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

1. 95% CIs, continued, factors affecting margin of error, and when to use which multiplier.
2. Bradley effect.
3. Statistical and practical significance, and longevity example.
4. Observational studies: association, confounding, and nightlights example.
5. Observational studies and experiments.
6. Experiments and aspirin example.
7. Random sampling and random assignment.
8. Blinding.
9. Portacaval shunt example.

Finish reading chapter 4.
http://www.stat.ucla.edu/~frederic/13/sum18 .
I changed the name of the "2SD method" in my notes from last lecture to the "1.96SE method".
HW2 is due Wed Aug22.

For possible tutoring in the Stats Club, see the UCLA Statistics Club facebook page.

Factors that Affect the Width of a Confidence Interval

Section 3.4

Factors Affecting Confidence Interval Widths

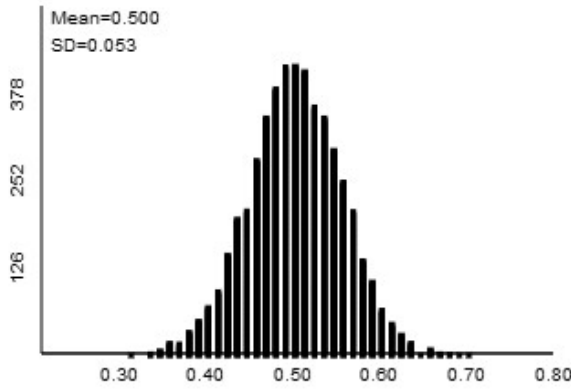
- **Level of confidence** (e.g., 90% vs. 95%)
  - As we increase the confidence level, we increase the width of the interval.
- **Sample size**
  - As sample size increases, variability decreases and hence the standard error will be smaller. This will result in a narrower interval.
- **Sample standard deviation**
  - A larger standard deviation, s, will yield a wider interval.
  - For sample proportions, wider intervals when $\hat{p}$ is closer to 0.5. s = √ [$\hat{p}$ (1-$\hat{p}$)].

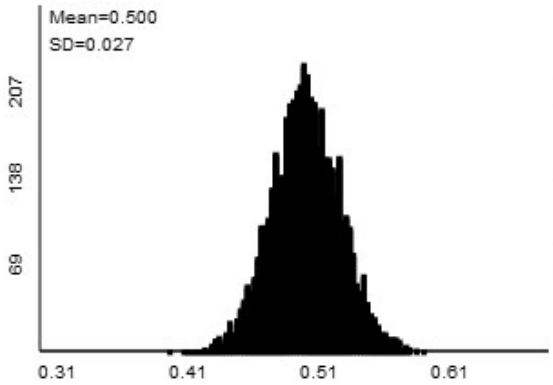
Level of Confidence

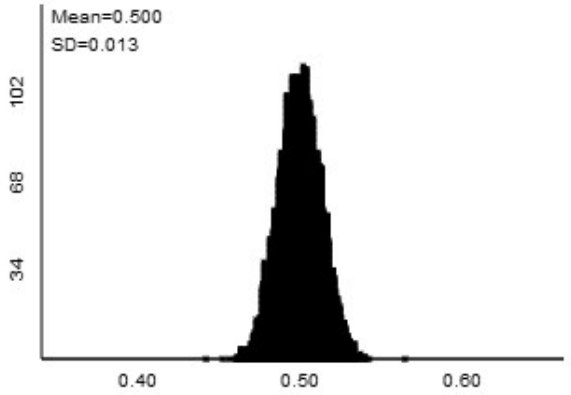
- If we have a wider interval, we should be more confident that we have captured the population proportion or population mean.
- We could see this with repeated tests of significance.
  - A higher confidence level corresponds to a lower significance level, and one must go farther to the left and farther to the right in our tables to get our confidence interval.

