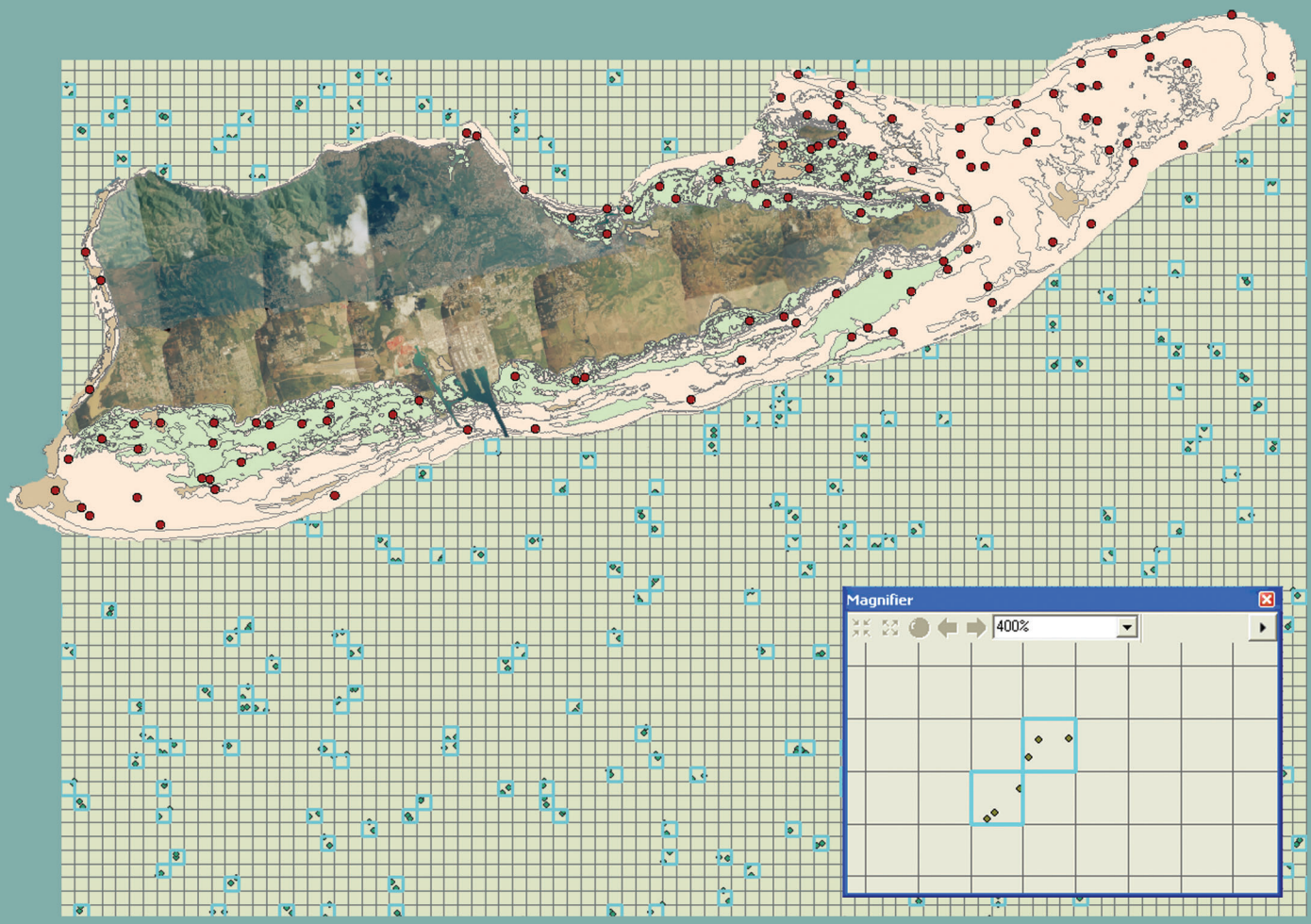
A User Guide for NOAA's CCMA Biogeography Team's Sampling Tool

A GIS Application for ArcGIS



May 2007

Eric Finnen and Charles Menza

NOAA/NOS/NCCOS/CCMA/Biogeography Team

OVERVIEW

NOAA's Center for Coastal Monitoring and Assessment Biogeography Team's Sampling Tool (Sampling Tool) provides users a means to efficiently develop sampling strategy for a population, be it people, animals, objects or processes, in a GIS environment. The Sampling tool was created for area-based sampling, that is, when the population and component sampling units are defined by known dimensions. Sample selection can be achieved using either a dataset defining all sample units or a defined sample area. The latter provides a means to perform probabilistic sampling when the definition of sampling units is impractical.

The Sampling Tool has two main functions: 1) to help select a sample from a population, and 2) to perform sample design analysis. When both of these functions are combined in an iterative manner, the tool effectively and simply achieves the goal of designing sample surveys to obtain accurate, high-precision estimates of population metrics at a minimum of cost using information gained from pervious surveys.

The choice of which design to use depends on sampling objectives, cost, expertise, and available data. Several sampling designs are available, including simple random sampling, stratified random sampling and multi-stage sampling. Data requirements for sample selection can be as simple as a polygon defining the area to be surveyed or as complex as a sample frame polygon dataset with strata definitions and strata variance estimates.

Sampling design analysis is used to determine the appropriate number of samples required for specific sampling objectives. Four distinct definitions of a sampling objective are provided, including a coefficient of variation and a confidence interval. For each definition, sample requirements are provided for a range of objectives. The results provide a quick and simple means to balance the advantages of increasing the number of samples with relaxing sampling objectives. Since sample requirements can be used as a sample design performance measure, results from multiple sampling designs can be used to compare design effectiveness.

Sampling design analyses require spatially explicit survey data that has already been collected or similar estimates. Currently, the tool can use point data to determine mean and variance estimates for a population or strata. These data are then used for further analyses.

Examples in the user manual are taken from reef fish surveys collected by the biogeo team. This survey uses a stratified random sampling design where strata are defined by a benthic habitat map to efficiently sample reef fish populations. It is important to note that this tool does not need to be limited to ecological application for which it was first created. The functions of this tool are based on well established sampling doctrine and may be used for any area-based sampling program.

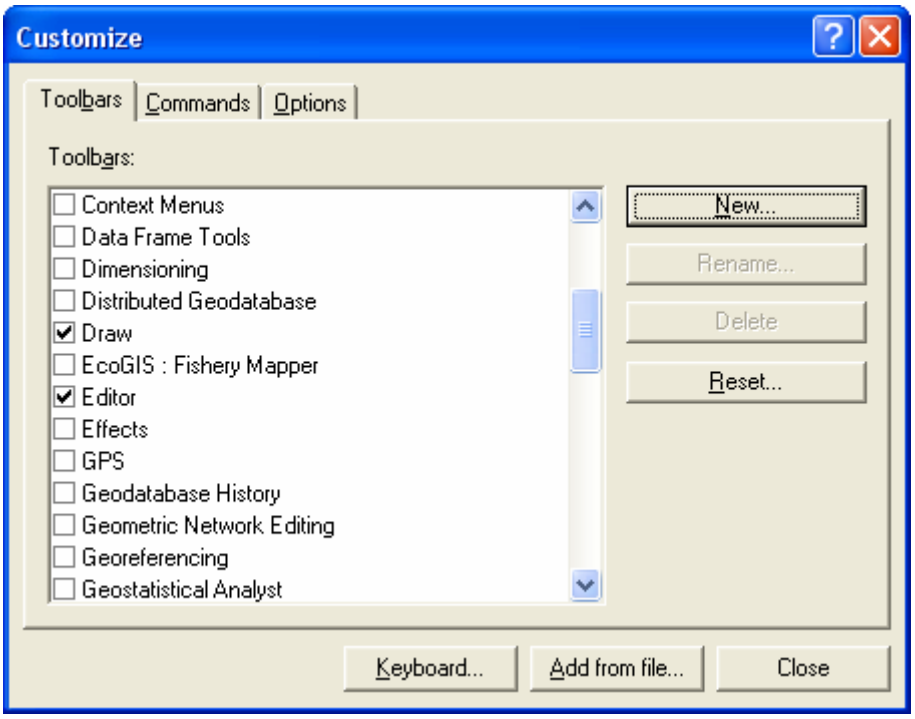
This tool is currently available for download on the ESRI website at http://arcscripts.esri.com/details.asp?dbid=15080.

REQUIREMENTS :
ArcGIS 9.2

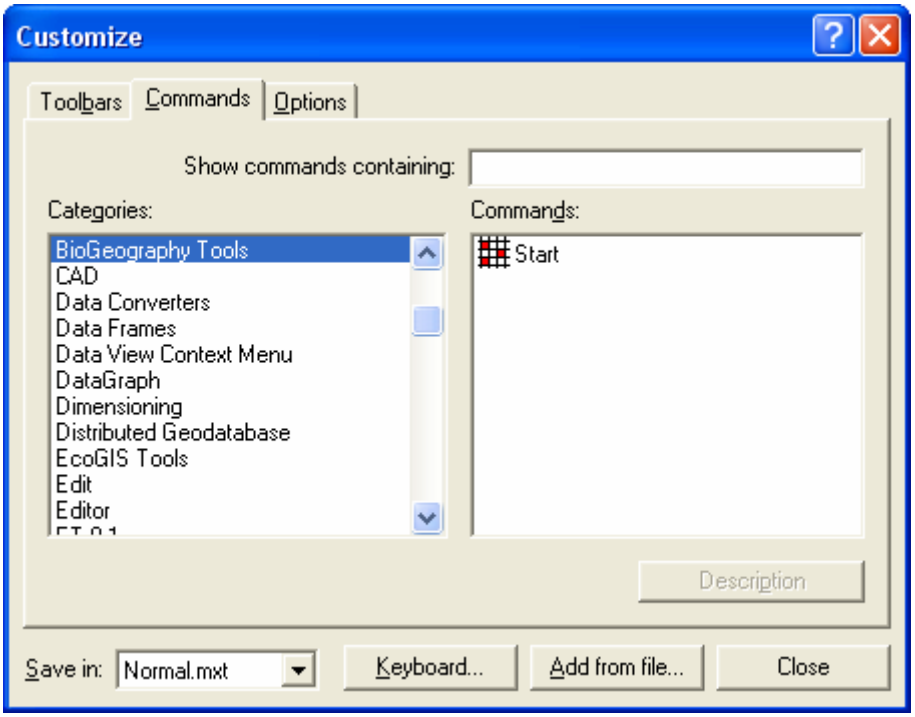
ArcGIS 9.2 Setup

Earlier versions of this application were a ArcGIS projects it is now currently a distributable DLL. The current version of the application is Sampler.dll newer versions will be named Sampler2.dll, Sampler3.dll etc.

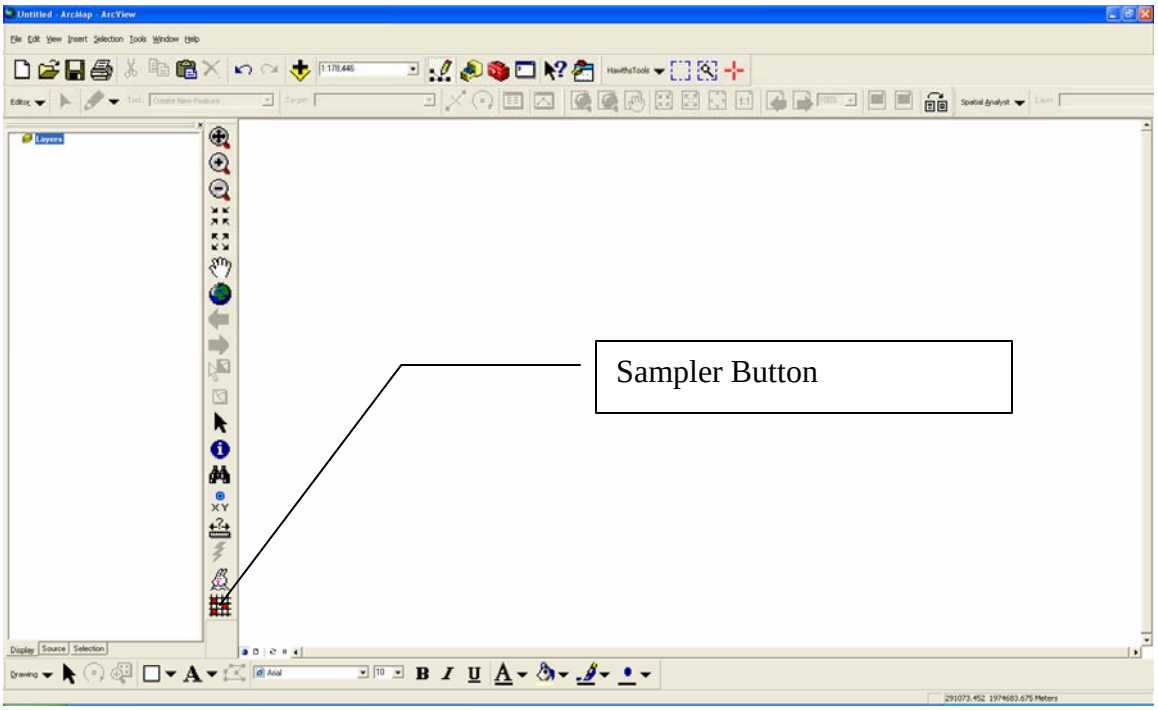
The zip file that was downloaded contains a dll that needs to be downloaded and placed in a permanent location on the hard drive. Click the TOOLS menu and navigate down to the CUSTOMIZE option, and click the ADD FROM FILE button. Navigate to the location on the hard drive where the dll has been unzipped and select it. This registers the dll within ArcGIS.



