Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

SKU: 9781305115347-SOLUTIONS

Chapter 2
Collecting Data Sensibly

2.1 This is an observational study. The treatments (length of stay) were determined by the condition of the patients. (In an experiment the patients would be assigned to the various lengths of stay by the investigators, usually using random assignment.)

2.2 This was an experiment, since the investigators (not the students) determined which discussion sections received the chocolate and which did not.

2.3 This was an experiment, since the professor (not the students) determined who was supplying a buying price and who was supplying a selling price.

2.4 **a** This is an observational study.

b No. It is quite possible, for example, that those children who averaged more than two hours of television viewing per day received, generally speaking, a less good education than those who did not, and that it is the less good education, and not the television viewing, that caused the lower reading scores.

2.5 **a** This is an experiment since it was decided by the researchers (in this case by random assignment) which participants would receive which treatments.

b Yes. Since the participants were randomly assigned to the treatments it is reasonable to conclude that receiving either real or fake acupuncture was the cause of the observed reductions in pain.

2.6 **a** This is an observational study.

b Yes. Since the researchers looked at a random sample of publically accessible MySpace web profiles posted by 18-year-olds, it is reasonable to generalize the stated conclusion to all 18-year-olds with publically accessible MySpace profiles.

c No, it is not reasonable to generalize the stated conclusion to all 18-year-old MySpace users since no users without publically accessible profiles were included in the study.

d No, it is not reasonable to generalize the stated conclusion to all MySpace users with publically accessible profiles since only 18-year-olds were included in the study.

2.7 **a** This is an experiment.

b Yes. Since the participants were randomly assigned to the treatments the researcher can reasonably claim that the music played was the cause of the higher rating.

2.8 It is quite possible, for example, that 3- and 4-year olds who drink something sweet once or twice a day generally speaking consume larger amounts of fat than those who do not, and that it is the fat that is causing the weight problems a year later, rather than the consumption of the sweet drinks.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.9 We are told that moderate drinkers, as a group, tended to be better educated, wealthier, and more active than nondrinkers. It is therefore quite possible that the observed reduction in the risk of heart disease amongst moderate drinkers is caused by one of these attributes and not by the moderate drinking.

2.10 **a** No. It is quite possible, for example, that women who choose to watch *Oprah* generally speaking have a more health oriented outlook than those who watch other daytime talk shows, and it is this health oriented outlook that causes the decrease in craving for fattening foods, not the watching of *Oprah*.

b Neither generalization would be reasonable since the survey was conducted on the DietSmart.com website. It is unlikely that users of this website would be representative of the population of women in the United States or of the population of women who watch daytime talk shows.

2.11 **a** The data would need to be collected from a simple random sample of affluent Americans.

b No. Since the survey included only affluent Americans the result cannot be generalized to all Americans.

2.12 It is possible, for example, that people who live in the South generally speaking eat less healthy food and exercise less than people who live in other regions of the country, and that the greater percentage of high blood pressure is caused not by living in the South but by the less healthy food and the lack of exercise.

2.13 The following is one possible method. Use a computer list of all the students at the college. Assign to each student a 10-digit decimal number. Sort the students according to the numbers assigned, smallest to largest. The first 100 students on the sorted list will form the sample.

2.14 Method 1: Using a computer list of the graduates, number the graduates 1-140. Use a random number generator on a calculator or computer to randomly select a whole number between 1 and 140. The number selected represents the first graduate to be included in the sample. Repeat the number selection, ignoring repeated numbers, until 20 graduates have been selected.
Method 2: Using a computer list of the graduates, number the graduates 001-140. Take the first three digits from the left hand end of a row from a table of random digits. If the three-digit number formed is between 001 and 140 inclusive, the graduate with that number should be the first graduate in the sample. If the number formed is not between 001 and 140 inclusive, the number should be ignored. Repeat the process described for the next three digits in the random number table, and continue in the same way until 20 graduates have been selected. (Three-digit numbers that are repeats of numbers previously selected should be ignored.)

2.15 The following is one possible method. Number the signatures 1-500. Use a random number generator on a calculator or computer to randomly select a whole number between 1 and 500. The number selected represents the first signature to be included in the sample. Repeat the number selection, ignoring repeated numbers, until 30 signatures have been selected.

2.16 **a** The 716 bicycle fatalities constitute *all* bicycle fatalities in 2008, and so the group represents a census.

b The average age of 41 years is a population characteristic, since it is the average for *all* bicycle fatalities in 2008.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.17 **a** The population is all American women.

b No. The sample included women only from Maryland, Minnesota, Oregon, and Pennsylvania. It is quite possible that the relationship between exercise and cognitive impairment is different for women in other states.

c No. As mentioned in Part (b), it is quite possible that the relationship between exercise and cognitive impairment is different for women in other states. For example, other states might contain a different racial mix from the four included in the study, and it is possible that women of different racial origins respond differently to exercise in terms of cognitive impairment.

d The inadequate sampling method used in this study is an example of selection bias. Women in the other 46 states were excluded from the study.

2.18 **a** Cluster sampling

b Stratified random sampling

c Convenience sampling

d Simple random sampling

e Systematic sampling

2.19 **a** Using the list, first number the part time students 1-3000. Use a random number generator on a calculator or computer to randomly select a whole number between 1 and 3000. The number selected represents the first part time student to be included in the sample. Repeat the number selection, ignoring repeated numbers, until 10 part time students have been selected. Then number the full time students 1-3500 and select 10 full time students using the same procedure.

b No. With 10 part time students being selected out of a total of 3000 part time students, the probability of any particular part time student being selected is 10/3000 = 1/300. Applying a similar argument to the full time students, the probability of any particular full time student being selected is 10/3500 = 1/350. Since these probabilities are different, it is not the case that every student has the same chance of being included in the sample.

