

Chapter 15 Solutions

15.1. The most important reason is (c); this is a convenience sample consisting of students with a particular interest in filmmaking which may make their opinions different from those of “typical” college-age adults. Anything we learn from this sample will not extend to the larger population

The other two reasons are valid (but less important) issues. Reason (a)—the size of the course, and large margin of error—would make the interval less informative, even if the sample were representative of the population. Reason (b)—nonresponse—is a potential problem with every survey, but there is no particular reason to believe it is more likely in this situation.

15.2. (a) The 95% confidence interval is $\bar{x} \pm 1.96\sigma/\sqrt{880} = 1.92 \pm 0.1209 = 1.80$ to 2.04 motorists. **(b)** The large sample size means that, because of the central limit theorem, the sampling distribution of \bar{x} is roughly Normal even if the distribution of responses is not. **(c)** Only people with listed phone numbers were represented in the sample, and the low response rate ($10.9\% \doteq \frac{5029}{45,956}$) means that even that group may not be well represented by this sample.

15.3. Any number of things could go wrong with this convenience sample. Depending on the time of day or the day of the week, certain types of shoppers would or would not be present.

15.4. (a) The three confidence intervals are given in the table on the right: $\bar{x} \pm z^*\sigma/\sqrt{n} \doteq 26.8 \pm 0.2933z^*$. **(b)** The margins of error, given in the “m.e.” column, increase as confidence level increases.

Conf. Level	z^*	m.e.	Interval
90%	1.645	0.4824	26.32 to 27.28
95%	1.960	0.5748	26.23 to 27.37
99%	2.576	0.7555	26.04 to 27.56

15.5. With $z^* = 1.96$ and $\sigma = 7.5$, the margin of error is $z^*\sigma/\sqrt{n} = 14.7/\sqrt{n}$. **(a)** and **(b)** The margins of error are given in the table. **(c)** Margin of error decreases as n increases. (Specifically, it is halved each time n quadruples.)

n	m.e.
100	1.47
400	0.735
1600	0.3675

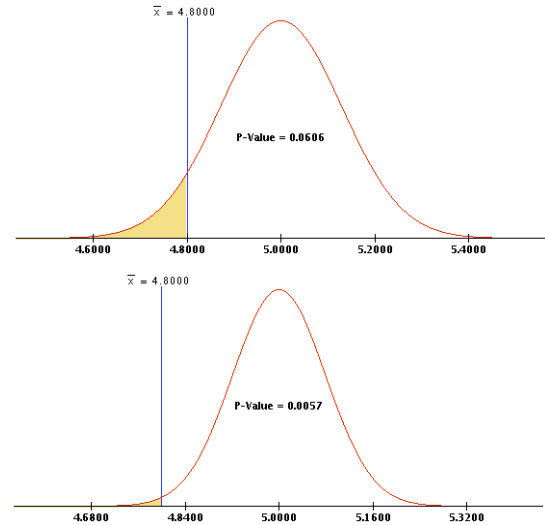
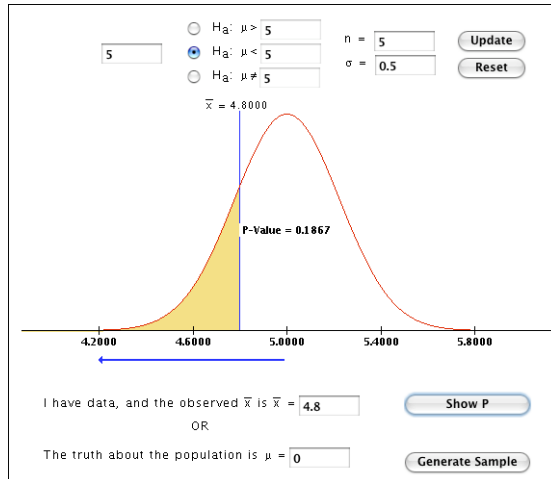
15.6. (a) Not included: this is a flaw in the sampling design. **(b)** Not included: this error arises from the sampling process. **(c)** Included: this random error is the only error addressed by confidence interval methods.

15.7. (a) $z = \frac{544-518}{114/\sqrt{50}} \doteq 1.61$ —not significant at the 5% level ($P = 0.0537$).

