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

1. 1.96 vs. t multiplier for 95% Cis.
2. Causation, observational studies, confounding, and smoking and Facebook examples.
3. Nightlights and nearsightedness.
4. Practical and statistical significance.

Read chapter 4.

HW2 is due Wed, Feb12, 1159pm. 2.3.15, 3.3.18, and 4.1.23.

Midterm is Mon Feb24 in class.

The course website is http://www.stat.ucla.edu/~frederic/13/W25 .

1. For CIs, when to use 1.96 from the normal, & when to use a multiplier based on the t distribution.

iid = independent and identically distributed.

if the observations are iid. and n is large, then

P(μ is in the range $\bar{x}$ +/- 1.96 σ/√n) ~ 95%.

If the observations are iid and normal, and σ is known, then

P(μ is in the range $\bar{x}$ +/- 1.96 σ/√n) ~ 95%.

If the obs. are iid and normal and σ is unknown, then

P(μ is in the range $\bar{x}$ +/- t_{mult} s/√n) ~ 95%.

where t_{mult} is the multiplier from the t distribution.

This multiplier depends on n.

For quantitative symmetric data, book says n ≥ 20 is large.

For proportions, need ≥ 10 of each type, in your sample.

2. Causation, observational studies, and confounding. Smoking and facebook examples.

Chapter 4

- Previously research questions focused on **one** proportion
  - What proportion of the time did Marine choose the right bag?
- We will now start to focus on research questions comparing **two** groups.
  - Are smokers more likely than nonsmokers to have lung cancer?
  - Are children who used night lights as infants more likely to need glasses than those who didn't use night lights?

- Typically we observe two groups and we also have two variables (like smoking and lung cancer).
- So with these comparisons, we will:
  - determine when there is an association between our two variables.
  - discuss when we can conclude the outcome of one variable causes a change in the other.

Observational studies and confounding.

Types of Variables

- When two variables are involved in a study, they are often classified as explanatory and response
- **Explanatory variable** (Independent, Predictor)
  - The variable we think may be causing or explaining or used to predict a change in the response variable. (Often this is the variable the researchers are manipulating.)
- **Response variable** (Dependent)
  - The variable we think may be being impacted or changed by the explanatory variable.
  - The one we are interested in predicting.

Roles of Variables

- Choose the explanatory and response variable:
  - Smoking and lung cancer
  - Heart disease and diet
  - Hair color and eye color

- Sometimes there is a clear distinction between explanatory and response variables and sometimes there isn't.

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?

3. 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.**

4. Statistical and Practical significance.

- *Statistically significant* means that the results are unlikely to happen by chance alone.
- *Practically important* means that the difference is large enough to matter in the real world.

Cautions

- Practical importance is context dependent and somewhat subjective.
- Well designed studies try to equate statistical significance with practical importance, but not always.
- Look at the sample size.
  - If very large, expect significant results.
  - If very small, don't expect significant results. (A lot of missed opportunities---type II errors.)

Longevity example.

According to data from the WHO (2014) and World Cancer Report (2014), the average number of cigarettes smoked per adult per day in the U.S. is 2.967, and in Latvia it is 2.853.

The sample sizes are huge, so even this little difference is stat. sig. (In the U.S., the National Health Interview Survey has $n > 87000$).

If you do not like cigarette smoke around you, should you move to Latvia?

The difference is statistically significant, but not practically significant for most purposes.