June 23
Grade 9

Grade 9 Mathematics Lesson Plan
Sample survey: Lessons that bring out the merits and the need for random sampling
* Revised version

Date: Thursday, June 23, 2016
15:05 - 15:55
Tokyo Gakugei University International Secondary School
Grade 9 Homeroom 3
Teacher: ARAI, Kenji

1 Introduction

1.1 About the Secondary Mathematics Research Group of Tokyo Gakugei University Attached Schools

This lesson is designed as a part of the research efforts of the Secondary Mathematics Research Group of Tokyo Gakugei University Attached Schools.

The purpose of this research group is as follows. Although mathematics education in Japan has emphasized the importance of the processes of mathematical exploration and the development of ways of thinking and abilities that are needed in those processes, mathematics teaching still focus almost exclusively on teaching of mathematical content. Of course, we are not saying that mathematical content is unimportant. However, even though "mathematical ways of observing and thinking" and "mathematical activities" have been long stressed in Japan, we wonder if mathematics lessons in Japanese classrooms clearly reflect the emphases.

To emphasize mathematical processes means to emphasize the process of using and creating mathematics. As an activity, that process may be grasped as "mathematical activity." The ways of observing and thinking utilized in that process is "mathematical ways of observing and thinking. In this way, our research group considers the entire processes involved in using and creating mathematics as "mathematical process."

When thinking about lessons that has "mathematical process" as the focus, questions such as the following will immediately arise. What does a lesson that focus on "mathematical process" look like? In such a lesson, how should we nurture students' ability for mathematical ways of observing and thinking? What ought the instruction that raises the quality of "mathematical process" be? How can we assess the quality of "mathematical process? How do we know when the quality is raised?

Therefore, the goal of our research group is the way mathematics lessons should be organized to raise the quality of "mathematical process," and this research lesson is a part of our efforts.

1.2 About the mathematical content of this lesson

It has been while since the usefulness and the importance of statistics, but how well are we teaching statistics so that students can actually experience its usefulness and importance. That question became the original of this particular research lesson.

In the current Japanese National Course of Study (COS), the treatment of statistics has been expanded. However, according to some recent reports, further enhancement of statistics education is necessary as our society becomes further globalized and information driven. One of the most important challenges facing the

mathematics education community is how to provide high quality statistics education to respond to this reality.

Although there are many reports and practical suggestions for teaching statistics, many issues still remain, particularly in the area of inferential statistics. For example, let's think about the topic of this unit, "sample survey." In order for students to experience the importance and the usefulness of a sample survey, students need to observe first hand the possibility of extracting the information about the population from a sample. In order for students to understand that the tendency observed in randomly selected samples is approximately the same as that of the population, it is necessary to grasp the tendency of the population. However, if we know the tendency of the population, there is no need for conducting a sample survey. On the other hand, if we do not know the tendency of the population then there is the need for a sample survey. However, in such a situation, it is difficult to argue that the tendency observed in samples is indeed similar to that of the population. I believe this "trade off" is the largest challenge in the unit, "Sample Survey."

With this challenge in mind, the proposal we want to present in this research lesson is the "construction of the unit, 'Sample Survey." Sugiyama (2009) said the following about the instruction of functions and statistics.

> I want the instruction of functions and statistics to communicate to students that they are both tools for solving problems or strategies for problem solving. ... We should not study statistics so that we can draw graphs or calculate various statistics. Rather, there should be a problem we want to solve, and we collect data and examine the problem situation using statistics so that we can solve the problem. We are not collecting data to do statistics.

We need to keep this spring in mind as we teach sample survey. We often study the mechanism of sample survey using simulations and random number generators. However, instead of using those tools to learn about sample survey, we should use sample surveys to solve a problem, and use those tools to examine the validity of the mechanism of sample survey. Thus, in this unit, students are presented with a situation in which sample survey is needed. Then, in order to discuss its methods and validity, we will tackle a task that is detached from the context, which will eventually contribute to the solution of the original problem.

