



한국외국어대학교
HANKUK UNIVERSITY OF FOREIGN STUDIES

주택과 구성원의 형태에 따른 가정 전력소비량의 영향요인 분석



제출일 : 2020. 12. 17

파이낸스어널리틱스 (목 789) - 권태연 교수님

컴퓨터 전자시스템공학부, 국제금융학과(이중)

201602023 오형석

I. 연구 배경 및 목적

OECD 회원국 중 한국이 2018 년 기준 국내총생산(GDP)(12 위)과 인구(27 위)에 비해 전력 소비부문에서 2019 년 기준 세계 6 위로 전력 사용량이 많았습니다. 2015 년부터 세계전력소비량 9 위에서 점점 오르기 시작하여 2019 년까지 6 위로 전력을 많이 소비하는 나라가 되었습니다.

우리나라는 2019 년 용도별(산업, 수송, 가정, 상업, 공공) 전력소비량에서는 산업용이 가장 많았고, 이어서 상업용 2 위, 가정용이 3 번째로 많았습니다. 또한 한국전력공사의 통계에 따르면 가정용 판매 전력량은 느린 속도로 줄고 있는 것을 볼 수 있습니다¹. 하지만 여전히 서비스업 다음으로 다른 부문(공공, 광업, 농업)에 비해 판매 전력량이 많은 상황입니다.

2016 년에 한국에너지공단에서 제공되는 자료에 따르면 한국의 1 인당 가정부문 에너지소비는 0.4toe¹로, 우리나라보다 경제규모가 큰 스페인(1 조 4 천 645 억달러)보다 더 많은 에너지를 사용한다고 발표했습니다.

이에 따라 우리에게 가장 밀접하다고 생각되는 가정의 형태와 여러 요소가 가정이 소비하는 전력에 어떤 영향을 끼치는지 분석해보고자 주제로 선정하게 되었습니다.

이 연구는 전국 시, 도의 2,520 가구를 대상으로 실시한 2018 가구에너지 상설 표본조사²를 데이터로 선정해, 참고하고자 하는 변수들의 결측치를 제외, 총 1,752 가구에 대한 표본데이터를 사용해 도출한 회귀 모델은 사회과학 부문에서 여러 실증적 자료로 활용될 수 있을 것으로 생각합니다.

¹ 석유환산톤 (toe, ton of oil equivalent) : 원유 1 톤이 연소할 때 발생하는 열량 약 107 kcal 로써, 1toe 는 연비 14km/L 인 자동 차는 서울과 부산을(410km) 22 번 왕복할 수 있는 양이다. 또한 한 달에 293kWh 의 전력을 사용하는 일반가정에서 약 1 년 3 개월간 쓸 수 있는 전력량과 도 같다. - 한국에너지공단

² 2018 가구에너지 상설표본조사 : 2011 년부터 통계청의 <통계작성승인>을 받아 8 년째 진행, 에너지경제연구원이 작성하고 (유)닐슨컴퍼니코리아 에서 전국 일반 2,520 가구를 표본으로 해 1 년주기로 작성하는 조사통계

II. 회귀분석

1. 분석자료

본 연구의 분석자료로 국가에너지통계 종합정보시스템(KESIS)에서 에너지경제연구원이 실시한 2018 가구에너지 상설표본조사 마이크로데이터를 사용했습니다. 이 데이터는 전국 각 시·도 2,520 가구를 표본 대상으로 한 패널조사 원본데이터로써 같이 첨부된 변수설명파일을 참고하였고 사용한 변수에 대한 자세한 내용은 <표 1>에 첨부하였습니다.

1) 종속변수

다양한 주택과 가정의 형태의 따른 전력소비량을 설명하기 위해 종속변수로 가정의 전력소비량을 종속변수로 지정했습니다. 종속변수인 전력 소비량은 각 가구의 일반전력, 심야전력을 합친 전력 변수의 양을 설정하였고 연간분석을 목표로 했기 때문에 “2017 년 전력 사용량의 합계”를 kW/h 의 단위로 사용했습니다. 그림<2>와 같이 분석 시 등분산성문제를 해결하기 위해 로그변환을 하였습니다.

2) 독립변수

전력소비량 예측을 위해 사용한 독립변수로 크게 주거공간에 관한 변수, 가정 특성에 관한 변수로 구분하였습니다.

