

# 공공건축물의 초기공사비 산정방법 연구

## Cost Estimating method for the Public Office building at the early stage

구원용\*, 김정곤\*\*, 이준석\*\*\*, 박형근\*\*\*\*

Koo, Won-Yong\*, Kim, Jung-Gon\*\*, Lee, Jun-Seok\*\*\*, Park, Hyeong-Geun\*\*\*\*

### 요약

이 연구는 프로젝트 기획단계에서 발주자가 실적자료를 활용해 공공건축물인 관공서의 총공사비를 견적하는 방법에 대한 것이다. 최근 공공건축물의 종류 및 용도가 다양화 되면서, 공사비 산정 시 고려해야할 요소가 증가하여 공사비 추정에 대한 논리적인 분석이 어려운 경우가 많다. 이로 인해 경우에 따라서는 건축물이 과다 또는 부실하게 계획되는 사례가 발생하고 있어 예산이 낭비되기 쉽다. 특히, 학교, 관공서, 체육시설 등의 공공건축공사에 BTL(Build Transfer Lease)방식이 도입되어 활성화되면서 건축물의 기획 단계부터 적정규모 산정 및 공사비에 대한 체계적 예측 및 관리가 필요하며, 효과적인 건축물의 기획을 위해서는 장기적인 건축물의 이용자 및 건축물의 수명주기(Life cycle)를 예측한 후 건축물의 규모를 산정하는 것이 바람직하다. 그러나 현실적으로 모든 변수들을 종합한 결과를 토대로 건축물을 기획하는 것은 용이한 일이 아니다. 따라서 본 연구에서는 발주자가 건축물의 기획 단계에서 공공건축물의 예산 수립을 쉽게 검토할 수 있는 방법론을 제공하고, 과거에 수행된 유사한 용도의 건축물 실적자료를 활용해, 총공사비를 견적할 수 있는 회귀분석모델을 제안하고 모델의 적용성검토를 위해 일본의 관공서 건축데이터를 통해서 검증하였다.

키워드: 발주자, 공공건축물, 견적, 회귀분석, 총공사비

## 1. 서론

### 1.1 연구의 배경 및 목적

최근 건설산업은 국제화와 더불어 건설기술의 발달, 프로젝트의 대형화, 그리고 시장원리의 도입에 따른 다양한 형태의 프로젝트 추진방식 및 관리방식의 확산으로 인해 빠르게 변화하고 있다. 공공건설 프로젝트에 있어서도 이러한 변화들로 인해 국가예산의 효율적인 집행과 관리가 요구된다. 따라서 발주자의 역할이 무엇보다 중요하며, 올바른 의사결정을 통해 기획부터 건축물에 대한 논리적인 접근이 필요하다. 그러나 대부분의 발주자는 전문지식과 경험 부족, 그리고 한정된 시간 등으로 인해 정확한 판단을 내리기 어려운 경우가 많다. 공공건축공사의 기획단계에서 발주자가 결정해야 하는 가장 중요한 정보 가운데 하나는 건축물의 디자인과 예산을 포함한 적정한 규모이다. 건축물의 적정한 규모산정을 통해 계획을 구체화하고 예산을 결정할

수 있으며, 규모산정을 위해서는 발주자가 이해하기 쉽고 신뢰할 수 있는 건축물 계획수립 모델이 필요하다. 이를 통해 발주자가 예측 가능한 오차범위 안에서 쉽게 공사비의 추정이 가능할 것이다. 일반적으로 공사비의 산정은 건설공사 수행의 여러 단계에서 이뤄지게 되는데, Hendrickson (1989)은 크게 설계견적(Design estimates)<sup>1)</sup>, 입찰견적(Bid estimates), 예산견적(Control estimates)으로 구분하였다. 각 견적 방법은 기본적으로 공사비산정절차 및 활용 목적에 따라서 방법이 달라 고유의 오차를 포함하는데, Hamilton<sup>2)</sup>은 이 오차를 그림1과 같이 정의 하였다. 기획단계에서는 주로 과거의 실적자료를 활용해 견적을 함으로 인해 데이터의 양과 질적인 수준 정도에 따른 오차가 수반된다. 따라서 견적의 오차를 없애는 것은 불가능에 가까우므로 건축물의 기획단계에서 발주자는 건축물의 규모와 예산을 동시에 고려하여, 총공사비를 상세견적에 가깝게 예측함으로써 오차를 줄이는 방법을 택해 결과를 얻게 된다. 그러나 건설프로젝트의 기획단계에서는 설계정보가 확정되지 않고, 정보가 한정되어 있어 총공사비를 예측하는 것은 매

\* 일반회원, 교토대학교 공학연구과 건축학전공 연구원, 공학박사 is.koo@archi.kyoto-u.ac.jp

\*\* 일반회원, 교토대학교 공학연구과 건축학전공 박사과정(교신저자) cs.kjg1207@archi.kyoto-u.ac.jp

\*\*\* 일반회원, 교토대학교 공학연구과 건축학전공 박사과정 cs.yoslove9@archi.kyoto-u.ac.jp

\*\*\*\* 종신회원, 한국건설기술연구원, 건설관리연구실 책임연구원, 공학박사 hgpark@kict.re.kr

1) 규모산정(Order of magnitude estimates), 기본견적(Preliminary or Conceptual estimates), 상세견적(Detailed or Definitive estimates), 그리고 기술견적(Engineer's estimates based on plans and specifications)으로 구분함

2) Albert Hamilton(1997), 견적오차는 규모산정: -30~+50(%), 기본견적: -15~+30(%), 그리고 상세견적: -5~+15(%)

