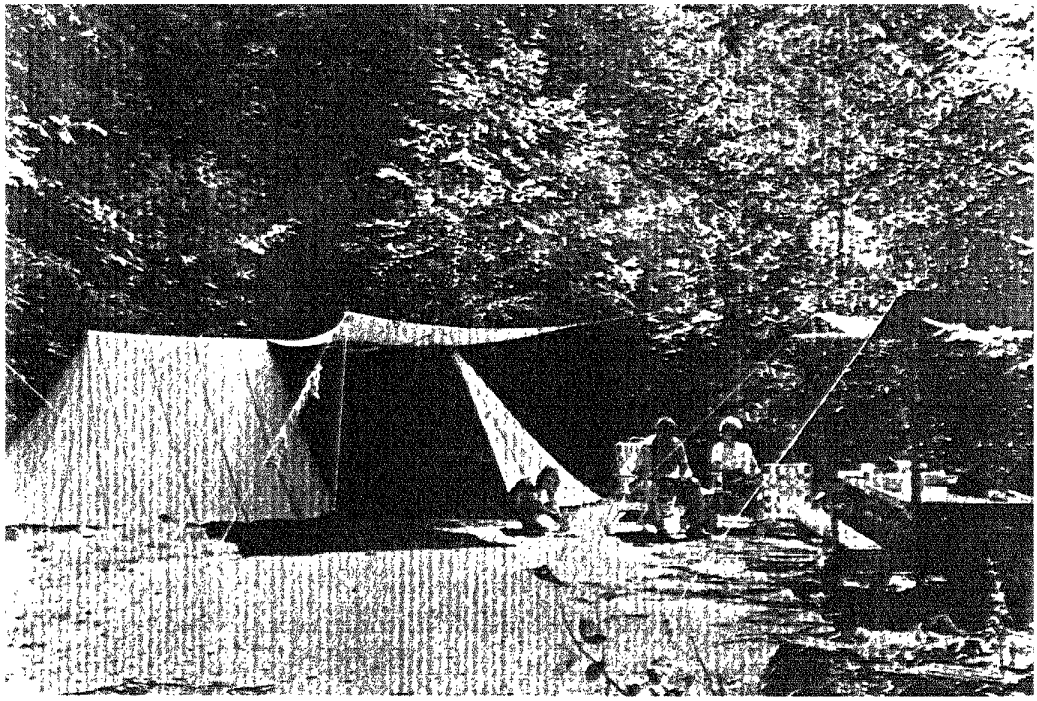
Estimating NUMBERS OF CAMPERS on Unsupervised Campgrounds

by J. Alan Wagar

U. S. FOREST SERVICE RESEARCH PAPER NE-18
1964
NORTHEASTERN FOREST EXPERIMENT STATION, UPPER DARBY, PA.
FOREST SERVICE, U. S. DEPARTMENT OF AGRICULTURE
RALPH W. MARQUIS, DIRECTOR



A typical family camping scene on Waterville campground, one of the areas included in a study of campsite use in the White Mountain National Forest.

The Author —

J. ALAN WAGAR, *a research forester specializing in forest recreation, served at the Northeastern Forest Experiment Station's research unit at Warren, Pa., in 1959-62, when the study reported here was made. Since then he has transferred to the Intermountain Forest and Range Experiment Station to become leader of a cooperative recreation research unit at Utah State University, Logan, Utah. He has a Bachelor's degree in forestry from the University of Washington, and Master's and Doctoral degrees in forestry from the University of Michigan.*

Estimating
NUMBERS OF CAMPERS
on Unsupervised Campgrounds

■

How Many Campers?

ANY agency that provides camping facilities needs an accurate record of campers as an aid to efficient planning and administration. However, complete counts of campers are expensive, especially at small unsupervised campgrounds. And although many of the methods now being used for estimating campers are of doubtful accuracy, sampling methods may offer a reliable solution. Sampling procedures that incorporate regression or ratio techniques look especially promising. When these techniques are used, area-wide fluctuations in campground attendance do not necessarily increase sampling error.

