

PROCEEDINGS OF THE KOREAN
DATA ANALYSIS
SOCIETY

July 7-8, 2022

at

School of Mathematics, Statistics and Data Science,
Sungshin University, Seoul, Korea
(<https://www.sungshin.ac.kr>)

THE KOREAN DATA ANALYSIS SOCIETY Founded 1998

PROCEEDINGS OF THE KOREAN DATA ANALYSIS SOCIETY July 7, 2022





한국자료분석학회

사무국: (우)31499 충남 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터 AI학과
전화: (041)540-5900 / Fax: (041)540-5908 / 회장 : 한상태 / 총무이사 : 이성건

문서번호 : KDAS-2201000

시행일자 : 2022. 7. 7.

(경유)

수 신 : 전 회원

참 조 :

선결			지	
	일자		시	
접 수	시간		결 재	
	번호			
담당자				

제목: 한국자료분석학회 2022년 하계 학술논문발표대회

1. 귀 회원님의 무궁한 발전을 기원하며, 본 학회의 발전을 위한 회원 여러분의 관심과 노고에 깊이 감사를 드립니다.
2. 본 학회 2022년도 하계 학술논문발표대회를 다음과 같이 개최함을 알려 드리오니 많은 참여 바랍니다.
 - 1) 개최일자 : 2022년 7월 7일(목) ~ 8일(금)
 - 2) 장 소 : 성신여자대학교 (서울)

한국자료분석학회장 한 상태



한국자료분석학회 2022년도 하계 학술논문발표대회 준비위원

조직위원회

공동위원장	한 상 태	호서대학교 빅데이터경영공학부
	강 현 철	호서대학교 빅데이터경영공학부
부 위원 장	이 학 배	연세대학교 응용통계학과
	김 기 환	고려대학교 빅데이터사이언스학부
	홍 재 범	부경대학교 경영학부
	이 영 섭	동국대학교 통계학과
	윤 성 민	부산대학교 경제학부
	박 찬 근	한국해양대학교 데이터사이언스전공
	박 승 열	(주) 케이스탯리서치 회장
	이 기 성	우석대학교 아동복지학과
위 원	이 성 건	성신여자대학교 수리통계데이터사이언스학부
	김 용 대	서울대학교 통계학과
	진 서 훈	고려대학교 빅데이터사이언스학부
	이 동 희	경기대학교 경영학부
자 문 위 원	강 관 중	동아대학교 수학과
	이 희 춘	상지대학교 컴퓨터데이터정보학과
	최 종 후	고려대학교 응용통계학과
	조 완 현	전남대학교 통계학과
	박 희 창	창원대학교 통계학과
	강 창 완	동의대학교 산업경영·빅데이터공학과

운영위원회

공동위원장	조 형 준	고려대학교 통계학과
	최 호 식	서울시립대학교 도시빅데이터융합학과
부위원장	조 장 식	경성대학교 정보통계학과
	김 지 수	가천대학교 간호학과
위 원	김 기 환	고려대학교 응용통계학과
	김 지 희	강원대학교 응급구조학과
	윤 상 복	동의대학교 도시공학과
	이 석 준	상지대학교 경영정보학과
	김 태 훈	경성대학교 경제금융물류학부

한국자료분석학회

2022년도 하계 학술논문발표대회

- 일시 : 2022년 7월 7일 ~ 8일
- 장소 : 성신여자대학교 (서울 성북구 돈암동)
- 공동주최 : 한국자료분석학회, 성신여자대학교 데이터사이언스센터
- 주관 : 성신여자대학교 수리통계데이터사이언스학부
- 후원: 한국연구재단, (주)케이스탯리서치, (주) GDS 컨설팅, 성신여자대학교

7월 7일 목요일

16:00 - 17:00 편집위원회 (수정관 A동 603호)

17:00 - 18:00 이사회 (수정관 A동 603호)

7월 8일 금요일

09:30 - 17:30 포스터발표 (수정관 4층 413호 플립드러닝실 앞)

09:30 - 11:00 튜토리얼 (성신관 110호) 좌장 : 이성건교수(성신여대)

- 제목 : 범주형 자료분석 - 교차분석과 시각화 중심으로
- 강사 : 한상태교수(호서대학교 빅데이터AI학과, 한국자료분석학회장)

11:10 - 12:00 초청 강연 (성신관 110호) 좌장 : 조형준교수(고려대)

- 제목 : 공정한 인공지능: 통계학의 새로운 도전
- 강사 : 김용대교수(서울대학교 통계학과)

12:10 - 12:30 총회 (성신관 110호)

12:30 - 14:00 중식 (교내식당)

14:00 - 15:30 논문발표 A1 - A6 (수정관 4층)

15:30 - 15:40 Coffee Break

15:40 - 17:10 논문발표 B1 - B6 (수정관 4층)

17:30 - 19:30 Banquet (교내식당)

논문발표 A1 (수정관 4층 405호)

경영경제

좌장 : 오명(부경대학교)

-
- 14:00 1. 한국 배출권거래시장 수요측면의 가격결정요인분석 1
| 우주희*(인하대학교 녹색금융대학원 녹색금융전공)
15:30 2. 제주 지역 사례를 통한 재난지원금 지급의 경제적 영향 분석 5
박수경*(강릉원주대학교 경제학과), 강영준(제주연구원)
3. 실물옵션(Real Option)을 활용한 Ohlson 기업가치평가모형 실증분석 연구 9
송유인*, 이용웅(한국외국어대학교 국제금융학과),
조용복(한국외국어대학교, 융합인재대학 ICT&AI Module)
4. 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익성에 관한 연구 13
오명*(부경대학교 글로벌자율전공학부)

논문발표 A2 (수정관 4층 406호)

경제

좌장 : 고선(중앙대학교)

-
- 14:00 1. 식민지기 교회와 인적자본 19
| 김동휘*(중앙대)
15:30 2. 청년연장과 자녀 결혼 21
오운서*(중앙대)
3. 학원 교습시간 규제와 학교 성적분포 25
고선(중앙대), 정유경*(중앙대)

논문발표 A3 (수정관 4층 407호)

간호

좌장 : 강세원(동서대학교)

-
- 14:00 1. 간호대학생의 심리적 독립, 자아존중감에 따른 대인관계 유능성에 관한 연구 ... 27
| 김민영*(경남정보대학교 간호학과)
15:30 2. 중년 여성의 갱년기 자각증상 스트레스 건강증진행위 및 삶의 질 31
박공주*(김해대학교 간호학과)
3. 유방암 진단 여성의 삶의 질 영향요인 35
전은미(배재대학교 간호학과), 강세원*(동서대학교 간호학과)

논문발표 A4 (수정관 4층 408호)

통계

좌장 : 김기환(고려대학교)

14:00	1. PLS 의사결정나무분석	39
15:30	2. CNN 및 Grad-CAM 기반의 케양병 감귤 이미지 분류모델 구축 및 점검	41
	3. CNN 기반 위성 이미지를 활용한 격자단위 인구 추정	43

유문성*, 김정환, 조형준(고려대학교 통계학과)

이동찬*, 변상영, 김기환(고려대학교 국가통계전공)

변상영*, 이동찬, 김기환(고려대학교 국가통계전공)

논문발표 A5 (수정관 4층 409호)

통계

좌장 : 최호식(서울시립대학교)

14:00	1. Self-supervised Classification for Functional Data	45
15:30	2. Forecasting Tail Risk of Time Series with Weighted Scoring Rules	47
	3. Learning Fair Representation via Distributional Contrastive Disentanglement	49
	4. EXoN: EXplainable encoder Network	53

황혜지*(서울시립대학교 도시빅데이터융합학과), 권순선(아주대학교 수학과),

안정연(KAIST, 산업시스템공학과), 박철우(KAIST, 수리과학과),

송경우(서울시립대학교 인공지능학과), 최호식(서울시립대학교 도시빅데이터융합학과)

홍성철*, 전종준(서울시립대학교 통계학과)

오창대*(서울시립대학교 인공지능학과), 원희지(고려대학교 컴퓨터학과),

소준혁(포스텍 컴퓨터학과), 김태로, 김예원(서울시립대학교 인공지능학과),

최호식(서울시립대학교 도시빅데이터융합학과), 송경우(서울시립대학교 인공지능학과)

안승환*(서울시립대학교 통계학과), 최호식(서울시립대학교 도시빅데이터융합학과),

전종준(서울시립대학교 통계학과)

논문발표 A6 (수정관 4층 412호)

통계

좌장 : 김태영(동의대학교)

14:00	1. Neural Temporal-Filter Network: 시계열 예측 정확도 향상을 위한	
	새로운 신경망 구조	55
15:30	2. 순환신경망을 이용한 서술형 강의평가 예측 연구	59
	3. Different Perspectives on Learning Statistics	63

심성현*, 강창완(동의대학교 산업경영빅데이터 공학과)

성진용*, 이제훈, 강창완, 최승배, 심성현(동의대)

김태영*(동의대학교 디그니타스교양교육연구소), 최승배(동의대학교 산업경영·빅데이터공학)

논문발표 B1 (수정관 4층 405호)

경영경제

좌장 : 이은광(서울시립대)

15:40 17:10	1. Predicting the Effect of the Fear and Greed Index for Digital Assets and Stocks on the Trends of Apartment Sales by Artificial Neural Network 67 박현욱*, 배성현 (성균관대학교)
	2. Decomposition of the dividend model: from the perspective of ESG and firm characteristics 71 김수인, 김인중*(홍익대학교 금융보험학과)
	3. SNS의 프라이버시 염려에 관한 연구: 정교화 가능성 모델을 중심으로 75 김유진*, 이형석(충북대학교 경영학부)
	4. 제한폭 확대 영향 연구 79 이은광*, 형남원(서울시립대학교 경제학과)

논문발표 B2 (수정관 4층 406호)

경제

좌장 : 고선(중앙대학교)

15:40 17:10	1. 같은 정보를 각각 그림과 표로 제시하면 소비자의 인지능력이 달라지는가?-가공식품 영양표시제도를 이용한 실험연구 83 허성운*(중앙대), 진현정(중앙대)
	2. 외국인 직접투자(FDI)의 유입이 경제성장에 미치는 영향 - OECD 가입국 대상 87 김민정*(중앙대)
	3. 위기 하의 호텔업의 기업부실위험과 부도모형 예측에 관한 연구 89 권유진*(중앙대), 박성용(중앙대)

논문발표 B3 (수정관 4층 407호)

농업/기상/통계

좌장 : 주용성(동국대학교)

15:40 17:10	1. 지역별 강설이 교통사고에 미치는 영향-강릉, 평창, 춘천을 대상으로 95 김현욱*, 심재관, 임병환(국립기상과학원 예보연구부 재해기상연구소)
	2. 양념채소(대파) 재배농가의 재배규모에 따른 경영성과 순환구조 분석 101 김민현*, 최정, 김도익, 손장환, 마은미(전라남도농업기술원)
	3. 토양 내 오염물질 농도 예측을 위한 베이지안 벌점 스플라인 107 주용성, 유태선*, 나중현, 윤태현, 이정호(동국대학교 통계데이터사이언스학과)

논문발표 B4 (수정관 4층 408호)

통계

좌장 : 정윤서(고려대학교)

15:40	1. Ensemble Variable Selection Using Genetic Algorithm	111
17:10	2. Application of the Conway-Maxwell-Poisson Hidden Markov models for analyzing traffic accident	115
	Hee-Young Kim*, Hyo-Jin Bae(Bigdata Science, Korea University), Su-Hyeon Kim(Department of National Statistics, Korea University)	
	3. SVD 기반 데이터 증강과 기계학습을 통한 웨이퍼 결함 패턴 분류	117
	서정원, 정윤서*(고려대학교 통계학과)	

논문발표 B5 (수정관 4층 409호)

통계

좌장 : 이동희(경기대학교)

15:40	1. 리만 좌표하강 알고리즘을 이용한 조각 측지선 적합	119
17:10	2. Bayesian Approach for Eliminating User Influence and Item Popularity Biases in User-Item Bipartite Networks	123
	정호현*(성신여자대학교 수리통계데이터사이언스학부)	
	3. ODIM: a method to identify inliers via inlier-memorization effect of deep generative models	127
	김동하*(성신여대 수리통계데이터사이언스학부), 황재성(SK Telecom), 김용대(서울대 통계학과)	
	4. 패널자료에서의 은닉마르코프모형을 이용한 사례 분석 연구	131
	이동희*(경기대학교 경영학전공)	

논문발표 B6 (수정관 4층 412호)

조사/미디어/경영

좌장 : 장안식(케이스텍),

15:40	1. 유튜브 악플 탐지를 위한 기계학습: 스택킹 앙상블 모델의 적용을 중심으로 137
	이신행*(중앙대학교 미디어커뮤니케이션학부)
17:10	2. 음원 스트리밍 플랫폼 선택속성의 중요도 분석에 관한 연구 141
	노태호*, 이형석(충북대학교 경영학부)
	3. 지역사회 안전수준을 측정하기 위한 사회안전지수의 개발 145
	정진우*, 장안식(케이스텍)
	4. 사회안전지수로 살펴본 사회안전 유형화와 다층모형의 적용 149
	장안식*, 정진우(케이스텍)

1. 간호사 소진 경감 중재 프로그램의 효과: 체계적 문헌고찰 및 메타분석	153
이미란(광주여자대학교 간호학과), 차지영(이화여자대학교 간호대학)	
2. 간호대학생의 비판적 사고성향, 자기주도적 학습능력과 학업적 자기효능감	155
오윤정, 이은미(남부대학교 간호학과)	
3. 간호대학생의 자기인식, 타인인식과 의사소통능력 간의 관계	159
오윤정, 이은미(남부대학교 간호학과)	
4. 간호대학생의 감성지능수준, 스트레스, 스트레스 대처능력 간 관계	163
이은미, 오윤정(남부대학교 간호학과)	
5. 간호대학생의 학년별 학업소진과 스트레스 대처능력 및 자기효능감 비교	167
이은미, 오윤정(남부대학교 간호학과)	
6. 중소병원간호사의 감성지능과 동료간 돌봄-배려행위가 조직몰입에 미치는영향	171
설진욱(운암한국병원), 박명숙(남부대학교 간호학과)	
7. 코로나 19로 인한 일상생활변화가 한국 청소년의 범불안장애에 미치는 영향	175
김은경(군산대학교 간호학과)	
8. 만성통증 노인환자의 주관적 건강상태에 따른 건강행위간 네트워크 분석	177
박세연(충남대학교 간호대학), 백승희(성신여자대학교 생애복지대학원 건강운동관리학과), 신지은(건양대학교 의과대학 정보의학교실)	
9. 응급실에 내원한 3대 중증응급환자의 양상변화(코로나19 이전과 이후 변화 중심으로)	181
나선경, 오미라(국립중앙의료원), 이원영(중앙대학교 의과대학)	
10. 한국 성인에서 체질량지수와 허리둘레를 기준으로 분류한 비만 유형별 건강행위 특성 분석	185
김선영(건양대학교병원 헬스케어데이터사이언스센터), 이명숙(성신여자대학교 바이오헬스융합학부), 신지은(건양대학교 의과대학 정보의학교실)	

11. 임상간호사의 일과 삶의 균형, 간호업무성고가 행복지수에 미치는 영향	189
남문희(대동대학교), 허현점(가야대학교)	
12. 코로나19로 인한 일상 및 행동 변화 분석	193
차효정(창원대학교 항노화헬스케어학과)	
13. 조건부 Quatember 확률화응답모형	197
이기성(우석대학교 아동사회복지학부), 홍기학(동신대학교 컴퓨터학과), 손창균(동국대학교 빅데이터·응용통계학과), 박근화(한국문화관광연구원), 홍성준(마사회 말산업연구소)	
14. Multinationality, Portfolio Coordination, and Downside Risk	201
송상철(China Technology Business University & Saint Joseph's University)	
15. 외식업 O2O 배달 앱 품질이 소비자의 이용의도에 미치는 영향: TAM과 USM 통합모델	207
이승환(한국항공대학교 경영학과)	
16. 연관성 분석 포트폴리오	211
정현중(한국투자증권, 투자 리서치)	
17. 재정데이터 내의 이상치 탐지를 위한 비지도 방법의 성능 비교	215
소희수, 박승리(마크애니)	
18. 한국 노인의 NISS 손상지수에 영향을 미치는 요인: 퇴원손상심층조사 자료를 이용하여	219
이은미(이화여자대학교 융합보건학과), 남영희(남서울대학교 보건행정학과)	

한국 배출권거래시장 수요측면의 가격결정요인분석

우주희¹, 신동훈²

1. 서론

본 연구는 2015년-2021년의 국내 할당 배출권(KAU)가격을 결정하는 요인에 대해 분석한 실증적 연구이다. 2021년 4월 기준 전세계에서 총 51개의 ETS가 시행중에 있으며 전 세계 배출량의 16.1%(8.73 GtCO₂e)를 통제한다. 국내 배출권 가격을 결정하는 요인으로 본 연구에서는 에너지변수, 경제변수, 기온변수를 고려한다. 에너지 가격이 상승하면 에너지 수요가 하락하여 배출권 가격이 하락할 것이다. 경제 변수의 경우 배출권 가격이 경기순응적일 것으로 예상하며 기온변수는 기온과 에너지사용량이 연관되어 에너지 수요 및 탄소배출량에도 영향을 미쳐 배출권 가격에 영향을 미칠 것이다.

2. 연구방법

2.1. 분석자료

탄소배출권 가격(KAU), 에너지 지표로 석유 가격(KO), 전력 가격, 경제 지표로 일반제조업 경기실사전망지수(KBSI_idus), 발전업종 경기실사전망지수(KBSI_elec), 기온 지표로 난방지수(HDD), 냉방지수(CDD)의 2015.01.12.부터 2021.12.28.까지의 7년간의 일일 시계열 데이터를 사용했다.

2.2. 모델 설정

배출권 가격을 결정하는 주 요인으로는 크게 에너지변수, 거시경제변수, 기후변수, 기타변수(제도적 요인, 대체상품요인, 거래량 등)이 있다. 본 연구에서는 탄소배출권 가격에 영향을 미치는 여러 요인 중 에너지변수, 경제변수, 기온변수만을 고려하여 상관관계를 추정하고자 한다. 본 연구에는 배출권 가격과 변수 간 장단기 동적 관계 분석을 위해 Pesaran and Shin (1999)과 Pesaran et al. (2001)이 제안한 ARDL(autoregressive distributed lag)모형을 사용한다. 모형의 최적 시차를 결정하기 위해 AIC(akaike information criterion)기준이 이용되었다. 추정된 ARDL 모형을 이용해 ARDL-bounds 검정을 시행하여 변수 간 공적분 관계 성립 여부와 장단기 동적 관계를 분석하며 검정에 이용되는 무제약 오차수정모형(UECM, unrestricted ECM)은 다음과 같다.

¹22212 인천광역시 미추홀구 인하로 100, 인하대학교 녹색금융대학원 녹색금융전공 석사과정.

E-mail: hellojh1120@naver.com

²(교신저자) 22212 인천광역시 미추홀구 인하로 100, 인하대학교 글로벌금융학과 부교수.

E-mail: dhshin@inha.ac.kr

$$\begin{aligned} \Delta KAU_t = & \alpha_0 + \sum_{k=1} \alpha_{1k} \Delta KAU_{t-k} + \sum_{i=0} \alpha_{2i} \Delta KO_{t-i} + \sum_{j=0} \alpha_{3j} \Delta KEL_{t-j} + \sum_{l=0} \alpha_{4l} \Delta KBSI_{t_{dus_{t-l}}} \\ & + \sum_{m=0} \alpha_{5m} \Delta KBSI_{t_{elec_{t-m}}} + \sum_{n=0} \alpha_{6n} \Delta HDD_{t-n} + \sum_{r=0} \alpha_{7r} \Delta CDD_{t-r} + \alpha_8 KAU_{t-1} + \alpha_9 KO_{t-1} \\ & + \alpha_{10} KEL_{t-1} + \alpha_{11} KBSI_{t_{dus_{t-1}}} + \alpha_{12} KBSI_{t_{elec_{t-1}}} + \alpha_{13} HDD_{t-1} + \alpha_{14} CDD_{t-1} + \epsilon_t \end{aligned}$$

3. 연구결과

3.1. 단위근 검정

ARDL(Auto Regressive Distributed Lag) 접근법에서 공적분 관계를 확인하기 위해 Pesaran and Shin(1999)과 Pesaran et al.(2001)에 의해 제안된 ARDL-bounds 검정법은 시계열 자료가 I(1)을 따를 때 사용할 수 있는 벡터오차수정모형(Vector Error Correction Model, VECM)에 비해 변수들이 $0 \leq d \leq 1$ 을 따르는 부분적분된 시계열에도 사용이 가능하다. 따라서 각 변수들의 정상성을 확인하기 위한 ADF 검정(Augmented Dickey-Fuller Test)과 PP 검정(Philips-Perron Test)을 시행한다. 단위근 검정 결과 기온변수(HDD, CDD)를 제외한 모든 변수가 I(1)변수임을 확인했다. 기온변수는 부분적분된 변수로 어떤 변수도 I(2) 또는 그 이상으로 적분되어 있지 않았다.

3.2. 공적분 검정 (ARDL-bounds 검정)

$H0$: $\alpha_8 = \alpha_9 = \alpha_{10} = \alpha_{11} = \alpha_{12} = \alpha_{13} = \alpha_{14} = 0$, 변수들은 공적분되어 있지 않다.

$H1$: $\alpha_8 \neq \alpha_9 \neq \alpha_{10} \neq \alpha_{11} \neq \alpha_{12} \neq \alpha_{13} \neq \alpha_{14} \neq 0$, 변수들은 공적분되어 있다.

변수 간 공적분 관계가 없다는 귀무가설은 F-통계량을 이용해 검정할 수 있다. 변수 사이에 장기적 상관관계가 존재하는 경우 F-통계량이 상한값(upper bounds critical value)보다 크다. 하한값(lower bounds critical value)보다 F-통계량이 작으면 귀무가설($H0$)을 기각할 수 없다(Pesaran et al. 1999). 공적분 관계를 보면 KAU15, KAU17, KAU21에서 공적분 관계가 성립하고 있었으며 KAU16, KAU18, KAU19, KAU20은 변수 간 장기 상관관계가 존재하지 않았다.

Table 1. Test results for ARDL model of K-ETS carbon price

Variable	Period of observation	F-statisic
KAU15	2015.10.07-2016.06.23	5.18
KAU16	2016.07.04-2017.06.30	1.56
KAU17	2017.02.06-2018.08.09	3.82
KAU18	2018.09.05-2019.09.30	1.46
KAU19	2019.05.21-2020.09.11	2.07
KAU20	2020.03.31-2021.08.10	1.59
KAU21	2021.07.28-2021.12.28	3.76

Note : upper bounds critical value is 2.27, lower bounds critical value is 3.28 at 5% level.

3.3. 장기관계 추정

Table 2는 각 상품 KAU15-KAU21에 대해 추정된 장기관계계수와 t-값을 나타낸다. 그 결과

KAU15의 가격이 전력가격, 냉방지수, KAU17의 가격의 경우 석유가격, KAU21 가격은 전력가격이 5% 수준에서 통계적으로 유의한 변수임을 확인할 수 있었다. 이는 탄소배출권 가격이 때때로 전력 가격에 영향을 받으며 석유가격, 냉방지수에도 영향을 받음을 암시한다. 특히 초기 시장이었던 KAU15에 비해 KAU19와 KAU21의 시장 성숙도가 더 높다는 가정 하에 탄소배출권 가격은 장기적인 관점에서 전력가격과 양(+)의 방향에서 움직이고 있음을 확인할 수 있다. 이론상으로 에너지가격인 전력 가격이 올라가면 에너지 수요가 감소하여 배출권 가격이 감소해야 하지만 이러한 결과는 우리나라 전기요금체계 특성상 에너지가격이 탄소배출권 가격에 잘 반영되고 있지 않음을 시사한다. 또한, 석유가격 역시 KAU17의 배출권 가격에 대해 양(+)의 상관관계를 보였다. 장기적인 관점에서 경제지표와 기온지표는 KAU15의 냉방지수를 제외하면 모든 상품, 기간에서 통계적 유의성을 가지지 못하는 것으로 나타났다.

Table 2. Estimated long-run coefficients of the K-ETS carbon price.

	KAU15	KAU16	KAU17	KAU18	KAU19	KAU20
KO	30.22	-4.84	18.11*	-1.56	-71.91	-66.15
KEL	-585.82*	7.30	16.38	111.61	637.89*	27.53*
KBSI_ELEC	-344.82	223.13	-43.56	687.12	327.82	444.49
KBSI_IDUS	-391.77	60.23	-51.88	424.99	673.22	-40.07
HDD	212.43	168.15	19.50	57.02	55.74	-84.28
CDD	-1955.43*	106.79	-76.34	561.18	-635.03	-154.75
C	105520.50*	-2205.27	17931.03	-78970.87	-45222.93	26058.46

Note : An asterisk(*) indicates significantce at 5% level.

3.4. 단기관계 추정

단기 관계 역시 상품에 따라 상이한 결과를 보였다. 에너지 가격의 경우 석유 가격은 모든 상품에 대해 어떤 단기 관계도 존재하지 않았다. 전력 가격의 경우 KAU15, KAU20, KAU21에서 5% 수준에서 통계적으로 유의했으며 장기관계와 동일하게 양(+)의 상관을 보였다. 경제 지표의 경우 KAU15, KAU17, KAU19에서는 발전업종의 산업전망지수가, KAU16, KAU20, KAU21에서는 일반업종의 산업전망지수가 유의한 독특한 형태를 확인할 수 있었다. 발전업종의 산업전망지수는 양(+)의 상관을 보인 데 반해 일반업종의 산업전망지수는 예상과 다르게 음(-)의 상관을 보였는데 이는 탄소배출권 거래의 주 수요자가 일반 제조업 회사보다는 발전업종회사이기 때문에 도출된 결과로 보인다. 기온변수의 경우 난방지수는 KAU15, KAU16, KAU21에서 유의했으며 냉방지수는 KAU15, KAU17, KAU18, KAU21에서 유의했다. 두 변수 모두 KAU16을 기점으로 난방변수의 경우 음(-)의 상관에서 양(+)의 상관으로, 냉방변수의 경우 양(+)의 상관에서 음(-)의 상관으로 변하는 패턴을 보였다. 이는 난방지수와 냉방지수가 기온변수를 대표한다기보다는 탄소배출권 제출 기한을 대표하고 있기 때문으로 추론이 가능하며 KAU16 상품을 기준으로 탄소배출권 거래 시기의 패턴이 변화했음을 의미한다.

4. 결론

본 연구는 한국의 배출권거래제 시장을 대상으로 국내 탄소배출권 가격의 수요 측면에서의 장

기 및 단기 상관관계를 실증분석하였다. 장기적으로는 KAU15에서 전력 가격(KEL)만이 예상한 방향으로 움직이고 있었으며 경제지표와 기온지표는 배출권 가격에 직접적인 영향을 미쳤다고 볼 수 없다. 탄소배출권 거래제는 온실가스 배출 감축을 목적으로 탄소배출권을 시장을 통해 거래하는 제도이다. 각 기업은 보유한 탄소배출권 및 배출량에 따라 배출권을 거래하게 되며 본 연구의 결과를 이용하여 향후 시장 예측에 참고할 수 있다.

References

- Banerjee, A., Dolado, J.J., Mestre, R. (1996). Error-correction mechanism tests for cointegration in a single-equation framework, *Journal of Time Series Analysis*, 19, 267 - 283.
- Pesaran, M.H., Shin, Y. (1999). An autoregressive distributed lag modelling approach to cointegration analysis. in: Strom, S. (ed.), Chapter 11, *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press, Cambridge.
- Pesaran, M.H., Shin, Y., Smith, R.J. (2001). Bounds testing approaches to the analysis of level relationships, *J. Appl. Econ.* 16, 289 - 326.

제주 지역 사례를 통한 재난지원금 지급의 경제적 영향 분석*

박수경¹, 강영준²

요 약

본 연구는 신용카드 일별 이용데이터를 이용하여 코로나 19 및 제1차 재난지원금 지급이 제주 지역 소비에 미친 영향을 분석하였다. 재난지원금 지급효과가 제주형과 정부형에 차이가 있었는지를 GMM 방법론을 이용하여 규명하며, VAR 방법론을 통해 동태적 영향력을 추정하였다. 실증 연구 결과 다음의 의미 있는 결과를 도출 할 수 있었다. 첫 번째, 코로나 19 전국 확진자 수의 증가는 제주도민의 카드 이용액을 유의하게 감소시켰다. 팬데믹 상황이 심각해질수록 제주도민의 소비 위축이 컸다는 것을 의미한다. 두 번째로, 제주방문 관광객 증가에 따른 관광객의 카드지출액 증가와 재난지원금 지급은 제주도민의 소득을 증가시켜 제주도민의 카드소비를 증가시킬 수 있음을 보여주었다. 일시(temporary) 소득으로 간주될 수 있는 재난지원금보다 경상 소득으로 간주될 수 있는 관광객 카드지출액 증가가 제주도민의 카드소비에 더 큰 영향을 준 것을 확인할 수 있었다. 또한 제주도민의 카드지출액은 재난지원금 지급 계획이 발표된 시점에 구조적 단절을 보여 실제 가치분 소득이 증가된 시점보다 사전에 증가하여 기대소득에 반응한 것을 보여주었다. 마지막으로 제주형 재난지원금 충격 발생 시 제주도민의 카드소비 증가효과는 즉각적이었던 반면, 정부형의 경우 카드소비 증가효과가 다소 지연되어 나타났다. 이는 정부형과 제주형의 지원금 지급 방법의 차이에서 기인했을 가능성을 보여주는 증거로 해석할 수 있다.

주요용어 : 코로나19, 긴급재난지원금, 제주특별자치도, 카드소비

1. 서론

코로나 19로 인해 피해를 본 가게 및 생산자를 구제하기 위해 각국 정부는 다양한 정책을 마련했으며, 한국 또한 중앙정부와 지방정부가 각 지역의 피해자 구제를 위한 재난지원금을 지급하는 정책을 추진하였다. 최근 재난지원금 지급에 대한 경제적 효과를 분석하고자 하는 많은 논문들이 발표되고 있다. 해외에서는 Baker et al.(2020), Coibion et al.(2020) 등이 있고, 국내에서는 Kim et al.(2020), 한종석 외 2인(2021) 등이 있다. 본 연구는 일별 재난지원금 지급액과 카드지출액 데이터를 이용하여, 제주지역을 대상으로 하는 중앙정부와 지역정부의 재난지원금 지급효과를 GMM을 이용하여 추정하고, 벡터자기회귀모형(Vector-Auto-Regressive model) 방법론을 이용하여 재난지원금 지급이 제주도민 소비에 미치는 시간별 효과를 측정해 본다.

*이 논문은 2021년도 강릉원주대학교 신입교원 연구비 지원에 의하여 연구되었음.

¹25457 강원도 강릉시 죽헌길 7, 국립 강릉원주대학교 경제학과 조교수. E-mail: soo62@gwnu.ac.kr

²63147 제주특별자치도 제주시 아연로 253, 제주연구원 미래산업·관광연구부 책임연구원. E-mail: yjkang@jri.re.kr

2. 연구자료

본 연구에서 사용된 자료는 크게 일별 카드데이터, 일별 재난지원금지금액, 일별 내국인 관광객 수의 세 가지로, 자료의 기간은 2020년 1월 1일부터 8월 31일까지 244일이다. 일별 카드데이터는 제주지역 BC카드 가맹점의 일별 매출액으로 제주도민과 내국인 관광객에 의해 발생한 매출액으로 구분하였다. 일별 재난지원금지금액은 정부가 지급한 제1차 재난지원금과 제주가 지급한 제1차 재난지원금으로 구분하였다. 그밖에도 코로나19확진자수와 일강수량을 추가 통제변수로 이용하였다.

원자료의 기초 통계량 분석결과는 Table 2와 같다. 제주도민 일일 카드이용액(CARD_JEJ)은 평균 2,060백만원으로, 내국인 일일 카드이용액(CARD_LAN) 1,359백만원보다 약 1.5배 높게 나타났다. 그러나 내국인 일일 카드이용액 표준편차는 435백만원으로 제주도민 일일 카드이용액 표준편차인 251백만원보다 약 1.7배 높게 나타났다. 제주형 재난지원금(SUB_JEJ)은 2020년 4월 21일부터 제주도민에게 지급이 시작되었고, 8월말까지 총 412억원이 지급되었다. 정부형 재난지원금(SUB_NAT)은 2020년 5월 8일부터 제주도민에게 지급이 시작되었고, 8월말까지 1,819억원이 지급되었다. 일별 내국인 관광객(TOUR_DOM) 수는 평균 26,518명이고, 최댓값은 49,219명이다.

Table 1. The description of the raw data

series name	Unit	Definition
CARD_JEJ	100M KRW / day	Daily BC card usage amount by Jeju residents in Jeju area
CARD_LAN	100M KRW / day	Daily BC card usage amount by Koreans in Jeju area, excluding Jeju residents
SUB_JEJU	100M KRW / day	Daily amount of first disaster aid paid to Jeju residents by local government
SUB_NAT	100M KRW / day	Daily amount of first disaster aid paid to Jeju residents by central government
TOUR_DOM	person / day	Number of domestic tourists who visited Jeju by day
c19_tot_inc	person / day	Number of daily COVID-19 cases
rain_day	mm / day	Daily precipitation

Table 2. Basic statistics of raw data

series	Unit	Mean	Std.dev	Min	Max	Summation
CARD_JEJ	100M KRW / day	2,060	251	984	3,787	502,586
CARD_LAN	100M KRW / day	1,359	435	712	2,465	331,714
SUB_JEJU	100M KRW / day	169	791	0	8,136	41,241
SUB_NAT	100M KRW / day	746	3,668	0	27,691	181,941
TOUR_DOM	person / day	26,518	9,287	0	49,219	6,470,487
c19_tot_inc	person / day	86	112	2	586	15,724
rain_day	mm / day	4	12	0	114	878

note) sample period : January 1, 2020 ~ August, 31, 2020

3. 모형과 분석결과

본 장에서는 제1차 제주형과 정부형 긴급재난지원금이 제주도민의 카드 소비에 미치는 영향을 GMM(Generalized Method of Moments)을 이용하여 도출한 이후 벡터자기회귀모형(Vector-Auto-Regressive model, 이하 VAR) 방법론을 이용하여 재난지원금 지급 시 제주도민의 카드이용액의 일별 반응을 동태적으로 살펴본다.

$$\text{GMM: } J(\beta, \widehat{W}_T) = \frac{1}{T} u(\beta)' \widehat{W}_T^{-1} u(\beta) \quad (1)$$

$$y_t = X_t \beta + u_t, y_t = c_t^{jeju}, X_t = [1 \quad y_t^{tour} \quad y_{jeju,t}^T \quad y_{nat,t}^T \quad n_t^{c19} \quad p_t^{rain}], \beta: (6 \times 1) \text{ coefficient vector, } u_t: \text{error-term, } E(Z_t u_t(\beta)) = 0$$

$$\text{VAR: } y_t = \nu_0 + \nu_1 t + A_1 y_{t-1} + \dots + A_p y_{t-p} + B_0 y_t^{exo} + u_t \quad (2)$$

$$y_t = (c_t^{jeju}, y_t^{tour}, y_{jeju,t}^T, y_{nat,t}^T)', c_t^{jeju}: \text{CARD_JEJU, } y_t^{tour}: \text{CARD_LAN, } y_{jeju,t}^T: \text{SUB_JEJU, } y_{nat,t}^T: \text{SUB_NAT, } A_i: (4 \times 4) \text{ coefficient matrix, } y_t^{exo} = [n_t^{c19} \quad p_t^{rain}]', n_t^{c19}: \text{c19_tot_inc, } p_t^{rain}: \text{rain_day, } B_0: (4 \times 2) \text{ coefficient matrix, } \nu_0 = (\nu_0^1, \dots, \nu_0^3)': \text{intercept vector, } \nu_1 = (\nu_1^1, \dots, \nu_1^3)': \text{constant trend vector, } u_t = (u_{1t}, \dots, u_{3t})': (3 \times 1) \text{ white noise vector, } E(u_t) = 0, E(u_t u_s') = \Sigma_u, E(u_t u_s') = 0 \text{ for } s \neq t$$

Table 3에서의 GMM 추정결과를 살펴보면, 관광객 카드 지출액, 정부형과 제주형 재난지원금 증가에 대해 통계적으로 유의하게 제주도민의 제주에서의 카드소비가 증가한 것을 확인할 수 있었다. 반면 전국의 코로나 확진자 수 증가나 강수량의 증가는 제주도민의 카드 이용액을 유의미하게 줄일 수 있음을 나타내고 있다.

Table 3. GMM results (adjusted sample: 4/21/2020~7/15/2020)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
y_t^{tour}	0.214	0.034	6.346	0.000
$y_{jeju,t}^T$	0.012	0.004	3.069	0.004
$y_{nat,t}^T$	0.014	0.002	6.278	0.000
n_t^{c19}	-0.050	0.012	-4.149	0.000
p_t^{rain}	-0.004	0.001	-2.921	0.007
constant	5.998	0.252	23.838	0.000

note1) GMM estimation weighting matrix: HAC (Bartlett kernel, Newey-West fixed

note2) Instrument specification: $c_t^{jeju}(-1), y_t^{tour}(-1), n_t^{c19}(-1), c_t^{jeju}(-2), y_t^{tour}(-2), n_t^{c19}(-2), c_t^{jeju}(-3), y_t^{tour}(-3), n_t^{c19}(-3), c_t^{jeju}(-4), y_t^{tour}(-4), n_t^{c19}(-4), c_t^{jeju}(-5), y_t^{tour}(-5), n_t^{c19}(-5), n_t^{c19}, n_t^{c19}(-1)$

note 3) J-statistic (Prob.): 10.854(0.541)

VAR 모형을 이용한 충격반응(Impulse Response) 분석 결과, 제주도민의 카드 소비액은 제1차 제주형 재난지원금과 제1차 정부형 재난지원금에 모두 양의 영향을 받는 것으로 나타났다. 제1차 제주형 재난지원금이 증가하면, 제주도민의 카드 소비액은 충격발생 2일차부터 4일차에 통계적으로 유의하게 증가하였다. 반면 제1차 정부형 재난지원금이 증가 충격 시, 제주도민의 카드 소비액은 충격발생 3일 이후부터 5일까지 3일간 유의하게 증가하는 것으로 나타났다.

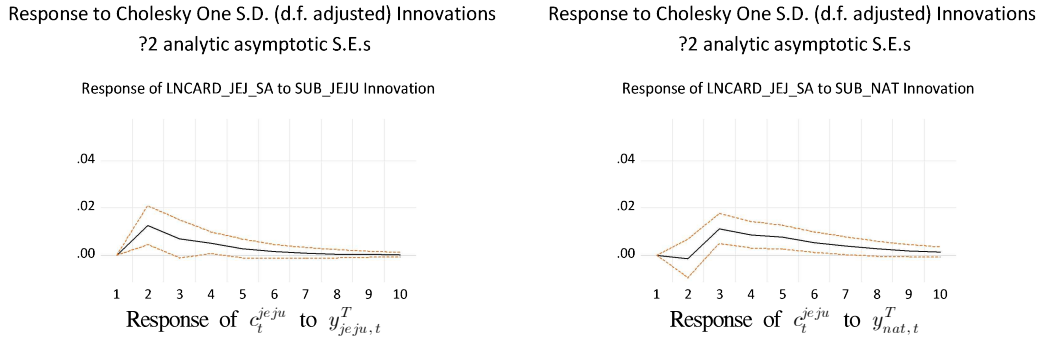


Figure 1. Results of impulse-response analysis

4. 결론

본 연구는 GMM과 VAR 모형을 이용하여 코로나 19 확진자 수 및 재난지원금 지급이 제주도민의 카드 소비에 미친 영향을 분석하였으며 결과는 다음의 함의를 준다. 첫째, 코로나 발생 이후 코로나 19 전국 신규 확진자 수의 증가는 제주도민의 카드 이용액을 유의하게 감소시켰다. 두 번째, 내국인 관광객의 카드이용액 증가와 제1차 제주형 및 정부형 재난지원금 지급은 각각 제주도민의 카드지출액을 통계적으로 유의하게 증가시켰다. 마지막으로 현금으로 지급된 제1차 제주형 재난지원금이 제주도민의 소비증가에 보다 즉각적으로 영향을 미친 반면, 카드 포인트로 지급된 제1차 정부형 재난지원금은 상대적으로 지연되어 효과가 나타나긴 했으나 영향의 지속 기간은 제주형보다 상대적으로 길게 나타났다.

References

- Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. *The Review of Asset Pricing Studies*, 10(4), 834-862.
- Coibion, O., Gorodnichenko, Y., & Weber, M. (2020). *Labor markets during the COVID-19 crisis: A preliminary view (No. w27017)*. National Bureau of economic research.
- Kim, Seonghoon, Kanghyock Koh, and Wonjun Lyou. "Do COVID-19 stimulus payments stimulate the economy? Evidence from card transaction data in South Korea." Evidence from Card Transaction Data in South Korea (September 29, 2020) (2020).
- Han, J. S., Kim, S. B., Chang Y. S.(2021), Economic Impacts of Universal Basic Income, *Journal of Korean Economic Analysis*, 27(1), 163-217.

실물옵션(Real Option) 평가방법을 활용한 올슨모형(Ohlson Model) 실증분석

송유인¹, 이용웅², 조용복³

요 약

올슨(Ohlson)의 초과이익모형은 회계장부 내의 정보를 기반으로 주식의 가치를 설명하고자 하는 기업가치평가모형이다. Ohlson 모형으로 주가를 설명하고자 할 때 초과이익(Residual Income, RI)을 정의하여 미래배당을 예측하고 이에 영향을 미치는 비회계 정보 변수가 필요한데, 많은 연구자들이 비회계 정보 변수를 정의하기 위한 연구를 지속하고 있다. 본 논문에서는 Ohlson 모형에 필요한 비회계 정보 변수를 실물옵션(Real Option) 가치 변수로 채택하고 코스닥 기업을 대상으로 실증 분석하였다. 그 결과 첫째, Ohlson 모형은 현재 초과이익이 발생하고 있는 우량기업과 중견기업에 대해서는 설명력이 높으나, 테크기업 및 초과이익이 발생하지 않는 기업군에 대해서는 설명력이 낮은 것으로 나타났다. 둘째, 실물옵션 변수는 모든 기업군에서 기존 Ohlson 모형에서 설명하지 못하고 있던 주식가치 부분을 설명해 주는 것으로 나타났다. 특히 초과이익이 발생하지 않는 기업군에서 Ohlson 모형의 설명력을 더욱 높여주는 것으로 나타났다.

주요용어 : 올슨모형, 실물옵션, 기업가치평가.

1. 서론

회계정보를 바탕으로 기업가치를 평가하려는 대표적인 연구로 Ohlson(1995)의 연구가 있다. Ohlson 모형은 객관적 정보인 회계정보에 기반하여 기업가치를 평가할 수 있는 모델로, 투자자의 의사결정에 회계정보가 유용한 역할을 하고 있음을 보여준다는 점에서 의의가 있다. Ohlson 모형에 의하면 주식가치는 크게 순자산 장부가액과 미래 초과이익의 예측인데, 이는 미래 이익과 순자산의 예측을 의미한다(김성기, 2003). 하지만, 미래 이익과 순자산을 계산하기 위해 현재 시점의 이익이 사용되기 때문에 아직 수익이 발생하지 않는 벤처기업이나 테크기업의 기업가치 평가에는 한계가 있다.

또한, 미래 초과이익에 영향을 미치는 기타 비회계정보인 미지의 값 v 를 찾기 위한 많은 연구가 수행되고 있다. 본 연구에서는 미지의 값 v 를 미래 초과이익을 보완하려는 값이 아닌 투자자들이 기업에 기대하는 “꿈의 크기”로 가정하고 실증분석하였다. 투자자의 꿈의 크기를 측정하기 위해서는 실물옵션평가방법(ROV: Real Option Valuation)을 도입하였다. ROV 모형은 Stewart Myers(1997)

¹02450 서울특별시 동대문구 이문로 107 한국외국어대학교 대학원 국제경영학과 박사과정.

E-mail: yooinsong@gmail.com

²17035 경기도 용인시 처인구 모현읍 외대로 81 한국외국어대학교 국제금융학과 교수.

E-mail: ywlee@hufs.ac.kr

³(교신저자) 30019 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부 강사.

E-mail: ybcho@korea.ac.kr

가 투자가치를 계산하기 위해 금융옵션에 실물을 적용한 개념으로 1973년 Black and Scholes에 의하여 제시된 OPM(Option Pricing Model)에 바탕을 두고 있으며, Luehrman(1998)은 ROV 모형을 응용하여 현실적으로 대응할 수 있는 7단계 방법론을 제시하였다(주병철, 2011).

2. 연구방법

Ohlson의 모형에 의하면 기업가치는 순자산 장부가치와 미래 초과이익의 현재가치의 합으로 나타낼 수 있다. 미래 초과이익에 대한 예측은 당기순이익과 배당금으로 계산되며, 기타 비회계정보 v 가 초과이익에 영향을 미친다.

$$V_t = BV_t + \alpha_1 RI_t + \alpha_2 v_t \quad (1)$$

많은 기존 실물옵션 연구에서 금융옵션의 변수 대응치로 매출액 등 현금 흐름을 사용하고 있으나, 그 경우 수익이 발생하지 않는 초기 기업의 기업가치에는 적용하기 힘들다. 따라서 본 연구에서는 수익이 발생하지 않는 기업들의 가치평가에 사용 가능하도록 아래와 같은 변수를 대응치로 사용하였다.

$$OV = S \cdot N(d_1) - \frac{K}{e^{r_f \cdot T}} \cdot N(d_2) \quad (2)$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r_f + \frac{\sigma^2}{2}\right)t}{\sigma \sqrt{t}}, \quad d_2 = d_1 - \sigma \sqrt{t}$$

여기서, OV = 옵션벨류

S = 주당가격

K = 주당자본금

σ = 연간일수익률의 표준편차 $\times \sqrt{250}$

r_f = 무위험이자율

t = 5년

본 연구에서는 기업 특성에 따른 Ohlson 모형의 설명력을 검증하고, OV 를 적용한 모형을 실증 분석하기 위해 식(1)~식(2)를 다음과 같은 모형으로 설정하였다.

$$\ln P_{it} = \beta_0 + \beta_1 \ln BPS_{it} + \beta_2 \ln RI_{it} + \epsilon_{it} \quad (\text{모형1})$$

$$\ln P_{it} = \beta_0 + \beta_1 \ln BPS_{it} + \beta_2 \ln RI_{it} + \beta_3 \ln OV_{it} + \epsilon_{it} \quad (\text{모형2})$$

$$\begin{aligned} \ln P_{it} = & \beta_0 + \beta_1 \ln BPS_{it} + \beta_2 \ln RI_{it} + \beta_3 \ln OV_{it} \\ & + \beta_4 \ln DPS_{it} + \beta_5 \ln (B/M)_{it} + \beta_6 \ln Leverage_{it} + \beta_7 \ln Beta_{it} + \epsilon_{it} \end{aligned} \quad (\text{모형3})$$

3. 실증분석 결과

본 실증분석에서는 2007년에서 2021년 코스닥 기업을 대상으로 기업군별 패널데이터를 구성하였다. 일반 기업과 재무제표의 차이를 보이는 금융업, 보험 및 연금업, 금융 및 보험 관련 서비스업 등 금융 관련 업종은 제외하였으며, 자본잠식 중이거나 거래정지 중인 기업, 연도 데이터는 주가에

비정상적인 영향을 줄 수 있어 제외하였다. 최종적으로 12,531개의 기업, 연도 데이터를 표본으로 선정하였다. 패널데이터는 한국거래소의 코스닥시장 공시상장관리 해설 기준 내 업종심사 기준에 따라 우량기업(Bluechip), 중견기업(Midsize), 벤처기업(Venture) 및 테크기업(Tech)으로 분류하였다. 단, 분석기간 중 업종분류가 변경된 기업의 경우 데이터 수집의 한계로 인해 최종시점의 업종분류 기준을 반영하였다. 또한, 수익성이 있는 기업과 없는 기업을 구분하여 RI+ 기업과 RI-기업으로 분류하였으며, 고정효과모형을 사용하여 패널회귀분석하였다.

실증분석 결과 Ohlson 모형은 잉여이익이 발생하고 있는 우량기업, 중견기업, 벤처기업에서는 기업의 주가를 잘 설명하고 있는 것으로 나타났다. 또한, BPS 변수는 기업 분류 및 수익성 여부와 상관없이 주가를 잘 설명하고 있다. 하지만, 미래 성장성이 중요한 테크기업과 RI가 음수인 우량기업 및 벤처기업은 Ohlson 모형이 주가를 잘 설명하고 있지 못한 것으로 나타났으며, RI가 음수인 중견기업의 경우 오히려 RI와 주가의 관계가 음의 상관관계로 나타났다. 즉, Ohlson 모형은 현재 당기순이익이 발생하는 기업들에 대해서는 주가 설명력이 높은 반면, 현재 수익이 나지 않는 기업들에 대해서는 주가를 잘 설명하지 못하는 한계점을 보여주었다.

Table 2는 비회계정보로 투자자들의 기업에 대한 기대감, 즉 “꿈의 크기”를 가정하고, 이를 측정하기 위한 변수로 OV를 추가하여 실증분석 한 결과이다. 그 결과 모든 기업분류에서 OV가 주가를 잘 설명하는 것으로 나타났으며, R-square 또한 개선되었다. 즉, 기존 Ohlson 모형에서 부족한 주가설명력을 보완할 수 있는 변수가 OV일 가능성을 보여주고 있다.

Table 3는 옵션밸류의 설명력을 추가 검증하기 위해 기존 많은 실증연구에서 사용되었던 변수들을 추가하여 강건성 테스트를 수행한 결과이다. 모든 기업분류군에서 OV는 주가와 유의미한 양의 상관관계를 보여주고 있는 것으로 나타났다. 또한, 기업규모와 시가총액의 가치관계를 나타내는 B/M변수도 유의미한 상관관계를 보여주고 있다.

Table 1. Fixed effect coefficients results of Model 1

	RI+				RI-			
	Bluechip	Midsize	Venture	Tech	Bluechip	Midsize	Venture	Tech
BPS	0.6715***	0.5056***	0.3740***	0.6745***	0.7782***	0.5338***	0.4793***	0.3407***
RI	0.1292***	0.0605***	0.0398**	0.0783	0.0177	-0.0229**	-0.0063	-0.0024
R2	0.4610	0.2989	0.1448	0.3719	0.4197	0.2661	0.1989	0.1182

Note 1: BPS is Book value per share. RI is residual income and is based on per share. All variables are taken natural logarithm.

Table 2. Fixed effect coefficients results of Model 2

	RI+				RI-			
	Bluechip	Midsize	Venture	Tech	Bluechip	Midsize	Venture	Tech
BPS	0.4488***	0.2137***	0.1399***	0.1185	0.3249***	0.1048***	0.0531	0.1473***
RI	0.0714***	0.0273***	-0.0047	0.0808	0.0095	-0.0070	0.0031	-0.0031
OV	0.4788***	0.4730***	0.4762***	0.7005***	0.5890***	0.6139***	0.6498***	0.6867***
R2	0.6556	0.5432	0.4241	0.7015	0.7055	0.6133	0.6183	0.6632

Note 1: BPS is Book value per share. RI is residual income and is based on per share. All variables are taken natural logarithm.

Table 3. Robustness test results

	RI+				RI-			
	Bluechip	Midsize	Venture	Tech	Bluechip	Midsize	Venture	Tech
BPS	0.7312***	0.4389***	0.3646***	0.1000	0.7136***	0.5224***	0.5101***	0.2831***
RI	0.0101*	-0.0073	-0.0265**	0.0939*	-0.0012	-0.0127**	-0.0000	0.0059
OV	0.2096***	0.3184***	0.3135***	0.5208***	0.2134***	0.3546***	0.3573***	0.5095***
DPS	0.0098***	0.0024	0.0076	0.0606	-0.0144*	-0.0099*	-0.0087	-
B/M	-1.1132***	-0.8744***	-0.8747***	-1.3745**	-1.0316***	-0.9736***	-1.1455***	-1.6858***
Leverage	-0.0223	-0.0064	0.0027	0.0689	0.0232	0.0366**	0.0257	0.0128
Beta	0.1516***	0.2290***	0.2555***	0.2515	0.0816**	0.1440***	0.1536***	0.1623***
R2	0.8565	0.7627	0.6874	0.7548	0.8897	0.8326	0.8354	0.7904

Note 1: BPS is Book value per share. RI is residual income and is based on per share. DPS is Dividends per share. B/M is Book to Market ratio. Leverage is calculated as Leverage = (Total Liabilities / Total Assets) X 100. Beta is the systematic market risk index. All variables are taken natural logarithm except for Beta.

4. 결론

배당할인모형에 기반한 기업가치평가모형들은 현실에서 널리 사용되고 있으나 미래 수익성을 예측해야 한다는 점에서 많은 불확실한 가정이 개입된다. Ohlson 모형은 미래 수익성을 회계정보로부터 예측하여 주가를 설명하기 때문에 많은 관심을 받아왔으나, 기존 배당할인모형과 마찬가지로 현재 이익이 나지 않는 회사의 미래이익 예측에는 한계점을 보여주고 있다.

본 연구에서는 회계정보가 설명하지 못하는 주가에 대해 투자자들의 기업의 미래에 대한 꿈의 크기가 반영되었을 것을 가정하였다. 꿈의 크기는 실물옵션을 활용하여 측정하였으며, 실증분석결과 실물옵션 변수가 기존 Ohlson 모형의 한계점을 보완하고 설명력을 높이는 새로운 변수가 될 가능성을 확인하였다. Ohlson 모형은 수익이 발생하고 있는 우량기업의 설명력은 높으나 기타 기업군에서는 상대적으로 설명력이 떨어졌다. 실물옵션 변수를 사용한 결과, 모든 기업군에서 R-square가 상승하였으며, 특히 RI+인 벤처기업 테크기업과 RI-기업들에 대한 R-square가 크게 높아졌다.

