4.5 The hypotheses are:

$$H_0: \quad \mu_A = \mu_B$$
$$H_a: \quad \mu_A \neq \mu_B$$

4.7 The hypotheses are:

$$H_0: \quad \mu = 50$$
$$H_a: \quad \mu < 50$$

4.16 (a) We define p_h to be the proportion with ADHD in the high pesticide group and p_l to be the proportion with ADHD in low pesticide group. The hypotheses are:

$$H_0: \quad p_h = p_l$$
$$H_a: \quad p_h > p_l$$

(b) No. Just because in the sample we have $\hat{p}_h > \hat{p}_l$, we can't assume that the same must be true for the population proportions.

(c) Assume the proportion of children with ADHD is the same for the high and low pesticide groups.

(d) Statistically significant results mean strong evidence against H_0, so the null hypothesis (that pesticide exposure is not related to the likelihood of an ADHD diagnosis) is no longer plausible.

(e) Children with high exposure to DAP are more likely to be diagnosed with ADHD than children with low exposure to the pesticide. (Note that these results come from an observational study rather than an experiment so we need to be sure that the statement of our conclusion does not imply causation.)

4.24 Notice that this is a test for a single proportion. We define p to be the proportion of ICU patients who are female. (We could also have defined p to be the proportion who are male. The test will work fine either way.) The hypotheses are:

$$H_0: \quad p = 0.5$$
$$H_a: \quad p \neq 0.5$$

4.25 We define μ to be the mean age of ICU patients. The hypotheses are:

$$H_0: \quad \mu = 50$$
$$H_a: \quad \mu > 50$$

4.28 (a) We define μ_r to be the mean metabolism rate for this species after taking a resveratrol supplement and μ_p to be the mean metabolism rate after taking a placebo. The hypotheses are:

$$H_0: \quad \mu_r = \mu_p$$
$$H_a: \quad \mu_r > \mu_p$$

(b) For species A, the sample mean for the resveratrol group is greater than the mean for the placebo group, but not by much and the distributions in the boxplots overlap quite a bit, so the difference may not be very significant. Thus we probably could not conclude that resveratrol increases metabolism for this species (although we can't be sure without doing a statistical test.)

(c) For species B, the sample mean for the resveratrol group is quite a bit greater than the mean for the placebo group. In fact, most of the values in the resveratrol sample are higher than almost all the values in the placebo sample, with less variability in both cases than the plots for species A. This gives stronger evidence that resveratrol increases the metabolism rate for species B. (Although, again, we cannot be sure without doing a statistical test.)

4.29 (a) We define μ to be the average amount of omega-3 in one tablespoon of this brand of milled flaxseed. Since the company is looking for evidence that the average is greater than 3800, the hypotheses are:

$$\begin{aligned} H_0 &: \quad \mu = 3800 \\ H_a &: \quad \mu > 3800 \end{aligned}$$

(b) We define μ as in part (a). Since the consumer organization is testing to see if there is evidence that the average is less than 3800, the hypotheses are:

$$\begin{aligned} H_0 &: \quad \mu = 3800 \\ H_a &: \quad \mu < 3800 \end{aligned}$$

4.30 This analysis does not involve a test because there is no claim of interest. We would likely use a confidence interval to estimate the average.

4.31 This analysis does involve a statistical test. The population parameter p is the proportion of people in the community living in a mobile home. The hypotheses are:

$$\begin{aligned} H_0 &: \quad p = 0.10 \\ H_a &: \quad p > 0.10 \end{aligned}$$

4.32 This analysis does not include a test because from the information in a census, we can find exactly the true population proportion.

4.33 This analysis does include a statistical test. This is a matched pairs experiment, so the population parameter is μ_D, the average difference of reaction time (right $-$ left) for all right-handed people. The hypotheses are $H_0 : \mu_D = 0$ vs $H_a : \mu_D < 0$. We could also write the hypotheses as $H_0 : \mu_R = \mu_L$ vs $H_a : \mu_R < \mu_L$.

4.34 This analysis does not include a test because there is no claim of interest. The analysis would probably include a confidence interval to give an estimate of the average reaction time.

4.35 This analysis does include a statistical test for a single proportion. The population parameter p is the proportion of people in New York City who prefer Pepsi and we are testing to see if the proportion who prefer Pepsi is greater than 50%. The hypotheses are:

$$\begin{aligned} H_0 &: \quad p = 0.5 \\ H_a &: \quad p > 0.5 \end{aligned}$$

4.36 This analysis does not include a statistical test. Since we have all the information for the population, we can compute the proportion who voted exactly and see if it is greater than 50%.

