First part of the class is dedicated to reviewing the Syllabus.

Syllabus Notes are located in Drive

Chapter 1

Overview and Descriptive Statistics

Section 1: Populations, Samples, and Processes

Terms:

**Population** includes all objects of interest.

For example: All individuals who received a B.S. in engineering during the most recent academic year.

**Sample** includes a subset of the population which is selected in some prescribed manner.

For example: A sample of all individuals who received a B.S. in engineering during the most recent academic year to obtain feedback about the quality of the engineering curricula.

**A variable** is any characteristic whose value may change from one object to another in the population.

For example: The gender of an engineering graduate (categorical), the age at which the individual graduated (numerical).

Types of data:

Data result from making observations either on a single variable or simultaneously on two or more variables.

(a) **Univariate data** set consists of observations on a single variable.

For example: The gender for 10 students can be M M F M F F F M M F.

(b) **Bivariate data** when observations are made on each of two variables.

For example: A (weight(kg), height(cm)) pair for each basketball player on a team, with the first observation as (72, 168), the second as (75, 212), and so on.

(c) In many **multivariate data** sets, some variables are numerical and others are categorical.

For example: The annual automobile issue of Consumer Reports gives values of variables as type of vehicle (small, sporty, compact, mid-size, large), city fuel efficiency (mpg), highway fuel efficiency (mpg), drivetrain type (rear wheel, front wheel, four wheel), and so on.

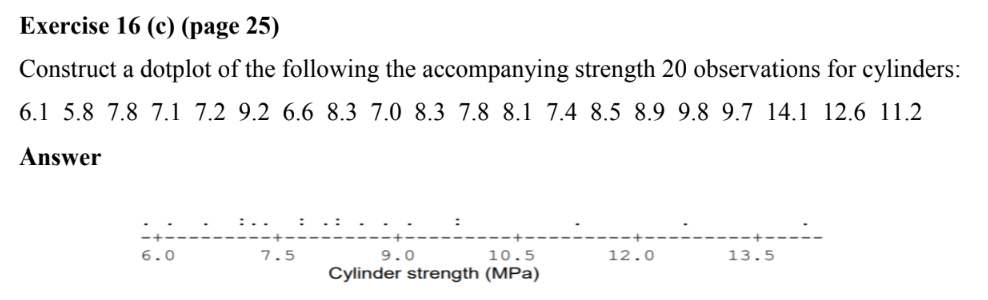
The discipline of statistics provides methods for organizing and summarizing data and for drawing conclusions based on information contained in the data.

Section 2: Pictorial and Tabular Methods in Descriptive Statistics

Given a data set consisting of n observations on some variable x, the individual observations will be denoted by x1, x2, ... , xn where x1 will be the first observation x2, the second, and so on.

**Dotplots**

When there are relatively few distinct data values. Each observation is represented by a dot above the corresponding location on a horizontal measurement scale. When a value occurs more than once, there is a dot for each occurrence, and these dots are stacked vertically.



**Stem-and-Leaf Displays**

Consider a numerical data set x1, x2, ... , xn for which each xi consists of at least two digits. Constructing a Stem-and-Leaf Display

In general, a display based on between 5 and 20 stems is recommended.

