Chapter 7
Data for Decisions

For All Practical Purposes: Effective Teaching

Because of the smaller class size, a teaching assistant can have a great influence on their students. You should see yourself as a role model for these undergraduate students. You may choose to interject quotes about the importance of education.

- "*Education is the movement from darkness to light.*" Allan Bloom
- "*Education is what survives when what has been learned has been forgotten.*" B.F. Skinner
- "*Education costs money, but then so does ignorance.*" Sir Claus Moser

A quote for you to consider in your teaching is by Joseph Joubert. He said, *"To teach is to learn twice."*

Chapter Briefing

A *sample* is a part of a large population. Hopefully, information obtained from the sample will allow us to draw conclusions about the *population*. In order for this to be possible, the sample must be representative of the population. *Random sampling* avoids the *bias* that can result when convenient samples are chosen or subjects choose themselves.

Experiments aim to show that certain treatments produce certain effects. Responses to treatment can be *confounded* with other variables so that no clear conclusions can be drawn. Good experimental design attempts to minimize confounding by comparing subjects who receive treatment, to a *control group*. Randomization is used to produce groups that are essentially similar prior to treatment. Random sampling eliminates bias but not variability. Differences unlikely to occur by chance provide evidence that the differences can be attributed to the effects of the treatment. Such differences are called *statistically significant*.

Statistical inference draws conclusions from data. These conclusions make estimates of a population parameter based on a statistic. Statistical inference is based on the idea that one needs to see how trustworthy a procedure is if it is repeated many times. An important method of inference is determining confidence intervals. With confidence intervals, we take as our estimate of an *unknown population parameter* the *appropriate sample statistic*. By revisiting the 68-95-99.7 rule, the 95% confidence interval is discussed. Confidence intervals are easy to obtain if the sampling distribution of our statistic is normal or approximately normal. Fortunately, this is the case with sample proportion and sample mean.

Being well prepared for class discussions with an understanding of the many terms in this chapter along with short examples will help students focus on the main topics presented in this chapter. Although calculations are needed in the chapter, a great deal of the chapter involves students applying the ideas to specific situations. In order to facilitate your preparation, the **Chapter Topics to the Point** has been broken down into **Data Analysis Versus Data Collection**, **Sampling**, **Bias**, **Simple Random Samples**, **Additional Bias**, **Studies**, **Study Design**, **Placebo Effect**, **Statistical Inference**, and **Confidence Intervals**. Any examples with solutions for these topics that do not appear in the text nor study guide are included in the *Teaching Guide*. You should feel free to use these examples in class, if needed.

The *Teaching Guide* includes the feature **Teaching the Calculator**. In this chapter the main problem area for students is using parentheses for calculations. Screen shots are included.

The last section of this chapter of *The Teaching Guide for the First-Time Instructor* is **Solutions to Student Study Guide ✎ Questions**. These are the complete solutions to the six questions included in the Student Study Guide. Students only have the answers to these questions, not the solutions.

Chapter Topics to the Point

⮷Data Analysis Versus Data Collection

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.

⮷Sampling

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**).

- A sample of people who choose to respond to a general appeal is called a **voluntary response sample**.
- A sample of people who are most easily available is called a **convenience sample.**

⮷Bias

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

- Voluntary response sampling is a likely source of bias.
- Convenience sampling is a likely source of bias.
- There may be gender-based bias, with women or men overrepresented in the sample.
- There may be economic class bias due to the location as to where the sample was taken.
- Responses can be strongly influenced by the wording of questions. By having leading questions or confusing questions, strong bias can be introduced

☝Teaching Tip

A good place to start class discussion is to ask students where they can envision other sources of bias in a sample. Many of the homework exercises in this chapter rely on students thinking about possible explanations as opposed to doing numerical calculations. So eliciting this type of classroom discussion should be beneficial to the students as they prepare to start their homework assignment.

Example

In a study of the shopping habits of adults, we asked 250 people as they exited a grocery store about their total purchase. What is the population? What is the sample? What kind of sample is it? Is there any possible bias?

Solution

The population is all adults that shop. The sample consists of those among the 250 who qualified as members of this population; for example, children would not be included, even if they were shopping. This is a convenience sample. There is possible gender-based bias in that grocery stores may be more frequented by woman rather than men.

⮷Simple Random Samples

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. 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.

In this chapter, Table 7.1 Random Digits will be referred to often. This table is printed below with a blank backside for you to tear out, if needed. Generally, students need to be told by instructions as to which line to start. The three-digit line label (left most column) is not part of the random digits.

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

Example

Use Table 7.1 to form a random sample of 5 students at Acme College from the entire population of 2350 AC students. Start on line 102.

