



**Technical Paper**

**Socio-Economic Indexes  
for Areas (SEIFA)**

**2016**

**2033.0.55.001**

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 10.30AM (CANBERRA TIME) TUESDAY 27 MARCH 2018

ABS Catalogue No. 2033.0.55.001

© Commonwealth of Australia 2018

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

1. INTRODUCTION	
1.1 What is SEIFA? .....	4
1.2 Purpose of technical paper.....	4
1.3 Some historical context .....	4
1.4 Features of SEIFA 2016.....	4
1.5 Interpretation of the indexes .....	5
2. CONCEPTUAL FRAMEWORK	
2.1 The concept of relative socio-economic advantage and disadvantage .....	6
2.2 Defining the concept behind each of the four indexes.....	6
3. INTRODUCTION	
3.1 The candidate list of variables .....	8
3.2 Constructing the variables.....	8
3.3 Description of candidate SEIFA variables.....	9
3.4 Basic exploratory analysis of variables .....	13
3.5 Candidate variable list for each index.....	13
4. CONSTRUCTION OF THE INDEXES	
4.1 Principal Component Analysis .....	15
4.2 Areas with no index scores.....	15
4.3 Step-by-step process .....	17
4.4 Technical details of each index: variables and loadings .....	19
4.5 Distributions of the indexes .....	24
4.6 Basic output: scores, ranks, deciles, and percentiles .....	26
4.7 Geographic output levels for SEIFA 2016.....	27
5. VALIDATION OF THE INDEXES	
5.1 Relationships between the indexes .....	29
5.2 Comparing 2011 and 2016 rankings .....	29
5.3 Validation of higher level area indexes.....	29
6. USING AND INTERPRETING SEIFA	
6.1 Broad guidelines on appropriate use .....	30
6.2 Choice of index .....	31
6.3 Using index scores for areas larger than SA1 .....	31
6.4 Mapping the indexes .....	32
6.5 Using the indexes as contextual variables in social analysis.....	32
6.6 Area-based quantiles versus population-based quantiles.....	32
References .....	34
Appendix – Variable Specifications .....	35

## 1. INTRODUCTION

### 1.1 WHAT IS SEIFA?

Socio-Economic Indexes for Areas (SEIFA) is a product developed by the ABS that ranks areas in Australia according to relative socio-economic advantage and disadvantage. The indexes are based on information from the five-yearly Census. SEIFA 2016 is based on Census 2016 data, and consists of four indexes, each focusing on a different aspect of socio-economic advantage and disadvantage, and being a summary of a different subset of Census variables.

Some common uses of SEIFA include:

- determining areas that require funding and services,
- identifying new business opportunities, and
- assisting research into the relationship between socio-economic disadvantage and various social outcomes.

### 1.2 PURPOSE OF TECHNICAL PAPER

This paper provides information on the concepts, data, and methods used to create SEIFA 2016. The paper also contains discussion of the correct interpretation and appropriate use of the indexes.

This paper is intended to be a comprehensive reference for SEIFA 2016. A basic user guide – SEIFA Basics – has also been prepared as part of this product release and can be viewed on the product web pages.

### 1.3 SOME HISTORICAL CONTEXT

A relative measure of socio-economic disadvantage was first produced by the ABS following the 1971 Census. Socio-Economic Indexes for Areas (SEIFA), in its present form, was first produced from the 1986 Census data.

### 1.4 FEATURES OF SEIFA 2016

This section highlights some important features of SEIFA 2016, and how they compare with SEIFA 2011.

SEIFA 2016 consists of the same four indexes as produced for SEIFA 2001, 2006 and 2011, each referring to the general population:

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

Since SEIFA is an established product, we have generally aimed to maintain consistency between SEIFA 2016 and the previous release. However, some changes have been made and are described below.

### **Updated geography standard**

- SEIFA 2016 uses the 2016 version of the Australian Statistical Geography Standard (ASGS). The structure of the 2016 ASGS is similar to the structure of 2011 ASGS. However, there have been updates to SA1 boundaries in many areas, in particular where there has been significant population growth. Also, State Suburbs (SSCs) and Postal Areas (POAs) are now constructed from Mesh Blocks rather than SA1s. For more information about the ASGS, see ABS (2016a) and ABS (2016b).
- SEIFA 2016 is not being released at the Statistical Local Area (SLA) level, since this is not part of the ASGS.
- There have been significant changes to some Local Government Areas (LGAs).
- Norfolk Island was included in the Australian Census for the first time in 2016.

### **Variables underpinning the indexes**

- Some variables were updated in line with new classification standards. Variables using cut-off values in their definitions, such as high and low income, were updated to use new cut-off values. For more information about how the cut-off values were selected, see sections 3.3.1 and 3.3.5. Census 2016 did not collect information about the type of internet access in dwellings, and so the DIALUP variable from SEIFA 2011 could not be considered for inclusion in SEIFA 2016.

### **Output**

- For the first time, SEIFA 2016 output is being made available in interactive maps on the ABS website.

## 1.5 INTERPRETATION OF THE INDEXES

To set some context for the rest of this paper, it is worth briefly touching on some important characteristics of the indexes:

- The indexes are assigned to areas, not to individuals. They indicate the collective socio-economic characteristics of the people living in an area.
- As measures of socio-economic conditions, the indexes are best interpreted as ordinal measures that rank areas. The index scores are based on an arbitrary numerical scale and do not represent a quantity of advantage or disadvantage. For ease of interpretation, we generally recommend using the index rankings and quantiles (e.g. deciles) for analysis, rather than using the index scores. Index scores are still provided in the output, and can be used for more sophisticated analyses.
- Each index is constructed based on a weighted combination of selected variables. The indexes are dependent on the set of variables chosen for the analysis. A different set of underlying variables would result in a different index.
- The indexes are primarily designed to compare the relative socio-economic characteristics of areas at a given point in time. It can be very difficult to perform useful longitudinal or time series analysis, and this sort of analysis should be undertaken with care.

There is more discussion of these points in Section 6.1.

## 2. CONCEPTUAL FRAMEWORK

### 2.1 THE CONCEPT OF RELATIVE SOCIO-ECONOMIC ADVANTAGE AND DISADVANTAGE

For SEIFA 2016, the concept of relative socio-economic advantage and disadvantage is the same as that used for SEIFA 2006 and 2011. That is, the ABS broadly defines relative socio-economic advantage and disadvantage in terms of *people's access to material and social resources, and their ability to participate in society*. This is described as 'broadly defined' in recognition of the many concepts that have emerged in the literature to describe advantage and disadvantage. The dimensions included in SEIFA are guided by international research, given the constraints of Census data. The Census does collect information on the key dimensions of income, education, employment, occupation, housing, and also some other miscellaneous indicators of advantage and disadvantage. These are the dimensions used for SEIFA to inform variable selection and are discussed further in Section 3.

Another point to note is that SEIFA measures relative advantage and disadvantage at an area level, not at an individual level. Area level and individual level disadvantage are separate though related concepts. Area level disadvantage depends on the socio-economic conditions of a community or neighbourhood as a whole. These are primarily the collective characteristics of the area's residents, but may also be characteristics of the area itself, such as a lack of public resources, transport infrastructure or high levels of pollution. However, it is important to remember that SEIFA is restricted to the information that is included in the Census.

It is recommended that users of SEIFA consider their research interests, the definition of each SEIFA index and the variables included in each index to determine the appropriate index to use. The ABS produces four indexes, each summarising a different subset of Census variables, because users may be interested in different aspects of socio-economic advantage and disadvantage. The next section provides more information on each of the four indexes included in SEIFA.

### 2.2 DEFINING THE CONCEPT BEHIND EACH OF THE FOUR INDEXES

This section gives a description of the concept behind each of the four indexes. For a list of the variables included in each index, see Section 4.4.5.

#### 2.2.1 The Index of Relative Socio-Economic Disadvantage

The IRSD summarises variables that indicate relative disadvantage. This index ranks areas on a continuum from most disadvantaged to least disadvantaged. A low score on this index indicates a high proportion of relatively disadvantaged people in an area. We cannot conclude that an area with a very high score has a large proportion of relatively advantaged people, as there are no variables in the index to indicate this. We can only conclude that such an area has a relatively low incidence of disadvantage.

#### 2.2.2 The Index of Relative Socio-Economic Advantage and Disadvantage

The IRSAD summarises variables that indicate either relative advantage or disadvantage. This index ranks areas on a continuum from most disadvantaged to most advantaged.

An area with a high score on this index has a relatively high incidence of advantage and a relatively low incidence of disadvantage. Due to the differences in scope between this index and the IRSD, the scores of some areas can vary substantially between the two indexes. For example, consider a large area that has parts containing relatively disadvantaged people, and other parts containing relatively advantaged people. This area may have a low IRSD ranking, due to its pockets of disadvantage. However, its IRSAD ranking may be moderate, or even above average, because the pockets of advantage may offset the pockets of disadvantage.

### 2.2.3 The Index of Economic Resources

The IER summarises variables relating to the financial aspects of relative socio-economic advantage and disadvantage. These include indicators of high and low income, as well as variables that correlate with high or low wealth. Areas with higher scores have relatively greater access to economic resources than areas with lower scores.

### 2.2.4 The Index of Education and Occupation

The IEO summarises variables relating to the educational and occupational aspects of relative socio-economic advantage and disadvantage. This index focuses on the skills of the people in an area, both formal qualifications and the skills required to perform different occupations. A low score indicates that an area has a high proportion of people without qualifications, without jobs, and/or with low skilled jobs. A high score indicates many people with high qualifications and/or highly skilled jobs.

### 3. THE DATA UNDERPINNING THE INDEXES

This section looks at the data used to construct the four indexes in SEIFA 2016. All data is from the 2016 Census of Population and Housing.

#### 3.1 THE CANDIDATE LIST OF VARIABLES

The candidate variable list from SEIFA 2011 was used for SEIFA 2016 with one exception: the type of internet access variable was not included in Census 2016, and so this was not available for inclusion in SEIFA 2016. The candidate variables fall into a multi-dimensional framework. The dimensions are:

- income variables,
- education variables,
- employment variables,
- occupation variables,
- housing variables, and
- other miscellaneous indicators of relative advantage or disadvantage.

Variables can relate to persons, families, or dwellings.

#### 3.2 CONSTRUCTING THE VARIABLES

##### ***Specifications***

The variables were expressed as proportion of units in an area with a specific characteristic. Depending on the variable, the unit may be a person, family, or dwelling. As each variable was expressed as a proportion, a numerator and denominator were required. The numerator for each variable was a subset of the denominator. In most cases, the numerator and denominator specifications were based on SEIFA 2011 specifications. Some minor changes were made to reflect updates to the Census 2016 variable coding. The Appendix contains detailed descriptions of the numerators and denominators used for all the SEIFA variables. Note that for convenience of presentation in the following sections, the variable proportions are expressed as percentages.

##### ***Place of Usual Residence***

A person may or may not be enumerated at their place of usual residence on Census Night. For all variables used in SEIFA 2016, a person's usual residence was used as the basis of analysis. Counts compiled on a 'place of usual residence' basis are appropriate for SEIFA, because they are less likely to be influenced by seasonal factors such as school holidays and snow seasons. However, it is important to understand that certain areas, for example SA1s in popular tourist destinations, may receive scores influenced by the specific time at which the Census is conducted. For instance, the 2016 Census was conducted in August 2016, which is during the high season for ski resorts and the townships in those areas. This means that these areas may have higher property rental prices, higher employment figures and greater income levels than if the Census were conducted in the low season.

