

응용통계학및실습

Final Project

< 최근 한국 영화관 분석 >

교수님: 임경필

학번: 20181256

이름: 김도현

1. 들어가며

필자는 어렸을 때부터 영화를 좋아했다. 재미있는 영화는 영화관에서 무조건 2번 이상 봐야만 했었다. 실제 한국 영화 산업은 엄청난 성장세를 이뤘다. 최근에 봉준호 감독의 2019년 작품인 '기생충'은 각종 아카데미 상을 휩쓸었다. 또한 우리 학교 동문이기도 한 박찬욱 감독은 2022년 작품인 '헤어질 결심'을 통해, 칸 영화제 감독상도 수상할 수 있었다.

하지만 최근 들어 영화관 산업은 긍정적으로 성장만 하지는 못하였다. 첫 번째 이유는 시대가 변함에 따른 OTT의 성장이다. OTT는 영화관에서 영화를 보는 것이 아니라 NETFLIX, WATCHA, COUPANG PLAY, TVING 등 월별 결제를 통해 무제한으로 영화 및 드라마를 시청할 수 있도록 하는 서비스를 애기한다. 처음에는 많은 사람들이 영화는 영화관에서 봐야 한다는 입장이 많았으나 또 하나의 사건이 터지게 된다. 바로 2020년 COVID-19의 등장이다. 이 전염병은 전 세계를 2년동안 팬데믹 상태로 만들었으며 의료 부문을 제외한 대부분의 산업의 성장세를 막았다. 영화 산업 역시 예외는 아니었다. 실내 활동에 인원제한이 생기면서 영화관에 손님이 끊기고, 적자를 이기지 못한 영화관들은 결국 영화 가격을 올리는 행동을 하게 되었다. 코로나 전 영화 가격은 일반석 2D 기준 8000원~10000원이었으나 코로나 이후 가격은 약 15000원이 되었다. 물가 상승의 요인도 있었으나 결국 영화관에서 관측하는 영화산업은 매우 축소되었다고 예측할 수 있다.

필자 역시 한국의 대표 영화관인 CJ_CGV 회사의 주식을 2020년에 샀었다. 이미 필자가 샀을 때 covid-19 직격탄을 맞은 후 충분히 저렴하다고 생각하였지만 계속된 적자, 영화관 산업의 축소로 인하여, 최근 1조원 유상증자를 했다. 영화관의 주가는 몇일 째 폭락하고 있다. 그래서 최근 영화관의 인식 및 실태 상황에 대해 자료 분석을 하고자 한다. 이는 최근 영화관 입장에서 가장 큰 사건 일 수 있는 COVID-19 전 후에 대한 영화관 자료 분석 역시 포함 하고자 한다. 해당 자료들은 한국 영화진흥위원회 KOFIC, 영화관입

장권통합전산망인 KOBIS 에서 가져온 것들이다.

2. 분석

우선 2020년 기준 covid 전인 2019년과 covid 후인 2022년에 대한 매출 차이를 통해, 전 세계적으로 영화관이 매출에 타격이 있었는지 proc ttest 를 통해 상위 10개 국가별 극장매출에 대한 분석을 해보고자 한다.

	미국	중국	일본	인도	영국	프랑스	한국	독일	호주	멕시코
2018	11,105.3	9,218.1	2,015.7	1,495.8	1,703.7	1,547.0	1,648.3	1,062.1	930.3	873.6
2019	10,657.0	9,301.9	2,395.9	1,621.5	1,597.7	1,591.3	1,642.8	1,146.4	853.8	989.2
2020	2,074.0	2,958.0	1,341.8	276.7	380.4	483.8	432.1	363.2	276.1	171.8
2021	4,272.0	7,332.8	1,474.5	500.7	765.8	756.6	510.4	441.4	454.6	369.6
2022	6,900.0	4,457.4	1,567.9	1,272.3	1,111.3	1,062.5	897.9	758.2	652.3	597.8

* 출처: Omdia

```
1 data movie;
2 do group='미국 ', '중국', '일본', '인도', '영국', '프랑스', '한국', '독일', '호주', '멕시코';
3 input before after @@;
4 output;
5 end;
6 cards;
7 10657 6900 9301.9 4457.4 2395.9 1567.9 1621.5 1272.3 1597.7 1111.3 1591.3
8 1062.5 1642.8 897.9 1146.4 758.2 853.8 652.3 989.2 597.8
9 ;
10 run;
11 proc ttest data=movie;
12 paired before*after;
13 run;
14
```

The TTEST Procedure					
Difference: before - after					
N	Mean	Std Dev	Std Err	Minimum	Maximum
10	1252.0	1637.5	517.8	201.5	4844.5

Mean	95% CL Mean	Std Dev	95% CL Std Dev
1252.0	80.5956 2423.4	1637.5	1126.3 2989.4

