응용통계연구 제 16권 1호, 2003년, pp. 169-179

동일조사에서 비율비교와 관련된 두 가지 흔한 오류

김현철 [1)]

요 약

동일조사에서 통계분석자가 흔히 범하는 오류 두 가지를 제시하였다. 하나는 일반적인 조사에서 발표된 조사결과로부터 두 비율의 비교를 시도할 때 범하기 쉬운 오류이고, 다른 하나는 중립적 응답 항목이 있을 때 통계전문가가 아닌 사람들 가운데에서 범하기 쉬운 잘못이다. 이런 오류들을 제시하고, 통계적으로 정확한 방법과 비교하여 잘못 사용하는 방법들이 갖는 문제들을 보여 줌으로써 교육 자료로 활용할 수 있도록 했다.

주요용어: 동일조사, 비율비교, 중립적 응답, 오차한계.

1. 문제 제기

동일조사에서 두 비율을 비교하는 문제는 모든 설문조사에서 거의 항상 다루어진다. 많은 경우에 명시적으로 비율비교를 실시하지 않는다 해도 사실상 암묵적으로 두 비율의 비교가 주요 관심사인 경우가 많다.

(예1) 선거와 관련된 여론조사를 했을 경우에 '표본크기가 1,000이고 따라서 신뢰도 95%에서 오차한계는 ±3.1%이다'라고 발표할 때는 그 이면에 어느 두 후보간의 지지율 비교를 전제로 하고 있는 것이다. 만약 A 후보의 지지율이 31%, B 후보의 지지율이 25.5%, C 후보의 지지율은 12%, 그리고 아직 결정하지 않았다는 응답이 31.5%였다고 하자. 이런 조사결과를 접할 때 많은 사람들은 과연 A 후보와 B 후보 간에 지지율의 차이가 존재하는가? 하는 점에 관심을 갖는다.

일단 이 조사의 오차 한계는 ±3.1%라고 발표되었으므로, 95% 신뢰수준에서, A 후보의 지지율은 (27.9 ~ 33.1)%이고, B 후보의 지지율은 (22.4 ~ 28.6)%이다. 따라서 두 후보의 지지율에는 통계적으로 차이가 있다고 말할 수 없게 된다. 즉 두 후보의 조사된 지지율의 차이가 오차한계의 길이인 6.2%를 넘지 못하면 신뢰도 95%에서 통계적으로 유의한 차이가 있다고 말할 수 없게 된다.

그러나 실제로 통계분석을 실시하는 과정을 살펴보면 여기에 두 가지 문제가 자리잡고 있다. 하나는 오차의 한계와 관련된 문제이고, 다른 하나는 아직 결정하지 않았다(실제로는 아직 결정하지 않았다는 중립적 응답과 무응답이 혼재된 경우가 많으며, 이것은 이 연구의 주제와는 별도의 문제이다)는 응답의 처리 방법과 관련이 있다.

1) (573-701) 전북 군산시 미룡동 산 68번지, 군산대학교 수리정보통계학부, 부교수
E-mail: kimhc@kunsan.ac.kr

조신섭(1998)은 8가지 기준에 따라 우리나라의 언론이 통계발표를 하는데 있어서의 문제점을 잘 정리하여 밝힌 바 있다. 그 연구에서 저자는 오차의 한계에 대한 발표는 비교적 잘 이루어지고 있다고 지적했지만, 아직도 통계조사의 결과를 발표하면서 이렇게 오차의 한계를 밝히는 수준도 완벽하게 이루어지지는 않는 것 같다. 한국갤럽(2002)이 2002년에 실시한 각종 여론조사결과를 언론사가 보도한 내용을 자사 홈페이지에 공개한 것을 살펴보면, 아직도 보도과정에서 오차의 한계를 누락시킨 것들을 발견할 수 있다. 심지어 강남준(2001)은 표본오차를 허구라고 주장하면서 언론보도에서 표본오차의 발표에 대해 매우 부정적인 주장을 하고 있기도 하다. 강남준의 주장은 표본오차와 비표본오차를 혼동하고 있는 주장으로 보인다. 물론 비표본오차가 발생할 수 있고 그 크기는 알 수가 없다는 점이 명시되어야겠지만, 그렇다고 표본오차를 밝히지 않아도 된다는 주장은 논리적이지 않다.