(b) $z = \frac{545-518}{114/\sqrt{50}} \doteq 1.67$ —significant at 5% level ($P = 0.0475$),

15.8. The full applet output for $n = 5$ is below on the left; on the right are the Normal curves drawn for $n = 15$ and $n = 40$. The reported P -values agree with the “hand-computed” values $z = \frac{4.8 - 5}{0.5/\sqrt{n}}$ and $P = P(Z < z)$ given in the table on the right.

n	z	P
5	-0.89	0.1867
15	-1.55	0.0606
40	-2.53	0.0057



15.9. The confidence intervals are given in the table on the right. In each case, the interval is $4.8 \pm 1.960 \left(\frac{0.5}{\sqrt{n}} \right)$.

n	Interval
5	4.362 to 5.238
15	4.547 to 5.053
40	4.645 to 4.955

15.10. (a) In a sample of size 500, we expect to see about 5 people who have a P -value of 0.01 or less [$5 = (500)(0.01)$]. These four *might* have ESP, or they may simply be among the “lucky” ones we expect to see. **(b)** The researcher should repeat the procedure on these four to see whether they again perform well.

15.11. $n = \left(\frac{(1.96)(7.5)}{1} \right)^2 \doteq 216.09$ —take $n = 217$.

15.12. $n = \left(\frac{(1.645)(60)}{10} \right)^2 \doteq 97.42$ —take $n = 98$.

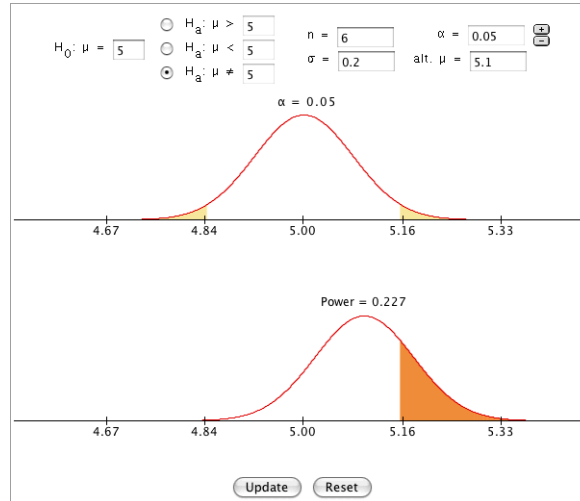
15.13. (a) If $\mu = 5.1$, the probability of (correctly) rejecting H_0 is 0.23. **(b)** This test allows too great a risk of mistakenly concluding that $\mu = 5$ when the product is unsuitable ($\mu = 5.1$). An unsuitable product would be sold about three-fourths of the time.

15.14. (a) To achieve higher power without changing α , we must increase the sample size. (Larger samples give more power against the same alternative.) **(b)** The power will increase. (Generally, power increases when α increases, and decreases when α decreases.) **(c)** The power against $\mu = 5.2$ would be greater: It is further from μ_0 (5), so the effect size is greater—it is easier to distinguish this alternative from the null hypothesis.

15.15. The powers reported by the applet are summarized in the following table.

	(a)	(b)	(c)		
n	Power	μ	Power	α	Power
6	0.227	5.1	0.227	0.05	0.227
12	0.405	5.2	0.684	0.10	0.336
24	0.684	5.3	0.955	0.25	0.528

One sample screenshot from the applet is shown on the right. Note that different software may give slightly different values. **(a)** As sample size increases (while α and the alternative remain the same), power increases. **(b)** Power is greater when the alternative considered is further away from μ_0 (while sample size and α remain the same). **(c)** Power increases when α increases (while sample size and the alternative remain the same).



15.16. For $\sigma = 0.2$, $\sigma = 0.1$, and $\sigma = 0.05$, the applet gives the power as 0.227, 0.684, and 0.998 (respectively). Power increases as variability decreases.

15.17. (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. Because these are “routine medical tests,” it seems reasonable to assume the subjects have no specific medical complaints, so that we choose H_0 to be “the patient is healthy.” However, for some conditions, the cost of mistakenly determining that someone is healthy might be too great, leading us to choose H_0 to be “the patient is ill.”

15.18. (a) Without a random sample, our results might be completely useless.

15.19. (c) This is a convenience sample consisting of the members of the basketball team, who are almost certainly all in fairly good physical condition.

15.20. (b) This voluntary response sample collects only the opinions of those who visit this site and feel strongly enough to respond.

15.21. (c) The announced margin of error accounts only for random sampling error.

15.22. (a) In describing an experiment, “uncontrolled” means that there was no control group.

15.23. (a) In order to detect small differences, we need large samples. (That is, small samples have low power against alternatives that are not very different from the null hypothesis.)

15.24. (c) We want $1.96\sigma/\sqrt{n} \leq 0.0005$, so $n \geq \left(\frac{(1.96)(0.001)}{0.0005}\right)^2 \doteq 15.36$ —use 16.

15.25. (a) The significance level is the probability of a Type I error—wrongly rejecting H_0 .

