

# Assessing Organizational Diversity with the Shannon Index

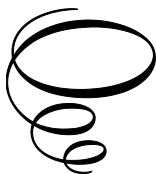


# Assessing Organizational Diversity with the Shannon Index

By

Salomón Alcocer Guajardo

**Cambridge  
Scholars  
Publishing**



Assessing Organizational Diversity with the Shannon 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

Copyright © 2023 by Salomón Alcocer Guajardo

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-9434-3

ISBN (13): 978-1-5275-9434-0

This book is dedicated posthumously to Dr. Claude Elwood Shannon  
(1916 – 2001) for his immeasurable contribution to the study  
of diversity by developing the index of entropy,  $-\sum p_i \ln p_i$ .



# CONTENTS

Exhibits.....	ix
New York City (NYC) Departments .....	xv
Acknowledgements .....	xviii
Preface .....	xix
Chapter 1 .....	1
Introduction	
Chapter 2 .....	12
Shannon Diversity Index	
Chapter 3 .....	20
Measurement of Gender Diversity	
Chapter 4 .....	42
Measurement of Ethnic Diversity	
Chapter 5 .....	77
Measurement of Age Diversity	
Chapter 6 .....	102
Composite Index of Organizational Diversity	
Chapter 7 .....	125
Analysis of Diversity with Ordinary Least Squares (OLS) Regression	
Chapter 8 .....	162
Analysis of Diversity with Robust Regression	
Chapter 9 .....	186
Analysis of Diversity with Tobit Regression	

Chapter 10 .....	207
Analysis of Diversity with Quantile Regression	
Chapter 11 .....	225
Analysis of Diversity with Ridge Regression	
Chapter 12 .....	246
Comparison of Statistical Methods	
Chapter 13 .....	260
Path Analysis of Public Sector Diversity	
Chapter 14 .....	293
Summary	
References .....	298
Glossary .....	307
Index .....	313



## EXHIBITS

Exhibit 2-1	Interpretation of Shannon (H) index scores	14
Exhibit 2-2	Interpretation of standardized H index scores	15
Exhibit 2-3	Calculation of the Shannon (H) gender diversity score for ACS for fiscal year 2024	16
Exhibit 2-4	Calculation of the unbiased Shannon ( $H_U$ ) gender diversity score for ACS for fiscal year 2019	16
Exhibit 3-1	Biased H ( $H_{GB}$ ) gender diversity scores for NYC departments for fiscal year 2019	21
Exhibit 3-2	Descriptive statistics for biased H ( $H_{GB}$ ) gender diversity scores for NYC departments for fiscal year 2019	25
Exhibit 3-3	Unbiased H ( $H_{GU}$ ) gender diversity scores for NYC departments for fiscal year 2019	28
Exhibit 3-4	Descriptive statistics for the unbiased H ( $H_{GU}$ ) gender diversity scores for NYC departments for fiscal year 2019	32
Exhibit 3-5	Comparison of biased ( $H_{GB}$ ) and unbiased ( $H_{GU}$ ) H gender diversity scores for NYC departments for fiscal year 2019	35
Exhibit 3-6	Paired z-test for equal means for biased ( $H_{GB}$ ) and unbiased ( $H_{GU}$ ) unstandardized H gender diversity scores for NYC departments for fiscal year 2019	38
Exhibit 4-1	Calculation of the biased ( $H_{EB}$ ) and unbiased ( $H_{EU}$ ) H ethnic diversity score for ACS for fiscal year 2019	44

Exhibit 4-2	Biased H ( $H_{EB1}$ ) ethnic diversity scores for NYC departments for fiscal year 2019	46
Exhibit 4-3	Descriptive statistics for biased H ( $H_{EB}$ ) ethnic diversity scores for NYC departments for fiscal year 2019	53
Exhibit 4-4	Unbiased H ( $H_{EU}$ ) ethnic diversity scores for NYC departments for fiscal year 2019	56
Exhibit 4-5	Descriptive statistics for unbiased H ( $H_{EU}$ ) ethnic diversity scores for NYC departments for fiscal year 2019	63
Exhibit 4-6	Comparison of biased ( $H_{EB}$ ) and unbiased ( $H_{EU}$ ) H ethnicity diversity scores for NYC departments for fiscal year 2019	66
Exhibit 4-7	Paired z-test for equal means for biased ( $H_{EB}$ ) and unbiased ( $H_{EU}$ ) H ethnic diversity scores for NYC departments for fiscal year 2019	73
Exhibit 4-8	Pearson correlation analysis of biased ( $H_{EB}$ ) and unbiased ( $H_{EU}$ ) H ethnic diversity scores for NYC departments for fiscal year 2019	74
Exhibit 4-9	Factor analysis (FA) of biased ( $H_{EB}$ ) and unbiased ( $H_{EU}$ ) H ethnic diversity scores for NYC departments for fiscal year 2019	75
Exhibit 5-1	Calculation of biased ( $H_{AB}$ ) and unbiased ( $H_{AU}$ ) H age diversity scores for ACS and EEPC for fiscal year 2019	79
Exhibit 5-2	Age distribution for each NYC department for fiscal year 2019	81
Exhibit 5-3	Biased ( $H_{AB}$ ) and unbiased ( $H_{AU}$ ) H age diversity scores for NYC departments for fiscal year 2019	87

