Causal Inference

Problem Set 1.

1. In children at risk for malaria, research is conducted in stimulating the immune system with micronutrients supplementation, an important one being vitamin A (e.g., Shankar et al., Lancet 1999, 354:203-9).

   Table 1 gives measurements of the malaria parasite Plasmodium falciparum (Pf), in $\log_e$ (counts / microL), for 6 children taking vitamin A supplementation and 6 children taking a control (placebo) treatment.

   Table 1: Parasite load [$\log_e$ (count/microL)].

Placebo	Vitamin A
8.62	0.06
1.48	1.72
8.93	2.19
9.57	7.32
2.65	7.53
7.30	7.62

   We wish to test Fisher's (sharp) null hypothesis of no treatment effect on these 12 children. Take the test statistic to be the difference in sample means. Assume that the assignment in Table 1 was a completely randomized experiment.

   (a) Calculate the exact p–value (two-tailed). [Hint: assuming the null hypothesis, write a program to find the 924 possible values of the test statistic under the assignment mechanism.]

   (b) Report the p–value obtained by drawing $1,000$ times from the distribution of the statistic under the null hypothesis and comparing the draws with the observed statistic.

   (c) Report the p–value by using a t-test.

   (d) In what aspects of the Rubin-Causal-Model (e.g., assignment mechanism, potential outcomes) is (b) an approximation to (a), and (c) and approximation to (a) ?

2. Now assume that prior to the experiment, researchers used a characteristic X (e.g., baseline severity measures) of the children to match them in 6 pairs (the 6 pairs being the 6 rows of Table 1), and then they assigned treatment at random within each pair. Answer 1(a)-1(d) for the pairwise assignment (For (c) use a paired t-test).

3. In analyzing the same data of Table 1, the p-values obtained in question 2 are higher than those in question 1. However, there are two main reasons why this comparison does not necessarily imply that pairwise randomization is a worse design generally. Give at least one of the reasons.

4. By writing the test statistic (difference in sample means) as an explicit function of the potential outcomes and the assignment indicators, show that, under complete randomization, the statistic is unbiased for the average treatment effect. Do the same for the pairwise randomization.

5. **Extra Credit.** Researchers are interested in obtaining some more information about the treatment effect. By assuming only that

   (a) the assignment in Table 1 was by complete randomization, and

   (b) $Y_i(1) = Y_i(0) + k_0$ for all i (additivity of treatment effect),

   find a 95% confidence interval for k_0, the treatment effect. [Hint: look at the 924 values in part 1(a).]

6. **Extra Credit.** In a study with complete randomization, is it possible for the data to provide evidence that the treatment effect is not additive ? If not, explain. If yes, describe briefly how you would use a statistical method to look for evidence against additivity, for example, in the context of the data in Table 1.