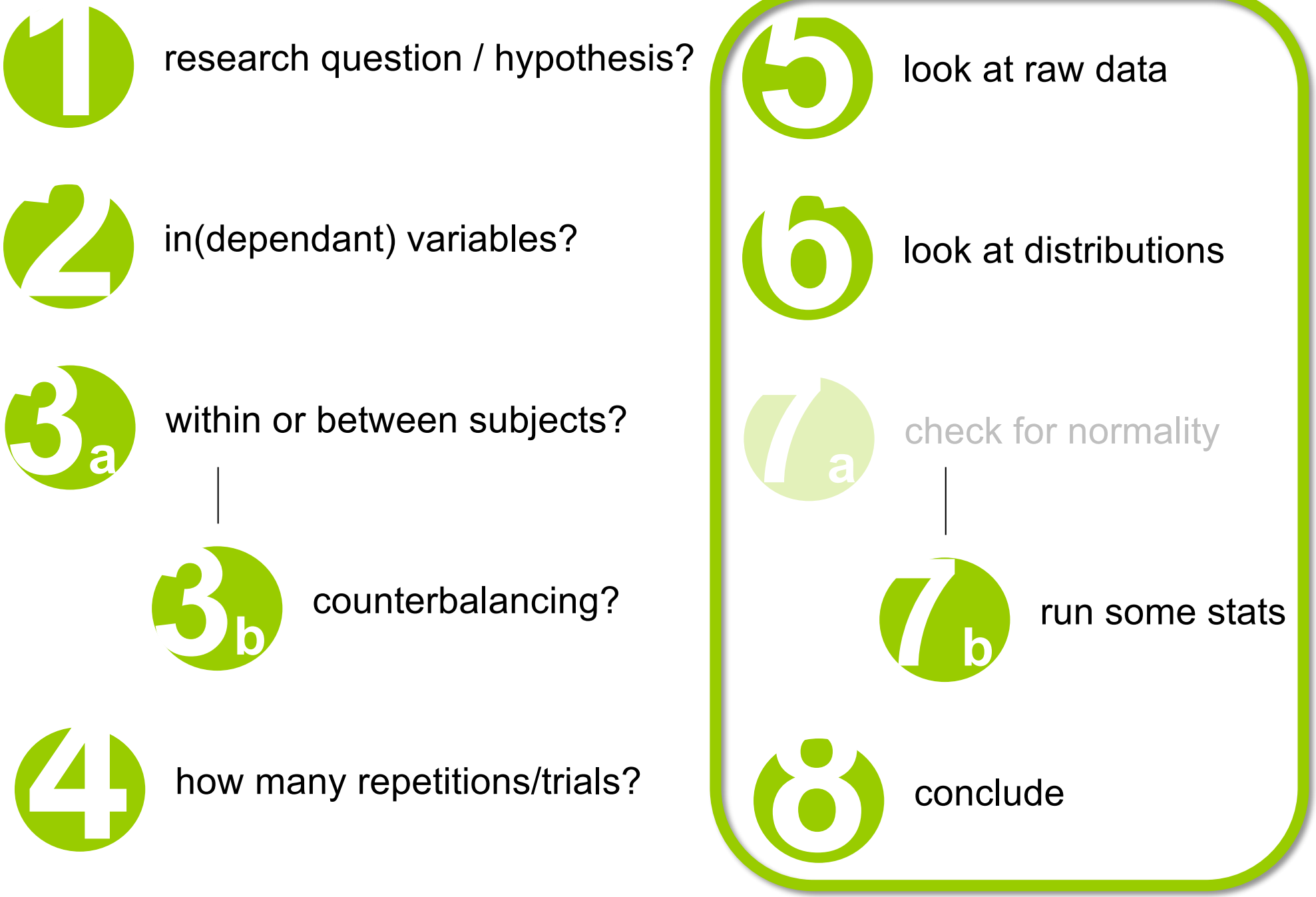
2, 7, 4, 1, 5, 7, 3, 8, 6, 4, 7 (size=11)

what is your best score (size of the list)?