4.52 (a) If the mean arsenic level is really 80 ppb, the chance of seeing a sample mean as high (or higher) than was observed in the sample from supplier A by random chance is only 0.0003. For supplier B, the corresponding probability (seeing a sample mean as high as B's when $\mu = 80$) is 0.35.

(b) The smaller p-value for Supplier A provides stronger evidence against the null hypothesis and in favor of the alternative that the mean arsenic level is higher than 80 ppb. Since it is very rare for the mean to be that large when $\mu = 80$, we have stronger evidence that there is too much arsenic in Supplier A's chickens.

(c) The chain should get chickens from Supplier B, since there is strong evidence that Supplier A's chicken have a mean arsenic level above 80 ppb which is unacceptable.

4.78 Only choice (c) "The probability of seeing data as extreme as the sample, when the null hypothesis, H_0, is true." matches the definition of the p-value. The other choices are common *misinterpretations* of a p-value. The p-value does *not* measure the probability of any hypothesis being true (of false) or the chance of making either type of error. It only measures how unusual the original data would be if the null hypothesis were true.

4.82 (a) The p-value (0.003) is small so the decision is to reject H_0 and conclude that the mean recall for sleep ($\overline{x}_s = 15.25$) is different from the mean recall for caffeine ($\overline{x}_c = 12.25$). Since the mean for the sleep group is higher than the mean for the caffeine group, we have sufficient evidence to conclude that mean recall after sleep is in fact better than after caffeine. Yes, sleep is really better for you than caffeine for enhancing recall ability.

(b) The p-value (0.06) is not less than 0.05 so we would not reject H_0 at a 5% level, but it is less than 0.10 so we would reject H_0 at a 10% level. There is some moderate evidence of a difference in mean recall ability between sleep and a placebo, but not very strong evidence.

(c) The p-value (0.22) is larger than any common significance level, so do not reject H_0. The placebo group had a better mean recall in this sample ($\overline{x}_p = 13.70$ compared to $\overline{x}_c = 12.25$), but there is not enough evidence to conclude that the mean for the population would be different for a placebo than the mean recall for caffeine.

(d) Get a good night's sleep!

4.86 (a) This is an experiment since the explanatory factor (cell phone "on" or "off) was controlled. The design is matched pairs, since all 47 participants were tested under both conditions. For each participant, we find the difference in brain activity between the two conditions.

(b) Randomization in this case means that the order of the conditions ("on" and "off) was randomized for all the participants. Cell phones were on the ears for both conditions to control for any lurking variables and to make the treatments as similar as possible except for the variable of interest (the radiofrequency waves).

(c) Using μ_{on} to represent average brain glucose metabolism when the cell phones are on and μ_{off} to represent average brain glucose metabolism when the cell phones are off, the hypotheses are:

$$\begin{aligned} H_0 &: \quad \mu_{on} = \mu_{off} \\ H_a &: \quad \mu_{on} \neq \mu_{off} \end{aligned}$$

Notice that since this is a matched pairs study, we could also write the hypotheses in terms of the average difference μ_D between the two conditions, with $H_0 : \mu_D = 0$ vs $H_a : \mu \neq 0$.

(d) Since the p-value is quite small (less than a significance level of 0.01), we reject the null hypothesis. There is significant evidence that brain activity is affected by cell phones.

(e) Both of these variables (brain glucose metabolism and amplitude of radiofrequency) are quantitative, so we use a scatterplot to graph the relationship.

(f) We are testing to see if the correlation ρ between these two variables is significantly different from zero, so the hypotheses are

$$\begin{aligned} H_0 &: \quad \rho = 0 \\ H_a &: \quad \rho \neq 0 \end{aligned}$$

where ρ is the correlation between brain glucose metabolism and amplitude of radiofrequency.

(g) This p-value is very small so we reject H_0. There is strong evidence that brain activity is correlated with the amplitude of the radiofrequency waves emitted by the cell phone.

6.10 (a) In this case, we have $p = 0.69$ and $n = 100$. The sample proportions will be centered at the population proportion of $p = 0.69$ so will have a mean of 0.69. The standard deviation of the sample proportions is the standard error, which is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.69(1-0.69)}{100}} = 0.046$$

(b) In this case, we have $p = 0.69$ and $n = 1000$. The sample proportions will be centered at the population proportion of $p = 0.69$ so will have a mean of 0.69. The standard deviation of the sample proportions is the standard error, which is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.69(1-0.69)}{1000}} = 0.015$$

Notice that the standard error is significantly less with a sample size of 1000 than it is with a sample size of 100.