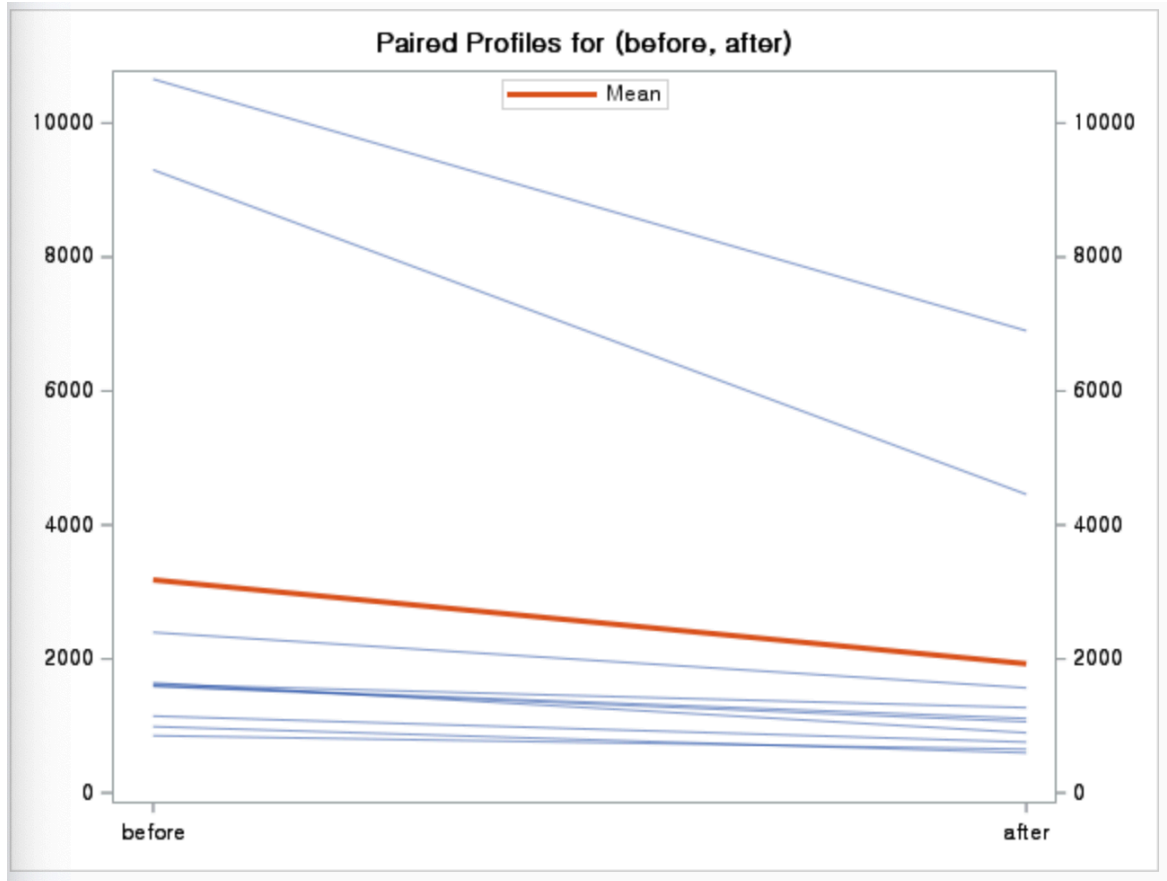
DF	t Value	Pr > t
9	2.42	0.0388

H_0 : 코로나 전과 코로나 후에 대한 전세계적인 영화관 매출 변화는 큰 차이가 없다.

H_1 : 코로나 전과 코로나 후에 대한 전세계적인 영화관 매출 변화는 차이가 있다.

해당 ttest 결과를 보면 $Pr > |t|$ 값이 0.0388 가 나왔다. 이는 p-value 가 0.0388 이기 때문에 4퍼센트 유의수준에서 H_0 를 기각 할 수 있다는 것이다. 즉 코로나 전과 코로나 후에 대해 세계적으로 영화관 매출에 변화가 생겼음을 알 수 있다.

추가적으로 양측 검정이 아닌 우측 검정을 비교해보자면 즉 H_1 : 코로나 전 이 코로나 후보다 영화관 매출이 더 높다. 라고 세운다면 t value가 양수 이기 때문에 해당 p-value의 절반으로 기각이 가능하다. 즉 약 2퍼센트 유의수준에서 기각이 가능함을 알 수 있다.



이 그래프를 통해서도 코로나 전 매출이 코로나 후 매출보다 평균적으로 높음을 알 수 있다. 즉 모든 영화관 매출은 확실히 covid-19 에 대한 타격을 입었다.

그러면 2020년을 기준으로 년도가 영화관 가격 인상률에 영향을 끼치는지 확인 해보도록 하겠다.

〈표 5〉 2018~2022년 각 국가별 전년 대비 영화 관람요금 인상률

(단위: %)

구 분	2018	2019	2020	2021	2022
미국	1.9	0.3	0.0	0.0	15.0
중국	2.9	4.8	0.1	8.8	4.2
일본	0.7	1.9	0.7	4.4	0.1
인도	13.0	19.3	16.5	-4.9	15.4
영국	-3.8	-1.4	-5.1	11.5	-0.9
프랑스	0.7	2.2	-2.1	2.5	0.7
한국	4.9	0.7	1.5	12.6	6.6
독일	-1.2	1.2	-3.3	6.2	5.8
호주	-1.9	4.6	-1.9	1.2	3.3
멕시코	5.8	7.5	5.2	15.3	0.7

* 출처: Omdia, Cinema Admissions & Box Office 4Q22, 로컬 화폐가치 기준(Average Ticket Price(Local)) 인상률

```

1 data movie;
2 do year='2018년','2019년','2021년','2022년';
3 input rate @@;
4 output;
5 end;
6 cards;
7 1.9 0.3 0.0 15.0
8 2.9 4.8 8.8 4.2
9 0.7 1.9 4.4 0.1
10 13.0 19.3 -4.9 15.4
11 -3.8 -1.4 11.5 -0.9
12 0.7 2.2 -2.1 2.5 0.7
13 4.9 0.7 12.6 6.6
14 -1.2 1.2 6.2 5.8
15 -1.9 4.6 1.2 3.3
16 5.8 7.5 15.3 0.7
17 ;
18 run;
19 proc glm;
20 class year;
21 model rate=year;
22 contrast '2019 vs 2022' year 0 1 0 -1;
23 contrast '2018~2019 vs 2021~2022' year 1 1 -1 -1;
24 means year/lsd;
25 run;
26 quit;

```

The GLM Procedure

Class Level Information		
Class	Levels	Values
year	4	2018년 2019년 2021년 2022년

Number of Observations Read	41
Number of Observations Used	41

The GLM Procedure

Dependent Variable: rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	120.546239	40.182080	1.26	0.3026
Error	37	1181.073273	31.920899		
Corrected Total	40	1301.619512			

R-Square	Coeff Var	Root MSE	rate Mean
0.092613	135.8617	5.649858	4.158537

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	3	120.5462395	40.1820798	1.26	0.3026

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	3	120.5462395	40.1820798	1.26	0.3026

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
2019 vs 2022	1	68.08050000	68.08050000	2.13	0.1526
2018~2019 vs 2021~2022	1	39.01821564	39.01821564	1.22	0.2760

H0: 년도에 따른 영화관 가격 인상률은 영향이 없다.

H1: 년도에 따른 영화관 가격 인상률은 영향이 있다.

