

1 Introduction

1.1 Notation and Nomenclature

- **Example Experiment 1: List View vs. Tile View**

Suppose that [Nike](#), the athletic apparel company is experimenting with their mobile shopping interface and they are interested in determining whether changing the user interface from *list view* to *tile view* (see Figure 1) will increase the proportion of customers that proceed to checkout.

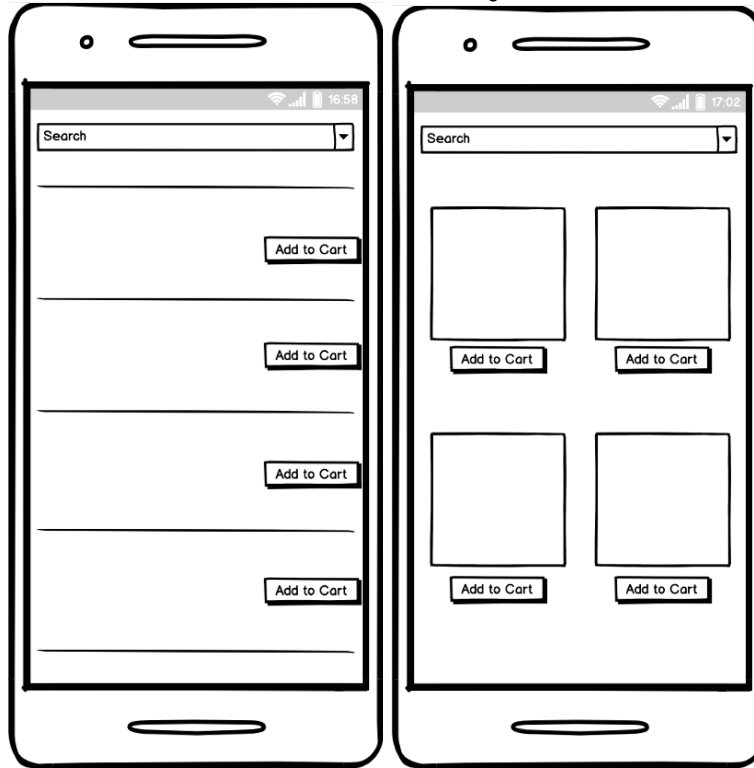


Figure 1: Nike Experiment. List View (left) vs. Tile View (right).

- **Example Experiment 2: Ad Themes**

Suppose that [Nixon](#), the watch and accessories brand, is experimenting with four different video ads that are to be shown on Instagram. The first has a surfing theme, the second has a rock climbing theme, the third has a camping theme, and the fourth has an urban professional theme. Interest lies in determining which of the four themes, on average, is watched the longest.

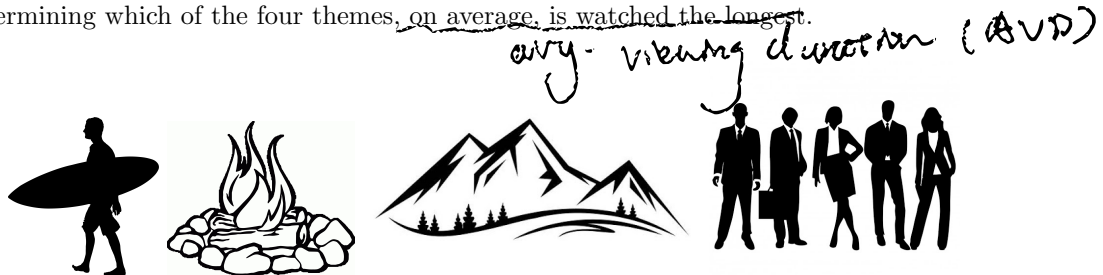


Figure 2: Nixon Experiment. Four different ad themes (from left to right: surfing, camping, rock climbing, urban professional).

Metric of Interest: The statistic the experiment is meant to investigate.

To be Max- or minimized

- E.g.:

- E.g.:
Click through rates Avg. time on page
Bounce rate 95th percentile page load time
(worst cases)

4) **Response Variable:** The variable of primary interest.

response variable: The variable of primary interest.
(what needs to be measured to calc. the MOI)

- Nike Ex: binary indicator of whether a customer checked out

- Nixon Exp (continuous) viewing duration

✓) **Factor:** The variable(s) of secondary interest. (IND)

- I.e. features, predators, invertebrates

Levels: The values that a factor takes on in an experiment.

Discount example:

Gap offers a discount for 13. 7. 14 days

Experimental Conditions: The unique combinations of levels of one or more factors.

