

온라인 조사를 통한 제 16대 대선 예측 및 분석

이 흥 철 (엔아이코리아)

김 원 용 (이화여대)



ISO 9001 국제 품질경영시스템 인증획득 기관

11

조사연구학회 춘계학술대회

1. 온라인 여론조사의 가능성



ISO 9001 국제 품질경영시스템 인증획득 기관

12

조사연구학회 춘계학술대회

1. 서 언

□ 조사에서 접근 방법의 타당도/효율성은 시대의 흐름에 따라 변화..

~1940년대 : 우편 조사

↓

1950년대~ : 개별면접조사

↓

1990년대~ : 전화조사

↓

2000년대~ : 온라인조사
핸드폰조사

▶ 접근 방법별 조사 비용 구성비 추세

(2000, ESOMAR Annual Report)

N ISO 9001 국제 품질경영시스템 인증획득 기관

| 3

조사연구학회 춘계학술대회

□ 한/미 온라인조사 시장의 성장 추세

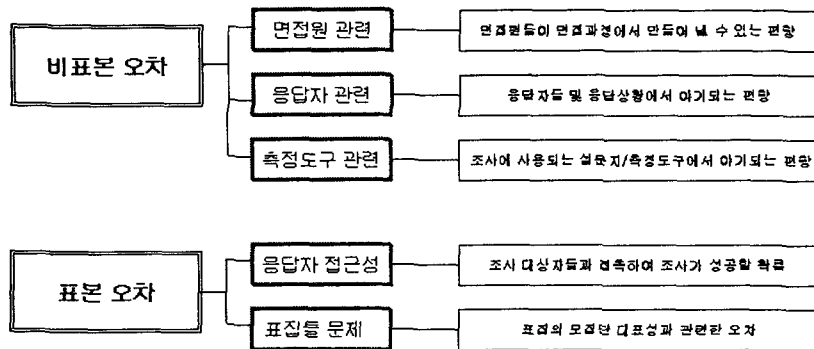
미국조사/온라인 조사시장규모 추이 →

← 국내조사/온라인 조사시장규모 추이

N ISO 9001 국제 품질경영시스템 인증획득 기관

| 4

□ 온라인 조사의 오차 요인



□ 조사 방법별 오차요인 비교

		개별면접조사	전화조사	홈조사	온라인조사
표본 오차	표본의 대표성(표집틀)	X	△	○	△
	응답자 접근성	X	△	○	△
비표본 오차	면접원 편향	X	△	△	○
	응답자 편향	X	△	○	○
	조사도구 편향	○	○	○	○
	응답상황 편향	X	X	△	○
기능성	특시성	X	△	△	○
	규모성	X	△	△	○
	경지성	X	△	X	○

● 매우우수 ○ 좋음 △ 나쁨 × 매우나쁨

□ 온라인조사의 모집단 대표성 확보방안

방법 1 : 오프라인상에서 모집단대표성을 갖는 패널 구축

→ 이들 대상 온라인 조사 (Knowledge Network 방식)

- 전체인구에서 확률적 표본추출에 의한 패널 구축
- 모든 패널들에게 MS사의 WebTV 셋톱박스 제공
- Web TV(인터넷 채널)를 통한 조사

방법 2 : 온라인상에서 패널구축후 조사

→ 분석과정에서 성향가중 모형적용 (Harris Interactive 방식)

- 패널구축은 비확률적 방식으로 구축
- 전화조사(확률적 표본추출)와 온라인조사 병행/비교
- 온라인조사결과에 성향가중치 조정 → 모수치에 맞게 조정

□ 2000년 미국 대선 예측 조사 자료 비교

	상대적인 득표 예측치*			예측오차		예측오차	
	고어	부시	거	(방법A)	순위	(방법B)	순위
선거결과	50	50	100				
ABC	48	52	100	2	12	3	12
CBS	51	49	100	1	4	2	1
CBS/NYT	47	53	100	3	15	5	15
CNN/USAToday(Gallup)	49	51	100	1	4	2	7
Fox(Opinion Dynamics)	50	50	100	0	1	0	1
Harris(InterneB)	50	50	100	0	1	0	1
Harris(phone)	50	50	100	0	1	0	1
Hotline	46	54	100	4	19	7	19
IBD/CNN/TIPP	49	51	100	1	4	2	5
ICR	49	51	100	1	4	3	7
Knowledge Networks	48	51	100	1	4	2	7
Manal College	47	53	100	3	15	6	15
NBC/WSJ/(Hart/Teeter)	48	52	100	2	12	4	12
Newweek	49	51	100	1	4	3	7
Pew	48	51	100	1	4	2	7
Rasmussen	47	53	100	3	15	6	15
Reuters/MBNBC(Zogby)	51	49	100	1	4	2	5
Voter.com(LakeGossas)	47	53	100	3	15	6	15
Washington Post	48	52	100	2	12	3	12
예측오차평균				1.6		3.5	

* 상대적인 득표율 예측치는 고어와 부시의 득표율 100%로 상계했을 때의 상대적인 득표율 비율을 나타냄.

