

공사비 예측시 이상값 존재하에서 데이터 처리 분석 방안

Methodology of data analyses under presence of outliers for estimating construction cost

오 세 대*

O, Se-Dae

허 영 기**

Huh, Young-Ki

Abstract

Statistical analyses with actual data are used in estimating construction cost for many years, but collected data could include factors that distort analytical results, namely outliers. To enhance reliability in predicting construction cost, the methodology, which is able to identify outliers and determine how to manage them, is needed. Actual costs obtained from 22 construction projects were studied. It is found that there is substantial disparity between results considering outliers and results not considering ones. Therefore, it is to identify outliers and apply an optimum process in estimating construction cost when actual data is used in statistical analysis.

키워드: outlier, 상관분석, 공사비

1. 서 론

1.1 연구의 배경 및 목적

건설프로젝트는 특정의 발주자에게 특정의 건설공사를 도급 받아 현장 시공하여 발주자에게 인도하는 One Time Project의 성격을 가지고 있다. 그러므로 건설프로젝트의 종류, 품질, 가격도 다양하며 표준화되지 않은 작업공정이 대부분이다. 또한 발주자의 선호도, 설계변경, 주변환경, 경제여건 등에 따라 수시로 당초계획이 변경되기 때문에 공사계약에 따른 불확실성이 많다. 이러한 상황은 전체 공사비에 대한 정확한 비용견적을 어렵게 하고 실제공사비와의 오차를 발생시켜 기업의 수익성과 이윤창출에 커다란 영향을 미치게 된다. 따라서 사례를 통한 데이터를 수집하여 비용예측 모델로 사용될 경우 공사비에 포함된 불확실성의 변수가 그대로 데이터에 포함되어 분석되므로 예측된 건축비의 신뢰성을 저하 시키게 된다.

건설 프로젝트의 공사비를 산정하는 체계는 크게 두 가지 관점에서 바라볼 수 있다. 우선, 발주자가 당해 프로젝트의 예산의 책정 및 입찰시 예정가격을 산정하기 위한 건적행위가 있다. 다른 하나는 시공계약자가 당해 사업의 수주와 시공을 관리하기 위해 수행하는 건적이 있다.

2004년 1월부터 국내 공공 건설사업에 도입된 실적공사

비제도는 바로 발주자가 당해 프로젝트의 예산의 책정 및 입찰시 예정가격을 산정하기 위한 건적행위에 관한 것으로 기존의 원가계산방식에 의해 구하던 건설 프로젝트의 공사비 산정방법, 근거, 체계를 일부 개선한 것이다.

실적공사비 적산방식 혹은 현행의 표준품셈에 의한 원가계산방식이 공공 건설사업에서 가지는 의미는 매우 크다고 할 수 있다. 공공 건설사업에서 예정가격은 단순히 발주자의 예산만이 아니라 시공사 입찰금액의 상한선이며, 적격심사제도상의 가격평가의 척도, 입찰금액 산출내역서의 작성 기준, 그리고 수량 증감으로 인한 계약금액 조정의 기준으로 활용되기 때문이다.^{6),7)}

실적공사비 적산방식에서 중요한 점은 표준 데이터베이스의 구축으로 다양한 변수와 각 공사별 조건에 따른 유사성을 가진 데이터의 조합을 통한 신뢰성높은 데이터의 수집을 통한 표준 데이터를 필요로 하게 된다.

사업성 분석단계에서 공사원가 산정은 기본 평면선정 및 동별배치, 법적 및 지리적 요건에 의한 층수산정, 용적율에 의한 주차장 설계여부, 분양성을 고려한 마감재의 수준, 지반조건에 따른 기초의 설계, 지자체의 요구사항 등 수 많은 설계변수들이 존재하고, 변수들은 지역적 조건과 판매방식에 따른 국민임대주택, 공공임대주택, 분양주택 등에 따라 다양한 모습으로 나타나고 있음에도 불구하고, 실적공사비 적산방식에서 수집된 데이터의 값의 분석은 단순통계적인 분석방법으로 평균값 및 중앙값 등으로 공사비를 예측하고 있는바, 수집된 데이터에 내포된 오류가 특별한 통계적인 신뢰성 검증을 통하지 않고 적용되고 있어¹⁾ 예측된 공사비의 신뢰성을 저하시키고 있다.

* 주저자, 공학석사, 한국에이비엔건설(주), 경영혁신팀 팀장

** 교신저자, 부산대학교 건축학부 조교수, 공학박사

(ykhuh@pusan.ac.kr)

이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음.

따라서 수집된 실적 데이터의 단순통계를 통해 적용된 사업비 예측의 신뢰성을 높이고, 실적 사업들의 건축비를 단순통계적인 분석을 통해 건축비 산정함으로 인해 발생하는 비합리성을 최소화하기 위한 단순하며 명료한 방법이 요구되어진다.

