Part I–Designing A Sampling Plan

Importance of Estimating Recreation Use

The USDA Forest Service is the largest provider of public outdoor recreation in the United States. The objective of recreation management in the Forest Service is to provide a quality recreation experience for the public while limiting their impacts on other resources to an acceptable level. Recreation managers need information to plan, monitor, and evaluate existing and future programs.

The Value of Visitor Use Information

To provide excellent recreation opportunities, the Forest Service must know its customers. As a minimum, managers like to know how many people use sites and areas, when they are used, and in what activities people participate. Knowing national forest customers helps balance supply and demand for recreation with other resources, and enables us to provide what people want.

Important factors in assessing the success of Forest and District recreation programs include customer satisfaction, visitor motivation, desired experience, impacts of visitor use on other resources, and the number of visitors participating in a variety of activities. Demand for recreation opportunities is measured by number of visitors. Latent demand (unfulfilled demand) may be identified by opportunities not available or through off-site regional surveys.

Recreation use counts may include total number of visits, or may describe how different groups use the Forest. Information such as length of stay, activity participation, activity duration, distribution of visitors, and trends are often desired. Accurate and reliable recreation use information assists all levels of the organization in making management decisions about which recreation opportunities to supply. It takes time, effort, and desire to gather accurate recreation use estimates. The process of estimating use correctly is guided by:

- *a properly designed sampling plan*
- *proper selection of measurement techniques*
- *data collection*
- *analysis of data and conversion to usable information*

Relationship to Upward Reporting Requirements

Estimated recreation use numbers are reported in the Forest Service national database. This database was called Recreation Information Management (RIM) until 1989

when the system was no longer required by the Washington Office. In late 1994, a new national database that contains recreation, facility, utility, and travel route information was released for national implementation. This database, named Infrastructure, contains information about estimated recreation use by activity, site, and area; as well as information about facility condition, location, and access. A list of activities can be found in Chapter 6 of the Forest Service Infrastructure User Guide (1994).

Definitions

Visits– an entry of one person upon a National Forest site or area to participate in one or more activities for an unspecified period of time. For example, one person staying for 1 week engaged in 20 activities is still one visit.

Activity Occasions –one individual participating in one activity during any reasonable portion or all of one day. For example, one individual participating in three different activities during the day is recorded as three activity occasions.

Recreation Visitor Day (RVD)– recreation use of national forest sites or areas, which aggregates 12 visitor hours. This may consist of one person for 12 hours, 12 persons for 1 hour, or any equivalent combination of continuous or intermittent use by individuals or groups.

Recreation use information in the Infrastructure database must be reported by visits, activity occasions, and recreation visitor days. This means that once raw data are collected they must be converted to this framework.

State and private organizations often request recreation data from the Forest Service national database. This information assists them in responding to public demand and trends, and assists in their long-term state and county planning processes. The Forest Service Washington Office and the U.S. Congress use national recreation information including recreation visitor days (RVDs) to make budget and policy decisions. Congress is especially interested in nationally designated areas such as Wilderness and Wild and Scenic Rivers.

The reliability and accuracy of recreation use data have usually been acceptable for developed sites such as campgrounds and ski areas, but poor for dispersed, semi-primitive, and rural areas. Accurate recreation use information is needed to do a quality job in national forest recreation management.

Technically, there is a difference between accurate and reliable data. Accuracy refers to how close to the actual population the sample estimates are. Reliability refers to consistency of measurements. Accuracy may be affected by an improperly designed or executed sampling plan. Reliability may be affected by malfunctioning equipment, observer bias, and other measurement factors.

The Sampling Plan

Accurate recreation use data collection begins with a well designed sampling plan. This chapter explains how to design a District, Forest, or Area sampling plan. Statistical terms and procedures such as population, sampling methodology, sampling unit, standard deviation, and computation of sample size for a desired amount of accuracy are explained.

Recreation database systems such as Infrastructure allow reporting over 80 recreation activities. Managers can also report by large areas or smaller increments of a large area (such as trail, boat launch, picnic area). A manager could easily become overwhelmed trying to report 80 activities at 60 different locations! Indeed, doing this on a yearly basis would be time and cost prohibitive. It is also unnecessary.