이렇듯 표본오차의 크기를 명확히 밝히는 것조차 완전하게 이루어지지 않고 있는 현실이다. 그러나 표본오차를 밝힌다 해도 문제는 남아있다. 앞의 (예1)에서 제시한 오차한계는 독립적인 하나의 모집단에서 얻은 표본비율로 모비율을 추정할 때 발생하는 오차한계의 최대값일 뿐 우리가 관심을 갖는 동일조사에서의 두 비율비교시 발생할 수 있는 오차한계가 아니기 때문이다.

한편, 설문조사를 실시하다 보면 중립적 의견을 갖는 응답자가 종종 발견된다.

(예2) 어떤 후보에 대한 선호도를 측정해 보기 위해 다음과 같이 단순한 질문을 만들었다고 하자.

질문) 귀하께서는 A후보가 마음에 드십니까?

1) 예 2) 아니오 3) 잘 모르겠다

이 예에서처럼 두 가지 이상의 선택 항목 중 어느 하나를 선택하게 하되 응답자의 의견이 중립적이거나 혹은 선호도가 비슷해서 선뜻 결정하기 어려운 응답자가 대답할 수 있는 항목을 추가할 필요가 있다. 이런 항목을 두지 않을 경우 중립적 응답자에게 필요 이상의 정신적 압박을 줄 수 있으며, 무성의한 응답, 혹은 무응답을 초래할 수도 있기 때문이다. 중립적 응답이 있을 때도 결국 응답비율을 비교하는 것이 일반적인 관심사이다. 그러나 통계전문가가 아닌 연구자들은 이런 중립적 응답이 포함될 경우 종종 조사 결과의 처리에 많은 부담을 느끼기도 하고 심지어는 임의로 처리함으로써 오류를 범하는 경우도 보게 된다.

우리는 두 비율을 비교하는 통계분석에서 오차한계를 정확하게 사용하지 않음으로써 발생할 수 있는 문제점을 2장에서 보여주려고 한다. 아울러 비전문가들이 중립적 응답 항목이 있는 질문의 통계 처리에서 잘 못 사용하기 쉬운 방법 3 가지를 제시하고, 비율비교에서의 오차한계를 비교함으로써 비전문가들이 유혹을 받는 방법들이 얼마나 적절하지 않은 방법인가를 3장에서 보여주고자 한다. 이런 결과는 통계 교육의 목적으로 사용될 수 있으며, 비전문가들에게 정확한 방법을 사용하도록 권장하는 목적으로도 사용될 수 있을 것이다.

2. 두 비율의 비교에서 오차의 한계

모집단의 크기를 N, 모집단 전체에서 응답항목 $i(i=1,\cdots,I)$를 선호하는 사람들의 수를 N_i, 그리고 표본의 크기를 n이라 하자. 또 표본추출 결과, 응답항목 i를 선호하는 사람들의 수를 X_i라고 하자. 만약 표본 추출률이 충분히 작다면 비복원추출은 복원추출을 대신할 수 있으므로 X_i들의 결합분포는 근사적으로 다항분포가 된다. 즉 특정한 응답항목의 응답비율이 증가하면 다른 응답항목의 응답비율은 감소하게 되는 음의 상관관계를 갖는다는 말이다.

이 때 응답항목 i를 선호하는 사람들의 표본비율을 $\hat{p_i}=\frac{X_i}{n},(i=1,\cdots,I)$ 이라 하면, Scott과 Seber(1983)는 신뢰도 95%에서 임의의 두 비율 p_i와 p_j사이의 차이에 대한 구간추정은 다음과 같음을 지적하였다.

$$p_i-p_j=\hat{p_i}-\hat{p_j}\pm 1.96\sqrt{\frac{\hat{p_i}+\hat{p_j}-(\hat{p_i}-\hat{p_j})^2}{n}} \tag{2.1}$$

따라서 두 표본비율간에 $1.96\sqrt{\frac{\hat{p_i}+\hat{p_j}-(\hat{p_i}-\hat{p_j})^2}{n}}$ 이상의 차이가 나야만 신뢰도 95%에서 통계적으로 두 비율간에 차이가 있다고 말할 수 있다.

