CHAPTER 14

14.1 odds = 1/12.

14.3 $\text{odds}_{\text{men}} \doteq 1.0635,$ and $\text{odds}_{\text{women}} \doteq 0.8667.$

14.5 For the men, log(odds) = ln(1.0635) $\doteq 0.6157$; log(odds) $\doteq -0.1431$.

14.7 If $x = 1$ for men and 0 for women, $\log(\text{odds}) = -0.1431 + 0.2047x$. (If vice versa, $\log(\text{odds}) = 0.0616 - 0.2047x$.) The odds ratio is 1.2272 (or 0.8149).

14.9 In Example 14.7, $b_1 = 0.660953$. $e^{0.660953} \doteq 1.936637,$ which rounds to 1.94.

14.11 (a) $\hat{p} \doteq 0.4606.$ (b) $\text{odds} \doteq 0.8539.$

14.13 (a) The model is y = log(odds) = $\beta_0 + \beta_1 x$. x = 1 if 25 or younger and 0 otherwise. **(b)** β_0 is the log(odds) of having used their cell phone to consult about a purchase in the last 30 days for people over 25; β_1 is the difference in log(odds) for those 25 or younger.

14.15 (a) n = 102,738. 3829 were rejected and 98869 were not. **(b)** log(odds) = −6.76382 + 4.41058x.

14.17 (a) Odds ratio $\doteq 82.3172$. **(b)** 64.7468 to 104.65572. **(c)** Recruits over 40 were about 82.3 times more likely to be rejected for service sue to bad teeth than recruits 20 or younger.

14.19 (a) G = 9413.21, $P < 0.0005$. **(b and c)** In each case, the hypotheses are $H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$.

Age Group	Confidence interval	z
21 to 25	(5.5941, 9.2247)	15.45
25 to 30	(13.5793, 22.1399)	22.82
30 to 35	(27.5384, 44.7252)	28.76
35 to 40	(41.3094, 66.8606)	32.25
Over 40	(64.7468, 104.65572)	36.01

14.21 For example, 647/78,639 = 0.82% of those 20 or younger were rejected. The rejection rate for those 20 to 25 was 1.96%, 1.96/0.82 = 2.39. That analysis would indicate the older recruits were about 2.39 times more likely to be rejected, while this analysis indicates an odds ratio of 6.806. The odds ratio is not directly comparable to the ratio of probabilities because it compares odds (the chance of the event happening divided by the chance of the event *not* happening).

14.23 Odds of being an exergamer are no TV: 0.1249, less than two hours of TV per day: 0.2595, at least two hours of TV per day: 0.4514. The fitted logistic regression equation is log(odds) = −1.34807 + 0.551742(> 2 hours) – 0.731368(No TV). The odds ratio for no TV to less than 2 hours of TV is 0.4813; for less than 2 hours of TV and at least two hours of TV is 1.7363. Both are significantly different from 0; 95% confidence interval for No TV is (0.020, 1.14) and for more than 2 hours (1.31, 2.30).

14.25 **(a)** The appropriate test would be a chi-square test with df = 4. **(b)** The logistic regression model has no error term. **(c)** H_0 should refer to β_1 (the population slope) rather than b_1 (the estimated slope). **(d)** The interpretation of coefficients is affected by correlations among explanatory variables.

14.27 **(a)** 2.845. **(b)** 6.140. **(c)** 9.925.

14.29 **(a)** $\hat{p}_{\text{low}} \doteq 0.0753$ and $\hat{p}_{\text{high}} \doteq 0.0899$. **(b)** 0.0814 and 0.0988. **(c)** −2.5083 and −2.3150.

14.31 2.1096 to 4.1080.

14.33 **(a)** $z \doteq 8.01$. **(b)** $z^2 \doteq 64.23$, which agrees with the value of X^2 given by SPSS and SAS. **(c)** For both the Normal and chi-square distributions, the test statistics are quite extreme, consistent with the reported *P*-value.

14.35 The odds favor a low tip from senior adults, those dining on Sunday, those who speak English as a second language, and French-speaking Canadians. Diners who drink alcohol and lone males are less likely to leave low tips. For example, for a senior adult, the odds of leaving a low tip were 1.099 (for a probability of 0.5236).

14.37 **(a)** $\hat{p}_{\text{hi}} \doteq 0.01648$ and $\text{odds}_{\text{hi}} \doteq 0.01675$, or about 1 to 60. **(b)** $\hat{p}_{\text{lo}} \doteq 0.00785$ and $\text{odds}_{\text{lo}} \doteq 0.00791$, or about 1 to 126. **(c)** The odds ratio is 2.1181.

14.39 **(a)** The estimated odds ratio is 2.1181; the odds-ratio interval is 1.28 to 3.51. **(b)** We are 95% confident that the odds of death from cardiovascular disease are about 1.3 to 3.5 times greater in the high-blood-pressure group.

14.41 **(a)** $\log(\text{odds}) = \beta_0 + \beta_1 x$, where $x = 1$ if the person is over 40, and 0 if the person is under 40. **(b)** p_i is the probability that the ith person is terminated; this model assumes that the probability of termination depends on age (over/under 40). **(c)** The estimated odds ratio is 3.859. A 95\% confidence interval for b_1 is 0.5409 to 2.1599. The odds of being terminated are 1.7 to 8.7 times greater for those over 40. **(d)** Use a multiple logistic regression model.

14.43 $\log(\text{odds}) = -1.7794 + 0.7068x$.

14.45 $\log(\text{odds}) = -0.9785 + 0.0614x$; the slope is significantly different from 0.

14.47 **(a)** $X^2 \doteq 33.65$ (df = 3), P = 0.0001.

(b) log (odds) = −6.053+0.3710 HSM + 0.2489 HSS + 0.03605 HSE. 95% confidence intervals: 0.1158 to 0.6262, −0.0010 to 0.4988, and −0.2095 to 0.2816. **(c)** Only the coefficient of HSM is significantly different from 0, though HSS may also be useful.

