

업무활동후 활동 참여 및 지속시간에 관한 연구

A Study on activity participation and duration after work activity Using Hazard-based models

공혜영

(연세대학교 도시공학과 석사과정)

정진혁

(연세대학교 도시공학과 부교수)

목 차

- | | |
|-----------|--------------|
| I. 서론 | 3. 기초자료분석 |
| II. 본론 | 4. 지속시간모형 추정 |
| 1. 자료의 수집 | III. 결론 |
| 2. 자료의 정리 | 참고문헌 |

I. 서론

개인이 여러 가지 활동에 어떠한 방식으로 시간을 배분하는가를 규명하기 위하여 많은 연구가 이루어졌다. 그 중 가장 주목할 만한 것은 효용이론에 바탕을 두고 개인은 사용가능한 시간을 모든 활동으로부터 창출되는 효용이 극대화하는 방향으로 배분한다는 가정을 한 것이었다.(Kitamura,1983) 그에 따라 목적지 선택 모형이나 활동시간예측모형에서 각 목적지별, 활동별, 개인별 효용을 미리 정의하고 제약조건하에서 개인이 활동을 통해 얻는 효용을 극대화시키는 해를 찾고자 하였다. 그러나 효용은 해당일의 상황과 필요에 의해 달라질 수 있으며 개인의 사회경제적 속성으로 그룹을 지어 효용을 정의할 경우 활동자체의 특성이나 같은 날 발생한 다른 활동들의 영향을 고려하지 못할 수 있다.

평일의 경우, 사람들은 출근, 등교 등의 의무적 활동 후 쇼핑, 여가 등의 재량활동과 귀가를 선택할 수 있다. 의무적 활동 후 수행하는 재량활동은 그 활동의 수나, 의무적 활동이 언제 마쳤는지에 따라 지속시간이 다르다는 것을 예상할 수 있다. 활동자체의 정해진 효용이 아닌 특

정일의 상황을 반영할 수 있도록 본 논문에서는 위험률기반 지속시간 모형을 활용하였다.

위험률기반 지속시간 모형은 수명측정과 산업공학 분야에서 지속시간자료에 대한 인과관계를 규명하기 위한 수단으로 사용되어 왔으며, 최근 교통분야에도 적용되고 있다.(Nam and Mannering, 1998)

본 논문에서는 위험률 기반 지속시간 모형을 기초로 활동지속시간을 추정하며, 활동지속시간에 영향을 미칠 것으로 판단되는 다양한 변수를 공변량(covariate)로 활용하여 기존의 효용극대화를 바탕으로 한 이론보다 다양한 요인을 고려할 수 있는 일반화된 모형을 구축하고자 한다.

II. 본론

1. 자료의 수집

본 논문에서는 서울시정개발연구원에서 실시한 2006년 서울시 가구통행실태조사 자료 중 평일자료를 바탕으로 330,424명의 일일통행자료를 구축하였다.

2. 자료의 정리

서울시 가구통행실태조사자료를 본 연구의 목적에 부합하는 자료로 만들기 위하여 통행단위로 정리되어있는 자료를 활동단위로 변경하였다. 변수는 가구속성, 개인속성, 통행속성, 활동속성으로 분류할 수 있으며 세부사항은 <표 1>과 같다.

<표 1> 변수의 정의 및 설명

| 구분 | 변수 | 설명 |
|--------------|-------------|--------------------|
| 가구 속성 | HNum | 가구원수 |
| | HKidnum | 미취학아동수 |
| | Hincome | 가구수입 |
| 개인 속성 | PRelate | 세대주와의 관계 |
| | PBirth | 출생년도 |
| | PSEX | 성별 |
| | PLicense | 운전면허 소유 여부 |
| | PJob | 직업 |
| | PEmployee | 고용형태 |
| | PStyle | 근무형태 |
| 통행 속성 | PTravel | 조사일의 통행 여부 |
| | Traveltime | 목적지까지 통행시간 |
| | Mode | 대중교통인지의 여부 |
| | Park | 주차요금 지불 여부 |
| | Card | 교통카드 사용 여부 |
| | Fare | 통행료 지불 여부 |
| 활동 속성 | Nexttravel | 다음 목적지까지 통행시간 |
| | Activityyno | 하루에 수행한 활동의 개수 |
| | Order | 활동의 순서 |
| | Activity | 활동의 종류 |
| | Starthour | 활동의 시작시간대 |
| | Preact | 현 활동 이전에 수행한 활동의 수 |
| | Nextact | 현 활동 다음에 수행할 활동의 수 |
| | Preact_dur | 이전에 수행한 활동의 총 시간 |
| | Nextact_dur | 다음에 수행할 활동의 총 시간 |
| Activity_dur | 현재 활동의 지속시간 | |

