J. Univ. Kuwait (Sci.) **21**, 1994

Statistical models for work trip rates of Kuwaiti households

GALAL M. SAID*, DAVID H. YOUNG** AND HASSAN K. IBRAHIM*

* *Department of Civil Engineering, Cairo University, Fayoum Branch, Fayoum, Egypt*

ABSTRACT

Most trip generation analysis procedures utilized in transport planning studies are based on either zonal regression or cross-classification analysis. There has been continuous criticisms of the quality of these techniques for a variety of reasons. The use of the Generalized Linear Model (GLM) framework in conducting household-level trip generation analysis has been the subject of some research, but has not been fully explored, and only a few applications have been reported. There are several advantages that the GLM approach can provide, including the variety of the statistical models that can be investigated and tested within the general framework, the consistency in the selection of the explanatory variables, and in the way low-frequency observations are handled.

This paper outlines briefly the generalized linear model framework and describes the various statistical model options it provides for the household-level trip generation analysis, which include regression, ANOVA and analysis of covariance models with various link functions and assumptions about the underlying distribution of trip observations in the classification cells. Data on work trips and household characteristics from Kuwait generated from an extensive home interview survey is utilized to demonstrate the practical nature of the proposed framework and the possibilities it offers. The application uses data of Kuwaiti households in three housing types. Work trips are found to be influenced by car ownership and the number of adults and children in the household. Some of the interaction effects are found to be significant.

INTRODUCTION

Conventional trip generation techniques such as the zonal regression analysis and cross-classification analysis have been widely reported in the literature [FHWA 1976; Hutchinson 1974; Kassoff & Deutschman 1969]. Most transport planning studies conducted in the sixties and seventies have used one of the two techniques. There have been some shifts towards more use of cross-classification analysis than the zonal regression approach because of the several advantages it offers (Wootton & Pick 1967; Stopher & McDonald 1983). Button (1976) and Dobson & McGarvey (1977)

Address for correspondence:

* Faculty of Engineering, Cairo University, Fayoum Branch, Fayoum, Egypt.

** Brunel—The University of West London, Uxbridge, Middlesex UBB 2PH, U.K.

reported on the differences and similarities of the two techniques.

Recently, some work has been reported on the difficulties in applying the cross-classification analysis approach in a routine manner (Stopher & McDonald 1983; Rickard 1989). Some suggestions overcoming these difficulties were put forward such as the use of the multiple-classification analysis method (Stopher & McDonald 1983), and the general linear model analysis of variance (GLANOVA) (Dobson 1976).

Said & Young (1990) used part of the data from home-interview surveys conducted in Kuwait related to only Kuwaiti households living in villas to show, with the aid of some illustrative runs of the computer package GLIM (Baker & Nelder 1978), the possibilities the regression specifications of the General Linear Model (GLM) can offer [Baker & Nelder (1978)]. The data used in this analysis are the ungrouped data and all possible combinations of two quantitative explanatory variables were used.

While Said & Young (1990) described the conceptual aspects of using the regression specification of GLM, the example used in the demonstration was of an illustrative nature and, as such, did not utilize fully the data available from the home interview surveys. The complete data provide for further more comprehensive analysis, both in detail and in the type of technique used.

This paper follows on the work of Said & Young (1990). The additional developments in this paper address:

1. The use of additional quantitative explanatory variables and the treatment of a mix of quantitative and qualitative variables in the regression specification of GLM.
2. The use of regression analysis of cross-classified data with quantitative and qualitative variables such as type of house.
3. The use of ANOVA in estimating household-level mean trip rates.
4. The assessment of the strength of each of the above approaches and the advantages they offer over current techniques.

In this study, different model forms are investigated based on the variables to be used in the analysis, the type of effects to be included (i.e. single, second order or higher order interaction effects) and the underlying distribution of the number of work trips by household within the classification cells.

The significance of developments in points 2 and 3 above arises because of the practical difficulties in the use of regression analysis of ungrouped data in forecasting, since it is not possible to forecast the spatial distribution of households at this level of detail.

1. DATA FOR KUWAITI HOUSEHOLDS

A background on Kuwait and the socio-economic characteristics of its residents has been reported in Said (1982) and Hutchinson & Said (1990). It was possible to identify seven different household groups based on three nationalities (Kuwaitis, Arabs and Asians), and three dwelling unit types for Kuwaiti households and two dwelling unit types for Arab and Asian households.

The data base used in this study is compiled from a home interview survey conducted in 1987 and 1988. The collected information covers both household and personal information. Furthermore, a travel diary for each individual was recorded.

Work trips along with some selected variables describing household and head of the household characteristics were compiled in a single master file with 6270 records for households of all nationality groups.

Because many variables describing household characteristics were available in the master file, there was a need to identify which of these variables should be included in the analysis in which the GLM framework is utilized. This has been achieved through the analysis of the correlation matrix. Furthermore, simple linear regression analysis of trip rates and each of the possible household variables has been concluded. The analysis has shown that for Kuwaiti households the most important three variables are the household size (measured by total number of persons), car ownership and number of adults in the household.

Each of the three variables (when treated separately) explained a good proportion of the variation in the number of trips. When both household size and the number of adults in the household are used in the same equation, the single effect of the household size had a negative sign but the combined effect of the two variables had positive signs and a regression coefficient that is significantly different from zero. Since the number of adults-variable expresses the number of grown-up persons of the household-size variable, it was decided that the use of a variable that complements the number of adults-variable is more appropriate than the use of the household size. Subsequently, the household-size variable has been replaced by a new variable that represents the number of children in the household. The single effect of this variable had a negative sign which is consistent with the expectation that the increase of the number of children allows for fewer adults being able to work.

The number of children (0–18 years) ranges from 0 to 15, car ownership levels range from 0 to 9 cars per household and the number of adults ranges from 1 to 12.

The grouped data are in the form of a cross-classification table. The classifying variables are grouped as follows:

Number of children:	0, 1–3, 4–7, 8–11 and 12–15
Car ownership:	0–1, 2–3, 4–6 and 7–9
Adults in the household:	1–2, 3–5, 6–8 and 9–12

The above groupings are based on earlier investigations of household distribution relative to the above variables, (Hutchinson & Said 1990; Said & Young 1990; Said 1992).

2. THE GENERALIZED LINEAR MODEL APPROACH

A number of generalized linear model representations for the Kuwait trip data is presented along with an outline of the associated statistical methods of analysis. The models reflect the dependence of average Kuwaiti household work trip rates on the qualitative factor (house type) and the three quantitative variables (x_1 = number of children in the household, x_2 = number of cars owned per household and x_3 = number of adults in the household). Three house types are used: villas, government housing (mostly villas of moderate space and quality provided through the National Housing Authority—NHA) and other housing types such as flat and old-type arabic houses (others).

