rounded up to the nearest integer. Thus, the desired sample size is

$$n = \frac{1068}{1 + 1067/3000} = 788,$$

rounded up to the nearest integer. ■

Exercises

7.4-1. Let X equal the tarsus length for a male grackle. Assume that the distribution of X is $N(\mu, 4.84)$. Find the sample size n that is needed so that we are 95% confident that the maximum error of the estimate of μ is 0.4.

7.4-2. Let X equal the excess weight of soap in a "1000-gram" bottle. Assume that the distribution of X is $N(\mu, 169)$. What sample size is required so that we have 95% confidence that the maximum error of the estimate of μ is 1.5?

7.4-3. A company packages powdered soap in "6-pound" boxes. The sample mean and standard deviation of the soap in these boxes are currently 6.09 pounds and 0.02 pound, respectively. If the mean fill can be lowered by 0.01 pound, \$14,000 would be saved per year. Adjustments were made in the filling equipment, but it can be assumed that the standard deviation remains unchanged.

(a) How large a sample is needed so that the maximum error of the estimate of the new μ is $\varepsilon = 0.001$ with 90% confidence?

(b) A random sample of size $n = 1219$ yielded $\bar{x} = 6.048$ and $s = 0.022$. Calculate a 90% confidence interval for μ.

(c) Estimate the savings per year with these new adjustments.

(d) Estimate the proportion of boxes that will now weigh less than 6 pounds.

7.4-4. Measurements of the length in centimeters of $n = 29$ fish yielded an average length of $\bar{x} = 16.82$ and $s^2 = 34.9$. Determine the size of a new sample so that $\bar{x} \pm 0.5$ is an approximate 95% confidence interval for μ.

7.4-5. A quality engineer wanted to be 98% confident that the maximum error of the estimate of the mean strength, μ, of the left hinge on a vanity cover molded by a machine is 0.25. A preliminary sample of size $n = 32$ parts yielded a sample mean of $\bar{x} = 35.68$ and a standard deviation of $s = 1.723$.

(a) How large a sample is required?

(b) Does this seem to be a reasonable sample size? (Note that destructive testing is needed to obtain the data.)

7.4-6. A manufacturer sells a light bulb that has a mean life of 1450 hours with a standard deviation of 33.7 hours. A new manufacturing process is being tested, and there is interest in knowing the mean life μ of the new bulbs. How large a sample is required so that $\bar{x} \pm 5$ is a 95% confidence interval for μ? You may assume that the change in the standard deviation is minimal.

7.4-7. For a public opinion poll for a close presidential election, let p denote the proportion of voters who favor candidate A. How large a sample should be taken if we want the maximum error of the estimate of p to be equal to

(a) 0.03 with 95% confidence?

(b) 0.02 with 95% confidence?

(c) 0.03 with 90% confidence?

7.4-8. Some college professors and students examined 137 Canadian geese for patent schistosome in the year they hatched. Of these 137 birds, 54 were infected. The professors and students were interested in estimating p, the proportion of infected birds of this type. For future studies, determine the sample size n so that the estimate of p is within $\varepsilon = 0.04$ of the unknown p with 90% confidence.

7.4-9. A die has been loaded to change the probability of rolling a 6. In order to estimate p, the new probability of rolling a 6, how many times must the die be rolled so that we are 99% confident that the maximum error of the estimate of p is $\varepsilon = 0.02$?

7.4-10. A seed distributor claims that 80% of its beet seeds will germinate. How many seeds must be tested for germination in order to estimate p, the true proportion that will germinate, so that the maximum error of the estimate is $\varepsilon = 0.03$ with 90% confidence?

7.4-11. Some dentists were interested in studying the fusion of embryonic rat palates by a standard transplantation technique. When no treatment is used, the probability of fusion equals approximately 0.89. The dentists would like to estimate p, the probability of fusion, when vitamin A is lacking.

(a) How large a sample n of rat embryos is needed for $y/n \pm 0.10$ to be a 95% confidence interval for p?

the cases where there were no failures. In fact, there were 17 previous launches in which no failures occurred. A scatter plot of the number of distressed O-rings per launch against temperature using data from all previous shuttle launches is given in Figure 6.5-3(b).

It is difficult to look at these data and not see a relationship between failures and temperature. Moreover, one recognizes that an extrapolation is required and that an inference about the number of failures outside the observed range of temperature is needed. The actual temperature at launch was 31°F, while the lowest temperature recorded at a previous launch was 53°F. It is always very dangerous to extrapolate inferences to a region for which one does not have data. If NASA officials had looked at this plot, certainly the launch would have been delayed. This example shows why it is important to have statistically minded engineers involved in important decisions.

These comments raise two interesting points: (1) It is important to produce a scatter plot of one variable against another. (2) It is also important to plot *relevant data*. Yes, it is true that some data were used in making the decision to launch the *Challenger*. But not all the relevant data were utilized. To make good decisions, it takes knowledge of statistics as well as subject knowledge, common sense, and an ability to question the relevance of information. ■

Exercises

6.5-1. Show that the residuals, $Y_i - \widehat{Y}_i$ $(i = 1, 2, \ldots, n)$, from the least squares fit of the simple linear regression model sum to zero.

