



Information Paper

An Introduction to Socio-Economic Indexes for Areas (SEIFA)

2006

Information Paper

**An Introduction to
Socio-Economic Indexes
for Areas (SEIFA)**

2006

**Brian Pink
Australian Statistician**

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) WED 26 MAR 2008

ABS Catalogue No. 2039.0

ISBN 0 642 47936 4

© Commonwealth of Australia 2008

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the Commonwealth. Requests and inquiries concerning reproduction and rights in this publication should be addressed to The Manager, Intermediary Management, Australian Bureau of Statistics, Locked Bag 10, Belconnen ACT 2616, by telephone (02) 6252 6998, fax (02) 6252 7102, or email: <intermediary.management@abs.gov.au>.

In all cases the ABS must be acknowledged as the source when reproducing or quoting any part of an ABS publication or other product.

Produced by the Australian Bureau of Statistics

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

CONTENTS

page

CHAPTER 1. INTRODUCTION

This chapter briefly introduces the 2006 SEIFA indexes, the concept SEIFA is measuring, and changes from the 2001 SEIFA. 4

CHAPTER 2. WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

This chapter describes the SEIFA numbers and briefly considers the most frequently asked questions about SEIFA. Some of these issues are discussed in greater detail in later chapters. 7

CHAPTER 3. HOW SEIFA IS CONSTRUCTED

SEIFA indexes are created by combining information selected from the Census. Each piece of information is first selected on the basis of a concept of relative socio-economic disadvantage; and then turned into a 'variable' which can be used in the index creation process. This chapter looks at the concept behind the indexes, how the information is selected and what a typical variable looks like. This chapter also looks briefly at principal components analysis. 17

CHAPTER 4. HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

This chapter aims to assist the interpretation of the index scores, using distributions, examples and maps. This chapter also briefly introduces the geographic areas used in SEIFA. 24

CHAPTER 5. HOW TO USE SEIFA

This chapter brings together some of the major issues surrounding the use of SEIFA that were discussed in previous chapters. This chapter is intended to briefly provide more detail about the use of SEIFA, but is not intended to replace a proper technical understanding of analysis. 32

CHAPTER 6. ENSURING SEIFA MEASURES DISADVANTAGE

This chapter briefly discusses some of the methods used to validate SEIFA in order to ensure SEIFA measures relative socio-economic advantage and disadvantage. 38

CHAPTER 7. FURTHER INFORMATION

References for further information. 40

APPENDIX

List of SEIFA Variables. 41

GLOSSARY

Brief definition of terms used in this paper. 44

ABBREVIATIONS

ABS	Australian Bureau of Statistics
ASGC	Australian Standard Geographical Classification
CD	collection district
IEO	Index of Education and Occupation
IER	Index of Economic Resources
IRSAD	Index of Relative Socio-economic Advantage and Disadvantage
IRSD	Index of Relative Socio-economic Disadvantage
LGA	local government area
POA	Postal Area
SD	statistical division
SEIFA	Socio-Economic Indexes for Areas
SLA	statistical local area
SSC	State Suburbs
SSD	statistical subdivision

SUMMARY OF CONTENTS

THIS RELEASE

This release of 2006 SEIFA supersedes the preliminary release.

CHAPTER 1. INTRODUCTION

This chapter briefly introduces the 2006 SEIFA indexes, the concept SEIFA is measuring, and changes from the 2001 SEIFA.

- SEIFA : BRIEF OVERVIEW
- ABOUT THIS PAPER
- BRIEF HISTORY OF SEIFA
- THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE
- WHAT IS A SEIFA INDEX?
- THE INDEXES
- SUMMARY OF CHANGES TO 2006 SEIFA

CHAPTER 2. WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

This chapter describes the SEIFA numbers and briefly considers the most frequently asked questions about SEIFA. Some of these issues are discussed in greater detail in later chapters.

WHAT ARE THE SOCIO-ECONOMIC INDEXES FOR AREAS (SEIFA) ?

- Where can I find the SEIFA indexes?
- In what format are the indexes available?

WHAT ARE THE FOUR SETS OF NUMBERS IN EACH INDEX?

- What is a SEIFA score?
- What is a SEIFA rank?
- What is a SEIFA decile?
- What is a SEIFA percentile?

THE GEOGRAPHIC AREAS

- What type of areas are available in SEIFA?
- Can I use other areas?
- Why is a SEIFA score not created for every CD?
- Are area boundaries comparable to previous years?
- How can a street have two different SEIFA scores?
- Why does my suburb have multiple SEIFA scores?

WHICH INDEX DO I USE?

- Index selection
- Why does the same area have a different rank for each index?
- Index of Relative Socio-economic Disadvantage (IRSD)
- Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)
- Index of Economic Resources (IER)
- Index of Education and Occupation (IEO)

OTHER FREQUENTLY ASKED QUESTIONS

- How do I match SEIFA to my data?
- For how many years are SEIFA indexes available?
- Can I compare SEIFA across time?
- Is everyone disadvantaged in a low score area?
- What information was / was not used to construct SEIFA?
- Can I create my own index?
- Why do CDs in different deciles have the same score?

SUMMARY OF CONTENTS *continued*

CHAPTER 3. HOW SEIFA IS CONSTRUCTED

SEIFA indexes are created by combining information selected from the Census. Each piece of information is first selected on the basis of a concept of relative socio-economic disadvantage; and then turned into a 'variable' which can be used in the index creation process. This chapter looks at the concept behind the indexes, how the information is selected and what a typical variable looks like. This chapter also looks briefly at principal components analysis.

THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE

- Disadvantage has no perfect measure
- SEIFA is a summary measure
- SEIFA is a relative measure
- Disadvantage is a social construct
- Disadvantage is subjective
- The international and ABS approaches

INFORMATION SELECTION

- Information selection
- Impact of PCA on selection
- Consistency and changes

VARIABLE CREATION PROCESS

PRINCIPAL COMPONENTS ANALYSIS (PCA)

EXAMINING THE VARIABLES

- Analysing the variable distributions
- Variable distributions and SEIFA scores

CHAPTER 4. HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

This chapter aims to assist the interpretation of the index scores, using distributions, examples and maps. This chapter also briefly introduces the geographic areas used in SEIFA.

WHAT DO THE INDEXES LOOK LIKE?

- Terms: scores, ranks, deciles and percentiles
- Index distributions and interpretation
- Index of Relative Socio-economic Disadvantage (IRSD) Distribution
- Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) Distribution
- Index of Economic Resources (IER) Distribution
- Index of Education and Occupation (IEO) Distribution

INTERPRETING SEIFA USING A COMPARISON OF TWO CDS

INTERPRETING SEIFA USING MAPS

INTRODUCTION TO GEOGRAPHY

- Creation of SEIFA for areas larger than a CD
- Further information on geographies

SUMMARY OF CONTENTS *continued*

CHAPTER 5. HOW TO USE SEIFA This chapter brings together some of the major issues surrounding the use of SEIFA that were discussed in previous chapters. This chapter is intended to briefly provide more detail about the use of SEIFA, but is not intended to replace a proper technical understanding of analysis.

HOW TO USE SEIFA

- SEIFA as a general tool
- Impact of PCA on SEIFA
- Comparing SEIFA over time
- SEIFA is an area level measure

EXAMPLES OF USE OF SEIFA

- Distributions for State, SD and SSD geographic areas
- Descriptive analysis
- Regression analysis and SEIFA: For more technical use
- Hints when using SEIFA to analyse survey data

FURTHER INFORMATION

CHAPTER 6. ENSURING SEIFA MEASURES DISADVANTAGE This chapter briefly discusses some of the methods used to validate SEIFA in order to ensure SEIFA measures socio-economic disadvantage.

CHECKING SEIFA

- Checking the indexes
- Checking the scores
- SEIFA and remoteness: An example

CHAPTER 7. FURTHER INFORMATION

SEIFA documentation

Geography documentation

Census documentation

APPENDIX: LIST OF SEIFA VARIABLES

Index of Relative Socio-economic Disadvantage

Index of Relative Socio-economic Advantage and Disadvantage

Index of Economic Resources

Index of Education and Occupation

GLOSSARY

CHAPTER 1 INTRODUCTION

CHAPTER 1 INTRODUCTION

This chapter briefly introduces the 2006 SEIFA indexes, the concept SEIFA is measuring, and changes from the 2001 SEIFA.

- SEIFA : BRIEF OVERVIEW
- ABOUT THIS PAPER
- BRIEF HISTORY OF SEIFA
- THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE
- WHAT IS A SEIFA INDEX?
- THE INDEXES
- SUMMARY OF CHANGES TO 2006 SEIFA

SEIFA : BRIEF OVERVIEW

SEIFA stands for Socio-economic Indexes for Areas. This suite of indexes ranks geographic areas across Australia in terms of their socio-economic characteristics. The SEIFA indexes are created by combining information collected in the five-yearly Census of Population and Housing (called the Census throughout this paper). There are four different indexes, each representing a slightly different concept. These concepts are abstract and difficult to measure, so the indexes aim to capture these abstract concepts by combining information that is related to the concept. For example, the Index of Relative Socio-economic Disadvantage uses information such as low income and low education as markers of relative socio-economic disadvantage.

The SEIFA indexes are rankings. Each index ranks different geographic areas of Australia according to a 'score' that is created for the area based on characteristics of people, families and dwellings within that area.

For all of the indexes, relative disadvantage is associated with a low number.

There are a number of ways the indexes can be used, such as targeting areas for business or services, 'demographic profiling', strategic planning, design of sample surveys, and social or economic research. The most appropriate index should be chosen for each application.

The 2006 SEIFA indexes are freely available on the ABS website, cat. no. 2033.0.55.001 Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only, 2006.¹

ABOUT THIS PAPER

This paper is an information manual. It is designed to explain SEIFA in non-technical terms. This paper is written so that each chapter is relatively self-contained for online access, and therefore may feel repetitive if all chapters are read together. Refer to Chapter 7 for references to more detailed and technical resources, especially the Socio-economic Indexes for Areas (SEIFA) Technical Paper (ABS cat. no. 2039.0.55.001, 2006) (called the Technical Paper throughout this paper).²

- | | |
|---------------------|--|
| This Chapter | provides an <i>Introduction</i> : it briefly overviews each of the indexes, the concept SEIFA is measuring, and changes from the 2001 SEIFA. |
| Chapter 2 | <i>What is SEIFA?</i> This chapter answers a number of <i>Frequently Asked Questions</i> : it describes the SEIFA numbers and indexes, and discusses the most frequently asked questions about SEIFA. |
| Chapter 3 | <i>How SEIFA is Constructed</i> : looks at the concept behind the indexes, how the information is selected and what a typical variable looks like. This chapter also looks briefly at principal components analysis. |

1 <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001>

2 <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0.55.001>

CHAPTER 1 INTRODUCTION *continued*

Chapter 4	<i>How to Interpret SEIFA Score Distributions:</i> assists the interpretation of the index scores, using distributions, examples and maps; and briefly introduces the geographic areas used in SEIFA.
Chapter 5	<i>How to Use SEIFA:</i> briefly discusses the use of SEIFA in analysis, bringing together some of the issues discussed in previous chapters.
Chapter 6	<i>Ensuring SEIFA Measures Disadvantage:</i> briefly discusses some of the methods used to ensure SEIFA measures socio-economic disadvantage.
Chapter 7	<i>Further Information:</i> provides references for related documentation.
Appendix	<i>List of SEIFA Variables:</i> lists the variables used to create each index.
Glossary	provides more detail on some of the terms discussed in this paper.

BRIEF HISTORY OF SEIFA

An index similar to SEIFA was first produced by the ABS from the 1971 Census. Substantial development of the indexes occurred in the late 1980s. SEIFA was first produced in its present form in 1990, using 1986 Census data, and consisted of five separate indexes. For the 2001 SEIFA, the variable selection process was revised and two indexes (the Urban and Rural indexes) were replaced with a single Index of Relative Socio-economic Advantage and Disadvantage. The same four indexes have been released for 2006 SEIFA and are overviewed later in Chapter 1 and discussed further in Chapters 2 and 4.

THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE

The concept of relative socio-economic disadvantage is abstract and difficult to capture. The SEIFA indexes are trying to capture socio-economic disadvantage, in terms of relative:

- access to material and social resources; and
- ability to participate in society.

As discussed in Chapter 3, there are two important issues to remember about SEIFA and the concept of relative socio-economic disadvantage:

- there is no perfect measure of disadvantage; and
- it is a summary measure that represents an 'average' of people and households in a particular area.

WHAT IS A SEIFA INDEX?

What is a SEIFA index?

SEIFA indexes are summary measures of a number of variables that represent different aspects of relative socio-economic disadvantage and/or advantage in a geographic area. For example, the Index of Relative Socio-economic Disadvantage uses information deemed to indicate a general level of relative socio-economic disadvantage. This index can then be used to show how one community compares to another community; that is, every area in Australia can be ranked using this summary measure of relative disadvantage.

How are the indexes created?

A summary measure for a particular community is created by combining information about the households and individuals who live in that area. The method used to combine this information is called Principal Components Analysis.

What information is used?

Because the indexes are strongly influenced by the choice of information used to create them, the information is selected with care. Information from the Census is only selected if it is deemed to reflect some element of relative socio-economic disadvantage. Each index aims to capture a different aspect of relative socio-economic disadvantage and therefore different information is used to construct each index.

CHAPTER 1 INTRODUCTION *continued*

THE INDEXES

Each index measures a slightly different aspect of socio-economic conditions within an area. Please refer to Chapters 2 and 4 for more detail on each index.

Disadvantage The Index of Relative Socio-economic *Disadvantage* (IRSD) is a general socio-economic index that summarises a range of information about the economic and social resources of people and households within an area. Unlike the other indexes, this index includes only measures of relative disadvantage. This index will be called the Index of Relative Disadvantage throughout this paper.

Advantage and Disadvantage The Index of Relative Socio-economic *Advantage and Disadvantage* (IRSAD) also summarises information about the economic and social resources of people and households within an area, however includes both relative advantage and disadvantage measures. This index will be called the Index of Relative Advantage and Disadvantage throughout this paper.

Economic Resources The Index of *Economic Resources* (IER) focuses on the general level of access to economic resources of people and households within an area.

Education and Occupation The Index of *Education and Occupation* (IEO) focuses on the general level of education and occupation-related skills of people within an area.

SUMMARY OF CHANGES TO 2006 SEIFA

As a general rule, every effort is made to keep SEIFA the same as the previous release. However, some changes are important or unavoidable. These changes are overviewed in later chapters and discussed in more detail in the Technical Paper.

New variables Some new data items that were introduced into the 2006 Census were analysed and included in the indexes. These new measures relate to broadband internet access, and a need for assistance with core activities.³ Some existing Census data items that had not been included in previous SEIFA indexes, were reconsidered for inclusion in some 2006 SEIFA indexes. For example, variables relating to housing costs and household over/undercrowding were analysed and included in the 2006 Index of Relative Advantage and Disadvantage.

Changed Census variables Some variables were changed in the 2006 Census, as detailed in the Technical Paper. The more important changed variables related to occupation and dial-up internet access.

Other major changes A significant change to SEIFA in 2006 is the use of Equivalised Household Income instead of the set of twelve family-structure income variables previously used. Both methods take into account the family structure of the household, such as the number of adults and children. However, equivalised household income captures more family types and is represented by only two variables in SEIFA. This change affects three of the indexes, particularly the Index of Economic Resources.

Previous SEIFA were created using Census information about people who were in an area *on Census Night*. For 2006, SEIFA instead uses information about people who *usually reside* in the area. This means 2006 SEIFA is representative of the usual residents of an area, rather than people who happened to be in the area on Census Night, such as travellers.

Index of Economic Resources The Index of Economic Resources has been revised to include variables that represent a wider range of economic resources, particularly aspects of relative wealth.

Excluded areas Some areas do not have a SEIFA score. For example, if there are too few respondents in an area, then a valid score cannot be calculated for the area. The rules for excluding areas have been revised and are described in Chapter 2.

³ The new Census questions on need for assistance were developed to provide an indication of whether people have a profound or severe disability. People with a profound or severe disability are defined as those people needing help or assistance in one or more of the three core activity areas of self-care, mobility and communication, because of a disability, a long term health condition (lasting six months or more) or advanced age. This is not a comprehensive measure of disability. Please refer to the Glossary for further information

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

INTRODUCTION AND MAIN POINTS

This chapter describes the SEIFA numbers and briefly considers the most frequently asked questions about SEIFA. Some of these issues are discussed in greater detail in later chapters.

- WHAT ARE THE SOCIO-ECONOMIC INDEXES FOR AREAS (SEIFA)?
- WHAT ARE THE FOUR SETS OF NUMBERS IN EACH INDEX?
- THE GEOGRAPHIC AREAS
- WHICH INDEX DO I USE?
- OTHER FREQUENTLY ASKED QUESTIONS

Main points of Chapter 2

Some common issues when using SEIFA, as well as some of the limitations of SEIFA, are listed below. These issues are discussed in this chapter, and in more detail in later chapters.

- It is difficult to capture an abstract concept like socio-economic disadvantage. Please refer to Chapters 1 and 3 for more information.
- The indexes do not capture every aspect of socio-economic disadvantage. Please refer to Chapter 5 for more information.
- Comparison with previous indexes is not recommended. Please refer to Chapter 5 for more information.
- The method used to create the indexes is called Principal Components Analysis. Every method has strengths and limitations. Please refer to Chapters 3 and 5 for more information.
- SEIFA is a summary of people in an area and does not apply to an individual person or dwelling. Every area has a diverse range of different people and dwellings. The SEIFA indexes represent the general level of socio-economic disadvantage of all the people in the area in which a person lives, not the person themselves. Please refer to Chapter 5 for more information.