** 방법A는 각 후보별 득표율 예측오차 평균, 방법B는 두 후보간의 차이값에 대한 예측오차

*** 자료원: Trugott (2001), Assessing Poll Performance in 2000 Campaign, *Polling Data Quarterly*, 63, 389-417.

□ 성향가중 모형

▶ Rosenbaum과 Rubin(1983)

- 비실험적 연구에서 인과적인 효과를 알아보기 위한 방법
(실험적 연구에서는 사전 처치변수에 대한 Randomization을 가정)
- 공변량변수(Covariate)에 의한 보정방식의 제약(Cochran, 1965)
- 조건확률에 기반한 가중모형
 - 결측치가 없는 완전한 자료세트에서,
 - 피험자 i ($i=1,2,3, \dots, n$)에 대한 가중점수는
 - 관찰된 공변량변수 벡터, x 가 사전확률로 주어진
 - 특정 처치집단($Z_i=1$ 혹은 0)의 조건부 확률로 표현될 수 있다.

$$\text{가중점수 } e(x_i) = \Pr(Z_i=1 | X_i=x_i)$$

- 위 식으로부터 주어진 $e=e(x)$ 에서 x 와 z 는 조건적으로 독립적임을 전제하면
(Rosenbaum과 Rubin(1983)의 Theorem 1)

$$\begin{aligned} \Pr(Z_1=z_1, \dots, Z_n=z_n | X_1=x_1, \dots, X_n=x_n) \\ = \prod e(x_i)^{z_i} \{1-e(x_i)\}^{1-z_i} \end{aligned}$$



□ 연구 과제

1. 성향가중치를 적용한 온라인조사와 전화조사/출구조사와의 비교.
2. 16대 투표행동 분석



II. 국내 온라인 선거예측 조사



1. 제 3회 지방선거 예측 조사

▶ 제 3회 지방선거를 위한 선거예측 조사에서 웹에어포리에는 인터넷 포털 사이트인 "다음 커뮤니케이션"의 활용 대상으로 실시한 온라인 여론조사 결과 오프라인 조사에 못지 않은 정확도를 보여 주었다.

【지역별 1, 2위 예측지 비교】

지 역	서울		부산		인천		경기		경원	
후 보	이명박	김민석	안상영	함여현	안상수	백상은	손학규	전 남	김권선	남동우
1위	53.6	42.9	63.1	19.6	60.0	29.9	60.0	35.5	77.6	22.4
2위	46.4	57.1	36.9	80.4	39.9	70.1	39.9	64.5	22.4	77.6
KBS	53.6	42.9	63.1	19.6	60.0	29.9	60.0	35.5	77.6	22.4
MBC	50.9	43.6	67.0	17.1	58.0	32.3	54.3	41.4	75.6	24.4
SBS	53.3	42.6	70.8	29.3	60.4	31.4	56.4	39.6	76.6	23.4



□ 서울지역 가구내 위치 구성비(2002. 5.)

	가구주	가구주배우자	가구주서녀	가구주부모	기타	계
인구센서스	42.3	29.0	17.5	3.0	8.2	100.0
온라인조사	43.7	34.0	18.6	2.5	3.3	100.0
전화조사	36.5	35.6	25.6	1.8	0.5	100.0

자료원 : 인구주택총조사=통계청 2000년 인구센서스 자료
 온라인조사=엔아이코리아 (2002. 5월, 유효표본 33,663명, 성향가중모형 적용)
 전화조사=코리아리서치/미디어리서치 (2002년 5월, 유효표본 1,600명)


2. 제 16대 대선 온라인 선거 여론 조사

- 패널모집: 다음커뮤니케이션의 활성화된 회원 250만명 대상 이메일 발송
- 패널규모: 141,813명
- 조사방법: 매 조사마다 이메일을 통해 조사(응답자 본인 확인)
- 조사시기 및 유효표본:

조사기간	유효표본	조사방법	표본 오차
08월 05일 ~ 08월 07일	30,848	온라인	±0.6
08월 18일 ~ 08월 20일	39,682	온라인	±0.5
09월 07일 ~ 09월 09일	36,957	온라인	±0.5
10월 12일 ~ 10월 14일	27,043	온라인	±0.6
11월 15일 ~ 11월 18일	33,969	온라인	±0.5
11월 22일 ~ 11월 24일	27,299	온라인	±0.6
12월 03일 ~ 12월 04일	29,051	온라인	±0.6
12월 10일 ~ 12월 12일	26,662	온라인	±0.6
12월 17일 ~ 12월 18일	25,629	온라인	±0.6
12월 19일 당일 해독조사	2,422	온라인	±2.0

