

where  $n$  is the number of observations. If  $j$  is an integer then the first quartile is the average of  $Y_{(j)}$  and  $Y_{(j+1)}$ :

$$\text{First quartile} = (Y_{(j)} + Y_{(j+1)})/2,$$

where  $Y_{(j)}$  is the  $j$ th sorted observation. For the sorted spider data,

$$j = (0.25)(16) = 4,$$

which is an integer. Therefore, the first quartile is the average of  $Y_{(4)}$  and  $Y_{(5)}$ :

$$\text{First quartile} = (2.31 + 2.37)/2 = 2.34.$$

If  $j$  is not an integer, then convert  $j$  to an integer by replacing it with the next integer that exceeds it (i.e., round  $j$  up to the nearest integer). The first quartile is then

$$\text{First quartile} = Y_{(j)},$$

where  $j$  is now the integer you rounded to.

The third quartile is computed similarly. Calculate

$$k = 0.75n.$$

If  $k$  is an integer, then the third quartile is the average of  $Y_{(k)}$  and  $Y_{(k+1)}$ :

$$\text{Third quartile} = (Y_{(k)} + Y_{(k+1)})/2,$$

where  $Y_{(k)}$  is the  $k$ th sorted observation. For the sorted spider data,

$$\boxed{k} \longrightarrow j = (0.75)(16) = 12,$$

which is an integer. Therefore, the third quartile is the average of  $Y_{(12)}$  and  $Y_{(13)}$ :

$$\text{Third quartile} = (3.00 + 3.09)/2 = 3.045.$$

If  $k$  is not an integer, then convert  $k$  to an integer by replacing it with the next integer that exceeds it (i.e., round  $k$  up to the nearest integer). The third quartile is then

$$\text{Third quartile} = Y_{(k)},$$

where  $k$  is the integer you rounded to.

The interquartile range is then

$$\text{Interquartile range} = 3.045 - 2.34 = 0.705 \text{ cm/s.}$$

The *interquartile range* is the difference between the third and first quartiles of the data. It is the span of the middle 50% of the data.

## The box plot

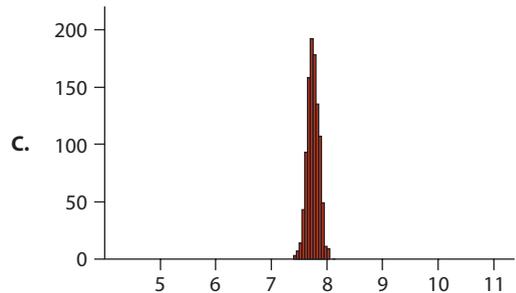
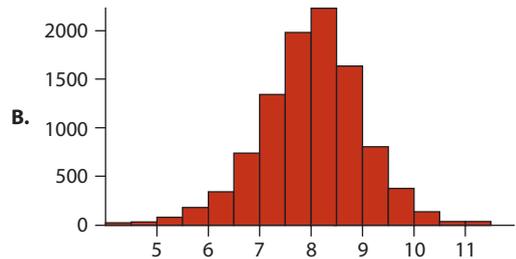
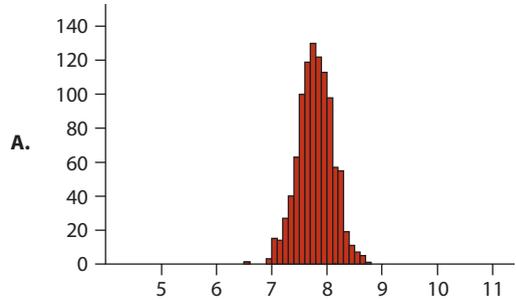
A **box plot** displays the median and interquartile range, along with other quantities of the frequency distribution. We introduced the box plot in Chapter 2. Figure 3.2-2

that when it flowers, it gives off a “powerful aroma of rotting fish and faeces” (Baeth 1996). The flowers smell this way because their principal pollinators are carrion beetles, who are attracted to such a smell. Baeth (1996) observed the number of carrion beetles (*Phaeochrous amplus*) that arrive per night to flowers of this species. The data are as follows:

51, 45, 61, 76, 11, 117, 7, 132, 52, 149

- a. What is the mean and standard deviation of beetles per flower?
- b. What is the standard error of this estimate of the mean?
- c. Give an approximate 95% confidence interval of the mean. Provide lower and upper limits.
- d. If you had been given 25 data points instead of 10, would you expect the mean to be greater, less than, or about the same as the mean of this sample?
- e. If you had been given 25 data points instead of 10, would you have expected the standard deviation to be greater, less than, or about the same as this sample?
- f. If you had been given 25 data points instead of 10, would you have expected the standard error of the mean to be greater, less than, or about the same as this sample?

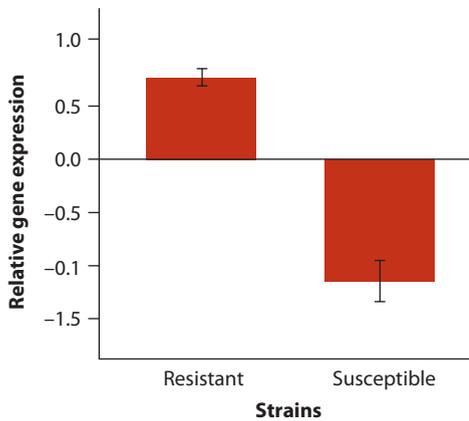
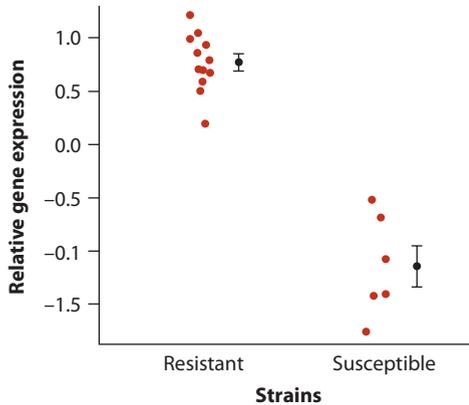
19. The following three histograms (A, B, and C) plot information about the number of hours of sleep adult Europeans get per night (Roenneberg 2012). One of them shows the frequency distribution of individual values in a random sample. Another shows the distribution of sample means for samples of size 10 taken from the same population. Another shows the distribution of sample means for samples of size 100.



- a. Identify which graph goes with which distribution.
- b. What features of these distributions allowed you to distinguish which was which?
- c. Estimate by eye the approximate population mean of the number of hours of sleep using the distribution for the data.
- d. Estimate by eye the approximate mean of the distributions of sample means.

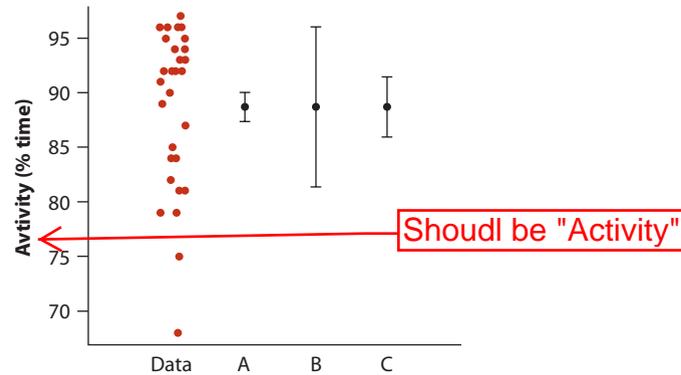
Beath (2x)

20. The following figure shows two alternative ways of presenting means with standard error bars in a graph. The data are from Daborn et al. (2002), who showed that elevated expression of the gene *Cyp6g1* in *Drosophila* causes resistance to DDT. Expression levels of the gene (relative to a standard) were measured in 12 resistant strains of *Drosophila* and 6 susceptible strains. Which graphical method is superior? Explain.



21. Is sleep necessary? To investigate, Lesku et al. (2012) measured the activity patterns of breeding pectoral sandpipers (*Calidris melanotos*) in the high Arctic in summer, when the sun never sets. The accompanying figure shows the observed percent time that individual males were awake and active in a 2008 field study. The data are on the left. To the right of the data are the sample mean (filled circle) and error bars for the standard deviation, the standard error of

the mean, and a 95% confidence interval for the mean (in no particular order).



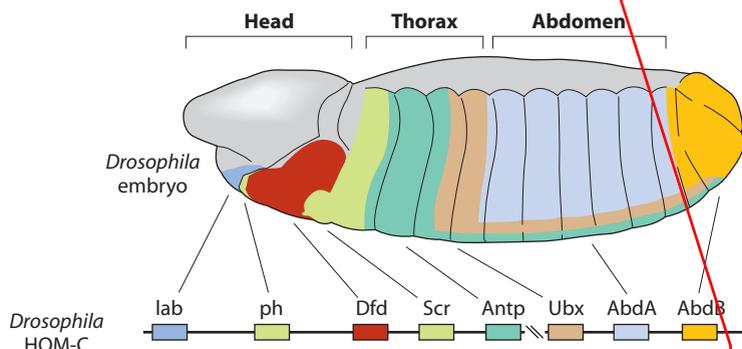
- Which of the error bars indicates the standard deviation?
  - Which error bar indicates the standard error of the mean?
  - Which error bar indicates a 95% confidence interval for the mean?
  - Estimate by eye the smallest plausible value for the mean activity (% time) of male pectoral sandpipers. Using this smallest plausible value, calculate approximately the maximum plausible number of hours (out of 24 hours) that males spend inactive or asleep.
22. How long do you hug somebody? Nagy (2011) measured the duration of spontaneous embraces at the 2008 Summer Olympic Games in Beijing, China. The data are the durations of hugs, in seconds, of athletes immediately after competing in the finals of an event. Hugs were either with their coach, a supporter (e.g., a team member), or a competitor. Descriptive statistics calculated from the data are in the following table. *n* refers to the sample size.

Relationship	Mean	Standard deviation	<i>n</i>
Coach	3.77	3.96	77
Supporter	3.16	2.76	75
Competitor	1.81	1.13	33

- According to the values in the table, which relationship group gets the longest hugs, on average, and which gets the briefest hugs?

- a. Draw a probability tree that describes all possible outcomes of such a survey and their probabilities.
- b. What is the overall probability that a randomly sampled respondent answers yes?
23. Imagine that a long stretch of single-stranded DNA has 30% adenine, 25% thiamine, 15% cytosine, and 30% guanine. (These make up the nucleotides of the DNA.) What is the probability of randomly drawing 10 adenines in a row in a sample of 10 randomly chosen nucleotides?
24. The *Hox* genes are responsible for determining the anterior–posterior identity of body regions (segments) in the developing insect embryo. Different *Hox* genes are turned on (expressed) in different segments of the body, and in this way they determine which segments become head and which thorax, which develop legs and which antennae. One surprising thing about the *Hox* genes is that they usually occur in a row on the same chromosome and in the same order as the body regions that they control. For example, the fruit fly *Drosophila melanogaster* has eight *Hox* genes located on a chromosome in exactly the same order as the body regions in which they are expressed, from head to tail (see figure below; Lewis et al. 1978; Negre et al. 2005). If the eight genes were thrown randomly onto the same chromosome, what is the probability that they would line up in the same order as the segments in which they are expressed?
25. The flour beetle, *Tribolium castaneum*, has 10 chromosomes, roughly equal in size, and it also has eight *Hox* genes (Brown et al. 2002; see Assignment Problem 24). If the eight *Hox* genes were randomly distributed throughout the genome of the beetle, what is the probability that all eight would land on the same chromosome?
26. A seed randomly blows around a complex habitat. It may land on any of three different soil types: a high-quality soil that gives a 0.8 chance of seed survival, a medium-quality soil that gives a 0.3 chance of survival, and a low-quality soil that gives only a 0.1 chance of survival. These three soil types (high, medium, and low) are present in the habitat in proportions of 30:20:50, respectively. The probability that a seed lands on a particular soil type is proportional to the frequency of that type in the habitat.
- a. Draw a probability tree to determine the probabilities of survival under all possible circumstances.
- b. What is the probability of survival of the seed, assuming that it lands?
- c. Assume that the seed has a 0.2 chance of dying before it lands in a habitat. What is its overall probability of survival?
27. A flycatcher is trying to catch passing bugs. The probability that it catches a bug on any given try is 20%.
- a. What is the probability that it catches its first bug on its fourth try?
- b. ~~What is the probability that it catches its first bug after at least four failures, assuming that it keeps trying until it is successful?~~

FIGURE FOR PROBLEM 24



**Change to:** What is the probability that at least four failures occur before the flycatcher has its first success?

The observed number of left-handed flowers (6) is less than the expected number from the null hypothesis (6.75), so we begin the calculation of the  $P$ -value by determining the probability of obtaining *six or fewer* left-handed offspring, assuming that the null distribution is true:

$$\Pr[\text{number of left-handed flowers} \leq 6] = \Pr[6] + \Pr[5] + \cdots + \Pr[0].$$

Summing these probabilities (given here by the height of the bars in Figure 6.4-1) yields

$$\Pr[\text{number of left-handed flowers} \leq 6] = 0.471.$$

This sum yields only the probability under the left tail of the null distribution (Figure 6.4-1). Because the test is two-sided, we also need to account for outcomes at the other tail of the distribution that are as unusual as or more unusual than the outcome observed. The most straightforward method to obtain  $P$  is to multiply the above sum by two (Yates 1984), yielding

$$\begin{aligned} P &= 2 \Pr[\text{number of left-handed flowers} \leq 6] \\ &= 2 \times 0.471 = 0.942. \end{aligned}$$

The  $P$ -value is quite high, and there is a high probability of getting data like these when the null hypothesis is true. The  $P$ -value is not less than or equal to the conventional significance level  $\alpha = 0.05$ , so our conclusion is that the null hypothesis is *not rejected*.