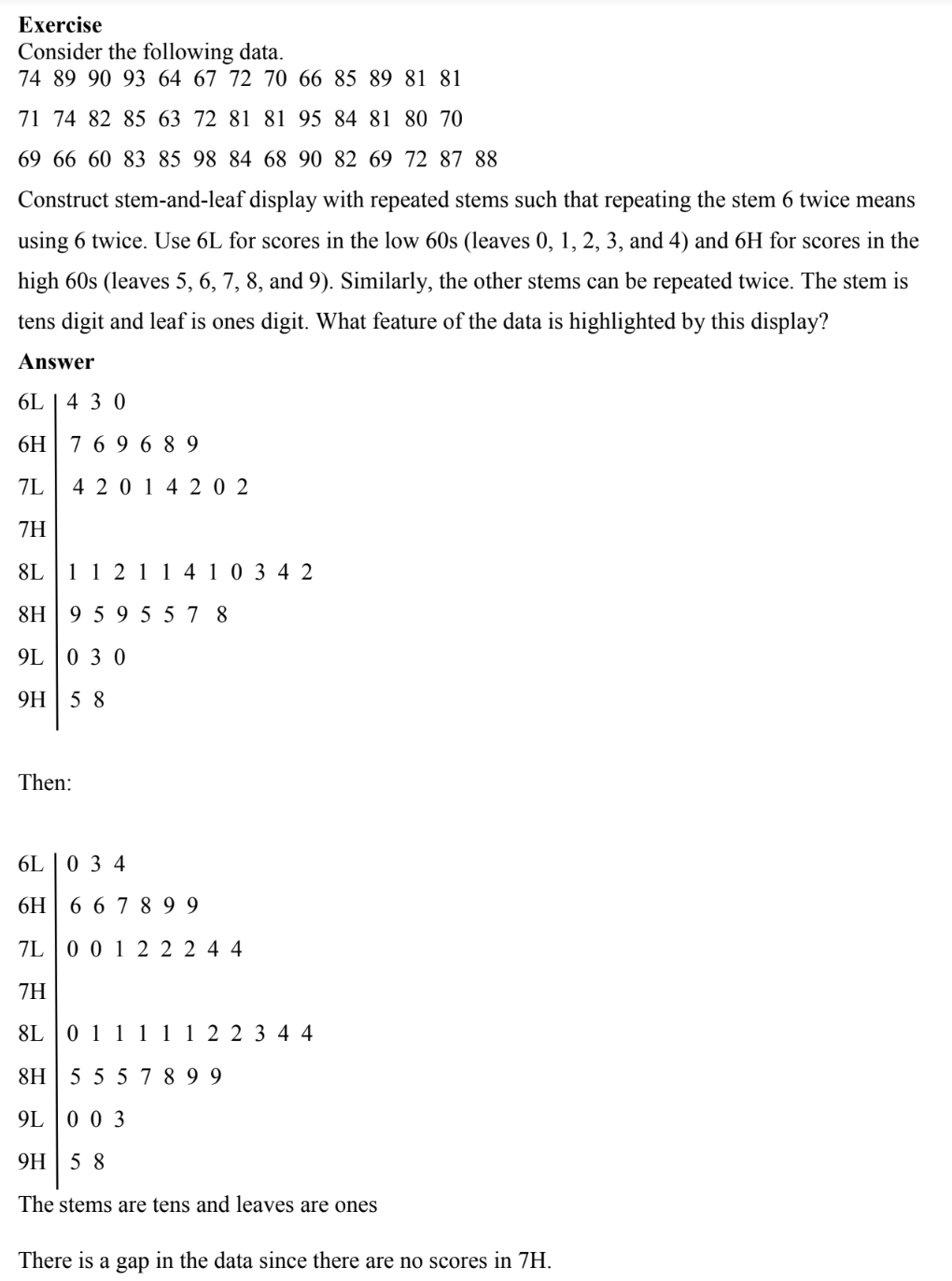
1. Select one or more leading digits for the stem values. The trailing digits become the leaves.

2. List possible stem values in a vertical column.

3. Record the leaf for each observation beside the corresponding stem value.

4. Indicate the units for stems and leaves in the display.

For example: The number 83 expresses as 8|3 if the stem is tense. The number 32.6 would be represented as 3|2.6 if the stem is tens.



**Histograms:**

A **numerical variable is discrete** if its set of possible values either is finite or else can be listed in an infinite sequence. It results from counting, in which case possible values are 0, 1, 2, 3, . . . or some subset of these integers.

A **numerical variable is continuous** if its possible values consist of an entire interval on the number line. It arises from making measurements. For example, if x is the pH of a chemical substance, then x could be any number between 0 and 14 such as 7.0, 7.03, 7.032, and so on.

**The frequency of any particular x** value is the number of times that value occurs in the data set. **The relative frequency of a value** is the fraction or proportion of times the value occurs and it is calculated as: (number of times the value occurs/number of observations in the data set)

**For example:** Data set consists of 200 observations on x = the number of courses a college student is taking this term. If 70 of these x values are 3, then frequency of the x value 3 is 70 and relative frequency of the x value 3 = 70⁄200 = 0.35.

