In situations where the *sample* variances are nearly equal, it is tempting to assume that the population variances are nearly equal as well. This assumption is not justified, however, because it is possible for the sample variances to be nearly equal even when the population variances are quite different. Computer packages often offer a choice of assuming variances to be equal or unequal. The best practice is to assume the variances to be unequal unless it is quite certain that they are equal. See the discussion in Section 5.6.

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be samples from *normal* populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

If σ_X and σ_Y are known to be equal, then, to test a null hypothesis of the form $H_0: \mu_X - \mu_Y \le \Delta_0$, $H_0: \mu_X - \mu_Y \ge \Delta_0$, or $H_0: \mu_X - \mu_Y = \Delta_0$:

- Compute $s_p = \sqrt{\dfrac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}}$.
- Compute the test statistic $t = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{s_p\sqrt{1/n_X + 1/n_Y}}$.
- Compute the P-value. The P-value is an area under the Student's t curve with $n_X + n_Y - 2$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1: \mu_X - \mu_Y > \Delta_0$	Area to the right of t
$H_1: \mu_X - \mu_Y < \Delta_0$	Area to the left of t
$H_1: \mu_X - \mu_Y \ne \Delta_0$	Sum of the areas in the tails cut off by t and $-t$

Exercises for Section 6.7

1. A crayon manufacturer is comparing the effects of two kinds of yellow dye on the brittleness of crayons. Dye B is more expensive than dye A, but it is thought that it might produce a stronger crayon. Four crayons are tested with each kind of dye, and the impact strength (in joules) is measured for each. The results are as follows:

Dye A:	1.0	2.0	1.2	3.0
Dye B:	3.0	3.2	2.6	3.4

a. Can you conclude that the mean strength of crayons made with dye B is greater than that of crayons made with dye A?

b. Can you conclude that the mean strength of crayons made with dye B exceeds that of crayons made with dye A by more than 1 J?

2. In a study of the relationship of the shape of a tablet to its dissolution time, 6 disk-shaped ibuprofen tablets and 8 oval-shaped ibuprofen tablets were dissolved in water. The dissolve times, in seconds, were as follows:

Disk:	269.0	249.3	255.2	252.7	247.0	261.6
Oval:	268.8	260.0	273.5	253.9	278.5	289.4
	261.6	280.2				

Can you conclude that the mean dissolve times differ between the two shapes?

3. The article "Influence of Penetration Rate on Penetrometer Resistance" (J. Oliveira, M. Almeida, et al., *Journal of Geotechnical and Geoenvironmental Engineering*, 2011:695–703) presents measures of

penetration resistance, expressed as a multiple of a standard quantity, for a certain fine-grained soil. Fifteen measurements taken at a depth of 1 m had a mean of 2.31 with a standard deviation of 0.89. Fifteen measurements taken at a depth of 2 m had a mean of 2.80 with a standard deviation of 1.10. Can you conclude that the penetration resistance differs between the two depths?

4. The article "Time Series Analysis for Construction Productivity Experiments" (T. Abdelhamid and J. Everett, *Journal of Construction Engineering and Management*, 1999:87–95) presents a study comparing the effectiveness of a video system that allows a crane operator to see the lifting point while operating the crane with the old system in which the operator relies on hand signals from a tagman. Three different lifts, A, B, and C, were studied. Lift A was of little difficulty, lift B was of moderate difficulty, and lift C was of high difficulty. Each lift was performed several times, both with the new video system and with the old tagman system. The time (in seconds) required to perform each lift was recorded. The following tables present the means, standard deviations, and sample sizes.

	Low Difficulty		
	Mean	Standard Deviation	Sample Size
Tagman	47.79	2.19	14
Video	47.15	2.65	40

	Moderate Difficulty		
	Mean	Standard Deviation	Sample Size
Tagman	69.33	6.26	12
Video	58.50	5.59	24

	High Difficulty		
	Mean	Standard Deviation	Sample Size
Tagman	109.71	17.02	17
Video	84.52	13.51	29

a. Can you conclude that the mean time to perform a lift of low difficulty is less when using the video system than when using the tagman system? Explain.

b. Can you conclude that the mean time to perform a lift of moderate difficulty is less when using the video system than when using the tagman system? Explain.

c. Can you conclude that the mean time to perform a lift of high difficulty is less when using the video system than when using the tagman system? Explain.

