Lab 12: Sampling and Interpolation

What You’ll Learn:
- Systematic and random sampling
- Stratified sampling
- Majority filtering
- A few basic interpolation methods

Data for the exercise are in the L12 subdirectory. Helpful flowcharts are shown at the end of these instructions.

What You’ll Produce: two maps, each with four panels. One map will have panels (i) with the shaded relief original DEM surface, with sample points, and (ii) through (iv) panels with the interpolated surface and contours for IDW, trend surface, and spline interpolation.

The second map will have the same panel as the first, but will have error surfaces and summary statistics corresponding to each of panels ii through iv.

Background: Theory is covered in Chapter 12 (Spatial Estimation) and 10 (Raster Analysis) of the GIS Fundamentals textbook.

Sampling and Interpolation in ArcMap
We’ll practice the mechanics of developing and applying sample points in ArcMap, and use them to extract data from a DEM and interpolate a surface. We’ll then compare the interpolated surface to the original DEM to calculate an error surface, and extract error statistics from a different set of sample points from the same DEM.

We’ll apply both systematic and random sampling. We’ll also develop and apply a stratification layer, because sometimes you want to stratify your sample, which means you wish to increase sample density in some portion of your area, using a map of zones, or strata.

Add the DEM layer, chirdem, a region in southeast Arizona and apply a color scheme to highlight topography in the DEM below the hillshade surface. Note the variation in topography, where there is a significant amount of change, where there is little change. Video: Setup

To keep your work organized, select your “layers” data frame, right click and select copy. Then paste the copied layer 3 times into your map, through Edit - Paste. Each frame then should be renamed:

Original,
Inverse Distance
Trend Surface
Spline Stratified
Apply a fixed scale throughout. Right click on each data frame name, then **Properties → Data Frame**, and set the extent to a Fixed Scale of 1:275,000.

Now add a hillshade to the Original data frame (**ArcToolbox**→**Spatial Analyst Tools**→**Surface** →**Hillshade**). Use an altitude of 25 degrees and model shadows. Make the hillshade semi-transparent on top of the chirdem DEM. (see flowchart Lab 12, Upper left frame, Map 1 at end of these instructions)

Switch to the Layout view and arrange the layers as below:
(hint: use landscape orientation and make sure the data frame windows DO NOT overlap)

---

**Systematic Sampling and IDW Interpolation**

Activate your Inverse Distance data frame

We'll first perform a systematic (grid) sampling, and then an inverse distance interpolation (**video**: Systematic sample, flowchart: Lab 12 Upper right frame, Map1) (see end of this exercise)

Open **ArcToolbox -> Data Management Tools -> Feature Class -> Create Fishnet** (see at right)
In the window that appears, navigate to a directory and name your output fishnet something like *fishnet1000*

Set the template to extent of *chirdem*

Specify an x and y cell size of 830

Use the measure tool to estimate the number of rows (should be something like 23,680/830 or about 29), and columns (28,500/830, or about 34)

Leave other values at defaults; make sure box in lower left is checked to add labels. Note, we only use the labels, not the fishnet, so remove the *fishnet1000*, but keep the labels in the display.

To sample point values, use **ArcToolbox** (**Spatial Analyst Tools** -> **Extraction** -> **Extract Values to Points**)

Specify input point feature (*fishnet1000_label*, from above), *chirdem* as the input raster, and an output point features of something like *sys1000*.

This should generate a sample set similar to that shown to the above. Remove the *fishnet1000_label* data layer.

Inspect the table for *sys1000*, noting the attributes and their values. You should have about 1000 rows in the table, one row for each data point. Note the column labeled *RASTERVALU* filled with elevation values.

Now, perform an interpolation via

**ArcToolbox** (**Spatial Analyst Tools**: **Interpolation** **IDW**)

Specify the *sys1000* input point file, the column containing the sampled DEM values as the Z value field as input (here,
\textit{Rastervalu}, a power of between 1 and 3, \textit{(your choice)} a variable search radius type, between 10 and 15 points, \textit{(your choice)}, and an output cell size of 30.

Note that you should name the output raster something that makes sense, but also within the limits on naming conventions. Here we named it \textit{IDWSysP2_12}

Make sure the IDW interpolated surface is in the Inverse Distance data frame.

Now generate contours for the interpolated surface (\textbf{ArcToolbox} $\rightarrow$ Spatial Analyst Tools $\rightarrow$ Surface $\rightarrow$ Contours).

Specify the IDW surface as input, a 100 meter contour interval, 0 base contour, and a z-factor of 1. Name the output contours appropriately.

Load the sample points and contours into your new data frame, appearing similar to the figure at right.

\textbf{Random Sampling and Trend Surface Interpolation (videos: RandomTrend, Flowchart Lab 12 lower left frame Map 1)} (see end of this exercise)

Activate the Trend Surface data frame.

Generate a set of random sampling points: \textbf{ArcToolbox} $\rightarrow$ Data Management Tools $\rightarrow$ Feature Class $\rightarrow$ Create Random Points

Specify appropriate file for the output location, name the point feature class something like \textit{ran1000}, leaving the constraining feature class blank, and set the constraining extent to \textit{chirdem}.

