

Chapter 2 Notes - Statistics

Matt Warner

0.1 Types of Data

Terminology

- If we measure or record
 - one observation from each individual or object we have univariate data
 - two observations from each individual or object we have bivariate data
 - more than two observations from each individual or object we have multivariate data (e.g. height, weight, gender, blood pressure, and cholesterol level of each person)
- Types of Data
 - Categorical or qualitative - non-numerical observations that may be placed in categories.
 - Numerical or quantitative - observations that are numbers.
- A numerical data set is
 - Discrete if its set of possible values is a finite set or a countable infinite set (i.e. an infinite sequence with a first value, second value, etc.)
 - Continuous if its set of possible values is an infinite set that forms an interval on the number line (usually associated with measuring)

Discrete vs Continuous

Continuous Data:

It can take on any value in an interval

Data that can be measured (i.e. speed of a car)

Discrete Data:

It can only have specific values.

Data that can be counted (i.e number of books)

0.2 Bar Charts and Pie Charts

Frequency distribution

Definition

The natural summary measures for a categorical data set are the number of times each category occurred and the proportion of times each category occurred. These values are usually displayed in a table as in Table 2.1

Class	Frequency	Relative frequency
Mammals	202	0.101
Birds	222	0.111
Reptiles	266	0.133
Amphibians	552	0.276
Fishes	468	0.234
Insect	290	0.145
Total	2000	1.000

Note:-

The shape of a distribution may be symmetric or skewed. The center of a distribution refers to the position of the majority of the data, and measures of variability indicate the spread of the data. The *variability* (or dispersion) of a distribution describes how much the measurements vary, as well as how compact or how spread out the data are.