

## <problem description>

2022년 한 해 동안 제주도를 여행한 여행객들을 계절, 상업, 내외국민에 따른 비용에 의문이 들어 anova분석을 통해 어떤 요소에서 여행객들의 소비가 상관있는지 알아보려함

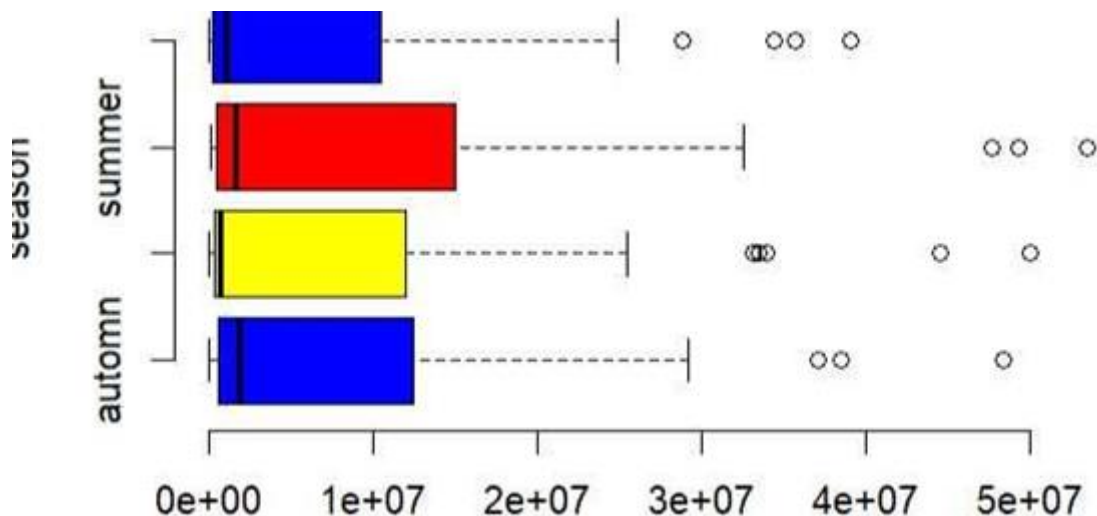
## <데이터 출처>



## # 계절별 비용 통계적 분석

	season	Mean	Max	Min	n	Sd
	<chr>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1	autumn	17390696.	50658227	410546	18	14963082.
2	spring	18074295.	51158568	421149	18	15743063.
3	summer	20780650.	54981123	527583	18	17674569.
4	winter	15208020.	39352393	428385	18	13094322.

## # 계절별 비용 시각화



두 데이터를 종합해보면 제주도는 여름시즌에 여행객들의 비용이 많이 지기는 하나 이것은 우연에 의한 것일 수 있으므로 4집단의 데이터를 아노바 분석하여 통계적으로 유의미한 차이가 있는 지 알아보겠다.

#### <ANOVA analysis>

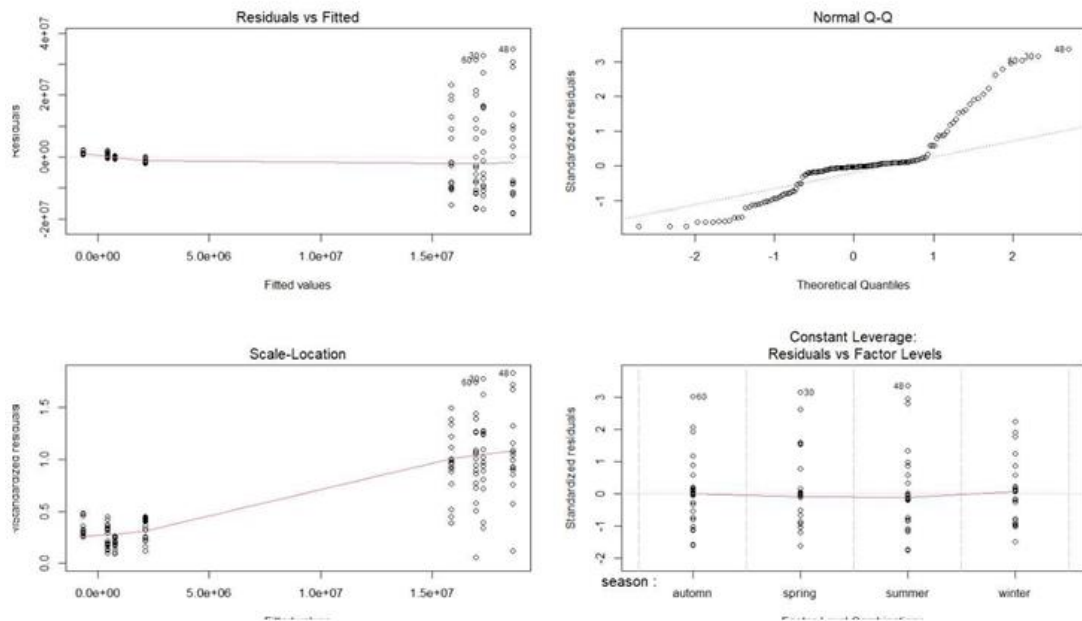
```
> #3. 계절별 여행 비용 아노바 분석
> season.aov=aov(cost~season,data = jeju)
> summary(season.aov)
              Df    Sum Sq   Mean Sq F value Pr(>F)
season          3 1.425e+14  4.749e+13   0.262  0.853
Residuals     140 2.539e+16  1.813e+14
> TukeyHSD(season.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = cost ~ season, data = jeju)
```

```
Fit: aov(formula = cost ~ season, data = jeju)

$season
      diff      lwr      upr    p adj
spring-autumn  341799.9 -7910846 8594446 0.9995510
summer-autumn  1694977.4 -6557668 9947623 0.9506272
winter-autumn -1091338.1 -9343984 7161308 0.9859669
summer-spring  1353177.5 -6899468 9605823 0.9738838
winter-spring -1433138.0 -9685784 6819508 0.9692324
winter-summer -2786315.4 -11038961 5466330 0.8163123
```

