Name:_______________________ Date:_________________________

AP Statistics: Chapter 5 Assignment: (Solution)

1. Match each definition on the right with the term on the left.

i) Observational Studies	I	a) Dividing the population into sub-groups, then choosing a simple random sample from each group.
ii) Stratified Random Sample	A	b) Using Statistics to answer specific questions of interest.
iii) Sampling	K	c) A gathering of information from the entire population
iv) Probability Sampling Design	L	d) Applying a treatment to a group in order to answer a specific question about that treatment's effect(s).
v) Multistage Samples	H	e) When a sample suffers from undercoverage or nonresponse due to the behaviour of the interviewer or respondent.
vi) Experiment	D	f) When every member of the population has an equal chance of being selected.
vii) Bias	G	g) A systemic error in the way the sample represents the population
viii) Statistical Inference	B	h) Choosing successively smaller groups within the population.
ix) Census	C	i) Asking questions about pre-existing groups and/or situations
x) Response Bias	E	j) A group of people who share a common characteristic.
xi) Sample Design	M	k) Selecting a part of a population to answer a question about the whole population.
xii) Population	J	l) Using random chance to select a sample.
xiii) Simple Random Sample	F	m) The method by which the sample is selected.

2a) Why is this an observational study and not an experiment?

The student can appeal to any of the three reasons in judging this study is not an experiment

1. There is no random assignment of subjects to treatments
2. There are no imposed treatments
3. Existing data is used

b) Two varaibles are confounded if their effect on the number of new cavities cannot be distinguished from one another. The student must mention not only that the confounding variables may affect the outcome but that they have differential effects within two groups. For instance: confounding would occur if patients who eat an apple a day differ from those who eat less than one apple a week on some variable that is related to dental health. In this example, diet or general level of health are examples of what might be confounding variables. For example, it is possible that people who eat an apple a day are more nutrition conscious and have a more health diet in general than those who eat one or fewer apples per week, and this might explain the observed difference in dental health.

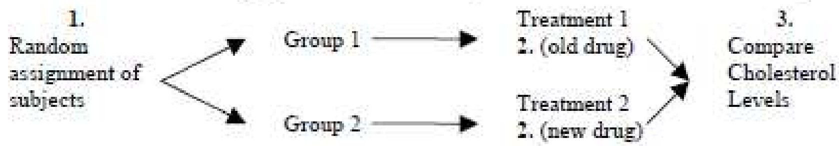
3.

(a) Describes an experimental design that includes:

1. Random assignment of volunteers to the treatment groups
2. Identification of treatment groups as old drug and new drug
3. Indication that a comparison or measurement of cholesterol levels should be made

OR

The student may give a detailed diagram that addresses the three parts:



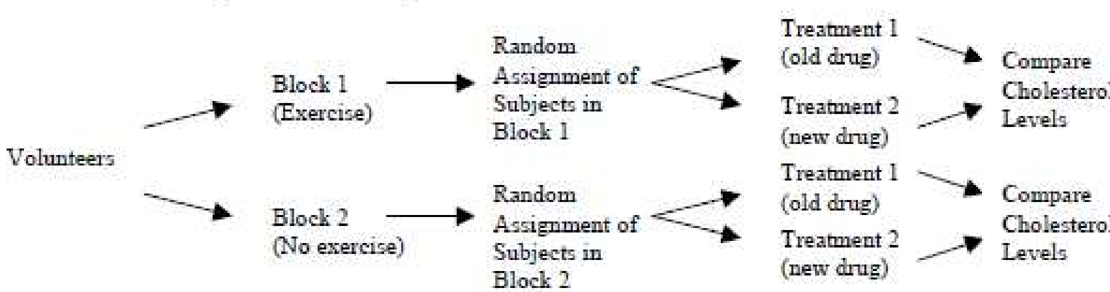
Note: In part (a), it is incorrect to use the terminology "treatment" and "placebo" for the treatment groups. It is considered correct to use "old drug" and "new drug", and "placebo," if a third group is used, for the treatment groups.

