THE APPLICATION OF A LOGISTIC REGRESSION MODEL FOR PREDICTING PREFERENCES OF TRANSPORT SYSTEM USERS

A. BRZEZIŃSKI[1], K. BRZEZIŃSKI[2], T. DYBICZ[3], Ł. SZYMAŃSKI[4]

Within the INMOP 3 research project, an attempt was made to solve a number of problems associated with the methodology of modelling travel in urban areas and the application of intermodal models. One of these is the ability to describe the behaviour of transport system users, when it comes to making decisions regarding the selection of means of transport and searching for relationships between travel describing factors and the decisions made in regard of means of transport choice.

The paper describes a probabilistic approach to the determination of modal split, and the application of a logistic regression model to determine the impact of variables describing individual and mass transport travels on the probability of selecting specific means of transport. Travels in local model of Warsaw city divided into 9 motivation groups were tested, for which ultimately 8 models were developed, out of which 7 were deemed very well fitted (obtained $pseudoR^2$ was well above 0.2).

Keywords: transport system, travel modelling, modal split, logistic regression, logit model

1. INTRODUCTION

[1] Warsaw University of Technology, Faculty of Civil Engineering, Institute of Roads and Bridges, ul. Lecha Kaczyńskiego 16, 00-637 Warsaw, mail: a.brzezinski@il.pw.edu.pl

[2] Warsaw University of Technology, Faculty of Civil Engineering, Institute of Roads and Bridges, ul. Lecha Kaczyńskiego 16, 00-637 Warsaw, mail:k.brzezinski@il.pw.edu.pl

[3] Warsaw University of Technology, Faculty of Civil Engineering, Institute of Roads and Bridges, ul. Lecha Kaczyńskiego 16, 00-637 Warsaw, mail:t.dybicz@il.pw.edu.pl

[4] TransEko Design-Consulting Office, l.szymanski@transeko.pl

The ability to describe the behaviour of transport system users in terms of the decisions they make regarding the selection of means of transport is the basis for developing travel models, which are used for analysing and predicting traffic. It is associated with searching for relationships between travel describing factors and the decisions made in relation to means of transport. The decisions as to which form of transport to choose are determined by various factors related to the travel itself, as well as the individual preferences of transport system users. When developing modal split models, it is impossible, due to data availability, to take into account all of the factors, therefore, it must be recognized that the choice mechanism shall be determined by uncontrolled, random factors. Therefore, the application of a probabilistic approach is justified.

A correct description of a modal split is very important, especially when analysing individual and public transport variant investments. Estimating passengers transferred between systems is required i.a. in projects applying for financing from EU programs and is evaluated by assessing bodies (e.g. CUPT and Jaspers). The growing demand for the use of traffic models enforces the need to develop and strengthen the methods for their construction, improving reliability and functionality.

The issue of choosing means of transport was addressed in the INMOP research project ("The principles of predicting road traffic, taking into account other means of transport"), implemented during the period of 1 February 2016 - 30 April 2019, commissioned by the National Centre For Research and Development (NCBiR) and the General Directorate for National Roads and Motorways (GDDKiA)1. The project deals with a hierarchical approach to travel modelling and predicting, i.e., concerns the methods of modelling traffic at national, regional and local levels. INMOP 3 is attempting to solve a number of detailed problems, also associated with modelling travel in urban areas, where the issue of intermodality of a model and modal split is particularly important [1], [2].

2. ANALYSIS METHOD - LOGISTIC REGRESSION

In a general approach, travels are divided into two categories: via means of individual and mass transport. Therefore, a response (dependant variable) in a model has a two-point distribution. Whereas, independent variables can adopt both, the continuous (travel time), discrete (number of changes), and even binary (whether a travel is to a paid parking zone) values. A tool, which enables describing the relation between such various independent variables and a dependent variable with two-point distribution is logistic regression. This analysis method found wide application in medicine,