References

- Ohlson, J. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11, 661-687.
- Kim, S. K. (2003). Equity valuation and disclosure of accounting information. *The Institute of Management Research*, SNU, 37(1), 1-27.
- Ju, B-C. (2011). A valuation of Kosdaq ventures using real option. *Journal of Industrial Economics and Business* 24(2), 1027-1048.
- Luehrman, T. A(1998). Investment opportunities as real option: Getting Started on the numbers. *Harvard Business Review*, Jul 01, 1-14.
- Black, F. and M. Scholes(1973). The pricing of option and corporate liabilities. *Journal of Political Economy*, 81(3), 637-654.

현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익성에 관한 연구

오명¹

요약

본 논문은 일본 주식시장을 대상으로 하여 모멘텀 현상이 나타나는지를 분석하였다. 특히, 현금흐름 뉴스와 할인율 뉴스 상황에서 모멘텀 전략의 수익성을 비교하였다. 추가로 시장 상승기 때 모멘텀 현상이 더 강하게 나타나는지를 확인하였다. 실증분석 결과를 요약하면 다음과 같다. 첫째, 일본 주식시장에서는 모멘텀 수익성이 약하게 나타났다. 둘째, 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익률을 분석한 결과, Celiker, Kayacetin, Kumar, Sonaer(2016)의 결과와 반대로 현금흐름 뉴스보다 할인율 뉴스 상황 하에서 모멘텀 효과가 더 강하게 나타났다. 셋째, 전체 시장이 상승기일 때 모멘텀 전략의 수익성이 더 높게 나타나 Cooper, Gutierrez, Hameed (2004)의 연구 결과를 지지한다.

주요용어 : 모멘텀 수익률, 로그-선형 벡터 자기회귀 모형, 현금흐름 뉴스, 할인율 뉴스, 시장 상황.

1. 서론

모멘텀 현상의 존재에 관하여는 Jegadeesh, Titman(1993)에 의하여 잘 알려져 왔다. Hong, Lim, and Stein(2000)은 미국 주식시장에서는 모멘텀 현상이 재무분석가 수(analyst coverage)가 적은 기업에서 뚜렷하게 나타난다고 보고하였다. Zhang(2006)은 수익률 변동성 및 현금흐름 변동성이 클수록 모멘텀 수익성이 더 높다는 것을 보여 주었다. Cooper, Gutierrez, Hameed(2004)는 미국 주식시장에서는 모멘텀 현상이 시장 상황에 기인한다고 하였다. 즉, 시장이 상승기일 경우 모멘텀 효과가 강하게 작용한다는 것을 보여 주었다. 신흥국 주식시장에서도 모멘텀 현상이 나타난다는 사실이 확인되었다. Chui, Wei, Titman(2000)은 한국, 홍콩, 싱가포르 등 아시아 주식시장을 분석대상으로 하였고 장부가치/시장가치 비율이 낮은 기업, 규모가 작은 기업, 회전율이 높은 기업에서는 모멘텀 전략에 의하여 유의적인 수익을 창출한다는 것을 보여주었다. Jung, Ryu(2021)와 Ohk, Lee(2020)는 그들의 연구결과를 지지한다.

Chui et al.(2010)은 개인주의 성향이 낮은 동아시아와 일본 주식시장에서는 모멘텀 전략의 수익성이 존재하지 않는다고 하였다. Asness, Moskowitz, Pedersen(2013)은 일본 주식시장에서는 모멘텀 요인이 유의적이지 않다고 하였다. 이러한 기존 연구들에서는 일본 주식시장에서는 모멘텀 전략에 의하여 유의적인 수익을 창출할 수 없다는 실증적 근거를 보여 주었다.

Celiker, Kayacetin, Kumar, and Sonaer(2016)는 Campbell(1991)의 로그-선형 벡터 자기회귀 모형을 이용하여 예기치 못한 시장수익률을 현금흐름 뉴스(cash-flow news, CF)와 할인율 뉴스(discount-rate

¹48513 부산광역시 남구 용소로 45, 부경대학교 글로벌자율전공학부 조교수. E-mail: wuming0530@pknu.ac.kr

news, DR)로 분해시켜 뉴스 상황(states) 하에서 모멘텀 효과의 차이에 대해 분석하였다. 그들은 미국 주식시장에서는 과거의 누적 현금흐름 뉴스가 양(+)일 때, 높은 모멘텀 수익성이 존재한다는 것을 보여 주었다. 이에 본 연구는 Campbell(1991)과 Celiker, Kayacetin, Kumar, and Sonaer(2016)에 의하여, 일본 주식시장에서 예기치 못한 시장수익률을 현금흐름 뉴스와 할인율 뉴스로 분해하고 각 뉴스의 상황에 따른 모멘텀 수익성을 측정하여 비교하고자 한다. 그리고 모멘텀 현상의 원인을 일본 주식시장의 특성과 제도적 차이로 설명한다. 추가로 Cooper, Gutierrez, Hameed(2004)처럼 시장 상승기 때 모멘텀 현상이 뚜렷하게 나타나는지를 확인한다. 실증 분석한 결과는 다음과 같다. 첫째, 일본 주식시장에서는 모멘텀 현상이 약하게 나타나고 있음을 알 수 있다. 둘째, 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익률을 분석한 결과, Celiker, Kayacetin, Kumar, Sonaer(2016)의 결과와 반대로 현금흐름 뉴스보다 할인율 뉴스 상황 하에서 모멘텀 효과가 더 강하게 나타났다. 셋째, 전체 시장이 상승기일 때 모멘텀 전략의 수익성이 더 강하게 나타난다는 것을 알 수 있다. 이는 Cooper, Gutierrez, Hameed(2004)의 연구 결과를 지지한다. 즉, 시장 상승기 때에 투자자들은 과대 확신(overconfidence)하는 경향이 보이며, 이로 인해 모멘텀 수익률을 발생시킨다는 것이다.

본 논문은 미국과 시장 특성이 다른 일본 주식시장을 대상으로 하여 현금흐름 뉴스와 할인율 뉴스 상황에서 모멘텀 전략의 수익성을 비교하였고, 실증 결과를 일본 주식시장의 제도적 차이와 시장 특성으로 설명하였다는 데에 있어서 학술적 의의를 가진다.

제2장은 연구에서 사용하는 자료와 방법론이고 제3장은 모멘텀 수익성에 대한 실증 결과를 보여 주고 있다. 마지막 장은 본 연구의 결론 부분이다.

2. 연구자료 및 방법론

2.1. 연구자료

본 논문은 도쿄 증권거래소와 오사카 증권거래소에 상장된 주식들을 연구 대상으로 하여 뉴스 상황 하에서의 모멘텀 수익률을 분석하였다. 표본기간은 1997년 1월부터 2019년 6월까지이다. 연구에 필요한 현금배당이 반영된 수정 주가 및 시가총액 데이터는 DataStream에서 추출하였다. 현금흐름 뉴스와 할인율 뉴스를 구하기 위하여 DataStream에서 시장수익률(TOPIX 지수), 배당수익률 및 1개월 예금금리 데이터를 추출하였다. Table 1은 현금흐름 뉴스와 할인율 뉴스를 구하기 위하여 사용한 월별 시장수익률(TOPIX 지수 수익률, RM), 전체시장의 배당수익률(DY) 및 상대적 수익률(relative rate, RR) 등 내생변수들의 기초 통계량이다.

Table 1. Descriptive statistics

	Mean	Median	Max.	Min.	Stdev.
RM	0.0015	0.0054	0.1297	-0.1898	0.0512
RR	-0.0001	-0.0002	0.0077	-0.0041	0.0015
DY	0.0147	0.0158	0.0285	0.0055	0.0059

2.2. 방법론

2.2.1. 로그-선형 벡터 자기회귀 모형(log-linear VAR model)

Campbell(1991)은 로그-선형 모형에 벡터 자기회귀 모형을 결합하여 예기치 못한 주식 수익률을 현금흐름 뉴스와 할인율 뉴스로 분해하였다. 본 연구에서 사용한 Campbell(1991)의 로그-선형 모형은 다음과 같다.

$$\begin{aligned} r_{t+1} - \mathbb{E}t r_{t+1} &= (\mathbb{E}t+1 - \mathbb{E}t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathbb{E}t+1 - \mathbb{E}t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &= N_{CF,t+1} - N_{DR,t+1} \end{aligned}$$

r_{t+1} 은 t+1월의 시장수익률이고 $\mathbb{E}t$ 는 t월에서의 기대치이다. ρ 는 미래 배당 변화율과 수익률에 대한 할인율을 의미하고 d_{t+1} 은 t+1월 배당의 자연로그 값을 나타낸다. 이 수식은 미래 배당의 기대치가 증가하면 예기치 못한 수익이 발생하지만 미래 수익률의 기대치가 증가하면 예기치 못한 손실이 발생한다는 것을 나타낸다. Campbell(1991)은 미래 배당 기대치의 변동을 현금흐름 뉴스(N_{CF})로, 미래 수익률 기대치의 변동을 할인율 뉴스(N_{DR})로 정의하였다. $e1$ 은 첫 번째 요소가 1이고 나머지가 0인 벡터이다. λ 는 $\rho I(I - \rho I)^{-1}$ 을 나타낸다. 여기서 I 은 단위행렬을 나타내고 I 은 계수 행렬을 나타낸다.

$$\begin{aligned} N_{DR,t+1} &= e1' \lambda u_{t+1} \\ N_{CF,t+1} &= (e1' + e1' \lambda) u_{t+1} \end{aligned}$$

2.2.2. 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익성의 측정

Celiker et al.(2016)에 의하여, 본 연구에서는 현금흐름 뉴스와 할인율 뉴스의 상황을 다음과 같이 정의하였다. 과거 12개월 동안의 누적 현금흐름 뉴스가 양(+)일 경우 t월은 상승_현금흐름 상황(Up_CF states)으로, 음(-)일 경우 t월은 하락_현금흐름 상황(Down_CF states)으로 정의한다. 또한 과거 12개월 동안의 누적 할인율 뉴스가 양(+)일 경우 t월은 상승_할인율 상황(Up_DR states)으로, 음(-)일 경우 t월은 하락_할인율 상황(Down_DR states)으로 정의한다. 본 연구에서는 모멘텀 수익성을 추정하기 위하여 Daniel, Moskowitz(2016)의 방법을 사용하였다. 이에 본 연구에서는 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 포트폴리오 수익률을 측정하여 모멘텀 현상의 원인을 규명한다.

3. 실증분석 결과

3.1. 모멘텀 수익성

Table 2에서는 뉴스 상황을 고려하지 않을 때의 모멘텀 수익률을 보여주고 있다. 즉, Daniel, Moskowitz(2016)에 의하여 구성된 모멘텀 포트폴리오 수익률(WML)의 평균치이다. 괄호 안의 값은 t-값을 나타낸다. 일본 주식시장에서는 모멘텀 현상이 약하게 나타나고 있음을 알 수 있다.

Table 2. Momentum returns

Loser	2	3	4	5	6	7	8	9	Winner	WML
1.459%	1.035%	0.909%	1.002%	0.834%	0.875%	0.943%	1.026%	1.160%	1.922%	0.463%
(6.178)	(5.605)	(5.208)	(6.540)	(5.497)	(5.654)	(6.234)	(6.378)	(6.513)	(8.068)	(1.882)

3.2. 뉴스 상황에 따른 모멘텀 수익성

Table 3은 현금흐름 뉴스와 할인율 뉴스 상황 하에서 일본 주식시장의 모멘텀 수익률을 보여 주고 있다. Panel A는 상승_현금흐름 상황 및 하락_현금흐름 상황 하에서의 모멘텀 수익률 평균치를 나타내고 Panel B는 상승_할인율 상황과 하락_할인율 상황에서의 모멘텀 수익률 평균치를 나타낸다. Panel A에서는 상승_현금흐름 상황에서 모멘텀 전략의 수익성이 존재한다는 것을 보여주고 있다. Panel B에서는 상승_할인율 상황에서 모멘텀 현상이 존재함을 보여준다. Panel A와 B를 비교해 볼 때 Celiker, Kayacetin, Kumar, Sonaer(2016)의 결과와 반대로 현금흐름 뉴스보다 할인율 뉴스 상황 하에서 모멘텀 효과가 더 강하게 나타난다는 것을 확인할 수 있다. 미국의 경우 대부분 기업들이 안정적인 배당정책을 실행하기에 배당의 신호효과가 강하고, 분산된 소유권 구조를 가지고 있어 정보비대칭 현상이 심하다. 또한 부채 비율도 낮기에 투자자들은 할인율 뉴스보다 현금흐름 뉴스에 더 민감할 수 있다. 하지만 일본의 경우 기업들이 대부분 불안정한 배당정책을 실시하고 집중화된 소유권 구조를 가지고 있으며 부채 비율도 높다. 이러한 시장 특성 하에서 일본 주식시장에서 투자자들은 오히려 할인율 뉴스에 더 민감할 수 있고 할인율 뉴스 상황 하에서 모멘텀 현상이 더 강하게 나타날 수 있다.

추가로 본 연구에서는 모멘텀 수익성이 시장 상황에 따라 어떻게 다른지를 분석하였다. 과거 12개월 동안의 누적 시장수익률이 0보다 크면 t월을 상승기로 정의하였다. 분석 결과에 따르면, 시장 상승기 때에 모멘텀 전략의 수익성이 더 크다는 것을 알 수 있다. (Table 4 생략)

Table 3. Momentum returns conditional on cash flow and discount rate news

Panel A. Cash-flow states			
	Loser	Winner	WML
Up_CF states	1.333%	1.899%	0.566%
	(5.334)	(6.090)	(1.755)
Down_CF states	1.670%	1.961%	0.291%
	(3.517)	(5.344)	(0.770)
Up-down	-0.337%	-0.062%	0.275%
	(-0.628)	(-0.128)	(0.554)
Panel B. Discount-rate states			
	Loser	Winner	WML
Up_DR states	0.652%	2.090%	1.438%
	(2.385)	(4.999)	(3.414)
Down_DR states	2.095%	1.790%	-0.305%
	(5.897)	(6.610)	(-1.114)
Up-down	-1.443%	0.301%	1.743%
	(-3.217)	(0.603)	(3.469)

4. 결론

본 연구는 일본 주식시장을 연구대상으로 하고 예기치 못한 시장수익률을 현금흐름 뉴스와 할인율 뉴스로 분해하여 각 뉴스 하에서의 모멘텀 수익성을 비교 분석하였다. 그리고 시장 상승기에 모멘텀 현상이 뚜렷하게 나타나는지를 확인하였다. 실증분석 결과에 따르면 첫째, 일본 주식시장에서는 모멘텀 현상이 약하게 나타났다. 둘째, 현금흐름 뉴스와 할인율 뉴스 상황에 따른 모멘텀 수익률을 분석한 결과, 현금흐름 뉴스보다 할인율 뉴스 상황 하에서 모멘텀 효과가 더 강하게 나타났다. 마지막으로 시장이 상승기일 때 모멘텀 전략의 수익성이 더 강하게 나타났다.

식민지기 교회와 인적자본

김동휘¹

1. 서론

인적자본은 개인과 경제의 발전에 중요한 요소이다. 베버는 프로테스탄트 윤리가 자본주의 정신을 가져왔다고 주장했다. (Weber, 1905) 오늘날 학자들은 천주교보다 더 저축하는 개신교에 의해 자본축적이 일어났다는 그의 주장에 대해서 똑같은 프로이센 자료를 이용하여 사실 여부를 확인했다. 300년 전 마틴 루터에 의해 대거 설립된 초등학교 교육이나 중앙정부의 민족차별정책이 당시 프로이센 국민의 문해율과 경제적 소득의 차이를 초래했다고 주장한다. (Becker, Woessmann, 2009), (Kersting, Wohnsiedler, Wolf, 2020) 비슷한 시기에 중국에서 활동한 개신교 선교사들에 의해 세워진 병원과 학교가 중국의 도시화와 현대 경제 성장에 관여했다(Bai, Kung, 2015), (Chen, Wang, 2022).

이 연구는 1884-1930년의 한국 자료를 이용해서 개신교 선교사 영향을 받은 교회들이 지역의 인적자본에 미친 영향을 2SLS로 분석하였다. 지역에 교회가 한 개 늘어날 때마다 지역의 문해율과 초등학교 취학률이 증가하였다. 교회에서 성경을 읽기 위해 한글을 가르치면서 여성의 문해율을 높였다. 또, 교회는 사람들의 여성 교육에 대한 인식을 개선했다.

선교사들이 활동한 선교기지들로부터의 최단 거리를 이용하여 인과관계를 식별하였다. 이 시기에 들어와서 활동한 교단 중 어느 정도의 규모를 자랑하는 곳은 장로교 네 군데와 감리교 두 군데이다. 1909년 선교사들 간의 암묵적 관할구역을 나눈 후 구역별로 선교사들의 활동 거점인 선교기지를 설립하였다. 선교기지는 선교사와 그 가족들의 생활 공간, 선교사들이 운영하는 학교와 의료 선교의 본거지인 병원 이 세 개가 모인 장소이다. 선교사들은 교회가 밀집한 지역보다는 교단 관할 구역의 지리적인 중심에 선교기지를 설립했다.

교회는 1905-1909년 사이에 폭발적으로 성장했다. 조선 왕정은 외국인과 외국 종교에 대해 적대적이었지만 시간이 지나면서 오히려 중앙정부로부터 선교를 도움받기도 했다. 1907년 평양대부흥을 통해 교회 성장 속도가 정점을 찍었다. 그러나 1910년에 들어서 총독부에 의해 교회의 성장은 느려졌다가 1919년 삼일운동 이후 교회의 2차 확장한 시기가 찾아온다. 1928년 이후부터 교회는 정체 혹은 침체기를 맞이한다.

조선총독부는 1911년 “조선교육령”을 공포하고 1918년 “3면 1교제”에서 1928년 “1면 1교제”로 보통학교를 지역에 늘리기 시작했다. 보통학교는 총독부의 지원과 지역 유지의 기부금으로 세워졌다. 학교 운영비는 지방세와 학교 수업료로 충당되었고 학교 수업료는 계속 증가하였다. (Park, 2012) 남학생 수 증가 속도는 전체 학생 수 증가 속도와 유사하지만, 여학생 수 증가 속도는 매우 낮았다. 전체 학생 수가 가파르게 증가했지만 1930년 취학률은 12.64%으로 늘어나는 교육에 대한 수요에 비해 보통학교 수는 부족했다.

¹06974 서울특별시 동작구 흑석로 84, 중앙대학교 경영경제대학 경제학과 석사과정.

E-mail: twosnail@gmail.com

교회는 지역의 교육에 관한 관심을 늘렸다. 교회에서 가르치는 프로테스탄트 윤리에 의해 개인들이 자기 계발에 관심을 두기 시작했다. 19세기 후반 여성에 대한 폐쇄적인 인식과 도시 지역을 제외하고 분석한 강건성 분석을 통해 확인할 수 있다. 교회의 남성 문해율에 대한 2SLS 계수가 0을 기각하지 못한다는 점과 교회가 여성 문해율과 여성 취학률을 높였다는 것이 그 증거이다.

References

- Weber, M. (1905). Die portestantische Ethik und der 'Geist' des Kapitalismu, *Archiv fur Sozialwissenschaft und Sozialpolitik*
- Becker, S.O., Woessmann, L. (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History, *The Quarterly Journal of Economics*, 136(1), 531-596.
- Kersting, F., Wohnsiedler, L., Wolf, N. (2020). Weber Revisited: The Protestant Ethic and the Spirit of Nationalism, *The Journal of Economic History*, 80(3), 710-745.
- Bai, Y., Kung, J. (2015). Diffusing Knowledge While Spreading God's Message: Protestantism and Economic Prosperity in China, 1840-1920, *Journal of European Economic Association*, 13(4), 669-698.
- Chen, Y., Wang, H., Yan, S. (2022). The long-term effects of Protestant activities in China, *Journal of Comparative Economics*.

부모의 정년이 자녀의 혼인에 미치는 영향

오윤서¹

1. 서론

이 논문에서는 2013년 법적 정년 연령 60세 의무화를 준-자연실험(Quasi-Natural Experiment)으로 응용하여 부모의 정년이 자녀의 결혼에 미치는 영향을 분석한다. 2013년 5월 22일 고령자 고용법 개정으로 만 60세 정년 규정이 권고조항이 아닌 의무 조항으로 변경되었다. 2013년 개정 직후 정년 연장 미 적용 시 처벌의 대상이 되는 것은 아니며 사업장 규모별로 단계적 적용을 위한 계도기간을 두었다.

2. 자료

연구에는 고령화 연구패널 1~8차 자료를 사용하였다. 분석 대상은 정년 제도의 대상이 되는 단일정년제를 채택한 직장에 다니는 법 개정(2013년) 이전에 정년이 60세 미만이었던 정규직 근로자로 한정하였다. 정년제를 채택한 직장에 다니지만 정규직이 아닌 경우, 정규직이지만 정년제를 채택한 직장에 다니지 않는 경우, 정년제 채택 직장 정규직 근로자이지만 정년이 60세 이상인 근로자는 정년 연장의 대상이 되지 않기 때문에 배제하였다. 자녀 관측치의 중복 생성을 막기 위해 부모의 성별을 남성으로 제한하였다. 자녀가 부모에 대한 부가 설문으로 들어가는 자료의 특성상 부모 중 한 사람을 기준으로 표본을 생성하지 않으면 동일 자녀에 대해 관측치가 중복으로 생성될 가능성이 있다. 자녀의 결혼에 대한 분석이기 때문에 혼인이 가능한 연령으로 자녀의 나이를 제한하였다. 자녀의 나이는 실질적으로 결혼이 실질적으로 불가능한 유아, 어린이, 청소년을 배제하기 위해 법적 혼인 가능 연령인 만 18세 미만인 자녀를 제외하고 혼인 적령기를 지난 만 45세 이상의 자녀 또한 표본에서 배제하였다.

자녀를 관측치 단위로 자료를 재구성하고 부모의 정년 및 직업에 관련된 정보를 연결한 결과 자녀 524명으로 분석 대상이 구성되었고, 각 자녀가 1~8차 조사 년도에 대해 선택적으로 응답한 결과 총 관측치 수는 3,065개이고 자료는 불균형 패널(unbalanced panel data)이라고 할 수 있다.

3. 방법

정년 60세 의무화 이전에는 정년이 대부분 55~58세인 상황이고 만 60세 이상의 정년을 갖고 있는 경우에는 정년 60세 의무화가 적용되지 않기 때문에 정년연령이 60세로 의무화가 될 경우, 기존에 정년이 60세 미만이었던 상용직 근로자들은 정년이 늘어나는 처치(Treatment)를 받게 된다. 처

¹06974 대한민국 서울시 동작구 흑석로 84, 중앙대학교 경제학부 대학원생. E-mail: dbstjstar@naver.com

치 여부는 근로자들 자신이 결정할 수 있는 것이 아니라 도입되는 시기에 아직 은퇴를 안 하고 재직 중이어야만 받게 되기 때문에 출생 연도에 따라 무작위로 처치가 된다고 할 수 있다. 따라서 특정 출생 연도를 기점으로 정년 연장 처치를 받게 되는지 여부가 나뉘게 되는데 그 기준은 1956년과 1957년생이다. 2013년 5월 22일 정년 연장 의무화 법 개정이 된 이후 계도기간이 끝나고 단계적 의무 도입이 처음으로 시작되는 해는 2016년이다. 2016년에 1957년생은 만 59세가 되기 때문에 정년 연장 효과를 보는 첫 출생 코호트이다.

1956년생과 1957년생을 기점으로 처치 여부가 나뉘는 방법을 이용하는 방법은 이중차분법이 있다. 이중차분법을 적용하기 위해서는 시점의 외생성이 담보되어야 하는데 법적 정년 연장은 제도의 시행 시점인 의무 도입 시점을 개인들이 정할 수 없기 때문에 처치 시점의 외생성이 확보되었다고 할 수 있다. 만약 1956년 이전 출생자 집단과 1957년 이후 출생자 집단의 자녀들에게 결혼 확률 차이가 있다면 이는 정년 연장의 효과로 정년이 자녀의 결혼에 미치는 영향을 식별할 수 있다.

$$y_{it} = \gamma_0 + \sum \delta_{0t} T_t + \sum \gamma_{1t} Treated_i^* T_t + \gamma_2 X_{it} + \mu_i + \epsilon_{it}$$

위 식은 패널자료 형식을 고려한 일종의 이벤트 분석으로(Event analysis) 확장형 이중차분 모형이다. y_{it} 은 종속변수를 대표하는 변수로 자녀 i 의 조사년도 t 년도의 자녀의 결혼 여부를 나타낸다. T_t 은 조사년도 더미로 시간고정 효과항이다. $Treated_i^* T_t$ 은 이중차분의 효과를 나타내는 교차항이다. 개인 i 의 처치여부를 나타내는 더미와 조사년도 더미변수를 각 조사년도에 대해 교차항을 구한다. X_{it} 은 자녀 i 의 연령, 근로 여부, 자녀 i 의 부모 연령 등의 개인 특성변인들의 벡터인 통제변수를 의미한다. μ_i 은 개인의 고정효과이다.

4. 실증분석 결과

확장형 이중차분 모형에서 기준 연도(Baseline)은 법 개정 직전 연도인 2012년도이다. Table 1의 1열은 전체 표본이다. 남녀 자녀 통합 표본의 경우 이중차분의 효과가 있는 경우는 2020년도이다. 2020년도에 8.3%p 결혼확률의 증가 효과가 나타났다. 남녀를 따로 분석할 경우 결혼확률 변화는 단순형 이중차분과 일관된 결과가 나왔다. Table 1의 2열을 보면 자녀가 여성인 경우 2014~2020년 모두 결혼확률의 증가가 나타났다. 다만 처치의 기준 시점은 계도기간이 종료되는 2016인데 2014년도에도 결혼확률의 증가가 있는 것을 보아 법 개정 이후로 계도기간에 단계적으로 도입했을 때 영향을 받은 것으로 추정된다. Table 1의 3열인 자녀가 남성인 경우에는 2016년과 2018년에 결혼확률 감소 효과가 있었다.

Table 1. The effect of the mandatory retirement age extension on the child's marriage

	(1)	(2)	(3)
Treated*2006	0.0269(0.0472)	0.00535(0.0678)	0.0460(0.0643)
Treated*2008	0.00427(0.0405)	-0.00731(0.0588)	0.0326(0.0542)
Treated*2010	0.0265(0.0387)	0.0456(0.0560)	0.0201(0.0520)
Treated*2014	0.0428(0.0365)	0.132**(0.0550)	-0.0316(0.0475)
Treated*2016	0.0216(0.0373)	0.182*** (0.0563)	-0.106** (0.0484)
Treated*2018	0.0517(0.0388)	0.266*** (0.0591)	-0.112** (0.0499)
Treated*2020	0.0830**	0.278***	-0.0778

Table 1. The effect of the mandatory retirement age extension on the child's marriage(continued)

	(1)	(2)	(3)
Age of the parent	0.00938*(0.00506)	0.0118(0.00777)	0.00762(0.00622)
1 if at most parent enrolled from high school	-0.0266(0.0298)	-0.0307(0.0449)	-0.0378(0.0372)
1 if at least parent enrolled in college	-0.0275(0.0250)	-0.0608(0.0393)	-0.00328(0.0299)
1 if the parent lives in cities	0.0139(0.0272)	-0.00650(0.0403)	0.0194(0.0350)
Net wealth of parent (unit : 10M won)	-0.000151(0.000255)	-0.0000213(0.000348)	-0.000463(0.000369)
1 if parent owns house	0.0239(0.0244)	-0.0298(0.0384)	0.0598**(0.0304)
Age	0.0483*** (0.00354)	0.0609*** (0.00565)	0.0324*** (0.00436)
1 if at most enrolled from high school	0.201*** (0.0418)	0.175*** (0.0634)	0.220*** (0.0543)
1 if at least enrolled in college	-0.0273(0.0328)	-0.0567(0.0470)	0.0429(0.0442)
1 if first born	0.0177(0.0234)	-0.00990(0.0357)	0.0393(0.0286)
1 if working	-0.0902*** (0.0154)	-0.200*** (0.0216)	0.0514** (0.0222)
1 if women	0.120*** (0.0212)		
Constant	1.590*** (0.221)	1.765*** (0.340)	1.244*** (0.271)
Observation	3,065	1,367	1,698
Number of child's id	524	241	287

The dependent variable is an indicator variable having value 1 if the child is married. Standard errors are in the parentheses. *** p<0.01, ** p<0.05, * p<0.1.

5. 결론

본 연구에서는 한국의 결혼 문화에 수반되는 큰 지출을 평탄화(smoothing)하기 위해 정년 전에 자녀를 결혼시키려는 합리적 행동이 실제로 존재하는지 실증분석을 통해 확인하였다. 고령화 연구 패널 자료를 이용하여 부모의 정년이 자녀의 결혼과 어떻게 연관되는지에 대해 분석하였다.

분석 결과 부모의 정년 연장은 자녀의 성별에 따라 다르게 영향을 미쳤다. 남성 자녀의 경우는 부모가 정년 연장이 될 경우 결혼을 늦추었다. 확장형 이중차분으로 분석한 경우 2016년에는 법 개정 직전 연도에 비해 10.6%p 결혼확률이 감소하였고, 2018년의 경우는 11.2%p 감소하였다.

References

Ahn, T. H. (2010). Employment Status of Men and the Family Formation, *Quarterly Journal of labor Policy*, 10(3), 34-64. (in Korean).

Cho, Y., Kim, K., Kim, S., Jeon, S. (2018). Heuristic Lee-Carter model for long-term marital rate prediction, *Journal of the Korean Data Analysis Society*, 20(1), 125-136. (in Korean).

Cho, Y., Kim, K., Jeon, S., Kim, S. (2018). A study on adjustment of transition rates between marital status, *Journal of the Korean Data Analysis Society*, 20(1), 137-147. (in Korean).

Choi, P., Min, I. (2015). Labor force status and employment quality, and marriage event for young workers: applying the discrete-time hazard model, *Korean Journal of Population Studies*, 38(2), 57-83. (in Korean).

Hwang, J. (2016). Housewife, “gold miss,” and equal: the evolution of educated women’s role in Asia and the U.S., *Journal of Population Economics*, 29(2), 529-570.

- Kang, J., Go, S. (2019). Family Background, Economic Activity, and Marriage Decision of the Korean Youth, *Journal of the Korean Data Analysis Society*, 21(6), 3007-3015. (in Korean).
- Kim, K.-W., Choe, S., Jeon, S. (2015). A study on modeling of first marriage, divorce, and remarriage, *Journal of the Korean Data Analysis Society*, 17(5B), 2447-2459. (in Korean).
- Kim, S. J. (2015). Why have Marriages been Delayed?, *Korean Journal of Labor Economics*, 44(2), 57-81. (in Korean).
- Lee, C. (2018). Did pro-natal policy in Korea fail?: a decomposition of fertility change from 2000 to 2016, *Korean Review of Economic Studies*, 66(3), 5-42. (in Korean).
- Lee, J. K., Kim, B. H. (2015). The prospect and meaning of marriage among unmarried women in Korea: heterogeneity across educational attainment groups, *Journal of Korean Women's Studies*, 31(4), 41-85. (in Korean).
- Nam, K. H., Kim, D. I. (2016). Analysis of the Determinants of Women's Marriage and Fertility, *Journal of Women and Economics*, 13(2), 25-52. (in Korean).
- Son, H. L., Song, H. J. (2018). The Empirical Analysis of the Relationship between the Expenditure and the Revenue of the Family Events in South Korea using NaSTaB, *Korean Journal of Public Finance*, 11(4), 147-166. (in Korean).
- Statistics Korea (2022). *2021 Marriage and Divorce Statistics*, Statistics Korea. (in Korean).
- Yi, S. H., Lee, S. H. (2011). A cause Analysis of Low Fertility and Population Aging in Korea : Evidence from Economic Factors in Marriage Determination, *Economic Analysis*, 17(3), 131-166. (in Korean).
- Yeoum, S.-G., Park, C.-S. (2015). The analysis of the structural relationship between self-actualization, gender equality awareness, and marriage value in unmarried women, *Journal of the Korean Data Analysis Society*, 17(3), 1633-1651. (in Korean).
- Yoo, H. J., Hyun, S. M. (2010). The effects of economic resources on marriage-delaying, *Korean Journal of Population Studies*, 33(1), 75-101. (in Korean).

학원 교습시간 규제와 학교 성적분포

정유경¹, 고선²

1. 서론

학원 교습시간 규제는 우리나라 17개 광역시도 교육청에서 조례로 시행중인 제도이다. 해당 정책의 도입 목적은 사교육 과열 해소, 자기주도 학습시간 증대, 성장기 청소년의 수면시간 보장 및 건강 증진에 있다. 교습시간 규제는 조례로써 제정 및 시행되며 규제 시각은 21시부터 24시까지 지역 및 교급에 따라 다르다. 해당 조례는 2007년 도입 이후 현재까지 시행되고 있으며 현재까지도 교육감 선거마다 공약으로 등장할 만큼 사회적 관심사가 높은 정책이다. 그러나 초기부터 해당 정책에 대해 교육계, 학부모 및 학생, 학원 관계자들 간에 찬반 논란이 있어왔다. 찬성 측에서는 정책이 소기의 목적을 달성할 것으로 기대한 반면, 반대 측에서는 주말 반으로 이동하는 등 풍선 효과가 나타날 수 있다는 점, 규제 대상이 아닌 과외로 수요가 전환됨으로써 소득 계층 간 교육 격차가 더욱 심화될 수 있다는 점 등을 우려하였다. 이러한 사회적 질문에 답을 하기 위해서는 실증 분석 연구가 필요하다.

한편 학원과 같은 사교육이 교육성과에 미치는 영향은 언제나 관심사가 되어왔으나, 인과효과로서의 연구 결과를 얻기란 쉽지 않다. 교육성과에 영향을 미치는 요인들을 알아보기 위해 무작위 배정 등과 같은 방법을 이용하기는 현실적으로 어렵다. 하지만 자연실험(Natural Experiments)에 해당하는 상황을 이용한다면 충분히 인과효과를 구할 수 있다. 학원 교습시간 제한 조례는 교육청에서 자발적으로 조례를 제정한 것이 아니라 근거법이 입법됨으로써 의무로써 제정된 점, 처치대상자인 학생이 조례의 결정권자가 아닌 점 등을 고려하였을 때 좋은 자연실험 사례에 해당한다. 따라서 해당 정책이 교육성과에 미친 영향을 인과효과로서 해석할 수 있을 것으로 기대된다. 본 연구는 이러한 점들에 착안하였다.

연구에 사용한 자료는 에듀데이터 2009~2011년 자료이다. 먼저 학교정보공시 데이터로부터 지역과 같은 학교 관련 정보들을 이용하였고 여기에 학업성취도평가(KICE) 및 수능(CSAT)등 학생 개인별 성적 자료를 연결하였다. 아울러 특목고, 특성화 학교 등을 제외한 일반계 고등학교 재학생만을 분석대상으로 한정하였다. 학업성취도평가는 2009년까지 고등학교 1학년을, 2010년부터는 고등학교 2학년을 대상으로 하는 시험이다. 학원시간규제가 교육성과에 미친 영향을 알아보기 위해 2009년 고등학교 입학 코호트를 분석 대상으로 정한다. 다시말해 2009년 고등학교에 입학한 학생들은 2009년, 2010년에 학업성취도 평가를, 2011년에 수능 시험을 치른다. 해당 코호트가 고등학교 재학 중인 2011년에 교습 시간이 단축되었으므로, 이 코호트의 성적 변화를 교육성과로 하여 규제 시각이 개정된 지역과 그렇지 않은 지역의 교육성과를 비교하는 것이 본 연구의 목표다. 따라서

¹06974 대한민국 서울특별시 동작구 흑석로 84, 중앙대학교 경제학과 박사과정. E-mail: yubliss42@naver.com

²(교신저자) 06974 대한민국 서울특별시 동작구 흑석로 84, 중앙대학교 경제학과 부교수.

E-mail: sungo@cau.ac.kr

분석은 2009년 학업성취도 평가 및 2011년 수능을 연결한 표본(이하 KICE2009_CSAT2011), 2010년 학업성취도 평가 및 2011년 수능을 연결한 표본(이하 KICE2010_CSAT2011)으로 두 번 이루어진다. 또한 학업성취도평가와 수능 성적의 점수 척도가 다르기 때문에 모두 Z 점수로 변환하였다.

분석 방법론은 규제 시각 개정 여부가 지역 및 연도별로 다르기 때문에 이종차분법(Difference-in-differences)을 적용하였다. 처치는 2011년 규제시각이 앞당겨 개정된 것으로 정의한다. 이에 따라 처치집단은 2011년 규제 시각이 개정된 지역 즉 광주, 경기, 대구, 인천이며 통제집단은 규제 시각을 단축하지 않은 나머지 지역이다.

추정 결과에 따르면 규제 시각을 앞당겨 개정하면 국어, 수학, 영어 점수가 약 0.01~0.07점 유의하게 상승하는 것을 확인할 수 있다. 다시 말해 규제 시각을 단축한 지역의 학생들의 교육 성과가 좋아졌다. 성적분포를 구간으로 나누어 살펴보면 국어는 하위권 성적이 상승했으며 상위권에 해당하는 백분위수 70~100% 구간에서는 약 0.02점 유의하게 하락하였다. 영어도 비슷한 경향을 보인다. 반면 수학은 최상위권 및 최하위권을 제외한 모든 구간에서 골고루 성적이 상승하였다. 추정 결과는 규제 시각을 앞당기면 하위권 학생들의 교육성과가 좋아지고 있다는 점을 공통적으로 보여준다.

본 연구는 에듀데이터의 학업성취도평가 및 수능 성적 자료를 이용하여 학원교습제한시각이 앞당겨 개정될 경우 국어, 수학, 영어 평균 성적이 대체로 상승하며 이러한 상승은 주로 하위권에서 나타나는 것을 확인하였다. 다시 말해 규제 시각을 개정한 지역에서의 교육성과가 개선되었다. 그러나 본 연구는 몇 가지 한계점이 존재한다. 우선 이러한 성적 상승을 설명할 매커니즘에 대해 생각해 볼 필요가 있다. 가령 규제의 대상이 아닌 과외를 했기 때문에 성적이 상승한 것인지, 혹은 학원 교습 시간이 짧아진 대신 생산성이 개선되었을 수 있다. 향후 에듀데이터에 학생 개인단위 데이터가 추가된다면 본 연구의 한계점을 보완할 수 있을 것으로 기대한다.

간호대학생의 심리적 독립, 자아존중감에 따른 대인관계 유능성 연구

김민영¹

1. 서론

1.1. 연구의 필요성

대학생은 그동안 통제된 환경에서 벗어나 다양한 인관관계를 형성하고 대인관계의 깊이와 폭이 확장되는 변화를 경험하게 된다(An, Chung, 2016). 최근 한국대학교육협의회에서 약 2,600명의 대학생을 대상으로 실시된 ‘2018년 대학생 정신건강 및 대학생활 적응에 관한 실태조사’ 결과, 대학생의 약 73.3%가 대인관계에 있어 어려움을 호소하는 것으로 나타나 대학생의 대인관계 문제가 심각하다는 것을 알 수 있다.

특히 간호대학생은 임상실습, 이론수업, 국가고시 등으로 학업에 집중해야 하는 부담감이 크고 대인관계를 다지는 시간이 상대적으로 부족하여 다양한 심리적 어려움을 겪고 있다(Yoon, 2015). 간호대학생의 경우 성숙한 대인관계 형성은 실습과 수업 뿐 아니라 간호업무를 원활히 수행할 수 있는 바탕이 되므로 대인관계 역량을 갖출 필요가 있다.

대인관계와 관련이 있는 요인 중 심리적 독립(psychological independence)은 부모와의 관계를 유지하면서 부모의 의존에서 벗어나 사고하고 행동하는 것을 의미한다(Josselson, 1988). 부모와 심리적으로 독립하는 과정을 통해 자기감, 자율성, 그리고 정체감이 형성되었을 때 원활한 대인관계 형성이 가능하며 만약 그렇지 못할 경우에 즉, 부모와의 심리적 독립에 어려움을 경험할수록 타인과 원활한 상호작용을 하기 어려워, 대인관계 문제를 상대적으로 더 경험할 가능성이 높다(Lee, Park, 2008; Han, Lee, 2015).

자아존중감은 개인이 스스로를 능력이 있고, 중요하고, 성공적이고, 가치있다고 믿는 정도와 이를 어느 정도 인정하느냐하는 태도를 말한다(Chung, 2009). 대학생들이 함양해야 할 중요한 심리적인 요소인 자아존중감은 진로나 취업 등 다양한 대학생활의 적응에 있어서 개인이 갖는 긍정적인 부분이며(Song, 2019), 간호대학생의 높은 자아존중감은 대상자 간호에 긍정적 영향을 주는 중요한 요소이다.

간호대학생에 있어 대인관계 유능성은 수업과 업무 특성상 무엇보다 중요하다. 임상실습교육 및 간호실무현장에서 접하게 되는 다양한 대상자들과 관계를 형성하게 되므로 대인관계의 폭과 깊이를 키워나가야 하며, 나아가 질 높은 간호를 제공하기 위해서 대인관계 유능성을 발달시켜야 한다(Lee, Chae, 2012).

이에 따라 본 연구에서는 심리적 독립과 자아존중감에 따른 간호대학생의 대인관계 유능성을 알아보고 간호학적 중재 개발의 기초자료로 활용하고자 한다.

1.2. 연구 목적

본 연구는 간호대학생의 심리적 독립, 자아존중감 및 대인관계 유능성을 확인하고 이들의 상관관계를 파악하고자 한다.

2. 연구방법

2.1. 연구설계

본 연구는 간호대학생이 심리적 독립, 자아존중감, 대인관계 유능성과 이들 간의 상관관계를 파악하기 위한 서술적 조사연구이다.

2.2. 연구 대상 및 자료 수집

본 연구는 2021년 5월 K대학교 간호대학의 1,2학년 학생들을 대상으로 연구를 진행하였다. 연구대상자의 수는 G Power 3.1 프로그램을 이용하여 유의수준(α) .05, 검정력($1-\beta$) 80%, 효과의 크기 (f_z) .15, 독립변수 9개로 산출해 본 연구에 부합하는 적정 표본 크기는 128명이 추출되었으며, 탈락률 20%를 고려하여 총 172부의 설문지를 배부하여 100% 회수되었다. 이중 응답이 불충분하거나 불성실한 19부를 제외하고 최종적으로 분석에 포함된 표본수는 153명이었다.

2.3. 연구 도구

2.3.1. 심리적 독립 측정도구

Hoffman(1984)의 심리적 독립척도를 정은희(1992)가 번안하고 수정한 5점 척도를 사용하였으며, Cronbach's α 계수는 .888이었다.

2.3.2. 자아존중감 측정도구

Rosenberg(1965) 자아존중감 척도를 한기백(2013)이 번안한 것을 사용하였으며, Cronbach's α 계수는 .906이었다.

2.3.3. 대인관계 능력 측정도구

Schlein과 Guerney(1971)의 대인관계변화척도를 한국 실정에 맞게 이형득과 문선모(1982)가 번안한 대인관계 변화척도를 사용하였다.

2.4. 자료 분석 방법

본 연구의 자료 분석 방법은 SPSS program 21을 이용해 통계처리 하여 다음과 같이 분석하였다.

- 1) 연구 대상자의 일반적 특성은 실수와 백분율로 처리하여 분석하였다.
- 2) 연구 대상자의 심리적 독립, 자아존중감, 대인관계 유능성은 평균, 표준편차, 최소값, 최대값을 산출하였다.
- 3) 일반적 특성에 따른 대인관계 유능성의 차이를 파악하기 위해 t-test와 ANOVA를 이용하여 분석하였다.

4) 심리적 독립, 자아존중감, 대인관계 유능성의 상관관계를 파악하기 위해 Pearson's correlation coefficient로 분석하였다.

3. 연구결과

3.1. 연구대상자의 심리적 독립, 자아존중감과 대인관계 유능성

연구대상자의 심리적 독립은 평균 2.71 ± 0.49 점이었고, 자아존중감은 5만점에 평균 3.61 ± 0.50 점, 대인관계 유능성은 5점 만점에 평균 3.08 ± 0.35 점으로 나타났다.

3.2. 대상자의 일반적인 특성에 따른 대인관계 유능성

대상자들의 일반적 특성에 따른 대인관계 유능성은 나이($F=3.688$, $p=.007$)를 제외하고는 대인관계 유능성에 차이를 보이지 않았다.

3.3. 대상자의 심리적 독립, 자아존중감과 대인관계 유능성과의 상관관계

대상자의 심리적 독립, 자아존중감과 대인관계 유능성간의 상관관계를 분석한 결과, 심리적 독립과 자아존중감 및 자아존중감과 대인관계 유능성이 통계적으로 유의한 상관관계를 나타내었다. 심리적 독립과 자아존중감($r=.469$, $p<.001$)은 정의 상관관계로 심리적 독립이 높을수록 자아존중감이 높게 나타났으며, 자아존중감과 대인관계 유능성($r=.165$, $p<.041$)은 자아존중감이 높을수록 대인관계 유능성이 높게 나타났다.

4. 논의

본 연구에서 일반적 특성에 따른 대인관계 유능성을 살펴본 결과 연령에 따라 유의한 차이가 있었다. 이는 연령이 높을수록 대인관계에 대한 경험이 상대적으로 많아 대인관계 유능성이 높다고 자각하고 있는 것으로 판단되며, 대인관계형성을 위한 시간적 노력이 대인관계 유능성에 영향을 미치는 결과라 생각된다. 따라서 대인관계 형성을 위해 시간을 투자하여 노력해야하며, 간호대학생을 위한 중재가 마련되어야 할 것이다.

대상자의 심리적 독립, 자아존중감과 대인관계 유능성을 살펴본 결과, 심리적 독립은 5점 척도에서 2.71점으로 대학생을 대상으로 한 연구의 3.85점(Park, 2020)보다 낮게 나타났다. 심리적 독립에 있어 자신에 대한 이해와 믿음은 무엇보다 중요한 부분으로서 자기자신의 신뢰를 높이고 자율적 의사결정을 할 수 있는 능력을 향상시킬 수 있다. 특히 간호대학생의 경우 자신에 대한 신뢰와 믿음은 학교생활 뿐 아니라 간호현장의 업무에 수행에 있어 중요한 부분을 차지하고 있으므로 심리적 독립의 필요성이 강조된다 할 수 있다.

자아존중감은 5점 만점에 3.61점으로 간호대학생을 대상으로 한 연구의 2.95점(Lee et al., 2017)보다 높게 나타났으나 대학생을 대상으로 한 연구의 3.60점(Park et al., 2019)과 비슷하게 나타났다. 이는 최근 간호대학생의 과도한 수업에 대한 부담감과 스트레스로 인해 자아존중감이 다소 감소하고 있다고 판단되며 교육과정에서 간호대학생의 전공에 대한 인식 및 만족도를 높여 정체성 확립

을 위해 노력해야 할 것으로 생각된다. 자아존중감은 높을수록 학생 스스로가 가치 있는 사람으로 여기고 학교 생활에 만족감을 느끼는 것으로 나타나(Oh, 2014) 간호대학생의 자아존중감을 높여 추후 전문간호의 진로를 위한 정체성 확립에도 노력해야 할 것이다.

대인관계 유능성은 5점 척도에서 3.08점으로 간호대학생을 대상으로 실시한 연구에서 3.68점(Kwak, Yi, 2016), 대학생의 3.42점(Seo, Kim, 2020) 보다 낮은 것으로 나타났는데, 이는 간호대학생의 수업, 실습, 국가고시 등의 다양한 학업 내용에 따라 대인관계의 원활한 관계를 요구받고 있으나 스스로 인식하고 있는 대인관계 유능성은 낮게 나타난 것이라 추론된다. 대인관계는 간호대학생뿐 아니라 나아가 전문간호인에게 조직 성과와 관련된 중요한 요인이므로(Seo et al., 2017) 대인관계 유능성을 위한 중재적 전략을 수립할 필요가 있다. 따라서 간호대학생의 과도한 부담감을 줄이고 효율적인 학습 수단을 활용하여 대인관계 유능성을 향상시킬 수 있는 기반을 마련해야 할 것이다.

심리적 독립, 자아존중감, 대인관계 유능성과의 상관관계를 살펴본 결과, 심리적 독립과 자아존중감 및 자아존중감과 대인관계 유능성이 유의한 상관관계를 보였다. 심리적 독립과 자아존중감은 심리적 독립이 높을수록 자아존중감이 높은 것으로 나타나 부모와의 심리적 독립은 현재의 자아존중감에 직접적인 영향을 미치는 부분이라고 한 연구(Jeon, Kwon, 2011)와 같은 결과를 보였다. 이는 부모로부터의 심리적 독립으로 인해 자율성이 증대되고, 개인의 긍정적 자아개념 형성에 영향을 주게 되는 것으로 생각된다. 따라서 인생의 주인은 자기 자신이라는 인식을 강조하여 부모로부터의 심리적 독립을 위한 의지를 촉진함으로써 자아존중감의 향상을 기대할 수 있고, 심리적 독립을 바탕으로 자아존중감을 향상시키기 위한 교육과 프로그램을 학교와 지역사회가 서로 연계하여 실시하면 더 많은 효과를 가져올 것이라 기대된다. 이를 위해 학교에서는 긍정적 대화법 및 발표를 통해 자아존중감을 강조해야 할 것이며(Adler, Fagley, 2005), 사회에서는 대인관계 조화를 위한 프로그램 개발 등을 통해 관심을 기울이고 적극 동참해야 할 것이다.

자아존중감과 대인관계 유능성의 관계에서는 자아존중감이 높을수록 대인관계 유능성도 높은 것으로 나타나 간호대학생을 대상으로 한 Choi, Kim(2019)의 연구결과와 일치했다. 자아존중감은 대학생들의 안정감을 유지시키고 미래에 대한 직업적 확신과 자기발전을 확대할 수 있는 긍정적 요소로서 간호대학생의 추후 간호 전문직에도 긍정적 영향을 줄 수 있으므로 자아존중감의 향상은 꼭 필요한 요소이다(Yu, et al., 2014). 긍정적 자아존중감은 대인관계에 있어 더 적극적으로 대처할 수 있으며(Kendall, Bloomfield, 2005), 그 관계를 잘 유지할 수 있다(Lim, Park, 2013). 따라서 임상실습 등을 통해 다양한 대상자와 의료진들과의 원활한 관계형성이 요구되는 간호대학생은 대인관계형성이 무엇보다 중요하며 이를 위해 자아존중감을 향상시켜야 하며, 의사소통 관련 수업 및 프로그램 등을 통한 노력의 기회가 제공되어야 한다(Lee, Park, 2013).

중년 여성의 갱년기 자각증상, 스트레스, 건강증진행위 및 삶의 질

박공주¹

요 약

본 연구의 목적은 중년 여성의 갱년기 자각증상, 스트레스, 건강증진행위가 삶의 질에 미치는 영향을 확인하기 위하여 시행되었다. 자료수집은 P시에 거주하는 만 40세에서 60세 이하의 여성 200명을 대상으로 2022년 3월1일부터 3월29일까지 자료를 수집하였다. 수집된 자료는 SPSS 26.0 프로그램을 활용하여 기술통계, t-test, ANOVA, 상관관계와 다중회귀로 분석하였다. 연구결과, 삶의 질의 영향요인은 월수입($\beta=-.359, p<.001$), 갱년기 자각증상($\beta=-.168, p=.004$), 스트레스($\beta=-.122, p=.029$), 건강증진행위($\beta=.602, p<.001$)로 나타났으며, 설명력은 58.6%로 나타났다. 본 연구를 통하여 중년 여성의 삶의 질을 향상시키기 위한 다양한 프로그램이 다뤄져야 할 것이다.

주요용어 : 중년 여성, 갱년기, 스트레스, 건강증진행위, 삶의 질.

1. 서론

1.1. 연구의 필요성

국내 인구의 기대 수명 증가로 65세 이상 여성은 2020년 현재 총 여성의 15.7%를 차지하지만 2050년에는 약 39.8%에 도달할 것으로 추정되고 있다(통계청, 2020). 이러한 인구의 고령화 과정에서 여성에게 건강과 관련한 중요한 관심사는 급격한 신체적 변화가 일어나는 중년기를 잘 극복하고 활기찬 노년기를 준비하는 것이라 할 수 있다(이정화 외 2014). 중년기로 진입하는 중요 지표인 폐경은 에스트로겐 분비 감소로 인한 월경의 영구적인 중단으로 정의된다(Mirhaghjou et al., 2016). 대부분 45~55세의 나이에 폐경을 경험하는데(Greenblum et al., 2013) 이 시기를 기준으로 심혈관 질환, 대사내분비질환, 우울증, 수면장애 등의 만성질환이 증가한다(통계청, 2020). 또한 중년 여성은 가족 내에서 핵심적인 역할을 수행하고 있기 때문에 중년 여성의 건강문제는 단순히 개인차원의 문제가 아니라 가족 구성원의 건강에 직·간접적인 영향을 미칠 수 있다(김애경, 2010).

1.2. 연구목적 및 대상

본 연구는 중년 여성의 삶의 질에 영향을 미치는 요인을 파악하고자 함이며 대상자는 P시에 거주하는 만 40세에서 60세 이하의 여성을 연구대상으로 연구를 진행하였으며, 총 230명의 대상자 중 설문에 불성실하게 응답한 30명의 대상자를 제외한 결과 총 200명을 최종 연구대상자로 선정하였다.

