# Understanding Statistics for HCI and Related Disciplines

## Part 4 – So What?

making sense of results

Alan Dix

http://alandix.com/statistics/



## look at the data

don't just add up the numbers!



## look at the data

eyeball the raw data
are there anomalies, extreme values?
does it match your model?

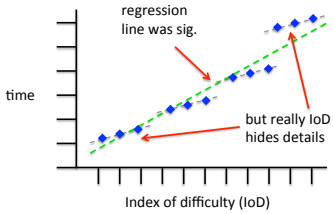
but remember
- randomness can be misleading
- data is not truth!



## example: Fitts's Law data

easy to jump straight to IoD (log distance/size)

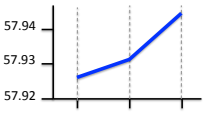
but sometimes more complex!






## choice of baseline

helps see small differences
… but magnifies them




good for rhetoric …
… but may be misleading
e.g. old aspirin advert

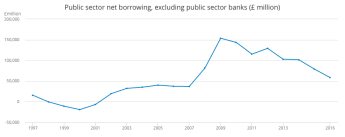

https://pixabay.com/photos/laboratory-analysis-chemistry-2815641/



## and basepoint – where do you start?


trough to peak

peak to trough




or annual average?

https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicsectorfinance/timeseries/dzls/pusf



## what have you *really* shown?

stats are about the measure,
but what does it measure



## what have you *really* shown

- think about the conditions
  - are there other explanations for data?
- individual or population
  - small #of groups/individuals, many measurements
  - sig. statistics => effect reliable for each individual
  - but are individuals representative of all?
- systems vs properties



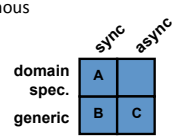
## a little story …

BIG ACM conference – 'good' empirical paper

looking at collaborative support for a task X

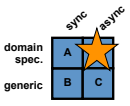
three pieces of software:

A – domain specific software, synchronous
B – generic software, synchronous
C – generic software, asynchronous






## experiment




sensible quality measures
reasonable nos. subjects in each condition

significant results p<0.05
domain spec. > generic
asynchronous > synchronous

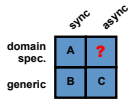


conclusion: really want asynchronous domain specific



## what's wrong with that?

interaction effects
gap is interesting to study
not necessarily good to implement




more important …
if you blinked at the wrong moment …

NOT independent variables
three different pieces of software
like experiment on 3 people!
say system B was just bad


B < A


B < C

## what went wrong?

borrowed psych method
- … but method embodies assumptions
- single simple cause, controlled environment

interaction needs ecologically valid experiment
- multiple causes, open situations

what to do?
- understand assumptions and modify





## diversity – individual/task

good *for* not just good



## don't just look at average!

e.g. overall system A lower error rate than system B

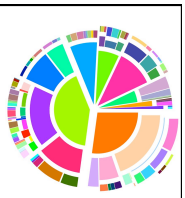
but … system B better for experts

| | | error rate | |
|---|---|---|---|
| | number | system A | system B |
| novice | 30 | 3.7% | 7.4% |
| expert | 10 | 9.6% | 2.7% |
| all | 40 | 5.2% | 6.2% |

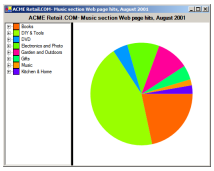


## … and tasks too

e.g. PieTree
(interactive circular treemap)



unfolding hierarchical text view good for finding small things



exploding Pie chart good for finding large things



more important to know
who or what
something is good for



## mechanism

from what happens
to how and why



## mechanism

quantitative and statistical
- what is true end to end
- phenomena

qualitative and theoretical
- why and how
- mechanism



## generalisation

empirical data
- at best interpolate

understanding mechanism allows:
- extrapolation
- application in new contexts



## example: mobile font size

early paper on fonts in mobile menus:
- well conducted experiment
- statistically significant results
- conclusion gives best font size

but … a menu selection task includes:
1. visual search (better big fonts)
2. if not found scroll/page display (better small fonts)
3. when found touch target (better big fonts)

no single best size – the balance depends on menu length, etc.

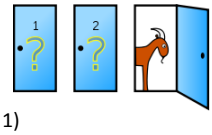


## complex issues

probability can be hard!



## Monty Hall problem




3 doors, one has car, two goats

contestant chooses one (say door 1)
Monty Hall opens one of the remaining doors with goat

should contestant change their mind?

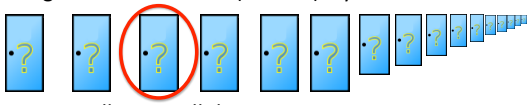
even mathematicians get confused!!

https://en.wikipedia.org/wiki/Monty_Hall_

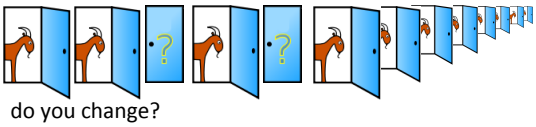


## tip: make the numbers extreme

imagine a million doors (one car) – you choose one



Monty Hall opens all the rest except one



do you change?



## lots of real examples!

DNA in court

- say DNA accuracy 1 in 100,000
- case 1:
  - person murdered after arguing with friend
  - friend matches DNA at scene
  - convincing evidence? ✓
- case 2:
  - person murdered, only clue DNA at scene
  - find person after police DNA database search
  - convincing evidence? ✗





## building for the future

adding to the discipline

## building for the future

- repeatability
  - comparisons more robust than measures
  - RepliCHI
- meta analysis
- data publishing and open science