2.20 Convenience samples are, by nature, very unlikely to be representative of the population.

2.21 **a** The pages of the book have already been numbered between 1 and the highest page number in the book. Use a random number generator on a calculator or computer to randomly select a whole number between 1 and the highest page number in the book. The number selected will be the first page to be included in the sample. Repeat the number selection, ignoring repeated numbers, until the required number of pages has been selected.

b Pages that include exercises tend to contain more words than pages that do not include exercises. Therefore, it would be sensible to stratify according to this criterion. Assuming that 20 non-exercise pages and 20 exercise pages will be included in the sample, the sample should be selected as follows. Use a random number generator to randomly select a whole number between 1 and the highest page number in the book. The number selected will be the

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

first page to be included in the sample. Repeat the number selection, ignoring repeated numbers and keeping track of the number of pages of each type selected, until 20 pages of one type have been selected. Then continue in the same way, but ignore numbers corresponding to pages of that type. When 20 pages of the other type have been selected, stop the process.

c Randomly select one page from the first 20 pages in the book. Include in your sample that page and every 20th page from that page onwards.

d Roughly speaking, in terms of the numbers of words per page, each chapter is representative of the book as a whole. It is therefore sensible for the chapters to be used as clusters. Using a random number generator randomly choose three chapters. Then count the number of words on each page in those three chapters.

e Answers will vary.

f Answers will vary.

2.22 It is not the *proportion* of voters that is important, but the *number* of voters in the sample – and 1000 is an adequate number.

2.23 The researchers should be concerned about nonresponse bias. Only a small proportion (20.7%) of the selected households completed the interview, and it is quite possible that those households who did complete the interview are different in some relevant way concerning Internet use from those who did not.

2.24 **a** This was a convenience sample. The study simply used students from one psychology class.

b The students in this psychology class, and at this particular small college, are unlikely to be representative of the set of all college students in the U.S. Also, since the survey was about illegal drug use, we cannot be sure that the students were giving truthful answers to the questions.

2.25 First, the participants in the study were all students in an upper-division communications course at one particular university. It is not reasonable to consider these students to be representative of all students with regard to their truthfulness in the various forms of communication. Second, the students knew during the week's activity that they were surveying themselves as to the truthfulness of their interactions. This could easily have changed their behavior in particular social contexts and therefore could have distorted the results of the study.

2.26 **a** No. The games used in the study were the 20 most popular for each of three games systems and therefore are not likely to be representative of the population of all video games with respect to the number of violent interactions.

b No. The sample is not likely to have been representative of the population of all video games with respect to the number of violent interactions and therefore it would not be reasonable to generalize the researchers' conclusion to all video games.

2.27 First, the people who responded to the print and online advertisements might be different in some way relevant to the study from the population of people who have online dating profiles. Second, only the *Village Voice* and Craigslist New York City were used for the recruitment. It is quite possible that people who read that newspaper or access those websites differ from the population

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

in some relevant way, particularly considering that they are both New York City based publications.

2.28 No. First, only 132 of the 1260 students who were sent the survey responded, and it is quite likely that those who chose to respond differ in some relevant way from those who did not. (This is nonresponse bias.) Second, those students who did respond might not have been fully honest about their financial condition. (This is response bias.)

2.29 **a** Yes. It is possible that students of a given class standing tend to be similar in the amounts of money they spend on textbooks.

b Yes. It is possible that students who pursue a certain field of study tend to be similar in the amounts of money they spend on textbooks.

c No. It is unlikely that stratifying in this way will produce groups that are homogeneous in terms of the students' spending on textbooks.

2.30 The individuals within each stratum should on the whole be similar in terms of the topic of the study. This is true of the proposed strata in Scheme 2, since it is likely that college students will on the whole be similar in their opinions regarding the possible tax increase; likewise nonstudents who work full time will on the whole be similar in their opinions regarding the possible tax increase, and nonstudents who do not work full time will on the whole be similar in their opinions regarding the possible tax increase. Scheme 1, however, is not suitable since we have no reason to believe that people within the proposed first-letter-of-last-name strata will be similar in terms of their attitudes to the possible tax increase. Similarly the suggested stratification in Scheme 3 is very unlikely to produce homogeneous groups.

2.31 It is not reasonable to generalize these results to the population of U.S. adults since the people who sent their hair for testing did so voluntarily. It is quite possible that people who would choose to participate in a study of this sort differ in their mercury levels from the population as a whole.

2.32 Different subsets of the population might have responded by different methods. For example, it is quite possible that younger people (who might generally be in favor of continuing the parade) chose to respond via the Internet while older people (who might on the whole be against the parade) chose to use the telephone to make their responses.

2.33 **a** Binding strength

b Type of glue

c The extraneous variables mentioned are the number of pages in the book and whether the book is bound as a hardback or a paperback. Further extraneous variables that might be considered include the weight of the material used for the cover and the type of paper used.

