

CHAPTER OVERVIEW

3: Exploring Data

Introduction

After an experiment has been completed, the collection of observations needs to be summarized or described, a process now referred to as **data exploration**. How many mice grew tumors in the treated versus control group? Did all mice respond to treatment? What is the typical cost of a new home in [Hilo, HI](#)? How large of a difference is there between the most expensive home and the middle? These questions require simple, basic summary statistics or **descriptive statistics** and informative **data visualizations**.

In this section we introduce two aspects of **summary statistics** expected in any report of a **data set**. Summary statistics provide a brief overview of relevant characteristics of observations in the data set to provide the reader to get a quick but meaningful look at the data set. Data set refers to a collection of data, usually a collection of related, ordered observations, measurements, and related information.

Summary statistics vary according to the needs of the reporting vehicle, but may generally address two characteristics of the data set

1. Central tendency or the description of the middle of the data.
2. Dispersion, the variability around the middle of the data set.

We introduce statistical graphics, a specialized kind of data visualization, in [Chapter 4 – How to Report Statistics](#). Statistical graphs are utilized to describe data sets, but also to communicate statistical inference, which we address formally beginning in [Chapter 7 – Probability and Risk Analysis](#).

Most of you have been asked at some point to calculate the average or the standard deviation. We will provide these again, but with additional statistics. Textbooks may present **calculator formulas** — nothing wrong with them, although we have to watch **significant figures**. But computer statistical packages generally do not use these formulations — that's why I present the formulas throughout the book to help define the statistical concept, not as a way to necessarily calculate the statistic by hand calculation. I haven't investigated this last point in any systematic way, but, because of access by scientists to increasingly powerful computers since the 1980s, I doubt anyone in the business of data analysis in the biological sciences has much use of the hand calculator.

Note to BI311 students:

On homework, quizzes and exams, you may be asked to calculate these descriptive statistics. I will provide you with formulas that illuminate the definitions of the statistics rather than enhance their computation.

[3.1: Data types](#)

[3.2: Measures of central tendency](#)

[3.3: Measures of dispersion](#)

[3.4: Estimating parameters](#)

[3.5: Statistics of error](#)

[3.6: Chapter 3 References and Suggested Reading](#)

This page titled [3: Exploring Data](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.