Modeling Household Weekend Activity Durations in Calgary

By

Ming Zhong

Research Associate, Ph.D., P.Eng.
Dept. of Civil Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta
Canada T2N 1N4

Phone: (403) 220-4820
Fax: (403) 282-7026
Email: ming.zhong@ucalgary.ca

John Douglas Hunt

Professor, Ph.D., P.Eng.
Dept. of Civil Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta
Canada T2N 1N4

Phone: (403) 220-8783
Fax: (403) 282-7026
Email: jdhunt@ucalgary.ca

Paper prepared for presentation

at the **Emerging Best Practices in Urban Transportation Planning (B)** Session

of the 2005 Annual Conference of the
Transportation Association of Canada
Calgary, Alberta

September 19th – 21st, 2005

Table of Contents

Modeling Household Weekend Activity Durations in Calgary

By Ming Zhong and John Douglas Hunt

Abstract

A large-scale survey for household weekend activity and related travel was completed recently in the City of Calgary. The data include detailed information of traveler and activity, such as personal type (e.g., adult worker or senior), employment status (fulltime or part-time), annual income, gender, activity type (e.g., shopping or sociality), activity duration, and starting & ending time of each activity. A micro-simulation based choice behavior model has been used in the previous city planning tasks. The model is capable of simulating complete travel behavior of individuals by considering travel purpose, travel mode, itinerary, activity durations, and even group influences. Previously, the simulation was done using a Monte Carlo process with sampling distributions based on weighted sample of observed durations. Simulations based on such "static" distributions, however, can not be used to analyze the influences of various policies (e.g., changes in transit fare) and travel conditions (congestion or easier accessibility) to household activities in a dynamic environment. This study is an initiative for modeling the relationship between activity durations and various influencing factors (e.g., personal type, employment status, and income level, etc.). Especially, hazard and survival functions are specified for each type of activity and individual personal type. The results show high degree of fit. It is believed that these models would be useful for travel-related policy analysis in the future modeling framework. (Total 222 words)

Introduction

Travel demand modeling has been intensively focused on forecasting individual activities and related travel patterns on weekdays to damp traffic congestions occurring mostly at morning or evening peak commuting hours. Literature review indicates that many studies have been carried out for investigating weekday travels, but less effort has been placed on weekends until recently [1-4]. However, as travel demand keeps increasing and infrastructure construction is getting constrained, congestions take place in many big cities over weekends. A previous study [5] shows that weekend household travels are comparable to those of weekdays, in terms of the number of trips made and trip lengths. It is clear that people have distinct travel behaviors on weekends because most do not have commitment to work and thus be able to participate in various other activities (e.g., maintenance shopping, sociality, and entertainment etc.). Hence, based on these reasons, it seems that special attention should be given to weekend travel as well.

A large scale of survey for weekend household activities was completed recently in the City of Calgary, Canada. The purpose of the survey is to collect enough data for both short-term operational analysis and long-term planning. The data obtained from the survey provides excellent opportunities to analyze weekend activity and related travel behaviors, and is expected to provide insights for future policy makings.

In this study, a general picture of weekend household activities in the City of Calgary is first presented. Then, a variety of hazard/duration models are analyzed and specified for various activities and demographic groups. The analyzed activities include travel related activity (e.g., dropping off or picking up a person), working, schooling, shopping, sociality, eating, entertainment, exercise, religious activities, and out-of-town travel. Demographic groups are studied for each type of activity, including AO (Adult non-worker), AWNC (Adult worker who needs car), AWNNC (Adult worker who doesn't need car), KEJS (Elementary or junior high school students), PSS (Post-secondary students), Sen (Seniors), SHS (Senior high school students), and YO (Young other). The analyses are applied to individual activity types and demographic groups to account for the heterogeneity in the data. For each activity and demographic group, hazard/duration models based on different distributions (e.g., lognormal or Weibull) are explored and best-fit models are specified.

The rest of this paper is organized as follows. First, a literature review for duration/hazard models is presented, then the study data are presented, followed by study models and results, and finally major findings and conclusions are given at the end of this paper.

Review of Duration/Hazard Models

The statistical analysis of what are called as lifetime, survival time, or failure time data has long been an important topic in many areas, such as biomedical, engineering, and social sciences [6]. It is known as duration modeling or hazard modeling. For example, Bartholomew [7] used duration models to study the lifetime distribution of equipment. Prentice [8] compared the effects of two chemotherapy treatments in prolonging survival time of 40 advanced lung cancer patients. Nevertheless, duration models have been more and more widespread used in transportation area [9-16].

The key element of duration modeling is hazard functions, which indicate the way that the risk or probability of failure varies with age or time [6]. A hazard function *h(t)* that expresses the probability of the occurrence of an event during a certain time interval, say t to $t+\Delta$t, given that the event has not occurred before the beginning of the interval. The conditional probability of duration starting or ending plays an important role as the probability indicates that an event starts or terminates depends on the length of time or the duration has lasted.

Let T be a nonnegative random variable representing the lifetime of individuals in some population. Here only a continuous variable T is assumed (discrete T can be accommodated by considering the discretization as a result of segmentation of continuous time into several discrete intervals), as this is the case for most applications. Let the probability density function of T is *f(t)* and the cumulative distribution function (CDF) is *F(t)*. Then we have:

$$F(t) = P(T < t) \quad [1]$$

Where T is a random time variable and t is some specific time. In the case of household activities, the cumulative distribution function is defined to indicate the probability of an activity would end before some specified time, t. Then the probability density function (PDF) of *f(t)* can be obtained:

$$f(t) = \frac{dF(t)}{dt} \quad [2]$$

Which provides unconditional distribution of duration T. The survival function (SF), *S(t)*, is then can be defined as:

$$S(t) = P(T \geq t) = 1 - F(t) \quad [3]$$

The above survival function represents the probability that the duration in a state T will be greater than or equal to some specific time t. The hazard function can then be expressed as a function of the probability density function *f(t)*, the cumulative distribution function *F(t)* and survival function *S(t)*, as shown in the following equation:

$$h(\mathrm{t}) = \lim_{\Delta t \to 0} \frac{p(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(\mathrm{t})}{1 - F(\mathrm{t})} = \frac{f(t)}{S(t)} \quad [4]$$

The hazard, h(t), gives the rate at which events (e.g., ending an activity) are occurring at time t, given that the event has not occurred up to time t.