2. 연구방법

2.1. 연구도구

1) 갱년기 자각증상

갱년기 자각증상은 Sarrel(1995)이 개발한 Menopause Symptom Index(MENSI)로 한숙희(2004)가 20 문항으로 수정·보완한 것을 본 연구에 맞게 일부 수정한 도구를 구성하여 사용하였다. 5점 Likert 척도로 ‘매우 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점으로 점수가 높을수록 갱년기에 대한 본인의 자각증상이 높음을 의미한다.

2) 스트레스

스트레스는 Tomioka et al(1997)이 개발한 스트레스 척도인 DHS-W에 대해 박상희 등(2001)이 한국인에 맞게 번안한 한국판 일상생활 스트레스 도구(DHS-KW)에 대해 일부 내용을 변경하여 사용하였다. 5점 Likert 척도로 ‘매우 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점으로 점수가 높을수록 스트레스가 높음을 의미한다.

3) 건강증진행위

건강증진행위는 Walker, Sechrist, Pender(1987)가 개발한 HPLP(Health Promotion LifeProfile)를 김금자(2000)가 수정·보완한 도구를 사용하였다. 5점 Likert 척도로 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점으로 점수가 높을수록 건강증진행위의 실천정도가 높음을 의미한다.

4) 삶의 질

삶의 질은 WHO가 개발한 WHOQOL-BREF를 민성길 등(2002)이 번안한 한국판 WHOQOL-BREF를 사용하였다. 5점 Likert 척도로 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점으로 점수가 높을수록 삶의 질이 높음을 의미한다.

2.2 자료분석

본 연구는 SPSS Win 26.0을 이용하여 대상자의 특성과 갱년기 자각증상, 스트레스, 건강증진행위, 삶의 질은 기술통계, t-test, ANOVA로 실시하였다. 대상자의 갱년기 자각증상, 스트레스, 건강증진행위, 삶의 질의 관계는 pearson's correlation coefficients로 분석하였다. 삶의 질에 미치는 요인을 파악하기 위해 multiple regression analysis로 분석하였다.

3. 연구결과

3.1. 대상자의 일반적 특성

본 연구 대상자의 연령은 40대 94명(47%), 50대 74명(37%), 60대 32명(16%)이었으며, 학력은 고등학교졸업 37명(18.55), 대학교졸업 163명(81.5%)으로 나타났고, 결혼상태는 미혼 23명(11.5%), 기혼

177명(88.5%)으로 나타났다. 대상자의 종교는 기독교 118명(59.0%)으로 가장 많았고, 직업은 전문직이 128명(64.0%)으로 가장 많았다. 월수입은 300~400만원 46명(23.0%), 200~300만원 53명(26.5%)이었고, 수면시간은 6~7시간 105명(52.5%)으로 가장 많았으며, 성격유형은 감성적이다 48명(23.0%), 보수적이다 49명(24.5%), 행동적이다 59명(28.0%)순으로 나타났다.

3.2. 일반적 특성에 따른 갱년기 자각증상, 스트레스, 건강증진행위, 삶의 질

대상자의 일반적 특성에 따른 갱년기 자각증상은 결혼상태($t=6.17, p=.01$), 종교 ($F=3.15, p=.04$), 수면시간($F=2.66, p=.04$), 성격유형($F=4.41, p=.02$)로 통계적으로 유의한 차이가 있었다. 사후 분석결과, 종교는 불교집단이 무교집단보다 유의하게 높았고, 수면시간은 6~7시간, 7~8시간 집단이 8시간 이상 집단보다 유의하게 높게 나타났으며, 성격유형은 이성적인 집단이 행동적인 집단보다 유의하게 높았다. 대상자의 일반적 특성에 따른 스트레스는 학력($t=12.36, p=.01$), 종교 ($F=3.13, p=.04$), 월수입($F=4.66, p=.01$), 수면시간($F=6.62, p<.01$), 성격유형($F=3.25, p=.01$)으로 통계적으로 유의한 차이가 있었다. 사후 분석결과, 종교는 불교집단이 무교집단보다 유의하게 높았고, 월수입은 100~200만원, 600만원이상 집단이 400~500만원 집단보다 유의하게 높게 나타났으며, 수면시간은 6~7시간 집단이 7~8시간 집단보다 높았으며, 성격유형은 감성적인 집단이 진취적인 집단보다 유의하게 높았다. 대상자의 일반적 특성에 따른 건강증진행위는 결혼상태($t=7.59, p=.06$), 종교($F=4.62, p=.01$), 월수입 ($F=5.09, p=.01$), 수면시간($F=10.89, p<.01$), 성격유형($F=5.24, p<.01$)이 통계적으로 유의한 차이가 있었다. 사후 분석결과, 종교는 무교집단이 불교집단보다 유의하게 높았고, 월수입은 300~400만원 집단이 유의하게 높았으며, 수면시간은 4~5시간 집단이, 성격유형은 보수적인 집단이 유의하게 높았다. 대상자의 일반적 특성에 따른 삶의 질은 연령($F=8.22, p<.01$), 결혼상태($t=16.42, p<.01$), 월수입 ($F=8.25, p<.01$), 수면시간($F=15.18, p<.01$), 성격유형($F=3.35, p=.01$)이 통계적으로 유의한 차이가 있었다. 사후 분석결과, 연령은 40대, 50대가 유의하게 높았고, 월수입은 100~200만원, 200~300만원, 300~400만원 집단이 600만원이상 집단보다 높았으며, 수면시간은 4~5시간 집단이 유의하게 높았고, 성격유형은 보수적이고 행동적인 집단이 유의하게 높았다.

3.3. 갱년기 자각증상, 스트레스, 건강증진행위, 삶의 질 상관관계

갱년기 자각증상은 스트레스($r=.486, p<.001$)와 유의한 정적 상관관계를 보였고, 건강증진행위 ($r=-.139, p=.050$), 삶의 질($r=-.305, p<.001$)과 유의한 부적 상관관계를 보였다. 스트레스는 삶의 질 ($r=-.238, p<.001$)과 유의한 부적 상관관계를 보였고, 건강증진행위는 삶의 질($r=.606, p<.001$)과 유의한 정적 상관관계를 보였다.

3.4. 삶의 질에 영향을 미치는 요인

갱년기 자각증상, 스트레스, 건강증진행위와 함께 다중회귀분석을 투입한 결과 월수입($\beta=-.359, p<.001$), 갱년기 자각증상($\beta=-.168, p=.004$), 스트레스($\beta=-.122, p=.029$), 건강증진행위($\beta=.602, p<.001$)가 영향을 미쳤으며, 이들 변인에 의한 설명력은 58.6%($F=33.82, p<.001, R^2=.569, Adj-R^2=.586$)였다.

4. 결과

본 연구는 중년여성의 갱년기 자각증상, 스트레스, 건강증진행위가 삶의 질에 미치는 영향을 파악하기 위한 서술적 조사연구이다. 연구 결과 연구대상자의 갱년기 자각증상은 평균 59.07 ± 12.93 점으로 나타났고, 스트레스는 평균 81.70 ± 11.55 점, 건강증진행위는 평균 93.91 ± 18.67 점, 삶의 질은 평균 62.57 ± 11.26 점으로 나타났다. 삶의 질에 미치는 영향요인은 월수입($\beta = -.359, p < .001$), 갱년기 자각증상($\beta = -.168, p = .004$), 스트레스($\beta = -.122, p = .029$), 건강증진행위($\beta = .602, p < .001$)로 나타났으며, 설명력은 58.6%로 나타났다. 본 연구를 통하여 중년 여성들이 건강한 삶의 개선을 위한 다양한 프로그램이 다뤄져야 할 것이다.

References

- 김금자(2000). 중년남성의 건강증진행위 예측모형, *한양대학교 대학원 석사학위 논문*.
- 김애경(2010). 중년 여성의 양생과 건강관련 삶의 질. *여성건강간호학회지*, 16(3), 297-306.
- 김지순, 안숙희(2016). 중년기 이후 여성에서 삶의 질 저하에 영향을 미치는 요인; 폐경 형태와 대사증후군 위험요인을 중심으로. *여성건강간호학회지*, 22(4), 275-286.
- 이정화, 김정혜, 김경덕(2014). 중년 여성의 갱년기증상과 삶의 질의 관계에 대한 인지기능의 매개효과. *한국산학기술학회지*, 15(7), 4439-4448.
- 한숙희(2004). 중년 여성의 폐경증상과 폐경 관리 및 삶의 질. *이화여자대학교 대학원 석사학위 논문*.

유방암 진단 여성의 삶의 질 영향요인

전은미¹, 강세원²

요 약

본 연구는 유방암 진단 여성의 삶의 질에 영향을 미치는 요인을 확인하고자 2019~2020년 국민건강영양조사 자료의 유방암 진단 여성 101명을 대상으로 조사하였다. 수집된 자료는 SPSS/WIN 25 프로그램을 이용하여 실수와 백분율, 평균, 표준편차, 집단 간 비교는 교차분석과 일반선형분석을 이용한 t-test, ANOVA를 이용하였으며, 사후검정은 Bonferroni 방법으로 분석하였다. 유방암 진단 여성의 삶의 질에 영향을 미치는 요인은 일반선형분석을 이용한 다중회귀분석을 이용하였다. 연구결과 인구학적 특성에 따른 삶의 질은 연령($F=6.151$, $p=.001$), 결혼상태($F=13.547$, $p=.001$), 교육수준($F=3.301$, $p=.021$), 경제활동($t=2.400$, $p=.017$)에서 통계적으로 유의한 차이가 있는 것으로 나타났으며, 건강특성에 따른 삶의 질은 건강검진($t=2.000$, $p=.046$)과 우울($t=-2.310$, $p=.022$)에서 통계적으로 유의한 차이가 있었다. 대상자의 삶의 질 측정결과 EQ-5D Index는 1점 만점에 0.93 ± 0.12 점, 주관적 건강상태는 5점 만점에 3.19 ± 0.80 점, 스트레스 인지 정도는 4점 만점에 2.90 ± 0.85 점, 체질량지수는 평균 24.09 ± 3.53 점이었으며, 삶의 질은 주관적 건강상태($r=-0.413$, $p<.001$), 체질량지수($r=-0.247$, $p<.013$)와 통계적으로 유의한 상관관계가 있는 것으로 나타났으며, 삶의 질에 영향을 미치는 변수는 주관적 건강상태로 주관적 건강상태가 좋을수록($t=-2.13$, $p=0.034$) 삶의 질이 높았고 설명력은 40.9%였다.

주요용어 : 유방암, 여성, 삶의 질, 주관적 건강상태.

1. 서론

1.1 연구의 필요성

최근 암검진 및 치료기술의 발전 등으로 암 환자의 생존률이 지속적으로 향상되고 있어 2018년 국가암등록통계 기준 암유병자는 201만명, 5년 상대생존율은 70.3%로 암환자 10명 중 7명은 5년 이상 생존하는 것으로 볼 수 있다(Ministry of health and welfare, 2021). 또한 초기 암 치료를 종료한 암 생존자들이 빠르게 증가하고 있을 뿐 아니라 치명적 질환이던 암은 생존을 향상으로 만성질환으로 인식되고 있어, 암 생존자를 위한 지속적 관리가 필요하다(Baek, Choi, 2018). 특히 유방암은 15~34세 여성에게 두 번째로 발생률이 높은 암으로(Ministry of health and welfare, 2021), 5년 상대 생존율이 91.5%로 꾸준히 증가하고 있다. 유방암은 여성에게 잘 발생하며 수술 후에는 유방의

¹Professor, Department of Nursing, Pai Chai University, 155-40 Baejae-ro, Seo-gu, Daejeon, 35345, Korea.
E-mail: charminggold@pcu.ac.kr

²Associate professor, Department of nursing, Dongseo University, 47 Jurye-ro, Sasang-gu, Busan, 47011, Korea.
E-mail: nursmile@gdsu.dongseo.ac.kr

상실로 좌절감과 수치심을 느끼는 질환으로 여성의 삶의 질에 영향을 미치게 된다(Wildes, Miller, Majors, Ramirez, 2009). 유방암 환자의 삶의 질에 관한 연구를 살펴보면 연령, 소득수준, 신체적, 정서적 기능을 포함하여 통증, 오심 피로 우울증상, 불안 등이 삶의 질에 영향을 미치는 것으로 나타났다(Park, Jun, Lee 2018; Park, Hong, 2017). 유방암 환자들의 삶의 질에 영향을 미치는 변수에 관한 연구는 대부분 부정적인 것에 맞추어져 있지만 부정적인 경험을 오히려 긍정적으로 해석하고 수용하며 성장하는 경우도 있는 것으로 나타났다(Jang et, al. 2013; Park, Jun, Jung, 2013). 이에 본 연구는 최근 국민건강영양조사 자료를 이용하여 국내 유방암 진단 여성의 삶의 질에 영향을 미치는 요인에 대한 과학적 근거를 확인하고 규명함으로써 유방암 진단 여성의 삶의 질 향상을 위한 프로그램 개발을 위한 기초자료를 제공하고자 한다.

2. 연구방법

2.1. 연구설계

유방암 진단 여성의 삶의 질 영양요인을 확인하기 위해 국민건강영양조사 제 8기에 해당하는 2019~2020년 자료를 통합가중치를 적용하여 이차분석한 서술적 조사연구이다.

2.2. 연구대상 및 자료수집

국민건강영양조사 제8기(2019~2020년)조사에 해당하는 전체 대상은 15,469명이었고, 그중에 여성은 8,326명이며, 본 연구에서는 유방암을 진단받은 여성 101명을 대상으로 하였다.

2.3. 연구도구

2.3.1. 인구사회학적 및 건강관련 특성

인구사회학적 특성은 연령, 결혼상태, 교육수준, 경제적 활동, 가구소득을 이용하였고 건강관련 특성으로는 걷기, 근력운동, 유산소운동, 건강검진, 우울, 경구피임약복용, 음주, 흡연, 수면시간을 이용하였다.

2.3.2. 주관적 건강상태, 스트레스 인지, 체질량지수

주관적 건강상태는 “매우 좋음”에서 “매우 나쁨”까지 5점 척도로 구분되었고 점수가 높을수록 주관적인 건강상태는 부정적으로 인지하고 있다. 스트레스 인지는 “대단히 많이 느낀다”에서 “거의 느끼지 않는다”까지 4점 척도로 구분하였고, 점수가 높을수록 스트레스 인지가 낮은 것을 나타낸다.

2.3.3. 삶의 질

삶의 질은 EuroQol-5 Dimension (EQ-5D)을 이용 하였다. EQ-5D는 삶의 질을 측정하는 보편적인 도구이며, 질 가중치를 부여하여 환산된 점수를 나타낸다. 점수는 최저 0점에서 최고 1점까지의 범위를 나타내며, 1점에 가까울수록 삶의 질이 높은 것을 의미한다.

2.4. 자료분석방법

유방암 진단 여성의 인구사회학적, 건강관련 특성과 주관적 건강상태, 우울, 스트레스, 삶의 질은 실수와 백분율, 평균, 표준오차로, 집단 간 비교는 교차분석과 일반선형분석을 이용한 t-test, ANOVA를 이용하였으며, 사후검정은 Bonferroni 방법으로 분석하였다. 유방암 진단 여성의 삶의 질에 영향을 미치는 요인은 일반선형분석을 이용한 다중회귀분석을 이용 하였다.

3. 연구결과

3.1. 연구결과

3.1.1. 대상자의 인구학적 특성에 따른 삶의 질 차이

연구대상자의 인구학적 특성에 따른 삶의 질은 연령($F=6.151$, $p=.001$), 결혼상태($F=13.547$, $p=.001$), 교육수준($F=3.301$, $p=.021$), 경제활동($t=2.400$, $p=.017$)에서 통계적으로 유의한 차이가 있는 것으로 나타났다.

3.1.2. 대상자의 건강특성에 따른 삶의 질 차이

대상자의 건강특성에 따른 삶의 질은 건강검진($t=2.000$, $p=.046$)과 우울($t=-2.310$, $p=.022$)에서 통계적으로 유의한 차이가 나타났다.

3.1.3. 대상자의 삶의 질, 주관적 건강상태, 스트레스 인지정도 및 체질량지수

대상자의 삶의 질 측정결과 EQ-5D Index는 1점 만점에 0.93 ± 0.12 점으로 나타났고, 대상자의 주관적 건강상태는 5점 만점에 3.19 ± 0.80 점, 스트레스 인지정도는 4점 만점에 2.90 ± 0.85 점, 체질량지수는 평균 24.09 ± 3.53 점이었다.

3.1.4. 대상자의 주관적 건강상태, 스트레스 인지 및 체질량지수와 삶의 질 간의 상관관계

삶의 질은 주관적 건강상태($r=-0.413$, $p<.001$), 체질량지수($r=-0.247$, $p<.013$)와 통계적으로 유의한 상관관계가 있는 것으로 나타났다.

3.1.5. 대상자의 삶의 질에 영향을 미치는 요인

유방암 진단 여성의 삶의 질에 영향을 미치는 요인을 확인하기 위하여 나이, 결혼상태, 교육수준, 건강검진여부, 주관적 건강상태, 체질량지수를 독립변수로 투입하였다. 삶의 질에 영향을 미치는 변수는 주관적 건강상태로 주관적 건강상태가 좋을수록($t=-2.13$, $p=0.034$) 삶의 질이 높았고 설명력은 40.9%였다.

References

- Baek, S. Y., & Choi, J. Y. (2018). A comparative study on the characteristics of cancer patients and cancer survivors. *Asian Oncology Nursing*, 18(1), 11-20. <https://doi.org/10.5388/aon.2018.18.1.11>
- Ministry of Health and Welfare, National Central Cancer Registry & National Cancer Center. (2021, April 30). Annual report of cancer statistics in Korea in 2018. Retrieved October 9, 2021, from <https://www.ncc.re.kr/cancerStatsView.ncc?bbsnum=558&searchKey=total&searchValue=&pageNum=1>
- Korea Central Cancer Registry. (2020). National Cancer Center : Report of cancer statistics in Korea in 2018. Sejong : *Ministry of health and Welfare*.
- Jang, S. H., Lee, D. B., Choi, U. J., Lee, K. M., Lee, S. Y. (2013). A Study of the Depression Symptoms and the Quality of Life in Patients with Breast Cancer in a University Hospital. *Korean Journal of Psychosomatic Medicine*. 1.21(1), 11-17.
- Park, K., Jun, J. S., Jung, S. Y. (2013). The Effects of Korean Mindfulness Based Stress Reduction Program on the Psychological Symptoms, Sleep and Quality of Life, in Korea Breast Cancer Patients. *The Korean Journal of Stress Research*. 1.21(3), 249-262.
- Park, L., Jun, S., Lee, J. A.. (2018). Quality of Life and Associated Factor among Cancer Survivors in Korea. *Korean Journal of Health Promotion* 18(1), 7-14. DOI: 10.15384/kjhp.2018.18.1.7
- Park, J. A., Hong, J. Y. . (2017). Factors Influencing Quality of life in Adult Cancer Patients: The Sixth Korea National Health and Nutrition Examination Survey (KNHANES IV-2), 2014. *Journal of the Korea Academia-Industrial cooperation Society*, 18(5), 382-90.DOI: 10.5762/KAIS.2017.18.5.382

PLS 의사결정나무 분석

Wenxing Yu¹, 김정환², 조형준³

요약

현대 생물학 자료 분석에서 자료가 많은 수의 다중공선성 예측 변수들과 작은 표본 크기를 가진 다변량 반응변수로 구성될 때 일반적인 회귀분석방법으로는 여러 가지 문제가 발생한다. 이러한 경우, 예측변수와 반응변수 사이의 회귀모형 적합을 위해 일반최소제곱(OLS) 방법 대신 부분최소제곱(PLS) 방법으로 추정한다. 그러나 자료가 복잡한 비선형 관계가 있을 경우 PLS 회귀모형에 의한 추론 및 예측하기에 어려움이 있다. 우리는 PLS와 의사결정나무의 장점을 결합하여 자료의 복잡한 비선형 관계를 감지하고 정확하게 예측하기 위해 PLS의사결정나무 분석 방법을 제안한다. 우리가 제안한 PLS의사결정나무는 자료가 많은 수의 다중공선성 예측변수들을 가지고 작은 표본 크기를 가진 일변량 및 다변량 반응변수 모두에 적용될 수 있다. 제안된 분석 방법의 성능은 각각 일변량 및 다변량 반응변수의 모의실험 자료와 실제 자료로 확인할 수 있다.

주요용어 : 다중공선성, 복잡 비선형, 부분최소제곱(PLS), 의사결정나무.

¹02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 박사과정. E-mail: wenxing1004@korea.ac.kr

²02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 박사과정. E-mail: kjhsjy0215@korea.ac.kr

³02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 교수. E-mail: hj4cho@korea.ac.kr

CNN 및 Grad-CAM 기반의 궤양병 감귤 이미지 분류모델 구축 및 점검

이동찬¹, 변상영², 김기환³

요약

하드웨어의 성능 및 정보처리 기술이 급격히 발전하면서 비정형 데이터의 처리 및 가치 창출에 관한 관심이 증가하고 있다. 이를 위한 다양한 인공지능 아키텍처들이 개발되고 있으며, 모델의 의사결정 분기점이 기하급수적으로 늘어나면서 큰 성능의 개선이 이루어지고 있다. 그러나 복잡한 모델 구조는 연구자의 결과 해석 용이성을 저해하는 주요한 원인이 되며, 모델 성능의 발전 속도와는 달리 설명 능력에 대해서는 진척이 더딘 실정이다. 설명 가능한 인공지능, 이하 XAI(eXplainable Artificial Intelligence)는 위와 같은 문제를 해결하기 위해 등장하였으며, 모델의 블랙박스를 이해 가능한 수준으로 분해하여 해석 가능성 및 신뢰도 제고에 도움을 준다. 본 연구에서는 CNN(Convolutional Neural Network) 모델을 사용하여 궤양병 감귤 이미지 분류 문제에 접근하였으며, 최종적으로 설계한 모델은 약 97% 수준의 정확도를 보였다. 이후 모델의 신뢰도 제고 및 개선 방향 판단을 위해 XAI 기법 중 하나인 Grad-CAM(Gradient-weighted Class Activation Mapping)을 적용하였으며, 이를 통해 구축한 모델이 최종적인 판단을 내리는데 주요한 역할을 한 이미지의 특정 영역을 파악하는 과정을 진행하였다. 점검 결과 이미지 외곽의 형태가 객체와 구분이 되지 않아 영향을 크게 받는 경우 및 특정 객체의 고유한 형태가 주요한 오분류 원인으로 감지되었다.

주요용어 : CNN, Grad-CAM, XAI.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 국가통계전공 박사과정.

E-mail: tozjs72@korea.ac.kr

²30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 국가통계전공 박사과정.

E-mail: wooaaaaa@korea.ac.kr

³30019 세종특별자치시 세종로 2511, 고려대학교 경제통계학부 국가통계전공 교수. E-mail: korpen@korea.ac.kr

CNN 기반 위성 이미지를 활용한 격자단위 인구추정

변상영¹, 이동찬², 김기환³

요 약

인구는 경제성장에 큰 영향을 미치기 때문에, 국가의 인구를 파악하는 것은 매우 중요한 일이며, 정확한 인구를 파악하기 위해 다양한 인구추정 방법이 연구되고 있다. 최근 인구추정을 위해 항공사진이나 위성 이미지를 활용한 연구가 활발히 진행되고 있다. 특히 야간불빛을 관측한 위성 이미지는 특정 지역의 실측자료 없이도 도시 개발과 경제성장 등을 파악할 수 있는 장점이 있어 기초 데이터가 부족한 저개발국 및 개발 도상국의 연구에 유용하게 활용되고 있다. 북한은 폐쇄적인 특성으로 인해 기초 데이터가 매우 부족하다. 통일연구원(2020)에 따르면, 북한 인구센서스는 1993년과 2008년 두 차례 실시하였지만, 두 센서스 모두 상이한 두 가지 인구수가 제공되어 데이터 왜곡의 가능성으로 신뢰성 측면에서 의문이 제기되고 있다. 북한의 인구를 정확히 파악하는 것은 한반도 통일정책의 수립에 매우 중요한 의미를 지닌다. 이에 본 연구에서는 북한의 인구를 추정하기 위해 CNN 기반 주간 위성 이미지를 활용한 격자 단위 인구추정 방법론을 제안하였다. 기존의 위성 이미지를 활용한 북한 인구추정에서는 위성 이미지를 보조자료로 활용하였으나, 본 연구에서는 위성 이미지만을 활용하여 인구를 추정하였다. 주간 위성 이미지는 Land Viewer에서 제공하는 Landsat 8 위성이 관측한 3, 4월의 위성 이미지를 활용하였다. 모델의 평가는 예측값과 실제값의 차이를 나타내는 평균 제곱근 오차(RMSE)를 이용하여 평가하였다. 대한민국을 기반으로 한 모델의 적합 평가는 만족스러운 결과가 나타났으나, 북한 대상의 인구추정 결과는 위성 이미지가 가지는 한계점으로 인해 부족한 부분이 있는 것으로 나타났다. 위성 이미지를 개선한다면 보다 개선된 결과가 나올 것으로 기대된다.

주요용어 : 북한 센서스, 위성 이미지, CNN, 격자인구, Land Viewer.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 박사과정.

E-mail: wooaaaaa@korea.ac.kr

²30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 박사과정.

E-mail: tojs72@korea.ac.kr

³30019 세종특별자치시 세종로 2511, 고려대학교 경제통계학부 국가통계전공 교수. E-mail: korpen@korea.ac.kr

Self-supervised classification for functional data^{*}

황혜지¹, 권순선², 안정연³, 박철우⁴, 송경우⁵, 최호식⁶

요약

보행데이터는 뇌성마비 환자의 치료에 따른 재활의 정상성을 평가하기 위해 수집한 자료이다. 함수형(functional) 자료구조를 가진 보행데이터를 학습시, 시간에 따른 의존성을 모형에 반영할 필요성이 있다. 본 연구에서는 최근 높은 텍스트 이해 능력을 갖춘 거대 언어모델에 주목하고, 언어모델을 함수형 데이터분석에 활용하는 효과적인 방법을 제시하고자 한다. 거대 언어모델은 토큰들의 순서를 학습하는 매우 효과적인 방법이나, 대량의 문장과 고수준의 계산자원을 요구하므로 이를 효율적으로 추정하는 함수추정방법이 필요하다. 연구에서는 함수형 데이터를 범주화하여 언어 모델의 입력으로써 토큰의 개념을 정의하고, 어텐션 기반의 자기지도 학습을 통해 함수형 데이터의 분포를 학습하는 방안을 제시한다. 아울러, 거대 언어모형에 필요한 데이터를 확보하는 방안으로 trend filtering을 통해 데이터를 증강하는 기법을 제안한다. 또한, 관찰된 자료와 증강된 자료의 대비를 통한 잡음대조추정방법(noise contrastive estimation, NCE)을 도입하여, 추정의 효율성을 높이는 사전학습방법을 제안한다. 실제 보행자료를 통해 제안된 방법이 이상치에 둔감하고 자료의 측정시점의 정렬에 덜 의존함을 확인할 수 있었으며, 다운스트림 작업인 뇌성마비 환자의 운동기능 정상/비정상 분류성능을 향상시킴을 확인하였다.

주요용어 : BERT, trend filtering, data augmentation, NCE.

^{*}본 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A1B05028565, 2017R1E1A1A03070345, 2021R1A6A1A10044950).

¹02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 석사과정.
E-mail: hyez@uos.ac.kr

²16499 경기도 수원시 영통구 월드컵로 206, 아주대학교 수학과 교수. E-mail: qrio1010@ajou.ac.kr

³34141 대전광역시 유성구 대학로 291, KAIST 산업및시스템공학과 교수. E-mail: jyahn@kaist.ac.kr

⁴34141 대전광역시 유성구 대학로 291, KAIST 수리과학과 교수. E-mail: parkcw2021@kaist.ac.kr

⁵02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 조교수.

E-mail: kyungwoo.song@uos.ac.kr

⁶(교신저자) 02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 부교수.

E-mail: choi.hosik@uos.ac.kr

Reference

- Wang, J. L., Chiou, J. M. and Müller, H. G. (2016). Functional data analysis, *Annual Review of Statistics and Its Application*, 3, 257-295.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Agresti, A. (2012). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Kim, S.-J., Koh, K., Boyd, S. P., and Gorinevsky, D. (2009). l1-trend filtering. *SIAM Rev.*, 51:339 - 360.
- SHIYANG, L., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In NeurIPS.
- Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285 - 323.
- Müller, H. G. (2005). Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32(2), 223-240.
- Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*.

Forecasting Tail Risk of Time Series with Weighted Scoring Rules

Sungchul Hong¹, Jong-June Jeon²

Abstract

In resource management or financial field, extreme events such as downpour, drought or bankruptcy are main research topics since they may incur severe problems. Many studies have built on tails in distribution of resource and return to forecast future events for a provision plan. This study especially focuses on tail risk such as Value-at-Risk(VaR) in time series to predict resource exhaustion. Using weighted continuous ranked probability score(CRPS) as loss function we impose larger weights to lower quantiles. To tackle intractable integral in loss function we exploit monotonic linear splines which parameterize the quantile function and it gives rise to closed form of the integral. In our methods, deep learning models like transformers return the parameters of monotonic linear spline that allow for quantile estimation. In numerical studies, we compare probabilistic forecasters and their property in lower quantile estimation with real-world time series datasets.

Keywords : Probabilistic forecast, Quantile estimation, Value-at-Risk, Deep learning.

¹Graduate Student, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, Korea. E-mail: sc.hong@uos.ac.kr

²(Corresponding Author) Associate Professor, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, Korea. E-mail: jj.jeon@uos.ac.kr

분리된 공정 표현 학습을 위한 분포 대조*

오창대¹, 원희지², 소준혁³, 김태로⁴, 김예원⁵, 최호식⁶, 송경우⁷

요약

공정한 표현을 학습하는 것은 알고리즘의 공정성(fairness)을 달성하거나, 민감한 정보에 대한 비판향화(debiasing)를 실현하는 데에 중요하다. 대부분의 기존 연구들은 불변성을 주입하기 위해 적대적인 표현학습법에 의존하지만, 적대적 학습법은 상대적으로 불안정한 훈련과정을 겪는 것으로 알려져 있으며, 이는 데이터 표현의 공정성과 예측력 사이의 균형을 해칠 수 있다. 본 연구에서는 분포 대조적 변이형오토인코더를 통한 공정 표현 학습이라는 새로운 접근법 FarconVAE를 제안하며, FarconVAE는 데이터에 표현에 대한 잠재 공간이 민감 정보에 관한 부분과 비 민감한 정보에 관한 부분으로 분리되게끔 유도한다. 그 후, FarconVAE는 잠재표현들의 분포를 대조함으로써 비 민감 잠재벡터들끼리 가까워지도록 하면서도 민감 잠재벡터들끼리는 서로 멀어지게, 또한 민감 잠재벡터와 비 민감 잠재벡터들끼리도 서로 멀어지게 강제한다. 본 연구에서는 분포 대조 학습을 위해 가우시안(Gaussian) 커널과 스튜던트-t(Student-t) 커널로부터 영감을 받은 새로운 종류의 대조 손실함수를 이론적인 분석과 함께 제안하며, 데이터 표현상의 분리를 더욱 증폭시키기 위해 새로운 교환-재구성(swap-reconstruction) 손실을 추가적으로 활용한다. 테이블, 이미지, 텍스트 등 다양한 정형/비정형 데이터들에 대한 공정성 평가, 사전학습 모델 비판향화, 도메인 일반화 실험들에서 FarconVAE는 우수한 성능을 보임을 확인하였다.

주요용어 : 공정성, 분리, 불변 학습, 대조학습, 변이형 오토인코더.

1. 서론

최근 머신러닝(machine learning) 알고리즘들의 놀라운 성공에 힘입어 사회의 다양한 분야에서 데이터기반 의사결정 시스템(data-driven decision making system)을 도입하고 있으며, 범죄예측이나 신

*본 논문은 IITP(2021-0-02067, 50%)와 NRF(2021R1F1A1060117, 50%)의 지원을 받아서 연구된 것임.

¹02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 석사과정.

E-mail: bnormal16@uos.ac.kr

²02841 서울특별시 성북구 안암로 145, 고려대학교 컴퓨터학과 석사과정. E-mail: gmlwl1026@korea.ac.kr

³37673 포항시 남구 청암로 77, 포항공과대학교 컴퓨터공학과 석사과정. E-mail: junhyukso@postech.ac.kr

⁴02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 석사과정.

E-mail: rlaxofh123@uos.ac.kr

⁵02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 석사과정.

E-mail: yeyewon12@uos.ac.kr

⁶02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과 부교수.

E-mail: choi.hosik@uos.ac.kr

⁷02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 조교수.

E-mail: kyungwoo.song@uos.ac.kr

용예측 등의 민감하면서 중요한 문제에 대해서도 머신러닝 모델을 활용하는 경우가 늘어나고 있다. 이때, 대부분의 머신러닝 모델들은 주어진 데이터셋에 대하여 잘 추출된 데이터 표현 벡터들 위에 단순한 예측기를 붙여 예측을 수행하게 되는데 이 때문에 머신러닝 알고리즘의 성능은 데이터의 벡터 표현(Bengio et al. 2013)에 크게 좌우된다. 한편, 인간에 의해 수집되는 다양한 데이터셋들에는 관심 있는 작업 수행에 관한 핵심적인 정보들 외에 불필요한 노이즈(noise)나 민감한 개인 정보(sensitive information) 등이 포함되어있을 수 있으며, 이러한 원치 않는 정보들에 대해 불변성(invariance)을 가지면서도 관심 작업과 관련된 데이터의 핵심적인 특성을 잘 포함하는 표현을 학습하는 것은 최근 중요한 문제로 대두되고 있다.

신경망 모델은 표현학습(representation learning)에 능통하다고 알려져 있으며, 신경망으로부터 학습된 데이터 표현은 고차원의 원본 데이터들에 대하여 내제된 핵심정보들을 잘 응축시킨 저차원의 벡터들로 볼 수 있다. 이러한 신경망기반 방법들은 이미지(He et al. 2016), 텍스트(Devlin et al. 2019), 테이블데이터(Xu et al. 2019) 등의 많은 도메인들에서 유의미한 성능 향상을 보여주었으나, 최근연구들에서 신경망이 공정성 및 도메인 일반화를 달성하는 데에 어려움을 겪는다는 것이 밝혀졌다(Arjovsky et al. 2019, Sarhan et al. 2020). 이는 신경망이 학습되는 과정에서 관심 과업과 관련된 중요한 정보들을 데이터 표현에 포함시킬 뿐만 아니라 민감한 개인정보나 불필요한 노이즈까지 흡수하기 때문이며, 학습된 데이터 표현에 존재하는 이러한 원치 않는 정보들 때문에 우리의 모델은 의사결정 시스템에서 불공평한 예측을 유도하거나, 데이터의 분포 변화(distribution shift) 상황에서 올바른 예측을 수행하는 데에 실패할 수 있다.

머신러닝 모델에서 민감 정보 혹은 허위적 상관관계를 제거하기 위한 많은 방법론들이 제안되어왔으며, 적대적 표현학습(adversarial representation learning) 방식이 대표적인 접근 방향이다(Roy, Boddeti, 2019). 이는 데이터 표현에 타겟 속성에 대한 정보는 최대한으로 보존하면서도 민감 속성에 대한 정보 누출을 최소화 하도록 상충하는 두 목적 식을 적대적으로 최적화하는 방법으로, 일단 학습이 완료되면 좋은 성능을 보여주나, 학습되는 과정이 불안정하다는 단점이 있다. 대안으로는 분리된 표현학습(disentangled representation learning)이 있으며, 데이터 표현 벡터들의 각 차원들이 오직 하나의 속성만을 암시하도록 벡터표현공간을 분리(disentanglement)하는 방법론이다(Creager et al. 2019). 그러나 많은 분리된 표현학습 방법론들은 표현 분리 목적의 달성을 위해 부분적으로 적대적 학습 손실을 활용하고 있으며, 이 역시 훈련 간에 불안정성을 야기한다. 또 다른 대안으로 불변 표현학습(invariant representation learning) 방법들(Arjovsky et al. 2019, Sagawa et al. 2020)이 있으나, 이들 역시 학습되는 표현에 민감 속성에 관한 정보가 포함되지 않을 것이라는 보장은 없다.

본 연구에서는, 상기 기술한 기존 방법론들의 모든 문제들을 회피하는 분포적 대조학습 기반의 분리된 공정한 표현을 얻는 방법론을 제시하며, 이를 위해 커널 기반의 새로운 분포대조 손실함수와 함께 교환-재구성 기법을 제안한다. 또한 다양한 종류의 데이터들에 대한 공정성 평가, 사전학습 모델 비편향화, 도메인 일반화 실험 결과를 통해, 제안하는 방법론의 유효성을 입증하였다.

Reference

- Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: *A review and new perspectives*. IEEE transactions on pattern analysis and machine intelligence 35, 8 (2013), 1798 - 1828.

- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 770 - 778.
- Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proceedings of NAACL-HLT*. 4171 - 4186.
- Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. 2019. Modeling tabular data using conditional gan. *Advances in Neural Information Processing Systems* 32 (2019).
- Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. 2019. Invariant risk minimization. *arXiv preprint arXiv:1907.02893* (2019).
- Mhd Hasan Sarhan, Nassir Navab, Abouzar Eslami, and Shadi Albarqouni. 2020. Fairness by learning orthogonal disentangled representations. In *European Conference on Computer Vision*. Springer, 746 - 761.
- Proteek Chandan Roy and Vishnu Naresh Boddeti. 2019. Mitigating information leakage in image representations: A maximum entropy approach. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2586 - 2594.
- Elliot Creager, David Madras, Jörn-Henrik Jacobsen, Marissa Weis, Kevin Swersky, Toniann Pitassi, and Richard Zemel. 2019. Flexibly fair representation learning by disentanglement. In *International conference on machine learning*. PMLR, 1436 - 1445.
- Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. 2020. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731* (2020).

Distribution Contrast for Disentangled Fair Representation*

*Changdae Oh¹, Heeji Won², Junhyuk So³, Taero Kim⁴,
Yewon Kim⁵, Hosik Cho⁶, Kyungwoo Song⁷*

Abstract

Learning fair representation is crucial for achieving fairness or debiasing sensitive information. Most existing works rely on adversarial representation learning to inject some invariance into representation. However, adversarial learning methods are known to suffer from relatively unstable training, and this might harm the balance between fairness and predictiveness of representation. We propose a new approach, learning FAir Representation via distributional CONTRastive Variational AutoEncoder (FarconVAE), which induces the latent space to be disentangled into sensitive and nonsensitive parts. We first construct the pair of observations with different sensitive attributes but with the same labels. Then, FarconVAE enforces each non-sensitive latent to be closer, while sensitive latents to be far from each other and also far from the non-sensitive latent by contrasting their distributions. We provide a new type of contrastive loss motivated by Gaussian and Student-t kernels for distributional contrastive learning with theoretical analysis. Besides, we adopt a new swap-reconstruction loss to boost the disentanglement further. FarconVAE shows superior performance on fairness, pretrained model debiasing, and domain generalization tasks from various modalities, including tabular, image, and text.

Keywords : Fairness, Disentanglement, Invariant Learning, Contrastive Learning, Variational Autoencoder.

*This paper is researched with the support of IITP(2021-0-02067, 50%) and NRF(2021R1F1A1060117, 50%).

¹M.S. student, Department of Artificial Intelligence, University of Seoul, 163, Seoulsiripdae-ro, Dondaemun-gu, Seoul, 02504, Korea. E-mail: bnormal16@uos.ac.kr

²M.S. student, Department of Computer Science and Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: gmlw11026@korea.ac.kr

³M.S. student, Department of Computer Science and Engineering, POSTECH, 77, Cheongam-ro Nam-gu, Pohang, 37673, Korea. E-mail: junhyukso@postech.ac.kr

⁴M.S. student, Department of Artificial Intelligence, University of Seoul, 163, Seoulsiripdae-ro, Dondaemun-gu, Seoul, 02504, Korea. E-mail: rlxofh123@uos.ac.kr

⁵M.S. student, Department of Artificial Intelligence, University of Seoul, 163, Seoulsiripdae-ro, Dondaemun-gu, Seoul, 02504, Korea. E-mail: yeyewon12@uos.ac.kr

⁶Associate Professor, Department of Urban Big Data Convergence, University of Seoul, 163, Seoulsiripdae-ro, Dondaemun-gu, Seoul, 02504, Korea. E-mail: choi.hosik@uos.ac.kr

⁷(Corresponding Author) Assistant Professor, Department of Artificial Intelligence, University of Seoul, 163, Seoulsiripdae-ro, Dondaemun-gu, Seoul, 02504, Korea. E-mail: kyungwoo.song@uos.ac.kr

EXoN: EXplainable encoder Network

SeungHwan An¹, Hosik Choi², Jong-June Jeon³

Abstract

We propose a new semi-supervised learning method of Variational AutoEncoder(VAE) which yields a customized explainable latent space by EXplainable encoder Network(EXoN). Customization means a manual layout design of latent space for specific labeled data. The EXoN provides two useful guides for customization of VAE by a conceptual center and a multi-modality of the latent space. The latent subspace produced by EXoN can be easily investigated by a simple measure. To improve the performance of our VAE in a classification task without the loss of performance as a generative model, we employ a new semi-supervised classification method called ‘SCI(Soft-label Consistency Interpolation)’. It is found that both the classification loss and the Kullback-Leibler divergence play a crucial role in constructing explainable latent space and the variability of generated samples from our proposed model depends on a specific subspace, called ‘activated latent subspace’. Our numerical results with MNIST and CIFAR-10 datasets show that EXoN produces an explainable latent space and reduces the cost of investigating representation patterns on the latent space.

Keywords : Variational AutoEncoder, explainability, customization, semi-supervised learning.

¹Integrated Master’s & Doctoral Program Student, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, Korea. E-mail: dkstmdghks79@uos.ac.kr

²Associate Professor, Graduate School, Department of Urban Big Data Convergence, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, Korea. E-mail: choi.hosik@uos.ac.kr

³(Corresponding Author) Associate Professor, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, Korea. E-mail: jj.jeon@uos.ac.kr

Neural Temporal-Filter Network: 시계열 예측 정확도 향상을 위한 신경망 구조*

심성현¹, 강창완²

요약

본 연구는 시계열 예측 문제에서 예측 정확도를 향상시키기 위해 Neural Temporal-Filter Network(NTFNet) 라고 불리는 새로운 시계열 신경망 구조를 제안한다. NTFNet에 포함된 네트워크 셀의 내부 구조에는 Variational Mode Decomposition(VMD) 기법을 통해 입력된 시계열 데이터를 분해하는 Temporal-Filter와 이를 통해 분해된 값들과 이전 시점에 학습된 은닉 상태와의 관계를 학습하기 위한 Temporal-Attention Gate로 구성되어 있다. 제안한 NTFNet의 성능을 확인하기 위해 단변량 시계열 데이터 및 다변량 데이터를 이용하여 비교 실험을 수행하였다. 그 결과 예측에 활용된 여러 딥러닝 방법들과 비교하였을 때 단기 및 장기 예측에서 모두 평균 15% 수준으로 예측성능이 향상되었음을 보인다.

주요용어 : 시계열 예측, 딥러닝, Neural Temporal-Filter Network.

1. 서론

시계열 예측 문제는 데이터 분석 영역에서 오랜 기간 동안 다루어진 연구 주제이다 (Karim et al., 2020). 전통적으로 Auto-regressive Integrated Moving Average (ARIMA), Vector Auto Regression (VAR) 과 같은 통계적 접근법이 많이 활용되어 왔고, 최근 인공지능 기술의 발전과 함께 Recurrent Neural Network (RNN), Long-Short Term Memory (LSTM), Gate Recurrents Unit (GRU)와 같은 딥러닝 기반의 예측 방법론들이 활발히 제안되고 있다 (Hua et al., 2019).

시계열 예측이 어려운 이유는 예측하고자 하는 종속 변수에 대한 독립변수를 정의하기 어려울 뿐 아니라, 개별 독립변수들이 종속 변수에 영향을 주는 시점이 상이하기 때문이다 (Apaydin, 2020). 이러한 이유로 최근 시계열 예측 분야에서 시계열 분해 기법을 결합한 하이브리드 딥러닝 모델 (Kamal, et al. 2020)과 잠재적인 시계열 데이터들 간의 패턴을 학습하는 어텐션 기법을 결합한 딥러닝 모형들이 사용되고 있다 (Huang et al., 2011).

본 연구는 단변량 및 다변량 시계열 예측의 정확도 향상을 위해 Neural Temporal-Filter Network (NTFNet)라고 불리는 새로운 네트워크 구조를 제안한다. 제안된 네트워크에는 시계열 데이터를 분

*본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성 (GrandICT 연구센터) 사업의 연구결과로 수행되었음(IITP-2022-2020-0-01791).

¹(교신저자) 47340 부산광역시 부산진구 엄광로 176, 동의대학교 산업경영빅데이터 공학과 조교수.

E-mail: ssh@deu.ac.kr

²47340 부산광역시 부산진구 엄광로 176, 동의대학교 산업경영빅데이터 공학과 교수. E-mail: cwkwang@deu.ac.kr

해하는 Temporal-Filter 와 은닉상태와 입력 데이터와의 관계를 학습하기 위한 Temporal-Attention Gate로 구성되어 있다.

이어서 기술할 본 논문의 구성은 다음과 같다. 2장에서는 제안된 네트워크 구조를 설명한다. 3장과 4장에서는 실험을 위한 데이터 소개와 실험 방법을 소개하고 실험 결과에 대해서 설명한다. 마지막으로 5장에서는 본 연구의 결론과 추가 연구사항에 대해서 논의한다.

2. 제안 네트워크 구조

본 연구에서 제안하는 Neural Temporal-Filter Network(NTFNet)라고 불리는 새로운 네트워크 구조를 제안한다. 제안된 NTFNet은 t 시점의 시계열 데이터 내에 존재하는 시계열 패턴(X_t)으로 분해하는 Temporal-Filter와 분해된 시계열 패턴(X_t)과 각 시계열 패턴에 대응하는 은닉상태 (H_{t-1})와의 관계를 학습하기 위한 Temporal-Attention Gate로 구성되어 있다. NTFNet의 구조는 Figure 1과 같다.

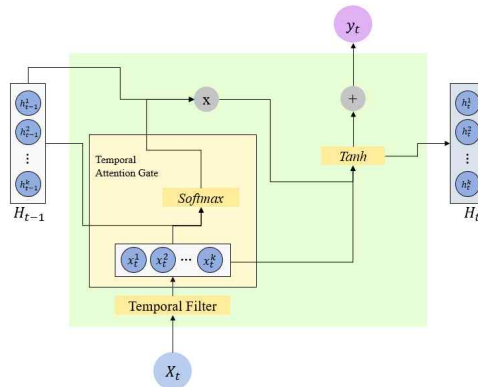


Figure 1. Cell Architecture of Neural Temporal-Filter Network

Temporal-Filter에서는 입력된 시계열 데이터를 k 개의 성분으로 분해하는 역할을 수행한다. 이때 분해된 성분은 시계열 데이터에 내재된 특정 패턴 단위로 분리되는 데이터를 의미한다. 시계열 데이터를 특정 패턴 단위로 분해하기 위해, 본 연구에서는 Variational Mode Decomposition(VMD) 방법을 활용하였다. 즉 Temporal-Filter의 입력 값은 t 시점의 입력된 시계열 데이터를 의미하고, t 시점의 분해된 시계열 데이터 $X_t = x_t^1, x_t^2, \dots, x_t^k$ 가 출력된다.

Temporal-Attention Gate에서는 t 시점의 입력 데이터를 패턴 단위로 분해한 시계열 분해 데이터 (X_t)와 $t-1$ 시점에서 학습된 은닉상태와(H_{t-1})의 관계를 학습하기 위한 게이트이다. 이때 활성화 함수로 Softmax 함수를 활용하여 X_t 와 H_{t-1} 사이의 관계를 학습하며, 식 (1)과 같이 학습된다.

$$\alpha = \text{softmax}(W_{XH}X_t + W_{HH}H_{t-1} + b_t) \quad (1)$$

최종적으로 k 개의 시계열 데이터로 분해된 X_t 와 Temporal-Attention Gate를 통해 학습된 α 를 이용하여 예서는 t 시점의 은닉상태(H_t)를 학습하며, 식 (2)와 같이 학습된다. 여기서 \odot 는 Hadamard product 연산자이다.

$$H_t = \tanh(W_{XH}X_t + W_{HH}(\alpha \odot H_{t-1}) + b_H) \quad (2)$$

3. 실험자료 및 방법

본 연구에서 제안하는 NTFNet의 성능을 비교하기 위해서, 2개의 단변량 시계열 데이터(UTD1, UTD2) 와 2개의 다변량 시계열 데이터(MTD1, MTD2) 대해서 비교 실험을 수행하였다. 먼저 2개의 단변량 데이터는 해운 분야에서 선박의 운임가격을 결정하기 위해 활용되는 발틱해운거래소에서 제공하는 BDI 데이터와 유럽의 트랜잭션 로그에서 제공하는 유럽 연합 배출권 거래 데이터를 이용하였다. 다변량 시계열 데이터에 대한 성능 비교 실험을 위해서, 본 연구에서는 야후파이낸스에서 제공하는 11개 국가의 주가지수 데이터와 세계무역기구가 제공하는 6개국의 주요 수출입물량 데이터를 활용하였다.

NTFNet과의 비교실험을 위해 단변량 시계열 데이터에는 LSTM과의 예측 성능을 비교하였다. 다변량 시계열 데이터에는 LSTM과 어텐션 기법이 적용된 LSTNet-A와 DSANet과의 비교실험을 수행하였다. 장단기 예측 성능을 비교하기 위해서 각각 1-Step, 6-Step, 12-Step 간격을 예측하였다. 실험에 사용된 학습 및 검증 데이터의 분리 비율은 7대 3을 사용하였고, 30번의 반복실험을 통해 그 성능을 비교하였다. 검증데이터의 성능을 비교하기 위해서 평균제곱근오차 (RMSE)와 평균절대비오차(MAPE)를 사용하였다.

4. 실험 결과

Table 1과 Table 2는 각각 단변량 시계열 데이터 및 다변량 시계열 데이터를 활용한 검증 데이터의 예측 오차의 평균과 표준편차를 보여준다. NTFNet의 단변량 시계열 데이터에 대한 예측성능을 비교하기 위해서 LSTM과 비교 실험을 수행하였고, 이때 두 모형 모두 512개의 은닉상태를 가지는 1개의 Layer로 구성된 모델을 사용하였다. 이 때 2개의 단변량 데이터 (UTD1, UTD2)에서 단기 뿐 아니라 장기 에서도 모두 제안된 NTFNet의 성능이 좋은 것으로 나타났다.

다변량 시계열 데이터를 이용한 비교실험에서는 LSTM 대비 50%정도 예측 성능이 개선되었으며, Attention 기법이 적용된 LSTNet-A와 DSANet에 비해서 단기예측 및 장기 예측에서 최대 15%로 예측 오차가 감소하였다.

Table 1. Results of prediction of univariate time-series data

모델	척도	UTD1			UTD2		
		1-Step	6-Step	12-Step	1-Step	6-Step	12-Step
NTFNet*	RMSE	21.09 (0.030)	95.08 (0.045)	163.5 (0.059)	0.690 (0.006)	1.515 (0.009)	2.375 (0.019)
	MAPE	1.613 (0.002)	6.802 (0.004)	11.47 (0.009)	1.911 (0.022)	4.320 (0.034)	6.868 (0.051)
LSTM	RMSE	30.62 (0.081)	118.1 (0.078)	210.1 (0.152)	0.868 (0.010)	1.901 (0.014)	2.824 (0.026)
	MAPE	1.984 (0.006)	8.172 (0.018)	15.01 (0.059)	2.291 (0.035)	5.401 (0.055)	8.531 (0.082)

예측오차의 평균 (표준편차)

Table 2. Results of prediction of multivariate time-series data

모델	척도	MTD1			MTD2		
		1-Step	6-Step	12-Step	1-Step	6-Step	6-Step
NTFNet*	RMSE	97.55 (2.180)	230.53 (3.347)	430.03 (4.57)	9945.1 (434.1)	19485.1 (618.8)	21319.9 (755.1)
	MAPE	0.91 (0.004)	1.951 (0.008)	4.04 (0.026)	5.22 (0.18)	9.31 (0.30)	11.69 (0.57)
LSTM	RMSE	202.98 (6.540)	618.04 (28.50)	814.05 (40.12)	12696.1 (462.3)	27641 (935.6)	28092.3 (1294.5)
	MAPE	2.011 (0.190)	6.301 (0.210)	8.31 (0.790)	7.99 (0.28)	15.33 (0.90)	16.78 (1.04)
LSTNet-A	RMSE	106.91 (3.590)	247.19 (5.650)	440.16 (8.76)	10256.3 (545.9)	20738.9 (698.0)	22021.5 (945.1)
	MAPE	1.11 (0.120)	2.19 (0.250)	3.81 (0.254)	5.34 (0.20)	9.45 (0.41)	12.01 (0.60)
DSANet	RMSE	105.34 (3.200)	245.21 (5.540)	432.7 (6.01)	10159.0 (504.6)	19690.5 (648.4)	21316.2 (886.6)
	MAPE	1.05 (0.101)	2.00 (0.150)	3.73 (0.290)	5.30 (0.19)	9.41 (0.33)	11.99 (0.59)

예측오차의 평균 (표준편차)

5. 결론 및 향후 연구

본 연구는 Neural Temporal-Filter Network(NTFN)라고 불리는 새로운 네트워크 구조를 제안한다. 새롭게 제안된 NTFNet에는 입력된 시계열 데이터를 분해하기 위한 Temporal-Filter와 분해된 시계열 데이터와 이전 시점에서 학습된 은닉상태와의 관계를 학습하기 위해 Temporal-Attention Gate로 구성되어 있다. 최종적으로 NTFNet는 Temporal-Filter를 통해 분해된 시계열 성분들과 Temporal-Attention Gate 통해 학습된 정보를 고려하여 은닉상태를 학습하는 방식으로 파라미터를 업데이트한다. 이러한 접근 방법은 시계열 데이터에 내재된 패턴을 분리하고, 분해된 패턴과 은닉상태의 값들과의 관계를 고려하여 학습하기 때문에 더 좋은 예측 성능을 보장할 수 있다. NTFNet의 우수성은 비교실험을 통해 입증 되었으며, 2개의 단변량 시계열 데이터와 2개의 다변량 시계열 데이터를 이용한 비교 실험 결과에서 기존 방법론에 비해서 평균 15%에 가까운 예측성능 향상을 보여주었다.