By designing a District- or Forest-wide sampling plan, only certain sites and/or activities need to be measured each year. Also, sampling representative areas may need to occur only every few years. If managers notice significant changes in use or facility condition, resampling may be necessary. A written sampling plan will also provide consistency of data collection from year to year even if personnel change. This makes the data more reliable. Knowing when and where to sample is as important as knowing how to sample. Designing a statistically valid multi-year sampling plan to collect desired information can save District personnel money, effort, frustration, and time. The following steps will be used to design the Sampling Plan:

Step 1 - Map Local Knowledge
Step 2 - Define Sampling Objectives
Step 3 - Define Priorities
Step 4 - Define the Population, Sample, Sampling Frame, and Sampling Unit
Step 5 - Select Sampling Design
Step 6 - Determine Sample Size and Compute Summary Statistics

Step 1 - Map Local Knowledge

The first step in designing a District or Forest sampling plan is to assemble field people who are knowledgeable and a map of the area for which use estimates are needed. **Step 1** is needed to identify **where** you plan to collect the visitor use data. Other helpful tools include recreation maps, a

list of recreation activities or Recreation Resource Information System (RRIS) reports, Forest Plans, Wilderness Plans, and other special area plans.

Using local knowledge, list the major recreation activities that occur in the area. Next, identify on the map where the hubs of activity for each recreation category occur. You may wish to further refine this by identifying heavy, medium, and light use areas for each category. Color pencils on a clear overlay placed on top of the area map work well (Figure 1). If your map becomes too crowded, consider creating several overlays which may include:

- *summer recreation use*
- *winter recreation use*

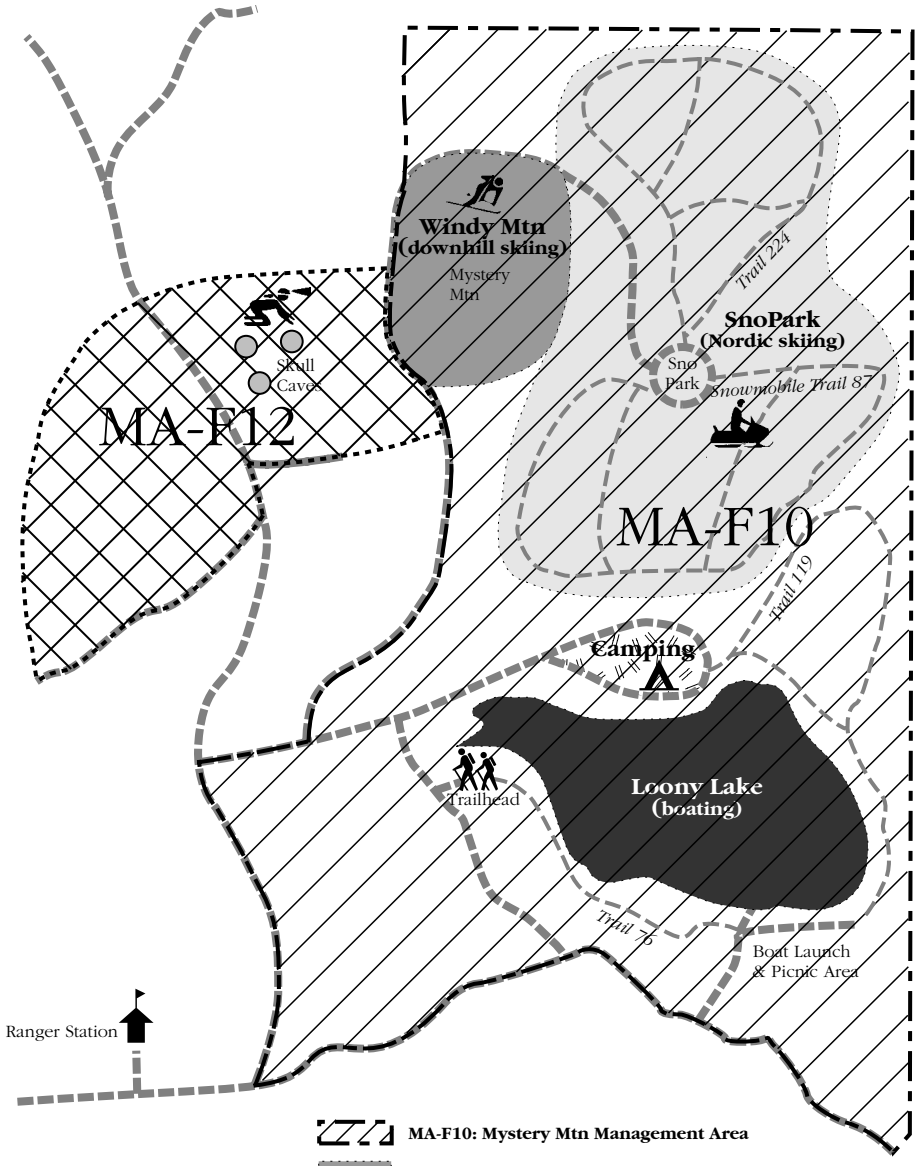


Figure 1.–Big Foot Ranger District Recreation Activity Map.

- *Recreation Opportunity Spectrum (ROS) class (urban, rural, roaded natural, semi-primitive motorized, semi-primitive non-motorized, and primitive)*
- *Forest plan management areas*

A Geographical Information System (GIS) will aid in assembling this information.

Step 2 - Define Sampling Objectives

There are many sources of information to help define recreation use sampling objectives. Defining the sampling objectives informs managers **why** visitor use data is being collected. Consult the Forest Monitoring Plan—what types of recreation use questions need to be answered? Consult your Ranger, Supervisor's Office, adjacent land management agencies, State and County cooperators, and the private sector—what types of information do they ask for? Though a separate issue and not part of the sampling objectives, during this stage it is possible to identify volunteers and partners in the private sector who may be willing to help collect data, especially if some of it is useful to them (Figure 2).

Big Foot Ranger District Recreation Use Sampling Plan—1995-2000

Objectives: Improve recreation use estimates of major activities on Big Foot Ranger District by the year 2000.

Identify the major recreation activities by type and amount of use that occur on Mystery Mountain Management Area.

Identify use at Windy Mountain Ski Resort. Is use exceeding design capacity? If so, what types of facility expansion would benefit both visitors, the local economy, and the environment?

With the recent designation of Skull Cave as a National Cave Resource, monitor spelunking use to determine effects of use on caves' ecosystem.

Identify whether or not visitors to Loony Lake are satisfied. Are the facilities adequate and do they appear well maintained? Did the area feel too crowded or were encounters within tolerable limits? Are customers willing to pay a fee to camp in the area? If so, what do they expect for their money that is not there now, and how much would they pay? Profile average Loony Lake visitor by age, sex, income, residence, number of previous visits, average length of stay.

Identify how much hunting use, driving for pleasure, and trail use occurs on District.

Responsibilities: District Resource Assistant will oversee implementation with the aid of seasonals and partners. Windy Mountain Ski Resort will pay 50% of cost for customer satisfaction survey and provide two personnel to be trained for interviews. Mulligan County Tourism Association will provide four people with vehicles to assist in hunting counts. District Engineers will provide and install traffic counters. District Resource Assistant will calibrate and validate counters. District Database Manager will obtain RECUSE computer programs, install, and assist seasonal employees in use of program.

District Recreation Planner will summarize data at end of each fiscal year, input data into Infrastructure database, and write summary report of data results that answers objectives identified above.

Figure 2. Example—Big Foot Ranger District Sampling Objectives.

Step 3 - Define Priorities

Due to limited time, money, and personnel, every question cannot be answered every year. Assume there are 3 to 5 years to answer all the important questions and list the first 3 years worth of priorities. Write the three-year sampling plan, listing activities to sample, locations to sample, and which year and season of the year these activities will be sampled (Figure 3).

Step 4 - Define the Population, Sample, Sampling Frame, and Sampling Unit

To understand why it is important to design a statistically sound sampling plan, a basic understanding of statistics is necessary. In very simple terms, statistics allow you to say the following: "I am 95 percent confident that between December 15, 1994, and March 30, 1995, between 1,250 and 1,340 people snowmobiled at Mystery Mountain." This is what statistics and sampling are all about. By identifying objectives, determining the population to describe, and identifying the sampling frame and sampling unit, the margin of error for recreation use reports is greatly reduced. A brief description of these important terms follows. Once they are understood, the sampling objectives can be very narrowly focused.