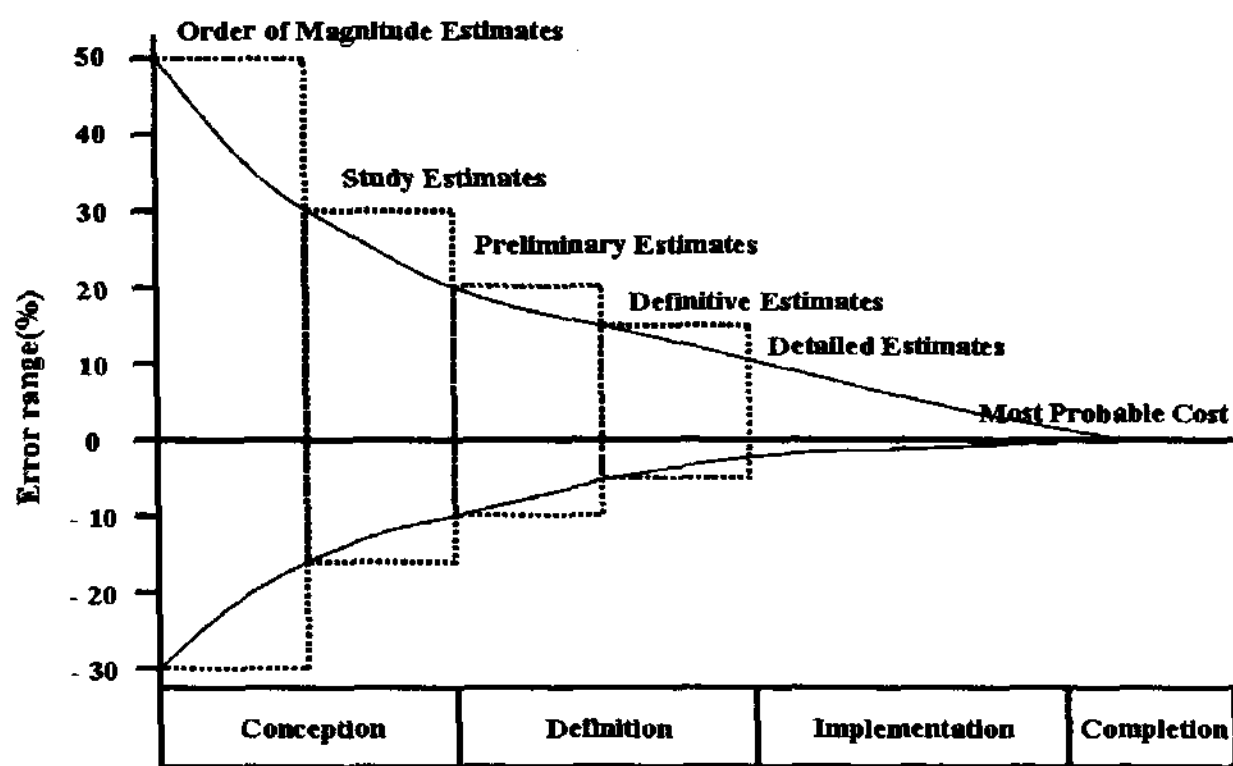


그림1. 건적단계별 오차범위(Albert Hamilton, 1997)

우 어려운 것이 현실이다. 일반적인 근사적 방법으로써 단위면적당 공사비로 총공사비를 구하는 방법이 많이 사용되어 오고 있으나 오차가 큰 단점이 있다. 그러므로 본 연구에서는 건설프로젝트의 기획단계에서 적절한 예산규모를 쉽게 산정하기 위해, 발주자가 과거에 수행한 비슷한 유형의 건축물 실적자료와 데이터마이닝 수법을 활용해 실제 건설 프로젝트에 소요되는 총공사비를 쉽게 예측할 수 있는 회귀모델을 제안하고자 한다. 또한 실적자료에 포함된 이상치의 영향을 최소화하기 위한, 부호제약조건을 회귀모델을 실적자료에 적용하여 예측결과를 검토하였다.

### 1.2 관련연구 현황

공사비 건적 방법에는 개산건적과, 상세건적<sup>3)</sup>이 있다. 전자는 프로젝트의 기획단계 및 기본설계단계에서 발주자가 타당성 검토를 하는 경우에 용이하며, 후자는 실시설계 이후 도면이 확정지어지면 이루어지게 된다. 따라서 기획단계에서는 개산건적기법을 이용하게 된다. 개산건적 방법을 적용방법에 따라 분류하면, 통계적 방법을 활용해 단위면적당 가격을 기준으로 하는 방법, 공종별로 건축물의 부위를 구별하고 시간, 지역 등의 영향을 지수화 하여 단가를 결정하는 방법, 그리고 회귀분석을 이용한 통계적 추정방법 등이 있다. 공사비 추정연구로서 김선국(2000), 조훈희(2002), 김광희(2004) 등은 인공지능망과 유전자 알고리즘과 같은 발견적수법(Heuristic method)을 사용하였고, 김광희(2004), 안성훈(2005)은 공사비 예측에 사례기반추론 기법을 적용하였다. 그러나 발견적수법의 경우는 입력변수와 결과의 인과관계에 대한 설명이 명확치 않고 활용이 쉽지 않아 발주자가 기획단계에서 사용하는데 무리가 있다. 김기동(1991)은 공동주택의 공사비를 결정하는 회귀분석식을 개발해 실적데이터로부터 설계 및 시공에 활용할 수 있는 방법을 연구하였고, 전재열(2003)은 실적공사비 단가의 효율성을 위해 공종, 부위별 그리고 물량의 영향을 고려해서 단가를 결정하는 회귀분석을 연구하였으며, 박종원(2004)은 군사시설의 적정예산 결정을 위해 회귀분석 방법을 사용하였다. 그리고 손보석(2007)은 공종분류 이외에 BE(building elements) 및

3) Boyd C. Paulson, Jr.(1995), 상세건적(detailed estimates)은 단위가격(unit-price)과 고정가격(fixed-price) 방식이 있으며, 대형 프로젝트 및 공공건설에는 단위가격 방식이 주로 적용됨

층별 수량 분류를 추가하고 회귀분석 방법을 통해 공사수량 변화에 적용 할 수 있는 모델을 연구하였다. 이러한 관련연구는 공사비를 구성하는 주요 요소를 정의하고 요소를 중심으로 분석한 결과를 토대로 한 부분적인 공사비연구에 치중한 경향이 있으며, 본 연구에서 제안하는 총공사비규모 산정 방법에 대한 연구는 최근에야 이뤄지고 있다.

### 1.3 연구의 방법 및 범위

일반적으로 건설프로젝트는 규모가 크고, 수많은 요소들이 관련되어 있기 때문에 건축물 실적자료에는 의도하지 않은 노이즈나 이상치(outlier)가 포함되는 경우가 많다<sup>4)</sup>.

