

Cash Transfers and Cognitive Development: treatment effect

Estimar los efectos de un tratamiento para cada individuo mediante modelos de Machine Learning para optimizar el éxito del tratamiento

Borrego Burón, José Manuel

Magre i Pont, Jaume

Roura i Cubí, Antonino

Sanginés-Uriarte Dooly, Joel



UNIVERSITAT DE
BARCELONA



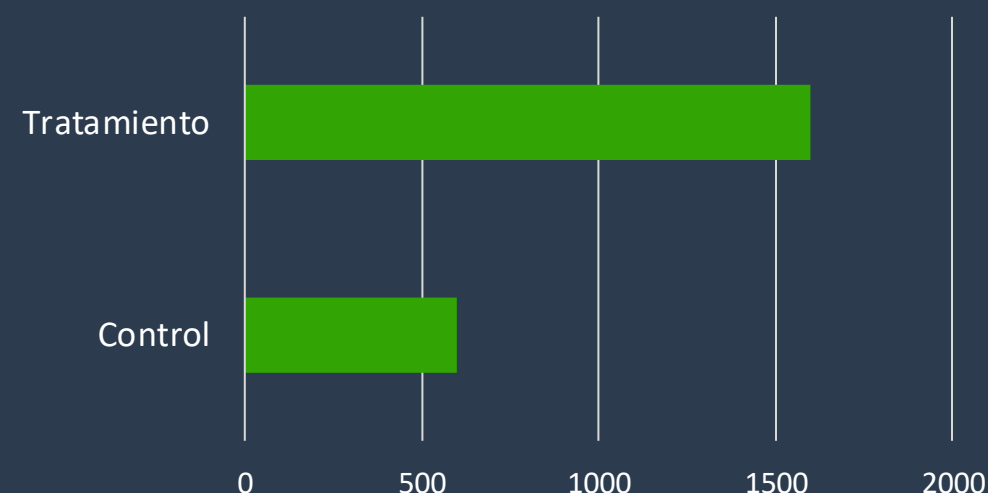
PROGRAMA *ATENCIÓN A CRISIS*



- Promoción del desarrollo cognitivo en la primera infancia.
- Ayudas económicas a hogares pobres en 2005-2006 en Nicaragua.
- Enmarcado dentro de un estudio de su impacto en las familias: datos económicos sociopersonales, de hábitos de vida y de desarrollo psicomotriz de los niños.

