Chapter 10

1-Sample Location Problems

The basic ideas associated with statistical inference were introduced in Chapter 9. We developed these ideas in the context of drawing inferences about a single population mean, and we assumed that the sample was large enough to justify appeals to the Central Limit Theorem for normal approximations. The population mean is a natural measure of centrality, but it is not the only one. Furthermore, even if we are interested in the population mean, our sample may be too small to justify the use of a large-sample normal approximation. The purpose of the next several chapters is to explore more thoroughly how statisticians draw inferences about measures of centrality.

Measures of centrality are sometimes called location parameters. The title of this chapter indicates an interest in a location parameter of *one* population. More specifically, we assume that $X_1, \ldots, X_n \sim P$ are independently and identically distributed, we observe a random sample $\vec{x} = \{x_1, \ldots, x_n\}$, and we attempt to draw an inference about a location parameter of P. Because it is not always easy to identify the relevant population in a particular experiment, we begin with some examples. Our analysis of these examples is clarified by posing the following four questions:

1. What are the experimental units, i.e., what are the objects that are being measured?

2. From what population (or populations) were the experimental units drawn?

3. What measurements were taken on each experimental unit?

4. What random variables are relevant to the specified inference?

For the sake of specificity, we assume that the location parameter of interest in the following examples is the population median, $q_2(P)$.

Example 10.1 A recycled printing paper is supposed to have a basis weight of 24 pounds for 500 basic sheets[1] and a caliper (thickness) of 0.0048 inches per sheet. To determine if the paper is being correctly manufactured, a sample of sheets is drawn and the caliper of each sheet is measured with a micrometer. For this experiment:

1. An experimental unit is a sheet of paper. Notice that we are distinguishing between experimental units, the objects being measured (sheets of paper), and units of measurement (inches).

2. There is one population, viz., all sheets of paper that might be produced by the designated manufacturing process.

3. One measurement (caliper) is taken on each experimental unit.

4. Let X_i denote the caliper of sheet i. Then $X_1, \ldots, X_n \sim P$ and we are interested in drawing inferences about $q_2(P)$, the population median caliper. For example, we might test $H_0 : q_2(P) = 0.0048$ against $H_1 : q_2(P) \neq 0.0048$.

Example 10.2 A drug is supposed to lower blood pressure.[2] To determine if it does, a number of hypertensive patients are administered the drug for two months. Each person's mean arterial pressure is measured before and after the two month period. For this experiment:

1. An experimental unit is a patient.

2. There is one population of hypertensive patients. (It may be difficult to discern the precise population that was actually sampled. All hypertensive patients? All Hispanic male hypertensive patients who live in Houston, TX? All Hispanic male hypertensive patients who live in Houston, TX, and who are sufficiently well-disposed to the medical

[1] A basic sheet of bond paper is 17 inches by 22 inches, so 500 basic sheets is equivalent to 2000 letter-size sheets.

[2] Blood pressure is the force exerted by circulating blood in the body's large arteries. Arterial blood pressure is measured by a sphygmomanometer, usually in units of millimeters of mercury (mm Hg). Systolic arterial pressure is the maximum pressure attained during a cardiac cycle, diastolic arterial pressure is the minimum, and mean arterial pressure is the average.

establishment to participate in the study? In published journal articles, scientists are often rather vague about just what population was actually sampled.)

3. Two measurements (mean arterial pressure before and after treatment) are taken on each experimental unit. Let B_i and A_i denote the mean arterial pressures of patient i before and after treatment.

4. Let $X_i = B_i - A_i$, the decrease in mean arterial pressure for patient i. Then $X_1, \ldots, X_n \sim P$ and we are interested in drawing inferences about $q_2(P)$, the population median decrease. For example, we might test $H_0 : q_2(P) \leq 0$ against $H_1 : q_2(P) > 0$.

Example 10.3 Nancy Solomon and Thomas Hixon investigated the effect of Parkinson's disease (PD) on speech breathing.[3] Nancy recruited 14 PD patients to participate in the study. She also recruited 14 normal control (NC) subjects. Each NC subject was carefully matched to one PD patient with respect to sex, age, height, and weight. The lung volume of each study participant was measured. For this experiment:

1. An experimental unit was a matched PD-NC pair.

2. The population comprises all possible PD-NC pairs that satisfy the study criteria.

3. Two measurements (PD and NC lung volume) were taken on each experimental unit. Let D_i and C_i denote the PD and NC lung volumes of pair i.

4. Let $X_i = \log(D_i/C_i) = \log D_i - \log C_i$, the logarithm of the PD proportion of NC lung volume. (This is not the only way of comparing D_i and C_i, but it worked well in this investigation. As explained in Section 7.6, ratios can be difficult to analyze and logarithms convert ratios to differences. Furthermore, lung volume data tend to be skewed to the right. As in Exercise 7.7.4, logarithmic transformations of such data often have a symmetrizing effect.) Then $X_1, \ldots, X_n \sim P$ and we are interested in drawing inferences about $q_2(P)$. For example, to test the theory that PD restricts lung volume, we might test $H_0 : q_2(P) \geq 0$ against $H_1 : q_2(P) < 0$.

[3] N. P. Solomon and T. J. Hixon. Speech breathing in Parkinson's disease. *Journal of Speech and Hearing Research*, 36:294–310, April 1993.

This chapter is divided into sections according to distributional assumptions about the X_i and (consequently) the location parameter of interest.

10.1 If the data are assumed to be normally distributed, then we will be interested in inferences about the population's mean. Because normal distributions are symmetric, the population mean is also the population median.

10.2 If the data are only assumed to be continuously distributed (not necessarily normally, or even symmetrically distributed), then we will be interested in inferences about the population median.

10.3 If the data are assumed to be continuously and symmetrically distributed (but not necessarily normally distributed), then we will be interested in inferences about the population's center of symmetry.

Each section is subdivided into subsections, according to the type of inference (point estimation, hypothesis testing, set estimation) at issue.

10.1 The Normal 1-Sample Location Problem

In this section we assume that $P = \text{Normal}(\mu, \sigma^2)$. As necessary, we will distinguish between cases in which σ is known and cases in which σ is unknown.

10.1.1 Point Estimation

Because normal distributions are symmetric, the location parameter μ is the center of symmetry and therefore both the population mean and the population median. Hence, there are (at least) two natural estimators of μ, the sample mean $\bar{X}_n$ and the sample median $q_2(\hat{P}_n)$. Both are consistent, unbiased estimators of μ. We will compare them by considering their *asymptotic relative efficiency* (ARE). A rigorous definition of ARE is beyond the scope of this book, but the concept is easily interpreted.

If the true distribution is $P = \text{Normal}(\mu, \sigma^2)$, then the ARE of the sample median to the sample mean for estimating μ is

$$e(P) = \frac{2}{\pi} \doteq 0.64.$$

This statement has the following interpretation: for large samples, using the sample median to estimate a normal population mean is equivalent to randomly discarding approximately 36% of the observations and calculating the

sample mean of the remaining 64%. Thus, the sample mean is substantially more efficient than is the sample median at extracting location information from a normal sample.

In fact, if $P = \text{Normal}(\mu, \sigma^2)$, then the ARE of *any* estimator of μ to the sample mean is ≤ 1. This is sometimes expressed by saying that the sample mean is *asymptotically efficient* for estimating a normal mean. The sample mean also enjoys a number of other optimal properties in this case. The sample mean is unquestionably the preferred estimator for the normal 1-sample location problem.

10.1.2 Hypothesis Testing

If σ is known, then the possible distributions of X_i are

$$\left\{\text{Normal}(\mu, \sigma^2) : -\infty < \mu < \infty\right\}.$$

If σ is unknown, then the possible distributions of X_i are

$$\left\{\text{Normal}(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0\right\}.$$

We partition the possible distributions into two subsets, the null and alternative hypotheses. For example, if σ is known then we might specify

$$H_0 = \left\{\text{Normal}(0, \sigma^2)\right\} \quad \text{and} \quad H_1 = \left\{\text{Normal}(\mu, \sigma^2) : \mu \neq 0\right\},$$

which we would typically abbreviate as $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$. Analogously, if σ is unknown then we might specify

$$H_0 = \left\{\text{Normal}(0, \sigma^2) : \sigma > 0\right\}$$

and

$$H_1 = \left\{\text{Normal}(\mu, \sigma^2) : \mu \neq 0,\ \sigma > 0\right\},$$

which we would also abbreviate as $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

More generally, for any real number μ_0 we might specify

$$H_0 = \left\{\text{Normal}(\mu_0, \sigma^2)\right\} \quad \text{and} \quad H_1 = \left\{\text{Normal}(\mu, \sigma^2) : \mu \neq \mu_0\right\}$$

if σ is known, or

$$H_0 = \left\{\text{Normal}(\mu_0, \sigma^2) : \sigma > 0\right\}$$

and

$$H_1 = \left\{\text{Normal}(\mu, \sigma^2) : \mu \neq \mu_0,\ \sigma > 0\right\}$$

if σ in unknown. In both cases, we would typically abbreviate these hypotheses as $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$.

The preceding examples involve two-sided alternative hypotheses. Of course, as in Section 9.4, we might also specify one-sided hypotheses. However, the material in the present section is so similar to the material in Section 9.4 that we will discuss only two-sided hypotheses.

