Chapter 7 Sampling Distributions of Estimates

Exercises for Section 7.2.1

1. **(a)** To halve $\text{sd}(\overline{X})$ we need $2^2 = 4$ times as many observations which is 40 in total or 30 in addition to the 10 we already have.

(b) To reduce $\text{sd}(\overline{X})$ to one third of its original size we need $3^2 = 9$ times as many observations which is 90 in total or 80 in addition to the 10 we already have.

(c) To reduce $\text{sd}(\overline{X})$ to one 9th of its original size we need $9^2 = 81$ times as many observations which is 810 in total or 800 in addition the 10 we already have.

2. Let X be the monthly profit. Then $X \sim \text{Normal}(\mu = 10, \sigma = 3.5)$ so that $\overline{X} \sim \text{Normal}(\mu_{\overline{X}} = 10, \sigma_{\overline{X}} = \frac{3.5}{\sqrt{6}} = 1.4289)$. Using these values for $\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$, $\text{pr}(\overline{X} > 8.5) = 0.8531$ (computer).

Exercises for Section 7.2.2

1. We can assume approximate Normality as $n = 50$. Hence $\overline{X}$ is approximately Normally distributed with $\mu_{\overline{X}} = 100$ and $\sigma_{\overline{X}} = \frac{15}{\sqrt{50}} = 2.1213$ giving $\text{pr}(\overline{X} < 97) \approx 0.07865$. (8%)

2. Let $\overline{X}$ be the average service time. We can assume approximate Normality as $n = 50$.

(a) As $\overline{X}$ is approximately Normally distributed with $\mu_{\overline{X}} = 3.1$ and $\sigma_{\overline{X}} = \frac{1.2}{\sqrt{50}} = 0.16971$), $\text{pr}(\overline{X} < 3.3) \approx 0.8807$.

(b) Total $= T = 50\overline{X}$ is approximately Normally distributed with mean $\mu_T = 50 \times 3.1 = 155$ and standard deviation $\sigma_T = \sqrt{50} \times 1.2 = 8.4853$. Thus, $\text{pr}(T < 150) \approx 0.2778$.

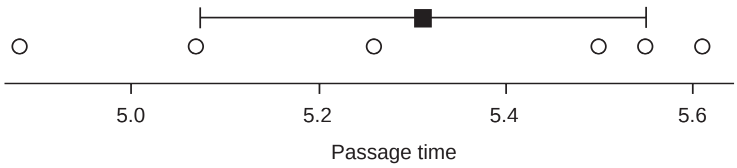
3. **(a)** Although n is only 28 we shall assume we can use the Normal approximation. Then, $\overline{X} \sim \text{Normal}(\mu_{\overline{X}} = 620.6, \sigma_{\overline{X}} = \frac{241.5}{\sqrt{28}} = 45.6392)$ so that $\text{pr}(\overline{X} \geq 795.1) \approx 0.0001$.
No, we do not believe they are the same. Testosterone levels seem to be higher in smokers. We used the mean and standard deviation for nonsmokers in our calculations. The probability is almost zero that a sample of nonsmokers of this size would have a mean testosterone level as high as we observed in our sample of smokers.

(b) Assuming that the serum level X for nonsmokers has, approximately, a Normal $(\mu = 620.6, \sigma = 241.5)$ distribution, the proportion above 795.1 is given by $\text{pr}(X \geq 795.1) = 0.2350$.

(c) No, as these answers deal with different quantities. In (a) we are dealing with a sample mean (average) of measurements taken from 28 men. If we repeatedly took such samples, the sample mean would almost never be 795.1 or bigger (approximately 9999 samples in every 10,000 taken would give a value of $\overline{x}$ *smaller* than 795.1). In (b) we are dealing with the behavior of single individuals. Single individuals quite often give a value of 795.1 or bigger (in fact 23.5% of individuals do). The reason why this is happening is that individual measurements are more variable than averages are.

Exercises for Section 7.2.3

1. (a)



(b) $\text{se}(\overline{x}) = \frac{s_X}{\sqrt{n}} = 0.1195$. Now $\overline{x} \pm 2\text{se}(\overline{x}) = 5.3117 \ \pm \ 2 \times 0.1195$, or $[5.07, 5.55]$. This has been added to plot above.

(c) Since $\frac{5.8 - 5.3117}{0.1195} = 4.09$, the value 5.8 is 4.09 standard errors above the sample mean from the data.

2. (a) $\text{se}(\overline{x}) = 57.696$. The two-standard-error interval is $\overline{x} \pm 2\text{se}(\overline{x}) = 795.1 \pm 2 \times 57.696$, or $[679.7, 910.5]$. It is a fairly safe bet that the true value is somewhere between 680 and 911 ng/dL.

(b) No, the value 620.6 ng/dL is not plausible as, since $\frac{620-6-795.1}{57.696} = -3.14$, this value is more than 3 standard errors below the estimate we obtained from our data. Alternatively, we could reach this conclusion by noting that 620.6 lies outside our 2-standard-error interval for the true mean.

3. Using a 2 standard-error interval, $\overline{x} \pm 2\text{se}(\overline{x}) = \overline{x} \pm 2\frac{s_X}{\sqrt{n}} = 250 \pm 2 \times \frac{50}{\sqrt{40}}$, i.e., $[234.2, 265.8]$. It is a fairly safe bet that the true average daily rate lies somewhere between \$234 and \$266.

Exercises for Section 7.3.1

1. (a) Since $\widehat{p} \pm 2\text{se}(\widehat{p}) = 0.39 \pm 2\sqrt{\frac{0.39 \times 0.61}{90}}$, or $[0.29, 0.49]$, it is a fairly safe bet that the true percentage of music students with fathers in the highest socioeconomic group lies somewhere between 29% and 49%.

(b) No, the value 0.23 or 23% is not plausible as, since $\frac{0.23 - 0.39}{0.05141} = -3.1$, the value 0.23 is 3.1 standard errors below our data estimate.
Alternatively, we could reach the same conclusion by noting that 0.23 lies outside our two-standard-error interval for the true value.

2. *Professionals*: $0.32 \pm 2\sqrt{\frac{0.32 \times 0.68}{2280}}$, i.e., $[0.30, 0.34]$. It is a fairly safe bet that the true percentage of Yahoo users who are professionals lies somewhere between 30% and 34%.
University: $0.4 \pm 2\sqrt{\frac{0.4 \times 0.6}{2280}}$, i.e., $[0.38, 0.42]$. It is a fairly safe bet that the true percentage of Yahoo users who have been to university lies somewhere between 38% and 42%. [All of this assumes that we are making inferences from a random sample of Yahoo users.]

3. We could work from a two-standard-error interval, namely, $\widehat{p} \pm 2\text{se}(\widehat{p}) = 0.48 \pm 2\sqrt{\frac{0.48 \times 0.52}{825}}$, or $[0.445, 0.515]$, and then argue that there are values of p above 50%

(namely, 0.5 to 0.515) that are plausible values because they lie in the interval. Alternatively, we could argue that as 0.5 (50%) is only $(0.5-0.48)/\sqrt{0.48 \times 0.52/825} = 1.15$ standard errors above our data estimate of 0.48, some values of p greater than 0.5 are plausible.

4. **(a)** The two-standard-error interval is $\widehat{p} \pm 2\text{se}(\widehat{p}) = 0.176 \pm 2\sqrt{\frac{0.176 \times 0.824}{8000}}$, or $[0.167, 0.185]$. It is a fairly safe bet that the true percentage of American women who had been raped lies somewhere between 16.7% and 18.5%.

(b) The two-standard error interval is $0.216 \pm 2\sqrt{\frac{0.216 \times 0.784}{1323}}$, or $[0.19, 0.24]$, indicating that somewhere between 19% and 24% of all American women who had been raped were under 12 when first raped. [Note that the interval in (b) is considerably wider than that in (a) indicating greater uncertainty about the true percentage.]

(c) The two-standard error interval is $0.558 \pm 2\sqrt{\frac{0.558 \times 0.442}{1323}}$, or $[0.531, 0.585]$, indicating that somewhere between 53% and 59% of all American women who had been raped were under 18 when first raped.

Exercises for Section 7.5

1. **(a)** $\text{se}(\widehat{p}_{11} - \widehat{p}_{13}) = \sqrt{\text{se}(\widehat{p}_{11})^2 + \text{se}(\widehat{p}_{13})^2} = \sqrt{0.031^2 + 0.030^2} = 0.04314$.

(b) $\frac{0.474-0.303}{0.04314} = 3.96$. The two sample proportions are nearly 4 standard errors apart clearly signalling that the corresponding true proportions smoking at least once in grades 11 and 13 are different. There is a real drop off.

