

선형회귀분석을 바탕으로 은평 한옥마을 방문객들의 경향성 파악 및 방문 장소 추천 시스템 제작

연구자: 신상우, 김준석, 박준서,
신민성, 이준서, 임건, 조민경

《 초 록 》

서울시 은평구 진관동에 위치한 은평한옥마을은 은평구의 대표적인 관광 명소로, 단순히 한옥뿐만이 아닌 근처 북한산의 아름다운 절경, 명사인 진관사 등 다양한 관광지가 존재해, 많은 관광객의 발길이 끊이지 않는 곳이다. 따라서 은평한옥마을을 찾는 관광객들의 취향과 방문 목적에 따라 은평한옥마을의 관광 코스는 매우 다양하게 나타난다. 그러나 방문 목적을 정하지 않은 상태에서 의도치 않게 은평한옥마을을 방문하게 되거나 은평한옥마을 근처에 존재하는 관광지에 어떤 것들이 있는지 모르는 사람들의 경우에는 관광 코스를 정하기 어려운 측면이 존재한다. 더불어, 은평한옥마을을 방문한 사람들이 느끼는 만족도 역시 나이, 성별, 동행 인원수 등의 조건에 따라 매우 상이하다.

따라서 본 연구는 은평한옥마을을 방문한 적이 있는 방문객들의 관광지 선호 경향성 데이터를 나이, 성별, 동행 인원수 등의 조건에 따라 정리하고, 정리한 데이터를 바탕으로 선형 회귀 분석을 이용해 은평한옥마을의 관광지 추천 시스템을 개발하였다. 제안한 분석 방법을 통해 은평한옥마을을 방문하는 사람들의 나이, 성별, 동행 인원수에 따라 방문객의 만족도를 조사하고, 각 방문객들이 데이터를 입력할 경우 각 장소별 만족도를 추천해주는 프로그램을 선형회귀모델링을 바탕으로 고안한다.

목 차

I. 서론	03
1. 연구 동기	03
2. 연구 목적	03
II. 이론적 배경	03
1. 선형 회귀 분석	04
2. 가변수	04
III. 연구 방법	04
1. 연구내용 및 방법	04
I) 설문지 내용	
II) 설문지 조사 방법	
III) 독립변수와 종속변수의 설정	
IV) 데이터 분석 방법	
V) 사용한 R 코드 및 R library	
IV. 결과 및 고찰	10
1. 설문 결과	
2. 분석 결과	
3. 사용자로의 적용	
V. 결론 및 제언	25
1. 결론 및 제언	25
VI. 참고 문헌	26

I. 서론

1. 연구 동기

은평구에 위치한 은평한옥마을은 은평구의 대표적인 관광 명소로, 단순히 한옥뿐만이 아닌 근처 북한산의 아름다운 절경, 명사인 진관사 등 다양한 관광지가 존재해, 많은 관광객의 발길이 끊이지 않는 곳이다. 따라서 은평한옥마을을 찾는 관광객들의 취향과 방문 목적에 따라 은평한옥마을의 관광 코스는 매우 다양하게 나타난다. 그러나 방문 목적을 정하지 않은 상태에서 의도치 않게 은평한옥마을을 방문하게 되거나 은평한옥마을 근처에 존재하는 관광지에 어떤 것들이 있는지 모르는 사람들의 경우에는 관광 코스를 정하기 어려운 측면이 존재한다. 더불어, 은평한옥마을을 방문한 사람들이 느끼는 만족도 역시 나이, 성별, 동행 인원수 등의 조건에 따라 매우 상이하다.

본 연구는 이에 착안하여, 관광객이 자신의 정보를 입력 시 각 장소별 만족도를 추천해주는 프로그램을 개발함에 목표를 둔다.

2. 연구 목적

본 연구는 은평한옥마을 관광 만족도 향상을 위하여 방문자 맞춤형 관광 장소 추천 프로그램을 제작하는 것을 목적으로 한다. 문화체육관광부의 '관광사업체 기초통계조사'에 따르면 전체 관광편의시설업에서 한옥체험업 업체가 차지하는 비중은 2009년에는 0.7% 수준이었지만, 2017년 한옥체험업체가 관광편의시설업에서 차지하는 점유율은 32%를 넘어섰다. 한옥체험 관광의 인기가 상승세를 그리는 가운데, 은평한옥마을에서는 한옥 체험뿐만 아니라, 경치 좋은 북한산을 보며 등산, 카페 방문, 진관사 방문 등 다양한 것들을 할 수 있기 때문에 관광객 수가 꾸준히 증가하고 있다. 그러한 관광객의 특성을 고려하여 새롭게 은평한옥마을에 방문하는 사람들을 위해 누적된 방문자들의 만족도를 제공하는 프로그램을 만드는 것이 목적이다. 이후 은평한옥마을을 방문한 사람들의 장소별 선호도의 경향성을 파악하여 미래의 관광지로서의 은평한옥마을 개발에 대한 제안을 하고자 한다.

II. 이론적 배경

1. 선형 회귀 분석

회귀분석은 여러 자료들 간의 관계성을 수학적으로 추정하고 설명하는 자료 분석 방법으로 인과관계가 상정된 모델을 구현할 수 있다. 회귀분석을 통해 종속변수와 독립변수 간에 선형관계가 존재하는지, 종속변수에 영향을 미치는 독립변수가 유의미한지와 영향력의 정도를 추정된 회귀모형을 통해 종속변수의 예측치를 알 수 있다. 회귀모형의 형태에 따라 하나의 종속변수에 대해 독립변수가 하나인 경우를 단순회귀분석, 하나의 종속변수에 대해 독립변수가 둘 이상인 경우를 다중회귀분석이라 한다.

선형회귀는 종속 변수 y 와 한 개 이상의 독립변수 x 와의 상관관계를 모델링하는 회귀 분석 기법이다. 한 개의 독립변수가 있으면 단순 선형 회귀, 둘 이상이면 다중 선형 회귀라 한다. 선형 회귀는 선형 예측 함수를 사용해 회귀식을 모델링하여, 알려지지 않은 파라미터는 데이터로부터 추정할 수 있다. 선형 회귀는 일차함수의 공식을 가진다. 선형 회귀 분석의 오차를 최소화하는 방법은 최소제곱법이다. 최소제곱법은 기울기와 y 절편을 알아 내 기존 데이터를 기반으로 최선의 선을 그려내는 방법으로, 어떤 계의 해방 정식을 근사적으로 구하며, 근사적으로 구하려는 해와 실제 해의 오차의 제곱의 합이 최소가 되는 해를 구하는 방법이다.

2. 가변수

가변수(Dummy variable)란 독립변수를 0과 1로 변환한 변수를 의미한다. 일반적인 경우 그 사실 여부에 대해 예/아니오로 확인 가능한 질적 변수(예: 남자인가? 대학원을 졸업했는가?)는 회귀 분석에 직접 투입하는 것이 불가능하다. 이러한 질적 변수를 회귀분석에 사용하기 위해 그 가부를 0 혹은 1의 숫자 형태로 대응시킨 변수를 가변수라 한다.

Ⅲ. 연구 방법

1. 연구내용 및 방법

1) 설문지 내용

- 1) 카페 만족도(1~10 사이의 수, 소수점 가능, 예시 : 4.2, 8.9) 방문하지 않았다면 0으로 작성해주세요.
- 2) 한옥박물관 만족도(1~10 사이의 수, 소수점 가능, 예시 : 4.2, 8.9) 방문하지 않았다면 0으로 작성해주세요.
- 3) 한옥 구경 만족도(1~10 사이의 수, 소수점 가능, 예시 : 4.2, 8.9) 방문하지 않았다면 0으로 작성해주세요.
- 4) 등산 만족도(1~10 사이의 수, 소수점 가능, 예시 : 4.2, 8.9) 방문하지 않았다면 0으로 작성해주세요.
- 5) 나이대(10대면 10, 50대 이상은 50으로 입력해주세요)
- 6) 일행 인원수(4명 이상이면 4)

은평한옥마을을 방문한 방문객들을 대상으로 하였다. 종속변수로 설정된 항목들에 대한 장소별 만족도값을 얻어내었다. 각 연령대별과 방문 인원에 따라 결과값을 분류한 후 경향성을 분석하여 장소별 추천에 대한 데이터로 활용하였다. 성별, 10대부터 50대까지의 연령대 및 1명부터 4명 이상의 방문 인원을 설정하였고, 1~10까지의 만족도 구간에서 응답하게 하였으며 방문하지 않은 경우는 0에 응답하도록 설정하였다.

