**Stat 490, Spring 2025** Due: Tuesday February 4, 11:59 PM

Homework 1

*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.

**Problem 1**

The owner of a Chinese restaurant wants to improve the taste of fried rice they serve in their lunch boxes (free with a lunch entree) to improve customer satisfaction. He believes that the key ingredient that is crucial to the taste is the brand of soy-sauce, so he wants to find out whether the improvement he is seeking can be achieved by using a more expensive brand of sauce rather than the existing one. If the evidence in favor of superiority of the new brand is not strong enough, then he would rather stick to the existing brand and think about other means to improve the quality. His plan of action is to prepare two large portions of fried rice, one using the existing brand of soy-sauce, and the other using the new brand. Fifty individuals participating in the study will be fed one of these two portions of fried rice.

(i) What is the experimental factor here? Which is its treatment level and which is its control level?

(ii) Who or what constitutes the experimental units?

(iii) What do you think should be an appropriate response or outcome variable and how can it be measured?

(iv) Discuss whether SUTVA is reasonable in this experiment, and discuss the choice of units (e.g., people eating at the same table or each person individually).

(v) What can the experimenter do to avoid the complications arising from possible violations of SUTVA?

**Problem 2**

Suppose, in the context of the fried rice experiment described in Problem 1, the owner of the restaurant feels that soy-sauce is not the only determining factor and the following other factors need to be considered when designing the experiment: (i) brand of rice (existing versus new) (ii) the amount of soya sauce per portion (currently 0.5) and (iii) the rice-to-water ratio (currently 1:2).

(a) How many factors are under consideration? Which among these are qualitative and which are quantitative?

(b) For which of these factors are the levels already determined? For each of the other factors, suggest how many levels you should be choosing. Give reasons for your answer and give examples of the levels that can be chosen.

(c) Based on your answer to part (b), write down notation representing potential outcomes assuming that there is no interference among units of the population

**Problem 3**

Suppose you want to conduct a randomized experiment to compare two keyboards, $A$ and $B$, with respect to their typing efficiencies. Your resources consist of (i) a one-page 400-word manuscript and (ii) twenty individuals who are willing to help you implement the experiment by taking part in typing the manuscript. The response variable ($Y$) is the time taken to type the manuscript. Assume that each individual is willing to spend as much time as it takes to type the manuscript but only once.

(a) What are the experimental units in this experiment?

(b) Write down a natural *estimand* for this problem in terms of potential outcomes. Is this the only possible estimand? If not, provide some alternatives.

(c) Explain how you would assign the 20 individuals to one of the keyboards if:

(i) All the 20 individuals are males of roughly same age with no formal training in typing.

(ii) The 20 individuals comprise 10 males and 10 females. None of the individuals have any formal training in typing.

Table 1: Pre-test typing speeds of 20 individuals

| Individual | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sex | F | M | F | F | F | M | M | F | M | M |
| WPM | 30 | 34 | 21 | 33 | 22 | 22 | 30 | 22 | 20 | 20 |
| Individual | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Sex | M | M | F | F | M | F | M | F | M | F |
| WPM | 32 | 30 | 23 | 30 | 22 | 22 | 21 | 17 | 20 | 30 |

(iii) The 20 individuals comprise 10 males and 10 females. Although none of them has a formal training in typing, you ask each one of them to take a simple typing pre-test and you record the data (see Table 1) on their typing speed measured as words per minute (WPM).

(d) Write down how many randomizations are possible in cases (i)-(iii) above.