** FINAL PROJECT 2008 미네소타 주 대선 투표

목자 Contents

O1 연.

연구문제

02

분석 결과

03

분석 결론

01 연구문제

- 연구 배경과 의미

1) 연구배경과 의미

미네소타 주의 각 카운티별로 오바마의 투표율이 더 높은지에 대한 관련 공변량을 고려하는 이항 로지스틱 분석으로 추후 대선에서 저소득층 비율, 실업률 등을 고려하여 대선 정책을 결정하는데 도움을 줄 것이라 기대

2) 선행 자료 검토

미국은 우리나라와 달리 선거인단 제도를 채택하여, 인구비례로 주마다 선거인단을 뽑는다. 실제 대통령 및 부통령에게 투표하는 것은 선거인단이기 때문에, 대부분 유권자들(시민)의 의견과 일치 하지만 그렇지 않은 경우도 존재한다.

오바마(민주당)	매케인(공화당)
 1) 증세, 임금인상, 공공서비스 확충을 통한 부의 불균형 해소 2) 미국산업 및 근로자들의 일자리 보호 3) 기업 및 고소득층 증세를 통한 재정적자 해소 4) 최저임금 인상 5) 중산층 및 저소득층 지원을 통한 소비위축 완화 	1) 전반적 감세정책을 통한 기업활동 활성화 2) 미국 근로자들의 국제경쟁력 제고 3) 정부 지출 자제를 통한 재정적자 해소 4) 일자리 창출을 통한 고용불안 해소

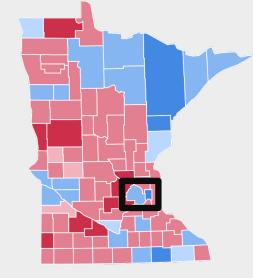
O1 연구문제

- 자료소개 및 가설 설정

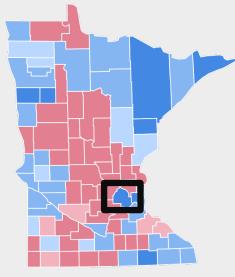
1) 자료소개

mn08.csv	mn08_income.csv	mn04.csv
County : 카운티 이름 Obama : 오바마 득표 수 McCain : 매케인 득표 수 pct_Obama : 오바마 득표율 pct_rural : 시골 거주 비율 medHHinc : 수입의 중앙값 unemp_rate : 실업자 비율 pct_poverty : 저소득층 비율 medAge2007 : 2007년 나이 중위수 medAge2000 : 2000년 나이 중위수 Gini Index : 소득격차	Rank : 소득 순위 Per_capital_income : 1인당 소득 Median_household_income : 가구당	County : 카운티 이름 pct_Kerry : 캐리 득표율 Kerry : 캐리 득표 수 pct_Bush :부시 득표율 Bush : 부시 득표 수 Num_vote_2004 : 투표 수
pct_native : 토착주민 비율	*출처 : 위키피디아	*출처 : 위키피디아
파생변수	파생변수	파생변수
Num_vote : 투표수 InFavor : 오바마 승리 여부	income : 카운티 별 소득 poverty : 저소득층 수	InFavor_04 : 캐리 승리 여부

2) 가설 설정



2004년 미네소타 주 투표 결과



2008년 미네소타 주 투표 결과

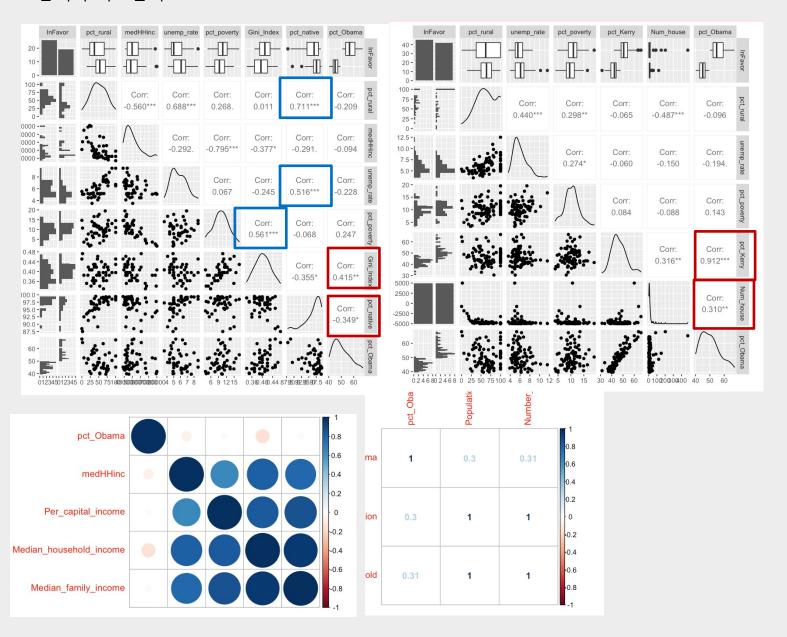
오바마의 득표율은 지난 대선의 민주당 후보의 득표율, 카운티 별 저소득층 비율, 시골 거주비율, 실업자 비율과 관련이 있는가?