Population. To define the population, determine exactly what is to be measured. In most cases, the population will be the number of people per day in a specified time period. If the time period of interest is December 15

Big Foot Ranger District

Year	*(Where)* **Location**	*(What)* **Activity**	*(When)* **Time Period**	*(How)* **Sampling Design**	*(How many)* **Sample Size***	*(What needed)* **Equipment & Technique**
1	Mystery Mtn. MA-F10	Drive for Pleasure	May 15, 1995 Nov. 1, 1995	Stratified random	20 week days 15 weekend days	Traffic Counter With Observation
	Windy Mtn. Resort	Downhill Skiing	Nov. 25, 1995 Mar. 30, 1995	Systematic with random start	22 week days 18 weekend days	Lift Ticket Tallies
	Loony Lake	Boating	April 15, 1995 Sept. 15, 1995	Stratified Random	12 week days 10 weekend days	Traffic Counter With Camera
	Loony Trailhead	Hiking, Mtn. Biking, Horse	June 15, 1995 Oct. 30,1995	Stratified Random 9 weekend days	15 week days	Trail Registration With Observation
2	Skull Caves MA-F12	Spelunking	April 1, 1995 Nov. 15, 1996	Simple Random	32 days	Trail Registration With Traffic Counter
	Windy Mtn.	Cross-country Skiing	Nov. 15, 1996 Mar. 30, 1997	Stratified Random	22 days	Observation Counts
	Big Foot District	Hunting (big game)	Nov. 4, 1996 Dec. 3, 1996	Cluster	7 days	Observation Counts
	Loony Lake CG (boat launch and picnic area)	Determine major activities and customer satisfaction	May 15, 1996 Sept. 30, 1996	See CUSTOMER Report Card Instructions	180 people	Customer Report Card Survey
		Overnight Camping	May 15, 1996 Sept. 30, 1996	Stratified Random	30 days	Fee Receipts
3-5	Big Foot District	Driving for Pleasure	May 15, 1997 Nov. 1, 1997	To be determined	To be determined	To be determined
	Loony Lake	Sail Boarding	April 15, 1997 Sept. 15, 1997	To be determined	To be determined	To be determined
	Wasatch MA-F11	OHV use	June 1, 1998 Oct. 30, 1998	To be determined	To be determined	To be determined

** Sample size is for example only and has not been statistically calculated for this example.*

Figure 3.–Example–Three-Year Sampling Plan.

to March 30, the population is the average number of people over 108 days. Keep in mind, a population is not just a collection of sampling units. It is a collection of measurements on the sampling unit.

If management wants to know about customer satisfaction, the population of interest. Managers might want to know about both the people who came to snowmobile on Mystery Mountain and those people who chose not to snowmobile there. This question is complicated to answer because the population is difficult to find. How would you find and interview people who thought about going to Mystery Mountain but went elsewhere? In cases that will involve complicated, off-site sample designs, it is best to hire a professional consultant, or contact a university or research station.

Sample. Samples are subsets of the population that are used to generalize to the population as a whole. Therefore, the sample must be representative of the population. A representative sample means that the characteristics of the sample are a close approximation to the characteristics of the population. For example, if we are interested in estimating use on a trail, we would sample all types of uses, spread across the entire year to obtain both winter and summer trail activities.

Sampling Frame. A sampling frame is the complete list of all possible sampling units in the population of interest from which the sample is drawn (Bailey, 1982). A sample cannot be more accurate than the sampling frame from which it is drawn. In most cases on the Ranger District, the sampling frame will be the time period for which recreation use estimates are needed (i.e., April 1 to September 30). This can be further refined into smaller time periods. Many managers suggest breaking the days into three or four time periods during daylight hours.

Sampling Unit. The objects of study are the sampling unit. The sum total of all the units of analysis is called the population (Bailey, 1982). If we want to describe the average number and type of people using the trail from June 15 to September 30, the sampling unit is days. If we want to describe non-observable visitor characteristics such as customer satisfaction, income, or minority status, the sampling unit is visitors.

If done with care, sampling the population can be highly accurate, and save time and money. The more accurate the sample, the more likely it closely resembles the population. This means that every element of the population had the same chance to be in the sample and the probability of the element to be selected is known. For example, if the defined population was 100 days and the sample size was 25, then we know every element of the population had a 1 in 4 chance of being in the sample.

Step5-Select Sampling Design

You have now identified where and when sampling will occur. The next step is to decide **how** to do the sampling; sampling design is simply an estimator plus a sampling rule. Data can be collected in a statistically random pattern, or can be collected in a systematic pattern such as every fifth day. Keep in mind, managers are interested in the measurements taken on a sampling unit and not the sampling unit itself.

