PACIFIC SOUTHWEST
Forest and Range
Experiment Station

FOREST SERVICE
U.S.DEPARTMENT OF AGRICULTURE
P.O. BOX 245, BERKELEY, CALIFORNIA 94701

NUMBER OF PINS IN TWO-STAGE STRATIFIED SAMPLING FOR ESTIMATING HERBAGE YIELD

William G. O'Regan C. Eugene Conrad

USDA Forest Service
Research Note PSW-300
1975

O'Regan, William G., and C. Eugene Conrad.
1975. Number of pins in two-stage stratified sampling for estimating herbage yield. USDA Forest Serv. Res. Note PSW-300, 4 p. Pacific Southwest Forest and Range Exp. Stn., Berkeley, Calif.

In a two-stage stratified procedure for sampling herbage yield, plots are stratified by a pin frame in stage one, and clipped. In stage two, clippings from selected plots are sorted, dried, and weighed. Sample size and distribution of plots between the two stages are determined by equations. A way to compute the effect of number of pins on the variance of estimated herbage yield has been developed. It was tested by using a sampling budget and a simulated population of *Bromus mollis* (soft chess), and worked reasonably well with other annual plant populations. But application to other plant communities should be made with prudence.

Oxford: 268.5–015.5.

Retrieval Terms: range sampling; herbage yield; statistical samples; two-stage sampling; San Joaquin Experimental Range.

Conrad and O'Regan[1] recently reported that two-stage stratified sampling using a pin frame at the first stage can result in a 15 percent reduction in the variance of estimated mean weight compared with simple random sampling with the same budget. In that study a simulated population of *Bromus mollis* (soft chess) and a budget of 6000 man-minutes were used in the example.

In two-stage stratified sampling with a pin frame, plots are chosen at random for the first-stage sampling. At each plot the frame is mounted, the pins lowered, and the number of pins touching soft chess (or bare ground near soft chess) is recorded. These "hits" (from 0 to 8 in the example) are used to form strata. All plots are clipped and bagged at the first stage. At the second stage, herbage from plots chosen at random in each stratum is sorted, dried, and weighed.

Using the appropriate formulae (given below) an estimated mean and estimated variance of the mean are calculated.

The question naturally arises, what is the cost effective number of pins to be used in the frame?

From the earlier study, we know the variances at zero and eight pins. In this note, using the same simulated soft chess population, we determine minimum variances for 1 through 7 pins.

In reducing from p + 1 to p pins per frame, the pin to be removed is chosen at random. The following definitions were used in developing this procedure for computing variance:

p = number of pins after reduction by one pin, $0 \leq p \leq 7$.

h = number of "hits," stratum name (number), $0 \leq h \leq p$.

Π_{ph} = proportion of element in h^{th} stratum when there are p pins per frame.

θ_{ph} = the population mean weight per element in the h^{th} stratum when there are p pins per frame.

σ_{ph}^2 = the variance of weight per element in stratum h when there are p pins per frame.

The following relationships are implied. Define

$$q = p + 1$$
$$i = h + 1$$

Then,

$$\Pi_{ph} = \frac{q-h}{q}\Pi_{qh} + \frac{i}{q}\Pi_{qi}$$

$$\theta_{ph} = \frac{q-h}{q\Pi_{ph}}\Pi_{qh}\theta_{qh} + \frac{i}{q\Pi_{ph}}\Pi_{qi}\theta_{qi}$$

$$\sigma^2_{ph} = \frac{q-h}{q\Pi_{ph}}\Pi_{qh}[\sigma^2_{qh} + (\theta_{qh} - \theta_{ph})^2] + \frac{i}{q\Pi_{ph}}\Pi_{qi}[\sigma^2_{qi} + (\theta_{qi} - \theta_{ph})^2]$$

Sequential application of these formulae to the simulated stratified population of soft chess (table 1) produced the eight simulated stratified populations for $0 \leq p \leq 7$ (table 2).

For p pins (p + 1 strata) the relationships

$$\Sigma^p_{h=0}\Pi_{ph} \equiv 1$$

$$\theta_p = \Sigma^p_{h=0}\Pi_{ph}\theta_{ph}$$

$$V_{pM} = \Sigma^p_{h=0}\Pi_{ph}\sigma_{ph}^2$$

$$V_{pN} = \Sigma^p_{h=0}\Pi_{ph}(\theta_{ph} - \theta_p)^2$$

$$\sigma_p^2 = V_{pM} + V_{pN}$$

continue to hold, and $\theta_p \equiv \theta$ and $\sigma^2 \equiv \sigma^2$, as they must.

Other definitions are:

A = number (large) of elements (plots) in the population

N_p = number of plots chosen at random from A

N_{ph} = observed number of plots in h^{th} stratum, p = number of pins

M_p = number of plots chosen from N_p for second stage sampling

M_{ph} = number of plots chosen at random from N_{ph} for second stage sampling, p pins

$$M_{ph} = \frac{N_{ph}}{N_p}M_p$$

W_{phi} = weight of the given species on i^{th} element, h^{th} stratum, p pins

C_N = cost per plot for first stage sampling[2]

C_M = cost per plot for second stage sampling

$B = M_pC_M + N_pC_N$ = budget, constant over all p

Travel time between plots is so small that it is assumed to be zero.

