7.7.10 $n = 13$

7.7.11 $|\mu_1 - \mu_2|/\sigma = 1.2$

7.7.12 The probability of a Type II error is 7%. If taking folic acid is truly associated with the risk of autism, then a study of this size has a 93% chance of finding that out (by finding significant evidence for HA when alpha is 0.05).

7.8.1 **(a)** Because of the small sample sizes, we require that the data comes from a normal population for our t-test to be valid (Condition 2). Unfortunately, the presence of outliers in each group suggest that the sperm concentrations are not from a normal population.

(b) With large samples (n_1 and $n_2 > 20$), a t-test would be valid even if the data is not from a normal population (Condition 2).

(c) These small Shapiro-Wilk test P-values support our visual assessment of normality: there is significant evidence that the data does not come from a normal population.

(d) A log transformation could make the data appear more normal and thus permit us to use a test that requires normality, such as the t-test.

7.8.2 The sample SDs are:

Heart disease: $s_1 = (850)\sqrt{8} = 2404$

Controls: $s_2 = (640)\sqrt{8} = 2217$

The fact that the mean (3840) for the heart disease patients is only about 1.6 SDs, greater than zero, together with the fact that the observed variable cannot be negative, casts doubt on the normality condition for the population. Because the sample sizes are rather small, such a departure from normality can lead to poor performance by the t test.

7.8.3 The data are not from two independent samples.

7.9.1 **(a)** False; the P-value is the probability of data as unusual as those obtained *if* H_0 is true.

(b) True; reject H_0 since the P-value is less than α.

(c) False. We should reject H_0, but don't know the chance that we would reject H_0 in repeated experiments – we do not know if H_0 is true.

(d) True; this is the interpretation of a P-value.

• **7.10.1** We consult Table 6 with $n = 7$ and $n' = 5$.

(a) $U_s = 26$. From Table 6, the smallest critical value is 27, which corresponds to a nondirectional P-value of 0.149. Since $U_s < 27$, it follows that $P > 0.149$.

(b) $U_s = 30$. From Table 6, the P-value is 0.048.

(c) $U_s = 35$. From Table 6, the P-value is 0.0025.

7.10.2 $n = 8$, $n' = 4$.

(a) $U_s = 25$. From Table 6, the P-value is 0.154.

(b) $U_s = 31$. From Table 6, the P-value is 0.0081.

(c) $U_s = 32$. From Table 6, the P-value is 0.0040.

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

H_0: Toluene has no effect on dopamine in rat striatum
H_A: Toluene has some effect on dopamine in rat striatum

Let 1 denote toluene and let 2 denote control. The ordered arrays of observations are as follows:

Y_1:					1911		2314	2464		2781 2803 3420
Y_2:	1397	1803	1820	1843		1990			2539	

For the K_1 count, we note that there are four Y_2's less than the first Y_1; there are five Y_2's less than the second Y_1; there are five Y_2's less than the third Y_1; and there are six Y_2's less than the fourth, fifth, and sixth Y_1. Thus,

$$K_1 = 4 + 5 + 5 + 6 + 6 + 6 = 32.$$

For the K_2 count, we note that there are no Y_1's less than the first, second, third, or fourth Y_2; there is one Y_1 less than the fifth Y_2; and there are three Y_1's less than the sixth Y_2. Thus,

$$K_2 = 0 + 0 + 0 + 0 + 1 + 3 = 4.$$

To check the counts, we verify that

$$K_1 + K_2 = 32 + 4 = 36 = (6)(6) = (n_1)(n_2).$$

The Wilcoxon-Mann-Whitney test statistic is the larger of the two counts K_1 and K_2; thus $U_s = 32$.

Looking in Table 6 under $n = 6$ and $n' = 6$, we find that for a nondirectional alternative, the entry in the 0.05 column is 31, for which the P-value is 0.041, and the entry in the 0.025 column is 33, for which the P-value is 0.015. Thus, the P-value is bracketed as

$$0.015 < \text{P-value} < 0.041.$$

At significance level $\alpha = 0.05$, we reject H_0, since $P < 0.041 < 0.05$. We note that K_1 is larger than K_2, which indicates a tendency for the Y_1's to be larger than the Y_2's. Thus, there is sufficient evidence ($0.015 < P < 0.041$) to conclude that toluene increases dopamine in rat striatum.