We shall use the following notation:

$x_{11}, x_{12}, \ldots, x_{1b}$	= the b observed values of x_1
$x_{21}, x_{22}, \ldots, x_{2c}$	= the c observed values of x_2
$x_{31}, x_{32}, \ldots, x_{3d}$	= the d observed values of x_3
cell (i, j, k, l)	= cell corresponding to the grouping of households of house type i, $x_1 = x_{1j}$, $x_2 = x_{2k}$, $x_3 = x_{3l}$
N	= total sample of households
n_{ijkl}	= households in cell (i, j, k, l), $\Sigma_i \Sigma_j \Sigma_k \Sigma_l n_{ijkl} = N$
Y_{ijklm}	= daily work trips observed for the m^{th} household in cell (i, j, k, l)
μ_{ijkl} and σ^2_{ijkl}	= mean and variance of household work trip numbers for population of households in cell (i, j, k, l)

The key objective is to set up a suitable statistical model for μ_{ijkl} to represent its dependence on the factors.

A priori, the within-cell variances $\{\sigma^2_{ijkl}\}$ are expected to increase as the means $\{\mu_{ijkl}\}$ increase and this invalidates the assumption of variance homogeneity which is made when applying the least squares method of analysis to the untransformed trip rates. As reported in Said & Young (1990), the ratios of the observed cell variances to the observed cell means were found to be approximately one, indicating that the distribution of trip rates within cells is approximately Poisson. This leads to two distinct possible approaches in the modelling and subsequent analysis of the data.

The first approach is based on stabilization of variance by working with the square root transformed observations $Y^*_{ijklm} = (1 + Y_{ijklm})^{1/2}$. Letting μ^*_{ijkl} denote the true mean for the transformed trip rates in cell (i, j, k, l), we write $Y^*_{ijklm} = \mu^*_{ijkl} + \epsilon^*_{ijklm}$ where $E(\epsilon^*_{ijklm}) = 0$ and $\text{var}(\epsilon^*_{ijklm}) = \sigma^2_*$ for all i, j, k, l, m, the variance being constant after transformation. If $\bar{Y}^*_{ijkl}$ denotes the observed mean of the transformed trip rates for cell (i, j, k, l) we have

$$\bar{Y}^*_{ijkl} = \mu^*_{ijkl} + \bar{\epsilon}^*_{ijkl} \tag{1}$$

where $\bar{\epsilon}^*_{ijkl} = \Sigma_m \epsilon^*_{ijklm}/n_{ijkl}$ with $E(\bar{\epsilon}^*_{ijkl}) = 0$, $\text{var}(\bar{\epsilon}^*_{ijkl}) = \sigma^2_*/n_{ijkl}$. The linear model representation for μ^*_{ijkl} is taken as

$$\begin{aligned} \mu^*_{ijkl} = \beta_{0i} &+ \beta_{1i} x_{1j} + \beta_{2i} x_{2k} + \beta_{3i} x_{3l} + \beta_{4i} x_{1j} x_{2k} \\ &+ \beta_{5i} x_{1j} x_{3l} + \beta_{6i} x_{2k} x_{3l} + \beta_{7i} x_{1j} x_{2k} x_{3l} \end{aligned} \tag{2}$$

The inclusion of the two- and three-variable interaction terms allows for departure from additivity of the effects of the quantitative variables x_1, x_2 and x_3. It is also seen that the regression coefficients are allowed to change with i, the index for house type.

We refer to the model given by (2) as the full model. It is fitted by weighted least squares by minimizing $\Sigma_i \Sigma_j \Sigma_k \Sigma_l n_{ijkl} (\bar{Y}^*_{ijkl} - \mu^*_{ijkl})^2$ where μ^*_{ijkl} satisfies (2). Letting $\hat{\mu}^*_{ijkl}$ denote the (weighted least squares) estimate of μ^*_{ijkl}, the goodness of fit of the model is measured by the deviance which for the above model is the weighted residual sum of squares $D = \Sigma_i \Sigma_j \Sigma_k \Sigma_l n_{ijkl} (\bar{Y}^*_{ijkl} - \hat{\mu}^*_{ijkl})^2$ with $V = a(bcd - 7)$ degrees of freedom, there being $abcd$ cell means and $7a$ β-coefficients in the model.

For interpretation of the results, estimates of the cell means for the untransformed trip rates are given by

$$\hat{\mu}^*_{ijkl} = \hat{\mu}^{*^2}_{ijkl} - 1 + \hat{\sigma}^2_* \tag{3}$$

where $\hat{\sigma}^2_* = D/V$ is the estimate of σ^2_* based on the fitted model.

Reduced model forms arise naturally from (2) by considering hypotheses of interest. For example, to test the hypothesis of no interaction effects involving x_1, x_2 and x_3 leads to the model

$$\mu^*_{ijkl} = \beta_{0i} + \beta_{1i}x_{1j} + \beta_{2i}x_{2k} + \beta_{3i}x_{3l} \tag{4}$$

If it is further assumed that house type has no effect, the new reduced model is

$$\mu^*_{ijkl} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2k} + \beta_3 x_{3l} \tag{5}$$

and so on. An examination of the increases in deviance as reduced models are considered, augmented by approximate F-tests, can be used to find a suitable model which is both concise and gives an adequate fit to the data.

In the second approach, specific use is made of the assumption of a Poisson for the within-cell trip rates. Thus Y_{ijklm} is now taken to be an observation from a Poisson distribution with mean μ_{ijkl}, $m = 1, \ldots, n_{ijkl}$. To ensure that the mean is positive, we use the logarithmic model

$$\begin{aligned}\log \mu_{ijkl} = {} & \beta_{0i} + \beta_{1i}x_{1j} + \beta_{2i}x_{2k} + \beta_{3i}x_{3l} + \beta_{4i}x_{1j}x_{2k} \\ & + \beta_{5i}x_{1j}x_{3l} + \beta_{6i}x_{2k}x_{3l} + \beta_{7i}x_{1j}x_{2k}x_{3l}\end{aligned} \tag{6}$$

In the analysis, the cell total number of trips $Y_{ijkl} = \Sigma_m Y_{ijklm}$ is used. This has a true mean $\mu'_{ijkl} = n_{ijkl}\mu'_{ijkl}$ satisfying

$$\log \mu'_{ijkl} = \log n_{ijkl} + \log \mu_{ijkl} \tag{7}$$

where $\log \mu_{ijkl}$ is given by (6). The model is fitted by the method of maximum likelihood and the calculations can be performed using the GLIM package by specifying a Poisson error distribution, a logarithmic link function for the means and declaring $\log n_{ijkl}$ as an offset variable. The goodness of fit of the model is measured by the deviance which, for the Poisson model is

$$D = 2\Sigma_i \Sigma_j \Sigma_k \Sigma_l \{ Y_{ijkl} \log (Y_{ijkl}/\hat{\mu}'_{ijkl}) - (Y_{ijkl} - \hat{\mu}'_{ijkl})\} \tag{8}$$

where $\hat{\mu}'_{ijkl}$ is the maximum likelihood estimate of μ_{ijkl} (see McCullagh & Nelder 1983)). If the model is correct, then D approximately has the chi-square distribution with $a(bcd - 7)$ degrees of freedom. Reduced models can be fitted and assessed by examining the increases in deviance using chi-square tests.