---

1 The INMOP 3 project is implemented by a consortium of technical universities: the Institute of Roads and Bridges at the Warsaw University of Technology (leader) and the Cracow University of Technology (partner)

finances, engineering issues, etc. Everywhere, where the answer to a question is yes or no. Logistic regression requires the establishment of a so-called linear function linking a predictor Eq. (2.3) with probability. A straight line cannot be used, as in the case of linear regression, for once, because the model would predict, in certain ranges, probabilities smaller than zero or higher than one. Linking functions in the logit Eq. (2.2) or probit (normal distribution quantile function) are usually used for this purpose. Whereas, modelling decision probability with a logit enables simpler result interpretation. Due to the above, a decision-making process regarding the selection of means of transport may be described with the use of a logistic regression model (*logit model*). In general, a regression function form is following [2][2]:

$$(2.1) \qquad Y \sim B(1, p)$$

$$(2.2) \qquad p = E(Y|X) = \frac{e^{(\beta X)}}{1 + e^{(\beta X)}}$$

where:

$B(1, p)$ is a binomial distribution with success probability p, and βX is a linear combination of independent variables and parameters β_i.

Eq. (2.2) assumes adopting a linking function in the form of a logit function. The aforementioned linear combination (linear predictor), which is used in the logit function, may be presented as follows:

$$(2.3) \qquad V = \sum_{x_i \in X} \beta_i x_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

The hence presented relationship between probability and independent variables enables easy interpretation of the results. A positive parameter for a given independent variable means increasing success probability and vice versa. Since the analysed issue in a natural way provides variables strictly associated with defined means of transport (individual and mass transport), breaking the logit function into a difference between two functions, defined as means of transport utility (V_{mt} – mass transport utility function, V_{it} – individual transport utility function) is beneficial.

$$(2.4) \qquad V = V_{mt} - V_{it}$$

---

[2] The denotations were adjusted in terms of consistency with the rest of the text.

The adopted notation enables transforming the expression to probability, so that with the use of the above function, it is easy to express individual transport choice probability, as well as mass transport choice probability. The probability that a transport system user selects mass transport for a given route p_{mt}, estimated with the logistic regression model, shall be in the form:

$$(2.5) \qquad p_{mt} = \frac{e^{V_{mt}-V_{it}}}{1+e^{V_{mt}-V_{it}}}$$

after the transformation, by multiplying the numerator and denominator by $e^{V_{ti}}$ we get:

$$(2.6) \qquad p_{mt} = \frac{e^{V_{mt}}}{e^{V_{it}}+e^{V_{mt}}}$$

Whereas, the probability that a transport system user selects individual transport for a given route p_{it}, will be in the form:

$$(2.7) \qquad p_{it} = 1 - p_{mt} = 1 - \frac{e^{V_{mt}-V_{it}}}{1+e^{V_{mt}-V_{it}}} = \frac{1}{1+e^{V_{mt}-V_{it}}}$$

after the transformation, by multiplying the numerator and denominator by $e^{V_{it}}$ we get:

$$(2.8) \qquad p_{it} = \frac{e^{V_{it}}}{e^{V_{it}}+e^{V_{mt}}}$$

With such a notation, result interpretation is as follows. Positive parameters β in the function V_{mt} mean increasing probability p_{mt} with increasing i-value of the independent variable. Whereas positive parameters in the function V_{it} decrease probability p_{mt}, hence increasing probability p_{it}. The estimation of parameters β is performed by maximizing the likelihood function L:

$$(2.9) \qquad L(X_1, \dots X_n, \beta) = \prod_{i=1}^{n} p(X_i)^{Y_i}[1 - p(X_i)]^{1-Y_i}$$

Unlike linear regression, searching for logistic regression parameters is a non-linear issue and requires an iterative approach. Algorithms implemented in statistical software, for example the R package[3] are used for this purpose.

---

[3] http://finzi.psych.upenn.edu/R/library/stats/html/

Good model matching is important in logistic regression, which is why it should contain only significant independent variables. The parameter significance was tested with the use of the Wald test (W) [3], which is obtained through dividing the estimator of a given parameter by its standard estimation error (id. $SE(\hat{\beta})$):

$$(2.10) \qquad\qquad W = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

A hypothesis with parameter β significance was tested, assuming that it was significantly different from zero. Assuming the null hypothesis $\beta = 0$, W has normal distribution $N(0,1)$. A parameter was deemed significant, if at a 95% confidence interval, 0 will not be in the critical set. The assumption regarding variable independence is not necessary from a formal point of view, but the existence of strong collinearity may lead to numerical problems. The adopted collinearity measure is the variance inflation factor:

$$(2.11) \qquad\qquad VIF_j = \frac{1}{1-R_j^2}$$

where:

VIF_j– variance inflation factor for an independent variable predictor j, j– dependent variable (by other variables, while in a logistic regression model - predictor), R_j^2– a coefficient of multiple correlation between a given variable and other variables in a model.