II) 설문지 조사 방법

데이터 수집 방법은 두 가지로 하였다. 소셜 네트워크 서비스(Instagram)를 이용해 설문지 링크를 보낸 후, 설문조사를 요청하여 자료를 수집하는 방법과 은평한옥마을의 지역을 세 군데로 나누어 돌아다니며 설문지를 통해 대면 설문을 요청하는 방법을 이용했다. 수집한 데이터의 표본의 크기는 약 340개의 크기를 가진다. 설문조사 문항을 통해 피설문자의 성별, 나이, 방문자 수, 거주 지역 데이터를 수집하여 피설문자에 대한 정보를 받고 종속변수로 설정된 설문 문항은 카페, 등산, 진관사 방문, 박물관 방문, 한옥 방문 만족도를 1부터 10까지의 실수 범위로 정하여 설문 문항을 구성하였다.

III) 독립변수와 종속변수의 설정

받은 독립변수의 경우 성별, 나이대, 동행 인원수로 설문을 받은 것을 확인할 수 있다. 본 연구의 설문조사에서의 독립변수는 모두 범주형 변수이다. 이에 본 연구는 성별, 나이대의 경우 가변수 개념을 활용해 마치 연속형 변수처럼 보이도록 할 것이며, 동행 인원수의 경우 범주형 변수로 볼 수 있지만, 이는 연속형 변수라고 해도 사실상 무방하므로(동행 인원수의 경우 결국 자연수의 모든 데이터를 가질 수 있다.) 설문조사에서 받은 데이터 그대로 변수를 설정하여 회귀분석을 진행했다. 본 연구의 목표인 성별과 나이대, 동행 인원수에 따라 각 장소별 만족도가 어떻게 변하는지를 확인하는 것은 별다른 문제가 없다.

	타임스탬프	sex	age	number	cafe	museum	hanok	mountain	jinkwan	...10
1	2022-09-18 17:14:59	0	10	1	7.200000	4.5	9.000000	2.300000	4.400000	NA
2	2022-09-18 17:15:48	0	10	1	7.200000	4.5	9.000000	2.300000	4.400000	NA
3	2022-10-19 20:33:37	0	10	1	8.000000	0.0	6.000000	6.000000	8.000000	NA
4	2022-10-19 20:34:40	0	10	1	7.000000	0.0	5.000000	6.000000	7.000000	NA
5	2022-10-19 20:34:52	0	10	1	7.500000	8.0	8.000000	6.500000	9.000000	NA
6	2022-10-19 20:35:01	0	10	1	9.540000	0.0	9.330000	10.000000	0.120000	NA
7	2022-10-19 20:35:44	0	10	1	0.000000	0.0	6.000000	0.000000	8.000000	NA
8	2022-10-19 20:35:58	0	10	1	0.000000	0.0	5.000000	0.000000	6.000000	NA
9	2022-10-19 20:36:59	0	10	1	0.000000	0.0	1.000000	0.000000	1.000000	NA

Showing 1 to 9 of 346 entries, 10 total columns

```

Console Terminal Jobs
R 4.1.1 ~ /
> sapply(지역연계_새_프레_1, class)
$timestamp
[1] "POSIXct" "POSIXt"

$sex
[1] "numeric"

$age
[1] "character"

$number
[1] "numeric"

$cafe
[1] "numeric"

$museum\r\n
[1] "numeric"

$hanok
[1] "numeric"

$mountain
[1] "numeric"

$jinkwan
[1] "numeric"
    
```

그림 1 가변수 활용을 위한 변수종류 확인

종속변수의 경우 만족도로, 1부터 10까지의 소수점까지 포함한 범위에서의 데이터를 받았다. 이에 만족도의 경우 선형회귀를 사용할 수 있는 연속형 종속변수로 간주할 수 있으며, 본 연구에서 선형회귀를 사용했다.

다음 사진은 성별, 나이대, 동행 인원 수 구분 없이 측정한 각 장소별 평균 만족도이다. 가장 아래 노란색 박스 부분이 평균 만족도 값이다.

E	F	G	H	I
커피 만족도(1~10 사이의 수 방문하지 않았다면 0으로 작성	한옥박물관 만족도(1~10 사이의 수 방문하지 않았다면 0으로 작성	한옥 구경 만족도(1~10 사이의 수 방문하지 않았다면 0으로 작성	등산 만족도(1~10 사이의 수 방문하지 않았다면 0으로 작성	진관사 만족도(1~10 사이의 수 방문하지 않았다면 0으로 작성
7.8	8	9	6.7	8.2
7.8	7	9	0	0
6.8	0	9	0	0
6.7	7	8	0	0
8.9	0	7	0	0
8.6	7	6	3	6
4.5	6	7	7	4
6.7	4	6	5	8
4	5	6	4	6
6	5	8	5	4
5	7	6	5	4
5	4	7	3	4
5	4	7	4	5
6.7	4	6	5	8
4	5	6	4	6
9	6	7.8	0	6
7.8	6.5	7.8	4.5	4.6
4	6	6	7	3
6.7	4	6	5	8
4	5	6	4	6
7.197815628	5.745771144	6.849262627	6.274950918	6.347720358

그림 2 장소별 평균 만족도 결과값

IV) 데이터 분석 방법

본 연구의 데이터 분석 방법은 위의 이론적 배경에서 언급한 선형회귀를 사용한다. 본 연구에서 선형회귀를 사용한 이유는, 선형회귀의 경우 종속변수인 만족도 데이터가 연속형 데이터이므로 일반적으로 선형회귀를 실시할 수 있기 때문이다.

V) 사용한 R 코드 및 R library

그래프를 위해 ggplot2 라이브러리를 사용하였다.

다음은 선형회귀에서 사용한 R 코드이다.

```
#begin_code
sa<- read.csv("GLim_v3.csv&")
head(sa)
attach(sa)
sa= sa[sa[,5]!=0,]
pairs(sa[,2:5])
sex=as.factor(sex)
age=as.factor(age)
summary(sex)
summary(age)
par(mfrow=c(2,2))
plot(sex,cafe,xlab="sex",ylab="cafe")
plot(age,cafe,xlab="age",ylab="cafe")
plot(as.factor(number),cafe,xlab="number",ylab="cafe")
```

```

lm.out1=lm(cafes~sex+age+number)
summary(lm.out1)
lm.out2=lm(cafes~sex+age+number+sex:age+sex:number+age:number)
summary(lm.out2)
#-----
sa<- read.csv("GLim_v3.csv&")
head(sa)
attach(sa)
sa= sa[sa[,6]!=0,]
pairs(sa[,2:5])
sex=as.factor(sex)
age=as.factor(age)
summary(sex)
summary(age)
par(mfrow=c(1,1))
plot(sex,museum,xlab="sex",ylab="museum")
plot(age,museum,xlab="age",ylab="museum")
plot(as.factor(number),museum,xlab="number",ylab="museum")
lm.out1=lm(museum~sex+age+number)
summary(lm.out1)
lm.out2=lm(museum~sex+age+number+sex:age+sex:number+age:number)
summary(lm.out2)
#-----
sa<- read.csv("GLim_v3.csv&")
head(sa)
attach(sa)
sa= sa[sa[,7]!=0,]
pairs(sa[,2:5])
sex=as.factor(sex)
age=as.factor(age)
summary(sex)
summary(age)
par(mfrow=c(1,1))
plot(sex,hanok,xlab="sex",ylab="hanok")
plot(age,hanok,xlab="age",ylab="hanok")
plot(as.factor(number),hanok,xlab="number",ylab="hanok")

lm.out1=lm(hanok~sex+age+number)
summary(lm.out1)
lm.out2=lm(hanok~sex+age+number+sex:age+sex:number+age:number)
summary(lm.out2)

#-----
sa<- read.csv("GLim_v3.csv&")
head(sa)
attach(sa)
sa= sa[sa[,8]!=0,]
pairs(sa[,2:5])
sex=as.factor(sex)
age=as.factor(age)
summary(sex)
summary(age)
par(mfrow=c(1,1))

