Intro to Experiment Design and Recommendation Engines

<https://youtu.be/gXlqR86h0yI>

**Experimental Design & Recommendation Engines**

The upcoming course is comprised of two broad topics: **experimental design** and **recommendation engines**.

**Experimental Design**

Within the **experimental design** portion of this course, there are three lessons:

**I. Concepts of Experiment Design**  
Here you will learn about what it means to run an experiment, and how this differs from observational studies. Topics include not only what to include in an experimental design, but also what to watch out for when designing an experiment.

**II. Statistical Considerations in Testing**  
In this lesson, you will learn about statistical techniques and considerations used when evaluating the data collected during an experiment. It is expected that you come into the course with knowledge about inferential statistics; the content here will see you applying that knowledge in different ways.

**III. A/B Testing Case Study**  
Here, you will put your skills to work to analyze data related to a change on a web page designed to increase purchasers of software.

At the end of these lessons, you can complete an optional **Portfolio Exercise** in coordination with our industry partner Starbucks. In the project, you will learn more about how Starbucks conducts experiments. You will also get access to data related to a hiring screen that was once conducted by Starbucks to test ideas related to experimental design and statistical metrics.

**Recommendation Engines**

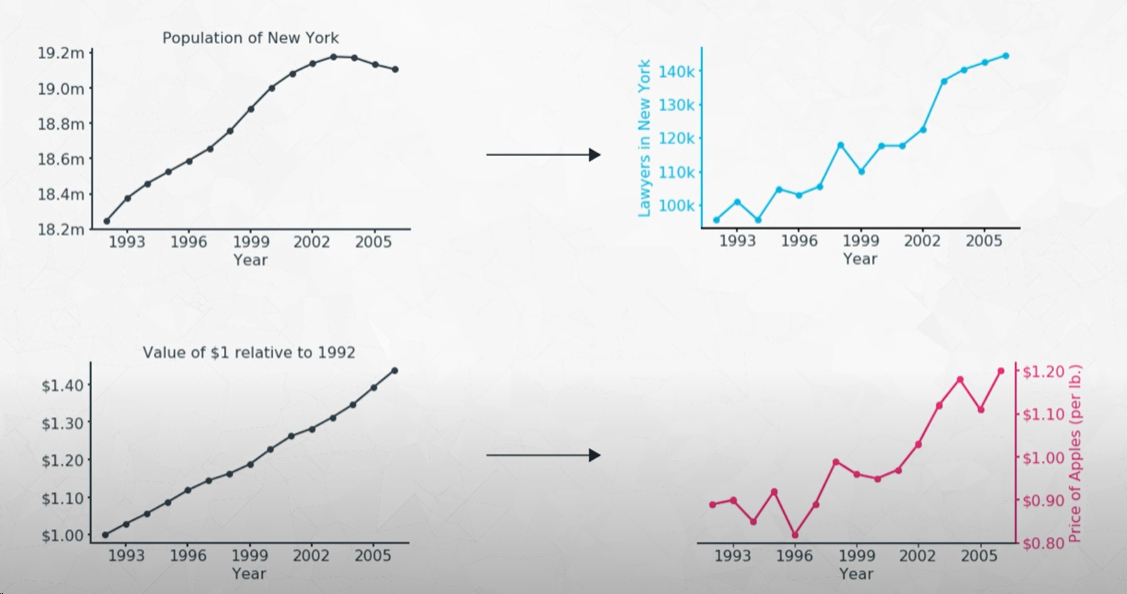
The **recommendation engines** portion of the course has two lessons:

**I. Introduction to Recommendation Engines**  
In this lesson, you will learn about the main ideas associated with recommendation engines. This includes techniques and measures of effectiveness.

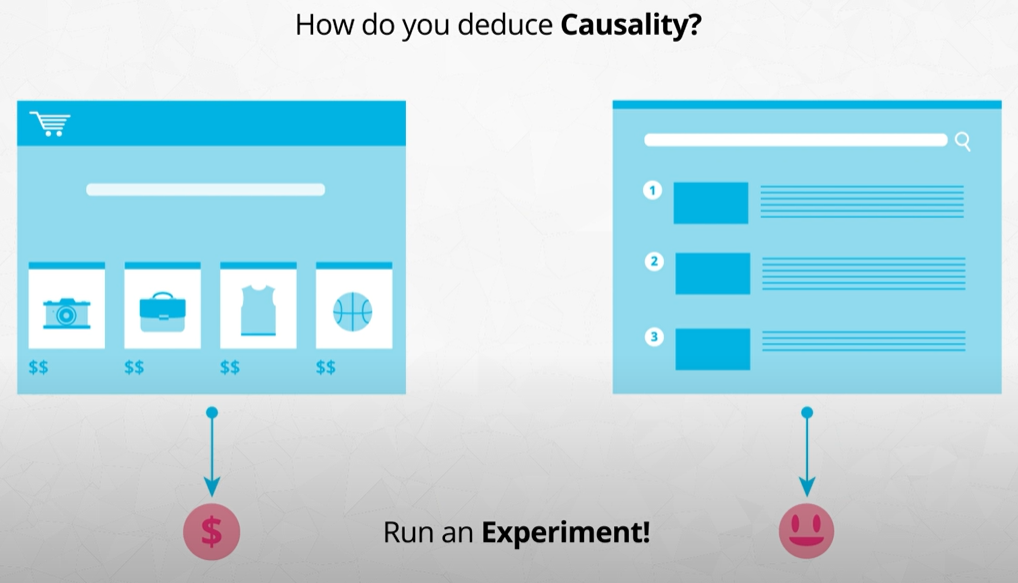
**II. Matrix Factorization for Recommendations**  
Extending on the previous lesson, you will learn about one of the most popular techniques for recommendation engines known as FunkSVD. You will also complete a class that brings together a number of techniques to make recommendations for a number of different scenarios.

