Notice: Only the questions with the score written have been scored.

1 [30 pts] Sarah would like to know the average sleep hours of all high school students in Korea. Therefore, she selected 1,000 students randomly and measured their sleep hours.

- Population: a collection of sleep hours of all high school students in Korea
- Sample: a collection of sleep hours of 1000 students randomly selected by Sarah
- Parameter: average sleep hours of all high school students in Korea
- statistic: sample mean of 1000 randomly selected students

- Correct all: **(+30 points)**
- Correct more than 2 definitions: **(+15 points)**
- Otherwise: **(+0 points)**

2 (a) Since shopping patterns may differ on Friday and Saturday, it is hard to determine each promotion's actual effect. To clarify this experiment, we could offer one promotion on a Friday and the other on the following Friday. Similarly, one promotion on a Saturday and the other on the following Saturday.

(b) Similar to (a), launching two campaigns in different states may adversely affect the precise comparison of such two campaigns. To control this, we could launch both campaigns in (separate) parts of the same state.

(c) A control is needed for comparison; if we compare this year's yield to last year's, we will not know how much of the difference can be attributed to changes in the economy. So instead, we should compare the new strategy's yield with another investment portfolio that used the old strategy.

3 (a) This is SRS(simple random sample): all students have the same chance with the probability $\frac{5}{55}$.

(b) Since people visiting the site decide whether they answer the online poll, it is a voluntary response sample.

(c) Since this experiment divides the sampling frame into distinct groups(male and female), it is a stratified random sample.

4 (a) "Population" cannot be selected; If it is selected by a method, it is a sample. However, according to the scenario, all individuals are selected as SRS, so 'all individuals' are sample. (In fact, in this case, we have a census rather than a sample.)

(b) In fact, "Dihydrogen monoxide" is H_2O. If respondents answer any level of concern about the danger of dihydrogen monoxide, they may misunderstand this substance. Conceivably, the respondent knew that the question was about water and had concerns arising from a bad experience of flood damage or near-drowning. However, misunderstanding seems to be more likely.

(c) This scenario relies on students' honesty in class. Instead of this method(raise their hands), an anonymous survey could obtain more precise data about cheating. In a public setting like this, it would be surprising if there were any raised hands (even though there are likely to be at least a few cheaters in the room).

5 [30 pts: 6 pts each]

(a) Whether or not they had diabetes

(b) The explanatory variable is height. It is categorical in this case because either someone is "short" or they are "tall".

(c) It must be an observational study because we cannot randomly assign someone their height.

(d) No, this is not an experiment.

(e) They need to sample randomly. If they were randomly sampled, they could generalize the results to all adult men in the population.

6 [40 pts for (a), (d), (e), and (g)] Answers with R.code and explanation will be shown in HW1(problem6) html file.