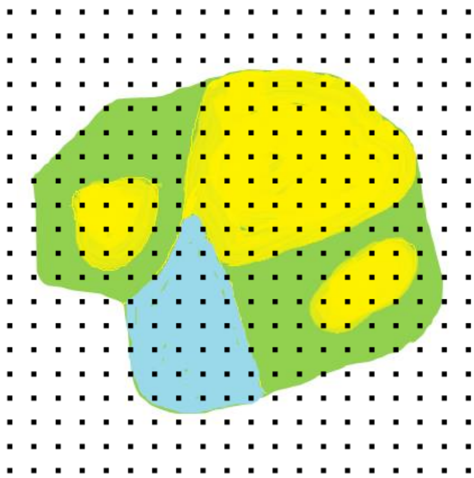
[Workshop] Practical Basic Statistics

Area Measuring

Let's learn basic statistics

through various "Area Measuring"



Issei Jinguji

Former JICA Expert

AGRICULTURAL STATISTICS PLANNING AND DESIGNIG

JICA Tsukuba International Center

14 August 2018

Contents

1. Introduction

Yesterday, you learned basic statistics through “Stone Cutting” which is an interesting experiment on statistical survey. Today, you learn area calculations, or area measuring surveys. In general, when you need data on planted area, you distribute a printed questionnaire of a certain format to local offices, which are asked to fill it out and report. In this case, it does not matter how they fill out the questionnaire. Therefore they often fill in items without properly researching or investigating, such as a reliable survey based on the random sampling and actual measuring method. And those results are often unreliable.

As you have already experienced through the stone cutting survey, if you want to obtain reliable statistics, you should adopt a reliable method based on the statistical manner with measuring. But it seems very difficult to apply such a method to a planted area survey in a big target region.

Recently, a simple method for planted area surveys was developed in Japan. It is called “Dot sampling method”. The sampling is conducted on Google Earth. The detail will be shown by Mr. Kenji Kamikura later.

Here, before you will learn the Dot sampling method, you learn various traditional area measuring methods. It might be roundabout, but it will bring you a lot of fruits in the near future.

2. Measuring for geometric shapes

First we would like to show you a review lesson. Do you remember the following calculation methods which you learned a long years ago when you were students? It might be hard work to apply those methods to an actual planted area survey. But let’s learn step by step

As you know, the area of a shape is a number that tells how many square units are needed to cover the shape. This is the basic concept of area. You can find an area by drawing a shape on graph paper, and counting the squares inside the shape. But it is not efficient when you calculate the area of geometric shapes such as a rectangle or a triangle. Therefore you can use area formulas instead.

(1) Squares and Rectangles

The area of a squares (or a rectangle) is the length (a) times the width (b).

S = a× b

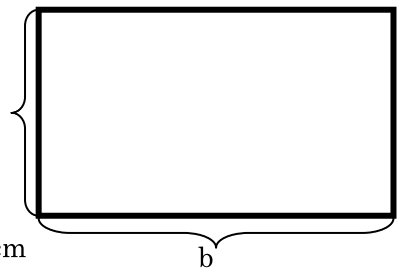


Figure 4: Rectangle

Area calculation　$S = a \times b = 2.7 \times 4.7 = 12.7 cm^2$

(2) Triangles

The area of a triangle is the side (a) times the height (h) and divided by two. The formula is shown below. When you use this formula, you have to make a perpendicular line (h) against a side (a).

$S = (a \times h)/2$.

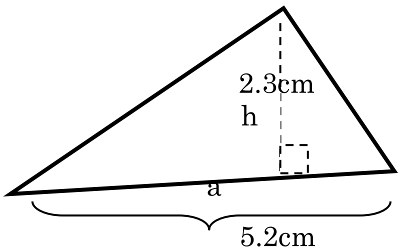


Figure 5: Triangle-1

Area calculation: $S = (5.2 \times 2.3)/2 = 6.0 cm$

Next one is another formula to calculate the area of triangle, which is called “Heron’s formula”. When you calculate the area using the formula, you don’t have to draw a perpendicular line. You just measure the lengths of three sides.