enter it at

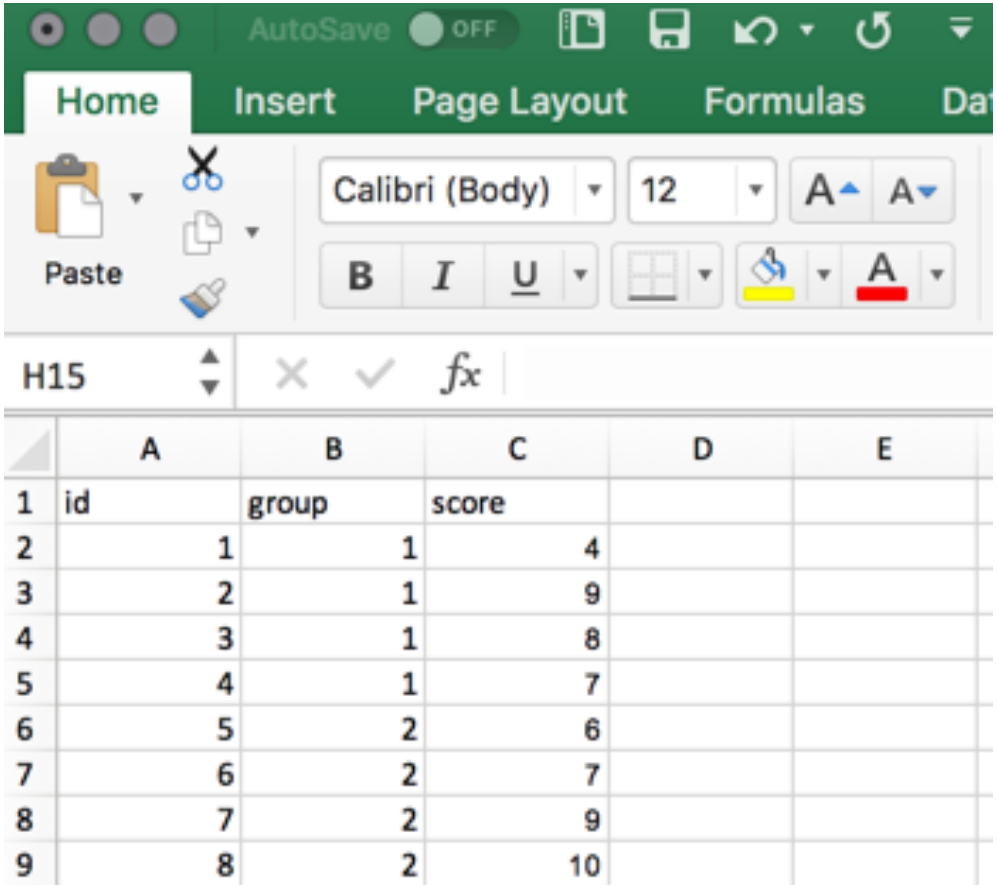
**https://tinyurl.com/COMS10011**

# let’s first look at the results


1
research question / hypothesis?
2
in(dependant) variables?
3a
within or between subjects?
3b
counterbalancing?
4
how many repetitions/trials?
5
look at raw data
6
look at distributions
7a
check for normality
7b
run some stats
8
conclude

# look at raw data

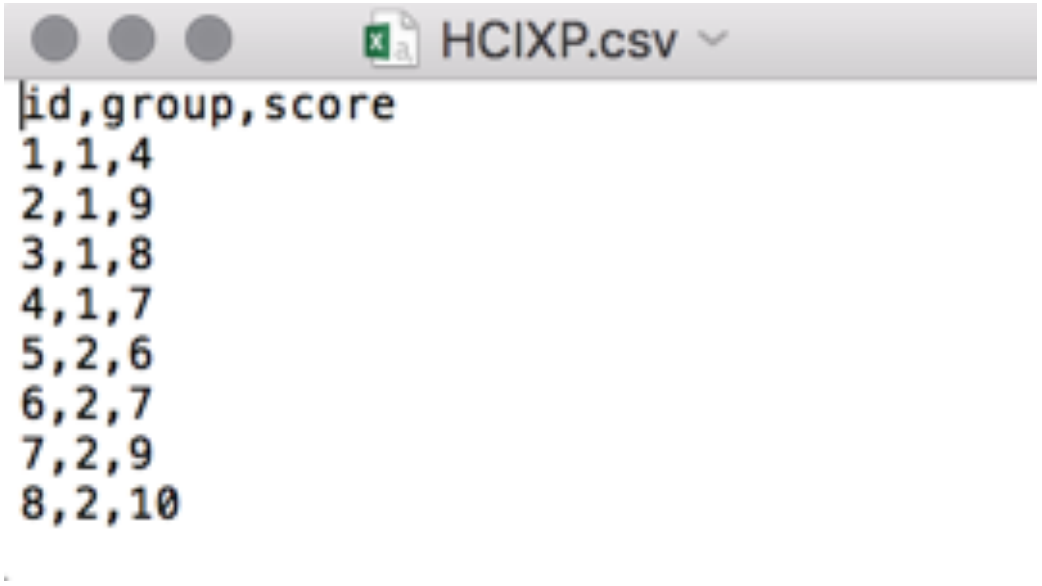
let’s put everything in a table (excel is great for that)




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | id | group | score | | |
| 2 | 1 | 1 | 4 | | |
| 3 | 2 | 1 | 9 | | |
| 4 | 3 | 1 | 8 | | |
| 5 | 4 | 1 | 7 | | |
| 6 | 5 | 2 | 6 | | |
| 7 | 6 | 2 | 7 | | |
| 8 | 7 | 2 | 9 | | |
| 9 | 8 | 2 | 10 | | |

