Understanding Statistics for HCI and Related Disciplines

Part 1 – Wild and Wide

concerning randomness and distributions

Alan Dix

http://alandix.com/statistics/

An 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 cultivation; 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?

your choice?

why?

✍ ...........................................

✍ ...........................................

✍ ...........................................


which did you choose?

day 1  day 2  day 3

chance or omen?

which did you choose?

day 1 – really random

empty squares & overfull squares

day 1 – really random

which did you choose?

day 2 – random but not uniform

clumped towards the middle

day 2 – random but not uniform

day 3 – too uniform

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

day 3 – too uniform
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 ≠ \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 $2 \times \sqrt{100}$
- 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 $2 \times \sqrt{30}$
- 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 $1.5 \times \sqrt{1000}$
- That is about +/- 50 people, or +/- 2.5%

N.B exact formula variance $\sim n p (1-p)$

... 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

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

Increase sample size
or reduce variability

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!**

<table>
<thead>
<tr>
<th>Year</th>
<th>FT mark</th>
<th>FT mark</th>
<th>PT mark</th>
<th>PT mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>75</td>
<td>80</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>2016</td>
<td>80</td>
<td>85</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

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

(2) coins automatically added to rows

(1) press to toss a coin

(3) clear coins to play again

(2) coins automatically added to rows

(1) press to toss a coin

(3) clear coins to play again

more (virtual) coin tossing

set number of coins and trials

summary statistics area

row counts

experiment with biased or non-independent coins

more (virtual) coin tossing

set number of coins and trials

summary statistics area

row counts

experiment with biased or non-independent coins

virtual two horse races

virtual race area

row counts

experiment with biased or non-independent coins

virtual two horse races

virtual race area

row counts

experiment with biased or non-independent coins

virtual two horse races

virtual race area

row counts

experiment with biased or non-independent coins
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:
  e.g. satisfaction rating (?average rating?)
  nominal/categorical:
  e.g. menu item chosen (%File=Font)/2 = File?

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)

Expected median income is £17,000 per week.

[Graph showing income distribution with an asymmetric tail]

From: Sustainable Development Indicators, Office of National Statistics, July 2014

long tail

Expected mean income is £17,000 per week.

[Graph showing income distribution with a long tail on the right]

[Highlighted area on the graph with a question mark]

what happens here?

From: Sustainable Development Indicators, Office of National Statistics, July 2014

long tail (ctd)

[Graph showing income distribution for specific groups]

average company director
Prime Minister
University of Bath vice-chancellor
Wayne Rooney

[Income levels labeled: £100,000 to £200,000 per week]

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)

From: Sustainable Development Indicators, Office of National Statistics, July 2014
approximations

may approximate one type of distribution with another
esp. using Normal

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

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

power law data is NOT Normal even when averaged

small number of frequent elements

long tail unbounded variance