1.2 연구내용 및 방법

실적 사업비의 분석을 통한 공사비의 예측시 수집된 실적 사업비 데이터의 경우 각각의 공사별 변수에 따른 특징에 의해 데이터의 일관성과 유사성에 있어 불확실성을 내포하고 있다. 분석을 위해 수집된 데이터의 일관성과 유사성의 결여는 이상값(outlier)의 존재로 결부되어진다.^{1),5)}

실적 사업비를 토대로한 통계치 분석을 통해 사업비의 예측에 있어 이상값의 존재에 유무에 따른 자료의 처리는 예측된 사업비의 신뢰성에 매우 심각한 영향을 미치게 된다.^{2),4),5)}

따라서 본 연구에서는

- 실적 사업비를 토대로 사업비 예측시 존재할 수 있는 이상값의 의미를 알아보고
- 이상값의 처리방안을 기존 문헌을 통해 조사하고
- 기존의 건축비 산정에서 특별하게 고려되지 않았던 이상값의 처리에 대한 방안을 검토하여 실적데이터를 적용한 사업비의 예측시 적용될 수 있는 최적의 비용예측산정의 방안을 제시하고자 한다.

2. 개요

2.1 이상값의 개요

이상점이란 한 변수 또는 둘 이상의 변수에 대해 정상범위 밖으로 아주 동떨어진 관측값을 지니 잔차의 값이 매우 커진 사례를 의미한다. 이상점들은 회귀식을 추정하는 과정에서 매우 강한 영향력을 미친다.

이상값은 자료에 영향을 주는 특징에 따라서 Additive Outlier(AO), Innovational Outlier(IO), Outliers in long Patch(PO), Level-shift Outlier(LS), Temporary(또는 Transient)-Change Outlier(TC)등으로 구분되어진다. AO는 특정시점에서만 나타나는 이상값이며, 대부분 명확한 이유를 설명하기 어렵다. 외부적인 영향 즉 값을 잘못 읽는다거나 기입을 잘못하여 생기는 경우가 많아서 관측 이상값(observational outlier)이라고도 불린다. IO는 특정기간동안 값이 영향을 받는 경우이고, TC는 IO의 특별한 예로 값의 영향이 지속적으로 감소하는 경우에 해당한다.

이상점들이 생겨나는 이유는 다음과 같다.

- ① 부정확한 자료입력
- ② missing value를 잘못 처리했을 경우 (컴퓨터 프로그램을 작성하는 과정에서 결측값을 규정하지 못하여 이것이 실제값으로 읽혀졌을 경우)
- ③ 표집과정에서의 착오로 모집단에 포함되지 않는 사례가 표본으로 선택되었을 경우
- ④ 모집단이 정규분포가 아니라 편포를 이루어 극한값들이 존재하는 경우

하지만 이상점을 제외하거나 변형시키는 것은 주의를 해야 한다. 이상점들은 일반적으로 간과되기 쉬운 자료의 유의미한 특성에 기인하는 것일 수도 있고, 이 경우 분석대상에 대한 매우 중요한 정보의 실마리를 제공할 수 있다.

2.2 이상값의 확인 및 처리 방법

이상점들의 검색은 1차적으로 개개변수의 편포(skewness)나 z점수를 확인하는 과정에서 이루어지나 회귀분석에서의 이상점들의 발견은 잔차를 분석하는 과정에서 이루어진다. 잔차의 산점도를 그려보면 다른 점들로부터 동떨어져 극한값을 지닌 이상점을 쉽게 발견할 수 있다.

- ① 이상값은 크기가 $n \geq 10$ 인 자료의 도수분포가 종형일 때, 이 자료에서 측정값의 z-점수가 3이상인 것들은 모두 이상값으로 판단하거나, 자료의 제1사분위수(Q_1), 제3사분위수(Q_3) 및 사분위간 범위인 $IQR = Q_3 - Q_1$ 을 계산하여 $Q_1 - 3 \cdot IQR$ 보다 작거나 $Q_3 + 3 \cdot IQR$ 보다 큰 값을 가진 측정값들을 자료에서 찾아 이상값으로 판단한다.

- ② 산점도를 이용한 이상값의 확인은 다음 그림1과 같이 직선을 이루고 있는 점들의 범위에서 지나치게 벗어나 있는 값을 이상값으로 판단할 수 있다.

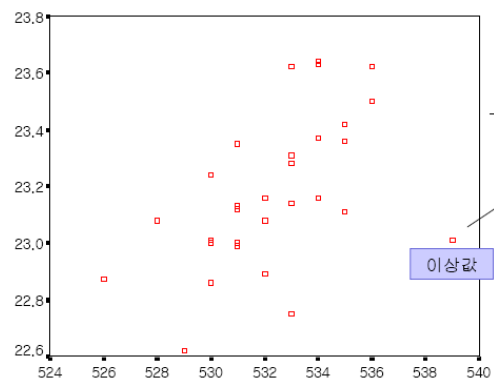


그림 1. 이상값 확인 산포도 예

- ③ Box-Plot을 이용한 이상값의 확인 방법은 다음과 같다. 그림 2는 box-plot의 의미를 보여주며, 그림 3은 Statpro 프로그램을 이용한 box-plot의 수행을 보여준다.