이러한 이상치들이 총공사비의 예측에 미치는 영향을 독립적으로 분석하는 것은 어렵지만, 이상치의 영향을 적게 받는 예측모델을 개발함으로써 이상치의 영향을 감소시킬 수 있다. 또한, 실적데이터는 데이터 항목들 간에 밀접한 관계를 가지고 있고 공사비는 각 계층 간에 상, 하위 항목들이 높은 상관관계를 갖는 계층구조로 이루어진 경우가 많아 데이터구조를 바탕으로 실적데이터에서 최적의 예측 모델을 만들기 위한 입력변수들을 선정하는 것은 어렵다. 본 연구에서는 현재 일반적으로 회귀모델에서 사용하고 있는 최소자승법보다 이상치의 영향을 적게 반영할 수 있도록 절대오차를 이용한 회귀모델을 적용하고, 합리적인 변수선택을 통해 공사비예측성능을 향상시키는 방법을 검토하고 있다. 각 예측모델의 성능평가는 모델의 학습성능이 아닌 예측모델의 예측성능을 기준으로 비교하였다<sup>5)</sup>. 연구에 사용된 데이터는 일본의 관공서건축 데이터로서 동일한 기관에서 발주한 유사형태의 건축물 자료이다. 본 연구에 이용한 모든 건축물 자료는 동일한 표준설계지침 및 규격화된 설계프로세스를 통해서 설계가 이루어지므로, 데이터 간에는 상당히 유사성이 높다. 예측모델개발에 사용된 건축물은 일본 전역에 과거 수십 년간 지어진 78개의 관공서들로서, 표1과 같이 다양한 크기의 실적데이터를 고려해 데이터가 다양성을 지니고 있어 예정공사비의 추정 시 높은 예측성능과 보편성을 기대할 수 있을 것으로 판단하였다. 또

표1. 연구에 사용된 건축물 데이터 범위

데이터 항목	데이터 범위	단위
총공사비	2 ~ 33	억엔
본관 연면적	771 ~ 35,192	m <sup>2</sup>
본관 지상층수	2 ~ 5	층
본관 지하층수	0 ~ 2	층
본관 1층 면적	617 ~ 7814	m <sup>2</sup>
공지면적	852 ~ 7,799	m <sup>2</sup>
부속건물 연면적	0 ~ 2,030	m <sup>2</sup>

4) 노이즈와 이상치는 명확하게 구분하기 힘들며, 노이즈는 오퍼레이터의 데이터 입력 과정에서 발생하는 실수 또는 시스템의 에러에 의한 의도하지 않은 데이터라고 할 수 있고, 이상치는 전체 데이터의 일반적 성질과 확연히 다른 성질을 가지고 있는 데이터로 정의된다. 본 연구에서는 전처리 단계에서 노이즈에 대해 충분히 처리되었다는 가정 하에 이상치만 고려함

5) 모델의 학습 성능은 예측모델이 사용된 데이터를 얼마나 잘 설명하고 있는지에 대한 평가이며, 모델의 예측성능은 미지의 데이터에 대한 예측모델의 정확성에 대한 척도임

한 공사비는 공사가 진행된 시기와 지역에 의한 물가의 차이가 변수로 작용하게 되므로, 실적데이터를 모델개발에 사용하기 위해서는 물가상승의 영향을 데이터로부터 제거해만 한다. 여기서는 데이터 전처리단계에서 실적데이터를 1990년의 도쿄(東京)의 물가를 기준한 물가지수와 지역지수로서 보정하여 사용하였다.(具源龍, 加藤直樹2005)

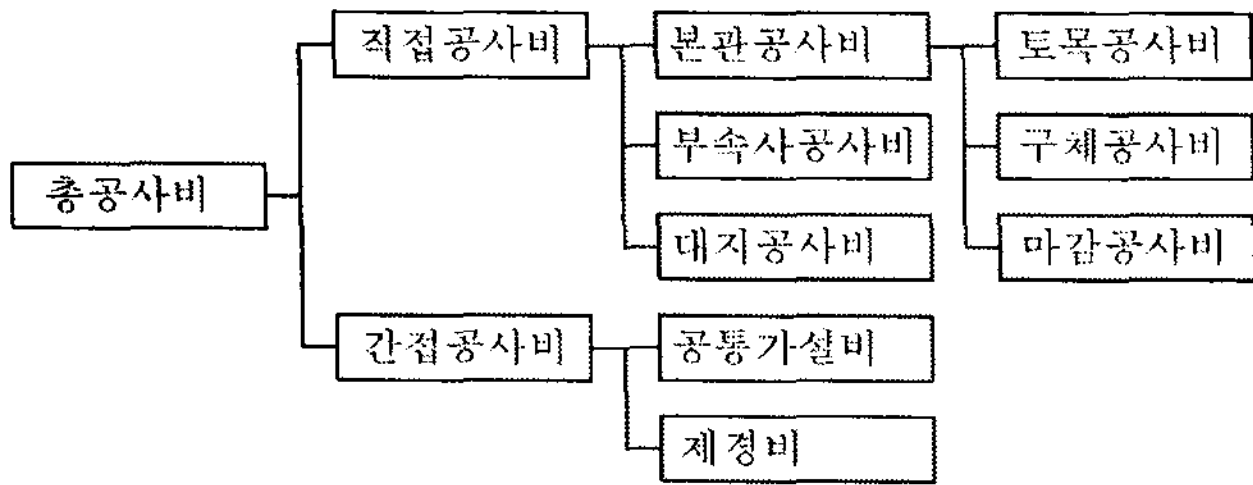


그림2. 총공사비의 구성

## 2 제안방법

### 2.1 부호제약조건의 회귀모델

일반적으로 선형회귀모델은 종속변수  $y$ 에 관한 정보를 가지고 있는  $p$ 개의 독립변수  $x_1, x_2, \dots, x_p$ 를 이용해 식(1)의 선형함수로 표현할 수 있다.

$$y = \sum_{i=1}^p \beta_i x_i + \beta_0 \quad (1)$$

여기서,  $\beta_0$ 와  $\beta_i$ 는 회귀계수이며,  $\beta_i (i = 0, \dots, p)$ 를 구하는 방법으로는 다양한 수법이 제안되어 있지만, 주로 실제값과 예측값의 잔차  $r_j$ 의 제곱을 최소화함으로써 회귀계수를 구하는 최소자승법이 일반화되어 있다<sup>6)</sup>.

