USGS
science for a changing world

National Park Inventories: Statistical Methods

Paul Geissler
USGS Biological Resources

Contents

Introduction

This is one of a series of papers requested by the NPS Natural Resource Inventory and Monitoring Program to assist Parks with their natural resource inventories. It is intended to provide you with an overview of statistical methods for inventories, expanding on the information available in NPS-75 (NPS 1994). It provides an overview of the steps necessary for a natural resource inventory, concentrating on developing a species list for the Park and estimating species richness (number of species present), on estimating relative abundance, and on developing maps of the observed and predicted distributions of species. Planning and conceptual models, monitoring statistics, data quality and methods for specific taxa will be addressed in other papers in this series. You are encouraged to consult the references for more detailed information.

What is an inventory?

The *Natural Resources Inventory and Monitoring Guideline NPS-75* (NPS 1992) identifies inventory as the fourth step in the inventory and monitoring process (Figure 1). Before embarking on an inventory, you should (1) define management goals and objectives, (2) evaluate available data and (3) develop a conceptual model. These steps

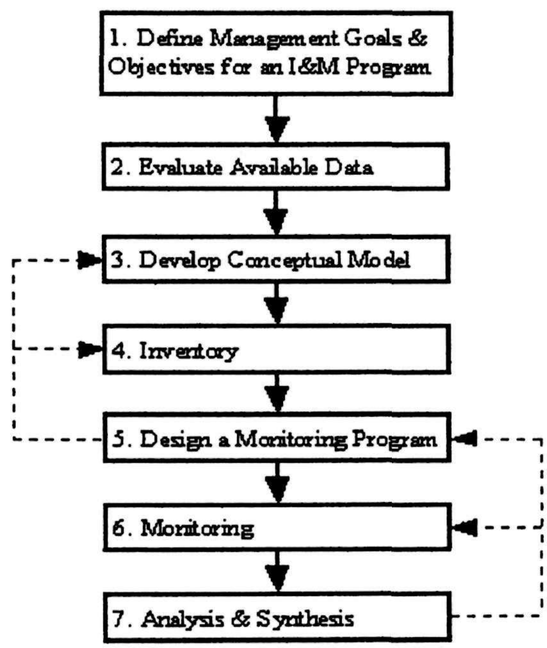

Figure 1 Conceptual Model of the National Park Service Inventory and Monitoring (I&M) Program (NPS 1992)

should provide a clear understanding of the purpose of the inventory and of how it will be used to further the objectives of the Park. These issues are discussed in another paper in this series on planning and conceptual models for National Park inventories. Another important issue not addressed in this paper is data quality and the extensive efforts required to edit and check data.

NPS-75 describes the inventory step (Phase I) for biological resources as follows [This paper will not address chemical and geophysical resources]:

Historical data base A comprehensive picture of the present state of knowledge of the Park's biota and related factors, and the history of these resources, should be compiled from existing source materials, with as much information as possible (or at least references to that information) transferred to computer data bases (text and map based). This information is typically summarized in manuscripts, published material, maps, photographs, museum records, knowledge of visiting scientists and park staff, and, in some cases, existing geographic information system data bases. Of special importance are records of rare but significant events (e.g., floods, fires, major landslides, volcanic eruptions). A bibliographic database referencing the body of published and significant unpublished documentation on park biota and related resources is essential. Critical review of this historical database is imperative to analyze the status and quality of existing natural resource information about the park and its environs.

