

## Big Picture

---

Collect Data (experiment or sample survey)

↓  
Exploratory Data Analysis

↓  
Formal Statistical Inference

- 
- We need to collect trustworthy data and be able to judge the quality of data produced by others in order to be able to do formal statistical inference (generalize results in samples).
  - We do not want to base conclusions or inferences on **anecdotal evidence**.
  - Sources of available data:
    - *Statistical Abstract of the United States*
    - Government databases
    - Internet

## Observation vs. Experiment

---

- An **observational study** observes individuals and measures variables of interest but does not attempt to influence the individuals' responses.
- An **experiment** imposes some treatment on individuals in order to observe their responses.

## Samples

---

- A **population** is the collection of individuals (do not have to be people) we are interested in studying.
- A **sample** is a subset of the population.
- If we can measure every individual in our population we have **census**.

---

**Example:** Some people believe that exercise raises the body's metabolic rate for as long as 12 to 24 hours enabling us to continue to burn off fat after our workout has ended. In a study of this effect, subjects were asked to walk briskly on a treadmill for several hours. Their metabolic rates were measured before and 12 hours after the exercise.

**Question:** Was this study an experiment? Why or why not?

**Question:** What are the explanatory and response variables?

## Design of Experiments

---

### Terminology:

- The individuals on which the experiment is done are called the **experimental units**. When the units are human beings, they are called **subjects**.
- A specific experimental condition applied to the units is called a **treatment**.
- The explanatory variables that you are think may explain the effect of the treatments are called **factors**. The specific values of each of the factors are called **levels**.
- An **experimental design** describes how the treatments are assigned to the experimental units.

**Example:** What are the effects of repeated exposure to an advertising message? The answer depends both on the length of the ad and on how often it is repeated. An experiment is conducted with undergrad students of OSU to investigate this question. All subjects viewed a 60-minute episode of a television show that included ads for a new ice cream. Some subjects saw a 30-second ad; others, a 90-second version. The same ad was repeated 1, 3, or 5 times during the program. After viewing, all subjects answered questions about their recall of the ad and their intention to try the ice cream.

- What are the experimental units?
  - the subjects are OSU undergrads
- What is the response variable?
  - recall of the ad and intention to try the ice cream
- What are the factors?
  - length of the ad and number of repetitions
- What are the levels?
  - for length the levels are 30 and 90 seconds
  - for repetitions the levels are 1, 3, and 5 times

What are the treatments? Make a diagram to layout the treatments.

		Factor B - Repetitions		
		1 time	3 times	5 times
Factor A Length	30 seconds	1	2	3
	90 seconds	4	5	6

**Example:** Suppose we want to answer the following question:  
*Does aspirin help prevent heart attacks?*  
Let's design a good experiment that can answer this question.

**Explanatory variable:** aspirin  
**Response variable:** number of heart attacks

In our study, over 20,000 doctors volunteered to be subjects.  
**Q:** Should all the subjects take an aspirin a day?

No, we want to determine if people who take aspirin have fewer heart attacks than people who do not take aspirin. We need to observe subjects not taking aspirin.

→ Experiments should compare treatments rather than attempt to assess a single treatment in isolation, i.e. experiments should be **comparative**.

**Definition:** Subjects who are not taking the aspirin are in the **control group**.

- Question:** To assign subjects to groups, should we
- a. let the subjects decide themselves which group to be in
  - b. try to balance the groups using our judgment
  - c. let a chance process (such as flipping a coin) decide

---

Say we let them decide themselves. If the aspirin group has fewer heart attacks, do we have evidence that the aspirin was responsible for it?

No, maybe those who chose to take aspirin were in poorer health to begin with, or all of the people who chose to take it are more concerned with their health and also take vitamins.

The effect of aspirin would be "mixed up" with the effect of these other variables. This is an example of **confounding**. We won't know whether changes in the explanatory variable or changes in lurking variables *caused* changes in the response variable.

---

Suppose we get an "expert" to divide the subjects using his judgment and he happens to be interested in the study. Consciously, or consciously, our expert may "stack" the groups in a way that favors a certain result and his *bias* could introduce lurking variables. For example, he may assign all of the overweight men to the no aspirin group.

**Definition:** The design of a study is **biased** if it systematically favors certain outcomes.

---

→ So we should use randomization. It will eliminate bias due to self-selection or the judgment of the experimenter.

Benefits of randomization:

- keep division free of bias
- scatter potential lurking variables

For example, we won't get all the old people in one group and all the young people in the other group. The two groups are similar (except for chance variation).

---

Now we know we want to do a **randomized comparative experiment** (an experiment that uses both comparison and randomization).

**Question:** Once we randomly assign subjects to groups, should we

- tell them which group they are in
- not tell them which group they are in

---

If Mr. Tension knows that he is not getting aspirin, he might think that he is going to have a heart attack and consequently be worried, raise his blood pressure, and increase his risk of a heart attack.

