계량경제학 Term paper

**영화 흥행에 영향을 주는 요인 분석**

**경제학과 20170547 문예진**

**경제학과 20170567 서정화**

**역할분담**

**문예진 : 자료수집 및 가공, Term paper 작성 및 정리**

**서정화 : 자료 수집, eviews 분석, Term paper 작성**

**목차**

**1.** 서론

1-1. 연구의 동기 및 목적

1-2. 모형의 설정과정

**2.** 실증분석 모형

2-1. 추정모형

2-2. 사용자료

3. 실증분석 결과

3-1. 설정모형에 대한 회귀분석

3-2. 부적합한 변수의 제거

3-3. 다중공선성 검정

3-4. 이분산 검정

3-5. 자기상관 검정

4. 주요추정결과 및 예측

5. 결론

1. **서론**

1. 연구의 동기 및 목적

2018년 기준, 한국의 1인당 연간 영화 관람횟수는 4.18회로, 미국의 3.7회보다 높은 상황이다. 영화산업 매출과 총 관객 수는 매년 증가하는 추세로, 영화 산업이 커지고 있다는 것을 알 수 있다. 2019년 월평균 약 160편의 영화가 개봉하는데, 과연 이렇게 다양한 영화들 중 흥행하는 영화는 어떻게 결정되는지, 어떤 요소들이 영화의 흥행에 영향을 미치는지에 의문이 들었다. 이에 따라 흥행 요소들을 예상해보고, 예상이 합리적인지, 또한 추측한 요소들이 관객 수를 끌어올리는데 얼마나 영향력이 있는지를 분석하고자 하였다.

2. 모형의 설정과정

­­“영화의 흥행요소는 창조의 영역, 배급의 영역, 홍보의 영역, 경쟁의 영역으로 분류가 가능한다” [[1]](https://d.docs.live.net/11f30f74b3d3639d/%EB%AC%B8%EC%84%9C/%EA%B3%84%EB%9F%89%EA%B2%BD%EC%A0%9C%ED%95%99%20%EB%B0%9C%ED%91%9C%20%EA%B3%84%ED%9A%8D%EC%84%9C.docx#_ftnref1) 전희국, 『영화 흥행 실적 예측을 위한 빅데이터 전처리 (Big Data Preprocessing for Predicting Box Office Success)』, KIISE, 제20권 제12호(2014.12)

문헌자료를 참고하여 각 영역에 해당하는 독립변수를 설정하였다.

⇒ “영화의 흥행요소는 창조의 영역, 배급의 영역, 홍보의 영역, 경쟁의 영역으로 분류가 가능하다.”(전희국;2014)

창조의 영역에는 영화의 창조성을 평가하는 네이버 관람객 평점과 로튼 토마토 지수를 독립변수로 사용할 수 있다. ‘배급’의 영역과 ‘경쟁’의 영역으로 스크린 수와 동월 개봉작 수는 활용하고, ‘홍보’ 부문은 인스타그램 해시태그 수를 사용한다. 종속 변수는 초기에 매출액으로 설정했으나, 이는 시간에 따른 티켓의 가격 변동이 반영되기 때문에 관객 수로 설정하였다.

1. **실증분석 모형**
2. 추정모형

1-1. 변수설정

* 종속변수 Y : log(관객수)

가장 바람직한 흥행지표는 매출액/투자비 비율이라고 생각되지만 투자액은 투자사의 대외비로 분류되어 있어 정보의 접근이 어렵다. 매출액을 종속변수로 두었더니 물가변동의 영향을 반영하지 못하여, 관객수를 종속변수로 삼는다.

영화를 선정한 기준은 탑 200 누적매출액이고, y변수의 분산 및 단위가 매우 크기 때문에, 로그 선형 모형 채택하였다.

⇒ 가장 바람직한 흥행지표는 투자비 대비 매출액 비율이지만, 투자액은 대외비로 분류되어 있는 정보이기 때문에 투자사에서 발표하지 않으면 정보의 접근이 어렵다. 그리고 누적 매출액은 물가변동의 영향을 받기 때문에, 관객수를 종속변수로 삼는다. 또한 변수의 분산 및 단위가 다른 변수에 비해 매우 크기 때문에 로그 선형 모형을 채택하였다.

