

Review of Part III – Gathering Data**R3.1. What design?**

The researchers performed a prospective observational study, since the children were identified at birth and examined at ages 8 and 20. There were indications of behavioral differences between the group of “preemies”, and the group of full-term babies. The “preemies” were less likely to engage in risky behaviors, like use of drugs and alcohol, teen pregnancy, and conviction of crimes. This may point to a link between premature birth and behavior, but there may be lurking variables involved. Without a controlled, randomized, and replicated experiment, a cause-and-effect relationship cannot be determined.

R3.2. What design?

A retrospective observational study was performed. There may be a link between tea drinking and survival after a heart attack. Other variables, like overall health and education might also be involved. Since lurking variables may be involved, a controlled, randomized, and replicated experiment must be performed to determine whether or not a cause-and-effect relationship exists between tea drinking and post heart attack survival.

R3.3. What design?

The researchers at the Purina Pet Institute performed an experiment, matched by gender and weight. The experiment had one factor (diet), at two levels (allowing the dogs to eat as much as they want, or restricted diet), resulting in two treatments. One of each pair of similar puppies was randomly assigned to each treatment. The response variable was length of life. The researchers were able to conclude that, on average, dogs with a lower-calorie diet live longer.

R3.4. What design?

The officials used a random sample. The population is all homes on the property tax list. The parameter of interest is level of radon contamination. The officials’ procedure is not clear, but if they make an effort to get some houses from each area in the city, the sample is stratified by area. If the procedure is followed carefully, the officials can use the results of the sample to make inferences about the radon levels in other houses in the county.

R3.5. What design?

This is a completely randomized experiment, with the treatment being receiving folic acid or not (one factor, two levels). Treatments were assigned randomly and the response variable is the number of precancerous growths, or simply the occurrence of additional precancerous growths. Neither blocking nor matching is mentioned, but in a study such as this one, it is likely that researchers and patients are blinded. Since treatments were randomized, it seems reasonable to generalize results to all people with precancerous polyps, though caution is warranted since these results contradict a previous study

R3.6. What design?

The research team performed a retrospective observational study. There is evidence that the date of first flowering has generally advanced over the last 47 years, but there may be other variables besides climate change that can account for this. The assertion of the researchers is speculative.

R3.7. What design?

The fireworks manufacturers are sampling. No information is given about the sampling procedure, so hopefully the tested fireworks are selected randomly. It would probably be a good idea to test a few of each type of firework, so stratification by type seems likely. The population is all fireworks produced each day, and the parameter of interest is the proportion of duds. With a random sample, the manufacturers can make inferences about the proportion of duds in the entire day’s production, and use this information to decide whether or not the day’s production is suitable for sale.

R3.8. What design?

The researchers used a completely randomized experiment, with three treatments that each subject received, so this could be considered matching. The treatments are receiving the spoiler, incorporating the spoiler into the first paragraph, and no spoiler (control treatment). The response variable is the rating that participants gave each story. Blinding is not mentioned. Results should be generalizable to the public, though since only college students were used, caution should be exercised.

R3.9. What design?

This is an observational retrospective study. Researcher can conclude that for anyone's lunch, even when packed with ice, food temperatures are rising to unsafe levels.

R3.10. What design?

The medical researchers performed a retrospective observational study. The data were gathered from pre-existing medical records. The study does not *prove* that there is no long-term risk of prostate cancer associated with having a vasectomy, but it does provide evidence to that effect.

R3.11. What design?

This is an experiment, with a control group being the genetically engineered mice who received no antidepressant and the treatment group being the mice who received the drug. The response variable is the amount of plaque in their brains after one dose and after four months. There is no mention of blinding or matching. Conclusions can be drawn to the general population of mice and we should assume treatments were randomized. To conclude the same for humans would be risky, but researchers might propose an experiment on humans based on this study.

R3.12. What design?

The artisan is performing an experiment. There are 2 factors (glaze type and temperature). The glaze type has 4 levels, and the temperature has 3 levels, resulting in 12 treatments (the different combinations of glazes and temperatures). There is no mention of randomization. The response variable is apparent age of the pottery. Assuming that the evaluator is unbiased, the artisan can make a conclusion about the best combination of glaze and temperature.

R3.13. What design?

