

엑셀 2002 통계함수의 번역 오류

李 貞 庸¹⁾

Translation errors in the statistical funtions of the Korean Excel 2002 software are analyzed into three categories: Pure translation mistakes; Vague translation; Inconsistent translation. These errors have not been corrected at least for the last ten years and therefore, it is suspicious that the Korean agents for foreign software companies are more interested in marketing with monololy power than in improving quality of their products

Keywords: Translation errors, Statistical Function, Excel

1. 서 론

미국 Microsoft (이하 MS)사의 MS-Office에 들어있는 엑셀(Excel)이 기초 통계학의 교육에 점점 많이 사용되고 있다. 2003년 4월 14일 현재 교보문고의 도서검색에서 ‘통계’와 ‘엑셀’ 또는 ‘Excel’의 키워드가 들어 있는 도서가 이미 24 권이나 된다.

본 논문은 엑셀이 통계의 교육 및 연습에 널리 이용되나 원본 소프트웨어를 한글화 하면서 많은 오류를 범했을 뿐 아니라 한국인 정서에 맞지 않는 번역으로 많은 학생들이 혼란을 겪고 있는 몇가지 예를 들고자 한다. 아울러 이러한 오류가 10년 이상 계속 고쳐지지 않았음을 지적하여 한국의 외국 sw 대리 회사가 sw의 판매에만 우선하고 그 질의 개선에는 무관심함을 비판하고자 한다.

제 2 절에서 엑셀의 역사를 개관한 후, 3 절에서 번역상의 오류들을 완전한 오류, 비명료성, 비일관성으로 나누어 지적한다. 제 4 절에서 결론으로 한국의 sw 대리회사들의 무관심을 비판한 후 개선 방향을 제시한다.

2. 엑셀의 역사

엑셀은 표계산(spread sheet)용 sw이다. 최초의 표계산용 sw는 1978년에 개발된 ‘비지칼크(VisiCalc)’이며 이후 개인용 컴퓨터의 운영체제로 사용된 DOS용 ‘Lotus 1-2-3’가 폭넓게 사용되다가 유사한 ‘QuattroPro’가 나왔으며, 이 후 Windows 운영체제가 개발된 이후 MS사에서 개발한 엑셀이 가장 널리 사용되고 있다. <표 1>에서 보는 바와 같이 가장 최근에 발표된 것은 Excel 2002이다. 그리고 엑셀의 최초 한글화는 버전 2.1였으며 1990년 9월에 처음 시판되었다.²⁾ 따라서 한글 엑셀은 시장에 나온지 10년 이상이 되었다. 한편 이러한 외국산 sw에 경쟁하

1) 인하대학교 경제학부 교수. joylee@inha.ac.kr

2) (주)마이크로소프트에 의함.

여 한글과 컴퓨터사는 자사의 문서편집용 sw인 ‘한글’과 ‘넥스소프트’ 사가 만든 표계산용 sw인 ‘넥셀’을 결합하여 ‘한글 오피스 2003’을 2002년 11월에 발표하였다.

<표 1> 엑셀 버전의 역사

발표 년도	엑셀 버전
1985	Excel 1.0
1987	Excel 2.0 for MS-DOS
1990	Excel 3.0
1992	Excel 4.0 for Windows 3.1.
1993	Excel 5.0
1995	Excel 7.0 for Windows 95/NT, Excel 95라 부름.
1997	Excel 97 for Windows
1999	Excel 2000 for Windows
2001	Excel 2002 for Windows XP

* Excel 6.0은 없으며 Excel 7.0부터는 MS-Office 버전과 같이 불립

출처: <http://www.j-walk.com/ss/excel/>

3. 한글 엑셀의 통계분석에 있는 번역 오류

MS-Office의 엑셀에는 총 80개의 통계함수와 19개의 통계분석 도구가 있다. 전문적인 통계 분석용 sw인 SAS나 SPSS, 또는 MINITAB 보다는 성능 면에서 뒤떨어지지만 데이터를 쉽게 변환할 수 있으며, 통계분석 외에 MS-Office의 다른 기능과³⁾ 함께 사용할 수 있으므로 간단한 통계분석에는 매우 유용한 sw이다. 한글화된 엑셀의 번역오류를 완전한 오류, 애매성, 비일관성이라는 관점으로 정리한다.