* 독립변수 X

영화 흥행에는 참신함(Creativeness), 분배정도(Distribution), 화제성 및 대중성(Marketing), 경쟁력(Competition) 등이 영향을 미친다고 가정한다.[[1]](https://d.docs.live.net/11f30f74b3d3639d/%EB%AC%B8%EC%84%9C/%EA%B3%84%EB%9F%89%EA%B2%BD%EC%A0%9C%ED%95%99%20%EB%B0%9C%ED%91%9C%20%EA%B3%84%ED%9A%8D%EC%84%9C.docx#_ftn1)

: 로튼 토마토 지수

로튼 토마토는 영화 관련 웹사이트 가운데 하나로, 주로 비평가 위주의 평점을 매기는 곳이다. 때문에 본 연구에서 예술성,참신함을 위한 지표로 사용하였다.

② : 스크린 수/100 → 분배정도 (스크린(코비스)

해당 영화 상영 기간 중 최대 스크린수, 전국 영화 전체 스크린수

③ : 인스타그램 해시태그 수/1000000 → 화제성

④ : 네이버 관람객 평점 → 대중성

⑤ : 개봉 달 동시상영 영화개수 → 경쟁력

- 가변수 D

: 책 또는 만화로 원작이 존재할 경우, 원작의 팬 층이 영화 매출에 영향을 줄 것이다. Series는 영화의 시리즈 여부, Source는 원작 소설/만화/영화의 존재 여부를 가변수로 설정한 것이다.

경쟁과 시리즈 물인경우, 음의 관계, 그 외에는 양의 관계를 가질 것이다. 특히 스크린수가 가장 큰 영향을 줄 것으로 예측한다.

2. 사용자료

* 데이터는 www.kobis.or.kr 에 공개된 자료에서, 누적매출액이 높은 200개의 영화를 선정하였다.

자료의 기초적인 분석

① 평균, 오차, 최댓값, 최솟값

사용한 데이터의 요약 통계량은 다음과 같다.

|  | 관객수 | 스크린수/100 | 네이버 평점 | 같은 달에 개봉한 영화개수 | 인스타그램 해시태그수  /1000000 | 로튼토마토 | 원작여부\_시리즈 | 원작여부\_소설/만화 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 평균 | 6094689 | 10.64175 | 8.5298 | 2.01 | 0.197899555 | 50.585 | 0.225 | 0.27 |
| 표준 편차 | 2862957 | 4.872114 | 0.739446 | 0.913275 | 1.213714884 | 38.79547 | 0.41863 | 0.445074 |
| 최소값 | 3280151 | 2.38 | 5.58 | 1 | 0.000018 | 0 | 0 | 0 |
| 최대값 | 17615658 | 28.35 | 10 | 4 | 12.721432 | 100 | 1 | 1 |

[자료1]

② 변수간 상관관계 수정하고 자료2 : 음, 양 상관관계 명시

|  | 관객수 | 스크린수/100 | 네이버 평점 | 같은 달에 개봉한 영화개수 | 인스타그램 해시태그수  /1000000 | 로튼토마토 | 원작여부\_시리즈 | 원작여부\_  소설/만화 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 관객수 | 1 |  |  |  |  |  |  |  |
| 스크린수  /100 | 0.381294 | 1 |  |  |  |  |  |  |
| 네이버 평점 | 0.215633 | 0.170525 | 1 |  |  |  |  |  |
| 같은 달에 개봉한 영화개수 | 0.04614 | 0.039523 | 0.014885 | 1 |  |  |  |  |
| 인스타그램 해시태그수  /1000000 | 0.095491 | -0.03627 | -0.12119 | -0.04778 | 1 |  |  |  |
| 로튼토마토 | 0.12748 | 0.321845 | 0.298057 | -0.02343 | -0.06415 | 1 |  |  |
| 원작여부\_  시리즈 | -0.12822 | 0.19023 | -0.11901 | -0.13735 | -0.07953 | 0.227316 | 1 |  |
| 원작여부\_  소설/만화 | -0.12343 | 0.076162 | 0.085213 | -0.0314 | -0.03735 | 0.022237 | -0.32769 | 1 |

⇒ (순서 바꿈) 1. 사용자료

영화는 KOBIS에 공시된 자료 중 역대 누적매출액 순위 1위부터 200위로 선정하였으며, 2019년 11월 27일에 검색한 자료를 토대로 진행하였다. 각 변수 자료의 요약 통계량은 다음과 같다.