## Interpreting a nonsignificant result

What does failure to reject  $H_0$  mean? Can we conclude that the null hypothesis is true? Sadly, we can't, because it is always possible that the true value of the proportion  $p$  differs from  $1/4$  by a small or even moderate amount. You can tell this from the 95% confidence interval for  $p$ , the proportion of left-handed flowers in the population of possible offspring from the cross. Using methods we will detail in Chapter 7, we calculated this interval to be:

$$\boxed{0.10 \quad 0.41}$$

$$\cancel{0.11} < p < \cancel{0.42}.$$

This result indicates that  $1/4$  falls between the lower and upper limits for  $p$ , and so it is certainly among the most-plausible values for this proportion. However, many other possible values for the proportion also fall within this most-plausible range. In other words, it's possible that  $p$  is  $1/4$ , but it is also possible that  $p$  differs from  $1/4$  even though  $H_0$  was not rejected, because the power of the test was limited by the relatively small sample size of 27.

How, then, do we interpret the result? A valid interpretation is that the data are *compatible* or *consistent* with the null hypothesis; in other words, the data are compatible with the simple genetic model of inheritance of handedness in the mud plantain. There is no need or justification to build more complex genetic models of inheritance for flower handedness. A time may come when a new study—one with a

The formula to calculate  $\chi^2$ , first introduced in Section 8.2, gives us

$$\chi^2 = \frac{(530 - 585.3)^2}{585.3} + \frac{(1332 - 1221.4)^2}{1221.4} + \frac{(582 - 637.3)^2}{637.3} = 20.04.$$

Next, we need to calculate the number of degrees of freedom for our test. There are three categories, which would ordinarily leave us with two degrees of freedom. However, we needed to estimate one parameter from the data to generate the expected frequencies (the probability of boys,  $p$ ). Using this estimate costs us an extra degree of freedom.<sup>8</sup> As a result, the number of degrees of freedom is

$$df = 3 - 1 - 1 = 1.$$

The critical value for the  $\chi^2_1$  distribution having one degree of freedom and a significance level  $\alpha = 0.05$  is 3.84 (see Statistical Table A). Because 20.04 is further into the tail of the distribution than 3.84, we know that  $P < 0.05$ ; therefore, we reject the null hypothesis. If we probe Statistical Table A further, we find that 20.04 is greater even than the critical value corresponding to  $\alpha = 0.001$ , so  $P < 0.001$ . A statistics package on the computer gave us a more exact value,  $P = 1.2 \times 10^{-7}$ .  $7.7 \times 10^{-6}$

These data show that the frequency distribution of the number of boys and girls in two-child families does *not* match the binomial distribution. This means that one of the assumptions of the binomial distribution is not met in these data. Either the probability of having a son varies from one family to the next, or the individuals within a family are not independent of each other, or both.

What is the reason for the poor fit of the binomial distribution to the number of boys in two-child families? Is the sex of the second child not independent of that of the first, as we've assumed? Are parents manipulating the sex of their children? One likely explanation is that many parents of two-child families having either no boys or two boys are unsatisfied with not having at least one child of each sex and decide to have a third child, thus "removing" their family from the set of two-child families.

## 8.6 Random in space or time: the Poisson distribution

When the dust settled after the 1980 explosion of Mount St. Helens, spiders were among the first organisms to recolonize the moonscape-like terrain. They dropped out of the airstream and grew fat on insects that arrived in the same way. Let's imagine the frequency distribution of spider landings across the landscape. What would it

8. Why do we lose another degree of freedom? Using the data to estimate a parameter of the probability model of the null hypothesis (here, the binomial distribution) unfairly improves the fit of the model to the data. We compensate for this by removing one degree of freedom for the test statistic for every parameter we estimate. If we estimate too many parameters, then the number of degrees of freedom would drop to zero and we couldn't do the test.

Using the standard formula for the  $\chi^2$  statistic, we compute

$$\chi^2 = \frac{(13 - 5.88)^2}{5.88} + \frac{(15 - 10.00)^2}{10.00} + \frac{(16 - 14.03)^2}{14.03} + \dots = 23.93.$$

We have six degrees of freedom for this test, accounting for the one parameter,  $\mu$ , that we had to estimate from the data:

$$\begin{aligned} df &= (\text{Number of categories}) - 1 - (\text{Number of parameters estimated from the data}) \\ &= 8 - 1 - 1 = 6. \end{aligned}$$

The critical value for  $\chi_{0.05,6}^2$  is 12.59 (see Statistical Table A). Our  $\chi^2$  statistic of 23.93 is further into the tail of the distribution than this critical value is, so our  $P$ -value is less than 0.05. More specifically, we can see that  $P < 0.001$ , because 23.93 is also greater than 22.46, the critical value corresponding to  $\alpha = 0.001$ . We reject the null hypothesis, therefore, and conclude that extinctions in this fossil record do *not* fit a Poisson distribution.

### Comparing the variance to the mean

How can we describe the way that a pattern deviates from the Poisson distribution? One unusual property of the Poisson distribution is that the variance in the number of successes per block of time (the square of the standard deviation) is equal to the mean ( $\mu$ ). In an observed frequency distribution, if the variance is greater than the mean, then the distribution is clumped. There are more blocks with many successes, and more with few successes, than expected from the Poisson distribution. If the variance is less than the mean, then the distribution is dispersed. The ratio of the variance to the mean number of successes is therefore a measure of clumping or dispersion.

For the extinction data, the sample mean number of extinctions is 4.21. The sample variance in the number of extinctions is

$$s^2 = \frac{(0 - 4.21)^2(0) + (1 - 4.21)^2(13) + (2 - 4.21)^2(15) + \dots}{76 - 1} = 13.72.$$

Because the sample variance (13.72) greatly exceeds the sample mean (4.21), the ratio of the variance to the mean (~~3.56~~) is greater than 1, and we conclude that the distribution of extinction events in time is highly clumped. That is, extinctions tend to occur in bursts (mass extinctions) rather than randomly or evenly in time.

3.26

## 8.7 Summary

- The  $\chi^2$  goodness-of-fit test compares the frequency distribution of a discrete or categorical variable with the frequencies expected from a probability model.

to hospitals differs between weekend and weekday. State the advantages and disadvantages of both.

- c. State the null and alternative hypotheses for such a test.
  - d. Test the hypotheses. State your conclusions clearly.
17. Truffles are a great delicacy, sending thousands of mushroom hunters into the forest each fall to find them. A set of plots of equal size in an old-growth forest in Northern California was surveyed to count the number of truffles (Waters et al. 1997). The resulting distribution is presented in the following table. Are truffles randomly located around the forest? If not, are they clumped or dispersed? How can you tell? (The mean number of truffles per plot, calculated from these data, is 0.60.)

Number of truffles per plot	Frequency
0	203
1	39
2	18
3	13
> 3	15

18. The anemonefish *Amphiprion akallopisos* lives in small groups that defend territories consisting of a cluster of sea anemones, among the tentacles of which the anemonefish live (think *Nemo*). In a field study of the species at Aldabra Atoll in the Indian Ocean, Fricke (1979) noticed that in a territory tended to have several males but just one female. Based on his counts, 20 territories of exactly four adult fish would have the following frequency distribution of female numbers.

males

females

Number of males	Number of females	Number of territories
0	4	0
1	3	20
> 2	< 3	0
Total		20

1

- a. Estimate the mean number of females per territory having four fish. Provide a standard deviation for this estimate.

- b. Does the number of females in territories having four fish have a binomial distribution? Show all steps in carrying out your test.
- c. If the number of females in territories does not have a binomial distribution, what is the likely statistical explanation (i.e., what assumption of the binomial distribution is likely violated)?
- d. Optional: Can you suggest a biological explanation for a non-binomial pattern?

19. Hurricanes hit the United States often and hard, causing some loss of life and enormous economic costs. They are ranked in severity by the Saffir–Simpson scale, which ranges from Category 1 to Category 5, with 5 being the worst. In some years, as many as three hurricanes that rate a Category 3 or higher hit the U.S. coastline. In other years, no hurricane of this severity hits the United States. The following table lists the number of years that had 0, 1, 2, 3, or more hurricanes of at least Category 3 in severity, over the 100 years of the 20th century (Blake et al. 2005):

Number of hurricanes Category 3 or higher	Number of years
0	50
1	39
2	7
3	4
> 3	0

- a. What is the mean number of severe hurricanes to hit the United States per year?
  - b. What model would describe the distribution of hurricanes per year, if they were to hit independently of each other and if the probability of a hurricane were the same in every year?
  - c. Test the fit of the model from part (b) to the data.
20. In snapdragons, variation in flower color is determined by a single gene (Hartl and Jones 2005). *RR* individuals are red, *Rr* (heterozygous) individuals are pink, and *rr* individuals are white. In a cross between heterozygous individuals, the expected ratio of red-flowered: pink-flowered: white-flowered offspring is 1:2:1.

The  $P$ -value for Fisher's exact test is the sum of the probabilities of all such extreme tables under the null hypothesis of independence, including equally or more extreme tables on the other tail of the distribution of tables. There are also confidence intervals for odds ratios based on small samples using the same logic as Fisher's exact test (see Agresti 2002).

We applied Fisher's exact test to the data in Table 9.5-1, using a statistical program on the computer, and we found that  $P < 0.0001$ . Thus, we can reject the null hypothesis of independence. Vampire bats evidently preferred the cows in estrus. The reasons for this are not clear.<sup>9</sup>

$10^{-10}$

## 9.6 G-tests

The  $G$ -test is another contingency test often seen in the literature. The  $G$ -test is almost the same as the  $\chi^2$  test except that the following test statistic is used:

$$G = 2 \sum_{row=1}^r \sum_{column=1}^c \text{Observed}(row, column) \times \ln \left[ \frac{\text{Observed}(row, column)}{\text{Expected}(row, column)} \right],$$

where "ln" refers to the natural logarithm. Under the null hypothesis of independence, the sampling distribution of the  $G$ -test statistic is approximately  $\chi^2$  with  $(r - 1)(c - 1)$  degrees of freedom.

For the data in Example 9.4 on fish infection and bird predation, for example,

$$\begin{aligned} G &= 2 \left( 1 \ln \left[ \frac{1}{17.0} \right] + 49 \ln \left[ \frac{49}{33.0} \right] + 10 \ln \left[ \frac{10}{15.3} \right] + 35 \ln \left[ \frac{35}{29.7} \right] \right. \\ &\quad \left. + 37 \ln \left[ \frac{37}{15.7} \right] + 9 \ln \left[ \frac{9}{30.3} \right] \right) \\ &= 71.45 \end{aligned}$$

71.45

This  $G$ -test statistic would be compared to the  $\chi^2$  distribution with

$$df = (r - 1)(c - 1) = 2.$$

Again, we would strongly reject the null hypothesis of independence.

The  $G$ -test is derived from principles of likelihood (see Chapter 20) and can be applied across a wider range of circumstances than the  $\chi^2$  contingency test. While the  $G$ -test is preferred by some statisticians, it has been shown to be less accurate for small sample sizes (Agresti 2002), and it is used somewhat less often than the  $\chi^2$  contingency test in the biological literature. The  $G$ -test has advantages, though, when analyzing complicated experimental designs involving multiple explanatory variables (see Sokal and Rohlf 1995 or Agresti 2002).

9. It has been speculated that, by drinking the blood of cows in heat, the bats minimize their intake of hormones, which are in higher concentration in the blood of non-estrous cows and may act as a birth control pill for bats, preventing reproduction.

26. The crab spider, *Thomisus spectabilis*, sits on flowers and preys upon visiting honeybees, as shown in the photo at the beginning of the chapter. (Remember this the next time you sniff a wild flower.) Do honeybees distinguish between flowers that have crab spiders and flowers that do not? To test this, Heiling et al. (2003) gave 33 bees a choice between two flowers: one had a crab spider and the other did not. In 24 of the 33 trials, the bees picked the flower that had the spider. In the remaining nine trials, the bees chose the spiderless flower. With

these data, carry out the appropriate hypothesis test, using an appropriate approximation to calculate  $P$ .

27. The following table lists the mean and standard deviation of several different normal distributions. In each case, a sample of 20 individuals was taken, as well as a sample of 50 individuals. For each sample, calculate the probability that the mean of the sample was less than the given value.

TABLE FOR PROBLEM 27

Mean	Standard deviation	Value	$n = 20$ : $\Pr[\bar{Y} > \text{value}]$	$n = 50$ : $\Pr[\bar{Y} > \text{value}]$
-5	5	-5.2		
10	30	8.0		
-55	20	-61.0		
12	3	12.5		



ent times or on different sides of the body or of the plot. In the two-sample design, each group constitutes an independent random sample of individuals. In both cases, we make use of the  $t$ -distribution to calculate confidence intervals for the mean difference and test hypotheses about means. However, the two different study designs require alternative methods for analysis. We elaborate on the reasons in Section 12.1.

## 12.1 Paired sample versus two independent samples

There are two study designs to measure and test differences between the means of two treatments. To describe them, let's use an example: does clear-cutting a forest affect the number of salamanders present? Here we have two treatments (clear-cutting/no clear-cutting), and we want to know if the mean of a numerical variable (the *number of salamanders*) differs between them. "Clear-cut" is the treatment of interest, and "no clear-cut" is the control. This is the same as asking whether these two variables, *treatment* (a numerical variable) and *salamander number* (a categorical variable), are associated.

We can design this kind of a study in either of two ways: a two-sample design (see the left panel in Figure 12.1-1) or a paired design (see the right panel in Figure 12.1-1). In the *two-sample* design, we take a random sample of forest plots from the population and then randomly assign either the clear-cut treatment or the no-clear-cut treatment to each plot. In this case, we end up with two independent samples, one from each treatment. The difference in the mean number of salamanders between the clear-cut and no-clear-cut areas estimates the effect of clear-cutting on salamander number.<sup>2</sup>

In the *paired* design, we take a random sample of forest plots and clear-cut a randomly chosen half of each plot, leaving the other half untouched. Afterward, we count the number of salamanders in each half. The mean difference between the two sides estimates the effect of clear-cutting.

Each of these two experimental designs has benefits, and both address the same question. However, they differ in an important way that affects how the data are analyzed. In the paired design, measurements on adjacent plot-halves are not inde-

2. We are assuming that you have persuaded an enlightened forest company to go along with this experiment. It is more likely that the forest company has already clear-cut some areas and not others, and you must compare the two treatments after the fact with an observational study. Both the experimental approach and the observational design yield a measure of the association between clear-cutting and salamander number, but only the experimental study can tell us whether clear-cutting is the *cause* of any difference (Chapter 14). The study design issues are the same, however: whether to choose forest plots randomly from each treatment (the two-sample design) or to randomly choose forest plots that straddle both treatments, making it possible to compare cut and uncut sites that are side by side (the paired design).