Sample Size

- We know as sample size increases, the variability (and thus standard deviation) in our null distribution decreases







n = 90 (SD = 0.054)

n = 361 (SD = 0.026)

n = 1444 (SD = 0.013)

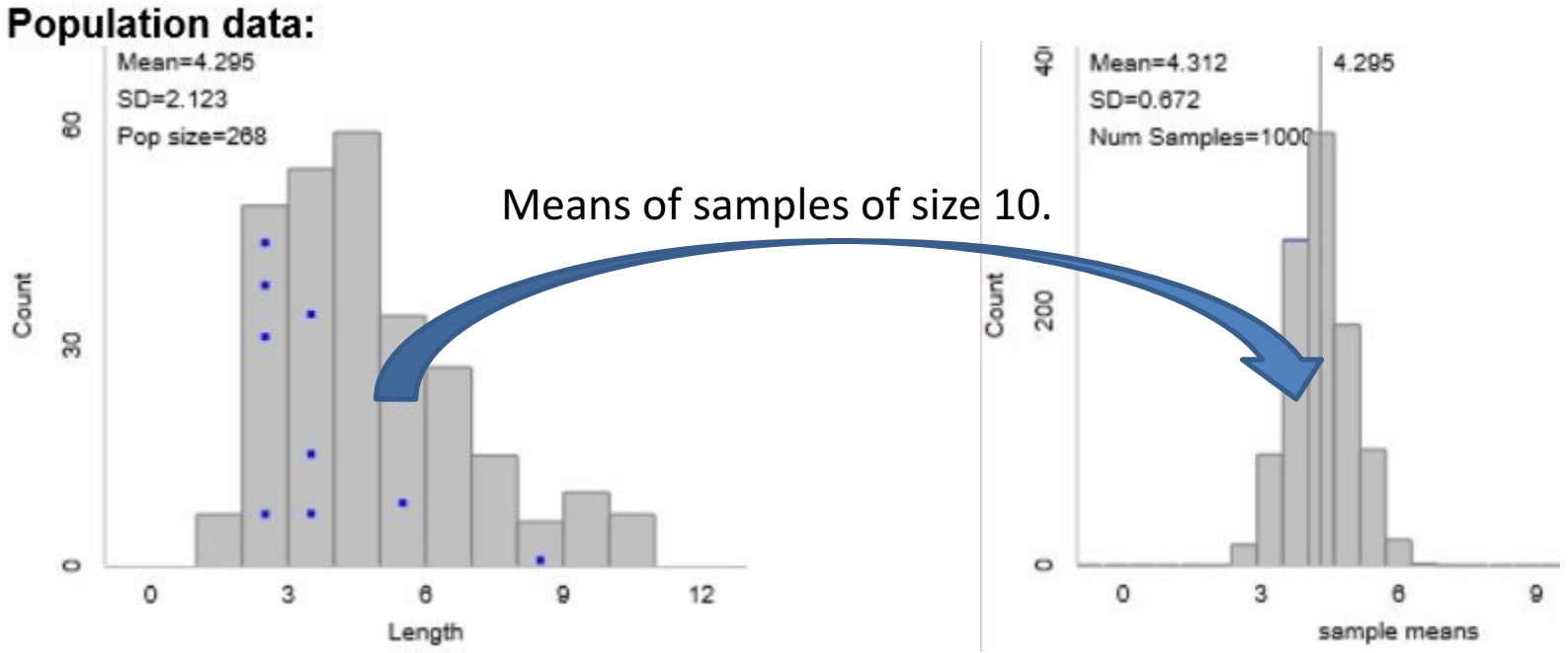
Sample size	90	361	1444
SD of null distr.	0.053	0.027	0.013
Margin of error	2 x SD = 0.106	2 × SD = 0.054	2 × SD = 0.026
Confidence interval	(0.091, 0.303)	(0.143, 0.251)	(0.171, 0.223)

Sample Size

- With everything else staying the same, increasing the sample size will make a confidence interval narrower.
- The observed sample proportion is the midpoint. That won't change as n changes.
- Margin of error is a multiple of the standard deviation so as the standard deviation decreases, so will the margin of error.

Standard Error

- Suppose we are taking repeated samples of a population.
- How do we estimate what the SE will be? $s/\sqrt{n}$.

