Denis Ostroushko - PUBH 7420 - HW1

Problem 1

1 - A

Let four treatments be assigned the following symbols for representation randomization assignment:

- NNC + patch = T1
- NNC + no patch = T2
- VLNC + patch = T3
- VLNC + no patch = T4

To get random treatment assignments we will use this rule:

- assign to T1 if a random number is 1 or 5
- assign to T2 if a random number is 2 or 6
- assign to T3 if a random number is 3 or 7
- assign to T4 if a random number is 4 or 8
- discard a random number if it is 0 or 9

A list of random numbers provided to us:

```
 [1] 5 3 7 2 3 2 2 3 4 6 3 2 8 7 5 1 4 2 4 4
```

Corresponding assignments to the entire list of numbers:

```
 [1] "T1" "T3" "T3" "T2" "T3" "T2" "T2" "T3" "T4" "T2" "T3" "T2" "T4" "T3" "T1"
[16] "T1" "T4" "T2" "T4" "T4"
```

First ten assignments:

```
 [1] "T1" "T3" "T3" "T2" "T3" "T2" "T2" "T3" "T4" "T2"
```

1 - B

Let two treatments and control be assigned the following symbols for representation randomization assignment:

- NNC + no patch = C
- VLNC + no patch = T1
- VLNC + patch = T2

We want to have 1:1.5:1.5 ratio, which will randomize more participants to the two VLNC groups. Taking the ratios to whole numbers, we will work with 2:3:3 ratio.

To get random treatment assignments we will use this rule: * assign to C if a random number is 1 or 2 * assign to T1 if a random number is 3,5,7 * assign to T2 if a random number is 4,6,8 * discard 0 and 9

Corresponding assignments to the entire list of numbers:

```
 [1] "T1" "T1" "T1" "C"  "T1" "C"  "C"  "T1" "T2" "T2" "T1" "C"  "T2" "T1" "T1"
[16] "C"  "T2" "C"  "T2" "T2"
```

First ten assignments:

```
 [1] "T1" "T1" "T1" "C"  "T1" "C"  "C"  "T1" "T2" "T2"
```

1 - C

Two possible block sizes are 3 and 6. We can extend this to further multiples of three, such as 9,12,15, etc… but for this exercise I will stick with 3 and 6.

Design:

- For blocks of size 3 we will have the following combinations of C, T1, T2 (Table 1):
- Ideally we want to have three of these combinations (due to the limited number of random numbers between 0 and 9)
- So, I will sample three random numbers from 1 to 6 and use three block options that were randomly selected. Setting seed. randomly selected numbers are 2, 3, 5
- Combinations to be used for schedule (Table 2):
- Similarly, for bock of size 6 we will have the following options. Each of C, T1, and T2 is represented in candidate blocks twice (Table 3). I will print out the first five options out of the total 6:

Table 1: Examples of blocks of size 3

			Block ID
C	T1	T2	1
C	T2	T1	2
T1	C	T2	3
T1	T2	C	4
T2	C	T1	5
T2	T1	C	6

Table 2: Selected block desgins for blocks of size 3

			Block ID
C	T2	T1	2
T1	C	T2	3
T2	C	T1	5

Table 3: Examples of blocks of size 6

						Block ID
C	C	T1	T1	T2	T2	1
C	C	T1	T2	T1	T2	2
C	C	T1	T2	T2	T1	3
C	C	T2	T1	T1	T2	4
C	C	T2	T1	T2	T1	5

Table 4: Selected block desgins for blocks of size 6

						Block ID
C	T1	T2	C	T1	T2	13
C	T1	T2	T2	T1	C	18
C	T2	C	T1	T1	T2	19
T2	C	T2	T1	C	T1	71
T2	T1	C	C	T1	T2	73
T2	T1	T2	C	T1	C	83

- For these block options, I will select six random numbers between 1 and 90. These blocks will be used for randomization schedule: 13, 18, 19, 71, 73, 83
- So, these blocks will be used when we sample a number corresponding to a block of size six (Table 4):

We have a total of nine blocks, randomization schedule will be crated using this scheme:

- use block of size 3 ID number 2 when random number is 1. Next three study subjects will be assigned using a sequence from this block
- use block of size 3 ID number 3 when random number is 2. Next three study subjects will be assigned using a sequence from this block
- use block of size 3 ID number 5 when random number is 3. Next three study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 13 when random number is 4. Next six study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 18 when random number is 5. Next six study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 19 when random number is 6. Next six study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 71 when random number is 7. Next six study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 73 when random number is 8. Next six study subjects will be assigned using a sequence from this block
- use block of size 6 ID number 83 when random number is 9. Next six study subjects will be assigned using a sequence from this block
- random number 0 will be discarded

Randomization schedule according to the rule, first 10 observations printed:

```
 [1] "C"  "T1" "T2" "T2" "T1" "C"  "T1" "C"  "T2" "C"
```

1 - D

- Step 1:
  - when random number = 0, let block size be 3
  - when random number = 1, let block size be 6
  - …
  - when random number = 9, let block size be 30
- Step 2:
  - for each block size option, find all possible combinations of Treatments and Control that correspond to 1:1:1 ratio
  - using random number sampling, select one of the options for each block and use results
  - This is the same algorithm that I described in part (C)
- Step 3:
  - Apply randomization scheme to the random number schedule and get a schedule, same idea as part (C)
  - This works best for trials that have 100s of participants.

Problem 2

- Source one: patients were invited to participate in trial, which may lead to the self-selection bias. People who are willing to participate in the study may be fundamentally different from those who decide to not participate. For example, they may be older, retired, and have more time to participate, which drives their decision.

Problem 3

3 - A

If, the ratios were 2:2:1, and there were 750 patents, then study planners desired to have 300 patients in IR and GR groups, and 150 patients in the NNC group

3 - B

We need to consider the total number of combinations of stratum resulting from a factorial design: 2 (for Age) $\times$ 2 (for menthol status) $\times$ 8 (for cites) = 32. Therefore, we will need 32 schedules.

3 - C

We need to ensure a 2:2:1 ratio, so a block of size five would be the smallest we can employ.

3 - D

Primary disadvantages are:

- Over stratification and impact on statistical power: We will have, on average, $750/32 =\sim 23.4$ participants in each stratum, which is a small number of people which may lead to highly variable estimator of the differences in the treatment results.
- Such level of stratification is administratively burdensome
- We will need to use multivariate regression models to adjust for prognostic variables still in order to do inference on such data.

What we can do instead:

- do random sampling of potential subjects for the study from each of these pre-specified stratum combinations
- With large enough sample size benefits of stratification will diminish. For our scenario, it might be the case that we need to recruit more subjects, and maybe stratify on just one prognostic variable.