Unbiased Sampling Rules

Sometimes the size of the population to be sampled is known; other times it is not known. If the population to be sampled is the number of days in the winter season, then, the population size is known. When the size of the population is known, statistical computations are more accurate and probability of the sample representing the population can be computed. Unbiased sampling methods include random sampling, systematic sampling, stratified random sampling, and cluster sampling.

Random Sampling. Random sampling means that every observation in the population has an equal chance of being selected. To eliminate bias and make sample selection truly random, use of a random number table is suggested. The goal is to eliminate bias. For example, if the sampling unit is days and the sampling frame is all days from June 1 to September 30, but Sundays were not included because they are everyone's days off, then the sample is ***biased***, and not random. Assume that all days are included in the lottery, and as each day is selected for sampling it is not replaced. Sampling without replacement increases all the remaining elements chance of selection. For most Forest Service purposes this factor is negligible. Random sampling without replacement is called ***simple random sampling*** (Bailey, 1982).

There are many ways to conduct a simple random sample of the population. One can assign each element of the population a number, then use a random number table to select each element. If days are the sampling unit, and the sampling frame is days from December 15 to June 15, one could cut up a calendar and put each day into a hat. Then, shake the hat and pick numbers from the hat until all needed sampling days are chosen.

Systematic Sampling. Another method of sampling commonly used is systematic sampling, also called 1 in k sampling. For example, the sampling unit is fee receipts and there are 500 fee receipts in the population; the desired sample size is 100. To obtain 100 samples, k would be 5. After randomly selecting from the first five envelopes, every fifth envelope is chosen for the sample, until 100 envelopes are chosen. The problem here, however, is that the laws of random chance are not allowed to operate.

Systematic sampling is more dependent on the adequacy of the sampling frame than is simple random sampling. The list of the population to sample must not be biased by ordering, such as alphabetization, ranking, sorting, etc. Simple random sampling is generally preferred for accuracy over systematic sampling. However, systematic sampling is more practical in that it is less work, provides more information per dollar, is simpler to perform, and may reduce error.

Stratified Random Sampling. A stratified sample is obtained by separating the population observations into non-overlapping groups, called strata, then selecting a simple random sample from within each stratum. Each stratum must be mutually exclusive; no sampling unit appears in more than one group. Stratification is useful only when the variation within strata is less than the variation between strata.

Stratified sampling is very useful. For example, many campgrounds have a large difference between weekend and weekday use. By stratifying the population to be sampled into Monday through Thursday and Friday through Sunday, a more accurate and potentially useful recreation use estimate may be obtained. Once the desired sample size is determined, the sample days are allocated to each stratum. This is more accurate since there are more weekdays in the sampling time period than weekend days. A random or systematic sample within each stratum is then taken. The point to be made here is that the same level of reliability associated with estimates can be achieved with fewer observations, saving time and money.

Cluster Sampling. A cluster sample is a simple random sample in which each sampling unit is a collection, or cluster, of observations. Unlike stratified sampling, in which every stratum is sampled, cluster sampling involves visiting a subset of clusters or primaries. Cluster sampling, sometimes called area sampling, is generally used when it is impossible or impractical to construct a sampling frame in which the sampling units are the sampling elements themselves. For example, if the management objective is to estimate total recreation use on trails on a District, the trailheads within each management area may be clustered together. Of the 10 clusters resulting from this grouping, only four will be sampled. Meaning, every trailhead in each of the four clusters will be sampled on randomly selected sample days. The remaining six clusters will not be sampled at all. The sample results will be applied to all 10 clusters. This is a very good sampling design, but sample size and formulas are complicated.

The above example illustrates the weakness of this sample design. If the clusters are not homogeneous, there is the possibility that the sample may not be representative of the population as a whole. The advantage of this design is time savings.

Biased or Nonprobability Sampling Rule

When the size of the population is not known it is more difficult to predict how close the sample is to representing the entire

population. Past research has shown that District managers have the most trouble accurately measuring dispersed recreation use. Best guesses and local observation are usually used to estimate this type of use. When the population size is unknown, and the probability of the element being selected in the sample is unknown, this is called nonprobability sampling. Nonprobability sampling methods include convenience sampling, list sampling, purposive sampling, and snowball sampling. The point to remember is that nonprobability sampling methods are biased, and as a result, irreproducible.