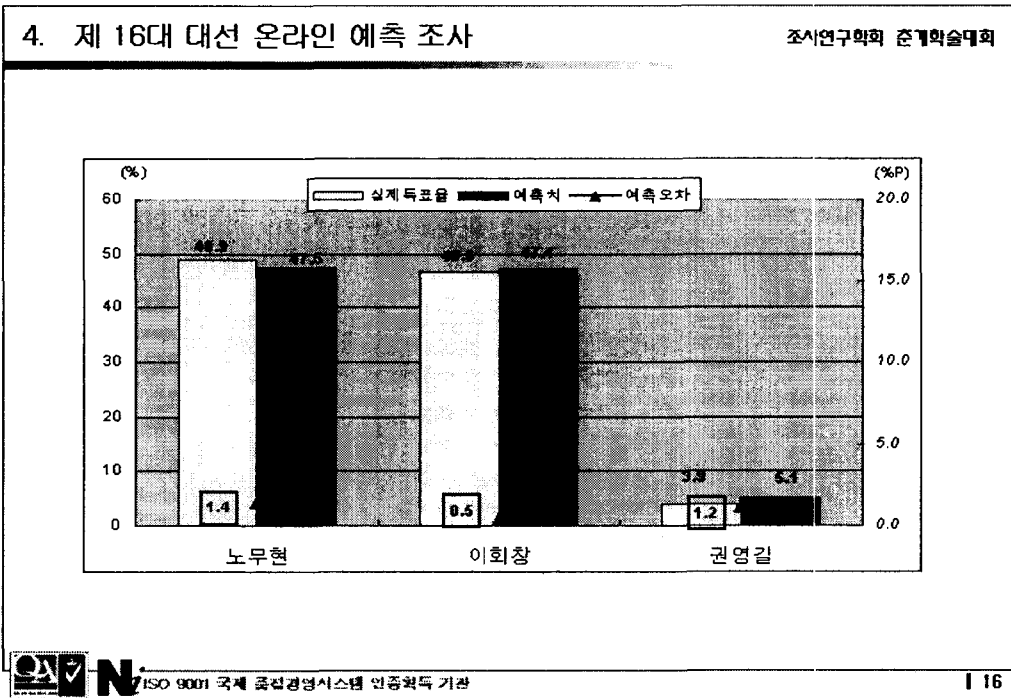
3. 제 16대 대선 여론조사 비교 조사연구학회 춘계학술대회

조사시기	유효표본	조사기관	식파기관	조사방법	노무현	이회창	권영길	기타	미도유보
8/5~8/7	30,848	연아이코리아		온라인	33.4	39.7	-	10.2	16.7
8/8	1,200	TNS	888	전 화	41.2	41.3	-		17.5
8/12	1,000	TNS	문화일보	전 화	26.1	43.0	4.6		26.3
8/19	1,030	길벗	조선일보	전 화	29.9	39.3	9.4		21.4
8/18~8/20	39,682	연아이코리아		온라인	32.2	40.6	-	7.9	19.5
8/19	1,030	길벗	조선일보	전 화	29.9	39.3	9.4		21.4
8/7~8/8	36,957	연아이코리아		온라인	27.7	36.7	3.3	-	32.3
8/8	1,553	KRC	동아일보	전 화	25.1	35.1	3.1		36.7
10/12~10/14	27,043	연아이코리아		온라인	34.2	39.6	5.0	-	22.2
10/18	1,000	TNS	문화일보	전 화	38.8	42.8	3.3		15.1
11/13~11/15	26,271	연아이코리아		온라인	36.9	39.6	4.1	1.5	19.0
11/14	1,517	KRC	동아일보	전 화	33.2	41.4	-	-	25.4
11/16~11/18	33,968	연아이코리아		온라인	37.1	39.2	4.0	1.5	19.2
11/17	640	길벗	MBC	전 화	38.3	42.3	-		19.4
11/22~11/24	27,299	연아이코리아		온라인	42.2	37.0	4.3	8.9	15.7
11/26	1,000	길벗	KB8	전 화	43.5	37.0	2.2	1.8	15.5



