2012년 대통령선거에서 광학식투표분류에 따른 후보자간 상대적 불균등성 규명

2017.4.29 | (번역・주석) 강세진_새사연 이사 | wisemaninspace@daum.net

이 보고서는 2017년 4월에 진행된 MPSA(midwest political science association) Annual Conference 2017에 발표된 논문 "A Measure to Detect Between-Candidate Relative Inequality Generated by Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea"의 일부를 발췌하여 번역하고 주석을 단 것입니다. 원문에 대한 자세한 정보는 http://www.mpsanet.org/에서 얻을 수 있으며, 원저자는 다음과 같습니다.

- HeeKyoung Chun, Department of Epidemiology, Georgia Southern University, PO Box 8015, Statesboro, USA. hchun@georgiasouthern.edu
- Pierre-Jerome Bergeron, Department of Mathematics and Statistics, University of Ottawa, Canada. pierrejerome@gmail.com
- HyunSeung Kim, Project BOO Inc. 268 Chungjeongro 3rd 11 St., Seodaemungu, Seoul, South Korea. n2mart@gmail.com
- OuJoon Kim, Project BOO Inc. 268 Chungjeongro 3rd 14 St., Seodaemungu, Seoul, South Korea. oujoon.k@gmail.com
- Hwashin Hyun Shin* 17 , Department of Mathematics and Statistics, Queen's University, 48 University Ave. Kingston, ON. Canada, K7L 3N6. hhshin@mast.queensu.ca
- *교신저자(Corresponding author)

【원문 : 논문제목 및 저자 정보】

A Measure to Detect Between-Candidate Relative Inequality Generated by Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea

HeeKyoung Chun, Department of Epidemiology, Georgia Southern University, PO Box 8015, Statesboro, USA.1-513-410-1608 hchun@georgiasouthern.edu

Pierre-Jerome Bergeron, Department of Mathematics and Statistics, University of Ottawa, Canada. pierrejerome@gmail.com

HyunSeung Kim, Project BOO Inc. 268 Chungjeongro 3rd St., Seodaemungu, Seoul, South Korea.82-10-3096-3628 n2mart@gmail.com

OuJoon Kim, Project BOO Inc. 268 Chungjeongro 3rd St., Seodaemungu, Seoul, South Korea. 82-10-5218-3395 oujoon.k@gmail.com

Hwashin Hyun Shin*, Department of Mathematics and Statistics, Queen's University, 48 University Ave. Kingston, ON. Canada, K7L 3N6. 1-613-741-0117 hhshin@mast.queensu.ca

*Corresponding author

Department of Mathematics and Statistics, Queen's University,

48 University Ave. Kingston, ON.

Canada, K7L 3N6

613-741-0117

Email: hhshin@mast.queensu.ca

1. 연구의 배경

여러 나라에서 전자투표를 채용하고 있으며, 이러한 투표방식의 소프트웨어, 하드웨어 및 운용상의 문제로 인해 개표결과가 뒤바뀌는 오류들이 나타나고 있다[1,14].

대한민국 선거관리위원회(이하 선관위)에 따르면, 2002년 이후 대통령선거에서 광학식투표지분류기(이하 광학분류기)가 쓰이고 있다[5,10]. 이 광학분류기에 의해 분류되는 투표지는 크게 분류표(기계에 의해 각 후보자별로 분류)와 미분류표(기계에 의해 분류되지 않아서 사후에 공무원, 교사, 개표위원 등 개표위원에 의해 수작업으로 분류)로 나뉜다[11]. 광학분류기가 투표지를 집계하는 주요 수단이므로, 개표가 제대로 이뤄지고 있는지 검증함에 있어서 미분류표와 분류표를 비교하는 것이 필요할 것이다.

2012년 18대 대통령선거에서 상위 두 후보의 득표율은 52% 대 48%였는데, 광학분류기로 분류한 결과 96%의 분류표와 4%의 미분류표가 발생하였다. 상위 두 후보를 비교하여 살펴보면(이하 동일), 두 분류집단 간에 후보간 상대적 불균등이 존재함을 알 수 있다. 여당이었던 1번 후보자는 분류표에서는 161개 선거구(64%)에서 앞섰으며, 미분류표에서는 208개 선거구(83%)에서 앞섰다.

이 연구에서는 2.2절에서 정의한 상대득표율(K)을 통해 각 선거구별 득표율을 비교하였다. 직관적으로, 유효투표지가 광학분류기에 의해 미분류될 가능성이 후보에 상관없이 동일하다고 한다면, K값은 1에 가까워야 한다. 하지만 분석결과, 249개 선거구(99.2%)에서 K값이 1보다 크게 나타났으며, 따라서 미분류표에서 1번 후보자는 2번 후보자에 비해 항상 상대적으로 많은 득표를 한 것이다. 이러한 1번 후보자에 대한 미분류표에서의 뚜렷한 편향은 분류표에서 2번 후보자가 1번 후보자에 비해 더 많은 득표를 올린 선거구에서도 동일하게 발생한다는 점이 이 연구를 수행하게 된 주요 이유이다. 또한 상대득표율 K로 두 후보자 간의 차이를 설명하는 것이 이 연구의 목표 중에 하나이다.

이 연구의 목적은 다음 세 가지이다. 첫째, 분류표와 미분류표의 이론적 분포를 도출하고, 상대득표율 K의 기대값과 분산을 구한다(2장). 둘째, 18대 대통령선거의 사례분석을 시행하고, 2007 및 2002년 대통령선거의 결과와 비교한다. 그래서 전체 선거구에 대한 상대득표율 K를 검토하고, 실제 투표결과를 바탕으로 회귀모형(※ 전국모형)을 추정하여, 추정된 계수값(K)이 선거에서 이기는 것에 미치는 영향을 분석하였다(3장). 셋째, 시뮬레이션을 통하여, 광학분류기와 개표위원에 의해 집계되는 실제 과정에서 어떻게 전국모형과 같은 결과가 나오도록 조작할 수 있는지 살펴보고, 광학분류기의 기계적 오류의 가능성과 비교한다(4장). 이 연구에서는 발견하기 어려운 광학분류기에 의한 잘못된 개표분류의 원인을 논의하고, 이를 방지하기 위한 방법을 제안하였다(5장). 결과적으로, 이 연구를 통해 선거에서 광학분류기를 사용하는 것이 위험하다는 결론을 얻었다.

【원문 : 연구의 배경】

1. Introduction and Motivation

Many countries use electronic voting systems and such systems have shown result-changing errors through problems with software, hardware and procedures [1,14].

According to the National Election Commission (NEC) of South Korea, presidential election in South Korea has used the op-scan counters for marked paper ballots since 2002 [5,10]. The op-scan counters first sort out the paper ballots into two groups: **classified** (sorted to each candidate by the op-scanners) versus **unclassified** (unsorted first by the op-scanners but sorted later manually by counting officials which include of government officials, teachers, commissioners, etc. [11]). Since the op-scan counters are the main tools for vote counting, it is necessary to examine the unclassified against classified for a post-election investigation on their proper operations.

The 18th presidential election in South Korea in 2012, which was a close election with 52% versus 48% for the top two candidates, used the op-scan counters, producing classified (96% of total votes) and unclassified ballots (4% of total votes). Focusing on the top two candidates (hereafter), we noticed a between-candidate relative inequality in the two groups. Candidate 1 from the incumbent party won the vote counts among the classified votes in 161 districts (64%) and won the unclassified vote counts in 208 districts (83%).

We compared vote ratios in each districts using a relative ratio K, defined in section 2.2. Intuitively, if valid ballots have an equal chance of being unclassified by the op-scan counters regardless of candidates, the K-value should be close to 1. It turned out however that the K-value was larger than one in 249 districts (99.2%), and thus candidate 1 always obtained relatively higher votes than candidate 2 from the unclassified group. This apparent favor toward candidate 1 in the unclassified became the main motivation of this study as this unexpected favor for candidate 1 occurred even in electoral districts where candidate 2 received more votes than candidate 1 among the classified. One of the study purposes is to explain the observed difference between two candidates in terms of the relative ratio K.

We set up three study objectives. First, we derive theoretical distributions of the classified and unclassified ballots, and find the theoretical expectation and variance of the proposed relative ratio K (Section 2). Second, we introduce a case study on the 18th presidential election and compare the election results with two previous elections in 2007 and 2002. We then examine the relative ratio K of all districts, construct a national model for the apparent voting pattern, and analyze the impact of the model parameter (K) on winning the election (Section 3).Third, we do a simulation to demonstrate how the national model could be implemented based on the practical process of paper ballot counting by both op-scan counters and counting officials and compare to

a systematic op-scan bias (Section 4). We, discuss source of undetectable misdistributions by the op-scan counters, and suggest some bias prevention methods (Section 5). Finally, we conclude with warnings on the op-scan counters in elections.

2. 분류표와 미분류표의 이론적 분포

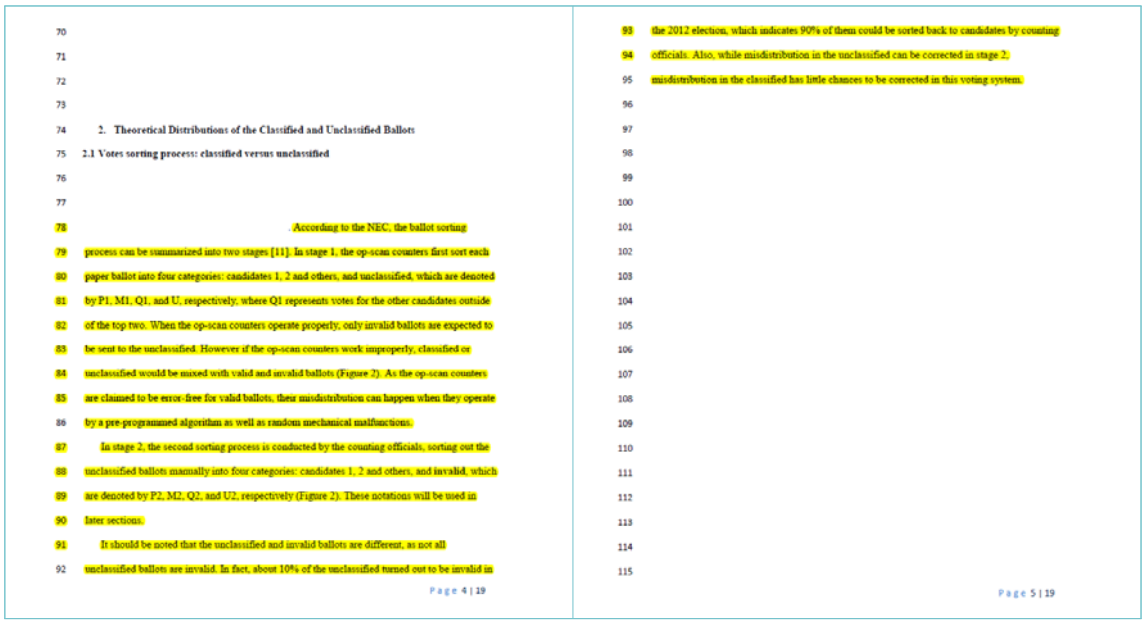
2.1 개표절차 : 분류 및 미분류

선관위에 따르면 개표절차는 크게 두 단계로 요약된다[11]. 첫 번째 단계에서, 광학분류기에 의해 투표지는 1번 후보자(P1), 2번 후보자(M1), 기타 후보자(Q1), 미분류(U)의 4개 부류로 분류된다. 광학분류기가 제대로 작동한다면, 오로지 무효표만 미분류로 보내질 것이다. 그런데, 만약 광학분류기가 제대로 작동하지 않는다면, 분류표 또는 미분류표에 유효표와 무효표가 섞일 수 있다(그림 2). 광학분류기는 유효표 분류에 오류가 없도록 요구될 것이므로, 정상적인 알고리즘에 따라 작동될 경우 잘못된 분류는 무작위적인 기계적 오차 정도만 발생할 수 있다.

두 번째 단계에서, 개표위원의 수개표를 통해 미분류표를 1번 후보자(P2), 2번 후보자(M2), 기타 후보자(Q2), 무효표(U2)로 분류한다(그림 2).

이때, 미분류표와 무효표가 다르다는 점에 유의할 필요가 있는데, 미분류표 전체가 무효표는 결코 아니다. 실제로, 2012년 대선에서 미분류표 중 최종적으로 무효표로 분류된 것은 약 10%인데, 이는 미분류표의 90%가 개표위원에 의해서 각 후보별 유효표로 재분류되었다는 것을 나타낸다. 한편, 이런 개표절차에서는, 두 번째 단계에서 미분류표로 잘못 분류된 것은 바로 잡을 수 있으나, 분류표로 잘못 분류된 것을 바로잡을 수 있는 기회가 적다는 것에도 주목할 필요가 있다.

【원문 : 개표절차 및 비분류표 발생 가능성】



2. Theoretical Distributions of the Classified and Unclassified Ballots

2.1 Votes sorting process: classified versus unclassified

According to the NEC, the ballot sorting process can be summarized into two stages [11]. In stage 1, the op-scan counters first sort each paper ballot into four categories: candidates 1, 2 and others, and unclassified, which are denoted by P1, M1, Q1, and U, respectively, where Q1 represents votes for the other candidates outside of the top two. When the op-scan counters operate properly, only invalid ballots are expected to be sent to the unclassified. However if the op-scan counters work improperly, classified or unclassified would be mixed with valid and invalid ballots (Figure 2). As the op-scan counters are claimed to be error-free for valid ballots, their misdistribution can happen when they operate by a pre-programmed algorithm as well as random mechanical malfunctions.

In stage 2, the second sorting process is conducted by the counting officials, sorting out the unclassified ballots manually into four categories: candidates 1, 2 and others, and invalid, which are denoted by P2, M2, Q2, and U2, respectively (Figure 2). These notations will be used in later sections.

