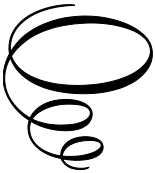
Assessing Organizational Diversity with the Simpson Index

Assessing Organizational Diversity with the Simpson Index

By

Salomón Alcocer Guajardo

Cambridge
Scholars
Publishing



Assessing Organizational Diversity with the Simpson Index

Series: Assessing Diversity in Nonprofit, Private, and Public Organizations

By Salomón Alcocer Guajardo

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-9264-2
ISBN (13): 978-1-5275-9264-3

This book is dedicated posthumously to Dr. Edward Hugh Simpson (1922 – 2019) for creating the diversity index ($S = 1 - \sum p^2$) that is the focus of this book and for his invaluable contribution to the study of diversity.

CONTENTS

EXHIBITS

NYC DEPARTMENTS

NYC Department	Acronym
Administration for Children's Services	ACS
Board of Corrections	BOC
Board of Election	BOE
Borough President-Bronx	BP-BX
Borough President-Brooklyn	BP-BK
Borough President-Manhattan	BP-MAN
Borough President-Queens	BP-QNS
Borough President-Staten Island	BP-SI
Business Integrity Commission	BIC
Campaign Finance Board	CFB
City Commission on Human Rights	CCHR
Civilian Complaint Review Board	CCRB
Conflicts of Interest Board	COIB
Department for the Aging	DFTA
Department of Buildings	DOB
Department of City Planning	DCP
Department of Citywide Administrative Services	DCAS
Department of Consumer Affairs	DCA
Department of Correction	DOC
Department of Cultural Affairs	DCLA
Department of Design and Construction	DDC
Department of Education	DOE
Department of Environment Protection	DEP
Department of Finance	DOF
Department of Health/Mental Hygiene	DOHMH

Department of Homeless Services	DHS
Department of Info Tech and Telecomm	DOITT
Department of Investigation	DOI
Department of Parks and Recreation	PARKS
Department of Probation	DOP
Department of Records and Information Service	DORIS
Department of Sanitation	DSNY
Department of Small Business Services	SBS
Department of Transportation	DOT
Department of Youth and Community Development	DYCD
District Attorney - Bronx County	DA-BX
District Attorney - Kings County	DA-BK
District Attorney - Manhattan	DA-MAN
District Attorney - Queens County	DA-QNS
District Attorney - Richmond County	DA-SI
District Attorney – Special Narcotics	DA-NARC
Equal Employment Practices Commission	EEPC
Financial Information Services Agency	FISA
Fire Department	FDNY
Housing Preservation and Development	HPD
Human Resources Administration / Social Services	HRA
Independent Budget Office	IBO
Landmarks Preservation Committee	LPC
Law Department	LAW
MAYORALTY	MAYORALTY
Municipal Water Finance Authority	MWFA
New York City Council	COUNCIL
New York City Fire Pension Fund	FDNYPF
New York City Police Pension Fund	NYCPPF
New York City Tax Commission	NYCTAX

NYC Civil Service Commission	NYCCSC
NYC Employees Retirement System	NYCERS
NYC Health + Hospitals	NYCHH
NYC Housing Authority	NYCHA
Office of Administrative Trials and Hearings	OATH
Office of Collective Bargaining	OCB
Office of Emergency Management	NYCEM (OEM)
Office of Payroll Administration	OPA
Office of the Actuary	ACTUARY
Office of the City Clerk	CLERK
Office of the Comptroller	COMPTROLLER
Office of the Public Advocate (PA)	PA
Offices of the Public Administrators	PUBADMIN
Police Department	NYPD
School Construction Authority	SCA
Taxi and Limousine Commission	TLC
Teachers Retirement System	TRS

ACKNOWLEDGEMENTS

I wish to thank the Sophie Edminson, Amanda Millar, Adam Rummens, and the production staff at Cambridge Scholars Publishing for their assistance with making this book a reality. I also wish to thank Steve Santiago for his immense and unwavering encouragement and support. This book would not be possible without them.