On the other hand, many patients respond favorably to any treatment, even a dummy treatment such as a sugar pill (called a **placebo**), presumably because of trust in the doctor and expectations of a cure. This is called the **placebo effect**.

The effects of the aspirin may get **confounded** with the placebo effect if subjects know which group they have been assigned to.

Subjects are said to be **blind** if they do not know which treatment they are receiving.

---

**Question:** If we give aspirin to the treatment group and a sugar pill to the control group, which group(s) experiences the placebo effect?

- The control group
- The treatment group
- Both groups

---

Evaluators will need to diagnose possible heart attacks in subjects and report data to the researchers.

**Question:** To aid these evaluators, should we let them know whether the patient was getting treatment?  
No, the evaluators may be biased.

For example, if a doctor gets money from the company manufacturing aspirin, she may actually be more likely to diagnose a subject in the control group with "minor chest pains" as having a heart attack than a subject in the treatment group.

Statistics 528 - Lecture 11  
Prof. Kate Calder 19

---

So we have a **double blind experiment** – neither the subjects nor the personnel evaluating them know which treatment any subject received.

A double-blind study eliminates the conscious and unconscious bias that subjects and evaluators may have concerning the treatments being studied.

Statistics 528 - Lecture 11  
Prof. Kate Calder 20

---

**Diagram of the experimental design**

```

    graph LR
      RA[Random Assignment] --> G1[Group 1  
11,000 physicians]
      RA --> G2[Group 2  
11,000 physicians]
      G1 --> T1[Treatment 1  
Aspirin]
      G2 --> T2[Treatment 2  
Placebo]
      T1 --> OA[Observe heart attacks]
      T2 --> OA
  
```

Statistics 528 - Lecture 11  
Prof. Kate Calder 21

---

We call this type of designs **completely randomized** – all the experimental units are allocated at random among all the treatments.

Completely randomized designs can compare any number of treatments.

**Example:** An experiment is conducted to compare a new variety of corn called opaque-2 with normal corn as food for chicks. The researchers decide to serve each of the two types of corn at two protein levels: 15% protein and 20% protein. They feed each diet to 10 one-day-old chicks and record their weight gains after 21 days.

Statistics 528 - Lecture 11  
Prof. Kate Calder 22

---

**Questions**

- What are the experimental units and the response variable in this experiment?
- How many factors are there? How many treatments? Use a diagram to lay out the treatments.
- How many experimental units does the experiment require?

Statistics 528 - Lecture 11  
Prof. Kate Calder 23

---

- Use a diagram to describe a completely randomized design for this experiment.

```

    graph LR
      WAC[Weigh All Chicks] --> RA1[Random Assignment]
      RA1 --> CV[20 Chicks - Normal Variety]
      RA1 --> CO[20 Chicks - Opaque-2]
      CV --> RA2[Random Assignment]
      CO --> RA3[Random Assignment]
      RA2 --> C15[10 chicks - 15% protein]
      RA2 --> C20[10 chicks - 20% protein]
      RA3 --> C15[10 chicks - 15% protein]
      RA3 --> C20[10 chicks - 20% protein]
      C15 --> MWG[Measure Weight Gain]
      C20 --> MWG
  
```

Statistics 528 - Lecture 11  
Prof. Kate Calder 24

---

If in a randomized, comparative experiment we find differences in average response of the groups, we can say that this difference is either due to the differences in treatments (**statistical significance**), or due to **natural variability in the units**.

The final idea of experimental design is to use **replication**.

- For example, if we had just two subjects available and we randomly assign one to each treatment, the treatment effect would be confounded with the differences between the subjects.
- If the treatments are given to many experimental units, however chance variation is reduced.

Statistics 528 - Lecture 11  
Prof. Kate Calder 25

---

- **Control/Comparison** - control of the effects of lurking variables on the response, most simply by comparing treatments
- **Randomization** - the use of impersonal chance to assign experimental units to treatments.
- **Replication** - repeating the experiment on many units to reduce chance variation in the results

Statistics 528 - Lecture 11  
Prof. Kate Calder 26

### Statistical Significance

---

If we observe a difference in responses that is so large that it would *rarely occur by chance*, then we say the difference is **statistically significant**.

If we observe statistically significant differences among the groups in a comparative randomized experiment (with double-blinding), we have a good evidence that the treatments actually caused these differences.

Statistics 528 - Lecture 11  
Prof. Kate Calder 27

### Cautions

---

- **Generalizing**: The results of an experiment only apply to people or units that are similar to the subjects or units used in the experiment.  

For example, a study uses males in their 40's as subjects and concludes that taking aspirin daily reduces the risk of having a heart attack. Can we conclude that taking aspirin daily will reduce the risk of heart attack for women? What about for men in their 80's? **NO**
- **Lack of Realism**: The subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study.  

For example, when political scientists study how voters form opinions of candidates, participants know they are part of an experiment and often watch commercials or read pamphlets for fake candidates.

Statistics 528 - Lecture 11  
Prof. Kate Calder 28