Power Analysis for Geographic Representativeness

M. D. Schmill, T. Oates

University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250
Telephone: 410-455-3017
Email: schmill@umbc.edu, oates@umbc.edu

Abstract

In many disciplines, researchers attempt to make global inferences from a sample of local case studies. In those cases, especially where the sample size is small, it is necessary to demonstrate that the sample is *unbiased*, a property we call “representativeness”. Most statistical tests are designed to detect bias, not the lack of it, and it is an error to simply accept the null hypothesis when a statistical test fails to reject it. The GLOBE system helps researchers demonstrate representativeness by augmenting traditional statistical tests (such as X^2) with *power analysis*.

Keywords: geographic representativeness, power analysis, spatial analytics, sampling bias, land change science

1. Introduction

An important notion to those who conduct synthesis studies, in which collections of local case studies are connected to global data in order to produce globally relevant conclusions, is that of *representativeness*. Representativeness is the extent to which conditions of interest (for example, population density) at the local study sites fit the global patterns for the same variables.

A related statistic, which we call *representedness*, measures the extent to which the conditions at an arbitrary geographic region are properly represented in the global range by the collection of case studies. Where representativeness allows for a quick judgment of whether or not site selection for your synthesis study appears to contain bias, representedness can be used to, among other things, render a map, allowing for quick judgments to be made on where, geographically, the bias exists.

In prior work, we presented GLOBE[1], an online visual-analytical tool for assisting land change scientists in conducting representativeness analyses (Schmill, Gordon, Oates, Magliocca, & Ellis, 2014). The GLOBE system provides several methods for assessing bias, including X^2 analysis and Monte Carlo methods based on *f*-divergence measures such as Jensen-Shannon Distance (Lin, 1991). Each of these methods tests the hypothesis that the distributions of variable values are the same for the collection of local case studies as it is for a global range. These methods return a *p* value, which is the probability of incorrectly rejecting the null hypothesis (and concluding bias).

These analytics are very helpful in detecting and visualizing bias, and additional features of GLOBE allow researchers to address bias, either by searching for more cases or by exporting case weightings for use in weighted analyses. However, if one wishes to conclude representativeness (practical equivalence of the collection and global

---

[1] http://globe.umbc.edu/

distributions), it is a mistake to do so if the test simply fails to conclude bias (Cohen, 1996). This is because the *power* of the test – the probability of correctly rejecting the null hypothesis if it is false – has not been established.

In this paper, we introduce a new GLOBE tool that allows researchers to perform power analysis on case study collections to make a case for representativeness, and how power, like representativeness, can be visualized geographically. We also show how the relationship between sample size and power can be used to illuminate what is required to address tests with weak statistical power.

2. Representativeness

Some background on representativeness analysis is necessary to understand how power analysis will work. While there are a number of ways to implement representativeness, we will focus here on our X^2-based method. Suppose the researcher has collected a set of case studies that concentrate on the impact of high agricultural intensity in rice villages. They would like to summarize their findings and substantiate a claim that their findings generalize to all places where rice is cultivated. Now suppose that a paper reviewer criticizes the work, claiming that the sample of cases is biased towards sites that are easy to access.

In this case, a representativeness analysis considers one variable, *accessibility*[2], over the range of terrestrial Earth where rice cultivation is non-zero. Since X^2 requires categorical data, the user specifies a mechanism for making accessibility, a continuous variable, discrete. They can choose GLOBE's recommendation, or use their own. They might choose cut-off points for three categories: low, medium, and high, where "high" indicates that the site is very accessible. The representativeness analysis will compute the multinomial distribution of the accessibility categories for the case collection and compare it to the distribution of GLOBE land units meeting the criterion of non-zero rice cultivation.

Pearson's X^2 test can compute from these two distributions a *p* value that can be used to make a judgment of bias. It, however, does not consider *where* the bias lies. For this, we use a related measure which we call r*epresentedness,* which re-imagines the 3-category multinomial as a series of three 2-category binomials, each being assessed using X^2 and the probability of a single category versus "everything else". With representedness, the researcher can zero in on where the bias lies, both in variable space and geographically, by rendering a map. If the X^2 test for the "high" accessibility range shows a significant difference, the criticism is confirmed, and the author needs to address the bias.

When bias is not detected, however, the researcher has more work to do. They must further demonstrate that the lack of bias is not due to a weak test or insufficient data.

3. Power Analysis

The primary result provided by a classical statistical test like X^2 or a Monte Carlo analysis is the probability of a type I error, denoted α – the probability of incorrectly rejecting the null hypothesis H_0 (thereby incorrectly accepting the alternative hypothesis).

[2] Measured as the distance in km to the nearest city (Nelson, 2008).

A type II error occurs when a test *fails* to reject H_0 when an actual difference is present, and its probability is denoted β. The statistical power of a test is equal to (1-β).

Calculating power is not always straightforward, and its value depends on the amount of data, the amount of effect considered significant, and the significance value α. It is particularly challenging when considering the multinomial case, as it is not sufficient to specify the *amount* of effect (bias), but also *how* the effect is realized in the multinomial. While we have implemented a Monte Carlo method for evaluating power in the multinomial case for various specified effect types, it is not an easy analysis for a user who is not expert in statistics.

