6. Case studies from 'Group A'

• The case studies of 'Group A' are simple real case studies where the answer is well known.

⇝ But, there will be some interesting statistical (analytics) and scientific issues to be discussed.

⇝ The method and approach to the data analysis are (hopefully) straightforward.

◇ In what follows, we present six case studies.

A.1 Length of visits of users entering a web site

• Once a user enters a web site how many pages or links within the web site does that user visit?

$\rightsquigarrow$ The answer to this question may suggest actions to improve the web site.

$\rightsquigarrow$ If similar 'distributions' for the number of pages visited per user are observed at different web sites, then maybe some laws can be established for all web sites.

$\rightsquigarrow$ Research efforts in this area are directed at finding these laws.

$\rightsquigarrow$ This is a small part of the current effort to understand human behaviour on the web.

• A client provides you with a data set giving a random sample of the length of visits of users entering his company's web site during one day.

$\rightsquigarrow$ The length of the visit is an estimate of the total number of pages seen by each user and is based on web server logs, thus it counts only pages recorded by the web server.

$\rightsquigarrow$ Pages cached in the user's browser or in a cache proxy server are unknown.

• The client thinks that those users who visited more than 100 pages are probably crawlers or maybe different people logged into the same IP address.

$\rightsquigarrow$ Therefore, the client proposes to exclude those users.

$\diamond$ What the clients would like from you is an 'exploratory data analysis' of these data and he would like also to know what 'distribution model' could be used for the user's behaviour on his company's web site.

Data

- Source:

  `www.statoo.com/DATA/AC/Awebsite.txt`

- Size: $50'000$ rows, 1 column (*i.e.* variable).

- Variable:
  - `Length`: numerical variable summarising the length of the visit to the web site.

"On the Internet, nobody knows you're a dog."

A.2 Effects of anaesthetics

• A study was carried out at a hospital to compare the effects of four different types of anaesthetic as used in major operations.

$\rightsquigarrow$ 80 patients undergoing a variety of operations were randomly assigned to one of the four anaesthetics and a variety of observations were taken on each patient both before and after the operation.

$\rightsquigarrow$ The client is interested on just one of the measurements made, namely the time, in minutes, from the reversal of the anaesthetic until the patient opened the eyes.

$\diamond$ The client would like to know if there is any evidence of differences between the effects of the four anaesthetics?

Data

- Source:

  `www.statoo.com/DATA/AC/Aanaesthetics.txt`

- Size: 20 rows, 4 columns (*i.e.* variables).

- Variables:
  - `A`: time, in minutes, from reversal of anaesthetic till the eyes open for each of the patients treated by anaesthetics A;
  - `B`: same time for the patients treated by anaesthetics B;
  - `C`: same time for the patients treated by anaesthetics C;
  - `D`: same time for the patients treated by anaesthetics D.

A.3 Testing a manufacturer's product claim

• A manufacturing client makes the claim that its device, when placed on a motor vehicle engine, would cause hydrocarbon, carbon monoxide (CO) and carbon dioxide (CO_2) emissions to steadily decrease.

$\rightsquigarrow$ In order to investigate these claims, tests were conducted using the company's device on one of the manufacturer's cars.

$\rightsquigarrow$ Hydrocarbon, CO and CO_2 emissions were measured for the car owned by the manufacturer.

$\rightsquigarrow$ The car was operated under the same conditions on 13 different days and 4 replicates of the experiment were performed on each of the testing days.

• The device was not placed on the engine until after the second day's measurements and remained on the engine while measurements were made on 9 additional days.

$\rightsquigarrow$ The device was then removed from the engine and 2 additional sets of measurements were made.

$\diamond$ The client thinks that the test days and the replicates are not important and he would like know if there is an effect due to the device on any or all of the three emissions.

Data

- Source:

  `www.statoo.com/DATA/AC/Aproductclaim.txt`

- Size: 52 rows, 6 columns (*i.e.* variables).

