Example 3

Two toothpastes are being studied for effectiveness in reducing the number of cavities in children. There are 100 children available for the study.

A) How do you assign the subjects?

Randomly divide children into two groups. Pull names out of hat

B) What do you measure?

Number of cavities before specific toothpaste and after using

C) What baseline data should you know about?

D) What factors might confound this experiment?

Dietary habits, economic status

E) What would be the purpose of a randomization in this problem?

To try and "balance out" the variables that could affect the number of cavities

There is going to be a lottery for candy in class today.

Use the random number generator to select 3 winners

Do this 3 times and write out how you used the random number generator to select the people.

11164	36318	75061	37674	26320	75100	10431
21215	91791	76831	58678	87054	31687	93205
10438	44482	66558	37649	08882	90870	12462
36792	26236	33266	66583	60881	97395	20461
73944	04773	12032	51414	82384	38370	00249

1	Aideyan, Olufemi
2	Breaux, Eric
3	Burns, Ty
4	Byun, Young-Joo (Brian)
5	Cheung, Monica
6	Choi, Justin
7	Doan, Bao
8	Fonts, Algy
9	Ghosh, Moyna
10	Griffin-Bey, Sabionn
11	Kim, Se Yeon
12	LaRoche, Daniel
13	Lee, Vivian
14	Leveille, Sophie
15	Ma, Jung Hee (Christina)
16	Mehta, Rajvee

17	Miller, Rebecca
18	Miller, Sarah
19	Padman, Neal
20	Park, Yeji (Lillian)
21	Rahman, Saadman
22	Shah, Avani
23	Sheikh, Firas
24	Thackston, William (Jeffrey)
25	Valentine, Brianna
26	Velazquez Rodriguez, Eri
27	Wang, Yue (Lydia)
28	Weber, Hedaya
29	Widjaja, Jonathan
30	Wiltshire, Charis
31	Ye, Wenxiong

Who wins the candy?

Who wins the candy?

Who wins the candy?

Chapter 13

Day 1

Observational Studies

- In an observational study, researchers don't *assign* choices; they simply observe them.
  - The text's example looked at the relationship between music education and grades.
  - Since the researchers did not assign students to get music education and simply observed students "in the wild," it was an observational study.

Slide 13 - 2

Observational Studies (cont.)

- Because researchers in the text example first identified subjects who studied music and then collected data on their past grades, this was a retrospective study.
- Had the researchers identified subjects in advance and collected data as events unfolded, the study would have been a prospective study.

Slide 13 - 3

Observational Studies (cont.)

- Observational studies are valuable for discovering trends and possible relationships.
- However, it is not possible for observational studies, whether prospective or retrospective, to demonstrate a causal relationship.

Slide 13 - 4

Basic Parts of Experiments

- Experimental units – individuals on which experiment is done
  - Subjects – experiment units that are human beings
- Treatment – specific experimental condition applied to units
  - Factors – the explanatory variables in the experiment
  - Level – the combination of specific values of each of the factors

Principles of Experimental Design

The basic principles of statistical design of experiments are

1. **Control** the effects of lurking variables on the response, most simply by comparing two or more treatments.
2. **Replicate** each treatment on many units to reduce chance variation in the results.
3. **Randomize**—use impersonal chance to assign experimental units to treatments.

Basic Principles of DoE

- Control
  - Overall effort to minimize variability in the way the experimental units are obtained and treated
  - Attempts to eliminate the confounding effects of extraneous variables (those not being measured or controlled in the experiment, aka lurking variables)
- Randomization
  - Rules used to assign the experimental units to the treatments
  - Uses impersonal chance to assign experimental units to treatments
  - Increases chances that there are no systematic differences between treatment groups
- Replication
  - Use enough subjects to reduce chance variation
  - Increases the sensitivity of the experiment to differences between treatments

Experimental Variability

Any experiment is likely to involve three kinds of variability:

- Planned, systematic variability. This is the kind we want since it includes the differences due to the treatments.
- Chance-like variability. This is the kind our probability models allow us to live with. We can estimate the size of this variability if we plan our experiment correctly.
- Unplanned, systematic variability. This kind threatens disaster! We deal with this variability in two ways, by randomization and by blocking. Randomization turns unplanned, systematic variation into planned, chance-like variation, while blocking turns unplanned, systematic variation into planned, systematic variation.

The management of these three sources of variation is the essence of experimental design.

