Homework_1

Ruoyuan Qian

2019/9/17

Problem 1

a)

I found that there are two rows of variable names which may cause some errors when reading the data into R. As a result, I use `skip=` to skip first row and read others into R.

This is the summary of demographics and co-morbidities in Excersice.xlsx.

Table 1: Demographics and co-morbidities

	Control (N=36)	Intervention (N=36)
Age		
- Mean (SD)	51.500 (10.809)	53.583 (9.581)
- Median	51.000	55.500
Gender		
- female	20 (55.6%)	20 (55.6%)
- male	16 (44.4%)	16 (44.4%)
Race		
- African American	22 (61.1%)	31 (86.1%)
- Hispanic	14 (38.9%)	5 (13.9%)
Depression		
- no	23 (63.9%)	26 (72.2%)
- yes	13 (36.1%)	10 (27.8%)
Smokes		
- no	31 (86.1%)	31 (86.1%)
- yes	5 (13.9%)	5 (13.9%)
HTN		
- no	16 (44.4%)	14 (38.9%)
- yes	20 (55.6%)	22 (61.1%)
T2DM		
- no	17 (47.2%)	23 (63.9%)
- yes	19 (52.8%)	13 (36.1%)

b)

b) i

The table 2.1 showing below is what I drew through R, and I save it as ".csv" file as well, but it is not very readable, so I edited it by Excel and made the table 2.2.

Table 2.1: Pre/Post changes in metabolic parameters (output directly from Rstudio)

```
##     X Variable       Intervention.N.36      Intervention.N.36.
## 1   1                          baseline              six month
## 2   2      SYS          133.64 ± 15.11          125.06 ± 15.44
## 3   3              134 ( 121.5 - 144 )    124 ( 116.75 - 135 )
## 4   4    delta           -8.58 ± 17.17
## 5   5      DIA            75.44 ± 9.10           74.58 ± 12.37
## 6   6                74.5 ( 69 - 81 )        74 ( 65 - 80.5 )
## 7   7    delta            -0.86 ± 8.30
## 8   8      BMI            31.97 ± 6.58            31.21 ± 6.13
## 9   9          29.25 ( 27.375 - 34.1 ) 29.15 ( 26.8 - 32.875 )
## 10 10    delta            -0.76 ± 1.44
## 11 11      HDL           50.17 ± 11.85           50.17 ± 13.07
## 12 12                47.5 ( 40 - 60 )     48.5 ( 43 - 60.25 )
## 13 13    delta             0.00 ± 8.09
## 14 14      LDL          102.94 ± 33.84          100.50 ± 30.39
## 15 15            109 ( 75.25 - 124.5 )     95 ( 76.5 - 120.5 )
## 16 16    delta           -2.44 ± 21.27
## 17 17      GLU          116.64 ± 74.91          107.14 ± 38.65
## 18 18             94 ( 83.75 - 116.5 )    95.5 ( 85.25 - 129 )
## 19 19    delta           -9.50 ± 57.36
##              Control.N.36            Control.N.36.
## 1                baseline                six month
## 2          133.47 ± 15.94           130.14 ± 14.35
## 3   131 ( 122.5 - 143.5 )      127.5 ( 120 - 140 )
## 4           -3.33 ± 14.81
## 5            77.14 ± 9.66             75.69 ± 7.54
## 6       76 ( 68.75 - 85 )         76.5 ( 69 - 82 )
## 7           -1.44 ± 10.11
## 8            34.23 ± 6.16             34.51 ± 5.97
## 9  33.4 ( 29.6 - 37.575 ) 33.05 ( 30.425 - 37.55 )
## 10            0.28 ± 0.97
## 11          48.33 ± 13.70            45.19 ± 10.78
## 12    43.5 ( 39 - 54.25 )         43.5 ( 38 - 52 )
## 13           -3.14 ± 6.91
## 14          99.83 ± 29.06            93.61 ± 27.47
## 15 104 ( 88.25 - 112.25 )   96.5 ( 77.5 - 110.25 )
## 16          -6.22 ± 23.12
## 17         128.97 ± 73.86           126.61 ± 63.96
```

