Landscape and Urban Planning 112 (2013) 118–120

Contents lists available at SciVerse ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Comment

Comment on: "Spatial variation among green building certification categories: Does place matter?" by Cidell and Beata

Svetlana Pushkar*

Ariel University Center of Samaria, Department of Civil Engineering, Faculty of Engineering, P.O.B. 3, Ariel 40700, Israel

HIGHLIGHTS

- ► LEED green building credits have a hierarchical (correlated) design.
- ► Ignoring the multilevel modeling procedures leads to the Pseudoreplication ANOVA.
- ► Pseudoreplication leads to the inaccurate appraisal of LEED credits.
- ► Proper statistical assumptions are needed to develop sustainable landscape design.

ARTICLE INFO

Article history:
Received 21 April 2012
Received in revised form 2 October 2012
Accepted 29 November 2012
Available online 27 December 2012

Keywords:
Sustainable landscape
LEED
Three-stage sampling
Pseudoreplication ANOVA
Neofisherian significant assessment

ABSTRACT

In a paper recently published in this journal, Cidell and Beata (2009) used a one-way ANOVA test with a post hoc procedure (Bayesian model selection) to assess the importance of spatially sensitive certification of LEED (Leadership in Energy and Environmental Design) standards. However, this analysis is inappropriately applied as it fails to recognize dependencies and hierarchies among the observation units, which result in pseudoreplication. Using an alternative, three-stage sampling method based on the neoFisherian paradigm, this paper reanalyzed the spatial sensitivity of the LEED categories. Contrary to Cidell and Beata's findings, the reanalysis shows that the indoor environmental quality (aspatial) category seems to be more preferable than the sustainable sites and energy and atmosphere (spatially sensitive) categories for LEED projects in the United States.

1. Introduction

In a research paper recently published in this journal, Cidell and Beata (2009) studied the role of geographic location in the application of green building certification criteria. The authors used a parametric one-way ANOVA test with a post hoc procedure (Bayesian model selection) to compare United States Environmental Protection Agency (EPA) regions in terms of general LEED (Leadership in Energy and Environmental Design) categories (case 1), spatially specific LEED categories (case 2), the six general LEED categories for the EPA regions (case 3), and the eleven spatially defined LEED credits for the EPA regions (case 4). Based on their analysis, Cidell and Beata (2009, p. 142) conclude that their findings reflect "the importance of place in the growing green building field, and underlines the need for more spatially sensitive certification standards."

However, the authors apply the one-way ANOVA test under the assumption that their observational units are not correlated. I believe this to be a critical flaw in their understanding of the LEED categories and by ignoring this statistical assumption, the one-way ANOVA test in all four cases leads to an artificially inflated degrees of freedom due to pseudoreplication (Hurlbert, 1984, 2009; Hurlbert & White, 1993; Picquelle & Mier, 2011). In this commentary, I hope to clarify the issues and present an alternative approach aimed at providing a more accurate appraisal of LEED categories.

2. Reconceptualizing the LEED study design and sampling

It is a well-known fact that LEED, developed by the U.S. Green Building Council (USGBC), has become the international standard for rating the sustainability of buildings. LEED contains six categories: sustainable sites (SS), water efficiency (WE), energy and atmosphere (EA), materials and resources (MR), indoor environmental quality (IEQ) and innovation in design (ID). Each LEED category contains several LEED credits (LEED, 2009). The correlated nature of the LEED categories are a result of dependencies in the calculation of LEED credits from different LEED categories.

* Tel.: +972 547 776 428.
E-mail addresses: sveta_pushkar@yahoo.com, svetlanap@ariel.ac.il

0169-2046/$ – see front matter
http://dx.doi.org/10.1016/j.landurbplan.2012.11.009

Table 1
Wilcoxon signed rank multiple tests among six correlated LEED categories. Probability (P), sample size (n = 10, the US EPA regions are defined as the primary sampling unit).

