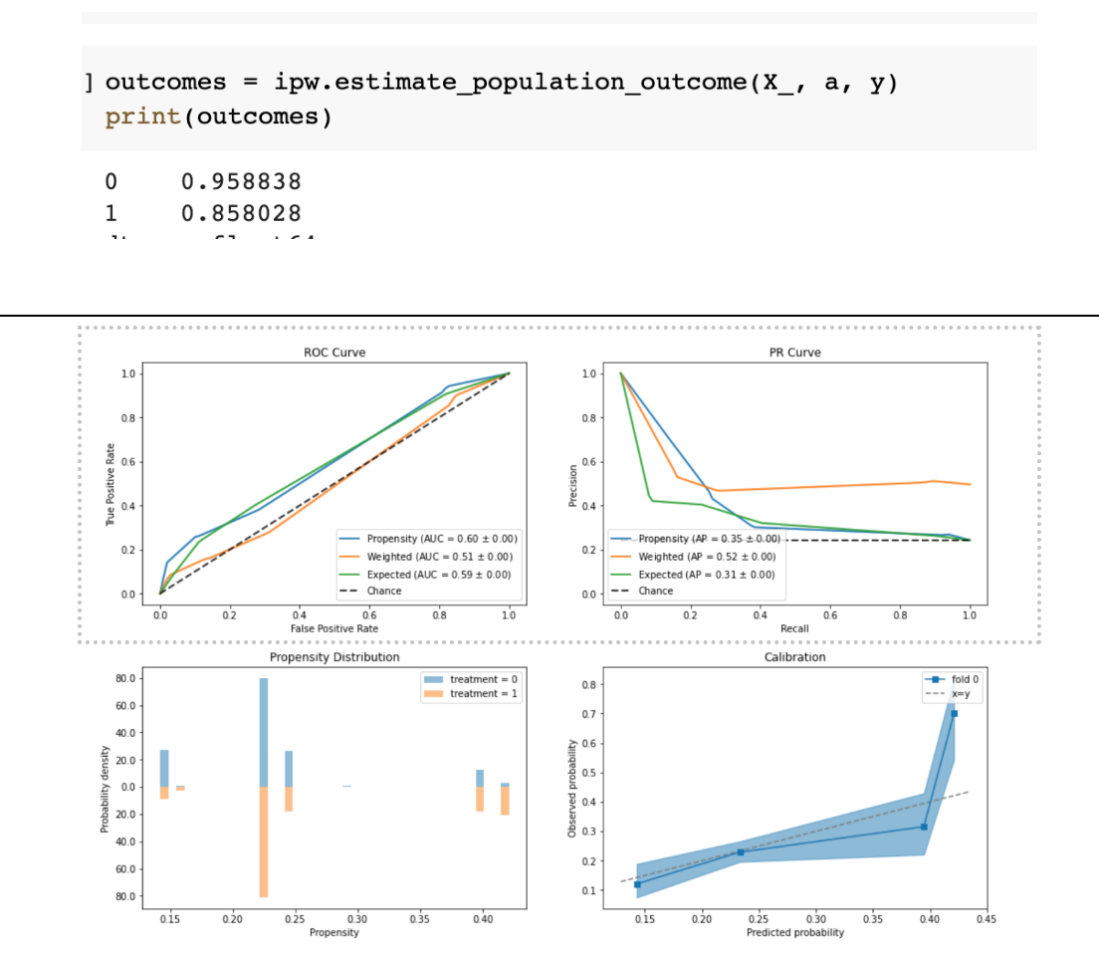
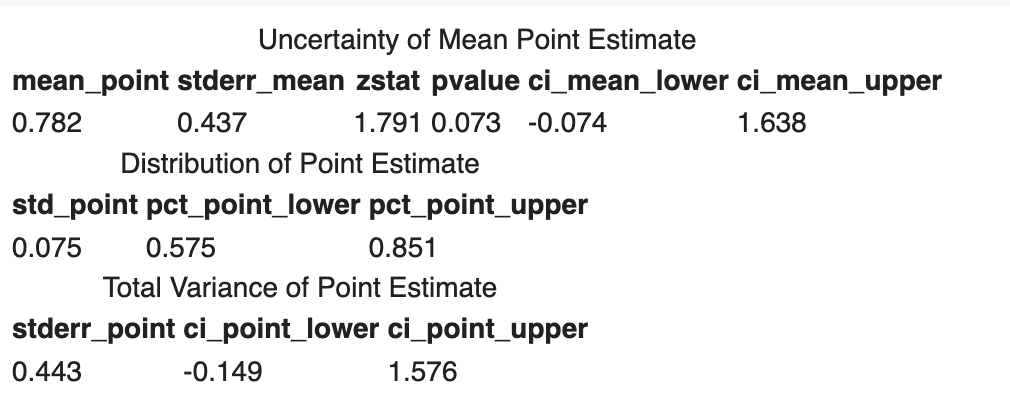
**Impact analysis of services/interventions on some outcomes using Causal Inference**