Geography An important aspect of Phase I is the location of resources with respect to a modern, high resolution base map series (e.g., USGS 7.5' quads) and coordinate system (e.g., Universal Transverse Mercator). Whether or not a GIS database has been developed at this time, resources should be mapped to the accuracy and precision appropriate for eventual entry into a GIS database. The map selected should be the largest scale, i.e., have the highest spatial fidelity, appropriate to the subject. An accurate and comprehensive representation of the park landscape and its vicinity at a known time is compiled from aerial and ground photography, satellite data, and land survey. It is most important that the "study area" include surrounding lands that influence park biota. This "study area" is determined early in the inventory process because it is used to define the universe within which data are collected. A primary data theme to be mapped is vegetation communities and other land cover, usually derived from remote sensing prior to detailed fieldwork; such a map is invaluable in stratifying other kinds of sampling. It is important to use a standard method of mapping vegetation so that results can be compared with other parks and other areas and results can be aggregated with the results from other parks. For more information on vegetation mapping see the web pages of Federal Geographic Data Committee (FGDC) Vegetation Subcommittee (http://biology.usgs.gov/fgdc.veg/) and the Gap Analysis Program (GAP) (http://www.gap.uidaho.edu/gap/). Qualitative biotic community descriptions of the map classes are developed based on preliminary ground-truthing.

Species During Phase I (inventory), the presence of as many species as possible is established. The ultimate goal is to establish an accurate inventory of all major life forms within a park, but this is a long-range goal. The historical database can be a starting point, but an inventory must be based on empirical data, because Phase I information must be reliable and accurate enough to serve as the basis for an effective monitoring program. Development of inventory priorities may be based upon criteria including taxonomic group; legal status (e.g., endangered); endemism; non-native species; species legally or illegally taken; species characterizing communities; heroic species; and species described in enabling legislation. Ordinarily, all vascular plants and vertebrate animals are included in Phase I inventory, as well as other species of interest or importance (e.g., lichens in tundra communities; marine mollusks; gypsy moths and other non-native insect pests; invasive species, major pathogens). Distributions of plant and animal species of special interest or concern are also to be developed in Phase I but may be greatly expanded in Phase II (monitoring step) as new systematic surveys are conducted. The collection and storage of voucher specimens is mandatory to document inventory information.

Purposes and Objectives

It is important to recognize that inventories and monitoring have different purposes and objectives although the two phases may overlap. Statistical methods for monitoring will be covered in a separate paper.

Inventory Purposes

- Document the occurrence, location, and current condition of physical habitat (water, air, soil, etc.) and major associated taxa (biota)
- Identify locally rare or threatened and endangered species, locating fragile (or rare) ecosystems and potential "indicator species"
- Assess the full range of populations, ecosystem components, processes, and stresses from which to subsample for later long-term monitoring

Inventory Objectives

- Map of physical resources
- List of species occurring in Park
- Map of known and predicted locations for each species
- Estimates of relative abundance for each species

Monitoring Purposes

- Provide information necessary for rational management decisions
- Detect changes and quantify trends in resource conditions
- Establish the range of natural variation in populations, communities and ecosystems
- Provide information on linkages between changes in resource conditions and their causes

Monitoring Objectives

- Detect major threats to natural resources in time to take effective action
- Provide Park management with an information base to support their decisions
- Measure effects of management actions
- Suggest causal relationships
- Provide early-warning system to detect threats
- Assess status of Park natural resources
- Evaluate specific impacts

Inventory Methods

I will describe the extremes of planned and unplanned searches, but intermediate degrees of planning can be used.

Unconstrained Searches

These are unplanned and often unorganized searches for the occurrence of animal and plant species. Observers are free to look wherever and whenever they wish, taking full advantage of their skills. Unconstrained searches include incidental observations, made during other activities. Typically, there is no attempt to range the Park, search rare habitats, find specific species or taxa, or to make a complete list of species occurring in the Park. Unconstrained searches are easy to conduct and provide inexpensive, preliminary species list for planning a more rigorous inventory. However, without a planned, systematic search of at least a random sample of Park habitats, species may be missed because their habitat was not searched and a biased picture of the distribution and abundance of the Park's biota may result.

Advantages

- fastest way to compile a list of species occurring within the Park
- makes use of the available knowledge and the skills of staff biologists
- minimizes planning and record keeping activities
- provides estimates of species richness (total number of species present)

Disadvantages

- effectiveness and bias depends on the skills of the observers
- rare habitats and inaccessible areas may be missed
- systematic information about the distribution of species within the Park is not available
- does not provide estimates of changes in the size of populations

Completeness

The Lincoln-Peterson Index (Seber 1982) can be used to estimate species richness (number of species present including those you have not found yet) if two species lists are generated independently at two different times or by two different groups of observers. You can compare the number of species you found with the estimated number in the Park to see how complete your species list is. You will want to make separate lists for major taxa such as birds, fish and vascular plants.