Parametric hazard functions

Various parametric families of models are used in the analysis of lifetime data. Among univariate models, a few distributions demonstrate their usefulness and have been applied in a wide range of situations [6]. These distributions are Exponential, Weibull, Log-logistic, Log-normal, extreme value (Gumbel) and Gamma. The four most common distributions are summarized below:

(1) Exponential:

The exponential distribution has the following PDF and SF:

$$f(t) = \lambda e^{-\lambda t} \qquad t > 0,$$
$$S(t) = e^{-\lambda t}$$

Then the hazard function is obtained by:

$$h(t) = \frac{f(t)}{S(t)} = \lambda \qquad [5]$$

So the exponential distribution is characterized by a constant hazard function. Because the assumption of a constant hazard function is very restrictive, the model's applicability is quite limited in reality.

(2) Weibull:

The Weibull distribution may be the most widely used duration models. It has a PDF and SF as follows:

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} e^{-(\lambda t)^{\beta}} \qquad t > 0,$$
$$S(t) = e^{-(\lambda t)^{\beta}}$$

$$h(t) = \lambda\beta(\lambda t)^{\beta-1} \qquad [6]$$

Where λ and β are positive rate and scale parameters. Note that the hazard function is monotone increasing if $\beta > 1$, decreasing if $\beta < 1$, and constant if $\beta = 1$. The model is fairly flexible and has simple expressions for the PDF, SF and hazard functions lead to its popularity.

(3) Log-Normal:

The log-normal distribution has been used as a model in diverse applications in engineering, medicine and other areas. The lifetime T is said to be log-normally distributed if $Y = log\ T$ is normally distributed, say with mean μ, variance σ^2, and PDF [6]:

$$f(t) = \frac{1}{(2\pi)^{1/2}\sigma t} \exp[-\frac{1}{2}(\frac{\log t - u}{\sigma})^2] \quad t > 0$$

The survivor and hazard functions for the log-normal distribution involve the standard normal distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} e^{-\frac{u^2}{2}} du$$

Then the log-normal survival function can be written as:

$$S(t) = 1 - \Phi(\frac{\log t - u}{\sigma})$$

The hazard function is then obtained:

$$h(t) = \frac{\frac{1}{(2\pi)^{1/2}\sigma t}\exp[0.5(\frac{\log t - u}{\sigma})^2]}{1 - \int_{-\infty}^{t} \frac{1}{(2\pi)^{1/2}} \exp(-\frac{u^2}{2}) du(\frac{\log t - u}{\sigma})} \quad [7]$$

(4) Log-Logistic:

The log-logistic distribution has PDF of the form:

$$f(t) = \frac{(\beta / \alpha)(t / \alpha)^{\beta - 1}}{[1 + (t / \alpha)^{\beta}]^2}, \quad t > 0, \quad \alpha, \beta > 0$$

The survival function and hazard function are:

$$S(t) = [1 + (t / \alpha)^{\beta}]^{-1}$$

$$h(t) = \frac{(\beta / \alpha)(t / \alpha)^{\beta - 1}}{[1 + (t / \alpha)^{\beta}]} \quad [8]$$

When $\beta > 1$, the hazard first increases and then decreases. When $0 < \beta \leq 1$, the hazard decreases with duration.

In addition to these four popular models, extreme value distribution is used in some cases where lifetime variable *T* follows a Weibull distribution, but we want to analyze Y = log *T*. In this case, Y follows an extreme value distribution. Gamma distribution is not used as a lifetime model as much as the four models mentioned above, due to the complexity of its hazard function. However, it does fit a variety of data. It also arises in some situations involving sum of a series of independent and identically distributed exponential random variables [6]. In such a case, however, the most challenging issue is to address that considered exponential random variables do have a same distribution parameter λ.

Nonparametric hazard functions

The above hazard functions are fully parametric and could be applied to questions with sound theoretical foundation. However, in some cases, if little or no knowledge of the functional form of the hazard is available, one might use a non-parametric approach in which there is no assumption involved concerning the underlying distribution of the baseline hazard. However, analyst can not use any explanatory variable in the model if a nonparametric approach is chosen. The most popular nonparametric method is the Kaplan-Meier Estimator. For such an approach, the duration scale is split into small discrete periods and by assuming a constant hazard within each period, one can then estimate the continuous-time step function hazard shape. The Kaplan-Meier Estimator is particularly good in situations in which there are a small number of groups and we want to know that if they share same survival distributions. There are several methods (such as Log-Rank and the Wilcoxon method) for such tests. Also, the nonparametric shape obtained from the Kaplan-Meier Estimator can be used to empirically test the assumed parametric baseline shapes [6].

Semi-parametric hazard functions

When there is no clear choice concerning hazard functions, another safer approach is to use semi-parametric hazard models. In these models, there is no distributional assumption for the baseline hazard and leave it arbitrarily, but assumption is made concerning the functional form specifying how the external covariates interact with the baseline hazard in the model. Two parametric forms are usually employed to accommodate the effect of external covariates on the hazard at any time: the proportional hazard form and the accelerated lifetime form [6, 9]. Here only proportional hazard form will be introduced due to its popularity.