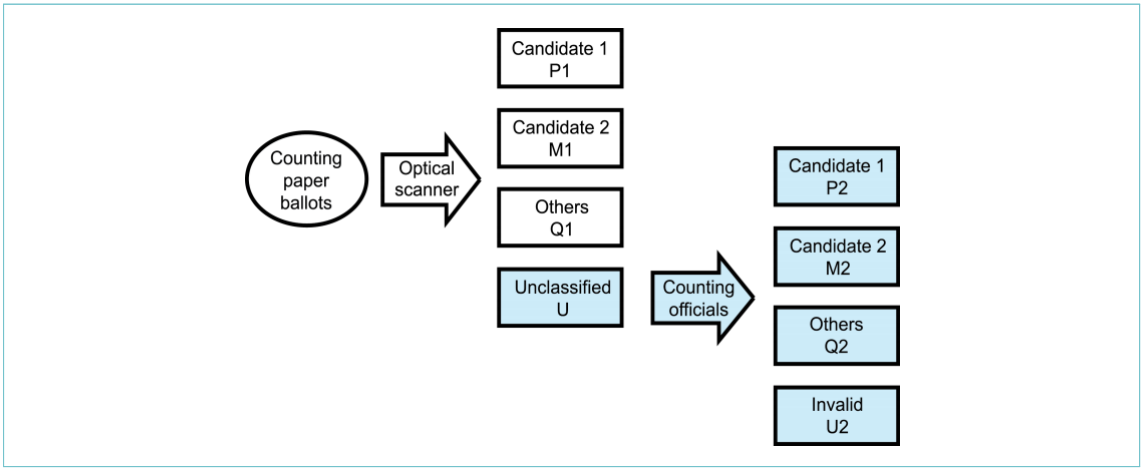
It should be noted that the unclassified and invalid ballots are different, as not all unclassified ballots are invalid. In fact, about 10% of the unclassified turned out to be invalid in

Page 4 | 19

the 2012 election, which indicates 90% of them could be sorted back to candidates by counting officials. Also, while misdistribution in the unclassified can be corrected in stage 2, misdistribution in the classified has little chances to be corrected in this voting system.

Page 5 | 19

그림 2. 미분류표 상의 유효표 및 무효표 개념



자료 : "A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea" In Event: *Poster Session: Methods for Studying Comparative Politics* (http://www.mpsanet.org/)

2.2 상대득표율(K) : 후보자간 상대적 불균등 측정 지표

임의의 선거구에 대해서, K_C를 분류표에서의 두 후보자 간 득표비 P1/M1, K_U를 미분류표에서의 두 후보자 간 득표비 P2/M2, K를 앞의 두 득표비의 비율인 상대득표율이라 정의하면, K는 그림 2의 개념에 따른 분류표와 미분류표의 함수로 정리될 수 있다.

$$K = K_U/K_C = (P2/M2)/(P1/M1)$$

미분류표에 포함되어 있던 유효표, 즉 광학분류기에 의해 그림 2의 P2, M2로 분류되었던 투표지만 고려해보자. 투표지가 충분히 공평하게 디자인되었다면, 미분류표의 유효표는 후보자와 무관하게 무작위적으로 발생하였을 것이다. 상대득표율 K는 알 수 없는 이유로 인해 광학분류기에 의해 미분류표로 구분된 유효표의 후보자간 상대적 불균등을 측정하는 지표이다.

2.2.1 K의 기대값: E[K]=1

만약 유효표가 미분류되는 것이 무작위적이라면, 즉 후보자에 상관없다면, 각 후보자의 유효표가 미분류될 확률은 같아야 하며, 이를 $Pr(U_P)=Pr(U_M)$이라 쓸 수 있다. 또한 두 후보의 총득표를 각각 P=P1+P2, M=M1+M2이라 할 수 있다. 각 유효표는 독립적으로(서로 관계없이) 분류되거나 미분류될 것이므로, P1, P2, M1, M2 모두 다음과 같은 이항분포{역주1}를 따른다고 할 수 있다.

- P1 ~ B(P, 1-r), 여기에서 B는 이항분포, 확률 $r = Pr(U_P)$.
- P2 ~ B(P, r)
- M1 ~ B(M, 1-r), 여기에서 $r = Pr(U_M) = Pr(U_P)$.
- M2 ~ B(M, r)

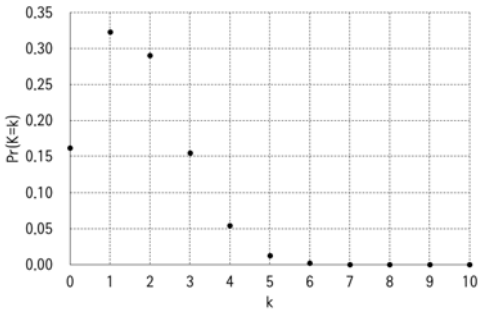
【역주1 : 이항분포】

이항분포는 어떤 것을 독립적(※각 시도끼리 어떠한 관련 없이)으로 여러 번 시도할 때, 즉 베르누이 시행일 때, 관찰하는 결과가 각 시도에서 나타날 확률이 일정하다고 여겨질 때의 확률분포이다. 그것이 성공하느냐, 실패하느냐 등의 두 가지로 나뉘기 때문에 이항(二項)이라고 표현한다. 수식으로는 달성횟수를 K, 시도횟수를 n, 한 번 시도할 때 달성 확률을 p라고 하여, K ~ B(n, p)라고 적는다.

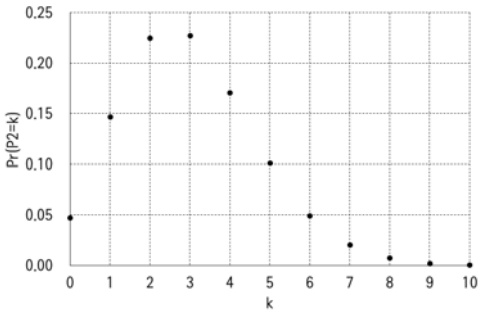
대표적인 예는 주사위를 던져서 어떤 숫자가 나오는 경우의 분포이다. 만약 1이 나오는 것을 살펴본다면, 한 번 시도할 때 1이 나올 확률은 p=1/6, 만약 10번 시도한다면 n=10, 1이 나올 횟수를 K라 하면 K ~ B(10, 1/6)이라 정리할 수 있다. 이때, K는 0부터 10까지의 경우를 지닌다. 즉 한 번도 안 나올 수도 있고, 10번 모두 나올 수도 있다. 이런 경우를 K=k, k=0, 2, 3, …, n이라고 하면, 각 경우의 확률 Pr(K=k)은 다음과 같다.

$$\Pr(K=k)=\binom{n}{k}p^k(1-p)^{n-k},\ \text{이때}\ \binom{n}{k}=\frac{n!}{k!(n-k)!}.$$

위 수식을 바탕으로 주사위를 10번 던져서 1이 나오는 횟수별 확률을 계산하여 분포도를 그리면 다음과 같다. 한 번도 안 나올 확률이 15%가 넘으며, 4번 이상 나올 확률은 5% 보다 조금 큰 정도에 불과하다.



이 논문에서 P2 ~ B(P, r)이라는 것은 1번 후보자의 미분류표 P2는 미분류표가 될 확률 r과 전체 시도횟수에 해당하는 전체 득표 P로 규정되는 이항분포를 따를 것이라는 의미이다. 만약 각 투표지가 미분류표가 될 확률이 3%라고 한다면, 100표 중에서 미분류표가 되는 각 경우(0에서 100표)의 확률은 다음과 같을 것이다. 즉, 미분류표가 하나도 안 나올 확률은 약 5%이며, 1표 가량 나올 확률은 약 15%, 2표가 나올 확률은 약 23%, 3표가 나올 확률도 약 23%, 10표 이상 나올 확률은 1%에도 미치지 못한다.



0 < p < 1, E는 기대값, $P_x(t)$가 확률생성함수일 때, X ~ B(n, p)이면,

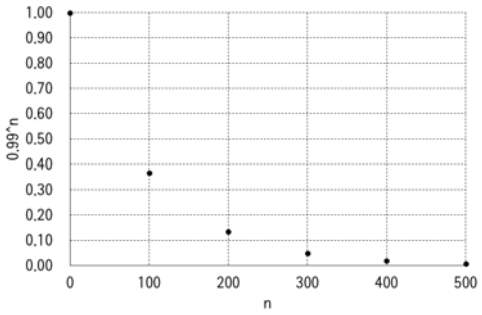
$E\left[\frac{1}{X+a}\right]=\int_0^1 t^{a-1}\cdot P_x(t)dt$이다. 따라서 q = 1 - p이면,

$E\left[\frac{1}{X+1}\right]=\int_0^1 t^0\cdot(q+pt)^n dt=\frac{1-q^{n+1}}{(n+1)p}$이다. 만약 n이 충분히 크면, 0 < q < 1일 때 q^n는 0에 수렴한다.{역주2} 그러므로 n이 충분히 크면, $E\left[\frac{1}{X+1}\right]=\frac{1}{(n+1)p}\simeq\frac{1}{np}$이 성립한다. 따라서 X가 충분히 크면, $E\left[\frac{1}{X+1}\right]\approx E\left[\frac{1}{X}\right]\approx\frac{1}{np}$임을 알 수 있다.

이를 2012년 18대 대선에 적용하면, 충분히 큰 P1, M2에 대해 $E\left[\frac{1}{P1}\right]=\frac{1}{P(1-r)}$, $E\left[\frac{1}{M2}\right]=\frac{1}{Mr}$이 성립한다.

【역주2 : 절대값이 1보다 작은 실수의 거듭제곱】

절대값이 1보다 작은 실수를 거듭제곱하면 제곱하는 횟수가 증가할수록 0에 가까워진다. 0.99를 제곱하는 경우를 살펴보면 다음과 같다. 100회만 제곱하여도 0.4보다 작아지며, 200회 정도에도 0.1에 수렴한다. 500회 정도 제곱하면 0.00657, 거의 0이라고 해도 무방한 수치가 된다.



1번 후보자의 분류표와 미분류표는 2번 후보자의 분류표와 미분류표에 대해서 독립적이므로,

$E[K]=E\left[\frac{P2/M2}{P1/M1}\right]=E\left[\frac{P2}{P1}\cdot\frac{M1}{M2}\right]=E\left[\frac{P2}{P1}\right]\cdot E\left[\frac{M1}{M2}\right]$이다. 상수 P와 M을 대입하면,

$E\left[\frac{P2}{P1}\right]=E\left[\frac{P-P1}{P1}\right]=E\left[\frac{P}{P1}-1\right]=P\cdot E\left[\frac{1}{P1}\right]-1=\frac{P}{P\cdot(1-r)}-1=\frac{r}{1-r}$,

$E\left[\frac{M1}{M2}\right]=E\left[\frac{M-M2}{M2}\right]=M\cdot E\left[\frac{1}{M2}\right]-1=\frac{M}{M\cdot r}-1=\frac{1-r}{r}$,

$E\left[\frac{P2}{P1}\right]\cdot E\left[\frac{M1}{M2}\right]=\frac{r\cdot(1-r)}{(1-r)\cdot r}=1$로 정리된다.

따라서 E[K] = 1이다.

충분히 큰 임의의 선거구에서 근사적으로 E[K]=1이라는 점에 주목할 필요가 있다. 이러한 이론적 기대값과 실제 K값의 차이는 각 후보자별로 미분류표가 얼마나 편향적으로(특정 후보자에게 얼마나 더 많이) 생성되는지를 나타내며, 따라서 미분류표의 후보자간 불균등을 측정할 수 있도록 해준다. 만약 차이가 무시할 만하다면 무시할 만한 편향(통계적으로 유의미한 편향이 없음)을 의미한다. 그렇지 않다면, 광학분류기가 정확하고 공평하게 작동되지 않았다는 합리적인 의심을 품을 수 있다.

【원문 : 상대득표율 K의 정의 및 이론적 기대값】

2.2 A proposed measure (K) of between-candidate relative inequality

For each district we let K_C, K_U and K denote three ratios, where K_C is a ratio of the two candidates' received vote counts (or rates), candidate 1/candidate 2, from the classified, K_U is that from the unclassified, and K is the relative ratio of the two ratios. Thus K is a function of the classified and unclassified:

$$K=K_U/K_C=(P2/M2)/(P1/M1),$$

using the notations in Figure 2.

We now focus on valid ballots only (excluding invalid ballots), which are unclassified by the op-scan counters such as P2 and M2 in Figure 2. As long as the paper ballot is designed fairly as shown in Figure 1, those valid ballots unclassified should be generated at random, regardless of candidate. We propose the relative ratio K as a measure of between-candidate relative inequality of their valid ballots unclassified by op-scan counters due to unknown reasons.

2.2.1 Theoretical expectation of K: E[K]=1

If valid ballots are unclassified at random, which is fair, the probability of candidate 1 or 2's valid vote to be sent to the unclassified should be the same, noted $Pr(U_D)= Pr(U_M)$. Let P=P1 + P2 and M=M1 + M2, where P and M are constants representing the total received votes of the two candidates, respectively. Since each valid vote will be either classified or unclassified independently, we know P1, P2, M1 and M2 all follow binomial distributions as follows:

- $P1 \sim B(P, 1-r)$, where B represents a binomial distribution with a probability $r = Pr(U_P)$.
- $P2 \sim B(P, r)$
- $M1 \sim B(M, 1-r)$, where $r= Pr(U_M)=Pr(U_P)$.
- $M2 \sim B(M, r)$

It is known that if $X \sim B(n, p)$, $E\left[\frac{1}{X+a}\right] = \int_0^1 t^{a-1} \cdot P_X(t)dt$, where $0 < p < 1$, E represents expectation, and $P_X(t)$ is the probability generating function. We thus have

$E\left[\frac{1}{X+1}\right] = \int_0^1 t^0 \cdot (q+pt)^n dt = \frac{1-q^{n+1}}{(n+1)p}$; where $q = 1-p$) [2]. If $n \to \infty$, *then* $q^n \to 0$ *as* $0 < q < 1$. Thus $E\left[\frac{1}{X+1}\right] = \frac{1}{(n+1)p} \cong \frac{1}{np}$ as $n \to \infty$.

