

統計敎育과 用語 標準化를 위한 考察

崔 泳 浩*

요 약

本 論文에서는 1900年 경부터 現在에 이르기까지 國內 統計 敎育의 歷史의 背景을 調查 研究 하였다. 아울러, 統計 用語의 標準化를 위하여 理論統計學과 敎育統計學의 用語를 比較 研究하였다.

끝으로, 統計用語 標準化를 위하여 몇가지 提案을 提示하면서 研究를 마감하였다.

1. 序 論

우리나라의 敎育과정에 統計科目이 최초로 편성된 것은 1906년 4월 26일 “勅令 第 56 號 實業學校令”에 의거한 實業敎科目 내용에서 찾을 수 있다.[3] 그 후 1945년 우리나라가 日帝로 부터 解放되고, 서울大學校를 비롯한 명문 사립대학 등이 미국식 學制에 의한 高等教育(대학敎育)이 시작된 1950년 부터 1960에 이르는 시기에는 統計學 강의가 주로 社會科學 分野(예컨대, 경제학, 敎育학 등……)와 應用科學 分野(예컨대, 농학, 의학등……)에서 그들 학문연구의 道具科目으로 행 해졌으리라 생각되며, 이론 보다는 통계적 방법론과 응용사례 등을 통한 敎育이 이루어져 왔음을 그 당시 敎수들의 저서들을 통하여 소급추정하여 볼 수 있겠다.

그 당시 통계학 敎育을 담당하여 강의하여 온 것으로 알려진 저명한 敎수들로서는 서울大學校 農科大學의 김준보 敎수와 이태현 敎수로서 대상학생은 주로 농학과, 농업경제학과 학생들로 생각된다.

그리고 이 분들이 執筆한 著書들은 각각 現代推測統計學[1]과, 現代生物統計學[5] 등으로서 경제학도와 농학도의 연구 및 실험을 위한 도구과목으로서 통계학의 이론 보다는 방법론과 응용이론을 그 내용으로 담고 있는 것으로 볼 수 있다.

그리고 경상계의 경우 이정환 敎수로서 그의 저서 新統計學[4]의 내용은 통계적 추정, 검정 시계열 및 指數등 경제통계를 그 내용으로 삼고 있다.

* 全州大學校 數學敎育科.

한편, 1950년대 중반에 서울대 수학과와 수학교육과의 경우 최지훈교수와 박한식 교수에 의하여 기초통계, 수리통계 및 확률론 강의가 시작 되었으며 사립대학의 경우 동국대와 성균관대의 경우 홍성해 교수와 김영돈 교수에 의하여 부분적으로나마 통계학 강의가 이루어져 왔으며 이것이 이론통계학 교육의 시작이었고 그후 통계학과 또는 응용통계학과가 개설되어 통계학 전공교육의 기원이 된 것으로 사료된다.

그러므로, 본 연구에서는 우리나라의 통계학 교육과 통계용어의 표준모형 연구가 그 목적인 만큼 우선 국내 교육학계에서 사용하고 있는 교육통계학의 통계용어와 타 분야에서 이미 사용되고 있는 통계용어를 비교 연구함으로써 보다 본래의 뜻에 적합하고 학문 연구와 교육에 표준이 될 수 있는 용어제정에 기초자료와 문제를 제시하고자 한다.

다른 한편, 교육 통계분야의 경우, 서울대 정범모 교수가 그 효시라고 볼 수 있다. 그의 著書 教育心理의 統計的 方法[6]은 그 이론이나 방법론이 정연하고 우리나라 통계용어(Statistical Terminology)제정에 큰 공헌을 한 것으로 간주된다. 그 후, 그의 학문을 계승한 많은 교육학자들이 교육통계학의 주류를 이루며 후학들, 특히 2세 교육을 담당하게 된 교사들을 교육하여 왔다.[2]

따라서, 중 고교에서 통계를 가르치는 수학교사들도 교육통계의 영향을 받게 되었음도 사실이다.

이제까지의 통계교육에 있어서 특히, 교육통계의 통계용어가 현직교사들에게 대부분 대물려 졌으리라고 하겠다.

2. 統計用語의 由來

앞서도 언급한 바 있듯이 1950년대부터 1960년대에 이르기까지 통계학 교육은 타 전공교육의 도구과목으로서 특히 교육학에서 교육심리와 교육평가 등의 도구과목으로서의 강의와 교육이 활발히 이루어져 왔다고 하겠다.