(b) When conducting a nondirectional test, we must check directionality. In this case, we note that K_1 is larger than K_2, which indicates a tendency for the Y_1's to be larger than the Y_2's, which is what the directional alternative predicts. We proceed as in part (a), except that we use the "directional" tail probabilities. Thus, $0.015/2 <$ P-value $< 0.041/2$. We reject H_0 and conclude that there is sufficient evidence ($0.0075 < P < 0.0205$) to conclude that toluene increases dopamine in rat striatum.

7.10.4 Let 1 denote experimental (to be hypnotized) and 2 denote control.

(a) H_0: Ventilation is not differently affected by the "experimental" and the "control" conditions
H_A: Ventilation is differently affected by the "experimental" and the "control" conditions

$K_1 = 53$, $K_2 = 11$, $U_s = 53$. With $n = 8$ and $n' = 8$, Table 6 has entries for $U_s = 51$, for which the P-value is 0.050, and for $U_s = 54$, for which the P-value is 0.021. Thus, $0.021 < P < 0.050$. H_0 is rejected. There is sufficient evidence ($0.021 < P < 0.050$) to conclude that ventilation rate tends to be higher under the "experimental" condition than under the "control" condition.

7.10.5 (a) $U_s = 9$. With $n = n' = 3$, $U_s = 9$ (under the 0.10 heading in Table 6) gives P-value = 0.100.

(b) $U_s = 16$. With $n = n' = 4$, $U_s = 16$ (under the 0.05 heading in Table 6) gives P-value = 0.029.

(c) $U_s = 25$. With $n = n' = 5$, $U_s = 25$ (under the 0.01 heading in Table 6) gives P-value = 0.0079.

7.10.6 (a) H_0: There is no sex difference in preening behavior
H_A: There is a sex difference in preening behavior

For $n = n' = 15$, the largest critical value is 179, for which the P-value is 0.0049. It follows that $P < 0.01$, so H_0 is rejected. There is sufficient evidence ($P < 0.0049$) to conclude that females tend to preen longer than males.

(b) H_0: There is no sex difference in preening behavior ($\mu_1 = \mu_2$)
H_A: There is a sex difference in preening behavior ($\mu_1 \neq \mu_2$)

$t_s = (2.127 - 4.093)/0.7933 = -2.48$. With $df = n_1 + n_2 - 2 = 28$, Table 4 gives $t_{0.01} = 2.467$ and $t_{0.005} = 2.763$, so that $0.01 <$ P-value < 0.02. Formula (6.7.1) yields df = 15.1 and the conservative approach of $df = \min\{n_1 - 1, n_2 - 1\}$ gives df = 14. For either of these df values we get $0.02 < P < 0.04$. In any case, H_0 is not rejected, since $P > 0.01$. There is not sufficient evidence to conclude that there is a sex difference in preening behavior.

(c) Both tests require independent, random samples. The condition required for the t test but not for the Wilcoxon-Mann-Whitney test is that the population distributions are normal. The frequency distribution for the females is highly skewed, due to the two large observations of 10.7 and 11.7. This casts doubt on the normality condition.

(d) $K_1 = 0 + 0 + 0 + 0 + 0 + 0 + .5 + 1 + 1.5 + 1.5 + 2 + 2 + 3.5 + 5 + 8.5 + 10 = 35.5$
$K_2 = 5.5 + 8 + 3(11.5) + 3(13) + 13.5 + 14 + 5(15) = 189.5$

where 1 denotes male and 2 denotes female.

7.10.7 (a) $U_s = 25$

(b) There is sufficient evidence ($P = 0.004$) to conclude that benzo(a)pyrene concentrations tend to be higher in group-housed mice than in singly housed mice.

(c) A directional alternative is valid in this case because the researchers were investigating the hypothesis that licking or biting other mice leads to *increase* benzo(a)pyrene concentration. If access to other mice affects benzo(a)pyrene concentration, the effect would be to increase the concentration; a decrease in concentration is not plausible.

