

# 6시그마 프로젝트에서의 통계기법의 오용

안병진

건국대학교 응용통계학과

## The misuse of statistical methods in Six Sigma projects

Byoung Jin Ahn

Dept. of Applied Statistics, Konkuk University

**Key Words** : Six Sigma, Statistical thinking, Statistical methods

### Abstract

Many companies in Korea are paying close attention to Six Sigma to get better business results. And there are numerous success stories on the use of statistical methods in solving problems, improving processes and cutting costs. However, there are many cases in which statistical methods are misused or inappropriately applied. In this paper, the pitfalls to avoid in using statistical methods are discussed. Also the role of statistics in Six Sigma is examined.

## 1. 서론

6시그마 경영은 고객만족을 위해 모든 프로세스에서 과학적 통계기법을 적용하여 높은 품질수준을 이룩함으로써 경영성과에 기여하는 경영혁신활동으로 볼 수 있다. 우리나라에서는 1990년대 후반부터 제조업을 중심으로 도입하기 시작하여 많은 기업이 추진하고 있으며 서비스, 금융, 공공부문까지 그 적용 영역이 확대되고 있다.

6시그마 경영이 확산됨에 따라 통계적 사고(statistical thinking)에 바탕을 둔 통계기법의 활용은 점차 증가 추세에 있다. 또한 정보화가 진척됨에 따라 기업은 많은 데이터를 확보하게 되고 이에 따른 데이터 분석

의 필요성 역시 증가하고 있다.

김재주(1981), 박성현 외 1인(1991), 박성현(2001)등에서도 알 수 있는 바와 같이 과거에도 통계기법이 품질문제해결을 위하여 중요한 역할을 해왔지만 통계마인드의 부족, 통계 소프트웨어의 부족, 통계전문인력의 부족 등으로 통계기법의 활용이 어려움을 겪어 왔던 것도 사실이다. 다시 말해서 통계는 너무 어렵다는 선입견 때문에 잘 사용하지 않는 경향이 있었고, 데이터를 얻어도 그 데이터가 무엇을 의미하는지를 잘 모르고, 데이터를 보고 어떠한 조치를 취해야 하는지를 모르는 경우가 많았다. 뿐만 아니라 데이터는 통계분석을 통해 지식으로 전환될 때 비로소 의미를 갖게 됨에

도 불구하고 제대로 활용되지 못하고 사장되는 경우도 많았다.

과거에는 통계기법의 활용이 활발하지 않다는 것이 문제였지만 점차로 통계기법의 활용이 증가함에 따라서 통계기법의 오용이 문제가 되고 있다. 통계기법은 문제해결을 위한 수단이지 목적이 아님에도 불구하고 형식적으로 적용하거나 획일적으로 적용하기도 하고 잘못 적용하기도 한다. 예를 들어 QC 7가지 도구만으로 충분히 해결될 문제를 통계로 포장하는 경우도 있고, 고급통계기법이 필요한 경우에 지식의 부족으로 사용하지 못하는 수도 있다.

기업에서 6시그마 경영을 추진할 때 도입 초기에는 통계의 오용이 큰 문제가 안될지 모르겠지만 통계기법에 대한 잘못된 실수는 조직 내에서 반복되고 확산되는 경향이 있기 때문에 통계기법의 활성화를 위축시키지 않는 범위 내에서 초기에 차단할 필요가 있다. 통계기법의 오용은 통계에 대한 신뢰성을 손상시키고 그 결과 6시그마 경영의 저해요인 이 될 수 있다. 이 논문에서는 6시그마 프로젝트를 수행하는 과정에서 통계기법을 사용할 때 범하기 쉬운 실수 및 유의할 점들을 정리해 보고자 한다.

## 2. 6시그마 경영에서의 통계학의 역할

이 절에서는 통계오용의 문제를 다루기 전에 6시그마 경영의 핵심요소인 측정의 중요성, 통계적 사고, 문제해결기법, 고객만족 등과 통계학은 어떠한 관계를 갖고 있는지를 알아보고, 6시그마 경영에서의 통계학의 역할에 대하여 검토해보기로 한다.

Stein(2001)은 “측정이야말로 기업의 비즈니스를 관리하기 위해 사용하는 필수도구로서, 6시그마는 측정의 중요성을 인식하는 최근의 관리철학이다.”라고 측정의 중요성을 강조하였다. 기업이 자신들의 프로세스를 숫자로 나타낼 수 없다면 프로세스 자체를 이해하지 못하고 있는 것이고 그 결과 프로세스를 관리할 수 없게 된다. 기업의 전략과 방침에 연계된 성과측정기준에 따라 프로세스를 측정하는 것이야말로 품질을 개선하고, 고객만족을 증진시키는 핵심적 수단이다. 다시 말해서 믿을 만한 데이터를 적기에 필요한 사람에게 제공할 수 있다는 것은 기업경쟁력에 매우 중요한 요소이며, 이를 위해서는 통계학이 필요하게 된다. Britz 외 3인(1997)에서 볼 수 있듯이 통계적 사고는 다음과 같은 기본원리에 바탕을 둔 사고체계이다.

