A Study on Sampling Methods in Rice Yield Survey

Hong Nai Park*

The Ministry of Agriculture and Fishery had carried out a random sample survey on rice yield by the method of plot cutting since 1959 to obtain correct statistics on rice production. There are important problems connected with the determination of optimum size and shape of sample plot, variation of yield between plots and field and variation in the time needed for different types of plot cutting.

This kind of research has been performed in India by Dr. P. V. Sukhatme and Dr. V. G. Panse. Also the same kind of research has been done in Japan. These researches show that the optimum procedures are much different in Japan than in India. In Korea, the optimum procedures may well be different from either of these countries. Although the government is proceeding with the crop yield survey, it does not attempt to investigate plot problems but emphasize total rice yield estimation.

This particular research was conducted on a scale small enough that the author could personally supervise all of the work.

1. SELECTION OF SAMPLE

This survey was conducted in 10 *myons* of Hwasong-*gun* Kyonggi-*do*. The *myon* is an administrative unit similar to the township in the U.S.A., and the *gun* is similar to a county. There are 19 *myons* in Hwasong-*gun* and the 10 selected *myons* belong to the "paddy area" of the *gun*. These 10

* Professor of Statistics, College of Natural Science, Seoul National University.

myons are defined as the universe of this survey. It is customary in Korea that agricultural area is generally divided into paddy, paddy and dry field and dry field area. The paddy area refers to the *myons* in which the proportion of paddy area to the total cultivated area is more than 60 per cent, paddy and dry field area refers to the *myons* in which the proportion of paddy area to the total cultivated land is between 40 and 60 per cent and dry field area refers to the *myons* in which this proportion is less than 40 per cent. Each *myon* was considered as a stratum, and the sample was selected by using a stratified three-stage sampling method. The sampling plan consisted in selecting 2 villages from each *myon* and 2 fields from each sample village. In each sample field, two random points were selected. Each sampling unit was selected by using probability proportionate to the paddy field area. Based on each selected random point, 18 different plots were selected. Out of the 18 plots, 12 plots were of 4 sizes for each of 3 shapes of plots. The sizes were 0. 5 *pyong*, 1 *pyong*, 2 *phong* and 3 *pyong*. (A *pyong* is the conventional measure of area in Korea and is approximately 36 square feet in U.S. units.) The 3 shapes were the rectangle, triangle and circle. The other 6 plots of the 18 were defined by taking 25, 50 and 75 clumps of rice plants along a single row and by taking 10, 20, 30 clumps along three consecutive rows.

The reasons for adopting 4 types of plot shapes—rectangle, triangle, circle and row—are that the rectangle and circle methods have been the existing official methods in Korea and that the triangle and row methods have been investigated in other countries.

The major factor causing bias of an estimate in such yield surveys seems to have been the problem of faulty inclusion or exclusion of the stumps on the border line of the plot. The planting method in Korea is generally in a straight row, similar to the Japanese way of planting, except in mountainous areas. If the row cutting method is adopted, it would be expected that the chance to cause such bias will be very small. It was reported that the row cutting method was most appropriate in Japan according to their surveys. In the yield survey in India, a comparatively large size plot was adopted and in Japan, a small size of plot with 1 *pyong* circle and 1 *pyong* single-row cutting was adopted. In this survey, in consideration of the planting system, the small sizes of plots less than 3 *pyong* would seem to be appropriate.

2. SURVEY METHOD

The field work was carried out by 10 field workers who were students of the College of Agriculture, and 10 helpers from the villages. Before harvest time, the field workers made a list of farms in each sample village and of the paddy fields in these farms.

Selection of Plots

The selection of plots in the sample field was carried out by the following steps:

1) One corner of the sample field was randomly selected and through a point near this corner, two perpendicular lines were established in such a manner that the entire field was included between the lines. The lines were used as the axes of a rectangular coordinate system.

2) Two random points were selected in the sample field by selecting the coordinates from a table of random numbers.