한편 대부분의 조사에서 발표되는 최대 허용오차는 식 (2.2)처럼 계산하여 사용하고 있다. 따라서 조사 결과에 상관없이 표본크기가 주어지면 항상 오차의 한계가 같은 것처럼 발표된다. 그래서 (예1)에서 표본크기가 1,000 이면 신뢰도 95%에서 오차의 한계는 ±3.1% (혹은 근사적으로 ±3.0%)라고 발표되고, 만약 표본크기가 500 이면 같은 기준에서 오차의 한계는 ±4.4% 라고 발표되는 것이다.

$$\pm 1.96\sqrt{\frac{0.5^2}{n}} \tag{2.2}$$

따라서 두 표본비율간에 $1.96\sqrt{\frac{0.5^2}{n}}\times 2$의 차이가 나야만 신뢰도 95%에서 통계적으로 두 비율간에 차이가 있다고 말할 수 있다. 이 오차한계는 각 모비율을 서로 다른 독립적인 표본에서 얻은 표본비율을 이용하여 추정할 때 발생하는 오차의 한계로 두 비율의 차이를 비교할 때는 적합하지 않다.

다음 그림(2.1)을 보자. 이 그림은 표본크기가 1,000일 때 두 응답항목, i, j 사이의 비율차이에 대한 오차 한계를 곡면으로 나타내고, 일반적인 표본조사에서 발표하는 표본크기만으로 계산한 오차의 한계를 평면으로 그린 것이다. 이 그림을 살펴보면, 식 (2.2)의 오차한계가 보수적인 의사결정이라는 측면에서 어느 정도 수용될 수 있을지 모르지만, 두 비율 중에서 어느 하나라도 0.5에서 멀어지면 식 (2.2)에 비해 식 (2.1)의 오차한계는 급격히 작아진다. 따라서 두 비율의 차이를 비교하고자 하는 사람들이 여론조사에서 발표되는 오차의 한계를 기준으로 의사결정한다면 오류를 범할 수 있게 된다.

(예3) 위의 (예1)의 결과를 식 (2.1)에 따라 다시 정확한 오차한계를 구해보자. 이 값은 약 4.6%가 된다. 그리고 A, B후보의 지지율 차이는 5.5%로 추정되었으므로 A, B 후보의 지지율 사이에는 신뢰도 95%에서 통계적으로 유의한 차이가 있다고 말할 수 있다. 이런 결과는

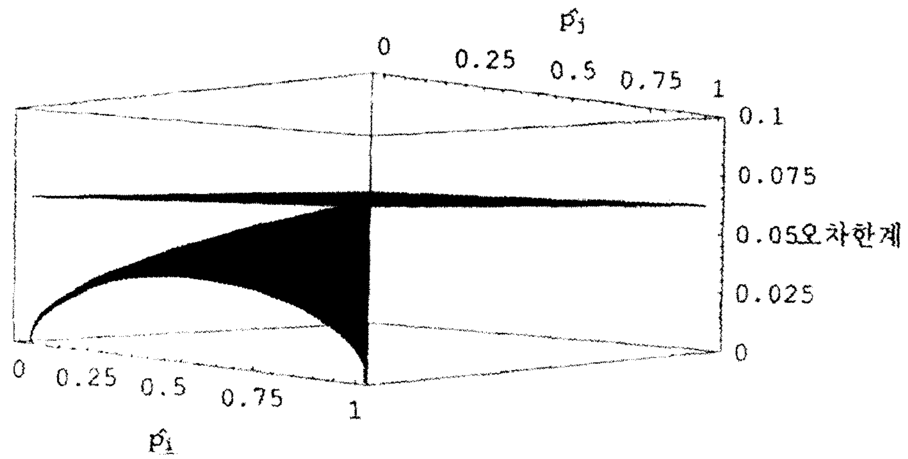

그림 2.1: 두 비율의 차이에 대한 오차한계(n=1,000)

앞서 제시한 결론과 서로 반대되는 것이다.

