



27th Australasian Transport Research Forum, Adelaide, 29 September – 1 October 2004

Paper title: A GIS approach to transit buffer problem analysis

Author(s) name(s): Sekhar Somenahalli, Wen Long Yue and Michael A P Taylor

Organisation(s): Transport Systems Centre, University of South Australia

Contact details:

Postal address: GPO Box 2471, Adelaide SA 5001

Telephone: +61 8 8302 1855

Facsimile: +61 8 8302 1880

email: Sekhar.somenahalli@unisa.edu.au

Abstract (200 words):

The issue of producing socio-economic profiles for transit service area buffers when the census collection districts (CDs) are not wholly contained has been investigated with regard to several contributing variables in Adelaide, South Australia. The analysis was carried out by grouping the CDs, according to the existing level of transit access of dwelling units, distance from the city centre, and direction (whether the CD is located in northern part or southern part of city etc.) of travel. This research focused on understanding the influence of these factors on the transit dependent socio-economic profile of the dwelling units. The results suggest that, within a certain distance (in the northern statistical sub division it is 17 km) there is no difference between the transit dependent social economic profile of CDs with full transit access (wholly contained) and those with partial transit access (partially contained). This research has outlined the process for calculating the threshold distance in Northern Adelaide to help other researchers prepare socio-economic profiles of transit service areas, which contain full and partial access CDs.

Introduction

The concept of area coverage expresses how many people live within a reasonable walking distance of a transit route. The area within a 5-minute walk from a bus stop is traditionally considered the primary service area. The standard definition (Bus route and schedule planning guidelines, NCHRP report #69, 1980) of service area is the region within 0.25 mile (0.4 kilometres) of a route or stop in a transit service. Using GIS software it is easy to build these service areas or transit buffers around the routes. Sekhar, Coffee and Yue (2001) described the methods by which population can be accurately estimated within transit route buffers. The next step is to estimate the socio-economic characteristics of those people who live within this buffered area.

Planning new transit routes and refining existing routes often requires the calculation of the transit dependant socio-economic characteristics of potential passengers within transit service areas. The most common method used to derive the social profile for potential passengers is to create buffers around the transit routes and use Census data collected by the statistics bureau, which in this case is the Australian Bureau of Statistics Census of Population and Housing 1996. The census data are based upon the Census Collection Districts (CDs) which are the smallest aerial units provided by the Australian Bureau of Statistics (ABS) population census. The CD is primarily designed by the ABS as the workload for one collector, but does take account of property boundaries and natural features, and is usually 300-400 households. However the CD boundary seldom matches the transit buffer boundary and this requires the development of methods for allocating partial CD populations to the transit buffers (Sekhar *et al* 2001). This research builds upon the earlier work of Sekhar *et al* 2001 and focuses upon the issues of applying statistical tests (both parametric and non parametric) to understand and solve the allocation of populations and socio-economic characteristics when CDs are only partially contained within the transit buffer area. The analysis is carried out by grouping the CDs according to the existing level of transit access of dwelling units, distance from the city centre and the direction (north, south east or west of the city) to understand the influence of these factors on the transit dependent socio-economic profile of dwelling units.

To assess how the populations are affected by changing a transport route (for example, does cutting a particular bus route affect certain groups more than others), Werner, R (1988) used non-parametric tests. This research focuses on such statistical tests (both parametric and non parametric) to understand and solve the wholly/partial buffer problems.

Study area and data base development

The study area (Figure 1) comprises the Adelaide Statistical Division (ASD). Adelaide is the capital city of the state of South Australia (SA) in Australia. The database for this study was developed from the following resources.

- Census data from Australian Bureau of Statistics, Census of Population and Housing, 1996.
- Transit route information from the SA Passenger Transport Board
- Cadastral database for the city from the Department of Environment and Heritage.
- Cadastral attribute data (valuation) from the Department of Administrative and Information Services.