14.49 (a) $X^2 \doteq 19.2256$, $\text{df} = 3$, $P = 0.0002$. **(b)** $X^2 \doteq 3.4635$, $\text{df} = 2$, $P = 0.1770$. **(c)** High school grades (especially HSM and, to a lesser extent, HSS) are useful, while SAT scores are not.

CHAPTER 15

15.1 The ranks for Group A are 3, 4, 8, 2, and 1. For group B: 6, 9, 10, 5, and 7.

15.3 To test whether the two groups of spas have the same distribution, we find $W = 18$.

15.5 $\mu_W = 27.5$ and $\sigma_W \doteq 4.7871$. $z \doteq -1.98$, which gives $P \doteq$ 0.0477; with the continuity correction, $z \doteq -1.88$, for which $P \doteq 0.0601$.

15.7 (a) For exergamers, no TV: 2.14%, some TV: 56.94%, more than 2 hours TV: 40.93%. For non-exergamers, no TV: 5.22%, some TV: 67.03%, more than 2 hours: 27.75%. **(b)** $\chi^2 = 20.068$, df = 2, $P < 0.0005$.

15.9 Now, rank all the members of the class together. Taking Group 1 to be the women, the sum of their ranks is 3 + 6 + 7 + 9 + 11 = 36.

15.11 $\mu_W = 30$ and $\sigma_W \doteq 5.4772$.

15.13 (a) W = 1351159.5, z = 8.5063, P < 0.0001. **(b)** For those who are civically engaged, 66.1% have at least some college, and 39.2% are college graduates. The corresponding percents for those who are not civically engaged are 49.7% and 25.9%.

15.15 Men and women are not significantly different (W = 1421, P = 0.6890). The t test assumes Normal distributions; with small samples, this might be risky; in the full data set, the men's distribution is skewed, and the women's distribution has a near-outlier.

15.17 (a) Normal quantile plots are not shown. No departures from Normality were seen. **(b)** For testing $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 > \mu_2$, we have $\bar{x}_1$ = 0.676, $s_1 \doteq 0.119$, $\bar{x}_2$ = 0.406, $s_2 \doteq 0.268$. Then, t = 2.06, which gives P = 0.047 (df = 5.52). **(c)** We test: H_0: Scores for both groups are identically distributed vs. H_a: High-progress children systematically score higher for which we find W = 36 and $P \doteq$ 0.0473.

15.19 (a) The 5 high-progress readers have ranks 8, 9, 4, 7, and 10. **(b)** $W = 38$; under H_0, $\mu_W = 27.5$ and $\sigma_W \doteq 4.7871$. **(c)** $z \doteq 2.09$, $P = 0.0183$. **(d)** The tied observations have ranks 4.5 and 8.5.

15.21 We test H_0: service scores and food scores have the same distribution; H_a: service scores are higher. Ranking service minus food differences gives $W^+ = 20$.

15.23 $\mu_{W^+} = 14$ and $\sigma_{W^+} \doteq 5.9161$. $z \doteq 0.93$ and $P \doteq$ 0.1762.

15.25 W^+ = 35.

15.27 $\mu_{W^+} = 18$. $\sigma_{W^+} \doteq 7.1414$.

15.29 (a) $n = 20$. **(b)** $W^+ = 192.0$. **(c)** $P = 0.001$. **(d)** The estimated median is 2.925. With $n = 20$, the median should be the average of the middle two absolute value differences.

15.31 $W^+ = 119$. $P = 0.001$.

15.33 (a) Six of the seven subjects rated drink A higher. **(b)** $t = 2.01$, $P = 0.091$. **(c)** $W^+ = 26.5$, $P = 0.043$. **(d)** The new data point is an outlier, which may make the t procedure inappropriate. This also increases the standard deviation of the differences, which makes t insignificant. The Wilcoxon test is not sensitive to outliers, and the extra data point makes it powerful enough to reject H_0.

15.35 (a) The distribution is right-skewed but has no outliers. **(b)** $W^+ = 31$ and $P = 0.556$.

15.37 We want to compare the attractiveness scores for $k = 5$ independent samples (the 102, 302, 502, 702, and 902 friend groups of subjects). Under the null hypothesis for ANOVA, each population is $N(\mu, \sigma)$. An F test is used to compare the group means. The Kruskal-Wallis test only assumes a continuous distribution in each population, and uses a chi-square distribution for the test statistic.

15.39 Minitab gives the median of each group, and the average rank for each group. Using the "adjusted for ties" values, $H = 17.03$, $P = 0.002$.

15.41 (a) $H = 9.12$ (df = 2), $P = 0.010$. **(b)** ANOVA yielded $F=7.72$ (df 2 and 42), $P = 0.001$. The ANOVA evidence is slightly stronger, but (at $\alpha = 0.05$) the conclusion is the same.

15.43 (a) The diagram should show 10 rats assigned to each group; apply the treatments, then observe bone density. **(b)** Stemplots suggest greater density for high-jump rats and a greater spread for the control group. **(c)** $H = 10.66$ and $P = 0.005$. ANOVA assumes Normal distributions with the same standard deviation and tests whether the means are all equal. Kruskal-Wallis tests whether the distributions are the same (but not necessarily Normal).
(d) There is strong evidence that the high-jump group has the highest average rank (and the highest density), the low-jump group is in the middle, and the control group is lowest.

15.45 $H = 45.35$, df = 2, $P < 0.0005$.

15.47 (a) Verizon's mean and median service times (1.73 and 1 hr) are quite a bit less than CLEC's (4.8 and 5 hr). **(b)** The distributions are sharply skewed, and the sample sizes are quite different, so a t test is not reliable. The Wilcoxon rank-sum test gives $W = 4778.5$, and $P = 0.0026$.