Now we can calculate the test statistic,  $t$ . According to the null hypothesis, the two means are equal; that is,  $\mu_1 - \mu_2 = 0$ . With this information, we can find

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{SE_{\bar{Y}_1 - \bar{Y}_2}} = \frac{2.29}{0.527} = 4.35.$$

We need to compare this test statistic with the distribution of possible values for  $t$  if the null hypothesis were true. The appropriate null distribution is the  $t$ -distribution, and it will have  $df_1 + df_2 = 153 + 29 = 182$  degrees of freedom. The  $P$ -value is then

$$P = 2 \Pr[t > 4.35].$$

Using a computer, we find that this  $t$ -value corresponds to

$$P = 0.000023.$$

Since  $P < 0.05$ , we reject the null hypothesis.

We reach the same conclusion using Statistical Table C. The critical value for  $\alpha = 0.05$  for a  $t$ -distribution with 182 degrees of freedom is

$$t_{0.05(2),182} = 1.97.$$

The  $t = 4.35$  calculated from these data is much further into the tail of the distribution than this critical value, so we reject the null hypothesis. In fact, the  $t$  calculated for these data is further in the tail of the distribution than all values given in Statistical Table C, including those for  $\alpha = 0.0002$ . From this information we may conclude that  $P < 0.0002$ . Based on these studies, there is a difference in the horn length of lizards eaten by shrikes, compared with live lizards. It is possible that shrikes avoid or are unable to capture the lizards with the longest horns. We can't be certain that this explains the difference in horn length, because this is an observational study. To infer causation, we would need to carry out a controlled experiment that manipulates lizard horn lengths.

## Assumptions

The two-sided  $t$ -test and two-sample confidence interval for a difference in means are based on the following assumptions:

- Each of the two samples is a random sample from its population.
- The numerical variable is normally distributed in each population.
- The standard deviation (and variance) of the numerical variable is the same in both populations.

We've heard the first two assumptions before—they are required for the one-sample  $t$ -test—but the third assumption is new. The two-sample methods we have just discussed are fairly robust to violations of this assumption. With moderate sample sizes (i.e.,  $n_1$  and  $n_2$  both greater than 30), the methods work well, even if the standard deviations in the two groups differ by as much as threefold, as long as the sample sizes of the two groups are approximately equal. If there is more than a threefold dif-

$$\text{Formula: } (\bar{Y}_1 - \bar{Y}_2) \pm \frac{t_{\alpha(2), df}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \text{ where } df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}\right]}$$

where  $df$  is rounded down to the nearest integer.

### Welch's approximate $t$ -test

**What is it for?** Tests whether the difference between the means of two populations is equal to a null hypothesized value when the standard deviations are unequal.

**What does it assume?** Both samples are random samples. The numerical variable is normally distributed within both populations.

**Test statistic:**  $t$

**Distribution under  $H_0$ :**  $t$ -distribution. The number of degrees of freedom are fewer than in the case of the two-sample  $t$ -test. See the previous entry on Welch's confidence interval for difference between two means for the formula for  $df$ .

$$\text{Formula: } t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

### $F$ -test

**What is it for?** Tests whether the variances of two populations are equal.

**What does it assume?** Both samples are random samples. The numerical variable is normally distributed within both populations.

**Test statistic:**  $F$

**Distribution under  $H_0$ :**  $F$ -distribution with  $n_1 - 1$ ,  $n_2 - 1$  degrees of freedom.

$H_0$  is rejected if  $F \geq F_{\alpha(2), n_1 - 1, n_2 - 1}$ . The quantity  $F_{\alpha(1), k - 1, N - k}$  is the critical value of the  $F$ -distribution corresponding to the pair of degrees of freedom ( $n_1 - 1$  and  $n_2 - 1$ ). Critical values of the  $F$ -distribution are provided in Statistical Table D.  $F$  is compared only to the upper critical value, because  $F$ , as computed below, always puts the larger sample variance in the top (the numerator) of the  $F$ -ratio.

$$\text{Formula: } F = \frac{s_1^2}{s_2^2}$$

where  $s_1^2$  is the larger sample variance and  $s_2^2$  is the smaller sample variance.

This should not be a ratio. The numerator (top part) should be MULTIPLIED BY the denominator (bottom part) not divided by it.

were not actually used on the next person.) The researchers measured how much hot sauce each man added, as a stand-in for aggression, because it correlates with the amount of pain inflicted on the next person. Do the two groups differ in the mean amount of hot sauce they add to the water? Here is a summary of the results:

Group	Sample size	Mean hot sauce added	SD hot sauce added
Gun handlers	15	13.61	8.35
Toy handlers	15	4.23	2.62

- a. Assuming that the amount of hot sauce added per person is normally distributed in each group, would an ordinary two-sample *t*-test be an appropriate test for this analysis?
  - b. If not, what would be an appropriate method to use?
29. Refer to Practice Problem 17. How big is the estimated difference between the means of the antibody and control treatments? Use a confidence interval to calculate the most-plausible range of values for the difference in mean percent.
30. *Spot the flaw.* There are two types of males in bluegill sunfish. Parental males guard small territories, where they mate with females and take care of the eggs and young. Cuckolder males do not have territories or take care of young. Instead, they sneak in and release sperm when a parental male is spawning with a female, thereby fertilizing a portion of the eggs. A choice experiment was carried out on juvenile sunfish to test whether offspring from the two types of males (fertilized by parental male vs. fertilized by cuckolder male) are able to discriminate between kin (siblings) from non-kin using odor cues. The researchers used a two-sample method to test the null hypothesis that fish are unable to discriminate between kin and non-kin. This null hypothesis was not rejected for offspring from parental males. However, the same null hypothesis was rejected for offspring from cuckolder males. The researchers concluded that offspring of cuckolder males are more able to discriminate

kin from non-kin than are offspring of parental males. What is wrong with this conclusion? What analysis should have been conducted?

31. Rutte and Taborsky (2007) tested for the existence of “generalized reciprocity” in the Norway rat, *Rattus norvegicus*. That is, they asked whether a rat that had just been helped by a second rat would be more likely itself to help a third rat than if it had not been helped. Focal female rats were trained to pull a stick attached to a tray that produced food for their partners but not for themselves. Subsequently, each focal rat’s experience was manipulated in two treatments. Under one treatment, the rat was helped by three unfamiliar rats (who pulled the appropriate stick). Under the other treatment, focal rats received no help from three unfamiliar rats (who did not pull the stick). Each focal rat was exposed to both treatments in random order. Afterward, each focal rat’s tendency to pull for an unfamiliar partner rat was measured. The number of pulls in a given period (in pulls/min.) by 19 focal female rats after both treatments is given below.

Focal rat	After receiving help	After receiving no help
10	0.43	0.29
11	0.86	0.14
12	0.57	0.29
20	0.86	0.57
30	0.29	0.29
31	1.14	0.71
32	0.57	0.29
33	0.86	0.86
34	1.43	0.86
40	0.86	0.86
41	0.57	0.43
42	0.86	1.00
43	0.00	0.86
50	1.00	0.43
51	0.86	1.00
52	0.86	0.00
60	1.86	1.14
61	0.86	0.71
62	0.86	0.71

one-

/ 62

- a. Draw a graph to illustrate the data. What trend is evident?

The best estimate of the slope is

$$b = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})} = \frac{167.5548}{5088.8447} = 0.03293.$$

should be a  
superscript 2 here

Calculating the intercept as before, and rounding, we get the least-squares regression line as

$$Y = 1.20 + 0.033X,$$

which can also be written as

$$\text{Log stability} = 1.20 + 0.033(\text{number of species}).$$

This line has a positive slope, as shown in Figure 17.3-1. The estimate of slope ( $b = 0.033$ ) indicates that log stability of biomass production rises by the amount 0.033 for every species added to plots.

To calculate the standard error of the slope, we need the mean square residual:

$$\begin{aligned} \text{MS}_{\text{residual}} &= \frac{\sum_i (Y_i - \bar{Y})^2 - b \sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{n - 2} \\ &= \frac{24.8149 - 0.03293(167.5548)}{161 - 2} \\ &= 0.12137. \end{aligned}$$

Thus, the standard error of  $b$  is

$$\text{SE}_b = \sqrt{\frac{\text{MS}_{\text{residual}}}{\sum_i (X_i - \bar{X})^2}} = \sqrt{\frac{0.12137}{5088.8447}} = 0.004884.$$

We now have all of the elements needed to calculate the  $t$ -statistic:

$$t = \frac{b - \beta_0}{\text{SE}_b} = \frac{0.03293 - 0}{0.004884} = 6.74.$$

We must compare this  $t$ -statistic with the  $t$ -distribution having  $df = n - 2 = 161 - 2 = 159$  degrees of freedom. Using a computer, we find that  $t = 6.74$  corresponds to  $P = 2.7 \times 10^{-10}$ , so we reject the null hypothesis. We reach the same conclusion if we use the critical value for the  $t$ -distribution with  $df = 159$  (Statistical Table C):

$$t_{0.05(2),159} = 1.97.$$

Since  $t = 6.74$  is greater than 1.97,  $P < 0.05$ , and we reject  $H_0$ . In other words, increasing the number of plant species in plots increases the stability of plant biomass production of the ecosystem. A 95% confidence interval for the population slope, calculated from the formula in Section 17.1, is

$$0.0233 < \beta < 0.0426.$$

indicating that the estimate of slope has fairly tight bounds.

previously under “Regression slope”), and  $t_{\alpha(2), n-2}$  is the two-tailed critical value of the  $t$ -distribution having  $df = n - 2$ .

### The $t$ -test of a regression slope

**What is it for?** To test the null hypothesis that the population parameter  $\beta$  equals a null hypothesized value  $\beta_0$ .

**What does it assume?** Same as the assumptions for the regression slope.

**Test statistic:**  $t$

**Distribution under  $H_0$ :**  $t$ -distributed with  $n - 2$  degrees of freedom.

**Formula:**  $t = \frac{b - \beta_0}{SE_b}$ , where  $SE_b$  is the standard error of  $b$  (see the formula given previously under “Regression slope”).

### The ANOVA method for testing zero slope

**What is it for?** To test the null hypothesis that the slope  $\beta$  equals zero, and to partition sources of variation.

**What does it assume?** Same as the assumptions for the regression slope.

**Test statistic:**  $F$

**Distribution under  $H_0$ :**  $F$  distribution.  $F$  is compared with  $F_{\alpha(1), 1, n-2}$ .

Source of variation	Sum of squares	$df$	Mean squares	$F$ -ratio
Regression	$SS_{\text{regression}} = \sum_i (\hat{Y}_i - \bar{Y})^2$	1	$\frac{SS_{\text{regression}}}{df_{\text{regression}}}$	$\frac{MS_{\text{regression}}}{MS_{\text{residual}}}$
Residual	$SS_{\text{residual}} = \sum_i (Y_i - \hat{Y}_i)^2$	$n - 2$	$\frac{SS_{\text{residual}}}{df_{\text{residual}}}$	
Total	$SS_{\text{total}} = \sum_i (Y_i - \bar{Y})^2$	$n - 1$		

There should be an over bar over this Y

### $R$ squared ( $R^2$ )

**What is it for?** Measuring the fraction of the variation in  $Y$  that is “explained” by  $X$ .

**Formula:**  $R^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}}$ ,

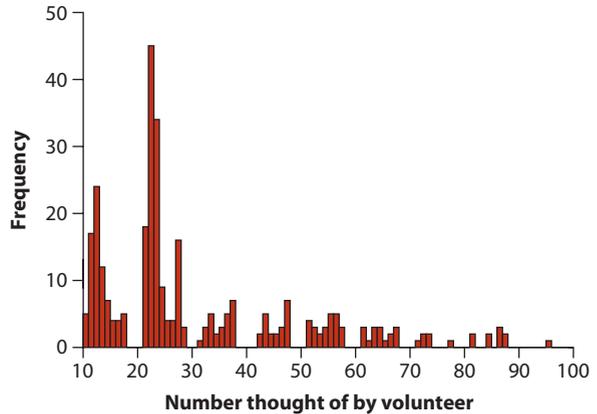
where  $SS_{\text{regression}}$  is the sum of squares for regression and  $SS_{\text{total}}$  is the total sum of squares.

every member of the audience to think of a two-digit number. After a show of pretending to read their minds, the performer states a number that a surprisingly large fraction of the audience was thinking of. This feat would be surprising if the people thought of all two-digit numbers with equal probability, but not if people everywhere tend to pick the same few numbers. Figure 19.1-1 shows the numbers chosen independently by ~~350~~ volunteers (Marks 2000). Are all two-digit numbers selected with equal probability?

/ 315

**FIGURE 19.1-1**

The distribution of two-digit numbers chosen by volunteers.



According to the histogram in Figure 19.1-1, the two-digit numbers chosen don't seem to occur with equal probability at all. A large number of people chose numbers in the teens and twenties, with few choosing larger numbers. Almost nobody chose multiples of ten. Let's use these data to test the null hypothesis that every two-digit number occurs with equal probability. The hypotheses are as follows.

$H_0$ : Two-digit numbers are chosen with equal probability.

$H_A$ : Two-digit numbers are not chosen with equal probability.

To analyze this problem, our first thought might be to use a  $\chi^2$  goodness-of-fit test (Chapter 8). There are 90 categories of outcome (each of the 90 integers between 10 and 99). The expected frequency of occurrence of each category is  $350/90 = 3.89$ , and the observed frequencies are those shown in Figure 19.1-1. The resulting value of the test statistic is  $\chi^2 = 1231.0$ . If this number exceeds the critical value for the null distribution at  $\alpha = 0.05$ , then we can reject  $H_0$ .

