An Introduction to Statistics for Librarians (Part One): Types of Data

Caitlin J. Bakker, MLIS, AHIP[a]

[a]Discovery Technologies Librarian, Dr. John Archer Library, University of Regina, Regina, Saskatchewan, Canada, https://orcid.org/0000-0003-4154-8382, caitlin.bakker@uregina.ca

Librarians usually aren't statisticians. Most of us haven't taken a statistics course and might not feel comfortable talking about things like p-values and confidence intervals. At the same time, most of us also want or need to know that results are "real." Whether we're assessing our own programs and services, conducting research, or reading scientific articles, it's helpful to understand some of the basics of statistics. This series in "The Research Mentor" column will guide you through the process of identifying the data type, choosing a statistical test, and interpreting the results, but it is not meant to be comprehensive or conclusive. Whenever possible, it's best to find a trained statistician who you can consult with about your specific project.

Identifying the Data Type

The first step to deciding what type of statistical test is appropriate is to think about what type of outcomes you are reviewing. There are two broad categories of data: categorical and continuous. Within these categories, there are four types of data: categorical data can be classified as (1) nominal or (2) ordinal, while continuous data can be classified as (3) interval or (4) ratio.

This table introduces the four different data types, which we'll go over in more depth below.

Data Type	Description	Examples
Nominal	Categorical data that can't be ordered in any meaningful way. These are data without any quantitative value.	Eye color, hair color, type of car, ice cream flavor, nationality
Ordinal	Categorical data that can be ranked or ordered meaningfully.	Likert scales (most to least likely, strongly disagree to strongly agree), course grades (letter grades A, B, C, D, F), class rank
Interval	Continuous data that don't have a natural zero.	Temperature in Fahrenheit or Celsius; MCAT, GRE, or SAT scores
Ratio	Continuous data that have a natural zero.	Temperature in Kelvin, height, weight, age, scores on a test (percentage correct)

Categorical Data Types: Nominal and Ordinal

Categorical data are also called discrete and refer to outcomes where there are a fixed number of possible values. Demographic data, such as gender, age group, or educational level are categorical data. There are two specific types of categorical data: (1) nominal and (2) ordinal. **Nominal** data have no order, while ordinal data are ordered. For example, if you surveyed a group of medical students and asked them which specialty they were planning on pursuing, their answers would be nominal data because you couldn't rank family medicine above surgery or below obstetrics. The answers are all different, but there is no rank or order. If you surveyed those same students and asked them to say whether they used certain library resources a lot, a little, rarely, or not at all, those answers would be **ordinal** because we could order them from most to least used.

Continuous Data Types: Interval and Ratio

Continuous data are numerical and can have any value between a minimum and maximum. Questions correct on an exam, or the final percentage of correct answers would be continuous, since a student could get anywhere from 0% to 100%. Like ordinal data, continuous data are ordered. They are also evenly spaced. The difference between 5% and 10% is the same as the difference between 10% and 15%. This is not true of ordinal data. The difference between "strongly disagree" and "disagree" on a Likert scale may not be the same as the difference between "disagree" and "neutral."

There are two types of continuous data: (1) interval and (2) ratio. The difference between these two lies in what zero means for each. With **ratio** data, zero would mean an absence of that variable. This is called a "natural zero," where zero means, "there isn't any of this." The example above of student scores on an exam would be ratio data. If a student scored 0%, it would mean that they got no questions right on the exam. In contrast, those same students' MCAT scores would be interval data since the possible scores for the MCAT range from 472 to 528. Absolute zero isn't used as a reference point since getting no questions right wouldn't result in a score of 0. Therefore, zero has a different meaning in **interval** data than it does in ratio data. This means that while both ratio and interval data can be added and subtracted, only ratio data can be divided and multiplied.

Why Does This Matter?

Knowing the type of data is essential to choosing the right statistical test. Trying to conduct a test designed for interval or ratio data on ordinal data would lead to incorrect results and possibly false conclusions. In the next column in this series, we'll discuss which statistical tests may be right for which type of data.

Suggested Readings

Bowers D. First things first – the nature of data. In: Medical statistics from scratch: An introduction for health professionals. 4th ed. Hoboken (NJ): Wiley; 2019. p. 3–14.

Daniel WW, Cross CL. Measurement and measurement scales. In: Biostatistics: A foundation for analysis in the health sciences. 10th ed. Hoboken (NJ): Wiley; 2019.

Greenhalgh T. How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests. BMJ. 1997 Aug 9;315(7104):364–6. doi: 10.1136/bmj.315.7104.364.

Stewart A. Types of data. In: Basic statistics and epidemiology: A practical guide. 4th ed. London (England): CRC Press; 2016. p. 17–9.