15.49 For meat, $W = 15$ and $P = 0.4705$, and for legumes, $W = 10.5$ and $P = 0.0433$ (or 0.0421, adjusted for ties).

15.51 (a) *Bihai*-red, *bihai*-yellow, and red-yellow. **(b)** $W_1 = 504$, $W_2 = 376$, $W_3 = 614$. All *P*-values are reported as 0 to four decimal places. **(c)** All three comparisons are significant at the overall 0.05 level (and would even be significant at the overall 0.01 level).

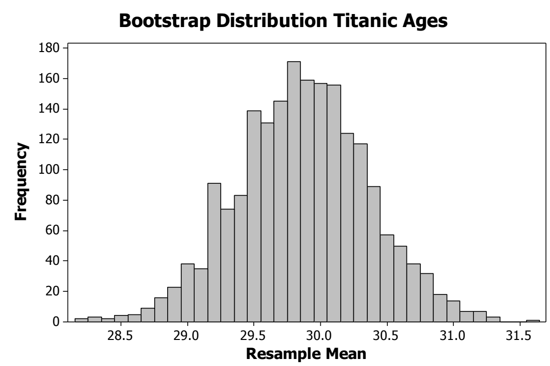
CHAPTER 16

16.5 The mean of the bootstrap distribution is about 75 percent; it is roughly Normal, but the peak in the center is too high to be truly Normal. The Normal quantile plot is straight, except for some straggling at the low end. Based on these plots, the bootstrap distribution is approximately Normal.

16.7 (a) The standard deviation of the bootstrap distribution will be approximately $s/\sqrt{n}$. **(b)** Bootstrap samples are done *with* replacement. **(c)** You should use a sample size *equal* to the original. **(d)** The bootstrap distribution is created by sampling with replacement from the original sample, not the population.

16.9 The bootstrap distribution is remarkably Normal, considering the original sample values. The center is about $420. The histogram is shown at the top of the next column.

16.11 Starting from a right-skewed distribution, the bootstrap distribution is approximately Normal and centered a bit less than 30 years old.



16.13 Bootstrap standard errors will vary. **(a)** s = 13.8; $s/\sqrt{n} = 13.8/\sqrt{60} = 1.782$. **(b)** s = 221.465, so SE = $221.465/\sqrt{186} = 16.239$.

16.15 (a) This distribution looks more strongly skewed than the one in Exercise 16.14 (it would look even more skewed if some of the calls above 600 seconds had been in the sample). **(b)** The bootstrap standard error for the sample of 80 calls is about 36.8. For the sample of 10 calls, the standard error is about 29. This is smaller (not larger), possibly due to the lack of the extremely long calls in the small sample.

16.17 (a) The bootstrap distribution is skewed; a t interval might not be appropriate. **(b)** The bootstrap t interval is $\bar{x} \pm t^{*}\mathrm{SE}_{\mathrm{boot}}$, where $\bar{x}$ = 354.1 sec, t* 2.0096 for df = 49, and SEboot is typically between 39.5 and 46.5. **(c)** The interval reported in Example 7.11 was 266.6 to 441.6 seconds.

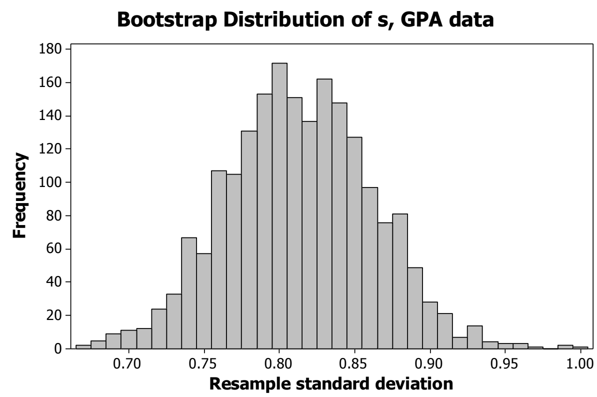
16.19 The summary statistics given in Example 16.6 include standard deviations $s_1 = 0.859$ for males and $s_2 = 0.748$ for females, so SED $= 0.13263$. The standard error reported by the bootstrap routine in Example 16.6 was 0.1327419 (very close).

16.21 The interval is 93.419 to 131.181.

16.23 (a) $\bar{x} = 29.881$, df = 1045, $t^* = 1.9622$, $SE_{boot} = 0.445$, so the interval is 29.008 to 30.754 years. **(b)** The standard deviation of the original sample is $s = 14.414$, so the interval is 29.006 to 30.756. The intervals are almost identical, only differing in the third decimal place.

16.25 (a) The bootstrap bias is typically between −4 and 4, which is small relative to $\bar{x} = 196.575$ min. **(c)** $SE_{\bar{x}} = 38.2392$, while SE_{boot} ranges from about 35 to 41. The usual *t* interval is 120.46 to 272.69 min.

16.27 The tails of this distribution are not Normal. We should hesitate to use a *t* confidence interval for the population standard deviation.



16.29 (a) The data appear to be roughly Normal, though with the typical random gaps and bunches that usually occur with relatively small samples. It appears from both the histogram and quantile plot that the mean is slightly larger than zero, but the difference is not large enough to rule out a *N*(0, 1) distribution. **(b)** The bootstrap distribution is extremely close to Normal with no appreciable bias. **(c)** $SE_{\bar{x}} = 0.1308$, and the usual *t* interval is −0.1357 to 0.3854.

16.31 (a) The sample standard deviation is $s = 4.4149$ mpg. **(b)** The typical range for SE_{boot} is between 0.55 and 0.65. **(c)** SE_{boot} is quite large relative to *s*, suggesting that *s* is not a very accurate estimate. **(d)** There is substantial negative bias and some skewness, so a *t* interval is probably not appropriate.