The regression approaches that have been described are easily modified in order to handle the more practical case in which the values of the quantitative variables x_1, x_2 and x_3 are grouped. This is done by using variable values equal to the central values of the groups.

The use of ANOVA models for handling trip rates of cross-classified household data case has been proposed in the literature (see for example Dobson (1976)). This approach leads to a modification of the models for the means. For example the model given by (2) now takes the known ANOVA model for four factors

$$\begin{aligned}\mu^*_{ijkl} = {} & \mu + A_i + B_j + C_k + D_l + (AB)_{ij} + (AC)_{ik} + (AD)_{il} \\ & + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + (ABC)_{ijk} + (ABD)_{ijl} \\ & + (ACD)_{ikl} + (BCD)_{jkl} + (ABCD)_{ijkl}\end{aligned} \tag{9}$$

A similar representation is used for the logarithmic model of the mean given by (6).

The fitting methods and techniques for comparing goodness of fit of models are the same as those described earlier. ANOVA models ignore the quantitative nature of the variables x_1, x_2 and x_3. Consequently, models containing more parameters will be needed to obtain the same degree of precision for fitted values as a corresponding regression model.

Kuwaiti households living in the three housing types will be considered in the rest of this paper. The models to be investigated for ungrouped data include regression models using transformed trip data and Poisson models with logarithmic link. Both regression and ANOVA models are used with the cross-classified (grouped) data.

3. RESULTS OF THE REGRESSION ANALYSIS OF UNGROUPED DATA

Said & Young (1990) described the results of illustrative runs using the regression specifications of GLM and only two quantitative variables: household size and car ownership.

In this section we report the findings from the statistical analysis when the three quantitative variables x_1, x_2 and x_3 identified earlier are used in the ungrouped format. It has been pointed out earlier that the household size variable has been replaced by a variable describing the number of children in the household, and a new variable describing the number of adults has been introduced. Furthermore, a new qualitative variable representing the house type is now considered.

Two approaches are examined, the weighted least square analysis of the transformed trip rates which are calculated using the square root transformation, and the maximum likelihood analysis of the untransformed trip rates using a Poisson regression model with a logarithmic link function for the means.

3.1. A MODEL BASED ON TRANSFORMED TRIP RATES

The starting model for the statistical analysis is the full model given by (2). To examine the effect of house type, this model was fitted together with the reduced models

$$\mu^*_{ijkl} = \beta_{0i} + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_3 X_{3l} + \beta_4 X_{1j} X_{2k} + \beta_5 X_{1j} X_{3l} + \beta_6 X_{2k} X_{3l} + \beta_7 X_{1j} X_{2k} X_{3l} \qquad (10)$$

$$\mu^*_{ijkl} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_3 X_{3l} + \beta_4 X_{1j} X_{2k} + \beta_5 X_{1j} X_{3l} + \beta_6 X_{2k} X_{3l} + \beta_7 X_{1j} X_{2k} X_{3l} \qquad (11)$$

The model given by (10) allows for a house type effect on the average level of the trip rate but takes the effects of x_1, x_2 and x_3 to be the same within each house type. The reduced model given by (11) implies that there are no house type effects at all. Fits of the models given by (2), (10) and (11) gave deviances equal to 109.9, 112.4 and 112.6 with 971, 985 and 987 degrees of freedom, respectively. The F-statistic for comparing the models given by (2) and (10) has value $\{(112.4 - 109.9)/(985 - 971)\}/(112.4/971\} = 1.54$ and the F-statistic for comparing the models given by (10) and (11) has value $\{(112.6 - 112.4)/2\}/\{112.4/985\} = 0.88$. If these values are referred to the upper critical values of the $F_{14,971}$ and $F_{2,985}$ distributions, the results are not significant even at the 10% level, so there is no evidence of any house type effects.

The model form given by (11) is now taken as the new full working model. Reduced models obtained by setting $\beta_7 = 0$, $\beta_6 = \beta_7 = 0$, etc. have been fitted. The deviances and associated degrees of freedom are shown in Table 1 for a selected subset of the models. Values of the F statistics for comparing the fits of appropriate pairs of models are also given. Values greater than 3.84 which is the upper 5% critical value of the $F_{1,\infty}$ distribution indicate that the relevant β coefficient should not be excluded from the model.

The results show that the three-variable interaction term $\beta_7 X_{1j} X_{2k} X_{3l}$ may be excluded, together with the two-variable interaction terms $\beta_4 X_{1j} X_{2k}$ and $\beta_5 X_{1j} X_{3l}$. However, the comparison of the models which only differ in the inclusion or exclusion of $\beta_6 X_{2k} X_{3l}$ shows that this term should not be excluded. Also it is seen that although all interactions involving X_1 are not significant, the term $\beta_1 X_{1j}$ should not be excluded from the model. We therefore conclude that the model

$$\mu^*_{ijkl} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_3 X_{3l} + \beta_6 X_{2k} X_{3l} \tag{12}$$

for the means of the transformed trip rates is the most concise model. The fit of this model gives the estimates $\hat{\beta}_0 = 1.224$, $\hat{\beta}_1 = -0.0153$, $\hat{\beta}_2 = 0.0471$, $\hat{\beta}_3 = 0.0240$, $\hat{\beta}_6 = 0.00661$ and deviance $= 112.8$. Using (3), the estimated means for the untransformed data are

$$\hat{\mu}_{ijkl} = (1.224 - 0.0153 X_{1j} + 0.0471 X_{2k} + 0.0240 X_{3l} + 0.00661 X_{2k} X_{3l})^2 - 0.886 \tag{13}$$

3.2. A POISSON MODEL

The untransformed trip rates are taken to have approximately a Poisson distribution with a logarithmic link function for the means, the initial model being given by (6). The effect of house type was first examined by comparing this model with the reduced models

$$\log \mu_{ijkl} = \beta_{0i} + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_3 X_{3l} + \beta_4 X_{1j} X_{2k} + \beta_5 X_{1j} X_{3l} + \beta_6 X_{2k} X_{3l} + \beta_7 X_{1j} X_{2k} X_{3l} \tag{14}$$

$$\log \mu_{ijkl} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_3 X_{3l} + \beta_4 X_{1j} X_{2k} + \beta_5 X_{1j} X_{3l} + \beta_6 X_{2k} X_{3l} + \beta_7 X_{1j} X_{2k} X_{3l} \tag{15}$$

Table 1. Deviances and F statistic values for analysis of transformed trip rates (ungrouped data).

