

Workshop Statistics: Discovery with Data, Second Edition

Topic 13: Designing Studies

Activity 13-7: Pet Therapy

- (a) The explanatory variable is pet ownership. The response variable is the recovery rate of the patients (whether or not the patient recovers).
- (b) This is an observational study. The researcher passively observes and records information on the patients' recovery. To be an experiment the researcher would've needed to determine who owned a pet.
- (c) Yes, there is a group without pets to compare to.
- (d) No, this study does not make use of randomization. The researcher did not randomly assign the patients to different treatment groups.
- (e) No, we cannot conclude that owning a pet has a therapeutic effect for heart attack survivors. There may be lurking variables that will have an effect on the response variable. We cannot conclude causation with an observational study.
- (f) This study could be a controlled experiment if the researcher used randomization to determine whether or not the patient owned a pet. In this case the researcher would *actively* impose the treatment on the subjects. The experimenter would then hope to see the direct effect of pet ownership on the recovery rate of heart attack patients.
- (g) This is debatable – is it feasible for us to tell someone to own a pet or not?

Activity 13-8: Foreign Language and SAT (*cont.*)

- (a) A controlled study would need to randomly assign students to different treatment groups, i.e.: foreign language study or no foreign language study. This will ensure that hidden confounding variables (i.e.: verbal aptitude and foreign language) will balance out in the long run between treatment groups.
- (b) It may not be ethical to carry out this study because it may have a negative effect on some students' SAT scores and we may not be able to control which classes they take.

Activity 13-9: Winter Heart Attacks

- (a) This is an observational study because researchers *passively* observed and recorded information on the heart attack deaths in Los Angeles.
- (b) A possible confounding variable could be the weather. An alternative explanation could be that the extremely cold winters on the east coast may have had an effect on the amount of heart attacks occurring in December and January.
- (c) The Los Angeles study does not prove that holiday stress causes an increase in holiday deaths. A possible confounding variable could be shorter days during the winter (less sunlight).

Activity 13-10: Friendly Observers

- (a) This study is an experiment since the researcher actively imposed which type of observer was present.
- (b) This study makes use of comparison because two different treatment groups were compared by the researcher, i.e.: participant only wins \$3 or Participant and observer win \$3.
This study makes use of randomization because the subjects were randomly assigned to two different treatment groups.
This study makes use of blindness because the subjects were not told whether or not they received the real treatment or the placebo; i.e.: the subjects were not told which group they were in.
- (c) Time to complete the game: quantitative variable. Whether or not the threshold was beaten: categorical binary variable Whether or not the observer was to share in the prize or not: categorical binary variable.
- (d) Time to complete the game: response variable. Whether or not the threshold was beaten: response variable Whether or not the observer was to share in the prize or not: explanatory variable.
- (e) Group A: $3/12 = 0.25$ Group B: $8/11 = 0.727$
- (f) Yes, the researchers conjectured that having an observer in the room will decrease a subject's performance on a skill based task (i.e.: playing a video game). There is evidence of this since group A performed worse than group B.
- (g) On the table, randomly assign the letters (the subjects) to either group one or group two. You could look up two digit numbers between 01 and 23. When one of these numbers comes up, assign that subject to group one. Do this until 12 subjects have been obtained. The rest of the subjects will be assigned to group B.
- (h) A disadvantage of flipping a coin to assign groups is that there is no way to assure an even number in both groups. For example, heads may come up 19 times and tails may only come up 4 times.

Activity 13-11: Children's Television Viewing (*cont.*)

- (a) This could be considered an experiment because the schools *activity imposed* the curriculums at the two schools (instead of letting the students choose their own). However, it is not a well-designed experiment since students were not randomly assigned to the two schools.
- (b) The explanatory variable is the amount of television viewing. This is a quantitative variable.
- (c) The findings in this study are more conclusive than if it were an observational study because is a controlled experiment, and we can measure the change in response after the curriculum. The researchers used several forms of control to help minimize the effects of extraneous variables on the response variable. However, it would be hard to conclude "cause and effect" since there may be some confounding variables related to which school a child attends, such as social economic status, that also affects the response.