2 Name of the Unit Data Analysis III

3 Outline of the Unit

3.1 Problem for the Unit

Rieko is tired of having so many homework and projects. She thought the reason she is so busy with the assignments is because she is a student at the ISS. She wanted to examine the hypothesis, "Students at the ISS is busier than other lower secondary school students in Tokyo." She decided to use the hours of sleep as the indicator for "busy-ness" and decided to survey the ISS students and other lower secondary school students in Tokyo. However, when she submitted her research plan, the teacher gave it back to her saying this will not be a valid survey. What is wrong with this plan? How should she conduct this survey?

Survey Plan

Survey 1
Ask all 120 9th grade students at the ISS about the number of hours they typically sleep.

Survey 2
Ask friends who go to the same *juku* about the number of hours they typically sleep.

The topic of the problem for this unit is to "investigate the hours of sleep for lower secondary school students in Tokyo." Because our school has been designated as an IB･SSH･SGH school, many students complain that they have too many assignments. By making this familiar issue as the research question, we can focus on the examination of the methods to conduct this survey.

In order to address this research question, students must first experience the need for a sample survey. To help students experience the need for a sample survey, we will have students think about the challenges of surveying the entire population. Some students may think it is feasible to survey all lower secondary school students. However, there are 815 lower secondary and secondary schools in Tokyo. Therefore, they would need to survey the sleeping hours of about 310,000 students. From this fact, students are likely to think about ways to predict the average sleeping hours for the whole population from a smaller set of data. Thus, they realize the importance of sample survey, and how to create smaller groups (samples) from the population will become their new task.

The next discussion will be how to construct samples, and it is anticipated the following 2 points will become the issues.

The first issue is the sampling method. In the current COS and textbooks, the simple random sampling method using a random number table is often shown. This method is easy to make sense since it simply pick samples randomly from the target population. On the other hand, there are many issues with this method such as needing to have the roster of the entire population or the inability to make use of critical information about the population. As a result, this method is not typically used. In the actual sample surveys, methods such as cluster sampling, 2-stage sampling and stratified sampling are used (Table 1). In this particular problem, these ideas can be readily developed. Since the population in this problem is all lower secondary school students in Tokyo, instead of randomly selecting samples from the entire population, for example, students can think about sampling after dividing the population based on regions (central Tokyo, suburbs, and mountain region) or grade levels. Thus, we believe this problem situation allows students to construct sampling methods that can be actually applied.

The second issue is the sample size. It is anticipated that the question of how many students to sample so that the sample will be a good model for the population

will naturally arise. Generally speaking, the sample size is determined based on the population variance. However, since the population variance is usually unknown, an approximation based on past surveys is often used. Concerning the population variance and the determination of the sample size, Fukui (2013) stated the following:

> In either way, because it is difficulty to know the accurate population variance, we must use an estimate. ... Thus, we cannot deny that the determination of the sample size is an approximation. Thus, the sample size calculated with a formula should be considered as a guideline, not as a definitive number.

This suggests that the sample size should be determined as a consensus of the group that is conducting a sample survey. It is easy to imagine situations where creating the sample of the size determined by the formula is not feasible because of budgetary limitation or other reasons. Thus, how to guide the class discussion so that each group can reach a consensus on the sample size will be an important instructional consideration. We plan to emphasize the goal of the survey. Students want to argue that their sleep hours is less than other lower secondary school students in Tokyo. Thus, we want to guide them to discuss what will be an acceptable error range to make their argument. That will focus their discussion on the sample size, and it will be an opportunity for them to understand that statistical analyses often depend on purposes of surveys.

