Chapter 6 Solutions

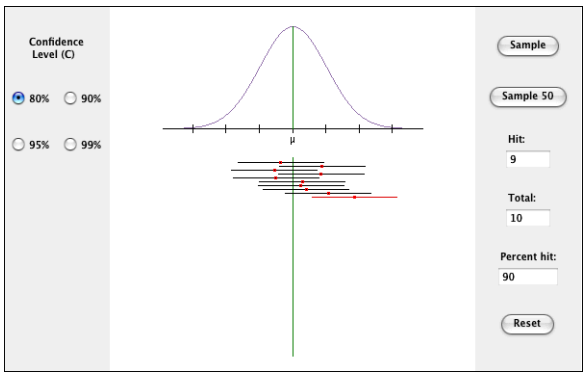
6.1. $\sigma_{\bar{x}} = \sigma/\sqrt{36} = \frac{\$2}{6} \doteq \$0.33$.

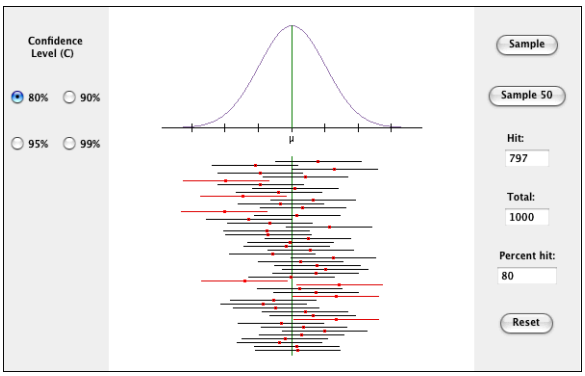
6.2. Take two standard deviations: $\frac{\$2}{3} \doteq \$0.67$.

6.3. As in the previous solution, take two standard deviations: $\frac{\$2}{3} \doteq \$0.67$.

Note: *This is the whole idea behind a confidence interval: Probability tells us that $\bar{x}$ is usually close to μ. That is equivalent to saying that μ is usually close to $\bar{x}$.*

6.4. Shown below are sample output screens for (a) 10 and (b) 1000 SRSs. In 99.4% of all repetitions of part (a), students should see between 5 and 10 hits (that is, at least 5 of the 10 SRSs capture the true mean μ). Out of 1000 80% confidence intervals, nearly all students will observe between 76% and 84% capturing the mean.





6.5. The standard error is $s_{\bar{x}} = \frac{s}{\sqrt{100}} = 0.4$, and the 95% confidence interval is $83.5 \pm (1.96)(4/\sqrt{100}) = 83.5 \pm 0.7840 = 82.7160$ to 84.2840.

6.6. A 99% confidence interval would have a larger margin of error; a wider interval is needed in order to be more confident that the interval includes the true mean. The 99% confidence interval is $83.5 \pm (2.576)(4/\sqrt{100}) = 83.5 \pm 1.03040 = 82.4696$ to 84.5304.

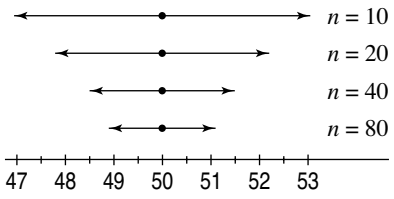
6.7. $n = \left(\frac{1.96 \times \$10{,}500}{\$900}\right)^2 \doteq 522.88$—take $n = 523$.

6.8. The margin of error $(1.96 \times \frac{\$10{,}500}{\sqrt{n}})$ would be *smaller* than \$900 with a *larger* sample, and *larger* with a *smaller* sample. If 1000 respond, the margin of error would be $(1.96)(\$332.04) \doteq \$650.80$; if $n = 450$, the margin of error would be $(1.96)(\$494.97) \doteq \$970.15$.

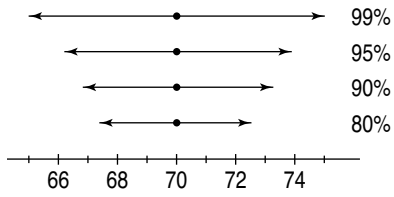
6.9. The response rate is $\frac{1468}{13000} \doteq 0.1129$, or about 11.3%. The reported margin of error is probably unreliable because we know nothing about the 88.7% of organizations that did *not* respond; they may be more (or less) male-dominated than those that responded.

6.10. (a) The 95% confidence interval is $93 \pm 11 = 82$ to 104. (The sample size is not needed.) **(b)** Greater than 11: A wider margin of error is needed in order to be more confident that the interval includes the true mean.

6.11. The margins of error are $1.96 \times 5/\sqrt{n}$, which yields 3.0990, 2.1913, 1.5495, and 1.0957. (And, of course, all intervals are centered at 50.) Interval width decreases with increasing sample size.



6.12. The margins of error are $z^* \times 12/\sqrt{36} = 2z^*$. With z^* equal to 1.282, 1.645, 1.960, and 2.576, this yields 2.564, 3.290, 3.920, and 5.152. (And, of course, all intervals are centered at 70.) Increasing confidence makes the interval wider.



6.13. To estimate the mean importance of recreation to college satisfaction, the 95% confidence interval is $7.5 \pm (1.96)(4.1/\sqrt{2673}) = 7.5 \pm 0.1554 = 7.3446$ to 7.6554.

6.14. Greater confidence requires a wider interval: The 99% confidence interval is $7.5 \pm (2.576)(4.1/\sqrt{2673}) = 7.5 \pm 0.2043 = 7.2957$ to 7.7043.

6.15. To estimate the mean importance of professor quality and interaction, the 95% confidence interval is $8.7 \pm (1.96)(3.5/\sqrt{2673}) = 8.7 \pm 0.1327 = 8.5673$ to 8.8327.

6.16. We must assume that the 2673 students were chosen as an SRS (or something like it). The non-Normality of the distribution is not a problem; we have a very large sample, so the central limit theorem applies.

6.17. For mean TRAP level, the margin of error is 2.29 U/l and the 95% confidence interval is $13.2 \pm (1.96)(6.5/\sqrt{31}) = 13.2 \pm 2.29 = 10.91$ to 15.49 U/l.

6.18. For mean OC level, the 95% confidence interval is $33.4 \pm (1.96)(19.6/\sqrt{31}) = 33.4 \pm 6.90 = 26.50$ to 40.30 ng/ml.

6.19. Scenario B has a smaller margin of error. Both samples would have the same value of z^* (1.96), but the value of σ would be smaller for (B) because we would have less variability in textbook cost for students in a single major.

Note: *Of course, at some schools, taking a sample of 100 sophomores in a given major is not possible.*

6.20. For the mean monthly rent for unfurnished one-bedroom apartments in Boston, the 95% confidence interval is $\$1400 \pm (1.96)(\$220/\sqrt{10}) = \$1400 \pm \$136.36 = \$1263.64$ to \$1536.36.