Specify 1000 as “The Number of Points” and 45 meters as “The Minimum Allowed Distance.” Do not check Create Multipoint Output. We set the value to
45 to avoid points in the same cells. The cells are 30 m wide, or about 43 m on the diagonal, so if all points are at least 45 m apart, no two points will fall in the same cell.

Now sample the elevation values from chirdem as with the systematic layer, via ArcToolbox \(\rightarrow\) Spatial Analyst Tools \(\rightarrow\) Extraction \(\rightarrow\) Extract Values to Points. Name the output something like \textit{ran1000_with_elevation}.

Remove the ran1000 data layer.

Calculate and interpolation surface, this time using the ArcToolbox \(\rightarrow\) Spatial Analyst Tools \(\rightarrow\) Interpolation \(\rightarrow\) Trend (see menu to the left).

Save to a permanent output data set named something like \textit{trend_p3}.

Make sure you use the correct column/attribute for the Z value field of the point data set and a cell size of 30.

Set the output polynomial order to 3, and the regression type to linear.

Run the interpolation, and then load the output interpolated surface into your data view.

Create and add contour lines, as you did for the for the IDW interpolation, above, to yield an output that looks something like that shown at the right.
Stratified Random Sampling of a Raster Layer
(Video: StratRan1.mov, flowchart: Lab 12 lower right frame, Map 1)
(as shown in map at end of exercise 1)

Sometimes we want to vary the sampling frequency across a map. Here, we’ll place more samples in areas that are steeper. First we’ll create three zones, or strata, and then we’ll assign samples based on these strata. Samples will be assigned proportional to both the area of the strata, and the relative steepness, with more samples in steeper strata.

Our strata boundaries will be based on slope, filtered to create larger, more generalized areas.

Activate your Spline Stratified data layer.

Calculate the slope for chirdem (ArcToolbox ➔ Spatial Analyst Tools ➔ Surface ➔ Slope), specifying slope in degrees, and a z-factor set to 1, saving to a permanent dataset named something like "slope_deg."

Create your strata: reclassify slope_deg using Spatial Analyst extension to group slopes into three classes, of 0-1.5, 1.5-18, 18 and above. We chose these values to yield acceptably balanced classes. Usually you stratify for some threshold of an attribute, e.g., slopes above which you can’t build, or elevations where you’re unlikely to find a resource of interest.

Note, you want a permanent reclassification, via reclass, to a new data set, and not just a change in the symbology for the data set. Name the output something like rcslope.

Now we want to reduce some of the “speckle” in the strata (rcslope), the single pixel or long, thin areas. We really don’t need to do this here, but sometimes we do, and this is a good opportunity to practice filtering.

Use ArcToolbox ➔ Spatial Analyst Tools ➔ Generalization ➔ Majority Filter.

Specify a FOUR as the Number of neighbors to use and the HALF replacement threshold. Run the filter five successive times using these options, with the output from each run used as the input to the next run. At each step, note the reduction in speckle as the data are smoothed.
Notice how the small speckles are removed in the final majority filter output, but the long, thin reaches in valleys remain. To generalize further and remove these, you may apply a median filter, via ArcToolbox (➡️) ➔ Spatial Analyst Tools ➔ Neighborhood ➔ Focal Statistics.

You may specify the last majority filtered layer as input, with a circular or rectangular neighborhood, and a 9 cell height and width.

Be sure to select the MEDIAN statistics type, and ignore NoData in calculations. Notice how this substantially generalizes the data, removing the long, thin features. (Video: StratRan2.mov)

Convert the final smoothed raster to a vector layer (ArcToolbox (➡️) ➔ Conversion Tools ➔ From Raster ➔ to Polygon, remember to Simplify polygons), name it something like “Strata.” Use symbology (change the colors) to display the results which should similar to what is shown to the right.

Now we must create the stratified sampling points.

We would like to have a total of approximately 1000 sample points, with 10 times as many sample points in the steep areas (red, at right) as in the flat (green areas). We’d like three times as many samples in the intermediate areas (yellow) as the flat areas.

So, the relative weightings are 10 for the red, 3 for the yellow, and 1 for the green. The total sums to 14, so that means 10/14ths, 3/14ths, and 1/14th.

This means we want about 1000*10/14 or 714 samples in the red, 1000*3/14 or 214 samples in the yellow, and 1000*1/14 or 72 samples in the green.

We can achieve this by distributing these samples over the polygons, based on the polygon area relative to the total area for the strata. For example, in my layer above, the largest yellow polygon has an area of 116.3 square kilometers, and the total area in yellow is 192.5 square kilometers. So, this largest yellow polygon should get 214 * 116.3/192.5, or 129 sample points.

We multiply the number of points for the strata by the polygon area, and divide it by the total area of the strata.
How do we get the total strata area? Remember, by opening the table for the data layer, then creating a new column (Options→Add field; name it “area”, float, precision 18, scale 6), calculating the area into this new column (right click the column heading, then Calculate geometry, sq km), then summarizing by strata ID, in this case by right clicking on the GRIDCODE column heading, then clicking on Summarize, then clicking on the plus to open statistics for area, checking Sum, specifying an output table, and clicking on OK.