1231.0 /

315/90  
= 3.5

But here we run into a problem. The expected frequency of ~~3.89~~ for each category violates the requirements of the  $\chi^2$  test that no more than 20% of the categories should have expected values less than five. As a result, the null distribution of the test statistic  $\chi^2$  is not a  $\chi^2$  distribution, so we cannot use that distribution to calculate a  $P$ -value. How do we determine the null distribution of our test statistic so that we can get a  $P$ -value?

3.5

One possible solution is to use computer simulation to generate the null distribution for the test statistic. Here's how it's done, in five steps:

1. Use a computer to create and sample an imaginary population whose parameter values are those specified by the null hypothesis. In the case of the

**TABLE 19.1-1** A subset of the results of a simulation. Each row has the first 12 of 350 numbers randomly sampled from an imaginary population in which all numbers between 10 and 99 occur with equal probability. The last column has the  $\chi^2$  statistic calculated on each simulated sample of 350 numbers.

315 /

/ 315

Simulation number	Simulated samples of 350 numbers (the first 12 numbers of each sample are shown)												Test statistic $\chi^2$	
1	32	42	27	74	78	86	98	71	28	50	41	54	...	86.1
2	38	63	98	36	88	10	74	35	62	52	90	48	...	80.5
3	52	45	59	44	44	25	94	29	27	64	24	47	...	95.4
4	69	55	31	42	35	48	52	37	91	40	67	67	...	78.4
5	11	66	32	11	76	96	73	20	64	40	37	49	...	87.7
6	46	15	87	23	92	18	91	26	31	23	40	51	...	96.9
7	61	35	58	33	58	82	67	95	16	64	59	64	...	95.4
8	47	54	16	39	91	68	49	57	10	21	79	51	...	106.7
9	21	84	30	66	81	13	16	18	81	91	52	95	...	92.3
10	88	26	48	44	34	72	89	14	98	35	99	54	...	100.5
...														...

/ 315

mentalist's numbers, simulating a single sample involves drawing 350 two-digit numbers between 10 and 99 at random and with equal probability. Each simulated sample must have 350 numbers, because 350 is the sample size of the real data. The first row of Table 19.1-1 lists the first 12 numbers of our first simulated sample of 350 numbers.

/ 315

/ 315 / 315

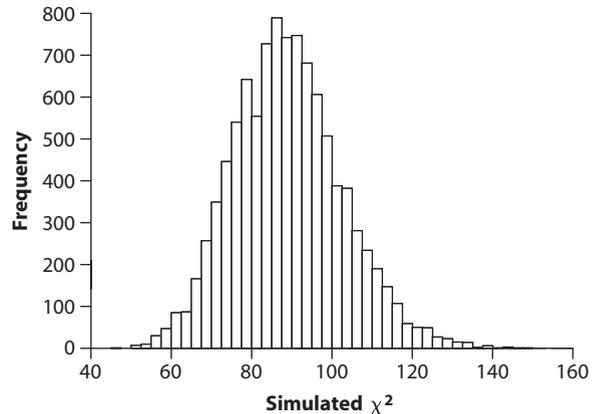
/ 315

2. Calculate the test statistic on the simulated sample. For the number-choosing example, we have decided to use  $\chi^2$  as the test statistic. This statistic does not necessarily have a  $\chi^2$  distribution under  $H_0$ , because of the small expected frequencies, but  $\chi^2$  is still a suitable measure of the fit between the data and the null hypothesis. In our first simulated sample, the  $\chi^2$  value turned out to be 86.1 (Table 19.1-1).
3. Repeat steps 1 and 2 a large number of times. We repeated the simulated sampling process 10,000 times, calculating  $\chi^2$  each time. (Typically a simulation should involve at least 1000 replicate samples.) Table 19.1-1 shows a subset of outcomes for the first 10 of our simulated samples.
4. Gather all of the simulated values for the test statistic to form the null distribution. The distribution of simulated test statistics can be used as the null distribution of the estimate. The frequency distribution of all 10,000 values for  $\chi^2$  that we obtained from our example simulations is plotted in Figure 19.1-2. This is our approximate null distribution for the  $\chi^2$  statistic.
5. Compare the test statistic from the data to the null distribution. We use the simulated null distribution to get an approximate  $P$ -value. In an ordinary  $\chi^2$  goodness-of-fit test, the  $P$ -value for the  $\chi^2$  statistic is the probability under the null distribution of obtaining a  $\chi^2$  statistic as large or larger than the observed

**FIGURE 19.1-2**

The null distribution for the  $\chi^2$  statistic based on 10,000 simulated random samples from an imaginary population conforming to the null hypothesis (Example 19.1). The test statistic calculated from the data,  $\chi^2 = 1231.0$ , is far greater than for any of the simulated samples.

1231.0 /



1231.0 /

value of the test statistic. The same is true with a simulated null distribution for  $\chi^2$ : the  $P$ -value is approximated by the fraction of simulated values for  $\chi^2$  that equal or exceed the observed value of  $\chi^2$  (in our case,  $\chi^2 = 1231.0$ ). According to Figure 19.1-2, *none* of the 10,000 simulated  $\chi^2$  values exceeded the observed  $\chi^2$  statistic. This means that the approximate  $P$ -value is less than 1 in 10,000 (i.e.,  $P < 0.0001$ ).<sup>1</sup> To be more precise, we would have to run more simulations.

These results show that when people choose numbers haphazardly, the outcome is highly non-random. This can make mentalists appear to have telepathic powers when the only power they possess is that of statistics.

## 19.2 Bootstrap standard errors and confidence intervals

The **bootstrap** is a computer-intensive procedure used to approximate the sampling distribution of an estimate. Bootstrapping creates this sampling distribution by taking new samples randomly and repeatedly *from the data themselves*. Unlike simulation, the bootstrap is not directly intended for testing hypotheses. Instead, the bootstrap is used to find a standard error or confidence interval for a parameter estimate. The bootstrap is especially useful when no formula is available for the standard error or when the sampling distribution of the estimate of interest is unknown.

Recall from Section 4.1 that the sampling distribution is the probability distribution of sample estimates when a population is sampled repeatedly in the same way. The standard error is the standard deviation of this sampling distribution. In principle,

1. We can't say  $P = 0$ , because we might find a more extreme value from the null distribution if we ran more simulations.

## Confidence intervals by bootstrapping

The approximate sampling distribution generated by the bootstrap can also be used to calculate an approximate confidence interval for the population parameter. We present the most commonly used method here.<sup>7</sup> The bootstrap  $1 - \alpha$  confidence interval can be obtained from the bootstrap sampling distribution by finding the points that separate  $\alpha/2$  of the distribution into each of the left and right tails. In other words, an approximate 95% confidence interval ranges from the 0.025 quantile to the 0.975 quantile of the bootstrap sampling distribution.

For example, let's compute a 95% confidence interval for the population median asymmetry using the bootstrap sampling distribution displayed in Figure 19.2-2. To determine the lower bound of the 95% confidence interval, we must find the 0.025 quantile—the value that was greater than or equal to the 250th (i.e.,  $0.025 \times 10,000$ ) sorted bootstrap replicate estimate. In our example, after sorting the 10,000 bootstrap replicate estimates from small to large, the 250th sorted value was  $-0.075$ . To determine the upper bound of the confidence interval, we must find the value that is less than or equal to 2.5% of the sorted bootstrap estimates—in other words, the 9751st measurement in the sorted bootstrap estimates (note that 250 out of 10,000 values are equal to or greater than the 9751st estimate). For the chimp brain data, this number is 0.31. As a result, the 95% bootstrap confidence interval for the population median asymmetry of Brodmann's area 44 in chimps is

$$-0.075 < \text{median} < 0.31.$$

This is a relatively wide confidence interval, indicating that the data are consistent with a broad range of possible values for the population median asymmetry of Brodmann's area 44 in chimps. This asymmetry of brain structure, thought to be so important for language development in humans, might have a median as low as zero (or very slightly larger on the left side than on the right side of the brain) or as large as 0.31 in our closest relative.<sup>8</sup>

right // left //

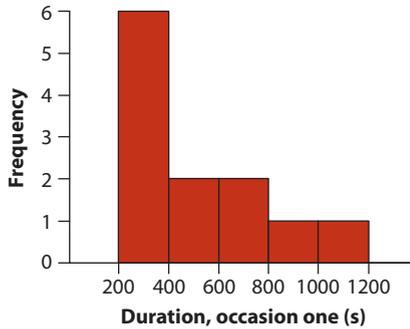
## Bootstrapping with multiple groups

Bootstrapping can be used for just about any statistic that can be estimated and for which conventional methods are unavailable. For example, bootstrapping may be used to obtain standard errors and confidence intervals for measures of the difference between groups. Here we show a bootstrap method to compare two groups whose frequency distributions are not normally distributed and might not have the same shape. The data from each group is resampled separately, and the bootstrap sampling uses the original sample sizes for each group. This bootstrap procedure generates “new” data sets that mimic sampling repeatedly from the original populations.

7. Consult the excellent book *An Introduction to the Bootstrap*, by Efron and Tibshirani (1993), for more details and other options.

8. The 95% confidence interval for *mean* asymmetry is similarly broad, but it doesn't overlap zero.

18. (a) A histogram is appropriate, though a box plot would also suffice.



- (b) Positive skew.

## Chapter 3

1. (a)  $n = 9$ . (b)  $\sum_{i=1}^n Y_i = 1179$ .

(c)  $\frac{\sum_{i=1}^n Y_i}{n} = 131$  mm Hg.

(d)  $\sum_{i=1}^n (Y_i - \bar{Y})^2 = 2036$ .

(e)  $s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1} = 254.5$  mm Hg<sup>2</sup>

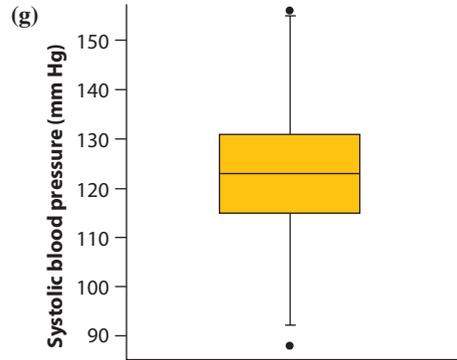
(f)  $s = \sqrt{s^2} = \sqrt{254.5} = 15.95$  mm Hg.

(g)  $CV = 100\% \frac{s}{\bar{Y}} = 12.2\%$ .

2. (a) There are 101 data points, an odd number, and so the middle data point will be the 51st. In the ordered set of data, this is 123. (b)  $k = 0.75 \times 101 = 75.75$ . Rounding up, we want the 76th data point. This is 131. (c)  $j = 0.25 \times 101 = 25.25$ . Rounding up, we want the 26th data point. This is 115. (d) The interquartile range is the upper quartile minus the lower quartile, or  $131 - 115 = 16$ . (e) 1.5 times the *IQR* is 24. The upper quartile (131)

24

plus 24 is 155. No. Therefore there will be a dot outside the “whiskers.” (f)  $115 - 24 = 91$ . No.



- (g) 3. (a) The number of doctors studied who sterilized a given percentage of patients under age 25. (b) The median would be more informative. The mean would be sensitive to the presence of the outlier (sterilizing more than 30% of female patients under 25), whereas the median is unaffected. (c) Yes, because there is a clear outlier. 4. (a) Box plot. (b) Median body mass of the mammals in each group. (c) The first and third quartiles of body mass in each group. (d) “Extreme values,” those lying farther than 1.5 times the interquartile range from the box edge. (e) Whiskers. They extend to the smallest and largest values in the data, excluding extreme values (those lying farther than 1.5 times the interquartile range from the box edge). (f) Living mammals have the smallest median body size (with a log body mass of about 2), and mammals that went extinct in the last ice age had the largest median size (around 5.5). Mammals that went extinct recently were intermediate in size, with a median of around 3.2. (g) The body size distribution of living mammals is right-skewed (long tail toward larger values), whereas the size frequency distribution in mammals that went extinct in the last ice age is left-skewed. The frequency distribution of sizes is nearly symmetric in mammals that went extinct recently. (h) One way to answer this is to use the interquartile range as a measure of spread, in which case the mammals that went extinct in the last ice age have the lowest spread. Another way to compare spread is to calculate the standard deviation of log body size in each group, but we

## Chapter 5

1. (a) They are mutually exclusive because each respondent can select only one answer. Therefore, two cannot occur. (b)  $\Pr[\text{very repulsive or somewhat repulsive}] = \Pr[\text{very repulsive}] + \Pr[\text{somewhat repulsive}] = 0.30 + 0.20 = 0.50$ . (c)  $\Pr[\text{not especially delicious}] = 1 - \Pr[\text{especially delicious}] = 1 - 0.01 = 0.99$ .

2. (a) 0.48. 0.52. (Probability tree is below.)

(b) Events are mutually exclusive:  $0.08 + 0.01 = 0.09$ . (c)  $\Pr[\text{somewhat delicious or especially delicious} | \text{man}] = 0.09$ . (d)  $\Pr[\text{somewhat delicious or especially delicious} | \text{woman}] = 0.06 + 0.01 = 0.07$ .

(e)

Sex	Response	Probability
Woman	Somewhat delicious	0.0312
	Especially delicious	0.0052
	Other answer	0.4836
Man	Somewhat delicious	0.0384
	Especially delicious	0.0048
	Other answer	0.4368

(f)  $\Pr[\text{somewhat delicious or especially delicious}] = \Pr[\text{woman}] \Pr[\text{somewhat delicious or especially delicious} | \text{woman}] + \Pr[\text{man}] \Pr[\text{somewhat delicious or especially delicious} | \text{man}] = (0.52)(0.07) + (0.48)(0.09) = 0.0796$ .

3. (a)  $\Pr[\text{HPV or Chlamydia}]$ . (b)  $\Pr[\text{HPV or Chlamydia}] = \Pr[\text{HPV}] + \Pr[\text{Chlamydia}] - \Pr[\text{HPV and Chlamydia}]$ . (c)  $\Pr[\text{HPV}] = 0.24 + 0.04 = 0.28$ .  $\Pr[\text{Chlamydia}] = 0.02 + 0.04 = 0.06$ .  $\Pr[\text{HPV or Chlamydia}] = 0.28 + 0.06 - 0.04 = 0.30$ .

