TigerSAMPLING: Sampling Lab

Introduction to TigerSAMPLING

The Bol'shaya Koshka (Russian for big cat) Reserve is a newly created animal reserve that was uniquely developed to help endangered species prosper. This 10,000 acre wild animal reservation was selected because an abundance of Siberian tigers have been found in the area. The diverse terrain of the reserve provides a wide variety of habitats for many different species of animals.

Since the tigers in this area are much more abundant than any other area in the world, they are starting to draw a significant number of researchers to the region. Your primary responsibility will be to help these researchers as they study the tigers and then incorporate the results of their research into a system to identify the best management practices for this reserve.

An important component of monitoring endangered species is to understand the age distribution of the population. Shifts in the distribution could indicate potential issues in sustaining the population. Additionally, initial research will be geared at simply determining how many tigers are in the preserve including where they live. Eventually, the data collection will extend to issues of tiger health issues amd tracking the population to ensure this endangered species continues.

At this preliminary stage, you will be exploring the preserve and collecting the first sample of tigers.

TASK #1: Background research on the Amur tiger

Read about the *Amur Tiger*. A short article is found at (http://www.wcsrussia.org/en-us/wildlife/amurtiger.aspx#.U-oTSygZS9c) (if this link does not work, go to the Wildlife Conservation Society Russia home page at: http://www.wcsrussia.org/en-us/home.aspx#.U-oTwSgZS9c and find Amur Tigers under the "Wildlife" menu). In your reading, consider the following questions:

1. Where do tigers live? What sort of terrain, climate and so forth is their habitat?
2. How dense would you expect tigers to be in the preserve?
3. What, if any, characteristics of tiger behavior will impact collecting data on them?
4. What tiger data would you want to collect in conducting future studies?
5. Based on the reading, what would be your strategy for finding and collecting data on tigers in the preserve? How would you be able to determine if your sample is large enough and represents the population of tigers well enough to be able to use in answering questions about things like the age distribution of tigers?

Play the tutorial for the TigerSAMPLING game briefly so you are familiar with the game controls. The game is found at the web site: http://statgames.tietronix.com/TigerSampling/ and select **Load Tutorial**. If you forget commands anytime during game play, you can hit the "p" key to pause the game and see game instructions.

Collect Tiger Data using TigerSAMPLING. Go to http://statgames.tietronix.com/TigerSampling/ and enter a PlayerName and GroupName (The "PlayerName" is a secret name, any combination of letters and numbers with no spaces. Do not use your name or a term that will identify you or your group. All group members should use the same "PlayerName"). The "GroupName" will be provided by your instructor. Select **Load Mission 1** and then **DataSet1**. Use the Full Screen option to see the entire game on your computer screen. You can type "**p**" to **pause** anytime while playing the game. This will allow you to review all the controls, exit the game and save your data.

TASK #2: Preliminary data analysis

For this task we will examine the tiger data collected to determine if it is appropriate to use in developing a model for tiger age prediction and potentially other research questions of interest. Your TigerSAMPLING dataset contains several variables:

- Age (in years) – a quantitative variable that ranges from 0 to 20
- NoseBlackProportion – the percent of the tiger's nose that is black
- Weight (in pounds)
- Sex (Female or Male)
- Size – the length of the tiger (in inches)
- Pad – the width of the back paw (in inches)
- Region – one of 6 regions of the preserve (A, B, C, D, E and F) defined by terrain features

After you have collected a dataset, answer the following questions

6. Do a preliminary inspection of the data – are there any missing observations? Are all of the tigers different or did you collect from the same tiger more than once? Are there any other noticeable data entry problems?

7. Produce summary statistics and graphs for each of the variables in the data set. Boxplots or histograms and means/standard deviation should be used for continuous variables. For categorical variables (Sex and Region) frequencies in each category and bar charts may help.

8. Based on these summaries, do you think there are issues with your sample of tigers? Are there tigers in the preserve that may not be adequately represented in the sample? From the reading in task one we know, for example, that the age range for Amur tigers is typically 0-20 years old.

TASK #3: Further data analysis – looking at regions

9. Examining the sample further, are there differences between tigers in different regions? Produce summary statistics for one of the variables that measure tiger size by region (weight, size or pad). What do you observe? If there are differences, is this a concern in using the data in models to answer research questions? Why or why not? If there are differences, what are possible explanations?

10. If you have a large enough sample size, conduct a statistical test to determine if the average ages differ between region A and D. Do you have statistical evidence of a difference?

11. Repeat the analysis in #8 (and #9) but instead of looking at age, compare gender in the different regions. Note that gender is not a continuous variable so different statistics/graphs will be needed.

One of the first simple research questions of interest is determining the difference in characteristics between male and female tigers.

12. In your sample, is there a difference in size between male and female tigers (you may choose one indicator of size)?

13. What factors could impact the analysis in question #11? Hint: think about the role of each of the other predictors in the data set. What would be a better approach to answer the question based upon your data?

TASK #4: Summarizing results and conclusions

14. Based on the results and observations made in this lab, write a short summary about the appropriateness of your sample for use in research and making inferences about the population. Address at a minimum the questions:
    - Is your sample representative of the tiger population more generally? If not, why? If so, give support for this answer.
    - Would your sample be appropriate for answering questions like estimating the average size or age of tigers in the preserve?
    - If another sample were taken, what recommendations would you make for collecting the data? Are there dangers with simply tranquilizing tigers "randomly"?

A final note about random sampling and bias: all is not lost if a sample has issues such as regional differences, or data that differs on some other characteristic between comparison groups. Sometimes, more complicated models can account for or "adjust" for these disparities. A multivariable model, for example, might suffice. Of course, a better option is to carefully plan the sampling to ensure bias is not present. A simple random sample is not always "best" and other sampling schemes are possible. Review additional statistical studies or consult a statistician prior to collecting or analyzing data that could have these issues.