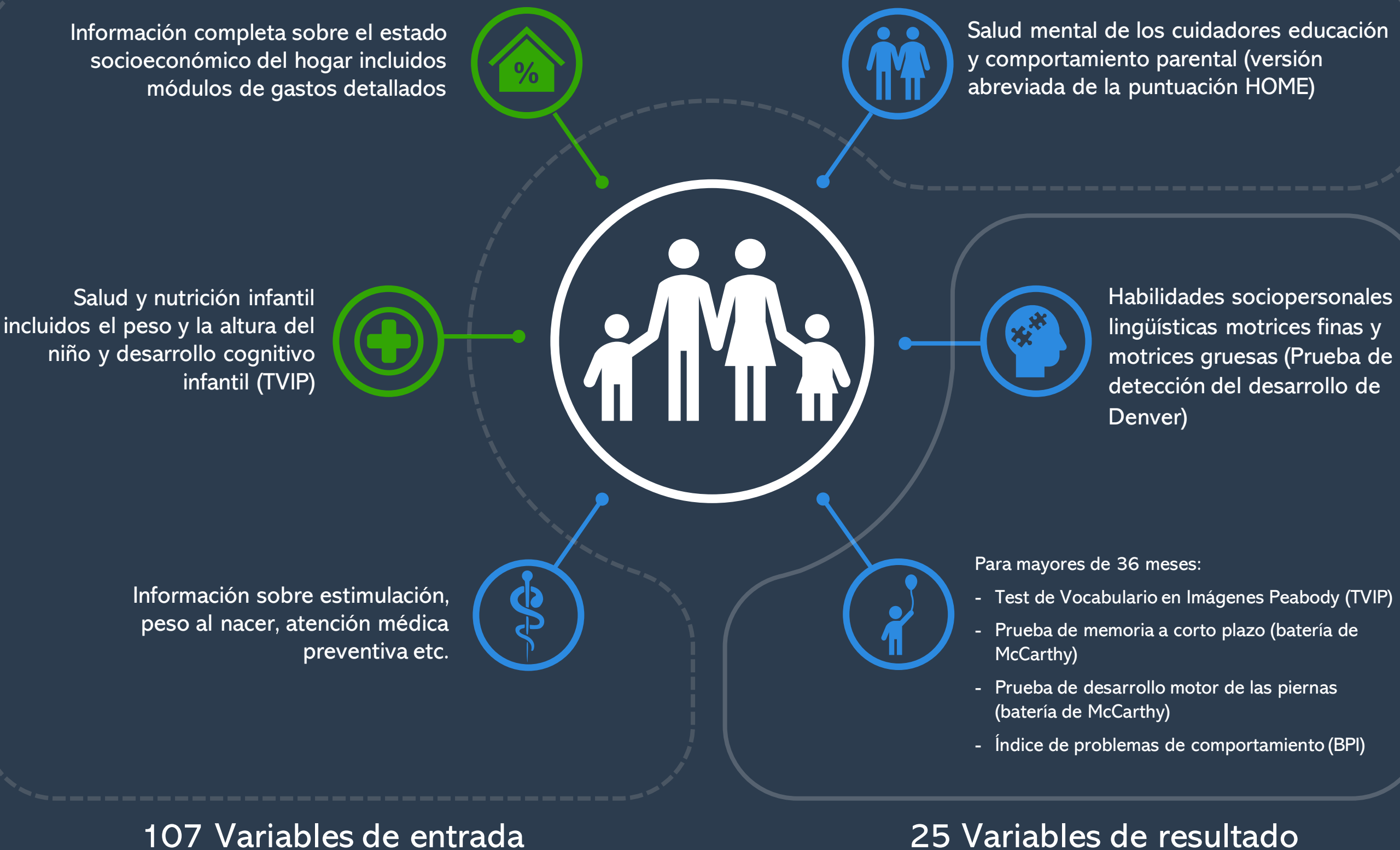


FAMILIAS INCLUÍDAS



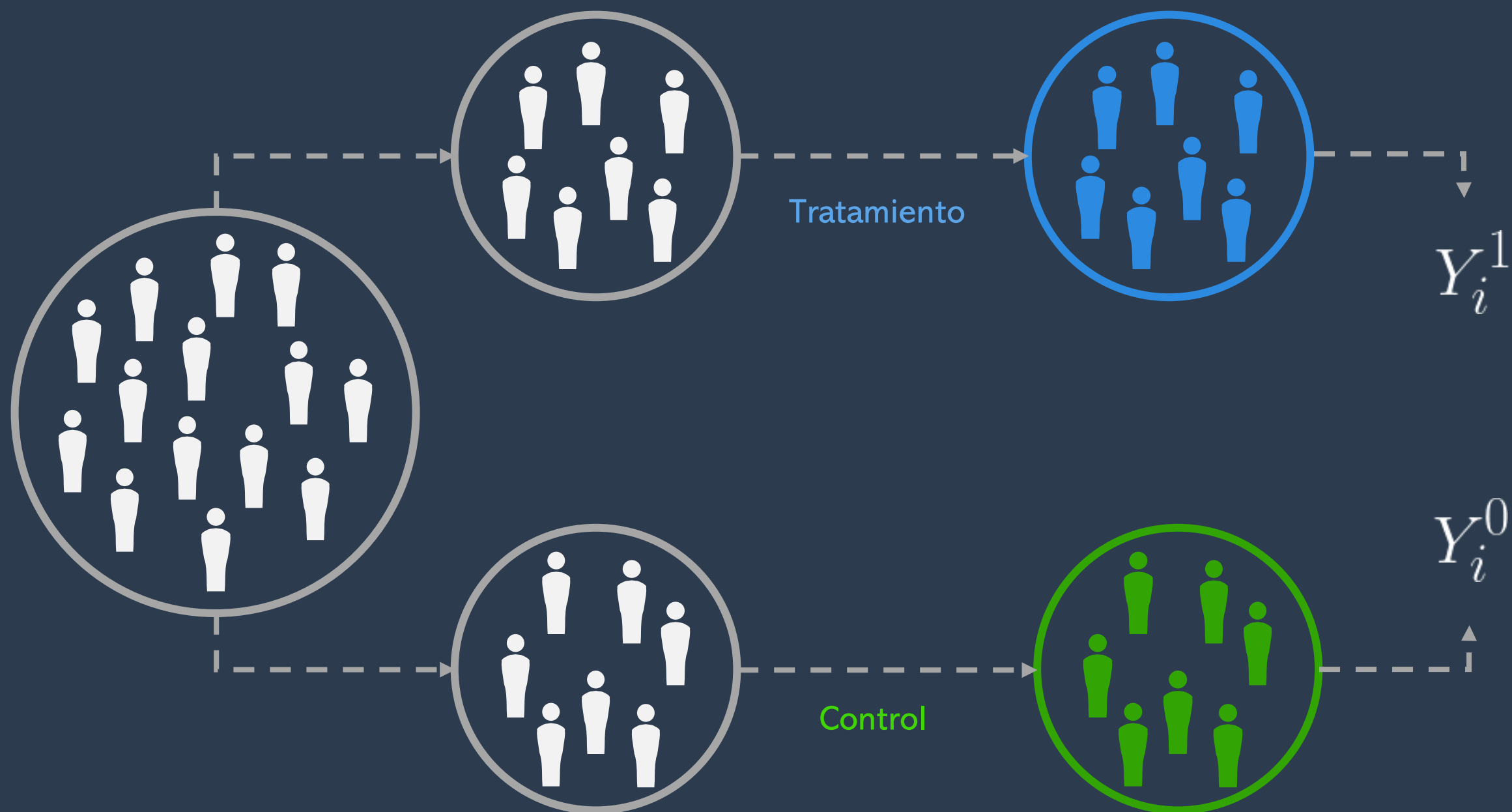
TEMPORIZACIÓN







EVALUACIÓN DEL EFECTO: *RANDOMIZED CONTROLLED TRIAL*



$$ATE_i = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$$



EVALUACIÓN DEL EFECTO



Objetivo

Identificar la heterogeneidad en el efecto al tratamiento para optimizar su respuesta



