Understanding Statistics for HCI and Related Disciplines

Part 1 – Wild and Wide

concerning randomness and distributions

Alan Dix

http://alandix.com/statistics/

unexpected wildness of random

just how random is the world?

raindrops and horse races

a story

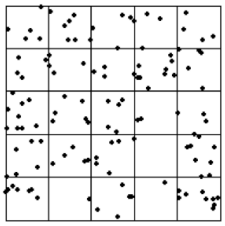
In the far off land of Gheisra there lies the plain of Nali. For one hundred miles in each direction it spreads, featureless and flat, no vegetation, no habitation; except, at its very centre, a pavement of 25 tiles of stone, each perfectly level with the others and with the surrounding land.

The origins of this pavement are unknown – whether it was set there by some ancient race for its own purposes, or whether it was there from the beginning of the world.

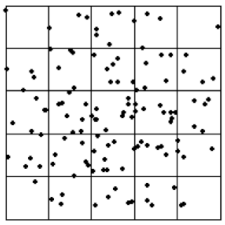
Rain falls but rarely on that barren plain, but when clouds are seen gathering over the plain of Nali, the monks of Gheisra journey on pilgrimage to this shrine of the ancients, to watch for the patterns of the raindrops on the tiles. Oftentimes the rain falls by chance, but sometimes the raindrops form patterns, giving omens of events afar off.

Some of the patterns recorded by the monks are shown on the following pages. Which are mere chance and which foretell great omens?

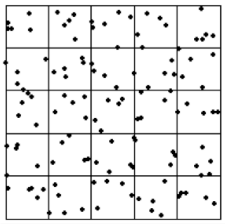
day 1



day 2

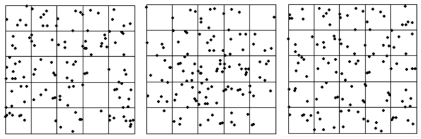


day 3



your choice?

day 1 day 2 day 3



which are by chance ...
and which are unusual?

why?

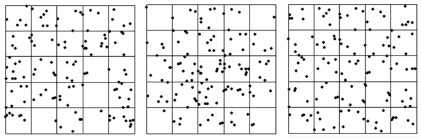
✍ .....

✍ .....

✍ .....

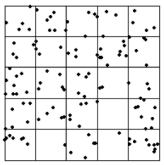
which did you choose?

day 1 day 2 day 3



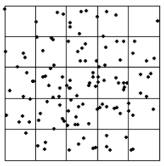
chance or omen?

day 1 – really random



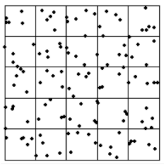
empty squares & overfull squares

day 2 – random but not uniform



clumped towards the middle

day 3 – too uniform



every square has 5 rain drops
too good to be true!

two horse races

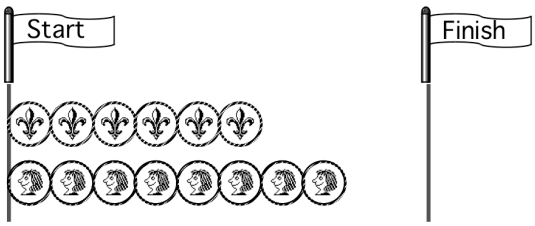
toss 20 coins

add the heads to one row
the tails to a second

the winner is the first row to 10

before you start:
what do you think will happen?

the race



did you get a clear winner?
or was it neck and neck?

the world is very random

probability head = 0.5

number of heads $\neq \frac{\text{number of tosses}}{2}$

lessons

apparent differences may be chance

real data has some bad values

e.g. Mendel's sweet peas and
electron charge discovery
were both too good!

quick (and dirty!) tip

for survey or other count data
do square root times two

Quick (and dirty!) Tip [1]

estimate variation of survey data
(for categories with small response)

- survey 1000 people on favourite colour
- say 10% of sample say it is blue
- this represents 100 people
- variation is roughly 2 x square root of 100
- that is about +/- 20 people, or +/- 2%

Quick (and dirty!) Tip [2]

For response nearer top end the same, but use the number who did not answer

- Say 85% of sample of 200 say colour is green
- 15% = 30 people did not say green
- variation is roughly 2 x square root of 30
- that is about +/- 11 people, or +/- 5.5%
- real value anything from 80 to 90 %

Quick (and dirty!) Tip [3]

If nearer 50:50 (e.g. presidential election!) then this is a little over estimate, use 1.5 instead