7.10.8 Let 1 denote joggers and let 2 denote fitness program entrants.

H_0: There is no difference in resting blood concentration of HBE between joggers fitness program entrants
H_A: There is a difference in resting blood concentration of HBE between joggers fitness program entrants

$K_1 = 93.5$, $K_2 = 71.5$, $U_s = 93.5$. With $n = 15$ and $n' = 11$, the entry under the 0.20 heading for a nondirectional alternative is 108, with a P-value of 0.198. This is the smallest entry listed, so the next smaller possible value, which is 107, has a P-value greater than 0.20. The data here give $U_s = 93.5$, so the P-value is greater than 0.20 and H_0 is not rejected. There is insufficient evidence ($P > 0.20$) to conclude that there is a difference in resting blood concentration of HBE between joggers fitness program entrants.

7.10.9 (a) Both normal probability plots appear fairly linear which is consistent with the pattern we'd expect to see if the data came from a normal population. Thus, there is no apparent evidence of abnormality in either sample…or more precisely, there is no apparent evidence that the sample data came from abnormal populations.

(b) Even though we don't have evidence for abnormality, we should not mistake this for compelling evidence that the data come from a normal population. We should not conclude that the data are normally distributed.

(c) If the data are normally distributed, using the Wilcoxon-Mann-Whitney test would result in a loss of power. That is, we would have a reduced chance of detecting a significant difference between the mean HBE levels between joggers and fitness participants, should such a difference exist between the two populations.

(d) If the data are not normally distributed, using the two-sample t-test might yield an incorrect P-value. The test might not be valid.

(e) If we had good reason to believe that HBE levels were normally distributed (e.g., evidence from other studies), then using the t-test could be appropriate and would be the most powerful test to use. If we did not have additional information regarding the normality of HBE levels, it would be wiser (i.e., safer) to use the valid, though less powerful, Wilcoxon-Mann-Whitney test.

7.10.10 No. From Table 6 we can see that the P-value is greater than 0.186 (which is the P-value when the test statistic is 111). Thus we do not have evidence for H_A.

7.S.1 **(a)** $SE_1 = 9.6 / \sqrt{12} = 2.771$; $SE_2 = 10.2 / \sqrt{13} = 2.829$.
$\sqrt{2.771^2 + 2.829^2} = 3.96$.

(b) $SE_1 = 2.7 / \sqrt{22} = 0.576$; $SE_2 = 1.9 / \sqrt{19} = 0.436$.
$\sqrt{0.576^2 + 0.436^2} = 0.722$.

(c) $SE_1 = 1.2 / \sqrt{5} = 0.537$; $SE_2 = 1.4 / \sqrt{7} = 0.529$.
$\sqrt{0.537^2 + 0.529^2} = 0.754$.

• **7.S.2** The null and alternative hypotheses are

H_0: Mean platelet calcium is the same in people with high blood pressure as in people with normal blood pressure ($\mu_1 = \mu_2$)
H_A: Mean platelet calcium is different in people with high blood pressure than in people with normal blood pressure ($\mu_1 \neq \mu_2$)

The standard error of the difference is

$$SE_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{16.1^2}{38} + \frac{31.7^2}{45}} = 5.399.$$

The test statistic is

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{(\bar{y}_1 - \bar{y}_2)}} = \frac{168.2 - 107.9}{5.399} = 11.2.$$

From Table 4 with df = 45 + 38 -2 = 81 ≈ 80, we find the critical value $t_{0.0005} = 3.416$. The tail area is doubled for the nondirectional test. Thus, the P-value is bracketed as $P < 0.001$. (Formula (6.7.1) yields df = 67.5, but the P-value is still bracketed as $P < 0.001$.)
Since the P-value is less than α (0.01), we reject H_0. There is sufficient evidence ($P < 0.001$) to conclude that mean platelet calcium is higher in people with high blood pressure than in people with normal blood pressure.

7.S.3 The mean difference in blood pressure is $\mu_1 - \mu_2$, where 1 denotes high blood pressure and 2 denotes normal blood pressure.

The standard error of the difference is

$$SE_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{31.7^2}{45} + \frac{16.1^2}{38}} = 5.399.$$

The critical value $t_{.025}$ is found from Student's t distribution with df given by Formula (6.7.1) as df = 67.5 ≈ 70. Table 4 gives $t_{70,0.025} = 1.994$. The 95% confidence interval is

$\bar{y}_1 - \bar{y}_2 \pm t_{0.25} SE_{(\bar{y}_1 - \bar{y}_2)}$
(168.2 - 107.9) ± (1.994)(5.399).