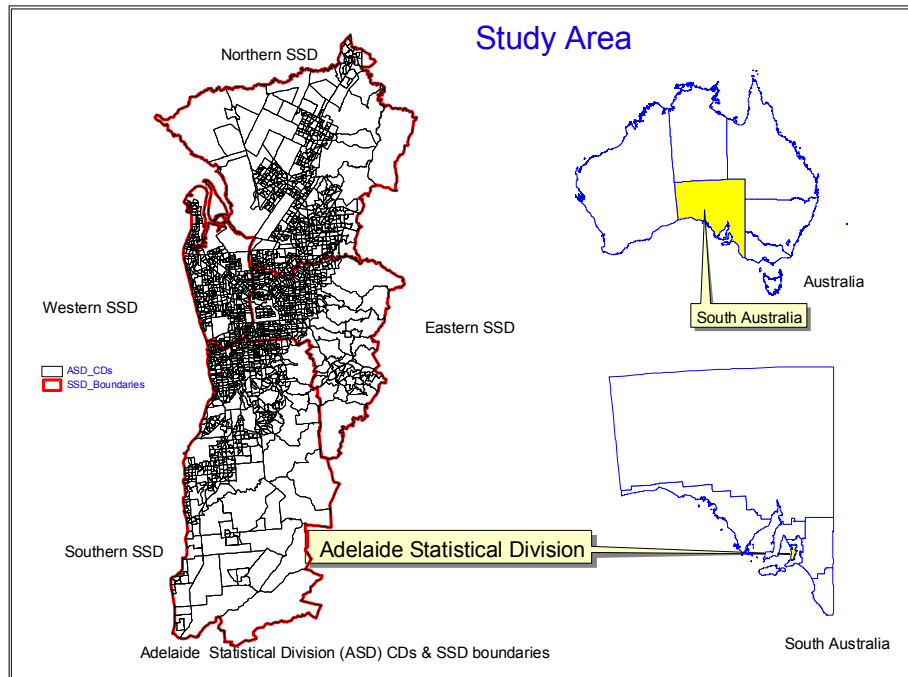


Figure 1 The study area

Statement of the problem and key definitions

It is often required to calculate the transit dependent socio-economic characteristics of the potential transit clientele along the transit routes, which will help in planning new routes or refining existing routes. The most common method is to create transit buffers around the transit routes and derive the socio-economic profile of the people who are inside the buffer polygons. Transit dependent characteristics are then derived from the 1996 census for the Census Collection Districts (CD). However, the CD boundaries seldom match the transit buffer boundary. Hence, we require a method for allocating socio-economic characteristics for CDs, which are partially contained within the transit service area. This problem of allocation issue is illustrated in Figure 2, which also shows the dwelling unit information obtained from DCDB-valuation data and the transit buffer.

Consider the task of estimating the transit dependent characteristic such as percentage of low-income households in the two CDs i.e. CD number 1 and CD number 2 (as displayed in Figure 2). From the ABS Census, CD number 1 has 25 percent of households with low-incomes. Since all the households are inside the transit buffer and the number of low-income households is known, low-income households can be estimated by multiplying this percentage (taken from census) with the number of households (from DCDB-valuation).

Similarly, consider the task of estimating the transit dependent characteristics of the households belonging to CD number 2 that are within the transit buffer. As per the ABS census, this CD has 20 per cent of low-income households. The number of dwelling units inside the buffer can be estimated using GIS software. The real problem is how to calculate the number of low-income households belonging to this CD, which are inside the buffer, since not all-dwelling units are inside the transit buffer. Is it realistic to apply the entire CD

percentage to those dwellings, which are inside the buffer? This problem is addressed in this research.

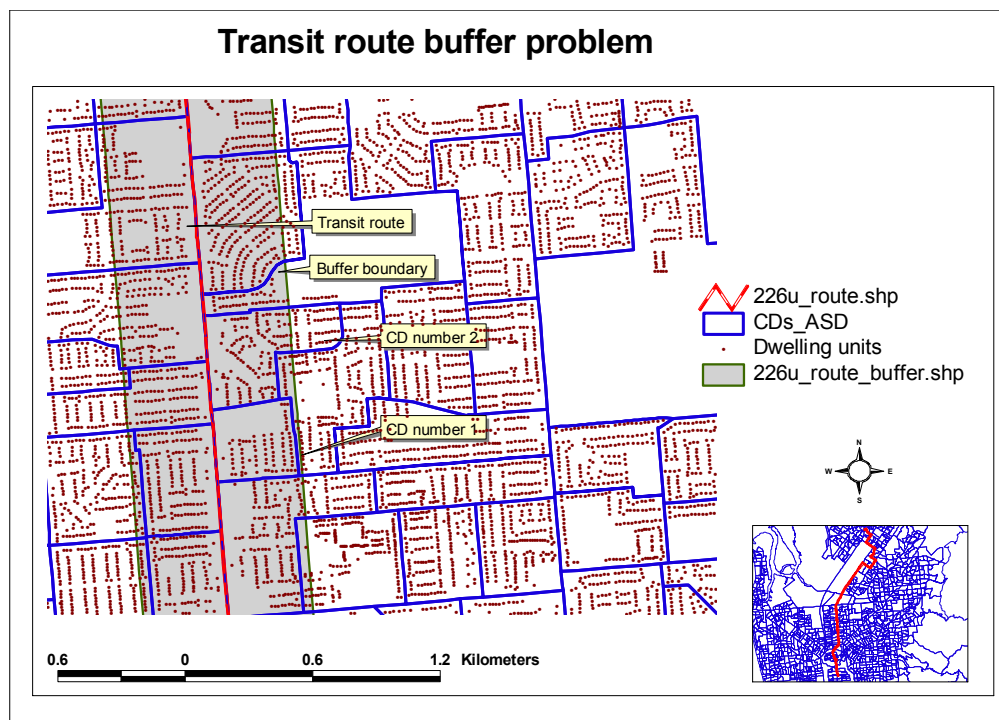


Figure 2 Transit buffer along a route, Census Districts and Dwelling units

Key definitions

Census Collection Districts (CDs): The Census Collection Districts (CDs) are the smallest aerial units provided by the Australian Bureau of Statistics (ABS) population census. The CD is primarily designed by the ABS as the workload for one collector, but does take account of property boundaries and natural features, and is usually 300-400 households. The numbers of households differ between CDs but in urban areas the CDs are far more uniform in size (area) and households than fringe or rural CDs, which are often large in area in an attempt to capture 300-400 households. The Adelaide Statistical Division (ASD) contains 2141 such CDs.

Wholly (full access) and partially contained (partial access) CDs: Wholly contained (full access) CDs have all dwelling units within the transit service buffer (400 metres euclidean distance from any transit route). Partially contained (partial access) CDs are intersected by the 400-metre transit buffer and contain a proportion of the CD dwellings.

Low Income Households (HHs): Households with a weekly income of less than \$300 (Australian dollars) per week. This definition is based on the ABS classification.