Table 1 Sampling Methods and Their Uses (Fukui, 2013)

Methods	Sampling procedure	Uses
Simple Randome Sampling	Randomly select samples from the entire population.	Not used much.
Cluster Sampling	Make the population into clusters, then select a cluster and survey all members of the cluster.	Some surveys use this method.
2-Stage Sampling	Make the population into clusters, then select a cluster. From the cluster, randomly select samples to be surveyed.	Commonly used in surveys that involve "household" and "individual."
Stratified Sampling	Make the population into a number of stratified groups. Then, select samples from each of the stratified groups.	Surveys involving business entities often utilize this method.
Stratified 2-Stage Sampling	Make the population into a number of stratified group. Then, in each stratified group, conduct the 2-stage sampling to create the sample to be surveyed.	Commonly used in surveys that involve "household" and "individual."
Sampling with Probability Proportion	In the methods involving clusters, the selection of clusters to be surveyed is made proportional to the size of the clusters.	Some surveys use this method.

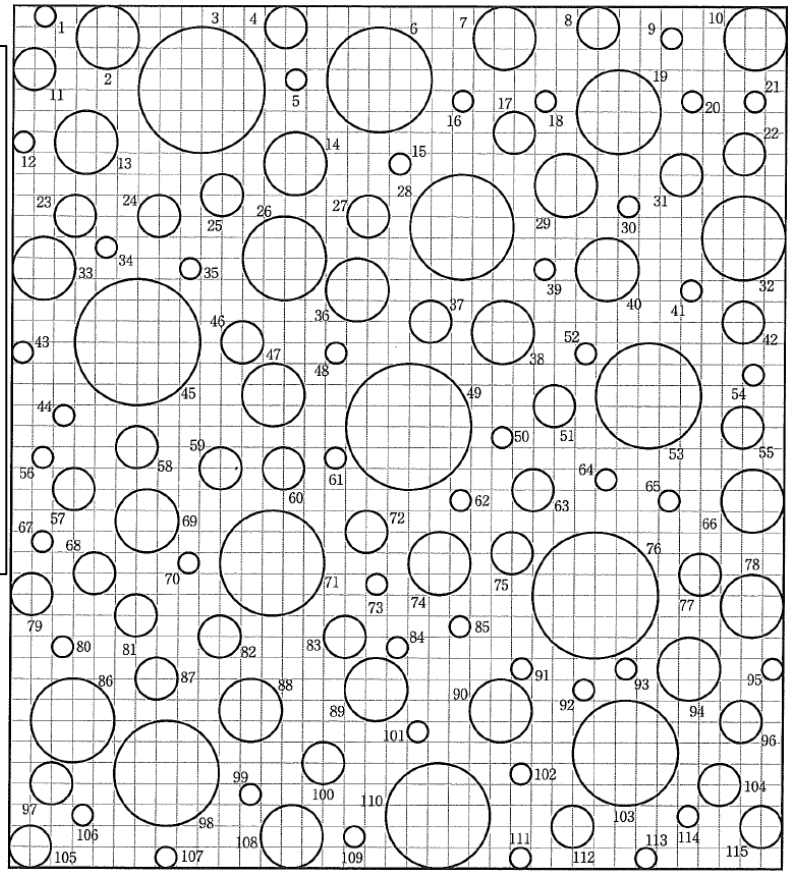
3.2 Goal of the Unit

The goal of the unit is for students to generate the idea of estimating population tendencies from a sub-group of the population and to devise methods for such a survey.

4 Task and Instruction in Today's Lesson

4.1 Task for the Lesson

In the figure shown on the right, there are 115 circles, and we want to investigate the average (mean) of their area. However, to find the area of all of the circles is rather tedious. Is there a way to determine the average of the area of these circles without actually calculating the area of all of the circles?



The task for this lesson comes from Investigation 2, "Can we trust sample survey?" in "Ch. 3 Foundations of Statistics" in the school-developed textbook, *TGUISS Mathematics 4*. Before this lesson, students learn about sample survey and randome sampling. The goal of this task is for students to actually experience whether or not random sampling is really effective.

First, after displaying the figure, students will be asked to estimate the average area of these circles. Then students will verify whether random or non-random sampling will give them an estimate that is closer to the average area of the circles in the population. The sample size will be 5. The larger sample size will improve the accuracy of random sampling but it is also the case for non-random sampling. On the other hand, if we use the sample size less than 4, it may be more difficult to value the merit of random sampling.