$$S = \sqrt{s(s-a)(s-b)(s-c)} \qquad \text{Where } s = (a+b+c)/2$$

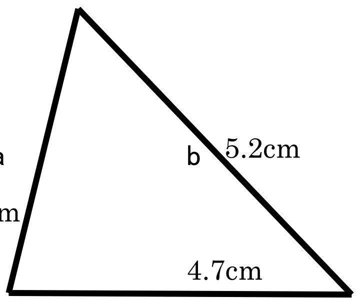


Figure 6: Triangle-2

Area calculation

$$s=\frac{(4.1+5.2+4.6)}{2}=6.85$$

$$S=\sqrt{6.85\times(6.85-3.9)\times(6.85-5.2)\times(6.85-4.6)}=\sqrt{75.02}=8.67\text{cm}^2$$

3. Measuring for complicated shapes

When you meet such complicated or twisted shapes which are shown below, you should use offset methods. The methods consist of basically three types, namely, a rectangular offset method, a crossed offset method and triangular offset method. Those methods are used not only a sheet of paper but also on the grounds.

(1) Rectangular offset method

The rectangular offset method is to draw a rectangle on the target figures. The rectangle should be drawn on the target figure just to be equal size with the area of the target figure. The offset lines are drawn with an eye-adjusting method.

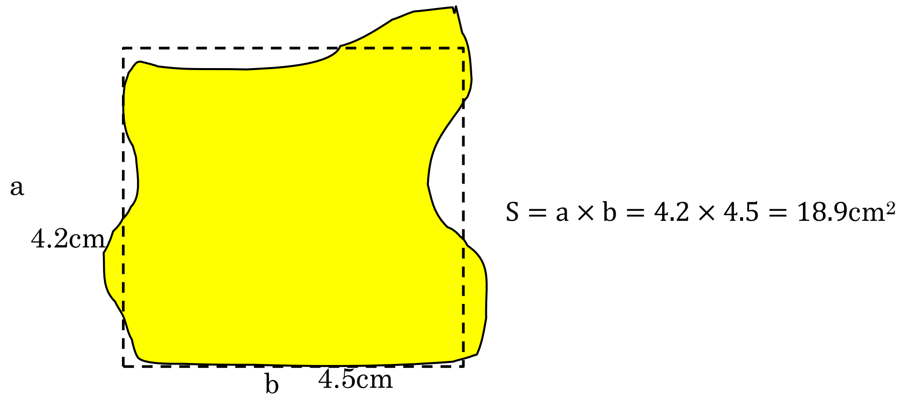
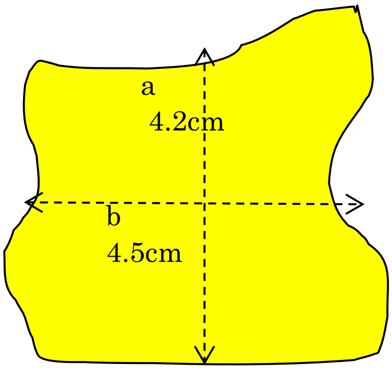


Figure 7: Rectangular offset method

(2) Crossing offset method

Next example shows a crossing offset method which is a kind of the rectangular offset method shown above.



$S = a \times b = 4.2 \times 4.5 = 18.9\text{cm}^2$

Figure 8: Cross method

(3) Triangular offset method

Next example shows a triangular offset method with Heron's formula

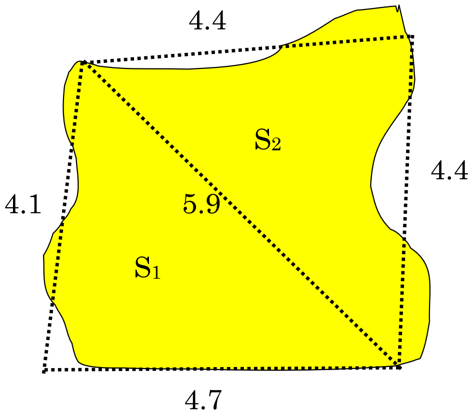


Figure 9: Triangular offset method

Area calculation

S_1 : $s = \frac{(4.1+4.7+5.9)}{2} = 7.35$

$$S_1 = \sqrt{7.35 \times (7.35-4.1) \times (7.35-4.7) \times (7.35-5.9)} = \sqrt{91.79} = 9.58\text{cm}^2$$

S_2 : $s = \frac{(4.4+4.4+5.9)}{2} = 7.35$

$$S_1 = \sqrt{7.35 \times (7.35-4.4) \times (7.35-4.4) \times (7.35-5.9)} = \sqrt{92.75} = 9.63\text{cm}^2$$

$S = S_1 + S_2 = 9.58 + 9.63 = 19.21\text{cm}^2$

(4) Section sheet method

When the area of a complicated shape is drawn on a sheet of paper, it can be measured putting a section sheet on it, which is made of transparent plastic. This method attempts to count the number of square units within the complicated target shape dividing two groups, namely complete squares and incomplete squares groups. For this reason, it can be said that the method is faithfully based on the concept of area ("the area of a shape is a number that tells how many square units needed to cover the shape") as we mentioned above. Therefor this method is taught at a primary school to understand the basic concept of area.

