Statistics

Week 1: Collecting Data (Chapter 3)

ESD, SUTD

Term 5, 2017

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Established in collaboration with MIT

Outline

Excel

If you use *Excel* on a Mac:

Download StatPlus:mac LE for free from AnalystSoft, and then use it while running *Excel*.

It should work similarly to the Analysis ToolPak.

Unmatched count

For the 1991 Race and Politics Survey, see the paper

J. H. Kuklinski and M. D. Cobb, *Racial Attitudes and the "New South"*, The Journal of Politics, **59** (1997), p323–349.

Some conclusion:

- For the non-South regions, unmatched count failed to give meaningful results (many people upset about everything). Other techniques gave racial discrimination at about 10%.
- For the South, unmatched count gave meaningful results. Racial discrimination at around 40%.

Randomized response

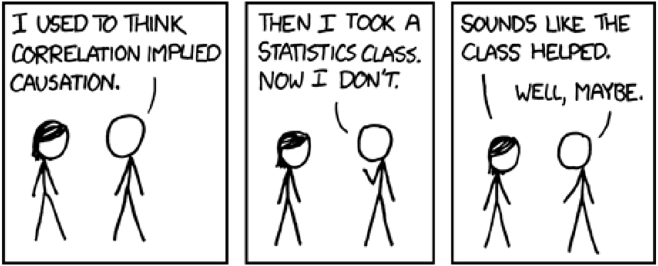
See pdf file on *eDimension*.

For this question, you can partially *check* your answer by considering some simple (boundary) cases:

- What would be the proportion of YES if the consumption rate was 0%?
- What would be the proportion of YES if the consumption rate was 100%?

Being able to check your own work is extremely important.

From http://xkcd.com/552/



Outline

Sampling

In many applications, the population is so large that sampling is the only practical option.

In some applications, destructive testing (such as crash tests) is required, so sampling is the only feasible option.

Before you sample, you need to:

- Define the population of concern.
- Specify what you want to measure.
- Specify a sampling method.
- Determine the sample size.

Questions

1. Why do we have to stick to the sample size? (Why can't we stop whenever we want?)

2. Why do we have to specify what to measure? (Multiple testing.)

Answers

1. So we can't 'cheat' by taking advantage of random fluctuations – for instance, if we wish to show that a coin is not fair by tossing it repeatedly, then stopping at a convenient point when there are significantly more H's than T's would be cheating.

2. Multiple testing – for instance, suppose we simultaneously test for 100 rare diseases, each has only a 1% chance of randomly occurring. Assuming that the diseases are independent, then detecting at least one of the diseases has probability as high as

$$1 - \left(1 - \frac{1}{100}\right)^{100} \approx 0.63.$$

Note that this probability is independent of any experiment we might wish to perform.

Sampling methods

Convenience sampling: use a sample that is readily available. (Not accurate. E. g. many psychological studies are done on psychology students.)

Snowball sampling: the first respondent refers to a friend. This friend refers to the next friend, etc.

Simple random sampling: a sample of size n is drawn from a population of size N without replacement, such that each possible sample of size n has the same chance of being chosen.

Demo: we can do this in Excel by creating a list of random numbers and sorting them.

More sampling methods

Systematic sampling: select every kth unit (useful for items coming off an assembly line).

Cluster sampling: divide the population into heterogeneous clusters (e. g. geographic areas), then draw simple random samples from each one. This method saves cost.

Stratified sampling: divide the population into homogeneous groups/strata (e. g. ethnicity, age group), then draw simple random samples. This method is more accurate.

For example, the General Household Survey 2015 used stratified sampling based on dwelling types and planning areas. Within each group, systematic sampling was used.

Exercise

How would you sample people to determine the proportion of various natural hair colours in a European country?

Hint: define the target population.

Possible design:

- Observe people from streets or dwellings (pick a range of different locations), or photos from a database. Do not use (say) Internet surveys, since self-reporting may be unreliable .
- Target the 20-40 year old age group, since people much older than this may be balding or going gray, and people much younger than this may experience darkening of hair colour.
- Target males, since they are less likely to dye their hair.

Outline

Treatment and placebo

Often, the aim of an experiment is to determine the effectiveness of a particular *treatment*.