(c) The two-standard-error interval is $\widehat{p}_{11} - \widehat{p}_{13} \pm 2\text{se}(\widehat{p}_{11} - \widehat{p}_{13}) = 0.474 - 0.303 \pm 2 \times 0.04314$, or $[0.09, 0.26]$. It is a fairly safe bet that the true percentage of grade 11 students using cigarettes at least once was larger than the corresponding percentage for grade 13 students by somewhere between 9 and 26 percentage points.

(d) For discussion. Some possibilities include students tending to experiment in grade 11 and then not continuing to smoke, the possibility that the group of students who stop attending school between grades 11 and 13 includes a higher proportion of the smokers, or that there are different social dynamics in different years of students.

2. We will denote observations 1–6 as being from group 1 and observations 7–29 as being from group 2. Then, $\overline{x}_1 = 5.3117$, $s_1 = 0.2928$: $\overline{x}_2 = 5.4835$, $s_2 = 0.1904$.

(a) $\text{se}(\overline{x}_2 - \overline{x}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.2928^2}{6} + \frac{0.1904^2}{23}} = 0.1260$.
As the two sample means are only $\frac{5.4835-5.3117}{0.126} = 1.36$ standard errors apart, we have no evidence that the corresponding true means differ. Thus we have no evidence that changing the wire changed the quantity the experiment was measuring.

(b) The two-standard-error interval is $\overline{x}_2 - \overline{x}_1 \pm 2\text{se}(\overline{x}_2 - \overline{x}_1) = 5.4835 - 5.3117 \pm 2 \times 0.126$, or $[-0.08, 0.42]$. The quantity being measured after changing the wire could be anywhere between 0.08 units smaller and 0.42 units bigger than the quantity being measured before the wire was changed. This includes the possibility of "no change."

3. $\widehat{p}_{before} - \widehat{p}_{after} = 0.31 - 0.20 = 0.11$. Then,

$$\begin{aligned} \text{se}(\widehat{p}_{before} - \widehat{p}_{after}) &= \sqrt{\frac{\widehat{p}_{before}(1-\widehat{p}_{before})}{n_{before}} + \frac{\widehat{p}_{after}(1-\widehat{p}_{after})}{n_{after}}} \\ &= \sqrt{\frac{0.31 \times 0.69}{186} + \frac{0.2 \times 0.8}{97}} = 0.05291. \end{aligned}$$

(a) As $\frac{0.31-0.20}{0.05291} = 2.08$, the sample proportions are more than 2 standard errors apart. It is a fairly safe bet that there was a real change.

(b) The two-standard-error interval is $0.11 \pm 2 \times 0.0529$, or $[0.004, 0.216]$. It is a fairly safe bet that the percentage having a one-time encounter in the past 3 months decreased by somewhere between about 0.4 and 22 percentage points.

(c) The population of people who use the clinic "frozen" at the two different points in time. We were assuming that the sets of people questioned were random samples from these two populations.

(d) We would tend to believe that it was the anouncement if there were no other changes we could think of which might have affected the types of people visiting the clinic or their behavior. We would want to investigate what else had changed over this period.

4. (a) In each case, $\overline{x} \pm 2\text{se}(\overline{x}) = \overline{x} \pm 2\frac{s_X}{\sqrt{n}}$ gives a two-standard-error interval for the true mean for the relevant group.
None: $620.6 \pm 2 \times \frac{241.5}{\sqrt{62}}$, or $[559, 682]$.
$1-30$: $715.6 \pm 2 \times \frac{248.0}{\sqrt{31}}$, or $[627, 805]$.
$31-70$: $795.1 \pm 2 \times \frac{305.3}{\sqrt{28}}$, or $[680, 910]$.

(b) (i) The two-standard-error interval is $\overline{x}_{31-70} - \overline{x}_{none} \pm 2\text{se}(\overline{x}_{31-70} - \overline{x}_{none}) = 795.1 - 620.6 \pm 2\sqrt{\frac{305.3^2}{28} + \frac{241.5^2}{62}}$, or $[44, 305]$, which places the true mean testosterone level for the 31–70 group as being somewhere between 44 and 305 ng/dL higher than the true mean for nonsmokers. It is not plausible that there is no difference.

(ii) The two-standard-error interval is $\overline{x}_{1-30} - \overline{x}_{none} \pm 2\text{se}(\overline{x}_{1-30} - \overline{x}_{none}) = 715.6 - 620.6 \pm 2\sqrt{\frac{248.0^2}{31} + \frac{241.5^2}{62}} = 95 \pm 108$, or $[-13, 203]$. The true mean testosterone level for the 1–30 group is somewhere between being 13 ng/dL below and 305 ng/dL higher than the true mean for nonsmokers. This includes the possibility that there is no difference at all. It is plausible that there is no difference.

(c) We can't use individual two-standard-error intervals for making comparisons when there is overlap as the combined variation is not taken into account properly – see Section 7.5.3.

(d) We have an observational study and not a controlled experiment. There could be some other variable related to both smoking and testosterone that is causing the relationship we see. The causal influence could even be in the other direction. Perhaps high-testosterone men are more likely to take up smoking.

(e) No, we cannot reach any such conclusion. The two-standard-error intervals refer only to the true *means* and say nothing about any other aspect of the distributions. We see from Fig. 7.5.2 that there is actually a great deal of overlap in the testosterone levels of the groups.

Review Exercises 7

1. The first 2 parts of this question concern ideas in Section 6.4.4 (see the paragraph entitled *Independent individuals versus Clones*).

(a) Here $Y = Sum = \sum_{i=1}^{7} X_i$ so that $\mu_Y = n\,\mu_X = 7 \times 87 = 609$ and $\sigma_Y = \sqrt{n}\,\sigma_X = \sqrt{7} \times 23 = 60.85$.

(b) Here $W = 7X$ so that $\mu_W = 7 \times \mu_X = 7 \times 87 = 609$ and $\sigma_W = 7 \times \sigma_X = 7 \times 23 = 161$.

(c) For a sample mean from a random sample of size $n = 36$, $\mu_{\overline{X}} = \mu_X = 87$ and $\sigma_{\overline{X}} = \frac{\sigma_X}{n} = \frac{23}{6} = 3.83$.

(d) Normal; the central limit theorem.

(e) It would have the same mean and smaller spread. More technically, as $144 = 4 \times 36$, the standard deviation of $\overline{X}$ for a sample of size 144 is one half as large as it is for a sample of size 36. If the original distribution was extremely non-Normal the distribution might also be more Normal looking.

3. **(a)** $X \sim \text{Binomial}(n = 870, p = 0.79)$.

(b) Suppose the true value of p really was $p = 0.79$. Then $\text{sd}(\widehat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.79 \times 0.21}{870}} = 0.01381$. Now $\frac{0.3897 - 0.79}{0.01381} \approx -29$. The observed data value is 29 standard deviations below 0.79. This would virtually never happen if the selection was random. We do not believe that the jury drawing was at random.

(c) Suppose the true value of p really was $p = 0.06$. Then $\text{sd}(\widehat{P}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.06 \times 0.94}{405}} = 0.01180$. Since $\frac{0.037 - 0.06}{0.0118} = -1.95$ the observed data value is almost 2 standard deviations below 0.06. Values this far away would seldom occur if hiring was at random.

(d) The first recommendation questions the randomness of the jury selection process whereas the second questions whether a particular factor, race in this case, was involved in the teacher-hiring process. The first does not attempt to apportion blame to any factor in the event that the selection process was not random; the second does tend to attribute blame on racial discrimination as a primary factor in the event that the selection process was not random. The first limits "suspect" to applying to a specific group, namely social scientists while the second makes no such limitation.

(e) A prima facie case is a case considered strong enough that it must be answered in court. While the occurrence of gross statistical disparities is an indication that the selection of teachers is not random, the factors that are used as criteria in the hiring process may not be necessarily based on any desire to discriminate.
Reasons for the decision: Past experience in other areas may have pointed to the practice of discrimination, and the strong opinion is an expression of the Court's disapproval in Law. Discrimination is hard to prove so that the opinion expressed tends to shift the burden onto proving nondiscrimination.
Reasons against: In the absence of further evidence, it is dangerous to blame just one factor in the actual hiring process.
Suppose, for example, that hiring was done blindly on the basis of qualifications. It might be that several factors including past discriminatory practices at other levels may have led to black teachers having lower qualifications on average. Basing hiring decisions on qualifications would then lead to a lower proportion of black teachers being hired. Would this make using qualifications discriminatory? Should the problem of black teachers being under-represented be attacked using the hiring process or at other levels? Political processes have long struggled with issues like this without reaching solutions that everybody can live with. So we leave it for you to decide!

5. **(a)** Let $\widehat{P}$ be the proportion of voters who voted for her. Then $\widehat{P}$ is approximately Normally distributed with $\mu_{\hat{P}} = p = 0.6$ and $\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6\times 0.4}{n}}$.