So the confidence interval is (49.5,71.1) or $49.5 < \mu_1 - \mu_2 < 71.1$ nM.

Alternatively, we could use df = 45 + 38 -2 = 81 ≈ 80, in which case the critical value is $t_{80,0.025} = 1.990$. This gives an interval of

(168.2 - 107.9) ± (1.990)(5.399).

So the confidence interval is (49.6,71.0) or $49.6 < \mu_1 - \mu_2 < 71.0$ nM.

• **7.S.4** No; the t test is valid because the sample sizes are rather large.

7.S.5 **(a)** H_0: Mechanical milking does not produce different cell count than manual milking ($\mu_1 = \mu_2$)
H_A: Mechanical milking produces higher cell count than manual milking ($\mu_1 > \mu_2$)

$t_s = (1215.6 - 219.0)/427.54 = 2.33$. With df = 18, Table 4 gives $t_{0.02} = 2.214$ and $t_{0.01} = 2.552$. Formula (6.7.1) yields df = 9.2 ≈ 9; with df = 9, Table 4 gives $t_{0.025} = 2.262$ and $t_{0.02} = 2.398$. Using either df value, $P < 0.05$ and H_0 is rejected. There is sufficient evidence to conclude that mechanical milking produces higher cell count than manual milking. (The data support the investigator's suspicion.)

(b) H_0: Mechanical milking does not produce different cell count than manual milking
H_A: Mechanical milking produces higher cell count than manual milking

$U_s = 69$. The shift in the data is in the direction predicted by H_A. With n = n' = 10, the non-directional P-value is 0.190 for $U_s = 68$ and 0.089 for $U_s = 73$. Thus, $0.0498 < P < 0.095$, we do not reject H_0. There is insufficient evidence ($0.0498 < P < 0.095$) to conclude that mechanical milking produces higher cell count than manual milking. (The data do not support the investigator's suspicion.) [Note that this contradicts the conclusion from part (a).]

(c) Both tests require independent, random samples. The condition required for the t test but not for the Wilcoxon-Mann-Whitney test is that the population distributions are normal. The frequency distribution for the mechanical group is highly skewed, with some observations (2996 and 3452) that are *much* greater than the others. This casts doubt on the normality condition.

(d) $K_1 = 10 + 7 + 0 + 10 + 10 + 7 + 1 + 10 + 4 + 10 = 69$
$K_2 = 3 + 2 + 1 + 2 + 2 + 3 + 5 + 5 + 5 + 3 = 31$

where 1 denotes mechanical and 2 denotes manual.

7.S.6 (a) $t_s = 3.69$

(b) We reject H_0 because the P-value is smaller than 0.01. We have sufficient evidence to say that stress tends to retard plant growth.

7.S.7 Let 1 denote control and let 2 denote stress.

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{2.13^2}{13} + \frac{1.73^2}{13}} = 0.7611.$$

$(30.59 - 27.78) \pm (2.064)(0.7611)$ (df = 24)

$(1.24, 4.38)$ or $1.24 < \mu_1 - \mu_2 < 4.38$ cm.

(a) The confidence interval indicates that the difference is at least 1 cm and so is "horticulturally important."

(b) According to the confidence interval, the difference could be greater or smaller than 2 cm, so the data do not indicate whether the difference is "horticulturally important."

(c) The confidence interval indicates that the difference is less than 5 cm and so is not "horticulturally important."