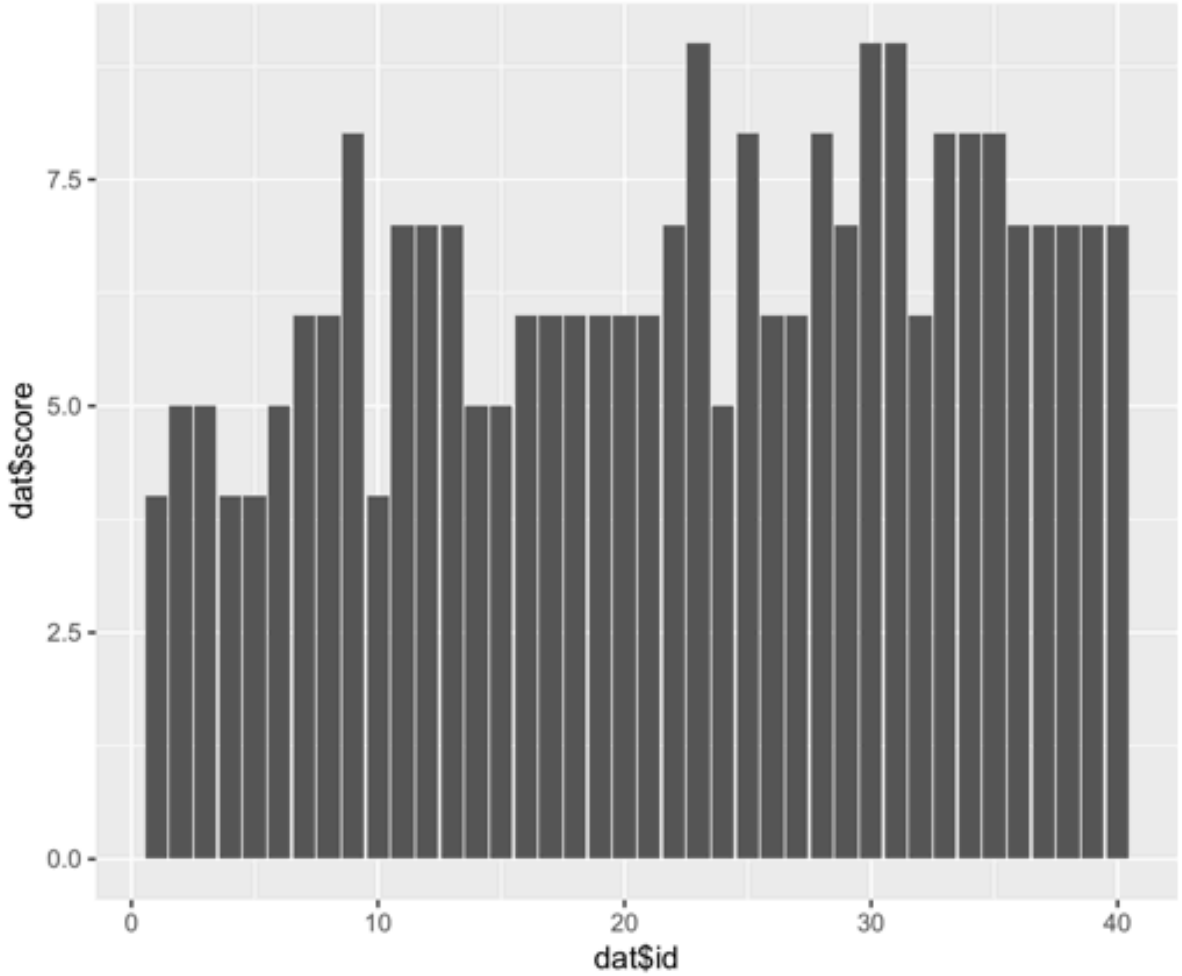
save your file as a .csv (comma separated virgule is a format to store tables as text files)

you can open csv with excel, text file an many other software



```
dat = read.csv("HCIXP.csv", header = TRUE)
print(dat) # look at the file in R
```