Note that

$$\frac{\#\text{species on list 1}}{\#\text{species in Park}} = \frac{\#\text{species on both lists}}{\#\text{species on list 2}}$$

Say that the first list includes 80% of the species in the park. Then if you have an independent second list of species, then 80% of the species on the second list will also be on the first list.

Rearranging,

$$\begin{pmatrix}\#\text{species}\\ \text{in Park}\end{pmatrix} = \begin{pmatrix}\#\text{species}\\ \text{on list 1}\end{pmatrix} \frac{(\#\text{species on list 2})}{(\#\text{species on both lists})}$$

More sophisticated estimators are available which relax some of the assumptions and allow for more than two lists, but a complex estimator would negate the simplicity of this quick-and-dirty approach.

Example

Knowledgeable volunteers can often be used to help Park staff compile species lists. The following description comes from a newspaper report. See http://www.mp1-pwrc.usgs.gov/blitz/kenilw.html for more information.

Scientists Invade NE Park
24-Hour Survey Strives
To Log Nature's Diversity
By *D'Vera Cohn*
Washington Post Staff Writer
Thursday June 6 1996
Page J01The Washington Post

"The corner of Kenilworth Park in Northeast had been transformed into an outdoor laboratory: metal tables strewn with papers, laptop computers, microscopes and insect specimens pinned into display cases. The occasion was a 'bio-blitz,' in which more than 90 scientists tried to catalogue all living species, from bacteria to bats, in the 700-acre park across the Anacostia River from the National Arboretum. They were given 24 hours to spot, trap, net or identify by sound as many of the flora and fauna as they could in the park and Kenilworth Aquatic Gardens. Organizers hoped for at least 750. By their 5 p.m. deadline Saturday, they were close, having logged more than 70 species of birds, 150 plants and trees, a dozen species of fish and more than 500 insects, including 50 species of parasitic wasp, 11 butterflies and a dozen dragonflies. Also noted were a half-dozen mammals, from white-footed mouse to white-tailed deer. Not yet tallied were earthworms and snails, crustaceans, snakes, salamanders, frogs and more."This was not a game-show trick but a serious scientific enterprise. Bird and butterfly groups sponsor annual counts, but blitz organizers said theirs may be the first attempt to draw up a broad inventory of species. Such a catalogue, they said, could contribute to basic knowledge of how the region's biological community functions.

"Organizers said they also want to prove that the Anacostia River, given up for dead by many people, harbors a rich array of life. Park officials said they would use the blitz to update their species list and identify any rare ones that need protection. Park officials hope to use the species list 'to learn a lot about the biological diversity that we have'"

Stratified Random Sampling
These are planned searches, designed to optimize the sampling effort and cover representative areas of Park. Physical and logistical constraints on the search areas are recognized and incorporated into the sampling design, and knowledge of species' habitat requirements is used to optimize the search efforts.

<u>Advantages</u>:
- Provides maps of areas considered to be potential habitat, with relative abundance estimates and confidence limits
- Search time is optimized to provide as much information as possible, while covering rare and inaccessible habitats
- Estimates of species richness (total number of species present) do not require as strong assumptions as unconstrained searches

<u>Disadvantages</u>:
- Requires more planning and record keeping
- Requires use of advanced statistical techniques

<u>Steps</u>

1. Define management goals and objectives.

2. Evaluate available data.

3. Identify park-specific inventory objectives using a conceptual model of the major ecosystem components, processes, and stresses interacting in the park.

4. Decide on taxa and areas to be inventoried.

5. Determine appropriate habitat, times, and methods to inventory the selected taxa.

6. Divide the areas and times for the inventory into relatively homogenous strata, considering the taxa, habitat, and season, so that as much of the differences as possible are between and not within strata .