- I.E.: treatments, variants, buckets, splits
- E.C. are combinations of levels

Experimental Units: That which is assigned to the experimental conditions, and on which the response variable is measured.

Nike example: Nike app users

Nixon example: Instagram users

* in online experiments, the unit is a person
but it doesn't have to be
(can be a pair of persons)

1.2 Experiments versus Observational Studies

- An **experiment** is composed of a collection of conditions defined by purposeful changes to one or more factors.
- The goal is to identify and quantify the differences in response variable values across conditions.
- In determining whether a factor significantly influences a response, like whether a video ad's theme significantly influences its AVD, it is necessary to understand how experimental units respond when exposed to each of the corresponding conditions.

- BUT... ideally, observe the same units in different conditions
but we can only observe each unit in a single condition

- **Counterfactual:** the hypothetical and unobservable value of a unit's response in a condition to which they were not assigned.

alternative reality (response)

- Because counterfactual outcomes cannot be observed, we require a proxy.

Instead, assign different units to diff. conditions
approx. of response

- Ideally the only difference between the units in each condition is the fact that they are in different conditions.

as homogeneous as possible
guaranteed by randomization

- The key here is that the factors are purposefully controlled in order to observe the resulting effect on the response. This facilitates causal conclusions.

- In an **observational study**, on the other hand, there is no measure of control in the data collection process

factors are all observed
not controlled

- This hinders our ability to establish causal connections between the factor(s) and the response variable. However, sometimes we have no choice.

– **Unethical Experiment 1:** In evaluating whether smoking causes lung cancer, it would be unethical to have a 'smoking' condition in which subjects are forced to smoke.

– **Unethical Experiment 2:** In dynamic pricing experiments, it would be unethical to show different users different prices for the same product.

– **Unethical Experiment 3:** In social contagion experiments, it would be unethical to show some social network users consistently negative content and others consistently positive content. [But Facebook did this anyway.](#)

– **Unethical Experiment 4:** Mozilla conducted an investigation in which the company was interested in determining whether Firefox users that installed an ad blocker were more engaged with the browser. However, it would have been unethical to force users to install an ad blocker and so they were forced to perform an observational study with propensity score matching instead. [Read the paper here.](#)

A controlled experiment
is not always
the best

	Advantages	Disadvantages
Experiment	causal inference is clear	expensive, risky, unethical (maybe)
Observational Study	low cost, low risk no ethical concerns	causal inference is unclear

1.3 QPDAC: A Strategy for Answering Questions with Data

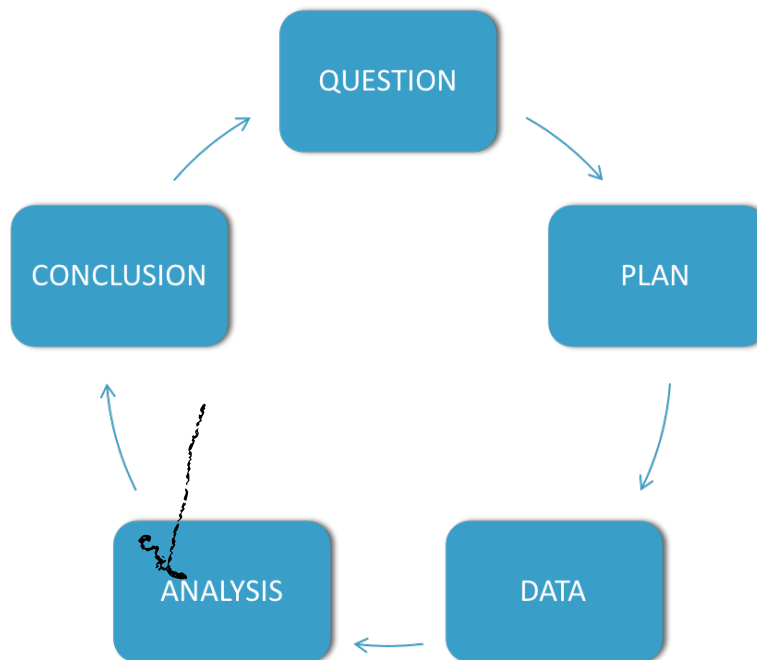


Figure 3: QPDAC Cycle

Question: Develop a clear statement of the question that needs to be answered.

concrete, widely communicated, quantifiable (in terms of NOI)