2.34 **a** Use two IQ tests (Test 1 and Test 2) of equal levels of difficulty. Randomly select 50 students from the school. Randomly assign the 50 students to two groups, Group A and Group B. The experiment will be conducted over two days. On the first day, students in Group A will do an IQ test without listening to any music and students in Group B will do an IQ test after listening to a Mozart piano sonata. On the second day the activities of the two groups will be switched. For each student decide randomly whether he/she will take Test 1 on the first day

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

and Test 2 on the second day, or vice versa. All conditions (temperature, time of day, amount of light, etc. – everything apart from the presence of the music or not) should be kept equal between the two days and between the two groups. In particular, the students taking the test without listening to the music should nonetheless sit quietly in a room for the length of time of the piano sonata. At the end of the experiment the after-music IQ scores should be compared to the no-music IQ scores.

b Yes, the fact that all conditions are kept the same is direct control.

c By having each student take IQ tests under *both* experimental conditions we are using a matched pairs design, and matched pairs is a form of blocking.

d The students were randomly assigned to the two orders of the treatments (no music then music, or music then no music). Also it was decided randomly for each student whether Test 1 or Test 2 would be taken first.

2.35 The following is one possible method. Write the names of the subjects on identical slips of paper, and place the slips in a hat. Mix the slips, and randomly select ten slips from the hat. The names on those ten slips are the people who will dry their hands by the first method. Randomly select a further ten slips from the hat. The names on those ten slips are the people who will dry their hands by the second method. The remaining ten people will dry their hands by the third method.

2.36 Random assignment should have been used to determine, for each cyclist, which drink would be consumed during which break.

2.37 **a** Blocking

b Direct control

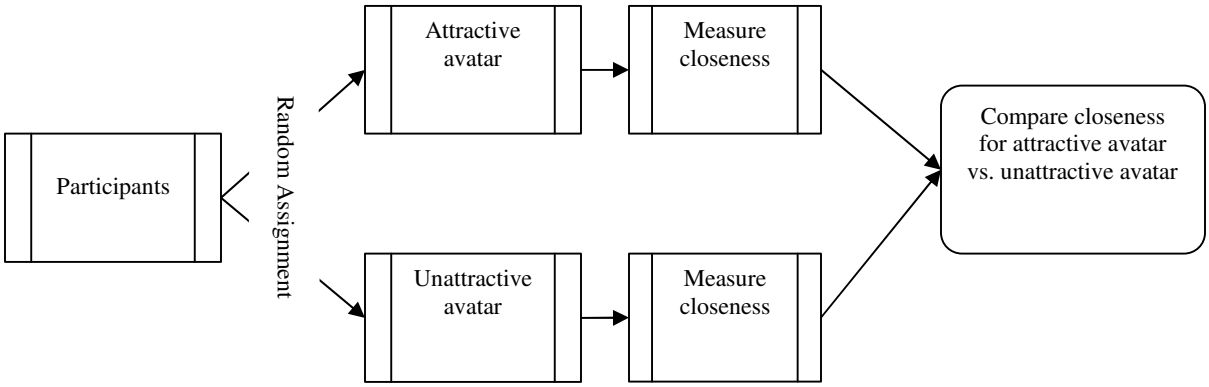
2.38 We rely on random assignment to produce comparable experimental groups. If the researchers had hand-picked the treatment groups, they might unconsciously have favored one group over the other in terms of some variable that affects the subjects' ability to deal with multiple inputs.

2.39 The figure shows that comparable groups in terms of age have been formed. (Differences between the age distributions in the groups can be seen: there are one or two children in the LR group who are younger than all those in the OR group, and also there is a greater number of children over 14 years old in the LR group. It is inevitable that there will be *some* differences between the groups under random assignment.)

2.40 **a** If the participants had been able to choose their own avatars, then it is quite possible, for example, that people with a lot of self confidence would tend to choose the attractive avatar while those with less self confidence would tend to choose the unattractive avatar. Then, if the same result was obtained as the one described in the paper, it would be impossible to tell whether the greater closeness achieved by those with the attractive avatar came about as a result of the avatar or as a result of those people's greater self confidence.

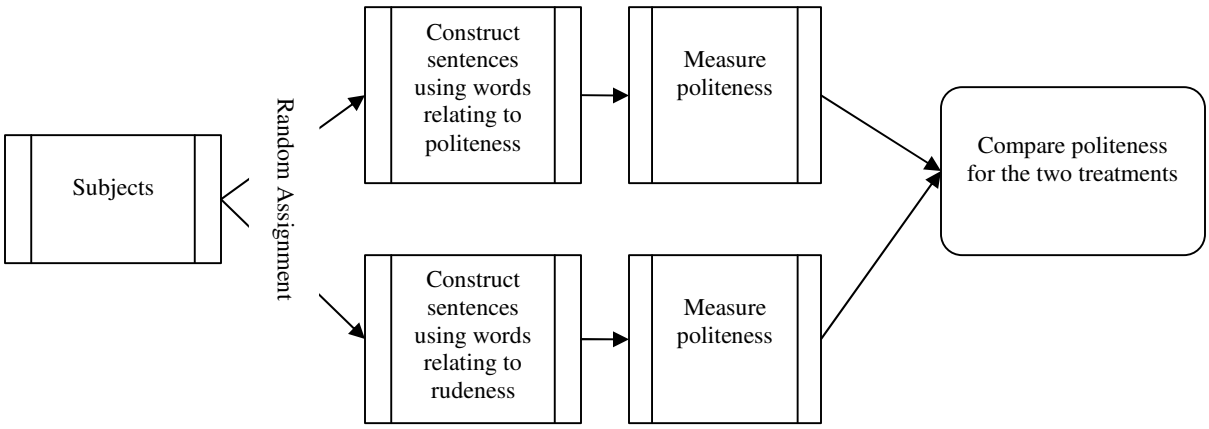
Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