##### ***Not stated and not applicable***

We excluded records with 'Not stated' and 'Not applicable' values (for the particular variable) from both the numerator and denominator counts. For details, see the Appendix.



# SEIFA TECHNICAL PAPER

---

## 3.3 DESCRIPTION OF CANDIDATE SEIFA VARIABLES

This section contains a description of each variable on the candidate variable list. There is a brief discussion of how each variable relates to our definition of relative socio-economic advantage or disadvantage. The tables containing the variable descriptions also state whether the variable is an indicator of relative advantage (adv) or relative disadvantage (dis). Each subsection corresponds to one of the socio-economic dimensions listed in Section 3.1.

### 3.3.1 Income variables

Table 3.1 List of income variables

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>INC_LOW</b>	<b>% People with stated annual household equivalised income between \$1 and \$25,999 (approx. 1st and 2nd deciles) (dis)</b>
<b>INC_HIGH</b>	<b>% People with stated annual household equivalised income greater than or equal to \$78,000 (approx. 9th and 10th deciles) (adv)</b>

Income is an important economic resource, and is a core component of our notion of relative socio-economic advantage and disadvantage (outlined in Section 2.1). Income variables are used in all the SEIFA indexes except the Index of Education and Occupation. The income variables used equivalised household income. Equivalisation is a process in which household income is adjusted by an 'equivalence scale', based on the number of adults and children in the household. The SEIFA variables using equivalised household income are calculated from the Census 2016 Equivalised Total Household Income variable (HIED).

The low income variable has been defined for SEIFA 2016 to capture approximately the first and second deciles of the equivalised household income distribution, excluding negative and nil income. That is, those people living in dwellings with equivalised household income between \$1 and \$499 per week (\$1 to \$25,999 per year). Much of the low income decile was a strong indicator of disadvantage, but people reporting negative and nil incomes tended to have profiles with less association with disadvantage. The cut-off of \$78,000 for the high income variable was chosen to approximately capture the highest income quintile (top 20%).

### 3.3.2 Education variables

Table 3.2 List of education variables

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>ATUNI</b>	<b>% People aged 15 years and over attending university or other tertiary institution (adv)</b>
<b>ATSCHOOL</b>	<b>% People aged 15 years and over attending secondary school (adv)</b>
<b>CERTIFICATE</b>	<b>% People aged 15 years and over whose highest level of educational attainment is a Certificate Level III or IV qualification (dis)</b>
<b>DEGREE</b>	<b>% People aged 15 years and over whose highest level of educational attainment is a bachelor degree or higher qualification (adv)</b>
<b>DIPLOMA</b>	<b>% People aged 15 years and over whose highest level of educational attainment is an advanced diploma or diploma qualification (adv)</b>
<b>NOEDU</b>	<b>% People aged 15 years and over who have no educational attainment (dis)</b>
<b>NOYEAR12ORHIGHER</b>	<b>% People aged 15 years and over whose highest level of educational attainment is Year 11 or lower (includes Certificate Levels I and II; excludes those still at secondary school) (dis)</b>

---

## SEIFA TECHNICAL PAPER

---

Education is an important domain when considering socio-economic advantage and disadvantage because the skills people obtain through school and post-school education can increase their own standard of living, as well as that of their community. Certificate Levels I and II are regarded as a lower educational attainment than year 12 schooling, and are grouped in the NOYR12ORHIGHER variable, as opposed to the CERTIFICATE variable. This specific educational hierarchy was based on the ABS publication *Education and Work Australia, May 2011* (ABS, 2011a). Note also that the CERTIFICATE variable is an indicator of relative disadvantage in SEIFA. It is true that having a certificate qualification gives a person an advantage over someone with no qualifications. However, at an area level, a high proportion of people with certificate qualifications correlates with other disadvantaging characteristics (e.g. lower skilled occupations).

### 3.3.3 Employment variables

Table 3.3 List of employment variables

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>UNEMPLOYED</b>	<b>% People in the labour force who are unemployed (dis)</b>
<b>UNEMPLOYED1</b>	<b>% People aged 15 and over who are unemployed (dis)</b>

For most people, employment is the main source of their income. Employment can also contribute to social participation and self-esteem. An unemployment variable is included in all of the SEIFA indexes. The standard unemployment variable (UNEMPLOYED) is calculated as the number of unemployed people divided by the number of people in the labour force (the unemployment rate). The variable used in the Index of Economic Resources (UNEMPLOYED1) is the number of unemployed people divided by the entire adult population of the area. This enables us to distinguish the unemployed from those employed and those not in the labour force, as the latter two groups were found to have significantly higher average wealth.

### 3.3.4 Occupation variables

Table 3.4 List of occupation variables

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>OCC_DRIVERS</b>	<b>% Employed people classified as Machinery Operators and Drivers (dis)</b>
<b>OCC_LABOUR</b>	<b>% Employed people classified as Labourers (dis)</b>
<b>OCC_MANAGER</b>	<b>% Employed people classified as Managers (adv)</b>
<b>OCC_PROF</b>	<b>% Employed people classified as Professionals (adv)</b>
<b>OCC_SALES_L</b>	<b>% Employed people classified as Low-Skill Sales Workers (dis)</b>
<b>OCC_SERVICE_L</b>	<b>% Employed people classified as Low-Skill Community and Personal Service Workers (dis)</b>
<b>OCC_SKILL1</b>	<b>% Employed people who work in a Skill Level 1 occupation (adv)</b>
<b>OCC_SKILL2</b>	<b>% Employed people who work in a Skill Level 2 occupation (adv)</b>
<b>OCC_SKILL4</b>	<b>% Employed people who work in a Skill Level 4 occupation (dis)</b>
<b>OCC_SKILL5</b>	<b>% Employed people who work in a Skill Level 5 occupation (dis)</b>

Occupation plays a significant part in determining socio-economic advantage and disadvantage. The ability to accumulate economic resources varies greatly with occupation type. The SEIFA 2016 occupation variables have been classified using ANZSCO – *Australian and New Zealand Standard Classification of Occupations, Version 1.2* (ABS, 2013).

Each occupation in ANZSCO is assigned a skill level ranging from 1 (highest) to 5 (lowest), which indicates the range and complexity of the set of tasks performed in a particular occupation. These skill levels were used as the basis of the occupation variables in the Index of Education and Occupation. The aim was to include broad categories of both advantaging and disadvantaging occupations, which complement the education variables by introducing the aspect of vocational skills. For the IRSD and the IRSAD, we used the ANZSCO major groups in conjunction with the skill levels to construct the occupation variables. This was done to identify occupations, or groups of occupations, which contribute to relative advantage or disadvantage at an area level. Using the major groups as well as the skill levels also helped to maintain consistency with SEIFA 2006 and SEIFA 2011.

### 3.3.5 Housing variables

Table 3.5 List of housing variables<sup>(a)</sup>

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>FEWBED</b>	<b>% Occupied private dwellings with one or no bedrooms (dis)</b>
<b>HIGHBED</b>	<b>% Occupied private dwellings with four or more bedrooms (adv)</b>
<b>HIGHMORTGAGE</b>	<b>% Occupied private dwellings paying more than \$2,800 per month in mortgage repayments (adv)</b>
<b>HIGHRENT</b>	<b>% Occupied private dwellings paying more than \$470 per week in rent (adv)</b>
<b>LOWRENT</b>	<b>% Occupied private dwellings paying less than \$215 per week in rent (excluding \$0 per week) (dis)</b>
<b>MORTGAGE</b>	<b>% Occupied private dwellings owning the dwelling they occupy (with a mortgage) (adv)</b>
<b>OVERCROWD</b>	<b>% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard) (dis)</b>
<b>OWNING</b>	<b>% Occupied private dwellings owning the dwelling they occupy (without a mortgage) (adv)</b>
<b>SPAREBED</b>	<b>% Occupied private dwellings with one or more bedrooms spare (based on Canadian National Occupancy Standard) (adv)</b>

(a) All dwelling variables excluded dwellings whose inhabitants all usually resided elsewhere, whose inhabitants were all under 15, or which could not be classified due to insufficient information. For numerator and denominator specifications, see Appendix A.

Having an adequate and appropriate place to live is fundamental to socio-economic wellbeing. There are many aspects to housing that affect the quality of people's lives. Dwelling size, cost and security of tenure are all important in this regard, and are therefore considered in SEIFA. Housing size is measured by the variables FEWBED, HIGHBED, OVERCROWD and SPAREBED. The variable FEWBED measures dwellings with one or no bedrooms, whilst the variable HIGHBED measures dwellings with four or more bedrooms. The variable OVERCROWD measures dwellings that do not have enough bedrooms for their occupants. Conversely, the variable SPAREBED measures dwellings that have one or more bedrooms spare for their occupants. These last two variables are calculated using the Canadian National Occupancy Standard, which determines housing appropriateness using the number of bedrooms and the number, age, sex and relationships of household members - for more information, refer to the Explanatory Notes for Housing Occupancy and Costs, 2015-16 (ABS, 2017). Housing cost is measured in SEIFA using reported mortgage or rent payments. The cut-offs for the high and low groups were based on the ranges corresponding to the top and bottom quintiles. The high housing cost variables (HIGHMORTGAGE, HIGHRENT) are indicators of relative advantage, because they indicate greater financial capacity, as well as higher quality housing or locational advantage. The low housing cost variable (LOWRENT) is an indicator of relative disadvantage, for similar reasons.

Owning a house, with or without a mortgage, is an indicator of advantage. First, owning a house implies security of tenure. For many Australian households, the family home is their most valuable asset. Owning with a mortgage indicates the financial capacity to make repayments, as well as the possession of a future asset. The denominator of the mortgage and rent variable proportions is based on all households in an area.

The Census captures limited household information, and does not for instance capture housing affordability, housing stress, dwelling value and dwelling quality. Although some variables, such as number of bedrooms and amount of rent or mortgage payments, may provide a proxy in some instances, their relationship to dwelling quality and dwelling value is not uniform across all areas. Due to this lack of comparability we have not attempted to construct these variables.

### 3.3.6 Other indicators of relative advantage or disadvantage

With the information available to us from the Census, there are additional variables we can construct related to socio-economic advantage and disadvantage that do not fall into the main domains of education, occupation, housing or employment. These variables are discussed below.

Table 3.6 List of other indicators of relative advantage or disadvantage <sup>(a)</sup>

---

<i>Variable mnemonic</i>	<i>Variable description</i>
<b>CHILDJOBLESS</b>	% Families with children under 15 years of age and jobless parents (dis)
<b>DISABILITYU70</b>	% People aged under 70 who need assistance with core activities due to a long-term health condition, disability or old age (dis)
<b>ENGLISHPOOR GROUP</b>	% People who do not speak English well (dis)
<b>HIGHCAR</b>	% Occupied private dwellings that are group occupied private dwellings (dis)
<b>LONE</b>	% Occupied private dwellings with three or more cars (adv)
<b>NOCAR</b>	% Occupied private dwellings that are lone person occupied private dwellings (dis)
<b>NONET</b>	% Occupied private dwellings with no cars (dis)
<b>ONEPARENT</b>	% Occupied private dwellings with no Internet connection (dis)
<b>SEPDIVORCED</b>	% Families that are one parent families with dependent offspring only (dis)
<b>UNINCORP</b>	% People aged 15 and over who are separated or divorced (dis)
	% Occupied private dwellings with at least one person who is an owner of an unincorporated enterprise (adv)

(a) All dwelling variables excluded dwellings whose inhabitants all usually resided elsewhere, whose inhabitants were all under 15, or which could not be classified due to insufficient information. For numerator and denominator specifications see Appendix A.