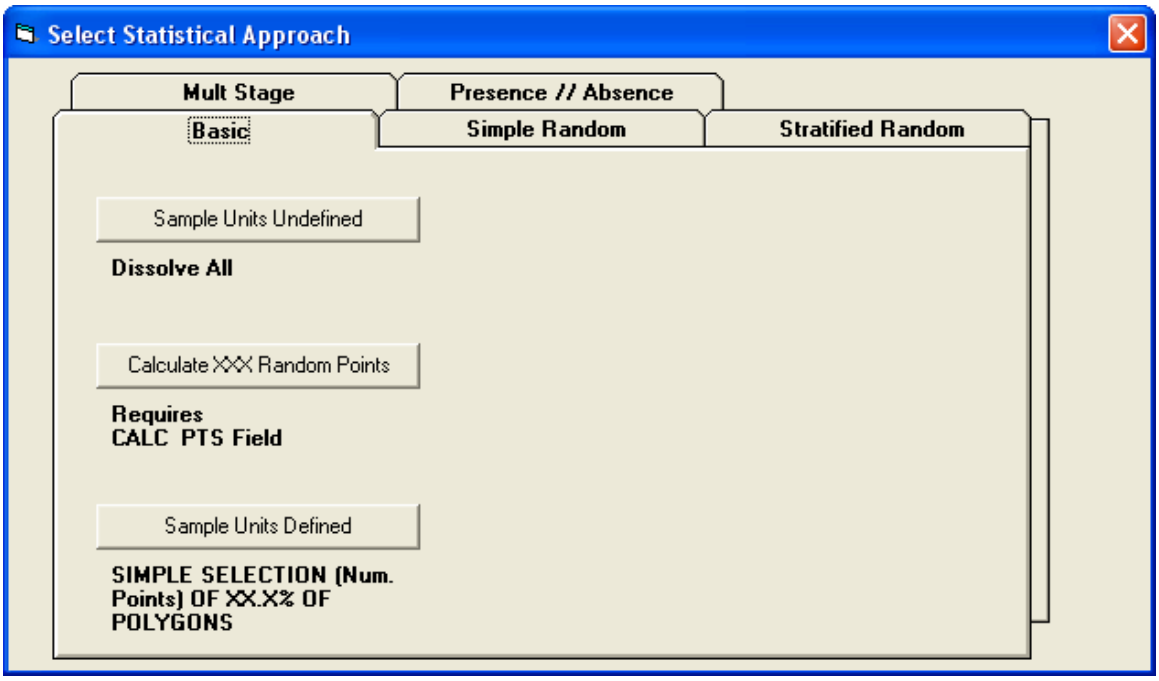
The customize menu should still be open, click the COMMANDS TAB on the interface . Then navigate to the Biogeography Tools Category on the left panel. This reveals the following interface



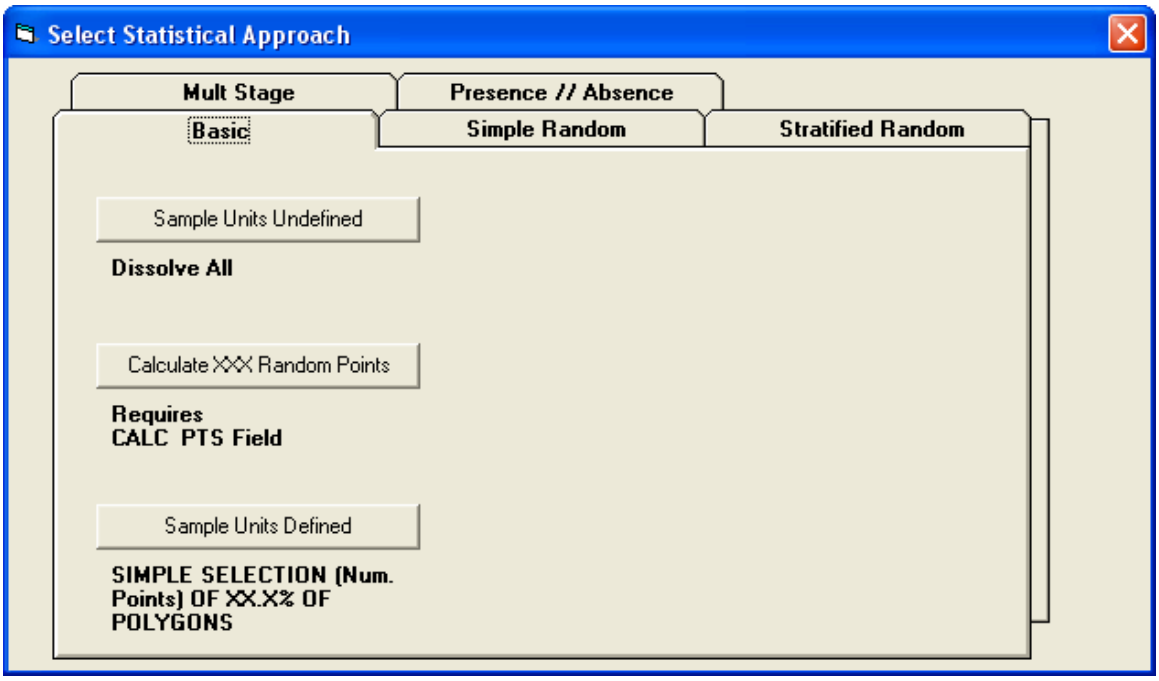
Select the start icon from the commands section in the right panel, drag and drop it onto a toolbar location on the interface.



Clicking the button brings up the following form with five tabs.



Basic Tab



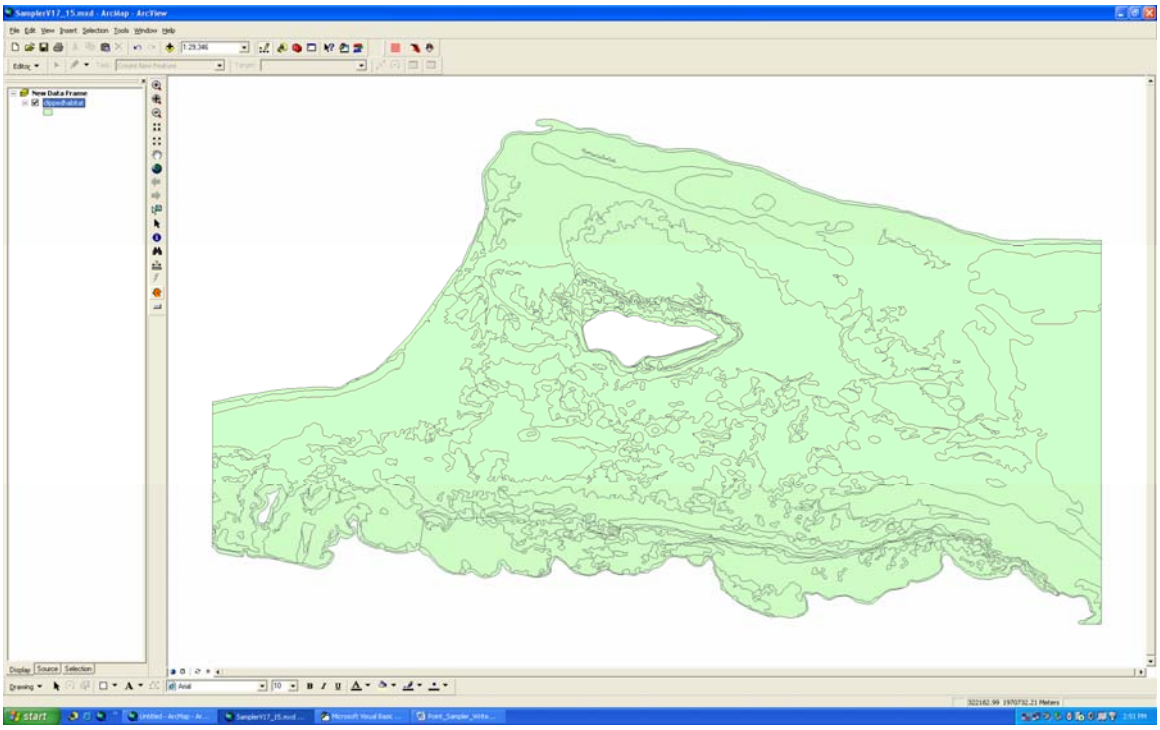
Sample Units Undefined

Definition:

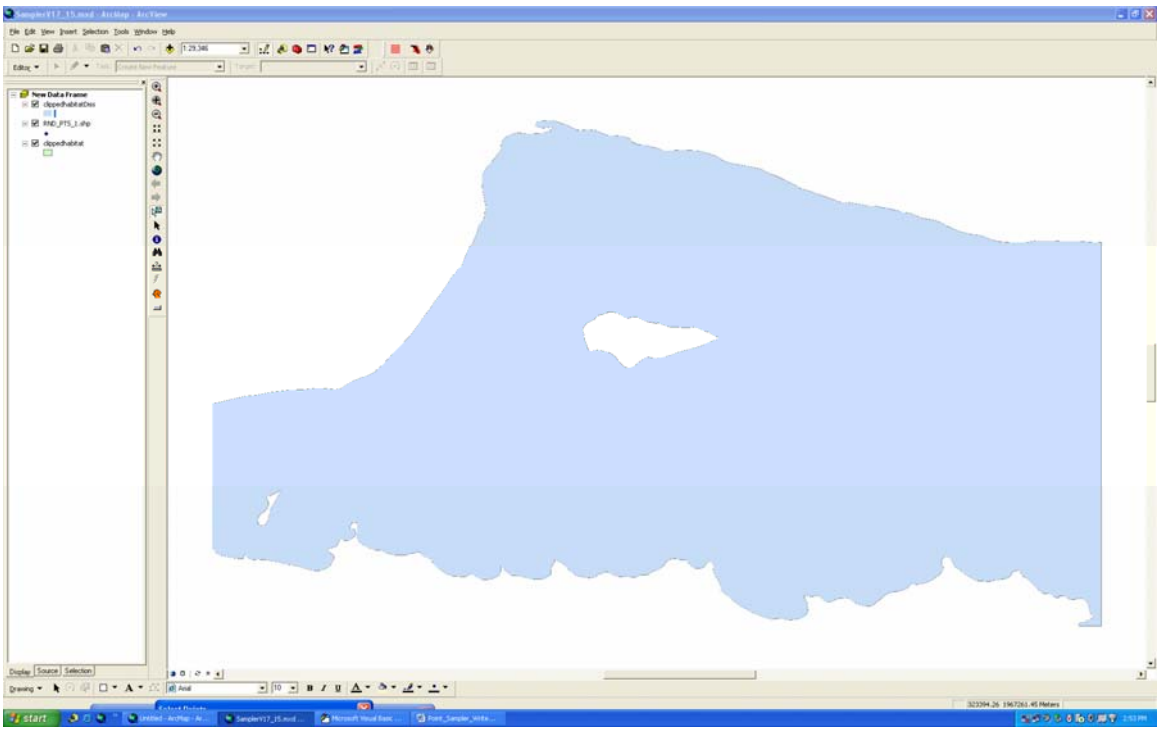
This button simply takes the extents of the polygon layer, dissolves all records into one, and then creates a point shapefile and randomly distributes a XX (user defined) number of points within that dissolved polygon..

Walkthrough:

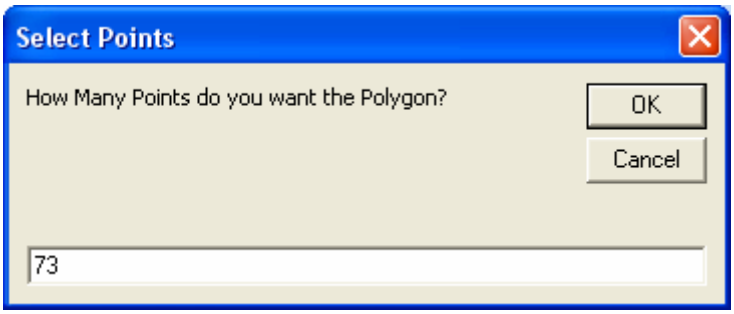
The sample units undefined button requires that the user have a polygon layer in the first (top) position in the Table Of Contents (TOC).



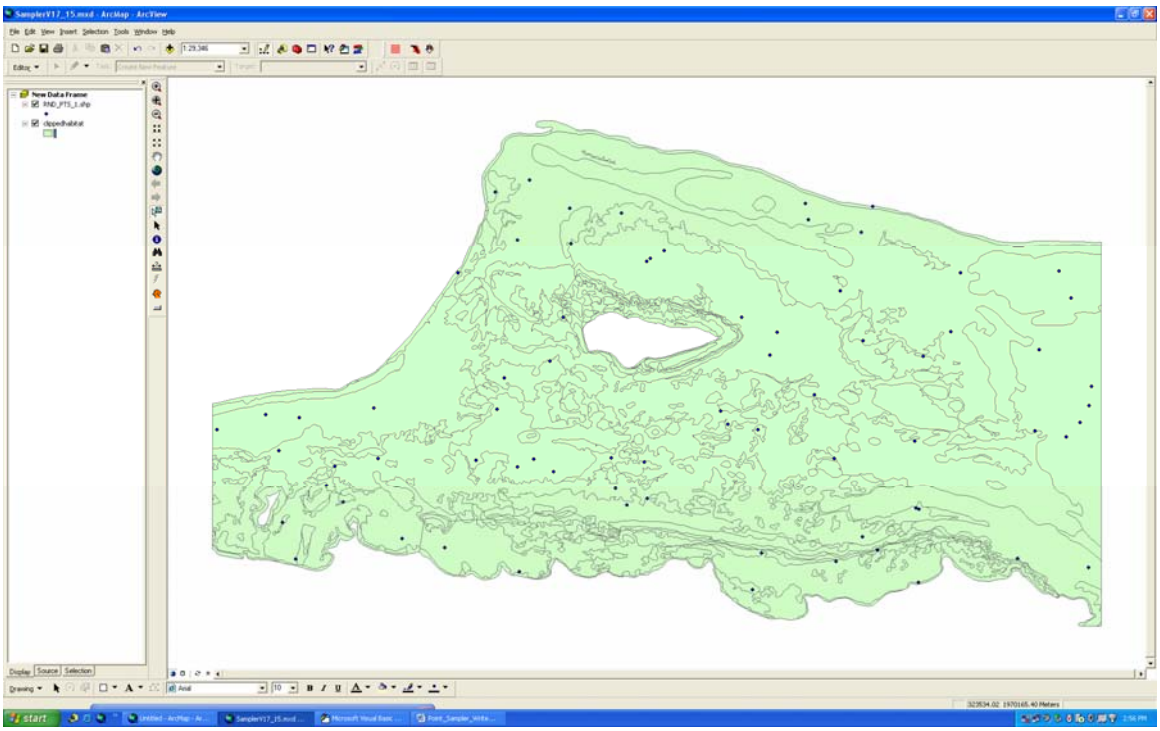
It then dissolves (in the background) that polygon dataset to one polygon and its determines its extents (min/max X/Y).



It then asks for the name and path of the point's dataset which is to be created and prompts for how many randomly distributed polygons you want in the dataset.



It then deletes the dissolved datasets and randomly distributes those number of points (73 in our example) within the constraints of the polygon



Calculate XX Random Points

Definition:
This button places XX points (user defined in CALC_PTS field) in each polygon table row from the layer we are analyzing

Walkthrough:

The sample units undefined button requires that the user have a polygon layer in the first (top) position in the Table Of Contents (Table Of Contents).

Attributes of STC_allhabitat_UTM_Dissolve

FID	Shape *	HABITAT	CALC_PTS
0	Polygon	Coral Reef and Colonized Hardbottom	2
1	Polygon	Other Delineations	12
2	Polygon	Submerged Vegetation	33
3	Polygon	Uncolonized Hardbottom	17
4	Polygon	Unconsolidated Sediments	84

Record: 3 Show: All Selected Records (0 out of

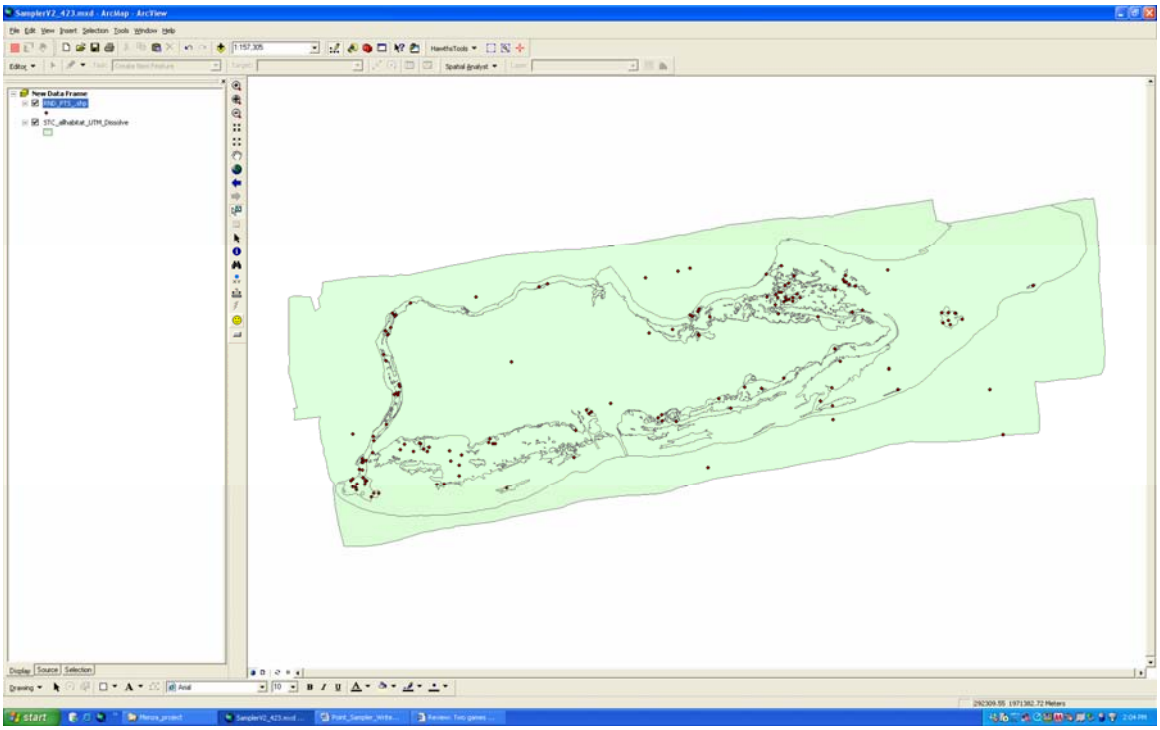
If the field doesn't exist the user receives a prompt

ArcMap

This button works by generating random points in each polygon. The number or random points is defined by the user in a field called CALC_PTS. This field does not appear to be in your dataset

OK

This set shows 148 points distributed according to the table above.