Number	Terms in Model	df	Dev.	F-Statistic
1	$x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3$	987	112.6	
2	$x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3$	988	112.6	$F_{2,1} = 0.000$
3	$x_1, x_2, x_3, x_1x_2, x_1x_3$	989	114.4	$F_{3,2} = 15.79^*$
4	$x_1, x_2, x_3, x_1x_2, x_2x_3$	989	112.7	$F_{4,2} = 0.88$
5	$x_1, x_2, x_3, x_1x_3, x_2x_3$	989	112.8	$F_{5,2} = 1.78$
6	x_1, x_2, x_3, x_1x_2	990	114.5	$F_{6,4} = 15.80^*$
7	x_1, x_2, x_3, x_1x_3	990	114.7	$F_{7,5} = 16.66^*$
8	x_1, x_2, x_3, x_2x_3	990	112.8	$F_{8,4} = 0.88$
9	x_2, x_3, x_2x_3	991	118.1	$F_{9,8} = 46.51$

(a) $F_{4,2}$ for example, is the F-statistic for comparing model number 4 with model number 2 and equals $[(112.7 - 112.6)/(989 - 988)]/(112.6/988) = 0.88$.
(b) Values of F-statistics are compared with 3.84 (significance level = 5%).

Fits of the models given by (6), (14) and (15) by maximum likelihood gave deviances equal to 887.5, 904.1 and 905.1 with 971, 985 and 987 degrees of freedom, respectively. The X^2 statistic for comparing the models given by (6) and (14) has value $904.1 - 887.5 = 16.6$ with 14 degrees of freedom. Similarly, the X^2 statistic for comparing the models given by (14) and (15) has value 1.0 with 2 degrees of freedom. Since $X^2_{14}(0.9) = 21.06$ and $X^2_2(0.9) = 4.61$, neither value is significant even at the 10% level, so again there is no evidence of any house type effect. Taking (15) as the new full working model, reduced models obtained by setting subsets of the regression coefficients equal to zero were fitted. The results for a selected subset of the models are displayed in Table 2. These show that only the interaction terms $\beta_7 X_{1j} X_{2k} X_{3l}$ and $\beta_5 X_{1j} X_{3l}$ can be excluded, so the model

$$\log \mu_{ijkl} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2k} + \beta_4 X_{1j} X_{2k} + \beta_6 X_{2k} X_{3l} \tag{16}$$

is adopted. The fit for this model gives the estimates $\hat{\beta}_0 = -0.574$, $\hat{\beta}_1 = -0.672$, $\hat{\beta}_2 = 0.173$, $\hat{\beta}_3 = 0.139$, $\hat{\beta}_4 = 0.0082$, $\hat{\beta}_6 = -0.0091$. The estimated means for the untransformed trip rates are, therefore, given by:

$$\begin{aligned}\hat{\mu}_{ijkl} = \exp(&-0.574 - 0.0672X_{1j} + 0.173X_{2k} + 0.139X_{3l}\\ &+ 0.0082X_{1j}X_{2k} - 0.0091X_{2k}X_{3l})\end{aligned} \tag{17}$$

The estimated means given by (13) and (17) have been calculated for a wide variety of combinations of values of x_1, x_2, x_3 and gave very similar values, indicating that the choice between the two models is not of crucial importance. This finding will be demonstrated in the next section dealing with grouped data.

4. REGRESSION ANALYSIS OF GROUPED DATA

The regression analysis using ungrouped data utilized all combinations of values of the three explanatory variables within each house type, for which at least one trip rate observation was available. As a consequence, the mean trip rates $\bar{Y}_{ijkl}$ are in many cases based on very small cell household frequencies. Since forecasts would be needed for future population numbers of households occurring for combinations of specified values of the explanatory variables, it is advantageous to have a fairly broad grouping of the values of the variables. The grouping specified in section 2 has been used which

Table 2. Deviances and X^2 statistic values for analysis of trip rates (poisson model for ungrouped data).

Number	Terms in Model	df	Dev.	X^2
1	x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$	987	906.0	
2	x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3	988	907.0	$X^2_{2,1} = 1.0$
3	x_1, x_2, x_3, x_1x_2, x_1x_3	989	914.2	$X^2_{3,2} = 7.2$
4	x_1, x_2, x_3, x_1x_2, x_2x_3	989	907.1	$X^2_{4,2} = 0.1$
5	x_1, x_2, x_3, x_1x_3, x_2x_3	989	913.6	$X^2_{5,2} = 6.5$
6	x_1, x_2, x_3, x_1x_2	990	914.4	$X^2_{6,4} = 7.3$
7	x_1, x_2, x_3, x_1x_3	990	919.6	$X^2_{7,5} = 6.0$
8	x_1, x_2, x_3, x_2x_3	900	915.8	$X^2_{8,5} = 2.2$

(a) $X^2_{4,2}$ for example, is the X^2-statistic for comparing model number 4 with model number 2 and equals $907.1 - 907.0 = 0.1$.

(b) Values of X^2-statistics are compared with 3.84 (significance level = 5%)

leads to a total of $5 \times 4 \times 4 = 80$ cells. House type specification is now ignored, following the results of the analyses described in section 3.

Table 3 shows the observed mean trip rates, with the cell frequencies of households being shown in parentheses. Ten of the 80 cells have zero frequencies and 19 other cells contain less than 5 households. Figure 1 illustrates these trip rates graphically. Both the table and figure represent the results of a routine application of the cross-classification approach to trip generation. Zero- and low-frequency outlying cells pose problems. While it is not possible to calculate trip rates for zero-frequency cells, the low-frequency cells can themselves have unreliable trip rate estimates.

For grouped data, the models and the associated statistical analyses are the same as those discussed in section 4 for the ungrouped data. The values of the explanatory variables are taken as the mid-point of the ranges selected for each variable. The broad findings are the same as those described in section 3 for ungrouped data.

For the transformed trip rates, it was found that all interactions could be excluded from the model except the term $\beta_6 X_{2k} X_{3l}$. The equation for the estimates of the cell means was found to be

$$\hat{\mu}_{jkl} = (1.253 - 0.0144X_{1j} + 0.0391X_{2k} + 0.0187X_{3l} + 0.0072X_{2k}X_{3l})^2 - 0.827 \quad (18)$$

For the fit of the Poisson model with logarithmic link for the means, only the three-variable interaction term $\beta_7 X_{1j} X_{2k} X_{3l}$ and the two-variable interaction term $\beta_5 X_{1j} X_{3l}$ could be excluded. The equation for the estimated cell mean trip rates was found to be

$$\hat{\mu}_{jkl} = \exp(-0.483 - 0.0643X_{1j} + 0.1496X_{2k} + 0.1204X_{3l} + 0.00804X_{1j}X_{2k} - 0.00589X_{2k}X_{3l}) \quad (19)$$

Table 3. Trip rates and household frequencies for Kuwaiti households.

