4.2 Observational Studies and Experiments

ADHD Linked to Lead and Mom's Smoking, by Karen Barrow (February 1, 2007)

> *A mother's smoking during pregnancy and exposure to lead significantly increases her child's risk for developing attention deficit hyperactivity disorder (ADHD), say researchers. In fact, as many as one third of cases of ADHD in children are linked to exposure to tobacco smoke and lead before birth, giving moms yet another reason to quit smoking during pregnancy.*
>
> *For the study, researchers from Cincinnati Children's Hospital Medical Center surveyed over 4,700 children between the ages of 4 and 15 and their parents. Over 4 percent of the children included had ADHD. The researchers found that those children whose mother smoked during pregnancy were over twice as likely to develop ADHD than a child whose mother had not smoked. In addition, a child who had been exposed to lead, giving them high lead blood levels, were four times as likely to have ADHD, as compared to a child with low lead levels in his blood.*

Based on this study, should we conclude that smoking during pregnancy *causes* an increase in the likelihood that a child develops ADHD? Explain.

Explain the concept of confounding in the context of this study.

Is there any way to *prove* that smoking causes ADHD?

Read 231-233 *Read word-for-word*

What are some differences between an observational study and an experiment?

What's the difference between an explanatory variable and a response variable?

Designing Experiments

Suppose we wanted to design an experiment to see if caffeine affects pulse rate.

Here is an initial plan:

- measure initial pulse rate
- give each student some caffeine
- wait for a specified time
- measure final pulse rate
- compare final and initial rates

What are some problems with this plan? What other variables are most likely to be sources of variability in pulse rates?

There are several steps we should take to solve these problems.

1. The first step is to include a ____________________________ that does not receive caffeine so we have something to compare to. Otherwise, any pulse-raising (or lowering) event that occurs during the experiment would be confounded with the caffeine. For example, an amazing stats lecture during the waiting period would certainly raise pulse rates, making it hard to know how much of the pulse increase was due to the caffeine.

Read 233-235 *Define in terms of caffeine experiment first, and then read.*

Briefly define the following terms:

- Treatment
- Experimental units
- Subjects
- Factor
- Level

HW: page 230 (37–42) page 253 (45–55 odd)

4.2 Experiments

The caffeine experiment, continued…

2. The second step is to make sure that the two groups (caffeine and non-caffeine) are as similar as possible at the beginning of the experiment.

We ___________________ subjects to treatments to create roughly equivalent groups. The random assignment balances the effects of other variables among the treatments groups. We must ALWAYS randomize since there will always be other variables we cannot control or that we do not consider. Randomizing guards against what we don't know and prevents people from asking "But what about this variable?"

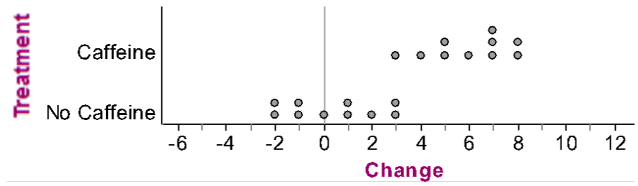
How do we randomize? What is a completely randomized design? (Read 236–238)

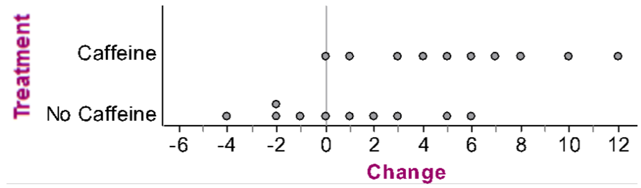
___________________ means ensuring that there are an adequate number of experimental units in each treatment group so that differences in the effects of the treatments can be distinguished from chance differences between the groups.

> Note: Replication can also refer to repeating the experiment with different subjects. This can help us feel more confident applying the results of our experiment to a ___________________.