6.a. Different inventory methods are required for different taxa (e.g., breeding birds and stream macroinvertebrates).

6.b. Within each of these strata, the species lists may be quite different in different habitats. Consequently, it is advantageous to select a separate sample of locations within each habitat type (stratification). The precision of the estimates will be improved by making separate estimates for each habitat and then combining the estimates. The precision is judged by the variability of the samples, and the variability is reduced by separating the samples into homogenous strata. For example, you may want to conduct point counts for birds during the breeding season and select a separate sample in each of the major vegetation types, because the species lists and relative abundances are likely to be quite different in different habitats. However, you would want a separate sampling plan for stream macroinvertebrates, selecting separate samples by stream type and elevation.

7. Roughly map the occurrence of the major habitat types in the Park, for each of the taxa that will be inventoried separately (e.g., breeding birds and stream macroinvertebrates). Identify areas that will be excluded either because the taxon does not occur there or because the area cannot be inventoried. Note that no information will be available for the species occurring in excluded areas.

8. Identify potential search plots or paths for each of the taxa in each of the major habitat types using the maps developed in the last step.

8.a. In accessible areas with relatively homogeneous habitat, a systematic grid with a random start could be used.

8.b. For species with very specialized habitat requirements and a spotty distribution, potential plots may be limited to patches of appropriate habitat.

8.c. Particularly steep terrain in mountainous areas may constrain searches to existing trail or drainage corridors..

8.d. Each situation will be different, but one should identify a relatively large number of feasible sample plots or paths and then randomly select those to be inventoried.

8.e. Remember that your sample will only be as representative as the list (sampling frame) from which you selected the sample.

8.f. If you exclude sites (e.g., inaccessible ones) from the list, you may be excluding species.

8.g. Remember that random sampling does not mean that you have to sample at impossible locations. You can use your knowledge of the area to:

- Stratify habitat types
- Exclude areas that are unimportant or inaccessible (you will not get any

information about excluded areas.)

- Reduce travel time

9. Randomly select search plots or paths using a table of random numbers. Separately list the potential search plots or paths in each stratum, defined by different taxa and major habitat type. Separate estimates will be made for each stratum and then combined to make estimates for the Park. To maximize the precision of the estimates, the number of plots or paths selected in each stratum should be

- proportional to the size of the stratum,
- proportional to the stratum standard deviation, and
- inversely proportional to the square root of cost (time to travel to site and search it).

You will need at least two plots or paths in each stratum to estimate a variance. The optimal number in each stratum can be estimated as follows (Thompson 1992:108).

$$n_h = \frac{(c - c_0) a_h s_h / \sqrt{c_h}}{\sum_{k=1}^{L} \left(a_k s_k / \sqrt{c_k} \right)}$$

$$c = c_0 + c_1 n_1 + c_2 n_2 + \cdots + c_L n_L$$

where

c = total cost
c_0 = fixed cost (travel to site, etc.)
c_h = cost of inventorying a plot or path in stratum h, including travel to the site
n_h = number of independent plots or paths counted in stratum h
a_h = stratum area
s_h = stratum standard deviation of counts
L = number of strata

This allocation requires estimates of c_h and s_h, which may not be available before the inventory is conducted. However, approximate values may be available from other inventories. In any event, estimates from the inventory will help plan subsequent monitoring.

10. Conduct an inventory on the randomly selected plots or paths, recording habitat information on where a species was and was not found.

11. Estimate species richness (number of species present in the Park) using capture/recapture methods (Nichols and Conroy 1996, Boulinier, et al. In Press, see http://www.mbr-pwrc.usgs.gov/ software.html.and select COMDYN -community dynamics). These methods apply to both animal and plants. A "capture" or "recapture" in this context refers to the occurrence of a species on a plot or transect. The results of the inventory include a species list for each selected plot or path and in some cases the number of individuals detected. These estimators use the frequency distribution for the number of species observed on exactly 1 plot or path, exactly 2 plots or paths, etc. to estimate the number of species present but not observed on any plot or path. These estimators permit detection probabilities to vary among species. , Boulinier, et al. (In Press) recommend using their program COMDYN (Hines, et al. In Press) or the model M_h with the program CAPTURE (Rexstad and Burnham 1991). They also discuss the estimation of species richness in other sampling situations. Species richness should probably be estimated separately for each of the major taxa. Strata should be pooled because strata estimates cannot be added to obtain an estimate for the Park.