For a binomial test of proportions, however, an analytical solution exists and there is only one way to express an effect on the proportion. The formula for 2-sided equality of 2 proportions, is as follows (S., J., & H., 2007):

$$1-\beta = \Phi\left(z - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) + \Phi\left(-z - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)$$

Where Φ is the standard normal cumulative probability function, Φ^{-1} is the standard normal quantile function, and

$$z = \frac{p - p_0}{\sqrt{\frac{p(1-p)}{n}}}$$

Here, p is the proportion of the category being assessed in the global data, p_0 is the biased proportion, and n is the sample size. The formula can also be solved for n to compute the sample size required to achieve a desired power given α.

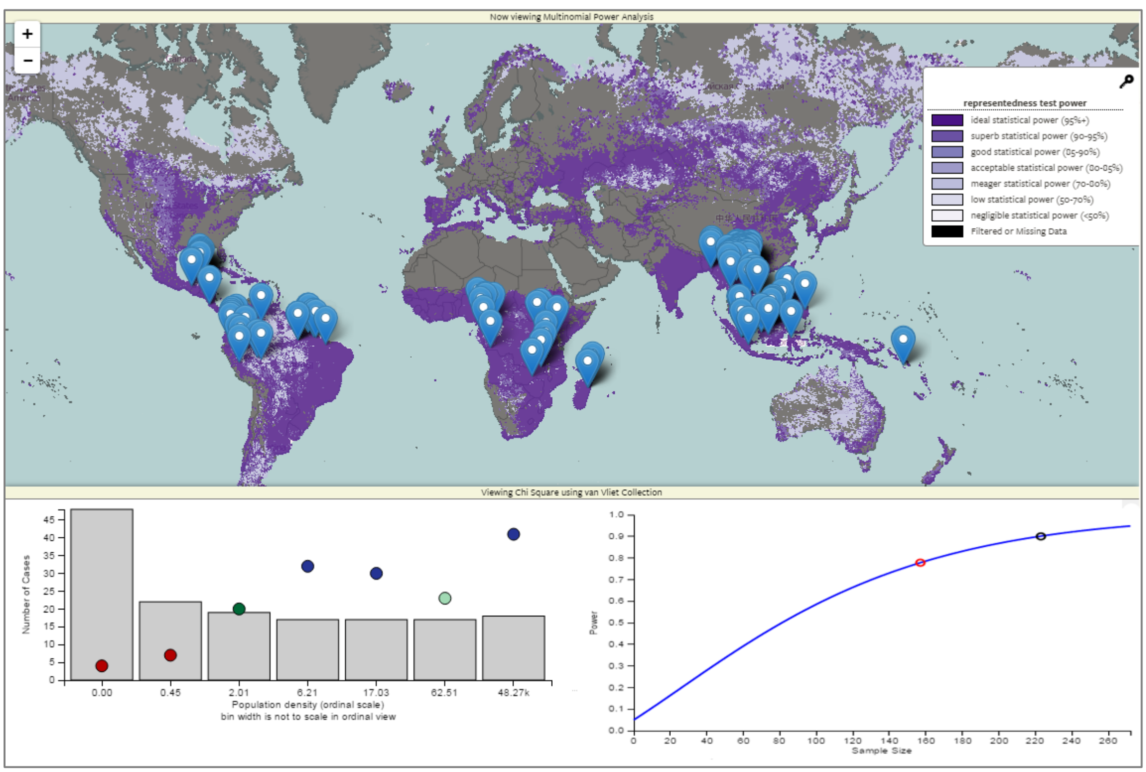

Figure 1. Screen shots from a GLOBE power analysis of 7 ranges of population density at 157 deforestation study sites. At top is a geographic representation of per-category

analysis power. At bottom left is a histogram showing representedness of the seven variable ranges, and at bottom right is a power curve for the lowest power category.

Figure 1 shows a GLOBE screenshot. At top is geographic representation of power. GLOBE land units (100 km2 hexagonal cells comprising a discrete global grid) are rendered as a map layer, each cell being colored according to the statistical power for the category to which it belongs. Darker shades of purple represent better statistical power, while white represents unacceptably weak conditions. These maps can be generated in real-time for exploration at arbitrary zoom levels, allowing the user to pinpoint geographic regions where the test is least likely to identify more subtle levels of bias. Below are plots of representedness (left) and a power curve (right), relating n to the statistical power of the weakest condition of the multinomial. This places a lower bound on power for the user's test and indicates what sample size is required to achieve the desired power.

4. Summary

It is a common statistical error to conclude a null hypothesis is true when a statistical test fails to reject it. This is especially true when researchers without expert statistical knowledge are attempting to demonstrate an unbiased sample – a task very common to Earth and geographic sciences where local studies are considered in aggregate to make global inferences. The power analysis tools and visualizations provided by GLOBE help researchers to conduct better representativeness analyses, and, consequently, to make more correct inferences.

5. Acknowledgements

This material is based upon work supported by the US National Science Foundation under grant NSF # 1125210, and cosponsored by the Global Land Project (GLP; www.globallandproject.org). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

6. References

Cohen, P. (1996, October). Getting what you deserve from data. *IEEE Expert, 11*(5).

Lin, J. (1991, January). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 145-151.

Nelson, A. (2008). Estimated travel time to the nearest city of 50,000 or more people in year 2000. *Global Environment Monitoring Unit – Joint Research Centre of the European Commission*.

S., C., J., S., & H., a. W. (2007). *Sample Size Calculations in Clinical Research.* Taylor & Francis.

Schmill, M. D., Gordon, L. M., Oates, T., Magliocca, N. R., & Ellis, E. C. (2014). GLOBE: Analytics for Assessing Global Representativeness. *COM.Geo '14: Proceedings of the 5th International Conference on Computing for Geospatial Research and Applications* (pp. 25-32). Washginton, D.C.: IEEE.