Chapter 7
Data for Decisions

Chapter Objectives

Check off these skills when you feel that you have mastered them.

- ☐ Identify the population in a given sampling or experimental situation.
- ☐ Identify the sample in a given sampling or experimental situation.
- ☐ Explain the difference between a population and a sample.
- ☐ Analyze a sampling example to detect sources of bias.
- ☐ Identify several examples of sampling that occur in our society.
- ☐ Select a numbering scheme for a population from which a random sample will be selected and use a table of random digits to select that random sample.
- ☐ Explain the difference between an observational study and an experiment.
- ☐ Recognize the confounding on the effects of two variables in an experiment.
- ☐ Explain the difference between the experimental group and the control group in an experiment.
- ☐ Design a randomized comparative experiment and display it in graphical form.
- ☐ Explain what is meant by statistically significant.
- ☐ Describe the placebo effect.
- ☐ Discuss why double blindness is desirable in an experiment.
- ☐ Define statistical inference.
- ☐ Explain the difference between a parameter and a statistic.
- ☐ Identify both the parameter and the statistic in a simple inferential setting.
- ☐ Compute the sample proportion when both the sample size and number of favorable responses are given.
- ☐ Using an appropriate formula, calculate the standard deviation of a given statistic.
- ☐ Explain the difference between the population mean and the sample mean.
- ☐ Given a sample proportion and sample size, list the range for a 95% confidence interval for the population proportion.
- ☐ Calculate differing margins of error for increasing sample sizes.
- ☐ Discuss the effect of an increased sample size on the statistic’s margin of error.

Guided Reading

Introduction

Numbers are used in a myriad of ways to describe the world we live in. Our social and economic concerns, science and ecology, politics, religion, health, recreation, any area of human activity is better understood by the collection and analysis of data. It is vitally important to know how to produce trustworthy data, and how to draw reliable conclusions from them. This is the role of statistics, the science of data handling.

🗝 Key idea

In Chapters 5 and 6, **data analysis** was explored. Graphs and numbers were produced to represent a set of data. In this chapter you will explore **how to produce data** that can be trusted for answering specific questions. Then in turn answering questions with a certain degree of confidence is **statistical inference**.

Section 7.1 Sampling

🗝 Key idea

In statistical studies, we gather information about a small, partial group (a **sample**) in order to draw conclusions about the whole, large group we are interested in (the **population**).

Example A

In a study of the smoking habits of urban American adults, we ask 500 people their age, place of residence, and how many cigarettes they smoke daily. What is the population and what is the sample?

Solution

The population is all Americans who live in a large city and are old enough to be classified as adults. The sample consists of those among the 500 interviewed who qualify as members of this population; for example, children would not be included, even if they smoke.

Section 7.2 Bad Sampling Methods

🗝 Key idea

Systematic error caused by bad sampling methods may lead to a **biased** study favoring certain outcomes.

Example B

Customers at a supermarket are sampled to determine their opinion about a volatile political issue. Can you identify a possible source of bias in such a survey?

Solution

There may be gender-based bias, with women overrepresented in the sample. Depending on the location of the market, there may also be a bias according to economic class, education level, political affiliation, etc.

🗝 Key idea

A sample of people who choose to respond to a general appeal is called a **voluntary response sample**. This voluntary response is a likely source of bias.

Example C
Television viewers are invited to call an 800 number to report their opposition to a bill to increase state gasoline taxes. Why might this survey be biased?

Solution
There is a high likelihood that a disproportionately large number of people angry about a potential tax increase will take the trouble to call to register their opposition.

Question 1

Consider the following.

a. convenience sample b. voluntary response sample c. bias

Which of the above expressions/word would **not** be used to fill in a blank for the following?

1) To determine the food preferences of students, a staff member surveys students as they exit a local bar. This type of sample is a _____.

2) A survey on the benefits of jogging is conducted outside a sporting-goods store. This is an example of _____.

Answer
b

Section 7.3 Simple Random Samples

Key idea
We can use a **simple random sample** (**SRS**) to eliminate bias. This is the equivalent of choosing names from a hat; each individual has an equal chance to be selected.

Example D
To choose a sample of five cards from a deck of 52 cards, you shuffle the cards and choose the first, third, fifth, seventh, and ninth card. Will this lead to a simple random sample?

Solution
Yes, if the deck has been thoroughly shuffled. After shuffling, any given card is equally likely to occupy any given position in the deck.

Key idea
A two-step procedure for forming a SRS using a table of random digits is:

Step 1: Give each member a numerical label of the same length.

Step 2: Read from the table strings of digits of the same length as the labels. Ignore groups not used as labels and also ignore any repeated labels.