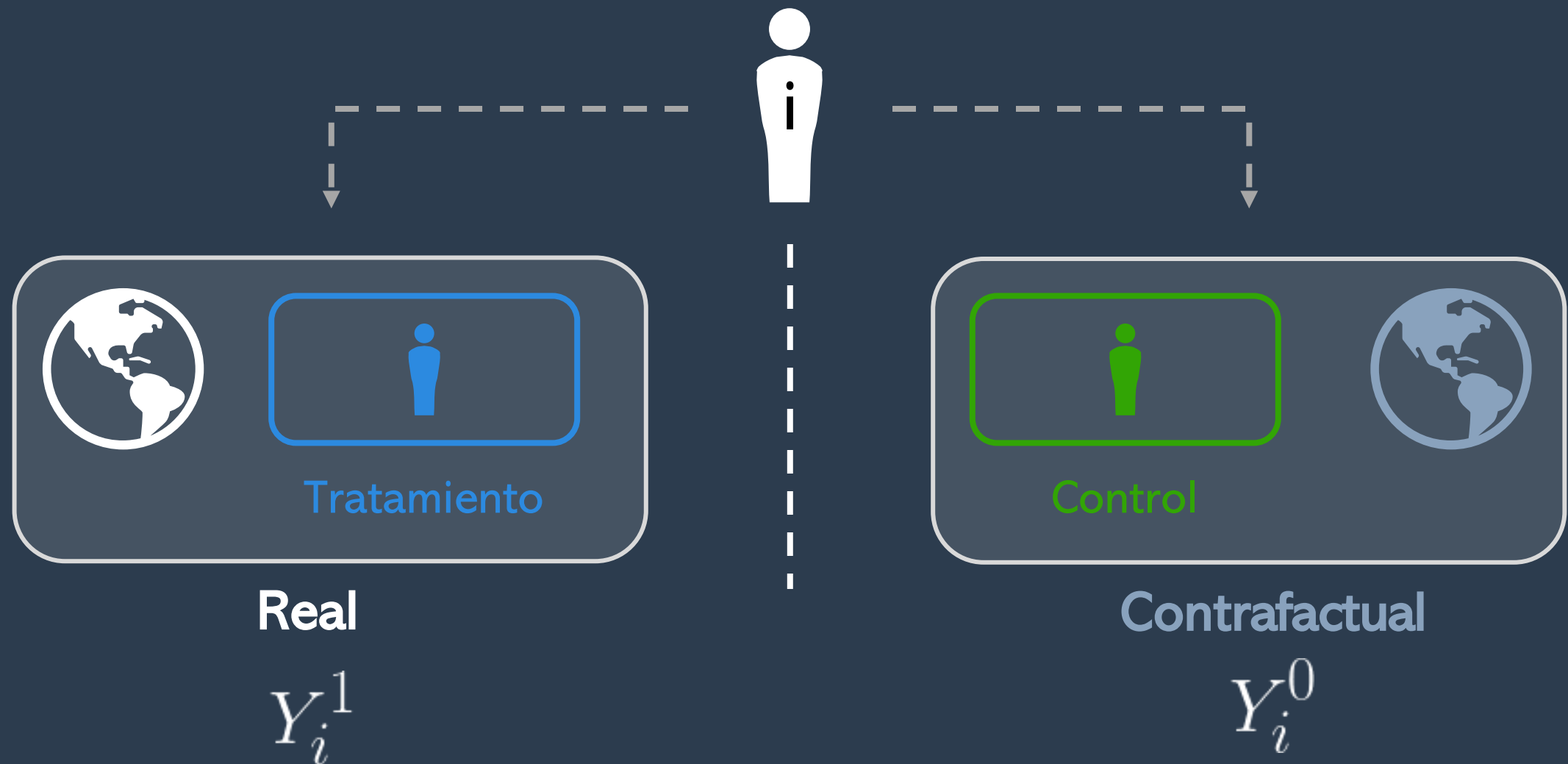
Tratamiento

Aplicado en los subgrupos que más se benefician



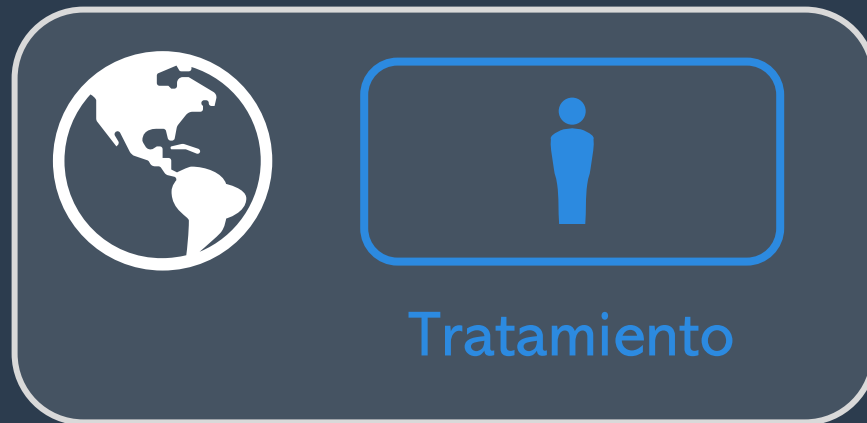
EVALUACIÓN DEL EFECTO INDIVIDUAL

$$\text{ITE}_i = Y_i^1 - Y_i^0$$

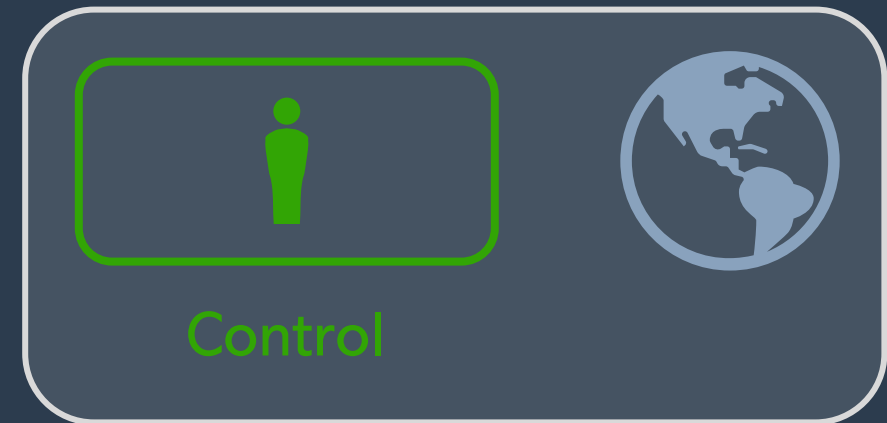




EVALUACIÓN DEL EFECTO INDIVIDUAL



Real



Contrafactual



$$ITE_i = Y_i^1 - Y_i^0$$



METODOLOGÍA



	Y	T	Var1	Var2
0	8.24	1	0.25	7
1	4.78	0	0.65	6
...