12. Select an additional sample and repeat the inventory if the desired proportion of species estimated to be in the Park (say 80%)

has not been detected in the initial inventory.

13. Estimate relative abundance for each species for the park as a whole and for each habitat type (stratum). Many inventory methods such as bird point counts provide an index of relative abundance as well as species lists. Relative abundance estimates and confidence intervals for each stratum are the mean and confidence limits for the counts on each plot or path in the stratum. The estimate for the Park weights the strata abundance estimates by the area of the strata.

$$\bar{y}_h = \sum_i y_{hi} / n_h$$

$$\bar{y} = \left(\sum_h a_h \bar{y}_h\right) / \left(\sum_h a_h\right)$$

$$s_h^2 = \frac{\sum_i (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

$$= \frac{\sum_i y_{hi}^2 - \frac{\left(\sum_i y_{hi}\right)^2}{n_h}}{n_h - 1}$$

$$V(\bar{y}_n) = s_h^2 / n_h$$

$$V(\bar{y}) = \left(\sum_h a_h^2 V(\bar{y}_h)\right) / \left(\sum_h a_h\right)^2$$

$$CI(\bar{y}) = \bar{y} \pm t_{\alpha,df} \sqrt{V(\bar{y})}$$

$$df = \sum_h (n_h - 1)$$

Where

i = plot or path
h = stratum
y_{hi} = count or measurement on plot or path in stratum h. If multiple counts are made, use the mean because they are not independent.
n_h = number of counts in stratum h
a_h = area of stratum h
$\bar{y}_h$ = relative abundance (mean count) for stratum h
$\bar{y}$ = relative abundance (mean count) for the Park
s_h^2 = variance of counts in stratum h
V() = variance of
CI() = confidence interval for
$t_{,df}$ = tabular t value for significance (type I error rate) and df degrees of freedom

14. Use discriminant analysis to predict species' occurrence. Use GIS to map areas of the park where each species is predicted to occur.

Discriminant analysis finds a function
$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots$
that maximized the F-test between the sites where the species was found and where it was not, where the x's are habitat variables (Overall and Klett 1971). This is the function that best separates or discriminates between sites where the species was found and not found. When scaled, the canonical discriminant function is distributed as a unit normal distribution. You can look up any site's score in a Z table to find the predicted probability that the species will be found at the site with the specified environmental values. If $z>0$, there is a greater than 50% change of finding the species. If the habitat variables are available in a GIS system, one can map the areas where the species is predicted to occur ($z>0$), expanding the usefulness of the inventory beyond the original sample.

15. Use inventory results to help plan monitoring program.

Examples

Species Richness

Species richness (number of species present in the Park) can be estimated on the Internet, using capture/recapture methods (Boulinier, et al. In Press). Using an Internet browser, find http://www.mbr-pwrc.usgs.gov/software.html. Scroll down to COMDYN -community dynamics (it is hidden in the grass) and click on "Run Program." Then scroll down to "Species Richness Computations" and click on "SPECRICH2." The program will ask for the number of sites/occasions (number of plots or paths) and then for

f(i) = number of species observed at exactly i sites/occasions and

n(i) = number of species observed at site/occasion i.

I used the following artificial data with 5 plots to illustrate the program:

Plot	Species Observed				
1	A	B			
2			C		
3		B		D	
4	A			D	
5	A				E

The program needs the following counts:

i	f(i)	n(i)
1	2	2
2	2	1
3	1	2
4	0	2
5	0	2

Here f(1)=2 because species C and E were observed on 1 plot, f(2)=2 because species B and D were observed on 2 plots, and f(3)=1 because species A was observed on 3 plots. Also n(1)=2 because 2 species (A and B) were observed on plot 1, n(2)=1 because 1 species(C)was observed on plot 2, etc.