6.21. No; This is a range of values for the mean rent, not for individual rents.

Note: *To find a range to include 95% of all rents, we should take* $\mu \pm 2\sigma$ *(or more precisely,* $\mu \pm 1.96\sigma$*), where* μ *is the (unknown) mean rent for all apartments, and* σ *is the standard deviation for all apartments (assumed to be \$220 in Exercise 6.20). If* μ *were equal to \$1400, for example, this range would be \$960 to \$1840. However, because we do not actually know* μ*, we estimate it using* $\bar{x}$*, and to account for the variability in* $\bar{x}$*, we must widen the margin of error by a factor of* $\sqrt{1 + \frac{1}{n}}$*. The formula* $\bar{x} \pm 2\sigma\sqrt{1 + \frac{1}{10}}$ *is called a* prediction interval *for future observations. (Usually, such intervals are constructed with the t distribution, discussed in the Chapter 7, but the idea is the same.)*

6.22. If the distribution were roughly Normal, the 68–95–99.7 rule says that 68% of all measurements should be in the range 13.8 to 53.0 ng/ml, 95% should be between -5.8 and 72.6 ng/ml, and 99.7% should be between -25.4 and 92.2 ng/ml. Because the measurements cannot be negative, this suggests that the distribution must be skewed to the right. The Normal confidence interval should be fairly accurate nonetheless because the sample mean $\bar{x}$ will still be roughly Normal because of the central limit theorem.

6.23. **(a)** For the mean number of hours spent on the Internet, the 95% confidence interval is $15.1 \pm (1.96)(5/\sqrt{1200}) = 15.1 \pm 0.2829 = 14.8171$ to 15.3829 hours. **(b)** No; this is a range of values for the mean time spent, not for individual times. (See also the comment in the solution to Exercise 6.21.)

6.24. **(a)** To change from hours to minutes, multiply by 60: $\bar{x}_m = 60\bar{x}_h = 906$ and $\sigma_m = 60\sigma_h = 300$ minutes. **(b)** For mean time in minutes, the 95% confidence interval is $906 \pm (1.96)(300/\sqrt{1200}) = 906 \pm 16.9741 = 889.0259$ to 922.9741 minutes. **(c)** This interval can be found by multiplying the previous interval (14.8171 to 15.3829 hours) by 60.

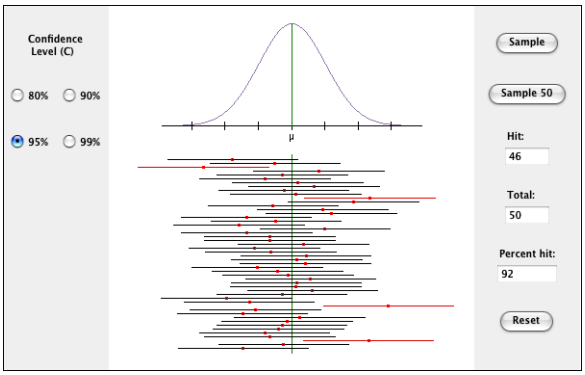
6.25. **(a)** The standard deviation of $\bar{x}$ is $\sigma_{\bar{x}} = \sigma/\sqrt{200} \doteq 2.1920$ cal/day. **(b)** The probability is about 0.95 that $\bar{x}$ is within 4.3840 cal/day (two standard deviations) of the population mean μ. **(c)** About 95% of all samples will capture the true mean of calories consumed per day in the interval $\bar{x} \pm$ 4.3840 cal/day. Some students may use 1.96 rather than 2 and answer 4.2964. (See also the comment in the solution to Exercise 6.3.)

6.26. **(a)** The standard deviation of the mean is $\sigma_{\bar{x}} = \frac{3.5}{\sqrt{20}} \doteq 0.7826$ mpg. **(b)** For the mean fuel efficiency, the 95% confidence interval is $43.17 \pm (1.96)(3.5/\sqrt{20}) = 43.17 \pm 1.5339 =$ 41.6361 to 44.7039 mpg.

6.27. Multiply by $\frac{1.609 \text{ km}}{1 \text{ mile}} \cdot \frac{1 \text{ gallon}}{3.785 \text{ liters}} \doteq 0.4251 \frac{\text{kpl}}{\text{mpg}}$. This gives $\bar{x}_{\text{kpl}} = 0.4251\bar{x}_{\text{mpg}} \doteq$ 18.3515 and margin of error $1.96(0.4251\sigma_{\text{mpg}})/\sqrt{20} \doteq 0.6521$ kpl, so the 95% confidence interval is 17.6994 to 19.0036 kpl.

6.28. One sample screen is shown below, along with a sample stemplot of results. The number of hits will vary, but the distribution should follow a binomial distribution with $n = 50$ and $p = 0.95$, so we expect the average number of hits to be about 47.5. We also find that about 99.7% of individual counts should be 43 or more, and the mean hit count for 30 samples should be approximately Normal with mean 47.5 and standard deviation 0.2814—so almost all sample means should be between 46.66 and 48.34.

44	00
45	0000
46	00
47	00000000
48	000000
49	000
50	00000



6.29. $n = \left(\frac{1.96 \times 6.5}{2.0}\right)^2 \doteq 40.58$—take $n = 41$.

6.30. If we start with a sample of size k and lose 20% of the sample, we will end with $0.8k$. Therefore, we need to *increase* the sample size by *25%*—that is, start with a sample of size $k = 1.25n$—so that we end with $(0.8)(1.25n) = n$. With $n = 41$, that means we should initially sample 52 subjects.

6.31. $n = \left(\frac{1.645 \times \$220}{\$50}\right)^2 \doteq 52.39$—take $n = 53$.

6.32. **(a)** For the mean of all repeated measurements, the 98% confidence interval is $10.0023 \pm (2.326)(0.0002/\sqrt{5}) = 10.0023 \pm 0.0002 = 10.0021$ to 10.0025 g.
(b) $n = \left(\frac{(2.326)(0.0002)}{0.0001}\right)^2 \doteq 21.64$—take $n = 22$.

6.33. The number of hits has a binomial distribution with parameters $n = 5$ and $p = 0.95$, so the number of misses is binomial with $n = 5$ and $p = 0.05$. We can therefore use Table C to answer these questions. **(a)** The probability that all cover their means is $0.95^5 \doteq 0.7738$. (Or use Table C to find the probability of 0 misses.) **(b)** The probability that at least four intervals cover their means is $0.95^5 + 5(0.05)(0.95^4) \doteq 0.9774$. (Or use Table C to find the probability of 0 or 1 misses.)

6.34. This is probably not a confidence interval; it is not intended to give an estimate of the mean income, but rather it gives the range of incomes earned by all (or most) telemarketers working for this company.

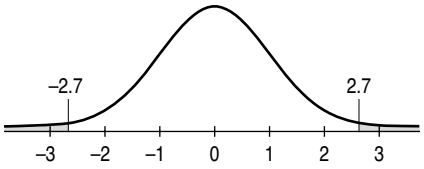
6.35. (a) We can be 95% confident, but not *certain*. **(b)** We obtained the interval 56% to 62% by a method that gives a correct result (that is, includes the true mean) 95% of the time. **(c)** For 95% confidence, the margin of error is about two standard deviations (that is, $z^* = 1.96$), so $\sigma_{\text{estimate}} \doteq 1.5\%$. **(d)** No; confidence intervals only account for random sampling error.