(b) When $n = 200$, $\sigma_{\hat{P}} = \sqrt{\frac{0.6\times 0.4}{200}} = 0.03464$ so that $\text{pr}(\widehat{P} > 0.5) = 0.998$.

7. **(a)** $\mu_X = \text{E}(X) = \sum x\,\text{pr}(x) = 1 \times \frac{20}{38} + (-1) \times \frac{18}{38} = \frac{2}{38} = 0.05263$.
$\sum (x - \mu_X)^2 pr(x) = (1 - 0.05263)^2 \times \frac{20}{38} + (-1 - 0.05263)^2 \times \frac{18}{38} = 0.99723$.
$\sigma_X = \text{sd}(X) = \sqrt{0.99723} = 0.9986 \approx 1$.

(b) **(i)** We are working with a sum from a random sample of size $n = 50$ from this distribution so $\text{E}(Sum) = n\,\mu_X = 50 \times 0.05263 = 2.6315$ and $\text{sd}(Sum) = \sqrt{n}\,\sigma_X = \sqrt{50} \times 0.9986 = 7.0612$.
(ii) $\text{E}(\overline{X}) = \mu_X = 0.05263$ and $\text{sd}(\overline{X}) = \frac{\sigma_X}{\sqrt{n}} = \frac{0.9986}{\sqrt{50}} = 0.14122$.

(c) Basically already done in (b). **(i)** $\mu_{Sum} = n \times 0.05263$ and $\sigma_{Sum} = \sqrt{n} \times 0.9986$.
(ii) $\mu_{\overline{X}} = 0.05263$ and $\sigma_{\overline{X}} = \frac{0.9986}{\sqrt{n}}$.

(d) The casino makes money if the average winnings from the 50 bets exceeds \$0. The gambler makes money if the average is less than \$0. When $n = 50$ we have $\overline{X} \sim \text{Normal}(\mu_{\overline{X}} = 0.05263, \sigma_{\overline{X}} = 0.14122)$ so that $\text{pr}(\overline{X} < 0) = 0.3547$. Thus, after 50 bets, approximately 35% of gamblers have made money.

(e) $\overline{X} \sim$ Normal $(\mu_{\overline{X}} = 0.05263, \sigma_{\overline{X}} = \frac{0.9986}{\sqrt{n}})$. When $n = 1000$, $\sigma_{\overline{X}} = 0.031579$ and $\text{pr}(\overline{X} < 0) = 0.0478$. After 1000 bets, approximately 1 person in 20 has made money. When $n = 100,000$, $\sigma_{\overline{X}} = 0.003158$ and $\text{pr}(\overline{X} < 0) \approx 10^{-62}$. Essentially no one has made money (except the casino of course). The probability is even smaller after a million bets.

(f) We have drastically rounded the data in the following table as we only want to give a broad picture. We have used 3 standard deviations about the mean each time so we are getting the range of values within which the result will fall 99.7% of the time. These intervals relate to the casino's winnings.

n		Average	Total
	$\mu_X \pm 3\frac{\sigma_X}{\sqrt{n}}$		$n\mu \pm 3\sqrt{n}\sigma_X)$
50	0.0526 ± 0.4237	[-0.37,0.48]	[-18.5,24]
1000	0.0526 ± 0.0947	[-0.04,0.15]	[-40,150]
100,000	0.0526 ± 0.0095	[0.043,0.062]	[4300,6200]
1,000,000	0.0526 ± 0.0030	[0.050,0.056]	[50,000, 56,000]

9. We use $\text{sd}(\widehat{P}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4\times0.6}{n}}$. We have to think about what n should be.

(a) Units are couples and we have $n = 100$ of them so $\text{sd}(\widehat{P}) = \frac{0.4\times0.6}{\sqrt{100}} = 0.024$.

(b) Units are individuals and we have $n = 200$ of them so $\text{sd}(\widehat{P}) = \frac{0.4\times0.6}{\sqrt{200}} = 0.0170$.

(c) The reality will be between the two extremes of (a) and (b). Individuals within a married couple do not necessarily think alike but they are more likely to think alike than are any two random individuals. Thus the variability in $\widehat{P}$ values will be larger than predicted by (b) and smaller than predicted by (a). We would tend to favor (a) which overstates the uncertainty.

11. (a) $\text{se}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{(0.66)(0.34)}{943}} = 0.0154$.

(b) $\text{se}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{0.5\times0.5}{172}} = 0.0381$.

(c) $\frac{0.0381}{0.0154} = 2.47$ times bigger.

(d) $\text{se}(\widehat{p}_1 - \widehat{p}_2) = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}} = \sqrt{\frac{0.67\times0.33}{221} + \frac{0.5\times0.5}{172}} = 0.0495$.

(e) $\sqrt{\frac{0.67\times0.33}{221}} = 0.0316$. The ratios are $\frac{0.0495}{0.0316} = 1.57$ times bigger and $\frac{0.0495}{0.0381} = 1.30$ times bigger.

(f) It will be too small for looking at proportions calculated from subsets of the data and for looking at differences between proportions. The poll will appear to be more accurate than it really is in these situations. See Section 8.5.3 for a detailed discussion of these issues.

13. (a) The proportion of the sample of downtown cap wearers who wear their caps backwards is $\widehat{p}_{dntn} = \frac{174}{407} = 0.4275$. The corresponding two-standard-error interval is $0.4275 \pm 2\sqrt{\frac{0.4275\times0.5725}{407}}$ giving $[0.378, 0.477]$. This suggests that the true proportion of downtown cap wearers who wore their caps backwards at this time was somewhere between about 38% and 48%.

(b) For business school cap wearers, the proportion of the sample who wore their caps backward is $\widehat{p}_{bus} = \frac{107}{319} = 0.3354$. The standard error of the difference in sample proportions, $\text{se}(\widehat{p}_{dntn} - \widehat{p}_{bus}) = \sqrt{\frac{0.4275\times0.5725}{407} + \frac{0.3354\times0.6646}{319}} = 0.03606$. As $\frac{(0.4275-0.3354)}{0.03606} = 2.55$, the two sample proportions are more than 2 $\text{se}(\widehat{p}_{dntn} - \widehat{p}_{bus})$ apart suggesting that the true proportions are in fact different.

(c) No. They do not conflict as $\text{se}(\widehat{p}_{dntn} - \widehat{p}_{bus}) < \text{se}(\widehat{p}_{dntn}) + \text{se}(\widehat{p}_{bus})$. See Section 7.5.3 for a discussion.

(d) Yes. As the individual intervals do not overlap we know that the two-standard-error interval for the difference will not contain zero in each case (see Section 7.5.3).

(e) If we just observe individuals moving past some location, we will often get people in groups who are likely to do similar things like wearing their caps the same way, thus violating the independence assumption. We might also be getting a biased sample. The types of people walking past the locations we choose to observe may tend to behave differently from those found at other locations. Locations would have to be sampled carefully to avoid this.

15. (a) $\text{se}(\overline{x}_{bottle} - \overline{x}_{breast}) = \sqrt{\frac{15.18^2}{90} + \frac{17.39^2}{210}} = 2.00.$
Since $\frac{(103.0-92.8)}{2.00} = 5.1$, the sample mean for breast-fed babies $\overline{x}_{breast}$ is more than 5 standard errors bigger than the sample mean for bottle-fed babies, so we conclude that the true mean IQ for breast-fed babies is larger than the true mean IQ for bottle-fed babies.

(b) The two-standard-error interval for the true difference $\mu_{breast} - \mu_{bottle}$ is given by $(103.0 - 92.8) \pm 2 \times 2.00$, or $[6.2, 14.2]$. This suggests that the true mean IQ for breast-fed babies is larger than that for bottle-fed babies by somewhere between 6 and 14 units.

(c) Preterm, low-birth-weight babies in the catchment areas of these special-care units fitting the profiles that trigger referral to these units.

(d) No, because this is observational data. Mothers chose to breast feed or not to breast feed. There was no random assignment. It may be, for example, that higher IQ mothers are more likely to choose to breast feed than lower IQ mothers and it is this (or one of a host of other possible differences) that leads to the observed IQ differences in the babies.

(e) Yes, as it partially answers the objection we raised in (d). We might expect mothers who wanted to breast-feed but could not (for physical reasons) to have similar IQs to those that wanted to and could. We note that the IQs of the babies of the wanted-to-but-couldn't group were similar to the IQs of the bottle-fed babies, and not to the breast-fed babies. This strengthens the impression that it is the breast-feeding itself that is causing the difference we are seeing.

17. (a) The three answers are: (i) Those in the "humane" group should score higher than those in the "inhumane" group.
(ii) In the humane group, there should be a trend downwards from "typical" to "atypical" (we would be more affected by what we saw if we thought it was typical). In the inhumane group, the reverse should hold. (iii) We would expect the "control" group reactions would fall in between those of the humane and inhumane "no-information group".

