

Statistics for HCI
Part 3 – Gaining Power
the dreaded ‘too few participants’

Alan Dix
<https://alandix.com/statistics/chi2022/>

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Statistics for HCI – CHI 2022 – Alan Dix

1

don't you hate it when ...

=== Reviewer 2
Although the premise seems sound there are too few participants to be able to draw meaningful conclusions.

=== Reviewer 3
P= 0.673 not significant

Statistics for HCI – CHI 2022 – Alan Dix

2

statistical power – what is it?

if there is a real effect
how likely are you to be able to detect it?

avoiding false negatives

Statistics for HCI – CHI 2022 – Alan Dix

3

increasing power

standard approach ...
add more participants

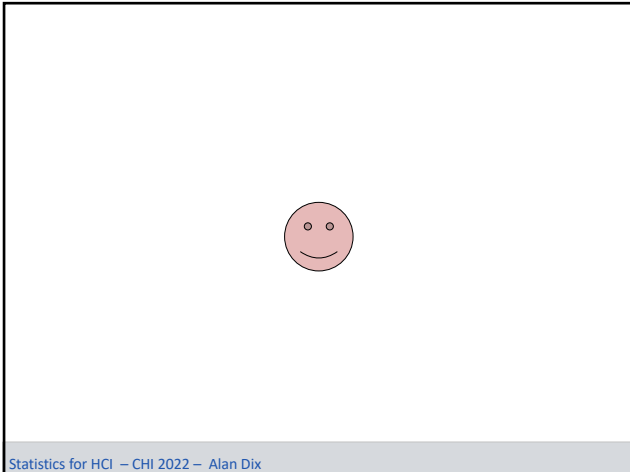
but not the only way!

can get more power ...
but often sacrifice a little generality
need to understand and explain

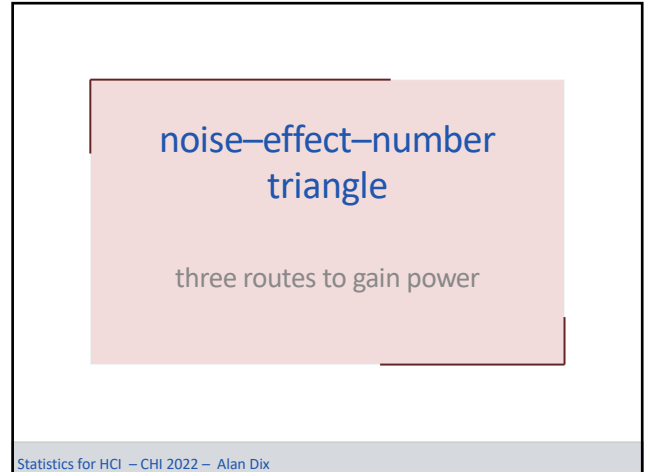
with great power comes great responsibility ;-)

Statistics for HCI – CHI 2022 – Alan Dix

4



5



6

recall Stats 101 (for simple data)

σ – standard deviation of data
often 'noise' – things you can't control or measure
e.g. individual variability

s.e. – standard error of mean (s.e.)
the accuracy of your estimate (error bars)

s.e. = σ / \sqrt{n} (if σ is an estimate $\sqrt{n-1}$)
to half standard error you must quadruple number.

Statistics for HCI – CHI 2022 – Alan Dix

7

effect size

how big a difference do you want to detect?
call it δ

the accuracy (s.e.) needs to be better than δ

$$\delta \gg \sigma / \sqrt{n}$$

Statistics for HCI – CHI 2022 – Alan Dix

8

number of participants

$$\delta \gg \sigma / \sqrt{n}$$

effect size noise

Statistics for HCI – CHI 2022 – Alan Dix

9

noise–effect–number triangle

to gain power
address any of these

not just more subjects!

Statistics for HCI – CHI 2022 – Alan Dix

10

general strategies

increase number

- the standard approach ... but ...
- square root often means very large increases

reduce noise

- control conditions (physics approach)
- measure other factors and fit (e.g. age, experience)

increase effect size

- manipulate sensitivity (e.g. photo back of crowd!)