As seen in equation 11, a variance inflation factor may adopt values from 1 to infinity. The better a given predictor is explained by other variables, the higher the value it adopts. From the point of view of the model, it is favourable for independent variables to be characterized by low values VIF (no more than 4-5). The model was assessed in terms of fitting, as well as through verifying the effectiveness in practical applications (the second assessment is not subject of this analysis). The goodness of fit was measured using $pseudoR^2$ suggested by McFadden [4]:

$$(2.12) \qquad\qquad pseudoR^2 = 1 - \frac{\ln(L_c)}{\ln(L_0)}$$

where:

L_c–likelihood of the full model, L_0– likelihood of the intercept model.

Similar R^2 in linear regression $pseudoR^2$ adopts values from 0 to 1, and the higher the value, the better the fitting of the model. However, it should not be interpreted in the same manner. Usually $pseudoR^2$ reaches lower values and a model is evaluated as well fitted with values in the 0.2-0.4 range.

3. CASE STUDY – LOGIT MODEL PARAMETER ESTIMATION

The analysis method described above, with the use of logistic regression was applied when developing a local travel model in the Warsaw metropolitan area. The process of estimating the logit model utilized a set of data obtained during commuting behaviour studies (surveys within the Warsaw Traffic Census, WBR 2015). The studies enabled developing a database regarding the travels Warsaw residents perform via various means of transport on typical weekdays4. In the database, each travel is a separate record and is deemed an independent observation (which meets the assumptions of a logit model). The travels were divided into 9 motivation groups (reasons for travelling) with similar behaviour: home - work (H-W), home - school (H-S), home - university (H-U), home - other (H-O), home - WOH5 (WOH), work – WOH (W-WOH), work – other (W-O), other – work (O-W), other.

17 variables were considered in the estimation: DIS travel time, TT0 – driving time over an unloaded network, TTC – driving time over a loaded network, x5 – travel to a paid parking zone, x6 – option to use a car (variable depending the district, tabl. 1), JRT – mass transport (MT) travel time, RTI – mass transport (MT) travel time, excluding walking to stops, IVT – the time of the sole travel in MT, NTR – number of changes in a travel by MT, SFQ – shuttling time of MT, WKD – walking distance in MT, WKT – walking time in MT, ACT – starting stop walking time, EGT – end stop walking time, ACD – starting stop walking distance, EGD – end stop walking distance, TWT – change-over time. As per [5], the factors affecting the choice of a particular means of transport were divided into four categories: the characteristics of the transport system itself, individual features of a traveller, characteristics of the particular travel (e.g. distance) and land development.

Table 1. The possibility to use a household car, depending per district, based on WBR 2015

	The possibility to use a car in the household		
	Yes	No	Total

---

4 Surveys carried out in a group of 17 000 respondents (36 519 surveys) were analysed. The most strongly represented surveys in the motivation groups were: home-work 7865 surveys, home-other 3004 surveys, other-home 3211 surveys.
5 WOH – Large-Format Commercial Facility