An unbiased estimator for θ is given by

$$\hat{\theta}_p = \frac{1}{N_p}\Sigma^p_{h=0}\frac{N_{ph}}{M_{ph}}\Sigma^{M_{ph}}_{i=1}W_{phi}$$

with variance

$$V(\hat{\theta}_p) = \frac{V_{pM}}{M_p} + \frac{V_{pN}}{N_p}$$

which is a minimum when[1]

$$M_p = \frac{B\sqrt{V_{pM}/C_M}}{(\sqrt{C_NV_{pN}} + \sqrt{C_MV_{pM}})}$$

$$N_p = \frac{B\sqrt{V_{pN}/C_N}}{(\sqrt{C_NV_{pN}} + \sqrt{C_MV_{pM}})}$$

which implies

$$V(\hat{\theta}_p) = [\sqrt{V_{pN}C_N} + \sqrt{V_{pM}C_M}]^2/B$$

Optimum allocations of a budget of 6000 man-minutes for $0 \leq p \leq 8$ were developed, and the related M_p and N_p as well as the resulting minimized variance of $\hat{\theta}_p$ were computed (table 3). After three pins per frame—and certainly after five pins per frame—the variance of the estimator shows only slight reduction.

In practice, M_{ph} and N_p will be rounded to integers and $M_{ph} \geq 2$. In our experience, however, these constraints would not seriously affect the answer.

APPLICATION

Although the results presented here are peculiar to the situation described in the example, we have found that the procedure applies reasonably well to other populations of annual plants at the U.S. Forest Service's San Joaquin Experimental Range, in central California. We suggest however, that application to other plant communities be made with prudence. If the method at stage one is not based on a pin frame or is not analogous to a pin-frame method, the results of this note may not be applicable.

NOTES

[1]Conrad, C. Eugene, and William G. O'Regan. 1973. Two-stage stratified sampling to estimate herbage yield. USDA Forest Service Res. Note PSW-278, 5 p. Pacific Southwest Forest and Range Exp. Stn. Berkeley, California.

[2]For C_N throughout this paper, please read C_{pN}.

Table 1–*Simulated population values for* Bromus mollis *obtained from 310 plots measured by using a two-stage stratified sampling procedure at the San Joaquin Experimental Range, California*

Parameter	Strata (h)									
	0	1	2	3	4	5	6	7	8	Total
Π_{8h}[1]	0.289	0.212	0.161	0.121	0.088	0.061	0.039	0.021	0.008	1.000
θ_{8h}[2]	1.350	3.375	4.875	8.170	9.400	9.912	12.693	13.385	15.900	5.214
σ_{8h}[3]	3.551	11.886	14.758	33.310	32.199	31.214	54.442	87.682	40.330	32.446

[1]Π_{8h} = proportion of the population in the h^{th} stratum.
[2]θ_{8h} = actual mean weight in grams/element in the h^{th} stratum.
[3]σ_{8h} = variance of the h^{th} stratum.

Table 2–*Population parameters for nine stratified populations based on the simulated* Bromus mollis *population of table 1; number of strata (p + 1, where p is the number of pins per frame), mean weight (θ_p) within stratum variance (V_{pM}), between stratum variance (V_{pN}), and population variance (σ_p^2).*

p	θ_p	V_{pM}	V_{pN}	σ_p^2
0	5.214	32.445	0	32.445
1	5.214	28.227	4.219	32.446
2	5.214	25.532	6.914	32.446
3	5.214	23.659	8.786	32.445
4	5.214	22.280	10.167	32.447
5	5.213	21.212	11.219	32.431
6	5.213	20.371	12.068	32.439
7	5.213	19.644	12.790	32.434
8	5.214	18.956	13.490	32.446

Table 3–*Number of pins per frame (p), unit costs at first and second stage (C_N,C_M), optimum number of plots at first and second stage (N_p,M_p) and resulting minimum variance (min $V(\hat{\theta})$) for the simulated stratified populations of* Bromus mollis *of table 2.*

p	C_N	C_M	Opt. N_p	Opt. M_p	min V ($\hat{\theta}_p$)
0	0	68	0	88	0.368
1	13/12	68	258	84	.352
2	18/12	68	287	82	.336
3	23/12	68	291	80	.326
4	28/12	68	286	78	.320
5	33/12	68	278	77	.316
6	38/12	68	270	76	.314
7	43/12	68	262	74	.313
8	48/12	68	255	73	.312

The Authors

C. EUGENE CONRAD is a research botanist assigned to the Station's work unit on flood and sediment reduction in steep unstable brushlands, with headquarters at Glendora, California. He is a graduate of Oregon State College (B.S. degree in general agriculture, 1956; M.S. degree in range management, 1959). He joined the Forest Service, and the Station staff in 1961. **WILLIAM G. O'REGAN** is the Station biometrician, with headquarters in Berkeley. He earned B.S. (in agricultural economics, 1949) and Ph.D. (1962) degrees at the University of California, Berkeley. He has been a member of the Station staff since 1957.

GPO 689–233/4423