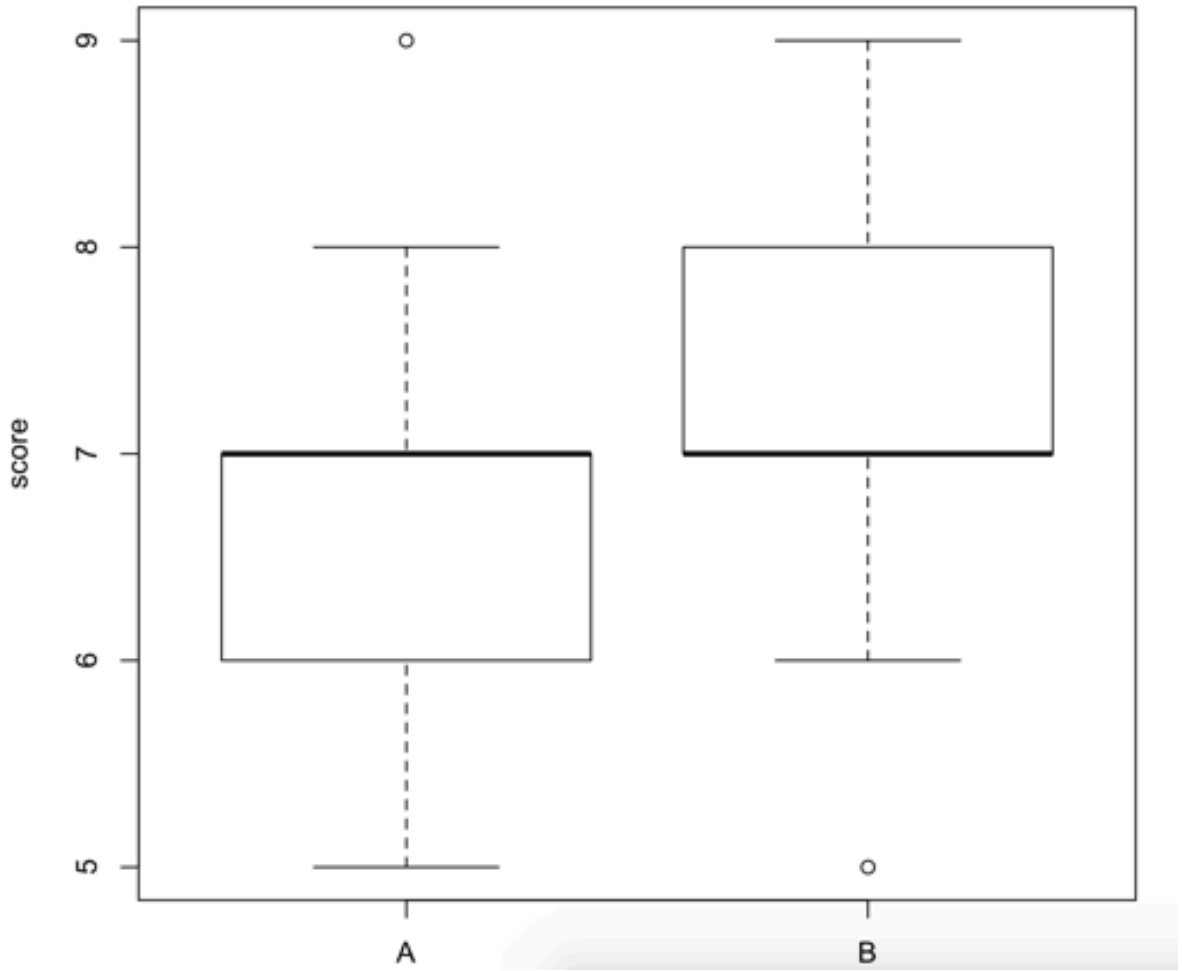
```
dat = read.csv("HCIXP.csv", header = TRUE)
print(dat) # look at the file in R

library(ggplot2)

ggplot(dat, aes(x = dat$id, y = dat$score)) +
geom_bar(stat = 'identity', position = 'dodge')
```




first: does the data look ok?

search for bugs, fatigue effect, learning effect
or outliers (>3 times std) = remove / redo xp

```
plot(score ~ group, data = dat)
```

