



# **베이지안 분위 회귀분석을 이용한 서울시 골목상권 외식업종 매출액 추정**

182STG18 이하경



# CONTENTS

## 001 데이터 설명

- 서울시 우리마을가게 상권분석서비스

## 002 분석 모형 설정

- JAGS 우도함수 직접 지정
- 김스변수선택

## 003 분석 결과

- 분위에 따른 회귀계수 비교
- 분위에 따른 최종 선택변수 비교

## 004 결론



001

데이터설명

---

# 서울시 '우리마을가게 상권분석서비스'

data.seoul.co.kr



- 서울시 '열린데이터광장'에 개방되어 있는 공공데이터
- 2018년 1분기~3분기의 자료만을 정제하여 사용
- 10가지 외식업종으로 한정

(한·중·일·양식음식점, 분식전문점, 치킨전문점,  
패스트푸드점, 제과점, 커피·음료, 호프·간이주점)

- 1000개의 표본을 추출하여 사용

## 상권-추정매출

- 분기별 월평균 추정매출
- 주중/주말 매출 비율
- 시간대별 매출 비율
- 성별 매출 비율
- 연령대별 매출 비율
- 상권 영역 내 점포수

## 상권-소득소비

- 상권 영역 별 월평균 소득
- 상권 영역 별 총지출(소비)

## 상권-인구

- 상권 영역 별 총 유동인구(월)
- 상권 영역 별 총 상주인구(월)

# 설명변수와 반응변수 정의

RESPONSE AND PREDICTORS



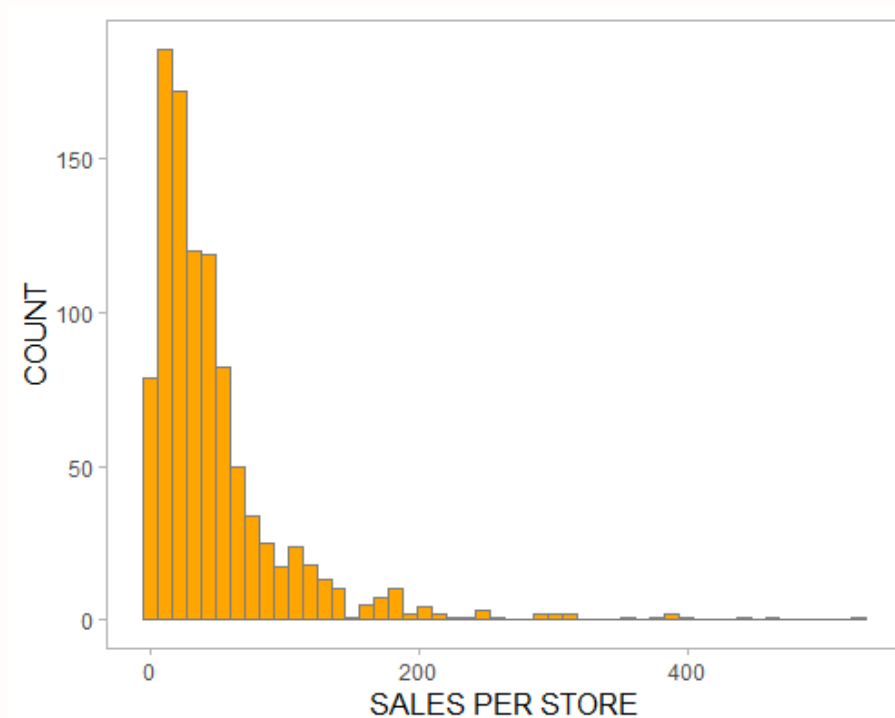
구분	변수		단위	비고
상권별 추정매출 (분기)	Y	점포당 월 평균 매출 금액	100만원	업종별 월평균매출액 / 점포 수
	X1	주말 매출 비율	%	
	X2	11~14시 매출 비율		
	X3	14~17시 매출 비율		
	X4	17~21시 매출 비율		
	X5	21시~24시 매출 비율		
	X6	여성 매출 비율		
	X7	10대 매출 비율		
	X8	20대 매출 비율		
	X9	30대 매출 비율		
	X10	점포 수	개	
상권별 유동인구(월)	X11	총 유동인구 수	명	분기별 평균 사용
상권별 상주인구(월)	X12	총 상주인구 수	명	분기별 평균 사용
상권별 소득소비(월)	X13	월 평균 소득 금액	원	분기별 평균 사용
	X14	총 소비(지출) 금액	원	분기별 평균 사용