(1) 완전한 오류

엑셀의 통계 함수에 대한 설명은 그 함수를 지정했을 때 나타나는 ‘함수마법사’의 창에 있는 함수설명과, 그 함수에 대한 도움말을 구체적으로 요구했을 때 나타나는 ‘도움말’ 설명의 두 가지가 있다. 오류가 있는 함수의 원문과 번역문을 비교한다.

VARPA

원문: “Calculates variance based on the entire population, including logical values and text”.

3) 예를 들면, 통계분석의 문서편집 (MS-Word)이나 프리젠테이션 (MS-Power Point) 기능

번역문: “전체 인구의 평방 편차(논리값과 텍스트 포함)를 구합니다.”

이것의 정확한 번역은 “모집단의 분산(논리값과 텍스트 포함)을 구합니다.”일 것이다. 여기서 모집단 ‘population’을 ‘인구’라 번역한 것 하나만으로도 그 번역의 수준을 충분히 알 수 있다.⁴⁾ 통계에 관한 sw를 한글화함에 있어서 통계학을 모르거나 비전문가에게 의뢰한 것으로 의심된다. ‘평방편차’⁵⁾라는 어려운 단어를 사용한 것도 그렇다. 분산이라는 의미로 평방편차라는 용어를 사용하고 있지만 구태여 어렵고 구태의연한 말로 정확히 표현한다면 ‘평균 평방편차’ 이어야 한다.

VARA

원문: “Estimates variance based on a sample, including logical values and text”.

번역문: “모집단의 평방 편차(논리값과 텍스트 포함)를 구합니다.”

원문대로라면 ‘표본으로부터 모집단의 분산을 추정합니다.’라고 해야 할 것을 모집단의 분산⁶⁾을 구하는 것으로 번역하고 있다. 이것의 도움말에는 “표본의 분산을 예측합니다.”라고 하여 영어의 ‘estimate’를 ‘예측’이라 번역하고 있다.

FORECAST

이 함수의 도움말에는 선형회귀 (linear regression)를 선형‘회기’라 하여 철자법이 틀려있다. 이러한 오류가 10년이 되었다는 것이 놀랍다.

아와 같은 오류는 엑셀과 같은 방대한 sw의 번역에 생길 수 있다고 생각할 수 있다. 그러나 문제의 심각성은 이러한 오류가 앞의 <표 1>에서 알 수 있는 바와 같이 과거 10 여년간 최소 일곱 번의 새로운 버전⁷⁾이 나왔음에도 불구하고 고쳐지지 않았으며, 가장 최근에 나온 엑셀 2002에서도 이러한 오류가 그대로 담습되어 있다는 사실에 있다. 과거의 것을 그대로 베꼈다는 사실을 증명하는 것이다. 더더욱 경악스런 사실은 (미국산) 엑셀에 필적하는 ‘순수 국산 스프레드시트’라고 주장하며 2002년 11월에 내어놓은 ‘넥셀’ 조차 위의 오류를 똑같이 반복하고 있어서 그 개발의 과정을 쉽게 짐작할 수 있다.

(2) 비명료성

애매한 번역으로 사용자를 혼란시키는 경우이다. 엑셀의 회귀분석의 결과로 나온 분산분석표 <표 2>를 보자. 회귀에 의한 제곱평균과 잔차에 의한 제곱평균을 나눈 F 비가 있고, 그 옆에 ‘유의한 F’라 이름 붙여져 있다. 계산된 0.00057은 자유도 (2, 3)에서 계산된 확률

-
- 4) 이상하게도 도움말의 내용은 “모집단 전체의 분산을 구합니다.”라고 정확하게 번역되어 있다.
 - 5) 사실 ‘평방편차’라는 통계용어는 한국통계학회에서 출간한 통계용어 사전이나 통계학 용어집에는 없다.
 - 6) 평방 편차를 계속 분산으로 오인하고 있음을 상기
 - 7) 엑셀의 한글화는 버전 2.1부터였음.

$P(F > 216.6667)$ 이다. 이것을 '유의한 F'라고 한 것은 무슨 의미인지 알 수 없다.