(b) Describes an experimental design that includes:

1. Creating blocks based on level of exercise or cholesterol level, or creating blocks using age, diet, gender, or any other factor plausibly related to cholesterol level **with explanation** (i.e., block on gender because males and females may respond differently)
2. Random assignment of subjects to treatments within blocks

OR

The student may give a detailed diagram that addresses the two parts as long as the blocking factor is described.



Note: No credit will be given in part (b) if a student does not use blocking in his/her design even though they randomize correctly.

Note: Crossover designs or matched-pairs designs that incorporate the idea of blocking are acceptable.

(c) Clearly explains a double blind experiment—neither the subjects nor those administering the drugs or monitoring results know which of the two drugs is being used.

An answer of yes without explanation receives no credit.

An answer of no could receive credit if the design described in part (b) does not allow for double-blinding.

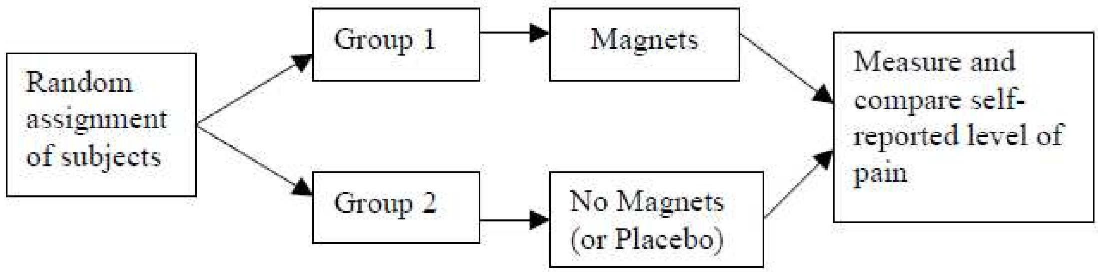
4.

Solution

Part (a):

1. Two treatments: magnets and no magnets (or magnets and placebo). Subjects in the no magnet group would be handled in the same way as the magnet group, but there would be no magnets embedded in the pads used.
2. There must be random assignment of subjects to treatments (or treatments to subjects). How the randomization would be carried out does not need to be specified, but it must be clear *what* is being randomized.
3. Variable measured: Self-reported level of pain or reduction in pain.

The design may be described by a diagram, but the treatments and the variable measured must be included and the randomization must be very clear.

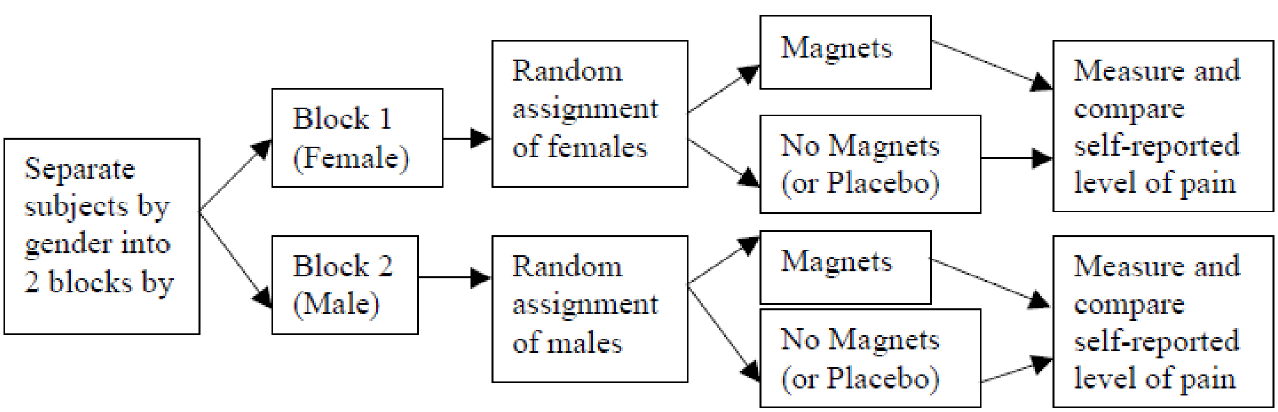


Part (b): Either one of the following approaches is acceptable.