Households with 'no' vehicle: Households, which do not own any motor vehicle.

Gaussian distribution: When many independent random factors act in an additive manner to create variability, data will follow a bell-shaped distribution (Figure 3) called the Gaussian distribution.

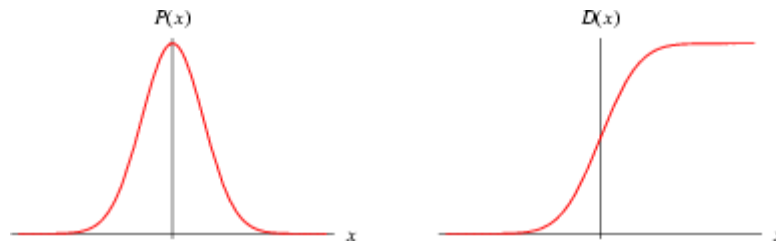


Figure 3 - Shape of distribution (Wolfram research, 2002)

The Gaussian probability distribution, with mean μ , and standard deviation σ is a normalized Gaussian function. This distribution is also called the Normal distribution, which is of the form:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{Equation 1}$$

Where $P(x) dx$ gives the probability that a variate with a Gaussian distribution takes on a value in the range $[x, x+dx]$. The cumulative distribution $D(x)$, which gives the probability variate will assume a value of $\leq x$ is then the integral of the Gaussian function (Wolfram Research, 2002). This Gaussian distribution has some special mathematical properties that form the basis of many statistical tests.

Key Objective

To estimate the population and derive transit dependant socio-economic characteristics within the transit buffer area from the smallest available spatial unit (i.e. CDs) characteristics.

Hypothesis

The following hypothesis can be set up: the transit service area buffer socio-economic dependent characteristics of the CDs, which are wholly contained within the buffer (i.e. CDs with every dwelling unit having transit access) do not differ from the socio-economic characteristics of CDs, which are partially contained within the transit buffer (i.e. CDs with partial transit access).

Methodology

The aim of this research is to determine whether there is a significant difference between the two CD groups under consideration (wholly or partially contained CDs). The objective is not to prove that the two groups are indistinguishable, but to identify that at a particular level of probability, the two groups are closer to each other. There are two statistical tests applied in the analysis, which are useful in comparing two groups of data.

- t- test (parametric test); and
- Mann-Whitney test (non-parametric test).

6 A GIS approach to transit buffer problem analysis

The first step is to obtain the data regarding full access and partial access CDs. Then, create a 400 metre buffer along all transit routes and overlay the DCDB-valuation dwelling theme. Subsequently, select CDs where all dwelling units are inside the transit buffer (i.e. full transit access CDs) and also select the CDs in which only portion of dwelling units are located inside the transit buffer (partial access CDs).

The next step is to check whether these two datasets meet the parametric test assumptions. If they meet the required assumptions for a parametric test, then, t-test is applied, otherwise the Mann-Whitney test (non parametric test) is applied. The steps in this procedure are summarised in Figure 4.

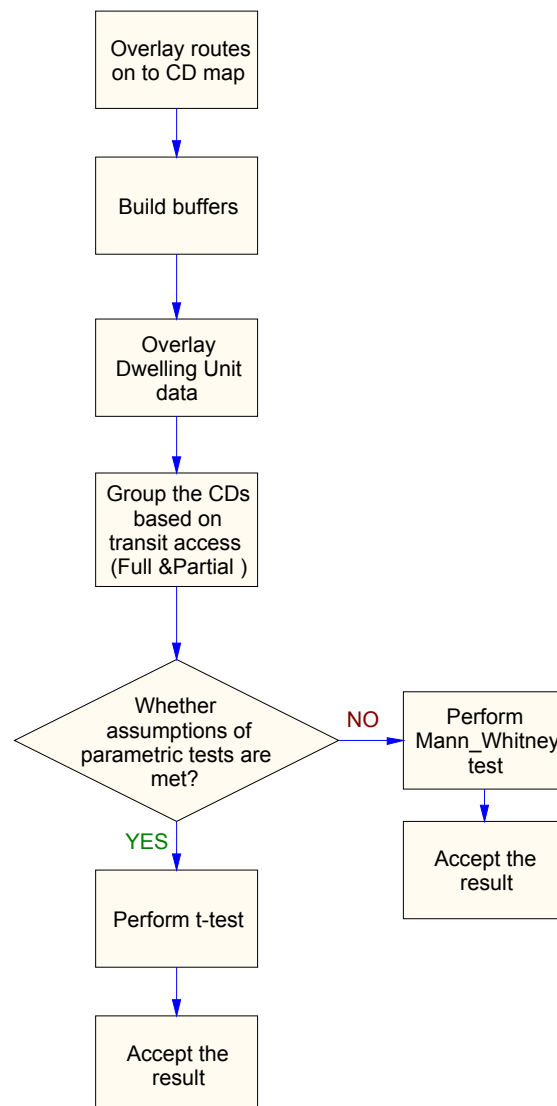


Figure 4 Approach Steps

Next task is to group the CDs based upon direction and distance for the following two transit dependent characteristics:

- The percentage of low-income households (HHs); and
- Households (HHs) with no vehicle.

Check for parametric test assumptions

The required conditions for parametric tests are that, the data distribution should be Normal (Gaussian) and that there should not be significant difference between the standard deviation of the two data groups. The first condition is tested using the Kolmogorov-Smirnov (K-S) test, while 'F' test is used to test the second condition.

