# DATA SELECTION MINI LECTURE

Contents

[DATA SELECTION MINI LECTURE 1](#_Toc73970095)

[THE TWO MAIN FLAVORS OF DATA: QUALITATIVE AND QUANTITATIVE1 1](#_Toc73970096)

[QUALITATIVE FLAVORS: BINOMIAL DATA, NOMINAL DATA, AND ORDINAL DATA1 1](#_Toc73970097)

[QUANTITATIVE FLAVORS2 2](#_Toc73970098)

[SOURCES: 3](#_Toc73970099)

Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things. However not all data is created the same way, so it is essential to understand what type of data you have and the kinds of questions your data can help you answer.

## THE TWO MAIN FLAVORS OF DATA: QUALITATIVE AND QUANTITATIVE1

At the highest level, two kinds of data exist: ***quantitative*** and ***qualitative***.

1. ***Qualitative*** data deals with characteristics and descriptors that can't be easily measured, but can be observed subjectively—such as smells, tastes, textures, attractiveness, and color.
2. ***Quantitative*** data deals with numbers and things you can measure objectively: dimensions such as height, width, and length. Temperature and humidity. Prices. Area and volume.

 Broadly speaking, when you measure something and give it a number value, you create quantitative data. When you classify or judge something, you create qualitative data. So far, so good. But this is just the highest level of data: there are also different types of quantitative and qualitative data.

## QUALITATIVE FLAVORS: BINOMIAL DATA, NOMINAL DATA, AND ORDINAL DATA1

1. ***Binomial*** data place things in one of two mutually exclusive categories: right/wrong, true/false, or accept/reject. Occasionally, you will get a box of jelly beans that contains a couple of individual pieces that are either too hard or too dry. If I went through the box and classified each piece as "Good" or "Bad," that would be binary data. I could use this kind of data to develop a statistical model to predict how frequently I can expect to get a bad jelly bean.
2. ***Nominal*** data: we assign individual items to named categories that do not have an implicit or natural value or rank. If I went through a box of jelly beans and recorded the color of each in my worksheet, that would be nominal data.
3. **O*rdinal*** data: items are assigned to categories that do have some kind of implicit or natural order, such as "Short, Medium, or Tall."  Another example is a survey question that asks us to rate an item on a 1 to 10 scale, with 10 being the best. This implies that 10 is better than 9, which is better than 8, and so on. 


## QUANTITATIVE FLAVORS2

There are two main types of quantitative data, which is also referred to as numeric data: ***interval and ratio****.*

1. ***Interval:*** there is order and the difference between two values is meaningful. **Interval data can have negative numbers.**

*Examples: temperature (Fahrenheit or Celcius), pH, SAT score (200-800), credit score (300-850).*

1. **Ratio*:*** it has all the properties of an interval variable, and also has a clear definition of 0.0. When the variable equals 0.0, there is none of that variable. **In other words, there are no negative numbers for ratios.**

*Examples: enzyme activity, dose amount, reaction rate, flow rate, concentration, pulse, weight, length, temperature in Kelvin (0.0 Kelvin really does mean “no heat”), survival time.*

**Interval** and **Ratio** variables can be further split into two types: **discrete** and **continuous**. As a general rule, *counts* are discrete and *measurements* are continuous.**1**

1. ***Discrete data*** is a count and can only take whole numbers. For instance, the number of children (or adults, or pets) in your family is discrete data, because you are counting whole, indivisible entities: you can't have 2.5 kids, or 1.3 pets.
2. ***Continuous* *data***, on the other hand, could be divided and reduced to finer and finer levels. For example, you can measure the height of your kids at progressively more precise scales—meters, centimeters, millimeters, and beyond—so height is continuous data.



***NOTE: We will only use discrete datasets for our activity, because other types of datasets could be averages and it would be problematic trying to use them in a 5-point summary.***

# SOURCES:

1. Understanding Qualitative, Quantitative, Attribute, Discrete, and Continuous Data Types

<https://blog.minitab.com/en/understanding-statistics/understanding-qualitative-quantitative-attribute-discrete-and-continuous-data-types>

1. KNOWLEDGEBASE - ARTICLE #1089 - What is the difference between ordinal, interval and ratio variables? Why should I care?

<https://www.graphpad.com/support/faq/what-is-the-difference-between-ordinal-interval-and-ratio-variables-why-should-i-care/>

This material is based upon work supported by the National Science Foundation under Grant No.1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.