

통계학 용어 - 어떻게 만들어 갈 것인가?

이 용구¹⁾, 허 명회²⁾, 신 봉섭³⁾

1. 통계학 용어 국문화의 발전사

통계학은 서양에서 시작된 학문으로 대부분 영문용어에서 기원한다. 또한 학문적 속성상 통계학은 자체 학문으로 존재하는 동시에 다른 학문 분야에서 연구방법론으로 이용되어 왔기 때문에 다중적 용어체계를 갖고 있다. 이런 이유에 더하여 특히 우리나라에서는 대부분의 학문분야와 마찬가지로 통계학도 과거에는 일본용어를 차용하였으나 1980년대 이후 미국유학을 한 교수들이 다수가 되면서 영문용어에서 추출한 국문용어의 사용이 퍼지게 되었다. 그 결과 하나의 원어에 대하여 2개 이상의 국문용어가 혼용되는 등 다소 혼란스러운 상황이 전개되었다. 이 문제를 해결하고자 한국 통계학회에서는 1987년 통계학 용어의 통일과 개념의 정립을 위해 통계용어사전을 발간하였다 (통계용어사전, 1987). 이 용어사전은 Kendall and Buckland (1982)를 근간으로 모두 3000개의 통계학 용어를 번역·수록함으로써 통계학 용어의 국문화에서 기초가 되었다.

그 후 기존의 통계학 용어를 정비하고 더욱 많은 용어를 개발하고자 1997년에 통계교육상담연구회 주관으로 통계학 용어집이 발간되었다 (통계학용어집, 1997). 국문-영문 용어의 대조표로 구성된 이 통계학 용어집의 의미는 기존용어를 정비하고 국문용어화의 원칙을 제시하였다는 데 있다. (국문용어화의 원칙에 대하여는 2절에서 기술한다.) 또한 1987년의 통계용어사전에 비해 1200개의 용어를 추가하였다. 또한 작업 결과를 엑셀 파일로 남김으로써 효율적인 후속 작업을 가능하게 하였다. 그 후 통계학회 창립 30주년 기념사업으로 2001년 통계교육상담연구회가 다시 주관이 되어 '통계학용어 개정 및 온라인화 위원회'를 구성하여 통계학 용어 보완 및 온라인화 작업을 하였다. 이 위원회에는 통계학의 각 영역별 특성을 고루 반영하기 위하여 통계교육상담연구회, 조사통계연구회, 통계계산연구회, 통계청 등이 작업에 참여하였다. 또한 지속적인 용어발굴과 개정을 위하여 웹(web)에 통계학 용어의 국문-영문 대조표를 올림으로써 온라인화 작업을 하였다. 이에 따라 한국통계학회 통계교육상담연구회 홈페이지(stat.anyang.ac.kr/sec/)에서 통계학 용어를 검색할 수 있게 되었다. 또한 1997년의 통계학 용어집에 비해 800개의 용어를 추가함으로써 약 5000개의 용어를 수록하게 되었다. 결과적으로 통계학용어에 관심을 가진 사람은 누구나 5000개의 통계학 용어에 대하여 쉽게 온라인 상에서 국문-영문 대조를 할 수 있게 되었다.

2. 1997년 통계학 용어의 국문화 지침

우리나라의 통계학 용어는 앞 절에서 설명한 바와 같이 복잡한 뿌리를 가지고 있다. 1997년 통계학 용어집 작업에서는 그동안 통계학 용어에 내재되어 있던 여러가지 문제점들을 해결하는데

1) 중앙대학교 응용통계학과 교수, E-mail : leeyg@cau.ac.kr

2) 고려대학교 통계학과 교수, E-mail : stat420@mail.korea.ac.kr

3) 안양대학교 정보통계학과 부교수, E-mail : bsshin@aycc.anyang.ac.kr

역점을 두었다. 1987년도 통계학용어사전에 누락되었으나 통계학계에 널리 이용되고 있던 bootstrap, coverage, inference, stem-and-leaf, urn model 등과 같은 용어들을 포함하는 외에, 몇 가지 잘못 사용되고 있던 중요한 용어들을 바로 개정하였는데, 예를 들면 sampling distribution을 '표본분포'에서 '표집분포' 또는 '표본추출분포'로 하였다. 그리고 일본식 한자어에서 우리에게 보다 친숙한 한자어로 바꾸었는데, 예를 들면 likelihood를 '우도(尤度)'에서 '가능도'로, 그리고 bias를 '편의(偏倚, 편기)'에서 '편향'으로 바꾸었다. 또한 순수한 한글 용어를 시도하였으며 (예컨대 double blinding을 '이중 눈가림'으로), 수학 용어에 대하여는 대한수학회에서 발간한 수학용어집을 최대한으로 참고하였다 (수학용어집, 1994). (Markov → 마르코프, Hermite → 에르미트, positive definite → 양의 정부호 등). 그리고 널리 쓰이고 있는 용어에 대하여는 대안을 제시하였다 (예컨대 null hypothesis 귀무가설 → 영가설, alternative hypothesis 대립가설 → 대안가설 등).

