

統計와 數字

金 然 亨
(全州大 統計學科)

우리들은 나서부터 죽을 때까지 統計와 밀접한 관계를 맺고 統計에 應用되거나 통계를 응용하면서도 너무 흔히 쓰이기 때문에 어떻게 應用되며 어떻게 응용하고 있는지를 특별히 생각하지 않고 일상 생활을 하고 있다. 또한 統計에 밝은 사람들이란 수자를 많이 기억하여 외우고 있는 것으로 생각하고 있다.

사실 수자를 기억하는 것은 필요한 일이지만 이에는 限度가 있는 것이다. 가령 우리나라의 1987년도의 GNP는 기억하고 있다 하여도 1980년의 GNP 수자까지 기억하고 있는 것은 無理이기 때문이다. 統計는 수없이 많고 또 시시각각으로 變化하는 것이기 때문에 이러한 수 많은 수자를 기억하는 것은 사실상 무리인 것이다.

數字를 기억하고 있으면 편리한 것이므로 기억할 수 있는 한, 많이 기억하는 것은 좋은 일이다. 그러나 統計에서 가장 중요한 일은 數字를 외는 것이 아니고, 數字를 보고 생각하는 것이며 사물을 판단하는 일이라. 흔히 우리가 數字에 강한 사람이라는 것은 수자를 함부로 암기하여 기억하는 것이 아니고 그 수자는 어떠한 의미를 가지고 있는가를 잘 이해하고 있는 사람인 것이다.

수자로 되게 하는 필요한 조건에는 두 가지가 있다. 첫째는 普遍性의 조건이며, 둘째로는 再現性의 조건이다. 이 두 조건을 갖추게 되면 객관적인 수자가 된다. 전자인 보편성의 조건이라는 것은 어떤 현상을 기술하기 위하여 수자를 사용하면, 여기서 만들어진 수자를 누가 보아도 납득할 수 있는 것을 말한다. 후자의 재현성의 조건은 같은 것을 두세 번 조사할 때, 몇 번이라도 같은 수자를 할당할 수 있는 것을 말한다. 이 두 가지의 조건을 충족하면 수자가 객관적으로 되는 것이다.

그러면 統計와 수자를 어떻게 구분할 것인가? 일반적으로 數字에는 數學의 수자와 統計의 수자가 있다. 수학에서 취급하는 수자는 추상적·일반적인 數를 말하는 것이고, 統計에서 취급하는 수자는 구체적·경험적인 수를 의미한다. 여기서 統計란 실질적이고 구체적인 대상, 즉 實體(entity)를 설명해 주는 數集團으로서 실험 또는 경험에 의해서 얻는 정보(information)를 數量的으로 기록한 수자를 말하는 것이다. 예를 들면 어느 대학교에 다니는 어떤 學生의 統計學 점수가 90 점이고 경제학 점수가 95 점이라 한다면, 여기서 90 과 95 란 수자 그 자체는 그 학생의

統計學 및 經濟學의 점수를 나타내고 있다는 것 이외에 더 이상의 의미를 갖고 있지 않기 때문에 統計資料로서는 아무런 의미가 없는 것이다. 그러나 그 대학교에 다니는 학생 전체의 統計學 및 經濟學 점수를 조사하여 수자로 나타냈다면 이것은 統計資料로서 의미가 있는 것이다. 그러므로 統計의 수자는 항상 그 배후에 구체적·경험적인 實體를 수반하는 수자가 되어야 한다.

그렇다면 어떠한 統計가 바람직한 統計라 할 수 있는가? 誤差 없는 統計가 생산되는 것이 바람직한 것이다. 물론 誤差라는 개념은 여러 가지 의미가 포함되어 임의적으로 정의하기는 어렵겠지만, 일반적으로 標本誤差와 非標本誤差로 구분하여 생각해 보기로 한다. 표본오차란 부분을 조사하여 전체를 알고자 하는 과정에서 일어나는 필연적인 정보 부족과 우연성과 편의 등에서 발생하는 오차를 말하며, 非標本誤差란 標本設計 理論에서 나타나지 않는 보이지 않는 오차라고 할 수 있다. 비표본오차는 여러 가지 요인에 의해 생길 수 있는데, 예를 들면 데이터의 측정·수집·처리 과정에서 생기는 실수, 불합리한 증화, 응답자의 잘못된 응답 등의 요인이 있다.

따라서 바람직한 통계로서는 위에서 설명한 바와 같이 우선 誤差가 적어야 할 것이다. 그리고 또한 비용이 가급적이면 적게 드는 것도 배놓을 수 없을 것이다. 비용이 적게 투입되고 시간도 적게 들며, 나아가 최소의 調査로 최대의 效果를 얻을 수 있다면 이는 곧 바람직한 統計가 될 수 있을 것이다.

統計集團이란 특정한 사물을 관찰하는 단위의 集合體를 말한다. 여기서 특정한 사물이라는 것은 세 가지의 환경, 즉 첫째로 시간적인 환경, 둘째는 장소적인 환경, 셋째로 속성적인 환경을 말한다. 한국인 또는 '88년에 살고 있는 한국인은 통계 집단으로 되나, 시간적인 환경 없이 그냥 막연하게 한국인이라고 할 경우, 이는 곧

統計集團으로 될 수 없는 것이다. 장소적인 환경은 일정한 地域的 범위와 지리적 구역내에 한하여 조사하는 것을 의미하고, 속성적 환경은 예컨대 물가 조사에서 중요한 상품의 가격만을 조사하든가 또는 사업체 노동 실태 조사에서 5인 이상의 종업원을 고용하는 사업체만을 조사하는 것과 같다.

그리고 통계 집단의 모든 단위는 적어도 한 가지의 성질을 공유해야 하며, 이 공유 성질을 제외하고는 무수한 점에 있어서 다른 성질을 가지고 있어야 한다. 逆으로 말하면, 集合體의 단위가 한 가지 성질도 공유하지 않거나 또는 그 성질이 모두 같으면 통계 집단이 되지 않는다는 것이다. 예를 들면 한국이라는 집단은 한국의 국적이라는 경험적 포식을 공유하고 있고 이 집단을 구성하는 각 개인은 性別, 연령, 직업, 주거지, 신장 등의 많은 성질에 있어서 상이하므로 일정한 시점에서의 한국인은 統計集團으로 되는 것이다.

이와 같이 집단화함으로써 개개의 관찰만으로는는 경향이나 法則性이 없는 것으로 보이지만 全體로서, 즉 집단화해서 볼 때에는 일정한 안정된 傾向이나 法則性이 나타나는 것이다.

이상의 통계와 통계 집단에 관한 개념의 인식을 차치하더라도 실제로 있어서는 統計의 重要性이 경시되는 경우가 많은 것 같다. 어떤 정책 개발을 위하여 特性 計數가 진요하게 필요한 경우에는 統計의 重要性이 새삼스럽게 認識되고 강조되지만 이것이 끝나면 인식권 밖으로 사라지는 경우가 허다하다. 어떤 의미에서는 統計가 기초적으로 그리고 언제나 중요한 것이기 때문에 그 중요성이 평상시에는 소홀하게 되는 것인 지도 모르겠다.

어쨌든 統計의 重要性은 누구나 모두 認識하고 강조하는 것이겠지만 平常時에도 各界에서 統計의 重要度를 깊이 인식해 주었으면 한다. *