CHAPTER 14

Section 14.1

1. We require indices r and s so that $\sum_{k=r}^{s-1} b(k;38,.25)$ is .95 (or just slightly over). Starting at .25(38) = 9.5 and working outward, trial and error finds that $\sum_{k=5}^{14} b(k;38,.25) = .9416$ but $\sum_{k=4}^{14} b(k;38,.25) = .9579$. Thus indices $r = 4$ and $s - 1 = 14$ (aka $s = 15$) meet our requirement. So, a general 95% CI for $\eta_{.25}$ when $n = 38$ is (Y_4, Y_{15}). From the original 38 salaries, the 4th and 15th in increasing order are y_4 = \$55,000 and y_{15} = \$61,000. So, with 95% confidence, the population 25th percentile of civil engineering starting salaries is between \$55,000 and \$61,000.

3. We require indices r and s so that $\sum_{k=r}^{s-1} b(k;40,.5)$ is .95 (or just slightly over). Start at index .5(40) = 20 and work outward using trial and error. Eventually we find $\sum_{k=14}^{26} b(k;40,.5) = .9615$ and all "shorter" sums have probability strictly less than .95. Thus $r = 14$, $s - 1 = 26$ (aka $s = 27$), and the general CI formula for the population median is (Y_{14}, Y_{27}).

5. Let $\tilde{\mu}$ = the true median house price in Houston (in \$1000s). We test H_0: $\tilde{\mu} = 197$ vs H_a: $\tilde{\mu} > 197$ using the one-sample sign test. If H_0 is true, the number of house prices over \$197,000 in a random sample of 25 homes should follow a binomial distribution with $n = 25$ and $p = .5$. In the data provided, 19 house prices exceed \$197,000, so the upper-tailed P-value is

$$P(19 \textit{ or more} \text{ prices} > 197 \text{ when } \tilde{\mu} = 197) = \sum_{k=19}^{25} \binom{25}{k} (.5)^k (1-.5)^{25-k} = 1 - B(18; 25, .5) = .007$$

Since $.007 < \alpha = .05$, H_0 is rejected at the .05 level. The data provide convincing statistical evidence that the true median home price in Houston exceeds Texas' statewide median of \$197,000.

7. Let p = the true proportion of all young children with social/emotional problems whose score would improve (i.e., score change < 0) after the physical activity regimen. The hypotheses of the one-sample sign test are H_0: $p = .5$ vs H_a: $p > .5$. Ignoring the one "tie" and working with the 24 other children in the sample, 17 of the 24 saw improvements (again, score change < 0). The one-sided P-value is

$$P(K \geq 17 \text{ when } K \sim \text{Bin}(24, .5)) = \sum_{k=17}^{24} \binom{24}{k} (.5)^k (1-.5)^{24-k} = 1 - B(16; 24, .5) = .032$$

Since $.032 < \alpha = .05$, H_0 is rejected at the .05 level. The data provide convincing statistical evidence that more than half of all such children would experience improvement after the physical activity regimen.

9. The proposed hypotheses are equivalent to a one-sample sign test. Let p = the true proportion of individuals who would perceive a longer time for the shorter exam (positive difference) in this experiment. Then the stated hypotheses are equivalent to H_0: $p = .5$ vs H_a: $p > .5$. With 109 observed "successes" out of 130, the one-sided P-value is

$P(K \geq 109 \text{ when } K \sim \text{Bin}(130, .5)) = \sum_{k=109}^{130} \binom{130}{k} (.5)^k (1-.5)^{130-k} \approx 0$. Given the very large sample size, the one-proportion z test is also appropriate, and the resulting test statistic value is $z = 7.72$. Either way, H_0 is resoundingly rejected — the data provide overwhelming evidence that subjects will typically (i.e., more than half the time) perceive the shorter/tougher test to take longer to complete, even though it doesn't.