16.33 (a) The distribution of $\bar{x}$ is *N*(80, 6.325). **(b)** SE_{boot} ranges from about 6.55 to about 6.75. **(c)** For $n = 40$, SE_{boot} ranges from about 3.1 to 3.9. For $n = 160$, SE_{boot} ranges from about 1.25 to about 1.45.

16.35 Answers will vary depending on the original samples.

16.37 (a) Both graphs indicate the bootstrap distribution is right skewed. **(b)** The 95% t confidence interval is $23.26 \pm (2.0096)SE_{boot}$. **(c)** The 95% bootstrap percentile confidence interval is about 16 days to about 31 days.

16.39 The bootstrap distribution is right skewed; an interval based on *t* would not be appropriate. The original sample had the statistic of interest $\hat{\theta} = 1.21$. The bootstrap distribution had a sample mean a bit higher because the bias is 0.045. The standard deviation of the bootstrap distribution

(SE_{boot}) is 0.2336. The BCa confidence interval is (0.766, 1.671), which is located about 0.11 higher than the regular bootstrap interval which was (0.653, 1.554).

16.41 (a) The bootstrap percentile and *t* intervals are very similar, suggesting that the *t* intervals are acceptable. **(b)** Every interval (percentile and *t*) includes 0.

Typical ranges	
t lower	−0.16 to −0.11
t upper	0.36 to 0.41
Percentile lower	−0.17 to −0.09
Percentile upper	0.35 to 0.42

16.43 The 95% *t* interval given in Example 16.5 was 2.794 to 3.106. The 95% bootstrap percentile interval is 2.793 to 3.095, as given in Example 16.8. The differences are relatively small relative to the width of the intervals, so they do not indicate appreciable skewness.

16.45 One set of 1000 repetitions gave the BCa interval as (0.4503, 0.8049). We see on the bootstrap distribution is left-skewed, and that there was one high outlier as well. The lower end of the BCa interval typically varies between 0.42 and 0.46 while the upper end varies between 0.795 and 0.805. These intervals are lower than those found in Example 16.10.

16.47 The percentile interval is shifted to the right relative to the bootstrap *t* interval. The more accurate intervals are shifted even further to the right.

Typical ranges	
t lower	114 to 127
t upper	266 to 279
Percentile lower	127 to 140
Percentile upper	267 to 298
BCa lower	137 to 152
BCa upper	292 to 371

16.49 Which interval students will report will vary. Typical ranges are given at right. We note that the locations of the intervals are quite different (especially the low ends); this mirrors what was seen in Exercise 16.47. These intervals are much lower than those in Exercise 16.47.

Typical ranges	
SE_{boot}	28 to 30
t lower	48 to 52
t upper	178 to 181
Percentile lower	60 to 67
Percentile upper	169 to 177
BCa lower	66 to 72
BCa upper	176 to 195

16.51 The low endpoint of the BCa interval ranges from about −0.427 to −0.372 and the upper end ranges from about 0.087 to 0.133. This interval is comparable to the -0.416 to 0.118 found in Example 16.6.

16.53 (a) The bootstrap distribution is left-skewed (with outliers); simple bootstrap inference is inappropriate. **(b)** The percentile interval is consistently about 0.978 to 0.994, while the BCa interval is consistently 0.970 to 0.993. These agree fairly well, but the BCa interval is shifted right and a bit shorter. We do have significant evidence that the correlation is not 0.

16.55 (a) The regression line is $\text{Rating} = 26.724 + 1.207$ PricePerLoad. **(b)** The ends of the bootstrap distribution do not look very Normal, a *t* interval may not be appropriate. **(c)** The typical standard error of the slope is 0.2846. With $t_{22} = 2.074$, the typical confidence interval would be 1.207 ± (2.074)(0.2846), or 0.6167 to 1.7973. All these intervals seem to be located higher.

Typical ranges	
SE_{boot}	0.228 to 0.242
t lower	0.705 to 0.734
t upper	1.68 to 1.71
Percentile lower	0.83 to 0.86
Percentile upper	1.74 to 1.81
BCa lower	0.71 to 0.80
BCa upper	1.62 to 1.69

16.57 The regression equation is $\text{Debt2010} = 0.05 + 1.06\text{Debt2009}$. **(a)** The scatterplot indicates a possible increase in variability with increasing 2009 debt, as well as possible outliers (especially the point at upper right, which was Greece). The Normal plot also indicates potential outliers on both ends of the distribution. **(b)** We see indications of high outliers in both the histogram and the Normal quantile plot of the bootstrap distribution. **(c)** The standard confidence interval for the slope is 1.06 ± (2.04)(0.03027), 0.998 to 1.122. This interval is close to all the bootstrap intervals, but a bit narrower.

16.59 No, because we believe that one population has a smaller spread, but in order to pool the data, the permutation test requires that both populations be the same when H_0 is true.

16.61 Enter the data with the score given to the phone and an indicator for each design. We have hypotheses $H_0: \mu_1 = \mu_2$ and $H_a: \mu_1 \neq \mu_2$. Resample the design indicators (without replacement) to scramble them. Compute the mean score for each scrambled design group. Repeat the process many times. The *P*-value of the test will be the proportion of resamples where the resampled difference in group means is larger than the observed difference (in absolute value).

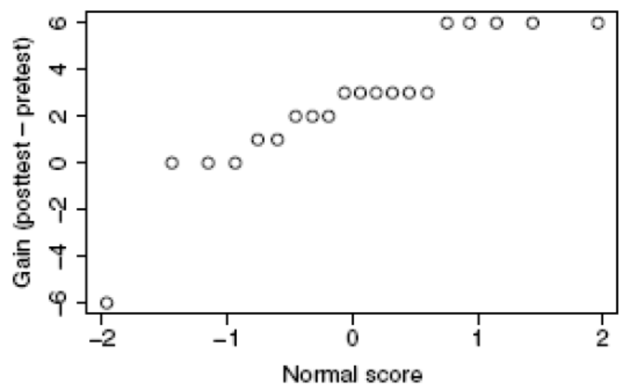
16.63 If there is no relationship, we have $H_0: \rho = 0$. We test this against $H_a: \rho \neq 0$. We can resample one of the variables, say screen satisfaction (without replacement), and compute the correlation between that and the original scores for keyboard satisfaction. Repeat the process many times. The proportion of times the correlation in the resamples is larger than the original (in absolute value) is the *P*-value for the test.