b



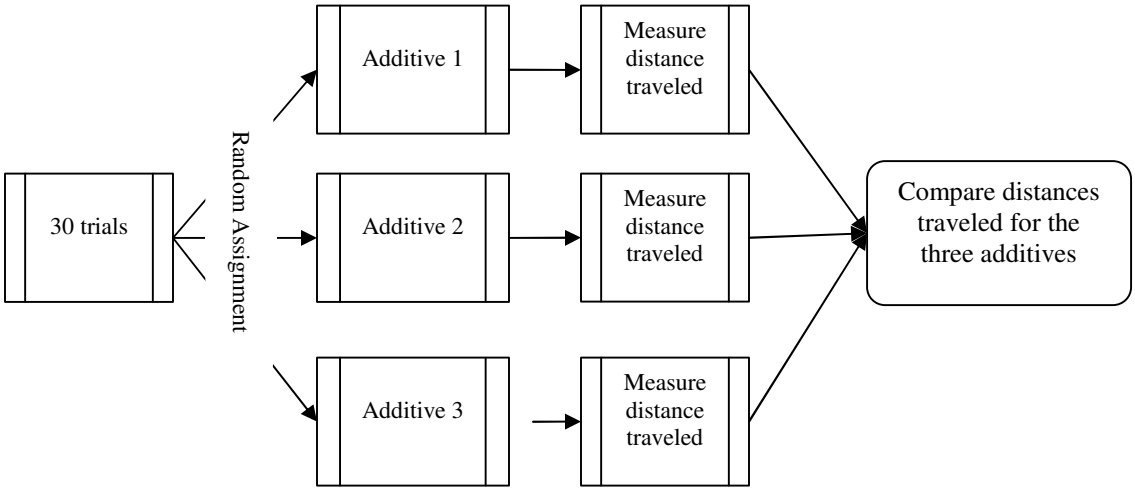
2.41 We rely on random assignment to produce comparable experimental groups. If the researchers had hand-picked the treatment groups, they might unconsciously have favored one group over the other in terms of some variable that affects the subjects' ability to learn through video gaming activity.

2.42



Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.43



2.44 How many bottles of water were used in the experiment? Were the bottles identical? Was there a control group in which the bottles had identical labels to those used in the treatment group but without the positive words? Were the bottles randomly assigned to the groups? Was the state of the water measured both before and after the experiment? Did the people who measured the water's structure know which bottles were in which group?

2.45 **a** This is an experiment. The researchers assigned the students to the treatments.

b No. The subjects were randomly assigned to the treatments, but there was no random selection from a population of interest.

c Yes.

d If the performance was *significantly* better for the group that read the material in the landscape orientation then, yes, for the set of subjects used in the experiment, reasoning was improved by turning the screen to landscape orientation. The subjects were randomly assigned to the treatment groups, and a significant difference in the results of the two groups means that the observed difference is unlikely to have occurred by chance.

e No. The students used in the study were all undergraduates taking psychology at a large university, and therefore cannot be considered to be representative of any larger population.

2.46 **a** The treatments are the names – Ann Clark and Andrew Clark – given to the participants.

b The response variables are the participants' answers to the questions given.

c Selecting a random sample of 1161 voters and giving them the female name, and then selecting a second random sample 1139 voters and giving them the male name, is exactly equivalent to selecting a random sample of 2300 voters, and then randomly assigning 1161 of them to the female name and the remainder to the male name. (It is assumed here that in the study given in the question it was ensured that there was no overlap between the two samples.)

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.47 **a** Red wine, yellow onions, black tea

b Absorption of flavonols

c Alcohol tolerance, amount of flavonols in the diet (other than from red wine, yellow onions, and black tea), gender

2.48 Suppose that an experiment is conducted in which people are given either a drug or a placebo, and that those who are given the drug do significantly better, on average, than those who are given the placebo. Since both groups experience the placebo effect (the psychological effect of taking a tablet) we are able to attribute the greater improvement of those who took the drug to the chemicals in the drug. However we are unable to tell just how much of a placebo effect is being experienced by all the subjects. By adding a control group (a group that is given nothing) and comparing the results for this group with the results for the placebo group we can measure the extent of the placebo effect.

2.49 **"**Blinding" is ensuring that the experimental subjects do not know which treatment they were given and/or ensuring that the people who measure the response variable do not know who was given which treatment. When this is possible to implement, it is useful that the subjects do not know which treatments they were given since, if a person knows what treatment he/she was given, this knowledge could influence the person's perception of the response variable, or even, through psychological processes, have a direct effect on the response variable. If the response variable is to be measured by a person other than the experimental subjects it is useful if this person doesn't know who received which treatment since, if this person *does* know who received which treatment, then this could influence the person's perception of the response variable.

2.50 Answers will vary.

2.51 **a** In order to know that the results of this experiment are valid it is necessary to know that the assignment of the women to the groups was done randomly. For suppose, for example, that the women were allowed to choose which groups they went into. Then it would be quite possible, for instance, that women who are particularly social by nature, and therefore whose health would be enhanced by any regular social gathering, would choose the more interesting sounding art discussions, while those less social by nature (and therefore less likely to be helped by social gatherings) would choose the more conventional discussions of hobbies and interests. Then it would be impossible to tell whether the stated results were caused by the discussions of art or by the greater social nature of the women in the art discussion group.

b Suppose that all the women took part in weekly discussions of art, and that generally an improvement in the medical conditions mentioned was observed amongst the subjects. Then it would be impossible to tell whether these health improvements had been caused by the discussions of art or by some factor that was affecting all the subjects, such as an improvement in the weather over the four months. By including a control group, and by observing that the improvements did not take place (generally speaking) for those in the control group, factors such as this can be discounted, and the discussions of art are established as the cause of the improvements.