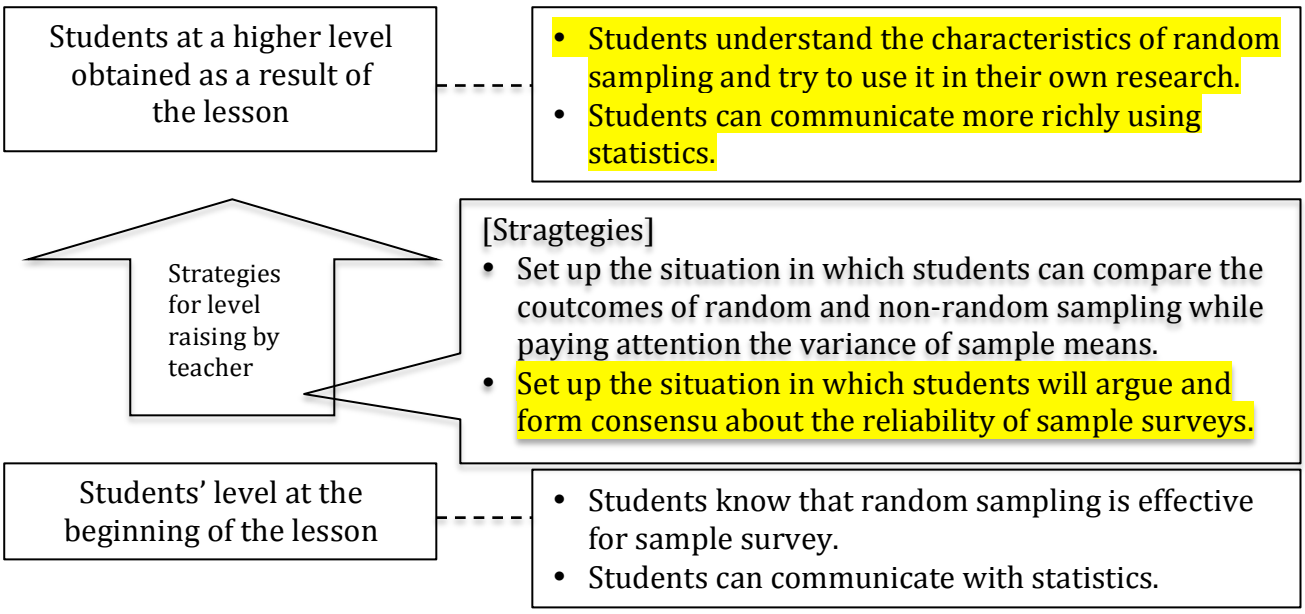
4.2 Goals of the Lesson and Instructional Strategies

There are two major goals for today's lesson.

One goal is for students to experience and understand the merits of random sampling and develop the disposition to make use of what they learned in their own projects in the future. One strategy to achieve the goals is to set up the problem situation in which students can easily compare the outcomes of random and non-random sampling. In particular, we want students to pay attention to the variance of sample means. By observing the distribution of sample variances, students will identify characteristics of both random and non-random samplings. From the identified characteristics, we hope to lead to the discussion of which method is appropriate for a sample survey or how we should employ the selected method to make it appropriate.

By doing so, students can make use of what they learn in this lesson when they tackle their own research question in the future.

The second goal is the improvement of students' ability to use statistics in communication. It has been suggested that "the importance of statistics as a tool for global communication has been increasing" (Watanabe, 2014, p. 35). "Communication" has become an important focul point in mathematics lessons recently. In fact, one of the assessment standards for the MYP mathematics is "communication." One of the goals of the mathematics department is "developing the ability to communicate effectively using mathematics." This also applies to statistics. In today's lesson, we will intentionally set up the situation in which students will argue and develop a consensus on the reliability of sample surveys based on their experimental data. We hope that communication using statistics will become a routine part of students' reasoning through today's lesson. We also hope that the quality of their discussion will become even higher as they think about the issue of uncertaintity.



