Designing an Experiment (part 2)

14

Probability and Statistics

COMS10011

Dr. Anne Roudaut
csxar@bristol.ac.uk

research question / hypothesis?

in(dependant) variables?

within or between subjects?

counterbalancing?

how many repetitions/trials?

look at raw data

look at distributions

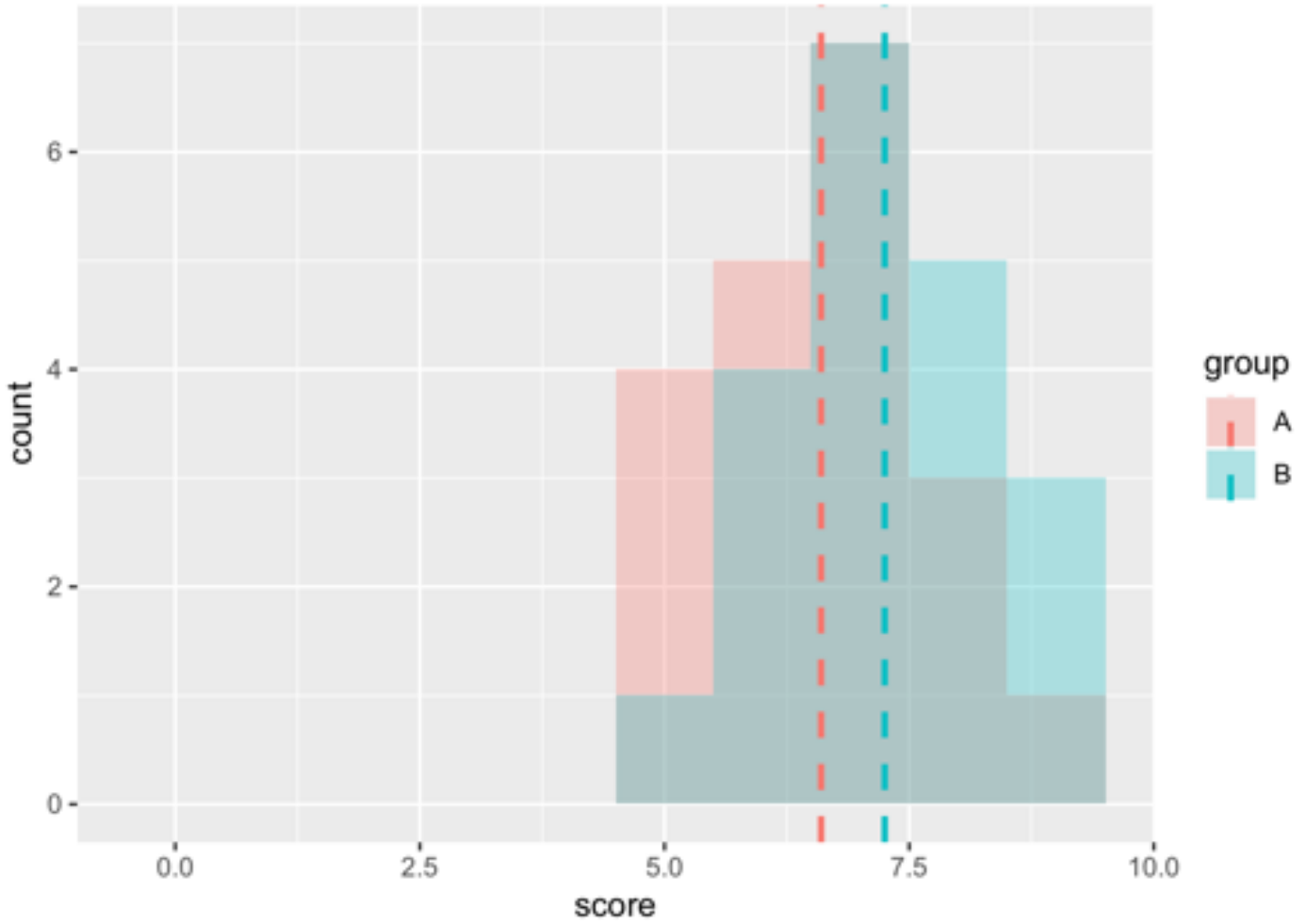
we will see why

check for normality

run some stats

so far we know t-test

conclude

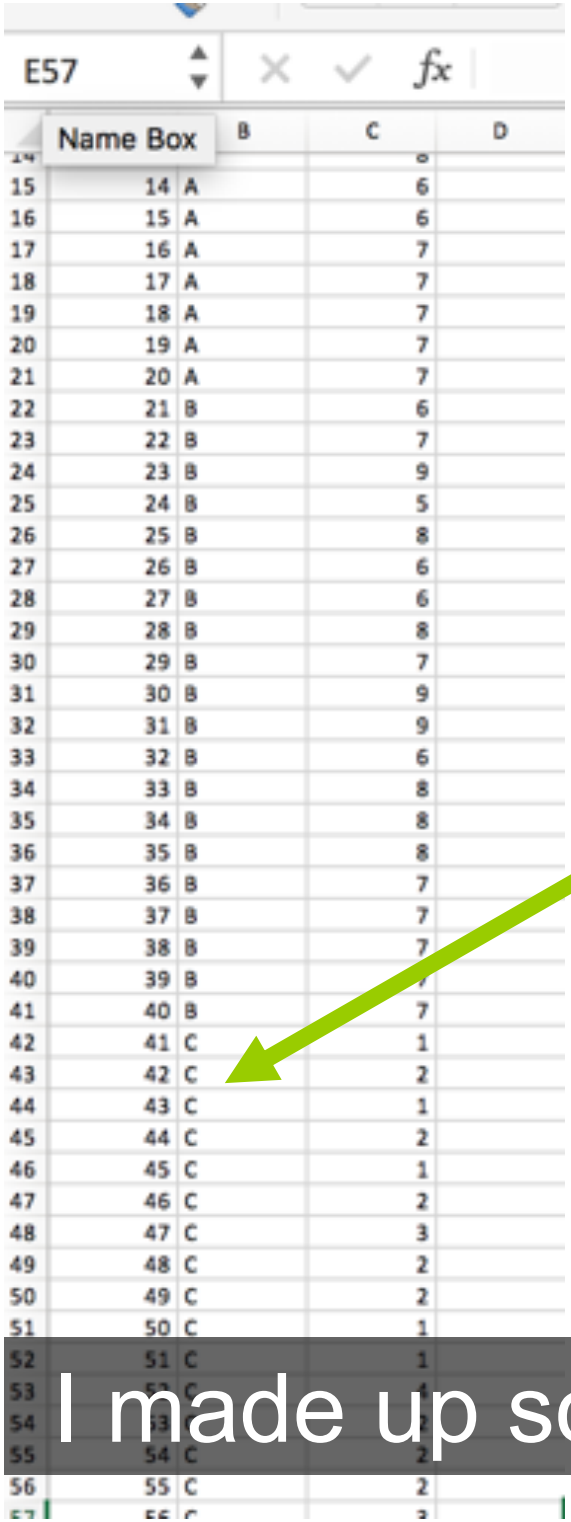


P>0.05. We did not find evidences of statistically differences between the two groups chocolate vs. no chocolate

let’s
complexify a little

in our xp, let’s add a 3rd imaginary group

they get a slap if they had the smallest memorisation score
(obviously not ethical so let’s keep this hypothetical!)



15	14	A	6
16	15	A	6
17	16	A	7
18	17	A	7
19	18	A	7
20	19	A	7
21	20	A	7
22	21	B	6
23	22	B	7
24	23	B	9
25	24	B	5
26	25	B	8
27	26	B	6
28	27	B	6
29	28	B	8
30	29	B	7
31	30	B	9
32	31	B	9
33	32	B	6
34	33	B	8
35	34	B	8
36	35	B	8
37	36	B	7
38	37	B	7
39	38	B	7
40	39	B	
41	40	B	7
42	41	C	1
43	42	C	2
44	43	C	1
45	44	C	2
46	45	C	1
47	46	C	2
48	47	C	3
49	48	C	2
50	49	C	2
51	50	C	1
52	51	C	1
53		C	
54			
55	54	C	2
56	55	C	2