The formula for the area calculation is shown as follows:

Total number of square units = (Number of complete square units × 1) + ((Number of incomplete square units × 0.5)

Area of the target figure = Area per square (unit) × Total number of square units.

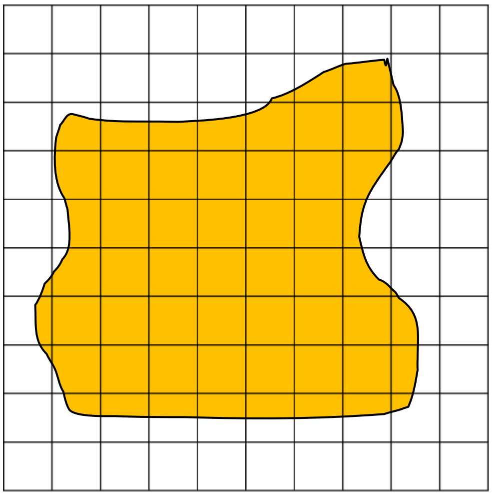


Figure 10: Complicated shape and section sheet method

Area calculation

Area per square (unit) = 0.64 × 0.64 = 0.4096cm^2

Total number of square units= (30 × 1) + (33 × 0.5) = 46.5

Area of the target figure = 0.4096 × 46.5 = 19.0cm^2

(5) Dot-grid sheet method

The area of a complicated shape can be measured with a dot-grid scale instead of the section paper method. When you use this method, first, you count the number of dots within the target figure, after that you multiply the area per dot by the number of dot.

This is a traditional area measuring or calculation method, but it is not known well, because it might be used in a special field, such as "measuring survey" on maps. But this method is powerful and useful for the area calculations. If you use this method, you don't have to measure length, width, height (Compare to Figure4, 5), or three sides (Compare to Figure 6) with scale or have to draw offset lines (See Figure 7, 8, 9). You just count the number of dots in a target figure.

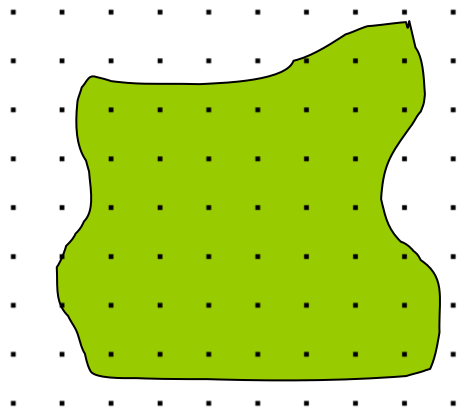


Figure 11: Complicated shape and dot-grid sheet method

Area calculation

Area per dot:= 0.64×0.64 = 0.4096cm^2

Number of sample dot within the figure=45

Area of the target figure =0.4096×45=18.4cm^2

(6) Comparison of the results obtained five area calculation methods

The graph below shows the comparison of the results obtained the above five area calculations (See Figure 7, 8, 9, 10 and 11). The result shows each method is good.