- 모든 일은 서로 연관된 프로세스를 통하여 이루어진다.
- 모든 프로세스에는 변동(variation)이 존재한다.
- 변동을 이해하고 줄이는 것이 관건이다.

“모든 일은 프로세스를 통하여 이루어진다”는 말은 결과를 완벽하게 하기 위해서는 과정이 완벽해야 한다는 의미이며, 이는 품질운동의 바탕이 되는 말이다. 또한 ‘서로 연관된’이라는 문구는 부분 최적화의 위험성을 지적하고 있다. 다시 말해서 각 프로세스는 서로 연관되어 있기 때문에 한 프로세스의 최적화는 전체의 최적화와 일치하지 않을 수 있다. 이러한 이유로 6시그마 경영은 고객이 중시하는 CTQ(critical to quality)를 우선 파악하고, CTQ와 관련된

모든 프로세스를 연결하여 이를 개선하기 위한 프로젝트를 수행하자는 것을 기본 전략으로 삼고 있다.

“모든 프로세스에는 변동이 존재한다”는 말은 현실세계를 보는 관점을 나타내 준다. 품질문제는 물리적 법칙과 같이 결정적인 것이 아니고 다양하고 무질서한 형태로 나타나는 경우가 많기 때문에 통계학이 유용하게 활용된다.

“변동을 이해하고 줄이는 것이 관건이다”는 말은 행동의 중요성을 나타내 준다. 프로세스의 변동은 대부분 데이터의 형태로 표현되어질 수 있고, 데이터를 제대로 이해하기 위해서는 통계기법이 필요하게 된다. 변동을 줄이기 위한 올바른 조치는 변동에 대한 올바른 이해가 선행될 때 비로소 가능하게 된다.

여기서 변동을 줄이자는 것은 프로세스의 변동을 줄이자는 것이지 더 나은 변화를 위해 조직이 감수해야하는 변동까지 줄이자는 것은 아니다. Pyzdek(2001)에서 언급된바와 같이 6시그마 경영에는 역설이 있다. 프로세스의 변동을 줄이기 위해서는 조직에서의 변동이 먼저 확대되어야 한다는 것이다. 다시 말해서 프로세스의 변동을 줄이기 위해서는 사고의 혁신이 필요하며, 사고의 혁신은 창의성을 통해 이루어지고, 창의성은 프로젝트 팀으로 하여금 새로운 시도를 용인하는 기업분위기가 조성될 때 가능해진다. 새로운 시도는 조직에 새롭고 다양한 변동을 야기하고, 심지어 실패하는 경우 새로운 혼란과 비용을 야기할 수도 있다. 예를 들어 여유 인력이 부족한 기업이 6시그마 경영을 추진하는 경우 프로젝트 수행자는 프로젝트에 몰입할 수 있는 시간적 여유가 없게되고, 그 결과 프로젝트는 형식적으로 진

행되기 쉽고 변화와 성과는 미미하게 된다. 또한 프로젝트 때문에 원래하고 있던 업무에 손실이 생겨 프로젝트 성과보다 파악 안되고 있는 업무손실이 더 클 수도 있다. 조직의 변동이 확대될 수 있을 때 프로세스의 변동이 축소될 수 있다는 것이 6시그마 경영의 역설이다. 따라서 6시그마 경영을 도입하는 과정에서 일어나는 변화로 인한 충격과 조직 안에서의 변동을 감당하기 어려운 기업은 6시그마 경영을 추진할 때 강약과 완급을 조절하여야 한다.

통계적 사고 이외에도 6시그마 경영에서 특히 강조되는 다음과 같은 마음가짐 역시 통계학과 밀접한 관계를 갖고 있다.

- 문제가 발생하기 전에 일을 처음부터 올바르게 하자.
- 결과보다는 원인을 관리하자.
- 예감과 경험을 뛰어넘어 근거 갖고 일 하자.

일을 처음부터 제대로 하면 문제도 발생하지 않고 문제해결 비용도 들어가지 않겠지만 현실적으로 그렇지 않은 경우가 많다. 만약 어떤 문제가 추후에 발생할 것인가를 사전에 파악하고 예측할 수 있다면 문제해결 비용을 대폭 줄일 수 있을 것이다. 이러한 경우 데이터 또는 실험에 바탕을 둔 통계기법은 예측기능을 갖고 있기 때문에 중요한 역할을 하게 된다.

프로세스의 결과에 영향을 주는 원인을 파악하고 관리할 수 있다면 문제발생을 예방할 수 있고 비용도 줄일 수 있겠지만 원인을 관리하기 위해서는 전제조건이 필요하다. 다시 말해서 원인과 결과 사이의 함수관계를 알아야 원인을 관리할 수 있다. 이러한 함수관계를 파악하기 위해서도 예를

들면 회귀분석과 같은 통계기법이 중요한 역할을 한다.

효율적인 개선이 이루어지기 위해서는 경험과 예감에만 의존하는 것이 아니라 데이터에 기초한 통계기법을 이용하여 근거를 갖고 의사결정 하여야 한다. 데이터 없이 의사결정을 하면 불확실성을 극복할 수 없기 때문에 시행착오를 유발하기 쉽고, 그 결과 많은 문제해결 비용이 발생하게 된다.