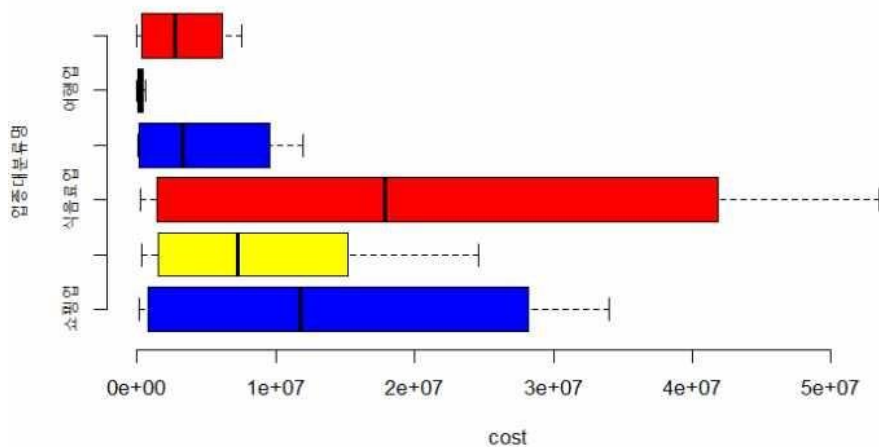
분석 결과 계절별로 여행객들의 여행비용은 4집단이 유의미한 차이를 가지고 있지 않는다. 또한 정규성이 크게 떨어져 이번에는 업종별로 여행객들이 쓰는 비용을 알아보겠다.



#### <업종별 비용 통계분석>

업종대분류명	Mean	Max	Min	n	Sd
<chr>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1 쇼핑업	14413826.	34000000	217481	24	14140104.
2 숙박업	8960343.	24600000	350200	24	8214812.
3 식음료업	21924954.	53500000	232754	24	21722359.
4 여가서비스업	4775526.	12000000	39300	24	4890200.
5 여행업	261702.	603574	25231	24	212148.
6 운송업	3253894.	7514411	34499	24	3007419.

#### <업종별 여행비용 시각화>



두 데이터를 보면 업종별로 여행객들이 차이가 크게 나타나는 것으로 추측할 수 있다. 이 차이가 통계적으로 유의미한 수치인지 anova분석을 통해 확인해 보겠다.