3) Two perpendicular lines, parallel to the axes, were established through each random point, thus dividing the area surrounding each point into 4 quadrants.

4) The four shapes of plots (rectangle, triangle, circle and row) were randomly allocated to the four quadrants defined in this way.

5) As shown in the diagram 1, a rectangle of dimensions 9 *cheok* by 12 *cheok* was defined in one quadrant and sub-divided into 4 sub-plots of 0.5,

1, 2 and 3 *pyong*s. (A *cheok* is a Korean linear measure equivalent to 1.00582 feet, U.S., and a *pyong* is 36 square *cheok*.) The length of the vertical and horizontal lines of the plot is the same as shown in the table 1.

6) The triangle plot was sub-divided into 4 sub-plots of 0.5, 1, 2 and 3 *pyong*. The length of horizontal, vertical and diagonal lines of each plot is the same as shown in the table 1. In order to establish accurate plots, frames of plots of various sizes made of inelastic string were prepared. The frames were set on the paddy plant.

7) The marked point at 8 *cheok* and 8 *cheok* as shown diagram 1 was taken as the center of the circles and four concentric circles were drawn with the radius of 2.3, 3.4, 4.8 and 5.9 *cheok* so as to make circles of 0.5, 1, 2 and 3 *pyong*s in area.

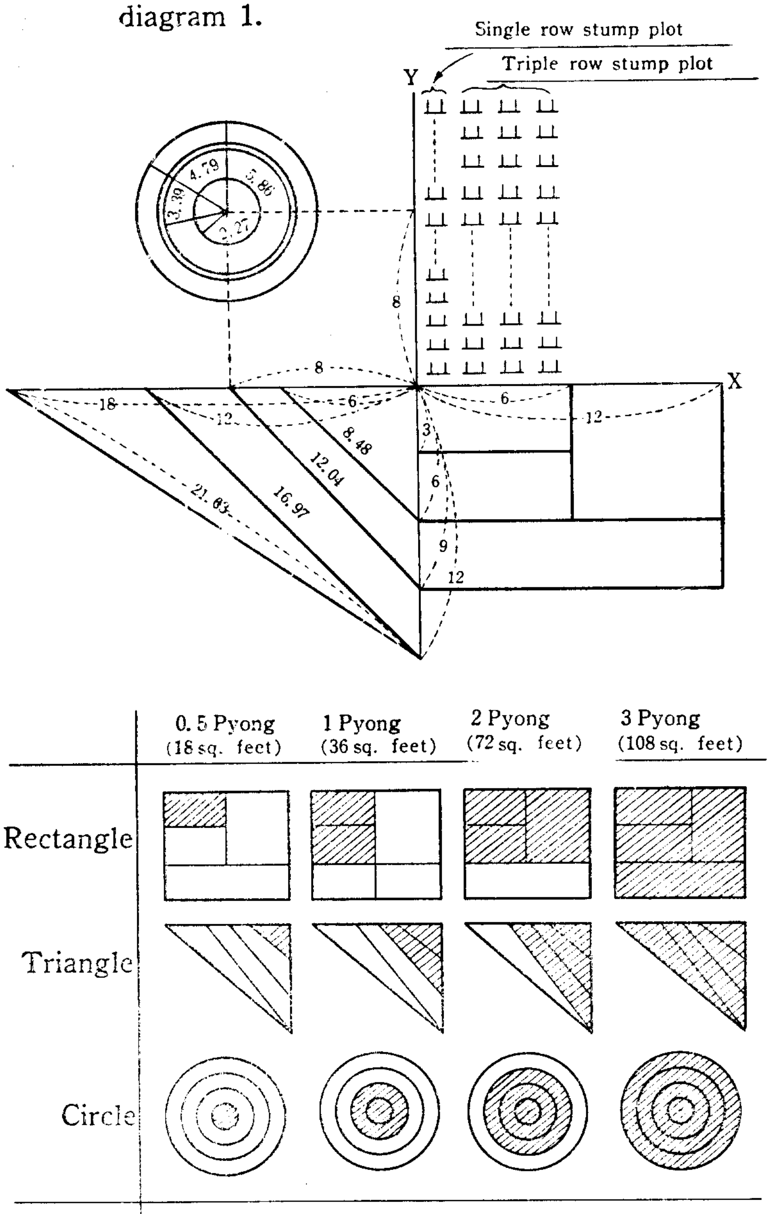
8) Finally, the row stumps at the last quadrant were marked.