Taken from In Introduction to the Design and Analysis of Experiments, George Cobb (1998)

Steps in Experimental Design

- Identify the problem to be solved
- Determine the Factors that Affect the Response Variable
- Determine the Number of Experimental Units
  - Time
  - Money
- Determine the Level of Each Factor
  - Control – fix level at one predetermined value
  - Manipulation – set them at predetermined levels
  - Randomization – tries to control the effects of factors whose levels cannot be controlled
  - Replication – tries to control the effects of factors inherent to the experimental unit
- Conduct the Experiment
- Test the claim (inferential statistics)

Statistically Significant

Statistical Significance

An observed effect so large that it would rarely occur by chance is called **statistically significant.**

Large sample sizes can force results that statistically significant but are not practically significant

Example: Milk consumption doubles your risk for a certain type of cancer from 1 in 10 million to 1 in 5 million

Remember our definition of unusual results (less than a 5% chance of occurrence

Analyzing Experiments Template

Topic	Answers
Research Question:	*What is the question the researchers are trying to answer?*
Subjects / Experimental Units:	*What are the experimental units?*
Explanatory Variable(s) / Factor(s):	*Type of variable: Quantitative or Categorical*
Treatment(s):	*What are the Factor(s) and their Levels?*
Response Variable(s):	*Type of variable: Quantitative or Categorical*
Experimental Design Description:	*Using words or diagrams describe the experimental design*
Experimental Design Principles:	*Explain how these design principles apply in this study*
Control:	*Eliminate confounding effects of extraneous variables*
Randomization:	*No systematic difference between the groups*
Replication:	*Reducing role of chance in results*
Blocking:	*If blocking used, describe the blocking / why it was used.*
Blinding:	*If blinding used, describe it in context.*
Concerns:	*What concerns about the experimental design?*
Statistical Analysis Technique(s):	*What statistical analysis techniques are appropriate?*
Conclusions:	*What conclusions can be drawn from the study?*

Experimental Problem Outline

- **Experimental Units** – what are our experimental units
- **Response Variable** – what are we measuring and how to determine good vs bad results
- **Explanatory Variables** – what other variables are we measuring, or changing to affect the response
  - These should include any factors and their levels
- **Assignment to Groups (blocking)** – how do you randomly assign experimental units into groups
  - Must be detailed enough for someone to duplicate
- **Assignment of Treatments** – how do you assign treatments to experimental units
  - Must be detailed enough for someone to duplicate
  - Double blindness can be discussed here if appropriate

ADHD Linked to Lead and Mom's Smoking, by Karen Barrow (February 1, 2007):

> *A mother's smoking during pregnancy and exposure to lead significantly increases her child's risk for developing attention deficit hyperactivity disorder (ADHD), say researchers. In fact, as many as one third of cases of ADHD in children are linked to exposure to tobacco smoke and lead before birth, giving moms yet another reason to quit smoking during pregnancy.*
>
> *For the study, researchers from Cincinnati Children's Hospital Medical Center surveyed over 4,700 children between the ages of 4 and 15 and their parents. Over 4 percent of the children included had ADHD. The researchers found that those children whose mother smoked during pregnancy were over twice as likely to develop ADHD than a child whose mother had not smoked. In addition, a child who had been exposed to lead, giving them high lead blood levels, were four times as likely to have ADHD, as compared to a child with low lead levels in his blood.*

Based on this study, should we conclude that smoking during pregnancy *causes* an increase in the likelihood that a child develops ADHD? Explain.

Explain the concept of confounding in the context of this study.

What are some differences between an observational study and an experiment?

What's the difference between an explanatory variable and a response variable?

What is a lurking variable? What two problems can they cause?