Example E
Describe how to use the table of random digits to form a random sample of 75 students at Hypothetical University from the entire population of 1350 HU students.

Solution
Assign each HU student a four digit numerical label, 0001–1350, making sure that no label is assigned twice. Then starting anywhere in the random digit table, read strings of 4 consecutive digits, ignoring repetitions and unassigned strings, until 75 assigned labels are obtained. The students with those labels are the sample.

Example F

A teacher wants to randomly poll her students regarding whether they liked a certain project or not. There are 25 students in the class and he wants to poll five of them. Starting at line 105 use the partial table (Table 7.1 of your text) to find the five random students.

Adam	Faiz	Kevin	Patty	Victoria
Billy	Gwen	Leo	Quinn	Wally
Cassy	Heidi	Mary	Rachel	Xavier
Daniel	Iliana	Nadia	Sarah	Yaffa
Edwin	Jacob	Ottis	Thomas	Zeki

TABLE 7.1 Random Digits

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915

Solution

Step 1: Give each student a label of the same numerical length.

01 Adam	06 Faiz	11 Kevin	16 Patty	21 Victoria
02 Billy	07 Gwen	12 Leo	17 Quinn	22 Wally
03 Cassy	08 Heidi	13 Mary	18 Rachel	23 Xavier
04 Daniel	09 Iliana	14 Nadia	19 Sarah	24 Yaffa
05 Edwin	10 Jacob	15 Ottis	20 Thomas	25 Zeki

Step 2: Use the table starting on line 105 looking at groups of digits of length 2.

TABLE 7.1 Random Digits

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915

Thus, the students she would poll are 07 = Gwen, 19 = Sarah, 14 = Nadia, 17 = Quinn, and 13 = Mary.

Question 2

Redo Example F, but start on line 102. In alphabetical order, who would be the third student that the teacher would select to poll?

Answer

Quinn

Section 7.4 Cautions about Sample Surveys

Key idea

Even a sound statistical design cannot guard against some of the pitfalls associated with statistical experiments. For example, **nonresponse** can be a cause of bias in an experiment, as can the artificial environments created for some experiments and **undercoverage**, by not including in samples certain parts of the population. Responses can be strongly influenced by the **wording of questions**. By having leading questions or confusing questions, strong bias can be introduced.

Example G

The Highway Patrol in a state decides to estimate the average speed of drivers using the fast lane. Using their patrol cars, they get behind them and record their speed which would be the same as the car in front of them. Is there any bias?

Solution

Yes, there is bias. Most likely the drives will slow down upon seeing the patrol car. This will result in overall lower average speed.

Section 7.5 Experiments

Key idea

An **observational study** is a passive study of a variable of interest. The study does not attempt to influence the responses and is meant to *describe* a group or situation.

Key idea

An **experiment** is an active trial of an imposed *treatment* and its *effects*. The study is meant to observe whether the treatment causes a change in the response.

Example H

Which is an experiment and which is an observational study?

a) You ask a sample of smokers how many cigarettes they smoke daily, and measure their blood pressure.
b) You select a sample of smokers and measure their blood pressure. Then you ask them to reduce their smoking by 5 cigarettes a day; after 3 months you recheck their blood pressure.

Solution

a) This is an observational study. We are passively observing and measuring.
b) This is an experiment. We are actively influencing the behavior of the subjects.

🗝 Key idea

When designing an **uncontrolled study**, care must be taken to avoid **confounded variables**. Confounding variables are variables whose effects on the outcome cannot be distinguished from one another.

🗝 Key idea

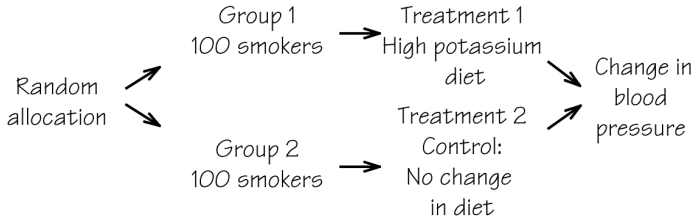
We can reduce the effect of confounded variables by conducting a **randomized comparative experiment**. The sample for the experiment is matched by a **control group**, with subjects assigned randomly to the treatment or the control group. Since personal choice can be a source of bias, the subjects should be randomly chosen for each group.

Example I

How would you design a simple randomized comparative experiment to test the effect of a high-potassium diet on smokers' blood pressure? Assume you have 200 smokers who have agreed to participate in the experiment.

Solution

From the group of 200 smokers who have agreed to participate, randomly select 100 to try the high-potassium diet. The other 100 will serve as the control group, and will make no change in their diet. Measure the blood pressure of each subject at the beginning and end of the testing period, and compare changes in the two groups.