(b) We will react to sample means more than two standard errors apart as providing evidence that differences between the corresponding true means exist. When sample means are less than 2 standard errors apart it is plausible that no true differences exist. In such situations we will say that we have not demonstrated a

true difference.
We have not demonstrated a true difference between any of the three subgroups shown the inhumane portrayal.
We have not demonstrated a true difference between any of the subgroups shown the humane portrayal.
We can conclude that the control group scored lower on average than the "typical" and "no-information" subgroups in the humane group and scored higher than the typical subgroup of the inhumane group.
Under "no-information" conditions, those seeing the humane portrayal have been demonstrated to score higher on average than those shown the inhumane portrayal.
Under typical conditions, those seeing the humane portrayal have been demonstrated to score higher on average than those shown the inhumane portrayal.
Even when told that the behavior is atypical, those seeing the humane portrayal have been demonstrated to score higher on average than those shown the inhumane portrayal.
The data supports contention (i) in (a), fails to demonstrate (ii), and partly supports (iii).

(c) Use larger samples with groups for each sex.

(d) Are there sex or age differences in the participants? What effect does the sex of the guard have? etc.

***19. (a)** Our estimate of the number visiting the doctor in Canada is $\hat{t}_{can} = 0.1 \times 27 \times 10^6$ (proportion times population size). Similarly, $\hat{t}_{usa} = 0.2 \times 250 \times 10^6$ and $\hat{t}_{mex} = 0.6 \times 90 \times 10^6$. Thus the total number visiting the doctor is $0.1 \times 27 \times 10^6 + 0.2 \times 250 \times 10^6 + 0.6 \times 90 \times 10^6$.
The total population size is $27 \times 10^6 + 250 \times 10^6 + 90 \times 10^6 = 367 \times 10^6$. The proportion visiting the doctor is thus

$$\widehat{p} = \frac{0.1 \times 27 \times 10^6 + 0.2 \times 250 \times 10^6 + 0.6 \times 90 \times 10^6}{367 \times 10^6}.$$

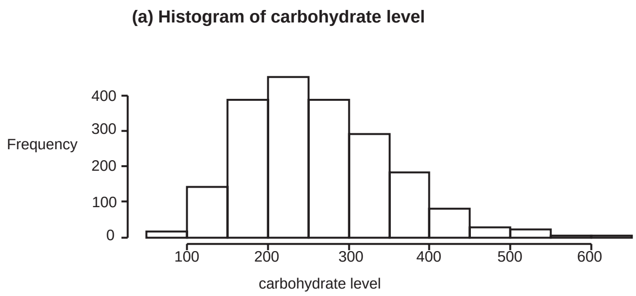
This gives $\frac{106.70}{367.00} = 0.297$.

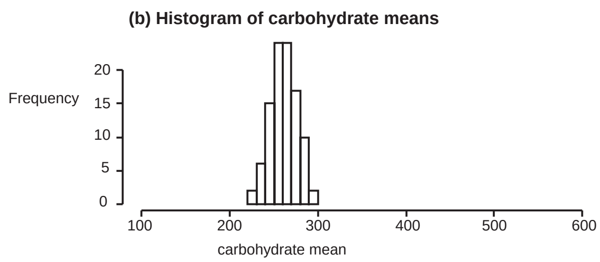
(b) Cancelling the 10^6 terms in the displayed equation above, we see that $\widehat{p} = \frac{27}{367} \times 0.1 + \frac{250}{367} \times 0.2 + \frac{90}{367} \times 0.6$ which is of the form $a_{can}\widehat{p}_{can} + a_{usa}\widehat{p}_{usa} + a_{mex}\widehat{p}_{mex}$.

(c) $\text{sd}(\widehat{p}) = \sqrt{\text{sd}(a_{can}\widehat{p}_{can})^2 + \text{sd}(a_{usa}\widehat{p}_{usa})^2 + \text{sd}(a_{mex}\widehat{p}_{mex})^2}$
$= \sqrt{a_{can}^2\text{sd}(\widehat{p}_{can})^2 + a_{usa}^2\text{sd}(\widehat{p}_{usa})^2 + a_{mex}^2\text{sd}(\widehat{p}_{mex})^2}$
Thus,
$\text{se}(\widehat{p}) = \sqrt{a_{can}^2\text{se}(\widehat{p}_{can})^2 + a_{usa}^2\text{se}(\widehat{p}_{usa})^2 + a_{mex}^2\text{se}(\widehat{p}_{mex})^2}$
$= \sqrt{a_{can}^2\frac{\widehat{p}_{can}(1-\widehat{p}_{can})}{n_{can}} + a_{usa}^2\frac{\widehat{p}_{usa}(1-\widehat{p}_{usa})}{n_{usa}} a_{mex}^2\frac{\widehat{p}_{mex}(1-\widehat{p}_{mex})}{n_{mex}}}$
$= \sqrt{(\frac{27}{367})^2 \times \frac{0.1\times0.9}{1000} + (\frac{250}{367})^2 \times \frac{0.2\times0.8}{1000} + (\frac{90}{367})^2 \times \frac{0.6\times0.4}{1000}}$
$= 0.00944$

Note about simulation exercises: Simulation is a random process. Every time you do it, you get different answers. The results your simulations produce for questions 21–23 will differ from ours in detail but should be fairly similar.

21. (a) Using the Chi-square distribution with 17 degrees of freedom, 1000 "carbohydrate levels" were generated and a histogram of the data follows. It is similar in shape to Fig. 6.1.1.

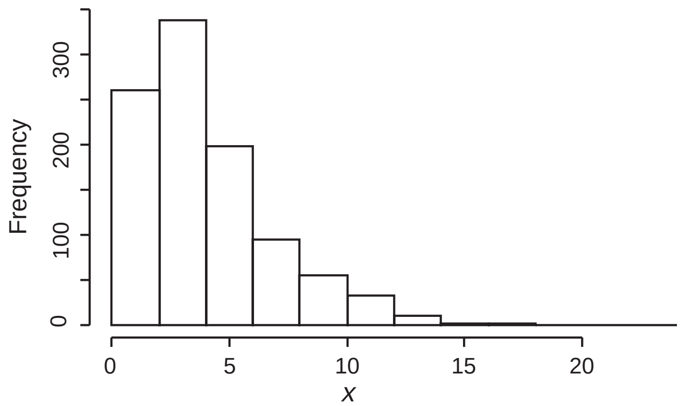




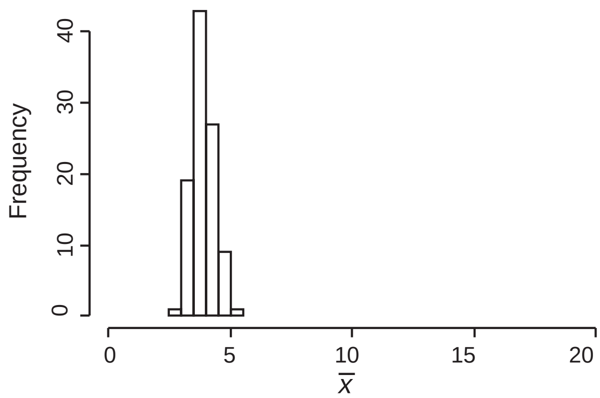
(b) 100 samples of size 25 were generated and a histogram of the 100 sample means is given above. It is plotted against the same horizontal scale as the plot for individual observations above it. We note that the histogram of sample means has a much smaller spread, showing that sample means are much less variable than individual observations. The histogram of the individual observations is very skewed while that of the sample means is much more bell shaped and Normal looking (a consequence of the central limit effect).

(c) Another 1000 carbohydrate levels were generated using a Chi-square distribution with 4 degrees of freedom. The histogram of the 1000 Chi-square random numbers and the histogram of the 100 sample means from samples of size 25 are given below. We observe the same sort of behavior as above.

Histogram of 1000 Chisquare(4) observations



Histogram of 100 means from samples of size 25



23. (a) For you to answer. Many people think that there should be no difference because the true percentage is 50% in both cases. This exercise is intended to bring home to you in a fairly concrete way that sample proportions from large samples are less variable than proportions from small samples, and thus are less likely to give a value as far away from the true proportion (0.5) as 0.7.

(b) By generating 10 Binomial($n = 1, p = 0.5$) random numbers, a simulation of the 10-question test follows:

0 0 0 1 0 0 1 0 1 0

(c) A simulation of 15 people guessing answers to the 10-question test is given below. Each column relates the results for one "person."