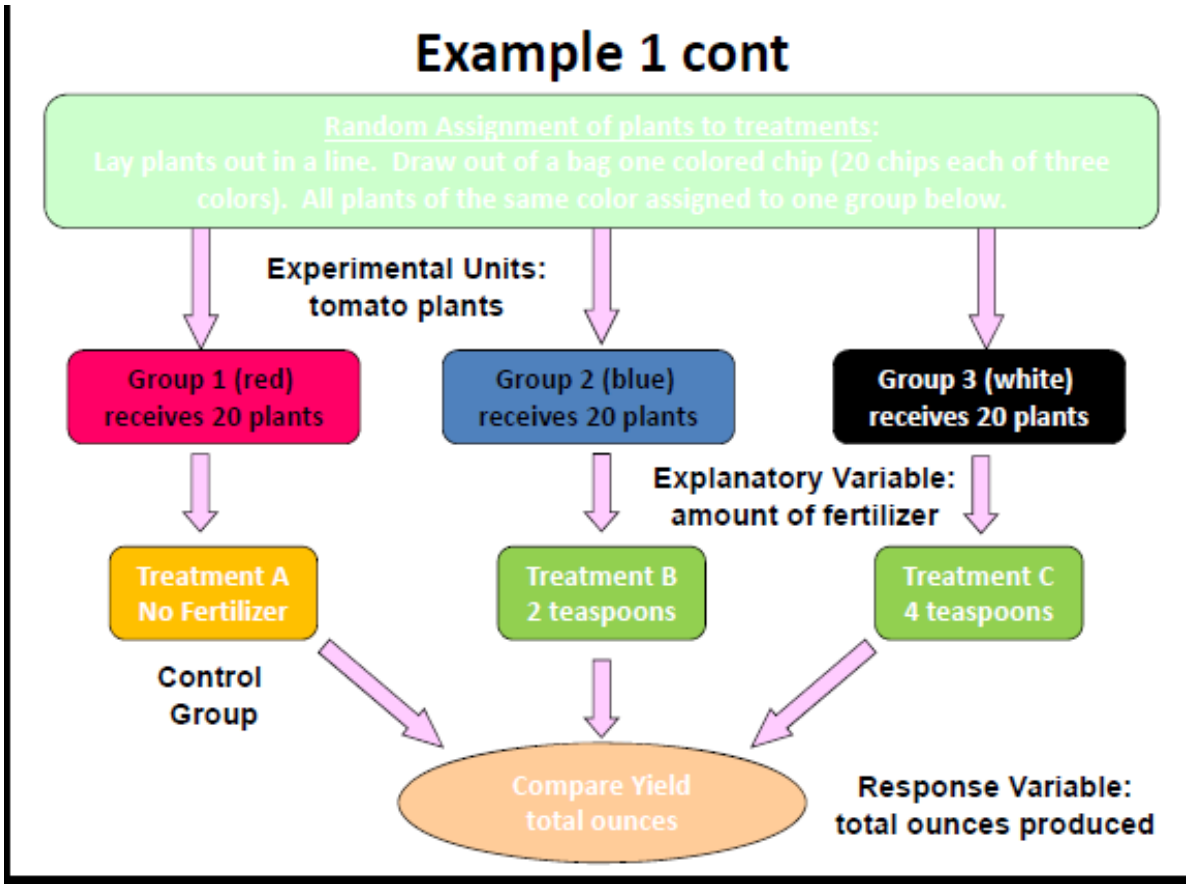
Example 1

Draw a picture detailing the following experiment:

A statistics class wants to know the effect of a certain fertilizer on tomato plants. They get 60 plants of the same type. They will have two levels of treatments, 2 and 4 teaspoons of fertilizer. Someone suggests that they should use a control group.

The picture should include enough detail for someone unfamiliar with the problem to understand the problem and be able to duplicate the experiment.

Example 1 cont



Example 2

A baby-food producer claims that her product is superior to that of her leading competitor, in that babies gain weight faster with her product. As an experiment, 30 healthy babies are randomly selected. For two months, 15 are fed her product and 15 are feed the competitor's product. Each baby's weight gain (in ounces) was recorded.

A) How will subjects be assigned to treatments?

No details given. Poor description of random selection
Two random selections taking place:
getting the 30 and then assigning them to the two products

B) What is the response variable?

Baby's weight gain in ounces

C) What is the explanatory variable?

Baby food brands

Example 3

Two toothpastes are being studied for effectiveness in reducing the number of cavities in children. There are 100 children available for the study.

A) How do you assign the subjects?

Randomly divide children into two groups. Pull names out of hat

B) What do you measure?

Number of cavities before specific toothpaste and after using

C) What baseline data should you know about?

D) What factors might confound this experiment?

Dietary habits, economic status

E) What would be the purpose of a randomization in this problem?

To try and "balance out" the variables that could affect the number of cavities

Example 4

We wish to determine whether or not a new type of fertilizer is more effective than the type currently in use. Researchers have subdivided a 20-acre farm into twenty 1-acre plots. Wheat will be planted on the farm, and at the end of the growing season the number of bushels harvested will be measured.

A) How do you assign the plots of land?

Blocking?? Before randomly assigning plots

B) What is the explanatory variable?

Types of fertilizer

C) What is the response variable?

Number of bushels of wheat harvested

D) How many treatments are there?

Two – new fertilizer and old (possibly none as a control group)

E) Are there any possible lurking variables that would confound the results?

Soil composition, rainfall, animal destruction effects

Example 4

We wish to determine whether or not a new type of fertilizer is more effective than the type currently in use. Researchers have subdivided a 20-acre farm into twenty 1-acre plots. Wheat will be planted on the farm, and at the end of the growing season the number of bushels harvested will be measured.

A) How do you assign the plots of land?

Blocking?? Before randomly assigning plots

B) What is the explanatory variable?

Types of fertilizer

C) What is the response variable?

Number of bushels of wheat harvested

D) How many treatments are there?

Two – new fertilizer and old (possibly none as a control group)

E) Are there any possible lurking variables that would confound the results?

Soil composition, rainfall, animal destruction effects