1. Saying yes and indicating how they would alter the design: Separating the subjects into the two gender groups and then randomizing subjects to treatments within each group. This may also be described using a diagram, as shown below, but the blocking factor and randomization must be clearly indicated.

   OR

2. Saying no and describing why. For example, indicating that the randomization in (a) should equalize the effects of gender in the two groups or assuming gender does not have a strong effect and since the sample size is large
   OR
   providing a good explanation for *why* gender does not have a strong differential effect on the outcome.



5.

Intent of Question

The primary goals of this question are to assess a student's ability to: (1) identify the response variable, treatments, and experimental units in a study; (2) critique the use of randomization and replication; (3) recognize and explain why a particular variable is a confounding variable.

Solution

Part (a):

The response variable was the amount of draft. The two treatments were the standard hitch and the new hitch. The experimental units were the two large plots of land.

Part (b):

Yes, the two hitches (treatments) were randomly assigned to the two plots (experimental units).

Part (c):

No, each treatment (type of hitch) was applied to only one experimental unit (plot of land). Replication is used to repeat the treatments on different experimental units so general patterns can be observed. There is no replication in this study.

Part (d):

Although 25 measurements were taken at different locations in the two plots, each hitch was used in one plot (experimental unit) only. Thus, if a difference in the draft is observed we will not know whether the difference is due to the hitch or the plot. In statistical language, the treatments (hitches) are confounded with the plots.

5. When a tractor pulls a plow through an agricultural field, the energy needed to pull that plow is called the draft. The draft is affected by environmental conditions such as soil type, terrain, and moisture.

A study was conducted to determine whether a newly developed hitch would be able to reduce draft compared to the standard hitch. (A hitch is used to connect the plow to the tractor.) Two large plots of land were used in this study. It was randomly determined which plot was to be plowed using the standard hitch. As the tractor plowed that plot, a measurement device on the tractor automatically recorded the draft at 25 randomly selected points in the plot.

After the plot was plowed, the hitch was changed from the standard one to the new one, a process that takes a substantial amount of time. Then the second plot was plowed using the new hitch. Twenty-five measurements of draft were also recorded at randomly selected points in this plot.

(a) What was the response variable in this study?

Identify the treatments.

What were the experimental units?

(b) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was randomization used properly in this study? Justify your answer.

(c) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was replication used properly in this study? Justify your answer.

6. (d) Plot of land is a confounding variable in this experiment. Explain why.

Solution:

a. The student can appeal to any of three reasons in judging this study not an experiment:

1. there is no random assignment of subjects to treatments;
2. there are no treatments imposed;
3. existing data is being used.

b. Two variables are confounded if their effect on the number of new cavities cannot be distinguished from one another. The student must mention not only that the confounding variables may affect the outcome but that they have differential effects within the two groups. For instance: confounding would occur if patients who eat an apple a day differ from those who eat less than one apple a week **on some variable that is related to dental health**. In this example, diet or general level of health are examples of what might be confounding variables. For example, it is possible that people who eat an apple a day are more nutrition conscious and have a more healthy diet in general than those who eat one or fewer apples per week, and this might explain the observed difference in dental health.

