



# 서울시 카페 상권 분석

Gibbs 변수 선택을 이용한 로지스틱 회귀 모형

김미성

Data

구분	변수명	설명	유형	출처
반응변수	y	1년 내 폐업 여부	(0, 1)	서울 열린데이터 광장 <small>&amp; 전제지연, 똑똑연정</small>
설명변수	month	개업월	범주형	
	gu	행정구	범주형	
	floor	층 구분 (1층 / 2층 이상 / 지하 / 지하+지상 / 1층+2층 이상)	범주형	
	total_floor	층수	연속형	
	total_area	총면적	연속형	
	franchise	프랜차이즈 여부	(0, 1)	
	near_cafe	주변 카페 수	연속형	
	near_fran_cafe	주변 프랜차이즈 카페 수	연속형	
	near_rest	주변 음식점 수	연속형	

## Data

구분	변수명	설명	유형	출처
설명변수	income	소득분위	연속형	서울시 우리마을가게 상권분석서비스
	household	가구 수	연속형	
	floating_pop	유동인구 수	연속형	
	living_pop	주거인구 수	연속형	
	office_pop	직장인구 수	연속형	
	female	여성인구 비율	연속형	서울 열린데이터 광장
	crosswalk	횡단보도 수	연속형	서울 열린데이터 광장 (위도, 경도)
	parking_lot	주차장 수	연속형	
	subway	지하철역 수	연속형	
	kospi	KOSPI 지수	연속형	Yahoo Finance
	blog	Naver 블로그 게시물 수	연속형	웹 크롤링

## Gibbs Variable Selection (GVS)

Model  $y_i \sim \text{Ber}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i \boldsymbol{\beta}$

Prior  $N\left(0, 100 \times \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{MLE})\right)$

Pseudo-prior  $N\left(\hat{\boldsymbol{\beta}}_{MLE}, \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{MLE})\right)$

## Gibbs Variable Selection (GVS)

Model  $y_i \sim \text{Ber}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i \boldsymbol{\beta}$

Prior  $N\left(0, 100 \times \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{MLE})\right)$

Pseudo-prior  $N\left(\hat{\boldsymbol{\beta}}_{MLE}, \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{MLE})\right)$

```
model{
  for(j in 1:k){ gbeta[j] <- gamma[j]*beta[j] }

  for(i in 1:n){
    y[i] ~ dbern(p[i])
    logit(p[i]) <- inprod( X[i, 1:k], gbeta[1:k] )
  }

  for(j in 1:k){
    gamma[j] ~ dbern(0.5)
    beta[j] ~ dnorm( mu.beta[j], tau.beta[j] )
    mu.beta[j] <- (1-gamma[j])*beta.mle[j]
    tau.beta[j] <- gamma[j]/(100*var.beta.mle[j]) +
      (1-gamma[j])/var.beta.mle[j]
  }
}
```

## 가능한 모형에 대한 사후확률 비교

$\gamma$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	사후확률
M1	0	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0.0003333
M2	0	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0.0003333
M3	0	0	1	0	1	1	0	0	1	1	1	0	0	1	0	1	0	0	1	1	0.0003000
M4	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	1	0	0	1	0	0.0003000
M5	0	1	1	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	1	1	0.0003000
M6	0	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	0	0.0003000
M7	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0.0003000
M8	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	0.0002667
M9	0	0	1	0	1	1	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0.0002667
M10	0	1	1	0	1	1	1	0	1	0	0	0	1	1	0	1	0	0	0	1	0.0002667

## 가능한 모형에 대한 사후확률 비교

$\gamma$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	사후확률
M1	0	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0.0003333
M2	0	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0.0003333
M3	0	0	1	0	1	1	0	0	1	1	1	0	0	1	0	1	0	0	1	1	0.0003000
M4	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	1	0	0	1	0	0.0003000
M5	0	1	1	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	1	1	0.0003000
M6	0	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	0	0.0003000
M7	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0.0003000
M8	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	0.0002667
M9	0	0	1	0	1	1	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0.0002667
M10	0	1	1	0	1	1	1	0	1	0	0	0	1	1	0	1	0	0	0	1	0.0002667

