We compute $\overline{X} = 39.01133$. We do not need the value of s, because we know that $\sigma = 0.010$. Since the population is normal, $\overline{X}$ is normal even though the sample size is small. The null distribution is therefore

$$\overline{X} \sim N(39.00,\ 0.010^2)$$

The z-score is

$$z = \frac{39.01133 - 39.000}{0.010/\sqrt{6}} = 2.78$$

The P-value is 0.0054, so H_0 can be rejected.

Summary

Let $X_1, \ldots, X_n$ be a sample from a *normal* population with mean μ and standard deviation σ, where σ is unknown.

To test a null hypothesis of the form $H_0: \mu \leq \mu_0$, $H_0: \mu \geq \mu_0$, or $H_0:\ \mu = \mu_0$:

- Compute the test statistic $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with $n - 1$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	***P*-value**
$H_1: \mu > \mu_0$	Area to the right of t
$H_1: \mu < \mu_0$	Area to the left of t
$H_1: \mu \neq \mu_0$	Sum of the areas in the tails cut off by t and $-t$

- If σ is known, the test statistic is $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$, and a z test should be performed.

Exercises for Section 6.4

1. Each of the following hypothetical data sets represents some repeated weighings of a standard weight that is known to have a mass of 100 g. Assume that the readings are a random sample from a population that follows the normal curve. Perform a t test to see whether the scale is properly calibrated, if possible. If impossible, explain why.

a. 100.02, 99.98, 100.03

b. 100.01

2. A geologist is making repeated measurements (in grams) on the mass of a rock. It is not known whether the measurements are a random sample from an approximately normal population. Below are three sets of replicate measurements, listed in the order they were made. For each set of readings, state whether the assumptions necessary for the validity of the t test appear to be met. If the assumptions are not met, explain why.

a. 213.03 212.95 213.04 213.00 212.99
213.01 221.03 213.05

b. 213.05 213.00 212.94 213.09 212.98
213.02 213.06 212.99

c. 212.92 212.95 212.97 213.00 213.01 213.04 213.05 213.06

3. A new centrifugal pump is being considered for an application involving the pumping of ammonia. The specification is that the flow rate be more than 5 gallons per minute (gpm). In an initial study, eight runs were made. The average flow rate was 6.5 gpm and the standard deviation was 1.9 gpm. If the mean flow rate is found to meet the specification, the pump will be put into service.

a. State the appropriate null and alternate hypotheses.

b. Find the P-value.

c. Should the pump be put into service? Explain.

4. A certain manufactured product is supposed to contain 23% potassium by weight. A sample of 10 specimens of this product had an average percentage of 23.2 with a standard deviation of 0.2. If the mean percentage is found to differ from 23, the manufacturing process will be recalibrated.

a. State the appropriate null and alternate hypotheses.

b. Compute the P-value.

c. Should the process be recalibrated? Explain.

5. The article "Influence of Penetration Rate on Penetrometer Resistance" (G. Gagnon and J. Doubrough, *Canadian Journal of Civil Engineering*, 2011:741–750) describes a study in which twenty 2-L specimens of water were drawn from a public works building in Bridgewater, Nova Scotia. The mean lead concentration was 6.7 μg/L with a standard deviation of 3.9 μg/L.

a. The Health Canada guideline states that the concentration should be less than 10 μg/L. Can you conclude that the water in this system meets the guideline?

b. A stricter guideline is being considered, which would require the concentration to be less than 7.5 μg/L. Can you conclude that the water in this system meets this guideline?

6. A new process for producing a type of novolac resin is supposed to have a mean cycle time of 3.5 hours per batch. Six batches are produced, and their cycle times, in hours, were

3.45 3.47 3.57 3.52 3.40 3.63

Can you conclude that the mean cycle time is greater than 3.5 hours?