Exhibit 5-4	Descriptive statistics for biased $H$ ( $H_{AB}$ ) age diversity scores for NYC departments for fiscal year 2019	92
Exhibit 5-5	Descriptive statistics for unbiased $H$ ( $H_{AU}$ ) age diversity scores for NYC departments for fiscal year 2019	94
Exhibit 5-6	Comparison of biased ( $H_{AB}$ ) and unbiased ( $H_{AU}$ ) $H$ age diversity scores for NYC departments for fiscal year 2019	97
Exhibit 5-7	Paired z-test for equal means for biased ( $H_{AB}$ ) and unbiased ( $H_{AU}$ ) $H$ age diversity scores for NYC departments for fiscal year 2019	99
Exhibit 6-1	Calculation of composite organizational diversity (OD) scores for ACS for fiscal year 2019	104
Exhibit 6-2	OD scores based on unstandardized biased $H$ ( $H_B$ ) diversity coefficients for NYC departments for fiscal year 2019	107
Exhibit 6-3	OD scores based on unstandardized unbiased $H$ ( $H_U$ ) diversity coefficients for NYC departments for fiscal year 2019	111
Exhibit 6-4	Descriptive statistics for biased and unbiased OD scores for NYC departments for fiscal year 2019	115
Exhibit 6-5	Comparison of OD scores for NYC departments for fiscal year 2019	118
Exhibit 6-6	Paired z-test for equal means for biased and unbiased unstandardized OD scores for NYC departments for fiscal year 2019	121
Exhibit 6-7	Pairwise Pearson correlation analysis of biased and unbiased OD scores for NYC departments for fiscal year 2019	121

Exhibit 6-8	Factor analysis (FA) of biased and unbiased OD scores for NYC departments for fiscal year 2019	122
Exhibit 7-1	Assumption of normality for OLS regression	128
Exhibit 7-2	Distribution of biased H ( $H_{GB}$ ) gender diversity scores by total employees	130
Exhibit 7-3	Scatterplot of biased H ( $H_{GB}$ ) gender diversity scores by total employees	133
Exhibit 7-4	Pearson correlation matrix for independent variables	135
Exhibit 7-5	Theoretical model for age, ethnic, and gender diversity for NYC departments	144
Exhibit 7-6	OLS regression analysis for age, ethnic, gender and organizational diversity for NYC departments for fiscal year 2019	152
Exhibit 8-1	Robust regression process	163
Exhibit 8-2	Robust regression analysis for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	169
Exhibit 8-3	Comparison of OLS and robust regression findings for biased ( $H_B$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	176
Exhibit 8-4	Comparison of OLS and robust regression findings for unbiased ( $H_U$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	180
Exhibit 9-1	Tobit regression analysis for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	193

Exhibit 9-2	Comparison of OLS and Tobit regression findings for biased ( $H_B$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	199
Exhibit 9-3	Comparison of OLS and Tobit regression findings for unbiased ( $H_U$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	203
Exhibit 10-1	Quantile regression analysis for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	212
Exhibit 10-2	Comparison of OLS and quantile regression findings for biased ( $H_B$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	216
Exhibit 10-3	Comparison of OLS and quantile regression findings for unbiased ( $H_U$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	219
Exhibit 11-1	Ridge regression process	227
Exhibit 11-2	Ridge regression analysis for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	232
Exhibit 11-3	Comparison of OLS and ridge regression findings for biased ( $H_B$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	237
Exhibit 11-4	Comparison of OLS and ridge regression findings for unbiased ( $H_U$ ) scores for age, ethnic, gender, and organizational diversity for NYC departments for fiscal year 2019	241
Exhibit 12-1	Comparative analysis of findings for age diversity for NYC departments for fiscal year 2019	248

