Using a Logistic Regression Equation to Model Wolf Habitat Preference in Yellowstone National Park

(Due April 8th, 2003 before 4:00 pm)

Client: Manager, Natural Resources Division, Yellowstone National Park.

Problem Statement:

In this exercise your job is to estimate where an alpha pair of wolves, soon to be released, will likely establish territory. You will apply a hypothetical logistic regression equation using ArcView GIS to perform a spatial modeling analysis to produce a new map theme. This new map theme will outline probable areas where the alpha pair of wolves will likely show a preference to resettle.

Background:

The release will be a late-winter soft release. A soft release means that the wolves will be penned in the release area for two months to allow them to become accustomed to the area in hopes they will not travel too far from the original release point. The release area will be northeast of Fishing Bridge.

Logistic regression is useful for situations where you want to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where some of the independent variables are dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model. Logistic regression is applicable to a broader range of research situations than discriminant analysis.

In order for the results to be easily understood, you will convert the logit values to probability values. Probability values are much easier to understand because they range from 0 to 1. Approaching values of 1 representing an area of high probability of wolf preference.

The hypothetical equation:

\[
\text{logit}(P) = -0.847 + (1.386 \times [\text{Elk}]) + (0.201 \times [\text{H}]) + (-2.197 \times [\text{Pack}])
\]

-0.847 is a constant that is necessary in a logistic regression model. Basically, it represents the logit of the probability that an area is good wolf habitat without considering any variables.

1.386 is a coefficient (importance ratio) that will be multiplied by the value from the elk distribution theme. It is a larger number because it is a stronger predictor of wolf habitat than human avoidance.

0.201 is the coefficient to be multiplied by the values from the human avoidance theme.

-2.197 is the coefficient for the existing wolf packs and is negative because there is an inverse relationship between new wolf habitat and existing pack territory.
**Deliverable(s):**

A short and concise interpretation of the probability map generated focusing on the additive effect of the predictive coefficients in the logistic regression equation. The final cartographic output will be, at a minimum, an area class map with the overall final probability values, the wolf den release area, and the 2 highest probability areas and the boundary shape file. Your interpretation will focus on the justification of your final predicted areas for resettlement. A flow chart is not a required deliverable but you may wish to insert a histogram graphic of the final probability values.

Your report is to be submitted electronically via e-mail. All maps and other supporting figures are to be inserted into the final document. Documents must be smaller than 2 Mb to be accepted through the York mail servers.

**New Application Topics Covered:** ArcTool Box Projection Wizard; Editing ArcView’s theme tables; Using the Geoprocessor extension; Using the field calculator; and, Using the Query Builder.

**The Project Data:**

The dataset for this project has been provided by ESRI, 1998. All theme are UTM 12N NAD27 unless otherwise indicated (see below). Specifically, these themes are:

- Wolfpacks.shp - polygon theme showing approximate boundaries of existing wolf pack territories during Summer 1998;
- Campsites.shp - point theme of campsites in Yellowstone National Park;
- Den-locs.shp - point theme of existing wolf den locations during Summer 1998;
- These points have been altered to not reveal the exact locations of wolf dens.
- Lakes_utm11.shp - Yellowstone's lakes and open water (UTM 11N NAD27);
- Elk.shp - polygon theme showing theoretical elk distribution for Summer 1998. The exact distribution of elk is not made public for their protection. By selecting vegetation types that elk prefer from the Yellowstone National Park landcover data was used to create this theme. The selected polygons were buffered by 400 meters;
- Roaddist.shp - polygon theme showing distance from roads in the park. This theme was created using the Buffer Wizard to buffer roads by 1000 and 2000 meters. Then, this theme was “unioned” with the park boundary and assigned utility measures.

Note: The data used in this exercise is intended only for this exercise and is not an accurate representation of species distribution in Yellowstone National Park.
Suggested Steps:

Overview:

You will first prepare the three themes critical to the analysis, human avoidance (Humans), elk distribution (Elk), and existing wolf territories (Wolfpacks). Then you will use a predictive model, in the form of an equation, to predict wolf habitat based on human avoidance, elk distribution, and wolf territories. The predictive model uses logistic regression, so the output will be logit values (also known as natural logarithms). In order for the results to be easily understood, you will convert the logit values to probability values. Probability values are much easier to understand because they range from 0 to 1.