1. 추정모형

① 종속변수 : 가장 바람직한 흥행지표는 투자비 대비 매출액 비율이지만, 투자액은 대외비로 분류되어 있는 정보이기 때문에 투자사에서 발표하지 않으면 정보의 접근이 어렵다. 하지만 누적 매출액을 종속변수로 두게 되면, 물가변동의 영향을 우리가 제거하기 어렵기 때문에 시계열 변화에 영향을 받지 않는 관객수를 종속변수로 삼는다. 또한 변수의 분산 및 단위가 다른 변수에 비해 매우 크기 때문에 로그 선형 모형을 채택하였다.

② 독립변수 : 인스타그램 해시태그 수는 홍보의 영역을 대표한다. 인스타그램에서 ‘#영화제목’로 검색했을 때 나오는 게시글 수를 자료로 사용하였고, 1단위는 100만개로 조정하였다.

③ 독립변수 : 네이버 관람객 평점은 영화의 창조성에 대한 관람의 평가와 선호를 보여준다. 네이버 영화에서 검색했을 때 나타나는 평점을 수집하였고, 1단위는 1점이다.

④ 독립변수 : 같은 달에 개봉한 영화의 개수를 나타내는 지표로, 경쟁의 영역을 평가하기 위해 사용되었다.

⑤ 독립변수 : 로튼 토마토 지수로, 영화의 창조성에 대한 비평가들의 평가와 선호를 보여준다.

⑥ 독립변수 : 스크린 수를 나타내는 변수로, 배급 부분을 수치화한 것이다. 1단위는 100개이다.

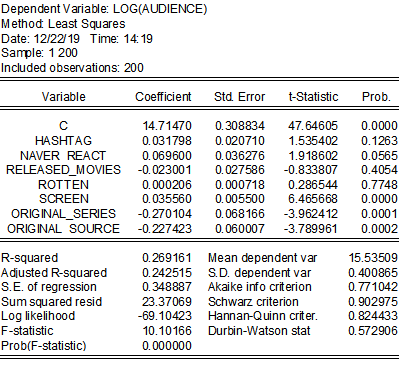
⑦ 가변수 : Series는 영화의 시리즈 여부를 가변수로 측정, 절편의 변화에 이를 반영하였다. Source는 원작 소설 혹은 원작 만화가 존재 여부를 가변수로 설정한 것이다. 시리즈가 존재하면, 원작 소설 혹은 만화가 존재하면로 계산한다.

: 책 또는 만화로 원작이 존재할 경우, 원작의 팬 층이 영화 매출에 영향을 줄 것이다. Series는 영화의 시리즈 여부, Source는 원작 소설/만화/영화의 존재 여부를 가변수로 설정한 것이다.

추정 모형에 대해 동월 개봉작의 수은 관람객수와 음의 관계, 그 외에는 양의 관계를 가질 것으로 예상했다. 특히 스크린 수가 가장 큰 영향을 줄 것으로 예상한다.

**III. 실증분석 결과**

1. 설정모형에 대한 회귀분석



[자료3]

위에서 설정한 회귀모형으로 회귀분석 한 결과 [자료3] 과 같은 OLS 추정치를 얻을 수 있었다.

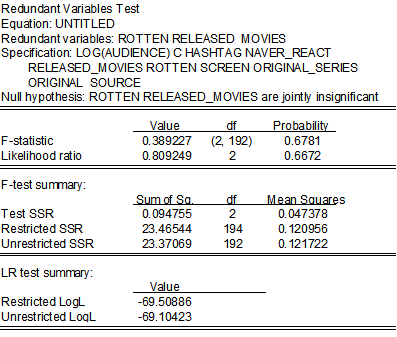
먼저, screen수, 네이버 평점, 원작여부(시리즈물, 소설,만화 및 영화)가 10% 유의수준에서 유의한 변수로 나타났다. 인스타그램 해쉬태그도 10% 유의수준에서는 기각되지만, 15% 수준으로 높일 경우 유의미한 변수로 취급된다.

반면에, 같은 달에 개봉한 영화의 수를 나타내는 Released\_movies와 로튼토마토 평가를 나타내는 Rotten변수의 p-value가 각각 0.41 과 0.77 로 높은 값을 보이며 10% 15% 유의검정에서 모두 유의적이지 않은 것으로 나타났다.