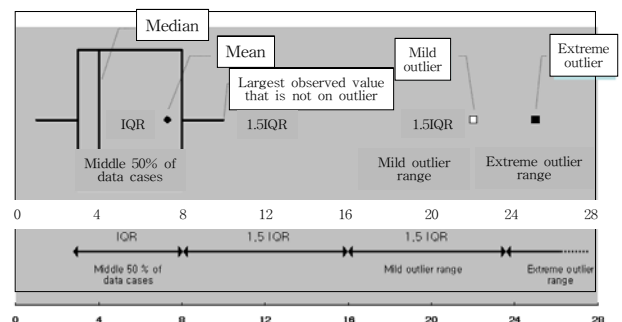


그림 2. Box-Whisker Plot-예

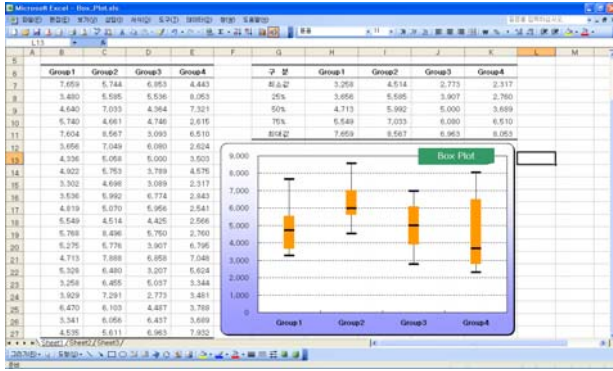


그림 3. 이상값 분석 box-plot 예

④ F-통계량 산점도

다변량 이상값 탐색 방법으로 Yang & Lee(1987)가 제안한 F - 통계량은 다음식과 같다.

$$F_j = \frac{n-p-1}{p} \times \left[\frac{1}{1 - n \cdot s_{jj}^2 / (n-1)^2} - 1 \right]$$

위식에서 F_j 통계량은 분자의 자유도가 p 이고 분모의 자유도가 $n - p - 1$ 인 $F(p, n - p - 1)$ 분포에 근사하게 된다. 이때 유의수준 0.05 에서 다변량 이상값 검정은 검정통계량의 값 F_j 가 기각치 $F(0.05, p, n - p - 1)$ 보다 클 때 자료벡터 y_j 가 다변량 이상값이라고 판정하게 된다.

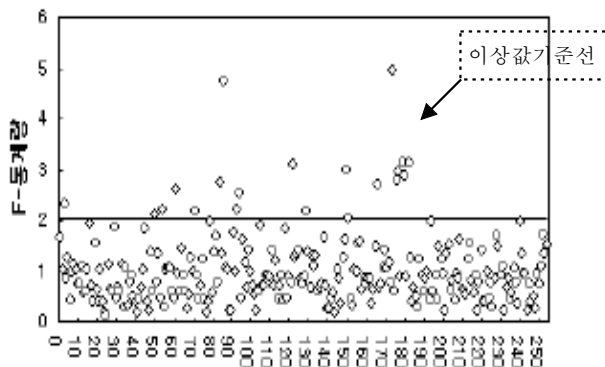


그림 4. F-통계량 산점도 예

2.3 임대주택 공사비산정 코스트 모델

1) 국민임대주택 공사비 추정방법

① 공사비 추정 절차

공사비 분석을 위해 필요한 자료들 중 민간업체들의 자료는 접근이 매우 어려운 면이 있어 비교적 접근이 가능한 주택공사에서 기 발주한 사업들을 선택하여 건축비 내역을 분석하고, 이의 결과를 지방자치단체의 사업실적과 비교·검토하는 것을 기본 방법론으로 선택하여 다음과 같은 추

정 절차를 거친 코스트 예측방안을 제시하고 있다.

② 데이터 수집을 위한 분석대상사업 선정

2004년 주택공사에서 발주한 39개 지구 사업에서 20개 사업을 발주지역, 세대수(약 400~800), 지하주차장 면적등을 고려하여 무작위로 선택하였으며, 2004년 발주된 사업에는 수도권 지역이 포함되지 않아 별도로 2개 사업을 선정하였으며, 수도권(서울) 지역에서 선택된 두 개 사업은 2000년 및 2002년 발주된 사업으로, 물가상승 및 마감재수준 상향조정 등을 고려하여 2004년 시점으로 환산한 금액을 분석에 사용하였다. 2004년 시점에서의 환산은 주택공사에서 현재 사용하고 있는 공동주택 공사비 지수를 활용하여 공사비를 추정하였으며, 총 분석대상사업은 22개사업으로 하였다.

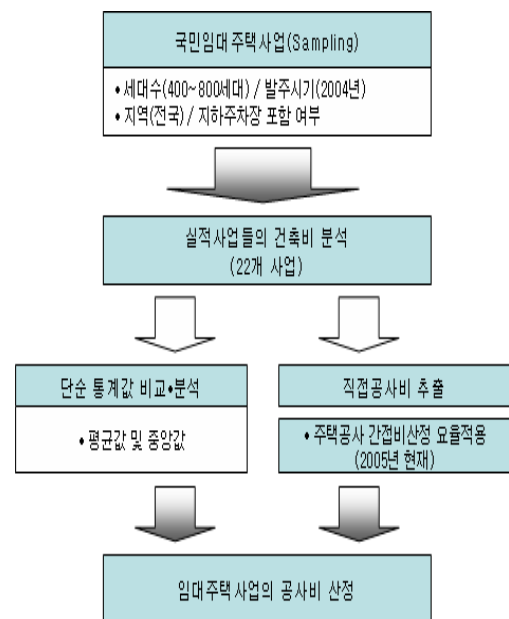


그림 5. 국민임대 주택사업의 건축비 추정 절차

2) 국민임대주택 건축비 분석

위 그림에서와 같이 실적사업들의 건축비 분석에 있어 단순 통계분석을 통해 평균값과 중앙값으로 자료검토하고 있다.