3. 기초 자료 분석

1) 일반현황

서울시 전체의 가구통행실태조사 대상은 총 96,727가구이며, 가구원수는 330,424명으로 나타났다. 가구원의 직업분포를 살펴보면 직장인이 43.3%로 가장 많았고, 학생이 25.1%, 미취학아동을 제외한 주부 및 무직이 24.1%, 미취학아동이 6.5%를 차지하는 것으로 나타났다.

<표 2> 서울시 가구통행실태조사 일반현황

| 전체 가구수 (가구) | 가구원수 (명) | 직업분류별 가구원수 (명) | | | |
|-------------|----------|----------------|--------|---------|---------|
| | | 직장인 | 학생 | 주부 및 무직 | 미취학 아동수 |
| 96,727 | 330,424 | 143,113 | 83,041 | 79,547 | 21,518 |
| | 100% | 43.3% | 25.1% | 24.1% | 6.5% |

2) 통행현황

조사대상 가구의 평일통행을 살펴보면, 남성의 경우 통행자가 80.6%, 비통행자가 19.4%를 나타냈으며 여성의 경우 통행자가 65.5%, 비통행자가 34.5%로 나타나 남성의 통행비율이 더 높은 것으로 나타났다.

<표 3> 서울시 통행자 및 비통행자 분포

| 구분 | 통행자 | | 비통행자 | | 계 | |
|-----|---------|---------|---------|---------|---------|---------|
| | 가구원 (명) | 구성비 (%) | 가구원 (명) | 구성비 (%) | 가구원 (명) | 구성비 (%) |
| 남 | 130,600 | 80.6 | 31,480 | 19.4 | 162,080 | 100.0 |
| 여 | 109,636 | 65.5 | 57,723 | 34.5 | 167,359 | 100.0 |
| 미응답 | 57 | 5.8 | 928 | 94.2 | 985 | 100.0 |
| 계 | 240,293 | 72.7 | 90,131 | 27.3 | 330,424 | 100.0 |

3) 활동현황

직업이 있는 143,113명 중 평일에 통행을 한 사람은 126,574명(88.4%)이며, 통행을 하지 않은 사람은 16,539명(11.6%)이다. 통행을 한 직장인

중 afterwork activity를 한 사람을 각각의 활동에 대해 정리하면 <표 4>와 같다.

<표 4> 서울시 Afterwork 활동자 및 비활동자 분포

| 구분 | 활동자 | | 비활동자 | | 계 | |
|----|---------|---------|---------|---------|---------|---------|
| | 가구원 (명) | 구성비 (%) | 가구원 (명) | 구성비 (%) | 가구원 (명) | 구성비 (%) |
| 쇼핑 | 1,423 | 1.1 | 125,151 | 98.9 | 126,574 | 100.0 |
| 여가 | 2,305 | 1.8 | 124,269 | 98.2 | 126,574 | 100.0 |
| 기타 | 2,518 | 2.0 | 124,056 | 98.0 | 126,574 | 100.0 |
| 전체 | 5,883 | 4.6 | 120,691 | 95.4 | 126,574 | 100.0 |

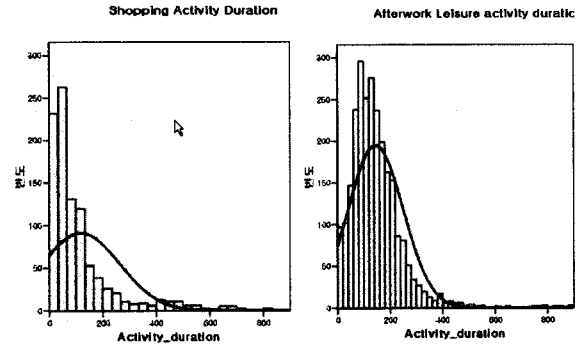
각각의 활동에 대하여 afterwork activity활동의 횟수와 전체 통행한 직장인에 대한 비율은 <표 5>와 같다.