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	0	1	0	1	0	1	1	0	1
1	1	0	0	0	0	0	1	0	1	0	0	1	1	0
1	1	1	0	0	0	1	0	1	1	1	1	0	1	1
1	1	1	1	1	0	1	0	1	1	0	1	0	1	0
1	1	1	0	1	0	1	0	1	0	0	1	1	0	0
0	1	1	0	1	1	1	1	1	1	1	0	0	1	0
0	1	0	0	1	1	1	1	0	1	1	0	0	1	1
0	0	0	1	0	1	1	0	1	1	0	0	0	0	0
0	0	1	1	0	0	1	1	0	0	0	1	1	0	1
1	0	0	0	1	0	1	0	1	1	1	0	0	1	0

Adding the number of answers each person got right (i.e., counting the onesdown a column, or equivalently, adding down the column) gives

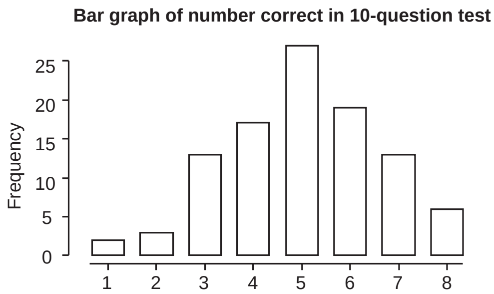
6 7 6 4 6 4 8 5 6 8 4 5 4 6 4

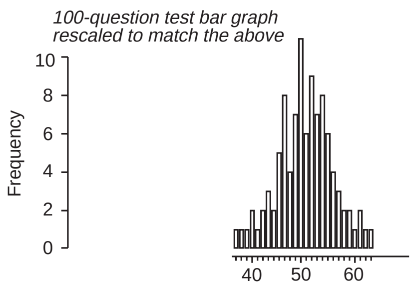
Thus, our 1st person got 6 correct, our 2nd got 7 correct, our 3rd got 6 correct, and so on.

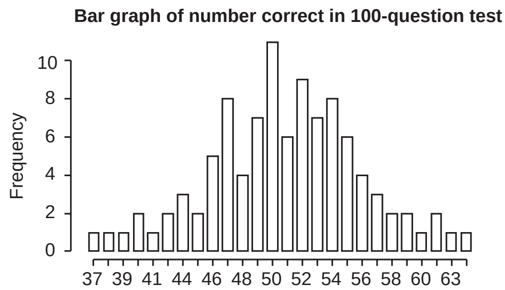
(d) The four conditions required for the Binomial model to be valid are satisfied. By generating 15 Binomial$(n = 10, p = 0.5)$ random numbers, the number of correct answers for 15 people randomly guessing in the 10-question test follows:

6 6 6 6 8 5 5 5 4 3 4 6 6 7 4

(e) Simulations of 100 people taking the 10-question test and the 100-question test are displayed in the bar graphs below.







The proportion of people who got at least 7 of the 10 questions correct was $\frac{19}{100} = 0.19$ and the proportion of people who got at least 70 of the 100 questions correct was $\frac{0}{100} = 0$. (The highest mark we observed was 64.)

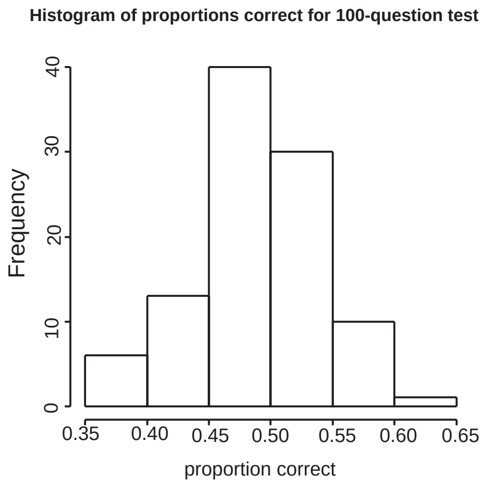
(f) The theoretical result is $\text{sd}(\widehat{P}) = \sqrt{\frac{p(1-p)}{n}}$ which is proportional to $\frac{1}{\sqrt{n}}$. This tells us that the variability in the values of $\widehat{P}$ decreases as the sample size increases.

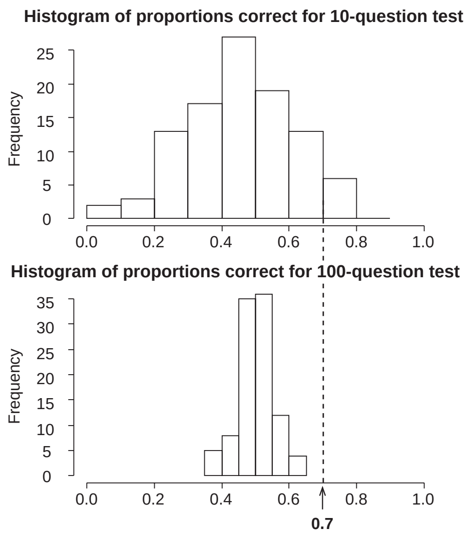
(g) For the people doing the 100-question test in (e) the 100 values of "proportion correct" are as follows:

```
0.55 0.50 0.57 0.51 0.48 0.56 0.49 0.51 0.41 0.57 0.47 0.50 0.38 0.64 0.52
0.50 0.51 0.46 0.52 0.49 0.63 0.50 0.53 0.52 0.47 0.50 0.49 0.47 0.53 0.50
0.59 0.59 0.54 0.58 0.55 0.49 0.40 0.61 0.37 0.54 0.49 0.58 0.52 0.54 0.44
0.49 0.51 0.51 0.46 0.55 0.44 0.48 0.40 0.55 0.53 0.39 0.48 0.50 0.50 0.56
0.61 0.42 0.52 0.52 0.55 0.52 0.57 0.42 0.46 0.54 0.56 0.52 0.47 0.53 0.47
0.46 0.54 0.47 0.44 0.50 0.53 0.60 0.45 0.48 0.53 0.55 0.49 0.47 0.54 0.56
0.51 0.45 0.50 0.47 0.54 0.46 0.52 0.50 0.53 0.54
```

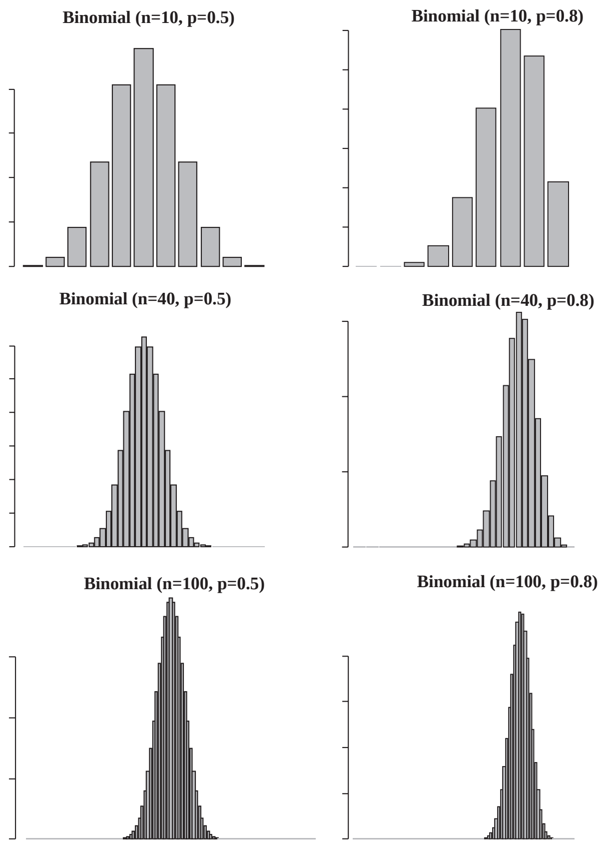
The standard deviation of these sample proportions is 0.053. This is very similar to the theoretical standard deviation of $\widehat{P}$ when $p = 0.5$, namely $\text{sd}(\widehat{P}) = \sqrt{\frac{0.5(1-0.5)}{100}} = 0.05$.

(h) A histogram of the proportions from (g) is given on the left below. The histogram is bell shaped but not completely symmetrical. [More interesting histograms are given on the right. These show the decrease in variability in the $\widehat{P}$ values with the larger sample size (number of questions). As the variability about the true value of 0.5 or 50% contracts, the proportion of values above 0.7 (70%) decreases.]





(i) Bar graphs of Binomial probabilities of the Binomial$(n, p = 0.5)$ for $n = 10$, $n = 40$, and $n = 100$ are given below. All are symmetrically bell shaped. We note the decreasing spread as n increases. On the right we have the same thing for $p = 0.8$. The bar graph is quite skewed when $n = 10$ but is symmetrical by the time $n = 100$.


