MATH 1051H (S61): Lecture #02

A randomized controlled trial.

‘Perhaps H.G. Wells was right when he said “Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.’”

- Samuel S. Wilks, President of the American Statistical Association, December 28, 1950

“I keep saying the sexy job in the next ten years will be statisticians. People think I’m joking, but who would’ve guessed that computer engineers would’ve been the sexy job of the 1990s?”

- Hal Varian, The McKinsey Quarterly, January 2009

Case Study: Treating Chronic Fatigue Syndrome

Treating Chronic Fatigue Syndrome

Objective. Evaluate the effectiveness of cognitive-behavior therapy for chronic fatigue syndrome.

Participant pool. 142 patients who were recruited from referrals by primary care physicians and consultants to a hospital clinic specializing in chronic fatigue syndrome.

Actual participants. Only 60 of the 142 referred patients entered the study. Some were excluded because they didn't meet the diagnostic criteria, some had other health issues, and some refused to be a part of the study.

Study Design

Patients randomly assigned to treatment and control groups, 30 patients in each group:

Treatment: Cognitive behavior therapy — collaborative, educative, and with a behavioral emphasis. Patients were shown on how activity could be increased steadily and safely without exacerbating symptoms.

Control: Relaxation — No advice was given about how activity could be increased. Instead progressive muscle relaxation, visualization, and rapid relaxation skills were taught.

Results

The table below shows the distribution of patients with good outcomes at 6-month follow-up. Note that 7 patients dropped out of the study: 3 from the treatment and 4 from the control group.

Proportion with good outcomes

	Good outcome		
	Yes	No	Total
Groups			
Treatment	19	8	27
Control	5	21	26
Total			
	24	29	53

Treatment Group: 19/27 = 0.70 = 70%

Control Group: 5/26 = 0.19 = 19%

Understanding the results

Do the data show a "real" difference between the groups?

Suppose you flip a coin 100 times. While the chance a coin lands heads in any given coin flip is 50%, we probably won't observe exactly 50 heads. This type of fluctuation is part of almost any type of data generating process.

The observed difference between the two groups (70 - 19 = 51%) may be real, or may be due to natural variation.

Since the difference is quite large, it is more believable that the difference is real.

We use statistical tools to determine if the difference is so large that we should reject the notion that it was due to chance.

Generalizing the results

Are the results of this study generalizable to all patients with chronic fatigue syndrome?

No. These patients had specific characteristics and volunteered to be a part of this study, therefore they may not be representative of all patients with chronic fatigue syndrome.

While we cannot immediately generalize the results to all patients, this first study is encouraging. The method at least works for patients with some narrow set of characteristics, and that gives hope that it will work, at least to some degree, with other patients.

Data Basics

Data matrix

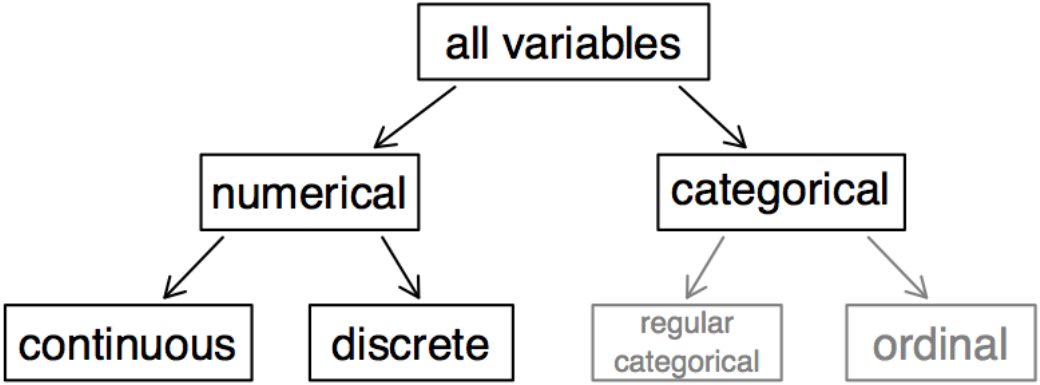
Data collected on students in a statistics class on a variety of variables

variable ↓

Stu.	gender	intro_extra	···	dread
1	male	extravert	···	3
2	female	extravert	···	2
3	female	introvert	···	4
4	female	extravert	···	2
⋮	⋮	⋮	⋮	⋮
86	male	extravert	···	3

← *observation*

Types of variables



Types of variables (cont.)