<표 5> 서울시 Afterwork 활동자의 활동 횟수

| 구분 | | 1 | 2 | 3 | 4 | 5 | 6 | 계 |
|----|-----|-------|-----|-----|-----|-----|-----|-------|
| 쇼핑 | 가구원 | 1,356 | 60 | 5 | 1 | 1 | 0 | 1,423 |
| | 비율 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 |
| 여가 | 가구원 | 2,193 | 102 | 8 | 1 | 1 | 0 | 2,305 |
| | 비율 | 1.7 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 1.8 |
| 기타 | 가구원 | 2,238 | 238 | 35 | 3 | 4 | 0 | 2,518 |
| | 비율 | 1.8 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| 전체 | 가구원 | 5,146 | 615 | 98 | 12 | 9 | 3 | 5,883 |
| | 비율 | 4.1 | 0.5 | 0.1 | 0.0 | 0.0 | 0.0 | 4.6 |

4) 지속시간 자료 분석

쇼핑, 여가 각각의 활동에 대한 지속시간의 빈도분포이다. 아래 그림에서 보는 것처럼 쇼핑, 여가 모두 200분 이내에 대부분이 종료되는 것을 알 수 있으며 쇼핑활동은 30~60분 사이에, 여가활동은 100~150분 사이에 지속시간이 집중해 있는 것을 알 수 있다.



<그림 1> 쇼핑,여가활동의 지속시간 빈도분포

5) 상관분석

자료의 정리작업을 수행한 후 각 변수들과 지속시간과의 관계를 피어슨 상관계수를 통해 측정하였다.

<표 5> 쇼핑지속시간과 각 변수의 상관관계

| 구분 | 변수 | 상관계수 | 유의확률 |
|-------------|------------|------------|----------|
| 가구 속성 | HNum | -0.018 | 0.566 |
| | HKidnum | -0.010 | 0.760 |
| | Hincome | -0.021 | 0.502 |
| 개인 속성 | PRelate | -0.071* | 0.025 |
| | PBirth | -0.101** | 0.002 |
| | Psex | -0.142** | 0.000 |
| | PLicense | -0.026 | 0.415 |
| | PJob | -0.097** | 0.002 |
| | PEmployee | 0.040 | 0.207 |
| | PStyle | 0.020 | 0.535 |
| 통행 속성 | Traveltime | 0.068* | 0.032 |
| | Mode | -0.012 | 0.703 |
| | Park | -0.034 | 0.288 |
| | Card | 0.030 | 0.344 |
| | Fare | 0.039 | 0.219 |
| | Nexttravel | 0.100** | 0.002 |
| | 활동 속성 | Activityno | -0.233** |
| Order | | -0.182** | 0.000 |
| Starthour | | -0.469** | 0.000 |
| Preact | | -0.182** | 0.000 |
| Nextact | | -0.160** | 0.000 |
| Preact_dur | | -0.411** | 0.000 |
| Nextact_dur | | -0.153** | 0.000 |

<표 6> 여가지속시간과 각 변수의 상관관계

| 구분 | 변수 | 상관계수 | 유의확률 |
|----------|-------------|----------|-------|
| 가구 속성 | HNum | -0.031 | 0.199 |
| | HKidnum | -0.016 | 0.508 |
| | Hincome | 0.021 | 0.398 |
| 개인 속성 | PRelate | -0.028 | 0.243 |
| | PBirth | 0.002 | 0.930 |
| | PSex | -0.033 | 0.178 |
| | PLicense | 0.047 | 0.053 |
| | PJob | -0.004 | 0.880 |
| | PEmployee | -0.008 | 0.738 |
| | PStyle | -0.002 | 0.932 |
| 통행 속성 | Traveltime | 0.025 | 0.296 |
| | Mode | 0.032 | 0.183 |
| | Park | -0.054* | 0.206 |
| | Card | -0.018 | 0.454 |
| | Fare | -0.041 | 0.095 |
| | Nexttravel | 0.074** | 0.002 |
| 활동 속성 | Activityno | -0.206** | 0.000 |
| | Order | -0.173** | 0.000 |
| | Starthour | -0.384** | 0.000 |
| | Preact | -0.173** | 0.000 |
| | Nextact | -0.141** | 0.000 |
| | Preact_dur | -0.304** | 0.000 |
| | Nextact_dur | -0.164** | 0.000 |

4. 위험률 기반 지속시간 모형

위험률 기반으로 한 지속시간의 모델링은 중도절단자료와 시간의존적 변수를 전통적인 회귀방법에 비해 상대적으로 간단하고 유연하게 모델링 할 수 있는 장점을 가지고 있다.