일반적으로 산술평균(arithmetic mean)을 이용한 값을 사용할 경우 이상값에 의해 평균값에 큰 영향을 미쳐 정확성을 떨어뜨리고 불확실성을 내포하게 된다. 중앙값(median, 중위값)을 사용할 경우 이상값에 민감하게 반응하지 않으나 통계적 성질이 복잡하여 이론 전개나 수리적인 분석이 어렵게 된다. 특히 장기간에 걸친 실적 데이터를 이용할 경우 특정한 패턴이나 이상값이 존재할 수밖에 없기 때문에 특정한 패턴이나 이상값의 영향에 민감하지 않는 중앙값(중위값)을 예측값으로 사용하는 것이 정확한 예측이 가능해진다.

이상값이 존재하는 상황에서는 평균값을 이용하여 분석할 경우 오류를 포함하고 있게 된다.

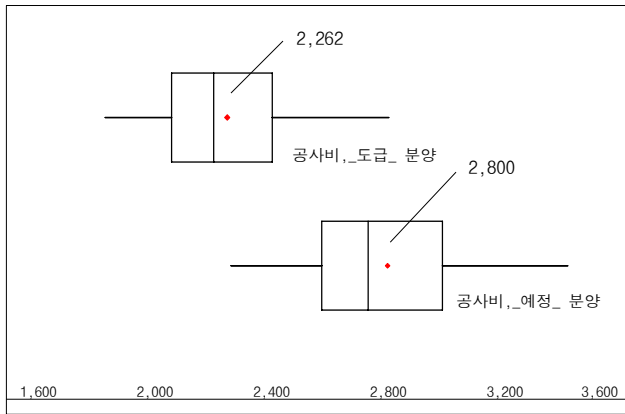


그림 6. Box-Whisker Plot- 22개 사업의 공사비

위 그림6은 실적사업비의 단순통계를 통해 추출된 값을 이용하여 분석한 결과 이상값이 존재하지 않는 것으로 판단하고 있다.

그러나 그림 7과 같이 분석사업대상의 공사항목별 사업비를 분석할 경우 이상값이 존재하고 있음을 알 수 있다. 즉, 총공사비의 평균을 비교할 경우 평균값의 단순통계 비교시 이상값이 존재하지 않는 것처럼 판단되나, 실제로 총공사비의 항목을 이루고 있는 세부항목에서 이상값이 존재하고 있으므로 단순평균값을 이용한 판단은 오류를 내포하고 있다고 할 수 있다.

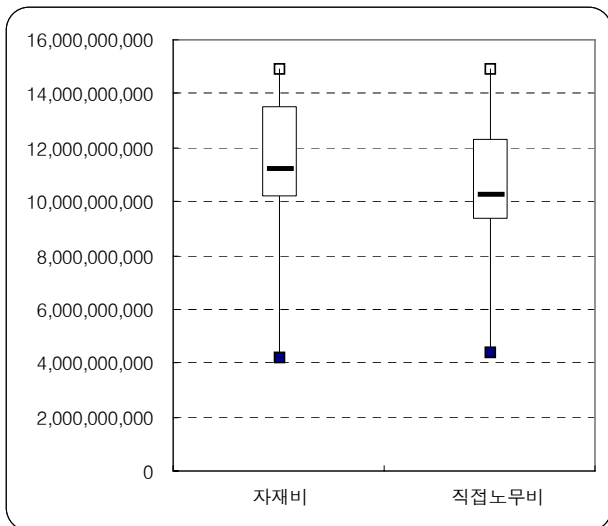


그림 7. 이상값 분석 Box-Plot 예

따라서 이상값의 존재에 대한 불확실성의 변수를 고려한 공사비 예측을 위한 이상값의 확인과 처리방법이 요구되어진다.

3. 이상값 처리 모델

3.1 이상값 처리과정



그림 8. 이상값 처리 과정

수집되는 사업비의 값들은 각각의 값에 대한 자료분포의 속성이 서로 다른 것으로 나타났기 때문에 사업비의 값을 단순 합산하는 산술평균값으로 사업비를 추정하는 데이터로 하게 될 경우, 각 항목별 사업비 속성을 충분히 반영하지 못하게 된다. 표준편차가 작은 분포에서 작은 차이는 유의한 차이가 될 수 있는 반면에 표준편차가 큰 분포에서는 어느 정도 이상의 차이가 아니면 유의한 차이가 될 수 없는 경우가 발생하기 때문이다. 또한 단순 산술평균값의 적용 방법은 각 사업비간의 복잡한 상관관계를 고려하지 못하는 점이다. 각 사업비간이 서로 독립적이거나 혹은 상관관계가 작을 경우 산술평균값을 이용하는 방법은 적절할 수도 있으나, 높은 상관관계가 내재된 경우에는 평가의 신뢰성에 문제가 제기된다. 서로 상관관계가 높을 경우에는 중복성이 높아지고, 또한 합산할 경우 특정 값이 상쇄될 수도 있기 때문이다.