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- **gender** — categorical
- **sleep** — numerical, continuous
- **bedtime** — categorical, ordinal
- **countries** — numerical, discrete
- **dread** — categorical, ordinal (could also be used as numerical)

Why Do We Need to Care?

The type of data determines the type of analysis that you can perform and the statistics that make sense. Example: Top hockey players in the NHL according to goals (regular season) in 2017-2018.

- The mean number of goals is a value that makes sense. The variable *Number of Goals* is numerical.
- The mean jersey number is a value that does not make sense. The variable *Jersey Number* is not numerical.
- If considering jersey numbers, the proportion of players with even jersey numbers is a value that makes sense. The variable *Odd or Even Jersey Number* is categorical.

Player	Goals	Number
Alex Ovechkin	49	8
Patrick Laine	44	29
William Karlsson	43	71
Evgeni Malkin	42	71
Eric Staal	42	12
Connor McDavid	41	41
Tyler Seguin	40	91
Anders Lee	40	27
Nikita Kucherov	39	86
Nathan MacKinnon	39	29
Mean (average):	41.9	46.5

Practice

What type of variable is a telephone area code?

1. numerical, continuous
2. numerical, discrete
3. categorical
4. categorical, ordinal

Practice

What type of variable is a telephone area code?

1. numerical, continuous
2. numerical, discrete
3. *categorical*
4. categorical, ordinal

Numerical Data

Numerical (or quantitative) data are *numbers* representing counts or measurements.

- Weights of athletes
- Number of siblings
- GPA

Working with Numerical Data

We can further distinguish between numerical data by breaking them into two types:

- **discrete** numerical data (integers)
- **continuous** numerical data (real numbers)

Categorical Data

Categorical (or qualitative) data are *names or labels* (categories!)

- country codes for telephones (e.g., USA and Canada are 1)
- Social Insurance Numbers (e.g., 516 248 917 - Canadian system)
- car colours (red, blue, green, … fuscia?)

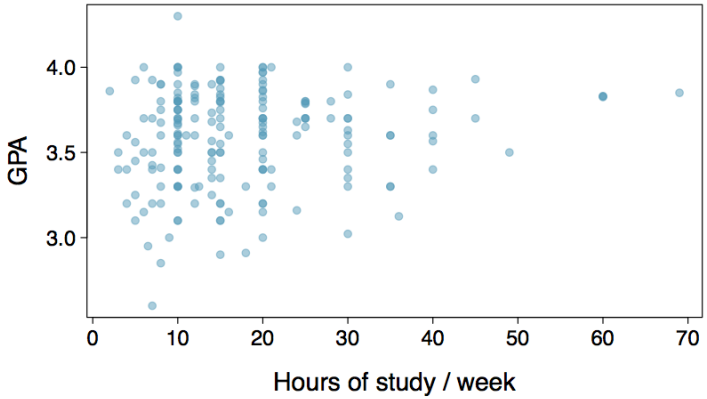
Ordinal Data

Ordinal Data are categorical data which have a natural order structure

- days of the week (Sunday, Monday, …)
- months of the year (January, February, …)
- letter grades (A, B, C, D, F)
- ranking scheme (Excellent, Good, Fair, Poor)

Relationships between variables

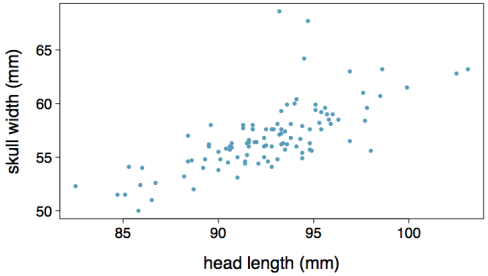
Does there appear to be a relationship between the hours of study per week and the GPA of a student?