2.52 **a** It is very possible that the nurses might have preconceptions about the two forms of surgery in terms of the amount of pain and nausea caused. Therefore, if the nurses know which children have been given which form of surgery, this might affect the amounts of medication

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

they give. By making sure that the nurses do not have this knowledge, this possible effect is avoided.

b Since the incisions made under the two procedures are different the patients and/or their parents would know which method had been used.

2.53 We will assume that only four colors will be compared, and that only headache sufferers will be included in the study.

Prepare a supply of "Regular Strength" Tylenol in four different colors: white (the current color of the medication, and therefore the "control"), red, green, and blue. Recruit 20 volunteers who suffer from headaches. Instruct each volunteer not to take any pain relief medication for a week. After that week is over, issue each volunteer a supply of all four colors. Give each volunteer an order in which to use the colors (this order would be determined randomly for each volunteer). Instruct the volunteers to use one fixed dose of the medication for each headache over a period of four weeks, and to note on a form the color used and the pain relief achieved (on a scale of 0-10, where 0 is no pain relief and 10 is complete pain relief). At the end of the four weeks gather the results and compare the pain relief achieved by the four colors.

2.54 **a** Randomly assigning 852 children to the book group and the rest to the control group consists of randomly selecting 852 to be in the book group and putting the remaining children in the control group.

b If no control group had been included in the study, then the only results available to the researchers would be the reading scores of the children who had been given the reading books. There would be no way of telling whether these scores were any better than the scores for children who were not given reading books.

2.55 Of the girls, randomly assign 350 to the book group and 350 to the no-book group. (You could do this by starting with a computer list of the names of the 700 girls. Assign to each name a random 10-digit number. Sort the names according to the numbers, from smallest to largest. The first 350 names on the sorted list are assigned to the book group, and the remainder to the no-book group.) Using a similar method, randomly assign 315 of the boys to the book group and 315 to the no-book group.

2.56 Suppose that the dog handlers and/or the experimental observers had known which patients did and did not have cancer. It would then be possible for some sort of (conscious or unconscious) communication to take place between these people and the dogs so that the dogs would pick up the conditions of the patients from these people rather than through their perception of the patients' breath. By making sure that the dog handlers and the experimental observers do not know who has the disease and who does not it is ensured that the dogs are getting the information from the patients.

2.57 **a** If the judges had known which chowder came from which restaurant then it is unlikely that Denny's chowder would have won the contest, since the judges would probably be conditioned by this knowledge to choose chowders from more expensive restaurants.

b In experiments, if the people measuring the response are not blinded they will often be conditioned to see different responses to some treatments over other treatments, in the same way as the judges would have been conditioned to favor the expensive restaurant chowders. It

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

is therefore necessary that the people measuring the response should not know which subject received which treatment, so that the treatments can be compared on their own merits.

2.58 This describes the placebo effect. The implication is that the experiments have included patients who have been given a placebo in place of the antidepressants.

2.59 **a** A placebo group would be necessary if the mere thought of having amalgam fillings could produce kidney disorders. However, since the experimental subjects were sheep the researchers do not need to be concerned that this would happen.

b A resin filling treatment group would be necessary in order to provide evidence that it is the material in the amalgam fillings, rather than the process of filling the teeth, or just the presence of foreign bodies in the teeth, that is the cause of the kidney disorders. If the amalgam filling group developed the kidney disorders and the resin filling group did not, then this would provide evidence that it is some ingredient in the amalgam fillings that is causing the kidney problems.

c Since there is concern about the effect of amalgam fillings it would be considered unethical to use humans in the experiment.

2.60 Answers will vary.

2.61 **a** This is an observational study.

b In order to evaluate the study, we need to know whether the sample was a random sample.

c No. Since the sample used in the Healthy Steps study was known to be nationally representative, and since the paper states that, compared with the HS trial, parents in the study sample were disproportionately older, white, more educated, and married, it is clear that it is not reasonable to regard the sample as representative of parents of all children at age 5.5 years.

d The potential confounding variable mentioned is what the children watched.

e The quotation from Kamila Mistry makes a statement about cause and effect and therefore is inconsistent with the statement that the study can't show that TV was the cause of later problems.

2.62 Answers will vary.

2.63 Answers will vary.

2.64 Study 1

1. Observational study
2. No
3. No
4. No. The fact that calcium takers were more common among the heart attack patients implies mathematically that calcium takers were more likely to be heart attack patients than non-calcium takers. However, it is quite possible that people who take a calcium supplement very often also take another supplement, and it is this other supplement that is causing heart attacks, not the calcium.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

5. No. The hospital at which this study was conducted cannot be considered to be representative of any larger population.

Study 2
1. Observational study
2. Yes
3. No
4. No. It is quite possible that people who take a calcium supplement very often also take another supplement, and it is this other supplement that is causing heart attacks, not the calcium.
5. The conclusions can only be generalized to the population of people living in Minneapolis who receive Social Security.

Study 3
1. Experiment
2. Yes
3. No
4. No
5. No reasonable conclusion can be drawn from the study.

Study 4
1. Experiment
2. No
3. Yes
4. Yes
5. No. The participants were volunteers.

2.65 By randomly selecting the phone numbers, calling back those where there are no answers, and asking for the adult in the household with the most recent birthday the researchers are avoiding selection bias. However, selection bias could come about as a result of the fact that not all Californians have phones. Also, by selecting the adult with the most recent birthday certain birth dates are being favored, and it could be suggested that different attributes of people born at different times of the year could introduce further selection bias. Further to that, there is always the concern that people might not answer truthfully. This is response bias.