이상값의 처리과정은 그림 8과 같이 이루어질 필요성이 있다. 특히 이상값은 상관성 분석과 밀접한 연관을 가지고 있다 할 수 있다. 일반적으로 데이터간의 상관성이 높은 값의 경우 이상값으로 판단되어 제거될 경우 전체 데이터에 미치는 영향이 매우 높아진다. 반대로 이상값이며 상관성이 낮은 값의 경우 제거되더라도 전체값에 미치는 영향은 낮아 특별히 이상값의 분석을 통한 이상값의 제거 과정을 거치지 않는다 하여도 큰 영향력은 없다고 할 수 있다.

3.2 모델 분석

본 연구에서 적용된 모델 사업대상은 상기 국민임대주택 사업비 산정에 적용된 공사비를 대상으로 이상값의 처리 모델을 적용하여 분석하였다.

표 1에서와 같이 수집된 데이터의 사업비 내역을 분석하고 이상값의 존재여부를 확인하기 위해 최대, 최소값을 구하고 각각의 분위수를 구하였다. 계산된 값을 이용하여 그림 9와 같은 결과를 도출하였으며 그 결과 각 내역별로 존

표 1. 공사비 항목별 공사비 분석(단위:원)

Project ID	공사비 항목									이상값 존재 빈도
	자재비	직접노무비	간접노무비	직접경비	기타경비	일반관리비	이윤	부가가치세	지급자재비	
P1010	9,799,348,771	9,329,092,219	849,361,704	1,053,696,185	953,777,491	803,628,000	771,463,000	958,068,000	2,447,226,000	
P1020	14,449,086,096	13,848,507,097	1,300,635,991	1,851,896,723	1,413,178,974	1,195,560,000	1,164,398,000	1,444,770,000	3,954,594,000	1
P1030	11,157,069,133	11,624,441,750	1,099,868,010	984,405,361	1,128,828,634	947,344,000	938,956,000	1,115,512,000	4,426,427,000	
P1040	12,193,468,131	11,629,736,805	1,093,336,073	1,626,353,694	1,188,543,733	1,014,516,000	991,865,000	1,219,207,000	3,755,274,000	
P1050	13,678,476,241	12,928,591,239	1,220,513,057	1,218,387,217	1,321,241,667	1,104,364,000	1,057,142,000	1,367,691,000	747,348,000	
P1060	11,230,060,930	9,513,749,516	894,228,308	1,084,375,469	1,035,309,233	868,077,414	804,323,402	1,122,912,838	743,685,344	
P1070	13,799,258,562	10,897,672,408	1,005,855,159	1,857,572,991	1,254,291,175	1,051,512,000	963,361,000	1,379,786,000	0	1
P1080	14,915,360,446	14,153,896,094	1,241,737,944	1,429,631,276	1,382,621,026	1,156,023,000	1,163,130,000	1,491,368,000	4,406,446,000	2
P1090	9,827,787,447	10,266,504,391	932,912,027	1,077,968,372	1,003,616,707	965,791,000	843,728,000	947,494,000	3,635,571,000	
P1100	14,572,680,864	14,938,317,816	1,354,018,341	1,517,358,113	1,482,735,581	1,413,917,000	1,227,845,000	1,420,807,000	4,887,218,000	6
P1110	12,971,309,578	13,301,252,440	1,249,442,030	1,415,069,465	1,303,839,835	1,263,224,000	1,098,726,000	1,256,086,000	4,226,351,000	
P1120	8,098,438,237	7,227,252,269	701,565,147	707,854,392	712,488,332	728,743,000	598,811,000	794,589,000	525,782,000	
P1130	10,981,639,377	9,678,594,625	906,893,831	939,745,743	1,030,205,213	983,646,000	812,844,000	1,098,034,000	834,133,000	
P1140	6,163,556,001	5,386,357,373	521,746,031	622,943,047	539,363,265	552,528,000	456,862,000	616,284,000	572,814,000	1
P1150	11,912,306,421	9,625,534,519	904,063,645	1,247,251,685	1,066,755,704	1,032,261,000	825,379,000	1,168,697,000	843,276,000	
P1160	10,972,025,330	10,037,595,535	910,289,196	1,607,045,093	1,054,924,770	1,028,317,000	868,387,000	1,067,514,000	975,829,000	
P1170	14,778,505,602	12,518,614,370	1,182,674,126	1,072,643,975	1,349,274,451	1,295,771,000	1,046,375,000	1,445,078,000	1,054,097,000	
P1180	10,162,449,780	9,243,241,011	865,478,451	1,181,058,715	968,853,899	936,801,000	790,630,000	1,016,135,000	927,128	
P1190	10,724,724,850	10,290,940,368	934,654,461	1,060,684,723	1,049,803,518	1,010,898,000	860,051,000	1,038,215,000	3,432,765,000	
P1200	11,944,603,097	10,304,648,681	922,091,543	1,171,458,401	1,127,923,107	1,068,688,000	875,632,000	1,169,009,000	0	1
P1210	10,424,777,740	9,307,300,997	862,469,887	1,564,742,132	997,059,222	846,349,596	813,386,481	1,042,477,774	2,008,182,323	
P1220	4,233,737,607	4,393,751,525	411,844,806	773,462,865	429,679,754	374,830,066	381,574,126	423,373,761	1,756,614,916	7
	자재비	직접노무비	간접노무비	직접경비	기타경비	일반관리비	이윤	부가가치세	지급자재비	
최소값	4,233,737,607	4,393,751,525	411,844,806	622,943,047	429,679,754	374,830,066	381,574,126	423,373,761	0	
25%	10,228,031,770	9,375,256,543	872,665,915	1,055,443,320	998,698,594	885,258,311	806,453,551	1,021,655,000	744,601,008	
50%	11,193,565,031	10,278,722,380	927,501,785	1,176,258,558	1,060,840,237	1,012,707,000	864,219,000	1,119,212,419	1,405,355,958	
75%	13,501,684,575	12,296,394,979	1,161,972,597	1,495,426,404	1,291,452,670	1,095,445,000	1,032,747,500	1,339,789,750	3,725,348,250	
최대값	14,915,360,446	14,938,317,816	1,354,018,341	1,857,572,991	1,482,735,581	1,413,917,000	1,227,845,000	1,491,368,000	4,887,218,000	
평균값	11,317,757,738	10,474,799,684	971,167,262	1,230,254,802	1,081,559,786	983,763,140	879,766,773	1,118,323,108	2,056,116,396	

