
Helsel’s Book: A Clever Thing About Supposedly Nothing

We have all seen them in textbooks, papers, or our own research datasets—the dreaded “less than” data, those environmental (most often chemical analysis) data that are “less than” some detection or reporting limit or some other small limit. In the interpretation and particularly in the statistical analysis of environmental data, the “less than” data or “nondetects” are but a nuisance to the practitioner: our beloved software programs apparently have no use (at least that most of us know of) for “less than” data. We ignore them or simply zero them out. They are nothing, rien, NADA. My statistics software, for example, automatically excludes all spreadsheet cells with character-type content (such as “<5.0” or “less than 5.0”) when I run the summary statistics on a list of reported values. I confess that on more than one occasion I have converted those “nondetect” data to zeros in my data tables to at least acknowledge that a measurement was taken, yet avoid the hassles of nonnumeric entries. The logic typically goes something like “they are so small that they are—for practical considerations—like zero.” The almost equivalent, but more “scientific” approach recommended by citable sources, such as USEPA and the Army Corps of Engineers, is to replace the “less than” data with a value of one-half of the reporting limit before proceeding with standard statistical analyses of the dataset.

Were you ever bothered by such a simplification? Then read on. In his introduction, Dennis Helsel makes a convincing and illustrative, yet simple, bid for taking a different approach: the tragedy of NASA’s space shuttle Challenger may well have been avoided if engineers had more carefully considered their “nondetects”. Helsel argues that in the case of the Challenger the consequences of ignoring nondetects were “a tragic one-time loss of life. For environmental sciences, the consequences are likely more chronic and continuous.” The motivation for this book was to provide a college textbook or practitioner’s resource that has the potential to jump-start the analysis of nothing (“NADA” being the initials of the rather clever book title) into something that should be practiced much more widely.

Statisticians call datasets with data that are less than (or larger than) a specified numeric value a “censored” dataset. These types of datasets not only arise from analyzing environmental concentrations, but are common in many other disciplines. It is therefore not surprising that statisticians have found ways to properly work with censored data. The only surprising thing is that they are not commonly used in the environmental field. This is even more surprising in light of the fact that statistical software installed on many of our computers is actually capable of handling censored data. The book introduces, in a very intuitive manner, the three basic approaches to dealing with censored data (substitution, maximum likelihood estimation [MLE], and nonparametric methods). The next chapter, a very enlightened and well-researched discussion follows on various definitions of reporting limits typically used in analytical laboratories. The author explains the differences between detection limit and quantitation limit, and other terms used in the context of reporting limits. The book also contains numerous references to papers with example applications, reviews, and method comparisons. An overview is provided of how to deal with censored data with standard statistical software. I found the material immensely practical, as short and simple as it is.

The remainder of the book is a “how-to” guide to basic statistical procedures with censored data, such as summary statistics, computing interval estimates, and graphical representations. The book even includes a chapter when all data are below the reporting limit. All of the “how-to” chapters describe the methods used under each of the three basic approaches (substitution, MLE, nonparametric methods), together with a discussion on when the approaches are most appropriate under what circumstances.

While the book does not delve much into the background or derivation of equations, it does contain material for readers familiar with the above techniques. It provides essential theoretical background needed to understand and compute results accurately. The author makes it easy to compare each method with procedures and examples described in the book.

Each chapter starts out with a motivational example and an overview of the chapter. All methods are carried out with example datasets, which are included in an appendix and can be obtained from a companion website. Computational procedures and results are shown using specific statistical software program (Minitab, State College, PA), but can easily be compared with those of other statistical software. The book also contains numerous example exercises and computer model exercises needed to understand and compute relevant statistics.

The author hence makes it easy to compare each method with procedures and examples described in the book. Each chapter starts out with a motivational example and an overview of the chapter. All methods are carried out with example datasets, which are included in an appendix and can be obtained from a companion website. Computational procedures and results are shown using specific statistical software program (Minitab, State College, PA), but can easily be compared with those of other statistical software. The book also contains numerous example exercises and computer model exercises needed to understand and compute relevant statistics.

In statistics, analysis of censored data is referred to as “reliability analysis,” “failure time analysis,” or “failure time analysis for grouped data.” The reason for this is readily explained in the Helsel’s book. If your statistics software handles these types of analyses, Helsel’s book is a great resource to get you started on applying them to environmental data.

Before dealing with statistical methods, the book introduces four chapters that discuss fundamental and common methods of dealing with censored datasets. The first chapter introduces, in a very intuitive manner, the three basic approaches to dealing with censored data (substitution, maximum likelihood estimation [MLE], and nonparametric methods). The author explains various definitions of reporting limits typically used in analytical laboratories. The author explains the differences between detection limit and quantitation limit, and other terms used in the context of reporting limits. The book also contains numerous references to papers with example applications, reviews, and method comparisons. An overview is provided of how to deal with censored data with standard statistical software. I found the material immensely practical, as short and simple as it is.

The remainder of the book is a “how-to” guide to basic statistical procedures with censored data, such as summary statistics, computing interval estimates, and graphical representations. The book even includes a chapter when all data are below the reporting limit. All of the “how-to” chapters describe the methods used under each of the three basic approaches (substitution, MLE, nonparametric methods), together with a discussion on when the approaches are most appropriate under what circumstances.

While the book does not delve much into the background or derivation of equations, it does contain material for readers familiar with the above techniques. It provides essential theoretical background needed to understand and compute results accurately. The author makes it easy to compare each method with procedures and examples described in the book.

Each chapter starts out with a motivational example and an overview of the chapter. All methods are carried out with example datasets, which are included in an appendix and can be obtained from a companion website. Computational procedures and results are shown using specific statistical software program (Minitab, State College, PA), but can easily be compared with those of other statistical software. The book also contains numerous example exercises and computer model exercises needed to understand and compute relevant statistics.

The author hence makes it easy to compare each method with procedures and examples described in the book. Each chapter starts out with a motivational example and an overview of the chapter. All methods are carried out with example datasets, which are included in an appendix and can be obtained from a companion website. Computational procedures and results are shown using specific statistical software program (Minitab, State College, PA), but can easily be compared with those of other statistical software. The book also contains numerous example exercises and computer model exercises needed to understand and compute relevant statistics.