7. Specifications call for the wall thickness of two-liter polycarbonate bottles to average 4.0 mils. A quality control engineer samples 7 two-liter polycarbonate bottles from a large batch and measures the wall thickness (in mils) in each. The results are: 3.999, 4.037, 4.116, 4.063, 3.969, 3.955, and 4.091. It is desired to test $H_0: \mu = 4.0$ versus $H_1: \mu \neq 4.0$.

a. Make a dotplot of the seven values.

b. Should a Student's t test be used to test H_0? If so, perform the test. If not, explain why not.

c. Measurements are taken of the wall thicknesses of seven bottles of a different type. The measurements this time are: 4.004, 4.225, 3.924, 4.052, 3.975, 3.976, and 4.041. Make a dotplot of these values.

d. Should a Student's t test be used to test $H_0: \mu = 4.0$ versus $H_1: \mu \neq 4.0$? If so, perform the test. If not, explain why not.

8. As part of the quality-control program for a catalyst manufacturing line, the raw materials (alumina and a binder) are tested for purity. The process requires that the purity of the alumina be greater than 85%. A random sample from a recent shipment of alumina yielded the following results (in percent):

93.2 87.0 92.1 90.1 87.3 93.6

A hypothesis test will be done to determine whether or not to accept the shipment.

a. State the appropriate null and alternate hypotheses.

b. Compute the P-value.

c. Should the shipment be accepted? Explain.

9. The article "Approximate Methods for Estimating Hysteretic Energy Demand on Plan-Asymmetric Buildings" (M. Rathhore, A. Chowdhury, and S. Ghosh, *Journal of Earthquake Engineering*, 2011: 99–123) presents a method, based on a modal pushover analysis, of estimating the hysteretic energy demand placed on a structure by an earthquake. A sample of 18 measurements had a mean error of 457.8 kNm with a standard deviation of 317.7 kNm. An engineer claims that the method is unbiased, in other words, that the mean error is 0. Can you conclude that this claim is false?

10. Refer to Exercise 12 in Section 5.3. Can you conclude that the mean penetration resistance is greater than 2.5?

11. Refer to Exercise 13 in Section 5.3. Can you conclude that the mercury content is less than 0.3 ppm?

12. The following MINITAB output presents the results of a hypothesis test for a population mean μ.

```
One-Sample T: X

Test of mu = 5.5 vs > 5.5

                                                95%
                                               Lower
Variable   N     Mean     StDev   SE Mean      Bound      T       P
X          5   5.92563  0.15755   0.07046    5.77542   6.04   0.002
```

a. Is this a one-tailed or two-tailed test?
b. What is the null hypothesis?
c. Can H_0 be rejected at the 1% level? How can you tell?
d. Use the output and an appropriate table to compute the P-value for the test of $H_0: \mu \geq 6.5$ versus $H_1: \mu < 6.5$.
e. Use the output and an appropriate table to compute a 99% confidence interval for μ.

13. The following MINITAB output presents the results of a hypothesis test for a population mean μ. Some of the numbers are missing. Fill them in.

```
One-Sample T: X

Test of mu = 16 vs not = 16

Variable    N      Mean    StDev   SE Mean       95% CI         T        P
X          11   13.2874     (a)     1.8389   ( (b), (c) )     (d)    0.171
```

6.5 Large-Sample Tests for the Difference Between Two Means

We now investigate examples in which we wish to determine whether the means of two populations are equal. The data will consist of two samples, one from each population. The basic idea is quite simple. We will compute the difference of the sample means. If the difference is far from 0, we will conclude that the population means are different. If the difference is close to 0, we will conclude that the population means might be the same.

As an example, suppose that a production manager for a manufacturer of industrial machinery is concerned that ball bearings produced in environments with low ambient temperatures may have smaller diameters than those produced under higher temperatures. To investigate this concern, she samples 120 ball bearings that were manufactured early in the morning, before the shop was fully heated, and finds their mean diameter to be 5.068 mm and their standard deviation to be 0.011 mm. She independently samples 65 ball bearings manufactured during the afternoon and finds their mean diameter to be 5.072 mm and their standard deviation to be 0.007 mm. Can she conclude that ball bearings manufactured in the morning have smaller diameters, on average, than ball bearings manufactured in the afternoon?