## 가능한 모형에 대한 사후확률 비교

$\gamma$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	사후확률
M1	0	1	1	0	1	1	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0.0003333
M2	0	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0.0003333
M3	0	0	1	0	1	1	0	0	1	1	1	0	0	1	0	1	0	0	1	1	0.0003000
M4	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	1	0	0	1	0	0.0003000
M5	0	1	1	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	1	1	0.0003000
M6	0	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	0	0.0003000
M7	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0.0003000
M8	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	0.0002667
M9	0	0	1	0	1	1	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0.0002667
M10	0	1	1	0	1	1	1	0	1	0	0	0	1	1	0	1	0	0	0	1	0.0002667



## 변수 선택 방법 비교 및 최종 모형

방법	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	개수
GVS		●	●		●	●	●				●		●			●			●	●	10
MLE		●	●		●	●								●							5
stepAIC	●	●	●		●	●	●	●						●	●	●	●				11

## 변수 선택 방법 비교 및 최종 모형

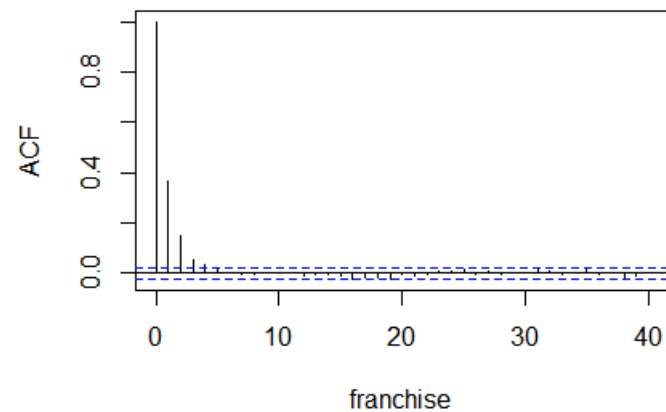
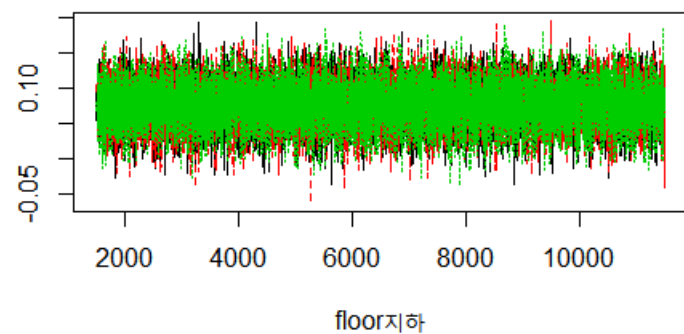
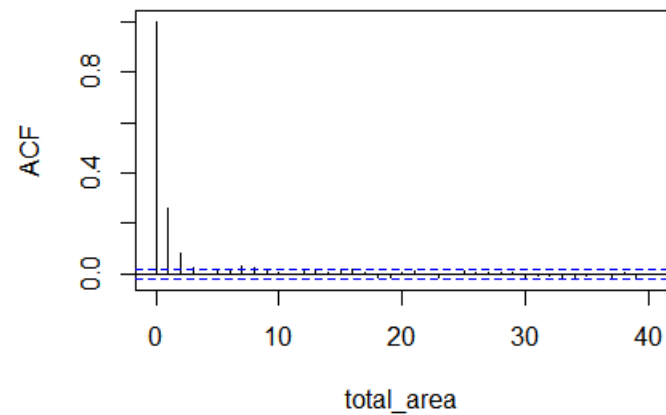
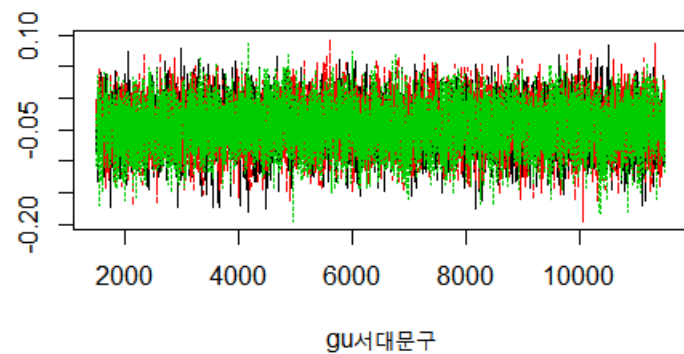
방법	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	개수
GVS		●	●		●	●	●				●		●			●			●	●	10
MLE		●	●		●	●								●							5
stepAIC	●	●	●		●	●	●	●						●	●	●	●				11



