

R언어를 통한 ANOVA 기반 학년별 음식점 만족도 차이

양윤태**, 김재경*, 이주협*, 손상길*, 윤진원*, 심재창*

**안동대학교 창의융합학부(인문)

*안동대학교 컴퓨터공학과

e-mail : jcshim@andong.ac.kr, ingkykor@naver.com

ANOVA-Based Grade Satisfaction Difference Test in R

Yoon Tae Yang**, Jaegyeong Kim*, Juhyeop Lee*, Sanggil Son*, Jinwon Yoon*,
Jaechang Shim*

*Dept of Computer Engineering, Andong National University

**Dept of Creative Fusion Andong National University

요 약

본 논문에서는 R언어를 이용한 국립 안동대학교 주변 음식점에 대해 학년별 만족도 차이를 보이는지 ANOVA를 기반으로 만족도 차이의 유무를 조사한다. ANOVA와 Duncan LSR Test에 대하여 정의를 기술하였으며 학년별 만족도 차이를 알아보기 위해 ANOVA를 활용하여 학년별에 따라 만족도 차이를 보이는지를 확인하였고 사후검증으로 Duncan LSR Test를 활용하여 만족도 차이가 어느 그룹 간에서 나타나는지를 다중 비교하였다.

1. 서론

디자인 평킹 기법을 활용하여 대학 새내기 시절과 복학생 시절, 어떤 음식점을 선택할지 어영부영하던 시절을 생각하면서 앞으로 국립 안동대학교에 입학, 복학하게 될 안동대 재학생들에게 안동대 주변 음식점 만족도를 바탕으로 음식점에 대한 자료를 만들어 입학생, 복학생들이 음식점을 선택하는 데 있어서 좋은 참고 자료가 될 수 있을 것 같아 이번 프로젝트 주제를 “R언어를 통한 ANOVA 기반 국립 안동대학교 주변 음식점에 대한 학년별 음식점 만족도 차이 분석”으로 선정하였다. ANOVA란 분산분석으로 불리며 명목 척도로 측정된 독립변수와 등간 또는 비율척도로 측정된 종속변수 사이와 관계를 연구하는 통계기법이다. 분산분석은 둘 이상의 집단들 간에 어떤 변수의 평균 점수에 차이가 있는지를 검정하는 것이라고 보면 된다. 분산분석은 하나의 범주형 독립변수와 종속변수 간의 관계를 분석하는 일원 분산분석과 둘 이상의 독립변수들을 함께 고려했을 때 이들이 종속변수에 미치는 효과를 분석하는 이원 분산분석으로 분류가 가능하다. 이번 프로젝트에서는 일원 분산분석을 통해서 음식점 A를 대표적으로 선정하여 범주형 독립변수와 종속변수 간의 관계를 분석하여 집단들 간에 만족도 차이가 있는지를 분석하도록 한다.

2. 분석 과정

2.1.1 국립안동대학교 학년 별 음식점 만족도 설문결과

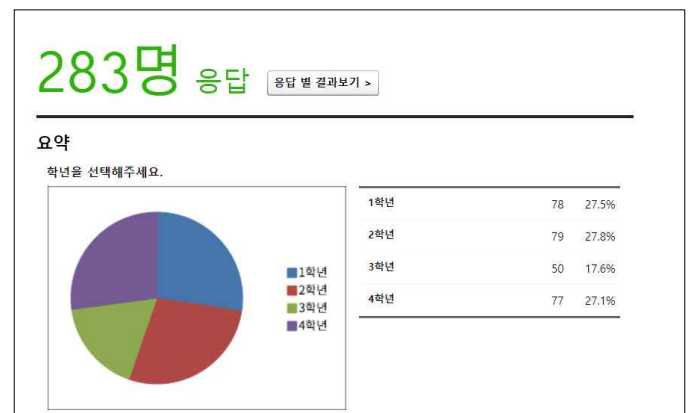


그림 1. 네이버 오피스 폼 기반의 설문조사 결과

네이버 오피스 폼의 설문조사를 활용하여 국립 안동대학교 주변 음식점에 대해 재학생 학년별 만족도를 조사하였다. 음식점은 음식점 A~R까지 총 18 지점의 음식점 만족도를 조사하였다. 그 결과 그림 1에서 설문조사의 결과를 확인할 수 있다.

2.1.2 R-Studio에 설문조사 csv파일 읽어오기

```
> setwd("c:/tmp")
> getwd()
[1] "c:/tmp"
> anusatis <- read.csv("anusatis.csv")
```

그림 2. 설문결과 csv파일을 R-Studio로 읽어오기

그림1의 결과를 csv파일로 저장하여 setwd("c:/tmp") 명

령어를 통해 R-Studio로 읽어온다. anusatis 변수에는 설문 결과 csv파일의 데이터가 할당된다. 그림 2는 외부로부터 csv파일을 불러와 anusatis.csv 데이터를 변수 anusatis에 할당하는 소스코드를 나타낸다.

2.2 tapply()를 이용한 anusatis 데이터 요약

```
> tapply(anusatis$restaurant_A, anusatis$grade, summary)
$`3rd_grade`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.0    3.0    4.0    3.8    5.0    5.0

$`4th_grade`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   4.00   4.00   4.26   5.00   5.00

$freshman
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   3.000   4.000   3.744   5.000   5.000

$sophomore
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   3.000   4.000   3.718   4.750   5.000
```

그림 3. anusatis 데이터 요약 결과

tapply() 함수를 이용하여 anusatis 데이터 중 음식점 A에 대한 학년별 만족도의 최댓값, 최솟값, 평균값을 확인할 수 있다. 그림 3은 tapply가 anusatis의 데이터를 요약해준 결과를 나타낸다.