🗝 Key idea

A well-designed experiment is one that uses the principles of **comparison** and **randomization**: comparison of several treatments and the random assignment of subjects to treatments.

🗝 Key idea

If subjects are randomly assigned to treatments, we can be confident that any differences among treatment groups that are too large to have occurred by chance are **statistically significant**. Small differences between groups in a study can be due to random variation, but statistically significant differences are too large to be attributable to chance and are reliable evidence of a real effect of the factors being studied.

Section 7.6 Thinking about Experiments

🗝 Key idea

The **placebo effect** is a special kind of confounding in which a patient responds favorably to any treatment, even a placebo (fake treatment).

🗝 Key idea

To avoid the placebo effect and any possible bias on the part of the experimenters, use a **double-blind** experiment, so neither subjects nor investigators know which treatment an individual is receiving.

Example J

How would you design a double-blind experiment to test the effect of a vitamin supplement on smokers' blood pressure?

Solution

Randomly assign labels 001 – 200 to your subjects. Using the labels, randomly choose 100 of the subjects to receive the vitamin supplement (say, in pill form), while the other group receives an indistinguishable placebo. The list of which group each subject belongs to is kept confidential; neither subjects nor experimenters know who is taking the real supplement until the experiment is over and the data have been recorded. This way, neither psychological factors nor unconscious bias on the part of the experimenters can play a role.

🗝 Key idea

A **prospective study** is an observational study that records slowly developing effects of a group of subjects over a long period of time.

🗝 Key idea

Only experimentation can produce fully convincing statistical evidence of **cause and effect**.

🗝 Key idea

Experiments like samples have weaknesses, in particular, they can **lack realism**. This would mean that it is hard to say exactly how far the results of the experiment can be applied.

Section 7.7 Inference: From Sample to Population

🗝 Key idea

Using a fact about a sample to estimate the truth about the whole population is called **statistical inference**. We are inferring conclusions about the whole population based on data from selected individuals. Statistical inference only works if the data comes from a random sample or a randomized comparative experiment. A sample should resemble the population, so that a sample statistic can be used to estimate a characteristic of the population.

🗝 Key idea

A **parameter** is a number that describes the **population**. A parameter is a fixed value, but we generally do not know what it is.

Key idea

A **statistic** is a number that described a **sample**. This value can change from sample to sample. A statistic is often used to estimate an unknown parameter.

Question 3

Consider the following.

a. statistic b. sample c. parameter

Which of the above expressions/word would **not** be used to fill in a blank for the following?

A random sample of 10 bags of flour has a mean weight of 24.9 pounds, less than the mean weight 25.05 pounds of all bags produced.

1) In this example. 25.05 is called a _____.
2) In this example, 24.9 is called a _____.

Answer

b

Key idea

If you have a simple random sample of size *n* from a large population and a count of success (such as agreeing with a survey question) in the same population then the **sample proportion of successes**, $\left(\hat{p}\right)$, is the following quotient.

$$\hat{p} = \frac{\text{count of successes in sample}}{n}$$

$\hat{p}$ is a statistic. The corresponding population proportion parameter is *p*.

Key idea

Results of a survey will vary from sample to sample. The margin of error given for a national sample indicates how close that result is to the truth. If the population result would fall in 95% of all samples drawn using the same method, we say we have 95% confidence that the truth about the population falls within this margin of error.

Example K

A random sample of 150 people are asked if they own dogs, and 58 of them say yes. What would you estimate the percentage of dog owners to be in the general population?

Solution

The sample proportion is $\hat{p} = \frac{58}{150} \approx 0.387 = 38.7\%,$ the actual population proportion may differ somewhat, but is reasonably likely to be fairly close to that of the sample. Thus, our best estimate is 38.7%.

Question 4

Suppose you conduct a telephone poll of 1250 people, asking them whether or not they favor mandatory sentencing for drug related crimes. If 580 people say "yes," what is the sample proportion $\hat{p}$ of people in favor of mandatory sentencing?

Answer

$\hat{p} = 46.4\%$

🗝 Key idea

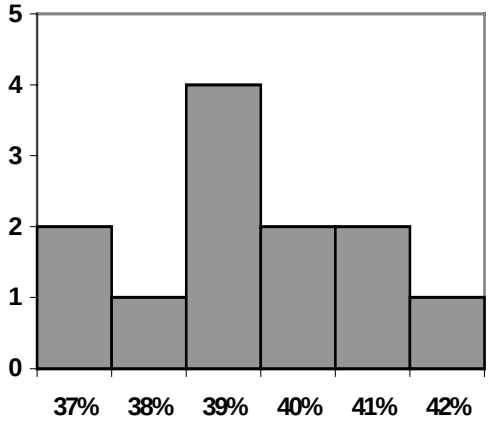
Statistical inference is based on the idea that one needs to see how trustworthy a procedure is if it is repeated many times. Results of a survey will vary from sample to sample; this is called sampling variability. So to answer the question as to what would happen for many samples, we do the following.

