Name: Aaron Marks

Geography 364—Lab Assignment 4 Answer Sheet

Regression Models

Due Date: Friday, November 19th, 7 p.m., 2010

Total points available: **100**

Type your answers to questions in the spaces provided. **DO NOT** turn in hand-written answers.

Q1 1) Identify two variables in the data set that you think might help to estimate species diversity on an island; 2) Provide an explanation of why these variables would be useful. Indicate whether you expect a *positive* or *negative* relationship in each case.

Variable 1: Area **5 pts**

Variable 2: Number of species

Two variables that may estimate species diversity would include the number of species and the area. To me the idea of diversity is many different types of species in a specific area, A large island may be just as diverse as a small one depending on the number of species found within such areas, and vice versa. I would expect a larger island to contain more species than a smaller island, and thus be more diverse, which would relate to a positive relationship, or correlation. **5 pts**

Q2 Based on an examination of the correlations, are you happy to proceed to building a regression model based on one of the variables you suggested in your answer to Q1? If so, why, and which *one* variable do you intend to use, and why? If not, why, and which other variable do you intend to use and why? **10 pts**

I am happy with my results to proceed with making a regression model based on area because in theory a larger or smaller area, should determine the large or small number of species.

Name: Aaron Marks

Q3 You should have decided which is your dependent variable and which is your independent variable. Tell us which is which:

Dependent variable: Number of species **2.5 pts**

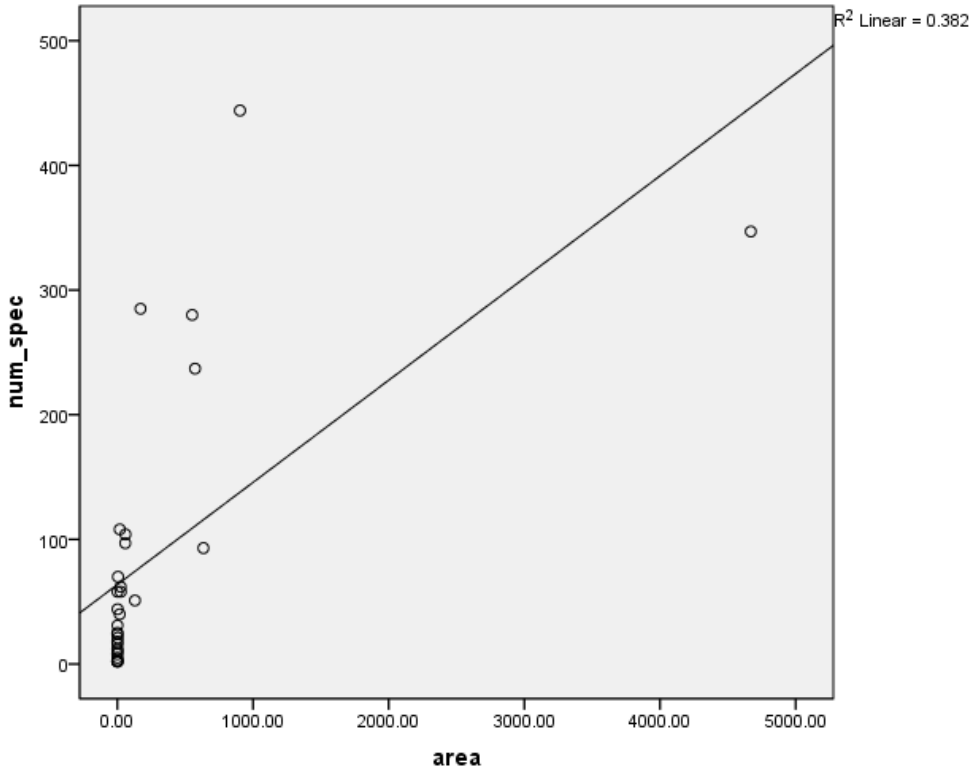
Independent variable: Area **2.5 pts**

Q4 From the regression results, write down the equation that represents your model: **5 pts**

$y = a + bx$

Number of species = 63.783 +0.082(area)