귀무가설 $H_0: a_1 = a_2 = 0$ (a_1, a_2 는 회귀계수) 하에서 그러한 결과 (F 비)가 나올 '확률'이 0.00057이라는 의미인데 마치 유의한 'F의 값'으로 해석하기 쉽게 번역을 했다. 원문은 'significance F'이다. 이것은 차라리 'F 비의 유의도'라고 번역했어야 한다.

그런데 이러한 애매성은 원문의 'significance F'에게도 책임이 있다. 그 아래 개별 독립변수 X1과 X2의 유의성을 검정하는 t 통계량 옆에는 'significance t'에 해당하는 '확률을 'P-value'라 썼기 때문에 이것과 일관성을 유지하자면 원문도 'significance F'를 'P-Value'라 했어야 한다.

<표 2> 분산 분석의 결과

	자유도	제곱합	제곱 평균	F 비	유의한 F
회귀	2	5098.039	2549.02	216.6667	0.00057
잔차	3	35.29412	11.76471		
계	5	5133.333			
	계수	표준 오차	t 통계량	P-값	하위 95%
Y 절편	10.19608	6.243797	1.632993	0.200976	-9.67449
X 1	86.27451	6.792356	12.70171	0.001053	64.65818
X 2	4.607843	0.74018	6.225302	0.008357	2.252258

(3) 비일관성

오류는 아니지만 설명의 일관성이 없어서 사용자를 혼란하게 하는 경우이다. 엑셀의 도움말과 같은 문장은 너무 엄격한 형식이어서는 안되겠지만 그렇더라도 같은 종류의 기능을 서로 다른 방법으로 기술하면 사용자에게 혼란을 일으킨다.

함수명	함수의 설명	해당 도움말
TDIST	스튜던트 t-분포 값을 구합니다	소표본 t-분포의 백분율(확률값)을 구합니다. TDIST(x,degrees_freedom,tails) X: 분포의 확률값을 계산하려는 변량의 값입니다.

t-분포를 함수의 설명에서 스튜던트 t 분포라 부르고, 도움말에서는 소표본 t-분포라 했다. t-분포가 흔히 소표본의 평균에 관한 가설검정에 사용되기는 하지만 꼭 소표본의 평균만을 위해 사용되지는 않는 것으로 원문에는 없는 '소표본'이라는 수식어를 붙인 것은 불필요하다. 또한 사용 구문에서 x를 '분포의 확률값을 계산하려는 변량의 값'이라 번역한 것은 너무나 직역한 것이다. 차라리 그냥 't 값'이라 하는 것이 훨씬 명료하다.

함수명	함수의 설명	해당 도움말
CHIDIST	카이 제곱 분포의 단측 검정 확률을 구합니다.	카이 제곱 분포의 단측 검정 확률을 구합니다. • CHIDIST = $P(X>x)$ 와 같이 계산하며 이 때 X 는 χ^2 분포를 하는 변수입니다.
FDIST	두 데이터 집합에 대해 F확률분포값(분포도)을 구합니다	F-확률 분포 값을 반환합니다. 이 함수를 사용하여 두 데이터 집합의 산포도가 서로 다른지 확인할 수 있습니다. • FDIST는 $FDIST=P(F<x)$ 로 계산되며, 여기서 F 는 F 분포를 갖는 임의의 확률 변수입니다.

위의 예에서 χ^2 분포와 F 분포의 함수 설명에 전자는 단측 검정이라 설명을 붙인 반면, 후자는 설명이 없다. F 분포의 도움말에 두 데이터 집합이 서로 다른 산포도를 갖는지를 알아볼 수 있다는 설명이 있으나 데이터 ‘집합’이나 ‘산포도’와 같은 번역은 틀렸다. ‘두 집단간에 분산이 같은가에 대한 검정을 할 수 있다’라고 해야 할 것이다. 엑셀의 비일관적인 설명은 이 두 분포의 확률범위를 나타내는 $CHIDIST = P(X>x)$ 와 $FDIST=P(F<x)$ 에 극명하게 나타나 있다. 둘 다 분포의 꼬리 부분의 확률을 구하는 것이므로 일관성을 유지하려면 $FDIST=P(F>x)$ 로 부등호의 방향이 바뀌어야 한다⁸⁾.

엑셀에서 계산되는 확률의 범위 표시가 일반적 상식과는 거꾸로 된 경우는 t-검정에도 나타난다. 통계분석 도구의 등분산 가정 t 검정 결과표인 <표 3>을 보자.