주거공간에 관한 독립변수로 면적(m²), 층수, 방 개수 그리고 범주형 변수로 주택형태를 단독주택기준, 연립/다세대주택과 아파트로 구분하였고, 완공 년도를 1970 년 이전을 기준으로 2010 년이후 사이에 10 년씩 나누어 총 6 개에서 하나를 빼 5 개의 범주를 가지고 있는 변수로 정했습니다.

가정특성에 관한 변수로는 가족구성원 수를 비롯해 전 월년 동월 전기요금 확인여부를 4 가지범주로 나눈 변수, 월평균소득을 200 만원미만부터 600 만원이상까지 4 가지로 나눈 범주형 변수로 구성하였습니다.

2. 분석과정

앞에서 정리한 자료를 분석하기 전, 면적과 방 개수, 가족 구성원은 양의 영향을 줄 것으로 예측하였고, 앞서 연구한 것³을 토대로 단독, 연립/다세대, 아파트 순서로 전력 소비량이 감소할 것으로 생각했습니다. 건물 준공 연도 또한 같은 연구에서 유의하게 나왔기 때문에 변수로 선정하였고, 전기요금을 전과 비교하는 가구가 더 전력소비량이 적을 것이라고 예측했습니다. 고소득 가구일수록 전력소비량이 상대적으로 많아 전기요금에 민감하게 반응한다는 선행 연구⁴를 참고하여 월평균 소득을 변수로 선정하였습니다.

이후 전체 독립변수로 모델링한 결과 <표 2>^{iv}와 같은 결과를 보였습니다. 현저히 낮은 R^2 값과 유의하지 않은 독립 변수를 제거하고자 R의 step 함수를 활용해 단계적 선택법(Stepwise Selection)으로 아래 <표 3>와 같이 변수를 추려 보았습니다. 이후, 잔차의 독립성, 정규성과 등분산성을 <그림 2>처럼 검증하였습니다. 그 결과로 다음과 같은 최종 모형이 나오게 되었습니다.

3. 최종모형

분석을 거쳐 도출한 모델은 다음과 같습니다.

<표 3>

Parameter	Estimate	Std. Error	T-Value	Pr(> t)	
α_0 (Intercept)	7.6012698	0.0491824	154.552	< 2e-16	***
x_1	0.0037111	0.0004973	7.463	1.33e-13	**
x_2	-0.0077052	0.0025075	-3.073	0.002153	***
x_3	0.0770974	0.0213554	3.610	0.000315	***
x_4	-0.1567629	0.0321872	-4.870	1.22e-06	***
x_5	-0.1467518	0.0284411	-5.160	2.76e-07	***
x_{11}	0.0312576	0.0091499	3.416	0.000650	***

Adj- R^2 =0.13

P-Value : < 2.2e-16

$$y_i = 7.601 + 0.004x_1 - 0.008x_2 + 0.077x_3 - 0.157x_4 - 0.147x_5 + 0.031x_{11} + \varepsilon_i$$

x_1 = 면적

x_2 = 층

x_3 = 방 개수

x_4 = 연립/다세대주택 (주택형태 기본변수: 단독주택)

x_5 = 아파트 (주택형태 기본변수: 단독주택)

x_{11} = 가족

³ 김민경 (2013). 서울시 주거용 건물의 전력 소비량 추정모델. 서울도시연구, 14(2), 179-192

⁴ Korean Energy Economic Review Volume 14, Number 3, November 2015 : pp. 27~81

III. 결론

이번 연구프로젝트를 통하여 전국 각 가정의 전력 사용량에 각 설정한 변수들이 어떤 영향을 끼치는지 살펴보았습니다. 먼저 단계적 변수선택에 있어서 제외되었던 세가지 범주형 변수에 대해 생각해 보았습니다. 첫째로 준공연도는 <표 2>에서 볼 수 있듯이 준공 연도가 오래될수록 더 양(+)의 영향이 크게 나타나는 것으로 보였으나 준공 연도를 구분하여 분석하기에는 자료가 너무 적기 때문에 변수에서 제외했다고 생각됩니다. 전기요금 고지서 확인 여부 또한 확인하는 빈도가 높은 가정일수록 전력 사용량의 음(-)의 영향이 나타났으나 같은 이유로 탈락된 것으로 보입니다. 마지막으로 월평균소득에 대해서 기본변수인 200 만원 이하에 상대적으로 200~400, 400~600 만원은 ($0.05 < P\text{-value} < 0.1$) 수준의 유의성을 보였지만 월평균소득 만으로는 전력사용량을 설명하는데 있어 각 구성원에 대한 물리적 자료가 부족했던 것으로 보입니다.