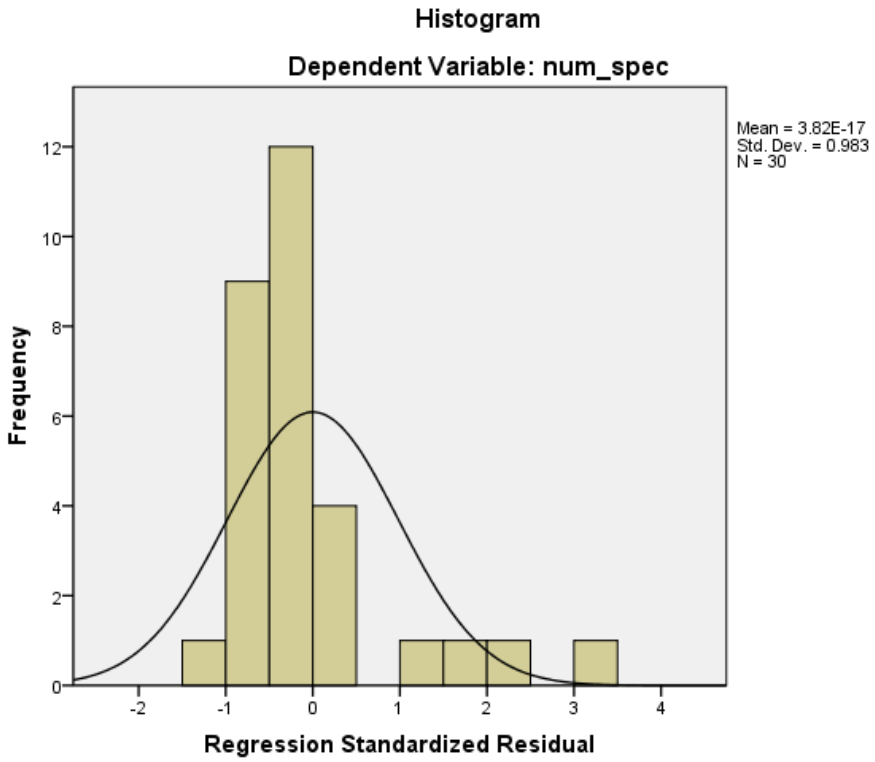
Q5 Make a single scatter plot of your dependent and independent variables. Attach this graph with a line drawn on it representing your model. **5 pts**



Q6 Based on the histogram of the model error or residuals, would you say that the error is normally distributed? (Just describe the distribution and why it appears normal or not to you). Attach a screenshot of the histogram. **10 pts**

Based on this histogram it would appear that the error is normally distributed. While there is a high count of values just below 0, there are a few outliers out past 2

and 3. Also the sig value given for the T-statistic are .001 and .000



Q7 What is the model's R^2 value? **5 pts**

R^2 = .382

Q8 Examine the ANOVA table. (1) Is the model statistically significant? (2) State which element in this table leads you to this conclusion. **5 pts**

The model is significant because the regression coefficients are significantly different from 0 and have a p value of .000 which is <.05.

Q9 Comment on anything in the Coefficients table that you feel is important (for example, note the coefficient *t* test *p*-values, and/or their 95% confidence intervals: what do they tell you about the model?) **10 pts**

The T values and their respective p values show that the corresponding coefficients are significant. As for the 95% confidence interval, these ranges include 0 and are therefore significant.

Name: Aaron Marks

Q10 PASW (SPSS) will identify any unusually bad errors in the model in a table labeled **Casewise Diagnostics**. This table also identifies the large error cases by their 'case number', i.e. their row number in the data view. **15 pts**

Does your model have any unusually large errors? If so, (1) determine which island(s) they relate to and (2) suggest reasons that the error might have occurred—is there anything unusual about the island, for example (as far as you can tell)?

There is a large error on case number 25 which is Santa Cruz. The reason behind this is the expected value of the area was 137.86 and the actual value was 903.82. The most unusual part about this island is it is the one with the highest number of species yet isn't the largest island. It has more species than an island more than 4x its size.

Q11 (1) What does your answer to Q10 tell you about the usefulness of your model? **10 pts**

(2) Discuss which other variables (either in the data set or ones that might make a difference, which are *not* in the data set) that you think would be useful in understanding biodiversity in the Galapagos.

For only having 1 error I would say it's a pretty good model, with just the one exception. Other variables may include the things such as temperature, rainfall or more importantly the land cover. An island with more buildings and parking , or crops, may effect the actual area species can live in. A large open island may have more species than a similar sized island that has been developed.

Q12 Finally, discuss how useful you think the regression modeling process is in improving your understanding of biodiversity in the Galapagos. **10 pts**

Besides the one exception it was almost a perfect model that could be used in understanding the number of species related to the size or area of an island. This can be more useful for data that is much less obvious than species compared to area such as things from voting to pollution to interest in outer space or waste generation. It's a nice model to be able to use to find out significant correlations between variables.