Section 14.2

11. We test H_0: $\mu = 100$ vs H_a: $\mu \neq 100$. The test statistic is s_+ = sum of the ranks associated with the positive values of $(x_i - 100)$, and we reject H_0 at significance level .05 if $s_+ \geq 64$ (from Table A.11, $n = 12$, with $\alpha/2 = .026$, which is close to the desired .025) or if $s_+ \leq \frac{12(13)}{2} - 64 = 78 - 64 = 14$.

x_i	$(x_i - 100)$	ranks
105.6	5.6	7*
90.9	–9.1	12
91.2	–8.8	11
96.9	–3.1	3
96.5	–3.5	5
91.3	–8.7	10
100.1	0.1	1*
105.0	5.0	6*
99.6	–0.4	2
107.7	7.7	9*
103.3	3.3	4*
92.4	–7.6	8

$s_+ = 27$, and since 27 is neither ≥ 64 not ≤ 14, we do not reject H_0. There is not enough evidence to suggest that the mean is something other than 100.

13. Let μ = true mean pH level at this site on the Sacramento River. We test H_0: $\mu = 7.3$ vs H_a: $\mu > 7.3$. The test statistic is s_+ = sum of the ranks associated with the positive values of $(x_i - 7.3)$, and we reject H_0 at significance level .05 if $s_+ \geq 24$ (from Table A.11, $n = 7$, with $\alpha = .055$ close to the desired .05).

x_i	$(x_i - 7.3)$	ranks
7.20	–0.10	4
7.24	–0.06	2
7.31	0.01	1*
7.38	0.08	3*
7.45	0.15	5*
7.60	0.30	6*
7.86	0.56	7*

$s_+ = 1 + 3 + 5 + 6 + 7 = 22$, which is *not* ≥ 24, so H_0 is *not* rejected at the .05 level. The data do not provide convincing statistical evidence that the true mean pH level at this site exceeds 7.3.

15. The data is paired, and we wish to test H_0: $\mu_D = 0$ vs H_0: $\mu_D \neq 0$. With $n = 12$ and $\alpha = .05$, H_0 should be rejected if either $s_+ \geq 64$ or if $s_+ \leq 14$.

d_i	–.3	2.8	3.9	.6	1.2	–1.1	2.9	1.8	.5	2.3	.9	2.5
rank	1	10*	12*	3*	6*	5	11*	7*	2*	8*	4*	9*

$s_+ = 72 \geq 64$, so H_0 is rejected at level .05. In fact for $\alpha = .01$, the critical value is $c = 71$, so even at significance level .01 H_0 would be rejected.

17.

a. Let μ_D = true mean difference in prawn eaten (2nd trial minus 1st trial) for the population of all female cleaner fish under these conditions. We test H_0: $\mu_D = 0$ vs H_a: $\mu_D < 0$. With $n = 8$, we will reject H_0 at the .055 level if $s_+ \leq \frac{8(8+1)}{2} - 30 = 6$.

Female	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Diff.	–0.043	–0.182	–0.011	–0.179	–0.167	0.022	–0.096	–0.046
Signed rank	–3	–8	–1	–7	–6	2	–5	–4

Here, $s_+ = 2 \leq 6$, and so H_0 is rejected. At the .05 (really, .055) level, the data provide convincing statistical evidence that female fish eat less of their preferred food, on average, after having been chased by a male cleaner fish.

b. Now let μ_D = true mean difference in prawn eaten for the population of all *male* cleaner fish under these conditions. We test H_0: $\mu_D = 0$ vs H_a: $\mu_D > 0$ and reject H_0 at the .055 level if $s_+ \geq 30$. Because $s_+ = 28 < 30$, H_0 cannot be rejected at this level. We are not convinced that male cleaner fish eat more prawn, on average, the second time around.