최종 모형의 설명변수

행정구, 층 구분, 총면적, 프랜차이즈 여부, 주변 카페 수, 가구 수,  
주거인구 수, 횡단보도 수, KOSPI 지수, 블로그 게시물 수

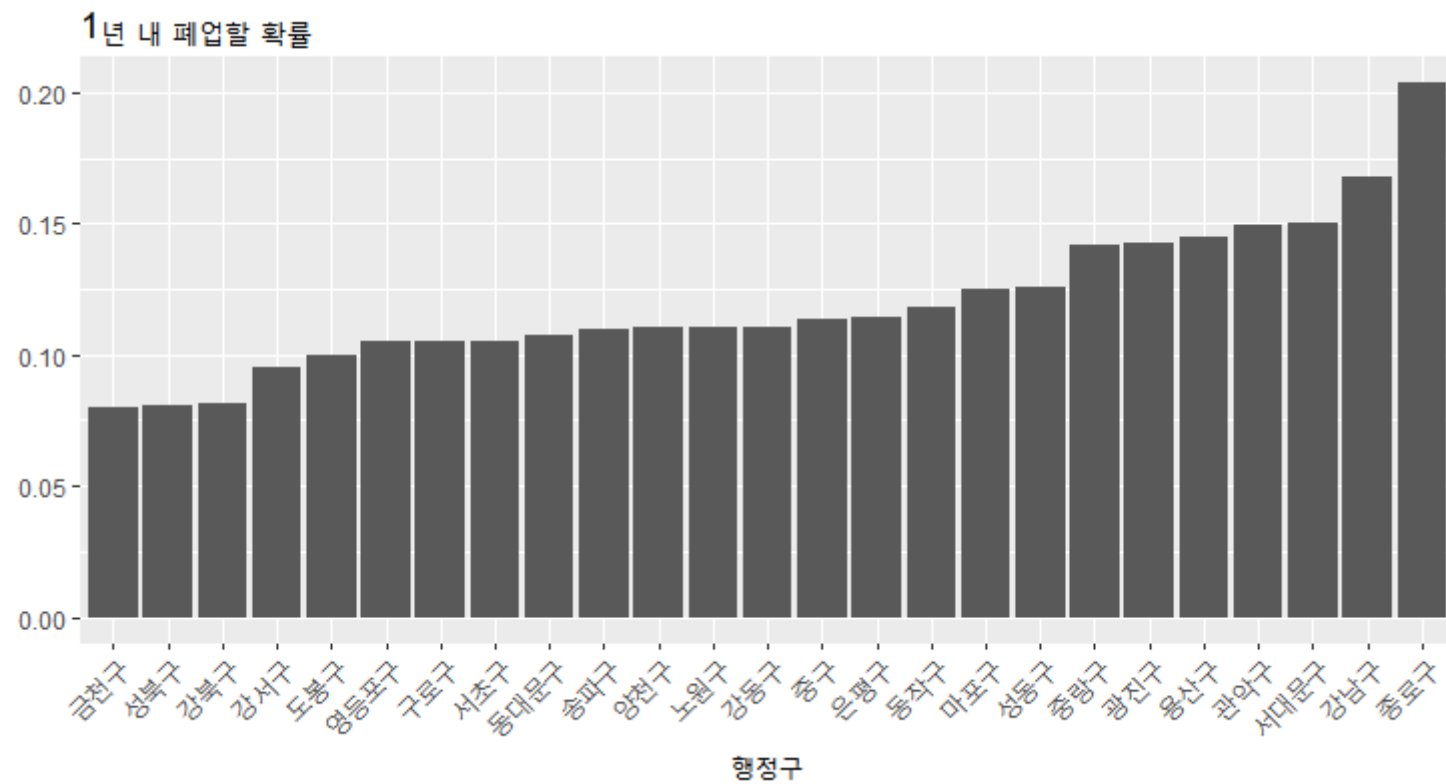
## 로지스틱 회귀 모형의 MCMC 수렴 진단



