United States
Department of
Agriculture

Forest Service

Pacific Northwest
Research Station

General Technical
Report
PNW-GTR-456
March 1999

Sampling and Estimating Recreational Use

Timothy G. Gregoire and Gregory J. Buhyoff

Authors

TIMOTHY G. GREGOIRE is with the School of Forestry and Environmental Studies, Yale University, 360 Prospect Street, New Haven, Connecticut 06511; and GREGORY J. BUHYOFF is with the Department of Forestry (0324), College of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. This work was performed under PNW P.O. 43-04H1-7-0159.

Abstract

Gregoire, Timothy G.; Buhyoff, Gregory J. 1999. Sampling and estimating recreational use. PNW-GTR-456. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 39 p.

Probability sampling methods applicable to estimate recreational use are presented. Both single- and multiple-access recreation sites are considered. One- and two-stage sampling methods are presented. Estimation of recreational use is presented in a series of examples.

Keywords: Sampling strategies, confidence intervals, stratified sampling.

Summary

This report is intended to provide guidance on the execution of sampling strategies to estimate recreational use in settings such as National Forests. The intended audience for the material presented here is resource managers, supervisors, and others in positions of responsibility for the allocation and maintenance of recreational resources. Informed management requires a credible assessment of how frequently and intensely the recreation resource is used. The procedures described in this report are tools for that purpose. The presentation is intended to be specific enough in the essential characteristics that these sampling strategies can be applied generally to the tremendous diversity of recreation sites and forms of recreation supported by them. No attempt has been made to anticipate special features of some recreation sites or attributes of certain forms of recreation that would presume some modification of the basic designs. As a consequence, this report serves more as a primer than as a comprehensive guide.

The orientation of this report is quite general, too, in that we do not focus on particular types of recreational-use areas or on types of recreational use, except for sake of example. Because of the differing opportunities for sampling that are presented, we distinguish broadly between overnight recreational-use areas and others. For lack of a better or established terminology, we refer to the latter as “transient-recreational-use areas.” For transient-recreational-use areas, we present sampling strategies for areas with a single access or entry point separately from those suitable for areas with multiple access or entry points.

Our concern is with estimating recreational use, as measured, for example, by the number of weekday recreationists on a particular hiking trail. We have not attempted to devise strategies to acquire information that cannot be observed by the person conducting the sample; that is, the strategies we propose are nonintrusive by design and therefore insufficient to estimate, for example, the average age of recreationist on a trail or to determine whether two hikers are members of a group hiking together. For such purposes, the sample selection methods we present may remain applicable, but the methods to elicit information would not.

This page has been left blank intentionally.
Document continues on next page.

Introduction

This report summarizes sampling strategies to estimate recreational use, such as average number of hikers on a park or forest trail. The sampling and estimation plans described in this report are aimed at producing credible estimates of visitor use based on observed activity. Sampling to estimate duration of recreational use is a more involved task than that considered here.

It would be futile to enumerate all possible forms of recreation amenable to the sampling strategies presented in this report, but hiking and overnight camping are two forms that provide useful foci for discussion. Within a 24-hour-day sampling frame, hiking along a specific trail can be regarded as a transient form of recreation, whereas overnight camping is not. We broadly categorize sites as being a transient-use type or an overnight-recreational-use type. Accordingly, sampling strategies for the two categories of recreation sites are presented separately.

For recreation sites with a single access point (for example, a closed loop trail with a lone trail head), the sampling ought to be conducted at that point. The neatness of the sampling and estimation procedures deteriorates for sites having multiple access points. We therefore treat the two types of sites separately, and we account for both the sampling and estimation details applicable to each.

Each sampling plan has a random component, essential to guard against biased estimation of recreational use. In other words, the sampling is probabilistic and therefore ensures that each recreationist using the site has an objective chance of being included in the sample. This feature, lacking in self-selected samples of names written in visitor registries, is crucial to the statistical validity of the results. Once a day and time are selected for sampling a recreation site, it is imperative that sampling be conducted (barring physical harm to the sampler, which may result from flash floods, lightning strikes, or similar harsh weather and environmental conditions). Convenience sampling must be avoided absolutely; for example, it is unacceptable to avoid sampling on a randomly selected but rainy day.

Measurement error is an abiding concern in any sampling effort. The methods proposed in this report require that a trained observer record recreational use. Measurement error in this context is an incorrect count of the number of recreationists or a misrecorded count of same. While more expensive, perhaps, than relying on mechanical or optical counters of visitor activity, the employment of a trained observer is more reliable and informative. The control of measurement error within the probability sampling framework imparts a scientific credibility to the system that will bear up under close scrutiny.

The probabilistic sampling plan coupled with the proposed estimation procedures ensures that recreational use will be unbiasedly estimated. The variance of the resulting estimate is largely a function of sampling intensity and frequency. Past sample information may provide data that can be used to plan the size of future samples at similar sites. This topic is addressed further in "Planning a Sample of Recreational Use."

The stratified random sampling methods suggested in this report are straightforward to administer and can be applied virtually without modification to any recreation area. We suggest, for example, that sampling be planned separately for weekends versus weekdays, whenever recreational use varies much between these two weekly strata.

If recreation or management activities occur much more, or less, frequently on holidays, then the resource manager (RM) might consider creating a third stratum, and plan the sampling effort for that stratum separately from the other two.

Stratification is suggested as a way not only to obtain separate estimates for each meaningfully defined stratum but also to increase the precision of the resulting estimates. Although more sophisticated plans can be devised and hold the promise of estimating recreational use more precisely, they are not suggested because they lack generality, are less intuitively plausible to persons untrained in the nuances of statistical sampling, and are less easily transported and implemented elsewhere. For multipurpose surveys, simplicity of sample design is a virtue.

We focused principally on estimating the average number of persons engaging daily in specific forms of recreation (hiking, biking, camping), during some period stipulated by the RM. This period, which will comprise the sampling frame of sample days, can be a month, a year, or longer. We refer hereafter to this designated period as a “season.” Presumably, the season will be an interval having relevance to the management of the recreational resource. However it is defined, we explain how to conduct the sample and show how to prorate the sample data to obtain an estimate of recreational use applicable to that period, either daily or seasonally.

We have woven several numerical examples into the text. Their purpose is to take some of the mystery out of the symbolic formulas for estimators of recreational use by demonstrating how the sample data actually are manipulated. The data are fictitious, as the examples needed to be short to encourage verification by those using the procedures. A simple computer spreadsheet should be sufficient to process not only the data provided in the examples but actual sample data as well.

Selecting Sample Days

We restricted the sample selection method to simple random sampling without replacement of days. In this section, we describe two methods that use a computer spreadsheet to select a simple random sample without replacement (SRS). Alternative methods of selecting a SRS may, of course, be used instead.

Suppose that the RM wishes to estimate recreational use for weekends separately from weekdays. The technique of poststratification can be used to do this when stratification has not been done prior to sampling. Poststratification procedures are exceedingly valuable in “after the fact” analyses of sample data. The treatment of poststratification is beyond the scope of this report, however. We suppose instead that the RM has the liberty to select sample days, by strata, beforehand. In the following, we first outline a procedure to select stratified random samples for the particular case where the RM is interested in estimating annual recreation (population of 365 or 366 days). Then we broaden this to the case where the time period of interest is defined differently, such that the population consists of some number of days implied by the definition, which we show as N.

Suppose that the RM had been interested in sampling during the 1997 calendar year, in which there were 111 weekend days, including holidays, and 254 weekdays. Based on considerations shown in the appendix, we assumed that the RM decided to sample 6 percent of the weekend and 3 percent of the weekdays throughout the year. After rounding to the nearest integer value, the RM needs to randomly select (in the sense of selecting with equal probability) $0.06 \times 111 = 7$ weekend days and $0.03 \times 254 = 8$ weekdays. In a spreadsheet, generate a column sequence of numbers from 1 to 365. Most spreadsheets have the facility to generate such a sequence automatically, and the chore of entering 365 numbers can be avoided. In an adjacent column, generate a series of uniform random numbers from the spreadsheet's function list.

With the spreadsheet's sorting function, rearrange both columns in order of increasing value of the random numbers in the second column.

From the sorted list, select from the top of the first column the first 8 days that correspond to weekdays. Then select the first 7 days that are weekend days. You will have chosen a SRS of 8 weekdays from the pool of 254 and a SRS of 7 weekend days from the pool of 111.

Many spreadsheets have a SRS function that can be used instead. If so, it will be necessary to list separately the weekdays from the weekend days, and to separately invoke the SRS function on each list. We think the procedure described first is easier but want to point out the option.

Once a day has been selected in advance, recreational use **must** be recorded on that day. Sampling must not be postponed owing to inclement weather, unless the sampler's safety is endangered. In that case, another day for sampling must be selected at random; in particular, it is not acceptable to choose the day immediately following the initially selected day. An advantage of the selection procedure using the sorted list of random numbers is that replacement sample days can be chosen easily by going further down the sorted list.

Now consider the case when the period of interest is the season spanning Memorial Day to Thanksgiving Day. Measured in days, the length of this period was $N = 180$ days altogether, in calendar year 1997; of these, there were $N_1 = 126$ weekdays and $N_2 = 54$ weekend days and holidays. We suppose that $\alpha_1 = 0.03$ remains the proportion of weekdays to be sampled; that is, the percentage of weekdays to be selected is $100\% \times \alpha_1 = 3\%$, as before. Similarly, $\alpha_2 = 0.06$ remains the proportion of weekend days to be sampled, so that the percentage of weekend days to be selected is $100\% \times \alpha_2 = 6\%$.

Regardless of the numbers attached to the symbols, the general formula for computing the actual numbers of weekdays to be selected is,

$$n_1 = \alpha_1 N_1 ,$$

where α_1 represents the sampling proportion from the stratum and N_1 represents the number of days in stratum 1. Similarly for stratum 2, the general formula for computing the actual numbers of weekend days to be selected is,

$$n_2 = \alpha_2 N_2 ,$$

where α_2 represents the sampling proportion from this stratum and N_2 represents the number of days in stratum 2 . Using the hypothetical numbers put forth for this example, these formulas solve to

$$n_1 = \alpha_1 N_1 = 0.03 \times 126 = 4 \text{ weekdays}$$

and

$$n_2 = \alpha_2 N_2 = 0.06 \times 54 = 3 \text{ weekend days ,}$$

after rounding the answers to the closest integer value.

To select the actual days to conduct the sampling, use the same spreadsheet algorithm outlined above when the entire calendar year was the sampling frame. Instead of $N = 365$ as in that case, here the column sequence needs to extend from 1 to $N = 180$, where day 1 implicitly represents Memorial Day. More generally, day 1 represents the initial day of the period of interest.