ISO 9001 국제 품질경영시스템 인증획득 기관

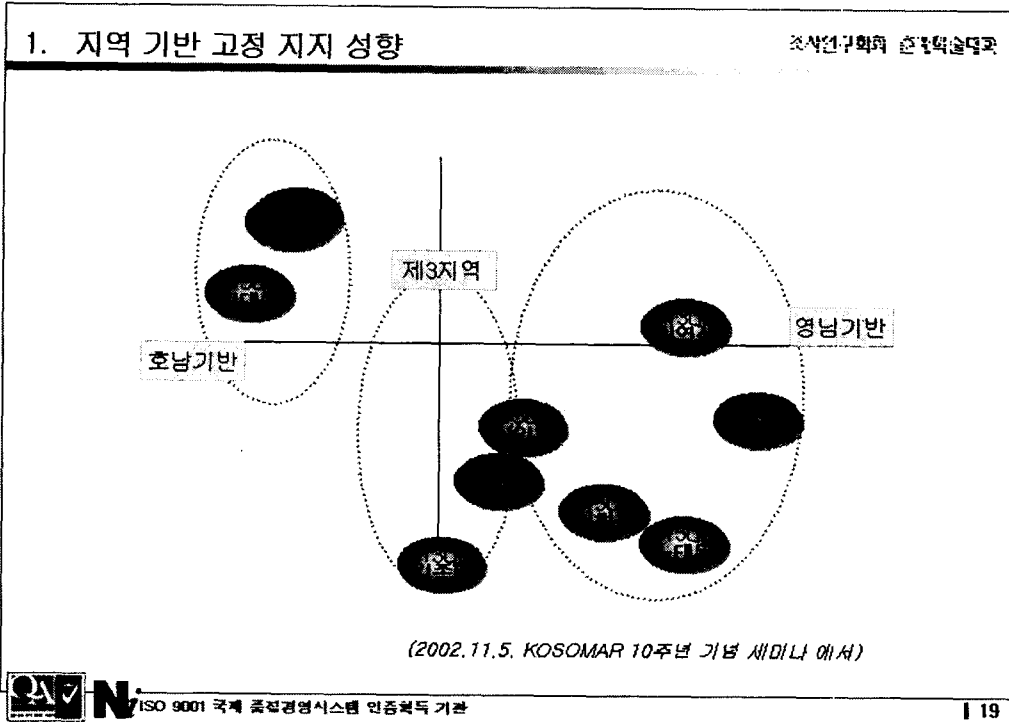
| 15



□ 방송사 출구조사/전화조사와의 비교

조사 개요			후보 득표율		예측 오차		
조사기관	접근방법	표본수	노무현 (%)	이회창 (%)	허용오차 (% P)	노무현 (% P)	이회창 (% P)
실제 득표율			48.9	48.5			
엔아이코리아	온라인 조사	2,422	47.5	47.1	<±2.0	-1.4	+0.5
KBS-미디어리서치	출구조사	40,000	49.1	46.8	<±0.5	+0.2	+0.2
MBC-KRC	출구조사	70,000	48.4	46.9	<±0.4	-0.5	-0.3
SBS-TNS코리아	전화조사	20,000	48.2	46.7	<±0.7	-0.7	+0.1
KBS-한국갤럽	전화조사	2,500	46.2	48.5	<±2.0	-2.7	+1.9

III. 16대 대선 투표행동



□ 13, 14, 15, 16대 대선 비교

조사연구학회 춘계학술대회

	13대				14대				15대				16대		
	노무현	김연삼	김대중	김종권	김연삼	김대중	김주연	김대중	이회창	이회창	이회창	노무현	이회창	김연삼	
전국	27.9	23.0	27.0	22.1	36.4	37.7	18.0	44.9	40.9	12.8	13.2	46.9	48.8	3.3	
서울	30.0	29.1	32.6	8.2	36.4	37.7	18.0	44.9	40.9	12.8	51.5	45.2	3.3		
수도권	41.0	28.1	22.1	33.7	36.6	31.9	22.3	39.1	35.7	23.5	50.9	44.6	4.5		
강원/제주	57.3	26.3	10.9	55.2	41.1	19.7	23.9	27.9	41.6	21.4	45.7	49.7	4.6		
충청권	33.1	20.1	11.9	34.6	36.9	27.8	27.3	43.9	27.4	26.6	53.0	41.7	5.3		
호남권	7.3	1.0	91.3	0.3	3.5	93.4	1.8	95.6	2.6	1.1	93.9	5.0	1.2		
경북권	68.1	26.6	2.5	2.4	62.5	8.9	17.3	13.1	66.9	17.7	20.3	75.8	3.9		
경남권	36.6	53.7	6.9	2.6	72.8	10.9	8.9	13.4	54.1	30.5	29.5	65.6	4.9		

ISO 9001 국제 품질경영시스템 인증획득 기관

20

2. 후보 단일화의 효과 얼마나?

조사연구학회 춘계학술대회

處. 단일화 후 지지율 상승

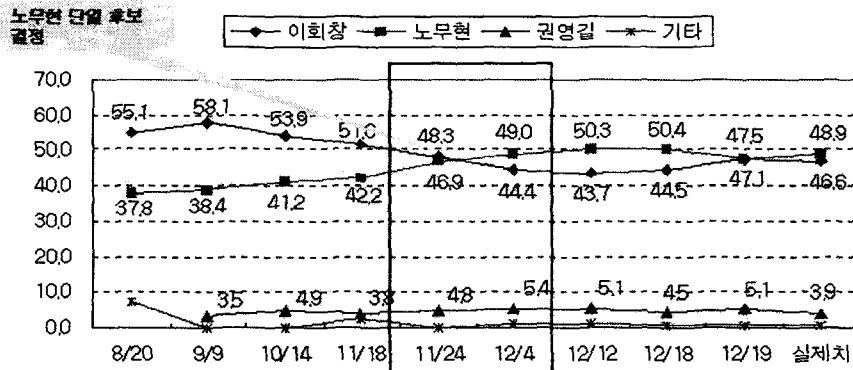
.....
 당초 점 후보를 지지한 유권자의 58.9%가
 노 후보 지지로 선회, 단일화의 시너지 효과가 돋보였다.
 이 후보에게는 24.6%가 갔다.