		% of N in a row	% of N in a row	Size in the sample	Size in the population	% of N in a row
Total		65.4%	34.6%	17000	1686339	100.0%
Residential district	Bemowo	74.2%	25.8%	1280	120757	100.0%
	Białołęka	82.9%	17.1%	1144	103421	100.0%
	Bielany	60.7%	39.3%	1017	112508	100.0%
	Mokotów	67.1%	32.9%	2028	213472	100.0%
	Ochota	62.4%	37.6%	794	85871	100.0%
	Praga-Południe	62.9%	37.1%	1508	167059	100.0%
	Praga-Północ	40.4%	59.6%	614	77822	100.0%
	Rembertów	88.3%	11.7%	266	22400	100.0%
	Śródmieście	49.3%	50.7%	1272	117587	100.0%
	Targówek	58.5%	41.5%	902	104424	100.0%
	Ursus	72.8%	27.2%	693	56730	100.0%
	Ursynów	77.6%	22.4%	1620	143836	100.0%
	Wawer	81.2%	18.8%	818	69311	100.0%
	Wesoła	83.7%	16.3%	295	23468	100.0%
	Wilanów	90.6%	9.4%	372	27871	100.0%
	Włochy	70.0%	30.0%	497	37271	100.0%
	Wola	51.5%	48.5%	1350	136021	100.0%
	Żoliborz	54.4%	45.6%	530	66511	100.0%

The variables used for the needs of the conducted analyses, cover the factors associated with features of individual travels to a large degree. Therefore, it should be assumed that it is possible to realistically represent the probability of selecting a particular means of transport. Factors associated with land development and a transport system would be of greater significance when comparing two different areas, with different transport systems. Whereas within Warsaw, their impact should not be significant. However, one should not expect full effectiveness of the obtained model, since many significant factors associated with individual travel features shall not be taken into account.

Given the adopted assumptions, the best model form was sought after. Subsequent proposals of the model involved testing the statistical significance of independent variables and whether there was any collinearity between them. In the course of the analyses, the distance independent variable parameter was rejected, due to the obvious strong correlation with other variables (e.g. individual and mass transport travel times). It was decided to reject the distance in favour of travel times, because distance is neutral in terms of the type of transport and qualifying it to any of the utility functions could be misleading. Moreover, this lead to using a larger number of independent variables, with their interpretation allowing to draw valuable conclusions (for example, walking times to bus stops).

The analyses were supported with experience in transport modelling, which improved the search for relevant utility function forms and enabled the development of a more intuitive and interpretation-friendly model. The results for the final search phase, in which the models were practically

determined and had been evaluated, are shown below. The adopted independent variable denotations are shown in table 2.

Table 2. The denotations of independent variables and parameters adopted in the model (key)

Independent variable denotation	Parameter denotation B	Independent variable description	Utility function
-	β_0	absolute term	V_{it}
TTC	β_3	individual transport travel time over a loaded network	
x5	β_5	is the travel to a parking zone	
x6	β_6	the option to use a car	
IVT	β_8	the time of sole mass transport (MT) travel	V_{mt}
NRT	β_9	the number of changes in mass transport - MT	
WKD	β_{11}	walking distance	
ACT	β_{13}	walking time to starting stop	
EGT	β_{14}	walking time to end stop	

Discrete numerical denotations are a consequence of rejecting some variables at an earlier stage of the analysis.

4. RESULT ANALYSIS AND MODEL EVALUATION

Results for selected motivation groups are shown in tabular form (table 3). Column β shows estimated values of model parameters. The p-value column means a test statistics value regarding the significance of a given model parameter, according to the Wald test (formula 10). It should be interpreted as the level of significance, at which a parameter can be deemed statistically significant. A 5% level of significance was adopted; therefore, a lower p-value enables considering a parameter as significant.

The last column shows the variance inflation factor (formula 11). Values closest to unity are the objective. Exceptionally, one pair of independent variables, exhibiting a rather significant collinearity, was allowed in the model. For obvious reasons, the travel time in individual transport is correlated with the mass transport travel time (the bigger the distance, the both times are longer). The independence of variables is not, however, a formal requirement but rather a practice, which allows to avoid numerical errors when fitting a model, and misinterpretations. Since the relationship only affects the results of these two variables, there is no fear of misinterpretation in this regard. Whereas thanks to leaving both variables in the model, both the mass, as well as individual, transport utility functions contain the travel time of relevant means of transport. The last row shows $pseudoR^2$ as per formula 12. Values higher than 0.2 were considered as very good model fitting.