(c) In this case, we have $p = 0.75$ and $n = 100$. The sample proportions will be centered at the population proportion of $p = 0.75$ so will have a mean of 0.75. The standard deviation of the sample proportions is the standard error, which is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75(1-0.75)}{100}} = 0.043$$

(d) In this case, we have $p = 0.75$ and $n = 1000$. The sample proportions will be centered at the population proportion of $p = 0.75$ so will have a mean of 0.75. The standard deviation of the sample proportions is the standard error, which is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75(1-0.75)}{1000}} = 0.014$$

Notice that the standard error is significantly less with a sample size of 1000 than it is with a sample size of 100.

6.12 We compute the standard errors using the formula:

$$\begin{aligned}
p = 0.8: \quad & SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.8(0.2)}{100}} = 0.040 \\
p = 0.5: \quad & SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(0.5)}{100}} = 0.050 \\
p = 0.3: \quad & SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(0.7)}{100}} = 0.046 \\
p = 0.1: \quad & SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.1(0.9)}{100}} = 0.030
\end{aligned}$$

The largest standard error is at a population proportion of 0.5 (which represents a population split 50-50 between being in the category we are interested in and not being in). The farther we get from this 50-50 proportion, the smaller the standard error is. Of the four we computed, the smallest standard error is at a population proportion of 0.1.

6.14 In each case, we determine whether $np \geq 10$ and $n(1-p) \geq 10$.

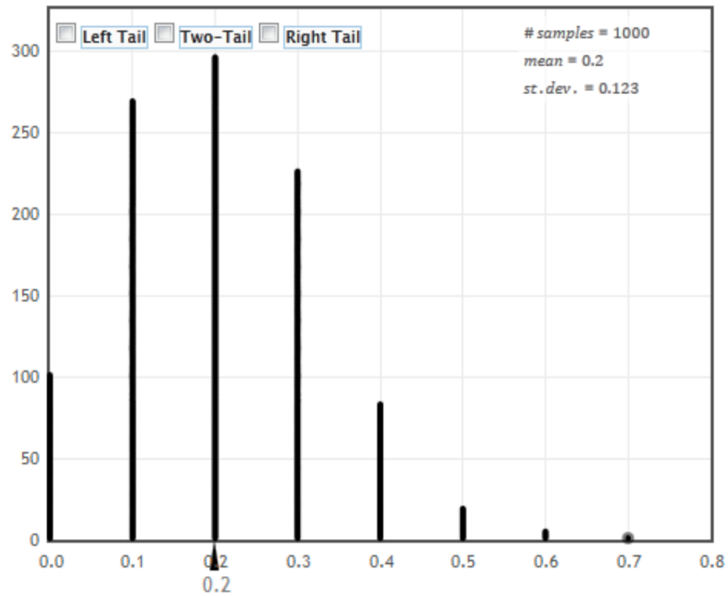
(a) No, the conditions do not apply, since $np = 80(0.1) = 8 < 10$.

(b) No, the conditions do not apply, since $np = 25(0.8) = 20$ but $n(1-p) = 25(1-0.8) = 5 < 10$.

(c) Yes, the conditions apply, since $np = 50(0.4) = 20$ and $n(1-p) = 50(1-0.4) = 30$.

(d) Yes, the conditions apply, since $np = 200(0.7) = 140$ and $n(1-p) = 200(1-0.7) = 60$.

6.18 (a) The graph below shows the sample proportions for 1000 samples of size 10 which were simulated from a population where $p = 0.2$. The distribution is not symmetric since the left tail is truncated at zero. This is not surprising since $np = 10 \cdot 0.2 = 2$ which is well less than the desired value of 10 that is needed in the Central Limit Theorem. The mean of these simulated sample proportions is 0.20 and the standard error is 0.123. Answers will vary with different simulations but should be approximately 0.20 and 0.126, respectively.



(b) The Central Limit Theorem for proportions says the expected mean of the sample proportions is p, which is 0.2, and the expected standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{10}} = 0.126$$

These are both very close to those obtained with simulations. Note that even though the sample size is not large enough to give a normal distribution, the mean and standard error of the sample proportions still agree with the Central Limit Theorem.

