

# **The Ecological Fallacy: How To Spot One And Tips On How To Use One To Your Advantage**

Hume F. Winzar, Griffith University

## **Abstract**

The ecological fallacy is a common and little understood error in the interpretation of statistical data wherein inferences about individuals are based on the aggregate of the group from which they belong. The paper overviews the importance of avoiding the error and illustrates the ease with which mistakes in inference can be made by examining a paper presented at the last ANZMAC conference, and by demonstrating with artificial data representing Hofstede's cultural dimensions. It concludes with an appeal for caution when considering the combination of aggregate data with our surveys of individuals.

## **Background**

*“The ecological fallacy is a widely recognised error in the interpretation of statistical data, whereby inferences about the nature of individuals are based solely upon aggregate statistics collected for the group to which those individuals belong. This fallacy assumes that all members of a group exhibit characteristics of the group at large”* (wikipedia, nd).

This paper is prompted by several papers presented at conferences over the last few years, including ANZMAC, where presenters accepted the ecological fallacy. The ecological fallacy also is something one sees often in published literature, especially in attempts to include Hofstede's cultural dimensions in a cross-national study. The paper seeks to explain what it is to become victim to the ecological fallacy, the consequences of making the error, and finally, by inference, some techniques for presenting otherwise useless results that will often snow the naïve reviewer or audience.

The term “Ecological Fallacy” was coined by Robinson in 1950 when he highlighted the correlation between % illiterate and % African-American in the US. Across the “Nine Regions”  $r = 0.95$ , across the 50 States  $r = 0.77$ , and at the Individual level  $r = 0.20$  (Robinson, 1950). How could it be that at the aggregate level the correlation was very high yet at the individual level the correlation was almost irrelevant? This is an important issue when government policy and grant money hinge on the decision. More recently we have seen statistics which show that those US states with the lowest levels of college graduates are most likely to vote Republican (Morford, 2004), yet the wealthy (and presumably well-educated) are regarded as the Republican base.

In 19<sup>th</sup> century Europe Durkheim (1897) pointed out that suicide rates were higher in countries that were more heavily Protestant. The logical inference for many readers was that the social conditions under Protestantism promoted suicide. This “Common Interpretation” of Weber's Protestant Work Ethic (Weber, 1958) argues that industrial capitalism is facilitated by Protestantism and occurred in predominantly Protestant countries earlier as a result. An examination of Durkheim's aggregate data seems to

support such a conclusion. More recently however, Delacroix & Nielson, (2001) show that measures of Industrial Capitalism have little relationship to Protestantism within countries, and even less at the regional level (when such measures are available). They suggest that the “common interpretation” is a result of anecdote and mere observation of 20<sup>th</sup> century prosperity.

In the 1970s economic commentators argued that Individualist economies, defined by Hofstede’s cultural dimensions, were economically superior. The logical inference was that Collectivism holds Asian countries back economically (Landes, 1998). Then in the 1990s similar commentators argued that Collectivism was responsible for growth of Tiger Economies (Zurndorfer, 2004). They couldn’t both be right.

So what are the origins of such ecological myths? Possibly authors have a political or social agenda. Perhaps it’s a rationalisation in which authors are simply reading the raw figures and making up a plausible story to fit the results. Regardless of the reasons for such errors of inference, we can see the effects of the ecological fallacy in which we identify statistical relationships at the aggregate level which do not accurately reflect the corresponding relationship at the individual data level. It’s an easy mistake to make. The following section briefly examines a recent paper, presented at the 2004 ANZMAC conference that demonstrates the ecological fallacy commonly presented in the business disciplines.

### ***The mistake of using Country as an indicator of individuals’ cultural values.***

A paper presented to this conference last year can serve as a convenient example of the ease and readiness with which we make the ecological fallacy. Malai and Speece (2004) made an extensive study that investigated, amongst other issues, the cultural impact on Perceived Customer Value. The relevant hypotheses were that:

*H3: National culture individualism has a significant positive effect on perceived brand name value, and*

*H4: National culture individualism has a significant positive effect on perceived service quality value.*

More than 600 international travellers from six countries were surveyed using an instrument presented in four languages on Customer Loyalty, Perceived Service Quality and Perceived Brand-Name Value. The culture construct, here Individualism, simply used Hofstede’s score for the countries of the survey respondents. In other words, instead of asking a series of questions designed to measure each respondent’s level on the Individualism-Collectivism scale, it was assumed that all subjects from the same country had exactly the same score. The subsequent analysis thus involved regression of national averages against respondent-level scores from a completely different data set. It is hardly surprising then that the best result in the study was about 7% explained variance ( $R^2$  would be even less if the more correct Bonferroni statistics had been calculated). This is a terrible waste – After about 40 questions had been asked of so many respondents, it would be a small marginal cost to add, say, five more questions to gain a short-form measure of Individualism/Collectivism and other cultural dimensions (Dorfman & Howell, 1988).