The program estimated that 6 species were present with a standard error of 1.75. The 95% confidence interval would be between 5 and 9 species, as we know there are at least 5 species present because we observed them.

Relative Abundance

The SYSTAT program RABUND.SYC (see program listings) uses the following input data sets for a simple artificial example:

```
file: COUNT.SYD
HABITAT$ COUNT
 A         35
 A         26
 A         27
 A         38
 A         45
 B         18
 B         21
 B         18
 B         13
 B         16

file: AREA.SYD
HABITAT$ AREA
 A        1000
 B        2000
```

The output is listed below. There is a separate row (case) for each habitat type (stratum) and for the Park-wide estimates. The variables are:

```
HABITAT$   habitat (stratum)
NU1COUNT   number of independent
           counts
ME1COUNT   mean count(relative
           abundance)
CU1COUNT   upper confidence limit
CL1COUNT   lower confidence limit
SE1COUNT   standard error
AREA       area
DF         degrees of freedom

Case number   HABITAT$   NU1COUNT
 ME1COUNT     CU1COUNT   CL1COUNT
 SE1COUNT        AREA         DF

          1    A          5.00000
  34.20000    44.03191    4.36809
   3.54119  1000.00000    4.00000

          2    B          5.00000
  17.20000    20.86238   13.53762
   1.31909  2000.00000    4.00000

          3    Park      10.00000
```

```
 22.86667    26.26101   19.47232
  1.47196 3000.00000    8.00000
```

Discriminant Analysis
I used Fisher's iris data as an example, because it is supplied with SYSTAT and other statistical packages. Pretend that Fisher's iris data are inventory data, where SEPALLEN, SEPALWID, PETALLEN, and PETALWID are habitat variables measured at each site, and where a species was found for records for species 1 and not found for other records.

Running the SYSTAT program DISC.SYC (see program listings) resulted in a printout that included the following:

```
Canonical discriminant
functions
---------------------------


                   1
Constant   -1.56822
SEPALLEN    0.48138
SEPALWID    1.77043
PETALLEN   -1.63781
PETALWID   -0.41899

Canonical scores of group
means
------------------------------
N -2.207
Y  4.413
```

Thus the Canonical discriminant function z = -1.56822 + 0.48138 SEPALLEN + 1.77043 SEPALWID - 1.63781 PETALLEN - 0.41899 PETALWID. Substituting the data from the first record (site) z = -1.56822 + 0.48138(5.1) + 1.77043 (3.5) - 1.63781(1.4) - 0.41899(0.2) = 4.41

Depending on how the data are set-up, the discriminant analysis can either predict the presence or absence of a species. Check the canonical scores of the group means to see whether the program is predicting the presence or absence. In this example the Y (found=yes) mean is positive (4.413), so it is predicting species presence. If the program is predicting the absence of the species, change the sign of the discriminant function.

If the canonical score is positive, the species is predicted to occur at that location. Looking this value up in a Z (normal cumulative distribution) table, the probability of the species occurring at the location is near 1. If the habitat variables are available in a GIS system, one can map the areas where the species is predicted to occur ($P>0.50$, $z>0$).

References

Boulinier, T., J.D. Nichols, J.R. Sauer, J.E. Hines, and K.H. Pollock. In press. Estimating species richness to make inference in community ecology: the importance of heterogeneity in species detectability as shown from capture-recapture analyses of North American breeding bird survey data

Hines, J.E., T. Boulinier, J.D. Nichols, J.R. Sauer and K.H. Pollock. In Press. COMDYN: software to study the dynamics of animal communities using a capture-recapture approach.

Levey, P.S. and S. Lemeshow. 1991. Sampling of populations: methods and applications. Wiley, New York.