[한국일보] 2002년 11월 26일 (화) 17:29

조사연구학회 춘계학술대회

□ 11월 24일 후보단일화가 결정된 이후 노 후보 지지율 반전

<엔아이코리아 온라인 조사 예측치 변화율 추이>



□ 후보 단일화 효과

		12월 4일 조사				
		사례수	노무현	이회창	권영길	기타
11월 24일 조사	노무현	(3988)	25.3	4.7	0.5	0.0
	이회창	(5102)	4.3	34.6	0.2	0.0
	정몽준	(2361)	11.5	6.2	0.3	0.0
	권영길	(446)	2.0	0.9	0.5	0.0
	기타	(59)	0.2	0.2	0.0	0.0
태도유보		(1098)	4.9	3.4	0.1	0.0

Z=40.4, p < 0.001

	유입	유출	전체
노무현	23.0	5.3	17.7
이회창	15.5	4.5	11.0
권영길	1.1	2.9	-1.8
기타	0.0	0.5	-0.5

		12월 18일 조사				
		사례수	노무현	이회창	권영길	기타
11월 24일 조사	노무현	(3662)	25.5	1.5	0.6	0.1
	이회창	(4495)	1.5	32.0	0.5	0.1
	정몽준	(2743)	14.4	5.0	1.1	0.3
	권영길	(416)	0.8	0.3	2.0	0.0
	기타	(48)	0.2	0.0	0.0	0.1
태도유보		(1837)	8.1	4.7	0.9	0.2

Z=52.3, p < 0.001

	유입	유출	전체
노무현	25.0	2.2	22.6
이회창	11.7	2.1	9.6
권영길	3.1	1.1	2.0
기타	0.6	0.3	0.3

• 후보단일화 효과 : 노의 순 유입비율 : 17.7%



ISO 9001 국제 품질경영시스템 인증획득 기관

3. 정몽준 '지지철회' 효과

□ 정몽준 '지지철회' 표심 영향 적었다?

정몽준 '지지철회' 표심 영향 적었다

...

결국 '정몽준 파문'은 정 대표 지지자의 일부 이탈과 함께 노 당선자의 표를 잠식시키는 효과가 동시에 작용해 '제로섬'으로 마무리된 것으로 보인다.

...

[한겨레] 2002년 12월 20일 (금) 21:09



ISO 9001 국제 품질경영시스템 인증획득 기관

□ 가설1: 젊은층의 집결효과 있었나?

정몽준 지지 철회, 노무현 후보 당선에 오히려 긍정적 영향

...
 000는 "정 대표 지지철회로 정 대표 지지자들이 일부 투표에 참여하지 않고 실질적으로 노 후보 득표율을 대략 2% 정도 낮춘 것으로 추정되지만 이보다 노 후보 지지자들을 결집시키는 효과가 컸던 것으로 판단된다"고 말했다.

000는 "출구조사 결과 오후 3시30분 중간집계부터 노 후보가 역전하기 시작했는데 이때 20~30대 유권자들이 출구조사 대상에서 눈에 띄게 늘었다"고 말했다. 000도 "투표율이 저조하게 나오자 노 후보 지지자들이 결집한 것으로 판단된다"며 "이전까지 노 후보 지지자들의 충성도가 그리 높지는 않았으나 위기의식을 많이 느낀 것 같다"고 말했다.

[연합뉴스] 2002년 12월 19일 23시 31분



ISO 9001 국제 품질경영시스템 인증 획득 기관

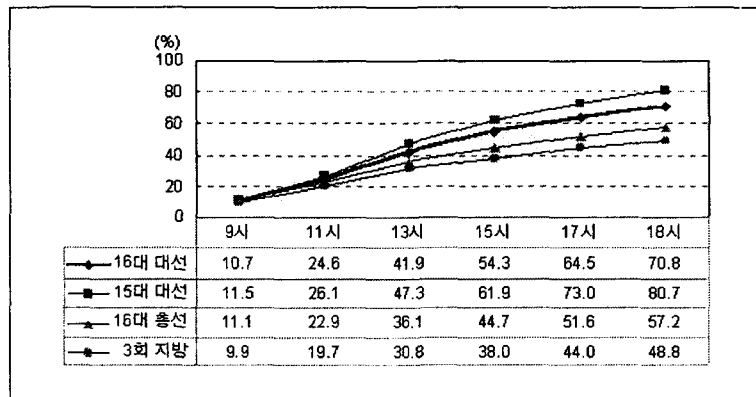
> 지역별 투표율 분석 - 역대 대선과 비교 전지역 투표율 최저

	투표율(%)			지역	투표율(%)		
	16대	15대	14대		16대	15대	14대
합계	70.8	80.7	81.9				
서울	71.4	80.5	81.4	강원	68.4	78.5	81.5
부산	71.2	78.9	83.2	충북	68.0	79.3	81.3
대구	71.1	78.9	78.5	충남	66.0	77.0	78.9
인천	67.8	80.0	80.3	전북	74.6	85.5	85.2
광주	78.1	89.9	89.1	전남	76.4	87.3	85.6
대전	67.6	78.6	80.3	경북	71.6	79.2	80.6
울산	70.0	81.1	-	경남	72.4	80.3	84.6
경기	69.6	80.6	80.4	제주	68.6	77.1	80.3



