Basic Issues in Experimental Design

Chapter 3 of "Empirical Methods for Artificial Intelligence”, Paul Cohen

Content

- Experiments and control
- Variables
  - extraneous and noise variables
- Four spurious effects
  - floor, ceiling, order, and regression effects
- Sampling bias
- Applicability to our work and discussion

Experiments

Aim: Discover the causal relationship between factor *X* and *Y*. This is usually done by looking at how *x* (representing *X*) influences *y* (representing *Y*).

Observation

- *x* cannot be directly manipulated
- *x*: predictor
- *y*: response

Manipulation

- *x* can be directly manipulated
- *x*: independent
- *y*: dependent

Control

Purpose: Rule out alternative explanations for the results.
Method: Control all plausible alternatives.

Treatment condition:

x & *everything else* $\rightarrow y_t$

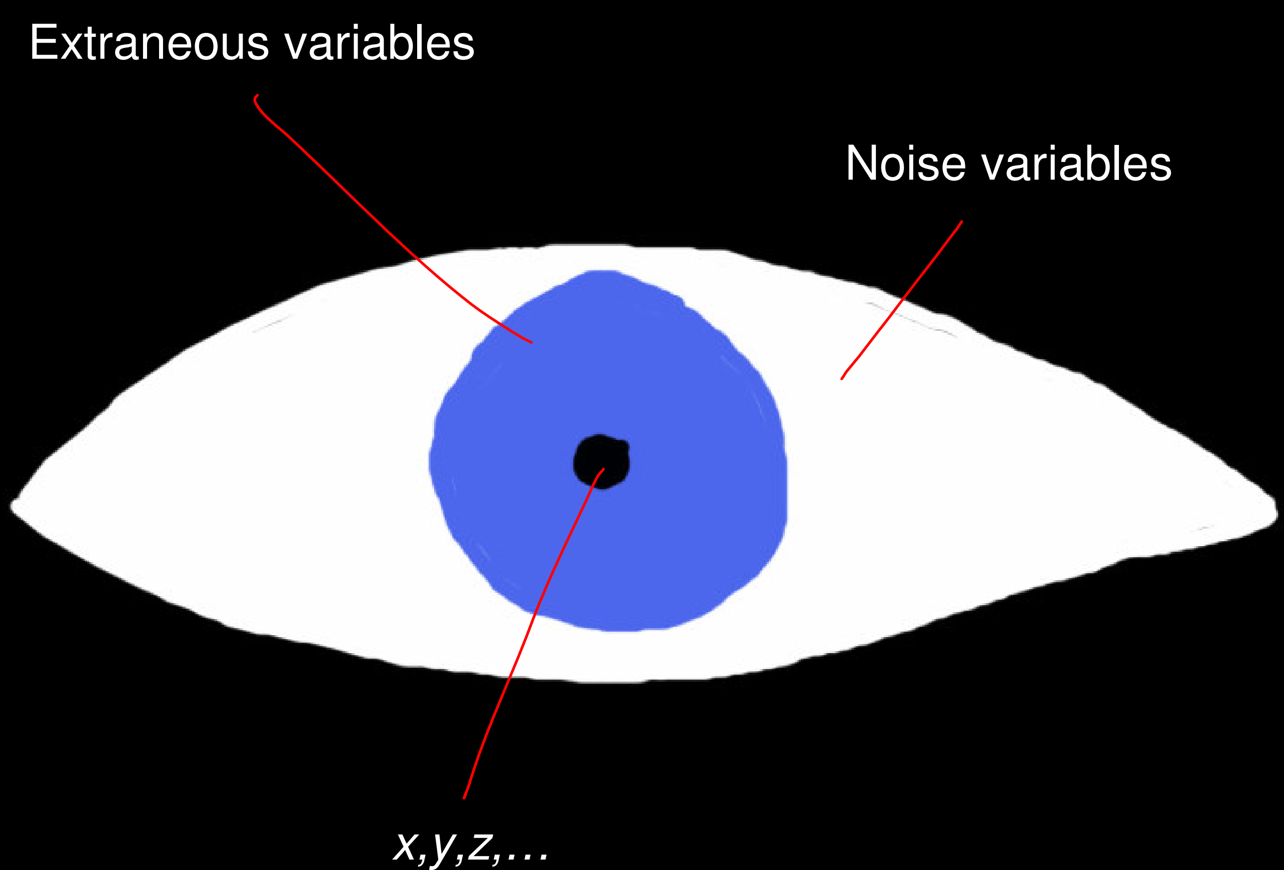
Control condition:

everything else $\rightarrow y_c$

We have to show that:

$y_t \neq y_c$

Variables



Variables

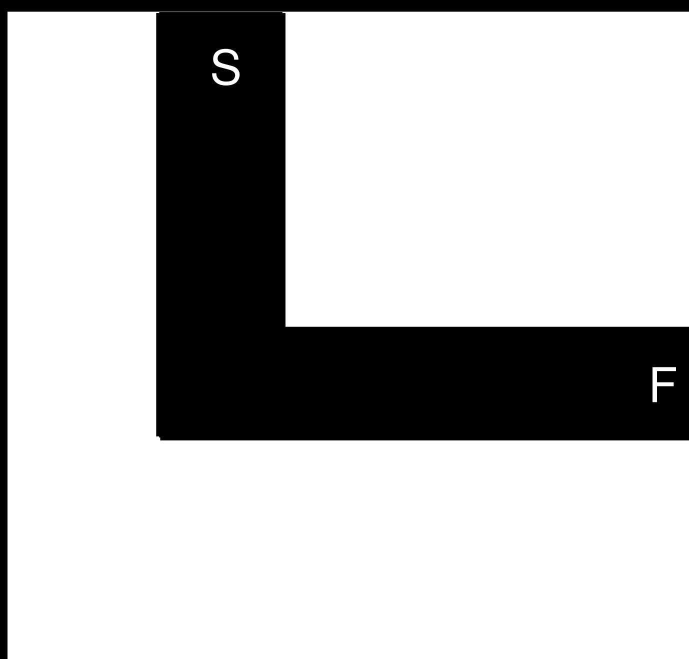
- We cannot control all variables
- Extraneous variables
  - plausible causes
  - controlled directly
- Noise variables
  - assumed to have negligible effect
  - give raise to variation
  - controlled through random sampling

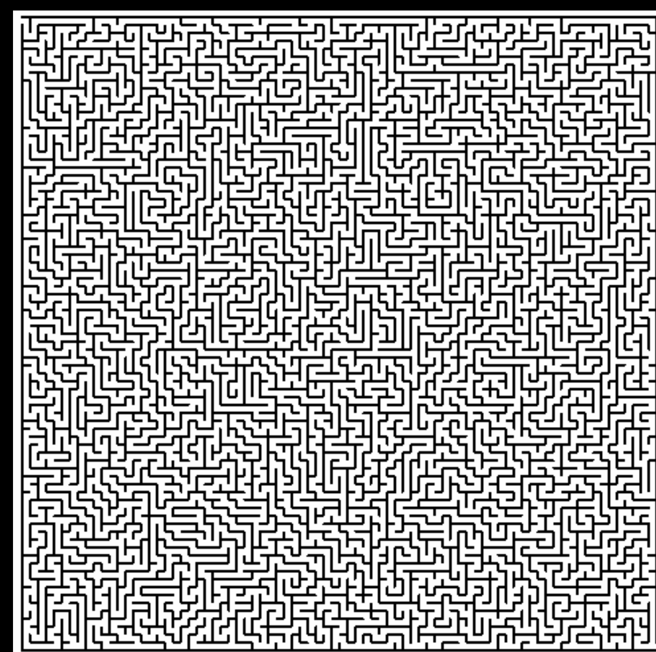
Four Spurious Effects

- or four things to be aware of when designing experiments

The Ceiling and the Flooring Effect

Are problems so easy or so hard that the results are trivial?





Regression Effect

Random samples from a distribution tend to regress towards the mean.

During debugging/fine tuning, don't work only with the experiments for which the performance is lowest.

- Because, if there is some non-deterministic component, the performance is likely to get better, "magically".

The Order Effect

- If the order of successive trails influence the outcome (and it usually does even when it is not obvious), take care:
  - sensors/actuators change calibration.
  - for software, warm-up runs are usually done to cancel out effects of caching and disk buffering.
- Solution: *counterbalancing* or a subset of all possible orderings.

Sampling Bias

Differences between a sample and the population it represents should result only from random chance.