Can you spot anything unusual about any of the data points?

Practice

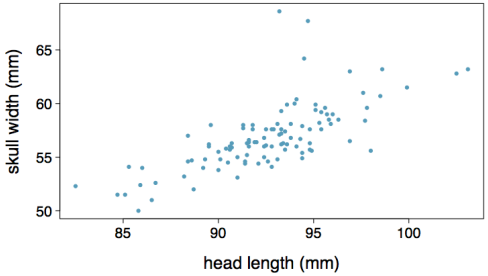
Based on the scatterplot, which of the following statements is correct about the head and skull length of possums?



1. There is no relationship between head length and skull width, i.e. the variables are independent.
2. Head length and skull width are positively associated.
3. Skull width and head length are negatively associated.
4. A longer head causes the skull to be wider.
5. A wider skull causes the head to be longer.

Practice

Based on the scatterplot, which of the following statements is correct about the head and skull length of possums?



1. There is no relationship between head length and skull width, i.e. the variables are independent.
2. *Head length and skull width are positively associated.*
3. Skull width and head length are negatively associated.
4. A longer head causes the skull to be wider.
5. A wider skull causes the head to be longer.

Associated versus Independent

- When two variables show some connection with one another, they are called **associated variables**.
  - Associated variables can also be called **dependent variables**, and vice-versa
- If two variables are not associated, i.e., there is no evident connection between the two, then they are said to be **independent**.

Overview of Data Collection Principles

Populations and Samples

Research Question Can people become better, more efficient runners on their own, merely by running?

Population of Interest All people

Sample Group of adult women who recently joined a running group

Population to which results can be generalized Adult women, if the data are randomly sampled

Anecdotal evidence and early smoking research

- Anti-smoking research started in the 1930s and 1940s when cigarette smoking became increasingly popular. While some smokers seemed to be sensitive to cigarette smoke, others were completely unaffected.
- Anti-smoking research was faced with resistance based on **anecdotal evidence** such as “My uncle smokes three packs a day and he’s in perfectly good health”: a limited sample size that might not be representative
- It was concluded that “smoking is a complex human behavior, by its nature difficult to study, confounded by human variability.”
- In time researchers were able to examine larger samples of cases (smokers), and trends showing that smoking has negative health impacts became much clearer.

Census

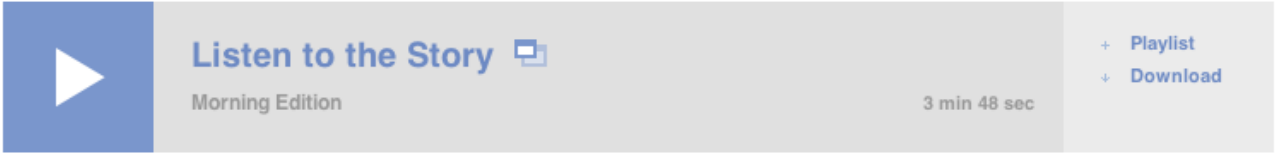
- Wouldn't it be better to just include everyone and "sample" the entire population?
  - This is called a **census**
- There are problems with taking a census:
  - It can be difficult to complete a census: there always seem to be some individuals who are hard to locate or hard to measure. And these difficult-to-find people may have certain characteristics that distinguish them from the rest of the population.
  - Populations rarely stand still. Even if you could take a census, the population changes constantly, so it's never possible to get a perfect measure.
  - Taking a census may be more complex than sampling.

Illegal Immigrants Reluctant To Fill Out Census Form

by PETER O'DOWD

March 31, 2010 4:00 AM

from KJZZ



There is an effort underway to make sure Hispanics are accurately counted in the 2010 Census. Phoenix has some of the country's "hardest-to-count" districts. Some Latinos, especially illegal residents, fear that participating in the count will expose them to immigration raids or government harassment.