- Say 50% of sample of 2000 say yes to question
- 50% = 1000 people did not say green
- variation is roughly 1.5 x square root of 1000
- that is about +/- 50 people, or +/- 2.5%

N.B exact formula variance ~ n p (1-p)

… but really

far more important:

the fairness of the sample
self-selection for surveys
the phrasing of the question

bias and variability

is it fair?
is it reliable?

bias

systematic effects
that skew results one way or other
e.g. choice of LinkedIn vs. Snapchat survey
WEIRD (Western, Educated, Industrialized, Rich, and Democratic)

bias persists no matter your sample size
but can sometimes be corrected
e.g. sample variance … hence the n-1

bias vs. variability

increase sample size or reduce variability

high variability with no bias
– measurements equally likely to be more or less than real value, but may be far away in either direction
– poor estimate of right thing

low variability with bias
– Measurements consistent with one another, but systematically one way
– good estimate of wrong thing

eliminate bias or model it

independence and non-independence

do buses really come in threes?

independence

different kinds of independence:
- measurements (normal kind)
- factor effects
- sample prevalence

non-independence often increases variability
… but may cause bias too

independence – measurements

measurements may be related to each other:

order effects
sequential measures on same person (learning, etc.)

'day effects'
e.g. sunny day improves productivity

experimenter effects
your mood effects participants

all increase variability

independence – factor effects

Relationships and correlations between things you are measuring (or failing to measure)

can confuse causality

e.g. death rates often higher in specialist hospitals

… because more sick people sent there

even reverse effects

… Simpson Paradox

Simpson's Paradox

You run a course with some full-time and some part-time students

You calculate:

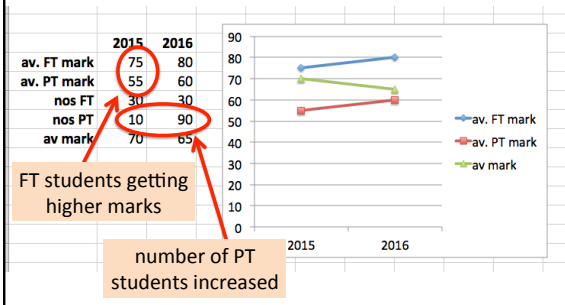
- average FT student marks increase each year
- average PT student marks also increase each year

University complains your marks are going down

Who is right?

Simpson's paradox – you both are!

	2015	2016
av. FT mark	75	80
av. PT mark	55	60
nos FT	30	30
nos PT	10	90
av mark	70	65



FT students getting higher marks

number of PT students increased

independence – sample

the way you obtain your sample

internal – subjects related to each other

e.g. snowball sampling

if not external – increases variance

external – subject choice related to topic

e.g. measuring age based on mobile app survey

may create bias

crucial question

Is the way I've organised my sample likely to be independent of the thing I want to measure?

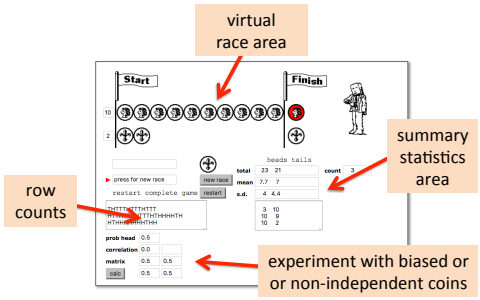
e.g. Fitts' law with different colour targets. Are student subjects likely to behave differently than the general population?

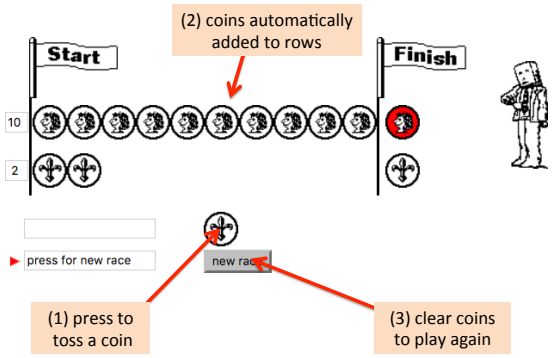
play!