PREFACE

The use of diversity indices to study demographic, occupational, and social heterogeneity in nonprofit, private, and public organizations has increased enormously over the past 20 years or so. Many diversity-centered studies utilize indices originally designed to measure biodiversity to assess age, ethnic, gender, and other types of heterogeneity in organizational and workforce settings. Concomitant with the adaption of ecological-based diversity indices such as Simpson's diversity index, the use of diversity indices developed for communications and other fields such as Shannon's H index has become more prevalent. While the adaptation of diversity indices designed to measure heterogeneity in communications, ecology, and other fields to organizational settings helps to assess and understand demographic or social diversity, the application of diversity indices to measure and analyze heterogeneity in organization settings has been given little empirical attention.

Because little empirical research has been devoted to examining the adaptation and application of diversity indices to measure and analyze demographic or social diversity in organizations, this book and its companion books address fundamental analytical and measurement issues and questions that arise when diversity indices are applied to demographic and employment data to obtain measures of heterogeneity. The issues and questions addressed in this book series include the following:

- How is measurement bias addressed by a particular diversity index?
- How is the number of categories used for a demographic or social characteristic taken into account by a particular diversity index?
- What are the statistical properties of a distribution of scores of a particular diversity index when it is applied to demographic and employment data?
- What is the appropriate statistical method that should be used based on the distribution of scores obtained by a particular diversity index?
- What is the maximum value of diversity that is obtainable by a particular diversity index?

Although the issues and questions addressed in this book series are fundamental to carrying out empirical research, they are often ignored or taken for granted by practitioners and researchers alike.

This book series consists of 9 books. They are the following:

- *Assessing Organizational Diversity with the Simpson Index* applies the Simpson diversity index to demographic and employment data reported by New York City (NYC) departments in fiscal year 2019. This book focuses on the application and analysis of Simpson diversity formulas for calculating biased and unbiased measures of demographic heterogeneity.
- *Assessing Organizational Diversity with the Shannon Index* applies the Shannon diversity index to the same demographic and employment data used in the first book. This book focuses exclusively on the application and analysis of Shannon diversity formulas for calculating biased and unbiased measures of demographic heterogeneity.
- *Assessing Organizational Diversity with the Heip Index* applies the Hill, Heip, Hurlbert, and Sheldon diversity indices of evenness to the data used in the first and second books. Practically speaking, the Hill, Heip, Hurlbert, and Sheldon indices are modifications of the Shannon index of diversity. From a statistical standpoint, the Shannon-based indices possess statistical properties that are superior to the Shannon index. Similar to the first and second books, this book focuses on the application and analysis of the indices in regard to measuring of demographic heterogeneity in organizations.
- *Assessing Organizational Diversity with the Smith and Wilson Indices* applies the Smith and Wilson (SW) indices to the same data used in the previous books. This book also uses other Simpson-based diversity indices to complement the SW indices such as the Ray and Singer index of concentration. The SW indices are modifications of the Simpson ($D = 1- \sum p^2$) diversity index and assess diversity. Similar to the previous books, this book applies the Simpson-based indices to the same demographic employment data to measure demographic heterogeneity in organizations.
- *Assessing Organizational Diversity with the McIntosh Index* applies the McIntosh evenness index to the same demographic and employment data used in the previous books. This book focuses exclusively on the analysis of diversity scores obtained by the McIntosh index. Because this index takes into account the number of groups used to categorize a demographic or social characteristic

of interest and the size of the workforce simultaneously, the diversity scores contain less measurement bias and have a greater degree of compatibility in comparison to the other diversity index covered in other books.

- *Assessing Organizational Diversity with the Index of Qualitative Variation* (IQV) applies the Mueller and Schuessler IQV to the same demographic and employment data used in books 1 - 5. Because the IQV is not invariant to ordering sequences, this book focuses on the application and analysis of heterogeneity scores obtained from the different ordering sequences of the data. Similar to the McIntosh evenness index presented in the 5th book, the IQV takes into account jointly the size of the workforce and the number of groups used in the categorization of the demographic or social characteristic of interest.
- *Assessing the Validity of Diversity Indices* compares all of the indices used in each book jointly and uses factor analysis to determine whether they assess the same or different aspects of demographic or social diversity. Pearson pairwise correlation analyses also are performed to assess the statistical associations amongst the diversity indices. Statistical analyses for equality of means are performed as well.
- *Assessing Organizational Diversity with Quantile Regression* applies quantile regression analysis to several of the diversity indices presented in the book series. This book performs quantile regression analyses at the 25th, 50th, 75th, and 90th percentiles for age, ethnic, and gender diversity.
- *Assessing Organizational Diversity with Structural Equation Modeling* (SEM) focuses exclusively on causal modeling. This book focuses on the development and analysis of a structural equation model for several diversity index discussed in the series. In so doing, the analyses treat age, ethic, and gender diversity as an intervening or mediation variables of organizational performance.