- Take a large number of random samples from the same population.
- Calculate $\hat{p}$ for each sample.
- Make a histogram of $\hat{p}$.
- Examine the distribution for shape, center, and spread, as well as outliers or other deviations.

Example L

Suppose the previous dog survey was conducted simultaneously by twelve investigators, each sampling 150 people, leading to the following twelve percentages of dog owners: {39%, 37%, 37%, 39%, 40%, 38%, 41%, 40%, 39%, 41%, 42%, 39%}. Sketch a histogram for this data and discuss the features of the distribution.

Solution



A sampling histogram will generally display a regular pattern with two important features: The results will be centered symmetrically around a peak, the true population value. The spread of the data will be tighter for large sample sizes, wider for small ones.

🗝 Key idea

The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population. This is the ideal pattern if we looked at all same-size possible samples of the population. For a simple random sample of size n from a large population that contains population proportion p the sampling distribution for $\hat{p}$ is approximately normal, with mean p, and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. The formula for the standard deviation of $\hat{p}$ shows that the spread of the sampling distribution is about the same for most sample proportions; it depends primarily on the sample size.

Example M

Suppose 24% of all college students think that textbook prices are reasonable. If you take a random sample of 2000 college students, what is the standard deviation of $\hat{p}$? Round to four decimal places.

Solution

Convert 24% to decimal form, 0.24. Standard deviation is as follows.

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.24(1-0.24)}{2000}} = \sqrt{\frac{0.24(0.76)}{2000}} \approx 0.0095$$

Question 5

Suppose that in the political poll from Question 4, the true population proportion is p = 45%. What is the standard deviation of the sampling distribution?

Answer

The standard deviation of $\hat{p}$ is approximately 0.0141.

Section 7.8 Confidence Intervals

Key idea

We cannot know precisely a true population parameter, such as the proportion p of people who favor a particular political candidate. To make an estimate, we interview a random sample of the population and calculate a statistic of the sample, such as the sample proportion, $\hat{p}$, favoring the candidate in question. Since $\hat{p}$ is close to normal in its distribution, we will consider the 95 part of the 68-95-99.7 rule which indicates that 95% of all samples of $\hat{p}$ will fall within two standard deviations of the true population proportion, p. This leads us to the **95% confidence interval for *p***, which is quite accurate for large values of n. It is as follows.

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Key idea

The **margin of error** of a survey gives an interval that includes 95% of the samples and is centered around the true population value. The margin of error is $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Example N

Suppose the results of the dog survey are announced as follows: “The percentage of people who own dogs is 39%, with a margin of error of 4%”. Can you say the following with reasonable (about 95%) certainty of being right?

a) At least (that is, not less than) 39% of people own dogs.
b) At most (that is, not more than) 45% of people own dogs.

Solution

a) No. The true percentage of dog owners is just as likely to be below 39% as above it.
b) Yes. With a 4% margin of error, the true percentage is almost certain to be within the 35%–43% range, and is thus highly likely to be less than 45%.

Example O

In a political poll, 695 potential voters are asked if they have decided yet which candidate they will vote for in the next election. Suppose that 511 say "yes."

a) Estimate the proportion of *undecided* voters.

b) Find a 95% confidence interval for this estimate.

Solution

a) The number of voters in the sample who have made up their minds is 511. Thus, the number of undecided voters in the sample is $695-511=184$. The sample proportion of voters is therefore as follows.

$$\hat{p}=\frac{184}{695}\approx 0.265=26.5\%$$

b) The 95% confidence interval for this estimate can be calculated as follows.

$$\hat{p}\pm 2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}=0.265\pm 2\sqrt{\frac{0.265\left(1-0.265\right)}{695}}=0.265\pm 2\sqrt{\frac{0.265\left(0.735\right)}{695}}\approx 0.265\pm 0.033$$

$$0.265-0.033=0.232=23.2\% \text{ to } 0.265+0.033=0.298=29.8\%$$

Rounding off we get an interval of (23%, 30%).

Question 6

In a college survey, 847 students were asked if they thought the cost of tuition ias reasonable. 521 said that they felt it was reasonable.

a) Estimate the proportion of students that believe the cost of tuition is reasonable.

b) Find a 95% confidence interval for this estimate.

Answer

a) 61.5%

b) (58.2%, 64.8%)

Homework Help

☞To assist you in your homework, a copy of Table 7.1 appears after this section.

Exercises 1 – 2
Carefully read Section 7.1 before responding to these exercises.