The CHILDJOBLESS variable is defined as the proportion of families with children under 15 years old and jobless parents. The variable could be an indicator for entrenched disadvantage since children who grow up in jobless families may be more likely to experience intergenerational unemployment and diminished opportunities to participate in society.

The disability variable (DISABILITYU70) provides an indication of the physical or health aspects of socio-economic disadvantage. It is based on the Census question on need for assistance, which was 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, long term health condition (lasting six months or more) or old age. Disability limits employment opportunities, and possibly access to community 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, as was done for SEIFA 2011.

A lack of fluency in English may limit employment opportunities and the ability to participate in society.

A car is both a material resource and a means of transport that enables greater freedom. A limitation of the NOCAR variable is that the need for a car varies depending on the remoteness of the area and access to public transport.

Having an internet connection allows access to information and services and may demonstrate a certain level of financial capability.

A past analysis of wealth data collected by the ABS showed that lone person households have lower average wealth (per person) than other household types. A higher proportion of lone person households in an area is correlated with lower ability to access economic resources beyond what is measured by the equivalised household income variables. An analysis of group households yielded a similar conclusion – an association with low wealth. A high proportion of unincorporated enterprise owners was found to correlate with high wealth and access to economic resources. These three variables were used only in the Index of Economic Resources.

One parent households are disadvantaged compared with other household types, because of the need to simultaneously provide and care for dependants. Apart from having lower equivalised household incomes, one parent families also have lower rates of employment and labour force participation, lower rates of home ownership and higher incidence of financial stress, as compared to couple family households – see, for example, *Australian Social Trends, 2007* (ABS, 2007). There are significant correlations at the area level between the number of one parent families and many indicators of relative socio-economic disadvantage. The same patterns are evident for areas with high proportions of people who are separated or divorced.

### 3.4 BASIC EXPLORATORY ANALYSIS OF VARIABLES

The Census data was converted into the SEIFA variable proportions, as defined in section 3.2. Summary statistics proportions were analysed to identify significant changes since 2011. Overall, there were no unexpected changes to the SEIFA variable proportions.

### 3.5 CANDIDATE VARIABLE LIST FOR EACH INDEX

Table 3.7 shows the candidate variable list for each index. The candidate list includes all variables considered for inclusion in an index before the principal component analysis stage (discussed in section 4). The final list of variables included in each index can be found in table 4.10 in section 4.4.5.

## SEIFA TECHNICAL PAPER

Table 3.7 Candidate variable list for each index, by socio-economic dimension

<i>Dimension</i>	<i>Index of Relative Socio-Economic Disadvantage</i>	<i>Index of Relative Socio-Economic Advantage and Disadvantage</i>	<i>Index of Economic Resources</i>	<i>Index of Education and Occupation</i>
Income	INC_LOW	INC_HIGH INC_LOW	INC_HIGH INC_LOW	
Education	NOYR12ORHIGHER NOEDU CERTIFICATE	NOYR12ORHIGHER NOEDU CERTIFICATE ATUNI DIPLOMA DEGREE		NOYR12ORHIGHER NOEDU CERTIFICATE ATUNI DIPLOMA DEGREE ATSCHOOL UNEMPLOYED
Employment	UNEMPLOYED	UNEMPLOYED	UNEMPLOYED1	
Occupation	OCC_LABOUR OCC_DRIVERS OCC_SERVICE_L OCC_SALES_L	OCC_LABOUR OCC_DRIVERS OCC_SERVICE_L OCC_SALES_L OCC_PROF OCC_MANAGER		OCC_SKILL1 OCC_SKILL2 OCC_SKILL4 OCC_SKILL5
Housing	LOWRENT OVERCROWD FEWBED	LOWRENT OVERCROWD HIGHBED HIGHRENT HIGHMORTGAGE OWNING SPAREBED	LOWRENT OVERCROWD MORTGAGE HIGHBED HIGHRENT HIGHMORTGAGE OWNING	
Other	CHILDJOBLESS ONEPARENT NOCAR DISABILITYU70 ENGLISHPOOR SEPDIVORCED NONET	CHILDJOBLESS ONEPARENT NOCAR DISABILITYU70 ENGLISHPOOR SEPDIVORCED NONET HIGHCAR	UNINCORP ONEPARENT NOCAR GROUP LONE	

Note – the Appendix contains the definitions of each variable listed in this table.

Note – the variables listed in this table are not the final list of variables included in the indexes. For the final list, see table 4.10 in Section 4.4.5.

### 4. CONSTRUCTION OF THE INDEXES

This section describes the methods used to construct the indexes, some important technical specifications of each index, and some basic output.

#### 4.1 PRINCIPAL COMPONENT ANALYSIS

Each index is a weighted sum of SEIFA variables. As with past versions of SEIFA, principal component analysis (PCA) is used to determine the weights. This section introduces some technical concepts related to PCA to assist the reader understand the SEIFA index construction process. Some references are given at the end of this section for readers interested in a comprehensive discussion of PCA.

PCA is a technique that involves summarising a large number of correlated variables into a set of new uncorrelated components, each of which is a linear combination of the original variables. There are as many principal components as there are variables. If the original variables are highly correlated, much of the variation can be summarised by a reduced set of components, hence enabling some easier analysis. The first principal component accounts for the largest proportion of variance in the original dataset, with each following component explaining less of the variance. The principle component used for each SEIFA index is the one that can be interpreted as best explaining the variation in the concept of advantage and disadvantage for that index. For the four indexes in SEIFA 2016, the first principal component was used to create the index.

The PCA procedure gives an eigenvalue for each component, which indicates the amount of variance in the original data explained by the component. The proportion of variance explained by a principal component is its eigenvalue divided by the sum of all the eigenvalues. Each variable in the analysis will be correlated with each component. This correlation is called the loading. Loadings help to interpret which aspects of advantage and disadvantage a component may represent. The loadings are also useful in comparing results obtained from different sets of original variables (such as for the four indexes in SEIFA). Loadings for each index are presented in the following sections.

In order to generate the component scores (otherwise known as raw scores) the loading is converted to a weight by dividing it by the square root of the eigenvalue. The product of the weight and standardised variable values are summed to produce the raw scores. The raw scores for each component will then have variance equal to the eigenvalue for that component. We then rescale the raw scores to a mean of 1,000 and standard deviation of 100 to create a new set of scores that are the index scores in SEIFA - this process is known as "standardisation".

More detailed explanations of PCA can be found in Jolliffe (1986) and O'Rourke (2005).

#### 4.2 AREAS WITH NO INDEX SCORES

Some SA1 areas do not receive an index score, either due to low populations or poor quality data. The criteria used to identify these areas are called 'exclusion rules'. SEIFA 2016 uses a similar exclusion rule framework as SEIFA 2011, with the aim of obtaining a reliable index score for as many areas as possible.

The 2016 exclusion rules use a two-phase approach:

- The first phase excludes areas (SA1s) that should not receive a SEIFA score because of the type of area, confidentiality, and reliability concerns (e.g. low population or low response rates for particular key variables).

## SEIFA TECHNICAL PAPER

- The second phase excludes areas (SA1s) by looking specifically at the variables included in each index. For each SA1, if any of the variables have a low denominator count, it is deemed that there is not enough data to support a reliable calculation of an index score for that area.

Some additional comments on the exclusion rule framework:

- The first phase rules are applied before PCA, and the second phase rules are applied after PCA and the list of variables is finalised. Section 4.3 provides details on how this is implemented.
- SA1s excluded in the first phase will be excluded for all four indexes. The number of SA1s excluded in the second phase may be different for each index, because they have different sets of variables.
- Following on from the point above, an area can receive a score for one index and not another depending on the make-up of its variables.
- The low denominator cut-off of 6 is chosen based on past practice and a judgement on how many responses are required to calculate a reliable value for an area.
- The exclusion of areas is based on both the confidentialised and unconfidentialised counts for each SEIFA variable to ensure the confidentiality of respondents is upheld and the reliability of the indexes is maintained.

The specific exclusion rules and the number of areas meeting each rule are summarised in table 4.1. Note that areas might fall into multiple categories, and this is why the column sum does not equal the final total number of excluded areas.

Table 4.1 Summary of excluded areas

	<i>Total SA1s excluded - First phase</i>
Population = 0	1,275
No Usual Address SA1	9
Offshore, Shipping SA1	24
Population > 0 and ≤ 10	798
Employed persons ≤ 5	2,086
Classifiable <sup>(a)</sup> occupied private dwellings ≤ 5	2,233
People in private dwellings ≤ 20%	1,699
<b>Total excluded due to any of the rules above</b>	<b>2,380</b>
<i>Index</i>	<i>Total SA1s excluded - Second phase</i>
IRSD	115
IRSAD	115
IER	91
IEO	12

(a) These are dwellings where the type of household living in the dwelling could be determined during the Census collection process. For more information, see *2901.0 - Census of Population and Housing: Census Dictionary, 2016* (ABS, 2016b).

The proportions of excluded SA1s are similar to those for SEIFA 2011.



### 4.3 STEP-BY-STEP PROCESS

With the preceding two sections providing context, a step-by-step process for constructing the indexes is presented below.

#### **Step 1. Creating the initial variable list**

Given the available data, we created a list of variables related to our definition of relative socio-economic advantage and disadvantage.

#### **Step 2. Constructing the variables**

We created all variables as proportions at the SA1 level (e.g. ‘% people aged 15 and over with no post-school qualifications’). We then standardised these proportions to a mean of 0 and a standard deviation of 1. The standardisation was used to prevent variables with larger prevalence, or larger ranges, from having a disproportionate influence on the index.

#### **Step 3: Applying first phase exclusion rules**

We excluded areas (SA1s) that should not receive an index score because of the type of area, confidentiality, and reliability concerns. See table 4.1 for specific rules.

#### **Step 4: Calculating the correlation matrix**

We set to missing any variables that have denominators less than our prescribed cut-off of 6. Note that we did not exclude areas based on this cut-off at this stage in the process – this occurred at step 9. We calculated the correlation matrix and used pairwise deletion when areas (observations) contain missing values. Pairwise deletion is a method for dealing with missing data. The maximum number of non-missing values for each pair of variables is used in the calculation of the correlation matrix. This is in contrast to listwise deletion in which entire records (areas in our case) are removed from the analysis if any of their variables have missing values. Given the number of observations in our dataset and the low prevalence of missing values, the use of pairwise deletion had very little impact on the correlation matrix, however it did enable a convenient way of implementing our second phase exclusion rules (see step 9).

#### **Step 5. Removing very highly correlated variables**

We removed highly correlated variables to avoid over-representing any specific socio-economic characteristic. When two variables had a correlation coefficient greater than 0.8 in absolute value, and were measuring conceptually similar aspects of advantage or disadvantage, we generally removed one of them. However, we applied some discretion, depending on the particular variables and the size of the correlation.

#### **Step 6. Conducting the initial PCA**

Using the correlation matrix, we conducted principal component analysis (PCA) to obtain the loading for each variable on the first principal component.

#### **Step 7. Removing low loading variables**

We excluded variables with loadings less than 0.3 in absolute value, on the grounds that they were not strong indicators of relative advantage or disadvantage. This limit is an accepted level in the PCA literature (see Jolliffe, 1986) and has been used in past releases of SEIFA. We removed variables one at a time, starting with the lowest loading variable.