Binomial (n=10, p=0.5)
Binomial (n=10, p=0.8)
Binomial (n=40, p=0.5)
Binomial (n=40, p=0.8)
Binomial (n=100, p=0.5)
Binomial (n=100, p=0.8)

Chapter 8 Confidence Intervals

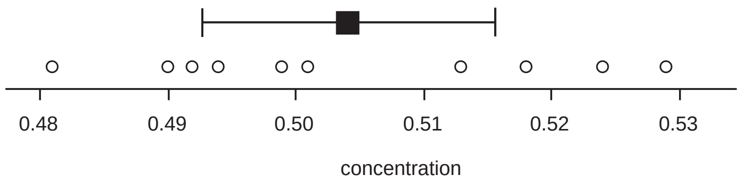
Exercises for Section 8.2

For all of these problems, we will be using the following formula $\overline{x} \pm t\text{se}(\overline{x}) = \overline{x} \pm t\frac{s_X}{\sqrt{n}}$ to construct our confidence intervals. "Confidence interval" is frequently abbreviated to "CI." The confidence intervals in problems 1 and 2 can be also be generated automatically using a package like Minitab or Excel (see Section 10.1.1). Of greatest importance is learning to interpret the intervals in the context of the particular data set.

1. We have $df = n - 1 = 5$ so that the t multiplier for a 95% CI is $t = 2.5706$. The 6 observations have sample mean and standard deviation given by $\overline{x} = 5.3117$ and $s_X = 0.2928$ respectively. The resulting 95% CI is $5.3117 \pm 2.5706 \times \frac{0.2928}{\sqrt{6}}$ or approximately [5.00, 5.62]. From this data, we can say with 95% confidence that the true mean density of the earth is somewhere between $5.0\,\text{g/cm}^3$ and $5.6\,\text{g/cm}^3$.

2. **(a)** We have $df = n - 1 = 9$ so that the t multiplier for a 95% CI is $t = 2.2622$. The 10 observations have sample mean and standard deviation given by $\overline{x} = 0.5041$ and $s_X = 0.0160$ respectively. The resulting 95% CI is $0.5041 \pm 2.2622 \times \frac{0.0160}{\sqrt{10}}$ or approximately [0.493, 0.516]. With 95% confidence, the true nitrate ion concentration is somewhere between $0.49\,\mu\text{g/mL}$ and $0.52\,\mu\text{g/mL}$.

(b)



3. **(a)** **(i)** Here $df = n - 1 = 61$ so for a 95% CI, $t = 1.9996$. The resulting CI is $620.6 \pm 1.9996 \times \frac{241.5}{\sqrt{62}}$ or approximately [559, 682]. With 95% confidence, the true or population mean testosterone level for nonsmokers is somewhere between 559 ng/dL and 682 ng/dL.

(ii) Here $df = n - 1 = 27$ so for a 95% CI, $t = 2.0518$. The resulting CI is $795.1 \pm 2.0518 \times \frac{305.3}{\sqrt{28}}$ or approximately [677, 913]. With 95% confidence, the true mean testosterone level for the 31–70 per day group is somewhere between 677 ng/dL and 913 ng/dL.

(b) When $df = 27$, $t = 1.7033$ for a 90% CI and $t = 2.7707$ for a 99% CI. The resulting CIs for the true mean testosterone level are:

(i) (90% CI) $795.1 \pm 1.7033 \times \frac{305.3}{\sqrt{28}}$, or [697, 893].

(ii) (99% CI) $795.1 \pm 2.7707 \times \frac{305.3}{\sqrt{28}}$, or [635, 955].

Exercises for Section 8.3

1. In Example 8.3.1, $n = 200$ and $\widehat{p} = 0.7$. Our CI formula is $\widehat{p} \pm z\text{se}(\widehat{p}) = \widehat{p} \pm z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$. The value of the multiplier z depends upon the confidence level.

(a) (90% CI) $0.7 \pm 1.6449 \times \sqrt{\frac{0.7\times 0.3}{200}}$, or [0.647, 0.753].

(b) (99% CI) $0.7 \pm 2.5758 \times \sqrt{\frac{0.7\times 0.3}{200}}$, or [0.617, 0.783].

2. The 95% CI is $0.36 \pm 1.96 \times \sqrt{\frac{0.36\times 0.64}{139}}$, or approximately [0.280, 0.440]. With 95% confidence, the true (or population) proportion of Hispanic people who have been pulled over on the roads by the police is somewhere between 28% and 44%.

Exercises for Section 8.4

We are using the confidence interval formula for a difference between two means from independent samples, namely $\overline{x}_1 - \overline{x}_2 \pm t\,\mathrm{se}(\overline{x}_1 - \overline{x}_2) = \overline{x}_1 - \overline{x}_2 \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

1. **(a)** We will use $df = min(5, 22) = 5$. For a 95% CI the t multiplier is $t = 2.5706$. Our 95% CI is $5.3117 - 5.4835 \pm 2.5706 \times \sqrt{\frac{0.2928^2}{6} + \frac{0.1904^2}{23}}$, or [−0.496, 0.152].

(b) The two-standard-error interval is $5.3117 - 5.4835 \pm 2 \times \sqrt{\frac{0.2928^2}{6} + \frac{0.1904^2}{23}}$, or [−0.424, 0.080], which is narrower.

(c) From either interval there is no evidence of a difference between the two true means, as the interval contains zero.

2. We use $df = min(30, 27) = 27$. For a 95% CI, $t = 2.0518$. Our 95% CI is $715.6 - 795.1 \pm 2.0518 \times \sqrt{\frac{248^2}{31} + \frac{305.3^2}{28}}$, or [−229, 70]. With 95% confidence, the true mean testosterone level for 1–30 per day smokers falls somewhere between smaller than that for 31–70 per day smokers by 229 ng/dL and larger by 70 ng/dL. This includes the possibility that there is no difference at all.

Exercises for Section 8.5.1 and 8.5.2

The confidence intervals asked for in these exercises are all for differences between two proportions and are of the form $\widehat{p}_1 - \widehat{p}_2 \pm z\,\mathrm{se}(\widehat{p}_1 - \widehat{p}_2)$. The formula used for the standard error is as given in Table 8.5.5. We will not repeat these formulas here but simply indicate whether we are dealing with a sampling situation (a), (b), or (c) as depicted in Fig. 8.5.1.

1. **(a)** Situation (c). **(b)** Situation (a). **(c)** Situation (a). **(d)** Situation (b). **(e)** Situation (a).

2. **(a)** $0.51 - 0.21 \pm 1.96 \times \sqrt{\frac{0.51+0.21-(0.51-0.21)^2}{500}}$, or [0.23, 0.37].
With 95% confidence, the population percentage of 15- to 17-year-olds who know a student who sells illegal drugs is bigger than the percentage who know a teacher who uses illegal drugs by somewhere between 23 and 37 percentage points.

(b) $0.51 - 0.22 \pm 1.96 \times \sqrt{\frac{0.51\times 0.49}{500} + \frac{0.22\times 0.78}{500}}$, or [0.23, 0.35].
With 95% confidence, the population percentage of 15- to 17-year-olds who know a student who sells illegal drugs is bigger than the corresponding percentage for 12- to 14-year-olds by somewhere between 23 and 35 percentage points.

(c) $0.35 - 0.23 \pm 1.96 \times \sqrt{\frac{0.35\times0.65}{822} + \frac{0.23\times0.77}{500}}$, or [0.07, 0.17].
With 95% confidence, the population percentage of principals who think students can use marijuana every weekend and still do well at school is bigger than the corresponding percentage for 15- to 17-year-olds by somewhere between 7 and 17 percentage points.

(d) $0.21 - 0.13 \pm 1.96 \times \sqrt{\frac{0.21+0.13-(0.21-0.13)^2}{500}}$, or [0.03, 0.13].
With 95% confidence, the population percentage of 12- to 14-year-olds who are most likely to hang out with friends after school is bigger than the percentage who go home and watch TV by somewhere between 3 and 13 percentage points.

(e) $0.22 - 0.16 \pm 1.96 \times \sqrt{\frac{0.22\times0.78}{500} + \frac{0.16\times0.84}{500}}$, or [0.01, 0.11].
With 95% confidence, the population percentage of 12- to 14-year-olds who are most likely to hang out with friends after school is bigger than the corresponding percentage of 15- to 17-year-olds by somewhere between 1 and 11 percentage points.

3. (a) [Situation (b)] $0.59 - 0.25 \pm 1.96 \times \sqrt{\frac{0.59+0.25-(0.59-0.25)^2}{1000}}$, or [0.29, 0.39].
With 95% confidence, the population percentage of New York voters who supported Clinton was bigger than the percentage who supported Dole by somewhere between 29 and 39 percentage points.

(b) [Situation (a)] $0.33-0.29\pm1.96\times\sqrt{\frac{0.33\times0.67}{1000} + \frac{0.29\times0.71}{1000}}$, or [−0.001, 0.08]. With 95% confidence, the population percentage of voters who supported Dole in New Jersey was somewhere between being the same as the percentage in Connecticut and being larger by 8 percentage points than in Connecticut.