```

```

plot(sex,mountain,xlab="sex",ylab="mountain")
plot(age,mountain,xlab="age",ylab="mountain")
plot(as.factor(number),mountain,xlab="number",ylab="mountain")
lm.out1=lm(mountain~sex+age+number)
summary(lm.out1)
lm.out2=lm(mountain~sex+age+number+sex:age+sex:number+age:number)
summary(lm.out2)
#-----
sa<- read.csv("GLim_v3.csv&")
head(sa)
attach(sa)
sa= sa[sa[,9]!=0,]
pairs(sa[,2:5])
sex=as.factor(sex)
age=as.factor(age)
summary(sex)
summary(age)
par(mfrow=c(1,1))
plot(sex,jinkwan,xlab="sex",ylab="jinkwan")
plot(age,jinkwan,xlab="age",ylab="jinkwan")
plot(as.factor(number),jinkwan,xlab="number",ylab="jinkwan")
lm.out1=lm(jinkwan~sex+age+number)
summary(lm.out1)
lm.out2=lm(jinkwan~sex+age+number+sex:age+sex:number+age:number)
summary(lm.out2)
#end

```


IV. 결과 및 고찰

1. 설문 결과

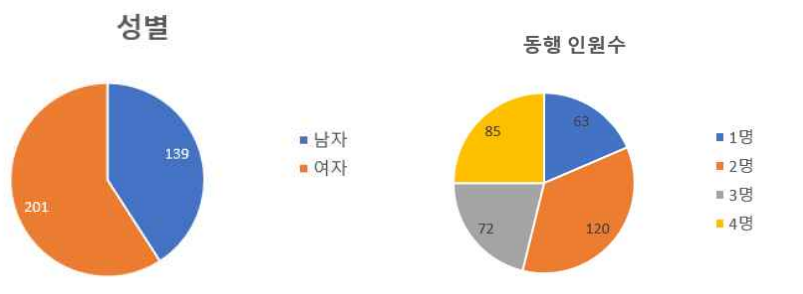


그림 3 성별 비율

그림 4 동행 인원수 비율

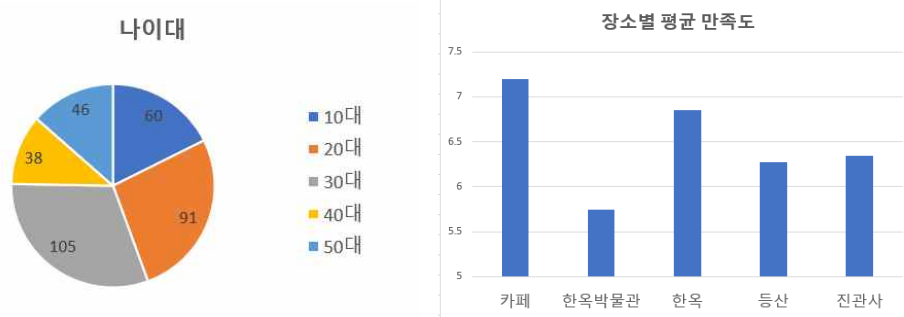


그림 5 나이대 비율

그림 6 장소별 평균 만족도값

위는 설문지 수합 이후 자료로, 남녀 응답 비율과 동행 인원수에 따른 응답 비율과 장소별 평균 만족도 및 나이대별 응답 인원수 비율을 나타낸다.

```
import pandas as pd
import tkinter.filedialog as fopener

fileName = fopener.askopenfilename()
data = pd.read_excel(fileName).T.to_dict()

def parseKey(_str):
    a, b = _str[0], _str[2]
    if a == "타":
        return "타임스탬프"
    elif a == "성":
        return "성별"
    elif a == "나":
        return "나이"
    elif a == "일":
        return "인원수"
```

```

elif a == "카":
    return "카페"
elif a == "한":
    if b == " ":
        return "한옥"
    else:
        return "박물관"
elif a == "등":
    return "등산"
elif a == "진":
    return "진관사"

```

```

data = [ { parseKey(k): v for k, v in u.items() } for u in data.values() ]
keys = { keyName: list({ x[keyName] for x in data }) for keyName in ['성별', '나이', '인원수'] }
user = { keyName: None for keyName in keys }

```

```

for keyName, oList in keys.items():
    print(f"{keyName} 을(를) 선택해주세요")
    for i, option in enumerate(oList):
        print(f"[{i+1}] {oList[i]}")

    temp = oList[int(input('\n'))-1]
    print(f"{temp} 을(를) 선택하였습니다\n")
    user[keyName] = temp

```

```

places = ['카페', '박물관', '한옥', '등산', '진관사']
stats = { k: 0 for k in places }
cnt = 0

```

```

for u in data:
    _match = 0
    for keyName, key in user.items():
        if u[keyName] == key:
            _match += 1
    if _match == 3:
        cnt += 1
        for k in places:
            stats[k] += u[k]

```

```

print(f"총 {cnt} 개의 응답이 있습니다")

```

```

if cnt != 0:
    res = { k: v / cnt for k, v in stats.items() }
    for k, v in res.items():
        print(f"{k} 에 대한 만족도: {v}")

input("Press any key to continue")

```

위 코드는 지금까지 받은 구글 설문조사 결과를 정리한 엑셀에서 데이터를 읽어와 사용자가 자신의 성별, 나이, 동행 인원수를 입력했을 때(성별 2가지, 나이대 5가지, 동행 인원수 4가지 경우로 총 40가지의 경우 설정) 사용자의 케이스와 같은 케이스의 장소별 만족도를 소수점 둘째 자리에서 반올림하여 보여준다.

```

Python 3.10.5 (v3.10.5:f377153967, Jun 6 2022, 12:36:10) [Clang 13.0.0 (clang-1300.0.29.30)] on darwin
Type "help", "copyright", "credits" or "license()" for more information.

===== RESTART: /Users/sangwoo/Downloads/main.py =====
성별 을(를) 선택해주세요
[1] 남자
[2] NaT
[3] 여자

3
여자 을(를) 선택하였습니다
나이 을(를) 선택해주세요
[1] NaT
[2] 40.0
[3] 10.0
[4] 50.0
[5] 20.0
[6] 30.0

5
20.0 을(를) 선택하였습니다
인원수 을(를) 선택해주세요
[1] NaT
[2] 1.0
[3] 2.0
[4] 3.0
[5] 4.0

5
4.0 을(를) 선택하였습니다

총 14 개의 응답이 있습니다
카페 에 대한 만족도:7.5
박물관 에 대한 만족도:4.0
한옥 에 대한 만족도:6.7
등산 에 대한 만족도:3.4
진관사 에 대한 만족도:3.8
Press any key to continue|