Within a large sampling area—for example, in a National Forest—the numbers of campers on many campgrounds may rise and fall in uniform response to such diverse factors as weather, day of the week, vacation periods, insect activity, and fishing seasons. Where this common pattern exists, the number of campers on a sample group of campgrounds may be related to campers on all campgrounds by regression and ratio techniques.

To test these techniques, a study was made on the White Mountain National Forest in New Hampshire from July 10 to September 10, 1961. Campers were counted each evening, according to

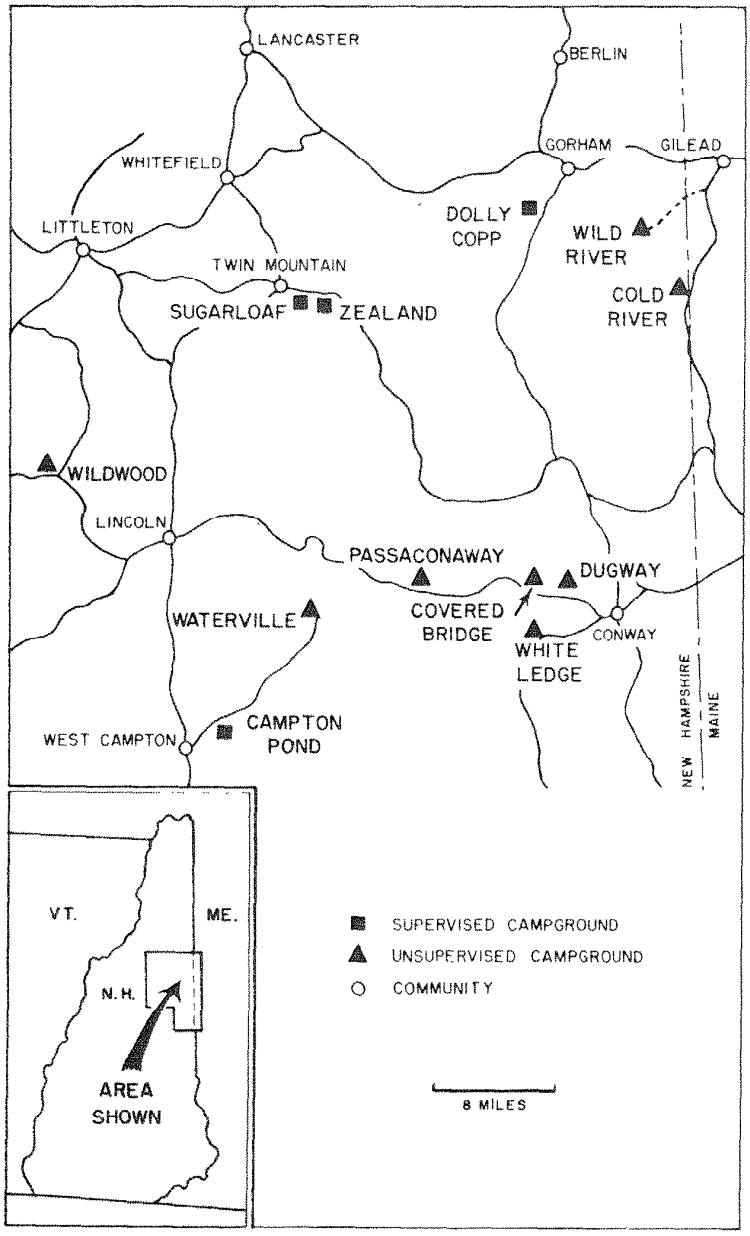


Figure 1.—Locations of campgrounds included in this study, on the White Mountain National Forest, New Hampshire.

a random schedule, on one or more of eight unsupervised campgrounds (fig. 1) grouped into sampling units as follows:

Sampling Unit A.—Wild River and Cold River Campgrounds—21 campsites.

Sampling Unit B.—White Ledge Campground—39 campsites.

Sampling Unit C.—Wildwood and Waterville Campgrounds—48 campsites.

Sampling Unit D.—Dugway, Covered Bridge, and Passaconaway Campgrounds—60 campsites.