http://www.npr.org/templates/story/story.php?storyId=125380052

Exploratory analysis to inference

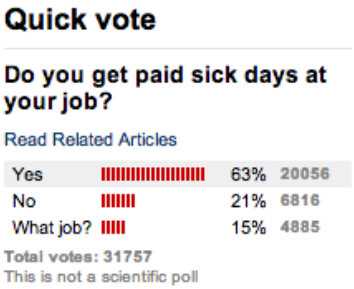
- Sampling is natural.
- Think about sampling something you are cooking - you taste (examine) a small part of what you're cooking to get an idea about the dish as a whole.
- When you taste a spoonful of soup and decide the spoonful you tasted isn't salty enough, that's **exploratory analysis**.
- If you generalize and conclude that your entire soup needs salt, that's an inference.
- For your inference to be valid, the spoonful you tasted (the sample) needs to be representative of the entire pot (the population).
- If your spoonful comes only from the surface and the salt is collected at the bottom of the pot, what you tasted is probably not representative of the whole pot.
- If you first stir the soup thoroughly before you taste, your spoonful will more likely be representative of the whole pot.

Sampling bias

Non-response: If only a small fraction of the randomly sampled people choose to respond to a survey, the sample may no longer be representative of the population.

Voluntary response: Occurs when the sample consists of people who volunteer to respond because they have strong opinions on the issue. Such a sample will also not be representative of the population.

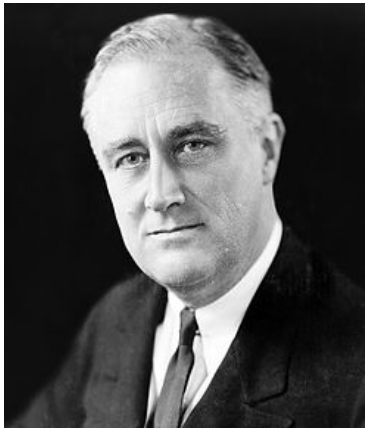
Convenience sample: Individuals who are easily accessible are more likely to be included in the sample.



Sampling bias example: Landon versus FDR

A historical example of a biased sample yielding misleading results. In 1936, Alf Landon became the Republican presidential nominee, opposing the re-election of Franklin Delano Roosevelt, a Democrat.

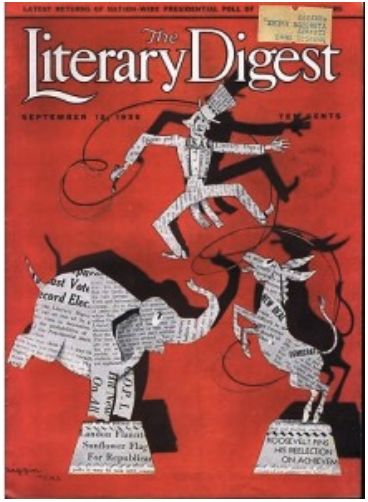
Landon (GOP)



FDR (DEM)

The Literary Digest Poll

- The Literary Digest polled about 10 million Americans, and got responses from about 2.4 million.
- The poll showed that Landon would likely be the overwhelming winner and FDR would get only 43% of the votes.
- Election result: FDR won, with 62% of the votes (and 98.5% of the electoral votes — the most lop-sided electoral vote victory in US history).
- The magazine was completely discredited because of the poll, and was soon discontinued.



The Literary Digest Poll — what went wrong?

- The magazine had surveyed
  - its own readers,
  - registered automobile owners, and
  - registered telephone users.

These groups had incomes well above the national average of the day (remember, this is Great Depression era) which resulted in lists of voters far more likely to support Republicans than a truly typical voter of the time, i.e., the sample was not representative of the American population at the time.

Large samples are preferable, but …

- The Literary Digest election poll was based on a sample size of 2.4 million, which is huge, but since the sample was biased, the sample did not yield an accurate prediction.
- Back to the soup analogy: If the soup is not well stirred, it doesn't matter how large a spoon you have, it will still not taste right. If the soup is well stirred, a small spoon will suffice to test the soup.

Practice

A school district is considering whether it will no longer allow high school students to park at school after two recent accidents where students were severely injured. As a first step, they survey parents by mail, asking them whether or not the parents would object to this policy change. Of 6,000 surveys that go out, 1,200 are returned. Of these 1,200 surveys that were completed, 960 agreed with the policy change and 240 disagreed. Which of the following statements are true?