The intuition that underlies testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ was discussed in Section 9.4:

- If H_0 is true, then we would expect the sample mean to be close to the population mean μ_0.
- Hence, if $\bar{X}_n = \bar{x}_n$ is observed far from μ_n, then we are inclined to reject H_0.

To make this reasoning precise, we reject H_0 if and only if the significance probability

$$\mathbf{p} = P_{\mu_0}\left(\left|\bar{X}_n - \mu_0\right| \geq \left|\bar{x}_n - \mu_0\right|\right) \leq \alpha. \tag{10.1}$$

The first equation in (10.1) is a formula for a significance probability. Notice that this formula is identical to equation (9.2). The one difference between the material in Section 9.4 and the present material lies in how one computes $\mathbf{p}$. For emphasis, we recall the following:

1. The hypothesized mean μ_0 is a fixed number specified by the null hypothesis.
2. The estimated mean, $\bar{x}_n$, is a fixed number computed from the sample. Therefore, so is $|\bar{x}_n - \mu_0|$, the difference between the estimated mean and the hypothesized mean.
3. The estimator, $\bar{X}_n$, is a random variable.
4. The subscript in P_{μ_0} reminds us to compute the probability under $H_0 : \mu = \mu_0$.
5. The significance level α is a fixed number specified by the researcher, preferably before the experiment was performed.

To apply (10.1), we must compute $\mathbf{p}$. In Section 9.4, we overcame that technical difficulty by appealing to the Central Limit Theorem. This allowed us to approximate $\mathbf{p}$ even when we did not know the distribution of the X_i, but only for reasonably large sample sizes. However, if we know that $X_1, \ldots, X_n$ are normally distributed, then it turns out that we can calculate $\mathbf{p}$ exactly, even when n is small.

Case 1: The Population Variance is Known

Under the null hypothesis that $\mu = \mu_0$, $X_1, \ldots, X_n \sim \text{Normal}(\mu_0, \sigma^2)$ and

$$\bar{X}_n \sim \text{Normal}\left(\mu_0, \frac{\sigma^2}{n}\right).$$

This is the exact distribution of $\bar{X}_n$, not an asymptotic approximation. We convert $\bar{X}_n$ to standard units, obtaining

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \text{Normal}\,(0,1)\,. \tag{10.2}$$

The observed value of Z is

$$z = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}.$$

The significance probability is

$$\begin{aligned}
\mathbf{p} &= P_{\mu_0}\left(|\bar{X}_n - \mu_0| \geq |\bar{x}_n - \mu_0|\right) \\
&= P_{\mu_0}\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| \geq \left|\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right|\right) \\
&= P\left(|Z| \geq |z|\right) \\
&= 2P\left(Z \geq |z|\right).
\end{aligned}$$

In this case, the test that rejects H_0 if and only if $\mathbf{p} \leq \alpha$ is sometimes called the 1-*sample z-test*. The random variable Z is the *test statistic*.

Before considering the case of an unknown population variance, we remark that it is possible to derive point estimators from hypothesis tests. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, the test statistics are

$$Z(\mu_0) = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}.$$

If we observe $\bar{X}_n = \bar{x}_n$, then what value of μ_0 minimizes $|z(\mu_0)|$? Clearly, the answer is $\mu_0 = \bar{x}_n$. Thus, our preferred point estimate of μ is the μ_0 for which it is most difficult to reject $H_0 : \mu = \mu_0$. This type of reasoning will be extremely useful for analyzing situations in which we know how to test but don't know how to estimate.

Case 2: The Population Variance is Unknown

Statement (10.2) remains true if σ is unknown, but it is no longer possible to compute z. Therefore, we require a different test statistic for this case. A natural approach is to modify Z by replacing the unknown σ with an estimator of it. Toward that end, we introduce the test statistic

$$T_n = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}},$$

where S_n^2 is the unbiased estimator of the population variance defined by equation (9.1). Because T_n and Z are different random variables, they have different probability distributions and our first order of business is to determine the distribution of T_n.

We begin by stating a useful fact.

Theorem 10.1 *If* $X_1, \ldots, X_n \sim Normal(\mu, \sigma^2)$, *then*

$$\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 / \sigma^2 \sim \chi^2(n-1).$$

The χ^2 (chi-squared) distribution was described in Section 5.5 and Theorem 10.1 is closely related to Theorem 5.3.

Next we write

$$\begin{aligned} T_n &= \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \cdot \frac{\sigma/\sqrt{n}}{S_n/\sqrt{n}} \\ &= Z \cdot \frac{\sigma}{S_n} = Z/\sqrt{S_n^2/\sigma^2} \\ &= Z/\sqrt{\left[(n-1)S_n^2/\sigma^2\right]/(n-1)}. \end{aligned}$$

Using Theorem 10.1, we see that T_n can be written in the form

$$T_n = \frac{Z}{\sqrt{Y/\nu}},$$

where $Z \sim \text{Normal}(0,1)$ and $Y \sim \chi^2(\nu)$. If Z and Y are independent random variables, then it follows from Definition 5.7 that $T_n \sim t(n-1)$.

Both Z and $Y = (n-1)S_n^2/\sigma^2$ depend on $X_1, \ldots, X_n$, so one would be inclined to think that Z and Y are dependent. This is usually the case, but it turns out that they are independent if $X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma^2)$. This is another remarkable property of normal distributions, usually stated as follows:

Theorem 10.2 *If $X_1, \ldots, X_n \sim Normal(\mu, \sigma^2)$, then $\bar{X}_n$ and S_n^2 are independent random variables.*

The result that interests us can then be summarized as follows:

Corollary 10.1 *If $X_1, \ldots, X_n \sim Normal(\mu_0, \sigma^2)$, then*

$$T_n = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t(n-1).$$

Now let

$$t_n = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}},$$

the observed value of the test statistic T_n. The significance probability is

$$\mathbf{p} = P_{\mu_0}\left(|T_n| \geq |t_n|\right) = 2P_{\mu_0}\left(T_n \geq |t_n|\right).$$

In this case, the test that rejects H_0 if and only if $\mathbf{p} \leq \alpha$ is called *Student's 1-sample t-test*. Because it is rarely the case that the population variance is known when the population mean is not, Student's 1-sample t-test is used much more frequently than the 1-sample z-test. We will use the `R` function `pt` to compute significance probabilities for Student's 1-sample t-test, as illustrated in the following examples.

Example 10.4 Suppose that, to test $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ (a 2-sided alternative), we draw a sample of size $n = 25$ and observe $\bar{x} = 1$ and $s = 3$. Then $t = (1-0)/(3/\sqrt{25}) = 5/3$ and the 2-tailed significance probability is computed using both tails of the $t(24)$ distribution, i.e., $\mathbf{p} = 2 * \texttt{pt}(-5/3, \texttt{df} = 24) \doteq 0.1086$.

Example 10.5 Suppose that, to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ (a 1-sided alternative), we draw a sample of size $n = 25$ and observe $\bar{x} = 2$ and $s = 5$. Then $t = (2-0)/(5/\sqrt{25}) = 2$ and the 1-tailed significance probability is computed using one tail of the $t(24)$ distribution, i.e., $\mathbf{p} = 1 - \texttt{pt}(2, \texttt{df} = 24) \doteq 0.0285$.

10.1.3 Set Estimation

As in Section 9.5, we will derive confidence intervals from tests. We imagine testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ for every $\mu_0 \in (-\infty, \infty)$. The μ_0 for which $H_0 : \mu = \mu_0$ is rejected are implausible values of μ; the μ_0 for which $H_0 : \mu = \mu_0$ is not rejected constitute the confidence interval. To accomplish this, we will have to derive the critical values of our tests. A significance level of α will result in a confidence coefficient of $1 - \alpha$.

Case 1: The Population Variance is Known

If σ is known, then we reject $H_0 : \mu = \mu_0$ if and only if

$$\mathbf{p} = P_{\mu_0}\left(\left|\bar{X}_n - \mu_0\right| \geq \left|\bar{x}_n - \mu_0\right|\right) = 2\Phi\left(-\left|z_n\right|\right) \leq \alpha,$$

where $z_n = (\bar{x}_n - \mu_0)/(\sigma/\sqrt{n})$. By the symmetry of the normal distribution, this condition is equivalent to the condition

$$1 - \Phi\left(-\left|z_n\right|\right) = P\left(Z > -\left|z_n\right|\right) = P\left(Z < \left|z_n\right|\right) = \Phi\left(\left|z_n\right|\right) \geq 1 - \alpha/2,$$

where $Z \sim \text{Normal}(0, 1)$, and therefore to the condition $|z_n| \geq q_z$, where q_z denotes the $1 - \alpha/2$ quantile of Normal$(0, 1)$. The quantile q_z is the critical value of the two-sided 1-sample z-test. Thus, given a significance level α and a corresponding critical value q_z, we reject $H_0 : \mu = \mu_0$ if and only if (iff)

$$\begin{aligned} & \left|\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right| = |z_n| \geq q_z \\ \text{iff} \quad & |\bar{x}_n - \mu_0| \geq q_z\sigma/\sqrt{n} \\ \text{iff} \quad & \mu_0 \notin \left(\bar{x}_n - q_z\sigma/\sqrt{n},\ \bar{x}_n + q_z\sigma/\sqrt{n}\right) \end{aligned}$$

and we conclude that the desired set of plausible values is the interval

$$\left(\bar{x}_n - q_z\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + q_z\frac{\sigma}{\sqrt{n}}\right).$$

Notice that both the preceding derivation and the resulting confidence interval are identical to the derivation and confidence interval in Section 9.5. The only difference is that, because we are now assuming that $X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma^2)$ instead of relying on the Central Limit Theorem, no approximation is required.