뿐만 아니라, 교육통계학은 우리나라 통계학교육 특히, 통계용어의 語源 또는 起源(Origin)이었다고 할 수 있을 만큼 많은 기여를 한 것도 사실이다.

그러나 통계용어가 많은 학자들의 참여와 제정과정을 거침이 없이 과도적으로 써온 것들이 그대로 정착된 느낌도 없지는 않다고 하겠다.

한편, 교육 통계학 이외의 타분야 저술들을 통하여 통계용어들을 살펴보면 대개의 경우 日書를 통하여 공부 하였거나 통계학 교재를 편저 하였을 경우 일본식 用語標記를 그대로 옮겨 적어서 그 것이 우리 일상 통계용어로 정착되었으리라고 볼 수 있다.

이와 같은 과도기적 통계교육 시대를 통하여 탄생한 다양한 통계용어들은 통계학을 공부하는 초심자 들과 통계실무가 들이 통계적 이론과 그 응용을 이해하는데 많은 어려움과 混頓(Chaos)를 빚어 왔으리라고 생각된다.

그러던 중 1960년대 초에, 美國의 서베이 리서치 사(Survey-Research 社), Rice 社長

이 이끌고, Tepping 박사를 단장으로 하는 주한 통계고문단이 내한하여 국내 통계학자들과 공동으로 1961년에 제정 간행 한 바 있는 統計用語集(Statistical Vocabulary)이 우리나라 통계용어 표준화의 처음이라고 하겠다. [8]이 用語集은 우리나라에서 최초로 統計用語를 標準化함으로써 장차 통계학의 이론과 방법론을 발전 시키고 보다 체계적이고 객관적으로 통계학을 교육함으로써 학문적인 발전을 기함은 물론, 통계실무 및 통계행정을 개선하는데 기여 하게된 계기를 마련한 시도였다고 하겠다.

그 후 1988년, 한국 통계학회가 주선하여 통계용어사전[8]을 편찬 간행하기에 이르렀다.

한편, 1950년대에 태동(胎動)한 통계학 교육이 1980년대 말에 이르기까지 40여년 동안 많은 발전을 거듭하여왔고 동국대와 고려대에 통계학과와 응용통계학과가 1960년대 초에 신설 인가되어 명실상부한 독립전공 교육으로서의 통계학 교육이 시작된 이래 현재에는 국내 50여개의 통계학과와 석, 박사 과정에서 많은 통계학과와 석학들을 교육배출하기에 이르렀다.

그런데, 우리의 현실은 아직도 통계학 용어표준화가 이루어 지지 않아 이론통계학자와 응용통계학자, 통계 실무가 사이에 학문적 공동연구와 통계적 방법론에 있어서 의사소통에 많은 어려움을 줄 만큼 混線을 초래하고 있음도 사실이다.[9]

그러므로, 본 연구를 통하여 우선, 교육통계학의 대표적인 통계용어와 기호(Notation)들을 비교 고찰함으로써 가장 이상적이라고 생각되는 용어 및 기호를 선정하고 이러한 과정을 통하여 통계용어의 표준화에 대한 필요성을 제기하고 또, 다음 연구에 기초자료를 제공함으로써 통계학 연구와 이용의 디딤돌(礎石)로 삼고자 한다.

3. 統計用語의 比較

우리가 현재 상용하고 있는 통계용어의 대부분은 앞서도 지적 하였듯이 日書를 통하여 교육받은 저자들이나 그들의 서적(Text)들에서 무리가 없는 것이면 그대로 답습하여 받아들인 것이 그 주류(主流)를 이루고 있고 또 다른 한 조류는 교육통계학자와 경제통계학자들이 사용하여온 용어들이 정착된 것으로 사료된다.

뿐만 아니라, 1960년초 주한 통계고문단이 제정한 통계용어들이 어느정도 표준화 된 용어으로써 오늘날 통계학자들과 실무가들이 많이 이에 따르고 있다 하겠다.

그러나 교육통계학의 경우 그 表現들이 독특한 바 있어 이제부터 대표적인 용어들을 선정 비교코자 한다.

그리고 그들이 사용하고 있는 記號(Notation)들 또한 소개 비교함으로써 이들에 관하여도 다소의 의견을 제안코자 한다.

