

A Study on the History of Statistics in the Early Twentieth Century Focused on Statistical Tests and Psychology

20세기 전반기 통계학사에 대한 연구:
통계적 검정과 심리학을 중심으로

JO Jae Keun 조재근

It was not until the early twentieth century that statistics emerged as an independent academic discipline. The developments of statistical theory and methods would not have been possible without heated controversies among founding fathers. One of them, controversy on the statistical test between R. A. Fisher and J. Neyman, E. S. Pearson had been very fierce and long-lasting. On the other hand it was in the early twentieth century that psychologists began to utilize statistical test which was a hybrid of tests developed by Fisher and Neyman-Pearson. By considering the history of fields such as psychology, we can see distinctive characteristics specific to the history of statistics.

Keywords: history of statistics, history of psychology, significance test, hypothesis test; 통계학의 역사, 심리학의 역사, 유의성검정, 가설검정.

MSC: 01A61, 97K80 ZDM: A30

1 서론

지금부터 약 반 세기 전, 영국의 대표적인 통계학자 가운데 한 사람이었던 M. G. Kendall (1907-1983)은 통계학의 과거를 회고하고 미래를 전망하는 글을 한 편 쓴 바 있다. 그 글에서 그는 이론통계학의 역사를 통틀어 커다란 창조적인 움직임이 세 차례 있었다고 주장했다. i) 17, 18세기에 인구 자료에 대한 추론에서 시작하여 정치산술(political arithmetic)을 창안한 것; ii) 19세기에 빈도 분포의 법칙, 그리고 통계적 관계에 대한 연구가 명확한 모습으로 나타난 것; iii) 그리고 20세기 전반기의 수학적 통계학 연구가

이 연구는 2010년도 경성대학교 연구년 지원에 의하여 수행되었음.

JO Jae Keun: Dept. of Informational Statistics, Kyung Sung Univ. E-mail: jkjo@ks.ac.kr

Received on May 17, 2013, revised on July 30, 2013, accepted on Aug. 9, 2013.

그것이였다[25, p. 182]. Kendall이 말한 20세기 전반기의 수학적 연구란 물론 K. Pearson, R. A. Fisher, J. Neyman, E. S. Pearson 등을 대표로 영국에서 시작되고 미국에서 크게 성장한 연구를 일컫는다. Kendall만 20세기 전반기 통계학의 역사에서 수학적 연구의 역할을 강조한 것은 아니다. Kendall보다 약 한 세대 뒤에 Bradley Efron은 20세기 전체의 역사를 되돌아보는 글에서 1900년부터 1950년대까지를 수학적 이론의 시대(theory era)라고 부르고 그 이후를 방법론의 시대(methodology era)라고 불렀다[8].

사실 20세기가 막 시작될 즈음만 하더라도 통계학은 나름의 이론적인 체계를 갖춘 학문 분야로 대접받지 못하고 있었다. 물론 17세기 이래 다양한 통계적 방법이 개발되어 쓰이고 있었지만 대부분 경제나 사회 현상을 연구하는 여러 분야에서 제각각 보조적인 역할만 하고 있었으므로 당시까지 통계학은 일종의 ‘이론 없는 활용의 시기’를 오래 지속하고 있었던 셈이다. 그런 분위기가 크게 바뀐 시기가 바로 20세기 전반기였다. 체계적인 이론 연구가 이루어지자 통계학은 개별 활용 분야마다에 종속되었던 모습에서 벗어나 독립적인 위치에 오르게 되었고 대단히 넓은 분야에서 이전과 매우 다른 방식으로 활용되기 시작하였다. 그 시기는 17세기 중반 Graunt의 정치산술, 그리고 Pascal과 Fermat의 확률 연구가 등장하면서 시작된 통계학의 역사 전체를 통틀어 보았을 때, 이론과 응용 모든 면에서 통계학이 급성장한 최초의 시기라고 할 수 있다. 그만큼 복잡해진 20세기 전반기의 통계학사를 간단히 정리하는 것은 어떻게 해도 억지스러운 일일 수밖에 없겠지만 일단 20세기 통계학의 역사가 그 이전의 역사와 달랐던 점을 크게 셋으로 정리해 볼 수 있겠다.

첫째, 19세기가 끝날 때까지 통계학은 독립적인 과학의 한 분야가 아니었고 기본적으로 다른 분야를 위해 데이터를 생산하고 정리하는 도구 역할을 했었다. 하지만 20세기가 시작되고 얼마 지나지 않아 통계학은 이전까지의 주변적인 역할에서 벗어나 학문 세계에서 뚜렷하게 독립적인 지위를 차지하게 되었다. 통계학의 여러 갈래 가운데 특히 1920-30년대부터 그러한 변화의 중심 역할을 한 것은 역시 수학적 통계학 연구였다.

둘째, 전체 학문 세계에서 통계학의 활용 분야가 19세기와 비교할 수 없을 정도로 넓어졌고 그 역할 또한 보조적인 도구의 역할을 넘어 대단히 중요해졌다. 통계학이 천문학이나 측지학 등의 자연현상을 관측한 데이터를 분석하는 역할을 했던 이전과 비교했을 때 특히 사회과학 분야에서 그러한 변화가 두드러졌다.

셋째, 품질관리나 정치 여론조사, 마케팅 조사 등 20세기에 등장한 새로운 영역에서 볼 수 있듯이 산업분야에서 통계학의 역할이 크게 확대되었다. 특히 이전과 달랐던 점은 그러한 경향이 제2차 세계대전을 계기로 군사 분야에서 두드러졌다는 점이다.

많은 연구가 나온 것은 아니지만 이처럼 전문화되고 다양화된 20세기 이후의 통계학에 대해 주제별로 혹은 인물별로 다룬 역사 연구들이 나온 바 있다[36, 18, 40]. 그런데 20

세기 통계학이 깊이와 폭 모든 측면에서 방대한 규모이다보니 통계학자들이 쓴 그러한 연구들에서는 앞서 언급한 이론적 발달과 그 응용 사이의 관계에 초점을 두고 살피기는 어려웠던 것 같다. 이 글에서는 위에서 제시한 20세기 통계학의 세 가지 특징 가운데 처음 두 가지를 중심으로 살펴볼까 한다. 즉 수리통계학을 중심으로 통계학이 독립적인 과학의 한 분야로 성장하는 과정과 사회과학 분야에서 통계학 이론과 방법이 활용되는 과정을 함께 살펴봄으로써 20세기 이전까지 통계학이 다른 분야와 맺었던 관계와 그 이후의 관계를 비교해볼 수 있을 것이다.

통계학의 연구 주제들 가운데 20세기 전반기의 수리통계학 연구에서 가장 큰 비중을 차지한 것은 역시 추정 문제였다. 그런데 대체적으로 사회과학 분야들에서는 통계적 추정법보다는 검정법을 주로 이용하는 분야들도 많았다. 따라서 우리는 사회과학, 그 중에서도 범위를 좁혀서 심리학에서 통계적검정을 활용한 역사를 중심으로 살펴보려 한다. 경제학, 사회학 등 여러 사회과학 분과학자들 가운데 심리학은 20세기 전반기에 등장한 분산분석법을 위시한 여러 통계적 검정법들을 가장 적극적으로 받아들인 분야 중 하나였을 뿐만 아니라 내부에서 통계학적 방법론을 둘러싼 논의가 가장 활발했던 분야 중 하나였기 때문이다. 물론 필자의 개인적인 선택에 의존해야하는 이러한 작업은 그 한계가 뚜렷할 수밖에 없지만, 20세기가 시작된 후 약 반세기 동안 통계학의 역사를 짧게 훑어보고 싶어 하는 분들에게는 작은 도움이 될 수도 있을 것이다.

널리 알려져 있다시피 20세기가 시작되자마자 나온 K. Pearson, W. Gosset 등의 연구에 이어 통계적 검정법 연구에서 큰 획을 그은 인물은 R. A. Fisher, J. Neyman, E. S. Pearson 등이었다. 어찌 보면 그러한 대가들이 20세기 전반기에 수학적으로 통계적 검정을 연구한 역사와, 심리학에서 통계적 검정법을 활용한 역사를 같은 글 안에서 함께 다루는 것은 너무 단순한 주제처럼 보일 수도 있겠다. 오늘날의 통계학 전문가들이 볼 때 '20세기 전반기에 경험적 연구를 했던 사회과학 연구자들은 통계학의 대가들이 만들고 수학적 이론으로 뒷받침한 검정 방법을 그대로 빌려다 썼을 것'처럼 보인다. 게다가 알고 보면 통계학 내부적으로 수학적 이론 연구에 힘입어 독립성이 강화되고 다른 한편으로 통계학 외부적으로 통계적 방법이 사회과학을 위시한 여러 분야에서 보편적인 과학 연구 방법으로 널리 쓰이게 된 두 가지 과정은, 둘 사이에 그리 큰 시간적 격차가 없이 진행되었다. 따라서 그 두 과정은 먼저 통계학의 대가들이 획기적인 연구를 내놓은 다음, 사회과학을 비롯한 여러 분야의 연구자들이 그 성과를 각 분야에서 활용하는 식으로 매끄럽게 연결된 것처럼 보인다. 하지만 본문에서 보겠지만 그 과정이 늘 그렇게 진행되지만은 않았다. 20세기 전반기, 특히 1930년대에서 1950년대 사이의 역사를 조금 세밀하게 살펴보면 통계학 바깥의 사람들이 통계학적 방법을 자신의 분야에서 활용하면서 항상 통계학자들이 내놓은 연구들을 그대로 받아들여 단순히 적용하지만은 않았음을

알 수 있다. 특히 통계적 검정의 경우가 대표적인 사례라고 할 수 있다.

