The process

A process is nothing more than a series of steps that describe how to proceed about achieving some goal. In this case, the goal is to answer some kind of question using a quantitative research methodology.

Processes can vary in how detailed they are, and consequently on how flexible. We should always try to structure the things we do as detailed as possible. Perhaps it is not currently possible, but by always trying we can do it as soon as it can. In the specific case of the research process, there are very good reasons for this.

**Practical reasons:** Using a detailed structured research process increases efficiency; that is, it allows to achieve the same result with less resources. Specifically, doing research requires first the design of the research, and then execution. A process can help design faster by avoiding trial and error for every new project. At the same time, a process allows one to foresee future needs and perhaps even to automate them such that implementation also becomes more efficient. Finally, a process allows one to be more consistent such that results are replicable and mistakes can be more easily identified and corrected. To see under which conditions it makes sense to a process consider [the following theoretical generalization].

**Philosophical reasons:** In philosophy, the field of Epistemology studies the nature of knowledge; how can we be sure we “know” something. There we can find the observation that correctly answering a question today does not imply that one is knowledgeable, as that could have simply been chance. A process acts as a deductive argument. If it is valid and today its conclusion was true, then the validity can be verified tomorrow and it can be reasonably inferred that it will be truthful tomorrow as well. This simply cannot be argued if the correct prediction was not rooted in a deductive argument; a process, in this case.

In the following I describe the process I follow when conducting research. It is not a completely detailed process, and as such is still sufficiently flexible to be applied to a wide array of fields like Marketing, Operations, Economics, Politics and the hot new topic of Data Science.

# The process:

**Problem Definition:**

In this stage there are two problems to define: the decision problem, and the research problem. An appropriate definition of the former is essential to achieving an appropriate and, most importantly, meaningful definition of the later. In the context of Marketing, for example, the Marketing Decision Problem would be something like “how to allocate advertising spending across channels”. Only after that is known can one realize that the research problem is to evaluate the revenue efficiency of each advertising channel and then use that to identify the allocation that maximizes revenue.

A practical tip to achieve this is by focusing on developing a first version of a mathematical model which, if true, would give the decision maker the answer they are looking for. It is not that mathematical models are the only way to go, but the indirect benefits of developing one are substantial.

First, one is forced to decide on a number of (dependent) variables of interest. Most likely there should be only one, but it might be the case that some clients will have a few. Some because of legitimate reasons, others because they simply aren’t in the position to consider some variables as explanatory and take them as dependent. For example, Brand Awareness is not a dependent variable in the grander scheme, revenue is; but if the client is the Advertising Manager then that’s what he’s in charge of.

Second, a list of potential explanatory variables for each dependent variable is agreed upon. These are also likely the ones that the client has control over and leaving them out would render the results unable to aid in the decision problem.

Third, an operational definition of all variables is also agreed upon, even if susceptible to change later on. For example, Brand Loyalty might be defined as the portion of customers who re-purchase.

This first mathematical model is probably very simple, but it can serve to get the researcher and the client on the same page as to what to expect. It leaves both parties confident that if that result is achieved, the project is a success.

The first step requires defining the problem. The practical tip in this case is to discuss it in terms of an analytical model; preferably mathematical. The reason for this is that it forces everyone involved to make decisions that will strongly simplify the rest of the process. For example, the most important first step is defining what the dependent variable of interest is. It will most likely be the case that only one variable will be of utmost interest. For example, in the case of business it should be profits. However, there are situations in which more than one is of interest. For example, a business might care about generating profits and being socially responsible. It is very likely these two are interrelated, in which case their structure needs to be studied. However, it could be the case that the decision make cared independently about two dependent variables. Assuming that’s the case with the previous example, after finding an operational variable for social responsibility, the company would need to decide on weights for each of them. This is something that happens often when the client is only in charge of part of the business, say of Marketing. In this case, they might be in the situation of caring about two or more variables which are related in the big picture, but not in the small one, i.e. from his perspective. For example, caring about customer satisfaction and brand awareness; these two would presumably lead to profits, but from the perspective of marketing they follow two completely different processes.

The main message here is that forcing a client to discuss in the framework of an analytical model requires them to provide us with their preferences. After this, the rest of the process can be summarized as calculating the utility of each alternative and selecting the one that maximizes it.

It is important to know that by no one model is necessarily better than another. Surely, economic theory is to be preferred, but an argument needs to be made. The best one is that it is sufficiently wide and with good predictive ability. When necessary, theories that are based on observations or outcomes are to be preferred, behaviors when dealing with people and results when working with business operations. Psychological concepts like Satisfaction and Image can be useful only if combined with observable frameworks.

Also remember that models are a simplification of the world. The reasons why this is a good thing are explained in the previous section arguing the benefits of processes. In short, they allow one to discern between what’s known and what is not, and to focus on what’s known. In this sense, there are infinitely many things that can be considered when solving the problem of a client. However, one should proceed evolutionarily; starting simple, and adding layers of complexity. This will also help simplifying the work with the client because they will be able to follow logically what is happening, and seeing how the predictive success of the model is increasing gradually. What normally happens is that clients hear economic models and start pulling out criticisms of them. Even if one gives in and starts developing a model from scratch, this model will be understandable if built slowly from the simple to the more complex. When a client says that one is not considering the emotional effect of advertising while evaluating the advertising ROI, one need only point out that the model is necessarily incomplete, that it perhaps can be improved, but that the non-emotional component is real and is as reported. Ideally, the client will understand that, in time, all of their suggestions will be tested and if real incorporated into the model.