4.3 Unit Plan

Analysis of Data (Total of 5 lessons)

1) Sample survey and population survey ..... 1 lesson
   - Need for and the meaning of sample survey
   - Characteristics and differences between population survey and sample survey
2) Random sampling and sample survey ..... 2.5 lessons
   [Today: 2 of 2.5]
   - Need for and methods of random sampling
   - Sample mean and population mean
   - Sample size
3) Summary ..... 1.5 lessons
   - Sample survey and methods in the context of the original problem
   - Merits and limitations of sample survey
   - Assigning a report

4.4 About the Students

Although the students in Homeroom 3 can complete project assignments individually, they tend to seek the solutions as they discuss the assignment with their peers. Many students engage in projects deeply and often come with questions that extend the original assignments outside of regular lesson periods. On the other hand, there is a mixture of students who are very talented in mathematics and those who are struggling with the basic mathematical understanding and procedures. Thus, care must be taken when discussing sophisticated mathematical ideas.

So far, these students have learned all basic descriptive statistics ideas except distribution/standard deviation/correlation. Although students have not studied probability, some students may use the term, probability, in today's lesson. However, they may be using the term more as an everyday term, therefore, we will treat it to simply mean likelihood of occurrence and avoid technical discussion.

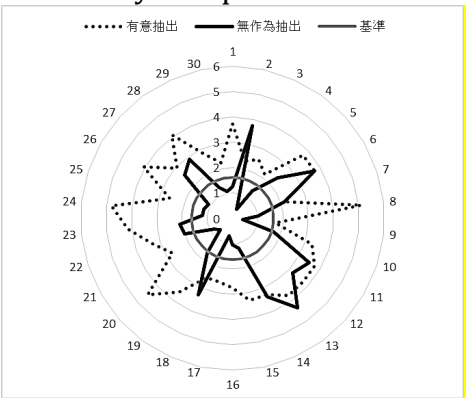
4.5 Flow of the Lesson

min.	Learning Activity (*Hatsumon* (T), Anticipated response (S))	Points of consideration
5	1. Introduction T1: In the last lesson, you have calculated average are of circles you picked, using random and non-random selection methods. Today, we will be analyzing the data you collected. • Display the data sets for random and non-random selections.	• The numbers in this data set are (radius)^2. • Let students know that the "No." is not their roster numbers.

	Non-Random	Random
1	3.75	1.25
2	2.3	3.75
3	2.55	0.4
4	2.15	1.35
5	3.75	2.4
6	3.75	3.75
7	2.15	2.15
8	5.1	1
9	1.7	0.4
10	3.3	1.7
11	3.75	3.5
12	3.75	3.2
13	3.75	4.35
14	3.3	3.4
15	3.3	1.2
16	2.75	1.05
17	2.55	0.7
18	2.55	3.3
19	3.6	1.7
20	4.5	0.65
21	2.75	0.85
22	3.3	2
23	4.15	2.1
24	4.8	1.2
25	2.6	1.2
26	4.05	1.1
27	2.95	2.55
28	4.1	2.9
29	2.8	1.35
30	2.25	1.1