Exercises 3 – 6
Carefully read Section 7.2 before responding to these exercises. Your answers may differ from a classmate in terms of describing reasons for bias or giving examples. Try to imagine yourself in the situation described in the exercise before responding.

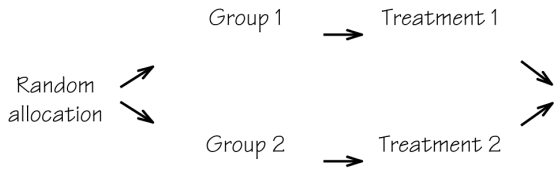
Exercises 7 – 10 & 13
Carefully read Section 7.3 before responding to these exercises. You may choose to make some copies of Table 7.1 to write on. Be very careful using the table by taking your time. Take the time to check your answer twice. If you write on the table in pencil and plan to reuse it, make sure to erase any stray marks so that they will not interfere with using the table again.

Exercises 11 – 12
Look carefully at Table 7.1 as you respond to these exercises.

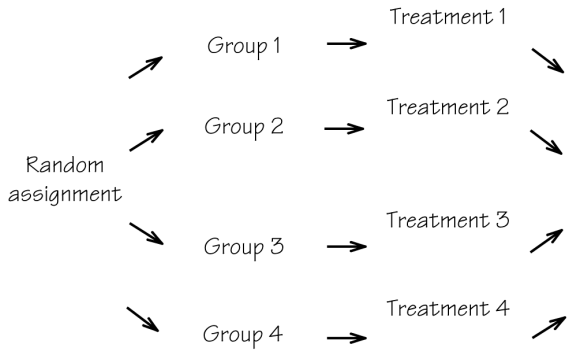
Exercise 14
Carefully read Section 7.3 before responding to this exercise. Pay particular attention to the definition of a simple random sample.

Exercises 15 – 17
Carefully read Section 7.4 before responding to these exercises. Read the scenario in each question carefully and try to imagine yourself in the situation described in an exercise before responding.

Exercises 18 – 27 & 29 – 30
Carefully read Section 7.5 before responding to these exercises. You will need Table 7.1 for Exercises 23, 24, 26, 27, 29, and 30. For Exercises 22, 23, 24, and 29, you will need to draw using the following as a template.

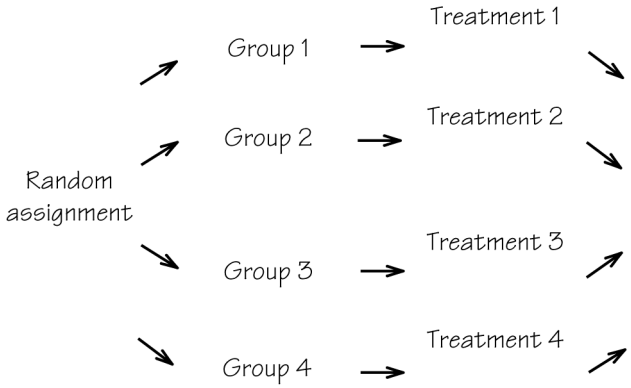


For Exercise 25, you will need to draw using the following as a template.



For Exercise 27, you will need to draw using the following as a template.

Exercises 28 & 31 – 35
Carefully read Section 7.6 before responding to these exercises. You will need Table 7.1 for Exercise 31. For Exercise 31, you will need to draw using the following as a template.

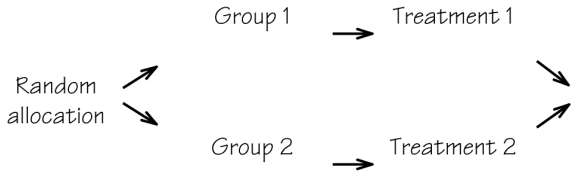


Exercises 36 – 41
Carefully read Section 7.7 before responding to these exercises. Look carefully at Table 7.1 as you respond to Exercise 41. You will need a calculator with the square root feature for Exercises 38 – 40. Make sure you know the requirements as to how much work should be shown (including steps for rounding) for your homework.

Exercises 42 – 53
Carefully read Section 7.8 before responding to these exercises. You will need a calculator with the square root feature for Exercises 42 – 46. Make sure you know the requirements as to how much work should be shown (including steps for rounding) for your homework. In Exercise 52, if we let E be the margin of error, then we have $E = 2\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}}$. Consider what happens to this formula when you need to find $\dfrac{E}{2}$.

Exercise 54

Carefully read Section 7.5 before responding to this exercise. You will need to draw using the following as a template.



Exercise 55

Carefully read Section 7.8 before responding to this exercise.

Exercise 56

Carefully read Sections 7.3 and 7.4 before responding to this exercise. You will need Table 7.1 for this exercise.