2.3.1 ANOVA 분석

```
> aov(restaurant_A ~ grade, data = anusatis)
Call:
aov(formula = restaurant_A ~ grade, data = anusatis)

Terms:
              grade Residuals
Sum of Squares   14.90976 255.47186
Deg. of Freedom         3      279

Residual standard error: 0.9569063
Estimated effects may be unbalanced
```

그림 4. ANOVA 분석

세 집단 이상일 때 사용하는 ANOVA 분석을 활용하여 음식점 A에 대한 1, 2, 3, 4학년(범주형 독립변수) 네 집단의 만족도 차이(종속변수)에 영향이 있는지를 분석한다. 그림 4는 anusatis 데이터의 내부 요인 음식점 A에 대한 만족도와 학년 간의 관계를 ANOVA를 이용하여 분석한 것을 나타낸다.

2.3.2 ANOVA 결과 요약

```
> aovmodel <- aov(restaurant_A ~ grade, data = anusatis)
> summary(aovmodel)
              Df Sum Sq Mean Sq F value Pr(>F)
grade          3  14.91   4.970    5.428 0.00122 **
Residuals     279 255.47   0.916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

그림 5. ANOVA 결과 요약

그림 5는 ANOVA 분석한 결과를 요약해준 것을 나타낸다. 그림 5를 통해서 확인한 결과 Pr(>F)의 값이 0.00122이므로 0.05보다 작기 때문에 값이 유의미하다는 뜻이다. 따라서 음식점 A에 대한 학년별 만족도에 차이를 보인다는 것을 확인할 수 있다.

2.4 Duncan LSR Test를 이용한 다중비교

```
> install.packages("agricolae")
Error in install.packages : Updating loaded packages
> library("agricolae")
> duncan.test(aovmodel, "grade", alpha = 0.05, console = TRUE)

Study: aovmodel ~ "grade"

Duncan's new multiple range test
for restaurant_A

Mean Square Error: 0.9156698

grade, means

      restaurant_A      std r Min Max
3rd_grade    3.800000 1.0497813 50  1  5
4th_grade    4.259740 0.8491884 77  1  5
freshman     3.743590 0.9728242 78  1  5
sophomore    3.717949 0.9789663 78  1  5

Groups according to probability of means differences and alpha level( 0.05 )

Means with the same letter are not significantly different.

      restaurant_A groups
4th_grade    4.259740    a
3rd_grade    3.800000    b
freshman     3.743590    b
sophomore    3.717949    b

Restarting R session...
```

그림 6. Duncan LSR Test를 위한 Duncan.test() 함수

그림 8. 학년 간 만족도 차이

본 논문에서는 R언어를 활용한 ANOVA 기반의 국립 안동대학교 주변 음식점에 대해 안동대 재학생 학년별 만족도 차이 검증 대하여 기술하였다. 주변 음식점들 중 음식점 A를 대표적으로 설정하여 학년별 만족도 차이를 검증하였다. 그림 4를 통해 음식점 A에 대하여 학년별 만족도 차이가 있음을 확인(유의미)하였고 그림 6, 7을 통해 사후검증으로 학년 간의 만족도 차이가 어느 그룹에서 나타나는지를 비교하였다. 그 결과 그림 6, 7을 통해 음식점 A에 대하여 대학생 4학년과 1, 2, 3학년의 만족도 차이가 있음을 확인할 수 있었다. 그림 8은 4학년과 1, 2, 3학년이 음식점 A에 대해 만족도 차이를 보이는 것을 나타낸다.

4. 기대효과

“국립 안동대학교 주변 음식점에 대한 학년별 만족도 차이가 있는가?”에 대한 검증을 통해서 학년별로 선호하는 음식점에 대해 만족도아 함께 만족도 차이가 있음을 보이면서 앞으로 입학하게 될 국립 안동대학교 신입생들과 복학생들에게 음식점 선택에 있어서 혼동과 시간 낭비 없는 효율적인 음식점 선택을 도와줄 수 있다.

참고문헌

- [1] 네이버카페 알언어 <https://cafe.naver.com/rlanguage>
- [2] KSI, “분산분석 (ANOVA)”
http://6025.co.kr/bbs/board.php?bo_table=cust_in&wr_id=17
- [3] R다중비교 - Duncan's LSR(Least Significant Range) test, <https://rfriend.tistory.com/133>

그림 7. Duncan LSR Test를 위한 LDuncan() 함수

그림 4를 통해서 학년에 따른 음식점 A에 대한 만족도 차이를 보이는 것을 확인하였다. 이에 사후분석으로 “어느 학년에서 만족도 차이를 보이는가?”에 대해 다중 비교를 하였다. 다중 비교 방법으로는 Tukey, Duncan, Scheffe Test 등이 있는데 이 중 Duncan's LSR Test를 활용하여 다중 비교하였다. Dauncan's LSR Test는 평균수에 관계 없이 결정한 유의수준을 유지하면서 모든 평균 간의 비교하는 것을 말한다.(DMRT) 그림 6은 duncan.test() 함수를 이용하여 다중비교 한 것을 나타낸다. 그림 7은 LDuncan() 함수를 이용하여 다중비교 한 것을 나타낸다. 그림 6, 7을 통하여 다중비교를 해본 결과 4학년과 1, 2, 3학년 간에 만족도 차이를 보이는 것을 확인할 수 있다.

3. 결론

```
DUNCAN TEST TO COMPARE MEANS

Confidence Level: 0.95
Dependent Variable: restaurant_A
Variation Coefficient: 24.61859 %

Independent Variable: grade
Factors Means
4th_grade 4.25974025974026 a
3rd_grade 3.8 b
freshman 3.74358974358974 b
sophomore 3.71794871794872 b
```