$$\text{Minimize} : \sum_j (r_j)^2 \quad (2)$$

그러나  $L^2$ 회귀모델은 잔차  $r_j$ 의 제곱의 합을 최소화하기 때문에 데이터에 이상치가 포함된 경우 이상치에 의한 영향이 커지는 특성이 있다. 본 연구에서는 이러한 문제를 해결하기 위해서, 식(3)의 절대오차최소법을 도입하였다. 절대오차최소법은 잔차의 절대값의 합을 최소화함으로써 회귀계수를 구하는 방법으로 최소자승법에 비해 상대적으로 이상치의 영향을 적게 받으며, 선형계획법을 이용해 계산할 수 있다<sup>7)</sup>.(Robert J. Vanderbei 2001)

$$\text{Minimize} : \sum_j |r_j| \quad (3)$$

더욱이 본 연구에서는 회귀식의 사용성에 주목하여, 회귀계수들의 부호와 관련된 부호제약조건의  $L^1$ 회귀모델을 제안하고자 한다. 일반적으로 사용되는 회귀모델에서는 각 입력변수의 계수들의 부호에 대해서는 특별히 제약이 없다. 그러나 이러한 회귀식을 실무에 적용하기 위해서는 만들어진 회귀식이 사용자관점에서 얼마나 쉽게 이해하고 사용할 수 있는지가 중요하다. 만약 회귀식에서 계수들의 부호가 수치들과 관련성 없이 다르게 나타나는 경우, 회귀분석의 장점이라고 할 수 있는 설명가능성이 떨어지므로, 입력변수

와 출력변수간의 상관계수의 부호를 기준으로 입력변수의 계수의 부호를 결정하는 제약조건을 추가함으로써 회귀모델의 설명가능성을 높이고자 한다. 부호제약조건의  $L^1$ 회귀모델은 다음과 같은 방법으로 구할 수 있다. 식(2)의  $L^1$ 회귀모델은 식(3)과 같은 선형계획 문제로서 정의될 수 있다.

$$\text{Minimize} : \sum_j r_j \quad (4)$$

$$s.t \quad -r_j \leq b_j - \sum_i a_{ij} x_i \leq r_j$$

식(4)의 제약조건에서  $a_{ij} x_i$ 를 회귀계수의 부호결정을 위해 회귀계수의 부호가 양인  $\alpha_{kj} x_k$ 와  $\beta_{lj} x_l$ 로 나누면 식(5)와 같다.

$$\text{Minimize} : \sum_j r_j \quad (5)$$

$$s.t \quad -r_j \leq b_j - \left( \sum_k \alpha_{kj} x_k + \sum_l \beta_{lj} x_l \right) \leq r_j$$

$$x_k \geq 0$$

$$x_l < 0$$

식(5)에서  $x_k$ 는 종속변수와 상관계수가 양수인 독립변수의 계수이고,  $x_l$ 은 종속변수와 상관계수가 음수인 독립변수의 계수이다.  $j$ 는 입력데이터의 실적자료를 나타낸다.

### 2.2 변수선택

회귀모델은 사용되는 독립변수의 조합에 따라서 그 예측 성능이 다르다. 특히 상관관계에 있는 변수들을 독립변수로 사용하는 경우, 다중공선성<sup>8)</sup> 등의 문제가 자주 발생해 예측모형의 성능을 저하시킬 수 있다. 이러한 문제들은 많은 독립변수 중에서 유용한 소수항목의 정선을 통해 예측모델의 학습에 사용하는 변수를 선택함으로써 해결할 수 있으며, 이론적 기반이 있는 경우에는 그 기반에 따라서 변수선택을 하면 된다. 그렇지 않은 경우 탐색적 방법으로 변수를 선택해야 한다. 적절한 변수의 조합에 대한 전역탐색은  $n$ 개의 변수에 대해서  $2^n - 1$ 개의 경우의 수가 발생하고, 변수의 개수가 일정수준을 넘어서면 현실적으로 계산이 불가능하다. 본 연구에서는 변수 선택의 방법으로, 변수증가법과 Correlation Feature Subset(CFS)법을 사용하여 예측실험 및 성능 검토를 실시하였다. 변수증가법은 최초단계에서 종속변수와 상관계수가 가장 큰 독립변수를 중회귀식의 생성에 사용하며, 다음 단계에서는 남아있는 독립변수를 중회귀식에 대입해 모형을 생성한 후, 예측성능의 개선이 가장 크게 이루어진 독립변수를 선택하고, 성능개선이 이루어지는 동안 이 단계를 반복해 나간다. CFS법은 종속변수와 상관계수가 높으면서 다른 독립변수와의 상관계수가 낮은 독립변수를 좋은 독립변수로 보고, 이러한 변수들을 우선적으로 선택하는 수법이다.(M. A. Hall and L. A. Smith 1997) 식(6)은 설명변수의 최적조합을 탐색해주는

<sup>6)</sup> 최소자승법에 의한 회귀모델을  $L^2$ 회귀모델이라고 함

<sup>7)</sup> 절대오차 최소법에 의한 회귀모델을  $L^1$ 회귀모델이라고 함

<sup>8)</sup> 다중공선성(Multicollinearity): 변수들의 상관관계로 인해 발생하는 문제로서 일반적으로 상관관계가 클수록 크게 발생함

식으로써  $G_s$ 가 최대인 경우를 탐색해 변수를 선택한다.

$$G_s = \frac{\overline{pr_{ci}}}{p\sqrt{p+p(p-1)\overline{r_{ii}}}} \quad (6)$$

여기서,  $p$ 는 설명변수의 개수,  $\overline{r_{ci}}$ 는 설명변수와 목적변수의 상관계수의 평균, 그리고  $\overline{r_{ii}}$ 는 설명변수간의 상관계수의 평균이다.