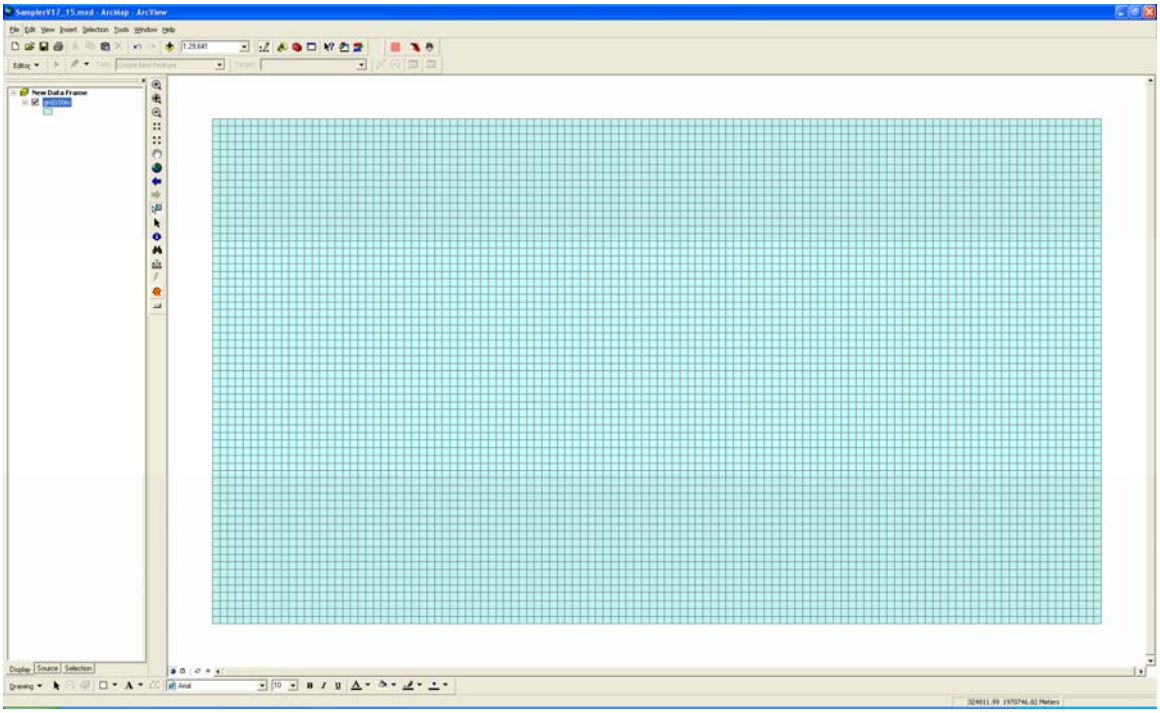
Sample Units Defined

Definition:

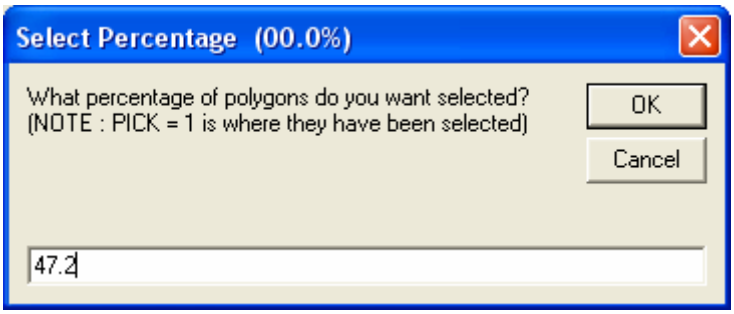
This button randomly selects XX.X % (user defined) of features within a polygon and within that dataset and flags them.

Walkthrough:

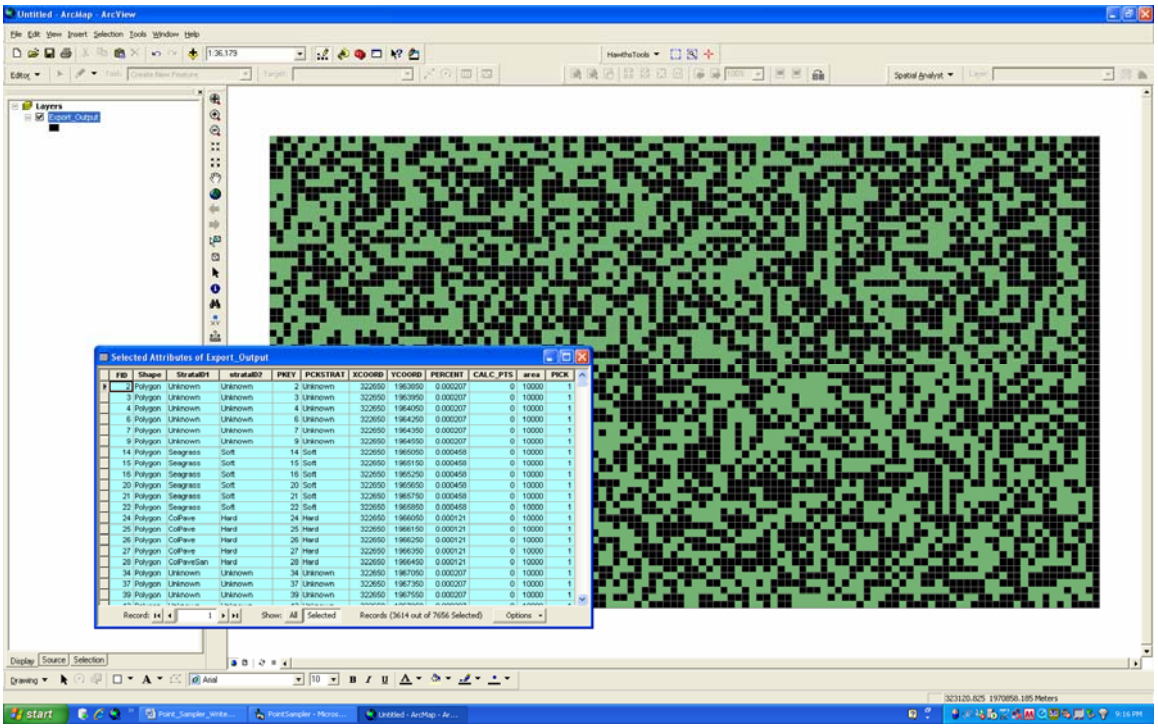
The sample units undefined button requires that the user have a polygon layer in the first (top) position in the Table Of Contents (TOC).



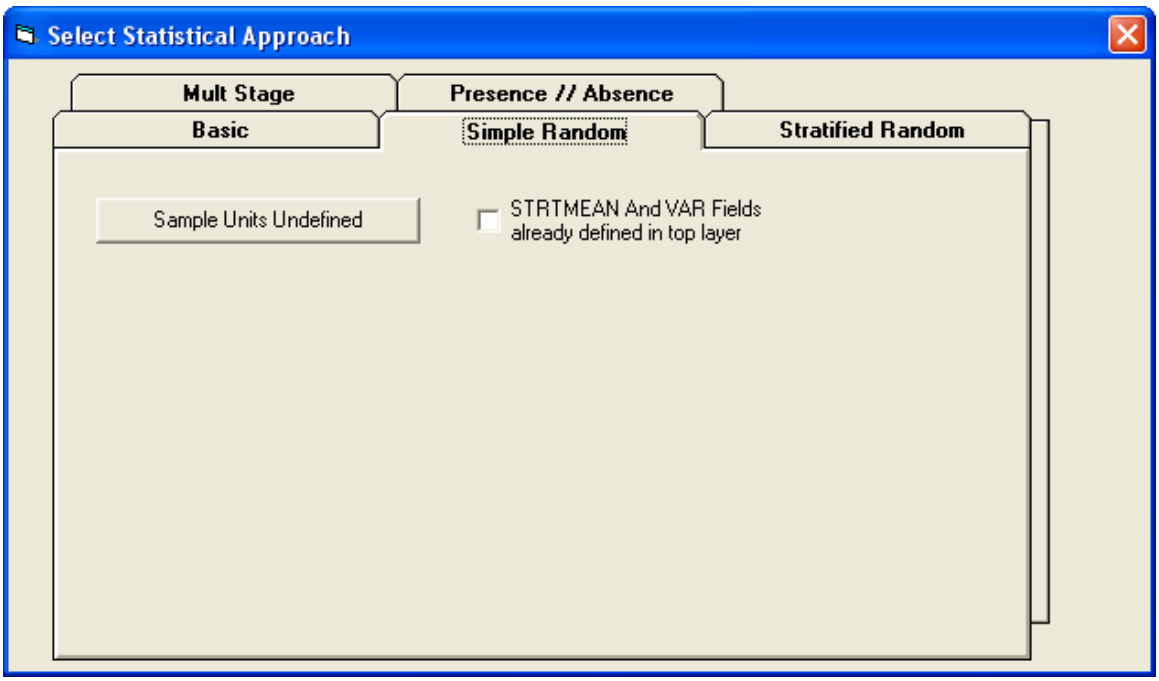
It then asks what percentage of the polygons do you want selected. It then selects that percentage. PLEASE NOTE: The selected polygons are where the PCK_FLD is equal to 1.



The following screen captures shows the affected percentage of polygons that had been selected. In this case it was 3614 features of 7656. (green is where PICK = 1)



Simple Random Tab



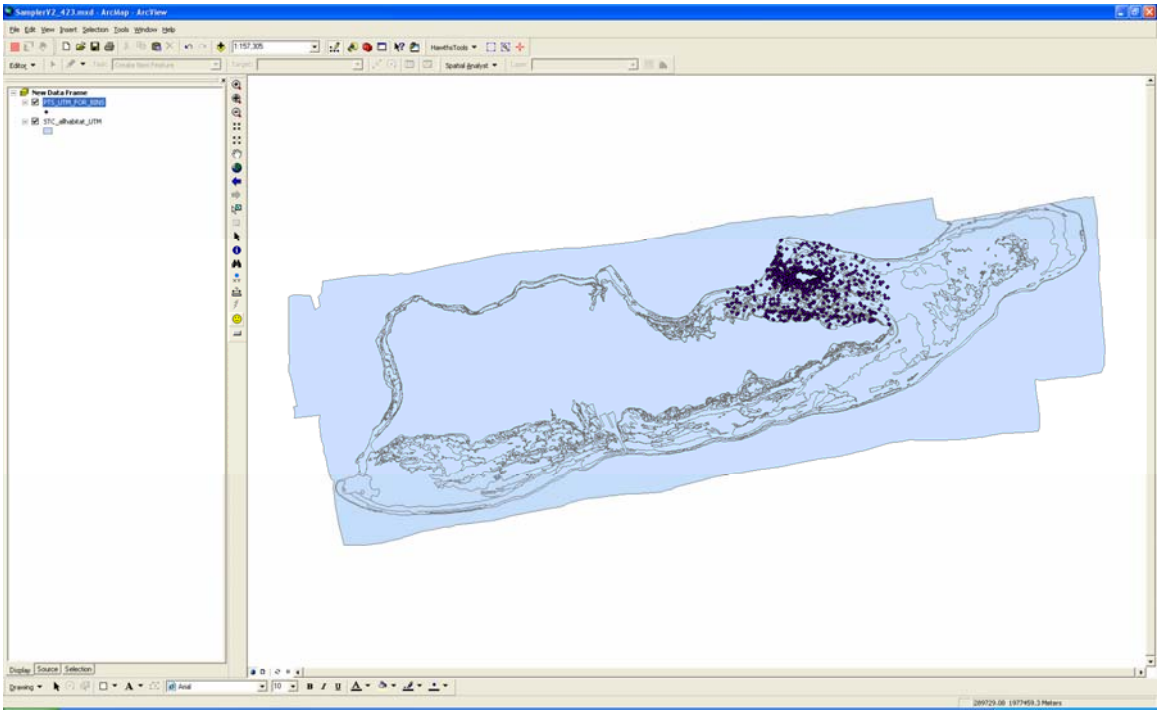
Definition:

Simple random sampling is used to divide the population into several smaller sub populations. The units undefined imply that the source strata layer is irregular. This approach dissolves the polygon into one single data record and then applies the point observations.

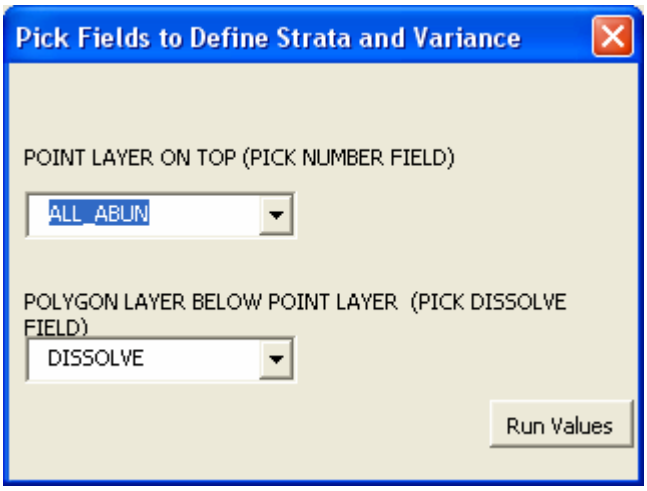
Walkthrough:

The sample units undefined button is located on the Simple Random Tab. This process requires a point and polygon dataset loaded. The point data needs to correspond to known observations in the area. The primary process involves creation of an appropriate StrataMean (STRTMEAN) field and Variance (VAR) field. This can be skipped if the check box next to the button is enabled and if your data has those fields predefined.

The sample units undefined button requires that the user have a point (observation) layer in the first (top) position in the TOC and a polygon layer in the next position below that.

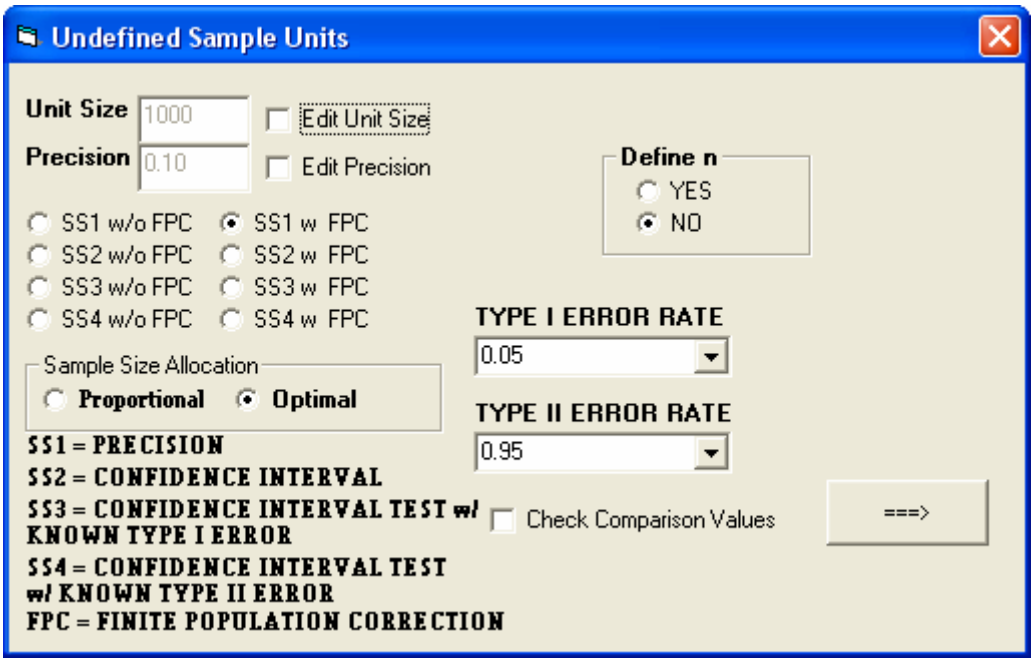


Clicks the Sample Units Undefined button. Please Note : that the Polygon has already been dissolved into one record and as such there is only one selection available in the second combo box.