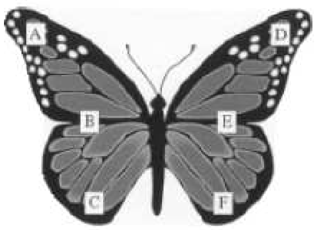
Note:
There are many possible examples of confounding variables. Any reasonable example of a confounding variable is acceptable, as long as a good explanation is given and the connection between the confounding variable and group membership is clear. Lack of a definition here can be rectified by a response in (c) that demonstrates a clear understanding of the concept of confounding variable.

c. No, because it is not an experiment, and cause-and-effect conclusions cannot be drawn from an observational study.

OR

No, because there are possible confounding variables.

1. Researchers often mark wildlife in ordre to identify particular individuals across time or space. A study of butterfly migration is designed to determine which lcoation on the butterlies' wings is best for marking. The six possible locations are those shown as A through F in the figure below. The butterfly in the figure is a monarch. Because marks in certain locations may be more likly to attract predators or cause problems than marks in other locations, the goal is to determine whether the six marking lcoations result in equivalent chances of successful migrations. To test this, researchers plan to mark 3,600 butterflies and release them, then count how many arrive displaying each marking location at the end of the migratory path.



A) Briefly describe a method you could use to assign the marking locations if you wanted to ensure that exactly 500 butteflies were marked in each location

B) Briefly describe a method you could use to assign the marking locations if you wanted to be independent from one butterfly to the next, and wanted each lcoation assigned with a probability 1/6 each time.

Solution:

(a) A complete response should try to make the six treatment groups look as nearly alike as possible, except for the treatment, so as to balance out any characteristics (seen or unseen) that might affect chances of successful migration.

Some ways to do this are as follows:

- Randomly allocate the 3600 butterflies to groups of 600 for each marking location. A correctly described method does not necessarily have to be practical. Although it would be difficult, the 3600 butterflies could be each assigned a unique number and then a random number table could be used to select 600 for position A, and so on. A more practical randomization could be accomplished by putting 3600 slips of paper n a bag, 600 for each letter. As each butterfly is captured, a slip is randomly selected from the bag (without replacement) to indicate the marking location for the butterfly.
- Divide the 3600 butterflies into groups smaller than 600 and then randomly assign an equal number of groups to each marking location.
- Systematically assign marking locations A through F to successive groups of 6 butterflies until all 3600 have been assigned.

Note: Responses for (a) should be read carefully to assess understanding of randomness. For example, only indicating an allocation of 600 butterflies per location is an incorrect response. Only stating that 600 butterflies are randomly assigned to each marking location is an incomplete response.

(b) Generate six possible outcomes independently, each with probability 1/6, perhaps by tossing a balanced die. Assign one number to each location, such as 1 = A, 2 = B, and so on. Toss the die and assign the marking location sequentially to the 3600 butterflies as they are caught. The order of assignment does not matter.

5. High cholesterol level in people can be reduced by exercise or by drug treatment. A pharmaceutical company has developed a new cholesterol-reducing drug. Researchers would like to compare its effects to the effects of the cholesterol-reducing drug that is currently available on the market. Volunteers who have a history of high cholesterol and who are currently not on medication will be recruited to participate in a study.

(a) Explain how you would carry out a completely randomized experiment for the study.

(b) Describe an experimental design that would improve the design in (a) by incorporating blocking.

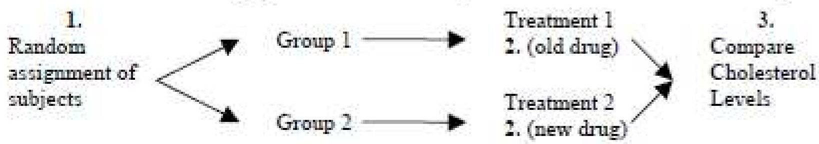
(c) Can the experimental design in (b) be carried out in a double blind manner? Explain.