6시그마 경영에서 통계기법이 활용되는 분야로는 고객요구조사, 제품설계 및 개발, 공정능력평가, 제품수명검사, 통계적 공정관리, 측정시스템분석, 검사 등이 있으며, 문제해결을 위하여 프로젝트를 수행하는 과정에서 통계기법의 역할을 정리해 보면 다음과 같다.

- 문제를 구체적으로 파악하고 정의하는데 도움을 준다.
- 새로운 데이터가 필요한 경우 어떤 데이터를 수집할 것인가를 결정하여 최소의 비용으로 신뢰할 수 있는 데이터를 수집할 수 있도록 계획한다.
- 현재 갖고 있거나 새로 얻어진 데이터에 대한 타당성을 평가한다.
- 데이터를 이해하기 쉽고 의미 있는 형태로 정리 요약한다.
- 데이터를 통계분석을 통하여 유용한 지식으로 전환한다.
- 데이터를 바탕으로 불확실성을 극복함으로써 올바른 의사결정을 위한 토대를 마련해준다.

6시그마경영의 핵심인 고객만족을 위해서도 통계적 기법은 중요한 역할을 한다. 고객만족을 위해서는 고객의 소리 (voice of

customer : VOC) 가 필수적 요소이며, 고객의 소리는 대부분 데이터를 통해 들을 수 있다. 또한 고객에 대한 올바른 지식은 데이터에 대한 통계적 분석을 통해 이루어진다. 이와 같이 통계학을 토대로 고객을 제대로 이해할 수 있을 때 비로소 기업은 고객에 대한 지식을 구성원들과 공유할 수 있게 되고 고객가치를 위한 행동을 취할 수 있게 된다.

지금까지 살펴본 바와 같이 6시그마 경영이 진척되어감에 따라 통계학은 단순한 문제해결을 위한 방법론에서 한 걸음 나아가 품질문제를 이해하고 의사 소통할 수 있는 언어로서 자리매김해 나가고 있다.

### 3. 프로젝트를 수행할 때 통계가 빠지기 쉬운 함정

6시그마 프로젝트를 수행하는 과정에서 통계기법은 교육과 훈련을 통하여 제대로 이해하는 사람들이 적재적소에 올바르게만 사용한다면 문제해결에 있어 강력한 수단이 될 수 있다. 그러나 통계기법을 형식적으로 사용하거나 왜곡되게 사용한다면 잘못된 결론에 도달할 수도 있고 사용하지 않는 것만 못하다.

6시그마 프로젝트는 대부분 정의단계, 측정단계, 분석단계, 개선단계, 관리단계라는 5단계의 정형화된 절차를 따르게 된다. 지금부터 각 단계에서 널리 사용되는 주요 통계 기법에 대하여 저지르기 쉬운 실수 및 유의사항을 정리해 보기로 한다. 일반적인 통계오용에 대한 사례들은 Campbell(1974), 한국통계학회(1991), 박 성현(1996) 등에서 찾아볼 수 있다.

### 3.1 성과 측정을 위한 데이터 수집

“측정 없이는 개선도 없다.” 는 말과 같이 6시그마 경영에서 측정은 중요한 역할을 하지만 쓸모도 없고 필요하지도 않은 측정을 계속하는 경우 측정관료주의가 생겨나 ‘측정을 위한 측정’ 이 되고, 소위 DRIP (data rich information poor) 상태에 빠질 수 있다. 이를 극복하기 위해서는 체계적인 측정 관리가 필요하며 이 때 유의할 점들을 정리해 보기로 하자.

- 데이터 수집이 처음부터 제대로 이루어지기 위해서는 모든 사람에게 동일한 의미를 제공해 주고, 일관성과 신뢰성을 확인시켜주며, 측정범위를 나타내 주는 운영정의(operational definition)를 분명히 할 필요가 있다.
- 데이터 수집계획을 수립할 때는 목적이 구체적이고 명확해야 하며, 예를 들면 다음과 같은 기준에 의해 데이터의 특성을 사전에 정리해 놓는 것이 바람직하다.
  - 원인 데이터, 결과 데이터
  - 제어 가능한 변수, 잡음(noise), SOP(standard operating procedure)
  - 이산형 데이터(discrete data) , 연속형 데이터(continuous data)
  - 망목특성(nominal-the-best), 망소특성(smaller-the-better), 망대특성(larger-the-better)

또한 추후 분석을 위해서는 가능하면 원 인 데이터와 결과 데이터를 함께 구하고, 데이터에 층별을 위한 중요요소를 포함시키는 것이 편리하다.

· 데이터 수집의 핵심은 적은 비용으로 목적에 맞고 믿을 만한 데이터를 수집하는 것이다. 이를 위해 유의할 점들을 정리하면 다음과 같다.

- 연구대상인 모집단은 명확히 정의되어야 하며 조사과정에서 왜곡되어서는 안 된다. 예를 들어 고객 만족도 조사를 할 때 잘못된 고객들로부터 정보를 얻는 것은 치명적일 수 있다.
- 가능하면 표본설계를 이용하여 데이터의 양을 줄일 수 있도록 한다.
- 구체적인 지침과 데이터 수집 방법, 시기, 장소 등을 명확히 하여 데이터의 일관성을 유지해야 한다.
- 체크시트나 조사표를 사전에 준비하여 데이터수집을 용이하게 해야한다.
- 데이터 수집규모가 큰 경우에는 예비 조사를 실시하여 사전에 수집과정의 타당성을 점검해 보아야 한다.

