Chapter 4

Hypothesis Testing

Obtaining a point estimate for an unknown parameter using the techniques in Chapter 2, and an associated interval estimate using the techniques in Chapter 3, is one type of statistical inference. The experimenter obtains a best guess for the unknown parameter (the point estimate) and a measure of the precision of the point estimate (the interval estimate). But not all statistical inference is numerical in nature.

There is a second broad type of statistical inference known as *hypothesis testing* which is introduced in this chapter. Hypothesis testing is a formal approach that has been devised by statisticians for testing whether data supports or refutes a conjecture (that is, a hypothesis) concerning an unknown population parameter. Over the years it has become the standard way of drawing conclusions from a data set. Hypothesis testing contains more elements than point and interval estimation; these elements will be defined and illustrated in Section 4.1.

4.1 Elements of hypothesis testing

Hypothesis testing is used to answer a question that is posed concerning a parameter. We begin with the definition of a hypothesis.

Definition 4.1 A *hypothesis* is a statement concerning an unknown parameter.

As a simple illustration, consider a coin with probability p of turning up heads when flipped, where p is an unknown but fixed parameter satisfying $0 < p < 1$. The hypothesis associated with the coin being fair is

$$p = \frac{1}{2}.$$

The hypothesis associated with the coin being biased is

$$p \neq \frac{1}{2}.$$

The hypothesis associated with the coin being biased in favor of turning up heads is

$$p > \frac{1}{2}.$$

These hypotheses could be tested with a coin flipping experiment.

Now consider a somewhat more complex example. A researcher has developed a new drug that allegedly extends the average survival time of a patient with liver cancer more than the current drug in use. Let μ_X be the population mean survival time of a liver cancer patient that uses the current drug. Let μ_Y be the population mean survival time of a liver cancer patient that uses the new drug. The hypothesis that the two drugs perform equally well in terms of the population mean survival time of a liver cancer patient is

$$\mu_X = \mu_Y.$$

The hypothesis that the new drug performs better than the current drug in terms of the mean survival time of a liver cancer patient is

$$\mu_X < \mu_Y.$$

This second hypothesis is often known as the *research worker's hypothesis* because this is the hypothesis that the experimenter is trying to show. Hypothesis testing parallels the scientific method. The two pillars of the scientific method are theory and experimentation. Formulating a hypothesis, as we have in the liver cancer drug example, corresponds to the development of a theory. In order to test the hypothesis, or theory, an experiment is conducted. In the case of the two drugs, the survival times $X_1, X_2, \ldots, X_n$ of n liver cancer patients taking the current drug would be collected. Likewise, the survival times $Y_1, Y_2, \ldots, Y_m$ of m liver cancer patients taking the new drug would also be collected. Care should be taken in the way that the patients are placed into the two groups so that the survival times in the two groups are not influenced by other factors such as age, gender, or the progress of the disease. This process is known to statisticians as *stratified sampling*. In addition, hypothesis testing in medicine often uses a *double-blind experiment* in which both the patient and the physician are unaware of whether the current drug or the new drug is being used. This procedure is used to ward off the *placebo effect*, which can bias the results of the experiment. A hypothesis test can then be conducted to assess the effectiveness of the new drug relative to the current drug.

The time-tested approach in hypothesis testing is to test one hypothesis, known as the null hypothesis, versus a second hypothesis, known as the alternative hypothesis. These two competing hypotheses are defined next.

Definition 4.2 A *hypothesis test* is a statistical inference technique for deciding between two competing propositions concerning an unknown parameter. The first proposition, known as the *null hypothesis*, is denoted by H_0. The second proposition, known as the *alternative hypothesis*, is denoted by H_1. The decision between H_0 and H_1 is based on a set of data values and is stated as either "reject H_0" or "fail to reject H_0."

Stating the null and alternative hypothesis is one of the first steps in conducting a hypothesis test. Some further details associated with H_0 and H_1 are given below.

- Some authors use the notation H_a or H_A for the alternative hypothesis. The more traditional H_1 is used here because of some helpful notational advantages.

- The null hypothesis has an inherently different nature than the alternative hypothesis because it is assumed to be true until enough evidence in the data causes it to be rejected. The null hypothesis is hardly ever true in any practical problem; it is almost always an approximation.

- The null hypothesis H_0 tends to be "no effect," "no difference," or "no change." The alternative hypothesis, on the other hand, is often the hoped-for conclusion of the test.