Exhibit 12-2	Comparative analysis of findings for ethnicity diversity for NYC departments for fiscal year 2019	252
Exhibit 12-3	Comparative analysis of findings for gender diversity for NYC departments for fiscal year 2019	255
Exhibit 13-1	Talent acquisition and organizational diversity	263
Exhibit 13-2	Theoretical causal model for age, ethnic, and gender diversity for NYC departments	265
Exhibit 13-3	Pearson correlation matrix for independent variables and diversity scores	271
Exhibit 13-4	OLS regression findings for age diversity for NYC departments for fiscal year 2019	272
Exhibit 13-5	Robust regression findings for age diversity for NYC departments for fiscal year 2019	275
Exhibit 13-6	Hypothesized causal model for age diversity for NYC departments	277
Exhibit 13-7	OLS regression analysis for age diversity with mediating variables for NYC departments for fiscal year 2019	283
Exhibit 13-8	OLS regression causal model for age diversity for NYC departments	285
Exhibit 13-9	Robust regression analysis for age diversity with mediating variables for NYC departments for fiscal year 2019	286
Exhibit 13-10	Robust regression causal model for age diversity for NYC departments	288

# 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 & 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 & Telecomm	DOITT
Department of Investigation	DOI
Department of Parks & Recreation	PARKS
Department of Probation	DOP
Department of Records & Information Service	DORIS
Department of Sanitation	DSNY
Department of Small Business Services	SBS
Department of Transportation	DOT
Department of Youth & 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 & 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 & 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 & Limousine Commission	TLC
Teachers Retirement System	TRS

## ACKNOWLEDGEMENTS

I wish to thank Sophie Edminson, Amanda Millar, Mhairi Nicol, 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 continued immense and unwavering encouragement and support. This book would not be possible without them.

## PREFACE

During the past 20 years or so, the use of diversity indices to study demographic, occupational, and social heterogeneity in nonprofit, private, and public organizations has increased enormously. Many diversity-centered and -focused studies adapt biodiversity and ecological indices to assess and measure age, ethnic, gender, and other types of heterogeneity in organizational and workforce settings. Concomitant with the adaptation of ecological-based diversity indices such as Simpson's ( $D = 1 / \sum p^2$ ) diversity index, the use of diversity indices developed for communications and other fields such as Shannon's  $H (-\sum p \ln p)$  index has become more prevalent. Although 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, little empirical attention has been given to the statistical properties of the diversity indices used to measure and analyze heterogeneity in organizations.

Because little empirical research has been devoted to examining the adaptation and use of diversity indices to measure and analyze demographic (or social) diversity in organizations, this book series addresses 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 addressed 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 to use 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 for 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, Hurlburt, Heip, and Sheldon diversity indices to the data used in the first and second books. The Hill, Heip, Hurlbert, and Sheldon diversity indices are modifications of the Shannon index of diversity. From a statistical standpoint, these diversity indices possess statistical properties that are superior to the Shannon index. Like the first and second books, this book focuses on the application and analysis of the indices with respect to measuring 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 companion books. In addition to applying the SW indices, other Simpson-based indices such as the Ray and Singer (RS) index of concentration are presented in the book. The SW and RS indices are modifications of the Simpson ( $D = 1 - \sum p^2$ ) diversity index and assess demographic heterogeneity. This book applies the Simpson-based indices to the same data used in previous books 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 on the analysis of diversity scores obtained by the McIntosh index. Because the index incorporates 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 indices covered in previous companion 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 the previous companion books. 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. Like the McIntosh evenness index presented in the 5th book, the IQV takes into account jointly the number of groups used in the categorization of the demographic (or social) characteristic of interest and the size of the workforce.
- *Assessing the Validity of Diversity Indices* compares the indices used in each companion 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 each of the diversity indices presented in this book series. This volume 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 volume focuses on the development and analysis of a structural equation model for each diversity index discussed in the book series. In so doing, the analyses treat age, ethnic, and gender diversity as an intervening (or mediating) variable of organizational performance.