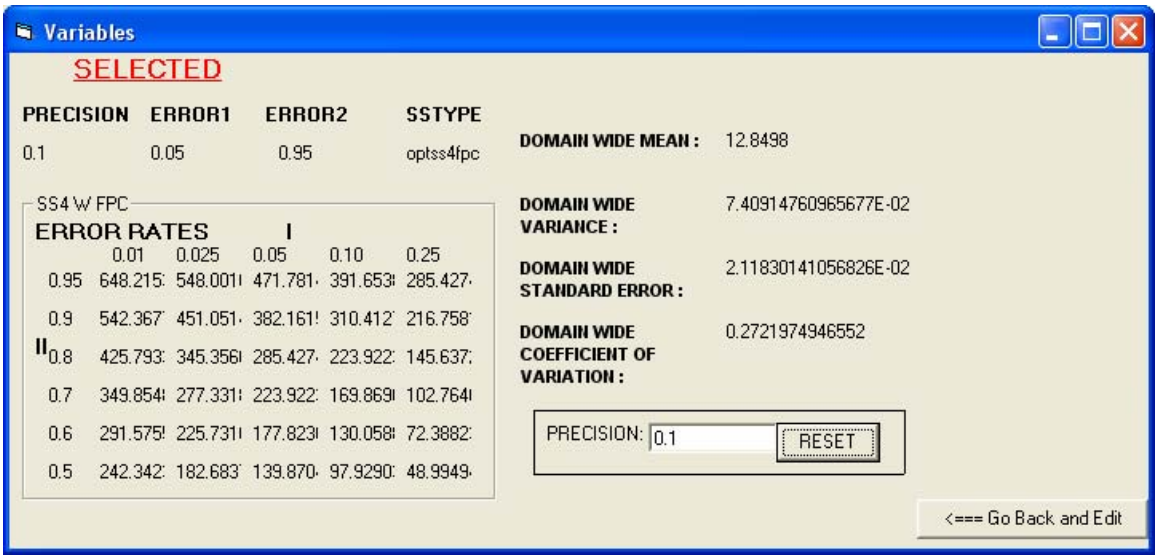


Clicking the Run Values button will run a dissolve on the polygon dataset using the selected strata and prompts the user if there are any acceptable or unacceptable values (zeros in the VAR or STRTMEAN field)

This brings up the form to select the appropriate sampling strategy and if needed the selection of appropriate Error Rates. If the Check Comparison Values button is selected (this is optional) it will open a form which shows the sample sizes for various error rates and precision.

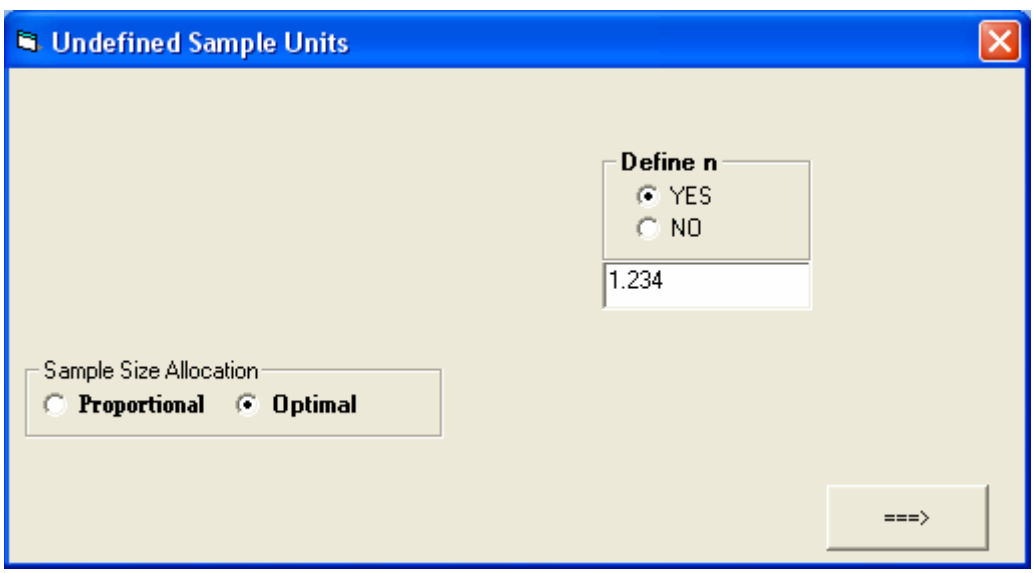


This is shown on the Variables form. The top of the form shows what the user had selected. The user then inputs a new precision in the textbox and clicks the RESET button. This updates the values below. The user can then click <- GO BACK AND EDIT. To make any changes they feel are necessary on the form above.



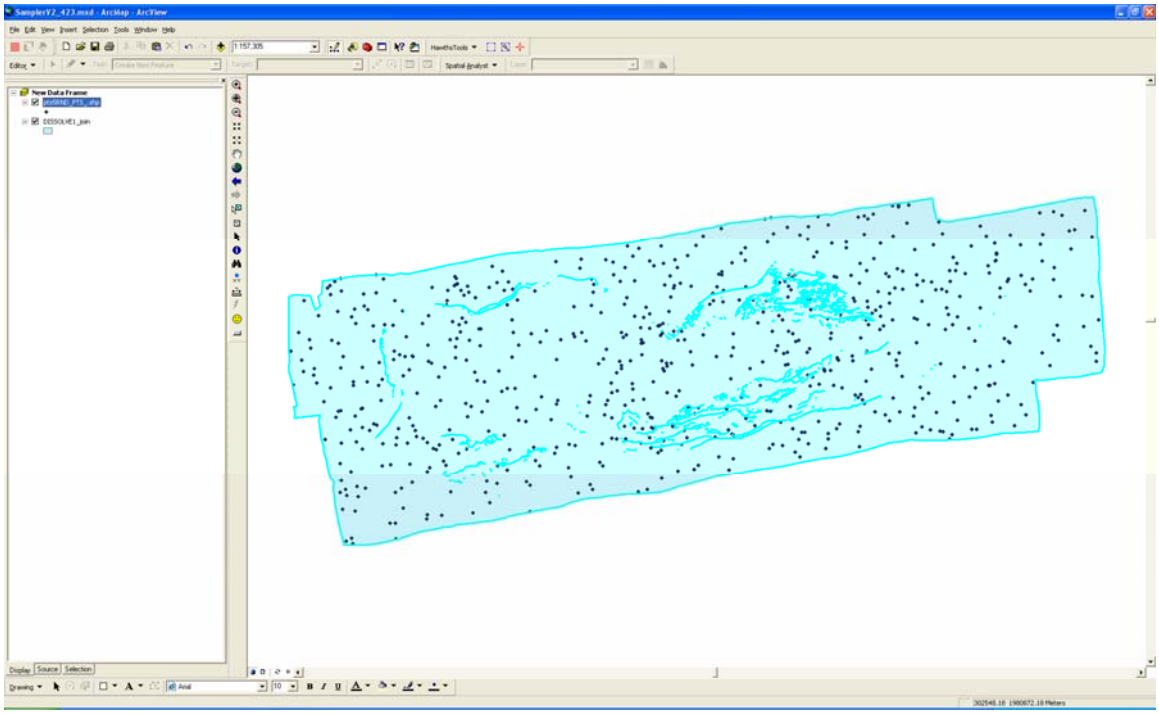
Optional Approach:

The user also has the ability to skip all the sampling approaches and manually set n for all fields, by clicking the **Define n** Button

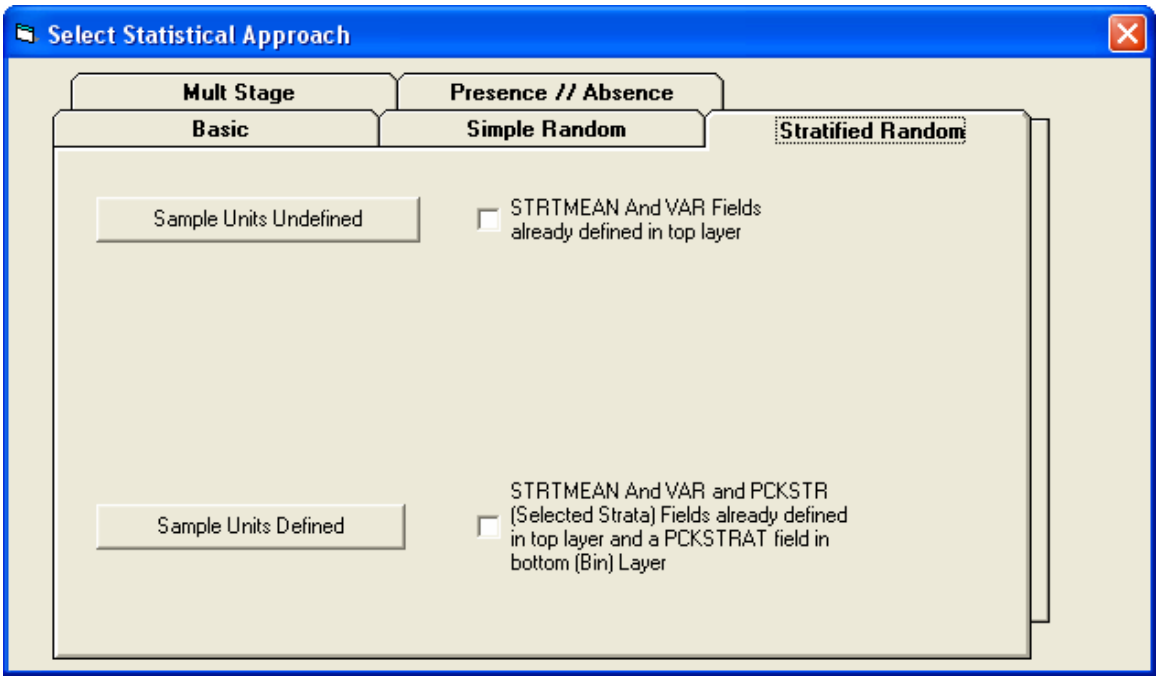


After which they define the sample size allocation and generate the sampling locations. This ends the optional approach.

After the appropriate variables are selected and CONTINUE is clicked on the UNDEFINED SAMPLE UNITS form. This then calculates how many sampling points need to be calculated for each strata that was selected. It also generates that number of points and randomly distributes them in each strata.


SamplerV2_423.mxd - ArcMap - ArcView
File Edit View Insert Selection Tools Window Help
1:157,305
HawthsTools
Editor
Spatial Analyst
New Data Frame
DISSOLVE1_join
Display Source Selection
Drawing
Arial
10
302540.16 1900672.18 Meters

Stratified Random Tab



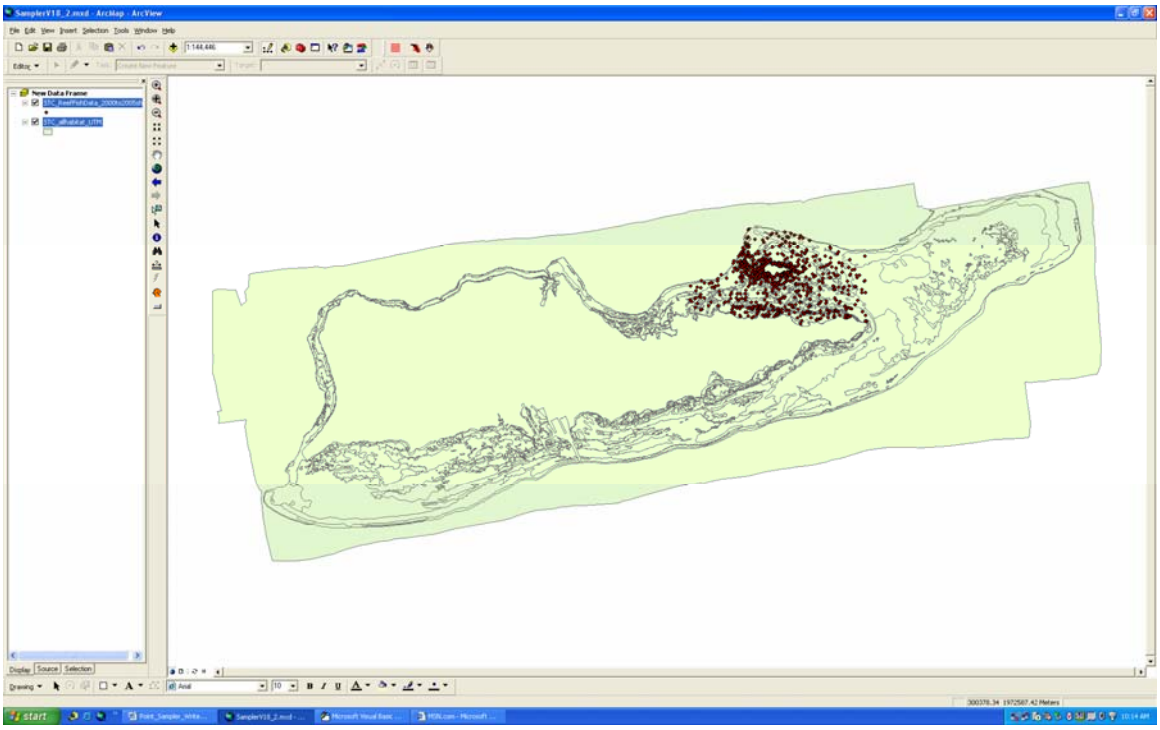
Sample Units Undefined

Definition:

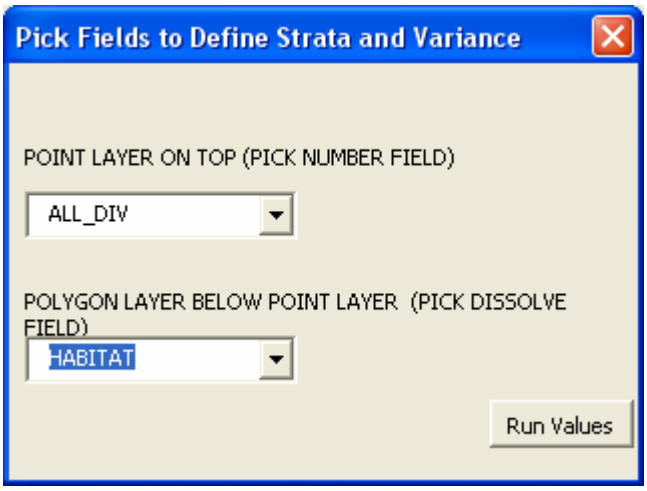
Stratified random sampling is used when one wants to divide the population into several smaller sub populations. The units undefined imply that the source strata layer is irregular.

Walkthrough:

The sample units undefined button is located on the Simple Random Tab. This process requires a point and polygon dataset loaded. The point data needs to correspond to known observations in the area. The primary process involves creation of an appropriate StrataMean (STRTMEAN) field and Variance (VAR) field. This can be skipped if the check box next to the button is enabled and if your data has those fields predefined.