Also for large X, we see $E\left[\frac{1}{X+1}\right] \approx E\left[\frac{1}{X}\right] \cong \frac{1}{np}$.

Applying to the 18th presidential election in 2012, we have

$E\left[\frac{1}{P_1}\right] \cong \frac{1}{P(1-r)}$ and $E\left[\frac{1}{M_2}\right] \cong \frac{1}{Mr}$ for large P1 and M2.

Since candidate 1's classified versus unclassified votes are independent from candidate 2's,

$$E[K] = E\left[\frac{P_2/_{M_2}}{P_1/_{M_1}}\right] = E\left[\frac{P_2}{M_2}\cdot\frac{M_1}{P_1}\right] = E\left[\frac{P_2}{P_1}\right]\cdot E\left[\frac{M_1}{M_2}\right]$$

As P and M are constants, we get

$$E\left[\frac{P_2}{P_1}\right] = E\left[\frac{P-P_1}{P_1}\right] = E\left[\frac{P}{P_1}-1\right] = N\cdot E\left[\frac{1}{P_1}\right]-1 = \frac{P}{P(1-r)}-1 = \frac{r}{(1-r)},$$

$$E\left[\frac{M_1}{M_2}\right] = E\left[\frac{M-M_2}{M_2}\right] = M\cdot E\left[\frac{1}{M_2}\right]-1 = \frac{M}{Mr}-1 = \frac{1-r}{r},$$

$$E\left[\frac{P_2}{P_1}\right]\cdot E\left[\frac{M_1}{M_2}\right] = \frac{r(1-r)}{(1-r)r} = 1.$$

Therefore $E[K] = 1$.

Note that $E[K] = 1$ for any electoral district if its size (number of voters) is large enough to use the asymptotic approach. The difference between the theoretical expectation and observed K-

value is an indicator of systematic bias in generation of the unclassified votes of each candidate and offers a measure of between-candidate relative inequality with respect to the unclassified ballots. Negligible difference implies negligible bias. Otherwise, one can raise a reasonable doubt on the accurate and fair operation of op-scan counters.

2.2.2 선거구 크기에 따른 상대득표율 K의 분산

일반적으로 $E\left[\frac{1}{(X+1)^a}\right]$을 이항변수 X와 1보다 큰 상수 a에 대해서 정리할 수 없으나, Cribari-Neto 등[3]이 제시한 다음과 같은 근사적 해를 사용할 수 있다.

X가 충분히 클 때, $E\left[\frac{1}{(X+1)^a}\right]=E\left[\frac{1}{X^a}\right]=(np)^{-a}+\left(\frac{a-1}{2p}-\frac{a+1}{2}\right)\frac{\Gamma(a+1)}{\Gamma(a)}\frac{1}{n^{a+1}p^a}$.

P1 ~ B(P, 1-r), M2 ~ B(M, r)이므로, 위 식에서 다음을 얻을 수 있다.

$E\left[\frac{1}{(P1)^2}\right]\cong\frac{1}{(P(1-r))^2}\left(1+\frac{1}{P(1-r)}-\frac{3}{P}\right)$, $E\left[\frac{1}{(M2)^2}\right]\cong\frac{1}{(Mr)^2}\left(1+\frac{1}{Mr}-\frac{3}{M}\right)$.

만약 Y와 Z가 독립변수라면 분산은 다음과 같다.

$Var(Y)=E[Y^2]-E[Y]^2$,

$Var(YZ)=Var(Y)Var(Z)+Var(Y)E[Z]^2+Var(Z)E[Y]^2$.

위 식에 상대득표율 K를 대입하면 다음과 같은 근사식을 얻을 수 있는데

$Var(K)\cong\frac{(P+M)r(1-r)+1}{PM(r(1-r))^2}\cong\frac{(P+M)}{PM(1-r)}$, 이를 살펴보면 상대득표율의 분산은 미분류율 r뿐만 아니라 선거구에서의 각 후보의 득표수(또는 선거구의 규모)의 영향도 받는다. 즉 상대득표율의 분산은 선거구가 커질수록 작아지며, 이는 선거구의 규모가 커질수록 K값이 기대값에 수렴해야 함을 의미한다. 이런 상대득표율의 특성은 광학분류기를 사용하는 모든 경우에 적용될 수 있다.

2.2.3 상대득표율 K의 정규성 검정(Lognormal test)

지금까지 상대득표율의 평균과 분산을 확인하였다. 이 절에서는 상대득표율 K의 확률분포를 시뮬레이션 분석을 통해 확인하고자 한다. 상대득표율이 항상 양수이므로, 로그정규분포에 부합하는지를 검토하였다. 즉 만약 K값이 로그정규분포를 따른다면, Log(K)값을 바탕으로 관측된 K값과 이론적 기대값의 차이를 검정할 수 있다. i선거구의 상대득표율 K_i의 평균을 1, 분산을 V_i라 하면, 로그정규식 $\log(K_i)\sim N(\mu_i,\Sigma_i)$에서 $\mu_i=-\frac{1}{2}log(1+V_i)$, $\Sigma_i=\log(1+V_i)$이다. 여기에서 V_i가 충분히 작을 경우 $\log(1+V_i)\approx V_i$이므로, $\mu_i=-\frac{1}{2}V_i$, $\Sigma_i=V_i$이다. 이를 활용하여 상대득표율 K에 대한 시뮬레이션 자료가 로그정규분포를 따르는지 검정하였다.

시뮬레이션 자료는 (1) 선거구의 크기는 1,000명부터 100,000명까지 1000명씩 증가, (2) 1번 후보자의 득표율은 0.2부터 0.8까지 0.1씩 증가, (3) 미분류율은 0.02부터 0.15까지 0.01씩 증가하게 생성하였다. 위 세 가지 조건을 조합한 5,000번의 시뮬레이션을 수행하여, 각 시뮬레이션 별 log(K)에 대해 Shapiro-Wilk[13]의 정규성 검정{역주3}을 하였다. 그 결과 선거구의 크기가 10,000명 이상, 1번 후보자 득표율 0.3 이상, 미분류율이 0.03 이상일 때 상대득표율이 로그정규분포를 따르는 것으로 나타났다. 이런 조건이 만족되지 않을 경우에는 Fisher의 정확확률검정{역주4}을 통해 두 후보간 불균등성을 직접 검정할 수도 있다.

【역주3 : Shapiro-Wilk 정규성 검정】

표본이 정규분포를 따르는지 검정하는 방법의 하나이며, 검정통계량은 다음과 같다.

$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, 여기에서 $x_{(i)}$는 i번째 순서통계량, $\bar{x}$는 표본평균, a_i는 표준정규분포에서 유도되는 상수값이다. 이 검정에서 귀무가설은 표본이 정규분포를 따른다는 것이며, 실제 검정은 다음 사례와 같이 이뤄지는데, 이 사례에서 검정통계량 W값은 0.99628, p값은 0.80235이므로 표본이 정규분포를 따른다는 귀무가설을 기각하기 어렵다. 만약 p값이 0.05 미만이라면 신뢰수준 95%에서 표본이 정규분포를 따른다는 귀무가설을 기각할 수 있을 것이다.

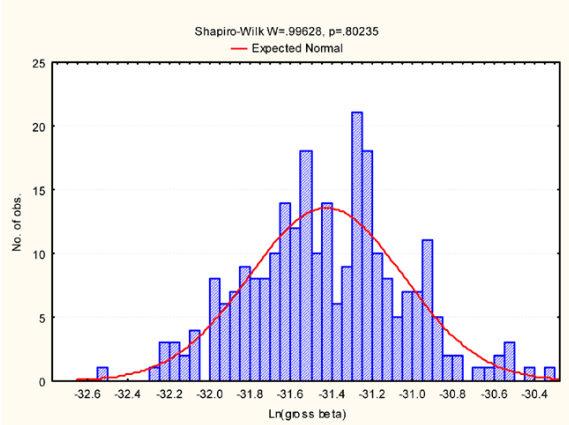


Figure B-3. Test of log normality for Arco gross beta.

자료: http://www.gsseser.com/annuals/2003/AppendixB.htm

【역주4 : Fisher의 정확확률검정】

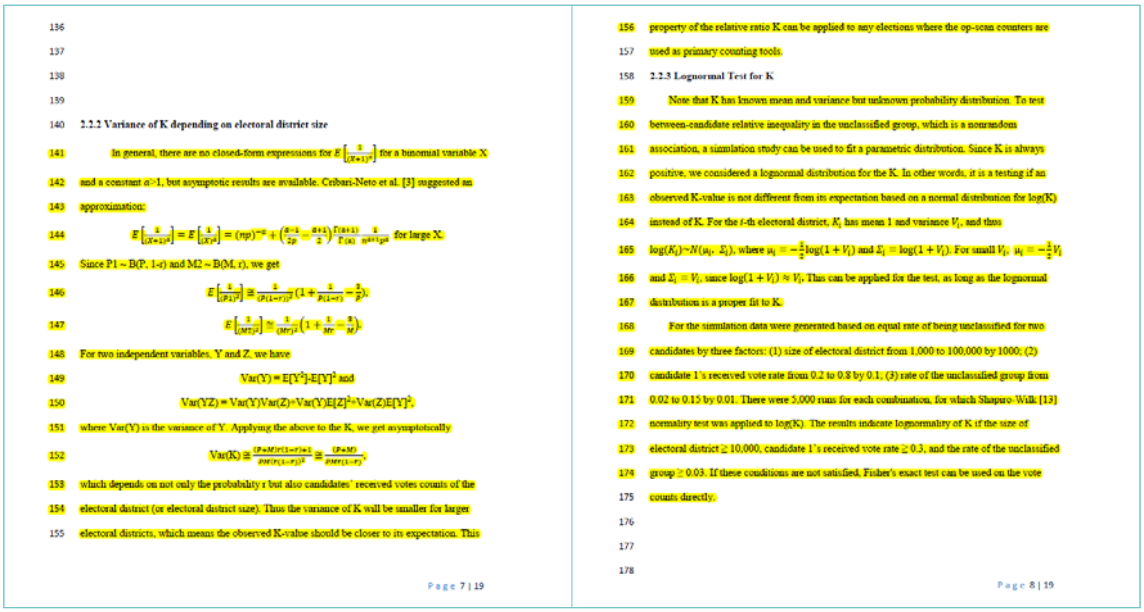
표본수가 적은 경우, 2개의 범주로 분류된 자료에서 범주별 통계적 차이를 검정하는 기법이다. 다음과 같은 자료가 있을 때 정확확률(p-value)은 $p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$로 직접 계산된다.

		범주 1		계
		true	false	
범주2	yes	a	b	a+b
	no	c	d	c+d
계		a+c	b+d	n

위 자료를 이 논문의 주제와 연관 지어 다음과 같다고 가정할 경우 정확확률 p는 약 0.017이다. 이는 분류표의 후보자별 득표율과 미분류표의 후보자별 득표율이 같을 확률이 1.7%라는 의미이며, 95% 신뢰수준에서 미분류표에서의 득표율이 분류표에서의 득표율과 다르다고 해석할 수 있다.

	분류표	미분류표	계
후보1	51	9	60
후보2	48	1	49
계	99	10	109

【원문 : 상대득표율 K의 분산 및 정규성 검정】



2.2.2 Variance of K depending on electoral district size

In general, there are no closed-form expressions for $E\left[\frac{1}{(X+1)^2}\right]$ for a binomial variable X and a constant $a>1$, but asymptotic results are available. Cribari-Neto et al. [3] suggested an approximation:

$$E\left[\frac{1}{(X+1)^a}\right] = E\left[\frac{1}{(X)^a}\right] = (np)^{-a} + \left(\frac{a-1}{2p} - \frac{a(a+1)}{2}\right)\frac{\Gamma(a+1)}{\Gamma(a)}\frac{1}{n^{a+1}p^a} \text{ for large X.}$$

Since P1 ~ B(P, 1-r) and M2 ~ B(M, r), we get

$$E\left[\frac{1}{(P1)^2}\right] \cong \frac{1}{(P(1-r))^2}\left(1+\frac{3}{P(1-r)}-\frac{3}{P}\right),$$

$$E\left[\frac{1}{(M2)^2}\right] \cong \frac{1}{(Mr)^2}\left(1+\frac{3}{Mr}-\frac{3}{M}\right).$$

For two independent variables, Y and Z, we have

$$Var(Y) = E[Y^2]-E[Y]^2 \text{ and}$$

$$Var(YZ) = Var(Y)Var(Z)+Var(Y)E[Z]^2+Var(Z)E[Y]^2,$$

where Var(Y) is the variance of Y. Applying the above to the K, we get asymptotically

$$Var(K) \cong \frac{(P+M)(1-r)+1}{PMr(1-r)^2} \cong \frac{(P+M)}{PMr(1-r)},$$

which depends on not only the probability r but also candidates' received votes counts of the electoral district (or electoral district size). Thus the variance of K will be smaller for larger electoral districts, which means the observed K-value should be closer to its expectation. This property of the relative ratio K can be applied to any elections where the op-scan counters are used as primary counting tools.

Page 7 | 19

2.2.3 Lognormal Test for K

Note that K has known mean and variance but unknown probability distribution. To test between-candidate relative inequality in the unclassified group, which is a nonrandom association, a simulation study can be used to fit a parametric distribution. Since K is always positive, we considered a lognormal distribution for the K. In other words, it is a testing if an observed K-value is not different from its expectation based on a normal distribution for log(K) instead of K. For the i-th electoral district, K_i has mean 1 and variance V_i, and thus $\log(K_i) \sim N(\mu_i, \Sigma_i)$, where $\mu_i = -\frac{1}{2}\log(1+V_i)$ and $\Sigma_i = \log(1+V_i)$. For small V_i, $\mu_i = -\frac{1}{2}V_i$ and $\Sigma_i = V_i$, since $\log(1+V_i) \approx V_i$. This can be applied for the test, as long as the lognormal distribution is a proper fit to K.