Pr>F 를 확인해보면 0.3026 임을 알 수 있다. 이는 약 31퍼센트 유의수준으로 H0가 기각이 가능하다. 즉 년도에 따른 영화관 가격 인상률은 큰 영향이 없음을 알 수 있다. 그러면 2019년과 2022년 2개만 보고 비교하기 위해 contrast 명령어 옵션을 사용하였다. 여기서는 약 Pr>F 가 0.1526으로 약 16퍼센트 유의수준에서 년도에 따른 영화관 가격 인상률은 영향이 없다고 기각이 된다. 이 역시 생각보다 기각 수준이 높음을 알 수 있다.

데이터로만 비교하면 년도에 따른 영화관 가격 인상률은 크게 관련이 없다고 할 수 있다. 하지만 여기서 주의해야 할 점은 물가 상승률, 그리고 환율에 대한 비교 값이 빠져 있음을 유의해야 한다.

그리고 한국에 영향을 가장 많이 끼치는 미국, 중국, 일본, 한국 4개 국가만 가지고 년도별 가격 인상률 비교판단을 해본다면 결과는 상당히 달라진다.

```
data movie;
do year='2018년', '2019년', '2021년', '2022년';
input rate @@;
output;
end;
cards;
1.9 0.3 0.0 15.0
2.9 4.8 8.8 4.2
0.7 1.9 4.4 0.1
4.9 0.7 12.6 6.6
;
run;
proc glm;
class year;
model rate=year;
contrast '2019 vs 2022' year 0 1 0 -1;
contrast '2018-2019 vs 2021-2022' year 1 1 -1 -1;
means year/lsd;
run;
quit;
```

The GLM Procedure					
Dependent Variable: rate					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	71.4725000	23.8241667	1.25	0.3366
Error	12	229.5450000	19.1287500		
Corrected Total	15	301.0175000			

R-Square	Coeff Var	Root MSE	rate Mean
0.237436	100.2554	4.373643	4.362500

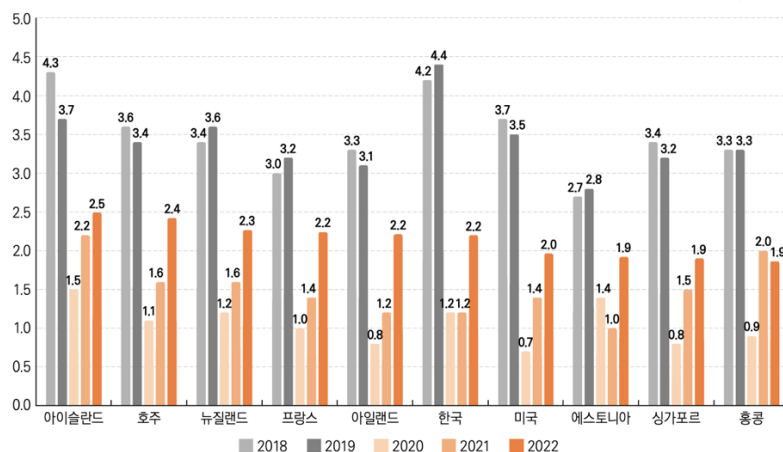
Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	3	71.47250000	23.82416667	1.25	0.3366

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	3	71.47250000	23.82416667	1.25	0.3366

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
2019 vs 2022	1	41.40500000	41.40500000	2.16	0.1670
2018~2019 vs 2021~2022	1	70.56000000	70.56000000	3.69	0.0789

여기서 중요하게 봐야할 데이터는 contrast를 통해 나오는 2018~2019년, 2021~2022년 이다. 결국 코로나 전 2개의 년도와 코로나 이후 2개의 년도를 비교하면 Pr>F를 보면 0.0789이기 때문에 약 8퍼센트 유의수준에서 H0가 기각 가능하므로 코로나 기준 전후에 따른 년도별 가격의 인상률이 어느 정도 영향이 있다는 것을 알 수 있다.

하지만 영화관의 매출을 분석하기 위해 이보다 더 중요한 것은 국가별 1인당 관람횟수인 것 같다.




```

1 data movie;
2 do year='2019년', '2022년';
3 input count @@;
4 output;
5 end;
6 cards;
7 3.7 2.5 3.4 2.4 3.6 2.3 3.2 2.2 3.1 2.2 4.4 2.2 3.5 2.0
8 2.8 1.9 3.2 1.9 3.3 1.9
9 ;
10 run;
11 proc anova data=movie;
12 class year;
13 model count=year;
14 means year/lsd;
15 run;
16 quit;