WHAT ARE THE SOCIO-ECONOMIC INDEXES FOR AREAS (SEIFA)?

SEIFA is a suite of four summary measures that have been created from Census information. Each index summarises a different aspect of the socio-economic conditions in an area, and therefore summarises a different set of social and economic information. The indexes can be used to explore different aspects of socio-economic conditions by geographic areas. For each index, every geographic area in Australia is given a SEIFA number which shows how relatively 'disadvantaged' that area is compared with other areas in Australia.

The four indexes in SEIFA 2006 are:

- Index of Relative Socio-economic *Disadvantage* (IRSD);
- Index of Relative Socio-economic *Advantage and Disadvantage* (IRSAD);
- Index of *Economic Resources* (IER); and
- Index of *Education and Occupation* (IEO).

Where can I find the SEIFA indexes?

The SEIFA indexes are available free of charge from the ABS website, cat. no. 2033.0.55.001 Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only, 2006.⁴

⁴ <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001>

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

In what format are the indexes available?

The SEIFA indexes are available in Excel spreadsheet format.

- A separate downloadable spreadsheet is available for each of the five geographical areas: Collection District (CD), Statistical Local Area (SLA), Local Government Area (LGA), Postal Area (POA) and State Suburbs (SSC).⁵
- Each spreadsheet contains the four SEIFA indexes that have been created for that geographical area, each on a separate worksheet. For example, the CD level spreadsheet contains the four CD level IRSD, IRSAD, IER and IEO indexes on separate worksheets. There is also a Summary worksheet containing the scores of all the indexes. The CD level spreadsheet also includes a worksheet with geographic concordances.
- The worksheet for each index shows:
 - the geographic area code. For example, the CD level spreadsheet lists the CD codes;
 - the number of people who usually resided in the area on Census Night;
 - the set of four index numbers, being the SEIFA score, rank, decile and percentile numbers (as described below) for the area; and
 - a State rank, a State decile and a State percentile number, which are applicable only within each State/Territory. Please refer to the Glossary for further details on these State numbers.
- For example, the CD level index for the Index of Economic Resources has a list of 37,457 CDs, with the IER score, IER rank, IER decile and IER percentile number for each CD.
- Previously, SEIFA has released separate indexes for Statistical Subdivisions (SSD), Statistical Divisions (SD) and States. These indexes have not been created for 2006 SEIFA, however a spreadsheet tool is available that shows population distributions of CD scores within these areas. This tool is discussed further in Chapter 5.

WHAT ARE THE FOUR SETS OF NUMBERS IN EACH INDEX?

SEIFA numbers

SEIFA is a tool that compares an area with other areas. It can be likened to a football premiership table. Just as a team that finished the season with 40 points cannot claim to be twice as good as a team that finishes the season with 20 points, so an area that has a SEIFA score of 1200 cannot claim to be twice as advantaged as an area that scored 600.

Each of the four indexes are released in four forms: as a 'score', as a 'rank', as a 'decile' and as a 'percentile'. Therefore, for each index, every geographical area is given a score number, a rank number, a decile number and as a percentile number. These numbers all rank areas in order of disadvantage, but are used for different purposes.

What is a SEIFA score?

A SEIFA score is created using information about people and households in a particular area. This score is standardised against a mean of 1000 with a standard deviation of 100. This means that the average SEIFA score will be 1000 and the middle two-thirds of SEIFA scores will fall between 900 and 1100 (approximately). (Refer to the Glossary for more on standardisation). A SEIFA score provides more information and is used for more sophisticated analysis. Deciles should be used for most analyses.

What is a SEIFA rank?

To determine the SEIFA rank, all the areas are ordered from lowest score to highest score. The area with the lowest score is given a rank of 1, the area with the second-lowest score is given a rank of 2 and so on, up to the area with the highest score which is given the highest rank, being 37,457 for a collection district (CD) index.

What is a SEIFA decile?

Deciles divide a distribution into ten equal groups. In the case of SEIFA, the distribution of scores is divided into ten equal groups. The lowest scoring 10% of areas are given a decile number of 1, the second-lowest 10% of areas are given a decile number of 2 and so on, up to the highest 10% of areas which are given a decile number of 10.

⁵ 2006 SEIFA indexes for CDs, SLAs, LGAs and POAs were released on 26 March 2008, with SSC indexes subsequently released.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

What is a SEIFA percentile? Percentiles divide a distribution into 100 equal groups. In the case of SEIFA, the distribution of scores is divided into 100 equal groups. The lowest scoring 1% of areas are given a percentile number of 1, the second-lowest 1% of areas are given a percentile number of 2 and so on, up to the highest 1% of areas which are given a percentile number of 100. SEIFA percentiles are provided to allow users to create their own groupings, such as quartiles (which contain 25% of CDs).

Refer to the Glossary for information on State ranks, State deciles and State percentiles.

THE GEOGRAPHIC AREAS

What type of areas are available in SEIFA?

The 2006 indexes are available for the following geographic areas:

- Collection District (CD);
- Statistical Local Area (SLA);
- Local Government Area (LGA);
- Postal Area (POA); and
- State Suburb (SSC).

Each type of area has its own set of scores, ranks, deciles and percentiles.

The smallest area for which SEIFA is available is a Census Collection District (CD). A CD is equivalent to a group of suburban blocks, roughly 250 households in a urban area. However, in a rural area, a CD can represent far fewer households. Both SLAs and LGAs are standard geographic areas which are based on CDs. POA is not an ASGC standard geographic area, but have been created by the ABS to match *Australia Post* postcodes as closely as possible. SSC is not an ASGC standard geographic area, but have been created by the ABS to match suburbs as closely as possible.⁶ Chapter 4 and the Glossary discuss these geographic areas further.

Can I use other areas?

If you wish to analyse a different type of area other than those available for SEIFA, then you may wish to create your own index. This is discussed later in this chapter.

Why is a SEIFA score not created for every CD?

It is not always meaningful to give a CD a score. For example, the CD may be an airport or a large office block, with no-one actually residing in the area. In addition, there may be very few people in the area, or only a few who responded to the Census questions. If there are only a few people or responses, then it becomes difficult to calculate a reliable score.⁷ Additionally, those who did respond may not be representative of the area as a whole. Confidentiality issues can also arise when there are only a few people in an area.

Areas were not given a score if they fell into one of the following categories:

- *No usual residents* (which accounted for 49% of excluded CDs);
- *Number of people* was ten or fewer (which accounted for another 13% of excluded CDs);
- Number of *employed* was five or fewer;
- Proportion of people *not responding* to the following Census questions (as appropriate) was 70% or more: occupation, labour force status, type educational institution, or non-school qualifications;
- Proportion of people in households for which *equivalised household income* could not be calculated was 70% or more.
- Number of *occupied private dwellings* was five or fewer;
- Proportion of people residing in *non-private dwellings* was 80% or more; or
- The area was classified as *off-shore, shipping or migratory*.

Please refer to the Technical Paper for further information.

⁶ POA and SSC are not part of the Australian Geographic Standard Classification (ASGC).

⁷ For small populations, each person greatly impacts the area's SEIFA score. For example, if there are five people in an area, then each person is equal to 20%. If one person in the area is a 'Manager', then an area with 100 people must have 20 'Managers' to be equivalent. One more 'Manager' would matter a lot to the smaller area, but little to the larger area. These two areas cannot really be compared.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

Are area boundaries comparable to previous years? SEIFA uses ABS geographic standards to determine area boundaries. However, these standards change over time for various reasons. For example, when new housing subdivisions are built or government area boundaries are revised, standards must be revised to match. There can be many changes in the five years since SEIFA indexes were last released. For example, in 1996 there were 34,422 CDs in Australia, with 37,207 CDs in 2001, and 38,704 CDs in 2006. Around 14% of 2006 CDs cannot be directly compared to 2001 CDs. Chapter 7 lists references for more information on ABS geographic standards.

How can a street have two different SEIFA scores? Because collection districts are based on suburban blocks, the blocks on either side of the same street could fall into different collection districts. If a street is long enough, it could also span two collection districts.

Why does my suburb have multiple SEIFA scores? Each CD is given its own SEIFA score. However, because CDs do not have names (only a code), it is not immediately obvious where a particular CD is located in Australia. As discussed above, CDs can be matched to other geographic areas that do have names. For example, a CD may be matched back to the SLA of PARRAMATTA (C) - INNER. There are two important points to note. First, as discussed above, separate SEIFA indexes are created for different geographies (CD, SLA, LGA, POA, SSC). It is important to be clear about which index you are using. If you are using the SLA index, then there will be only one score for the SLA of PARRAMATTA (C) - INNER. However, if you are using the CD index, then there are many CDs within an SLA, and each CD will have its own SEIFA score. For example, there are 53 CDs within the SLA of PARRAMATTA (C) - INNER. Second, the SLA name itself can cause confusion if it shares the same name as a suburb, such as the example used here. Please refer to Chapter 4 for a map of CD scores in an example SLA, and a discussion of the geographies used in SEIFA.

WHICH INDEX DO I USE?

Index selection

Four SEIFA indexes are available, and the choice of index depends on the analysis to be undertaken. There are two general indexes of socio-economic disadvantage, an index for economic resources and an index for education and occupation. It is important to be clear about the concept you are trying to capture before selecting an index. In addition to the discussion below, please also refer to Chapter 4, and to the Appendix for a full list of variables included in each index.

Why does the same area have a different rank for each index? Each index aims to capture a slightly different aspect of relative disadvantage and is constructed using different variables. It is therefore possible for the same area to have different rankings on each index. For example, it is possible for an area to rank relatively highly in the Disadvantage index, but not in the Advantage and Disadvantage index, because these indexes include different variables.

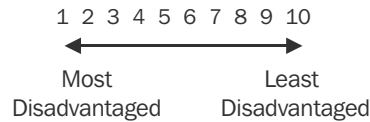
CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

Index of Relative Socio-economic Disadvantage (IRSD)

This index is a general socio-economic index that summarises a wide range of information about the economic and social resources of people and households within an area. Because this index focuses on disadvantage, only measures of relative disadvantage are included. This means that, unlike the other indexes, a high score (or decile) reflects a *relative lack of disadvantage* rather than relative advantage, as shown in figure 2.1.

FIGURE 2.1 IRSD, Interpretation of Deciles



This index summarises 17 different measures, such as low income, low education, high unemployment and unskilled occupations.

A *low score* indicates relatively greater disadvantage in general. For example, an area could have a low score if there are (among other things):

- many households with low income, many people with no qualifications, or many people in low skilled occupations.

A *high score* indicates a relative lack of disadvantage in general. For example, an area may have a high score if there are (among other things):

- few households with low incomes, few people with no qualifications or in low skilled occupations.

This index is preferred in situations where the user:

- wants to look at disadvantage and lack of disadvantage; and
- wants a broad measure of disadvantage, rather than a specific measure (such as low income).

An example would be where a user:

- wants to ensure an allocation of funds goes to disadvantaged areas.

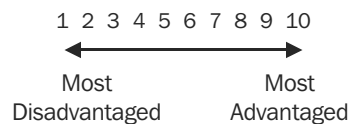
We do not recommend using this index if the user:

- is analysing information that was used to create this index, such as Indigenous status;⁸
or
- wants to look at both advantage and disadvantage.

Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)

This index is also a general socio-economic index that was created using measures of relative disadvantage (similar to those used in the Index of Relative Disadvantage), as well as measures of relative advantage. As shown in figure 2.2, a high score (or decile) reflects relative advantage.

FIGURE 2.2 IRSAD, Interpretation of Deciles



There are 21 measures included, such as: low or high income, internet connection, occupation and education. This index does not include Indigenous status.

A *low score* indicates relatively greater disadvantage and a lack of advantage in general. For example, an area could have a low score if there are (among other things):

- many households with low incomes, or many people in unskilled occupations; AND
- few households with high incomes, or few people in skilled occupations.

⁸ By definition there will be a relationship between an index and the information that was used to create it. Note however, that the information used to create SEIFA were area-level measures; that is, they were represented as a proportion of people or households within an area. This means that the information used by your analysis may be conceptually different to the information used in SEIFA, if it is a person-level measure for example. This note applies to all the indexes.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) *continued*

A *high score* indicates a relative lack of disadvantage and greater advantage in general.

For example, an area may have a high score if there are (among other things):

- many households with high incomes, or many people in skilled occupations; AND
- few households with low incomes, or few people in unskilled occupations.

This index is preferred in situations where the user:

- is not looking only at disadvantage and lack of disadvantage;
- wants advantage to offset any disadvantage in an area;
- is using a variable in their analysis, such as Indigenous status, that has been included in the construction of another index but not this index; or
- is unable to identify a specific aspect of disadvantage, such as income, that is important to their particular analysis.⁹

An example would be where a user:

- considers the issue being examined to be affected by both advantage and disadvantage. For example, this index would be suitable if the user is looking at health status and anticipates disadvantaged people to have worse health and advantaged people to have better health. If disadvantaged people are expected to have worse health, but it is less-disadvantaged people (rather than advantaged people) that are expected to have better health, then the Index of Relative Disadvantage would be more appropriate to use; or
- is analysing information that is not included in the index, such as home ownership or Indigenous status.

We do not recommend using this index if the user:

- is analysing information that has already been included in the index, such as access to broadband internet connection; or
- is only interested in disadvantage.

Index of Economic Resources (IER)

This index reflects the economic resources of households within an area. There are 15 variables included to measure a wide range of concepts, such as: household income, housing expenditures (e.g. rent) and wealth (e.g. home ownership). Some of these are new to the 2006 IER, which introduced additional measures to better capture economic resources, such as unemployment and ownership of an unincorporated enterprise. However, some important information, such as superannuation, was not included as it was not collected in the Census. This index does not include education or occupation measures. Similar to the Index of Relative Advantage and Disadvantage, this index includes measures that capture both 'low' and 'high' access to economic resources.

A *low score* indicates a relative lack of access to economic resources in general. For example, an area may have a low score if there are:

- many households with low income, or many households paying low rent; AND
- few households with high income, or few owned homes.

A *high score* indicates relatively greater access to economic resources in general. For example, an area may have a high score if there are:

- many households with high income, or many owned homes; AND
- few low income households, or few households paying low rent.

This index is preferred in situations where the user:

- is looking specifically at access to economic resources and the associated ability to participate in society.

We do not recommend using this index if the user:

- is only interested in disadvantage, as this index measures both advantage and disadvantage; or
- is interested in a more specific measure, such as income; or
- is interested in a more general measure, such as the Index of Relative Advantage and Disadvantage; or

⁹ A general measure of disadvantage is not suitable for every analysis.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

Index of Economic Resources (IER) *continued*

- is also analysing information that was included in the index, such as household income.

Index of Education and Occupation (IEO)

This index reflects the general level of education and occupation-related skills of people within an area. There are nine measures included in this index. The education information in this index includes qualifications achieved and whether further education is being undertaken. The occupation information in this index includes occupations that require a high level of skills, occupations that require a low level of skills, as well as unemployment. This index does not include any income measures. This index includes both 'low' and 'high' measures of education and occupation.

This index is highly correlated with the Index of Relative Disadvantage and the Index of Relative Advantage and Disadvantage, but has only a medium correlation with the Index of Economic Resources (which is not unexpected as these two indexes have only one variable in common).

A *low score* indicates relatively lower education and occupation status of people in the area in general. For example, an area could have a low score if there are:

- many people with few qualifications, or unemployed or in low skilled occupations;
- AND
- few people with a high level of qualifications or in highly skilled occupations.

A *high score* indicates relatively higher education and occupation status of people in the area in general. For example, an area could have a high score if there are:

- many people with higher education qualifications or in highly skilled occupations;
- AND
- few people without qualifications or in low skilled occupations.

This index is preferred in situations where the user:

- is interested in occupation and education variables; or
- is interested in an index that does not include income; or
- is interested in investigating the economic resources and the education and occupation separately.¹⁰

We do not recommend using this index if the user:

- is only interested in disadvantage (not advantage); or
- is interested in a more general measure of disadvantage and advantage; or
- is analysing information that has already been included in the index, such as unemployment.

The index scores are described further in Chapter 4.

¹⁰ The 2006 Index of Economic Resources was created using a different set of variables to those used to create the Index of Education and Occupation. These indexes share one variable, unemployment, however this variable is defined differently to match the concept of each index. However, these indexes are correlated.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

OTHER FREQUENTLY ASKED QUESTIONS

How do I match SEIFA to my data?

First, ensure that the SEIFA geographic areas match the geography areas of your data set. For example, if your analysis uses Census Collection Districts (CDs) then use the SEIFA index for CDs. Use the smallest area that your data set allows, preferably CDs.

Second, it should be noted that 2006 SEIFA geographic boundaries are based on the 2006 Australian Standard Geographic Classification (ASGC); 2001 SEIFA boundaries are based on the 2001 ASGC, and so on. Therefore, the 2006 SEIFA is based on information about people and households within the 2006 ASGC boundaries. Boundary changes mean that information based on the boundaries of one ASGC release will not necessarily be consistent with another ASGC release. Check which ASGC your survey data used. Always check the geographic standard used in the collection of your data, as differences can substantially affect comparisons.