6.36. The new design can be considered an improvement if the mean response to the survey is greater than 4. The null hypothesis should be H_0: $\mu = 4$; the alternative hypothesis could be either $\mu > 4$ or $\mu < 4$. The first choice would be appropriate if we want the default assumption to be that the new design is *not* an improvement; that is, we will only conclude that the new design is better if we see compelling evidence to that effect. Choosing H_a: $\mu < 4$ would mean that the default assumption is that the new design is at least as good as the old one, and we will stick with that belief unless we see compelling evidence that it is worse.

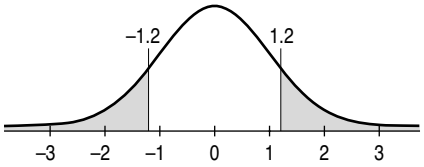
Students who are just learning about stating hypotheses might have difficulty choosing the alternative for this problem. In fact, either one is defensible, although the typical choice in such cases would be $\mu > 4$—that is, we give the benefit of the doubt to the old design and need convincing evidence that the new design is better.

6.37. If μ is the mean DXA reading for the phantom, we test H_0: $\mu = 1.4$ g/cm^2 vs. H_a: $\mu \neq 1.4$ g/cm^2.

6.38. $P(Z > 2.7) = 0.0035$, so the two-sided P-value is $2(0.0035) = 0.0070$.



6.39. $P(Z < -1.2) = 0.1151$, so the two-sided P-value is $2(0.1151) = 0.2302$.



6.40. For a two-sided alternative, z is statistically significant at $\alpha = 0.05$ if $|z| > 1.96$.

6.41. For a one-sided alternative (on the positive side), z is statistically significant at $\alpha = 0.05$ if $z > 1.645$.

6.42. For $z^* = 2$ the P-value would be $2P(Z > 2) = 0.0456$, and for $z^* = 3$ the P-value would be $2P(Z > 3) = 0.0026$.

Note: *In other words, the Supreme Court uses α no bigger than about 0.05.*

6.43. (a) $z = \frac{27-25}{5/\sqrt{25}} = 2$. **(b)** For a one-sided alternative, $P = P(Z > 2) = 0.0228$. **(c)** For a two-sided alternative, double the one-sided P-value: $P = 0.0456$.

6.44. The test statistic is $z = \frac{0.522 - 0.5}{0.316/\sqrt{100}} \doteq 0.7$. This gives very little reason to doubt the null hypothesis ($\mu = 0.5$); in fact, the two-sided P-value is $P \doteq 0.4839$.

6.45. Recall the statement from the text: "A level α two-sided significance test rejects ... H_0: $\mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ." **(a)** Yes, 30 is in the 95% confidence interval because $P = 0.08$ means that we would not reject H_0 at $\alpha = 0.05$. **(b)** No; 30 is not in the 90% confidence interval because we would reject H_0 at $\alpha = 0.10$.

6.46. **(a)** If the alternative is two-sided, the answer is *yes;* see the quote from the text in the solution to Exercise 6.45 for an explanation. If H_a is one-sided, then the answer depends on its direction: If H_a is $\mu > 68$, the answer is *no*. If H_a is $\mu < 68$, the answer is *yes* because the given interval suggests $\bar{x} = 61$ and the standard error of the mean is about 2 (or less if the interval were constructed using the t distribution rather than the Normal distribution), so $\bar{x}$ is 3.5 standard errors from 68. **(b)** Regardless of the alternative, we would not reject H_0 because 62 falls well inside the confidence interval.

6.47. **(a)** Yes, we reject H_0 at $\alpha = 0.05$. **(b)** No, we do not reject H_0 at $\alpha = 0.01$. **(c)** We have $P = 0.026$; to reject, we need $P < \alpha$.

6.48. **(a)** No, we do not reject H_0 at $\alpha = 0.05$. **(b)** No, we do not reject H_0 at $\alpha = 0.01$. **(c)** We have $P = 0.074$; to reject, we need $P < \alpha$.

6.49. **(a)** One of the one-sided P-values is half as big as the two-sided P-value (0.03); the other is $1 - 0.03 = 0.97$. **(b)** Suppose the null hypothesis is H_0: $\mu = \mu_0$. The smaller P-value (0.03) goes with the one-sided alternative that is consistent with the observed data; for example, if $\bar{x} > \mu_0$, then $P = 0.03$ for the alternative $\mu > \mu_0$.

6.50. **(a)** The standard deviation of the sample mean is $12/\sqrt{20}$. **(b)** The null hypothesis should be a statement about μ, not $\bar{x}$. **(c)** $\bar{x} = 48$ would not make us inclined to believe that $\mu > 54$. **(d)** Even if we fail to reject H_0, we are not sure that it is true.

Note: *That is, "not rejecting H_0" is different from "knowing that H_0 is true." This is the same distinction we make about a jury's verdict in a criminal trial: If the jury finds the defendant "not guilty," that does* not *necessarily mean that they are sure he/she is innocent. It simply means that they were not sufficiently convinced of his/her guilt.*

6.51. **(a)** Hypotheses should be stated in terms of the population mean, not the sample mean. **(b)** The null hypothesis H_0 should be that there is no change. **(c)** A small P-value is needed for significance; $P = 0.99$ gives no reason to reject H_0.

6.52. **(a)** We are checking to see if the proportion p increased, so we test H_0: $p = 0.7$ vs. H_a: $p > 0.7$. **(b)** The professor believes that the mean μ for the morning class will be higher, so we test H_0: $\mu = 72$ vs. H_a: $\mu > 72$. **(c)** Let μ be the mean student response (for the population of all students). We are trying to determine if students are neutral about the change, or if they have an opinion about it, with no preconceived idea about the direction of that opinion, so we test H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$.

6.53. **(a)** If μ is the mean score for the population of credit-by-exam students, then we test H_0: $\mu = 26$ vs. H_a: $\mu \neq 26$ because we have no prior belief about whether credit-by-exam students will do better or worse. **(b)** If μ is the mean time to complete the maze with rap music playing, then we test H_0: $\mu = 20$ seconds vs. H_a: $\mu < 20$ seconds because we believe rap music will make the mice finish more quickly. **(c)** If μ is the mean area of the apartments, we test H_0: $\mu = 460$ ft^2 vs. H_a: $\mu < 460$ ft^2, because we suspect the apartments are smaller than advertised.

6.54. **(a)** If p_m and p_f are the proportions of (respectively) males and females who like marketing best, we test H_0: $p_m = p_f$ vs. H_a: $p_m > p_f$. **(b)** If μ_A and μ_B are the mean test scores for each group, we test H_0: $\mu_A = \mu_B$ vs. H_a: $\mu_A > \mu_B$. **(c)** If ρ is the correlation between amount of credit card debt and self-esteem, we test H_0: $\rho = 0$ vs. H_a: $\rho < 0$.

Note: *In each case, the parameters identified refer to the respective* populations, *not the samples*.

6.55. **(a)** H_0: $\mu = \$62{,}500$; H_a: $\mu > \$62{,}500$. **(b)** H_0: $\mu = 2.6$ hr; H_a: $\mu \neq 2.6$ hr.

6.56. **(a)** For H_a: $\mu > \mu_0$, the P-value is $P(Z > 1.34) = 0.0901$.
(b) For H_a: $\mu < \mu_0$, the P-value is $P(Z < 1.34) = 0.9099$.
(c) For H_a: $\mu \neq \mu_0$, the P-value is $2P(Z > 1.34) = 2(0.0901) = 0.1802$.