Statistics for HCI – CHI 2022 – Alan Dix

11

Statistics for HCI – CHI 2022 – Alan Dix

12

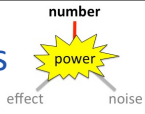
subjects

control or manipulate who

Statistics for HCI – CHI 2022 – Alan Dix

13

more subjects or trials



more subjects
– average out *between* subject differences

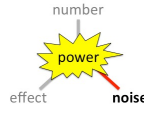
more trials
– average out *within* subject variation
e.g. Fitts' Law experiments

... but both both may need lots
e.g. to reduce noise by 10, need 100 times more

Statistics for HCI – CHI 2022 – Alan Dix

14

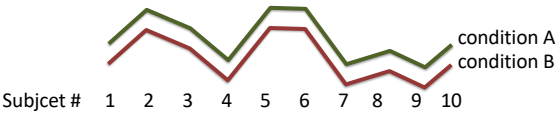
within subject/group



cancels out between subject variation

helpful if effect reasonably consistent
but between subject variability high

may cause problems with order effects, learning



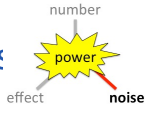
Subject # 1 2 3 4 5 6 7 8 9 10

condition A
condition B

Statistics for HCI – CHI 2022 – Alan Dix

15

narrow/matched users



aims to reduce between subject variation

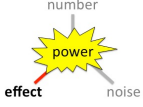
choose subjects who are very similar to each other or in some way have matched
e.g. balance gender, skills

allows between subject experiments
how do you know what is important to match?

Statistics for HCI – CHI 2022 – Alan Dix

16

targeted user group



aims to increase the effect size


choose group of users who are likely to be especially affected by the intervention

e.g. novices or older users

but ... generalisation to other users will be by theoretical argument not empirical data

Statistics for HCI – CHI 2022 – Alan Dix

17



Statistics for HCI – CHI 2022 – Alan Dix

18

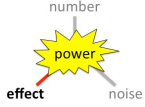
tasks

control or manipulate what

Statistics for HCI – CHI 2022 – Alan Dix

19

distractor tasks

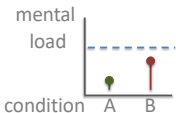


aim to saturate user's cognitive resources so make them more sensitive to intervention

e.g. count backwards while performing task

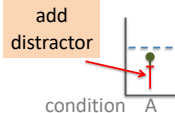
helpful when coping mechanisms mask effects

mental load



condition A B

add distractor

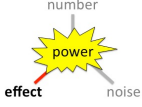


condition A B

Statistics for HCI – CHI 2022 – Alan Dix

20

targeted tasks



design a task that will expose effect of intervention

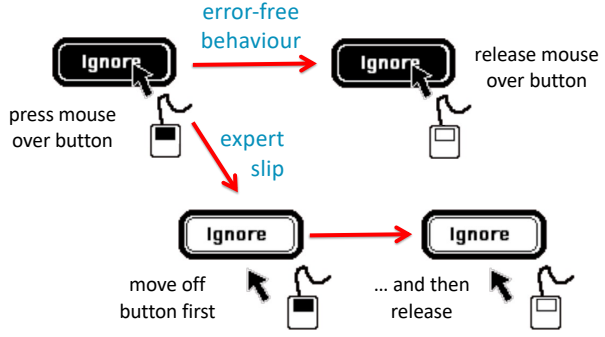
e.g. trouble with buttons paper (expert slip)

... but ... care again with generalisation!

Statistics for HCI – CHI 2022 – Alan Dix

21

example: trouble with buttons



press mouse over button

Ignore

error-free behaviour

Ignore

release mouse over button

expert slip

Ignore

move off button first

Ignore

... and then release

Statistics for HCI – CHI 2022 – Alan Dix

22

trouble with buttons (2)

novices:

- work more slowly – less likely to make slip
- notice lack of semantic feedback – so they recover

experts:

- act quickly – so make more slips
- focused on next action, so miss feedback

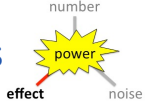
problem:

- experts slips don't happen often ... never in experiments
- needed to craft task to engineer expert slips

Statistics for HCI – CHI 2022 – Alan Dix

23

demonic interventions



extreme version create deliberately nasty task!

e.g. 'natural inverse' steering task

Fitts' Law-ish experiment

added artificial errors to cause overshoots

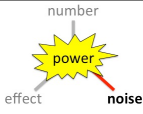
... but ... again generalisation

... and ... subjects may hate you!

Statistics for HCI – CHI 2022 – Alan Dix

24

reduced vs. wild



in the wild has lots of extraneous effects
= noise!

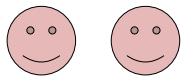
control environment => lab or semi-wild

reduced task
e.g. scripted use in wild environment

reduced system
e.g. mobile tourist app with less options

Statistics for HCI – CHI 2022 – Alan Dix

25



Statistics for HCI – CHI 2022 – Alan Dix

26