**Preliminary Exam Questions: Dr. Dinsmore**

**Marty Simonson**

**March 27, 2020**

1. **Most sampling strategies are comprised of two primary elements: an estimator and a sampling design. Together, these two elements can be used to attain a desirable balance between precision and bias. Focusing on bias, the three primary sources are statistical bias, selection bias, and measurement bias. Provide a definition and example(s) of each and then rank and justify these biases according to *your view* of the extent to which they have, or currently are, detracting from the quality of data guiding fisheries science. [2-3 pages]**

The terms accuracy and precision are too often used interchangeably when describing statistical results of experiments. Fine-scale instruments and rigorous protocols may result in precise data, but statistics and subsequent inferences about the natural world would be false if too much bias were introduced in data collection or analysis. Therefore, understanding the sources of bias and what methods are necessary to mitigate bias is a critical component of research study design. In the context of natural resource management, the three most significant sources of bias are statistical bias, selection bias, and measurement bias.

Statistical bias occurs when sample statistics do not reflect the true population parameter. There are numerous sources for statistical bias, among them are latent (unobserved) variables and overfitting of models. The most significant source of statistical bias occurs when not enough care and skill are used in performing statistical tests and experiments from which the data are derived. For example, count data of species often follows a Poisson distribution, but data of this type also have variances that increase proportionally to the mean; failure to account for that fact will introduce statistical bias to analyses that assume equal variances between samples and groups.

A biased sample is not representative of the population from which it was drawn. If we are to make reasonably valid estimates of population values based on samples, the measurements must be representative of the population of interest. We may never be certain that a sample is unbiased, because if we were certain then we would have the population-level data and there would be no need for estimation. In ecological statistics, biased samples can arise when selectivity of capture gears is not properly accounted for, or when measurements at sample sites vary in time. For instance, taking a dissolved oxygen measurement in a lake every day at noon will not represent the true dissolved oxygen concentration profile throughout a 24-hour cycle, and a study focusing on daily plant respiration and photosynthesis will not have an accurate representation of that data, despite the possibility of being very precise.

Measurement bias is a result of faulty or uncalibrated measuring instruments and human error. Even the most sophisticated aquatic measurement probe cannot supply exact solute concentrations, and if it could the measurements are still subject to error in use and interpretation, both of which result in distorted data. For an example of the latter, I once had a technician in Toledo that rarely recorded fish lengths that ended in 5 or 0; they must have thought recording those lengths would appear like they were rounding the data too often when in fact the opposite was true. The implications of measurement bias can be problematic, however, if the objective of a study is to make *relative* comparisons, assumptions of equal variance still meets the criteria for unbiased estimation.

The type of bias with the largest detraction from the quality of data in fish science would be selection bias. Generalizations based on biased samples lead to false conclusions; if sampling design is biased then an unlimited number of samples may be taken that do not reflect the true population parameter. Further, fisheries sampling gear is highly size- and species-selective, and a substantial amount of background knowledge is required to adequately construct a robust sampling design. Agency-based work may sometimes be limited by the 8:00 AM to 5:00 PM work schedule, when fish have different activity pattens and distributions than they do at dawn and dusk. Finally, the greatest source of selection bias come from human judgement. The subjective nature of study design and model selection is a high-risk source of introducing bias to analyses.

The type of bias with the intermediate level of detraction from quality fish data is measurement bias. Day-to-day changes in lakes and lack of training and experience of technicians are the other significant contributions to measurement bias. One of the largest issues I’ve had with my data is compiling my database (in metric) with the Iowa DNR data, which is recorded in Imperial units. Further, sometimes on the same data sheet a fish will be weighed in decimal pounds, while another is weighed in decimal ounces; catching these errors and correcting/censoring data has been a frustrating experience, to say the least.

Statistical bias, then, is the least significant source of unreliable and inaccurate data in my opinion. Methods have been developed to account for latent environmental variables, and tests for autocorrelation of errors provide insight into the equal variance assumption of the most widely applied statistics. Finally, sensitivity analyses can be applied to test if changing one predictor variable at a time has significant influence on the study outcome. Therefore, thoughtful statistical design can prevent bias from becoming a problem in the analysis phase. Through my research experience I’ve been exposed to each of the sources of bias, and drawing on that experience I am confident in saying that bias of fisheries data is most likely introduced through subjective human decisions about sample design and model selection; measurement and statistical bias can be accounted for and corrected but poorly obtained data lasts forever.