<표 3> t-검정: 등분산 가정 두 집단

	변수 1	변수 2
평균	164.9	146.9
분산	194.9889	162.9889
관측수	10	10
공동(Pooled) 분산	178.9889	
가설 평균차	0	
자유도	18	
t 통계량	3.008462	
$P(T \leq t)$ 단측 검정	0.003773	
t 기각치 단측 검정	1.734063	
$P(T \leq t)$ 양측 검정	0.007545	
t 기각치 양측 검정	2.100924	

계산된 t 값이 3.008462 이므로 단측검정을 위해 자유도 18인 t 분포의 오른쪽 꼬리 중 t 가

8) 이 부분은 원문 엑셀의 오류이다.

3.008462보다 큰 확률은 0.003773이고, 양측검정을 위해 t 가 3.008462보다 크거나 -3.008462보다 작을 확률은 그 두 배인 0.007545이다. 그렇다면 전자의 범위 표시는 $P(T>t)$ 이어야 하고, 후자는 $P(|T|>t)$ 이어야 한다.

번역의 문제가 아니라도 원본 엑셀에도 애매한 경우는 많다. 예를 들면 VARP(모집단의 분산계산)와 VARA(표본분산의 추정)의 도움말에 똑 같은 자료로써 계산의 예를 들었는데 분명히 이 예는 모집단의 자료가 아니라 표본의 자료로 보아야 한다. 또 분산계산의 공식을 $\frac{\sum(x-\mu)^2}{n}$ 대신에 $\frac{n\sum x^2 - (\sum x)^2}{n^2}$ 으로 제시한 것은 분산이 편차의 자승의 평균이라는 개념을 학생들에게 보여주지 못한 선택이다.

4. 결론

엑셀은 그 사용의 편리성으로 세계 각국의 언어로 번역되어 전세계적인 사무자동화용 SW가 되었다. 한국에도 1990년에 처음 한글화가 된 이후 새로운 버전이 나올 때마다 계속 번역되어 보급되고 있다. 그러나 한글 엑셀에는 번역의 오류가 10년 이상 고쳐지지 않고 반복되고 있는 것이 있으며 더욱이 다른 회사의 제품에 똑 같은 오류가 나타나고 있다. 이러한 사실의 의미는 다음과 같이 정리된다.

첫째, 국내 SW 외국의 SW를 급하게 번역하여 판매하는 것에만 신경을 썼지 그 질을 높이는 데에는 무관심했다. 우리는 많은 경우 외국산 SW에 대한 한글판 설명이 너무도 불분명하고 우리의 정서에 맞지 않게 번역이 되어 차리리 원문을 억지로라도 그대로 읽어보는 편이 이해가 빠른 경우를 경험한다. SW의 사용법에 대한 번역은 그것의 배경이 되는 이론을 잘 아는 전문가가 실제로 그 SW를 사용해 보면서 번역해야 한다. 그러나 그 유명한 엑셀의 경우조차도 적어도 이 논문에서 지적한 통계분석 부분에 관해서는 지난 10년간 그렇지 못했음이 드러났다.

둘째, 같은 오류가 다른 회사 제품에서도 재현되고 있다는 의미는 SW의 기술적 개발, 즉 프로그래밍에만 신경을 쓰지 사용자와의 의사소통 방법인 도움말의 내용에는 독자적인 투자를 하지 않는다는 의미이다. 이것은 자칫 법적인 문제로도 비화될 가능성이 있다.

셋째, 같은 오류가 10년 이상 계속되고 있다는 사실은 SW 사용자와 SW 개발자간에 피드백이 활발하지 못하다는 의미이다. 이러한 문제를 해결하기 위해서는 개발자가 스스로 사용자 그룹을 만들어 적극적으로 그들의 의견을 수집하고 그것을 다음 버전의 개발에 반영할 수 있는 시스템을 만들어야 할 것이다.

<참고문헌>

- http://www.j-walk.com/ss/excel/
- (주)마이크로 소프트(2002), Office 2002
- 한글과 컴퓨터(2003), 한글 Office 2003
- 한국통계학회(1994), 통계용어사전, 자유아카데미
- 한국통계학회(1997), 통계학 용어집, 자유아카데미