The intention here is not to single out just one paper for vilification: This is an example of papers one can see in conferences in all of the business disciplines each

year, and in well regarded journals (e.g. Volkema, 2004). One can sympathise with the temptation to simplify a study and take advantage of the contributions of others: *“This person is of this background. Experience, or a noted authority, says that on average such people have these attributes. Therefore I shall infer that this person has those same attributes.”* It’s convenient and saves time and generally helps us deal with the world. Unfortunately it’s similar logic that says that women are bad drivers, Muslims are terrorists, French are arrogant, and Americans are rich and stupid. Except for the ones we’ve actually met.

Occasionally there is little way around risking the ecological fallacy. The finest level measurement available from the Australian census data is at the level of “collection district” which is the aggregation of about 225 households. Census data from Australia’s 37,000 collection districts often is combined with industry data, or customers’ address data in order to gain a broader understanding of a customer base. The underlying assumptions are that people who reside in the same area live similar lives and that what is true for the collection district is also true for the individual. Neither assumption holds up completely, of course, but often it can be taken as a “rough approximation” of the characteristics of group members. And it can be more useful than not making the approximation at all. To the extent that such analysis works it is because each collection district is small, increasing the likelihood of homogeneity in the aggregated data observation. When the aggregation level is large and heterogeneous then we’re setting ourselves up for trouble.

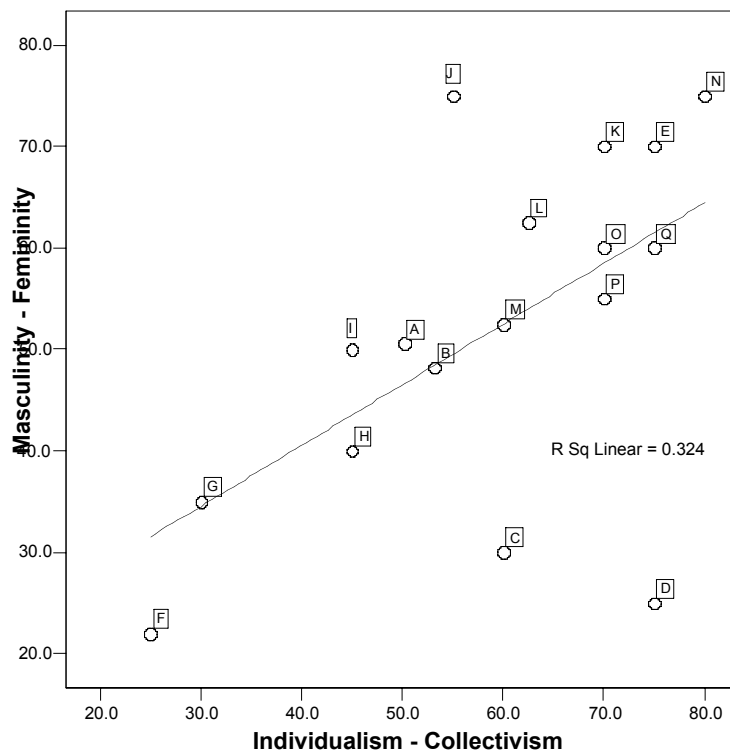
### **An example of fallacious ecological inference**

Figure\_1 illustrates average scores for seventeen different countries on two Hofstede Dimensions, Masculinity/Femininity and Individualism/Collectivism. It is clear that the mean scores are positively correlated:  $r=0.57$  ( $\alpha<0.05$ ). The logical, but incorrect, inference is that M/F and I/C are positively correlated across all members of this sample.