6.57. **(a)** For H_a: $\mu > \mu_0$, the P-value is $P(Z > -1.73) = 0.9582$.
(b) For H_a: $\mu < \mu_0$, the P-value is $P(Z < -1.73) = 0.0418$.
(c) For H_a: $\mu \neq \mu_0$, the P-value is $2P(Z < -1.73) = 2(0.0418) = 0.0836$.

6.58. Recall the statement from the text: "A level α two-sided significance test rejects ... H_0: $\mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ." **(a)** No, 30 is not in the 95% confidence interval because $P = 0.04$ means that we would reject H_0 at $\alpha = 0.05$. **(b)** No, 30 is not in the 90% confidence interval because we would also reject H_0 at $\alpha = 0.10$.

6.59. **(a)** Regardless of the alternative, we would not reject H_0 because 24 falls well inside the confidence interval. **(a)** If the alternative is two-sided, the answer is *yes;* see the quote from the text in the solution to Exercise 6.58 for an explanation. If H_a is one-sided, then the answer depends on its direction: If H_a is $\mu > 30$, the answer is *no*. If H_a is $\mu < 30$, the answer is *yes* because the given interval suggests $\bar{x} = 25.5$ and the standard error of the mean is about 1.5 (or less, if the interval were constructed using the t distribution rather than the Normal distribution), so $\bar{x}$ is 3 standard errors from 30.

6.60. The study presumably examined malarial infection rates in two groups of subjects—one with bed nets and one without. The observed differences between the two groups were so large that they would be unlikely to occur by chance if bed nets had no effect. Specifically, if the groups were the same, and we took many samples, the difference in malarial infections would be so large less than 0.1% of the time.

6.61. $P = 0.27$ means that we have little reason to doubt that the purity is the same, but that does not mean we are sure that the purity levels are the same. See the comment in the solution to Exercise 6.50.

6.62. If the presence of pig skulls were not an indication of wealth, then differences similar to those observed in this study would occur less than 1% of the time by chance.

6.63. Even if the two groups (the health and safety class, and the statistics class) had the same level of alcohol awareness, there might be some difference in our sample due to chance. The difference observed was large enough that it would rarely arise by chance. The reason for this difference might be that health issues related to alcohol use are probably discussed in the health and safety class.

6.64. Even if scores had not changed over time, random fluctuation might cause the mean in 2005 to be different from the 2000 mean. However, in this case the difference was so great that it is unlikely to have occurred by chance; specifically, such a difference would arise less than 5% of the time if the actual mean had not changed. We therefore conclude that the mean did change from 2000 to 2005.

6.65. While there was some difference in average scores between 2000 and 2005, that difference was so small that it could have occurred by chance even if population mean scores had not changed in that time.

6.66. If μ is the mean north-south location, the hypotheses are H_0: $\mu = 100$ vs. H_a: $\mu \neq 100$. For testing these hypotheses, we find $z = \frac{99.74 - 100}{58/\sqrt{584}} \doteq -0.11$. This is not significant—$P = 2(0.4562) = 0.9124$—so we have no reason to doubt a uniform distribution based on this test.

6.67. If μ is the mean east-west location, the hypotheses are H_0: $\mu = 100$ vs. H_a: $\mu \neq 100$ (as in the previous exercise). For testing these hypotheses, we find $z = \frac{113.8 - 100}{58/\sqrt{584}} \doteq 5.75$. This is highly significant ($P < 0.0001$), so we conclude that the trees are not uniformly spread from east to west.

6.68. For testing these hypotheses, we find $z = \frac{10.2 - 8.9}{2.5/\sqrt{6}} \doteq 1.27$. This is not significant ($P = 0.1020$); there is not enough evidence to conclude that these sonnets were not written by our poet. (That is, we cannot reject H_0.)

6.69. (a) $z = \frac{132.2 - 115}{30/\sqrt{25}} \doteq 2.87$, so the P-value is $P = P(Z > 2.87) = 0.0021$. This is strong evidence that the older students have a higher SSHA mean. **(b)** The important assumption is that this is an SRS from the population of older students. We also assume a Normal distribution, but this is not crucial provided there are no outliers and little skewness.

6.70. (a) H_0: $\mu = 9.5$ mg/dl vs. H_a: $\mu \neq 9.5$ mg/dl. **(b)** $z = \frac{9.57 - 9.5}{0.4/\sqrt{160}} \doteq 2.21$, so the P-value is $P = 2P(Z > 2.21) = 0.0272$. This is pretty strong evidence that μ is different from (greater than) 9.5 mg/dl. **(c)** The 95% confidence interval is $9.57 \pm (1.96)(0.4/\sqrt{160}) = 9.57 \pm 0.062 = 9.508$ to 9.632.

6.71. **(a)** H_0: $\mu = 0$ mpg vs. H_a: $\mu \neq 0$ mpg, where μ is the mean difference. **(b)** The mean of the 20 differences is $\bar{x} = 2.73$, so $z = \frac{2.73 - 0}{3/\sqrt{20}} \doteq 4.07$, for which $P < 0.00005$. We conclude that $\mu \neq 0$ mpg; that is, we have strong evidence that the computer's reported fuel efficiency differs from the driver's computed values.

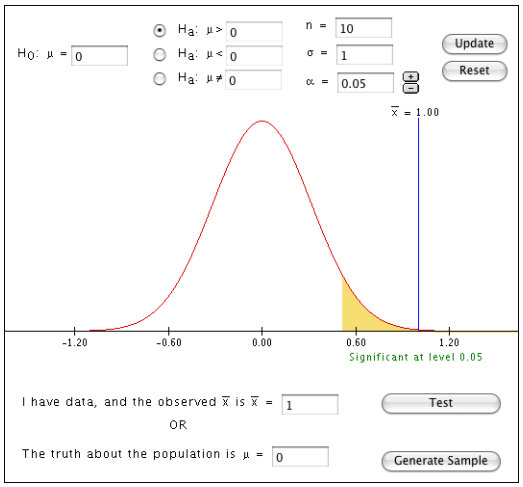
6.72. A debt of \$11,400 in 1997 would be equivalent to a debt of (\$11,400)(1.12) = \$12,768 in 2002, so the increase in debt is \$18,900 − \$12,768 = \$6132. The standard deviation (\$1900) and the details of the computation are shown in Example 6.14. With the same hypotheses given in that example, the test statistic is:

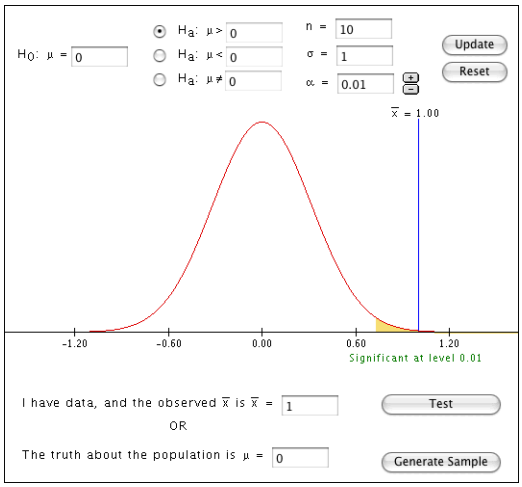
$$z = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard deviation}} = \frac{\$6132 - \$0}{\$1900} \doteq 3.23$$

The P-value is $P = P(Z \geq 3.23) \doteq 0.0006$; recall that, for the unadjusted data, the P-value was 0.00004. Not surprisingly, adjusting for inflation weakens the evidence against H_0, but it is still quite sufficient to conclude that college loan debt increased from 1997 to 2002.