Adults in the household	Car ownership	Number of children under 18 years				
		0	1–3	4–7	8–11	12–15
1–2	0–1	0.59 (17)	0.62 (68)	0.62 (120)	0.43 (35)	0.00 (1)
	2–3	1.18 (17)	1.08 (148)	1.07 (203)	0.55 (38)	0.00 (2)
	4–6	1.00 (1)	0.86 (7)	1.09 (11)	0.50 (2)	–
	7–9	–	–	–	–	–
3–5	0–1	0.86 (14)	0.73 (30)	0.54 (57)	0.53 (40)	0.45 (11)
	2–3	1.27 (33)	1.16 (149)	0.93 (207)	0.70 (81)	0.48 (19)
	4–6	1.91 (11)	1.67 (86)	1.65 (83)	1.25 (24)	2.11 (9)
	7–9	3.00 (1)	2.50 (2)	2.17 (6)	2.50 (2)	–
6–8	0–1	2.00 (2)	0.80 (5)	1.88 (8)	0.88 (8)	0.33 (6)
	2–3	2.33 (3)	1.72 (47)	1.61 (62)	1.34 (41)	1.20 (20)
	4–6	3.00 (12)	2.45 (98)	2.20 (107)	1.78 (37)	1.80 (10)
	7–9	–	3.00 (15)	3.11 (18)	4.33 (6)	3.50 (2)
9–12	0–1	3.00 (1)	3.50 (2)	1.00 (2)	2.00 (2)	0.00 (1)
	2–3	2.00 (1)	–	1.55 (11)	1.25 (4)	1.00 (6)
	4–6	–	3.00 (11)	3.58 (26)	3.10 (21)	2.70 (10)
	7–9	5.00 (1)	4.39 (18)	4.35 (17)	3.42 (12)	4.00 (3)

– indicates 0 household frequency
(17) cell frequency; households

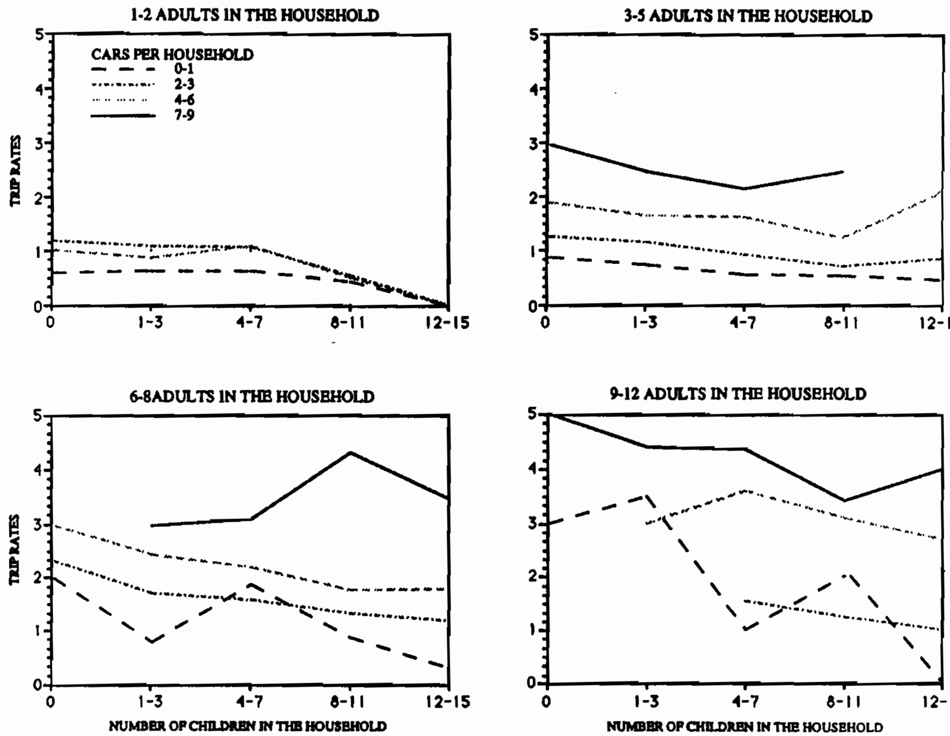

Fig. 1. Mean trip rate estimates based on the routine use of the cross-classification analysis procedure.

The model fits in (18) and (19) for the grouped data are very close to the model fits in (13) and (17) of the ungrouped data.

Table 4 shows the fitted means using (18) and (19). Figure 2 shows graphically the model fits using (18). Table 4 and Fig. 2 show that the model estimates of the mean trip rates are similar in magnitude. The agreement between the model estimates and the observed mean trip rates in Table 3 is generally good, for cells with high household frequencies. Unlike the case when the routine application of the cross-classification analysis procedure is used, the fitted models produce mean trip rate estimates for all cells in the cross classification including cells with zero observed household frequencies.

The two fitted models show consistent increase of mean trip rate with increase in the car ownership and number of adults in the household. The estimates of the mean trip rates decrease systematically as the number of children in individual households increases. The negative signs associated with the single effect of the number of children variable are consistent with expectations. An increase in the number of children results in fewer female adults joining the labour force.

5. ANOVA OF GROUPED DATA

The results obtained by applying ANOVA techniques to the grouped data are presented in this section. The model given by (9) is the full ANOVA model and its estimated form reproduces exactly the observed mean trip rates given in Table 3. We

