(b) We are 95% confident that the mean calorie content is at least 275 calories, which exceeds the 252 calories listed on the package. There is compelling evidence that the value listed on the package is wrong.

(c) If we assume that the number of calories is normally distributed with mean equal to 252 calories, then ½ of the packages would have fewer calories than advertised (since the mean and median are equal for normally distributed data). Thus it makes sense that the manufacturer would desire the mean number of calories to exceed the number reported on the label.

6.S.21 (a) False. The confidence interval includes zero, so we are not confident that μ_1 and μ_2 are different.

(b) True. This is what a confidence interval tells us.

(c) False. The confidence interval is used to make an inference about the difference between μ_1 and μ_2; it does not tell us about individual data points (such as the length of hospitalization for a nitric oxide infant).

6.S.22 (a) False. We know that $\bar{y}_1 - \bar{y}_2$ is 6.9. The CI is for $\mu_1 - \mu_2$.

(b) False. This would only be true if we knew that $\mu_1 - \mu_2$ were exactly 6.9; but in that case we wouldn't be constructing a confidence interval to estimate $\mu_1 - \mu_2$.

(c) True. The confidence interval includes zero, so we would retain the null hypothesis of "no effect" if we conducted a hypothesis test.

6.S.23 In order to be 100% confident that our interval includes the parameter we need to include all real numbers in the interval.

CHAPTER 7
Comparison of Two Independent Samples

7.1.1 The randomization test does not provide compelling evidence for a difference in pulse rates between the genders. 120 of 252 or about 48% of the randomizations yielded differences in means that were larger than the difference observed in our samples of five men and women. Thus, our observed difference is quite consistent with the types of differences one might see do to randomness alone.

7.1.2 (a)

	Randomization									
GITH	1	2	3	4	5	6	7	8	9	10
42.3	L	L	L	N	L	L	N	L	N	N
51.5	L	L	N	L	L	N	L	N	L	N
53.7	L	N	L	L	N	L	L	N	N	L
53.1	N	L	L	L	N	N	N	L	L	L
50.7	N	N	N	N	L	L	L	L	L	L
Normal Mean	49.17	48.97	49.70	52.77	48.17	48.90	51.97	48.70	51.77	52.50
Low Mean	51.90	52.20	51.10	46.50	53.40	52.30	47.70	52.60	48.00	46.90
Difference	**-2.73**	**-3.23**	-1.40	**6.27**	**-5.23**	**-3.40**	**4.27**	**-3.90**	**3.77**	**5.60**

- (b) Nine of the ten randomizations (in bold above) yield differences that are at least as large in magnitude as -2.73.

(c) A difference of means as large as -2.73 would occur 9 out of 10 times randomly, thus this difference does not provide good evidence that dietary chromium affects GITH liver enzyme activity.

7.1.3 (a)

	Randomization									
Colonies	1	2	3	4	5	6	7	8	9	10
30	C	C	C	S	C	C	S	C	S	S
36	C	C	S	C	C	S	C	S	C	S
66	C	S	C	C	S	C	C	S	S	C
76	S	C	C	C	S	S	S	C	C	C
27	S	S	S	S	C	C	C	C	C	C
16	S	S	S	S	S	S	S	S	S	S
C mean	44.0	47.3	57.3	59.3	31.0	41.0	43.0	44.3	46.3	56.3
S mean	39.7	36.3	26.3	24.3	52.7	42.7	40.7	39.3	37.3	27.3
Difference	**4.3**	**11.0**	**31.0**	**35.0**	**-21.7**	-1.7	2.3	**5.0**	**9.0**	**29.0**

	Randomization									
Colonies	11	12	13	14	15	16	17	18	19	20
30	C	C	S	C	S	S	C	S	S	S
36	C	S	C	S	C	S	S	C	S	S
66	S	C	C	S	S	C	S	S	C	S
76	S	S	S	C	C	C	S	S	S	C
27	S	S	S	S	S	S	C	C	C	C
16	C	C	C	C	C	C	C	C	C	C
C mean	27.3	37.3	39.3	40.7	42.7	52.7	24.3	26.3	36.3	39.7
S mean	56.3	46.3	44.3	43.0	41.0	31.0	59.3	57.3	47.3	44.0
Difference	**-29.0**	**-9.0**	**-5.0**	-2.3	1.7	**21.7**	**-35.0**	**-31.0**	**-11.0**	**-4.3**

(b) 16 of the 20 randomizations (in bold above) yield differences that are at least as large as 4.3 in magnitude.