회귀분석 결과 추려진 독립변수는 모두 통계적으로 유의함을 보였고 가정의 전력 소비량은 면적, 방 개수, 가족 구성원 수에 각각 0.3%, 7.7%, 3.1%의 양(+)의 영향을 끼치는 것으로 나타났습니다. 이어서 한 층이 높아질 때 마다 0.7%의 전력사용량이 감소하는 것으로 나타났습니다. 마지막으로 주택의 유형이 연립/다세대주택이 단독주택에 상대적으로 15%낮게 나타났으며 아파트의 경우도 14.7% 낮게 나타났습니다. 결과적으로 독립변수가 종속변수를 설명하는 설명력($R^2 = 0.13$)이 너무 낮다고 지적될 수도 있습니다⁵. 그러나 본 프로젝트의 목적이 종속변수에 대한 핵심 영향요인 탐색과 예측을 위한 모델링이 아닌, 종속변수에 대한 독립변수들의 유의미한 영향력을 확인하려는 것이므로, 유의성 검증에서 유의미한 것으로 분석된 모든 변수가 $p < 0.01$ 로 매우 높은 수준의 유의성을 나타내어 본 연구의 타당성을 입증하였습니다.

⁵ 설명력(R^2)의 기준은 학자들마다 차이가 있다. Cohen(1988)의 기준에 따르면 R^2 는 large($0.26 \leq R^2$), middle($0.13 \leq R^2 < 0.26$), small ($0.02 \leq R^2 < 0.13$)로 분류된다. Falk 와 Miller(1992)의 기준에 따르면 R^2 값이 0.10 이상이면 설명력이 적절한 것으로 본다. 본 연구의 위계적 회귀분석의 설명력은 0.107 로 Cohen 의 기준에 따르면 설명력이 낮으나, Falk 와 Miller 의 기준인 10%를 충족하여 본 연구는 유의미하다고 볼 수 있다

IV. 한계점 및 소감

본 연구를 수행하면서 한계점으로 인식될 만한 것들이 몇 가지가 존재하였습니다.

첫 번째는, 여러 참고문헌을 참고한 결과 가구의 전력소비량통계는 도시가구와 도시 외 가구의 전력소비행태가 매우 이질적이기 때문에 따로 구분하여 분석하는 것이 바람직하다고 되어있었습니다. 하지만 연구에 활용한 자료에는 서울, 광역시, 그 외 시도로 구분하고 있었기 때문에 대상에 대해 모호했던 점이 큰 한계라고 생각합니다.

두 번째로 이번 연구에서는 가족 구성원 수를 독립변수로 두어 연구를 하였지만 단지 구성원의 수가 아닌 구성원의 물리적 특성(나이, 인구, 직업, 교육수준, 남, 여 비율)에 따라 전력 소비 패턴을 고려한다면 전력소비량 결정요인에 대해 다른 유의미한 결과를 도출할 수 있을 것이라고 생각합니다.

세 번째로 에너지 공단의 자료에 따르면 전력, 가스, 등 여러 에너지소비량은 산업의 경제활동성, 계절별 기온변화, 국가 정책, 등 여러 요인에 따라 변동가능성이 있다고 합니다. 그 중 기온의 변화에 따른 가정의 냉, 난방의 수요가 직접적인 영향이라고 생각하였지만 이 또한 연간 데이터를 통해 분석하는 것이 목표였기 때문에 시계열 요소가 포함되지 않은 최종 모델의 독립변수들 만으로는 전력사용량을 설명하는데 한계가 존재할 것으로 생각합니다.