6.73. For (b) and (c), either compare with the critical values in Table D or determine the P-value (0.0091). **(a)** H_0: $\mu = 1.4$ mg vs. H_a: $\mu > 1.4$ mg. **(b)** Yes, because $z > 1.645$ (or because $P < 0.05$). **(c)** Yes, because $z > 2.326$ (or because $P < 0.01$).

6.74. A sample screen (for $\bar{x} = 1$) is shown below on the left. As one can judge from the shading under the Normal curve, $\bar{x} = 0.5$ is not significant, but 0.6 is. (In fact, the cutoff is about 0.52, which is approximately $1.645/\sqrt{10}$.)

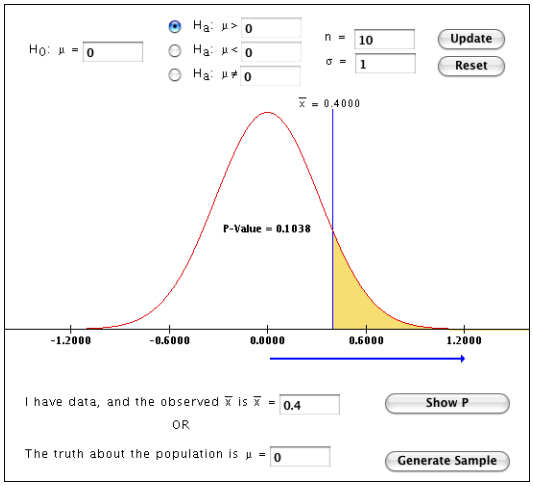




6.75. See the sample screen (for $\bar{x} = 1$) above on the right. As one can judge from the shading under the Normal curve, $\bar{x} = 0.7$ is not significant, but 0.8 is. (In fact, the cutoff is about 0.7354, which is approximately $2.326/\sqrt{10}$.) Smaller α means that $\bar{x}$ must be farther away from μ_0 in order to reject H_0.

6.76. A sample screen (for $\bar{x} = 0.4$) is shown on the right. The P-values given by the applet are listed in the table below; as $\bar{x}$ moves farther away from μ_0, P decreases.

$\bar{x}$	P	$\bar{x}$	P
0.1	0.3745	0.6	0.0287
0.2	0.2643	0.7	0.0136
0.3	0.1711	0.8	0.0057
0.4	0.1038	0.9	0.0022
0.5	0.0571	1	0.0008



6.77. When a test is significant at the 1% level, it means that if the null hypothesis were true, outcomes similar to those seen are expected to occur less than once in 100 repetitions of the experiment or sampling. "Significant at the 5% level" means we have observed something that occurs in fewer than 5 out of 100 repetitions (when H_0 is true). Something that occurs "less than once in 100 repetitions" also occurs "fewer than 5 times in 100 repetitions," so significance at the 1% level implies significance at the 5% level (or any higher level).

6.78. Something that occurs "fewer than 5 times in 100 repetitions" is not necessarily as rare as something that occurs "less than once in 100 repetitions," so a test that is significant at 5% is not necessarily significant at 1%.

6.79. Using Table D or software, we find that the 0.0025 critical value is 2.807, and the 0.0005 critical value is 3.291. Therefore, if $2.807 < |z| < 3.291$—that is, either $2.807 < z < 3.291$ or $-3.291 < z < -2.807$—then z would be significant at the 0.5% level, but not at the 0.1% level.

6.80. As $3.091 < 3.1 < 3.291$, the two-sided P-value is between $2(0.001) = 0.002$ and $2(0.0005) = 0.001$. (Software tells us that $P \doteq 0.002$, consistent with the observation that z is close to 3.091.)

6.81. As $0.35 < 0.674$, the one-sided P-value is $P > 0.25$. (Software gives $P = 0.3632$.)

6.82. Because $1.282 < 1.37 < 1.645$, the P-value is between $2(0.05) = 0.10$ and $2(0.10) = 0.20$. From Table A, $P = 2(0.0853) = 0.1706$.

6.83. Because the alternative is two-sided, the answer is the same for $z = -1.37$: $-1.282 > -1.37 > -1.645$, so Table D says that $0.10 < P < 0.20$, and Table A gives $P = 2(0.0853) = 0.1706$.

6.84. **(a)** $z = \frac{541.4 - 525}{100/\sqrt{100}} = 1.64$. This is not significant at $\alpha = 0.05$ because $z < 1.645$ (or $P = 0.0505$). **(b)** $z = \frac{541.5 - 525}{100/\sqrt{100}} = 1.65$. This *is* significant at $\alpha = 0.05$ because $z > 1.645$ (or $P = 0.0495$). **(c)** Fixed-level significance tests require that we draw a line

between "significant" and "not significant"; in this example, we see evidence on each side of that line. The 5% significance level is a guideline, not a sacred edict. P-values are more informative ways to convey the strength of the evidence.

6.85. In order to determine the effectiveness of alarm systems, we need to know the percent of all homes with alarm systems, and the percent of burglarized homes with alarm systems. For example, if only 10% of all home have alarm systems, then we should compare the proportion of burglarized homes with alarm systems to 10%, not 50%.

6.86. Finding something to be "statistically significant" is not really useful unless the significance level is sufficiently small. While there is some freedom to decide what "sufficiently small" means, $\alpha = 0.25$ would lead the student to incorrectly reject H_0 one-fourth of the time, so it is clearly a bad choice.

6.87. The first test was barely significant at $\alpha = 0.05$, while the second was significant at any reasonable α.

6.88. $P = 0.90$ means that we have no reason to doubt the null hypothesis. Specifically, results like those observed in the sample would occur 90% of the time when H_0 is true.

6.89. A significance test answers only Question b. The P-value states how likely the observed effect (or a stronger one) is if H_0 is true, and chance alone accounts for deviations from what we expect. The observed effect may be significant (very unlikely to be due to chance) and yet not be of practical importance. And the calculation leading to significance *assumes* a properly designed study.

6.90. The study may have rejected $\mu = \mu_0$ (or some other null hypothesis), but with such a large sample size, such a rejection might occur even if the actual mean (or other parameter) differs only slightly from μ_0. For example, there might be no practical importance to the difference between $\mu = 10$ and $\mu = 10.5$.

6.91. We expect more variation with small sample sizes, so even a large difference between $\bar{x}$ and μ_0 (or whatever measures are appropriate in our hypothesis test) might not turn out to be significant. If we were to repeat the test with a larger sample, the decrease in the standard error might give us a small enough P-value to reject H_0.

6.92. Based on the description, this seems to have been an experiment (not just an observational study), so a statistically significant outcome suggests that vitamin C is effective in preventing colds.