Solution

Assign each AC student a four digit numerical label, 0001–2350, making sure that no label is assigned twice. Read strings of 4 consecutive digits, ignoring repetitions and unassigned strings, until 5 assigned labels are obtained. The students with those labels are the sample.

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

☝Teaching Tip

In order to familiarize students with Table 7.1, ask them to describe features of the table such as the width (it is 40 digits wide). Because it is easy to make mistakes in counting, students should have some kind of system to double check themselves as they go along. In the last example, since the string size was four digits, we should always start at the first digit of a new line as we progress down the table. Another feature students should notice is that the digits are grouped in order to make the reading easier. They should not consider the space between the groups of five as a forced place to begin looking for a string of digits. Similarly, if your string-length does not divide 40 (such as 3 or 6), the end of the line is not a terminal place.

Additional Bias

Random sampling will eliminate bias in the choice of a sample from a list of the population. It is difficult, however, to have a complete and totally accurate list of a population. Even a sound statistical design cannot guard against some of the pitfalls associated with statistical experiments.

- **Nonresponse** can be a cause of bias in an experiment.
- **Artificial environments** can be a cause of bias in an experiment.
- **Undercoverage**, by not including in samples certain parts of the population can cause bias.

Studies

- 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.
- 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.
- A **prospective study** is an observational study that records slowly developing effects of a group of subjects over a long period of time.
- An **uncontrolled study** is a study that lacks a comparison (i.e., control) group.
- A **controlled study** is a study that has a comparison (i.e., control) group.

☝Teaching Tip

A sign of an observational study is the lack of imposed treatment on the subjects.

Study Design

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

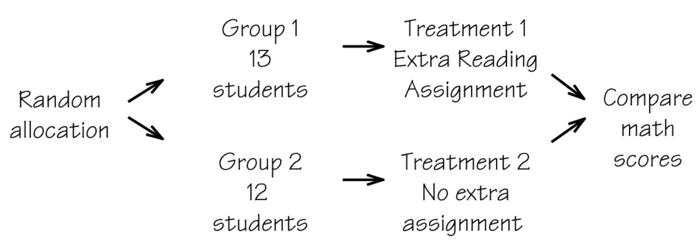
Example

A teacher has the following students in her class.

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

The teacher wants to determine if giving students extra reading assignments has any effect on a math test that includes word problems. The 25 students in the class are given a math pre-test. 13 students will be given the extra reading assignment and 12 will not (control group). Each group will be given a math post-test and the results will be compared. Outline the design of the experiment and begin on line 104 to determine the 12 students to be the control group.

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 104 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

The random labels in order are 01, 02, 07, 08, 13, 14, 15, 17, 18, 19, 21, and 25. Thus, the control group is Adam, Billy, Gwen, Heidi, Mary, Nadia, Ottis, Quinn, Rachel, Sarah, Victoria, and Zeki.

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

Placebo Effect

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

Teaching Tip

When one increases the number of factors in an experiment, the number of groups also increase. Forming a table is often helpful to show how many groups and the type of groups that are needed.

Example

Describe the subject groups for a combined test of a high-potassium supplement and a multi-vitamin supplement on smokers' blood pressure.

Solution

Form four equal groups (if possible) using a random procedure. Give the groups treatments according to the following table:

	High-potassium supplement	**Placebo**
Multi-vitamin supplement	**Group 1** Multi-vitamin supplement High-potassium supplement	**Group 2** Multi-vitamin supplement Placebo
Placebo	**Group 3** High-potassium supplement Placebo	**Group 4** Placebo Placebo

The experiment should be double-blind, so no participant knows who belongs to which group.

Statistical Inference

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.

☝Teaching Tip

Students can easily confuse the following two definitions and in turn may confuse down the way what formulas to apply. You may choose to make an effort to state them several times and write them on the board.

- A **parameter** is a number that describes the **population**. A parameter is a fixed value, but we generally do not know what it is.
- A **statistic** is a number that described a **sample**. A statistic is often used to estimate an unknown parameter.

If you have a simple random sample of size *n* from a large population and a count of successes (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*.

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.

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 $\hat{p}$ has the following features.

- It is approximately normal.
- The mean is *p*.
- The standard deviation is $\sqrt{\frac{p(1-p)}{n}}$.

☝Teaching Tip

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. You may choose to ask students to visualize what happens to the formula for standard deviation as *n* increases and ask what this would mean.

Confidence Intervals

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 sample 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.

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

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 $2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$.

Example

In a fall semester student poll, 450 potential spring semester students are asked if they have decided which classes they would take. Suppose that 212 say "yes."

a) Estimate the proportion of who have decided about their spring classes.
b) Find a 95% confidence interval for this estimate.