16.65 (a) The observed difference in means is $\frac{57+53}{2} - \frac{19+37+41+42}{4} = 20.25$. **(d)** Out of 20 resamples, the number which yields a difference of 20.25 or more has a binomial distribution with $n = 20$ and $p = 1/15$, so a *P*-value is likely to be between 0 and 0.2. **(e)** Only one resample gives a difference of means greater than or equal to the observed value, so the exact *P*-value is 1/15.

16.67 (a) $H_0: \mu_{EE} = \mu_{LE}$ and $H_a: \mu_{EE} \neq \mu_{LE}$ (the question of interest in Example 7.16 was whether we could conclude the two groups were not the same). **(b)** $\bar{x}_{EE} = 11.56$, $s_{EE} = 4.306$, $\bar{x}_{LE} = 5.12$, $s_{LE} = 4.622$. $t = 2.28$, df = 8, $P = 0.0521$. At the 0.05 level, we fail to reject the null hypothesis that the two groups have the same mean weight loss. **(c)** Many repetitions of the permutation test found typical *P*-values between 0.048 (where the null hypothesis would be rejected at the 0.05 level) to 0.07 (where the null would not be rejected). **(d)** Many repetitions of the boot procedures found a minimum value for the low end of the BCa interval 0.35.

16.69 (a) The two populations should be the same shape, but skewed; (clearly non-Normal) so that the *t* test is not appropriate. **(b)** Either test is appropriate if the two populations are both Normal with the same standard deviation. **(c)** We can use a *t* test, but not a permutation test, if both populations are Normal with different standard deviations.

16.71 (a) We test $H_0: \mu = 0$ versus $H_a: \mu > 0$, where μ is the population mean difference before and after the summer language institute. We find $t \doteq 3.86$, df = 19, and $P \doteq 0.0005$. **(b)** The quantile plot (below) looks odd because we have a small sample, and all differences are integers. **(c)** The *P*-value is almost always less than 0.002. Both tests lead to the same conclusion: The difference is statistically significant (the language institute did help comprehension).



16.73 (a) We have $H_0: \rho = 0$ versus $H_a: \rho \neq 0$. **(b)** The observed correlation is *r* = 0.671. We create permutation samples and observe the proportion with correlations at least 0.671 in absolute value. You should find a *P*-value 0.002 or less. We'll conclude there is a correlation between price and rating for laundry detergents.

16.75 For testing $H_0: \sigma_1 = \sigma_2$ versus $H_a: \sigma_1 \neq \sigma_2$, the permutation test *P*-value will almost always be between 0.065 and 0.095. In the solution to Exercise 7.105, we found $F \doteq 1.50$ with df = 29 and 29, for which $P \doteq 0.2757$—three or four times as large. In this case, the permutation test *P*-value is smaller, which is typical of short-tailed distributions.

16.77 For the permutation test, we must resample in a way that is consistent with the null hypothesis. Hence we pool the data—assuming that the two populations are the same—and draw samples (without replacement) for each group from the pooled data. For the bootstrap, we do not assume that the two populations are the same, so we sample (with replacement) from each of the two datasets separately, rather than pooling the data first.

16.79 (a) We will test $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$, (males are coded as 1 in the data file). The observed mean for males was 2.7835 and the observed mean for females was 2.9325. We seek the proportion of resamples where the absolute value of the difference was at least 0.149. The *P*-value should generally be between 0.25 and 0.32. **(b)** We test $H_0: \sigma_1/\sigma_2 = 1$ versus $H_a: \sigma_1/\sigma_2 \neq 1$. The observed ratio is $0.8593/0.7477 \doteq 1.149$. The *P*-value is generally between 0.235 and 0.270.

16.81 The bootstrap distribution looks quite Normal, and (as a consequence) all of the bootstrap confidence intervals are similar to each other, and also are similar to the standard (large-sample) confidence interval: 0.0981 to 0.1415.

16.83 (a) The standard test of $H_0 : \sigma_1 = \sigma_2$ versus $H_a : \sigma_1 \neq \sigma_2$ leads to F = 0.3443 with df 13 and 16; $P \doteq 0.0587$. **(b)** The permutation P-value is typically between 0.02 and 0.03. **(c)** The P-values are similar, even though technically, the permutation test is significant at the 5% level, while the standard test is (barely) not. Because the samples are too small to assess Normality, the permutation test is safer. (In fact, the population distributions are discrete, so they cannot follow Normal distributions.)

16.85 The 95% *t* interval and percentile interval (roughly 2.75 to 3.02) are roughly equal when rounded to two decimal places. In Example 16.8, the percentile interval was 2.793 to 3.095 (a bit higher and narrower than these) and the t interval was 2.80 to 3.10 (again, narrower). This is at least in part explained by eliminating more observations on either end with the 25% trim.

16.87 (a) The correlation for males is 0.4657. Because the bootstrap distribution does not look Normal, we'll focus on the percentile interval. The lower end of the percentile intervals ranged from 0.269 to 0.286, with the upper end ranging from 0.623 to 0.625. **(b)** The correlation for females is 0.3649. Again focusing on the percentile interval, the intervals are wider. The low end of the percentile interval ranged from 0.053 to 0.081 while the upper end ranged from 0.581 to 0.604. **(c)** The bootstrap distribution of the differences in correlations is very Normal. It is also clear that 0 will be included in the interval (the interval for one bootstrap set was (−0.2426, 0.4229)). All intervals examined had a low end between −0.22 and −0.24 with high end between about 0.40 and 0.42. We can conclude that there is no significant difference in the correlation between high school math grades and college GPA by gender.