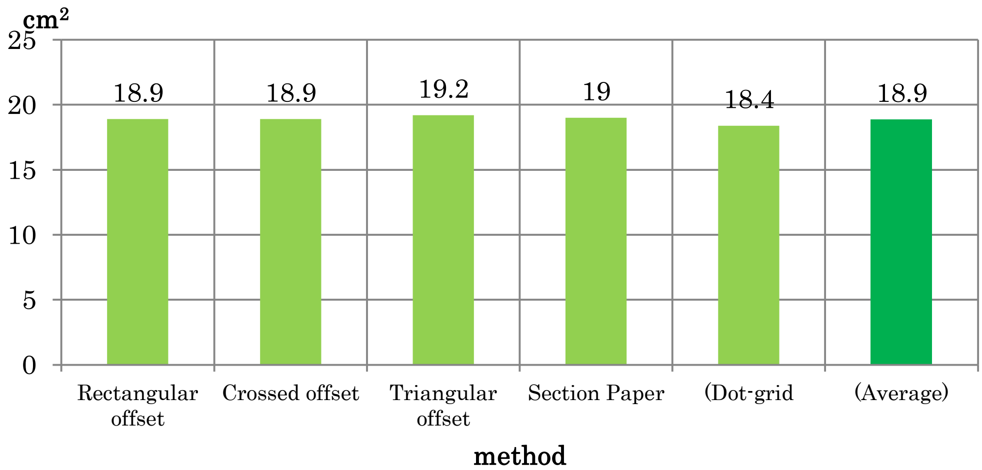


Figure 12: Comparison of the five area calculation methods

4. Measuring for complicated shapes by category

(1) Section sheet method

The figure shown below is more complicated than the figure shown above since it is divided into some segmentations. When you need to calculate the area by color (category/attribute), you can calculate the areas applying above method. But, it seems difficult and boring a bit.

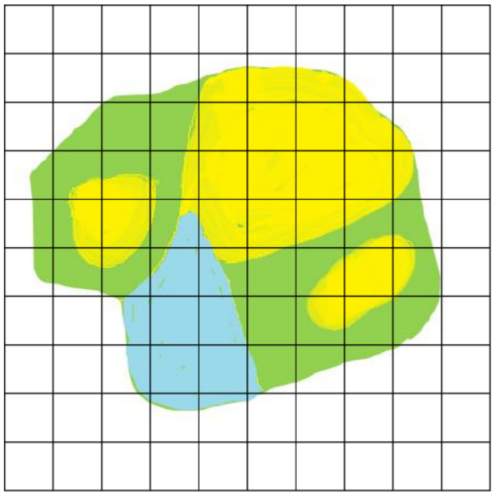


Figure 13: More complicated shape and section sheet method

Area calculation

Area per square (unit):= $0.64 \times 0.64 = 0.4096\text{cm}^2$

Number of squares=47

Area of the target figure $=0.4096 \times 47 = 19.3\text{cm}^2$

The calculation table for the number of squares is shown below.

Table 2: Area calculation table for the section sheet method

Category	Number of complete square (Weight:1)	Number of incomplete square (Weight: 0.5)	Adjusted number of complete square	Percentage (%)
Green	1	16	17.5	37.2
	1	15		
Yellow	0	6	20	42.6
	7	13		
	0	7		

Blue	3	13	9.5	20.2
Total	12	70	47.0	100.0

(2) Dot-grid sheet method

If you adopt the dot-grid method, you can measure the area by color easily. The method is the same as the former method (Figure 11). You will be able to calculate the area by category quickly. You will find out it is like to measure with “a two dimensional scale”. This is a strong method especially for the area calculation of complicated figures as you have known already.

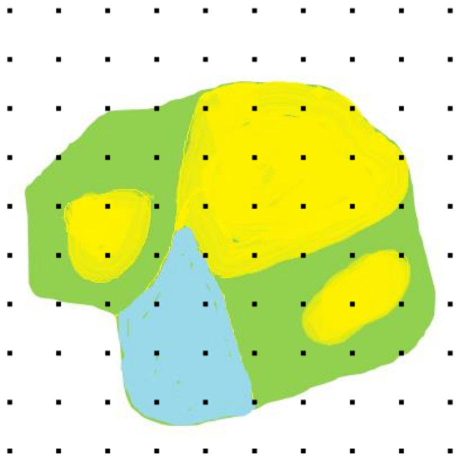


Figure 14: More complicated shape and dot-grid sheet method

An example of the area calculation by color is shown on the table below.

Table 3: Area calculation table for dot-grid sheet method

Category	Number of sample dots		Area per dot	Area by color
	n_i	%	k	$\hat{T} = n_1 \times k$
Green	19	42.2	0.4096	7.8
Yellow	20	44.5		8.2
Blue	6	13.3		2.5
Total	45	100.0	0.4096	18.4

Note 1: Special area measuring tools

We don't explain those methods in detail, just introduce a bit.

(1) Planimeter

A planimeter is a machine to measure complicated figures by tracing the edge (border line) of those on maps.