```

그림 7 위 코드 실행 결과

2. 분석 결과

1) 카페

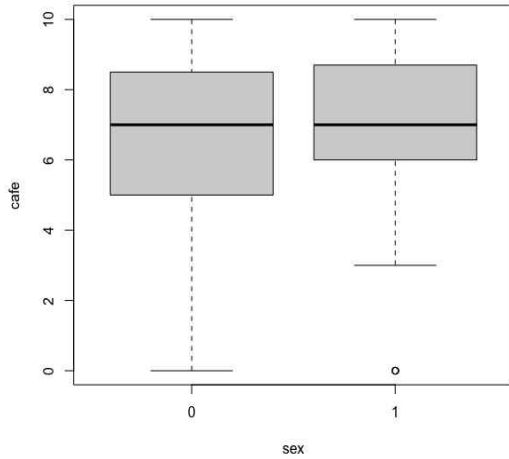


그림 8 성별별 카페 만족도 boxplot

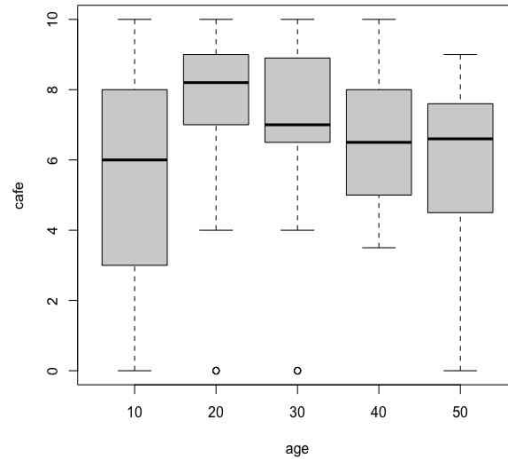


그림 9 나이대별 카페 만족도 boxplot

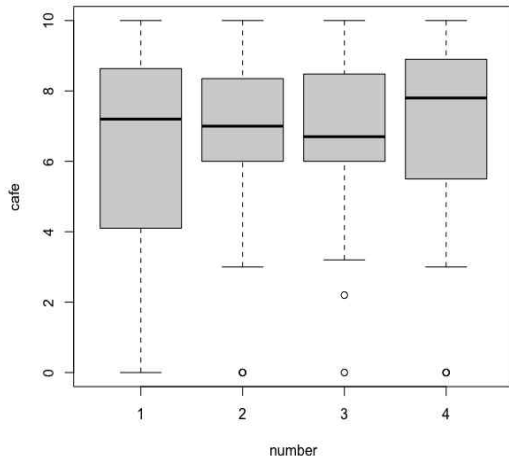


그림 10 방문 인원수별 만족도 boxplot

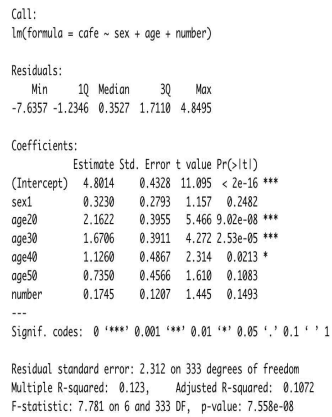


그림 11 카페 만족도 선형회귀 결과

위 사진의 선형회귀의 summary를 보면 우선 기준 대상이 10대이면서 남자인 것을 확인할 수 있다. 이 기준으로부터 만족도의 차이가 다른 집단의 경우 어느 정도인지를 위 그림에서 확인할 수 있다. 위 결과를 해석하면 성별에 따른 차이가 유의하게 존재하지는 않으며, 10대와 20대, 30대, 40대와 50대의 차이는 유의하게 존재하는 것을 확인할 수 있다. 또한 50대와 10대의 차이, 동행 인원수에 따른 만족도 값의 차이가 존재하지만 유의하지는 않은 것을 알 수 있다.

10대와 20대와의 만족도의 차이는 2.1622 만점으로 드러났으며(소숫점 다섯 째 자리에서 반올림한 결과임.) 이는 20대의 만족도가 평균적으로 2.1622만큼 더 높다는 의미로 받아들일 수 있다. 따라서 20대의 카페 만족도가 10대의 카페 만족도보다 높음을 알 수 있다.

10대와 30대의 만족도의 차이는 1.6706만큼의 차이를 보여 통계적으로 유의한 차이를 드러내고, 비교적 30대의 카페 만족도가 10대의 카페 만족도보다 높음을 확인할 수 있다.

위 그래프 자료는 box plot으로 위아래 회색 영역이 전체 데이터의 50%가 들어가 있으며, 75%, 25%의 자료를 이 범위 내에 포함한다. 정가운데의 진한 검정색 선은 중위수(median)이며, 이 범위 위 아래로 전체 데이터의 개수의 25%씩을 포함한다.

위 해석 방법을 토대로 해석한다면, age-cafe box plot에서부터 알 수 있듯이 비교적 10대에 비해 20, 30대가 카페 만족도가 높은 것으로 드러났으며, 40, 50대 이상은 10대보다 조금 높지만 유의한 차이는 아닌 것을 볼 수 있다. 또한 동행 인원수 및 성별에는 큰 영향을 받지 않는 것으로 나타남을 확인할 수 있다.

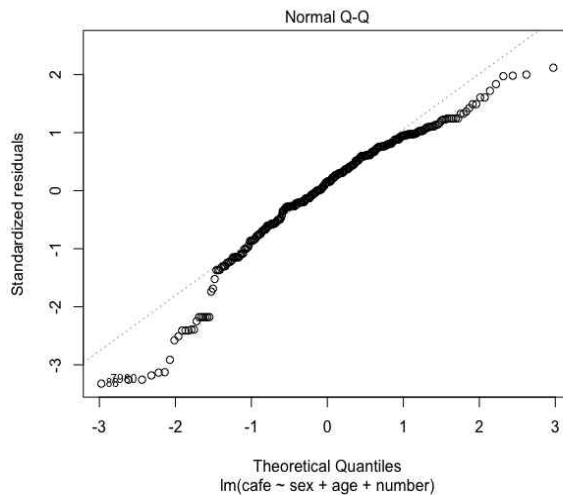


그림 12 선형회귀 정규성 확인

위 그림을 통해 선형회귀의 한 가지 가정인 정규성을 확인할 수 있다. residual이 정규 분포를 따르는 standard residual이 우상향하는 직선상에 모여있음을 통해 볼 수 있다.

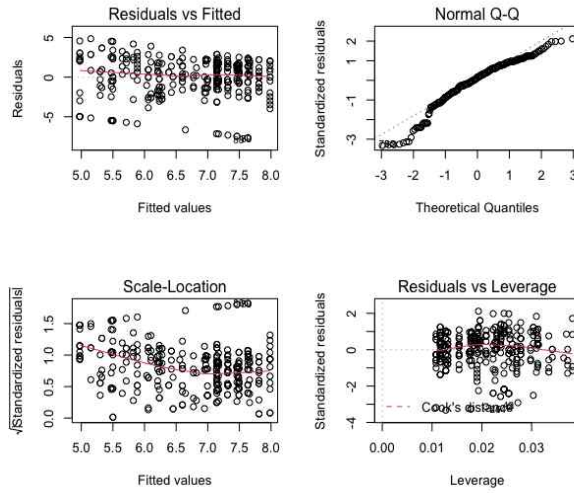


그림 13 선형회귀 정규성 확인

위의 분석들을 확인해보자. 좌측 두 개의 그래프는 residual(잔차)들이 등분산성을 띠고 있는지 확인하기 위한 그래프이다. 왼쪽 두 그래프에서 특별한 경향성이 나타나지 않는 것으로 보아 둘은 등분산성 가정을 만족함을 확인할 수 있다.