The researchers performed an experiment. There is one factor (gene therapy), at two levels (gene therapy and no gene therapy), resulting in two treatments. The experiment is completely randomized. The response variable is heart muscle condition. The researchers can conclude that gene therapy is responsible for stabilizing heart muscle in laboratory rats.

R3.14. What design?

Observational prospective study. Since the study was based on a large number of adults from eight states over a long period of time, it is reasonable to conclude that smoking and bladder cancer are associated.

R3.15. What design?

The orange juice plant depends on sampling to ensure the oranges are suitable for juice. The population is all of the oranges on the truck, and the parameter of interest is the proportion of unsuitable oranges. The procedure used is a random sample, stratified by location in the truck. Using this well-chosen sample, the workers at the plant can estimate the proportion of unsuitable oranges on the truck, and decide whether or not to accept the load.

R3.16. What design?

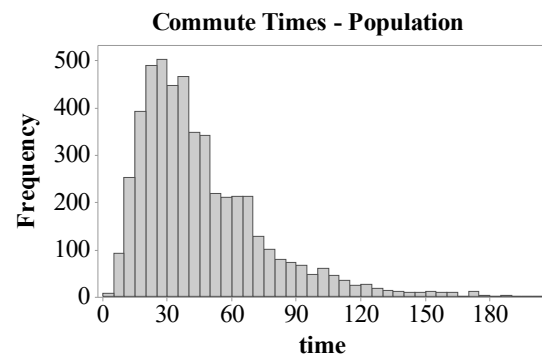
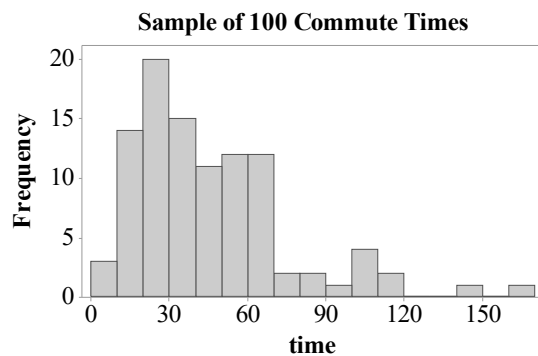
The soft drink manufacture is sampling in order to determine whether or not the machine that caps the bottles is working properly. The population is all of the bottle cap seals. The parameter of interest is the whether or not the bottles are sealing properly. They are using a systematic sample, checking bottles at fixed intervals. If any bottles in the sample are not sealed properly, they can tell that the machine may need adjustment or repair.

R3.17. What design?

Observational retrospective study, performed as a telephone-based randomized survey. Based on the excerpt, it seems reasonable to conclude that more education is associated with a higher Emotional Health Index score, but to insist on causality would be faulty reasoning.

R3.18. What design?

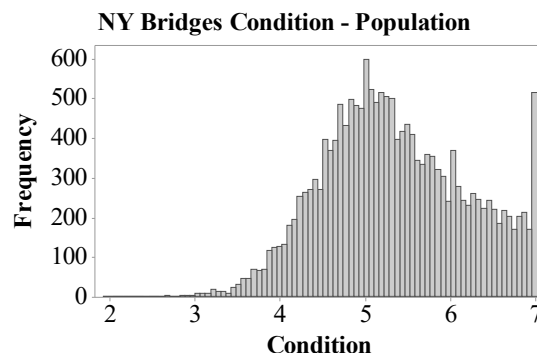
This statistics professor is performing an experiment, blocked by whether or not the students have taken calculus. However, there is probably no randomization, since students usually select their own courses. Hopefully, the two sections contain similar groups of students. There is one factor (use of software), at two levels (software and no software), resulting in two treatments. The response variable is the final exam score. The experiment incorporates blinding, since the graders do not know which students used software and which did not. The professor can decide if computer software is beneficial, and if so, determine whether or not calculus students perform differently than those who have not had calculus.