Examples: medical treatments, diets, new ways of teaching, different work conditions, different production techniques, . . .

Problem: the *placebo* effect. A placebo is a simulated and otherwise ineffectual treatment.

Frequently, a patient given a placebo treatment (without knowing it) will have a perceived or actual improvement.

In medicine, common placebos include inert tablets (such as sugar pills) or inert injections.

Examples of placebo

- Placebos that are perceived to be more expensive tend to work better.
- When (falsely) told that a placebo has a certain smell/taste, some patients start to believe that they can smell/taste it.
- When (falsely) told that a placebo has a side-effect (e. g. numbs pain, or causes a rash), some patients actually experience the side-effect.
- Red placebo pills work better as stimulants while blue pills work better as depressants (e. g. sleeping pills).
- Even renaming a medication will temporarily make it more effective due to the novelty.

Examples of placebo

Occasionally, the placebo effect may be advantageous, for instance if a clinic runs out of pain killers.

For a real life example of the placebo effect, watch www.youtube.com/watch?v=udJ31KKXBKk from 1:50 onwards.

There are many related effects, such as the Hawthorne effect: any change in work conditions will temporarily increase productivity, due to the novelty and the perception that the workers are getting attention.

Control group

Therefore, it is important in a study to have a *control group*, which receives either no treatment, the standard treatment, or an ineffectual (placebo) treatment, whichever is most appropriate. This group provides a baseline for comparison.

The control group and treatment group should be allocated randomly.

A study is called *double blind* if both the researcher and the subject are kept unaware of which group they belong to. This removes psychological effects and is more accurate.

However, using placebos in an experiment raises ethical questions as it can be seen as a form of deception. Therefore, any research that involves human subjects must be reviewed and approved by the appropriately authority (at SUTD: the IRB).

Real life examples

- Salk polio vaccine trial: $> 200\,000$ people each in placebo and treatment groups. '6 sigma' (the meaning will be explained later).
- For many vaccines, roughly the same percentage of people in placebo and treatment groups experience side effects.
- 'Although there have been reports of an MSG-sensitive subset of the population, this has not been demonstrated in placebo-controlled trials.'

  From M. Freeman, *Reconsidering the effects of monosodium glutamate: a literature review*, J Am Acad Nurse Pract, **18** (2006), p482-486.

Exercise

How would you design a study to test whether video games improve one's reflexes?

Think about: whether the study should be observational or experimental, whether to use control groups, etc.

Answer

An observational study may establish a positive correlation between video games and reflexes, but one possible explanation for this correlation may be that people with better reflexes are better at video games, and hence play them more often. Since we are interested in whether videos games *improve* one's reflexes, such a study is not ideal.

An experimental study is preferred. Subjects should have no prior experience with video games. They should be randomly divided into a control group and a treatment group. Ideally, the two groups should be as similar as possible.

The treatment group gets to play video games, while the control group doesn't. After a predetermined period of time, the reflexes of both groups should be measured and compared.

(We can also *block* by the age group of the test subjects.)

Outline

Randomized block design

Example: suppose we want to compare three medical treatments, A, B and C, for their effectiveness.

It is known that the treatments may affect people of different ages differently.

We are only interested in A, B and C (the *treatment factor*), not in the ages (the *noise factor*).

How to design an experiment? As in stratified sampling, put people of the same age group in a ‘block’. Randomize treatment within each block.

11–20	21–30	31–40	41–50	51–60
A	B	C	A	B
C	A	B	C	C
B	C	A	B	A

Why randomize? To reduce the effects of any hidden variables not accounted for by blocking.

Latin square design

What if there are two noise factors?

Example: suppose we want to compare four types of air fresheners (labeled 1, 2, 3, 4). The tests are done in 4 different rooms and in 4 different months. Both the room and the month have an effect on air quality.

	R_1	R_2	R_3	R_4
M_1				
M_2				
M_3				
M_4				

We can fill in the grid as if it were a (simplified) sudoku puzzle. This is an example of a *Latin square* design.

Latin square answer

One possible design is:

	R_1	R_2	R_3	R_4
M_1	1	2	3	4
M_2	2	3	4	1
M_3	3	4	1	2
M_4	4	1	2	3

Each number appears exactly once in every row and column.