**Biostatistics: Homework #1**

1. Cluster Randomization: For some studies, particularly if the intervention involves physicians, there is a choice you need to make about whether to randomize patients or to randomize physicians. Assume that you wish to evaluate a weight-loss intervention that is introduced to patients during a primary care visit. Success with this intervention depends on **both** the primary care physician and the patient. Usually, in carrying out such a study, you will have access to many patients but relatively few primary care physicians. Below are two ways in which you can design such a study.

Simple Patient Randomization: For every patient who comes in and is eligible for the study, a coin is flipped. If the coin is “heads”, then the primary care physician enrolls the patient in the weight loss intervention. If the coin is “tails”, then the patient is treated in the usual way. In both cases, you follow the patient to see how much weight the patient has lost at 12 months.

1. Assuming that the coin is fair (i.e. there was a 50% chance of “heads”), what is the probability that a patient receives the weight loss intervention? (2 points)
2. Is this a randomized trial? Why? (2 points)
3. Are the patient outcomes independent of each other? Why? (2 points)
4. If the coin was not fair (i.e., there was a 67% chance of “heads”) would it still be accurate to call this a randomized trial? Why? (2 points)

Physician Cluster Randomization: For every primary care physician in the study, a fair coin is flipped. If the coin is “heads”, then all of the physician’s patients receive the weight loss intervention. If “tails”, then all of the physician’s patients receive usual care.

1. Argue that every patient has a 50% chance of receiving the computerized intervention. (2 points)
2. Are the patient outcomes independent of each other? Why? (2 points)
3. What are the advantages and disadvantages of using simple patient randomization? (2 points)
4. What are the advantages and disadvantages of using physician cluster randomization? (2 points)
5. Block Randomization**:** In most randomized studies, patients are assigned to Treatment A versus Treatment B using a scheme called block randomization. First, the investigator chooses an even number as a “block size”, usually 6 or 8. For this problem though, we will assume a block size of 4. Then, in each group of 4 patients, 2 will be assigned to Treatment A, and 2 will be assigned to Treatment B. Therefore, in the first 4 patients, there are 6 possible treatment assignment patterns: {AABB}, {BBAA}, {ABAB}, {BABA}, {ABBA} and {BAAB}; and each pattern has probability 1/6 of being used. This is repeated for the next 4 patients, and the next 4, and so on.
6. What is the probability that the first patient will be assigned to Treatment A? (2 points)
7. What is the probability that the second patient will be assigned to Treatment A? (2 points)
8. If the first patient is assigned to Treatment A, what is the probability that the second patient will be assigned to Treatment A? (2 points)
9. Are the treatment assignments independent within a block? (2 points)
10. Are the treatment assignments independent between blocks? (2 points)
11. Name one advantage of having a small block size (think practically, not statistically). (2 points)
12. Name one disadvantage of having a small block size (practical, not statistical). (2 points)
13. Independence**:** In a cohort of 100 people: 12 have diabetes alone; 22 have high blood pressure alone; 18 have both diabetes and high blood pressure; and the remaining 48 have neither diabetes nor high blood pressure. Are these data suggestive of independence or dependence between diabetes and high blood pressure? (6 points)
14. A 3-armed randomized trial was carried out in patients with rheumatoid arthritis (RA). The goal of the trial was to improve RA symptoms through either cognitive behavioral therapy (CBT) or relaxation response therapy (RR), when compared to an attention control. Outcome measures for the 3 groups were collected at baseline, at 6 months and at 12 months. One outcome measure was the time it took for a patient to walk 50 feet. Another outcome was the physician’s rating of joint swelling.
15. How would you display the results of this study for the walk time outcome? Describe (or roughly sketch, if that’s easier) the graphical presentation and any summary statistics you would present. Do not worry about testing for this assignment. (5 points)
16. Another outcome that was collected (same 3 arms and same 3 time-points) was a physician rating of a patient’s joint: 0=no swelling; 1=detectable synovial thickening without loss of bony contours; 2=loss of distinctness of bony contours; 3=bulging synovial proliferation with cystic characteristics. How would you present these data? (5 points)
17. A longitudinal cohort study was carried out in children (aged birth to 18 years) with hypertrophic cardiomyopathy. Our current interest is in growth, as reflected by the heights of these children at the time of diagnosis. The children’s heights were converted to z-scores using normative data from the CDC, and children who died in the first 6 months of life had a mean height z-score of -1.2 (median -1.0).
18. Explain briefly what it means to convert heights to z-scores, and why this was necessary here. (5 points)
19. From the summary data, how would you describe the heights of children who died within 6 months, when talking to a non-technical audience who might not understand z-scores. (5 points)
20. The following figure shows graphical information about glucose values in the Framingham dataset, according to the level of education (the first group had education level missing; the second group, educ1=1 is lowest; the last group, educ1=4 is highest).



* 1. Describe in words the distribution of glucose values for educ1=1. Please comment on the mean, median, range, and skewness. If it is helpful, the diamond represents the mean glucose value, and the horizontal line is at the median. (5 points)
  2. Based on slide #7 of lecture #1 from Dr. Orav’s slide deck or slide 60 of lecture #1 from Dr. Healy’s slide deck, what test do you think you will need to run in order to determine whether glucose is related to education level? (5 points)
  3. Without doing a test, do you think glucose is related to education level? (5 points)