The maximum number of stumps forming a single straight row in the case of planting is generally 30-25 stumps. Therefore, the length of the sub-plot was set at 25 stumps in the single row method. In the row method, the length and area of the sub-plot were measured. In the row method, the estimate of yield by cutting the rows forming single row seems to be more rational than the estimate of yield by cutting the stumps locating in the certain area of the plot as used in Japan.

Table 1. Side Length

Size (*pyong*) \ Shape	Rectangle		Triangle			Circle radius (*cheok*)
	Vert. (*cheok*)	Horiz. (*cheok*)	Vert. (*cheok*)	Horiz. (*cheok*)	Diag. (*cheok*)	
0.5	3	6	6	6	8.48	2.27
1	6	6	9	8	12.04	3.39
2	6	12	12	12	16.97	4.79
3	9	12	12	18	21.63	5.86

diagram 1.


Single row stump plot
Triple row stump plot
Y
X
4.79
3.86
3.39
2.27
8
8
18
12
6
6
12
3
6
9
12
8.48
12.04
16.97
21.63
0.5 Pyong (18 sq. feet)
1 Pyong (36 sq. feet)
2 Pyong (72 sq. feet)
3 Pyong (108 sq. feet)
Rectangle
Triangle
Circle

Plot Cutting Procedure

Plot cutting was carried out by the following steps:

1) A rectangle plot of 0.5 *pyong* was set on the paddy plant by using the string frame and poles mentioned above. The stumps clearly surrounded by the frame were cut and the paddy cut was arranged into bundles. The stumps located near the border of the frame were not cut at this time.

2) The stumps located near the border of the plot were then cut and each stump laid down on the root of the stump with the best possible attention.

3) The string frame set on the paddy was moved down to the surface of the ground and the rectangle plot of 0.5 *pyong* was established. In order to check the accuracy of the plot, the length of each side of the rectangle was measured again.

4) With the string frame moved down to the surface of the ground, it was judged whether or not the stumps located near the border of the plot were included in the plot.

The procedure is the same with the plot of other rectangle plots triangle and circle as the case of 0.5 *pyong* rectangle plot.

The procedure in the row method is relatively simple because counting and cutting of a given number of the stumps are enough. But in the row method, the interval between stumps and the length of the row in the sub-plots have to be measured.

After the cutting of the sample plot was completed, the sample field was completely cut. The cutting and threshing of the sample field were carried out by the farm having the cultivated land and the field workers observed these operations and measured the yield.

Threshing and Measurement

The procedure of threshing generally conducted by farms is to thresh with a threshing machine and eliminate dust by using winnower. These machines are good to thresh a large amount but are not appropriate to thresh a small amount. Handling a small amount of paddy, different instruments (*Chonchi* instead of threshing machine and *kee* instead of winnower) are more appropriate in order to obtain the accurate yield. In this survey, threshing machine and winnower were used for the paddy of the 40 sample fields but *kee* and *chonchi* were used for the paddy of the sample plots. In measuring yield, grain was not weighted but its volume was measured.

Survey of Area

After the cutting and threshing of sample plots and fields were completed the planted area of the sample field was surveyed. The record of the paddy field held by farms includes the area of the border and the harvested area is not known by the farm having harvested. The area of the field was measured by the triangulation method. The triangulation method is to divide the area into many triangles by using survey poles, to measure the lengths of the sides of those triangles, and to calculate the area of all the triangles.