- The language used in the decision associated with the conclusion of the hypothesis test, that is, "reject H_0" or "fail to reject H_0," is important. There is a subtle difference between the conclusions "fail to reject H_0" and "accept H_0." The words "fail to reject H_0" imply that there is not enough statistical evidence in the sample to reject H_0 in favor of H_1. The null hypothesis H_0 has not been confirmed—there just simply isn't enough evidence in the data to reject it.

- Some authors refer to H_0 and H_1 as "statistical hypotheses."
- In the examples considered in this chapter, the data values will always be a random sample.

The first example of a hypothesis test is used to decide whether an unknown parameter θ assumes one value or another. To keep the mathematics simple, only a single observation, denoted by X, is collected to decide between the null and alternative hypotheses.

Example 4.1 A single observation X is collected to test the hypothesis

$$H_0 : \theta = 1$$

versus

$$H_1 : \theta = 3,$$

where θ is an unknown positive parameter from the continuous population distribution described by the probability density function

$$f(x) = 2\theta^{-2} x e^{-(x/\theta)^2} \qquad x > 0.$$

Conduct a hypothesis test associated with the data value $x = 2.2$.

The population distribution is the Rayleigh distribution. The null and alternative hypotheses completely define the population probability distribution of X. The population probability density functions associated with H_0 and H_1 are graphed in Figure 4.1. Let x be a realization of the random variable X. Because of the support of the population probability distribution, x will always be positive. It is clear from Figure 4.1 that a smaller value of x is evidence in favor of H_0 and a larger value of x is evidence in favor of H_1. Now consider the particular value of x collected in this case, that is, $x = 2.2$. The data value is out in the far right-hand tail of the probability density function associated

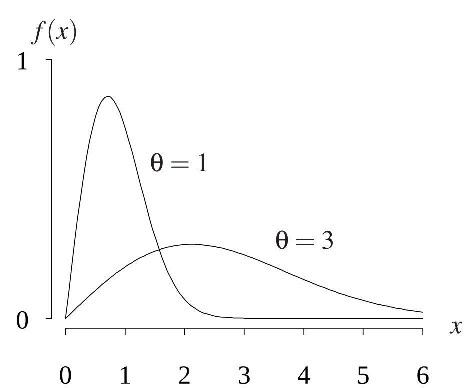

Figure 4.1: Population probability density functions for $\theta = 1$ and $\theta = 3$.

with $\theta = 1$, but in the center of the probability density function associated with $\theta = 3$, so it is reasonable to conclude that H_0 should be rejected in this case. The probability distribution associated with $\theta = 1$ is unlikely to have produced $x = 2.2$. This conclusion was drawn in a rather intuitive and unscientific manner by appealing to Figure 4.1, so the next example establishes a cutoff value for x for deciding between H_0 and H_1.

The null and alternative hypotheses in the previous example completely defined the probability density function of X. A hypothesis that completely defines a probability distribution is known as a simple hypothesis, which is defined next.

Definition 4.3 A hypothesis that completely specifies a probability distribution is called a *simple hypothesis*. A hypothesis that does not completely specify a probability distribution is called a *composite hypothesis*.

The next example considers a simple null hypothesis and a simple alternative hypothesis to test a hypothesis concerning the parameter θ from a special case of the beta distribution.

Example 4.2 A single observation X is collected to test the simple null hypothesis

$$H_0 : \theta = 1/2$$

versus the simple alternative hypothesis

$$H_1 : \theta = 3,$$

where θ is an unknown positive parameter from the continuous population distribution described by the probability density function

$$f(x) = \theta x^{\theta - 1} \qquad 0 < x < 1.$$

Analyze an appropriate decision rule based on the single data value.

The null and alternative hypotheses are both simple because they completely define the population probability distribution of X. The population probability density functions

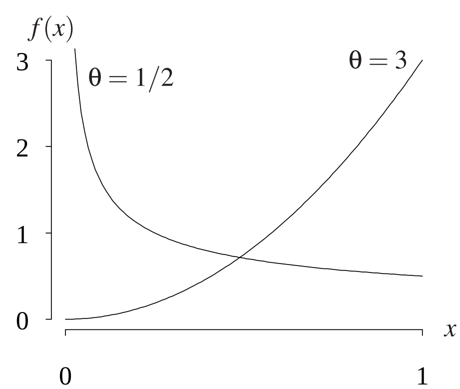

Figure 4.2: Population probability density functions for $\theta = 1/2$ and $\theta = 3$.