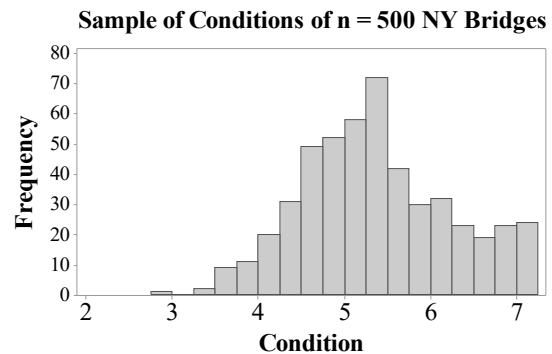
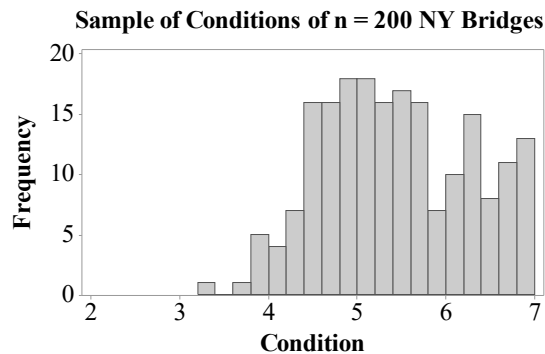
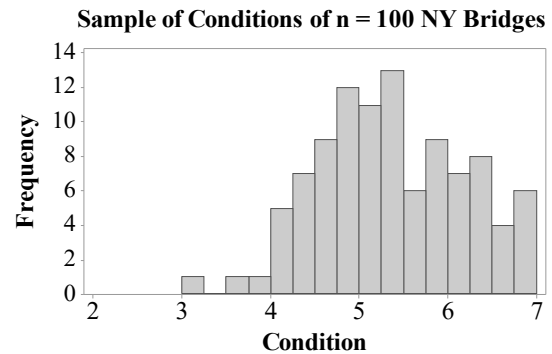
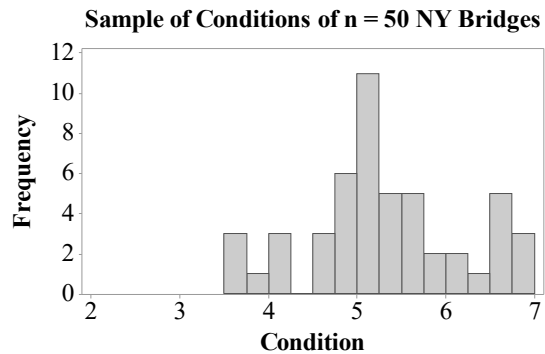
R3.19. Commuter sample.

The sample shows skewness to the right that is in the population. The sample suggests there may be high outliers, but the population has a smooth tail and no outliers.

R3.20. Samples of bridges.

Answers will vary. Results of larger samples will resemble the population distribution more closely. We can make a decent estimate of the center from each histogram, but shape, as well as extreme values (like the mode at 7) are harder to see in the samples.



R3.20. (continued)**R3.21. Alternate day fasting.**

- 100 obese persons.
- Restrictive diet, alternate fasting diet, control (no special diet).
- Weight loss (reported as % of initial weight).
- No, it could not be blinded because participants must know what they are eating.
- No, it is not necessary that participants be a random sample from the population. All that is needed is that they be randomly assigned to treatment groups.

R3.22. Cell phone risks.

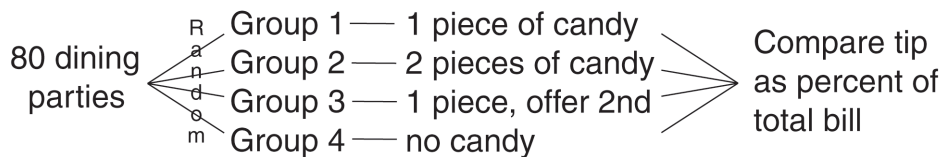
- This is an experiment, since treatments were imposed on randomly assigned groups. There is one factor (radio waves), at three levels (digital cell phone radio waves, analog cell phone radio waves, and no radio waves), resulting in three treatments. The response variable is the incidence of brain tumors.
- The differences in the incidence of tumors between the groups of rats were not great enough to convince the researchers that the differences were due to anything other than sampling variability.
- Since the research was funded by Motorola, there may have been bias present. The researchers may have felt pressure to declare cell phones safe, and introduced bias, intended or not, into the experiment.

R3.23. Tips.

