Experimentation Overview

Why Experiment in a Business?

A Business Question

We wish to increase prices for first time renewals. How much should we increase the price in order maximize our revenue?

Keep in mind, We are influencing customer's choices to stay or go:

At Original Pricing

At New Pricing	Retain	Churn
Retain	A%	B%
Churn	C%	D%

A% = Customers that will stay regardless
B% = Most likely 0%
C% = Customers churn due to higher price
D% = Customers that would churn anyway

An Analysis + Insight

Looking at all of our historical 1st time renewals, lets evaluate the % churn associated with various signup and increase prices.

For example,

- Signup $1 RPS w/ increase to $5 RPS = t% churn
- Signup $1 RPS w/ increase to $7 RPS = s% churn
- …
- Signup $1 RPS w/ no increase = u% churn

Therefore,

We can optimize our *short term* cash by identifying the best combination of signup price and price increase where :

Incremental cash from A% > **Lost cash from C%**

The Limitation

- We have not observed all signup + increase combinations (or if we have we don't have enough samples most likely)
- Products (or features) at various times
- Competitive landscape at various times
- Quality of support + platform (i.e. outages)
- Seasonality in churn
- Economy is changing
- Industry is changing
- Etc…..

Directionally, we could "subtract" off the average churn we see compared to various price increases, but this will not adjust for many of the items above and only identifies correlation NOT CAUSE and EFFECT.

Without cause and effect, you can't make the inference:

"I am confident that if we increase prices by $y then we will see C% churn, but still make money"

Experimentation and EIG

Business Question	Confounding in JUST Observation
1. **Cross sell opportunities:** Determine which customers are likely to buy ecommerce products that have not already.	1. Seasonality, self motivated customers, promotions, customer lifecycles, competitor pricing and landscape, "new awareness" (site positioning, alternative wording etc.)
2. **Product feature optimization:** Determine combination of features (or value proposition) of a new cloud hosting product that are most attractive to prospective customers or segments of customers.	2. Seasonality, competitor, survey results vs. actual behavior (*talk about more*), alternative product offerings and pricing.
3. **Improve online content to reduce support call volumes:** Determine what content and how to augment it in order to contain customer's on our site (not requiring an inbound call) through self-service support.	4. Seasonality, other drivers of calls, "stubborn" customers, contact channel preference, correlation vs. cause and effect
3. **Price increase opportunity:** Determine how much we can increase prices at renewal (lets say 1st time renewal for now) and increase revenue enough to offset the increase in churn.	3. Seasonality, incomplete view of start price + end price (price increase), promotions, customer type, changing processes and conditions, other factors of churn not related to price

Quick Ramp Up - Experimentation

General Description

Experimental Design is: An information-gathering exercise where variation is present (whether under the full control of the design or not), from which one can evaluate the "true" relationship between a stimuli and response or influences on a process to measure:

1. The effect on an experimental unit or process due to a cause or intervention.
2. Deeper understanding of a process too complex to observe with out structure.

In other words: We design an experiment to minimize "confounding" variables to identify the "true" relationship between a stimuli and response.
(i.e. price increase and churn)

Considerations for when not to use

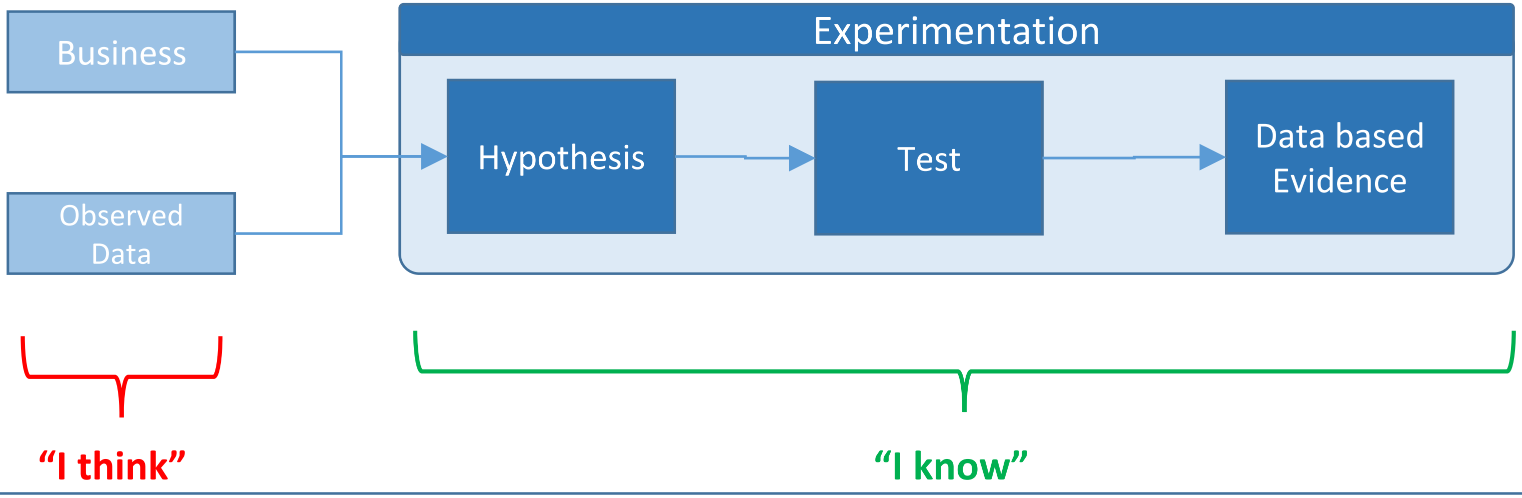
There are times to NOT push experimentation when it's not welcomed:

1. Looking for a directional answer or generating a hypothesis
2. Objective is historical description not prediction or causation
3. Insight requires fast delivery with no time to test
    1. Consider post implementation testing
4. Process is non-flexible, meaning it is extremely difficult to implement a useful experiment
5. Resources are sparse – data collection, development etc.
6. Unethical or not legal

In the Decision Process

Business Objective:

Going past "description" and into the realm of "prediction"



Experimental Design Process – High level

1 What are you trying to "prove"

Determine the stimuli and response you are interested in understanding as well as defining the scope of the experiment:

- Experimental unit
- Process
- Hypothesis
- Response
- Desired output / potential action

2 Identify confounders

Identify confounders of the response – what would influence the response variable other than the intended stimuli or intervention:

- What can you control
- What can't you control

3 Match a Design

Couple your experimental objective and the type of confounders to determine an experimental design structure to employ

- A/B or control/intervention?
- Blocking
- Factorial
- More advanced (response surface, mixture etc.)

We will talk more about these later

4 Identify Arms and Interventions

Whether there is a control and intervention group or multiple "arms" / groups, experimental units must be assigned to these groups and then a method of stimuli intervention determined:

- Experimental units assigned randomly to groups
- Stimuli methodically introduced

5 Execute

Perform the experiment and collect the data.

Once designed, this is usually in the hands of the business owners and data collection systems.

6 Evaluate

Analytically summarize and measure the results of the experiment. This will vary depending on the stimuli, response, objective and design.

In many cases, statistical methods are employed to determine "true" differences vs. noise or randomness in measurement / process.

We will talk more about these later

Analytical Experimental “Approximation”

There are ways of approximating an experiment with observed data, unfortunately it is completely dependent on the type of data that has been collected and the nuances of the process (i.e. confounding variables).

Taking the price increase example, if:

- We have historically observed all signup price and increase combinations we are interested in
- We had enough observations in in each sample set
- Knew exactly when promotions, campaigns and save interventions occurred
  - Or knew that they were not occurring
  - Or occurred a the same exact rate historically as they were going to into the future

Then: We could use evaluation techniques to determine how much churn resulted from signup price and increase combinations for current sets of customers, thus allowing optimization.

Examples of commonly used methods with observed data:

- Back testing - develop a model which describes the stimuli and response and “score” past (or a random subset) observations to test how accurate your representation is
- Forecasting - use a cut point in time where prior data is to “train” your model and future data is to “validate” your model.
- Quasi-testing – “train” model with up to date data and “test” on soon, but not yet collected data.
- Propensity Matching

Experimentation By Industry (Examples)

Engineering + Agriculture

Mechanistic systems with multiple inputs and one or multiple outputs that can be "optimized" or have to be within spec to be useable (not considered waste). These tend to have less noise presence and be governed by physical laws.

Examples?

Health + Medicine

Improvement of treatment and prevention through randomized control (or multi-intervention) trials. Usually, observing subjects for long periods of time where noise is overwhelming and difficult to control for.

Examples?

Behavioral Psychology

Studying observable behavior (often of humans) and drivers of this behavior influenced by various stimuli (which the subjects are explicitly aware of or not).

Examples?

Generalize

Process

Pretty much any system that has an input and output can be viewed as a process. In the process there are 1 or more known (or unknown) influences on variation. In the pursuit of improving a process (optimizing an output based on some requirement) experimental methods can be employed to guide towards success

Examples at EIG?