**The relative frequency (percentage**) is calculated by multiplying a relative frequency by 100.

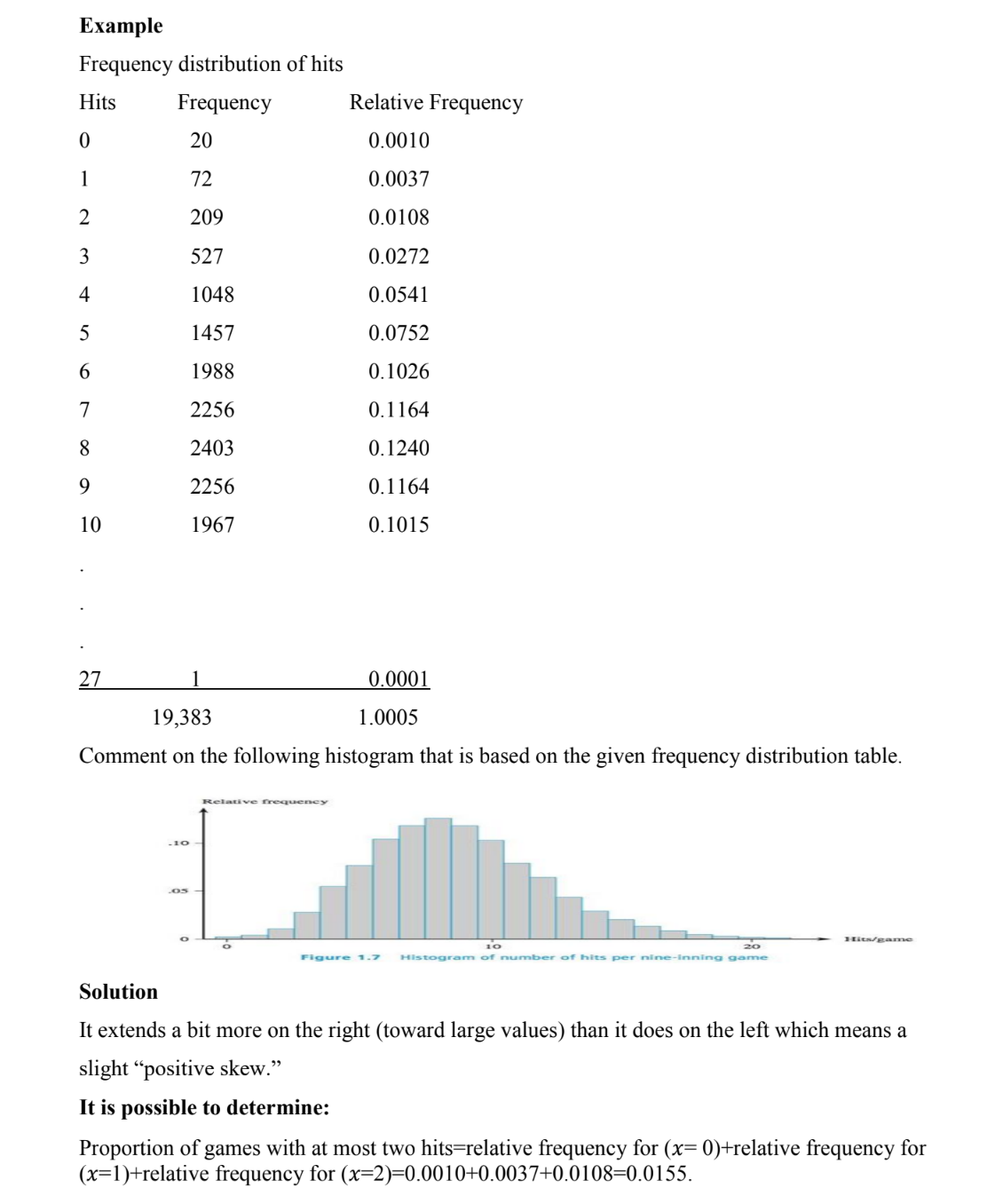
**For example:** The(70⁄200) × 100 = 35% of the students in the sample are taking three courses.