Example 10.6 Suppose that we desire 90% confidence about μ and $\sigma = 3$ is known. Then $\alpha = 0.10$ and $q_z \doteq 1.645$. Suppose that we draw $n = 25$ observations and observe $\bar{x}_n = 1$. Then

$$1 \pm 1.645\frac{3}{\sqrt{25}} = 1 \pm 0.987 = (0.013, 1.987)$$

is a 0.90-level confidence interval for μ.

Case 2: The Population Variance is Unknown

If σ is unknown, then it must be estimated from the sample. The reasoning in this case is the same, except that we rely on Student's 1-sample t-test.

As before, we use S_n^2 to estimate σ^2. The critical value of the 2-sided 1-sample t-test is q_t, the $1 - \alpha/2$ quantile of a t distribution with $n - 1$ degrees of freedom, and the confidence interval is

$$\left(\bar{x}_n - q_t \frac{s_n}{\sqrt{n}},\ \bar{x}_n + q_t \frac{s_n}{\sqrt{n}}\right).$$

Example 10.7 Suppose that we desire 90% confidence about μ and σ is unknown. Suppose that we draw $n = 25$ observations and observe $\bar{x}_n = 1$ and $s = 3$. Then $q_t = \texttt{qt}(.95, \texttt{df} = 24) \doteq 1.711$ and

$$1 \pm 1.711 \times 3/\sqrt{25} = 1 \pm 1.027 = (-0.027, 2.027)$$

is a 90% confidence interval for μ. Notice that the confidence interval is larger when we use $s = 3$ instead of $\sigma = 3$.

10.2 The General 1-Sample Location Problem

In Section 10.1 we assumed that $X_1, \ldots, X_n \sim P$ and $P = \text{Normal}(\mu, \sigma^2)$. In this section, we again assume that $X_1, \ldots, X_n \sim P$, but now we assume only that the X_i are continuous random variables.

Because P is not assumed to be symmetric, we must decide which location parameter to study. The population median, $q_2(P)$, enjoys several advantages. Unlike the population mean, the population median always exists and is not sensitive to the influence of outliers. Furthermore, it turns out that one can develop fairly elementary ways to study medians, even when little is known about the probability distribution P. For simplicity, we now denote the population median by θ.

10.2.1 Hypothesis Testing

It is convenient to begin our study of the general 1-sample location problem with a discussion of hypothesis testing. As in Section 10.1, we initially consider testing a 2-sided alternative, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. We will explicate a procedure known as the *sign test*.

The intuition that underlies the sign test is elementary. If the population median is $\theta = \theta_0$, then when we sample P we should observe roughly half

the x_i above θ_0 and half the x_i below θ_0. Hence, if we observe proportions of x_i above/below θ_0 that are very different from one half, then we are inclined to reject the possibility that $\theta = \theta_0$.

More formally, let $p_+ = P_{H_0}(X_i > \theta_0)$ and $p_- = P_{H_0}(X_i < \theta_0)$. Because the X_i are continuous, $P_{H_0}(X_i = \theta_0) = 0$ and therefore $p_+ = p_- = 0.5$. Hence, under H_0, observing whether $X_i > \theta_0$ or $X_i < \theta_0$ is equivalent to tossing a fair coin, i.e., to observing a Bernoulli trial with success probability $p = 0.5$. The sign test is the following procedure:

1. Let $\vec{x} = \{x_1, \ldots, x_n\}$ denote the observed sample. If the X_i are continuous random variables, then $P(X_i = \theta_0) = 0$ and it should be that each $x_i \neq \theta_0$. In practice, of course, it may happen that we do observe one or more $x_i = \theta_0$. For the moment, we assume that $\vec{x}$ contains no such values.

2. Let
$$Y = \#\{X_i > \theta_0\} = \#\{X_i - \theta_0 > 0\}$$
be the test statistic. Under $H_0 : \theta = \theta_0$, $Y \sim \text{Binomial}(n; p = 0.5)$. The observed value of the test statistic is
$$y = \#\{x_i > \theta_0\} = \#\{x_i - \theta_0 > 0\}.$$

3. Notice that $EY = n/2$. The significance probability is
$$\mathbf{p} = P_{\theta_0}\left(\left|Y - \frac{n}{2}\right| \geq \left|y - \frac{n}{2}\right|\right).$$
The sign test rejects $H_0 : \theta = \theta_0$ if and only if $\mathbf{p} \leq \alpha$.

4. To compute $\mathbf{p}$, we first note that
$$\left|Y - \frac{n}{2}\right| \geq \left|y - \frac{n}{2}\right|$$
is equivalent to the event
   (a) $\{Y \leq y \text{ or } Y \geq n - y\}$ if $y \leq n/2$;
   (b) $\{Y \geq y \text{ or } Y \leq n - y\}$ if $y \geq n/2$.

   To accommodate both cases, let $c = \min(y, n - y)$. Then
$$\mathbf{p} = P_{\theta_0}(Y \leq c) + P_{\theta_0}(Y \geq n - c) = 2P_{\theta_0}(Y \leq c) = \texttt{2*pbinom(c,n,.5)}.$$

Example 10.8(a) Suppose that we want to test $H_0 : \theta = 100$ versus $H_1 : \theta \neq 100$ at significance level $\alpha = 0.05$, having observed the sample

$$\vec{x} = \{98.73, 97.17, 100.17, 101.26, 94.47, 96.39, 99.67, 97.77, 97.46, 97.41\}.$$

Here $n = 10$, $y = \#\{x_i > 100\} = 2$, and $c = \min(2, 10 - 2) = 2$, so

$$\mathbf{p} = \texttt{2*pbinom(2,10,.5)} = 0.109375 > 0.05$$

and we decline to reject H_0.

Example 10.8(b) Now suppose that we want to test $H_0 : \theta \leq 97$ versus $H_1 : \theta > 97$ at significance level $\alpha = 0.05$, using the same data. Here $n = 10$, $y = \#\{x_i > 97\} = 8$, and $c = \min(8, 10 - 8) = 2$. Because large values of Y are evidence against $H_0 : \theta \leq 97$,

$$\begin{aligned} \mathbf{p} &= P_{\theta_0}(Y \geq y) = P_{\theta_0}(Y \geq 8) = 1 - P_{\theta_0}(Y \leq 7) \\ &= \texttt{1-pbinom(7,10,.5)} = 0.0546875 > 0.05 \end{aligned}$$

and we decline to reject H_0.

Thus far we have assumed that the sample contains no values for which $x_i = \theta_0$. In practice, we may well observe such values. For example, if the measurements in Example 10.8(a) were made less precisely, then we might have observed the following sample:

$$\vec{x} = \{99, 97, 100, 101, 94, 96, 100, 98, 97, 97\}. \tag{10.3}$$

If we want to test $H_0 : \theta = 100$ versus $H_1 : \theta \neq 100$, then we have two values that equal θ_0 and the sign test requires modification.

We assume that $\#\{x_i = \theta_0\}$ is fairly small; otherwise, the assumption that the X_i are continuous is questionable. We consider two possible ways to proceed:

1. Perhaps the most satisfying solution is to compute all of the significance probabilities that correspond to different ways of counting the $x_i = \theta_0$ as larger or smaller than θ_0. If there are k observations $x_i = \theta_0$, then this will produce 2^k significance probabilities, which we might average to obtain a single $\mathbf{p}$.

2. Alternatively, let $\mathbf{p}_0$ denote the significance probability obtained by counting in the way that is most favorable to H_0 (least favorable to

H_1). This is the largest of the possible significance probabilities, so if $\mathbf{p}_0 \leq \alpha$ then we reject H_0. Similarly, let $\mathbf{p}_1$ denote the significance probability obtained by counting in the way that is least favorable to H_0 (most favorable to H_1). This is the smallest of the possible significance probabilities, so if $\mathbf{p}_1 > \alpha$ then we decline to reject H_0. If $\mathbf{p}_0 > \alpha \geq \mathbf{p}_1$, then we simply declare the results to be equivocal.

Example 10.8(c) Suppose that we want to test $H_0 : \theta = 100$ versus $H_1 : \theta \neq 100$ at significance level $\alpha = 0.05$, having observed the sample (10.3). Here $n = 10$ and $y = \#\{x_i > 100\}$ depends on how we count the observations $x_3 = x_7 = 100$. There are $2^2 = 4$ possibilities:

possibility	$y = \#\{x_i > 100\}$	$c = \min(y, 10 - y)$	$\mathbf{p}$
$y_3 < 100, y_7 < 100$	1	1	0.021484
$y_3 < 100, y_7 > 100$	2	2	0.109375
$y_3 > 100, y_7 < 100$	2	2	0.109375
$y_3 > 100, y_7 > 100$	3	3	0.343750

Noting that $\mathbf{p}_0 \doteq 0.344 > 0.05 > 0.021 \doteq \mathbf{p}_1$, we might declare the results to be equivocal. However, noting that only 1 of the 4 possibilities lead us to reject H_0 (and that the average $\mathbf{p} \doteq 0.146$), we might conclude—somewhat more decisively—that there is insufficient evidence to reject H_0. The distinction between these two interpretations is largely rhetorical, as the fundamental logic of hypothesis testing requires that we decline to reject H_0 unless there is compelling evidence against it.