### 2.3 회귀모델의 평가

예측모델의 평가는 데이터마이닝의 중요한 과정 중 하나이다. 모델의 성능을 평가하는 기준으로는 학습 성능과 예측 성능이 있으며, 학습 성능은 입력데이터로 적합한 모델을 만든 다음 학습데이터에 대한 학습의 정도로 모델의 성능을 평가하는 방법으로 간단하게 성능평가를 할 수 있는 장점이 있다. 하지만 반드시 학습 성능이 모델의 예측 성능의 좋은 지표가 된다고 할 수는 없다. 그에 비해 예측 성능으로 모델을 평가하는 방법은 일반적으로 입력데이터를 학습데이터와 검증데이터로 나누어서 학습데이터로 예측모형을 만든 다음, 검증데이터의 평균예측오차로 예측모델의 성능을 나타낼 수 있다. 그런데 이 방법은 학습데이터와 검증데이터를 나누는 방법에 따라서 예측모델의 평가가 달라지는 경우가 있다. 여기서는 모델의 예측 성능 평가방법으로서 모델의 학습과 검증에 leave-one-out법(Ian H. Witten and Eibe Frank 2005)을 이용하여 보다 객관적으로 평가하고자 한다. leave-one-out법은  $n$ 개의 입력데이터에 대해서  $n-1$ 개의 실적데이터를 학습데이터로 해서 모델을 생성하고, 나머지 실적데이터로 검증을 하는 방법을  $n$ 개의 실적데이터 전체에 대해서 실시해 구한 평균을 예측모형의 평균예측오차로 인식한다. 이 방법은 데이터의 규모가 아주 클 경우에는 적용하기 힘들지만, 소규모데이터에 적용하기에 적합한 수법이다. 이 방법을 적용하면 개별적인 실적자료에 대한 예측오차의 경향을 분석할 수 있다.

## 3. 실험 및 결과 고찰

### 3.1 입력데이터의 설정

건축물의 실적데이터로부터 총공사비 산정에 필요한 정

보를 변수로 선정하였으며, 78개의 RC(Reinforced concrete) 구조건물의 실적데이터로부터 그림3과 같이 16개의 항목을 입력변수로, 총공사비를 출력변수로 정하였다. 변수로는 본관과 부속건물의 면적과 관련된 변수가 및 건물의 공간구성과 관련된 외벽면적과, 용적 등도 사용되었다. 또한 건물의 평면이 정해질 경우, 표준설계지침안에 따라서 개략적으로 정해지는 외벽의 창호면적도 이 단계에서 사용가능한 설계정보로 보고 입력변수로 사용하였다.

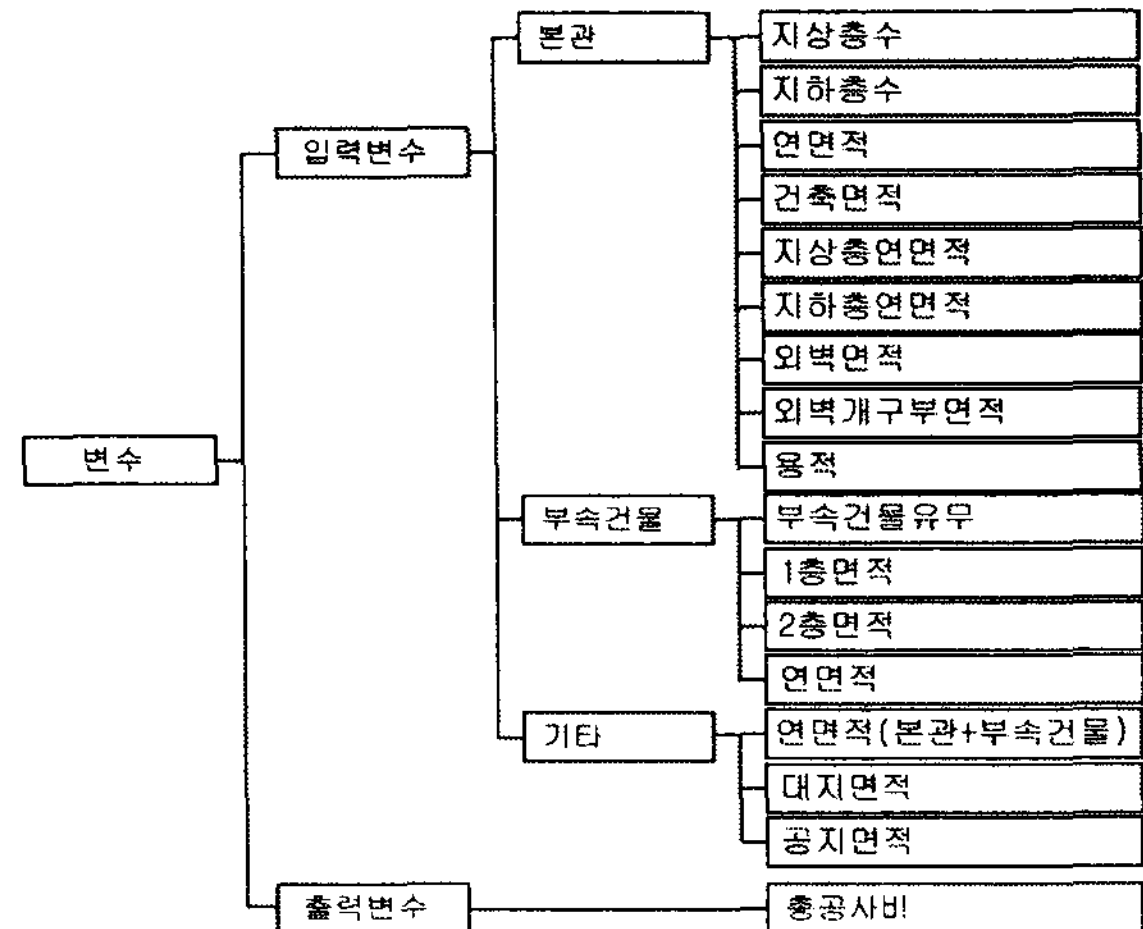


그림3. 입력변수 구성

### 3.2 총공사비 예측모델의 실험 및 결과

실험을 위해 변수증가법, CFS, 그리고 모든 변수를 고려하는 3가지 방법으로 변수를 선택 한 후에  $L^1$ 회귀법과  $L^2$ 회귀법을 적용하는 실험을 수행하여, 표2의 실험결과와 표3의 회귀분석결과를 얻었다. 표2의 실험 결과로부터, 2.3절에서 서술한 것과 같이 예측 성능과 모델 성능 사이에 명확한 상관관계가 나타나고 있지 않음을 알 수 있었다. 따라서 본 연구에서 모델 평가의 기준으로 삼은 예측 성능에 의한 모델평가방법에 의의가 있음을 알 수 있다. 또한 실험결과에서  $L^1$ 회귀모델에 변수증가법을 적용한 실험4가 가장 좋은 예측 성능을 나타내었다.  $L^2$ 회귀모델에 변수증가법을 적용한 실험2에 대해서 비율로는 12% 감소되는 성과를 얻었다. 그런데,  $L^1$ 회귀모델을 적용해서 얻어진 실험4의 회귀식은 예측 성능에 있어서는 가장 우수하지만, 회귀식에 음의 부호를 가진 계수들이 존재하여 논리적으로 맞지 않는다.