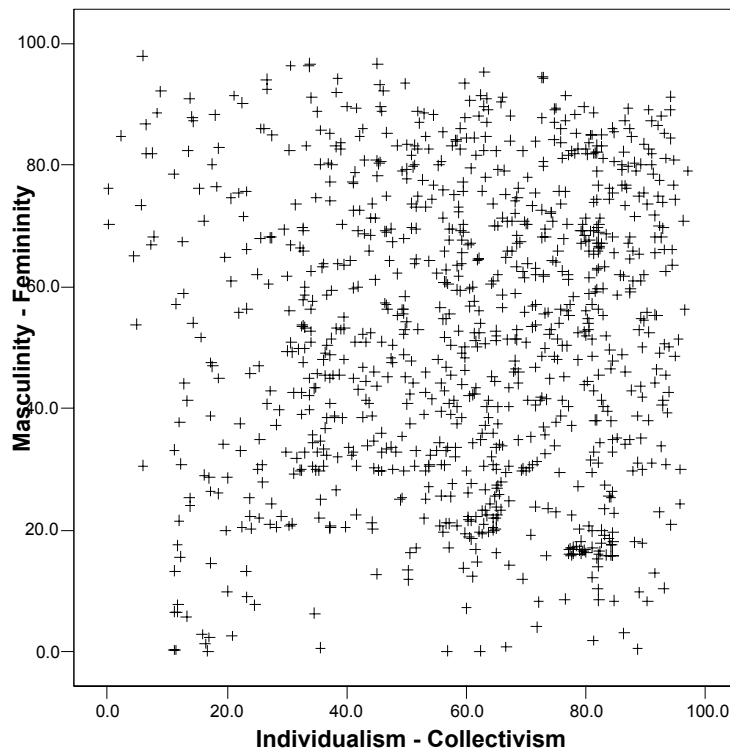
We might expect then that a person with a high I/C score would also have a high M/F score. Interestingly, a correlation of these two variables across all respondents in this data set shows a correlation coefficient of just  $r=0.05$  ( $\alpha<0.12$ ) as illustrated in Figure\_2.

This counterintuitive result becomes even more interesting when we discover that within each country the correlations are actually negative! Overall the correlation between M/F and I/C within countries on average is  $-0.332$ , as illustrated in Figure\_3. In contrast to the inference drawn from the mean country data illustrated in Figure\_1 we see in Figure\_3 that those respondents who have the higher I/C scores are likely to have the lower M/F scores within their own countries.

**Figure 1: Mean M/F and I/C scores for 17 Countries**



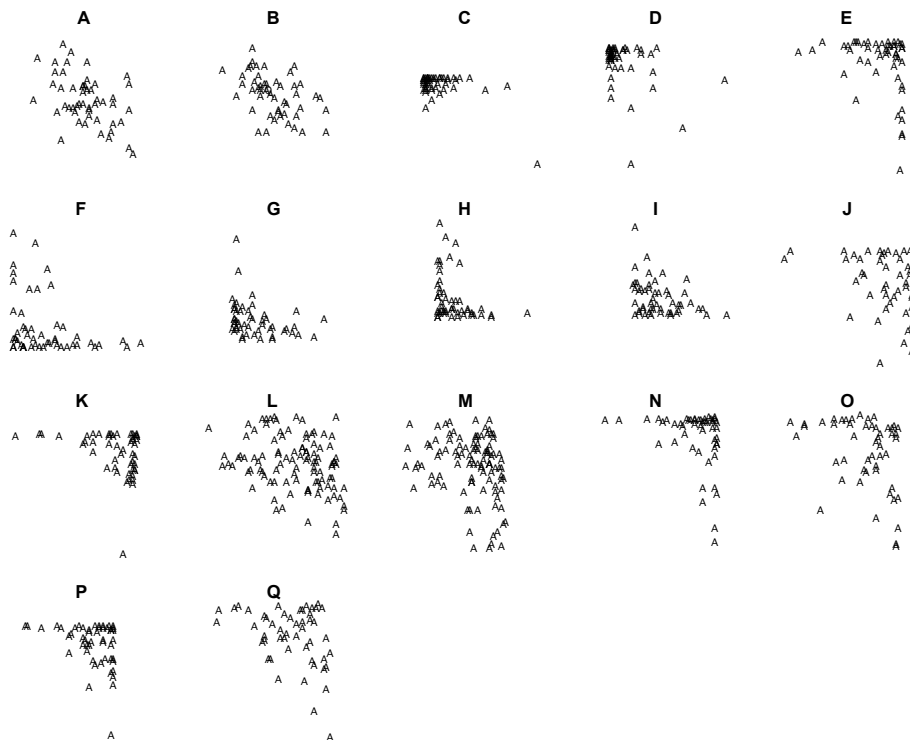
**Figure 2: All respondents from all countries**



These results are not in conflict. They illustrate the danger of attributing relationships found at one level of aggregation to another level of aggregation.

These data are artificial, created to illustrate a point, but they are not unrealistic. Hofstede himself has taken pains to warn clients and fellow scholars of the danger of the ecological fallacy. He points out, for example, that attributes with low correlations at the aggregate level are highly correlated at the individual level (e.g. Masculinity-Femininity Vs Individualism-Collectivism in Hong Kong), and attributes highly correlated at the aggregate level have low correlations at the individual level (e.g. Power Distance Vs Individualism-Collectivism in most countries) (Hofstede, 2001, 2004).