ii) [✓]한옥박물관

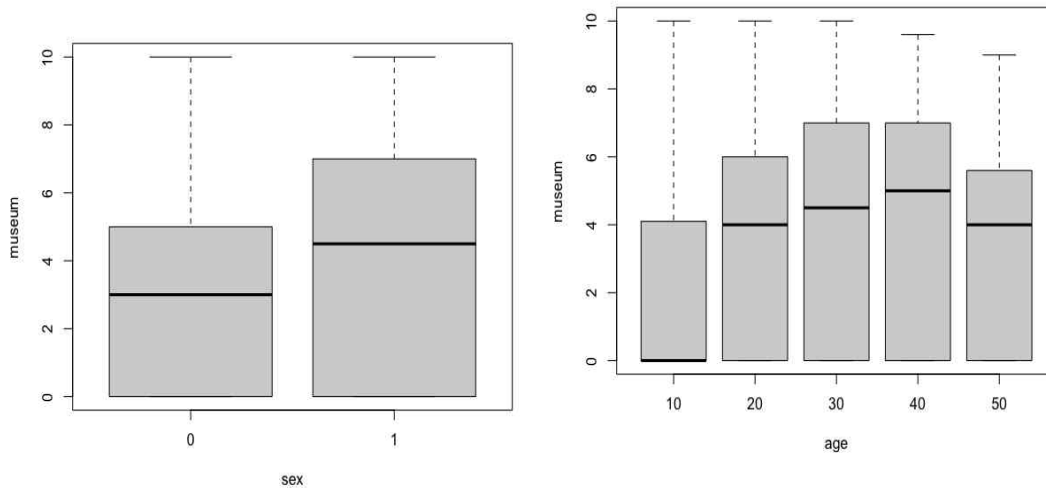


그림 14 한옥박물관 성별별 만족도 boxplot

그림 15 한옥박물관 나이대별 만족도

boxplot

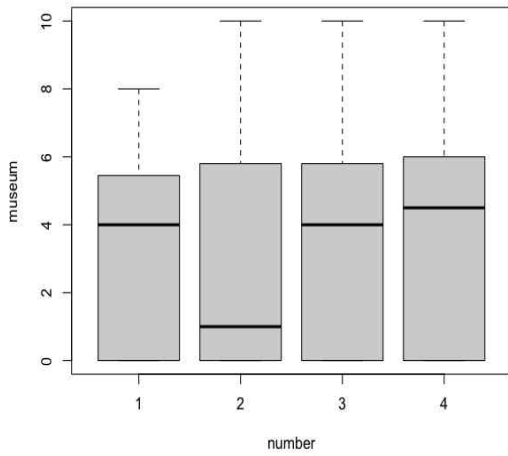


그림 16 한옥박물관 방문자수 별 만족도 boxplot

```
Call:
lm(formula = museum ~ sex + age + number)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7111 -2.6917  0.2596  2.2907  8.1432

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5100     0.5736   0.889 0.374603
sex1         0.9707     0.3702   2.622 0.009135 **
age20        1.4345     0.5242   2.736 0.006548 **
age30        1.8836     0.5183   3.634 0.000323 ***
age40        2.2132     0.6450   3.431 0.000677 ***
age50        1.8425     0.6051   3.045 0.002513 **
number       0.3367     0.1600   2.104 0.036122 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.064 on 333 degrees of freedom
Multiple R-squared:  0.09749, Adjusted R-squared:  0.08123
F-statistic: 5.995 on 6 and 333 DF, p-value: 5.728e-06
```

그림 17 한옥박물관 선형회귀 결과

위 linear model 분석을 통해 도출할 수 있는 결론은 다음과 같다. 역시 기준은 남자, 10대가 되었으며 성별, 나이대, 동행 인원 수에 따라 모두 통계적으로 유의한 차이가 있음을 확인할 수 있다. 성별에 따른 차이는 여자의 경우 남자보다 약 0.9707 만큼의 한옥 박물관 만족도가 높음을 확인할 수 있으며, 20대, 30대, 40대, 50대 모두 10대보다 각각 1.4345, 1.8836, 2.2132, 1.8425 만큼의 한옥박물관 만족도가 높음을 확인할 수 있다. 또한 동행 인원수에 따라 만족도가 증가함도 유의한 차이를 확인할 수 있었는데, 동행 인원수가 1명 증가할 때마다 만족도는 0.3367씩 증가하며 이는 한옥박물관의 경우 동행 인원수가 많아질수록 만족도가 증가한다는 결론을 얻을 수 있다.

아래 그림을 통해 정규성을 확인할 수 있다. residual이 정규분포를 따르는 standard residual이 우상향하는 직선상에 모여있음을 통해 확인하고, 이는 박물관 만족도의 선형 회귀가 정당함을 확인시킨다.

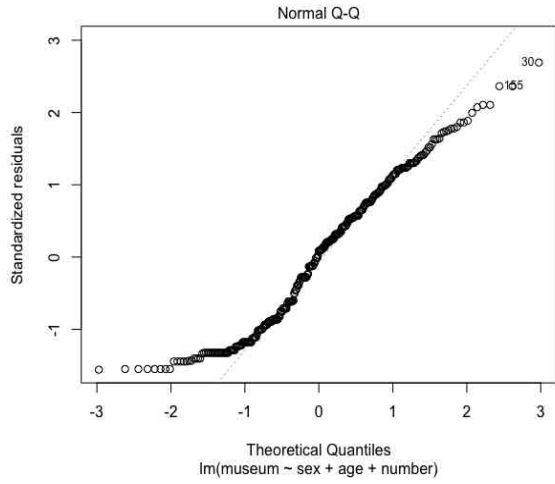


그림 18 선형회귀 정규성 확인

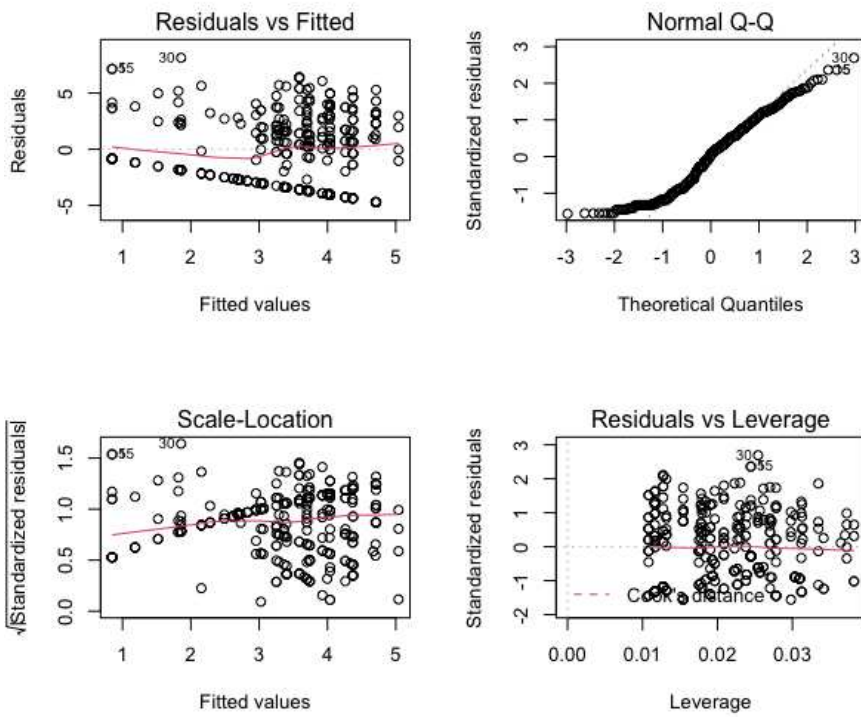


그림 19 선형회귀 가정 확인

역시 왼쪽 두 그래프를 통해 등분산성을 만족함을 확인할 수 있다.

iii) 한옥 구경

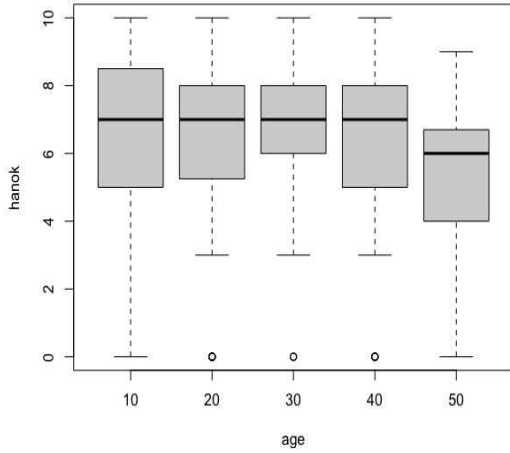


그림 20 한옥 구경 나이대별 만족도 boxplot

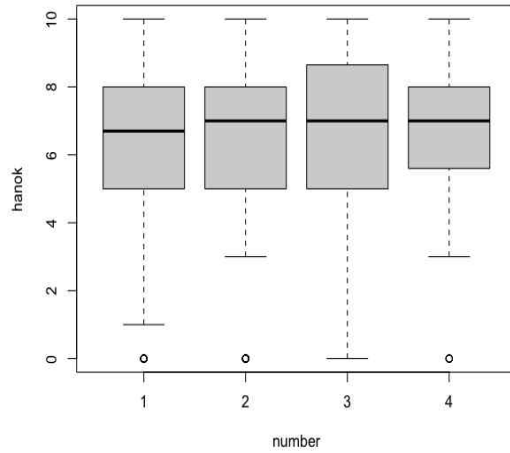


그림 21 한옥박물관 방문자수별 만족도 boxplot

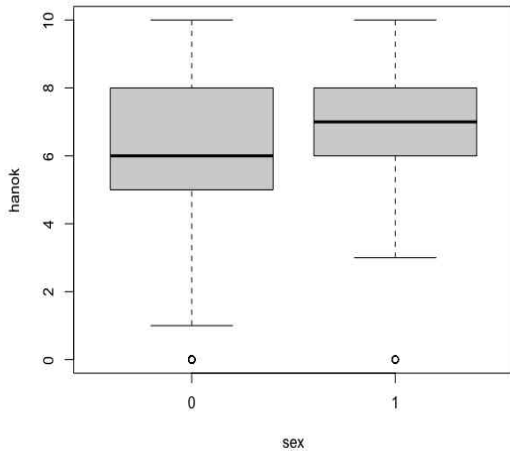


그림 22 한옥박물관 성별별 만족도 boxplot

```
Call:
lm(formula = hanok ~ sex + age + number)

Residuals:
    Min       1Q   Median       3Q      Max
-7.060 -1.153  0.323  1.626  4.295

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.30726   0.43821  14.393 < 2e-16 ***
sex1         0.68635   0.28279   2.427 0.015753 *
age20       -0.57703   0.40051  -1.441 0.150606
age30       -0.21087   0.39598  -0.533 0.594725
age40       -0.69435   0.49277  -1.409 0.159749
age50       -1.64944   0.46230  -3.568 0.000413 ***
number       0.09243   0.12226   0.756 0.450167
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.341 on 333 degrees of freedom
Multiple R-squared:  0.07531, Adjusted R-squared:  0.05865
F-statistic: 4.52 on 6 and 333 DF, p-value: 0.0002018
```