(c) [Situation (a)] $0.28 - 0.2 \pm 1.96 \times \sqrt{\frac{0.28\times0.72}{1000} + \frac{0.2\times0.8}{1000}}$, or [0.04, 0.12].
With 95% confidence, the population percentage of Americans worried about difficulties in getting health care was larger than the corresponding percentage for Canadians by somewhere between 4 and 12 percentage points.

(d) [Situation (c)] $0.38 - 0.32 \pm 1.96 \times \sqrt{\frac{0.38+0.32-(0.38-0.32)^2}{1000}}$, or [0.01, 0.11]. With 95% confidence, the population percentage of New Zealanders who think recent changes have harmed the quality of health care was larger than the percentage who believe the system should be rebuilt by somewhere between 1 and 11 percentage points.

(e) The people in the UK, which spends least on health care, seem happiest with their system.

Exercises for Section 8.6

1. (a) Take $n \geq \left(\frac{1.96}{.02}\right)^2 \times 0.5 \times 0.5 \approx 2401$.

(b) Take $n \geq \left(\frac{1.96}{.02}\right)^2 \times 0.15 \times 0.85 \approx 1225$.

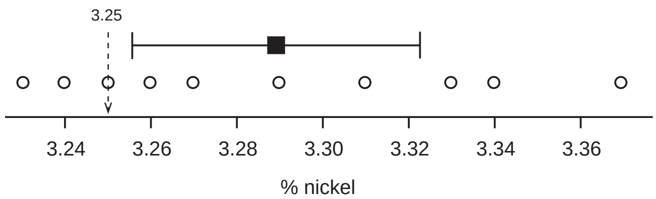
(c) Take $n \geq \left(\frac{1.96}{.02}\right)^2 \times 0.85 \times 0.15 \approx 1225$.

2. (a) Take $n \geq \left(\frac{1.96\times.07559}{.025}\right)^2 \approx 35$.

(b) Take $n \geq \left(\frac{1.96\times.44049}{.025}\right)^2 \approx 1193$.
The difference is so big because thiol measurements are much more variable in the rheumatoid population than they are in the normal population so substantially more people must be sampled from the rheumatoid population to estimate the population mean to the same level of precision.

Review Exercises 8

1. (a)



No, the dot plot looks well behaved.

(b) $\overline{x} = 3.289$, $s_X = 0.04701$

(c) With $df = n - 1 = 9$, $t = 2.262$ for a 95% CI. The resulting 95% CI for the true percent nickel content is $3.289 \pm 2.262 \times \frac{0.04701}{\sqrt{10}}$, or [3.26%, 3.32%]. Yes, there is evidence that this batch differs from previous batches as the usual mean nickel content for previous batches (3.25%) lies outside the 95% confidence interval for the true mean nickel content of this batch.

(d) Added to the plot in (a) above.

The interval is now $3.289 \pm 2.022691 \times \frac{0.04701064}{\sqrt{40}}$, or [3.27, 3.30].

If the multiplier did not change, the width of the confidence interval would halve. The multiplier also gets slightly smaller with the increase in *df*, so with more significant figures, you will see that the width of the new interval for $n = 40$ is slightly less than half the width of the interval for $n = 10$.

***(e)** Take $n \geq \left(\frac{1.96\times0.04701064}{.015}\right)^2 \approx 38$

3. (a) A single piece of paper may look like it would take less time to answer. If there was a difference, we would expect the single sheet version would have a higher response rate.

(b) We have independent samples (situation (a) in Fig. 8.5.1) so we use the corresponding formula in Table 8.5.5 for $\text{se}(\widehat{p}_1 - \widehat{p}_2)$. We have $\widehat{p}_1 = 0.36$, $\widehat{p}_2 = 0.3$, $n_1 = 220$, and $n_2 = 220$, giving $\text{se}(\widehat{p}_1 - \widehat{p}_2) = \sqrt{\frac{0.36\times0.64}{220} + \frac{0.30\times0.70}{220}} = 0.044742$.

(90% CI) $0.36 - 0.30 \pm 1.6449 \times 0.044742$, or [−0.014, 0.13]. With 90% confidence, the true response rate for the one-sheet version is somewhere between being 1.4 percentage points lower than for the two-sheet version and 13 percentage points higher.

(95% CI) $0.36 - 0.30 \pm 1.96 \times 0.044742$, or [−0.03, 0.15]. When we change to a 95% CI our interval becomes wider (less precise). With 90% confidence, the true response rate for the one-sheet version is somewhere between being 3 percentage points lower than for the two-sheet version and 15 percentage points higher.

(c) Since both intervals contain zero, the change in printing format may make no difference to the response rate. We would be inclined to use the two-sided version in accordance with our intuition and the slight suggestion of an increased response rate given by the data. We note that both response rates were quite low.

***5.** **(a)** Take $n \geq \left(\frac{2.5758}{.04}\right)^2 \times 0.5 \times 0.5 \approx 1037$.

(b) Take $n \geq \left(\frac{2.5758}{.04}\right)^2 \times 0.4 \times 0.6 \approx 995$ (which is only slightly smaller).

(c) Take $n \geq \left(\frac{2.5758}{.04}\right)^2 \times 0.6 \times 0.4 \approx 955$ (which is the same as for (b)).

(d) Practical questions including how to sample students from the many schools in a city. Definitional questions such as how to handle students whose parents were never legally married but have now split. (Whether to include them depends on the purpose of the survey. Are you interested in legalities or whether the students are living with both parents?)

7. **(a)** We would plot the data using box plots to compare groups (as these groups are quite large). We would also look at stem-and-leaf plots or histograms to look at distributional shape.

(b) We use $df = \min(n_1 - 1, n_2 - 1) = 89$ in determining the size of the multiplier t. The only difference between the 95% confidence interval and the two-standard-error interval we calculated in problem 15(b) in Review Exercises 7 is that we are now using $t = 1.9870$ standard errors rather than 2 standard errors. Not surprisingly, we get virtually identical intervals. Our 95% CI for the true difference in means is $103.0 - 92.8 \pm 1.9870 \times \sqrt{\frac{17.39^2}{210} + \frac{15.18^2}{90}}$, or [6.2, 14.2]. Breast-fed babies have IQs that are higher on average than bottle-fed babies by somewhere between 6 and 14 points.

***(c)** Take $n \geq (\frac{1.96 \times 17.39}{1})^2 \approx 1162$.

(d) The babies were all pre-term and very small, and only from special care units in several areas in England. The results may be special to this population.

(e) It is an observational study in which mothers chose whether to breast feed. The study does not demonstrate that the effect is causal.

(f) The CI would change to $103.0 - 92.8 \pm 2.144787 \times \sqrt{\frac{17.39^2}{210} + \frac{15.18^2}{15}}$, or [1.4, 19.0]. The interval has become more than twice as wide.

(g) This problem is very similar to Example 6.4.2. We want pr$(X < Y)$ where $X \sim$ Normal$(\mu_X = 103.0, \sigma_X = 17.39)$ and $Y \sim$ Normal$(\mu_Y = 92.8, \sigma_Y = 15.18)$. pr$(X < Y) =$ pr$(X - Y < 0) =$ pr$(W < 0)$. Here $W = X - Y$ has mean $\mu_W = 103.0 - 92.8 = 10.2$ and standard deviation $\sigma_W = \sqrt{17.39^2 + 15.18^2} = 23.0834$. Using these values, pr$(W < 0) = 0.3293$. For any two randomly selected babies, there is approximately 1 chance in 3 that the bottle-fed baby will have a higher IQ. The confidence interval is only talking about the difference between the means and says nothing about any other aspect of the distribution. In fact, there is substantial overlap between the IQ distributions for both groups.

9. **(a)** The data suggests that ex-smokers have healthier eating patterns on average than smokers, both when we look within manual workers and when we look within non-manual workers. Similarly, non-manual workers seem to have healthier eating patterns on average than manual workers, both when we look within smokers and within ex-smokers.