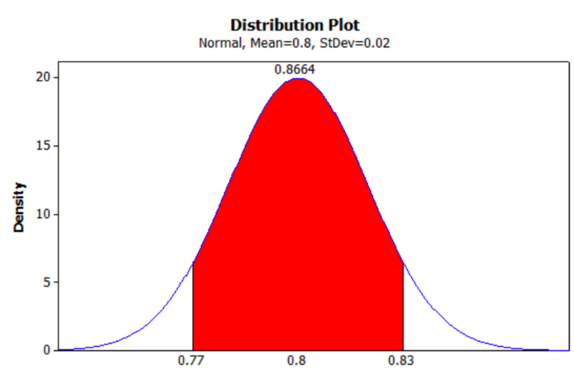
6.22 Using *StatKey* or other technology to create a bootstrap distribution, we see for one set of 1000 simulations that $SE = 0.014$. (Answers may vary slightly with other simulations.) Using the formula from the Central Limit Theorem, and using $\hat{p} = 0.753$ as an estimate for p, we have

$$SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{0.753(1-0.753)}{1000}} = 0.014$$

We see that the bootstrap standard error and the formula match very closely.

6.24 In the example about on time arrivals in the text, we see that the standard error for the proportion in this case (with $n = 400$ and $p = 0.8$) is $SE = \sqrt{0.8(1-0.8)/400} = 0.020$ and we see that it is appropriate to use a normal distribution.

To find the chance that an ontime proportion for a sample of 400 flights is within 0.03 of the true $p = 0.80$, we find the area between $0.80 - 0.03 = 0.77$ and $0.80 + 0.03 = 0.83$ for a $N(0.80, 0.02)$ curve. This area, 0.8664, is shown in the figure below. About 87% of samples will produce an ontime proportion that is at least this close (within ± 0.03) to the ontime proportion for all flights.



Note that we could also convert the endpoints to z-scores

$$z_1 = \frac{0.77 - 0.80}{0.02} = -1.5 \qquad \text{and} \qquad z_2 = \frac{0.83 - 0.80}{0.02} = 1.5$$

and find the area between -1.5 and 1.5 for a standard N(0,1) density.

6.34 The sample size is definitely large enough to use the normal distribution. For a confidence interval using the normal distribution, we use

$$\text{Sample statistic} \pm z^* \cdot SE.$$

The relevant sample statistic for a confidence interval for a proportion is $\hat{p} = 0.20$. For a 99% confidence interval, we have $z^* = 2.576$, and the standard error is $SE = \sqrt{\hat{p}(1-\hat{p})/n}$. The confidence interval is

$$\begin{array}{rcl} \hat{p} & \pm & z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \\ 0.20 & \pm & 2.576 \cdot \sqrt{\dfrac{0.20(0.80)}{1000}} \\ 0.20 & \pm & 0.033 \\ 0.167 & \text{to} & 0.233 \end{array}$$

We are 99% confident that the proportion of US adults who say they never exercise is between 0.167 and 0.233. The margin of error is $\pm 3.3\%$.

6.36 The sample size is definitely large enough to use the normal distribution. For a confidence interval using the normal distribution, we use

$$\text{Sample statistic} \pm z^* \cdot SE.$$

The relevant sample statistic for a confidence interval for a proportion is $\hat{p} = 0.83$. For a 95% confidence interval, we have $z^* = 1.96$, and the standard error is $SE = \sqrt{\hat{p}(1-\hat{p})/n}$. The confidence interval is

$$\begin{array}{rcl} \hat{p} & \pm & z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \\ 0.83 & \pm & 1.96 \cdot \sqrt{\dfrac{0.83(0.17)}{1000}} \\ 0.83 & \pm & 0.023 \\ 0.807 & \text{to} & 0.853 \end{array}$$

We are 95% confident that the proportion of adults who believe that children spend too much time on electronic devices is between 0.807 and 0.853. The margin of error is 0.023. Since the lowest plausible value for p in the confidence interval is 0.807, it is *not* plausible that the proportion is less than 80%. Since 0.85 is within the plausible range in the confidence interval, it is plausible that the proportion is greater than 85%.

6.40 (a) In each case, we use

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

for the confidence interval, using $z^* = 1.96$ for a 95% confidence interval. In every case, $n = 970$.

- For percent updating their status, we have:

$$0.15 \pm 1.96 \cdot \sqrt{\frac{0.15(0.85)}{970}} = 0.15 \pm 0.022 = \text{the interval from } 0.128 \text{ to } 0.172.$$

  We are 95% confident that between 12.8% and 17.2% of Facebook users update their status in an average day.

- For percent commenting on another's post, we have

$$0.22 \pm 1.96 \cdot \sqrt{\frac{0.22(0.78)}{970}} = 0.22 \pm 0.026 = \text{the interval from } 0.194 \text{ to } 0.246.$$

We are 95% confident that between 19.4% and 24.6% of Facebook users comment on another's post in an average day.

- For percent commenting on another's photo, we have

$$0.20 \pm 1.96 \cdot \sqrt{\frac{0.20(0.80)}{970}} = 0.20 \pm 0.025 = \text{the interval from } 0.175 \text{ to } 0.225.$$

We are 95% confident that between 17.5% and 22.5% of Facebook users comment on another's photo in an average day.