References

- Karim, F., Majumdar, S. & Darabi, H. 2020. Adversarial attacks on time series. *IEEE transactions on pattern analysis and machine intelligence*, 43, 3309-3320.
- Hua, Y., Zhao, Z., Li, R., Chen, X., Liu Z. & Zhang, H. 2019. Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, 57, 114-119.
- Apaydin, H., Sattari, M. T., Falsafian, K. & Prasad, R. 2021. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. *Journal of Hydrology*, 600, 126506.
- Huang, S., Wang, D., Wu, X. & Tang, A. Dsanet: Dual self-attention network for multivariate time series forecasting. *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019. 2129-2132.

딥러닝을 이용한 서술형 강의 평가 분류 연구

성진용¹, 이재훈², 최승배³, 심성현⁴, 강창완⁵

요약

대학에서의 강의평가 제도는 교육의 질을 향상시키는 여러 가지 방법 중 하나이다. 강의평가는 일반적으로 객관식 평가문항과 서술형 평가 문항을 통해 학생들의 수업에 대한 인식을 도출하는데 있다. 객관식 평가는 신뢰도 및 타당도 측면에서 안정된 강의평가 정보를 제공한다는 장점이 있으며 반면에 서술형 평가는 자유로운 의견 서술을 통해 교육의 질을 개선할 수 있는 유의한 피드백을 제공한다는 장점이 있다. 강창완 외 3인(2021)은 텍스트마이닝을 이용하여 D대학 서술형 강의평가 자료를 기계학습 방법을 이용하여 분석한 바 있다. 본 연구에서는 서술형 강의평가 자료 분석의 후속 연구로 딥러닝을 통해 기존 모형의 정분류율을 개선하고, 유용한 정보를 도출해내고자 한다. 이를 위해 첫째, 서술형 강의평가 텍스트 자료에 대해 긍정적 평가인지 부정적 평가인지를 예측 분류하는데 형태소 분석 결과를 기반으로 딥러닝 기법인 RNN, LSTM 방법을 적용하였다. 둘째, Word Piece 임베딩을 통한 딥러닝 기법인 BERT 방법을 적용하였다. 분석용 데이터셋(70%)를 이용하여 최적모형을 찾고, 최적모형을 검증용 데이터셋(30%)에 적용하여 분석하였으며 분석 결과, 검증용 데이터에 대한 예측력 측면에서 BERT 방법의 정분류율이 96%로 나타나 나이브베이즈(85%), RNN, LSTM(88%) 방법에 비해 가장 좋은 성능을 보여 서술형 강의평가 분류모델에 적합한 것으로 확인하였다.

주요용어 : 기계학습, 딥러닝, 나이브 베이스 방법, RNN, LSTM, BERT.

1. 서론

최근 대학에서는 강의평가 제도를 통해 교육의 질을 향상시키고 있다. 강의평가는 보통 정량적 평가 문항과 강의에 대한 의견을 기술하는 서술형 평가 문항으로 이루어져 있으며 정량적 평가 문항의 경우는 신뢰도와 타당도 관점에서 잘 계획된 측정도구이므로 안정되게 강의 교육의 질을 평가할 수 있다. 반면 서술형 평가 문항의 경우는 질적 접근을 통한 데이터 특성을 가지고 있으며, 학생들의 강의에 대한 인식을 직접적이고 구체적으로 파악할 수 있는 중요한 자료라고 할 수 있

¹47340 부산광역시 부산진구 엄광로 176 동의대학교 대학원 수학정보통계학과, 석사과정.
E-mail: wlsdyd3194@naver.com

²47340 부산광역시 부산진구 엄광로 176 동의대학교 대학원 수학정보통계학과, 석사과정.
E-mail: dynamic973124@naver.com

³47340 부산광역시 부산진구 엄광로 176 동의대학교 산업경영빅데이터공학전공, 교수.
E-mail: csb4851@deu.ac.kr

⁴47340 부산광역시 부산진구 엄광로 176 동의대학교 산업경영빅데이터공학전공, 조교수. E-mail: ssh@deu.ac.kr

⁵(교신저자) 47340 부산광역시 부산진구 엄광로 176 동의대학교 산업경영빅데이터공학전공, 교수.
E-mail: cw kang@deu.ac.kr

다. 그러나 서술형 강의평가와 같은 대용량의 텍스트 자료 분석에 대한 분석은 비정형데이터 특성을 갖고 있어 분석 방법이 다소 제한적이다. 최근 빅데이터(Big Data) 기술의 발전에 따라 다양한 분석 방법이 개발되고 있으며, 비정형 데이터를 이용해 텍스트 문서로부터 의견과 감정을 도출하여 패턴을 분석한다. 강창완 외 3인(2021)은 이러한 필요성에 따라 D 대학교 서술형 강의평가의 긍/부정 분류를 텍스트마이닝 기법 중 나이브베이즈 방법을 적용한 바 있다. 본 연구에서는 D 대학교의 서술형 강의평가 분석 방법의 시간·물리적인 노력을 감소시킬 수 있는 개선안 모색과 더 나아가 과학적인 분석 결과를 도출을 위한 딥러닝 기법의 활용성 및 효과성을 검토한다(Kim et al, 2021). 구체적인 연구 목적은 서술형 강의의 평가 자료들이 긍정적인 평가인지, 부정적인 평가인지를 예측 분류하는데 있으며, 이에 대한 적용을 위해 나이브 베이즈(Naive Bayes), RNN(Recurrent Neural Network), LSTM(Long Short-Term Memory), BERT(Bidirectional Encoder Representations from Transformers) 방법을 고려하였다.

본 연구의 구성은 2장에서 텍스트마이닝과 본 연구에서 사용한 관련 분석 방법에 대해서 개략적으로 소개하고, 3장에서는 서술형 강의평가 데이터 구조와 분석 목적, 그리고 분석 결과를 제시한다.

2. 텍스트마이닝

텍스트 마이닝은 비정형, 반정형 데이터에 대하여 자연어 처리 기술과 문서 처리 기술을 적용하여 유용한 정보를 추출, 가공하는 목적으로 하는 기술로서 최근 기계학습(machine learning) 방법을 이용하여 광범위하게 활용되고 있다.

2.1 RNN 방법을 이용한 문서분류

RNN은 말 그대로 순환 모형을 가지며, 가장 기본적인 인공신경망 시퀀스 모델이다. RNN은 동일한 활성화 함수를 한 시퀀스의 모든 요소마다 적용하여 출력 결과가 이전의 계산 결과에 영향을 받으며, 음성, 문자와 같은 순차적으로 등장하는 데이터 처리에 적합한 모델로 알려져 있다. RNN은 시퀀스 길이와 관계없이 인풋과 아웃풋을 받아들일 수 있는 네트워크 구조이기 때문에 필요에 따라 다양하고 유연하게 구조를 만들 수 있다는 장점이 있지만, 비교적 짧은 시퀀스만 효과적으로 처리한다는 한계점이 존재하고 이를 장기 의존성 문제(the problem of Long-Term Dependencies)라고 한다(RNN Tutorial-Part1). RNN의 일반적인 구조는 Figure 1과 같다.

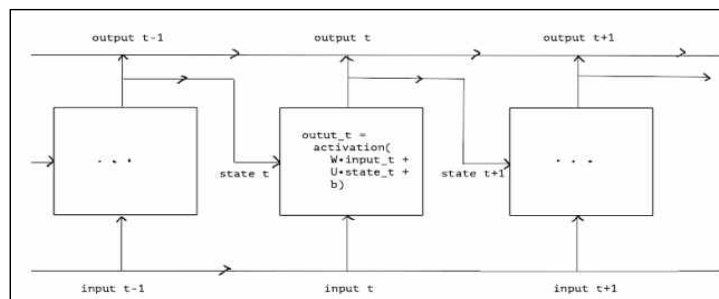


Figure 1. Structure of RNN

2.2. LSTM 방법을 이용한 문서분류

LSTM은 RNN의 장기 의존성 문제를 극복하고자 순환 신경망의 변형인 모형이다. 기존 RNN의 구조에서 데이터를 계산하는 각 길목에 입력 게이트, 망각 게이트, 출력 게이트가 추가되어 각 상태 값을 메모리 공간 셀에 저장하고, 데이터를 접하는 게이트 부분을 조정하여 불필요한 연산, 오차 등을 줄여 장기 의존성 문제를 일정부분 해결하였다(Hochreiter, S.&Schmidhuber, J., 1997).

2.3. BERT 방법을 이용한 문서분류

BERT는 구글이 공개한 사전 훈련(pre-trained)된 언어 모델로서 트랜스포머(transformer)를 이용하여 구현되었고 위키피디아 (25억 단어)와 BooksCorpus(8억 단어)와 같은 레이블이 없는 텍스트데이터로 사전 훈련된 모델이다. 특히 BERT는 2018년 공개된 이후로 자연어처리 분야에서 최고의 성능을 보여주고 있다(wikidocs.net/115055). BERT가 최고 성능을 보여주는 이유는 레이블이 없는 방대한 데이터로 사전 훈련된 모델을 가지고, 레이블이 있는 다른 작업에서 추가 훈련과 함께 하이퍼파라미터를 재조정하여 모델을 이용하기 때문이다. BERT의 기본 구조는 트랜스포머의 인코더를 쌓아올린 구조이며 본 연구에서는 한글 BERT를 이용하였다.

3. 데이터 분석

3.1. 데이터 소개

본 연구에서 사용한 데이터는 D 대학교 2021년 1학기 서술형 강의평가 데이터로서 총 10,061개의 평가 문장으로 구성되어 있다. 사용한 변수는 text 변수와 target 변수만을 이용하였고 여기서 target 변수는 이항 분류(긍정 평가= S, 부정 평가=C)로 라벨링하였다.

3.2. 분석 결과

먼저, 나이브베이즈, RNN, LSTM으로 주어진 데이터를 군집화하고 실제 목표(target)변수와 얼마나 일치하는지 정분류한 결과를 제시하였다(Table 1). 여기서 분석은 R 프로그램으로 분석하였다.

Table 1. Accuracy of methods

		accuracy	sensitivity	specificity
Naive Bayes method	train data	0.846	0.60	0.92
	test data	0.842	0.60	0.91
RNN	train data	0.887	0.81	0.97
	test data	0.865	0.67	0.92
LSTM	train data	0.886	0.86	0.95
	test data	0.866	0.71	0.91
BERT	train data	0.950	0.98	0.94
	test data	0.946	0.97	0.86

한편, RNN, LSTM을 분석할 때 단어 임베딩을 하게 되는데 NIA 형태소 사전을 활용하여 다음과 같은 형태의 단어 임베딩 행렬을 만든다. RNN, LSTM 분석 절차는 다음과 같다.

- 단계 1) R의 KoNLP 패키지 안에 있는 SimplePos22를 사용하여 형태소 분석을 시행한다.
- 단계 2) 단계 1 결과와 NIA 형태소 사전과 조인(Join)하여 단어 임베딩을 한 데이터를 만든다.
- 단계 3) 전처리 작업이 끝난 데이터를 문서-단어 임베딩 행렬(document-term matrix)로 만든다.
- 단계 4) RNN, LSTM을 수행하고 정분류율을 확인한다.

본 연구에서 사용한 나이브 베이즈, RNN, LSTM 방법 적용은 주어진 문서 치료를 학습용 데이터와 검증용 데이터로 7:3 비율로 분할하고 학습용 데이터를 이용하여 모형을 학습시킨 후 검증용 데이터에 적용함으로써 모형 성능이 얼마나 좋은지 평가하게 된다. RNN과 LSTM 방법은 형태소 분석을 활용한 문서 분류 결과로써 검증용 데이터의 정분류율이 대략 0.87로 확인되어지지만 민감도와 특이도 값이 LSTM이 더 높아 RNN 보다는 좋은 예측 분류 모델이라고 판단된다.

한편, 형태소 분석을 활용하지 않은 BERT 방법을 적용한 정분류율은 0.95로 나타나 본 연구에서 고려한 모든 모형 중 가장 좋은 예측모형으로 나타났다.

Different Perspectives on Learning Statistics

Taeyoung Kim¹, Seungbae Chor²

Abstract

At the most graduate schools of education around the globe, students are required to take basic Statistics courses for their future research including thesis and dissertation even though some of students do not have any background of Statistics. The current study investigates how do they narrate and reflect upon challenges to study Statistics via qualitative methods. Considering different strata, international and native, this study examines how they differently react to this issue. In-depth interviews were used as the primary data collection and a focus group interview with international students was further employed. Preliminary results suggest that the international students' most challenging factor is language issue rather than the subject itself, while domestic students are more interested in how the subject is taught.

Keywords : Quantitative methods, In-depth interview, Statistics education, Statistical literacy, Language barrier.

1. Introduction

One of the founders of modern Statistics, Karl Pearson, said "Statistics is the grammar of science." As this popular statistical quote, Statistics is taught in almost all research-based schools around the world. Even though some of them don't have independent "Statistics department", basic statistics courses are provided as a general requirement or an elective. Like Mathematics and Science literacy, statistical literacy has also been emerging field. Former American Statistical Association(ASA) President Katherine Wallman defined statistical literacy as "including the cognitive abilities of understanding and critically evaluating statistical results as well as appreciating the contributions statistical thinking can make"(Wallman, 1993). Also, in the era of big data, the need for Statistics is growing as a tool of dealing with big data.

Accordingly, Statistics education recently emerged as a field that puts emphases on the teaching and learning of statistics at all level, and in both formal and informal educational

¹Research Professor, Dignitas Institute for Liberal Education, Dong-eui University, 176 Eomgwang-ro, Busan, Korea, 47340. E-mail: tkim33@deu.ac.kr

²Professor, Industrial Management & Big Data Engineering Major, Dong-eui University, 176, Eomgwang-ro, Busanjin-gu, 47340. E-mail: csb4851@deu.ac.kr

settings(Garfield & Ben-Zvi, 2008). Also, there has been a strong need for research on statistics education which are designed to offer implications for instruction and future research across several disciplines such as psychology, mathematics education, educational psychology, and statistics(Garfield & Ben-Zvi, 2007). To keep pace with this change, a couple of institutions opened a graduate program that provides a degree track in statistics education(University of Minnesota(<http://www.cehd.umn.edu/EdPsych/Programs/QME/stats-intro.html>), and University of Florida(<http://education.ufl.edu/statistics-education/>)).

Adding to prior research conducted with participants of graduate students in education, psychology or social sciences(e.g. Finney, 2000; Earley, 2001), this paper presents the experiences of graduate students, who are required to take introductory statistics courses for degrees and their future research including thesis and dissertation even though some of them do not have any background of Statistics. For those cases, this study wanted to listen to their voices in terms of challenges to take the courses and to understand statistical concepts.

2. Literature review

The demand for learning Statistics has been increased. Mills(2004) mentioned that “the teaching and learning of statistics has impacted the curriculum in every level of education. The NCTM Principles and Standards(NCTM, 2000) includes a content standard that emphasizes statistical reasoning for Pre-K through grade 12 and as a result, many states now include and emphasize statistical thinking in their statewide curriculum guidelines. Student enrollments in AP statistics courses are increasing each year and there are many ASA-sponsored programs and workshops to support K-12 teachers and administrators(i.e., BAPS, Adopt-A-School).” In postsecondary education, many students are required to take a statistics course, irrespective of their major. Therefore, the basic understanding of statistical concepts and has become increasingly important for them(Mills, 2004).

There has been a change on how to teach statistics, to students of different ages as well as in a wide range of fields(Mills, 2004). In 1992, for example, the American Statistical Association(ASA) and the Mathematical Association of America(MAA) made a joint committee to study the instruction of introductory statistics. The key recommendations were to emphasize statistical reasoning and notions less theoretically, and to foster active learning(Cobb, 1992). According to Moore(1997), a shift on the teaching of statistics calls for researchers and teachers to focus on the synergy between content, pedagogy, and technology. That is, content and pedagogy should be highly affected by technology as well as students should be dynamic participants allocated with organized activities that emphasize statistical concepts and ideas which are not mathematical in nature.

In relation to affective factors concerning statistics education, a lot of efforts have been made to explore what impact a student’s statistical performance. Mills(2004) lists a variety of

surveys designed to measure student attitudes: Statistics Attitude Survey (Roberts & Bilderback, 1980); Attitude Toward Statistics Scale (Wise, 1985); Statistics Attitude Scale (McCall, Belli, & Madjidi, 1990); Multifactorial Scale of Attitudes Toward Statistics (Auzmendi, 1991); Survey of Attitudes Toward Statistics (Schau, Dauphinee, & Del Vecchio, 1995). In more recent years, Fullerton and Kendrick (2013) sought to measure attitude towards statistics with a nationally representative sample of students of advertising majors with Students' Attitudes toward Statistics Scale (STATS) (Sutarso, 1992).

In addition, another body of literature talks about the students' anxiety towards statistics and quantify how anxiety influences performance using different measures (e.g. Revised Mathematics Anxiety Rating Scale (Plake & Parker, 1982); Statistical Anxiety Rating Scale (Cruise, Cash, & Bolton, 1985); Statistics Anxiety Inventory (Zeidner, 1991)). From prior research, Hopkins, Hopkins, and Glass (1996) found that students' attitudes toward statistics consist of anxiety, cynicism, fear, and contempt.

Also, Bui and Alfaro (2011) argued that "particular concern for many professors of statistics is anxiety associated with taking a course in statistics. Studies that have examined statistics anxiety have shown that a student's competence and success in a statistics course is influenced by their attitudes toward the course (Kottke, 2000)." Baloglu (2004) indicated that "statistics anxiety is a relatively new construct and is related to but different from math anxiety." Although both mathematics anxiety and statistics anxiety accompany the stress that students feel when coping with mathematical reasoning, Baloglu (2004) states that "statistics anxiety is distinct in that students also have apprehension involving the verbal reasoning and manipulation of the mathematical symbols that are required in understanding statistics."

The findings from prior research introduced above display that statistics education has been continuously emphasized and it created much discussion on the teaching and learning of statistics in every level of education. Also, a number of research demonstrated students' affective domains involved in learning statistics such as statistics anxiety and attitude toward statistics, and attempted to measure those constructs using a variety of measures.

Although there have been much prior studies in relation of statistics education, most of them followed quantitative approach and the participants in the studies were mainly undergraduates from diverse majors. Filling the research gap, this research sought to perform an in-depth study through a qualitative method with participants of graduate students by involving the use of the interview as the primary method and listening to their stories. It comprises a preliminary descriptive examination of the perceptions and experiences of students in the graduate school of education an Association of American Universities (AAU) institution. Also, with a few exceptions (e.g. Bell (1988)), not much literature have compared the different aspects of international and domestic students when learning statistics. In this respect, the current research allocated the participants into either international or domestic groups. Finally, to further examine the difference within international students, a focus group interview followed.

References

- Baloglu, M.(2004). Statistics anxiety and mathematics anxiety: Some interesting differences. *Educational Research Quarterly*, 27(3), 38-48.
- Bell, J. A.(1998). International students have statistics anxiety too! *Education*, 118(4), 634.
- Bogdan, R. C., & Biklen, S. K.(2007). *Research for education: An introduction to theories and methods*. New York: Pearson.
- Bui, N. H., & Alfaro, M. A.(2011). Statistics anxiety and science attitudes: Age, gender, and ethnicity factors. *College Student Journal*. 45(3), 573-585.
- Earley, M.A.(2001). Improving statistics education through simulations: the case of the sampling distribution. *Paper presented at the Annual Meeting of the Mid-Western Educational Research Association*. Chicago, Illinois, October 24 - 27.
- Finney, S.J.(2000). The meaning of success for students in statistical methods courses: A phenomenological study. *Paper presented at the annual meeting of the American Educational Research Association*, New Orleans.
- Kottke, J. L.(2000). Mathematical proficiency, statistics knowledge, attitudes toward statistics, and measurement course performance. *College Student Journal*, 34(3), 334-334.
- Moore, D. S.(1997). New pedagogy and new content: The case of statistics. *International statistical review*, 65(2), 123-137.
- Morgan, D.(1997). *Focus groups as qualitative research*. 2nd edition. Thousand Oaks, CA: SAGE.
- NCTM.(2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Plake, B. S., & Parker, C. S.(1982). The development and validation of a revised version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement*, 42(2), 551-557.
- Wallman, K. K.(1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Wise, S. L.(1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*,45(2), 401-405.

Predicting the Effect of the Fear and Greed Index for Digital Assets and Stocks on the Trends of Apartment Sales by Artificial Neural Network

Hyun-Wook Park¹, Sung-Hyeon Bae²

Abstract

The psychology of stock buyers and sellers greatly affects stock prices, and fluctuating stock prices affect another asset, real estate prices. In addition, since the current cryptocurrency market is becoming huge, it is necessary to analyze the real estate market psychology with the psychological index of cryptocurrency. In this study, an artificial neural network model was implemented to predict the trend of apartment sales supply and demand in Korea based on the digital asset fear and greed index and the stock fear and greed index. The model is implemented in a form that minimizes the mean squared error by using a multi-layer neural network(Activation function: Swish).

Keywords : Fear and greed index, Apartment sales trend, ANN, Swish;

1. Introduction

Historically, changes in the stock market have had an impact on real estate asset management, and those two are closely related. In particular, changes in the stock market, such as the IMF, Lehman, and COVID-19, have caused major changes in the situation and price of real estate sales in Korea. In addition, the current market capitalization of domestic cryptocurrencies was estimated at 55.2 trillion won as of March 2022, which is similar to 55.53 trillion won of NAVER, the 4th largest stock market capitalization in Korea as of January 2022. Accordingly, the purpose of this study is to investigate the effect of the fear and greed index, which indicates investment sentiment and investment trends in the stock market and cryptocurrency market, on the trend of supply and demand of apartments in Korea.

2. Data and Methods

In this study, the digital asset fear and greed index of Dunamu Datavalue and the fear

¹(First author) B.S. in Computer Education, Sungkyunkwan University. 25-2, Seonggyungwan-ro, Jongno-gu, Seoul, 03063, Korea. E-mail: printpark@skku.edu

²B.S. in Computer Education, Sungkyunkwan University. 25-2, Seonggyungwan-ro, Jongno-gu, Seoul, 03063, Korea. E-mail: sunghyeon86@gmail.com

and greed index of CNN Business were selected, and the period is 2021.01~2022.05. The trend of supply and demand for apartment sales by the Korea Real Estate Board was used as a dependent variable, which is an indicator of the superiority of supply and demand in the apartment sales in Korea. CNN's fear and greed index is set based on 7 indicators (Market momentum, Stock price strength, Stock price breadth, Put and call options, Market volatility, Safe haven demand, Junk bond demand). Dunamu's index is calculated using the Upbit market index, Volatility-Volume score, and Momentum score. The apartment sales supply and demand index of the Korea Real Estate Board ranges from 0(high supply) to 200(high demand), so the value was adjusted to 0~100 to match the previous fear and greed index. Market volatility (VIX), an important indicator of the fear and greed index, represents the expected price movement sentiment over the next 30 days. According to 'Factors Affecting the Days on Market of Residential Real Estate in Seoul', the cumulative contract rate for apartments in Seoul is 58% on the 30th day. Therefore, we assumed that the sentiment on stocks and digital assets will affect the supply and demand for real estate sales in about one month. Therefore, we matched the fear and greed index of stock and digital assets with the trend of apartment sales supply and demand one month later.

The model was implemented through a multi-layer neural network with Tensorflow and Keras. Swish($f(x) = \frac{1}{1+e^{-x}}x$, $f'(x) = f(x) + \text{sigmoid}(x)(1-f(x))$) was used for the activation function. Two or three hidden layers and their accompanying units and proper epochs were determined through the loss test below, and the final selected model summary is as follows.

Table 1. Fear and Greed Index

	Extreme Fear	Fear	Neutral	Greed	Extreme Greed
Digital Asset (Dunamu)	0~20	20~40	40~60	60~80	80~100
Stock (CNN)	0~25	25~45	45~55	55~75	75~100
Apartment (Korea Real Estate Board)	0~50		50~100		

Table 2. Loss test of each model

First hidden layer units	16	32	32	64	64	64	128	128	16	32	64	128	256
Second hidden layer units	16	32	64	32	64	128	64	128	16	32	64	128	256
Third hidden layer units	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	16	32	64	128	256
Loss	8.27	7.98	7.49	7.65	7.70	7.63	7.91	8.24	8.09	7.57	7.45	7.29	7.56

※(Loss: Mean squared error, Activation function: Swish, Validation_split: 0.3, epochs: 1,000)

Table 3. Model Summary

Layer(type)	Output Shape		Param #
Input layer	(None, 2)		0
First hidden layer	Dense	(None, 128)	384
	Batch Normalization	(None, 128)	512
	Activation	(None, 128)	0
Second hidden layer	Dense	(None, 128)	16512
	Batch Normalization	(None, 128)	512
	Activation	(None, 128)	0
Third hidden layer	Dense	(None, 128)	16512
	Batch Normalization	(None, 128)	512
	Activation	(None, 128)	0
Output layer	(None, 1)		129
Total params: 35,073			
Trainable params: 34,305			
Non-trainable params: 768			

3. Results

As a result of performing deep learning with the above model, model prediction became possible. The above model predicted the trend of apartment sales supply and demand in Korea through the digital asset fear and greed index and stock fear and greed index in 2022.04.19~2022.05.02. The comparison between the predicted value and the actual value is as follows.

Table 4. Predicted and actual value of trends in the supply and demand of apartment sales in Korea

Date	05-17	05-18	05-19	05-20	05-23	05-24	05-25	05-26	05-27
Prediction	48.2827	48.2277	48.1850	47.9241	47.9028	47.7621	47.5350	47.7563	47.7362
Actual	47.0310	47.0262	47.0214	47.0165	47.0117	46.9998	46.9879	46.9760	46.9642

4. Conclusion

In this study, we analyzed and predicted the trends in the supply and demand of apartment sales in Korea through the multi-layer neural network analysis of digital asset fear and greed index and stock fear and greed index. Although there are some errors, the trend predicted by the artificial neural network model and the trend of the actual data were similar with a downward trend. As a result, if there is a longer period of fear and greed index, and a more sophisticated deep learning model, it will be possible to roughly predict the trend of supply and demand of apartment sales in Korea one month or a certain period from the time of the investigation. Although this study focused on investor psychology, it seems necessary to add and analyze various factors that affect real estate sales, such as changes in policy and interest rates.

References

- Ramachandran. P., Zoph. B., Le. Quoc. V. (2017). Searching for activation functions, *Google Brain*, 5-10.
- Li. C. A., Wang. J. C.. (2013). The influences of greed and fear on fund performance, *The International Journal of Business and Finance Research*, 7(5), 47-55.
- Dunamu Datavalue Team. (2021). Digital assets fear and greed index methodology and rules version 2.0, 3-32. (in Korean)
- Sim. S. Y. (2019). Global digital asset market trend, *Korea Capital Market Institute*, 1-6. (in Korean).
- Ko. J. S., Choi. S. H., Noh. S. C. (2019). Factors Affecting the Days on Market of Residential Real Estate in Seoul, *Journal of the Korean Housing Studies Review*, 27(1), 05-30. (in Korean)

Decomposition of the Dividend Model: from the Perspective of ESG and Firm Characteristics

Su-In Kim¹, Injoong Kim²

Abstract

Over the past decade, firms of good ESG ratings, B+ and above, have shown consistently higher propensities to pay dividends compared to the poor ESG firms. The counterfactual decomposition of Maarten(2010) decomposes the dividend difference between the good and the poor ESG group into the part that is attributable to firm characteristics and the residual part that is contributed by the pure ESG group effect. Our results indicate that there exists a significant difference in dividend payments between the two ESG groups that can not be explained by the traditional dividend model of Fama and French(2001). Good ESG firms are generally more mature, larger, profitable and less risky firms, all of which are the typical firm characteristics of dividend payers. Our hypothetical decomposition result suggests that the explanatory power of the firm characteristics is 87.6% and the remaining 12.4% is the ESG group effect. In more recent years, this portion of the ESG group effect has increased to 18.4% and particularly, the ESG group effect is most pronounced for the KOSDAQ firms for which the ESG group effect occupies 45.9% of the total effect. This suggests that acquiring good ESG ratings can make noticeable difference in adopting more shareholder friendly dividend policy.

Keywords : dividend, ESG, decomposition, firm characteristic, KOSPI, KOSDAQ.

1. Introduction

Environmental, social, and governance(ESG) criteria are a set of non-financial standards for a company's behavior and they are becoming more important in recent years. This paper investigates the pure effect of the ESG ratings on dividend policy by decomposing the mean difference in dividends into the part that is contributed by the firm characteristics and the residual part that is attributable to the pure ESG group effect.

¹Assistant Professor, Accounting Major, College of Business Management, Hongik University, 2639 Sejong-ro, Sejong, 30016, Korea. E-mail: suinkim@hongik.ac.kr

²(Corresponding Author) Assistant Professor, Finance & Insurance Major, College of Business Management, Hongik University, 2639 Sejong-ro, Sejong, 30016, Korea. E-mail: kij@hongik.ac.kr

2. Empirical Method

This study analyzes ESG rated firms' dividend payments for the period of 2011-2021 based on the following Jann (2008) and Maarten's (2010) mean decomposition technique.

$$E[Y_A] - E[Y_B] = \{E[X_A] - E[X_B]\}'\beta^* + E[X_A]'(\beta_A - \beta^*) + E[X_B]'(\beta^* - \beta_B) \quad (1)$$

The first term on the right hand side of the equation represents the portion contributed by the heterogeneity of firm characteristics and the remaining terms capture the ESG group effect. The base line dividend forecast model of $Y = X'\beta + \epsilon$, in which the dividend predicting explanatory vector X contains the following firm characteristics, is based on the traditional dividend model of Fama and French(2001).

$$\Pr(\text{Payer} = 1) = \text{Logit}(\alpha + \beta_1 \text{Size} + \beta_2 \text{Profitability} + \beta_3 M/B + \beta_4 \text{Risk} + \beta_5 RE/TE) + \epsilon \quad (2)$$

3. Main Results

Figure 1 plots the proportion of dividend payers for the good ESG and the poor ESG group over the period from 2011 to 2021. Firms of higher ESG ratings, B+ and above, show consistently higher propensity to pay dividends compared to the poor ESG counterpart.

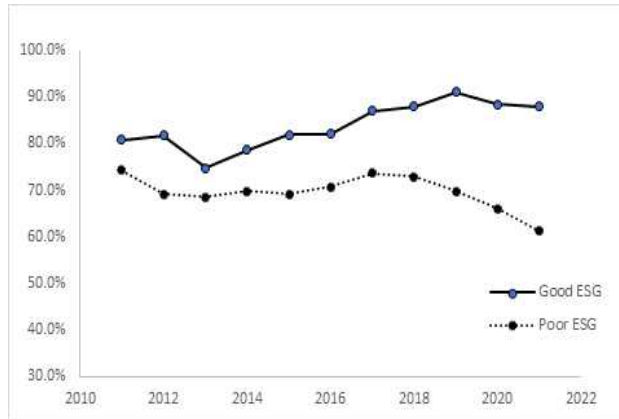


Figure 1. Proportion of dividend payers

The systematic difference in dividend payments in Figure 1, of course, will be largely influenced by the difference in firm characteristics of each ESG group. Table 1 summarizes the average of firm characteristic variables from the traditional dividend forecast model and tests the statistical difference between the two groups. As we expect, firms that belong to the good ESG group are more mature, larger, profitable, and less risky firms, all of which are the typical characteristics of dividend payers.

Table 1. Means of firm characteristic variables in each ESG group

	Dividend	Size	Profitability	M/B	Risk	RE/TE
Good ESG	0.852	21.926	0.049	1.552	3.325	0.558
Poor ESG	0.694	19.715	0.032	1.559	3.863	0.408
Total	0.728	20.201	0.036	1/557	3.745	0.441
Good – Poor	0.158*	2.211*	0.017*	-0.007	-0.538*	0.150*
P-value	(<0.001)	(<0.001)	(<0.001)	(0.897)	(0.005)	(<0.001)

T-tests for the mean difference between the good and the poor ESG group are reported.

Table 2. Logistic regressions of Good/Poor ESG group

Predictors	All Sample		Good ESG		Poor ESG	
	Coefficient	Std err	Coefficient	Std. err.	Coefficient	Std. err.
Intercept	-6.614*	1.045	-3.995*	1.573	-5.970*	1.247
Size	0.226*	0.027	0.092*	0.037	0.293*	0.040
Profitability	13.993*	0.922	19.473*	2.747	13.025*	0.981
M/B	-0.080*	0.027	-0.180*	0.057	-0.049*	0.020
Risk	-0.244*	0.030	-0.134*	0.051	-0.241*	0.032
RE/TE	2.539*	0.116	2.170*	0.301	2.632*	0.130
Log likelihood	-2,550.7		-756.2		-2,050.8	

Note: The market to book ratio, M/B represents the growth opportunity and the retained earnings over total equity, RE/TE measures the lifecycle of a firm.

Table 3. Factual and counter-factual proportions of dividends

Firm characteristics	Entire sample		KOSPI		KOSDAQ	
	Poor ESG	Good ESG	Poor ESG	Good ESG	Poor ESG	Good ESG
Poor ESG	0.694	0.756	0.701	0.757	0.575	0.656
Good ESG	0.862	0.852	0.871	0.851	0.683	0.857

Table 4. Decomposition of ESG and firm characteristic effect

Panel A. Log odds ratios						
	Entire sample		KOSPI		KOSDAQ	
	Estimated coefficient	Bootstrap Std. error	Estimated coefficient	Bootstrap Std. error	Estimated coefficient	Bootstrap Std. error
Total effect	0.929*	0.071	0.891*	0.077	1.489*	0.430
Firm characteristic1	0.616*	0.102	0.606*	0.090	1.146*	0.437
Residual1	0.313*	0.115	0.285*	0.103	0.344*	0.148
Firm characteristic2	1.011*	0.077	1.055*	0.074	0.465*	0.125
Residual2	-0.082	0.059	-0.164	0.090	1.024*	0.401
Panel B: Proportions of firm characteristics						
Method1	0.663*	0.114	0.680*	0.107	0.769*	0.279
Method2	1.089*	0.108	1.184*	0.116	0.312*	0.117
Aggregate	0.876*	0.089	0.932*	0.092	0.541*	0.176
ESG group effect	0.124		0.068		0.459	

Table 2 presents the link between dividend payments and firm characteristics. The logit regression results are consistent with the previous findings. Firm size, profitability and the lifecycle measure(RE/TE) are positively related with dividends, while risk and the investment opportunity are negatively related with dividends. Over the whole sample period of 2011~2021, the actual proportion of dividend payments is 85.2% for the good ESG group and 69.4% for the poor ESG group, respectively. Table 3 decomposes the difference in the dividend payments into the effect of firm characteristics and the residual effect of the ESG ratings. We mainly focus on the ESG effect(horizontal movement) and the most dramatic effect of ESG can be observed for the KOSDAQ subsample, where the dividend probabilities increase from 57.5% to 65.6% and from 68.3% to 85.7% for the poor and the good ESG firm characteristic group, respectively.

Table 4 measures the contribution of the ESG group effect by decomposing the mean difference of dividends between the good and the poor ESG group. For example, based on the equation 1), out of the 0.929 total effect, the difference in firm characteristics explains 0.616/0.929, or 66.3% by the method 1 and the residual portion 0.313/0.929, or 33.7% is attributable to the residual part. In the actual estimation of the equation 1), the unobservable ideal vector β^* is replaced with the observable vector $\beta^* = \hat{\beta}_A$ and $\beta^* = \hat{\beta}_B$, for the method 1 and 2. The average of the two methods, 87.6% is attributable to the difference in firm characteristics and the remaining 12.4% is the ESG group effect.

For the subsample of the 2017~2021, the above ESG group effect increases to 18.4%, 10.8% and 46.0% from 12.4%, 6.8% and 45.9%, respectively.

4. Conclusion

Our results suggest that there exists systematic difference in firm characteristics between the good and the poor ESG firms. Good ESG firms are at the mature stage of firm lifecycle, and they are generally larger, more profitable and less risky firms compared to the poor ESG counterparts. The systematic difference in firm characteristics explains 87.6 % of the dividend difference and the remaining 12.4% is the role of ESG ratings. This figure increases to 18.4% for the more recent year sample. It is likely that non-financial ESG criteria can drive firms to adopt more shareholder friendly payout policy.

References

- Lim, B. K., Kim, C. K. (2022). The effect of consumer centered management on financial performance and ESG level, *Journal of The Korean Data Analysis Society*, 24(2), 811-822. (in Korean).

SNS의 프라이버시 염려에 관한 연구:정교화 가능성 모델을 중심으로

김유진¹, 이형석²

요약

본 연구는 소셜네트워크서비스(SNS)에서 정교화 가능성 모델을 적용하여 이용자들의 정보 프라이버시에 관한 염려에 영향을 미치는 요인을 파악하여 가설을 도출하였다. SNS의 정보 프라이버시 염려에 영향을 미치는 요인으로 중심경로에는 정보품질, 서비스품질, 프라이버시 정책 세 가지 요인을 도출하였으며, 주변경로에는 지각된 SNS 앱 명성, 낙관주의의 두 가지 요인을 도출하였다. 본 연구를 통해 SNS 기업들이 사용자의 정보 프라이버시 보호 행동을 이해하는 데에 도움을 줄 것으로 본다. 또한 정교화 가능성 모델을 SNS 프라이버시 연구에 적용하였다는 데 의의가 있다.

주요용어 : SNS, 프라이버시 염려, 정교화 가능성 모델.

1. 서론

스마트폰의 일상화와 함께 소셜 네트워크 서비스(social network service: 이하 SNS)는 하나의 문화로 자리 잡았다. 이용자들은 SNS를 통해 이용자들 간 의사소통 외에도 다양한 정보를 공유, 인맥 확대 등을 통해 사회적 관계를 생성하고 강화한다. 하지만 현재에는 메타로 사명이 변경된 페이스북의 개인정보 유출 사건뿐만 아니라 온라인에서 특정인의 개인정보를 찾아내어 유포하는 일명 ‘신상털기’가 심각한 사회 문제로 주목받으면서 SNS 이용자들의 프라이버시에 관한 염려가 점차 높아지고 있다(Lim, Lee, 2019). 이용자들의 프라이버시에 대한 염려가 점차 높아지는 상황에서 이용자들이 개인정보 제공과 관련된 의사결정을 할 때 어떤 요인이 중요한 역할을 하는지 파악해 볼 필요가 있다(Kim, Kim, 2016).

따라서 본 연구에서는 기존의 SNS 및 프라이버시 연구를 검토한 뒤, 정교화 가능성 모델을 중심으로 SNS에서 프라이버시 염려에 영향을 미치는 요인을 파악하여 연구모형 및 가설을 도출하였다. 본 연구를 바탕으로 SNS 기업들이 이용자의 프라이버시 염려를 이해하고 효과적으로 관리하는데 도움을 줄 것으로 기대한다.

2. 이론적 배경 및 가설

¹(제1저자) 28644 충청북도 청주시 서원구 충대로1, 충북대학교 대학원 경영학과 석박사통합과정수료.

E-mail: yjk824@naver.com

²(교신저자) 28644 충청북도 청주시 서원구 충대로1, 충북대학교 경영대학 경영학부 교수.

E-mail: hyungle@chungbuk.ac.kr

2.1. SNS 환경에서 정보 프라이버시 염려

기존 온라인 환경에서 사용자의 프라이버시 피해 경험은 웹사이트에 제공한 자신의 개인정보나 금융정보 등이 유출되어 오용 및 남용될지도 모른다는 염려를 유발한다(Kim, Kim, 2013). 하지만 SNS 환경은 기존 온라인 환경의 정보 프라이버시 염려와는 다른 시각으로 바라보아야 한다. SNS 사용자들은 자발적으로 자신의 일상과 관련된 정보를 업로드하고 다른 사용자들과 커뮤니케이션을 하며 관계를 유지하기 때문이다. 수많은 사용자들이 서로 연결되어 있어 개인이 공개하는 정보의 범위가 매우 광범위하기 때문에 개인정보가 유출될 경우 여러 가지 심각한 문제가 발생할 확률도 매우 높다(Park, Kwahk, 2019). 이처럼 SNS 환경에서 정보 프라이버시 염려는 사용자가 인지하는 위험과 같은 신념을 형성할 때 중요한 역할을 할 뿐만 아니라, 프라이버시 행동에 대한 의사결정에 직접적인 영향을 미친다(Min, Hwang, 2016).

2.2. 정교화가능성모델

Petty, Cacioppo(1986)이 제시한 정교화 가능성 모델은 메시지가 중심 경로(central route)와 주변 경로(peripheral route)라는 두 가지 경로를 통해 처리된다고 가정하는 설득의 이중 경로 모델이다. 이때 메시지의 질이나 관여도와 같은 요소에 의해 메시지가 어떤 경로로 처리되는지 결정된다. 메시지가 높은 정교화 수준을 가진다면 중심경로를 통해 메시지가 주의를 기울여 정보를 처리한다. 반면, 메시지가 낮은 정교화 수준을 가진다면 주변 경로를 통해 주의를 기울이지 않고 정보를 처리한다. 또한 중심 경로는 내부적 평가를 통해 메시지와 관련된 정보를 처리하는 반면 주변 경로는 외부 단서를 중심으로 처리한다(Kroenung, Eckhardt, 2017).

정교화 가능성 모델은 정보시스템 연구에서도 확장되었다. Bhattacharjee, Clive(2006)은 정교화 가능성 모델을 바탕으로 IT 사용 의도에 있어서 중심 및 주변 경로를 비교하였다. Zhou(2012)는 모바일 बैं킹에서 사용자의 초기 신뢰 프로세스를 정교화 가능성 모델로 설명하였다. 뿐만 아니라, Zhou(2017)는 위치기반 서비스 사용자의 프라이버시 염려를 효과적으로 이해하기 위해 중심 및 주변 경로에 대한 영향 요인에 프라이버시 조절 요인을 조절변수로 포함하여 연구를 진행하였다. Gu, Xu, Xu, Zhang, Ling(2017)은 모바일 앱 다운로드의 프라이버시 염려를 정교화 가능성 모델로 설명하였다. Kim, Oh(2019)은 정교화 가능성 모델을 바탕으로 정보 처리와 프라이버시 역설에 관한 연구를 진행하였다. 따라서 본 연구는 정교화 가능성 모델의 선행 연구를 바탕으로 중심경로에는 정보품질, 서비스 품질, 프라이버시 정책이라는 세 가지 요인을, 주변경로에는 지각된 SNS앱 명성, 낙관주의라는 두 가지 요인을 도출하여 다음과 같은 가설을 설정한다.

- H1: 정보 품질은 프라이버시 염려에 유의한 영향을 미칠 것이다.
- H2: 서비스 품질은 프라이버시 염려에 유의한 영향을 미칠 것이다.
- H3: 프라이버시 정책은 프라이버시 염려에 유의한 영향을 미칠 것이다.
- H4: 지각된 SNS앱 명성은 프라이버시 염려에 유의한 영향을 미칠 것이다.
- H5: 낙관주의는 프라이버시 염려에 유의한 영향을 미칠 것이다.

3. 연구모형

본 연구는 선행 연구를 바탕으로 SNS 사용자의 정보 프라이버시 염려에 영향을 미치는 구성개념으로 정교화 가능성 모델에서 도출한 정보 품질, 서비스 품질, 프라이버시 정책, 지각된 SNS 앱 명성, 낙관주의로 설정하고 정보 프라이버시 염려와의 인과관계를 설정하였다. 앞서 기술한 가설들에 대한 연구모형은 Figure 1과 같다.

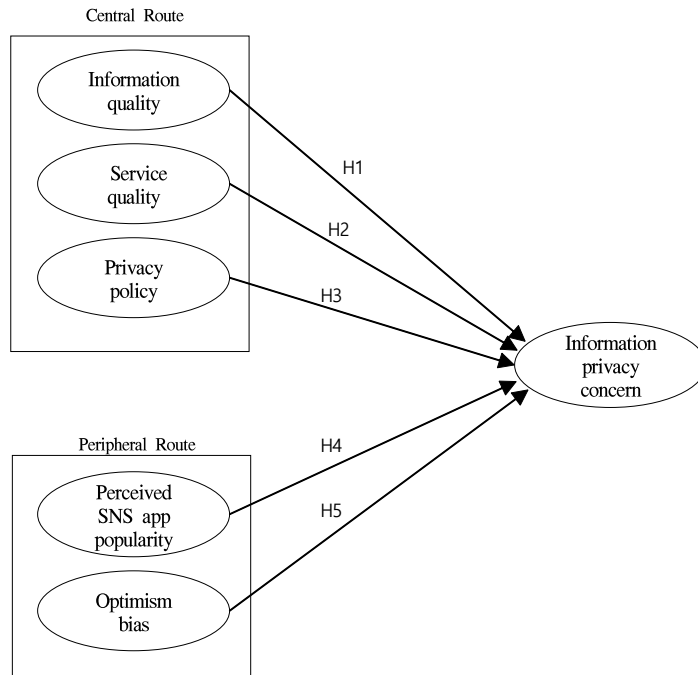


Figure 1. 연구모형

4. 연구 제언

본 연구는 기존 선행연구들을 바탕으로 SNS 사용자의 프라이버시 염려에 영향을 미치는 요인들을 도출하여 그들 사이의 관계에 관한 가설을 설정하였다. SNS 사용자들의 불안감이 점점 커지는 상황에서 본 연구를 통해 기업들은 SNS 사용에 영향을 미치는 요인들을 파악하고 이해함으로써 이들을 어떻게 관리해야 하는지 알 수 있다. 정교화 가능성 모델을 SNS 환경에 적용한 본 연구는 학술적으로도 의미가 있을 것으로 본다.

앞으로 본 연구의 가설들을 실증적으로 분석하기 위해 SNS 사용자들을 대상으로 설문조사를 통해 자료를 수집하여 인과관계에 대한 가설들을 검증할 것이다. 이러한 통계적 가설검정 과정을 통해 SNS 사용 시 이용자들의 정보 프라이버시 염려에 중요한 영향을 미치는 요인들이 무엇인지 확인할 수 있다.

References

- Bhattacharjee, A., Sanford, C. (2006). Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model, *MIS Quarterly*, 30(4), 805-825.
- Gu, J., Xu, Y., Xu, H., Zhang, C., Ling, H. (2017). Privacy Concerns for Mobile App Download: An Elaboration Likelihood Model Perspective, *Decision Support Systems*, 94, 19-28.
- Kim, J. K., Kim, S. H. (2013). Privacy Protection Behavior of Online User: Based on Privacy Paradox Perspective, *The Journal of Internet Electronic Commerce Research*, 13(1), 41-64. (in Korean).
- Kim, J. K., Oh, D. W. (2019). A Study on the Dual Process in the Information Processing and Privacy Paradox, *The Journal of Internet Electronic Commerce Research*, 19(4), 1-21. (in Korean).
- Kim, S. H., Kim, J. K. (2016). A Study on Factors Influencing Privacy Decision Making on the Internet: Focus on Dual-Calculus Model, *The Journal of Informaion Systems*, 25(3), 197-215. (in Korean).
- Kroenung, J., Eckhardt, A. (2017). What Determines User Attitudes in IS Research? A Meta-analytic Structural Equation Modeling Approach, *Journal of Information Technology Theory and Application*, 18(3), 34-65.
- Lim, T., Lee, H. S. (2019). Factors Affecting Intention to Disclose Personal Information and Continuous Usage Intention in Social Network Service, *The Journal of Internet Electronic Commerce Research*, 19(1), 17-38. (in Korean).
- Min, H. H., Hwang, G. H. (2016). The effect of Privacy Factors on the Provision Intention of Individual Information from the SNS Users, *Journal of Digital Convergence*, 14(12), 1-12. (in Korean).
- Park, J. Y., Kwahk, K. Y. (2019). Motivational Factors Affecting Self-Disclosure Behavior of SNS Users, *Korean Management Review*, 48(2), 561-587. (in Korean).
- Petty, R. E., Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion, *In communication and persuasion*, Springer, New York:NY, 1-24.
- Zhou, T. (2012). Understanding Users' Initial Trust in Mobile Banking: An Elaboration Likelihood Perspective, *Computers in Human Behavior*, 28, 1518-1525.
- Zhou, T. (2017). Understanding Location-Based Services Users' Privacy Concern: An Elaboration Likelihood Model Perspective, *Internet Research*, 27(3), 506-519.

제한폭 확대 영향 연구

이은광¹, 형남원²

요 약

시장안정화장치 중 하나인 가격제한폭은 불공정거래, 시장 효율성, 투자행태 및 감시비용 등 시장 전반에 영향을 주며, 국내를 포함한 아시아 전역에서 현재까지 시행되고 있다. 본문에서는 제한폭 확대에 의해 변동성이 어떻게 변화하는지를 식별함으로써 확대 이전에 발생한 정보열위 투자자의 과반응을 억제하지 못해 변동성이 증가하는지, 아니면 확대 이전 투자자의 불안심리에 의해 조성된 균형가격의 불확실성을 완화해 주어 변동성이 감소(또는 변화없음)하는지 확인하였다. 그 결과, 제한폭 확대될 경우 확대 이전의 변동성 수준에 비해 총변동성과 체계적 위험에 감소한 것으로 나타나, 과반응가설을 따르지 않은 것으로 나타났다. 이는 과반응에 의해 조성된 불안심리로 인한 가격폭락의 촉진과 시장변동성의 증가를 방지해주는 혜택이 시장참여자에게 제공됨을 시사한다.

주요용어 : 가격제한폭, 이변량 독립적 분류, 가치 가중 포트폴리오, 변동성, 과반응가설.

1. 서 론

1987. 10. 19. 미국발 블랙먼데이 발생 이후 시장안정화를 위한 미국 정부의 노력으로 많은 제도 마련과 그 효과에 대해 경제학자들이 연구를 진행해 왔다. 이 중 제한폭 제도는 상장증권의 공정한 가격형성을 도모하고 급격한 시세변동에 따른 투자자 피해방지와 공정한 거래질서를 확립하기 위해 제도를 도입한 이래로 정률제 기준, 총 7차례에 걸쳐 제한폭을 확대해 왔다. 제한폭의 긍정적인 견해로는 주문 불균형을 비롯한 여러 거시경제적인 이유로 발생한 과도한 위험으로부터 가격제한폭에 의해 손실 일부를 방지해준다. 또한, Ma, Rao, Sear(1989)가 제시한 과반응 가설에 따르면 주가가 제한폭에 도달한 경우 냉각기간을 제공하여 정보우위 투자자의 전략을 동일하게 취한 정보열위 투자자에게 패닉에 따른 과반응을 줄이며, 시장정보를 재평가 및 투자전략을 보완함으로써 결과적으로 변동성을 정상수준으로 줄여준다. 반면, 부정적인 견해로 제한폭이 확대될수록 제한폭의 완충이 완화됨에 따라 주문불균형을 비롯한 여러 거시경제적인 요인으로 인해 발생할 수 있는 시장참여자들의 투자손실은 제한폭 확대 이전보다 커질 수 있다. 그리고 만일 제한폭이 과반응을 줄여준다면, 제한폭을 확대한 경우 투자자들의 과반응을 억제하지 못해 변동성이 커질 수 있다.

이렇듯 연구자마다 주장하는 바도 다양하며, 시장효율성 측면에서 제한폭이 도입되거나 변경된

¹02504 서울특별시 동대문구 서울시립대로163 서울시립대학교 일반대학원 경제학과 박사과정.

E-mail: lee2137@uos.ac.kr

²(교신저자) 02504 서울특별시 동대문구 서울시립대로163 서울시립대학교 경제학과 교수.

E-mail: nhyung@uos.ac.kr

다고 하여 변동성이 효과적으로 통제되는지 알기 어렵다. 이에 본 저자는 과반응 가설검정을 기반한 실증분석을 통해 제한폭 확대가 내포한 경제적 의미와 제한폭 확대가 시장참여자들에게 어떤 기능하는지를 밝혀내고자 한다. 정책효과 식별을 위해 처치군과 비교군의 변동성을 비교하는 방식에 이어, 변동성에 영향을 줄 수 있는 기업특성, 레버리지, 표본대상, 표본기간, 변동성의 특성(비대칭성, 비교시점 및 과소추정 사례)들을 통제하여 정책효과를 식별해 낸 점, 그리고 제한폭이 본질적으로 어떤 위험에 영향을 주었는지 밝혀낸 점들은 분명 남다른 의미를 가질 것이다.

2. 자료 및 분석모형

본문에서는 모든 정률제 확대사례 7건 중에서 처치군과 대조군을 비교하기 위한 목적으로 특정 시장에서만 제한폭이 확대된 1998. 03. 02.(Kospi, 8→12%), 1998. 05. 25.(Kosdaq, 8→12%), 1998. 12. 07.(Kospi, 12→15%), 2005. 03. 28.(Kosdaq, 12→15%) 네 건의 사례와 2015. 06. 15.(Kospi & Kosdaq, 15→30%)의 최근 사례를 포함하였으며, 그 중에서도 외생적 충격이 적은 2005. 03. 28. 사례를 중점적으로 보았다.