19. The paired differences are –1.3, –86.7, –120.6, –198.0, –31.0, +25.8, –201.2, –18.3.

a. The signed ranks are –1, –5, –6, –7, –4, +3, –8, –2, from which $s_+ = 3$. With $n = 8$, we reject H_0: $\mu_D = 0$ in favor of H_0: $\mu_D < 0$ at the .05 (really, .055) level if $s_+ \leq 8(8+1)/2 - 30 = 6$ from Table A.11. Therefore, with $s_+ = 3 \leq 6$, H_0 is rejected. Equivalently, with the aid of software, the P-value is $P(S_+ \leq 3) = .021$. *This test procedure assumes that the population distribution of differences is at least symmetric.*

b. The mean and sd of the eight differences are $\bar{d} = -78.9$ and $s_D = 87.7$. So, the test statistic value is $t = \frac{-78.9 - 0}{87.7/\sqrt{8}} = -2.54$. At df = 8 – 1 = 7, the P-value is $P(T \leq -2.54) = .019$. Therefore, at the .05 level, H_0 is rejected. *This test procedure assumes that the population distribution of differences is normal.*

c. The P-values were .035 (sign test), .021 (Wilcoxon signed-rank test), and .019 (paired t test). As is typical, the P-value decreases with more powerful tests. But, all three tests agree that H_0 is rejected at the .05 level, and the sign test has the fewest assumptions.

21. From Table A.11 with $n = 7$, $c = 28$ and $7(7 + 1)/2 - c + 1 = 1$ are the indices of the pairwise averages that form a CI for μ with confidence level $100(1 - 2(.008))\% = 98.4\% \approx 99\%$. That is, we require the smallest and largest pairwise averages, which are easily $\bar{x}_{(1)} = (7.20 + 7.24)/2 = 7.22$ and $\bar{x}_{(28)} = (7.60 + 7.86)/2 = 7.73$. Therefore, the desired CI is (7.22, 7.73).

23. The paired differences for the $n = 8$ female cleaner fish are –.043, –.182, –.011, –.179, –.167, .022, –.096, and –.046. From Table A.11 with $n = 8$, $c = 26$ and $8(8 + 1)/2 - c + 1 = 11$ are the indices of the pairwise averages that form a CI for μ_D with confidence level $100(1 - 2(.023))\% = 95.4\% \approx 95\%$. Among the 36 pairwise averages, the 11th and 26th in order are $\bar{x}_{(11)} = -.1745$ and $\bar{x}_{(26)} = -.0110$. Therefore, the desired 95% CI for μ_D is (–.1745, –.0110).

Section 14.3

25. The ordered combined sample is 163(y), 179(y), 213(y), 225(y), 229(x), 245(x), 247(y), 250(x), 286(x), and 299(x), so $w = 5 + 6 + 8 + 9 + 10 = 38$. With $m = n = 5$, Table A.13 gives the upper tail critical value for a level .05 test as 36 (reject H_0 if $w \geq 36$). Since $38 \geq 36$, H_0 is rejected in favor of H_a.

27. The hypotheses of interest are H_0: $\mu_1 - \mu_2 = 1$ vs. H_a: $\mu_1 - \mu_2 > 1$, where 1 (X) refers to the original process and 2 (Y) to the new process. Thus 1 must be subtracted from each x_i before pooling and ranking. At level .05, with $m = n = 8$ H_0 should be rejected in favor of H_a if $w \geq 84$.

x – 1	3.5	4.1	4.4	4.7	5.3	5.6	7.5	7.6
rank	1	4	5	6	8	10	15	16
y	3.8	4.0	4.9	5.5	5.7	5.8	6.0	7.0
rank	2	3	7	9	11	12	13	14

Since $w = 1 + 4 + 5 + 6 + 8 + 10 + 15 + 16 = 65 < 84$, H_0 is not rejected.

29.

a.

