Blocking and Clustering

1. Blocking: when, in a randomized experiment, you take units and make sure that if one given unit is assigned to treatment, then another unit is assigned to control. Used to reduce the size of the differences that can arise by chance. Increases statistical power given an experiment with same sample and effect size.
   1. If some variables are related to the outcome, restrict ourselves to randomizations that keep treatment and control similar. Randomize in blocks.
      1. Blocking reduces standard error. Standard error for 2 blocks:
      2. Unless the probability of assignment to the treatment group is identical for every block, pooling observations across blocks will produce biased estimates of the overall ATE.
   2. Benefits of blocking
      1. Allows for more precision (efficiency) by not conducting randomization where covariates are very imbalanced.
      2. Estimator will produce an estimate that is closer to the true causal effect.
2. Clustering: when multiple units always have to be in treatment or control together at the same time. Increases standard error.
   1. Cluster: the level where the treatment is assigned.
      1. Outcomes can be observed at more fine-grained levels.
      2. There may be a difference in cluster-average outcomes.
   2. Perils of ignoring clustering
      1. Not capturing the right spread.
   3. Larger differences between treatment and control outcomes will happen by chance. To account for this, when conducting randomization inference, match the experiment’s assignment process
      1. If assignment happens at random, randomization inference needs to as well
      2. If assignment happens with clusters, randomization inference needs to as well
      3. If assignment happens with blocks, randomization inference needs to as well
   4. Power is relatively worse when average between cluster differences are larger, within cluster differences are smaller, and the number of clusters is small. Converse is true too.