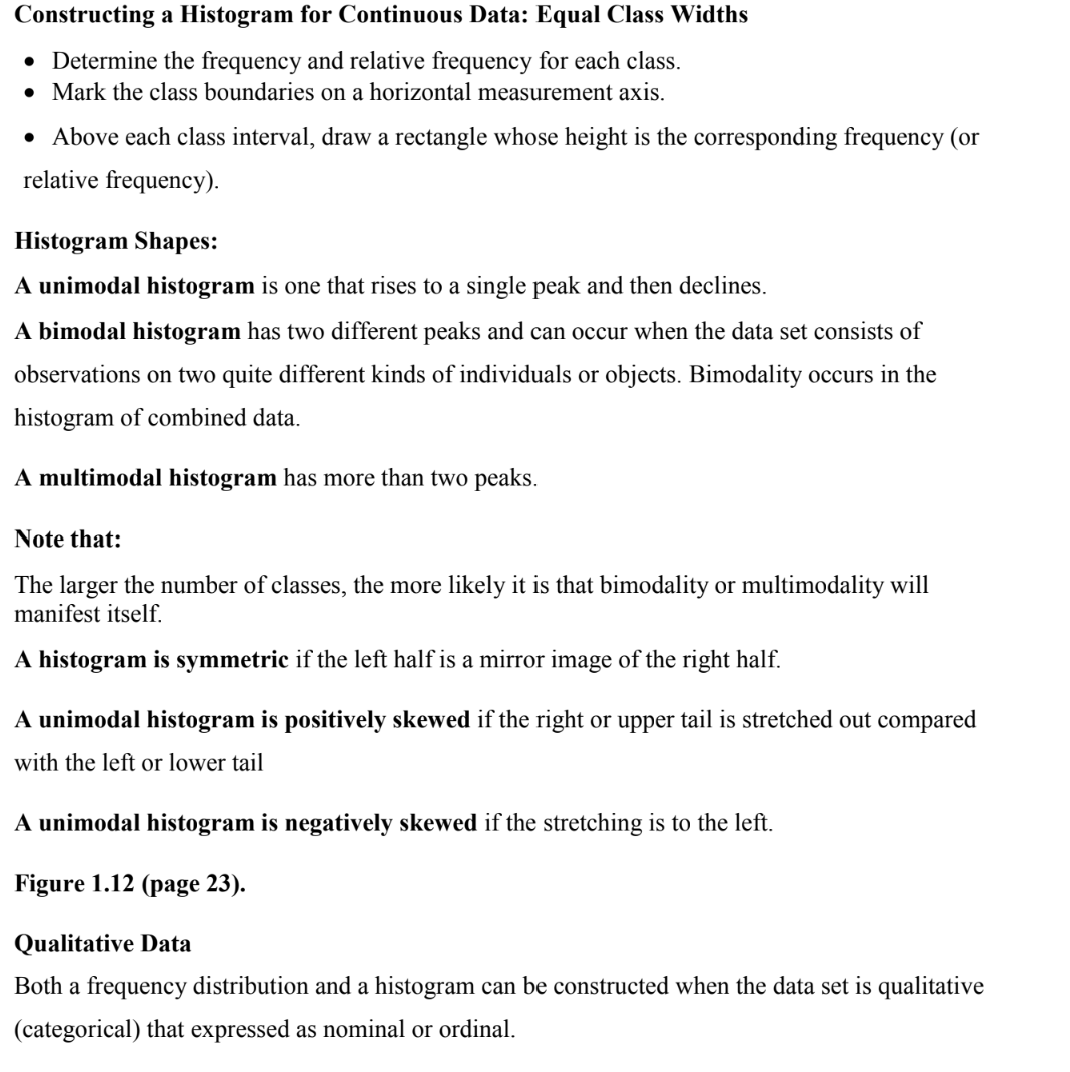
Note that:

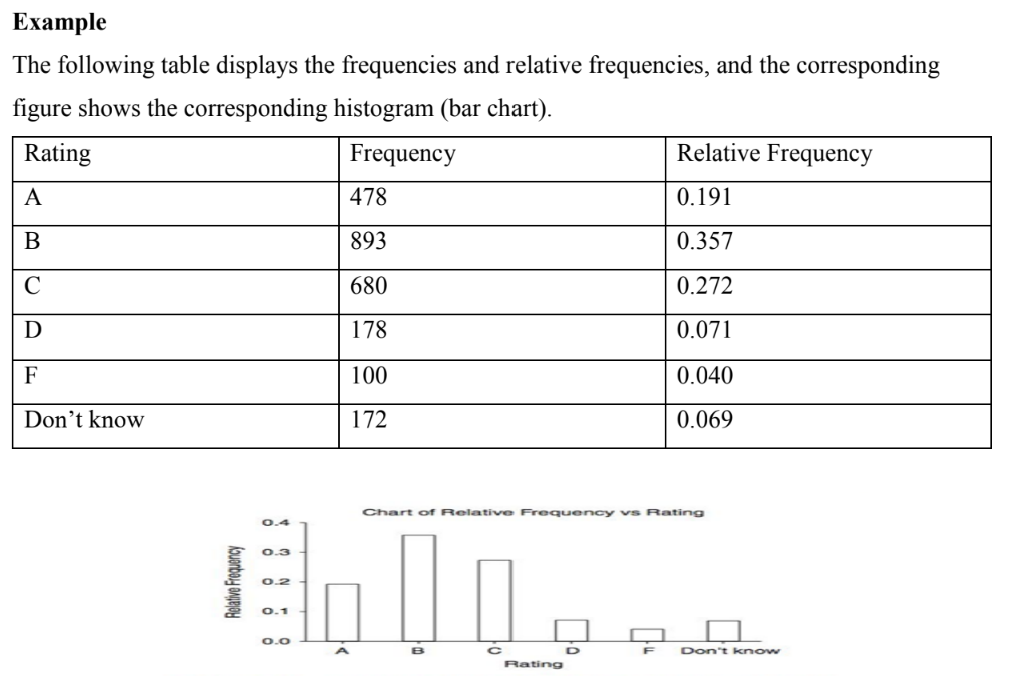
* The relative frequencies are usually of more interest than the frequencies.
* The relative frequencies should sum to 1, but in practice the sum may differ slightly from 1 because of rounding.
* A frequency distribution is a tabulation of the frequencies and/or relative frequencies.

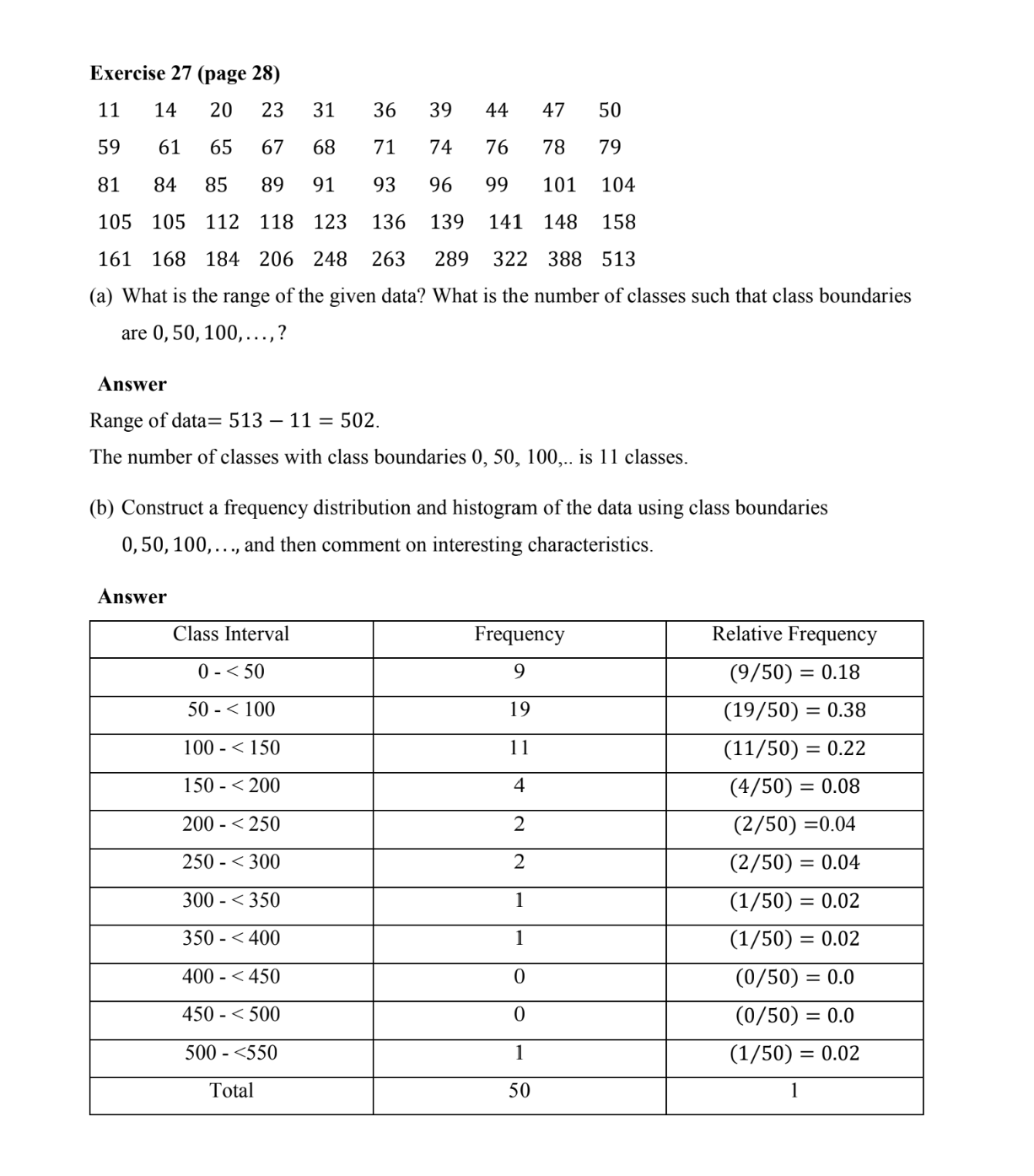
**Constructing a Histogram for Discrete Data**