10.2.2 Point Estimation

Next we consider the problem of estimating the population median. A natural estimate is the plug-in estimate, the sample median. Another approach begins by posing the following question: For what value of θ_0 is the sign test least inclined to reject $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta \neq \theta_0$? The answer to this question is also a natural estimate of the population median.

In fact, the plug-in and sign-test approaches lead to the same estimation procedure. To understand why, we focus on the case that n is even, in which case $n/2$ is a possible value of $Y = \#\{X_i > \theta_0\}$. If $|y - n/2| = 0$, then

$$\mathbf{p} = P\left(\left|Y - \frac{n}{2}\right| \geq 0\right) = 1.$$

We see that the sign test produces the maximal significance probability of $\mathbf{p} = 1$ when $y = n/2$, i.e., when θ_0 is chosen so that precisely half the

observations exceed θ_0. This means that the sign test is least likely to reject $H_0 : \theta = \theta_0$ when θ_0 is the sample median. (A similar argument leads to the same conclusion when n is odd.)

Thus, using the sign test to test hypotheses about population medians corresponds to using the sample median to estimate population medians, just as using Student's t-test to test hypotheses about population means corresponds to using the sample mean to estimate population means. One consequence of this remark is that, when the population mean and median are identical, the "Pitman efficiency" of the sign test to Student's t-test equals the asymptotic relative efficiency of the sample median to the sample mean. For example, using the sign test on normal data is asymptotically equivalent to randomly discarding 36% of the observations, then using Student's t-test on the remaining 64%.

10.2.3 Set Estimation

Finally, we consider the problem of constructing a $(1 - \alpha)$-level confidence interval for the population median. Again we rely on the sign test, determining for which θ_0 the level-α sign test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ does not reject H_0.

The sign test rejects $H_0 : \theta = \theta_0$ if and only if

$$y(\theta_0) = \#\{x_i > \theta_0\}$$

is either too large or too small. Equivalently, the sign test declines to reject H_0 if and only if θ_0 is such that the numbers of observations above and below θ_0 are roughly equal.

To determine the critical value for the desired sign test, we suppose that $Y \sim \text{Binomial}(n; 0.5)$. We would like to find k such that $\alpha = 2P(Y \leq k)$, or $\alpha/2 = \texttt{pbinom}(k, n, 0.5)$. In practice, we won't be able to solve this equation exactly. Binomial distributions are discrete; hence, `pbinom(k,n,0.5)` only attains certain values. We will use the `qbinom` function plus trial-and-error to find k such that $\texttt{pbinom}(k, n, 0.5) \approx \alpha/2$, then modify our choice of α accordingly.

Having determined a suitable (α, k), the sign test rejects $H_0 : \theta = \theta_0$ at level α if and only if either $y(\theta_0) \leq k$ or $y(\theta_0) \geq n - k$. We would like to translate these inequalities into an interval of plausible values of θ_0. To do so, it is helpful to sort the values observed in the sample.

Definition 10.1 *The* order statistics *of* $\vec{x} = \{x_1, \ldots, x_n\}$ *are any permutation of the* x_i *such that*

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}.$$

If $\vec{x}$ contains n distinct values, then there is a unique set of order statistics and the above inequalities are strict; otherwise, we say that $\vec{x}$ contains ties.

Thus, $x_{(1)}$ is the smallest value in $\vec{x}$ and $x_{(n)}$ is the largest. If $n = 2m + 1$ (n is odd), then the sample median is $x_{(m+1)}$; if $n = 2m$ (n is even), then the sample median is $[x_{(m)} + x_{(m+1)}]/2$.

For simplicity we assume that $\vec{x}$ contains no ties. If $\theta_0 < x_{(k+1)}$, then at least $n - k$ observations exceed θ_0 and the sign test rejects $H_0 : \theta = \theta_0$. Similarly, if $\theta_0 > x_{(n-k)}$, then no more than k observations exceed θ_0 and the sign test rejects $H_0 : \theta = \theta_0$. We conclude that the sign test does not reject $H_0 : \theta = \theta_0$ if and only if θ_0 lies in the $(1-\alpha)$-level confidence interval

$$\left(x_{(k+1)}, x_{(n-k)}\right).$$

Example 10.8(d) Using the 10 observations from Example 10.8(a), we endeavor to construct a 0.90-level confidence interval for the population median. We begin by determining a suitable choice of (α, k). If $1-\alpha = 0.90$, then $\alpha/2 = 0.05$. The `R` command `qbinom(.05,10,.5)` returns $k = 2$. Next we experiment:

k	`pbinom`$(k, 10, 0.5)$
2	0.0546875
1	0.01074219

We choose $k = 2$, resulting in a confidence level of

$$1 - \alpha = 1 - 2 \cdot 0.0546875 = 0.890625 \doteq 0.89,$$

nearly equal to the requested level of 0.90. Now, upon sorting the data (the `sort` function in `R` may be useful), we quickly discern that the desired confidence interval is

$$\left(x_{(3)}, x_{(8)}\right) = (97.17, 99.67).$$

10.3 The Symmetric 1-Sample Location Problem

Again we assume that $X_1, \ldots, X_n \sim P$. In Section 10.1 we made a strong assumption, that $P = \text{Normal}(\mu, \sigma^2)$. The procedures that we derived under the assumption of normality, e.g., the sample mean estimator and Student's 1-sample t-test, are efficient when P is normal, but may not perform well if P is not normal. For example, the sample mean is sensitive to the effect of outliers.

In Section 10.2 we made a weak assumption, that the X_i are continuous random variables with pdf f. The procedures that we derived, e.g., the sample median estimator and the sign test, perform well in a variety of circumstances, but are somewhat inefficient when P is normal.

The present section describes one possible compromise between the normal methods of Section 10.1 and the general methods of Section 10.2. To the assumption that the X_i are continuous random variables, we add the assumption that their pdf, f, is symmetric. The assumption of symmetry permits comparison of procedures for means and procedures for medians, as the population mean and the population median necessarily equal the population center of symmetry, which we denote by θ. We follow the same reasoning that we deployed in Section 10.2, first describing a test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, then using it to derive point and set estimators of θ.

10.3.1 Hypothesis Testing

Let $D_i = X_i - \theta_0$, the amount by which X_i exceeds the hypothesized center of symmetry. The numerator of Student's 1-sample t statistic is the sample mean of the D_i, whereas the sign test statistic is the number of observations for which $D_i > 0$. We now describe a third test based on $D_1, \ldots, D_n$. The *Wilcoxon signed rank test* extracts less information from $D_1, \ldots, D_n$ than does Student's 1-sample t-test, but more information than does the sign test.

Because the X_i are continuous random variables, each $P(D_i = 0)$ and each $P(|D_i| = |D_j|) = 0$ for $i \neq j$. Therefore, with probability one, there is a unique ordering of the absolute differences:

$$|D_{i_1}| < |D_{i_2}| < \cdots < |D_{i_n}| .$$

The Wilcoxon signed rank test is based on this ordering. (In practice we may encounter samples with $d_i = 0$ and/or $|d_i| = |d_j|$, in which case we will consider multiple orderings.)

Let R_i denote the rank of $|D_i|$, i.e., $R_{i_1} = 1$, $R_{i_2} = 2$, etc. Consider two possible test statistics,

$$T_+ = \sum_{D_{i_k} > 0} k = \sum_{D_i > 0} R_i,$$

the sum of the "positive ranks," and

$$T_- = \sum_{D_{i_k} < 0} k = \sum_{D_i < 0} R_i,$$

the sum of the "negative ranks." Because

$$T_+ + T_- = \sum_{k=1}^{n} k = n(n+1)/2,$$

T_- is determined by T_+ (and vice versa), so it suffices to restrict attention to T_+.

If θ_0 is the population center of symmetry, then each D_i is equally likely to be positive or negative and

$$ET_+ = \sum_{i=1}^{n} iP\left(D_i > 0\right) = \sum_{i=1}^{n} i/2 = n(n+1)/4.$$

The Wilcoxon signed rank test rejects $H_0 : \theta = \theta_0$ if and only if we observe T_+ sufficiently different from ET_+.

Let $\vec{x} = \{x_1, \ldots, x_n\}$ denote the observed sample and let $d_i = x_i - \theta_0$ denote the observed differences. For now, we assume that $d_i \neq 0$ and $|d_i| \neq |d_j|$. Let t_+ denote the observed value of T_+. Then the Wilcoxon signed rank test rejects $H_0 : \theta = \theta_0$ if and only if

$$\mathbf{p} = P_{H_0}\left(|T_+ - n(n+1)/4| \geq |t_+ - n(n+1)/4|\right) \leq \alpha.$$

The challenge lies in computing $\mathbf{p}$. To illustrate the nature of this challenge, we consider a simple example.