1. Some of the mailings may have never reached the parents.
2. The school district has strong support from parents to move forward with the policy approval.
3. It is possible that majority of the parents of high school students disagree with the policy change.
4. The survey results are unlikely to be biased because all parents were mailed a survey.

(a) Only 1 (b) 1 and 2 (c) 1 and 3 (d) 3 and 4 (e) Only 4

Practice

A school district is considering whether it will no longer allow high school students to park at school after two recent accidents where students were severely injured. As a first step, they survey parents by mail, asking them whether or not the parents would object to this policy change. Of 6,000 surveys that go out, 1,200 are returned. Of these 1,200 surveys that were completed, 960 agreed with the policy change and 240 disagreed. Which of the following statements are true?

1. Some of the mailings may have never reached the parents.
2. The school district has strong support from parents to move forward with the policy approval.
3. It is possible that majority of the parents of high school students disagree with the policy change.
4. The survey results are unlikely to be biased because all parents were mailed a survey.

(a) Only 1 (b) 1 and 2 (c) *1 and 3* (d) 3 and 4 (e) Only 4

Sampling Errors

There are a number of ways things can go wrong. Some examples:

- non-response
- self-selection
- framing bias
- sensitive topics
- interviewer bias
- timing

More details in a **handout** that will be posted on Blackboard.

Explanatory and Response Variables

- To identify the explanatory variable in a pair of variables, identify which of the two is suspected of affecting the other:

$$\text{explanatory variable} \overset{\text{might affect}}{\longrightarrow} \text{response variable}$$

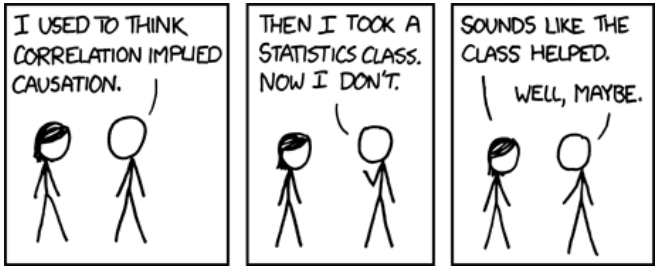
- Labeling variables as explanatory and response does not guarantee the relationship between the two is actually causal, even if there is an association identified between the two variables. We use these labels only to keep track of which variable we suspect affects the other.

Methods of Collecting Data

Observational study: Researchers collect data in a way that does not directly interfere with how the data arise, i.e. they merely "observe", and can only establish an association between the explanatory and response variables.

Experiment: Researchers randomly assign subjects to various treatments in order to establish causal connections between the explanatory and response variables.

If you're going to walk away with one thing from this class, let it be "correlation does not imply causation".



Observational studies and sampling strategies

New study sponsored by General Mills says that eating breakfast makes girls thinner

By ALEX DOMINGUEZ, Associated Press

Girls who regularly ate breakfast, particularly one that includes cereal, were slimmer than those who skipped the morning meal, according to a study that tracked nearly 2,400 girls for 10 years.

Girls who ate breakfast of any type had a lower average body mass index, a common obesity gauge, than those who said they didn't. The index was even lower for girls who said they ate cereal for breakfast, according to findings of the study conducted by the Maryland Medical Research Institute. The study received funding from the National Institutes of Health and cereal-maker General Mills.

"Not eating breakfast is the worst thing you can do, that's really the take-home message for teenage girls," said study author Bruce Barton, the Maryland institute's president and CEO.

The fiber in cereal and healthier foods that normally accompany cereal, such as milk and orange juice, may account for the lower body mass index among cereal eaters, Barton said.

The results were gleaned from a larger NIH survey of 2,379 girls in California, Ohio and Maryland who were tracked between ages 9 and 19. Results of the study appear in the September issue of the Journal of the American Dietetic Association.

Nearly one in three adolescent girls in the United States is overweight, according to the association. The problem is particularly troubling because research shows becoming overweight as a child can lead to a lifetime struggle with obesity.

As part of the survey, the girls were asked once a year what they had eaten during the previous three days. The data were adjusted to compensate for factors such as differences in physical activity among the girls and normal increases in body fat during adolescence.