When holidays are partitioned separately into a third stratum, as mentioned in the "Introduction," the formulas presented above extend rather naturally; namely, the RM must determine the number of holidays, N_3, comprising this third stratum, as well as the sampling fraction, α_3, to be included in the sample. The number of days selected from stratum 3 is

$$n_3 = \alpha_3 N_3 .$$

The above discussion is framed in terms of a sampling day, yet it is unrealistic to assume that this coincides with a 24-hour day when recreational use of the site takes place. The RM must decide, for each particular site, the period during the 24-hour day when recreational use of the site takes place. This will differ, perhaps, with the form of recreational use being monitored. In some cases the 14-hour period from 6:00 a.m. to 8:00 p.m. may be sufficient; in other cases, it may be excessive or insufficient. In principle, anything less than a 24-hour sampling day invites measurement (observer) error. We assume that the RM chooses the sampling day knowledgeably to minimize the magnitude of measurement error.

As an aside, it might be worthwhile to note that the cost of sampling can be computed rather quickly in the case where sampling costs do not differ among strata: it costs as much to station a sampler during the week as it does during the weekend or on a holiday. In this case, the total size of the sample is $n = n_1 + n_2$ or $n = n_1 + n_2 + n_3$, depending on the number of strata one decides to use. If c represents the cost of sampling for a single day, then the overall cost of sampling is just

$$C = c \times n .$$

Sampling Strategy for Transient-Recreational-Use Areas

Single Access Sites

For sake of discussion, we presume that there are two types of recreational use one wishes to monitor—backpacking and biking—separately across two strata—weekdays and weekend days. Our use of two different forms of recreational use does not imply more or fewer forms cannot be monitored. Our use of two is solely for example.

On randomly selected days, a person will observe and record use of the site. This person should be located at the access point to the site. Ideally, this person should record recreational use for the entire portion of the day during which recreational activities are likely to occur. If sampling cannot be conducted for an entire day, then a two-stage sampling approach should be adopted, wherein days are selected randomly as described above, and then a predetermined block of hours within the day is selected randomly. Details of the two-stage approach are discussed in "Two-Stage Selection," on page 14.

The observer stationed at the access point should record only traffic entering the site. The data recording form, whether paper or electronic, should have fields to indicate the recreation site, sampling location within site, date, and the beginning and ending times of observation.

Each person entering the trail should be recorded. At a minimum, separate tallies by recreational use must be maintained, in this case backpacking and biking. Further information may prove useful for management purposes; for example, tallying recreationists by both use and gender, providing gender can be discerned with negligible measurement error.

Notation

As established in "Selecting Sample Days," N, N_1, and N_2 represent, respectively, the number of days in the period of interest, in the subset of the period of interest encompassing weekdays, and in the subset encompassing weekend days. Similarly, n, n_1, and n_2 represent the number of sample days.

To distinguish among the number of individual backpackers tallied on the different sample days in stratum 1, we introduce this notation:

$$
\begin{aligned}
I_{1,1}(\text{backpackers}) &= \text{observed number of backpackers on sample day 1 in stratum 1} \\
&\vdots \\
I_{1,k}(\text{backpackers}) &= \text{observed number of backpackers on sample day k in stratum 1} \\
&\vdots \\
I_{1,n_1}(\text{backpackers}) &= \text{observed number of backpackers on sample day } n_1 \text{ in stratum 1}
\end{aligned}
$$

Similarly for the tallies of backpackers in stratum 2, we have the following:

$$
\begin{aligned}
I_{2,1}(\text{backpackers}) &= \text{observed number of backpackers on sample day 1 in stratum 2} \\
&\vdots \\
I_{2,k}(\text{backpackers}) &= \text{observed number of backpackers on sample day k in stratum 2} \\
&\vdots \\
I_{2,n_2}(\text{backpackers}) &= \text{observed number of backpackers on sample day } n_2 \text{ in stratum 2}
\end{aligned}
$$

The numbers of bikers tallied throughout the sample can be symbolized in a parallel fashion:

$$
\begin{aligned}
I_{1,1}(\text{bikers}) &= \text{observed number of bikers on sample day 1 in stratum 1} \\
&\vdots \\
I_{1,k}(\text{bikers}) &= \text{observed number of bikers on sample day k in stratum 1} \\
&\vdots \\
I_{1,n_1}(\text{bikers}) &= \text{observed number of bikers on sample day } n_1 \text{ in stratum 1}
\end{aligned}
$$

For stratum 2,

$$\begin{aligned} I_{2,1}(\text{bikers}) &= \text{observed number of bikers on sample day 1 in stratum 2} \\ &\vdots \\ I_{2,k}(\text{bikers}) &= \text{observed number of bikers on sample day k in stratum 2} \\ &\vdots \\ I_{2,n_2}(\text{bikers}) &= \text{observed number of bikers on sample day } n_2 \text{ in stratum 2} \end{aligned}$$

Estimating formulas by stratum—An estimate of the number of backpackers per day in stratum 1 is given by,

$$\hat{I}_1(\text{backpackers}) = \frac{1}{n_1}\sum_{k=1}^{n_1} I_{1,k}(\text{backpackers})$$

$$= \frac{I_{1,1}(\text{backpackers}) + \ldots + I_{1,k}(\text{backpackers}) + \ldots + I_{1,n_1}(\text{backpackers})}{n_1} ,$$

so that an estimate of the total number of backpackers using this site in stratum 1 is this average multiplied by the number of days in the stratum:

$$\hat{T}_1(\text{backpackers}) = N_1 \times \hat{I}_1(\text{backpackers})$$

An estimate of the number of backpackers per day in stratum 2 is given by,

$$\hat{I}_2(\text{backpackers}) = \frac{1}{n_2}\sum_{k=1}^{n_2} I_{2,k}(\text{backpackers})$$

$$= \frac{I_{2,1}(\text{backpackers}) + \ldots + I_{2,k}(\text{backpackers}) + \ldots + I_{2,n_2}(\text{backpackers})}{n_2} .$$

The estimate of total number of backpackers in stratum 2 is,

$$\hat{T}_2(\text{backpackers}) = N_2 \times \hat{I}_2(\text{backpackers}) .$$

An estimate of the average number of bikers per day in stratum 1 is given by,

$$\hat{I}_1(\text{bikers}) = \frac{1}{n_1}\sum_{k=1}^{n_1} I_{1,k}(\text{bikers})$$

$$= \frac{I_{1,1}(\text{bikers}) + \ldots + I_{1,k}(\text{bikers}) + \ldots + I_{1,n_1}(\text{bikers})}{n_1} \ ,$$

so that the estimate of the total number of bikers using this site in stratum 1 is this average multiplied by the number of days in the stratum,

$$\hat{T}_1(\text{bikers}) = N_1 \times \hat{I}_1(\text{bikers})$$

An estimate of the average number of bikers per day in stratum 2 is given by,

$$\hat{I}_2(\text{bikers}) = \frac{1}{n_2}\sum_{k=1}^{n_2} I_{2,k}(\text{bikers})$$

$$= \frac{I_{2,1}(\text{bikers}) + \ldots + I_{2,k}(\text{bikers}) + \ldots + I_{2,n_2}(\text{bikers})}{n_2} \ .$$

The estimate of total number of bikers in stratum 2 is,

$$\hat{T}_2(\text{backpackers}) = N_2 \times \hat{I}_2(\text{bikers}) \ .$$

Estimating formulas for both strata combined—Because the estimated total number of backpackers in both strata is the sum,

$$\hat{T}(\text{backpackers}) = \hat{T}_1(\text{backpackers}) + \hat{T}_2(\text{backpackers}) \ ,$$

the average number of backpackers per day in both strata is estimated as

$$\hat{I}(\text{backpackers}) = \hat{T}(\text{backpackers}) / N \ .$$

For bikers a similar computation is used:

$$\hat{T}(\text{bikers}) = \hat{T}_1(\text{bikers}) + \hat{T}_2(\text{bikers}) .$$

The average number of bikers per day in both strata is estimated as

$$\hat{I}(\text{bikers}) = \hat{T}(\text{bikers}) / N .$$

Estimating the total number of recreationists per season is straightforward:

$$\hat{T} = \hat{T}(\text{backpackers}) + \hat{T}(\text{bikers}) .$$

The average number of both types of recreationists per day is estimated by

$$\hat{I} = \hat{T} / N .$$

Example 1—Management needed to estimate the amount of backpacking and biking recreation on the Spruce Run Trail throughout the 1997 season from Memorial Day to Thanksgiving Day. During this period, sampling was conducted on $n_1 = 4$ weekdays and $n_2 = 3$ weekend days. The observed use is summarized in the following tabulation:

Recreationist	**Stratum 1: weekdays**				**Stratum 2: weekends**		
Backpackers	13	16	12	11	28	32	22
Bikers	8	6	8	10	10	18	14

From these daily tallies of recreational use, the average daily number of weekday backpackers can be estimated by,

$$\hat{I}_1(\text{backpackers}) = \frac{1}{4}(13 + 16 + 12 + 11) = 13 .$$

Similarly, the estimate of the average daily number of weekend day backpackers is,

$$\hat{I}_2(\text{backpackers}) = \frac{1}{3}(28 + 32 + 22) = 24 .$$

The estimates of biking use are computed as,

$$\hat{I}_1(\text{bikers}) = \frac{1}{4}(8 + 6 + 8 + 10) = 8 ,$$

and

$$\hat{I}_2(\text{bikers}) = \frac{1}{3}(10 + 18 + 14) = 14 .$$

Because $N = 180$ days in this period in 1997, the average daily number of backpackers is estimated to be,

$$\hat{I}(\text{backpackers}) = \frac{1}{180}(126 \times 13 + 54 \times 24) = 16.3 \text{ backpackers} ,$$

where 126 and 54 represent, respectively, the number of days in the weekday and weekend strata. Seasonally, the estimated number of backpackers is,

$$\hat{T}(\text{backpackers}) = 2934 .$$

Likewise, the average daily number of bikers is,

$$\hat{I}(\text{bikers}) = \frac{1}{180}(126 \times 8 + 54 \times 14) = 9.8 \text{ bikers} .$$

Seasonally, the estimated number of bikers is,

$$\hat{T}(\text{bikers}) = 1764 .$$

The total number of both types of recreationists throughout the season is,

$$\hat{T} = 4698 .$$

The average daily number of backpackers and bikers is

$$\hat{I} = 26.1 .$$

Variance Estimation and Confidence Intervals

Stratum estimates—An estimate of the variance of $\hat{I}$(backpackers) depends on the extent to which the individual daily tallies, $I_{1,1}, I_{1,2}, ..., I_{1,n_1}$ vary around $\hat{I}$(backpackers). Indeed, it is computing the variance of an estimate of recreation visits and constructing a confidence interval for the underlying population value of the number of visits that necessitate keeping the daily tallies separate. As a computational aid, we suggest the these deviations be computed separately as,

$$\begin{aligned}
\mathrm{DI}_{1,1}(\text{backpackers}) &= \mathrm{I}_{1,1}(\text{backpackers}) - \hat{\mathrm{I}}_1(\text{backpackers}) \\
&\vdots \\
\mathrm{DI}_{1,k}(\text{backpackers}) &= \mathrm{I}_{1,k}(\text{backpackers}) - \hat{\mathrm{I}}_1(\text{backpackers}) \\
&\vdots \\
\mathrm{DI}_{1,n_1}(\text{backpackers}) &= \mathrm{I}_{1,n_1}(\text{backpackers}) - \hat{\mathrm{I}}_1(\text{backpackers}) .
\end{aligned}$$

The estimated variance of $\hat{I}$ is designated as $\hat{\mathrm{var}}[\hat{I}_1(\text{backpackers})]$ and can be computed by the following formula:

$$\hat{\mathrm{var}}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right] = \left(\frac{1}{n_1} - \frac{1}{N_1}\right) \frac{\sum_{k=1}^{n_1} \mathrm{DI}_{1,k}^2(\text{backpackers})}{n_1 - 1} .$$

The estimated standard error of $\hat{\mathrm{I}}_1$ is $\sqrt{\hat{\mathrm{var}}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right]}$.