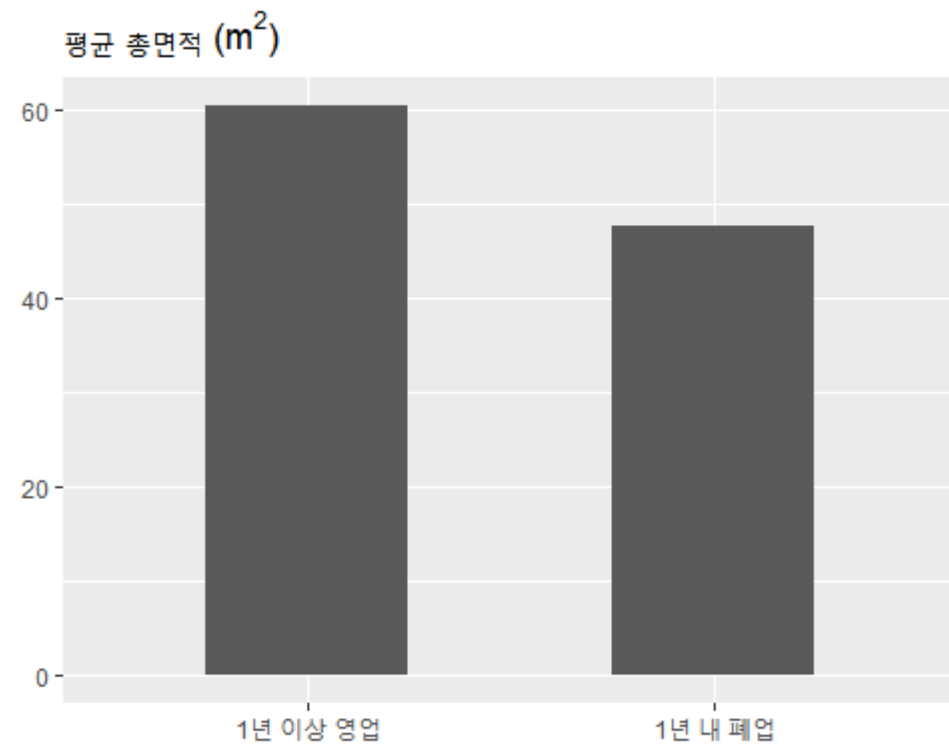
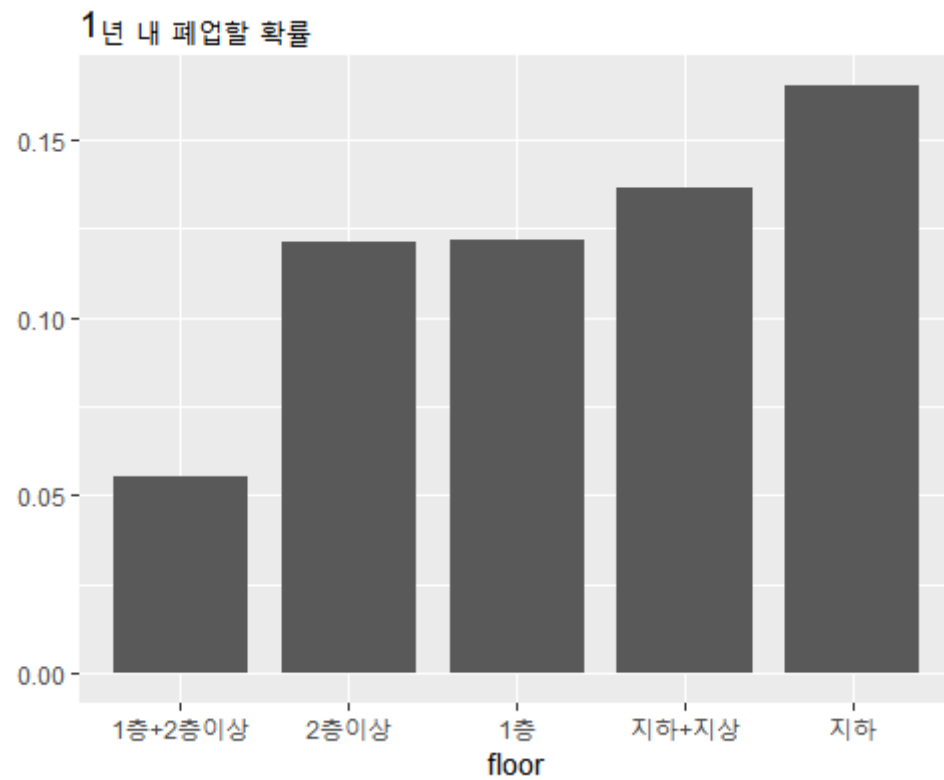
Gelman 상수

$1.002823 < 1.1$

## 카페가 위치한 행정구에 따른 평균 폐업률



## 층 구분 및 총면적과 폐업률의 관계



---

감사합니다