앞서 언급했다시피 현대적인 통계적 검정법은 K. Pearson과 W. Gosset의 경우를 제외하고 나면 흔히 ‘유의성검정 (significance test)’ 이라고 불리는 Fisher의 검정법과 ‘가설검정 (hypotheses test)’ 이라고 불리는 Neyman-Pearson의 검정법이 등장한 1920-1930년대부터 시작된다. 또한 그 두 가지 검정법의 개발자인 Fisher와 Neyman이 수십 년에 걸쳐 과연 어느 쪽 방법이 더 과학적인 것인가를 두고 양보 없이 대립하고 논쟁을 계속한 것도 통계학 전문가들에게는 매우 잘 알려진 사실이다. 양쪽의 대립은 두 대표자가 모두 세상을 떠난 이후까지도 완전히 해소되지 않았으므로 Neyman의 제자이자 오랜 동료였던 E. L. Lehmann이 1990년대에 들어서까지도 양쪽 입장이 그렇게 대립적인 것만은 아니었으며 서로 손잡을 수 있는 부분도 많다는 주장을 하면서 화해를 모색할 정도였다[28].

한편, 후대의 심리학자들이 주요 심리학 학술지에 실린 논문을 대상으로 한 표본조사를 통해 밝힌 바에 따르면, 경험적 연구 가운데 통계적 검정법을 이용한 연구가 1950년대 초반에 이미 80%를 넘을 정도였다고 한다. 그렇다면 당시의 심리학 연구자들이 이용한 검정법은 Fisher의 것이었을까, Neyman-Pearson의 것이었을까? 본문에서 보겠지만 그들이 이용한 검정법은 누구의 것도 아닌 두 가지 검정법을 혼합한 잡종 검정법 (hybrid test)이었다. 20세기 전반기에 나온 심리학 연구 논문은 물론 당시에 널리 읽힌 심리통계학 교과서에서는 통계적 검정을 누가 개발했는지 밝히지 않은 채 마치 당대의 모든 통계학자들이 유일한 한 가지 검정법만을 인정한다는 듯이 잡종 검정법을 소개했던 것이다. 각 검정법을 개발한 통계학자들이 서로 다른 것으로 간주하고 치열한 논쟁을 벌였던 두 가지 검정법을 심리학 연구자들이 모든 차이와 논쟁을 무시하고 단일한 것으로 받아들여 널리 이용한 역사는 심리학의 역사에서 뿐만 아니라 통계학의 역사에서도 매우 흥미로운 사례로서 통계학자들에게 이론과 응용 사이의 관계를 다시 생각할 필요성을 제기한다. 어찌 보면 20세기 전반기라는 시기는 통계학이 변창하기 시작한 시기인 한편, 통계학자들이 연구하는 통계학과 바깥에서 활용하는 통계학이 서로 다른 길을 걷기 시작한 시기일지도 모른다는 것이다. 이런 문제의식이 얼핏 서로 멀어 보이는 20세기 전반기 수리통계학과 통계적 검정, 그리고 심리학의 역사를 하나의 글에서 함께 생각하게 된 이유이다.

2 본론

앞서 밝혔듯이 이 글은 통계학자들이 만든 검정법을 심리학자들이 나름의 방식으로 변형하여 활용한 사례를 통해 20세기 전반기 통계학의 한 가지 특징을 살피는 글이다. 이를 위해 우리는 먼저 20세기 전반기의 통계학이 이론적인 체계나 활용 분야 측면에서 그

이전까지의 통계학에 비해 크게 달랐던 점을 살펴볼 것이다. 그 시작점을 어디로 잡든 20세기가 시작되기 전까지 통계학은 독립적인 학문 분야라고 보기 어려웠고 자연과학이든 사회과학이든 다른 분야를 위해 데이터를 수집하여 표나 그래프로 나타내고 이론적 뒷받침 없이 데이터를 분석하는 수준을 크게 넘어서지 못했다. 따라서 당시까지의 통계학과 가까운 분야는 수학보다는 천문학, 경제학, 사회학 등 데이터가 나온 개별 분야들이었다. 그러나 19세기 말엽에서 20세기가 시작될 무렵부터 영국을 중심으로 통계학은 그 이론과 활용 모두에서 비약적인 변화를 겪게 되었다. 그 변화의 중심은 수학적 통계학 연구였는데 그러한 변화 과정이 아주 부드럽게 진행된 것은 아니었다. 수학 중심의 새로운 통계학은 각국 통계학회 내부에서부터 반대에 부딪혔고 나름의 갈등을 겪으면서 자리를 잡아갔다. 그러한 과정을 통해 통계학은 이전에 갖추지 못했던 이론적인 체계를 갖춘 독립적인 학문 분야로 성장하였다. 그에 따라 심리학을 위시한 여러 분야에서 통계학적 방법을 적극적으로 이용하게 됨은 물론 계량심리학, 계량경제학 등의 이름으로 각 분야에서 새로운 연구영역들이 생기기에 이르렀던 것이다.

아래 본문에서 우리는 먼저 1930년대의 미국을 중심으로 수리통계학 연구자들이 새로운 학술지와 학회를 만들면서, 통계학을 다른 분야를 위한 도구도 아니고 수학의 하위분야도 아닌 독립적인 과학으로 만들어갔던 과정을 살펴볼 것이다. 그들의 활동은 말할 것도 없이 R. A. Fisher의 연구를 필두로 1920년대 이후에 나온 기념비적인 연구 성과들에 힘입은 것이었다. 이어서 우리는 사회과학, 그 가운데에서 심리학과 검정을 중심으로 통계학적 방법이 활용된 사례를 살펴볼 것이다.

일찍이 19세기 후반부터 나온 F. Galton, K. Pearson의 연구, 그리고 20세기 초에 나온 Fisher의 통계학 연구는 유전학이나 농학 문제를 연구하는 과정에서 나온 것이었다. 한편 최소제곱법은 Pearson과 Fisher의 연구가 나오기 훨씬 이전인 19세기 초부터 천문학을 위시한 여러 분야에서 널리 쓰인 바 있다. 하지만 천문학이든 유전학이든 농학이든 그 연구대상은 모두 자연현상이었다. 20세기에 접어들어 이후 통계학이 이전과는 크게 다른 모습으로 급성장하는 과정은 곧 새로운 통계학적 이론과 방법이 등장하는 과정이었을 뿐 아니라 그것을 활용하는 분야가 급속히 늘어나는 과정이기도 하였다. 그 가운데에서도 특히 사람들이 모여 만든 사회를 연구하는 여러 사회과학 분야에서 통계학 이론과 방법을 널리 활용하기 시작하였다. 우리가 통계학의 역사를 영국의 정치산술에서 시작하든 독일의 국상학에서 시작하든 통계학은 이른 시기부터 오늘날 사회과학 분과 학문들이 다루는 여러 현상과 밀접한 관련을 맺고 있었다. 한편 확률에 대한 연구는 B. Pascal, P. Fermat, C. Huygens, J. Bernoulli같은 수학자들이 사회 현상이 아닌 도박 문제를 푸는 과정에서 시작되었다. 하지만 그들 역시 순수하게 수리적인 연구에만 머물렀던 것은 아니었다. 예컨대 베르누이는 18세기 초에 이미 확률 연구를 여러 사회

현상에 적용할 수 있으리라 전망한 바 있으며, 19세기 초 확률 연구에서 큰 획을 그은 라플라스 역시 확률을 사회과학 연구에 활용할 것을 강조한 바 있다[2, 26].

하지만 베르누이나 라플라스의 시도에도 불구하고 확률, 통계학을 사회과학에 활용하는 작업은 오랫동안 순조롭지 못했다. 19세기 초에 이미 최소제곱법이나 중심극한정리 등 중요한 이론과 데이터 분석방법들이 등장하여 자연과학 분야에서 널리 이용된 바 있었지만 사회과학에서는 그렇지 못했다. 그 중요한 이유 가운데 하나는 당시의 확률, 통계 이론과 데이터 분석법들은 동일한 천문, 자연현상을 반복적으로 관측해서 얻은 데이터를 분석하기 위해 나온 것이었던 반면, 사회 데이터는 그렇지 않았기 때문이다. 한 가지 천문 현상을 여러 차례 반복 관측한 천문학 데이터는 동질적인 것이라고 간주하는데 아무런 문제가 없었지만 서로 다른 사람들로 이루어진 사회의 데이터를 적절한 범주에 따라 구분해서 천문학 데이터처럼 동질적인 데이터로 만들기는 어려웠던 것이다[42, chap. 5]. 단일한 자연 현상과 달리 서로 다른 사람들이 모인 사회의 현상을 연구 대상으로 삼는 학문이 부닥칠 수밖에 없는 이러한 장벽을 넘어서기 위해 A. Quetelet를 비롯한 19세기 통계학의 대가들이 여러 가지 방식으로 해결책을 모색한 바 있다. 하지만 19세기가 끝날 때까지 이 문제에 대한 답은 제대로 나오지 않았다. 이에 따라 사회 연구에 사용되는 확률, 통계학은 19세기가 끝날 무렵까지 오랫동안 ‘숫자 헤아리기’와 표나 그래프로 데이터를 요약하고 정리하거나 응용 분야별로 서로 다른 방식으로 데이터를 분석하는 수준을 넘어서지 못했다. 이처럼 확률, 통계학을 사회과학에 적용하는 과정은 순탄한 과정이 전혀 아니었던 것이다[16, chap. 14, 15, 16].

