What Affects Happiness?

Nathan Sanders, Yunzhe Li

11/16/18

Abstract

Happiness of people on country level is a very important measure for government to develop and determine new policies for next year. It is necessary to construct a model to help predict it. Country's economy, government's corruption, health, freedom, contribution of family, generosity, and location all possibly have influence on the happiness of people. All of them should be considered in the prediction model.

Introduction

Background

An organization, Gallup conducts a survey called the Gallup World Poll for around 155 countries. In this poll they use simple random sampling to pick their samples. They survey roughly 1000 people for each country, 2000 for large countries like China and Russia. Very rarely in some countries they have 500-1000 surveys. They also only conduct surveys on adults, who are age 15 and up. To account for places that do not have an easy access to phones, they used two types of surveys. One survey is face to face, which takes roughly an hour, and the other one is through the phone which takes roughly thirty minutes. Sustainable Development Solutions Network, another organization, uses this poll and summarizes questions from the poll into six factors which include, Economy (GDP), Family, Health, Freedom, Trust, and Generosity. More details about those factors will be discussed later.

Objective

Construct a prediction model for a given country. In order to do this we need to do conduct exploratory data analysis about the data first.

Data Description

Data set is downloaded from Kaggle and it is originally provided by Sustainable Development Solutions Network. There are three csv files, each of them corresponds to a year, from 2015 to 2017. For each file, there are more than 150 countries with 12 different variables.

- Country: 155+ countries.

- Region: 10 different regions.

- Variation: There are three different measures for the variation for happiness score. (Standard Error, Confidence Interval, Whisker Low and Whisker High.)

- Happiness Score: Score that is scored from 0 - 10.

- Happiness Rank: The rank of a country in happiness.

- GDP per capita: Gross Domestic Product.

- Family: A value that determines how questions in the Family section in the Poll relate to happiness score.

- Health: A value that determines how questions in the Health section in the Poll relate to happiness score.

- Freedom: A value that determines how questions in the Freedom section in the Poll relate to happiness score.

- Trust in Government: A value that determines how questions about the individuals trust in government relate to happiness score.

- Generosity: A value that determines how questions that relate to generosity in the Poll relate to happiness score.

- Dystopia Residual: The value of the difference between a theoretical country's happiness score and the Countries' happiness score.

Plan & Methodology

First we want to do exploratory data analysis to determine if year factor and region factor have effect on the data. Then we will check if six variables have strong linear relationship with happiness score, and if there is correlation between them. We will use step-wised model selection method to finalized our linear model for happiness score and six factors. PCA might also be used to determine the amount of contribution of each factor. Finally, we will use this model to predict the happiness score for a given country with a prediction interval.

- Exploratory Data Analysis.

- Two-way ANOVA for testing the effects of Region and Year.

- Linear regression for a full model with all six different factors.

- Model selection base on full model. Construct a model for the most important factors.

- Possibly PCA to find the linear combination of different factors.