- 탐색적 자료분석

candidate	InFavor_08	mean(pct_rural)	mean(medHHinc)	mean(unemp_rate)	mean(pct_poverty)	mean(medAge2007, na.rm = T)	mean(medAge2000)
<chr></chr>	<int></int>	<dbl></dbl>	<dbl></dbl>	<dbl></dbl>	<dbl></dbl>	<dbl></dbl>	<dbl></dbl>
McCain	0	62.955556	48851.867	6.1911111	9.733333	37.488462	37.742222
Obama	1	64.000000	46700.429	5.8095238	10.819048	37.323810	38.538095
compare	1	1.044444	-2151.438	-0.3815873	1.085714	-0.164652	0.795873

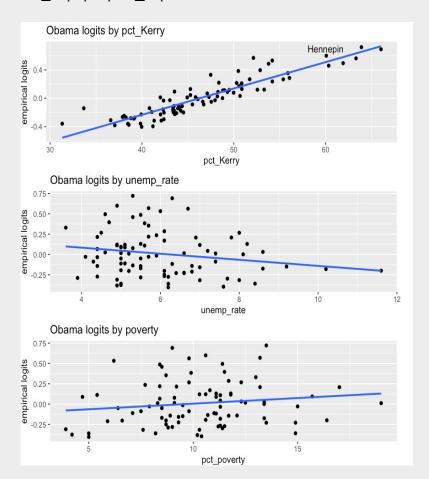
- •수입의 중앙값은 매케인을 지지하는 지역이 2151.438 높았다.
- •실업자의 비율은 매케인을 지지하는 지역이 0.38% 높았다.
- •저소득층의 비율은 오바마를 지지하는 지역이 1.08%높았다.
- •2000년 나이의 중앙값의 평균은 오바마를 지지하는 지역이 0.8정도 높았다.
- •소득 격차는 오바마를 지지한 지역이 0.016 높았다. (결측치 제거 후)
- •토착 주민 비율의 평균은 매케인을 지지한 지역이 1.46% 높았다.
- •전체적으로 투표수, 인구수의 평균이 많이 차이를 보이는데 오바마를 지지한 지역의 인구수 평균이 2배 이상 높다.
- •평균 소득에는 크게 차이가 없는 것으로 보인다.

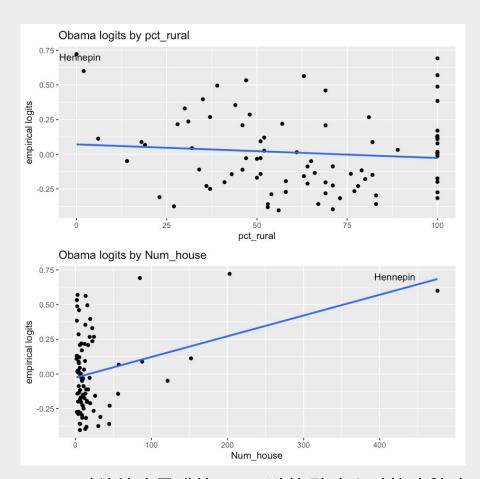
- 탐색적 자료분석



O2 분석결과

- 탐색적 자료분석





Pct_Kerry를 제외하고, 유의한 변수를 대체한 변수들이라 이상치가 존재하고, log 변환 결과 유의하지 않아 제거 선형관계가 뚜렷하다고 보긴 어렵지만, 선형성을 가정하고 분석 진행