Solution

a) The sample proportion of students is as follows.

$$\hat{p} = \frac{212}{450} \approx 0.471 = 47.1\%$$

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

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 0.471 \pm 2\sqrt{\frac{0.471\left(1-0.471\right)}{450}} = 0.471 \pm 2\sqrt{\frac{0.471\left(0.529\right)}{450}} \approx 0.471 \pm 0.047$$

$$0.471 - 0.047 = 0.424 = 42.2\% \text{ to } 0.471 + 0.047 = 0.518 = 51.8\%$$

We get an interval of (42.2%, 51.8%). Rounding this interval would be (42%, 52%)

Teaching Tip

To gain greater certainty in your statistical estimate, you must increase the confidence interval accordingly. That is, precision goes down as certainty goes up. To avoid this loss of precision, you must use a larger sample size in your poll. You may choose to warn students about rounding along with discussing this idea. In the last example, if you had to round the interval say, (34.7%, 37.2%), the 34.7% would be rounded to 34% and the 37.2% would be rounded up to 38.

Teaching Tip

You may choose to summarize some of the ideas of this section by stating that statistical evidence not based on experiments can show that an effect is present without showing *why* the effect is present. Even when data is produced and analyzed carefully, there is always the chance that the conclusions drawn from the data are incorrect. The strength of statistical inference is in setting confidence levels as high as we deem necessary.

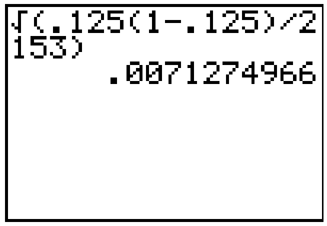
Teaching the Calculator

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

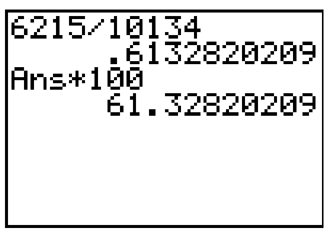
In a political poll, 10134 potential voters are asked if they have decided yet which candidate they will vote for in the next election. Suppose that 6215 say "yes."

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.

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

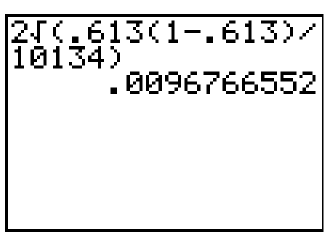


$\hat{p}$ would be approximately $0.613 = 61.3\%$

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

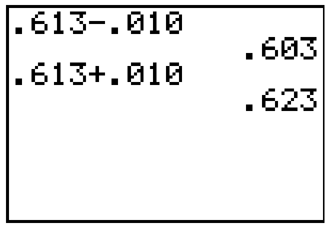
$$\hat{p} \pm 2\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 0.613 \pm 2\sqrt{\frac{0.613(1-0.613)}{10{,}134}}$$

Calculate $2\sqrt{\frac{0.613(1-0.613)}{10{,}134}}$ 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 $(60.3\%, 62.3\%)$.

Solutions to Student Study Guide ✎ Questions

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 ____________.

Solution

Voluntary response sample (choice b) is not used.

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

Question 2

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 102 use Table 7.1 to find the five random students. In alphabetical order, who would be the third student that the teacher would select to poll?

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

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 102 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 19 = Sarah, 17 = Quinn, 09 = Iliana, 22 = Wally, and 13 = Mary. Placing these students in alphabetical order, Quinn is the third student the teacher would poll.

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 _____________.

Solution

Sample (choice b) is not used.

1) In this example. 25.05 is called a **parameter**.
2) In this example, 24.9 is called a **statistic**.

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?

Solution

The sample proportion is $\hat{p} = \frac{580}{1250} = 0.464 = 46.4\%$, 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 46.4%.

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?

Solution

Convert 45% to decimal form, 0.45. Standard deviation is as follows.

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.45(1-0.45)}{1250}} = \sqrt{\frac{0.45(0.55)}{1250}} \approx 0.0141$$

Question 6

In a college survey, 847 students were asked if they thought the cost of tuition is 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.

Solution

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

$$\hat{p} = \frac{521}{847} \approx 0.615 = 61.5\%$$

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

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.615 \pm 2\sqrt{\frac{0.615(1-0.615)}{847}} = 0.615 \pm 2\sqrt{\frac{0.615(0.385)}{847}} \approx 0.615 \pm 0.033$$

$$0.615 - 0.033 = 0.582 = 58.2\% \text{ to } 0.615 + 0.033 = 0.648 = 64.8\%$$

We get an interval of (58.2%, 64.8%).