Example 10.9 We test $H_0 : \theta = 10$ versus $H_1 : \theta \neq 10$ with a significance level of $\alpha = 0.15$. We draw $n = 4$ observations, obtaining a sample of $\{12.4, 11.0, 11.3, 11.7\}$. The corresponding differences are $\{2.4, 1.0, 1.3, 1.7\}$, each of which is positive, and the sum of the positive ranks is $t_+ = 4+1+2+3 = 10$. Because $0 \leq T_+ \leq 10$ and $ET_+ = n(n+1)/4 = 5$, the significance probability is

$$\mathbf{p} = P_{H_0}\left(|T_+ - 5| \geq |10 - 5|\right) = P_{H_0}\left(T_+ = 10\right) + P_{H_0}\left(T_+ = 0\right).$$

To compute $\mathbf{p}$, we require the pmf of the discrete random variable T_+ under the null hypothesis that $\theta_0 = 10$ is the population center of symmetry.

Under $H_0 : \theta = 10$, each D_i is equally likely to be positive or negative; hence, each of the $2^4 = 16$ sign patterns in Table 10.1 are equally likely. The pmf of T_+ is as follows:

k	0	1	2	3	4	5	6	7	8	9	10
$16P(T_+ = k)$	1	1	1	2	2	2	2	2	1	1	1

R_1	R_2	R_3	R_4	T_+
+	+	+	+	10
+	+	+	-	6
+	+	-	+	7
+	+	-	-	3
+	-	+	+	8
+	-	+	-	4
+	-	-	+	5
+	-	-	-	1
-	+	+	+	9
-	+	+	-	5
-	+	-	+	6
-	+	-	-	2
-	-	+	+	7
-	-	+	-	3
-	-	-	+	4
-	-	-	-	0

Table 10.1: Behavior of T_+ under $H_0 : \theta = \theta_0$ with $n = 4$ observations.

Thus,

$$\mathbf{p} = P_{H_0}(T_+ = 10) + P_{H_0}(T_+ = 0) = \frac{1}{16} + \frac{1}{16} = 0.125.$$

Because $\mathbf{p} \leq \alpha$, we reject $H_0 : \theta = 10$.

Although conceptually straightforward, it is evidently cumbersome to compute significance probabilities for the Wilcoxon signed rank test unless n is quite small. For $n = 20$, there are $2^{20} = 1,048,576$ possible sign patterns! Rather than compute exact significance probabilities, we consider two ways of approximating $\mathbf{p}$:

1. Simulation.

   Using `R`, it is easy to generate random sign patterns and compute the value of T_+ for each pattern. The observed proportion of sign patterns for which

   $$|T_+ - n(n+1)/4| \geq |t_+ - n(n+1)/4|$$

   estimates the true significance probability. (The more sign patterns that we generate, the more accurate our estimate of $\mathbf{p}$.) I wrote an `R`

function, `W1.p.sim`, that implements this procedure. This function is described in Appendix R and can be obtained from the web page for this book.

2. Normal Approximation.

   It turns out that, for n sufficiently large, the discrete distribution of T_+ under $H_0 : \theta = \theta_0$ can be approximated by a normal distribution.

   Theorem 10.3 *Suppose that $X_1, \ldots, X_n$ are symmetric continuous random variables with center of symmetry θ_0. Let T_+ denote the test statistic for the Wilcoxon signed rank test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Under H_0, $ET_+ = n(n+1)/4$, $VarT_+ = n(n+1)(2n+1)/24$, and*

$$P_{H_0}(T_+ \leq c) \to P\left(Z \leq \frac{c - ET_+}{\sqrt{VarT_+}}\right)$$

   as $n \to \infty$, where $Z \sim Normal(0, 1)$.

   I wrote an `R` function, `W1.p.norm`, that uses Theorem 10.3 to compute approximate significance probabilities. This function is described in Appendix R and can be obtained from the web page for this book.

Example 10.9 (continued) Five replications of the simulation procedure `W1.p.sim(4,10)` resulted in estimated significance probabilities of 0.117, 0.114, 0.128, 0.124, and 0.125. The normal approximation of 0.100, obtained from `W1.p.norm(4,10)`, is surprisingly good.

Example 10.10 Suppose that $n = 20$ observations produce $t_+ = 50$. Two replications of `W1.p.sim(20,50,10000)` resulted in estimated significance probabilities of 0.0395 and 0.0366. The normal approximation, obtained from `W1.p.norm(20,50)`, is 0.0412.

Finally, we consider the case of samples with $d_i = 0$ and/or $|d_i| = |d_j|$. We will estimate the average significance probability that results from all of the plausible rankings of the $|d_i|$. To do so, we simply modify the simulation procedure described above by subjecting each x_i to small random perturbations. The $|d_i|$ are recalculated on the perturbed samples, each of which results in a unique ordering. By choosing the magnitude of the perturbations sufficiently small, we can guarantee that each complete ordering generated from a perturbed sample is consistent with the partial ordering

derived from the original sample. I wrote an `R` function, `W1.p.ties`, that implements this procedure. This function is described in Appendix R and can be obtained from the web page for this book.

Example 10.11 To test $H_0 : \theta = 5$ versus $H_1 : \theta \neq 5$ at $\alpha = 0.05$, the following x_i were observed:

1.5	9.7	3.9	7.6	8.0	7.3	5.0	9.7	2.3	2.3
6.6	9.4	8.6	7.7	8.4	2.7	9.1	5.3	3.1	9.4

The corresponding $d_i = x_i - 5$ are the following:

−3.5	4.7	−1.1	2.6	3.0	2.3	0.0	4.7	−2.7	−2.7
1.6	4.4	3.6	2.7	3.4	−2.3	4.1	0.3	−1.9	4.4

Ranked by $|d_i|$, we observe the following signed ranks:

$\text{sign}(d_i)$	?		+		−		+		−	
d_i	0.0	<	0.3	<	1.1	<	1.6	<	1.9	<
r_i	1		2		3		4		5	
$\text{sign}(d_i)$	−/+		+		−/−/+		+		+	
d_i	2.3, 2.3	<	2.6	<	2.7, 2.7, 2.7	<	3.0	<	3.4	<
r_i	6/7		8		9/10/11		12		13	
$\text{sign}(d_i)$	−		+		+		+/+		+/+	
d_i	3.5	<	3.6	<	4.1	<	4.4, 4.4	<	4.7, 4.7	
r_i	14		15		16		17/18		19/20	

The value of t_+ depends on

1. whether $r_7 = 1$ is counted as a positive or a negative rank;
2. whether $d_6 = +2.3$ is ranked 6 or 7; and
3. whether $d_{14} = +2.7$ is ranked 9, 10, or 11.

Thus, t_+ might be as small as

$$\begin{aligned} t_+ &= 2 + 4 + 6 + 8 + 9 + 12 + 13 + 15 + 16 + 17 + 18 + 19 + 20 \\ &= 20 \cdot 21/2 - (1 + 3 + 5 + 7 + 10 + 11 + 14) = 159 \end{aligned}$$

or as large as

$$\begin{aligned} t_+ &= 1 + 2 + 4 + 7 + 8 + 11 + 12 + 13 + 15 + 16 + 17 + 18 + 19 + 20 \\ &= 20 \cdot 21/2 - (3 + 5 + 6 + 9 + 10 + 14) = 163. \end{aligned}$$

We might use `W1.p.sim` to estimate, or `W1.p.norm` to approximate, the significance probabilities associated with $t_+ = 159$ and $t_+ = 163$. Alternatively, we might use `W1.p.ties` to estimate the average significance probability associated with all of the possible ways of assigning signed ranks to these data. Five replications of `W1.p.ties(x,5)` resulted in estimated **p**-values of 0.035, 0.024, 0.038, 0.033, and 0.020. As each $\mathbf{p} \leq \alpha$, it is reasonable to reject $H_0 : \theta = 5$.

10.3.2 Point Estimation

In Section 10.2.2 we showed that the sample median is the value of θ_0 that the sign test is least inclined to reject as the population median. Now we derive an estimator, $\hat{\theta}$, of the population center of symmetry, θ. We do so by determining the value of θ_0 for which the Wilcoxon signed rank test is least inclined to reject $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta \neq \theta_0$.