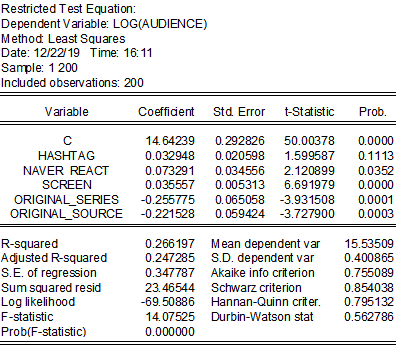
⇒ 위의 회귀 모형을 회귀분석 한 결과 [자료2]와 같은 OLS 추정치를 얻을 수 있다. 먼저 스크린 수(Screen), 네이버 평점(NaverReact), 시리즈물 여부(Series), 원작 소설 및 만화 여부(Source) 변수는 10%유의수준에서 유의한 변수로 나타났다. Hashtag도 10% 유의수준에서는 기각되지만, 15%수 준으로 높일 경우 유의미한 변수로 취급된다???. 이는 Hashtag 변수가 영화 흥행을 제대로 설명하지 못한다는 것으로 해석할 수 있다.

반면에, 같은 달에 개봉한 영화 수(ReleasedMovies)와 로튼 토마토 지수(Rotten)는 p-value 값이 각각 0.4054와 0.7748 로 높은 값을 보이며 10% 15% 의 유의검정에서 모두 유의적이지 않은 것으로 나타났다.

2. 부적합한 변수의 제거



[자료4]



[자료5]

초기의 모형에서 released\_movies와 rotten에 대해 부적합 변수에 대한 검정을 시행한 결과 [자료4]와 [자료5]와 같이 나타났다.

먼저, F-statistic의 유의확률을 나타내는 probaility가 10%, 15%의 유의검정에서 모두 기각되지 않기때문에, released\_movies와 rotten변수는 유의미하지 않다 라는 가정은 기각되지 않는다.

따라서, released\_movies와 rotten변수는 유의미하지 않다.

정보 기준(information criteria)을 고려해 자료를 살펴보면, AIC(Akaike infromation criterion) 값이0.7710 에서 7551 로, SC(Schwarz criterion) 값이 0.9030에서 0.8540로 하락하는 것을 관찰할 수 있다. 따라서 변수 released\_movies와 rotten을 제외한 모형이 더 적합한 모형임을 알 수 있다.

또한 나머지 변수 screen, naver\_react, hashtag, orignial\_series, original\_source는 각각 유의확률 0.0000, 0.1113, 0.0001, 0.0003 의 값을 보이면서 15% 유의수준에서 변수가 유의적이지 않다는 귀무가설을 모두 기각할 수 있다. 따라서 위의 네 변수는 모두 유의적인 변수임을 알 수 있다.

따라서 다음과 같은 새로운 회귀분석 식을 도출하였다.

log(Revenue) = c + b1screen +b2naver\_React+b3hastag+b4original\_sereis+b5original\_source +ei

⇒ 초기 모형 OLS 분석 [자료2]에 따르면 ReleasedMovies와 Rotten 변수는 유의미하지 않기 때문에 제거하는 것이 좋을지에 대한 검정을 시행해야 한다. 두 변수에 대해 부적한 변수 검정을 시행하면 [자료3]과 [자료4]로 나타난다. [자료3]에 따르면, F-통계량의 유의확률이 0.6781로, α=10%,15%의 유의검정 모두에서 기각되지 않는다. 즉 두 변수 모두 유의미하지 않다는 가정은 기각되지 않는다.

정보 기준 Information Criteria를 고려하여 자료를 살펴보면, AIC(Akaike Information Criterion) 값이 0.7710에서 0.7551로 감소하였고 SC(Schwarz Criterion) 값 또한 0.9030에서 0.8540으로 감소하는 것을 알 수 있다. 이는 ReleasedMovies와 Rotten 변수를 제거한 모형이 더 적합한 모형임을 보여준다. 나머지 변수 (Screen, NaverReact, Hashtag, OriginalSeries, OrinigalSource)의 유의확률은 각각 (0.0000, 0.1113, 0.0001, 0.0003)으로 [자료2]의 유의확률보다 조금 증가했지만, α=15%의 유의수준에서 귀무가설이 기각되는 것을 알 수 있다. 따라서 네 변수는 모두 유의적이라고 해석할 수 있다. 이에 따라 새로운 회귀식을 도출하면 다음과 같다.