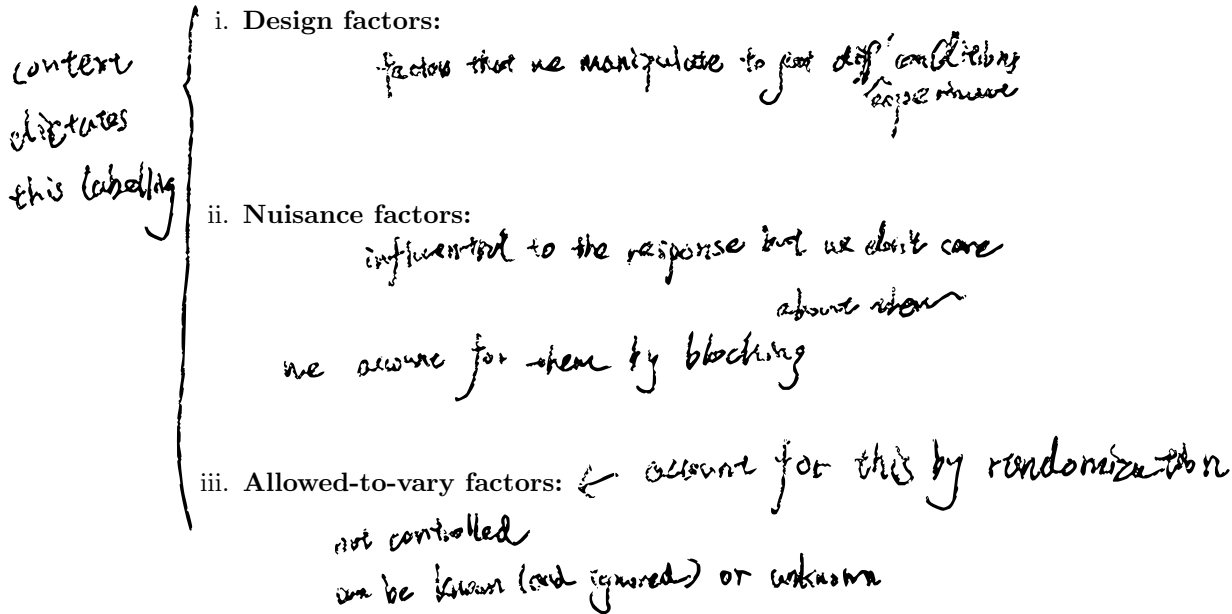
- Nike Example:

- Nixon Example:

Plan: In this stage, the experiment is designed and all pre-experimental questions should be answered.

- Choose the response variable: *calculated by Q. and NOI*

- Choose the factor(s): brainstorm all factors that might influence the response and make decisions about whether and how they will be controlled in the experiment.



- Choose the experimental units:
- Choose the sample size and sampling mechanism.

↓
how many units per condition?

Data: In this stage the data are collected according to the **Plan**. It is extremely important that this step be done correctly; the suitability and effectiveness of the analysis relies on the data being collected correctly. "Garbage in, garbage out".

Analysis: In this stage the **Data** are statistically analyzed to provide an objective answer to the **Question**.

- This is typically achieved by way of estimating parameters, fitting models, and carrying out statistical hypothesis tests
- If the experiment was well-designed and the data were collected correctly, this step should be straightforward

Conclusion: In this stage the results of the **Analysis** are considered and one must draw conclusions about what has been learned.

- These conclusions should then be clearly communicated to all parties involved in – or impacted by – the experiment

1.4 Fundamental Principles of Experimental Design

Randomization refers both to the manner in which experimental units are selected for inclusion in the experiment and the manner in which they are assigned to experimental conditions.

Thus we have two levels of randomization:

- The first level of randomization exists to ensure the sample of units included in the experiment is representative of those that were not.

due to difficulty of including all units & risky experiments!

- The second level of randomization exists to balance the effects of extraneous variables not under study (i.e., the allowed-to-vary factors).

minimizes counterfactuals and make comparison easy

Replication refers to the existence of multiple response observations within each experimental condition and thus corresponds to the situation in which more than one unit is assigned to each condition.

- Assigning multiple units to each condition provides assurance that the observed results are genuine, and not just due to chance.

- For instance, consider the Nike experiment introduced previously. Suppose the CORs in the *list view* and *tile view* conditions were respectively 0.5 and 1.

- How much replication is needed?

Power analysis, type I & II errors

Blocking is the mechanism by which nuisance factors are controlled for.

- To eliminate the influence of nuisance factors we hold them fixed during the experiment.

- Thus we run the experiment at fixed levels of the nuisance factors, i.e., within blocks.

- **Example:** Consider an email promotion experiment in which the primary goal is to test different variations of the message in the subject line with the goal of maximizing 'open rate'. However, suppose that it is known that 'open rate' is also influenced by the time of day and the day of the week that the email is sent.

Optional Exercises:

- Definitions: 1, 2, 3, 4