National Park Service (NPS). 1992. Natural resources inventory and monitoring guideline. NPS-75

Nichols, J.D. and M.J. Conroy. 1996. Estimation of species richness. Pages 226-234 *in* D.E. Wilson, F.R. Cole, J.D. Nichols, R. Rudran and M. Foster, editors. Measuring and monitoring

biological diversity. Standard methods for mammals. Smithsonian Institution Press, Washington, D.C.
Overall, J.E. and C.J. Klett. 1971. Applied multivatiate analysis. McGraw-Hill.
Rexstad, E. and K. Burnham. 1991. Users guide for interactive program CAPTURE. Abundance estimation of closed animal populations. Colorado State University, Fort Collins, Colorado
Scheaffer, R.L., W.I. Mendenhall and L.Ott. 1996. Elementary survey sampling. Duxbury.
Seber, G.A.F. 1982. Estimation of animal abundance and related parameters. Macmillan, New York.
Snedecor, G.W. and W.G. Cochran. 1980. Statistical methods. IA State Univ. Press, Ames,, Chapter 21.
Steel, R.G.D. and J.H. Torrie. 1980. Principles and procedures of statistics. McGraw-Hill, New York. Chapter 25.
Thompson, S.K. 1992. Sampling. Wiley, New York.

Please send comments and suggestions to:
Paul Geissler
USGS Biological Resources
12100 Beech Forest Road
Laurel, MD, USA 20708-4038
Tel. 301-497-5780, FAX -7584
E-mail: Paul_Geissler@usgs.gov
Revised August 30, 1998

Program Listings

These programs and sample data may be downloaded from
http://www.mp1-pwrc.usgs.gov/fgim/npsim.htm

RABUND.SYC - Relative Abundance

```
remark RABUND.SYC
remark SYSTAT program for relative abundance estimation
use
basic
new
remark files starting with "tmp" are temporaty & can be deleted
remark "list" statements (except last) can be removed to save paper
remark response variable is named "count", change as needed
remark habitat (stratum) identification variable is named "habitat$"
remark area variable is named area, change as needed.
remark change next statement to your data file
use e:\work\nps\istat\count
save tmp
sort habitat$
list
remark ---------------
remark output variables for count are named:
remark mean=me1count,number=nu1count,standard error=se1count
remark lower and upper confidence limits=cl1count,cu1count
stats
use tmp
by habitat$
save tmp1 / AG
remark change 95% confidence interval in next statement
stats count / mean sem n conif95
```

```
remark ---------------
remark change next statement to your habitat area file
use e:\work\nps\istat\area
save tmp
sort habitat$
remark ---------------
save tmp2
merge tmp1 tmp / habitat$
list
remark ---------------
use tmp2
save tmp3
let e=area*me1count
let df=nu1count-1
let v=area*area*se1count*se1count
list
run
remark ---------------
use tmp3
save tmp4
stats / sum
remark ---------------
use tmp4
save tmp5
let habitat$="Park"
let me1count=e/area
let se1count=sqr(v/area/area)
remark change 95% confidence interval in next 2 statements to 1-alpha/2
let cu1count=me1count+se1count*tif(0.975,df)
let cl1count=me1count-se1count*tif(0.975,df)
list
run
remark ---------------
save e:\work\nps\istat\est
append tmp3 tmp5 / union
drop e,v,statistc$
list
```

DISC.SYC - Discriminant Analysis

```
remark - file: DISC.SYC - Sample SYSTAT program to estimate predict
species
distributions.
remark - Change the paths on the next 2 statements to those on your
computer.
basic
Use 'd:\systat\data\iris.syd'
save e:\work\nps\istat\iris.syd
remark - Pretend that Fisher's iris data are inventory data, where
SEPALLEN,
SEPALWID,
remark - PETALLEN, and PETALWID are habitat variables, and a species was
found on
remark - records for species 1 and not found on other records.
Remark - Although it is not realistic, the data are easily available.
if species=1 then let FOUND$='Y'
```

```
else let FOUND$='N'
run
discrim
save tmp / scores,data
model FOUND$ = SEPALLEN SEPALWID PETALLEN PETALWID
print medium
estimate
use tmp
save e:\work\nps\istat\discrim.syd
remark - prob is probability of detecting species at site
let prob=zcf(score(1),0,1)
list
run
```