Our derivation relies on a clever representation of t_+. First, given θ_0, we order the observations:

$$x_{(1)} < \cdots < x_{(k)} < \theta_0 < x_{(k+1)} < \cdots < x_{(n)}$$

Let $d_j = x_{(j)} - \theta_0$ and let r_j denote the rank of $|d_j|$, so that

$$t_+ = r_{k+1} + \cdots + r_n.$$

For $j = k+1, \ldots, n$,

$$|d_j| = \left|x_{(j)} - \theta_0\right| = x_{(j)} - \theta_0 = d_j.$$

Hence, if $i \leq j$, then

$$d_i = x_{(i)} - \theta_0 \leq x_{(j)} - \theta_0 = d_j = |d_j|$$

and

$$\begin{aligned} |d_i| \leq |d_j| \quad &\text{iff} \quad -\left(x_{(i)} - \theta_0\right) \leq x_{(j)} - \theta_0 \\ &\text{iff} \quad \left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) \geq 0. \end{aligned}$$

It follows that

$$\begin{aligned} r_j &= \#\left\{i : i \leq j,\, |d_i| \leq |d_j|\right\} \\ &= \#\left\{i : i \leq j,\, \left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) \geq 0\right\}. \end{aligned}$$

Finally, notice that, if $i \leq j \leq k$, then

$$\left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) < 0.$$

Hence, we can write the Wilcoxon test statistic, the sum of the positive ranks, as

$$\begin{aligned} t_+ = \sum_{j=k+1}^{n} r_j &= \sum_{j=k+1}^{n} \#\left\{i : i \leq j, \left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) \geq 0\right\} \\ &= \sum_{1}^{n} \#\left\{i : i \leq j, \left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) \geq 0\right\} \\ &= \#\left\{(i,j) : i \leq j, \left(x_{(i)} - \theta_0\right) + \left(x_{(j)} - \theta_0\right) \geq 0\right\} \\ &= \#\left\{(i,j) : i \leq j, (x_i - \theta_0) + (x_j - \theta_0) \geq 0\right\}. \end{aligned}$$

We know that $H_0 : \theta = \theta_0$ is most difficult to reject when θ_0 is such that $t_+ = ET_+ = n(n+1)/4$. There are $n(n+1)/2$ (i,j) pairs, so the Wilcoxon signed rank test will be least inclined to reject θ_0 when $(x_i - \theta_0) + (x_j - \theta_0)$ is positive for half the pairs and negative for the other half. This condition is equivalent to the condition that $2\theta_0$ is the median of the pairwise sums, $x_i + x_j$, or that θ_0 is the median of the pairwise averages, $(x_i + x_j)/2$.

The pairwise averages, $(x_i + x_j)/2$ for $1 \leq i \leq j \leq n$, are called the *Walsh averages*. The median of the Walsh averages is the estimator, $\hat{\theta}$, of the population center of symmetry, θ, that corresponds to the Wilcoxon signed rank test. This estimator is an excellent compromise between the sample mean and the sample median. For any symmetric distribution with finite variance, the asymptotic relative efficiency (ARE) of $\hat{\theta}$ to $\bar{X}$ is guaranteed to be at least 0.864. For normal distributions, the ARE is $3/\pi \doteq 0.955$, so using $\hat{\theta}$ instead of $\bar{X}$ entails only a slight loss of efficiency. For some symmetric distributions, the ARE is considerably greater than 1.

Although cumbersome to calculate by hand, the median of the Walsh averages is easily computed in `R`. I wrote a function, `W1.walsh`, that does so. This function is described in Appendix R and can be obtained from the web page for this book.

Example 10.11 (continued) To estimate the center of symmetry, we compute the median of the Walsh averages.

```
> W1.walsh(x)
[1] 6.3
```

10.3.3 Set Estimation

In Section 10.2.3 we constructed confidence intervals for the population median by including θ_0 if and only if the sign test failed to reject θ_0 as a plausible population median. We follow the same reasoning here. Let θ denote the population center of symmetry. We use the Wilcoxon signed rank test to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ at significance level α. The set of θ_0 that are not rejected are the plausible populations centers of symmetry, a $(1 - \alpha)$-level confidence interval for θ.

As in the preceding section, we represent the Wilcoxon test statistic, T_+, as the number of Walsh averages that exceed θ_0. Because the Wilcoxon signed rank test rejects $H_0 : \theta = \theta_0$ if and only if T_+ is either sufficiently large or sufficiently small, we deem θ_0 a plausible value of θ if and only if there are sufficient numbers of Walsh averages below and above θ_0. Thus, the desired confidence interval for θ must consist of those θ_0 for which at least k Walsh averages are $\leq \theta_0$ and at least k Walsh averages are $\geq \theta_0$. The quantity k is determined by the level of confidence that is desired. As with the confidence intervals for population medians that we derived from the sign test, not all confidence levels are possible.

The fact that we must approximate the distribution of the discrete random variable T_+ under $H_0 : \theta = \theta_0$ complicates our efforts to construct confidence intervals. We proceed in three steps:

1. Use the normal approximation to guess a reasonable value of k, i.e., a value for which $P(T_+ \leq k - 1) \approx \alpha/2$. Recall that

$$\begin{aligned} P\left(T_+ \leq k - 1\right) &= P\left(T_+ \leq k - 0.5\right) \\ &= P\left(\frac{T_+ - ET_+}{\sqrt{\operatorname{Var} T_+}} \leq \frac{k - 0.5 - ET_+}{\sqrt{\operatorname{Var} T_+}}\right) \\ &\approx P\left(Z \leq \frac{k - 0.5 - ET_+}{\sqrt{\operatorname{Var} T_+}}\right), \end{aligned}$$

   where $Z \sim \text{Normal}(0, 1)$; hence, given $\alpha \in (0, 1)$, a reasonable value of k is obtained by solving

$$P\left(Z \leq \frac{k - 0.5 - ET_+}{\sqrt{\operatorname{Var} T_+}}\right) = \frac{\alpha}{2},$$

   resulting in

$$k = 0.5 + ET_+ - q_z\sqrt{\operatorname{Var} T_+} = 0.5 + \frac{n(n+1)}{4} - q_z\sqrt{\frac{n(n+1)(2n+1)}{24}},$$

   where $q_z = \texttt{qnorm}(1 - \alpha/2)$.

2. Use simulation to estimate the confidence coefficients, $1 - 2P(T_+ \leq k - 1)$, associated with several reasonable choices of k.

3. Finalize the choice of k (and thereby the confidence coefficient) and construct the corresponding confidence interval. The lower endpoint of the interval is the kth Walsh average; the upper endpoint is the $[n(n+1)/2 + 1 - k]$th Walsh average.

I implemented these steps in the `R` function `W1.ci`, described in Appendix R and available from the web page for this book. This function returns a 5×4 matrix. The first column contains possible choices of k, the second and third columns contain the lower and upper endpoints of the corresponding confidence interval, and the fourth column contains the estimated confidence coefficients.

Example 10.11 (continued) To construct a confidence interval with confidence coefficient $1 - \alpha \approx 0.90$ for θ, the population center of symmetry, I obtained the following results:

```
> W1.ci(x,.1,10000)
      k Lower Upper Coverage
[1,] 59  5.25  8.00   0.9195
[2,] 60  5.30  7.95   0.9158
[3,] 61  5.30  7.85   0.9083
[4,] 62  5.35  7.85   0.8913
[5,] 63  5.35  7.85   0.8851
```

The estimated confidence coefficient for $k = 61$ (which happens to be the value of k produced by the normal approximation) is nearest 0.90, so the desired confidence interval is $(5.30, 7.85)$.

10.4 Case Study: Deficit Unawareness in Alzheimer's Disease

Many clinical descriptions of Alzheimer's disease (AD) include the phenomenon of anosognosia, an unawareness of cognitive deficit. One way to obtain experimental evidence of this phenomenon is to perform a *predicted performance experiment*, the hallmark feature of which is that each AD patient is asked to predict his performance on a cognitive task. The hope is that one can infer deficit unawareness from discrepancies between actual

and predicted performance, but one must first control for other possible explanations of such discrepancies. In this section we describe a predicted performance experiment introduced by McGlynn and Kaszniak (1991), adopt a measure of deficit unawareness proposed by Trosset and Kaszniak (1996), and analyze data reported by Kaszniak, DiTraglia and Trosset (1993).[4]

The most basic predicted performance experiment simply examines the discrepancy between a patient's predicted and actual performance on a cognitive task. In such an experiment, the exerimental unit is an individual patient. Two measurements are made on each patient: the patient's prediction of his performance, denoted *ppp*, and the patient's actual score, denoted *pscor*. Following McGlynn and Kaszniak (1991), we measure the discrepancy between these two quantities by forming their ratio, $ppp/pscor$. As described in Section 7.6, we then transform the ratios to an additive scale, e.g., by computing $\log_2(ppp/pscor)$.

A fatal flaw with the basic experiment described above is that one cannot attribute overpredicted performance, $\log_2(ppp/pscor) > 0$, to deficit unawareness. Overpredicted performance *might* mean that AD patients are not aware of impaired cognitive function, but it might also mean that they tend to underestimate the difficulty of the task. To distinguish these possibilites, one must construct a more elaborate experiment.

Suppose that we ask each subject to predict his own performance (*ppp*) and the performance of his spousal caregiver (*ppc*), then measure the actual performance of both patient (*pscor*) and caregiver (*cscor*). In this experiment, the experimental unit is a patient-caregiver pair and four measurements are made on each pair. One can compare the extent to which the patient overpredicted his own performance and the extent to which the patient overpredicted his caregiver's performance, e.g., by computing

$$\log_2 \left[(ppp/pscor) \div (ppc/cscor) \right].$$

Yet even knowing that the patient overpredicted his own performance while correctly predicting his caregiver's performance is insufficient to infer deficit unawareness. Suppose that $ppp/pscor = 2$ and $ppc/cscor = 1$. These results

[4]S. M. McGlynn and A. W. Kaszniak (1991). When metacognition fails: impaired awareness of deficit in Alzheimer's disease. *Journal of Cognitive Neuroscience*, 3:183–189.

M. W. Trosset and A. W. Kaszniak (1996). Measures of deficit unawareness for predicted performance experiments. *Journal of the International Neuropsychological Society*, 2:315–322.

A. W. Kaszniak, G. DiTraglia and M. W. Trosset (1993). Self-awareness of cognitive deficit in patients with probably Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 15:32. Abstract.