• 7.S.8 The null and alternative hypotheses are

H_0: Stress has no effect on growth
H_A: Stress tends to retard growth

The data are already arrayed in increasing order. We let Y_1 denote control and Y_2 denote stress. For the K_1 count, we note that there is one Y_2 less than the first Y_1; there are ten Y_2's less than the second Y_1; there are twelve Y_2's less than the third, fourth, fifth, sixth, and seventh Y_1; there are twelve Y_2's less than the eighth Y_1 and one equal to it; and there are thirteen Y_2's less than the ninth through thirteenth Y_1. Thus,

$$K_1 = 1 + 10 + 12 + 12 + 12 + 12 + 12 + 12.5 + 13 + 13 + 13 + 13 + 13 = 148.5.$$

For the K_2 count, we note that there are no Y_1's less than the first Y_2; there is one Y_1 less than the second through tenth Y_2; there are two Y_1's less than the eleventh and twelfth Y_2; and there are seven Y_1's less than the thirteenth Y_2 and one equal to it. Thus,

$$K_2 = 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 2 + 2 + 7.5 = 20.5.$$

To check the counts, we verify that

$$K_1 + K_2 = 148.5 + 20.5 = 169 = (13)(13) = (n_1)(n_2).$$

To check the directionality of the data, we note that $K_1 > K_2$, which suggests a tendency for the Y_1's to be larger than the Y_2's, which would indicate that stress retards growth. Thus, the data do deviate from H_0 is the direction specified by H_A.

The Wilcoxon-Mann-Whitney test statistic is the larger of the two counts K_1 and K_2; thus $U_s = 148.5$.

Looking in Table 6 under n = 13 and n' = 13, we find that for a directional alternative, the largest entry is 139, for which the non-directional P-value is 0.0042. Thus, the P-value is bracketed as

$P < 0.0021$.
At significance level $\alpha = 0.01$, we reject H_0, since $P < \alpha$. There is sufficient evidence ($P < 0.0021$) to conclude that stress tends to retard growth.

7.S.9 We reject H_0 because the P-value is smaller than 0.10. There is sufficient evidence to conclude that biodiversity is greater along the Vermilion River than along the Black River.

7.S.10 Let 1 denote positive response and let 2 denote no response.

H_0: Ovarian pH is not related to progesterone response
H_A: Ovarian pH is related to progesterone response

$K_1 = 1.5$, $K_2 = 106.5$, $U_s = 106.5$ With n = 18, n' = 6, and a nondirectional alternative, the largest entry is 95, for which the P-value is 0.0044. Thus, the P-value is bracketed as

$P < 0.0044$.
At significance level $\alpha = 0.05$, we reject H_0, since $P < \alpha$. There is sufficient evidence ($P < 0.0044$) to conclude that ovarian pH is lower among responders to progesterone than among nonresponders.

7.S.11 (a) H_0: $\mu_1 = \mu_2$ and H_A: $\mu_1 \neq \mu_2$

(b) The small P-value means that if the two groups really were the same then it would be very unlikely that the sample pH means would differ by as much as it did in this study.

(c) We reject H_0 in favor of H_A. There is strong evidence that mean oocyte pH is lower when there is a positive response to progesterone than when there is no response.

7.S.12 The lack of a statistically significant difference in weight gain does not show that the new diet is as good as the standard diet. (Evidence to this effect could be obtained from either a confidence interval or an analysis of the power of the test; see the next exercise.)

7.S.13 If the standard deviation is about 20% of the mean and the difference between the means is about 10%, then

$$\frac{|\mu_1 - \mu_2|}{\sigma} = \frac{0.10}{0.20} = 0.5.$$

To achieve power of 0.50, Table 5 indicates that 32 animals per group would be required. Since 25 < 32, the power of this study was less than 0.50. Thus, the study did *not* have a reasonably good chance of detecting a population mean deficiency of 10% on the new diet.

7.S.14 **(a)** Two of the patients contributed two observations each to the data set. Thus, there is hierarchical structure, so the t test is not appropriate.

(b) No. The Wilcoxon-Mann-Whitney test, like the t test, requires that the observations within a sample be independent of each other, so the Wilcoxon-Mann-Whitney test is not appropriate.

7.S.15 Let 1 denote low chromium and let 2 denote normal.

H_0: Low chromium diet does not affect GITH ($\mu_1 = \mu_2$)
H_A: Low chromium diet does affect GITH ($\mu_1 \neq \mu_2$)

$\bar{y}_1 = 51.75,\ s_1 = 5.526$

$\bar{y}_2 = 53.17,\ s_1 = 4.123$

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{5.526^2}{14} + \frac{4.123^2}{10}} = 1.970.$$

$t_s = (51.75 - 53.17)/1.970 = -0.72$. Formula (6.7.1) gives df = 21.9; $t_{22,0.20} = 0.858$, so $P > 0.40$. Using a computer, we get P = 0.48. Thus, we do not reject H_0. There is insufficient evidence (P = 0.48) to conclude that low chromium diet affects GITH in rats.