표 2. 상관성분석

공사비 항목	자재비	직접노무비	간접노무비	직접경비	기타경비	산재보험료	안전관리비기초액	산업안전보건관리비	퇴직공제부금	고용보험료	국민건강보험료	국민연금보험료	일반관리비	이윤	손해보험료	부가가치세	지급자재비	지급자재비납품도
자재비	1.0000																	
직접노무비	0.9418	1.0000																
간접노무비	0.9397	0.9952	1.0000															
직접경비	0.6939	0.6532	0.6410	1.0000														
기타경비	0.9841	0.9803	0.9772	0.7089	1.0000													
산재보험료	0.9419	1.0000	0.9958	0.6539	0.9803	1.0000												
안전관리비기초액	0.4590	0.5597	0.5590	-0.0367	0.4914	0.5579	1.0000											
산업안전보건관리비	0.9822	0.9861	0.9831	0.6956	0.9951	0.9863	0.4980	1.0000										
퇴직공제부금	0.9724	0.9486	0.9348	0.7210	0.9736	0.9480	0.4308	0.9748	1.0000									
고용보험료	0.9536	0.9509	0.9527	0.7641	0.9749	0.9516	0.3583	0.9700	0.9572	1.0000								
국민건강보험료	0.1833	0.0088	-0.0177	0.0995	0.0941	0.0069	-0.2325	0.1045	0.2775	0.1102	1.0000							
국민연금보험료	0.1833	0.0088	-0.0177	0.0995	0.0941	0.0069	-0.2325	0.1045	0.2775	0.1102	1.0000	1.0000						
일반관리비	0.9390	0.9382	0.9353	0.6160	0.9558	0.9376	0.5397	0.9402	0.9247	0.9143	0.0379	0.0379	1.0000					
이윤	0.9582	0.9959	0.9906	0.7105	0.9903	0.9959	0.5114	0.9926	0.9654	0.9681	0.0469	0.0469	0.9458	1.0000				
손해보험료	0.2791	0.3373	0.3866	0.2372	0.3157	0.3415	0.1441	0.3381	0.2394	0.3660	-0.1539	-0.1539	0.1333	0.3253	1.0000			
부가가치세	0.9983	0.9355	0.9345	0.7001	0.9795	0.9357	0.4371	0.9804	0.9705	0.9543	0.2037	0.2037	0.9215	0.9529	0.3085	1.0000		
지급자재비	0.2970	0.5558	0.5388	0.2816	0.4172	0.5565	0.3873	0.4483	0.4054	0.4117	-0.2161	-0.2161	0.3709	0.5191	0.2378	0.2855	1.0000	
지급자재비납품도	-0.3574	-0.3469	-0.3540	0.0438	-0.3379	-0.3458	-0.5958	-0.3409	-0.2749	-0.2049	0.1097	0.1097	-0.4075	-0.3238	-0.1621	-0.3451	-0.0254	1.0000

재하는 이상값의 존재를 확인하였다. 공사비에서 가장 큰 비중을 차지하는 자재비와 직접노무비의 경우에 중앙값이나 평균값보다 상대적으로 이상값이 큰 값의 차이를 보이며 존재하였다.

공사비 항목간의 상관성을 알아보기 위해 표 2와 같이 상관성분석을 실시한 결과 공사비의 높은 비율을 차지하는 자재비, 직접노무비의 경우 상관성이 매우 높은 것으로 조사되었으며, 기타 공사비 항목의 경우에도 상호간의 상관성이 높게 나타나 이상값의 확인과 제거에 있어 단순통계를 이용한 비용예측결과에 높은 영향을 미칠것임을 알 수 있다.

3.3 분석결과

표 3의 결과와 같이 확인된 이상값을 제거한 후 나타난 통계값에서와 같이 평균값의 경우 모든 사업비 항목에서 1%이상의 차이를 보였으며, 이상값의 영향을 비교적 적게 받는 중앙값의 경우에도 사업비 항목 비율이 가장 높은 자재비와 상대적으로 높은 비중을 차지하고 있는 지급자재비의 경우 중앙값 마저도 차이를 보였다.