1) 위험함수

지속시간모형에서 기본 개념이 되는 것은 위험함수이다. t시점에서의 위험함수는 t시점까지 생존한 사람이 t시점 바로 직후 순간적으로 사망할 조건부 확률로 정의할 수 있으며 다음의

식으로 표현할 수 있다.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= \frac{1}{\Pr(T \geq t)} \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}$$

위의 식에서 T 는 생존시간이며, $\Delta t \rightarrow 0$ 로서 t 시점 직후의 순간적 사건을 표현하고 있다. 또한 위험함수는 생존함수인 $S(t)$ 와 밀도함수인 $f(t)$ 로 정의될 수 있으며, 이에 따라 누적분포함수로 표현할 수 있다.

$$h(t) = \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{-dS(t)/dt}{S(t)} = \frac{-\ln S(t)}{dt}$$

누적분포함수 $F(t) = \Pr[T < t]$ 즉, 주어진 시간 t 보다 앞서 지속을 끝낼 확률을 나타낸다고 할 수 있다.

2) 모수적 모형

모수적 모형은 지속시간 자료에 대해서 특정 분포를 따른다는 가정을 하고 파라미터를 추정하는 방법이다. 모수적 모형에는 비례위험모형과 가속 실패시간모형이라는 두 접근방법이 폭넓게 사용되고 있다. 비례위험모형은 회귀파라미터를 설명할 수 있는 장점과, 기저함수의 형태를 설명하는데 쉬운 장점을 가지고 있기 때문에 실패시간 연구에서 폭넓게 사용되었다.

모수적 모형의 대표적인 분포에는 지수, 와이블, 로그-로지스틱, 로그정규 분포가 있으며, 이 중 와이블 분포의 위험함수는 여러 형태를 포함하기 때문에 널리 사용되고 있다.

와이블 분포는 지수분포의 일반화된 형태로, 지수분포의 위험함수가 시간 t 에 독립적인 것과는 달리 종속적이며, 와이블 분포의 위험함수, 생존함수, 밀도함수는 다음과 같다.

$$h(t) = \lambda P(\lambda t)^{P-1}$$

$$S(t) = \exp(-\lambda t^P)$$

$$f(t) = \lambda P(\lambda t)^{P-1} \exp(-\lambda t^P) \quad (\lambda > 0, P > 0)$$

3) 비례위험 모형

비례위험의 형태는 기저함수에 외부공변량의 영향을 곱한 것으로 설명된다. 즉, 지속시간에 영향을 미치는 외부공변량을 위험함수에 포함시키기 위해 기저함수에 곱한 형태로 표현되며, 이 경우 일정한 비율로 증가한다고 가정하는 것으로 다음의 식과 같이 표현할 수 있다.

$$h_1(t) = \exp(B_i \times X_i) h_0(t)$$

$h_0(t)$ = 외부공변량의 영향을 고려하지 않은 위험함수

$h_1(t)$ = 외부공변량 X_i 를 가진 case의 관찰시간 t 에서의 위험함수

B_i = 외부공변량 X_i 의 regression coefficient

4. 지속시간 모형 추정

1) 지속시간 모형 설정

본 논문에서는 활동지속시간을 추정하는데 있어 조건부 확률 개념의 위험률 기반 지속시간 모형을 사용하였다. 위험함수는 활동의 지속시간에 영향을 주는 여러 요인들의 영향을 설명할 수 있다는 장점이 있다. 본 논문에서 사용한 위험률기반 지속시간 모형(hazard-based duration model)에는 종속변수의 분포를 가정할 수 있는 모수적모형(Parametric model)과 종속변수의 분포를 가정하지 않는 비모수적모형(nonparametric model)이 있다. 모수적 모형에서 주로 사용되는 분포는 지수분포, 와이블 분포, 로그-로지스틱 분포, 로그-정규 분포가 사용된다. 이 중 와이블 분포의 위험함수는 비교적 단순한 형태를 지니면서도 여러 증감의 경우를 표현할 수 있는 장점을 가지고 있어, 활동지속시간 자료를 와이블분포로 가정하였다. 또한, 지속시간에 영향을 미치는 공변량(covariates)를 모형내에 적용시키기 위한 방법으로 위험함수에 공변량을 적용시키면 위험함수가 항상 비례적인 관계에 있게 되는 비례위험률모형(proportional hazard model)을 가정하였다.