Confidence intervals for average number of recreationists visiting a site—An approximate 90-percent confidence interval for I_1(backpackers) is constructed as,

$$\hat{\mathrm{I}}_1(\text{backpackers}) \pm \mathrm{t}_{df}^{0.90} \sqrt{\hat{\mathrm{var}}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right]} ,$$

where $t_{df}^{0.90}$ is the 90*th* percentile point from the Student's *t* distribution with $df = n_1 - 1$ degrees of freedom.

An alternative expression of the confidence interval which many find more appealing is one for which the

$$\pm \mathrm{t}_{df}^{0.90} \sqrt{\hat{\mathrm{var}}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right]}$$

portion is expressed as a percentage of $\hat{I}_1$(backpackers):

$$\hat{I}_1(\text{backpackers}) \pm E\% \ ,$$

where $E\% = \left\{ \dfrac{t_{df}^{0.90} \sqrt{\hat{\text{var}}\left[\hat{I}_1(\text{backpackers}) \right]}}{\hat{I}_1(\text{backpackers})} \right\} 100\%.$

For the weekend stratum,

$$\hat{\text{var}}\left[\hat{I}_2(\text{backpackers}) \right] = \left(\frac{1}{n_2} - \frac{1}{N_2} \right) \frac{\sum_{k=1}^{n_2} DI_{2,k}^2(\text{backpackers})}{n_2 - 1} \ ,$$

where $DI_{2,k}$(backpackers) = $I_{2,k}$(backpackers) - $\hat{I}_2$(backpackers), and so on.
An approximate 90-percent confidence interval for I_2(backpackers) is constructed as,

$$\hat{I}_2(\text{backpackers}) \pm t_{df}^{0.90} \sqrt{\hat{\text{var}}\left[\hat{I}_2(\text{backpackers}) \right]}$$

where $t_{df}^{0.90}$ is the 90th percentile point from the Student's t distribution with $df = n_2 - 1$ degrees of freedom. Equivalently, one can express the confidence interval as

$\hat{I}_2$(backpackers) $\pm$ E% where $E\% = \left\{ \dfrac{t_{df}^{0.90} \sqrt{\hat{\text{var}}\left[\hat{I}_2(\text{backpackers}) \right]}}{\hat{I}_2(\text{backpackers})} \right\} 100\% \ .$

Combined strata estimates—For both strata combined,

$$\hat{\text{var}}\left[\hat{I}(\text{backpackers}) \right] = \frac{1}{N^2} \left\{ N_1^2 \times \hat{\text{var}}\left[\hat{I}_1(\text{backpackers}) \right] + N_2^2 \times \hat{\text{var}}\left[\hat{I}_2(\text{backpackers}) \right] \right\} \ .$$

An approximate 90-percent confidence interval for I(backpackers) is constructed as,

$$\hat{I}(\text{backpackers}) \pm t_{df}^{0.90} \sqrt{\hat{\text{var}}\left[\hat{I}(\text{backpackers}) \right]}$$

where $t_{df}^{0.90}$ is the 90*th* percentile point from the Student's *t* distribution with $df = n_1 - n_2 - 2$ degrees of freedom. An equivalent expression is $\hat{\text{I}}(\text{backpackers}) \pm \text{E}\%$,

$$\text{where} \;\; \text{E}\% = \left\{ \frac{t_{df}^{0.90} \sqrt{\hat{\text{var}}\left[\hat{\text{I}}(\text{backpackers})\right]}}{\hat{\text{I}}(\text{backpackers})} \right\} 100\% \; .$$

Example 2—Continuing from example 1, a few spreadsheet manipulations provided,

$$\hat{\text{var}}\left[\hat{\text{I}}_1(\text{backpackers})\right] = \left(\frac{1}{4} - \frac{1}{126}\right)\left(4.\overline{6}\right) = 1.13$$

$$\hat{\text{var}}\left[\hat{\text{I}}_2(\text{backpackers})\right] = \left(\frac{1}{3} - \frac{1}{54}\right)\left(25.\overline{3}\right) = 7.98$$

$$\hat{\text{var}}\left[\hat{\text{I}}(\text{backpackers})\right] = \frac{1}{180^2}\left[126^2 \times 1.13 + 54^2 \times 7.98\right] = 1.27 \; .$$

Approximate 90-percent confidence intervals are constructed with $t_3^{0.90} = 2.353$, $t_2^{0.90} = 2.290$, and $t_5^{0.90} = 2.015$:

Parameter	**Approximate 90-percent confidence interval**
$\text{I}_1(\text{backpackers})$	13 ± 2.5 backpackers ≡ 13 backpackers ± 19%
$\text{I}_2(\text{backpackers})$	24 ± 8.2 backpackers ≡ 24 backpackers ± 34%
$\text{I}(\text{backpackers})$	16.3 ± 2.3 backpackers ≡ 16.3 backpackers ± 14%

The corresponding calculations for bikers are:

$$\hat{\text{var}}\left[\hat{\text{I}}_1(\text{bikers})\right] = \left(\frac{1}{4} - \frac{1}{126}\right)\left(2.\overline{6}\right) = 0.65$$

$$\hat{\text{var}}\left[\hat{\text{I}}_2(\text{bikers})\right] = \left(\frac{1}{3} - \frac{1}{54}\right)\left(16.0\right) = 5.04$$

$$\hat{\text{var}}\left[\hat{\text{I}}(\text{bikers})\right] = \frac{1}{180^2}\left[126^2 \times 0.65 + 54^2 \times 5.04\right] = 0.77$$

Approximate 90-percent confidence intervals are constructed with $t_3^{0.90} = 2.353$, $t_2^{0.90} = 2.290$, and $t_5^{0.90} = 2.015$:

Parameter	Approximate 90-percent confidence interval
I_1(bikers)	8 ± 1.9 bikers ≡ 8 bikers ± 34%
I_2(bikers)	14 ± 6.5 bikers ≡ 14 bikers ± 47%
I(bikers)	9.9 ± 1.8 bikers ≡ 9.9 bikers ± 18%

Confidence intervals for totals—Because

$$\hat{T}_1(\text{backpackers}) = N_1 \times \hat{I}_1(\text{backpackers}),$$

then

$$\widehat{\text{var}}\left[\hat{T}_1(\text{backpackers})\right] = N_1^2 \times \widehat{\text{var}}\left[\hat{I}_1(\text{backpackers})\right],$$

and an approximate 90-percent confidence interval for T_1(backpackers) is just

$$\hat{T}_1(\text{backpackers}) \pm N_1 \times t_{df}^{0.90}\sqrt{\widehat{\text{var}}\left[\hat{I}_1(\text{backpackers})\right]},$$

where $df = n_1 - 1$.

Similarly,

$$\widehat{\text{var}}\left[\hat{T}_2(\text{backpackers})\right] = N_2^2 \times \widehat{\text{var}}\left[\hat{I}_2(\text{backpackers})\right]$$

and

$$\hat{T}_2(\text{backpackers}) \pm N_2 \times t_{df}^{0.90}\sqrt{\widehat{\text{var}}\left[\hat{I}_2(\text{backpackers})\right]},$$

when $df = n_2 - 1$.

Finally,

$$\widehat{\text{var}}\left[\hat{T}(\text{backpackers})\right] = \widehat{\text{var}}\left[\hat{T}_1(\text{backpackers})\right] + \widehat{\text{var}}\left[\hat{T}_2(\text{backpackers})\right]$$

and

$$\hat{\mathrm{T}}(\text{backpackers}) \pm \mathrm{N} \times t_{df}^{0.90}\sqrt{\hat{\mathrm{var}}\left[\hat{\mathrm{I}}(\text{backpackers})\right]}\ ,$$

where $df = n_1 + n_2 - 2$.

A parallel development holds when estimating the total number of recreational bikers, namely $\mathrm{T}(\text{bikers})$.

Example 3—Using the results from example 2, the following intervals for $\mathrm{T}_1(\text{backpackers})$, $\mathrm{T}_2(\text{backpackers})$, and $\mathrm{T}(\text{backpackers})$ are obtained from the above formulas:

Parameter	Approximate 90-percent confidence interval
$\mathrm{T}_1(\text{backpackers})$	1638 ± 315 backpackers ≡ 1638 backpackers ± 19%
$\mathrm{T}_2(\text{backpackers})$	1296 ± 445 backpackers ≡ 1296 backpackers ± 34%
$\mathrm{T}(\text{backpackers})$	2934 ± 409 backpackers ≡ 2934 backpackers ± 14%

Notice that the ± (percentage) terms are identical to those in confidence intervals for $\mathrm{I}_1(\text{backpackers})$, $\mathrm{I}_2(\text{backpackers})$, and $\mathrm{I}(\text{backpackers})$.

Two-Stage Selection

If the staffing and expense of sampling recreational use for an entire day are prohibitive, we suggest a two-stage approach, wherein each potential day of sampling is partitioned into fixed blocks of time. For a 14-hour sampling day, four 3.5-hour blocks might make a logical partition. Three 5-hour blocks, likewise, would make sense for a 15-hour sampling day. The two-stage selection procedure consists of the random selection of a day, followed by a random selection of a block of time within that day. We envision that the number of blocks into which each day is partitioned will be rather small; for example, two to four blocks per day. Sampling of recreational use on the trail then proceeds as previously explained.