따라서 확인된 결과에 의해 단순통계값을 이용한 평균값을 적용하여 실제 사업비 분석을 통해 비용을 예측하거나 비용분석 적용시 정확한 분석에 영향을 미칠 수 있다.

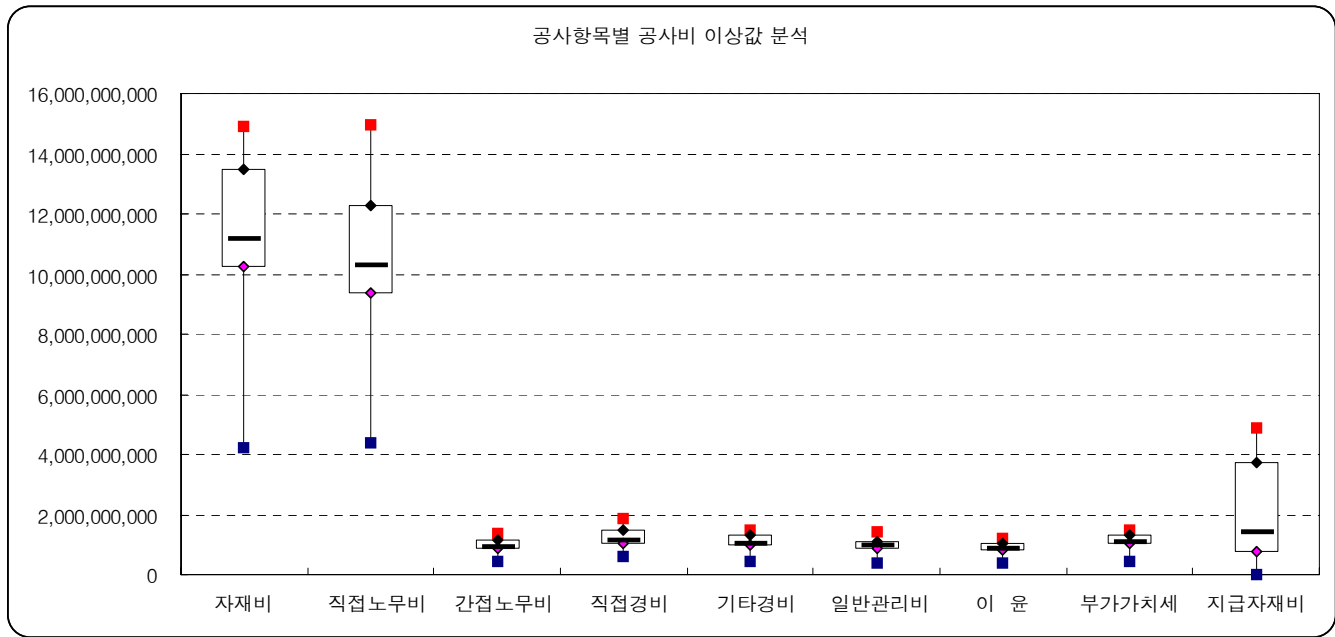


그림 9. 공사항목별 공사비 이상값 분석

표 3. 공사항목별 공사비 이상값 분석 결과(단위:원)

이상값 제거 전		자재비	직접노무비	간접노무비	직접경비	기타경비	일반관리비	이윤	부가가치세	지급자재비
	최소값	6,163,556,001	5,386,357,373	521,746,031	707,854,392	539,363,265	552,528,000	456,862,000	616,284,000	927,128
	25%	10,424,777,740	9,467,585,191	887,040,844	1,058,937,589	1,001,977,336	919,620,104	810,713,850	1,032,695,000	790,740,500
	50%	11,230,060,930	10,278,722,380	927,501,785	1,176,258,558	1,060,840,237	1,012,707,000	864,219,000	1,119,212,419	1,756,614,916
	75%	13,678,476,241	11,851,956,196	1,120,569,539	1,451,562,985	1,266,678,340	1,077,607,000	1,005,492,500	1,283,987,250	3,695,422,500
	최대값	14,915,360,446	14,153,896,094	1,300,635,991	1,857,572,991	1,413,178,974	1,295,771,000	1,164,398,000	1,445,078,000	4,426,427,000
	평균값	11,655,092,030	10,555,676,185	979,990,831	1,229,538,293	1,094,094,998	992,702,101	887,272,494	1,134,418,331	2,123,544,353
이상값 제거 후		자재비	직접노무비	간접노무비	직접경비	기타경비	일반관리비	이윤	부가가치세	지급자재비
	최소값	4,233,737,607	4,393,751,525	411,844,806	622,943,047	429,679,754	374,830,066	381,574,126	423,373,761	0
	25%	10,228,031,770	9,375,256,543	872,665,915	1,055,443,320	998,698,594	885,258,311	806,453,551	1,021,655,000	744,601,008
	50%	11,193,565,031	10,278,722,380	927,501,785	1,176,258,558	1,060,840,237	1,012,707,000	864,219,000	1,119,212,419	1,405,355,958
	75%	13,501,684,575	12,296,394,979	1,161,972,597	1,495,426,404	1,291,452,670	1,095,445,000	1,032,747,500	1,339,789,750	3,725,348,250
	최대값	14,915,360,446	14,938,317,816	1,354,018,341	1,857,572,991	1,482,735,581	1,413,917,000	1,227,845,000	1,491,368,000	4,887,218,000
	평균값	11,317,757,738	10,474,799,684	971,167,262	1,230,254,802	1,081,559,786	983,763,140	879,766,773	1,118,323,108	2,056,116,396
차이값	50%	36,495,898	0	0	0	0	0	0	0	351,258,958
	평균값	337,334,292	80,876,501	8,823,569	-716,508	12,535,212	8,938,961	7,505,721	16,095,223	67,427,957