Solution:

(a) Describes an experimental design that includes:
1. Random assignment of volunteers to the treatment groups
2. Identification of treatment groups as old drug and new drug
3. Indication that a comparison or measurement of cholesterol levels should be made

OR

The student may give a detailed diagram that addresses the three parts:

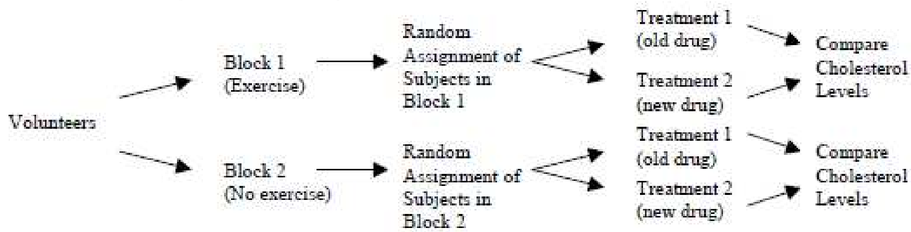


Note: In part (a), it is incorrect to use the terminology "treatment" and "placebo" for the treatment groups. It is considered correct to use "old drug" and "new drug", and "placebo," if a third group is used, for the treatment groups.

(b) Describes an experimental design that includes:
1. Creating blocks based on level of exercise or cholesterol level, or creating blocks using age, diet, gender, or any other factor plausibly related to cholesterol level **with explanation** (i.e., block on gender because males and females may respond differently)
2. Random assignment of subjects to treatments within blocks

OR

The student may give a detailed diagram that addresses the two parts as long as the blocking factor is described.



Note: No credit will be given in part (b) if a student does not use blocking in his/her design even though they randomize correctly.

Note: Crossover designs or matched-pairs designs that incorporate the idea of blocking are acceptable.

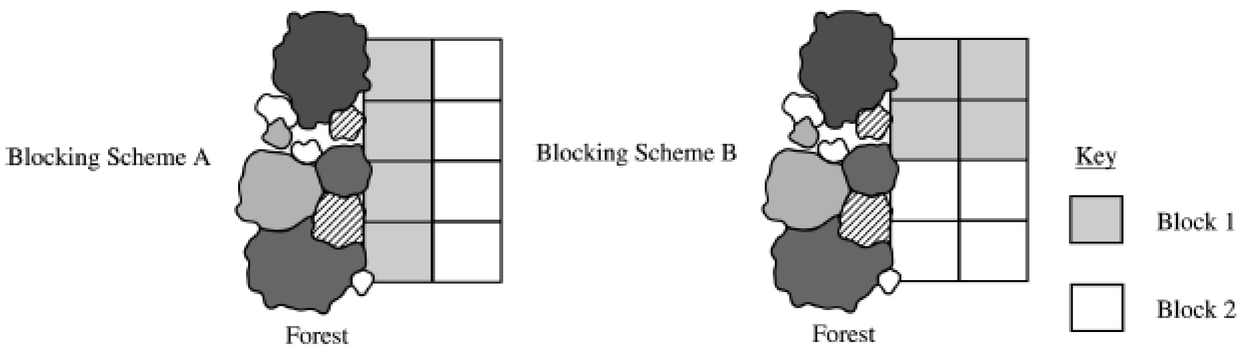
(c) Clearly explains a double blind experiment—neither the subjects nor those administering the drugs or monitoring results know which of the two drugs is being used.

An answer of yes without explanation receives no credit.

An answer of no could receive credit if the design described in part (b) does not allow for double-blinding.

4. Students are designing an experiment to compare the productivity of two varieties of dwarf fruit trees. The site for the experiment is a field that is bordered by a densely forested area on the west (left) side. The field has been divided into eight plots of approximately the same area. The students have decided that the test plots should be blocked. Four trees, two of each of the two varieties, will be assigned at random to the four plots within each block, with one tree planted in each plot.

The two blocking schemes shown below are under consideration. For each scheme, one block is indicated by the white region and the other block is indicated by the gray region in the figures.



(a) Which of the blocking schemes, A or B, is better for this experiment? Explain your answer.

