**Observational studies and experiments**

- **Observational study:** Data collection does not directly interfere with how the data arise, and can only establish an association between the explanatory and response variables.
- **Experiment:** Researchers randomly assign subjects to various treatments in order to be able to establish causal connections.
- If you’re going to walk away with one thing from this class, let it be “correlation does not imply causation”.

![http://xkcd.com/552/](http://xkcd.com/552/)

---

**Confounding variables**

A study that surveyed a random sample of otherwise healthy adults found that people are more likely to get muscle cramps when they’re stressed. However the study also noted that people drink more coffee and sleep less when they’re stressed.

**What type of study is this?**

**What is the conclusion of the study?**

**Can this study be used to conclude a causal relationship between increased stress and muscle cramps?**

---

**Principles of experimental design**

- **Control:** Compare treatment of interest to a control group.
- **Randomize:** Randomly assign subjects to treatments.
- **Replicate:** Within a study, replicate by collecting a sufficiently large sample. Or replicate the entire study.
- **Block:** If there are variables that are known or suspected to affect the response variable, first group subjects into *blocks* based on these variables, and then randomize cases within each block to treatment groups.
More on blocking

We would like to design an experiment to investigate if increased stress causes muscle cramps:
- Treatment: increased stress
- Control: no or baseline stress
- It is suspected that the effect of stress might be different on younger and older people: block for age.

Why is this important? Can you think of other variables to block for?

Difference between blocking and explanatory variables

- Explanatory variables (factors) are conditions we can impose on the experimental units.
- Blocking variables are characteristics that the experimental units come with, that we would like to control for.
- Blocking is like stratifying, except used in experimental settings when randomly assigning, as opposed to when sampling.

Random assignment vs. random sampling

<table>
<thead>
<tr>
<th>Ideal experiment</th>
<th>Random assignment</th>
<th>No random assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random assignment</td>
<td>Causal conclusion, generalized to the whole population.</td>
<td>No causal conclusion, correlation statement generalized to the whole population.</td>
</tr>
<tr>
<td>Generalizability</td>
<td></td>
<td>Good observational studies</td>
</tr>
<tr>
<td>No random assignment</td>
<td>Causal conclusion, only for the sample.</td>
<td>No causal conclusion, correlation statement only for the sample.</td>
</tr>
<tr>
<td>Causation</td>
<td></td>
<td>Good observational studies</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>Good observational studies</td>
</tr>
</tbody>
</table>

Poll

A study is designed to test the effect of light level and noise level on exam performance of students. The researcher also believes that light and noise levels might have different effects on males and females, so wants to make sure both genders are represented equally under different conditions. Which of the below is correct?

(a) There are 3 explanatory variables (light, noise, gender) and 1 response variable (exam performance)
(b) There are 2 explanatory variables (light and noise), 1 blocking variable (gender), and 1 response variable (exam performance)
(c) There is 1 explanatory variable (gender) and 3 response variables (light, noise, exam performance)
(d) There are 2 blocking variables (light and noise), 1 explanatory variable (gender), and 1 response variable (exam performance)
Application exercise: 1.2 Scientific studies in the press

In this activity you will read a short article and answer a few of question about a study. It is possible that the article doesn’t provide the relevant information to answer certain question(s). If so, refer to the original study (linked). You do not need to read the entire paper, in fact, you are certainly not expected to understand all of it. Simply find the information that will help you answer the questions.

- What are the cases?
- What is (are) the response variable(s) in this study?
- What is (are) the explanatory variable(s) in this study?
- Does the study employ random sampling? How about random assignment?
- Is this an observational study or an experiment? Explain your reasoning.
- Can we establish a causal link between the explanatory and response variables?
- Can the results of the study be generalized to the population at large?

Poll

What is the main difference between observational studies and experiments?

(a) Experiments take place in a lab while observational studies do not need to.
(b) In an observational study we only look at what happened in the past.
(c) Most experiments use random assignment while observational studies do not.
(d) Observational studies are completely useless since no causal inference can be made based on their findings.

Project ideas - observational studies

- **1 numerical**: Is the average number of hours Americans spend relaxing after work different than the European average of 3 hours/day?
  [Data: Number of hours relaxing after work]
- **1 categorical**: Estimate the percentage of North Carolina residents who live below the poverty line and are planning to vote Republican in the most recent presidential election.
  [Data: Vote Republican - yes, no]
- **1 numerical and 1 categorical**: Is there a relationship between mom’s working status during the first 5 years of the child’s life and the child’s education?
  [Data: Number of years of education of child; Mom’s working status - yes, no]
- **2 categorical**: Do racial minority groups in North Carolina have less access to health care coverage?
  [Data: Ethnicity - white, minority; Health coverage - yes, no]

Project ideas - experiments

- **1 numerical and 1 categorical**: Is there a relationship between memory and distraction? Randomly assign 20 students to two groups: one group memorizes a list of words while also listening to music, another group memorizes the same words in silence. Compare average number of words memorized in the two groups.
  [Data: Number of words memorized; Group - treatment, control]
- **2 categorical**: Is there a relationship between learning and distraction? Randomly assign a group of students to two groups: one group studies a concept while also listening to music, the other group studies in silence using the same materials. Then test whether or not they learned the concept.
  [Data: Whether or not the students learned the concept - yes, no; Group - treatment, control]