CHAPTER 4

Inferences on a Single Population

CONTENTS

■ Example 4.1: How Accurately Are Areas Perceived?

The data in Table 4.1 are from an experiment in perceptual psychology. A person asked to judge the relative areas of circles of varying sizes typically judges the areas on a perceptual scale that can be approximated by

$$\text{judged area} = a(\text{true area})^b.$$

For most people the exponent b is between 0.6 and 1. That is, a person with an exponent of 0.8 who sees two circles, one twice the area of the other, would judge the larger one to be only $2^{0.8} = 1.74$ as large. Note that if the exponent is less than 1 a person tends to underestimate the area; if larger than 1, he or she will overestimate the area. The data shown in Table 4.1 are the set of measured exponents for 24 people from one particular experiment (Cleveland *et al.*, 1982). A histogram of this data is given in Figure 4.1.

It may be of interest to estimate the mean value of b for the population from which this sample is drawn; however, because we do not know the value of the population standard deviation we cannot use the methods of

Statistical Methods, DOI: 10.1016/B978-0-12-374970-3.00004-4

Table 4.1 Measured Exponents

0.58	0.63	0.69	0.72	0.74	0.79
0.88	0.88	0.90	0.91	0.93	0.94
0.97	0.97	0.99	0.99	0.99	1.00
1.03	1.04	1.05	1.07	1.18	1.27

Note: Reprinted with permission from the American Statistical Association.