#### **Step 8. Conducting PCA on the reduced list of variables**

We conducted a PCA on the reduced variable list, and if any other variables loaded below 0.3, we repeated steps 7 and 8.

### Step 9. Finalise list of variables in index and apply second phase exclusion rules

Once we knew the final list of variables in the index, we could exclude any SA1s that had any of their variable denominators less than our prescribed cut-off of 6.

### Step 10. Calculating and standardising component/index scores

We derived the first principal component scores for each SA1 by taking the product of each standardised variable with its respective weight, then taking the sum across all variables. Note that the weight for each variable was calculated by dividing the loading by the square root of the eigenvalue.

$$Z_{SA1} = \sum_{j=1}^p \frac{L_j}{\sqrt{\lambda}} \times X_{j,SA1}$$

where

$Z_{SA1}$	=	raw score for the SA1
$X_{j,SA1}$	=	standardised variable of the $j$ -th variable for the SA1
$L_j$	=	loading for the $j$ -th variable
$\lambda$	=	eigenvalue of the principal component
$p$	=	total number of variables in the index

For convenience of presentation, we then rescaled the raw scores to a mean of 1,000 and standard deviation of 100 to create a new set of scores that are the SA1 index scores in SEIFA.

Note that the principal components are arbitrary with respect to their sign (positive or negative), so we set the sign of the weights and loadings so that they make intuitive sense. That is, we gave advantage indicators positive weights and loadings, and disadvantage indicators negative weights and loadings. Accordingly, high scores indicate relative advantage, and low scores indicate relative disadvantage. This is consistent with previous editions of SEIFA.

### Step 11. Creating higher geographic level indexes

We constructed indexes for geographies higher than the SA1 level using population weighted averages of the constituent SA1s. We used the following formula:

$$INDEX_{AREA} = \frac{\sum_{i=1}^n (INDEX_{SA1_i} \times POP_{SA1_i})}{POP_{AREA}}$$

where

$INDEX$	=	Index score for each SA1 or higher level area
$POP$	=	Population for each SA1 or higher level area
$n$	=	Total number of SA1s (with index scores) in the higher level area

The higher level area population is the sum of the populations from the constituent SA1s that received an index score. Populations in excluded SA1s are not included in this calculation.

## SEIFA TECHNICAL PAPER

Although the higher level indexes were constructed from standardised SA1 level indexes, they were not standardised themselves. Therefore the higher level area indexes do not necessarily have a mean of 1,000 or standard deviation of 100. Only SA1s with index scores were used to create the higher level indexes. In a small number of cases, where a higher level area contains a number of SA1s that were excluded, its index score may not be a good representation of its entire population.

For this reason, the output spreadsheets provide the proportion of each higher area level population that was in excluded SA1s. In general, we encourage users conducting analysis at higher level areas to keep in mind that the indexes were constructed at the SA1 level, and to consider using the distribution of SA1s within the higher level areas, rather than just the one index score for each higher level area. This is further discussed in Section 6.3.

### 4.4 TECHNICAL DETAILS OF EACH INDEX: VARIABLES AND LOADINGS

This section gives the results of the principal component analysis carried out for each index, including variable loadings and percentage of variance explained. We also list the variables initially considered for inclusion but removed due to high correlations with other variables or weak loadings.

#### 4.4.1 Index of Relative Socio-Economic Disadvantage

The IRSD summarises variables that indicate relative disadvantage at the SA1 level, according to the concept described in Section 2.2.1. The final variable list and corresponding loadings are shown below in table 4.2.

Table 4.2 Final IRSD variables and loadings

<i>Variable mnemonic</i>	<i>Variable loading</i>	<i>Variable description</i>
INC_LOW	-0.91	% People with stated annual household equivalised income between \$1 and \$25,999 (approx. 1st and 2nd deciles)
CHILDJOBLESS	-0.83	% Families with children under 15 years of age who live with jobless parents
NONET	-0.79	% Occupied private dwellings with no internet connection
NOYR12ORHIGHER	-0.77	% People aged 15 years and over whose highest level of education is Year 11 or lower. Includes Certificate I and II
UNEMPLOYED	-0.75	% People (in the labour force) unemployed
OCC_LABOUR	-0.74	% Employed people classified as 'labourers'
LOWRENT	-0.73	% Occupied private dwellings paying rent less than \$215 per week (excluding \$0 per week)
ONEPARENT	-0.67	% One parent families with dependent offspring only
DISABILITYU70	-0.67	% People aged under 70 who have a long-term health condition or disability and need assistance with core activities
SEPDIVORCED	-0.55	% People aged 15 and over who are separated or divorced
OCC_DRIVERS	-0.54	% Employed people classified as Machinery Operators and Drivers
OCC_SERVICE_L	-0.53	% Employed people classified as Low Skill Community and Personal Service Workers
NOCAR	-0.49	% Occupied private dwellings with no cars
OVERCROWD	-0.46	% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)
NOEDU	-0.43	% People aged 15 years and over who have no educational attainment
ENGLISHPOOR	-0.30	% People who do not speak English well

The 2016 IRSD index explains 43% of the total variance of the variables in the final variable list. The corresponding percentages for previous indexes are: 44% (2011 IRSD), 39% (2006 IRSD) and 33% (2001 IRSD).

## SEIFA TECHNICAL PAPER

### Removal of highly correlated variables

Of the variables considered for the IRSD, there were no two variables that had a correlation coefficient greater than 0.8 in absolute value.

### Removal of low loading variables

Table 4.3 shows the variables that were dropped from the IRSD because their loading was below our prescribed cut-off of 0.3 in absolute value. The variables are shown in the order they were removed, with the loadings from the iteration when they were removed.

Table 4.3 IRSD variables removed due to low loadings

Variable mnemonic	Variable loading	Variable description
FEWBED	-0.16	% Occupied private dwellings with one or no bedrooms
CERTIFICATE	-0.19	% People aged 15 years and over whose highest level of educational attainment is a Certificate III or IV qualification
OCC_SALES_L	-0.27	% Employed people classified as Low Skill Sales

### 4.4.2 Index of Relative Socio-Economic Advantage and Disadvantage

The IRSAD summarises variables that indicate either relative socio-economic advantage or disadvantage, according to the concept described in Section 2.2.2. The final variable list and corresponding loadings are shown below in table 4.4.

Table 4.4 Final IRSAD variables and loadings

Variable mnemonic	Variable loading	Variable description
INC_LOW	-0.89	% People with stated annual household equivalised income between \$1 and \$25,999 (approx. 1st and 2nd deciles)
NOYR12ORHIGHER	-0.85	% People aged 15 years and over whose highest level of education is Year 11 or lower. Includes Certificate I and II
OCC_LABOUR	-0.79	% Employed people classified as 'labourers'
NONET	-0.78	% Occupied private dwellings with no internet connection
CHILDJOBLESS	-0.76	% Families with children under 15 years of age who live with jobless parents
DISABILITYU70	-0.69	% People aged under 70 who have a long-term health condition or disability and need assistance with core activities
UNEMPLOYED	-0.66	% People (in the labour force) unemployed
ONEPARENT	-0.65	% One parent families with dependent offspring only
LOWRENT	-0.64	% Occupied private dwellings paying rent less than \$215 per week (excluding \$0 per week)
OCC_DRIVERS	-0.62	% Employed people classified as Machinery Operators and Drivers
SEPDIVORCED	-0.60	% People aged 15 and over who are separated or divorced
OCC_SERVICE_L	-0.54	% Employed people classified as Low Skill Community and Personal Service Workers
CERTIFICATE	-0.36	% People aged 15 years and over whose highest level of educational attainment is a certificate III or IV qualification
NOEDU	-0.34	% People aged 15 years and over who have no educational attainment
NOCAR	-0.33	% Occupied private dwellings with no cars
OVERCROWD	-0.33	% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)

## SEIFA TECHNICAL PAPER

OCC_SALES_L	-0.32	% Employed people classified as Low Skill Sales
ATUNI	0.36	% People aged 15 years and over at university or other tertiary institution
HIGHBED	0.44	% Occupied private dwellings with four or more bedrooms
HIGHRENT	0.47	% Occupied private dwellings paying rent greater than \$470 per week
OCC_MANAGER	0.47	% employed people classified as Managers
DIPLOMA	0.50	% People aged 15 years and over whose highest level of education attainment is a diploma qualification
OCC_PROF	0.71	% Employed people classified as Professionals
HIGHMORTGAGE	0.72	% Occupied private dwellings paying mortgage greater than \$2,800 per month
INC_HIGH	0.83	% People with stated annual household equivalised income greater than \$78,000 (approx 9th and 10th deciles)

The 2016 IRSAD index explains 38% of the total variance of the variables in the final variable list. The corresponding percentages for previous indexes are: 39% (2011 IRSAD), 44% (2006 IRSAD) and 41% (2001 IRSAD).

### Removal of highly correlated variables

The variable DEGREE had high correlations with NOYR12ORHIGHER (-0.84) and OCC\_PROF (0.91). This suggested that the proportion of people in an area with a degree was explained by other variables in the index. Therefore DEGREE was dropped.

### Removal of low loading variables

Table 4.5 shows the variables dropped from the IRSAD because of low loadings. The variables are shown in the order they were removed, with the loadings from the iteration when they were removed.

Table 4.5 IRSAD variables removed due to low loadings

Variable mnemonic	Variable loading	Variable description
FEWBED	-0.14	% Occupied private dwellings with one or no bedrooms
OWNING	0.22	% Occupied private dwellings owning dwelling without a mortgage
ENGLISHPOOR	-0.23	% People who do not speak English well
SPAREBED	0.28	% Occupied private dwellings with one or no bedrooms
HIGHCAR	0.26	% Occupied private dwellings with three or more cars

### 4.4.3 Index of Economic Resources

The IER focuses on the financial aspects of relative socio-economic advantage and disadvantage, according to the concept described in Section 2.2.3. The final variable list and corresponding loadings are shown below in table 4.6.

Table 4.6 Final IER variables and loadings

Variable mnemonic	Variable loading	Variable description
INC_LOW	-0.77	% People with stated annual household equivalised income between \$1 and \$25,999 (approx. 1st and 2nd deciles)
NOCAR	-0.73	% Occupied private dwellings with no cars
LOWRENT	-0.72	% Occupied private dwellings paying rent less than \$215 per week (excluding \$0 per week)
LONE	-0.66	% Occupied private dwellings who are lone person occupied private dwellings
ONEPARENT	-0.63	% One parent families with dependent offspring only

## SEIFA TECHNICAL PAPER

UNEMPLOYED1	-0.54	% People aged 15 years and over who are unemployed
OVERCROWD	-0.51	% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)
GROUP	-0.37	% Occupied private dwellings who are group occupied private dwellings
OWNING	0.36	% Occupied private dwellings owning dwelling without a mortgage
UNINCORP	0.52	% Dwellings with at least one person who is an owner of an unincorporated enterprise
INC_HIGH	0.55	% People with stated annual household equivalised income greater than \$78,000 (approx 9th and 10th deciles)
MORTGAGE	0.67	% Occupied private dwellings owning dwelling (with a mortgage)
HIGHMORTGAGE	0.68	% Occupied private dwellings paying mortgage greater than \$2,800 per month
HIGHBED	0.74	% Occupied private dwellings with four or more bedrooms

The 2016 IER index explains 38% of the total variance of the variables in the final variable list. The corresponding percentages for previous indexes are: 39% (2011 IER) and 35% (2006 IER).