Behavior

Putting human behavior at the center of a decision based process, allows for similar inferences and experimentation as any process. However, the stimuli and corresponding behavior (outcome) has more variation associated with it and many more industry and department based applications.

Examples at EIG?

Basic Experimental Designs

Type	Description	When to Use
A/B Test & RT or RCT (Randomized Control Trial)	2 or more arms are identified, usually with 1 control (non-intervention) arm and 1 or more intervention arms. Random assignment of experimental units are to reduce bias with an additional evaluation to ensure pre-test balance.	This is your bread and butter testing design. It is “simple” and allows for easier explanation to business owners. Best for testing “intervention” vs. control or alternative as a whole – not for identifying the influence of a series of stimuli or changes. Also usually best in production situations
Factorial Designs Full or Fractional	**Full**: One or more factors, each with discrete possible values or "levels", and whose experimental units take on all possible combinations of these levels across all such factors. **Fractional**: only a subset of level combinations are taken to reduce trials necessary to make desired inferences	Allows the investigator to study the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable. RCT is like a single factorial design where the levels are Intervention or no intervention
Blocking	blocking is the arranging of experimental units in groups (blocks) that are similar to one another. In most cases, units within blocks are similar but between blocks are different. Reducing known variability is exactly what blocking does. **Example?**	Its principle lies in the fact that a certain variability cannot be overcome and is confounding the detection of how your inputs are influencing the output or response. Usually blocking is necessary in systematic or environmental circumstances that have limitations of control or adjustment.

More advanced methods exist – we will not get into them here

Considerations of Experimental Unit Assignment

The Objective: of experimental unit assignment is to allow for as much variation to be equally balanced across the arms / groups of the design (or in each trial). In ensuring "balance" you are in essence reducing (or eliminating) bias for one group or type of experimental unit to perform differently than others for reasons outside of the factors you have attempted to control and design for.

How to prove: the confirmation that you have a balanced study prior to execution is some form of magnitude and statistical test across the arms that have been assigned based on the outcome measure you are out to positively influence (i.e. if testing NPS, with a control and intervention group, you ensure that the NPS scores are directionally and statistically similar prior to test execution).

A common method: The experimental designs we have discussed are aimed at controlling variation in the process or behavior by either:

- **RT (RCT):** Randomly assigning across arms to attempt to eliminate bias and "balance" or evenly allocate variation to each group.
- **Factorial:** Random assignment combined with ensuring enough trials are down to estimate the variation seen across factorial combinations (we will talk more about this later)
- **Blocking:** Creating blocks or groups for similar experimental units so that we don't have to worry about difficult or impossible to control variation

The reason why "randomization" is used to take the selection process out of the experimenter where unknown or intended bias may originate. An additional level of robustness is called "stratified randomization" where there may be known "types" of experimental units in which case you want to distributed each type to the arms or groups of your study (i.e. Shared vs. Dedicated Hosting when attempting to x-sell a product) - this can actually be a form of "Blocking" design as well.

Evaluation – in Group Comparison (i.e. RCT)

Evaluation

- Evaluating an experiment can come in various forms, we will not go into all of them.
- Logically, evaluation occurs by testing the null hypothesis (Ho) by collected data (from the experiment) and determining if the evidence contradicts the null hypothesis.
  - Basic example:
    - Ho: Price increases will not increase churn
    - Goal: determine if that is false – if so, to what degree.
- To determine if the data you collected contradicts the null hypothesis, generally statistical techniques are used.
  - Which technique to use depends on the corresponding statistical assumptions which are affected by design type, data type, distribution, and potentially other factors

Statistical Step

Chi-Square Test

Test of Distribution

- Used with a discrete response variable (usually focused on binary, trinary)
- Evaluates the distribution of both "groups" of the experiment and determines if the observed frequencies are different than the expected (i.e. intervention vs. control)

See R code example

Purchase Rates

	Purchase	No Purchase
Intervention	100 (10%)	900 (90%)
Control	10 (1%)	990 (99%)

T-Test

Test of Centrality

- Used a continuous response variable (i.e. numeric and can take many -> infinite values)
- Evaluations the "middle point" of both "groups" of the experiment and determines the center points of both groups are enough away given observed variation (noise).

See R code example
*assuming normal distribution for now

Average Purchase Amount

	Average	Standard Deviation	Sample Size
Intervention	$21.2	$0.57	10
Control	$15.4	$0.56	10