2.66 Answers will vary.

2.67 We rely on random assignment to produce comparable experimental groups. If the researchers had hand-picked the treatment groups, they might unconsciously have favored one group over the other in terms of some variable that affects the ability of the people at the centers to respond to the materials provided.

2.68 Yes. It seems quite possible that "rate of talk" would be related to gender and therefore that gender would be a sensible blocking variable.

2.69 **a** Observational study

b It is quite possible that the children who watched large amounts of TV in their early years were also those, generally speaking, who received less attention from their parents, and it was the lack of attention from their parents that caused the later attention problems, not the TV-watching.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.70 **a** Observational study

b No. As Mr. Adamson pointed out, it is quite possible that women who drink a lot of soda generally have less healthy lifestyles than those who do not, and that it's the unhealthy lifestyles that are causing the diabetes, not the consumption of soda or fruit punch.

2.71 **a** It cannot be concluded from the results of the study that being single causes an increased risk of violent crime.

b It is possible, for example, that the kind of people who get married are the kind of people, generally speaking, who are more careful about where they walk alone. It is this carefulness that is reducing the likelihood of being a victim of violent crime, not the fact that these people are married.

2.72 Nonresponse bias: it is likely that those who responded differed in some important way from those who did not. Also there is a possibility of response bias in that those who *did* respond might not have been answering truthfully.

2.73 All the participants were women, from Texas, and volunteers. All three of these facts tell us that it is likely to be unreasonable to generalize the results of the study to all college students.

2.74 **a** The design is good in that it includes a large number of doctors (adequate replication) and directly controls the clothing and gestures. In order for the design to be correct it is important to know that the doctors were randomly assigned to the eight tapes.

b A greater number of actors should have been used. It is possible that particular attributes of the individual actors (apart from their race and gender) might have influenced the doctors' decisions, while use of a greater number of actors would have reduced the effect of such individual differences.

2.75 **a** The extraneous variables identified are gender, age, weight, lean body mass, and capacity to lift weights. They were dealt with by direct control: all the volunteers were male, about the same age, and similar in weight, lean body mass, and capacity to lift weights.

b Yes, it is important that the men were not told which treatment they were receiving, otherwise the effect of giving a placebo would have been removed. If the participants *were* told which treatment they were receiving, then those taking the creatine would have the additional effect of the mere taking of a supplement thought to be helpful (the placebo effect) and those getting the fake preparation would not get this effect. It would then be impossible to distinguish the influence of the placebo effect from the effect of the creatine itself.

c Yes, it would have been useful if those measuring the increase in muscle mass had not know who received which treatment. It is possible that, through having this knowledge, the people would have been unconsciously influenced into exaggerating the increase in muscle mass for those who took the creatine.

2.76 **a** The treatments are standing and squatting. The response variable is the amount of the tip (probably expressed as a percentage of the amount of the bill).

b There are many variables affecting the amount of the tip, and most of these are dealt with by the use of the coin to decide which treatment to use. However extraneous factors relating to

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

the behavior of the server can be directly controlled. For example, the server can attempt to keep constant such things as his/her manner and his/her speed of service.

c Blocking would be appropriate for some extraneous variables that cannot be dealt with using direct control. For example, the waiter could keep note of whether the party is a family, a group of friends, or an individual, and of the time of day at which the meal is taking place.

d Any extraneous variable could be a confounding variable if, by bad luck, the result of the random assignment is that certain values of the variable are associated with one of the treatments and other values of the variable with the other treatment. Suppose, for example, we believe that groups tip more generously than individuals. Then if groups turn out to be significantly more frequent than individuals when the server is standing, and vice versa when the server is squatting, then "group/individual" would be a confounding variable.

e The random assignment of customers to the treatments makes the issue described in Part (d) unlikely, so long as a large number of customers is used in the experiment.

2.77 **a** The design could be completely randomized or could involve blocking. The following is a completely randomized design.

Divide the plot into a 4 by 4 grid consisting of 16 equally sized square subplots. Number the subplots 1-16. Use a random number generator to select integers between 1 and 16 inclusive. Ignoring repeats, the subplots represented by the first four integers will receive undisturbed native grasses. The subplots represented by the following four integers will receive managed native grasses. The subplots represented by the following four integers will receive undisturbed nonnative grasses. The remaining four subplots will receive managed nonnative grasses.

b Some possible confounding variables are the amount of light a subplot receives, the amount of moisture in a subplot, and whether or not a subplot is on the boundary of the grid. (One of these variables, amount of light, for example, will actually *be* a confounding variable if one particular type of grass is assigned to subplots with more light than the other types of grass.)

c This is an experiment, since the treatments (the different types of grass) are assigned to the subplots by the investigators, rather than using areas of land that already have the types of grass mentioned.

2.78 **a** Answers will vary. The most basic experimental design would be as follows. First number the 100 locations in the kiln. Randomly assign 50 of the locations to receive the first type of clay. The remaining 50 locations will receive the second type of clay. Fire the tiles in the kiln. After firing is complete, compare the proportions of cracked tiles for the two types.

Alternatively, select 50 pairs of locations where the temperature is expected to be the same for the two locations in each pair. For each pair, randomly assign one location to the first type of clay and the other location to the other type of clay. Fire the tiles and compare the proportions of cracked tiles for the two types.

b In the first design, the extraneous variable temperature is dealt with by randomly assigning the locations to the clay types. In the second design it is dealt with by blocking by temperature.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

2.79 **a** There are many possible designs. We will describe here a design that blocks for the day of the week and the section of the newspaper in which the advertisement appears. For the sake of argument we will assume that the mortgage lender is interested in advertising on only two days of the week (Monday and Tuesday) and that there are three sections in the newspaper (A, B, and C). We will refer to the three types of advertisement as Ad 1, Ad 1, and Ad 3.