6.93. (a) If SES had no effect on LSAT results, there would still be some difference in scores due to chance variation. "Statistically insignificant" means that the observed difference was no more than we might expect from that chance variation. **(b)** If the results are based on a small sample, then even if the null hypothesis were not true, the test might not be sensitive enough to detect the effect. Knowing the effects were small tells us that the statistically insignificant test result did not occur merely because of a small sample size.

6.95. **(a)** $z = \frac{508-505}{100/\sqrt{100}} = 0.3$, so $P = P(Z > 0.3) = 0.3821$. **(b)** $z = \frac{508-505}{100/\sqrt{1000}} \doteq 0.95$, so $P = P(Z > 0.95) = 0.1711$. **(c)** $z = \frac{508-505}{100/\sqrt{10000}} = 3$, so $P = P(Z > 3) = 0.0013$.

6.96. The interval is $508 \pm (2.576)(100/\sqrt{n})$. $n = 100$: 482.24 to 533.76. $n = 1000$: 499.85 to 516.15. $n = 10,000$: 505.42 to 510.58.

6.97. With all 50 states (plus Puerto Rico) listed in the table, we have information about the entire population in question; no statistical procedures are needed (or meaningful).

6.100. When many variables are examined, "significant" results will show up by chance, so we should not take it for granted that the variables identified are really indicative of future success. In order to decide if they are appropriate, we should track this year's trainees and compare the success of those from urban/suburban backgrounds with the rest, and likewise compare those with a degree in a technical field with the rest.

6.102. We expect 50 tests to be statistically significant: Each of the 1000 tests has a 5% chance of being significant, so the number of significant tests has a binomial distribution with $n = 1000$ and $p = 0.05$, for which the mean is $np = 50$.

6.103. $P = 0.00001 = \frac{1}{100,000}$, so we would need $n = 100{,}000$ tests in order to expect one P-value of this size (assuming that all null hypotheses are true). That is why we reject H_0 when we see P-values such as this: It indicates that our results would rarely happen if H_0 were true.

6.104. Using $\alpha/6 \doteq 0.008333$ as the cutoff, the fourth ($P = 0.008$) and sixth ($P < 0.001$) tests are significant.

6.105. Using $\alpha/12 \doteq 0.004167$ as the cutoff, we reject the fifth ($P = 0.001$), sixth ($P = 0.004$), and eleventh ($P = 0.002$) tests.

6.106. The power of this study is far lower than what is generally desired—for example, it is well below the "80% standard" mentioned in the text. Twenty percent power for the specified effect means that, if the effect is present, we will only detect it 20% of the time. With such a small chance of detecting an important difference, the study should probably not be run (unless the sample size is increased to give sufficiently high power).

6.107. A larger sample gives more information and therefore gives a better chance of detecting a given alternative; that is, larger samples give more power.

6.108. The power for $\mu = -5$ is 0.82, the same as the power for $\mu = 5$ because both alternatives are an equal distance from the null value of μ. (The symmetry of two-sided tests with the Normal distribution means that we only need to consider the size of the difference, not the direction.)

6.109. The power for $\mu = 70$ will be higher than 0.5, because larger differences are easier to detect.

6.110. The power is approximately 0.061.

Note: *At the time these solutions were prepared, the* Power *applet did not allow the sample size to be more than 100. For $n = 100$, the power is about 0.052.*

A search of the Web might reveal other power calculators. In the past, a good one was `http://calculators.stat.ucla.edu/powercalc`*, but at this time of this writing, that Web site was down "due to technical issues."*

6.111. The power is approximately 0.99. (See the note in the previous solution. For $n = 100$, the power is about 0.401.)

6.112. **(a)** $\alpha = P(\text{Type I error}) = P(\bar{x} > 26 \text{ when } \mu = 25)$

$$= P\left(Z > \tfrac{26-25}{50/\sqrt{900}}\right) = P(Z > 0.6) = 0.2743.$$

(b) $P(\text{Type II error when } \mu = 28) = P(\bar{x} \le 26 \text{ when } \mu = 28)$

$$= P\left(Z \le \tfrac{26-28}{50/\sqrt{900}}\right) = P(Z < -1.2) = 0.1151.$$

(c) $P(\text{Type II error when } \mu = 30) = P(\bar{x} < 26 \text{ when } \mu = 30)$

$$= P\left(Z \le \tfrac{26-30}{50/\sqrt{900}}\right) = P(Z < -2.4) = 0.0082.$$

(d) The sample size ($n = 900$) is so large that the mean will be very close to Normal.

6.113. $z \ge 2.326$ is equivalent to $\bar{x} \ge 450 + 2.326(100/\sqrt{500}) \doteq 460.4$, so the power is

$$P(\text{reject } H_0 \text{ when } \mu = 460) = P(\bar{x} \ge 460.4 \text{ when } \mu = 460)$$

$$= P\left(Z \ge \tfrac{460.4-460}{100/\sqrt{500}}\right) \doteq P(Z \ge 0.09) = 0.4641.$$

This is quite a bit less than the "80% power" standard.

6.114. **(a)** $P(\text{Type I error}) = P(X = 0 \text{ or } X = 1 \text{ when the distribution is } p_0) = 0.2.$

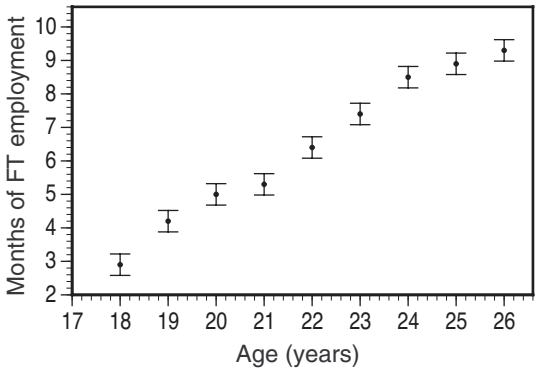
(b) $P(\text{Type II error}) = P(X > 1 \text{ when the distribution is } p_1) = 0.6.$

6.115. **(a)** H_0: The patient is healthy (or "the patient should not see a doctor"); H_a: The patient is ill (or "the patient should see a doctor"). A Type I error is a false positive—sending a healthy patient to the doctor. A Type II error means a false negative—clearing a patient who should be referred to a doctor. **(b)** One might wish to lower the probability of a false negative so that most ill patients are treated, especially for serious diseases that require fast treatment. On the other hand, if resources (for example, money or medical personnel) are limited, or for less serious health problems, lowering the probability of false positives might be desirable.

Note: *For (a), there is no clear choice for which should be the null hypothesis in this case; we could also choose H_0 to be "the patient is ill." This choice may also be affected by the factors considered in part (b).*

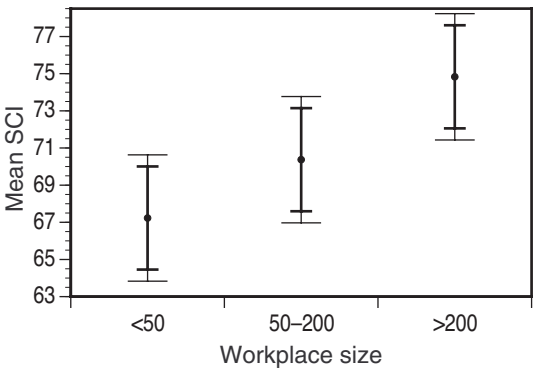
6.116. **(a)** Because all standard deviations and sample sizes are the same, the margin of error for all intervals is $(1.96)(4.5)/\sqrt{750} \doteq 0.3221$ months. The confidence intervals are listed in the table below. **(b)** Plot below. **(c)** The mean number of months of full-time employment shows a steady increase over time, at an average rate of about 0.8 months per year.