3. RESULT

Estimate

Table 2 shows the average yield in *hop* per *pyong* and percentage bias for plots of different sizes, compared with average yield in the whole field.

There may exist a general tendency that the estimate of average yield decreases as the plot size increases except in the row stump cutting methods. Also it will be seen that the bulk of bias rate varies with narrow range. The bias rates occuring in the rectangle plot cutting run from 1.84 to 3.53%. In the triangle, from 1.84 to 3.07%. In the circle, from 2.07 to 3.45%. In the single row stump cutting method, from 0.69 to 1.34%. In the triple

Table 2. Average Yield of Rice in *hop* per *pyong* with Coefficient of Variation of the Estimate and Percentage Bias for Plots of Different Sizes

Size of plot (*pyong*)	No. of plots	Rectangle			Triangle			Circle		
		Average yield	c.v (%)	Bias (%)	Average yield	c.v (%)	Bias (%)	Average yield	c.v (%)	Bias (%)
0.5	80	27.02	2.98	3.53	26.90	3.53	26.90	2.93	307	27.00
1.0	80	26.74	3.15	2.45	26.85	3.09	2.87	27.13	2.74	3.95
2.0	80	26.58	3.07	1.84	26.76	3.06	2.54	26.64	3.08	2.07
3.0	80	26.59	2.86	1.88	26.58	3.24	1.84	26.6[illegible]	3.14	2.26

Av. size of plot (*pyong*)	No. of plots	Single row			Av. size of plot (*pyong*)	Triple row		
		Av. yield	c.v (%)	Bias (%)		Av. yield	c.v (%)	Bias (%)
0.410	80	26.28	2.25	0.69	0.498	24.24	3.16	−7.13
0.821	80	26.45	2.35	1.34	0.966	25.77	2.74	−1.26
1.238	80	26.36	2.43	0.99	1.431	26.17	2.56	0.26

* For the whole field cutting

Number of fields:	40 fields
Average size of field:	372.8 *pyong*
Average yield:	26.10 *hop*
c.v. of the estimate:	3.26%

row, from 0.26 to 7.13%.

The estimates of average yield in the single row stump plots are not consistently in the direction of either over-estimation or under-estimation. The estimate of average yield in the triple row stump method decreases as plot size decreases. It suggests that the bias of the estimated average yield does not occur always in the direction of over-estimation, but under-estimation. The estimate of average yield from the first triple row stump plot shows unusually large estimate of bias, which may be the result of some non-sampling error other than the bias due to plot size.

According to Table 2, the biases (if any) are relatively smaller in the row stump methods than in the others with the exception of the first triple row stump plot. This suggests that when the row stump method is adopted the bias resulting from the erroneous inclusion of plants near the boundary of the sample plot may not occur or be reduced, especially if the stump row is clearly visible. However, it might be that there is a possibility of some kind of bias even for the row stump method. Actually, there is some difficulty in deciding which stump belongs to the row in cases where rice stump row is not clear. Some such cases occur near the boundary of a field where, ususally, the plants located not so regularly as in the interior of the field. Also, some biases might have occured in measuring the length and width of a plot, especially for the triple-row stump plots near the boundary of a field.

It can be seen that both the largest value of bias rate 7. 13% and the smallest value 0. 26% appear in the triple row plot cutting. And the bias rates among the rectangular, triangular and circular plots are almost the same. Also that the small bias rate in average appears in the single row plots. To compare the precision of the estimate for different plots, table 2 shows the relation between the size of plot and the coefficient of variation of estimates of average yield.