For the simulation data were generated based on equal rate of being unclassified for two candidates by three factors: (1) size of electoral district from 1,000 to 100,000 by 1000; (2) candidate 1's received vote rate from 0.2 to 0.8 by 0.1; (3) rate of the unclassified group from 0.02 to 0.15 by 0.01. There were 5,000 runs for each combination, for which Shapiro-Wilk [13] normality test was applied to log(K). The results indicate lognormality of K if the size of electoral district $\geq$ 10,000, candidate 1's received vote rate $\geq$ 0.3, and the rate of the unclassified group $\geq$ 0.03. If these conditions are not satisfied, Fisher's exact test can be used on the vote counts directly.

Page 8 | 19

3. 사례분석: 2012년 18대 대통령선거

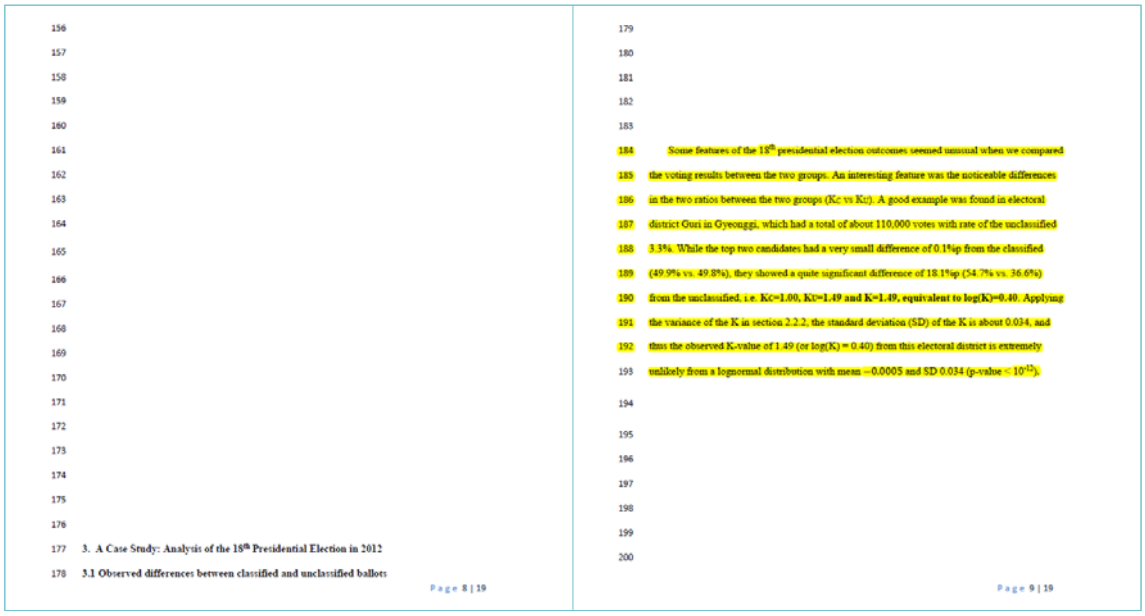
3.1 분류표와 미분류표의 실제 차이

분류표와 미분류표를 비교해보면 18대 대선결과에서 보여지는 수치는 비정상적으로 여겨진다. 두 그룹에서의 후보자간 득표비(K_C vs K_U)에 뚜렷한 차이가 있다. 경기도의 구리시가 좋은 사례인데, 약 110,000명이 투표를 하여, 3.3%의 미분류표가 발생하였다. 분류표에서 상위 두 후보의 득표율차이는 0.1%(49.9% vs 49.8%)로 매우 작았으나, 미분류표에서는 18.1%(54.7% vs 36.6%)의 뚜렷한 차이를 보이고 있는데, 즉 K_C=1.00, K_U=1.49, K=1.49, Log(K)=0.40이다. 2.2.2절에서 유도한 수식으로 K의 표준편차를 구하면 0.034이므로 이 선거구의 K값 1.49는 $\mu = -0.0005$, $\Sigma = 0.034$인 로그정규분포를 가정할 때 매우 나타나기 어려운 수치이다(p-value < 10^{-12})[역주5].

【역주5 : p-value < 10^{-12}】

K가 $\mu = -0.0005$, $\Sigma = 0.034$인 로그정규분포를 따른다면, K값이 1.49일 확률이 0.0000000001%보다 작다는 의미이다. 정상적인 수치가 아니라는 것을 의미한다.

【원문 : 분류표와 미분류표의 실제 차이】



3. A Case Study: Analysis of the 18th Presidential Election in 2012

3.1 Observed differences between classified and unclassified ballots

Page 8 | 19

Some features of the 18th presidential election outcomes seemed unusual when we compared the voting results between the two groups. An interesting feature was the noticeable differences in the two ratios between the two groups (K_C vs K_U). A good example was found in electoral district Ouri in Gyeonggi, which had a total of about 110,000 votes with rate of the unclassified 3.3%. While the top two candidates had a very small difference of 0.1%p from the classified (49.9% vs. 49.8%), they showed a quite significant difference of 18.1%p (54.7% vs. 36.6%) from the unclassified, i.e. K_C=1.00, K_U=1.49 and K=1.49, equivalent to log(K)=0.40. Applying the variance of the K in section 2.2.2, the standard deviation (SD) of the K is about 0.034, and thus the observed K-value of 1.49 (or log(K) = 0.40) from this electoral district is extremely unlikely from a lognormal distribution with mean −0.0005 and SD 0.034 (p-value < 10^{-12}).

Page 9 | 19

3.2 최근 대선에서 상대적득표 차이

표 1은 3개 선거구에서의 두 그룹(분류표 및 미분류표)간 차이를 나타낸다. 3개 선거구에서 16대 및 17대 대선의 K_C와 K_U를 비교해보면, K값은 이론적 기대값인 1에 가깝다(1.02 및 1.04). 예를 들어, G선거구에서 상위 두 후보는 분류표에서 각각 36.3%, 56.5%, 미분류표에서 각각 33.7%, 50.3%를 득표하였다. Y선거구에서는 분류표에서 각각 16.9%, 59.6%, 미분류표에서 각각 17.5%, 59.4%를 득표하였다.

반면에 18대 대선에서는 같은 3개 선거구에서 1보다 큰 K값(1.35-1.44)을 보이고 있다. 18대 대선의 총 251개 선거구에서 상대득표율(K)은 0.97~2.17이며, 평균은 1.48이다.

표 1. 2002, 2007, 2012대선에서 3개 선거구의 개표결과

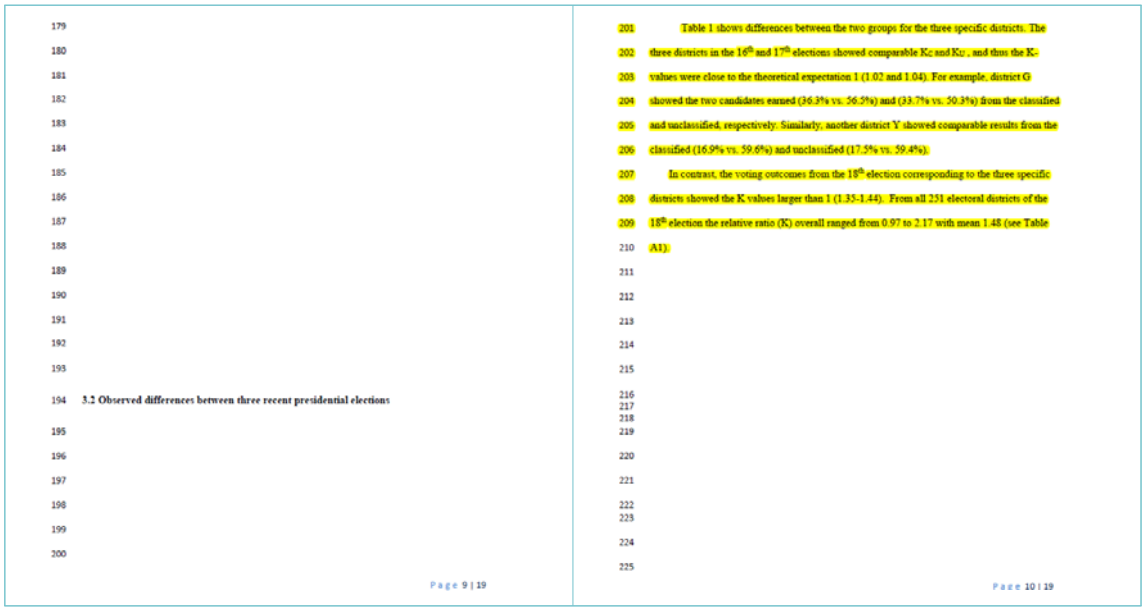
Voting results for three specific districts from three presidential elections in 2002, 2007, 2012

	Votes from the classified (sorted by op-scanners)		Votes from the unclassified (sorted by counting officials)		
Election year and district	candidate 1 (P_C)[a]	candidate 2 (M_C)[a]	candidate 1 (P_U)[b]	candidate 2 (M_U)[b]	**K=K_U/K_C**
16th in 2002, district G	36.3%	56.5%	33.7%	50.3%	**1.04**
17th in 2007, district N	23.1%	47.7%	25.1%	50.7%	**1.02**
district Y	16.9%	59.6%	17.5%	59.4%	**1.04**
18th in 2012, district G	40.2%	59.4%	39.9%	43.7%	**1.35**
district N	46.3%	53.3%	45.0%	36.0%	**1.44**
district Y	51.9%	47.8%	54.4%	36.6%	**1.37**

[a] P_C= (votes for candidate 1/total votes); M_C= (votes for candidate 2/total votes) from classified
[b] P_U= (candidate 1/total votes); M_U= (candidate 2/total votes) from unclassified; K_U= P_U /M_U

자료 : "A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea" In Event: *Poster Session: Methods for Studying Comparative Politics* (http://www.mpsanet.org/)

【원문 : 최근 대선에서 상대적득표 차이】



3.2 Observed differences between three recent presidential elections

Page 9 | 19

Table 1 shows differences between the two groups for the three specific districts. The three districts in the 16th and 17th elections showed comparable K_C and K_U, and thus the K-values were close to the theoretical expectation 1 (1.02 and 1.04). For example, district G showed the two candidates earned (36.3% vs. 56.5%) and (33.7% vs. 50.3%) from the classified and unclassified, respectively. Similarly, another district Y showed comparable results from the classified (16.9% vs. 59.6%) and unclassified (17.5% vs. 59.4%).

In contrast, the voting outcomes from the 18th election corresponding to the three specific districts showed the K values larger than 1 (1.33-1.44). From all 251 electoral districts of the 18th election the relative ratio (K) overall ranged from 0.97 to 2.17 with mean 1.48 (see Table A1).

Page 10 | 19

3.3 전국모형

상대득표율 K는 개표가 독립적으로 이뤄지는 개별 선거구에 적용하도록 고안한 것이다. 이에 더하여, 이 연구에서는 18대 대선의 모든 선거구에 대한 전국상대득표율 K_N에 대해서도 다루었다.

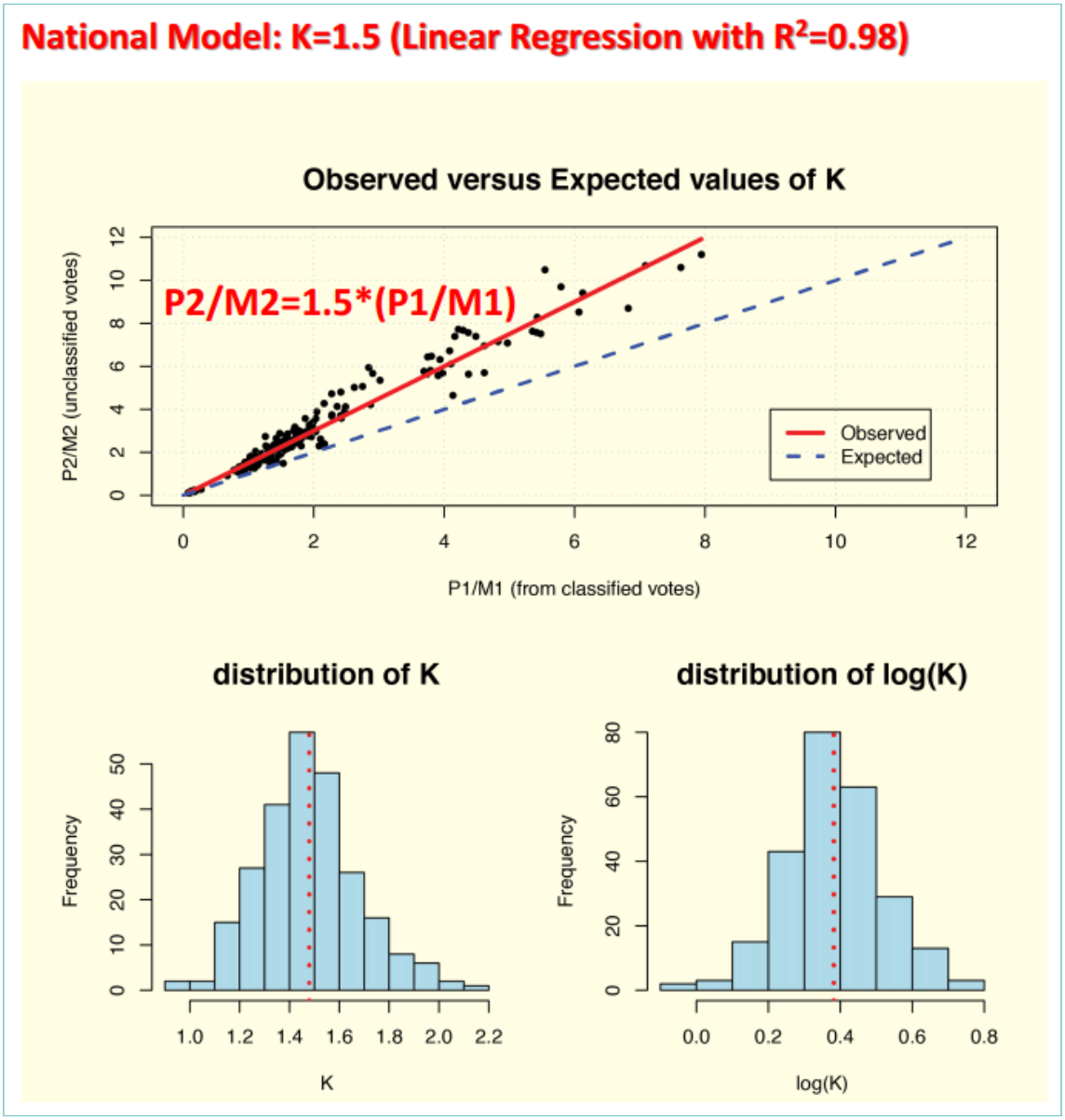
3.3.1 18대 대선 251개 선거구에 대한 전국상대득표율

$K=K_U/K_C$이므로, 우선 $K_U=P2/M2$와 $K_C=P1/M1$ 사이의 상관성을 검토하였다. K_C를 x축, K_U를 y축에 두어 251개 선거구의 산점도를 그리면 그림 3과 같은데, 여기에서 보여지는 기울기가 전국상대득표율 K_N을 의미한다.