그러다가 19세기가 끝나고 20세기에 접어들면서 통계학과 사회과학의 관계는 상당히 빠른 속도로 새로운 단계로 접어들게 된다. 그 관계를 살필 때 우리가 주목해야 할 것은 당시 학문 연구의 전반적인 흐름이다. 학문의 역사에서 19세기 말부터 20세기 전반에 걸친 시기는 여러 분야들이 나름의 영역과 방법론을 갖추고 물리학으로 대표되는 자연과학에 버금가는 지위를 얻기 위해 노력했던 시기였다. 특히 사변적인 이론 중심의 성격이 강했던 사회과학 분야로서는 종전까지의 모습에서 벗어나 연구 대상이나 연구 방법론을 ‘과학’으로 격상시키는 것이 시급한 과제였다. 그러므로 당연히 이미 수 세기 전부터 상당히 확고한 지위를 누리고 있던 물리학을 위시한 자연과학 분야들을 본받으려는 시도들이 여러 사회과학 분야에서 각각 다른 방식으로 나타났다. 앞서 보았듯이 측정하고 데이터를 얻는 과정으로부터 사회과학 분야들이 ‘과학’이 되기 위해 넘어야 할 장벽이 있었고 여러 사회과학 분야들이 그러한 장벽을 넘어 보다 더 경험과학에 가까운 면모를 갖추기 위해 모색한 길 중 하나가 바로 통계학과 가까워지는 것이었다. 그렇게 된 이유는 물론 당시에는 이미 통계학이 19세기까지 보여준 모습과는 크게 다른 모습으로 변신하였기 때문이었다. 20세기 전반기에 K. Pearson, R. A. Fisher, J. Neyman 등의

대가들 덕분에 크게 성장한 통계학은 과학적 방법을 찾던 사회과학 연구자들의 관심을 끌기에 충분했던 것이다[42, 37, 18]. 이전의 통계학과 비교했을 때 그들의 연구에서 찾을 수 있는 가장 두드러진 특징을 꼽는다면 우리는 그것을 수리통계학(mathematical statistics)이라 부를 수 있을 것이다.

아래에서 조금 더 자세히 살피겠지만 통계학 내부에서 수리통계학을 중점적으로 연구하는 통계학자들의 단체와 그들의 논문을 실는 독자적인 학술지가 생긴 것은 1930년대였는데 추정과 함께 통계적 검정법은 당시부터 상당히 중요한 연구주제였다. 특히 통계학자들이 발표한 검정법은 사회과학 연구자들, 그 중에서도 실험심리학자들로부터 큰 호응을 얻은 결과 1940년대부터 심리학 연구방법론에서 대단한 비중을 차지하게 되었다. 그런데 통계학 내부에서 통계적 검정은 단일한 논리체계를 가진 유일한 방법이 전혀 아니었다. 20세기 전반기의 대표적인 통계학자인 R. A. Fisher와 J. Neyman-E. S. Pearson을 중심인물로 검정의 목적과 절차를 둘러싸고 진행되었던 논쟁은 상당히 오래 지속되었을 뿐 아니라 자못 치열한 것이었다. 즉 20세기 전반기에 통계학이 얻게 된 독자적인 학문의 지위는 통계학 연구에 높은 수준의 수학기론을 적극적으로 활용한 것에 덧붙여, 통계학의 본질적인 역할에 대해서부터 서로 다른 생각을 가진 통계학자들이 다양한 주장을 내세우며 논쟁을 벌인 결과였던 것이다. 그런데 20세기 중반 이후 심리학자들이 자신들의 데이터를 분석하기 위해 널리 활용한 통계적 검정법은 그와 달랐다. 심리학자들은 Fisher와 Neyman-Pearson의 방법을 혼합한 검정법을 널리 이용하였는데 그 결과 심리통계 교과서나 논문에서는 통계학자들 사이의 이견과 논쟁을 찾아볼 수 없게 되었다. 통계적 검정법이 사회과학에서 활용되는 과정에서 통계학 내부와 외부의 모습은 사뭇 달랐던 것이다.

2.1 수리통계학의 발달

앞서 언급했듯이 통계학이 수학이나 물리학, 경제학, 사회학 등과 나란히 독립적인 하나의 전문적인 학문분야로 성장한 것은 20세기 이후의 일이다. 그 이전에도 이미 J. Bernoulli, de Moivre, P. Laplace, K. Gauss 등의 확률 연구가 나와 있었고 1830년대부터는 영국을 필두로 여러 국가의 여러 도시에 통계학회들이 생겨 통계조사를 중심으로 한 다양한 활동을 펼친 바 있었다. 하지만 그러한 활동은 20세기 이후의 통계학과는 꽤 거리가 먼 것이었으므로 20세기가 시작되기 이전까지는 통계학회들 역시 오늘날의 학회들과 매우 다른 모습이었다.

어떤 학문이든 대학을 중심으로 독자성을 지닌 하나의 학문분야로 자리 잡는 과정에서 전문학술지의 존재, 그리고 그 분야에 대한 관심을 공유하는 사람들의 모임인 학술단체의 존재는 대단히 중요한 역할을 한다. 특히 근대 학문의 역사에서 학술지와 학회는 각

분야가 나름의 정체성을 형성하고 성장하는 데 매우 중요한 역할을 해왔다. 그런데 통계학의 역사 속에서 국가별 통계학회들과 (오늘날 International Statistical Institute(ISI)의 전신에 해당하는) 국제학회, 그리고 통계학 학술지들의 역할을 살펴보면 우리가 일반적으로 생각하는 다른 학문분야들의 모습과 상당히 다른 독특한 면모를 볼 수 있다. 일례로 구미 각국에서는 1830년대부터 통계학회가 생기기 시작했는데 대부분의 모임이 주요 도시별 단체로 시작되었다. 하지만 각 도시에 생긴 통계학회들은 대부분 오늘날의 학회처럼 학자들이 모여 연구, 토론하고 학회지를 내는 활동을 하지 않았다. 그나마 나중에 Royal Statistical Society(RSS)로 바뀌게 되는 영국의 런던통계학회(Statistical Society of London(SSL))는 학회가 생긴지 얼마 뒤인 1838년부터 정기적으로 학회지를 내기 시작하였지만 미국통계학회(American Statistical Association(ASA))의 경우는 그렇지 않았다. 1839년에 보스턴에서 창립된 미국통계학회가 오늘날 *Journal of the American Statistical Association(JASA)*의 전신에 해당하는 학회지를 내기 시작한 것은 학회 창립이후 무려 50년이 지난 때인 1889년부터였다. 영국, 미국 할 것 없이 오랫동안 학회지에 실린 글들의 주제가 오늘날의 통계학회들 학회지에 실리는 글의 주제와 사뭇 달랐음은 물론이다. 50년간 학회지도 내지 않았으므로 미국통계학회가 창립 후 그동안 어떤 활동을 했는지 세부적으로 알 수 있는 길은 사실상 없고, 초창기 ASA에 대해 우리가 알 수 있는 것은 회원의 규모, 역대 회장 명단 정도에 그친다. 미국보다 모든 면에서 앞서있던 영국 런던통계학회의 학회지에 실린 글들, 그리고 직업이 의사였던 인물(Edward Jarvis)이 1852년부터 무려 32년간 미국통계학회의 회장을 맡았다는 점 등으로부터 미루어 미국통계학회의 회원들 역시 영국의 경우와 마찬가지로 인구, 보건, 노동, 교육 등 여러 사회적 주제에 대한 데이터를 수집하는 일에 관심을 가졌으리라 짐작된다[23, p. 9-11]. 또한 19세기말 이후부터는 여러 통계학회의 회원 수도 늘어나는 등 성장하는 모습을 보이지만 학회 회원의 구성을 살펴보면 20세기가 시작되고 나서 한동안까지도 경제학을 비롯한 사회과학 분야를 배경으로 가진 사람이 대다수였다.

오늘날에 되돌아보면 다소 뜻밖의 일로 보일 수도 있지만 수학과 통계학의 관계 역시 19세기 끝 무렵까지 서로 멀었다. 영국에서 런던수학회가 생긴 것은 런던통계학회보다 30여 년 늦은 1865년이었고 프랑스수학회는 1872년, 미국수학회의 전신인 뉴욕수학회는 미국통계학회보다 거의 50년이나 늦은 1888년에야 생겼다. 통계학교수나 통계학과가 전무한 시기였으므로 통계학회의 회원 구성은 수학연구자들의 단체였던 수학회와 다를 수밖에 없었고, 19세기가 끝날 때까지도 통계학은 자연이나 사회 현상에 대한 데이터를 모으고 정리해서 다른 분야의 연구를 돕는 도구를 넘어서지 못했다. 이러한 역사는 각국의 통계 단체들이 원래 수학 단체들처럼 학문적인 연구와 토론을 위한 학자들의 모임과 거리가 멀었음을 잘 보여준다. 19세기까지의 통계학사는 ‘통계’라는 이름을 붙인

오랜 역사를 지닌 학회가 존재하고 학회지도 존재했지만 ‘통계학’이라는 분야가 나름의 학문으로 대접받지 못했던 역사이기도 하였다.

그러한 분위기에서 통계학이 이전까지의 주변적인 모습을 벗고 학문 세계에서 독자적인 지위를 얻기 위해서는 데이터 수집과 정리 단계를 넘어 수학기론을 바탕으로 한 수준 높은 체계를 갖추어야 했다. 그러한 변화를 일으킨 인물을 든다면 단연 R. A. Fisher가 으뜸일 것이다. 1920년대 이후에 발표된 Fisher의 논문과 책들 덕분에 통계학은 비로소 나름의 이론체계를 지닌 독자적인 학문으로 탄생할 수 있었다. 하지만 그런 과정이 순탄했던 것은 아니다. 통계학에 대한 수학적 연구는 1920년대에 영국에서 시작되어 1930년대에 대서양을 건너 미국에서 꽃을 피우게 된다. 1930년에 첫 호가 나온 *Annals of Mathematical Statistics(AMS)*와 1935년에 창립된 *Institute of Mathematical Statistics(IMS)*가 그러한 변화의 중심이었다. 그 이전인 1920년대 중반부터 미국 통계학계에서는 수학적 이론을 연구하는 수리통계학자들과 ASA의 주류 회원들 사이에 통계학의 성격, 통계학 교육의 방향, 학회의 진로 등을 두고 갈등이 드러나기 시작했다. IMS가 창립될 당시 주역 중의 한 사람이었던 Allen T. Craig는 20세기 초 미국 통계학의 모습을 회고하면서 “1920년 이전까지만 하더라도 수리통계학이라는 이름으로 새로이 개발되고 있던 과학적 추론법에 대해 진지하게 관심을 가진 사람을 미국 대학의 수학과에서 채용한 곳은 여섯 군데에도 미치지 못했다.”고 말했다. Craig에 따르면 1920년대에 접어든 이후부터 수리통계학을 연구하는 사람이 늘었지만 당시 미국통계학회(ASA)의 대다수 회원들은 그들의 학술지(JASA)에 의미를 알 수 없는 수학기호가 포함된 논문을 싣는데 반대했고, 미국수학회 회원들은 수리통계학 연구를 수상쩍은 것으로 치부했다고 한다. H. Carver, L. Rietz, H. Hotelling 등 AMS와 IMS를 만드는 데 주도적인 역할을 한 사람들은 대부분 수학 학위를 갖고 통계학을 연구한 사람들이었는데 그들이 쓴 수리통계학 논문들은 JASA에서 환영받는 논문이 아니었던 것이다. 그 결과 수리통계학 연구자들이 ASA 산하에 새로운 학술지 AMS를 만들었지만 거기에 실리는 논문을 읽을 수 없었던 다수 ASA 회원들은 학술지 발행비용을 내지 않으려했고 그러한 갈등의 결과 수리통계학자들은 ASA와 별도로 IMS라는 독자적인 단체를 만들기에 이르렀다[6, 34]. 오랜 역사를 지닌 통계학회와 학회지가 있음에도 불구하고 수리통계학 연구자들이 새로운 학회지와 새로운 학회를 따로 만들게 되었던 당시 미국 통계학계의 분위기를 엿볼 수 있는 사례를 하나만 든다면 ASA의 회장이었던 Fisher의 강연을 들 수 있다(R. A. Fisher가 아니고 미국 경제학자 Irving Fisher이다). 1932년 그는 통계학회 회장의 자격으로 “경제학에 봉사하는 통계학(Statistics in the Service of Economics)”이라는 강연을 했다. 이런 제목은 오늘날 통계학회 회장의 강연 제목으로는 생각하기 어려운 것으로서 경제학자들이 다수였던 당시의 분위기를 잘 보여준다[9]. 초창기 AMS