그리고 가격제한폭은 증가시점을 기준으로 조정되었으므로, 본문에서는 추가분할, 유무상증자, 배당금 등이 조정된 수정종가 수익률을 사용하였다. 이를 바탕으로 증시별 지수수익률과, 기업특성에 따른 변동성의 차이를 통제하기 위해 각 증시별 포트폴리오 단위로 3×3 이변량 독립적 분류(3:4:3 비율을 적용함)를 진행한 이후에 기업규모를 가중한 포트폴리오 수익률들을 각각 구성하여 AR(0-5)-EGARCH(1,1,1) 모형에서 BIC를 기준으로 최적모형을 선별하여 비교하였다.

여기서 종속변수는 모두 정상적인 조건(stationary condition)을 충족하며, 이때 과반응 반대가설을 검정한 무조건부 분산 차이(①)는 다음과 같다.

$$\text{평균방정식: } y_t = \alpha + \phi \sum_{i=0}^5 y_{t-i} + \varepsilon_t, \quad \varepsilon_t = z_t \sqrt{h_t}, \quad \varepsilon_t | \Omega_{t-1} \sim N(0, h_t)$$

$$\text{분산방정식: } \ln h_t = \omega + \gamma(|\xi_{t-1}| + E|\xi_{t-1}|) + \theta \xi_{t-1} + \beta \ln h_{t-1} + \delta D_t,$$

$$\xi_t \equiv \frac{\varepsilon_t}{\sqrt{h_t}}, \quad g(\xi_t) = \theta \xi_t + \gamma(|\xi_t| - E|\xi_t|), \quad D_t = \begin{cases} 0, & \text{확대 전} \\ 1, & \text{확대 후} \end{cases}$$

$$H_0: \exp\left(\frac{\delta}{1-\beta}\right) = 1, \quad H_1: \exp\left(\frac{\delta}{1-\beta}\right) < 1 \cdots \text{①}$$

과반응 반대가설 : 제한폭 확대 시 변동성은 증가하지 못함

H0 : 제한폭 확대 시 변동성은 변화 없을 것이다.

H1 : 제한폭 확대 시 변동성은 감소할 것이다.

3. 실증분석 및 해석

Ma, Rao, Sear(1989)가 제시한 과반응 가설에 대해 반대한 Fama(1989)와 Kyle(1988)는 확대 전 증시 내 조성된 시장참여자들의 불안심리가 제한폭 확대에 의해 완화되어 균형가격에 대한 불확실성이 감소함으로써 변동성이 줄어든다고 주장하였다. 본문은 과반응 반대가설을 검정하기 위해 국내

증시를 대상으로 Dataguide(Ver. 5)에서 제공한 일별자료 중에서 제한폭 확대된 시점들을 기준으로 표본기간을 차례로 확장하였다. 그리고 높은 레버리지를 가지는 금융주를 배제하고 남은 상장회사, 총 2,761개를 대상으로 Stata(Ver. 15. 1)로 분석하였다.

Table 1. Estimated results of unconditional variance difference between Kospi and Kosdaq

Indexes	①_pi_120	①_daq_120	①_pi_240	①_daq_240
1998. 03. 02.	-0.67(0.11) ***	-0.05(0.05)	-0.34(0.12) ***	0.49(0.49)
1998. 05. 25.	1.17(0.49) ***	-0.12(0.37)	0.00(0.01)	-0.55(0.15) ***
1998. 12. 07.	0.43(0.32) *	1.47(0.67) **	0.17(0.31)	4.53(1.47) ***
2005. 03. 28.	-0.10(0.31)	-0.61(0.10) ***	-0.30(0.12) ***	-0.38(0.18) **
2015. 06. 15.	1.40(0.53) ***	2.83(6.26)	1.13(0.34) ***	1.67(0.85) **

Note : This table is for the Kospi and Kosdaq, the sampling period is divided into 120 days (60 business days before and after the limit) to 240 days. ① was estimated by the Delta method, standard errors in huber/white/sandwich estimator are in parentheses. and ***, **, and * are significant at the 1%, 5%, and 10% significance levels, respectively.

Kospi에서 제한폭이 확대된 1998. 03. 02. 사례를 보자면 Kospi의 무조건부 분산 차이는 -0.67에서 5% 유의수준 내에서 통계적으로 유의했으며, Kosdaq의 무조건부 분산 차이는 -0.05이나 통계적으로 유의하지 않아 제한폭이 확대된 Kospi의 무조건부 분산 차이가 비교적 작게 나타났다. 또한 Kospi에서 확대된 1998. 12. 07. 사례와 Kosdaq에서 확대된 1998. 05. 25., 2005. 03. 28. 사례에서도 ‘처치군-대조군 비교방식’을 통해 해석할 경우 제한폭 확대가 시행되지 않은 증시에 비해 가격제한폭 확대를 시행한 증시에서 총 변동성이 상대적으로 감소함을 보였다. 이는 표본기간을 ±120로 확장된 경우에서도 동일한 양상으로 나타났다.

Table 2. Robustness results of unconditional variance difference between Kospi and Kosdaq

port	①_pi_120	①_daq_120	①_pi_240	①_daq_240
1,1	-0.85(0.04) ***	-0.95(0.71) *	-0.76(0.14) ***	1.18(5.35)
1,2	4.3E+13(4.6E+16)	-0.69(0.10) ***	0.19(0.58)	1.77(5.36)
1,3	-0.21(0.19)	-0.78(0.28) ***	-0.31(0.64)	-0.10(0.19)
2,1	-3.7E-5(5.3E-5)	-0.62(0.11) ***	0.05(0.17)	-0.63(0.40) *
2,2	0.43(0.96)	-0.80(0.19) ***	0.40(0.66)	0.19(0.37)
2,3	4.99(29.99)	-0.69(0.10) ***	-0.64(1.66)	-0.21(0.27)
3,1	-0.40(3.84)	-0.38(0.15) ***	-0.49(0.08) ***	-0.86(0.82)
3,2	-0.15(0.72)	-0.80(0.05) ***	-0.11(0.45)	-0.56(0.11) ***
3,3	-0.67(0.19) ***	0.39(0.56)	0.55(1.17)	-0.02(0.06)

Note : This table is for portfolios in a Kospi and Kosdaq on March 28, 2005, the sampling period is divided into 120 days (60 business days before and after the limit) to 240 days. Portfolio is [i=Firm size, j=B/M, 1=low, 3=high], ① was estimated by the Delta method, standard errors in huber/white/sandwich estimator are in parentheses. and ***, **, and * are significant at the 1%, 5%, and 10% significance levels, respectively.

Table 2에서는 Kim, Limpaphayom(2000), Kim Liu and Yang(2013) 주장처럼 기업특성이 변동성에 영향을 줄 수 있으므로 제한폭 확대에 의한 각 증시별 포트폴리오 총 변동성 및 고유변동성 변화에 대해서도 1998년 외환위기에 비해 외생적 충격이 상대적으로 적은 2005. 03. 28. 사례에 한해 분석하였다. 분석결과 Kosdaq 증시 내 port_3,3들의 소속기업 1개로 매우 적은 포트폴리오를 제외하면 port_1,2, port_1,3, port_2,1, port_2,2, port_2,3, port_3,2에서는 Kosdaq의 제한폭이 확대되었음에

도 KOSPI 대비 Kosdaq의 무조건부 분산 차이가 감소하는 경향을 보여, 포트폴리오에서도 대체로 제한폭 확대에 따라 총 변동성이 감소한다는 강건한 결과를 보였다. 다만 표본기간을 확장한 경우에서 결과가 달리 나타난 이유는 제한폭 확대는 지속성을 가져 제한폭 확대 효과는 존재함에도 불구하고, 다른 거시경제적 충격이 개입됨으로 인해 제한폭 확대 효과가 희석되는 것으로 나타났다.

Table 3. Estimated results for unsystematic risk in Fama-French 3 factor model

port	①_pi_120	①_daq_120	①_pi_240	①_daq_240
1,1	-0.47(0.08) ***	1.15(13.06)	-0.82(0.20) ***	-0.45(0.51)
1,2	1.7E-4(7.1E-5)	-0.67(0.09) ***	5.9E-5(4.7E-5)	0.29(0.43)
1,3	-0.71(0.59)	-0.02(0.13)	9.3E+4(1.6E+7)	0.32(0.39)
2,1	-0.35(0.15) **	9.3E-4(2.8E-4) ***	-0.12(0.11)	-0.39(0.11) ***
2,2	-0.71(0.12) ***	-0.67(0.20) ***	-0.37(0.13) ***	0.12(0.18)
2,3	-1.00(0.02) ***	-0.75(0.44) **	-0.27(0.25)	-2.3E-3(0.03)
3,1	-0.46(0.36) *	-0.66(0.09) ***	-0.04(0.16)	-0.34(0.17) **
3,2	-0.47(1.34)	-0.73(0.08) ***	2.1E-5(3.6E-5)	-0.65(0.11) ***
3,3	-0.73(0.06) ***	0.15(0.51)	0.22(0.23)	-0.07(0.23)

Note : This is the result of comparing idiosyncratic volatility within the three-factor model proposed by Fama and French. ① was estimated by the Delta method, standard errors in huber/white/sandwich estimator are in parentheses. and ***, **, and * are significant at the 1%, 5%, and 10% significance levels, respectively.

엄경식·강형철·이윤재(2008)은 제한폭 제도를 채택한 거래소와 정책당국자의 주요목표가 분산 가능한 고유변동성에 의해 초래되는 시장의 불확실성을 감소하므로, 고유변동성의 변화를 식별해내는 것이 중요하다고 밝혀, Table 3 고유변동성에 대한 결과도 제시하였다. 분석결과, 대조군 증시에 비해 처치군 증시 내 port_1,1, port_2,1, port_2,2, port_2,3, port_3,3에서 유의미하게 증가하였고, port_1,2, port_3,1, port_3,2에서는 감소하였으며, port_1,3에서는 유의미한 차이를 보이지 않아, 제한폭 확대 이후 포트폴리오들의 고유변동성을 비교한 결과에서는 뚜렷한 패턴을 보이지 않았다.

자산가격결정이론에 따르면 주식수익률의 위험은 총 변동성을 중심으로 체계적 위험과 고유위험으로 나뉘는데, Table 2와 Table 3 결과를 바탕으로 체계적 위험의 관점에서 바라볼 경우, 제한폭의 확대는 대조군에 비해 처치군 증시에서 체계적 위험의 큰 감소가 발생하는 것으로 나타났다.

4. 결론 및 시사점

본 연구는 시장효율성 측면에서 가격제한폭 확대가 주가 변동성에 어떤 영향을 미치는지에 대해 과반응가설을 검정함으로써 제한폭 확대가 국내 증시에 유효한 기능을 수행하는지 판단할 수 있다. 실증분석 결과, 가격제한폭 확대가 시행된 이후 120영업일에서 처치군 증시의 변동성이 감소하는 경향을 보여 과반응가설을 지지하지 못한다고 주장한 Kim(2001)의 분석결과와 유사함을 보였다. 이는 Fama(1989)와 Kyle(1988) 주장에 따라 확대 전 증시 내 조성된 시장참여자들의 불안심리가 완화되어 균형가격에 대한 불확실성이 감소하였으나, 제한폭 확대 이후에는 균형가격에 관한 불확실성이 감소하여 주가는 균형가격 도달에 이로우며, 불안심리로 인한 가격폭락의 촉진과 시장 변동성 증가를 방지해주어, 제한폭 확대는 시장참여자에게도 유익한 혜택이 제공되며 이로 인해 효율적인 시장 조성에 이로운 제도임을 시사한다.

같은 정보를 <그림> VS <표>로 제시하면 소비자의 인지능력이 달라지는가? - 가공식품 영양표시제도를 이용한 실험 연구 -

허성윤¹, 진현정²

요 약

이 연구는 정보의 내용과 양이 같은 상황에서 정보를 각각 <그림>과 <표>로 제시했을 때, 소비자의 인지능력이 달라지는지를 확인할 목적으로 수행하였으며, 이 과정에서 영양표시제도 라는 정책적 도구를 사용하였다. 소비자 1,453명을 대상으로 온라인 웹 실험을 실시하였고 시판 중인 <표>형식의 영양표시와 <그림>형식의 영양표시를 제시하고 간단한 O,X퀴즈를 수행하였다. <그림>유형의 정보는 현행 <표>유형의 정보와 비교하여 정보의 양이 동일한 가운데, 소비자가 더 빠르고 더 정확하게 인지할 수 있음을 확인하였으며, 특히, 주부와 고연령대 소비자에게는 상대적으로 더 효과적인 수 있음을 보였다.

1. 서론

1.1. 연구의 필요성

전 세계적으로 비만 및 당뇨병, 고혈압, 고콜레스테롤혈증 등 식습관과 관련된 질병이 주요 사회문제가 되고 있다. 우리나라의 비만유병률은 20년 동안 남성에서 급격히 증가하여 5명 중 2명이 비만이며, 여자는 4명 중 1명이 비만이다. 특히 남성 30~40대와 여성 65세 이상의 비만유병률이 가장 높으며, 20년 동안 지속적으로 증가함. 고콜레스테롤혈증 유병률은 2005년 이후 남녀 모두 약 3배 이상 증가하여 5명 중 1명이 고콜레스테롤혈증 유병자이며 한국 또한 식습관과 관련된 질병문제가 사회 문제로 대두되고 있다(질병관리청, 2020).

정부는 우리국민의 건강한 식생활을 유도하고 소비자의 알권리를 위해, 가공식품의 영양성분 표시제도를 운영 중에 있으나, 국민건강영양조사에 따르면, 실제 가공식품 선택시 영양표시를 이용한다는 소비자는 전체 소비자의 30% 초반 수준에 머물러 있다.

언론 보도에 따르면 식약처가 실시한 ‘2020년 영양표시 인식활용조사’에서 우리 국민들의 영양표시제도에 대한 만족도는 37%에 그쳤으며, 전체 응답자의 27%가 ‘영양표시를 이해하기 어렵다’고 응답했고, ‘영양표시에 있는 수치의 의미를 모른다’는 응답이 가장 높았던 것으로 전해지고 있다.³

이처럼 정부는 정책적 노력을 기울이며 영양표시 제도의 이용률을 높이고자 하지만, 소비자는

*이 논문은 한국농촌경제연구원으로부터 데이터를 제공받아 분석하였음.

¹(제1저자) 06974 서울특별시 동작구 흑석로 84, 중앙대학교 일반대학원 경제학과 박사과정.

E-mail: heo0411@cau.ac.kr

²(교신저자) 06974 서울특별시 동작구 흑석로 84, 중앙대학교 경제학부 교수. E-mail: hyunjin@cau.ac.kr

³메디컬투데이. 2021.10.21. 『알고보니 콜레스테롤 초과비율 174배?...대기업 PB식품 영양표시 부적합률 15%』

해석이 어렵고, 의미를 잘 모르는 등의 이유로 어려움을 겪고 있는 것으로 보인다. 따라서 현행 제도가 국민의 눈높이에 맞는지, 소비자들이 정보에 노출되었을 때 인지능력을 먼저 살펴볼 필요가 있다.

1.2. 연구의 목적

이 연구의 목적은 정보의 내용과 양이 같은 상황에서 정보를 각각 <그림>과 <표>로 제시했을 때, 소비자의 인지능력이 달라지는지를 확인하기 위함이며, 영양표시제도 라는 정책적 도구를 사용함으로써 분석의 결과에 따라 정책적 시사점을 도출할 수 있을 것으로 기대한다.

2. 연구방법

2.1. 분석 도구

우리나라는 일부 가공식품 품목(과자, 빵, 초콜릿류 등)에 대해 정해진 표시대상 성분(열량, 탄수화물 등)을 표시를 하도록 『식품위생법』에 그 근거를 두고 있다. 식품제조업자는 다음의 <표>형태의 도안 중 한 가지를 선택하여 성분별 함량과 기준치를 표시하여야 한다.



note: For each form, 1) per total content (1 package), 2) per 100g (ml), 3) per unit content

Figure 1. Various examples of nutrition fact

한편, 이외에 영양성분을 주표시면에 하려는 경우에는 다음의 <그림>형태의 표시도안을 사용하여 표시해야 하고, 이 경우, 위 <표> 형태의 성분 표시는 생략할 수 있다. 주표시면(제품 전면부) 도안은 어디까지나 권고 사항이므로, 강제성이 없으며 일부 시리얼 및 뉴트리션바 제조사에서 도입을 했으나 주변에서 흔하게 볼 수는 없는 상황이다.



Figure 2. Front of package standard design

2.2. 분석자료

분석에는 2018년 한국농촌경제연구원의 “농식품 표시정보 현황 및 활용도 제고방안” 연구 수행 과정에서 수집한 자료를 활용하였다. 조사는 조사전문기관 엠브레인에 의뢰하였으며, 2018년 9월 27일부터 10월 1일까지 구조화된 설문지를 이용한 온라인 웹 실험 방식으로 이루어졌다. 총 1,453명을 대상으로 실험을 실시하였으며, 성, 연령, 지역에 비례한 유의할당 추출법을 사용하였다.

2.3. 실험설계

실험의 구성은 실제 마트에서 시판 중인 가공식품 중, <표>형식의 영양표시와 <그림>형식의 영양표시를 수집하여 스캔과정을 거쳐 컴퓨터 화면에 제시하였으며, 각 유형별로 문항의 유형은 모두 같고 숫자와 성분명을 달리하여 간단한 O,X퀴즈 형태로 제시하였다.

동일 응답자가 순차적으로 <표>와 <그림> 두 가지 모두에 대해 답하도록 설계하였으며, 학습효과가 나타날 것을 우려하여 제시 순서는 랜덤하게 배정하였다(응답자의 50%는 <표>를 먼저 응답 후 <그림>응답, 나머지 50%는 <그림>을 먼저 응답 후 <표>에 응답).

B1. 제시된 영양표시를 보고 각 문항에 답하여 주세요. (A형)



	그렇다	아니다
1. 이 제품의 1봉지는 64g이다	<input type="radio"/>	<input type="radio"/>
2. 1봉지를 섭취하면 328kcal이다	<input type="radio"/>	<input type="radio"/>
3. 1봉지의 포화지방은 16g이 들어있다	<input type="radio"/>	<input type="radio"/>
4. 1봉지의 나트륨은 200mg이 함유되어 있다	<input type="radio"/>	<input type="radio"/>
5. 1봉지의 탄수화물 함유량은 일일 기준치의 7%이다	<input type="radio"/>	<input type="radio"/>
6. 이 제품의 총중량은 192g이며 이를 다 먹으면 984kcal이다	<input type="radio"/>	<input type="radio"/>

B1. 제시된 영양표시를 보고 각 문항에 답하여 주세요. (B형)



	그렇다	아니다
1. 이 제품의 1회 제공량은 30g이다	<input type="radio"/>	<input type="radio"/>
2. 1회 제공량 만큼 섭취하면 117kcal이다	<input type="radio"/>	<input type="radio"/>
3. 1회 제공량의 트랜스 지방은 0.3g이 들어있다	<input type="radio"/>	<input type="radio"/>
4. 1회 제공량의 당류는 26.4g이 함유되어 있다	<input type="radio"/>	<input type="radio"/>
5. 1회 제공량의 단백질 함유량은 일일 기준치의 8%이다	<input type="radio"/>	<input type="radio"/>
6. 이 제품의 총 중량은 320g이며 이를 다 먹으면 1,248kcal이다	<input type="radio"/>	<input type="radio"/>

Figure 3. Experimental questions

3. 분석 방법 및 결과 요약

<표>와 <그림>의 응답결과를 각각 정답 수, 소요시간, 정답당 소요시간을 계산하여 t-test를 통해 평균값이 차이를 비교하였다. <표>의 경우 정답 수는 총 6개 문항 중 평균 5.0문항, <그림>은 평균 5.42문항으로 나타났다. 둘 간의 차이는 0.43문항이며, t값이 통계적으로 유의하였다. 소요시간의 경우, 6문항을 응답하는데 <표>는 평균 81.0초, <그림>은 67.6초가 소요되었으며, 13.4초의 차이를 보였다. 이 같은 차이는 통계적으로 유의하다. 정답 개당 소요시간은 <표> 16.99초, <그림> 12.77초로 4.22초의 차이가 발생하였으며, 모두 통계적으로 유의미한 차이인 것으로 나타났다. 즉, <그림> 형태의 영양표시유형은 <표>형태의 영양표시유형 보다 더 정확하고 더 빠르게 인지하는 것을 확인하였다.

다음으로는 <표> 유형 영양표시의 정답 수, <표> 유형 정답 소요 시간(초), <그림> 유형 영양표시의 정답 수, <그림> 유형 정답 소요 시간(초)를 종속변수로 SUR 모형(Seemingly Unrelated Regression Model)을 사용하여 <그림>과 <표>의 인지능력에 영향을 미치는 요인을 분석하였다.

SUR 분석의 주요 결과로는 먼저 영양표시제도가 반복적인 경험에 의해 인지능력 개선이 가능함을 확인하였다. <그림>은 주요 식품구매자인 주부들의 정확도를 증가시키고 시간을 단축시키는데 효과적임을 알 수 있었으며, 연령대가 높을수록 <그림>이 <표>에 비해 상대적으로 효과적인 제도임을 확인할 수 있었다. 이와 같은 차이는 SUR 모형의 교차방정식 제약식 검정을 통해 유의성을 충족하였다.

4. 결론 및 시사점

현재 영양표시제도는 소비자단체의 요구와 생산자 단체의 저항에 직면해 있다. 소비자는 더 많은 정보를 원하지만, 생산자에게는 이러한 요구가 규제로 작용하기 때문이다. 그 가운데 소비자의 영양표시제도 이용률은 낮은 수준에 머물러 있으며, 해석이 어렵고 의미를 모르겠다는 불만이 나오고 있는 상황이다.

이 연구의 분석결과 <그림>유형의 정보는 현행 <표>유형의 정보와 비교하여 정보의 양이 동일한 가운데, 소비자가 더 빠르고 더 정확하게 인지할 수 있음을 확인하였다. 특히, 주부와 고연령대 소비자에게는 상대적으로 더 효과적일 수 있음을 보였다. 따라서 현재 정부가 자율에 맡기고 있는 <그림>형태의 표시도안의 사용을 단계적으로 확대 적용한다면, 현재 수준보다 더 나은 활용이 가능하며, 신호등 표시 등 다른 대안적 표시방법에 비해 업계의 우려도 덜 할 것으로 생각한다. 또한 <그림>형태의 표시도안 사용을 통해, 소비자의 인지 부담을 경감함으로써, 제도의 활용률 제고에도 기여할 수 있을 것으로 기대한다.

References

- Medical Today. (2021.10.21.). I found out that the cholesterol excess ratio is 174 times?... 15% of non-conformity with nutrition labeling of PB food from large corporations.
- Ministry of Food and Drug Safety. (2017). Standards for labeling food.
- Ministry of Food and Drug Safety. Step-by-step nutrition labeling guidelines.

외국인 직접투자(FDI)와 경제자유도가 경제성장에 미치는 영향 - OECD 가입국 대상 -

김민정¹

1. 서론

FDI는 1995년 WTO의 출범 이후로 활발하게 이루어지고 있고 이에 따라 경제 성장의 긍정적인 역할을 한다는 긍정적인 연구결과와 부정적인 연구결과 모두 존재한다. 이에 FDI의 유입이 경제성장에 어떠한 영향을 미치는지에 대해서 명확하게 결론을 내릴 수 없다.

이에 본고는 FDI의 유입과 경제자유도가 OECD 국가의 경제성장률에 미치는 영향에 대해서 분석하였다. 또한, 경제자유도가 경제성장에 미치는 영향에 대해서 알아보기 위해서 분석하였다. OECD 국가를 선진국과 중진국 개도국으로 분류함으로써 국가 경제성장 규모에 따른 차이점을 살펴보고자 하였다. 더 나아가 경제성장률뿐만 아니라 1인당 경제성장률에 미치는 영향까지 살펴봄으로써 FDI의 유입과 경제자유도가 실제 국민의 삶의 질에 어떠한 영향을 미쳤는지에 대해서 살펴보았다.

중속변수로는 생산가능인구, 환율, 수출, 기업적자유도, 투자의 자유도, 무역의 자유도를 통해서 분석하고자 하였다. 본고에선 POLS 모형, FEM 모형, REM 모형을 사용하였으며, Hausman 검정을 통해서 검정을 실시하였다.

선진국의 경우 생산가능인구와 FDI가 긍정적인 영향을 미쳤으며, 수출과 환율은 경제성장에 부정적인 영향을 미쳤다. 경제자유도는 사업규제자유도만 긍정적인 영향을 미쳤으며, 투자자유도와 무역자유도는 부정적인 영향을 미쳤다.

중진국의 경우 생산가능인구와 수출은 경제성장에 부정적인 영향을 미쳤으며, FDI와 환율은 비교적 긍정적인 영향을 미쳤다. 경제자유도는 모두 부정적인 영향을 미쳤다.

개도국의 경우 생산가능인구, FDI, 환율은 경제성장에 긍정적인 영향을 미치며, 수출은 부정적인 영향을 미친다. 경제자유도에선 사업규제자유도와 무역자유도는 부정적인 영향을 미치며, 반대로 투자의 자유도는 긍정적인 경제성장에 긍정적인 영향을 미친다.

FDI 유입이 1인당 경제성장률에 미치는 영향에 대해서 위 와 같이 선진국, 중도국, 개도국으로 나눠서 분석하였다. 그 결과 선진국에선 생산가능인구, FDI, 환율이 상승할수록 경제성장률에 긍정적인 영향을 미쳤으며, 수출은 경제성장률에 부정적인 영향을 미쳤다. 사업규제자유도, 투자자유도, 무역자유도는 부정적인 영향을 미쳤다.

중진국의 경우 생산가능인구와 수출은 경제성장률에 부정적인 영향을 미쳤지만, 반대로 FDI와 환율은 경제성장에 긍정적인 영향을 미쳤다. 사업규제자유도 역시 경제성장에 긍정적인 영향을 미쳤다. 이와 반대로 투자자유도와 무역자유도는 부정적인 영향을 미쳤다.

¹06974 서울특별시 동작구 흑석로 84, 중앙대학교 경제학과 석사과정. E-mail: 12rlaalswjd34@naver.com

마지막으로 개도국의 경우 생산가능인구, FDI 유입은 경제성장에 긍정적인 영향을 미쳤지만, 환율과 수출은 경제성장에 부정적인 영향을 미쳤다. 사업규제자유도와 투자자유도는 경제성장에 긍정적인 영향을 미치는 반면에 무역 자유도는 부정적인 영향을 미쳤다.

FDI의 유입은 대체적으로 GDP뿐만 아니라 1인당 GDP에도 대체적으로 긍정적인 영향을 미쳤다. 이와 반대로 수출은 GDP뿐만 아니라 1인당 GDP에도 부정적인 영향을 미쳤는데 이와 관련된 인과관계는 발견하지 못하였다.

생산가능인구는 긍정적인 영향을 미치는 국가들도 있지만, 부정적인 영향을 미치는 국가들도 있기에 이에 관련된 연구를 조금 더 면밀하게 살펴보아야 한다.

위기 하의 호텔업의 기업실패와 기업부실위험 예측에 관한 연구*

권유진¹, 박성용²

요 약

많은 선행 연구는 호텔의 기업 존속 여부가 기업의 재무구조와는 무관하며 주로 지리적 위치나 서비스에 의존한다고 주장한다. 하지만 코로나19 위기와 같이 기업의 재무건전성이 무너진 유례없는 상황에서 호텔업은 고정비용을 지급하기 위해 융통 가능한 자금이 필요하게 되었고, 유동성문제와 지급불능문제로 인해 도산 위기에 처해졌다. 본 연구는 팬데믹 상황으로 인해 유동성이 얼어붙은 금융시장에서 영업이익 감소로 이어질 수 있는 호텔업의 코로나19 위기와, 비슷한 유형의 범지구적 재난에 대한 호텔업의 기업부실위험을 설명하기 위해 재무건전성과 생존가능성 간의 연관성을 분석한다. 본 연구는 코로나19 위기와 같은 위기 상황에서 상대적으로 고정 비용보다 변동 비용의 비율이 높은 기업이 영업활동 감소에 대응하여 비용을 더 신속적으로 잘 조정할 수 있다는 가정 하에, 각 비용별로 영업레버리지를 추정하고, 호텔업의 부도를 로짓모형으로 예측한다. 분석 결과, 매출액 증가율, 자기자본비율, 현금-총비용 비율, 중간투입비용의 영업레버리지, 영업비용의 영업레버리지, 총자산 그리고 호텔의 위치를 도입한 로짓모형이 2008년 글로벌 금융위기 때의 호텔업 부도와 유동성문제, 지급불능문제를 가장 잘 예측하였다. 글로벌 금융위기의 재무변수로 추정된 모형을 사용하여 2021-2022년 호텔업 중 부도 혹은 기업부실위험에 직면할 기업을 예측한 결과, 관측치의 24.87%, 총 60개의 호텔업이 기업실패 혹은 기업부실위험에 직면할 것으로 예측되었다.

주요용어 : COVID-19, 호텔업 부도 예측 모형, 영업레버리지.

1. 서론

팬데믹 상황과 금리 인상으로 인해 호텔업은 자금 조달의 어려움과 지급능력 부족으로 재무건전성이 악화되었고, 기업의 생존을 보장하기 어려운 상황이다. 몇몇 연구들이 재무변수를 사용하여 위기 상황에서의 호텔업의 생존에 관하여 분석하였지만, COVID-19 위기의 호텔업의 생존을 연구는 스트레스 검정과 같은 시나리오 분석에 그치고 있어, 실제 COVID-19 발발 시기인 2020년의 상황을 반영한 결과를 파악하기에는 한계가 있다(Crespi-Cladera, Martín-Oliver, Pascual-Fuster, 2021). 이에 본 연구는 국내 호텔업의 재무변수와 재무구조를 사용하여 위기 하의 호텔 부도 예측모형을 설정하고 COVID-19 위기로 인해 예측되는 기업실패와 기업부실위험을 예측하는 것을 목적으로 한다.

*이 논문은 2021년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹06974 서울특별시 동작구 흑석로 84, 중앙대학교 경제학과 석사과정. E-mail: yujin.kwonn@gmail.com

²(교신저자) 06974 서울특별시 동작구 흑석로 84, 중앙대학교 경제학과 교수. E-mail: sungpark@cau.ac.kr

2. 실증 연구 방법 및 모형설정

2.1 영업 레버리지

영업 레버리지는 고정자산을 보유함으로써 고정비용을 부담하는 정도를 뜻한다. 본 연구에서는 영업레버리지로 추정된 고정비용의 비율에 따라 호텔업의 기업부실위험의 차이를 분석하고자 한다. 비용의 유형에 따라 고정 비용과 변동 비용에 대한 구조는 다르므로 영업레버리지는 각각 다르게 추정된다. 중간투입비용과 영업비용에는 고정비용 요소와 변동비용 요소가 혼합되어 있으므로, 중간투입비용과 영업비용에 대해 다음 이원고정효과모형(Two-way fixed effect model) 추정식을 사용하여 영업이익의 변화에 대한 각 비용 유형별 탄력성을 분석하고, 비용 유형별 고정비용 요소와 변동비용 요소를 구분한다.

$$\ln Cost_{it}^j = \beta^j \ln Operating Income_{it} + \eta_i + \gamma_t + \epsilon_{it}$$

여기서 i 와 t 는 각 기업과 연도를 의미하고, j 는 $j = \{\text{Intermediate inputs, Operating expenses}\}$ 의 비용 유형을 나타낸다. ϵ_{it} 은 오차항이며, η_i 와 γ_t 는 각각 기업과 시간의 고정효과이다. 계수 추정치는 영업이익의 변화에 대한 각 유형의 비용 탄력성을 의미하며, 주어진 영업이익의 변화에 대한 각 기업 및 각 비용 유형의 상대적인 변화를 추정한다. Table 1은 영업이익 변화에 따른 각 비용 유형별 탄력성을 나타낸다.

Table 1. Elasticity of each cost type to change in operating income

Dependent variable:	ln(Intermediate inputs)	ln(Operating expenses)
ln(Operating income)	0.9914*** (0.019)	0.4323*** (0.009)
fixed effect	Two-way fixed effect	Two-way fixed effect
N	3106	3933

Note. The number in parentheses means the robust standard deviation, and *** in the estimate indicates that the estimate is statistically significant under the significance level of 0%

본 연구에서 추정된 OL^j 은 고정비용을 부담하는 정도를 뜻하므로 총비용 대비 고정비용으로 정의된다. 따라서 OL^j 은 $\frac{FC}{FC+VC}$ 로 나타난다. 여기서 FC 는 고정비용, VC 는 변동 비용을 의미한다. 고정효과 η_i 와 γ_t 는 기업의 영업이익에 의존하지 않는 고정비용 요소이고, 영업 외 비용 또한, 이자비용과 감가상각비와 같이 영업이익에 의존하지 않는 고정비용이므로, 영업레버리지(OL^j)는 다음과 같이 추정된다.

$$OL^j = \frac{\eta_i + \gamma_t + \ln(\text{Non Operating Expenses}_{it})}{\ln(\text{Total Expenses}_{it})}$$

기존 문헌에 의하면, 중간투입비용의 변동비용 요소는 이외 다른 비용의 변동비용 요소보다 크다(Nicolau, 2005). 중간투입비용의 변동비용 요소가 크다는 것은 고정비용 요소가 작음을 뜻하므로, 영업레버리지가 상대적으로 작다는 것을 의미한다. 따라서 두 비용의 영업레버리지의 전체 추정치를 구하기 위해, 두 영업레버리지의 가중평균을 구한다. Table 2는 각 비용 유형별로 추정된 영업레버리지의 기초통계량을 나타낸다.

Table 2. Operating leverage descriptive statistics

	Mean	SD	Min	P25th	P50th	P75th	Max
Intermediate inputs	0.3116	0.022	0.1608	0.2993	0.3146	0.3268	0.3883
Operating expenses	0.6805	0.028	0.5847	0.6614	0.6768	0.6968	0.8350
Weighted leverage	0.4959	0.132	0.2418	0.3850	0.4385	0.6291	0.8008

Note. The weighted average of operating leverage was calculated by weighting the book price of intermediate inputs and operating expenses in the financial statements and income statements

2.4 로짓 부도예측 모형

2007년부터 2013년까지의 글로벌 금융위기의 재무변수 데이터를 사용하여 부도예측모형을 추정하고, 추정된 모형을 사용하여 2022년 호텔업의 실패와 기업부실위험을 예측한다. 본 연구의 부도예측모형은 다음 식과 같이 추정된다.

$$\Pr(y = 1) = f\left(GrOpIncome_{it}, \frac{Cash_{it-1}}{Expenditure_{it-1}}, \frac{Capital_{it-1}}{Assets_{it-1}}, OL_{it-1}^j, Control_{it-1}\right),$$

Table 3. Logit model of bankruptcy during 2007-2013

Pr(y_{it}) = 1, if the firm is out of business during 2007-2013	Average marginal effects				
	(1)	(2)	(3)	(4)	(5)
Growth Operating incomeit	-0.163*** (0.046)	-0.144*** (0.041)	-0.167*** (0.039)	-0.190*** (0.010)	-0.179*** (0.044)
Cash/Total expensesit-1	-0.062 (0.055)	-0.075 (0.065)	-0.075 (0.055)	-0.054 (0.040)	-0.043 (0.040)
Capital Adequacy ratioit-1	-0.075*** (0.019)	-0.063*** (0.017)	-0.056*** (0.017)	-0.043** (0.018)	-0.053*** (0.017)
Intermediate inputsit-1		-0.195 (0.409)		0.227 (0.456)	
Operating expensesit-1			1.807*** (0.351)	1.689*** (0.384)	
Weighted averageit-1					0.240*** (0.067)
ln(Assetsit-1)	-0.058*** (0.010)	-0.036*** (0.008)	-0.020** (0.008)	-0.037*** (0.010)	-0.048*** (0.010)
Id(Seouli)				-0.005** (0.023)	-0.060** (0.010)
Id(Gangwoni)				0.070*** (0.025)	0.066** (0.008)
Pseudo R2	0.142	0.102	0.142	0.197	0.185
AIC	584.67	635.7	608.86	555.96	562.28

Note: The number in parentheses means the robust standard deviation, and ***, **, * in the estimate indicates that the estimate is statistically significant under the significance level of 1%, 5% and 10%

본 모형은 2008 글로벌 금융위기 당시 실패한 기업들이, COVID-19 위기와 동일한 문제(영업레버리지, 유동성문제, 지급불능문제)로 실패했음을 전제한다. 종속변수 y 는 t 기의 i 기업의 부도 기업을 식별하는 더미 변수이다. Cultrera and Brédart (2016)은 매출액이 높을수록 파산 위험이 줄어들어 기업의 생존이 증가함을 분석했다. 본 연구에서는 영업 활동의 변화에 대한 기업 실패의 민

감도를 설명하기 위해 매출액의 증가율을 나타내는 $GrOpIncome_{it}$ 을 독립변수로 포함한다.

$\frac{Cash_{it-1}}{Expenditure_{it-1}}$ 는 현금 보유액을 총비용으로 나눈 재무비율로, 자산의 유동성 및 비용 구조를

설명한다. 또한 총자산 대비 자기자본의 비율을 의미하는 자기자본비율 $\frac{Capital_{it-1}}{Assets_{it-1}}$ 와, $j =$

{Intermediate inputs, Operating expenses}에 해당하는 j 비용에 대해 추정된 영업레버리지 OL^j 가 독립변수에 포함된다. 또한 통제변수 $Control_{it}$ 를 모형에 포함하여 기업 실패를 설명할 수 있는 다른 가능성을 통제한다. 통제변수로는 총자산과 호텔의 지리적 위치를 포함한다. Table 3은 총 5가지 모형 설정으로 로짓 부도예측모형을 추정한 결과이다.

3. 분석 결과

3.1 표본 내 기업실패 및 부실위험기업 예측

2008 글로벌 금융위기(2007-2013) 동안 본 연구의 표본에는 449개의 호텔이 있었고 그 중 63개는 기업실패에 직면하였다. 본 연구에서는 추정된 로짓 부도예측모형이 호텔업의 기업실패를 예측하는 데 유용한 도구인지 분석한다. 최종 모형은 Table 3의 모형(4)가 채택되었다. Table 4는 모형 (4)를 사용하여 기업실패를 예측한 혼동행렬이다. 2008 글로벌 금융위기에서 살아남은 기업의 약 80%가 로짓 부도예측모형에 의해 생존할 것으로 예측되고, 실패한 기업의 약 30%가 로짓 부도예측모형에 의해 도산할 것으로 예측된다. 모형 전체의 정확도는 80% 이상으로 모형 (4)가 기업실패를 적절히 예측함을 확인할 수 있다.

Table 4. Firm survival and failure for 2007-2013

		Observed		
		Survive	Failure	Total
Prediction	Survive	776 (93.61%)	131 (71.98%)	907 (89.71%)
	Failure	53 (6.39%)	51 (28.02%)	104 (10.29%)
	Total	829 (100%)	182 (100%)	1011
Accuracy: 81.80%				
Pearson chi-squared(1)=73.323				
p-value=0.000				

Note: The number in parenthesis is proportion(%) in terms of observed.

3.2 COVID-19로 인한 기업실패 및 기업부실위험 예측

본 절에서는 추정된 부도예측모형을 사용하여, COVID-19로 인해 기업실패와 기업부실위험에 직면하게 될 호텔업에 대한 예측을 제시한다. 총 243개 호텔업의 2020년과 2021년 관측치 453개를 대상으로 2021년과 2022년의 호텔업의 기업실패와 기업부실위험 예측을 수행하였다. Table 5는 예측 결과를 나타낸다.

Table 5. Firm failure and risk for 2021-2022

2021-2022		
Forecasts		
Count		
None	Failure or risk	Total
143	100	243
(58.85%)	(41.15%)	(100%)

2021년에서 2022년에 기업실패 및 기업부실위험을 겪을 것이라고 예상되는 기업은 41.15%로 100개의 기업이 추정된 모형을 통해 예측되었다. 2008 글로벌 금융위기 기간 동안 기업실패 및 부실위험을 겪을 것이라고 예측한 기업의 약 60%가 기업실패 및 부실위험을 겪었으므로, 미루어 추정해 보았을 때, 총 관측치의 24.87%, 약 60개의 호텔업이 COVID-19 위기 동안에 어려움을 겪을 것으로 예상된다.

강설이 지역별 교통사고에 미치는 영향: 강릉, 평창, 춘천을 중심으로*

김현욱¹, 심재관², 임병환³

요약

교통사고와 기상의 관계는 많은 연구가 이루어졌으나 상충되는 결과가 많고 명확한 판단을 내리기 어렵기 때문에 각 지역의 특징에 따라 지역맞춤형 교통사고 정책이 필요하다. 본 연구에서는 강릉, 평창, 춘천을 대상으로 강설이 교통사고에 미치는 영향력을 분석하고, 교통사고의 발생가능성이 50% 이상되는 강설 임계값을 추정하기 위해 기상청의 ASOS 강설 자료와 한국도로교통공단의 교통사고 자료를 활용했다. 피어슨 상관분석으로 교통사고에 가장 많은 영향을 주는 누적 강설 시간을 확인하고 로지스틱회귀분석으로 지역별 강설의 영향력과 교통사고 위험 임계값을 추정했다. 또한 휴일을 상세 구분함으로써 지역별 교통사고의 특징을 확인했다. 그 결과, 지역별로 휴일 전일, 휴일, 마지막휴일에 교통사고의 발생이 상이한 특징을 확인했다. 또한 강설이 증가할수록 교통사고 발생가능성은 지역별로 강릉 약 1.4배, 평창 약 1.3배, 춘천 약 1.7배 증가했다. 마지막으로 교통사고의 발생 가능성이 50%이상인 강설 임계값은 강릉 약 6.5 cm, 평창 약 13.7 cm, 춘천 약 4.7 cm로 추정되었다. 이 연구의 결과는 겨울철 교통영향예보의 기초 자료로 활용 가능할 것으로 판단된다.

주요용어 : 강설, 교통사고, 위험수준, 영향력

1. 서론

교통사고는 인적요인, 도로환경요인, 기상환경요인, 교통환경요인, 차량요인 등 복합적 요인에 의해 시공간적으로 발생한다. 이 중에서 교통사고에 영향을 주는 기상은 강우, 강설, 안개 등 다양하며(Keay et al. 2006; Abdel-Aty et al. 2011; Morgan et al. 2011; Kim et al. 2013), 운전자의 시인성, 도로의 마찰력에 결정적 역할을 하므로 도로 이용시 안전성과 이동성에 큰 영향을 주는 요인이다. 그러나 기상이 도로 안전에 미치는 영향에 대한 연구 주제는 상당한 관심을 끌었지만(Karlaftis and Yannis, 2010), 동시에 상충되는 결과를 보이고 있다(Drousu et al. 2020). 특히, 강설은 교통사고와 밀접한 연관이 있지만(Knapp, 2000; Lee et al. 2000; Maze et al. 2005; Qju and Nixon, 2008), 교통사

*이 연구는 기상청 국립기상과학원 「기상업무지원기술개발연구」 “재해기상 목표관측·분석·활용기술 개발”의 지원으로 수행되었습니다.

¹(교신저자) 25457 강원도 강릉시 죽헌길7 강릉원주대학교 산학협력관 707호, 국립기상과학원 예보연구부 재해기상연구소 선임연구원 E-mail: darzhyun@korea.kr

²25457 강원도 강릉시 죽헌길7 강릉원주대학교 산학협력관 707호, 국립기상과학원 예보연구부 재해기상연구소 기상연구사 E-mail: srsjk@korea.kr

³25457 강원도 강릉시 죽헌길7 강릉원주대학교 산학협력관 707호, 국립기상과학원 예보연구부 재해기상연구소 소장 E-mail: weatherman@korea.kr

고의 증감에 대한 영향은 연구마다 상이하다. Qiu et al.(2008)은 2008년 4개의 연구 결과를 이용한 메타분석을 통하여 강설이 강우보다 높은 영향을 준다고 보고했다. Aemal et al.(2000)은 Iowa주에서 폭설의 집중도와 지속성이 충돌비율과 관계가 있는 것을 밝혔다. 캐나다 오타와에서는 강설시 교통사고 부상자가 20% 증가했고, 한국에서는 9% 증가했다(Andrey et al. 2003; Lee, 2013). Lee(2013)의 연구에서 강설과 교통사고는 역 U자형 패턴을 보였다. Eisenberg et al(2004)는 강설이 증가할수록 교통사고는 증가하지만, 치명률은 감소한다고 보고했다. 게다가, 두번째 이후 발생하는 강설의 경우, 교통사고는 증가하지만 치명률은 감소하며, 첫번째 강설의 경우, 두번째 강설에 비해서 치명률이 높았다. 반면에, Fridstorm et al.(1995)와 Brown and Baass(1997), and Kim et al(2013)은 강설시 교통사고가 감소한다고 설명했고, Drousu et al.(2020)은 루마니아 지역에서 강설과 교통사고는 연관이 없다고 보고했다. 이와같이 상충되는 연구 결과가 나타나는 것은 연구가 수행된 각 국가와 지역은 다른 결과를 설명할 수 있는 고유한 기후와 지역적 특징을 가지고 있기 때문이다. 교통사고는 물리적, 사회적, 법, 행동양식, 운송수단 등에 의해 영향을 받는 복합적인 현상이며(Lee, 2013), 기상요소의 정의, 연구방법, 자료에 의해서도 그 영향이 다르게 측정될 수 있다(Mills et al. 2011). 그러므로 이 연구에서는 강원도 지역 시군구를 대상으로 강설이 지역별 교통사고에 미치는 영향력과 교통사고 발생가능성이 높은 강설량을 제시하는 것을 목적으로 한다.

2. 자료 및 방법

이 연구는 2011년부터 2020년까지 기상청의 ASOS 강설 관측 자료와 한국도로교통공단의 교통사고 자료를 이용하여 분석했다. 교통사고 자료는 경찰에서 조사 처리된 교통사고 정보를 기반으로 구축된 국가공식통계 자료인 경찰 DB와 보험사 등 기관별로 수집된 교통사고 정보를 통합하여 교통사고 총량 파악 및 학술 연구용 통계자료인 통합 DB로 구분된다. 경찰 DB는 국가공식통계자료인 만큼 내용이 상세한 장점을 가지고 있지만 자료의 양이 통합 DB에 비해 적다는 단점을 가지며, 통합 DB는 자료의 양이 많은 장점이 있으나 내용이 경찰 DB에 비해 상세하지 못한 단점을 가지고 있다. 본 연구에서는 경찰 DB와 통합DB를 합하여 기상개황이 정확히 기록된 자료를 대상으로 연구를 진행했다.

연구에 사용된 변수는 일별 자료로 정리하여 사용했다. 강설 변수는 1, 3, 6, 9, 12, 15, 18, 21, 24시간 누적 강설을 계산하고 일별 최댓값을 추출하여 사용했다. 교통사고 변수는 국가통계포털의 자동차등록대수를 활용하여 표준화했으며, 시간 자료를 합하여 일별자료로 구축했다. 교통사고 자료는 고속국도, 차대사람 사고 자료는 사고의 특징이 다를 것으로 판단하고 제외했으며, 일별 ASOS의 강설관측과 교통사고의 기상개황이 일치하는 자료를 선별하여 사용했다. 또한, 일별 교통사고의 특징이 다를 것으로 판단하고 휴일을 상세 구분하여 평일, 휴일전일, 휴일, 마지막휴일로 구분하고 강원도 지역의 해안, 산지, 내륙 지역의 차이를 보기 위해 강릉, 평창, 춘천 지역을 대상으로 연구를 수행했다.

본 연구에서는 강설과 교통사고의 관계와 지역별 강설이 교통사고에 미치는 영향력, 교통사고 발생가능성이 높은 강설량을 분석하기 위해 피어슨상관계분석과 로지스틱회귀분석을 사용했다.

3. 결과

Table 1은 지역별 누적강설과 교통사고의 상관관계를 보여준다. 지역별로 휴일을 상세 구분하여 분석한 결과, 강릉은 평일, 휴일전일과 마지막 휴일에 상관성이 높았고, 평창은 평일과 휴일 전일, 춘천은 평일, 휴일과 마지막 휴일에 상관성이 높았다. 이와 같이 지역별로 상이한 결과는 지역특징에 따른 교통에 따라 다양한 결과를 보여줄 수 있음을 나타낸다. 이 결과는 추후 심도 있는 연구를 통해서 다양한 지역특징이 교통사고에 미치는 영향력 확인과 지역 맞춤형 교통 정책의 필요성을 제기한다.

또한, 연구지역의 강설과 교통사고의 상관성을 살펴보면 강릉은 휴일 전일 0.67, 마지막 휴일 0.58, 평일 0.53 순으로 1시간 강설과 높은 양의 상관관계를 나타냈다. 평창은 휴일 전일 0.48, 평일 0.42 순으로 3시간 강설과 높은 양의 상관관계를 나타냈다. 마지막으로 춘천은 마지막휴일 0.60, 휴일 0.52, 평일 0.47 순으로 높은 양의 상관관계를 나타냈다.

Table 1. Correlation between snowfall and traffic accidents by day

		snow_max_1h	snow_max_3h	snow_max_6h	snow_max_9h	snow_max_12h	snow_max_15h	snow_max_18h	snow_max_21h	snow_max_24h
Gangneung	weekday	0.53	0.53	0.52	0.52	0.51	0.51	0.51	0.50	0.50
	before_holiday	0.67	0.66	0.62	0.61	0.61	0.64	0.64	0.64	0.64
	holiday	-0.04	-0.06	-0.07	-0.07	-0.07	-0.10	-0.10	-0.10	-0.10
	last_holiday	0.58	0.56	0.55	0.61	0.61	0.61	0.61	0.61	0.66
Pyeongchang	weekday	0.42	0.42	0.41	0.37	0.36	0.35	0.35	0.35	0.34
	before_holiday	0.43	0.48	0.48	0.47	0.47	0.46	0.45	0.44	0.44
	holiday	0.26	0.25	0.24	0.25	0.24	0.24	0.24	0.24	0.24
	last_holiday	0.27	0.22	0.22	0.22	0.21	0.21	0.21	0.20	0.18
Chuncheon	weekday	0.47	0.45	0.43	0.43	0.43	0.42	0.42	0.42	0.43
	before_holiday	0.00	-0.13	-0.13	-0.13	-0.13	-0.13	-0.13	-0.14	-0.15
	holiday	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
	last_holiday	0.60	0.60	0.59	0.59	0.59	0.59	0.59	0.59	0.59

Table 2. Logistic regression analysis results of regional snowfall and traffic accidents

		Variable	B	S.E.	Wald	df	Significance level	Exp(β)
Gangneung	snow_max_1hr	.328	.090	13.183	1	.000	1.388	
	constant	-1.253	.364	11.826	1	.001	.286	
Pyeongchang	snow_max_3hr	.251	.059	18.098	1	.000	1.286	
	constant	-2.767	.322	73.970	1	.000	.063	
Chuncheon	snow_max_1hr	.538	.155	12.044	1	.001	1.713	
	constant	-.931	.345	7.268	1	.007	.394	

Table 2는 지역별 강설과 교통사고의 로지스틱회귀 분석 결과이다. Table 1의 결과에 따라 상관성이 가장 높은 변수를 선택했고, 휴일을 세분화하여 자료의 수가 부족한 관계로 상관성이 높은 날짜를 통합하여 연구를 수행했다. 모든 연구지역에서 강설과 교통사고의 로지스틱회귀모형은 적합했으며, 변수는 통계적으로 유의미하게 분석되었다(P-value <0.05). 강릉은 1시간 강설이 증가할수록 교통사고의 발생 가능성이 약 1.4배, 춘천은 약 1.7배, 평창은 3시간 강설이 증가할수록 교통사고 발생가능성이 약 1.3배 증가하는 것으로 나타났으며, 각 지역별 교통사고 위험모델 식은 (1)~(6)과 같다. 또한, 각 지역에서 교통사고 발생가능성이 50% 이상인 평균 강설량은 강릉 약 6.5 cm, 평창 약 13.7 cm, 춘천 약 4.7 cm로 분석되었다.

$$Gangneung = -1.253 + 0.328 \times snow_max_1hr \quad (1)$$

$$Accident\ Risk = EXP(Gangneung)/(1 + EXP(Gangneung)) \quad (2)$$

$$Pyeongchang = -2.767 + 0.251 \times snow_max_1hr \quad (3)$$

$$Accident\ Risk = EXP(Pyeongchang)/(1 + EXP(Pyeongchang)) \quad (4)$$

$$Chuncheon = -0.931 + 0.538 \times snow_max_1hr \quad (5)$$

$$Accident\ Risk = EXP(Chuncheon)/(1 + EXP(Chuncheon)) \quad (6)$$

4. 결론

이 연구는 2011년부터 2020년까지 기상자료와 교통사고 자료를 활용하여 강릉, 평창, 춘천에서 강설이 교통사고에 미치는 영향력 분석과 교통사고의 발생가능성이 높은 강설 임계값을 추정하기 위해 수행되었다. 연구 결과 교통사고에 대한 강설의 영향력은 지역별, 날짜별로 상이한 것으로 확인되었다. 이와 같은 결과는 지역별로 맞춤형 교통사고 정책의 필요성뿐만 아니라 특정 이벤트로 인한 통행량의 증감에 따른 대응정책의 마련이 필요함을 시사한다. 본 연구의 결과는 현재 기상청에서 시행되고 있는 영향예보를 지원하여 겨울철 교통사고 4단계 위험수준을 구분하기 위한 기초자료로써 충분한 역할을 할 수 있을 것으로 판단된다.

