Examining Clinical Trials By Countries Through Principal Component Analysis

Description of the Topic:

Using Principal Component Analysis, I will examine what characteristics (through indices and indicators) in a country cause a pharmaceutical company to conduct clinical trials in that country. The goal is to determine what factors contribute most to a country's likelihood of being chosen to host clinical trials. The analysis will be conducted in three ways; one will be using all the data. The second analysis will break the data into two time periods. The third analysis will look at individual pharmaceutical companies choices of clinical trial location sites. I am conducting different analyses because I will then be able to compare the indices and determine which are the most influential in whether a country is chosen. The overarching research question that I am trying to answer is: what combination of indices and weightings provides the best measure of the number of clinical trials a country is likely to host?

Background Research of Related Work:

For this research project there are three areas of background research that apply. The first area is the exploration of the diffusion of clinical trials in new countries. It has already been observed that there is a shifting of the location of pharmaceutical research and development.[1] The U.S.A. is still a leader in pharmaceutical research and development, followed by other modern westernized countries such as the UK and Germany. However, other non-"traditional" countries are beginning to host pharmaceutical research and development.[2] When looking at the locations of clinical trials, countries such as the Czech Republic, Hungary, and Estonia are beginning to conduct and host many clinical trials.[3] The current literature has established that clinical trials are becoming more globally widespread and are no longer restricted to a few choice countries.

The second area of background concerns other sectors that have looked at diffusion and why. Northeastern Professor Rodine-Hardy has done a similar analysis using event history analysis and survival models to look at different indices and the effect they have on the liberalization of telecommunication companies.[4] Rodine-Hardy primarily examines the relationship between liberalization and membership in international organizations. The difference between this research project and her work will be the idea of working on a spectrum. In Rodine-Hardy's work, a survival analysis was necessary because membership and liberalization were one-time events. This project will instead look at indices and the number of clinical trials; both of which can vary and are not limited to a binary option.

The third area of background material concerns indices and machine learning techniques for comparing countries. Previous work using indices and principal component analysis has been used to compare the different markets of countries.[5] The goal of this type of analysis is to identify similar countries that may not usually be associated as like-countries, as well as to help identify weaknesses or deficiencies in countries that may be affecting their markets. Other principal component analyses have been conducted using indices to measure the type of political economy in countries and cluster like-countries together based on the principal component analysis.[6] The goal of the political economy based clustering is to find more optimal means by

which to describe countries by using multiple indices rather than describing countries solely through a single index.

Data Source:

There are two data sources for clinical trials that I will be using. First, ClinicalTrials.gov is a database of clinical trials maintained by the U.S. National Institutes of Health. [7] It has data from 2008 through the present. Second, ADIS is a paid for database of clinical trials from Springer which was provided to me by Professor Dunlap.[8] The ADIS database has clinical trial data spanning 1996 to 2006. Furthermore, I will use indices from the Worldwide Governance Indicators,[9] World Health Organization's Global Health Observatory,[10] and the World Bank Indicators.[11]

Algorithms and Code Sources:

In this research project I will use Principal Component Analysis from the psych[12] package in R and the GPArotation[13] package in R. Additionally, I will use the ggplot2[14] package in R for plotting.

Evaluating The Success of the Project

The success of the project will be evaluated based on whether a Principal Component Analysis can be created that works over both time and for multiple companies. The primary goal is to find certain indices that can generally predict the likelihood of a country being chosen as a host for clinical trials. Success can then be measured in how much variation the chosen indices can explain when examining the likelihood of a country being chosen. Success can also be partially evaluated in the number of indices needed in the principal component analysis to explain a majority of the variation; the fewer indices needed, the more likely that these indices have a real world relationship with the likelihood of being a clinical trial host and are not just the product of over-fitting.

References:

[1] Cockburn, Iain M., and Matthew J. Slaughter. "The global location of biopharmaceutical knowledge activity: new findings, new questions." *Innovation Policy and the Economy, Volume 10*. University of Chicago Press, 2010. 129-157.

[2] Thiers, Fabio A., Anthony J. Sinskey, and Ernst R. Berndt. "Trends in the globalization of clinical trials." *Nature Reviews Drug Discovery* 7.1 (2008): 13-14.

[3] Thiers, Fabio A., Anthony J. Sinskey, and Ernst R. Berndt. "Trends in the globalization of clinical trials." *Nature Reviews Drug Discovery* 7.1 (2008): 13-14.

[4] Rodine-Hardy, Kirsten. "Globalization, International Organizations, and Telecommunications." *Review of Policy Research* 32.5 (2015): 517-537.

[5] Grein, Andreas F., S. Prakash Sethi, and Lawrence G. Tatum. "A Dynamic Analysis of Country Clusters, the Role of Corruption, and Implications for Global Firms." *Department of Marketing and International Business. URL: idec. gr/iier/new. com* (2008).

[6] Breunig, Christian, and John S. Ahlquist. "Country Clustering in Comparative Political Economy." (2009).
[7] ClinicalTrials.gov, https://clinicaltrials.gov/
[8] Springer, http://www.springer.com/gp/adis
[9] Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. "Governance matters VIII: aggregate and individual governance indicators, 1996-2008." *World bank policy research working paper* 4978 (2009).
[10] World Health Organization, http://apps.who.int/gho/data/node.imr
[11] World Bank, http://data.worldbank.org/indicator
[12] https://cran.r-project.org/web/packages/psych/psych.pdf
[13] https://cran.r-project.org/web/packages/GPArotation/GPArotation.pdf
[14] https://cran.r-project.org/web/packages/ggplot2/index.html