### **Removal of highly correlated variables**

No variables were dropped based on high correlations.

### **Removal of low loading variables**

Table 4.7 shows the variable dropped from the IER because of a low loading.

Table 4.7 IER variable removed due to low loadings

<i>Variable mnemonic</i>	<i>Variable loading</i>	<i>Variable description</i>
HIGHRENT	-0.07	% Occupied private dwellings paying rent greater than \$470 per week

#### 4.4.4 Index of Education and Occupation

The IEO summarises variables related to educational qualifications and vocational skills, according to the concept described in Section 2.2.4. The final variable list and corresponding loadings are shown below in table 4.8.

Table 4.8 Final IEO variables and loadings

<i>Variable mnemonic</i>	<i>Variable loading</i>	<i>Variable description</i>
NOYR12ORHIGHER	-0.87	% People aged 15 years and over whose highest level of education is Year 11 or lower. Includes Certificate I and II
OCC_SKILL5	-0.81	% Employed people who work in a Skill Level 5 occupation
OCC_SKILL4	-0.77	% Employed people who work in a Skill Level 4 occupation
CERTIFICATE	-0.55	% People aged 15 years and over whose highest level of educational attainment is a certificate III or IV qualification
UNEMPLOYED	-0.55	% People (in the labour force) unemployed
NOEDU	-0.31	% People aged 15 years and over who have no educational attainment
OCC_SKILL2	0.35	% Employed people who work in a Skill Level 2 occupation
ATUNI	0.48	% People aged 15 years and over at university or other tertiary institution
DIPLOMA	0.51	% People aged 15 years and over whose highest level of education attainment is a diploma qualification
OCC_SKILL1	0.89	% Employed people who work in a Skill Level 1 occupation

## SEIFA TECHNICAL PAPER

The 2016 IEO index explains 41% of the total variance of the variables in the final variable list. The corresponding percentages for previous indexes are: 47% (2011 IEO), 52% (2006 IEO) and 46% (2001 IEO).

### Removal of highly correlated variables

DEGREE (% People aged 15 years and over with a degree or higher qualification) had high correlations with NOYR12ORHIGHER (-0.84) and OCC\_SKILL1 (0.83). It was decided that the proportion of people with a degree was already well explained by the index, and DEGREE was removed.

The OCC\_SKILL1 variable had high correlations with OCC\_SKILL2 (0.84) and OCC\_SKILL4 (-0.81). However, we decided to retain this variable in the list of candidate variables, as this may help with the interpretability of the IEO index.

### Removal of low loading variables

Table 4.9 shows the variable dropped from the IEO because of a low loading.

Table 4.9 IER variable removed due to low loadings

Variable mnemonic	Variable loading	Variable description
ATSCHOOL	0.01	% People aged 15 years and over who are still attending secondary school

### 4.4.5 Summary of variables included in indexes

Table 4.10 below shows the final set of variables included in each index.

Table 4.10 List of variables in each index, by socio-economic dimension

Dimension	Index of Relative Socio-Economic Disadvantage	Index of Relative Socio-Economic Advantage and Disadvantage	Index of Economic Resources	Index of Education and Occupation
Income	INC_LOW	INC_HIGH INC_LOW	INC_HIGH INC_LOW	
Education	NOYR12ORHIGHER NOEDU	NOYR12ORHIGHER NOEDU CERTIFICATE ATUNI DIPLOMA		NOYR12ORHIGHER NOEDU CERTIFICATE ATUNI DIPLOMA UNEMPLOYED
Employment	UNEMPLOYED	UNEMPLOYED	UNEMPLOYED1	
Occupation	OCC_LABOUR OCC_DRIVERS OCC_SERVICE_L	OCC_LABOUR OCC_DRIVERS OCC_SERVICE_L OCC_SALES_L OCC_MANAGER OCC_PROF		OCC_SKILL1 OCC_SKILL2 OCC_SKILL4 OCC_SKILL5
Housing	LOWRENT OVERCROWD	LOWRENT OVERCROWD HIGHRENT HIGHBED HIGHMORTGAGE	LOWRENT OVERCROWD OWNING MORTGAGE HIGHBED HIGHMORTGAGE	
Other	CHILDJOBLESS ONEPARENT DISABILITYU70 ENGLISHPOOR NOCAR SEPDIVORCED NONET	CHILDJOBLESS ONEPARENT DISABILITYU70 NOCAR SEPDIVORCED NONET	UNINCORP ONEPARENT LONE GROUP NOCAR	

### 4.5 DISTRIBUTIONS OF THE INDEXES

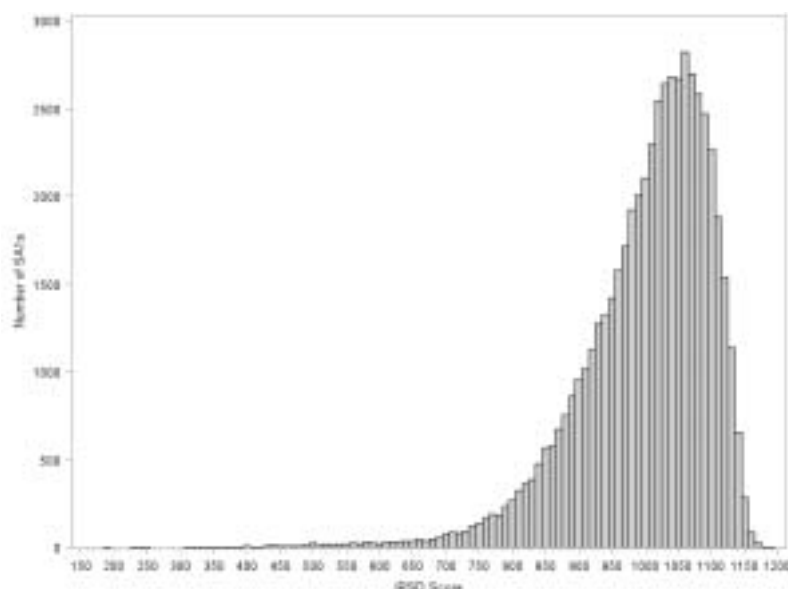
This section presents frequency histograms for each index at the SA1 level. The index distributions have generally similar shapes to those from SEIFA 2011.

#### 4.5.1 Index of Relative Socio-Economic Disadvantage

The IRSD distribution shown in figure 4.11 has a very long left tail. The values range from about 188 to 1186. This index contains only disadvantage indicators, so there is more scope to distinguish between disadvantaged areas than advantaged areas.

The steep peak for this distribution means that there will be little difference in the scores of SA1s in the middle deciles, and so the characteristics related to the IRSD variables may not vary much across SA1s in these middle deciles.

Figure 4.11 IRSD score distribution

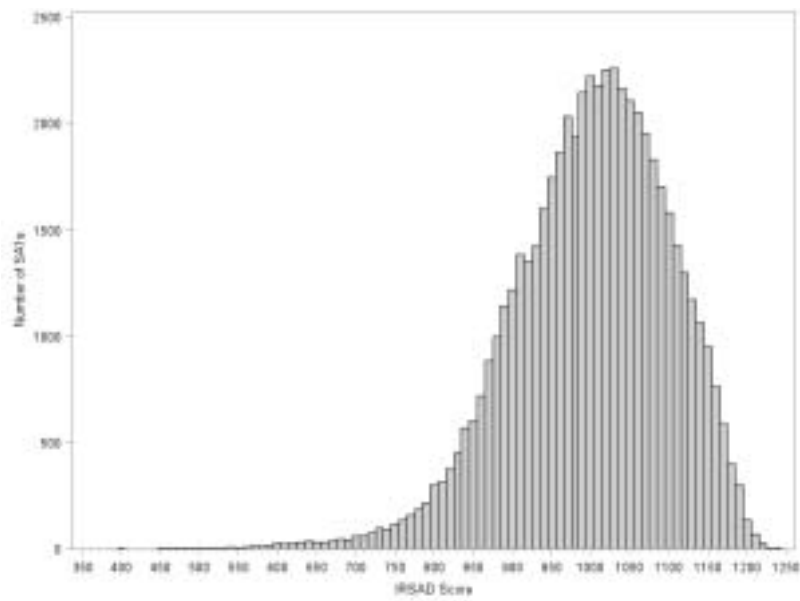


#### 4.5.2 Index of Relative Socio-Economic Advantage and Disadvantage

The scores for IRSAD range from 400 to 1239. The right-hand slope is not as steep in the IRSAD distribution as it is in the IRSD distribution. This means that the IRSAD scores of SA1s in the upper deciles are more spread out than the IRSD scores in these deciles, and this index has a greater ability to differentiate between the more advantaged areas.



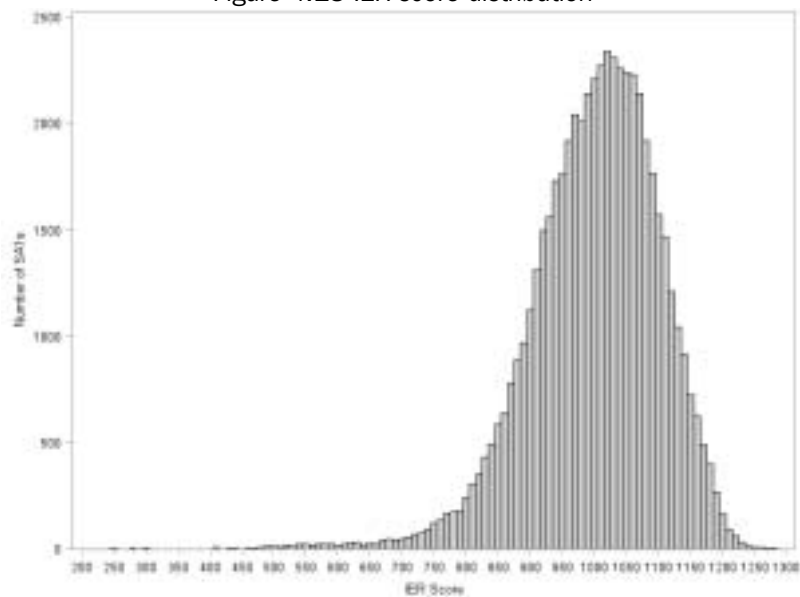
Figure 4.12 IRSAD score distribution



#### 4.5.3 Index of Economic Resources

The scores for IER range from 245 to 1281.

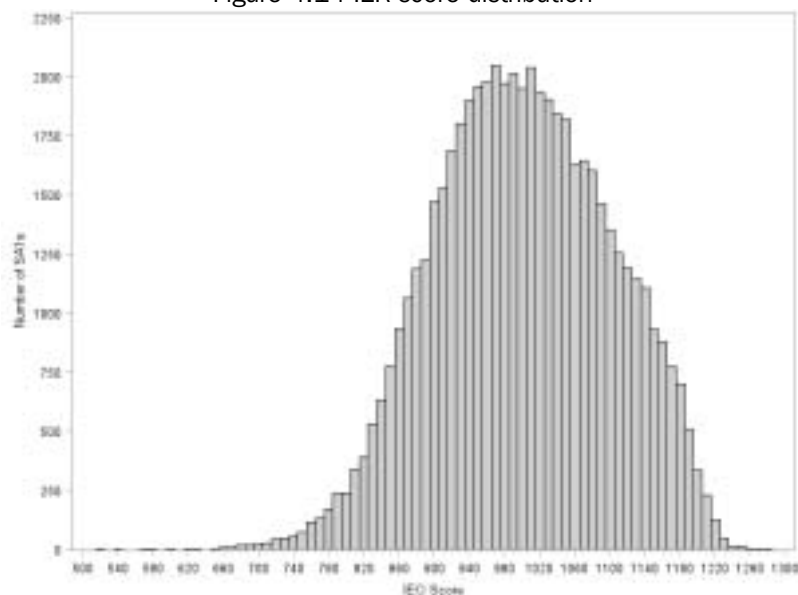
Figure 4.13 IER score distribution



#### 4.5.4 Index of Education and Occupation

The scores for IEO range from 523 to 1283.