표2. 총공사비 예측실험 결과

회귀수법	$L^2$ 회귀		$L^1$ 회귀		부호제약조건의 $L^1$ 회귀		CFS	
	모든변수	변수증가법	모든변수	변수증가법	모든변수	변수증가법		
변수선택 실험	실험1	실험2	실험3	실험4	실험5	실험6	실험7	
예측오차	MAE	147,241,953	117,127,343	602,985,648	103,025,398	131,026,401	105,509,761	133,375,581
	MRE	0.13	0.12	0.53	0.09	0.12	0.09	0.14
	MAX(RE)	0.54	0.46	31.87	0.42	0.48	0.42	0.52
	중상관계수	0.91	0.94	0.17	0.95	0.93	0.95	0.93
학습오차	MAE	258,779,828	549,661,111	90,891,545	100,490,345	597,415,402	642,194,813	114,835,719
	MRE	0.28	0.49	0.08	0.09	0.56	0.60	0.12
	MAX(RE)	0.91	0.77	0.45	0.42	0.81	0.83	0.48
	중상관계수	0.94	0.94	0.95	0.95	0.91	0.92	0.96

\* MAE : Mean Absolute Error(평균절대오차), MRE : Mean Relative Error(평균상대오차), RE: Relative Error(상대오차)



본관 연면적의 경우 연면적과 총공사비는 일반적으로 비례하나, 회귀식에서는 계수가 음의 부호를 가지고 있어 회귀식의 설명가능성이 떨어진다. 부호제약조건을 L<sup>1</sup>회귀모

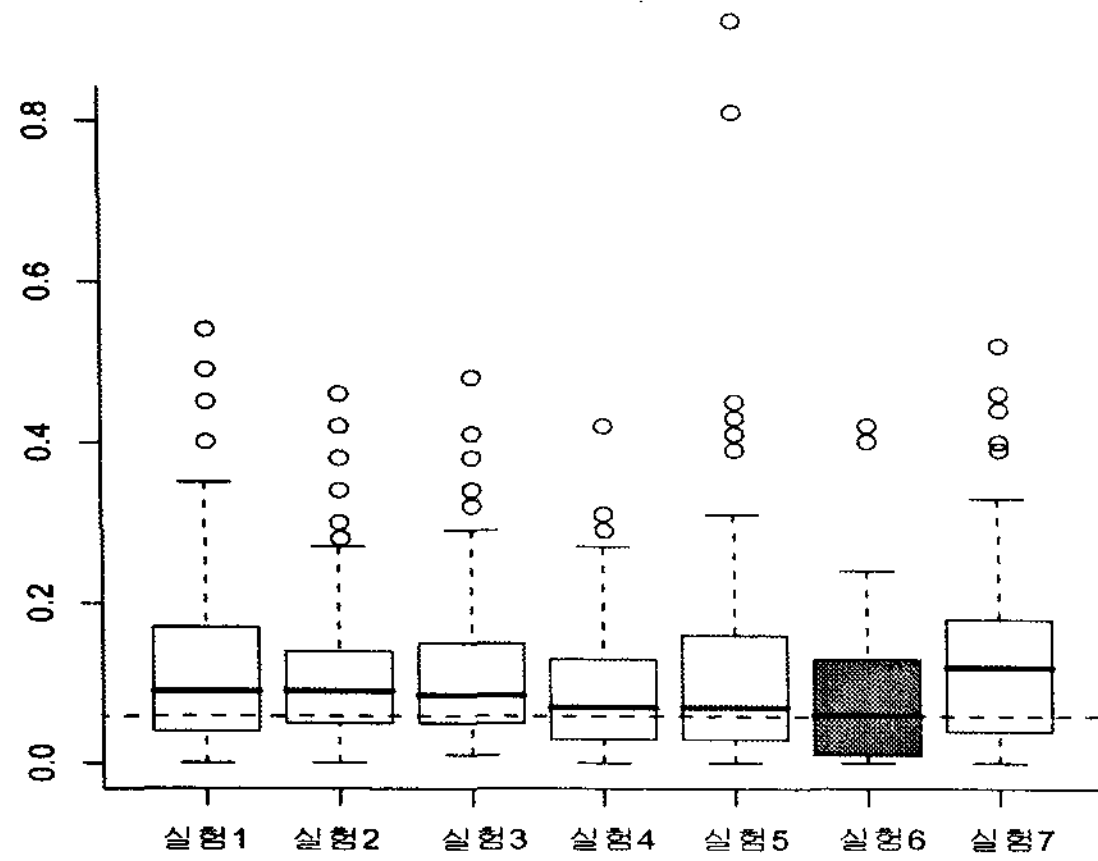


그림4. 예측실험 상대오차의 box-whisker plot

델을 사용한 실험6의 결과는 실험4와 비교했을 때, 그림4에서 볼 수 있듯이 예측성능이 크게 저하되지 않으면서도 논리적인 부호를 가짐으로써 계수들이 실험4에 비해서 설명가능성이 높다고 볼 수 있다. 실험6은 leave-one-out법과 변수증가법의 조합에 의해서 모델의 평가가 이루어지기 때문에 회귀식에 많은 계산량이 요구된다. 실적자료수가  $n$ , 설명변수의 개수가  $p$ 인 경우 변수증가법에 의해서  $q$ 개의 변수가 선택되어져서 만들어진 회귀식은  $np!/(p-q-1)!$ 회의 계산이 필요하다. 실험7은 부호제약조건을 L<sup>1</sup>회귀모델에 CFS법에 의한 변수선택을 적용해 계산한 결과로서 실험6에 비해서 계산량이 적다. 예측성능과 학습 성능은 양쪽 모두 일정수준의 성능을 보여주고 있어서, 실무에서 계산시간 등의 문제로 인해 변수선택법을 사용하기 곤란한 경우에 좋은 대안이 될 수 있다.