These unsupervised campgrounds included a wide range of conditions. At one extreme, Wild River Campground had 7 campsites, was accessible only by an unsurfaced and unmaintained road, and was 10 miles from the nearest village—a community of 50. By contrast, White Ledge Campground had 39 campsites, was adjacent to a paved main route, and was only 5 miles from the nearest town—a community of 1,200. Such dissimilarity tends to make sampling difficult.

From early summer until September 4, as standard administrative procedure, campers were already being registered on four supervised campgrounds—Dolly Copp, Campton Pond, Sugarloaf, and Zealand.

Relating Use of Unsupervised Campgrounds to Supervised Campgrounds

Where attendance at supervised campgrounds is already being recorded, attendance at unsupervised campgrounds can be estimated with few additional counts if unsupervised camping follows the same patterns as supervised camping. To provide data for relating unsupervised use to supervised use, campers were counted simultaneously on all 8 unsupervised campgrounds on 18 randomly selected evenings during the last hour before darkness, from July 10 to September 4. Sixteen of these counts coincided with records of supervised camping. Total counts of unsupervised camping, and the corresponding records of supervised camping,

(table 1) were used to compute the following regression (fig. 2):

$$Y = 51.432 + .309\ X \qquad S_{y.x} = 52.42$$

where:

X = the attendance at supervised campgrounds for one night.
Y = the estimate of unsupervised campers for the same night.
$S_{y.x}$ = the standard error of the estimate.

The regression estimate of unsupervised camping for the 57 nights between July 10 and September 4 is:

$$\Sigma Y = \Sigma(51.432 + .309\ X)$$
$$= 57\ (51.432) + .309\ (\Sigma X)$$

From table 1, $\Sigma X = 77,848$, and the estimate for the period is computed as 26,986 camper nights for the 8 unsupervised campgrounds.

The original plan was to determine a variance for each night's estimate of campers. Confidence intervals for the entire period were then to be computed from the sum of these nightly variances. However, this procedure was abandoned because nightly variances are not independent and therefore, according to statistical theory, are not additive. C. Allen Bickford (personal communication) demonstrated that $\sigma^2_{\Sigma Y} = N^2\sigma^2_{\bar{Y}}$ and recommended that variance for the entire season be computed by the formula:

$$S^2_{\Sigma Y} = N(N - n)\ S^2_{y.x} \left\{ \frac{1}{n} + \frac{(\bar{X}_N - \bar{X}_n)^2}{\Sigma x^2} \right\}$$

where:

N = the number of nights in the entire period (57 in this case).
n = the number of nights on which total counts were made (16 in this case).
Σx^2 = the sum of the squared deviations from the mean for the data collected on the n total-count nights.

This is the formula (7.13) for $S^2_{\bar{Y}}$ given by Cochran,[1] multiplied by N^2 and the finite population correction $(1 - n/N)$. Con-

[1] Cochran, William G. SAMPLING TECHNIQUES. John Wiley & Sons, Inc., New York, 320 pp., 1953.

fidence intervals based on the variance given by this formula are:

$$\text{Regression estimate} \pm t \sqrt{\text{variance}}$$

= 26,986 ± 1,374 camper nights at 95 percent confidence (error of 5.1 percent).

= 26,986 ± 1,907 camper nights at 99 percent confidence (error of 7.1 percent).

For the data plotted in figure 2, ratio estimation appears as reasonable a procedure as regression estimation. Following Cochran (1953), R = y/x where y and x are sums of sample data—in this case sums of unsupervised and supervised campers, respectively, for the 16 total-count nights (table 1). Thus, R = 7,796/22,574 = 0.34535. The ratio estimate of unsupervised campers for July 10 to September 4 equals R times the total number of supervised campers for the period:

= 0.34535 (77,848) = 26,885 camper nights.