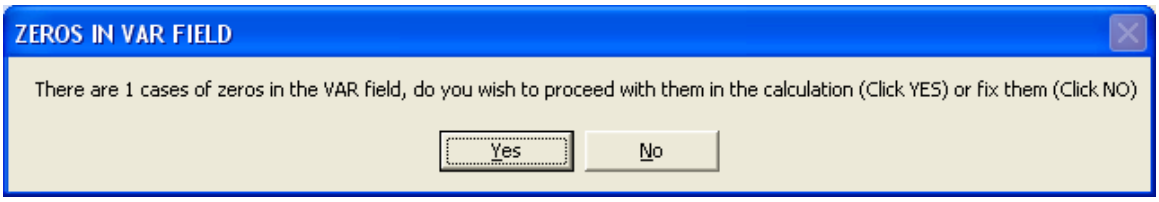


Clicking the button without having the appropriate check box that STRTMEAN and VAR are defined checked, brings up the following form which asks the user to select which number field from the point analysis to use and to select the field from the polygon layer for analysis.

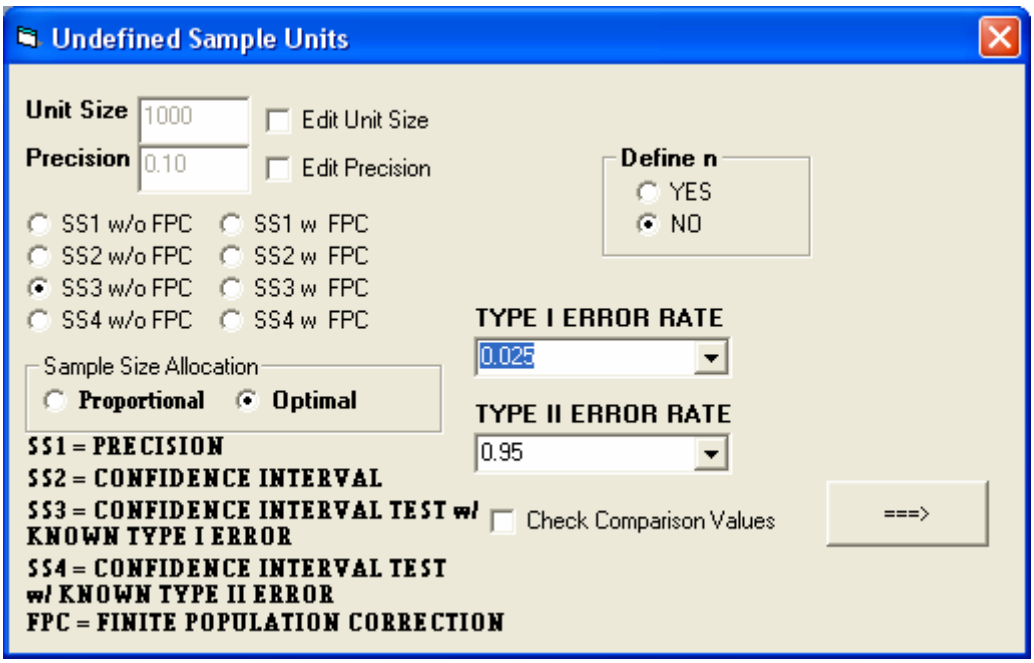


Clicking the Run Values button will run a dissolve on the polygon dataset using the selected strata and prompts the user if there are any acceptable or unacceptable values (zeros in the VAR or STRTMEAN field).

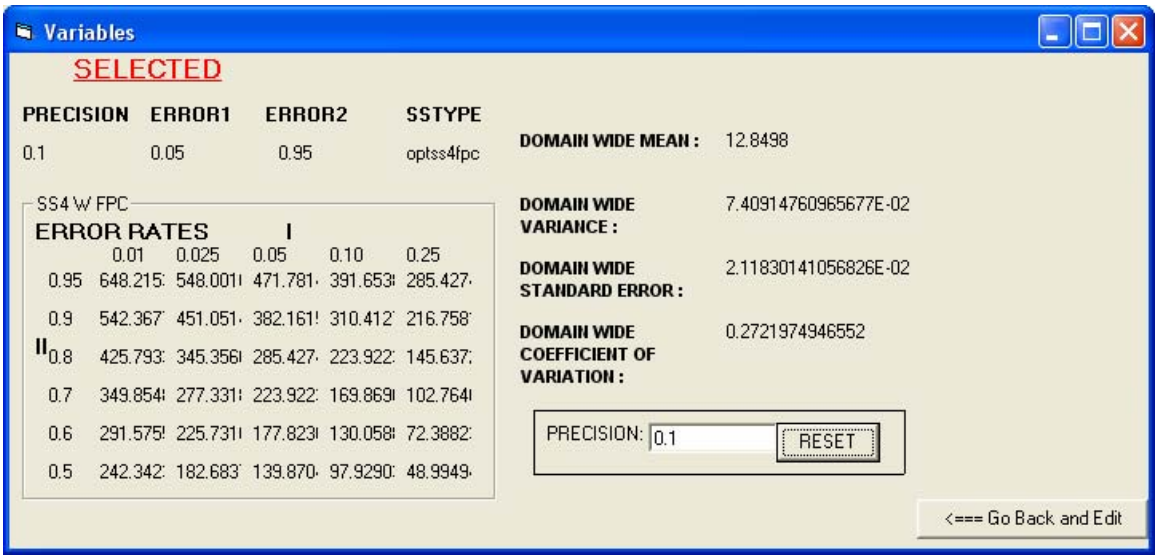
PLEASE NOTE : After this step this process is identical to the process for the Sample Units Undefined button on the Simple Random Tab above.



This brings up the form to select the appropriate sampling strategy and if needed the selection of appropriate Error Rates. If the Check Comparison Values button is selected (this is optional) it will open a form which shows the sample sizes for various error rates and precision.

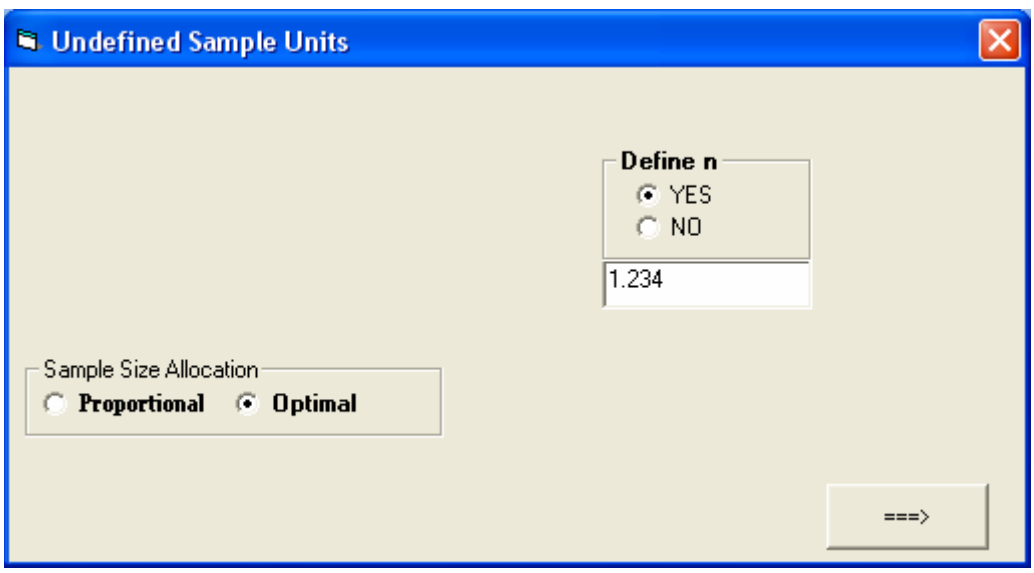


This is shown on the Variables form. The top of the form shows what the user had selected. The user then inputs a new precision in the textbox and clicks the RESET button. This updates the values below. The user can then click <- GO BACK AND EDIT. To make any changes they feel are necessary on the form above.



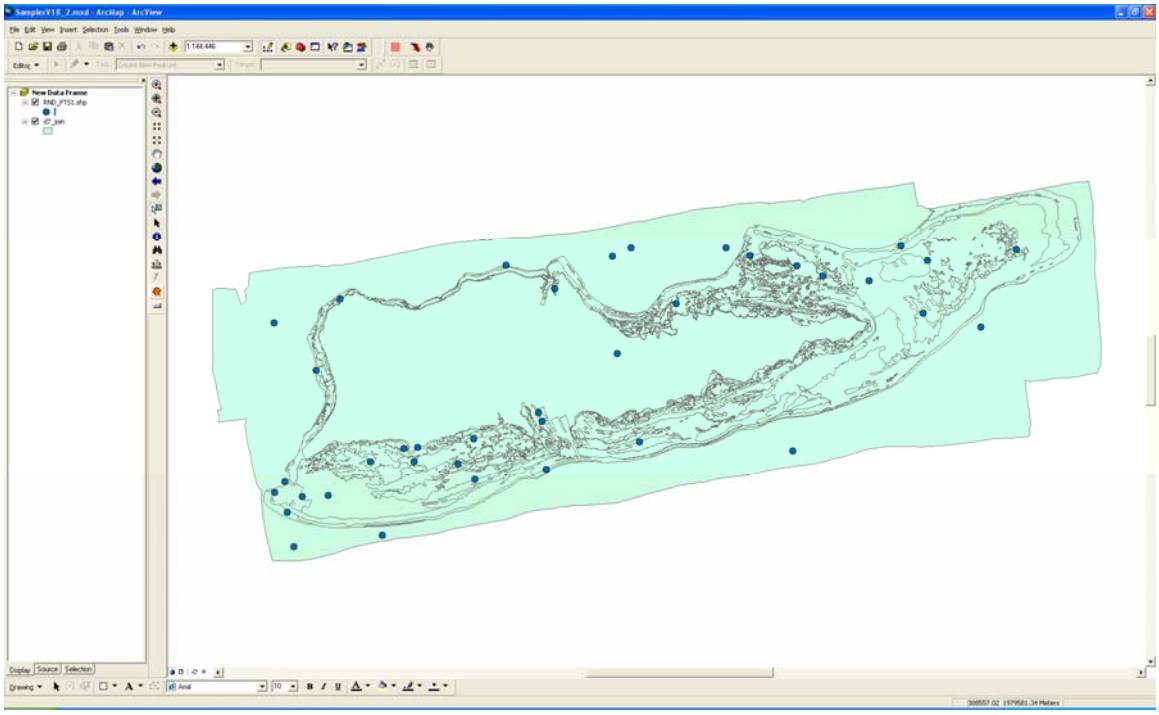
Optional Approach :

The user also has the ability to skip all the sampling approaches and manually set n for all fields, by clicking the **Define n** Button



After which they define the sample size allocation and generate the sampling locations. This ends the optional approach.

After the appropriate variables are selected and CONTINUE is clicked on the UNDEFINED SAMPLE UNITS form. This then calculates how many sampling points need to be calculated for each strata that was selected. It also generates that number of points and randomly distributes them in each strata.

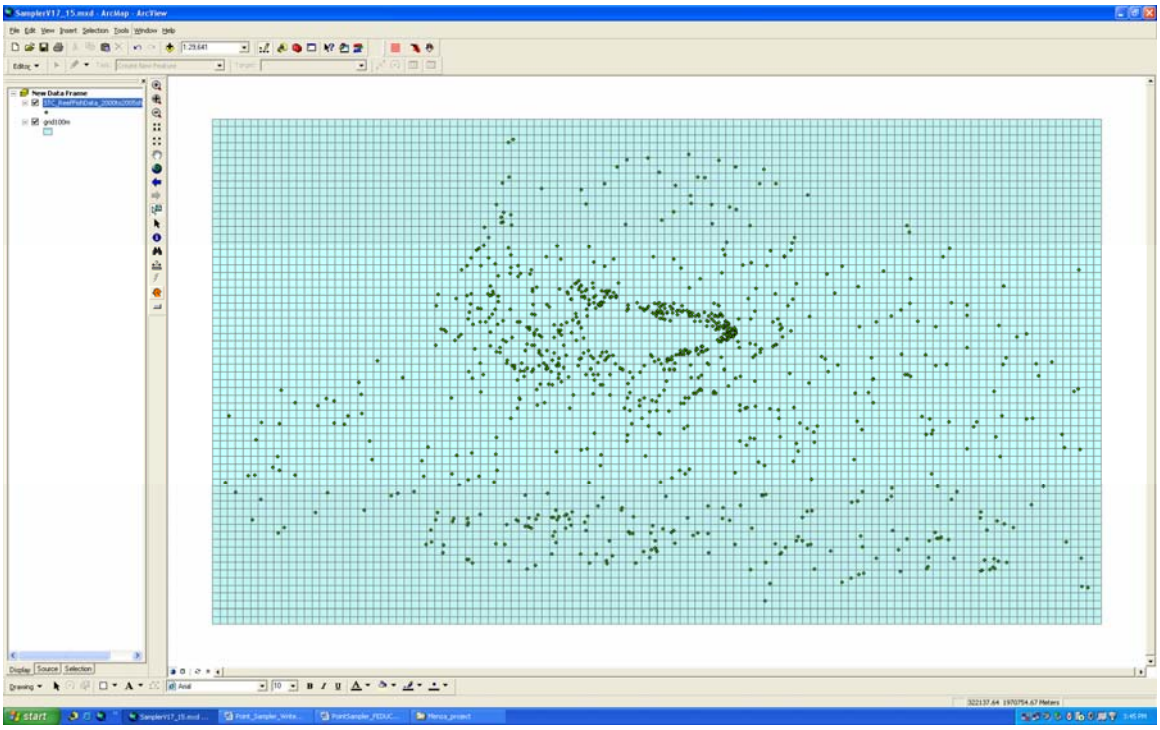


Sample Units Defined

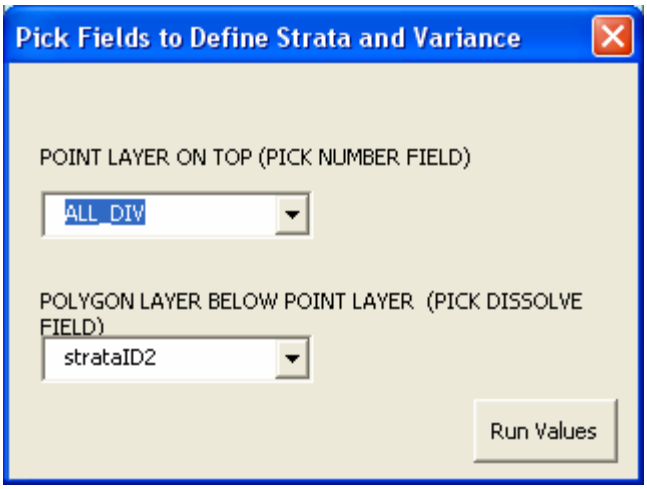
Definition:
Stratified random sampling is used when one wants to divide the population into several smaller sub populations. The units defined imply that the source strata layer is defined or “regular”. This approach analyzes the layer as a dissolved set of features to determine the group data. These values are then applied to the undissolved layer and polygons are selected and flagged.

Walkthrough:

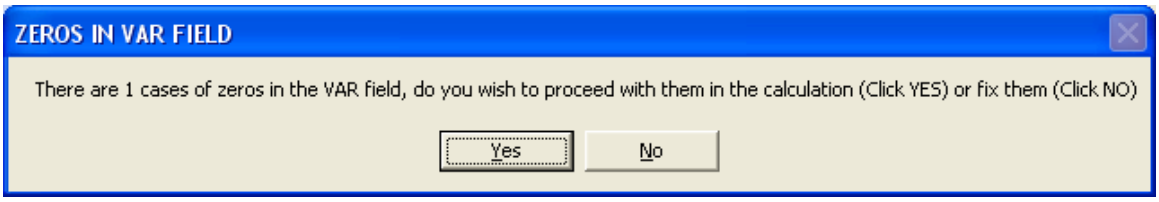
The Sample Units Defined button is located on the Simple Random Tab. This process requires a point and polygon dataset loaded. The polygon dataset needs to have an appropriate field for strata definition. The point data needs to correspond to known observations in the area. The primary process involves creation of an appropriate StrataMean (STRTMEAN) field and Variance (VAR) field. This can be skipped if the check box next to the button is enabled and if your data has those fields predefined.