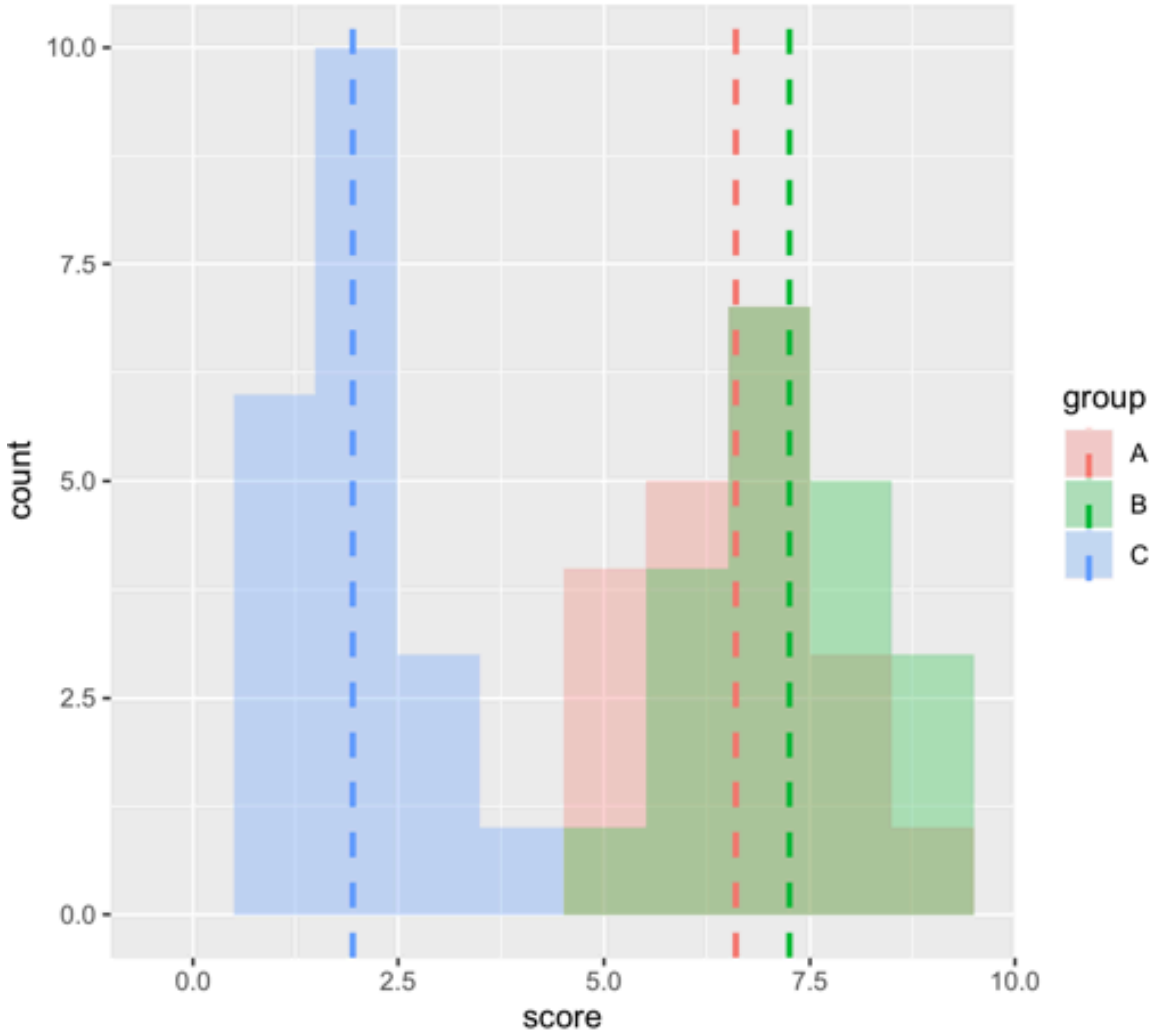
Group C: “slap”

I made up some data

```
# Find the mean of each group
dat = read.csv("HCIXP-anova.csv", header = TRUE)
cdat <- ddply(dat, "group", summarise,
score.mean=mean(score))
cdat
```

```
  group score.mean
1     A       6.60
2     B       7.25
3     C       1.95
```

```
# 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?

can we use t-tests?

(3 tests to compare group 1 with 2, 2 with 3 and 1 with 3)

-> yes but use Bonferoni correction

significance level not 0.05 anymore but 0.05 / number of comparisons performed (here 3) so 0.016

```
# Use a t-test (two-tails, unpaired)

# (we already know A vs B not significative) so we need to
do

t.test(dat$score[dat$group == "A"], dat$score[dat$group ==
"C"], alternative = "two.sided")
```

t = 14.753, df = 34.591, p-value < 2.2e-16