For purposes of continuity and compatibility, each diversity index is subjected to the same statistical analyses. The IQV, McIntosh evenness, Shannon, Simpson, and SW indices are of special focus 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 address 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

Diversity (or *integration*) indices have been used to assess the level of demographic (or social) heterogeneity in nonprofit, private, or public organizations since the early 1970s (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). For the most part, the application of diversity indices to aggregate demographic employment data has centered on measuring age, ethnic, or gender heterogeneity. More recently, diversity indices have been used to assess concepts such as educational and occupational diversity. In most of the previous studies, workforce diversity has served as a dependent variable. More recent studies, however, 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). Consistent with the companion books in the series, 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**

As stated in *Assessing Organizational Diversity with the Simpson Index* (Guajardo, 2023), several indices of diversity (or variation) have been used to measure demographic (or social) 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; Nachmias and Rosenbloom, 1973);
- McIntosh evenness index (e.g., Guajardo, 2013 and 2015);

- Shannon (Teachman) index (e.g., Choi, 2010; Choi and Rainey, 2010); and,
- Simpson (Blau or Lieberman) diversity index (e.g., Guajardo, 2014; Pitts, 2005; Starks, 2009).

Except for the Shannon index, each index has a theoretical distribution of scores ranging from 0 to 1. An index score of 0 means the absence of diversity (or heterogeneity), and a score of 1 means absolute diversity. Of the indices mentioned above, this book focuses exclusively on the Shannon 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:

$$D_E = \frac{N - \sqrt{\sum n^2}}{N - \frac{N}{\sqrt{S}}}$$

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.

This index has a distribution of scores ranging from 0 to 1. As stated above, 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 (1961) developed the IQV to assess demographic heterogeneity (or variation) in a community setting (see 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 (see Mueller, Schuessler, and Costner, 1970; Wilcox, 1967). Symbolically, the IQV index takes the following form (see Mueller, Schuessler, and Costner, 1970; Wilcox, 1967):

$$IQV = \frac{\text{Total observed differences}}{\text{Maximum Possible Differences}} = \frac{\sum f_i f_j}{\frac{s(s-1)}{2} \times \left(\frac{F}{s}\right)^2}$$

where



$f$  represents the frequency (or number) of individuals;  
 $s$  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. Theoretically, a score of 0 indicates a lack of heterogeneity (i.e., homogeneity), and a score of 1 indicates absolute heterogeneity (see Mueller, Schuessler, and Costner, 1970). *Assessing Organizational Diversity with the Index Qualitative Variation* focuses exclusively on the application of the index to public organizations.

Shannon (1948) created the H index 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  $k$ th category. Unlike the McIntosh and IQV 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 ( $H_M = \ln(n) = \ln(5) = 1.609$ ) when ethnicity is categorized into 5 groups. The following formula is used to standardize H index scores:  $H_s = \frac{-\sum [p \ln(p)]}{\ln(n)}$ . The Shannon index is the focus of this book.

As stated in *Assessing Organizational Diversity with the Simpson Index* (Guajardo, 2023), the Simpson diversity index is used 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 characteristics. As discussed in the companion book, Simpson's index has a distribution of scores ranging from 0 to 1. As developed by Simpson, 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, when ethnicity is categorized into 5 groups, the maximum diversity score is 0.80 ( $S_M = \frac{k-1}{k} = \frac{4-1}{5} = 0.80$ ). Simpson's index is represented frequently by the following formula:  $D_A = 1 - \sum p^2$ , where  $p$  represents the percent of individuals in a particular category or group. The index also is represented

as follows:  $D_B = \frac{\sum n(n-1)}{N(N-1)}$ , 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:  $D_s = \frac{D}{\frac{(k-1)}{k}}$ , where  $D$  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 diversity 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 Shannon diversity scores impact the findings when statistical assumptions are violated.

As stated in *Assessing Organizational Diversity with the Simpson Index* (Guajardo, 2023), this book series 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 satisfy 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, most 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  $\ln(n)$  when the Shannon index is used. 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 for analyzing a distribution of diversity scores that has a restricted quantitative continuous range of values in comparison to OLS regression.

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 published articles and books that address organizational (or workforce) diversity, few have focused specifically on the adaptation and 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 publications 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). Because 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 with respect to normality, skewness, and other distribution characteristics are often taken for granted or glossed over at best.

Due to the paucity of research on the application of diversity indices to demographic 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 satisfies 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 with respect 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 with respect to measuring heterogeneity as accurately as possible when there are categories with missing data?

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

## **Application of the Shannon Index**

This book applies the Shannon index to employment and workforce data reported by NYC departments for fiscal year 2019. NYC departments are used to highlight the measurements issues that are likely to be confronted in other employment setting regardless of whether the organizations are nonprofit, private, or public institutions. The Shannon diversity index is presented in Chapter 2. The traditional Shannon index formula ( $H = -\sum p \ln p$ ) is applied to categorized demographic employment data to illustrate how diversity scores are calculated for data presented as percentages. In addition, the formula for obtaining unbiased Shannon 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 Shannon index entails the use of the traditional formula and 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 the scores are compared statistically for equality.