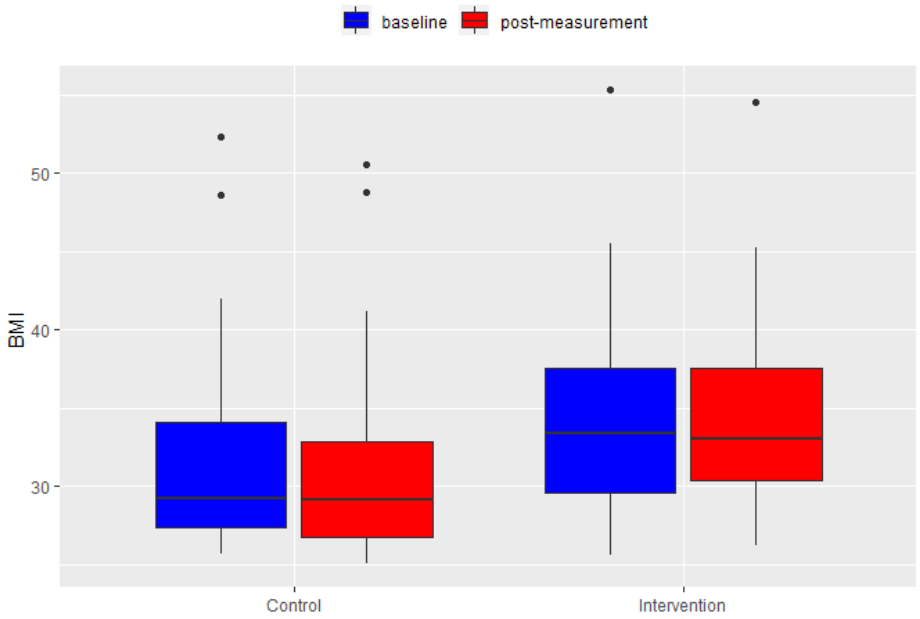
```
## 18     98 ( 81.75 - 139 )    106.5 ( 85 - 145.75 )
## 19           -2.36 ± 51.22
```

Table 2.2: Pre/Post changes in metabolic parameters (edited by Excel based on Table 2.1)

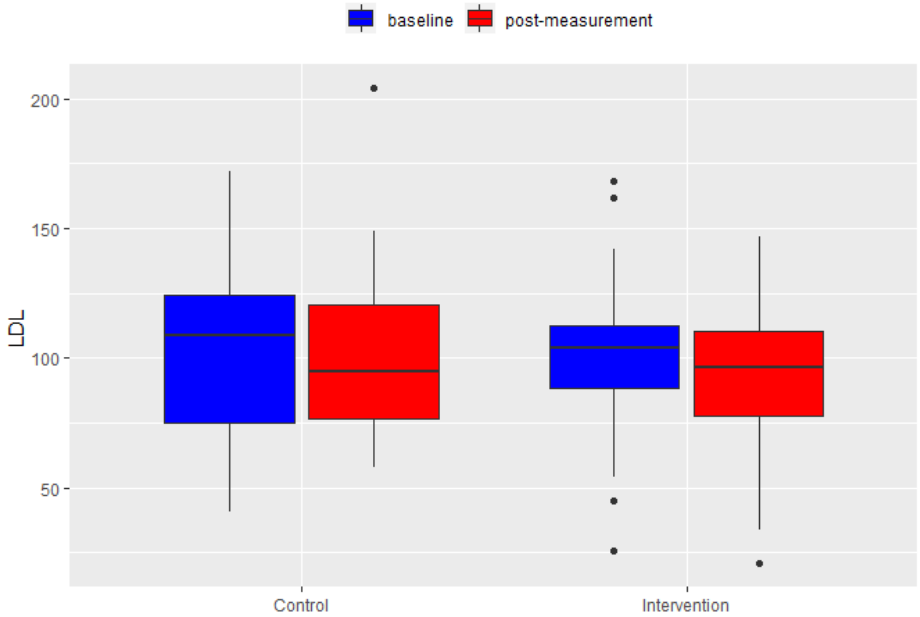
Variable	Intervention,N=36		Control,N=36	
	baseline	six month	baseline	six month
SYS	133.64 ± 15.11 134 (121.5 - 144)	125.06 ± 15.44 124 (116.75 - 135)	133.47 ± 15.94 131 (122.5 - 143.5)	130.14 ± 14.35 127.5 (120 - 140)
delta	-8.58 ± 17.17		-3.33 ± 14.81	
DIA	75.44 ± 9.10 74.5 (69 - 81)	74.58 ± 12.37 74 (65 - 80.5)	77.14 ± 9.66 76 (68.75 - 85)	75.69 ± 7.54 76.5 (69 - 82)
delta	-0.86 ± 8.30		-1.44 ± 10.11	
BMI	31.97 ± 6.58 29.25 (27.375 - 34.1)	31.21 ± 6.13 29.15 (26.8 - 32.875)	34.23 ± 6.16 33.4 (29.6 - 37.575)	34.51 ± 5.97 33.05 (30.425 - 37.55)
delta	-0.76 ± 1.44		0.28 ± 0.97	
HDL	50.17 ± 11.85 47.5 (40 - 60)	50.17 ± 13.07 48.5 (43 - 60.25)	48.33 ± 13.70 43.5 (39 - 54.25)	45.19 ± 10.78 43.5 (38 - 52)
delta	0.00 ± 8.09		-3.14 ± 6.91	
LDL	102.94 ± 33.84 109 (75.25 - 124.5)	100.50 ± 30.39 95 (76.5 - 120.5)	99.83 ± 29.06 104 (88.25 - 112.25)	93.61 ± 27.47 96.5 (77.5 - 110.25)
delta	-2.44 ± 21.27		-6.22 ± 23.12	
GLU	116.64 ± 74.91 94 (83.75 - 116.5)	107.14 ± 38.65 95.5 (85.25 - 129)	128.97 ± 73.86 98 (81.75 - 139)	126.61 ± 63.96 106.5 (85 - 145.75)
delta	-9.50 ± 57.36		-2.36 ± 51.22	

b) ii

Picture 1: Boxplot of BMI variable



Picture 2: Boxplot of BMI variable



b) iii

According to table of Pre/Post changes in metabolic parameters, structured exercise program has a larger impact on the `Systolic`, `Glu` and `BMI` variables, so it means that the program may affect `Systolic`, `Glu` and `BMI` more greatly compared with the impact on the other variables.