**Context and Problem:**

Imagine that you start extending a credit offer to certain customers to try and boost sales. You see that purchases among those customers have increased during the time you made the special offer, but can you conclude it’s because of the credit offer?

To find out, perhaps you try to compare the “before” and “after” states among customers who received the offer. Or perhaps you try comparing sales between customers receiving the offer and customers not receiving it (known as naïve attribution). Perhaps your stakeholders are now challenging you with these analyses, however, because of double counting or other confounding effects they can observe or already know about.

So how can you confidently conclude that differences in sales are due solely to the credit offer? What would have happened if the interventions were not done? Causal inference can help answer these questions.

For decades, industries and NGOs have used causal inference in the form of [randomized control trials](https://www.unicef-irc.org/KM/IE/impact_7.php) (RCTs). As powerful as these techniques are, however, they are also expensive. Fortunately, there is an alternative approach. Today, the wide availability of machine learning and large datasets makes causal inference — done with observational data or in quasi-experiments (a research design that tests for a cause-effect relationship without randomly assigned groups) — practical in a variety of industries without having to rely on the more expensive RCTs.

Causal inference can be helpful in several related situations. A basic one is analyzing the impact of Humanitarian services/interventions, which is inherently a “treatment effect ” problem — one in which the intervention (or “treatment,”) has a causal effect on an outcome variable (such as Food consumption score, using of latrines, etc...). The treatment effect can be measured at the population, treated group, subgroup, and individual levels.

Building further on this example, a treatment effect for each individual unit under study is defined as the difference between two potential outcomes: One outcome if the unit is exposed to the treatment and another outcome if the unit is exposed to the control. In practice, the individual treatment effect is unobservable because individual units can be either in the treatment or the control group, but not both. As a result, counterfactuals — what would have happened without an intervention — are the basis of causal inference, and the estimation of counterfactuals poses the biggest challenges but also provides the greatest opportunities in various scenarios and helps provide answer on the appropriate services/interventions in the case of a multi-treatment scenario.

Causal inference bridges the gap between prediction and decision-making. This is useful because prediction models alone are of no help when reasoning what might happen if we change a system or take an action, even for prediction models with extremely high accuracy. This is because going from prediction to a decision is not always straightforward. A typical supervised machine learning algorithm optimizes for the difference between actual values and their predicted values, but a decision based on such a prediction is not always one that maximizes the intended outcome when we take action. The very act of decision-making based on a prediction model may change the environment in ways that put us into untested territory, dampening the predictive power of the model. For example, suppose a data scientist builds an accurate model to predict who is most food insecure, and then a Humanitarian program initiates offers or campaigns for these households or individuals. Now the problem is divided into two separate problems, solved by different teams, resulting in local optimal solutions. The action taken on potential food insecure might change the environment, and therefore the optimal outcome. Instead of predicting who is likely to be food insecure as one problem and leaving campaign effectiveness to the humanitarian program as a second problem, we can use causal inference to predict the best action to improve food behavior for each individual or household.

**Database:** The data used come from RANOWASH Madagascar

**Repository: the documented notebook**

[WASH\_services (1).ipynb - Colaboratory (google.com)](https://colab.research.google.com/drive/1ryyk_JFllgKVSnGOgnwOH0g4U0wSNs5h#scrollTo=woRu-ztEjVi6)

**Process**:

**Estimation**

1. If you are assuming uncounfoundedness and then use estimation methods under such assumption, then identify clearly the Treatment (s) (T), the outcome (o), the covariates (X), the confounders (C) and use one of the sorresponding methods in the graph



1. If unconfoundedness can’t be verified, use Instrumental variables based methods in the graph above

**Model validation**

1. Audit the model and check for it’s validity using models from graph below

