

## 통계적으로 사고하라

### 읽기와 쓰기 능력만큼 중요한 통계적 사고

정보화 시대라는 표현처럼 현대인들은 예전에 비해 훨씬 많은 정보를 가지게 되었다. 인터넷과 IT기술의 발달로 현대인들이 개별적으로 얻을 수 있는 정보의 양도 많아졌고, 그런 지식 정보를 분석하고 효율적으로 다루는 지식 경쟁력이 각별히 중시되는 시대가 되었다. 개인이나 조직의 경쟁력은 바로 정보 데이터를 효과적으로 다룰 수 있는 능력, 분석하고 적절한 판단을 내리는 능력, 그리고 미래를 예측하는 능력에 있다고 해도 과언이 아니다. 여기서 통계는 자료의 효과적 수집과 분석 그리고 객관적이고 명확한 척도를 제공할 뿐만 아니라 현대사회를 살아가는 사람들과 정보를 공유하고 소통하기 위한 중요한 수단이기도 하다.

특히 우리나라는 예전에 비해 민주화의 진전과 자유 경쟁 체제의 도입에 따라 사회 구성원들, 예를 들면 정부, 정당, 사회단체, 기업, 가계주체들 간의 이해관계가 복잡해지고 있다. 그리고 이해관계가 다양해짐에 따라 각 주체들이 내세우는 주장들의 충돌도 그만큼 많아지고 있다. 예를 들어 현재 진행되고 있는 한미 FTA의 경우를 보더라도 정부와 시민단체의 주장은 정반대로 크게 대립하고 있다. 이들 각자는 자신들의 주장의 '객관성'과 '사실성'의 근거로 통계와 데이터를 활용한다. 동일한 출처의 통계자료가 각기 다른 해석 과정을 거쳐 서로 상반된 주장들을 뒷받침하는 근거로 이용하는 아이러니가 발생하고 있다.

통계자료는 일반인들에게는 '사소하게' 보일지 모르지만, 전문가들에게는 '민감하고', '결정적인' 가정과 해석의 차이에 따라 상반된 분석결과를 만들어 낸다. 따라서 주장의 타당성을 증명하는 데 통계자료를 활용하거나, 통계자료를 근거로 한 주장의 타당성을 검토할 수 있는 '통계 자료 이해' 능력은 대다수의 현대인들에게 필수적으로 요구된다. 타임머신을 처음으로 생각해낸 SF 소설가 허버트 조지 웰즈의 말<sup>1)</sup>이 정보화 시대에서 지식사회로 가는 요점에 현실화되고 있는 것이다.

---

1) 통계적인 사고는 유능한 시민이 되기 위해 읽기와 쓰기 능력과 마찬가지로 반드시 갖추어야 할 능력이다. -Herbert George Wells

## 통계의 기본 원리에 익숙해지기

<테블스 에드버킷>이라는 영화가 있다. 키아누 리브스, 알 파치노, 샤를리즈 테론 등이 출연한 영화였는데, 소도시에서 승승장구하던 변호사 키아누 리브스가 뉴욕의 큰 변호사 사무실에 고액의 연봉으로 스카우트 된 후 자신의 숨씨를 처음 보이는 장면이 있다. 판사가 형사재판에서 피고의 유무죄를 판결하는 우리나라와는 달리 미국에서는 유무죄 여부를 일반 시민들 중에서 무작위로 차출된 배심원들이 판결하는 배심원 제도를 취하고 있다. 피고측의 변호사들이 1차 선정된 배심원들 중 '배심원으로 적합하지 않다'고 배심원들을 제외시키는 장면이 나오는데, 키아누 리브스는 배심원들의 복장, 구두 등을 세심하게 관찰하고 자신의 의뢰인에게 불리한 의견을 제시할 가능성이 높은 배심원들을 죽집게처럼 찾아낸다. 그리고 이를 통해 불리했던 재판을 뒤집는 묘기를 보여준다. 즉, 1차 선정된 배심원들 중에 자신의 의뢰인에게 불리한 배심원들을 잘 찾아내서 제외함으로써 재판을 이긴다. 여기서 중요한 통계적 원리 몇 가지를 발견할 수 있다.

배심원들은 전체 시민을 대표하여 선정된 사람들이다. 다시 말하면 이들의 의견이 전체 시민의 의견을 대표하는 것이다. 배심원들은 전체 시민들 중 일부이다. 전체 시민의 의견을 일부 시민의 판단을 통해 의사결정하는 것이 미국 사법제도의 원리이다. 전체를 대표하는 일부를 통해 전체를 판단하는 것이 중요한 원리이기도 하다. 통계자료를 분석한다는 것은 전체 중에 일부인 표본 자료를 선택하고 분석함으로써 전체가 어떠하리라는 추측을 하는 것이다.

선정된 배심원들은 각자 다른 성향을 가지고 있다. 보통의 극적인 영화에서처럼 '진실에 눈을 뜬 배심원들이 만장일치로 주인공의 편을 들어주는' 경우는 현실에서는 잘 발생하지 않는다. 우리나라의 경우를 봐도 법률 관련 최후판정관인 대법원, 헌법재판소에서의 평결은 대개 5대 4, 또는 6대 3으로 사안에 대한 찬성/반대의 의견이 갈리는 것을 자주 볼 수 있다. 현실에서도 우리가 파악해야 하는 대상들은 이처럼 단일화된 모습이 아니라, 서로 다른 값을 가지며 다양성을 가지고 있다. 그렇다면 그렇게 다양성을 가지고 있는 사항들에서 사실 전체를 보기 위해서는 어떻게 해야 할까? 우선은 대상이 가지고 있는 다양성을 인정해야 한다. 그리고 그 다양성을 이해하는 방법에 익숙해져야 한다.