Kolmogorov – Smirnov (K-S) test: This is also called the normality test as it tests for deviations from Gaussian (Normal) distributions. The Kolmogorov-Smirnov test, named after the statisticians Kolmogorov and Smirnov who developed it, is a simple method for testing whether there is a significant difference between an observed frequency distribution and theoretical frequency distribution (Richard and Rubin, 1991). The K-S statistic quantifies the discrepancy between the distribution of the data and an ideal Gaussian distribution; a larger value denotes a larger discrepancy. It is not informative by itself, but is used to compute a 'p' value (GraphPad Software Inc, 1998). The p value from the normality test answers this question: If a sample is randomly drawn from a Gaussian population, what is the probability of obtaining a sample that deviates as much from Gaussian distribution as this sample does? More precisely, the p value answers this question: if the population was really Gaussian, what is the chance that a randomly selected sample of this size would have a KS distance as large, or larger, as observed (GraphPad Software Inc, 1998)?

F test to compare variances: The 'F' test is used to test the second condition (that the two populations will have the same standard deviation and thus same variance). If the two populations have the same variance, what is the chance that the selected samples whose ratio of variance is as far from 1.0 (or further) as observed in the experiment? A sufficiently small p value suggests that variance is different. However, critical p value depends on the sample sizes of the two datasets under comparison.

Parametric and Non parametric tests: The parametric test used in this study is a 't-test' and the non-parametric test used in this study is the Mann-Whitney test (Zuwaylif, 1980).

t- test: The t-test is a parametric test used to test the significance of the difference between the means of two samples, which come from normal distributions with a common variance. In this test the standard error of the difference between the two means is estimated using Equation 2 where μ_1 and μ_2 are the means, s_1 and s_2 are the standard deviations, and n_1 and n_2 are the sample sizes.

$$\sigma_{(\mu_1-\mu_2)} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)} \quad \text{Equation 2}$$

Once the t-value is computed it is compared with the table of significance to test whether the ratio is large enough to determine the difference between the groups is a chance finding. To test for significance, the alpha level is normally set at a risk level of 0.05. This means that five times out of a hundred there is a possibility to find statistically significant difference between the means even if no difference (i.e., by 'chance') exists. In the t-test, the degrees of freedom (df) are the sum of the persons in both groups minus two. When the alpha level is set and the degrees of freedom calculated, the t-value is compared with the values in a standard table to determine whether the t-value is large enough to be significant. If the t value is significant, then the difference between the means for the two groups is different.

Mann-Whitney test: Certain assumptions regarding the sample populations are required when using parametric tests. For example, the use of t-distribution to test the significance of the difference between the means of two small samples requires that the two samples are independent samples selected from normally distributed populations with equal variances. Similarly, the use of F-distributions to compare the means of several populations requires that the samples be drawn from normal distributions with equal variances. When these stringent assumptions about the distributions of parent population can't be met, parametric tests are no longer reliable; and some alternative tests, called 'distribution free' tests, are utilised. However, non-parametric tests do not utilise all the information provided by the sample and in situations where both parametric and non-parametric tests are applicable, non-parametric tests have less statistical power and are thus at a greater risk of selecting inappropriate hypotheses. Hence, the Mann-Whitney test is used when the assumptions required for parametric tests are not met.

In situations where the assumptions regarding the distribution of the parent population are in doubt, it is better to use a non-parametric test such as the Mann-Whitney test. This is based on the idea that if the two independent samples are drawn from the same population, then the average ranks of the two samples should almost be equal. The Mann-Whitney test is useful for comparing two populations on the basis of independent random samples. By assuming that the two population distributions have the same shape, the test is used to determine whether the two populations have identical location or shape. This test is applied to the data in this study, as it is appropriate to test the null hypothesis that there is no difference in the socio-economic profiles of the wholly and partially contained CD groups. It is argued that this test is useful for such problems because the test does not assume normally distributed observations and can be applied to skewed variables. The Mann-Whitney test consists of the following steps (Zuwaylif, 1980):

Step 1: Combine the scores of both groups, then assign rank 1 to the lowest score in the combined data, rank 2 to the second highest, rank 3 to the third highest, and so on until all values are ranked.

Step 2: Sum the ranks for each group. The rank sum is denoted by R. The sum of the first group ranks is R_1 , and the sum of the second group ranks is R_2 .

Step 3: Determine the value of Mann-Whitney U statistic using the following equation, where n_1 and n_2 are the number of samples of the two groups.

$$U = n_1 \times n_2 + n_1 \left(\frac{n_1 + 1}{2} \right) - R_1 \quad \text{Equation 3}$$

or

$$U = n_1 \times n_2 + n_2 \left(\frac{n_2 + 1}{2} \right) - R_2 \quad \text{Equation 4}$$

Step 4: Determine the mean and standard deviation of the statistic U. The mean of statistic U, denoted by $E(U)$, which is calculated by the following equations:

$$E(U) = \frac{n_1 + n_2}{2} \quad \text{Equation 5}$$

The standard deviation of U, denoted by σ_U , is

$$\sigma_U = \sqrt{\left(n_1 \times n_2 \left(\frac{n_1 + n_2 + 1}{12} \right) \right)} \quad \text{Equation 6}$$

Step 5: Calculate the ‘z’ score value using the following equation:

$$z = \frac{(U - E(U))}{\sigma_U} \quad \text{Equation 7}$$

Extract the probability (p) value from the table of z scores (Zuwaylif, 1980). If the result of this computation shows a ‘z’ score with a high probability (i.e. $p > 0.05$), it implies that the medians of the two groups are not significantly different.