4. (a)  $\Pr[\text{cancer} | \text{smoker}] = 0.172$ .

(b)

Smoking	Cancer	Probability
Smoker	Lung cancer	0.089
	Not	0.431
Nonsmoker	Lung cancer	0.006
	Not	0.474

(c)  $\Pr[\text{smoker and cancer}] = (0.52)(0.172) = 0.089$ . (d)  $\Pr[\text{smoker and cancer}] = \Pr[\text{smoker}] \Pr[\text{cancer} | \text{smoker}] = (0.52)(0.172) = 0.089$ .

Yes. (e)  $\Pr[\text{nonsmoker and no cancer}] = \Pr[\text{nonsmoker}] \Pr[\text{no cancer} | \text{nonsmoker}] = (0.48)(0.987) = 0.474$ .

5. (a)  $\Pr[\text{smoker} | \text{cancer}] = \Pr[\text{cancer} | \text{smoker}] \Pr[\text{smoker}] / \Pr[\text{cancer}] = 0.089 / 0.095 = 0.937$ . (b)  $\Pr[\text{cancer}] = 0.089 + 0.006 = 0.095$ . (c)  $\Pr[\text{smoker} | \text{cancer}] = 0.089 / 0.095 = 0.937$ .
6. (a) 5/8. (b) 1/4. (c) 7/8 (either in this case means pepperoni or anchovies or both). (d) No (some slices have both pepperoni and anchovies). (e) Yes. Olives and mushrooms are mutually exclusive. (f) No.  $\Pr[\text{mushrooms}] = 3/8$ ;  $\Pr[\text{anchovies}] = 1/2$ ; if independent,  $\Pr[\text{mushrooms and anchovies}] = \Pr[\text{mushrooms}] \times \Pr[\text{anchovies}] = 3/16$ . Actual probability = 1/8. Not independent. (g)  $\Pr[\text{anchovies} | \text{olives}] = 1/2$  (two slices have olives, and one of these two has anchovies). (h)  $\Pr[\text{olives} | \text{anchovies}] = 1/4$  (four slices have anchovies, and one of these has olives). (i)  $\Pr[\text{last slice has olives}] = 1/4$  (two of the eight slices have olives; you still get one slice—it doesn't matter whether your friends pick before you or after you). (j)  $\Pr[\text{two slices with olives}] = \Pr[\text{first slice has olives}] \times \Pr[\text{second slice has olives} | \text{first slice has olives}] = 2/8 \times 1/7 = 1/28$ . (k)  $\Pr[\text{slice without pepperoni}] = 1 - \Pr[\text{slice with pepperoni}] = 3/8$ . (l) Each piece has either one or no topping.
7.  $\Pr[\text{encounter and success}] = \Pr[\text{encounter}] \Pr[\text{capture} | \text{encounter}] = (0.035)(0.40) = 0.014$ .
8. Of 273 trees, 45 trees have cavities, so the probability of choosing a tree with a cavity is  $45/273 = 0.165$ .
9. (a)  $\Pr[\text{vowel}] = \Pr[A] + \Pr[E] + \Pr[I] + \Pr[O] + \Pr[U] = 8.2\% + 12.7\% + 7.0\% + 7.5\% + 2.8\% = 38.2\%$ . (b)  $\Pr[\text{five randomly chosen letters from an English text spell "STATS"}] = \Pr[S] \times \Pr[T] \times \Pr[A] \times \Pr[T] \times \Pr[S] = 0.063 \times 0.091 \times 0.082 \times 0.091 \times 0.063 = 2.7 \times 10^{-6}$ . (Each draw is independent, but all must be successful to satisfy the conditions, so we must multiply the probability of each independent event.) (c)  $\Pr[2 \text{ letters from an English text} = "e"] = 0.127 \times 0.127 = 0.016$ .
10. (a)  $\Pr[A_1 \text{ or } A_4] = \Pr[A_1] + \Pr[A_4] = 0.06 + 0.03 = 0.09$ . (b)  $\Pr[A_1 \text{ and } A_1] = \Pr[A_1] \Pr[A_1] = 0.06 \times 0.06 = 0.0036$ . (c)  $\Pr[\text{not } (A_1 \text{ and } A_1)]$

See next page - answers to (c) and (d) need to be swapped

- $= 1 - \Pr[A_1 \text{ and } A_1] = 1 - 0.0036 = 0.9964$ . **(d)**  $\Pr[A_1 A_3] = (0.06)(0.84) + (0.84)(0.06) = 0.1008$ .
- (e)**  $\Pr[\text{two individuals not } A_1 A_1] = \Pr[\text{not } A_1 A_1] \Pr[\text{not } A_1 A_1] = 0.9964 \cdot 0.9964 = 0.9928$ .
- (f)**  $\Pr[\text{at least one of two individuals is } A_1 A_1] = 1 - \Pr[\text{neither is } A_1 A_1] = 1 - 0.9928 = 0.0072$ .
- (g)**  $\Pr[\text{three individuals have no } A_2 \text{ or } A_3 \text{ alleles}] = \Pr[\text{six alleles are not } A_2 \text{ or } A_3] = (1 - 0.84 - 0.03)^6 = (0.13)^6 = 0.0000048$ .
- 11. (a)**  $\Pr[\text{no dangerous snakes}] = \Pr[\text{not dangerous in the left hand}] \times \Pr[\text{not dangerous in the right hand}] = 3/8 \times 2/7 = 6/56 = 0.107$ .
- (b)**  $\Pr[\text{bite}] = \Pr[\text{bite} \mid 0 \text{ dangerous snakes}] \Pr[0 \text{ dangerous snakes}] + \Pr[\text{bite} \mid 1 \text{ dangerous snake}] \Pr[1 \text{ dangerous snake}] + \Pr[\text{bite} \mid 2 \text{ dangerous snakes}] \Pr[2 \text{ dangerous snakes}]$ .  
 $\Pr[0 \text{ dangerous snakes}] = 0.107$  [from part (a)].  
 $\Pr[1 \text{ dangerous snake}] = (5/8 \times 3/7) + (3/8 \times 5/7) = 0.536$ .  
 $\Pr[2 \text{ dangerous snakes}] = 5/8 \times 4/7 = 0.357$ .  
 $\Pr[\text{bite} \mid 0 \text{ dangerous snakes}] = 0$ .  
 $\Pr[\text{bite} \mid 1 \text{ dangerous snake}] = 0.8$ .  
 $\Pr[\text{bite} \mid 2 \text{ dangerous snakes}] = 1 - (1 - 0.8)^2 = 0.96$ .  
 Putting these all together:  
 $\Pr[\text{bite}] = (0 \times 0.107) + (0.8 \times 0.536) + (0.96 \times 0.357) = 0.772$ .
- (c)**  $\Pr[\text{defanged} \mid \text{no bite}] = \frac{\Pr[\text{no bite} \mid \text{defanged}] \Pr[\text{defanged}]}{\Pr[\text{no bite}]}$
- $\Pr[\text{no bite} \mid \text{defanged}] = 1$ ;  $\Pr[\text{defanged}] = 3/8$ ;  
 $\Pr[\text{no bite}] = \Pr[\text{defanged}] \Pr[\text{no bite} \mid \text{defanged}] + \Pr[\text{dangerous}] \Pr[\text{no bite} \mid \text{dangerous}]$   
 $= (3/8 \times 1) + [5/8 \times (1 - 0.8)] = 0.5$ .
- So,  $\Pr[\text{defanged} \mid \text{one snake did not bite}] = (1.0 \times 3/8)/(0.5) = 0.75$ .
- 12. (a)**  $\Pr[\text{all five researchers calculate 95\% CI with the true value}]$ ? Each one has a 95% chance, all samples are independent, so  $\Pr = (0.95)^5 = 0.774$ . **(b)**  $\Pr[\text{at least one does not include true parameter}] = 1 - \Pr[\text{all include true parameter}] = 1 - 0.774 = 0.226$ .
- 13. (a)** 0.99. **(b)**  $\Pr[\text{cat survives seven days}] = \Pr[\text{cat not poisoned one day}]^7 = (0.99)^7 = 0.932$ . **(c)**  $\Pr[\text{cat survives a year}] = \Pr[\text{cat not poisoned one day}]^{365} = (0.99)^{365} = 0.026$ .  
**(d)**  $\Pr[\text{cat dies within year}] = 1 - \Pr[\text{cat survives year}] = 1 - 0.026 = 0.974$ .
- 14. (a)** Of the 1347 people who did not have HIV, 129 tested positive. Therefore the false-positive rate is  $129/1347 = 0.096$ . **(b)** Of the 170 people with HIV, 4 tested negative. The false-negative rate is  $4/170 = 0.024$ . **(c)** Use Bayes' theorem:  $\Pr[\text{HIV} \mid \text{positive test}] = \Pr[\text{positive test} \mid \text{HIV}] \Pr[\text{HIV}] / \Pr[\text{positive test}] = (166/170)(170/1517) / ((129+166)/1517) = 0.56$ .
- 15.** Sampling a *Wnt*-responsive cell has probability 0.09, whereas the probability is 0.91 of a nonresponsive cell.
- (a)**  $\Pr[\text{WWLWWW}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}$ . **(b)**  $\Pr[\text{WWWWWL}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}$ .  
**(c)**  $\Pr[\text{LWWWWW}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}$ . **(d)**  $\Pr[\text{WLWLWL}] = 0.09^3 \times 0.91^3 = 5.5 \times 10^{-4}$ .  
**(e)**  $\Pr[\text{WWWLLL}] = 0.09^3 \times 0.91^3 = 5.5 \times 10^{-4}$ . **(f)**  $\Pr[\text{WWWWW}] = 0.09^6 = 5.3 \times 10^{-7}$ .  
**(g)**  $\Pr[\text{at least one nonresponsive cell}] = 1 - \Pr[\text{WWWWW}] = 1 - 0.09^6 = 0.9999995$ .
- 16.**  $\Pr[\text{next person will wash his/her hands}] = \Pr[\text{wash} \mid \text{man}] \times \Pr[\text{man}] + \Pr[\text{wash} \mid \text{woman}] \times \Pr[\text{woman}] = 0.74 \times 0.4 + 0.83 \times 0.6 = 0.794$ .
- 17. (a)**  $\Pr[\text{one person not blinking}] = 1 - \Pr[\text{person blinks}] = 1 - 0.04 = 0.96$ .  
**(b)**  $\Pr[\text{at least one blink in 10 people}] = 1 - \Pr[\text{no one blinks}] = 1 - (0.96)^{10} = 0.335$ .

The 6 should be a superscript

Add parentheses  
around 0.9964 2x

$= 1 - \Pr[A_1 \text{ and } A_1] = 1 - 0.0036 = 0.9964.$  (d)  $\Pr[A_1 A_3] = (0.06)(0.84) + (0.84)(0.06) = 0.1008.$

(e)  $\Pr[\text{two individuals not } A_1 A_1] = \Pr[\text{not } A_1 A_1] \Pr[\text{not } A_1 A_1] = 0.9964 \cdot 0.9964 = 0.9928.$

(f)  $\Pr[\text{at least one of two individuals is } A_1 A_1] = 1 - \Pr[\text{neither is } A_1 A_1] = 1 - 0.9928 = 0.0072.$

(g)  $\Pr[\text{three individuals have no } A_2 \text{ or } A_3 \text{ alleles}] = \Pr[\text{six alleles are not } A_2 \text{ or } A_3] = (1 - 0.84 - 0.03)^6 = (0.13)^6 = 0.0000048.$

11. (a)  $\Pr[\text{no dangerous snakes}] = \Pr[\text{not dangerous in the left hand}] \times \Pr[\text{not dangerous in the right hand}] = 3/8 \times 2/7 = 6/56 = 0.107.$

(b)  $\Pr[\text{bite}] = \Pr[\text{bite} \mid 0 \text{ dangerous snakes}] \Pr[0 \text{ dangerous snakes}] + \Pr[\text{bite} \mid 1 \text{ dangerous snake}] \Pr[1 \text{ dangerous snake}] + \Pr[\text{bite} \mid 2 \text{ dangerous snakes}] \Pr[2 \text{ dangerous snakes}].$

$\Pr[0 \text{ dangerous snakes}] = 0.107$  [from part (a)].