	Water efficiency (WE)	Innovation and design (ID)	Materials and resources (MR)	Energy and atmosphere (EA)	Sustainable sites (SS)	Indoor environmental quality (IEQ)
Median	258	299	441	506	554	699
$IQR_{[75-25]}$	179	256	374	358	453	563
WE	x					
ID	0.0020	x				
MR	0.0020	0.0020	x			
EA	0.0020	0.0020	0.0059	x		
SS	0.0020	0.0020	0.0039	0.4922	x	
IEQ	0.0020	0.0020	0.0020	0.0039	0.0020	x

For example, the credit of Construction Waste Management (MR category) in a rubric "Related credits" declares: "if an existing building is found to contain contaminated substances, such as lead or asbestos . . . see the following credit: Brownfield Redevelopment" (SS category), which intends to rehabilitate damaged sites (LEED, 2009, p. 358).

Recently, Picquelle and Mier (2011) recommended the method of two-stage or higher levels of hierarchical sampling for studying phenomena that may naturally occur in groups. In the case where both the role of geography in green buildings and spatially specific LEED categories in U.S. are to be evaluated, a three-stage sampling method should be used. Ignoring the sampling structure in such an analysis results in "sacrificial pseudoreplication" (Hurlbert, 1984, 2009). According to Hurlbert (2009), this type of pseudoreplication is committed when multiple correlated observations (sub-units) within an observation unit are supposedly analyzed as independent observation units. This error results in artificially inflated degrees of freedom, giving the illusion of having a more powerful statistical conclusion (Hurlbert, 2009; Picquelle & Mier, 2011).

Picquelle and Mier (2011, p. 2) defined a 'primary sampling unit' as "an element within a 'sampling frame' that is sampled and is statistically independent of other sampling units within the frame. The 'sampling frame' is defined as the collection of all elements (primary sampling units) accessible for sampling in the population of interest." Sub-units from the primary sampling unit in two-stage or higher levels of hierarchical sampling are defined as 'correlated observations.'

In reconceptualizing the Cidell and Beata (2009) study as a three-stage sampling design, the EPA regions are defined as the primary sampling unit (the first stage), the U.S. is defined as the sampling frame, the LEED categories are defined as sub-units (the second stage), and the LEED credits are defined as secondary sub-units (the third stage). This contrasts with cases 3 and 4 of the Cidell and Beata (2009) study, where the observations of the second and third stages were considered as a simple random sample. Ignoring the multilevel structure of these data in modeling procedures leads to the Pseudoreplication ANOVA (Picquelle & Mier, 2011) and an incorrect understanding of their scope in the LEED studies. This has significant implications in developing realistic and cost-effective strategies for sustainable landscape design.

3. Reanalysis of LEED study data

I conducted a reanalysis of the LEED data studied by Cidell and Beata (2009) using a neoFisherian paradigm as an alternative to Bayesian model selection for drawing statistical conclusions. Hurlbert and Lombardi (2009, p. 311) recommend using neoFisherian significant assessment (NFSA) to interpret the "sign and magnitude of statistical effects." They also state that investigators should always report exact P values in tabular form. In this case, the statistical analysis I used in this paper followed the "three-valued logic" recommended by Hurlbert and Lombardi (2009, p. 324): "in a two-group case, one examines the P value yielded by a significance test for a difference between groups and concludes one of three things: the difference between the true population means seems to be negative, it seems to be positive, or it cannot confidently be stated to be either so judgment is reserved or suspended... three-valued logic is thus an important element in the neo-Fisherian paradigm."

Based on the NFSA, Hurlbert and Lombardi (2012, p. 15) clearly show that no comparison-wise error rate (CWERs), set-wise (or family-wise) type I error rates (SWERs), or false discovery rates (FDRs) of any sort can be justified. They further argue that "neither the classical nor the newer multiple comparison procedures based on fixed maximum potential set-wise error rates are helpful to the cogent analysis and interpretation of scientific data" (Hurlbert & Lombardi, 2012, p. 1). Therefore, I chose a two-sided Wilcoxon signed rank test for conducting tests among the LEED categories and report the results independently (Table 1).