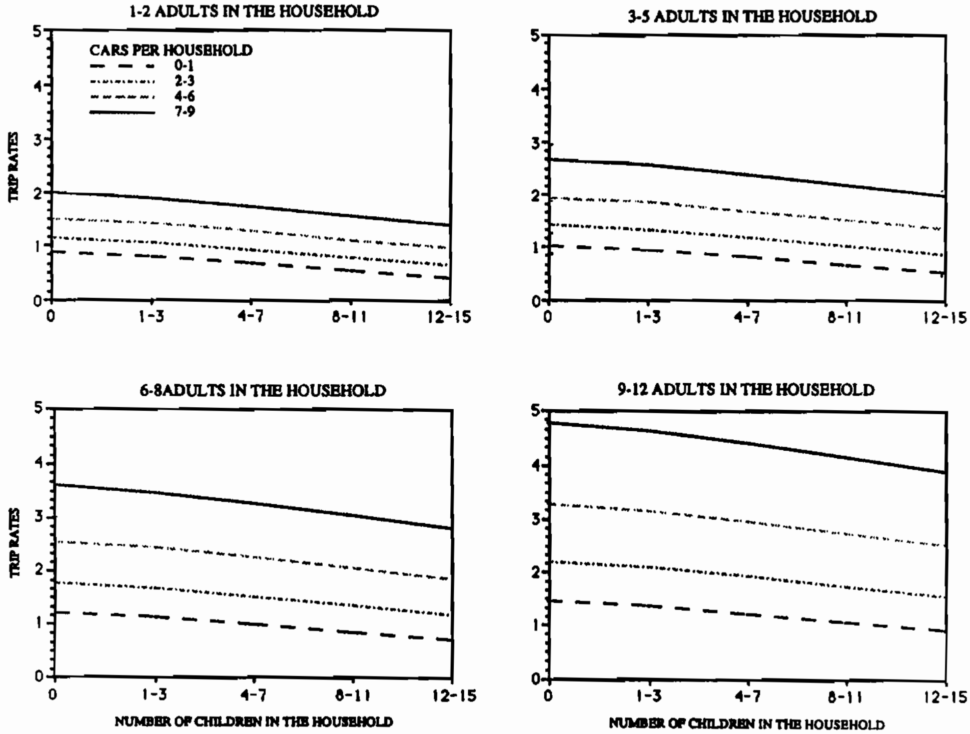

Fig. 2. Mean trip rate estimates of cross-classified household groups using a regression model of transformed trips.

wish to find the simplest reduced form of the model which adequately represents the data.

The regression analysis has shown that the house type effect is negligible. The starting model for the cell means based on the transformed trip rates is

$$\mu^*_{jkl} = \mu + B_j + C_k + D_l + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + (BCD)_{jkl} \qquad (20)$$

B_j is the main effect of the jth level for number of children, C_k is the effect of the kth level of car ownership, D_l is the effect of the lth level of number of adults, and $(BC)_{jk}$, $(BD)_{jl}$, $(CD)_{kl}$, $(BCD)_{jkl}$ represent the interactions.

The statistical analysis of the deviances obtained when the model given by (20) and its reduced forms are fitted shows that the only significant interaction is $(CD)_{kl}$, so the model

$$\mu^*_{jkl} = \mu + B_j + C_k + D_l + (CD)_{kl} \qquad (21)$$

is adopted. The least squares estimates of the parameters are

$$\hat{\mu} = 1.331$$

$$\hat{B}_j = (0, -0.070, -0.094, -0.191, -0.199) \quad \text{for } j = 1, \ldots, 5$$

$$\hat{C}_k = (0, 0.155, 0.130, 0.578) \quad \text{for } k = 1, \ldots, 4$$

$$\hat{D}_l = (0, 0.019, 0.197, 0.438) \quad \text{for } l = 1, \ldots, 4$$

Table 4. Estimated mean trip rates based on fits using grouped data of (i) classical model for transformed trip rates, (ii) poisson model for untransformed trip rates

		Number of children under 18 years									
Adults in the household	Car ownership	0 (i)	0 (ii)	1–3 (i)	1–3 (ii)	4–7 (i)	4–7 (ii)	8–11 (i)	8–11 (ii)	12–15 (i)	12–15 (ii)
1–2	0–1	0.88	0.79	0.80	0.70	0.68	0.57	0.54	0.44	0.41	0.35
	2–3	1.15	1.05	1.07	0.96	0.93	0.82	0.78	0.68	0.64	0.57
	4–6	1.51	1.49	1.43	1.42	1.28	1.30	1.11	1.18	0.96	1.06
	7–9	2.00	2.28	1.90	2.27	1.74	2.27	1.56	2.26	1.38	2.25
3–5	0–1	1.03	1.06	0.95	0.94	0.82	0.76	0.67	0.59	0.53	0.46
	2–3	1.42	1.37	1.33	1.25	1.18	1.07	1.02	0.89	0.87	0.74
	4–6	1.95	1.88	1.86	1.78	1.69	1.63	1.52	1.48	1.34	1.34
	7–9	2.67	2.74	2.57	2.73	2.38	2.72	2.18	2.71	1.99	2.70
6–8	0–1	1.21	1.51	1.13	1.34	0.49	1.08	0.84	0.84	0.70	0.66
	2–3	1.76	1.88	1.67	1.72	1.51	1.47	1.34	1.22	1.17	1.02
	4–6	2.53	2.46	2.42	2.34	2.24	2.15	2.05	1.94	1.85	1.75
	7–9	3.59	3.41	3.46	3.40	3.26	3.39	3.03	3.38	2.80	3.36
9–12	0–1	1.44	2.28	1.36	2.02	1.21	1.63	1.05	1.27	0.90	1.00
	2–3	2.19	1.72	2.09	2.48	1.92	2.12	1.73	1.77	1.55	1.48
	4–6	3.27	3.39	3.15	3.22	2.95	2.95	2.73	2.67	2.52	2.41
	7–9	4.78	4.41	4.64	4.40	4.41	4.38	4.15	4.36	3.90	4.34

$$(\widehat{CD})_{kl} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.031 & -0.007 & -0.301 \\ 0 & 0.208 & 0.199 & 0.259 \\ 0 & -0.038 & 0.037 & 0 \end{pmatrix} \quad \text{for } k = 1, \dots, 4 \text{ and } l = 1, \dots, 4$$

The observed deviance is 5.56. These estimates can be used in the fitted model

$$\hat{\mu}_{jkl} = (\hat{\mu} + \hat{B}_j + \hat{C}_k + \hat{D}_l + (\widehat{CD})_{kl})^2 - 0.891 \tag{22}$$

to provide estimates of the cell means μ_{jkl} for the untransformed data.

As a second approach, let $Y_{\cdot jkl}$ and $n_{\cdot jkl}$ denote respectively, the total number of trips and total number of households in cell (j, k, l). Then $Y_{\cdot jkl}$ is taken to approximately have a Poisson distribution with mean $n_{\cdot jkl}\,\mu_{\cdot jkl}$ where

$$\mu_{\cdot jkl} = \exp(\mu + B_j + C_k + D_l + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + (BCD)_{jkl}) \tag{23}$$

Chi-square tests applied to the differences between the deviances when the model given by (23) and its reduced forms are fitted show that the only significant interaction is $(CD)_{jk}$. Maximum likelihood estimates of the parameters were found to be

$$\hat{\mu} = -0.311$$

$$\hat{B}_j = (0, -0.170, -0.232, -0.454, -0.456) \quad \text{for } j = 1, \dots, 5$$

$$\hat{C}_k = (0, 0.553, 0.479, 0.732) \quad \text{for } k = 1, \dots, 4$$

$$\hat{D}_l = (0, 0.064, 0.711, 0.1262) \quad \text{for } l = 1, \dots, 4$$

$$(\widehat{CD})_{kl} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.083 & -0.237 & -0.870 \\ 0 & 0.494 & 0.165 & 0.050 \\ 0 & -0.612 & 0.300 & 0 \end{pmatrix} \quad \text{for } k = 1, \ldots, 4 \text{ and } l = 1, \ldots, 4$$

These estimates can be used to provide estimates of the cell mean trip rates using

$$\hat{\mu}_{jkl} = \exp(\hat{\mu} + \hat{B}_j + \hat{C}_k + \hat{D}_l + (\widehat{CD})_{kl}) \qquad (24)$$

The negative signs of the single effects involving the number of children in the household enforces the initial findings that this variable has a structural effect on household trip rates and that more children in the household results in fewer work trips. The estimated mean trip rates were calculated using (22) and (24). The results show that the two ANOVA models provide similar estimates with the exception of a few cells in the highest car ownership category.