What type of study is this: observational or experiment?

"Girls who regularly ate breakfast, particularly one that includes cereal, were slimmer than those who skipped the morning meal, according to a study that tracked nearly 2,400 girls for 10 years. […] As part of the survey, the girls were asked once a year what they had eaten during the previous three days."

This is an **observational study** since the researchers merely observed the behavior of the girls (subjects) as opposed to imposing treatments on them.

What is the conclusion of the study?

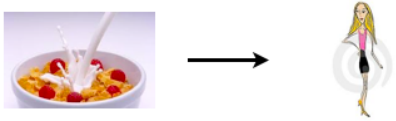
There is an association between girls eating breakfast and being slimmer.

Who sponsored the study?

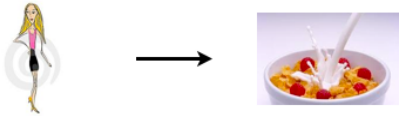
General Mills (cereal manufacturer)

3 Possible Explanations

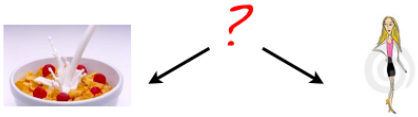
1. Eating breakfast causes girls to be thinner.



1. Being thin causes girls to eat breakfast.



2. A third variable is responsible for both. What could it be? An extraneous variable that affects both the explanatory and the response variable and that make it seem like there is a relationship between the two is called a **confounding variable**.



Prospective versus Retrospective Studies

A **prospective study** identifies individuals and collects information as events unfold.

- **Example**: The Nurses Health Study has been recruiting registered nurses and then collecting data from them using questionnaires since 1976.

Retrospective studies collect data after events have taken place.

- **Example**: Researchers reviewing past events in medical records.

Obtaining Good Samples

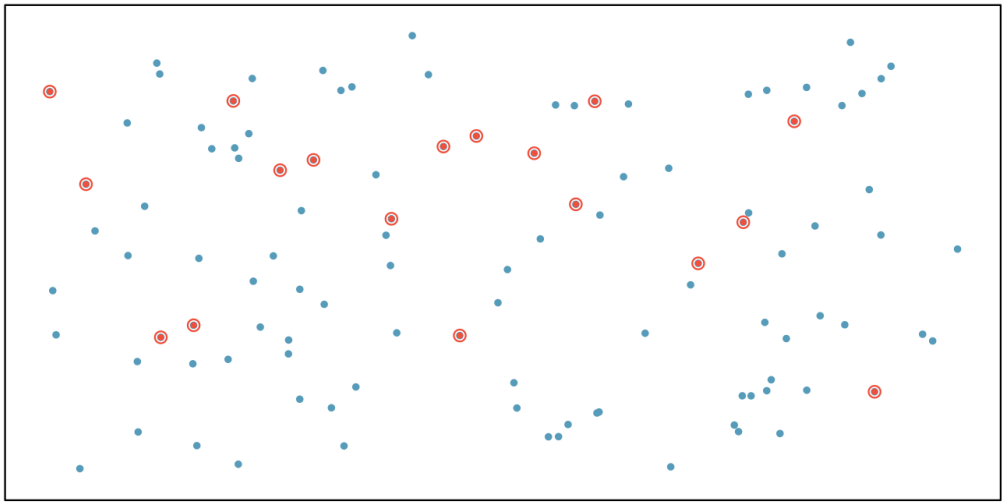
Almost all statistical methods are based on the notion of implied randomness.

If observational data are not collected in a random framework from a population, these statistical methods – the estimates and errors associated with the estimates – are not reliable.

Most commonly used random sampling techniques are **simple**, **stratified**, and **cluster** sampling.