4. 모델이용 건축비 분석결과

실적 공사비를 이용한 분석에 있어 데이터에 내포된 이상값의 처리 방안을 주택사업의 공사비 산정에 적용하여 이상값의 존재여부를 확인하고 이상값의 처리에 따른 공사비의 영향에 대한 처리결과는 다음 표 4와 같다.

표4의 결과와 같이 실적 데이터를 이용한 공사비 예측시 이상값은 상관성 분석과 밀접한 연관을 가지고 있으므로 전체 데이터에 미치는 영향을 고려하여 이상값의 제거 유무에 따라 평균값은 1~2%의 차이를 보여주는 것으로 분석되었다. 평균값의 차이값에서 보듯이 비록 큰 값의 차이

표 4. 이상값 처리 결과(단위:원)

항목	예정가		도급가	
	평균값	중앙값	평균값	중앙값
outlier 포함	2,796,716	2,734,996	2,258,931	2,216,350
outlier 제거	2,764,704	2,740,145	2,223,936	2,221,546
차이값	32,012	-5,149	34,995	-5,196

를 보이지 않았으나 분석된 값을 이용하여 공사비의 산정과 같은 경우에 사용될 경우 전체 사업비에 미치는 영향을

고려할 경우 무시될 수 없는 값의 차이라 할 수 있다.

5. 결론 및 추후 연구과제

본 연구에서는 실제 사업비 또는 실적 데이터를 이용하여 사업비를 분석하거나 공사비를 예측함에 있어 수집된 데이터에 내포되어 존재하는 이상값의 확인 방법과 제거에 대해 알아보았으며, 실제 이상값이 존재할 경우 발생할 수 있는 문제점에 대해서도 실제 사업비 분석을 통해 알아보았다. 또한 이상값이 존재할 경우 이상값의 처리를 통해 사업비분석에 있어 신뢰성과 타당성을 높이하고자 하였다.

그 결과 유사성 높은 실적 데이터의 수집이 상대적으로 어려운 건축사업비와 같은 경우 실제 사업비를 이용한 데이터의 경우 이상값이 존재할 확률이 높음을 확인할 수 있었다. 또한 이상값의 존재여부와 처리방안은 사업비 분석을 위해 사용되어지는 각 항목의 초기 레벨에서부터 적용되어야 신뢰성이 높아짐을 알 수 있었다. 특히 단순 평균값을 이용한 비교의 경우 이상값의 영향을 중앙값을 이용한 비교값보다 더 많은 영향을 받으므로 비용항목의 초기 레벨(낮은레벨)에서부터 적용되어야 할 필요가 있다.

이상값과 상관관계의 연관성에서 비용항목간의 상관성이 높은 값의 경우 이상값의 존재여부가 전체 사업비 분석에 매우 큰 영향을 보인 반면, 상대적으로 상관성이 낮은 비용의 경우 비록 이상값이 존재하여도 이상값의 존재가 사업비 분석에 있어 큰 영향을 보이지 않았다. 따라서 신속한 비용분석의 처리분석을 위해서는 주성분의 개념에 입각한 상대적으로 상관성이 높으며 사업비 비율이 높은 항목을 분석하여 사업비 분석에 적용하여도 큰 오류가 없다고 할 수 있다.

본 연구에서는 이상값의 확인과 제거를 위해 단순하며 이용이 간편한 방법을 적용하였으나, 더 신뢰성 높은 분석을 위해서는 F-산포도 검증이나 로버스트 분석 등의 건축비 분석의 적용방안이 요구되어 진다.

참고문헌

1. 성웅현(2005), "사업성 종합지수를 이용한 기술의 사업성 상대 등급 평가에 관한 연구", 지식경영연구 제6권 제2호
2. 남경두(1995), "Performance Comparison of Neural Networks and Univariate Linear Regression Models in the Presence of an Outlier", 건국대학교 경제경영연구소 상경연구, 제20집
3. 신관호(2006), "국가간 위험분담의 경제적 요인 분석", 금융학회지 제11권 제1호
4. 이명재, 이진용, 김규범, 원종호(2005), "국가지수 관측소 측정자료의 이상값 분석", 한국지하수토양환경학회지 Vol. 10, No 1.
5. 성웅현, 양동우(2005), "중소벤처기업의 기술가치평가를 위한 할인을 추정에 관한 연구", 지식경영연구 제6권 제1호
6. 박효열, 송용식, 김선국(2003), "공동주택 사업비 산정을 위한 표준DB 연구", 대한건축학회논문집 구조계 19권 6호
7. 박효열, 송용식, 김선국(2003), "공동주택 실적공사 표준 DB를 이용한 COST MODEL", 대한건축학회논문집 구조계 20권 5호