Exercise 57

Carefully read Section 7.7 before responding to this exercise. Also, recall the 68-95-99.7 rule as you work this exercise.

TABLE 7.1 Random Digits

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915
115	61041	77684	94322	24709	73698	14526	31893	32592
116	14459	26056	31424	80371	65103	62253	50490	61181
117	38167	98532	62183	70632	23417	26185	41448	75532
118	73190	32533	04470	29669	84407	90785	65956	86382
119	95857	07118	87664	92099	58806	66979	98624	84826
120	35476	55972	39421	65850	04266	35435	43742	11937
121	71487	09984	29077	14863	61683	47052	62224	51025
122	13873	81598	95052	90908	73592	75186	87136	95761
123	54580	81507	27102	56027	55892	33063	41842	81868
124	71035	09001	43367	49497	72719	96758	27611	91596
125	96746	12149	37823	71868	18442	35119	62103	39244
126	96927	19931	36809	74192	77567	88741	48409	41903
127	43909	99477	25330	64359	40085	16925	85117	36071
128	15689	14227	06565	14374	13352	49367	81982	87209
129	36759	58984	68288	22913	18638	54303	00795	08727
130	69051	64817	87174	09517	84534	06489	87201	97245
131	05007	16632	81194	14873	04197	85576	45195	96565
132	68732	55259	84292	08796	43165	93739	31685	97150
133	45740	41807	65561	33302	07051	93623	18132	09547
134	27816	78416	18329	21337	35213	37741	04312	68508
135	66925	55658	39100	78458	11206	19876	87151	31260
136	08421	44753	77377	28744	75592	08563	79140	92454
137	53645	66812	61421	47836	12609	15373	98481	14592
138	66831	68908	40772	21558	47781	33586	79177	06928
139	55588	99404	70708	41098	43563	56934	48394	51719
140	12975	13258	13048	45144	72321	81940	00360	02428
141	96767	35964	23822	96012	94591	65194	50842	53372
142	72829	50232	97892	63408	77919	44575	24870	04178
143	88565	42628	17797	49376	61762	16953	88604	12724
144	62964	88145	83083	69453	46109	59505	69680	00900
145	19687	12633	57857	95806	09931	02150	43163	58636
146	37609	59057	66967	83401	60705	02384	90597	93600
147	54973	86278	88737	74351	47500	84552	19909	67181
148	00694	05977	19664	65441	20903	62371	22725	53340
149	71546	05233	53946	68743	72460	27601	45403	88692
150	07511	88915	41267	16853	84569	79367	32337	03316

Do You Know the Terms?

Cut out the following 25 flashcards to test yourself on Review Vocabulary. You can also find these flashcards at http://www.whfreeman.com/fapp7e.

Chapter 7 **Data for Decisions** **Bias**	**Chapter 7** **Data for Decisions** **95% confidence interval**
Chapter 7 **Data for Decisions** **Confounding**	**Chapter 7** **Data for Decisions** **Control group**
Chapter 7 **Data for Decisions** **Convenience sample**	**Chapter 7** **Data for Decisions** **Double-blind experiment**
Chapter 7 **Data for Decisions** **Experiment**	**Chapter 7** **Data for Decisions** **Margin of error**
Chapter 7 **Data for Decisions** **Nonresponse**	

An interval computed from a sample by a method that captures the unknown parameter in 95% of all possible samples. When we calculate the interval for a single sample, we are 95% confident that the interval captures the unknown parameter.	**A systematic error that tends to cause the observations to deviate in the same direction from the truth about the population whenever a sample or experiment is repeated.**
A group of experimental subjects who are given a standard treatment or no treatment (such as a placebo).	**Two variables are confounded when their effects on the outcome of a study cannot be distinguished from each other.**
An experiment in which neither the experimental subjects nor the persons who interact with them know which treatment each subject received.	**A sample that consists of the individuals who are most easily available, such as people passing by in the street. A convenience sample is usually biased.**
As announced by most national polls, the margin of error says how close to the truth about the population the sample result would fall in 95% of all samples drawn by the method used to draw this one sample.	**A study in which treatments are applied to people, animals, or things in order to observe the effect of the treatments.**
	Some individuals chosen for a sample cannot be contacted or refuse to participate.