마지막 느낀 점으로 이번 프로젝트를 수행하면서 전공수업에서 듣지 않은 회귀분석과 다양한 통계적 지식을 습득함으로써 통계적 지식을 필요로 하는 인공지능을 공부하는데 굉장한 도움이 될 것으로 생각합니다. 또한 이 수업의 목표가 단순히 이전에 수강했던 데이터 시각화가 아닌 설정된 가설과 연구결과의 검증 등 결론을 도출하기위한 통계적으로 근거를 제시하는 방법습득에 목표가 있다고 생각하였는데 앞으로 연구, 또는 프로젝트, 보고서를 수행한다면 R을 사용하여 이 수업에서 배운 것을 비롯한 다양한 분석기법을 통해 결론을 뒷받침하는 추가적인 도구로 사용할 수 있을 것입니다.

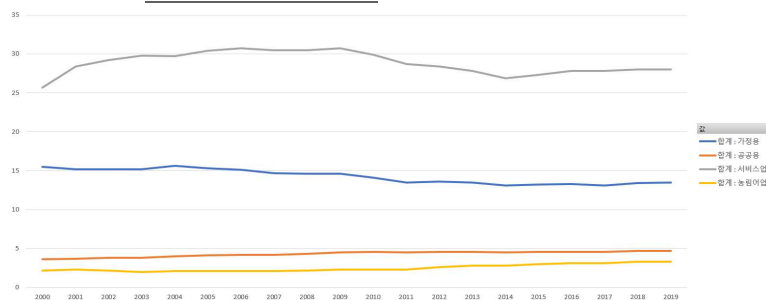
참고 자료

- 김민경 (2013). 서울시 주거용 건물의 전력 소비량 추정모델. 서울도시연구, 14(2), 179-192
- Korean Energy Economic Review Volume 14, Number 3, November 2015 : pp. 27~81.
- Cohen, J.(1988), Statistical Power Analysis for the Behavioral Sciences(2nd Ed.), Lawrence Erlbaum Associates, Inc.

- 에너지관리공단, 2012, "건물(가정부문) 에너지 소비 행태 분석을 위한 DB 구축 및 활용: Part 1.

에너지관리공단 DB 를 활용한 가정의 에너지소비행태에 관한 연구," 에너지관리공단 연구보고서.