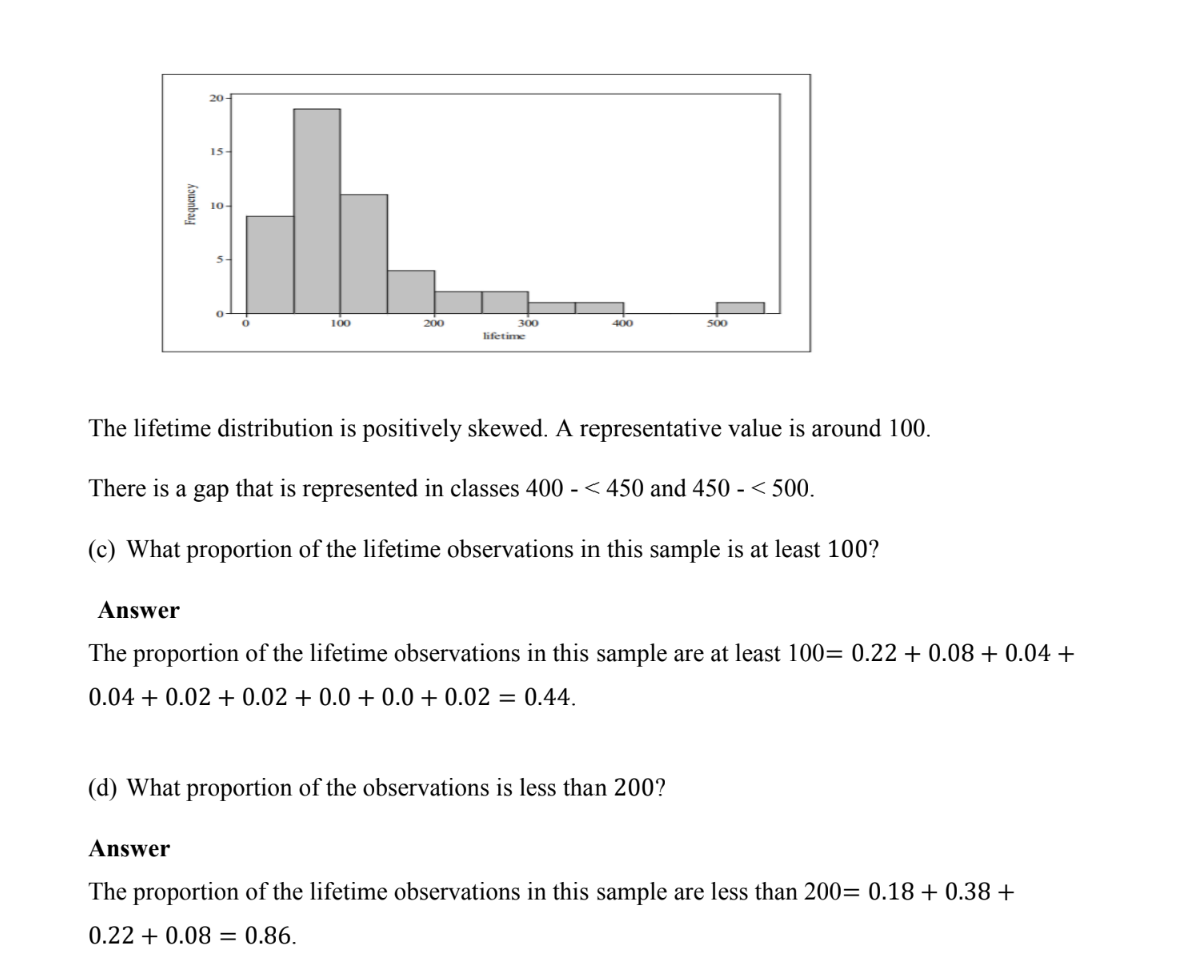
First, determine the frequency and relative frequency of each x value. Then mark possible x values on a horizontal scale. Above each value, draw a rectangle whose height is the relative frequency (or alternatively, the frequency) of that value.