# The Marketing Research Process:

* Problem definition:
  + Environmental context of the problem:
    - Discussion with decision makers
    - Secondary data evaluation
  + Marketing Decision Problem:
  + Marketing Research Problem:
* Research Approach Developed:
  + Objective theoretical framework
    - Suggests key dependent and independent variables, and possibly also mediation variables.
    - Suggests operationalization of key variables; practical means of measuring them.
    - Suggests a research design; whether causal or descriptive.
    - Suggests how to select the sample of the population
    - Guide the selection of data analysis and interpretation of results
    - Contextualizes the findings of this research with those of previous research and other sources of knowledge.
  + Analytical model
    - Verbal
    - Graphical
    - Mathematical
  + Research questions: more specific questions about the problem that, if answered, satisfy the decision-maker in need of research.
  + Hypotheses: Implied by the analytical model, are assertions about the data that can be proven to be true or false.
* Research Design Developed:
  + Steps:
    - Define Information Needed (From theoretical framework, analytical model, research questions and hypotheses)
    - Decide if Exploratory (Data Mining, only on Panel) or Conclusive: Descriptive, or Causal
      * Descriptive: Pictures of the reality. A summary of descriptive statistics or metrics.
        + Cross-sectional (single)
        + Time Series (multiple cross-sections)
        + Longitudinal (panels)
      * Causal: using theoretical models, or experimentation to establish cause-effect relationships.
    - Design the sequence of techniques of understanding and measurement
    - Construct an appropriate form of data collection
    - Specify the sampling process and sampling size
    - Develop a plan of quantitative data analysis
* Secondary Data Collection:
  + Extract
    - Internal
      * Structured: metrics and KPIs already used for other purposes.
      * Semi-structured: operational or transactional data and others that would need some processing
      * Unstructured: text data like customer complaints, as well as other data collected with very foreign purposes.
    - External
      * Structured: Public databases, syndicated services
      * Unstructured: data on the web, could be text or require the combination of multiple data points to create an observation.
  + Transform:
  + Load:
* Primary Data Collection:
  + Sample or Experiment: Identify and select individuals for primary research
  + Fieldwork and Data Collection:
  + Data Preparation:
* Data Analysis:
  + Data Mining: Mindless correlations
  + Data Science
    - Descriptive Analysis
    - Model
      * Econometrics: Focused on causal inference
      * Machine Learning: Focused on accuracy of prediction
* Report Preparation and Presentation:

The Data Science Process: Notice that Data Science will always work with secondary information. Therefore, Data Science lives on before marketing research. The previous steps, however, will still need to be undertaking. The first step is answering a question, and then realizing that the data is there and doing “Data Science”. Organizations should try to gather as much data as possible, yes, but they shouldn’t expect to use it. They are saving on marketing research, and making better decisions. That’s where the value is. Lets not get it twisted. It might be the case that recording some data is pointless.

A Researcher gets research project requests that can be ad-hoc, for which a specific detailed and structured process is not applicable, or non-ad-hoc, for which such a process is applicable. The Researcher gets these ad-hoc projects with probability , and non-ad-hoc with probability .

The researcher has to decide whether to use or not a structured detailed process. For simplicity, assume that it doesn’t cost anything to develop said process, and that it is appropriate for the project at hand.

Let us assume that time is the only resource needed for taking on a project. Then, having a process would presumably reduce the amount of time needed to complete the project. However, this process would only be applicable to non-ad-hoc projects.

Assuming that the Researcher cares about minimizing project completion time because the results would be equivalent since the process is appropriate, then the expected time of completion using a detailed process is given by:

Here, refers to the amount of time it would take to do a non-ad-hoc project without the process. Similarly, refers to the amount of time it takes to do the ad-hoc project. The parameter represents how much time can be reduced by using the process.

We can assume both projects done without process take the same amount of time, that is . This implies that whether or not a process is intrinsically ad-hoc doesn’t matter, but rather what’s important is the way that it is undertaken.

Now, the decision rule is to choose to use processes if . To answer this, it is useful to find the threshold at which one is indifferent between the alternatives; that is, when the time is equal. This is done simply by equating both sides of the equation:

The factor is what determines whether or not it makes sense to use a process. The two parameters, and represent the likelihood of encountering an ad-hoc project and how much time is saved by using the process. Assuming that , and both types of projects are equally likely, then . If , then there is no benefit of doing a project with the process. The time is the same with or without. However, the moment , then it makes sense to use a process; assuming that Of course, as , needs to increase to compensate for it.

A function between and can be used to describe how much the latter needs to increase for each increase in the former.

The purpose of this example is twofold: to demonstrate the point of when using a process makes sense, and to serve as a first example of applying the process here described.