· 표본오차(sampling error)는 표본의 크기가 증가하면 감소하지만, 비표본 오차(non-sampling error)는 오히려 증가할 수도 있다. 다음과 같은 비표본 오차는 철저한 관리를 통하여 줄여야 한다.

- 적절하지 않은 표본추출 방법을 사용한다든지, 혹은 표집틀(sampling frame)이 불완전하여 조사단위의 일부가 누락 또는 중복되어 발생하는 오차
- 응답자의 부재, 부정확한 응답 또는 무응답(non-response)에 의한 오차
- 조사원의 훈련부족 또는 조사원과 응답자의 상호관계에서 발생하는 조사 현장에서의 오차

- 조사원이 잘못 기재 하거나 기록된 설문지를 분석하기 위하여 처리하는 과정에서 발생하는 오차
- 고객요구조사나 만족도 조사를 위해 설문지를 활용하는 경우 다음과 같은 점에 유의하여야 한다.
  - 응답자의 기본적인 특성을 나타내는 성별, 연령, 직업 등과 같은 배경변수를 조사항목에 포함시킴으로써 분석할 때 동질적 집단의 특성을 파악할 수 있게 한다.
  - 사전에 무응답자의 처리기준과 처리 방법을 구체적으로 결정해 둔다.
  - 문항을 제대로 작성하여야 한다. 예를 들면 질문에 모호한 단어의 사용, 두 개의 질문을 동시에 하고 있는 이중 질문, 질문자체가 응답자에게 특정한 방향을 암시하는 질문, 가정을 포함하고 있는 질문 등은 피해야 한다.
  - 조사결과를 전산처리하기 편리하도록 문항 및 응답항목을 숫자로 부호화 하는 것이 좋다.
- 사전에 데이터가 제대로 수집되도록 노력하여야 하지만 분석하고 활용하기 전에 수집된 데이터를 평가할 필요가 있다.
  - 신뢰하기 어려운 데이터를 바탕으로 분석한 후 이상한 결과가 나오면 통계가 잘못되었다고 하는 경우가 있다. 데이터는 거짓말을 하지 않으며 나쁜 데이터도 없다. 사람의 나쁜 의도와 태만이 있을 뿐이다.
  - 특별한 품질 개선 활동 없이도 작업

현장에서 데이터의 신뢰성만 향상 시키면 품질이 좋아지는 경우가 종종 있다.

- 주요 잠재 원인 변수 X에 대한 측정시스템분석은 반응변수 또는 결과변수인 Y에 대한 측정시스템분석 만큼 중요하다.
- 잡음이 측정결과에 영향을 주는 경우, 어렵지만 고정시키거나 영향을 줄이기 위한 노력이 필요하다.

### 3.2 가설검정

통계적 가설검정은 서로 상반된 주장이나 추측을 가설로 설정한 후, 표본데이터를 통하여 어느 것이 보다 타당한지를 가려내는 절차이다. 6시그마 프로젝트를 수행할 때는 개선 전후의 성과를 비교하거나 예를 들면 정규성 검정과 같이 데이터의 타당성을 검토할 때 사용되고, 주로 분석단계에서 근본원인에 대한 타당성을 입증하고자 할 때 많이 활용된다. 가설검정을 할 때의 유의사항을 정리해 보면 다음과 같다.

- 가설검정이 적절하지 않은 상황에서 함부로 사용해서는 안 된다. 예를 들면 프로토타입(prototype)에 대한 데이터를 바탕으로 서로 다른 환경인 미래 제품 성능에 대한 가설검정을 하는 것은 위험하다.
- 사용되고 있는 데이터의 타당성을 확인하여야 한다. 예를 들면 모분산을 모르는 경우 모평균에 대한 검정을 하고자 할 때 무조건 t-검정을 사용하는 경향이 있는데 만약 데이터가 정규성에서 많이 벗어나 있다면 비모수 검정이 바람직 할 수도 있다.

- 가설을 제대로 설정해야 한다. 귀무가설과 대립가설을 거꾸로 설정하여 엉뚱한 결론을 내리는 경우도 있다.
- 상황에 적합한 검정방법을 택해야 한다. 예를 들면 배제해야 할 기타변동요인이 존재하는 경우 두 모집단간의 비교 (2-sample t) 보다는 짝지워진 비교 (paired t)가 효율적이다.
- 비용과 검출력을 고려하여 적절한 표본의 수와 데이터 수집 방법을 결정하여야 한다. 예를 들어 정규성 검정을 할 때 표본의 수가 너무 적으면 곤란하고 통상 30개 이상이 적절하다. 이 경우는 표본의 수가 많더라도 데이터 수집기간이 장기간인 경우는 중심이동을 초래하게 되어 정규분포를 따르지 않게 되는 수가 많다.
- 유의확률 p값과 유의수준을 단순 비교하여 기각여부를 기계적으로 판단해서는 안 된다. 예를 들어 p값이 유의 수준보다 약간 크게 나온 경우는 추가적인 데이터를 확보하여 검출력을 향상시킨 후 유의성에 대한 판단을 내리는 것이 바람직 할 수도 있다. 또한 유의 수준도  $\alpha=0.05, 0.01$  만 사용할 수 있는 것이 아니고 제 1종 오류로 인한 손실이 심각하지 않고 제 2종 오류로 인한 손실이 크다고 여겨지는 경우에는  $\alpha=0.1$  을 사용할 수도 있다.
- 기각여부를 결정한 후 해석에 유의하여야 한다. 예를 들어 귀무가설을 기각하지 못했을 때는 귀무가설이 옳다는 뜻이 아니