에 실린 논문들 중에는 유럽에서 나온 논문을 재수록하거나 독일어, 프랑스어 논문을 영어로 번역한 것들도 있었다. 즉 다른 학문 분야들과 마찬가지로 1930년대만 하더라도 미국 통계학은 대서양 건너편에 있는 나라들에 비해 뒤쳐진 모습이었지만 2차 세계대전을 거치면서 그러한 상황은 급속하게 바뀌게 되었다.

따라서 우리는 R. A. Fisher의 중요한 연구가 나온 1922년 또는 1925년, 또는 미국에서 IMS가 생긴 1935년 또는 IMS가 독자적인 힘으로 AMS를 발행하게 된 1938년을 수리통계학이라는 새로운 학문 분야가 뚜렷한 지위를 얻고 통계학의 모습을 크게 바꾼 시점으로 잡을 수 있겠다[23, 24]. 영국에서는 미국처럼 별도의 수리통계학 학회가 생기지는 않았지만 RSS에서는 비슷한 시기에 종래의 사회, 경제 통계 분야의 글을 신던 학술지 이외에 수학적 논문을 싣는 새로운 학술지(오늘날의 *JRSS-B*)를 새로 발행하게 된다. 1930년대는 대서양의 양쪽에서 통계학이 새로운 모습을 갖춘 시기라고 하겠다. 어찌 보면 그러한 과정은 사회과학으로부터 통계학이 독립하는 과정이라고도 볼 수 있겠지만 그렇다고 해서 사회과학 연구자들이 새로운 통계학에 대해 모두들 비판적이기만 했던 것은 아니었다. 20세기 전반기는 여러 사회과학 분야들 역시 독자적인 학문으로 체계를 갖추려하던 시기였으므로 보다 과학적인 연구를 위해 통계학 분야의 새로운 성과들에 주목하고 적극 활용하려는 사람들도 나타났다. 20세기가 시작될 때까지 일종의 상하 위계관계에 있던 사회과학과 통계학은 1920-30년대에 나온 수학적 연구에 힘입어 이전과 매우 다른 새로운 관계로 접어들게 되었던 것이다. 이전까지 한계가 뚜렷했던 통계학의 역할이 이 시기에 비로소 훨씬 중요해졌음은 물론이다. 한편, 사회과학 연구자를 비롯하여 통계학의 성과를 활용하는 사람들이 통계학의 새로운 성과 모두를 필요로 했던 것도 물론 아니었다. 당연히 그들은 수학적인 이론보다는 방법론에 주목했고 그 중에서도 Fisher와 Neyman-Pearson의 검정법은 심리학자를 중심으로 가장 널리 이용된 방법이었다. 다음 절부터 우리는 논의의 폭을 대폭 좁혀서 통계적 검정과 그 활용에 대해 살펴보려한다.

2.2 Fisher와 Neyman-Pearson의 통계적 검정법

통계학의 역사를 연구하는 학자들은 대개 통계적 검정의 역사가 18세기 초 영국의 의사였던 John Arbuthnot (1667-1735)이 발표한 신생아 성비에 대한 연구로부터 시작되었다고 평가한다[42, p. 438-441]. 이후 Daniel Bernoulli, Laplace 등의 연구가 나왔고[18, chap. 4, 14] 20세기에 접어든 이후에는 K. Pearson, W. Gosset의 연구가 등장하여 대표본 뿐 아니라 소표본 문제에 대해서도 검정법이 쓰이기 시작했다. 이어서 1925년에는 Fisher의 책 *Statistical Methods for Research Workers*를 통해 유의성검정(test of significance)이 처음 소개되었고 이어서 1928년과 1933년에 Neyman-Pearson

의 가설검정(test of hypotheses) 연구가 나왔다[28]. 사실 처음 양쪽의 연구가 나왔을 때만 하더라도 분위기는 적대적이기보다는 서로에 대해 우호적이었다. 하지만 이후 논쟁이 진행되면서 양쪽은 논리적으로나 감정적으로 대단히 날카롭게 대립하게 되었다. 논쟁은 주로 Fisher와 Neyman(E. S. Pearson은 논쟁에 소극적인 편이었다) 사이에서 진행되었는데 별다른 타협점을 찾지 못한 채, 사실은 서로 자신의 입장을 더욱 굳히면서, 오랫동안 지속되었다.

널리 알려져 있다시피 1920년대 초부터 Fisher는 추정 문제에 대해 최적성(optimal-ity in estimation)을 연구하였다. 하지만 그는 자신의 유의성 검정을 최적 검정이라는 측면에서 연구하지는 않았던 반면, Neyman과 Pearson은 그들의 논문 제목에 있는 'the most efficient tests of statistical hypotheses' 라는 구절에서 볼 수 있듯이 검정 문제에 대해 최적성의 문제를 다루었다. 따라서 Neyman과 Pearson 당사자들은 물론 Lehmann을 위시한 그들의 후계자들은 Neyman-Pearson의 가설검정이 Fisher의 유의성 검정법 연구를 심화, 발전시켰다고 판단했다. 하지만 이후 몇 십 년 동안 이어진 양쪽의 논쟁은 최적성이라는 통계학적 기준에 머무르지 않고 검정의 절차는 어떠해야할지, 과연 어떤 것이 과학적인 방법인지, 확률을 어떻게 볼 것인지, 검정 결과를 어떻게 해석할 것인지 등 통계학의 근본(foundation of statistics)에까지 닿는 많은 무거운 주제를 담고 있었다.

여기서 철학적인 논의는 제외하고 두 방법의 절차만 간단히 비교해본다면 Fisher는 귀무가설만 인정했을 뿐 대립가설을 세우는 것에 반대했고 따라서 검정의 오류를 두 가지 종류로 나누는 것도 받아들이지 않았다. 그는 Neyman-Pearson의 검정법이 품질 관리 같은 좁은 영역에서는 쓸 수 있겠지만 널리 과학적인 연구 방법이 될 수는 없다고 주장했다. 물론 오늘날 p값이라고 불리는 것은 Fisher의 것인데 그는 p값을 가설들의 채택, 기각을 가르는 기준으로 삼기보다는 가설을 반증(disprove)할 수 있는 통계적 근거(statistical evidence)의 척도로 생각했다. 즉 그는 검정법을 두 가설 중 하나를 선택하는 최종적인 결정의 방법이 아니라 데이터로부터 추론해가는 일련의 과정 가운데 한 단계 정도로, 다시 말해 타당하지 못한 귀무가설을 하나씩 버려나가는 귀납적 과정이 과학적 연구방법이라고 생각하였다. 또한 Fisher는 관습적으로 유의수준 0.05를 기준으로 귀무가설을 기각하거나 채택하는 것에 대해서도 통계학이라는 것이 그렇게 단순한 논리로 이루어진 것이 아니라면서 반대하였다. 무엇보다 그는 스스로를(사실 대단히 유명한 유전학자이기도 하였다) 자연과학자로 생각하였는데, 통계적 검정법은 연구자가 그 문제에 대해 제대로 아는 바가 없을 때에만 사용해야할, 따라서 그리 신빙성을 두기 어려운 연구방법이라고 주장하였다. 연구 주제에 대한 깊은 이해가 우선해야지, 문제를 제대로 파악하지도 못한 상태에서 이루어지는 통계학적 검정 때문에 그 연구가 과학적

인 연구가 되는 것은 전혀 아니라는 것이다[11, 12]. 논쟁이 진행되는 과정에서 Fisher는 자신의 유의성 검정법을 ‘inductive inference’라고 주장했고 Neyman은 자신의 가설검정법을 ‘inductive behavior’라고 불렀는데[33, 19] 이와 같은 양측의 대립에 대해 후세 통계학자들은 어떻게 평가했을까? 베이즈 통계학의 대표자 중 한 사람이었던 L. J. Savage는 이를 빈도학과 내부의 논쟁 중에 가장 으뜸가는 논쟁으로 평가했다.