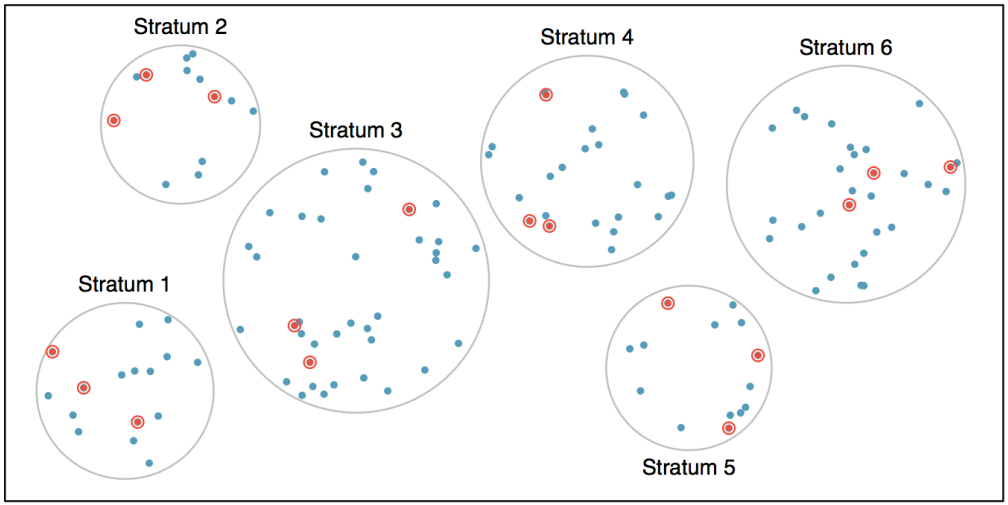
Simple Random Sample

Randomly select cases from the population, where there is no implied connection between the points that are selected.



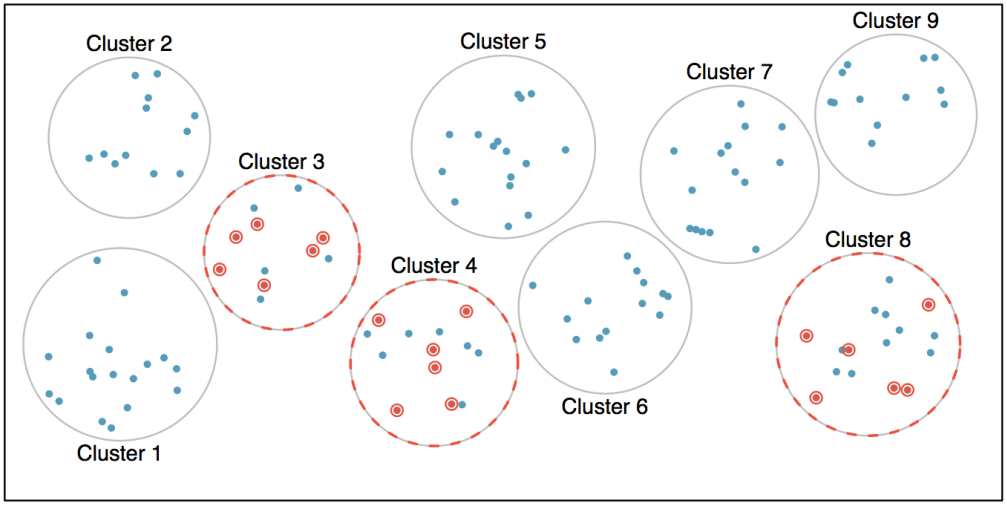
Stratified Sample

Strata are made up of similar observations. We take a simple random sample from each stratum.



Cluster Sample

Clusters are usually not made up of homogeneous observations, and we take a simple random sample from a random sample of clusters. Usually preferred for economical reasons.



Practice

A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments. Which approach would likely be the *least* effective?

1. Simple random sampling
2. Cluster sampling
3. Stratified sampling
4. Blocked sampling

Practice

A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments. Which approach would likely be the *least* effective?

1. Simple random sampling
2. Cluster sampling
3. *Stratified sampling*
4. Blocked sampling

Conclusion and Note

The only way to avoid the potential bias of samples is to select the sample **randomly**. Aside from Simple Random Sampling, the other techniques mentioned can be used to assist in this.

Handout posted to Blackboard reviewing the topic, and you can read Chapter 1.4.2 in our text.