특히, 본 연구는 교육통계용어와 이론통계분야의 통계용어 사이에 교량을 놓음으로서 통계연구자들의 이해를 촉진하고 통계적 방법론의 적용에 기여함을 내부적인 목적으로 삼고자 한다.

3.1 記述統計學分野

조사 대상 개체들의 特性(Characteristic)을 나타내 주는 “變數(Variable) 또는 變量(Variate Value)”을 變因으로 부르므로 이 用語에서 연속변인, 이산변인 등의 용어들이 파생한다. 그리고, “度數(Frequency)”는 頻度로 命名함으로 “度數分布(Frequency-Distribution)”은 頻度分布로 불리워 지고 특히, “散布度(Dispersion)”의 경우 變散度로 “分散(Variance)”는 變量으로 명명함으로써 Variable과 혼돈이 야기 될 우려도 있다.

또, 이 표현에서 “分散分析(Analysis of Variance)”는 變量分析이라고 불리워 지게 된다. 그밖에도 歪度(Skewness)와 尖度(Kurtosis)는 각각 偏布度와 聳度로 명명하여 사용되어 오고 있다.

3.2 推測統計學分野

조사대상의 집합인 “母集團(Population)”은 全集, 그 부분집합인 標本(Sample)”은 標集으로 명명함으로써 標本分布(Sampling Distribution)는 標集分布, “小標本(Small Sample)”은 小標集으로 각각 불리워진다.

특히, “假說檢定(Hypothesis Test)”를 “假說檢證”으로 “有意水準(Significance-Level)”은 有意度로 각각 사용하여 오고 있다.

3.3 記號의 比較

이론통계학에서는 母數(Parameter)와 統計量(Statistic)을 구분하여 母數의 경우 희랍文字(Greek Letter)로 표기하고, 후자의 경우는 라틴文字(Alphabet)로 표현하는 것이 불문률(不文律)로 되어 있다.

예컨대, 母平均과 母分散 그리고 母相關係數를 각각 μ , σ^2 그리고 ρ 로 나타내는데 반하여 標本으로부터의 標本平均, 標本分散 그리고 標本相關係數를 각각 \bar{X} , S^2 그리고 R 로 표기하는 것이 관례이다.

그런데, 교육통계와 응용통계학의 경우 이들의 기호를 엄격히 구분하지 않고 혼용하는 경우가 있으며 아주 일반화하여 平均은 M (Mean의 약자)로 표기한다.

한편, 두 母集團의 비교에서는 M , N 등으로 母平均을 구분하여 표기한다.

따라서, 標準誤差(Standard Error)의 경우 σ_x , $\sigma_{x_1-x_2}$ 등은, σ_M , σ_{M-N} 등으로 쓴다. 이상의 내용을 정리 요약하면 다음 <표1>과 같다.

〈표. 1〉 用語와 記號의 比較

| 區分 | 原 語 | 統計用語集 | 敎育統計 | 과 생 어 |
|-----------------------|---------------------|---|-------------------------------|-----------------------------|
| 記 述 統 計 學 | Variable Frequency | 變數(數) 芻수(度數) | 變因頻度 | 連續變因 또는 離散變因 頻度分布(度數分布) |
| | Dispersion Variance | 散布度 分散 | 變散度變量 | 變量 分布(分散分析) |
| | Skewness | 歪 度 | 偏布度 | |
| | Kurtosis | 尖 度 | 聳 度 | |
| 推 測 統 計 學 | Population | 母集團 | 全 集 | 標集 分布(標本分布) |
| | Sample | 標 本 | 標 集 | |
| | Small Sample | 小標本 | 小標集 | |
| | Test | 檢 定 | 檢 證 | <i>t</i> -檢證(<i>t</i> -檢定) |
| | Significance Level | 有意水準 | 意義度 | |
| 記 號 | Normal Dstribution | $N(\mu, \sigma_2)$ | " | 正常分布(定規分布) |
| | 母平均 | μ | <i>M</i> | |
| | 標本平均 | \bar{X} | <i>M</i> | |
| | 標準誤差 | $\sigma_2 \quad \sigma_{\bar{x}-\bar{y}}$ | $\sigma_M \quad \sigma_{M-N}$ | |
| | 中位數 | \tilde{X} | <i>Md_n</i> | |

4. 統計用語 標準化를 위한 提言

앞서의 3절에서 보아 알 수 있듯이 이론 통계학과 敎育 통계학 사이에 동일한 用語들이 서로 다른 表現으로 작명(作名)되어 많은 불편과 혼선을 빚게 되고, 통계학자와 敎育자 및 통계 실무자들 사이에 학문적 의사소통에 지장을 초래하게 되어 그 敎育과 보급에 많은 장애가 되어 왔다고 사료된다.