Generar *dataset* contrafactual y predecir el *outcome*

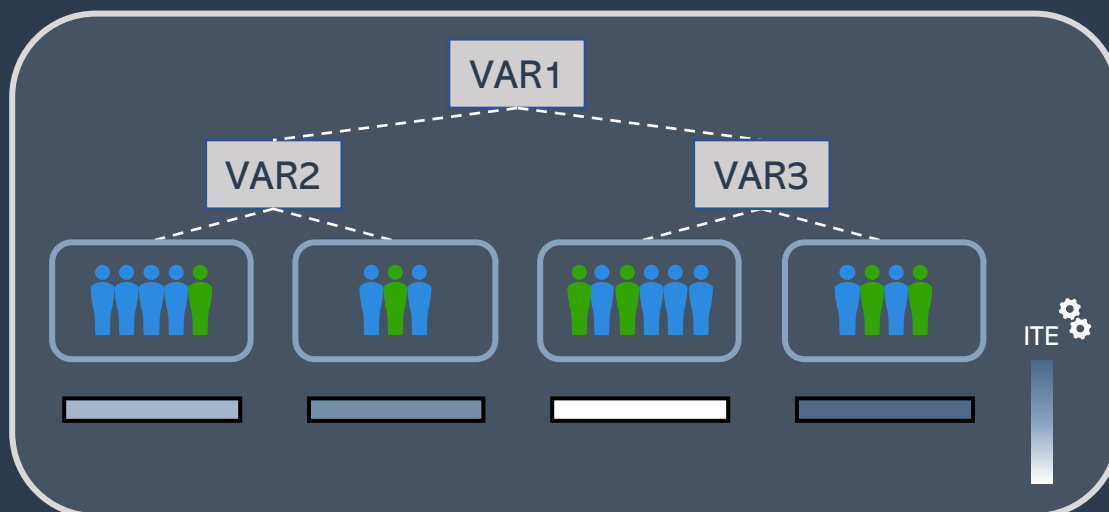
	Y	T	Var1	Var2
0	<u>7.89</u>	0	0.25	7
1	<u>4.91</u>	1	0.65	6
...



↓

$$ITE_i = Y_i^1 - Y_i^0$$

← -





LIMPIEZA DE DATOS

Datos brutos

	Var1	Var2
0	-	-
1	-	-
...

4511 x 198 var.

Selección inicial de variables

	Var1	Var2
0	-	-
1	-	-
...

4511 x 51 var.

Omisión registros
NaN > 80% var.

Omisión registros
age_transfer < -11

Codificación de
variables categóricas

s1male_head_05
s3atollet_hh_05
s3awater_access_hh_05
s3aelectric_hh_05
s4p6_vitamina_i_05
s4p7_parasite_i_05
s11ownland_hh_05

Omisión registros
por valores atípicos

cons_food_pc_05 com_control_05
cons_tot_pc_05 pranimalprot_f_05
s3ap24_hitime_h_05 prfruitveg_f_05
bweight

Imputación según
unidad familiar

prfruitveg_f_05 cons_food_pc_05
prstap_f_05 ed_mom
pranimalprot_f_05 yrsedfath
s2mother_inhs_05 vitamiron_06
propfood_05

Imputación *NaN* por
media/moda

Normalización

s1age_head_05 age_transfer
s3ap23_stime_h_05 bweight
s3ap24_hitime_h_05 ed_mom
s3ap25_hitime_h_05 com_haz_05
cons_food_pc_05 com_waz_05
cons_tot_pc_05 com_tvip_05
yrsedfath com_control_05

Dataset final

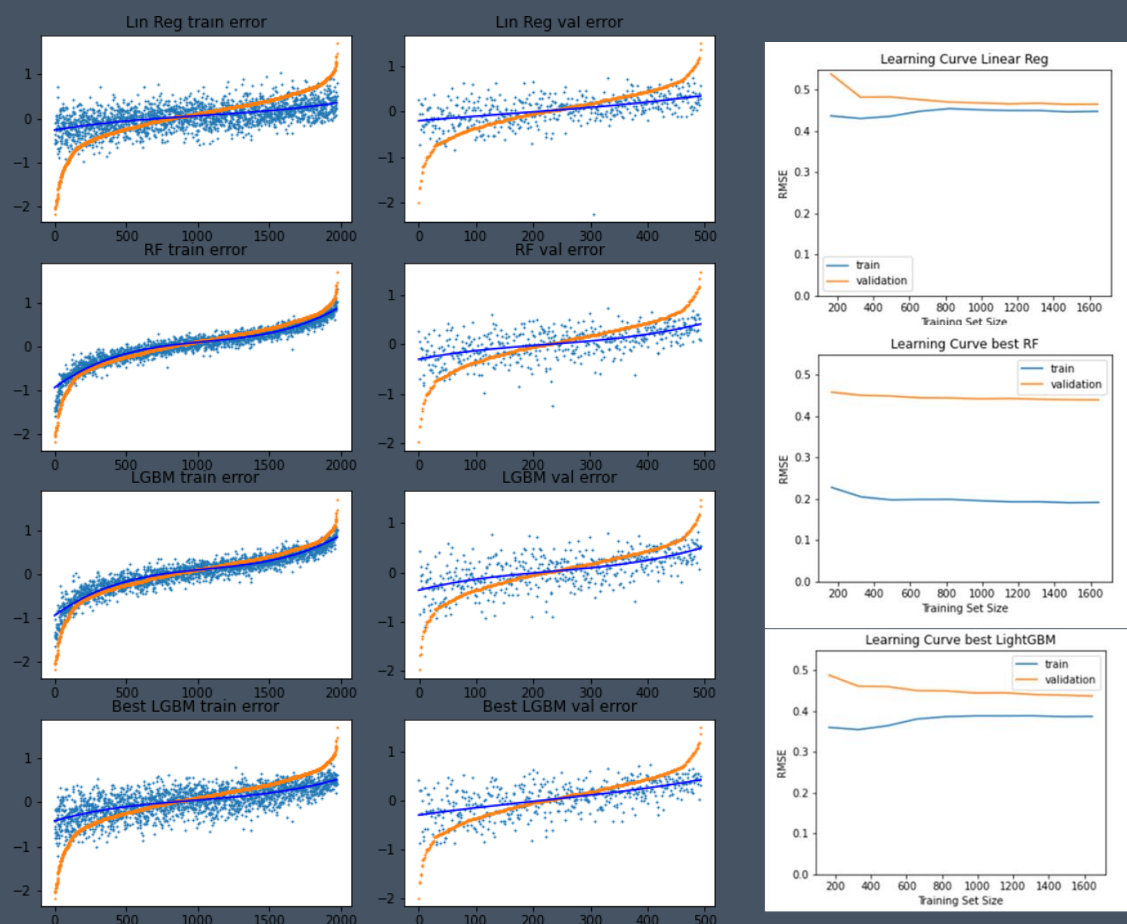
	Var1	Var2
0	-	-
1	-	-
...