**Figure 3: Relationship between M/F & I/C within each country**



## Lessons to be learned

So what lessons can we learn from a proper consideration of the Ecological Fallacy? The issue goes beyond simply applying Hofstede's cultural values arbitrarily to any cross-national data set we might come across. As Fiske (2002) has pointed out, there are other, possibly more pressing, problems with that practice. He argues that much of the research on individualism and collectivism:

- ❑ treats nations as cultures and culture as a continuous quantitative variable;
- ❑ conflates all kinds of social relations and distinct types of autonomy;
- ❑ ignores contextual specificity in norms and values;
- ❑ measures culture as the personal preferences and behaviour reports of individuals;
- ❑ rarely establishes the external validity of the measures used;
- ❑ assumes cultural invariance in the meaning of self-reports and anchoring and interpretation of scales; and
- ❑ reduces culture to explicit, abstract verbal knowledge.

The ecological fallacy is contained in much of such research, and such criticisms can be equally levelled at other research where aggregate data has been applied to the individual level. This paper then is an appeal to researchers to consider carefully the data they are dealing with. Solutions to integrating data of different levels of aggregation are available. Hierarchical regression, which can be constructed as a special class of SEM or as a Mixed Model in SAS or SPSS, or with specialised software such as MLwiN, (Centre for Multilevel Modelling, 2004) HLM6 (SSI, 2005) or MixREG (Hedeker, 2005), permit regression models within aggregated data. A more sophisticated and accurate technique is demonstrated by Gary King (1997) but requires considerable technical expertise at present. All these approaches however require large data sets with many examples of aggregated data, not just two to six different countries. If you are dealing with survey data then it is very risky to add organisational, or country-level data to the analysis.

## References

- Centre for Multilevel Modelling (2004), Online : <http://multilevel.ioe.ac.uk/> (Accessed August 2005)
- Delacroix, J. & Nielson, F. (2001), The Beloved Myth: Protestantism and the rise of Industrial Capitalism in Nineteenth-Century Europe, *Social Forces*, 80,2,509-553.
- Durkheim E. (1897), *Le Suicide*. F. Alcan, Paris. English translation by J.A. Spalding, 1951, Free Press, Collier-MacMillan, Toronto, Canada
- Dorfman, P.W. and Howell, J.P. (1988), "Dimensions of Culture and Effective Leadership Patterns: Hofstede Revisited," in R.N. Farmer and E.G. McGain (eds.), *Advances in International Comparative Management*, Vol. 3, pp. 127-150.
- Fiske, Alan P. (2002), Using Individualism and Collectivism to Compare Cultures- A Critique of the Validity and Measurement of the Constructs: Comment on Oyserman et al., *Psychological Bulletin*, 128,1,78-88
- Hedeker, D. (2005) MixREG: mixed-effects linear regression, Online: <http://tigger.uic.edu/~hedeker/mix.html>
- Hofstede, G. (2001), *Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thousand Oaks CA: Sage Publications.
- Hofstede, G. (2004), Geert Hofstede. Online: <http://www.geert-hofstede.com/> (Accessed August 2005)
- King, G. (1997), *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.
- Landes, D.S. (1998), *The wealth and poverty of nations: Why some are so rich and some so poor*. Norton.
- Robinson, W.S. (1950), Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15,3,351-357
- SSI (Scientific Software International) (2005), Online: <http://www.ssicentral.com/hlm/> (Accessed August 2005)
- Malai, V. and Speece, M. (2004), Perceived Customer Value Relationships and Cultural Impacts on its. *Proceedings of ANZMAC 2004: Wellington*.
- Morford, M. (2004), Down With Fancy Book Learnin': What's it mean that the big cities and college towns of America all voted blue? *San Francisco Chronicle*, (November 12) Online: <http://www.sfgate.com/> (Accessed January 2005)

Volkema, R.J. (2004), Demographic, cultural, and economic predictors of perceived ethicality of negotiation behavior: A nine-country analysis, *Journal of Business Research*, 57, 69-78.

Weber, M. (1958), *The Protestant Work Ethic and the Spirit of Capitalism*, Talcott Parsons (Tr.) Scribner. (Original work published 1905)

Wikipedia: The free online encyclopedia. Online: <http://www.wikipedia.org/> (Accessed May 2005)

Zurndorfer, H.T. (2004), Confusing Confucianism With Capitalism: Culture As Impediment And/Or Stimulus To Chinese Economic Development, *Proceedings of the Third Global Economic History Network Meeting*, Konstanz, Germany, (3-5 June)