16.89 (a) There were 32 poets, who died at an average age of 63.19 years (s = 17.30), with median age 68. There were 24 nonfiction writers, who died at an average age of 76.88 years (s = 14.10), with median age 77.5. Side-by-side boxplots clearly show that poets seem to die younger. Both distributions are somewhat left-skewed, and the non-fiction writer who died at age 40 is a low outlier. **(b)** Testing $H_0 : \mu_N = \mu_P$ against $H_a : \mu_N \neq \mu_P$ gives t = 3.26 with P = 0.002 (df = 53). A 95% confidence interval for the difference in mean ages is (5.27, 22.11). **(c)** The bootstrap distribution is symmetric, and seems close to Normal, except at the ends of the distribution, so a bootstrap t interval should be appropriate. The low ends of the bootstrap interval are typically between 5.26 and 5.82; the high ends are typically between 21.26 and 22.15. One particular interval seen was (5.40, 21.36). Note this interval is a bit narrower than the two-sample *t* interval.

16.91 The permutation test for the mean ages returns a P-value of 0.006 (comparable to the 0.002 from the t test). The 99% confidence interval for the P-value is between 0.0002 and 0.0185. We can determine there is a difference in mean age at death between poets and nonfiction writers.

16.93 All answers (including the shape of the bootstrap distribution) will depend strongly on the initial sample of uniform random numbers. The median M of these initial samples will be between about 0.36 and 0.64 about 95% of the time; this is the center of the bootstrap *t* confidence interval. **(a)** For a uniform distribution on 0 to 1, the population median is 0.5. Most of the time, the bootstrap distribution is quite non-Normal **(b)** SE_{boot} typically ranges from about 0.04 to 0.12 (but may vary more than that, depending on the original sample). The bootstrap *t* interval is therefore roughly $M \pm 2SE_{boot}$. **(c)** The more sophisticated BCa and tilting intervals may or may not be similar to the bootstrap *t* interval. The *t* interval is not appropriate because of the non-Normal shape of the bootstrap distribution, and because SE_{boot} is unreliable for the sample median (it depends strongly on the sizes of the gaps between the observations near the middle).

16.95 See Exercise 8.55 for more details about this survey. The bootstrap distribution appears to be close to Normal; bootstrap intervals are similar to the large-sample interval (0.3146 to 0.3854).

16.97 (a) This is the usual way of computing percent change: $89/54 - 1 = 0.65$. **(b)** Subtract 1 from the confidence interval found in Exercise 16.94; this typically gives an interval similar to 0.55 to 0.75.

16.99 (a) The mean ratio is 1.0596; the usual *t* interval is $1.0596 \pm (2.262)(0.02355) \doteq 1.0063$ to 1.1128. The bootstrap distribution for the mean is close to Normal, and the bootstrap confidence intervals are usually similar to the usual *t* interval, but slightly narrower. Bootstrapping the median produces a clearly non-Normal distribution; the bootstrap *t* interval should not be used for the median. **(b)** The ratio of means is 1.0656; the bootstrap distribution is noticeably skewed, so the bootstrap *t* is not a good choice, but the other methods usually give intervals similar to 0.75 to 1.55. **(c)** For example, the usual *t* interval from part (a) could be summarized by the statement, “On average, Jocko’s estimates are 1% to 11% higher than those from other garages.”

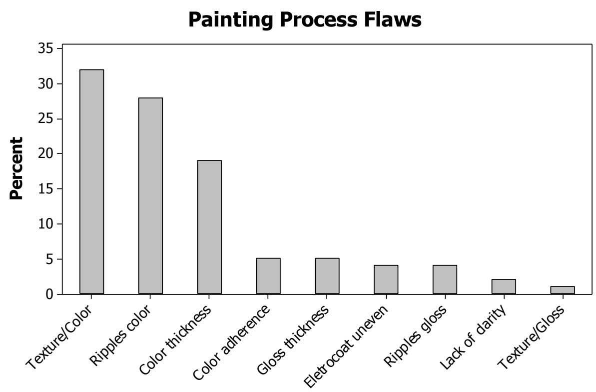
CHAPTER 17

17.1 Answers will vary.

17.3 Answers will vary.

17.5 The center line is at μ = 90 seconds. The control limits should be at about 57.8 and 122.2 seconds.

17.9 The most common problems are related to the application of the color coat; that should be the focus of our initial efforts.



17.11 Possible causes could include delivery delays due to traffic or a train, high demand during special events, and so forth.

17.13 (a) For n = 10, the new control limits for the $\bar{x}$ chart are 1.905 and 2.095 inches (the center line will not change). The new center line for the s chart is 0.09727 inch and the control limits will be $B_5\sigma =$ 0.076 inch and $B_6\sigma = 0.1669$ inches. **(b)** For n = 2, the new control limits for the $\bar{x}$ chart are 1.7879 and 2.212 inches. The center line for the s chart will be 0.07979 inch with control limits 0 and 0.2606 inch. **(c)** There are 2.54 centimeters to an inch, so the centerline for the $\bar{x}$ chart is 5.08 cm with control limits 4.699 and 5.461 cm. The new center line for the s chart will be 0.234 cm with control limits 0 and 0.5304 cm.

17.15 (a) For the x chart, the center line is at μ = 1.014 lb; the control limits should be about 0.9811 and 1.0469 lb. **(b)** The center line for the s chart 0.01684 lb, and the control limits are 0 and 0.04324 lb. **(d)** The s chart is in control, but there were signals on the $\bar{x}$ chart at samples 10 and 12.

17.17 (a) The center line is at μ = 11.5 Kp; the control limits should be at 11.2 and 11.8 Kp. **(c)** Set B is from the in-control process. The process mean shifted suddenly for Set A; it appears to have changed on about the 11th or 12th sample. The mean drifted gradually for the process in Set C.