3. 2001년 통계학 용어의 국문화 지침

2001년도 통계학 용어의 국문화는 기본적으로 1997년도 지침에 준하여 작업하였다. 2001년도 작업에서 특히 중점을 두어 실시한 사항들을 정리하면 다음과 같다.

첫째, 새로운 용어라고 해서 모두 포함하지는 않았다. 즉, 기존 용어의 조합으로 구성이 가능한 경우는 제외하였다. 예컨대 bivariate exponential distribution은 bivariate(이변량)와 exponential distribution(지수분포)이 있기 때문에 포함하지 않았다.

둘째, 가능한 한 원어를 직역하고자 하였다. 예컨대 within-subjects factor는 '개체내 인자'로, sequential design은 '순차설계'로 하였다.

셋째, 영어식 표현인 피동형 용어는 가급적 피하고자 하였다. (예를 들면 accelerated bias-corrected bootstrap method는 '편향이 가속 수정된 붓스트랩 방법'이 아니라 '가속편향수정 붓스트랩 방법'으로 하였다.

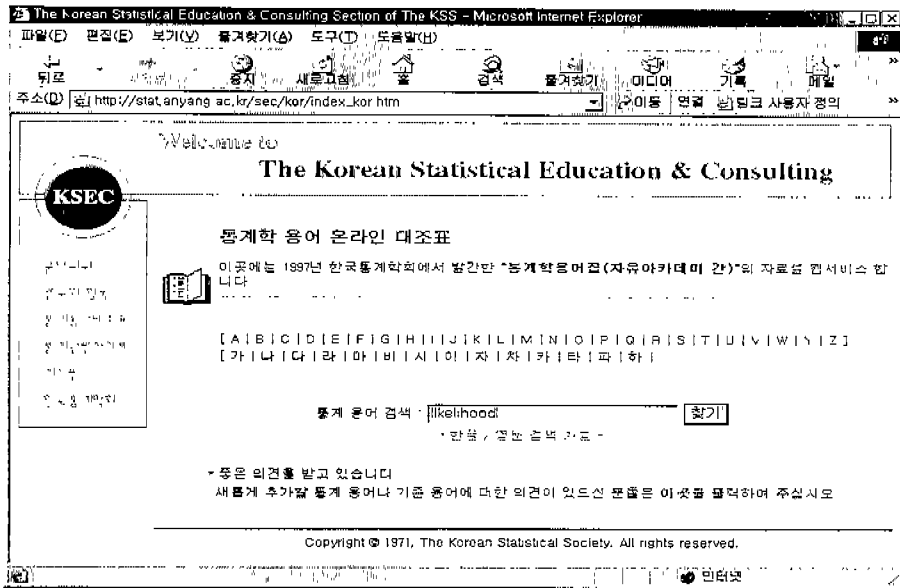
넷째, 가능한 1997년에 발간된 통계학 용어집의 용어를 최대한 활용하고자 하였다. 잦은 개정은 혼란을 국문 용어의 정착을 저해할 것이기 때문이다.

다섯째, 띄어쓰기에 있어서 한자어끼리는 붙여쓰고 (예: absolutely continuous → 절대연속(의)) 외래어는 띄어쓰도록 하였다 (예: bootstrap method → 붓스트랩 방법).