For purposes of continuity and compatibility, each diversity index is subjected to the same statistical analyses. Special interest is placed on the IQV, McIntosh evenness, Shannon, Simpson, and SW indices in this book series because they have been used in previous research on demographic or social diversity in nonprofit, private, or public organizations.

This book series is written for practitioners and researchers in human resources and other fields that are interested in measuring and analyzing

demographic, occupational, or social heterogeneity in organizations. The purpose of the book series is to addresses measurement and analytical issues that practitioners and researchers alike are likely to face when they apply a particular diversity index to demographic and employment data provided by a nonprofit, private, or public organization. As such, this book series should serve as a reference for selecting the diversity index that is best suited for measuring and analyzing heterogeneity in an organizational setting. This book series also should serve as a reference for selecting the statistical method that is best suited for analyzing the distribution of scores obtained by the diversity index of choice.

CHAPTER 1

INTRODUCTION

Since the early 1970s, diversity (or *integration*) indices have been used to assess the level of demographic or social heterogeneity in nonprofit, private, or public organizations (e.g., Akram, Abrar ul Haq, Natarajan, and Chellakan, 2020; Boehm, Kunze, and Bruch, 2014; Choi, 2010; Gazley, Chang, and Bingham, 2010; Grabosky and Rosenbloom, 1975; Guajardo, 2014; Moon and Christensen, 2020; Nachmias and Rosenbloom, 1973). The application of diversity indices to aggregate demographic employment data has centered on measuring age, ethnic, or gender heterogeneity. In the majority of the previous studies, workforce diversity has served as a dependent variable. More recent studies have treated workforce diversity as an independent variable which influences organizational performance (e.g., Gazley, Chang, and Bingham, 2010; Khan, Khan, and Senturk, 2019; Lee-Kuen, Sok-Gee, and Zainudin, 2017; Pitts, 2005). This book takes the position that workforce diversity such as age, ethnic, and gender heterogeneity is an *intervening variable* that influences organizational performance (e.g., Guajardo, 2014; Pitts, 2006).

Indices of diversity and variation

Several indices of diversity (or variation) have been used to measure demographic or social diversity (or heterogeneity) in organizations. They include the following:

- Index of qualitative variation (IQV) or measure of variation (MV; e.g., Grabosky and Rosenbloom, 1975; Kellough, 1990; Kim, 1993; Mueller, Schuessler, and Costner, 1970; Nachmias and Rosenbloom, 1973; Wilcox, 1967);
- McIntosh evenness index (e.g., Guajardo, 2013 and 2015);
- Shannon (Teachman) index (e.g., Choi, 2010; Choi and Rainey, 2010); and,
- Simpson (Blau or Lieberson) diversity index (e.g., Guajardo, 2014; Pitts, 2005; Starks, 2009).

Each index has a theoretical distribution of scores that ranges from 0 to 1. With the exception of the Shannon index, a score of 0 indicates the absence of diversity (or heterogeneity) and a score of 1 indicates absolute diversity. This book focuses solely on the Simpson diversity index.

Briefly, McIntosh (1967) created the diversity index of evenness to assess the level of similarity amongst a group of species with different characteristics living in the same community. The McIntosh evenness index is represented as follows: $D_E = \frac{\mathbb{N} - \sqrt{\sum n^2}}{N - \frac{N}{\sqrt{s}}}$, where N represents the total number of individuals, n represents the number of individuals within a particular group, and S represents the number of distinct groups in a community or organization.

D_E has a distribution of scores that ranges from 0 to 1. According to McIntosh (1967), a score of 0 indicates an absence of diversity and a score of 1 indicates absolute diversity. *Assessing Organizational Diversity with the McIntosh Index* focuses exclusively on the application of the index to public organizations.