2) 지속시간 예측모형 추정

본 연구에서는 활동지속시간에 영향을 미치는 다양한 요인을 모형에 반영하기 위하여 변수를 크게 가구속성, 개인속성, 통행속성, 활동속성으로 분류하였다. 독립변수의 사용은 상관분석결과를 바탕으로 선정하여 모형의 추정에 반영하였다. 수행한 상관분석 결과 상관관계가 크지 않은 것으로 나타난 변수라도 기존연구에서 유의한 변수로 나타났거나, 연구자의 판단에 비추어 지속시간의 예측에 필요한 변수라 판단될 경우 변수를 모형의 추정에 사용하였다.

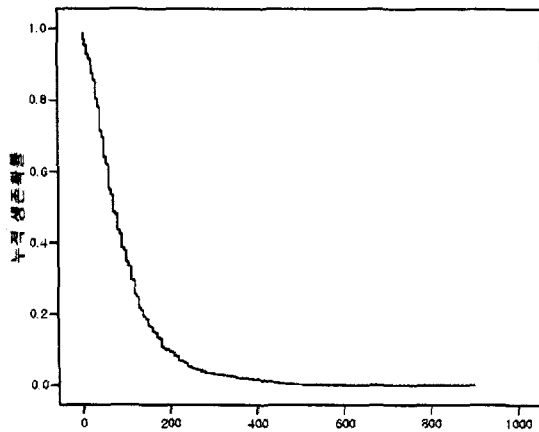
3) 지속시간 예측모형 추정결과

쇼핑활동의 지속시간을 추정한 결과 활동을 시작한 후 한시간이 경과하면 45%, 2시간이 경과하면 74%의 사람들이 활동을 끝마치고, 3시간 이내에 쇼핑활동에 참여한 사람의 99%가 활동을 종료한다.

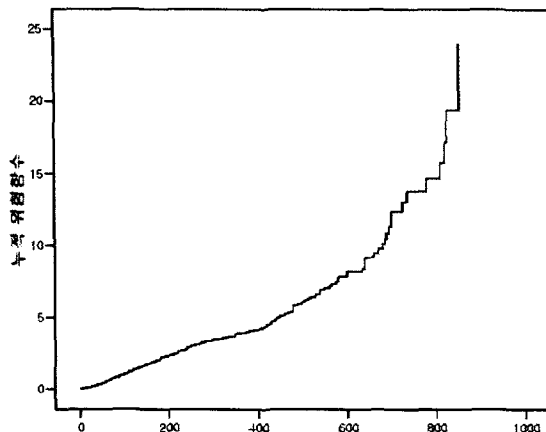
쇼핑활동에 대한 지속시간 모형을 구축하기 위하여 외부공변량의 regression coefficient를 추정한 결과 통행의 수, 활동의 수, 활동시작시간, 다음에 해야 할 활동들의 활동시간에 대한 양의 값을 나타내어, 각 공변량의 값이 커질수록 위험률이 높아져 지속시간은 짧아지는 것을 확인하였다. 그러나 활동을 하기 위하여 이동한 시간이 길거나 활동의 순서가 클수록 활동지속시간은 짧아질 것이라는 일반적인 생각과는 달리 이동한 시간이 길수록, 활동순서가 클수록 지속시간이 길어지는 것을 확인하였다.

<표 7> 쇼핑지속시간 모형 추정결과

| 구분 | 변수 | B | 유의확률 | Exp(B) |
|-------------------------------|-------------|--------|-------|-----------|
| 통행 속성 | Traveltime | -0.004 | 0.034 | 0.996 |
| | Travelno | 0.088 | 0.014 | 1.092 |
| 활동 속성 | Activityno | 0.128 | 0.021 | 1.137 |
| | Order | -0.193 | 0.000 | 0.824 |
| | Starthour | 0.193 | 0.000 | 1.213 |
| | Nextact_dur | 0.002 | 0.000 | 1.002 |
| log-likelihood at zero | | | | 11605.525 |
| log-likelihood at convergence | | | | 11220.926 |
| degree of freedom | | | | 6 |
| χ^2 | | | | 384.297 |



