Question a) Compare the means of baseline characteristics (X's) of eligible vs. ineligible children in treatment villages, to evaluate the balance of the groups. Interpret the results.

The result gives us the balance of pre observed characteristics limited to observations within treatment groups. Within the treatment groups, the treatment was not given in random assignment therefore, we don't expect all the baseline covariates to be balanced. In this comparison, there are few covariates that reject the null hypothesis and show significant levels above 5%. Those are [gender hohsex dist_sec hidalgo mich puebla queret sanluis].

Question b) Construct estimates of the propensity score of treatment (eligibility) among children in treated villages. Then, (i) evaluate the balance of observations in the treated and control groups by blocks of the estimated propensity score, and (ii)construct histograms or empirical estimates of the pdf of the estimated propensity score for children in eligible vs. ineligible households in the treated villages.

*For this question I constructed 3 different sets of covariates to construct 3 different estimated propensity scores. Out of 3 models, we can pick the best one by visually inspecting the overlap in histogram and also checking the balance between estimated propensity scores in blocks. However, it is worthy to mention that none of these models satisfy the balancing property and therefore none of them are "ideal". By visually inspecting the pdf and histogram of estimated pscores, a set of covariates formed by their significant p-value (est_pscore_psig) have the best overlap. The intuition would indicate that the best overlapping model would also have the most balanced blocks. This is proven to be true as well. We regress our pscore by blockid and see psig model has the lowest tvalue although all 3 models have t value that rejects the null hypothesis.

Question C)Construct estimates of the ATT using inverse propensity score weights based on the best set of covariates determined in step (b).Compare these estimates to the non-experimental estimates generated from the simple differences of eligible and ineligible children (in Assignment 1).

Using inverse propensity score weights, we estimated the average treatment effect with 3 models constructed. Comparing the estimated average treatment effect from non-experimental data, we see that our "best" model from part b is actually the closest to it. This is contradicting our intuition as we know the average treatment effect from non experimental models is a "bad" estimate and our best model of propensity score matching should be better than this. The other 2 model's estimate does follow general intuition but we know the model has bad overlap and high tvalue. The reason why my best model gives out a bad estimate could be that it failed to have balanced pscore in blocks and therefore failed to reduce the downward bias or even worsened it.