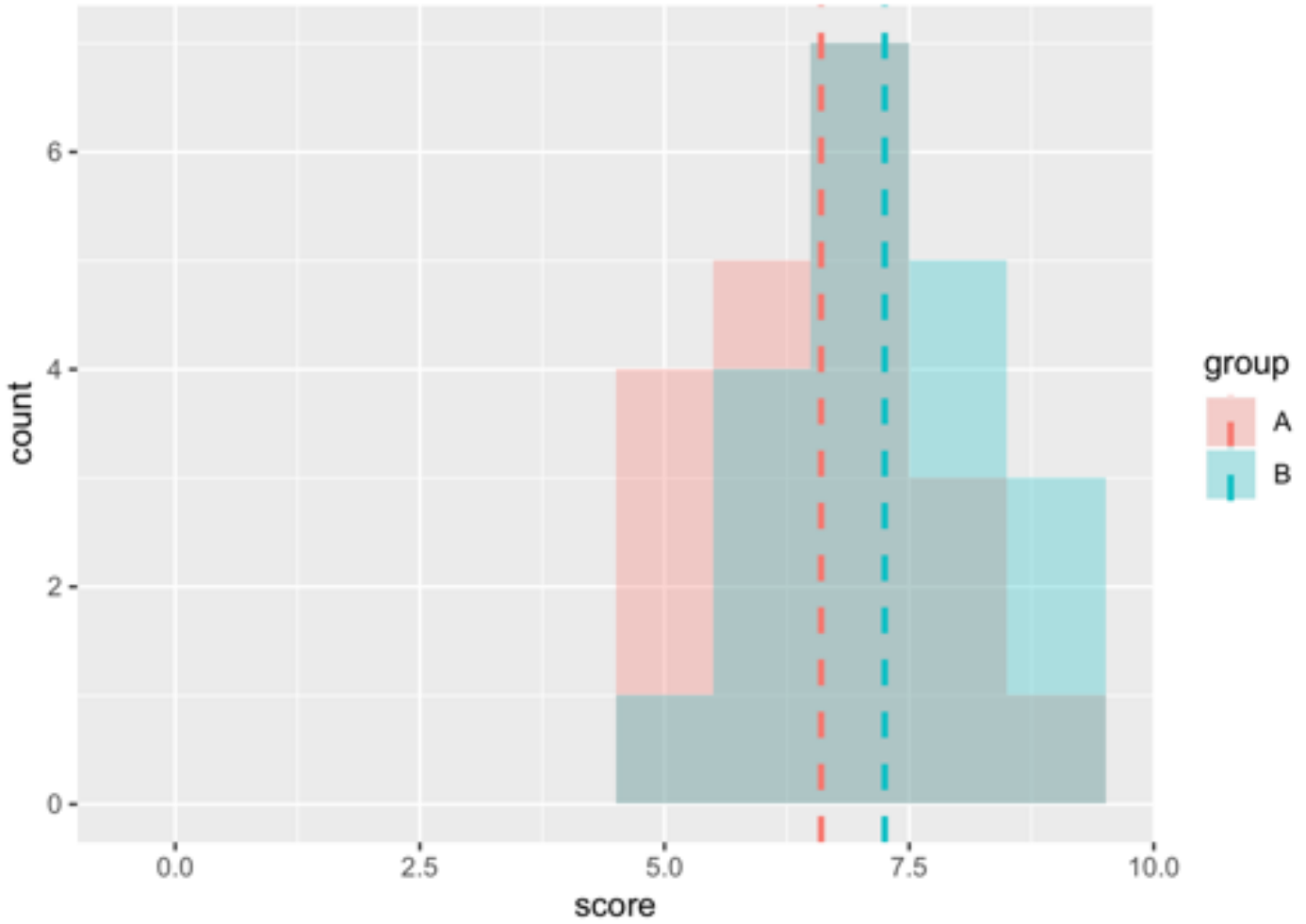


# look at histograms

```
# Find the mean of each group
library(plyr)
cdat <- ddply(dat, "group", summarise,
score.mean=mean(score))
cdat
```

```
  group score.mean
1     A       5.60
2     B       7.25
```

```
# Overlaid histograms with means
ggplot(dat, aes(x=score, fill=group)) +
geom_histogram(binwidth=1, alpha=.3, position="identity")
+ geom_vline(data=cdat, aes(xintercept=score.mean,
colour=group), linetype="dashed", size=1) +
expand_limits(x = 0, y = 0)
```




your gut feeling: are these groups different?

are these distributions likely to have happen by chance?

… is this the results of the factor (chocolate)?

# use a statistic test

```
# Use a t-test (two-tails, unpaired)
t.test(dat$score[dat$group == "A"], dat$score[dat$group
=="B"], alternative = "two.sided")

        Welch Two Sample t-test

        data:  dat$score[dat$group == "A"] and
        dat$score[dat$group == "B"]
        t = -1.8185, df = 37.982, p-value = 0.07688
        alternative hypothesis: true difference in means is
        not equal to 0
        95 percent confidence interval:-
        1.37361001  0.07361001
        sample estimates: mean of x mean of
        y  6.60      7.25
```

**“We could not find any significance differences!”**

**p-value = 0.07**

is is enough to say that the two groups are different?

-> nope, not under significant level of 0.05

can we say that the two groups are same then?

-> nope, can only prove things are different, but not that they are the same

# conclude

if p was lower than significance level we could say:

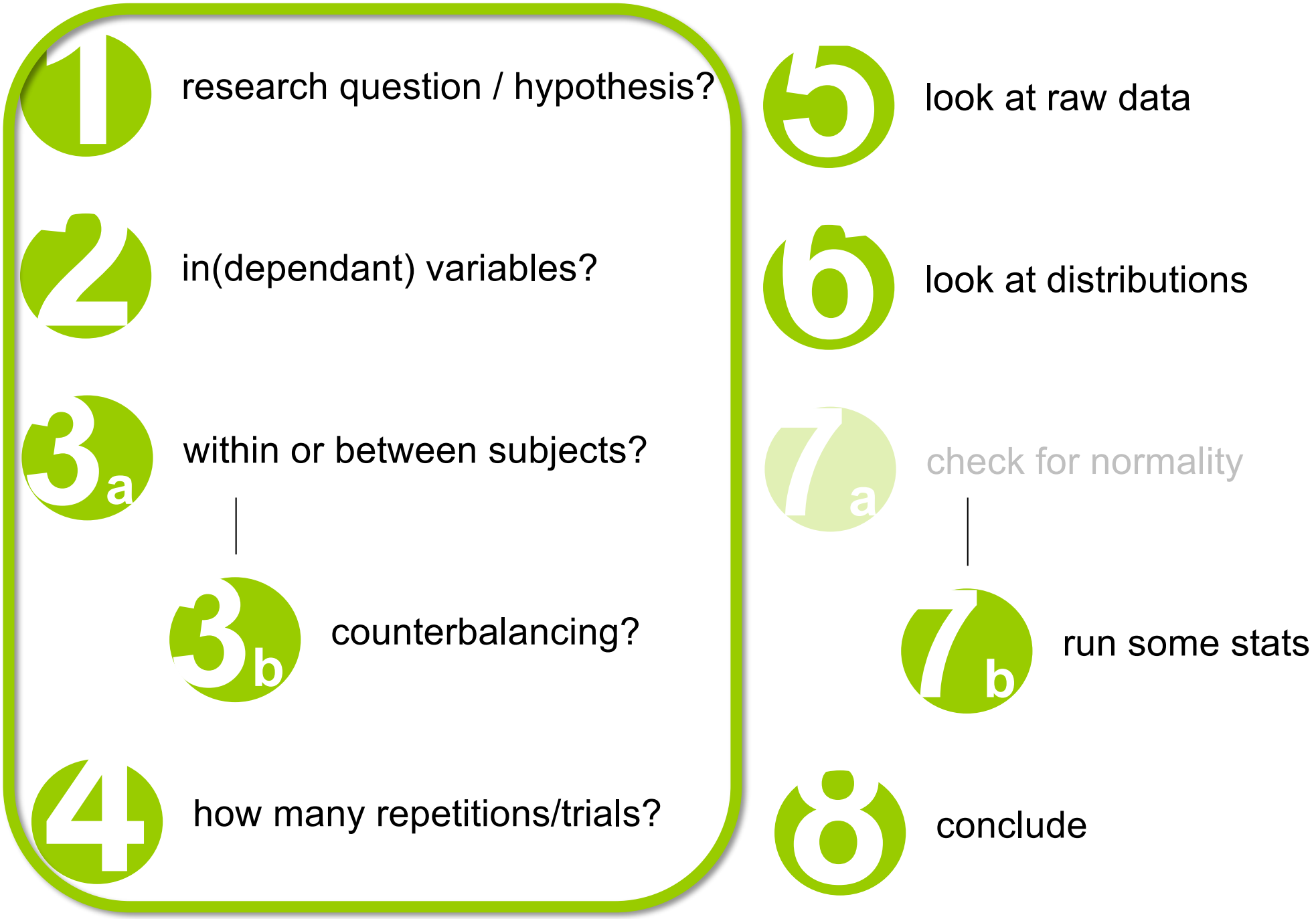
“a student t-test showed significant difference between the two group (two-tailed $t(46)=4.520$, $p < 0.005$)”