Age (years)	Average months worked
18	2.58 to 3.22
19	3.88 to 4.52
20	4.68 to 5.32
21	4.98 to 5.62
22	6.08 to 6.72
23	7.08 to 7.72
24	8.18 to 8.82
25	8.58 to 9.22
26	8.98 to 9.62



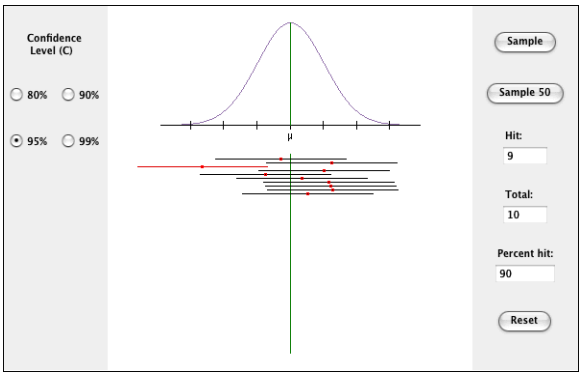
6.117. **(a)** Because all standard deviations and sample sizes are the same, the margin of error for all intervals is $(1.96)(19)/\sqrt{180} \doteq 2.7757$. The confidence intervals are listed in the table below. **(b)** The plot below shows the error bars for the confidence intervals of (a), and also for part (c). The limits for (a) are the thicker lines which do not extend as far above and below the mean. **(c)** With $z^* = 2.40$, the margin of error for all intervals is $(2.40)(19)/\sqrt{180} \doteq 3.3988$. The confidence intervals are listed in the table below and are shown in the plot (the thinner lines with the wider dashes). **(d)** When we use $z^* = 2.40$ to adjust for the fact that we are making three "simultaneous" confidence intervals, the margin of error is larger, so the intervals overlap more.

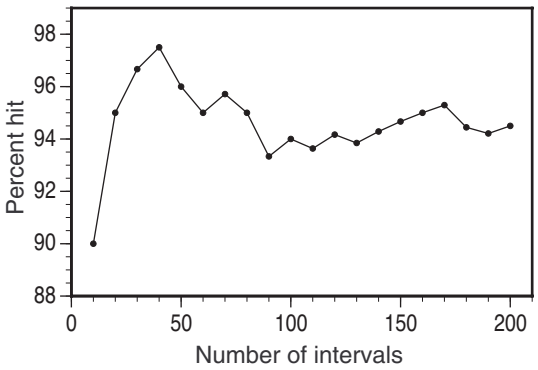
Workplace size	Mean SCI
< 50	64.45 to 70.01
50–200	67.59 to 73.15
> 200	72.05 to 77.61
< 50	63.83 to 70.63
50–200	66.97 to 73.77
> 200	71.43 to 78.23



6.118. Shown below is a sample screenshot from the applet and an example of what the resulting plot might look like. Most students (99.7% of them) should find that their final proportion is between 0.90 and 1; 90% will have a proportion between 0.925 and 0.975.

Note: *For each n (number of intervals), the number of "hits" would have a binomial distribution with $p = 0.95$, but these counts would not be independent; for example, if we knew there were 28 hits after 30 tries, we would know that there could be no more than 38 after 40 tries.*

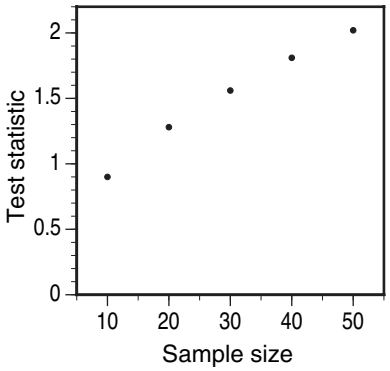


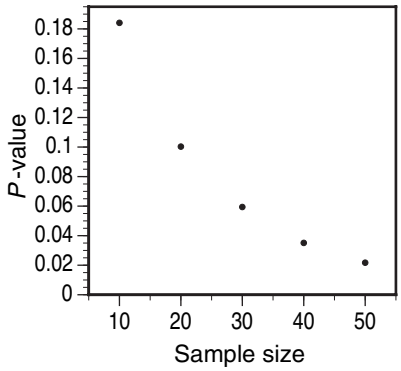


6.119. A sample screenshot and example plot are not shown but would be similar to those shown above for the previous exercise. Most students (99.4% of them) should find that their final proportion is between 0.84 and 0.96; 85% will have a proportion between 0.87 and 0.93.

6.120. For $n = 10$, $z = \frac{1-0}{14/\sqrt{10}} \doteq 0.90$, for which $P = 0.1841$. For the other sample sizes, the computations are similar; the resulting table and graphs are shown below. We see that sample size increases the value of the test statistic (assuming the mean is the same), which in turn decreases the size of the P-value.

n	z	P
10	0.90	0.1841
20	1.28	0.1003
30	1.56	0.0594
40	1.81	0.0351
50	2.02	0.0217





6.121. **(a)** $\bar{x} = 5.\overline{3}$ mg/dl, so $\bar{x} \pm 1.960\sigma/\sqrt{6}$ is 4.6132 to 6.0534 mg/dl. **(b)** To test H_0: $\mu = 4.8$ mg/dl vs. H_a: $\mu > 4.8$ mg/dl, we compute $z = \frac{\bar{x}-4.8}{0.9/\sqrt{6}} \doteq 1.45$ and $P \doteq 0.0735$. This is not strong enough to reject H_0.

Note: *The confidence interval in (a) would allow us to say without further computation that, against a two-sided alternative, we would have $P > 0.05$. Because we have a one-sided alternative, we could conclude from the confidence interval that $P > 0.025$, but that is not enough information to draw a conclusion.*

6.122. **(a)** The 90% confidence interval is $145 \pm (1.645)(8/\sqrt{15}) = 145 \pm 3.3979 = 141.6021$ to 148.3979 mg/g. **(b)** Our hypotheses are H_0: $\mu = 140$ mg/g vs. H_a: $\mu > 140$ mg/g. The test statistic is $z = \frac{145-140}{8/\sqrt{15}} = 2.42$, so the P-value is $P = P(Z > 2.42) = 0.0078$. This is strong evidence against H_0; we conclude that the mean cellulose content is higher than 140 mg/g. **(c)** We must assume that the 15 cuttings in our sample are an SRS. Because our sample is not too large, the population should be Normally distributed, or at least not extremely nonnormal.

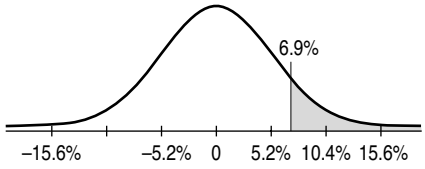
6.123. **(a)** The stemplot is reasonably symmetric for such a small sample.