15.26. (b) The power is the probability of rejecting H_0 when a specific alternative ($\mu = 3$, in this case) is true.

15.27. (c) Power is the probability of detecting deviation from the null hypothesis.

15.28. We would like to know that the 148 respondents were chosen at random from all general managers of three-star and four-star hotels. We would also like to know the rate of nonresponse.

15.29. Were these random samples? How big were the samples? If the samples were very large, a significant result might not indicate an important difference.

15.30. $n = \left(\frac{(2.576)(0.78)}{0.2}\right)^2 \doteq 100.93$ —take $n = 101$.

15.31. (a) These results come from a convenience sample, rather than a random sample. Women sampled from a shopping mall do not represent the larger population of all women. (b) While this sample might not represent *all* women, it may be reasonable to view the women in the sample as an SRS of women who shop at large suburban malls.

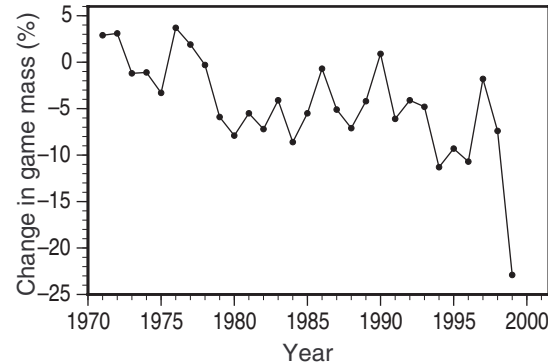
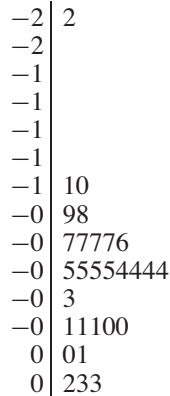
15.32. $n = \left(\frac{(1.96)(3000)}{1000}\right)^2 \doteq 34.57$ —take $n = 35$.

15.33. Many people might be reluctant to discuss such intimate details of their personal lives (although this may be somewhat offset by some who might be inclined to brag or exaggerate). The margin of error allows only for random sampling error, not bias.

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

15.35. The effect is greater if the sample is small. Quoting from the text (page 396), “If the sample is large, \bar{x} will be close to Normal even if individual measurements are strongly skewed [However,] outliers can distort the results of inference.” Outliers have this effect because \bar{x} is not resistant, but when the sample is large, the impact on \bar{x} is less.

- 15.36. (a)** A stemplot is shown; a histogram could also be used. The distribution has a low outlier, which makes confidence interval methods unreliable. **(b)** The timeplot shows a decreasing trend over time, so we should not treat these 29 observations as coming from a single population. (For example, the mean for an observation made in 1971 would not be the same as an observation made in 1999.)



- 15.37.** Opinion—even expert opinion—unsupported by data is the weakest type of evidence, so the third description is level C. The second description refers to experiments (clinical trials) and large samples; that is the strongest evidence (level A). The first description is level B: stronger than opinion, but not as strong as experiments with large numbers of subjects.
- 15.38.** A significance test answers only question (b). The P -value states how likely the observed effect (or a stronger one) is if chance alone is operating. 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.
- 15.39. (a)** Margin of error decreases. **(b)** The P -value decreases (the evidence against H_0 becomes stronger). **(c)** The power increases (the test becomes better at distinguishing between the null and alternative hypotheses).
- 15.40.** 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.
- 15.41. (a)** With $\bar{x} = 9.524$, we have $z = \frac{9.524-8}{2/\sqrt{5}} \doteq 1.7$, for which $P = 0.0891$. This is not significant at the 5% level. **(b)** The sample size is small, so the test has low power. As a result, it will often mistakenly fail to reject H_0 .
- 15.42. (a)** The difference observed in the study would occur in less than 1% of all samples if the two populations actually have the same proportion remaining on welfare. **(b)** The interval is constructed using a method that is correct (i.e., contains the actual proportion) 95% of the time. **(c)** No: treatments were not randomly assigned, but instead were chosen by the mothers. Mothers who choose to attend a job-training program may be more inclined to get themselves off of welfare.