References

- Abdel-Aty, M., Ekaram, A., Huang, H., Choi, K. (2011) A study on crashes related to visibility obstruction due to fog and smoke, *Accid. Anal. Prev.*, 2011. 43(5):1730-1737
- Aemal, J., Ketith, K., Karen, L., Leland, D. (2000) Safety implications of snowstorms on interstate highways, TRB paper, NO. 00-0313
- Andrey J., Mills, B., Vandermolen, J. (2003) Temporal analysis of weather-related collision risk for Ottawa, Canada:1990-1998, presented at Annual Meeting of the Transportation Research Board, Washington, D.C.
- Eisenberg D. (2004) The mixed effects of precipitation on traffic crashed, *Accid Anal Prev.*, 36(4):637-647
- Eisenberg D., Warner, K. E. (2005) Effects of snowfalls on motor vehicle collisions, injuries, and fatalities, *Am J Public Health*, 95(1): 120-124
- Keay, K., Simmonds I., (2006) Road accidents and rainfall in a large Australian city. *Accid Anal Prev*, 38(3):445-454
- Kim H. U., Han S. O., Kim K. H. (2013) An analysis of correlation between weather and traffic accidents in Korea. *Journal of Korean Meteorological Society* 10(193), 404-405

- Knapp, K. K., Kroeger, D., Giese, K. (2000) Mobility and Safety Impacts of Winter Storm Events in a Freeway Environment, Center for Transportation Research and Education, Iowa State University
- Lee, Y. T., Kim Y. J., Kim D. H., Yang G. W. (2013) Analysis of Seasonal Variation Effect of the Traffic Accidents on Freeway, Journal of Korean Society of Transportation, 18(5): 7-16
- Maze T. H., Agarwal, M., Burchett, G. (2005) Whether Weather Matters to Traffic Demand, Traffic Safety, Journal of the Transportation Research Board, 1948:170-176
- Mills B. N., Andrey, J., Hambly, D. (2011) Analysis of precipitation-related motor vehicle collision and injury risk using insurance and police record information for Winnipeg, Canada, J Safety Res, 42(5):383-3902
- Morgan A., Mannering, F. L. (2011) The effects of road-surface conditions, age, and gender on driver-injury severities. *Accid Anal Prev.*, 43(5):1852-1863
- Qiu L., Nixon, W. A. (2008) Effects of adverse weather on traffic crashes: systematic review and meta-analysis, *Transportation Research Record*, 2055:139-146

양념채소(대파) 재배농가의 재배규모에 따른 경영성과 순환구조 분석*

김민현¹, 최정², 김도익³, 손장환⁴, 마은미⁵

1. 서론

역사적으로 농업이 중요하지 않은 시기가 없었다. 산업화 시대가 도래하면서 잠시나마 그 중요성이 낮아진 적은 있었으나 4차 산업혁명이 현실화되면서 농업분야가 더욱 더 중요하게 자리 잡고 있다. 또한, 국제정세에서 에너지, 반도체, 식량은 중요한 국가안보 자산이 되었다. 한국사회에서도 농업의 위상은 점차적으로 향상되고 있으며, 국민들도 농업의 인식도가 높아지면서 도시민들의 새로운 일자리로 부각되고 있다. 인식도 향상과 일자리 창출로서 농업의 지속성을 보이기 위해서는 무엇보다도 농업인의 소득 창출과 향상이 뒷받침되어야 한다. 농가소득은 농업소득, 농외소득, 이전소득, 비경상소득으로 구분되며, 농업소득은 농업경영 활동의 결과로 얻어지는 소득을 의미한다(김동신 외, 2016). 농업소득은 생산량, 농가수취가격(농산물 판매가격), 경영비 변동과 각 요인 변동의 상호작용에 따라 변화한다(전지연 외, 2016). 농업소득에 영향을 미치는 요인에 대한 연구를 보면, 농업경영조직 유형, 경영 규모, 지대별 입지, 전·겸업 유형(최재혁 외, 2005), 정보화 수준(남수연 외, 2007), 가구원의 교육수준(정진화 외, 2013), 농업법인 조직 참여(김동신 외, 2016), 생산성 효과, 규모 효과, 가격 효과(윤지원 외, 2016), 귀농연차(전익수, 2019) 등이 많거나 높을수록 농업소득의 효과가 더 높게 나타났다고 하였다. 이렇듯 다양한 측면에서 연구가 진행되고 있다. 농업경영에서 지속 가능한 경영을 위해서는 경영성과를 비교하여 현재 본인의 위치를 진단하고 합리적인 경영계획을 수립하는 것이 중요하다. 이를 위해서는 재배 중인 작목에 대한 수량, 가격, 경영비, 소득 등 수익성을 측정할 수 있는 지표가 필요함에 따라 농촌진흥청에서는 1977년부터 농산물 소득조사(2020년 기준 50개 작목)를 실시하여 농축산물 소득자료집을 발간하고 있으며, 2002년부터 공식 정부통계(제143002호)로 승인되어 정확도와 신뢰도를 공인받고 있다(농촌진흥청, 2021). 농업소득 분석은 인구통계 및 재배면적을 조사하고, 총수입(생산량×농가평균수취가격)과 경영비 등 조사하여 23개 항목에 대한 정보를 제공하고 있다. 각각의 정보는 서로 단절된 것이 아니라 유기적·단계적으로 다음과 같은 경로로 구조화 할 수 있겠다. 농업소득의 단계별 구조는 ‘①재배면적→②재포기간→③경영비→④총수입→⑤소득→⑥순수익→①재배면적’으로 순환적으로 형성할 수 있을 것이다. 하지만, 농업소득 분석은 재배면적에 관계없이 10a당을 기준으로 총수입, 경영비, 생산비, 소득, 순수익 정보를 제시하고 있다. 10a를 기준면적으로 수익성 지표를 제시하다 보니 경영

*본 결과물은 농촌진흥청의 지원을 받아 연구된 것임(PJ015765).

¹58214, 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 연구사. Email: kimh9107@korea.kr

²58214, 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 연구관. Email: chadongi01@korea.kr

³58214, 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 연구관. Email: doik1020@korea.kr

⁴58214, 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 연구관. Email: son1168@korea.kr

⁵58214, 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 연구사. Email: mem83@korea.kr

규모(재배면적)가 농업소득에 영향을 미치는 지를 알 수가 없다. 이에 재배면적에 따른 농업소득이 차이가 있음을 파악해 볼 필요가 있다. 따라서, 본 연구에서는 다음과 같은 연구문제를 제시할 수 있다. 첫째, “조사 작목은 농업소득의 단계별 구조를 순환적으로 형성하고 있는가?”이다. 둘째, “이러한 구조는 조사 작목의 재배면적별로 동일하게 형성하고 있는가?”이다. 셋째, “재배면적별로 구조별 영향력은 차이가 있는가?”이다. 연구대상 작목은 대파를 선정하였는데, 채소수급 안정 사업 대상작목으로 유통 및 가격 안정이 중요하고, 양념채소는 산지 집중도가 높고 노지재배 위주로 기상여건에 따른 생산량 변동 폭이 크기 때문에 가격 변동성과 소득 변동성이 크기(전지연 외, 2016) 때문이다.

2. 연구대상 및 연구방법

연구대상은 농촌진흥청 농산물 소득조사 대상농가이며, 연구 대상자는 5,300명이 선정되었다. 선정된 농산물 소득조사 대상농가의 조사 작목에 대한 경영성과 분석을 위해 전문 조사요원이 농가가 작성한 경영기록장 데이터와 구조화된 조사표를 통한 면접조사 결과를 활용하였다. 분석 작목인 대파는 2018년도부터 2020년도까지 213명이 소득조사 대상농가로 선정되어 조사되었다. 년도별로 보면, 2018년 63명, 2019년 71명, 2020년 79명이다. 시·도별로 보면, 강원도 44명, 경기도 42명, 경상북도 9명, 전라남도 76명, 전라북도 42명이다.

Table 1. 지역별 조사대상자 분포

metropolitan	n	metropolitan	n	metropolitan	n
강원도	44	경상북도	9	전라북도	42
경기도	42	전라남도	76	전체	213

주1. 소득조사 대상농가는 년도별, 지역별로 변동이 발생하였음

조사대상자의 재배경력과 재배규모를 살펴보면, 총 영농경력은 평균 27년이고, 대파 재배경력은 평균 19년이다. 총 경지면적은 평균 50,071m²이고, 총 대파 재배면적은 30,971m², 대파 임차면적은 13,257m²이다. 총 대파 재배면적을 구간으로 보면, 1만m²이하 65명, 3만m²이하 77명, 3만m²초과 71명이다. 또한, 연간 교육시간은 평균 26시간으로 나타났다.

Table 2. 조사대상자 재배경력, 재배면적 기초통계량

		단위 : 년, m ² , 시간				
	item	n	min	max	m	sd
경력	총 영농경력	200	3	52	27	12
	대파 재배경력	200	3	48	19	10
면적	총 경지면적	213	1,980	263,010	50,071	44,449
	총 대파 재배면적	213	1,650	231,000	30,971	34,283
	대파 임차면적	116	990	66,000	13,257	12,649
	교육시간	52	2	231	26	50

분석방법은 SPSS 21.0 통계패키지를 활용하였고, 자료의 기초자료분석을 위한 기술통계 분석 및 차이검증을 실시하고, 연구문제를 검증하기 위해 회귀분석을 실시하였다.

3. 연구결과

년도, 재배규모에 따른 대파 소득조사 농가의 수익성 지표 차이를 검증하기 위해 일원배치분산 분석을 실시하였다. 년도별 총수입, 단가 등 지표 평균치를 비교했을 때 차이가 있으나, 통계적으로 유의미한 결과만을 보면 다음과 같다. 먼저, 년도별 차이를 보면 총수입은 2018년 2,929,019원, 2019년 2,831,026원, 2020년 3,511,459원으로 차이(F=6.578**)를, 판매단가는 2018년 1,043원, 2019년 897원, 2020년 1,191원으로 차이(F=7.996***)를 보였다. 반면, 수량이나 경영비, 생산비, 소득, 순수익은 통계적으로 유의한 차이를 보이지 않았다. 재배규모에 따른 차이를 분석한 결과, 경영비(F=7.604**), 생산비(F=9.046***), 순수익(F=5.455**)이 차이를 보이고 있다. 경영비는 중(1,515,924원)·소(1,369,070원)·대(1,119,470원), 생산비는 소(3,137,825원)·중(2,667,303원)·대(2,016,105원), 순수익은 대(948,816원)·중(447,294원)·소(113,439원) 순으로 경영성과가 높다는 것을 알 수 있다.

Table 3-1. 년도, 재배규모에 따른 수익성 지표별 분산분석 결과

No	item	m	F	item	m	F
총수입	18년도	2,929,019	6.578**	1만㎡이하(소)	3,251,264	0.856
	19년도	2,831,026		3만㎡이하(중)	3,114,597	
	20년도	3,511,459		3만㎡초과(대)	2,964,921	
단가	18년도	1,043	7.996***	1만㎡이하(소)	1,051	0.931
	19년도	897		3만㎡이하(중)	1,101	
	20년도	1,191		3만㎡초과(대)	996	
경영비	18년도	1,307,571	0.435	1만㎡이하(소)	1,369,070	7.604**
	19년도	1,308,039		3만㎡이하(중)	1,515,924	
	20년도	1,392,302		3만㎡초과(대)	1,119,470	
생산비	18년도	2,418,848	1.840	1만㎡이하(소)	3,137,825	9.046***
	19년도	2,438,314		3만㎡이하(중)	2,667,303	
	20년도	2,862,521		3만㎡초과(대)	2,016,105	
순수익	18년도	510,171	0.534	1만㎡이하(소)	113,439	5.455**
	19년도	392,712		3만㎡이하(중)	447,294	
	20년도	648,938		3만㎡초과(대)	948,816	

주1. *p<0.050, **p<0.010, ***p<0.001

주2. 수량, 소득은 통계적으로 무의미(p>0.050)하여 삭제함.

연구문제에서 제기된 경로관계에서의 영향관계를 분석하기 위해 경로관계의 변수간 단순회귀분석을 실시하였다. 총 대파 재배면적과 재포기간 간은 통계적으로 유의미($\beta=0.204$, $t=3.007^{**}$, $R^2=0.041$)하고, 재포기간과 경영비 간은 통계적으로 유의미($\beta=-0.234$, $t=-3.491^{**}$, $R^2=0.055$)하고, 경영비와 총수입 간은 통계적으로 유의미($\beta=0.451$, $t=7.332^{***}$, $R^2=0.203$)하고, 총수입과 소득 간은 통계적으로 유의미($\beta=0.867$, $t=25.256^{***}$, $R^2=0.751$)하고, 소득과 순수익 간은 통계적으로 유의미($\beta=0.462$, $t=7.563^{***}$, $R^2=0.213$)하고, 순수익과 대파 총 재배면적 간은 통계적으로 유의미($\beta=0.198$, $t=2.922^{**}$, $R^2=0.039$)하게 나타났다. 단계별 구조에 따른 변수 간 회귀분석 결과 계수값이 모두 통계적으로 유의하므로 대파는 농업소득의 단계별 구조를 형성하고 있다고 할 수 있다.

Table 3-2. 농업소득 단계별 구조 간 영향관계 분석 결과

independent	dependent	B	β	t	F	R^2	adj R^2
총 대과 재배면적	재포기간	0.001	0.204	3.007**	9.039	0.041	0.037
재포기간	경영비	-638,513	-0.234	-3.491**	12.189	0.055	0.050
경영비	총수입	0.902	0.451	7.332***	53.756	0.203	0.199
총수입	소득	0.775	0.867	25.256***	637.873	0.751	0.750
소득	순수익	0.614	0.462	7.563***	57.191	0.213	0.210
순수익	총 대과 재배면적	0.004	0.198	2.922**	8.536	0.039	0.035

재배면적별로 단계적 구조를 형성하는 지 파악하기 위해 재배면적에 따라 단계별 구조 간 영향관계를 분석한 결과, 모든 면적에서 총 재배면적과 재포기간, 순수익과 총 재배면적 간은 영향관계, 대면적에서 재포기간과 경영비 간이 통계적으로 유의하지 않았으며($p>0.050$), 나머지는 통계적으로 유의미($p<0.050$)하게 나타났다. 재배면적과 재포기간이 무의미한 이유는 본 분석에서는 면적을 대상으로 하였기 때문으로 판단되어 진다. 통계적으로 유의미($p<0.050$)한 결과만을 보면, 경영비와 총수입은 중($\beta=0.598$, $t=6.374***$, $R^2=0.358$)·대($\beta=0.438$, $t=4.074***$, $R^2=0.192$)·소($\beta=0.250$, $t=2.068*$, $R^2=0.063$), 총수입과 소득은 대($\beta=0.945$, $t=24.250***$, $R^2=0.894$)·소($\beta=0.932$, $t=20.591***$, $R^2=0.869$)·중($\beta=0.768$, $t=10.244***$, $R^2=0.590$), 소득과 순수익은 대($\beta=0.552$, $t=5.546***$, $R^2=0.305$)·소($\beta=0.498$, $t=4.594***$, $R^2=0.248$)·중($\beta=0.393$, $t=3.650***$, $R^2=0.154$) 순으로 영향관계를 보이고 있다.

Table 3-3. 재배면적별 농업소득 단계별 구조 간 영향관계 분석 결과

independent	dependent	area	B	β	t	F	R^2	adj R^2
재포기간	경영비	소	-593,566	-0.276	-2.300*	5.288	0.076	0.062
		중	-866,062	-0.241	-2.121*	4.501	0.058	0.045
		대	-48,390	-0.027	-0.224	0.050	0.001	-0.014
경영비	총수입	소	0.686	0.250	2.068*	4.275	0.063	0.048
		중	0.932	0.598	6.374***	40.623	0.358	0.349
		대	1.332	0.438	4.074***	16.594	0.192	0.180
총수입	소득	소	0.909	0.932	20.591***	423.994	0.869	0.867
		중	0.616	0.768	10.244***	104.937	0.590	0.584
		대	0.856	0.945	24.250***	588.074	0.894	0.892
소득	순수익	소	0.662	0.498	4.594***	21.102	0.248	0.236
		중	0.554	0.393	3.650***	13.322	0.154	0.143
		대	0.631	0.552	5.546***	30.753	0.305	0.295

주1. 대과총면적→재포기간, 순수익→대과총면적은 통계적으로 무의미($p>0.050$)하여 삭제함.L

4. 결론

본 연구는 양념채소(대과) 재배농가의 재배면적별 농업소득 분석을 통해 경영비 및 소득에 대한 시사점을 도출하고자 하였다. 첫째, 조사 작목은 농업소득의 단계별 구조를 순환적으로 형성하고 있었다. 둘째, 재배면적별로 소득 단계별 구조는 경영비, 총수입, 소득, 순소득 단계에서만 동일하게 형성하고 있었다. 셋째, 재배면적별로 구조별 영향력 차이를 보면, 경영비와 총수입은 중·대·소,

총수입과 소득은 대·소·중, 소득과 순수익은 대·소·중 순으로 영향관계를 보인다는 점에서 경영비는 중규모, 대규모, 소규모 순이며, 총수입, 소득, 순수익은 대규모, 소규모, 중규모 순이라는 것을 알 수 있다. 본 연구의 한계점은 두 가지이다. 첫째, 3년간의 데이터를 분석하다보니 패널이 정규화되지 못했다는 점이다. 둘째, 표본 데이터가 300개를 넘지 못하여 AMOS를 활용한 구조방정식모델을 위한 집단별 차이검증을 못했다는 점이다. 하지만, 본 연구에 대한 의의를 보면, 첫째, 농업소득의 각각의 영역이 농업소득 단계별 구조를 순환적으로 형성하고 있다는 점을 확인했다는 점이다. 둘째, 재배면적별로 각각 영역에서 경영성과 금액이 차이가 있다는 점과 양념채소는 대규모가 경영성과 측면이나 영향력 측면에서 높다는 것을 도출했다는 점이다.

References

- Kim, D. S., Cho, D-H., Yeo, C.(2016). The effects of participation of farmers in agricultural corporation on agricultural income, *Korean Journal of Local Government & Administration Studies*, 30(4), 161-184.
- Nam, S-Y., Cho, J-K., LEE, S-W.(2007). The Effects of Eco-friendly Farming and Digital Competence on Agricultural Income, *Journal of Rural Development*, 29(5), 45-62.
- Yoon, J-W., Kwon, O. S.,(2016). Decomposing Agricultural Income of Korean Farm Households, *Journal of Rural Development*, 39(4), 29-49.
- Jun, I.(2019). Effects of Returning-to-farm Related Policies on Agricultural Income of the Return Farmers by the Year of Returning. *Journal of Agriculture & Life Science*, 53(2), 163-171.
- Jeon, J-Y, Park, J.(2016). A Decomposition of Income Variability: A Case Study on Rice and Spicy Vegetables Farms, *Journal of Rural Development*, 39(3), 23-48.
- Jung, J. H., Cho, H. J.,(2013). The Effect of Schooling of Farm Household Members on Household Income, *Journal of Agricultural Education and Human Resource Development*, 45(1), 1-23.
- Choi, J-H., Ko, S. N.(2005). Determinants of Farm Household Income in Korean, *Journal of Industrial Economics and Business*, 18(3), 1139-1159.

토양 내 오염물질 농도 예측을 위한 베이지안 벌점 스플라인

유택선¹, 나종현², 윤태현³, 이정호⁴, 김준명⁵, 김한석⁶, 권만재⁷, 주용성⁸

1. 서론

토양오염은 현대 사회의 크나큰 문제점으로 자리 잡고 있다. 토양오염은 인간의 생산 활동과 공장을 통해 토양에 배출된 유해폐기물, 중금속 등과 같은 오염물질에 의해 발생되며, 토양오염은 토양뿐만 아니라 토양 내 지하수와 같은 수자원에도 악영향을 미쳐 생태계와 더불어 인간의 삶 전반에 해로운 영향을 주고 있다. 이와 같은 토양오염의 심각성으로 인하여 최근 토양오염 관련 법령과 정책이 수립되고 있으며, 각종 토양오염정화 사업이 전국적으로 진행되고 있다.

토양오염정화를 위해서는 조사 부지의 토양 내 오염물질의 분포를 정확히 파악해야 한다. 이를 위해서는 조사 부지의 가능한 모든 연속된 지점들에서 시료를 채취해야 한다(Cochran 1977, Thompson 2002). 그러나 이는 현실적으로 불가능하기 때문에 조사 부지 내에서 시료 채취 지점을 고르게 선정하고 각 지점에서 채취된 시료를 통해 얻은 자료를 이용하여 크리깅(kriging)과 같은 공간 예측 모형을 통해 시료가 채취된 지점이 아닌 지점들을 예측하고, 이를 통해 조사 부지의 토양 내 오염물질의 분포를 추정한다. 따라서 자료를 이용하여 공간 예측의 정확성을 극대화 시키는 것은 토양 내 오염물질의 분포를 파악함에 있어서 중요한 문제이다. 하지만 자료가 충분하지 않은 경우, 토양 자료의 특성상 이상치(anomaly)를 포함할 가능성이 높기 때문에 이상치의 영향을 크게 받아 공간 예측의 정확성이 감소할 수 있다. 이러한 문제점으로 크리깅과 같은 공간 예측 모형은 토양 내 오염물질의 분포를 정확히 예측함에 있어서 근본적인 한계가 존재한다(Kim et al.2012). 조사 부지의 토양 내 오염물질에 대한 사전 정보가 존재하는 경우, 사전 정보를 공간 예측에 활용하여 자료가 충분하지 않은 경우에 대한 공간 예측의 정확성을 개선시킬 수 있다.

본 연구에서는 전국토양측정망 자료를 사전 정보로 이용하여 관심 지역의 토양 내 오염물질 농도 예측을 위한 베이지안 벌점 스플라인 모형(Bayesian penalized spline)을 제안한다. 또한, 제안된 방법의 성능을 평가하기 위해 비교모형으로 사전정보를 사용하지 않는 벌점 스플라인, 공간 통계 모형인 정규 크리깅(ordinary kriging) 그리고 딥러닝 모형인 신경망 모형(neural network)을 이용하였다.

¹04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계데이터사이언스학과 석사과정.

²04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계데이터사이언스학과 석사과정.

³04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계데이터사이언스학과 석사과정.

⁴04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계데이터사이언스학과 석사과정.

⁵04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계정보기술연구소 연구원.

⁶02841 서울특별시 성북구 안암로 145, 고려대 지구환경과학과 연구교수.

⁷02841 서울특별시 성북구 안암로 145, 고려대 지구환경과학과 교수.

⁸(교신저자) 04620 서울특별시 중구 필동로 1길 30, 동국대학교 통계데이터사이언스학과 교수.

2. 연구방법

2.1. 분석 자료

본 연구의 분석 자료는 사전 정보로 사용할 자료로서 환경부에서 전국적으로 운영하는 토양측정망에서 2018년부터 2020년까지 측정된 자료에서 중복된 지점과 물질 미검출 지점을 제외한 2325개의 토양측정망에 대한 물질 농도 자료를 이용하였고, 공간 예측을 진행할 관심 지역의 자료로서 남양주시에 있는 한 지역의 6000m²(100m×60m) 크기의 조사 부지의 표토를 24개의 시료 채취 지점에서 채취한 시료를 통해 구한 물질 농도 자료를 이용하였다. 분석에 사용된 물질로, 니켈, 구리, 비소, 아연, 납으로 총 5가지 물질을 사용하였다. 또한, 분석 시에는 각 물질 농도값의 분포가 크게 치우쳐져 있는 문제점이 존재하여 각 물질 농도값에 로그를 취한 값을 분석에 사용하였다. 그리고 각 자료의 관측값의 좌표의 경우, 전국토양측정망 자료는 분석의 편의를 위해 관심 지역의 중심지점의 좌표를 기준으로 중심화(centering)를 하였고, 관심 지역의 자료는 좌표의 규모(scale)가 전국토양측정망 자료에 비해 너무 작은 관계로 표준화하여 분석에 사용하였다.

2.2. 제안 모형

본 연구에서는 조사 부지의 토양 내 오염물질 농도 예측을 위해 베이지안 벌점 스플라인 모형을 제안한다. 본 모형은 두 단계에 걸쳐 구축되며, 첫 번째 단계에서는 전국토양측정망 자료를 바탕으로 벌점 스플라인 모형을 구축한 뒤 전국 토양 오염물질 농도 예측 평면을 예측한다. 벌점 스플라인 모형을 선형혼합모형(linear mixed model) 형태로 나타낼 수 있다(Ruppert, Wand, Carroll, 2005).

$$y_i = \beta_0 + b_i' \beta_K + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2), \quad \beta_K \sim N_K(0, \sigma_\beta^2 I_K)$$

여기서 y_i 는 i 번째 지점의 물질 농도를 의미하며 K 는 기저함수벡터 구축에 사용되는 매듭(knot)의 수를 나타낸다. 기저함수벡터 $b_i = (b_{i1}, b_{i2}, \dots, b_{iK})'$ 는 자료의 i 번째 x-y좌표를 각 매듭좌표를 기준으로 이변량 정규분포 기저함수 $N(0, \sigma_b^2 I_2)$ 를 이용하여 변환된 것이다. 이변량 정규분포의 조절모수 σ_b^2 은 10-fold 교차 검증을 통해 설정하였다. β_0 는 고정효과 모수를 나타내고 β_K 는 임의효과 모수벡터를 나타내며 각 임의효과는 서로 독립이고 분산 모수 σ_β^2 인 정규분포를 따른다. 오차항 ϵ_i 는 평균이 0, 분산이 σ_ϵ^2 인 정규분포를 따르며, 임의효과들과는 독립이라고 가정한다. 위 모형의 모수를 제한된 최대우도추정법(restricted maximum likelihood; REML)으로 추정된 뒤 관심 지역의 맞춤값 \hat{y} 과 표준오차 $se(\hat{y})$ 을 구한다.

두 번째 단계에선 첫 번째 단계에서 구한 \hat{y} 과 $se(\hat{y})$ 을 사전 정보로 사용하여 베이지안 벌점 스플라인 모형을 구축하여 관심 지역의 공간 예측을 수행한다.

$$v_i = \gamma_0 + z_i' \gamma_M + \eta_i, \quad \eta_i \sim N(0, \sigma_\eta^2), \quad \gamma_M \sim N_M(0, \sigma_\gamma^2 I_M)$$

여기서 v_i 는 i 번째에 물질의 농도를 의미하며 첫 번째 단계와 마찬가지로 여기서 M 은 기저함수벡터 구축에 사용되는 매듭(knot)의 수를 나타낸다. $z_i = (z_{i1}, z_{i2}, \dots, z_{iM})'$ 은 이변량 정규분포 기저함수를 이용하여 변환된 기저함수벡터이다. γ_0 는 고정효과 모수이고 γ_M 은 임의효과 모수벡터를 나타내

며 각 임의효과는 서로 독립이고 분산 모수 σ_γ^2 인 정규분포를 따른다. 오차항 η_i 는 평균이 0, 분산이 σ_η^2 인 정규분포를 따르고 임의효과들과는 독립이라고 가정한다. 본 모형은 베이지안 구조를 따르므로 각 모수에 대해 사전분포를 가정한다.

$$\gamma_0 \sim N(\hat{y}, se(\hat{y})^2), \sigma_\eta^2 \sim \Gamma(0.01, 0.01), \sigma_\gamma^2 \sim \Gamma(0.1, 0.1)$$

위 모형의 각 모수는 깁스 표본추출(Gibbs sampling) 알고리즘을 이용하여 추정하였다.

2.3. 분석 결과

제안 모형의 성능평가를 위해 사전정보를 사용하지 않는 벌점 스플라인, 정규 크리깅, 신경망 모형을 비교모형으로 이용하였다. 정규 크리깅은 R 패키지 *atakrig*의 *autoKrige* 함수를 이용하여 주어진 자료에 대하여 자동으로 적합하였고, 신경망 모형은 R 패키지 *neuralnet*를 이용하여 적합하였다. 성능평가는 관심 지역의 5가지의 물질 농도에 대하여 각 모형 별로 훈련용 자료의 크기가 8, 12, 16일 때의 교차검증을 200번 반복하여 계산된 RMSE(root mean squared error)의 평균 비교를 통해 이루어졌으며, 각 모형의 RMSE의 평균은 다음 Table 1과 같다.

Table 1. mean of RMSE by methods after cross-validation with 200 iterations

variables	sample size of Train dataset	model			
		Bayesian penalized spline	penalized spline	ordinary kriging	deep neural network
As	8	0.2947	0.3189	0.3077	0.4189
	12	0.2958	0.3105	0.3051	0.3857
	16	0.2929	0.3129	0.3067	0.3532
Cu	8	0.2760	0.2935	0.3037	0.3499
	12	0.2515	0.2582	0.2899	0.3254
	16	0.2476	0.2479	0.2669	0.3092
Ni	8	0.1833	0.1994	0.1874	0.2453
	12	0.1720	0.1820	0.1856	0.2101
	16	0.1693	0.1703	0.1767	0.1980
Pb	8	0.1812	0.1967	0.1882	0.2429
	12	0.1685	0.1809	0.1835	0.2168
	16	0.1640	0.1705	0.1776	0.2062
Zn	8	0.3139	0.3234	0.3310	0.4325
	12	0.3097	0.2948	0.3271	0.3916
	16	0.3162	0.2964	0.3169	0.3691

대부분의 상황에서 제안 모형의 성능이 관심 지역에 대한 공간 예측 시 가장 좋은 성능을 보였다. 특히 공간 예측에 쓰인 훈련용 자료의 크기가 작을수록 타 모형들에 비해 제안 모형의 성능이 좋은 것으로 나타났다. 다만, 아연에 대하여 훈련용 자료의 크기가 12, 16인 경우 타 모형에 비해 공간 예측의 성능이 좋지 않은 것으로 나타났는데, 이는 전국토양측정망 자료를 이용하여 구한 관심 지역의 맞춤값 \hat{y} 이 관심 지역의 물질 농도 표본평균값과 상이하여 사전 정보로서 공간 예측의 정확성 개선에 도움을 주지 못한 것으로 판단된다.

3. 결론

본 연구는 전국토양측정망 자료를 사전 정보로 이용하여 관심 지역의 토양 내 오염물질 농도 예측을 위한 베이지안 벌점 스플라인 모형(Bayesian penalized spline)을 제안하였다. 타모형들과의 성능평가 결과, 사전 정보가 관심 지역의 자료와 상이한 경우를 제외하고, 자료가 적을수록 사전 정보를 이용한 제안 모형이 타모형들에 비해 공간 예측의 정확성이 좋음을 확인할 수 있었다. 따라서 토양오염정화를 위한 조사 부지의 토양 내 오염물질의 분포를 파악하고자 할 때, 조사 부지에 대한 사전 정보가 있는 경우 제안 모형을 사용할 수 있음을 제언한다.

Reference

- Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.
- Thompson, S. K. (2002). *Sampling, 2nd Edition*. Wiley, 367.
- Kim, H.-R., Kim, K.-H., Yun, S.-T., Hwang, S.-I., Kim, H.-D., Lee, G.-T., & Kim, Y.-J. (2012). Evaluation of Geostatistical Approaches for better Estimation of Polluted Soil Volume with Uncertainty Evaluation. *Journal of Soil and Groundwater Environment*, 17(6), 69 - 81. <https://doi.org/10.7857/JSGE.2012.17.6.069>
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression* (No. 12). Cambridge university press.

Ensemble Variable Selection Using Genetic Algorithm

Seogyoung Lee, Martin Seunghwan Yang, Jongkyeong Kang, SeungJun Shin¹

Abstract

Variable selection is one of the most crucial tasks in supervised learning, such as regression and classification. The best subset selection is straightforward and optimal but not practically applicable unless the number of predictors is small. In this article, we propose directly solving the best subset selection via the genetic algorithm(GA), a popular stochastic optimization algorithm based on the principle of Darwinian evolution. To further improve the variable selection performance, we propose to run multiple GA to solve the best subset selection and then synthesize the results, which we call ensemble GA(EGA). The EGA significantly improves variable selection performance. In addition, the proposed method is essentially the best subset selection and hence applicable to a variety of models with different selection criteria. We compare the proposed EGA to existing variable selection methods under various models, including linear regression, Poisson regression, and Cox regression for survival data. Both simulation and real data analysis demonstrate the promising performance of the proposed method.

Keywords : Genetic Algorithm, Ensemble Learning, Generalized Linear Model, Cox Regression.

1. Introduction

Variable selection has become of great interest in supervised learning, such as regression and classification in various applications. Suppose we are given a pair of random variables $(Y, \mathbf{X}) \sim F_{Y|\mathbf{X}}$ where Y is a univariate response, $\mathbf{X} = (X_1, \dots, X_p)^T$ is a p -dimensional predictor, and $F_{Y|\mathbf{X}}$ denotes their joint distribution. The goal of variable selection is to identify an index set S defined as

$$S = \{j \mid F_{Y|\mathbf{X}} \text{ functionally depends on } X_j, j = 1, \dots, p\}$$

where $F_{Y|\mathbf{X}}$ is conditional distribution of Y given \mathbf{X} .

Variable selection often assumes the relationship between Y and \mathbf{X} is completely determined through a linear function of \mathbf{X} , i.e., $\beta^T \mathbf{X}$ where $\beta = (\beta_1, \dots, \beta_p)^T$ is a parameter of interest. To be more precise, we assume

¹Department of Statistics, Korea University. E-mail: sfamily4@korea.ac.kr

$$Y \perp \mathbf{X} \mid \boldsymbol{\beta}^T \mathbf{X}$$

where \perp denotes the statistical independence. The assumption above covers a wide variety of models, including linear regression, generalized linear model, Cox regression, and single index model, to name a few. Under the assumption above, S is equivalently written as

$$S = j \mid \beta_j \neq 0, j = 1, \dots, p$$

In this article, we focus on the parametric models under(1.2), and let $\ell(\boldsymbol{\beta}; Y, \mathbf{X})$ denote a log-likelihood function of the model. In population level, it is natural to assume that the true model with informative variables in S achieves the optimal efficiency. That is, we have

where $\mathbf{v}_u = v_j \mid j \in U^T$ for a given vector $\mathbf{v} = (v_1, \dots, v_p) \in R^p$ and an index set $U \subseteq \{1, \dots, p\}$. Given a set of data $(y_i, \mathbf{x}_i), i = 1, \dots, n$, the best subset selection \hat{S} solves an empirical version of (1.3) given as follows.

$$\hat{S} = \arg \max_{S \subseteq \{1, \dots, p\}} \frac{1}{n} \sum_{i=1}^n \ell(\hat{\boldsymbol{\beta}}_S; y_i, \mathbf{x}_{i,S}) - C \cdot |S|,$$

where $\hat{\boldsymbol{\beta}}_S$ denotes the maximum likelihood estimator(MLE) of $\boldsymbol{\beta}_S$, i.e, $\hat{\boldsymbol{\beta}}_S = \arg \max_{\boldsymbol{\beta}_S} \frac{1}{n} \sum_{i=1}^n \ell(\boldsymbol{\beta}_S; y_i, \mathbf{x}_{i,S})$.

We remark that the second term is necessary to prevent over-fitting in the sample level. Here, C is a non-negative constant, and the popular choice of C include $\log n$ and 2 which are known as Bayesian Information Criterion(BIC) and Akaike Information Criterion(AIC), respectively.

Although it seems straightforward, BSS is a combinatorial problem and thus infeasible unless p is restrictively small. To tackle this, one can solve Best subset selection objective function sequentially. Sequential variable selection methods include forward selection and backward elimination and have become greatly popular in practice due to their simplicity. However, the sequential methods are inevitably sub-optimal since they cannot search the entire space of the model.

In the meantime, the penalization has been another canonical approach of the variable selection since the seminal least absolute shrinkage and selection operator(LASSO) proposed by Tibshirani(1996). The penalized variable selection solves the following problem to seek a sparse estimator $\hat{\boldsymbol{\beta}}$.

$$\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^n \ell(\boldsymbol{\beta}; y_i, \mathbf{x}_i) - \sum_j p_{\lambda}(|\beta_j|)$$

where $p_{\lambda}(\beta)$ denotes the sparsity-pursuing penalty function with λ being a non-negative tuning parameter that controls the degree of sparsity. Popular choices of the penalty function include, but not are not limited to LASSO, elastic net(Zou and Hastie; 2005), group lasso(Yuan and

Lin; 2006), smoothly clipped absolute deviance(SCAD, Fan and Li; 2001) penalty, and minimax concave penalty(MCP, Zhang; 2010) penalty. The penalized approach estimate \hat{S} as $\{j \mid \hat{\beta}_j \neq 0, j=1, \dots, p\}$ where $\hat{\beta}_j$ denotes the j th element of $\hat{\beta}$. The penalized variable selection is also closely connected to the best subset selection. It is well known that L_0 -penalized variable selection is equivalent to the best subset selection, and all the aforementioned penalties can be viewed as a continuous approximation of the L_0 -penalty.

Our work is motivated by the following question: Can we solve best subset selection directly? In this paper, we propose to solve BSS via genetic algorithm(GA), a popular stochastic search algorithm motivated by the principle of Darwinian evolution. GA is an effective tool for solving NP-hard optimization problems such as the best subset selection. See for example, Leardi and Gonzalez(1998); Broadhurst et al.(1997) and Niazi and Leardi(2012).

In the meantime, Zhu and Chipman(2006) proposed to parallel version of GA-based variable selection to enhance the variable selection by running multiple GA with several initial values and then synthesizing the multiple results. One can notice that their idea can be viewed as an ensemble variable selection.

The success of the ensemble variable selection depends on the diversity of base selectors to be aggregated for the final result. There are two ways to introduce the diversity in the ensemble variable selection. First type is generate randomness from the data by using resampling method like bootstrap or subsampling. See Zhu and Fan(2011); Wang et al.(2011), and Meinshausen and Bühlmann(2010). Another way to introduce the diversity is to use a stochastic base selector. The parallel version of GA-based variable selection proposed by Zhu and Chipman(2006) is a canonical example for this type. See also Xin and Zhu(2012) and Zhang et al.(2006).

In this article, we extend the idea of Zhu and Chipman(2006) to more general models, parametric models that satisfy $\{eqref{eq:linear}\}$ such as the generalized linear model, and cox-regression for censored responses as a special case. Although our proposal is the same as Zhu and Chipman(2006), except for extension beyond the linear regression, we focus more on the ensemble aspect of the method instead of the GA aspect. Namely, Zhu and Chipman(2006) understand their algorithm as a variant of GA, but we regard it as a version of ensemble variable selection. From our point of view, GA is not essential, and any stochastic variable selection method can be used as a base selector, which brings a natural extension of our method. For this reason, we call the method ensemble variable selection on the genetic algorithm(to be referred to as EGA hereafter) instead of the parallel GA algorithm named by Zhu and Chipman(2006).

The proposed EGA has a clear advantage in practice due to its conceptual simplicity. We demonstrate its promising performance through a simulation study and real data analysis under various models beyond the classical linear regression.

The remainder of the paper is organized as follows. In Section 2, we introduce GA and

describe how to solve the best subset selection problem using GA. In Section 3, the ensemble extension of the GA-based variable selection is described in great detail. Simulation studies to evaluate the finite sample performance of the EGA are conducted in Section 4, and illustration to real data examples is given in Section 5. Concluding remarks follow in Section 6.

Application of the Conway-Maxwell-Poisson Hidden Markov models for analyzing traffic accident^{*}

Hee-Young Kim¹, Hyo-Jin Bae², Su-Hyeon Kim³

Abstract

This paper documents the application of a hidden Markov model with Conway-Maxwell-Poisson(CMP) state-dependent distribution. The CMP distribution originally developed by Conway and Maxwell(1962) for modeling queues and service rates. And recently CMP has been re-introduced by Shmueli et al.(2005) for analyzing count data subjected to over- and under-dispersion. Over the past half century, Hidden Markov Models(HMMs) have become increasingly popular tools for modeling time series data where, at each point in time, a hidden state process selects among a finite set of possible distributions for the observations. We model the distributions of the traffic accident covering the period 2020, comprising T=366 observations in total, using CMP-HMM and Poisson-HMM, which allows for overdispersion and serial dependence. Our results indicate two-state CMP-HMM is improvement on the two-state Poisson-HMM by AIC and BIC. Also we present the full array of the conditional distributions, i.e., the distribution of X_t conditioned on all the other observations at all times other than t .

Keywords : Count data; Time series modeling; Conway-Maxwell-Poisson distribution; Hidden Markov model.

^{*}This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045707, NRF-2021R1F1A1048309) and Natural Science Foundation of Korea university (No. K2021811).

²(Corresponding Author) Associate Professor, Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700, E-mail: starkim@korea.ac.kr

³Undergraduate student, Department of Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700.

SVD 기반 데이터 증강과 기계학습을 통한 웨이퍼 결함 패턴 분류

서정원¹, 정운서²

요약

반도체 공정 중 EDS공정의 경우 반도체 칩들의 수율 및 정상 작동 여부를 판단하기 위한 공정이다. EDS 공정으로 얻은 웨이퍼맵(Wafer Map)을 통해 엔지니어들이 육안으로 결함 여부 및 결함 패턴을 구분하는 한편 합성곱 신경망(convolutional neural network, CNN) 딥러닝 모형을 통해 결함 패턴 분류를 자동화하기 위한 연구가 계속해서 이어지고 있다. 현재 웨이퍼맵 데이터의 경우 패턴 분류 모형을 적합하기 위한 양품 대비 불량품의 결함 패턴 별 데이터가 부족하다. 본 논문은 이 문제점을 해결하기 위한 방법으로 특이값 분해(singular value decomposition, SVD) 방법을 이용하여 데이터를 증강한다. 특이값 분해 방법의 데이터 증강 성능 평가를 위해 이미지 데이터의 잡음제거와 증강에 일반적으로 사용되는 잡음제거 오토인코더(denoising autoencoder, DAE)와 변분 오토인코더(variational autoencoder, VAE)의 성능을 비교한다. 성능 비교는 두가지 측면에서 이루어지며 데이터 증강에 소요되는 시간과 증강된 데이터로 적합한 분류 모형의 성능을 비교한다. 성능 비교를 위해 사용된 모형은 서포트 벡터머신(support vector machine, SVM)과 랜덤포레스트(random forest)와 같은 기계학습 기반의 모형으로 증강된 이미지 데이터로부터 공간적 특성을 추출한 후 변수로 활용한다. 성능 비교 결과 특이값 분해로 증강한 데이터로 적합한 분류기 모형의 성능이 가장 좋았고 증강 속도 또한 가장 빠름을 확인하였다.

주요용어 : 특이값분해, 딥러닝, 데이터 증강, 오토인코더, 공간적 특성.

¹02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 대학원생. E-mail: sjw2403@naver.com

²02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 부교수. E-mail: yoons77@korea.ac.kr

리만 좌표하강 알고리즘을 이용한 조각 측지선 적합

박관영¹, 신재경², 구자용³

요약

본 연구에서는 구면에서 관측한 데이터를 평활하는 문제를 해결하기 위한 조각 측지선 적합 방법을 고려한다. 구면 데이터가 복잡한 경향성을 보이는 경우 변화점을 감지할 수 있는 유연한 곡선 적합 방식이 요구된다. 이를 위해 유클리드 데이터에 대한 조각 선형회귀 모형을 구면으로 확장해서 조각 측지선 모형을 적합할 수 있다. 조각 측지선은 조정점의 개수와 위치에 의해 형태가 완전히 결정된다는 특징을 가진다. 인접한 측지선의 속도 벡터의 차이를 벌점화하는 성김 벌점화 방법을 통해 조정점의 개수를 데이터 적응적으로 선택하고자 한다. 구면 거리 함수를 이용해 정의한 손실함수와 조정점 선택을 위한 벌점항의 합을 목적함수로 하는 최적화 문제를 풀기 위해 리만 좌표하강 알고리즘을 도입한다. 수치적 연구의 결과를 통해 좌표하강 알고리즘의 효율성과 곡선 적합 방법의 우수성을 확인하였다. 또한 조각 측지선 적합 방법론을 확장할 수 있는 다양한 가능성을 검토하였다.

주요용어: 구면 데이터, 리만 기하학, 벌점화, 조각 측지선, 좌표하강 알고리즘.

1. 서론

구면 데이터 분석 방법은 기상학, 생물학 등 분야에서 발생하는 구면에서 직접 관측한 데이터와 심장병 관련 연구, 컴퓨터 비전, 생리학 등 제반 분야에서 발생하는 방향 데이터 분석을 위해 빈번하게 사용된다(Gould 1969; Jupp 1987, Su et al. 2012; Mardia and Jupp 2000). 특히 시간에 따라 노이즈와 함께 관측된 구면 데이터를 평활하는 문제는 구면 데이터 분석 방법론에서 중요한 위치를 차지하고 있다(Bak et al., 2022).

본 연구에서는 구면 데이터를 평활하는 문제를 해결하기 위한 조각 측지선 적합 방법을 고려한다. 측지선은 구면에서 직선의 역할을 하기 때문에(do Carmo 1992) 본 연구의 방법론은 조각 선형 모형의 기하학적 확장으로 이해할 수 있다. 조각 측지선 적합을 위해 최소제곱법을 다양체 상으로 일반화하여 구면 거리 함수에 기초한 이차 손실함수를 최소화하는 문제를 정식화한다. 이때 과적합을 막고 변화점을 감지하기 위한 정칙화 방법이 요구되는데, 본 연구에는 조각 측지선 적합 문제에서 조정점의 개수와 위치를 데이터 적응적으로 선택하기 위해 속도 벡터의 기반한 l_2 벌점

¹02844 서울특별시 성북구 보문로34다길 2, 성신여자대학교 수리통계데이터사이언스학부 조교수.

E-mail: kybak@sungshin.ac.kr

²02841 서울특별시 성북구 안암로 145, 고려대학교 정경대학 통계학과 박사과정.

E-mail: tsworud60@korea.ac.kr

³02841 서울특별시 성북구 안암로 145, 고려대학교 정경대학 통계학과 교수. E-mail: jykoo@korea.ac.kr

화 방식을 채택한다. 내재적 기하학에 기초한 최적화 문제를 해결하기 위해 리만 좌표하강 알고리즘을 도입한다. 수치적 연구를 통해 리만 좌표하강 알고리즘의 효율성을 확인하고 구면 곡선 적합 방법론의 성질을 살펴본다. 또한 제안한 방법론의 확장 가능성을 검토한다.

2. 조각 측지선 적합

구면에 두 점 u 와 v 를 연결하는 측지선은 다음과 같이 정의할 수 있다(do Carmo 1992; Bak et al. 2022).

$$\alpha(t; u, v) = \frac{\sin((1-t)\theta)}{\sin(\theta)}u + \frac{\sin(t\theta)}{\sin(\theta)}v, \quad t \in [0, 1]$$

여기서 $\theta = d(u, v) = \arccos(u^\top v)$ 는 구면 거리를 의미한다. 주어진 구면 자료에 유연한 곡선을 적합하기 위해 부분 구간에서 정의되는 측지선을 연결한 조각 측지선을 고려할 수 있다. 시간 구간 I 위에 매듭점 $\tau_1 < \dots < \tau_J$ 가 주어지면 이에 대응하는 부분 구간은 $I_j = [\tau_j, \tau_{j+1}]$, $j = 1, \dots, J-1$, 형태로 정의할 수 있다. 매듭점에 대응하는 구면에 조정점을 $\xi = (\xi_1, \dots, \xi_J)$ 로 표기하면 조각 측지선은 지시함수 $\in d(\cdot)$ 을 사용해서 다음과 같이 정의할 수 있다.

$$\gamma(t; \xi) = \sum_{j=1}^{J-1} \alpha\left(\frac{t - \tau_j}{\tau_{j+1} - \tau_j}; \xi_j, \xi_{j+1}\right) \in d(t \in I_j)$$

시간에 따라 관측된 구면 데이터는 $\{(t_n, y_n)\}_{n=1}^N$ 으로 나타낼 수 있다. 여기서 t_n 은 관측 시점을 나타내고 y_n 은 그에 대응하는 구면에 자료점을 나타낸다. 구면 데이터를 잘 적합하는 조각 측지선을 찾기 위해 유클리드 최소제곱 문제를 일반화하여 다음과 같은 솔실함수를 정의할 수 있다.

$$\ell(\xi) = \sum_{n=1}^N d^2(y_n, \gamma(t_n; \xi))$$

이때 조정점에 성감성을 부여하는 벌점항을 통해 과적합 문제를 해결하고자 한다. 이를 위해 속도 벡터의 차이를 벌점화하는 ℓ_2 그룹라쏘 방법 사용한다.

$$p(\xi) = \sum_{j=2}^{J-1} \|\alpha'(0; \xi_j, \xi_{j+1}) - \alpha'(1; \xi_{j-1}, \xi_j)\|_2$$

복잡도 모수 $\lambda > 0$ 을 도입하여 목적함수를

$$\ell^\lambda(\xi) = \ell(\xi) + \lambda p(\xi) \quad (1)$$

와 같이 정의하고, 목적함수를 최소화하는 조정점을

$$\hat{\xi}_\lambda = \operatorname{argmin} \ell^\lambda(\xi)$$

로 표기한다. 주어진 구면 데이터를 적합하는 조각 측지선은 최종적으로 $\gamma(\cdot; \hat{\xi}_\lambda)$ 의 형태로 정의한다.

2. 리만 좌표하강 알고리즘

목적함수 (1)을 최소화하기 위해 그래디언트에 기반한 좌표하강 알고리즘을 고려한다. 조정점은 유클리드 벡터로 주어지기 때문에 이를 업데이트 하는 것은 블록 좌표하강 알고리즘으로 이해할 수 있다(Tseng 2001). 이때 유클리드 그래디언트를 그대로 사용하는 경우 조정점이 구면에서 움직이지 않고 구면을 포함하는 유클리드 공간 안에서 움직이는 문제가 발생한다. 이 부분을 해결하기 위해 유클리드 그래디언트를 구면에 접평면으로 정사영하는 리만 그래디언트를 고려한다.

리만 그래디언트를 사용하면 조정점이 접평면 위에서 목적함수를 최소화하는 방향으로 업데이트된다. 이렇게 업데이트된 조정점을 지수 사상을 이용해 구면으로 옮겨 준 후 다음 조정점에 대한 업데이트를 진행한다. 이 과정을 모든 조정점에 대해 수행하며 알고리즘이 수렴할 때까지 반복한다. 업데이트 공식을 도식적으로 나타내면 다음과 같다.

$$\xi_j^{(k)} \leftarrow \text{Exp}_{\xi_j^{(k-1)}}[\xi_j^{(k-1)} - \eta_{k,j} \text{grad}_{\xi_j} \ell^\lambda(\xi_1^{(k)}, \dots, \xi_j^{(k)}, \xi_j^{(k-1)}, \dots, \xi_j^{(k-1)})], j = 1, \dots, J$$

여기서 $\xi_j^{(k)}$ 는 k번째 반복에서 j번째 조정점의 좌표를 나타내고 $\text{Exp}(\cdot)$ 함수는 지수 사상, grad 는 리만 그래디언트를 의미하며, $\eta_{k,j}$ 는 경사하강 방법을 사용할 때 학습 속도를 조절하는 역할을 한다. 고정된 복잡도 모수 λ 에 대해 리만 좌표하강 알고리즘을 적용해 $\hat{\xi}_\lambda$ 를 계산할 수 있다.

제안한 별점항을 사용하는 경우 λ 가 커짐에 따라 $\|\alpha'(0; \xi_j, \xi_{j+1}) - \alpha'(1; \xi_{j-1}, \xi_j)\|_2$ 값이 0으로 다가간다. λ 가 충분히 큰 경우 두 인접한 측지선은 같은 대원 위에 있다는 의미가 되므로 중간에 있는 조정점을 제거할 수 있다. 이렇게 성감성을 부여하는 별점화 방법을 통해 조정점에 개수로 표현되는 모형의 복잡도를 선택하게 된다. 모형 선택을 최적으로 진행하기 위해 여러 λ 값을 고려하여 그에 해당하는 곡선을 구한 후 구면 거리를 이용하는 아래와 같이 정의되는 BIC(Schwarz 1978)를 이용하여 모형을 평가한다.

$$\text{BIC}_\lambda = N \log \ell(\hat{\xi}_\lambda) + 3J_\lambda \log N$$

여기서 J_λ 는 주어진 λ 에 대응하는 조정점의 개수를 의미한다. BIC를 최소화 하는 λ 를 최적 복잡도 모수로 선택하고 이에 대응하는 조정점의 개수와 위치를 사용하여 최종 구면 적합 곡선을 도출한다.

3. 토의

본 연구에서는 리만 좌표하강 알고리즘을 이용하여 구면 데이터에 측지선 적합을 수행하는 별점화 방법론을 살펴보았다. 본 연구의 곡선 적합 방법과 알고리즘은 구면 기하학의 특징을 사용하지 않고 일반적인 리만 기하학에만 기초하고 있기 때문에 높은 확장 가능성을 가진다. 조각 측지선 적합 방법론은 두 가지 관점에서 직접적인 확장이 가능하다.

우선 조각 측지선 적합은 유클리드 공간에서 생각했을 때 조각 선형 모형과 대응하는 것으로 이해할 수 있다. 따라서 데이터에 복잡하고 곡선적인 경향성이 나타나는 경우 차수를 높여 구면에서 스플라인 곡선을 적합하는 방법을 개발할 수 있다. Bak et al.(2022), Crouch et al.(1999), Popier

and Noakes(2006)에서 확인할 수 있듯이 구면에서는 대수적 다항식이 정의되지 않지만 카스텔조 알고리즘을 일반화하여 재귀적으로 높은 차수의 베지어 곡선을 정의할 수 있다. 이를 통해 높은 차수의 스플라인을 정의하고 복잡한 구면 데이터 적합에 적용하는 별점화 방법론을 개발하는 연구를 고려할 수 있다.

제안한 방법이 구면 기하학에 의존하지 않기 때문에 양정치 행렬 공간, 3차원 직교군과 같은 여러 다양체 공간에서 관측되는 데이터를 분석하는데 응용할 수 있다(Absil et al., 2009). 리만 그래디언트의 형태를 유도하여 곡선 적합 방법을 확장하면 예측, 변화점 감지 등 다양체 데이터 분석을 요하는 여러 응용 분야에서 활용할 수 있을 것으로 기대한다.