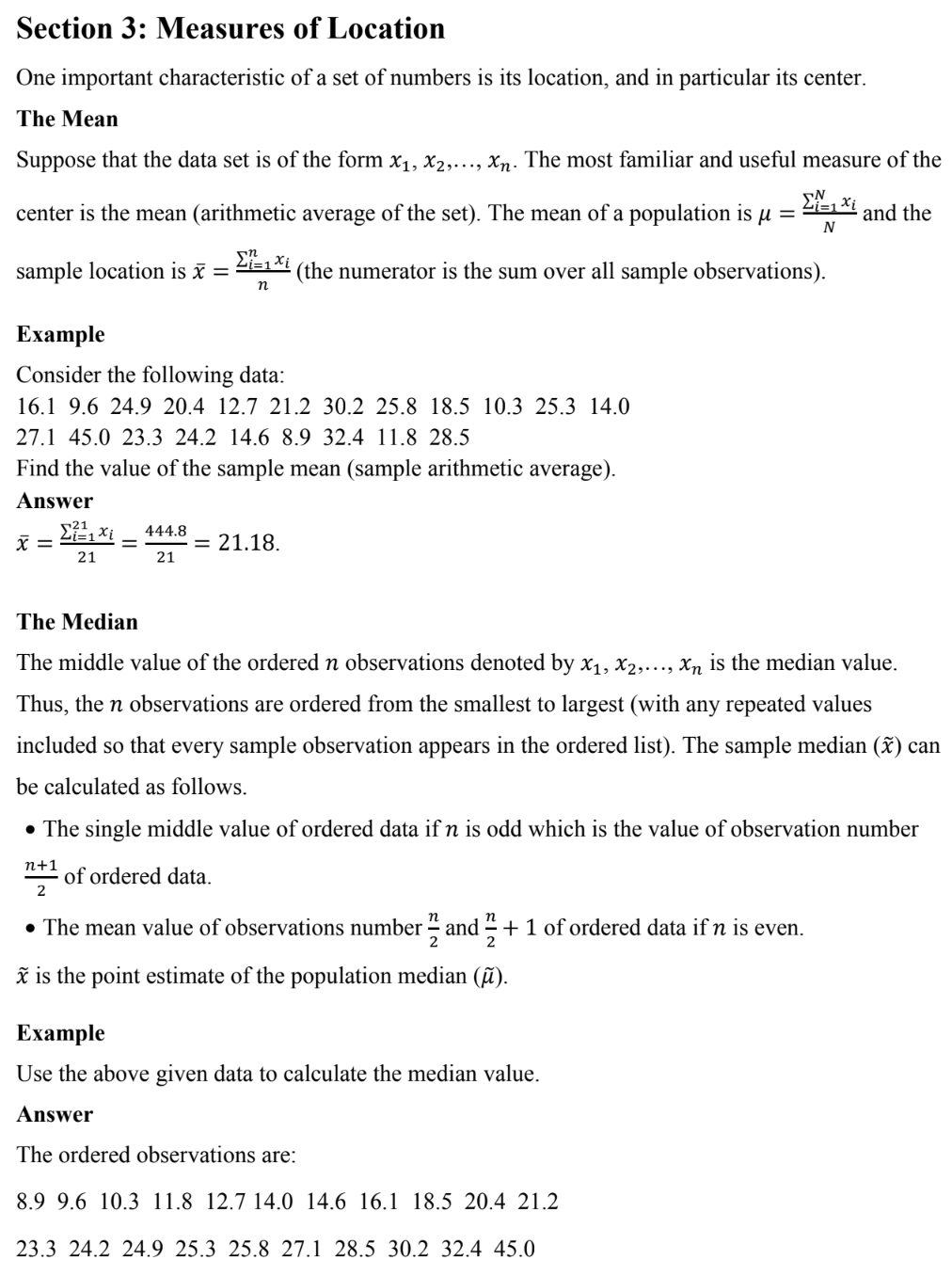


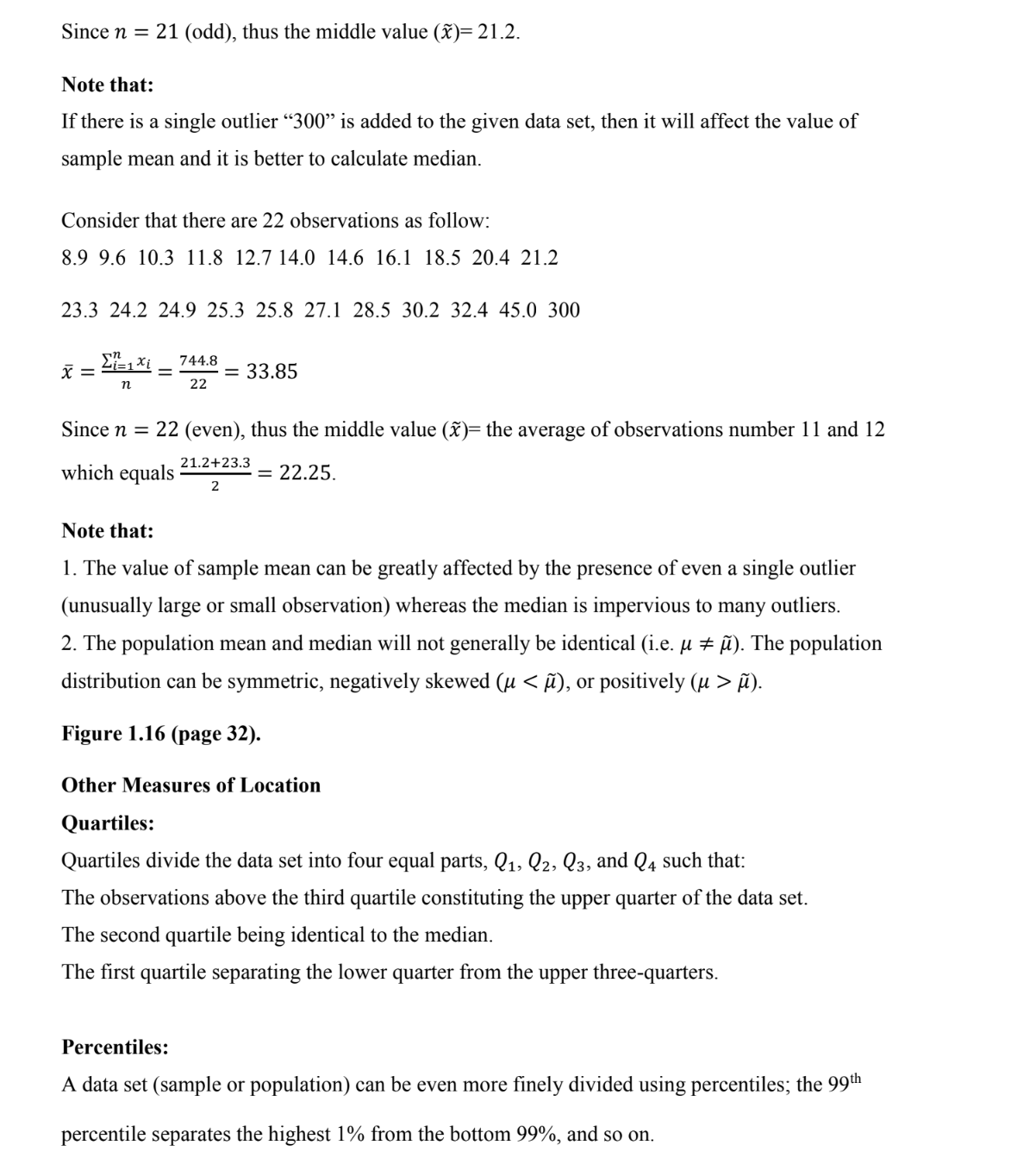












**CONTINUE LECTURE NOTES LATER**