Figure 4.14 IER score distribution



#### 4.6 BASIC OUTPUT: SCORES, RANKS, DECILES, AND PERCENTILES

The output is presented in spreadsheets and is available for free at [www.abs.gov.au](http://www.abs.gov.au) under the catalogue number 2033.0.55.001.

##### 4.6.1 Scores

The scores are a weighted combination of the selected indicators of advantage and disadvantage which have been standardised to a distribution with a mean of 1000 and standard deviation of 100. An area with all of its indicators equal to the national average will receive a score of 1000. The score for an area will increase if an area has: an indicator of advantage that is greater than the national average; or an indicator of disadvantage that is less than the national average. Conversely, the score for an area will decrease if an area has: an indicator of disadvantage that is greater than the national average; or an indicator of advantage that is less than the national average. Indicators which are further away from the national average have a larger impact on the score.

For areas larger than SA1, the scores are a population weighted average of constituent SA1 scores, as described in Step 11 of Section 4.3.

It is important to remember that the scores are an ordinal measure (discussed in more detail in Section 6.1.2), 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.

##### 4.6.2 Ranks, Deciles, and Percentiles

Using the scores, other measures are derived that are easier to interpret and more appropriate to use in many situations. We have calculated ranks, deciles, and percentiles and included these in the output spreadsheets. These measures are defined below:

## SEIFA TECHNICAL PAPER

---

**Rank** – The areas are ranked in order of their score, from lowest to highest, with rank 1 representing the most disadvantaged area. Note that in the spreadsheets, rankings are provided on a national basis and also a state/territory basis. Note that the same set of scores is used for each ranking – the scores are not recalculated for each state/territory.

**Deciles** – All areas are ordered from lowest to highest score, the lowest 10% of areas are given a decile number of 1, the next 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. This means that areas are divided up into ten equal sized groups, depending on their score.

**Percentiles** – All areas are ordered from lowest to highest score, the lowest 1% of areas are given a percentile number of 1, the next 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. This means that areas are divided up into one hundred equal sized groups, depending on their score. Sometimes deciles and percentiles are referred to generally as quantiles. Other commonly used quantiles include quintiles and quartiles, although we have not included these in the output spreadsheets. They can be easily derived using the percentiles.

### 4.7 GEOGRAPHIC OUTPUT LEVELS FOR SEIFA 2016

The primary unit of analysis and the smallest area for which the indexes are available is the Statistical Area Level 1 (SA1). This is the recommended unit of analysis for SEIFA 2016.

For a selection of geographic areas larger than SA1, scores have been calculated by taking population-weighted averages of constituent SA1 scores. The output spreadsheets also contain some information about the distribution of SA1 index scores within larger areas. This enables users to consider the socio-economic diversity that can exist within a larger area.

Table 4.15 below summarises the output available at the different geographic levels.

Table 4.15 Geographic output summary for SEIFA 2016

---

<i>Geographic unit</i>	<i>Index score</i>	<i>SA1 distribution information</i>
Statistical Area level 1 (SA1)	Yes	N/A
Statistical Area level 2 (SA2)	Yes	Yes
Statistical Area level 3 (SA3)	No	Yes
Statistical Area level 4 (SA4)	No	Yes
Local Government Area (LGA)	Yes	Yes
State Suburb (SSC)	Yes	Yes
Postal Area (POA)	Yes	Yes
Commonwealth Electoral Division (CED)	No	Yes
State Electoral Division (SED)	No	Yes

## SEIFA TECHNICAL PAPER

---

Note – for the geographies larger than SA1, and not in the ASGS (LGAs, SSCs and POAs), a best fit correspondence of SA1s to the larger geographies was used. Local Government Areas (LGAs), State Suburbs (SSCs) and Postal Areas (POAs) are constructed from Mesh Blocks in the 2016 version of the ASGS. In some cases, particularly for certain SSCs with small populations, the SA1 boundaries do not correspond closely to the higher level area. For this reason, SEIFA scores for SSCs and POAs with small populations should be used with caution, as the scores may have been calculated from populations that do not correspond closely with the actual population in the area. The ABS Maps product on the ABS website may be useful for identifying areas that do not correspond closely to the SA1 structure – see <http://stat.abs.gov.au/itt/r.jsp?ABSMaps>.

The output spreadsheets contain specific references to the ABS publications from which the geography classifications and correspondences have been sourced.

## 5. VALIDATION OF THE INDEXES

Once the indexes are calculated, they are checked to ensure that they are measuring the desired concept and that the results generally make sense. This validation is important to establish the credibility of the indexes and identify any issues that may have been missed in the construction of the indexes. The methods used to validate SEIFA 2016 include:

- comparison of SEIFA 2016 rankings with 2011 rankings
- identification of the drivers of change from SEIFA 2011 to 2016, and
- seeking review from internal and external experts.

### 5.1 RELATIONSHIPS BETWEEN THE INDEXES

We examined SEIFA for internal consistency by looking at the correlations between the indexes. Table 5.1 shows the rank correlation matrix. All correlations are in the expected directions and show significant relationships. The IRSD is very highly correlated with the IRSAD (0.97).

Table 5.1 Spearman's rank correlation matrix

---

<i>Dimension</i>	<i>IRSD</i>	<i>IRSAD</i>	<i>IER</i>	<i>IEO</i>
<b>IRSD</b>	<b>1.00</b>			
<b>IRSAD</b>	<b>0.97</b>	<b>1.00</b>		
<b>IER</b>	<b>0.83</b>	<b>0.76</b>	<b>1.00</b>	
<b>IEO</b>	<b>0.83</b>	<b>0.92</b>	<b>0.52</b>	<b>1.00</b>

The indexes that measure specific dimensions of advantage and disadvantage (IER and the IEO) have a lower correlation with the other indexes. The IER includes variables associated with high and low wealth that are not included in the other indexes. The IEO focuses solely on educational qualifications, employment and vocational skills.

The IER and the IEO are positively correlated, but the correlation is much weaker than between the other indexes (0.52). There is a significant difference between the concepts measured by these two indexes, and they do not share any common variables.

### 5.2 COMPARING 2011 AND 2016 RANKINGS

The SA1 scores from 2016 were checked against comparable areas from 2011, where possible, to identify areas with large changes and determine whether these changes were plausible. Some changes are to be expected, particularly in areas with high population growth and areas that have been affected by economic changes in the region. This process did not identify any results that seemed unrealistic.

### 5.3 VALIDATION OF HIGHER LEVEL AREA INDEXES

Most of the validation was focused on the SA1 level indexes because SA1s are the primary unit of analysis and indexes for higher level areas (e.g. SA2) are population weighted averages of the SA1 scores. However, we conducted basic validation checks on any higher level area indexes that we produced.

### 6. USING AND INTERPRETING SEIFA

This section provides some advice and information to assist in the appropriate use of SEIFA, and to help users gain the most value from the product.

#### 6.1 BROAD GUIDELINES ON APPROPRIATE USE

##### 6.1.1 Area level indexes

The indexes are assigned to areas, not to individuals. They indicate the collective socio-economic characteristics of the people living in an area. A relatively disadvantaged area is likely to have a high proportion of relatively disadvantaged people. However, such an area is also likely to contain some people who are relatively advantaged. When area level indexes are used as proxy measures of individual level socio-economic advantage and disadvantage, many people are likely to be misclassified. This is known as the ecological fallacy. Wise and Mathews (2011) conducted an investigation into the extent of this issue as it relates to SEIFA.

##### 6.1.2 Ordinal indexes

As measures of socio-economic level, the indexes are best interpreted as ordinal measures. They can be used to rank areas, and are also useful to understand the distribution of socio-economic conditions across different areas. Also, the index scores are on an arbitrary numerical scale. The scores do not represent some quantity of advantage or disadvantage. For example, we cannot infer that an area with an index value of 1000 is twice as advantaged as an area with an index value of 500.

For ease of interpretation, we generally recommend using the index rankings and quantiles (e.g. deciles) for analysis, rather than using the index scores. Index scores are still provided in the output, and can still be used for analysis when appropriate.

For more information on index scores, rankings, and quantiles, see Section 4.6.

##### 6.1.3 Importance of the underlying variables

Each index is constructed based on a weighted combination of selected variables. The indexes are dependent on the set of variables chosen for the analysis. A different set of underlying variables would result in a different index. At the same time, because of the large number of variables in each index, removing or altering one variable will not usually have a large effect.

Users should consider the aspect of socio-economic advantage and disadvantage in which they are interested, and examine the underlying set of variables in each index (see Sections 3 and 4). This will allow them to make an informed decision on whether an index is appropriate for their particular purpose. Section 6.2 provides some tips on choosing which of the four indexes to use.

### 6.1.4 Issues with longitudinal or time series analysis

The indexes are designed to compare the relative socio-economic characteristics of areas at a given point in time, not to compare individual areas across time. When considering longitudinal or time series analysis using indexes from different Census years, there are a number of issues that potentially make the analysis very difficult to interpret:

- The constituent variables and variable weights for the index are likely to have changed.
- The boundaries and numbers of relevant small areas may have changed.
- The distribution of the standardised index values will have changed (e.g. a score of 800 does not represent the same level of disadvantage in different years).
- There are changes in the way the variables are defined. For example, from SEIFA 2006 onwards, the indexes have been calculated using the characteristics of an area's usual residents, rather than those of the people in the area on Census Night.

For these reasons, it can be very difficult to perform useful longitudinal or times series analysis, and it should not be attempted without due consideration of the issues. If comparisons over time are being made, we recommend the use of quantiles (e.g. deciles) rather than ranks or scores.

### 6.2 CHOICE OF INDEX

Depending on the aim or context of the analysis, one of the SEIFA indexes may be more appropriate than the others. Below are some aspects to be considered:

- The concept and variables underlying each index. The concepts behind each index are described in Section 2.2. The final variable lists for each index are in Section 4.4.
- The degree to which the four indexes are correlated with each other – this is discussed in Section 5.1.
- The IRSD ranks areas on a continuum from most disadvantaged to least disadvantaged, while the other three indexes (IRSAD, IER, IEO) rank areas on a continuum from most disadvantaged/least advantaged to most advantaged/least disadvantaged.
- The IRSD and IRSAD are more general measures in the sense that they summarise variables from a wider range of socio-economic dimensions. The IER and IEO are more targeted measures aimed at capturing narrower concepts.
- Simpler measures, such as income or employment status, may be more appropriate than SEIFA for some analysis. For an in-depth discussion on choosing a socio-economic measure, see ABS (2011b).

### 6.3 USING INDEX SCORES FOR AREAS LARGER THAN SA1

As discussed in Section 6.1.1, the fact the indexes are area level measures means that they will mask some diversity at finer levels of disaggregation. In some applications of the indexes, it may be important to identify diversity of socio-economic characteristics within areas.