(b) Even though the students have decided to block, they must randomly assign the varieties of trees to the plots within each block. What is the purpose of this randomization in the context of this experiment?

Part (a):

Blocking scheme A is preferable because it creates homogeneous blocks with respect to forest exposure. That is, plots in the same block have similar exposure to the forest.

Part (b):

Randomization of varieties of trees to the plots within each block should reduce any possible bias due to confounding variables, such as fertility or moisture, on the productivity of the two types of dwarf trees.

OR

Randomization of varieties of trees to the plots within each block should even out (or equalize) the effect of other characteristics of the plots that might be related to the productivity of the trees.

2. A manufacturer of boots plans to conduct an experiment to compare a new method of waterproofing to the current method. The appearance of the boots is not changed by either method. The company recruits 100 volunteers in Seattle, where it rains frequently, to wear the boots as they normally would for 6 months. At the end of the 6 months, the boots will be returned to the company to be evaluated for water damage.

   (a) Describe a design for this experiment that uses the 100 volunteers. Include a few sentences on how it would be implemented.

   (b) Could your design be double blind? Explain.

Solution

Part (a):

A **paired design** is used in which each subject receives a pair of boots where one boot is treated with the new method and the other with the current method.

Subjects should be randomly assigned to one of two groups. Group 1 would have the new method applied to the right boot; group 2 would have the new method applied to the left boot.

OR

For each subject, whether the new method is applied to the right or left boot is determined at random.

OR

A **crossover design** is used in which each subject receives a pair of boots, both of which were treated with one treatment. The boots are used for three months and then exchanged for a second pair of boots, both of which were treated with the other treatment. These boots are then used for the next three months.

Subjects should be randomly assigned to one of two groups. One group receives boots with the new treatment first and the other group receives boots with the current method first.

NOTE: Additional appropriate blocking schemes are considered extraneous.

Part (b):

The design could be double blind, as long as both the *subjects* and the person *evaluating* the boots for water damage do not know which boots were treated with the new method and which were treated with the current method.

NOTE: If the student does something unexpected in part (a) and gives a design that actually cannot be double blind, then part (b) could be considered correct provided the response explains why the design could not be double blind.

3. A preliminary study conducted at a medical center in St. Louis has shown that treatment with small, low-intensity magnets reduces the self-reported level of pain in polio patients. During each session, a patient rested on an examining table in the doctor's office while the magnets, embedded in soft pads, were strapped to the body at the site of pain. Sessions continued for several weeks, after which pain reduction was measured.

   A new study is being designed to investigate whether magnets also reduce pain in patients suffering from herniated disks in the lower back. One hundred male patients are available for the new study.

   (a) Describe an appropriate design for the new study. Your discussion should briefly address treatments used, methods of treatment assignment, and what variables would be measured. Do not describe how the data would be analyzed.

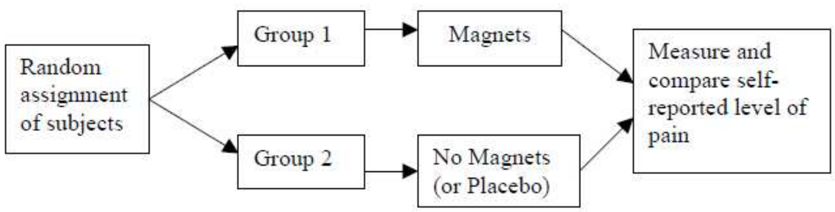
   (b) Would you modify the design above if, instead of 100 male patients, there were 50 male and 50 female patients available for the study? If so, how would you modify your design? If not, why not?

Solution

Part (a):

1. Two treatments: magnets and no magnets (or magnets and placebo). Subjects in the no magnet group would be handled in the same way as the magnet group, but there would be no magnets embedded in the pads used.
2. There must be random assignment of subjects to treatments (or treatments to subjects). How the randomization would be carried out does not need to be specified, but it must be clear *what* is being randomized.
3. Variable measured: Self-reported level of pain or reduction in pain.