- Variables:
  - `day`: test day $(1, 2, \ldots, 13)$;
  - `rep`: replicate $(1, 2, 3, 4)$;
  - `dev`: device ($1 =$ 'present', $0 =$ 'absent');
  - `hc`: hydrocarbon (parts per million);
  - `co`: carbon monoxide, CO ($\%$ of volume);
  - `co2`: carbon dioxide, CO_2 ($\%$ of volume).

A.4 Comparing teaching methods

• The aim of the client's study is to compare teaching methods.

$\rightsquigarrow$ As such, 45 students were divided randomly into five equal-sized groups.

$\rightsquigarrow$ Two groups were taught by the currently used method (the 'control' method), and the other three groups by one the three new methods.

$\rightsquigarrow$ At the end of the experiment, all students took a standard test and the results (marks out of 30) were reported.

$\diamond$ The client would like to know what conclusions can be drawn about differences between teaching methods.

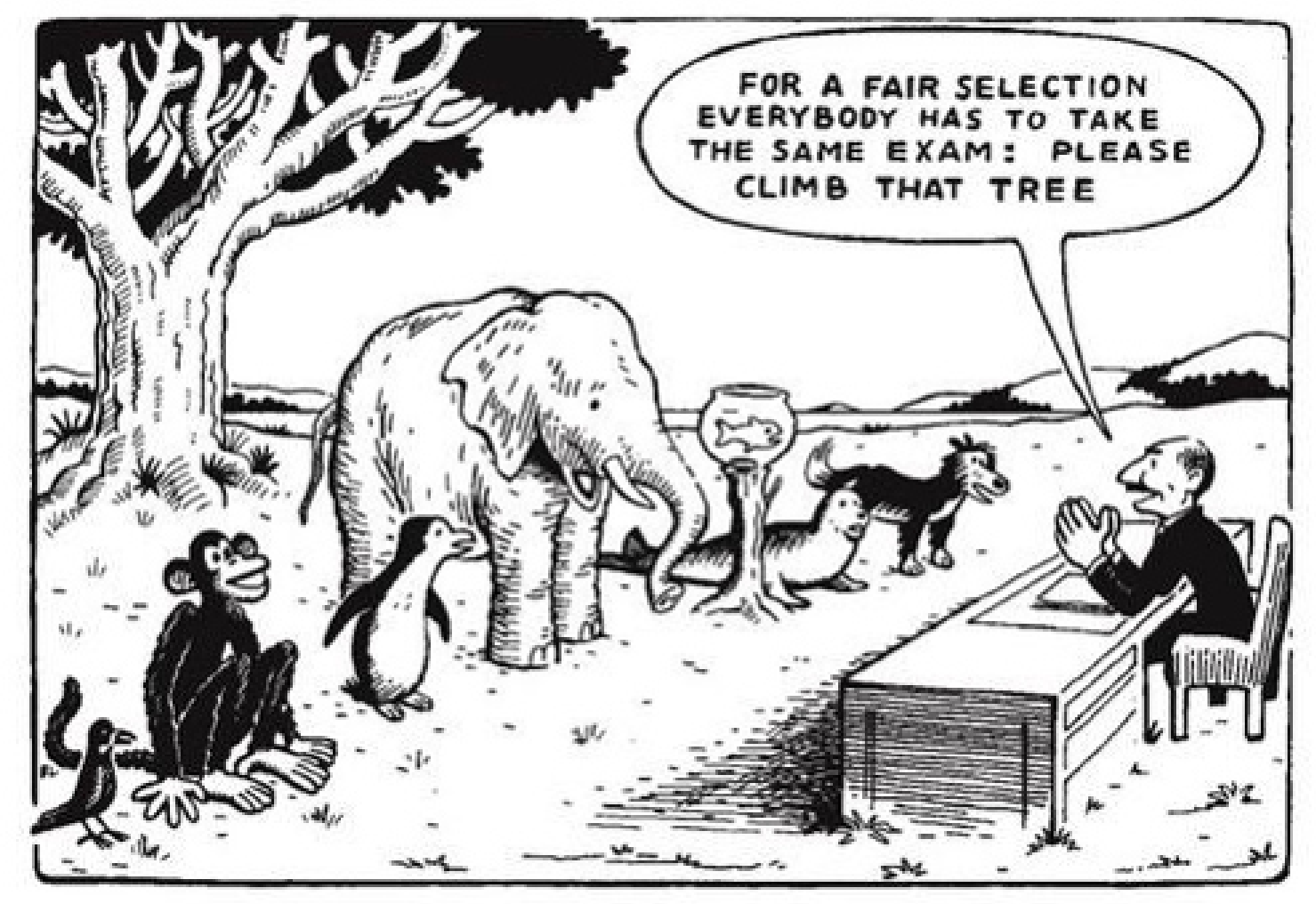
Data

- Source:

  `www.statoo.com/DATA/AC/Ateaching.xls`

- Size: 9 rows, 5 columns.

- Variables containing the results (marks out of 30):
  - `Group A (control)`: of the first 'control' group;
  - `Group B (control)`: of the second 'control' group;
  - `Group C (praised)`: of the group taught by the 'praised' method;
  - `Group D (reproved)`: of the group taught by the 'reproved' method;
  - `Group E (ignored)`: of the group taught by the 'ignored' method.



A.5 Job promotion discrimination

• An organisation may face a discrimination complaint. The plaintiff is a high-ranking employee who was not promoted after the last election of the chairman of a governing commission.

$\rightsquigarrow$ He believes that the reason for his non-promotion is that only candidates who contributed to the election campaign of the winning candidate for chairman of the governing commission were promoted. Other candidates like him who did not contribute to the campaign of the winning candidate for chairman of the governing commission were not promoted.

• Indeed, including the plaintiff, 10 employees were up for promotion. Seven were promoted and of these seven, six had made financial contributions to the campaign of the winning candidate for chairman of the governing commission. The remaining four did not contribute to the campaign.

• Moreover, the organisation's promotion procedure is based on the scores from a standard test that was taken by all 10 candidates. The scores are ranked and the promotion procedure requires that for each promotion slot, the successful candidate must be selected from those who are currently among the top three ranked candidates (including ties). This is applied sequentially until all the available promotion slots have been filled.

◇ The plaintiff was ranked number 4 and believes that this ranking should have been more than sufficient to obtain a promotion from the seven that were available. Since campaign contributions are not a requirement for promotion, the plaintiff claims that discrimination occurred.

Data

- Contribution by promotion:

	Promoted	Not Promoted
Contributed to Winner	6	0
Did Not Contribute to Winner	1	3

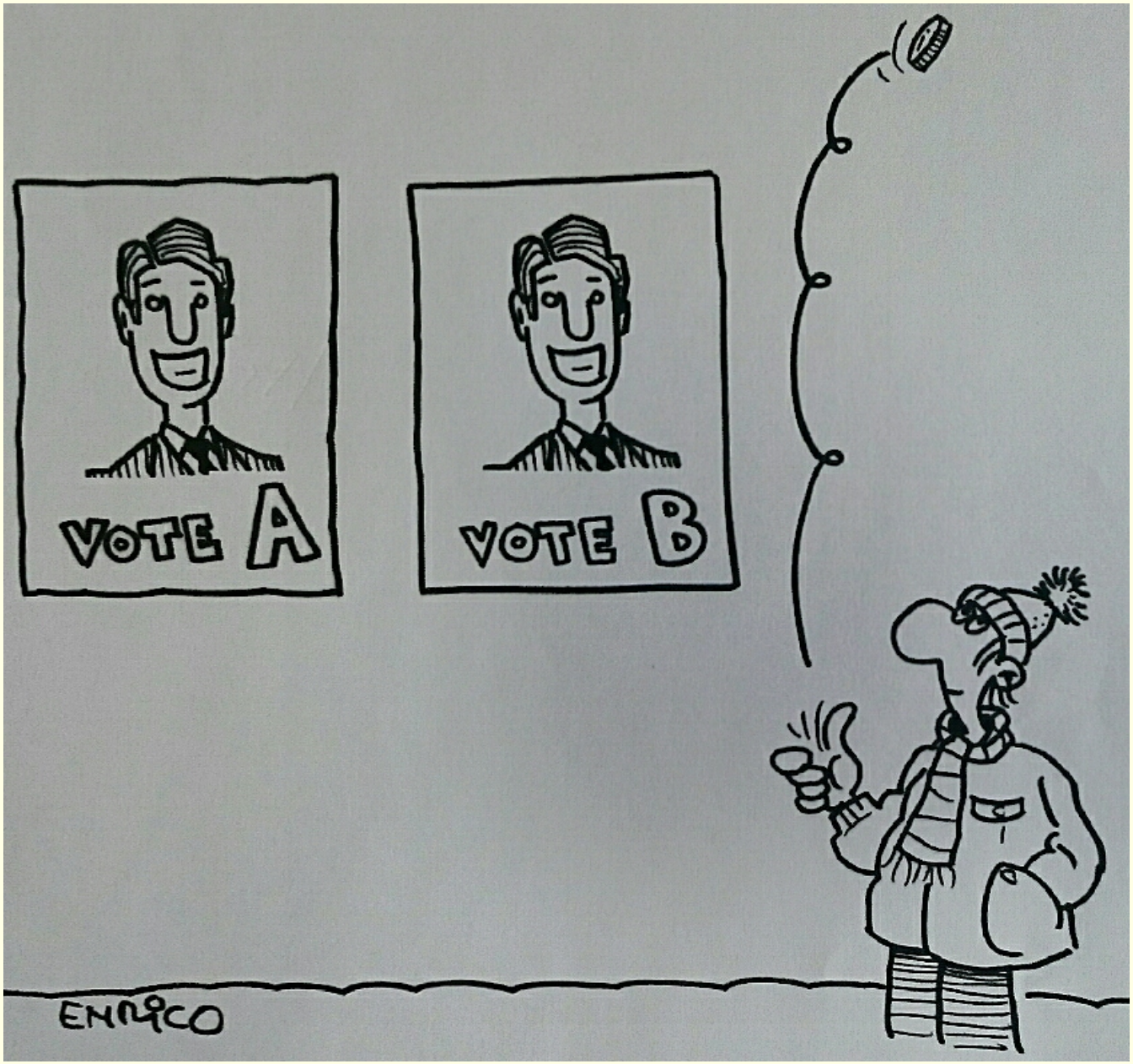
- Candidate ranking (1 = highest):

Candidate	A	B	C	D	E	F	G	H	I	J
Rank	1	2	3	4	5	6	7	8	8	8

• Additional information: employees who were not candidates for promotion were asked a question regarding whether they felt a positive or negative change in their job conditions after the election was held. They were also asked whether they made financial contributions to the campaign of the winning candidate.

$\rightsquigarrow$ Additional data ('Unknown' means that no response to that question was provided):

	Positive	Negative	Unknown
Contributed to Winner	4	0	2
Did Not Contribute to Winner	0	7	0
Unknown	1	9	2



A.6 Type of clothing

• The main focus of this study was to examine whether the interaction between student and teacher differed according to the type of clothing worn by a student.

$\rightsquigarrow$ The experiment consisted of the client observing several classes taught by one teacher and counting the number of interactions that each student had with the teacher.

$\rightsquigarrow$ An interaction was classified as either 'Negative' or 'Positive' according to the client's judgement of the teacher's response or reaction to the student.

$\rightsquigarrow$ Moreover, the interaction counts were cross-classified by the student's gender and clothing type.

$\diamond$ The client would like to know what conclusions can be drawn about differences between the interactions and whether there is any change due to gender or clothing type.

Data

- Source:

  `www.statoo.com/DATA/AC/Aclothing.xls`

- Size: 462 rows, 4 columns (*i.e.* variables).

- Variables:
  - `Clothing`: type of clothing worn by student (`Other` = 'unusual clothing', `Std` = 'standard gender-specific clothing', `Unisex` = 'unisex clothing');
  - `Gender`: gender of student (`Female` or `Male`);
  - `Count`: number of interactions observed;
  - `InteractionType`: type of interaction, from teacher's perspective (`NegInt` = 'negative interaction between student and teacher', `PosInt` = 'positive interaction between student and teacher').