Activity 13-12: Baldness and Heart Disease (*cont.*)

- (a) Occurrence of heart attacks: response variable; Baldness: explanatory variable
- (b) This is an observational study because the researchers *passively* observed and recorded information on the subjects' baldness. (As opposed to imposing baldness on people!)
- (c) You cannot conclude that baldness causes heart disease from this study since this was not a randomized experiment. There may be many confounding variables that also affect the response variable such as age: older men will have more baldness and also more incidences of heart attacks.

Activity 13-13: Gender and Lung Cancer (*cont.*)

- (a) This is an observational study because the researchers *passively* screened and recorded data on the smokers' health (as opposed to imposing the gender variable).
- (b) The observational units are the 1000 smokers older than 60 used in this study.
- (c) Gender: explanatory variable; Occurrence of lung cancer: response variable
Note, there are typos in these data. The problem should read: 459 women and 541 men with 19 of the women and 10 of the men suffering from lung cancer.
- (d) Men with cancer: $19/551 = .034$ Women with lung cancer: $29/459 = .063$
 $19/459 = .041$; $10/541 = .018$
- (e) **Answers will vary.** Potential confounding variables include diet and tendency to get check-ups for cancer (suppose men have better diets or suppose women are more likely to get check ups so the cancer is detected more often). Other biological or genetic factors could also be confounding variables here as well, but need to argue why they behave differently for men and women.

Activity 13-14: Reducing Cold Durations

- (a) This is an experiment because the researchers *actively* imposed different treatments of the patients.
- (b) The experimental units are the 104 subjects reporting to the lab with colds.
- (c) Amount of zinc nasal spray: explanatory variable Duration of cold: response variable
- (d) The researchers used blindness to control the experiment. The subjects were not told whether or not they were receiving the real treatment or the placebo to ensure that they would not respond differently due to the researcher's perceived expectations.

Activity 13-15: Religious Lifetimes

- (a) This is an observational study because the researchers *passively* observed and recorded information on people's religious practices and lifetimes, without imposing whether or not they attend services.
- (b) Attendance of religious ceremonies: explanatory variable; Life span: response variable
- (c) You cannot conclude that attending religious services will lengthen your life. A possible confounding variable may be health. Perhaps people who attend religious ceremonies take better care of their bodies, which may affect their life span.

(d) No, the size of the sample is the important consideration, not the relative size of the population.

Activity 13-16: Natural Light and Achievement (*cont.*)

(a) Researchers will randomly assign the students to two different treatment groups: 1 with high natural light, 1 with low natural light. Then the researchers will compare the responses of the two treatments.

(b) It would be difficult to carry out this experiment because there are ethical considerations that could prevent us from depriving students of natural light and also from possibly detrimentally affecting their education.

(c) John B. Lyons could say, "There is a causal relationship between daylight and achievement" if this was a well designed, controlled experiment.

Activity 13-17: Memory Experiment

(a) This is an experiment because the subjects were randomly assigned to two treatment groups, 1 with organized chunks of letters, and 1 without organized chunks of letters. Then the researchers compared the responses of the two treatment groups.

(b) Each student randomly received 1 of 2 different memory cards from the instructor. This is important as randomization should equalize many of the extraneous variables (e.g., people with good vs. poor memories) between the two groups. Thus, any differences we later observe between the two groups can be attributed to the treatment.

(c) - (g) **Answers will vary**

Activity 13-18: SAT Coaching

(a) Coaching program: explanatory variable; SAT score: response variable

(b) This is an observational study. The researcher *passively* observed and recorded information on the student's SAT scores instead of imposing whether or not someone takes an SAT course.

(c) We cannot conclude that the SAT coaching caused the improvements since this was not a controlled experiment. A lurking variable, such as the students' degree of motivation, may have affected the response variable as well as being related to whether or not they take an SAT course.