Experiments

Principles of Experimental Design

Studies where researchers assign treatments to cases are called **experiments**. If the assignment of treatments to cases (e.g., using a coin flip to determine which treatment a patient receives), the study is called a **randomized experiment**.

Randomized experiments have a series of four principles.

Principle 1: Controlling

Researchers assign treatments to cases, and do their best to **control** for other differences between groups.

Example: in a drug trial, patients may be asked to take a pill daily. Some may take the pill dry (ick!), some with just a sip of water, some with coffee, and others with juice. To **control** for the effect of accompanying liquid, a doctor may ask all patients to drink a 12 oz glass of water with the pill.

Principle 2: Randomization

Researchers **randomize patients** into treatment groups to account for variables that cannot be controlled.

Example: some patients are more susceptible to disease than others due to dietary habits. **Randomizing** patients into treatment/control groups helps even out these differences, possibly preventing accidental bias.

Principle 3: Replication

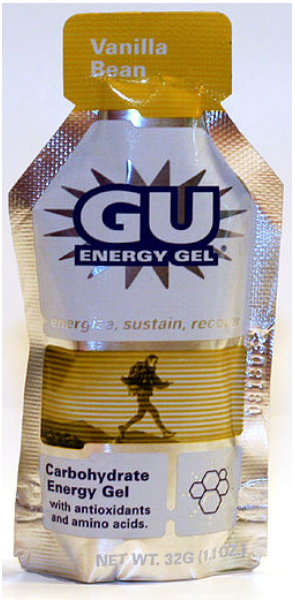
The more cases researchers observe, the more accurately they can estimate the effects of explanatory variables on response variables. In a single study, we **replicate** by collecting a sufficiently large sample. Additionally, scientists often replicate an entire study over again to verify earlier findings.

Principle 4: Blocking

Researchers sometimes know (or suspect) that variables other than the treatment influence the response. Under this situation, they may first group individuals by this variable, and then randomize cases within each block. This is known as **blocking**.

Example: If we were researching the effect of a drug on heart attacks, we might first split patients into high-risk and low-risk **blocks** (based on diet, physique, genetic screening, or some other approach), and *then* randomly assign half of each block to the control group, and the other half to the drug (treatment) group.

More on Blocking



We would like to design an experiment to investigate if energy gels makes you run faster:

- **Treatment**: energy gel
- **Control**: no energy gel

It is suspected that energy gels might affect pro and amateur athletes differently, therefore we block for pro status:

- Divide the sample to pro and amateur
- Randomly assign pro athletes to treatment and control groups
- Randomly assign amateur athletes to treatment and control groups
- Pro/amateur status is equally represented in the resulting treatment and control groups

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

Practice

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 equally represented in each group. Which of the below is correct?

1. There are 3 explanatory variables (light, noise, gender) and 1 response variable (exam performance)
2. There are 2 explanatory variables (light and noise), 1 blocking variable (gender), and 1 response variable (exam performance)
3. There is 1 explanatory variable (gender) and 3 response variables (light, noise, exam performance)
4. There are 2 blocking variables (light and noise), 1 explanatory variable (gender), and 1 response variable (exam performance)

Practice

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 equally represented in each group. Which of the below is correct?}

1. There are 3 explanatory variables (light, noise, gender) and 1 response variable (exam performance)
2. *There are 2 explanatory variables (light and noise), 1 blocking variable (gender), and 1 response variable (exam performance)*
3. There is 1 explanatory variable (gender) and 3 response variables (light, noise, exam performance)
4. There are 2 blocking variables (light and noise), 1 explanatory variable (gender), and 1 response variable (exam performance)

Difference between Blocking and 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.

More Experimental Design Terminology

Placebo: fake treatment, often used as the control group for medical studies

Placebo effect: experimental units showing improvement simply because they believe they are receiving a special treatment

Blinding: when experimental units do not know whether they are in the control or treatment group

Double-blind: when both the experimental units and the researchers who interact with the patients do not know who is in the control and who is in the treatment group

Practice

What is the main difference between observational studies and experiments?

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

Practice

What is the main difference between observational studies and experiments?

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

Random Assignment versus Random Sampling