Despite the fact that there are two variables in the home-school motivation group, collinearity was not checked. In this particular case, in order to avoid numerical errors, one x=IVT-TTC variable was created and only the IVT parameter was sought after. After calculating the parameter, the variable was divided into IVT and TTC, and then qualified to various utility functions. Due to the adopted convention (formula 4), the IVT parameter changed the sign, hence, the signs of IVT and TTC parameters are the same.

Table 3. Results of fitting a model

Independent variable denotation	For the home-work motivation group			For the home-other motivation group			For the home-school motivation group		
	Parameter value β	p-value	VIF	Parameter value β	p-value	VIF	Fβ	p-value	VIF
-	-2.3818	<2e-16	-	-3.5141	<2e-16	-	-0.768734	<2e-16	not applicable*
TTC	-0.0433	<2e-16	4.3370	-0.0332	<2e-16	5.1451	-0.045924	<2e-16	
x5	-0.4289	<2e-16	1.2708	-0.5231	<2e-16	1.2479	-	-	
x6	2.3991	<2e-16	1.1715	3.2221	<2e-16	1.1326	-	-	
IVT	-0.0331	<2e-16	4.5539	-0.0234	<2e-16	5.1248	-0.045924	<2e-16	
NRT	-0.3392	<2e-16	1.6583	-0.4689	<2e-16	1.9358	-	-	
WKD	-0.3672	<2e-16	1.1206	-0.8982	<2e-16	1.1638	-	-	
ACT	-0.0353	<2e-16	1.1610	-0.0729	<2e-16	1.1586	-	-	
EGT	-0.0333	0.0004	1.1701	-0.0411	<2e-16	1.1177	-	-	
model rating				model rating			model rating		
$pseudoR^2$	0.285			0.444			0.143		

The results obtained can be illustrated by charts. Below one can see example relationship between the probability of selecting mass transport estimated by a model in the home-work motivation group and selected independent variables. Grey points represent the probability determined for individual travels described with an individual set of parameters. Whereas the blue line represents the relationship between the probability of selecting mass transport and one selected variable, under the assumption of mean values of other variables.
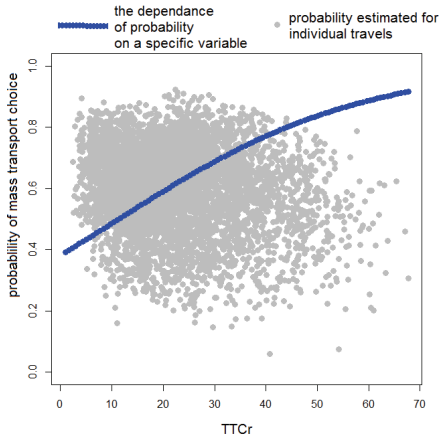
Fig.1 The impact of individual travel time in a morning peak loaded network on the probability of selecting mass transport in the case of the home-work motivation group
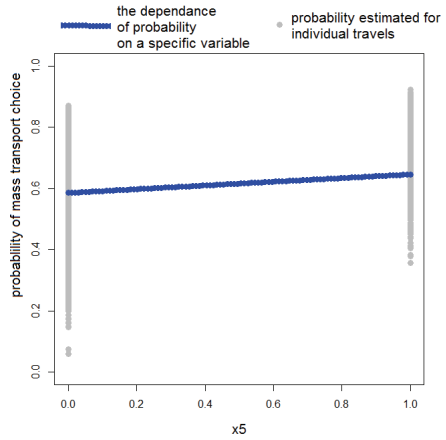

Fig. 2 The impact of travel destination parking zone on the probability of selecting mass transport in the case of the home-work motivation group
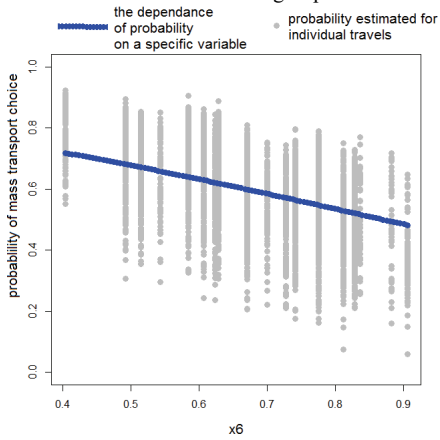