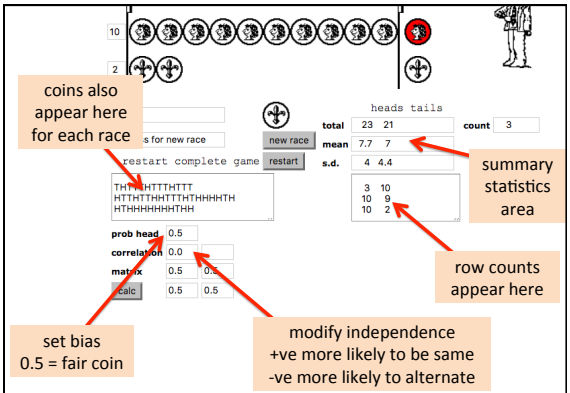
experiment with
bias and independence

http://www.meandeviation.com/tutorials/stats/morecoin.html

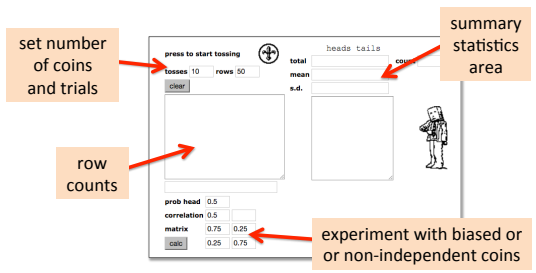
virtual two horse races

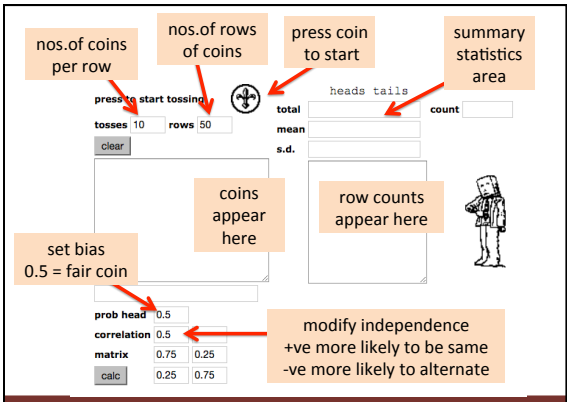




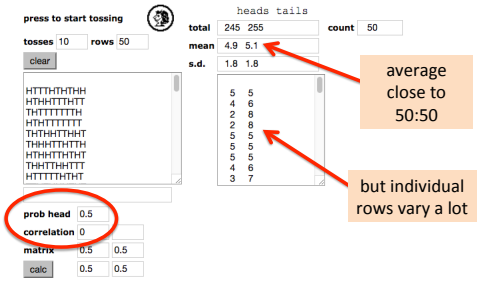


more (virtual) coin tossing

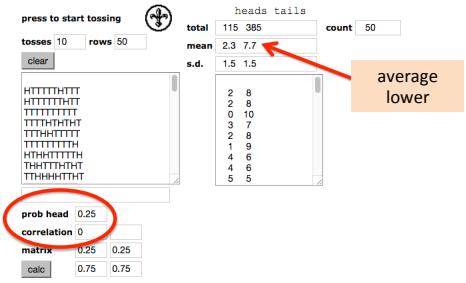




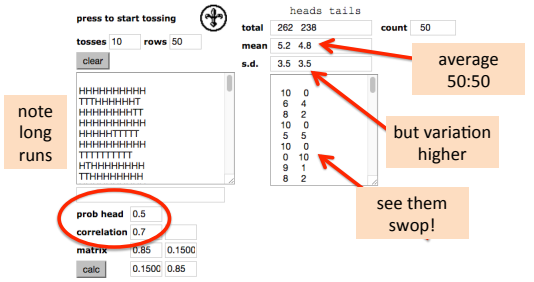
fair coin (independent tosses)



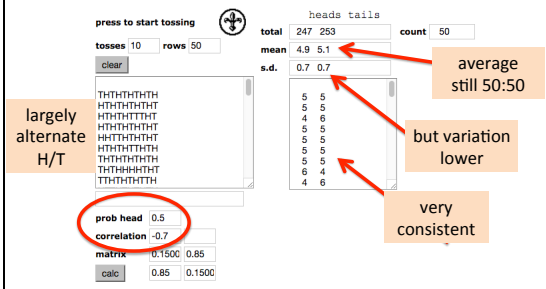
biased coin (independent)



positive correlation (runs)



negative correlation (alternation)



distributions

discrete or continuous
bounds and tails

what kind of data

continuous:

e.g. time to complete task (12.73 secs)

discrete ...

arithmetic:

e.g. number of errors (average makes some sense)

ordered/ordinal:

.g. satisfaction rating (?average rating?)