Analysis

Stage 1: Grouping of CDs based on transit access: At this stage, all the CDs in Adelaide Statistical Division (ASD) are grouped based on their access to transit. The first group consists of CDs in ASD which will have full transit access (wholly contained-in which case, every dwelling unit has access to transit route within 400 metres walking distance) and the second group consists of CDs that have partial transit access (partially contained-in which case, not all dwelling units have access to transit service within 400 metres walking distance). The steps involved in this operation are listed below:

Step 1: Build the Census District map for ASD from the census data of the ABS using any standard software. The current analysis uses ESRI’s GIS software ArcView 3.2.

Step 2: Merge all the transit routes and build a single theme containing the routes for all forms of transit.

Step 3: Build the euclidean buffer (route buffer) based upon 400 metres. It is common practice in transit planning to use 400 metres (0.25 mile) as the upper limit of walking distance to a transit stop (Hsiao, Lu, Sterling and Weatherford (1997)).

Step 4: Overlay the dwelling unit information obtained from the DCDB-valuation data.

Step 5: Calculate the number of dwelling units in each CD and group into the wholly and partially contained CDs (full and partial access to the transit service) (Figure 5). The buffer acts like a cookie-cutter and the CDs like dough. After the cookie-cutter is pushed into the dough, one can aggregate data inside the cut out piece and compare it with the rest of the uncut dough. Derive transit dependent socio-economic characteristics of each CD in these groups from the census data.

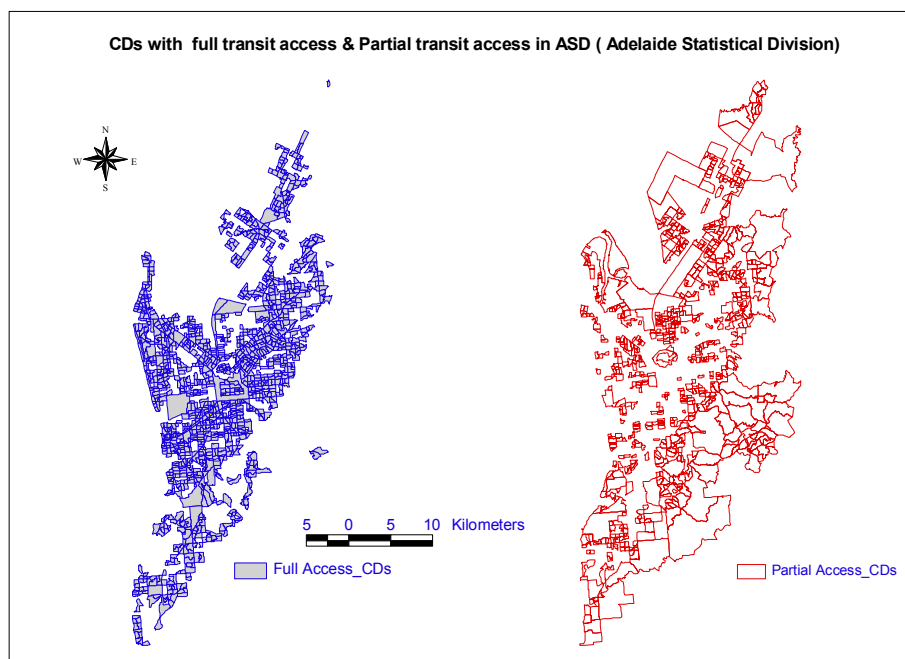


Figure 5 Full transit access and partial access CDs in the ASD

Step 6: Apply t-test to the two groups for any of the transit dependent characteristics only if the distribution is normal (which can be tested by Kolmogorov-Smirnov test) and there is no significant difference in the standard deviations between the two groups, otherwise perform the Mann-Whitney test, as this test is more appropriate if the conditions for t-test are not met.

Findings of the analysis of stage one: Table 1 and Table 2 show the results of t-test and Mann-Whitney test respectively. From these tables, it is clear that the means or medians of these two groups are statistically different.

Table 1 Check for parametric test assumptions for stage one of the analysis – All CDs in ASD

Description	The Groups	Number of CDs	Mean	Std. Deviation (SD)	Kolmogorov and Smirnov Test			Is the difference b/n the SDs Significant?
					KS	Passed Normality test?	Distribution	
% of Low_Inc HHs	Full transit access CDs	1373	25.48	10.66	0.0273	Yes	<i>Not* Gaussian</i>	F= 1.091 p= 0.107 <i>Not Significant</i>
	Partial transit access CDs	661	19.24	10.13	0.0772	No		
HHs with No vehicle	Full transit access CDs	1373	16.52	10.56	0.0024	No	<i>Not* Gaussian</i>	F= 1.413 p<0.0001 Extremely Significant
	Partial transit access CDs	661	10.36	8.88	0.1204	No		

*Note: Since the data failed the normality test and thus failed to satisfy the basic assumption of parametric tests, the Man-Whitney test was applied and the results are shown in Table 2 below.