### <업종별 비용 anova 분석>

```
> #3.업종별 cost를 아노바 분석
> jeju.aov=aov(cost~업종대분류명,data = jeju)
> summary(jeju.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
업종대분류명	5	7.765e+15	1.553e+15	12.07	1.06e-09 ***
Residuals	138	1.776e+16	1.287e+14		

---  
Signif. codes:  
0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
p adj

숙박업-쇼핑업	0.5570885
식음료업-쇼핑업	0.2038276
여가서비스업-쇼핑업	0.0433332
여행업-쇼핑업	0.0004174
운송업-쇼핑업	0.0108992
식음료업-숙박업	0.0016488
여가서비스업-숙박업	0.7966493
여행업-숙박업	0.0910693
운송업-숙박업	0.5064150
여가서비스업-식음료업	0.0000088
여행업-식음료업	0.0000000
운송업-식음료업	0.0000010
여행업-여가서비스업	0.7399730
운송업-여가서비스업	0.9972469
운송업-여행업	0.9424573

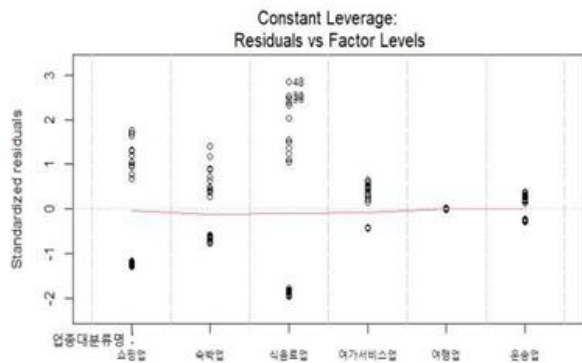
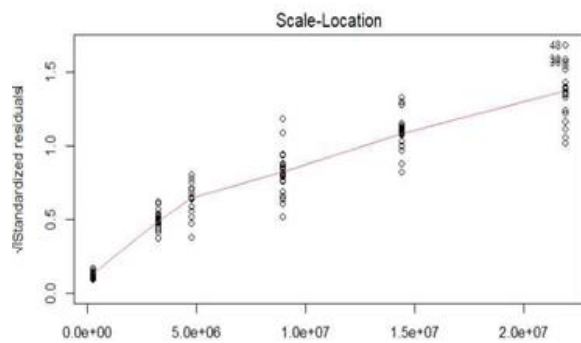
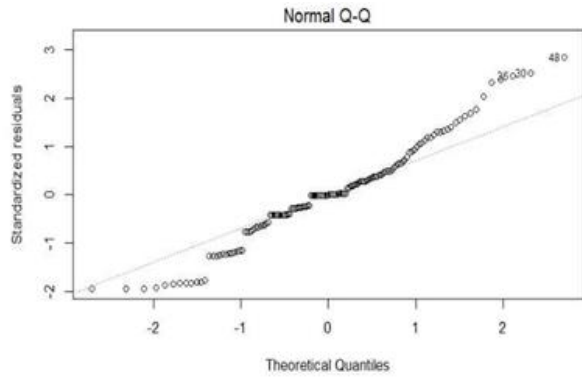
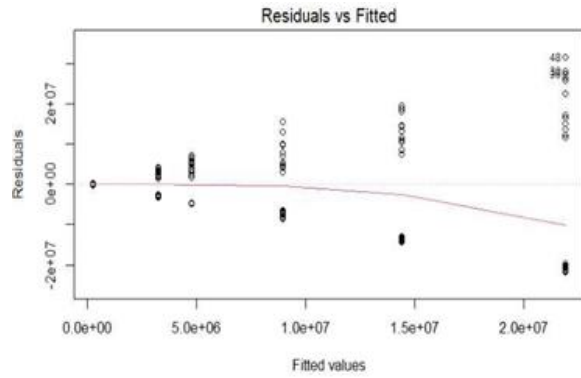
### <result 및 감상>

분석 결과, 업종별로 유의미한 소비 차이가 존재하는 것으로 나타났습니다.