ISO 9001 국제 품질경영시스템 인증 획득 기관

> 시간대별 투표율 분석-과거 선거와 특별한 차이가 보이지 않음

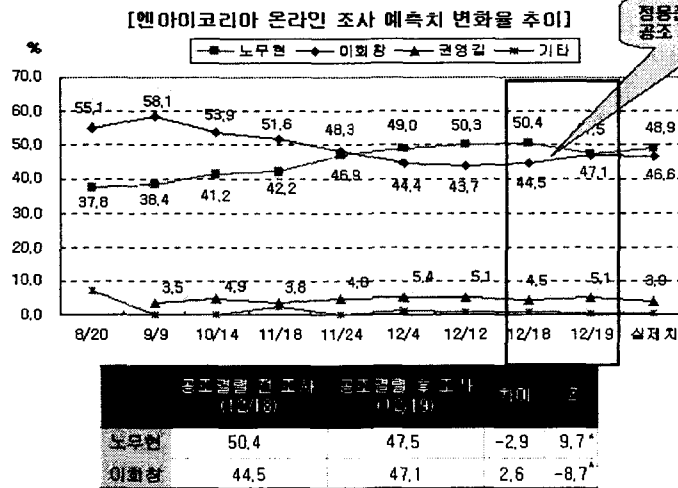


• 젊은층의 집결효과 있었다는 증거 없음



ISO 9001 국제 품질경영시스템 인증획득 기관

□ 가설2: 정몽준 지지철회 효과 정말 있었나?



ISO 9001 국제 품질경영시스템 인증획득 기관

조사연구학회 춘계학술대회

□ 노무현 유입된 표보다 유출된 표 더 많음

		12월 19일 조사				
		사례수	노무현	이회창	권영길	기타
12월 18일 조사	노무현	(467)	43.4	3.1	0.8	0.3
	이회창	(457)	1.0	45.8	1.9	-
	권영길	(35)	1.1	0.3	2.2	-
	기타	-	-	-	-	-

	유입	유출	전체
노무현	2.2	4.3	-2.1
이회창	3.4	2.9	+0.5

	노무현	이회창
멘아이코리아 최종 예측치와의 차이(%)	-2.9	2.6
패널 태도 변화 차이(%)	-2.1	0.5
"공조결렬"효과	2.5%P ±0.5	1.5%P ± 1.0