라 아무런 확신도 할 수 없다는 의미가 가깝다. 또한 통계적으로 유의하다고 해서 반드시 실제적 차이를 의미하는 것은 아니다. 제대로 판단하기 위해서는 통계적 결과와 전문 기술적 지식을 함께 고려하여야 한다.

### 3.3 공정능력 분석

공정이 관리상태 하에 있을 때, 그 공정에서 생산되는 제품의 품질변동이 어느 정도 인가를 나타내는 양이 공정능력이다. 공정능력에 대한 이해는 각종손실과 고객불만을 감소시키는 지름길이다.

공정능력이 올바르게 평가되기 위해서는 표본을 바르게 얻었는지, 측정시스템에 문제가 없는지를 검토하고, 데이터의 정규성을 검토하여 공정능력지수가 제대로 산출된 것임을 입증하여야 한다. 만약 데이터가 정규분포를 따르지 않으면 실제 공정능력보다 과대 또는 과소 평가 될 수 있으므로 데이터 변환이나 일반화된 공정능력지수의 사용을 검토하여야 한다.

공정능력 지수의 산출만을 위한 공정능력 평가가 되어서는 안 된다. 공정능력 (capability)과 공정성능(performance)평가를 통해 관리의 문제인지 또는 기술의 문제인지를 확인하여 프로젝트 진행방향을 올바르게 설정할 수 있는 기회로 삼아야 한다. 또한 공정능력 평가는 측정단계 뿐만 아니라, 관리단계에서도 목표의 달성수준을 측정하고 추후과제에 대한 가능성을 판단하는데도 사용될 수 있다.

### 3.4 상관분석

프로젝트를 수행할 때 변수들간의 상호

관련성에 관심을 갖게 되는 경우가 흔히 있다. 이 경우 두 변수 사이에 선형관계가 존재하는가를 분석하는 것이 상관분석이다. 상관분석을 위해서는 산점도와 상관계수를 함께 살펴보아야 하며, 산점도를 볼 때는 점들이 나타나는 형태뿐만 아니라 이상한 데이터가 있는지 살펴보아야 한다. 이상치(outlier)가 발견되면 원인을 규명하여 수정해 주도록 한다. 또한 점들이 몇 개의 집단으로 층별 되는지 살펴본다. 이 경우 몇 개의 모집단으로부터 데이터가 섞여 있는 경우이므로 해석에 유의하여야 한다.

표본상관계수를 이용하여 무 상관검정을 할 때 해석에 유의하여야 한다. 표본상관계수 값이 0에 가까운 값이 나온 경우 선형관계가 약하다는 뜻이지 두 변수 사이에 아무런 관계도 없다는 뜻은 아니다. 다시 말해서 두 변수 사이에 뚜렷한 곡선관계가 있는 경우에도 표본상관계수 값은 0에 가까운 값이 나올 수 있다. 무 상관검정결과 두 변수 사이에 상관관계가 있다는 것이 입증되었다고 해서 두 변수 사이에 반드시 인과관계가 있다는 것은 아니며, 두 변수 모두에 영향을 주고 있는 제3의 변수가 숨어있을 수도 있다. 상관분석을 할 때도 통계적인 의미가 있는 경우 현실적으로도 의미 있는지 전문 기술적 지식을 바탕으로 함께 검토해 보아야 한다.

### 3.5 회귀분석

회귀분석은 반응변수 Y와 설명변수 X들간의 관련성을 수학적 모형을 가정하고, 데이터로부터 이를 추정하여 분석하는 통계기법이다. 회귀분석은 Y와 X들간의 관계식을 도출하여 이를 예측하는데 활용하거나 Y에

대한 X의 역할을 설명하는 데 사용된다. 또한 Y에 중요한 영향을 주는 X를 선별하거나 바람직한 Y를 위한 X의 관리기준을 선정하는데도 사용된다. 회귀분석을 할 때 유의할 사항을 정리하면 다음과 같다.