(예4) 울산 MBC가 지난 4월 20~21 양일간 울산 시민 2,500명을 대상으로 실시한 대선 여론 조사 결과(강태봉 2002)는 노무현 후보와 이회창 후보의 지지율은 각각 42.5%와 39%였다고 발표했다. 이 조사에서 발표된 오차한계는 4%(±2%)였고, 따라서 두 후보 사이의 지지율에는 신뢰수준 95%에서 통계적으로 서로 차이가 있다고 말할 수 없다. 그러나 식 (2.1)에 의하면 두 후보간의 차이에 대한 정확한 오차한계는 약 3.5%이다. 이런 결과는 결국 두 사람의 지지율의 차이가 통계적으로 유의한가에 대해서 서로 상반된 결론에 도달하게 한다.

이상에서 우리는 하나의 질문에 대해 서로 다른 두 응답항목 간의 모비율 차이를 비교하는 문제를 다루었다. 일반적인 조사에서 흔히 발표하는 오차한계를 사용해서 두 비율의 차이를 비교한다면 두 표본비율 중 어느 하나라도 0.5에서 크게 벗어난 값을 가질 때는 두 비율의 차이에 대한 오차한계를 지나치게 과대 추정을 하기 때문에 정확한 방법으로 계산했을 때와 서로 다른 결론에 도달할 수 있음을 확인하였다.

3. 중립적 응답이 있을 때 두 비율의 차이

3.1. 잘못된 방법 세 가지

여기에서는 중립적 응답이 있을 때 두 비율의 비교 문제를 생각해 보자. 중립적 응답항목을 선택한 응답자가 있을 때도 여전히 주된 관심사는 두 비율 p_i와 p_j사이의 차이이다. 이 경우에도 역시 앞의 (2.1)에 따라 오차한계를 구해야 한다. 그러나 필자가 통계상담 과정에서 발견한 바에 따르면, 통계학 전공자가 아닌 사람들 가운데에서 가끔 시도하려는 유

혹을 받는 방법으로 관념적인 단순화에 기초한 세 가지 방법이 있다. 이들은 기본적으로 두 개의 독립표본에서 얻은 비율의 비교에 기초하고 있다. 여기서는 2장에서 정의한 표기법에 중립응답을 0번 응답항목로 추가하여 확장한 것을 사용하기로 하자. 즉 응답항목 i $(i = 0, \cdots, I)$에 대한 응답자 수를 X_i라고 했을 때, 중립응답자 수는 X_0이 되고, 모비율도 역시 p_0으로 나타내기로 하자.

첫 번째 방법은 중립적인 응답 항목에 응답한 사람들을 무응답으로 간주하여 표본에서 제외시키고 나머지 표본만을 사용하여 표본비율을 얻어 두 개의 독립표본에서 얻은 비율의 차이비교를 실시하는 방법이다. 이 경우의 오차한계는 다음 식 (3.1)과 같다. 이 때 두 비율의 차이에 대한 점추정량을 $\hat{p_i}^a - \hat{p_j}^a$이라 하면, 이 추정량은 $p_i - p_j$에 대해 편향추정량이 된다.

$$\hat{p_i}^a = \frac{X_i}{n^a}$$

$$\hat{p_j}^a = \frac{X_j}{n^a}$$

$$\pm 1.96\sqrt{\frac{\hat{p_i}^a(1-\hat{p_i}^a) + \hat{p_j}^a(1-\hat{p_j}^a)}{n^a}} \tag{3.1}$$

여기서 n^a는 중립적 응답을 제외한 나머지 응답항목에 대한 응답자의 합계이다.