Mueller and Schuessler developed the IQV to assess demographic heterogeneity (or variation) in a community or social setting (Mueller, Schuessler, and Costner, 1970; Wilcox, 1967). The IQV obtains a measurement of heterogeneity by dividing the total observed difference by the maximum possible differences (Mueller, Schuessler, and Costner, 1970; Wilcox, 1967). Symbolically, the IQV index takes the following form:

$$\text{IQV} = \frac{\text{Total observed differences}}{\text{Maximum Possible Differences}} = \frac{\sum f_j f_j}{\frac{n(n-1)}{2} \text{ x } (\frac{F}{n})^2}$$

The terms of the formula are defined as follows:

f represents the frequency (or number) of individuals;
n represents the number of social characteristics (i.e., groups); and,
F represents the total number of individuals.

The index has a distribution of scores ranging from 0 to 1. Similar to the Simpson and McIntosh indices, a score of 0 indicates a lack of heterogeneity (i.e., homogeneity) and a score of 1 indicates absolute heterogeneity (Mueller, Schuessler, and Costner, 1970). *Assessing Organizational Diversity with the Index of Qualitative Variation* focuses exclusively on the application of the index to public organizations.

Shannon (1948) created the H index of entropy (or uncertainty) for the communications field to obtain the *probability* of successive messages being independent of each other. Since its creation, the H index has been adapted to assess demographic or social diversity in organizations. The index is presented as follows: $H = -\sum[p_k * \ln(p_k)]$, where p is the proportion of individuals in the kth category. Unlike the IQV, McIntosh, and Simpson indices, the Shannon index has a distribution of scores ranging from 0 to $\ln(n)$, where *ln* represents the natural logarithm of a number and n represents the number of demographic or social categories (e.g., Harrison and Klein, 2007). For instance, the maximum value of H is 1.609 ($H_M = \ln(5) = 1.609$) when ethnicity is categorized into 5 groups. The following formula is used to standardize H index scores: $H_s = \frac{-\sum[pk*\ln(pk)}{\ln(n)}$. When the standardized form of the H index is used, the distribution of the standardized scores ranges from 0 to 1. The Shannon index is the focus of *Assessing Organizational Diversity with the Shannon Index*.

The Simpson diversity index is used the most frequently to assess demographic or social heterogeneity in organizations. Simpson (1949) created the diversity index to obtain the probability that two individuals chosen at random from the same community would share the same (or different) characteristics. Theoretically, Simpson's index has a distribution of scores ranging from 0 to 1. According to Simpson (1949), a score of 0 indicates the absence of diversity and a score of 1 indicates absolute diversity. In actuality, the maximum value for a particular demographic or social characteristic is determined by the following formula: $S_M = \frac{k-1}{k}$, where k is the number of categories or groups formed for the demographic or social characteristic of interest. For instance, the maximum diversity score is 0.80 ($S_M = \frac{k-1}{k} = \frac{4-1}{5} = 0.80$) when ethnicity is categorized into 5 groups. Simpson's index is represented frequently by the following formula: $S_A = 1-\sum p^2$, where p represents the percent of individuals in a particular category or group. The index also is represented as follows: $S_B = 1 - \frac{\sum n(n-1)}{N(N-1)}$, where n is the number of individuals in a category or group and N is the total number of individuals. Standardized Simpson scores are obtained by applying the following formula: $S_S = \frac{s}{\frac{(k-1)}{k}}$, where S equals S_A or S_B and k represents the number of categories or groups.

Statistical methods

Numerous statistical methods have been used to analyze the relationships amongst a set of organizational factors and Simpson, Shannon, or other scores. The statistical methods used in previous *cross-sectional studies* on demographic or social diversity in organizations include the following:

- Ordinary least squares (OLS) regression (e.g., Choi, 2010; Gazley, Chang, and Bingham, 2010; Kellough, 1990; Kim, 1993; Pitts, 2005);
- Tobit regression (e.g., Guajardo, 2016; Poulos and Doerner, 1996); and,
- Quantile regression (e.g., Guajardo, 2016).

These and other statistical methods are used in this book to illustrate how the distribution of Simpson diversity scores impact the findings when statistical assumptions are violated.

This book takes the position that the selection and use of a particular statistical method should be based on whether the distribution of diversity scores obtained by a particular index satisfies the underlying assumptions of the statistical method that is selected. At times, the selection of a particular method is based on academic tradition. For instance, the majority of the studies that analyze demographic diversity in public organizations use OLS regression to perform the multivariate analyses although the diversity indices presented above obtain scores with restricted quantitative continuous distributions that range from 0 to at least $\ln(n)$. By contrast, Tobit and quantile regression methods are used primarily in economics because these methods are better suited for variables with restricted distributions. Methodologically and statistically, Tobit, quantile, and other similar regression methods are better suited to analyze a distribution of diversity scores with a restricted quantitative continuous range of values. The selection and use of a particular statistical method to analyze the relationship amongst a set of organizational factors and demographic (or social) diversity is important because financial, human resources, and policy decisions are often based on the findings of a study.