251개 선거구 K값의 평균, 중앙값, 역분산가중평균은 각각 1.48, 1.47, 1.45이므로 대칭분포라 할 수 있다(그림 3). 따라서 K_N의 추정치는 1.5이다. 이를 수식으로 표현하면 다음과 같이 간략히 쓸 수 있다.{역주6}

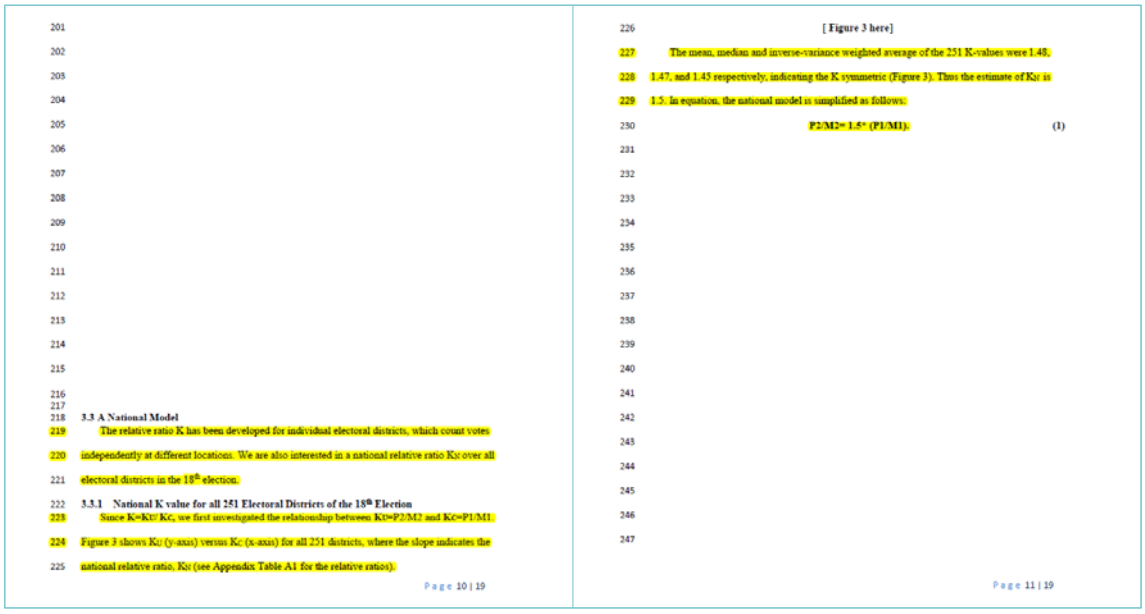
$$P2/M2=1.5*(P1/M1) \qquad (1)$$

그림 3. 전국모형: K=1.5 (선형회귀모형. R^2=0.98)



자료 : "A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea" In Event: *Poster Session: Methods for Studying Comparative Politics* (http://www.mpsanet.org/)

【원문 : 18대선 전국상대득표율】



3.3 A National Model

The relative ratio K has been developed for individual electoral districts, which count votes independently at different locations. We are also interested in a national relative ratio K_N over all electoral districts in the 18th election.

3.3.1 National K value for all 251 Electoral Districts of the 18th Election

Since K=K_U/ K_C, we first investigated the relationship between K_U=P2/M2 and K_C=P1/M1. Figure 3 shows K_U (y-axis) versus K_C (x-axis) for all 251 districts, where the slope indicates the national relative ratio, K_N (see Appendix Table A1 for the relative ratios).

Page 10 | 19

[Figure 3 here]

The mean, median and inverse-variance weighted average of the 251 K-values were 1.48, 1.47, and 1.45 respectively, indicating the K symmetric (Figure 3). Thus the estimate of K_N is 1.5. In equation, the national model is simplified as follows:

$$P2/M2 = 1.5^{*}(P1/M1). \quad (1)$$

Page 11 | 19

【역주6 : 전국모형에 대한 선관위의 공식적인 해명?】

그림 3에서 알 수 있듯이 전국모형 P2/M2=1.5*(P1/M1)은 결정계수(R^2)가 98%에 달할 정도로 설명력이 높다. 즉 이 모형을 적용하면 분류표에서의 득표차이만으로 미분류표에서의 득표차이를 거의 정확하게 예측할 수 있다. 그런데 이에 대한 선관위의 공식해명(2017.4.19; http://nec.go.kr)은 다음과 같다.

> 명확하지 않은 기표로 인하여 미분류 처리된 투표지의 원인은 여러 가지가 있겠으나 연령이 특히 중요한 요소로 작용하고 있음. 실제 제18대 대선 결과를 보면 노년층이 많은 시골지역(군단위)의 미분류율은 5% 초반대로 청년층이 많은 도시지역(시지역)의 2% 후반대 보다 1.8배 정도 높게 나타나는바, 이는 노년층의 투표에서 미분류표로 처리되는 비율이 청년층보다 더 높다는 것을 의미함.
>
> ※ 통계청의 2012년 인구통계를 보면 20대 이상 주민수에서 60대 이상이 차지하는 비율이 郡지역(39.4%)이 區지역(20.5%) 보다 1.9배 높음.
>
> 지난 대선에서 방송3사(KBS, MBC, SBS) 출구조사 결과 박근혜 후보자의 예상득표율은 20대에서는 33.7%, 30대는 33.1%인 반면, 50대에서는 62.5%, 60대 이상에서도 72.3%로 나타나 50대 이상 연령층에서 박근혜 후보자의 예상득표율이 높았다는 사실을 확인할 수 있고, 이를 통해 노년층의 투표지가 더 많이 미분류 처리되었을 것이라는 사실과 미분류된 투표지에서 박근혜 후보자의 상대득표율이 정상 분류된 투표지에서보다 더 높게 나올 수밖에 없었다는 것을 알 수 있음.
>
> ※ 노년층 지지율이 높은 후보자의 득표율이 미분류표에서 높아지는 현상은 다른 선거에서도 동일하게 나타남.

요약하자면, '1) 어르신들이 많은 선거구에서 미분류표가 많다(기표 실수가 많다). 2) 어르신들이 1

번 후보자를 지지하는 비율이 높기 때문에 당연히 미분류표에서 더 많이 득표하였다'는 것이다.

위 해명이 얼핏 그럴듯하게 들리지만, 만약 그렇다면 그림 3의 전국모형처럼 분류표에서의 득표차이로 미분류표에서의 득표차이를 예측할 수 있는 모형의 설명력이 높게 나올 수 없다. 선관위의 해명을 모형식으로 표현하면 다음과 같을 것이다.

P2/M2=a*노인인구비율+b

경험상 매우 높은 설명력을 보였던 P1/M1 대신에 노인인구비율을 넣었을 때 추정되는 a가 통계적으로 유의미할 가능성은 낮아 보인다. 영화 '더 플랜'의 영상에서 개표결과를 확인할 수 있는 18개 선거구의 개표결과를 바탕으로 선관위 해명 'P2/M2=a*노인인구비율(60세 이상)+b'의 a를 추정해보면 다음과 같다. 예상한대로 모형의 설명력은 0.027(2.7%)에 불과하며, 당연히 2.001로 추정된 a값의 유의확률도 0.513이나 된다. 이는 이 모형에서 a가 0일 확률이 51.3%에 달한다는 의미이다. 이 변수를 넣는 것이 쓸데없다는 통계적 설명이다. 재미있는 점은 y절편(b)이 1.534(p값 0.021)로 추정되었다는 것인데, 해석하자면 노인인구비율과 상관없이 P2/M2는 1.5로 고정되어 있다는 것이다. 즉 항상 1번 후보자가 2번 후보자에 비해 미분류표에서 1.5배 득표하고 있다는 것이다. 노인인구와 관계없다.

요약 출력

회귀분석 통계량	
다중 상관계수	0.165
결정계수(R-sq)	0.027
조정된 결정계수	-0.034
표준 오차	0.895
관측수	18.000

분산 분석

	자유도	제곱합	제곱 평균	F 비	유의한 F
회귀	1	0.358	0.358	0.447	0.513
잔차	16	12.808	0.801		
계	17	13.166			

	계수	표준 오차	t 통계량	P-값	하위 95%	상위 95%	하위 95.0%	상위 95.0%
Y 절편	1.534	0.600	2.557	0.021	0.262	2.805	0.262	2.805
X 1	2.001	2.994	0.669	0.513	-4.345	8.347	-4.345	8.347

통계적 검증을 위한 DB 구성

개표소	분류표 득표율		미분류표 득표율		(P2/M2)/(P1/M1) ※ 상대적 득표율(K)	2012년 60세 이상 인구비율
	박근혜(P1)	문재인(M1)	박근혜(P2)	문재인(M2)		
서울 성동구	86,412	93,598	2,243	1,631	1.49	0.162
부산 금정구	95,025	58,837	2,530	1,040	1.51	0.191
경기 군포시	77,673	91,361	2,738	2,294	1.40	0.156
경기 김포시	83,057	70,336	3,319	1,894	1.48	0.123
경기 남양주	160,389	152,755	5,932	3,834	1.47	0.112
경기 소사구	61,075	70,710	1,763	1,329	1.54	0.138
경기 오정구	46,948	54,826	1,745	1,323	1.54	0.126
경기 원미구	114,382	137,157	3,127	2,707	1.39	0.141
경기 분당구	156,328	137,280	4,209	3,014	1.23	0.119
경기 수정구	58,582	71,075	3,719	3,221	1.40	0.145
강원 강릉시	79,489	40,122	2,994	941	1.61	0.156

강원 고성군	10,904	5,229	997	431	1.11	0.208
강원 원주시	99,485	71,028	2,947	1,380	1.52	0.185
강원 삼척시	26,514	13,053	1,054	295	1.76	0.245
강원 속초시	27,619	16,323	918	304	1.76	0.213
강원 양구군	7,675	4,434	369	137	1.56	0.279
강원 양양군	11,151	5,300	382	147	1.24	0.297
전남 신안군	2,479	20,136	236	1,275	1.50	0.380

1) 개표수치는 영화 <더 플랜>에서 인용한 것이다. 화면상 수치를 옮겨 적은 것이라 잘못 기입 된 것이 있을 수 있으며, 전체 자료를 확인하지는 못하였다.
2) 60세 이상 인구비율은 2012년 연령별 주민등록인구 기준이다.

참고로, 위 자료를 바탕으로 'P2/M2=a*(P1/M1)+b'를 추정하면 다음의 그림과 같다. 18개 선거구만 넣었음에도 결정계수가 0.876에 달할 정도이다. 위 자료로 추정된 상대득표율 K는 1.468(p값 0.000)이다.

요약 출력

회귀분석 통계량	
다중 상관계수	0.936
결정계수	0.876
조정된 결정계수	0.868
표준 오차	0.320
관측수	18

분산 분석

	자유도	제곱합	제곱 평균	F 비	유의한 F
회귀	1	11.576	11.576	113.147	0.000
잔차	16	1.637	0.102		
계	17	13.213			

	계수	표준 오차	t 통계량	P-값	하위 95%	상위 95%	하위 95.0%	상위 95.0%
Y 절편	0.010	0.194	0.050	0.961	-0.401	0.420	-0.401	0.420
X 1	1.468	0.138	10.637	0.000	1.175	1.760	1.175	1.760

3.3.2 전국상대득표율이 개표결과에 미치는 영향

수식 (1)의 전국모형이 뜻하는 것처럼 분류표에서 누가 득표를 많이 했는지에 상관없이 1번 후보자가 2번 후보자에 비해 미분류표에서 일정한 비율로 더 많은 득표를 하였다면, 이런 의문을 품게 된다. '개표결과에 전국상대득표율이 미치는 영향은 무엇일까?'

의문을 해소하기 위해, L=P1/M1, P2/M2=K*(P1/M1)이라 하면, K≥1, L>0이다. 그러면 P2/M2=K*L이고, 따라서 P1=L*M1, P2=K*L*M2이다. 간략한 논의를 위해, 1번 후보자가 선거에서 이기는 상황만 고려해보았다. 1번 후보자가 2번 후보자에 비해 많은 득표를 하였다면, 이는 명확하게 (P1+P2)>(M1+M2)임을 뜻하며, 따라서 1번 후보자가 이기는 조건은 다음과 같다.