Third, match (as closely as possible) the *time* of your data collection with the year the SEIFA data was collected. For example, if your survey was undertaken in 2006, then use the 2006 SEIFA as this was based on data collected in the 2006 Census.

In practice, you will rarely be able to perfectly match the timing of the data collection and the ASGC boundaries between SEIFA and your survey.

For how many years are SEIFA indexes available?

A type of socio-economic disadvantage measure was first produced using information from the 1971 Census. In their present form, the indexes were first produced in 1990 (using 1986 Census data). Later indexes have been produced in 1994 (1991 Census), 1998 (1996 Census), 2003 (2001 Census) and 2008 (2006 Census). The indexes from 1986 onwards are available on the ABS website.

Can I compare SEIFA across time?

This is a common question and many users would like to be able to compare an area's SEIFA score over time, however this is not recommended. There are a number of reasons for caution:

- SEIFA is based on information measured at a certain point in time. The indexes are based on the complex interactions between this information. Because the world changes, this information also changes, which results in changes to the indexes.
- A change in the rank of a particular area is not necessarily a reflection of changes within that area. For example, if an area has a rank of 19,000 in 2001 and a rank of 20,000 in 2006, this cannot be interpreted as the area being less disadvantaged than before. This higher rank may be due to *other areas* becoming more disadvantaged, rather than to any change within that particular area.
- While consistency in the construction of SEIFA is important, some changes are made where important or necessary.
- As discussed earlier in Chapter 2, the total number of areas measured also changes with each release, affecting the ranks. The 2001 SEIFA CD ranks ranged between 1 - 35,695; the 2006 SEIFA CD ranks ranged between 1 - 37,457. SEIFA is based on geography standards which evolve over time to include new CDs: the 2001 ASGC has 37,209 CDs and the 2006 ASGC has 38,704 CDs. In addition, some of these CDs are excluded from SEIFA and the number of excluded CDs also changes.
- As discussed earlier in Chapter 2, the ASGC area boundaries can change. If an area boundary changes to include a different group of households, then the overall characteristics of the area may change even if all the other households in the area do not.

For these reasons, comparing SEIFA over time can lead to misleading results. If you must compare the SEIFA score of an area over time, it is best to state that an area had:

- a SEIFA decile of 4 in 2001, and 5 in 2006; or
- a SEIFA rank of 19,000 in 2001, and 20,000 in 2006; or
- a SEIFA score of 977 in 2001, and 990 in 2006.

We strongly recommend that users do *not* go further and state that an area improved by 13 'SEIFAs', or even that the area has 'improved' at all.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

Can I compare SEIFA across time? *continued*

We recommend the use of deciles, rather than ranks or scores. The ranks are impacted by the different number of CDs included in each release. For example, if a CD has a rank of 50 in 2001 and a rank of 100 in 2006, this could be simply due to the larger number of CDs included in the 2006 index, rather than any change in the area. In addition, using the scores may imply a level of accuracy that is not provided by a general measure of relative disadvantage like SEIFA. Ranks and scores are appropriate only for more technical use. Further, we recommend that the extreme deciles be compared, because these tend to be more stable over time.

This issue is briefly discussed again in Chapter 5.

Is everyone disadvantaged in a low score area?

No. It is not correct to use SEIFA to represent a person or a household.

A SEIFA score is an *average* of people and households within a given area, therefore not everyone in the area would have that same score. Using the football analogy, a team with some excellent players may not necessarily qualify for the finals. In the same way, an area that SEIFA rates as relatively disadvantaged could have some advantaged households; just as an area ranked as relatively advantaged could have some households that are relatively disadvantaged.

SEIFA provides summary information about the people in an area, not information about an individual person in an area. If you have person-level data and you want to use SEIFA, then you need to be careful in your interpretation. Area-level measures do provide some information about individual people, however this is only part of the story. Area-effects and personal information can both be relevant to your analysis.¹¹ SEIFA represents the general socio-economic status of the area in which an individual lives. SEIFA says nothing about the socio-economic status of a particular person or household.

For example, a person could be employed in a high-paying job but live in an area of generally high unemployment and low income. In this situation, the person may be quite well-off even though the area's SEIFA score indicates relative disadvantage. If used in a person-level analysis, this could be interpreted as: "the area in which the person lives is relatively disadvantaged". It is not correct to use SEIFA to say a particular person is relatively disadvantaged, because other data would be needed to draw this conclusion.

What information was / was not used to construct SEIFA?

Measures deemed to represent different aspects of relative socio-economic disadvantage were considered for inclusion in the indexes. The data items were sourced only from the Census as this is the best source of detailed information within small geographic areas. However, this approach limits the range of data available because some information is difficult or sensitive to request. Therefore, some information was not included in SEIFA, including:

- Access to *infrastructure*, which impacts on people's ability to participate in society. Examples include access to schools, community services, shops, transport and medical facilities.
- *Wealth* information (such as the value of assets) directly impacts on the economic resources of households and the ability of people to participate in society. However, this concept is difficult to measure. Although the 2006 Index of Economic Resources has included more information that reflect aspects of wealth than earlier indexes, wealth itself is not directly measured in the Census and is therefore not included in SEIFA.
- The *remoteness* of an area can impact on people's ability to participate in society. Remoteness is not included in 2006 SEIFA.
- The *cost of living* within an area impacts on a household's access to economic resources and ability to participate in society. However, the difference in the cost of living between areas is difficult to measure and is not captured in the Census.

¹¹ In the same way, the macroeconomy and the global economy can also be relevant to your analysis.

CHAPTER 2 WHAT IS SEIFA? FREQUENTLY ASKED QUESTIONS

continued

What information was / was not used to construct SEIFA?
continued

- Much of the information related to socio-economic disadvantage is related to the *age* of a person, such as income and education. The decision whether to account for age is determined by the particular analysis for which SEIFA is being used, and therefore accounting for age has not been done within SEIFA.¹² Each user needs to consider whether their particular research should account for age. The decision depends on the particular analysis being undertaken. Please refer to Age Standardisation in the Glossary.
- There are many other important socio-economic issues which are not measured in the Census, such as health, crime rates, pollution and financial stress.

This information must be included separately in an analysis if considered important.

In addition, there is some information captured by the Census that was not included in SEIFA. There are two reasons why this information was not included. First, a measure that was too closely related to other measures already in SEIFA was dropped from the index. Second, a measure that contributed very little to SEIFA was also dropped from the index. This is explained in more detail in Chapter 3 and in the Technical Paper.

Can I create my own index?

Some users wish to create their own index to use *different information* to that used in SEIFA. Census information is freely available via CData Online, and the Basic Community Profile Datapack which is available on a cost-recovery basis.¹³ However it is not possible to create a complex index like SEIFA using these products. First, SEIFA is created using unperturbed Census data, whereas these products have been perturbed to protect confidentiality. Second, only a selection of the Census data items are available in these products. Last, as discussed in Chapter 3, the variables used in SEIFA are complex, and variables of this complexity cannot be recreated using these products. More complex variables can be requested through Client Services¹⁴, also on a cost-recovery basis. However, this data will also be perturbed.

Some users wish to create their own index to use *different areas* to that used in SEIFA. As discussed earlier in this chapter, SEIFA indexes are created for five different types of area: Census Districts (CDs), Statistical Local Areas (SLAs), Local Government Areas (LGAs), Postal Areas (POAs) and State Suburbs (SSC). However, you can also use the SEIFA indexes to create scores for other types of area. Because CDs form the basis of all of the ASGC standard geography boundaries, CD scores can be combined to represent larger standard areas. To create a score for a standard area, use a population weighted average of the CDs within the area. Bear in mind that this method changes the distribution of the index. These issues are discussed further in the Glossary (under Population Weighting) and the Technical Paper.

Why do CDs in different deciles have the same score?

The decimal places are not shown for the SEIFA scores as this may imply a level of accuracy that is not provided by a general measure of relative disadvantage like SEIFA. Therefore, some CDs appear to have the same score because the decimal places are not shown, and due to the standard distribution chosen for SEIFA. As discussed earlier, SEIFA scores have been standardised to a mean of 1000 and standard deviation of 100. This means that two-thirds of the CDs, approximately 25,000 CDs, will have a SEIFA score between 900 and 1100. This choice of standard distribution means that some CDs look like they have the same score when the decimal places are not shown. Using the Index of Relative Disadvantage as an example, both Decile 1 and Decile 2 have several CDs with a score of 879. If the standard deviation was instead set to 1000, then these scores would then be 8790, 8791 and so on.

Please refer to Standardisation in the Glossary for more information.

¹² Adjusting for age was only undertaken for the variable measuring 'need for assistance'. Refer to the Glossary for more detail.

¹³ Another Census product, Table Builder, was not yet available at the time of publication, and is expected to be available on a cost-recovery basis.

¹⁴ <http://www.abs.gov.au/websitedbs/d3310114.nsf/home/Census%20Contacts>

CHAPTER 3 HOW SEIFA IS CONSTRUCTED

CHAPTER 3 HOW IS SEIFA CONSTRUCTED?

INTRODUCTION

SEIFA indexes are created by combining information selected from the Census. Each piece of information is first selected on the basis of a concept of relative socio-economic disadvantage; and then turned into a 'variable' which can be used in the index creation process. This chapter looks at the concept behind the indexes, how the information is selected and what a typical variable looks like. This chapter also looks briefly at principal components analysis.

- THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE
- INFORMATION SELECTION
- VARIABLE CREATION
- PRINCIPAL COMPONENTS ANALYSIS (PCA)
- EXAMINING THE RELATIONSHIP BETWEEN VARIABLES AND SCORES

THE CONCEPT OF SOCIO-ECONOMIC DISADVANTAGE

The concept of relative socio-economic disadvantage is neither simple nor well-defined. Based on international research and also on information collected in Census, we broadly define relative socio-economic advantage and disadvantage in terms of people's access to material and social resources, and their ability to participate in society. However, there are some important things to remember about the concept of relative socio-economic disadvantage.

Disadvantage has no perfect measure

No single measure can fully capture the concept of relative socio-economic disadvantage. It is important to be aware that the information used to create SEIFA is only *related* to disadvantage and does not perfectly measure disadvantage. For example, information on low income is used in three of the indexes because it is related to disadvantage; and high income is related to advantage or lack of disadvantage. However, low income does not guarantee disadvantage; it is only an indicator that a household might be disadvantaged. Some low income households may have access to other economic resources such as wealth or support from other households, or their low current income could reflect a temporary situation, such as a business or investment start up.

SEIFA is a summary measure

A SEIFA index summarises the characteristics of people and households within an area. A SEIFA score therefore reflects this group of people as a whole; it does not reflect any one person or household within that area. In addition, areas are often quite diverse and so can have both high income and low income households, for example. It is possible for a high income household to reside in a relatively disadvantaged area.

SEIFA is a relative measure

It is incorrect to state that an area with a low SEIFA score *is* disadvantaged. It can only be determined that an area is disadvantaged *relative* to other areas.

Disadvantage is a social construct

Disadvantage is defined by the society in which we live. While an area could be disadvantaged compared to another area within Australia, it could be more advantaged than other parts of the world.

It is also difficult to create a single measure which captures disadvantage across a diverse country like Australia. The concept of disadvantage can have a slightly different interpretation for different regions and sub-cultures. For example, in some ways regional areas may be considered more disadvantaged than metropolitan areas due to their remoteness. SEIFA does not focus on disadvantage in terms of what it means to a person living in the city, nor solely to a person living in regional areas. Because SEIFA is a general index, it aims to measure relative socio-economic disadvantage in terms of what it means to everyone in Australia.

Disadvantage is subjective

What does it mean to be disadvantaged? Every person you ask is likely to give a different response, basing their answers on their own perceptions and different criteria. For example, most will agree that income has an important relationship with disadvantage. However, some may argue that poor access to health care services or education is a

Disadvantage is subjective *continued* greater indicator of disadvantage than low income. Because there are many dimensions of socio-economic disadvantage, disadvantage is difficult to measure. SEIFA is limited to reflecting only information that is measured in the Census.

The international and ABS approaches International and Australian studies have looked at similar concepts, including social capital, deprivation, poverty, well-being and social exclusion. While these concepts are related to socio-economic disadvantage, they are also different. In addition, some indexes have been developed based on these different concepts, such as the *New Zealand Index of Deprivation*. Employment, education and financial well-being are three dimensions common to most of these indexes. Other dimensions commonly included are unemployment, housing stress, overcrowding, home ownership, family support, family breakdown, family type, lack of wealth (no car or telephone), low income, Indigenous status and foreign birth.

SEIFA aims to capture relative socio-economic disadvantage by selecting data based on these common dimensions. As discussed earlier, the ABS broadly defines relative socio-economic disadvantage in terms of people's access to material and social resources, and their ability to participate in society.

As there can be no one definitive measure of advantage/disadvantage, there are four separate indexes in SEIFA. As discussed in Chapters 1 and 2, each index aims to capture a different aspect of relative socio-economic disadvantage. Because each index uses different variables, an area can therefore have a different score for each index.

Importantly, any measure of disadvantage will only reflect the information from which it is made. This makes the variable selection process very important, as discussed in the 'Information selection and variable creation' section of this chapter.

INFORMATION SELECTION

Selection Any measure of disadvantage will only reflect the information from which it is made, which makes the information selection process very important. SEIFA indexes are created by combining information selected from the Census. Information from the Census is first selected on the basis of the concept of relative socio-economic disadvantage defined earlier in this chapter. The appropriate Census data items are then turned into a 'variable', which are used in the index creation process.

Information selection Information is taken only from the Census as this is the best source of detailed information within small geographic areas. However, this approach limits the range of information available.

Information was selected on the basis of *association* with relative disadvantage, rather than any assumption of cause and effect. For example, low income was selected because it is associated with relative disadvantage, not because it is either a 'cause' or an 'effect' of disadvantage.

Some information is included because of a *direct* association with relative disadvantage, such as low income. Other information is included because of an *indirect* association; this information acts as a proxy for important information that is unable to be captured using Census data. A good example is the variable measuring the proportion of people in an area who identify themselves as Indigenous. SEIFA includes this information as a proxy measure of relative disadvantage because it captures important information that is otherwise unavailable from the Census, such as relative health and life expectancy information.

In reality all data, both direct and indirect, provide a wealth of indirect information about relative socio-economic disadvantage. That is, in some way all included data acts as a proxy for relative disadvantage. In SEIFA, the indexes aim to explain relative disadvantage by measuring some common dimension underlying these proxies.¹⁵

¹⁵ Because the data are proxies of relative disadvantage, it is difficult to use theory to predetermine appropriate weights for the data.

Impact of PCA on selection

When selecting information, it is important to consider the method used to create the indexes. The method used, principal components analysis (PCA)¹⁶, creates a summary measure from a group of select variables. PCA assumes there is some common dimension underlying these variables, and creates a summary measure to capture this commonality. This summary measure represents a 'line of best fit' which is entirely dependent on the relationships (correlations) between the variables. Unlike some other methods, PCA *requires* the variables to be correlated. It is actually the correlations that SEIFA is trying to capture, because this commonality among the variables is that which is deemed to measure relative socio-economic disadvantage. However, we need to make sure that any particular aspect of relative socio-economic disadvantage is not overrepresented as this could bias SEIFA.

SEIFA aims to measure relative socio-economic disadvantage using a select group of variables, each of which measures a different concept of relative disadvantage. Information is selected to capture an aspect of relative disadvantage, however the indexes can be biased if too many variables are included that measure the same aspect. Deciding whether a variable is too similar is not straightforward because, as discussed above, it is precisely the similarities between the information that SEIFA aims to capture. This issue is illustrated by the following examples.

- The Disadvantage index includes both a low-rent variable and a public-housing variable, even though many households within public housing also pay relatively low rent. It was decided that there was enough difference between the aspects of relative disadvantage captured by each variable (at an area-level) to include both.
- Some households have to access the private rental market until they are able to obtain public housing. Also, the Commonwealth Government provides rental assistance to eligible private renters. In SEIFA, these households are unable to be separated from other households in the private rental market.

Consistency and changes

While consistency across SEIFA releases is very important, changes were made where necessary or to improve the quality of the indexes. Refer also to Chapter 1 and the Technical Paper.

- Census information changes over time with new data items and improvements to existing variables. Some new data items were considered for inclusion in SEIFA, such as need for assistance with core activities.¹⁷ Some SEIFA variables were re-specified due to changes to Census data items.
- Classification standards used by Census are updated over time. SEIFA variables were affected by changes to classification standards for occupation.
- Some SEIFA variables were redefined because society has changed. For example, access to broadband internet is considered an indicator of improved ability to participate in society. However, access to broadband internet was not collected in earlier Censuses.
- Some existing Census data items that were not previously included in SEIFA, or included only in one or two SEIFA indexes, were reconsidered for inclusion in the 2006 SEIFA indexes. For example, unemployment was introduced to the 2006 Index of Economic Resources.
- Important improvements to SEIFA were made, such as the use of equivalised household income, and counting people in their usual residence (rather than where they were on Census night).

Because the indexes are sensitive to changes, the information selection process is important.

¹⁶ Please refer to the Principal Components Analysis (PCA) section later in this chapter, and to the Glossary for more information.

¹⁷ Please refer to the Glossary for further information on need for assistance with core activities.

VARIABLE CREATION PROCESS

The process of creating SEIFA variables is as follows:

Step 1: Information selection

As discussed, Census data items that represent an aspect of relative socio-economic disadvantage were selected. The information may relate to individuals, families or households.