<그림 2> 쇼핑활동의 누적생존율



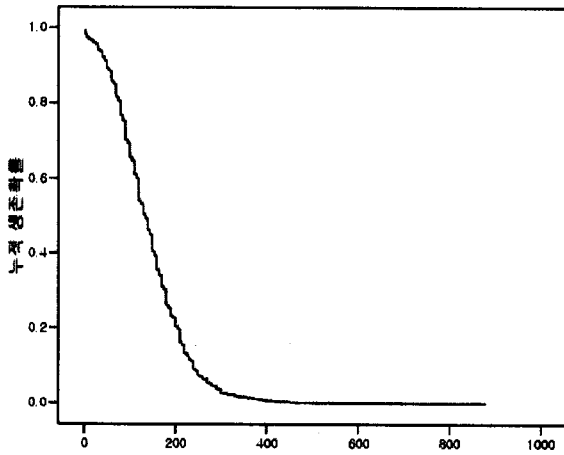
<그림 3> 쇼핑활동의 위험률

여가활동의 지속시간을 추정한 결과 활동을 시작한 후 한시간이 경과하면 14%, 2시간이 경과하면 46%, 3시간이 경과하면 74%, 4시간이 경과하면 91%의 사람들이 활동을 끝마치고, 6시간 이내에 쇼핑활동에 참여한 사람의 99%가 활동을 종료한다. 이를 바탕으로 쇼핑활동에 비해 여가활동의 지속시간이 길다는 것을 알 수 있다.

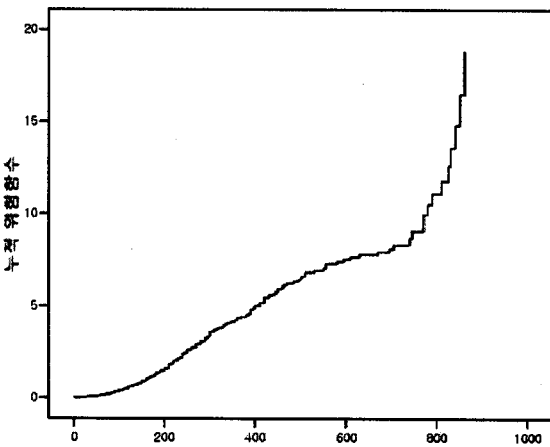
여가활동에 영향을 주는 공변량의 coefficient를 추정한 결과 쇼핑통행과는 달리 가구수입, 통행의 수, 다음 목적지까지의 이동시간, 활동 시작시간, 현재활동 전과 후에 수행한 활동들의 지속시간에 영향을 받는 것으로 나타났다. 가구수입이 높을수록 위험률이 낮아져 여가활동지속시간은 길어지며, 나머지 공변량은 공변량값이 커질수록 위험률은 높아졌다.

<표 8> 여가지속시간 모형 추정결과

| 구분 | 변수 | B | 유의확률 | Exp(B) |
|-------------------------------|-------------|--------|-------|-----------|
| 가구 | HIncome | -0.102 | 0.072 | 0.903 |
| 통행 속성 | Travelno | 0.094 | 0.000 | 1.098 |
| | Nexttravel | 0.002 | 0.079 | 1.002 |
| 활동 속성 | Starthour | 1.101 | 0.000 | 1.106 |
| | Preact_dur | 0.001 | 0.000 | 1.001 |
| | Nextact_dur | 0.005 | 0.000 | 1.005 |
| log-likelihood at zero | | | | 21848.723 |
| log-likelihood at convergence | | | | 21426.785 |
| degree of freedom | | | | 6 |
| χ^2 | | | | 430.344 |



<그림 4> 여가활동의 누적생존율



<그림 5> 여가활동의 위험율

참고문헌

1. 권상순, "Hazard-based duration model을 이용한 고속도로 돌발상황 지속시간 추정에 관한 연구", 석사학위논문, 경기대학교, 2002
2. 김현명, "Activity chain을 이용한 교통망 수요분석 모형", 2005
3. 서울특별시, "2006 서울시 가구통행실태조사 및 보완조사 최종보고서", 2007
4. 송경일 외, "SPSS for Windows를 이용한 생존분석", spss아카데미, 1999
5. Ryuichi Kitamura, "A model of daily time allocation to discretionary out-of-home activities and trip", Transportation Research part B, Vol.18B, pp255-266
6. Simon P. Washington, Matthew G. Karlaftis and Fred L. Mannering, "Statistical and econometric Methods for Transportation data analysis", CHAPMAN&HALL/CRC, 2003

III. 결론

효용극대화를 통한 시간배분의 경우 개인을 사회경제적 속성에 따라 그룹으로 나누어 그룹별 Utility Profile을 구성하고 효용을 극대화하는 방향으로 활동이 지속되었다. 그러나 각 변수와 지속시간의 상관관계에서 개인의 사회경제적 속성은 활동의 지속시간과 큰 관련이 없는 것으로 나타났으며 오히려 이동하는데 걸린 시간이나 하루 중 해야 할 활동 및 그 구성에 더 큰 영향을 받는 것으로 나타났다.