“빈도론자(frequentists)들 사이에 존재하는 분열 가운데 가장 큰 것은 Fisher 진영과 Neyman-Pearson 진영 사이의 분열이다.” [41, p. 577]

또한 Neyman의 제자이자 오랜 기간 동료였던 E. L. Lehmann은 Fisher와 Neyman이 모두 세상을 떠나고 난 후 “1908년 Gosset의 t-검정에서 시작된 현대적인 검정이론에 대해 Fisher는 새로운 패러다임을 창조해냈지만 검정통계량을 선택하는 것에 대해서는 생각하지 못했고 Neyman과 Pearson이 두 가지 오류와 함께 가장 좋은 검정법에 대한 이론을 내놓음으로써 그 문제를 해결했다.”[29, p. 1242-1243]고 평가했다. 즉 Lehmann은 Fisher와 Neyman-Pearson의 입장이 대립되는 것이라기보다는 Fisher가 시작한 통계적 검정법 연구를 Neyman과 Pearson이 이어받아 완성시킨 것으로 보았던 것이다.

한편, 그 두 가지 검정법을 여러 측면에서 검토하는 문제는 후대 통계학자들 뿐 아니라 과학철학자들의 연구주제가 되었다[7]. 하지만 이 글에서 그런 주제까지 논의하는 것은 필자의 능력을 벗어나는 일이다. 또한 철학적인 문제를 떠나 검정 절차와 해석에 있어서 양측의 차이에 대해서도 이미 여러 연구자들이 잘 정리해두었으므로 여기서 자세한 내용을 다시 반복할 필요도 없겠다(두 가지 방법의 차이를 간명하게 설명한 사례로 Huh [22], Oakes [35] 가운데 6장, Chow [4]의 Table 2.3(p. 21), Halpin and Stam [19], Hager [17] 등이 있고 양측의 여러 차이를 설명하면서도 두 검정법이 서로 보완적인 역할을 할 수 있다고 보는 Lehmann [29]도 보라). Table 1은 Fisher [10, 11], Neyman [33], 그리고 Halpin and Stam [19, p. 631], Hager [17, p. 254-256] 등을 참조하여 Fisher의 검정법과 Neyman-Pearson 검정법의 차이에 대해 간략히 정리한 것이다.

마지막으로 1930년대에 시작된 Fisher와 Neyman-Pearson의 대립적인 입장은 오늘날까지 여러 모습으로 나타나고 있는데, 가령 “1930년대 Neyman-Pearson의 연구가 나온 이래 통계학 이론을 지배한 것은 결정론적 패러다임이었다. 데이터를 근거로서 해석하는 것(interpreting data as evidence)이 아니라 여러 행동 방식 중에서 선택하는 것이 통계학 문제가 되었다.”라는 서문이 붙은 책(제목은 *Statistical Evidence*) 역시 비판적인 입장에서 과학 연구와 통계학의 관계를 다시 묻고 있다[38]. 그와 같은 ‘근거(evidence)’에 대한 Royall의 연구는 철학자, 통계학자, 생태학자들의 연구로 더욱 확장된 바 있다[43].

비교할 점	Fisher	Neyman-Pearson
가설 설정	귀무가설(H_0)만 설정. 제1, 2종 오류를 정의할 수 없음	귀무가설(H_0)과 대립가설(H_1) 설정
검정 절차		미리 정한 유의수준(α) 조건을 만족시키면서 검정력(power)을 가장 크게 만드는 검정통계량과 기각영역을 선택한 다음 데이터 수집
	데이터로부터 p값 계산	데이터로부터 검정통계량 값 계산
	p값이 충분히 작으면 '통계적으로 유의한 결과'가 나왔다고 판단.	검정통계량 값이 기각영역 안에 포함되면 H_0 를 기각하고 H_1 채택; 그렇지 않으면 H_0 채택
검정결과 해석	p값의 크기로 '근거의 강도(strength of evidence)'를 판단. 검정결과가 유의하면 H_0 가 틀렸거나 나오기 어려운 결과가 나왔다고 판단하고 추론을 계속한다.	확률에 바탕을 둔 검정은 가설의 참과 거짓에 대한 정보를 제공할 수 없다. 동일한 검정을 반복할 때 잘못된 결과가 나올 오류의 확률을 가지고 최적의 검정법을 찾아 두 가설 가운데 하나를 선택한다.
과학철학적 입장	귀납적 추론(inductive inference)	귀납적 행동(inductive behavior)
	H_0 를 반증(disprove)할 수는 있으나 증명하거나 채택할 수는 없다. 검정의 결과는 임시적인 것으로서 다시 고칠 수 있어야한다.	통계적 검정에서는 두 가지 오류의 확률을 고려하여 어떤 가설을 선택하는 행동을 해야한다
비판	검정통계량을 임의로 선택한다. 최적검정법(optimal test)라는 보장이 없다. 유의한 결과가 나와도 H_1 이 없으므로 구체적인 결론을 내릴 수 없다. 유의한 결과가 나오지 않으면 결론은 더욱 모호하다.	제품의 합격-불합격 판정을 할 때에나 쓸 수 있는 것으로서 일반적인 과학 연구에서의 추론 방법으로 적절하지 않다. 현실적인 과학실험에서는 동일한 검정을 반복해서 얻는 확률의 개념이 타당하지 않다.

Table 1. Fisher와 Neyman-Pearson 검정법의 비교

2.3 심리학과 통계적 검정

앞에서 우리는 19세기까지 연구를 보조하는 중속된 수준에서 여러 사회과학 분야들과 가까운 관계를 유지해왔던 통계학이 20세기 전반기에 수리통계학이 발달하면서 독자적인 입지를 굳혀가는 과정을 살펴보았다. 그렇다면 그때부터 통계학과 사회과학 사이의 거리가 멀어지기 시작했을까? 물론 새삼 답을 할 필요도 없이 그렇지 않았다. 수리통계학의 발달에 따라 통계학이 다른 분야에 봉사하는 도구를 넘어 독립적인 학문분야로 성장하는 과정은 한편으로 통계학 이론과 방법이 활용되는 분야가 대폭 확대되는 과정과 겹친다. 특히 여러 사회과학 분과 학문에서 그러한 경향이 두드러졌다. 하지만 통계학의 성격이 크게 달라진 만큼 통계학과 사회과학이 만나는 점점 역시 19세기의 모습과 같을 수는

없었다. 새로운 통계학에 주목한 사회학자, 경제학자, 심리학자들은 해당 분야 내부에서 계량적인 접근법이 강조되는 새로운 영역을 개척하게 된다. 계량경제학, 계량심리학 등의 새로운 분야가 학회와 학회지를 통해 새로 등장한 시기가 바로 1930년대였다. 즉 20세기 전반기의 통계학은 독자적인 학문으로서의 입지를 얻었을 뿐 아니라 여러 개별 학문 분야의 면모를 크게 바꾸는 역할도 하였던 것이다.

20세기 전반기에 경제학, 사회학, 심리학 등 여러 사회과학 분야의 연구자들은 통계학자들의 연구 성과를 나름의 방식으로 활용하는 한편 각 분야에 적절한 통계학적 이론과 분석법을 따로 만들어내기도 하였다. 대표적인 사례로 유전학자인 S. Wright의 경로분석(path analysis), 심리학자인 C. Spearman의 요인분석(factor analysis) 등을 들 수 있겠고, errors-in-variable model, errors-in-equation model 등 계량경제학자들이 만든 여러 모형들도 마찬가지이겠다. 사회과학 연구자들 가운데에서도 특히 경제학자들(당시의 주역들 대부분이 나중에 노벨경제학상을 받게 된다)은 1930년에 계량경제학회(Econometric Society)를 따로 만들고 학술지도 펴내는 등 상당히 두드러진 활동을 독자적으로 펼친 바 있다[31]. 따라서 20세기 전반기의 통계학과 경제학, 사회학 등의 관계는 이 글에서 함께 다루기에 너무 큰 주제이므로 그 역사는 다른 기회에 살필 수 있을 것이다. 여기서는 범위를 좁혀서 심리학과 통계학의 관계를 중점적으로 살펴보려한다.

근대 이후 흔히 뉴턴주의 물리학이라 일컬어지는 결정론적인, 또는 법칙정립적인 과학관이 널리 받아들여지면서 통계학은 오랫동안 천문학의 그늘을 벗어나지 못했다. 사실 정규분포에 바탕을 둔 이론들이 20세기에 들어와서까지도 종종 ‘오차 이론(error theory)’이라고 불렸으며 오늘날 표준편차에 해당하는 것 대신 ‘probable error’가 20세기 초까지도 널리 쓰인 사례 등이 단지 명목상의 흔적만은 아니었던 것이다. 그 이름들은 모두 천문학에서 시작된 것이었다. 천문학 연구자들이 통계학적 방법을 이용한 목적은 단일한 참값 또는 보편적인 법칙을 얻기 위해서였다. 그와 같은 천문학적 사고방식은 천문학자들이 통계학에 대해 흥미를 잃어버린 이후까지도 끈질기게 살아남았는데 항상 통계학이 진보하는데 도움이 되었던 것은 아니었다. 그러한 사고방식은 특히 사람들이 모여 이루어지는 사회 현상을 나타내거나 설명하는 데에 통계학적 방법이 적용되는 과정에서 상당한 억제작용을 하였다. 천문학에서는 단일한 값을 반복 관측한 결과들의 평균값을 좋은 추정값으로 택할 수 있었지만 사회 현상에서는 단일한 값 못지않게 값들의 산포라든가 분포가 중요했다.

심리학은 그런 분위기속에서 가장 앞서 통계학적 방법을 받아들인 분야라고 할 수 있다. 심리학의 역사에서는 페히너(Gustav Fechner, 1801-1887), 그리고 이어지는 분트(Wilhelm Wundt, 1832-1920)의 연구를 실험심리학, 또는 과학적 심리학의 시작이라고 평가한다 [3, 5, 42]. Fechner와 같은 실험심리학 연구자들은 조건을 통제하여 실험을

설계하고 (주로 자신을 실험대상으로 삼아) 데이터를 얻은 다음, 거기에 자연과학에서 이미 쓰이고 있던 확률, 통계 방법을 적용하기 시작했던 것이다. 그러한 전통을 이어받은 20세기 전반기 이후의 실험심리학자들은 심리학을 객관적인 과학으로 격상시키기 위해 확률과 통계학적 추론 방법을 적극 활용하였다. 즉 확률과 통계는 심리학의 추론 문제에 대해 객관적이고 기계적인 해결책을 제공함으로써 심리학이 라플라스가 생각했던 결정론적인 자연과학을 닮아가는 데 중요한 역할을 했던 것이다[13]. 그와 같은 19세기 중반 이후 심리학의 역사를 돌이켜 본다면 1920년대와 30년대에 Fisher의 유의성검정과 Neyman-Pearson의 가설검정이 등장한 이후 통계적 검정 방법을 가장 널리 활용한 분야 가운데 하나가 바로 심리학이었던 것이 우연은 아니었던 셈이다.

