Exercises for Section 6.11

1. A random sample of size 11 from a normal distribution has variance $s^2 = 96$. Test $H_0: \sigma^2 \leq 50$ versus $H_1: \sigma^2 > 50$.

2. A random sample of size 29 from a normal distribution has variance $s^2 = 24$. Test $H_0: \sigma^2 \geq 30$ versus $H_1: \sigma^2 < 30$.

3. Scores on an IQ test are normally distributed. A sample of 25 IQ scores had variance $s^2 = 64$. The developer of the test claims that the population variance is $\sigma^2 = 225$. Do these data provide sufficient evidence to contradict this claim?

4. A machine that fills beverage cans is supposed to put 12 ounces of beverage in each can. The variance of the amount in each can is 0.01. The machine is moved to a new location. To determine whether the variance has changed, 10 cans are filled. Following are the amounts in the 10 cans. Assume them to be a random sample from a normal population.

12.18	11.77	12.09	12.03	11.87
11.96	12.03	12.36	12.28	11.85

   Perform a hypothesis test to determine whether the variance differs from 0.01. What do you conclude?

5. A sample of 25 one-year-old girls had a mean weight of 24.1 pounds with a standard deviation of 4.3 pounds. Assume that the population of weights is normally distributed. A pediatrician claims that the standard deviation of the weights of one-year-old girls is less than 5 pounds. Do the data provide convincing evidence that the pediatrician's claim is true? (Based on data from the National Health Statistics Reports.)

6. The 2008 General Social Survey asked a large number of people how much time they spent watching TV each day. The mean number of hours was 2.98 with a standard deviation of 2.66. Assume that in a sample of 40 teenagers, the sample standard deviation of daily TV time is 1.9 hours, and that the population of TV watching times is normally distributed. Can you conclude that the population standard deviation of TV watching times for teenagers is less than 2.66?

7. Scores on the math SAT are normally distributed. A sample of 20 SAT scores had standard deviation $s = 87$. Someone says that the scoring system for the SAT is designed so that the population standard deviation will be $\sigma = 100$. Do these data provide sufficient evidence to contradict this claim?

8. One of the ways in which doctors try to determine how long a single dose of pain reliever will provide relief is to measure the drug's half-life, which is the length of time it takes for one-half of the dose to be eliminated from the body. A report of the National Institutes of Health states that the standard deviation of the half-life of the pain reliever oxycodone is $\sigma = 1.43$ hours. Assume that a sample of 25 patients is given the drug, and the sample standard deviation of the half-lives was $s = 1.5$ hours. Assume the population is normally distributed. Can you conclude that the true standard deviation is greater than the value reported by the National Institutes of Health?

9. Find the upper 5% point of $F_{7,20}$.

10. Find the upper 1% point of $F_{2,5}$.

11. An F test with five degrees of freedom in the numerator and seven degrees of freedom in the denominator produced a test statistic whose value was 7.46.
    a. What is the P-value if the test is one-tailed?
    b. What is the P-value if the test is two-tailed?

12. A broth used to manufacture a pharmaceutical product has its sugar content, in mg/mL, measured several times on each of three successive days.

Day 1:	5.0	4.8	5.1	5.1	4.8	5.1	4.8
	4.8	5.0	5.2	4.9	4.9	5.0	
Day 2:	5.8	4.7	4.7	4.9	5.1	4.9	5.4
	5.3	5.3	4.8	5.7	5.1	5.7	
Day 3:	6.3	4.7	5.1	5.9	5.1	5.9	4.7
	6.0	5.3	4.9	5.7	5.3	5.6	

    a. Can you conclude that the variability of the process is greater on the second day than on the first day?
    b. Can you conclude that the variability of the process is greater on the third day than on the second day?

13. Refer to Exercise 11 in Section 5.6. Can you conclude that the variance of the sodium content differs between the two brands?

14. Refer to Exercise 13 in Section 5.6. Can you conclude that the time to freeze-up is more variable in the seventh month than in the first month after installation?

6.12 Fixed-Level Testing

Critical Points and Rejection Regions

A hypothesis test measures the plausibility of the null hypothesis by producing a P-value. The smaller the P-value, the less plausible the null. We have pointed out that there is no scientifically valid dividing line between plausibility and implausibility, so it is impossible to specify a "correct" P-value below which we should reject H_0. When possible, it is best simply to report the P-value, and not to make a firm decision whether or not to reject. Sometimes, however, a decision has to be made. For example, if items are sampled from an assembly line to test whether the mean diameter is within tolerance, a decision must be made whether to recalibrate the process. If a sample of parts is drawn from a shipment and checked for defects, a decision must be made whether to accept or to return the shipment. If a decision is going to be made on the basis of a hypothesis test, there is no choice but to pick a cutoff point for the P-value. When this is done, the test is referred to as a **fixed-level** test.

Fixed-level testing is just like the hypothesis testing we have been discussing so far, except that a firm rule is set ahead of time for rejecting the null hypothesis. A value α, where $0 < \alpha < 1$, is chosen. Then the P-value is computed. If $P \le \alpha$, the null hypothesis is rejected and the alternate hypothesis is taken as truth. If $P > \alpha$, then the null hypothesis is considered to be plausible. The value of α is called the **significance level**, or, more simply, the **level**, of the test. Recall from Section 6.2 that if a test results in a P-value less than or equal to α, we say that the null hypothesis is rejected at level α (or $100\alpha\%$), or that the result is statistically significant at level α (or $100\alpha\%$). As we have mentioned, a common choice for α is 0.05.

Summary

To conduct a fixed-level test:

- Choose a number α, where $0 < \alpha < 1$. This is called the significance level, or the level, of the test.
- Compute the P-value in the usual way.
- If $P \le \alpha$, reject H_0. If $P > \alpha$, do not reject H_0.

Example 6.27

Refer to Example 6.1 in Section 6.1. The mean wear in a sample of 45 steel balls was $\overline{X} = 673.2\,\mu$m, and the standard deviation was $s = 14.9\,\mu$m. Let μ denote the population mean wear. A test of $H_0: \mu \ge 675$ versus $H_1: \mu < 675$ yielded a P-value of 0.209. Can we reject H_0 at the 25% level? Can we reject H_0 at the 5% level?