(L-1)*M1+(K*L-1)*M2>0 (2)

미분류표에 의해서 1번 후보자가 이기는 유일한 조건은 다음과 같다. 만약 L<1(즉, P1<M1 ※분류표에서는 2번 후보자가 이김)이면, (K*L-1)*M2>(1-L)*M2이다. 따라서 M2>0이므로, (K*L-1)>(1-L)*(M1/M2)이다. 부등식을 간략하게 하기 위해, x=M1/M2라 하면, x는 2번 후보자의 미

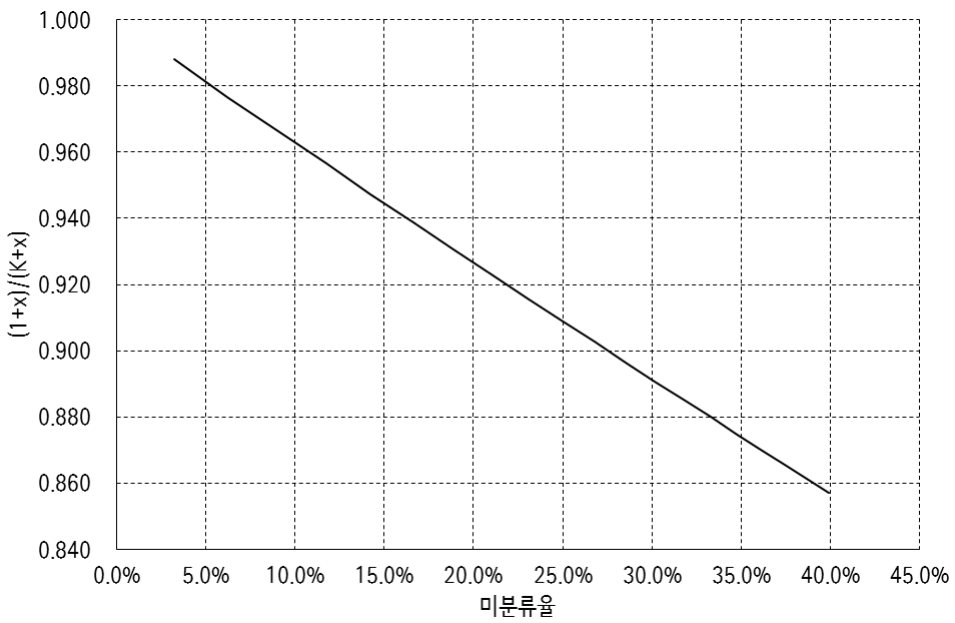
분류표(M2)에 대한 분류표(M1)의 비율을 뜻한다. 그렇다면, 식 (2), 즉 1번 후보자가 이기는 조건을 x와 K로 정리하여 L>(1+x)/(K+x)라 할 수 있다. 이 조건에 대해서 좀 더 논의하면 다음과 같다.

1번 후보자의 승리는 x와 K에 따라 결정된다. K를 상수(K=1.5)로 두고 x를 변수로 두어 K의 영향을 분석해보자. 2번 후보자의 분류표에서의 득표가 미분류표에서의 득표보다 크므로(즉, 18대 대선에서 M1>M2이므로), x의 최소값은 1이다(x≥1).

그림 4에서 보여지듯이(※ 이런 조건에서는), 수식 (1)의 전국모형(※ K=1.5)은 미분류표가 늘어날수록 1번 후보자가 이길 가능성을 높이게 된다.{역주7} 미분류표는 광학분류기에 의해서 발생하므로, K 추정치는 광학분류기의 사용과 연관되어 있는 것이다. K 추정치가 1번 후보자가 선거에서 이기는 것에 주는 영향은 1번 후보자가 분류표에서 지는 상황일 때 극대화될 것이다.

18대 대선에서, x는 약 35였으며, L약 0.99였으므로, 1번 후보자가 분류표에서 2번 후보자에 비해 적게 득표하였지만 미분류표에서 더 많이 득표하여 선거에서 이긴 것이다. 이는 K=1.5로 설정됨에 따른 결과일 것이며, 18대 대선에서의 1.5라는 추정치가 지니는 명확한 의미이다.

그림 4. 미분류율과 (1+x)/(K+x)의 관계

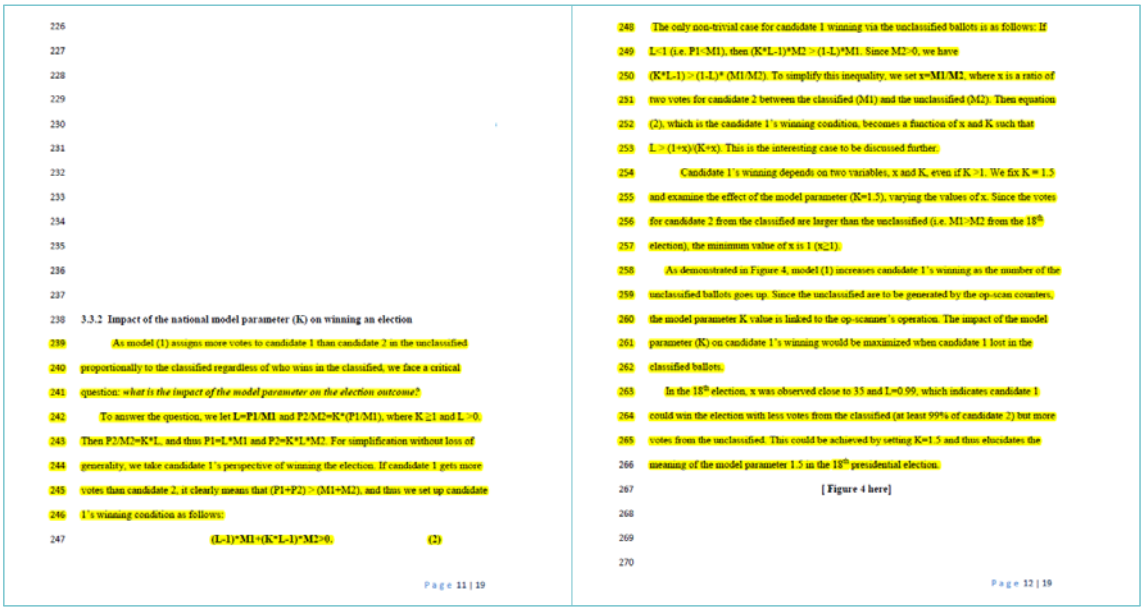


※ 원문에서 'Figure 4'를 확인하지 못하여, P1, M1을 상수로 두고, K=1.5일 때 M2에 따른 미분류율{=2.5M2/(2.5M2+M1+P2)}을 x축, (1+x)/(K+x)를 y축으로 두어 그래프를 그린 것임(※ 원문의 그림과 다를 수 있음). 미분류율이 증가함에 따라 (1+x)/(K+x)는 낮아지므로, 1번 후보자가 이길 가능성이 높아짐.

【역주7】

그림 4를 살펴보면, 분류표에서 1번 후보자가 2% 가량 뒤쳐지더라도(L=0.98) 미분류표가 5% 이상 발생하면, K=1.5이므로 최종적으로는 1번 후보자가 선거에서는 이기게 된다.

【원문 : 18대선 전국상대득표율】



3.3.2 Impact of the national model parameter (K) on winning an election

As model (1) assigns more votes to candidate 1 than candidate 2 in the unclassified proportionally to the classified regardless of who wins in the classified, we face a critical question: *what is the impact of the model parameter on the election outcome?*

To answer the question, we let L=P1/M1 and P2/M2=K*(P1/M1), where K ≥1 and L >0. Then P2/M2=K*L, and thus P1=L*M1 and P2=K*L*M2. For simplification without loss of generality, we take candidate 1's perspective of winning the election. If candidate 1 gets more votes than candidate 2, it clearly means that (P1+P2) > (M1+M2), and thus we set up candidate 1's winning condition as follows:

$$(L-1)*M1+(K*L-1)*M2>0. \quad (2)$$

Page 11 | 19

The only non-trivial case for candidate 1 winning via the unclassified ballots is as follows: If L<1 (i.e. P1<M1), then (K*L-1)*M2 > (1-L)*M1. Since M2>0, we have (K*L-1) > (1-L)* (M1/M2). To simplify this inequality, we set x=M1/M2, where x is a ratio of two votes for candidate 2 between the classified (M1) and the unclassified (M2). Then equation (2), which is the candidate 1's winning condition, becomes a function of x and K such that L > (1+x)/(K+x). This is the interesting case to be discussed further.

Candidate 1's winning depends on two variables, x and K, even if K >1. We fix K = 1.5 and examine the effect of the model parameter (K=1.5), varying the values of x. Since the votes for candidate 2 from the classified are larger than the unclassified (i.e. M1>M2 from the 18th election), the minimum value of x is 1 (x≥1).

As demonstrated in Figure 4, model (1) increases candidate 1's winning as the number of the unclassified ballots goes up. Since the unclassified are to be generated by the op-scan counters, the model parameter K value is linked to the op-scanner's operation. The impact of the model parameter (K) on candidate 1's winning would be maximized when candidate 1 lost in the classified ballots.

In the 18th election, x was observed close to 35 and L=0.99, which indicates candidate 1 could win the election with less votes from the classified (at least 99% of candidate 2) but more votes from the unclassified. This could be achieved by setting K=1.5 and thus elucidates the meaning of the model parameter 1.5 in the 18th presidential election.

[Figure 4 here]

Page 12 | 19

4. 시뮬레이션

4.1 시나리오 1 : 의도치 않은 기계적 편차

전국모형에서 추정된 K를 설명할 수 있는, 잠재적인 기계적 편차($\beta > 0$)가 존재한다면, 2.2.2절에서 논의한 개념에 따라 1번 후보자에 대해서 P2 ~ B(P, r+β)로 다시 정리할 수 있으며, 2번 후보자에 대해서는 M2 ~ B(M, r)로 정리할 수 있다. 여기에서 미분류율 r과 기계적 편차 β가 일정하다면, 전체적인 미분류율은 실제 개표결과와 같은 3.7%일 것이다. 따라서 1+β/r은 각 선거구에서 나타나는 K 값처럼 약 1.5여야 한다. 이 연구에서는 r=0.03, β=0.0145라 두고 시뮬레이션 하였다.

4.2 시나리오 2 : 고의적인 조작

다음과 같이 가정한다. (1) 1번 및 2번 후보자를 제외한 나머지 후보자에 대한 투표는 무시할 만큼 작다(전체 투표수의 0.37%). (2) 미분류표의 비율(R)은 선거구에 따라 다르지만 작은 수준이다(전국적으로 분류표:미분류표=96:4). (3) 미분류표의 10%가 무효표(R2)이다(전국적으로 모든 선거구에서 동일). (4) 미분류표에서 기타 후보들의 득표 또한 무시할 수 있을 만큼 작다(미분류표의 1.3%). (5) 광학분류기에 의해서 미분류표로 잘못 분류된 정상표는 선관위의 규정에 따라 개표위원에 의해 공정하게 재분류된다. (6) 분류표로 잘못 분류된 무효표 등은 개표위원에 의해 확인되지 않는다.{역주8}

【역주8】

분류표에 무효표나 다른 후보자의 표가 섞이는 경우, 개표위원에 의해 발각되지 않을 수 있는지에 대해서는 논란의 여지가 있을 수 있다. 당연히 선관위에서는 가능성을 부정하고 있다.

※ 미분류로 처리된 투표지는 모두 수작업으로 다시 분류하고, 분류된 투표지도 사람이 육안으로 모두 재확인함.

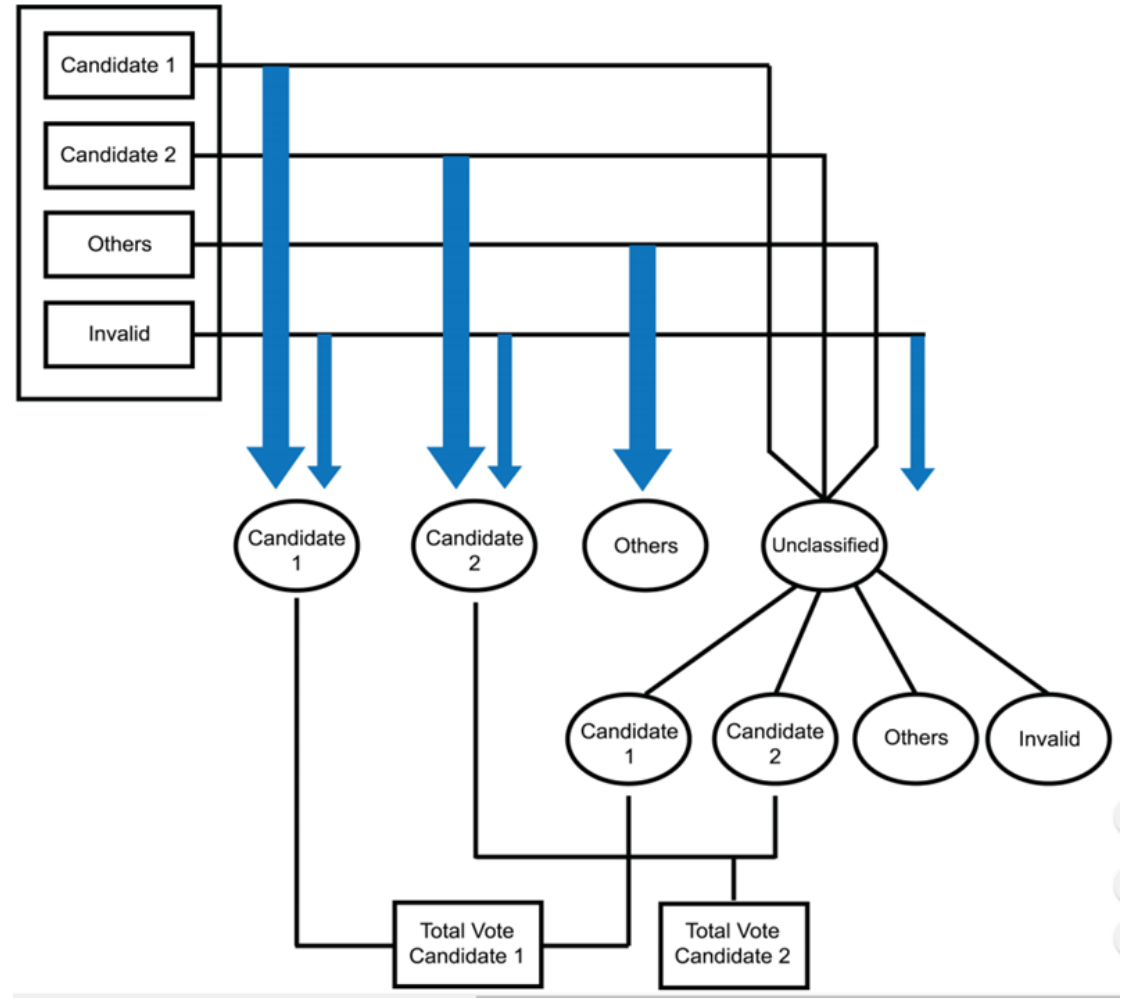
하지만, 영화 '더 플랜'을 보면 분류표를 확인하는 과정이 훑어보는 수준이었으며, 따라서 수백 장의 투표지 중에서 3% 가량 섞여있는 무효표 등을 제대로 확인하지 못할 가능성을 검증해야 한다.