X	rank	Y	rank
0.43	2	1.47	9
1.17	8	0.8	7
0.37	1	1.58	11
0.47	3	1.53	10
0.68	6	4.33	16
0.58	5	4.23	15
0.5	4	3.25	14
2.75	12	3.22	13

We verify that w = sum of the ranks of the x's = 41.

b. We are testing H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 < 0$. The reported P-value is $.0027 < .01$ so we reject H_0. There is evidence that the distribution of good visibility response time is to the left (or lower than) that response time with poor visibility.

31. Let μ_1 = true mean number of nonsynonymous mutations for all patients who experience durable benefit from Keytruda, and define μ_2 similarly for those not experiencing durable benefit. We test H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 > 0$. Technically $m = 7$ and $n = 9$ is beyond Table A.11, so we'll use the z approximation. The null distribution of W has mean $m(m+n+1)/2 = 59.5$ and variance $mn(m+n+1)/12 = 89.25$. To deal with the tie at value 300, we'll assign each one the average of their ranks; they occupy the 10th and 11th position in order, so each gets assigned rank 10.5. That makes the rank sum for the first sample $w = 7 + 9 + 10.5 + 12 + 13 + 14 + 16 = 81.5$. The z-value is $z = \frac{81.5 - 59.5}{\sqrt{89.25}} = 2.33$, so P-value $\approx 1 - \Phi(2.33) = .01$. At the .05 significance level, we reject H_0 (because .01 < .05) and conclude that patients experiencing durable clinical benefit from Keytruda do indeed have a higher average number of nonsynonymous mutations than those without a durable benefit.

33. For all three tests, we use the large-sample z approximation to the rank sum test. With $m = n = 40$, when H_0 is true the sampling distribution of W has mean $40(40 + 40 + 1)/2 = 1620$ and variance $40(40)(81)/12 = 10800$. The z-statistics and P-values for the three tests appear below.

Pain	Depression	Anxiety
$z = \frac{1475 - 1620}{\sqrt{10800}} = -1.40$	$z = \frac{1316 - 1620}{\sqrt{10800}} = -2.93$	$z = \frac{1171 - 1620}{\sqrt{10800}} = -4.32$
P-value = $\Phi(-1.40) = .0808$	P-value = $\Phi(-2.93) = .0017$	P-value = $\Phi(-4.32) < .0001$

Comparing each P-value to .01, we fail to reject the first null hypothesis and reject the other two. That is, the data do not convince us that dog therapy provides greater pain reduction than the control, but we are convinced that dog therapy provides both greater depression and anxiety reduction than the control. Having performed three tests at the .01 level, the chance of *at least one* type I error is no more than .01 + .01 + .01 = .03 (by Bonferroni's inequality).

35. $m = n = 5$ and from Table A.14, $c = 21$ and the 90% (actually 90.5%) interval is $(d_{ij(5)}, d_{ij(21)})$. The five smallest $x_i - y_j$ differences are –18, –2, 3, 4, 16 while the five largest differences are 136, 123, 120, 107, 87 (construct a table like the one in Section 14.3), so the desired interval is (16, 87).

Section 14.4

37. Let μ_i = true mean fasting C-peptide level (nmol/L) for the ith diabetes population ($i = 1, 2, 3, 4$). The hypotheses are $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs H_a: not all μ_i are equal. With $n = 26 + 32 + 65 + 17 = 140$, the Kruskal-Wallis test statistic value is

$$h = \frac{12}{140(140+1)} \sum J_i \left(\overline{r}_{i\cdot} - \frac{140+1}{2} \right)^2$$

$$= \frac{12}{140(140+1)} \left[26(72.8 - 70.5)^2 + 32(79.2 - 70.5)^2 + 65(56.4 - 70.5)^2 + 17(104.6 - 70.5)^2 \right] = 21.43$$

Based on a χ^2_{4-1} distribution, the P-value is < .0001. Thus, H_0 is rejected at any reasonable significance level — the data provide convincing statistical evidence that the mean FCP level is not the same for all types of diabetics.