키아누 리브스와 다른 변호사와의 차별점은 무엇일까? 변호사는 재판에 대한 각 배심원들의 의견을 직접 배심원들에게 물어볼 수는 없다. 하지만 영리한 변호사는 각 배심원들의 의견이 사람들의 어느 속성과 관련이 있는가에 대해 잘 파악하고 있는 사람이다. 주인공은 다른 사람들이 보지 못하는 의견과 관련이 많은 변수들을 미리 많이 파악하고 있다. 그리고 이들 중에 어느 것이 중요한 것인지도 알고 있다. 그래서 그런 관련 속성과 변수들, 예를 들어 배심원들의 복장, 구두 상태 등을 관찰하여 배심원들의 의견까지도 정확하게 예측한다. 원인과 결과의 관계를 파악하고 활용하는 것이 통계의 주요한 목적 중 하나이다. 원인 분석이란 중요한 원인을 잘 찾아내는 것과 그 원인의 개별적인 상태를 보고 결과의 값을 예측해내는 것을 포함하는 것이다.

하지만 예측 과정에는 불확실성이 따른다. 구두를 잘 손질하는 사람이 모두 부지런한 사람이라고 말할 수는 없다. 우연히 그날 구두를 닦고 올 수도 있는 것이다. '확률'을 이해하고 활용하는 법을 알고 있다면 이러한 불확실성에 대해 좀 더 잘 대처할 수 있다. 확률은 정확한 판단에도 사용된다. 명탐정 셜록 홈즈를 생각해보면 이해하기 쉽다. 탐정은 결과에 대한 정보를 모으고, 상황에 대해 여러 추론을 한 뒤 증거를 모아서 사건 정황을 추리해낸다.

## 통계적 사고

정보와 데이터의 양은 가히 홍수라 할 만큼 폭발적으로 증가하고 있다. 이런 시대에는 데이터를 처리하고 분석해 적시에 좋은 판단을 내리는 능력이 필수적이다. 그리고 그 능력은 기본적으로 자료를 바라볼 때 통계적으로 사고할 수 있느냐에 달려 있다. 수많은 데이터가 아무런 가치가 없는 무용지물로 전락해버리느냐 아니면 분석과 판단의 유용한 도구가 되느냐는 바로 통계적 사고가 그 해답을 쥐고 있다고 해도 과언이 아니다. 이 책을 쓰면서 늘 염두에 두었던 점은 이 책을 읽는 독자들이 바로 '통계적으로 사고하기'에 익숙해지기를 바라는 마음이었다. 몇 가지로 이를 위한 방법을 정리하면 다음과 같다.

첫째, 먼저 통계자료의 수집, 정리, 분석 방법 등의 원리와 그에 활용된 관련 용어들에 익숙해지기를 바란다. 원리와 용어에 익숙해진다는 것은 그만큼 새로운 세계로 들어간다는 것을 의미한다. 이것은 수학공부와 유사하다. 시험문제를 풀 때 용어가 낯설거나, 원리를 새로 외우면 제한된 시간 내에 문제를 풀 수 없다. 기본적인 용어와 공식은 미리 익숙하게 몸에 체득해 두어

야 한다. 새로운 통계 수치를 볼 때 기본 원리를 모른다면, 통계에 대한 막연한 거부감이 생겨 통계를 기피하게 된다. 통계학의 기본적인 개념들을 잘 이해함으로써 우리가 추구하는 바인 '현상' 과 '관계'를 좀 더 잘 파악할 수 있을 것이다. 이 책에서 제시하는 많은 사례들 속에서 원리와 용어에 친숙해지면, 통계 자료와 통계학에 대한 친근감과 응용 능력이 증대되리라 생각한다. 더불어 많은 그림과 표를 통해 통계분석에 깔려있는 기본 틀을 이해하면 자료에 대한 직관력이 증대될 것이다.

둘째, 나누어 보는 방법에 익숙해져서 '통계적 사고'의 능력이 증대되기를 바란다. 통계적 사고란 기본적으로 대상을 나누어서 구분해 보는 것이라고 필자는 생각한다. 다시 말해 본질적으로 대상이 가지고 있는 '다양성'을 염두에 두고, 자료를 나누어서 다양성이 발생하는 모양을 파악하여 대상을 이해하는 것이 통계적 사고의 기본 개념이다. 여기에 더해 다양성을 설명할 수 있는 적절한 구분자를 찾아주면 좀 더 대상을 잘 이해할 수 있다. 이런 사고를 통해 독자들은 통계자료를 근거로 하는 주장들의 타당성을 평가할 수 있는 능력이 생길 수 있을 것이다.

셋째, 다양한 관계에 대한 이해이다. 관계, 특히 특정 사안에 대해서 이를 설명하는 인과관계의 이야기들은 매우 복잡하고 뒤섞여 있어서 각각의 타당성을 판정하기 힘들고 많은 사람들이 종종 여기서 혼동을 일으킨다. 하지만 인과관계에 대한 정확한 분석은 정확하고 설득력 있는 의사결정을 내리는데 많은 영향을 준다. 여기에는 숫자 분석 이전에 기초적인 논리학에 관한 지식이 필요하므로, 이 책에서는 관련 있는 부분에 한해서 이를 포함하였다. 다양한 관계의 이해를 통해 인과관계에 관련된 주장의 타당성을 평가하고, 자료를 볼 때 새로운 시각을 가질 수 있었으면 한다.

『통계의 미학』 서문에서 발췌, 최제호, 동아시아 출판사