6.5-2. In some situations where the regression model is useful, it is known that the mean of Y when $X = 0$ is equal to 0, i.e., $Y_i = \beta x_i + \varepsilon_i$ where ε_i for $i = 1, 2, \ldots, n$ are independent and $N(0, \sigma^2)$.

(a) Obtain the maximum likelihood estimators, $\widehat{\beta}$ and $\widehat{\sigma^2}$, of β and σ^2 under this model.

(b) Find the distributions of $\widehat{\beta}$ and $\widehat{\sigma^2}$. (You may use, without proof, the fact that $\widehat{\beta}$ and $\widehat{\sigma^2}$ are independent, together with Theorem 9.3-1.)

6.5-3. The midterm and final exam scores of 10 students in a statistics course are tabulated as shown.

(a) Calculate the least squares regression line for these data.

(b) Plot the points and the least squares regression line on the same graph.

(c) Find the value of $\widehat{\sigma^2}$.

Midterm	Final	Midterm	Final
70	87	67	73
74	79	70	83
80	88	64	79
84	98	74	91
80	96	82	94

6.5-4. The final grade in a calculus course was predicted on the basis of the student's high school grade point average in mathematics, Scholastic Aptitude Test (SAT) score in mathematics, and score on a mathematics entrance examination. The predicted grades x and the earned grades y for 10 students are given (2.0 represents a C, 2.3 a C+, 2.7 a B–, etc.).

(a) Calculate the least squares regression line for these data.

(b) Plot the points and the least squares regression line on the same graph.

(c) Find the value of $\widehat{\sigma^2}$.

x	y	x	y
2.0	1.3	2.7	3.0
3.3	3.3	4.0	4.0
3.7	3.3	3.7	3.0
2.0	2.0	3.0	2.7
2.3	1.7	2.3	3.0

6.5-5. A student who considered himself to be a "car guy" was interested in how the horsepower and weight of a car affected the time that it takes the car to go from 0 to 60 mph. The following table gives, for each of 14 cars, the horsepower, the time in seconds to go from 0 to 60 mph, and the weight in pounds:

Horsepower	0–60	Weight	Horsepower	0–60	Weight
230	8.1	3516	282	6.2	3627
225	7.8	3690	300	6.4	3892
375	4.7	2976	220	7.7	3377
322	6.6	4215	250	7.0	3625
190	8.4	3761	315	5.3	3230
150	8.4	2940	200	6.2	2657
178	7.2	2818	300	5.5	3518

(a) Calculate the least squares regression line for "0–60" versus horsepower.

(b) Plot the points and the least squares regression line on the same graph.

(c) Calculate the least squares regression line for "0–60" versus weight.

(d) Plot the points and the least squares regression line on the same graph.

(e) Which of the two variables, horsepower or weight, has the most effect on the "0–60" time?

6.5-6. Let x and y equal the ACT scores in social science and natural science, respectively, for a student who is applying for admission to a small liberal arts college. A sample of $n = 15$ such students yielded the following data:

x	y	x	y	x	y
32	28	30	27	26	32
23	25	17	23	16	22
23	24	20	30	21	28
23	32	17	18	24	31
26	31	18	18	30	26

(a) Calculate the least squares regression line for these data.

(b) Plot the points and the least squares regression line on the same graph.

(c) Find point estimates for α, β, and σ^2.

6.5-7. The Federal Trade Commission measured the number of milligrams of tar and carbon monoxide (CO) per cigarette for all domestic cigarettes. Let x and y equal the measurements of tar and CO, respectively, for 100-millimeter filtered and mentholated cigarettes. A sample of 12 brands yielded the following data:

Brand	x	y	Brand	x	y
Capri	9	6	Now	3	4
Carlton	4	6	Salem	17	18
Kent	14	14	Triumph	6	8
Kool Milds	12	12	True	7	8
Marlboro Lights	10	12	Vantage	8	13
Merit Ultras	5	7	Virginia Slims	15	13

(a) Calculate the least squares regression line for these data.

(b) Plot the points and the least squares regression line on the same graph.

(c) Find point estimates for α, β, and σ^2.

6.5-8. The data in the following table, part of a set of data collected by Ledolter and Hogg (see References), provide the number of miles per gallon (mpg) for city and highway driving of 2007 midsize-model cars, as well as the curb weight of the cars:

Type	mpg City	mpg Hwy	Curb Weight
Ford Fusion V6 SE	20	28	3230
Chevrolet Sebring Sedan Base	24	32	3287
Toyota Camry Solara SE	24	34	3240
Honda Accord Sedan	20	29	3344
Audi A6 3.2	21	29	3825
BMW 5-series 525i Sedan	20	29	3450
Chrysler PT Cruiser Base	22	29	3076
Mercedes E-Class E350 Sedan	19	26	3740
Volkswagen Passat Sedan 2.0T	23	32	3305
Nissan Altima 2.5	26	35	3055
Kia Optima LX	24	34	3142

(a) Find the least squares regression line for highway mpg (y) and city mpg (x).

(b) Plot the points and the least squares regression line on the same graph.

(c) Repeat parts (a) and (b) for the regression of highway mpg (y) on curb weight (x).

6.5-9. Using an Instron 4204, rectangular strips of Plexiglas® were stretched to failure in a tensile test. The following data give the change in length, in millimeters