$\Pr[1 \text{ dangerous snake}] = (5/8 \times 3/7) + (3/8 \times 5/7) = 0.536.$

$\Pr[2 \text{ dangerous snakes}] = 5/8 \times 4/7 = 0.357.$

$\Pr[\text{bite} \mid 0 \text{ dangerous snakes}] = 0.$

$\Pr[\text{bite} \mid 1 \text{ dangerous snake}] = 0.8.$

$\Pr[\text{bite} \mid 2 \text{ dangerous snakes}] = 1 - (1 - 0.8)^2 = 0.96.$

Putting these all together:

$\Pr[\text{bite}] = (0 \times 0.107) + (0.8 \times 0.536) + (0.96 \times 0.357) = 0.772.$

Answer given as (d) here should move up to (c); and answer given for (c) starting on previous page) should be the answer for (d)

(c)  $\Pr[\text{defanged} \mid \text{no bite}] = \frac{\Pr[\text{no bite} \mid \text{defanged}] \Pr[\text{defanged}]}{\Pr[\text{no bite}]}$

$\Pr[\text{no bite} \mid \text{defanged}] = 1; \Pr[\text{defanged}] = 3/8;$

$\Pr[\text{no bite}] = \Pr[\text{defanged}] \Pr[\text{no bite} \mid \text{defanged}] + \Pr[\text{dangerous}] \Pr[\text{no bite} \mid \text{dangerous}]$   
 $= (3/8 \times 1) + [5/8 \times (1 - 0.8)] = 0.5.$

So,  $[\text{defanged} \mid \text{one snake did not bite}] = (1.0 \times 3/8)/(0.5) = 0.75.$

12. (a)  $\Pr[\text{all five researchers calculate 95\% CI with the true value}]?$  Each one has a 95% chance, all samples are independent, so  $\Pr = (0.95)^5 = 0.774.$  (b)  $\Pr[\text{at least one does not include true parameter}] = 1 - \Pr[\text{all include true parameter}] = 1 - 0.774 = 0.226.$

13. (a) 0.99. (b)  $\Pr[\text{cat survives seven days}] = \Pr[\text{cat not poisoned one day}]^7 = (0.99)^7 = 0.932.$  (c)  $\Pr[\text{cat survives a year}] = \Pr[\text{cat not poisoned one day}]^{365} = (0.99)^{365} = 0.026.$

(d)  $\Pr[\text{cat dies within year}] = 1 - \Pr[\text{cat survives year}] = 1 - 0.026 = 0.974.$

14. (a) Of the 1347 people who did not have HIV, 129 tested positive. Therefore the false-positive rate is  $129/1347 = 0.096.$  (b) Of the 170 people with HIV, 4 tested negative. The false-negative rate is  $4/170 = 0.024.$  (c) Use Bayes' theorem:  $\Pr[\text{HIV} \mid \text{positive test}] = \frac{\Pr[\text{positive test} \mid \text{HIV}] \Pr[\text{HIV}]}{\Pr[\text{positive test}]} = \frac{(166/170)(170/1517)}{((129+166)/1517)} = 0.56.$

15. Sampling a *Wnt*-responsive cell has probability 0.09, whereas the probability is 0.91 of a nonresponsive cell.

(a)  $\Pr[\text{WWLWWW}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}.$  (b)  $\Pr[\text{WWWWWL}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}.$

(c)  $\Pr[\text{LWWWWW}] = 0.09^5 \times 0.91 = 5.4 \times 10^{-6}.$  (d)  $\Pr[\text{WLWLWL}] = 0.09^3 \times 0.91^3 = 5.5 \times 10^{-4}.$

(e)  $\Pr[\text{WWWLLL}] = 0.09^3 \times 0.91^3 = 5.5 \times 10^{-4}.$  (f)  $\Pr[\text{WWWWW}] = 0.09^6 = 5.3 \times 10^{-7}.$

(g)  $\Pr[\text{at least one nonresponsive cell}] = 1 - \Pr[\text{WWWWW}] = 1 - 0.096 = 0.9999995.$

16.  $\Pr[\text{next person will wash his/her hands}] = \Pr[\text{wash} \mid \text{man}] \times \Pr[\text{man}] + \Pr[\text{wash} \mid \text{woman}] \times \Pr[\text{woman}] = 0.74 \times 0.4 + 0.83 \times 0.6 = 0.794.$

17. (a)  $\Pr[\text{one person not blinking}] = 1 - \Pr[\text{person blinks}] = 1 - 0.04 = 0.96.$

(b)  $\Pr[\text{at least one blink in 10 people}] = 1 - \Pr[\text{no one blinks}] = 1 - (0.96)^{10} = 0.335.$

## Chapter 6

- Statement (a) is correct. If the estimate that the test is based on is biased, the estimate will on average be different from the true value. As a result, the probability of rejecting the (true) null hypothesis is increased, and therefore the Type I error rate is increased.
- False. The Type 1 error rate is set by the experimenter, and it will be accurate provided the sample is a random sample.
- (a) Failing to reject a false null hypothesis. (b) The probability ( $\alpha$ ) used as a criterion for rejecting the null hypothesis; if the  $P$ -value is less than or equal to  $\alpha$ , then the null hypothesis is rejected, otherwise the null hypothesis is not rejected. (c) Setting a higher significance level,  $\alpha$ , such as raising it to 0.05 instead of 0.01, **decreases** / ~~increases~~ the probability of failing to reject a false null hypothesis.
- (a) True. (b) False. (c) False.
- (a)  $H_0$ : The rate of correct guesses is  $1/6$ . (b)  $H_A$ : The rate of correct guesses is not  $1/6$ .
- (a) Alternative hypothesis. (b) Alternative hypothesis. (c) Null hypothesis. (d) Alternative hypothesis. (e) Null hypothesis.
- (a) Lowers the probability of committing a Type I error. (b) Increases the probability of committing a Type II error. (c) Lowers power of a test. (d) No effect.
- (a) No effect. (b) Decreases the probability of committing a Type II error. (c) Increases the power of a test. (d) No effect.
- (a)  $P = 2 \times (\text{Pr}[15] + \text{Pr}[16] + \text{Pr}[17] + \text{Pr}[18]) = 0.0075$ . (b)  $P = 2 \times (\text{Pr}[13] + \text{Pr}[14] + \dots + \text{Pr}[18]) = 0.096$ . (c)  $P = 2 \times (\text{Pr}[10] + \text{Pr}[11] + \text{Pr}[12] + \dots + \text{Pr}[18]) = 0.815$ . (d)  $P = 2 \times (\text{Pr}[0] + \text{Pr}[1] + \text{Pr}[2] + \text{Pr}[3] + \dots + \text{Pr}[7]) = 0.481$ .
- Failing to reject  $H_0$  does not mean  $H_0$  is correct, because the power of the test might be limited. The null hypothesis is the default and is either rejected or not rejected.
- Begin by stating the hypotheses.  $H_0$ : Size on islands does not differ in a consistent direction from size on mainlands in Asian large mammals (i.e.,  $p = 0.5$ );  $H_A$ : Size on islands differs in a

consistent direction from size on mainlands in Asian large mammals (i.e.,  $p \neq 0.5$ ), where  $p$  is the true fraction of large mammal species that are smaller on islands than on the mainland. Note that this is a two-tailed test. The test statistic is the observed number of mammal species for which size is smaller on islands than mainland: 16. The  $P$ -value is the probability of a result as unusual as 16 out of 18 when  $H_0$  is true:  $P = 2 \times (\text{Pr}[16] + \text{Pr}[17] + \text{Pr}[18]) = 0.00135$ . Since  $P < 0.05$ , reject  $H_0$ . Conclude that size on islands is usually smaller than on mainlands in Asian large mammals.

- (a) Not correct. The  $P$ -value does not give the size of the effect. (b) Correct.  $H_0$  was rejected, so we conclude that there is indeed an effect. (c) Not correct. The probability of committing a Type I error is set by the significance level, 0.05, which is decided beforehand. (d) Not correct. The probability of committing a Type II error depended on the effect size, which wasn't known. (e) Correct.
- Their test almost certainly failed to reject  $H_0$ , because the 95% confidence interval includes the value of the parameter stated in the null hypothesis (i.e., 1).
- (a)  $H_0$ : Subjects pick the mother correctly one time in two ( $p = 1/2$ ),  $H_A$ : Subjects pick the mother correctly more than one time in two ( $p > 1/2$ ). (b) One-sided, because the alternative hypothesis considers parameter values on one side of the parameter value stated in the null hypothesis. This seems justified here because it is not feasible that sons would resemble their mothers less than randomly chosen women. (d)  $P = 0.1214 + 0.1669 + \dots + 0.000004 = 0.881$  (it is quicker to calculate as  $1 - (0.000004 + 0.00007 + \dots + 0.0708) = 0.881$ ). (e) Since  $P > 0.05$ , do not reject the null hypothesis. (f) Calculate a 95% confidence interval for  $p$ .

## Chapter 7

- (a)  $n$  independent trials, with each having the same probability of "success,"  $p$ . Yes. (b)  $p = 0.30$ ,  $n = 7$ . (c)  $\text{Pr}[5] = \binom{7}{5} (0.3)^5 (0.7)^2 =$

0.0250. (d)  $\Pr[6] = \binom{7}{6} (0.3)^6 (0.7)^1 = 0.0036$ .  $\Pr[7] = \binom{7}{7} (0.3)^7 (0.7)^0 = 0.0002$ . (e)  $\Pr[5 \text{ or more}] = \Pr[5] + \Pr[6] + \Pr[7] = 0.0250 + 0.0036 + 0.0002 = 0.0288$ .

2. (a)  $H_0$ : Probability of right turn equal to the probability of a left turn ( $p = 0.5$ );  $H_A$ : Probability of right turn does not equal the probability of a left turn ( $p \neq 0.5$ ). (b) Number of right ears is 19. (c)  $\Pr[19] = \binom{25}{19} (0.5)^{19} (0.5)^6 = 0.00528$ . (d) 20, 21, 22, 23, 24, 25. (e)  $\Pr[20] = \binom{25}{20} (0.5)^{20} (0.5)^5 = 0.00158$ ;  $\Pr[21] = \binom{25}{21} (0.5)^{21} (0.5)^4 = 0.00038$ ;  $\Pr[22] = 7 \times 10^{-5}$ ;  $\Pr[23] = 9 \times 10^{-6}$ ;  $\Pr[24] = 7 \times 10^{-7}$ ;

$\Pr[25] = 3 \times 10^{-8}$ . (f)  $\Pr[19 \text{ or more}] = \Pr[19] + \Pr[20] + \dots + \Pr[25] = 0.00528 + 0.00158 + 0.00038 + \dots = 0.0073$ . (g)  $P = 2(0.0073) = 0.0146$ . (h)  $P$  is the probability of a result as extreme or more extreme than that observed if the null hypotheses were true. (i) Reject  $H_0$ , because  $P < 0.05$ . More people use the right ear than use the left ear when listening to a stranger in the noisy nightclub.

3. (a) Estimated proportion returned  $= 101/240 = 0.421$ . (b)  $p' = \frac{X+2}{n+4} = \frac{101+2}{240+4} = 0.4221$

(c) The lower bound is  $p' - 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.4221 - 1.96\sqrt{\frac{0.4211(1-0.4211)}{240+4}} = 0.360$

(d) The upper bound is  $p' + 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.4221 + 1.96\sqrt{\frac{0.4211(1-0.4211)}{240+4}} = 0.484$

(e) Answers will vary. For example,  $p = 0.35$  and  $p = 0.45$  lie inside, whereas  $p = 0.5$  and  $p = 0.8$  lie outside. (f) It is likely that  $H_0: p = 0.5$  would be rejected, because 0.5 lies outside the most-plausible range of values indicated by the 95% confidence interval.

4. (a) 91/220 had cancer, so the estimated probability of a cast or crew member developing cancer is 0.414. (b) The standard error of the proportion is 0.033. [The square-root of  $(0.414)(1-0.414)/(220-1)$ ]. This quantity measures the standard deviation of the sampling distribution of the proportion. (c) Using the Agresti–Coull method to generate confidence intervals, we first calculate  $p' = (X+2)/(n+4) = (91+2)/(220+4) = 0.415$ . The lower bound for the 95% confidence interval is

$0.415 - 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.351$ . The upper bound is  $0.415 + 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.480$ . The

confidence intervals do not bracket the typical 14% cancer rate for the age group. It is unlikely that 14% is the true cancer rate for this group.

5. (a) 46/50 bills had cocaine, so the estimated proportion is 0.92. (b) The 95% confidence interval, calculated using the Agresti–Coull method:  $p' = 48/54 = 0.89$ ,  $0.805 < p < 0.973$ . (c) We are 95% confident that the true proportion of U.S. one-dollar bills that have a measurable cocaine content lies between 0.805 and 0.973.
6. (a) No, the probability of drawing a red card changes depending on the cards that have been drawn. (b) Yes, the sampling is with replacement, so the probability of success remains constant. (c) No, the probability of drawing a red ball will change with each draw. (d) Yes, the number of red-eyed flies in a sample from a *large* population can be described using the binomial distribution. (Because the population sampled is large, we will assume that the sampling of each individual has a negligible effect on the probability of drawing a red fly on the next draw.) (e) No, the individuals within different families may have different probabilities of having red eyes due to shared genetic differences.
7. (a) This estimate pools numbers from two different groups. Women with rosacea are not a random sample of the population. (b) If the control group women are a random sample,  $15/16 = 0.938$  have mites. (c)  $p' = 17/20 = 0.85$ . The 95% confidence interval is 0.69 to 1.00. (Proportions cannot be larger than 1, so

05 should be 0.5  
(add decimal)

0.0250. **(d)**  $\Pr[6] = \binom{7}{6} (0.3)^6 (0.7)^1 = 0.0036$ .  $\Pr[7] = \binom{7}{7} (0.3)^7 (0.7)^0 = 0.0002$ . **(e)**  $\Pr[5 \text{ or more}] = \Pr[5] + \Pr[6] + \Pr[7] = 0.0250 + 0.0036 + 0.0002 = 0.0288$ .

2. **(a)**  $H_0$ : Probability of right turn equal to the probability of a left turn ( $p = 0.5$ );  $H_A$ : Probability of right turn does not equal the probability of a left turn ( $p \neq 0.5$ ). **(b)** Number of right ears is 19. **(c)**  $\Pr[19] = \binom{25}{19} (0.5)^{19} (0.5)^6 = 0.00528$ . **(d)** 20, 21, 22, 23, 24, 25. **(e)**  $\Pr[20] = \binom{25}{20} (0.5)^{20} (0.5)^5 = 0.00158$ ;  $\Pr[21] = \binom{25}{21} (0.5)^{21} (0.5)^4 = 0.00038$ ;  $\Pr[22] = 7 \times 10^{-5}$ ;  $\Pr[23] = 9 \times 10^{-6}$ ;  $\Pr[24] = 7 \times 10^{-7}$ ;  $\Pr[25] = 3 \times 10^{-8}$ . **(f)**  $\Pr[19 \text{ or more}] = \Pr[19] + \Pr[20] + \dots + \Pr[25] = 0.00528 + 0.00158 + 0.00038 + \dots = 0.0073$ . **(g)**  $P = 2(0.0073) = 0.0146$ . **(h)**  $P$  is the probability of a result as extreme or more extreme than that observed if the null hypotheses were true. **(i)** Reject  $H_0$ , because  $P < 0.05$ . More people use the right ear than use the left ear when listening to a stranger in the noisy nightclub.

3. **(a)** Estimated proportion returned  $= 101/240 = 0.421$ . **(b)**  $p' = \frac{X + 2}{n + 4} = \frac{101 + 2}{240 + 4} = 0.4221$

**(c)** The lower bound is  $p' - 1.96 \sqrt{\frac{p'(1-p')}{n+4}} = 0.4221 - 1.96 \sqrt{\frac{0.4211(1-0.4211)}{240+4}} = 0.360$

**(d)** The upper bound is  $p' + 1.96 \sqrt{\frac{p'(1-p')}{n+4}} = 0.4221 + 1.96 \sqrt{\frac{0.4211(1-0.4211)}{240+4}} = 0.484$

**(e)** Answers will vary. For example,  $p = 0.35$  and  $p = 0.45$  lie inside, whereas  $p = 0.5$  and  $p = 0$  lie outside. **(f)** It is likely that  $H_0: p = 0.5$  would be rejected, because 0.5 lies outside the most-plausible range of values indicated by the 95% confidence interval.