(c) A difference as large as 4.3 colonies would occur 80% (16/20) of the time due to chance alone, thus there is no compelling evidence that the soap inhibits *E. coli* growth compared to the control.

• **7.2.1** **(a)** The observed t statistic is

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} = \frac{735 - 854}{38} = -3.13.$$

From Table 4 with df = 4, we find the critical values $t_{0.02} = 2.999$ and $t_{0.01} = 3.747$. Because t_s is between $t_{0.01}$ and $t_{0.02}$, the P-value must be between .02 and .04. Thus, the P-value is bracketed as $0.02 <$ P-value < 0.04.

(b) The observed t statistic is

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} = \frac{5.3 - 5.0}{0.24} = 1.25.$$

From Table 4 with df = 12, we find the critical values $t_{0.20} = .873$ and $t_{0.10} = 1.356$. Because t_s is between $t_{0.10}$ and $t_{0.20}$, the P-value must be between 0.20 and 0.40. Thus, the P-value is bracketed as $0.20 <$ P-value < 0.40.

(c) The observed t statistic is

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} = \frac{36 - 30}{1.3} = 4.62.$$

From Table 4 with df = 30, we find the critical value $t_{0.0005} = 3.646$. Because t_s is greater than $t_{0.0005}$, the P-value must be less than 0.001. Thus, the P-value is bracketed as P-value < 0.001.

7.2.2 **(a)** $t_s = (100.2 - 106.8)/5.7 = -1.16$.

$t_{10,0.20} = .879$; $t_{10,0.10} = 1.372$. Thus, $0.20 < P < 0.40$.

(b) $t_s = (49.8 - 44.3)/1.9 = 2.89$.

$t_{13,0.01} = 2.650$; $t_{13,0.005} = 3.012$. Thus, $0.01 < P < 0 .02$.

(c) $t_S = (3.58 - 3.00)/.12 = 4.83$.

$t_{19,0.0005} = 3.883$. Thus, P-value < 0.001.

• **7.2.3** **(a)** 0.085 < 0.10, which means that the P-value is less than α. Thus, we reject H_0.

(b) 0.065 >0 .05, which means that the P-value is greater than α. Thus, we do not reject H_0.

(c) Table 4 gives $t_{19,0.005} = 2.861$ and $t_{19,0.0005} = 3.883$, so $0.001 <$ P-value < 0.01. Since $P < \alpha$, we reject H_0.

(d) Table 4 gives $t_{12,0.05} = 1.782$ and $t_{12,0.04} = 1.912$, so $.008 <$ P-value < 0.10. Since $P > \alpha$, we do not reject H_0.

7.2.4 **(a)** Because $P > \alpha$, we do not reject H_0.

(b) Because $P < \alpha$, we reject H_0.

(c) $t_{5,0.04} = 2.191$ and $t_{5,0.03} = 2.422$, so $0.06 < P < 0.08$. Because $P < \alpha$, we reject H_0.

(d) $t_{16,0.04} = 1.869$ and $t_{16,0.03} = 2.024$, so $0.06 < P < 0.08$. Because $P > \alpha$, we do not reject H_0.

Remark concerning tests of hypotheses The answer to a hypothesis testing exercise includes verbal statements of the hypotheses and a verbal statement of the conclusion from the test. In phrasing these statements, we have tried to capture the essence of the biological question being addressed; nevertheless the statements are necessarily oversimplified and they gloss over many issues that in reality might be quite important. For instance, the hypotheses and conclusion may refer to a causal connection between treatment and response; in reality the validity of such a causal interpretation usually depends on a number of factors related to the design of the investigation (such as unbiased allocation of animals to treatment groups) and to the specific experimental procedures (such as the accuracy of assays or measurement techniques). In short, the student should be aware that the verbal statements are intended to clarify the *statistical* concepts; their *biological* content may be open to question.

7.2.5 **(a)** $t_s = 1.07$

(b) H_0:Hereford and Brown Swiss/Hereford cows gain the same amount of weight during a 78-period, on average. H_A: The mean 78-day weight gain of Hereford and Brown Swiss/Hereford cows differs across the breeds.

(c) We retain H_0 because the P-value is larger than 0.10. We have insufficient evidence to say that HH and SH have different mean weight gains.

7.2.6 $\bar{y}_1 - \bar{y}_2 = 3.49 - 3.05 = .44$.