ISO 9001 국제 품질경영시스템 인증획득 기관
| 29

조사연구학회 춘계학술대회

4. 40대의 표심이 당락 결정

40대는 날마다 변한다

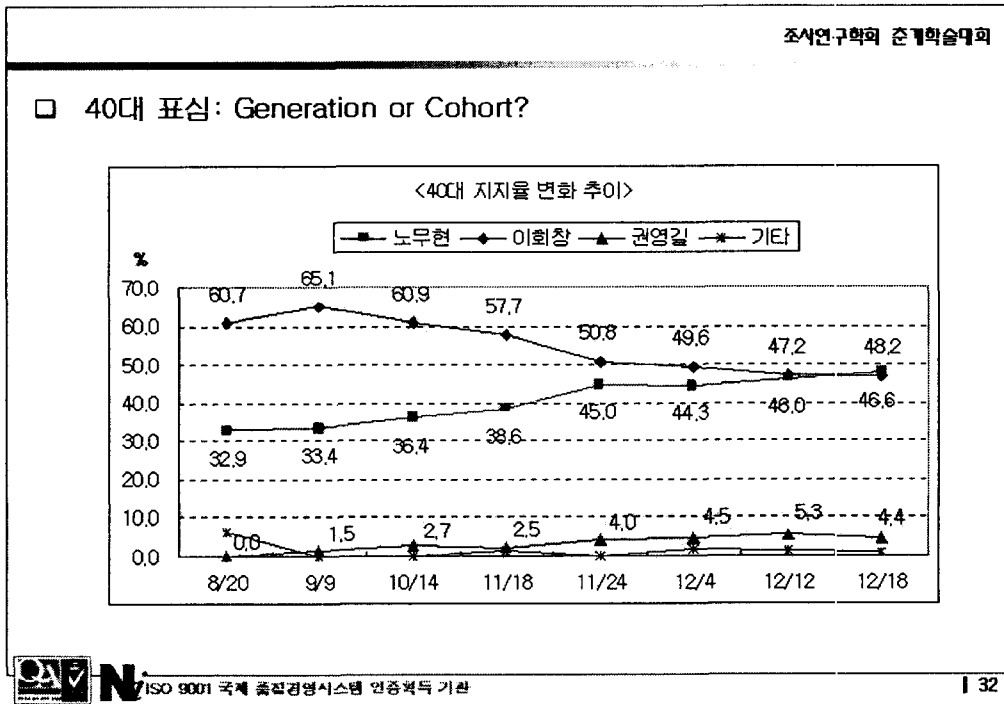
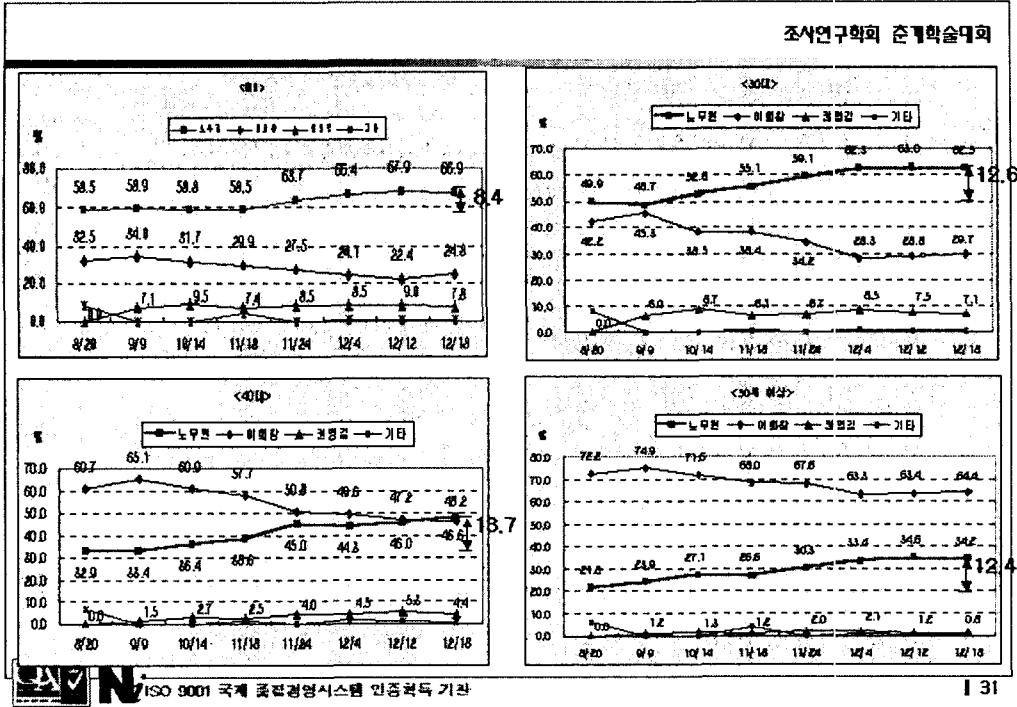
...

40대의 표심이 어떻게 요동치는 이유에 대해서는, 이들 심리 내부에 진보와 보수, 이상과 현실이 혼재돼 있기 때문이라는 것이 전문가들의 공통된 견해다. 사회 각 분야의 중심 축으로 일하는 탓에 안정성향을 보이면서도 70~80년대 민주화 격동기를 거친 세대라는 점에서 개혁을 지향하기도 하는 것이다.

...

[한겨레21] 2002년 12월 12일 (목) 14:18

ISO 9001 국제 품질경영시스템 인증획득 기관
| 30



▶ 연령구분에 대한 의사결정 나무 분석

- CHAID 사용
- 연령은 실수응답자료를 사용

	1	2	3	4	5	6	7	8	9	10	11	12
09/09 조사	20-22	22-28		28-31	31-35	35-38	39-43	43-48	48~			
10/14 조사	20-22	22-25	25-28	28-31	31-35	35-38	39-43	43-48	48~			
11/18 조사	20-22	22-27		27-30	30-34	34-38	38-42	42-47	47-50	50~		
11/24 조사	20-25		25-28	28-35		35-40		40-52		52~		
12/04 조사	20-22	22-25	25-31		31-35	35-38	39-43	43-48	48~			
12/12 조사	20-25		25-28	28-31	31-35	35-38	39-43	43-48	48-54	54~		
12/18 조사	20-28			28-32	32-37		37-42	42-46	46-50	50-53	53-58	58~



□ 39-42세 (37-43세)의 특성을 Cohort로 볼 수 있는가?