- 산점도도 그려보지 않고 단순회귀분석 하 것은 피해야 한다.
- 두 변수의 관계가 성질상  $X=0$  일 때는  $Y=0$  이라는 것이 명백할 때가 있다. 이 경우 원점을 지나는 회귀를 사용하는 것이 타당하다.
- 1차 다항 회귀모형이 적절하지 않은 경우 2차 다항 회귀모형이 타당한지 검토해 볼 필요가 있다.
- 데이터가 층별 변수를 포함하고 있는 경우에는 가변수(dummy variable)를 모형에 포함시켜 분석하는 것이 바람직하다.
- 실제현장에서는 CTQ의 형태로 주어지는 반응변수가 여러 개인 경우가 많다. 이 경우 각 반응변수별로 최적조건이 일치하지 않게 되면 최적조건을 결정하기 어렵게 된다. 다시 말해서 최적조건이 반응변수별로 다르게 되는 모순현상이 발생하게 된다. 이러한 경우 X들의 값을 조금씩 변화시켜가면서 Y들의 값이 어떻게 변화하여 가는가를 모니터링하여 최적조건을 구하거나, 흔히 쓰는 방법은 가장 중요한 반응변수에 대한 회귀식을 목적함수(objective function)로 놓고 나머지 반응변수에 대한 회귀식을 제약조건(constraints)으로 보아 최적화하는 방법



- 이다. 만일 해가 없는 경우는 제약조건을 완화하여 최적화하며, 구한 최적조건에 대해서는 재현성 실험을 통해 개선정도를 확인하여야 한다.
- 반응변수의 수가 하나이더라도 중심치 이탈의 문제와 산포의 문제를 동시에 갖고 있는 경우가 있다. 이러한 경우에는 흔히 각 실험 조건에서 반복측정을 통하여 여러 관측값을 구한 후 평균과 분산을 구하여 이 두 가지를 동시에 최적화하기도 한다.
  - 설명변수의 수가 많은 경우에는 변수선택을 하여야 한다. 만약 설명력이 약한 변수가 모형에 포함되면 모형이 복잡해지고 예측력이 떨어진다. 반면 중요한 변수가 누락 되면 분석의 목적을 달성하기 어려워진다. 따라서 상황에 적절한 판단기준을 택하여 설명변수를 선택함으로써 최적의 회귀식을 도출하여야 한다.
  - 회귀분석을 할 때 오차에 대하여 정규성, 독립성, 등분산성 등을 가정하는 경우가 많으며, 이러한 경우 가정이 충족되어야만 분석결과가 의미를 갖는다. 따라서 잔차 분석을 통하여 오차가 가정을 만족하는지 검토하는 것이 바람직하다.
  - 반응변수의 값 중에는 측정상의 과오나 실험상의 과오 또는 데이터를 입력할 때의 실수 등으로 인하여 그 값이 지나치게 크거나 작아서 이상치가 되는 경우가 있다. 이상치가 발견되면 적절한 대응책이 필요하며, 경우에 따라서 재 실험하거나 또는 제거시킨다.
  - 회귀분석의 결과는 소수의 측정값에 의해서도 크게 영향을 받을 수 있다. 이와 같은 영향 관측값(influential observation)을 식별함으로써 데이터의 특성을 파악할 수 있을 뿐만 아니라 분석결과를 해석하는데 도움을 얻을 수 있다.
  - 다공선성(multicollinearity)이란 중회귀 모형에서 여러 개의 설명변수들이 선형결합 관계를 맺고 있는 상태를 말한다. 다공선성이 존재하는 경우 설명변수들 사이에 설명효과가 분명치 않게 되고, 최소제곱법에 의해 추정되는 회귀계수들을 신뢰할 수 없게 된다. 계획된 실험을 통하여 얻은 데이터에는 다공선성이 존재하는 경우가 드물지만, 사무간접부문에서 관측된 데이터를 바탕으로 회귀 분석할 경우 다공선성이 있을 수 있으므로 유의하여야 한다.
  - 예측을 위하여 회귀선을 사용할 때 데이터 중심에서 지나치게 멀리 떨어진 점에서의 예측은 신뢰하기 어렵다.
  - 예전의 데이터에 근거를 둔 회귀직선이 반드시 새 데이터에도 맞다고 보기 어렵다.
  - 회귀직선을 표본이 추출된 모집단과 다른 모집단에서 예측하는데 사용하면 위험할 수 있다.
  - 회귀분석은 변수들간의 관계를 설명하는 것이지만 어떤 변수가 다른 변수에 원인을 제공하는 인과관계를 설명하는 것이 아님에 유의하여야 한다.
-

### 3.6 실험설계

실험설계란 시스템의 성격을 알기 위하여 실험을 수행할 때, 과연 어떤 식으로 실험을 수행해야 효율적인 데이터 수집이 가능한지, 또 어떻게 데이터를 처리해야 얻어진 데이터에 대한 분석 및 해석이 적절한지 등을 다루는 통계기법이다. 실험설계는 프로젝트를 수행하는 과정에서 잠재적 원인들 중에서 핵심적 소수(vital few)를 선별하고 선별된 핵심적 소수 X들 사이의 교호작용(interaction)을 파악하고자 할 때 사용된다. 또한 파악된 원인 X와 결과 Y 사이의 관계를 규명하고 Y를 최적화시키는 X의 조건을 결정할 때도 유용하다. 실험을 할 때 유의사항을 정리해 보면 다음과 같다.