그림 23 한옥박물관 선형회귀 결과

위 자료는 한옥 구경에 따른 만족도를 선형회귀모델링 한 자료이다. 성별, 50대 정도가 유의한 차이를 확인할 수 있을 뿐, 대체적으로 만족도가 비슷함을 확인할 수 있다.

비교적 여성이 남성보다 0.6864 만큼 높은 만족도를 보임을 확인할 수 있었으며, 50대의 경우 10대에 비해 1.65 정도 낮은 만족도를 보여 연령대가 높을 경우 한옥 구경에 대한 만족도는 낮아지는 것으로 드러났다. 20대를 제외한 30대, 40대, 50대 자료를 모두 살펴볼 경우 나이대가 증가함에 따라 만족도가 더욱 많이 감소함을 확인할 수 있으며, 이는 위의 age-hanok box plot에서도 확인할 수 있다. 다만 20대의 경우 만족도가 30대에 비해 만족도가 크게 낮아지는 경향을 보이는데, 이는 20대의 경우 카페 만족도가 30대에 비해 더 높아 비교적 한옥구경의 만족도를 낮게 준 영향 탓으로 보여진다.

아래 그림을 통해 정규성을 확인할 수 있다. residual이 정규분포를 따름은 standard residual이 우상향하는 직선상에 모여있음을 통해 확인하고, 이는 한옥 구경 만족도의 선형회귀가 정당함을 확인시킨다.

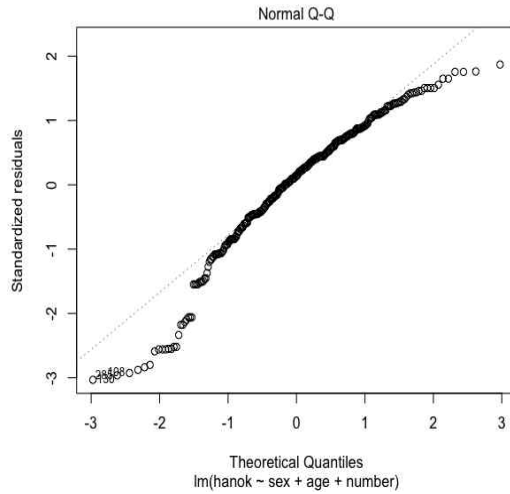


그림 24 선형회귀 정규성 확인

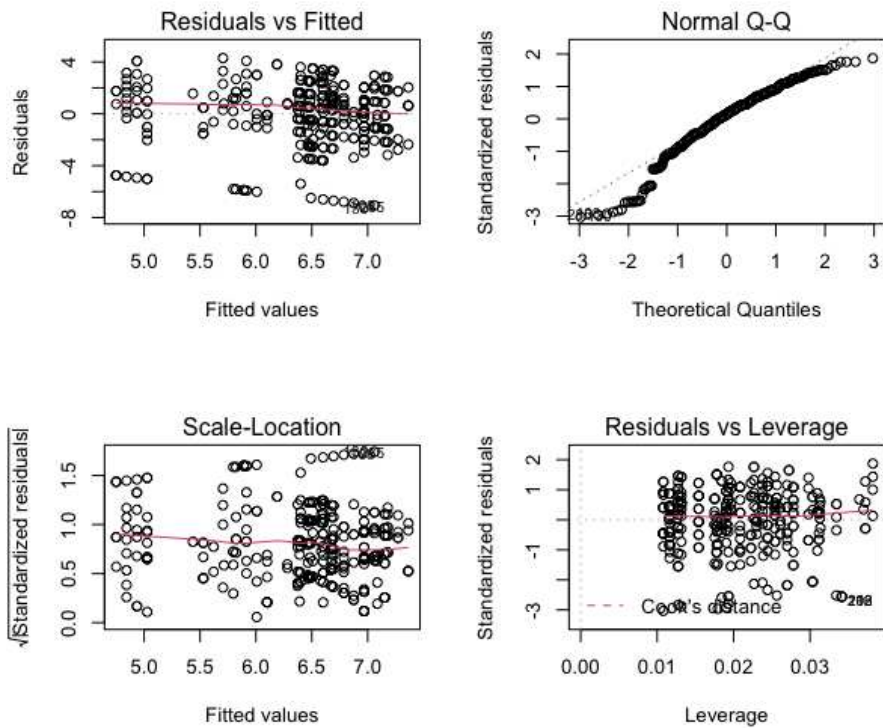


그림 25 선형회귀 가정 확인

역시 특별한 경향성이 보이지 않아 등분산성을 만족한다.