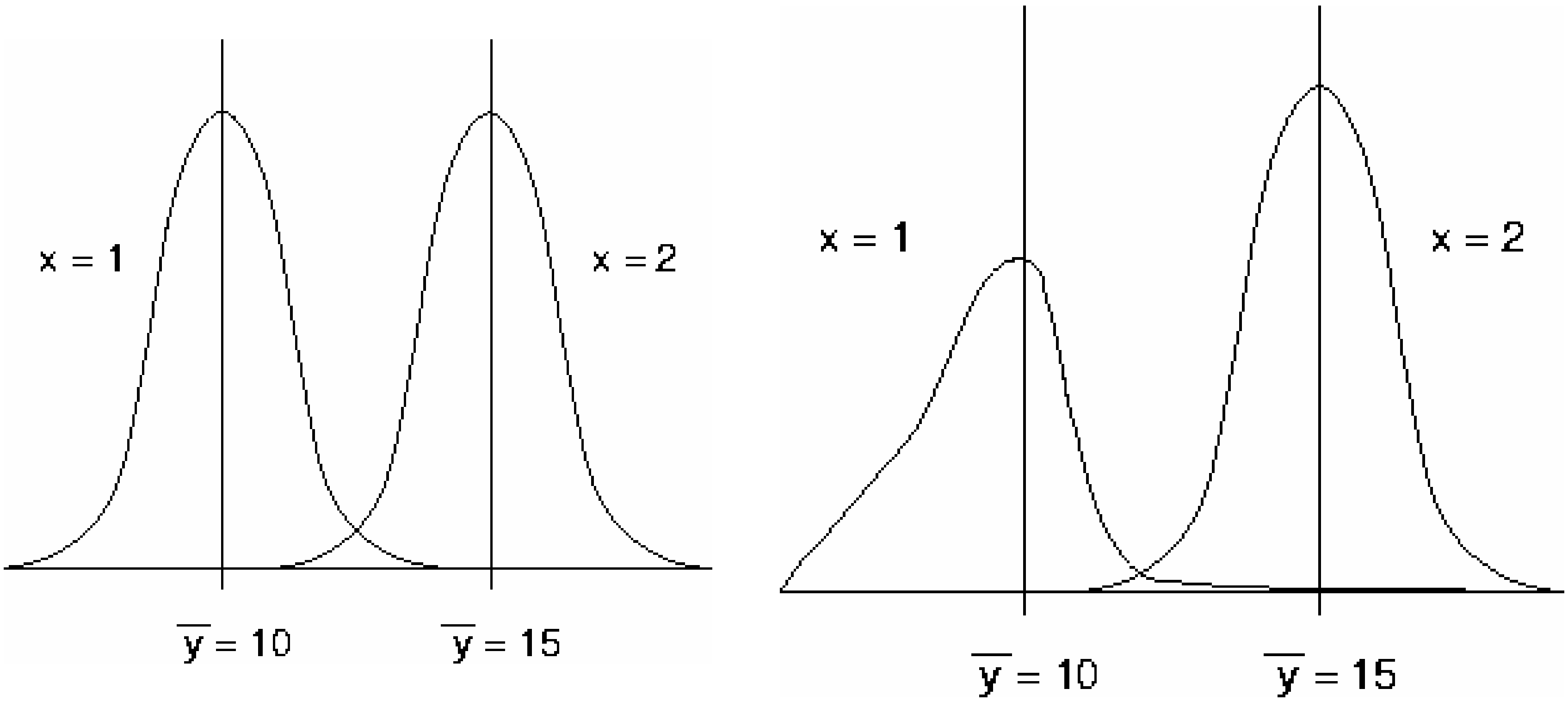
When differences arise for reasons other than chance, you have introduced sampling bias into your research.

Sampling Bias

Example -1936 US presidential election:

- Literary Digest: *2.000.000* opinions
  - prediction - Landon: 57%, Roosevelt: 43%.
- George Gallup: *300.000* opinions
  - prediction: Roosevelt would win.
- Roosevelt won
  - Literary Digest had only asked car-owners (middle- and upper-class).

Sampling Bias



Guidelines for Experimental Design

Make the *experimental procedure* explicit.

Make an example of a *data table*.

Make an example of the *analysis*.

Consider possible *results and interpretations*.

Make sure that you answer the right questions (preferably *research questions*).

Application to Our Work

- In the book, most examples are based on expert systems and planners.
- However, the effects and pitfalls are general.
- Can we find more examples of ceiling, flooring, regression and order effects?
- Don’t a lot of scientists use the *sampling bias* to get published?
  - You can be selective about which problems you present results for and show that your new, shiny method is better than the rest…