Step 2: Variable specification

For the purposes of the PCA method, all SEIFA variables were expressed as proportions. Every proportion was specified using past specifications and advice from the ABS Census area and relevant subject matter areas. Many decisions were made for every variable specified.

For example, the variable LOWRENT was defined as the proportion of households paying rent who pay less than \$120 per week. The numerator and denominator for this variable was specified as:

the number of renting occupied private dwellings who pay less than \$120 per week rent (exclude rent = \$0), divided by
the number of renting occupied private dwellings with stated rent.

This example illustrates a number of the issues that arose when specifying variables. In order to make these decisions, it was important to consider the specific concept the variable was aiming to capture and the impact of the alternative choices. Some of the issues in this example include:

- Whether to count people, families or dwellings. The proportion of people (in low rent households) can be different to the proportion of dwellings (that pay low rent).
- Whether the denominator should include the number of rented dwellings or the number of total dwellings. It is important to consider the impact of this decision for different situations, such as the case of a CD with no rented dwellings.
- Whether the denominator should include dwellings who did not state their rent payments. In this example, only those dwellings that stated their rental payments were included.
- Whether to include zero rent payments in the numerator. In this example, rented dwellings paying zero rent were not included in the numerator, because these dwellings tended to have different characteristics to dwellings paying low rent.
- How to determine the cut-off values. In this example, rental payment of \$120 per week was chosen because this represented 20 per cent of rented dwellings. This cut-off does not represent any presumed level of *absolute* disadvantage; it was selected to reflect the concept of *relative* disadvantage.
- While people could be counted within their CD of usual residence (if this information was provided), dwelling information was only recorded if at least one person was at home on Census Night. If the usual residents were elsewhere on the night (on holidays, for example), then no-one would have been at home to record *that* dwelling's characteristics.

Step 3: Missing values

All surveys and censuses contain missing information, either due to non-response or due to responses being illegible. While Census data processing imputed data for age, sex, marital status and usual residence, other information was coded to 'Not stated'. However, some data items, such as income, were considered crucial to the construction of SEIFA indexes. As discussed in Chapter 2, a SEIFA score was not created for a CD that had a significant proportion of missing data for these important variables. Therefore, if a CD had a high level of non-response for an important variable, then that CD was excluded from the analysis and no SEIFA scores were created for that CD. In some instances, such as the previous example, the variables themselves were specified to exclude a dwelling or person if some information was not stated.

Step 4: Data extraction

Data was extracted from the Census and used to create the numerators and denominators according to the SEIFA variable specifications.

CHAPTER 3 HOW SEIFA IS CONSTRUCTED *continued*

Step 5: Checking The numerators and denominators were checked against published State totals (where possible) or by being re-created using a different methodology.

Step 6: Variable creation Variables were created from the numerators and denominators. Summary statistics were created for each variable so that the variable distributions could be examined.

PRINCIPAL COMPONENTS ANALYSIS (PCA)

Once the variables have been created, they are used to create a score for every CD. Combined, this set of CD scores is called a SEIFA index.

This section briefly explains the method used to create the SEIFA indexes, called Principal Components Analysis (PCA). PCA is discussed further in the Glossary and in the Technical Paper. These summaries are in no way intended to replace a proper technical understanding of this technique.

The concept of relative socio-economic disadvantage is difficult to capture because it has many dimensions which are hard to measure. In SEIFA, PCA is used to create a summary measure of a group of characteristics. For example, the Index of Relative Socio-economic Disadvantage is a summary measure of a group of characteristics related to relative socio-economic disadvantage. There is no preconception about how important each characteristic is to the index. The importance of a characteristic is determined by the relationships between the characteristics themselves across all the areas. PCA uses these complex relationships to create a 'weight' for each characteristic. A SEIFA score can then be calculated for an individual area using that area's own characteristics and these weights.

EXAMINING THE RELATIONSHIP BETWEEN VARIABLES AND SCORES

Analysis of the variable distributions

This section aims to further explain the relationship between variables and CD scores.

Because the SEIFA variables were expressed as proportions, they could only have a value between zero and one. For example, if a CD has 100 dwellings with 20 earning low income, then the *LOW INCOME* variable value for that CD would equal $20/100 = 0.20$. Each CD would therefore have a set of values, one for every variable, between zero and one.

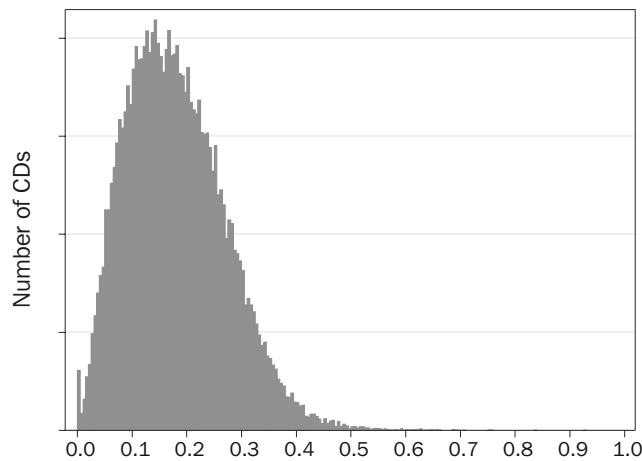
One way to describe a variable is to graph the value for every CD along a zero-one axis. This is called the variable's distribution. This distribution shows how many CDs have a variable value of 0.10, how many have a value of 0.11 and so on. Most of the 2006 SEIFA variables have a similar distribution to that shown in figure 3.1. These distributions are generally 'right-skewed' and have a low average value with a long right tail. The skewness of these variables is expected, for example many CDs have a low proportion of rented dwellings paying low rent, but some CDs have a high proportion.

The variables with the greatest skewness are: low rental payments; high rental payments; Indigenous status; no schooling; overcrowding; current university attendance; and renting from government or community organisation. These variables were also highly skewed in 2001.

Some variables have a more 'normal' distribution, which looks more symmetrical: no post-school qualifications; no year 12 education, no internet connection; mortgage home ownership; outright home ownership. The spare bedroom variable is left-skewed. This was a similar feature of the 2001 SEIFA variables.

Analysis of the variable distributions *continued*

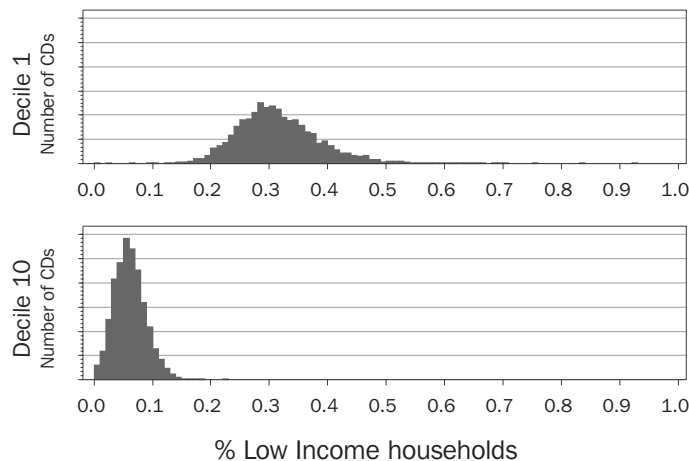
FIGURE 3.1 TYPICAL VARIABLE DISTRIBUTION, All CDs



Variable distributions and SEIFA scores

While figure 3.1 describes a variable's distribution for *all* CDs, we can also look at a variable's distribution for only those CDs with a low (or high) SEIFA score. For example, the distribution of the *Low Income* variable for CDs in Decile 1 (most disadvantaged) can be compared to the distribution of CDs in the Decile 10 (most advantaged).¹⁸ Figure 3.2 compares the distribution of *Low Income* variable values for CDs in Deciles 1 and 10 (for the Index of Relative Advantage and Disadvantage).¹⁹

FIGURE 3.2 LOW INCOME VARIABLE DISTRIBUTION, Decile 1 and Decile 10



There are two important points to note:

- The average value for CDs in Decile 1 is high compared to Decile 10. This means that, on average, CDs in the most disadvantaged IRSAD decile have a higher proportion of low income dwellings, than CDs in the most advantaged IRSAD decile.
- An individual CD can still have a value very different to this average. For example, it is possible for an individual CD in the most disadvantaged decile to have a *Low Income* value similar to CDs in the most advantaged decile.

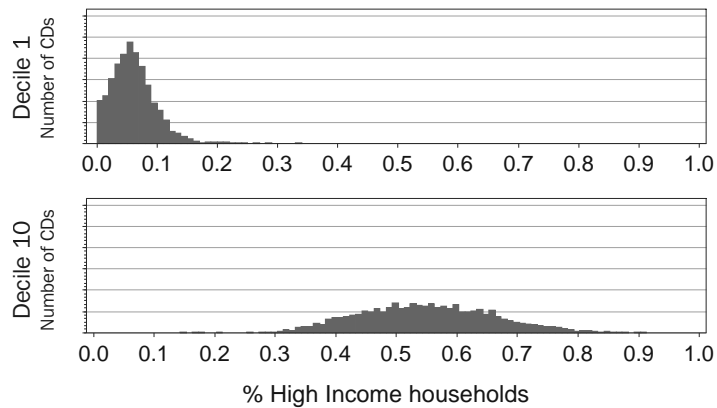
For comparison, figure 3.3 shows these top and bottom decile distributions for the *High Income* variable.

¹⁸ Each decile contains 10% of all CDs.

¹⁹ These are Variable distributions, which are not to be confused with the SEIFA index distributions shown in Chapter 4.

Variable distributions and SEIFA scores *continued*

FIGURE 3.3 HIGH INCOME VARIABLE DISTRIBUTION, Decile 1 and Decile 10



When comparing the LOW INCOME and HIGH INCOME distributions there are two more important points to note:

- The most disadvantaged decile CDs have a higher average value for the LOW INCOME variable than they do for the HIGH INCOME variable. This means that CDs with a relatively low SEIFA score will tend to have more low income dwellings than high income dwellings.
- SEIFA includes both LOW INCOME and HIGH INCOME variables because these are measuring different concepts.²⁰ For example, a CD with few low income dwellings may or may not have a high proportion of high income dwellings. Because a 'lack of low income dwellings' is different from 'many high income dwellings', both measures are included in SEIFA.

This section aimed to improve understanding of the relationship between the SEIFA index scores and the variables used to create them.

²⁰ Apart from the Disadvantage index, which includes only disadvantage variables.

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

INTRODUCTION

This chapter aims to assist the interpretation of the index scores, using distributions, examples and maps. This chapter also briefly introduces the geographic areas used in SEIFA.

- WHAT DO THE INDEXES LOOK LIKE?
- INTERPRETING SEIFA USING A COMPARISON OF TWO CDS
- INTERPRETING SEIFA USING MAPS
- INTRODUCTION TO GEOGRAPHY

WHAT DO THE INDEXES LOOK LIKE?

Terms: Scores, ranks, deciles and percentiles

SCORES: A score for a collection district (CD) is created by adding together the weighted characteristics of that CD. The scores for all CDs are then standardised to a distribution where the average equals 1000 and roughly two-thirds of the scores lie between 900 and 1100.²¹

This means that approximately 15% of CDs have a score lower than 900 with the remaining 85% of CDs having a score higher than about 900. Approximately 85% of CDs have a score lower than 1100 with the remaining 15% of CDs having a score higher than about 1100. The scores are used to rank the CDs, so care should be taken when comparing scores. For example, an area with a score of 500 is not twice as disadvantaged as an area with a score of 1000; it just had more markers of relative disadvantage.

RANKS: The CDs are ranked in order of their score, from lowest to highest.

DECILES: Decile 1 contains the bottom 10% of CDs, Decile 2 contains the next 10% of CDs and so on.

PERCENTILES: Percentile 1 contains the bottom 1% of CDs, Percentile 2 contains the next 1% of CDs and so on.

Index distributions and interpretation

Please refer to Chapter 2 for further discussion on the interpretation of each index.

Figures 4.1 - 4.4 show the distribution of CD scores for each index.²² Most of the CDs have scores somewhere in the middle, with a few areas having very high or very low scores. Because these are standardised scores, the mean score equals 1000 and approximately two-thirds of the scores lie between 900 and 1100.²³ Despite standardisation, the shape of the distribution and the minimum and maximum score are different for each index. These different distributions highlights the fact that there are many ways to capture relative socio-economic disadvantage.

Care must be taken when directly comparing individual CD scores, particularly where the scores are similar. It is more appropriate to compare deciles groups than scores of individual CDs. Figures 4.1 - 4.4 show the decile cut-offs marked along the top axis of the index distributions. However, for all four indexes, the middle deciles are close together. This means that care must be taken when comparing these middle deciles because the CDs in these deciles are not particularly disadvantaged or advantaged. The distributions of each index are described further below. For additional information on the interpretation of SEIFA scores refer to Chapter 2. For information on the use of SEIFA refer to Chapter 5.

21 Actually, around 74% of CDs lie between 900 and 1100 for the Index of Relative Disadvantage. Please refer to Standardisation in the Glossary for more information.

22 The horizontal axis shows the SEIFA score, and the vertical axis shows how many CDs have that score (37,457 CDs in total).

23 Refer to Standardisation in the Glossary.

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

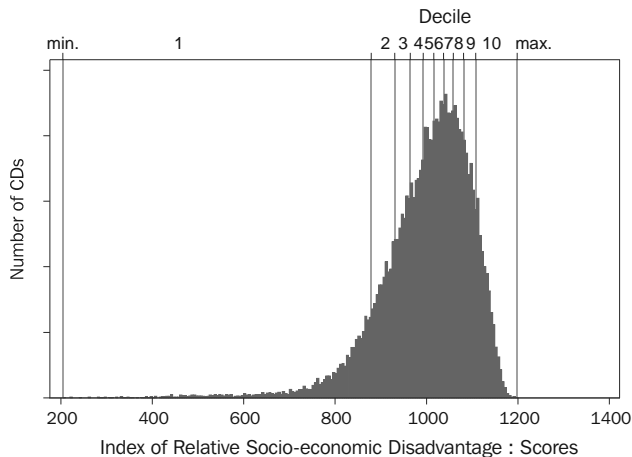
Index of Relative Socio-economic Disadvantage (IRSD) Distribution

Figure 4.1 shows the distribution of the Index of Relative Socio-economic Disadvantage, the index that includes *only* variables related to relative disadvantage (17 variables in total). The full list of variables appear in the Appendix, including:

- Proportion of low income households in the area;
- Proportion of people who do not speak English well;
- Proportion of households who pay low rent; and
- Proportion of people with no post-school qualifications.

This index is therefore appropriate for distinguishing between relatively disadvantaged areas. This is shown by the long left tail in the distribution. The decile cut-offs (marked along the top axis), show that the middle deciles are close together. It is not appropriate to use SEIFA indexes to distinguish between these deciles. This indicates that CDs with mid-range scores are neither particularly disadvantaged nor lacking disadvantage relative to other areas. This index is appropriate for users who are interested in the relative disadvantage of people in an area (lower deciles), and the relative lack of disadvantage of people in an area generally (upper deciles). Note that compared to the other indexes, the CDs in the upper deciles have more similar scores for this index. This is because these CDs lack indicators of disadvantage, and it will therefore be more difficult to distinguish between them using this index, compared to the other indexes.

FIGURE 4.1 IRSD SCORES HISTOGRAM



Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) Distribution

Figure 4.2 shows the distribution for the Index of Relative Socio-economic Advantage and Disadvantage. This index utilised 21 variables which are listed in the Appendix, including:

- Proportion of high income households in the area;
- Proportion of low income households in the area;
- Proportion of households with broadband internet connection; and
- Proportion of households without access to the internet.

The decile cut-offs again show that the middle scores are close together. However, the higher scores are more spread out than the Disadvantage index. Because this index also includes variables related to relative advantage, it is useful for distinguishing between advantaged areas (as well as disadvantaged areas). Therefore, this index is appropriate for users who are interested in relative advantage as well as disadvantage.

For the Disadvantage index, the higher deciles indicate *only* a relative lack of disadvantage. For the Index of Relative Advantage and Disadvantage, the higher deciles indicate relative advantage. A CD could have a high 'disadvantage' score (relative lack of disadvantage), but not necessarily a high 'advantage/disadvantage' score (relative advantage).