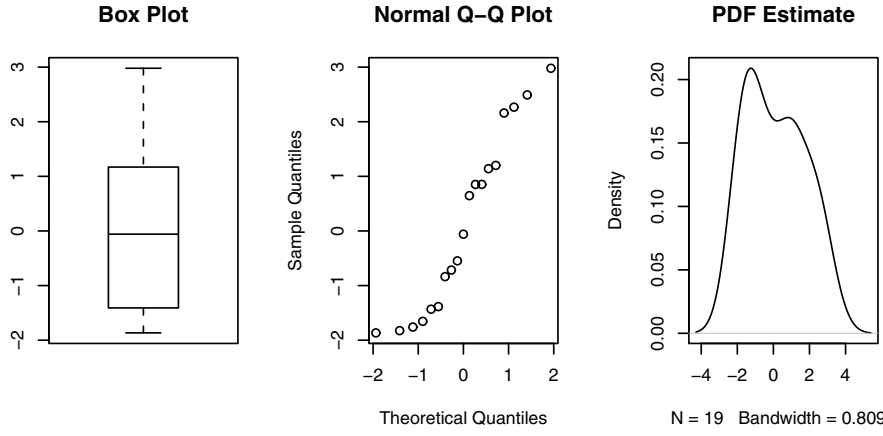

Figure 10.1: Three representations of $n = 19$ values of $\log_2(\text{CPA})$.

might obtain because AD patients are not aware of impaired cognitive function, but they might also obtain because of a general human tendency to overestimate one's own ability. Accordingly, McGlynn and Kaszniak (1991) proposed an even more elaborate predicted performance experiment.

In the McGlynn-Kaszniak experiment, the experimental unit is a patient-caregiver pair. The patient predicts his own performance (*ppp*) and his caregiver's performance (*ppc*). The caregiver predicts her own performance (*cpc*) and the patient's performance (*cpp*). Finally, the actual performance of both patient (*pscor*) and caregiver (*cscor*) is measured. Thus, six measurements are made on each experiment unit.

It is not at all obvious how to combine the six measurements in the McGlynn-Kaszniak experiment to obtain a single measure of deficit unawareness. Trosset and Kaszniak (1996) discussed several flawed possibilities and proposed the Comparative Prediction Accuracy (CPA) measure,

$$\text{CPA} = \frac{(ppp/pscor) \div (ppc/cscor)}{(cpc/cscor) \div (cpp/pscor)}.$$

They envisioned analyzing log(CPA) by performing Student's 1-sample t-test. For ease of interpretation, we replace log(CPA) with $\log_2(\text{CPA})$.

Table 10.2 contains data from a predicted performance experiment reported by Kaszniak, DiTraglia and Trosset (1993). This particular experiment involved $n = 19$ AD patient-caregiver pairs and used a generative

i	PMMSE	*ppp*	*pscor*	*ppc*	*cscor*	*cpc*	*cpp*	CPA	$\log_2$(CPA)
1	26	10	7	10	18	15	5	2.2041	1.1402
2	21	6	21	6	11	10	10	0.2744	−1.8658
3	25	8	10	16	19	6	6	1.8050	0.8520
4	19	5	8	8	17	20	2	0.2822	−1.8251
5	15	5	8	10	17	10	8	1.8062	0.8530
6	19	2	5	20	20	10	6	0.9600	−0.0589
7	18	5	6	35	20	25	5	0.3175	−1.6554
8	23	10	17	15	16	10	5	0.2953	−1.7599
9	18	8	16	18	16	30	25	0.3704	−1.4330
10	20	5	8	20	14	10	5	0.3828	−1.3853
11	13	3	2	30	26	15	7	7.8867	2.9794
12	22	6	7	6	21	8	5	5.6250	2.4919
13	27	12	11	15	26	12	12	4.4694	2.1601
14	10	20	5	30	19	10	5	4.8133	2.2670
15	19	8	13	10	17	10	5	0.6840	−0.5479
16	19	6	13	7	12	12	10	0.6086	−0.7164
17	23	10	13	10	23	8	4	1.5651	0.6462
18	21	6	15	8	12	6	7	0.5600	−0.8365
19	8	12	4	24	14	8	3	2.2969	1.1997

Table 10.2: Data from a predicted performance experiment on $n = 19$ AD patient-caregiver pairs. This experiment used a generative naming task. The quantity PMMSE is the patient's score on the Mini-Mental State Examination, a widely used measure of AD severity (lower scores indicate greater impairment). The quantities *ppp*, *pscor*, *ppc*, *cscor*, *cpc* and *cpp* are the six measures from which CPA (Comparative Prediction Accuracy) is computed.

naming task. The final column contains the values of $\log_2$(CPA), the sample $\vec{x}$ that we will analyze using techniques for 1-sample location problems.

A box plot, a normal probability plot, and a kernel density estimate constructed from $\vec{x}$ are displayed in Figure 10.1. As is so often the case in practice, none of the methods that we have described is entirely satisfying. The tails in the normal probability plot suggest systematic departures from normality, but even the weaker assumption of symmetry appears suspect. On the other hand, the departure from symmetry is not so great that one can be very enthusiastic about forgoing procedures based on the Wilcoxon signed rank test for less powerful procedures based on the sign test. In

such situations, one should do several things. First, one should generate several simulated data sets to obtain a sense of how sampling variation may be affecting one's interpretation of the diagnostic displays in Figure 10.1. Second, one should perform different analyses under different assumptions. If different procedures yield comparable conclusions, then one can only be encouraged that the conclusions are valid. These practices are explored in Exercise Set B in Section 10.5.

It is important to appreciate that reasonable minds may disagree about the best ways to analyze actual data. In practice, one cannot know the true distribution from which the sample was drawn—one can only make reasonable assumptions about it. Even professional statisticians may disagree about what assumptions are most reasonable. For the present $\vec{x}$, my personal preference is to use procedures based on an assumption of symmetry, i.e., procedures derived from the Wilcoxon signed rank test.

Assuming symmetry, let θ denote the population center of symmetry. Because positive values of $\log_2(\text{CPA})$ provide nominal evidence of deficit unawareness, it is natural to test $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$. The expected sum of the positive ranks is $ET_+ = n(n+1)/4 = 95$. The sum of the positive ranks is $t_+ = 103$, computed by the following R commands:

```
> pos <- which(x>0)
> d <- abs(x)
> r <- rank(d)
> tplus <- sum(r[pos])
```

Is $t_+ = 103$ enough larger than $ET_+ = 95$ to reject $H_0 : \theta \leq 0$ in favor of $H_1 : \theta > 0$? The significance probability for testing $H_0 : \theta = 0$ against the two-sided alternative $H_1 : \theta \neq 0$ is approximately 0.77, obtained as follows:

```
> W1.p.sim(19,103,100000)
[1] 0.76859
```

The significance probability for testing $H_0 : \theta \leq 0$ against the one-sided alternative $H_1 : \theta > 0$ is half the two-sided signficance probability, too great to reject $H_0 : \theta \leq 0$ at the conventional significance level of $\alpha = 0.05$.

Finally, to construct a two-sided confidence interval for θ with confidence coefficient $1 - \alpha \approx 0.90$, we obtain the following results:

```
> W1.ci(x,.05,100000)
      k      Lower     Upper Coverage
[1,] 53 -0.5568219 0.7970628  0.91157
[2,] 54 -0.5478818 0.7753275  0.90406
```

```
[3,] 55 -0.5068864 0.7732281  0.89421
[4,] 56 -0.5063870 0.7496210  0.88785
[5,] 57 -0.5045537 0.7491217  0.87826
```

The estimated confidence coefficient for $k = 54$ slightly exceeds 0.90, so a reasonable confidence interval for θ is approximately $(-0.548, 0.775)$. Considering that these numbers are (base 2) logarithms, this is quite a range of values. The corresponding range of plausible CPA values is approximately $(0.684, 1.712)$.

10.5 Exercises

Problem Set A

1. Assume that $n = 400$ observations are independently drawn from a normal distribution with unknown population mean μ and unknown population variance σ^2. The resulting sample, $\vec{x}$, is used to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ at significance level $\alpha = 0.05$.

   (a) What test should be used in this situation? If we observe $\vec{x}$ that results in $\bar{x} = 3.194887$ and $s^2 = 104.0118$, then what is the value of the test statistic?

   (b) If we observe $\vec{x}$ that results in a test statistic value of 1.253067, then which of the following `R` expressions best approximates the significance probability?

      i. `2*pnorm(1.253067)`
      ii. `2*pnorm(-1.253067)`
      iii. `1-pnorm(1.253067)`
      iv. `1-pt(1.253067,df=399)`
      v. `pt(1.253067,df=399)`

   (c) True or False: if we observe $\vec{x}$ that results in a significance probability of $\mathbf{p} = 0.03044555$, then we should reject the null hypothesis.

2. A device counts the number of ions that arrive in a given time interval, unless too many arrive. An experiment that relies on this device produces the following counts, where `Big` means that the count exceeded 255.

251	238	249	Big	243	248	229	Big	235	244
254	251	252	244	230	222	224	246	Big	239

Use these data to construct a confidence interval for the population median number of ions with a confidence coefficient of approximately 0.95.

Problem Set B The following exercises elaborate on the case study explicated in Section 10.4. The sample, $\vec{x}$, is the final column in Table 10.2.