otherwise:

“we did not find any significant results”

cannot conclude, no evidences to show that having chocolate rewards improve memorisation

let’s go
backward a little


1
research question / hypothesis?
2
in(dependant) variables?
3a
within or between subjects?
3b
counterbalancing?
4
how many repetitions/trials?
5
look at raw data
6
look at distributions
7a
check for normality
7b
run some stats
8
conclude

# research question::

a statement that identifies a phenomenon to be studied

in our xp: I believe that **rewards improve memorization skills**

… suggested by <*insert smart guess*>

# hypotheses::

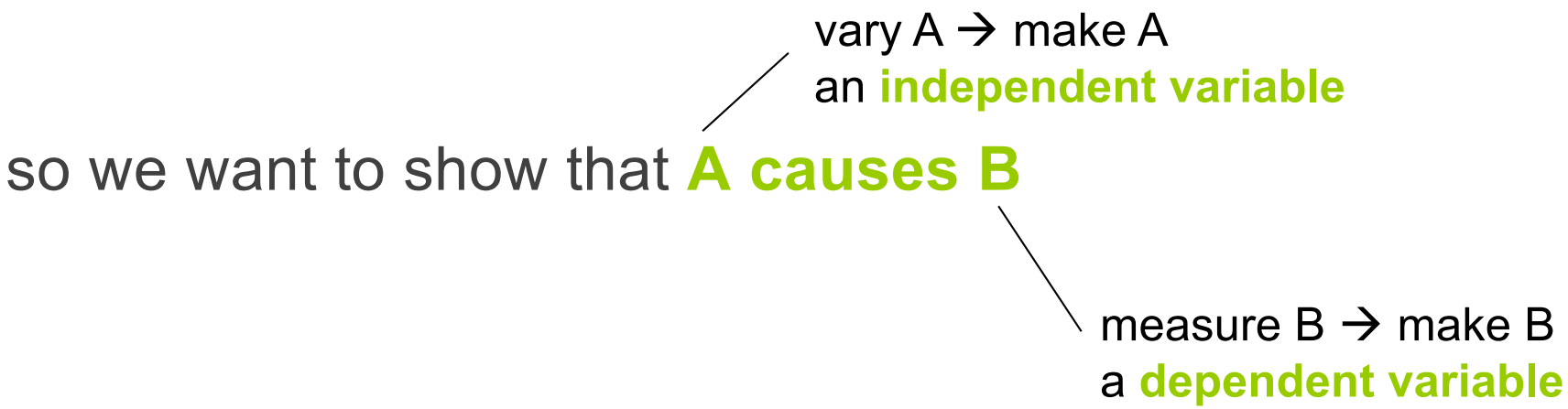
statement of the predicted relationship between at least two experimental variables

**provisional answer to a research question**

in our xp: **group chocolate will have a higher memorisation score than group with no reward**

# (in)dependent variable ::

the **dependent variable** is the event studied and expected to change whenever the **independent variable** is altered


vary A → make A
an independent variable
so we want to show that A causes B
measure B → make B
a dependent variable

in our xp?