(b) All confidence intervals calculated here are 95% CIs for differences between proportions from independent samples (situation (a) in Fig. 8.5.1). All are calculated using $\widehat{p}_1 - \widehat{p}_2 \pm 1.96\,\text{se}(\widehat{p}_1 - \widehat{p}_2)$ where $\text{se}(\widehat{p}_1 - \widehat{p}_2) = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$.
Using only non-manual workers, the following are 95% confidence intervals for differences in true proportions between ex-smokers who consume the item (e.g., breakfast) and smokers who consume the item. The sample sizes are always $n_1 = 517$ and $n_2 = 404$.
Breakfast: $\widehat{p}_1 = 0.824$, $\widehat{p}_2 = 0.629$, and the CI is [0.14, 0.25].
Brown bread: $\widehat{p}_1 = 0.536$, $\widehat{p}_2 = 0.345$, and the CI is [0.12, 0.25].
Fresh Fruit: $\widehat{p}_1 = 0.776$, $\widehat{p}_2 = 0.594$, and the CI is [0.12, 0.24].
Fried food: $\widehat{p}_1 = 0.162$, $\widehat{p}_2 = 0.282$, and the CI is [−0.17, −0.07].
We have clearly demonstrated that the ex-smokers do better than the non-smokers when it comes to both having breakfast and having a healthy breakfast. For example, with 95% confidence, the true percentage of exsmokers consuming breakfast is larger than that for smokers by somewhere between 14 and 25 percentage points. The other intervals are all read similarly. The only exception is the last interval which tells us that, with 95% confidence, the true percentage of exsmokers consuming fried food is smaller than that for smokers by somewhere between 7 and 17 percentage points.

11. **(a)** 95% CI: $0.48 \pm 1.96 \times \sqrt{\frac{0.48\times0.52}{2700}}$, or [0.461, 0.500]. With 95% confidence, the true percentage of Independents who voted Republican was somewhere between 46% and 50%.

(b) We are comparing proportions from independent samples (situation (a) in Fig. 8.5.1). Our 95% CI for $p_{94} - p_{98}$ is $0.55 - 0.48 \pm 1.96 \times \sqrt{\frac{0.55\times0.45}{2700} + \frac{0.48\times0.52}{2700}}$, or [0.043, 0.097]. With 95% confidence, the true percentage of Independents who voted Republican in 1998 was smaller than that in 1994 by somewhere between about 4 and 10 percentage points.

(c) We are comparing proportions from independent samples (situation (a) in Fig. 8.5.1). Our 95% CI for $p_{4year} - p_{postgrad}$ is
$0.53 - 0.45 \pm 1.96 \times \sqrt{\frac{0.53\times0.47}{2700} + \frac{0.45\times0.55}{1800}}$, or [0.050, 0.110]. With 95% confidence, the true percentage of 4-year college graduates who voted Republican was larger than that for people who have done postgraduate study by somewhere between about 5 and 11 percentage points.

(d) 95% CIs for population proportions of ethnic group voting Republican:
White [0.539, 0.561]; Black [0.09, 0.13]; Hispanic [0.31, 0.39]; Asian [0.32, 0.52].
The 95% CI for difference between population proportions of Asians and Hispanics voting Republican (situation (a) comparison):
$0.42 - 0.35 \pm 1.96 \times \sqrt{\frac{0.42\times0.58}{100} + \frac{0.35\times0.65}{500}}$, or [−0.04, 0.18].

13. The actual margin of error associated with $\widehat{p} = 0.03$ is $1.96 \times \sqrt{\frac{.03\times.97}{1000}} \approx 0.0106$, which is very close to 1%.

15. **(a)** When $df = n - 1 = 38$, for a 95% CI we use $t = 2.024$. The resulting 95% CI is (using the formula in answer to 14(a)): $10.97 \pm 2.024 \times \frac{2.67}{\sqrt{39}}$, or [10.1, 11.8]. With 95% confidence the true mean score under control conditions lies somewhere between 0.1 and 11.8.

(b) We will use $df = \min(n_1 - 1, n_2 - 1) = 17$, so for a 95% CI we use $t = 2.110$. The resulting 95% CI is (using the formula in answer to 14(b)): $13.28 - 10.97 \pm 2.110 \times \sqrt{\frac{1.9^2}{18} + \frac{2.67^2}{39}}$, or [1.0, 3.6]. With 95% confidence, the true mean score under "humane/no info." conditions is larger than it is under control conditions by somewhere between about 1 and 6.

(c) Interval set up is as for (b). The resulting 95% CI is: $10.97 - 10.44 \pm 2.110 \times \sqrt{\frac{2.67^2}{39} + \frac{2.43^2}{18}}$, or $[-0.98,\ 2.04]$. As zero lies in this interval we cannot tell whether there is a true difference or in what direction such a difference lies. What we can say with 95% confidence is that the true mean score under control conditions lies somewhere between being smaller than it is under "inhumane/no info." conditions by approximately 1 and and being larger by 2.04.

***17** **(a)** For independent samples (situation (a) in Fig.8.5.1), the margin of error is z standard errors or $z\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$. When $n_1 = n_2 = n$, this becomes $\frac{z}{\sqrt{n}}\sqrt{\widehat{p}_1(1-\widehat{p}_1) + \widehat{p}_2(1-\widehat{p}_2)}$. Solving $\frac{z}{n}\sqrt{\widehat{p}_1(1-\widehat{p}_1) + \widehat{p}_2(1-\widehat{p}_2)} \le w$ for n gives the desired expression.

(b) Since taking $\widehat{p} = 0.5$ maximizes $\widehat{p}(1-\widehat{p})$, we should use $\widehat{p}_1 = 0.5$ and $\widehat{p}_2 = 0.5$.

(c) Using these values we get $n \ge \left(\frac{z}{w}\right)^2 \times 0.5 = \frac{1}{2}\left(\frac{z}{w}\right)^2$.

(d) Arguing as in (a), we get get $n \ge \left(\frac{z}{w}\right)^2 \times [\widehat{p}_1 + \widehat{p}_2 - (\widehat{p}_1 - \widehat{p}_2)^2]$. The biggest value this can take also happens when $\widehat{p}_1 = 0.5$, and $\widehat{p}_2 = 0.5$ giving $n \ge \left(\frac{z}{w}\right)^2$.

19. **(a)** 100 samples of 9 "male heights" following a $N(\mu = 174\text{cm}, \sigma = 6.57\text{cm})$ distribution were generated. For each sample, a 95% confidence interval for the true mean was obtained using the formula $\overline{x} \pm t\,\text{se}(\overline{x})$. Here $n = 9$ so that $df = 8$ and $t = 2.306$. The 100 95% confidence intervals were:

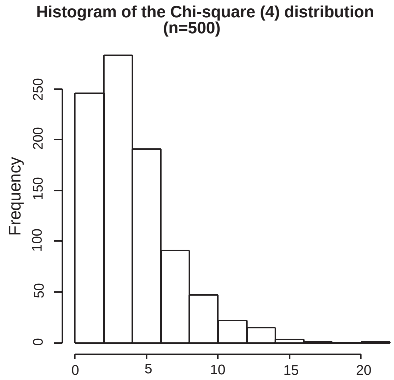
```
(172.3962, 178.6724); (174.4476, 183.1760); (173.8045, 178.3554);
   .....      .....      .....     .....       .....     .....
(169.0617,176.4371); (169.2078, 177.2485);  (170.8897, 182.4987).
```

The proportion of our intervals that contain the true mean 174 is $\frac{96}{100} = 0.96$. The average width of the 100 intervals is 9.66.

(b) This time 100 samples of 25 male heights were generated and, for each sample, a 95% confidence interval for the true mean was obtained. We now have $df = 24$ and $t = 2.064$. The proportion of the intervals that contain the true mean 174 is $\frac{94}{100} = 0.94$. The average width of the 100 intervals is 5.39, which is 4.27 less than than the average width of the intervals from (a).

[We expect the intervals to have approximately a 95% coverage no matter what the sample size is. The length of the intervals, however, is proportional to $\frac{1}{\sqrt{n}}$ (apart from a minor *df* effect), so the intervals get shorter as n increases. (See Fig. 8.1.4).]

21. (a) A histogram of 500 observations from a Chi-square distribution with 4 degrees of freedom is shown below.



(b) 100 samples of size 9 were generated. For each sample a 95% confidence interval was obtained using the formula $\overline{x} \pm t\,\text{se}(\overline{x})$. Here, $df = n - 1 = 8$ and $t = 2.306$. The one hundred 95% confidence intervals were:

```
(1.5800846, 7.845032); (1.9556367, 6.000979); (1.6196485, 5.745535);
   ......     .....       .....     .....       .....      .....
(2.0805119, 5.502391); (1.8812885, 4.266043); (1.2595665, 3.630293).
```

The proportion of our intervals that contained the true mean (4) was $\frac{88}{100} = 0.88$.

(c) We next generated 100 samples of size 25. Here, $df = n - 1 = 24$ and the t multiplier is $t = 2.0639$. In this case the proportion of our intervals that contained the true mean (4) was $\frac{93}{100} = 0.93$.

(d) When 1000 samples of size 9 and 1000 samples of size 25 were generated, the proportion of intervals that contained the true mean (4) was $\frac{903}{1000} = 0.903$ and $\frac{946}{1000} = 0.946$ respectively. For a discussion of the issues involved here, see page 411 of the book.

(e) It is working very well by the time $n = 25$ and it is not too terrible even at $n = 9$.