

## Task 4.6 Building and mapping a small area deprivation index for health needs assessment

The purpose of this exercise is to construct and map a health-oriented deprivation index for a study area in southern England. The exercise demonstrates the multiple stages of data retrieval and manipulation in the creation of such indices, compared to the relative ease of mapping the results. The index to be constructed uses the same variables as the Carstairs index but will not be standardized and calibrated using exactly the same reference data.

### **Data:**

Provided within this compressed file is a Shapefile **HantsOA01** containing a boundary dataset for 2001 census output areas in the Hampshire, Southampton and Portsmouth local government areas in southern England, extracted from data freely distributed on DVD by the Office for National Statistics. The approximate area of the map can be explored online at

<http://local.live.com/default.aspx?v=2&cp=51.046574~-1.198883&style=r&lvl=10&tilt=-90&dir=0&alt=-1000&scene=4318724> The license documents **Conditions of supply of 2001 Census Output Area Boundaries.doc** and **OS\_OA\_License.doc** specify the conditions of use of these datasets.

### **Exercise:**

The paper by Morgan and Baker (2006) describes the construction of a Carstairs deprivation index for areas known as wards in England and Wales by the Office for National Statistics. The analysis described in the paper uses raw, unpublished census data that are not available to researchers outside the statistical agencies. In this exercise, we shall produce a Carstairs-based index for census output areas using published census data from the Neighbourhood Statistics website. Before commencing this exercise, you should read the paper by Morgan and Baker (2006).

**Task 1:** What are Morgan and Baker's (2006) reasons for selecting a Carstairs deprivation index and the types of health-related analysis which they were intending to perform?

Note the four census variables which make up the Carstairs index.

*See last page for answers.*

Census data need to be retrieved from the Neighbourhood Statistics website <http://www.neighbourhood.statistics.gov.uk>. The site allows spreadsheet tables to be downloaded which contain census variables for a single topic, with variables as the

columns and geographical areas as the rows. From the front page, select the option to 'view or download data by topic'. This will present a series of topic categories, of which one is '2001 Census: Census Area Statistics'. Opening up this option in the topic tree will present a range of topic statistics available from the 2001 census for England and Wales. Each topic area will need to be selected and downloaded separately. Select the topics 'Economic Activity (UV28)'; 'National Statistics Socio-economic Classification (UV31)', 'Cars or Vans (UV62)' and 'Persons per Room - Household Residents (UV83)'. Each in turn should be selected to download, using the 'NeSS Geography Hierarchy South East' (as our study area falls within the South East region) and a .csv file saved to your local PC, resulting in four compressed data files each containing all the data for one topic area. It is recommended that you give each downloaded file a unique name relating to its content such as 'UV62\_cars'. Within each of these compressed files are a series of spreadsheets containing data for the South East region. Each file contains data for one level of the census geography hierarchy, of which the output areas are the lowest. The files containing output area data have names identifying the topic they contain and ending in '...GeoPolicy\_SE\_OA'. These are the files required for our further analysis.

These files contain many census variables and output areas which are not required for the present analysis and it is therefore necessary to select the rows and columns of interest and to combine them into a single index value for each output area within the study area. The 'OA\_CODE' column in each file contains the output area identification codes. We require only the records for output areas with codes beginning 24 (the County of Hampshire) and those beginning 00MS and 00MR (the Cities of Southampton and Portsmouth). These may also be identified by their names in the 'CTY\_NAME' and 'LA\_NAME' fields. The data columns within each file include a denominator variable (such as 'All households' and the variable(s) of interest such as 'No car or van'. The relevant variables for the Carstairs index for the output areas in the study area should be extracted and pasted into a new spreadsheet and percentage values for the four Carstairs variables produced. Unemployment as a percentage of the entire population should be used in place of male unemployment used in the Morgan and Baker paper, which describes the methodology for calculating the Carstairs index, and provides a table of national mean and standard deviation values for each variable (Table 2). Although these relate to their ward level data, they may be used here for the purposes of this exercise. Alternatively, you can calculate the means and standard deviations for each variable for the output areas in your own data. Standardized scores are produced for each of the four Carstairs variables by subtracting the mean (either the means for your data and just for the study area, or the national means in Table 2 of the Morgan and Baker paper) and dividing by the standard deviation (either the standard deviations for the study area, or the standard deviations in the Morgan and Baker paper); these standardized scores are then summed to produce the combined index.

**Task 2:** Summarise the calculation of the Carstairs index and produce a table of summary statistics for each Carstairs variable and for your overall index for the output areas in the study area, showing mean, maximum and minimum values.

*See last page for answers.*

Display the boundary dataset in ArcGIS. You will find that there is a field in the Shapefile called 'COA\_code' which contains census output area codes of the type '24UNGH0009'. This field matches the 'OA\_CODE' column in the census data files from Neighbourhood Statistics and will allow the census data to be matched to the boundaries.

Join your four census variables and combined score to the map and examine each of the distributions, making reference to the online topographic mapping as necessary to identify the principal urban areas.

**Task 3:** Based on the discussion in Morgan and Baker (2006), what relationships might you expect to find between your map and patterns of health in this study area?

*See last page for answers.*

### **Reference:**

Morgan, O. and Baker, A. (2006) Measuring deprivation in England and Wales using 2001 Carstairs scores *Health Statistics Quarterly* 31, 28-33

<http://www.ons.gov.uk/ons/rel/hsq/health-statistics-quarterly/no--31--autumn-2006/measuring-deprivation-in-england-and-wales-using-2001-carstairs-scores.pdf>

### **Suggested answers to tasks:**

**Task 1:** Morgan and Baker give several reasons for using the Carstairs index. One reason was that by using such a census-based measure, they could calculate deprivation index values for both England and Wales. Another reason was that one of the alternative possible indicators, the Index of Multiple Deprivation (IMD, released for various years starting with 2004 through to 2015), contains a health domain as one of its component parts. There is something circular in the logic of looking at the relationship between health and deprivation, when the deprivation measure itself is in part based on health data.

**Task 2:** Here are some summary statistics for the constituent standardized scores of the Carstairs index:

	Unemployment	Overcrowding	No car	Low Social Class	Carstairs
Mean	0.00	0.00	0.00	0.00	0.00
Minimum	-1.15	-0.76	-1.30	-1.96	-5.16
Maximum	8.45	10.26	4.38	3.83	15.66

Note that these scores have been calculated using the means and standard deviations for the data for this exercise (for Portsmouth, Southampton and Hampshire) – not the national means given in Table 2 of Morgan and Baker. If you used the national means given in Morgan and Baker, your results will be different and you will not have a set of means that are all zero. If you used national means and standard deviations, your summary table should hopefully look something like this:

Variable	Mean	Max. values	Min. values
Unemployment	-0.91961	3.16898112	-1.5119363
Overcrowding	-1.05468	4.14721555	-1.4385965
Car Ownership	0.334121	6.13721805	-1.3815789
Low Social Class	-0.24421	3.27337342	-2.0417288
Carstairs Index	-1.88438	9.154893089	-6.37384

Table 1 - Results of calculation of Carstairs Index

Task 3: The authors suggest in their conclusions that ‘Carstairs deprivation index has been shown to perform well in explaining variations in health measures’ – in other words, one would expect to find a significant positive relationship between the Carstairs index and measures such as infant mortality. They suggest that this relationship may be stronger than for other variables, such as social class. Even though not all households living in a deprived area will be deprived, nonetheless, there is often an ‘area effect’ whereby even better-off households in deprived areas are at greater risk of ill health.