**independent variable** = group type (nothing vs. chocolate)

**dependent variable** = memorization score

everything else should be a…

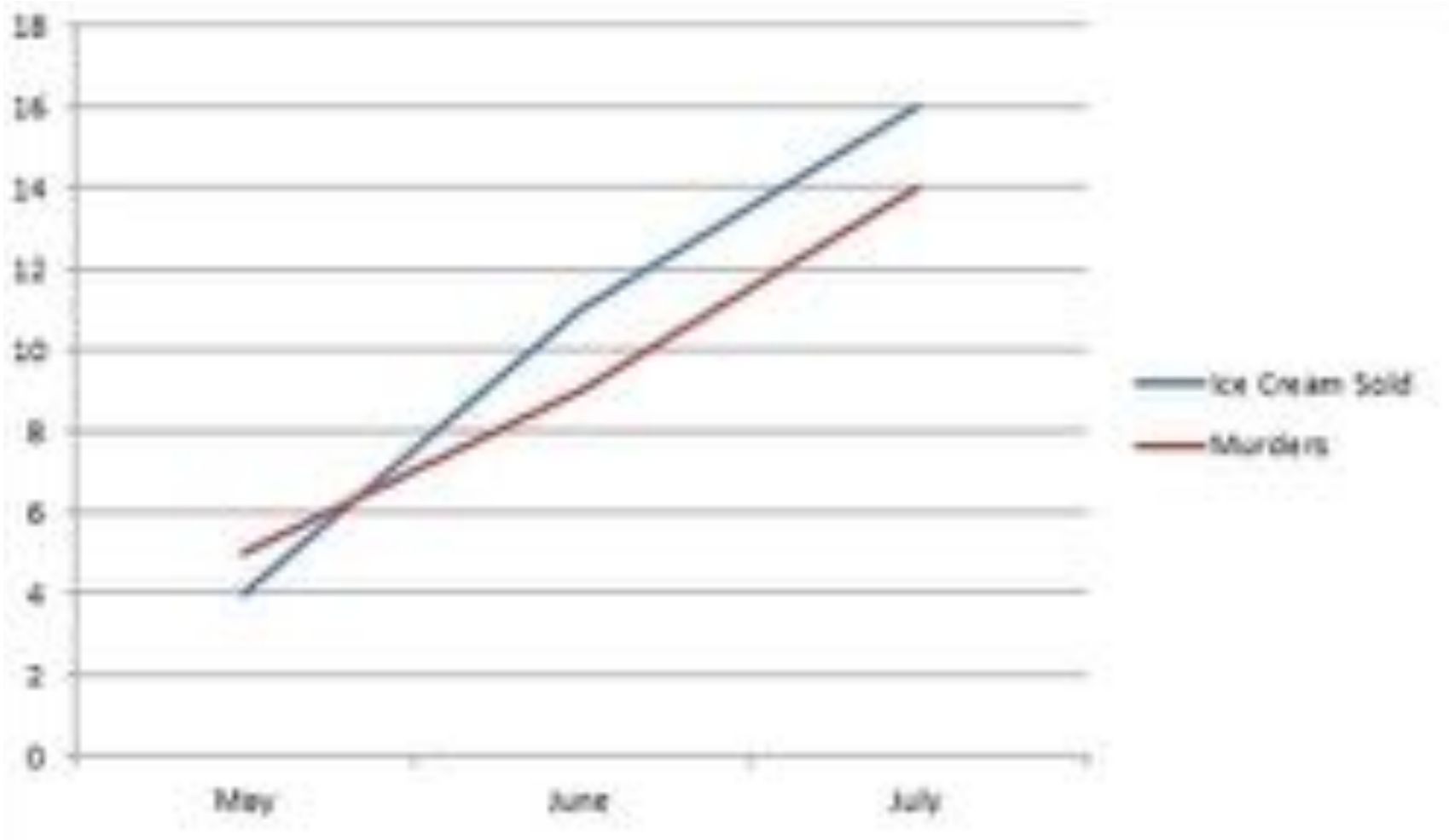
# controlled variable ::

the variables that are kept constant to prevent their influence on the effect of the independent variable on the dependent

avoid…

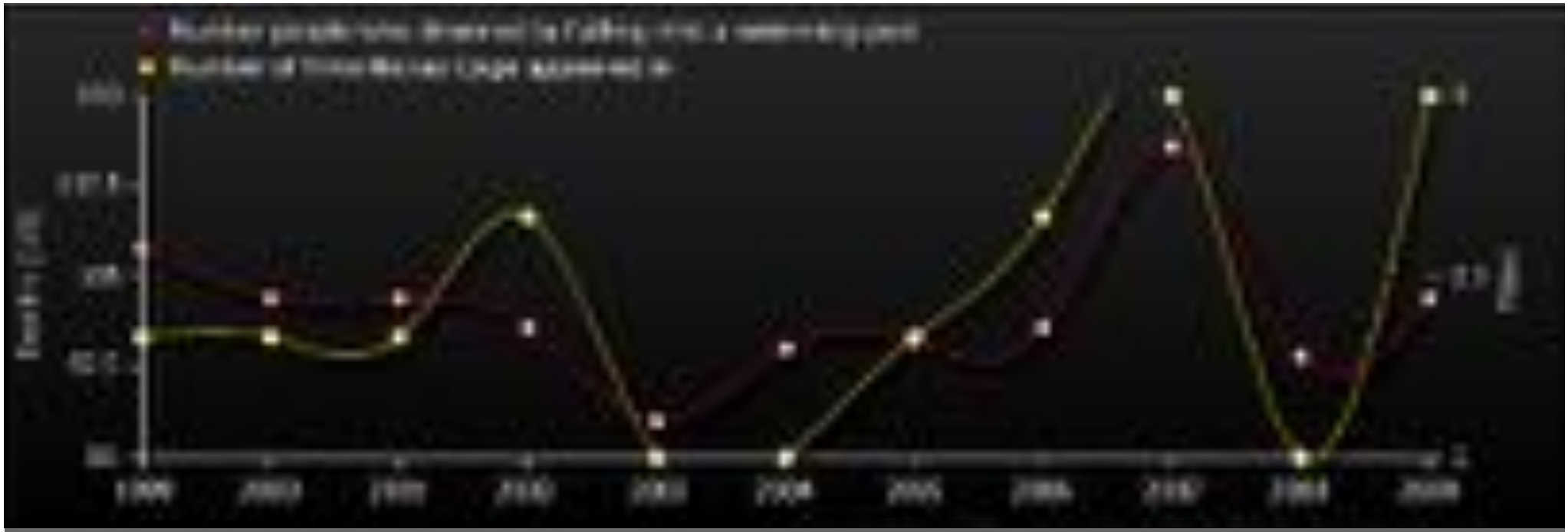
# confounding variable ::