- 실험설계를 적용하기 적절한 상황인지부터 먼저 판단해야 한다. 예를 들어 공정이 관리 상태에 있지 않다면 통계적 공정관리와 같은 시도가 먼저 이루어져야 한다. 공정이 관리 상태가 된 후에야 비로소 근본 원인을 규명하고, 산포를 줄일 수 있는 공정조건을 찾기 위한 실험이 효과적 일 수 있다.
- 비효율적 실험의 대표적인 예로서 OFAT (one factor at a time)와 ‘한번의 실험으로 모든 것을 얻으려는 시도’를 들 수 있다. 최적조건을 찾기 위해 나머지 요인은 고정시키고 한 요인씩 변화시켜 가면서 최적조건을 찾아가는 OFAT는 실제 반응 표면의 모양이 산봉우리 모양을 띄고 있지 않거나 요인간에 교호작용이 있는 경우에는 엉뚱한 결론에 도달할 수 있다. 반면에 ‘한번의 실험으로 모든 것을

얻으려는 시도’ 역시 위험하다. 왜냐하면 실험을 하는 과정에서 실험결과에 따라 최적조건이 있으리라 기대되는 영역이 변할 수도 있고, 주요 요인이 탈락되거나 추가될 수도 있으며, 실험의 목적 자체가 바뀌어질 수도 있다. 따라서 적정 규모의 실험을 축차적으로 실시하는 것이 위험 부담이 적고 효율적이다.

- 실험을 수행하는 팀원간에 또는 관련 부서와 실험의 범위, 목적, 접근방식, 역할 분담 일정, 제한요소 등에 대한 합의가 이루어진 후 실험을 수행해야 효율적이다.
- 실험의 목적에 따라 적절한 설계방법을 선택해야 한다. 예를 들어 선별실험이라면 일부실험시범을 구상해야 할 것이고, 최적화 실험이라면 반응표면 분석 등과 같은 좀더 정밀한 실험이 적절할 것이다.
- 복잡한 실험과 고급분석방법은 피하는 것이 좋다. 실험이 복잡하면 결과해석이 어려워 지고, 적절한 통계분석이 없는 경우도 있다. 현장에서는  $2^3$ 요인배치,  $3^2$ 요인배치만으로도 충분한 경우가 많다.
- 요인, 수준, 반응변수의 선택에 있어서 문제에 대한 기술적 지식을 활용해야 하며, 결과해석 역시 통계분석에만 의존할 필요가 없다.
- 실험을 할 때 선택된 요인이외에 다른 조건들은 될 수 있는 대로 동일하게 유지시켜야 하는 것이 원칙이지만 관측값에 영향을 줄 수 있는 외부요인이 있을 수

있다. 외부요인 중 피할 수 있는 것은 블록화를 통해 제거하며, 피할 수 없는 요인은 랜덤화를 통하여 상쇄시키는 것이 원칙이다. 또한 실험자가 조절하기 어려운 계량변수인 공변수(covariate)가 있다면 모형에 포함시켜 분석의 정밀도를 높여야 한다.

- 랜덤화 방법에 따라 분석방법이 달라진다. 예를 들어 지분설계(nested design)에 의해 얻은 데이터를 교차실험 형식으로 정리한 후 잘못 분석하는 경우가 있다.
- 실험에서 고려하고 있는 요인이 고정요인(fixed factor)인지 또는 랜덤요인(random factor)인지에 따라 분석방법이 달라진다. 예를 들어 혼합모형의 경우를 고정모형인 경우와 같이 분산분석 하거나, 랜덤요인의 경우 의미가 전혀 없는 다중비교를 하는 경우가 있다.
- 얻어진 데이터에 대하여 추가 분석이 충분히 이루어지지 않음으로써 정보가 낭비되는 수가 있다. 예를 들어 같은 데이터에 회귀분석을 적용하면 구체적인 변수간의 관계식, 최적조건 등을 알아낼 수 있는 상황임에도 불구하고 단지 분산분석만 실시 함으로써 유의성만 검토하고 그치는 경우가 있다.
- 통계적 유의성과 실용적 유의성이 반드시 일치하는 것은 아니다. 예를 들어 최적의 공정조건을 찾아냈더라도 장비를 교체해야 하고 막대한 비용이 든다면 교체하지 않는 선에서 최적조건을 찾아볼 필요가 있다.

- 실험이 끝난 후 적절한 조치를 취하고 관리 단계로 넘어가야 한다. 예를 들면 개선 되었다는 증거가 확실한지 여부, 역기능 발생여부, 추가적 실험의 필요성 등을 충분히 검토할 필요가 있다.

프로젝트를 수행하는 과정에서 통계기법을 사용할 때 통계분석 결과가 그 동안 믿어 왔던 것과 상반된 결과가 나오면 당혹스럽게 된다. 이때 데이터에 의문을 갖는 것은 좋지만 무조건 마음을 닫고 무시하는 것은 옳지 않으며, 예를 들어 다음 사항을 검토해 볼 필요가 있다.

- 데이터는 믿을만한가?
- 데이터에 이상치나 영향 관측값이 포함되어 있지 않은가?
- 결론을 내리기에 충분한 양의 데이터를 확보하고 있는가?
- 분석을 위해 필요한 가정은 충족되고 있는가?
- 상황에 적절한 분석방법을 택하고 있는가?

다소 중복되는 말이지만 믿을 수 있는 데이터를 바탕으로 적절한 통계기법을 올바르게 사용할 때 비로소 통계기법은 가치를 갖는다.