This results in a calculated area of:
- 268.3 square kilometers for the flat (gridcode = 1) strata
- 192.5 square kilometers for the intermediate (gridcode = 2) strata, and
- 211.2 square kilometers for the steep (gridcode = 3) strata.

Your numbers may be slightly different, but should be with a few percent of these areas if you used the methods we described above.

WRITE YOUR NUMBERS DOWN, YOU WILL NEED THEM LATER.

We now want to calculate the number of samples per polygon. (Video: StratRan3.mov)

Samples per polygon may be computed many ways, but we'll:

1) Open the strata layer attribute table and add a new long integer field named samp_num
2) Select all polygons for a given strata (gridcode)
3) Multiply the total number of points for this stratum (e.g., 714 for the steep strata, gridcode 3) by the area of the polygon, divided by the total area of the strata (in this case 211.2). Repeat this process for strata 2 & 1.

You'll notice in the calculation shown at right that I've added 0.5, and applied the
Int () function. We can only assign whole sample points, and add the 0.5 for rounding.

Do this for all three strata, substituting the appropriate areas and number of samples for the strata.

Note that your numbers for the strata area and relative number of samples may be different than those shown if you apply a different set of generalization parameters.

Next perform **Selection→Clear Selected Features.** (Video: StratRan4.mov)

Now, to create the random points. We use the same tool as before, **ArcToolbox -> Data Management Tools -> Feature Class -> Create Random Points.**

Specify an output location and file name, something like strat_ran, and a Constraining Feature Class as the vector layer called "strata"

Make sure to select the **Field** radio button to specify the number of points, and use the number value column you calculated (samp_num in this example).
This should generate a sample set that looks something like that to the right, with a higher sampling density in the steeper areas.

Now you should apply this stratified random sample to one last interpolation method.

First, use ArcToolbox -> Spatial Analyst Tools -> Extraction -> Extract Values to Points to assign the chirdem elevation values to each sample point (refer to the previous instructions if need be). Name the output file strat_ran_with_elevation.

Next, estimate a surface using a spline interpolation routine, found in the Arc Tools( ), Spatial Analyst Toolbox→Interpolation→Spline.

Specify the strat_ran_with_elevation as the input for the point features, with the chirdem elevation value field (called rastvalu) as the Z value.

Specify an output raster, and set the cell size to equal that of chirdem (which is 30).

Use a Tension Spline type, with the default weight of 0.1, and use 20 points for the spline.

After running the tool, add this output to your data frame. Calculate contours, and display these and the sample points on your layout. Arrange the layout so it looks approximately like that in the figure below, with an appropriate title, labels, scalebar, name, uniform elevation legend across all three interpolation layers, and north arrow, and turn this in. This is MAP 1.
Interpolation and Sampling Methods

By
Any Geode Student

GIS Fundamentals
Lab 12

Original Surface
Inverse Distance Weighted Systematic Sample

Trend Surface Random Sample
Spline Stratified Random Sample
MAP2
Now, calculate error surfaces for each of the three interpolations. Do this by using the ArcToolbox→Spatial Analyst Tools→Map Algebra→Raster Calculator, and subtract the interpolated raster from chirdem for each of the three methods.

Array the error surfaces in a second layout, similar to that below, to turn in. Note you should use a uniform error range for the symbology across all three error surfaces. (Hint: use Properties→Symbology, stretched, type minimum-maximum. Enter the same extreme high and low values in each frame using the “edit High Low Values” on the Symbology screen.)
Lab 12
Upper left frame
Map 1

Start

chirdem
DEM

Spatial
Analyst
dervive
Hillshade

Set Scale to
1:275,000

Lab 12
Map 1
(upper left frame)

Hillshade of
chirdem

End
Lab 12
Upper right frame
Map 1

1. Start

2. chridem
   DEM

3. Step 1
   ArcToolbox ->
   Data Management Tools ->
   Feature Class ->
   Create Fishnet

4. Step 2
   ArcToolbox ->
   Spatial Analyst Tools ->
   Extraction ->
   Extract Values to Points

5. Step 3
   Spatial Analyst
   Interpolate Inverse
   Distance Method
   cell size = 30
   power=1-3 (choice)
   variable search 10-15 (choice)

6. Step 4
   IDW/Smoothing 12
   Spatial Analyst derive contours

7. Step 5
   100 meter contours of
   IDW/Smoothing 12

   Set Scale to
   1:275,000

End
Lab 12
lower left frame
Map 1

Start

Chirdem DEM

ArcToolbox -> Data Management Tools -> Feature Class -> Create Random Points

ram1000_with_elevation

Step 1

ArcToolbox -> Spatial Analyst Tools -> Extraction -> Extract Values to Points

Step 2

Spatial Analyst interpolate
Trend
cell size = 30
polynomial = 3
regression = linear

Step 3

trend_p3

Spatial Analyst derive contours

Step 4

100 meter contours of trend_p3

Step 5

Lab 12 Map 1 (lower left frame)

Set Scale to 1:275,000

End