이제 심리학에서 검정이 어떻게 활용되기 시작했는지 살펴보자. 물론 20세기에 통계학자들이 검정법을 만들어내기 전에도 심리학자들은 실험 데이터를 분석해왔지만 당시만 하더라도 대개 실험장치를 상세히 설명하고 집단이 아닌 개별적 결과를 표준화되지 않은 방식으로 설명하는 데 그쳤다. 통계를 제시하는 경우에도 오늘날과 같은 분석법이 없었으므로 대푯값을 위시한 기술통계 결과를 보여주는 정도에 머물렀고 데이터로부터 가설에 대한 추론을 하는 경우는 볼 수 없었다. 그러다가 1940년대 무렵부터 심리학에서 ‘추론혁명(inference revolution)’이라 부를 만한 변화가 방법론 쪽에서 등장하게 된다. 대략 1950년대 중반까지 지속된 그 ‘혁명’은 심리학을 방법론의 수준에서 통일시키는 역할을 하게 되었는데 그 결과 통계학적 추론이 심리학에서 필수불가결한 도구로 자리잡기에 이른다[13]. 그 중에서도 NHST(Null Hypothesis Significance Test)라고 흔히 불리는 검정법이 데이터로부터 가설에 대해 추론하는 유일한 도구로 확고한 위치를 차지하게 되었다. 심리학자이자 과학사가인 Gerd Gigerenzer의 연구에 따르면 그 과정에서 대표적인 심리통계학자들이 쓴 교과서들이 중요한 역할을 하였는데 거의 모든 교과서 저자들이 통계학 내부에서 진행되었던 치열한 논쟁에 대해서는 언급조차 하지 않은 채 검정법을 소개하였다고 한다. 결과적으로 심리학을 공부하는 사람들은 통계학 내부에서는 아무런 논쟁이나 갈등 없이 단일한 검정법을 만들어 쓰고 있는 것으로 믿게 되었고 대략 1960년대부터 NHST 방법은 심리학 연구에서 거의 성서와 같은 독보적인 권위를 누리게 되었다는 것이다. 일례로 심리학자들이 1911년부터 1998년까지 미국심리학회(American Psychological Association)에서 발행한 12가지 학술지를 대상으로 표본조사한 결과 해당 학술지에 실린 경험적 연구논문 가운데 82.4%가 통계적 검정법을 이용했으며 그 중에서도 p값을 이용한 논문은 1930년대 초에 나타나기 시작하여 1940년대 전반에 40%를 넘어섰고 1950년대 후반에는 86%를 넘어섰다고 한다[20].

Gigerenzer는 20세기 심리학자들이 통계적 검정법을 이처럼 적극적으로 활용한 이유는 그들이 물리학에서와 마찬가지로 심리학에서도 객관성(objectivity)과 결정론(de-

terminism)을 추구한 때문이라고 주장했다[13, 14, 15]. 어쨌든 통계학자들이 개발한 방법을 어떤 식으로든 다른 분야의 전문가들이 활용하여 그 분야 학문이 발전하는데 도움이 되었다면 다행일 것이다. 그런데 Gigerenzer의 연구 가운데 우리가 주목할 것은 것처럼 통계학적 검정법을 널리 활용한 것이 심리학에 득이 된 것만은 아니라고 평가하는 점이다. 무엇보다도 그는 오늘날까지 심리학을 위시한 온갖 통계학 교과서에서 검정 방법이 단 하나만 있는 것으로 소개하고 있는 것을 비판한다. 앞서 살펴보았듯이 Fisher의 유의성검정과 Neyman-Pearson의 가설검정은 수십 년 동안 서로 대립해온 서로 다른 검정법이었는데도 심리학자들은 유일한 방법만 존재하는 것처럼 간주하고 검정법을 이용했다는 것이다. 무엇보다 Gigerenzer가 강조해서 문제 삼는 것은 교과서에 실리는 검정 방법이 두 가지 대립되는 방법 가운데 하나를 택한 것이 아니라, 각 방법을 만든 당사자들이 다르다고 주장했던 차이점을 무시한 채 양쪽을 버무려 만든 잡종검정법이라는 점이다.

그가 잡종검정법이라고 일컫는 것은 오늘날 여러 통계학 교과서에서도 흔히 볼 수 있는 것으로서 그 설명 순서는 대략 다음과 같이 요약해볼 수 있겠다.

1. 귀무가설 H_0 와 대립가설 H_1 을 설명한다.
2. 제1종 오류와 제2종 오류(때로는 검정력까지 포함하여)에 대해 설명한 다음 유의수준(α)을 설명한다.
3. 적절한 검정통계량의 값을 데이터로부터 얻고 그 값에 대한 p값을 구한다.
4. p값과 유의수준 α 값을 비교하여 p값 $< \alpha$ 이면 H_0 를 기각한다. 또 그 반대이면 H_0 를 채택하거나 H_0 를 기각하지 않는다.

어떤 교과서에서는 이 방법을 Fisher가 만든 유의성검정법이라고 부르고 또 어떤 책에서는 Neyman과 Pearson이 만든 가설검정법이라고 부르고 있다. 물론 그 두 가지 검정법은 서로 다르므로 그 둘을 따로 구분해서 소개하는 교과서도 있는데 한국어 책, 영어 책을 가릴 것 없이 통계학개론 수준에서는 보기 쉽지 않다. 어쨌든 Fisher가 만든 유의성검정법과 Neyman-Pearson의 가설검정법이 같은 것이 아니라면 위에서 소개한 네 단계 검정법을 Fisher의 것으로 소개하거나 Neyman-Pearson의 것으로 소개한다면 둘 중 한 쪽은 옳지 않을 것이다. 어느 쪽이 옳을까? 양쪽이 모두 옳기도 하고 양쪽이 모두 틀렸기도 하다. 위의 네 단계 검정법에는 Fisher의 것과 Neyman-Pearson의 것이 섞여있기 때문이다. 위에서 귀무가설 H_0 와 대립가설 H_1 을 세우는 것은 Neyman-Pearson의 것이다. 먼저 나온 Fisher의 검정법에서는 오늘날 귀무가설(영가설, null hypothesis)이라 불리는 것만 있었고 연구가설(대립가설, alternative hypothesis)은

없었다.

이처럼 검정의 역할이나 검정 과정, 그리고 확률에 대한 생각까지 양쪽의 주장이 여러 면에서 달랐음에도 불구하고 심리학 교과서를 비롯한 여러 곳에서는 그 둘을 결합한 방법을 소개해왔고 사실 지금도 그렇게 하고 있다. Gigerenzer는 그런 잡종검정법이 널리 쓰이게 된 데에는 몇 가지 힘이 작용했다고 보고 있다. 먼저 심리학, 사회학, 교육학 등 여러 분야의 교과서를 쓴 사람들은 통계학자들 사이에 나타난 다양한 생각이나 논쟁은 물론 검정법을 만든 통계학자들의 이름조차 아예 소개하지 않음으로써 통계학을 추상적인 진리이자 빈틈없는 추론법으로 보이게 만들고 잡종검정법을 유일한 과학이론인 것처럼 보이도록 만들었다고 한다. 심리학사 연구자들이 초기의 주요한 심리통계학 교과서들(1940년에서 1950년 사이에 초판이 나온 Lindquist, Edwards, McNemar 등의 교과서)을 검토한 결과 Gigerenzer가 문제 삼고 있는 것들을 그대로 확인할 수 있었다고 한다[19]. 그러한 교과서 저자들은 통계학자들 사이에 날카로운 이견대립이 존재하고 있는 주제에 대해 것처럼 출처도 밝히지 않은 채 잡종검정법만을 소개하면서 대개 ‘심리학자들이 통계적 검정법을 제대로 이해하는 데 필요한 수학 지식을 갖추고 있지 못하기 때문’이라는 이유를 들었다는 것이다.

또한 연구자들과 주요 학술잡지 편집진에서는 잡종검정법을 과학적이고 객관적인 이론으로 떠받들면서 연구의 궁극적인 목적은 유의한 결과를 얻는 것이며 유의하지 않은 결과는 실험이 잘못된 것이므로 학술지에 실릴 수 없다는 분위기를 만들어냈다고 한다 (검정 결과가 유의함을 나타내는 ***와 같은 별표들로 가득한 연구들을 오늘날도 흔히 볼 수 있다). 그 결과 1960년대 이후 NHST는 통계학 그 자체로 심리학 연구에서 중추적인 역할을 맡게 되었고 실제적으로 유일한 통계학적 도구로 자리 잡게 되었다고 한다. 그리고 당사자들이 세상을 떠날 때까지 화해하지 못했던 Fisher와 Neyman-Pearson의 대립되는 입장은 대립이라기보다는 연속적인 발전의 과정으로 여겨졌다고 한다. 속을 들여다보면 상당히 소란하였던 통계학의 역사를 심리통계 교과서들은 조용히 누적적으로 일관성 있게 발전해온 역사로 묘사했던 것이다. 그 결과 심리학 연구에서는 점점 연구를 설계하고 연구할 가치가 있는 가설을 찾는 것보다는 유의한 검정결과를 얻는 쪽을 더 강조하게 되었고 이론이 중심이 되는 연구보다는 데이터로부터 귀납적인 연구를 강조하는 분위기로 흐르게 되었다고 한다. Gigerenzer는 그와 같은 변화가 심리학을 위해서도 바람직하지 않았을 뿐 아니라 다양한 통계분석 이론과 방법 가운데 NHST라는 한 가지 방법이 지나친 비중을 차지하게 된 결과 통계학의 역동적인 면모도 가려버리는 역작용도 하였다고 보았다.