References

- Absil, P.A., Mahony, R., Sepulchre, R. (2009). *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton.
- Bak, K.-Y. (2022). Intrinsic spherical smoothing method based on generalized Bézier curves and sparsity inducing penalization, *Journal of Applied Statistics*.
- Crouch, P., Kun, G., Leite, E.S. (1999). The de Casteljau algorithm on lie groups and spheres, *Journal of Dynamical and Control Systems*, 5, 397-429.
- do Carmo, M.P. (1992). *Riemannian Geometry*, Birkhäuser, Boston.
- Gould, A.L. (1969). A regression technique for angular variates, *Biometrics* 25, 683-700.
- Jupp, P.E., Kent, J.T. (1987). Fitting smooth paths to spherical data, *Journal of the Royal Statistical Society: Series C*, 36, 34-46.
- Su, J., Dryden, I.L., Klassen, E., Le, H., Srivastava, A. (2012.) Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds, *Image and Vision Computing*, 30, 428-442.
- Mardia, K.V., Jupp, P.E. (2000). *Directional Statistics*, Wiley.
- Popier, T., Noakes, L. (2006). C^2 spherical Bézier splines, *Computer Aided Geometric Design*, 23, 261-275.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461 - 464.
- Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization, *Journal of optimization theory and applications*, 109(3), 475 - 494.
- Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49-67.

Bayesian Approach for Eliminating User Influence and Item Popularity Biases in User-Item Bipartite Networks^{*}

Hohyun Jung¹

Abstract

User-item bipartite networks consist of users and items, where edges indicate the interactions of user-item pairs. We propose a Bayesian generative model to the user-item bipartite network that can measure the two types of rich-get-richer biases: item popularity and user influence biases. Furthermore, the model contains a novel measure of an item, namely the item quality that can be used in the item recommender system. The item quality represents the genuine worth of an item when the biases are removed. We apply the proposed model to Flickr user-tag and Netflix user-movie networks to yield remarkable interpretations of the rich-get-richer biases. We further discuss genuine item quality using Flickr tags and Netflix movies, considering the importance of bias elimination.

Keywords : Bayesian inference, bipartite networks, item recommendation, rich-get-richer, social network.

1. Introduction

In this study, we propose a Bayesian generative model applied to the communities with a user-item bipartite network structure, focusing on the elimination of rich-get-richer biases. The main contributions are summarized as follows:

1) The model measures the sizes of user influence and item popularity in the community to understand the rich-get-richer mechanism. To the best of our knowledge, this would be the first work to analyze the rich-get-richer by considering the conjunction of the two factors. A community with a high user influence effect would show more social interactions, while a community with a high item popularity effect would have an internal community structure that provides sufficient information about items, for example, item rankings and trending items.

2) The model measures the genuine quality of the item after user influence and item popularity biases have been removed. Influential users such as highly-subscribed YouTubers

^{*}This research was supported by the National Research Foundation of Korea (NRF-2021R1G1A109410312).

¹Assistant Professor, School of Mathematics, Statistics and Data Science, Sungshin Women's University, 34 Da-gil, Bomun-ro, Seongbuk-gu, Seoul, 02844, Korea. E-mail: hhjung@sungshin.ac.kr

can easily publicize their favored items, and those items will receive benefits. The genuine item quality could be useful in determining key items that can be loved by users for a long time. In addition, by recommending qualitatively excellent items, overall user activity would increase in the community.

3) We apply the proposed model to the Flickr user-tag and the Netflix user-movie networks to analyze the characteristics of the rich-get-richer mechanism and the item quality considering rich-get-richer biases.

2. Methods

We denote the time-series user-item bipartite networks by B_0, B_1, \dots, B_T . The edge connection probability $\theta_{i,t}$ of item i at time t is given by

$$\theta_{i,t} = \text{sigmoid}(\beta_0 + \beta_q q_i + \beta_r r_{i,t} + \beta_s s_{i,t}), \quad t = 1, 2, \dots, T$$

where the item quality q_i of item i is assumed to follow the standard normal distribution $q_i \sim N(0,1)$, the item popularity $r_{i,t}$ of item i at time t is set to recent degree increase of items, and the user influence score $s_{i,t}$ of item i at time t is set to recent degree increase of users. The coefficients β_r and β_s measure the size of rich-get-richer biases corresponding to item popularity and user influence, respectively.

We employ Bayesian inference using diffuse priors for model parameters and standard normal priors for item qualities.

3. Results

Flickr is an image hosting service as well as an online community created in 2004. We use crawled datasets (Görlitz, Sergej, Steffen, 2008) from January 1, 2004, to January 1, 2006. Netflix provides internet entertainment services for watching movies and TV shows. The Netflix data (Bennett, Lanning, 2007) are obtained from November 11, 1999, to December 31, 2005. We consider the data up to January 1, 2005, as initial data for both data, and construct Flickr user-tag and Netflix user-movie bipartite networks with a time interval of 28 days and $T=12$. Also, we analyze the users and items with at least 100 connections at the last time point $t=12$. We have 29,798 users and 14,231 items on Flickr, and 236,117 users and 16,795 items on Netflix.

Table 1 and Table 2 show the parameter estimation results and high-quality items for each data.

Table 1. The parameter estimation results of the proposed model on the real data.

Network	Posterior Statistics	β_0	β_q	β_r	β_s
Flickr	Mean (Estimate)	-7.2091	0.9676	0.0614	0.0789
	Median	-7.2093	0.9675	0.0615	0.0788
	Standard Deviation	0.0053	0.0018	0.0006	0.0013
Netflix	Mean (Estimate)	-8.7172	1.9734	0.0596	0.2780
	Median	-8.7168	1.9729	0.0596	0.2780
	Standard Deviation	0.0038	0.0026	0.0001	0.0006

Table 2. High-quality items for each data.

Rank	Flickr		Netflix	
	Quality	Degree Increase	Quality	Degree Increase
1	stem	stem	1675	9810
2	sharp	sharp	4754	11713
3	vega	vega	13526	15356
4	batteries	timer	10105	4918
5	timer	batteries	13508	2087

4. Conclusion

We apply the proposed model to Flickr user-tag and Netflix user-movie bipartite networks. It turns out that the rich-get-richer effect caused by item popularity and user influence is clearly operative in both networks. Flickr turns out to have small rich-get-richer biases. The effect of user influence in Netflix is estimated to be large. We estimate the genuine item quality considering rich-get-richer biases. Top 5 Netflix high-quality and high-degree-increase items show significant differences since the degree increase of movies watched by influential users could be overrated. The item qualities can be useful when recommending items to the interested users, and we believe that it would help community growth.

References

- Görlitz, O., Sergej, S., and Steffen, S. (2008) PINTS: peer-to-peer infrastructure for tagging systems, *International Workshop on Peer-to-Peer Systems*, 19, 1855641-1855660.
- Bennett, J., Lanning, S. (2007). The Netflix Prize, *Proceedings of KDD Cup and Workshop*, 2007, 35

ODIM: a method to identify inliers via inlier-memorization effect of deep generative models

Dongha Kim¹, Jaesung Hwang², Yongdai Kim³

Abstract

Identifying whether a given sample is an outlier or not is a significant issue in various real-world domains. Many trials have developed outlier detection methods, but they mainly presumed no outliers in the training data set. This study considers a more general situation where training data contains some outliers, and any information about inliers and outliers is not given. We propose a powerful and efficient learning framework to identify inliers in a training data set using deep neural networks. We start with a new observation, called the inlier-memorization effect, that when we train a deep generative model with data contaminated with outliers, the model first memorizes inliers before outliers. Exploiting this finding, we develop a new method called Outlier Detection via the Inlier-Memorization effect(ODIM). The ODIM only requires a few updates; thus, it is fast and efficient. We also provide a data-adaptive strategy to find the optimal number of updates, which makes the ODIM applied to real domains at ease. We empirically demonstrate that our method can refine inliers successfully in both tabular and image data sets.

Keywords : Anomaly detection, inlier-memorization effect, importance weighted autoencoders.

1. Introduction

Outlier(also anomaly) is an observation that differs significantly from other observations, and outlier detection(OD) is the task of identifying whether a given sample is an outlier or not. OD is used as a preprocessing step in supervised learning to filter out anomalous training samples which may degrade the performance of a predictive model. Also, OD itself has various applications such as fraud detection, fault detection, and defect detection in images.

Many researchers have devoted to OD, and various methods for OD can be roughly

¹Assistant Professor, School of Mathematics, Statistics and Data Science, Sungshin Women's University, 34 Da-gil, Bomun-ro, Seongbuk-gu, Seoul, 02844, Korea E-mail: dongha0718@sungshin.ac.kr

²Manager, SK Telecom, 264, Pangyo-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea.
E-mail: postechiminuru@gmail.com

³Professor, Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. E-mail: ydkim0903@gmail.com

categorized into three categories: 1) Supervised outlier detection (SOD) requires label information about whether each training sample is inlier(also normal) or outlier(also abnormal) and solves the two-class classification task. 2) Semi-supervised outlier detection(SSOD) refers to methods which assume all training data being inliers and construct patterns or models for the inliers. SSOD can be interpreted as the one-class classification task since any information of outliers is not used during the training procedure. 3) Unsupervised outlier detection(UOD) deals with training data including some outliers but no label information about anomalousness is available. Most anomaly detection tasks in practice are related to UOD since existence of outliers in massive data are unknown in advance. Thus, in this study, we concentrate on developing a UOD learning framework.

We start with finding a new and interesting observation that deep neural networks memorize inliers first and outliers next in a training phase. When we train a deep generative model with training data that includes outliers, the reduction of inliers' loss values is prioritized to outliers at early updates. We call this observation the inlier-memorization(IM) effect. The IM effect occurs simply because, in the early learning stage, decreasing the loss values of inliers rather than outliers is a more beneficial direction to reduce the overall loss function, which will be explained more rigorously in the subsequent section.

From the IM effect, we propose a novel and simple but powerful method called Outlier Detection via the IM effect(ODIM) to identify inliers from a given training data set.

We train a deep generative model with a log-likelihood-based approach such as VAE(Kingma and Welling, 2013) and IWAE(Burda et al., 2016) for a few updates and based on per-sample loss distribution, data with small loss are regarded as inliers.

Here, the key to our method's success is choosing the optimal update number to utilize the IM effect to the max since the IM effect is sensitive to how long the training proceeds. To do this, at each update, we fit the Gaussian mixture model with two clusters to the per-sample loss distribution and evaluate the Wasserstein distance to measure how apart the two clusters are distributed. We chase the distances as the update goes on and find the optimal update point.

Our method has several advantages over existing OD methods. First, ODIM is simple and efficient in terms of computational time and resources because it requires only few iterations in the training phase to identify inliers. Thus, even if we get multiple loss scores from models trained with different initializations to achieve the ensemble score, it would not take much time. Second, ODIM is agnostic to the types of data. Note that most existing methods specialize in a specific data domain as tabular or image. Our method does not require specific data types and hence can be applied to areas such as medicine, finance, and production processes without any hamper, which allows our method to be utilized in many real-world industries. Third, our method is generally more accurate in identifying inliers than other baseline methods. With extensive experiments over about 20 data sets, including tabular and

image data, we show that ODIM is competitive or superior to other proposed methods.

There is a study that utilizes a phenomenon similar to the IM effect. Wang et al.(2019) suggested a method utilizing so-called inlier-priority effect via the self-supervised learning framework. Generating more than a hundred artificial classes with a pre-specified annotation strategy, they succeeded in identifying inliers with high accuracy. The method in Wang et al.(2019) works well, but it can not analyze tabular data since the way it annotates the training data using a self-supervised learning method is specialized for image data. It can be a critical drawback considering that most practical domains handle tabular data. Moreover, there exist image data domains, such as wafer data in the semiconductor industry, that are hard to generate labels with existing annotation techniques.

The key contributions of this work are: 1) We find a new finding called the IM effect that deep generative models memorize inliers prior to ourliers at early learning stages. 2) We provide a simple but powerful UOD learning framework called ODIM to identify inliers from a given unlabeled training data mixed with anomalous samples. 3) We empirically demonstrate the superiority and efficiency of our method by analyzing various benchmark data sets.

References

- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Yuri Burda, Roger B. Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *Proceedings of 4th International Conference on Learning Representations, ICLR*, 2016.
- Siqi Wang, Yijie Zeng, Xinwang Liu, En Zhu, Jianping Yin, Chuanfu Xu, and Marius Kloft. Effective end-to-end unsupervised outlier detection via inlier priority of discriminative network. *Proceedings of Advances in Neural Information Processing Systems, NeurIPS*, 2019.

패널자료에서의 은닉마르코프모형을 이용한 사례 분석 연구

이동희¹

요 약

패널조사는 동일 집단이나 개인에 대해 동일한 항목에 대해 매번 반복 측정된 결과들을 다루게 되는데, 전 과정에 걸쳐 특정 항목에 대한 지속적인 변화를 압축적으로 설명하는 것은 쉽지 않은 일이다. 본 연구에서는 은닉마르코프모형을 이용한 패널조사 자료에 대한 분석 방법을 소개하고, 한국미디어패널조사 자료를 이용한 분석 과정과 결과를 살펴보았다.

주요용어 : EM-알고리즘, 기계학습, 마르코프 연쇄, 베이지안 네트워크

1. 서론

가구나 국가, 기업 등으로부터 여러 번에 걸쳐 반복적으로 관찰된 횡단면 자료값들을 합쳐 놓은 것을 패널자료(panel data) 혹은 경시적 자료(longitudinal data)라 부른다(Baltagi, 2008). 대부분의 패널자료는 일련의 가구나 개인들로 이루어진 집단을 대상으로 여러 회에 걸친 조사를 통해 수집되는데, 짧게는 단기간에 걸친 2회의 조사로부터 길게는 20년 이상의 기간에 걸쳐 조사가 이루어지기도 한다. 패널자료의 경우 동일한 개체나 집단으로부터 시간흐름에 따라 반복적으로 관찰되기 때문에 이러한 반복측정으로부터 파생하는 상관성이 흔히 나타난다. 따라서 패널자료를 이용한 모형의 통계적 추론을 위해서는 이와 같은 현상을 반영할 필요가 있으며, 이를 고려한 분석을 통해 분석 대상들 간의 시간흐름에 따른 유의미한 변화들을 분석 대상내의 변화들로부터 구분해 낼 수 있어야 한다(Hsiao, 2003). 이러한 목적에서 패널자료를 이용한 대부분의 분석은 반응변수와 설명변수들 간의 인과관계를 패널자료가 갖는 특성들을 반영하여 살펴보는 것에 한정돼 있다.

본 연구에서는 패널조사를 통해 나타나는 지속적으로 관찰된 특정 항목에 대한 변화 자체를 파악하기 위한 효과적인 분석 방법으로 은닉마르코프모형(hidden Markov model)을 소개하고자 한다. 은닉마르코프모형은 베이지안 네트워크(Bayesian network)이란 이름으로도 알려져 있는 동태적 분석(dynamic analysis)방법의 하나로, 실제로 관찰된 변수들 간의 관계를 마르코프 연쇄(Markov chain)적 특징을 가정한 은닉변수(hidden variable)들을 통해 순차적으로 나타난 현상들을 체계적으로 설명할 수 있는 모형이다(Zucchini and MacDonald, 2009). 은닉마르코프모형은 1970년대부터 음성 인식 분야에서 응용되기 시작하면서 최근에는 기계학습(machine learning) 분야에서 필기체 인식, 동작 인식, 문장 인식 등을 위해 널리 사용되는 대표적인 모형 가운데 하나로 알려져 있다(MacDonald and Zucchini, 2007).

패널조사는 앞서 언급했듯이 동일한 개체로부터 동일한 항목을 반복 측정하기 때문에, 각 항목

¹16227 경기도 수원시 영통구 광고산로 154-42, 경기대학교 경영학과 교수. Email: dhl@kgu.ac.kr.

에 대한 특정 개체들의 선택 가능한 조합은 패널조사의 실시 횟수와 선택할 수 있는 항목의 수에 따라 기하급수적으로 늘어나게 된다. 따라서 패널조사를 통해 수집한 자료를 통해 시간 흐름에 따른 변화를 파악하기 위해서는 실제 관찰된 수많은 조합들 가운데 의미 있는 패턴을 찾는 것이 필요하며, 이때 은닉마르코프모형은 이러한 현상을 분석하기 위한 적절한 대안 가운데 하나이다. 본 연구에서는 패널자료를 대상으로 은닉마르코프모형을 적용함으로써 변수들 간의 인과관계 외에 반복적인 패널조사를 통해 나타난 의미 있는 패턴을 찾을 수 있는 방법으로 은닉마르코프모형을 소개하는 한편, 2010년 시범조사를 시작한 이래 현재까지 진행되고 있는 한국미디어패널조사를 통해 수집된 자료를 대상으로 한 분석 사례를 제시하고자 한다.

2. 은닉마르코프모형

패널조사에서 대부분의 조사항목은 범주형 변수로 구분할 수 있다. 즉 특정 항목들 가운데 하나를 선택하는 것이다. 예를 들어 패널조사를 통해 동일 개체들을 대상으로 4개 가운데 하나를 선택할 수 있는 문항을 제시하여 10회 걸쳐 조사를 진행했다면, 나타날 수 있는 모든 패턴의 수는 410가지가 될 것이다. 단순히 매회 조사에서 특정 항목의 비율만을 조사한다면 각 시점에서 전체적인 해당 항목의 비율만을 살펴볼 수 있을 뿐, 조사 대상자의 선택이 시간 변화에 따라 어떤 변화를 보이는지에 대해서는 현실적으로 모든 가능한 패턴 조합을 고려하여 파악하기란 불가능하다. 여기서는 다음과 같이 패널조사를 통해 관찰된 범주형 일변량 변수에 대해 가장 간단한 형태의 은닉마르코프모형의 구성과 추정과정에 대해 설명해보고자 한다.

범주형 반응변수 Y_i 가 T 회에 걸쳐 반복적으로 관찰된 결과를 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT})'$ 라 하고, $i=1, \dots, n$, 즉 매회 조사에서 n 개의 관찰값들로 이루어져 있다고 하자. 즉 \mathbf{Y} 는 n 개의 개체의 어떤 특성을 T 회 반복 측정된 결과를 포함하고 있는 반응변수이다. 이때 살펴보고자 하는 것이 특정 개체의 특성이 보여주는 시간에 따른 변화라면 Bartolucci et al.(2007)이 제안한 다음과 같은 방법을 사용할 수 있다.

우선 i 번째 개체의 시간 변화에 따른 관찰 결과인 벡터 \mathbf{Y}_i 에 대응하는 관찰되지 않은 잠재과정(latent process) $\mathbf{U}_i = (U_{i1}, \dots, U_{iT})'$ 이 있음을 가정하고, 이러한 \mathbf{U}_i 가 주어졌을 때, \mathbf{Y}_i 가 조건부 독립임을 가정하자. 즉 동일한 관찰대상에 대해 시차에 따라 나타나는 변화들에 대해서 잠재변수 \mathbf{U}_i 를 이용하여 설명하고자 하는 것이다. 이를 위해 이러한 잠재과정이 상태공간(state space) $\{1, \dots, k\}$ 에서의 1차 마르코프과정(1st-order Markov process)을 따른다고 하면, $t > 2$ 인 모든 경우에서 잠재변수(latent variable) U_{it} 는 U_{it-1} 이 주어졌을 때, U_{it}, \dots, U_{it-2} 들과 조건부 독립이다. 이때 사용된 마르코프 가정을 통해 반응변수 \mathbf{Y}_i 에서 나타나는 시간에 따른 변화를 반영할 수 있게 되며, 비교적 다른 방법들에 비해 사용상의 제약이 거의 없고, 쉽게 상황을 설명할 수 있다. 이와 같은 마르코프 가정에 기반하여 다음과 같은 모형화가 가능하다.

이 모형의 모수들은 조건부 반응확률들로 나타낼 수 있다. 즉 i 번째 관찰개체에 대한 t 시점에서의 잠재변수가 주어졌을 때의 반응변수의 조건부확률 $P(Y_{it} = y | U_{it} = u)$ 를 $\phi_{yu}^{(t)}$ 라고 하자. 이때 $t=1, \dots, T$, $u=1, \dots, k$, $y=1, \dots, l$ 이다. 이때 i 번째 관찰개체에 대한 최초 시점에서의 초기확률(initial probability), $P(U_{i1} = u)$ 을 $\pi_u^{(1)}$ 이라 나타내고, $t \neq 1$ 인 경우의 전이확률(transition probability),

$P(U_{it} = v | U_{it-1} = u)$ 를 $\pi_{vu}^{(t)}$ 로 표현하면 다음과 같이 모형에서 필요한 확률에 대한 모수들을 나타낼 수 있다. 이때 $v = 1, \dots, k$ 이다.

이와 같이 표현된 확률들을 이용하여 잠재변수 U_i 의 분포를 나타내면 다음과 같다.

$$P(U_i = \mathbf{u}) = \pi_u^{(1)} \prod_{t>1} \pi_{u_t|u_{t-1}}^{(t)}$$

여기서 $\mathbf{u} = (u_1, \dots, u_T)$ 이다. 마찬가지로 U_i 가 주어졌을 때의 Y_i 의 조건부확률은 다음과 같이 표현할 수 있다.

$$P(Y_i = \mathbf{y} | U_i = \mathbf{u}) = \prod_t \phi_{y_t|u_t}^{(t)} \quad (1)$$

따라서 범주형 반응변수 Y_i 의 주변부분포(marginal distribution) $f(\mathbf{y})$ 는 잠재변수와 반응변수 간의 결합확률분포를 $f(\mathbf{u}, \mathbf{y}) = P(Y_i = \mathbf{y} | U_i = \mathbf{u})P(U_i = \mathbf{u})$ 로 나타낼 수 있으므로, 다음과 같이 구할 수 있다.

$$\begin{aligned} f(\mathbf{y}) &= P(Y_i = \mathbf{y}) \\ &= \sum_{\mathbf{u}} f(\mathbf{u}, \mathbf{y}) \\ &= \sum_{u_1} \phi_{y_1|u_1}^{(1)} \pi_u^{(1)} \sum_{u_2} \phi_{y_2|u_2}^{(2)} \pi_{u_2|u_1}^{(2)} \dots \sum_{u_T} \phi_{y_T|u_T}^{(T)} \pi_{u_T|u_{T-1}}^{(T)} \end{aligned} \quad (2)$$

여기서 $\mathbf{y} = (y_1, \dots, y_T)^T$ 이다. 이때 주변부분포를 구하는 과정에서 상태공간과 반복측정 시점에 따라 잠재변수들로 이루어진 벡터 \mathbf{u} 와 관련된 모든 가능한 k^T 경우들에 대해 합을 구해야 하므로 상당한 계산이 이루어져야 한다. 따라서 효율적인 연산을 위해 이를 그대로 사용하는 것이 아니라 전향적 재귀(forward recursion)와 행렬연산 등을 이용한다. 행렬표현을 이용한 전향적 재귀를 이용한 연산과 관련한 자세한 내용은 Bartolucci et al.(2007)의 연구에 제시돼 있다.

이제 이와 같은 모형의 추정에 대해 살펴보도록 하자. 식 (2)의 형태로 반응변수의 주변부분포를 나타낼 수 있으므로 이를 이용하여 최대가능도추정(maximum likelihood estimation) 방법을 적용할 수 있다. n 개의 관찰개체들로 이루어진 표본 가운데 i 번째 개체의 반응변수를 \mathbf{y}_i 라 하면, 모형의 로그가능도함수(log-likelihood function)는 다음과 같이 나타낼 수 있다.

$$l(\boldsymbol{\theta}) = \sum_i \log(f(\mathbf{y}_i)) \quad (3)$$

여기서 $\boldsymbol{\theta}$ 는 모형에 포함된 모수 벡터이다. 이 식은 다음과 같은 형태로 표현이 가능하다.

$$l(\boldsymbol{\theta}) = \sum_{\mathbf{y}} n_{\mathbf{y}} \log(f(\mathbf{y}_i)) \quad (4)$$

이때 $n_{\mathbf{y}}$ 는 표본에서 나타나는 동일한 \mathbf{y} 들의 반응값들에 따른 빈도를 나타내는데, 이와 같이 로그가능도함수를 변환을 한 이유는 공변량을 갖지 않는 경우에 연산횟수를 줄여 식 (3)에 비해 보다 효율적으로 계산을 할 수 있기 때문이다. 로그가능도함수를 최대화하여 모수벡터 $\boldsymbol{\theta}$ 의 추정량, 즉 최대가능도추정량(maximum likelihood estimates)을 얻을 수 있다. 특히 직접적으로 관찰되지 않는 잠재변수가 모형에 포함돼 있기 때문에 이러한 모형에 대한 최대가능도추정은 Dampster et

al.(1977)이 제안한 EM알고리즘(Expectation-Maximization Algorithm)방법을 이용하여 수행할 수 있다 (Bartolucci, 2006). 즉 반응변수 \mathbf{y}_i 와 관련된 부분은 완전자료(complete data)로 간주하면, 잠재변수 \mathbf{u}_i 는 결측자료(missing data)로 구분할 수 있다. 식 (4)를 이용하여 완전자료에 대한 로그가능도함수를 다음과 같이 표현할 수 있다.

$$l^*(\theta) = \sum_{\mathbf{u}} \sum_{\mathbf{y}} m_{\mathbf{u}\mathbf{y}} \log[f(\mathbf{u}, \mathbf{y})]$$

여기서 $m_{\mathbf{u}\mathbf{y}}$ 는 잠재변수 \mathbf{u} 의 상태공간에서의 각 상태와 반응변수 \mathbf{y} 의 각 항목들 간의 교차빈도를 나타낸다. 완전자료의 로그가능도 함수는 다음과 같이 정리할 수 있다.

$$l^*(\theta) = \sum_t \sum_u \sum_y \tilde{a}_{uy}^{(t)} \log[\phi_{y|u}^{(t)}] + \sum_u a_u^{(1)} \log(\pi_y) + \sum_{t>1} \sum_u \sum_v a_{uv}^{(t)} \log[\pi_{v|u}^{(t)}] \quad (5)$$

이때 $a_u^{(t)}$ 는 t 시점의 잠재 상태공간에서 나타난 개체들의 발생빈도이며, $a_{uv}^{(t)}$ 는 t 시점에서 잠재 상태공간 내의 상태 u 로부터 또 다른 상태 v 로 변화한 개체들의 발생빈도이다. $\tilde{a}_{uy}^{(t)}$ 는 잠재 상태에서 반응변수 y 에 대응되는 개체들의 발생빈도인데, 이들 빈도들은 모두 알 수 없으므로 다음과 같은 두 단계로 이루어진 EM알고리즘을 이용하여 최대가능도 추정결과를 얻을 수 있다.

● E-단계(Expectation-step): 이미 관찰된 값들과 현재 상태에서의 모수추정결과를 이용하여 식 (5)에 포함된 알려지지 않은 빈도들에 대한 기댓값을 산출하여 로그가능도 함수의 기댓값을 구한다. 이들 빈도들에 대한 기댓값은 다음의 식들을 통해 얻을 수 있다.

$$\hat{a}_u^{(t)} = \sum_{\mathbf{y}} n_{\mathbf{y}} r^{(1)}(u|\mathbf{y}), \quad \hat{a}_{uv}^{(t)} = \sum_{\mathbf{y}} n_{\mathbf{y}} \sum_{t>1} r^{(t)}(u,v|\mathbf{y}), \quad \hat{a}_{uy}^{(t)} = \sum_{\mathbf{y}} n_{\mathbf{y}} I(y^{(t)} = y) r^{(t)}(u|\mathbf{y})$$

여기서 $I(\cdot)$ 은 지시함수(indicator function)를, $r^{(t)}(u|\mathbf{y}) = P(U_{it} = u | \mathbf{Y}_i = \mathbf{y})$ 를 나타내며, $r^{(t)}(u,v|\mathbf{y}) = P(U_{it-1} = u, U_{it} = v | \mathbf{Y}_i = \mathbf{y})$ 이다.

● M-단계(Maximization-step): 앞에서 구한 기댓값들을 이용해 얻어진 로그가능도 함수 (4)를 최대화하여 모수 θ 의 추정값을 얻는다. 로그가능도함수의 최대화 과정에서 모수 θ 는 모두 상대빈도, 즉 확률들이며, 로그가능도함수의 최대화를 통한 확률들의 추정값은 모두 폐쇄식으로 나타나므로 쉽게 얻을 수 있다.

이와 같은 EM알고리즘을 통해 얻어진 모수 θ 의 추정값이 일정 수준 이하로 수렴할 때까지 반복하여 최대가능도추정값을 얻을 수 있다. 다만 잠재변수가 포함된 혼합물모형(mixture model)에서 나타나는 가능도함수의 다점(multimodal) 특성으로 인하여 EM알고리즘을 통해 얻어진 최종 추정값이 대역최대값(global maximum)이 아닐 수 있다. 따라서 다수의 초기값을 무작위로 선정하여 다수의 로그가능도 함수들에 대한 최대화과정을 거쳐 얻어진 복수의 추정결과 가운데 가장 큰 값을 선택하여 최종 추정값으로 사용하는 방법을 통해 보완할 필요가 있다(Berchtold, 2004).

References

- Baltagi, B.H. (2008). *Econometric Analysis of Panel Data (4th edition)*, John Wiley & Sons.
- Bartolucci, F. (2006). Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities, *Journal of the Royal Statistical Society: series B*, 68, 155-178.
- Bartolucci F., Pennoni F., Francis B. (2007). A latent Markov model for detecting patterns of criminal activity, *Journal of the Royal Statistical Society: series A*, 170(1), 151-132.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1), 1-22.
- Hsiao, C. (2003). *Analysis of Panel Data*, Cambridge University Press.
- Liang, K.-Y., Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, 73(1), 13-22.
- MacDonald, I.L., Zucchini, W. (1997). *Hidden Markov and other Models for Discrete-Valued Time Series*, Chapman and Hall.
- Zucchini, W., MacDonald, I.L. (2009). *Hidden Markov Models for Time Series*. CRC Press.

유튜브 악플 탐지를 위한 기계학습: 스택킹 앙상블 모델의 적용을 중심으로

이신행¹

요 약

본 연구는 기계학습을 이용한 자동화 방식의 악플 분류 모델을 구현하는데 있어 단일 알고리즘의 적용이 갖는 한계에 천착했다. 이에 복수의 알고리즘 조합으로 각 알고리즘의 장점을 취하고 단점을 보완해 악플 분류의 성능을 개선하는 스택킹 앙상블 모델을 제안하였다.

주요용어 : 악플 분류, 기계학습, 스택킹 앙상블.

1. 서론

본 연구는 기계학습을 이용한 자동화 방식의 악플 분류 모델을 구현하는데 있어 단일 알고리즘의 적용이 갖는 한계에 천착했다. 특히, 분류 알고리즘을 선택할 때 정밀도와 재현율의 상호보완적 성격에 따라 어느 한 지표만을 기준으로 모델의 성능을 개선하는데 제약이 있을 뿐만 아니라 알고리즘별로 우수한 평가 지표가 달라질 수 있어 분류 성능의 극대화를 달성하는데 장애가 되는 점에 주목했다. 가령, 악플로 분류되지 말아야 하는 댓글이 악플로 잘못 분류되는 확률을 평가하는 정밀도에서 우수하거나 아니면 악플을 악플이 아닌 댓글로 잘못 분류하는 확률을 평가하는 재현율에서 우수한 알고리즘이 각기 다른 경우에 단일 알고리즘에 기반한 악플 분류 모델의 성능 개선은 한계에 부딪히게 된다. 결국, 악플 탐지에 있어 모든 평가 지표에서 가장 우수한 성능을 보이는 단일 알고리즘이 존재하지 않을 경우 서로 다른 장점을 보이는 알고리즘들을 조합하여 정밀도와 재현율 모두 개선하는 방안이 요구된다.

본 연구는 이러한 문제를 해결하는 방안으로 복수의 단일 알고리즘이 분류한 결과를 조합해 이를 바탕으로 새로운 모델을 구축함으로써 악플 분류의 성능을 개선하는 이른바 스택킹 앙상블 모델을 활용해 그 결과를 평가해보고자 한다. 사실, 댓글과 같은 단문 텍스트를 분류하는 기계학습 모델의 성능 개선은 데이터 양의 확충과 품질 향상, 그리고 알고리즘의 초모수 조정 등의 작업으로 주로 이루어져 왔다(Vidgen, Yasseri, 2020). 반면 서로 다른 단일 알고리즘을 조합해 구성된 스택킹 앙상블(stacking ensemble) 모델을 활용해 텍스트 분류 성능을 개선하고자 한 연구는 미미했다(Eom, Kim, Choi, 2020). 이에 본 연구는 logistic regression(LR), Naïve Bayes(NB), random forest(RF), support vector machine(SVM) 등 텍스트 분류에 적용되고 있는 알고리즘들을 이용해 스택킹 앙상블 모델을 구성하고 그 성능을 평가함으로써 악플 탐지 모델을 활용한 연구에 기여하고자 한다.

¹06974 대한민국 서울특별시 동작구 흑석로 84, 중앙대학교 미디어커뮤니케이션학부 조교수.

E-mail: shinlee@cau.ac.kr

2. 연구 방법

2.1 데이터

본 연구는 유튜브에서 악플을 탐지하는 기계학습 모델의 성능 평가와 앙상블 학습을 이용한 성능 향상을 위해 다음의 과정으로 데이터를 수집했다. 우선, 유튜브에서 유명인에 대한 비난 혹은 조롱을 목적으로 제작된 콘텐츠를 주로 생산하는 일명 “사이버렉카” 채널을 선정해 인기 동영상에 달린 댓글을 수집했다. 선정된 채널은 “괴인협회”와 “악인전 - 인물소개” 등으로 2022년 3월 24일 현재 이 두 채널에 게시된 영상은 각각 216개와 67개이고 이 영상들에 대한 총 조회수는 각각 199,433,611회와 71,484,906회였다. 이러한 “사이버렉카” 채널들의 최신 영상 75개로부터 댓글을 웹 크롤링(web crawling)의 방식으로 총 59,999건을 수집했다.

수집한 댓글 데이터에 대해 악플 여부를 판별해 표시하는 라벨링 작업을 진행했다. 악플은 악의를 바탕으로 타인에 대한 언어적 폭력을 행사하는 개념으로 정의하고, 이에 대한 조작적 정의로써 타인에 대한 욕설과 비방, 조롱, 차별, 음담패설이 포함된 댓글로 설정했다(Kim, Lee, 2008). 그 결과, 총 59,999건의 댓글 중 약 4.75%인 2,851건이 악플로 분류됐다. 본 연구에서는 악플 사례수에 맞춰 악플이 아닌 댓글에서 임의로 표본을 재추출하는 undersampling 방법을 활용해 총 5,702건의 댓글로 구성된 데이터를 마련했다.

2.2 특성 추출

본 연구는 악플을 탐지하는 기계학습 모델을 구현하기 위해 다음의 과정을 거쳤다. 우선, 준비된 데이터의 댓글에서 악플을 분류하는데 있어 어휘적 특성을 주요 변인으로 삼았다. 그리고 이러한 특성을 추출하기 위해 자연어처리 기법을 이용한 형태소 분석을 적용했다. 특히, 악플은 댓글의 내용적 의미로 판별된다는 전제하에 언어의 의미를 구성하는 체언, 용언, 관형사, 부사 등의 품사에 해당되는 어휘(실질형태소)들만 남기는 텍스트 처리 과정을 거쳤다. 이를 위해 R에서 Kiwi(Korean Intelligent Word Identifier)를 적용하는 “elbird” 패키지를 활용했다. 그 결과, 전체 댓글에서 추출한 실질형태소는 총 9,131개의 어휘였고 각 댓글에서 평균 13.47회(SD=12.97) 등장하고 있음을 발견했다.

2.3 스택킹 앙상블 모델의 적용

본 연구는 최적의 악플 탐지 모델을 도출하기 위해 여러 개의 단일 분류 알고리즘을 사용하고 각 알고리즘의 장단점을 파악한 뒤 스택킹 앙상블 모델을 활용해 향상된 예측 성능의 모델을 도출하고자 한다. 스택킹 기법은 알고리즘들 각각의 장점은 취하고 단점은 보완하는 방식으로 조합해 단일 알고리즘 모델에 비해 향상된 분류 성능의 모델을 생성한다. 이에 본 연구는 댓글에서 악의적 요소를 탐지해 악플을 분류하는 모델을 도출하고자 LR, NB, RF, SVM 알고리즘들을 선택해 그 성능을 살핀 후 스택킹 기법을 활용해 단일 알고리즘들의 조합한 앙상블 모델을 생성하고 개선된 성능을 평가하고자 한다. 스택킹 앙상블 모델을 구현하는 데에는 R의 “stacks” 패키지를 활용했다.

3. 분석 결과

Table 1은 학습용 세트의 특성을 단일 분류 알고리즘이 학습한 기본 분류기(Base Classifier)의 성능을 10-겹 교차 검증으로 평가한 결과를 보여준다.

Table 1. Performance of base classifiers

Algorithm	Accuracy	Precision	Recall	F_1 score
Logistic Regression (Lasso regularization with penalty = 0.01)	0.785	0.860	0.681	0.760
Naïve-Bayes	0.514	0.738	0.288	0.247
Random Forest (Trees = 100)	0.775	0.828	0.695	0.756
Random Forest (Trees = 1000)	0.781	0.841	0.695	0.761
Support Vector Machines (Radial basis function kernel with cost = 0.25)	0.502	0.501	1	0.668
Support Vector Machines (Polynomial kernel with cost = 0.25)	0.734	0.717	0.776	0.745

스태킹 앙상블 모델은 각 단일 분류기가 예측한 결과를 조합해 최적의 결과를 도출하는 메타 학습 과정을 거친다. 이때 정규화 과정을 통해 예측에 기여도가 낮은 단일 분류기는 최종 모델 구성에서 제외한다. Table 2는 각 알고리즘 별로 초모수 조절을 통해 가장 뛰어난 성능을 보인 5개의 모델이 스택킹 앙상블 모델을 구성함에 있어 어떻게 활용되었는지를 보여주는 계수이다. 총 15개의 단일 분류기 모델 중 하나의 LR 모델, 두 개의 RF 모델, 그리고 하나의 SVM 모델이 스택킹 앙상블의 메타 분류 모델에 따른 악플 분류에 영향을 미치고 있음이 나타났다.

Table 2. Stacking coefficients of algorithms

Algorithm	ID	Hyperparameter	Coefficient
Logistic Regression	LR 1	Penalty = 0.000695, Tokens = 3,000	0
	LR 2	Penalty = 0.00113, Tokens = 3,000	0.564
	LR 3	Penalty = 0.00183, Tokens = 3,000	0
	LR 4	Penalty = 0.00298, Tokens = 2,000	0
	LR 5	Penalty = 0.00483, Tokens = 2,000	0
Random Forest	RF 1	mtry = 10, min_n = 6, Tokens = 3,000	0
	RF 2	mtry = 10, min_n = 2, Tokens = 5,000	4.97
	RF 3	mtry = 10, min_n = 3, Tokens = 5,000	6.84
	RF 4	mtry = 15, min_n = 6, Tokens = 5,000	0
	RF 5	mtry = 10, min_n = 5, Tokens = 4,000	0
Support Vector Machine	SVM 1	Cost = 0.0131, Tokens= 1,000	0
	SVM 2	Cost = 0.000977, Tokens= 1,000	0
	SVM 3	Cost = 0.000977, Tokens= 5,000	0.701
	SVM 4	Cost = 0.000977, Tokens= 2,000	0
	SVM 5	Cost = 0.000977, Tokens= 4,000	0

Table 3은 스택킹 앙상블 모델에 따른 악플 분류에 활용된 단일 알고리즘 모델의 성능 평가 결과이다. 정확도와 F_1 점수의 기준으로 가장 좋은 성능을 보여준 RF 모델은 재현율에 있어 다른 분

류 알고리즘에 비해 떨어지는 성능을 보였다. 그에 반해 LR 모델의 재현율은 훨씬 더 우수한 성능을 보였다. 이러한 결과는 악플 분류에 있어 단일 알고리즘으로 구성된 모델의 한계를 보여준다. RF 모델을 사용할 경우 악의성이 없는 댓글이 악플로 잘못 분류되는 사례는 적겠지만 악플로 분류되어야 하는 댓글들을 탐지해 분류하는데 실패하는 사례가 LR 모델에 비해 많을 것이기 때문이다. 반면 LR 모델은 RF 모델에 비해 악의성 있는 댓글을 악플로 분류하지 않는 사례는 적겠지만, 악플로 분류된 댓글 중 악의성이 없는 댓글이 포함될 확률이 크다는 문제가 예상된다.

Table 4는 단일 알고리즘 모델을 조합한 메타 분류기를 이용해 평가용 세트의 댓글에서 악플을 분류한 결과로 그 성능을 살펴본 결과이다. 단일 알고리즘에 기반한 악플 분류 모델의 성능이 평가 지표에 따라 다른 양상을 보였지만 단일 분류기 모델의 예측 결과를 학습한 스택킹 앙상블 모델은 정밀도와 재현율에 있어 서로 상충되는 예측 결과를 상호보완함으로써 정확도와 F_1 점수에서 평가 결과가 향상됐음을 발견했다. 이 결과는 악플 분류에 있어 정밀도와 재현율 중 어느 한쪽의 평가에만 우수한 단일 분류 모델의 한계를 극복하는 방안으로 스택킹 앙상블 모델이 활용될 수 있음을 보여준다.

Table 3. Performance of each algorithm in stacking ensemble model

Algorithm	Hyperparameter	Accuracy	Precision	Recall	F_1 score
LR 2	Penalty = 0.00113, Tokens = 3,000	0.779	0.779	0.776	0.777
RF 2	mtry = 10, min_n = 2, Tokens = 5,000	0.798	0.862	0.707	0.777
RF 3	mtry = 10, min_n = 3, Tokens = 5,000	0.791	0.857	0.695	0.768
SVM 3	Cost = 0.000977, Tokens= 5,000	0.763	0.782	0.725	0.753

Table 4. Performance of stacking ensemble model on testing dataset

Algorithm	Accuracy	Precision	Recall	F_1 score
Stacking Ensemble	0.809	0.839	0.762	0.798

References

- Eom, H., Kim, J., & Choi, S. (2020). Machine learning-based corporate default risk prediction model verification and policy recommendation: Focusing on improvement through stacking ensemble model. *Journal of Intelligence and Information Systems*, 26(2), 105 - 129. (in Korean).
- Kim, M., & Lee, J. (2008). The publicness of the internet and solutions for trolling. *Journal of Political Communication*, 9, 5-50. (in Korean).
- Vidgen, B., & Yasseri, T. (2020). Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17(1), 66 - 78.

음원 스트리밍 플랫폼 선택속성의 중요도 분석에 관한 연구

노태호¹, 이형석²

요 약

본 연구는 음원 스트리밍 플랫폼에서 AHP 기법을 활용하여 이용자들이 음원 스트리밍 플랫폼 선택 시 고려하는 요인을 파악하여 계층화하였다. 음원 스트리밍 플랫폼 선택속성 중 1차 요인으로 안정성, 디자인, 정보성, 편리성, 개인화로 구분하였으며, 각 하위 요인을 시스템 품질, 프라이버시, 매력적인 화면, 뛰어난 레이아웃, 음원 수, 최신 업데이트, 음질, 사용 방법, 검색 기능, 멀티 디바이스, 음원 추천 서비스, 나만의 재생목록으로 설정하였다. 본 연구를 통해 국내 음원 플랫폼 기업들의 차별화된 마케팅 전략을 수립하는 데 도움을 줄 수 있을 것으로 기대한다. 또한 음원 스트리밍 플랫폼 선택 시에 발생하는 의사결정과정을 또 다른 접근방식에서 바라보았다는 점에서 학술적 의의가 있다고 하겠다.

주요용어 : 음원 스트리밍 플랫폼, AHP.

1. 서론

최근 정보통신 기술 및 멀티 디바이스의 발전으로 인해 전 세계적으로 음악산업은 성장하고 있다. 정보통신기술의 발달은 음악 서비스 이용 방법을 오프라인 방식에서 온라인 방식으로 바뀌어왔다. 그중 모바일을 이용한 음원 스트리밍 시장이 성장하면서 국내에서는 음원 서비스 기업과 통신사가 제휴를 맺고 서비스를 제공하고 있다. SKT의 음원 스트리밍 플랫폼인 ‘멜론’이 2019년까지 국내 음원 스트리밍 플랫폼 1위를 차지하였으나, 2020년부터 ‘youtube’가 점유율을 앞서기 시작했으며 ‘youtube’의 음원 스트리밍 플랫폼인 ‘youtube music’은 2021년 점유율 17.2%로 전년도 대비 가장 큰 폭으로 성장하였다(Music Industry White Paper, 2021). 전세계 음원 스트리밍 플랫폼 중 1위 플랫폼인 ‘Spotify’가 국내에 진출하면서 국내 음원 시장의 경쟁은 더욱 치열해질 것으로 보인다. 이에 따라, 이용자들이 음원 스트리밍 플랫폼을 선택할 때 어떤 요인을 중요시하는지 알아볼 필요가 있다.

따라서 본 연구는 E-service, 음원 스트리밍과 관련된 기존 연구들을 검토하여 이용자들의 음원 스트리밍 플랫폼의 중요 선택 요인들을 추출하고 AHP 기법을 활용하여 선택요인들의 상대적 중요도 파악함으로써 의사결정 과정에서 활용할 수 있는 우선순위 요인들을 도출하였다. 본 연구를 바탕으로 음원 스트리밍 플랫폼의 경쟁력 확보와 동시에 이용자들의 인식 개선에도 도움을 줄 것으로 기대한다.

¹(제1저자) 28644 인천광역시 미추홀구 인하로 100, 인하대학교 녹색금융대학원 녹색금융전공 석사과정.

E-mail: hellojh1120@naver.com

²(교신저자) 28644 충청북도 청주시 서원구 충대로1, 충북대학교 경영대학 경영학부 교수.

E-mail: hyunglee@chungbuk.ac.kr

2. 이론적 배경

2.1. 음원 스트리밍 서비스

온라인 음원서비스는 인터넷을 통해 음원을 스트리밍과 다운로드 방식으로 제공하는 서비스로 디지털 자원을 이용한 전자상거래의 일종이다(Gang, Uhm, 2010). 스트리밍이란, 전체 데이터를 다운로드하여 재생하는 것이 아니고 일부 데이터를 다운받고 나서 흘러주고 동시에 사용할 수 있는 양만큼 데이터를 지속적으로 흘러주는 것이다(Park, 2015). 음원 스트리밍 서비스는 정보통신기술의 발달로 시기와 장소에 구애받지 않고 이용할 수 있어 이용자들의 음원 관련 산업에 대한 접근성은 과거보다 크게 상승했다. 유료 온라인 음원서비스에 대한 인식 변화, K-POP의 부흥, 각종 오디션 프로그램에 의한 일반인 참여 음악의 생산과 소비에 의해 음원시장 규모가 확대되었으며, 특히 스마트폰기기의 빠른 확산으로 인한 모바일 음악 시장의 성장은 음원 유통 구조를 근본적으로 변화시키고 있다.(Lee, 2012).

2.2. 계층분석과정(AHP)

AHP(Analytic Hierarchy Process)는 Thomas L, Saaty교수에 의해 1970년대 개발된 의사결정 방법론으로 계획, 분쟁 해결을 위한 우선순위를 판단하기 위해 수학적으로 완성된 방법론이다(Saaty, 1987). AHP 분석 기법은 정량적 요소뿐만 아니라 정성적인 요소까지 모두 살펴볼 수 있어 다양한 분야에서 널리 사용되고 있다(Park et al, 2017). 또한 쌍대비교가 가능하며, 다수 관계자의 의견을 함께 반영함으로써 보다 객관적인 평가가 가능하다는 장점이 있다. 이를 통해 비교적 복잡한 선택 속성들 간에 문제를 단순화시키고, 일방적으로 강조되거나 경시되는 문제를 해결해 줄 수 있다.

AHP 분석 기법은 E-service 관련 연구에도 적용되었다. Seo(2015)는 이리닝 품질에 영향을 주는 선택 요인들을 규명하고 학생과 교수가 어떤 선택요인을 중요시하는지 그 차이를 분석하기 위해 AHP를 활용하였다. Choe, Kim(2015)은 Kano 모형과 AHP를 활용하여 소셜 커머스의 서비스 품질 요인의 우선순위를 파악하고자 하였다. Park et al.(2019)은 온라인 동영상 플랫폼의 선택 속성 간 중요도 파악을 위해 AHP 기법을 사용하였다. Oh, Lee(2021)는 전자상거래 물류 서비스 품질의 고객 만족 요인을 파악하기 위해 AHP 기법을 활용하였다. 이와같이 본 연구에서는 AHP 분석 기법의 선행 연구를 바탕으로 선택속성 요인을 도출하여 계층화하였다.

3. 연구모형

본 연구는 선행 연구를 바탕으로 음원 스트리밍 플랫폼 선택속성의 여러 요인을 도출하여 계층화하였다. 음원 스트리밍 플랫폼의 선택속성의 1차 요인으로는 안정성, 디자인, 정보성, 편리성, 개인화로 구분하였으며, 하위 요인을 시스템 품질, 프라이버시, 매력적인 화면, 뛰어난 레이아웃, 음원 수, 최신 업데이트, 음질, 사용 방법, 검색 기능, 멀티 디바이스, 음원 추천 서비스, 나만의 재생 목록으로 설정하였다. 이러한 선택속성들을 정리하여 계층화한 것을 Figure 1로 나타내었다.

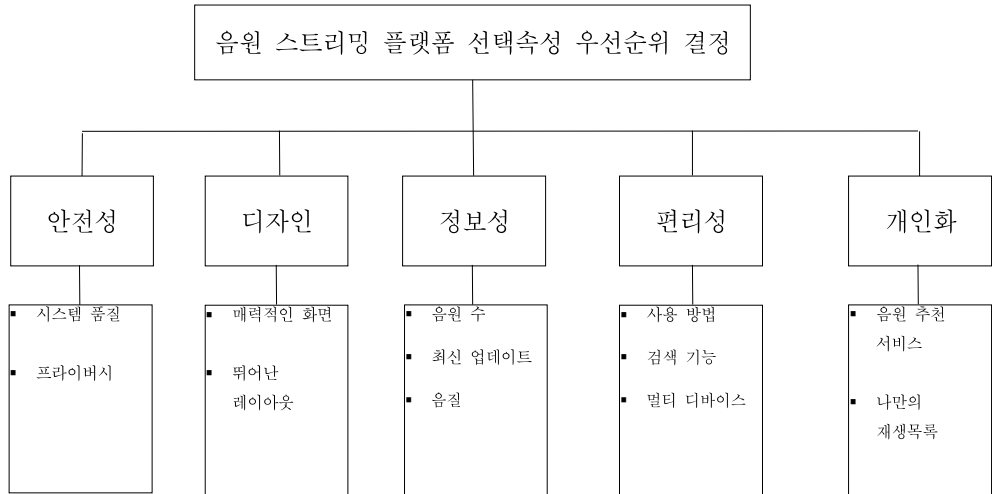


Figure 1. 연구모형

4. 연구 제언

본 연구는 선행 연구를 바탕으로 음원 스트리밍 플랫폼 선택 시에 고려되는 요인들을 도출하여 계층화하였다. 본 연구는 기존 연구와는 다르게 음원 스트리밍 플랫폼 선택 시에 발생하는 의사결정과정을 또 다른 접근방식에서 바라보았다는 점에서 의의가 있다. 또한 국내 음원 스트리밍 시장의 경쟁이 심화되는 현시점에서, 국내 음원 플랫폼 기업들의 차별화된 마케팅 전략을 수립하는 데 도움을 줄 수 있다.

향후에는 실제 음원 스트리밍 플랫폼 이용자들을 대상으로 설문조사를 실시하여 계층화된 속성들에 대한 가중치를 분석하여 우선순위를 파악할 것이다. 따라서 음원 스트리밍 플랫폼 선택 시 가장 중요한 요인과 후순위 요인들이 무엇인지 확인할 수 있으며, 성별, 연령별 차이를 분석하고자 한다.

References

- Choe, B. D., Kim, J. M. (2015). Prioritizing Factors Affecting Customer's Satisfaction in Social Commerce using Kano model and Analytical Hierarchy Process, *Korean Corporation Management Review*, 22(6), 109-224. (in Korean).
- Gang, S. M., Uhm, G. H. (2010). The Antecedents of Switching Cost and its Effect on Customer Loyalty in Digital Music Service Industry, *Asia pacific journal of information systems*, 20(2), 157-180. (in Korean).
- Lee, S. J. (2012). Examination of Factors Influencing Switching Intention in Mobile Music Service: focusing on Moderating Effects of Attractiveness of Alternatives and Switching Costs, *The Journal of the Korea Contents Association*, 12(10), 453-465. (in Korean).
- Oh, H. M., Lee, H. S. (2021). A Study on Cross-Border e-commerce with Logistics Service Quality Impact on Customer Satisfaction Using AHP Technique, *Korean Society of Transportation*, 39(2), 204-219. (in Korean).
- Park, I. K. (2015). A Study on the Usability of Graphic User Interface by the User Behavior in a Mobile Music Streaming App, *Design Convergence Study*, 14(2), 151-168. (in Korean).

- Park, J. E., Kim, S. W., Kim, E., Kim, S. C. (2017). A Multi-criteria Approach toward Identifying Determining Factors of T-Commerce Adoption : Comparing T-Commerce, TV Home Shopping, and Online Shopping, *Korean Journal of Broadcasting and Telecommunication Studies*, 31(1), 5-37. (in Korean).
- Park, S. H., Lee, D. S., Park, S. H. (2019). A study on Critical Factor of Selecting Online Video Platform by Using AHP, *Journal of Korean Society of Industrial and Systems Engineering*, 42(4), 173-182. (in Korean).
- Saaty, R. W. (1987). The Analytic Hierarchy Process - What It Is and How It Is Used, *Mathl Modelling*, 9(3-5), 161-176.
- Seo, C. G. (2015). Study on the e-Learning Quality Evaluation Using AHP, *The Journal of Internet Electronic Commerce Research*, 15(4), 337-347. (in Korean).