The experimental units are 18 issues of the newspaper (that is, 18 dates) consisting of Mondays and Tuesdays over 9 weeks. Use a random process to decide which three Mondays will receive advertisements in Section A, which three Mondays will receive advertisements in Section B, and which three Mondays will receive advertisements in Section C. Do the same for the nine Tuesdays. We have now effectively split the 18 issues into the six blocks shown below. (There are 3 issues in each block.)

Mon, Sect A	Mon, Sect B	Mon, Sect C
Tue, Sect A	Tue, Sect B	Tue, Sect C

Now randomly assign the three issues in each block to the three advertisements. (Ad 1 is then appearing on three Mondays, once in each section, and on three Tuesdays, once in each section. The same applies to Ad 2 and Ad 3.) The response levels for the three advertisements can now be compared (as can the three different sections and the two different days).

b Within each block, the three issues of the newspaper were randomly assigned to the three advertisements.

Online Exercises

2.80 **a** "forests are being destroyed…80 acres per minute"

b **"**vanishing tropical forests"

c **"**man-made extinction"

d "destruction of tropical forests"

2.81 Answers will vary.

2.82 Answers will vary.

2.83 Answers will vary.

2.84 Answers will vary.

2.85 Answers will vary.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

ACTIVITY 2.1: DESIGNING A SAMPLING PLAN

Background
In this activity, you will work with a partner to develop a sampling plan.

Suppose that you would like to select a sample of 50 students at your school in order to learn something about how many hours per week, on average, students at your school spend engaged in a particular activity (such as studying, surfing the web, or watching TV).

1. Discuss with your partner whether you think it would be easy or difficult to obtain a simple random sample of students at your school and then obtain the desired information from all of the students selected for the sample. Summarize your discussion by writing a few sentences explaining why you think it would be easy or difficult.

2. With your partner, decide how you might go about selecting a sample of 50 students from your school that, while it may not truly be a simple random sample, could be reasonably considered to be representative of the population of interest. Write a brief description of your sampling plan and be sure to point out the aspects of your plan that you think make it reasonable to argue that it will be representative.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.1 continued

3. Explain your plan to another pair of students. Ask them to critique your plan, pointing out any potential flaws they see in the plan. Write a brief summary of the comments you received. Now reverse roles, and provide a critique of the plan devised by the other pair.

4. Based on the feedback you received in step 3, would you modify you original sampling plan? If not, explain why this is not necessary. If so, describe how the plan would be modified.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

ACTIVITY 2.2: AN EXPERIMENT TO TEST FOR THE STROOP EFFECT

Background
In 1935, John Stroop published the results of his research into how people respond when presented with conflicting signals. Stroop noted most people are able to read words quickly and that they cannot easily ignore them and focus on other attributes of a printed word, such as text color.

For example, consider the following list of words:

green blue red blue yellow red

It is easy to quickly read this list of words. It is also easy to read the words even if the words are printed in color, and even if the text color is different than the color of the word. For example, suppose that the word "green" is printed in red ink, that the word "blue" is printed in yellow ink, and so on according to the table below. People can read the words in the colored list as quickly as they can read the list that isn't printed in color.

Word	**Color of Ink**
green	red
blue	yellow
red	blue
blue	green
yellow	red
red	green

However, Stroop found that if people are asked to name the text colors of the words in the list above (red, yellow, blue, green, red, green), it takes them longer. Psychologists believe that this is because the person responding has to inhibit a natural response (reading the word) and produce a different response (naming the color of the text).

If Stroop is correct, people should be able to name colors more quickly if they do not have to inhibit the word response, as would be the case if they were shown a list of colored rectangles.

1. Design an experiment to compare times to identify colors when they appear as text to times to identify colors when there is no need to inhibit a word response. Be sure to indicate how randomization is incorporated into your design. What is your response variable? How will you measure it? How many subjects will you use in your experiment and how will they be chosen?

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.2 continued

2. When you are satisfied with your experimental design, carry out the experiment. You will need to construct your list of colored words and a corresponding list of colored bars to use in the experiment. You will also need to think about how you will implement your randomization scheme.

3. Summarize the resulting data in a brief report that explains whether your findings are consistent with the "Stroop Effect".

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

ACTIVITY 2.3: MCDONALD'S AND THE NEXT 100 BILLION BURGERS

Background: The article "Potential Effects of the Next 100 Billion Hamburgers Sold by McDonald's" (*American Journal of Preventative Medicine* [2005]:379-381) estimated that 992.25 million pounds of saturated fat would be consumed as McDonald's sells its next 100 billion hamburgers. This estimate was based on the assumption that the average weight of a burger sold would be 2.4 oz. This is the average of the weight of a regular hamburger (1.6 oz.) and a Big Mac (3.2 oz.). the authors took this approach because

> "McDonald's does not publish sales and profits of individual items. Thus, it is not possible to estimate how many of McDonald's first 100 billion beef burgers sold were 1.6 oz hamburgers, 3.2 oz. Big Macs (introduced in 1968), 4.0 oz. Quarter Pounders (introduced in 1973), or other sandwiches."

This activity can be completed as an individual or as a team. Your instructor will specify which approach (individual or team) you should use.