1. 다중공선성 검정

설명 변수들 사이에 강한 상관관계를 보이는 다중공선성이 존재하는 경우, 설명력이 있는 변수일 경우에도 유의성이 없는 것으로 판단하게 되는 오류를 범할 가능성이 커지는 문제가 발생한다. 따라서 변수들 사이의 다중공선성의 존재여부를 점검하기 위해 새로운 회귀분석 식의 설명변수들 사이의 상관관계를 계산하였다.

|  | 스크린수/100 | 네이버 평점 | 인스타그램 해시태그수/1000000 | 원작여부\_시리즈 | 원작여부\_소설/만화 |
| --- | --- | --- | --- | --- | --- |
| 스크린수/100 | 1 |  |  |  |  |
| 네이버 평점 | 0.170525 | 1 |  |  |  |
| 인스타그램 해시태그수  /1000000 | -0.03627 | -0.12119 | 1 |  |  |
| 원작여부\_시리즈 | 0.19023 | -0.11901 | -0.07953 | 1 |  |
| 원작여부\_소설/만화 | 0.076162 | 0.085213 | -0.03735 | -0.32769 | 1 |

[자료6]

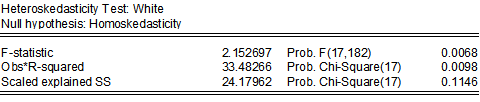
계산 결과 모형전체의 R^2 =0.2662 보다 높은 상관계수를 갖지 않으므로, 다중공선성의 문제가 존재하지 않다고 결론지을 수 있다.

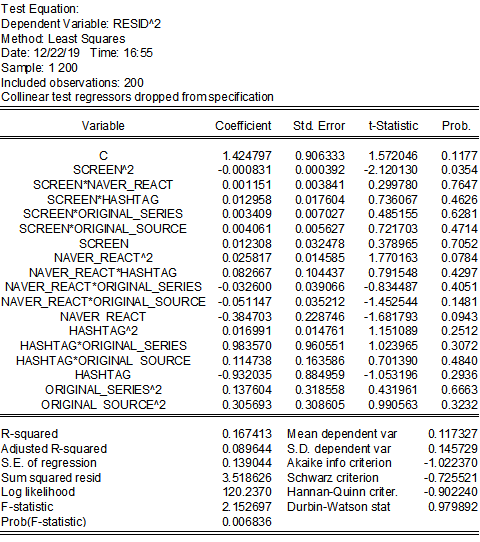
⇒ 다중공선성은 1) 높은 R^2값과 낮은 t-검정치가 동시에 나타날 경우 2) 변수 추가 혹은 제거 시 계수추정치의 불안정성이 나타날 경우 3) X\_i들 사이의 상관계수가 높을 경우 의심해볼 수 있다. 다중공선성이 존재하는 경우, 설명력이 있는 변수일 경우에도 유의성이 없는 것으로 판단하게 되는 오류를 범할 가능성이 커지는 문제가 발생한다. 데이터의 다중공선성의 존재 여부를 점검하기 위해 새로운 회귀 분석식의 설명 변수들 사이의 상관관계를 계산하였다. [자료5]에 따르면, 모형 전체의 R^2=0.2662 보다 높은 상관계수를 갖지 않기 때문에 다중공선성의 문제가 발생하지 않는다고 결론지을 수 있다.

4. 이분산 검정

모형의 오차항에 이분산이 나타나는 경우, 회귀계수 추정량의 분산이 커져, 효율성이 낮아지게 된다. 또한, 표준오차를 이용한 t 검정법, F-검정법 등의 통상적인 검정법으로 잘못된 판정을 야기할 수 있다.

따라서 White 이분산 검정법을 이용해 회귀모형의 이분산 존재 여부를 판단해보고, 이분산이 기)발생할 경우 이를 해결해보고자 하였다.

[자료7]

[자료8]

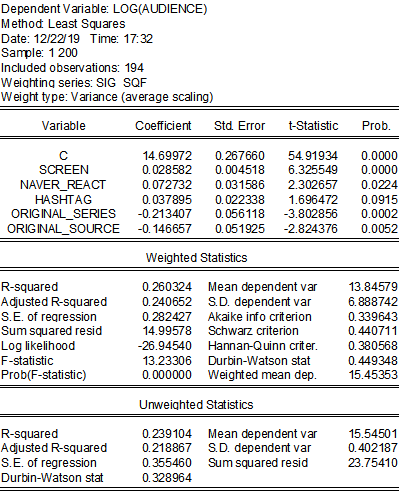
White의 이분산 검정법을 적용해 본 결과, 검정통계량 n\*R^2 = 33.48266 및 이에 대한 카이스퀘어 유의확률은 0.0282가 도출되었다. 5% 유의수준 하에서 검정통계량이 임계치인 카이스퀘어(17;5%) =27.587보다 큰 값을 보이고, 유의확률 또한 0.05보다 작은 값을 보이므로 이분산이 없다는 귀무가설이 기각됨을 알 수 있다.