Fig. 3 The impact of the possibility to use a car (in a household) on the probability of selecting mass transport in the case of the home-work motivation group
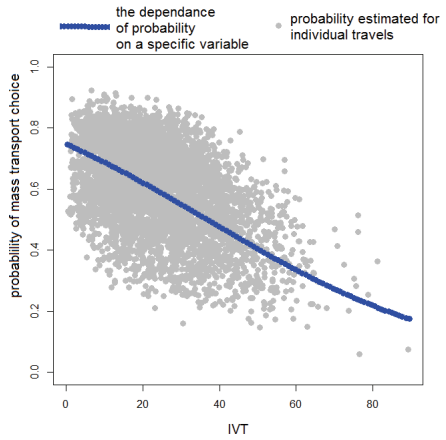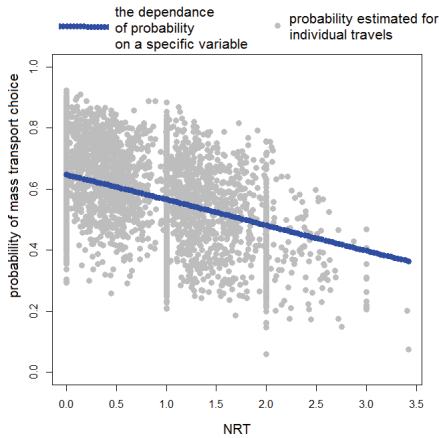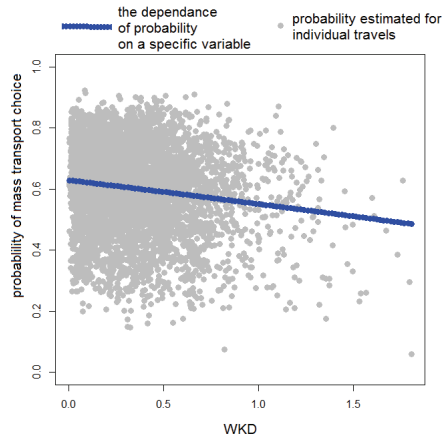

Fig.4 The impact of the mass transport travel time on the probability of selecting mass transport in the case of the home-work motivation group

Fig. 5 The impact of the number of changes on the probability of selecting mass transport in the case of the home-work motivation group

Fig.6 The impact of distance of movement on foot on the probability of selecting mass transport in the case of the home-work motivation group

As shown in the figures above, the results were consistent with intuition. Extending the travel time in individual transport or shortening in mass transit vehicles, increase the probability of selecting mass transport for a given travel. Whereas extending the mass transit travel time or the time of reaching a stop/station, decrease this probability.

5. CONCLUSIONS

The analysis with the use of a logistic regression model was conducted for travels in the Warsaw transport system, recorded in the form of a database, obtained based on a survey among the citizens. The objective was to test a logistic regression model for the determination of the impact of individual variables describing individual and mass transport travels on the probability of selecting specific means of transport. Travels divided into 9 motivation groups were tested, for which ultimately 8 models were developed (with the exception of the home-university group), out of which 7 were deemed very well fitted (obtained $pseudoR^2$ was well above 0.2). Whereas in the case of the home-school motivation group, the model used only 2 independent variables (due to numerical issues), resulting in $pseudoR^2$ at 0.143.

Based on the conducted analyses, it was concluded that a linear regression model enables the observation of certain relationships, however, it does not explain their causality. Nonetheless, it may be used for predicting the behaviour of transport system users and to indicate the direction of in-depth theoretical analyses, which could explain the observed relationships in a cause-effect manner. An

example of such proceedings may be a hypothesis based on the analysis of coefficients by ACT and EGT variables, describing the time of reaching the starting and final stops, respectively. Both coefficients have a negative sign, which means that increasing the time has a discouraging effect in terms of using mass transport. It is rather obvious, however, it should be noted that in ACT variable coefficient is, in the case of certain motivations, ca. 4 times bigger, which means that the time of reaching the starting stop has a greater impact. This can explain the observed behaviour involving transport system users during morning peak on routes from the city centre, often decide to travel by mass transport, even though, the conditions for individual transport are favourable (unloaded network). The reason may be the proximity of the initial stops.