Figure 15: Area calculation with a planimeter

(2) GIS (Geographic Information System)

GIS has a function to measure the area of land on electrical maps, which is like the planimeter explained above. Google Earth Pro is a kind of GIS.

Figure 16: Area calculation on Google Earth Pro

5. Dot-grid method as an area sampling survey

(1) Change from a measuring method to a sampling survey method

1) Theoretical background

By the way, above explanations are how to measure the size of area on various figures. You might have found out already that those were not statistics, but just actual measuring surveys using various measuring tools and methods. It is true in a sense.

However, as you have already found this area measuring method can be changed as a sampling survey. This method is called "Attribute survey" to estimate proportion by category. The attribute survey is one of the most important methods to generate reliable statistics. The dot grid plate is a type of sampling method. The description of below quotation is in case of the sample dots which are selected with a simple random sampling method. The location of each dot-grid point can be used as a sample point on a figure, which is selected at random using the systematic sampling method. The difference of the both should be notable.

In this connection, you should pay attention that the dot-grid method can estimate area "proportions" by counting the number of dots by category (color) only judging "yes (true)" or "no (false)". In this case, the figure as your target can be a binomial population in statistics.

For instance, a statistic text ("BASIC THEORY OF SAPMLING SURVEYS" MAFF of Japan 2001) which we learned says as follows:

"The binomial population is a distribution in which the values of a variable can be either 0 or 1. If "1" is assigned to those which meet a certain condition and "0" is assigned to those which do not, a distribution called a binomial population will be obtained. For example, the number males and females and the probability of getting "1" or not by throwing a die belong to this category. When n sample (sample size = n) are sampled from a binomial population, the distribution of the sum of sample value is called the binomial distribution. If, it is assumed that the probability that variable x is 1 in a binomial population is P and $1-\mathrm{P}=\mathrm{Q}$ the mean μ of this distribution will be P , the variance σ^2 will be $\mathrm{PQ}$and the standard deviation σ(the square root of the variance) will be $\sqrt{\mathrm{PQ}}$."

2) Adaptation for area survey

Assuming the area of target figure (W) will be given, the area by category ($\hat{T}$) is estimated by the proportion $\{\hat{P} = (n_i/n)\}$ times W. The formula is shown below.

$$\hat{T} = (n_i/n) \times W$$

Based on this idea, the area and precision of above figure (See Figure 14) can be calculated by color like Table 4. In this calculation, the area of target figure is 18.1 cm^2, which was calculated with planimeter measuring.

Table 4: Area calculation table for the dot sampling method

Category	Number of sample dots	Percentage (%)	Area (cm^2)	Precision				
				SE	CV	Compleate survey		Difference
	n_i	$\hat{P} = (n_i/n) \times 100$	$\hat{T} = (n_i/n) \times W$	SE = $\sqrt{\hat{P}\,(100 - \hat{P})/n}$	CV = $(SE/\hat{P}) \times 100$	Area by planimeter	Percentage (P)%	D = $\hat{P} - P$
Green	19	42.2	7.6	7.4	17.4	7.1	39.2	3.0
Yellow	20	44.5	8.1	7.4	16.6	8.0	44.2	0.3
Blue	6	13.3	2.4	6.3	27.0	3.0	16.6	-3.3
Total	45	100.0	18.1	-	-	18.1	100.0	0.0

W: 18.1cm^2 is calculated by a planimeter.

(2) High density dot-grid method for improving accuracy

It is known that a bigger sample size makes a more reliable estimate. It means the estimate becomes close to the true value, and the sampling error becomes smaller. Therefore, we tried to put a higher density dot-grid plate on the figure, and we counted the number of dots by color as follows. The sample size was 179 as the result.

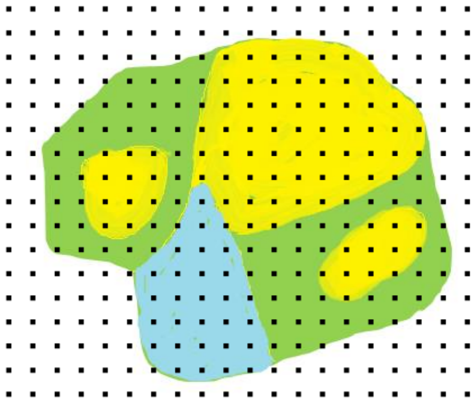


Figure 17: More complicated shape by a high density dot-grid sheet method

The result under the sample size is 179 is shown below. As the former sample size was 45, the new sample proportion is much closer to true value which is measured by a planimeter, and precision (SE, CV) is much better than when sample size was 45.