시나리오 2는 개표절차에 맞추어 두 단계로 시뮬레이션 한다. 첫 단계에서는, 가상의 투표지를 실제 개표결과와 유사하게 생성한다. 이를 위해 실제 총 투표수(분류표+미분류표+무효표), 후보자별 득표수, 전체 투표 중 미분류표의 비율을 적용한다. 가상 투표지는 이항분포에 따라 생성할 수 있다. 1000회 반복시행하여 모든 선거구의 실제 개표결과와 유사한 최적의 자료를 도출하였다. 둘째 단계에서는 개표기가 가상의 투표지를 분류하는 과정을 주어진 조건에 맞추어 시뮬레이션 한다. 조건부 베르누이 시행 (a)과 다항분포 (b)를 미리 정해진 확률에 따라 적용하여 다음과 같이 분류되도록 하였다.

a) 각 후보자의 유효표는 각 후보자의 분류표와 미분류표로 분류된다.

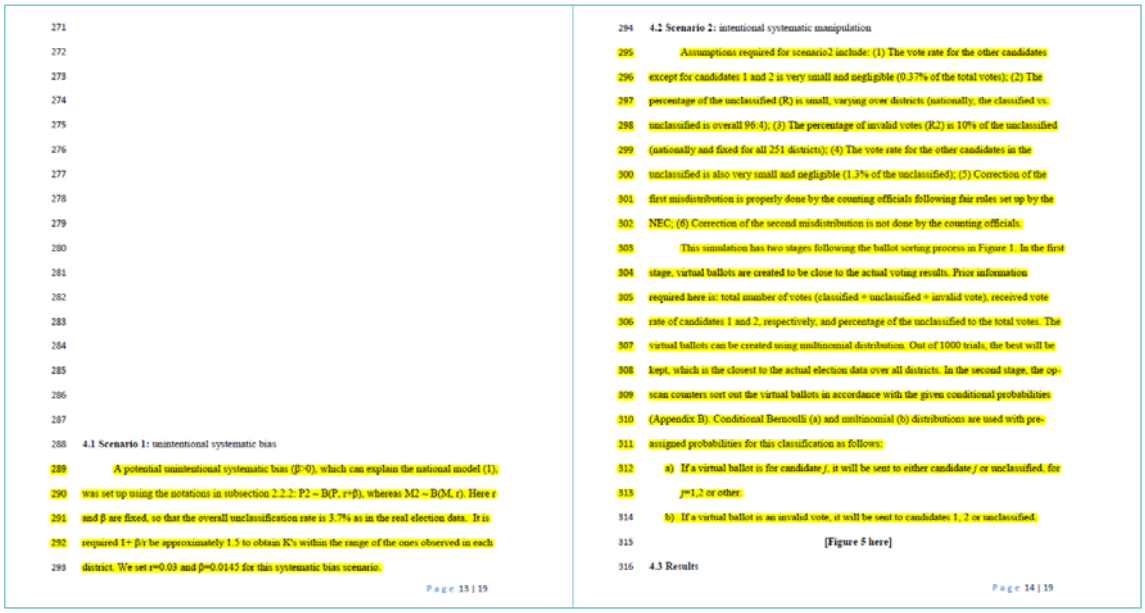
b) 가상투표지가 무효표이면 1번 및 2번 후보자의 분류표와 미분류표로 분류된다.

그림 5. 시뮬레이션 절차(시나리오 2)



자료 : "A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea" In Event: *Poster Session: Methods for Studying Comparative Politics* (http://www.mpsanet.org/)

【원문 : 시나리오 1, 2】



4.1 **Scenario 1:** unintentional systematic bias

A potential unintentional systematic bias ($\beta>0$), which can explain the national model (1), was set up using the notations in subsection 2.2.2: $P2 \sim B(P, r+\beta)$, whereas $M2 \sim B(M, r)$. Here r and β are fixed, so that the overall unclassification rate is 3.7% as in the real election data. It is required $1+\beta/r$ be approximately 1.5 to obtain K's within the range of the ones observed in each district. We set r=0.03 and β=0.0145 for this systematic bias scenario.

Page 13 | 19

4.2 **Scenario 2:** intentional systematic manipulation

Assumptions required for scenario2 include: (1) The vote rate for the other candidates except for candidates 1 and 2 is very small and negligible (0.37% of the total votes); (2) The percentage of the unclassified (R) is small, varying over districts (nationally, the classified vs. unclassified is overall 96:4); (3) The percentage of invalid votes (R2) is 10% of the unclassified (nationally and fixed for all 251 districts); (4) The vote rate for the other candidates in the unclassified is also very small and negligible (1.3% of the unclassified); (5) Correction of the first misdistribution is properly done by the counting officials following fair rules set up by the NEC; (6) Correction of the second misdistribution is not done by the counting officials.

This simulation has two stages following the ballot sorting process in Figure 1. In the first stage, virtual ballots are created to be close to the actual voting results. Prior information required here is: total number of votes (classified + unclassified + invalid vote), received vote rate of candidates 1 and 2, respectively, and percentage of the unclassified to the total votes. The virtual ballots can be created using multinomial distribution. Out of 1000 trials, the best will be kept, which is the closest to the actual election data over all districts. In the second stage, the op-scan counters sort out the virtual ballots in accordance with the given conditional probabilities (Appendix B). Conditional Bernoulli (a) and multinomial (b) distributions are used with pre-assigned probabilities for this classification as follows:

a) If a virtual ballot is for candidate j, it will be sent to either candidate j or unclassified, for j=1,2 or other.

b) If a virtual ballot is an invalid vote, it will be sent to candidates 1, 2 or unclassified.

[Figure 5 here]

4.3 Results

Page 14 | 19

4.3. 시뮬레이션 결과

기계적 편차를 가정한 시나리오 1의 경우, 유사한 K값이 나오도록 시뮬레이션 하였을 때, 미분류표의 비율이 각 선거구별 실제 결과와 일치하지 않는다. 이 시나리오를 검토하는 이유는 r+β가 허용오차 범위의 오류여서 광학분류기를 점검하는 과정에서 후보자별 편차를 선거 이전에 알 수 없었을 수도 있다는 것이다. 기계적 편차에 바탕을 둔 시뮬레이션 결과는 18대 대선의 결과를 제대로 구현하지 못하고 있는데(표 2), 미분류표의 비율이 선거구에 따라 일정치 않고 1번 후보자에 대한 편향도 균일하지 않다.(※ 실제 개표결과에서는 각 선거구별 편향이 로그정규분포를 따름)

이에 비해, 두 번째 시뮬레이션은 실제 개표결과와 비교적 유사한 결과를 보여준다(표 2). 선거구간의 인구규모의 차이가 있으므로, 시뮬레이션 결과를 검토할 때 251개 선거구 각각의 득표수보다는 득표율을 살펴보았다. 모든 선거구를 비교한 결과 시뮬레이션 결과가 실제 투표결과를 적절하게 구현하고 있음을 알 수 있으며, 따라서 전국모형과 같은 결과가 나오도록 어떻게 개표를 조작할 수 있는지 적절하게 보여주는 예시라 할 수 있다. 이 시뮬레이션 결과는 5% 유의수준에서 전체 251개 선거구 중 97% 선거구의 결과를 제대로 예측하였다. 이는 시나리오 2가 18대 대선에서 광학분류기가 어떻게 작동되도록 조작되었는지 적절하게 설명하고 있다는 의미이다.

표 2. 실제 개표결과와 시뮬레이션 결과

Results: Actual vs. Simulated Votes

251 districts combined	Total number of votes	Total unclassified votes	Votes from the classified		Votes from the unclassified		Invalid votes
			candidate 1	candidate 2	candidate 1	candidate 2	
Actual	**29,827,252**	**1,111,165**	**14,782,150**	**13,828,239**	**586,632**	**397,505**	**112,360**
Simulated 1	**29,827,252**	**1,229,495**	**14,683,046**	**13,797,770**	**685,822**	**427,442**	**112,570**
Simulated 2	**29,827,252**	**1,080,700**	**14,787,440**	**13,857,352**	**594,739**	**370,907**	**111,117**

자료 : "A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea" In Event: *Poster Session: Methods for Studying Comparative Politics* (http://www.mpsanet.org/)

【원문 : 시뮬레이션 결과】

The first simulation scenario involving a systematic bias, while providing similar K values, does not match the district level variability in the unclassification rates. The motivation is that $\varepsilon+\beta$ is within some tolerance limit so that the candidate level examination would be omitted in the calibration of the op-scan and the bias would go unnoticed before the election. The simulation results based on the systematic machine bias model do little support plausibility in the 18th election (Table 2), mainly because unclassified rates vary among districts and exhibit a non-uniform but unidirectional bias in favor of candidate 1.

In comparison, the second simulation shows the simulated votes are fairly close to the actual votes nationwide (Table 2). Considering different population size between large and small electoral districts, we evaluate simulation results for each of 251 electoral districts in received vote rates (%) rather than received vote counts (see Appendix Table C1 for all 251 districts). The comparable results over all districts imply that the simulation reasonably reflects the actual votes and thus can explain a way of model implementation. The simulation results turned out that 97% of 231 electoral districts showed very good predicted vote rates within ±3% of margin of acceptance (Appendix Table C2). This indicates the proposed scenario 2 can describe a plausible way by which the op-scan counters may have been operated in the 18th presidential election.

[Table 2 here]

Page 15 | 19

5. 분석결과의 논의

5.1 시뮬레이션 결과의 해석

수식 (1)의 전국모형과 같은 결과가 나오도록 하는 방법에는 여러 가지가 있을 수 있다. 이 논문에서 예시로 든 무효표를 유효표에 섞는 방법은, 투표결과를 바꿀 정도는 아닐 수 있다. 만약 2번 후보자의 표를 1번 후보자에게 섞는 등의 방법이 쓰인다면, 실제로 투표결과가 바뀔 수도 있지만, 발각되기도 쉬울 것이다. 무효표를 유효표로 섞는 방법이, 그 수는 적을지라도, 박빙의 승부에서 어떤 후보자의 패배 위험을 완화시키는 역할을 했을 수 있다.

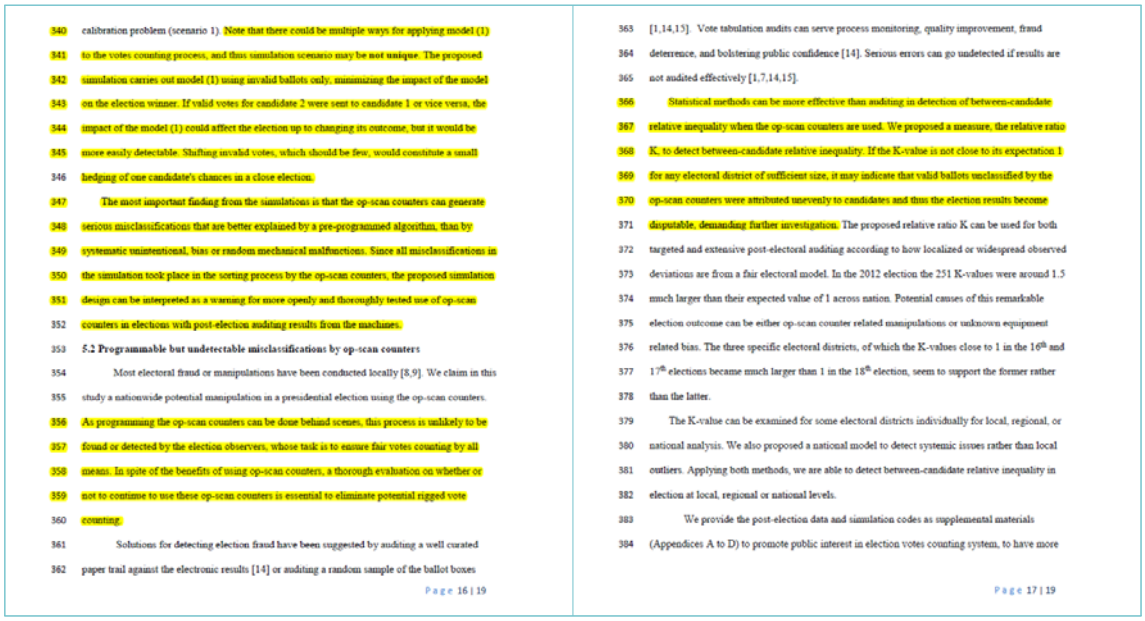
이 논문에서 시뮬레이션을 통해 밝힌 좀 더 중요한 문제는 광학분류기가 만들어 낸 잘못된 개표결과가, 고의적이지 않은 기계적 편차나 무작위적 오작동에 의한 것이 아니라, 사전에 조작된 프로그램에 의해 발생하였을 개연성이 훨씬 높다는 것이다. 시뮬레이션을 통해 이러한 잘못이 광학분류기에서 발생할 가능성이 높다는 결과가 나왔으므로, 광학분류기로 집계한 이전 선거결과를 재조사하는 등 좀 더 공개적이고 전반적인 검증을 거쳐 개표에 광학분류기를 사용할 것인지 여부를 정해야 할 것이다.

