experiments in HCI

- people ... people ... people
- how to avoid the dreaded 'n.s.'

what do you want to know?

- product A is better than product B
  - easy but boring (good for the glossies...)
- formative evaluation
  - what's good what's bad
  - statistics not always necessary
- small components and detail
  - similar to traditional psychology
- design principles and methods
  - building systems equally good ... arghh!!!!

raw materials

- people
  - on their own complex and variable
  - in groups ...
- computer interfaces
  - expensive to build
    ... but can you get away with a mock-up?

people are different

- skills, knowledge, expertise
- tiredness, illness, motivation
- especially when they think

$\Rightarrow$ high variability

$\Rightarrow$ no significant result ...

complex effects

- transfer effects
  - learning and interference
- time of day effects
  - tiredness, post-lunch dip, ...
  - aliasing with other effects
- expert slips
  - making them happen!

... so are contexts and groups

- field vs. laboratory
- social dynamics

  $\Rightarrow$ yet more variability!
- developing group relationships

  $\Rightarrow$ difficulties for pairing

subjects?

- getting them! – especially groups
- who are they?
  - psych. and CS undergrads – typical?
  - OK (?) for low level psychology
  - computer expertise, 2nd guessing, ...
- surveys – self selecting

statistics is the least of your problems!

solutions

- paired designs
  - beware transfer effects
- low-level measures
  - no thinking!
- qualitative as well as quantitative
  - backup if the stats fails!
  - give meaning to the numbers

low-level measures

- single user
  - keystroke timings, slips
- multi-user
  - conversational units, glances
- look for underlying effects
  - e.g. trouble with buttons
- data logging
  - long-term studies, vast data sets

combining data

- broad and shallow
  - survey, logging, large simple experiments
  - quantitative

  → *statistics*

✚ deep analysis of a few subjects
  - video logs, post-task walkthrough, matching, knowledge elicitation, anecdotal evidence
  - qualitative

  → *meaning*

what can you say?

① 80% of the sample had this problem
- true, but so what ...

② 80% of users will have this problem
- needs statistics, confidence interval etc.

③ at least one user has this problem
- make own judgement about generality

④ the problem happens for this reason
- needs deep data ...

do you need statistics?

- to make general assertions – YES
- to quote numbers – YES
- to understand why – need deep data

✔ plan for non-significance

✔ qualitative + quantitative