두번째 방법은 중립적인 응답을 한 응답자들을 다른 응답항목에 응답한 사람들에 할당해주는 방법인데, 할당 방법은 각 응답항목의 응답 비율로 나누어 주는 방법이다. 표본비율,

$$\hat{p_i}^b = \frac{X_i + X_0\frac{X_i}{n}}{n} = (1 + \frac{X_0}{n})\frac{X_i}{n}$$

$$\hat{p_j}^b = \frac{X_j + X_0\frac{X_j}{n}}{n} = (1 + \frac{X_0}{n})\frac{X_j}{n}$$

를 기초로 하여 $\hat{p_i}^b - \hat{p_j}^b$으로 점추정하고, (3.2)와 같은 방법으로 오차한계를 구하는 방법이다.

$$\pm 1.96\sqrt{\frac{\hat{p_i}^b(1-\hat{p_i}^b) + \hat{p_j}^b(1-\hat{p_j}^b)}{n}} \tag{3.2}$$

세번째 방법은 중립적인 응답을 한 응답자들은 나머지 응답항목에 각각 동일한 비율의 응답을 한 것으로 간주하고,

$$\hat{p_i}^c = \frac{X_i + \frac{X_0}{I}}{n}$$

$$\hat{p_j}^c = \frac{X_j + \frac{X_0}{I}}{n}$$

를 기초로 $\hat{p_i}^c - \hat{p_j}^c$으로 점추정하고, 오차한계는 (3.3)과 같은 식이 된다.

$$\pm 1.96\sqrt{\frac{\hat{p_i}^c(1-\hat{p_i}^c)+\hat{p_j}^c(1-\hat{p_j}^c)}{n}} \tag{3.3}$$

3.2. 잘못된 방법의 편향

우선 적절하지 않은 세 점추정량이 $p_1 - p_2$에 대해 비편향성을 만족하는지 여부를 확인해 보자. 먼저 첫 번째 방법에 대해서는,

정리 3.1 $\hat{p_i}^a - \hat{p_j}^a$는 $p_i - p_j$를 과대추정하는 편향추정량이다.

증명:

$$Bias[\hat{p_i}^a - \hat{p_j}^a] = E[\hat{p_i}^a - \hat{p_j}^a] - (p_i - p_j) = E[\frac{n}{n^a}\frac{X_i - X_j}{n}] - (p_i - p_j)$$

위 식에서 비록 n^a는 X_i들과 서로 독립이지 않으나, 항상 $\frac{n}{n^a} \geq 1$이어서 기대값 부분은 $(p_i - p_j)$보다 크거나 같게 되므로

$$Bias[\hat{p_i}^a - \hat{p_j}^a] \geq 0$$

두번째 방법에 대해서도,

정리 3.2 $\hat{p_i}^b - \hat{p_j}^b$는 $p_i - p_j$를 과대추정하는 편향추정량이다.

증명:

$$Bias[\hat{p_i}^b - \hat{p_j}^b] = E[\hat{p_i}^b - \hat{p_j}^b] - (p_i - p_j) = E[(1 + \frac{X_0}{n})\frac{X_i - X_j}{n}] - (p_i - p_j)$$

마찬가지로 비록 X_0는 X_i들과 서로 독립이지 않으나, 항상 $1 + \frac{X_0}{n} \geq 1$이어서 기대값 부분은 $(p_i - p_j)$보다 크거나 같게 되므로

$$Bias[\hat{p_i}^b - \hat{p_j}^b] \geq 0$$

처음 두 방법의 편향을 살펴보면 모두 중립적 의견을 갖는 사람들의 수가 커질수록 편향도 역시 커진다는 것을 알 수 있다.

한편 세번째 방법은 비록 $\hat{p_i}^c$와 $\hat{p_j}^c$는 각각 p_i과 p_j에 대해서는 편향추정량이지만, $p_i - p_j$에 대해서는 비편향추정량이 된다.

정리 3.3 $\hat{p_i}^c - \hat{p_j}^c$는 $p_i - p_j$에 대해 비편향추정량이다.