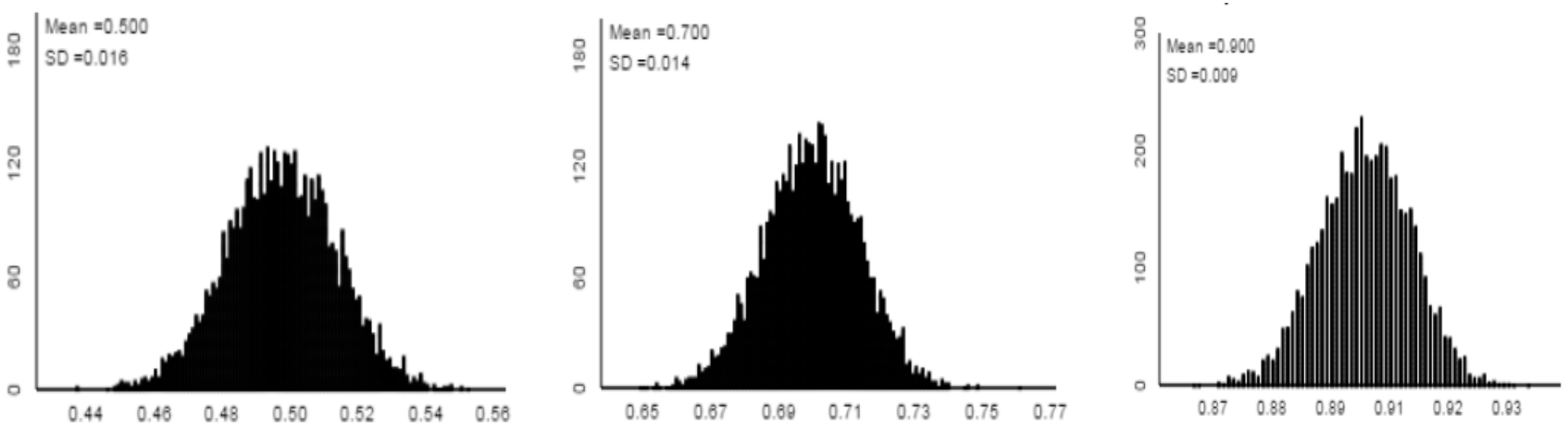


Standard Error

- The SE is approximated by $s/\sqrt{n}$.
- Remember that 1.96 s/√n is approximately the margin of error for a 95% confidence interval, so as the standard deviation of the sample data, s, increases so does the width of the confidence interval.
- Intuitively this should make sense, more variability in the data should be reflected by a wider confidence interval.

Value of $\hat{p}$
(or the value used for π under the null)

- As the value that is used under the null gets farther away from 0.5, the standard error decreases.



Formulas for Theory-Based Confidence Intervals

$$\hat{p} \pm multiplier \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \bar{x} \pm multiplier \times \frac{s}{\sqrt{n}}$$

- The width of the confidence interval increases as level of confidence increases (multiplier)
- The width of the confidence interval decreases as the sample size increases
- The value $\hat{p}$ also has a more subtle effect. The farther it is from 0.5 the smaller the width.
- The width of the confidence interval increases as the sample standard deviation s increases.

What does 95% confidence mean?

- If we repeatedly sampled from a population and constructed 95% confidence intervals, 95% of our intervals will contain the population parameter.
- Notice the interval is the random event here.

What does 95% confidence mean?

- Suppose a 95% confidence interval for a mean is 2.5 to 4.3. We would say we are 95% confident that the population mean is between 2.5 and 4.3.
  - Does that mean that 95% of the data fall between 2.5 and 4.3?
    - No
  - Does that mean that in repeated sampling, 95% of the sample means will fall between 2.5 and 4.3?
    - No
  - Does that mean that there is a 95% chance the population mean is between 2.5 and 4.3?
    - Not quite but close.

What does 95% confidence mean?

- What does it mean when we say we are 95% confident that the population mean is between 2.5 and 4.3?
  - It means that if we repeated this process (taking random samples of the same size from the same population and computing 95% confidence intervals for the population mean) repeatedly, 95% of the confidence intervals we find would contain the population mean.
  - P(confidence interval contains μ) = 95%.