39. Let μ_i = true mean fracture load (kN) for the ith loading point distance (i = 1, 2, 3). The hypotheses are $H_0: \mu_1 = \mu_2 = \mu_3$ vs H_a: not all μ_i are equal. The data are repeated here with their ranks.

Distance	Fracture load			
31.2 mm	4.78 [10]	4.41 [9]	4.91 [11]	5.06 [12]
36.0 mm	3.47 [5]	3.85 [8]	3.77 [7]	3.63 [6]
42.0 mm	2.62 [1]	2.99 [3]	3.39 [4]	2.86 [2]

From these, $\bar{r}_{1\cdot} = 10.5, \bar{r}_{2\cdot} = 6.5, \bar{r}_{3\cdot} = 2.5$ and $h = \frac{12}{12(12+1)} \sum 4(\bar{r}_{i\cdot} - 6.5)^2 = 9.85$. Based on a χ^2_{3-1} distribution, the P-value is roughly .007. Since .007 < .01, we reject H_0 and conclude that the true mean fracture load is not the same at these three distances. (The result should be obvious, since the ranks sort perfectly into the three rows: 1-4 at the bottom, then 5-8, then 9-12 at the top.)

41. As noted in the hint, the R_{ij}'s are simply a re-arrangement of the integers 1 to n. So, for a start, the mean rank is simply $\bar{R}_{\cdot\cdot} = \frac{1}{n}(1 + 2 + \cdots + n) = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$. Then, the double sum that defines SST can be rewritten: $\text{SST} = \sum\sum (R_{ij} - \frac{n+1}{2})^2 = \sum_{k=1}^{n} (k - \frac{n+1}{2})^2$. Expand the quadratic and use the formulas for the sum and sum-of-squares of the first n integers:

$$\text{SST} = \sum_{k=1}^{n} k^2 - 2\tfrac{n+1}{2}\sum_{k=1}^{n} k + n(\tfrac{n+1}{2})^2 = \frac{n(n+1)(2n+1)}{6} - 2\frac{n+1}{2}\frac{n(n+1)}{2} + n\frac{(n+1)^2}{4}$$

$$= \frac{2n(n+1)(2n+1) - 6(n+1)n(n+1) + 3n(n+1)^2}{12} = \frac{n^3 - n}{12} = \frac{n(n^2-1)}{12}$$

43.

a. First, determine the within-subject ranks (i.e., ranks 1, 2, 3 within each column):

	Subject					
Position	**1**	**2**	**3**	**4**	**5**	**6**
Neutral	1.28[1]	0.88[2]	0.69[3]	1.52[1]	0.83[2]	2.58[3]
Flexion	1.29[2]	0.76[1]	0.43[2]	2.11[3]	1.07[3]	2.18[2]
Extension	1.51[3]	1.12[3]	0.23[1]	1.54[2]	0.20[1]	1.67[1]

Let μ_i = true mean disc bulge (mm) at T11-T12 in the ith position (i = 1 for neutral, 2 for flexion, 3 for extension). The hypotheses are $H_0: \mu_1 = \mu_2 = \mu_3$ vs H_a: not all μ_i are equal. The rank averages of the three positions/rows are $\bar{r}_{1\cdot} = 12/6 = 2, \bar{r}_{2\cdot} = 13/6 = 2.1\overline{6}, \bar{r}_{3\cdot} = 11/6 = 1.8\overline{3}$, from which the test statistic value (based on I = 3 groups and J = 6 subjects) is $Fr = \frac{12(6)}{3(3+1)} \sum_{i=1}^{3} (\bar{r}_{i\cdot} - 2)^2 = 0.333$. Compared to a χ^2_{3-1} distribution, the P-value is very large (roughly .85 from software). Thus, H_0 is certainly not rejected. The data do not indicate that true mean disc bulge at T11-T12 varies by position.