Analyzing and measuring diversity in organizations

Despite the voluminous number of articles and books that have addressed organizational or workforce diversity, few studies have focused on the

application of diversity indices that are used to measure and analyze demographic or social heterogeneity in organizations (e.g., Biemann and Kearney, 2010; Guajardo, 2013 and 2015; Harrison and Klein, 2007; McDonald and Dimmick, 2003). Fewer studies have focused on assessing the reliability and validity of the heterogeneity scores obtained from applying diversity indices to employment data reported by nonprofit, private, or public organizations (e.g., Guajardo, 1996). As a consequence of the lack of extensive research on the adaptation and utilization of diversity indices designed to measure heterogeneity in communications (e.g., Shannon H index) or ecology (e.g., Simpson diversity index), the properties of the diversity scores obtained by a particular index in regard to normality, skewness, and other distribution characteristics are often taken for granted or glossed over at best.

Because there is a paucity of research on the application of diversity indices to employment data, this book addresses methodological and statistical issues and questions that should be addressed when assessing demographic or social diversity in the workplace. The issues and questions include the following:

- Does the use of a particular diversity index produce a distribution of scores that satisfy the underlying assumptions of the statistical method of choice?
- How does the categorization of the demographic or social characteristics of interest affect the reliability and validity of the diversity scores obtained by the chosen index?
- What are the distribution properties of the diversity scores in regard to the range of scores, to the skewness of the distribution, to the peakedness of the distribution, and to other related issues?
- What is the level of compatibility of the diversity scores amongst the organizations in the study in regard to measuring heterogeneity as accurately as possible when there are categories with missing data?

Addressing these and other measurement issues and questions prior to undertaking a statistical analysis of demographic or social diversity in organizations is critical for obtaining statistical findings that are unbiased, reliable, and stable.

Application of the Simpson Index

This book applies the Simpson index to employment and workforce data reported by NYC departments for fiscal year 2019. NYC departments are selected for this book series to highlight the measurements issues that are likely to be confronted in other employment settings regardless of whether the organizations are nonprofit, private, or public institutions. The Simpson diversity index is presented in Chapter 2. The traditional Simpson index formula ($S = 1 - \sum p^2$) is applied to qualitatively categorized demographic employment data to illustrate how diversity scores are calculated for data presented as frequencies or percentages. In addition, the formula for obtaining unbiased Simpson diversity scores is applied to the same data.

In Chapter 3, the index is applied to demographic data to obtain diversity scores for gender. The application of the Simpson index entails the use of the traditional formula ($S = 1 - \sum p^2$) and the use of the formula for calculating unbiased diversity scores. Standardized scores are obtained for each set of diversity scores as well. In addition to assessing the range of the distribution of the scores and to testing for normality and skewness, the distributions of scores are compared statistically for equality.

Chapter 4 applies the Simpson index to demographic data to obtain diversity scores for ethnicity. Because NYC departments categorize ethnicity in two different ways, four sets of unstandardized Simpson diversity scores are obtained. Biased and unbiased diversity scores are calculated for the binary categorization of ethnicity: White or minority. The second set of biased and unbiased diversity scores are calculated for the 5-group categorization of ethnicity: Asian, Black, Hispanic, Some other race (SOR), and White. Standardized coefficients are calculated for each set of unstandardized scores. The two different ways of categorizing ethnicity allow for assessing changes in the distributions of the scores. They also permit for testing the sets of diversity scores for equality.

In Chapter 5, diversity scores for age are obtained by applying the Simpson index to the demographic data. The issues of measurement compatibility and reliability arise due to calculating diversity scores for NYC departments with workforces where the range of ages is confined to a small number of age group categories. The issue of measurement compatibility and reliability is compounded when the formula for obtaining unbiased Simpson diversity scores is applied to the data. Standardized scores are calculated for each set of diversity coefficients. Statistical analyses for equality are performed in addition to testing for normality and skewness. Because the