It will be seen that there is no general tendency such as coefficient of variation of estimate of average yield decreases as plot size increases. As a whole, the c.v value varies with small range. The range of c.v is from 2. 86 to 3. 15% in the rectangle, from 2. 93 to 3. 24% in the triangle, from 2. 74 to 3. 14% in the circle, from 2. 25 to 2. 43% in the single row stump, from 2. 56 to 3. 16% in the triple row stump. The tendency of c.v appears irregularly. In the Rectangle shape of plot, the c.v value shows almost constant except the large value of 1 *pyong* plot. In the triangle, it shows also constant except the large value of 3 *pyong* plot. In the circle, it appears irregularly. In the single row stump, it increases as plot size increases. In the triple row stump, it decreases as plot size increases. In the triple row stump, it decreasesas polt size increases. As a whole, c.v value runs with small range from 3.24% occuring in the 3 *pyong* triangle plot to 2.25% occuring in the first single row stump plot. Such a small range of c.v indicates that there is no big difference of precision of the estimate among the different plots. It is remarkable that the c.v value 3.26% of the whole field cutting is relatively large. It is considered that the large value of c.v for whole field due to the different threshing process among farms. In fact, there are many different ways among farms concerning with threshing tools, threshing place, amounts of the grain loss, and carefulness in threshing work. Even it is so, still it may be said that average minimum c.v value appears in the single row plot cutting and furthermore the first single row plot has maximum precision out of 18 plots.

Test of Bias

To test of bias, comparison of estimates of average yield for different size of plots with whole field was made as shown table 3.

Table 3. **Test of Bias**

Size of plot	Rectangle			Triangle			Circle		
	Diff.	S.E	t	Diff.	S.E	t	Diff.	S.E	t
0.5	0.92	0.900	1.02	0.80	1.01	0.79	0.90	0.99	0.91
1.0	0.64	1.080	0.59	0.75	1.02	0.74	1.03	0.96	1.07
2.0	0.48	0.926	0.52	0.66	1.01	0.65	0.54	1.04	0.52
3.0	0.49	0.899	0.55	0.48	0.98	0.49	0.59	1.05	0.56

Av. size of plot	Single row			Av. size of plot	Triple row		
	Diff.	S.E	t		Diff.	S.E	t
0.413	0.18	0.92	0.20	0.498	−1.86	0.77	−2.41
0.821	0.35	1.04	0.34	0.966	−0.33	0.74	−0.44
1.238	0.26	0.96	0.27	1.431	0.07	0.83	0.09

The formula of the variance of the bias S_B^2 is defined as follows:

$$B=E\left(\frac{x}{y}-\frac{u}{w}\right)=\text{expectation of the bias}$$

$$S_B^2=S_{x/y}^2+S_{u/w}^2-2S_{x/y\cdot u/w}$$

$$=\left(E\frac{x}{y}\right)^2 V_x^2+V_y^2-2V_{xy}+\left(E\frac{u}{w}\right)^2+(V_u^2+V_w^2-2V_{uw})$$

$$-2\left(E\frac{x}{y}\right)\left(E\frac{u}{w}\right)(V_{xu}-V_{xw}-V_{uy}+V_{yw})$$

where $x=\sum x_{hijk}$=estimate of total yield based on a given plot

$y=\sum y_{hijk}$=estimate of total area based on a given plot

$u=\sum u_{hij}$ =estimate of total yield based on the whole field

$w=\sum w_{hij}$=estimate of total area based on the whole field

The result of the comparison test shows that none of the biases are significant. All the observed t values are less than 1 except for three plots—the first triple row stump plot, 1 *pyong* circle plot, and 0.5 *pyong* rectangle plot. The largest t value 2.409 occurred in the first triple row stump plot.

As shown in the table 4A, the result of yield survey in India indicates significant differences of bias for small sizes of plots. However, the result of this survey shows non-significant differences for small sizes of plots such as less than 3 *pyong*.

As shown in table 4B, the result in Japan shows significant differences of bias between 1 *pyong* circle plot and 1 *pyong* single row stump plot. However, there is no significant difference between circle plot and single row plot in this survey.

In summary, as mentioned previous paragraph, coefficient of variation shows the precisions for all kinds of plots are approximately same. Also it was observed that all kinds of plot cutting methods are free from bias. That is, the t test adopted here is weak in discriminating the significant differences among 18 plots.

Therefore, to decide an optimum plot among those 18 kinds of plots, further analysis may be required.