심리학을 비롯한 사회과학계에서 있었던 통계적 검정의 폐해를 둘러싼 논란은 우선 그 양만으로도 통계학자들을 놀라게 만들 만하다. 예컨대 책으로 나온 것만 보더라도

The Significance Test Controversy [32], *Statistical Inference: A Commentary for the Social and Behavioural Sciences* [36], *Statistical Significance: Rationale, Validity and Utility* [4], *The Cult of Statistical Significance* [44] 등을 들 수 있겠다. 연구의 양 뿐 아니라 특히 Oakes [35]와 같은 책에서 볼 수 있는 논의의 깊이와 폭 역시 통계학자들이 주목할 만하다. 책 말고도 논문 형태로 지난 반 세기 동안 여러 사회과학 분야의 전문 학술지에 실린 것들은 헤아리기 불가능할 만큼 많다[39, 20, 21]. 또한 통계적 검정법의 문제점 뿐 아니라 보다 넓게 통계학적 추론을 위한 이론과 방법에 대해 통계학자들이 단일한 통일된 입장을 가졌을 것이라는 사회과학자들의 생각을 비판하는 연구에 관심을 가진 사회과학 연구자도 Gigerenzer만 있었던 것은 아니다. 가령 심리학자인 M. Oakes는 확률에 대한 여러 가지 정의, 그리고 통계학적 추론에 대한 여러 입장들(Fisher, Neyman-Pearson, Bayesian, likelihood inference)을 검토한 다음, “통계학은 논쟁들로 넘쳐나는 분야이다. 연구하면 할수록 통계학적 추론의 토대는 바닥없는 구멍에 바탕을 두고 있는 것 같다는 생각을 하게 된다.”라고 말한 바 있다[35, p. 145].

2.4 통계학자들의 역할

앞에서 살핀 것처럼 심리학에서 통계적 검정법은 Fisher의 방법과 Neyman-Pearson의 검정법을 섞은 것이었다. Fisher의 ‘귀납적 추론(inductive inference)’이나 Neyman의 ‘귀납적 결정(inductive decision) 혹은 귀납적 행동(inductive behavior)’ 처럼 그 방법들을 처음 발표한 사람들은 각자 나름 과학에 대한 다른 견해를 갖고 자신들의 검정법을 옹호했었다. 하지만 실제로 그 두 가지는 누가 어떤 생각으로 만들었는지 언급도 되지 않은 채 20세기 중반부터 통계학자들이 이견 없이 합의한 유일한 통계적 검정법으로 심리학을 비롯한 여러 분야에서 널리 이용되었던 것이다.

여기서 우리가 질문해야 할 문제가 하나 있다. 지금까지의 대략 살펴본 바에 따르면 심리학에서 오랫동안 통계적 검정을 둘러싼 논란이 지속되는 동안 (처음 통계적 검정법을 만든 Fisher나 Neyman-Pearson을 제외하고는) 통계학자들은 사실상 거의 등장하지 않았다. 과연 통계학자들은 응용 분야에서 진행되는 논란을 보고 어떤 반응을 보였을까? 물론 통계적 검정법의 오용, 남용에 대한 우려는 일찍부터 나온 바 있다. 아주 이른 시기의 사례로 JASA에 실린 Berkson [1]의 글을 들 수 있겠다. 그는 “오늘날 주요 대학에서 가르치는 통계학은 거의 전적으로 유의성 검정으로 이루어져 있다.”[9, p. 325]고 진단한 다음, 통계적 검정법의 기본 원리를 논리적으로 비판하고 Fisher 자신의 검정 사례를 비롯하여 여러 예를 통해 검정법이 잘못 이용되고 있는 현실을 경고한 바 있다.

하지만 앞서 Gigerenzer의 지적에서 볼 수 있었던 잡종검정법에 대한 논의는 심리학 연구자들 사이에서 자못 활발하게 진행된 반면 통계학 내부, 특히 수리통계학자들 사

이에서는 거의 볼 수 없는 편이었다. 대표적인 사례로 1950년대 말에 *Testing Statistical Hypotheses*의 초판을 발표한 E. L. Lehmann은 책 맨 앞에 ‘통계적 추론과 통계적 결정 (statistical inference and statistical decision)’에 대해 단 한 페이지 분량으로 아주 짝막하게 언급하였을 뿐 Neyman의 제자답게 책 전체에 걸쳐 Fisher의 이론보다는 Neyman-Pearson의 가설검정법을 다루었다. 이러한 결과는 이론통계 연구자들이 통계적 검정이 실제로 다른 분야에서 어떻게 사용되고 있는지 염려할 필요 없이 검정이론을 연구할 수 있는 시대가 되었음을 잘 말해준다[27]. 심지어 심리학에서 ‘추론혁명’이 진행되던 시기에 활발히 활동하고 있던 논쟁의 당사자들 (Fisher, Neyman, Pearson) 들 역시 자신들의 방법을 옹호하는 논쟁은 이어가면서도 그 방법들이 다른 분야에서 어떻게 쓰이며 다른 분야 연구의 성격을 어떻게 바꾸어놓고 있는지에 대해서는 사실상 무관심했던 것 같다. 지금까지 살펴본 통계적 검정의 사례를 통해 우리는 수리적인 통계학 이론, 방법들과 그 활용 사이의 간극이 20세기 전반기에 이미 상당히 깊어졌음을 확인할 수 있다.

우리가 앞서 살펴본 Gigerenzer의 우려는 20세기 중반 이후 오늘날까지 통계학자들 또는 심리통계학자들이 쓴 통계학 교과서에 소개되는 검정법이 틀렸다는 데에 있지 않다. 그는 통계적 검정이라는 것이 Fisher와 Neyman 이래 과학 연구 절차에 대해 아무런 이견 없이 통일된 단일한 입장을 가진 매끈한 분야가 아니었음에도 불구하고, 적지 않은 통계전문가들과 통계적 방법을 이용하여 ‘과학적인 연구’를 하고 싶어 하는 여러 분야의 사람들은 통계학계 내부의 논란을 살피는 대신 정체가 모호한 통일된 방법을 만들어냈을 뿐 아니라 그 방법을 교과서와 학술지 편집과정을 통해 적극적으로 가르치고 활용했던 역사를 비판한 것이다. 따라서 20세기 전반기에 확립된 통계학의 넓은 활용 범위는 어쩌면 통계학자들의 힘만으로 이루어졌다기보다는 여러 통계학자들의 다양한 생각을 하나로 단순화시켜버리는 과정을 통해 만들어진 것인지도 모른다. 당연히 그 과정은 통계학에 온전히 득이 되지만은 않았을 것이다.

3 결론

이 글에서는 20세기 전반기에 수학적 이론 연구를 중심으로 통계학이 독립적인 학문으로 비약하는 과정, 그리고 그 결과로 나온 여러 통계학적 방법이 많은 새로운 분야로 널리 활용되는 과정을 사회과학, 특히 심리학에서 통계적 검정법이 활용된 역사를 사례로 선택해 살펴보았다. 그 결과 여러 측면에서 서로 다른 Fisher와 Neyman-Pearson의 검정법이, 나온 지 얼마 되지도 않았고 서로 간에 합의점 없이 논쟁 중이었던 당시부터 널리 이용되었음을 알게 되었다. 또한 1940년부터 이미 심리통계 교과서들은 두 가지 검정법을 적당히 혼합한 방법을 소개하기 시작하였고 이후 그 잡종 검정법이 심리통계

교과서와 학술지에서 객관적 연구를 담보하는 ‘유일한 과학적 방법’으로 군림해왔음을 보였다. 물론 이미 1940년대 초부터 수십 년 동안 통계학적 검정의 오용과 남용에 대해 통계학과 심리학 안팎에서 다양한 비판들이 제기되었지만 사실상 오늘날까지도 NHST는 널리 일상적으로 이용되고 있다.

이 글에서는 통계적 검정법의 오용, 남용을 문제 삼는 대신 통계학 내부에서는 오랫동안 대립해온 Fisher와 Neyman-Pearson의 검정법이 심리학을 비롯한 통계학 외부에서 단일한 과학적 연구방법으로 널리 쓰인 것에 주목하고 이로부터 통계학의 이론 연구와 응용 사이의 간극이 20세기 전반기에 벌어지기 시작되었다는 점을 강조하였다. 결과적으로 20세기 전반기라는 시기는 수학적 연구에 힘입어 통계학만 독립적인 전문 분야로 비약한 것이 아니라 통계적 방법을 활용하는 분야들 역시 제각각 여러 가지 방식으로 나름의 전문영역을 확보하게 된 특별한 시기였던 것 같다. 게다가 지금 되돌아보았을 때 통계학 이론 연구자들과 심리통계 연구자를 비롯한 응용 분야 연구자들이 서로 적극적으로 대화하고 논쟁을 벌인 흔적은 잘 보이지 않는다. 물론 논쟁이 없었던 것이 아니라 논쟁이 이론 통계 전문가들 내부에서만, 그리고 각 분야의 응용통계 전문가들 내부에서만 따로따로 진행된 것처럼 보인다는 말이다.