- a) The waiters performed an experiment, since treatments were imposed on randomly assigned groups. This experiment has one factor (candy), at two levels (candy or no candy), resulting in two treatments. The response variable is the percentage of the bill given as a tip.
- b) If the decision whether to give candy or not was made before the people were served, the server may have subconsciously introduced bias by treating the customers better. If the decision was made just before the check was delivered, then it is reasonable to conclude that the candy was the cause of the increase in the percentage of the bill given as a tip.
- c) “Statistically significant” means that the difference in the percentage of tips between the candy and no candy groups was more than expected due to sampling variability.

R3.24. Tips, take 2.

- a) A diagram of the tipping experiment appears below.



- b) This experiment has 1 factor (candy), at 4 levels (1 piece, 2 pieces, 1 piece with an additional piece offered, and no candy).
- c) 1 factor at 4 levels results in 4 treatments.
- d) The response variable is the percent of the total bill left as a tip.
- e) The diners were not aware that they were part of an experiment, so the experiment was blinded. This experiment did not use double-blinding, but there is probably no way to double blind this experiment, since there is no need to blind the evaluator of the response variable. Biased evaluation of the amount of tip left doesn't seem possible.
- f) If the waitress knew which tables were going to receive certain treatments, then she might have treated some tables better than others. The waitress should be unaware of the treatment until after the meal, to avoid the introduction of bias. (Note: This is still only single-blinding. Blinding does not refer to the number of people blinded, but rather the type of blinding employed. If the diners and the waitress are unaware of the assignment of treatments, it is for the same purpose, namely to keep the diners from being systematically influenced.)

R3.25. Timing.

There will be voluntary response bias, and results will mimic those only of the visitors to Sodahead.com and not the general U.S. population. The question is leading responders to answer “yes” though many might understand that the president’s timing for his vacation had nothing to do with the events of the week.

R3.26. Laundry.

- a) Answers may vary. Water (quality and temperature) and material can vary. These confounding variables may influence results. The treatments in the experiment must be in environments that are identical, with the exception of the factor being studied.
- b) These conditions are unrealistic. This will not help the experimenters determine how well *SparkleKleen* will work in the conditions for which it was intended.
- c) If the swatches are stained at the same time, the stains on the swatches washed later will have more time to “set in”, causing bias towards *SparkleKleen*. Also, unforeseen variables, like changes in water temperature or pressure won’t be equalized through randomization.
- d) The conditions under which the *SparkleKleen* was tested are unknown. There is no way to keep the conditions comparable. Furthermore, the company that produced *SparkleKleen* may not produce reliable data. They have a vested interest in the success of their product.

R3.27. How long is 30 seconds?

- a) There are 3 factors, at two levels each. Eyes (open or closed), Music (On or Off), Moving (sitting or moving).
- b) *Subject* is a blocking variable.
- c) Each of 4 subjects did 8 runs, so 32 in all.
- d) Randomizing completely is better to reduce the possibility of confounding with factors that weren’t controlled for.

R3.28. Cookies.

- a) There are 3 factors. Cooking time has two levels (10 or 15 min), temperature had three levels (325°, 375°, and 425°F), and number of chips had two levels (5 or 10).
- b) Blinding means that the person presenting the cookie or the person tasting the cookie doesn’t know which cookie is being tasted. They might have to literally blindfold the judge because some of the factors might be apparent from the cookie itself. It will hard to blind the experimenter, so it is probably not double blinded.
- c) Two levels of cooking time, times three levels of temperature, times two levels of numbers of chips is $2 \times 3 \times 2 = 12$ treatments. Each judge will have to taste 12 cookies.
- d) Rank is an ordered categorical variable.
- e) Answers may vary. Cookies vary by batch, so even if they try to control factors, there may be batch to batch variation. Tasting is very subjective and 12 cookies is a lot to taste.

R3.29. Homecoming.

- a) Since telephone numbers were generated randomly, every number that could possibly occur in that community had an equal chance of being selected. This method is “better” than using the phone book, because unlisted numbers are also possible. Those community members who deliberately do not list their phone numbers might not consider this method “better”!
- b) Although this method results in a simple random sample of phone numbers, it does not result in a simple random sample of residences. Residences without a phone are excluded, and residences with more than one phone have a greater chance of being included.
- c) No, this is not an SRS of local voters. People who respond to the survey may be of the desired age, but not registered to vote. Additionally, some voters who are contacted may choose not to participate.
- d) This method does not guarantee an unbiased sample of households. Households in which someone answered the phone may be more likely to have someone at home when the phone call was generated. The attitude about homecoming of these households might not be the same as the attitudes of the community at large.