증명:

$$Bias[\hat{p_i}^c - \hat{p_j}^c] = E[\hat{p_i}^c - \hat{p_j}^c] - (p_i - p_j) = E[\frac{X_i - X_j}{n} - \frac{\frac{X_0}{I} - \frac{X_0}{I}}{n}] - (p_i - p_j) = 0$$

3.3. 모의실험에 의한 신뢰도 비교

잘못된 세 가지 방법에 대하여 편향정도, 오차한계의 타당성 등을 비교해 보기 위해 모의실험을 수행하였다. 모집단의 크기는 100,000으로 했고, 이 모집단을 다시 1, 2번 응답 항목을 선호하는 사람들의 집단 1, 2와 중립적 의견을 가진 집단 0의 세 부집단으로 나누었다. 집단 1의 비율(p_1)은 0.5, 0.4, 0.3, 0.2의 네 가지로 정했으며, 집단 1과 집단 2의 구성비의 차이($p_1 - p_2$)는 0.05와 0.1로 하였다. 이렇게 모집단을 구성하고, 랜덤표본을 각각 크기(n) 1000, 2000의 두 크기로 추출하여 총 12(3 x 2 x 2)가지의 경우를 실험했다.

표 3.1: 각 방법의 오차한계 비교

p_1	p_2	n	점추정값 평균				평균 오차한계				
방법			0	1	2	3	0	1	2	3	이론
0.50	0.45	1000	0.051	0.054	0.053	0.051	0.060	0.045	0.044	0.044	0.060
		2000	0.050	0.052	0.052	0.050	0.043	0.032	0.031	0.031	0.043
	0.40	1000	0.100	0.111	0.110	0.100	0.058	0.046	0.044	0.044	0.058
		2000	0.100	0.111	0.110	0.100	0.041	0.032	0.031	0.031	0.041
0.40	0.35	1000	0.049	0.066	0.062	0.049	0.054	0.050	0.044	0.044	0.054
		2000	0.050	0.067	0.062	0.050	0.038	0.036	0.031	0.031	0.038
	0.30	1000	0.100	0.143	0.130	0.100	0.051	0.052	0.043	0.044	0.051
		2000	0.100	0.143	0.130	0.100	0.036	0.037	0.031	0.031	0.036
0.30	0.25	1000	0.050	0.091	0.072	0.050	0.046	0.059	0.043	0.044	0.046
		2000	0.050	0.091	0.073	0.050	0.032	0.042	0.030	0.031	0.032
	0.20	1000	0.100	0.199	0.150	0.100	0.043	0.061	0.042	0.044	0.043
		2000	0.100	0.200	0.150	0.100	0.031	0.043	0.030	0.031	0.031
0.20	0.15	1000	0.050	0.142	0.082	0.050	0.037	0.073	0.040	0.044	0.037
		2000	0.051	0.144	0.083	0.051	0.026	0.052	0.028	0.031	0.026
	0.10	1000	0.100	0.334	0.170	0.100	0.033	0.075	0.037	0.044	0.033
		2000	0.100	0.333	0.170	0.100	0.024	0.053	0.026	0.031	0.024

이렇게 얻은 표본에서 각 방법으로 $p_1 - p_2$에 대한 점추정값과 95%신뢰도에서의 오차한계를 계산했다. 이렇게 계산된 각 오차한계가 $p_1 - p_2$보다 크면 실제로는 $p_1 - p_2$가 0이 아닌데도 구간추정의 규칙에 따라 두 비율 사이에는 차이가 있다고 결정할 수 없으므로 잘못된 의사결정을 하게 된다. 따라서 이 오차한계가 $p_1 - p_2$보다 작고 또 그 오차한계로 구한 신뢰구간 내에 $p_1 - p_2$가 포함되면 옳바른 결론에 도달한 경우이므로, 각 조건에서 5,000번의 반복실험을 하여 앞의 조건에 해당하는 경우의 수를 반복수 5,000으로 나눈 값을 신뢰도로 하였다. 한편 점추정량과 오차한계에 대해서는 각각 평균이 얼마나 되는 지도 함께 계산했다. 이러한 모든 모의실험 과정은 GAUSS 2.0(Aptech Systems 1988)을 통해 이루어졌다.