7.S.16 Let 1 denote low chromium and let 2 denote normal.

H_0: Low chromium diet does not affect GITH
H_A: Low chromium diet does affect GITH

$U_s = 70.5$. n = 14, n' = 10, and a nondirectional alternative, the smallest entry is 93, under the 0.20 heading. Thus, $P > 0.20$ and we do not reject H_0. There is insufficient evidence ($P > 0.20$) to conclude that low chromium diet affects GITH in rats.

7.S.17 (See Exercise 7.S.15 for basic computations.)

(a) (51.75 - 53.17) ± (2.074)(1.970) (df = 21.9 ≈ 22)

(-5.5,2.7) or $-5.5 < \mu_1 - \mu_2 < 2.7$ thousand cpm/gm.

(b) All values in the confidence interval are smaller in magnitude than 8 thousand cpm/gm; thus the data support the conclusion that the difference is "unimportant."

(c) The confidence interval indicates that the difference could be larger in magnitude than 4 thousand cpm/gm or smaller; thus the data do not indicate whether the difference is "unimportant."

7.S.18 **(a)** Let 1 denote infected and let 2 denote uninfected.

H_0: Malaria is not related to stamina ($\mu_1 = \mu_2$)
H_A: Malaria is associated with decreased stamina ($\mu_1 < \mu_2$)

$\bar{y}_1 = 26.87,\ s_1 = 6.815;$

$\bar{y}_2 = 32.23,\ s_1 = 8.067.$

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{6.815^2}{15} + \frac{8.067^2}{15}} = 2.727.$$

$t_s = (26.87 - 32.23)/2.727 = -1.97$. Formula (6.7.1) gives df = 27.2; $t_{27,0.03} = 1.963$ and $t_{27,0.025} = 2.052$, so $0.025 < P < 0.03$. Using a computer, we get P = 0.0298. Thus, we reject H_0. There is sufficient evidence (P = 0.0298) to conclude that malaria is associated with decreased stamina.

(b) H_0: Malaria is not related to stamina
H_A: Malaria is associated with decreased stamina

$U_s = 155$; the data deviate from H_0 in the direction specified by H_A. With n = 15, n' = 15, the non-directional P-value is 0.098 for $U_s = 153$ and 0.045 for $U_s = 161$. Thus, $0.0225 < P < 0.049$ and we reject H_0. There is sufficient evidence ($0.0225 < P < 0.049$) to conclude that malaria is associated with decreased stamina.

7.S.19 **(a)** Let 1 denote amphetamine and let 2 denote control.

H_0: Amphetamine is not related to water consumption ($\mu_1 = \mu_2$)
H_A: Amphetamine is associated with decreased water consumption ($\mu_1 < \mu_2$)

$\bar{y}_1 = 129.375,\ s_1 = 27.850$

$\bar{y}_2 = 156.0,\ s_1 = 25.322$

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{27.850^2}{4} + \frac{25.322^2}{4}} = 18.82.$$

$t_s = (129.375 - 156)/18.82 = -1.415$. With $df = n_1 + n_2 - 2 = 6$ (Formula (6.7.1) yields df = 5.9), Table 4 gives $t_{0.20} = 0.906$ and $t_{0.10} = 1.440$, so $0.10 < P < 0.20$. (Using a computer, we get P = 0.104.) Thus, we do not reject H_0. There is insufficient evidence ($0.10 < P < 0.20$) to conclude that amphetamine is associated with decreased water consumption.

(b) H_0: Amphetamine is not related to water consumption
H_A: Amphetamine is associated with decreased water consumption

$K_1 = 4$, $K_2 = 12$, $U_s = 12$; the data deviate from H_0 in the direction specified by H_A. With n = 4, n' = 4, the smallest entry is 13, under the 0.20 heading. For the directional test, we divide the heading in half. Thus, P > 0.10 and we do not reject H_0. There is insufficient evidence (P > 0.10) to conclude that amphetamine is associated with decreased water consumption.