# 점포당 월평균매출액의 히스토그램

MONTHLY SALES PER STORE



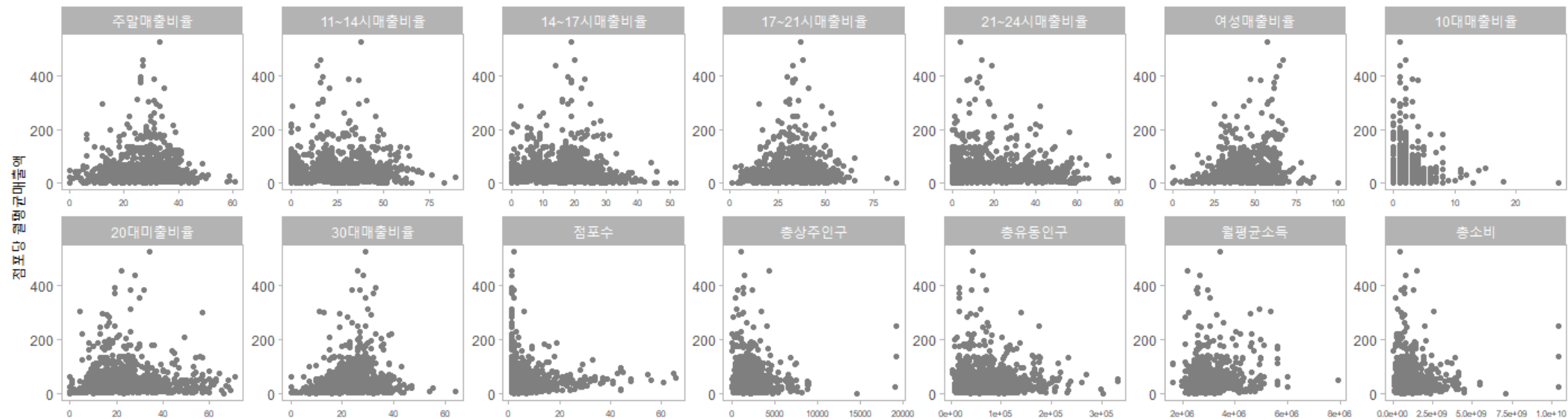
- 오른쪽으로 꼬리가 긴 비대칭적 분포



Min	Q1	Mean	Q2	Q3	Max
0.099	15.166	32.372	48.818	58.037	526.705

# 반응변수와 설명변수의 산점도

SCATTERPLOTS BETWEEN VARIABLES



- 설명변수 X에 따라 Y의 분산이 일정하지 않음 (등분산성 가정 X)



002

## 모형 설정

# JAGS를 활용한 분위 회귀계수의 사후표본 생성

- $p=0.1, 0.25, 0.5, 0.75, 0.9$ 의 총 5가지 분위수
- 비대칭 라플라스 (ALD) 분포의 밀도함수를 사용해 우도함수를 직접 지정 (0-1 기법)
- 김스변수선택(GVD)을 위해 Gamma의 사전 분포를 추가로 정의함
- 사전 설정

nChains	nAdapt	nUpdate	nIter	thin
3	10000	10000	20000	10





003

분석 결과

---

# 분위에 따른 회귀계수 비교

BAYESIAN QUANTILE REGRESSION



# 분위에 따른 회귀계수 비교

BAYESIAN QUANTILE REGRESSION

양의 영향



# 분위에 따른 회귀계수 비교

BAYESIAN QUANTILE REGRESSION

음의 영향



# GVS 변수선택 결과 비교

GIBBS VARIABLE SELECTION



\*분위(상대빈도)