R3.30. Youthful appearance.

- a) The differences in guessed age were greater than differences that could be explained by natural sampling variability.
- b) Dr. Weeks is implying that having sex caused the people to have a more youthful appearance. It seems more plausible that younger-looking people are more sexually active than older-looking people, because of their age.

R3.31. Smoking and Alzheimer's.

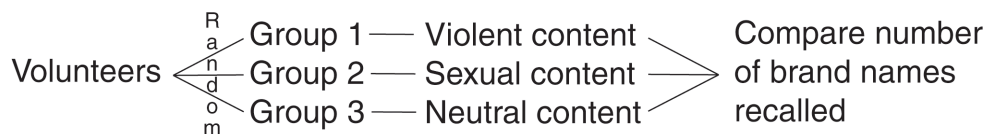
- a) The studies do not prove that smoking offers any protection from Alzheimer's. The studies merely indicate an association. There may be other variables that can account for this association.
- b) Alzheimer's usually shows up late in life. Since smoking is known to be harmful, perhaps smokers have died of other causes before Alzheimer's can be seen.
- c) The only way to establish a cause-and-effect relationship between smoking and Alzheimer's is to perform a controlled, randomized, and replicated experiment. This is unlikely to ever happen, since the factor being studied, smoking, has already been proven harmful. It would be unethical to impose this treatment on people for the purposes of this experiment. A prospective observational study could be designed in which groups of smokers and nonsmokers are followed for many years and the incidence of Alzheimer's disease is tracked.

R3.32. Antacids.

- a) This is a randomized experiment, blocked by gender.
- b) Experiments always use volunteers. This is not a problem, since experiments are testing response to a treatment, not attempting to determine an unknown population parameter. The randomization in an experiment is random assignment to treatment groups, not random selection from a population.
- c) Since the experiment is studying the effects of an antacid, the placebo may actually confound the experiment, since the introduction of *any* substance, even a sugar pill, into the digestive system may have an effect on acid reflux. (The use of some sort of placebo is always recommended, but in some cases it may be difficult to find a placebo that truly has no effect, beyond the expected "placebo effect", of course!)

R3.33. Sex and violence.

This experiment has one factor (program content), at three levels (violent, sexual, and neutral), resulting in three treatments. The response variable is the number of brand names recalled after watching the program. Numerous subjects will be randomly assigned to see shows with violent, sexual, or neutral content. They will see the same commercials. After the show, they will be interviewed for their recall of brand names in the commercials.



R3.34. Pubs.

- a) *Who* – 900 Englishmen; *What* – The researcher was interested in their reasons for going to the pub. *When* – not stated; *Where* – England; *Why* – The producers of Kaliber alcohol-free beer hoped to show that men went to the pub for reasons other than the alcohol. *How* – Researchers surveyed men regarding their reasons for going to the pub.
- b) The researcher surveyed believes that the population is all Englishmen.
- c) The most important omitted detail is the selection process. How did the researcher acquire the sample of men? Was randomness used? Is the sample representative of the population of Englishmen?
- d) Although not stated, it appears that the researcher simply took convenience samples of those in the pubs.
- e) The results may be biased. First of all, an alcohol-free beer producer funded the survey. Respondents may have felt subconscious pressure to indicate that alcohol was not the primary reason for going to the pub. Additionally, admitting that you go to the pub merely for the alcohol is a potentially embarrassing admission. The percentage of pub patrons who go for the alcohol may be significantly higher than 10%.

R3.35. Age and party 2008.

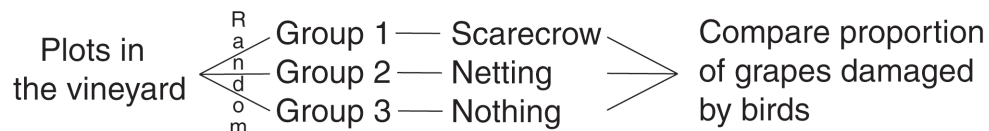
- a) The number of respondents is roughly the same for each age category. This may indicate a sample stratified by age category, although it may be a simple random sample.
- b) 1530 Democrats were surveyed. $\frac{1530}{4002} \approx 38.2\%$ of the people surveyed were Democrats.
- c) We don't know. If data were collected from voting precincts that are primarily Democratic or primarily Republican, that would bias the results. Because the survey was commissioned by NBC News, we can assume the data collected are probably reliable.
- d) The pollsters were probably attempting to determine whether or not political party is associated with age.