The proportional hazard form specifies the effect of external covariates to be multiplicative on an underlying hazard function [9]. The model defines the hazard rate at time *t*,

$$h(t,x) = h_0(t)\exp(\beta x) \qquad [9]$$

where $h_0(t)$ is the baseline hazard rate assuming that all covariates in x have a value of 0, and β is a corresponding vector of coefficients to be estimated. In the proportional hazard model, the effect of external covariates is to shift the entire hazard function based on characteristics of individual; the basic hazard function is assumed same for all members of a group.

The semi-parametric models partially relax the assumption of parametric relationship between various factors and resulting hazard rate. However, it should be noted that the multiplicative form of Equation [9] is a strong assumption and require careful checking in applications [6]. Moreover, it can be shown that if data censoring exists and the underlying survival distribution is known, the semi-parametric proportional hazard models do not produce efficient coefficient estimates [16].

Study Data and Primary Analyses

Dataset from recently completed household weekend travel survey in Calgary was used in this study. The data include detailed information of traveler and activities, such as personal type (e.g., adult worker or senior), employment status (fulltime or part-time), annual income level (1-10 levels), gender, age, household size (persons in household), driving capability (holding license or not), activity type, activity durations (in minutes), and starting & ending time of each activity. Totally there are 12,916 observations used in this study. By excluding those with missing durations, 12,882 observations are used in the following analysis.

The primary objective of this study is to examine and analyze the relationship, if any, between durations of various activities and people's socio-demographic characteristics collected in the data. In order to have a preliminary understanding to the data, various box-plots are used to discover possible relationship between activity durations and individual socio-demographics. Figure 1 shows two of them. Figure 1(a) presents the distribution of activity durations based on activity and personal types. At first, it can be found that there are large differences between mean levels of different activity durations. For example, the travel related activities (drop-by etc.) are usually about 1-2 minutes, whereas working typically last for over 200 minutes. Secondly, for each activity type, significant variability exists among activity durations of different personal types (e.g., adult non-worker and senior). Based on these findings it seems that analyses at least should be applied to individual groups classified based on these two variables to well handle the heterogeneity in the data.

Figure 1(b) shows the distribution of activity durations based on activity type and annual income level. It is expected that people with higher income would usually have different consumption patterns than those with lower income, for example, perhaps more social activity and longer entertainment/leisure durations. However, the plot does not show any obvious trend within individual activity types, as shown in Figure 1(b). It seems that income level does not significantly influence activity durations. Another finding from the above figures is that there are a large number of outliers existing in the data and activity durations tend to skew to the right in most cases. The finding is quite conform to those of many other studies, and indicates that models based on normal distribution may not be appropriate for the data, and the other types of models are needed to solve the problem at hand.

Figure 2 shows percentage of different weekend activities. It is clear that the dominant weekend activity is shopping, which is about 35% of all weekend activities. The following important activities are eating, entertainment/leisure, and sociality. The percentage of these activities is 12.5%, 12.4%, and 11.4% respectively. About 10% of activities are related to the work and 8% are travel related activities. The remaining activities are usually about 5% or less of the total, as shown in Figure 2. The predominance of shopping and the other important weekend activities (e.g., eating) and related intensity in a short period in the weekend afternoons indicate that a high traffic demand is possible at a given time, and thus could present significant challenges to urban traffic management systems (UTMS). The above results indicate that distinct traffic operation and control strategy than those for weekdays may be required to avoid congestions. Such strategy could be something like "giving priority to major corridor to shopping centers".

Study Models and Results

The first question encountered in modeling activity durations would be which kind of models should be used? As mentioned in the previous section, there are three families of models available for duration modeling: full-parametric, nonparametric, and semi-parametric. In this study, nonparametric approach is not used in duration modeling because they don't incorporate any parameters and can not be used in the policy analysis. However, it is used in this study to judge whether two groups could share a common survival function. Regarding full-parametric and semi-parametric, the question is to address that if a firm conclusion can be made to the underlying data distribution and if a parametric model can provide a good fit. As mentioned before, if these two conditions can be met, parametric models are preferred because of not only their better predictions, but also their wider acceptance among practitioners.

Based on the primary analyses from the previous section, it is clear that investigation should be applied to individual groups of different personal and activity type. First, investigation is applied to individual activity types, and best-fit parametric models are identified. Table 1 shows the best-fit models for them. The following 11 distributions available in MINITAB are tested against different activities: Weibull, lognormal, exponential, loglogistic, 3-parameter Weibull, 3-parameter lognormal, 2-parameter exponential, 3-parameter loglogistic, smallest extreme value, normal, and logistic. These distributions are fitted into data and the fitness is evaluated with adjusted Anderson-Darling test statistics (AD values in the table) and correlation coefficients (COR in the table). The criterion is to select the distribution with the lowest AD value or the highest COR value. The best-fit models selected for the 10 activities are: lognormal or 3-parameter lognormal for travel related activity, working, schooling, shopping, eating, entertainment/leisure, and religious, civic etc. Weibull or 3-parameter Weibull are identified as the best-fit model for social and out-of-town activity. 3-parameter loglogistic is identified as the best for the exercise activity. The best-fit models identified above have the high COR values and indicate good fits, as shown in Table 1. The COR values for best-fit models are 0.99 or 1.00, except for travel related activity and out-of-town activity. The reason for inferior goodness-of-fit for travel related activity is that most people reported their durations in integer minutes rather than "real spell" (e.g., 1.2 minutes). This results in that data are not normally distributed, but cluster onto certain points (e.g., 1 or 2 minutes). The small number of observations lead to the deteriorated fit for out-of-town activity. Totally there are only 15 observations for such activities in the dataset. The obtained high goodness-of-fits emphasize that parametric, rather than semi-parametric, models should be used.