Step 1 – Opening and preparing the Project file.

a) Open the wolf.apr project stored in your exercise directory. This project has been started for you and contains the following themes: Wolfpacks, Campsites, Roads, Den_locs, Elk, Boundary, Roaddist and a LandSat image used as a backdrop.

b) The data for Yellowstone National Park is stored in UTM 11N meters, so check your view properties’ map units and distance units accordingly.

c) Load the Geoprocessing extension so that you can perform polygon overlay operations.

d) Using the ArcTool Box – Projection Wizard, convert the lakes_utm11.shp to the projects default UTM grid. No change to the datum (i.e., NAD27) is required. Name the new shapefile Lakes.shp and add the feature file to the TOC. Change the legend and labeling properties appropriately.

Step 2 - Prepare human avoidance theme

One of the variables in the wolf prediction model is human avoidance. Wolves are more likely to be found in areas away from people. The utility measures assigned in this step are 1, 3, and 5 based on a scale of 1-9 (ordinal scales identify relative magnitudes, but they do not quantify exact differences between values). Notice in the view, the higher utility measures for Roaddist represent a greater distance from people. When the measures are applied in the logistic regression equation, they will be multiplied by 0.201. Areas farther away from humans having higher utility measures will have a higher value in the predictive model as well.

a) Create buffers around the campsites (i.e., features of a theme). Create a set of buffers (as multiple rings) at 1000 meters and a second set of buffers at 2000 meters. Create the buffers in a new theme called Campbuf.

b) Edit the Campbuf theme table (see “editing a table” in the ArcView Help Topics... menu). Add a field called Campcode to the theme table and assign a utility weight of 1 to the 1000-meter buffers and 3 to the 2000- meter buffers (see “Editing the values in a table” in the ArcView Help Menu).

c) Using the Geoprocessing Menu, Union Campbuf and the Roaddist theme to create a human avoidance theme called Humans. (see “Unioning features” in the ArcView Help Topics... menu). Open the Theme table for Humans.shp and review the results of the Union operation.

Now that you have distance from roads and campsite buffers together in one theme, you will need to consolidate the utility measures. Since campsites have more human activity, you will
allow those utility measures to take precedence over the road distance measures when the two buffers coincide. The exception to this is that polygons with a Roadcode of 1, which means that they are adjacent to roadways, will keep a utility measure of 1.

d) Create a new field in Humans called H_code.
e) Calculate H_code equal to Roadcode using the field calculator (see figure 1.0).
f) Using the Query Builder, Select the records in the Humans theme where the BufferDis (from Campbuf) does not equal 0 and Roadcode does not equal 1. Select “New Set” when the equation is built. You should at least 41 records selected. Using the Field Calculator, re-assign H_code to equal Campcode for those selected records. Save your edits and stop editing.
g) Classify your new Human theme by Unique Value using H_code. Ensure no records are selected in your tables before proceeding to step 3.

![Field Calculator](image)

*Figure 1.0 – Calculating the H_code field to equal the Roadcode field using the field calculator*

Step 3 - Prepare Elk and Wolfpacks themes and union with human avoidance

The best predictor of wolf habitat is elk distribution. Therefore, another variable used in your wolf prediction model is the presence of elk, the wolves’ prey. Your elk theme has no classes, it shows only the presence or absence of elk (i.e., a dichotomous variable). So, there is no reason to assign a formal utility measure. In this case, you assign what is known as a flag. A flag is a dummy value used to multiply by the coefficient in the equation to obtain a nonzero value. In this case, you will use a common flag value of 1.

An additional variable that may predict where new wolves will settle is current wolf territories. You will also assign a flag value of 1 to the polygons representing existing wolf pack boundaries.

a) Add a field to your Elk theme called Elk_flag and assign it a value of 1 for every record. Save and stop your edits.
b) Now using the Wolfpacks theme, add the new field called Pack_flag and assign it a value of 1 for all records. Save and stop your edits.

c) Union the Elk theme with the Wolfpacks theme. Then, union your results with the Humans theme. Name the new theme Predicted.

Step 4 - Apply a logistic regression model to the union theme

You will now apply the logistic regression equation to the *Predicted* theme to find areas that have a high likelihood of supporting a wolf pack.

a) Add a field to the Predicted theme called Sum. Define the field with 2 decimal places.

b) Review “Calculating a field's values” using the ArcView Help Topics. Menu. Now calculate the Sum field equal to the following expression:

\[-0.847 + (1.386 \times \text{[Elk_flag]}) + (0.201 \times \text{[H_code]}) + (-2.197 \times \text{[Pack_flag]})\]

Review the values in the Sum field to make sure you calculated the field correctly. The logit values should range from -2.84 to 1.54. Check the statistics for that attribute field to confirm the min/max values.