Iv) 등산

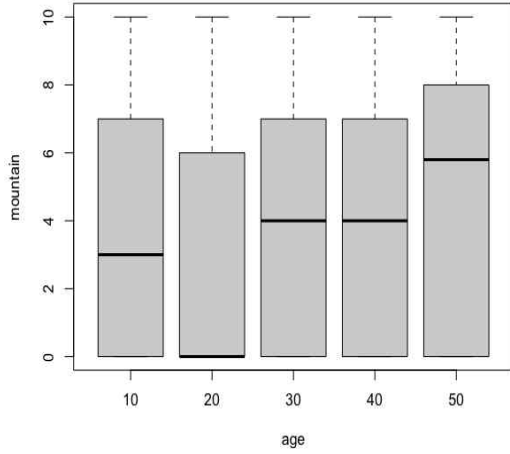


그림 26 등산 나이대별 만족도 boxplot

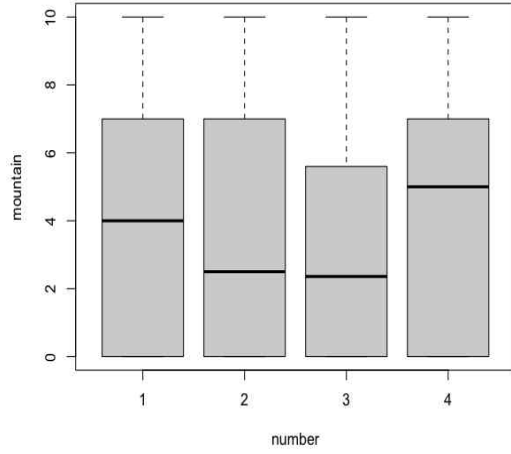


그림 27 등산 방문자수별 만족도 boxplot

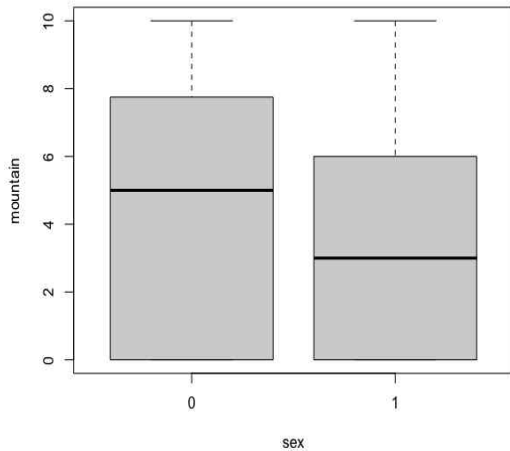


그림 28 등산 성별별 만족도 boxplot

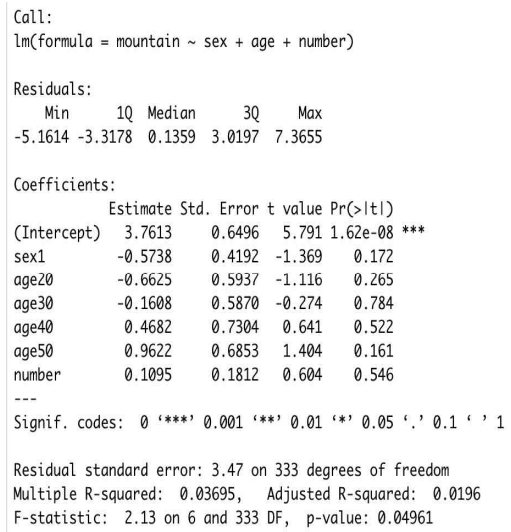


그림 29 등산 선형회귀 결과

위 자료는 등산 만족도에 대한 선형회귀모델링 자료이다. 위 자료에서는 통계적으로 유의한 차이를 가지는 집단을 확인하지 못했으며, 통계적으로 유의하지는 않지만 만족도를 비교했을 경우 20, 30대는 만족도가 10대에 비해 낮은 반면 40, 50대의 경우 10대에 비해 만족도가 높아짐을 확인할 수 있다. 이는 중장년층의 등산 선호도에 따른 영향으로 보인다.

성별에 따라 여자가 등산 만족도가 남자에 비해 0.5738 정도 낮은 것으로 확인되었지만, 이는 통계적으로 유의하지 않았으며 낮은 까닭은 비교적으로 남성의 등산 선호도가 높은 경향을 따름을 확인할 수 있다.

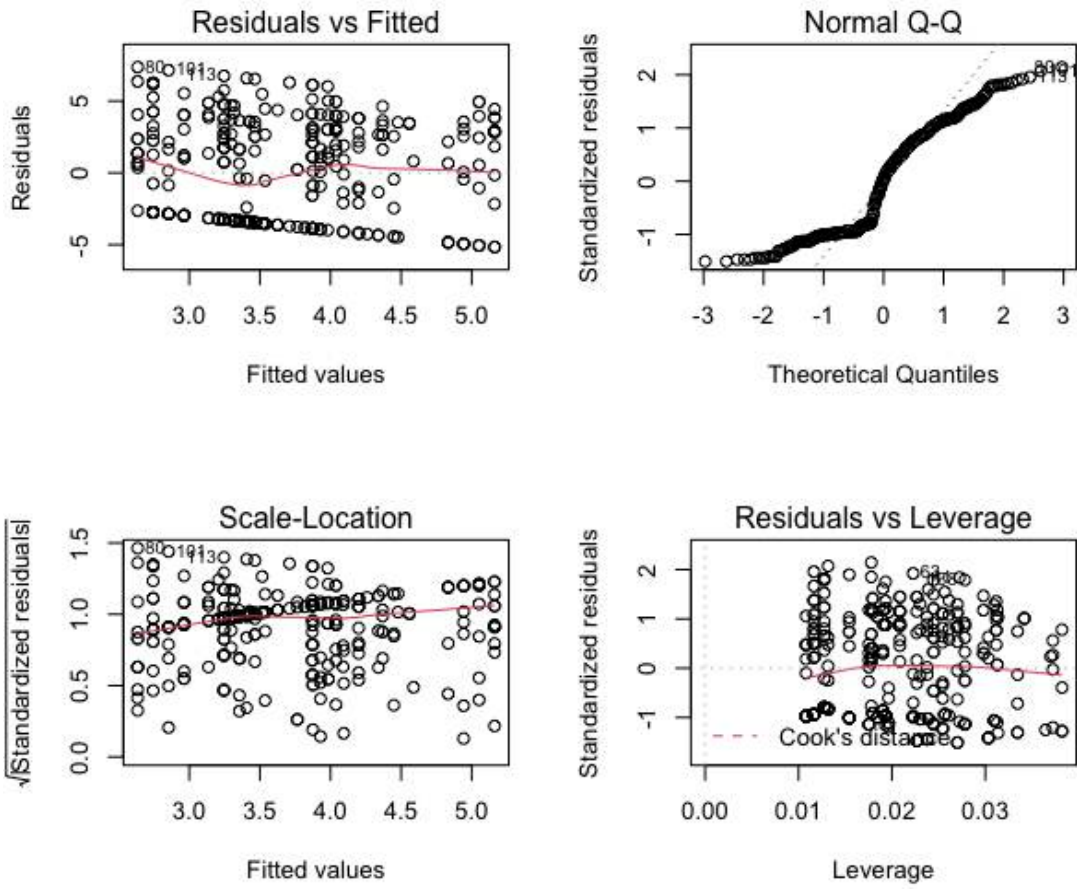


그림 30 선형회귀 가정 확인

위 자료로부터 잔차의 등분산성을 확인한다.

v) 진관사 방문

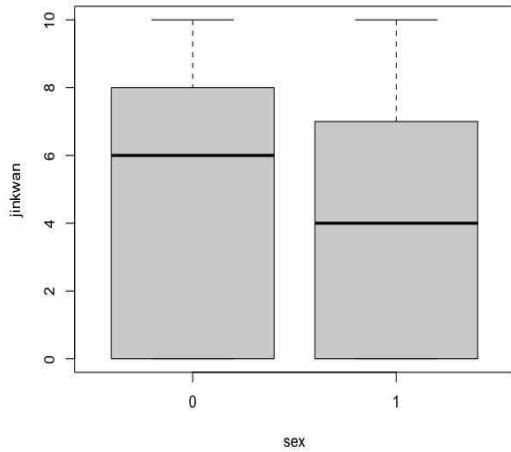


그림 31 진관사 방문 성별별 만족도 boxplot

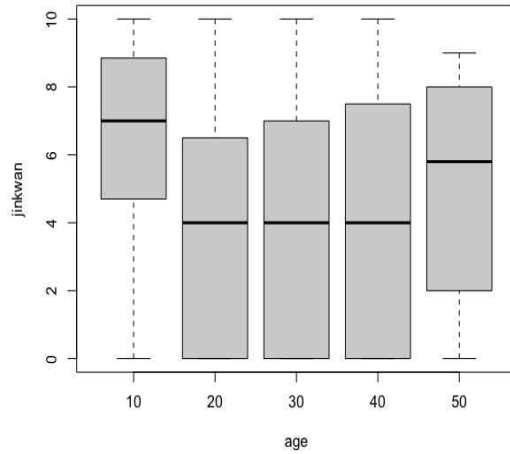


그림 32 진관사 방문 나이대별 만족도 boxplot

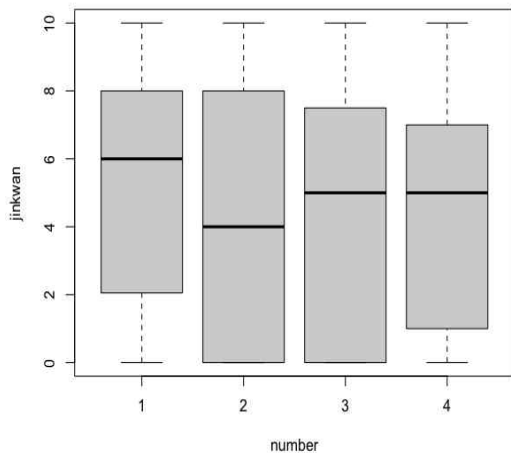


그림 33 진관사 방문 방문자수별 만족도 boxplot

```
Call:
lm(formula = jinkwan ~ sex + age + number)

Residuals:
    Min       1Q   Median       3Q      Max
-6.4694 -3.5642  0.4188  3.0672  6.5070

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.52204    0.64196  10.159 < 2e-16 ***
sex1         -0.07117    0.41428  -0.172  0.86370
age20        -2.79979    0.58674  -4.772  2.74e-06 ***
age30        -2.55188    0.58010  -4.399  1.47e-05 ***
age40        -2.09898    0.72189  -2.908  0.00389 **
age50        -1.52335    0.67725  -2.249  0.02514 *
number       -0.05269    0.17911  -0.294  0.76881
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.43 on 333 degrees of freedom
Multiple R-squared:  0.08248, Adjusted R-squared:  0.06594
F-statistic: 4.989 on 6 and 333 DF, p-value: 6.54e-05
```

그림 34 진관사 방문 선형회귀 결과

위의 자료는 진관사 만족도를 선형회귀모델링한 자료이다.