Figure 3 shows the empirical cumulative distribution function and fitted line based on 3-parameter lognormal distribution for the shopping activities. The similitude between these two lines again emphasizes the high goodness-of-fit achieved.

Because of the significance of shopping among all weekend activities, it is used as an example here to illustrate duration modeling of various activities in this study. Figure 4 shows the probability density function, probability plot, survival and hazard function for shopping activities based on 3-parameter lognormal distribution. The estimated parameters and calculated statistics are shown in the right of the figure. It can be seen from the probability plot in Figure 4 that the selected model fits the data very well, except for the part of lower left corner. Again, the imputed nature of reported values results in such distortion. The hazard function shows that about half of shopping activities have a duration of less than 25 minutes. The hazard rate increases dramatically when the duration approaches 25 minutes. After that, the hazard rate continues decreasing, which indicates that there is another group of consumers who tend to shop for a longer time. Based on the estimated parameters, a set of duration models can be set up and hazard/survival rate at any time can be readily calculated.

In order to investigate if different hazard/survival functions should be applied to subgroups based on personal types (e.g., AWNC subgroup), Kaplan-Meier method was used to check underlying distributions of the data. Figure 5 shows a nonparametric survival plot for social activities. The figure and calculated test statistics (Log-Rank and Wilcoxon, as shown in the figure) all indicate that the demographic subgroups are significant different from each other and individual hazard function should be specified. Table 2 shows AD and COR values of different models when tested on individual subgroups. It can be seen from the table that the best-fit models for the subgroups are quite consistent with that of the aggregate level when all subgroups are combined together. Out of 7 of 8 cases, 3-parameter Weibull distribution is identified as the best-fit models. There is only one exception that 3-parameter loglogistic distribution is identified as the best-fit model for YO group. Even in this case, the differences between these two distributions are quite small (the difference between AD values is 0.33 and that between COR values is only 0.01) and therefore, 3-parameter Weibull distribution still can be used to avoid complexity in the modeling process. Analyses also indicate that subgroups for the other activities (e.g., eating) share a common type of best-fit model (e.g., lognormal, but with different parameters) in most cases. Similarly, the hazard functions for activity durations of the other subgroup can be easily estimated based on the methods illustrated above.

Concluding Remarks

Studying household activities and corresponding travel patterns are important themes of activity-based transportation planning. Previous research has been focused on weekday commuting travel, but little has been placed on weekend travel. This paper studies the weekend household activities in the context of the City of Calgary, Canada. The attempts made in this study include identifying important influencing factors to various activities and specifying best-fit hazard/duration models for individual subgroups based on personal and activity type.

The analyses carried out in this study show that the following activities are dominant over weekends: shopping (35.2%), entertainment/leisure (12.4%), eating (12.5%), and sociality (11.4%). Among these, shopping is the most important activity and it is over one third of all weekend activities (Figure 2). The findings indicate that future research should pay more attention to these activities. Pattern analyses show that weekend activities mostly carry out in the afternoon. The findings indicate that there are different travel patterns on weekends and they deserve special attention. One of implications is that "specially designed" traffic operation and control strategy may have to be applied to accommodate such distinct traffic demand.

The applicability of parametric, nonparametric, and semi-parametric models are examined in this study. Nonparametric models don't incorporate any variable and therefore are not appropriate for policy analysis. However, they are useful for checking the underlying distributions and helpful for specifying appropriate parametric models. They are also used in this study to determine if subgroups could share a common survival function (Figure 5). The semi-parametric approach was not considered in the study because the competing parametric models show high goodness-of-fit. Study results clearly suggest that different parametric models should be specified for different types of activity (Table 1). The most frequently selected models are lognormal, followed by Weibull and loglogistic. This study also shows that the models selected at aggregate level (e.g., by activity type) are highly consistent with those selected at disaggregate level (e.g., subgroups based on personal type) (Table 2).

The “rounding” of reported durations existing in the data result in deteriorated fit and prediction power of the developed models. It is expected that the developed models could be more accurate if “real” durations would have been reported in the data. Such a problem could be solved by “imputing” these rounded observations into a normally distributed population. Future research is going to explore such an issue.

The hazard/duration models developed here are intended to be used in the future urban modeling framework. The aim is to replace previous “static” activity duration models. The models developed in this study explicitly consider many factors of household and individual members, but do not incorporate those by which various policy analyses can be made, such as transit fare and waiting time. Currently, this is limited by the data used. Future research should include these variables into the model when this information is available.

Acknowledgments

The authors are grateful towards Natural Science and Engineering Research Council (NSERC), Canada for their financial support, and the City of Calgary for the data used in this study.