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

Index of Relative
Socio-economic Advantage and
Disadvantage (IRSAD)
Distribution *continued*

FIGURE 4.2 IRSAD SCORES HISTOGRAM

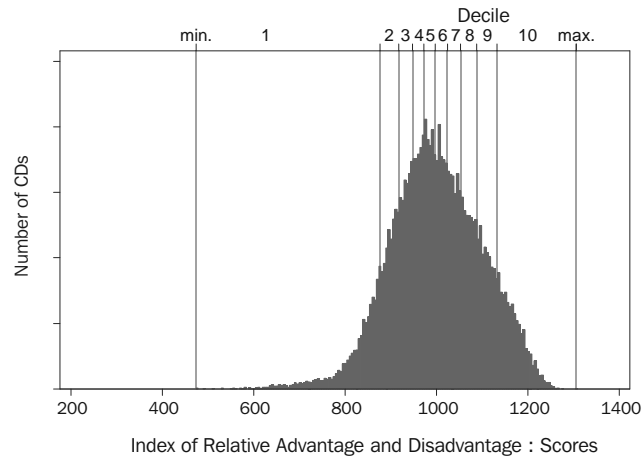


FIGURE 4.3 IER SCORES HISTOGRAM

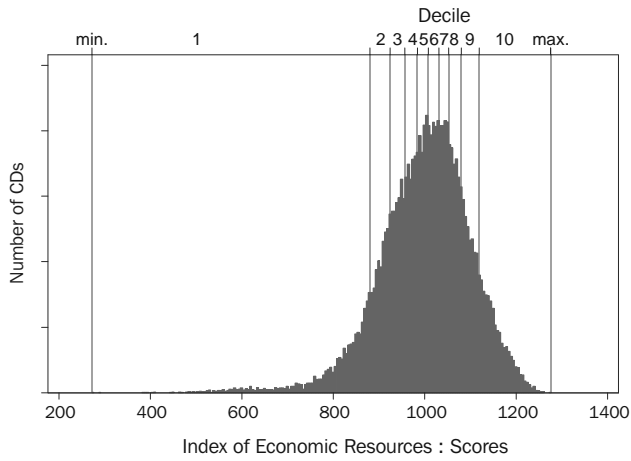
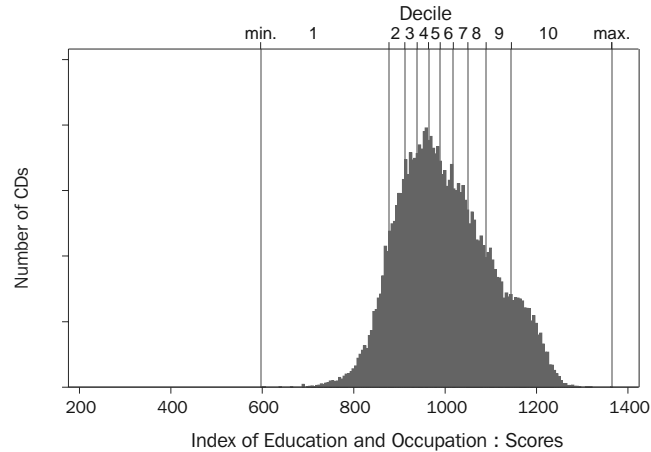


FIGURE 4.4 IEO SCORES HISTOGRAM



CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

Index of Economic Resources (IER) Distribution

Figure 4.3 shows the distribution for the Index of Economic Resources. This index also includes variables related to both relative disadvantage and advantage (15 in total). The variables are selected to represent 'access to economic resources'. These variables are listed in the Appendix, and include:

- Proportion of high income households in the area;
- Proportion of low income households in the area;
- Proportion of home owners in the area; and
- Proportion of unemployed people in the area.

This index is appropriate for users who are interested in comparing areas based on the overall access to economic resources of people in the area. While this index includes a broad range of information, it does not include information, such as superannuation, that is not measured by the Census.

Index of Education and Occupation (IEO) Distribution

Figure 4.4 shows the distribution for the Index of Education and Occupation. This index also includes variables related to both relative disadvantage and advantage (nine in total). The variables are selected to represent education and occupation. The education variables are designed to measure low or high levels of educational attainment of people in the area. The occupation variables are designed to measure people with jobs requiring low skills, few qualifications and have little control over one's own worktime; as well as people with jobs that tend to require high skills, high qualifications and are autonomous. These variables are listed in the Appendix, and include:

- Proportion of people with a diploma qualification;
- Proportion of people with no post-school qualifications;
- Proportion of people employed in a job classified in occupation-skill group 1 (such as managerial and professional jobs); and
- Proportion of people employed in a job classified in occupation-skill group 5 (such as some labouring and sales jobs).

This index is appropriate for users who are interested in comparing areas based on the qualifications and jobs of the people in the area.

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

INTERPRETING SEIFA USING A COMPARISON OF TWO CDS

Comparison This section aims to assist the interpretation of SEIFA scores, ranks and deciles by using two CDs located in Sydney as examples:

- A CD located in Manly (located at the Sydney Harbour entrance); and
- A CD located in Macquarie Fields (located on the south-west outskirts of Sydney).

It is important to note that there may be a diverse range of CDs within an area (as shown by the SLA map in figure 4.7). Similarly, there can be a diverse range of people and households within each CD. The CD selected for this example may not be representative of the area as a whole, or of an individual person living in the CD.

Figure 4.5 shows the scores, ranks and deciles for each of the four indexes for both the 'Manly' and 'Macquarie Fields' CDs.

FIGURE 4.5 COMPARISON OF SEIFA SCORES, RANKS, DECILES

	IRSD			IRSAD			IER			IEO		
	score	rank	decile	score	rank	decile	score	rank	decile	score	rank	decile
CD in Manly	1 116	34 598	10	1 180	36 254	10	1 094	31 519	9	1 196	36 272	10
CD in Macquarie Fields	1 000	16 116	5	989	17 345	5	996	16 865	5	970	15 858	5

The first thing to note is that each particular area has a different score for each index. It can be difficult to know at first glance whether a score is relatively high or low, especially since the range of scores is slightly different for each index (as shown by the distributions in figures 4.1 - 4.4).

CD in Manly For example, the CD in Manly has a different score for each index, although they do look roughly the same (IRSD 1116; IRSAD 1180; IER 1094; IEO 1196). All of these index scores are above 1100 (apart from IER), which means that for every index except IER, this CD has a higher score than at least 85% of the other CDs.²⁴ This is confirmed by the rank for each index, remembering that ranks range between 1 and 37,457 for CDs (IRSD 34,598; IRSAD 36,254; IER 31,519; IEO 36,272). This is also confirmed by the decile for each index (IRSD 10; IRSAD 10; IER 9; IEO 10). This means that this CD lies within the top 10% of CDs for IRSD, IRSAD and IEO; and in the top 20% of CDs for IER.

As a whole, the people in this area as a whole tend to lack disadvantage, are relatively advantaged, have relatively high access to economic resources and are relatively highly educated and skilled. An area will tend to have similar scores for every index, however this is not necessarily the case because each index is based on different information.

CD in Macquarie Fields The CD in Macquarie Fields also has roughly similar scores for each of the four indexes (IRSD 1000; IRSAD 989; IER 996; IEO 970). These scores are all at, or just below, the mean of 1000, which means they are in the mid range. This is confirmed by the ranks (IRSD 16,116; IRSAD 17,345; IER 16,865; IEO 15,858). This is also confirmed by the deciles (IRSD 5; IRSAD 5; IER 5; IEO 5).

As a whole, people in this area do not particularly lack disadvantage, are neither relatively advantaged or disadvantaged, have medium access to economic resources and have a mid level of education and skills relative to other areas. SEIFA is a general measure of relative socio-economic disadvantage that includes a wide range of variables. Therefore, care must be taken when comparing this CD to other CDs with scores in the mid range. As discussed in Chapter 5, areas may have similar scores, but for very different reasons due to the wide range of variables included in the index construction.

²⁴ Note that these proportions are approximate only. Please refer to Standardisation in the Glossary.

CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

CD in Macquarie Fields *continued*

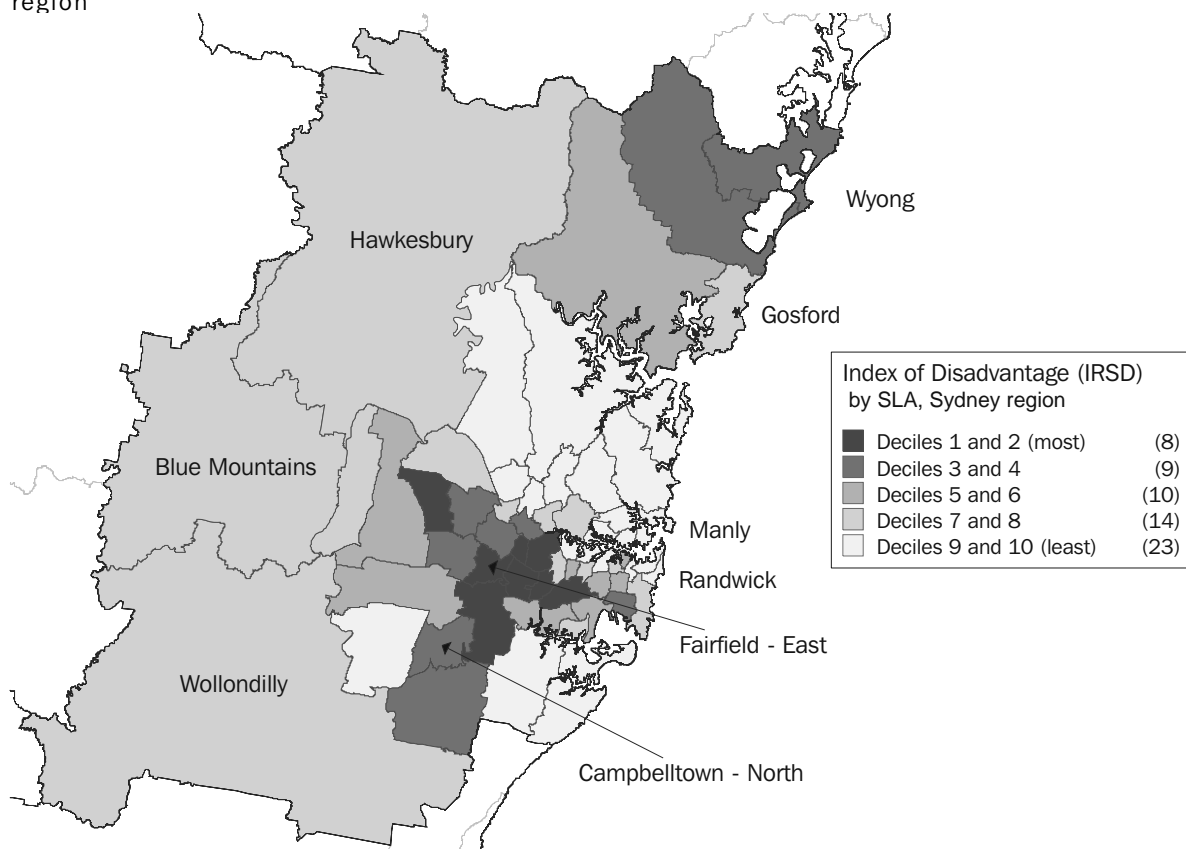
However, in this case, the individual CDs in this example can be compared because their scores are so different. The CD in Manly is generally more advantaged (and less disadvantaged) than the CD in Macquarie Fields, for each of the four indexes. Note that it is possible for a high (or low) income household to be located in either of these CDs.

INTERPRETING SEIFA USING MAPS

Maps Maps are useful as visual comparisons of areas for the SEIFA indexes. Figure 4.6 maps the Index of Relative Disadvantage for Statistical Local Areas (SLAs) within the Sydney region. This map shows SLAs with a score in the lower deciles (reflecting more disadvantage), such as LIVERPOOL EAST, as well as SLAs with a score in the highest deciles (reflecting less disadvantage), such as MANLY.

At the Australia-wide level, there are an equal number of SLAs in each decile. However, as shown on the legend, Sydney has many more SLAs in the highest deciles (less disadvantage) than in the lower deciles (more disadvantage).

FIGURE 4.6 INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE, SLA DECILES, Sydney region



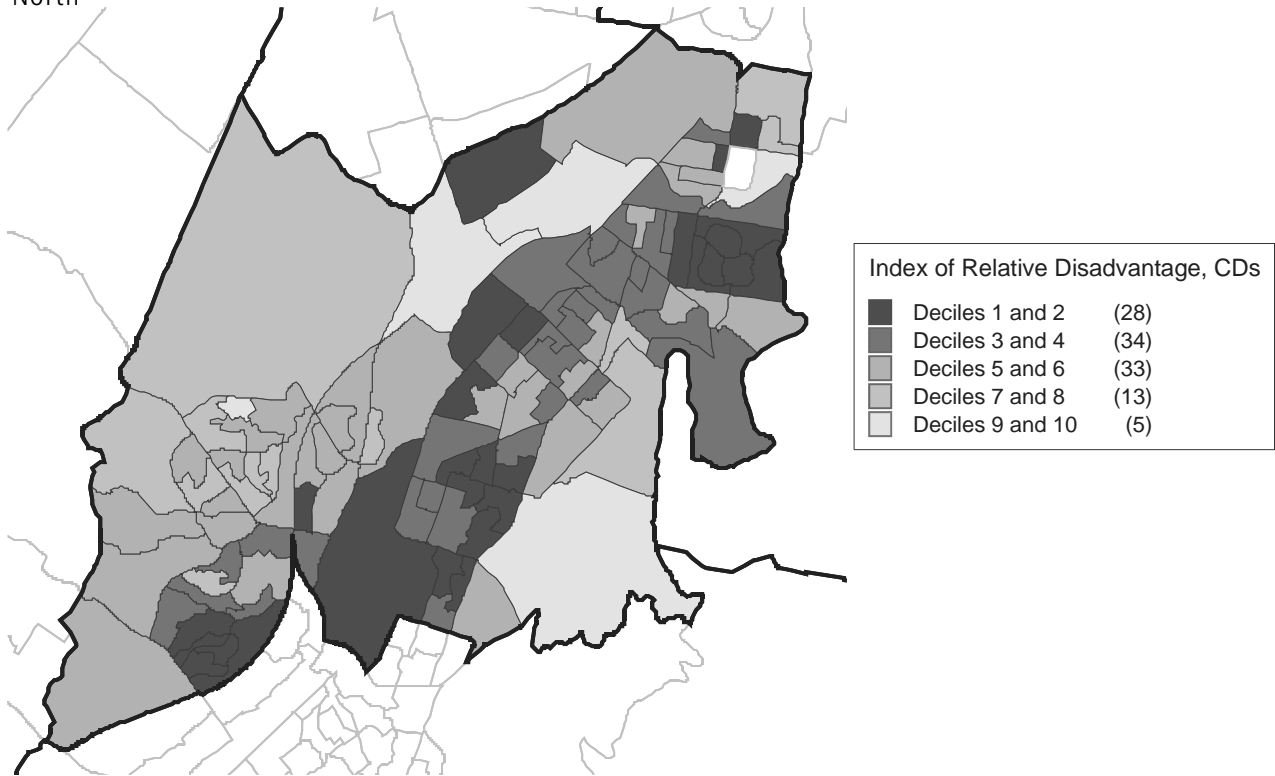
CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

Maps continued

While areas of relatively high or low disadvantage tend to be clustered together, it is important to note that there can be a lot of diversity within an area. This is highlighted by figure 4.7, which maps the Index of Relative Disadvantage for the collection districts within CAMPBELLTOWN NORTH STATISTICAL LOCAL AREA. The entire SLA has an IRSD decile of 3, however every CD decile is represented by the collection districts within this SLA. Because of this diversity, using the SEIFA indexes for CDs are generally more appropriate than for larger areas, such as SLAs. If using SEIFA indexes for larger areas, any diversity should be acknowledged.

FIGURE 4.7 INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE, CD DECILES, Campbelltown North



CHAPTER 4 HOW TO INTERPRET SEIFA SCORE DISTRIBUTIONS

continued

INTRODUCTION TO GEOGRAPHY

Geography SEIFA indexes are released for four levels of area. Collection Districts (CD), Statistical Local Areas (SLA) and Local Government Areas (LGA) are ASGC standard geographical areas. Postal Areas (POA) and State Suburbs (SSC) are not ASGC areas, however are commonly used geographical areas. These geographies are explained further in the Glossary. Refer also to Geographies in the Glossary.

Creation of SEIFA for Areas larger than a CD As discussed in Chapter 2, the SEIFA index scores were first created at a CD level. The index scores for other geographic levels (SLA, LGA, POA, SSC), were constructed using the scores of the CDs within these areas.²⁵ Even though a CD is the smallest type of area currently available, people and households within a CD can be quite different to each other. While SEIFA represents an *average* of all these different people, SEIFA does not represent the *individual* situation of each person. Larger areas are more likely to have a greater diversity of people and households. While SEIFA indexes are created for larger areas, the best use of SEIFA will be achieved at the CD level. Where SEIFA is required for analysing a larger area (such as SLA), we recommend carefully examining the distribution of CDs within these areas.

Care must be taken when interpreting the index scores at these other geographic levels. Refer to Population Weighting and Geographies in the Glossary for more information.

Further information on geographies Refer to the following references for further information on geographies.

- Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 : Chapter 1, Introduction (ABS cat. no. 2905.0)
- Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 : Chapter 4, Postal Areas (ABS cat. no. 2905.0)
- Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 : Chapter 5, State Suburbs (ABS cat. no. 2905.0)
- Census Dictionary, 2006 (Reissue) : Glossary, Collection District (CD) (ABS cat. no. 2901.0)
- Census Dictionary, 2006 (Reissue) : Glossary, Postal Areas (POA)
- Census Dictionary, 2006 (Reissue) : Glossary, State Suburbs (SSC)

²⁵ Indexes for these areas were constructed by taking the population-weighted mean of the CD score within the area, as explained further in the Glossary and the Technical Paper.

CHAPTER 5 HOW TO USE SEIFA

INTRODUCTION

This chapter brings together some of the major issues surrounding the use of SEIFA that were discussed in previous chapters. This chapter is intended to briefly provide more detail about the use of SEIFA, but is not intended to replace a proper technical understanding of analysis.

- HOW TO USE SEIFA
- EXAMPLES OF USE OF SEIFA

HOW TO USE SEIFA

It is important to have a good understanding of SEIFA before using it in an analysis.

SEIFA as a general tool

The concept of relative socio-economic disadvantage is neither simple nor well-defined. No single measure is able to completely capture this concept. As discussed previously, SEIFA uses the commonality between a wide range of information to represent relative socio-economic disadvantage. Therefore, the concept and the selection of information used to make the indexes are broad. This means that SEIFA is a *general* measure, which will therefore impact how SEIFA should be used in analysis.