i <그림 1> 용도별 판매 전력량, 한국전력공사 「한국전력통계」,2000~2019

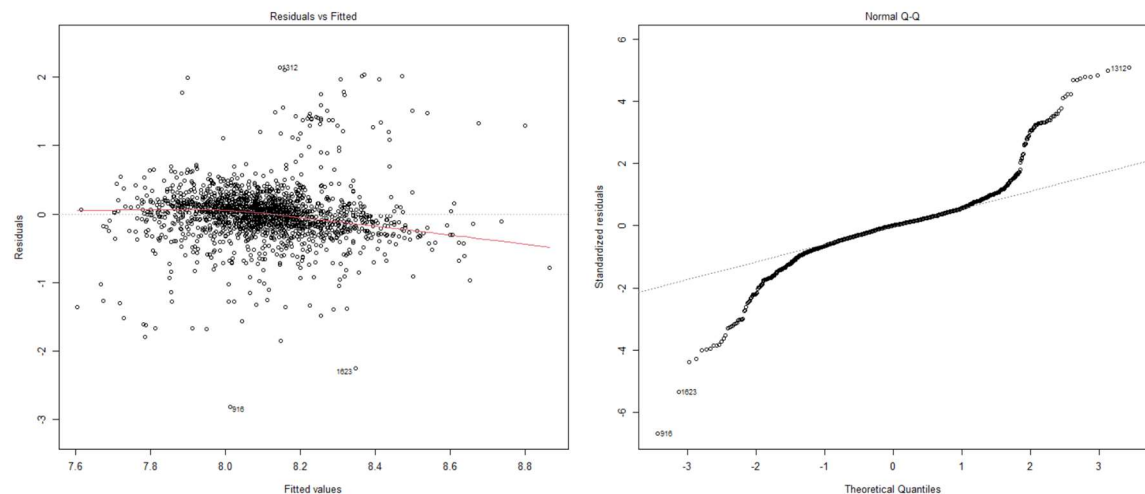


ii <표 1>

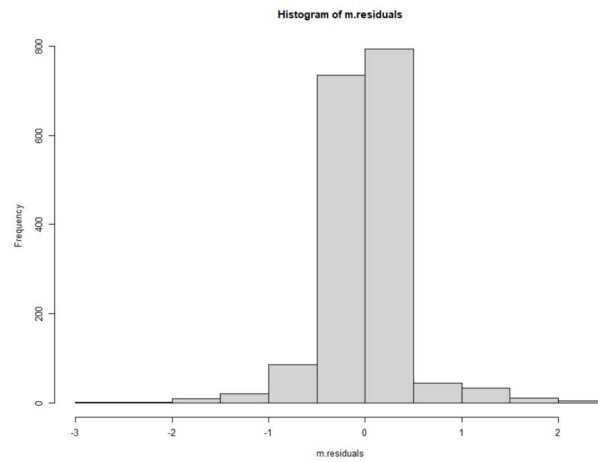
변수				값	비고
종속변수	연간 전력소비량 총계			Y_E	kW/h (log 변환)
독립변수	주거 공간	면적		x_1	m ²
		층수		x_2	연속형 변수 (층)
		방 개수		x_3	연속형 변수 (개)
		주택형태	단독주택	-	(주택형태 기본변수)
			연립/다세대	x_4	단독주택=0,다세대/연립주택=1
			아파트	x_5	단독주택=0 아파트=1
		완공 년도	1970 년이전	-	(건물연도 기본변수)
			1971-1979 년	x_6	~1970=0, 1971-1979=1
			1980-1979 년	x_7	~1970=0, 1980-1989=1
			1990-1979 년	x_8	~1970=0, 1990-1999=1
	2000-2009 년		x_9	~1970=0, 2000-2009=1	
		2010 년 이후	x_{10}	~1970=0, 2010~=1	
	가정 특성	가족 구성원 수		x_{11}	연속형 변수 (명)
		전 월년 동월 전기요금 확인 여부	매월확인함	-	(확인여부 기본변수)
			확인하는경우>안하는경우	x_{12}	매월확인 =0, 확인> :1
			확인안하는경우>하는경우	x_{13}	매월확인 =0, 확인안함> :1
			전혀 확인 안함	x_{14}	매월확인 =0, 전혀 확인안함=1
월평균소득		200 만원 미만	-	(월평균소득 기본변수)	
		200-400 만원 미만	x_{15}	~200=0, 200~400 = 1	
		400-600 만원 미만	x_{16}	~200=0, 400~600 = 1	
	600 만원 이상	x_{17}	~200=0, 600~ = 1		
상수항				a_0	-
계수				a_k	-

iii <그림 2>

잔차의 등분산성



잔차의 정규성



잔차의 독립성(Durbin-Watson 검정)

```
> durbinWatsonTest(m.residuals)
[1] 1.842005
```

iv <표 2>

Residual standard error: 0.4224 on 1722 degrees of freedom

Multiple R-squared: 0.138, Adjusted R-squared: 0.1295

F-statistic: 16.22 on 17 and 1722 DF, P-value: < 2.2e-16

Parameter	Estimate	Std. Error	T-Value	Pr(> t)	
a_0 (Intercept)	7.6012698	0.0491824	154.552	< 2e-16	***
x_1	0.0037111	0.0004973	7.463	1.33e-13	**
x_2	-0.0077052	0.0025075	-3.073	0.002153	***
x_3	0.0770974	0.0213554	3.610	0.000315	***
x_4	-0.1567629	0.0321872	-4.870	1.22e-06	***
x_5	-0.1467518	0.0284411	-5.160	2.76e-07	***
x_6	0.0791554	0.0712712	1.111	0.266886	
x_7	0.0233047	0.0554132	0.421	0.674128	
x_8	0.0260253	0.0535347	0.486	0.626930	
x_9	0.204847	0.0547293	0.374	0.708233	
x_{10}	-0.0109765	0.0599864	-0.183	0.854833	
x_{11}	0.0312576	0.0091499	3.416	0.000650	***
x_{12}	-0.0272605	0.0242716	-1.123	0.261533	
x_{13}	-0.0003743	0.0263300	-0.014	0.988661	
x_{14}	0.0373796	0.0536661	0.0697	0.486196	
x_{15}	0.0582438	0.0298232	1.953	0.050985	.
x_{16}	0.0621709	0.0329797	1.885	0.059581	.
x_{17}	0.0425685	0.0426226	0.999	0.318066	