17.19 For the s chart with n = 6, we have c_4 = 0.9515, B_5 = 0.029 and B_6 = 1.874, so the center line is 0.0009515 inch, and the control limits are 0.000029 and 0.001874 inch. For the $\bar{x}$ chart, the center line is μ = 0.87 inch, and the control limits are 0.8688 and 0.8712 inch.

17.21 For the $\bar{x}$ chart, the center line is 43, and the control limits are 25.91 and 60.09. The center line for the s chart is 11.9756, and the control limits are 0 and 25.02. The control charts show that sample 5 was above the UCL on the s chart, but it appears to have been special cause variation, as there is no indication that the samples that followed it were out of control.

17.23 (a) The process mean is the same as the center line: $\mu = 715$. The control limits are three standard errors from the mean, so $\sigma \doteq 23.3333$. **(b)** If the mean changes to $\mu = 700$, then $\bar{x}$ is approximately Normal with mean 700 and standard deviation $\sigma/\sqrt{4} \doteq 11.6667$, so $\bar{x}$ will fall outside the control limits with probability $1 - P(680 < \bar{x} < 750) = 0.0436$. **(c)** With $\mu = 700$ and $\sigma = 10$, $\bar{x}$ is approximately Normal with mean 700 and standard deviation $\sigma/\sqrt{4} = 5$, so $\bar{x}$ will fall outside the control limits with probability $1 - P(680 < x < 750) \doteq 1$.

17.25 The usual 3σ limits are $\mu \pm 3\sigma/\sqrt{n}$ for an $\bar{x}$ chart and $(c_4 \pm 3c_5)\sigma$ for an s chart. For 2σ limits, simply replace “3” with “2.” **(a)** $\mu \pm 2\sigma/\sqrt{n}$. **(b)** $(c_4 \pm 2c_5)\sigma$.

17.27 (a) Shrinking the control limits would increase the frequency of false alarms, because the probability of an out-of-control point when the process is in control will be higher (roughly 5% instead of 0.3%). **(b)** Quicker response comes at the cost of more false alarms. **(c)** The runs rule is better at detecting gradual changes. (The one-point-out rule is generally better for sudden, large changes.)

17.29 We estimate $\hat{\sigma}$ to be $s/0.9213 \doteq 1.1180$, so the x chart has center line $x = 47.2$ and control limits $x \pm 3\ \hat{\sigma}/\sqrt{4} \doteq 45.523$ and 48.877. The s chart has center line $s = 1.03$ and control limits 0 and $2.088\ \hat{\sigma} \doteq 2.3344$.

17.31 (a) The centerline for the s chart will be at $\bar{s} = 0.09325$. In Exercise 17.12, we had $n = 4$, so $c_4 = 0.9213$ and $\hat{\sigma} = 0.10122$. We’ll have UCL = 2.088(0.10122) = 0.21135 and LCL = 0. **(b)** None of the sample standard deviations were greater than 0.21135, so variability is under control. **(c)** The centerline for the $\bar{x}$ chart is $\bar{\bar{x}} \doteq 2.0078$. The limits are 1.85597 to 2.15963 inches. **(d)** These limits are slightly higher than those found in Exercise 17.12 (1.85 to 2.15 inches).

17.33 Sketches will vary.

17.35 (a) Average the 20 sample means and standard deviations and estimate μ to be $\hat{\mu} = \bar{\bar{x}} = 2750.7$ and σ to be $\hat{\sigma} = \bar{s}/c_4 = 345.5/0.9213$.= 375.0. **(b)** In the s chart shown in Figure 17.7, most of the points fall below the center line.

17.37 If the manufacturer practices SPC, that provides some assurance that the phones are roughly uniform in quality—as the text says, “We know what to expect in the finished product.” So, assuming that uniform quality is sufficiently high, the purchaser does not need to inspect the phones as they arrive because SPC has already achieved the goal of that inspection: to avoid buying many faulty phones. (Of course, a few unacceptable phones may be produced and sold even when SPC is practiced—but inspection would not catch all such phones anyway.)

17.39 The quantile plot does not suggest any serious deviations from Normality, so the natural tolerances should be reasonably trustworthy.

17.41 If we shift the process mean to 2500 mm, about 99% will meet the new specifications.

17.43 The mean of the 17 in-control samples is $\bar{\bar{x}} = 43.4118$, and the standard deviation is 11.5833, so the natural tolerances are 8.66 to 78.16.

17.45 Only about 44% of meters meet the specifications.

17.47 The limited precision of the measurements shows up in the granularity (stair-step appearance) of the graph. Aside from this, there is no particular departure from Normality.

17.49 The quantile plot, while not perfectly linear, does not suggest any serious deviations from Normality, so the natural tolerances should be reasonably trustworthy.

17.51 (a) (ii) A sudden change in the $\bar{x}$ chart: This would immediately increase the amount of time required to complete the checks. **(b)** (i) A sudden change (decrease) in *s* or *R* because the new measurement system will remove (or decrease) the variability introduced by human error. **(c)** (iii) A gradual drift in the $\bar{x}$ chart (presumably a drift up, if the variable being tracked is the length of time to complete a set of invoices).

17.53 The process is no longer the same as it was during the downward trend (from the 1950s into the 1980s). In particular, including those years in the data used to establish the control limits results in a mean that is too high to use for current winning times, and a standard deviation that includes variation attributable to the "special cause" of the changing conditioning and professional status of the best runners. Such special cause variation should not be included in a control chart.

17.55 LSL and USL are specification limits on the individual observations. They are *specified* as desired output levels, rather than being *computed* based on observation of the process. LCL and UCL are control limits for the averages of samples drawn from the process. The purpose of control limits is to detect whether the process is functioning "as usual," while specification limits are used to determine what percentage of the outputs meet certain specifications (are acceptable for use).