표 3.2: 각 방법의 신뢰도 비교

p_1	p_2	n	신뢰도			
방법			0	1	2	3
0.50	0.45	1000	0.355	0.502	0.508	0.507
		2000	0.604	0.717	0.719	0.731
	0.40	1000	0.889	0.810	0.796	0.849
		2000	0.951	0.782	0.769	0.849
0.40	0.35	1000	0.414	0.498	0.533	0.537
		2000	0.713	0.654	0.684	0.787
	0.30	1000	0.945	0.592	0.636	0.905
		2000	0.954	0.400	0.498	0.908
0.30	0.25	1000	0.546	0.441	0.542	0.584
		2000	0.832	0.458	0.589	0.847
	0.20	1000	0.949	0.184	0.407	0.950
		2000	0.946	0.036	0.191	0.947
0.20	0.15	1000	0.733	0.260	0.503	0.622
		2000	0.942	0.117	0.400	0.920
	0.10	1000	0.945	0.003	0.123	0.988
		2000	0.951	0.000	0.016	0.988

이상의 조건으로 실험한 결과는 표(3.1)과 (3.2)에 있다.

우선 표 (3.1)에서 방법 0은 Scott등(1983)이 제시한 식 (2.1)에 의한 결과인데, 모의실험 결과와 이론적인 결과가 잘 일치함을 볼 수 있다. 여기서 이론적인 결과란 (3.1)에서 이론이라고 되어있는 마지막 열의 값으로 표본비율이 아닌 모비율을 사용해서 식 (2.1)의 오차한계를 계산한 것이다. 또 방법 1, 2, 3은 차례대로 3.1절에서 소개한 잘못 사용하는 방법들을 나타낸다. 3.2절에서 이미 살펴보았듯이 방법 1, 2는 $p_1 - p_2$를 과대 추정하고 있으며, 과대 추정 정도는 방법 1이 방법 2에 비해 더욱 심각한 것으로 나타났다. 또 평균 오차한계를 살펴보면 식 (2.1)의 결과는 앞에서 보았듯이 각각의 비율이 0.5에서 멀어지면 멀어질수록 오차한계의 크기가 작아지나, 방법 1은 오히려 커지는 것으로 나타났다. 방법 2의 오차한계는 작아지긴 하나 각각의 비율값의 변동에 민감하게 달라지지 않는다. 방법 3은 소숫점 아래 넷째자리에서 반올림한 결과에 전혀 변동이 없었다.

이제 표 (3.2)에서 각 방법의 신뢰도를 비교해 보자. 이 실험은 신뢰도 95%를 가정한 실험이므로 신뢰도는 0.95에 가까울수록 정확한 병법이다. 우선 Scott등(1983)의 방법은 모비율 차이가 10% 일 때는 비교적 정확한 것으로 나타났으나, 두 비율의 차이가 0.05일 때는 두 비율이 모두 0.5에서 멀어질 수록 신뢰도가 95%에 근접해 가는 것을 발견할 수 있다. 그러나 잘못 사용하는 방법 1, 2의 경우에는 정 반대의 현상을 보여주고 있다.

여기서 한가지 흥미로운 사실은 방법 3의 실험 결과이다. 방법 3의 점추정은 사실 상 비편향추정량이므로 표 (3.1)의 결과에는 특이할 만한 사실이 없다. 그러나 표 (3.2)를 보면, 두 비율이 0.5에 가깝고 두 비율의 차이가 작으면(여기서는 0.05) 통계적으로 옳은 방법보다 오히려 신뢰도가 더 높다는 점이다. 이는 방법 0보다 오차한계가 작아서 추정된 비율의 차이보다 오차한계가 작은 값을 갖게 되어 결국 모비율 사이의 차이를 식별할 수있기 때문이다. 또 두 비율이 모두 작은 값을 가질 때는 방법 3의 오차한계가 정확한 방법에 비해 커져서 신뢰도 역시 0.95를 훨씬 상회하는 것을 발견했다.

4. 사례연구

저자는 지난 6월 8~9 양일간 K시 성인 남녀 1,000명을 대상으로 6.14 지방선거의 K시 시장 선거와 관련된 선호도 조사를 실시했다. 이 조사에서 오차 한계는 식 (2.2)에 따르면 ± 3.1%였다. 이 조사 결과 선호도에서 1위를 차지한 후보는 K후보로 39.2%였고, 2위를 차지한 후보는 H후보로 17.8%였다. 3위는 J후보로 8.6%, 무응답은 약 30%였으며, 나머지가 기타 군소 후보들을 지지하는 사람들이었다.