3. The third step is to make sure that the groups are treated in the exact same manner, except for the treatments. We do this by ___________________ other variables. In other words, we make sure these variables are the same for all groups. This is important for two reasons:

- Prevents confounding: For example, sugar is an important variable to consider because it may affect pulse rates. If one treatment group was given regular Coke (which has sugar) and the other treatment group was given caffeine free Diet Coke (which has no sugar), then sugar and caffeine would be confounded. If there was a difference in the average change in pulse rates of the two groups after receiving the treatments, we wouldn't know which variable caused the change, and to what extent. To prevent sugar from becoming confounded with caffeine, we need to make sure that members of both treatment groups get the same amount of sugar.
- Reduces variability: For example, the amount of soda consumed is important to consider because it may affect pulse rates. If we let subjects in both groups drink any amount of soda they want, the changes in pulse rates will be more variable than if we made sure each subject drank the same amount of soda. This will make it harder to identify the effect of the caffeine (i.e., our study will have less power). For example, the first set of dotplots show the results of a well-done experiment. The second set of dotplots show the results of an experiment where students were allowed to drink as much (or as little) soda as they pleased. The additional variability in pulse rate changes makes the evidence for caffeine less convincing.





Would weight be a confounding variable in this experiment?

It is also important that all subjects in both groups are _________ so that the expectations are the same for the subjects in both groups. Otherwise, members of the caffeine group might suffer from the ____________________. If the people measuring the response are also blind, the experiment is ____________________

Note: Not all experiments have a control group or use a placebo as long as there is comparison. For example, if you are testing a new drug, it is usually compared to the currently used drug, not a placebo. Also, you can do an experiment to compare four brands of paint without using a placebo.

SUMMARY: With randomization, replication, and control, each treatment group should be nearly identical, and the effects of other variables should be about the same in each group. Now, if changes in the explanatory variable are associated with changes in the response variable, we can attribute the changes to the explanatory variable *or the chance variation in the random assignment*.

Read 239–244

Alternate Example: *Dueling diets*
A health organization wants to know if a low-carb or a low-fat diet is more effective for weight loss. The organization decides to conduct an experiment to compare these two diet plans with a control group that is only provided with a brochure about healthy eating. Ninety volunteers agree to participate in the study for three months. Write a few sentences describing how you would implement a completely randomized design for this experiment. Explain how your design incorporates the principles of experimental design. Can your design be double-blind? Why was it important to include a control group?
HW: page 254 (57, 63–71 odd)

4.2 The Caffeine Experiment

Read 244

The results of an experiment are called ______________________________ if they are unlikely to occur by random chance. That is, if it is unlikely that the results are due to the possible imbalances created the random assignment.

For example, if caffeine really has no effect on pulse rates, then the average change in pulse rate of the two groups should be exactly the same. However, because the results will vary depending on which subjects are assigned to which group, the average change in the two groups will probably differ slightly. Thus, whenever we do an experiment and find a difference between two groups, we need to determine if this difference could be attributed to the chance variation in random assignment or because there really is a difference in effect of the treatments.

How can we determine if the results of our experiment are statistically significant?

HW: page 255 (59, 73, 76)

4.2 Blocking

Suppose that a mobile phone company is considering two different "keyboard" designs (A and B) for its new smart phone. The company decides to perform an experiment to compare the two designs using a group of 10 volunteers, where each volunteer will test one of the two designs. The response variable is typing speed, measured in words per minute.

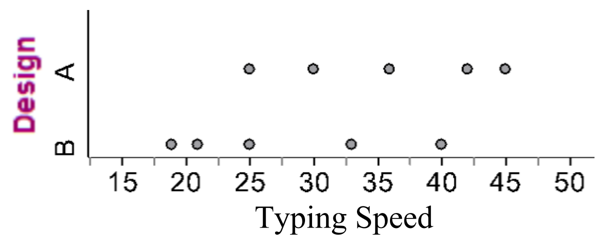
How should the company deal with the fact that some of the volunteers already use a smart phone while the remaining volunteers do not? They could use a completely randomized design and hope that the random assignment distributes the smart phone users and non-smart phone users about evenly between the group using design A and the group using design B. Even so, there might be a lot of variability in typing speed in both groups because some members of each group are much more familiar with smart phones than the others. This additional variability might make it difficult to detect a difference in the effectiveness of the two designs. What should the researchers do?

Because the company knows that experience with smart phones will affect typing speed, they could start by separating the volunteers into two groups—one with experienced smart phone users and one with inexperienced smart phone users. Each of these groups of similar subjects is known as a block. Within each block, the company could then randomly assign half of the subjects to use design A and the other half to use design B. To control other variables, each subject should be given the same passage to type while in a quiet room with no distractions. This randomized block design helps account for the variation in typing speed that is due to experience with smart phones.