When using an index at a geographic level higher than SA1 (e.g. SA2s and LGAs), we do have some scope to assess the diversity within that area by looking at its constituent SA1s. There is further discussion about assessing diversity within areas in Wise and Mathews (2011) and Radisich and Wise (2012). The second paper also proposes an additional measure that can be used to identify diverse larger areas. This measure is called the 'SA1-concentration score' and can identify the presence of disadvantaged SA1s within an overall advantaged large area.

To enable the analyses described above, an additional type of output has been released for SEIFA 2016. For all geographic levels higher than SA1 for which index scores are released, the corresponding SA1 distributions within those areas have been presented in spreadsheets.

As noted previously, SEIFA scores for SSCs and POAs with small populations should be used with caution, because the SA1 boundaries may not correspond closely to the higher level area. For more information, see section 4.7.

### 6.4 MAPPING THE INDEXES

Maps of the indexes are an excellent way of observing the spatial distribution of relative socio-economic advantage and disadvantage. Interactive maps of the SEIFA 2016 indexes are available on the SEIFA 2016 page on the ABS website.

### 6.5 USING THE INDEXES AS CONTEXTUAL VARIABLES IN SOCIAL ANALYSIS

SEIFA index scores are commonly merged onto a person level dataset based on the area in which that person resides. The indexes can then be used to help investigate the relationship between disadvantage or advantage and other variables of interest. This type of analysis can yield some very interesting findings; however, it is important to interpret the findings correctly. Some interpretive issues are discussed below.

A SEIFA index refers to the area in which a person lives. It is a contextual variable. It is incorrect to say that a person is very disadvantaged just because they live in a very disadvantaged area. It is true that living in a very disadvantaged area may disadvantage them to a certain extent, but it is possible that they are advantaged in other respects such as having a good education and earning a high income, and are therefore not typical of other residents in that area. The issue of diversity of individuals within areas is further investigated and discussed in Wise and Mathews (2011).

It is desirable to use the smallest geographic unit possible when merging an index to another dataset. In the case of SEIFA 2016, the SA1 is the smallest unit available, and if possible, SA1s should be derived on the dataset to which SEIFA scores are being appended.

### 6.6 AREA-BASED QUANTILES VERSUS POPULATION-BASED QUANTILES

In this paper the word 'quantiles' is used to collectively describe measures such as percentiles and deciles.

In the spreadsheets in which the indexes are presented, quantiles (percentiles and deciles) are presented in addition to the index scores and rankings, as described in Section 4.6. These quantiles are calculated based on dividing the number of areas into equal groups. These are called area-based quantiles.

An alternative way of defining the quantiles is to divide them into equal groups based on the number of people living in those areas. The quantiles would then contain an equal number of people (or at least as can be best achieved) in each group, rather than an equal number of areas. These are called population-based quantiles.



## SEIFA TECHNICAL PAPER

---

The ABS publishes area-based quantiles because they are easier to interpret, since SEIFA is an area-based measure. They also serve most analytical purposes. There are some instances in which the use of population-based quantiles is appropriate. Users can create their own population-based quantiles using information already available in the output spreadsheets. Population-based deciles are also available in TableBuilder. As mentioned above, population-based quantiles can be difficult to interpret, so users should take care in how they are applied. The population-based quantiles represent groups of individuals who live in similarly ranked areas, as opposed to groups of similarly ranked individuals.

### REFERENCES

- Australian Bureau of Statistics (2007) *Australian Social Trends, 2007*, cat. no. 4102.0 ABS, Canberra.
- (2011a) *Education and Work Australia, May 2011*, cat. no. 6227.0, ABS, Canberra.
- (2011b) *Information Paper: Measures of Socioeconomic Status*, cat. no. 1244.0.55.001, ABS, Canberra.
- (2013) ANZSCO – *Australian and New Zealand Standard Classification of Occupations, Version 1.2*, cat. no. 1220.0, ABS, Canberra.
- (2016a) *Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016*, cat. no. 1270.0.55.001, ABS, Canberra.
- (2016b) *Australian Statistical Geography Standard (ASGS): Volume 3 - Non ABS Structures*, cat. no. 1270.0.55.003, ABS, Canberra.
- (2016c) *Census of Population and Housing: Census Dictionary, 2016*, cat. no. 2901.0, ABS, Canberra.
- (2017) *Housing Occupancy and Costs, Australia, 2015–16*, cat. no. 4130.0, ABS, Canberra.
- Jolliffe, I.T. (1986) *Principal Component Analysis*, Springer–Verlag, New York.
- O’Rourke, N.; Hatcher, L. and Stepanski, E.J. (2005) *A Step-by-Step Approach to Using SAS for Univariate and Multivariate Statistics, Second Edition*, SAS Institute Inc., Cary, NC.
- Radisich, P. and Wise, P. (2012) “Socio-Economic Indexes For Areas: Robustness, Diversity Within Larger Areas and the New Geography Standard”, *Methodology Research Papers*, cat. no. 1351.0.55.038, Australian Bureau of Statistics, Canberra.
- Wise, P. and Mathews, R. (2011) “Socio-Economic Indexes For Areas: Getting a Handle on Individual Diversity Within Areas”, *Methodology Research Papers*, cat. no. 1351.0.55.036, Australian Bureau of Statistics, Canberra.

## APPENDIX

### VARIABLE SPECIFICATIONS

This appendix gives descriptions of each variable considered for inclusion in one of the 2016 indexes. The description of the variable proportion is followed by two bullet points; the first is a description of the numerator, the second is a description of the denominator. The square brackets contain specifications for creating the numerator/denominator from Census data items, according to the mnemonics used in the Census of Population and Housing: Census Dictionary, 2016 (ABS, 2016). The variables are arranged by socio-economic dimension.

Notes:

- The Skill Level for each occupation can be found in *ANZSCO – Australian and New Zealand Standard Classification of Occupations, Version 1.2* (ABS, 2013)
- Household composition was ‘not classifiable’ if the household: contained only visitors or persons aged under 15 years on Census night; or was determined to be occupied on Census Night but the collector could not make contact; or could not be classified because there was insufficient information on the Census form.
- The Canadian National Occupancy Standard determines housing appropriateness, using the number of bedrooms and the number, age, sex and relationships of household members. For more information refer to *Housing Occupancy and Costs, Australia, 2015–16* (ABS, 2017)

#### Income variables

<b>INC_LOW</b>	<p><b>% PEOPLE WITH STATED ANNUAL HOUSEHOLD EQUIVALISED INCOME BETWEEN \$1 AND \$25,999 (approx. 1st and 2nd deciles)</b></p> <ul style="list-style-type: none"> <li>▪ number of people living in classifiable occupied private dwellings with stated annual household equivalised income between \$1 and \$25,999 [HIED = 02–05]</li> <li>▪ number of people living in classifiable occupied private dwellings with stated household \equivalised income [HIED = 01–15]</li> </ul>
<b>INC_HIGH</b>	<p><b>% PEOPLE WITH STATED ANNUAL HOUSEHOLD EQUIVALISED INCOME GREATER THAN \$78,000 (approx. 9th and 10th deciles)</b></p> <ul style="list-style-type: none"> <li>▪ number of people living in classifiable occupied private dwellings with stated annual household equivalised income greater than \$78,000 [HIED = 11–15]</li> <li>▪ number of people living in classifiable occupied private dwellings with stated household equivalised income [HIED = 01–15]</li> </ul>

#### Education variables

<b>ATSCHOOL</b>	<p><b>% PEOPLE AGED 15 YEARS AND OVER WHO ARE STILL ATTENDING SECONDARY SCHOOL</b></p> <ul style="list-style-type: none"> <li>▪ number of people aged 15 years and over who are still attending secondary school [AGEP &gt; 14 and TYPP = 31, 32 33]</li> <li>▪ number of people aged 15 years and over (excluding educational institution attendance not stated) [AGEP &gt; 14 and TYPP ne &amp;&amp;, VV]</li> </ul>
<b>ATUNI</b>	<p><b>% PEOPLE AGED 15 YEARS AND OVER AT A UNIVERSITY OR OTHER TERTIARY INSTITUTION</b></p> <ul style="list-style-type: none"> <li>▪ number of people aged 15 years and over at university or other tertiary institution [AGEP &gt; 14 and TYPP = 50]</li> <li>▪ number of people aged 15 years and over (excluding educational institution attendance not stated) [AGEP &gt; 14 and TYPP ne &amp;&amp;, VV]</li> </ul>
<b>CERTIFICATE</b>	<p><b>% PEOPLE AGED 15 YEARS AND OVER WHOSE HIGHEST LEVEL OF EDUCATION IS A CERTIFICATE III or IV QUALIFICATION</b></p> <ul style="list-style-type: none"> <li>▪ number of people aged 15 years and over with a certificate III or IV qualification [HEAP = 51]</li> <li>▪ number of people aged 15 years and over (excluding highest level of education not stated) [HEAP ne 001, @@@, VVV, &amp;&amp;&amp;]</li> </ul>

## SEIFA TECHNICAL PAPER

---

DEGREE	<p>% PEOPLE AGED 15 YEARS AND OVER WHOSE HIGHEST LEVEL OF EDUCATION IS A BACHELOR DEGREE OR HIGHER</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over whose highest level of education is a bachelor degree or higher [HEAP = 1–3]</li><li>number of people aged 15 years and over (excluding highest level of education not stated) [HEAP ne 001, @@@, VVV, &amp;&amp;&amp;]</li></ul>
DIPLOMA	<p>% PEOPLE AGED 15 YEARS AND OVER WHOSE HIGHEST LEVEL OF EDUCATION IS AN ADVANCED DIPLOMA OR DIPLOMA</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over whose highest level of education is an advanced diploma or diploma [HEAP = 4]</li><li>number of people aged 15 years and over (excluding highest level of education not stated) [HEAP ne 001, @@@, VVV, &amp;&amp;&amp;]</li></ul>
NOEDU	<p>% PEOPLE AGED 15 YEARS AND OVER WHO HAVE NO EDUCATIONAL ATTAINMENT</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over whose highest level of education is no educational attainment [HEAP = 998]</li><li>number of people aged 15 years and over (excluding highest level of education not stated) [HEAP ne 001, @@@, VVV, &amp;&amp;&amp;]</li></ul>
NOYEAR12ORHIGHER	<p>% PEOPLE AGED 15 YEARS AND OVER WHOSE HIGHEST LEVEL OF EDUCATION IS YEAR 11 OR LOWER</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over whose highest level of education is year 11 or lower (includes certificate I and II qualifications; excludes those still at secondary school) [HEAP = 613, 621, 720, 721, 811, 812, 998 and TYPP ne 31, 32, 33]</li><li>number of people aged 15 years and over (excluding highest level of education not stated) [HEAP ne 001, @@@, VVV, &amp;&amp;&amp;]</li></ul>

### Employment variables

UNEMPLOYED	<p>% PEOPLE (IN THE LABOUR FORCE) WHO ARE UNEMPLOYED</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over who are unemployed and looking for work [LFSP = 4–5]</li><li>number of people aged 15 years and over in the labour force [LFSP = 1–5]</li></ul>
UNEMPLOYED1	<p>% PEOPLE AGED 15 YEARS AND OVER WHO ARE UNEMPLOYED</p> <ul style="list-style-type: none"><li>number of people aged 15 years and over who are unemployed and looking for work [LFSP = 4–5]</li><li>number of people aged 15 years and over (excluding labour force status not stated) [LFSP = 1–6]</li></ul>