3085 x 47 var.



PREDICCIÓN

Selección de modelo



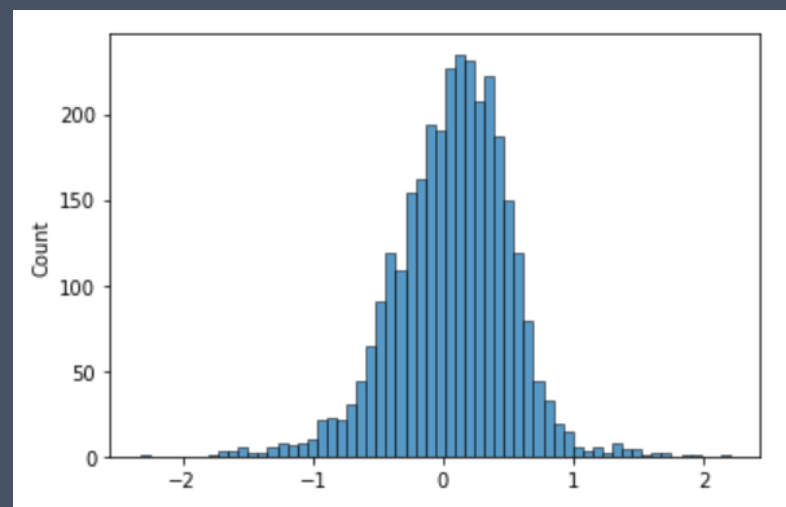
LightGBM

RMSE: 0.431

R^2 : 32.24%

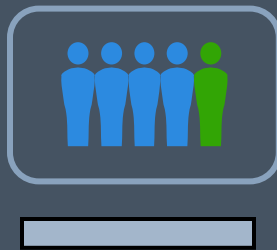
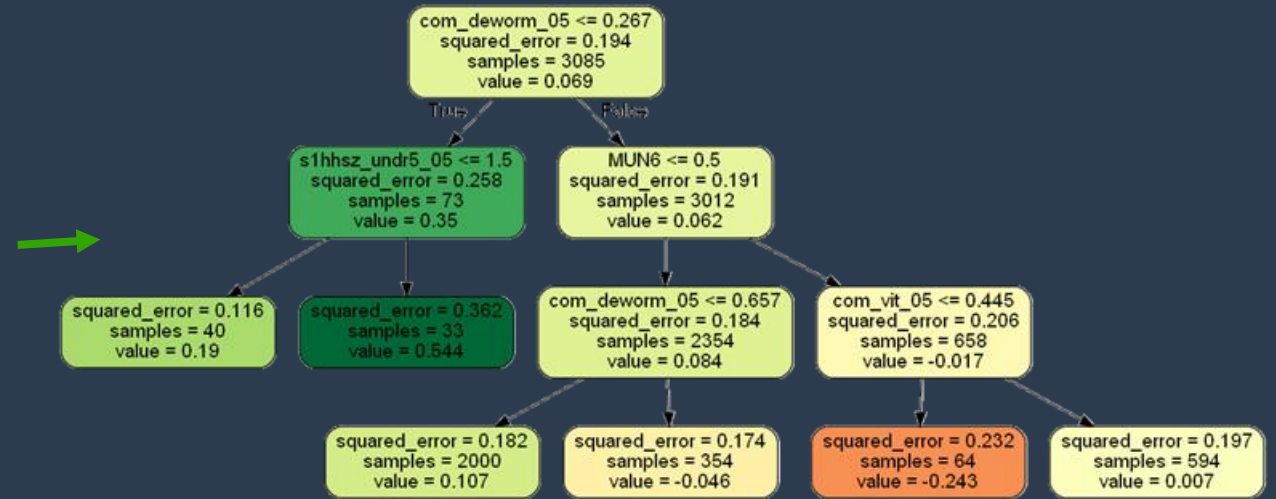
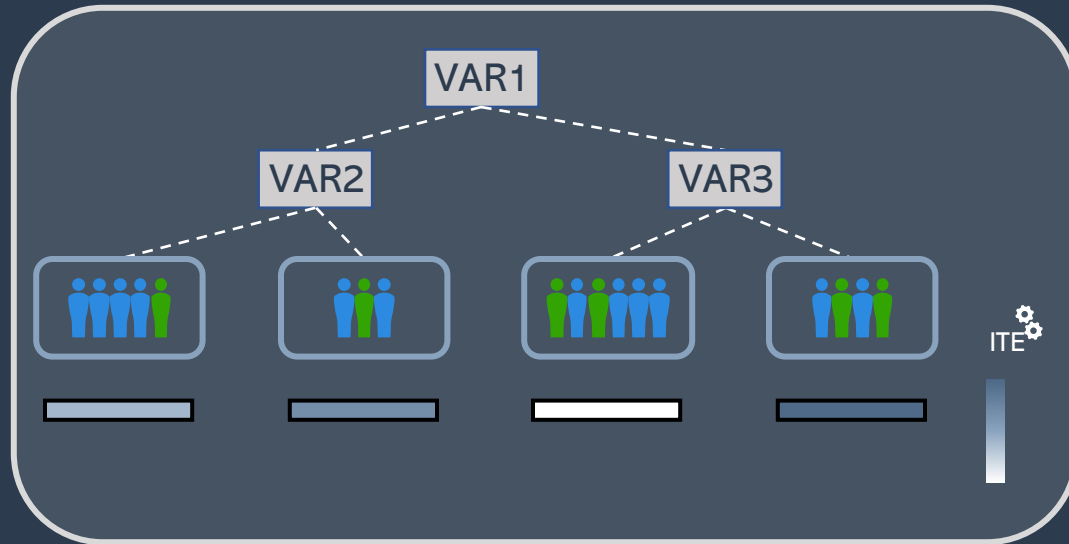
LightGBM.predict(X + 1-tr) = Contrafactual

$T_i = 0 \quad ITE_i = \tilde{Y}_1 - Y_0 \quad (\text{contrafactual-obs})$
 $T_i = 1 \quad ITE_i = Y_1 - \tilde{Y}_0 \quad (\text{obs-contrafactual})$