extraneous variables that **correlates with both** the dependent variable and the independent variable




ice cream consumption leads to murder

**counfounding** : weather temperature



number of people drowned by falling into a swimming-pool correlates with number of films Nicolas Cage appeared in

this is not about **correlation**

this is about how to show **causality,**
i.e., that **some A causes some B**

in our xp, do we have confounding variables?

**yes, it is not greatly designed :s**

gender, age, background, what you ate before, if you like chocolate or not, if you are competitive and want the others not to have chocolate, if some of the numbers are familiar to you etc.

what can we do about it?

- avoid them by controlling as much as you can in the environment
- if you cannot, make it an independent variable (e.g. gender)
- some are inherent *noise* (human individuality), use more participants to get *statistical power*

the goal of a quantitative study is to find
**a signal** in **a lot of noise**

# experimental design:

aims at maximizing your chances of **finding the signal** and not the noise

1. need to absolutely **avoid systematic biases**

(e.g., learning effect, fatigue). They give you **false results!**

2. **avoid random noise.** It makes your results non-significant. Clever experimental design is all about keeping the noise down

e.g. in our xp, I made you **practice before**!


# within vs. between?

within = all participants do same

between = participants do only certain conditions


suffer less user variation

statistical power with less participants

no biases from other conditions (e.g. transfer of learning)

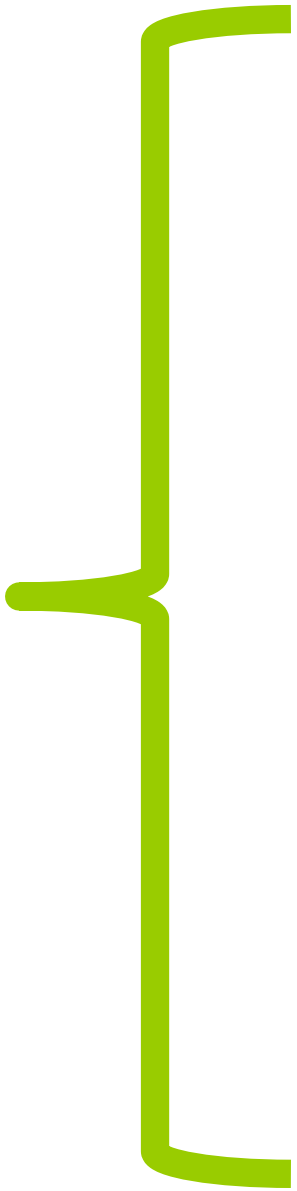
# within vs. between?

within = all participants do same

between = participants do only certain conditions

in our xp, it had to be **between subjects**
(because of the rewards)

participants did not do all conditions:
½ did the control condition
½ the reward condition



imagine a **within subjects** (test how fast we click an icon):

participants do all conditions:

they start with the trackpad

when finished they do the mouse

is it a good idea?

**nope -> learning effect**


# counterbalancing ::

a method of avoiding confounding among variables

**presenting conditions in a different order**

one approach to counterbalancing is to use a…

| | | |
|---|---|---|
| A | B | C |
| C | A | B |
| B | C | A |

# Latin square ::

an $n \times n$ array filled with $n$ different Latin letters, each occurring exactly once in each row and exactly once in each column.

# how many trials?

ideally make as much trials as you can to reduce noise but try to keep experiment around 30 min … max 40 min

in our xp, we did only one trial because of time constraint, but should have done more to **reduce noises**

# summary

research question / hypothesis?

in(dependant) variables?

within or between subjects?

counterbalancing?

how many repetitions/trials?

look at raw data

look at distributions

we will see why

check for normality

run some stats

so far we know t-test

conclude

1. Explain the eight steps to design and analyze an experiment
2. Explain what is a within or between subject experiment
3. Explain what is a controlled variable or a confounding variable
4. Explain the difference between correlation and causality

# take away

end