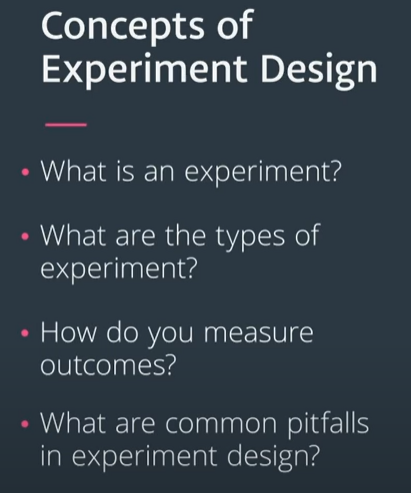
**Project**

At the end of this course, you will complete a project that uses data from the IBM Watson Studio platform to make recommendations for which articles a user should engage with! Hope you are excited to get started!

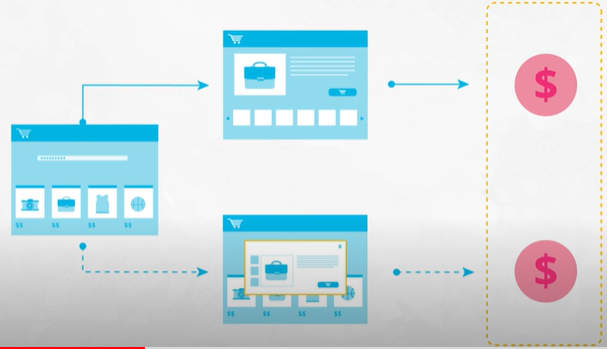


Correlation does not imply causation





What is an experiment Part I



Does the new feature bring more money?

* To test this an experiment is needed

Key features of an experiment:

1. **Comparison**: Comparison between groups (Webseite ohne Overlay vs. Webseite mit Overlay)
2. **Random assignement to groups**: Control other variables via randomization (by assigning visitors to one or the other webpage other variables like age, gender etc. should be equally distributed). The only practical feature between the groups should be the feature we care about.

If a true experiment cannot be run:

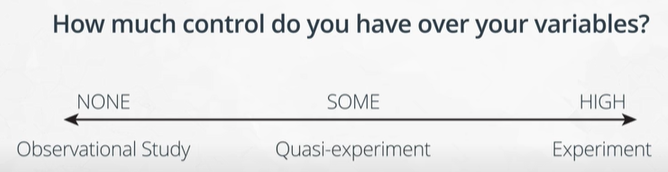
Reasons:

* Not having multiple groups
* Not being able to control random assignment to groups

Types of study depend on the amount of control over the variables in play.

High level of control 🡪 experiment.

No control 🡪 observational study



Example for observational study:

Hypothesis: Smoking leads to cancer, to conduct an experiment here is not possible, only observations are reasonable

<https://youtu.be/fH_xF5_SDCE>?

<https://youtu.be/PYzN1usi7QY>

## Types of Study

There are many ways in which data can be collected in order to test or understand the relationship between two variables of interest. These methods can be put into three main bins, based on the amount of control that you hold over the variables in play:

* If you have a lot of control over features, then you have an **experiment**.
* If you have no control over the features, then you have an **observational study**.
* If you have some control, then you have a **quasi-experiment**.

While the experiment is the main focus of this course, it's also useful to know about the other types of study so that you can use them in effective ways, especially if an experiment cannot be run.

### Experiments

In the social and medical sciences, an experiment is defined by comparing outcomes between two or more groups, and ensuring equivalence between the compared groups except for the manipulation that we want to test. Our interest in an experiment is to see if a change in one feature has an effect in the value of a second feature, like seeing if changing the layout of a button on a website causes more visitors to click on it. Having multiple groups is necessary in order to compare the outcome for when we apply the manipulation to when we do not (e.g. old vs. new website layout), or to compare different levels of manipulation (e.g. drug dosages). We also need equivalence between groups so that we can be as sure as possible that the differences in the outcomes were only due to the difference in our manipulated feature.