In the Cidell and Beata (2009) study (case 1), simple ANOVA tests cannot be used to evaluate P-values among 10 independent EPA regions due to the fact that the LEED categories are correlated observations (sub-units). Visual inspection of the scatterplot of observations (case 3) of both LEED categories and EPA regions shows that non-parametric tests should be used. Alternatively, I present the observations as sample size (n), medians (Mdn), interquartile ranges ($IQR_{[75-25]}$), and two-tailed P values.

In case 3, instead of using the one-way ANOVA test (Cidell & Beata, 2009), the Wilcoxon signed rank test should be used to evaluate the two-tailed P values among six correlated LEED categories. Ignoring the fact that the LEED categories/credits are interdependent leads to two major problems. The first is the statistical analysis is not testing the research hypothesis. The second is that interclass correlations between observations can lead to calculated P values that are either higher or lower than the true P value (Lazic, 2010). As a consequence of incorrect design of statistical analysis, there is a possibility to obtain incorrect results and conclusions. For example, in case 3 Cidell and Beata (2009) suggested, "(the) only statistically significant pairings were between WE and IEQ, and between IEQ and ID." In my reanalysis using the two-sided Wilcoxon signed rank test, the P value now suggests that the difference between all pairings of LEED categories seems to be positive, while the difference between EA and SS seems to be negative.

The reanalysis of the data in this paper reveals a monotonic increase in LEED points between all pairings when they are presented in the following sequence: WE, ID, MR, EA, SS, and IEQ. The IEQ category was the most preferable category for the LEED projects under consideration. In general, IEQ includes aspatial credits (for example, minimum indoor air quality performance, environmental tobacco smoke control, outdoor air delivery monitoring, etc.). Therefore, the IEQ category is preferable because it can be applied to every project. The credits for SS category (for example, site selection, brownfield redevelopment, etc.) as well as the credits for EA category (for example, green power) are spatial. As a result, these credits are the next preferable categories, because they can be implemented without significantly increasing the budget of projects constructed in urban areas.

4. Conclusions

The main conclusion of the analysis performed by Cidell and Beata (2009) highlights the importance of spatially sensitive certification categories. The present reanalysis revealed that the IEQ category seems to be the more preferable category than SS and EA (spatially sensitive categories) for the LEED projects under consideration due to aspatial applicability of this category. The disagreement between the conclusions follows from inaccurate usage of the appropriate statistical procedures in the study of Cidell and Beata (2009).

References

Cidell, J., & Beata, A. (2009). Spatial variation among green building certification categories: Does place matter? *Landscape and Urban Planning*, *91*, 142–151.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, *54*, 187–211.

Hurlbert, S. H. (2009). The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology*, *123*, 434–443.

Hurlbert, S. H., & Lombardi, C. M. (2009). Final collapse of the Neyman–Pearson decision theoretic framework and rise of the neoFisherian. *Annales Zoologici Fennici*, *46*, 311–349.

Hurlbert, S. H., & Lombardi, C. M. (2012). Lopsided reasoning on lopsided tests and multiple comparisons. *Australian and New Zealand Journal of Statistics*, *54*, 23–42.

Hurlbert, S. H., & White, M. D. (1993). Experiments with invertebrate zooplanktivores: Quality of statistical analyses. *Bulletin of Marine Sciences*, *53*, 128–153.

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BMC Neurosciences*, *11*, 5.

LEED. (2009). *LEED reference guide for green building design and construction*. Retrieved from https://new.usgbc.org/resources/leed-reference-guide-green-building-design-and-construction

Picquelle, S. J., & Mier, K. J. (2011). A practical guide to statistical methods for comparing means from two-stage sampling. *Fisheries Research*, *107*, 1–13.

United States Environmental Protection Agency. (2007). *Green buildings*. US EPA. Retrieved from http://www.epa.gov/greenbuilding/