The estimation formulas presented earlier are appropriate only for single-stage sampling wherein the entire day is selected. Rather than repeat so many formulas here, we describe a multiplicative adjustment to the previous formulas. Suppose that the sampling day is partitioned such that each block of the partition represents a fraction, f, of the entire day, and that a single block of hours within a sampling day constitutes the subsample. Then for both backpackers and bikers, the estimation formulas for $\hat{\mathrm{I}}_1$ and $\hat{\mathrm{I}}_2$ should be multiplied by the factor $1/f$. The estimation formulas for $\hat{\mathrm{I}}, \hat{\mathrm{T}}_1$, $\hat{\mathrm{T}}_2$, and $\hat{\mathrm{T}}$ remain unchanged.

Unless more than one block within a day is subsampled, there is no way to unbiasedly estimate the variance of $\hat{\mathrm{I}}_1$, $\hat{\mathrm{I}}_2$, $\hat{\mathrm{I}}, \hat{\mathrm{T}}_1, \hat{\mathrm{T}}_2$, and $\hat{\mathrm{T}}$. The variance estimators given in the preceding section can be used, but they will underestimate the actual amount of sampling variation. An obvious alternative is to subsample at least two of the blocks of time into which the sample day is partitioned. In view of the coarse partitioning of each sampling day that we envision, however, the rationale for subsampling at least two blocks within a day, rather than sampling the entire day, becomes rather strained.

If the two-stage approach is adopted, it is essential that **each** random selection of a sampling day be accompanied by a random selection of a time block within that day. It is not appropriate to randomly select a single time block, which then is used throughout all days of sampling.

It is important that the two-stage procedure, if adopted, be used throughout the entire sampling effort; moreover, the partitioning of the day into blocks of consecutive hours must remain constant. We envision a partition of each day, for example, into four blocks where the first block of time spans 6:00-9:30 a.m.; the second spans 9:30 a.m. to 1:00 p.m.; the third spans 1:00-4:30 p.m.; and the fourth spans 4:30-8:00 p.m. Under the two-stage selection, this partitioning of the day is fixed, and one of these four time blocks would be chosen at random. It is **not** acceptable, for example, to randomly choose *any* time of day as the start of a 3.5-hour block of time. It is crucial that sampling be conducted during the selected block of time. Subjective substitution of one time block for another (e.g., due to inclement weather) will inject bias into the estimates of recreational use.

It is **not** important that a two-stage procedure be used for all trails, but it would simplify record keeping if the same sample design were implemented consistently for all trails.

Multiple Access Sites

When recreationists can enter and exit a site at more than one access point, it is possible to unbiasedly estimate recreational use by stationing an observer at each point on the selected sampling days. Each observer would record the number of recreationists entering the site, presuming again that separate tallies would be maintained for backpackers and bikers. In this situation, recreationists passing by the access point would not be tallied, as they would have been tallied at the entry. Recreation use may be estimated by combining all the observers' tallies into a single tally, and then using the formulas given for "single-access sites." Very likely, this option is not feasible for sites having numerous access points.

A two-stage sample is an alternative when there are fewer observers available than there are access points. One consideration would be to randomly (that is, simple random sampling without replacement) assign the available observers among the access points in a two-stage sampling scheme. Suppose m represents the number of access points at which observers are stationed on a particular sampling day, and that not only is m identical for all strata but its value also remains constant throughout the sampling season. It is understood, however, that the set of selected m access points will differ from one sampling day to another.

Under this two-stage setup, suppose that

$$\text{M} = \text{the number of access points on the site}$$

so that

$$g = m/\text{M}.$$

Furthermore let

$I_{1,1,1}$(backpackers) = observed number of backpackers on sample day 1 in stratum 1 at the first of the subsampled access points

$\vdots$

$I_{1,1,j}$ (backpackers) = observed number of backpackers on sample day 1 in stratum 1 at the *j*th of the subsampled access points

⋮

$I_{1,1,m}$ (backpackers) = observed number of backpackers on sample day 1 in stratum 1 at the *m*th of the subsampled access points

Similarly for the tallies of backpackers in stratum 2, we have

$I_{2,1,1}$ (backpackers) = observed number of backpackers on sample day 1 in stratum 2 at the first of the subsampled access points

⋮

$I_{2,1,j}$ (backpackers) = observed number of backpackers on sample day 1 in stratum 2 at the *j*th of the subsampled access points

⋮

$I_{2,1,m}$ (backpackers) = observed number of backpackers on sample day 1 in stratum 2 at the *m*th of the subsampled access points

The tallies for bikers for each day, stratum, and access point can be denoted in the same way.

Let $\hat{I}_{1,k}$(backpackers) represent the estimated average number of individual backpackers for the *k*th sampling day in stratum 1, which is computed as,

$$\hat{I}_{1,k}(\text{backpackers}) = \frac{M}{m}\sum_{j=1}^{m} I_{1,k,j}(\text{backpackers}) .$$

Similarly for stratum 2,

$$\hat{I}_{2,k}(\text{backpackers}) = \frac{M}{m}\sum_{j=1}^{m} I_{2,k,j}(\text{backpackers}) .$$

As for the single-access point section, let I_1 and I_2 represent the average daily recreational use of the trail for strata 1 and 2, respectively. The two-stage estimation formulas for I_1 and I_2 are,

$$\hat{I}_1(\text{backpackers}) = \frac{1}{n_1}\sum_{k=1}^{n_1} \hat{I}_{1,k}(\text{backpackers})$$

and

$$\hat{I}_2(\text{backpackers}) = \frac{1}{n_2}\sum_{k=1}^{n_2} \hat{I}_{2,k}(\text{backpackers}) .$$

Formulas for $\hat{\mathrm{T}}_1$(backpackers), $\hat{\mathrm{T}}_1$(backpackers), $\hat{\mathrm{I}}$(backpackers), and $\hat{\mathrm{T}}$(backpackers), remain unaltered from those given in the preceding section.

The variance of $\hat{\mathrm{I}}_1$(backpackers) comes from two sources: the first-stage component of variance that results from observing recreational use on only a fraction, n_1 of N, of all the days of the season; and the second-stage component of variance that results from observing recreational use at only a fraction, g, of all of the M access points for the site on any given sampling day.

Just as when working up the variance estimators for the single-stage sampling, it proves convenient to compute deviations from the average. In particular, suppose the deviation of $\hat{\mathrm{I}}_{1,k,j}$(backpackers) from the estimated average number of backpackers using the site on the kth day of sampling is expressed as,

$$\mathrm{DI}_{1,k,j} = \mathrm{I}_{1,k,j}(\text{backpackers}) - \frac{1}{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{I}_{1,k,i} \ .$$

Because there are m access points in the subsample and n_1 sample days in the first stage of sampling, there are $m \times n_1$ deviations of the form $\mathrm{DI}_{1,k,j}$.

Additionally, suppose the deviation of $\hat{\mathrm{I}}_{1,k}$(backpackers) from the estimated daily average recreational use per access point is expressed as,

$$\mathrm{D}\hat{\mathrm{I}}_{1,k} = \hat{\mathrm{I}}_{1,k}(\text{backpackers}) - \hat{\mathrm{I}}_{1,}(\text{backpackers}) \ .$$

Because there are n_1 sample days, there are n_1 deviations of the form $\mathrm{DI}_{1,k}$.

Similarly the deviations in stratum 2 are designated as,

$$\mathrm{DI}_{2,k,j} = \mathrm{I}_{2,k,j}(\text{backpackers}) - \frac{1}{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{I}_{2,k,i} \ ,$$

and

$$\mathrm{DI}_{2,k} = \hat{\mathrm{I}}_{2,k}(\text{backpackers}) - \hat{\mathrm{I}}_2(\text{backpackers}) \ .$$

There are $m \times n_2$ deviations of the form $\mathrm{DI}_{2,k,j}$ and there are n_2 deviations of the form $\mathrm{D}\hat{\mathrm{I}}_{2,k}$.

With these, the variance of $\hat{\mathrm{I}}$(backpackers) is estimated as,

$$\widehat{\mathrm{var}}_{2s}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right] = \left(\frac{1}{n_1} - \frac{1}{N_1}\right)\frac{\sum_{k=1}^{n_1} \mathrm{D}\hat{\mathrm{I}}^2_{1,k}(\text{backpackers})}{n_1 - 1}$$

$$+\frac{\mathrm{n}_1\mathrm{M}^2}{\mathrm{N}_1}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{\sum_{k=1}^{n_1}\sum_{j=1}^{m} \mathrm{DI}^2_{1,k,j}(\text{backpackers})}{n_1\left(m-1\right)} \ .$$

Notice that the size of the first stage sample, n_1 appears in the denominator of both variance terms, whereas the second-stage sample size, m, appears only in the denominator of the second. This implies that the variance of a two-stage sample reduces more quickly by enlarging the first stage sample than the second. In other words, if there is no difference in the cost of either sampling on an additional day or sampling an additional access point on a given day, sample on an additional day to increase the size of the first-stage sample.

For stratum 2,

$$\hat{\mathrm{var}}_{2s}\left[\hat{\mathrm{I}}_2(\text{backpackers})\right] = \left(\frac{1}{n_2}-\frac{1}{N_2}\right)\frac{\sum_{k=1}^{n_2} \mathrm{D}\hat{\mathrm{I}}^2_{2,k}(\text{backpackers})}{n_2-1}$$

$$+\frac{\mathrm{n}_2\mathrm{M}^2}{\mathrm{N}_2}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{\sum_{k=1}^{n_2}\sum_{j=1}^{m} \mathrm{DI}^2_{2,\mathrm{k},\mathrm{j}}(\text{backpackers})}{n_2\,(m-1)}$$

As before,

$$\hat{\mathrm{T}}_1(\text{backpackers}) = \mathrm{N}_1 \times \hat{\mathrm{I}}_1(\text{backpackers})$$
$$\hat{\mathrm{var}}\left[\hat{\mathrm{T}}_1(\text{backpackers})\right] = \mathrm{N}_1^2 \times \hat{\mathrm{var}}\left[\hat{\mathrm{I}}_1(\text{backpackers})\right]$$

$$\hat{\mathrm{T}}_2(\text{backpackers}) = \mathrm{N}_2 \times \hat{\mathrm{var}}\left[\hat{\mathrm{I}}_2(\text{backpackers})\right]$$
$$\hat{\mathrm{var}}\left[\hat{\mathrm{T}}_2(\text{backpackers})\right] = \mathrm{N}_2^2 \times \hat{\mathrm{I}}_2(\text{backpackers})$$

and

$$\hat{\mathrm{T}}(\text{backpackers}) = \hat{\mathrm{T}}_1(\text{backpackers}) + \hat{\mathrm{T}}_2(\text{backpackers})$$
$$\hat{\mathrm{var}}\left[\hat{\mathrm{T}}(\text{backpackers})\right] = \hat{\mathrm{var}}\left[\hat{\mathrm{T}}_1(\text{backpackers})\right] + \hat{\mathrm{var}}\left[\hat{\mathrm{T}}_2(\text{backpackers})\right]$$

$$\hat{\mathrm{I}}(\text{backpackers}) = \hat{\mathrm{T}}(\text{backpackers})/\mathrm{N}$$
$$\hat{\mathrm{var}}\left[\hat{\mathrm{I}}(\text{backpackers})\right] = \hat{\mathrm{var}}\left[\hat{\mathrm{T}}(\text{backpackers})\right]/\mathrm{N}^2$$

All the above can be developed in parallel fashion to estimate the amount of bikers recreating on the site.