b. The rank means are now $\overline{r}_{1\cdot} = 219/105 = 2.086, \overline{r}_{2\cdot} = 222/105 = 2.114, \overline{r}_{3\cdot} = 189/105 = 1.800$, from which $Fr = \frac{12(105)}{3(3+1)}\sum_{i=1}^{3}(\overline{r}_{i\cdot} - 2)^2 = 6.34$. Compared to a χ^2_{3-1} distribution, the *P*-value is about .042. Thus, we reject H_0 at the .05 significance level and conclude, based on all 105 subjects, that true mean disc bulge at T11-T12 indeed varies by position.

c. Now define μ_i = true mean disc bulge (mm) at T4-T5 in the *i*th position. The hypotheses are unchanged. The new rank averages are 207/105 = 1.971, 221/105 = 2.105, and 202/105 = 1.924, and the updated test statistic value (same formula as in (b)) is $Fr = 1.85$. The *P*-value from a χ^2_{3-1} distribution is roughly .40, so that H_0 is *not* rejected (at .05 or any reasonable significance level). The data do not indicate that position affects true mean disc bulge at T4-T5.

45. There are $I = 10$ treatments (algorithms) and, coincidentally, $J = 10$ blocks (images). Let μ_i = the true Kapur entropy measurement for the *i*th algorithm ($i = 1, \ldots, 10$). The hypotheses are $H_0: \mu_1 = \cdots = \mu_{10}$ vs H_a: not all μ_i are equal. From the rank means provided, the test statistic value is $Fr = \frac{12(10)}{10(10+1)}\sum_{i=1}^{10}(\overline{r}_{i\cdot} - 5.5)^2 = 78.67$. Even at 9 df, the *P*-value associated with such a large chi-squared statistic is effectively zero, so H_0 is resoundingly rejected. The data provide clear evidence that the algorithms are *not* equally effective at minimizing Kapur's entropy measure. In particular, the four algorithms inspired by quantum computing (Q's in the name) have much lower rank means, suggesting they are far better at minimizing entropy.

47.

a. Within each block, the ranks are 1, 2, …, *J*. So, the mean rank within every block is $\overline{R}_{\cdot j} = \frac{1}{J}(1 + 2 + \cdots + J) = \frac{1}{J}\frac{J(J+1)}{2} = \frac{J+1}{2}$. But also, the collection of all ranks in Friedman's test procedure are 1, 2, …, *J* repeated *I* times (one for each treatment). So, the grand mean of all ranks is also just the average of 1, 2, …, *J*; i.e., $\overline{R}_{\cdot\cdot} = \frac{J+1}{2}$. Thus $\overline{R}_{\cdot j} - \overline{R}_{\cdot\cdot} = \frac{J+1}{2} - \frac{J+1}{2} = 0$, from which SSB = 0.

b. We've established that $\overline{R}_{\cdot\cdot} = \frac{J+1}{2}$. Within each block, the R_{ij}'s are just a rearrangement of the integers 1, 2, …, *J*. Thus

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}(R_{ij} - \overline{R}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\sum_{k=1}^{J}\left(k - \frac{J+1}{2}\right)^2 = I\sum_{k=1}^{J}\left(k - \frac{J+1}{2}\right)^2 = I\sum_{k=1}^{J}k^2 - 2I\frac{J+1}{2}\sum_{k=1}^{J}k + IJ\left(\frac{J+1}{2}\right)^2$$

$$= I\frac{J(J+1)(2J+1)}{6} - I(J+1)\frac{J(J+1)}{2} + IJ\left(\frac{J+1}{2}\right)^2 = \frac{IJ^3 - IJ}{12} = \frac{IJ(J^2-1)}{12}$$