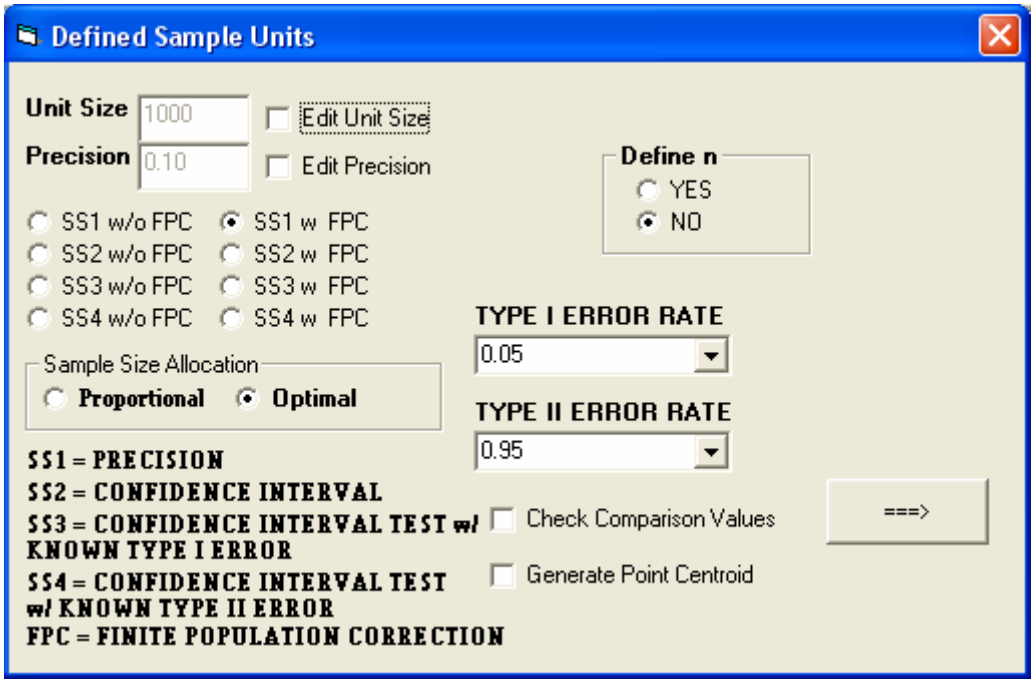
Once the button is pressed it prompts the user to select the appropriate number field from the point layer and strata field from the polygon layer.



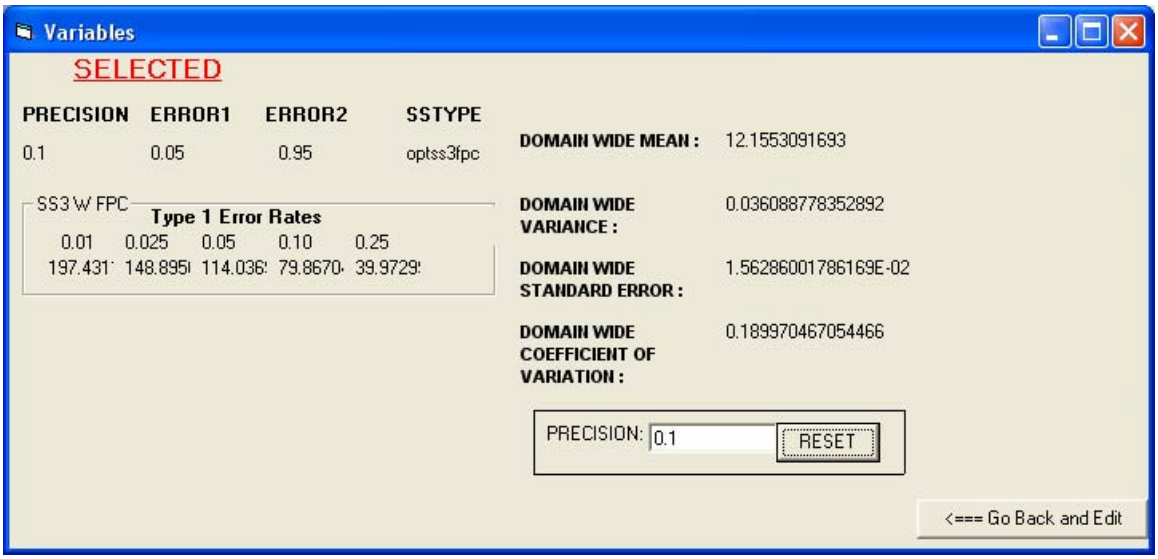
Clicking the Run Values button will run a dissolve on the polygon dataset using the selected strata and prompts the user if there are any acceptable or unacceptable values (zeros in the VAR or STRTMEAN field)



This brings up the form to select the appropriate sampling strategy and if needed the selection of appropriate error rates. If the Check Comparison Values button is selected (this is optional) it will open a form which shows the sample sizes for various error rates and precision.

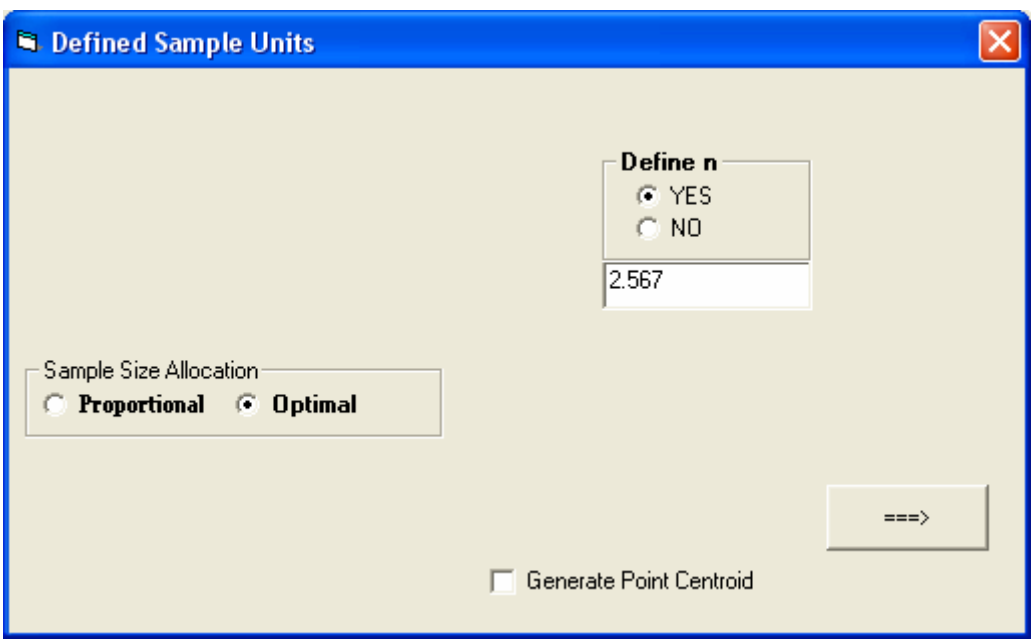


This is shown on the Variables form. The top of the form shows what was selected. The user then inputs a new precision in the textbox and clicks the RESET button. This updates the values below. The user can then click <- GO BACK AND EDIT. To make any changes they feel are necessary on the form above.



Optional Approach:

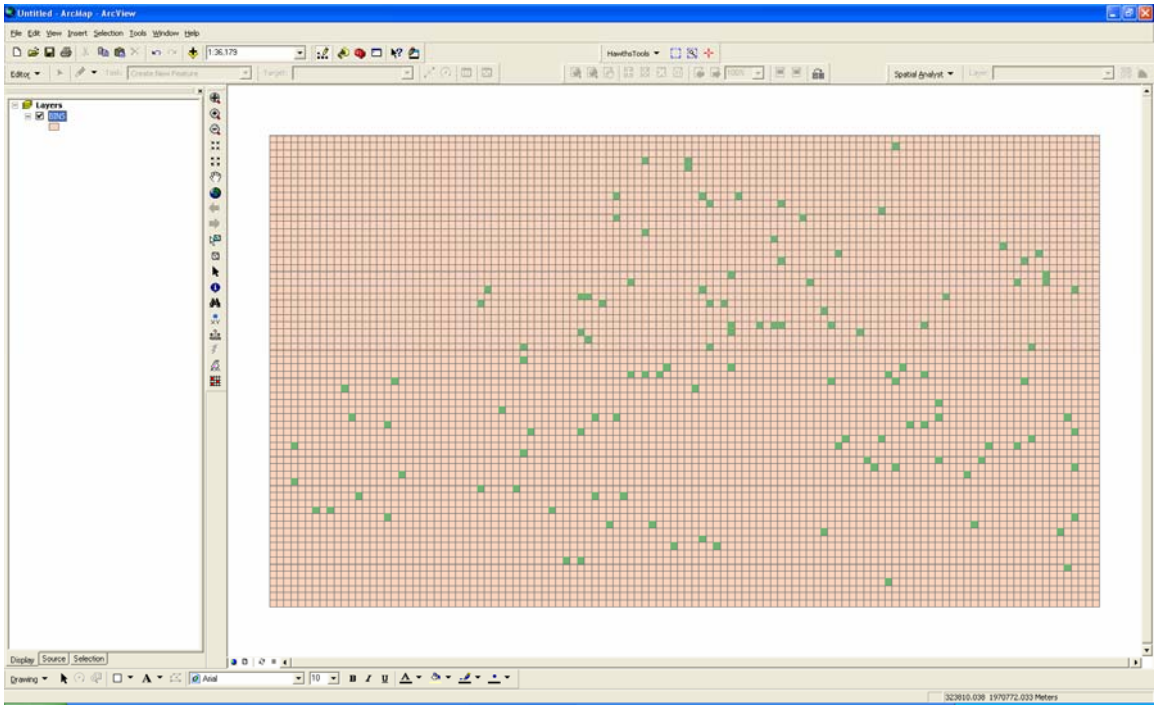
The user also has the ability to skip all the sampling approaches and manually set n for all fields, by clicking the **Define n** Button



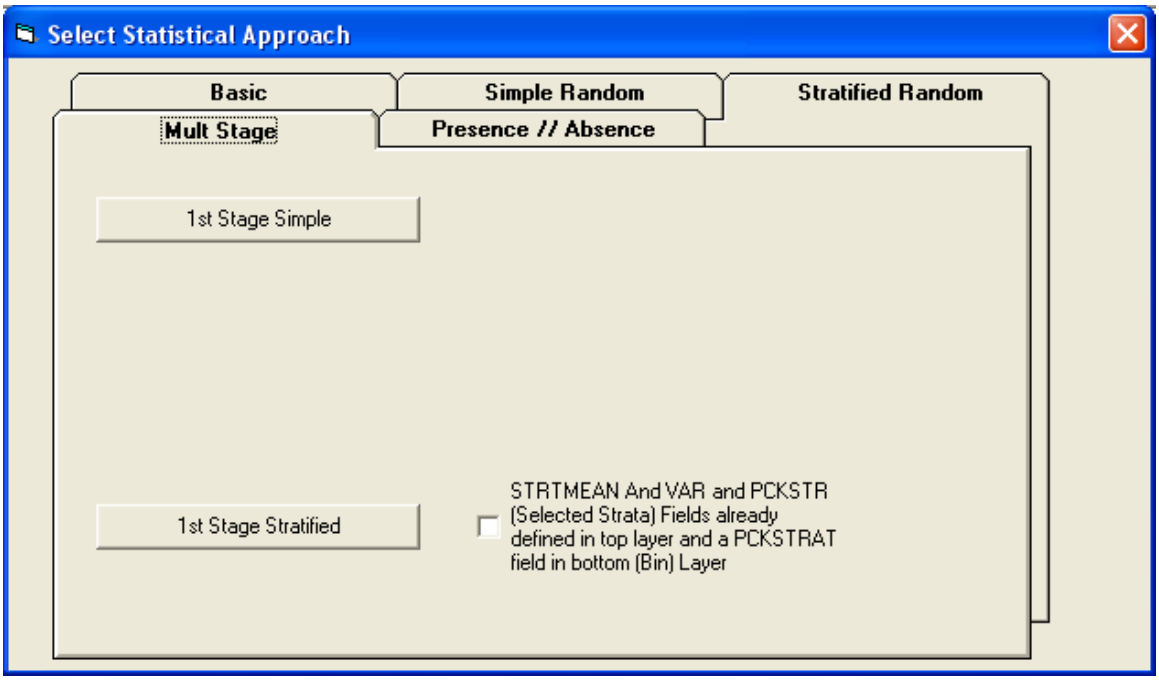
After which they define the sample size allocation and generate the sampling locations. This ends the optional approach.

After the appropriate variables are selected and CONTINUE is clicked on the DEFINED SAMPLE UNITS form, the number of sampling points for each strata selected is determined. Additionally, the number of points is determined and randomly distributed in each stratum.

When CONTINUE is clicked (without checking the variables) it processes the analysis and selects the appropriate number of polygons from the primary (original) polygon dataset. In the case of this analysis it was where the PICK field is equal to 1. This analysis randomly selected 116 of 7656 polygons (Green is selected).



Multistage Tab



1st Stage Simple

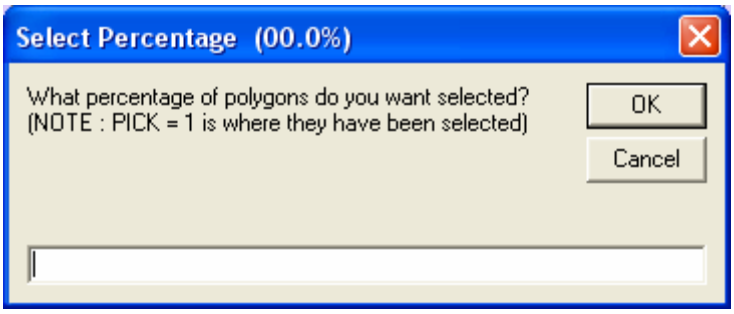
Definition:

This button randomly selects XX.X % (user defined) of features within a polygon and within that dataset and flags them. It then places XX number of points (user defined) in each of the polygons selected in the first step.