Step 5 - Calculate probability of wolf habitat preference

Now, to make the final values easier to understand, you will convert the logit values to probability values.

a) Add a field to the Predicted theme called Prob. Define the field with 2 decimal places.

b) Calculate the Prob field equal to the following expression:

\[\text{Number.GetEuler}^{\text{Sum}} / (1 + \text{Number.GetEuler}^{\text{Sum}})\]

Note: ArcView GIS does not have an Exp request, which is the easiest way to convert logits to probabilities. The ArcView Spatial Analyst extension does provide this request. As an alternative you could use Spatial Analyst and the expression: \(\text{Exp[Sum]} / (1 + \text{Exp[sum]})\)

c) Stop editing the theme table and save your edits.

Review the values in the Prob field. Probability values always range from 0 to 1.

d) Classify the Predicted theme in the view based on the Prob field and load the prob.avl legend file. Prob.avl is located in your lab data folder.

Notice that there is a class break at 0.5. A probability value of 0.5 means that there is an equal chance that wolves may or may not prefer those areas. The two classes represented by the darker greens show habitat which wolves may prefer. The lower three classes represent areas where there is little likelihood that the released alpha pair will establish territory.

Step 6 - Exclude lakes from final theme
You have successfully predicted suitable wolf habitat. But, before your analysis is complete, you also must take into account areas where you know the alpha pair will not settle—lakes. You will now flag the lake polygons as excluded.

a) Add a field to Lakes.shp called Excluded and assign it a value equal to -99. Save your edits and stop editing.
b) Union Lakes with Predicted. Name the final shapefile Wolves.
c) Create a new field in Wolves called Final_prob. Define the field with 2 decimal places. Calculate Final_prob equal to Prob.
d) Using the Query Builder, select polygons that have values equal to -99 in the Excluded field. You should have approximately 983 polygons selected.
e) Calculate Final_prob equal to -99 for these highlighted polygons. Stop editing the table and save your edits.
f) Classify the Wolves theme based on the Final_prob field using the prob.avl legend file. Add a new legend category and assign -99 to the null value and display it in the legend. Change the label for the null value class to Excluded Areas.

When you are finished, your view will look similar to the one below.

Figure 2.0 – Probability Surface after step 6
Calculating Polygon Area Values Using Avenue Scripts in ArcView GIS:

You will need to calculate the area for each polygon in the final probability theme. In your project data folder you will find a file called calcapl.ave. This is an avenue script file that you can run in ArcView to conduct geometric measurement on shapefiles. This script file calculates area and perimeter for polygon themes.

Follow these steps in the script window:

1) Under the /View/Properties.. ensure a Distance Units for this project is set;
2) Ensure the wolves.shp theme layer is select in the TOC;
3) Under the /Script menu “Load in a text file” and select calcapl.ave from the data folder;
4) Read through the description in the window and then compile the script using the compile button.
5) Now run the script using the run button.
6) Examine the attribute table of for the data theme the script operation was applied on. Notice the new Area and Perimeter fields. These field values are units area as specified in the View Properties dialogue menu. The records in the tables can be sorted in ascending or descending order by field using the tool bar.

(Note – on your desktop, the attribute table, map window and the script window must be open and visible for this script to work when you execute the command. This script is quite temperamental)

Step 7 – Refine the Analysis with Objective Facts

Now that you have completed your analysis, try to determine where the alpha pair will most likely establish territory. Wolves need at least 74 square kilometers of territory for hunting per pack.

We want to find areas with a final probability of .82 or greater with an area reported at 74 km² or greater. These areas should also be in good proximity to the release wolf den (i.e., Fishing Bridge). You should be able to narrow it down to 3 areas using the probability and area criteria and then down to 2 areas using the proximity to the release area (i.e., The Fishing Bridge Wolf Den).

Calculate the area for all of the polygons in the wolves.shp theme. Use the process of sorting attribute fields, selecting records, and then create a new shape file to finish the exercise. Make your own decisions to select 2 final polygons that are the most reasonable areas where you might expect the wolves to resettle based on the criteria given above. Ensure you discuss your reasoning in your lab report.

End.