Convenience or "Haphazard" Sampling. This sampling design is commonly used by Forest Service employees to estimate dispersed recreation use. It is also called accidental sampling. The data collector simply chooses the most convenient sampling unit. For example, to estimate recreation use at dispersed campsites, the number of people using these sites is recorded whenever the data collector happens to drive by the area on the way to other duties. There is no statistical randomness to the sample. It cannot be determined how close the sample estimates are to actual use of the dispersed sites.

Convenience sampling is generally undesirable. Visitors who are convenient to sample are the only ones who get into the convenience sample data. Visitors at more easily accessible trailheads, picnic areas, or boat launches may not be representative of all users.

List Sampling. List sampling is the nonprobability sampling equivalent of stratified sampling, with the added requirement that each stratum is represented in the sample in the same proportion as the entire population. For example, during a sampling frame of 120 days, there are 51 weekend days (42 percent of the sample frame) and 69 weekdays (58 percent of the sample frame). The determined sample size was 60 days. Sample days would be apportioned as 42 percent on weekends and 58 percent on weekdays, or 25 and 35 days respectively. However, now any 25 weekend days and any 35 week days would be sampled until the quota of sample days was filled. There is obviously room for sampling bias here, affecting the accuracy of the data collected.

Purposive Sampling. In purposive, or judgment sampling, investigators do not necessarily have a quota to fill, but neither do they just pick the most convenient sample day. Rather, the investigators use their own judgment about which sampling unit to choose from, and pick only those which best meet the purposes of the study.

The advantage of this sampling design is the use of local field knowledge. If the management objective is to define the characteristics of horseback riders on Big Foot Ranger District, the investigator may simply go to areas known to be used by horseback riders and select only horseback riders to interview.

Snowball Sampling. Snowball sampling has proved most useful for conducting observational research and community studies. It is conducted in stages. In the first stage, a few people with the desired characteristics are identified and interviewed. They are also asked to identify other people who have these same characteristics. The second stage of sampling involves interviewing these new people, who are also asked to identify other people. It snowballs from there.

If the management objective is to understand characteristics of off-road vehicle riders in the local community who use the District, this type of sampling design would be very useful. Once the first few riders are identified many more would be found. These surveys may reveal information not normally evident to managers.

Step 6 - Determine Sample Size and Compute Summary Statistics

Determining the appropriate sample size is one of the most critical steps in designing the sampling plan. In general, the larger the sample, the more likely it is that the measured use will represent the whole population. Regardless of the sample size, there is always some difference between the sample statistics and the population parameters they are estimating. This difference is called the sampling error. The key to estimating the most appropriate sample size is determining how much sampling error to accept. Factors to consider include cost and time required to achieve the desired sample size.

How many observations are needed in the sample to confidently generalize to the population? This depends on the management objectives and desired results. Recreation use estimates for planning and funding purposes should be accurate enough to compare with other use statistics. A confidence level of 67 percent may be sufficient for making broad policy decisions, while site specific or allocation decisions may require a 95 percent confidence level.

Sample size is derived through mathematical formulas. Do not be intimidated by these equations. Most people lose interest in executing a sound sampling plan at this point, fearing the torturous looking formulas. To encourage consistency of data collection and ease of use for the average recreation planner, a computer

program has been developed. The program, RECUSE, will allow you to design a simple or stratified random sample for several locations and any chosen time period. Once you have answered the simple questions, it will automatically select sample days. Included in the calculation will be the allowable standard error to be achieved, specified by the user. You can then collect and record data on these predetermined sample days. The data are then entered into the second part of the RECUSE program and statistics are calculated for you. The RECUSE program is stored in a Data General library. To retrieve the program, follow the instructions in Appendix J.

Even though the computer program will greatly simplify execution of the Sampling Plan, a basic understanding of some statistical terms is useful. These terms help explain how well the sample describes the population of interest.

Sample and Population Mean ($\overline{X}$ and μ)

Data are collected on randomly selected sample days. First, the number of people counted each sample day is recorded (X_1, X_2, X_3,...X_i). These numbers are then added together and divided by the total number of sample days. This gives the average number of people per day. This statistic is called the sample mean and is denoted $\overline{X}$ (bar *X*). The population mean, denoted by μ, is a single number that will not change. The sample mean varies depending on the sample days selected (Steel, Torrie, 1960).