Figure 3 compares the estimates of the mean trip rates of the fitted ANOVA model in (22) with the trip rates calculated through the routine application of the cross-classification analysis shown in Table 3. The comparison is done so that cells with low and high household frequencies are shown separately. ANOVA model estimates are generally consistent with those calculated in Table 3 based on the cross-classification analysis. Low-frequency outlying cells show the greatest discrepancies. The ANOVA estimates are obviously more reliable being based on the entire classification and not the particular cells with low frequencies.

Figure 4 shows the estimated mean trip rates of the fitted ANOVA model in (22). The figure illustrates that there is no longer a smooth relation between the estimated means and the car ownership level (see for example the trip rate estimate for 0–1 and 2–3 cars where the number of adults variable is 1–2 or 9–12). This occurs because the ANOVA approach ignores the quantitative levels of this explanatory variable and so tends to give estimated means shadowing the observed means more closely than was obtained by using the regression approach. However, the precision of the estimated means using the ANOVA approach is less than that using the regression approach.

6. CONCLUSION

When dealing with areas of diversified socio-economic features such as the presence of several structurally different household groups, current trip generation techniques fail to respond to the modelling needs that emerge in these circumstances. Said (1992) has indicated that the zonal regression analysis and household category analysis are bound to produce unsatisfactory results because of the loss of considerable variance when using the first approach and the poor reliability of trip rate estimates that would result from the second approach.

A statistical modelling framework that includes a variety of model options for estimating household-level trip generation rates has been proposed. The modelling framework is based on the general linear model and builds on the ideas of Nelder & Wedderburn (1972), Dobson (1983), and Said & Young (1990).

Household mean trip rates are modelled as linear functions of quantitative and qualitative variables describing household structure and life cycle, car ownership and house type characteristics. Six different models have been proposed and fitted to

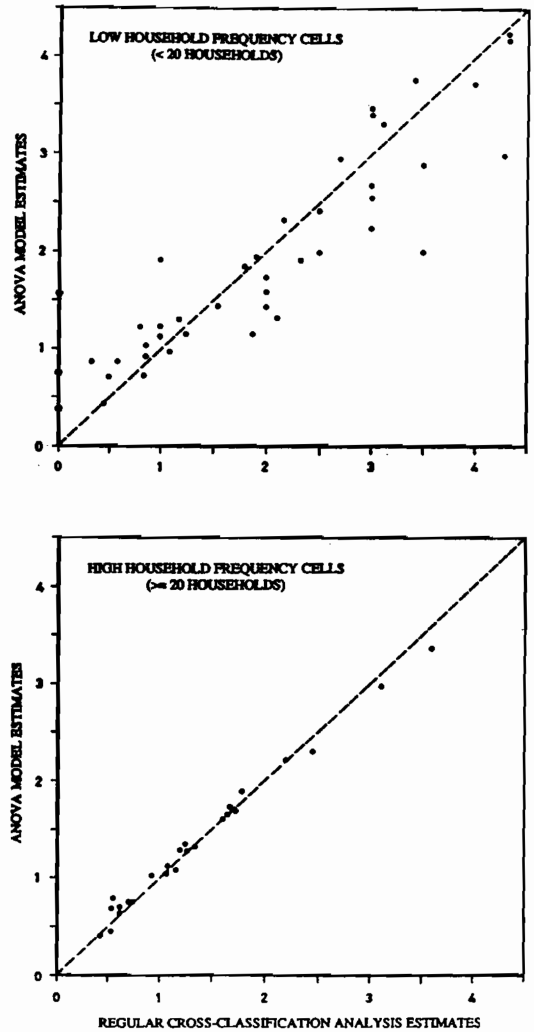

Fig. 3. A comparison between mean trip rate estimates of ANOVA and the routine cross-classification procedure.

observed data which are of the regression and ANOVA types. In each case, statistical tests were used to establish the significant single and higher order interaction effects that need to be included in the proposed models.

The proposed models used data for households classified into all possible levels of the explanatory quantitative and qualitative variables as well as households cross-classified according to selected groupings of these variables. The latter case is of practical significance since it will be difficult to forecast household numbers for all the possible levels of the explanatory variables if these models are to be used for forecasting and planning purposes.

The calibration of the models to the Kuwait data has indicated that the house type has no real effect on the mean trip rates of Kuwaiti households. The fits of both regression and ANOVA models have indicated that the single effects of the variables

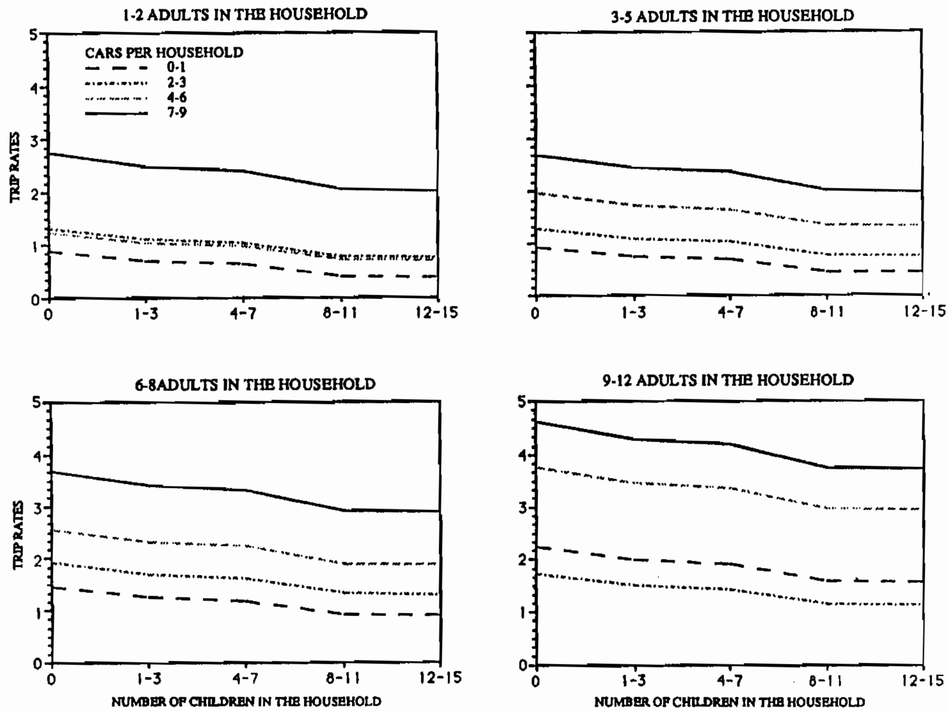

Fig. 4. Mean trip rate estimates of cross-classified household groups using ANOVA model of transformed trips.

describing the number of children and number of adults in the household, and car ownership are all significant as well as some of the first-order interaction effects of these variables. Household trip rates were found to decrease with the increase in the number of children in the household.