특히 여행객들은 운송업, 서비스업, 숙박업에서 비슷한 수준의 비용을 지출하고 있으며, 식음료업과 쇼핑업에서도 소비 패턴이 유사하다는 흥미로운 결과가 도출되었습니다. 개인적으로도 식음료업과 쇼핑업에서 지출한 금액이 거의 비슷했던 경험이 있어, 이 분석 결과가 매우 인상 깊었습니다.

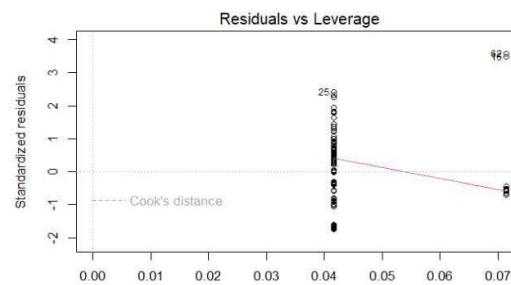
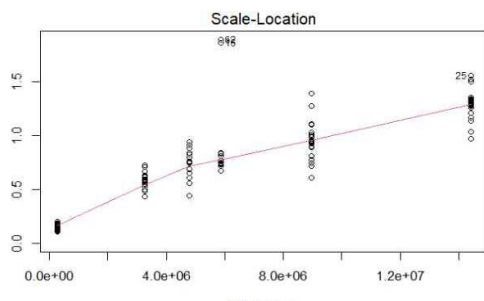
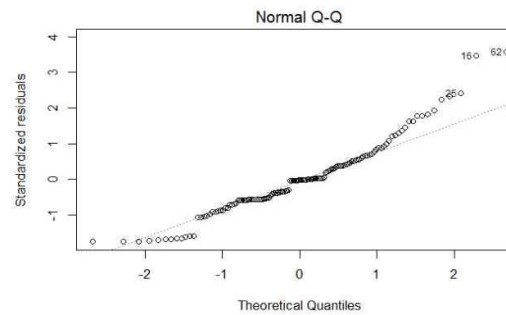
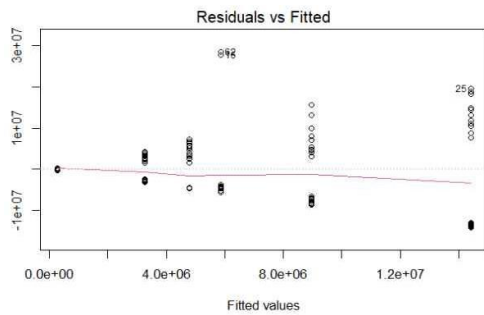
### <Conclusion>

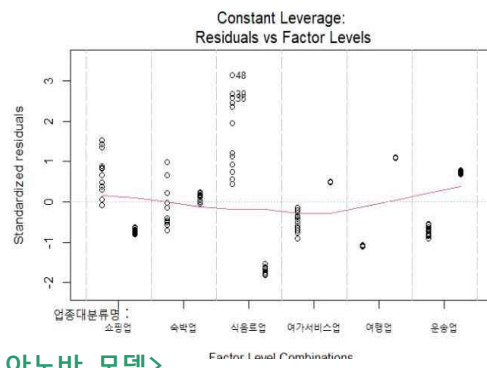
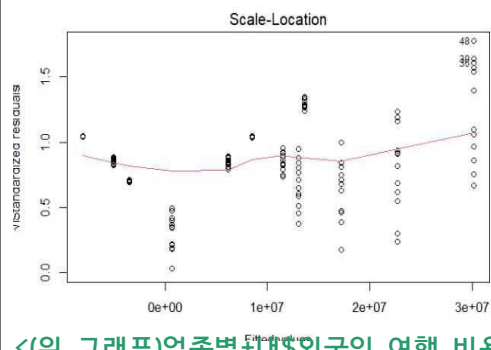
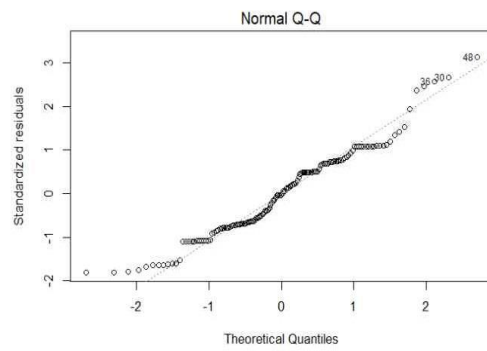
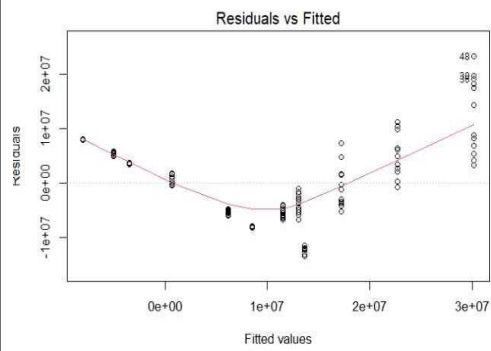
시각화를 통해 수행한 아노바(ANOVA) 분석에서는 이상치 제거의 필요성이 드러났으며, 또한 내국인과 외국인 요소를 추가 변수로 고려하여 총 3가지 아노바 모델을 구성하였습니다. 이러한 분석 확장을 통해 보다 정교한 통계 해석이 가능했고, 소비 유형에 영향을 미치는 요인들에 대한 깊은 통찰을 얻을 수 있었습니다.



(위 그래프) 업종별 여행 비용 아노바 모델 - 이상치 제거 전>

<아래그래프>업종별 여행 비용 아노바 모델 - 이상치 제거 후>





### <위 그래프)업종별+내\$외국인 여행 비용 아노바 모델>

정규성이 어느 정도있고 첫 번째 화면에 빨간색이 0에 가까이 붙는 이상치 제거 모델을 채택하려고 한다.