

Writing a Scope of Inference

My Guidelines:

1. Keep causal inference SEPARATE from inference to populations. In other words, do not refer to both within the same sentence. Initially start by using just two sentences (you can get more creative after you have nailed the two sentence approach).
 1. one sentence addressing whether causal inference is justified and why
 2. one sentence addressing whether any inference to a larger population is justified and why
2. Make sure you write the *Scope of Inference* specific to the language of the problem (*not* just generic statements). For example, DO NOT say “There was no randomization so cause and effect statements cannot be made.” Instead, for example, say “The lizards were not randomly assigned to diets, therefore we can not conclude that the fly diet caused the increase in average weight over the lizards on the tofu diet.”
 1. Say *what* was (or was not) randomly assigned to *what*, therefore...
 2. Say *what* was (or was not) randomly selected from *what*, therefore...
3. There is no need to refer to the statistical results - you can write a *Scope of Inference* based on the study design without yet knowing the results.

Some examples from *The Statistical Sleuth* (not all follow my guidelines)

- Because students were randomly assigned to type of questionnaire, one may infer the difference in creativity scores was *caused* by the difference in motivational questionnaires. Because the subjects were not selected randomly from any population, extending this inference to any students beyond this group is speculative. This deficiency, however, is minor; the causal conclusion is strong even if it applies only to the recruited subjects.
- Although there is convincing evidence that the males, as a group, received larger starting salaries than the females, this analysis alone cannot address whether this difference is attributable to sex discrimination because sex cannot be randomly assigned to employee. The evidence is consistent with discrimination, but other possible explanations cannot be ruled out; for example, the males may have had more years of previous experience. Inference cannot extend beyond the employees from this bank during this time period because there was not random selection from a larger group of employees.
- These twins were not randomly selected from general populations of schizophrenic and non-schizophrenic individuals. Tempting as it is to draw inferences to these wider populations, such inferences must be based on the assumption that these individuals are as representative as random samples are. Furthermore, the study is observational because individuals clearly cannot be randomly assigned to be schizophrenic or not, meaning that no causal connection between the left hippocampus volume and schizophrenia can be established from this study and analysis alone. In fact, the researchers had no theories about whether the abnormalities preceded the disease or resulted from it.

- We cannot conclude that longer humerus lengths among survivors enabled them to survive, because clearly humerus length was not randomly assigned to sparrows. Inference to sparrows beyond the sample is dangerous because the sparrows were not randomly selected from all sparrows in the region. In particular, the living sparrows brought to Bumpus were found in a specific area and were so stressed that they were easily collected. Inference to populations of similarly stressed sparrows is risky.