Table 4A. **Relation between Plot Size and Biases in Yield**

Size and shape of plot	Plot area		No. of plot	Average yield lbs/acre	Standard error lbs/acre	Percentage over est.
	sq. ft.	*pyong*				
whole field	—		108	1939.2	107.3	
50×20(Link)[2]	43560	12.1	108	1954.1	105.0	.8
3′ ⊙	28.29	.79	216	2025.9	125.8	4.5
2′ ⊙	12.57	.35	216	2113.2	129.1	9.0
5′ △	11.12	.31	216	2433.4	161.7	25.5

P.V. Sukhatme; "The problem of plot size in large scale yield survey"

Table 4B. **Bias in Circle Cutting Method**

Prefecture	No. of plot①	Weight of grain circular cutting②	Per "tsubo" stump cutting③	Bias ④ ②—③	Bias rate ④/③	Standard error	t-value
Yamagata	80	596	587	8.83	1.5	2.57	3.4
Nara	100	473	470	2.67	0.6	1.60	1.7
Tottoyi	100	506	501	5.24	1.0	1.82	2.9
Kochi	97	434	426	7.52	1.8	2.27	3.3
Chiba	377	498	492	5.91	1.2	1.03	5.7

Note; *Crop Survey in Japan*, M.A.F.

Regression Analysis

The result that none of the biases are significant means theoretically that the power of the above test is not great enough to reveal any biases that may exist. The following regression analysis provides more a powerful test than the individual comparison test to investigate the biases. The assumptions in this regression analysis are linearity of regression and homoscedasticity. Now, the hypothesis to be test is that regression cofficient is equal to zero.

Table 5. Regression Analysis

		Rectangle	Triangle	Circle	Single row	Triple row
Est'd regression coefficient	b	−0.1576	−0.1251	−0.1742	0.0016	−0.0482
Est'd standard error	S_b	0.0669	0.115	0.0809	0.00353	0.00625
	$t=b/S_b$	−2.356	−1.088	−2.153	0.4535	−7.712

As the result of table 5, the regression coefficient b's of the rectangle, circle and triple row stump plots are significant different from zero at 5% significant level even though they are non-significant in the result of the individual comparison test as seen table 3. However, the biases of triangle plots and the single row stump plot are non-significant. That is, regression analysis provides very useful information in determining the optimum plot and from this result, it is concluded that the triangle plots cutting and single row plot cutting are efficient methods. Connecting with the result of this analysis, the magnitude of bias should be discussed. The allowable maximum bias rate may be thought 3%. From Table 2, only triangle 0.5 *pyong* plot is not satisfied this condition among all triangle plots and single row plots.

Finally, to decide an optimum plot, further analysis may be required. For this purpose, cost function should be considered which is equally efficient as precision so that to lead to choice of smallest, most easily defined plot.

Table 6. Average Time Needed for Different Size of Plot Cutting

Size of plot	Rectangle	Triangle	Circle	Plot area	Single row	Triple row
0.5	20	20	13	1	14	13
1	34	33	19	2	20	19
2	60	61	30	3	28	25
3	74	78	37			

The time decreases as plot size increases. Circle plot cutting, single and triple row cutting needed less time for the operation to cut off approximately same area of plot than rectangle, triangle cutting.

In summary, regarding to bias from table 5, triangle plot and single row plot are independent from plot sizes. Regarding cost, circle, single row and triple row have minimum cost. And, regarding to c.v of average yield, single row plot has minimum value. In conclusion, it is said the first single row stump plot provides maximum precision, allowable bias and minimum cost among all kinds of plots and it is most appropriate, easily defind plot.

REFERENCES

[1] Sukhatme, P. V., "The Program of Plot Size in Large Scale Yield Surveys," *Journal of American Statistical Association,* Vol. 42 (1947).

[2] Sukhatme, P. V. and Panse, V. G., *Statistical Methods for Agriculture Workers,* New Dehli: Indian Council of Agriculture Research, 1957.