1. The authors of the article believe that the use of 2.4 oz. as the average size of a burger sold at McDonald's is "conservative", which would result in the estimate of 992.25 million pounds of saturated fat being lower than the actual amount that would be consumed. Explain why the authors' belief might be justified.

2. Do you think it would be possible to collect data that could lead to a value for the average burger size that would be better than 2.4 oz.? If so, explain how you would recommend collecting such data. If not, explain why you think it is not possible.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

ACTIVITY 2.4: VIDEO GAMES AND PAIN MANAGEMENT

Background: Video games have been used for pain management by doctors and therapists who believe that the attention required to play a video game can distract the player and thereby decrease the sensation of pain. The paper "Video Games and health" (*British Medical Journal* [2005]:122-123) states:

> "However, there has been no long term follow-up and no robust randomized controlled trials of such interventions. Whether patients eventually tire of such games is also unclear. Furthermore, it is not known whether any distracting effect depends simply on concentrating on an interactive task or whether the content of games is also an important factor as there have been no controlled trials comparing video games with other distracters. Further research should examine factors within games such as novelty, users' preferences, and relative levels of challenge and should compare video games with other potentially distracting activities."

1. Working with a partner, select one of the areas of potential research suggested in the passage from the paper and formulate a specific question that could be addressed by performing an experiment.

2. Propose an experiment that would provide data to address the question from part 1. Be specific about how subjects might be selected, what the experimental conditions (treatments) would be, and what response would be measured.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.4 continued

3. At the end of Section 2.3 there are 10 questions that can be used to evaluate an experimental design, Answer these ten questions for the design proposed in part 2.

4. After evaluating your proposed design, are there any changes you would like to make to your design? Explain.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

BONUS ACTIVITY 2.5: CLUSTER SAMPLING

Materials needed: One phone book for each group

Background: In this activity, you will work with a group of 5 – 8 students to develop and implement a sampling plan.

Entries in the phone book residential pages might include both an address and a phone number or just a phone number with no address. Suppose we would like to estimate the proportion of entries in the residential pages that do not list an address.

1. It would be very time consuming to take a simple random sample of entries from the residential pages of the phone book. Briefly discuss the difficulties in using simple random sampling for sample selection.

2. Another approach is to regard the residential part of the phone book as a collection of pages. We could then randomly select pages and include all entries on the selected pages in the sample. This method is called cluster sampling—the population is divided into subgroups called clusters (in this activity the phone book pages define the clusters) and then entire clusters are selected at random. Unlike stratified random sampling, cluster sampling works best when each cluster is as heterogeneous as possible, with each cluster mirroring the population as much as possible. Do you think the cluster sampling method described for sampling entries from the residential pages of the phone book will produce a sample of entries that is representative of the population? Explain

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.5 continued

3. Each individual in your group will select a page at random and then determine (1) the number of entries on that page and (2) the number of entries on the page that do not include an address. Describe the method you will use to select the pages to be examined.

4. Using the method described in step 3, have each team member select a page and determine the following

Number of entries on the page:

Number of entries with no address:

5. Combine the information from all of the individuals in your group in the following table and then compute the entries in the bottom row of the table by calculating the totals for the last two columns.

Team Member	**Number of entries**	**Number without an address**
Total		

6. Use the entries in the last row of the table above to calculate the sample proportion of entries that do not include an address.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.5 continued

7. Compare your group's sample proportion to the sample proportions of the other groups in your class. Are the proportions similar or do they differ substantially? Write a few sentences describing why the sample proportions from different groups are not all the same.

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

BONUS ACTIVITY 2.6: SPEED SORTING

Materials needed: 30 index cards, one die, and a stopwatch or a watch with a second hand for each group.

Background: A researcher has speculated that it will take longer for a person to sort large numbers than small numbers. To investigate, the following experiment has been proposed:

- Write the numbers 0 to 9 on index cards (one number per card). Write the numbers 1040 to 1049 on ten cards and the numbers 6070 to 6079 on ten cards. Keep the cards in three stacks, one for 0-9 (Stack 1), one for 1040-1049 (Stack 2), and one for 6070-6079 (Stack 3).
- For each volunteer, roll a die. If the result is 1 or 2, use the cards from Stack 1. If the result is 3 or 4, use the cards from Stack 2 and if the result is 5 or 6, use Stack 3.
- Mix the cards in the selected stack well, and then hand them to the volunteer.
- Ask the volunteer to arrange the cards in numerical order as quickly as possible and record the time required to do so.

1. This experiment compares three different experimental conditions—sorting the number 0-9, sorting 1040-1049, and sorting 6070-6079. Why do you think that the 1040-1049 condition and 6070-6079 condition were both included in the experiment?

2. Why is the random assignment of volunteers to experimental conditions (using the roll of a die) important?

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.6 continued

3. Working in a small group, carry out this experiment by finding 15 volunteers who are not in your class. For each volunteer, be sure to record the time required to sort and which stack of cards was used.

Volunteer	Stack Used	Time to Sort
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

4. Ask two other groups for their data so that you will have data from a total of 45 volunteers.

Stack Used	Time to Sort	Stack Used	Time to Sort	Stack Used	Time to Sort

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual

Activity 2.6 continued

5. Construct a dot plot of the 45 observations, using a different color to represent the times for each of the three experimental conditions. (For example, use blue for stack 1 times, red for stack 2 times, and green for stack 3 times.)

6. Write a few sentences commenting on any interesting patterns (or lack of patterns) that you see in the plot from step 5. In particular, is the data consistent with the researcher's theory?

Introduction to Statistics and Data Analysis 5th edition Peck Solutions Manual