Area calculation by color is shown on the table below.

Table 5: Area calculation table for the high density dot sampling method

Category	Number of sample dots	Percentage (%)	Area (cm^2)	Precision		Comparison	
				SE	CV	Complete-survey (%)	Difference
	n_i	$\hat{P} = (n_i/n) \times 100$	$\hat{T} = (n_i/n) \times W$	$SE = \sqrt{\hat{P}(100-\hat{P})/n}$	$CV = (SE/\hat{P}) \times 100$	Percentage	$D = \hat{P} - P$
Green	70	39.1	7.1	3.6	9.3	39.2	-0.1
Yellow	79	44.1	8.0	3.7	8.4	44.2	-0.1
Blue	30	16.8	3.0	2.8	16.6	16.6	0.2
Total	179	100.0	18.1	-	-	100.0	0.0

(3) Comparison of the four results and analyses

Next table shows comparison of the four results shown above, namely the section sheet method, dot sampling method, high density dot sampling method and complete survey method.

The estimate of proportion with the high density dot sampling method is closer to the value of the complete survey than other results.

Table 6: Comparison of the high density dot sampling method and others

	Section sheet method (%) (Figure13)	Dot sampling method (%) (Figure14)	High density dot sampling (%) (Figure17)	Complete survey(%) (Table4)
Green	37.2	42.2	39.1	39.2
Yellow	42.6	44.5	44.1	44.2
Blue	20.2	13.3	16.8	16.6
Total	100.0	100.0	100.0	100.0

It is interesting to try analyses using results of various surveys. This is also a field of basic statistics. From the data shown above table, you can draw graphs as follows.

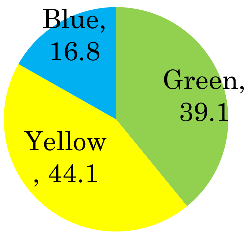


Figure18: Share of Color by the high density dot sampling survey

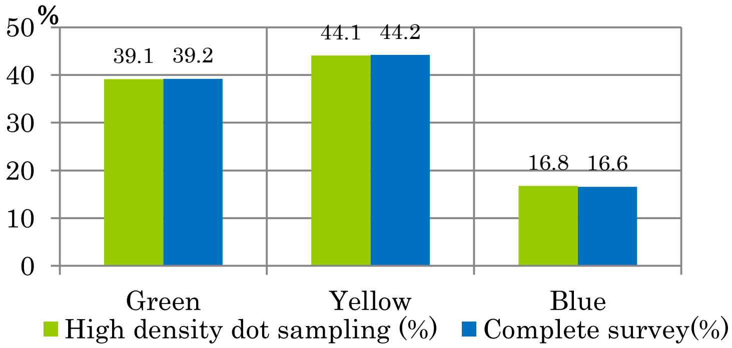


Figure19: Comparison of high density dot sampling survey and complete survey

Furthermore, you will find out that the each difference ($D = \hat{P} - P$) is quite smaller than the SE (Standard error) which shows the simple random sampling error (See Table 5). There are two reasons there, we think.

The first reason is that the dot-grid sample is selected by such a systematic random sampling method, which is a regular interval sampling method in area (or in distance). From a practical point of view, the method is stable, and it has a characteristic such as an actual measuring method with a two dimensional scale, as we mentioned above. You can say the first reason comes from the difference of sampling method.

The second reason is that each color part as an attribute in the figure is gathered into certain sizes. If the each part by color is divided into more small pieces and if they are completely scattered independently, as the result, even the systematic sampling method cannot exert its power, and bring in the same results as the simple random sampling method. You can say the second reason comes from the difference of population's characteristic.

Therefore, it suggests that you can share the systematic random sampling method as a dot grid method is better than the simple random sampling method under such conditions like Figure 17. Of course, in addition, you should not forget the systematic random sampling method is more convenient than the simple random sampling method from the practical view point.