References

1. Allison, M.L., S. Srinivasan, and C.R. Bhat. (2005). An Exploratory Analysis of Weekend Activity Patterns in the San Francisco Bay Area. Presented at *the 84th Transportation Research Board Annual Meeting*, Washington DC.
2. Bhat, C.R., and Lockwood, A. (2004). On Distinguishing Between Physically Active and Physically Passive Episodes and Between Travel and Activity Episodes: An Analysis of Weekend Recreational Participation in the San Francisco Bay Area. *Transportation Research Part A*, Vol. 38, No. 8, pp. 573-592.
3. Bhat, C.R., and Srinivasan, S. (2005). A Multidimensional Mixed Ordered-Response Model for Analyzing Weekend Activity Participation. *Transportation Research Part B*, Vol. 39, No. 3, pp. 255-278.
4. Sall, E.A., C.R. Bhat, and J. Reckinger. (2005). An Analysis of Weekend Work Activity Patterns in the San Francisco Bay Area. Presented at *the 84th Transportation Research Board Annual Meeting*, Washington D.C.
5. Federal Highway Administration (FHWA). (2004). 2001 National Household Transportation Survey. FHWA, US Department of Transportation (website: http://nhts.ornl.gov/2001/pub/STT.pdf, last accessed Feb. 18, 2005)
6. Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Inc., Hoboken, New Jersey.
7. Bartholomew, D.J. (1957). A Problem in Life Testing. *Journal of American statistics Association*, 52, 350-355.
8. Prentice, R.L. (1973). Exponential Survival with Censoring and Explanatory Variables. *Biometrika*, 60, pp. 279-288.
9. Bhat, C.R. (2000). Duration Modeling. *Handbook of Transport Modelling*, pp. 91-111, Edited by D. A. Hensher and K.J. Button, Elsevier Science.
10. Bhat, C. R., T. Frusti, H. Zhao, S. Schönfelder, and K.W. Axhausen. (2002). Intershopping Duration: An Analysis Using Multiweek Data. In the proceedings of the *81st Annual Meeting of the Transportation Research Board* (CD-ROM), TRB, Washington, D.C.
11. Doherty, S. T., M. Lee-Gosselin, K. Burns and J. Andrey. (2002). Household Activity Rescheduling in Response to Automobile Reduction Scenarios. *Transportation Research Record 1807*, pp. 174-182.
12. Ettema, D., A. Borgers and H. Timmermans. (1993). Simulation model of activity scheduling behavior. *Transportation Research Record 1413*, pp. 1-11.
13. Gärling, T., M. P. Kwan and R. G. Golledge. (1994). Computational-process modelling of household activity scheduling. *Transportation Research B* 25(5), pp. 355-364.
14. Hensher, D.A. and F.L. Mannering. (1994). Hazard-Based Duration Models and Their Application to Transport Analysis. *Transport Reviews*, 14, pp. 63-82.
15. Miller, E. J. and M. J. Roorda. (2003). Prototype Model of Household Activity/Travel Scheduling. *Transportation Research Record 1831*, pp. 114-121.
16. Mohammadian, A. and S.T. Doherty. (2004). A Hazard Model for the Duration of Time between Planning and Execution of an Activity. Presented at the *Conference on Progress in Activity-Based Analysis*, Vaeshartelt Castle, Maastricht, The Netherlands.

List of Figures and Tables

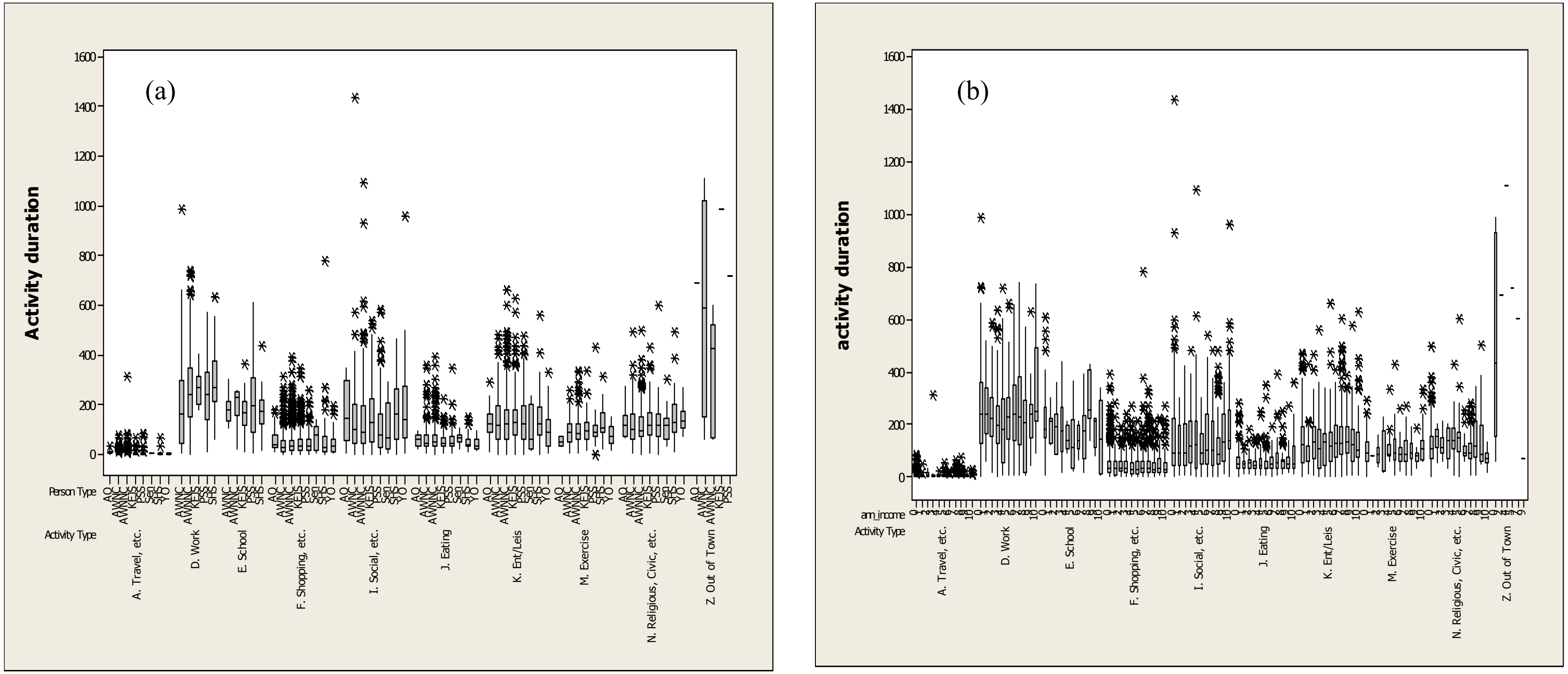

Figure 1 Box-plot of activity durations versus (a) personal and activity type (b) activity type and annual income level