Supplementary Exercises

49. Because we have two independent samples, the Wilcoxon rank-sum test is appropriate. Let μ_1 = true mean vessel patency (%) for all mice receiving PEGPH20, and define μ_2 similarly for the standard treatment. The hypotheses of interest are $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 > 0$. The rank sum for the first (x) sample is 5 + 6 + 7 + 8 = 26. Using Table A.13 with $m = n = 4$, the upper-tailed P-value is $P(S_+ \geq 26) = .014$. Since .014 < .05, H_0 is rejected at the .05 level, indicating that PEGPH20 indeed yields higher average vessel patency than the standard treatment. However, H_0 would not be rejected at the .01 level since .014 > .01. This is perhaps surprising because 26 is the highest possible s_+ value with $m = n = 4$, meaning that .014 is the lowest possible P-value with these sample sizes. It would be impossible with these small sample sizes to ever reject H_0 at any level below .014.

51. For both parts, use the large-sample z approximation to the Wilcoxon signed-rank test. When H_0 is true, the mean and variance of S_+ are $n(n+1)/4 = 50(50+1)/4 = 637.5$ and $50(51)(101)/24 = 10731.25$.

a. The test statistic value is $z = \dfrac{616 - 637.5}{\sqrt{10731.25}} = -0.21$, so the P-value is $2[1 - \Phi(|-0.21|)] \approx .834$. With such a large P-value, the data do not suggest a significant difference between the two pain measurements, at least on average, for the population of osteoarthritis patients. If the measurements ought to be the same both times ("reliability"), the data do not contradict a claim the sensory test is reliable.

b. The test statistic value is $z = \dfrac{814 - 637.5}{\sqrt{10731.25}} = 1.70$, so the P-value is $2[1 - \Phi(1.70)] \approx .089$. At the .10 significance level, H_0 is rejected, suggesting that on average the two pain measurements differ for healthy people. This might indicate a lack of reliability of the sensory test for the population of healthy patients.

53. Let μ_i = true mean axial stiffness for the ith plate length. The hypotheses are $H_0 : \mu_1 = \cdots = \mu_5$ vs H_a: not all μ_i are equal. The ranks and rank average are displayed below.

i								$\overline{r}_{i\cdot}$
1	1	2	3	4	5	10	24	7.00
2	6	8	9	13	17	21	22	13.71
3	11	12	15	16	18	20	25	16.71
4	7	14	19	26	29	32	33	22.86
5	23	27	28	30	31	34	35	29.71

The test statistic value is $h = \dfrac{12}{35(35+1)} \sum \left(\overline{r}_{i\cdot} - \dfrac{35+1}{2} \right)^2 = 20.21$. Since $20.21 \geq \chi^2_{.01,4} = 13.277$, H_0 is rejected at the .01 level. The data provide convincing evidence that mean axial stiffness varies by plate length (i.e., not all means are equal).

55. Let μ_i = population mean skin potential (mV) with the ith emotion (i = 1 for fear, etc.). The hypotheses are $H_0 : \mu_1 = \cdots = \mu_4$ vs H_a: not all μ_i are equal. Rank the values within each column/subject/block from 1 to 4, then determine the mean rank for each emotion/row. The resulting rank means are 3.375, 2.5, 2.375, and 1.75. Hence, $Fr = \dfrac{12(8)}{4(4+1)} \sum_{i=1}^{4} (\overline{r}_{i\cdot} - 2.5)^2 = 6.45$. Since $6.45 < \chi^2_{.05,3} = 7.815$, we fail to reject H_0 at the .05 level. The data do not provide convincing evidence that mean skin potential depends on which emotion a person is experiencing.

57.

Sample:	*y*	*x*	*y*	*y*	*x*	*x*	*x*	*y*	*y*
Observations:	3.7	4.0	4.1	4.3	4.4	4.8	4.9	5.1	5.6
Rank:	1	3	5	7	9	8	6	4	2

The value of W' for this data is $w' = 3 + 6 + 8 + 9 = 26$. At level .05, the critical value (Table A.13, m = 4, n = 5) for the upper-tailed test is c = 27 (α = .056). Since 26 is not ≥ 27, H_0 cannot be rejected at level .05.