```
# and

t.test(dat$score[dat$group == "B"], dat$score[dat$group ==
"C"], alternative = "two.sided")
```

t = 17.054, df = 34.971, p-value < 2.2e-16

In both case p_value < 0.016 so we can conclude!

Another test we can use when we have more than two groups to compare is an ANOVA

we have 3 different conditions (or 1 factor with 3 different levels) so we will do a one-way ANOVA

anova::

analyze of variance to compare multiple variables

one-way anova = one variable with multiple levels

two-way anova = two variables with multiple levels

```
# first we run the one-way anova
library(ez)
ezANOVA(dat,id,between=group,dv=score)
```

```
  Effect DFn DFd      F            p p<.05       ges
1  group   2  57 154.8886 9.056612e-24     * 0.8445923
```

ok something is going to be interesting here

```
# second, run the pairwise comparison

pairwise.t.test(dat$score,dat$group, paired=FALSE,
p.adjust.method="bonferroni")
```

```
  A       B
B 0.16    -
C <2e-16  <2e-16
```

here are significant differences

and we don't need to do the Bonferroni correction (already included)

we can write:

“A one-way ANOVA showed a significant effect on time for the variable Group ($F_{2,57}=154.88$, $p < 0.05$).”

and then:

“Post-hoc comparison t-tests (using Bonferoni correction) showed significant difference between the group C and the group A ($p<0.05$) and between group C and group B ($p<0.05$).”

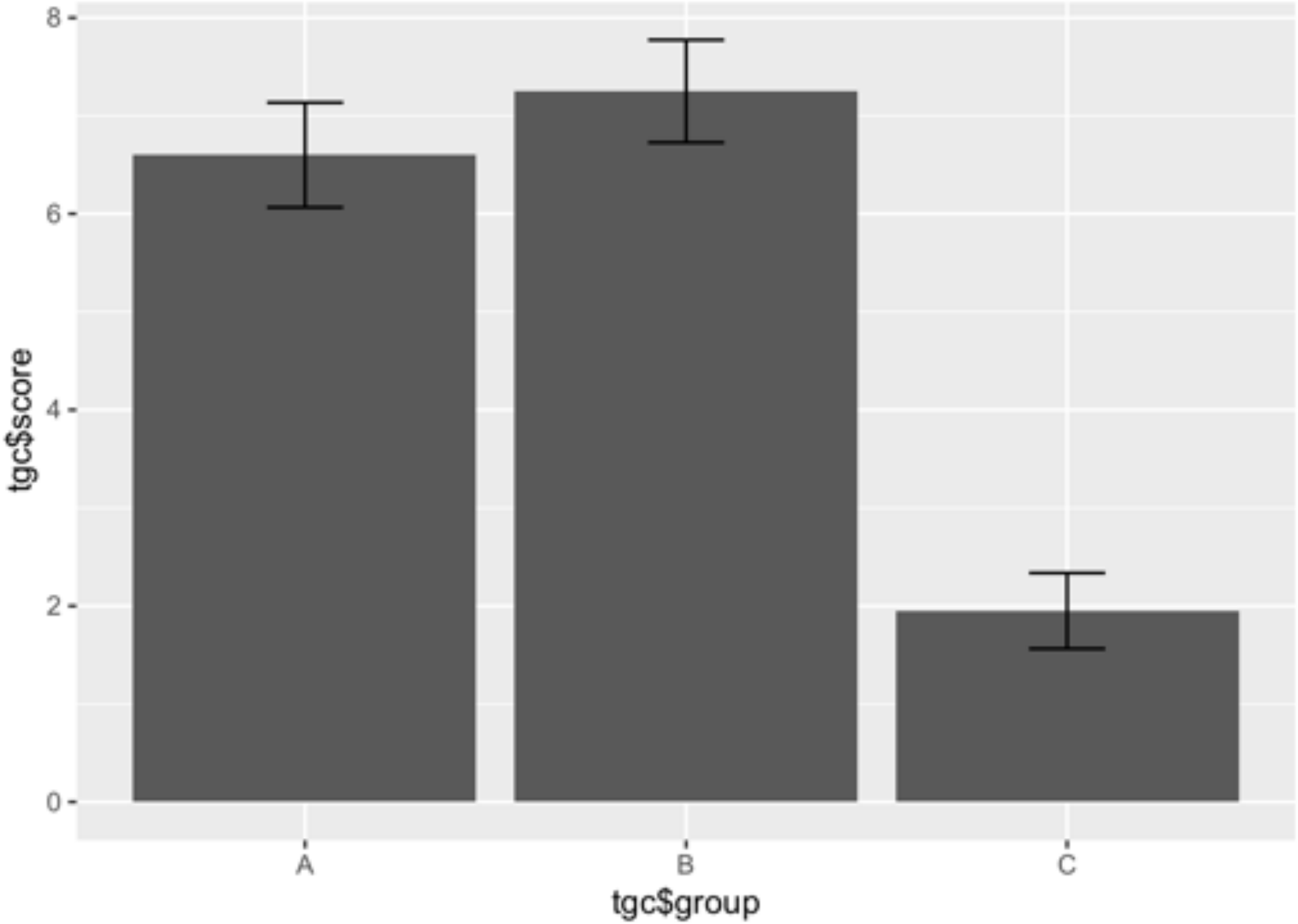
<you could also give means values to give more info>

one last thing you could find useful: how to make a graph with confident interval