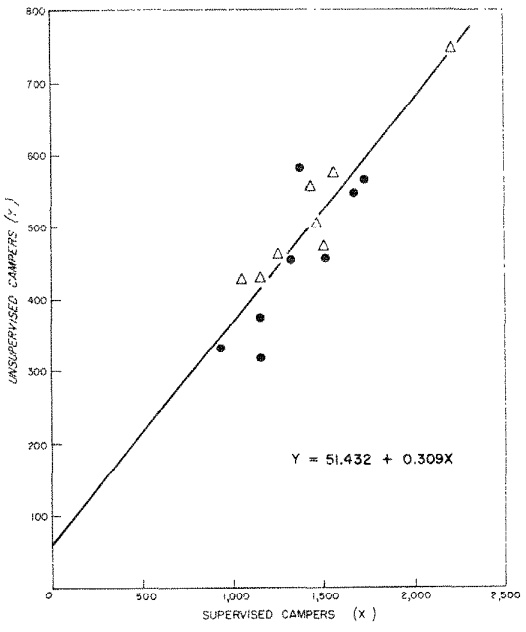


Figure 2.—Regression line for estimating unsupervised campers from camper registration on supervised campgrounds. Triangles represent data collected on weekend nights; round dots indicate data collected on week nights.

Table 1.—Summary of use-count data for campgrounds on the White Mountain National Forest, July 10 to September 10, 1961, in number of campers

(Underlined figures are values for nights when all units were counted)

Date	Unsupervised campgrounds, by sampling units—					Supervised campgrounds
	A	B	C	D	Total	
July						
10	..	..	..	179	..	984
11	17	..	..	..	..	1,102
12	..	..	84	..	..	1,114
13	31	124	80	97	332	930
14	..	120	..	..	..	872
15	6	135	120	166	427	1,056
16	..	..	62	..	..	985
17	28	153	48	144	373	1,151
18	..	178	..	..	..	1,263
19	..	..	70	..	..	1,199
20	..	183	..	..	..	1,296
21	33	..	..	..	..	1,293
22	..	..	..	263	..	1,544
23	31	136	91	215	473	1,506
24	24	..	..	..	..	1,479
25	..	..	..	228	..	1,493
26	..	171	..	..	..	1,553
27	39	152	55	211	457	1,515
28	..	..	77	..	..	1,453
29	47	144	118	264	573	1,569
30	28	..	..	..	..	1,448
31	..	..	..	207	..	1,546
August						
1	41	158	138	212	549	1,666
2	..	..	133	..	..	1,677
3	..	138	..	..	..	1,486
4	48	..	..	..	..	1,494
5	..	..	..	291	..	1,520
6	42	158	112	247	559	1,434
7	..	..	..	264	..	1,585
8	29	..	..	..	..	1,602
9	66	153	127	217	563	1,729
10	..	..	182	..	..	1,634
11	..	173	..	..	..	1,558
12	..	..	..	294	..	1,636
13	38	126	124	219	507	1,469
14	66	157	127	232	582	1,372
15	..	154	..	..	..	1,345

CONTINUED

Table 1.—(continued).

Date	Unsupervised campgrounds, by sampling units—					Supervised campgrounds
	A	B	C	D	Total	
16	..	..	..	247	..	1,388
17	..	..	112	..	..	1,236
18	..	..	137	..	..	1,348
19	55	..	..	..	..	1,353
20	38	116	64	208	426	1,167
21	..	62	..	..	..	1,162
22	..	..	..	237	..	1,206
23	28	..	..	..	..	1,262
24	..	..	67	..	..	1,307
25	28	77	122	227	454	1,312
26	22	88	120	230	460	1,283
27	17	..	..	..	..	1,206
28	..	..	79	..	..	1,191
29	27	49	73	166	315	1,191
30	..	66	..	..	..	1,254
31	37	..	..	..	..	1,137
September						
1	..	125	..	..	..	1,879
2	65	194	182	305	746	2,224
3	..	..	..	283	..	1,846
4	..	49	..	..	..	338
5	..	..	17	..	..	..
6	6	25	19	33	83	..
7	7	..	..	..	..	..
8	..	..	..	54	..	..
9	..	..	17	..	..	..
10	7	14	7	19	47	..