(a) $SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{0.4^2}{5}+\frac{0.4^2}{5}} = 0.2530$

$t_s = 0.44/0.2530 = 1.74.$

Table 4 gives $t_{8,0.10} = 1.397$ and $t_{8,0.05} = 1.860$, so $0.10 < \text{P-value} < 0.20$.

(b) $SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{0.4^2}{10}+\frac{0.4^2}{10}} = 0.1789.$

$t_s = 0.44/0.1789 = 2.46.$

Table 4 gives $t_{18,0.02} = 2.214$ and $t_{18,0.01} = 2.552$, so $0.02 < \text{P-value} < 0.04$.

(c) $SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{0.4^2}{15}+\frac{0.4^2}{15}} = 0.1461.$

$t_s = 0.44/.1461 = 3.01.$

Table 4 gives $t_{28,0.005} = 2.763$ and $t_{28,0.0005} = 3.674$, so $0.001 < \text{P-value} < 0.01$.

• **7.2.7 (a)** The null and alternative hypotheses are

H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$

where 1 denotes heart disease and 2 denotes control. These hypotheses may be stated as

H_0: Mean serotonin concentration is the same in heart patients and in controls
H_A: Mean serotonin concentration is not the same in heart patients and in controls

The test statistic is

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{(\bar{y}_1-\bar{y}_2)}} = \frac{3840-5310}{1064} = -1.38.$$

From Table 4 with df = 14, we find the critical values $t_{0.10} = 1.345$ and $t_{0.05} = 1.761$. Thus, the P-value is bracketed as $0.10 < \text{P-value} < 0.20$.
Since the P-value is greater than α (0.05), H_0 is not rejected. There is insufficient evidence ($0.10 < P < 0.20$) to conclude that serotonin levels are different in heart patients than in controls.

(b) $SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{850^2 + 640^2} = 1064.$

7.2.8 (a) H_0: mean tibia length does not depend on gender ($\mu_1 = \mu_2$)

H_A: mean tibia length depends on gender ($\mu_1 \neq \mu_2$)

$$SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{2.87^2}{60}+\frac{3.52^2}{50}} = 0.62056.$$

$t_s = (78.42 - 80.44)/0.62056 = -3.26$. $df = n_1 + n_2 - 2 = 108 \approx 100$. (Formula (6.7.1) gives df = 94.3.) Table 4 gives $t_{0.005} = 2.626$ and $t_{0.0005} = 3.390$; thus $0.001 < \text{P-value} < 0.01$, so we reject H_0. There is sufficient evidence ($0.001 < P < 0.01$) to conclude that mean tibia length is larger in females than in males.

(b) Judging from the means and SDs, the two distributions overlap substantially, so tibia length would be a poor predictor of sex.

(c) H_0 and H_A are as is part (a).

$$SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{2.87^2}{6}+\frac{3.52^2}{5}} = 1.962$$

$t_s = (78.42 - 80.44)/1.962 = -1.03$. $df = n_1 + n_2 - 2 = 9$. (Formula (6.7.1) gives df = 7.8.) Table 4 gives $t_{0.20} = 0.883$ and $t_{0.10} = 1.383$; thus $0.20 < \text{P-value} < 0.40$ and we do not reject H_0.

7.2.9 (a) H_0: $\mu_1=\mu_2$
(b) The dotplots provide visual evidence that H_0 is false.
(c) The tiny P-value means that if the two groups really were the same then it would be extremely unlikely that the sample mean MBF values would differ by as much as they did in this experiment.
(d) We reject H_0 in favor of H_A. There is strong evidence that mean MBF is higher for the hypoxia condition than for the Normoxia condition.

7.2.10 (a) H_0: mean thymus weight is the same at 14 and 15 days ($\mu_1 = \mu_2$)
H_A: mean thymus weight is not the same at 14 and 15 days ($\mu_1 \neq \mu_2$)

$$SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{8.73^2}{5}+\frac{7.19^2}{5}} = 5.06$$

$t_s = (31.72 - 29.22)/5.06 = 0.49$. $df = n_1 + n_2 - 2 = 8$. (Formula (6.7.1) gives df = 7.7.) Table 4 gives $t_{0.20} = 0.889$; thus P –value > 0.40, so we do not reject H_0. There is insufficient evidence ($P > 0.40$) to conclude that mean thymus weight is different at 14 and 15 days.

(b) According to the P-value found in part (a), the fact that $\bar{y}_1$ is greater than $\bar{y}_2$ could easily be attributed to chance.