Activity 13-19: Popcorn Smarts

(a) This can be considered an experiment because the researcher is *imposing* that the students eat popcorn.

(b) No, we cannot conclude that SmartFood is responsible for the improved test scores. A possible confounding variable may be study time. The students may have studied more for the final than the midterm.

(c) The teacher could randomly assign the students to two different treatment groups.

(d) If the students could select their group then we wouldn't know if those who eat popcorn did better because of the popcorn or because of some other variable that was

common to those in the popcorn group.

(e) **Answers will vary but should focus on randomization.** The teacher could randomly assign students a letter, either A or B. Group A could be the popcorn group and Group B could be the SmartFood group.

(f) The students should not know the purpose of the experiment, i.e., the study should be blind. The students in the control group should eat regular popcorn, that way students will not know if they are receiving the real treatment or the placebo.

Activity 13-20: Capital Punishment

(a) No, this is not a controlled experiment because the researcher did not impose which states have the death penalty.

(b) No, we cannot assume that the death penalty has a deterrent effect since this is just an observational study. There may be confounding variables which can also affect the response variable, such as the legal system in the state or the overall crime rate.

(c) No, we cannot make any conclusions about causal relationships because this is an observational study. There could be other variables that are masking the effect of the death penalty.

Activity 13-21: Literature for Parolees

(a) Literature Group: Committed crime: $6/32 = 0.1875$ Not committed crime: $26/32 = 0.815$

(b) Control Group: Committed crime: $18/40 = 0.45$ Not committed crime: $22/40 = 0.55$

(c) The experiment did not randomly assign parolees to the different treatment groups. Instead, qualifications must be met in order to get into the literature program. Perhaps literacy could be a confounding variable that would affect the instances of crime.

Activity 13-22: Parking Meter Reliability

(a) This is an observational study because Ellie *passively* observed and recorded information on parking meter reliability.

(b) Since the meters were randomly selected from Berkeley, we might be willing to generalize to Berkeley. However, since they were not randomly selected from all California parking meters, we wouldn't be willing to generalize the results to this population.

Activity 13-23: Therapeutic Touch

(a) This is an experiment. Emily imposed the treatment (her hand) on the subjects.

(b) Emily flipped a coin to decide which hand she would hold hers over.

(c) This study was not double blind. Emily was aware of which subjects received which treatment.

(d) No, Emily's sample was volunteers, not a random sample from all practitioners (though even if those who thought they could prove themselves didn't...)

Activity 13-24: Effectiveness of Gasoline Additive

- (a) Number the cars 1-30 and then look up two digits numbers until you have 18 different ones. These cars will be the treatment group, the rest will be the control group.
- (b) We could essentially block on each car. Have each car drive both with and without the additive. This allows us to make the comparison on each car, eliminating many of the other extraneous variables. We would randomly determine which gas they used first (some would use additive first, some would go without the additive first).

Activity 13-25: Prayers and Cell Phones

Answers will vary.

For example, (i) could randomly divide subjects into two groups. Have one group talk on the cell phone while driving, the other group not use a cell phone. Then look at the accident rates or compare performance on an obstacle course to see if the two groups behave differently. An observational study would not allow us to include causation. For example, those who use a cell phone may be less careful drivers in general and would get into more accidents anyway, regardless of the cell phone use.

Activity 13-26: Survey of Student Opinion

- (a) - (b) Answers will vary.
- (c) If everyone takes a random sample then the results from each sample should be similar from sample to sample.
- (d) This is an experiment because we imposed which version of the question they answered to see if the wording had an effect on their responses.

Activity 13-27: Survey of Student Opinion (cont.)

- (a) - (b) Answers will vary.
- (c) If we are told something is true, or supposed to make us feel better, we tend to believe it.
- (d) This is an experiment because we imposed which version of the question they answered to see if the wording had an effect on their responses.