Chapter 7 **Data for Decisions** **Observational study**	**Chapter 7** **Data for Decisions** **Parameter**
Chapter 7 **Data for Decisions** **Placebo effect**	**Chapter 7** **Data for Decisions** **Population**
Chapter 7 **Data for Decisions** **Prospective study**	**Chapter 7** **Data for Decisions** **Randomized comparative experiment**
Chapter 7 **Data for Decisions** **Sample**	**Chapter 7** **Data for Decisions** **Sample proportion**

A number that describes the population. In statistical inference, the goal is often to estimate an unknown parameter or make a decision about its value.	**A study (such as a sample survey) that observes individuals and measures variables of interest but does not attempt to influence the responses.**
The entire group of people or things that we want information about.	**The effect of a dummy treatment (such as an inert pill in a medical experiment) on the response of subjects.**
An experiment to compare two or more treatments in which people, animals, or things are assigned to treatments by chance.	**An observational study that follows two or more groups of subjects forward in time.**
The proportion $\hat{p}$ of the members of a sample having some characteristic (such as agreeing with an opinion poll question). The sample proportion from a simple random sample is used to estimate the corresponding proportion pin the population from which the sample was drawn.	**A part of the population that is actually observed and used to draw conclusions, or inferences, about the entire population.**

Chapter 7 **Data for Decisions** **Sampling distribution**	**Chapter 7** **Data for Decisions** **Simple random sample (SRS)**
Chapter 7 **Data for Decisions** **Statistic**	**Chapter 7** **Data for Decisions** **Statistical inference**
Chapter 7 **Data for Decisions** **Statistical significance**	**Chapter 7** **Data for Decisions** **Table of random digits**
Chapter 7 **Data for Decisions** **Undercoverage**	**Chapter 7** **Data for Decisions** **Voluntary response sample**

A sample chosen by chance, so that every possible sample of the same size has an equal chance to be the one selected.	**The distribution of values taken by a statistic when all possible random samples of the same size are drawn from the same population. The sampling distributions of sample proportions are approximately normal.**
Methods for drawing conclusions about an entire population on the basis of data from a sample. Confidence intervals are one type of this method.	**A number that describes a sample. A statistic can be calculated from the sample data alone; it does not involve any unknown parameters of the population.**
A table whose entries are the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 in a completely random order. That is, each entry is equally likely to be any of the 10 digits and no entry gives information about any other entry.	**An observed effect is statistically significant if it is so large that it is unlikely to occur just by chance in the absence of a real effect in the population from which the data were drawn.**
A sample of people who choose themselves by responding to a general invitation to give their opinions. Such a sample is usually strongly biased.	**The process of choosing a sample may systematically leave out some groups in the population, such as households without a television.**

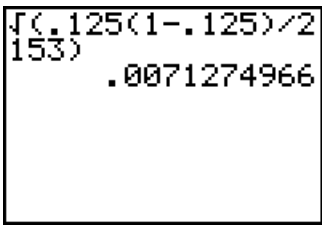
Learning the Calculator

Example 1

Calculate the standard deviation of $\hat{p}$ given that $p = 0.125$ and $n = 2153$.

Solution

Since $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.125(1-0.125)}{2153}}$, we can enter the following into the calculator.



Thus, the standard deviation is approximately 0.007.

Example 2

In a political poll, 10134 potential voters are asked if they have decided yet which candidate they will vote for in the next election. Suppose that 6215 say "yes."

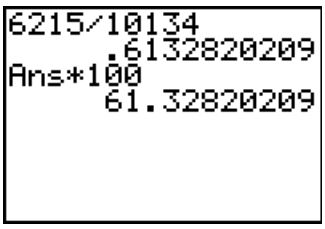
a) Estimate the proportion of decided voters.

b) Find a 95% confidence interval for this estimate.

Solution

a) The sample proportion of voters is therefore as follows.

$$\hat{p} = \frac{6215}{10{,}134}$$



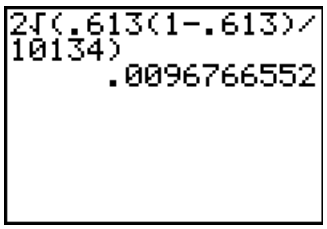
$\hat{p}$ would be approximately $0.613 = 61.3\%$

Continued on next page

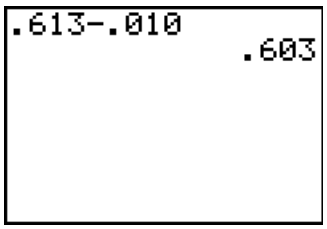
b) The 95% confidence interval for this estimate can be calculated as follows.

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 0.613 \pm 2\sqrt{\frac{0.613\left(1-0.613\right)}{10{,}134}}$$

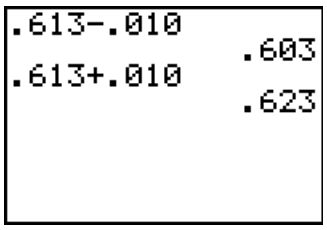
Calculate $2\sqrt{\frac{0.613\left(1-0.613\right)}{10{,}134}}$ first.