Chapter 4 applies the Shannon index to demographic data to obtain diversity scores for ethnicity. Because NYC department categorize ethnicity in two different ways, four sets of unstandardized Shannon 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 score 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 methods of categorizing ethnicity allow for assessing changes in the distribution of the scores and for testing the sets of diversity scores for statistical equality.

In Chapter 5, diversity scores for age are obtained by applying the Shannon index to the demographic data. The issues of measurement compatibility and reliable 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 Shannon 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 Shannon index does not incorporate the number of groups in the calculation of the diversity scores, the level of age diversity is inflated for small and midsize NYC departments.

A composite score of organizational diversity is developed in Chapter 6. This is done for illustrative purposes. Weighted and unweighted composite scores for organizational diversity are presented. The distributions of scores are tested for equality, normality, and skewness.

## **Analysis of demographic diversity**

OLS regression is presented in Chapter 7. In so doing, statistical models for age, ethnic, and gender diversity are developed and discussed. The assumptions underlying OLS regression are presented. A Pearson pairwise correlation analysis is undertaken as an initial step to screen for collinearity amongst the predictors. Regression diagnostics are performed to assess for multicollinearity. The OLS regression findings are compared to the statistical findings of other regression methods in subsequent chapters.

Chapter 8 focuses on the application of robust regression. The rationale for undertaking the robust regression analysis is to adjust for atypical diversity scores. The statistical models tested with OLS regression are tested with robust regression. Lastly, the statistical findings obtained with OLS and robust regression are compared and discussed.

Tobit regression is presented in Chapter 9. As stated above, the maximum value of a Shannon diversity score is determined by the number of categories created for the demographic (or social) characteristic of interest. As such, the maximum value of a particular distribution of Shannon diversity scores is determined by  $\ln(n)$ . For each Tobit regression, the lower (or left) limit is set at 0 and the upper (or right) limit is set by  $\ln(n)$  where  $n$  is the number of groups. Tobit regression is used to test the statistical models tested with OLS regression in Chapter 7. After discussing the Tobit regression findings, they are compared with the statistical findings obtained with OLS regression. Because the distributions of Shannon scores are restricted to a range between 0 and  $\ln(n)$ , Tobit regression is the preferred statistical method to use.

In Chapter 10 quantile regression is presented. The statistical analysis focuses on the median although quantile regression is capable of producing findings for multiple percentiles simultaneously. The rationale for presenting quantile regression is that satisfying the assumption of normality is not necessary. Secondly, unlike OLS regression which is influenced by atypical values, quantile regression is less prone to the influence of atypical values. For comparative purposes, the statistical findings obtained with OLS and quantile regression are discussed. *Assessing Organizational Diversity with Quantile Regression* presents a thorough application of quantile regression to different diversity scores obtained with the same demographic data.

Chapter 11 applies ridge regression to address the issue of collinearity amongst the predictors. In so doing, ridge regression is applied to the biased and unbiased Shannon diversity scores. After applying ridge regression to the data, the level of collinearity obtained by the method is compared to the level of collinearity obtained by the OLS regression models. Consistent with previous chapters, the OLS and ridge regression findings are compared.

Chapter 12 compares the statistical findings for age, ethnic, and gender diversity that were obtained with OLS, robust, Tobit, quantile, and ridge regression. The comparison begins with discussing the findings for age diversity. Specifically, the findings for the biased and unbiased diversity score are compared for each statistical model. Similar comparisons are made for the findings for ethnic and gender diversity.

Path analysis is presented in Chapter 13. The chapter begins with discussing the role of employee turnover and hiring with respect to changes in the level of demographic (or social) diversity in organizations. A theoretical causal model illustrating the relationships amongst organizational factors and diversity in terms age and ethnicity is provided. Testable causal models are developed based on a Pearson pairwise correlation analysis and on the statistical findings obtained by the regression analyses performed in previous chapters. Tobit regression is not used to perform the path analysis because the mediating variables have quantitative continuous distributions that are unrestricted in terms of their values. *Assessing Organizational Diversity with Structural Equation Modeling* (SEM) provides an in-depth presentation of path analysis and SEM. In the final book of the series, different diversity indices are analyzed and modeled with the use of the same demographic and employment data.

Lastly, Chapter 14 summarizes the key points of the book. The summary focuses on measurement and analytical issues. Measurement strengths and