Using a randomized block design allows us to account for the variation in the response that is due to the blocking variable. This makes it easier to determine if one treatment is really more effective than the other.

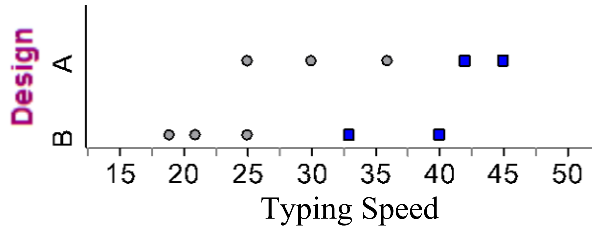
To see how blocking helps, let's look at the results of an experiment using 10 volunteers, 4 who already use a smart phone and 6 who do not. In the block of 4 smart phone users, 2 will be randomly assigned to use design A and the other 2 will be assigned to use design B. Likewise, in the block of 6 non-smart phone users, 3 will be randomly assigned to use design B and the other 3 will be assigned to use design B. Each of the 10 volunteers will type the same passage and the typing speed will be recorded.

Here are the results:

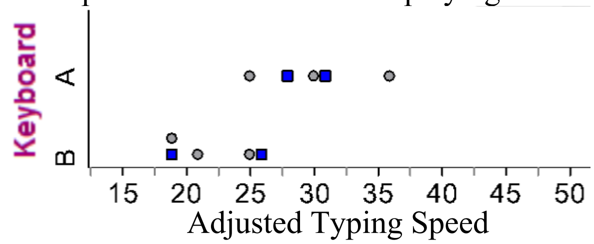


There is some evidence that design A results in higher typing speeds, but the evidence isn't that convincing. There is enough overlap in the two distributions that the differences might simply be due to the chance variation in random assignment.

If we compare the results for the two designs within each block, however, a different story emerges. Among the 4 smart phone users (indicated by the blue squares), design A was the clear winner. Likewise, among the 6 non-smart phone users (indicated by the gray dots), design A was also the clear winner.



The overlap in the first set of dotplots was due almost entirely to the variation in smart phone experience—smart phone users were generally faster than non-smart phone users, regardless of which design they used. In fact, the average typing speed for the smart phone users was 40 while the average typing speed for non-smart phone users was only 26, a difference of 14 words per minute. To account for the variation created by the difference in smart phone experience, let's subtract 14 from each of the typing speeds in the block of smart phone users to "even the playing field." Here are the results:



Because we accounted for the variation caused by the difference in smart phone experience, the variation in each of the distributions has been reduced. There is now almost no overlap between the two distributions, meaning that the evidence in favor of design A is much more convincing. *When blocks are formed wisely, it is easier to find convincing evidence that one treatment is more effective than another.*

Blocking in experiments is similar to stratification in sampling.

- Blocking accounts for a source of variability, just like stratifying. This means that blocking is another form of control.
- Blocks should be chosen like strata: the units within the block should be similar, but different than the units in the other blocks. You should only block when you expect that the blocking variable is associated with the response variable.
- Blocks, like strata, are not formed at random!

What are some variables that we can block for in the caffeine experiment? In general, how can we determine which variables might be best for blocking?

REVISED SUMMARY: Think about all possible sources of variability in the response variable. Control everything you can, block for the things you can measure but can't control, and randomly assign treatments within the blocks to balance out the effects of any remaining variables.

Alternate Example: *Microwave Popcorn*
A popcorn lover wants to know if it is better to use the “popcorn button” on her microwave oven or use the amount of time recommended on the bag of popcorn. To measure how well each method works, she will count the number of unpopped kernels remaining after popping. She goes to the store and buys 10 bags each of 4 different varieties of microwave popcorn (movie butter, light butter, natural, and kettle corn), for a total of 40 bags.

Explain why a randomized block design might be preferable to a completely randomized design for this experiment.

Outline a randomized block design for this experiment.

What is a matched pairs design? Could we use a matched pairs design for the caffeine experiment?

HW: page 257 (78, 80, 81, 87)