표 4.1: 사례분석 비교

방법	비율차이에 대한 점추정	오차한계
실제 차이(선거결과)	20.3%	
정확한 방법 0	21.4%	± 2.3%
잘못된 방법 1	30.6%	± 2.5%
잘못된 방법 2	29.0%	± 2.1%
잘못된 방법 3	21.4%	± 2.0%

따라서 선두 두 후보간의 추정된 지지율 차이는 21.4%였다. 이제 이 지지율 차이를 본문에서 설명한 정확한 방법과 잘못된 방법들에 의해 분석하여 그 결과를 비교해 보자. 비교 결과는 표 (4.1)에 있다. 이 결과에서 우리는 잘못된 방법 1, 2가 비율차이를 지나치게 과대 추정하고 있음을 다시 한번 확인할 수 있다. 또 잘못된 방법 1, 2로는 오차 한계를 추정하여 적용시켰을 때 예측 결과가 실제 결과를 크게 벗어난 것을 알 수있다. 이 사례에서도 잘못된 방법 3은 오차한계도 작고 좋은 방법처럼 보이나, 실제 신뢰수준을 알 수 없다는 점에서 좋은 방법이 아니다.

5. 결론

우리는 이제까지 동일조사에서 비율비교와 관련해서 발생할 수 있는 두 가지 오류에 대해 설명했다. 첫째는 두 비율을 비교하는데 있어서 조사결과를 발표할 때 함께 발표하는 최대 허용오차를 사용하게 되면 지나치게 보수적인 결과를 얻게됨을 설명했다. 물론 비표본

오차등으로 인해 실제 표본오차보다 더 큰 오차가 발생할 수있다 해도, 그것은 별개로 다루어져야 할 문제라는 점을 기억해야 한다. 둘째는 중립적 응답이 있는 설문조사에서 그 처리 방법에 대해서 몇가지 결과를 살펴보았다. 우리는 이런 결과들에서 비전문가들이 잘못 사용하기 쉬운 세 가지 방법이 큰 위험을 안고 있음을 확인할 수 있었다. 우리는 이 결과를 바탕으로 비전문가들이 정확한 방법인 식 (2.1)을 사용하도록 권장할 수 있었다. 다만 잘못된 방법 3의 경우에는, 비편향성을 만족할 뿐 아니라 경우에 따라서는 신뢰성을 높일 수도 있다는 사실이 발견되어 좀 더 체계적인 연구가 필요할 것으로 보인다.

참고문헌

[1] 강남준 (2001). 여론조사에서 기자의 역할, <관훈저널>, 2001년 가을호, 관훈클럽, 169-180.

[2] 강태봉 (2002), 울산MBC 대선관련 여론조사, <부산일보>, 4월 23일, 부산.

[3] 조신섭 (1998). 언론보도 사례를 통해 본 통계발표상의 문제, <응용통계연구>, 제12권 2호, 한국통계학회, 633-655.

[4] 한국갤럽 (2002). News, news, *http://egallup31.gallup.co.kr/News/*

[5] Aptech Systems Inc. (1988). *GAUSS Version 2.0 Systems and Graphics Manual.*

[6] Scott, A.J. and Seber, A.F. (1983). Difference of Proportions From the Same Survey, *The American Statistician*, Vol. 37, No. 4, 319-320.

[2002년 6월 접수, 2002년 10월 채택]

Some Statistical Issues to Compare the Two Proportions in a Sample Survey

Hyun Chul Kim [1)]

ABSTRACT

We suggest two types of misuses to analyze the same survey data. One is related with the fact that people may use the wrong bounds of error when they compare two proportions. And the other is related with that some non-statisticians are apt to use wrong methods when there is a neutral answer in a question. We suggest these methods and compare them with the statistically good method. It will be a good results in educational purpose.

Keywords: Sample Survey; Difference of Two Proportions; Neutral Responses; Error of Estimate.

1) Associate Professor, Faculty of Mathematics, Informatics and Statistics, Kunsan National University.
E-mail: kimhc@kunsan.ac.kr