### 3.3 의사결정트리를 이용한 예측오차분석

실험6에서 평균절대오차는 0.09, 중상관계수는 0.95의 높

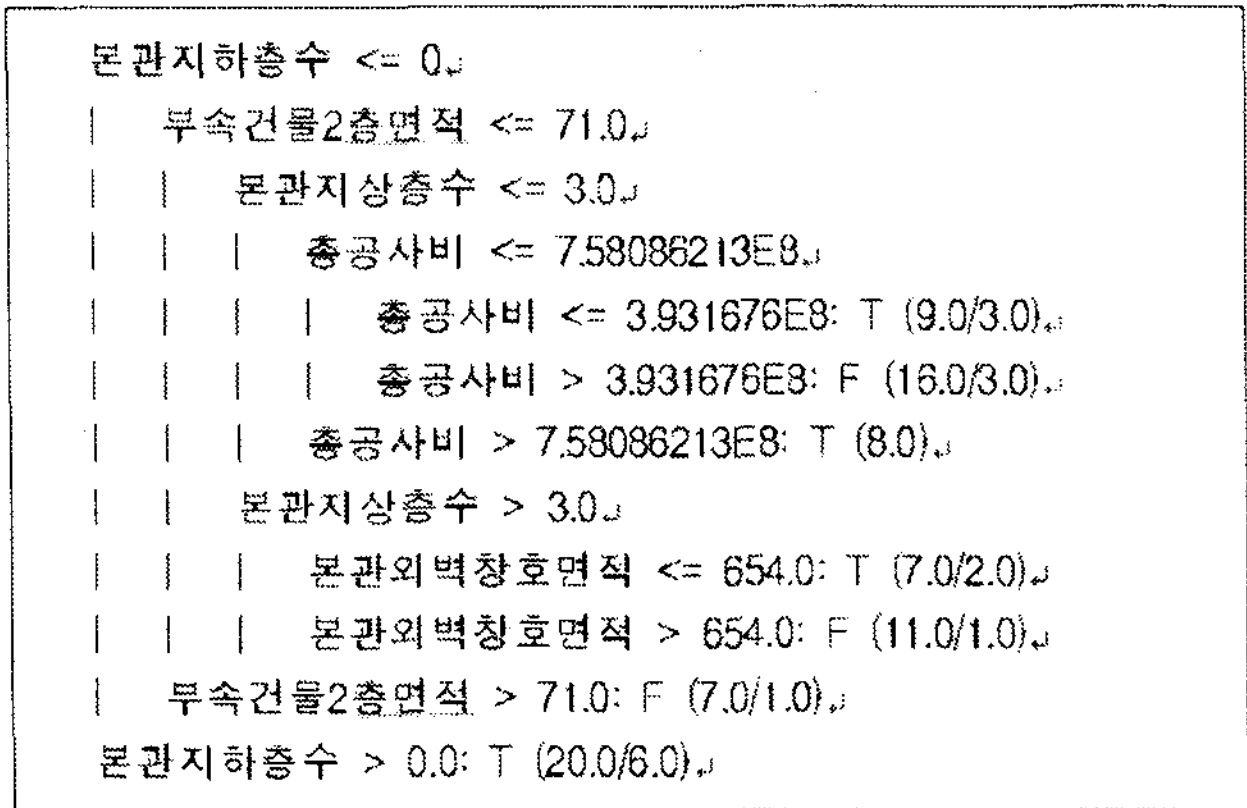


그림5. 의사결정트리를 이용한 예측오차의 분석결과  
 은 예측성능이 얻어졌다. 하지만 상대오차가 0.45로서 예측오차에 어느 정도의 편차가 존재했다. 본 연구에서는 이러한 예측오차의 편차에 대해서 분석을 실시하였다. 실험6의 예측결과를 대상으로 예측오차의 대소에 따라 동일한 크기로 그룹을 나눈 다음, 그림5의 의사결정트리(decision tree)를 통해 예측오차의 크기에 영향을 미치는 요소들에 대해 분석한 결과 79%의 정확도결과가 얻어졌다. 그림5에서 가장 중요한 의사결정 분기법칙으로써 본관지하층수가 사용되었다. 지하층이 있는 26개의 실적자료 가운데 20개의 실적자료에 대한 예측오차가 중앙치보다 큰 것으로 나타났다. 실제 지하층 공사는 토목공사를 수반하였으며, 토목공사는 각 공사현장의 토질과 다양한 조건에 의해 공사비의 변동이 컸다. 다음으로는 부속건물2층 면적, 본관지상층수, 총공사비, 본관외벽창호면적이 사용되었다

## 4. 결론

이 논문에서는 공공건축물의 실적데이터에 대해 데이터 마이닝 수법을 활용함으로써, 건축프로젝트의 기획단계에서 발주자의 예산산정 작업에 실용적으로 사용할 수 있는 회