```

The ANOVA Procedure

Class Level Information

Class	Levels	Values
year	2	2019년 2022년

Number of Observations Read	20
Number of Observations Used	20

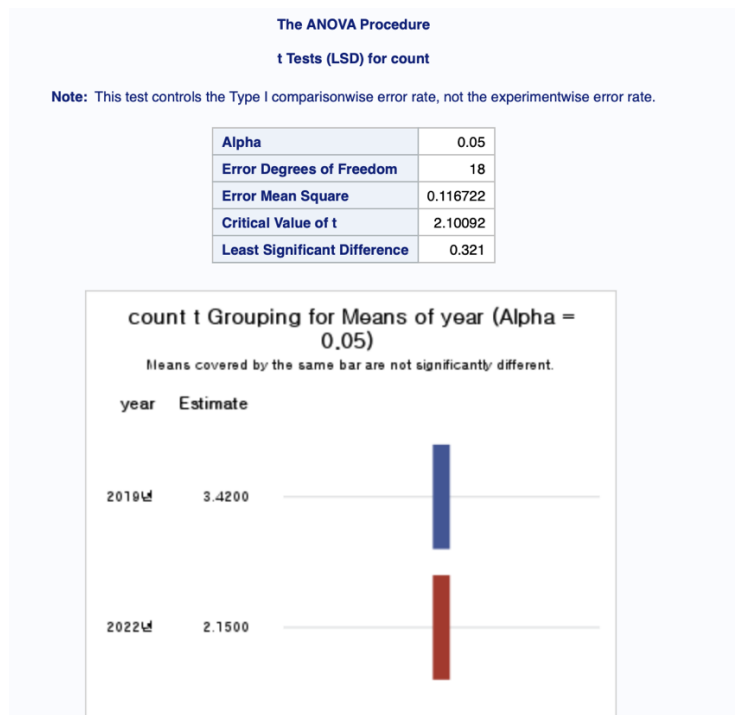
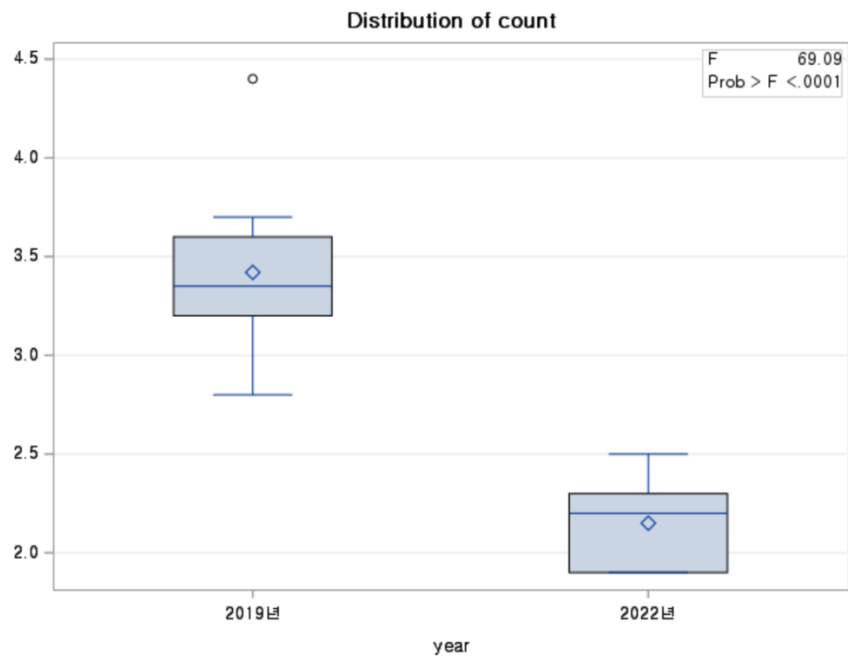
The ANOVA Procedure

Dependent Variable: count

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.06450000	8.06450000	69.09	<.0001
Error	18	2.10100000	0.11672222		
Corrected Total	19	10.16550000			

R-Square	Coeff Var	Root MSE	count Mean
0.793321	12.26737	0.341646	2.785000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
year	1	8.06450000	8.06450000	69.09	<.0001



H_0 : 년도에 따라 국가별 1인당 관람횟수에 영향이 없다.

H_1 : 년도에 따라 국가별 1인당 관람횟수에 영향이 있다.

이 분산분석에서는 약 1퍼센트 유의수준에서 H_0 가 기각됨을 알 수 있다. 코로나 후로 국가별 1인당 관람횟수가 영향이 있다는 것을 알 수 있다.

결론적으로 모든 국가의 영화관 산업은 코로나 이후로 성장성에 직격탄을 맞았고 아직까지 코로나 전으로 회복하지 못함을 알 수 있다.

그러면 현재 한국의 영화관들은 전국적으로 스크린을 적절하게 운영하고 있는지 궁금해져 다중 회귀분석을 진행하였다. 만약 스크린의 수가 인구 수와 관객 수 점유율에 비례하지 않는다면 스크린 운영을 잘못 하고 있는 것이기 때문이다.

〈표 8〉 2022년 광역단체별 매출액, 관객 수, 1인당 관람횟수

(단위: 개, 원, %, 명, 회)

구분	극장 수	스크린 수	좌석 수	관객 수	관객 수 점유율	매출액	매출액 점유율	평균 관람요금	인구 수	1인당 관람횟수
서울	90	591	94,487	28,924,221	25.6%	306,486,284,668	26.4%	10,596	9,428,372	3.07
부산	29	228	34,053	7,686,230	6.8%	77,592,126,987	6.7%	10,095	3,317,812	2.32
대구	22	145	20,912	5,124,982	4.5%	51,492,054,973	4.4%	10,047	2,363,691	2.17
인천	27	172	25,920	5,745,673	5.1%	60,017,187,502	5.2%	10,446	2,967,314	1.94
광주	17	128	17,473	3,740,228	3.3%	37,161,260,711	3.2%	9,936	1,431,050	2.61
대전	16	107	17,805	3,832,629	3.4%	40,045,566,788	3.5%	10,449	1,446,072	2.65
울산	9	56	8,583	2,399,252	2.1%	24,672,600,749	2.1%	10,283	1,110,663	2.16
세종	4	23	3,284	694,648	0.6%	7,199,052,411	0.6%	10,364	383,591	1.81
경기	139	854	116,306	28,481,402	25.2%	298,913,651,947	25.8%	10,495	13,589,432	2.10
강원	28	127	14,980	2,762,750	2.4%	26,151,431,882	2.3%	9,466	1,536,498	1.80
충북	21	119	15,981	3,070,645	2.7%	30,745,022,635	2.6%	10,013	1,595,058	1.93
충남	29	149	18,068	3,755,668	3.3%	37,848,185,011	3.3%	10,078	2,123,037	1.77
전북	29	139	19,277	3,311,508	2.9%	32,758,672,539	2.8%	9,892	1,769,607	1.87
전남	23	101	12,704	2,513,205	2.2%	23,876,999,762	2.1%	9,501	1,817,697	1.38
경북	33	152	17,422	3,739,796	3.3%	36,643,638,680	3.2%	9,798	2,600,492	1.44
경남	38	192	25,260	5,706,481	5.1%	55,615,771,055	4.8%	9,746	3,280,493	1.74
제주	7	39	4,165	1,315,744	1.2%	12,992,418,700	1.1%	9,875	678,159	1.94
합계	561	3,322	466,680	112,805,062	100.0%	1,160,211,927,000	100.0%	10,285	51,439,038	2.19

* 인구 수는 2022년 12월 31일 현재 행정안전부 주민등록인구통계(주민등록 인구 및 세대현황) 기준

```

data movie;
do region=1 to 17;
input screen rate people @@;
output;
end;
cards;
591 25.6 94 228 6.8 33 145 4.5 24 172 5.1 30
128 3.3 14 107 3.4 14 56 2.1 11
23 0.6 4 854 25.2 14 127 2.4 15
119 2.7 16 149 3.3 21 139 2.9 18 101 2.2 18
152 3.3 26 192 5.1 33 39 1.2 7
;
run;
proc reg data=movie;
model screen=rate people/stb;
run;
quit;