Table 2 Mann-Whitney test results for stage one of the analysis – All CDs in ASD

Description	The Groups	Number of CDs	Median	Sum of the ranks	Mann-Whitney U statistic		Two Tailed p Value
					U	U ¹	
% of Low_Inc HHs	Full transit access CDs	1373	25.15	942737	218764	442237	P < 0.0001 Extremely Significant
	Partial transit access CDs	661	17.17	437555			
HHs with No vehicle	Full transit access CDs	1373	15.60	955578	205923	455078	P < 0.0001 Extremely Significant
	Partial transit access CDs	661	7.95	424714			

This may be due to the fact that they are spread over a wide area in Adelaide. Based upon the size of Adelaide and the potential for this to influence the test outcomes, the CDs are further grouped based on the distance from the city centre.

Stage 2: Grouping the CDs based on transit access and distance from the city centre: All dwelling units or households in CDs that are less than two km from the city centre have access to a transit route. The majority of CDs within five km of the city centre have full transit access and hence the allocation and determination of socio-economic characteristics for CDs and transit buffers does not arise for these CDs. This group is not analysed further and the rest of the research is concerned with the CDs which are more than five km from the city centre. CDs are grouped at two km bands; 5-7 km, 7-9 km and 9-11 km, and the data are segregated based on transit access (full and partial).

The grouping of CDs in the two km bands may look arbitrary. However, this distance band of two km is chosen based on the average diameter of the CD, which is close to two km. The logic behind this is that the statistical tests valid for the group of CDs within the band of two km will hold good for all size of CDs in the study area.

Step 1: Build the CD map for the ASD from the ABS census data as explained in the previous stage.

Step 2: Identify the central CD, which in this case is the CD containing the GPO (General Post Office) and convert this shape to centroid and store it in one theme. Similarly, convert all shapes of the CDs (in ASD) to centroids and store them in another theme.

Step 3: Use an Avenue script (spider.ave in this case) to create a spider diagram based on the linear distances between the points contained in the two point themes. The output is a new theme (spider) with distances stored in the feature table. This table contains the distance from the centroid of each CD to the centroid of the central CD.

Step 4: Select all CDs with a distance to the city centre less than 5 km and discard from the analysis. Select the CDs in two km bands and group them for further analysis.

12 A GIS approach to transit buffer problem analysis

Step 5: Merge all the transit routes and build a single theme containing the routes of all forms of transit.

Step 6: Build the euclidean buffer (route buffer), parallel to the transit route.

Step 7: Overlay the dwelling unit information obtained from DCDB-valuation data.

Step 8: Group all CDs where all dwelling units have transit access. Formulate another group consisting of CDs in which only portion of dwelling units have transit access (Figure 6). Derive transit socio-economic dependent characteristics of each CD in these groups from the census data.

Step 9: Apply the statistical tests to these two groups for the transit dependent characteristics.

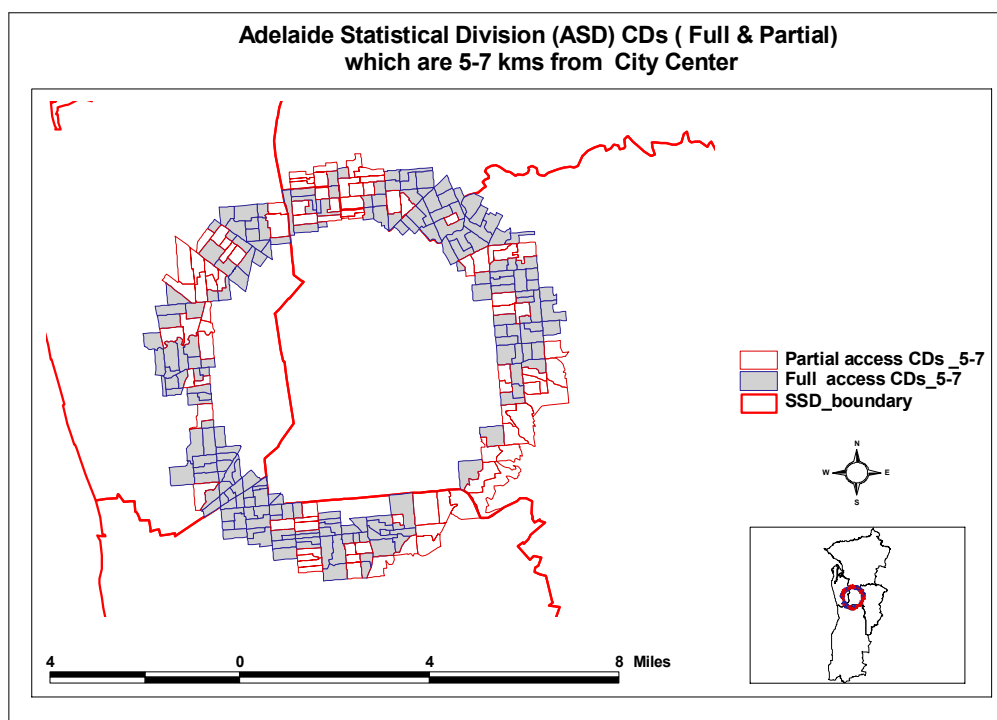


Figure 6 Full and partial access CDs in ASD, which are 5-7 km from city centre

Findings of analysis of stage two: Results for the 5-7 km group of CDs, are shown in Table 3 and Table 4. The medians and means of the two groups differ significantly.

