



M1-02: Experimental Design

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-02/>

Experimental Design

Statistical studies or experiments are done to see if a _____ has an effect on the outcome. This outcome is called the _____.

Consider a study based on the point of view of the researcher designing it:

- If the **researcher decides who gets the treatment**, the investigation is called an _____.
 - In a controlled experiment, researchers identify two groups:
Group 1:

Group 2:
- If the **researcher** just studies the effects of the treatment but **plays no part in choosing who gets the treatment** then the investigation is called an _____.

When designing experiments, the goal is to make the treatment group and control groups as alike as possible. There are many ways to divide the subjects into two groups.

However, randomization is best! Randomly dividing the subjects into the 2 groups is the most likely to make the treatment and control groups as alike as possible because it eliminates human _____. With enough subjects, differences average out. Not only differences that the researcher has identified as relevant, but on all characteristics, including the hidden ones that the researcher might not realize are important.

IDEAL Experimental Design

Randomized Controlled Double-Blind Experiment

Randomized controlled double-blind experiments have been a **gold standard** in the medical field for a while. They are also becoming more commonly used in other fields such as economics. A randomized controlled double-blind experiment must meet three criteria:

1. _____: Random assignment to treatment and control works best to make the treatment and control groups as alike as possible because it eliminates systematic differences (bias). With enough subjects, random differences average out.
2. _____: Make sure to have an explicit comparison group. Without one, you may see more positive results than what really exists.
3. _____: Neither the subjects nor those who are evaluating them know who is in the treatment and control group.



M1-02: Experimental Design

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-02/>

Why is Double-Blind Important?

1. Whether people think they have received the treatment can affect their response. To separate the effects of the actual treatment from the idea of treatment, the subjects shouldn't know which group they are in. In other words, they should be "blind" to this knowledge

This can be achieved by giving the control group a fake treatment called a _____. Since both groups believe they've been given the treatment, any differences in their responses can be attributed to the treatment itself and not the idea of treatment.

2. Knowing which subjects received the treatment and which did not can bias the people evaluating the results. To eliminate this bias, evaluators should be "blind" to this knowledge.

Puzzle #1: A study was done to test whether Ginkgo biloba (GB) could alleviate the symptoms of Alzheimer's and dementia. The 52-week study included 500 participants with Alzheimer's and dementia. Half were randomly assigned to either take GB daily or to take a sugar pill daily. Neither the subjects nor evaluators knew who was in each group. At the end of the study, there was significant evidence that GB improved cognitive performance and the social functioning of the patients for 6 months to 1 year

a) Who are the subjects? What is the treatment? What is the response?

b) Is there a control group? If so, who is in it?

c) Was there a placebo? If so, what was it?

d) Was this experiment double-blind?

e) What can we conclude and why?

Puzzle #2: A statistical study was done to test the effectiveness of a new diet pill this year at Carle Hospital. They wanted to determine if the diet pill worked to help patients lose weight after 6 months. They recruited volunteers for the study and decided to give all of them the diet pill. They wanted to compare their weight loss percentage to patients from last year who didn't take the diet pill.

a) Is this an ideal experimental design? If so, why? If not, why not?

b) How would you design this experiment if you were able to choose?