7.S.20 (a) False. The P-value for a test is the probability of getting data at least as extreme as those obtained, if H_0 is true; it is not the probability that the null hypothesis is true.

(b) True. The P-value for a test is the probability of getting data at least as extreme as those obtained, if H_0 is true, which is what this statement says.

(c) False. The probability that H_0 is rejected depends on the power of the test, which is not known. (*If* H_0 is true -- and we don't know if it is true or not -- and a new study is done *that uses* $\alpha = 0.04$, then there is a 4% probability that H_0 will be rejected.)

7.S.21 (a) False. The confidence interval gives a range that we infer covers $\mu_1 - \mu_2$. It does not tells us where the bulk of the data lie.

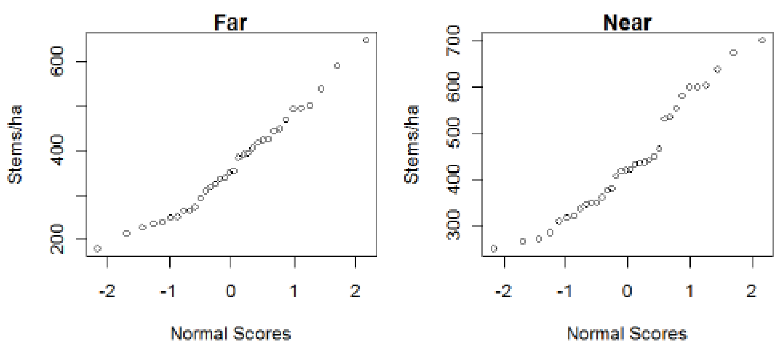
(b) True. This is what a confidence interval tells us.

(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 future values of $\bar{y}_1$ and $\bar{y}_2$. (*If* $\mu_1 - \mu_2$ were exactly equal to the $\bar{y}_1 - \bar{y}_2$ difference obtained in this study -- which is not at all likely -- then the confidence interval could be used to predict where future $\bar{y}_1 - \bar{y}_2$ differences would fall.)

(d) 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 number of calories in a particular entree).

7.S.22 True. (85 -(-27))/2 = 56 is the margin of error for 95% confidence.

7.S.23 (a) The normal probability plots show very mild skewness:

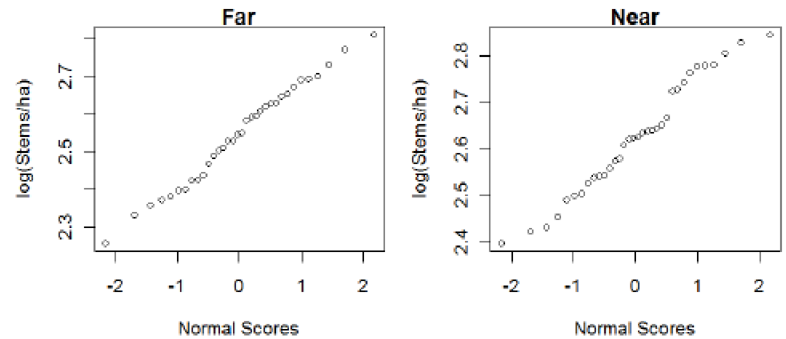


(b) The following computer output shows that the P-value is 0.0198.

```
      Two Sample t-test

data:  Stems by Type
t = -2.3883, df = 65.457, p-value = 0.01982
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -126.74451  -11.31431
sample estimates:
 mean in group far mean in group near
          368.4412           437.4706
```

(c) After taking logarithms of each observation, the normal probability plots look more linear:



The P-value for the t-test is now 0.0156, which is almost the same as the P-value before the transformation:

```
Two Sample t-test

data:  log10(Stems) by Type
t = -2.4832, df = 65.354, p-value = 0.01559
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.14068298 -0.01527030
sample estimates:
 mean in group far mean in group near
         2.546118           2.624095
```

(d) When the sample sizes are fairly large the t-test gives almost exactly the same P-value for data that are midly skewed as for data that are not skewed. That is, the t-test is robust.

7.S.24 (a) Let 1 denote andro and let 2 denote control.