### Occupation variables

OCC_DRIVERS	<p>% EMPLOYED PEOPLE CLASSIFIED AS MACHINERY OPERATORS AND DRIVERS</p> <ul style="list-style-type: none"><li>number of employed people classified as Machinery Operators and Drivers [OCCP = 7]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_LABOUR	<p>% EMPLOYED PEOPLE CLASSIFIED AS LABOURERS</p> <ul style="list-style-type: none"><li>number of employed people classified as Labourers [OCCP = 8]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_MANAGER	<p>% EMPLOYED PEOPLE CLASSIFIED AS MANAGERS</p> <ul style="list-style-type: none"><li>number of employed people classified as Managers [OCCP = 1]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_PROF	<p>% EMPLOYED PEOPLE CLASSIFIED AS PROFESSIONALS</p> <ul style="list-style-type: none"><li>number of employed people classified as Professionals [OCCP = 2]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_SALES_L	<p>% EMPLOYED PEOPLE CLASSIFIED AS LOW-SKILL SALES WORKERS</p> <ul style="list-style-type: none"><li>number of employed people classified as Low-Skill Sales Workers [OCCP = 6 and Skill Level = 5]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_SERVICE_L	<p>% EMPLOYED PEOPLE CLASSIFIED AS LOW-SKILL COMMUNITY AND PERSONAL SERVICE WORKERS</p> <ul style="list-style-type: none"><li>number of employed people classified as Low-Skill Community and Personal Service Workers [OCCP = 4 and Skill Level = 4–5]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>
OCC_SKILL1	<p>% EMPLOYED PEOPLE WHO WORK IN A SKILL LEVEL 1 OCCUPATION</p> <ul style="list-style-type: none"><li>number of employed people who work in a Skill Level 1 occupation [Skill Level = 1]</li><li>number of employed people with a stated occupation [OCCP = 1–8]</li></ul>

## SEIFA TECHNICAL PAPER

---

- OCC\_SKILL2** % EMPLOYED PEOPLE WHO WORK IN A SKILL LEVEL 2 OCCUPATION
- number of employed people who work in a Skill Level 2 occupation [Skill Level = 2]
  - number of employed people with a stated occupation [OCCP = 1–8]
- OCC\_SKILL4** % EMPLOYED PEOPLE WHO WORK IN A SKILL LEVEL 4 OCCUPATION
- number of employed people who work in a Skill Level 4 occupation [Skill Level = 4]
  - number of employed people with a stated occupation [OCCP = 1–8]
- OCC\_SKILL5** % EMPLOYED PEOPLE WHO WORK IN A SKILL LEVEL 5 OCCUPATION
- number of employed people who work in a Skill Level 5 occupation [Skill Level = 5]
  - number of employed people with a stated occupation [OCCP = 1–8]

### *Housing variables*

- FEWBED** % CLASSIFIABLE OCCUPIED PRIVATE DWELLINGS WITH ONE OR NO BEDROOMS
- number of classifiable occupied private dwellings with one or no bedrooms [BEDD = 00,01 and HHCD = 11–32]
  - number of classifiable occupied private dwellings with a stated number of bedrooms [BEDD ne &&, @@ and HHCD = 11–32]
- GROUP** % OCCUPIED PRIVATE DWELLINGS THAT ARE GROUP OCCUPIED PRIVATE DWELLINGS
- number of classifiable occupied private dwellings that are occupied by group households [HHCD = 32 and HHCD = 11–32]
  - number of classifiable occupied private dwellings [HHCD = 11–32]
- HIGHBED** % OCCUPIED PRIVATE DWELLINGS WITH FOUR OR MORE BEDROOMS
- number of classifiable occupied private dwellings with four or more bedrooms [BEDD = 04–30 and HHCD = 11–32]
  - number of classifiable occupied private dwellings with a stated number of bedrooms [BEDD ne &&, @@ and HHCD = 11–32]
- HIGHMORTGAGE** % OCCUPIED PRIVATE DWELLINGS PAYING MORE THAN \$2,800 PER MONTH IN MORTGAGE REPAYMENTS
- number of mortgaged classifiable occupied private dwellings with monthly mortgage repayments greater than \$2,800 [MRED = 2801–9999, HHCD = 11–32 and TEND = 1–7]
  - number of classifiable occupied private dwellings (excluding those with tenure not stated, mortgage not stated and rent not stated) [TEND ne &, @, MRED ne &&&&, RNTD ne &&&& and HHCD = 11–32]
- HIGHRENT** % OCCUPIED PRIVATE DWELLINGS PAYING MORE THAN \$470 PER WEEK IN RENT
- number of rented classifiable occupied private dwellings with rent payments greater than \$470 per week [RNTD = 471–9999, HHCD = 11–32 and TEND = 1–7]
  - number of classifiable occupied private dwellings (excluding those with tenure not stated, mortgage not stated and rent not stated) [TEND ne &, @, MRED ne &&&&, RNTD ne &&&& and HHCD = 11–32]
- LOWRENT** % OCCUPIED PRIVATE DWELLINGS PAYING LESS THAN \$215 PER WEEK IN RENT (EXCLUDING \$0 PER WEEK)
- number of rented classifiable occupied private dwellings with rent payments less than \$215 per week (excluding rent-free and renting from employer) [RNTD = 1–214 and HHCD = 11–32 and LLDD ne 51, 52]
  - number of classifiable occupied private dwellings (excluding those with tenure not stated, mortgage not stated and rent not stated) [TEND ne &, @, MRED ne &&&&, RNTD ne &&&& and HHCD = 11–32]
- OVERCROWD** % OCCUPIED PRIVATE DWELLINGS REQUIRING ONE OR MORE EXTRA BEDROOMS (BASED ON CANADIAN NATIONAL OCCUPANCY STANDARD)
- number of classifiable occupied private dwellings needing one or more extra bedrooms (based on Canadian National Occupancy Standard) [HOSD = 01-04 and HHCD = 11–32]
  - number of classifiable occupied private dwellings (excluding dwellings where housing utilisation cannot be determined or is not stated) [HOSD ne 10, &&, @@ and HHCD = 11–32]
- OWNING** % OCCUPIED PRIVATE DWELLINGS OWNING THE DWELLING THEY OCCUPY (WITHOUT A MORTGAGE)
- number of households owning the dwelling they occupy without a mortgage [TEND = 1 and HHCD = 11–32]
  - number of classifiable occupied private dwellings (excluding tenure not stated) [TEND ne &, @ and HHCD = 11–32]
- MORTGAGE** % OCCUPIED PRIVATE DWELLINGS OWNING THE DWELLING THEY OCCUPY (WITH A MORTGAGE)
- number of mortgaged classifiable occupied private dwellings [TEND = 2, 3, 6 and HHCD = 11–32]
  - number of classifiable occupied private dwellings (excluding tenure not stated) [TEND ne &, @ and HHCD = 11–32]

## SEIFA TECHNICAL PAPER

---

**SPAREBED** % OCCUPIED PRIVATE DWELLINGS WITH ONE OR MORE SPARE BEDROOMS (BASED ON CANADIAN NATIONAL OCCUPANCY STANDARD)

- number of classifiable occupied private dwellings with one or more spare bedrooms (based on Canadian National Occupancy Standard) [HOSD= 06-09 and HHCD = 11-32]
- number of classifiable occupied private dwellings (excluding dwellings where housing utilisation cannot be determined or is not stated) [HOSD ne 10, &&, @@ and HHCD = 11-32]

**LONE** % OCCUPIED PRIVATE DWELLINGS THAT ARE LONE PERSON OCCUPIED PRIVATE DWELLINGS

- number of classifiable occupied private dwellings that are occupied by lone person households [HHCD = 31]
- number of classifiable occupied private dwellings [HHCD = 11-32]

*Other indicators of advantage or disadvantage*

*Cars*

**HIGHCAR** % OCCUPIED PRIVATE DWELLINGS WITH THREE OR MORE CARS

- number of classifiable occupied private dwellings which had 3 or more registered motor vehicles at or near the dwelling [VEHD = 03-30 and HHCD = 11-32]
- number of classifiable occupied private dwellings (excluding number of vehicles not stated) [VEHD ne &&, @@ and HHCD = 11-32]

**NOCAR** % OCCUPIED PRIVATE DWELLINGS WITH NO CARS

- number of classifiable occupied private dwellings which did not have a registered motor vehicle at or near the dwelling [VEHD = 00 and HHCD = 11-32]
- number of classifiable occupied private dwellings (excluding number of vehicles not stated) [VEHD ne &&, @@ and HHCD = 11-32]

*Internet*

**NONET** % OCCUPIED PRIVATE DWELLINGS WITH NO INTERNET CONNECTION

- number of classifiable occupied private dwellings with no internet connection [NEDD = 2 and HHCD = 11-32]
- number of classifiable occupied private dwellings (excluding internet connection not stated) [NEDD ne &, @ and HHCD = 11-32]

*Other*

**CHILDJOBLESS** % FAMILIES WITH CHILDREN UNDER 15 YEARS OF AGE AND JOBLESS PARENTS

- number of families with children aged under 15 and jobless parents [FMCF = 21, 31 and LFSF = 16, 17, 19, 25, 26]
- number of families (excluding not applicable and not stated) [FMCF ne @@@@ and LFSF ne 06, 11, 15, 18, 20, 21, 27, @@]

**DISABILITYU70** % PEOPLE AGED UNDER 70 WHO NEED ASSISTANCE WITH CORE ACTIVITIES

- number of people aged under 70 years needing assistance in one or more of the three core activity areas of self-care, mobility and communication, because of a disability, long term health condition (lasting six months or more) or old age [AGEP < 70 and ASSNP = 1]
- number of people aged under 70 years (excluding need for assistance not stated) [AGEP < 70 and ASSNP = 1-2]

**ENGLISHPOOR** % PEOPLE WHO DO NOT SPEAK ENGLISH WELL

- number of people aged 5 years and over who speak English either not well or not at all [AGEP > 4 and ENGLP = 4, 5]
- number of people aged 5 years and over (excluding those who did not state their English proficiency or main language) [AGEP > 4 and ENGLP = 1-5]

**ONEPARENT** % FAMILIES THAT ARE ONE PARENT FAMILIES WITH DEPENDENT OFFSPRING ONLY

- number of families that are one parent families with dependent offspring only [FMCF = 3112, 3122, 3212]
- number of families [FMCF ne @@@@]

**SEPDIVORCED** % PEOPLE AGED 15 AND OVER WHO ARE SEPARATED OR DIVORCED

- number of people aged 15 years and over who are separated or divorced [MSTP = 3, 4]
- number of people aged 15 years and over [MSTP = 1-5]

**UNINCORP** % OCCUPIED PRIVATE DWELLINGS WITH AT LEAST ONE PERSON WHO IS THE OWNER OF AN UNINCORPORATED ENTERPRISE

- number of classifiable occupied private dwellings where at least one usual resident is the owner of an unincorporated enterprise [SIEMP = 5, 6, UAICP = 1 and HHCD = 11-32]
- number of classifiable occupied private dwellings [HHCD = 11-32]

## FOR MORE INFORMATION . . .

**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.

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

FAX 1300 135 211

EMAIL [client.services@abs.gov.au](mailto:client.services@abs.gov.au)

PHONE 1300 135 070

## FREE ACCESS TO STATISTICS

All ABS statistics can be downloaded free of charge from the ABS web site.

**WEB ADDRESS** [www.abs.gov.au](http://www.abs.gov.au)