```

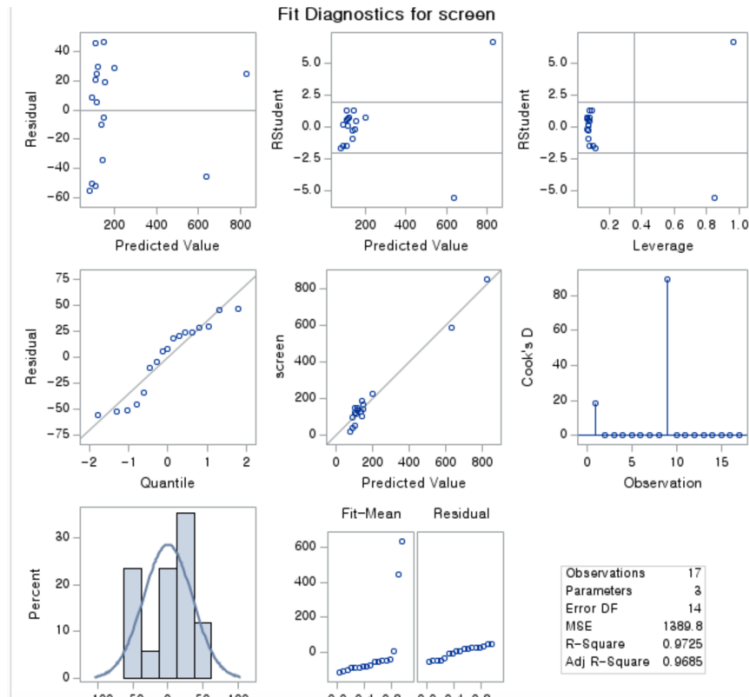
The REG Procedure
Model: MODEL1
Dependent Variable: screen

Number of Observations Read	17
Number of Observations Used	17

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	687575	343787	247.36	<.0001
Error	14	19457	1389.80534		
Corrected Total	16	707032			

Root MSE	37.28009	R-Square	0.9725
Dependent Mean	195.41176	Adj R-Sq	0.9685
Coeff Var	19.07771		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	69.42105	14.00565	4.96	0.0002	0
rate	1	31.60122	1.63308	19.35	<.0001	1.12806
people	1	-2.57347	0.60936	-4.22	0.0009	-0.24620



우선 intercept를 보면 1퍼센트의 유의수준으로 $b_0=0$ 이 기각됨을 알 수 있다. 즉 이 y절편이 0이 아님을 알 수 있다. 그리고 rate을 보면 1퍼센트 유의수준에서 $rate=0$ 이 기각됨을 알 수 있다. 즉 관객 수 에 대한 기울기가 존재한다. 그리고 people을 보면 1퍼센트 유의수준에서 $people=0$ 이 기각됨을 알 수 있다. 즉 사람 수 에 대한 기울기가 존재한다. Fit Diagnostics를 보면 Outlier가 2개정도 보이는데 observation을 보면 서울 지역 및 경기 지역이 가장 이 회귀분석에 영향이 크다. 그 이유는 서울 과 비교했을 때 스크린 수가 약 200개 이상 많음에도 불구하고 관객 수 점 유율은 더 낮기 때문이다.

전반적으로 R-square를 보면 1에 매우 가까운 0.9725가 나온다. 즉 직선에 거의 일치하고 있기 때문에 한국의 스크린 운영은 인구 수나 관객 수 점 유율을 잘 반영하고 있는 것 같다.

그럼 반대로 일인당 관람횟수에 영향을 미치는 것이 어떤 변수들이 있는지 확인하기 위해 다중회귀분석도 해보았다. 변수들은 극장 수, 스크린 수, 평균 관람요금, 인구 수 이다. 이때 인구 수는 $1/10000$ 으로 계산하였다. 그리고 변수 선택법은 backward 방식으로 SLSTAY=0.1로 약 10퍼센트 안에 있는 것들만 남아있도록 설정했다.

```

data movie;
do region =1 to 17;
input theater screen price people count @@;
output;
end;
cards;
90 581 10596 942 3.07
29 228 10095 331 2.32
22 145 10047 236 2.17
27 172 10446 296 1.94
17 128 9936 143 2.61
16 107 10449 144 2.65
9 56 10283 110 2.16
4 23 10364 38 1.81
139 854 10495 2.10
28 127 9466 153 1.80
21 119 10013 159 1.93
29 149 10078 212 1.77
29 139 9892 176 1.87
23 101 9501 181 1.38
33 152 9798 260 1.44
38 192 9746 328 1.74
7 39 9875 67 1.94
;
run;
proc reg data=movie;
model count=theater screen price people/selection=backward slstay=0.1;
run;

```

The REG Procedure
Model: MODEL1
Dependent Variable: count

Number of Observations Read	16
Number of Observations Used	16

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.7323 and C(p) = 5.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2087.11225	521.77806	7.52	0.0036
Error	11	763.03804	69.36709		
Corrected Total	15	2850.15029			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-166.54228	100.36155	191.01513	2.75	0.1252
theater	-0.05495	0.09850	21.59167	0.31	0.5881
screen	0.02008	0.01116	224.59176	3.24	0.0994
price	0.01678	0.00982	202.60533	2.92	0.1155
people	-0.01665	0.01100	158.81922	2.29	0.1584

Backward Elimination: Step 3

Variable price Removed: R-Square = 0.6069 and C(p) = 4.1499

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1729.87747	1729.87747	21.62	0.0004
Error	14	1120.27282	80.01949		
Corrected Total	15	2850.15029			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	4.45279	3.06098	169.33186	2.12	0.1678
screen	0.00220	0.00047241	1729.87747	21.62	0.0004

전체적으로 analysis of Variance 를 봤을 때 해당 선택된 변수들은 일인당 관람횟수에 영향이 있는 것으로 보인다. (H_0 를 약 1퍼센트 유의수준에서 기각) 이때 각각의 변수들이 어느정도 영향을 끼치는지 확인해보면 영화관 개수는 약 59퍼센트 유의수준, 영화 스크린 개수는 9.9퍼센트 유의수준에서 가격은 12퍼센트 유의수준, 인구수는 16퍼센트 유의수준에서 각 H_0 가 기각된다. (H_0 : 해당 변수들이 일인당 관람횟수에 영향을 미치지 않는다.)