Sums and averages	Unsupervised campgrounds, by sampling units—						Supervised campgrounds
	A	B	C	D	Total		
All nights	..	..	..	..	..	..	77,848 (57)[1]
Total count	628	2,159	1,727	3,412	7,926	7,796	22,574
nights	(18)	(18)	(18)	(18)	(18)	(16)	(16)
Other	323	1,419	1,037	2,547	..	..	..
nights	(11)	(11)	(12)	(11)	..	..	..
Aver.	32.793	123.379	92.133	205.483	..	..	..
per night	(29)	(29)	(30)	(29)	..	..	..

[1] Numbers in parentheses represent the number of nights of data included in the various sums and averages.

Variance is estimated as:

$$\frac{N(N-n)}{n} \cdot \frac{\sum_{i=1}^{n}(y_i - Rx_i)^2}{(n-1)} = 391{,}233.27$$

where:

N = number of nights in the entire period (57).
n = number of total-count nights in the period (16).
y_i = number of unsupervised campers on night i.
x_i = number of supervised campers on night i.

Confidence intervals based on the variance given by this formula are:

Ratio estimate $\pm$ t $\sqrt{\text{variance}}$

= 26,885 $\pm$ 1,333 camper nights at 95 percent confidence (error of 4.9 percent).

= 26,885 $\pm$ 1,843 camper nights at 99 percent confidence (error of 6.9 percent).

Thus, where they are applicable, ratio methods give estimates and confidence intervals comparable to those obtained from regression.

Estimating Number of Campers from a Sample

Where there is no registration of campers, camper numbers may be estimated from counts on a sample of campgrounds. To test the effectiveness of various samples, the 18 total counts of unsupervised camping (the underlined figures in table 1) were used to develop 4 regressions. These relate use on each sampling unit to total use on all unsupervised campgrounds and are as follows:

$Y_A = 181.634 + 7.415\ X_A$ $\quad S_{y.x} = 102.47$
$Y_B = 83.331 + 2.976\ X_B$ $\quad S_{y.x} = 86.24$
$Y_C = 113.374 + 3.408\ X_C$ $\quad S_{y.x} = 77.91$
$Y_D = 30.571 + 2.162\ X_D$ $\quad S_{y.x} = 52.94$

where:

X_A, X_B, X_C, and X_D = one-night counts of campers on Sampling Units A, B, C, and D respectively.

Y_A, Y_B, Y_C, and Y_D = estimates of total use for one night.

These equations and the data on which they are based are graphed in figures 3, 4, 5, and 6. Because of the large Y-intercepts in some of these graphs, ratio methods were not considered appropriate for estimates based on these sampling units.

On each night not devoted to a total count, use was counted on whichever sampling unit was indicated by a randomly determined schedule. These counts are the numbers not underlined in table 1. By using each of these counts in the appropriate regression, total use could be estimated for any night on which one of the sampling units had been counted.

The relationship $\Sigma Y = N.a + b\ (\Sigma X)$ is again applicable, but N now equals the number of nights, excluding total-count nights, on which a given sampling unit was counted. Using the sums of campers for all nights when total counts were not made (table 1), the use estimate for all but total-count nights would be

$$\begin{aligned} Y_A &= 11(181.634) + 7.415(\ 323) = 4393.019 \\ Y_B &= 11(83.331) + 2.976(1419) = 5139.585 \\ Y_C &= 12(113.374) + 3.408(1037) = 4894.584 \\ Y_D &= 11(30.571) + 2.162(2547) = 5842.895 \\ &\text{Total camper nights} = \overline{20{,}270} \end{aligned}$$

When use for the 18 total-count nights is added, the estimate of unsupervised camping for the period July 10 to September 10 is: 20,270 + 7,926 = 28,196 camper nights. This is an unbiased estimate of total use. However, confidence intervals must be rather awkwardly handled because the variances from regressions Y_A, Y_B, Y_C, and Y_D are not independent and cannot be added as was planned originally.

