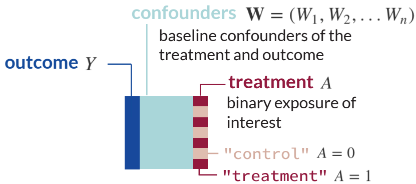
A VISUAL GUIDE TO G-COMPUTATION

G-computation is an **estimation technique for causal inference.** Here we use G-computation to estimate the **mean difference in an outcome**, adjusting for confounders. Under causal assumptions (not presented here) this is the **Average Treatment Effect (ATE)**, or the difference in outcomes if all observations had received treatment compared to if no observations had received treatment.

Data structure:

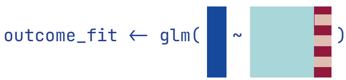


Estimand:

$$ATE = E[E[Y|A=1,\mathbf{W}] - E[Y|A=0,\mathbf{W}]]$$

Algorithm:

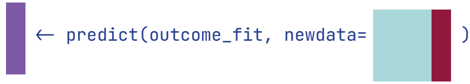
First, estimate the expected outcome for all observations, using confounders and treatment status as predictors.



$$E[Y|A,\mathbf{W}]$$

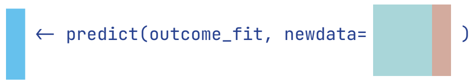
Then, use that model fit to predict every observation's outcome using:

1. Every treatment status set to "treatment"



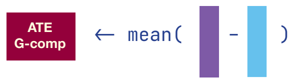
$$\hat{E}[Y|A=1,\mathbf{W}]$$

2. Every treatment status set to "control"



$$\hat{E}[Y|A=0,\mathbf{W}]$$

Finally, calculate the ATE by taking the average difference between the expected outcomes.



$$\hat{ATE} = \hat{E}[\hat{E}\ [Y|A=1,\mathbf{W}] - \hat{E}\ [Y|A=0,\mathbf{W}]]$$

The G-computation estimator relies on correct specification of the generalized linear model (GLM) for the final estimate to be consistent.

Although flexible machine learning models can be used for the outcome regression instead of GLMs, there is generally not theory to support a normal distribution of the estimator for computing confidence intervals, p-values, etc.

kathoffman/causal-inference-visual-guides
@kat_hoffman_