First, because SEIFA is only a broad measure of relative socio-economic disadvantage, a given analysis is likely to require additional information that is specific to the topic. For example, a study analysing the relationship between smoking and disadvantage may consider also including other information, such as the availability of doctors in the area, or the number of smokers in the household. Neither of these measures are included in SEIFA, but may be important to the particular analysis.

Second, two areas may have a similar score, but for very different reasons. For example, a low score in one area may be due to a high proportion of low income households, while a low score in a different area may instead be due to a high proportion of low-rent households.²⁶ Although both areas have a similar SEIFA score, only one of these variables may be important to the topic you are analysing. For example, you may wish to distinguish between two areas based on low income, but not necessarily on low-rent. It is therefore important to note that SEIFA is a general tool that is not specifically designed for any particular analysis.

Impact of PCA on SEIFA

The method used to construct the SEIFA indexes is called Principal Components Analysis. As discussed in Chapter 3, this method will impact on how SEIFA should be used in your analysis.

First, the index scores can be used to *rank* areas in terms of disadvantage, however other arithmetic relationships may not be meaningful. For example, an area with a score of 500 is *not* half as advantaged as an area with a score of 1000. Similarly, it is incorrect to use the size of the gap between the scores to compare levels of disadvantage. For example, the difference in disadvantage between two areas with scores of 500 and 600, is not the same as the difference between two areas with scores of 800 and 900.²⁷

Second, PCA creates indexes by combining information (variables) related to relative disadvantage. PCA uses the relationships between the variables themselves to determine the importance (weight) of each variable to the index. This means that every variable has a different level of importance in the SEIFA indexes, and this may be different to how important the variable is to your particular analysis.

Third, it is important to be aware of the variables used to construct the SEIFA indexes. Because the SEIFA indexes are summary measures of a wide range of variables, SEIFA itself will, by definition, be related to each of the variables separately. This means that your analysis is likely to find a relationship between SEIFA and any one of these variables.

²⁶ This example is overly simplistic. In reality scores are determined by a combination of information.

²⁷ This issue also impacts on the assumptions of regression analysis. If you are conducting regression analysis, ensure you are aware of the distribution of the scores (Chapter 4).

Impact of PCA on SEIFA
continued

It will be difficult to tell how much of this relationship is due simply to the inclusion of the variable in the construction of SEIFA. Although, SEIFA does only capture a part of the information contained in the variables. Nevertheless, the use of any of these variables alongside SEIFA in your analysis must be done with caution. The Appendix and the Technical Paper include a list of the variables used in each of the indexes.

Comparing SEIFA over time

As discussed in Chapter 2, comparing SEIFA scores over time is not recommended for a number of reasons.

- SEIFA is only a relative measure, not an absolute measure of socio-economic disadvantage. For example, *all* of the areas could have become less disadvantaged since the last release, however this would not be evident in SEIFA. An area may have a lower score than it did previously, however this could be due to changes in the other areas, rather than any change to that area.
- SEIFA is a snapshot of an area, with a five year gap between each release (every Census). Because the world changes between releases, the relationships between the variables will change. As these relationships are used to create SEIFA, the indexes will therefore change with each release.
- While consistency across SEIFA releases is very important, changes are made where necessary or important. In effect, this means that different releases of SEIFA capture a slightly different aspect of relative socio-economic disadvantage. For example, previous SEIFA indexes did not measure broadband internet access because this technology was not widely available. This information is now collected in the Census and is included in the Index of Relative Advantage and Disadvantage.
- Boundaries of geographic areas change between Censuses. There are good reasons for these changes and some changes can be significant.

For these reasons, we do not recommend comparing SEIFA over time. If you must do so, compare the extreme scores (such as the top and bottom deciles) rather than the mid-range scores. This is because, as shown by the index distributions in Chapter 4, the mid-range scores are similar, so any comparison over time for areas with these scores should be done with caution. Be aware how the above issues affect your analysis.

SEIFA is an area level measure

As discussed in Chapter 2, an area can have a diverse range of people. SEIFA scores represent an *average* for all people in an area. This affects the interpretation of SEIFA in your analysis. For example, it would be appropriate for an analysis of the relationship between SEIFA and people's health to refer to the association between a person's health status and the general level of relative socio-economic disadvantage of the people in their *area*. It would not be correct to discuss the relative disadvantage of the *person*, only of the general level of relative disadvantage of people and households in the area.

EXAMPLES OF USE OF SEIFA

Examples

For most analysis, we recommend the use of the SEIFA deciles because the SEIFA scores require a certain amount of technical knowledge for proper use. The actual SEIFA scores should only be used for more technical analysis. As shown below, SEIFA can be used in many ways, however each method has limitations.

Distributions for State, SD and SSD geographic areas

In 2006, SEIFA indexes have not been created for State, Statistical Division (SD) and Statistical Subdivision (SSD) geographic areas. However, for 2006 SEIFA, a spreadsheet tool is available that shows the distribution of SEIFA scores within these areas. This spreadsheet shows the distribution of people who are usual residents on Census Night, rather than the distribution of CDs.²⁸ For each index, a particular geographic area can be selected, and the tool will show the population distribution by CD-level scores for this area. This distribution can be compared to the distribution for Australia. The table output shows the number and proportion of usual residents within the CDs. The graph output shows the proportion of people in the area, by the score of the CD in which they

²⁸ The distribution of people is different to the distribution of CDs because there is a different number of people living in each CD.

Distributions for State, SD and SSD geographic areas *continued*

usually resided on Census Night. The colours used on the graph output roughly approximate the top and bottom CD deciles.

This spreadsheet tool is available on the ABS website with the SEIFA indexes under cat. no. 2033.0.55.001 Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only, 2006.

Figures 5.1 and 5.2 show the distribution of people by the score of the CD in which they usually resided on Census Night, for the Index of Relative Advantage and Disadvantage, within AUSTRALIA and within SYDNEY NORTHERN BEACHES SSD respectively. Note that figure 5.2 shows the distribution of people living in a CD with a certain score, *as a proportion of all the people usually residing in the SYDNEY NORTHERN BEACHES SSD on Census Night.*²⁹ This means that we can directly compare this population distribution with that of AUSTRALIA. Figure 5.2 shows that most of the usual residents of SYDNEY NORTHERN BEACHES SSD resided in CDs with high IRSAD scores, compared to the Australian distribution.

FIGURE 5.1 POPULATION DISTRIBUTION, BY IRSAD CD SCORES, AUSTRALIA
% people

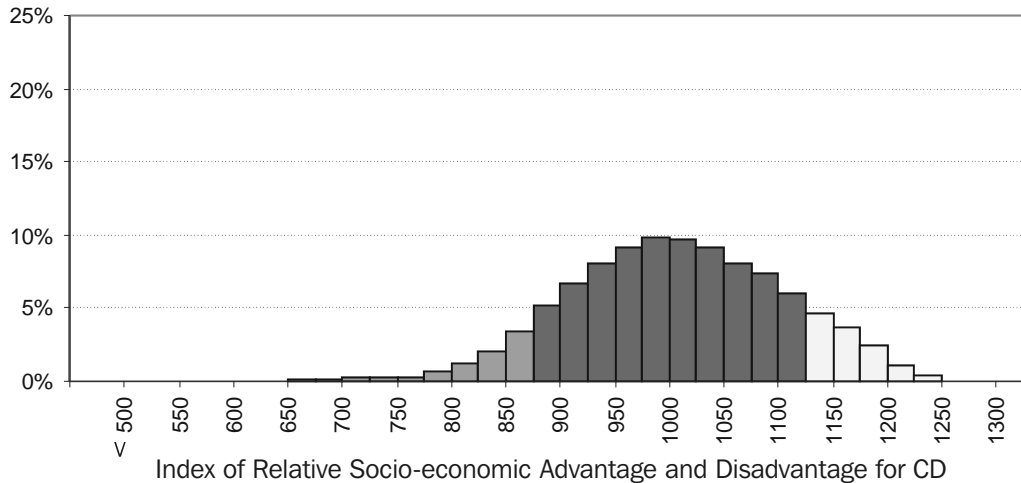
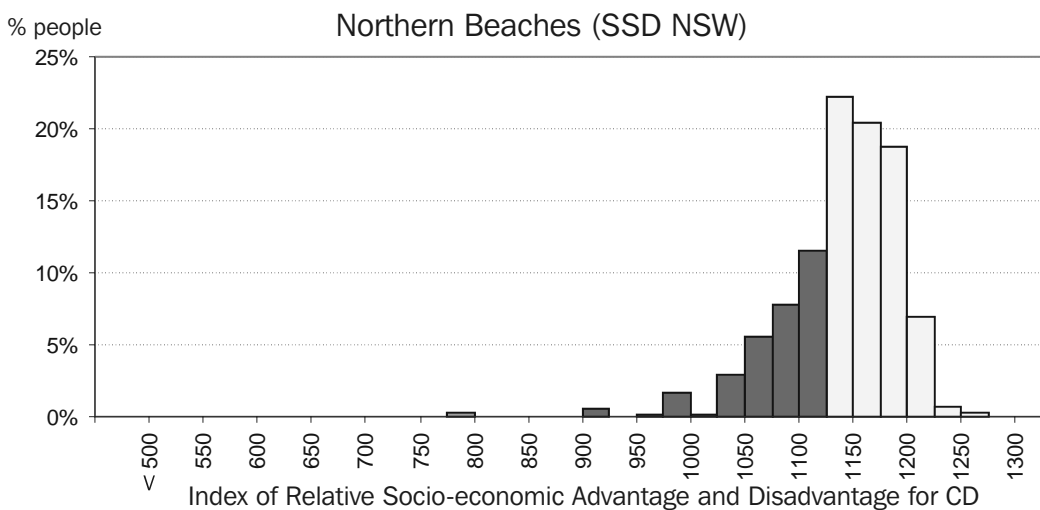


FIGURE 5.2 POPULATION DISTRIBUTION, BY IRSAD CD SCORES, SYDNEY NORTHERN BEACHES
% people



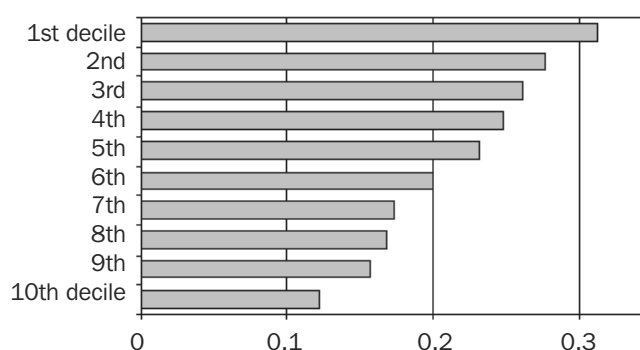
²⁹ Excluded CDs were not given a score. However, people in excluded CDs were counted in the denominator when calculating the proportions.

Descriptive analysis

This section provides a descriptive analysis of the association between SEIFA Index of Relative Socio-economic Disadvantage and health. A person's lifestyle choices, including diet, exercise and smoking habits can be related to the area where they live. For example, diet may be affected by the cost and availability of fruit and vegetables in local shops, or the location of fast-food outlets. Exercise may be affected by the availability of sports and recreation facilities in the area, or by safety concerns. There has been substantial research into the effects of neighbourhood disadvantage on the health of individual people. As measures of neighbourhood socio-economic status, the SEIFA indexes can be used to look at the relationship between neighbourhood disadvantage and health.

Figure 5.3 shows the proportion of adults who are daily smokers (using National Health Survey 2004-5, unpublished data), by SEIFA IRSD deciles. This figure indicates a relationship between smoking and SEIFA.

FIGURE 5.3 PROPORTION OF ADULTS WHO SMOKE DAILY, BY IRSD DECILES



There are a number of ways this figure can be interpreted, including:

- Compared to other areas, the proportion of adults who smoke daily is greater in areas with relatively more socio-economic disadvantage.
- On average, adults are more likely to report they smoke daily if they live in an area that generally has greater relative socio-economic disadvantage.
- The proportion of adults in the lowest decile (greatest disadvantage) who smoke daily is almost three times the proportion of adults in the highest decile (least disadvantaged) who smoke daily. Around 32% of adults who lived in areas with the lowest IRSD decile smoke daily, compared with around 12% of adults who lived in areas with the highest IRSD decile.

Figure 5.4 shows self-assessed general health (using National Health Survey 2004-5, unpublished data), by IRSD deciles.

FIGURE 5.4 PROPORTION OF ADULTS WITH SELF-ASSESSED HEALTH, BY IRSD DECILES

IRSD deciles	SELF-ASSESSED HEALTH	
	Fair or Poor	Good to Excellent
Deciles 1 and 2	23	77
Deciles 3 and 4	20	80
Deciles 5 and 6	15	85
Deciles 7 and 8	13	87
Deciles 9 and 10	11	89
Total	16	84

Descriptive analysis *continued*

This figure indicates a relationship between self-assessed health and SEIFA, which can be interpreted as:

- Compared to other areas, there are higher rates of self-assessed 'fair to poor' health (or lower rates of self-assessed 'good to excellent' health) in areas with generally more socio-economic disadvantage.
- On average, adults are more likely to report 'fair to poor' health (self-assessed) if they live in an area that generally has greater relative socio-economic disadvantage.
- The proportion of adults in the lowest deciles with 'fair to poor' health is twice the proportion of adults in the highest deciles. Around 11% of adults living in areas that generally lack relative socio-economic disadvantage self-assessed their health to be 'fair or poor', compared with 23% of adults living in areas with a generally high level of relative socio-economic disadvantage.

**Regression analysis and SEIFA:
For more technical use**

This section is not intended to describe regression analysis, but to highlight some issues when using SEIFA in regression analysis. There are alternative ways that SEIFA deciles can be used in regression analysis. For example, dummy variables can be created for each SEIFA decile, or just for the extreme deciles.

If using SEIFA in linear regression analysis, particular care must be taken, as such analysis is based on a certain set of statistical assumptions. These assumptions include the linearity assumption, for example, which assumes constant relationships between the analysed variables.³⁰ As an example, if SEIFA were used to analyse people's health, then the results of a regression analysis would be based on the assumption that a unit increase in SEIFA is related to a fixed amount of change in people's health. That is, the relationship between a change in health and a change in SEIFA remains the same, whatever the SEIFA score. However, the difference in disadvantage between SEIFA scores of 500 and 600 is not equal to the difference between scores of 900 and 1000. Therefore, this assumption will not hold and care should be used when applying SEIFA to regression analysis. Users should be aware of the distribution of index scores before using SEIFA in regression analysis.

**Hints when using SEIFA to
analyse survey data**

As discussed in Chapter 2, there are a few issues to consider when using SEIFA to analyse survey data.

- SEIFA refers to the general population of an *area*. Therefore, the individual households or people in the survey data must be able to be linked to an area. If each person-level record on the survey file has a Collection District number (CD), then the person's information (e.g. health) can be matched to the SEIFA score (or rank, or decile) using this CD number.
- It is important to select the SEIFA index that matches the geographic areas used in the survey. For example, if the survey refers to *Australia Post* postcodes, then use the SEIFA indexes for Postal Areas.³¹
- Because these geography standards change over time, it is important to use the SEIFA release that most closely corresponds to that used in the survey. Different releases of geography standards are able to be matched using the concordances available on the ABS website (refer to Further Information below). For example, if the survey was conducted in 2003, then it may have used a previous geography standard. This means that it may be best to use the 2001 release of SEIFA, so that the survey data and SEIFA will be referring to the same areas. Note that it is possible to use the 2006 release of SEIFA, as long as you analyse only those CDs that have not changed since the geographic standard used by the survey. However, it should be noted that using only part of the survey will impact the survey weights.³²

³⁰ This assumption also applies to transformed variables.

³¹ Refer to the geography discussion in Chapter 4.

³² For every survey, weights are created so that each person in the survey represents multiple people in the actual population. If only part of the sample is used in an analysis, these weights may no longer accurately represent the population.

Hints when using SEIFA to analyse survey data *continued*

- Surveys sample only a proportion of the population and may miss the people or households that are important to your particular analysis. For example, surveys are often able to sample relatively few people and households from very remote areas. In addition, not everyone will respond to survey questionnaires. This means that some information on the survey could be biased if certain groups of people tended not to respond. Similarly, while the Census aims to capture information about the entire population of Australia, not every person responds to every Census question. Some information from the Census could be biased if certain groups of people are less likely to be measured. For example, household income is not able to be calculated for households with an adult who was elsewhere on Census Night. This group of households may be important to your particular analysis.

FURTHER INFORMATION

References

Refer to the following papers for other examples of using SEIFA.

- SEIFA Information Paper, 2001 (ABS cat. no. 2039.0)
- SEIFA: Introduction, Use and Future Directions (Research Paper, Sep 2006, ABS cat. no. 1351.0.55.015)

CHAPTER 6 ENSURING SEIFA MEASURES DISADVANTAGE

CHAPTER 6 ENSURING SEIFA MEASURES DISADVANTAGE

INTRODUCTION

This chapter briefly discusses some of the methods used to validate SEIFA in order to ensure SEIFA measures relative socio-economic advantage and disadvantage.

CHECKING SEIFA

How well do the SEIFA indexes reflect relative disadvantage or advantage in towns and neighbourhoods across Australia? The SEIFA indexes are checked in a number of ways before release to ensure that the indexes reflect relative socio-economic disadvantage or advantage.