During analysis of data from this study, it became obvious that the same sampling unit should have been counted every night, and that using a different sampling unit each night was not good procedure. Because four sampling units were used, only approximate confidence limits can be placed on the total estimate of campers for the season. These limits are an approximation of

the confidence limits that would have resulted if each sampling unit had been the one on which nightly counts were made. This indicates the precision that might be expected for estimates based on sampling units of various sizes. The formula

$$S^2_{\bar{Y}} = N(N - n)S^2_{y.x}\left(\frac{1}{n} + \frac{(\overline{X}_N - \overline{X}_n)^2}{\Sigma x^2}\right)$$

would again be applicable if $\overline{X}_N$ were known. But campers were counted on each sampling unit only 29 or 30 times rather than on all N nights during the sampling period (N = 63 in this case). However, an assumption that is approximately correct allows confidence limits to be computed for each sampling unit as if it had been counted all 63 nights.

It has been assumed that the means of the 29 or 30 counts made on each sampling unit are equal to the means that would have been obtained if use had been counted on the same unit all 63 nights. In the formula given above, the term

$$\frac{(\overline{X}_N - \overline{X}_n)^2}{\Sigma x^2}$$

contributes a very minor part to the value of $S^2_{\bar{Y}}$ and inaccuracies from using the wrong value of $\overline{X}_N$ can be expected to be very small. By using an estimate of $\overline{X}_N$ (table 1) we can identify the approximate error that would have occurred if all sampling had been on one unit.

Using the above formula, if campers had been counted on all 63 nights of the season on a single sampling unit, approximate confidence limits would be as follows:

Sampling unit	*Campsites in unit*	*Estimate of camper nights*	*Confidence level*	*Error, percent*
A	21	28,196 ± 2,730	0.95	9.6
		28,196 ± 3,750	.99	13.2
B	39	28,196 ± 2,287	.95	8.1
		28,196 ± 3,142	.99	11.1
C	48	28,196 ± 2,070	.95	7.3
		28,196 ± 2,843	.99	10.1
D	60	28,196 ± 1,434	.95	5.1
		28,196 ± 1,970	.99	7.0

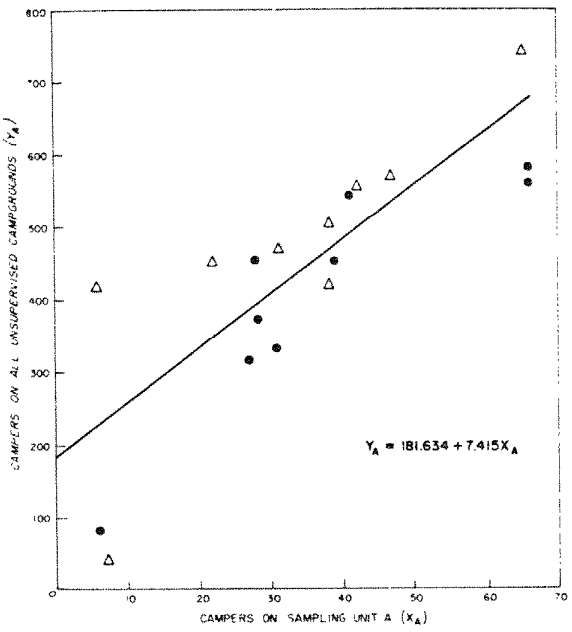


Figure 3.—Regression line for estimating total unsupervised camping from a count of campers on Wild River and Cold River Campgrounds (Sampling Unit A). Triangles represent data collected on weekend nights; round dots indicate data collected on week nights.