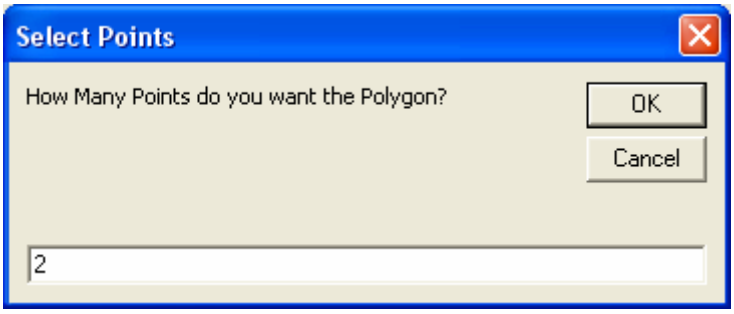
Walkthrough:

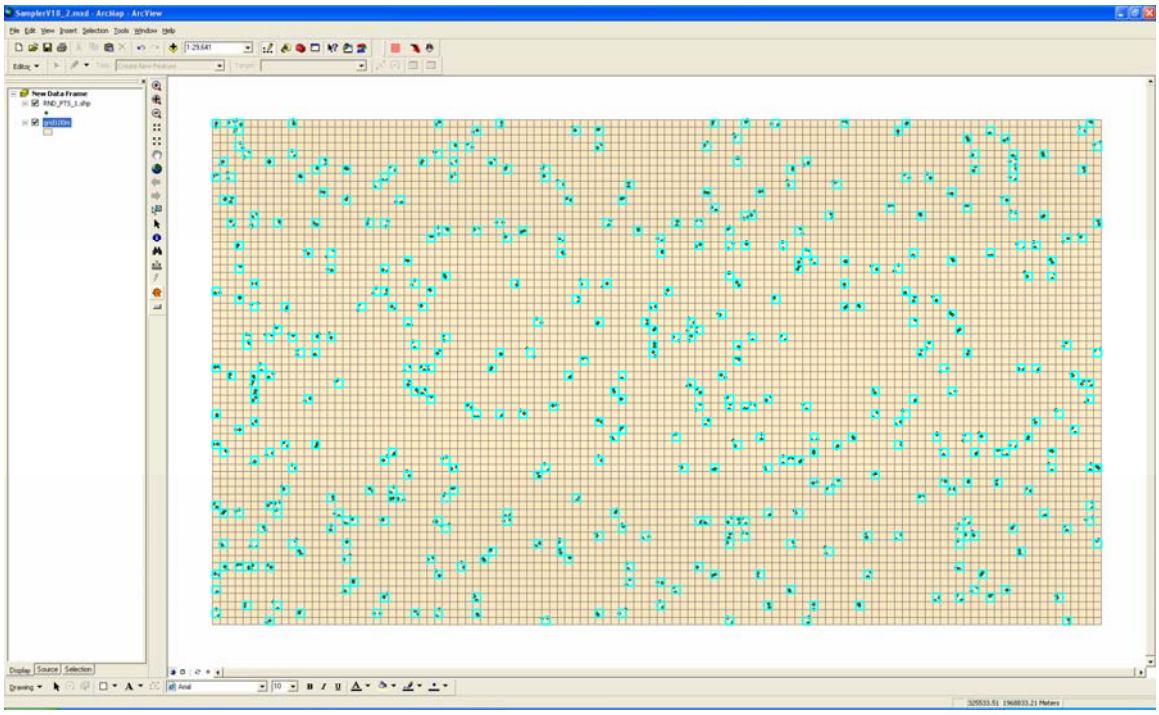
This is effectively the same procedure as the Sample Units Undefined button on the Simple Random Tab, except that it asks for how many randomly distributed points you wish to place in each polygon that is selected. The first prompt asks what percent of the polygons you want selected. The second prompt then asks how many random points do you want generated in each polygon

In this case we have selected 5.7 % of the polygons



and populated them with 2 points each.



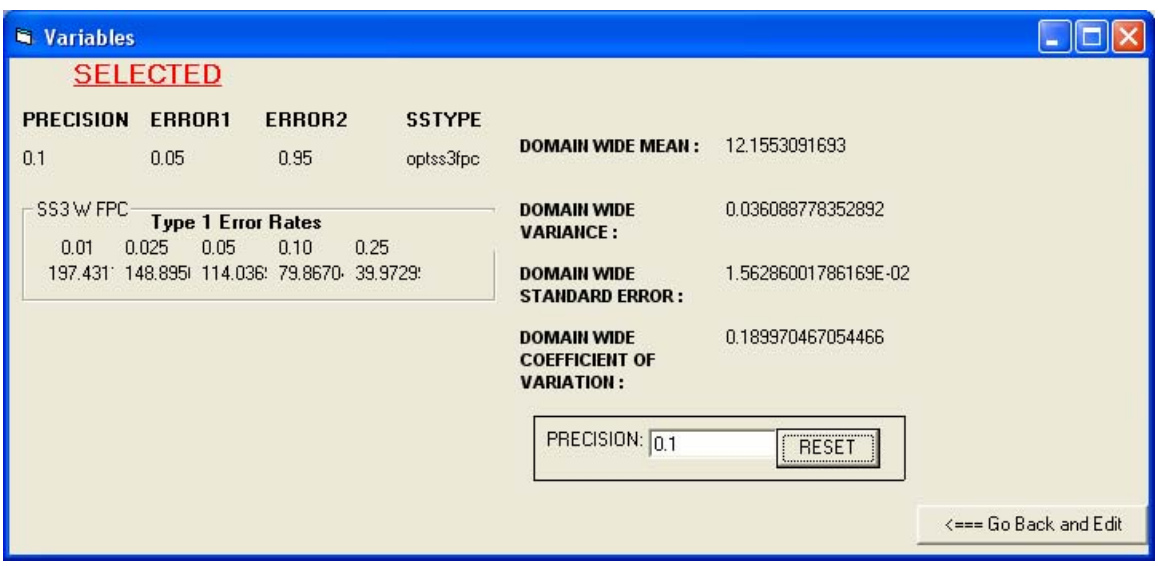


1st Stage Stratified

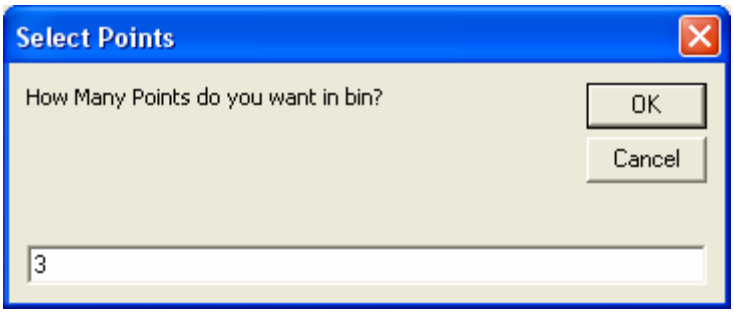
Definition:
Stratified random sampling is used when one wants to divide the population into several smaller sub populations. The units defined imply that the source strata layer is “regular”. This approach analyzes the layer as a dissolved set of features to determine the group data. These values are then applied to the undissolved layer and polygons are selected and flagged . The selected polygons then have XX number of points (user defined) in each of them.

Walkthrough:
This is effectively the same procedure as the sample units defined button on the Stratified Random Tab, except that it asks for how many randomly distributed points you wish to place in each polygon that is selected.

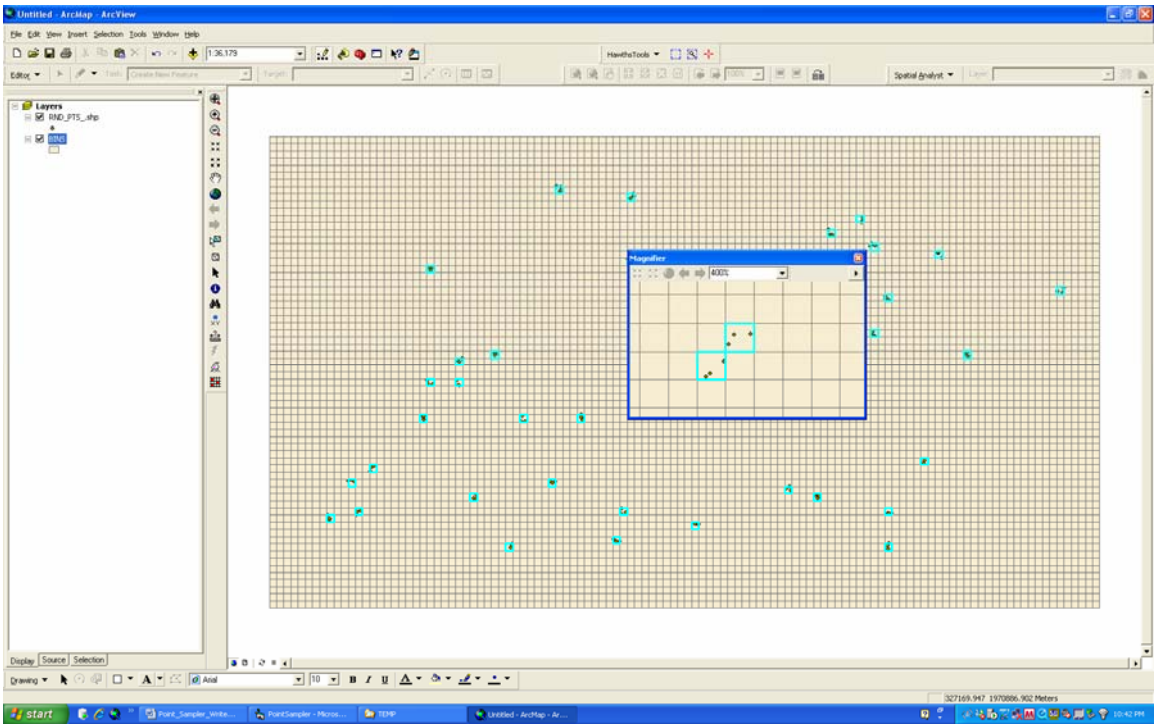
From the Variables form.



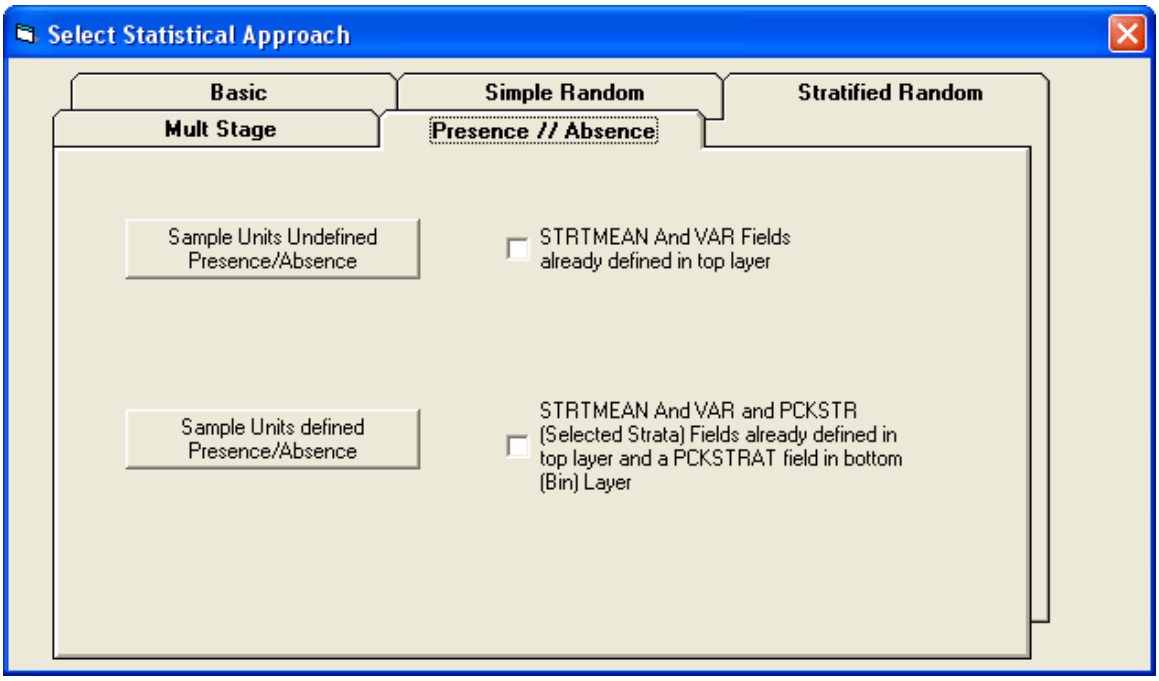
After the user has clicked the CONTINUE button on the next form it queries the user on how many points they want assigned in each polygon.



It then places 3 randomly assigned points in each polygon that was randomly selected.



Presence/Absence Tab



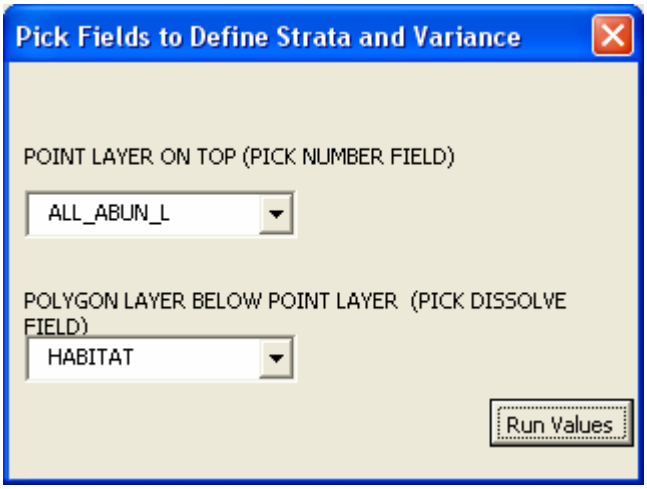
Sample Units Undefined Presence/Absence

Definition:

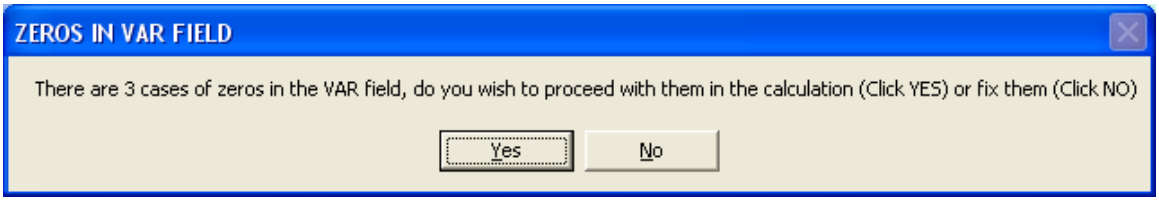
Presence/Absence Sampling for undefined area sampling is used when one wants to divide the population into several smaller sub populations. The units undefined is when the source strata layer is irregular and we are factoring in the presence or absence of observations into the calculations and not using the results of the observations.

Walkthrough:

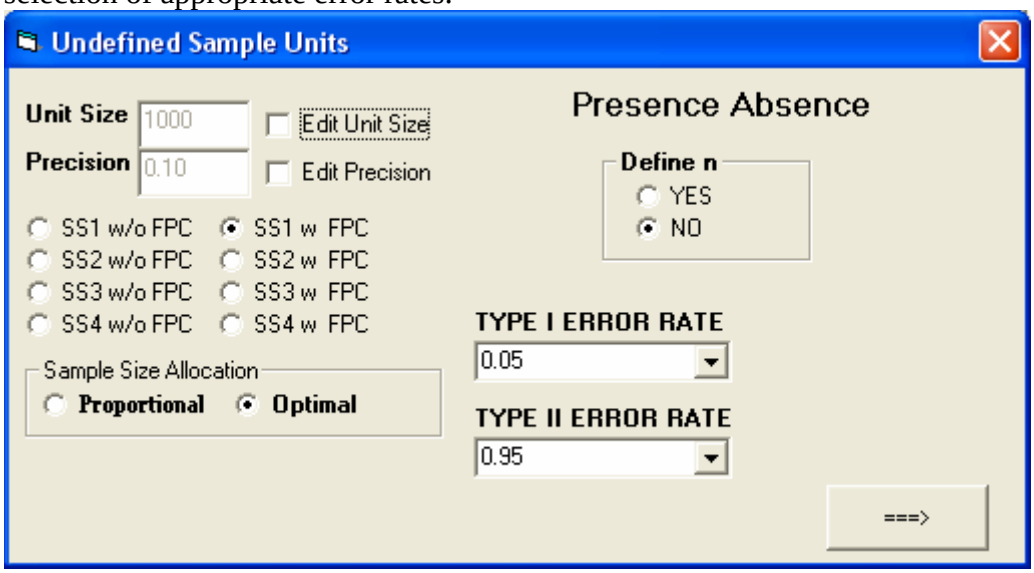
The Sample Units Undefined Presence/Absence button is located on the Presence//Absence tab. This process requires a point and polygon dataset loaded. The primary process involves creation of an appropriate StrataMean (STRTMEAN) field and Variance (VAR) field. This can be skipped if the check box next to the button is enabled and if your data has those fields predefined.