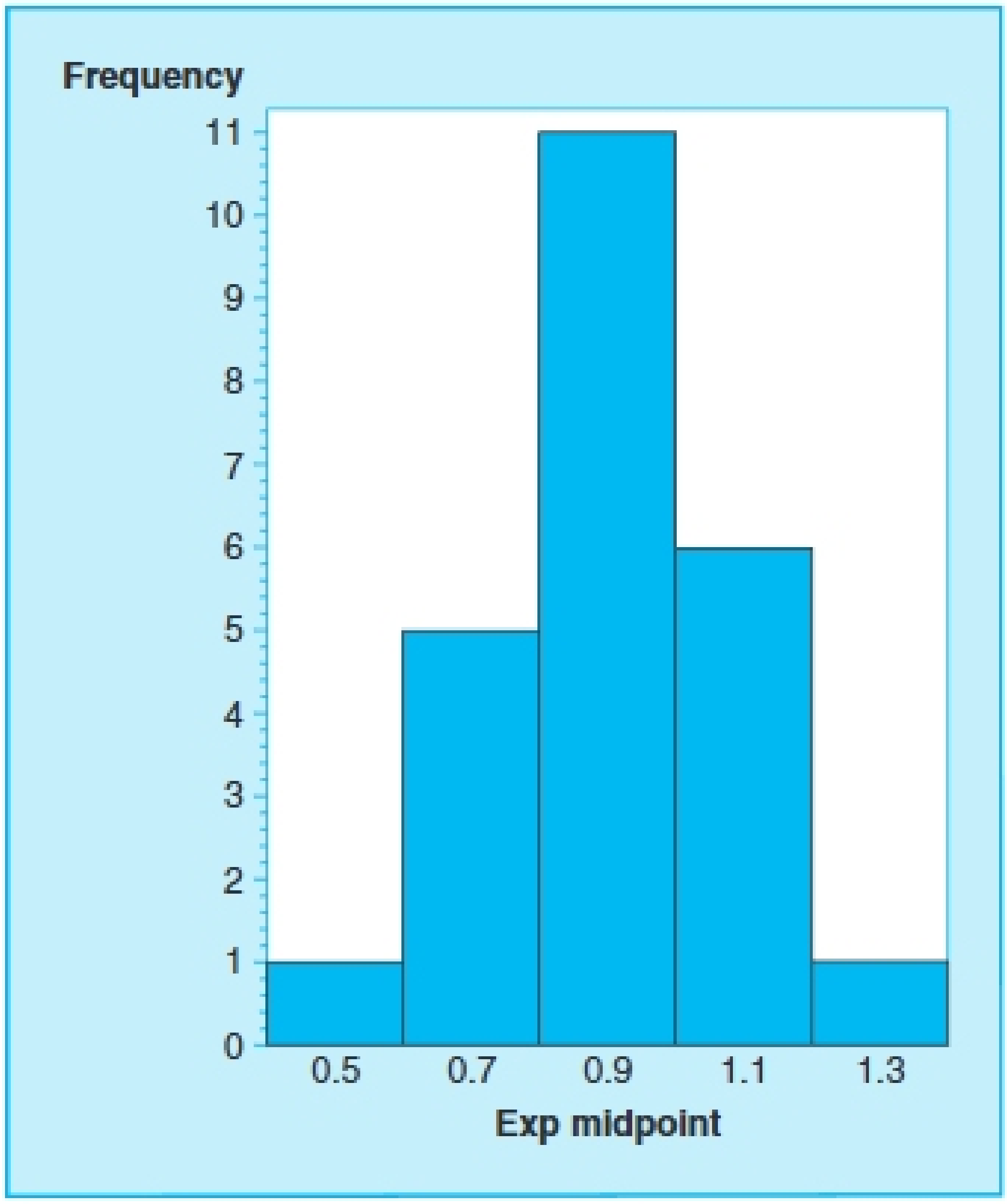

FIGURE 4.1
Histogram of Exponents in Example 4.1.

Chapter 3. Further, we might be interested in estimating the variance of these measurements as well. This chapter discusses methods for doing inferences on means when the population variance is unknown as well as inferences on the unknown population variance. The inferences for this example are presented in Sections 4.2 and 4.4. ■

4.1 INTRODUCTION

The examples used in Chapter 3 to introduce the concepts of statistical inference were not very practical, because they required outside knowledge of the population variance. This was intentional, as we wanted to avoid distractions from issues that were irrelevant to the principles we were introducing. We will now turn to examples that, although still quite simple, will have more useful applications. Specifically, we present procedures for

- making inferences on the mean of a normally distributed population where the variance is unknown,
- making inferences on the variance of a normally distributed population, and
- making inferences on the proportion of successes in a binomial population.

Increasing degrees of complexity are added in subsequent chapters. These begin in Chapter 5 with inferences for comparing two populations and in Chapter 6 with inferences on means from any number of populations. In Chapter 7 we present inference procedures for relationships between two variables through what we will refer to as the linear model, which is subsequently used as the common basis for many other statistical inference procedures. Additional chapters contain brief introductions to other statistical methods that cover different situations as well as methodology that may be used when underlying assumptions cannot be satisfied.

4.2 INFERENCES ON THE POPULATION MEAN

In Chapter 3 we used the sample mean $\bar{y}$ and its sampling distribution to make inferences on the population mean. For these inferences we used the fact that, for any approximately normally distributed population the statistic[1]

$$z = \frac{(\bar{y} - \mu)}{\sigma/\sqrt{n}}$$

has the standard normal distribution. This statistic has limited practical value because, if the population mean is unknown, it is also likely that the variance of the population is unknown.

In the discussion of the t distribution in Section 2.6 we noted that if, in the above equation, the known standard deviation is replaced by its estimate, s, the resulting statistic has a sampling distribution known as Student's t distribution. This distribution has a single parameter, called **degrees of freedom**, which is $(n - 1)$ for this case. Thus for statistical inferences on a mean from a normally distributed population, we can use the statistic

$$t = \frac{(\bar{y} - \mu)}{\sqrt{s^2/n}},$$

where $s^2 = \sum(y - \bar{y})^2/(n - 1)$.

---

[1] In Section 2.2 we adopted a convention that used capital letters to designate random variables and lowercase letters to represent realizations of those random variables. At that time we stated that the specificity of this designation would not be necessary after Chapter 3. Therefore, for this and subsequent chapters we will use lowercase letters exclusively.