따라서 설정한 모형에 이분산이 있다는 결과를 얻을 수 있다.

이분산문제를 해결하기 위해 실행가능한 일반최소자승법(FGLS)을 이용하였다.

⇒ 모형의 오차항에 이분산이 나타나는 경우, 회귀계수 추정량의 분산이 커져 추정치의 효율성이 낮아지게 된다. 또한 표준오차를 이용한 t-검정법, F-검정법 등의 통상적인 검정법으로는 잘못된 결론을 내릴 수 있기 때문에 사용을 지양해야 한다. 회귀 모형에 White 이분산 검정법을 적용하면 [자료6]과 [자료7]의 결과를 얻을 수 있다.

먼저, 검정통계량 n∙R^2=33.48266, 이에 대한 카이스퀘어 유의확률은 0.0098의 결과가 도출되었다. 이는 5% 유의 수준 하에서 임계치인 x^2 (17;5%)=27.587 보다 큰 값을 보이며, 유의확률 또한 0.05보다 작은 값은 보이므로 이분산이 없다는 귀무가설은 기각된다. 따라서 회귀 모형에 이분산이 있다는 결과를 얻을 수 있다.

[자료9]

FGLS를 적용한 결과, 조정된 결정계수 값은 0.2407이다. OLS 추정법으로 계산했을 때 도출된 조정된 결정계수 값 = 0.2189 에 비해 높아진 것을 확인할 수 있다. 또한 각 변수들의 p-value 값도 [자료 5]에 비해 모두 더 낮은 값을 보인다.

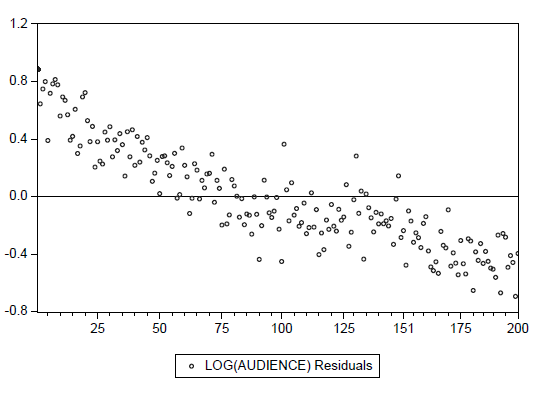
⇒ 이를 해결하기 위해 실행가능한 일반최소자승법(FGLS)을 이용하면 [자료8]을 얻을 수 있다. 그 결과, OLS으로 추정한 [자료5]에 비해 결정계수 와 조정된 결정계수 의 값이 (0.266197, 0.247285)에서 (0.260324, 0.240652) 미세하게 낮아졌다. 하지만 표준오차가 크게 감소하였고, 각 변수의 p-value 또한 낮아졌기 때문에 MSE 기준에 따라서

라는 결론을 내릴 수 있다. 즉 FGLS를 사용한 계수추정치가 더 좋은 추정치가 된다.

5. 자기상관 검정

자기상관 문제가 나타나게 되면, OLS 추정량은 불편성은 유지하지만 효율성을 얻지 못하고, 이에 따라 더 이상 BLUE가 되지 못한다. 또한 OLS 추정량의 분산이 하향 편의를 갖게 되어 t-통계량, F-통계량을 사용할 수 없다.

먼저 잔차그래프를 분석



Durbin Watson 통계량이 2에 수렴하고 있지 않아, 자기상관이 존재할 가능성은 있지만, 이러한 자기상관이 발생한 이유는, 누락변수때문일 확률이 높다.

누락변수를 찾아 자기상관을 해결하기 어려우므로 이는 본 연구의 한계점이라고 할 수 있다.