The design may be described by a diagram, but the treatments and the variable measured must be included and the randomization must be very clear.

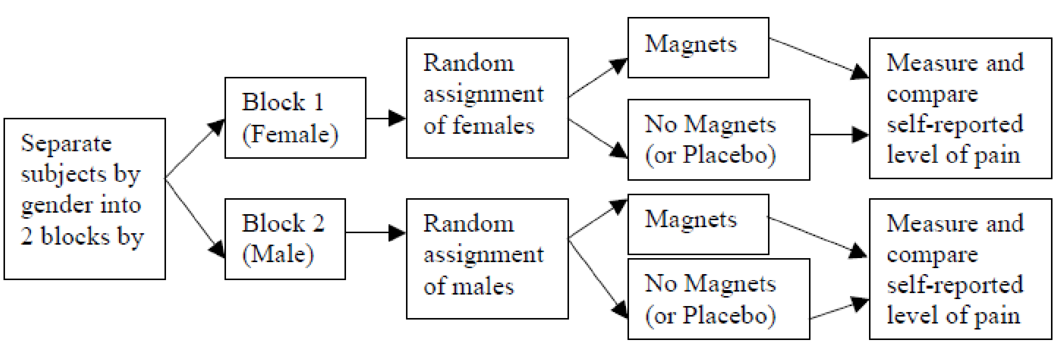


Part (b): Either one of the following approaches is acceptable.

1. Saying yes and indicating how they would alter the design: Separating the subjects into the two gender groups and then randomizing subjects to treatments within each group. This may also be described using a diagram, as shown below, but the blocking factor and randomization must be clearly indicated.

   OR

2. Saying no and describing why. For example, indicating that the randomization in (a) should equalize the effects of gender in the two groups or assuming gender does not have a strong effect and since the sample size is large
   OR
   providing a good explanation for *why* gender does not have a strong differential effect on the outcome.



5. When a tractor pulls a plow through an agricultural field, the energy needed to pull that plow is called the draft. The draft is affected by environmental conditions such as soil type, terrain, and moisture.

A study was conducted to determine whether a newly developed hitch would be able to reduce draft compared to the standard hitch. (A hitch is used to connect the plow to the tractor.) Two large plots of land were used in this study. It was randomly determined which plot was to be plowed using the standard hitch. As the tractor plowed that plot, a measurement device on the tractor automatically recorded the draft at 25 randomly selected points in the plot.

After the plot was plowed, the hitch was changed from the standard one to the new one, a process that takes a substantial amount of time. Then the second plot was plowed using the new hitch. Twenty-five measurements of draft were also recorded at randomly selected points in this plot.

(a) What was the response variable in this study?

Identify the treatments.

What were the experimental units?

(b) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was randomization used properly in this study? Justify your answer.

(c) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was replication used properly in this study? Justify your answer.

(d) Plot of land is a confounding variable in this experiment. Explain why.

Intent of Question

The primary goals of this question are to assess a student's ability to: (1) identify the response variable, treatments, and experimental units in a study; (2) critique the use of randomization and replication; (3) recognize and explain why a particular variable is a confounding variable.

Solution

Part (a):

The response variable was the amount of draft. The two treatments were the standard hitch and the new hitch. The experimental units were the two large plots of land.

Part (b):

Yes, the two hitches (treatments) were randomly assigned to the two plots (experimental units).

Part (c):

No, each treatment (type of hitch) was applied to only one experimental unit (plot of land). Replication is used to repeat the treatments on different experimental units so general patterns can be observed. There is no replication in this study.

Part (d):

Although 25 measurements were taken at different locations in the two plots, each hitch was used in one plot (experimental unit) only. Thus, if a difference in the draft is observed we will not know whether the difference is due to the hitch or the plot. In statistical language, the treatments (hitches) are confounded with the plots.