> To compute estimates of recreational use by gender (or any other sociodemographic breakdown that is recorded), substitute the "number of individuals" described on the preceding pages with the "number of individuals of the gender of interest."

Construction of confidence intervals for population parameters based on estimates from two-stage and more complex designs is problematic. Since the introduction of resampling techniques, such as the jackknife and the bootstrap, there has been a tremendous amount of research into the construction of confidence intervals using these tools. Any useful treatment of this topic is outside the scope of the present report, however.

Example 4—In this example we expand on the previous three examples by supposing that the data provided in example 1 were collected at one randomly selected access point on each of the 7 days of sampling (4 weekdays and 3 weekend days). These data are augmented with those collected on the same sampling days but from a second randomly selected access point; that is, in this example, we demonstrate how to estimate backpacker and biker activity on the Spruce Run Trail based on observations from a two-stage sample of $m = 2$ access points. Purely for the sake of example, we assume that the trail has a total of $M = 10$ access points from which $m = 2$ were selected for each day of sampling. Hence, the second-stage sampling fraction is $g = 0.2$.

The sample data are presented in table 1. The first value in each cell was obtained from one of the randomly selected access points on that sample day, and the other was obtained from the second of the selected access point.

Table 1—Two-stage sample observations for each day and access point

Recreationist	-------- Stratum 1: weekdays --------				-- Stratum 2: weekends --		
Backpackers	13, 17	16, 17	12, 9	11, 21	28, 35	32, 32	22, 28
Bikers	8, 11	6, 7	8, 11	10, 8	10, 12	18, 17	14, 19

Based on these observations,

$$\hat{I}_{1,1}(\text{backpackers}) = \frac{10}{2}(13 + 17) = 150 .$$

The remaining calculations of $\hat{I}_{1,k}$ and $\hat{I}_{2,k}$ for backpackers and bikers are summarized in table 2.

Table 2—Daily averages of recreational use within strata

Stratum	Backpackers	Bikers
Stratum 1:		
$\hat{I}_{1,1}$	150	95
$\hat{I}_{1,2}$	165	65
$\hat{I}_{1,3}$	105	95
$\hat{I}_{1,4}$	160	90
Stratum 2:		
$\hat{I}_{2,1}$	315	110
$\hat{I}_{2,2}$	320	175
$\hat{I}_{2,3}$	250	165

These values, in turn, enable one to estimate the average daily recreational use by stratum; for example,

$$\hat{I}_1(\text{backpackers}) = \frac{1}{4}(150 + 165 + 105 + 160) = 145 .$$

The remaining trail estimates of recreational use, $\hat{I}_1$, and $\hat{I}_2$, for backpackers and bikers are summarized in table 3.

Table 3—Estimated daily recreational use for entire trail[a]

Stratum	Backpackers	Bikers
Stratum 1:		
$\hat{I}_1$	145.0	86.25
Stratum 2		
$\hat{I}_2$	295.0	150.0

[a]Averages per access point are obtained by dividing these values by M =10.

Because $\hat{I}_{1,1}(\text{backpackers}) = 15.0$ and $\hat{I}_{2,1}(\text{backpackers}) = 31.5$, the deviations of each access point observation can be computed from the daily average as:

$$DI_{1,1,1} = 13 - 15 = -2$$

$$DI_{1,1,2} = 17 - 15 = 2$$

$$DI_{2,1,1} = 28 - 31.5 = -3.5$$

$$DI_{2,1,2} = 35 - 31.5 = 3.5$$

and so on, as summarized in table 4,

Table 4—Deviations of each observation (shown in table 1) from daily average at each access point

Stratum	------------- Backpackers -----------		---------------- Bikers ---------------	
Stratum 1:	Access point $j = 1$	Access point $j = 2$	Access point $j = 1$	Access point $j = 2$
$DI_{1,1,j}$	-2.0	2.0	-1.5	1.5
$DI_{1,2,j}$	-0.5	0.5	-0.5	0.5
$DI_{1,3,j}$	1.5	-1.5	-1.5	1.5
$DI_{1,4,j}$	-5.0	5.0	1.0	-1.0
Stratum 2:				
$DI_{2,1,j}$	-3.5	3.5	-1.0	1.0
$DI_{2,2,j}$	0.0	0.0	0.5	-0.5
$DI_{2,3,j}$	3.0	-3.0	-2.5	2.5

The sum of these squared deviations for each stratum and recreational use are shown in table 5.

Table 5—Squared deviations of each observation from the daily averages, summed within strata

Stratum	Backpackers	Bikers
Stratum 1	Sum of squared deviations, $DI_{1,k,j}$	Sum of squared deviations, $DI_{1,k,j}$
	63.0	11.5
Stratum 2	Sum of squared deviations, $DI_{2,k,j}$	Sum of squared deviations, $DI_{2,k,j}$
	42.5	15.0

The deviations denoted by $D\hat{I}_{1,k}$(backpackers) are the differences between each $\hat{I}_{1,k}$ value and $\hat{I}_1$. For $D\hat{I}_{1,1}$(backpackers),

$$D\hat{I}_{1,1}(\text{backpackers}) = 150 - 145 = 5.0\ ,$$

and for $D\hat{I}_{1,1}$(bikers),

$$D\hat{I}_{1,1}(\text{bikers}) = 95 - 86.25 = 8.75\ .$$

The remaining deviations for backpackers and bikers are displayed in table 6.

Table 6—Deviations of daily estimates from overall daily average within strata

Stratum	Backpackers	Bikers
Stratum 1:		
$D\hat{I}_{1,1}$	5	8.75
$D\hat{I}_{1,2}$	20	-21.25
$D\hat{I}_{1,3}$	-40	8.75
$D\hat{I}_{1,4}$	15	3.75
Stratum 2:		
$D\hat{I}_{2,1}$	20	-40
$D\hat{I}_{2,2}$	25	25
$D\hat{I}_{2,3}$	-45	15

The sum of squared deviations are needed in the variance formulas, along with the values of n_1, N_1, n_2, N_2, m, M, and g as follows:

$$\hat{\mathrm{var}}_{2s}\left[\hat{I}_1(\text{backpackers})\right] = \left(\frac{1}{n_1} - \frac{1}{N_1}\right)\frac{\sum_{k=1}^{n_1} \hat{D}I^2_{1,k}(\text{backpackers})}{n_1 - 1}$$

$$+ \frac{n_1 M^2}{N_1}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{\sum_{k=1}^{n_1}\sum_{j=1}^{m} DI^2_{1,k,j}(\text{backpackers})}{n_1(m-1)}$$

$$= \left(\frac{1}{4} - \frac{1}{126}\right)\frac{2250}{3} + \frac{(4)(100)}{126}\left(\frac{1}{2} - \frac{1}{10}\right)\frac{63}{4(1)}$$

$$= 181.5 + 20.0 = 201.5\ .$$

Thus, the standard error of $\hat{I}_1$(backpackers), expressed as a percentage of $\hat{I}_1$(backpackers), amounts to $\left(\sqrt{201.5}\ /\ 145.0\right)100\% = 9.8\%$.

Similarly,

$$\hat{\mathrm{var}}_{2s}\left[\hat{I}_2(\text{backpackers})\right] = \left(\frac{1}{3} - \frac{1}{54}\right)\frac{3050}{2} + \frac{(3)(100)}{54}\left(\frac{1}{2} - \frac{1}{10}\right)\frac{42.5}{3(1)}$$

$$= 480.1 + 31.5 = 511.6\ .$$

The standard error of $\hat{I}_2$(backpackers), expressed as a percentage of $\hat{I}_2$(backpackers), amounts to $\left(\sqrt{511.6}\,/\,295.0\right)100\% = 7.7\%$.

Turning attention to the estimated variance of the estimates of recreational use by bikers, we get

$$\hat{\mathrm{var}}_{2s}\left[\hat{I}_1(\text{bikers})\right] = \left(\frac{1}{n_1}-\frac{1}{N_1}\right)\frac{\sum_{k=1}^{n_1}\hat{D}I^2_{1,k}(\text{bikers})}{n_1-1}$$

$$+\frac{n_1 M^2}{N_1}\left(\frac{1}{m}-\frac{1}{M}\right)\frac{\sum_{k=1}^{n_1}\sum_{j=1}^{m}DI^2_{1,k,j}(\text{bikers})}{n_1(m-1)}$$

$$= \left(\frac{1}{4} - \frac{1}{126}\right)\frac{618.75}{3} + \frac{(4)(100)}{126}\left(\frac{1}{2} - \frac{1}{10}\right)\frac{11.5}{4(1)}$$

$$= 49.9 + 3.7 = 53.6\ .$$

The standard error of $\hat{I}_1$(bikers), expressed as a percentage of $\hat{I}_1$(bikers), amounts to $\left(\sqrt{53.6}\,/\,86.25\right)100\% = 8.5\%$.

Similarly,

$$\hat{\mathrm{var}}_{2s}\left[\hat{I}_2(\text{bikers})\right] = \left(\frac{1}{3}-\frac{1}{54}\right)\frac{2450}{2} + \frac{(3)(100)}{54}\left(\frac{1}{2} - \frac{1}{10}\right)\frac{15.0}{2(1)}$$

$$= 385.6 + 11.1 = 396.7\ .$$

The standard error of $\hat{I}_2$(bikers), expressed as a percentage of $\hat{I}_2$(bikers), amounts to $\left(\sqrt{396.7}\,/\,150.0\right)100\% = 13.3\%$.

Lastly for backpackers,

$$\hat{T}_1(\text{backpackers}) = 126 \times 145 = 18{,}270 \text{ per season during the week,}$$

and

$$\hat{T}_2(\text{backpackers}) = 54 \times 295 = 15{,}930 \text{ per season during the weekend,}$$

and

$$\hat{T}(\text{backpackers}) = 18{,}270 + 15{,}930 = 34{,}200 \text{ per season,}$$

with a daily average of

$$\hat{I}(\text{backpackers}) = 34{,}200/180 = 190 \text{ per day during the season.}$$

For bikers, the corresponding estimates are as follows:

$$\hat{T}_1(\text{bikers}) = 126 \times 86.25 = 10{,}868 \text{ per season during the week,}$$

and

$$\hat{T}_2(\text{bikers}) = 54 \times 150 = 8{,}100 \text{ per season during the weekend,}$$

and

$$\hat{T}(\text{bikers}) = 10{,}860 + 8{,}100 = 18{,}968 \text{ per season,}$$

with a daily average of

$$\hat{I}(\text{bikers}) = 18{,}968/180 = 105 \text{ per day during the season.}$$

Sampling Strategy for Overnight Recreational Use Areas

For camping at sites where access is controlled, it is not likely that any form of sampling will be required. We have in mind, for example, a campground with a staffed access point or gate where registration occurs and fees are paid upon entry. For such a site, recreational use statistics can be compiled from the registration information.