⇒ 자기상관 문제가 나타나게 되면, OLS 추정량은 불편성을 유지하지만 효율성을 얻지 못하고, 이에 따라 더 이상 BLUE가 되지 못한다. 또한 추정량의 분산이 하향 편의를 갖게 되어 t-통계량, F-통계량을 사용하는 것은 바람직하지 않다. 자기상관은 잔차그래프를 분석하거나 Durbin Watson 통계량을 통해 발생 여부를 알아볼 수 있다. 먼저 [자료9]의 잔차 그래프를 보면, 잔차가 일정한 추세 이루고 있는 것으로 보아 자기상관이 발생한다고 판단할 수 있다. 이는 [자료8]에서 Durbin Watson 통계량이 0.440711로, 2에 가깝게 수렴하지 않아 자기 상관이 존재한다는 결론을 내릴 수 있다. 이는 모형의 독립변수가 3개밖에 되지 않기 때문에 누락변수로 인한 자기상관으로 예상할 수 있다. 하지만 누락변수를 찾아 자기 상관을 해결하기 어렵기 때문에, 이는 분석의 한계점이라고 할 수 있다.

**IV. 주요추정결과 및 예측**

부적합변수 판정을 통해 초기에 설정한 설명변수들 중 released\_movies, rotten 변수를 제외한

log(audience) = c + b1screen +b2naver\_React+b3hashtag+b4original\_series+b5original\_source +ei

의 최종 회귀식을 도출해냈다. 위 회귀식에서 다중공선성의 문제는 나타나지 않았지만, 이분산이 존재하는 것으로 확인되어 FGLS를 적용해 회귀분석한 결과는 [자료 7]과 같다. 이를 통해 도출된 회귀계수를 대입하면 다음과 같은 추정치를 도출할 수 있다.

log(audience)=14.57155+ 0.038673\*Screen + 0.079458\*Naver\_react + 0.026334\*Hashtag -0.334983\*Original\_series-0.227844\*Original\_Source

위 결과를 통해 다음과 같은 사실을 알 수 있다.

1. screen수가 100개 더 증가함에따라 관객수는 약 3.8% 증가한다.
2. naver 평점이 1점 증가하면 관객수는약 7.9%증가한다.
3. 인스타그램 해시태그가 100만개 증가하면 관객수는약 2. 6% 더 증가한다.
4. 영화가 시리즈물의 후속편이면 관객수가 약 33.5% 감소한다.
5. 영화가 소설이나 영화, 만화 등의 원작이 있으면 관객수가 약 22.8% 감소한다.

따라서, 모든변수(가변수 제외)가 한 단위씩 증가하면, 관객수가 3.8%+7.9%+2.6% = 14.3% 증가하는 것을 알 수 있다.

⇒ 초기에 설정한 설명변수들 중 ReleasedMovies, Rotten 변수를 부적합 변수 판정을 통해 제외한 회귀식

은 다중공선성의 문제는 나타나지 않지만, 이분산이 존재하는 것으로 확인되었다. 이를 해결하기 위해 FGLS를 적용하여 회귀분석한 결과는 [자료8]과 같다. 이를 통해 도출된 회귀계수를 대입하면, 다음과 같은 추정치를 도출할 수 있다.

log(Audience) = 14.69972 + 0.037895 Hashtag + 0.072732 NaverReact + 0.028582 Screen -0.213407 Series -0.146657 Source

위 결과를 통해 다음과 같은 사실을 알 수 있다.

1) 인스타그램 해시태그가 100만개 증가하면 관객수는 약 3.8% 증가한다.

2) 네이버 관람객 평점이 1점 증가하면 관객수는 약 7.3% 증가한다.

3) 스크린 수가 100개 더 증가함에 관객수는 약 2.8% 증가한다.

4) 영화가 시리즈물의 후속편이면 관객수가 약 21.3% 감소한다

5) 영화가 소설이나 만화 등의 원작이 있으면 관객수가 약 14.7% 감소한다.

따라서 가변수를 제외한 모든 독립변수가 한 단위씩 증가하면, 관객수가 3.8%+7.3%+2.8% = 13.9% 증가한다는 것을 알 수 있다.

**V. 결론**

1. 결과 해석

영화의 흥행에 미치는 요소로 창조, 배급, 홍보, 경쟁의 영역으로 구분할 수 있다는 선행연구를 참고하여, 로튼 토마토 지수, 스크린수, 인사그램 해시태그 수 및 네이버 관람객 평점, 개봉 달 동시상영 영화개수를 영화 흥행에 영향을 미치는 독립변수로 선정하였다. 그리고 영화의 기대치에 영향을 미쳐 흥행에 영향을 줄 수 있는 원작에 대한 변수를 시리즈에 대한 후속물, 기존 내용을 각색한 경우로 나누어 가변수를 설정하였다.