17.57 For computing $\hat{C}_{pk}$, note that the estimated process mean (2750.7 mm) lies closer to the USL. **(a)** $\hat{C}_p \doteq 1.3028$ and $\hat{C}_{pk} \doteq 1.0850$. **(b)** $\hat{C}_p \doteq 0.8685$ and $\hat{C}_{pk} \doteq 0.6508$.

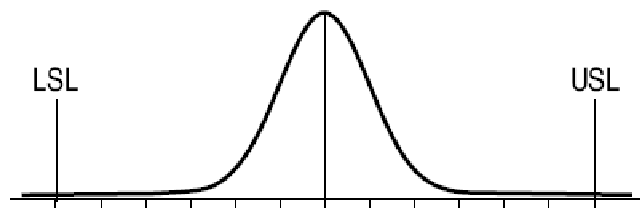
17.59 (a) $\hat{C}_p \doteq 1.1901$ and $\hat{C}_{pk} \doteq 1.0418$. (These were computed with the unrounded values of $\bar{x}$ and *s*; rounding will produce slightly different results.) **(b)** Customers typically will not complain about a package that was too heavy.

17.61 (a) $C_{pk} \doteq 0.5767$. 50% of the output meets the specifications. **(b)** LSL and USL are 0.865 standard deviations above and below to mean, so the proportion meeting specifications is $P(-0.865 < Z < 0.865) \doteq 0.6130$. **(c)** The relationship between C_{pk} and the proportion of the output meeting specifications depends on the shape of the distribution.

17.63 See also the solution to Exercise 17.47. **(a)** Use the mean and standard deviation of the 85 remaining observations: $\hat{\mu} = \bar{x} = 43.4118$ and $\hat{\sigma} = s = 11.5833$. **(b)** $\hat{C}_p \doteq 0.2878$ and $\hat{C}_{pk} = 0$. This process has very poor capability: The mean is too low and the spread too great. Only about 46% of the process output meets specifications.

17.65 We have $\bar{x} = 7.996$ mm and $s = 0.0023$ mm, so we assume that an individual bearing diameter X follows a $N(7.996, 0.0023)$ distribution. **(a)** About 91.8% meet specifications. **(b)** $\hat{C}_{pk} \doteq 0.5797$.

17.67 This graph shows a process with Normal output and C_p = 2. The tick marks are σ units apart; this is called "six-sigma quality" because the specification limits are (at least) six standard deviations above and below the mean.



17.69 For example, choosing six calls per shift gives an idea of the variability and mean for the shift as a whole.

17.71 The outliers are 276 seconds (sample 28), 244 seconds (sample 42), and 333 seconds (sample 46). After dropping those outliers, the standard deviations drop to 9.284, 6.708, and 31.011 seconds. (Sample #39, the other out-of-control point, has two moderately large times, 144 and 109 seconds; if they are removed, s drops to 3.416.)

17.73 (a) For those 10 months, there were 957 overdue invoices out of 26,350 total invoices (opportunities), so $\bar{p} = \frac{957}{26,350} = 0.03632$. **(b)** The center line and control limits are: CL = $\bar{p}$ = 0.03632, control limits: 0.02539 and 0.04725.

17.75 The center line is at the historical rate (0.0231); the control limits are 0.00295 and 0.04325.

17.77 The center line is at $\bar{p} = \frac{194}{38370} = 0.00506$; the control limits should be at −0.0004 (use 0) and 0.01052.

17.79 (a) The student counts sum to 9218, while the absentee total is 3277, so $\bar{p}$ = 0.3555 and n = 921.8. **(b)** The center line is p = 0.3555, and the control limits are:= 0.3082 and 0.4028. The p chart suggests that absentee rates are in control. **(c)** For October, the limits are 0.3088 and 0.4022; for June, they are 0.3072 and 0.4038.

17.81 (a) $\bar{p}$ = 0.008. We expect about 4 = defective orders per month. **(b)** The center line and control limits are: CL = $\bar{p}$ = 0.008, control limits −0.00395 and 0.01995 (We take the lower control limit to be 0.) It takes at least ten bad orders in a month to be out of control because (500)(0.01995) = 9.975.

17.83 (a) The percents do not add to 100% because one customer might have several complaints; that is, top priority should be given to the process of creating, correcting, and adjusting invoices, as the three most common complaints involved invoices.

17.85 Points above the UCL on an $\bar{x}$ (s) chart suggest that the process mean (standard deviation) may have increased. For the mean, an increase might signal a need to recalibrate the process in order to keep meeting specifications (that is, to bring the process back into control). An increase in the standard deviation typically does not indicate that adjustment or recalibration is necessary, but it will require re-computation of the $\bar{x}$ chart control limits.

17.87 We find that $\bar{s}$ = 7.65, so with c_4 = 0.8862 and B_6 = 2.276, we compute $\hat{\sigma}$ = 8.63 and UCL = 19.65. One point (from sample #1) is out of control.

17.89 (a) $C_p = 1.1423$. This is a fairly small value of C_p; if the mean wanders too far from 830, the capability will drop. **(b)** If we adjust the mean to be close to 830 mm × 10^{-4} (the center of the specification limits), we will maximize C_{pk}. C_{pk} is more useful when the mean is not in the center of the specification limits. **(c)** The value of $\hat{\sigma}$ used for determining C_p was estimated from the values of s from our control samples. These are for estimating short-term variation (within those samples) rather than the overall process variation.

17.91 (a) Use a p chart, with center line $\bar{p} = \frac{15}{5000} = 0.003$ and control limits 0 to 0.0194. **(b)** There is little useful information to be gained from keeping a p chart: If the proportion remains at 0.003, about 74% of samples will yield a proportion of 0, and about 22% of proportions will be 0.01. To call the process out of control, we would need to see two or more unsatisfactory films in a sample of 100.

17.93 Several interpretations of this problem are possible, but for most reasonable interpretations, the probability is about 0.3%.