nominal/categorical:

e.g. menu item chosen ((File+Font)/2 = Flml ?)

finite or unbounded

number of heads in 6 tosses

discrete, finite

number of heads until first tail

discrete, unbounded

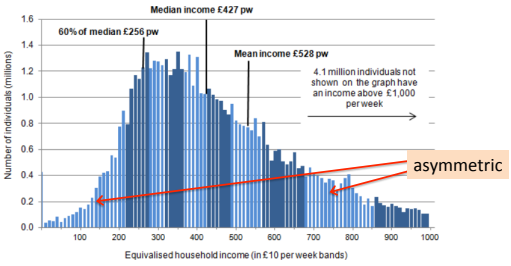
wait before next bus

continuous, bounded below (zero), … but not above!

difference between heights

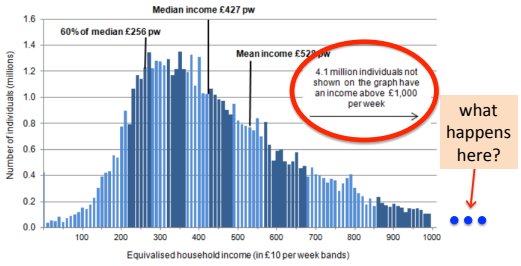
continuous, (sort of) unbounded

distribution graph (e.g. UK income 2011/12)



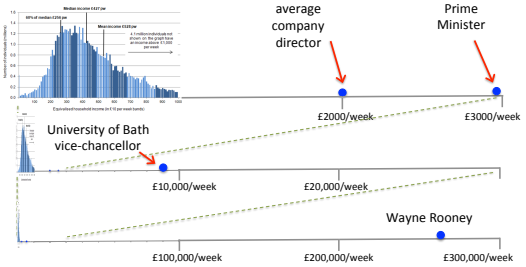
From: Sustainable Development Indicators, Office of National Statistics, July 2014
http://webarchive.nationalarchives.gov.uk/20160105183323/http://www.ons.gov.uk/ons/rel/wellbeing/sustainable-development-indicators/july-2014/sustainable-development-indicators.html

long tail



From: Sustainable Development Indicators, Office of National Statistics, July 2014
http://webarchive.nationalarchives.gov.uk/20160105183323/http://www.ons.gov.uk/ons/rel/wellbeing/sustainable-development-indicators/july-2014/sustainable-development-indicators.html

long tail (ctd)



From: Sustainable Development Indicators, Office of National Statistics, July 2014
http://webarchive.nationalarchives.gov.uk/20160105183323/http://www.ons.gov.uk/ons/rel/wellbeing/sustainable-development-indicators/july-2014/sustainable-development-indicators.html

one or two tails, …

what is your question?

do you care which direction?

is error rate *higher*?

– one tailed (discrete)

are completion times *different*?

– two tailed (continuous)

normal or not

approximations
central limit theorem
power law

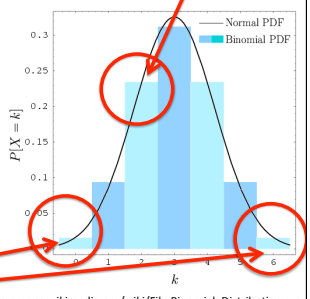
approximations

may approximate one type of distribution with another

esp. using Normal

continuous / discrete

bounded / unbounded



https://commons.wikimedia.org/wiki/File:Binomial_Distribution.svg

why is Normal normal?

central limit theorem

if you:

- average lots of things (or near linearly combine)
- around the same size (so none dominates)
- nearly independent
- and have finite variance

what?

then you get Normal distribution

non-Normal – what can go wrong?

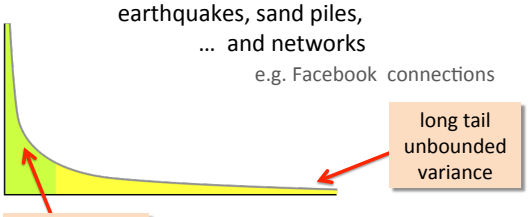
non-linearity – e.g. thresholds

+ve / -ve feedback
  snowflakes and clouds
  bi-modal exam marks

unbounded variance
  the more you sample the bigger the variation
  used to be rare e.g. wage/wealth distributions
    … but now Power Law …

power law – scale free

earthquakes, sand piles,
… and networks
e.g. Facebook connections



long tail unbounded variance

small number of frequent elements

power law data is NOT Normal even when averaged

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