REFERENCES

1. Brzeziński A., Dybicz T.. Szymański Ł, Włodarek P. Zastosowanie Krajowego Modelu Ruchu w analizie przebiegu dróg szybkiego ruchu przez Warszawę. Problemy komunikacyjne miast warunkach zatłoczenia motoryzacyjnego. Annały inżynierii ruchu i badań transportowych, 2017, t.1 (XI)
2. Brzeziński A., Dybicz T., Szymański Ł. Wykorzystanie KMR do modelowania ruchu w miastach na przykładzie Warszawy. Konferencja Naukowo-Techniczna Miasto i Transport 2018, Politechnika Warszawska, kwiecień 2018r
3. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2000). Applied logistic regression John Wiley & Sons.
4. Dobek A., Moliński K., Skotarczak E. Porównanie mocy testów Rao's score, Walda i ilorazu największej wiarogodności dla tablicy kontyngencji wymiaru (2xc). Biometrical Letters, 2015.
5. Domencich, Thomas A., and Daniel McFadden. Urban travel demand-a behavioral analysis. No. Monograph. 1975.
6. Davis, B. C. "Factors affecting choice of travel mode." Australian Road Research Board Conference Proc. Vol. 7. No. 2. 1975.

LIST OF FIGURES AND TABLES:

Rys. 4. Wpływ czasu jazdy transportem zbiorowym na prawdopodobieństwo wyboru transportu zbiorowego w przypadku grupy motywacji dom-praca

Fig. 5. The impact of the number of changes on the probability of selecting mass transport in the case of the home-Work motivation group

Rys. 5. Wpływ liczby przesiadek na prawdopodobieństwo wyboru transportu zbiorowego w przypadku grupy motywacji dom-praca

Fig. 6. The impact of distance of movement on foot on the probability of selecting mass transport in the case of the home-Work motivation group

Rys. 6. Wpływ odległości przejść pieszo na prawdopodobieństwo wyboru transportu zbiorowego w przypadku grupy motywacji Dom-Praca

Tab. 1. The possibility to use a household car, depending per district, based on WBR 2015

Tab. 1. Możliwość korzystania z samochodu znajdującego się w gospodarstwie domowym w zależności od dzielnicy na podstawie WBR 2015

Tab. 2. The denotations of independent variables and parameters adopted in the model (key)

Tab. 2. Przyjęte w modelu oznaczenia zmiennych objaśniających i parametrów (legenda)

Tab. 3. Results of fitting a model

Tab. 3. Wyniki dopasowania modelu

ZASTOSOWANIE MODELU REGRESJI LOGISTYCZNEJ DO PRZEWIDYWANIA PREFERENCJI UŻYTKOWNIKÓW SYSTEMU TRANSPORTOWEGO

Słowa kluczowe: system transportowy, modelowanie podróży, modal split, regresja logistyczna, model logitowy

STRESZCZENIE

Umiejętność opisania zachowań użytkowników systemu transportowego w zakresie podejmowanych decyzji dotyczących wyboru środka transportowego stanowi podstawę tworzenia modeli podróży, służących analizom i prognozowaniu ruchu. Wiąże się to z poszukiwaniem zależności pomiędzy czynnikami opisującymi podróże, a podejmowanymi decyzjami o wyborze środków transportu. Decyzje o tym, jaki rodzaj transportu wybrać są zdeterminowane różnymi czynnikami dotyczącymi samej podróży, ale również indywidualnymi preferencjami użytkowników systemu transportowego. Tworząc modele podziału zadań przewozowych nie sposób jest, ze względu na dostępność danych, uwzględnić wszystkie możliwe czynniki, zatem trzeba uznać, że o mechanizmie wyboru będą decydować czynniki niekontrolowane, losowe. Dlatego też uzasadnione jest stosowanie podejścia probabilistycznego.