2	034
2	
3	01124
3	6
4	3

(b) $\bar{x} = 30.4$ μg/l; $30.4 \pm (1.96)(7/\sqrt{10})$ gives 26.0614 to 34.7386 μg/l. **(c)** H_0: $\mu = 25$ μg/l; H_a: $\mu > 25$ μg/l. $z = \frac{30.4-25}{7/\sqrt{10}} \doteq 2.44$, so $P = 0.0073$. (We knew from (b) that it had to be smaller than 0.025). This is fairly strong evidence against H_0; the beginners' mean threshold is higher than 25 μg/l.

6.124. **(a)** The intended population is probably "the American public"; the population that was actually sampled was "citizens of Indianapolis (with listed phone numbers)." **(b)** Take $\bar{x} \pm 1.96s/\sqrt{201}$. Food stores: 15.22 to 22.12. Mass merchandisers: 27.77 to 36.99. Pharmacies: 43.68 to 53.52. **(c)** The confidence intervals do not overlap at all; in particular, the *lower* confidence limit of the rating for pharmacies is higher than the *upper* confidence limit for the other stores. This indicates that the pharmacies are *really* rated higher.

6.125. **(a)** Under H_0, $\bar{x}$ has a $N(0\%, 55\%/\sqrt{104}) \doteq N(0\%, 5.3932\%)$ distribution. **(b)** $z = \frac{6.9-0}{55/\sqrt{104}} \doteq 1.28$, so $P = P(Z > 1.28) = 0.1003$. **(c)** This is not significant at $\alpha = 0.05$. The study gives *some* evidence of increased compensation, but it is not very strong; similar results would happen about 10% of the time just by chance.



6.126. No: "Significant at $\alpha = 0.01$" *does* mean that the null hypothesis is unlikely, but only in the sense that the evidence (from the sample) would not occur very often if H_0 were true. There is no probability associated with H_0; it is either true or it is not.

Note: *Bayesian statistics views the parameter we wish to estimate as having a probability distribution; with that viewpoint, it would make sense to speak of "the probability that H_0 is true." This textbook does not take the Bayesian approach.*

6.127. Yes. That's the heart of why we care about statistical significance. Significance tests allow us to discriminate between random differences ("chance variation") that might occur when the null hypothesis is true, and differences that are unlikely to occur when H_0 is true.

6.128. If p is the probability that red occurs, we test H_0: $p = \frac{18}{38}$ vs. H_a: $p \neq \frac{18}{38}$.

6.129. For each sample, find $\bar{x}$, then take $\bar{x} \pm 1.96(4/\sqrt{12}) = \bar{x} \pm 2.2632$.

We "expect" to see that 95 of the 100 intervals will include 25 (the true value of μ); binomial computations show that (about 99% of the time) 90 or more of the 100 intervals will include 20.

6.130. For each sample, find $\bar{x}$, then compute $z = \frac{\bar{x}-25}{4/\sqrt{12}}$. Choose a significance level α and the appropriate cutoff point—for example, with $\alpha = 0.10$, reject H_0 if $|z| > 1.645$; with $\alpha = 0.05$, reject H_0 if $|z| > 1.96$.

If, for example, $\alpha = 0.05$, we "expect" to reject H_0 (that is, make the wrong decision) only 5 of the 100 times.

6.131. For each sample, find $\bar{x}$, then compute $z = \frac{\bar{x}-23}{4/\sqrt{12}}$. Choose a significance level α and the appropriate cutoff point (z^*)—for example, with $\alpha = 0.10$, reject H_0 if $|z| > 1.645$; with $\alpha = 0.05$, reject H_0 if $|z| > 1.96$.

Because the true mean is 25, $Z = \frac{\bar{x}-25}{4/\sqrt{12}}$ has a $N(0, 1)$ distribution, so the probability that we will accept H_0 is $P\left(-z^* < \frac{\bar{x}-23}{4/\sqrt{12}} < z^*\right) = P(-z^* < Z + 1.7321 < z^*) = P(-1.7321 - z^* < Z < -1.7321 + z^*)$. If $\alpha = 0.10$ ($z^* = 1.645$), this probability is $P(-3.38 < Z < -0.09) = 0.4637$; if $\alpha = 0.05$ ($z^* = 1.96$), this probability is $P(-3.69 < Z < 0.23) = 0.5909$. For smaller α, the probability will be larger. Thus we "expect" to (wrongly) accept H_0 about half the time (or more), and correctly reject H_0 about half the time or less. (The probability of rejecting H_0 is essentially the power of the test against the alternative $\mu = 25$.)

6.132. The listing below shows the statements ordered from largest to smallest means (i.e., with the highest level of agreement at the top) for both age groups. Lines are drawn in the list to show sets of similar complaints for both groups; for example, the top complaint is the same for both groups, and the next set is the same (although the order is different). Restaurateurs wishing to make their businesses attractive to older customers should probably pay the most attention to the top of this list. (Note that those statements that are in the bottom half of the list have ratings of just over 3 or less. On a 1-to-5 scale, "3" is a neutral response.)

Ages 50–64	Ages 65–79
I would rather be served ...	I would rather be served ...
I would rather pay the server ...	Tables are too close together.
Tables are too close together.	Print size is not large enough.
Print size is not large enough.	Background music is often too loud.
Most restaurants are too noisy.	I would rather pay the server ...
Background music is often too loud.	Most restaurants are too noisy.
Restaurants are too smoky.	Tables are too small.
Service is too slow.	Restaurants are too smoky.
Tables are too small.	Service is too slow.
Glare makes menus difficult to read.	Glare makes menus difficult to read.
Most restaurants are too dark.	It is difficult to hear the service staff.
It is difficult to hear the service staff.	Most restaurants are too dark.
Colors make menus difficult to read.	Colors make menus difficult to read.

In comparing the means, we use a two-sided alternative because we have no prior expectation that the differences should be either positive or negative. As the text says, the z statistics are found by taking the differences between the means and dividing by 0.08; for

example, the first is $z = \frac{2.75 - 2.93}{0.08} \doteq -2.25$, for which $P = 0.0244$. Those with P-values less than 0.05 are marked with an asterisk. Because we are performing 13 tests, a P-value below 0.05 should not be viewed as establishing significance. Using the Bonferroni approach (see Exercise 6.104), we would be willing to declare as significant any differences with $P < 0.05/13 \doteq 0.0038$; these are marked with two asterisks. Given the clustering of complaints noted in the list above, it is not too surprising that few differences are significant. Two of these three differences suggest that the older group may have more concerns about noise level and hearing.

	Statement	z	P	
Ambience:	Most restaurants are too dark.	−2.25	0.0244	*
	Most restaurants are too noisy.	−1.25	0.2113	
	Background music is often too loud.	−3.50	0.0005	**
	Restaurants are too smoky.	0.62	0.5353	
	Tables are too small.	−2.38	0.0173	*
	Tables are too close together.	−0.25	0.8026	
Menu design:	Print size is not large enough.	−1.13	0.2585	
	Glare makes menus difficult to read.	−2.50	0.0124	*
	Colors of menus make them difficult to read.	−2.38	0.0173	*
Service:	It is difficult to hear the service staff.	−4.38	0.0000	**
	I would rather be served than serve myself.	1.13	0.2585	
	I would rather pay the server than a cashier.	5.00	0.0000	**
	Service is too slow.	0.37	0.7114	