- For percent "liking" another's content, we have

$$0.26 \pm 1.96 \cdot \sqrt{\frac{0.26(0.74)}{970}} = 0.26 \pm 0.028 = \text{the interval from } 0.232 \text{ to } 0.288.$$

We are 95% confident that between 23.2% and 28.8% of Facebook users "like" another's content in an average day.

- For percent sending another user a private message, we have

$$0.10 \pm 1.96 \cdot \sqrt{\frac{0.10(0.90)}{970}} = 0.10 \pm 0.019 = \text{the interval from } 0.081 \text{ to } 0.119.$$

We are 95% confident that between 8.1% and 11.9% of Facebook users send another user a private message in an average day.

(b) The plausible proportions for those commenting on another's content are those between 0.194 and 0.246, while the plausible proportions for those updating their status are those between 0.128 to 0.172. Since these ranges do not overlap, we can be relatively confident that these proportions are not the same. A greater percentage comment on another's content than update their own status.

6.62 (a) If we let p_l be the proportion of left-handed lawyers, then we are testing test $H_0 : p_l = 0.10$ vs. $H_a : p_l \neq 0.10$.

(b) The sample proportion is $\hat{p}_l = 16/105 = 0.1524$. The test statistic is

$$z = \frac{0.1524 - 0.10}{\sqrt{\frac{0.10(1-.10)}{105}}} = 1.79.$$

The area above 1.79 in the standard normal curve is 0.037, so the p-value is $2(0.037) = 0.074$.

(c) We do not reject H_0 at the 5% significance level, and thus do not conclude that the proportion of left-handed lawyers differs from the proportion of left-handed Americans. At the 10% significance level we do reject H_0 and conclude that there is a higher percentage of left-handed lawyers.

6.64 We are conducting a hypothesis test for a proportion p, where p is the proportion of all MLB games won by the home team. We are testing to see if there is evidence that $p > 0.5$, so we have

$$H_0 : \quad p = 0.5$$
$$H_a : \quad p > 0.5$$

This is a one-tail test since we are specifically testing to see if the proportion is greater than 0.5. The test statistic is:

$$z = \frac{\text{Sample statistic} - \text{Null parameter}}{SE} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.549 - 0.5}{\sqrt{\frac{0.5(0.5)}{2430}}} = 4.83.$$

Using the normal distribution, we find a p-value of (to five decimal places) zero. This provides very strong evidence to reject H_0 and conclude that the home team wins more than half the games played. The home field advantage is real!

6.66 The sample proportion of questions having B as the correct answer is $\hat{p} = 90/400 = 0.225$. If all the choices were equally likely, we would expect each to be correct about 1/5, or 20%, of the time. If p represents the proportion of time B is the correct choice on all AP multiple choice questions, the hypotheses are:

$$H_0: \quad p = 0.20$$
$$H_a: \quad p > 0.20$$

The test statistic is:

$$z = \frac{\text{Sample statistic} - \text{Null parameter}}{SE} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.225 - 0.20}{\sqrt{\frac{0.2(0.8)}{400}}} = 1.25.$$

This is an upper-tail test, so the p-value is the area above 1.25 in a standard normal distribution. We find the p-value is 0.106. Even at a 10% level, we do not reject H_0. We do not find evidence that B is more likely to be the correct choice.

6.68 The proportion under the null hypothesis is p_0=0.5 and the sample size is $n = 25$. To see if the sample size is large enough for the proportion CLT to apply, we check that $np_0 = 25 \cdot 0.5 = 12.5 \geq 10$ and $n(1 - p_0) = 25(1 - 0.5) = 12.5 \geq 10$. Since both conditions are satisfied (under H_0 we expect at least 10 correct and 10 incorrect matches), we may use the test based on a normal distribution.
The proportion from the sample is $\hat{p} = 16/25 = 0.64$ and standardized test statistic is

$$z = \frac{0.64 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{25}}} = 1.40.$$

Since this is an upper tail alternative, $H_a : p > 0.5$, the p-value is the area under a standard normal curve in the tail above 1.40. Using technology or a table we find this area is 0.081.
While a p-value of 0.081 gives some evidence that the number of matches in the experiment is slightly unusual, it is not small enough, for example at a 5% significance level, to conclude that the null hypothesis of no dog/owner resemblance should be rejected. In other words, we did not find enough evidence to conclude that dogs resemble their owners. We reached a similar conclusion based on an estimated p-value of 0.1145 from the randomization distribution in the example in Chapter 4.