

## Introductory Statistics Tutorial

### Chapter 1 – What is Statistics?

In early 1997 Whitcoulls bookstores conducted a nation-wide survey. Whitcoulls' shoppers were invited to fill in a short survey. They were asked to list, in order, their three favourite books. Using the results, Whitcoulls published the list "New Zealand's 100 Favourite Books". The top twenty books from this list are given below.

No.	Book	No.	Book
1	The Lord of the Rings	11	April Fool's Day
2	The Power of One	12	Complete Winnie the Pooh
3	Pride and Prejudice	13	The Runaway
4	The Bible	14	Clan of the Cave Bear
5	Wild Swans	15	Long Walk to Freedom
6	The Horse Whisperer	16	Sleepers
7	Cross Stitch	17	Jane Eyre
8	Goosebumps	18	Gone With the Wind
9	The Bone People	19	Wuthering Heights
10	The Hobbit	20	The English Patient

1. Based on the information stated above, the two most obvious types of non-sampling errors that may be present in this survey are:
  - (1) self-selection bias and interviewer effects.
  - (2) random errors and non-response bias.
  - (3) self-selection bias and question effects.
  - (4) selection bias and self-selection bias.
  - (5) selection bias and chance errors.
  
2. A student wants to select and read any 4 books from this top twenty books list. Choose a simple random sample of 4 books for this student. To select the sample you must use the thirty random digits given below. Start at the beginning of the line of random digits given below.

87105    75663    05103    47781    00910    21112

The four books in the random sample are:

- (1) Wild Swans, April Fool's Day, The Hobbit, Goosebumps.
- (2) The Hobbit, Wild Swans, The Hobbit, The Power of One.
- (3) The Hobbit, Wild Swans, The Power of One, April Fool's Day.
- (4) Goosebumps, Cross Stitch, The Hobbit, Wild Swans.
- (5) Goosebumps, Cross Stitch, The Lord of the Rings, Wild Swans.

3. Consider the following three studies:

**Study 1:** An animal researcher was interested in cats' abilities to survive surprisingly high falls if they had time to twist round and prepare for the impact. Vets in New York City recorded incidents of cats falling out of apartment windows. The data was divided into three groups: cats that fell from one or two storeys above the ground; cats that fell from three to five storeys above the ground and cats that fell from six or more storeys above the ground. The proportion of cats that survived in each group was then compared.

**Study 2:** A random sample of 100 students is asked to keep a diary in which they record their clothing expenditures for the next three months. The expenditures of males and females are then compared.

**Study 3:** A sample of 50 shoppers at an appliance store is split into two groups. One group is shown a television commercial for a new range of appliances that has been filmed in the same style as previous television commercials for the store. The second group is shown a television commercial for the same new range of appliances that has been filmed in a totally new style. An hour after viewing the commercial, each of the shoppers was asked what they could recall about the new range of appliances and a score based on their recollection was recorded. The recall scores were then compared for the two groups.

- (i) For each study, describe what "treatment" is being compared and what "response" is being measured to compare the treatments.

**Study 1:**

**Study 2:**

**Study 3:**

- (ii) Which of the three studies would be described as experiments and which would be described as observational studies?

**Study 1:**

**Study 2:**

**Study 3:**

- (iii) For the studies that are observational, briefly explain why an experiment could not be carried out instead.



4. In 1950 two hundred employees from the Christchurch Firestone Tire and Rubber Company became part of a cancer study. These employees were observed until 1996 and any occurrences of cancer within this group were recorded. This study is **best** called:

- (1) a double-blind experiment.
- (2) a randomised experiment.
- (3) a sample survey.
- (4) a retrospective observational study.
- (5) a prospective observational study.

5. Which **one** of the following statements is **false**?

- (1) Non-sampling errors are often bigger than the random sampling errors in surveys.
- (2) People will sometimes answer a question differently for different interviewers.
- (3) Sophisticated sampling projections can always correct the results if the population you are sampling from is different to the one of interest.
- (4) Slight changes in the wording of questions can often make a big change to survey results.
- (5) Non-response can cause bias in surveys because non-respondents can behave differently from people who respond.

6. A TIME daily poll on the Internet invited readers to make a choice from a given list of options, in response to the following question:

*“Three times in the last five months, children went on killing sprees. What is fuelling this bizarre and tragic trend?”*

As of 2 June 1998, the largest proportion of respondents (29%) chose the option:

*“Nurture: The American family is crumbling; permissive parents are raising wild children.”*

We wish to use this percentage as an estimate of the proportion of all Americans who believe that *Nurture* is the cause.

Which **one** of the following is **not** a potential source of non-sampling error in this survey?

- (1) Question effects.
- (2) Self-selection bias.
- (3) Selection bias.
- (4) Non-response bias.
- (5) Transferring findings.

7. Two drugs are to be compared. A group of 20 people are each randomly allocated one of the two drugs. Neither the people who were treated nor the doctor who administered the drugs knew who got the drug. Which best describes this situation?

- (1) An observational study.
- (2) A double blind experiment.
- (3) A sample survey.
- (4) A case-control study.
- (5) A block design.