The estimates of the mean trip rates of the models using the cross-classified data based on regression and ANOVA model calibrations have been assessed and compared with the mean trip rate estimates based on a routine use of the cross-classification analysis procedure. A number of improvements were identified and include the more reliable estimates of trip rates of household groups in low-frequency outlying cells. Obviously, the ability to select classifying variables on sound statistical bases is a main feature of this approach.

The results of this study have been assembled in the form of charts that give the estimated mean trip rates of households of different categories. These rates could be used along with forecasts of future household numbers in order to estimate the work trips generated by individual zones.

There are potential extensions of this work. One extension is the treatment of Arab and foreign Asian household groups in Kuwait which are significantly different in their characteristics from Kuwaiti households and represent more than 60 percent of the total population. Another extension is the study of the stability of the household characteristics found to influence mean household trip rates. This is not a straight-forward task given the structural demographic and economic changes that have taken place in Kuwait over the last two decades and the changes that are bound to happen in the future. A provision must exist for the study of the implications of related policy changes that have an impact on the social mix. A third but related extension is finding ways to forecast household groups according to the selected characteristics.

The spatial and social differentiations of the nature and scale observed in Kuwait and other similar communities must have further implications for the requirements of subsequent transport modelling phases such as modal split and trip distribution analyses. These requirements need to be recognized and reflected in model design, calibration procedures and applications in planning. This needs to be done to a level of detail that is compatible and in harmony with that proposed for the trip generation phase.

ACKNOWLEDGMENTS

This research was funded by Kuwait University, Projects EV030 and EV048 and Kuwait Foundation for the Advancement of Sciences, Project 870803.

REFERENCES

Baker, R. & Nelder, J. 1978. GLIM Manual (release 3), Numerical Algorithms Groups, Oxford, U.K.

Button, K.J. 1976. Category analysis and multiple regression models of trip generation: A possible reconciliation, International Journal of Transport Economics, **3:** 19–27.

Dobson, R. 1976. The general linear model analysis of variance: Its relevance to transportation planning and research. Socio-Economic Planning Science **10:** 231–5.

Dobson, A.J. 1983. An introduction to statistical modelling. Chapman and Hall, London.

Dobson, R. and McGarvey, W.E. 1977. An empirical comparison of disaggregate category and regression trip generation analysis techniques. Transportation **6:** 287–307.

Federal Highway Administration, 1976. Guidelines for trip generation analysis, U.S. Department of Transportation, Washington, D.C.

Hutchinson, B.G. 1974. Principles of urban transport systems planning. McGraw-Hill, New York.

Hutchinson, B.G. and Said, G.M. 1990. Spatial differentiation, transport demands and transport model design in Kuwait. Transport Reviews **10(2):** 91–112.

Kassoff, H. and Deutschman, H. 1969. Trip generation: A critical appraisal. Highway Research Record 297, Highway (Now Transportation) Research Board, Washington, D.C., 15–30.

McCullagh, P. and Nelder, J.A. 1983. Generalized linear models, Chapman and Hall, London.

Nelder, J.A. and Wedderburn, R.W. 1972. Generalized linear models, Journal of the Royal Statistical Society A **135(3):** 470–85.

Rickard, J.M. 1989. Application of a generalized linear modelling approach to category analysis, Transportation Planning and Technology **13(2):** 121–9.

Said, G.M. 1992. Modelling household car ownership in the Gulf States. Journal of Transport Economics and Policy **26:** 121–38.

Said, G.M. 1982. An overview of Transport in Kuwait. Transport Reviews **2(4):** 321–48.

Said, G.M. and Young, D.H. 1990. A general linear model framework for estimating work trip rates of households in Kuwait, Transportation Research A, **24A(3):** 187–200.

Stopher, P. & McDonald, K. 1983. Trip generation by cross-classification: An alternative methodology. Transportation Research Record 944, Transportation Research Board, Washington D.C., 84–91.

Wootton, H.J. & Pick, G.W. 1967. A model for trip generation by households, Journal of Transport Economic and Policy **1:** 137–53.

(Received 13 January 1993, revised 25 January 1994)

نماذج إحصائية لمعدل رحلات العمل للأسر الكويتية

جلال مصطفى سعيد، ديفيد هاميلتون * وحسن خليل ابراهيم
قسم الهندسة المدنيه، كلية الهندسة، جامعة القاهرة ـ فرع الفيوم
الفيوم ـ جمهورية مصر العربية

خلاصة

تعتمد معظم نماذج توليد الرحلات المستخدمة في دراسات تخطيط النقل على اساليب الانحدار على مستوى اقسام منطقة الدراسة أو على اسلوب التصنيف العرضي. ولقد كان استخدام اسلوب النموذج الخطي العام موضوعا لبعض الابحاث في مجال تقدير معدلات رحلات الأسر ولكنه لم يتم الكشف بعد عن امكانياته العديدة في هذا المجال حيث احتوت المراجع على اعداد قليلة من التطبيقات. إن استخدام النموذج الخطي العام يقدم بعض المزايا في هذا المجال منها إنه يتيح المجال لاقتراح نماذج احصائية عديدة وتقييم واختبار كل منها ويتيح كذلك اختيار المتغيرات المستقلة بطريقة منطقية.

وفي هذا البحث يتم بصورة مختصرة وصف النموذج الخطي العام والنماذج الاحصائية المتعددة التي تتاح عند استخدامه لتقدير معدل الرحلات والتي تندرج تحتها نماذج الانحدار، ونماذج تحليل التغير مع ما يستتبع ذلك من امكانية افتراض بدائل لاشكال دالة الوصل وكذلك التوزيع الاحصائي المحتمل للمتغير المستقل والذي يعبر عن عدد الرحلات للأسرة.

ويستخدم البحث بيانات عن رحلات العمل وخصائص الأسر في الكويت والتي تم استخراجها من استقصاء منزلي وذلك لتوضيح الطبيعة العملية للأسلوب المقترح ولقد تم التعامل مع الأسر الكويتية بحسب نوع المسكن ووجد ان معدلات رحلات العمل تتأثر بمتغيرات تمثل ملكية السيارة للأسرة وعدد الأفراد البالغين وعدد الأطفال بها وقد وجد ان التأثير المشترك لبعض هذه المتغيرات هام في بعض الحالات.

*عنوان المراسله: قسم الرياضيات ـ جامعة غرب لندن ـ المملكة المتحدة