

Types of Data and the Appropriate Tests for them

- Data can be broadly separated into *categorical* and *numerical* data.
- Categorical data can be further separated into *ordinal* and *nominal* data (see Figure 1).
- The appropriate statistical test depends on which combination of data types are tested.

by ALEXANDER THORPE, QUENTIN F. GRONAU, & GARSTON LIANG

1 Data Types

When designing a study, it is important to consider the type of data you will collect, as this will also influence how you analyse your data. Below, we will discuss the different types of data. These definitions are true for both independent and dependent variables, and it is the combination of data types that determine which statistical tests are appropriate. Data can be broadly separated into two categories—*numerical*, where a data point can be represented as some kind of number value with mathematical properties, and *categorical*, which is any other kind of data. We will also present examples of each data type that are relevant to neuromodulation literature.

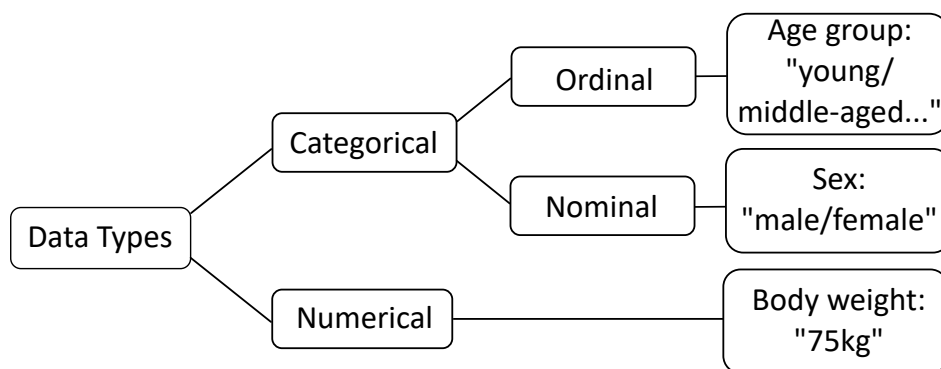


Figure 1: Different types of data with examples.

2 Numerical Data

When working with quantitative research methods, numerical data is the most common data type. It is also very common to work with numerical dependent variables, as we tend to measure outcomes with numerical scales. Numerical data is often further separated into two sub-categories—*discrete* or *interval*, where data points can only take certain values, and *continuous*, where data can take any numerical value. Another sub-category, *ratio*, refers to data where zero is the lowest possible value. For example, temperature is not expressed in ratio form, because values below zero can exist, but body weight is expressed in ratio form, because it is impossible to weigh less than 0kg. Despite these different subcategories and labels, all numerical data can be analysed using the same statistical methods, so there is no need to consider them when choosing an appropriate test.

3 Categorical Data

Categorical data, also known as *qualitative* data, refers to any data type that can be split into meaningful groups, but is not necessarily represented with numerical values. They are still present in quantitative research—grouping variables, for example, can be categorical. Categorical data can be separated into *ordinal* data, where the categories have an inherent order, and *nominal* data, where the categories have no set order.

3.1 Ordinal Data

Ordinal data refers to data whose categories can be arranged in some kind of order. An example of this would be age groupings. Participants in a study of chronic pain could be categorised as *young*, *adolescent*, *young adult*, *middle-aged*, and *elderly*. These groups could be ordered such that participants in one group are older than the preceding group, and younger than the succeeding group. This may be meaningful for a study's aims—if the study is examining the relationship between age group and incidence of chronic pain, the researchers may expect incidence of chronic pain to increase from one group to the next.

3.2 Nominal Data

Nominal data cannot be grouped in the same way as ordinal data. Classifiers such as sex or experimental group membership are nominal, as there is no inherent order to male/female or experimental/control. These groups can be compared, but no relationship between the group membership and any dependent variable could be inferred just by the ordering of the groups. A special case of nominal data is *binary* data, where data points belong to one of two mutually exclusive groups, and can be represented only as 0 or 1. This data is not numerical, as the numbers are used only to signify some property—one could just as easily use “true” and “false” or “yes” and “no”. For example, data indicating whether or not a participant meaningfully improved after treatment could be represented by binary values (1=improved, 0=did not improve). This data type can be analysed in ways other nominal data cannot.

