Stat 490, Spring 2025 Due: Tuesday March 18, 11:59 PM

Homework 3

Tirthankar Dasgupta

Collaboration policy: you are free to discuss the problems with others, though it is strongly recommended that you try the problems on your own first. Copying is not allowed, and write-ups must be your own explanations in your own words.

Computation: you can use appropriate R codes and functions posted on Canvas or write your own codes or use any software of your choice for computational purposes. You can also use any output of the R codes from the notes (posted February 28 onwards) directly as an input to your solutions and perform the rest of the computations manually.

Problem 1

The following table shows the post treatment LDL cholesterol for 12 participants with similar pre-treatment levels of cholesterol. The participants were grouped into two blocks by age and a balanced randomized block design (i.e., balanced completely randomized design within each block) was used to assign two treatments (0: existing and 1:new) to the participants.

Block (Age)	Unit (i)	$Y_i(0)$	$Y_i(1)$	W_i	y_i
$1(\leq 45)$	1	?	97	1	97
	2	99	?	0	99
	3	?	97	1	97
	4	97	?	0	97
	5	102	?	0	102
	6	?	99	1	99
Block mean		$\overline{y}_1(0) = 99.3$	$\overline{y}_1(1) = 97.7$		
Block variance		$s^2_{10} = 6.3$	$s^2_{11} = 1.3$		
$2\ (> 45)$	7	?	97	1	97
	8	?	98	1	99
	9	100	?	0	100
	10	99	?	0	99
	11	97	?	0	97
	12	?	101	1	101
Block mean		$\overline{y}_2(0) = 98.7$	$\overline{y}_2(1) = 98.7$		
Block variance		$s^2_{20} = 2.3$	$s^2_{21} = 4.3$		

(a) Using a Fisher randomization test with the test statistic $T_{\text{avg-uneqvar}}$, assess whether there is any differential treatment effect on any unit in block 1 against a two-sided alternative. To get full credit, show (i) how you calculated the observed value of the test statistic (ii) how you calculated the p-value and (iii) your conclusions in plain English.

(b) Using a Fisher randomization test, assess the plausibility of the sharp null hypothesis of zero treatment effect for each of the 12 participants, using a test statistic of your choice. To get full credit, explain (i) your choice of the test statistic and show (ii) how you calculated the observed value of the test statistic (ii) how you calculated the p-value and (iii) your conclusions in plain English.

(c) Use a Neymanian approach to obtain asymptotic 95% confidence intervals for the following estimands: (i) The average treatment effect within block 2 (ii) The difference of the average treatment effects within block 1 and block 2. To get full credit, show your calculations (manual or codes).

Problem 2

A scientist wants to test the effect of a diet A (treatment) versus a more traditional diet B (control). He has 20 animals available for the experiment. The animals differ with respect to age, sex and other characteristics.

(i) What are the drawbacks of using a completely randomized treatment assignment?

(ii) Suppose the experimenter forms ten pairs of animals, so that within each pair the animals appear "similar to each other" to the experimenter, although the experimenter does not share the precise information that lead to the pairing. The two animals within each pair are labeled 1 or 2. Within each pair, an animal gets either diet A or diet B and the other animal gets the other diet; allocation is decided by a flip of a fair coin. The response is the percentage decrease in the animals weight after two months of treatment. The results are shown in Table 1. What is the total number of randomizations possible in this experiment?

Table 1: Observed results of the diet experiment

Group	1	2	3	4	5
A	13.2 (1)	8.2 (2)	10.9 (1)	14.3 (1)	10.7 (2)
B	14.0 (2)	8.8 (1)	11.2 (2)	14.2 (2)	11.8 (1)
Group	6	7	8	9	10
A	6.6 (2)	9.5 (1)	10.8 (2)	8.8 (2)	13.3 (1)
B	6.4 (1)	9.8 (2)	11.3 (1)	9.3 (1)	13.6 (2)

(iii) Consider a randomized block design with five blocks, each consisting of 4 animals, two of which randomly receive diet A and the other two receive diet B. Give an example of a randomization that is possible in this design but not with this matched-pair design.

(iv) Express the data in the form of a table that shows the observed and the missing potential outcomes and also the observed assignment vector $\boldsymbol{w}$.

(v) Perform a randomization test to assess Fisher's sharp null hypothesis of no treatment effect. Clearly state the test statistic you use, report the p-value and state your conclusions. Include a histogram that shows your results.

(vi) Obtain an approximate 95% confidence interval for the average causal effect using Neyman's method.

Problem 3

Consider a matched-pair experiment with $N = 2B$ units, where the N units are paired into B blocks of size 2 each. Define $\overline{\tau}_b$ as the causal effect of the treatment in block b (i.e., the difference between the average potential outcomes under treatments 1 and 0 within block b) and $\overline{\tau}$ as the overall average causal effect over all N units.

(a) Let $\widehat{\overline{\tau}}_b$ denote the estimator based on the observed outcomes for the two individuals in block b. Prove that $\widehat{\overline{\tau}}_b$ is an unbiased estimator of $\overline{\tau}_b$.

(b) What is an unbiased estimator of $\overline{\tau}$ in terms of $\widehat{\overline{\tau}}_b$ for $b = 1, \ldots, B$? Why is it unbiased (provide a proof).