For overnight areas where access is neither controlled nor monitored, we propose that campsites within the campground of interest be sampled. It is not sufficient to observe entry through the access points to the overnight recreation area because entry on a particular day may represent multiple days of recreation. For our objective of estimating daily recreational use, it is necessary to sample and observe that use among the campsites on the selected days of sampling.

Sampling should be conducted at a time of day when campers are most likely to be at the campsite and awake. We suggest early morning or just before sunset as the best times. Once the time of day for campsite sampling is determined, it should be used consistently throughout the sampling effort.

It is unlikely that the number of individuals in each group of overnight recreationists can be observed without risking substantial measurement error in some cases. We therefore propose switching the focus of estimation from the number of recreating individuals to the number of groups of overnight recreationists.

Small Campgrounds

If the number of campsites is small enough to permit all to be observed on a selected day of sampling, then the estimation formulas provided for transient-recreational-use sites with single access points are applicable. To be specific, we tailor and present those formulas here.

As established earlier, N, N_1, and N_2, represent, respectively, the number of days in the season of interest, in the subset of the season encompassing weekdays, and in the subset encompassing weekend days. Similarly, n, n_1, and n_2 represent the number of sample days.

To distinguish among the *number of groups* tallied on the different sample days in stratum 1, we introduce this notation:

$$
\begin{aligned}
G_{1,1} &= \text{observed number of groups on sample day 1 in stratum 1} \\
&\vdots \\
G_{1,k} &= \text{observed number of groups on sample day k in stratum 1} \\
&\vdots \\
G_{1,n_1} &= \text{observed number of groups on sample day } n_1 \text{ in stratum 1}
\end{aligned}
$$

Similarly for the tallies of groups in stratum 2, we have the following:

$$
\begin{aligned}
G_{2,1} &= \text{observed number of groups on sample day 1 in stratum 2} \\
&\vdots \\
G_{2,k} &= \text{observed number of groups on sample day k in stratum 2} \\
&\vdots \\
G_{2,n_2} &= \text{observed number of groups on sample day } n_2 \text{ in stratum 2}
\end{aligned}
$$

Estimating the average number of groups per day by stratum—An estimate of the average number of groups per day in stratum 1 is given by,

$$\hat{G}_1 = \frac{1}{n_1} \sum_{k=1}^{n_1} G_{1,k} = \frac{G_{1,1} + \ldots + G_{1,k} + \ldots + G_{1,n_1}}{n_1} ,$$

so that an estimate of the total number of groups using this site throughout the season in stratum 1 is this average multiplied by the number of days in the stratum,

$$\hat{T}_1 = N_1 \times \hat{G}_1 .$$

An estimate of the average number of groups per day in stratum 2 is given by,

$$\hat{G}_2 = \frac{1}{n_2} \sum_{k=1}^{n_2} G_{2,k} = \frac{G_{2,1} + \ldots + G_{2,k} + \ldots + G_{2,n_2}}{n_2} ,$$

and the estimated total number of groups in stratum 2 throughout the season is,

$$\hat{T}_2 = N_2 \times \hat{G}_2 .$$

Estimating the average number of groups per day for both strata—Because the total number of groups in both strata is the sum,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 ,$$

the average number of groups in both strata is estimated as,

$$\hat{G} = \hat{T} / N .$$

Variance estimation and confidence intervals—An estimate of the variance of $\hat{G}_1$ depends on the extent to which the individual daily tallies, $G_{1,1}$, $G_{1,2}$,...,G_{1,n_1} vary around $\hat{G}_1$. As a computational aid, we suggest that these deviations be computed separately:

$$\begin{aligned} DG_{1,1} &= G_{1,1} - \hat{G}_1 \\ &\vdots \\ DG_{1,k} &= G_{1,k} - \hat{G}_1 \\ &\vdots \\ DG_{1,n} &= G_{1,n_1} - \hat{G}_1 \end{aligned}$$

The estimated variance of $\hat{G}_1$ is designated as $\hat{\text{var}}(\hat{G}_1)$ and can be computed by the following formula:

$$\hat{\text{var}}\left(\hat{G}_1\right) = \left(\frac{1}{n_1} - \frac{1}{N_1}\right) \frac{\sum_{k=1}^{n_1} DG_{1,k}^2}{n_1 - 1} .$$

For the weekend stratum,

$$\hat{\mathrm{var}}\left(\hat{\mathrm{G}}_2\right) = \left(\frac{1}{n_2} - \frac{1}{N_2}\right)\frac{\sum_{k=1}^{n_2} \mathrm{DG}_{2,\mathrm{k}}^2}{n_2 - 1} ,$$

where $\mathrm{DG}_{2,\mathrm{k}} = \mathrm{G}_{2,\mathrm{k}} - \hat{\mathrm{G}}_2$, and so on.

For the estimate, $\hat{\mathrm{G}}$, pertaining to both strata combined,

$$\hat{\mathrm{var}}\left(\hat{\mathrm{G}}\right) = \frac{1}{\mathrm{N}^2}\left[\mathrm{N}_1^2 \times \hat{\mathrm{var}}\left(\hat{\mathrm{G}}_1\right) + \mathrm{N}_2^2 \times \hat{\mathrm{var}}\left(\hat{\mathrm{G}}_2\right)\right]$$

$$= \frac{1}{\mathrm{N}^2}\left[\hat{\mathrm{var}}\left(\hat{\mathrm{T}}_1\right) + \hat{\mathrm{var}}\left(\hat{\mathrm{T}}_2\right)\right].$$

Confidence intervals for average number of groups of recreationists visiting a site—An approximate 90-percent confidence interval for G_1 is constructed as,

$$\hat{\mathrm{G}}_1 \pm t_{df}^{0.90}\sqrt{\hat{\mathrm{var}}\left(\hat{\mathrm{G}}_1\right)} \text{ or } \hat{\mathrm{G}}_1 \pm \mathrm{E\%} ,$$

where $t_{\mathrm{df}}^{0.90}$ is the 90th percentile point from the Student's t distribution with $df = n_1 - 1$ degrees of freedom, and where,

$$\mathrm{E\%} = \left(t_{df}^{0.90}\sqrt{\hat{\mathrm{var}}\left(\hat{\mathrm{G}}_1\right)} / \hat{\mathrm{G}}_1\right)100\% .$$

For the corresponding estimate of the seasonal total number of groups of recreationists in stratum 1,

$$\hat{\mathrm{T}}_1 \pm \mathrm{N}_1 \times t_{df}^{0.90}\sqrt{\hat{\mathrm{var}}\left(\hat{\mathrm{G}}_1\right)} \text{ or } \hat{\mathrm{T}}_1 \pm \mathrm{E\%} .$$

An approximate 90-percent confidence interval for G_2 is constructed as,

$$\hat{\mathrm{G}}_2 \pm t_{df}^{0.90}\sqrt{\hat{\mathrm{var}}\left(\hat{\mathrm{G}}_2\right)} \text{ or } \hat{\mathrm{G}}_2 \pm \mathrm{E\%}$$

where $t_{df}^{0.90}$ is the 90th percentile point from the Student's t distribution with $df = n_2 - 1$ degrees of freedom, and where,

$$E\% = \left(t_{df}^{0.90}\sqrt{\hat{\text{var}}\left(\hat{G}_2\right)} / \hat{G}_2\right)100\% \ .$$

For the corresponding estimate of the seasonal total number of groups of recreationists in stratum 2,

$$\hat{T}_2 \ \pm \ N_2 \ \times \ t_{df}^{0.90}\sqrt{\hat{\text{var}}\left(\hat{G}_2\right)} \ \text{ or } \hat{T}_2 \ \pm \ E\% \ .$$

An approximate 90-percent confidence interval for G is constructed as,

$$\hat{G} \ \pm \ t_{df}^{0.90}\sqrt{\hat{\text{var}}\left(\hat{G}\right)} \ \text{ or } \hat{G} \ \pm \ E\% \ ,$$

where $t_{df}^{0.90}$ is the 90th percentile point from the Student's t distribution with $df = n_1 + n_2 - 2$ degrees of freedom, and where,

$$E\% = \left(t_{df}^{0.90}\sqrt{\hat{\text{var}}\left(\hat{G}\right)} / \hat{G}\right)100\% \ .$$

For the corresponding estimate of the seasonal total number of groups of recreationists in stratum 2,

$$\hat{T} \ \pm \ N \ \times \ t_{df}^{0.90}\sqrt{\hat{\text{var}}\left(\hat{G}\right)} \ \text{ or } \hat{T} \ \pm \ E\% \ .$$

Example 5—In this example we suppose that the small Pine River campground has been sampled on $n_1 = 6$ weekdays and $n_2 = 4$ weekend days. The observed number of groups of campers for each of the days is shown in table 7. For this example, as in the previous ones, we suppose that the season of interest comprises $N = 180$ days, of which $N_1 = 126$ are weekdays and the remainder, $N_2 = 54$, are weekend days.

Table 7—Number of groups of campers observed on each sampling day at the Pine River campground

Stratum 1: weekdays	Stratum 2: weekends
18, 23, 17, 19, 25, 26	31, 29, 26, 34

Based on these observations,

$$\hat{G}_1 = \frac{1}{6}(18 + 23 + 17 + 19 + 25 + 26) = 21.\overline{3}$$

and

$$\hat{G}_2 = \frac{1}{4}(31 + 29 + 26 + 34) = 30.0 \, .$$

The variance of $\hat{G}_1$ is estimated by,

$$\hat{\text{var}}\left(\hat{G}_1\right) = \left(\frac{1}{n_1} - \frac{1}{N_1}\right)\frac{\sum_{k=1}^{n_1} DG_{1,k}^2}{n_1 - 1} = \left(\frac{1}{6} - \frac{1}{126}\right)\frac{73.\overline{3}}{5} = 2.33 \, ,$$

and that of $\hat{G}_2$ by,

$$\hat{\text{var}}\left(\hat{G}_2\right) = \left(\frac{1}{n_2} - \frac{1}{N_2}\right)\frac{\sum_{k=1}^{n_2} DG_{2,k}^2}{n_2 - 1} = \left(\frac{1}{4} - \frac{1}{54}\right)\frac{34.0}{3} = 2.62 \, .$$

Using a value of $t_5^{0.90} = 2.015$ a 90-percent confidence interval for G_1 is constructed as,

$$21.\overline{3} \pm 2.015\sqrt{2.33} = 21.\overline{3} \pm 3.08 = 21.\overline{3} \pm 14.4\% \, ,$$

and with $t_2^{0.90} = 2.353$, a 90-percent confidence interval for G_2 is constructed as,

$$30.0 \pm 2.353\sqrt{2.62} = 30.0 \pm 3.81 = 30.0 \pm 12.7\% \, .$$

Seasonal totals are estimated by,

$$\hat{T}_1 = N_1 \times \hat{G}_1 = 126 \times 21.\overline{3} = 2688$$

and

$$\hat{T}_2 = N_2 \times \hat{G}_2 = 54 \times 30.0 = 1620 \ .$$

A 90-percent confidence interval for T_1 is constructed as,

$$2688 \pm 2.015\sqrt{36991} = 2688 \pm 387.5 = 2688 \pm 14.4\% \ ,$$

and a 90-percent confidence interval for T_2 is constructed as,

$$1620 \pm 2.353\sqrt{7650} = 1620 \pm 205.8 = 1620 \pm 12.7\% \ .$$

Large Campgrounds

Large campgrounds are those that have many campsites or the campsites are too dispersed to permit observation of all of them on a selected day of sampling. Consequently, a subsample of campsites is selected and observed.