4 Appropriate Tests

How do you know which statistical test to use? It all depends on the combination of data types of your *independent variables*, or the variables being manipulated, and *dependent variables*, or variables that measure an outcome. Figure 2 provides a brief guide to which test is appropriate based on data types. For more detailed guides, see the “Further Reading” section below.

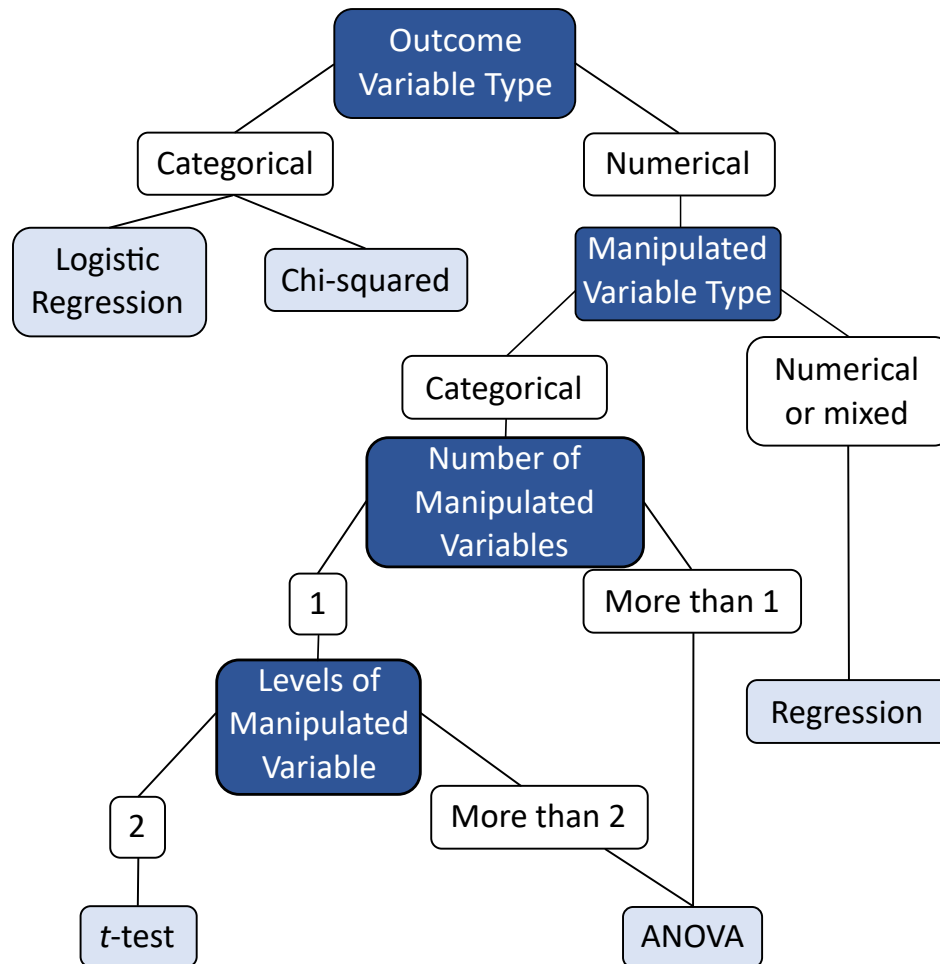


Figure 2: Appropriate tests for different combinations of variables and variable types.

5 Further Reading

Below are some open-source resources that go into more detail about the different data types, and how to choose appropriate tests for them.

- Swinscow, T. D. V. (1997). *Statistics at square one*. BMJ Publishing Group.
- Kebalepile, M. M., & Motshabi Chakane, P. (2022). Commonly used statistical tests and their application. *South African Journal of Anaesthesia and Analgesia*, 28(5), S80-84.
- Nayak, B. N., & Hazra, A. (2011). How to choose the right statistical test? *Indian Journal of Ophthalmology*, 59(2), 85-86.