“The principles of experimentation, illustrated by a psycho-physical experiment”
or
Sir Ronald Fisher’s Tea Tasting Lady

Marek J. Druzdzel

University of Pittsburgh
Department of Information Sciences
and Intelligent Systems Program

marek@sis.pitt.edu
http://www.pitt.edu/~druzdzel

Sir Ronald A. Fisher (1990). “The principles of experimentation, illustrated by a psycho-physical experiment”. In: R.A. Fisher. “The Design of Experiments”. Chapter 2, 11-26

Overview

- The problem
- Experimental design
- Null hypothesis
- Randomization
- Sensitiveness
- Concluding remarks

The problem

A lady can taste which was poured first: milk or tea. Can she really?

The experiment

- The lady (subject) will be presented with 8 cups of tea
  - four of each kind
  - in random order
  - one at a time
- She will be asked to taste each cup and say either “Milk first” or “Tea first.”

The math behind the experiment

- How many possible sequences are there?
- Choosing 4 objects out of 8:

$$\binom{8}{4} = \frac{8!}{4!\,(8-4)!} = \frac{8*7*6*5}{4*3*2*1} = 70$$

- How many cups to present?

cups milk first	cups tea first	p all right	p one wrong
4	4	1/70	16/70
3	3	1/20	9/20

Why do we compute this? We want to measure a level of surprise by the results.

Statistical significance

- Significance level is set a-priori by the experimenter.
- “Smallness of probability [required] … before he would be willing to admit that his observations had demonstrated a positive result.”
- “five per cent … a standard level of significance … They are prepared to ignore all results which fail to reach this standard.”
- Fisher states: “We may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result.”

Null hypothesis

- Results are divided into two classes with opposed interpretations: H_0 (called *null hypothesis*) and H_1.
- Null hypothesis can never be proved, but possibly disproved.
- Null hypothesis must be exact: "The lady cannot discriminate at all by taste whether the milk or tea was added first."
- The alternative hypothesis is its complement: "The lady can discriminate by taste whether the milk or the tea were added first."

Randomization

- Equalization must always, to some extend, be always incomplete: make sure that the frequency distribution is not violated
- Random order
- Every cup has equal chance
- Guarantee validity of test of significance (against those causes of disruption that have not been eliminated)

Factors in the experiment:

- thickness of cups
- quantities of milk
- strength of tea
- temperature

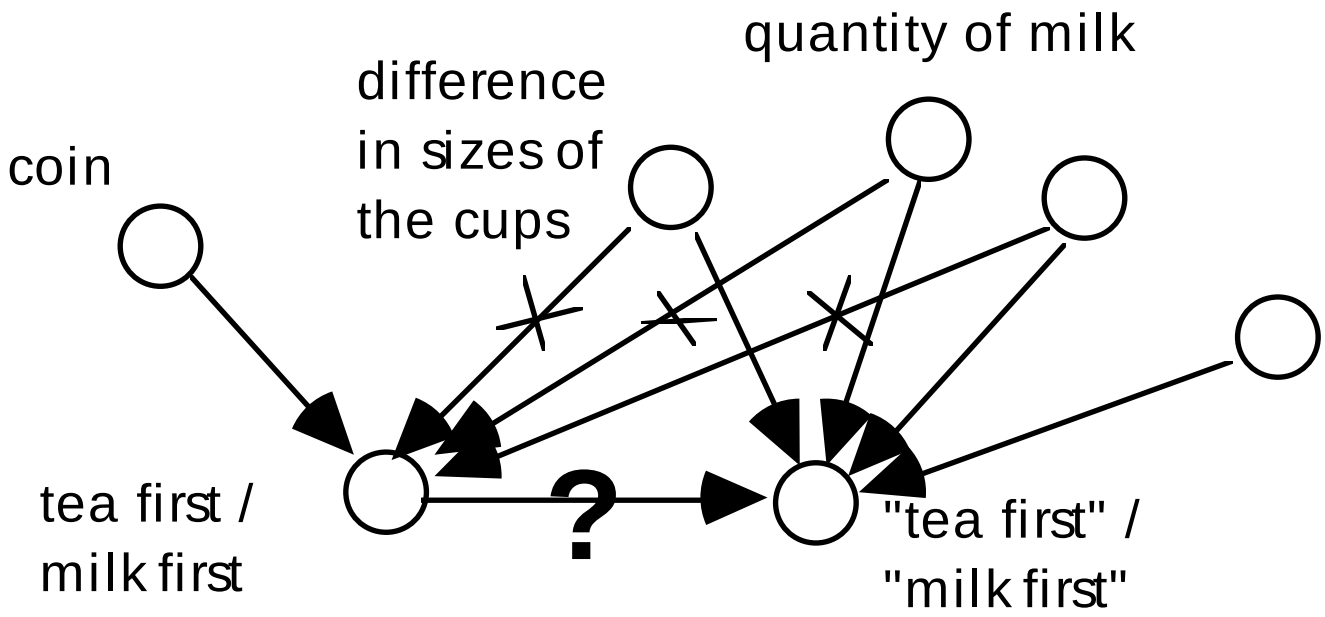
Sensitiveness

- Size of the experiment
- Repetition of the experiment
- Qualitative improvements:
  - reorganization of structure
  - refinements of technique

cups milk first	cups tea first	p all right	p one wrong
4	4	1/70	16/70
3	3	1/20	9/20
6	6	1/924	36/924

- Independent treatment of cups
- Unequal numbers of the two treatments
- Refinements may be important for a phenomenon

Causal graph for the design



- **Could there have been an alternative connection between tea first/milk first and "tea first"/"milk first"?**
- **If so, then it was broken by randomization, so this experiment is capable of excluding quite a number of graphs (except the ones that have a direct causal connection between independent and dependent variables.**

The principles of experimentation

Significance, null hypothesis, randomization, sensitiveness

Concluding remarks

- This is a fundamental paper. It introduces several ideas, the main being randomization.
- “Fisher's lady” really existed. She was the wife of a colleague of Fisher. She claimed that she could taste the difference in tea. The actual experiment has most probably never been conducted.
- How does 0.05 come about? There is an anecdote that Fisher was once asked at a party what an acceptable level of significance is. He said casually that 1/20 might do and this is how 0.05 has become a magical level.
- Fisher was a frequentist. This is another reason why he needed randomization: He needed it to obtain a distribution.