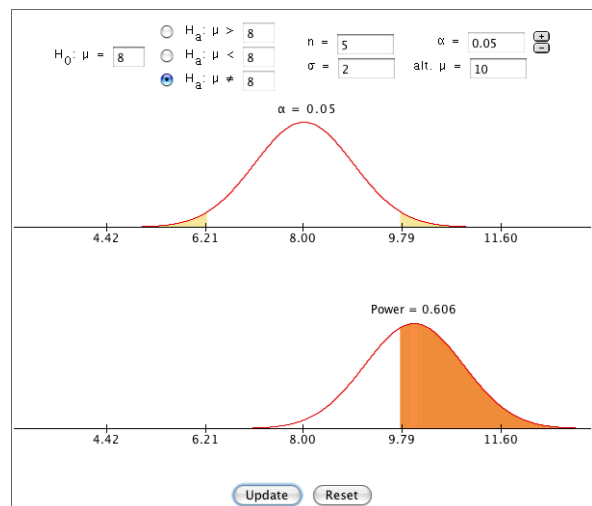
15.43. (a) “Statistically insignificant” means that the differences observed were no more than might have been expected to occur by chance even if SES had no effect on LSAT results. **(b)** If the results are based on a small sample, then even if the null hypothesis were not true, the test might not be sensitive enough to detect the effect. Knowing the effects were small tells us that the test was not insignificant merely because of a small sample size.

15.44. A low-power test may do a good job of *not* incorrectly rejecting the null hypothesis (that is, avoiding a Type I error), but it will often accept H_0 even when it is false, simply because it is difficult to distinguish between H_0 and “nearby” alternatives.

15.45. (a) This test has a 20% chance of rejecting H_0 when the alternative is true. **(b)** If the test has 20% power, then when the alternative is true, it will fail to reject H_0 80% of the time. **(c)** The sample sizes are very small, which typically leads to low-power tests.

15.46. (a) The significance level is the probability of a type I error, which was stated to be 0.04. **(b)** If surgery produced appreciably better results, this test would have detected that difference—that is, it would have given significant results—about 90% of the time.

15.47. The applet reports that the power of the test is 0.606. (Other software might report a slightly different value.)



15.48. (a) Because the alternative is $\mu > 0$, we reject H_0 at the 5% level when $z > 1.645$. **(b)** We reject H_0 when $3.162\bar{x} > 1.645$, or $\bar{x} > 0.5202$. **(c)** When $\mu = 0.8$, the power is

$$P(\text{reject } H_0) = P(\bar{x} > 0.5202) = P\left(\frac{\bar{x} - 0.8}{1/\sqrt{10}} < \frac{0.5202 - 0.8}{1/\sqrt{10}}\right) = P(Z > -0.88)$$

for which Table A gives 0.8106. (The answer given by the applet in Figure 15.2 is slightly different because of rounding.)

15.49. (a) The test statistic is $z = \frac{\bar{x} - 5}{0.2/\sqrt{6}}$. We reject H_0 at the 5% level if $z < -1.96$ or $z > 1.96$. (b) We reject H_0 if $\bar{x} > 5.16$ or $\bar{x} < 4.84$. These two numbers are $5 \pm (1.96)(0.2/\sqrt{6})$. In other words, we reject H_0 if \bar{x} is *not* between 4.84 and 5.16. (c) The power against $\mu = 5.1$ is

$$\begin{aligned} 1 - P(4.84 < \bar{x} < 5.16) &= 1 - P\left(\frac{4.84 - 5.1}{0.2/\sqrt{6}} < Z < \frac{5.16 - 5.1}{0.2/\sqrt{6}}\right) \\ &\doteq 1 - P(-3.18 < Z < 0.73) \\ &\doteq 1 - 0.7666 = 0.2334. \end{aligned}$$

(The answer given by the applet in the solution to Exercise 15.15 is slightly different because of rounding.)

15.50. $P(\text{Type I error}) = \alpha = 0.05$. $P(\text{Type II error}) = 1 - \text{Power} = 1 - 0.78 = 0.22$.

15.51. The power is $1 - P(\text{Type II error}) = 1 - 0.14 = 0.86$.

15.52. (a) $P(\text{Type I error}) = P(\bar{x} > 0) = P(Z > 0) = 0.50$. (b) $P(\text{Type II error when } \mu = 0.3) = P(\bar{x} \leq 0) = P\left(Z \leq \frac{0 - 0.3}{1/\sqrt{9}}\right) = P(Z \leq -0.9) = 0.1841$. (c) $P(\text{Type II error when } \mu = 1) = P\left(Z \leq \frac{0 - 1}{1/\sqrt{9}}\right) = P(Z \leq -3) = 0.0013$.

15.53. (a) In the long run, this probability should be 0.05. Out of 100 simulated tests, the number of false rejections will have a binomial distribution with $n = 100$ and $p = 0.05$. Most students will see between 0 and 10 rejections. (b) If the power is 0.808, the probability of a Type II error is 0.192, so in the long run, this probability should be 0.192. Out of 100 simulated tests, the number of false non-rejections will have a binomial distribution with $n = 100$ and $p = 0.192$. Most students will see between 10 and 29 non-rejections.