Under this two-stage sampling setup, suppose that,

S = the number of campsites in the site ,
s = the number of campsites sampled ,

and that on each day of sampling the fraction ,

h = s/S ,

of campsites are observed.

The estimation formulas mirror those given for two-stage sampling of access points for transient-recreational-use sites. We have tailored them to the current problem as presented below.

Let

$G_{1,1,1}$ = observed number of groups on sample day 1 in stratum 1 at the first of the subsampled campsites

⋮

$G_{1,1,j}$ = observed number of groups on sample day 1 in stratum 1 at the jth of the subsampled campsites

⋮

$G_{1,1,s}$ = observed number of groups on sample day 1 in stratum 1 at the sth of the subsampled campsites

Similarly for the tallies of groups in stratum 2, we have,

$G_{2,1,1}$ = observed number of groups on sample day 1 in stratum 2 at the first of the subsampled campsites

⋮

$G_{2,1,j}$ = observed number of groups on sample day 1 in stratum 2 at the *j*th of the subsampled campsites

⋮

$G_{2,1,s}$ = observed number of groups on sample day 1 in stratum 2 at the *s*th of the subsampled campsites

Let $\hat{G}_{1,k}$ represent the estimated average number of groups for the *k*th sampling day in stratum 1, which is computed as,

$$\hat{G}_{1,k} = \frac{1}{h}\sum_{j=1}^{s} G_{1,k,j} \;.$$

Similarly for stratum 2,

$$\hat{G}_{2,k} = \frac{1}{h}\sum_{j=1}^{s} G_{2,k,j} \;.$$

The two-stage estimation formulas for $\hat{G}_1$ and $\hat{G}_2$ are

$$\hat{G}_1 = \frac{1}{n_1}\sum_{k=1}^{n_1} \hat{G}_{1,k}$$

and

$$\hat{G}_2 = \frac{1}{n_2}\sum_{k=1}^{n_2} \hat{G}_{2,k} \;.$$

Formulas for $\hat{T}_1$, $\hat{T}_2$, $\hat{G}$, and $\hat{T}$ remain unaltered from those given in the preceding section.

The variance of $\hat{G}_1$ and $\hat{G}_2$ comes from two sources: there is the first-stage component of variance that results from observing recreational use on only a fraction, n of N, of all the days of the season; and there is the second-stage component of variance that results from observing recreational use at only a fraction, h, of all the S campsites on the site on any given sampling day.

It is convenient to compute deviations from the average. In particular, suppose the deviation of $G_{1,k,j}$ from the estimated average number of groups using the site on the kth day of sampling is expressed as,

$$DG_{1,k,j} = G_{1,k,j} - \frac{1}{s}\sum_{i=1}^{s} G_{1,k,i} \ .$$

Because there are s campsites in the subsample and n_1 sample days in the first stage of sampling, there are $s \times n_1$ deviations of the form $DG_{1,k,j}$.

Additionally, suppose the deviation of $\hat{G}_{1,k}$ from the estimated seasonal average, $\hat{G}_1$, is expressed as,

$$D\hat{G}_{1,k} = G_{1,k} - \hat{G}_1 \ .$$

Because there are n_1 sample days, there are n_1 deviations of the form $D\hat{G}_{1,k}$.

Similarly, the deviations in the weekend stratum, stratum 2, are designated as,

$$DG_{2,k,j} = G_{2,k,j} - \frac{1}{s}\sum_{i=1}^{s} G_{2,k,j} \ ,$$

and

$$D\hat{G}_{2,k} = G_{2,k} - \hat{G}_2 \ .$$

With these, the variance of $\hat{G}_1$ is estimated as,

$$\hat{\text{var}}_{2s}\left(\hat{G}_1\right) = \left(\frac{1}{n_1} - \frac{1}{N_1}\right)\frac{\sum_{k=1}^{n_1} D\hat{G}^2_{1,k}}{n_1 - 1} + \frac{n_1 S^2}{N_1}\left(\frac{1}{s} - \frac{1}{S}\right)\frac{\sum_{k=1}^{n_1}\sum_{j=1}^{s} DG^2_{1,k,j}}{n_1\left(s-1\right)} \ .$$

Likewise,

$$\hat{\text{var}}_{2s}\left(\hat{G}_2\right) = \left(\frac{1}{n_2} - \frac{1}{N_2}\right)\frac{\sum_{k=1}^{n_2} D\hat{G}^2_{2,k}}{n_2 - 1} + \frac{n_2 S^2}{N_2}\left(\frac{1}{s} - \frac{1}{S}\right)\frac{\sum_{k=1}^{n_2}\sum_{j=1}^{s} DG^2_{2,k,j}}{n_2\left(s-1\right)} \ .$$

As before,

$$\begin{aligned}
\hat{T}_1 &= N_1 \times \hat{G}_1 \\
\hat{\text{var}}\left(\hat{T}_1\right) &= N_1^2 \times \hat{\text{var}}\left(\hat{I}_1\right)
\end{aligned}$$

$$\begin{aligned}
\hat{T}_2 &= N_2 \times \hat{\text{var}}\left(\hat{I}_2\right) \\
\hat{\text{var}}\left(\hat{T}_2\right) &= N_2^2 \times \hat{\text{var}}\left(\hat{I}_2\right)
\end{aligned}$$

$$\begin{aligned}
\hat{T} &= \hat{T}_1 + \hat{T}_2 \\
\hat{\text{var}}\left(\hat{T}\right) &= \hat{\text{var}}\left(\hat{T}_1\right) + \hat{\text{var}}\left(\hat{T}_2\right)
\end{aligned}$$

$$\begin{aligned}
\hat{G} &= \hat{T}/N \\
\hat{\text{var}}\left(\hat{G}\right) &= \hat{\text{var}}\left(\hat{T}\right)/N^2
\end{aligned}$$

As mentioned earlier, construction of confidence intervals for population parameters based on estimates from two-stage and more complex designs is problematic. Any useful treatment of this topic is outside the scope of this report.

Planning a Sample of Recreational Use

We have repeatedly expressed confidence intervals in the alternate form,

$$\textit{estimated quantity} + \textit{percentage},$$

where *percentage* is the length of the interval on either side of the *estimated quantity* and expressed as a percentage of the *estimated quantity*. For example, from the “Single Access Sites” subsection under “Sampling Strategy for Transient-Recreational-Use Areas,” section, we showed the 90-percent confidence interval for the average number of individual backpackers on weekdays as,

$$\hat{I}_1(\text{backpackers}) \pm E\% ,$$

where,

$$E\% = \left[\frac{t_{df}^{0.90}\sqrt{\hat{\text{var}}\left[\hat{I}_1(\text{backpackers})\right]}}{\hat{I}_1(\text{backpackers})}\right] 100\% .$$

One reason for preferring intervals expressed in this fashion is that many people can better assimilate the information when it is in relative terms (as a percentage of the estimated value) rather than absolute terms. Another reason is the natural link this expression provides to the formula used to estimate the sample size needed in the future.

If a past sample brought us to within ± 12 percent of the target, one might wish to take a larger sample next time to achieve a "margin of error" of only 10 percent. Or the decision could be made to conserve resources and select a smaller sample next time, but one that will have a good chance of achieving a "margin of error" of at least 15 percent, say. These types of comparisons presume that the level of confidence, here 90 percent, remains constant.

Generally speaking, if $E\%$ continues to represent the "margin of error" achieved from a sample, then $E^*\%$ will denote the "margin of error" desired next time. Many sources refer to $E^*\%$ as an "allowable error," where the implication is that a sample is desired that is sufficiently large to allow the estimate to be within $E^*\%$ of the targeted value at least $(1-\alpha)$ 100 percent of the time, where $(1-\alpha)$ 100 percent is the desired confidence level.

For sample-size planning, it stands to reason that the smaller value of $E^*\%$ stipulated, the larger the sample needed to achieve this level of precision.

The other factor (other than the desired confidence level and $E^*\%$) affecting the size of the sample needed is the variation observed among the sampled values. More heterogeneous populations will exhibit greater variation among observations, and more homogeneous populations will exhibit less variation.

The variance among population values is not at all influenced by sampling, as it is a feature of the population itself. Conversely, the variance among population values does affect the size of the sample needed to achieve a stipulated margin of error, $E^*\%$. A well-known result, that many will find intuitively reasonable, is that the greater the heterogeneity of the population, the larger the sample that will be needed to achieve the desired margin of error.

Data from prior samples can be used to estimate this among-observations variation on a relative or percentage basis with a statistic known as the coefficient of variation. Again referring back to the estimation of $\hat{I}_1(\text{backpackers})$ from the "Single Access Sites" section, the "coefficient of variation of $I_1(\text{backpackers})$" can be estimated by,

$$CV_1\% = \left[\frac{\sqrt{\frac{1}{n_1 - 1} \sum_{k=1}^{n_1} DI_{1,k}^2(\text{backpackers})}}{\hat{I}_1(\text{backpackers})} \right] 100\% \; .$$

Similarly, the coefficient of variation of the average daily number of individual backpackers on weekends can be estimated by,

$$\mathrm{CV}_2\% = \left[\frac{\sqrt{\dfrac{1}{\mathrm{n}_2 - 1} \sum_{k=1}^{\mathrm{n}_2} \mathrm{DI}_{2,k}^2(\text{backpackers})}}{\hat{\mathrm{I}}_2(\text{backpackers})} \right] 100\% \; .$$

The stipulated margin of error, E*%, and the estimated coefficient of variation can be used jointly to estimate the size of the samples, say n_1* and n_2*, that will be needed to achieve E*%. As an initial estimate of n_1*, the following simplified formula is sufficient:

$$n_1^* = \left(\frac{t_{df}^{0.90} \times \mathrm{CV}_1\%}{\mathrm{E}^*\%} \right)^2 \; .$$

For this initial calculation, it is sufficient to choose a t-value of 1.645. Or if a 95-percent confidence level is preferred, choose a t-value of 1.96.