표3. 회귀분석 실험으로부터 얻어진 회귀식

회귀수법 변수선택 실험	L <sup>2</sup> 회귀 모든변수 실험1	L <sup>2</sup> 회귀 변수증가법 실험2	L <sup>1</sup> 회귀		부호제약조건을 L <sup>1</sup> 회귀		CFS 실험7
			모든변수 실험3	변수증가법 실험4	모든변수 실험5	변수증가법 실험6	
본관지상층수	$\beta_1$	21,968,305	7,966,480		2,856,360		
본관지하층수	$\beta_2$	177,777,896	82,556,658	98,996,200	76,818,600	92,883,500	
본관연면적	$\beta_3$	0	30,612	-12,782,100	-36,817,200		19,544
본관건축면적	$\beta_4$	-11,237		-101,997			85,030
본관지상층면적합계	$\beta_5$	94,992		-2,926,600			
본관지하층면적합계	$\beta_6$	79,954		-2,849,620		55,234	44,183
본관외벽면적합계	$\beta_7$	-654		9,782			
본관외벽창호면적합계	$\beta_8$	19,467		-31,779	-15,068	2,237	42,140
본관용적	$\beta_9$	-4,781		7,822	-11,573		
부속건물유무	$\beta_{10}$	-291,942		19,840,300			
부속건물1층면적	$\beta_{11}$	-367,740		-211,223		-92,267	
부속건물2층면적	$\beta_{12}$	287,052		-240,728			
부속건물연면적	$\beta_{13}$	0	72,432	-15,557,000	-36,904,200	84,953	80,616
건물연면적	$\beta_{14}$	0		15,772,800	36,965,900		56,243
대지면적	$\beta_{15}$	145,043		136,019	9,628	33,186	
공지면적	$\beta_{16}$	-41,575		-156,819			
상수향	$\beta_0$	124,673	157,410,329	62,879,600	105,691,000	79,891,400	119,912,000
							119,529,098

귀모델을 연구하였다. 일반적인 회귀분석의 단점을 보완하기 위해 부호제약조건의 L<sup>1</sup>회귀모델을 제안하고, 공공시설물의 실적데이터를 이용해 제안모델의 유용성을 검토하였다. 회귀식을 도출함에 있어서 부호제약에 의한 회귀식의 설명가능성을 높임으로써 사용자의 접근이 용이하게 하였다. 본 연구에서 제안한 부호제약조건의 L<sup>1</sup>회귀모델은 평균 예측정확도가 91%이었고, 의사결정트리를 이용한 각 실적자료별 오차경향분석은 79%의 정확도를 나타내었다. 그러므로 현재 조달청(2007)의 공공건축공사 낙찰률 평균이 71%인 점을 고려할 때 기획단계에서 발주자의 예산을 수립에 제안모델을 사용하는 것은 예측효과가 있다고 볼 수 있으며, 제안모델을 유사한 형태의 공공시설물 공사에 활용한다면 발주자가 기본적인 설계자료 없이도 쉽게 예산을 시뮬레이션 해볼 수 있어 부분 공사비가 아닌 예산 수립을 위한 총공사비 규모 산정에 적절할 것으로 판단된다.

### 참고문헌

1. 김광희 외 1인(2004). "유전자 알고리즘에 의한 신경망 구조의 최적화를 이용한 공동주택의 초기 공사비 예측에 관한 연구", 대한건축학회지, 대한건축학회, 제20권 제2호, pp.81-88
2. 김광희 외 1인(2004). "사례기반추론 기법을 이용한 공동주택 초기 공사비 예측에 관한 연구", 대한건축학회지, 대한건축학회, 제20권 제5호, pp.83-92
3. 김기동(1991). "우리나라 공동주택의 코스트모델 개발에 관한연구", 서울대학교 박사학위논문
4. 김선국 외 1인(2000) "신경망을 이용한 사무소 건물의 코스트 모델", 대한건축학회, 논문집 제16권 제9호, pp 59-66
5. 박종원(1995). "공사비지수적용을 통한 군 시설공사의 예산산정 모델", 서울대학교 석사학위논문
6. 손보석(2007). "영향변수에 따른 수량변화 분석을 이용한 기본설계단계의 개산견적 모델 개발", 한국건설관리학회지, 한국건설관리학회, 제8권 제2호, pp.155-166
7. 안성훈 외 1인(2005). "전문가지식을 활용한 공동주택 초기단계 공사비 예측에 관한 연구", 대한건축학회지, 대한건축학회, 제21권 제6호, pp.81-88
8. 전재열(2003). "실적자료 분석에 의한 적정 공사비 산정 방법의 전산화 알고리즘 구축에 관한 연구", 한국건설관리학회지, 한국건설관리학회, 제4권 제4호, pp.192-200
9. 조훈희(2002). "국내 건축공사비지수 개발 및 뉴럴 네트워크를 이용한 지수 예측방안에 관한 연구", 고려대학교 박사학위논문
10. 조달청(2007). 정부공사제도 개선방안, 정책보고서
11. Albert Hamilton(1997). *Management by Projects*, Thomas Telford House, pp. 250-254
12. Boyd C. Paulson, Jr.(1995). *Computer Application in Construction*, McGraw-Hill, pp.353-356
13. Hendrickson C., Au T.(1989). *Project management for construction*, Prentice-Hall, pp.125-127
14. M. A. Hall and L. A. Smith(1997). "Feature subset selection: a correlation based filter approach", Proc Fourth International Conference on Neural Information Processing and Intelligent Information Systems, Dunedin, New Zealand, pp.855-858
15. Robert J. Vanderbei(2001). *Linear Programming: Foundations and Extensions*, Springer, pp.195-196
16. Ian H. Witten and Eibe Frank(2005). *Data Mining: Practical Machine Learning Tools And Techniques*, Morgan Kaufmann Pub, pp.151-152
17. 具源龍, 加藤直樹(2005). "データマイニング手法を用いた建築プロジェクトの工事費予測モデルの提案": データ加工による精度の向上を目指して", 日本建築學會環境系論文集, 第598号

### Abstract

In this research, we studied an estimating method in client's sight to estimate the total construction cost which is based on the historical cost data at the early stage of the office buildings as a public phase. It is very difficult to analyze the estimation accurately and logically. When a client estimates a project, he/she has to consider there are many issues at the planning step, according as office buildings become gradually diversified as well as their roles continuously extended. Therefore, those are usually make problems for wasting the budget in accordance with the cost estimation errors. Moreover, many kinds of public construction projects, especially such as school, office, sports complex, and the others, have been invested the private finances defined as BTL(Build Transfer Lease) method that are required to manage the detailed process more strictly from initial planning. In order to make an effective planning, the long-term users amount and the building life cycle at the beginning of project should be considered previously and then it may enable to achieve an appropriate project plan. But actually considering overall variables in a building planning is impossible. Accordingly, suggesting a regression model based on the historical cost data from many similar types of office building to support client's role known as estimating the total cost at the early stage. And then performing the test against the proposed model to research the reasonability as using the historical cost data of Japan office buildings.

**Keywords :** Client, Office buildings, estimates, regression analysis, total cost