한편 본문에서 살펴본 통계적 검정의 사례를 통해 우리는 단지 이론 통계 연구자들과 통계학적 방법을 활용하는 다른 분야 사람들과의 소통이라는 문제 이외에 또 다른 문제도 생각해보게 된다. 통계학계 내부에서 전반적으로 합의한 적도 없는 논쟁의 대상인 방법들을 다른 응용 분야 사람들이 어떠한 이론적인 근거도 없이 혼합하여 교과서에 싣고 연구 논문에서 널리 이용할 경우 통계학자들은 어떤 대응을 할 것인가라는 문제가 그것이다. 네 가지 정도 대응방식을 생각할 수 있겠다. 무비판적으로 통계적 방법을 이용하는 사람들을 비판하면서 이론적으로 문제가 적은 다른 방법을 이용하도록 계몽하는 것이 하나이고, 외부에서 혼합해서 만든 그 새로운 잡종방법이 수십 년간 널리 쓰였다는 현실을 인정하고 그 방법을 뒷받침할 이론을 만들어 제공하는 것이 두 번째이겠다. 세 번째 방식은 Fisher의 입장과 Neyman-Pearson의 입장이 서로 매우 대립되는 상반된 입장이라는 양측의 주장을 검토해서 양쪽을 화해시킬 수 있는 공통분모를 찾는 것이고, 마지막 방식은 양쪽의 차이가 무시할 정도로 사소한 것이라고 보거나 두 입장의 연속성을 강조함으로써, 잡종 검정법이 활용되어온 현실을 문제없다고 인정하는 방식이다. 다음과 같은 주장에서 볼 수 있듯이 거칠게 말해서 Lehmann [29, 30]의 입장은 세 번째와 네 번째 모두에 해당할 듯하다.

가설검정이라는 꽤 응집력있는 주제를 만드는 데 있어서 Fisher와 Neyman-Pearson이 이바지한 바들은 서로 보완적인 것이었다. 검정 방법론을 개발하여 초석을 놓은 것이 Fisher라면 이어서 그 방법론을 뒷받침하는 이론을

만들어 낸 사람들이 Neyman과 Pearson이었다[30, p. 88].

한편 1940년대 이래 각종 검정법을 소개하고 이용해온 여러 분야의 전문가들, 그리고 오늘날 개론 수준의 교과서에서 바로 그 방법을 ‘유의성 검정’, ‘가설 검정’ 등 이름도 구분하지 않은 채 소개하고 있는 통계학자들의 생각은 네 번째 방식에 가까워보인다. 어쨌든 일종의 과학사회학적 주제라고도 볼 수 있는 이 문제에 대해서는 과학사학자와 과학사회학 연구자들이 관심을 갖고 연구할 만하겠다.

통계학자들이 쓴 교과서나 통계학 전문 학술지가 아닌 다른 분야의 전문가들이 쓴 글에서 검정을 비롯한 여러 낯익은 주제에 대한 글을 만나는 일은 통계를 공부하는 사람들에겐 특별한 느낌을 준다. 필자가 이 글을 쓰게 된 이유도 그런 느낌을 함께 나누고 다른 사람들의 생각도 듣고 싶었기 때문이다. 통계학 바깥의 글들을 읽다보면 흔히 통계학 전공자들이 ‘응용통계’라고 간단히 부르기도 하는 영역에서 통계학에 대해 논의해온 역사가 이미 상당하며 그 논의의 폭과 깊이 그리고 무엇보다 논의의 다양성이 놀라울 정도임을 확인하게 된다. 그 결과, 통계학적 방법을 활용하는 분야는 더 이상 통계학 전공자가 왈가왈부할 수 있는 영역이 아니고 각 응용 분야의 전문가들만이 논의할 수 있는 뚜렷한 ‘전문분야’가 된 것처럼 보이기도 한다. 하지만 다른 한편, 오늘날 허다히 많은 학문 분야들 가운데 지속적으로 아주 여러 분야에서 그와 같은 쟁점을 만들고 활발한 논의의 대상이 되는 분야가 통계학 말고는 또 없을 듯도 싶다. 바로 이런 모습은 20세기 전반기 이후 통계학만이 갖게 된 아주 독특한 면모일 것이다.

참고 문헌

1. J. Berkson, “Tests of significance considered as evidence”, *Journal of the American Statistical Association* 37(1942), 325–335.
2. J. Bernoulli, *The Art of Conjecturing*, (Korean translation by Jae Keun Jo), Jimanji, 2008. J. 베르누이, 「추측술」, 조재근 옮김, 지만지, 2008.
3. E. Boring, *A history of Experimental Psychology*, second edition, Prentice Hall, 1950.
4. S. L. Chow, *Statistical Significance: Rationale, Validity and Utility*, SAGE, 1996.
5. M. Cowles, *Statistics in Psychology: An Historical Perspective*, second edition, Lawrence Erlbaum, 2001.
6. A. T. Craig, Our Silver Anniversary, *Annals of Mathematical Statistics* 31(1960), 835–837.
7. M. Dickson and D. Baird, Significance Testing, *Handbook of the Philosophy of Science, Vol. 7., Philosophy of Statistics*(edited by Prasanta S. Bandyopadhyay and Malcolm Forster), Kluwer Academic Press, 2011.
8. B. Efron, The Statistical Century, *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*(edited by J. Panaretos), Lawrence Erlbaum, 2003, 29–44.

9. I. Fisher, "Statistics in the Service of Economics", *Journal of the American Statistical Association* 28(1933), 1–13.
10. R. A. Fisher, "The Logic of Inductive Inference(with discussions)", *Journal of the Royal Statistical Society* 98(1935), 39–82.
11. R. A. Fisher, "Statistical Methods and Scientific Induction", *Journal of the Royal Statistical Society(B)* 17(1955), 69–78.
12. R. A. Fisher, *Statistical Methods and Scientific Inference*, second edition, Harfner, 1959.
13. G. Gigerenzer, Probabilistic Thinking And The Fight Against Subjectivity, *The Probabilistic Revolution, Vol. 1: Ideas in History*, (edited by L. Krüger, L. Daston, M. Heidelberger) MIT Press, 1987, 11–33.
14. G. Gigerenzer, *Rationality for Mortals. How People Cope with Uncertainty*, Oxford University Press, 2008.
15. G. Gigerenzer and D. J. Murray, *Cognition as Intuitive Statistics*, L. Erlbaum Associates, 1987.
16. I. Hacking, *The Taming of Chance*, (Korean translation by Hye Kyung Jung), Badabooks, 2012. I. 해킹, 「우연을 길들이다」, 정혜경 옮김, 바다출판사, 2012.
17. W. Hager, The Statistical Theories of Fisher and of Neyman and Pearson: A Methodological Perspective, *Theory & Psychology* 23(2013), 251–270.
18. A. Hald, *A History of Mathematical Statistics from 1750 to 1930*, Wiley, 1998.
19. P. F. Halpin and H. J. Stam, "Inductive Inference or Inductive Behavior: Fisher and Neyman: Pearson Approaches to Statistical Testing in Psychological Research (1940–1960)", *The American Journal of Psychology* 119(2006), 625–653.
20. R. Hubbard and P. A. Ryan, "The Historical Growth of Statistical Significance Testing in Psychology—And Its Future Prospects", *Educational and Psychological Measurement* 60(2000), 661–681.
21. R. Hubbard and M. J. Bayarri, "Confusion over Measures of Evidence (p 's) versus Errors(α 's) in Classical Statistical Testing", *The American Statistician* 57(2003), 171–182.
22. Huh M. H., *Tests of Significance, On R. A. Fisher's Contributions to Statistics*, Freedom Academy, 1988, 49–57. 허명희, 「유의성 검정, R. A. Fisher 연구」, 자유아카데미, 1988, 49–57.
23. W. Hunter, "Drawing the Boundaries: Mathematical Statistics in 20th-century America", *Historia Mathematica* 23(1996), 7–30.
24. W. Hunter, "An Unofficial Community: American Mathematical Statisticians before 1935", *Annals of Science* 56(1999), 47–68.
25. M. G. Kendall, "On the Future of Statistics—A Second Look", *Journal of the Royal Statistical Society, Series A* 131(1968), 182–204.
26. P. S. Laplace, *Philosophical Essay on Probabilities*, (Korean translation by Jae Keun Jo), Jimanji, 2009. P. S. 라플라스, 「확률에 대한 철학적 시론」, 조재근 옮김, 지만지, 2009.
27. E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, 1959.

28. E. L. Lehmann, Introduction to Neyman and Pearson (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Breakthroughs in Statistics*, Springer, 1992, 67–72.
29. E. L. Lehmann, “The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?”, *Journal of the American Statistical Association* 88(1993), 1242–1249.
30. E. L. Lehmann, *Fisher, Neyman, and the Creation of Classical Statistics*, Springer, 2011.
31. M. S. Morgan, *The History of Econometric Ideas*, Cambridge University Press, 1990.
32. D. E. Morrison and R. E. Henkel, *The Significance Test Controversy: A Reader*, Aldine Publishing, 1970.
33. J. Neyman, “Inductive Behavior as a Basic Concept of Philosophy of Science”, *Revue de l’Institut International de Statistique* 25(1957), 7–22.
34. J. Neyman, The Emergence of Mathematical Statistics, *On The History Of Statistics And Probability*(edited by D. B. Owen), Marcel Dekker, 1976, 147–193.
35. M. W. Oakes, *Statistical Inference: A Commentary for the Social and Behavioural Sciences*, Wiley, 1986.
36. D. B. Owen, *On the History of Statistics and Probability*, Dekker, 1976.
37. T. M. Porter, *The Rise of Statistical Thinking: 1820–1900*, Princeton University Press, 1986.
38. R. M. Royall, *Statistical Evidence: Likelihood Paradigm*, Chapman & Hall/CRC, 1997.
39. W. W. Rozeboom, “The Fallacy of the Null-Hypothesis Significance Test”, *Psychological Bulletin* 57(1960), 416–428.
40. D. Salsburg, *The Lady Tasting Tea*, (Korean translation by Joong Yang Park), Freedom Academy, 2012. D. 샬스버그, 「통계학의 피카소는 누구일까—20세기 과학혁명을 이끈 통계학 영웅들의 이야기」, 박종양 옮김, 자유아카데미, 2011.
41. L. J. Savage, “The Foundations of Statistics Reconsidered”, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, University of California Press, 1961, 575–586.
42. S. M. Stigler, *The History of Statistics*, (Korean translation by Jae Keun Jo), Hangilsa, 2005. S. M. 스티글러, 「통계학의 역사」, 조재근 옮김, 한길사, 2005.
43. M. L. Taper and S. R. Lele, *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, Chicago University Press, 2004.
44. S. T. Ziliak and D. N. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, University of Michigan Press, 2008.