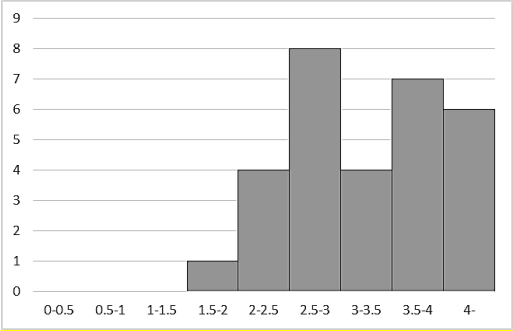
	S1-1: The data from non-random sampling are generally greater. S1-2: The data from random sampling are generally smaller. S1-3: There are people with the same data for both random and non-random sampling. S1-4: The data are more varied with randome samples. T2: Remember that the problem we are working on is to estimate the mean area of 115 circles. Did anyone actually calculate the mean? S2: About 1.63π. T3: That's correct. It is about 1.63π. Keeping that in mind, what do you notice when you look at this set of data or comparing it to your own results? S3-1: I think the values from random sampling are closer to 1.63. S3-2: I think there are many data that are far from 1.63 among non-random samples. S3-3: With either method, no one got the actual mean value. T4: Let's recall what we are doing. What is the purpose of sample surveys? S4: We want to estimate/infer the tendency of the whole population from a smaller set of data. T5: That's correct. In the context of today's problem, we are trying to estimate the mean area of the 115 circles based on the average of 5 circles. Keeping that goal in mind, which method of sampling is more reliable, random or non-random? I want you to come up with a group answer based on data. Make sure you can explain your reasoning. T6: I will hand out USB memory to each group. There is an Excel file with the data, and please use it in your group discussion.	• If no one calculated it, the teacher will give the mean.
15 (20)	2. Group problem solving To conduct a sample survey, which method of sampling is more reliable, random or non-random? Let's explain your reasons. S5-1: Conclude the method that has more values closer to 1.63. S5-2: Calculate the mean of the sample mean. Non-random: 3.27 Random sample: 1.92 S5-3: By calculating the mean of sample means, conclude that the random sampling is more effective because the difference is smaller. Non-random: |1.63 – 3.27| = 1.64 Random: |1.63 – 1.92| = 0.29	[Strategy] "What is the mean of the means?" "In what kinds of situations can we use the mean?

S5-4: Try to represent the data using graphs.

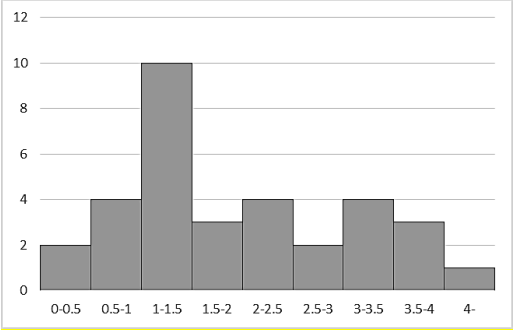


S3-1: Construct a histogram for sample mean for each method

Non-random sampling



Random sampling



S5-6: Display the distribution of sample means in a histogram. Conclude that the non-random selection which result in less variation is more effective.

S5-7: Display the distribution of sample means in a histogram. Because there is possibility that randome sampling might result in the sample mean of 4.35, conclude that the non-random selction is more effective.

S5-8: Display the distribution of sample means in a histogram. Conclude that the random sampling is more effective because the "peak" of the distribution is closer to the population mean.

[Strategy]
"Based on this graph, how do you explain which method is more reliable?"

- Construction of histograms and other graphs will be done with graphing calculators or using GeoGebra.

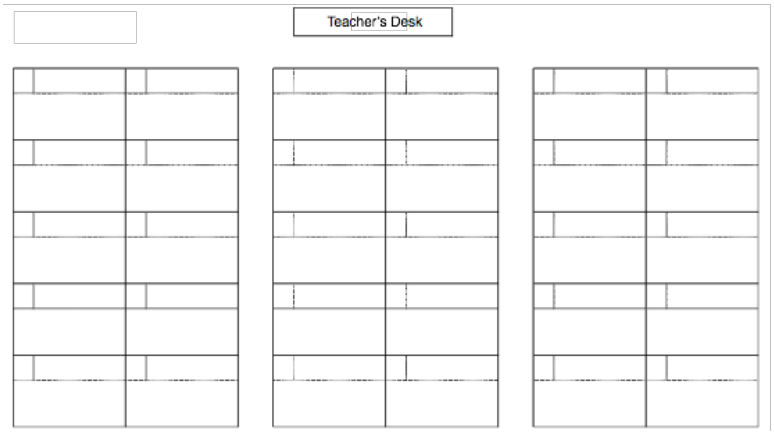
[Strategy]
"Is it possible to compare these methods using something other than variations?"