	p	주말	11·14시	14·17시	17·21시	21·24시	여성	10대	20대	30대	점포수	상주인구	유동인구	평균소득	총지출
GVS 동시선택 사후평균	0.1 (28%)	●	●	●	●		●			●	●	●		●	●
	0.25 (53%)	●	●	●	●		●		●	●	●	●	●	●	●
	0.5 (62%)	●	●	●	●		●		●	●	●	●	●	●	●
	0.75 (81%)	●	●	●	●	●	●	●	●	●		●	●	●	●
	0.9 (92%)	●	●	●	●	●	●	●	●	●	●	●	●	●	●
stepAIC		●	●	●		●	●		●	●		●		●	●

# GVS 변수선택 결과 비교

GIBBS VARIABLE SELECTION

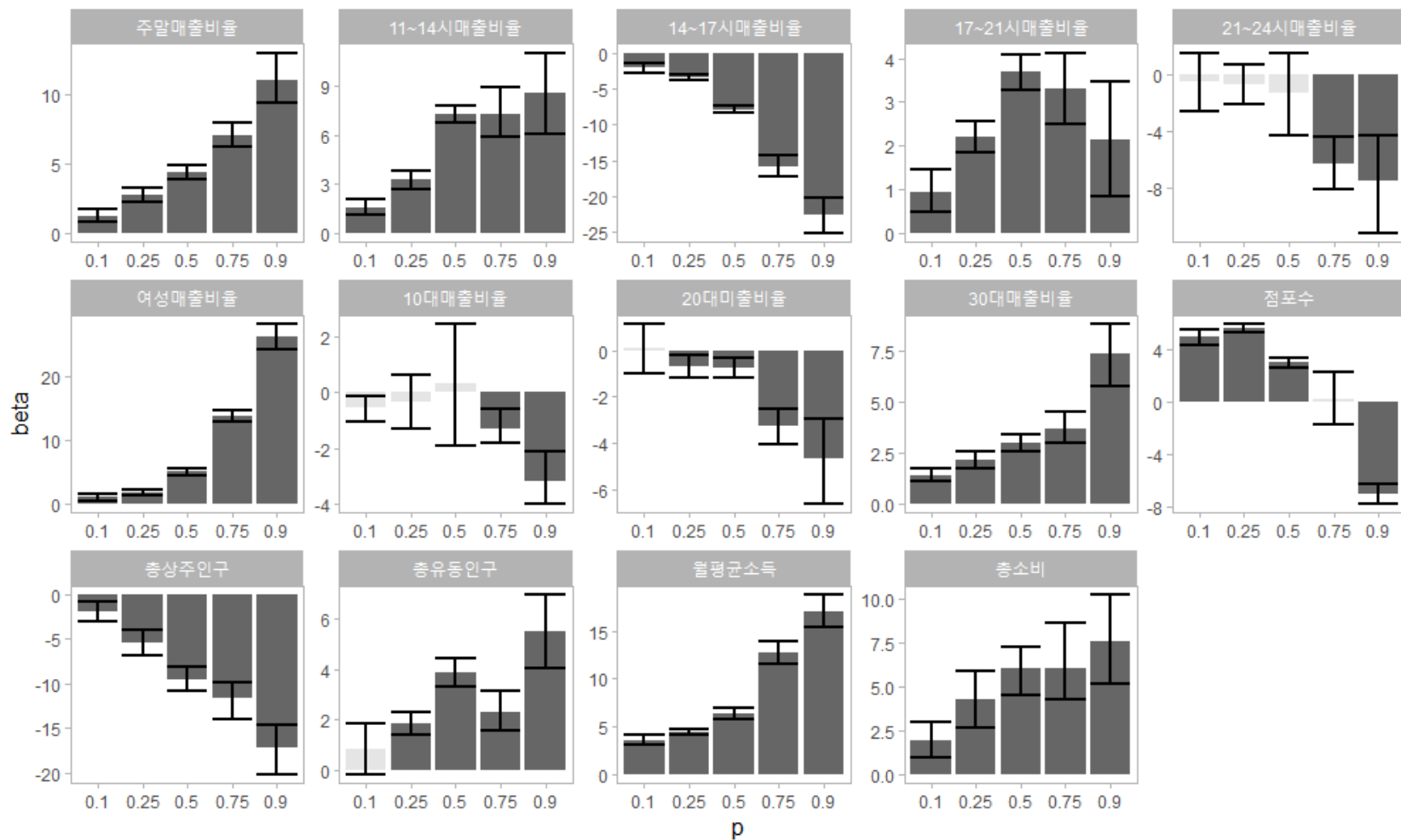


\*분위(상대빈도)

p		주말	11·14시	14·17시	17·21시	21·24시	여성	10대	20대	30대	점포수	상주인구	유동인구	평균소득	총지출
GVS 동시선택 사후평균	0.1 (28%)	●	●	●	●		●			●	●	●		●	●
	0.25 (53%)	●	●	●	●		●		●	●	●	●	●	●	●
	0.5 (62%)	●	●	●	●		●		●	●	●	●	●	●	●
	0.75 (81%)	●	●	●	●	●	●	●	●	●		●	●	●	●
	0.9 (92%)	●	●	●	●	●	●	●	●	●	●	●	●	●	●
stepAIC		●	●	●		●	●		●	●		●		●	●

# 베이지안 사후 추론: 사후평균 및 표준편차와 95% 신뢰구간

BAYESIAN INFERENCE



# 베이지안 사후 추론: 사후평균 및 표준편차와 95% 신뢰구간

BAYESIAN INFERENCE



	p=0.1		p=0.25		p=0.5		p=0.75		p=0.9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
INTERCEPT	10.438	0.240	18.047	0.224	33.319	0.232	58.191	0.388	103.853	0.663
주말	1.220	0.247	2.765	0.257	4.377	0.258	7.039	0.455	11.111	0.927