The equation for the sample mean is:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots X_i}{n}$$

Where: n = **number of sample days**

X_i = **number of people on day** n

Variance

During a recreation use season, the number of people hiking a trail will vary from day to day. On some days the use will be higher than the average (mean), other days use will be lower. It would take more observations to get a good estimate of average hiking use per day if use varies from 5 to 200 people than if the range of use were 30 to 40 people.

The measure of variation most commonly used by statisticians is the variance. Variance is denoted by S^2. The square root of the variance is the standard deviation (S). Note: a variance is not useful for description. It is an intermediate step when calculating the standard deviation.

To calculate the estimate from a simple random sample (Freese, 1962):

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left[\sum_{i=1}^{n} X_i\right]^2}{n}}{(n-1)}$$

Where: S^2 = **sample estimate of the population variance**

X_i = **the value of unit** i **in the sample**

n = **the number of units observed in the sample**

Σ = **the sum of**

Standard Error and Confidence Limits

Like the individual observations in a population, sample estimators are subject to variation. The average (mean) from a sample of 10 days will seldom be the same as the average obtained from sampling a different set of 10 days. It is desirable to know how much variation is expected among sample estimators. Just as the variance and standard deviation are measures of variation among individuals in a population; the variance of the estimate and standard error describe the variability among sample estimators. The standard error is merely a standard deviation, but among estimates rather than among individual observations.

Variability of the estimates depends on the sampling method, sample size, and variability of units in the population. Given the standard error, it is possible to establish limits that suggest how close we might be to the parameter being estimated. These are called confidence limits. The values given by the sample estimates plus or minus one standard error are called the 67 percent confidence limits. This means we have a 1 in 3 chance of including the parameter being estimated. This confidence interval is appropriate for the broad range decision making and non-site specific planning. By spreading the limits we can be more confident that they will include the parameter being estimated. The estimate plus or minus 1.96 standard errors will give limits that will include the parameter unless a 1 in 20 chance has occurred. These are called the 95 percent confidence limits. For a 1 in 100 chance of including the parameter, we use the 99 percent confidence limit of 2.576 standard errors (Freese, 1962). This method of computing confidence limits will give valid estimates for samples of 30 observations or more.

Coefficient of Variation

The coefficient of variation (CV) is the ratio of the standard deviation to the average (mean). The coefficient of variation puts the expression of variability on a relative basis. The coefficient of variation simplifies the job of estimating and remembering the degree of variability of different populations. Such information is useful in planning a sample survey (Freese, 1962). For a sample with a mean of 10 ($\overline{X} = 10$) and a

standard deviation of 4 (S = 4), we would estimate the coefficient of variation as:

$$CV = \frac{S}{\overline{X}} = \frac{4}{10} = 0.4 \text{ or } 40\%$$

Student's *t*-distribution

This statistic is used to compute a confidence interval about a population parameter. A Student's *t* table is used to look up the correct value. To use this table, you need to know the sample size and the alpha (α) level. If *n* is not known, an estimate can be used.

Range

The range is the difference between the largest and smallest observation in the sample. For example, snowmobiling counts on sample days were 98, 120, 110, 75, 119, and 111. The range for the sample is 75 to 120 or 45 people.

The Confidence Statement

For each random sample and any level of probability, a confidence interval is established about the population mean. This is where one states: "Snowmobiling on Mystery Mountain from December 15 to March 30 was between 6,750 and 8,250 people with 95% confidence." Confidence limits (L_1:L_2) are formed by adding and subtracting twice the standard error to and from the sample mean.

The equation to solve is:

(L_1:L_2) = Estimate ± (*t*) (standard error of estimate)

Sample Size

There are many formulas available to determine sample size depending upon the sampling design (simple, stratified, etc.). Two formulas are given here, one for simple random designs and one for stratified random sample designs, assuming that the data are being sampled from a normal distribution. If you wish more detailed information on any of these statistics, refer to the literature cited.

