A Study of the Relationship between Economic Development and Environmental Condition of Countries in the 21st Century

Author: Xuehan Liu, Senior at Department of Statistics, email: xuehanl@iastate.edu Project Advisor: Prof. Mervyn Marasinghe, Department of Statistics

Introduction

In the past decade, people have begun to pay considerable attention to the Earth's environment. In order to accelerate development, countries have cleared forests to build factories and let sewage into the rivers. Animals have lost their habitats and are threatening to become extinct. It appears to be that the growth of economy is accomplished by sacrificing the environment. Based on this background, I want to investigate whether evidence exists for the hypothesis that the economy and the environment impact each other on earth.

There will be two parts of analysis in this project. The first part focuses on economic status and threatened species, under two research questions:

- Is there any association between threatened species levels and GDP levels?
- Does the country with better economic status have better environment condition?

The second part studies on the threatened species along with both natural environmental indicators and economic developmental indicators, with one research question:

- Can the number of threatened species in a country be predicted from the country's natural environment indicators (such as forest area or annual fresh water withdrawal) and economic development indicators (such as GNI, household expenditure, and value of industry) ?

Data and Method

The data is extracted from the World Bank website. It contains various demographic, geographic, and environment-related variables collected on 249 subjects since the year 2000. However, there are only 196 countries (195, if not include Taiwan) in the world.

For Categorical Variables:

- Mosaic Plots
- Contingency Table
- Proportion Test
- Chi-square test for independence
- Log Transformation of the rare data

For Quantitative Variables:

- K-Mean Cluster Analysis, involving with principle component analysis
- Multiple Linear Regression (MLR)

Discussion

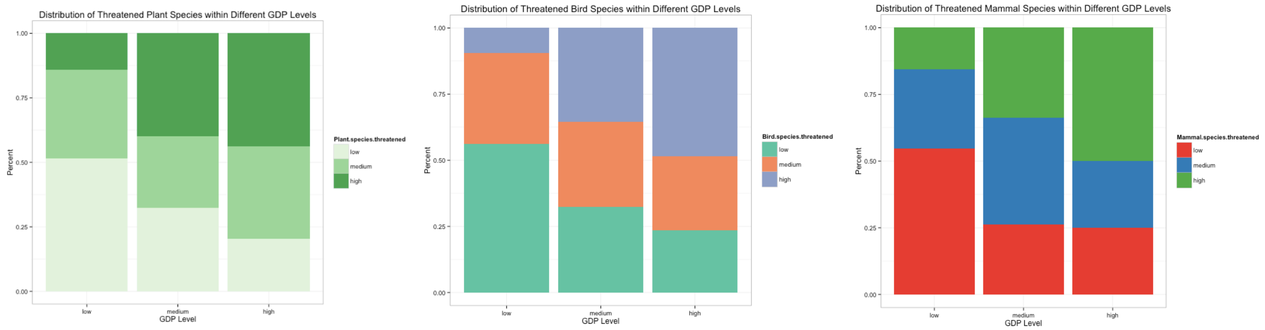
Excluded Potential Outliers :

- In both cluster analysis and MLR: **The United States, China, Japan, Germany (**extremely high values for economic indicators), and **Ecuador** (only excluded in clustering, because of extremely high number of threatened species).

- In MLR: **India, Russia, Australia, Brazil** (extremely high values for all natural environmental indicators),

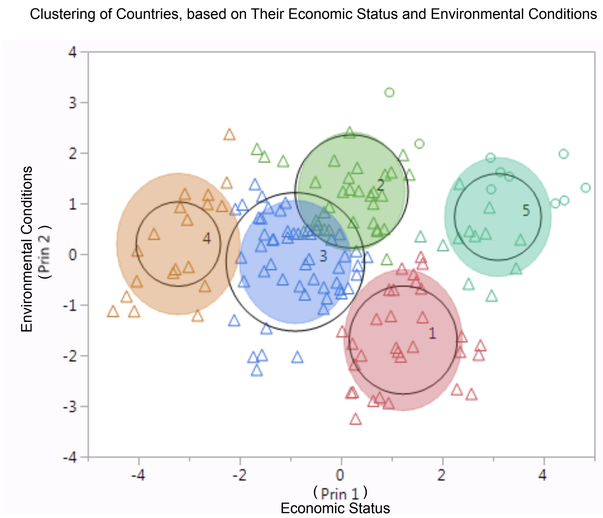
Result

Research Question One:



- Goodman-Kruskal Gamma for associations: for Plant vs. GDP is 0.389, for bird vs. GDP is 0.457, and for mammal vs. GDP is 0.415.
- The Goodman-Kruskal Gamma for all three contingency tables are moderate. There are moderate agreements between the level of GDP and the level of each of threatened plant, bird, and mammal species.
- As the level of GDP increase, the level of threatened plant, bird, and mammal species tend to increase by in a very moderate strength.

Research Question Two:



- The K-means cluster method uses the first two principle components to compute scores for each country and then plot the scores of 1st vs. 2nd principle components.
- Six variables: number of threatened plant, bird, and mammal species, household consumption expenditure, GNI, and value of industry.
- Prin 1: 1st principle component; Prin 2: 2nd principle component
- Score of a country for 1st principle component:
  *0.357*threatened bird + 0.298*threatened mammal + 0.284*threatened plant + 0.476*household expenditure + 0.487*GNI + 0.489*industry value*
- Score of a country for 2st principle component:
  *-0.465*threatened bird – 0.522*threatened mammal – 0.464 *threatened plant + 0.335*household expenditure + 0.331*GNI + 0.272*industry value*
- High score on Prin1=> high economic status
- High score on Prin2=> low number of threatened species => high environmental condition.
- Countries in cluster 2 and 5 have both relative high economic status and high environmental condition
- Countries in cluster 3 have medium level on both economic status and environmental condition.
- Countries in cluster 1 and 4 have high level on one, but low level on the other.

Research Question Three:

- Model fitted: $\log\hat{C}$= 4.64+ 0.51*log(household consumption expenditure)-0.69*log(GNI)+0.22*log(fresh water withdrawal)-0.14*log(arable land) +0.33*log(forest area)+0.10*log(value of industry)
- R-square=0.441. About 44.1% of the variation in threatened species can be explained by the model with variable forest area, fresh water withdrawal, household consumption expenditure, GNI, value of industry, after log transformation.
- This is not a very good model. Mainly because the R-square is too small, even though the all assumptions for the model is met.
- Try fit quadratic terms in the model to improve R-square, but the R-square goes down.
- Correlation between the response and explanatory variables are not very strong. The strongest correlation is 0.6628, between threatened species and forest area. Note the correlation is positive.
- Also try use other variables, such as percent forest area, CO2 emissions, and greenhouse gas emission, and agricultural land, to find a better model. Using these variables still fails to improve the R-square. Their correlations with response are even weaker than previous variables.
- Using data-based approach to predict the number of threatened species may not be valid.

Summary

- As the level of GDP increases, the level of threatened plant, bird, mammal for about half countries also tends to increase.
- About half countries have consistent level of both economic status and environment conditions. The other half countries have high level of one aspect and low level of the other aspect.
- In the data-based analysis, economic and natural environmental indicators are not ideal in predicting the number of threatened species, particularly for plant, bird, and mammal species.
- It is surprise that the variable forest area and the number of threatened species, particularly for plant, bird, and mammal have the inverse relationship.
- One thing can be continued in the future is that, in order to predicting the number of threatened species in a country, it should be plausible to study the threatened species in the biology thesis, in stead of modeling based on the available data, and then combing with statistical skills to find a model for prediction.

Reference

What are the Goodman-Kruskal statistics? (n.d.). Retrieved April 16, 2016, from http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/tables/other-statistics-and-tests/what-are-the-goodman-kruskal-statistics/

Data World Development Indicators| World DataBank. (n.d.). Retrieved March 20, 2016, from http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators#

Mongabay.com. (n.d.). Retrieved March 20, 2016, from http://rainforests.mongabay.com/03highest_biodiversity.htm

Smith, T. E. (n.d.). A SUGGESTED PROCEDURE FOR MULTIPLE REGRESSION ANALYSIS. ESE 302. Retrieved April 5, 2016, from http://www.seas.upenn.edu/~ese302/extra_mtls/Multiple_Regression.pdf