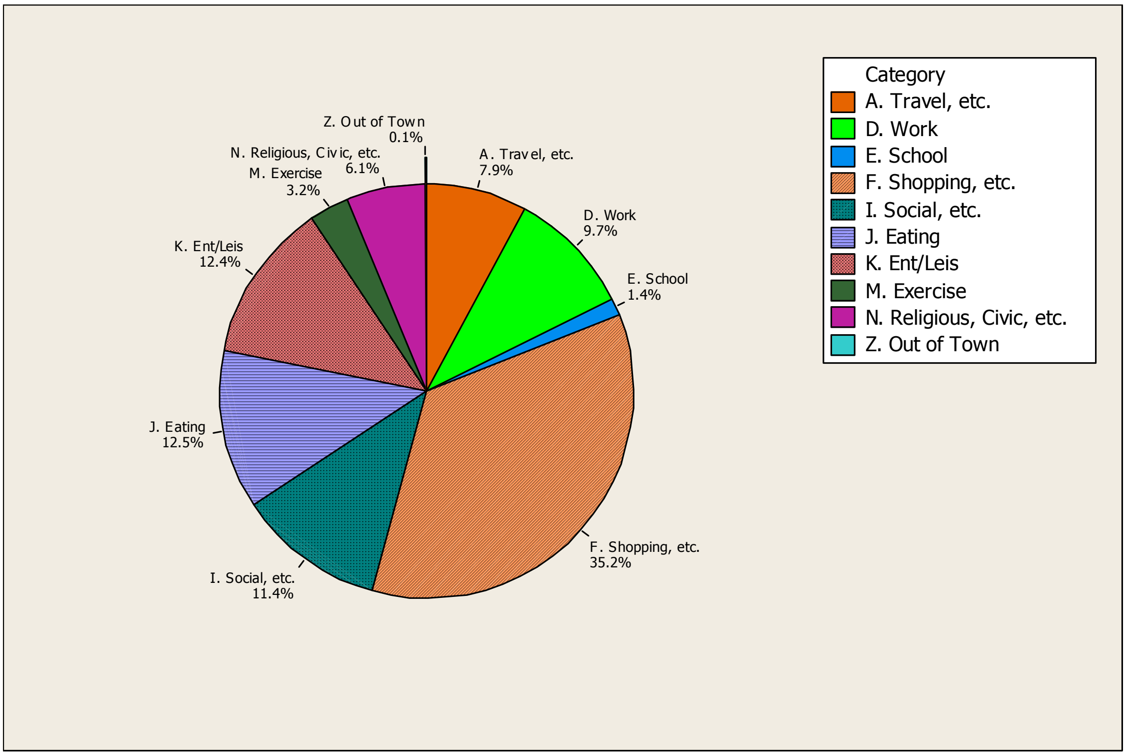

Figure 2 Percentage of different weekend activities

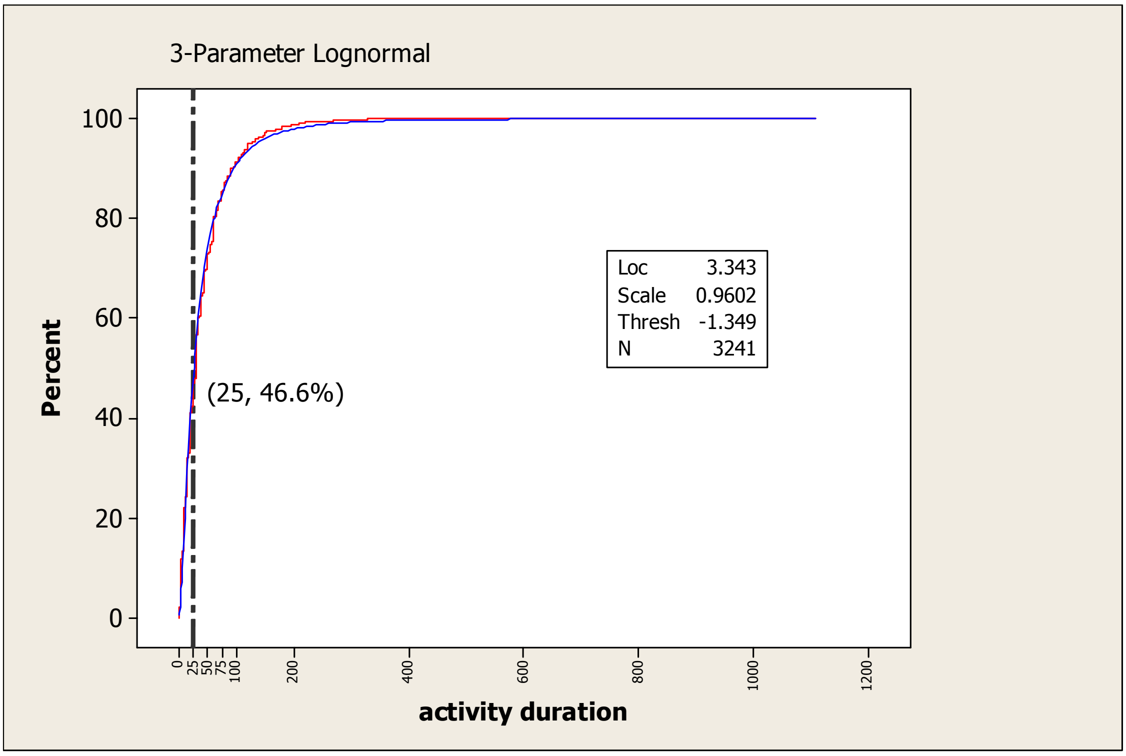

Figure 3 Empirical cumulative distribution function and fitted line for shopping activities