1. Refer to Figure 10.1. The tails in the normal probability plot do not appear to be consistent with an assumption of normality, but the lack of linearity might be due to sampling variation. To investigate whether or not this is the case, please do the following:
   (a) Use `rnorm` to generate four samples from a normal distribution, each with $n = 19$ observations.
   (b) Construct a normal probability plot for each simulated sample. Compare these plots to the normal probability plot of $\vec{x}$ in Figure 10.1.
   (c) Compute the ratio of the sample interquartile range to the sample standard deviation for $\vec{x}$ and for each simulated sample.
   (d) Reviewing the available evidence, are you comfortable assuming that $\vec{x}$ was drawn from a normal distribution? Do four simulated samples provide enough information to answer the preceding question?
2. Assuming that $X_1, \ldots, X_{19} \sim \text{Normal}(\mu, \sigma^2)$, use Student's 1-sample test to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ with a significance level of $\alpha = 0.05$. Construct a two-sided confidence interval for μ with a confidence coefficient of $1 - \alpha = 0.90$. Compare the results of these analyses to the results obtained in Section 10.4.
3. Assuming only that $X_1, \ldots, X_{19}$ are continuous random variables with population median θ, use the sign test to test $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ with a significance level of $\alpha \approx 0.05$. Construct a two-sided confidence interval for θ with a confidence coefficient of $1 - \alpha \approx 0.90$. Compare the results of these analyses to the results obtained in Section 10.4.

Problem Set C Table 10.3 displays a famous data set studied by Charles Darwin.[5] These data appear as Data Set 3 in *A Handbook of Small Data*

[5] C. Darwin (1876). *The Effect of Cross- and Self-Fertilization in the Vegetable Kingdom*, Second Edition. John Murray, London.

Sets, accompanied by the following description:

> Pairs of seedlings of the same age, one produced by cross-fertilization and the other by self-fertilization, were grown together so that the members of each pair were reared under nearly identical conditions. The aim was to demonstrate the greater vigour of the cross-fertilized plants. The data are the final heights [in inches] of each plant after a fixed period of time. Darwin consulted [Francis] Galton about the analysis of these data, and they were discussed further in [Ronald] Fisher's *Design of Experiments*.

Pair	Fertilized	
	Cross	Self
1	23.5	17.4
2	12.0	20.4
3	21.0	20.0
4	22.0	20.0
5	19.1	18.4
6	21.5	18.6
7	22.1	18.6
8	20.4	15.3
9	18.3	16.5
10	21.6	18.0
11	23.3	16.3
12	21.0	18.0
13	22.1	12.8
14	23.0	15.5
15	12.0	18.0

Table 10.3: Darwin's data on the heights (in inches) of cross- and self-fertilized seedlings.

1. Show that this problem can be formulated as a 1-sample location problem. To do so, you should:

   (a) Identify the experimental units and the measurement(s) taken on each unit.

(b) Define appropriate random variables $X_1, \ldots, X_n \sim P$. Remember that the statistical procedures that we will employ assume that these random variables are independent and identically distributed.

(c) Let θ denote the location parameter (measure of centrality) of interest. Depending on which statistical procedure we decide to use, either $\theta = EX_i = \mu$ or $\theta = q_2(X_i)$. State appropriate null and alternative hypotheses about θ.

2. Does it seem reasonable to assume that the sample $\vec{x} = (x_1, \ldots, x_n)$, the observed values of $X_1, \ldots, X_n$, were drawn from:

   (a) a normal distribution? Why or why not?

   (b) a symmetric distribution? Why or why not?

3. Assume that $X_1, \ldots, X_n$ are normally distributed and let $\theta = EX_i = \mu$.

   (a) Test the null hypothesis derived above using Student's 1-sample t-test. What is the significance probability? If we adopt a significance level of $\alpha = 0.05$, should we reject the null hypothesis?

   (b) Construct a (2-sided) confidence interval for θ with a confidence coefficient of approximately 0.90.

4. Now we drop the assumption of normality. Assume that $X_1, \ldots, X_n$ are symmetric (but not necessarily normal), continuous random variables and let θ denote the center of symmetry.

   (a) Test the null hypothesis derived above using the Wilcoxon signed rank test. What is the significance probability? If we adopt a significance level of $\alpha = 0.05$, should we reject the null hypothesis?

   (b) Estimate θ by computing the median of the Walsh averages.

   (c) Construct a (2-sided) confidence interval for θ with a confidence coefficient of approximately 0.90.

5. Finally we drop the assumption of symmetry, assuming only that $X_1, \ldots, X_n$ are continuous random variables, and let $\theta = q_2(X_i)$.

   (a) Test the null hypothesis derived above using the sign test. What is the significance probability? If we adopt a significance level of $\alpha = 0.05$, should we reject the null hypothesis?

(b) Estimate θ by computing the sample median.

(c) Construct a (2-sided) confidence interval for θ with a confidence coefficient of approximately 0.90.

Problem Set D The ancient Greeks greatly admired rectangles with a height-to-width ratio of

$$1 : \frac{1+\sqrt{5}}{2} = 0.618034.$$

They called this number the "golden ratio" and used it repeatedly in their art and architecture, e.g., in building the Parthenon. Furthermore, golden rectangles are often found in the art of later western cultures.

A cultural anthropologist wondered if the Shoshoni, a native American civilization, also used golden rectangles.[6] The following measurements, which appear as Data Set 150 in *A Handbook of Small Data Sets*, are height-to-width ratios of beaded rectangles used by the Shoshoni in decorating various leather goods.

0.693	0.662	0.690	0.606	0.570
0.749	0.672	0.628	0.609	0.844
0.654	0.615	0.668	0.601	0.576
0.670	0.606	0.611	0.553	0.933

We will analyze the Shoshoni rectangles as a 1-sample location problem.

1. There are two natural scales that we might use in analyzing these data. One possibility is to analyze the ratios themselves; the other is to analyze the (natural) logarithms of the ratios. For which of these possibilities would an assumption of normality seem more plausible? Please justify your answer.

2. Choose the possibility (ratios or logarithms of ratios) for which an assumption of normality seems more plausible. Formulate suitable null and alternative hypotheses for testing the possibility that the Shoshoni were using golden rectangles. Using Student's 1-sample t-test, compute a significance probability for testing these hypotheses. Would you reject or accept the null hypothesis using a significance level of 0.05?

[6] *Lowie's Selected Papers in Anthropology*, edited by C. Dubois, University of California Press, Berkeley, 1970.

3. Suppose that we are unwilling to assume that either the ratios or the log-ratios were drawn from a normal distribution. Use the sign test to construct a 0.90-level confidence interval for the population median of the ratios.

Problem Set E Developed in 1975, the drug captopril is used to treat hypertension. Table 10.4 displays data on the effect of captopril on blood pressure. Researchers measured the supine systolic and diastolic blood pressures of $n = 15$ patients with moderate essential hypertension, immediately before and two hours after administering caprotil.[7]

Patient	Systolic		Diastolic	
	before	after	before	after
1	210	201	130	125
2	169	165	122	121
3	187	166	124	121
4	160	157	104	106
5	167	147	112	101
6	176	145	101	85
7	185	168	121	98
8	206	180	124	105
9	173	147	115	103
10	146	136	102	98
11	174	151	98	90
12	201	168	119	98
13	198	179	106	110
14	148	129	107	103
15	154	131	100	82

Table 10.4: Blood pressures of $n = 15$ patients immediately before and two hours after receiving captopril.

We consider the question of whether or not captopril affects systolic and diastolic blood pressure differently.

[7]G. A. MacGregor, N. D. Markandu, J. E. Roulston, and J. C. Jones (1979). Essential hypertension: effect of an oral inhibitor of angiotension-converting enzyme. *British Medical Journal*, 2:1106–1109. These data appear as Data Set 72 in *A Handbook of Small Data Sets*.

1. Let SB and SA denote before and after systolic blood pressure; let DB and DA denote before and after diastolic blood pressure. There are several random variables that might be of interest:

$$X_i = (SB_i - SA_i) - (DB_i - DA_i) \tag{10.4}$$

$$X_i = \frac{SB_i - SA_i}{SB_i} - \frac{DB_i - DA_i}{DB_i} \tag{10.5}$$

$$X_i = \frac{SB_i - SA_i}{SB_i} \div \frac{DB_i - DA_i}{DB_i} \tag{10.6}$$

$$X_i = \log\left(\frac{SB_i - SA_i}{SB_i} \div \frac{DB_i - DA_i}{DB_i}\right) \tag{10.7}$$

   Suggest rationales for considering each of these possibilities.

2. Which (if any) of the above random variables appear to be normally distributed? Which appear to be symmetrically distributed? Identify the variable that you will use in subsequent analyses.

3. Choose a suitable measure of centrality, $\theta = EX_i = \mu$ or $\theta = q_2(X_i)$, for subseqient analysis. What value of θ corresponds to the possibility that captopril affects systolic and diastolic blood pressure equally?

4. Test the null hypothesis that captopril affects systolic and diastolic blood pressure equally. Compute a significance probability. Should the the null hypothesis be rejected at a significance level of $\alpha = 0.05$?

5. Construct a confidence interval for θ that has a confidence coefficient of (approximately) $1 - \alpha = 0.90$.