```
# first we run the one-way anova
library(Rmisc)
tgc <- summarySE(dat, measurevar="score",
groupvars=c("group"))
tgc
```

```
  group  N score        sd        se        ci
1     A 20  6.60 1.1424811 0.2554665 0.5346976
2     B 20  7.25 1.1180340 0.2500000 0.5232560
3     C 20  1.95 0.8255779 0.1846048 0.3863824
```

```
ggplot(data = tgc, aes(x = tgc$group, y = tgc$score)) +
geom_bar(stat = 'identity', position = 'dodge') +
geom_errorbar(aes(ymin= tgc$score - ci, ymax= tgc$score  +
ci), width=.2, position=position_dodge(.9))
```


8
6
4
2
0
tgc$score
A
B
C
tgc$group

let’s design another experiment

biggest cause disputes in UK

do you put milk in your cup of tea before or after the boiling water?

putting milk after incorrect: milk heat unevenly and proteins in it denature (clumping), affecting taste Dr Stapley of Loughborough University

let’s try to **reproduce the experiment** by Dr Stapley

which method is the best: adding milk before or after the boiling water … by “best” = a better taste

What is your hypothesis?

(a sentence that can derive a test)

What is your hypothesis?

(a sentence that can derive a test)

H = participants will prefer the taste of tea when the milk is put after the boiling water

What are the dependent and independent variables?

1. What are our two conditions?

What are the dependent and independent variables?

1. What are our two conditions?

IV = One cup is made with milk before

One cup is made with milk after

What are the dependent and independent variables?

1. What are our two conditions?

IV = One cup is made with milk before
One cup is made with milk after

2. What are you measuring and how?

What are the dependent and independent variables?

1. What are our two conditions?

IV = One cup is made with milk before
One cup is made with milk after

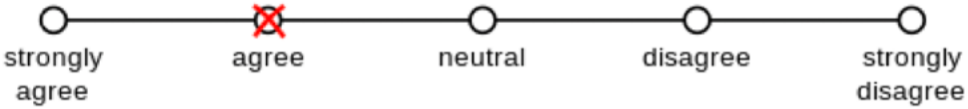
2. What are you measuring and how?

DV = tastiness … but how?

which is the best way?

[ ] ask them to rank two cup of tea

[x] ask them to evaluate teach on a scale

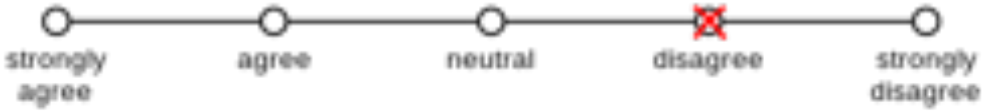


ranking = not good, does not indicate the scale, e.g. the different might be minimal

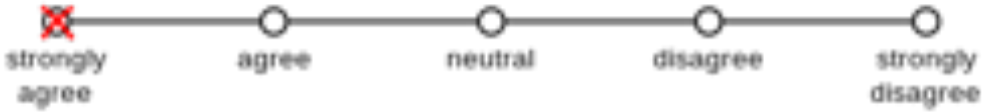
1. Wikipedia has a user friendly interface.

2. Wikipedia is usually my first resource for research.

3. Wikipedia pages generally have good images.



4. Wikipedia allows users to upload pictures easily.



if you want to collect subjective metric such as opinions, use **Likert Scale** = ordinal but treated as **continuous variable**

Likert scale::

psychometric response scale primarily used in **questionnaires** to obtain participant's preferences or degree of agreement with a statement (generally 5pt likert scale, also 7pt)

Agreement

- Strongly Agree
- Agree
- Undecided
- Disagree
- Strongly Disagree

Frequency

- Very Frequently
- Frequently
- Occasionally
- Rarely
- Never

Importance

- Very Important
- Important
- Moderately Important
- Of Little Importance
- Unimportant

Likelihood

- Almost Always True
- Usually True
- Occasionally True
- Usually Not True
- Almost Never True

writing good questions

Goal is to collect information that is:

Valid

measures the quantity that is supposed to be measured

Reliable

measures the quantity in consistent/reproducible manner

Unbiased

measures the quantity in a way that does not systematically under- or overestimate the true value

Discriminating

can distinguish adequately between respondents for whom the underlying level of the quantity or concept is different

How many cups of coffee or tea do you drink in a day?

No, ask for an answer in only one dimension, separate the question into two

(1) How many cups of coffee do you drink during a typical day?

(2) How many cups of tea do you drink during a typical day?