Equivalence between groups is typically carried out through some kind of randomization procedure. A **unit of analysis** is the entity under study, like a page view or a user in a web experiment. If we randomly assign our units of analysis to each group, then on the whole, we should expect the feature distributions between groups to be about the same. This theoretically isolates the changes in the outcome to the changes in our manipulated feature. Of course, we can always dig deeper afterwards to see if certain other features worked in tandem with, or against, our manipulation.

### Observational Studies

In an experiment, we exert a lot of control on a system in order to narrow down the changes in our system from one source to one output. Observational studies, on the other hand, are defined by a lack of control. Observational studies are also known as naturalistic or correlational studies. In an observational study, no control is exerted on the variables of interest, perhaps due to ethical concerns or a lack of power to enact the manipulation. This often comes up in medical studies. For example, if we want to look at the effects of smoking on health, the potential risks make it unethical to force people into smoking behaviors. Instead, we need to rely on existing data or groups to make our determinations.

We typically cannot infer causality in an observational study due to our lack of control over the variables. Any relationship observed between variables may be due to unobserved features, or the direction of causality might be uncertain. (We'll discuss this more later in the lesson.) But simply because an observational study does not imply causation does not mean that it is not useful. An interesting relationship might be the spark needed to perform additional studies or to collect more data. These studies can help strengthen the understanding of the relationship we're interested in by ruling out more and more alternative hypotheses.

### Quasi-Experiments

In between the observational study and the experiment is the quasi-experiment. This is where some, but not all, of the control requirements of a true experiment are met. For example, rolling out a new website interface to all users to see how much time they spend on it might be considered a quasi-experiment. While the manipulation is controlled by the experimenter, there aren't multiple groups to compare. The experimenter can still use the behavior of the population pre-change and compare that to behaviors post-change, to make judgment on the effects of the change. However, there is the possibility that there are other effects outside of the manipulation that caused the observed changes in behavior. For the example earlier in this paragraph, it might be that users would have naturally gravitated to higher usage rates, regardless of the website interface.

As another example, we might have two different groups upon which to make a comparison of outcomes, but the original groups themselves might not be equivalent. A classic example of this is if a researcher wants to test some new supplemental materials for a high school course. If they select two different schools, one with the new materials and one without, we have a quasi-experiment since the differing qualities of students or teachers at those schools might have an effect on the outcomes. Ideally, we'd like to match the two schools before the test as closely as possible, but we can't call it a true experiment since the assignment of student to school can't be considered random.

While a quasi-experiment may not have the same strength of causality inference as a true experiment, the results can still provide a strong amount of evidence for the relationship being investigated. This is especially true if some kind of matching is performed to identify similar units or groups. Another benefit of quasi-experimental designs is that the relaxation of requirements makes the quasi-experiment more flexible and easier to set up.

## Types of Experiment

## <https://youtu.be/7ihDj4M7EiU>

Most of the time, when you think of an experiment, you think of a **between-subjects** experiment. In a between-subjects experiment, each unit only participates in, or sees, one of the conditions being used in the experiment. The simplest of these has just two groups or conditions to compare. In one group, we have either no manipulation, or maintenance of the status quo. This is like providing a known drug treatment, or an old version of a website. This is known as the **control group**. The other group includes the manipulation we wish to test, such as a new drug or new website layout. This is known as our **experimental group**. We can compare the outcomes between groups (e.g. recovery time or click-through rate) in order to make a judgement about the effect of our manipulation. (Since we have an experiment, we'll randomly assign each unit to either the control or experimental group.) For web-based experiments, this kind of basic experiment design is called an **A/B test**: the "A" group representing the old control, and "B" representing the new experimental change.

We aren't limited to just two groups. We could have multiple experimental groups to compare, rather than just one control group and one experimental group. This could form an A/B/C test for a web-based experiment, with control group "A" and experimental groups "B" and "C".

If an individual completes all conditions, rather than just one, this is known as a **within-subjects** design. Within-subjects designs are also known as repeated measures designs. By measuring an individual's output in all conditions, we know that the distribution of features in the groups will be equivalent. We can account for individuals' aptitudes or inclinations in our analysis. For example, if an individual rates three different color palettes for a product, we can know if a high rating for one palette is particularly good compared to the others (e.g. 10 vs. 5, 6) or if it's not a major distinction (e.g. 10 vs. 8, 9).

Randomization still has a part in the within-subjects design in the order in which individuals complete conditions. This is important to reduce potential bias effects, as will be discussed later in the lesson. One other downside of the within-subjects design is that it's not always possible to pull off a within-subjects design. For example, when a user visits a website and completes their session, we usually can't guarantee when they'll come back. The purpose of their following visit also might not be comparable to their first. It can take a lot more effort in control in order to set up an effective within-subjects design.

### Side Note: Factorial Designs

Factorial designs manipulate the value of multiple features of interest. For example, with two independent manipulations "X" and "Y", we have four conditions: "control", "X only", "Y only", "X and Y". Experimental designs where multiple features are manipulated simultaneously are more frequently seen in engineering and physical sciences domains, where the system units tend to be under stricter control. They're less seen in the social and medical realms, where individual differences can impede experiment creation and analysis.