[Remark: A student has commented that it would not be surprising if μ_1 were actually less than μ_2, because in certain cases the thymus gland would be expected to shrink during embryonic development.]

• **7.2.11** The null and alternative hypotheses are

H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$

where 1 denotes flooded and 2 denotes control. These hypotheses may be stated as

H_0: Flooding has no effect on ATP
H_A: Flooding has some effect on ATP

The standard error of the difference is

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{0.184^2}{4}+\frac{0.241^2}{4}} = 0.1516.$$

The test statistic is

$$t_s = \frac{\bar{y}_1-\bar{y}_2}{SE_{(\bar{Y}_1-\bar{Y}_2)}} = \frac{1.190-1.785}{0.1516} = -3.92.$$

From Table 4 with df = $n_1 + n_2 - 2 = 6$ (Formula (6.7.1) yields df = 5.6), we find the critical values $t_{0.005} = 3.707$ and $t_{0.0005} = 5.959$. Thus, the P-value is bracketed as 0.001 < P-value < 0.01.

Since the P-value is less than α (0.05), we reject H_0. There is sufficient evidence (0.001 < P < 0.01) to conclude that flooding tends to lower ATP in birch seedlings.

7.2.12 (a) $t_s = 3.73$

(b) We reject H₀ because the P-value is smaller than 0.01. We have strong evidence that average increase in plasma volume is greater for albumin than for polygelatin.

7.2.13 H_0: Mean fall in cholesterol is the same on both diets ($\mu_1 = \mu_2$)
H_A: Mean fall in cholesterol is not the same on both diets ($\mu_1 \neq \mu_2$)

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{31.1^2}{10}+\frac{29.4^2}{10}} = 13.53.$$

$t_s = (53.6 - 55.5)/13.53 = -0.14$.

df = $n_1 + n_2 - 2 = 18$ and $t_{18,0.20} = 0.862$. Thus, P-value > 0.40, so we do not reject H_0. There is insufficient evidence (P > 0.40) to conclude that the two diets differ in their effects on cholesterol.

7.2.14 (a) True. We would reject H_0 because the P-value is less than α.

(b) True. We have significant evidence for H_A because the P-value is less than α.

(c) True. We would reject H_0 because the P-value is less than α.

(d) False. The P-value = 0.03 which is less than α=0.10, α, thus there is significant evidence for H_A

(e) True. This follows directly from the definition of a P-value.

(f) False. H_0 is either true or it is not, unfortunately we do not know which is the case. There is no probability associated with the "truthiness" of H_0 because there is no randomness in the statement. Contrast this to the statement: There is a three percent chance of observing a difference in sample means as large as the one observed, when H_0 is true. Sample means are random…they vary from sample to sample.

7.2.15 (a) True. We would reject H_0 because the P-value is less than α.

(b) True. We have significant evidence for H_A because the P-value is less than α.

(c) False. We do not reject H_0 because the P-value is greater than α.

(d) True. We do not have significant evidence for H_A because the P-value is not less than α.

(e) False. The P-value is the probability, under H_0, of getting a result as extreme as, or more extreme than, the result that was actually observed.

7.2.16 (a) H_0: The mean number of bacteria colonies after 24 hours is the same for ordinary soap solution as it is for sterile water solution.

(b) We retain H₀ because the P-value is larger than 0.10. We have insufficient evidence to say that two solutions differ.

7.2.17 H_0: mean height is the same for control and fertilized plants ($\mu_1 = \mu_2$)
H_A: mean height is not the same for control and fertilized plants ($\mu_1 \neq \mu_2$)

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{0.65^2}{28}+\frac{0.72^2}{28}} = 0.183$$

$t_s = (2.58 - 2.04)/0.183 = 2.95$. Using Table 4 with df=50 we have $t_{0.005} = 2.678$ and $t_{0.0005} = 3.496$. Thus 0.001 < P-value < 0.01, so we reject H_0. There is strong evidence (0.001 < P < 0.01) to conclude that the mean height of fertilized radish sprouts is less than that of controls.

7.2.18 (a) $t_s = -4.497$

(b) H_0: Redbacked and leadbacked salamandars have the same average tail length. H_A: The mean tail lengths of redbacked and leadbacked salamanders differ.

(c) We reject H₀ because the P-value is smaller than 0.05. We have sufficient evidence to say that redbacked and leadbacked salamanders have different mean tail lengths.