

# STA10003 Reading A – Bias and Critical Thinking

Statistics are part of our daily lives. We come across them in the media, when talking to friends, and even in statistics classes! Many companies use statistics to increase sales of their products, and particularly when it has been suggested that a 'study' was undertaken, the statistics can sound credible. However, all might not be as first thought ... the results might in fact be biased.

Thiên  
lêch

Bias is often defined as a result of a systematic error in the study. If bias is introduced, the results may not give a true indication or picture of the outcome. Consider the following poll results:

"Coke tastes better than Pepsi ... 76% prefer Coke over Pepsi" [online poll / unknown votes]

"Coke tastes better than Pepsi ... 65.05% prefer Coke over Pepsi" [online poll – 18925 votes]

"Coke tastes better than Pepsi ... 65.3% prefer Coke over Pepsi" [online poll – 24476 votes]

While it might be tempting to conclude that Coke does in fact taste better than Pepsi, and we might opt for the 'around 65% of people prefer Coke rather than Pepsi' ... these studies may be biased.

There are a multitude of ways bias can be introduced into a study, and two of the main forms relate to Sampling bias and Information bias.

## Sampling Bias: Thiên lệch mẫu

Sampling bias refers to how the sample was 'selected' for the study. For example, if we were to look at the Coke vs Pepsi information above, we can see that all of the studies involved an online poll. An online poll introduces bias in that it is assumed that all people can access the internet, and that they are interested in having their say. We would therefore need to question if 'people' who access the poll are in fact truly representative of 'all people'. If the sample is not a true representation of the population of interest [we might want to question what 'people' actually means – adults? Soft drink consumers?], then our study findings will only provide us with information about the sample taken. The aim of using a sample is so that we can infer [generalise] our study findings to the population.

To avoid sampling bias, a sample should be randomly selected from the population about which we wish to draw conclusions.

## Information Bias:

Information bias refers to how the information was collected and measured.

Firstly, did anyone with a vested interest conduct or sponsor the study? While Joe Bloggs [not his / her real name] was interested enough in the topic to conduct the online poll, we might want to question if he / she has any affiliation with either of the products. Perhaps Joe Bloggs is a staff member in the Research and Development department of one of the companies. Was Joe Bloggs approached, and subsequently paid, by one of the companies to conduct the poll?

A second consideration relating to information bias is how the data collected was measured. What / how questions were the questions actually asked? The polls shown above all contained only two response options: Prefer Coke / Prefer Pepsi. There was not an option, for example, for respondents to say that they prefer another brand, or that they don't drink soft drinks. This 'forces' the

# STA10003 Supplement A – Bias and Critical Thinking

respondent to make a choice – other than opting out of the poll entirely which leads us to question if the sample is truly representative of the population of interest.

When asking questions, the wording needs to be unambiguous. It should *not* contain leading or emotional language which encourages a particular response. Consider the following:

“Should concerned dog owners vaccinate their pets?”

The use of the word ‘concerned’ implies that people who may not vaccinate their pets don’t really care about them. Removing the ‘concerned’ from the question makes it clear and concise – and is not leading the respondent to agree.

“The company has wasted a lot of money on XYZ. Should this intolerable practice be allowed to continue?”

Again we have problems with the wording. There is an emotional statement to begin with, and the question is further compounded by the use of ‘this intolerable practice’. There are a number of ways to ask this type of question in a neutral manner. For example, asking how much the company should spend on XYZ means that the respondents are free to make their own judgement.

Question wording that can have multiple answers is also a problem. For example:

‘When was the last time you upgraded your computer and printer?’


If I purchased a new printer last week, but have been using the same computer for two years, it’s difficult to know how to answer this one! – and just as difficult to make any meaningful conclusion based on the answer [which part of the question was actually answered – the printer or the computer?].

Even when questions are asked in a neutral manner, there is also the possibility of Social Desirability bias. The concept of Social Desirability occurs when people respond in a way that they think that should, rather than what they actually think or do. A respondent may answer a question based on what their friends think, so as not to appear ‘different’, or they may over- or under-report instances of certain types of behaviour [for instance we might over-report instances of charity donations and under-report instances of drug use].

## ***Identifying possible sources of bias in a study:***

Whilst this information is not exhaustive, it should provide you with some techniques so that you can ‘Critically Think’ about statistics. It provides us with tools to make a judgement based on information, rather than on the statistics alone.

Since **our purpose is to avoid bias**, we need to ask:

- 
- How were the individuals / objects in the study selected?
  - What measurements were made? / What questions were asked?
  - Who conducted / sponsored the study?

We need to be on the alert for misleading statistics ...