Table 3 Check for parametric test assumptions for the two groups of CDs in ASD, which are 5-7 km from CBD

Description	The Groups	Number of CDs	Mean	Std. Deviation (SD)	Kolmogorov and Smirnov Test			Is the difference b/n the SDs Significant?
					KS	Passed Normality test?	Distribution	
% of Low_Inc HHs	Full transit access CDs	196	26.47	8.24	0.059	Yes	Gaussian	F= 1.739 P= 0.008 Significant*
	Partial transit access CDs	88	22.21	10.87	0.085	Yes		
HHs with no vehicle	Full transit access CDs	196	16.77	8.07	0.098	No	<u>Not Gaussian</u>	F= 1.089 P= 0.311 Not Significant
	Partial transit access CDs	88	12.87	8.42	0.084	Yes		

*Note: Since the data failed the normality test and thus failed to satisfy the basic assumption of parametric tests, the Man-Whitney test was applied and the results are shown in Table 4 below.

Table 4 Mann-Whitney test results for the two groups of CDs in ASD, which are 5-7 km from CBD

Description	The Groups	Number of CDs	Median	Sum of the ranks	Mann-Whitney U statistic		Two Tailed p Value
					U	U ¹	
% of Low_Inc HHs	Full transit access CDs	196	26.17	30372	6182.5	11066	<0.0001 Extremely Significant
	Partial transit access CDs	88	21.11	10099			
HHs with no vehicle	Full transit access CDs	196	15.97	30441	6113.5	11135	< 0.0001 Extremely Significant
	Partial transit access CDs	88	13.16	10030			

This analysis indicates that the distance of the CD from the city centre is not the only factor that constitutes in the formation of homogeneous CDs. To further refine the analysis, CDs are regrouped based on the distance of CD from the city centre and direction.

Stage 3: Grouping CDs based on transit access, CDs distance to city centre, and direction: For the next section of this analysis, the CDs are grouped based on transit access, distance

from the city centre, and direction (whether they are located in the northern or southern side of the Adelaide Statistical Division).

For this research the CDs in northern ASD are selected, and all the steps detailed in stage 2 are repeated. The distance and direction of CD are both controlled to reflect the reality of working with individual route transit buffers (Figure 7). The data are grouped into bandwidths based upon two km, 5-7 km, 7-9 km, 9-11 km, 11-13 km, 13-15 km, 15-17 km, and 17-19 km and the analysis is done for each grouping. Figure 7 displays Northern CDs in the 5-7 km band from city centre.

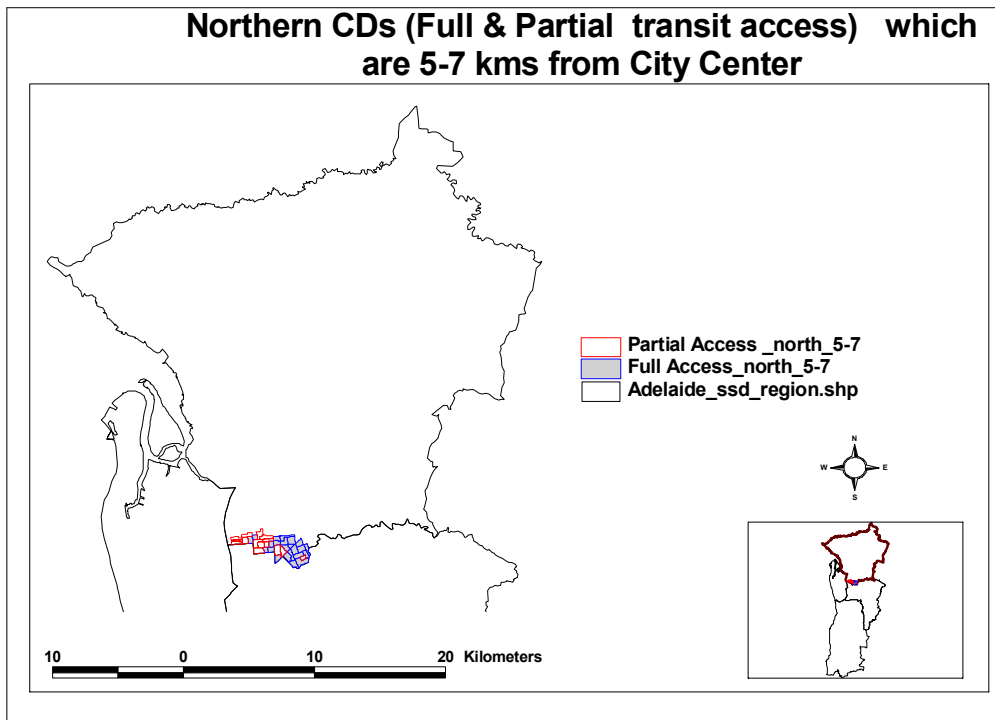


Figure 7 Full and partial access CDs in North, which are 5-7 km from City centre

Comments on the results of analysis of stage three: The results for the northern CDs in the distance range of 5-7 km (Table 5) show that there is no difference between the transit dependent social economic profile of CDs with full transit access, and with partial transit access. Hence, the buffer problem can easily be addressed for this group. A similar outcome is observed for the CD groups in the distance ranges of 7-9 km, 9-11 km, 11-13 km, 13-15 km, and 15-17 km. However, in the case of CDs which are beyond 17 km (Table 6 and Table 7) this trend did not continue. The analysis for that group showed a significant difference (mean or median) between the data in the full and partial access sets.