5. The Mastic tree (*Pistacia lentiscus*) is used in reforestation efforts in southeastern Spain. The article "Nutrient Deprivation Improves Field Performance of Woody Seedlings in a Degraded Semi-arid Shrubland" (R. Trubata, J. Cortina, and A. Vilagrosaa, *Ecological Engineering*, 2011:1164–1173) presents a study that investigated the effect of adding slow-release fertilizer to the usual solution on the growth of trees. Following are the heights, in cm, of 10 trees grown with the usual fertilizer (the control group), and 10 trees grown with the slow-release fertilizer (treatment). These data are consistent with the mean and standard deviation reported in the article. Can you conclude that the mean height of plants grown with slow-release fertilizer is greater than that of plants with the usual fertilizer?

Usual	17.3	22.0	19.5	18.7	19.5
	18.5	18.6	20.3	20.3	20.3
Slow-release	25.2	23.2	25.2	26.2	25.0
	25.5	25.2	24.1	24.8	23.6

6. Two weights, each labeled as weighing 100 g, are each weighed several times on the same scale. The results, in units of μg above 100 g, are as follows:

First weight:	53	88	89	62	39	66
Second weight:	23	39	28	2	49	

Since the same scale was used for both weights, and since both weights are similar, it is reasonable to assume that the variance of the weighing does not depend on the object being weighed. Can you conclude that the weights differ?

7. It is thought that a new process for producing a certain chemical may be cheaper than the currently used process. Each process was run 6 times, and the cost

of producing 100 L of the chemical was determined each time. The results, in dollars, were as follows:

New Process:	51	52	55	53	54	53
Old Process:	50	54	59	56	50	58

Can you conclude that the mean cost of the new method is less than that of the old method?

8. The article "Effects of Aerosol Species on Atmospheric Visibility in Kaohsiung City, Taiwan" (C. Lee, C. Yuan, and J. Chang, *Journal of Air and Waste Management*, 2005:1031–1041) reported that for a sample of 20 days in the winter, the mass ratio of fine to coarse particles averaged 0.51 with a standard deviation of 0.09, and for a sample of 14 days in the spring the mass ratio averaged 0.62 with a standard deviation of 0.09. Let μ_1 represent the mean mass ratio during the winter and let μ_2 represent the mean mass ratio during the summer. It is desired to test $H_0: \mu_2 - \mu_1 = 0$ versus $H_1: \mu_2 - \mu_1 \neq 0$.

a. Someone suggests that since the sample standard deviations are equal, the pooled variance should be used. Do you agree? Explain.

b. Using an appropriate method, perform the test.

9. The article "Wind-Uplift Capacity of Residential Wood Roof-Sheathing Panels Retrofitted with Insulating Foam Adhesive" (P. Datin, D. Prevatt, and W. Pang, *Journal of Architectural Engineering*, 2011:144–154) presents a study of the failure pressures of roof panels. A sample of 15 panels constructed with 8-inch nail spacing on the intermediate framing members had a mean failure pressure of 8.38 kPa with a standard deviation of 0.96 kPa. A sample of 15 panels constructed with 6-inch nail spacing on the intermediate framing members had a mean failure pressure of 9.83 kPa with a standard deviation of 1.02 kPa. Can you conclude that 6-inch spacing provides a higher mean failure pressure?

10. The article "Magma Interaction Processes Inferred from Fe-Ti Oxide Compositions in the Dölek and Sariçiçek Plutons, Eastern Turkey" (O. Karsli, F. Aydin, et al., *Turkish Journal of Earth Sciences*, 2008:297–315) presents chemical compositions (in weight-percent) for several rock specimens. Fourteen specimens (two outliers were removed) of limenite grain had an average iron oxide (Fe_2O_3) content of 9.30 with a standard deviation of 2.71, and seven specimens of limenite lamella had an average iron oxide content of 9.47 with a standard deviation of 2.22. Can you conclude that the mean iron oxide content differs between limenite grain and limenite lamella?