Checking the indexes

The first part of the checking process is to examine the construction of each SEIFA index and the indexes themselves. Each index includes a wide range of variables that are deemed to be related to the concept of relative socio-economic disadvantage of that index. A variable is considered for inclusion if it is conceptually sound, however, a variable is only included after further checks have been carried out. This involves checking if the variable weights make intuitive sense, and analysing the relationship between the variable with all other variables. If the variable adds little value to the index, or if the variable appears not to measure relative socio-economic disadvantage, then it is not included in the index construction. As discussed in Chapter 3, a variable is also not included if it is deemed to be too similar to other variables, as this could bias the index.³³ A group of academics and expert users also provided suggestions on conceptual and index construction issues, as well as recommendations for future SEIFA indexes.

The final indexes were checked using other data sources, including survey information on health, income, net worth as well as measures of remoteness. One example is briefly discussed in the 'SEIFA and remoteness: An example' section below. We also consider how similar the 2006 indexes are to previous SEIFA indexes.

Checking the scores

The next step is to check the actual index scores. Because it is not feasible to investigate each score individually (there are 37,457 scores for each of the four CD indexes, plus the index scores for the larger geographic areas), we instead look closely at areas which:

- have the highest and lowest SEIFA scores in each state or territory;
- have a substantially different rank to 2001; or
- have substantially different ranks across the four indexes.

Where particular CDs were identified through this process, we examined the Census variables to see why these CDs might stand out. We also used local knowledge to check the highest and lowest SEIFA scores.

SEIFA and remoteness: An example

The SEIFA indexes are checked using different data sources, such as measures of remoteness. We expect relative socio-economic disadvantage to be related to remoteness for various reasons, including a reduced ability to participate in society in very remote areas. We can also use survey data to analyse the relationship between remoteness and relevant information, and then use this analysis to check the relationship between remoteness and the SEIFA indexes. For this example, we compare the remoteness distribution of the Index of Economic Resources with the remoteness distribution of household income.³⁴ While the Index of Economic Resources captures more than just income information, we would expect similarities in their remoteness distributions.

Figure 6.1 shows the relationship between the Index of Economic Resources (IER) scores and the remoteness measure *Accessibility/Remoteness Index of Australia (ARIA)*.³⁵ The 25th, 50th and 75th percentile scores are shown.³⁶ The figure shows that MAJOR CITY IER scores tend to be higher than the other regions; and that REMOTE CD

³³ A list of variables is included in the Appendix.

³⁴ Net worth distribution was also checked. Refer to the Glossary for Net Worth definition.

³⁵ Maximum and minimum CD scores not shown for clarity.

³⁶ These percentiles indicate the upper limit of scores of the lowest 25%, 50% and 75% of CDs respectively.

CHAPTER 6 ENSURING SEIFA MEASURES DISADVANTAGE

continued

SEIFA and remoteness: An example *continued*

scores also tend to be slightly higher than OUTER REGIONAL area scores. Also, the VERY REMOTE CD scores are typically much lower than CD scores in the other regions.

For comparison, figure 6.2 shows the relationship between remoteness with household income, using the *Survey of Income and Housing* (2005-6, unpublished data).³⁷ Note that there are very few VERY REMOTE households actually surveyed, so these households are not shown on the figure. This figure shows a similar remoteness distribution for income. Again, household income tends to be slightly higher in MAJOR CITY households than other areas; and are again higher in REMOTE households than OUTER REGIONAL households. Although not shown, similar analysis was undertaken for various income measures and household net worth.

FIGURE 6.1 IER SCORES BY ARIA

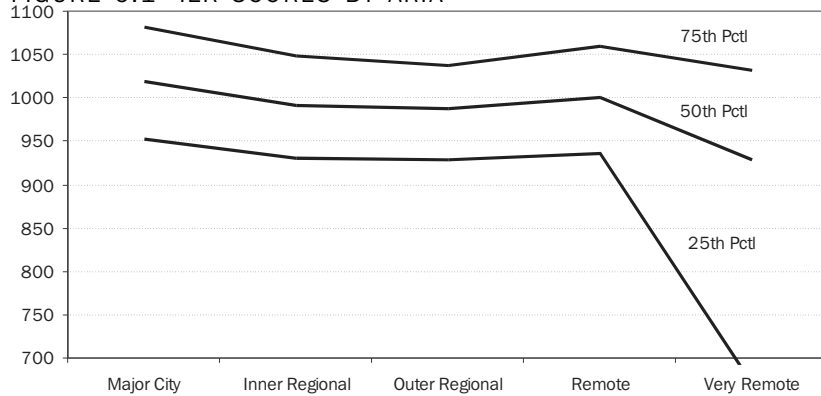
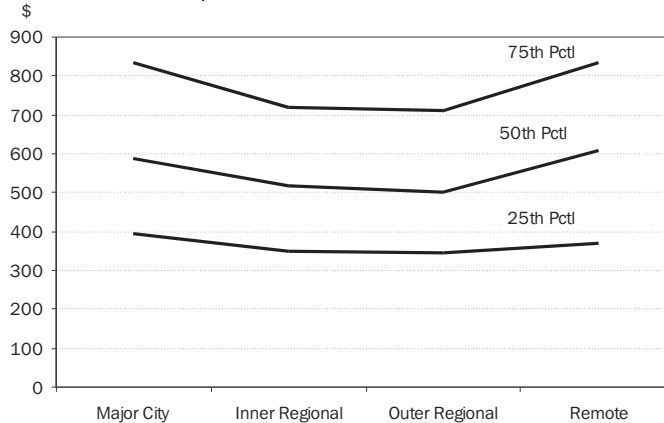


FIGURE 6.2 EQUIVALISED DISPOSABLE INCOME BY ARIA



³⁷ Maximum and minimum household scores not shown for clarity.

CHAPTER 7 FURTHER INFORMATION

CHAPTER 7 FURTHER INFORMATION

SEIFA DOCUMENTATION

Links to further SEIFA information:

- Socio-Economic Indexes for Areas (SEIFA), Australia - *Data only, 2006* (ABS cat. no. 2033.0.55.001)
- Socio-Economic Indexes for Areas (SEIFA): *Technical Paper 2006* (ABS cat. no. 2039.0.55.001)
- Socio-Economic Indexes for Areas (SEIFA): Information Paper, 2001 (ABS cat. no. 2039.0)
- *SEIFA: Introduction, Use and Future Directions* (Research Paper, Sep 2006, ABS cat. no. 1351.0.55.015)

GEOGRAPHY DOCUMENTATION

Links to further ABS Geography information:

- *Statistical Geography Volume 1* - Australian Standard Geographical Classification (ASGC), Jul 2006 (ABS cat. no. 1216.0)
- *Statistical Geography: Volume 2* - Census Geographic Areas, Australia, 2006 (ABS cat. no. 2905.0)
- Australian Standard Geographical Classification (ASGC), Jul 2007 (ABS cat. no. 1216.0)
- Australian Standard Geographical Classification (ASGC) *Concordances*, 01 Jul 2006 (ABS cat. no. 1216.0.15.002)
- *ABS Postal Area Concordances*, Aug 2006 (ABS cat. no. 2905.0.55.001)
- *Statistical Geography - Australian Standard Geographical Classification (ASGC), Digital Boundaries*, 2006 (ABS cat. no. 1259.0.30.002)
- *Census of Population and Housing: Census Geographic Area Digital Boundaries*, Australia, 2006 (ABS cat. no. 2923.0.30.001)
- *Census Dictionary, 2006 (Reissue) : Glossary, Collection District (CD)* (ABS cat. no. 2901.0)
- *Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 : Chapter 5, State Suburbs* (ABS cat. no. 2905.0)
- *Census Dictionary, 2006 (Reissue) : Glossary, State Suburbs (SSC)* (ABS cat. no. 2901.0)
- *Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 : Chapter 4, Postal Areas* (ABS cat. no. 2905.0)
- *Census Dictionary, 2006 (Reissue) : Glossary, Postal Areas (POA)* (ABS cat. no. 2901.0)

CENSUS DOCUMENTATION

Links to further Census information:

- *Census Dictionary, 2006 (Reissue)* (ABS cat. no. 2901.0)
- *2006 Census Non-Response Rates Fact Sheets, 2006* (ABS cat. no. 2914.0.55.001)
- *2006 Census Data by Location*
- *2006 Census Data*
- *2006 Census: New Product Briefs*
- *CData Online: Product Brief* (ABS cat. no. 2064.0)
- *Basic Community Profile Datapack, 2006* (ABS cat. no. 2069.0.30.001)
- *2006 Census Data Packs at a Glance, 2007 to 2012* (ABS cat. no. 2069.0)
- *2006 Census Data Packs - Product Description*
- *Census Contacts : Client Services*

APPENDIX: LIST OF SEIFA VARIABLES

APPENDIX - LIST OF SEIFA VARIABLES

INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE

Included Variables

- % Occupied private dwellings with *no internet connection*
- % Employed people classified as *Labourers*
- % People aged 15 years and over with *no post-school qualifications*
- % People with stated annual household equivalised *income between \$13,000 and \$20,799* (approx. 2nd and 3rd deciles)
- % Households *renting from Government or Community organisation*
- % People (in the labour force) *unemployed*
- % *One parent families* with dependent offspring only
- % Households paying *rent less than \$120* per week (excluding \$0 per week)
- % People aged under 70 who have a long-term health condition or disability and *need assistance* with core activities
- % Occupied private dwellings with *no car*
- % People who identified themselves as being of *Aboriginal and/or Torres Strait Islander origin*
- % Occupied private dwellings *requiring one or more extra bedrooms* (based on Canadian National Occupancy Standard)
- % People aged 15 years and over who are *separated or divorced*
- % Employed people classified as *Machinery Operators and Drivers*
- % People aged 15 years and over who *did not go to school*
- % Employed people classified as *Low Skill Community and Personal Service Workers*
- % People who *do not speak English well*

Variables Dropped

- % Employed people classified as *Low Skill Clerical and Administrative Workers*
- % Employed people classified as *Low Skill Sales Workers*
- % Occupied private dwellings with *one or no bedrooms*
- % People aged 15 years and over who *left school at Year 11 or lower*

INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE AND DISADVANTAGE

Included Variables

- % People aged 15 years and over with *no post-school qualifications*
- % Occupied private dwellings with *no internet connection*
- % People with stated annual household equivalised *income between \$13,000 and \$20,799* (approx. 2nd and 3rd deciles)
- % Employed people classified as *Labourers*
- % Households paying *rent less than \$120* per week (excluding \$0 per week)
- % People aged under 70 who have a long-term health condition or disability and *need assistance* with core activities
- % Employed people classified as *Machinery Operators and Drivers*
- % People (in the labour force) *unemployed*
- % *One parent families* with dependent offspring only
- % Households *renting from Government or Community organisation*
- % Employed people classified as *Low Skill Community and Personal Service Workers*
- % Occupied private dwellings *requiring one or more extra bedrooms* (based on Canadian National Occupancy Standard)
- % Occupied private dwellings with *no car*
- % Occupied private dwellings with *four or more bedrooms*
- % People aged 15 years and over *at university* or other tertiary institution
- % Households paying *mortgage greater than \$2,120* per month
- % Households paying *rent greater than \$290* per week
- % People aged 15 years and over with an *advanced diploma or diploma qualification*

APPENDIX: LIST OF SEIFA VARIABLES *continued*

- % Employed people classified as *Professionals*
- % Occupied private dwellings with a *broadband* internet connection
- % People with stated annual household equivalised *income greater than \$52,000* (approx 9th and 10th deciles)

Variables Dropped

- % Employed people classified as *Low Skill Sales Workers*
- % Households *owning dwelling* they occupy (without a mortgage)
- % People who *do not speak English* well
- % Occupied private dwellings with *three or more cars*
- % Occupied private dwellings with *one or more bedrooms spare* (based on Canadian National Occupancy Standard)
- % Employed people classified as *Managers*
- % People aged 15 years and over with *degree* or higher
- % People aged 15 years and over who *left school at Year 11 or lower*

INDEX OF ECONOMIC RESOURCES

Included Variables

- % People with stated annual household equivalised *income between \$13,000 and \$20,799* (approx. 2nd and 3rd deciles)
- % *One parent families* with dependent offspring only
- % Occupied private dwellings with *no car*
- % Households *renting from Government or Community organisation*
- % Households paying *rent less than \$120* per week (excluding \$0 per week)
- % People aged 15 years and over who are *unemployed*
- % Households who are *lone person households*
- % Occupied private dwellings *requiring one or more extra bedrooms* (based on Canadian National Occupancy Standard)
- % Households *owning dwelling* they occupy (without a mortgage)
- % Dwellings with at least one person who is an *owner of an unincorporated enterprise*
- % Households paying *mortgage greater than \$2,120* per month
- % Households owning dwelling (with a *mortgage*)
- % Households paying *rent greater than \$290* per week
- % People with stated annual household equivalised *income greater than \$52,000* (approx 9th and 10th deciles)
- % Occupied private dwellings with *four or more bedrooms*

Variables Dropped

- % Occupied private dwellings that are *improvised dwellings*
- % Households who are *group households*
- % Occupied private dwellings with *four or more cars*
- % Occupied private dwellings with *one or more bedrooms spare* (based on Canadian National Occupancy Standard)
- % Occupied private dwellings with *one or no bedrooms*

INDEX OF EDUCATION AND OCCUPATION

Included Variables

- % People aged 15 years and over who *left school at Year 11 or lower*
- % People aged 15 years and over with *no post-school* qualifications
- % Employed people who work in a *Skill Level 5* occupation
- % Employed people who work in a *Skill Level 4* occupation
- % People (in the labour force) *unemployed*
- % People aged 15 years and over with a *certificate* qualification
- % People aged 15 years and over *at university* or other tertiary institution
- % People aged 15 years and over with an *advanced diploma or diploma* qualification
- % Employed people who work in a *Skill Level 1* occupation

APPENDIX: LIST OF SEIFA VARIABLES *continued*

Variables Dropped

- % People aged 15 years and over who are *still attending secondary school*
- % Employed people who work in a *Skill Level 2* occupation
- % People aged 15 years and over who *did not go to school*
- % People aged 15 years and over with *degree* or higher

GLOSSARY

Introduction	This Glossary is intended to briefly provide more detail on some of the terms and concepts used throughout this paper, but is not intended to replace a proper technical understanding of these concepts.
AGE-STANDARDISATION	<p>Groups of people with different age structures tend to have different characteristics. For example, people in their 30s are likely to have a higher level of education than people in their 70s. Therefore, a neighbourhood with many people in their 30s is more likely to have a higher level of education in general than a neighbourhood with many people in their 70s. In SEIFA, a high level of education is considered to be relatively advantageous. This means that SEIFA ranks the younger neighbourhood as more advantaged, because of factors related to the different age profiles of the areas.</p> <p>Age-standardisation could have been used to directly compare neighbourhoods with different age-profiles. However, age-standardisation was generally not used in SEIFA because the decision of whether to use age-standardisation depends on the type of using SEIFA. For different analyses, the link between age and education may or may not matter. For example, this link may not matter if your analysis considers education to be relatively advantageous irrespective of age. In this case, it would not be appropriate to age-standardise the education variable.</p> <p>In SEIFA, adjusting for age was only undertaken for the variable measuring 'need for assistance'. This variable is directly linked to age because people are more likely to require assistance with core activities with increasing age. Some analysis would consider a 'need for assistance' to be important irrespective of age. However, SEIFA does not include people above the age of 70 in this measure, due to the impact this age group has on the variable.</p>
ASGC	Australian Standard Geographic Classification. Refer to Geographies in the Glossary.
CENSUS COLLECTION DISTRICT (CD)	The CD is the smallest available area in the Australian Standard Geographic Classification (ASGC) and the smallest area for which the SEIFA indexes are available. ³⁸ CDs are specifically designed for Census collection purposes and generally represent a reasonable workload for a Census collection officer. This means that CDs have different physical sizes and different population sizes. A CD might represent one apartment block in a city, or it might cover a vast outback area. CDs form the basis of all other geographies. Their boundaries can change for various reasons, including development on urban fringes, the implementation of higher density housing or changes in local government boundaries. There are 38,704 CDs in the 2006 ASGC (of which, 37,457 CDs were included in SEIFA). Refer to Geographies in the Glossary.
CENSUS DATA ITEM	Census collects information about a wide range of topics. This paper refers to this Census information as data items. For SEIFA, Census data items were used to create variables, which in turn, were used to create the SEIFA indexes.
CORRELATION	A correlation measures the linear relationship between two measures. A perfectly positive relationship would have a correlation of +1; a perfectly negative relationship would have a correlation of -1; no relationship would have a correlation of 0.
DECILE	<p>Deciles divide a distribution into ten equal groups. In the case of SEIFA, the distribution of scores is divided into ten equal groups. The lowest scoring 10% of areas are given a decile number of 1, the second-lowest 10% of areas are given a decile number of 2 and so on, up to the highest 10% of areas which are given a decile number of 10.</p> <p>To create the CD level deciles for example, because there are 37,457 CDs, the 3,746 CDs that have the lowest scores are given a decile number of 1, and so on.</p> <p>Note that State deciles are also provided; refer to State decile in the Glossary for more information.</p>
DISABILITY	Refer to Need for Assistance.

³⁸ Mesh Blocks are a smaller standard geographic area, however are currently experimental.