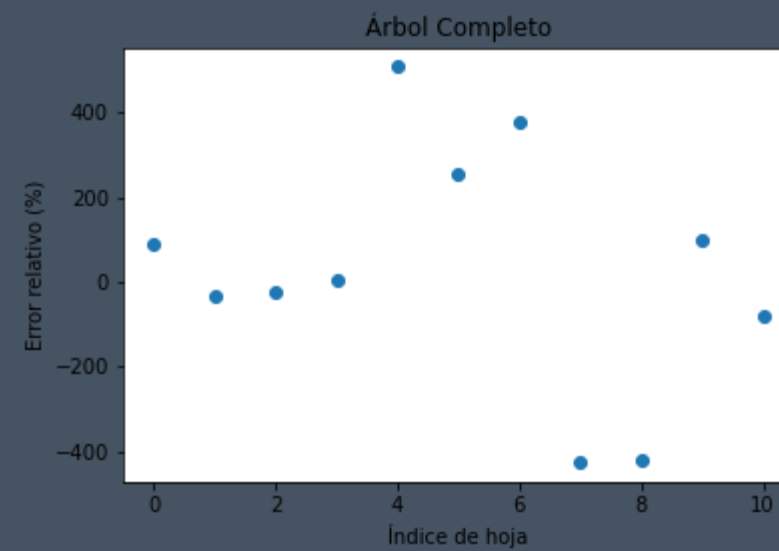




ANÁLISIS DE SUBGRUPOS

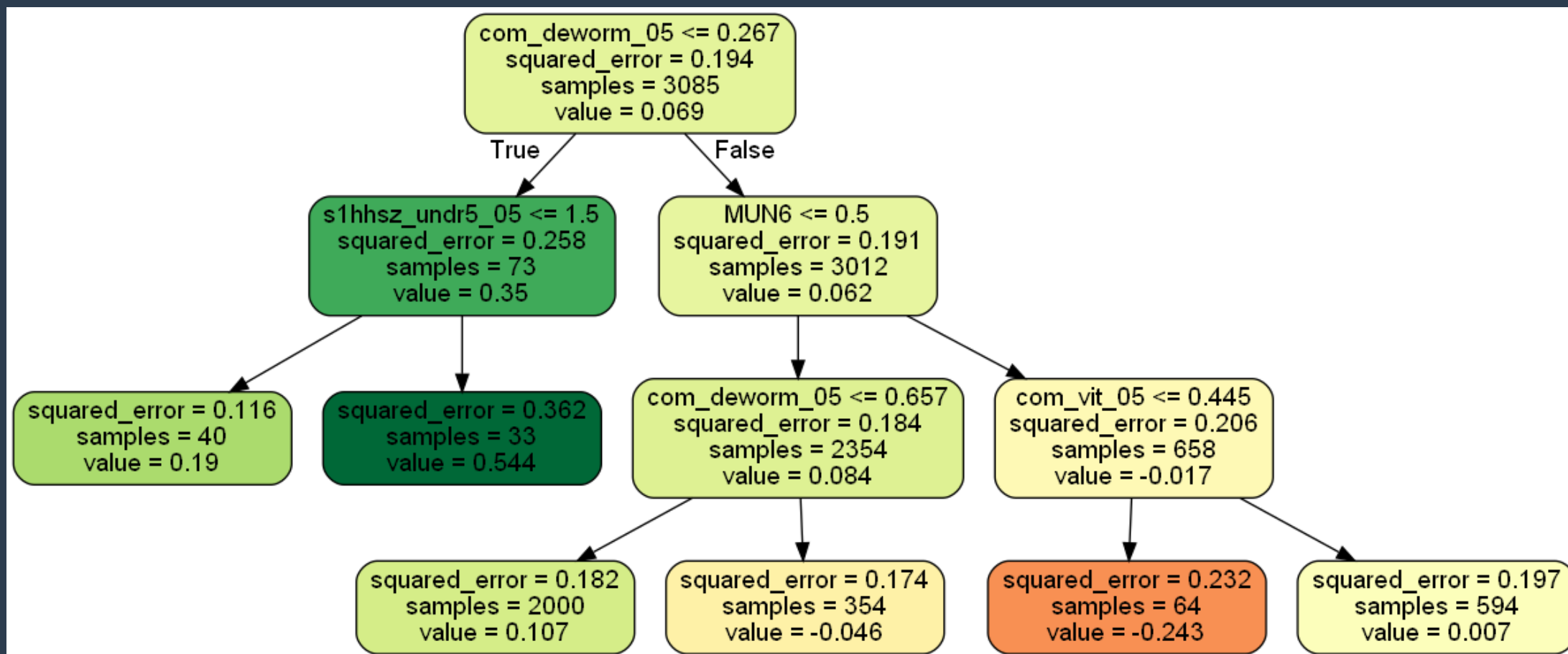


$$\overline{\text{ITE}} \stackrel{?}{=} \text{ATE}$$





INTERPRETACIÓN DE LOS *DECISION TREES*: COMUNITARIO