Clicking the Run Values button will run a dissolve on the polygon dataset using the selected strata and prompts the user is there are any acceptable or unacceptable values (zeros in the VAR or STRTMEAN field).

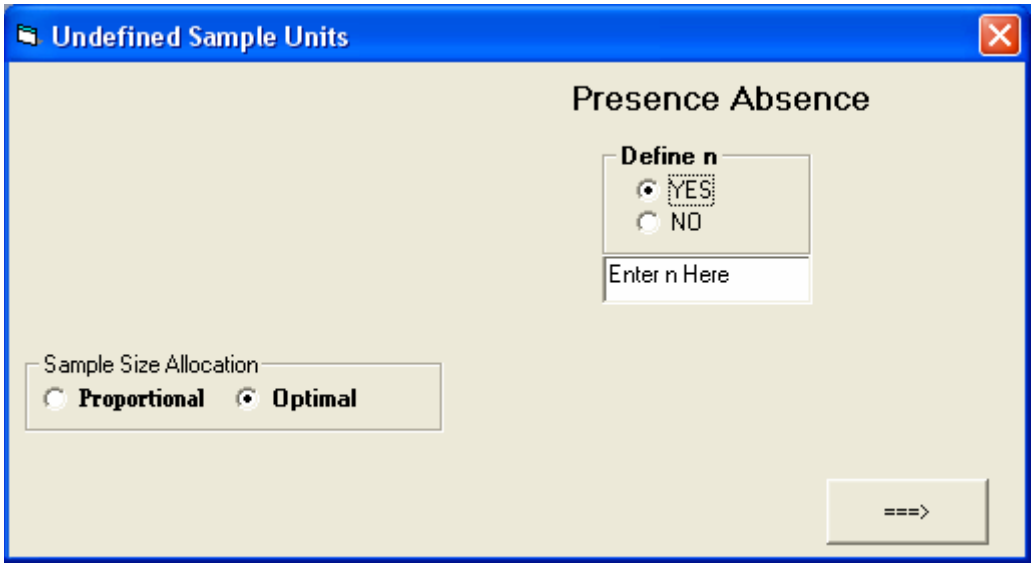


This brings up the form to select the appropriate sampling strategy and if needed the selection of appropriate error rates.



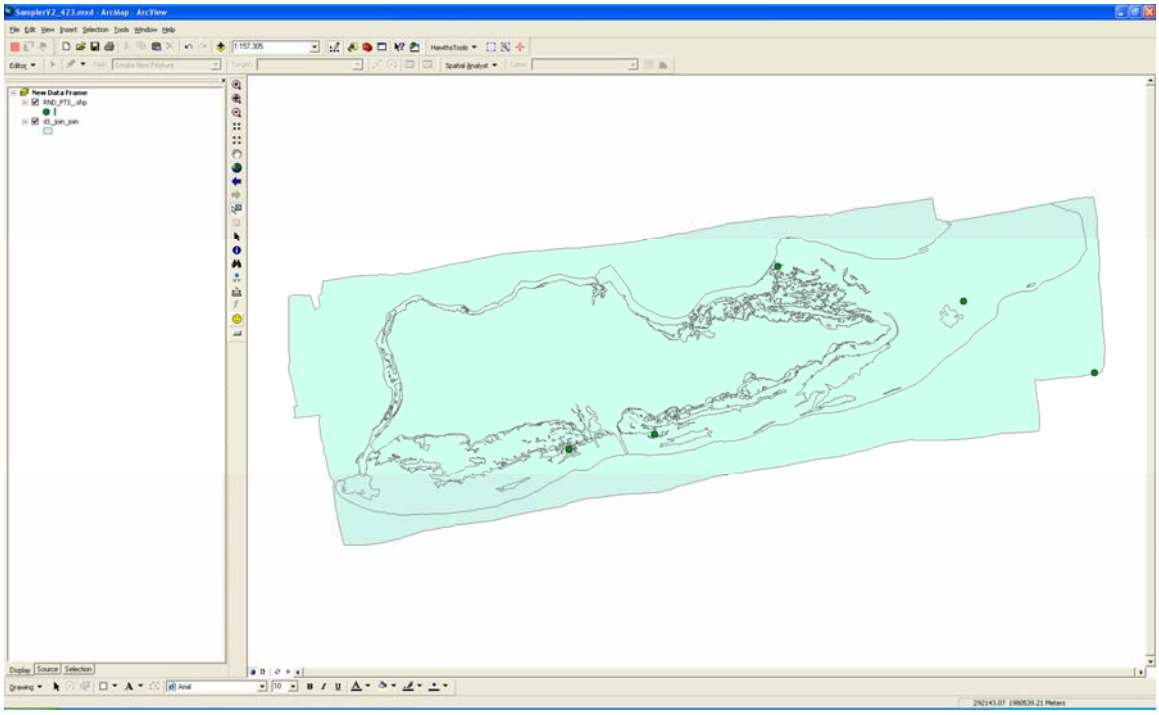
Optional Approach:

The user also has the ability to skip all the sampling approaches and manually set n for all fields, by clicking the **Define n** Button



After which they define the sample size allocation and generate the sampling locations. This ends the optional approach.

After the appropriate variables are selected and CONTINUE is clicked on the UNDEFINED SAMPLE UNITS PRESENCE ABSENCE form. This calculates how many sampling points need to be calculated for each stratum that was selected. It also generates that number of points and randomly distributes them in each stratum.



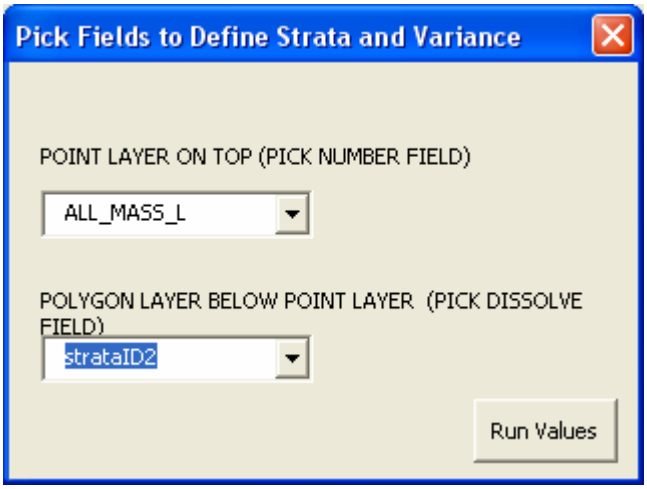
Sample Units Defined Presence/Absence

Definition:

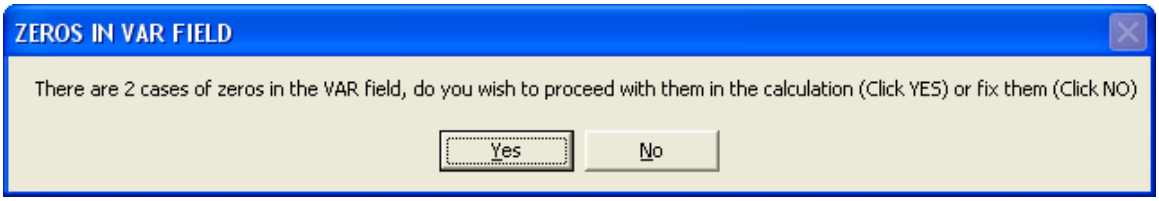
Presence/Absence Sampling for defined areas sampling is used when wants to divide the population into several smaller sub populations. The units defined are when the source strata layer is defined or “regular” (in our example they are 1000 square meter grids). This approach analyzes the layer as a dissolved set of features to determine the group data. In this analysis we are factoring in the presence or absence of observations into the calculations and not using the results of the observations. These values are then applied to undissolved layer and polygons are selected and flagged.

Walkthrough:

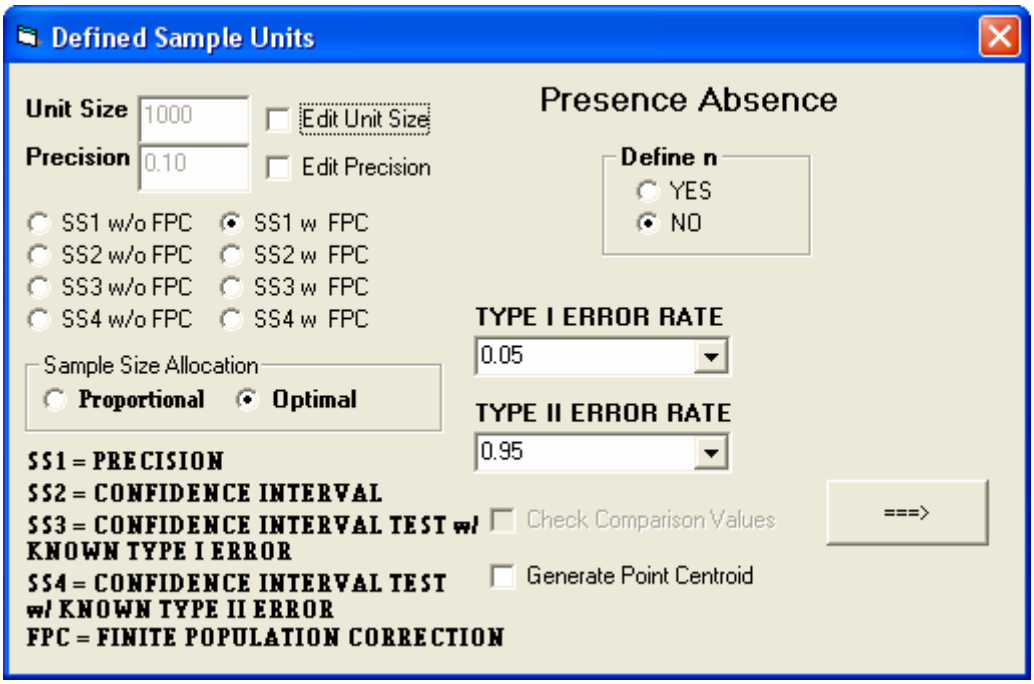
The Sample Units Defined Presence/Absence button is located on the Presence//Absence tab. This process requires a point and polygon dataset loaded. The primary process involves creation of an appropriate StrataMean (STRTMEAN) field and Variance (VAR) field. This can be skipped if the check box next to the button is enabled and if your data has those fields predefined.



Clicking the Run Values button will run a dissolve on the polygon dataset using the selected strata and prompts the user is there are any acceptable or unacceptable values (zeros in the VAR or STRTMEAN field).

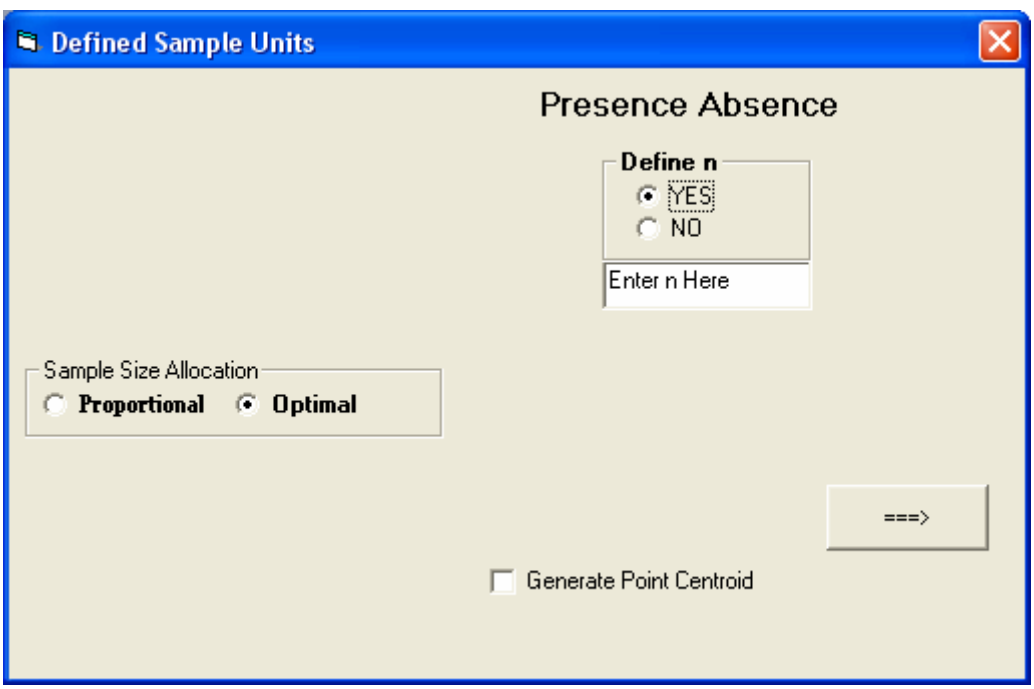


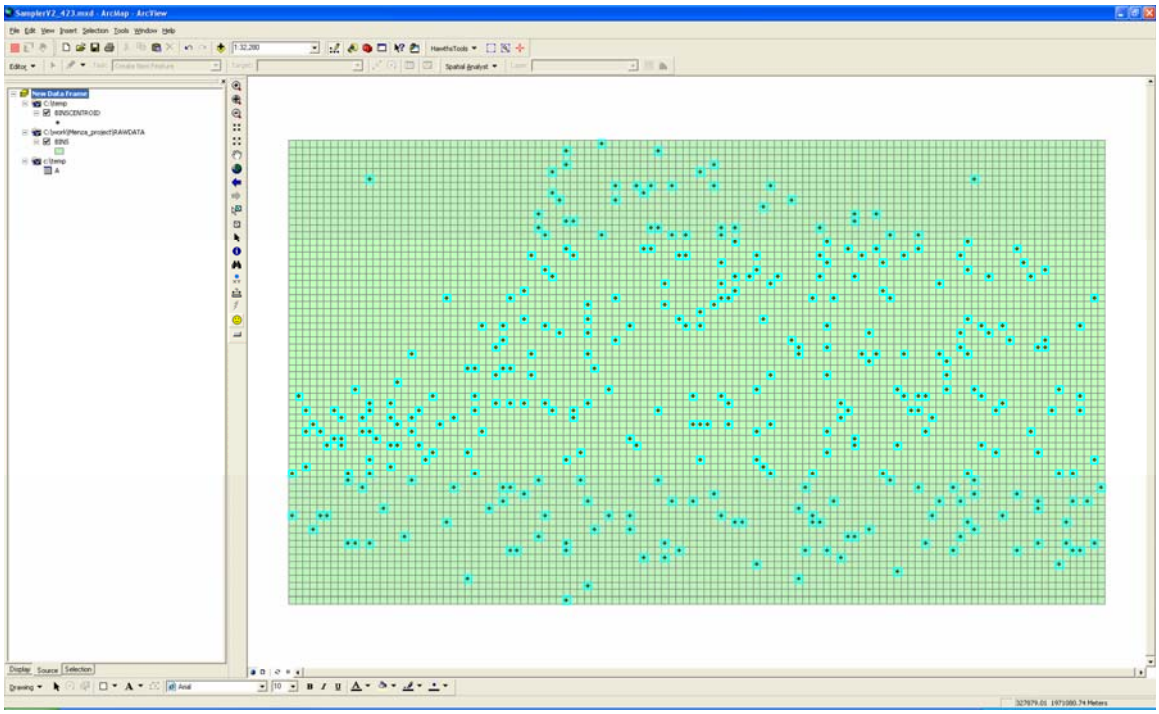
This brings up the form to select the appropriate sampling strategy and if needed the selection of appropriate error rates.



Optional Approach:

The user also has the ability to skip all the sampling approaches and manually set n for all fields, by clicking the **Define n** Button



After which they define the sample size allocation and generate the sampling locations. This ends the optional approach.

After the appropriate variables are selected and CONTINUE is clicked on the UNDEFINED SAMPLE UNITS PRESENCE ABSENCE form. This then calculates how many sampling points need to be calculated for each strata that was selected. It also generates that number of points and randomly distributes them in each strata.