## 4. 결론

6시그마 경영이 정착되어 감에 따라 통계기법의 활용 역시 활발해져 가고 있다. 대부분의 경우 통계기법은 문제해결을 위한 강력한 수단으로서 커다란 기여를 하고 있

지만 프로젝트를 수행할 때 그럴듯하게 보이기 위한 포장술로 전락된 경우도 있다. 통계학은 만병통치약도 아니며, 적재적소에 올바르게 사용될 때 비로소 의미가 있으므로 통계활용을 위축시키지 않는 범위 내에서 통계오용이 일어나지 않도록 유념하여야 한다. 6시그마 프로젝트를 진행하는 과정에서 통계기법을 사용할 때 유념해야 할 사항을 정리해 보면 다음과 같다.

- 믿을 수 없는 데이터를 바탕으로 한 통계 분석은 의미가 없다.  
(garbage in garbage out)
- 통계기법은 가능한 간단한 방법을 사용해야 한다.(keep it simple statistically ; KISS)
- 목적을 분명히 하고 필요할 때만 사용한다.
- 통계기법이 요구하는 원칙들을 지켜야 한다.
- 관련된 고유지식의 바탕 위에서 통계기법을 사용해야 한다.

통계가 활성화하기 위해서는 무엇보다 경영진의 통계에 대한 인식이 바뀌어야 한다. 기업의 경영진이 정보를 어떻게 수집하고 관리하며 활용하는가에 따라 기업의 미래가 달라질 수 있다는 인식을 갖고 통계가 기업의 일상적인 업무에 통합되어 운영될 수 있도록 노력할 때 비로소 통계는 기업의 경영성과에 이바지할 수 있게 될 것이다. 기업의 입장에서 통계를 활성화하기 위한 방안으로 다음과 같은 점을 고려해 볼 수 있다.

- 데이터 시스템을 정비할 필요가 있다.

데이터에 대한 신뢰성이 향상되면 기업이 갖고 있는 문제나 약점이 드러나게 되고, 그것을 수정하기 위한 행동이 시작될 때 성과가 나타나기 시작한다. 또한 통계분석이 즉시 가능할 수 있도록 SDB(satistical data base)를 구축할 수 있다면 성과의 속도는 더욱 가속화 될 것이다.

- 통계교육을 할 때 기본에 충실해야 한다.

통계의 기본원리나 과정은 생략한 채 통계 소프트웨어의 사용만을 강조하는 경우에는 통계에 대한 두려움을 없애는 데는 도움이 될지 모르지만 현장에서 응용이 잘 안되고 오용이 되는 주요원인이 될 수 있다. 역시 기본에 충실해야 일회성 교육으로 끝나지 않고 현장에서 응용이 가능해 진다.

- 통계전문가를 양성할 필요가 있다.

통계기법은 단기간의 교육으로 체득하기 어렵고 BB(black belt)나 MBB(master black belt)가 모든 통계기법에 정통하기는 현실적으로 어려운 측면이 있다. 6시그마 경영을 추진하는 일부기업에서 재무성과검증을 위해 FEA(financial effect analyst)를 양성하듯이 통계문제에 대한 전문적 충고와 조언을 제공해 줄 수 있는 SPA(statistical problem analyst)의 양성도 고려해 볼 필요가 있다. 이 때 SPA는 반드시 통계학 전공자일 필요는 없으면 현장과 6시그마를 이해하는 사람 중 통계학에 관심을 갖고 있는 사람이면 된다.

이와 같이 하여 데이터가 구성원들간에 의사 소통할 수 있는 언어로서 자리매김하

고, 통계기법이 문제해결을 위한 수단으로서 뿐만 아니라 조직을 관리하고 더 나은 경영성과를 얻기 위한 도구로서 인식될 수 있을 때 통계학은 기업에 제대로 정착될 수 있을 것이다.

Progress, vol.34, pp.29-32.

## 참고문헌

- [1] 김 재주(1981), “한국통계의 현황과 장래-공업통계”, 「통계학연구」, 한국통계학회, 10권, 창립10주년 기념호.
- [2] 박 성현, 김 재주(1991), “공업통계의 현황과 활성화 방안”, 「통계학 연구」, 한국통계학회, 20권, 창립20주년 기념호.
- [3] 박 성현(1996), “잘못 쓰기 쉬운 통계의 함정에 관한 고찰”, 「품질경영학회」, 제24권 제2호, pp.179-186.
- [4] 박 성현(2001), “공업통계의 과거·현재·미래”, 「한국통계학회 논문집」, 8권, 창립 30주년 기념호, pp.27-34.
- [5] 한국통계학회(1991), 「알고 보면 재미있는 통계이야기」, 자유아카데미.
- [6] Britz, G., Emerling, D., Mare, L.B., Hoel, R.W., Shado, T.(1997), “How to teach others to apply statistical thinking”, Quality Progress, pp.67-79.
- [7] Campbell, S.K.(1974), Flaws and fallacies in statistical thinking, Prentice-Hall, New Jersey.
- [8] Pyzdek, T.(2001), The six sigma handbook, McGraw-Hill.
- [9] Stein, Philip(2001), “Measurements for Business-using the metrology body of knowledge to enhance management decisions, business operation”, Quality