Memo: Concerned this matter, please remember the experiment of the stone cutting survey to estimate the average weight per stone using sampling methods. The standard error (SE) of the systematic sampling method closes to the SE of the simple random sampling method. You found that the systematic sampling method is not superior to the simple random sampling method. The reason why is the arrangement of stones in the population was done without considering the size (weight) or color of each stone. Therefore, unfortunately, the systematic sampling method could not show its power in the case.

(4) Operation on the GIS such as Google Earth

You have learned various area measuring methods through those experiments. We think that you might have been interested in the dot-grid method, especially.

Dot-grid sampling method is an old method. It was known that the method could be applied to an area survey on aerial maps. But those maps were so expensive. Therefore, it was difficult to use such expensive maps for a planted survey. But, the advent of "Google Earth" made it possible. Google Earth is a kind of GIS. Anyone can use it on the Internet.

Connected an old method to the latest technology, a new area survey method was born. We call the method "Dot sampling method", when the dot-grid method is used on Google Earth. And the method is developing day by day.

Note 2: Theoretical back ground (1)

Inferences for Population Proportions

A statistics book (NEIL A. WEISS, "ELEMENTARY STATISTICS". Addison-Wasley Publishing Company USA 1996) explains the attribute survey as follows:

DEFINITION: Population proportion and sample proportion

Consider a population in which each member is classified as either having or not having a special attribute, a so-called two category population. Then we use the following notation and terminology.

Population proportion, $\boldsymbol{P}$: The proportion (percentage) of entire population that has the specified attribute.

Sample proportion, $\hat{\boldsymbol{P}}$: The proportion (percentage) of a sample from the population that has the specified attribute

FORMULA: Sample proportion

A sample proportion, $\hat{\boldsymbol{P}}$ is computed using the formula

$$\hat{p} = \frac{n_1}{n}$$

Where n_1 denotes the number of members sampled that have the specified attribute and n denotes the sample size..

Correspondence between notations for means and proportions

	Parameter	Statistics
Means	μ	$\bar{X}$
Proportion	P	$\hat{\boldsymbol{P}}$

As we know, a sample mean, $\overline{X}$, can be used to make inferences about a population mean, μ. Similarly, a sample proportion, $\hat{P}$, can be used to make inferences about a population proportion P.

KEY FACT: The sampling distribution of the proportion

Suppose a large random sample of size n is to be taken from a two-category population with population $\hat{P}$. Then the random variable $\hat{P}$ is approximately normally distributed and has mean $\mu_{\hat{P}} = P$ and standard deviation $\sigma_{\hat{P}} = \sqrt{P(1-P)/n}$. In other word, probabilities for $\hat{P}$ are approximately equal to areas under the normal curve with parents P and $\sqrt{P(1-P)/n}$.

Note 3: Theoretical back ground (2)

The dot sampling method originates in the point sampling method. Dr. Frank Yates describes the point sampling method as follows (SAMPLING METHODS for CENSUS AND SURVEYS, London, 1949)

3.9 Sampling with probabilities proportional to size of units

"If we have areas demarcated on a map, such as fields, and a point is located at random on the map, the probabilities of the point falling within the boundaries of the different fields are clearly proportional to the areas of the field. Consequently, areas can be selected at random with probabilities proportional to their size by the simple procedure of taking random points on the map." (P.35)

"The principle has applications in agricultural surveys designed to determine the acreage and yield of different crops, total cultivated area, etc. All that is required for acreage is to determine the proportion of points which fall in areas of the given type. The method therefore particularly attractive when carrying out surveys of the area crops, etc., by aerial survey, by provided the different crops can be recognized on the photograph, since it avoids all the measurement of area which would be required if an ordinary random sample of areas were taken." (P.35)

6. Afterword

Under the title of "Practical Basic Statistics", we have shown you some interesting statistical experiments.