우리는 $\alpha=0.1$ 로 잡았으니 최종적으로 스크린 개수 변수가 일인당 관람횟수에 가장 큰 영향을 미친다는 것을 알 수 있다. 이때 β_0 는 약 17퍼센트 유의수준에서 $\beta_0=0$ 이다를 기각하므로 y 절편이 0일 것으로 예측된다.

년도별 영화(외국영화, 한국영화) 선호도 역시 궁금하였다. 이를 확인해보면 한국사람들의 영화 취향을 확인할 수 있을 것이다. 이는 `proc freq` 명령어를 사용하였다.

자세한 자료는 여기서 참고하였다.

<https://www.kobis.or.kr/kobis/business/stat/them/findYearlyTotalList.do>

```
data movie;
input year $ country $ count @@;
cards;
2019년 한국 51.0 2019년 외국 49.0
2020년 한국 68.0 2020년 외국 32.0
2021년 한국 30.1 2021년 외국 69.9
2022년 한국 55.7 2022년 외국 44.3
;
run;
proc freq data=movie;
weight count;
tables year*country/nocol chisq measures;
run;
```

FREQ 프로시저

빈도 백분율 행 백분율	테이블 year * country			
year	country			
	외국	한국	합계	
2019년	49	51	100	
	12.25	12.75	25.00	
	49.00	51.00		
2020년	32	68	100	
	8.00	17.00	25.00	
	32.00	68.00		
2021년	69.9	30.1	100	
	17.48	7.53	25.00	
	69.90	30.10		
2022년	44.3	55.7	100	
	11.08	13.93	25.00	
	44.30	55.70		
합계	195.2	204.8	400	
	48.80	51.20	100.00	

year * country 테이블에 대한 통계량

통계량	자유도	값	Prob
카이제곱	3	29.9268	<.0001
우도비 카이제곱	3	30.6551	<.0001
Mantel-Haenszel 카이제곱	1	1.1307	0.2876
파이 계수		0.2735	
우발성 계수		0.2638	
크레머의 V		0.2735	

통계량	값	ASE
감마	-0.0774	0.0733
Kendall의 타우-b	-0.0486	0.0461
Stuart의 타우-c	-0.0595	0.0564

H0: 년도별 한국사람의 영화 국적 선호도가 동일하다.

H1: 년도별 한국사람의 영화 국적 선호도가 동일하지 않다.

카이제곱을 확인해보면 약 1퍼센트 유의수준에서 H0를 기각함을 알 수 있다. 즉 년도별 영화 국적 선호도는 동일하지 않다는 것이다. 연관성 측도에서도 연관성 측도/ASE 가 2 이상을 넘지 못하기 때문에 연관성이 거의 없다.

그럼 최근 9년치 데이터를 기준으로 한국사람들의 한국영화 선호도가 50프로를 넘기는지 궁금하여 univariate 명령어를 사용해보았다.

```
data movie;
input rate @@;
cards;
53.7 51.8 50.9 51.0 68.0 30.1 55.7 35.9 52.3
;
run;
proc univariate data=movie mu0=50 alpha=0.05 cibasic;
var rate;
run;
```


UNIVARIATE 프로시저			
변수: rate			
적률			
N	9	가중합	9
평균	49.9333333	관측값 합	449.4
표준 편차	11.0413994	분산	121.9125
왜도	-0.5254744	첨도	0.8606086
제곱합	23415.34	수정 제곱합	975.3
변동계수	22.1122818	평균의 표준 오차	3.68046646
기본 통계 속도			
위치속도		변이속도	
평균	49.93333	표준 편차	11.04140
중위수	51.80000	분산	121.91250
최빈값	.	범위	37.90000
		사분위수 범위	2.80000
정규성 가정하 기본 신뢰 한계			
모수	추정값	95% 신뢰한계	
평균	49.93333	41.44616	58.42050
표준 편차	11.04140	7.45799	21.15279
분산	121.91250	55.62163	447.44058
위치모수 검정: Mu0=50			
검정	통계량	p 값	
스튜던트의 t	t -0.01811	Pr > t	0.9860
부호	M 2.5	Pr >= M	0.1797
부호 순위	S 6.5	Pr >= S	0.4961

H0: M=50

H1: M<50

p 값을 보면 0.9860이 나온다. 그리고 t는 음수값이다. 즉 H0를 (0.9860/2) *100 의 유의확률로 기각한다는 뜻이다. 이는 대략 50퍼센트 유의수준에서 기각된다는 말이기 때문에 H0를 기각하지 못한다. 즉 일반적으로 한국영화와 외국 영화에 대한 선호도는 비슷하다는 것을 알 수 있다. 평균에 대한 95퍼센트의 신뢰구간은 41.4~58.4 임을 확인할 수 있다.

이제 이원배치법을 사용하기 위해 영화 장르별, 영화 국적 별로 영화 선호도 차이가 있는지 확인해 보았다.

여기서 주의해야 할 점은 이에 대한 데이터를 구하기 어려워 필자가 직접 2021년, 2022년 여러 데이터를 종합하여 가상 데이터를 만들었다는 점을 인지하자!