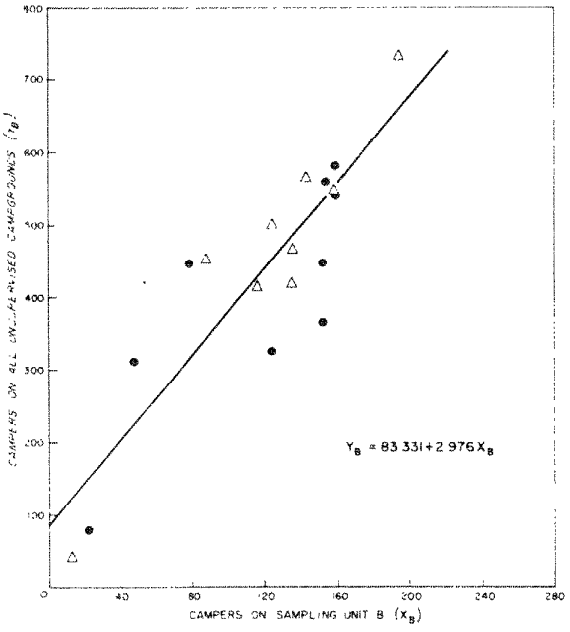
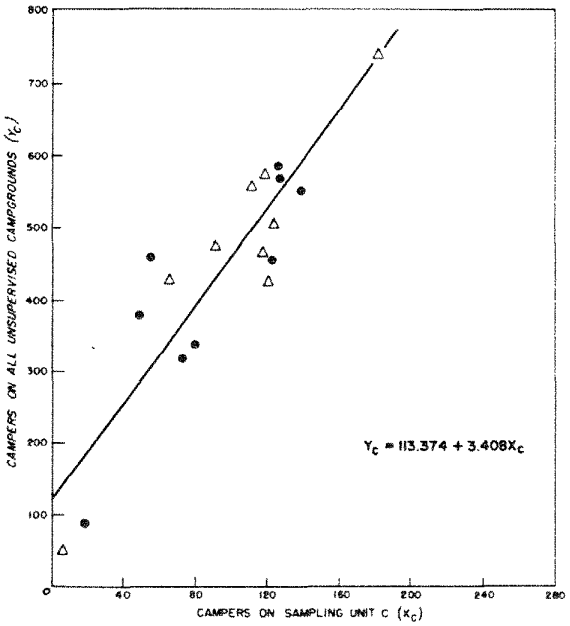


Figure 4.—Regression line for estimating total unsupervised camping from a count of campers on White Ledge Campground (Sampling Unit B). Triangles represent data collected on weekend nights; round dots represent data collected on week nights.

Figure 5. — Regression line for estimating total unsupervised camping from a count of campers on Wildwood and Waterville Campgrounds (Sampling Unit C). Triangles represent data collected on weekend nights; round dots indicate data collected on week nights.



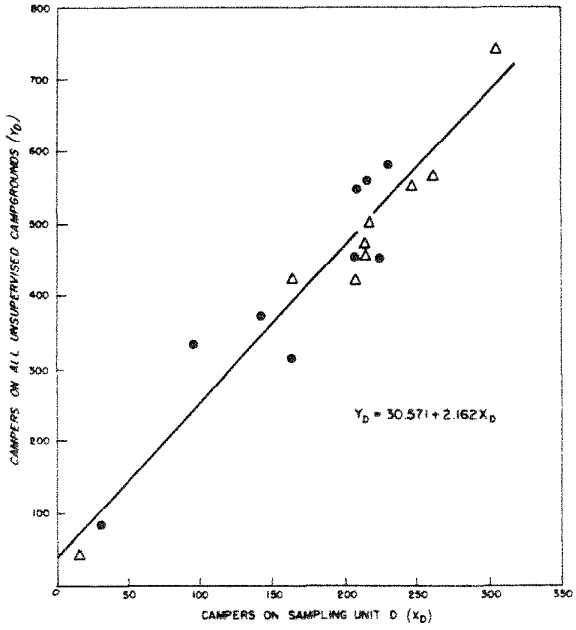


Figure 6. — Regression line for estimating total unsupervised camping from a count of campers on Dugway, Covered Bridge, and Passaconaway Campgrounds (Sampling Unit D). Triangles represent data collected on weekend nights; rounded dots indicate data collected on week nights.

Note that error decreases as sample size increases. This provides a basis for estimating the sample size necessary for a chosen level of precision.

Application of Results

Sampling is only one method of estimating numbers of recreationists, and consideration of other methods should not be overlooked just because they were not evaluated in this study. For example, in many situations improved methods of using vehicle counters might give better results at less cost. However, where vehicle counts are not meaningful, sampling offers an effective method of estimating recreational use.