What does 95% confidence mean?

- If we repeatedly sampled from a population and constructed 95% confidence intervals, 95% of our intervals will contain the population parameter.
- Notice the interval is the random event here.

For CIs, when to use 1.96, from the normal, and 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.

2. Cautions When Conducting Inference, and the controversial "Bradley Effect"

Example 3.5A

The “Bradley Effect”

- Tom Bradley, long-time mayor of Los Angeles, ran as the Democratic Party’s candidate for Governor of California in 1982.
  - Political polls of likely voters showed Bradley with a significant lead in the days before the election.
  - Exit polls favored Bradley significantly.
  - Many media outlets projected Bradley as the winner.
- Bradley narrowly lost the overall race.

The “Bradley Effect”

- After the election, research suggested a smaller percentage of white voters had voted for Bradley than polls predicted.
- A very large proportion of undecided voters voted for Deukmejian.

The “Bradley Effect”

- What are explanations for this discrepancy?
  - Likely voters answered the questions with a “social desirability bias”.
  - They answered polling questions the way they thought the interviewer wanted them to.
- Discrepancies in polling and elections has since been called the “Bradley effect”.
- It has been cited in numerous races and has included gender and other stances on political issues.

Clinton vs. Obama

- In the 2008 New Hampshire democratic primary
  - Obama received 36.45% of the primary votes.
  - Clinton received 39.09%.
- This result shocked many since Obama seemed to hold a lead over Clinton.
- USA Today/Gallup poll days before the primary, $n = 778$.
  - 41% of likely voters said they would vote for Obama.
  - 28% of likely voters said they would vote for Clinton.
- How unlikely are the Clinton and Obama poll numbers given that 39.09% and 36.45% of actual primary voters voted for Clinton and Obama?

Clinton vs. Obama

- We're assuming that the 778 people in the survey are a good representation of those who will vote.
  - The 778 people aren't a simple random sample.
- Pollsters used random digit dialing and asked if respondents planned to vote in the Democratic primary.
  - 9% (a total of 778) agreed to participate.
  - 319 said that they planned to vote for Obama and 218 for Clinton.

Clinton vs. Obama

Suppose we make the following assumptions:

1. Random digit dialing is a reasonable way to get a sample of likely voters.
2. The 9% who participated are like the 91% who didn't.
3. Voters who said they planned to vote actually voted in the primary.
4. Answers to who they say they will vote for match who they actually vote for.

Then we expect the sample proportion roughly to agree with the final vote proportion.

Clinton vs. Obama

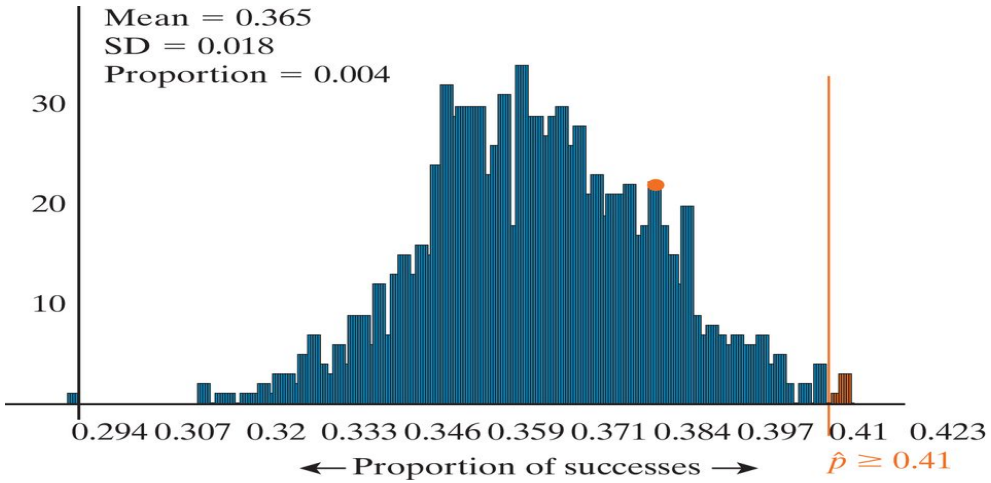
- One question is whether the proportion of likely voters who say they will vote for Obama is the same as the proportion of likely voters who actually vote for Obama (observed on primary day to be 0.3645).
- What would the Bradley Effect do in this case?
  - The proportion who say they will vote for Obama would be larger than 0.3645.