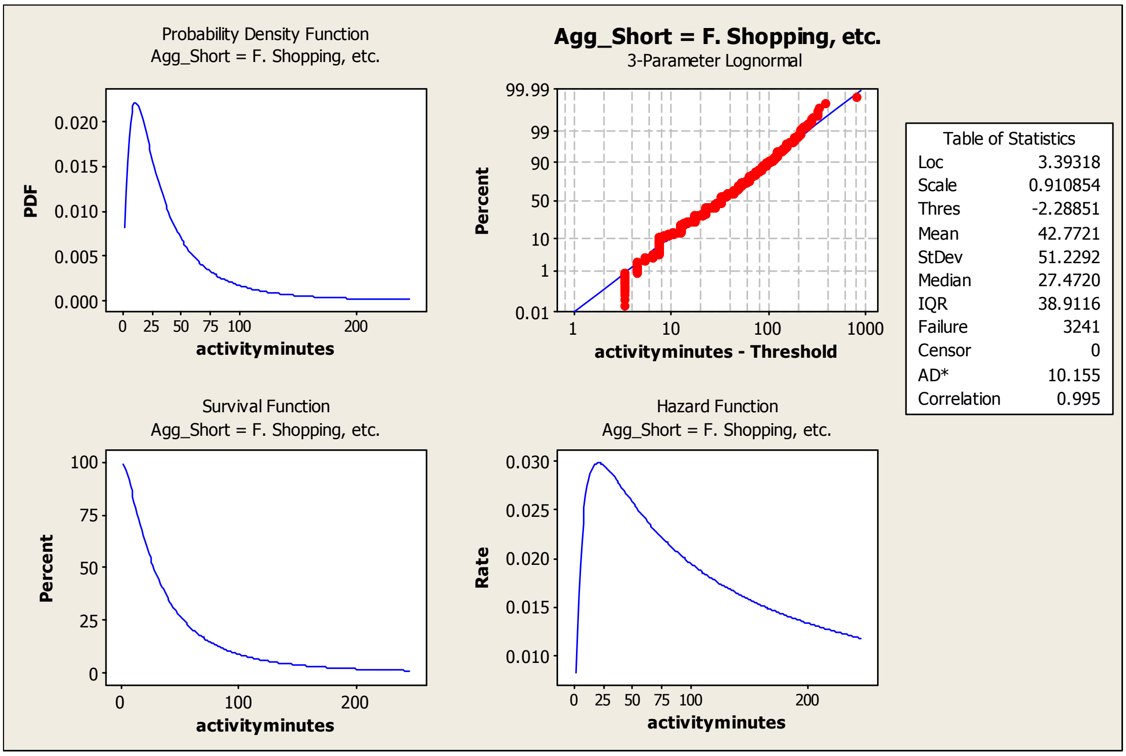

Figure 4 Distribution overview plot for weekend shopping activities

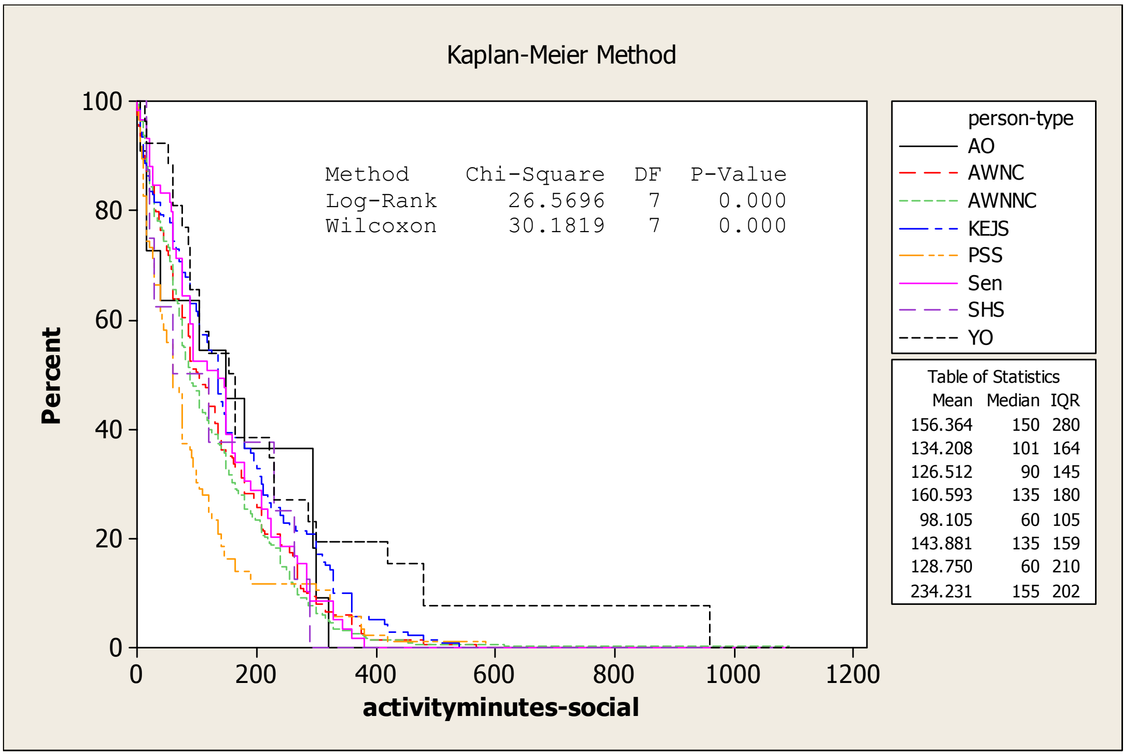

Figure 5 Nonparametric survival plots for social activities

Table 1 Best-fit models for individual activity type

	Travel Related		Work		School		Shopping		Social		Eating		Ent/Leis		Exercise		Religious etc.		Out of town	
Distribution	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR
Weibull	127.45	0.81	12.25	0.98	1.53	0.98	29.16	0.98	2.51	0.99	19.41	0.98	9.15	0.99	3.06	0.98	4.45	0.96	2.53	0.84
Lognormal	49.21	0.91	41.51	0.92	4.39	0.93	15.71	0.99	19.88	0.96	9.22	0.99	37.28	0.94	4.45	0.95	10.45	0.91	2.93	0.77
Exponential	133.57	*	83.53	*	18.69	*	18.14	*	7.18	*	96.98	*	80.46	*	49.57	*	93.35	*	3.66	*
Loglogistic	58.99	0.90	37.57	0.92	3.96	0.93	23.68	0.99	20.67	0.96	10.28	0.99	33.50	0.94	3.10	0.97	7.60	0.93	2.92	0.78
3-Parameter Weibull	98.24	0.84	3.90	0.98	0.69	0.99	13.05	0.99	2.62	0.99	17.67	0.98	4.49	0.99	3.29	0.99	3.47	0.99	2.36	0.91
3-Parameter Lognormal	50.63	0.92	3.10	0.99	0.67	0.99	10.16	1.00	6.02	0.99	8.12	0.99	2.75	1.00	1.49	0.99	1.54	1.00	2.33	0.89
2-Parameter Exponential	100.22	*	82.13	*	16.59	*	39.78	*	7.13	*	88.46	*	77.37	*	46.61	*	91.26	*	3.96	*
3-Parameter Loglogistic	61.07	0.90	5.33	0.98	0.86	0.99	23.02	0.99	12.83	0.98	10.46	0.99	4.75	0.99	0.76	1.00	2.08	0.99	2.30	0.90
Smallest Extreme Value	191.63	0.38	69.94	0.91	6.01	0.93	584.98	0.73	125.1	0.81	139.88	0.77	95.32	0.86	30.16	0.84	37.47	0.87	2.38	0.91
Normal	102.77	0.49	9.04	0.98	0.89	0.99	184.82	0.86	30.75	0.92	46.01	0.88	17.42	0.95	7.87	0.93	8.09	0.95	2.33	0.89
Logistic	99.84	0.52	10.52	0.97	0.87	0.98	177.40	0.87	34.19	0.92	43.68	0.89	16.08	0.96	6.73	0.94	8.31	0.96	2.29	0.90