R3.36. Bias?

- a) Barone claims that nonresponse bias exists in polls, since conservatives are more likely to refuse to participate than other groups. Nonresponse is less of an issue if it is believed that all groups fail to respond at the same rate.
- b) The population of interest is all adults in the United States.
- c) The column totals do not add up to 100%. There is information missing, and the discrepancies are too large to be attributed to rounding.
- d) The differences observed are similar to differences that may have been observed simply due to natural sampling variability. Differences of this size would be probable, even if no bias exists.

R3.37. Save the grapes.

This experiment has one factor (bird control device), at three levels (scarecrow, netting, and no device), resulting in three treatments. Randomly assign different plots in the vineyard to the different treatments, making sure to ensure adequate separation of plots, so that the possible effect of the scarecrow will not be confounded with the other treatments. The response variable to be measured at the end of the season is the proportion of bird-damaged grapes in each plot.



R3.38. Bats.

Answers may vary. This experiment has one factor (type of bat), at two levels (wooden and aluminum), resulting in two treatments. The response variable is the difference in the distance the ball is hit with each type of bat. Since players vary in their ability to hit the ball, a matched design should be used, with each batter hitting with both types of bats several times in a randomly chosen order. Find the difference for each batter, and then calculate the average difference in distance between the aluminum and wooden bats.

R3.39. Acupuncture.

- a) The “fake” acupuncture was the control group. In an experiment, all subjects must be treated as alike as possible. If there were no “fake” acupuncture, subjects would know that they had not received acupuncture, and might react differently. Of course, all volunteers for the experiment must be aware of the possibility of being randomly assigned to the control group.
- b) Experiments always use volunteers. This is not a problem, since experiments are testing response to a treatment, not attempting to determine an unknown population parameter. The randomization in an experiment is random assignment to treatment groups, not random selection from a population. Voluntary response is a problem when sampling, but is not an issue in experimentation. In this case, it is probably reasonable to assume that the volunteers have similar characteristics to others in the population of people with chronic lower back pain.
- c) There were differences in the amount of pain relief experienced by the two groups, and these differences were large enough that they could not be explained by natural variation alone. Researchers concluded that both proper and “fake” acupuncture reduced back pain.

R3.40 Fuel efficiency.

- a) There are two factors: Tire pressure at two levels (28 and 32 psi) and gasoline type at two levels (regular and premium).
- b) He would need to perform eight runs, two each at each of the four treatment combinations.
- c) Conditions (especially weather) may vary and may confound the effect of the gasoline types. To minimize this, randomize the treatment order.
- d) He would be introducing a blocking variable—the car (actually the car/driver combination).

R3.41. Security.

- a) To ensure that passengers from first-class, as well as coach, get searched, select 2 passengers from first-class and 12 from coach. Using this stratified random sample, 10% of the first-class passengers are searched, as are 10% of the coach passengers.
- b) Answers will vary. Number the passengers alphabetically, with 2-digit numbers. Bergman = 01, Bowman = 02, and so on, ending with Testut = 20. Read the random digits in pairs, ignoring pairs 21 to 99 and 00, and ignoring repeated pairs.

65 43 67 11	27 04
XX XX XX Fontana	XX Castillo

The passengers selected for search from first-class are Fontana and Castillo.

- c) Number the passengers alphabetically, with 3-digit numbers, 001 to 120. Use the random number table to generate 3-digit numbers, ignoring numbers 121 to 999 and 000, and ignoring repeated numbers. Search the passengers corresponding to the first 12 valid numbers generated.

R3.42. Internet speed.

- a) There are five factors, at two levels each—Computer (Mac or PC), Library (Main or Science), Time of Week (Weekday or Weekend), Time of Day (Before or after 5 PM), and Load (Running other jobs or not).
- b) $2 \times 2 \times 2 \times 2 \times 2 = 32$ treatments. They will need to perform 32 runs to run all the treatment levels. However, they will actually need to perform more runs than that, since each treatment combination will need to be replicated a number of times.
- c) They should perform all the runs in random order.