분석결과, 초기 예상과는 다르게 로튼 토마토 지수와 개봉 달 동시상영 영화개수는 영화의 흥행에 유의미한 영향력을 미치는 요소가 아니라고 판명되었다. 이는 일반 관람객들이 영화를 선정할 때 비평가의 의견을 주의깊게 여기지 않는다고 할 수 있다. 또한, 설정한 변수들 중에서 네이버평점이 영화 흥행에 가장 큰 영향을 미치고, 모든 변수가 증가한다고 해도 관객 수가 15%안팎으로 증가하는 것은 각 변수가 흥행에 미치는 해당영역의 대표값이 되는데에는 한계가 있다고 판단된다. 한편, 시리즈물이나 기존내용을 각색한 경우 구분없이 원본이 존재하면 관객 수가 크게 떨어지는 결과로 보아, 1편 또는 원작의 기대치를 속편이 충족시키지 못했거나, 이미 스토리를 알고 있어서 사람들이 찾아보지 않았을 거라고 추측해볼 수 있다. 결론적으로, 본 연구에 의하면 원작이 없는 새로운 스토리의 영화를 만드는 것이 흥행하는영화를 만드는 시작점이 될 것이고, 개봉 후 높은 네이버 평점을 받을 수 있도록 특히 신경써야 한다고 판단할 수 있다.

⇒ 영화의 흥행에 영향을 미치는 요소로 창조, 배급, 홍보, 경쟁의 영역으로 구분할 수 있다는 선행 연구를 참고하여 로튼 토마토 지수, 스크린 수, 인스타그램 해시태그 수 및 네이버 관람객 평점, 동월 개봉 영화 개수를 독립변수로 선정하였다. 또한 원작의 팬층이 영화의 흥행에 영향을 줄 수 있기 때문에 이 영향력을 시리즈 후속작인 경우와 소설 혹은 만화 원작이 존재하는 경우로 나누어 가변수를 설정하였다.

분석 결과, 초기 예상과는 다르게 로튼 토마토 지수와 개봉 달 동시 개봉 영화 개수는 영화의 흥행에 유의미한 영향력을 미치는 요소가 아니라고 판명되었다. 이는 일반 관람객들이 영화를 선택할 때, 비평가의 의견을 중요시하지 않는다고 해석할 수 있다. 가변수인 원작 팬층 존재 여부는 시리즈 후속작인지 소설 혹은 만화 원작이 존재하는지에 상관없이 관객 수를 크게 떨어뜨리는 결과를 보였다. 이는 1편 또는 원작에 대한 기대치를 속편이 충족시키지 못했거나, 이미 스토리를 알고 있기 때문에 영화관에서 선택하지 않았다고 해석할 수 있다. 연구의 한계점으로는 누락변수에 의한 자기상관을 해결하지 못했다는 점과 인스타그램 해시태그 변수가 영화 흥행을 제대로 설명하지 못한다는 점을 들 수 있다. 설정한 변수들 중 네이버 평점이 영화 흥행에 가장 큰 영향을 미치고, 가변수를 제외한 모든 변수가 1단위씩 증가한다고 해도 관객 수가 15% 안팎으로 증가하는 것 또한 각 변수가 흥행 요소의 각 영역에서 대표성을 띄지 못한다는 데에 한계가 있다고 판단된다.

결론적으로, 본 연구에 의하면 원작이 없는 새로운 영화를 만드는 것이 영화를 흥행시킬 수 있는 시작점이 될 수 있으며, 개봉 후 네이버 관람객 평점이 높은 수준으로 유지되도록 신경써야한다.

**VI. 부록**

1. 참고 사이트

https://[www.kobis.or.kr](http://www.kobis.or.kr/)

[www.rottentomatoes.com](http://www.rottentomatoes.com/)

[www.instagram.com](http://www.instagram.com)

<https://movie.naver.com>

1. 참고 문헌

* 남준우, 이한식, 「계량경제학(이론과 EViews/Excel 활용) 제 3판」
* 2011년, 홍문사 전희국, 『영화 흥행 실적 예측을 위한 빅데이터 전처리 (Big Data Preprocessing for Predicting Box Office Success)』, KIISE, 제20권 제12호(2014.12)
* 영화진흥위원회 영화정책연구원 「2018년 한국 영화산업 결산」, p.9-10
* 영화진흥위원회 영화정책 연구원 「2019년 11월 한국 영화산업 결산」, p.3

1. 사용자료