11. The article "Structural Performance of Rounded Dovetail Connections Under Different Loading Conditions" (T. Tannert, H. Prion, and F. Lam, *Can J Civ Eng*, 2007:1600–1605) describes a study of the deformation properties of dovetail joints. In one experiment, 10 rounded dovetail connections and 10 double rounded dovetail connections were loaded until failure. The rounded connections had an average load at failure of 8.27 kN with a standard deviation of 0.62 kN. The double-rounded connections had an average load at failure of 6.11 kN with a standard deviation of 1.31 kN. Can you conclude that the mean load at failure is greater for rounded connections than for double-rounded connections?

12. The article "Variance Reduction Techniques: Experimental Comparison and Analysis for Single Systems" (I. Sabuncuoglu, M. Fadiloglu, and S. Celik, *IIE Transactions*, 2008:538–551) describes a study of the effectiveness of the method of Latin Hypercube Sampling in reducing the variance of estimators of the mean time-in-system for queueing models. For the M/M/1 queueing model, ten replications of the experiment yielded an average reduction of 6.1 with a standard deviation of 4.1. For the serial line model, ten replications yielded an average reduction of 6.6 with a standard deviation of 4.3. Can you conclude that the mean reductions differ between the two models?

13. In an experiment to test the effectiveness of a new sleeping aid, a sample of 12 patients took the new drug, and a sample of 14 patients took a commonly used drug. Of the patients taking the new drug, the average time to fall asleep was 27.3 minutes with a standard deviation of 5.2 minutes, and for the patients taking the commonly used drug the average time was 32.7 minutes with a standard deviation of 4.1 minutes. Can you conclude that the mean time to sleep is less for the new drug?

14. Refer to Exercise 11 in Section 5.6. Can you conclude that the mean sodium content is higher for brand B than for brand A?

15. Refer to Exercise 12 in Section 5.6. Can you conclude that the mean permeability coefficient at 60°C differs from that at 61°C?

16. The following MINITAB output presents the results of a hypothesis test for the difference $\mu_X - \mu_Y$ between two population means.

```
Two-sample T for X vs Y

      N     Mean    StDev    SE Mean
X    10    39.31     8.71      2.8
Y    10    29.12     4.79      1.5

Difference = mu (X) - mu (Y)
Estimate for difference: 10.1974
95% lower bound for difference: 4.6333
T-Test of difference = 0 (vs >): T-Value = 3.25  P-Value = 0.003  DF = 13
```

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?

17. The following MINITAB output presents the results of a hypothesis test for the difference $\mu_X - \mu_Y$ between two population means. Some of the numbers are missing. Fill in the numbers for (a) through (d).

```
Two-sample T for X vs Y

      N     Mean    StDev    SE Mean
X     6    1.755    0.482    (a)
Y    13    3.239     (b)      0.094
Difference = mu (X) - mu (Y)
Estimate for difference: (c)
95% CI for difference: (-1.99996, -0.96791)
T-Test of difference = 0 (vs not =): T-Value = (d)  P-Value = 0.000  DF = 7
```

6.8 Tests with Paired Data

We saw in Section 5.7 that it is sometimes better to design a two-sample experiment so that each item in one sample is paired with an item in the other. In this section, we present a method for testing hypotheses involving the difference between two population means on the basis of such paired data. We begin with an example.

Particulate matter (PM) emissions from automobiles are a serious environmental concern. Eight vehicles were chosen at random from a fleet, and their emissions were measured under both highway driving and stop-and-go driving conditions. The differences (stop-and-go emission − highway emission) were computed as well. The results, in milligrams of particulates per gallon of fuel, were as follows:

	Vehicle							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Stop-and-go	1500	870	1120	1250	3460	1110	1120	880
Highway	941	456	893	1060	3107	1339	1346	644
Difference	559	414	227	190	353	−229	−226	236