Clinton vs. Obama

- State the Null and Alternative hypotheses.
  - Null: The proportion of likely voters who would claim to vote for Obama is 0.3645.
  - Alternative: The proportion of likely voters who would claim to vote for Obama is higher than 0.3645.

Clinton vs. Obama

- Simulation of 778 individuals randomly chosen from a population where 36.45% vote for Obama
- The chance of getting a sample proportion of 0.41 successes or higher is very small. 0.004.



Clinton vs. Obama

- Convincing evidence that the discrepancy between what people said and how they voted is not explained by random chance alone.
- At least one of the 4 model assumptions is not true.

Clinton vs. Obama

1. **Random digit dialing is a reasonable way to get a sample of likely voters**
   - Roughly equivalent to a SRS of New Hampshire residents who have a landline or cell phone
   - Slight over-representation of people with more than one phone

Clinton vs. Obama

2. **The 9% of individuals reached by phone who agree to participate are like the 91% who didn't**
   - 91% includes people who didn't answer their phone and who didn't participate
   - Assumes that respondents are like non-respondents.
   - The ***response rate*** was very low, but typical for phone polls
   - No guarantee that the 9% are representative.

Clinton vs. Obama

3. **Voters who said they plan to vote in the Democratic primary will vote in the primary.**
   – There is no guarantee.
4. **Respondent answers match who they actually vote for.**
   There is no guarantee.

Clinton vs. Obama

Because of the wide disparity between polls and the primary, an independent investigation was done with the following conclusions:

1. People changed their opinion at the last minute
2. People in favor of Clinton were more likely not to respond
3. The Bradley Effect
4. Clinton was listed before Obama on every ballot

These are examples of **nonrandom errors.**

3. 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.

Causation.

Chapter 4

Big Idea of 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?

Big Idea of Chapter 4

- 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 an outcome of the other.

4. 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.

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?

Night Lights and Nearsightedness

Example 4.1

Nightlights and Near-Sightedness

- 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 ≈ 0.105 78/232 ≈ 0.336 41/75 ≈ 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.**

5. 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.

6. 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.

7. 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, but 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.

8. Blinding.

Even in experiments, the treatment and control groups can be different in ways other than the explanatory variable. This is especially true when the response variable is somewhat subjective.

Pain is an example. One study found that 1/4 of patients suffering from post-operative pain, when given a placebo (just a pill of sugar and water) claimed they experienced "significant prompt pain relief".

8. Blinding.

People might not be able to judge their own levels of pain very well, and may be influenced by the belief that they have taken an effective treatment.

Thus in an experiment with such a response variable, researchers should ensure the subject does not know whether he or she received the treatment or the control. This is called blinding.

In a *double-blind* experiment, neither the subject nor the researcher recording the response variable knows the level of the explanatory variable for each subject, i.e. treatment or control.

9. Portacaval shunt example.

The following example shows the importance of doing a randomized controlled experiment.

The portacaval shunt is a medical procedure aimed at curbing bleeding to death in patients with cirrhosis of the liver.

The following table summarizes 51 studies on the portacaval shunt. The poorly designed studies were very enthusiastic about the surgery, while the carefully designed studies prove that the surgery is largely ineffective.

	Degree of enthusiasm		
Design	High	Moderate	None
No controls	24	7	1
Controls, but not randomized	10	3	2
Randomized controlled	0	1	3

9. Portacaval shunt example.

Why did the poorly designed studies come to the wrong conclusion?

A likely explanation is that in the studies where patients were not randomly assigned to the treatment or control group, by and large the healthier patients were given the surgery.

This alone could explain why the treatment group outlived the control group in these studies.

	Degree of enthusiasm		
Design	High	Moderate	None
No controls	24	7	1
Controls, but not randomized	10	3	2
Randomized controlled	0	1	3