21. H_0: $\mu_1 = \mu_2$. H_1: $\mu_1 \neq \mu_2$. Test statistic: $t = 32.773$. Critical values: $t = \pm 2.023$ (Tech: ± 1.994). P-value < 0.01 (Tech: 0.0000). Reject H_0. There is sufficient evidence to warrant rejection of the claim that the two populations have equal means. The difference is highly significant, even though the samples are relatively small.
23. 0.03795 lb $< (\mu_1 - \mu_2) <$ 0.04254 lb (Tech: 0.03786 lb $< (\mu_1 - \mu_2) <$ 0.04263 lb). Because the confidence interval does not include 0, there appears to be a significant difference between the two population means. It appears that the cola in cans of regular Pepsi weighs more than the cola in cans of Diet Pepsi, and that is probably due to the sugar in regular Pepsi that is not in Diet Pepsi.
25. a. The sample data meet the loose requirement of having a normal distribution. H_0: $\mu_1 = \mu_2$. H_1: $\mu_1 > \mu_2$. Test statistic: $t = 1.046$. Critical value: $t = 2.381$ (Tech: 2.382). P-value > 0.10 (Tech: 0.1496). Fail to reject H_0. There is not sufficient evidence to support the claim that men have a higher mean body temperature than women.
    b. $-0.31°\text{F} < (\mu_1 - \mu_2) < 0.79°\text{F}$. The test statistic became larger, the P-value became smaller, and the confidence interval became narrower, so pooling had the effect of attributing more significance to the results.
27. H_0: $\mu_1 = \mu_2$. H_1: $\mu_1 \neq \mu_2$. Test statistic: $t = 15.322$. Critical values: $t = \pm 2.080$. P-value < 0.01 (Tech: 0.0000). Reject H_0. There is sufficient evidence to warrant rejection of the claim that the two populations have the same mean.
29. a. H_0: $\mu_1 = \mu_2$. H_1: $\mu_1 < \mu_2$. Test statistic: $t = -3.002$. Critical value based on 68.9927614 degrees of freedom: $t = -2.381$ (Tech: -2.382). P-value < 0.005 (Tech: 0.0019). Reject H_0. There is sufficient evidence to support the claim that students taking the nonproctored test get a higher mean than those taking the proctored test.
    b. $-25.68 < \mu_1 - \mu_2 < -2.96$ (Tech: $-25.69 < \mu_1 - \mu_2 < -2.95$)

Section 9-4

1. Parts (c) and (e) are true.
3. The test statistic will remain the same. The confidence interval limits will be expressed in the equivalent values of km/L.
5. H_0: $\mu_d = 0$ cm. H_1: $\mu_d > 0$ cm. Test statistic: $t = 0.036$ (rounded). Critical value: $t = 1.692$. P-value > 0.10 (Tech: 0.4859). Fail to reject H_0. There is not sufficient evidence to support the claim that for the population of heights of presidents and their main opponents; the differences have a mean greater than 0 cm (with presidents tending to be taller than their opponents).
7. a. $\bar{d} = -11.6$ years b. $s_d = 17.2$ years
   c. $t = -1.507$ d. $t = \pm 2.776$
9. H_0: $\mu_d = 0$. H_1: $\mu_d \neq 0$. Test statistic: $t = -1.507$. Critical values: $t = \pm 2.776$. P-value > 0.20 (Tech: 0.2063). Fail to reject H_0. There is not sufficient evidence to support the claim that there is a difference between the ages of actresses and actors when they win Oscars.
11. 1.0 min $< \mu_d <$ 12.0 min. Because the confidence interval includes only positive values and does not include 0 min, it appears that the taxi-out times are greater than the corresponding taxi-in times, so there is sufficient evidence to support the claim of the flight operations manager that for flight delays, more of the blame is attributable to taxi-out times at JFK than taxi-in times at LAX.
13. H_0: $\mu_d = 0$. H_1: $\mu_d > 0$. Test statistic: $t = 2.579$. Critical value: $t = 2.015$. P-value < 0.025 (Tech: 0.0247). Reject H_0. There is sufficient evidence to support the claim that among couples, males speak more words in a day than females.
15. $-6.5 < \mu_d < -0.2$. Because the confidence interval does not include 0, it appears that there is sufficient evidence to warrant rejection of the claim that when the 13th day of a month falls on a Friday, the numbers of hospital admissions from motor vehicle crashes are not affected. Hospital admissions do appear to be affected.
17. H_0: $\mu_d = 0$. H_1: $\mu_d < 0$. Test statistic: $t = -1.080$. Critical value: $t = -1.833$. P-value > 0.10 (Tech: 0.1540). Fail to reject H_0. There is not sufficient evidence to support the claim that *Harry Potter and the Half-Blood Prince* did better at the box office. After a few years, the gross amounts from both movies can be identified, and the conclusion can then be judged objectively without using a hypothesis test.
19. $0.69 < \mu_d < 5.56$. Because the confidence interval limits do not contain 0 and they consist of positive values only, it appears that the "before" measurements are greater than the "after" measurements, so hypnotism does appear to be effective in reducing pain.
21. H_0: $\mu_d = 0$. H_1: $\mu_d \neq 0$. Test statistic: $t = -5.553$. Critical values: $t = \pm 1.990$. P-value < 0.01 (Tech: 0.0000). Reject H_0. There is sufficient evidence to support the claim that there is a difference between the ages of actresses and actors when they win Oscars.
23. H_0: $\mu_d = 0$. H_1: $\mu_d < 0$. Test statistic: $t = -1.560$. Critical value of t is between -1.671 and -1.676 (Tech: -1.673). P-value > 0.05 (Tech: 0.0622). Fail to reject H_0. There is not sufficient evidence to support the claim that among couples, males speak fewer words in a day than females.
25. H_0: $\mu_d = 6.8$ kg. H_1: $\mu_d \neq 6.8$ kg. Test statistic: $t = -11.833$. Critical values: $t = \pm 1.994$ (Tech: ± 1.997). P-value < 0.01 (Tech: 0.0000). Reject H_0. There is sufficient evidence to warrant rejection of the claim that $\mu_d = 6.8$ kg. It appears that the "Freshman 15" is a myth, and college freshman might gain some weight, but they do not gain as much as 15 pounds.

Section 9-5

1. a. No.
   b. No.
   c. The two samples have the same standard deviation (or variance).
3. The F test is very sensitive to departures from normality, which means that it works poorly by leading to wrong conclusions when either or both of the populations has a distribution that is not normal. The F test is not robust against sampling methods that do not produce simple random samples. For example, conclusions based on voluntary response samples could easily be wrong.