11-14시	1.568	0.266	3.274	0.295	7.254	0.267	7.305	0.770	8.582	1.290
14-17시	-2.066	0.331	-3.352	0.226	-7.882	0.259	-15.833	0.763	-22.613	1.233
17-21시	0.937	0.243	2.202	0.186	3.683	0.198	3.292	0.414	2.135	0.643
21시-24시	-	-	-	-	-	-	-6.309	0.926	-7.523	1.748
여성	0.934	0.298	1.709	0.273	5.050	0.275	13.825	0.480	26.296	0.998
10대	-	-	-	-	-	-	-1.321	0.337	-3.194	0.488
20대	-	-	-0.724	0.243	-0.748	0.219	-3.240	0.384	-4.649	0.932
30대	1.400	0.166	2.154	0.217	2.958	0.223	3.700	0.387	7.392	0.793
점포수	4.893	0.290	5.583	0.164	2.946	0.185	-	-	-7.018	0.367
상주인구	-1.928	0.527	-5.433	0.772	-9.626	0.685	-11.680	1.059	-17.217	1.395
유동인구	-	-	1.832	0.229	3.892	0.293	2.317	0.398	5.515	0.747
평균소득	3.530	0.261	4.394	0.155	6.275	0.307	12.687	0.618	17.099	0.875
총지출	1.956	0.522	4.237	0.827	6.033	0.687	6.064	1.133	7.624	1.265





004

결론

## 골목상권에 대한 베이지안 분위 회귀분석

- 실제 골목상권 자료에 적용 결과 매출액의 분위에 따라 설명변수의 영향력이 매우 다름
  - (+) 주말, 여성, 30대 연령의 매출 비율이 큰 상권영역의 점포당 매출액은 크게 나타나며 높은 분위일수록 영향력이 더 큼
  - (-) 오후(14~17시), 20대 연령의 매출 비율, 상권 내 동일 업종 점포 수, 상주 인구가 클수록 점포당 매출액은 적게 나타나며 높은 분위일수록 변수들의 영향력이 더 큼
- 조건부 평균에만 초점을 둔 평균 회귀분석에 비해 보다 다양한 정보를 얻게 되었음



감사합니다 :)