4. **(a)** 91/220 had cancer, so the estimated probability of a cast or crew member developing cancer is 0.414. **(b)** The standard error of the proportion is 0.033. [The square-root of  $(0.414)(1-0.414)/(220-1)$ ]. This quantity measures the standard deviation of the sampling distribution of the proportion. **(c)** Using the Agresti–Coull method to generate confidence intervals, we first calculate  $p' = (X + 2)/(n + 4) = (91 + 2)/(220 + 4) = 0.415$ . The lower bound for the 95% confidence interval is

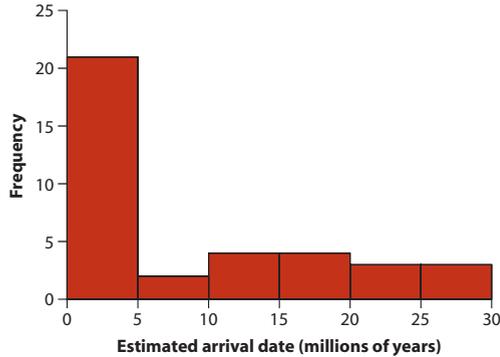
$0.415 - 1.96 \sqrt{\frac{p'(1-p')}{n+4}} = 0.351$ . The upper bound is  $0.415 + 1.96 \sqrt{\frac{p'(1-p')}{n+4}} = 0.480$ . The

confidence intervals do not bracket the typical 14% cancer rate for the age group. It is unlikely that 14% is the true cancer rate for this group.

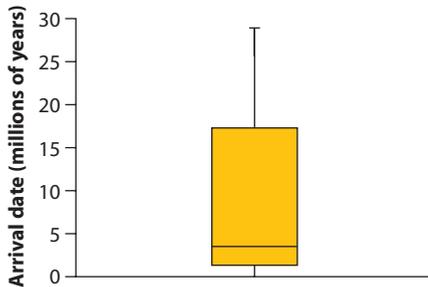
5. **(a)** 46/50 bills had cocaine, so the estimated proportion is 0.92. **(b)** The 95% confidence interval, calculated using the Agresti–Coull method:  $p' = 48/54 = 0.89$ ,  $0.805 < p < 0.973$ . **(c)** We are 95% confident that the true proportion of U.S. one-dollar bills that have a measurable cocaine content lies between 0.805 and 0.973.
6. **(a)** No, the probability of drawing a red card changes depending on the cards that have been drawn. **(b)** Yes, the sampling is with replacement, so the probability of success remains constant. **(c)** No, the probability of drawing a red ball will change with each draw. **(d)** Yes, the number of red-eyed flies in a sample from a large population can be described using the binomial distribution. (Because the population sampled is large, we will assume that the sampling of each individual has a negligible effect on the probability of drawing a red fly on the next draw.) **(e)** No, the individuals within different families may have different probabilities of having red eyes due to shared genetic differences.
7. **(a)** This estimate pools numbers from two different groups. Women with rosacea are not a random sample of the population. **(b)** If the control group women are a random sample,  $15/16 = 0.938$  have mites. **(c)**  $p' = 17/20 = 0.85$ . The 95% confidence interval is 0.69 to 1.00. (Proportions cannot be larger than 1, so

delete "- 1"

13. (a) Histogram shows a sharply right-skewed frequency distribution of ages, with the mode at a young age. There might be a second, low peak at intermediate ages.



- (b) Median appears to be between 0 and 5 million years ago (mya), whereas mean is between 5 and 10 mya. The mean is greater than the median because the distribution is right-skewed: the large values influence the mean more than the median. (c) Mean (8.66 mya) is indeed greater than the median (3.51 mya). (d) First quartile: 1.25 mya; third quartile: 17.30 mya; interquartile range: 16.05 mya. (e) Box plot:

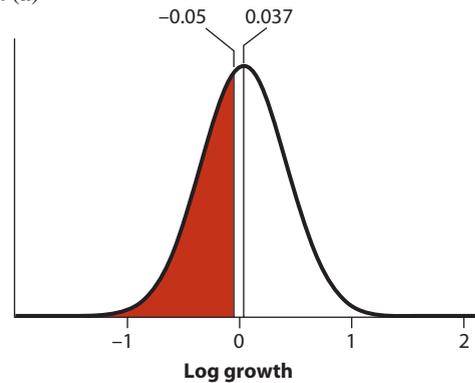


14. False. The given hypothesis is either right or wrong. Hypotheses aren't variables subject to chance. The  $P$ -value is a measure of how unusual the data are if the null hypothesis is true.
15. (a) 3D effects violate "Make patterns in the data easy to see." Small fonts violate "Draw graphical elements clearly." (b) Bar graph.
16. (a) Histogram of whale numbers,  $\chi^2$  goodness-of-fit test to the Poisson distribution. (b) Mosaic plot or grouped bar graph,  $\chi^2$  contingency test or Fisher's exact test. (c) Mosaic plot or grouped bar graph, Fisher's exact test. (d) Histogram,

$\chi^2$  goodness-of-fit test to the Poisson distribution. (e) Mosaic plot or grouped bar graph,  $\chi^2$  contingency test or Fisher's exact test. (f) Bar graph, binomial test (or  $\chi^2$  goodness-of-fit test if the sample size is large enough). (g) Bar graph, binomial test (or  $\chi^2$  goodness-of-fit test if the sample size is large enough).

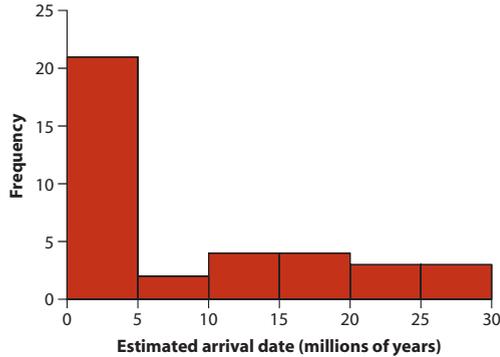
## Chapter 10

1. (a)

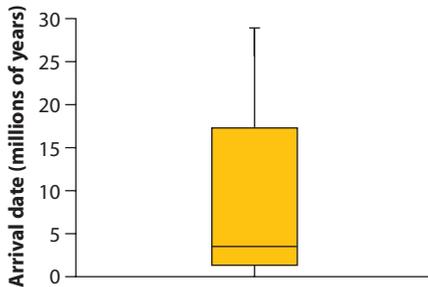


- (b)  $Z = (Y - \text{mean}) / (\text{standard deviation}) = (-0.05 - 0.037) / (0.385) = -0.226$ .
- (c)  $\Pr[Z < -0.226] = \Pr[Z > 0.226]$ , because the standard normal distribution is symmetrical around zero. (d) From Statistical Table B, 0.397.
- (e) Same as (d): 0.397. (f). Same as (e): 0.397.
2. (a)  $H_0$ : The proportion of people who are struck by lightning who are men is 0.5. (b) The mean should be  $np = 648(0.5) = 324$ .
- (c) The standard deviation should be  $\sqrt{np(1-p)} = \sqrt{648(0.5)(0.5)} = 12.7$ .
- (d) 531 is above the mean value in the null hypothesis (324), so we subtract  $1/2$  to get 530.5 with the continuity correction.
- (e)  $Z = (530.5 - 324) / 12.7 = 16.2$ .
- (f)  $\Pr[\text{number of men} \geq 531] = \Pr[Z \geq 16.2]$ . Using Statistical Table A, this is off the chart, so we know that this probability is  $< 0.00002$ .
- (g) The two-tailed value is twice the probability found in the previous step, so  $P < 0.00004$ .
- (h) We conclude that the proportion of men struck by lightning is greater than 0.5.

13. (a) Histogram shows a sharply right-skewed frequency distribution of ages, with the mode at a young age. There might be a second, low peak at intermediate ages.



- (b) Median appears to be between 0 and 5 million years ago (mya), whereas mean is between 5 and 10 mya. The mean is greater than the median because the distribution is right-skewed: the large values influence the mean more than the median. (c) Mean (8.66 mya) is indeed greater than the median (3.51 mya). (d) First quartile: 1.25 mya; third quartile: 17.30 mya; interquartile range: 16.05 mya. (e) Box plot:

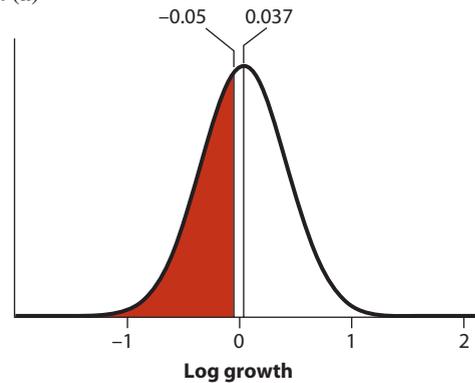


14. False. The given hypothesis is either right or wrong. Hypotheses aren't variables subject to chance. The  $P$ -value is a measure of how unusual the data are if the null hypothesis is true.
15. (a) 3D effects violate "Make patterns in the data easy to see." Small fonts violate "Draw graphical elements clearly." (b) Bar graph.
16. (a) Histogram of whale numbers,  $\chi^2$  goodness-of-fit test to the Poisson distribution. (b) Mosaic plot or grouped bar graph,  $\chi^2$  contingency test or Fisher's exact test. (c) Mosaic plot or grouped bar graph, Fisher's exact test. (d) Histogram,

$\chi^2$  goodness-of-fit test to the Poisson distribution. (e) Mosaic plot or grouped bar graph,  $\chi^2$  contingency test or Fisher's exact test. (f) Bar graph, binomial test (or  $\chi^2$  goodness-of-fit test if the sample size is large enough). (g) Bar graph, binomial test (or  $\chi^2$  goodness-of-fit test if the sample size is large enough).

## Chapter 10

1. (a)



- (b)  $Z = (Y - \text{mean}) / (\text{standard dev}) = (-0.05 - 0.037) / (0.385) = -0.226$
- (c)  $\Pr[Z < -0.226] = \Pr[Z > 0.226]$  because the standard normal distribution is symmetrical around zero. (d) From Statistical Table B, 0.40905. (e) Same as (d): 0.397. (f). Same as (e): 0.397.

2. (a)  $H_0$ : The proportion of people who are struck by lightning who are men is 0.5. (b) The mean should be  $np = 648(0.5) = 324$ . (c) The standard deviation should be  $\sqrt{np(1-p)} = \sqrt{648(0.5)(0.5)} = 12.7$ . (d) 531 is above the mean value in the null hypothesis (324), so we subtract  $1/2$  to get 530.5 with the continuity correction. (e)  $Z = (530.5 - 324) / 12.7 = 16.2$ . (f)  $\Pr[\text{number of men} \geq 531] = \Pr[Z \geq 16.2]$ . Using Statistical Table A, this is off the chart, so we know that this probability is  $< 0.00002$ . (g) The two-tailed value is twice the probability found in the previous step, so  $P < 0.00004$ . (h) We conclude that the proportion of men struck by lightning is greater than 0.5.

$\Pr[Z > 2.15] = 0.0158$

0.910

insert minus sign before 0.32

3. From Statistical Table B. (a) 0.090. (b) ~~0.090~~.  
 (c)  $\Pr[Z > -2.15] = 1 - \Pr[Z < -2.15] = 1 - \Pr[Z > 2.15] = 1 - 0.016 = 0.984$ .  
 (d)  $\Pr[Z < 1.2] = 1 - \Pr[Z > 1.2] = 1 - 0.11507 = 0.885$ .  
 (e)  $\Pr[0.52 < Z < 2.34] = \Pr[0.52 < Z] - \Pr[2.34 < Z]$ , because the first value is the area under the curve from 0.52 to infinity, and the second value is the area under the curve from 2.34 to infinity. The difference will be the area under the curve from 0.52 to 2.34.  
 $0.302 - 0.010 = 0.292$ .  
 (f)  $\Pr[-2.34 < Z < -0.52] = 0.292$ : the normal distribution is symmetrical on either side of 0.  
 (g)  $\Pr[Z < -0.93] = \Pr[Z > 0.93] = 0.176$ .  
 (h)  $\Pr[-1.57 < Z < 0.32] = (1 - \Pr[Z > 1.57]) - \Pr[Z > 0.32] = (1 - 0.058) - 0.37 = 0.567$ .
4. (a) To determine the proportion of men excluded, we convert the height limit into standard normal deviates.  $(180.3 - 177.0)/7.1 = 0.46$ .  $\Pr[Z > 0.46] = 0.323$ , so roughly one-third of British men are excluded from applying.  
 (b)  $(172.7 - 163.3)/6.4 = 1.47$ .  $\Pr[Z > 1.47] = 0.071$ , which is the proportion of British women excluded.  $1 - 0.071 = 0.929$  is the proportion of acceptable to MI5.  
 (c)  $(163.3 - 163.3)/7.1 = 0.44$  standard deviate above the height limit.  
 (d)  $\Pr[Z < 0.44] = 0.667$  is the normal distribution.  $i$  is bimodal, while  $iii$  is skewed. (b) All three would generate approximately normal distributions of sample means due to the central limit theorem.
6. (a)  $\Pr[\text{weight} > 0.5 \text{ kg}]$ : Transform to standard normal deviate:  $(5 - 3.296)/0.560 = 3.04$ .  $\Pr[Z > 3.04] = 0.00118$ .  
 (b)  $\Pr[3 < \text{birth weight} < 4]$ : Transform both to standard normal deviates, and subtract probabilities of the Z-values.  $(3 - 3.296)/0.560 = -0.53$ .  $(4 - 3.296)/0.560 = 1.28$ .  $\Pr[Z < -0.53] = 0.29806$ ,  $\Pr[Z > 1.28] = 0.10027$ .  $1 - 0.29806 - 0.10027 = 0.602$ . (c) 0.06681 babies are more than 1.5 standard deviations above, with the same fraction below, so 0.13362 of babies are more than 1.5 standard deviations in either direction. (d) First, transform 1.5 kg into normal standard deviates:  $1.5/0.560 = 2.68$  standard deviations.  $\Pr[Z > 2.68] = 0.00368$ . Since the distribution is symmetric, we