GLOSSARY *continued*

DISADVANTAGE	The terms 'disadvantage' and 'socio-economic disadvantage' are used interchangeably in this paper.
EQUIVALISED HOUSEHOLD INCOME	<p>For the same standard of living, a larger household will require more income than a smaller household. For example, a two-person household will require more income to maintain the same standard of living as a single person household. Equivalence scales are used to adjust household income measures by the size of the household, so that all households can be more readily compared. The Census uses the 'modified OECD' equivalence scale. There are some issues for SEIFA, including:</p> <ul style="list-style-type: none">■ Equivalised household income is created only for private dwellings in the Census. Therefore SEIFA does not capture income for people living in non-private dwellings.■ Equivalised household income is not created for households with an adult who did not state their income or is temporarily absent, accounting for around 11% of applicable households. <p>There are other issues surrounding Census income measures, including:</p> <ul style="list-style-type: none">■ Equivalised household income includes wages, salaries and other income, such as dividends and rental assistance. Tax and superannuation contributions were not removed, therefore this is <i>not</i> a disposable income measure. Therefore, while some households pay less tax than other households, this cannot be captured in SEIFA.■ In the Census, income is measured after expenses from rent income or business/farm income were removed. These expenses may be significant for some households, and may result in low or even negative income.³⁹ Previous ABS research has found that these households tend to have higher net worth and household expenditure than other households with similar income levels.⁴⁰ Because these differences are not captured in the Census, they are not captured in SEIFA.■ Some households may mistakenly under-report their income, for example, by not including their pension or dividend income. This can lead to bias in SEIFA if this under-reporting is not evenly distributed throughout the population. <p>SEIFA is a general measure of relative socio-economic disadvantage that captures more information than income alone.</p>

³⁹ Negative and zero equivalised household income was measured in the Census.

⁴⁰ Appendix 4 of Household Wealth and Wealth Distribution, Australia, 2003-04, cat. no. 6554.0.

GLOSSARY *continued*

GEOGRAPHIES

The figure below compares the physical size of CDs, POAs, SLAs and LGAs for the Inner Sydney region. Note that this figure is indicative only, as these geographies can look quite different elsewhere in Australia. These maps were created using the 2006 Census Data by Location on the ABS website.⁴¹

GEOGRAPHIES, INNER SYDNEY



LOCAL GOVERNMENT AREA (LGA)

A Local Government Area (LGA) is an area under the responsibility of an incorporated local government or Indigenous council. LGA boundaries can be changed by the State/Territory government. Consequently the ABS adjusts its CDs and SLAs to match these boundaries. LGAs are equivalent to one or more SLAs. There are 676 LGAs in the 2006 ASGC (of which, 661 LGAs were included in SEIFA). Refer to Geographies in the Glossary.

MEAN

A mean is an average; a measure of central tendency of a distribution. A mean SEIFA score can be calculated by adding the value of all scores and dividing this by the number of scores being added.

NEED FOR ASSISTANCE

The disability variable, which is new to SEIFA in 2006, provides an indication of the physical or health aspects of relative socio-economic disadvantage. It is based on the new Census questions on need for assistance, which were developed to provide an indication of whether people have a profound or severe disability. People with a profound or severe disability are defined as those people needing help or assistance in one or more of the three core activity areas of self-care, mobility and communication, because of a disability, a long term health condition (lasting six months or more) or

⁴¹ www.censusdata.abs.gov.au

GLOSSARY *continued*

NEED FOR ASSISTANCE <i>continued</i>	advanced age. For brevity in this paper, need for assistance is referred to using the term 'disability'. Note that the Census measure was designed to indicate the disability status of people in Australia according to geographic area, and for small groups within the broader population and is not a comprehensive measure of disability. Disability can limit employment opportunities, and consequently access to financial resources. For the purpose of indicating relative socio-economic disadvantage, we have limited the scope of the SEIFA disability variable to people aged under 70 (for more information see Appendix B of the Technical Manual).
NET WORTH	Refer to Wealth.
PERCENTILE	<p>Percentiles divide a distribution into 100 equal groups. In the case of SEIFA, the distribution of scores is divided into 100 equal groups. The lowest scoring 1% of areas are given a percentile number of 1, the second-lowest 1% of areas are given a percentile number of 2 and so on, up to the highest 1% of areas which are given a percentile number of 100. SEIFA percentiles are provided to allow users to create their own groupings, such as quartiles (which contain 25% of CDs).</p> <p>Note that State percentiles are also provided; refer to State Percentile in the Glossary for more information.</p>
POPULATION WEIGHTING	<p>SEIFA releases indexes for five different types of area: Census Districts (CDs), Statistical Local Areas (SLAs), Local Government Areas (LGAs), Postal Areas (POAs) and State Suburbs (SSCs). However, the SEIFA indexes can be used to create scores for other types of area. Because CDs form the basis of all of the standard geography boundaries, CD scores can be used to represent larger standard areas.</p> <p>To create a score for a standard area, use a population weighted average of the CDs within the larger area. First, multiply each CD score by the number of people within that CD, and then divide by the total number of people within the larger area. These values can then be added together for all the CDs within the larger area, which is equal the score of the larger area. Population counts for CDs (the number of usual residents in the CD) have been provided with the index scores. The method outlined here is the method used to create the index scores for SLAs, LGAs, POAs and SSCs in SEIFA.</p> <p>Once the scores have been created for the larger areas, the ranks, deciles and percentiles for these areas are then calculated. For example, once the SLA scores have been created, the SLAs are then ranked in order of their SLA score, and given a SLA rank number (between 1 and 1426). Then the SLAs are separated into ten groups and given a SLA decile number (between 1 and 10). The 10% of SLAs with the lowest SLA scores are given an SLA decile number of 1. That is, neither SLA ranks nor SLA deciles are created using population weighting directly from CD ranks and deciles.</p> <p>It is important to note that, because of this method of construction, the distribution of scores for these larger geographic areas will <i>not</i> be a standard distribution. For example, the mean SLA scores will <i>not</i> be 1000, just as the standard deviations will not be 100. Also, the SLA deciles do refer to 10% of SLAs, and have only an indirect relationship to the CD deciles. An individual SLA will contain multiple CDs, with a range of CD scores, ranks and deciles that will be different to the SLA score, rank and decile.</p> <p>Refer to the Technical Paper for further information on population weighting. Refer also to Geographies in the Glossary.</p>
POSTAL AREA (POA)	<p>Postal Areas are not an ASGC standard geography but are based upon one or more CDs in an attempt to match the postcodes used by <i>Australia Post</i> (at the time of the Census). Postcodes are used for delivering mail and in many cases have no specific boundaries. Because some surveys are based on postcode information rather than standard geographies, SEIFA is available in Postal Areas. However, SEIFA users need to be aware that Postal Areas and postcodes are not always good matches, and should use POAs with caution. For example, a CD can only be matched to single Postal Area even if it spans two postcodes. POAs cannot be matched at all to SLAs or LGAs, and may be smaller or</p>

GLOSSARY *continued*

POSTAL AREA (POA) *continued*

larger than an SLA. The Postal Area number is the same as the matched postcode.⁴² There are 2515 POAs in the 2006 ASGC (of which, 2474 POAs were included in SEIFA). Refer to Geographies in the Glossary.

PRINCIPAL COMPONENTS ANALYSIS

This section explains the method used to create the SEIFA indexes, called Principal Components Analysis (PCA). This brief overview is in no way intended to replace a proper technical understanding of this technique or its use, but will provide some guidance to facilitate the use of SEIFA.

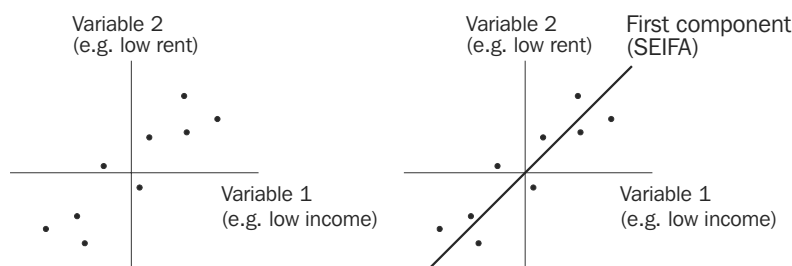
OVERVIEW

The concept of relative socio-economic disadvantage is difficult to capture because it has many dimensions and because these dimensions are hard to measure. In SEIFA, PCA is used to create a summary measure of a group of characteristics. For example, the Index of Relative Socio-economic Disadvantage (IRSD) is a summary measure of a group of characteristics related to relative socio-economic disadvantage. There is no preconception about how important each characteristic is to the index. The importance of a characteristic is determined by the relationships between the characteristics themselves across all the areas. These complex relationships are used to create a 'weight' for each characteristic. A SEIFA score can then be calculated for an individual area using that area's own characteristics and the characteristic's weights.

FURTHER EXPLANATION

SEIFA uses PCA to create 'components'. A component captures a common relationship between a group of variables. (Refer to Chapter 3 for the creation of 'variables' from characteristics.)⁴³ The figure below shows an example of two variables that are combined to produce a component.⁴⁴ The First component is a 'line of best fit' through the variables. SEIFA uses only this First component because it captures the most information.⁴⁵ This component becomes our measure of socio-economic disadvantage. This component is used to calculate the 'weights' of each variable, which depend on the importance of each variable to the component. A SEIFA 'score' for every area in Australia can then be determined.

To determine a SEIFA score for an area, we first multiply the variable weights by the variable values for that area, and then add together. As discussed in Chapter 2, the distribution of scores were then 'standardised', so that the average score is 1000 and approximately two-thirds of the scores lie between 900 and 1100. Chapter 4 shows the standardised distributions of the scores for each index. Refer to Standardisation in the Glossary.



⁴² Because postcodes are not named, Postal Areas are not given names (only the postcode number).

⁴³ Because variables are proportions, an area's variable value must have a value between 0 and 1. The variable values themselves are standardised.

⁴⁴ For example, these variables could be the low-income and no-schooling variables. SEIFA uses many variables, not just the two shown in this figure.

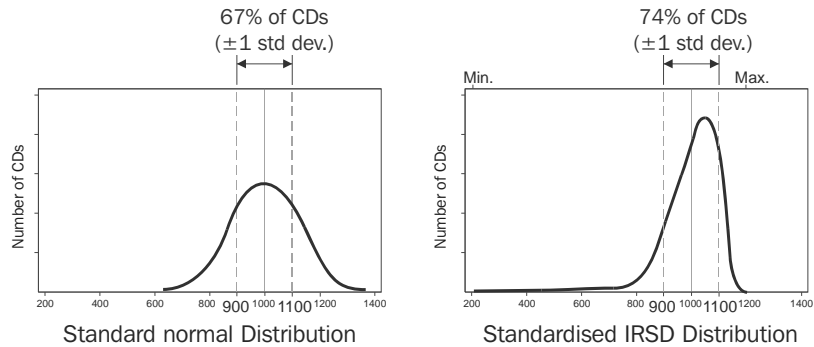
⁴⁵ There are as many components as there are variables, two in this example. However, SEIFA uses only the First component (and only this First component is shown in the figure). While this component captures the *most information possible by a single component*, this First component captures only a proportion of the information contained in the variables.

GLOSSARY *continued*

- RANK** To determine the SEIFA rank, all the areas are ordered from lowest score to highest score. The area with the lowest score is given a rank of 1, the area with the second-lowest score is given a rank of 2 and so on, up to the area with the highest score which is given the highest rank (37,457 for a collection district (CD) index). While two areas may appear to have the same score due to rounding, every area has an individual score and an individual rank. However, caution should be used when separating areas with similar scores and ranks.
- Note that State ranks are also provided; refer to State Rank in the Glossary for more information.
- REGRESSION ANALYSIS** A popular analytical technique that, as for all techniques, relies on a set of assumptions. Refer to a statistical text for details.
- SCORE** A SEIFA score is created using information about people and households in a particular area. A CD score is standardised against a mean of 1000 with a standard deviation of 100. This means that the average SEIFA CD score will be 1000 and the middle two-thirds of SEIFA scores will fall between 900 and 1100 (approximately). (Refer to Standardisation in the Glossary). A SEIFA score provides more information and is used for more sophisticated analysis. Ranks or deciles should be used for most analysis.
- STATISTICAL LOCAL AREA (SLA)** Statistical Local Areas (SLAs) are made up of one or more CDs. There are 1426 SLAs in the 2006 ASGC (of which, 1390 SLAs were included in SEIFA). Refer to Geographies in the Glossary.
- STANDARD GEOGRAPHIC AREAS** Refer to ASGC in the Glossary.
- STANDARDISATION (AGE)** Refer to Age-standardisation in the Glossary.
- STANDARDISATION** Using technical language, the standard distribution chosen for SEIFA has a "mean of 1000 and a standard deviation of 100". First, all of the CDs are ordered from the lowest to highest score. Second, all the scores are 'shifted' together so that the average area now has a new score of 1000. The areas are still in the same order, but they all have new scores spread around the average of 1000. The final stage changes how these scores are spread around the average. While they still remain in order, the scores are spread out (or condensed) so that two-thirds of the areas have 'standardised' scores somewhere between 900 and 1100; that is, approximately two-thirds of the scores lie within 100 either side of the average of 1000. This means that approximately 15% of CDs have a score lower than 900 with the remaining 85% of CDs having a score higher than 900. Approximately 85% of CDs have a score lower than 1100 with the remaining 15% of CDs having a score higher than 1100.
- It is important to note that the distributions of the SEIFA indexes are not exactly a normal distribution even though they have been standardised. However the above proportions are roughly the same. The figure below compares a standard distribution to the standardised distribution of the Index of Relative Socio-economic Disadvantage (IRSD). The distributions of the other indexes are more symmetrical, as shown in Chapter 4.
- Standardisation is useful when interpreting the scores and where the scores are used for more technical analysis. A different standard distribution could have been used, such as a mean of zero and a standard deviation of one. However, this would give some negative scores which may have been confusing and may have wrongly implied that areas with negative scores were 'disadvantaged'. When thinking about how to interpret standardised scores, it can be useful to bear in mind how the distribution would have looked if a different standard had been chosen.

STANDARDISATION *continued*

COMPARISON OF A STANDARD DISTRIBUTION AND THE STANDARDISED IRSD CD SCORE DISTRIBUTION



STATE SUBURBS (SSC)

Like postal areas, State Suburbs (SSC) are not an ASGC standard geography. SSCs are based upon one or more CDs in an attempt to match suburbs (at the time of the Census). However, unlike postal areas, State Suburbs do not cover all of Australia, although most of the population is covered. Refer to the geography links in Chapter 7 for further information about the areas covered by State Suburbs. Because some surveys are based on suburb information rather than standard geographies, SEIFA is available in State Suburbs. (Note that SLAs in Brisbane and other major urban areas in Queensland, Darwin and Canberra are aligned closely with suburbs.) However, SEIFA users need to be aware that State Suburbs and suburbs are not always good matches, and should use SSCs with caution. For example, a CD can only be matched to single State Suburb even if it spans two suburbs. Not all CDs are matched to State Suburbs, and these CDs are given an Unclassified SSC code. SSCs cannot be matched at all to SLAs or LGAs. SSCs have both a code and a name. SSC names are based on the most recent gazetted locality boundaries current at the time of a Census. There are 8464 SSCs in the 2006 ASGC, however not all of these will be included in SEIFA. Refer to Geographies in the Glossary. Note that the 2006 indexes for all geographies *except* SSC, were released on 26 March 2008.

STATE RANK, STATE DECILE, STATE PERCENTILE

SEIFA indexes are created for all areas within Australia. Therefore, the terms 'rank', 'decile' and 'percentile' used throughout this paper and the spreadsheets refer to the indexes based the number of areas across Australia. However, some users want to look only at areas within a certain state or territory. For these users, State ranks, State deciles and State percentiles have also been provided in the spreadsheets. These State numbers have *not* been created from scratch, but have instead been created using the Australia rank, decile and percentile as appropriate.

For example, the State ranks have been re-created for each state/territory, such that the lowest ranking CD in each of the eight states/territories is given a State rank of 1. This means, for example, that it is possible for a CD to have a rank of "79" (ranked 79th in Australia), and a State rank of "1" (ranked 1 in the state). On the spreadsheets, there will therefore be a total of eight State ranks equal to "1" (one for every state and territory). However, it is important to remember that a State rank of 1 *only* applies to that particular state/territory. That is, a CD with a State rank of "1" in NSW can have a very different SEIFA score (and level of relative socio-economic disadvantage) compared to a CD with a State rank of "1" in another state or territory.

Similarly, the State deciles have been created for each state/territory, such that the lowest 10% of areas in each of the eight States/Territories is given a State decile of 1. This means, for example, that it is possible for a CD to have a decile of "3", but a State decile of "1". A CD with a State decile of "1" in NSW can have a very different level of relative socio-economic disadvantage compared to a CD with a State decile of "1" in another state or territory.

Refer to Deciles, Ranks, Percentiles in the Glossary for more information.

GLOSSARY *continued*

- VARIABLES** SEIFA uses principal components analysis to create scores using a range of measures related to the concept of relative socio-economic disadvantage. These measures, called variables, are created using data items from the Census.
- WEALTH** Wealth is defined by different people in different ways. However, it is commonly interpreted as 'net worth', being a person's assets (such as shares and property) less their liabilities. In SEIFA, we extend this concept to include other aspects of wealth, such as access to lines of credit. However, the Census captures very little direct information on these measures. Therefore SEIFA uses proxy measures, such as ownership of an unincorporated enterprise.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au



2203900001063

ISBN 0 642 47936 4

RRP \$11.00