Prawidłowe opisanie procesu podziału zadań przewozowych jest bardzo ważne zwłaszcza przy analizach wariantowych inwestycji transportu indywidualnego i publicznego. Oszacowanie pasażerów przeniesionych pomiędzy systemami jest wymagane m.in. w projektach aplikujących o dofinansowanie z programów Unii Europejskiej i jest oceniane prze jednostki opiniujące (np. CUPT i Jaspers). Rosnące zapotrzebowanie na stosowanie modeli ruchu wymusza konieczność rozwijania i wzmacniania metod ich budowy, poprawiania wiarygodności i funkcjonalności.

Tematykę wyboru środka transportu podjęto w projekcie badawczym INMOP 3 („Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu") realizowanym w okresie 1 luty 2016 – 30 kwietnia 2019 r. na zamówienie Narodowego Centrum Badań i Rozwoju (NCBiR) oraz Generalnej Dyrekcji Dróg Krajowych i Autostrad (GDDKiA)[6]. Projekt zajmuje się hierarchicznym podejściem do modelowania i prognozowania podróży, tj. dotyczy metod modelowania ruchu na poziomie krajowym, regionalnym i lokalnym. INMOP 3 stawia sobie za cel podjęcie próby rozwiązania szeregu problemów szczegółowych, także związanych z modelowaniem podróży w obszarach zurbanizowanych, gdzie zagadnienie intermodalności modelu i podział zadań przewozowych jest szczególnie ważne.

W ramach projektu zastosowano model regresji logistycznej do przeprowadzenia analiz podróży odbywanych w systemie transportowym Warszawy zapisanych w formie bazy danych uzyskanej na podstawie badania ankietowego mieszkańców. Celem było przetestowanie modelu regresji logistycznej do ustalenia wpływu poszczególnych zmiennych opisujących podróże transportem indywidualnym oraz zbiorowym, na prawdopodobieństwo wyboru danego środka transportu. Podróże przebadania w podziale na 9 grup motywacji, dla których ostatecznie opracowano 8 modeli (z wyjątkiem grupy dom-uczelnia), z których 7 uznano za bardzo dobrze dopasowane (*pseudoR²* znacznie powyżej 0,2). Natomiast w przypadku grupy motywacji dom-szkoła w modelu wykorzystano tylko 2 zmienne objaśniające (ze względu na problemy numeryczne), w związku z czym *pseudoR²* wyniosło 0,143.

Na podstawie przeprowadzonych analiz stwierdzono, że model regresji liniowej pozwala na zaobserwowanie pewnych zależności, jednak nie wyjaśnia przyczynowości. Niemniej jednak może zostać wykorzystany do predykcji zachowań użytkowników systemu transportowego oraz wskazać kierunki pogłębionych analiz teoretycznych, które mogą wyjaśnić obserwowane zależności w sposób przyczynowo-skutkowy. Przykładem takiego postępowania może być hipoteza postawiona na podstawie analizy współczynników przy zmiennych ACT i EGT opisujących odpowiednio czas dojścia do przystanku początkowego i czas dojścia do przystanku końcowego. Obydwa współczynniki mają znak ujemny, co znaczy, że zwiększenie czasu wpływa zniechęcająco na korzystanie z transportu zbiorowego. Jest to dość oczywiste, jednak warto zauważyć, że współczynnik przy zmiennej ACT jest w przypadku niektórych motywacji ok. 4 razy większy, to znaczy, że czas dojścia do przystanku początkowego ma większy wpływ. Może to tłumaczyć obserwowane zachowania

---

polegające na tym, że w szczycie porannym w relacjach z centrum miasta, użytkownicy systemu transportowego często decydują się na podróż transportem zbiorowym pomimo tego, że warunki dla transportu indywidualnego są korzystne (nieobciążona sieć). Przyczyną takiego stanu rzeczy może być właśnie bliskość przystanków początkowych.

Przeprowadzone analizy potwierdziły, że model regresji liniowej może być wykorzystany do predykcji zachowań użytkowników systemu transportowego. Uzyskane wyniki okazały się zgodne z intuicją - wydłużenie czasu podróży w transporcie indywidualnym, lub skrócenie czasu jazdy w pojazdach transportu zbiorowego zwiększa prawdopodobieństwo wyboru transportu zbiorowego w danej podróży. Natomiast wydłużenie czasu przejazdu transportem zbiorowym, czy dojścia do przystanku, obniżają to prawdopodobieństwo.