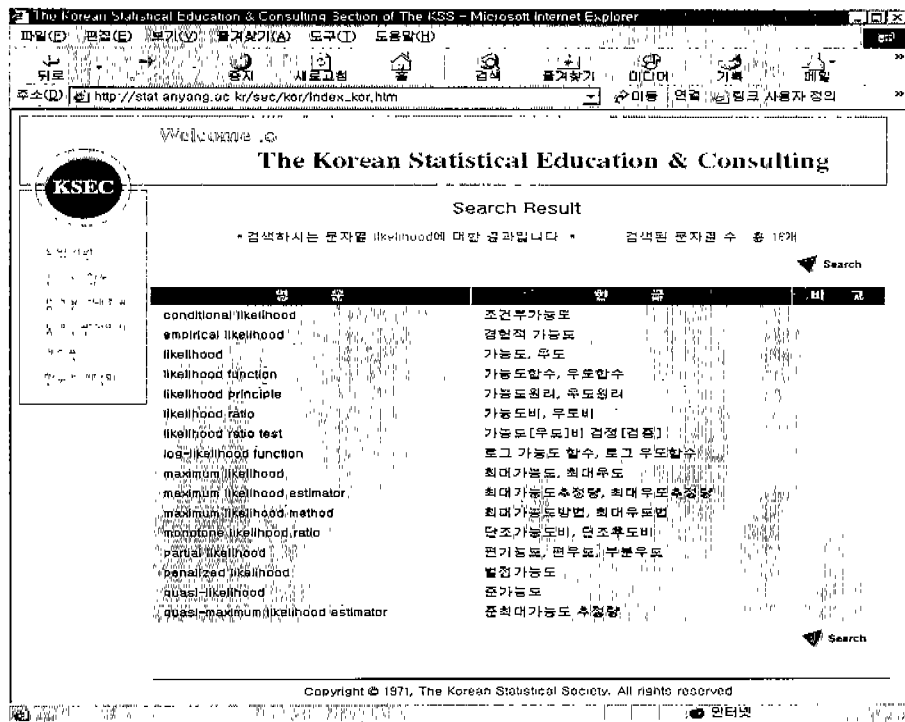
2001년도에는 위와 같은 원칙 하에서 지난 5년간 한국통계학회에서 발간된 모든 학술지, 즉 Journal of Korean Statistical Society(JKSS), 응용통계연구, 한국통계학회 논문집과 미국통계학회의 American Statistician에 게재된 논문의 중심용어(keywords)를 망라하는 방향으로 용어 추가작업을 하였다.

4. 통계학 용어의 검색

1987년의 통계용어사전과 1997년의 통계학 용어집은 책으로 출간된 것이기 때문에 편리성 측면에서 일부 제한이 있고 국문용어화의 지속적 전개가 어려웠다. 이에 따라 2001년 통계용어위원회에서는 통계학 용어의 온라인화 작업을 실시하여 한국통계학회 통계교육상담 연구회의 홈페이지(stat.anyang.ac.kr/sec/)에서 통계학 용어를 국문→영문, 영문→국문으로 쉽게 검색·확인할 수 있게 되었다. <그림 1>과 <그림 2>는 한국통계학회 홈페이지에서 통계학 용어를 검색하는 과정을 보여준다.



<그림 1> 통계교육상담연구회 홈페이지 (http://stat.anyang.ac.kr/sec/)에서의 용어검색 : 이 화면에서는 likelihood에 대한 국문용어를 찾고 있다.



<그림 2> 통계교육상담연구회 홈페이지 (계속) : 이 화면은 likelihood가 포함된 모든 영문용어에 대한 국문용어를 찾아 보여주고 있다.

5. 북한의 통계학 용어

한반도에서 냉전의 시대가 종식됨에 따라 앞으로 남북한의 빈번한 학술교류가 기대된다. 물론 통계학도 예외가 아닐 것이다. 현재 북한에서는 어떤 통계학 용어를 쓰고 있는지를 아는 것은 중요하며 우리가 차용할 만한 좋은 용어가 있다면 그렇게 하지 않을 이유가 없다. 따라서 북한의 통계학 용어에 대한 본격적 연구가 요구된다.

이 절에서는 북한에서 출간되어 사용되고 있는 다음 1권의 통계학 교재를 집중적으로 살펴봄으로써 북한의 통계학 용어에 대한 탐색을 하고자 한다.

· 조주경 (1989) <확률론> 김일성 종합대학 출판사, 평양.

저자인 조주경은 한 때 서울대학교 수학과에 재학한 바 있고, 김일성 대학 교수로 2000년 남북이산가족 교류로 서울과 부산을 왕래한 바 있으며 현재 김일성 종합대학 교수이다. 책 제목은 확률론이지만 상당부분 수리통계학을 다루고 있어, 우리로 치면 학부수준의 <확률 및 수리통계학 입문> 정도에 해당한다. 책은 통일원에서 운영하고 있는 북한자료센터 (광화문 우체국 건물내 소재)에서 구하여 열람하였다.

남북한 용어의 차이를 몇 가지로 구분하여 정리하기로 하겠다. 첫째로는 몇 가지 기본용어에서 다음과 같은 차이가 난다.