영화 장르별, 한국, 외국 영화 비율에 맞게 영화 선호도를 가상으로 데이터화 해보았다.

```

data movie;
do kind='액션' , '범죄' , '애니메이션' , '드라마' , '코미디' , '스릴러' , '기타' ;
do country='한국' , '외국' ;
do rep=2021,2022;
input count @@;
output;
end;
end;
end;
end;
cards;
38 40 65 71
22 24 2 1
2 3 15 17
15 13 1 2
5 6 1 1
4 5 1 1
7 8 6 6
;
run;
proc glm data=movie;
class kind country;
model count=kind country kind*country;
means kind country kind*country;
run;

```

The GLM Procedure

Class Level Information

Class	Levels	Values
kind	7	기타 드라마 범죄 스릴러 애니메이션 코미디
country	2	외국 한국

Number of Observations Read	28
Number of Observations Used	28

The GLM Procedure

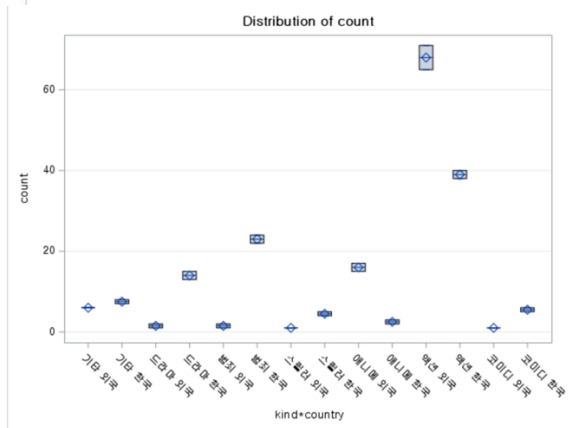
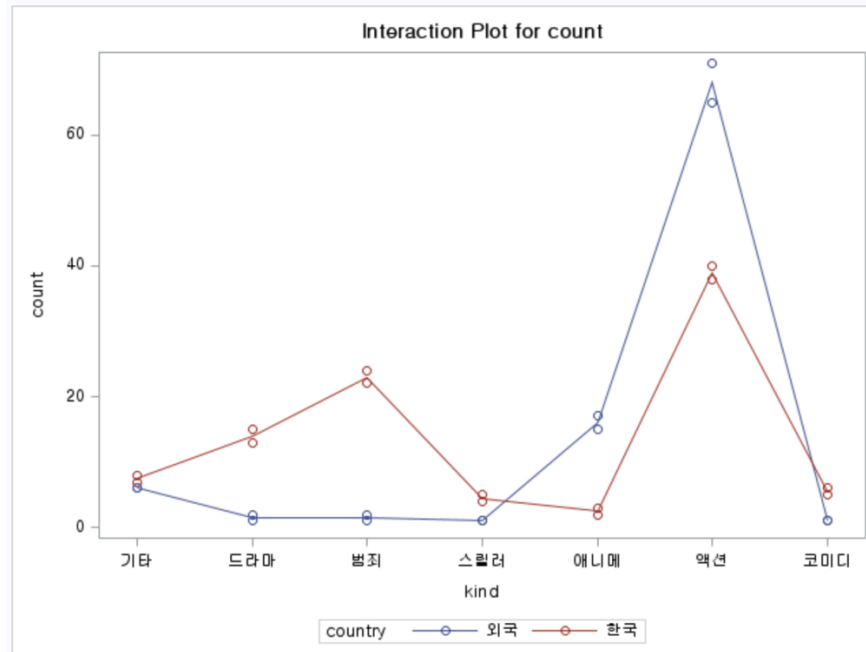
Dependent Variable: count

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	9351.428571	719.340659	347.27	<.0001
Error	14	29.000000	2.071429		
Corrected Total	27	9380.428571			

R-Square	Coeff Var	Root MSE	count Mean
0.996908	10.54945	1.439246	13.64286

Source	DF	Type I SS	Mean Square	F Value	Pr > F
kind	6	7674.928571	1279.154762	617.52	<.0001
country	1	0.142857	0.142857	0.07	0.7967
kind*country	6	1676.357143	279.392857	134.88	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
kind	6	7674.928571	1279.154762	617.52	<.0001
country	1	0.142857	0.142857	0.07	0.7967
kind*country	6	1676.357143	279.392857	134.88	<.0001



해당 데이터를 분석해보면 영화 종류에 따라 선호도가 확실히 차이가 있음을 알 수 있다. (H_0 : 영화 종류에 따른 선호도 차이는 없다. 약 1퍼센트 유의수준에서 기각) 반면 영화 국적만 봤을 때는 선호도에 대해 유의미한 차이가 없음을 알 수 있다. (H_0 : 영화 국적에 따른 선호도 차이는 없다. 약 80퍼센트 유의수준에서 기각) 그리고 교호작용을 살펴보면 두개를 동시에 보았을 때 역시 1퍼센트 유의수준에서 기각됨을 알 수 있다. 즉 교호작용도 어느정도 효과가 있는 것으로 보인다.

그러면 실제로 사람들이 영화관의 가격이 비싸다고 생각할까? 해당 자료는 20대 203명을 대상으로 영화관 영화 관람 비용이 비싼지에 대해 2020년 12월에 설문조사한 자료이다.

<https://www.kofic.or.kr/kofic/business/rsch/findPolicyDetail.do?boardNumber=39&policyNo=4506#none>

```
data movie;
input yesno$ count;
cards;
yes 152
no 51
;
run;
proc freq data=movie order=data;
weight count;
exact binomial;
tables yesno/ alpha= 0.05;
run;
```

FREQ 프로시저				
yesno	빈도	백분율	누적 빈도	누적 백분율
yes	152	76.38	152	76.38
no	47	23.62	199	100.00

이항비	
yesno = yes	
비율(P)	0.7638
ASE	0.0301
95% 신뢰하한	0.7048
95% 신뢰상한	0.8228
정확 신뢰한계	
95% 신뢰하한	0.6986
95% 신뢰상한	0.8210

H0: P = 0.5의 검정	
H0 하에서의 ASE	0.0354
Z	7.4433
단측 Pr > Z	<.0001
양측 Pr > Z	<.0001
정확 검정	
단측검정 Pr >= P	<.0001
양측 = 2*단측	<.0001

근사표준오차인 ASE의 크기가 작다는 것을 보아 해당 분석 결과는 적절하게 나왔다고 볼 수 있다. 현재 20대의 약 76센트의 비율이 영화관의 티켓 가격이 비싸다는 결과가 나왔다. 95퍼센트의 신뢰구간을 확인했을 때 역시 최소 10명중 7명 꼴은 영화관의 티켓가격이 비싸다고 생각하는 것이다.

```
data movie;
input gender$ yesno$ count @@;
cards;
남자 yes 26 남자 no 9
여자 yes 126 여자 no 42
;
run;
proc freq data=movie order=data;
weight count;
tables gender*yesno/ chisq fisher expected nopercnt nocol;
run;
```