Table 5 Unpaired t test results for the two groups of Northern CDs, which are 5-7 km from CBD

Description	The Groups	Mean	Number of CDs	Std. Deviation (SD)	Kolmogorov and Smirnov Test			Is the difference b/n the SDs Significant?	95% Confidence interval	Two Tailed P Value
					KS	Passed Normality test?	Distribution			
% of Low_Inc HHs	Full transit access CDs	28.56	23	7.99	0.075	Yes	Gaussian	F= 1.404 P= 0.231 Not Significant	Mean difference 4.026 Confidence Interval: -1.76 to 9.84	t= 1.410 (df=36) & P=0.167 Not Significant
	Partial transit access CDs	32.58	15	9.47	0.135	Yes				
HHs with No vehicle	Full transit access CDs	18.27	23	5.48	0.115	Yes	Gaussian	F= 2.021 P= 0.067 Not Significant	Mean difference 1.198 Confidence Interval: -3.16 to 5.56	t= 0.556 (df=36) & P=0.581 Not Significant
	Partial transit access CDs	19.47	15	7.79	0.158	Yes				

Table 6 Check for parametric test assumptions for the two groups of Northern CDs, which are 17-19 km from CBD

Description	The Groups	Number of CDs	Mean	Std. Deviation (SD)	Kolmogorov and Smirnov Test			Is the difference b/n the SDs Significant?
					KS	Passed Normality test?	Distribution	
% of Low_Inc HHs	Full transit access CDs	38	18.53	9.89	0.1346	Yes	Gaussian	F= 2.15 P= 0.018 Significant *
	Partial transit access CDs	30	12.61	6.74	0.1205	Yes		
HHs with no vehicle	Full transit access CDs	38	9.97	8.83	0.2296	No	Not Gaussian	F= 8.27 P < 0.001 Extremely Significant
	Partial transit access CDs	30	5.48	3.07	0.0941	Yes		

*Note: Since the data above failed either normality test or F test for equal variance thus failed to satisfy the basic assumption of parametric tests, Man-Whitney test was applied and results are shown in the next Table.

Table 7 Mann-Whitney test results for the two groups of Northern CDs that are 17- 19 km from CBD

Description	The Groups	Number of CDs	Median	Sum of the ranks	Mann-Whitney U statistic		Two Tailed P Value
					U	U ¹	
% of Low_Inc HHs	Full transit access CDs	38	16.69	1512	369	771	0.0133 <i>Significant</i>
	Partial transit access CDs	30	12.64	834			
HHs with no vehicle	Full transit access CDs	38	7.43	1477.5	403.5	736.5	P = 0.0404 <i>Significant</i>
	Partial transit access CDs	30	5.19	868.5			

Summary and Conclusions

The issue of producing socio-economic profiles for transit service area buffers when the CDs are not wholly contained has been investigated with regard to several contributing variables in Adelaide, South Australia. The analysis was carried out by grouping the CDs, according to the existing level of transit access of dwelling units, distance from the city centre, and direction (whether the CD is located in northern part or southern part of city etc.) of travel. This research focused on understanding the influence of these factors on the transit dependent socio-economic profile of the dwelling units. The results suggest that, within a certain distance (in the northern statistical sub division it is 17 km) there is no difference between the transit dependent social economic profile of CDs with full transit access (wholly contained) and those with partial transit access (partially contained). As 90 per cent of the CDs in the Northern Statistical Subdivision are within 17 km, the socio-economic profiles for transit service area buffers within this distance can be estimated with a confidence that the characteristics do not differ significantly between full and partial access CDs. One of the major outcomes of this research is that the importance of determining the distance within which the homogeneity of the transit dependent characteristics in the full transit and partial transit access CDs is maintained. This research has outlined the process for calculating the threshold distance in Northern Adelaide to help other researchers prepare socio-economic profiles of transit service areas, which contain full and partial access CDs. Since direction was such a significant variable, the threshold distance should be calculated for each direction.

Acknowledgments

The authors would like to acknowledge the valued contributions from Chris Dyson of Transport System Centre (TSC), University of South Australia. They would also like to thank the staff members of Passenger Transport Board of South Australia for their cooperation in connection with the data collection.

References

- Bus route and schedule planning guidelines* (1980) National Cooperative Highway Research Program 69, Transportation Research Board Washington D C: pp 1-97.
- Gilbert, N (1981) *Statistics*, 2nd edn. Saunders college publishing, Philadelphia.
- Hsiao, S Lu, J Sterling, J and Weatherford, M. (1997) *Use of geographic information system analysis of transit pedestrian access*, Transportation Research Record 1604, pp 50-59.
- Richard, I L and Rubin, D S (1991) *Statistics for Management*, Prentice - Hall International (UK) Editions 5th edn.
- Sekhar, S V C Coffee, N and Yue, W L (2001) Transit Route Service area analysis using GIS, *Journal of Eastern Asia Society for Transportation Studies*, 4(1), pp 171-185.
- The InStat Guide to Choosing and Interpreting Statistical Tests* (1998) A manual for GraphPad InStat Version 3, GraphPad Software Inc.
- Werner, R J (1988) Equity in Public Transit: A Test Statistic for Buffer Problems, in *Proceedings of the Eighteenth Annual ESRI User Conference*, San Diego: pp. 1-6.
- Wolfram research (2002) [online Accessed on April, 2001],
URL: <http://mathworld.wolfram.com/GaussianDistribution.html>
- Zuwaylif, F H (1980) *General Applied Statistics*, 3rd edn Addison Wesley, California.