Best-fit model	Lognormal		3-parameter lognormal		3-parameter lognormal		3-parameter lognormal		Weibull		3-parameter lognormal		3-parameter lognormal		3-parameter loglogistic		3-parameter lognormal		3-parameter Weibull	

Table 2 Goodness-of-fit tests for demographic subgroups of social activities

	Assumed Distributions																					
	1		2		3		4		5		6		7		8		9		10		11	
Personal Type	Normal		Exponential		2-parameter exponential		Weibull		3-parameter Weibull		Lognormal		3-parameter lognormal		Smallest extreme value		Logistic		Loglogistic		3-parameter loglogistic	
	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD	COR	AD
AO	0.943	1.60	*	1.79	*	2.048	0.964	1.53	0.965	1.53	0.931	1.72	0.945	1.58	0.919	1.97	0.932	1.67	0.926	1.79	0.935	1.65
AWNC	0.954	5.47	*	3.05	*	3.478	0.987	1.41	0.989	0.87	0.94	6.40	0.988	1.38	0.862	24.01	0.948	6.39	0.94	6.04	0.975	2.75
AWNNC	0.914	14.41	*	5.25	*	4.676	0.995	1.35	0.995	1.31	0.968	7.35	0.989	3.05	0.81	56.27	0.915	16.24	0.966	8.00	0.977	5.97
KEJS	0.964	3.22	*	3.84	*	3.846	0.985	1.58	0.985	1.46	0.938	5.48	0.985	1.20	0.882	15.70	0.956	3.82	0.937	5.53	0.972	2.01
PSS	0.864	7.60	*	2.04	*	2.364	0.991	0.82	0.994	0.63	0.984	1.03	0.991	0.73	0.747	24.49	0.869	6.84	0.981	1.20	0.984	1.06
Sen	0.938	1.86	*	1.73	*	1.756	0.96	1.77	0.983	1.67	0.962	1.73	0.963	1.72	0.897	2.27	0.93	1.90	0.955	1.77	0.957	1.76
SHS	0.972	1.40	*	3.27	*	3.166	0.981	0.71	0.993	0.52	0.928	1.87	0.988	0.70	0.906	5.83	0.963	1.65	0.932	1.84	0.978	1.02
YO	0.854	2.94	*	0.94	*	1.372	0.977	1.11	0.979	0.99	0.982	0.76	0.989	0.71	0.755	9.09	0.86	2.51	0.984	0.69	0.989	0.66
Average	0.93	4.81	N/A	2.74	N/A	2.84	0.98	1.29	**0.99**	**1.12**	0.95	3.29	0.98	1.38	0.85	17.45	0.92	5.13	0.95	3.36	0.97	2.11