FREQ 프로시저

빈도 기대값 행 백분율	테이블 gender * yesno			
	gender	yesno		
		yes	no	
	남자	26 26.207 74.29	9 8.7931 25.71	
		35		
	여자	126 125.79 75.00	42 42.207 25.00	
168				
합계		152	51	203

gender * yesno 테이블에 대한 통계량

통계량	자유도	값	Prob
카이제곱	1	0.0079	0.9294
우도비 카이제곱	1	0.0078	0.9295
연속성 수정 카이제곱	1	0.0000	1.0000
Mantel-Haenszel 카이제곱	1	0.0078	0.9295
파이 계수		-0.0062	
우발성 계수		0.0062	
크래머의 V		-0.0062	

Fisher의 정확 검정	
(1,1) 셀 빈도(F)	26
하단측 p값 Pr <= F	0.5406
상단측 p값 Pr >= F	0.6269
테이블 확률 (P)	0.1675
양측 p값 Pr <= P	1.0000

표본 크기 = 203

조금 더 자세하게 전체 203명 응답자 중 35명은 남자, 168명은 여자였다. 성별에 따라 영화 티켓이 비싸다고 생각하는지 알아보기 위해 `chisq` 옵션도 추가해서 독립표본에 의한 두 모비율 비교를 해보았다. 기대빈도가 5이하가 25프로 이상이라면 `fisher` 검정을 해야 하기 때문에 `fisher` 옵션도 사용해보았다.

H0 : Pa=Pb

H1 : Pa~Pb

카이제곱을 확인해보면 약 93퍼센트에서 H0가 기각됨을 확인할 수 있다. 이는 성별에 관계 없이 영화관 티켓 가격이 비싸다고 생각된다고 얘기할 수 있다.

3. 나오며

지금까지 영화관을 기준으로 각종 분석을 진행하였다. 확실히 covid-19은 전세계적으로 영화관의 매출에 큰 타격을 입혔음을 알 수 있다. 그리고 아직 까지 covid-19 이전의 영화관 매출 및 성장성을 회복하지 못했다고도 결론 내릴 수 있다. 여기서 중요한 것 중 하나는 일인당 영화 관람 횟수이다. 이 역시 전세계적으로 코로나 이전, 코로나 이후에서 분명한 차이가 보이는데 코로나로 인한 ott service 의 발달도 있겠지만 필자는 코로나 이후 영화 관람 비용 가격 증가로 인한 사람들의 영화관 인식 변화 기준으로 분석했었다. 이 역시 현재 영화 관람 비용이 비싸다고 생각하는 사람이 남녀를 불문하고 70퍼센트 정도이니 적절한 주장이라고 볼 수 있다.

2022년 한국 영화관에 대해서는 회귀분석을 통해 전체적으로 분석했다. 결과적으로 일인당 영화관람 횟수에 가장 큰 영향을 미치는 것은 스크린 개수임을 알 수 있었다.

영화관 매출을 올릴 수 있는 방법 관점에서, 이를 과대해석해서는 안되지만 그럼 일인당 영화관람 횟수가 적은 지역에 스크린 개수를 많이 설치한다면 어떻게 될까? 라는 궁금증도 들었다. 스크린 설치 비용 대비 영화관람 횟수 증가에 대한 자료가 있으면 어느정도 분석이 가능하지 않을까 생각한다.

영화 장르 및 영화 국적에 대한 한국사람 선호도 분석도 했다. 영화 국적은 한국사람들에게 큰 영향을 미치지 않은 것으로 보였다. 재미있는 영화가 있다면 영화 국적을 불문하고 본다는 의미이다. 하지만 영화 장르의 경우 액션이 가장 선호도가 높았는데 교호작용을 넣으면 또 추가 해석이 가능했다. 외국 영화의 경우 범죄 장르에 대한 선호도가 매우 낮았으나 한국 영화는 범죄 장르에 대한 선호도가 2번째로 높았다. 전체적인 영화 국적으로 영화 선택의 차이는 보이지 않지만 어떤 장르의 어떤 국적인지는 중요한 요소라고 볼 수 있다. 이는 한국 영화 제작사들이 범죄 장르에 대해 한국 사람들의 입 맛에 더 잘 맞추고, 자본이 많이 필요한 액션 영화는 외국의 대규모 자본이 들어간 외국 영화 제작사들이 액션 영화를 잘 만든다고 볼 수도 있을 것 같다.

몇 년 전만 하더라도 아침 조조 영화를 5천원에 보던 시절이 있었다. 물론 물가도 상승하기는 했지만 결국 코로나로 인해 영화관들의 사업 사정이 어려

워지면서 어쩔 수 없는 관람 비용 상승으로 어느새 평균 영화 티켓 1만원 이상이 되었다. 더 이상 심심하면 영화나 보러갈래? 라는 말이 이치에 맞지 않는 말이 되었을 수도 있다.

한편으로는 사람들의 영화를 고르거나 보는 눈이 높아져서 기존의 아무 상업영화나 찍어냈던 영화산업에서는 이제 퀄리티 높은 영화를 만들어내야만 사람들의 관심도와 수익을 올릴 수 있다는 것을 인지 했을 것이다. 해당 내용에 대해서는 분석을 하지 않았으나 나중에 기회가 된다면 영화 투자비용, 제작 기간 대비 수익률을 최근과 과거를 나누어 데이터를 비교해봐도 좋을 것 같다.

처음 이렇게 데이터를 추출하여 분석해보는 작업을 했는데 생각보다 데이터를 추출하는 것, 그리고 내가 원하는 정보만 뽑아 내는 것이 어려웠다. 그래도 이번 기말 프로젝트를 통해서 다시 한번 이번 학기에 배운 수업 내용들을 상기시키고 영화관에 대해서도 필자가 생각한 것이 맞는지 실제 데이터를 토대로 확인하는 과정이 재밌었다.

4. 참고 사이트 및 자료 출처

- <https://www.kofic.or.kr/kofic/business/main/main.do>
- <https://www.kobis.or.kr/kobis/business/main/main.do>
- 온라인 영화 시장 변화 및 산업 전망 분석 조사(정인숙, 2020-12)
- 영화티켓지수로 알아본 영화관람 가격 적정성 점검(김성희, 2022-10)
- 2022 년 한국영화산업 결산(손진아, 2023-04)