The 95% confidence interval for this estimate is 0.613 ± 0.010. First calculate $0.613 - 0.010$.



To calculate $0.613 + 0.010$, you can save some keystrokes by pressing 2nd then ENTER and then edit by using the ◄. Change the – to a +. Then press ENTER.



The 95% confidence interval for this estimate is $(60.3\%, 62.3\%)$.

Practice Quiz

1. A marketing firm interviewed 80 shoppers randomly selected from the 4000 customers at one of the mall's 45 stores yesterday. The sample in this situation is the

a. 45 stores.

b. 80 selected shoppers.

c. 4000 customers.

2. In an election for mayor, there are 3 candidates and 24,000 eligible voters. A newspaper interviews 240 voters as they come out of the polls. The population here is the

a. 240 voters interviewed.

b. 3 candidates.

c. 24,000 eligible voters.

3. A well-designed survey should minimize

a. bias.

b. randomness.

c. the placebo effect.

4. Here is a list of random numbers: 16807 64853 17463 14715. Use this list to choose 3 numbers from the set {1, 2, 3, . . . , 20}. What are the numbers?

a. 16, 7, 17

b. 16, 17, 15

c. 16, 17, 14

5. If Adam's sample statistic has a margin of error of 3% and Nadia's sample statistic has a margin of error of 5%, then

a. Nadia's estimate is biased.

b. Adam has more samples.

c. Nadia's experiment gave a higher sample estimate.

6. Five workers each sample 50 students to determine their favorite fast-food restaurant. Each worker returns with slightly different results. This is probably due to

a. bias.

b. sampling variability.

c. the use of a control group.

7. A drug test randomly selects one of three treatments for each participant. Neither the experimenter nor the participant knows which drug is chosen. This is an example of

a. a randomized comparative experiment.

b. an experiment which is not double-blind.

c. bad sampling methods.

8. To determine interest in a new paper towel, samples are mailed to 300 local residents, of which 120 prefer it to their current brand. What is the standard deviation of the sampling distribution of this statistic?

a. 2.83%

b. 3.65%

c. 4.47%

9. A random poll of 600 people shows that 60% of those polled are in favor of a new school building. Find a 95% confidence interval for the proportion of the residents in favor of a new school building.

a. 56% to 64%

b. 40% to 80%

c. 58% to 62%

10. An election poll of 628 voters showed that 412 of them approve of the policies of the leader. A 95% confidence interval for the proportion of voters who approve of the policies of the leader is _____.

a. 61.8% to 69.4%

b. 63.7% to 67.5%

c. 65.4% to 65.8%

Word Search

Refer to pages 281 – 282 of your text to obtain the Review Vocabulary. There are 20 hidden vocabulary words/expressions in the word search below. *Randomized comparative experiment*, *Nonresponse*, *95% confidence interval*, *Undercoverage*, and *Statistical inference* do not appear in the word search due to expression length. *Statistical significance*, *statistic*, *sample*, *sample proportion*, *experiment* and *double-blind experiment* appear separately in the word search. It should be noted that spaces and hyphens are removed.

O R T E R T E P I P A Q A P A I E D Y T O E F Z F
G S C E O B S E R V A T I O N A L S T U D Y T O W
N T E P E L P M A S E C N E I N E V N O C N J M B
T H F L E C A N N E L P M A S O R L E H E X P C T
N B F J X E R O R R E F O N I G R A M M T O H O A
E X E L P M A S M O D N A R E L P M I S B S O N A
M R O H F S M P S F V N E H I E E R R M N B J T M
I T B P R T E S P X H O N Q F T E S E W K I G R I
R G E T S I T E H S A M P L E P R O P O R T I O N
E S C E P Z E N S M N O L S X E E R X W N O S L O
P X A O V O R L H S I I M E O S E C E D S Z E G I
X I L A Y P E D R F O E G G V D E G D T Z W I R T
E L P M A S E S N O P S E R Y R A T N U L O V O A
E C N A C I F I N G I S L A C I T S I T A T S U L
M I P O I D E A D T J L A Z K K R H L B Y S A P U
E R X O L O G U Y G F M P F F X A T B L I X R E P
D Y O F I O B L H S E X S D Y A D R E A V A B E O
R C X N O I T U B I R T S I D G N I L P M A S H P
S S S T I G I D M O D N A R F O E L B A T R J L D
R W N P A I G A S N H Z O I M F G F U E G W R C C
T O O E B Y D U T S E V I T C E P S O R P U H D F
C F D A S C T V L C S C O N F O U N D I N G F M F
P O I V L G A G O A D X E I E D H T R E N I K J H
E E N S G R W X E Y S T P E Y C M M F N C A G E N
I F S T A T I S T I C E O E N E E Q E B G Y I C A

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____