H_0: Andro has no effect ($\mu_1 = \mu_2$)
H_A: Andfo has an effect ($\mu_1 \neq \mu_2$)

$$SE_{(\bar{Y}_1-\bar{Y}_2)} = \sqrt{\frac{13.3^2}{9}+\frac{12.5^2}{10}} = 5.94.$$

$t_s = (14.4 - 20.0)/5.94 = -0.94$. With df =16, Table 4 gives $t_{0.20} = 0.865$ and $t_{0.10} = 1.337$, so $0.20 < P < 0.40$. (Using a computer, we get P = 0.359.) Thus, we do not reject H_0. There is insufficient evidence ($0.20 < P < 0.40$) to conclude that andro affects lat pulldown strength.

(b) H_0: Andro has no effect ($\mu_1 = \mu_2$)
H_A: Andro has a positive effect ($\mu_1 > \mu_2$)

The data deviate from H_0 in the opposite direction from what H_A predicts. Thus, $P > 0.50$ and we do not reject H_0. There is insufficient evidence ($P > 0.50$) to conclude that andro increases lat pulldown strength.

7.S.25 A treatment was given to 244 subjects and the number of drinks for each subject was recorded for 7 days. The treatment subjects were compared to 238 control subjects. The null hypothesis that the treatment has no effect on drink consumption was rejected in a one-sided test. There is strong evidence (P = 0.0031) that the treatment is associated with a reduction in drink consumption. Moreover, we can be 95% confident that on average treatment subjects consume between .92 and 5.56 fewer drinks, per 7 days, than do control subjects.

7.S.26 (a) The explanatory variable is the drug received: AZT or Placebo.

(b) The response variable is the HIV status of babies.

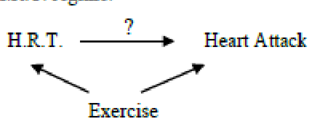
(c) The experimental units are HIV positive mothers.

7.S.27 (a) The explanatory variable is placement position (prone/supine).

(b) The response variable is the living status (dead/alive) of patients.

(c) The experimental units are patients.

7.S.28 As observational data, we cannot infer a causal relationship between H.R.T. and heart attack risk. A very reasonable explanation for the observed association is that women who are concerned with their health are more likely to exercise (and reduce their heart attack risk) as well as follow a H.R.T. regime.



7.S.29 This is a directional test and $\bar{y}_1 < \bar{y}_2$ so we take the P-value from the non-directional test and cut it in half, getting P-value = 0.065. Thus (iv) is correct: The P-value would be less than 0.13 and H_0 would be rejected if $\alpha = 0.10$.

7.S.30 (vi) The choice of α does not depend on n. As n goes down the SE goes up, which means that t goes down since $t = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$. This means that the P-value goes up.

7.S.31 A The 90% confidence interval excludes zero, so a test of H_0: $\mu_1 = \mu_2$ against H_A: $\mu_1 \neq \mu_2$ has a P-value less than 0.10. For testing H_0: $\mu_1 = \mu_2$ against H_A: $\mu_1 < \mu_2$ we note that the data deviate in the right direction (the CI is entirely to the negative side of zero), so the directional P-value is half of the non-directional P-value, so $P < 0.05$, so Miguel rejects H_0.

7.S.32 (a) H_0: $\mu_1=\mu_2$ and H_A: $\mu_1<\mu_2$

(b) We reject H_0 in favor of H_A. There is strong evidence that being given the sugar-free drink leads to less change in BMI.

7.S.33 (a) H_0: $\mu_1=\mu_2$ and H_A: $\mu_1>\mu_2$

(b) We (just barely) retain H_0. There is no sufficient evidence to conclude that wrist extension improves more over six weeks when the intervention is used.

(c) The non-directional P-value would be twice as large as the directional P-value, so if we fail to reject H_0 with a directional test for α =0.10, we will also not reject it for the non-directional test.

7.S.34 The first test says that there is not sufficient evidence to conclude that $\mu_1\neq\mu_2$ and the third test says that there is not sufficient evidence to conclude that $\mu_2\neq\mu_3$. However, "absence of evidence is not evidence of absence." It is quite possible that μ_1 and μ_3 differ, as the second t-test indicates, despite a lack of evidence in the other two tests of a difference.

7.S.35 If H_0 is true then the test statistic has a t distribution with 57 degrees of freedom. That is, it is a symmetric distribution, centered at zero, with SD approximately equal to 1.