What brand of computer do you own?

(A) IBM PC

(B) Apple

Avoid hidden assumptions

Make sure to accommodate all possible answers

Make each response a separate dichotomous item

Do you own an IBM PC? (Circle: Yes or No)

Do you own an Apple computer? (Circle: Yes or No)

Or allow for multiple responses

What brand of computer do you own? (Circle all that apply)

Do not own computer

IBM PC

Apple

Other

Have you had pain in the last week?

[ ] Never [ ] Seldom [ ] Often [ ] Very often

Make sure question and answer options match

Reword either question or answer to match

How often have you had pain in the last week?

[ ] Never [ ] Seldom [ ] Often [ ] Very Often

Where did you grow up?

Country

Farm

City

Avoid questions having non-mutually exclusive answers

Design the question with mutually exclusive options

Where did you grow up?

House in the country

Farm in the country

City

Which one of the following do you think increases a person's chance of having a heart attack the most? (Check one.)

[ ] Smoking[ ] Being overweight [ ] Stress

Encourage to consider each possible response to avoid the uncertainty of whether a missing item may represent either an answer that does not apply or an overlooked item

Which of the following increases the chance of having a heart attack?

Smoking:	[ ] Yes	[ ] No	[ ] Don't know
Being overweight:	[ ] Yes	[ ] No	[ ] Don't know
Stress:	[ ] Yes	[ ] No	[ ] Don't know

On a scale from 1 to 5, how fun did you have using our new system?

1. not at all 2. Not really 3.undecided 4. somewhat 5. very much

Avoid biased questions

Design the question with mutually exclusive options

On a scale from 1 to 5, how would you rate your experience with our new system?

1. not fun at all 2. Not really fun 3.undecided 4. somewhat fun 5. very much fun

Rank from 1 to 3 your preference in beverage

[ ] Tea [ ] Coffee [ ] Orange Jus

Avoid ranking at all cost and rather use Likert scales

On a scale from 1 to 5 rate how much you like the following beverages

Tea:	1. not at all	2. Not really	3.undecided	4. somewhat	5. very much
Coffee:	1. not at all	2. Not really	3.undecided	4. somewhat	5. very much
Orange jus:	1. not at all	2. Not really	3.undecided	4. somewhat	5. very much

writing good questions

H = participants will prefer the taste of tea when the milk is put after the boiling water

IV = One cup is made with milk before

One cup is made with milk after

DV = tastiness

On a scale of 1 to 5 rate the tastiness of this cup?

1 very not tasty 2 not tasty 3 undecided 4 tasty 5 very tasty

Is this a within or between subjects experiment?

Do they do each condition or only one?

Is this a within or between subjects experiment?

Do they do each condition or only one?

Between: taste buds compromised by drinking a cup -> better to take two different groups

(note in this case there is nothing to counterbalance)

How many repetitions?

How many repetitions?

Only one trial (same reason, taste buds compromised)

H = participants will prefer the taste of tea when the milk is put after the boiling water

IV = One cup is made with milk before

One cup is made with milk after

DV = tastiness

On a scale of 1 to 5 rate the tastiness of this cup?

1 very not tasty 2 not tasty 3 undecided 4 tasty 5 very tasty

Between subjects with one trial only

go on and try this with a friend (in 2 weeks)

https://tinyurl.com/statsBristol

tea
milk
water

don’t tell them how you made the cup

tea
water
milk

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

conclude

design the experiment in such way that the results will be **easy to analyze**

be sure you will be **able to perform the statistical** analysis

there are many R tutorials online!

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