위 자료에서 유의한 차이는 20대, 30대, 40대, 50대에서 확인할 수 있으며 20대, 30대, 40대, 50대의 경우 진관사의 만족도가 모두 10대보다 낮은 것으로 드러났다. 각각의 만족도가 2.7998, 2.5519, 2.0990, 1.5230만큼 낮은 것이 확인되었다. 이는 실제 예상치 못했던 결과이다. 보편적으로 나이가 높아질수록 만족도가 높아진다고 가설을 설정했지만, 이 가설에 부합하지 않으며 가설을 기각하고 통계적으로 유의한 결과인 20, 30, 40, 50대 모두 10대에 비해 만족도가 낮다는 결과를 채택하겠다.

아래 그림을 통해 정규성을 확인할 수 있다. residual이 정규분포를 따르는 standard residual이 우상향하는 직선상에 모여있음을 통해 확인하고, 이는 진관사 방문 만족도의 선형회귀가 정당함을 확인시킨다.

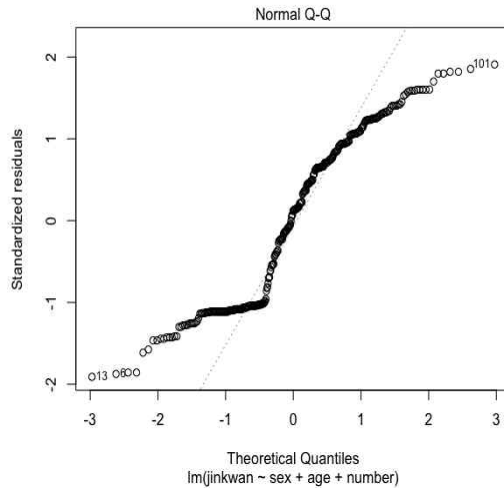


그림 35 진관사 방문 선형회귀 정규성 확인

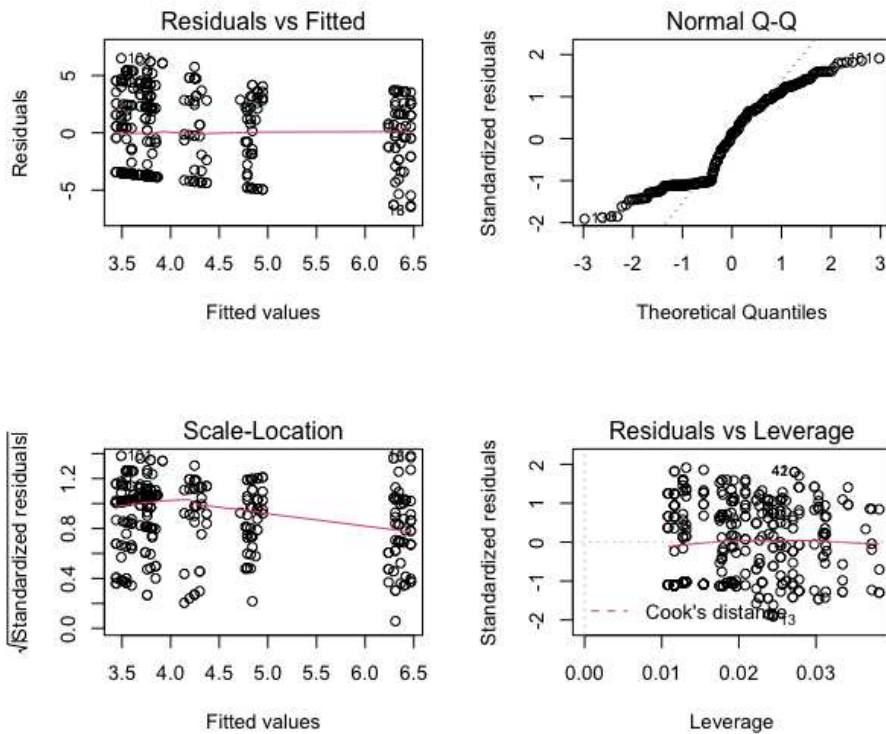


그림 36 진관사 방문 선형회귀 가정 확인

위 자료에서 왼쪽 두 그래프에서부터 잔차의 등분산성을 확인한다.

3. 사용자로의 적용

다음 자료는 실제 선형회귀모델링을 바탕으로 사용자가 자신의 정보를 입력했을 때 출력해주는 만족도 python 코드이다.

```
x1 = int(input('남자라면 0, 여자라면 1을 입력해주세요 : '))
k = int(input('\n나이대를 입력해주세요(50대 이상이면 50으로 입력해주세요) :'))
x7 = int(input('\n동행인원수를 입력해주세요(4인 이상이라면 4라고 입력해주세요) : '))

def SA(x1, x2, x3, x4, x5, x6, x7):
    return 4.8014 + 0.3230 * x1 + 2.1622 * x3 + 1.6706 * x4 + 1.1260 * x5 + 0.7350
    * x6 + 0.1745 * x7

if k == 10:

    print(SA(x1,1,0,0,0,0,x7))

elif k == 20:

    print(SA(x1,0,1,0,0,0,x7))

elif k == 30:

    print(SA(x1,0,0,1,0,0,x7))

elif k == 40:

    print(SA(x1,0,0,0,1,0,x7))

else:

    print(SA(x1,0,0,0,0,1,x7))
```

위 파이썬 코드를 실행 시 다음과 같은 작업이 수행된다.

```

Python 3.10.6 (v3.10.6:9c7b4bd164, Aug 1 2022, 17:13:48) [Clang 13.0.0 (clang-1300.0.29.30)] on darwin
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: /Users/matthewgunlim/Downloads/main2.py =====
남자라면 0, 여자라면 1을 입력해주세요 : 1

나이대를 입력해주세요(50대 이상이면 50으로 입력해주세요) :20

동행인원수를 입력해주세요(4인 이상이라면 4라고 입력해주세요) : 3
7.8101
>>>

```

그림 37 파이썬 코드 실행 시 수행되는 결과

실제 자신의 정보를 입력하면 각 장소별 방문했을 때의 예상되는 만족도 값을 선형회귀 모델에 근거하여 출력한다. 위의 python 코드는 카페 만족도 하나만의 만족도 값이다. 나머지 장소에 대한 만족도 값이 필요한 경우 위 각 장소별 결과 부분의 linear modeling 을 돌렸을 때의 coefficient를 다시 가져와 선형회귀모델을 다시 적용하면 된다.

V. 결론 및 제언

설문 결과를 바탕으로 장소별 만족도를 구하고, 방문객의 특성별로 평균 만족도를 분석했다. 설문 결과가 계속 누적될수록 더 많은 설문 결과가 생길 것이고, 그에 따라 최신 트렌드가 반영된, 더욱 정확한 만족도 및 선호도를 파악할 수 있다. 또한, 많은 설문 결과를 바탕으로 계절, 기후 등 더욱 다양한 척도를 기준으로 결과를 분석할 수 있을 것이며 이를 토대로 더욱 정교한 추천 알고리즘을 구축할 수 있을 것이다. 설문 결과를 계속 늘려나가 이용자가 더욱 만족할 수 있도록 도울 뿐만 아니라 관광지로서의 은평한옥마을 개발에 많은 도움이 될 것이라고 생각한다.

VI. 참고 문헌

<https://drive.google.com/drive/folders/0AOr57MERwE8HUK9PVA>

위 파일은 본 연구에서 사용한 데이터 자료이다.

가변수 . (n.d.).

<https://ko.wikipedia.org/wiki/%EA%B0%80%EB%B3%80%EC%88%98>.

선형 회귀 . (n.d.). https://ko.wikipedia.org/wiki/선형_회귀.

조인석, 김경호. (n.d.). *IPA기법을 활용한 관광객 재방문 의사에 관한 연구- 전주 한옥마을을 중심으로*. n.p.: 한국경영교육학회.

김연선. (2019). 전주 한옥마을 관광객의 관광동기가 만족도와 재방문 의도에 미치는 영향 연구. 한국콘텐츠학회논문지, 19(1), 196-204.