On the basis of the initial estimate, reestimate n_1* by using the following formula with a value for $t_{\mathrm{df}}^{0.90}$ corresponding roughly to n_1* – 1 degrees of freedom:

$$\mathrm{n}_1^* = \left[\left(\frac{\mathrm{E}^*\%}{t_{df}^{0.90} \times \mathrm{CV}_1\%} \right)^2 + \frac{1}{N_1} \right]^{-1} \; .$$

An estimate of n_2* can be computed in the same way.

These estimates of n_1* and n_2* need **not** be fine tuned any further. The reason for this cautionary remark is that even with the best of planning and execution of the sample, there is no guarantee that a sample of size n_1* (or n_2*) will necessarily achieve the desired E*% margin of error, except in a probabilistic sense. There is little sense in fretting over alternative values of n_1* or n_2* that differ by one, two, or similar inconsequential values.

In situations where no prior survey information exists, a small pilot survey can be implemented to get a rough idea of the coefficient of variation values. Barring this, an educated guess can be substituted.

The preceding development shows the same stipulated margin error for both the weekday and weekend strata. Depending on the relative importance of reliable estimates from the two strata, the RM may prefer to stipulate a different E*% value for each stratum.

Example 6—Needed samples sizes for backpackers to achieve a $E^*\% = 10$ percent margin of error at the same 90-percent confidence level:

Using the information from examples 1 and 2 as a source of prior information about the natural variation of backpacking recreation on the Spruce Run Trail, one estimates $CV_1\% = 16.6\%$ and $CV_2\% = 11.8\%$.

Thus as an initial estimate of the sample size in stratum 1 needed to achieve a $E^*\% = 10$ percent margin of error is computed as,

$$n_1^*(\text{backpackers}) = \left(\frac{1.645 \times 16.6\%}{10.0^*\%}\right)^2 = 7.5 \text{ days} \ .$$

Adjusting the t-value so that it is appropriate for roughly 6 degrees of freedom ($t_6^{0.90} = 1.943$) one computes,

$$n_1^*(\text{backpackers}) = \left[\left(\frac{10.0\%}{1.943 \times 16.6\%}\right)^2 + \frac{1}{126}\right]^{-1} = 9.6 \simeq 10 \text{ days} \ .$$

An initial estimate of the sample size in stratum 2 needed to achieve a $E^*\% = 10$ percent margin of error is computed as,

$$n_2^*(\text{backpackers}) = \left(\frac{1.645 \times 11.8\%}{10.0^*\%}\right)^2 = 3.8 \text{ days} \ .$$

Adjusting the t-value so that it is appropriate for roughly 3 degrees of freedom ($t_3^{0.90} = 2.353$) one computes,

$$n_2^*(\text{backpackers}) = \left[\left(\frac{10.0\%}{2.353 \times 11.8\%}\right)^2 + \frac{1}{54}\right]^{-1} = 6.7 \simeq 7 \text{ days} \ .$$

Shown below is a parallel set of computations for bikers, designed to estimate the needed sample size to achieve a $E^*\% = 10$ percent margin of error at the 90-percent confidence level.

Using the information from examples 1 and 2 as a source of prior information about the natural variation of biking recreation on the Spruce Run Trail, one estimates $CV_1\% = 20.4\%$ percent and $CV_2\% = 28.6\%$ percent.

Thus as an initial estimate of the sample size in stratum 1 needed to achieve a E*% = 10% margin of error is computed as,

$$n_1^*(\text{bikers}) = \left(\frac{1.645 \times 20.4\%}{10.0^*\%}\right)^2 = 11.3 \text{ days} \quad .$$

Adjusting the t-value so that it is appropriate for roughly 11 degrees of freedom ($t_{11}^{0.90}$ = 1.818), one computes,

$$n_1^*(\text{bikers}) = \left[\left(\frac{10.0\%}{1.818 \times 20.4\%}\right)^2 + \frac{1}{126}\right]^{-1} = 12.3 \simeq 12 \text{ days} \quad .$$

An initial estimate of the sample size in stratum 2 needed to achieve a E*% = 10 percent margin of error is computed as,

$$n_2^*(\text{bikers}) = \left(\frac{1.645 \times 28.6\%}{10.0^*\%}\right)^2 = 22.1 \text{ days} \quad .$$

Adjusting the t-value so that it is appropriate for roughly 21 degrees of freedom ($t_{21}^{0.90}$ = 1.721) one computes,

$$n_2^*(\text{bikers}) = \left[\left(\frac{10.0\%}{1.721 \times 28.6\%}\right)^2 + \frac{1}{54}\right]^{-1} = 16.7 \simeq 17 \text{ days} \quad .$$

Complicating Factors

Although the above approach to planning the needed sample size focuses on controlling the margin of error within each stratum, it does not control the overall margin of error of the combined estimate: For example,

$$\hat{I}(\text{backpackers}) = \frac{N_1 \times \hat{I}_1(\text{backpackers}) + N_2 \times \hat{I}_2(\text{backpackers})}{N} \quad .$$

An alternate approach to sample size planning in a stratified random setup is to estimate the combined size of the stratified sample, say n*, and then to allocate n* to n_1* and n_2*, according to some plan. Cochran (1977: sec. 5.9) provides the details. The drawback of this approach is that control is conceded over the margin of error one wishes to achieve in the individual strata.

For stratified samples aimed at providing estimates of multiple characteristics, proportional allocation of the overall size of sample usually has much to recommend it. With proportional allocation, the size of the sample in each stratum is proportional to the size of the stratum, for example, $n_1^* \propto N_1$ and so on. The conventional wisdom is that with proportional allocation, none of the various characteristics will be estimated optimally, but on the other hand, the characteristics will be sampled with a frequency proportional to their occurrence in the population.

When sampling recreational use, it might be better to make the allocation proportional to something other than the size of the stratum. Our rationale here is that the use of N_1 and N_2 to determine the allocation of the sample makes sense if the proportions N_1/N and N_2/N are good indicators of the relative recreational use among strata. However, if most of the recreational use occurs on weekends, a better planning tactic may be to set n_2^*/n equal to an educated guess of the proportional recreational use on weekends, instead of gearing it to the size of the weekend stratum as measured by the number of days within it. The following example illustrates this tactic.

Example 7—Past experience suggests that 60 percent of the recreational use of a trail by mountain bikers occurs on the weekend. In this case, a future survey of overall size $n^* = 40$ days may be allocated such that there are $n_2 = 24$ days (0.6×40) of sampling on weekends, and $n_1 = 16$ days during the week, a tactic that shifts the sampling effort to include more of the heavy use days than would be included under a conventional proportional allocation.

For surveys in which multiple characteristics (for example, trail use estimates by mountain bikers, hikers, overnight campers, by weekday and by weekend) are being estimated, the planning of future sample sizes to achieve desired levels of precision is problematic because the optimal sizes of samples within strata for one characteristic will rarely match those of another characteristic, and indeed may be quite opposite. Unless the RM mounts separate samples for each recreational use of interest, this problem is insurmountable.

It is evident that for any recreation survey encompassing not only multiple strata but also multiple recreational uses, planning of the future sample is far from straightforward. The essential problem is that it may not be possible to stipulate an overall sample size adequate to provide sufficiently precise estimates of multiple characteristics. Compromises will have to be made to avoid being swamped in a maze of costly, unipurpose surveys. The terms of this compromise cannot be determined statistically. An oft-cited solution is to base the sample size on the requirements for precise estimation of the most important characteristic, and to let that implicitly determine the precision with which the remaining characteristics can be estimated.

Epilogue

The sampling strategies presented in this report surely will get adapted and refined as experience with them is gained. All the sampling designs described in this report are either one- or two-stage stratified random designs, which may strike some as quite complicated already! Especially if the results of recreational-use surveys will be subjected to public and official scrutiny, we urge that the sampling design stay comparatively simple, as these designs are. All these methods provide unbiased estimates of recreational use, at least in principle. Measurement error is a biasing agent: hence, our introductory caution about making extra efforts to minimize this source of inaccuracy. Measurement error notwithstanding, all the methods presented in this report are free of the bias that is insinuated as a result of self-selection into the sample.

Having championed simplicity, we should admit that it is easy to envision special-purpose surveys of a particular recreational use that could be made much more efficient by a more complex design or estimator. For example, past information or auxiliary information on file might be used to select the sample days differentially depending on anticipated use, or the information might be incorporated into a ratio or regression estimator. For such special-purpose surveys, something other than the simple strategies proposed here may be justifiable. Still, we reemphasize our introductory remark that for multiuse surveys, simplicity of design is a virtue worth striving for.

We have attempted to provide a set of basic tools applicable to a broad range of recreational sites to estimate numerous recreational uses. Certainly there will be combinations of sites and uses for which these methods are insufficient. As experience with these mavericks is accumulated, so too will our knowledge of how to design strategies to meet the two-fold objective of simplicity and efficiency.

Literature Cited

Cochran, W.G. 1977. Sampling techniques. New York: Wiley. 428 p.

Appendix

Sample sizes, n^*, needed to achieve $E^*\%$ margin of error at various levels of $CV\%$:

	CV%			
E*%	**10**	**25**	**50**	**100**
5	13	70	271	1082
10	9	19	70	271
15	18	10	32	122
20	10	9	19	70

This page has been left blank intentionally.
Document continues on next page.

This page has been left blank intentionally.
Document continues on next page.

This page has been left blank intentionally.
Document continues on next page.

Gregoire, Timothy G.; Buhyoff, Gregory J. 1999. Sampling and estimating recreational use. PNW-GTR-456. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 39 p.

Probability sampling methods applicable to estimate recreational use are presented. Both single- and multiple-access recreation sites are considered. One- and two-stage sampling methods are presented. Estimation of recreational use is presented in a series of examples.

Keywords: Sampling strategies, confidence intervals, stratified sampling.

The **Forest Service** of the U.S. Department of Agriculture is dedicated to the principle of multiple use management of the Nation's forest resources for sustained yields of wood, water, forage, wildlife, and recreation. Through forestry research, cooperation with the States and private forest owners, and management of the National Forests and National Grasslands, it strives—as directed by Congress—to provide increasingly greater service to a growing Nation.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, gender, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326-W, Whitten Building, 14th and Independence Avenue, SW, Washington, DC 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.

Pacific Northwest Research Station
333 S.W. First Avenue
P.O. Box 3890
Portland, Oregon 97208-3890

U.S. Department of Agriculture
Pacific Northwest Research Station
333 S.W. First Avenue
P.O. Box 3890
Portland, OR 97208

Official Business
Penalty for Private Use, $300

do NOT detach label