Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Observational studies: association, confounding, and nightlights example.
2. Observational studies and experiments.
3. Experiments and aspirin example.
4. Random sampling, random assignment, and blocking.

Finish reading chapter 4.
http://www.stat.ucla.edu/~frederic/13/W23 .
HW2 is due Fri Feb10 at 2pm to statgrader@stat.ucla.edu or
statgrader2@stat.ucla.edu .

1. Observational Studies

- In observational studies, researchers *observe* and measure the explanatory variable but do not set its value for each subject.
- Examples:
  - A significantly higher proportion of individuals with lung cancer smoked compared to same-age individuals who don't have lung cancer.
  - College students who spend more time on Facebook tend to have lower GPAs.

Do these studies prove that smoking *causes* lung cancer or Facebook *causes* lower GPAs?

Nightlights and Nearsightedness

Example 4.1

Nightlights and nearsightedness

- Near-sightedness often develops in childhood
- Recent studies looked to see if there is an association between near-sightedness and night light use with infants
- Researchers interviewed parents of 479 children who were outpatients in a pediatric ophthalmology clinic
- Asked whether the child slept with the room light on, with a night light on, or in darkness before age 2
- Children were also separated into two groups: near-sighted or not near-sighted based on the child's recent eye examination

Night-lights and near-sightedness

	Darkness	Night Light	Room Light	Total
Near-sighted	18	78	41	137
Not near-sighted	154	154	34	342
Total	172	232	75	479

The largest group of near-sighted kids slept in rooms with night lights. It might be better to look at the data in terms of proportions.

Conditional proportions

$18/172 \approx 0.105$ $78/232 \approx 0.336$ $41/75 \approx 0.547$

Night lights and near-sightedness

	Darkness	Night Light	Room Light	Total
Near-sighted	**10.5%** 18/172	**33.6%** 78/232	**54.7%** 41/75	137
Not near-sighted	154	154	34	342
Total	172	232	75	479

- Notice that as the light level increases, the percentage of near-sighted children also increases.
- We say there is an **association** between near-sightedness and night lights.
- Two variables are **associated** if the values of one variable provide information (help you predict) the values of the other variable.

Night lights and near-sightedness

- While there is an association between the lighting condition and nearsightedness, can we claim that night lights and room lights *caused* the increase in near-sightedness?
- Might there be other reasons for this association?

Night lights and near-sightedness

- Could parents' eyesight be another explanation?
  - Maybe parents with poor eyesight tend to use more light to make it easier to navigate the room at night and parents with poor eyesight also tend to have children with poor eyesight.
  - Now we have a third variable of *parents' eyesight*
  - *Parents' eyesight* is considered a ***confounding variable.***
  - Other possible confounders? Wealth? Books? Computers?

Confounding Variables

- A **confounding variable** is associated with both the explanatory variable and the response variable.
- We say it is confounding because its effects on the response cannot be separated from those of the explanatory variable.
- Because of this, we can't draw cause and effect conclusions when confounding variables are present.

Confounding Variables

- Since confounding variables can be present in observational studies, we can't conclude causation from these kinds of studies.
- This doesn't mean the explanatory variable isn't influencing the response variable. **Association may not imply causation, but can be a pretty big hint.**

2. Observational studies versus Experiments

Section 4.2

Observational Studies vs. Experiments

- In an **observational study**, the researchers do not set the level of the explanatory variable for each subject. Typically each subject herself decides her level of the explanatory variable. Sometimes nature decides.
- For example, the researchers didn’t control which children slept with a night light on or not.
- Observational studies always have potential confounding variables present and these may prevent us from determining cause and effect.

Observational Studies vs. Experiments

- In an **experiment**, the researchers set the level of the explanatory variable for each subject.
- These levels may correspond to a treatment and control.
- Well designed experiments can control for confounding variables by making the treatment and control groups very similar except for what the experimenter manipulates.

3. Experiments and aspirin example.

Physicians' Health Study I (study aspirin's affect on reducing heart attacks.

- Started in 1982 with 22,071 male physicians.
- The physicians were **randomly assigned** into one of two groups.
    - Half took a 325mg aspirin every other day and half took a placebo.

Results

- Intended to go until 1995, the aspirin study was stopped in 1988 after finding significant results.
- 189 (1.7%) heart attacks occurred in the placebo group and 104 (0.9%) in the aspirin group. This is a 45% reduction in heart attacks for the aspirin group.
- What about confounding variables? Could the aspirin group be different than the placebo group in some other ways?
  - Did they have a better diet?
  - Did they exercise more?
  - Were they genetically less likely to have heart attacks?
  - Were they younger?

The Big Idea

- Confounding variables are often circumvented in experiments due to the **random assignment** of subjects to treatment groups.
- Randomly assigning people to groups tends to balance out all other variables between the groups.
- So confounding variables, including ones the researchers didn't anticipate, should be roughly equalized between the two groups and therefore should not be confounding.
- **Thus, cause and effect conclusions are sometimes possible in experiments through random assignment.** It must be a well run experiment though.

4. Random sampling and random assignment.

- With observational studies, **random sampling** is often done. This possibly allows us to make inferences from the sample to the population where the sample was drawn.
- With experiments, **random assignment** is done. This might allows us to conclude causation.

- The Physician's Health Study used random assignment. Did it also use random sampling?
- No, hardly any experiments use random sampling. Most get their subjects in other ways.
- The Physician's Health Study sent out invitation letters and questionnaires to all 261,248 male physicians between 40 and 84 years of age who lived in the United States.
- Of the 59,285 who were willing to participate in the trial, 26,062 were told they could not because of some medical condition or current medical treatment.

- So to what group can we generalize the results that taking aspirin can reduce heart attacks?
  - Just physicians in the study?
  - All male physicians between 40-84 years old?
  - All males physicians?
  - All males between 40-84 years olds?
  - All males?
  - Everyone between 40-84 years old?
  - Everyone?

Article Baseline Demographics After Random Assignment

Parameter	Placebo (n=129)	Uceris (n=128)
Mean age, years (range)	39.9 (12–68)	37.6 (13–66)
Men	77 (59.7)	70 (54.7)
Women	52 (40.3)	58 (45.3)
Mean disease duration (yrs)	6.3	5.5
Duration ≤1 year, n (%)	23 (17.8)	28 (21.9)
Duration >5 years, n (%)	51 (39.5)	44 (34.4)
Proctosigmoiditis	64 (49.6)	58 (45.3)
Left-sided colitis	44 (34.1)	37 (28.9)
Mean baseline UCDAI score	6.2	6.5
Mean baseline EI score	6.6	6.5
Prior mesalazine use	75 (58.1)	66 (51.6)
Prior sulfasalazine use	28 (21.7)	33 (25.8)

Sandborn WJ, Travis S, Moro L, Jones R, Gautille T, Bagin R, Huang M, Yeung P, Ballard ED 2nd Once-daily budesonide MMX® extended-release tablets induce remission in patients with mild to moderate ulcerative colitis: results from the CORE I study. *Gastroenterology* 2012 Nov;143(5):1218-26

Blocking and Random Assignment

- The goal in random assignment is to make the two groups as similar as possible in all ways other than the treatment.
- Sometime there are known confounders and you can block on (control for) these variables.
- For example, if our subjects consist of 60% females and 40% males, we can force each group to be 60% female and 40% male, using a matched pair design.
- Blocking makes sense when there are known confounders you want to control for. But randomly assigning subjects to groups makes them as similar as possible in terms of unknown confounders.