이와 같은 번거러움은 표준화 과정을 거쳐, 標準統計用語를 만들고 보급하므로써 시급히 극복해야 할 과제(課題)임을 밝혀 두는 바이다.

우선 이와 같은 표준화 작업으로서는 ;

첫째, 기존 통계용어들을 비교 연구 하므로써 합당한 것을 설정하여 표준용어로 지정한다. 예컨대, Sample은 標本보다는 標本을 이루고 있는 개체(또는 측정치)들의 集合(Set)이므로 개체(個體) 하나하나를 標本으로 부를때, 標本들의 集合이란 의미에서 그들을 압축한 뜻이 함축 하도록 한 『標集』이 標本보다는 바람직한 용어로 사료된다.

그러면 "Sampling Distribution"도 자연스럽게 標集分布로 정해진다.

둘째, 용어의 불합리성이 混頓(Chaos)을 초래하는 것은 제외되어야 한다. 예컨대, 응용통계의 경우 "Variance"를 變量이라 하는데 이것은 순수통계의 경우 Variable을 變數 또는 變量으로 호칭하므로 혼돈을 초래하게 되므로 권장하기 어려우며, 이것에서 파생되는 變量分析은 分散分析의 意味로서 모호하게 느껴질 수도 있다.

셋째, 통계학의 이론적 의미로 볼때 어긋나는 것은 제외되어야 한다. 예컨데, 통계적 가설검정의 경우 “Test”는 檢證보다는 檢定이 더 통계적 방법론을 잘 나타내 준다고 보겠다.

왜냐 하면, “Test”는 귀무가설 H_0 를 기각(Reject)할 것 인가? 아니면 H_0 를 채택(Accept)할 것인가? 라는 判定問題를 표본검사 결과에 입각하여 양자택일(兩者擇一)하는 일종의 意思決定 과정(Decision-Making Process)이므로 “檢査와 決定”을 압축한 용어인 檢定이 보다 합당하다고 하겠다.

넷째, 끝으로 아직 제정되지 않았거나 응용통계학에서 사용되고 있는 용어를 조사 검토하여 표준용어를 제정하여야 한다.

예컨데, 교육 통계학의 경우 要比(Critical Ratio)로 T -test를 한다. 이것은 이론통계학에서는 거의 찾아 볼 수 없는 固有言語이다.

뿐만 아니라 가설검정의 경우 응용통계학에서는 有意水準(Significant Level)보다는 P -값(P -Value)를 선호한다.

본 논문에서는 이상 몇 가지 사례를 통하여 통계용어 표준화의 긴요성을 제안하였으며, 이와 같은 표준화 작업은 계속되어야 하며 본 연구는 문제제기를 위한 한 계기에 불과하다고 하겠다.

참 고 문 헌

- (1) 김준보(1954), “현대추측통계학” 서울:민중서관
- (2) 백용덕(1986), “교육, 심리, 사회연구를 위한 통계적 방법”, 서울:형설출판사
- (3) 이재욱(1982), “갑오경장 이후 한일합병까지 수학교육에 관한 연구”, 인하대학교 대학원 석사논문
- (4) 이정환(1962), “신 통계학”, 서울:법문사
- (5) 이태현(1961), “현대생물통계학”, 서울:문운당
- (6) 정범모(1964), “교육심리의 통계적 방법”, 서울:배영사
- (7) 주한고문통계단(1962), “통계용어집”, 서베이 리서치 회사
- (8) 한국통계학회(1988), “통계용어사전”, 서울:자유아카데미
- (9) 허문렬(1984), “수리통계학”, 서울:경문사

A study for the standardization of statistics terminologies and statistics education in Korea

Choi Young Ho

Abstract

In this thesis, I have discussed the historical backgrounds of statistical education in Korea from 1900 to present.

And, I have compared statistical terminologies between theoretical statistics and applied statistics for the standardization of the statistical terminologies.

Finally, I present some suggestions for the standardization of terminologies in statistics.