Formula for Simple Random Samples (Steel and Torrie, 1960):

$$n = \frac{t^2 (CV)^2}{E^2}$$

Where:
- ***n*** = **sample size (note: use an initial guess for *n*, say n_0, to obtain *t* value)**
- ***t*** = **Student's *t* value from table**
- ***CV*** = **coefficient of variation in percent**
- ***E*** = **allowable error** $= \frac{\overline{X} - \mu}{X}(100)$

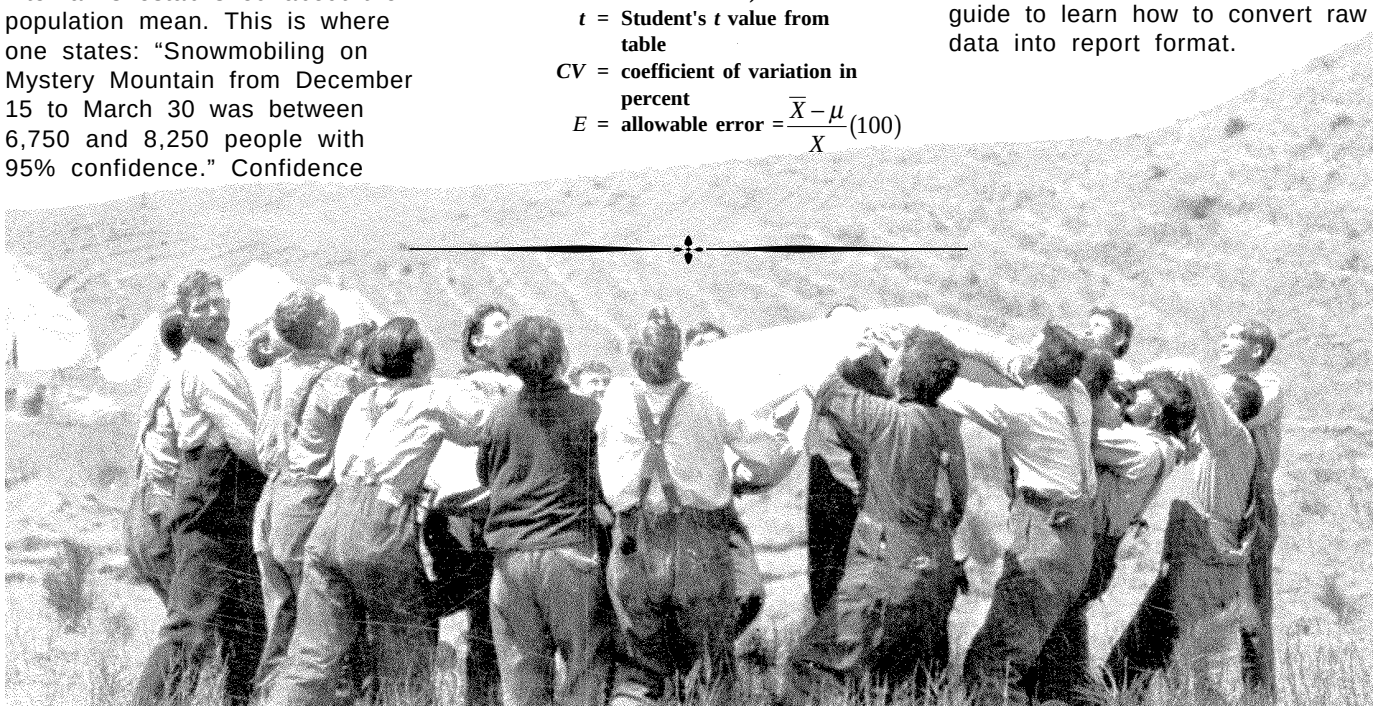
Formula for Stratified Random Samples:

$$n = \frac{t^2 \left[\sum_{i=1}^{L} W_i S_i \right]^2}{V_0 + \frac{1}{N} t^2 \sum_{i=1}^{L} W_i^2 S_i^2}$$

Where:
- ***n*** = **sample size**
- V_0 = **desired precision**
- S_i^2 = **estimate of variance in the *i* th stratum**
- S_i = **standard deviation in the *i* th stratum**
- ***t*** = **Student's *t* value from table**
- ***N*** = **total number of observations in population (or best guess)**
- N_i = **number of observations in the *i* th stratum**
- $W_i = \frac{N_i}{N}$
- ***L*** = **number of strata**

The Next Step

Now that the basic Sampling Plan is designed, read through Part II to select the most suitable sampling techniques and equipment. After the data are collected and analyzed read through the final section of this guide to learn how to convert raw data into report format.



End Part 3