Chapter 5

5.1 In an observational study, there is no attempt to control factors that might influence the variable of interest. In an experimental study, a factor (such as regular use of a fitness center) is controlled by randomly selecting who is exposed to that factor, thereby reducing the influence of other factors on the variable of interest.

5.2a The study is observational. The statistics practitioner did not randomly assign stores to buy cans or bottles.
b Randomly assign some stores to receive only cans and others to receive only bottles.

5.3 Randomly sample smokers and nonsmokers and compute the proportion of each group that has lung cancer.
b The study is observational. Experimental data would require the statistics practitioner to randomly assign some people to smoke and others not to smoke.

5.4a A survey can be conducted by means of a personal interview, a telephone interview, or a self-administered questionnaire.
b A personal interview has a high response rate relative to other survey methods, but is expensive because of the need to hire well-trained interviewers and possibly pay travel-related costs if the survey is conducted over a large geographical area. A personal interview also will likely result in fewer incorrect responses that arise when respondents misunderstand some questions. A telephone interview is less expensive, but will likely result in a lower response rate. A self-administered questionnaire is least expensive, but suffers from lower response rates and accuracy than interviews.

5.5 Five important points to consider when designing a questionnaire are as follows:

(1) The questionnaire should be short.
(2) Questions should be short, clearly worded, and unambiguous.
(3) Consider using dichotomous or multiple-choice questions, but take care that respondents needn't make unspecified assumptions before answering the questions.
(4) Avoid using leading questions.
(5) When preparing the questions, think about how you intend to tabulate and analyze the responses.

5.6a The sampled population will exclude those who avoid large department stores in favor or smaller shops, as well as those who consider their time too valuable to spend participating in a survey. The sampled

population will therefore differ from the target population of all customers who regularly shop at the mall.
b The sampled population will contain a disproportionate number of thick books, because of the manner in which the sample is selected.
c The sampled population consists of those eligible voters who are at home in the afternoon, thereby excluding most of those with full-time jobs (or at school).

5.7a The *Literary Digest* was a popular magazine in the 1920s and 1930s which had correctly predicted the outcome of many presidential elections. To help predict the outcome of the 1936 presidential election, the *Literary Digest* mailed sample ballots to 10 million prospective voters. Based on the results of the ballots returned, the magazine predicted that the Republican candidate, Alfred Landon, would defeat the Democratic incumbent, Franklin D. Roosevelt, by a 3 to 2 margin. In fact, Roosevelt won a landslide victory, capturing 62% of the votes.
b The main reason for the poll being so wrong was *nonresponse bias* resulting from a self-selected sample, causing the sample to be unrepresentative of the target population. (Only 2.3 million ballots were returned.) The second reason was *selection bias*, resulting from poor sampling design, causing the sampled population and the target population to differ. Most of those to whom a ballot was sent were selected from the *Literary Digest's* subscription list and from telephone
directories. These people tended to be wealthier than average and tended to vote Republican.

5.8a A self-selected sample is a sample formed primarily on the basis of voluntary inclusion, with little control by the designer of the survey.
b Choose any recent radio or television poll based on responses of listeners who phone in on a volunteer basis.
c Self-selected samples are usually biased, because those who participate are more interested in the issue than those who don't, and therefore probably have a different opinion.

5.9 We should ignore the results because this is an example of a self-selected sample.

5.10 No, because the sampled population consists of the responses about the professor's course. We cannot make draw inferences about all courses.

5.11 We used Excel to generate 40 three-digit random numbers. Because we
will ignore all randomly generated numbers over 800, we can expect to ignore about 20% (or about 8 to 10) of the randomly generated numbers. We will also ignore any duplications. We therefore chose to generate 40 three-digit random numbers, and will use the first 25 unique random numbers less than 801 to select our sample. The 40 numbers generated are shown below, with a stroke through those to be ignored.

6	357	456	449	862	154	55	412	475	430
999	912	60	207	717	651	10	294	327	165
576	871	990	354	390	540	893	181	496	870
738	820	32	963	160	32	231	86	970	46

5.12 We used Excel to generate 30 six-digit random numbers. Because we will ignore any duplicate numbers generated, we generated 30 six-digit random numbers and will use the first 20 unique random numbers to select our sample. The 30 numbers generated are shown below.

169,470	744,530	22,554	918,730	320,262	503,129
318,858	698,203	822,383	938,262	800,806	56,643
836,116	123,936	80,539	154,211	391,278	940,154
110,630	856,380	222,145	692,313	949,828	561,511
909,269	811,274	288,553	749,627	858,944	39,308

5.13 The operations manager can select stratified random samples where the strata are the four departments. Simple random sampling can be conducted in each department.

5.14 Use cluster sampling, letting each city block represent a cluster.

5.17a Sampling error refers to an inaccuracy in a statement about a population that arises because the statement is based only on sample data. We expect this type of error to occur because we are making a statement based on incomplete information. Nonsampling error refers to mistakes made in the acquisition of data or due to the sample observations being selected improperly.
b Nonsampling error is more serious because, unlike sampling error, it cannot be diminished by taking a larger sample.

5.18 Three types of nonsampling errors:

(1) Error due to incorrect responses

(2)Nonresponse error, which refers to error introduced when responses are not obtained from some members of the sample. This may result in the sample being unrepresentative of the target population.

(3)Error due to selection bias, which arises when the sampling plan is such that some members of the target population cannot possibly be selected for inclusion in the sample.

5.19 Yes. A census will likely contain significantly more nonsampling errors than a carefully conducted sample survey.