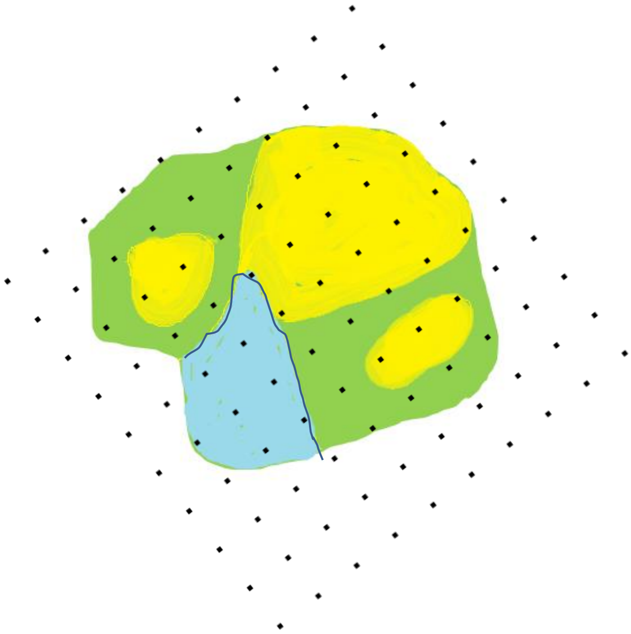
The first one is called "Stone cutting survey". You have found that random sampling method with weighing is really reliable through the experiment. Furthermore, you have learned how to calculate statistical values using the data (a sample variance) you got out from the experiment.

The second one is on the area measuring experiments. You have learned various measuring and calculation methods. The measuring method with dot-grid plate was changed to a sampling survey as an attribute survey. It can be a sampling survey method which is used on even Google Earth. Therefore, the dot-grid method has a possibility to make an actual land use surveys easy dramatically. You are going to learn the method practically soon. It will resolve a lot of problems which you face on the carrying to improve agricultural statistics.

I hope that those experiences on basic statistics for two days would be useful for you to make preparations of latter lectures and to improve agricultural statistics in the future.

7. Quiz [Basic]: Calculate the area and precision of the following figures by color on the tables.

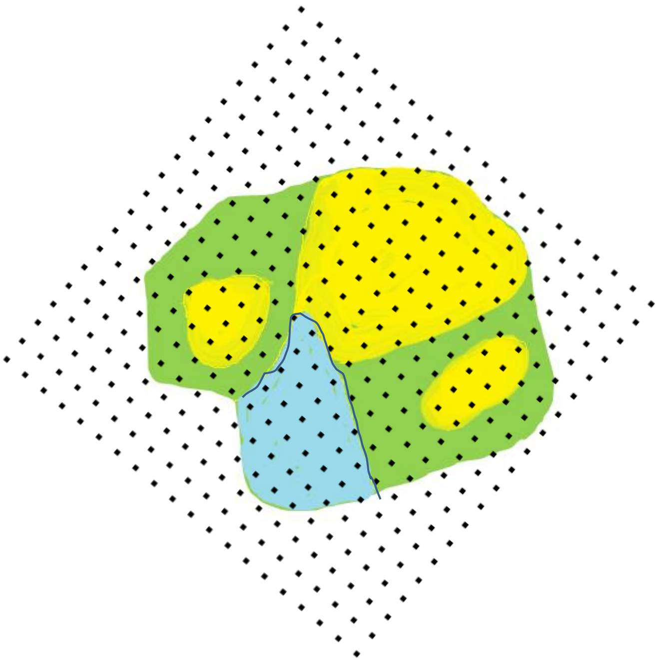
(1) Calculation with a low density dot-grid plate.



Category	Number of sample dots		Percentage (%)	Area (km^2)	Precision		Comparison	
					SE	CV	Complete survey	Difference
	n_i		$\hat{P} = (n_i/n) \times 100$	$\hat{T} = (n_i/n) \times W$	SE $= \sqrt{\hat{P}\,(100 - \hat{P})/n}$	CV $= (SE/\hat{P}) \times 100$	P	D $= \hat{P} - P$
Green							39.2	
Yellow							44.2	
Blue							16.6	
Total							100.0	

Note: W : Each of your country's area is tentatively assigned to W.

(2) Calculation with a high density dot-grid plate



Category	Number of sample dots		Percentage (%)	Area (km^2)	Precision: SE	Precision: CV	Comparison: Complete survey	Comparison: Difference
	n_i		$\hat{P} = (n_i/n) \times 100$	$\hat{T} = (n_i/n) \times W$	SE $= \sqrt{\hat{P}\,(100 - \hat{P})/n}$	CV $= (SE/\hat{P}) \times 100$	P	D $= \hat{P} - P$
Green							39.2	
Yellow							44.2	
Blue							16.6	
Total							100.0	

Note: W : Each of your country's area is tentatively assigned to W.

8. Summary of the Area measuring

1. Topic:
2. Date:
3. Reporter:

4. Purpose

5. Activities

6. Output

7. Impressions and suggestions