Moreover, there are some variables that changed more severely in control than in intervention group, such as `diastolic`, `HDL` and `LDL`. In other words, the program is not beneficial to the decline of `diastolic`, `HDL` and `LDL`.

As for the BMI in the study, in general, the values in intervention group are higher than the control group's while there is little difference within each group.

As for the LDL, for both groups, there is a trend that the values in post measurement are lower than the baseline's, and it is more obvious in control. Besides, in intervention, the values are more concentrated whereas values in control are more dispersed.

c)

It is an interventional clinical trial, and from the perspective of demographics, the two groups are basically balanced, which is favorable for the study.

However, there is some potential issues as well. For instance, the medians of baselines of BMI in two groups are of great difference, which may make the two groups incomparable. Besides, as for the LDL, the variances of two groups are extremely different, which may cause the similar problem like BMI.

Besides, the study dose not implement blinding, so there may contain a lot of bias from doctors, subjects and so on. What is worse, the control is non-participating, as a result, the new program in the study is just compared with blank control instead of standard control. Due to that, the study can only tell the difference between new program and non-participant but fails to test the difference between new program and established one. Finally, randomization is not designed in the study, which may become a big bias in grouping.

To sum up, the study contains plenty of bias and the result is not convincing enough and maintains huge limitation.

Problem 2

The prevalence of the DS: $P(D) = 0.001$

The sensitivity of triple test: $P(T \mid D) = 0.6$

The false-positive rate: $P(T \mid D^C) = 0.05$

The probability of positive predictive value:

$$P(D \mid T) = \frac{P(T \mid D) \cdot P(D)}{P(T \mid D) \cdot P(D) + P(T \mid D^C) \cdot P(D^C)} = \frac{0.6 \times 0.001}{0.6 \times 0.001 + 0.05 \times 0.999} = 0.012$$

The magnitude is so small compared with the cfDNA tests and biochemical screening test. Since prevalence plays an integral role in PPV, the possible reason that triple test obtains such low PPV is due to the different prevalence in different areas and different populations.

Problem 3

Primary information about the study:

The study was about association of midlife to late-life blood pressure patterns with incident dementia. It was a prospective population-based cohort study[1]. The sample size was 4761 patients. And the population of interest was midlife people with normotension, hypertension or hypotension. The exposures were normotension, hypertension and hypotension. There were totally 6 visits during the study. The main outcome of the study was dementia onset after fifth visit.

The results in the article said that Participants in the midlife and late-life hypertension group and in the midlife hypertension and late-life hypotension group obtained hazard ratio (HR), 1.49 with 95%CI, 1.06-2.08 and HR, 1.62 with 95%CI, 1.11-2.37, respectively, which means both groups contained statistical significantly higher risk than normotension group. To sum up, people having hypertension at midlife are more likely to obtain dementia in their late-life.

Pros:

1. The sample size is big enough, so the result is more reliable and authoritative than studies containing small sample sizes.
2. It is designed as cohort study rather than case-control study, so it contains less bias than the case-control study. Besides, it last for over 24 years, which is a relatively long time. As a result, because of both of reasons above, cause-effect is easier to established due to temporality.
3. The study excludes participants who met criteria for dementia before or at visit 5, which is a rigorous way to ensure that the study is examining the long-term relationship between blood pressure and dementia.
4. The study applies several measurements to diagnose the dementia, which will make the study more precise.
5. A proportional hazards assumption was tested before cox model, which is of great importance to choose the appropriate statistics model. Moreover, utilizing cox proportional hazards regression models, the study makes the best use of the data including the censoring.

Cons (potential bias):

1. The criteria for testing late-life normotension, hypertension and hypotension was just

induced at fifth visit, which may not be a valid measurement to group people. I mean that some people may be miss-grouped. As a result, such bias may affect the outcome.

2. The criteria fails to indicate the effect of medication, such as antihypertensive medication use, so it could be a confounder of the study and become a potential bias to affect outcome.
3. The study includes 4 U.S. communities, but the analysis dose not indicate the potential difference between communities and whether they can represent the whole population of interest.

Media link:

https://www.cbsnews.com/news/controlling-blood-pressure-may-help-prevent-dementia/

References:

[1]. Keenan A. Walker, P., et al., Association of Midlife to Late-Life Blood Pressure Patterns With Incident Dementia. JAMA Original Investigation, 2019.