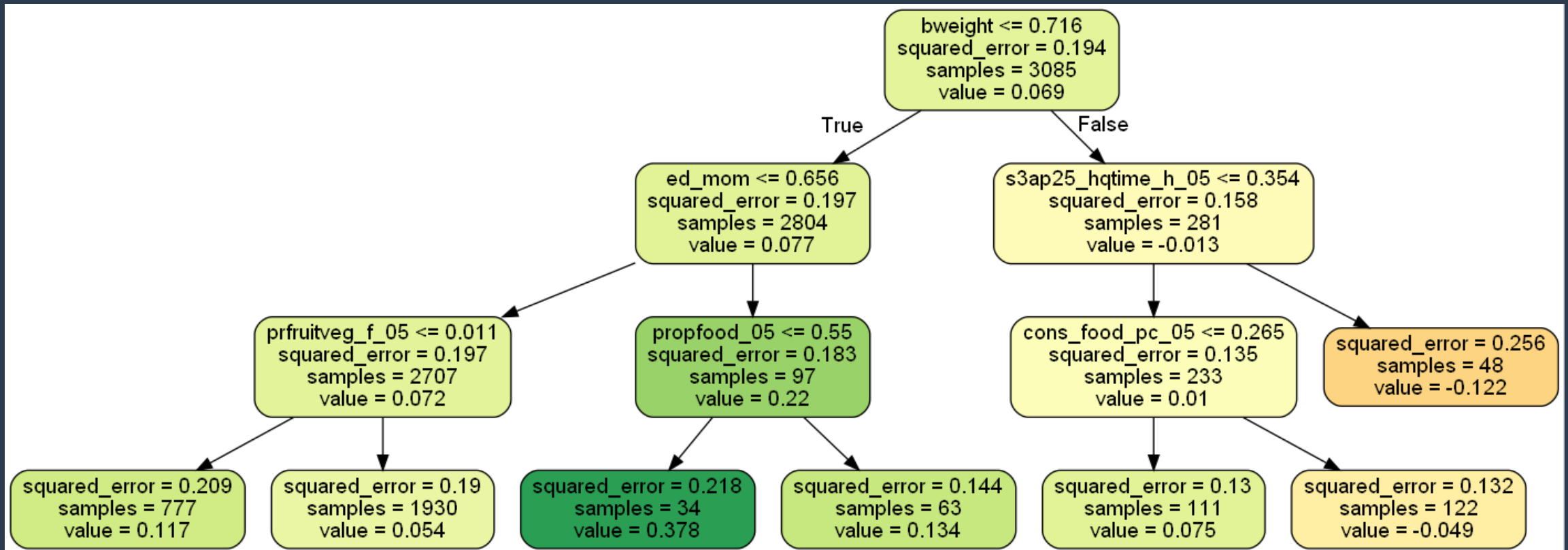


VARIABLES COMUNITARIAS IMPORTANTES

- Com_deworm_05
- MUN6



INTERPRETACIÓN DE LOS *DECISION TREES*: INDIVIDUAL

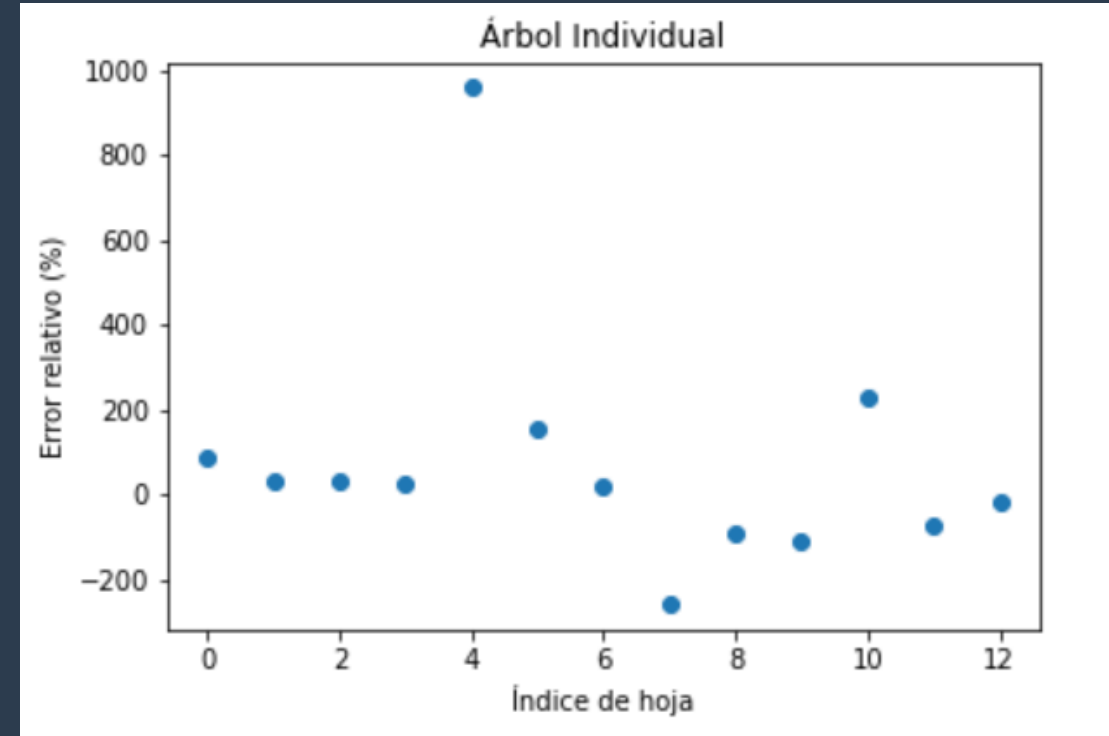
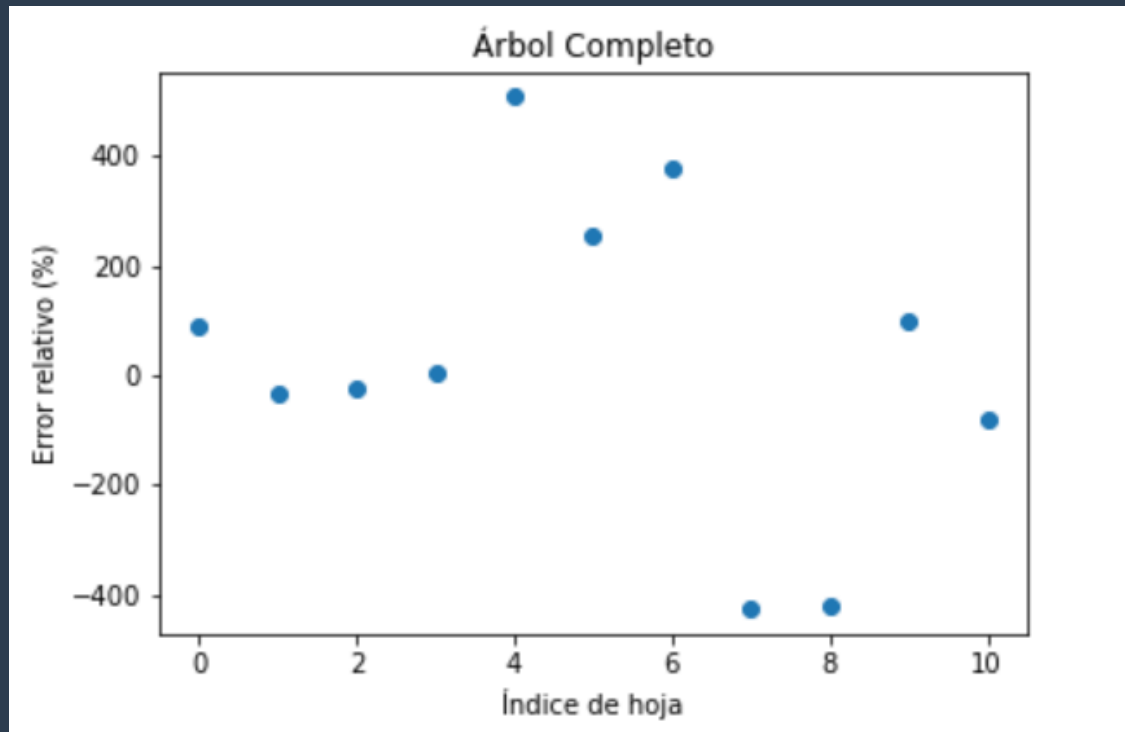