- multiply this by two to reflect the probability of being 2.68 standard deviations above or below the mean:  $(0.00368)(2) = 0.00736$ . (e) The standard error is the same as the standard deviation of the mean. It is equal to the standard deviation divided by the square root of  $n$ , or  $0.560/\sqrt{10} = 0.18 \text{ kg}$ . To find the probability that the mean of a sample of 10 babies is greater than 3.5 kg, we transform this mean into a Z-score.  $(3.5 - 3.296)/0.18 = 1.13$ .  $\Pr[Z > 1.13] = 0.129$ .
7. (a) The lower graph has the higher mean (ca. 20 vs. 10), whereas the upper graph has the higher standard deviation. (b) The lower graph has the higher mean (ca. 15 vs. 10) and the higher standard deviation (ca. 5 vs. 2.5).
8. The standard deviation is approximately 10, as the region within one standard deviation from the mean will contain roughly 2/3 of the data points.
9. (a) In a normal distribution, the modal value occurs at the mean, so the mode is 35 mm.  
 (b) A normal distribution is symmetric, so the middle data point is the mean, 35 mm.  
 (c) Twenty percent of the distribution is less than 20 mm in size. (Why? Normal distributions are symmetric, so if 20% of the distribution is 15 mm or larger than the mean, 20% must be 15 mm or smaller than the mean.)
10. (a) The lower distribution (ii) would have sample means that had a more normal distribution, because the initial distribution is closer to normal. Both distributions would converge to a normal distribution if the sample size were sufficiently large.  
 (b) The distribution of the sums of samples from a distribution will be normally distributed, given a sufficiently large number of samples.
11. (a)  $\Pr[Y \geq 180] = \Pr[Z > (179.5 - np)/\sqrt{np(1 - p)}] = \Pr[Z > (179.5 - (400)(0.4))/\sqrt{(400)(0.4)(1 - 0.4)}] = \Pr[Z > 1.99] = 0.023$ .  
 (b)  $\Pr[Y \geq 130] = \Pr[Z > (129.5 - (400)(0.4))/\sqrt{(400)(0.4)(1 - 0.4)}] = \Pr[Z > -3.11] = 1 - \Pr[Z > 3.11] = 0.999$ .  
 (c)  $\Pr[155 \leq Y \leq 170] = \Pr[Y \geq 155] - \Pr[Y \geq 171]$ .  $\Pr[Y \geq 155] = \Pr[Z > (154.5 - (400)(0.4))/\sqrt{(400)(0.4)(1 - 0.4)}] = \Pr[Z > -0.56] = 1 - \Pr[Z > 0.56] = 0.712$

$\Pr[Z > 0.32] - \Pr[Z > 1.57] = 0.37448 - 0.05821 = 0.316$

(0.713 if calculated with a computer).  $\Pr[Y \geq 171] = \Pr[Z > (170.5 - (400)(0.4))/\sqrt{(400)(0.4)(1 - 0.4)}] =$

$\Pr[Z > 1.07] = 0.142$ .  $\Pr[155 \leq Y \leq 170] = 0.69497 - 0.13136 = 0.570$ .

12. Average expected number of cancer victims  $= 0.14 \times 220 = 30.8$ . Standard deviation  $= \sqrt{np(1 - p)} = \sqrt{220(0.14)(0.86)} = 5.15$ . Convert actual number of victims, 91, into standard normal deviate:  $(91 - \frac{1}{2} - 30.8)/5.15 = 11.60$ . This is off the chart:  $P < 0.00002$ .

13.  $SE = s/\sqrt{n}$ .  $Z = (Y - \text{mean})/SE$ .

Mean	SD	Value	SE <sub>10</sub>	Z <sub>10</sub>	Pr( $\bar{Y} > \text{Value}$ )	SE <sub>30</sub>	Z <sub>30</sub>	Pr( $\bar{Y} > \text{Value}$ )
14	5	15	1.58	0.63	0.264	0.91	1.10	0.136
15	3	15.5	0.95	0.53	0.298	0.54	0.91	0.181
-23	4	-22	1.26	0.79	0.215	0.73	1.37	0.085
72	50	45	15.81	-1.71	0.956	9.18	-2.96	0.998

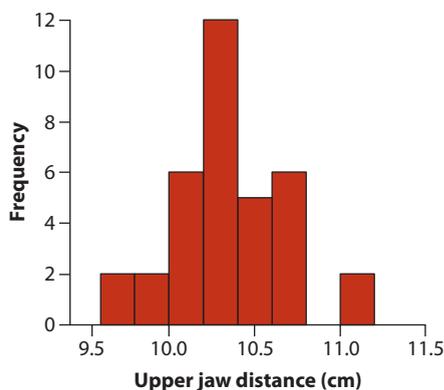
## Chapter 11

1. (a)  $n = 31$ . (b)  $\bar{Y} = 39.329$  m.  
 (c)  $s = 30.663$  m.  
 (d)  $SE_{\bar{Y}} = s/\sqrt{n} = 30.633/5.567 = 5.507$ .  
 (e)  $df = n - 1 = 30$ . (f)  $\alpha = 0.05$ .  
 (g)  $t_{0.05(2),30} = 2.04$ . (h) Variable has a normal distribution in the population, and sample is random. (i)  $\bar{Y} \pm t_{0.05(2),30}s$ ,  $28.09 < \mu < 50.56$ .  
 (j)  $H_0$ : The mean elevation change is 0 ( $\mu = 0$ ).  
 $H_A$ : The mean elevation change is not 0 ( $\mu \neq 0$ ). (k)  $t = (\bar{Y} - 0)/SE_{\bar{Y}} = 7.141$ .  
 (l) Variable has a normal distribution in the population, and sample is random. (m)  $t_{0.0001(2),30} = 4.48$ . Since  $t > 4.48$ ,  $P < 0.0001$  (on the computer,  $P = 6.06 \times 10^{-8}$ ). (n) Yes, the average elevational range shifted upward.
2. (a) The population variance. (b) Variable has a normal distribution in the population and sample is random. (c)  $s^2 = 940.23$ . (d)  $df = 30$ .  
 (e)  $\alpha = 0.05$ . (f)  $\chi_{0.025,30}^2 = 46.98$ .  
 (g)  $\chi_{0.975,30}^2 = 16.79$ .  
 (h)  $df s^2 / \chi_{0.025,30}^2 < \sigma^2 < df s^2 / \chi_{0.975,30}^2$ .  
 $600.4 < \sigma^2 < 1680.0$ .
3. (a)  $t_{0.05(2),11} = 2.20$ . (b)  $t_{0.05(2),31} = 2.04$ .  
 (c)  $t_{0.01(2),100} = 2.63$  (is the same as  $t_{0.01(2),120}$ ).  
 (d)  $t_{0.05(2),22} = 2.07$ . (e)  $t_{0.01(2),7} = 3.50$ .
4. (a)  $\chi_{0.025,11}^2 = 21.92$ .  $\chi_{0.975,11}^2 = 3.82$ .  
 (b)  $\chi_{0.025,31}^2 = 48.23$ .  $\chi_{0.975,31}^2 = 17.54$ .

2.04

Change  $s$  to  $SE_{\bar{Y}}$  (with a bar over the  $Y$ )

- (c)  $\chi_{0.005,100}^2 = 140$   
 (d)  $\chi_{0.025,22}^2 = 36.78$   
 (e)  $\chi_{0.005,7}^2 = 20.28$ .  $\chi_{0.995,7}^2 = 0.989$ .
5. The 99% confidence interval must be larger than the 95% confidence interval in order to have a higher probability of capturing the true mean.
6. (a) Answers will vary.

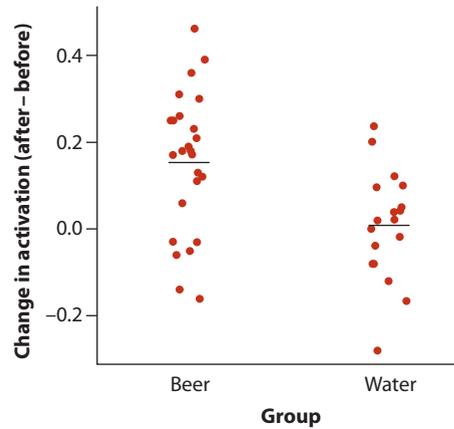


- (b) The sample mean is 10.32 cm and the standard error of the mean is 0.056 cm. The standard error estimates the standard deviation of the sampling distribution of sample means. (c) The 95% confidence interval for the mean is  $10.32 \pm (0.056 t_{0.05(2),34})$ .  $t = 2.03$ , so the 95% confidence interval is  $10.21 \text{ cm} < \mu < 10.44 \text{ cm}$ . (d) The variance is 0.11005. There are 35 individuals, so 34  $df$ . The  $\chi^2$  critical values are looked up in Statistical Table

was higher for *Pteronotus*, which is in the opposite direction as predicted by the model.

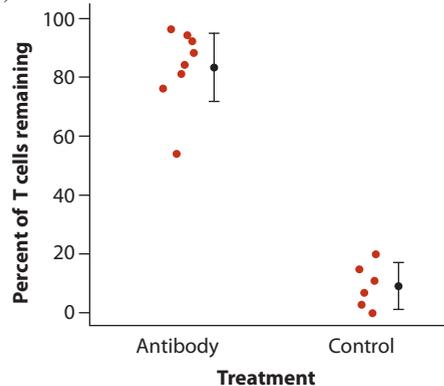
12. (a) We can use a paired  $t$ -test in this case, because the body temperature measurements are taken on the same individuals as the brain temperature measurements. To do this, we calculate the difference in temperature between brains and bodies for each ostrich, find the mean difference ( $0.648^\circ\text{C}$ ) and the standard error of the difference ( $0.116^\circ\text{C}$ ).  $H_0$ : Mean difference between brain and body temperature is zero.  $H_A$ : Mean difference is not zero.  $t = 0.648/0.116 = 5.6$  with 5  $df$ , which is greater than the critical value for  $\alpha(2) = 0.05$ , 2.57, so  $P < 0.05$ . We reject the null hypothesis of no difference between brain and body temperature ( $P = 0.0025$ ). (b) While our test is significant, the deviation is the opposite of that predicted from observations of mammals in similar environments: brains are hotter than bodies in ostriches, not cooler.
13. (a) *ii.* (b) *ii.* (c) From *ii.*, we can still be fairly confident that the groups are different, but we need to mentally double the size of the error bars to make this determination. (d) With sample sizes of 100, the standard errors will be one-tenth as great as the standard deviations. Graphs *i*, *ii*, and *iii* will be significantly different.
14. (a) Use a 95% confidence interval for the difference:  $-10.68 < \mu_d < -0.94$  mm. This assumes that the distribution of change in length in the population is normal and that the iguanas are a random sample. (b) All values within the most-plausible range are negative and include, e.g.,  $-1$  mm,  $-10$  mm, etc. The data are not consistent with no change or an increase in the mean. (c)  $16.6 < \sigma_d < 23.6$  mm. (d)  $H_0$ : The mean change in length is zero ( $\mu_d = 0$ ).  $H_A$ : The mean change in length is not zero.  $t = -2.38$ ,  $df = 63$ ,  $t_{0.05(2), 63} = 2.00$ . Reject  $H_0$  ( $P = 0.020$ ). Iguanas really did shrink.
15.  $H_0$ : Mean male aggressiveness is the same in the two groups.  $H_A$ : Mean aggressiveness is different. Two-sample  $t$ -test:  $s_p^2 = 0.498$ ,  $SE_{\bar{y}_1 - \bar{y}_2} = 0.252$ ,  $t = 9.85$ ,  $df = 44$ ,  $t_{0.05(2), 44} = 2.02$ . Since  $t > 2.02$ ,  $P < 0.05$ . Reject  $H_0$  ( $P < 10^{-10}$ ). Males are more aggressive when mated with a neighboring female.
16. (a) Strip chart, box plot, and multiple histograms. The strip chart is shown below. There is a trend for the activation to be greater in the beer

group. The beer group also appears to be more variable.



- (b) The standard deviations of the two groups are not hugely different (0.162 vs. 0.126) so ordinary two-sample  $t$ -test OK.  $H_0$ : Mean changes for beer and water are not different ( $\mu_1 - \mu_2 = 0$ ).  $H_A$ : Mean changes for beer and water are different ( $\mu_1 - \mu_2 \neq 0$ ).  $\bar{Y}_1 = 0.1544$ ,  $\bar{Y}_2 = 0.0081$ ,  $\bar{Y}_1 - \bar{Y}_2 = 0.1463$ ,  $n_1 = 25$ ,  $n_2 = 18$ ,  $s_1^2 = 0.02633$ ,  $s_2^2 = 0.01591$ ,  $s_p^2 = 0.02201$ ,  $SE_{\bar{y}_1 - \bar{y}_2} = 0.04586$ ,  $t = 3.191$ ,  $df = 41$ ,  $t_{0.05(2), 41} = 2.02$ . Since  $3.191 > 2.02$ ,  $P < 0.05$ . Reject  $H_0$  ( $P = 0.0027$ ).

17. (a)



- (b) Antibody mice:  $\bar{Y}_1 = 83.125$ ,  $s_1 = 13.58$ ,  $SE_{\bar{y}_1} = 4.80$ ,  $df_1 = 7$ ,  $t_{0.05(2), 7} = 2.36$ ,  $71.8 < \mu_1 < 94.5$ . Control mice:  $\bar{Y}_2 = 9.333$ ,  $s_2 = 7.50$ ,  $SE_{\bar{y}_2} = 3.06$ ,  $df_2 = 5$ ,  $t_{0.05(2), 5} = 2.57$ ,  $1.46 < \mu_2 < 17.20$ . (c) Added to the plot in part (a). (d) The null hypothesis would be