- 모형적합

Model	공변량	AIC	Pr(>Chi)			
Model 1	pct_Kerry	4611.9		A matrix: 9 × 2		
Model 2	pct_Kerry, unemp_rate	3604.3	1.456571e-221		2.5 %	97.5 %
Model 3	pct_Kerry, unemp_rate, pct_poverty	3011	1.774127e-131	(Intercept)	0.2363557	0.2768768
				pct_Kerry	1.0367237	1.0375980
Model 4 pct_Kerry, unemp_rate	pct_Kerry, unemp_rate, pct_poverty, pct_rural	2963.2	1.712851e-12	unemp_rate	0.9099195	0.9388199
not Komo uno moto not novembro est	not Kormi unomp roto not novortiv not rurol		3.52681e-46	pct_poverty	1.0096001	1.0226131
Model 5	Model 5 pct_Kerry, unemp_rate, pct_poverty, pct_rural, unemp_rate:pct_rural	2761.6		pct_rural	0.9951343	0.9967327
				Num_house	1.0000507	1.0000990
pct_Kerry, unemp_rate, pct_poverty, pct_rural unemp_rate:pct_rural, pct_poverty:pct_rural unemp_rate: pct_poverty			unemp_rate:pct_rural	1.0005688	1.0008262	
	unemp_rate:pct_rural, pct_poverty:pct_rural,	2758.4	0.02609455	pct_poverty:pct_rural	0.9999603	1.0000572
	unemp_rate . pot_poverty			unemp_rate:pct_poverty	0.9978081	1.0003904

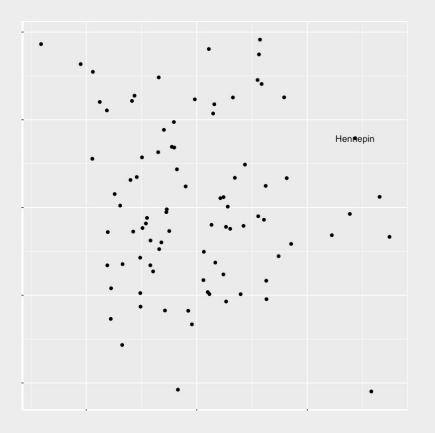
변수 간 다중 공선성: 7.8

적합결여가 있어 교호작용 항을 추가하였으나, 여전히 적합결여 발생

Model 6에서 pct_poverty:pct_rural, unemp_rate:pct_poverty는 교호작용이 유의하지 않음

- 1) 오바마의 득표율에 대한 시골 거주 비율은 저소득층 비율에 따라 통계적으로 유의하게 다르지 않음.
- 2) 오바마의 득표율에 대한 실업자 비율의 영향은 저소득층 비율에 따라 통계적으로 유의하게 다르지 않음.

- 모형적합



```
Call:
glm(formula = Obama/Num ~ pct_Kerry + unemp_rate + pct_poverty +
   pct_rural + unemp_rate:pct_rural, family = quasibinomial,
   data = final data, weights = Num)
Deviance Residuals:
    Min
                   Median
                                       Max
-10.5177
         -3.0625
                  -0.5144
                            3.5984
                                     9.5379
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)
                  -1.3068288 0.1143331 -11.430 < 2e-16 ***
pct_Kerry
                   0.0371635 0.0008244 45.079 < 2e-16 ***
unemp_rate
                  0.0114366 0.0021301
pct poverty
                                      5.369 7.38e-07 ***
                  pct rural
unemp_rate:pct_rural 0.0007717 0.0002610
                                       2.956 0.00408 **
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasibinomial family taken to be 23.3093)
   Null deviance: 92830.1 on 86 degrees of freedom
Residual deviance: 1887.5 on 81 degrees of freedom
AIC: NA
```

인구의 20%를 차지하고, 시골 거주 비율이 2%로 낮은 'Hennepin'이 이상치일까 확인해보았지만, 눈에 띄는 이상치는 아님

1-pchisq(2559,82) = 0으로, 표준오차 를 4.83배 팽창시켜, 과대산포를 조정했음 에도 적합결여는 해결되지 않음

03 분석 결과

- 최종 모델

최종 모델

$$\begin{split} \log(\frac{\hat{p}_i}{1-\hat{p}_i}) = &-1.307 + 0.037(\textit{pctKerry}) - 0.093(\textit{unemprate}) \\ + &0.011(\textit{pctpovert}) - 0.005(\textit{pctrural}) + 0.0008(\textit{unemprate}:\textit{pctrural}) \end{split}$$