It is fortunate that sampling precision depends primarily upon sample size rather than upon the size of the total population. Thus, if applied where there are many campsites, the regression or ratio estimation procedures described here should give acceptable precision with sampling rates that are economically feasible. For the data of this study, regression and ratio estimation proved much more efficient than random sampling. But there were only 168 campsites in the total population and only 63 nights in the sampling period. This meant that a rather high rate of sampling was required for collecting enough data to establish relationships. With longer camping seasons and greater numbers of campsites, the same levels of precision could be achieved with lower sampling rates. And where regressions (or ratios) prove to be usable for several years, sampling rates could average still lower.

Where it is possible to estimate numbers of campers on unsupervised campgrounds from actual registration on supervised campgrounds, this procedure may offer the most precise estimate for the least additional cost. For example, (1) if total counts on 18 nights were adequate for relating unsupervised use to records of supervised use, (2) if the relationship proved usable for 3 years, and (3) if each season lasted 100 days, then the required rate of sampling would average only 6 percent per year. At a confidence level of 95 percent, and for conditions as they existed

Campers at Dugway campground. Use of unsupervised campgrounds such as this one was compared with use of supervised campgrounds, where all campers were registered.

in this study, this rate of sampling would provide an estimate within approximately 5 percent of actual use.

However, it must be recognized that relationships between supervised and unsupervised camping will provide useful estimates only when camper numbers fluctuate in approximately the same patterns on both types of areas.

Where campers are not registered at all, it will be necessary to select a sample of campgrounds where nightly use is highly correlated with total nightly use on all campgrounds in the sampling area. Rather than count a different sample each night, as was done in this study, use should be counted every night on the same sample of campsites.

As an example of what sampling rates and precision might be expected, assume (1) that campground conditions will be as they existed in this study, (2) that a total of 500 campsites will be within the area being sampled, (3) that 18 total counts will establish an estimation equation that can be used for 3 years, and (4) that a sample of 24 campsites will be counted every night. Under these conditions the total rate of sampling (18 total counts plus nightly counts on the sample) would average approximately 10 percent per year. At a confidence level of 95 percent, this would provide an estimate within approximately 10 percent of actual use.

Procedures already in use at some areas might contribute to economy in sampling. Some campgrounds now have self-registration boards with a numbered peg for each campsite. Camping parties are required to fill out a card and hang it on the peg corresponding to the campsite they wish to occupy. At a few campgrounds (U.S. Forest Service 1961)[2] tickets must be purchased daily from a vending machine and displayed in the window of a locked box by the campsite. Where such procedures are in use, the proportion of people who fail to register or to buy tickets may be quite stable. If that is the case, total use probably can be estimated by regression or ratio methods from records of self registration or ticket sales. This would greatly reduce the costs of data collection and would allow higher rates of sampling.

Once use estimates and confidence limits are determined for individual sampling areas, they can be combined to give more precise estimates for broader regions. Thus, where the use estimate for sampling area i is represented by the expression:

$$\text{Estimate}_i \pm \text{Interval}_i$$

the combined estimate of use on N areas will be:

$$\sum_{i=1}^{N} \text{Estimate}_i \pm \sqrt{\sum_{i=1}^{N} (\text{Interval}_i)^2}$$

[2] United States Forest Service. CHARGING FOR USE OF NATIONAL FOREST RECREATION AREAS: AN ADMINISTRATIVE STUDY. U.S. Forest Serv. California Region, 13 pp., 1961.

For example, if seasonal use estimates on four hypothetical areas are:

Area	Estimate of camper nights	Error, percent
1	30,000 ± 3,000	10
2	40,000 ± 3,200	8
3	20,000 ± 2,400	12
4	50,000 ± 5,500	11

Then total use would be:

$$140{,}000 \pm \sqrt{(3{,}000)^2 + (3{,}200)^2 + (2{,}400)^2 + (5{,}500)^2}$$

$$= 140{,}000 \pm 7{,}434 \text{ camper nights } (5.31\text{-percent error}).$$

Thus the sampling methods described here would be efficient if applied on a large enough scale. This would require (1) a large number of campsites on one sampling area or (2) the combining of independent estimates from several sampling areas.

■