VARIABLES INDIVIDUALES IMPORTANTES

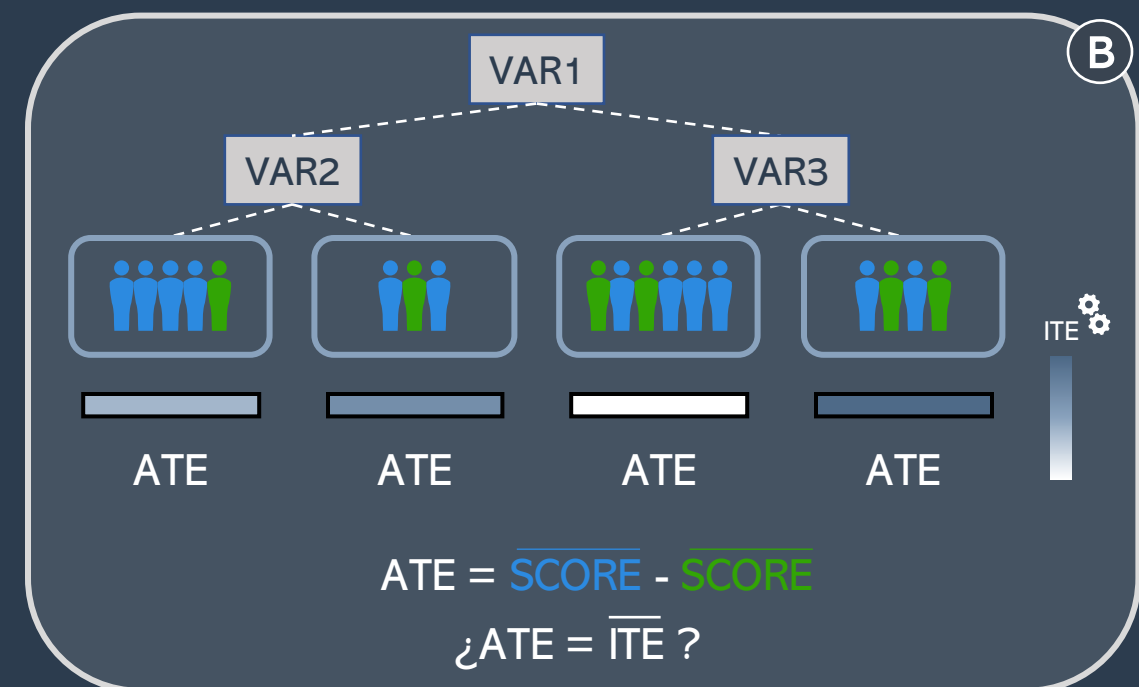
- bweight
- ed_mom
- s3ap25_hqtime_h_05



VALIDEZ DEL MÉTODO: COMPARACIÓN DE ITEs Y ATEs



- Error grande, signos parecidos. Subestimación.
- No invalida, pero tampoco lo respalda



LIMITACIONES DEL ESTUDIO

- Estudio base causal
- Variables con poco poder predictivo
- Número de observaciones bajo
- z_all_06

BIBLIOGRAFÍA

- [Athey and Imbens, 2015] Athey, S. and Imbens, G. (2015). Recursive partitioning for heterogeneous causal effects.
- [Case and Paxson, 2006] Case, A. and Paxson, C. (2006). Stature and status: Height, ability, and labor market outcomes.
- [Crépon et al., 2014] Crépon, B., Devoto, F., Duflo, E., and Pariente, W. (2014). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco.
- [Currie and Thomas, 1999] Currie, J. and Thomas, D. (1999). Early test scores, socioeconomic status and future outcomes.
- [Friedberg et al., 2018] Friedberg, R., Tibshirani, J., Athey, S., and Wager, S. (2018). Local linear forests.
- [Jacob, 2021] Jacob, D. (2021). Cate meets ml - conditional average treatment effect and machine learning. SSRN Electronic Journal.
- [Kwak and Kim, 2017] Kwak, S. K. and Kim, J. H. (2017). Statistical data preparation: Management of missing values and outliers.
- [Lamont et al., 2018] Lamont, A., Lyons, M. D., Jaki, T., Stuart, E., Feaster, D. J., Tharmaratnam, K., Oberski, D., Ishwaran, H., Wilson, D. K., and Horn, M. L. V. (2018). Identification of predicted individual treatment effects in randomized clinical trials. *Statistical Methods in Medical Research*, 27:142–157.
- [Macours et al., 2012] Macours, K., Schady, N., and Vakis, R. (2012). Cash transfers, behavioral changes, and cognitive development in early childhood: Evidence from a randomized experiment.
- [Yao, 2021] Yao, F. (2021). Machine learning with limited data.
- [Zhou et al., 2017] Zhou, D. P., Balandat, M., and Tomlin, C. J. (2017). Estimating heterogeneous treatment effects in residential demand response.