5.2 사전에 프로그램 되어도 알아채기 어려운 광학분류기의 잘못된 개표분류

분류기에 잘못된 프로그램을 심는 것은 눈앞에서 이뤄지는 일이 아니므로, 참관인 등에게 발견되거나 발각되지 않을 가능성이 높다. 광학분류기를 사용함에 따른 편익에도 불구하고, 잠재적인 개표조작 가능성을 제거하기 위해 광학분류기를 계속 사용할 것인지 여부를 철저하게 평가해야 한다.

분류기를 사용한 이후에는 통계적 검정 방법이 좀 더 효율적일 수 있다. 이 논문에서 제안한 상대득표율 K로 후보자간 상대적 불균등을 찾아볼 수 있다. 만약 K값이 1에 가깝지 않다면, 유효표가 후보자별로 균등하지 않게 미분류표로 흘러들어간다는 것을 의미하며, 따라서 선거결과를 좀 더 자세히 검증해야 한다.

【원문 : 분석결과의 논의】



calibration problem (scenario 1). Note that there could be multiple ways for applying model (1) to the votes counting process, and thus simulation scenario may be not unique. The proposed simulation carries out model (1) using invalid ballots only, minimizing the impact of the model on the election winner. If valid votes for candidate 2 were sent to candidate 1 or vice versa, the impact of the model (1) could affect the election up to changing its outcome, but it would be more easily detectable. Shifting invalid votes, which should be few, would constitute a small hedging of one candidate's chances in a close election.

The most important finding from the simulations is that the op-scan counters can generate serious misclassifications that are better explained by a pre-programmed algorithm, than by systematic unintentional, bias or random mechanical malfunctions. Since all misclassifications in the simulation took place in the sorting process by the op-scan counters, the proposed simulation design can be interpreted as a warning for more openly and thoroughly tested use of op-scan counters in elections with post-election auditing results from the machines.

5.2 Programmable but undetectable misclassifications by op-scan counters

Most electoral fraud or manipulations have been conducted locally [8,9]. We claim in this study a nationwide potential manipulation in a presidential election using the op-scan counters. As programming the op-scan counters can be done behind scenes, this process is unlikely to be found or detected by the election observers, whose task is to ensure fair votes counting by all means. In spite of the benefits of using op-scan counters, a thorough evaluation on whether or not to continue to use these op-scan counters is essential to eliminate potential rigged vote counting.

Solutions for detecting election fraud have been suggested by auditing a well curated paper trail against the electronic results [14] or auditing a random sample of the ballot boxes

Page 16 | 19

[1,14,15]. Vote tabulation audits can serve process monitoring, quality improvement, fraud deterrence, and bolstering public confidence [14]. Serious errors can go undetected if results are not audited effectively [1,7,14,15].

Statistical methods can be more effective than auditing in detection of between-candidate relative inequality when the op-scan counters are used. We proposed a measure, the relative ratio K, to detect between-candidate relative inequality. If the K-value is not close to its expectation 1 for any electoral district of sufficient size, it may indicate that valid ballots unclassified by the op-scan counters were attributed unevenly to candidates and thus the election results become disputable, demanding further investigation. The proposed relative ratio K can be used for both targeted and extensive post-electoral auditing according to how localized or widespread observed deviations are from a fair electoral model. In the 2012 election the 251 K-values were around 1.5 much larger than their expected value of 1 across nation. Potential causes of this remarkable election outcome can be either op-scan counter related manipulations or unknown equipment related bias. The three specific electoral districts, of which the K-values close to 1 in the 16th and 17th elections became much larger than 1 in the 18th election, seem to support the former rather than the latter.

The K-value can be examined for some electoral districts individually for local, regional, or national analysis. We also proposed a national model to detect systemic issues rather than local outliers. Applying both methods, we are able to detect between-candidate relative inequality in election at local, regional or national levels.

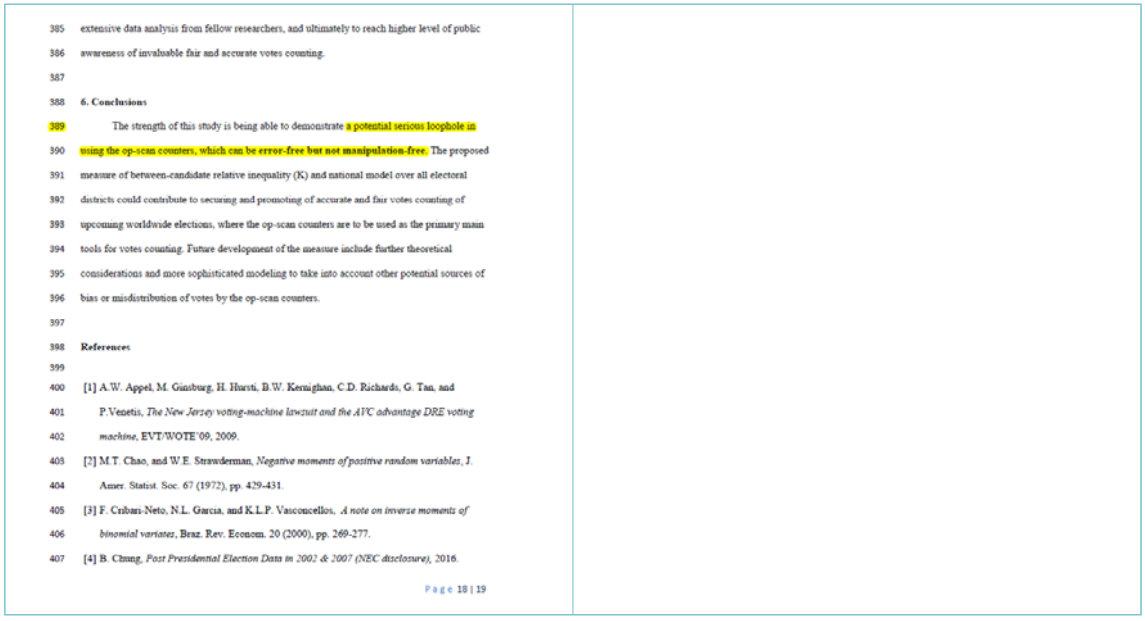
We provide the post-election data and simulation codes as supplemental materials (Appendices A to D) to promote public interest in election votes counting system, to have more

Page 17 | 19

6. 결론

분류기를 사용함에 있어서 심각한 문제는, 오차 없는 정확성을 기대할 수 있을지 모르지만, 의도적인 조작으로부터 자유로울 수 없다는 점이다. (※ 즉, 광학분류기는 '조작하라'는 지시까지 '오차 없이' 수행할 것이다.)

【원문 : 결론】



extensive data analysis from fellow researchers, and ultimately to reach higher level of public awareness of invaluable fair and accurate votes counting.

6. Conclusions

The strength of this study is being able to demonstrate a potential serious loophole in using the op-scan counters, which can be error-free but not manipulation-free. The proposed measure of between-candidate relative inequality (K) and national model over all electoral districts could contribute to securing and promoting of accurate and fair votes counting of upcoming worldwide elections, where the op-scan counters are to be used as the primary main tools for votes counting. Future development of the measure include further theoretical considerations and more sophisticated modeling to take into account other potential sources of bias or misdistribution of votes by the op-scan counters.

References

[1] A.W. Appel, M. Ginsburg, H. Hursti, B.W. Kernighan, C.D. Richards, G. Tan, and P.Venetis, *The New Jersey voting-machine lawsuit and the AVC advantage DRE voting machine*, EVT/WOTE'09, 2009.

[2] M.T. Chao, and W.E. Strawderman, *Negative moments of positive random variables*, J. Amer. Statist. Soc. 67 (1972), pp. 429-431.

[3] F. Cribari-Neto, N.L. Garcia, and K.L.P. Vasconcellos, *A note on inverse moments of binomial variates*, Braz. Rev. Econom. 20 (2000), pp. 269-277.

[4] B. Chung, *Post Presidential Election Data in 2002 & 2007 (NEC disclosure)*, 2016.

Page 18 | 19

※ 이 보고서는 2017년 4월에 진행된 MPSA(midwest political science association) Annual Conference 2017에 발표된 논문 “A Measure to Detect Between-Candidate Relative Inequality Generated by Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea”의 일부를 발췌하여 번역하고 주석을 단 것입니다. 원문에 대한 자세한 정보는 http://www.mpsanet.org/에서 얻을 수 있으며, 원문 내용에 대한 문의는 원저자에게 해주시기 바랍니다.

【원저자】

- HeeKyoung Chun, Department of Epidemiology, Georgia Southern University, PO Box 8015, Statesboro, USA. hchun@georgiasouthern.edu
- Pierre-Jerome Bergeron, Department of Mathematics and Statistics, University of Ottawa, Canada. pierrejerome@gmail.com
- HyunSeung Kim, Project BOO Inc. 268 Chungjeongro 3rd 11 St., Seodaemungu, Seoul, South Korea. n2mart@gmail.com
- OuJoon Kim, Project BOO Inc. 268 Chungjeongro 3rd 14 St., Seodaemungu, Seoul, South Korea. oujoon.k@gmail.com
- Hwashin Hyun Shin* 17 , Department of Mathematics and Statistics, Queen’s University, 48 University Ave. Kingston, ON. Canada, K7L 3N6. hhshin@mast.queensu.ca
- *교신저자(Corresponding author)

참고문헌 References

[1] A.W. Appel, M. Ginsburg, H. Hursti, B.W. Kernighan, C.D. Richards, G. Tan, and P.Venetis, *The New Jersey voting-machine lawsuit and the AVC advantage DRE voting machine*, EVT/WOTE'09, 2009.

[2] M.T. Chao, and W.E. Strawderman, *Negative moments of positive random variables*, J. Amer. Statist. Soc. 67 (1972), pp. 429-431.

[3] F. Cribari-Neto, N.L. Garcia, and K.L.P. Vasconcellos, *A note on inverse moments of binomial variates*, Braz. Rev. Econom. 20 (2000), pp. 269-277.

[4] B. Chung, *Post Presidential Election Data in 2002 & 2007* (NEC disclosure), 2016.

[5] B. Chung, *The Accusation Against 18th 408 Presidential Election Fraud*, Baweesol, 2013.

[6] J. Elklit and A. Reynolds, *A framework for the systematic study of election quality*, Democratization 12 (2005), pp.147-162.

[7] A.J. Feldman, A. Halderman, and E.W. Felten, *Security Analysis of the Diebold AccuVote-TS Voting Machine*, USENIX/ACCURATE EVT'07, 2007.

[8] P. Klimek, Y. Yegorov, R.Hanel, and S. Thurner, *Statistical detection of systematic election irregularities*, Proc. Natl. Acad. Sci. USA. 109 (2012), pp. 16469-16473.

[9] D. Kobak, S. Shpilkin, and M.S. Pshenichnikov, *Integer percentages as electoral falsification fingerprints*, Ann. Appl. Stat. 10 (2016), pp. 54-73.

[10] National Election Commission (NEC) of South Korea, *Op-scan counter is accurate and fast*, (2014), Available at http://blog.nec.go.kr/220008433572 .

[11] National Election Commission (NEC) of South Korea, *Handbook of Public Election Process*, 34-9760878-100012-14, 2014.

[12] Project Boo, *Post Presidential Election Data in 2012 (NEC disclosure)*, Feb.28 2017; data available at https://drive.google.com/open?id=0Bx0QzBOFfX95Mkl3MmcwUHhCMVk

[13] S.S. Shapiro and M.B.Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika 52 (1965), pp. 591-611.

[14] P.B. Stark, *Risk-limiting vote-tabulation audits: The importance of cluster size*, Chance 23 (2010), pp. 9-12.

[15] B. Wofford, *How to hack an election in 7 minutes*, Politico Magazine, 2016.

2017년 새사연 발간 보고서

2017년 4월 29일 현재

아젠다	발간일	제목	작성자
경제	01/03	진짜' 경제민주화로 ⑤ 하청 중소기업, 글로벌 중견 대기업 될 수 없나?	정승일
노동	01/09	2017 전망보고서 (1) : 노동시장 불안정성의 심화	송민정
경제	01/12	일본은행이 선택한 화폐적 해법, 2017년을 희망의 해로 만들 수 있을까?	송종운
복지	01/16	2017 전망보고서 (2) : '불통'에 멈춰버린 사회, 안전망을 세워야 한다	최정은
국내외 정세	01/23	2017 전망보고서 (3) : 국내외 정세, 대전환을 탐색하는 2017	박세길
세계경제	02/03	2017 전망보고서 (4) : 2017년 세계경제, "공포의 해"가 될 것인가?	송종운
마을	02/06	2017 전망보고서 (5) : 다가오는 건거의 계절, 마을살이의 운명은?	강세진
부동산	02/10	2017 전망보고서 (6) : 장기불황 초입에 들어서는 주택시장	권순형
보건의료	02/13	2017 전망보고서 (7) : 한국 보건의료 체계의 개혁, 더 이상 미룰 수 없다	고병수
종합	02/22	2017 전망보고서 (8) : 2017년 7대 분야를 전망하다	새사연
부동산	03/08	기업형 임대주택 정책 무엇이 문제인가? ①	권순형
부동산	03/10	기업형 임대주택 정책 무엇이 문제인가? ②	권순형
부동산	03/16	기업형 임대주택 정책 무엇이 문제인가? ③	권순형
부동산	03/23	기업형 임대주택 정책 무엇이 문제인가? ④	권순형
부동산	04/12	민달팽이주택협동조합이 3년을 버티며 남긴 고민들	황서연
정치	04/18	The Plan : 민주주의 깨트리기	강세진
사회정책	04/27	성장과 복지를 위한 사회정책, 한국사회 미래비전이 되어야 한다	이은경
정치	04/29	2012년 대통령선거에서 광학식투표분류에 따른 후보자간 상대적 불균등성 규명	강세진