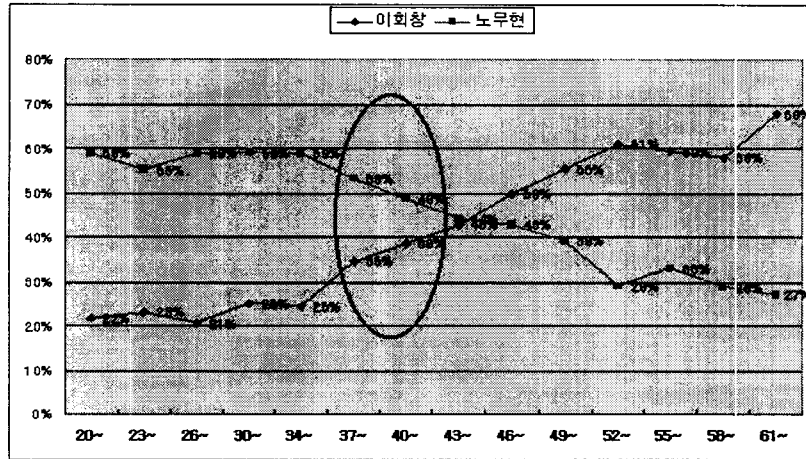


▶ 연령대별 출생시기와 대학생살시기

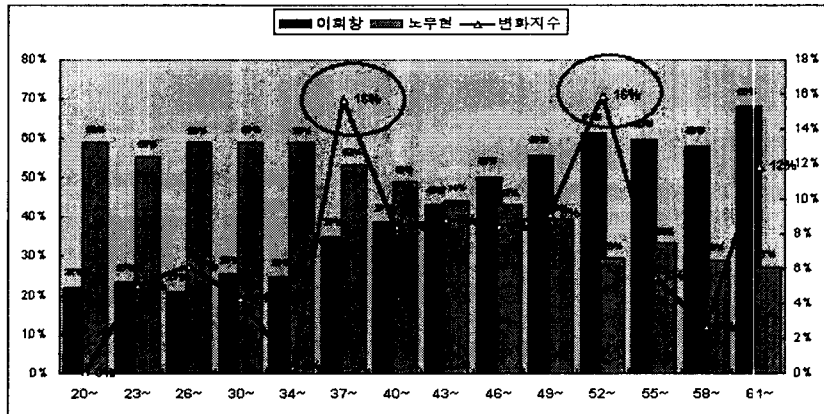
	20-28세	29-32세	33-35세	36-38세	39-42세	44-50세
1951-55						출생
1956-60						출생
1961-64				출생		
1965-68		출생	출생			
1970-74						
1975						대학
1976						대학
1977						대학
1978						대학
1978 10/25 12/12	출생					
1980						
1981						
1982						
1983						
1984						
1985						
1986						
1987						
1988						
1989						
1990						
1991-	대학	대학	대학			



▶ 후보 단순 지지율(12/18일 조사)

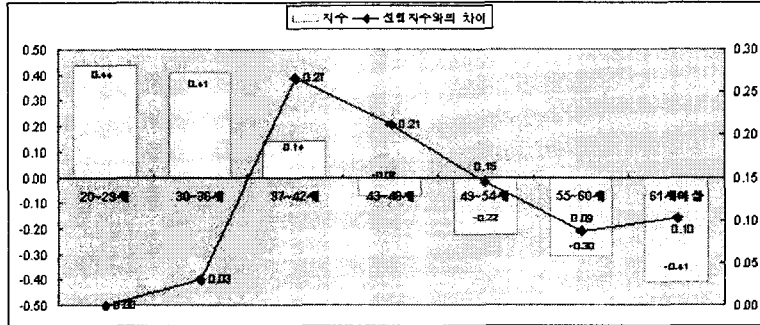


▶ 연령별 후보 지지율 차이의 변화율 (12/18 조사)



• 37-43세 들의 공유 경험이 세대내 고유한 Cohort 효과의 가능성 시사.

연령군집별 후보 지지패턴 지수 차이 (10/7~12/18 조사의 Aggregated 자료)



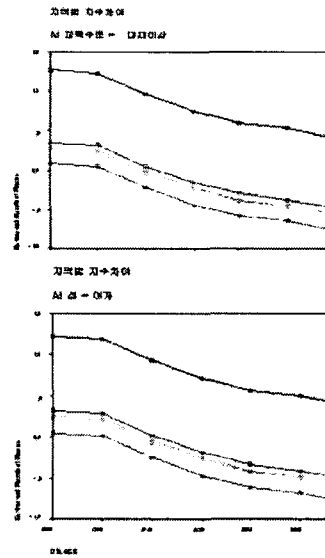
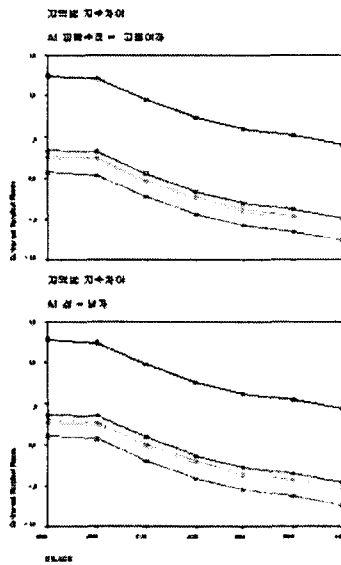
• 단, 지수=(노지지율-미지지율)/(노지지율+미지지율)
지수차이=지수_i - 지수_{i-1}

집단	n	1	2	3	4	5
11세미만	118	-0.407				
11-14세	148	-0.204	-0.204			
15-19세	140		-0.217	-0.217		
20-24세	140			-0.072		
25-29세	140				0.141	
30-34세	140					0.404
35-39세	140					0.459
40-44세	140					0.969

Tukey의 HSD



ISO 9001 국제 품질경영시스템 인증획득 기관



ISO 9001 국제 품질경영시스템 인증획득 기관