모든 변수가 유의하며: i 지역에서의 오바마 득표율의 추정치

•pct_Kerry에 대한 추정 오즈비, 즉 캐리 득표율에 대한 지수 계수는 $e^{0.037}=1.038$ 로, 지난 대선 캐리의 득표율(%)가 1% 올라갈 때마다, 오바마의 득표율은 3.8% 증가

•unemp_rate에 대한 추정 오즈비, 즉 실업자율에 대한 지수 계수는 $e^{-0.093}=0.911$ 로, 카운티의 실업자 비율이 1% 증가할 때마다, 오바마의 득표율은 8.9% 감소

•pct_poverty에 대한 추정 오즈비, 즉 저소득층 비율에 대한 지수 계수는 $e^{0.011}=1.012$ 로, 저소득층 비율이 1% 증가할 때마다 오바마의 득표율은 1.2% 증가

•pct_rural에 대한 추정 오즈비, 즉 카운티별 시골 거주 비율에 대한 지수 계수는 $e^{-0.005}=0.955$ 로, 시골 거주 비율이 1% 증가할 때마다 오바마의 득표율은 0.5% 감소

•unemp_rate: pct_rural 에 대한 해석은 실업률에 따라 시골거주비율이 투표에 미치는 영향이 통계적으로 유의하게 다르다는 의미인데, 실업률이 0일 때, -0.005 => 시골거주 비율이 1% 증가하면 오바마의 득표율은 0.5% 감소, 실업률이 1%일 때, -0.098=> 시골거주비율이 1% 증가하면 오바마의 득표율이 9.3% 감소한다는 의미이다.

	2.5 %	97.5 %
(Intercept)	0.2163295	0.3386566
pct_Kerry	1.0361885	1.0395425
unemp_rate	0.8774724	0.9458908
pct_poverty	1.0072876	1.0157337
pct_rural	0.9925882	0.9982474
unemp_rate:pct_rural	1.0002600	1.0012841

각 항은 통계적으로 유의함

O3 분석 결론

- 연구결과

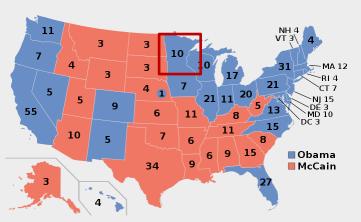
카운티별로 지난 대선에서 민주당의 후보의 득표율과 오바마의 득표율의 상관관계가 0.91로 매우 높은 것으로 보아, 이전에 지지하던 후보를 계속 지지하는 경향이 큼.

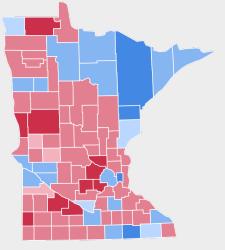
실업률에 따른 오바마의 득표율이 낮아지는 것으로 미루어보아, **매케인의 일자리 정 책의** 효과가 있었다고 추정해볼 수 있음.

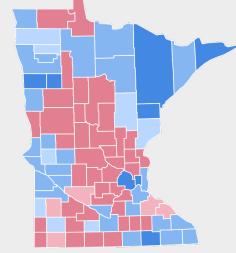
저소득층 비율이 유의한 공변량이라는 것으로 미루어보아, **오바마의 부의 불균형 해** 소 정책이 기여했다고 추정할 수 있음.

토착주민 비율과 관련이 높은 시골 거주 비율에 따른 오바마의 득표율이 낮아진 것은 지난 대선에 미네소타주의 각 카운티들 중 특히 시골에서 공화당의 지지율이 높았던

흐름이 이어짐.







- 03 분석 결론
- 결론의 정당성을 위한 가정과 한계점

종속변수와 관련이 있는 2가지 변수가 모두 절반 정도의 카운티에 결측값이 존재해, 이를 직접적으로 연구에 이용할 수 없음 (Gini_Index, pct_native)

따라서, 이를 대체한 변수들이 종속변수와 연관이 있다는 가정을 하고 분석을 진행

=> 관련 공변량으로 정확한 추정을 했다고 볼 수 없음

정책적 요소보다는 지난 대선의 영향을 많이 받는 선거였다는 점

적합 결여를 해결하지 못함

참고 문헌

미네소타주 1인당 평균 소득 (2006~2010)

https://en.wikipedia.org/wiki/List_of_Minnesota_locations_by_per_capita_income

2004년 미네소타 주 대선 결과

https://en.wikipedia.org/wiki/2004_United_States_presidential_election_in_Minnesota