[Strategy]
"Is their anything you can say by looking at the tendency of the whole distribution?"

	S5-9: Display the distribution of sample means in a histogram.	
25 (45)	3. Comparison and Discussion • Pick up the ideas of S5-3~7 (5-8,9) - the idea of making the data more visual. T6: What do you think about the idea, 5-6? Any comment? S6-1: But the "peak" for the random sampling is closer to the population mean. S5: It bothers me that there is a very large value in random sampling (3-3). S6: Since we are examining the pattern in sampling methods, I think we should look at the distribution as a whole. T7: What do you think about S5-8's idea? What do you think about focusing on the "peak" of each distribution? What does the peak mean in the context of this problem? S7-1: There will be a lot of data points in that interval. S7-2: The graph shows the distribution of sample means. So, it tells us how likely it is to get the mean in that interval visually. T8: What S7 says it is likely to get the sample mean greater than 2.5 with non-random sampling but the sample mean for random sampling is likely to be in the range of 1 ~ 1.5, isn't it? If that's the case, which method is more effective to estimate the population mean? S8-1: Random sampling. S8-2: We can't say either. S8-2: It may depend on the situation? T9: (To S8-2, 3) Why do you think so? S9-1: In today's problem, random sompling was better, but I think there may be situations where non-random sampling is better. S9-2: I'm not sure if I can actually apply the idea to a real survey. T10: (To S9-1) In what kinds of situation do you think non-random sampling works better? S10-1: In the circle problem, we may end up selecting all large circles, but we wouldn't do that if we are using non-random sampling. So, random sampling may give us the result is very different from the reality. So, non-rondom sampling may be better. S10-2: In a real survey, if we use random sampling, we may end up asking people from the same region. So, it's better to intentionally spread them out. T11: In our discussion earlier, someone mentioned that it is more likely to get a sample mean that is closer to the population mean. What do you think in relation to that idea? S11-1 (S10-1) I think we always need to keep in mind that	[Strategy] If no one represented the data visually using graphs, call on S5-1, or the teacher will suggest. [Strategy] If no one raises the idea S4 or S3-4, ask "Where will the population mean be in the graph?" [Strategy] If S5-8's idea does not come up, ask "If you were to write in the population mean in the graph, where would it go?" [Strategy] If the discussion does not follow T7-S7, ask students where in the distribution another (new) sample mean may be located? • Comparing to the problem context, try to answer the problem about the average area of the circle.

	sample mean will not necessarily be closer to the real value. S11-2 (S10-1): We need to keep in mind a sample mean may be very different from the real value. S11-3 (S10-2): Maybe we can intentionally divide the population up into regions then use random sampling in each region. S11-4 (S10-2): Maybe we can randomly sample first, then check how regions are distributed. T12: (To S10-2/S11-2) If you relate the context of the circle problem to the context of a survey, how do you think you can use the idea of sampling? S12-1: S11-3 S12-2: S11-4 S12-3: If we have some specific information about the population like we did with the circle problem, we can consider it in the sampling.	[Strategy] If students' ideas appear to lean toward non-random sampling, remind students about the idea that considered the mean distribution (T9).
5 (50)	4. Summarize T13: Let's summarize. The task for today's lesson was to think about which method of sampling is more reliable, random or non-random, and explain why. Based on our discussion today, one conclusion is that, based on the distribution of sample means, randome sampling is more likely to result in a value close to the actual value. On the other hand, in the real surveying situation, we may have to think about various issues. At this point, you probably have some additional questions and ideas. Keeping those things in mind, let's finish the lesson by writing your reflection. • Students write their reflection entries.	

4.6 Seating Chart



Note: Students will be working in groups of 4~5.

References: Omitted