영문용어	남한	북한
random variable	확률변수	우연량
random event	임의사건	우연사건
variance	분산	두제곱편차
permutation	순열	차례무이

우리의 경우, 영문의 'random'에 대한 국문 용어가 확률, 임의 등으로 그다지 마땅하지 않고 통일되어 있지도 않은 것이 실정인데 북한은 우연으로 통일해 쓰고 있는 것이 눈의 띈다. 우연(偶然)이 쉬운 한자어이므로 'random'에 대한 국문 용어로 우리의 확률이나 임의보다 좋다고 생각된다. 그러나 영문의 'variance'를 두제곱편차라고 하는 것은 수학적 정의를 용어화한 것으로 보이는데 우리의 분산(分散)이 개념을 담고 있으므로 북한의 두제곱편차보다 낫다고 생각된다. 'permutation'을 차례무이로 하는 것은 특이하다 (여기서 무이는 無二인 것으로 짐작된다).

둘째로 북한의 특이한 발음과 표기로 북한의 통계학 용어가 우리에게 낯설게 느껴진다. 다음은 몇 가지 예이다.

영문용어	남한	북한
type 1 error	제1종 오류	제1종 오유
continuous	연속	런속
above/below	위·아래	우·아래
example	예	례

셋째로 북한의 외래 통계학 용어를 살펴본 결과 다음과 같이 몇 가지 특이한 용어들이 발견되었다.

영문용어	남한	북한
Poisson distribution	포아송 분포	뵁송 분포 / 드뵁 분포
parameter	모수 / 파라미터	파라메터
moment	적률	모멘트
random vector	확률 벡터	우연 벡토르
Markov process	마르코프 과정	마르꼬브 과정
Monte-Carlo simulation	몬테칼로 시행	몬테-칼르로 (통계적) 시행법

외래어 표기에 있어 러시아어의 영향을 볼 수 있다. 'parameter'와 'moment'에서 보듯이 오히려 우리보다 외래어를 적극적으로 수용하는 측면도 있다.

다섯째, 북한 용어의 최대 특징은 순수한 한글화이다. 다음은 몇 몇 예이다.

영문용어	남한	북한
residual	잔차	나머지
complement	여사건	나머지 사건
discrete distribution	이산분포	띠엄 분포
coefficient	계수	결수
confidence interval	신뢰구간	믿음 구간
law of large numbers	대수의 법칙	큰 수의 법칙
frequency ration	빈도 비율	잡음을
right side	우변	오른변
piecewise smooth	조각별 평활	토막미끈
rejection region	기각역	버림 구역

결수, 믿음 구간, 큰 수의 법칙, 버림 구역 등은 우리가 차용을 검토해볼 만한 좋은 용어로 생각된다.

여섯째, 북한이 좋은 한글용어를 다수 사용하고 있다고 해서 한자 용어를 기피하는 것은 아니다. 다음은 몇 가지 예이다.

영문용어	남한	북한
joint event	곱사건	적사건
likelihood	가능도 / 우도	우도
unbiased	비편향, 불편	불편(不偏)
positive definite	양의 정부호	정값
replacement	복원	반환
without replacement	비복원	비반환

우도(尤度)나 불편(不偏)보다는 우리가 1997년도 용어화 작업시 쓰기로 한 가능도(可能度)와 비편향(非偏向)이 더 좋은 용어로 생각된다.

이상으로 복한의 용어를 개괄하여 본 결과, 꼭 좋은 용어만을 쓰는 것은 아니나 우리가 차용할 만한 용어가 다수 있으며 우리가 이해하여야 용어도 또한 다수 있음을 알 수 있었다. 향후 중고교 교재 등을 포함하여 포괄적인 자료수집과 분석이 필요하다.

6. 향후 통계학 용어 개발 방향

통계학 용어에 대한 작업은 크게 기존의 용어에 대한 개정·보완과 새로운 용어의 추가로 나눌 수 있다. 기존의 용어에 문제가 있으면 교육 상담 연구회 홈페이지의 게시판을 이용하여 공론화함으로써 수정·보완해 나가야 할 것이다. 또한 새로운 용어의 추가는 한국 통계학회에서 발간하는 3개 학술지(Journal of Korean Statistical Society, 응용통계연구, 한국통계학회논문집)에 수록된 용어들을 검색하여 필요한 용어를 확보하고 외국의 기초적인 통계학 학술지인 American Statistician, Journal of the American Statistical Association 등을 보조적으로 활용하는 방안이 있을 수 있다. 마지막으로, 그러나 무엇보다도 중요한 것은, 한국통계학회의 회원들이 통계학 용어의 국문화에 관심을 갖는 것임을 말하고 싶다.

참고문헌

- [1] 대한수학회 (1994). 수학용어집, 청문각.
- [2] 한국통계학회 (1987). 통계용어사전, 자유아카데미.
- [3] 한국통계학회 (1997). 통계학용어집, 자유아카데미.
- [4] Kendall, M.G. and Buckland, W.R. (1982). *A Dictionary of Statistical Terms*, The Fourth Edition. Longman Group, London and New York.