지역사회의 안전수준을 측정하기 위한 사회안전지수의 개발

정진우¹, 장안식²

요 약

사회가 복잡해지고 시민들의 욕구가 다양해짐에 따라 과거에는 체감하지 못했던 사회적 안전에 대한 시민들의 요구가 높아졌음에도 안전에 관한 기존의 지표들이 범죄, 사고 등 신체적 안전에 초점을 두고 있어 안전·안녕에 대한 시민들의 다양한 요구를 반영하는 데에 한계를 지니고 있으며, 안전 실태를 드러내는 통계를 중심으로 산출하여, 안전에 대한 시민들의 체감도를 반영하지 못한다는 제약이 있다. 이러한 한계를 보완하고자 사회안전에 대한 다양한 지표항목을 발굴하고, 지역주민들의 체감도를 측정해 이를 종합적으로 반영한 사회안전지수를 개발하였다. 사회안전지수는 안전에 관한 기존의 지표들과 비교할 때, 보다 다양한 층위의 요소들을 사회안전의 요소로 간주하여 지수산출에 포함하고, 해당 지역주민들의 안전에 대한 체감도를 반영했다는 점에서 기존 지표들과 차별점을 지닌다. 다만, 온라인 조사라는 제약으로 인해 전국의 모든 시군구를 분석하지 못했다는 한계가 있다.

1. 서론

급격한 사회변화로 인해 미래에 대한 불확실성이 커지고, 이로 인해 개인의 삶과 안녕을 위협하는 불안요인들이 증가하고 있다. 특히 코로나19 팬데믹을 겪으면서 개인 삶의 방식과 양태가 영향을 받으면서 변화하였고, 이 과정에서 과거에는 체감하지 못했던 사회적 안전에 대한 시민들의 요구가 높아졌다.

이러한 사회변화에 조응해 UN, OECD 등 국제기구 또는 대학, 민간기구에서도 개인과 사회의 안녕과 안전 등에 대한 다양한 지표들을 생산해 공표하고 있으며, OECD의 더 나은 삶 지수(Better Life Index), 영국 이코노미스트 인텔리전스 유닛(EIU)의 안전도시지수, 미국 존스홉킨스대학 보건안전센터, 핵위협방지구상(NTI) 등이 주도하는 세계보건안전지수(Global Health Security Index), 글로벌 안전과학회사 UL의 UL 글로벌 안전지수(UL Safety Index) 등이 여기에 해당한다.

한편, 우리나라 행정안전부가 공표하는 지역안전지수는 안전에 관한 각종 통계를 활용하여 자치단체별 안전수준을 계량화하여 1~5등급으로 구분한 것으로, 안전사고 사망자의 체계적 감축을 위해 안전관련 정책이나 사업을 결정하고 집행하는 지방자치단체의 관심과 자율적 개선 유도를 목적으로 하고 있다.

행정안전부의 지역안전지수는 지역안전을 교통사고, 화재, 범죄, 자연재해, 생활안전, 자살, 감염병 등의 7개 분야로 구분하고, 각 분야별 사망자수 및 발생건수 등 결과에 관한 항목을 위해지표, 사망자 발생 등의 원인이 될 수 있는 지표를 취약지표, 사망자 발생 등을 경감할 수 있는 지표를

¹06650 서울특별시 서초구 반포대로24길 76, ㈜케이스탯리서치 이사.

²06650 서울특별시 서초구 반포대로24길 76, 케이스탯 공공사회정책연구소 소장.

경감지표로 세분화하여 이를 인구 또는 행정구역 면적 등으로 표준화하여 산출한다.

지역안전지수는 지역안전을 위해지표, 취약지표, 경감지표로 구분하여 측정하고, 이를 종합하여 지수를 산출했다는 점에서 큰 의미를 지닌다. 하지만, 객관적인 통계만을 이용하여 지역안전수준을 측정하였고, 그 지역주민들이 체감하는 안전수준이 반영되지 않았다는 점에서 한계가 있다.

따라서 객관적 통계지표 외에 주민들이 체감하는 부문별, 영역별 안전수준을 반영한 지표의 필요성이 제기되며, 대한민국 국민이 살아가는 지역사회의 사회안전도와 주민들의 불안 정도를 진단하고, 불안 해소를 위한 국가와 지방자치단체의 역할을 조명하기 위해 사회안전지수를 개발하였고, 사회안전지수는 지방자치단체와 국가의 정책적 대응을 유도해, 국민의 불안요인을 약화 또는 해소하는 것을 목표로 한다.

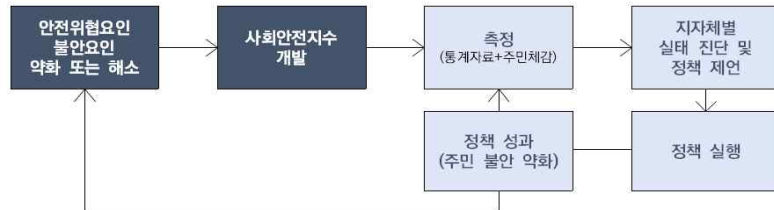


Figure 1. 사회안전지수의 개발 배경

2. 사회안전지수의 지수체계

먼저 시민들의 사회안전에 영향을 미치는 요인을 크게 △경제활동 △생활안전 △건강보건 △주거환경의 네 가지 차원으로 구분하고, 각 차원별로 객관적 안전수준을 보여주는 정량지표(객관적 지표)와 해당 차원에 대한 주민들의 체감도로 측정한 정성지표(주관적 지표)로 구성하였다.

각각의 차원은 다시 세부영역으로 세분화하였는데, 경제활동 차원은 △소득 △고용 △복지 △노후, 생활안전 차원은 △치안 △소방 △교통안전, 건강보건 차원은 △건강상태 △의료환경 △의료충족, 주거환경 차원은 △대기·환경 △주거·보육·교육 △문화·여가 △인구변동으로 구성하였다.

정량지표는 중앙정부 또는 지방자치단체, 공공기관에서 발표하는 공식 통계 및 조사자료를 활용하였으며, 4개 차원별 각 세부영역을 구성하는 정량지표항목은 총 44개로 구성하였다.

한편, 정성지표는 (주)케이스터리서치와 온라인 전문조사회사인 (주)피앰아이가 공동으로 전국 229개 시군구에 거주하는 만 18세 이상 성인남녀를 대상으로 실시한 <사회안전인식조사>를 통해 측정하였다.

<사회안전인식조사>의 조사대상자는 전국 226개 기초자치단체와 세종특별자치시, 제주시, 서귀포시에 거주하는 만 18세 이상 성인남녀이며, 조사방법은 구조화된 설문지를 이용한 온라인 설문조사이며, 지난 2021년 10월 6일부터 31일까지 약 3.5주 간 실시하였다.

전체 조사참여자 수는 32,762명이나, 229개 시군구 중 50표본 이상의 표본이 확보된 155개 시군구만 분석에 포함하였으며, 50표본을 확보하지 못한 74개 시군구는 분석에서 제외되었다.

<사회안전인식조사>의 조사항목은 총 35개 질문과 15개의 배경질문을 포함해 총 50개의 질문으로 구성하였다. 각 질문항목의 보기에는 7점 리커트 척도를 사용하였으며, 일부 항목은 역방향 척도를 제시하였으므로, 분석 시에 역방향으로 리코드하여 분석하였다.



Figure 2. 사회안전지수의 구성 체계

3. 사회안전지수의 산출과 공표

분석대상 155개 시군구를 대상으로 사회안전지수를 산출하기 위해서는 다음과 같은 과정을 거쳤다.

먼저 측정 단위가 다른 지표 항목들 간 합산이 가능하도록 각 지표 항목의 표준정규누적분포값을 계산하여 합산하는 표준화 과정을 거쳤다. 이렇게 변환한 지표 항목들의 점수를 정량지표와 정성지표로 구분하여, 각각 지표 항목 간 동일한 비중합산 가중치로 합산하였으며, 이를 수식으로 표현하면 아래와 같다.

$$s_i = \text{Normalize} \left(\sum_{n=1}^k \text{Standardize}(f_n) \right)$$

이 식에서 S_i 는 각 차원을 구성하는 정량지표 및 정성지표의 사회안전지수 점수를 의미하며, f_n 은 해당 차원의 정량지표와 정성지표를 구성하는 개별 지표값을 나타낸다. **Standardize**(Z-Score에 따른 표준화)는 각 개별 지표를 합산하기 위해, **Normalize**(정규화)는 비교의 수월성을 위해 정규화하는 과정을 의미한다.

다음으로, 각 차원의 정량지표와 정성지표를 각각 60%와 40%의 가중치를 적용해 합산하여 차원별 사회안전지수를 산출하였다.

이렇게 산출한 차원별 사회안전지수를 합산하여 최종 사회안전지수를 산출할 때에는 경제활동 차원 30%, 생활안전 차원 30%, 건강보건 차원 30%, 주거환경 차원 10%의 가중치를 적용하였으며, 이 과정을 도식화하면 Figure 3과 같다.

이렇게 산출한 시군구별 사회안전지수를 바탕으로, 분석대상 155개 시군구를 아래와 같은 기준에 따라 5개 등급으로 구분하여 제시하였다(Table 1).

최종적으로 분석 대상 155개 시군구 중 상위 50개 시군구에 대해서만 사회안전지수 점수와 순위를 머니투데이를 통해 공개하였다.

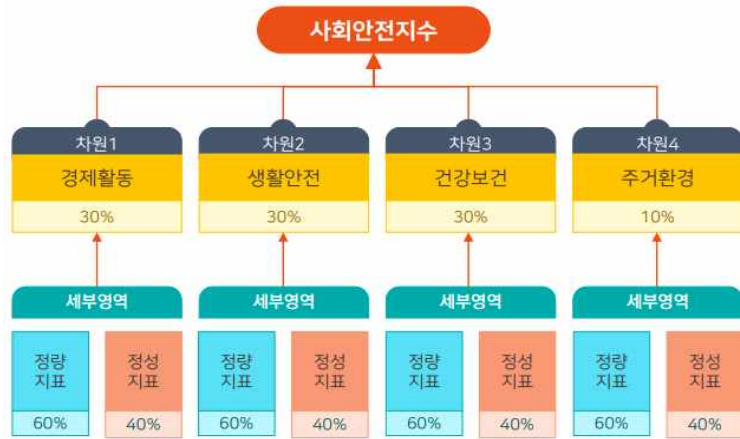


Figure 3. 사회안전지수의 산출

Table 1. 사회안전지수의 등급별 배분

A등급	B등급	C등급	D등급	E등급	계
30개	30개	35개	30개	30개	155개

4. 결론

안전에 대한 다양한 시민들의 욕구를 반영하고, 통계자료에 근거한 객관적 지표 외에 안전에 대한 시민들의 체감도를 반영하고자 사회안전지수를 개발하였다. 사회안전지수는 안전에 관한 기존의 지수들과 비교할 때, 보다 다양한 층위의 요소들을 사회안전의 요소로 간주하여 지수산출에 포함하고, 해당 지역주민들의 안전에 대한 체감도를 반영했다는 점에서 기존 지수들과 차별점을 지닌다. 다만, 온라인 조사는 제약으로 인해 전국의 모든 시군구를 분석하지 못했다는 한계가 있다.

사회안전지수로 살펴본 사회안전 유형화와 다층모형의 적용

장안식¹, 정진우²

1. 서론

왜 현대사회를 사는 우리들은 알 수 없는 불안에 의해서 고통 받고 있는가? 현대사회에서 안전은 시장의 힘에 의해 자유의지적인 선택과 자기 실천으로 바뀌었고, 그 부산물로서 불안이라는 것이 만들어졌다(Bauman, 2000). 존재론적 불안은 대부분의 인간이 그들의 행위에 대한 사회적·물질적 환경에서 자기 정체성의 유지·지속과 관련한 확신과 관련이 깊다(Giddens, 1990). 존재론적 불안은 그래서 때로는 인지적이고 경험적이라기보다는 감정적이고 무의식적으로 간주되기도 한다. 인간은 그들의 불확실성/의문으로부터 보호할 수 있는 방식으로 사회를 이해할 필요가 있고 그렇게 함으로써 존재의미를 부여한다.

우리는 주관적 인지를 객관적 실재로서 받아들인다. 현대인들은 예전에 비해 보다 유동적이고 변화가 많으며 하루살이와 같은 생활을 감내해야만 한다. 현대사회에서 불확실성은 사람들의 삶에서 중요한 요소로 자리매김했고, 이러한 삶들은 더욱 더 유동적이고 다양하게 되었다. 후기 근대성이 가져온 기회와 풍요, 그리고 선택과 자유의 확장은 근본적인 취약성과 불안, 그리고 불확실성이란 값을 치르게 만들었다(Sennet, 1998).

불안한 사회를 살아가는 사람들의 안전에 대한 요구는 끝이 없다(Zedner, 2003). 안전의 요구들은 더욱 더 많은 안전장치 혹은 예방책들을 고안했고, 우리가 가지고 있는 취약성에 대해서 더 많이 주지하게 되었다. 안전의 장치가 우리 삶에서 준비되어져 있지 않을 시 우리는 마치 벌거벗은 채로 사자굴에 들어가는 것처럼 느끼게 되었다(Loader, 2008). 일반적인 사회 불안의 요소들은 취약성과 무기력함 그리고 불확실성의 감정으로부터 나온다. 사회불안의 만연은 때때로 우리 주변을 타자화하고 주변화된 사회 집단들을 희생양으로 삼는 것을 정당화하고 사회통제를 강화하는 것을 정당화해준다.

사회안전 혹은 사회불안으로 측정되어져온 것들은 지금 처해 있는 사회적 환경의 특정 위험에 대한 인지된 위험보다는 더 폭넓게 퍼져 있는 어떠한 것이 그 근원에 있다. 사회안전 혹은 불안에 대한 대중적 감정은 그들의 삶의 공간과 환경 속에 배태되어 있다. 현대인들의 감정은 그들의 삶의 터전이 되는 지역구조적인 환경 속에 배태되어 있으며, 그들이 가지는 삶의 불안이나 사회적 불평등에 대한 보다 근본적인 관심은 사회안전/사회불안에 영향을 줄 수 있다.

사회안전에 대한 사람들의 감정이 다른 불안요인들을 표현하고 있다는 생각은 새로운 것이 아니다. 이러한 측면에서 보면, 일상적 안전의 의미로만 측정되어져온 사회안전은 어떤 특정 위험보다는 더 폭넓게 퍼져 있는 광범위한 어떠한 것에 그 근원을 두고 있다고 할 수 있다. 사회안전에 대한 평가는 지역사회에 대한 일차적 그리고 이차적 지식과, 관찰되는 물리적/사회적 무질서의 증

¹06650 서울특별시 서초구 반포대로24길 76, 케이스텍 공공사회정책연구소 소장.

²06650 서울특별시 서초구 반포대로24길 76, (주)케이스텍리서치 이사.

거들, 그리고 지역사회의 구성적 변화에서 파생되는 편견들로부터 유래한다. 사회의 경제적 미래는 사람들로 하여금 그들이 통제할 수 없는 어떠한 사건들에 대해서 취약하게 느끼게 할 수 있다. 사회안전과 관련된 지역 문제들은 은유적으로 표현된다. 사회안전을 위협하는 은유적 표현들은 실업이나 고용과 같은 경제적 상황, 정주여건과 같은 물리적 환경의 퇴락, 삶의 건강성을 위협하는 건강보건의 문제 등 다양한 사회적 신호들을 포함한다. 사회안전은 삶에 대한 은유로 나타날 수 있고, 이러한 방식을 통해서 개인들은 인지된 사회안전의 수준과 함께 자신들의 환경을 표현하고 이해한다.

현대 사회의 불확실성에서 파생되는 광범위한 불안은 일종의 전이된 불안이라고 할 수 있다. 불확실한 하루살이 인생을 감내해야만 하는 상황에서 사람들의 삶은 불확실성으로 점철될 수밖에 없고, 그러한 불확실성의 감정은 그들의 삶에서 ‘안전’을 위협한다. 삶에 배태되어 있는 불안의 증대는 결과적으로 ‘사회안전’ 수준의 격하로 나타나게 된다.

그렇다면 왜 “사회안전지수”? Social Security/Safety Index?

안전의 사전적 정의는 ‘편안하고 온전하여 탈이나 위험성이 없고 사건사고가 생길 위험이 없이 온전한 것’이다. ‘편안할 안’은 무사한 것으로 아무 일이 없다는 것이고 ‘온전할 전’은 온전히 갖추어져 있다는 의미로 해석할 수 있다. 또 다른 의미로서 ‘security’의 의미는 무언가로부터 떨어져 나간다는 의미의 ‘se’와 마음을 쓰고 염려하고 걱정한다는 의미의 ‘care’가 결합한 것이다. 즉, ‘security’는 걱정이 없고 걱정에서 자유롭다는 것을 의미한다. 결과적으로 ‘사회안전’이라는 의미는 신체적인 위험에서 자유롭다는 의미, 경제적인 위험에서 자유롭다는 의미, 그리고 불확실성으로 생기는 삶의 불안을 최소화할 수 있다는 의미로 확장된다.

“2021 사회안전지수(Korea Security Index 2021)”는 우리 사회의 안전과 불안감에 영향을 주는 생활안전, 경제활동, 건강보건, 주거환경 등 4대 분야를 조사·분석해 점수화한 평가지수다. 우리나라는 지난 반세기 동안 ‘압축적 근대화’를 겪으며 사회 모든 면에서 성장을 일구어왔다. 2021년 유엔 무역개발회의(UNCTAD)는 한국의 지위를 개발도상국에서 선진국 그룹으로 변경하였다. 이는 1964년 유엔무역개발회의가 설립된 이래 개도국에서 선진국 그룹으로 지위를 변경한 첫 사례이다. 반세기 조금 넘는 시간 동안 이만큼의 성장을 이룬 사례는 찾아보기 어렵다. 하지만 이러한 괄목할 만한 성장의 이면에는 높은 자살률(OECD 1위, 세계보건기구(WHO) 3위, 청년 자살 증가율 OECD 2위), 높은 노인빈곤율(OECD 회원국 중 1위) 등이 존재한다.

우리사회는 선진국에 접어들었고 부유해졌지만, 과연 얼마나 안전(혹은 불안)할까? “2021 사회안전지수(Korea Security Index 2021)”는 우리사회에서 사회안전이라는 주제가 의미 있게 논의될 수 있도록 사회안전을 구체화하고 측정하고자 한다. 급격한 사회변화로 인해 미래에 대한 불확실성이 커지고, 이로 인해 개인의 삶과 안녕을 위협하는 불안요인들이 증가하고 있다. 이러한 환경에서, “사회안전지수(Korea Security Index)”는 대한민국 국민이 살아가는 지역사회의 사회안전도와 주민들의 불안 정도를 진단하고, 불안 해소를 위한 국가와 지방자치단체의 역할을 조명하기 위해 기획되었다. 이를 위해, “사회안전지수(Korea Security Index)”는 사회안전의 개념을 경제, 생활안전, 건강보건, 정주환경 등의 영역으로 구분하여 측정하고 개념화하여 지수화를 시도하였다. “사회안전지수(Korea Security Index)”의 각 항목은 중앙정부 및 지자체, 공공기관에서 발표한 통계 및 조사 자료를 활용한 정량자료와 해당 시군구에 거주하는 주민들의 설문조사로 측정한 정성자료로 구성하였다.

2. 사회안전지수를 이용한 유형화 작업



Figure 1. 안전-격차-유형화

안전지수 격차
-
종합

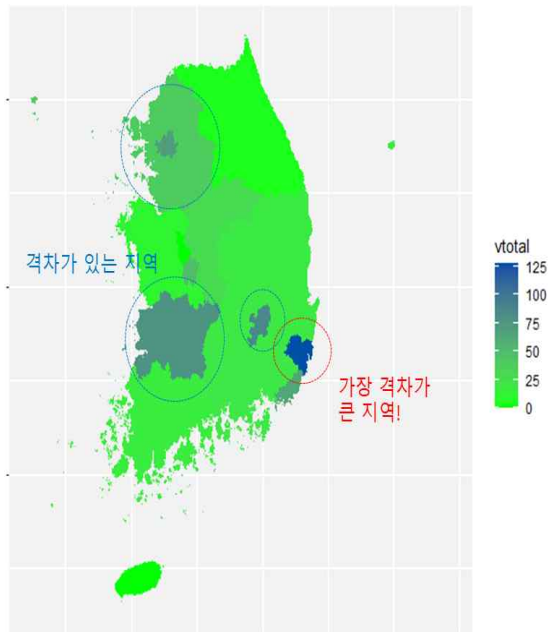


Figure 2. 안전지수 격차-종합

안전지수 격차-종합

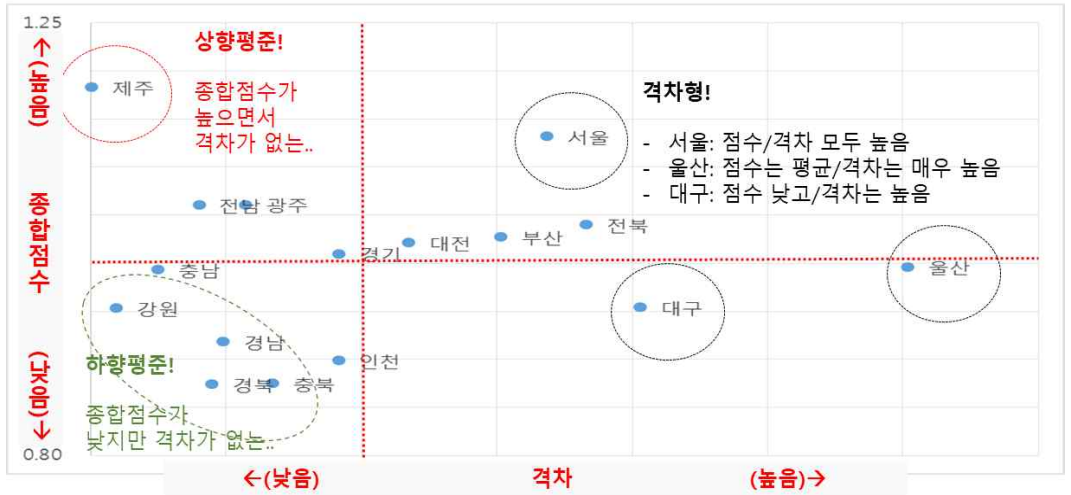


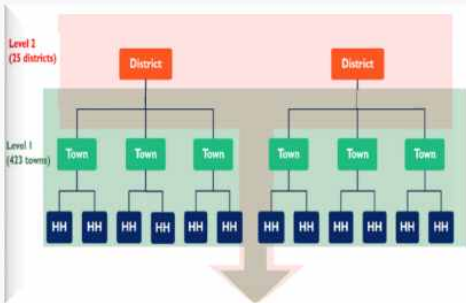
Figure 3. 안전지수 격차-종합

3. 사회안전지수에 다층모형 적용

- 사회안전지수와 이중 행정/민간 자료 결합방안 & 이중 자료의 융결합 방안
- 다층모형/다수분모텔(Multi-level Analysis)의 적용

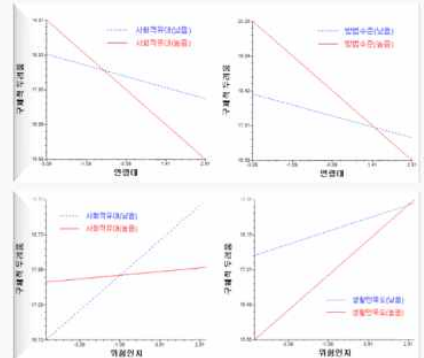
접근 방향

- 단일 층위에서의 데이터 결합방식에서 포착할 수 없는 특성에 대해 2개 이상의 층을 구성하는 다층 모형을 구성하고, 이를 통해 다양한 형태의 이중 데이터 결합을 위한 범위/영역을 확장함



적용 예시

- 연령의 효과는 부적이지만, 어떤 동네에 사느냐에 따라 그 효과가 다르게 나타남 (적용) 각 구별 맥락(자료)에 따라 개인이 느끼는 효과가 다르게 나타난다는 것을 분석적으로 보여줄 수 있음



간호사 소진 경감 중재 프로그램의 효과: 체계적 문헌고찰 및 메타분석

이미란¹, 차지영²

요약

본 연구는 간호사의 소진을 경감시키기 위해 수행된 중재가 간호사의 소진 경감에 미치는 효과를 확인하기 위한 체계적 문헌고찰 및 메타분석 연구이다. 문헌선정을 위해 7개의 국외 데이터베이스와 2개의 국내 데이터베이스를 통해 2011년 1월 1일부터 2020년 12월 31일까지 출판된 소진 관련 중재 연구를 검색하였다. 문헌검색 결과, 30개의 연구가 체계적 문헌고찰에 포함되었고 이 중 27개의 연구를 메타분석하였다. 체계적 문헌고찰 결과, 간호사의 소진을 줄이기 위한 중재로는 그룹으로 진행된 마음챙김 프로그램이 주로 이루어지고 있었다. 메타분석 결과, 27개의 연구 중 16개 연구에서 소진을 하나의 개념으로 측정하였으며 11개의 연구에서는 소진을 3개의 하위 개념(감정적 소진, 비인간화, 개인적 성취)으로 측정하였다. 소진을 하나의 개념으로 측정한 16편의 연구에 대한 메타분석결과 중재를 통해 간호사의 소진을 감소시키는 효과가 있는 것으로 나타났지만, 소진을 하위개념으로 나누어 본 11개의 연구에 대한 메타분석 결과에서는 중재가 감정적 소진과 비인간화 감소에는 효과가 있었으나 개인적 성취는 향상시키지 못했다. 따라서 간호사의 소진은 중재를 통해 완화될 수 있으며, 소진을 하위 개념으로 평가한 경우, 정서적 소진과 비인간화의 경감에는 효과가 확인되었으나 개인적 성취의 향상은 확인할 수 없었다.

주요용어 : 소진, 메타분석, 마음챙김, 간호사, 심리사회적 중재, 체계적 문헌고찰.

¹62396 광주광역시 광산구 광주여대길 69, 광주여자대학교 간호학과 조교수. E-mail: imr@kwu.ac.kr

²03760 서울특별시 서대문구 이화여대길 52, 이화여자대학교 간호대학 부교수. E-mail: chiyoung@ewha.ac.kr

간호대학생의 비판적 사고성향, 자기주도적 학습능력과 학업적 자기효능감

오운정¹, 이은미²

1. 서론

비판적 사고란 문제해결과 의사결정을 이끌어 내기 위한 과정으로, 빠르게 변하는 간호현장에서 의사결정에 전문성을 발휘할 수 있으며, 이러한 의사결정 결과는 임상판단력으로 나타난다. 자기주도적 학습능력이란 학습자 스스로가 학습과 정의 주도권을 갖고 자신의 학습욕구를 진단하고 목표를 설정하며 적합한 학습전략을 선택 및 실행하여 자신이 성취한 학습결과를 스스로 평가하는 과정을 말한다. 학업적 자기효능감은 학업 상황에서 과제를 수행하기 위해 조직하고 실행하는 자신의 능력에 대한 신념을 의미한다. 이는 학습의 결과인 학업 성취를 가장 많이 예측하는 변인으로써, 학습 행동의 여부를 결정하게 되는 동기로 작용한다. 본 연구는 간호대학생의 비판적 사고성향, 자기주도적 학습능력 및 학업적 자기효능감의 관계를 알아보고 학업적 자기효능감에 영향을 미치는 요인을 확인하기 위해 시도되었다.

2. 연구방법

2.1 연구설계 및 자료분석방법

본 연구는 간호대학생의 비판적 사고성향, 자기주도적 학습능력, 학업적 자기효능감을 파악하여 학습자의 학업능력을 증진시키기 위한 지도방안을 모색하기 위한 기초자료를 제공하기 위한 서술적 조사연구이다. 본 연구의 대상은 G광역시 소재 일개 4년제 대학의 간호대학생으로 선정하였다. 자료수집 기간은 2021년 5월 1일부터 5월 25일까지였으며, 자료수집은 자가보고식 질문지를 이용하였다. 연구대상자 수는 G power 3.1.2를 이용하였고, 유의수준 .05, 검정력(1-β) .90 효과크기 .15로 최소 141명으로 탈락률을 고려하여 185명을 선정하여 최종 170부를 최종분석에 이용하였다. 수집된 자료는 SPSS/Win 23.0 program을 이용하여 분석하였다. 대상자의 일반적 특성은 빈도와 백분율을 이용하였고, 비판적 사고성향, 자기주도적 학습능력 및 학업적 자기효능감의 차이는 Independent t-test, ANOVA를 이용하였다. 대상자의 비판적 사고성향, 자기주도적 학습능력, 학업적 자기효능감 간의 상관관계는 Pearson's correlation coefficient를 이용하였으며, 대상자의 학업적 자기효능감에 미치는 영향은 Stepwise multiple regression analysis를 이용하여 분석하였다.

2.2 연구 도구

비판적 사고성향은 Yoon(2004)이 개발한 도구로 총 27문항이며, 5점 척도로 점수가 높을수록 비

¹62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: netagnes@nambu.ac.kr

²62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: jjang9312@nambu.ac.kr

판적 사고성향이 높은 것을 의미하는데, 본 연구에서 Chronbach's $\alpha = .85$ 이었다. 자기주도적 학습능력은 Lee 등(2003)이 개발한 도구로 총 45문항이며, 5점 척도로 점수가 높을수록 자기주도적 학습능력이 높은 것을 의미하는데, 본 연구에서 Cronbach's $\alpha = .94$ 이었다. 학업적 자기효능감은 Kim과 Park(2001)이 개발한 도구로 총 28문항이며, 5점 척도로 점수가 높을수록 학업적 자기효능감이 높음을 의미하는데, 본 연구에서 Cronbach's $\alpha = .92$ 이었다.

3. 연구결과

본 연구는 대상자는 평균 연령은 21.1세였으며, 여학생이 148명(87.1%)이었다. 성격유형은 '내향형' 117명(68.8%), '외향형' 53명(31.2%)이었고, 학교성적은 '중' 98명(57.6%), '상' 33명(19.4%), '하' 39명(22.9%) 이었고, 학과만족도는 '만족' 81명(47.6%), '보통' 86명(50.6%)으로 나타났다. 선호하는 학습유형은 '강의' 122명(71.8%)였다. 대상자의 일반적 특성에 따른 학업적 자기효능감은 성별 ($t=4.05$, $p=.007$), 성격유형($t=-5.35$, $p<.001$), 학업성취도($F=7.32$, $p=.001$), 학과만족도($F=12.52$, $p<.001$), 선호하는 학습유형($F=14.65$, $p<.001$)에서 통계적으로 유의한 차이가 있었다(Table 1).

Table 1. Differences of Critical thinking, Self-directed learning ability, Academic self-efficacy by General Characteristics

Characteristics	Categories	n(%) or M(SD)	Critical thinking		Self-directed learning ability		Academic self-efficacy	
			M(SD)	t or F(p) Scheffe	M(SD)	t or F(p) Scheffe	M(SD)	t or F(p) Scheffe
Age(yr)	range 19-44	21.1(3.21)						
Gender	Male	22(12.9)	3.82(0.46)	2.72*	3.69(0.51)	2.486*	3.45(0.65)	4.05*
	Female	148(87.1)	3.57(0.40)	(.007)	3.41(0.49)	(.014)	2.94(0.53)	(.007)
Personality	Introverted	117(68.8)	3.52(0.39)	-3.65*	3.33(0.45)	-4.488	2.87(0.47)	-5.35
	Extroverted	53(31.2)	3.77(0.42)	(<.001)	3.69(0.52)	(<.001)	3.33(0.63)	(<.001)
School achievement	Well	33(19.4)	3.75(0.37)	5.88*	3.66(0.45)	11.473*	3.20(0.55)	7.32*
	Average	98(57.6)	3.62(0.43)	(.003)	3.49(0.52)	(<.001)	3.06(0.60)	(.001)
	Poor	39(22.9)	3.42(0.35)	a,b>c	3.15(0.36)	a,b>c	2.74(0.39)	a,b>c
Major satisfaction	Satisfactiona	81(47.6)	3.71(0.41)	6.11*	3.62(0.50)	11.436*	3.22(0.53)	12.52*
	Moderateb	86(50.6)	3.50(0.40)	(.003)	3.27(0.44)	(<.001)	2.81(0.54)	(<.001)
	Dissatisfactionc	3(1.8)	3.46(0.51)	a>b	3.48(0.78)	a>b	3.21(0.37)	a>b
Residential Form	Lecturea	122(71.8)	3.53(0.40)		3.35(0.47)		2.87(0.43)	
	case-based learningb	22(12.9)	4.02(0.43)	6.75*	3.96(0.48)	8.343*	3.78(0.65)	14.65*
	problem-based learningc	14(8.2)	3.67(0.38)	(<.001) d>a	3.48(0.53)	(<.001) b>d, c,d>a	3.25(0.57)	(<.001) b>a, a,b>d
	discussion-based learningd	12(7.1)	3.60(0.41)		3.44(0.50)		3.22(0.55)	

본 연구에서 학업적 자기효능감은 비판적 사고성향과 ($r=.64$, $p<.001$) 간에 정적 상관관계가 있었고, 자기주도적 학습능력($r=.78$, $p<.001$) 간에도 정적 상관관계가 있었다(Table 2).

본 연구에서 대상자의 학업적 자기효능감에 미치는 영향을 알아보기 위해 일반적 특성 중 통계적으로 유의한 차이가 있었던 학업성취도와 비판적 사고성향, 자기주도적 학습능력을 회귀분석식에

투입한 후 다중회귀분석을 실시한 결과 학업적 자기효능감에 영향을 미치는 요인은 자기주도적 학습능력($\beta=.78$, $p<.001$)로 나타났다(Table 3).

Table 2. Correlation of Critical thinking, Self-directed learning ability, Academic self-efficacy

	Critical thinking	Self-directed learning ability	Academic self-efficacy
Critical thinking	1		
Self-directed learning ability	.75* ($<.001$)	1	
Academic self-efficacy	.64* ($<.001$)	.78* ($<.001$)	1

Table 3. Factors Influencing Academic self-efficacy

Variable	B	S.E	β	t	p	AdjR ²	F	p
Self-directed learning ability	0.89	0.056	.78	16.25	$<.001$.609	263.96	$<.001$

4. 논의 및 제언

본 연구에서 학업적 자기효능감에 가장 높은 영향을 미치는 요인으로는 자기주도적 학습능력으로 나타났다. 이러한 결과를 토대로 간호대학생의 학업능력 향상을 위해서는 자기주도적 학습능력이 필요하며, 이를 강화하는 교육과정 개발과 다양한 보조프로그램 개발이 필요한 것으로 사료된다.

References

- J. Yoon. (2002), *Development of an Instrument for the Measurement of Critical Thinking Disposition In Nursing*, Unpublished master's thesis, Catholic University. Seoul.
- S. J. Lee, Y. K. Jang, H. N. Lee & K. Y. Park, (2003). A study on th development of life-skills: communication, problem solving, and self-directed learning (RR2003-15-03). *Korean Educational Development Institute*.
- A. Y. Kim, I. Y. Park, "Construction and validation of academic self-efficacy scale", *The Journal of Educational Research*, Vol. 39, No. 1, pp. 95-123, 2001.

간호대학생의 자기인식, 타인인식과 의사소통능력 간의 관계

오운정¹, 이은미²

1. 서론

간호사와 대상자의 관계형성과정을 의미하는 의사소통은 간호목적 달성의 최적의 수단이므로 간호교육과정에서 학습활동을 통한 대상자와 치료적 인간관계를 형성하고 적절한 의사소통과정으로 의료상황에 적합한 대처와 간호문제의 현명하고 적절한 의사소통과정으로 의료상황에 적합한 대처와 간호문제의 현명한 해결능력을 습득할 수 있어야 한다. 자기인식이란 자기 자신의 내적 사고나 감정에 주의를 기울이고, 타인에게 영향을 주는 사회적 대상으로서 자기를 인식하며, 다른 사람들에 의해 평가된다는 것에 대한 우려나 타인들의 존재에 대한 불편함을 인식하는 것을 의미한다. 타인인식이란 다른 사람에 대한 주의나 관심 및 의식으로, 다른 사람의 태도나 모습에 주의를 기울이고 표정이나 태도가 무엇을 의미하는가를 생각하며, 타인의 내적인 사고나 감정을 생각하는 것을 의미한다. 의사소통능력은 장시간의 훈련으로 향상시킬 수 있으므로 대학 교육과정에서부터 지속적인 교육이 필요하다. 본 연구는 간호대학생을 대상으로 자기인식, 타인인식과 의사소통능력의 관계를 파악하고, 간호학생의 의사소통 경험 과정에서 겪는 어려움과 이를 해결해 가는 과정을 포괄적이고 심도 있게 알아보고자 시도되었다.

2. 연구방법

2.1 연구설계 및 자료분석방법

본 연구는 간호대학생의 자기인식, 타인인식과 의사소통능력 간의 관계를 파악해보기 위해 시도된 서술적 조사연구이다. 본 연구는 G시에 소재한 S대학 간호학과 학생을 대상으로 본 연구의 목적 및 설문지 작성요령과 유의사항을 충분히 설명하고 설문지를 배부하였다. 자료수집은 2021년 9월부터 10월까지 이루어졌고, 대상자 선정 기준은 연구의 목적을 이해하고, 의사소통이 가능하며 연구 참여를 수락한 자로 최종 222부를 최종분석에 이용하였다. 수집된 자료는 SPSS/Win 23.0 program을 이용하여 분석하였다. 대상자의 일반적 특성은 빈도와 백분율을 이용하였고, 자기인식, 타인인식과 의사소통능력의 차이는 Independent t-test, ANOVA를 이용하였다.

2.2 연구 도구

자기인식은 Fenigstein 등(1975)의 자기인식검사(Self-Awareness Tests)를 수정 보완하여 은혁기(2001)가 사용한 총 20문항의 도구를 사용하였다. 타인인식 측정도구는 辻(1994)가 개발한 것을 은

¹62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: netagnes@nambu.ac.kr

²62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: jjang9312@nambu.ac.kr

혁기(2001)가 우리나라 청소년을 대상으로 신뢰도와 타당도를 검증한 총 21문항의 도구를 사용하였다. 의사소통능력을 파악하기 위해 본 연구에서는 Navran(1967)이 부부의사소통 검사를 위해 개발한 PCI(Primary Communication Inventory)의 내용을 최은연(2004)이 수정한 총 25문항의 도구를 사용하였다.

3. 연구결과

본 연구 결과 대상자의 성별은 남자 31명(14.0%), 여자 191명(86.0%)이었고, 대상자의 학년은 1학년 70명(31.5%), 2학년 58명(26.1%), 3학년 66명(29.7%), 4학년은 28명(12.6%)이었다. 종교 여부는 종교가 없는 대상자가 134명(60.4%)으로 많았고, 종교가 있는 대상자는 87명(39.2%)이었다. 경제 상태는 189명(85.1%)이 '중'으로 가장 많았고, 17명(7.7%)이 '상', 16명(7.2%)이 '하' 순으로 나타났다. 의사소통 교육은 156명(70.3%)이 교과목을 수강한 경험이 있고 66명(29.7%)이 수강한 경험이 없다고 했으며, 임상실습경험이 없는 대상자는 123명(55.4%), 임상실습경험이 있는 대상자는 99명(44.6%)이었다. 대상자의 학업성취수준에서 142명(64.0%)이 '중', 62명(27.9%)은 '상', 18명(8.1%)은 '하' 순으로 응답하였다(Table 1).

Table 1. Differences of Self-awareness, Other-awareness and Communication Ability by General Characteristics

Characteristics	Categories	n	Self-awareness		Other-awareness		Communication Ability	
			M±SD	t or F(p) Scheffe'	M±SD	t or F(p) Scheffe'	M±SD	t or F(p) Scheffe'
Grade	1st	70	3.28±.28	1.29(.278)	3.28±.47	1.53(.206)	3.44±.38	2.59(.053)
	2nd	58	3.33±.27		3.43±.43		3.40±.37	
	3rd	66	3.37±.33		3.45±.59		3.54±.41	
	4rd	28	3.38±.25		3.32±.47		3.60±.38	
Gender	Male	31	3.22±.23	-2.42(.016)	3.27±.50	-1.27(.205)	3.23±.41	-3.99(<.001)
	Female	191	3.35±.29		3.40±.50		3.52±.37	
Religion	Yes	87	3.41±.30	3.06(.002)	3.44±.58	1.27(.230)	3.55±.41	2.21(.028)
	No	134	3.29±.28		3.35±.44		3.43±.37	
	High	17	3.42±.32	1.02(.362)	3.44±.43	.12(.882)	3.51±.40	.18(.830)
Economic state	Moderate	189	3.33±.29		3.38±.52		3.48±.40	
	Low	16	3.29±.18		3.38±.39		3.43±.30	
Experience in taking a communication course	Yes	156	3.37±.28	2.52(.012)	3.42±.52	1.94(.053)	3.51±.39	1.54(.124)
	No	66	3.26±.29		3.28±.44		3.42±.48	
Clinical practice experience	Yes	99	3.38±.30	2.19(.029)	3.41±.56	.67(.502)	3.55±.40	2.54(.012)
	No	123	3.30±.28		3.36±.44		3.42±.37	
School achievement	Well ^a	62	3.44±.27	5.69(.004) a>b	3.48±.49	2.31(.102)	3.54±.47	1.31(.270)
	Average ^b	142	3.30±.28		3.36±.51		3.41±.34	
	Poor ^c	18	3.26±.32		3.22±.43		3.40±.41	

Table 2. Correlation of Self-awareness, Oher-awareness and Communication Ability

	Self-awareness	Oher-awareness	Communication Ability
Self-awareness	1		
Oher-awareness	.32 (<.001)	1	
Communication Ability	.39 (<.001)	.26 (<.001)	1

본 연구결과 자기인식과 타인인식은 통계적으로 유의한 양의 상관관계($r=.32, p<.001$)로 나타났으며, 자기인식과 의사소통도 양의 상관관계($r=.39, p<.001$)가 있는 것으로 나타났다. 또한 타인인식과 의사소통도 양의 상관관계($r=.26, p<.001$)가 있는 것으로 나타났다(Table 2).

4. 논의 및 제언

본 연구에서 자기인식과 타인인식 및 의사소통능력은 서로 양의 상관관계가 있는 것으로 나타났다. 이러한 결과를 토대로 간호대학생의 의사소통능력을 증진시키기 위해서는 간호교육과정에서 자기인식과 타인인식을 증진할 수 있는 방법의 모색이 필요하며, 자기인식 훈련을 포함한 자기인식을 증진시킬 수 있는 연구가 이루어져야 한다고 사료된다.

References

- A. Fenigstein, M. F. Scheier, A. H. Buss. (1975). Public and private self-consciousness. *Journal of Consulting and Clinical Psychology*, 43(4), 522-527.
- H. K. Eun, (2001). Adolescents' Self-awareness, Other-awareness, Interpersonal skills and the Satisfaction of interpersonal relationship : the Difference among the Grades and between the Sexes, *The Journal of youth Counseling Research*, 21(32), 136-157.
- 辻平治郎. (1994). 自己意識と他者意識. 北大路書房.
- Navran (1967). Communication and adjustment in marriage [Internet]. c2016[cited 2022 Jun 9], Available From: <http://onlinelibrary.wiley.com/doi/10.1111/j.1545-5300.1967.00173.x/abstract>.
- E. Y. Choi. (2004). *(The) effects of a communication skills group on high school students' self-concept and communication ability*, [Master's thesis]. Kosin University, Pusan.

간호대학생의 감성지능수준, 스트레스, 스트레스 대처능력 간 관계

이은미¹, 오윤정²

1. 서론

1.1. 연구의 필요성

간호대학생들은 타 학과에 비해 간호사 국가고시에 합격해야 한다는 심리적 압박감, 많은 학습량, 주기적인 임상실습 및 교내실습 등으로 인해 많은 스트레스를 경험하게 된다(Jung, 2006). 그 과정에서 낮은 임상환경에 적응해야하는 것 뿐 아니라 여러 사람과의 대인관계 및 전문적인 지식과 기술의 부족으로 어려움을 겪는다. 이러한 스트레스의 예방과 처리의 과정을 잘 수행하지 못하면 여러 심리적 부담감을 경험하게 되며, 이는 곧 학습능력과 수행능력의 저하를 일으켜 자신감 하락과 좌절감을 가져온다(Han, Bae, 2015).

스트레스 대처방식이 부적절하였을 시 학업이나 생활에 쉽게 적응하지 못하고 학업 스트레스로 인한 소진 및 학교생활의 어려움을 겪게 된다(Lee, 2003). 스트레스에 직면한 경우 상황을 변화시키려는 적극적 대처를 사용하면 소진을 줄여주지만 소극적 대처를 사용하면 오히려 소진을 가중시킨다(Song, Chae, 2014). 따라서 간호대학생이 학업 및 실습으로 인한 스트레스를 직면했을 때 적극적인 대처능력은 소진정도에 주요한 영향요소라고 볼 수 있다.

감성지능은 인간이 가진 심리적 성향 중 긍정적인 감정 성향이고 감정을 평가하고 효과적으로 조절할 줄 아는 능력이며 정서를 활용할 줄 아는 능력이다(Wong, Law, 2002). 감성지능이 높은 사람은 스트레스 상황에도 긍정적으로 평가하고 적극적인 문제 해결을 위해 노력하는 등 대처방식을 보이지만 감성지능이 낮은 사람은 정서적 갈등으로 인해 스트레스 정도가 더 높게 나타난다(Song, Chae, 2014). 이에 본 연구는 간호대학생들의 감성지능수준과 스트레스 정도 및 대처능력에 대해 파악하고 이들 간의 관련성을 확인하여 간호대학생의 스트레스 감소를 위해 감성지능수준 향상을 돕는 프로그램 개발의 필요성과 근거를 제시하고자 한다.

2. 연구방법

2.1. 연구설계

본 연구는 간호대학생의 감성지능수준, 스트레스, 스트레스 대처능력 간의 관계를 확인하기 위한 서술적 조사연구이다.

¹62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: jjang9312@nambu.ac.kr

²62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: netagnes@nambu.ac.kr

2.2. 연구대상 및 자료수집방법

본 연구의 대상은 G광역시 소재 일 대학교 간호학과에 재학 중인 학생들 중 의사소통에 문제가 없으며 본 연구의 목적을 이해하고 연구 참여에 자발적으로 서면 동의한 자로 선정하였다. 자료 수집은 2021년 9월 27일부터 10월 13일까지 자가보고식 설문지를 통해 이루어졌다. 대상자 수는 G*power 3.1.2 프로그램을 이용하여 상관관계분석을 위한 중간정도의 효과크기 .15, 검정력 80%, 유의수준 .05를 기준으로 예상 표본수를 산출하였고, 총 응답자 200명 중 자료입력이 불충분한 응답자 6명을 제외한 194명의 자료를 최종자료 분석 대상으로 선정하였다.

2.3. 연구도구

감성지능은 Wong과 Law(2002)가 개발한 도구를 Jung(2006)이 번안하여 수정한 도구로 총 16문항으로 구성되어 있다. 스트레스 정도는 Cohen, Kamarck와 Mermelstein(1983)이 개발한 지각된 스트레스 척도를 Park과 Seo(2010)가 번안하여 수정한 한국판 지각된 스트레스 척도로 총 10문항으로 구성되어 있다. 스트레스 대처능력은 Folkman과 Lazarus(1985)가 개발한 후 수정한 스트레스 대처방식 척도를 Song과 Chae(2014)가 번안한 도구로 총 42문항으로 구성되어 있다.

3. 연구결과

3.1. 대상자의 일반적 특성에 따른 감성지능수준, 스트레스, 스트레스 대처능력

연구 대상자의 일반적 특성에 따른 감성지능은 전공만족도($t=5.60$, $p=.004$)와 건강상태($t=3.00$, $p=.049$)에 따라 차이가 있는 것으로 나타났다. 스트레스 정도는 성별($t=1.73$, $p=.049$)에 따라 유의한 차이를 보였고, 스트레스 대처능력은 전공만족도($t=3.40$, $p=.036$)에 따라 차이가 있었다(Table 1).

3.2. 감성지능 수준과 스트레스와 스트레스 대처능력 간 상관관계

간호대학생의 감성지능 수준과 스트레스정도($r=-.449$, $p<.001$), 스트레스 정도와 스트레스 대처능력은($r=-.232$, $p=.047$) 유의한 음의 상관관계를 보였고, 감성지능 수준과 스트레스 대처능력($r=.520$, $p<.001$)은 유의한 양의 상관관계가 있는 것으로 나타났다(Table 2).

3.3. 감성지능 수준에 따른 스트레스 및 스트레스 대처능력

대상자의 감성지능을 사분위수를 기준으로 상위25%(삼사분위수)인 49명과 하위25%(일사분위수)인 49명으로 나누어 스트레스와 스트레스 대처방식의 차이를 살펴본 결과, 감성지능 수준이 상위 25%인 스트레스 정도는 감성지능 하위 25%인 학생에 비해 낮게 나타났다($t=4.81$ $p<.001$) 스트레스 대처방식에 있어서는 문제중심과 정서중심 두 가지로 나뉘는데 상위 25%인 문제중심 스트레스 대처방식은 하위 25%인 학생에 비해 높게 나타났고($t=-6.41$ $p<.001$), 정서중심 스트레스 대처방식 또한 상위 25%는 하위 25% 보다 높게 나타났다($t=-5.49$ $p<.001$)(Table 3).

4. 결론 및 제언

본 연구는 간호대학생들의 감성지능 수준과 스트레스 및 대처능력에 대해 파악하고 이들 간의 관련성을 확인하여 간호대학생의 스트레스 감소를 위해 감성지능 수준 향상을 돕는 프로그램 개발의 기초자료를 제공하고자 시도되었다. 본 연구결과 간호대학생의 감성지능 수준, 스트레스 정도, 스트레스 대처능력 간에 유의한 상관관계를 보였고, 감성지능 수준에 따라 스트레스 정도와 대처능력 정도의 차이가 있는 것으로 나타났다. 따라서 향후 간호대학생들의 감성지능을 높여 스트레스를 줄이거나 해결해나갈 수 있는 중재프로그램개발 및 그 효과를 검증하는 연구가 필요할 것으로 사료된다.

Table 1. Differences of Emotional Intelligence, Stress Level, Stress-Coping Ability by General Characteristics (N=194)

Variables	Categories	n(%) or M(SD)	Emotional intelligence		Stress level		stress-coping ability	
			M(SD)	t or F(p) Scheffe	M(SD)	t or F(p) Scheffe	M(SD)	t or F(p) Scheffe
Age(yr)		21.58(4.12)						
Gender	Female	157(80.5)	3.67(0.55)	-1.30 (.199)	2.12(0.55)	1.73 (.049)	3.43(0.34)	-1.56 (.875)
	Male	37(19.5)	3.54(0.46)		2.30(0.58)		3.34(0.34)	
Grade	1	72(36.4)	3.51(0.48)	0.86 (.462)	2.30(0.55)	0.15 (.933)	3.32(0.33)	1.21 (.309)
	2	64(32.8)	3.64(0.54)		2.25(0.60)		3.41(0.37)	
	3	58(29.8)	3.56(0.41)		2.25(0.58)		3.35(0.31)	
Religion	Yes	64(32.8)	3.52(0.45)	-.93 (.584)	2.25(0.54)	-.31 (.267)	3.35(0.35)	-.15 (.567)
	No	130(67.2)	3.59(0.49)		2.28(0.59)		3.36(0.33)	
Reason to choose the nursing	Aptitudes and interest	60(30.8)	3.66(0.51)	1.88 (.134)	2.21(0.60)	0.91 (.435)	3.41(0.36)	1.32 (.269)
	Parent or teacher's recommendation	61(31.4)	3.58(0.47)		2.22(0.54)		3.34(0.32)	
	High employment rate	62(31.8)	3.49(0.45)		2.37(0.57)		3.35(0.30)	
	Etc	11(5.6)	3.38(0.51)		2.29(0.65)		3.21(0.44)	
Satisfaction of major	Satisfaction ^a	42(21.5)	3.77(0.38)	5.60 (.004) b>a	2.11(0.54)	2.61 (.076)	3.47(0.33)	3.40 (.036) b,c<a
	Usually ^b	140(71.8)	3.52(0.48)		2.30(0.57)		3.32(0.32)	
	Dissatisfaction ^c	12(6.2)	3.39(0.59)		2.47(0.79)		3.34(0.44)	
Health status	Good ^a	102(52.3)	3.64(0.45)	3.00 (.049) b<a,c	2.18(0.56)	2.66 (.072)	3.36(0.32)	1.21 (.309)
	Usually ^b	83(42.6)	3.47(0.48)		2.37(0.57)		3.34(0.36)	
	Bad ^c	9(4.6)	3.65(0.66)		2.30(0.69)		3.51(0.29)	

Table 2. Relationship Emotional Intelligence, Stress Level, Stress-Coping Ability (N=194)

Variable	Emotional intelligence r(p)	Stress level r(p)	Stress-coping ability r(p)
Emotional intelligence	1		
Stress level	-.449(<.001)	1	
Stress-coping ability	.520(<.001)	-.232(.047)	1

Table 3. Stress Level, Stress-Coping Ability by Emotional Intelligence Level (N=98)

Variable	하위 25% 그룹 (n=49)	상위 25% 그룹 (n=49)	t(p)
	Mean(SD)	Mean(SD)	
Stress level	2.51(0.69)	1.73(0.44)	4.81(<.001)
Stress-coping ability(problem focus)	3.06(0.38)	3.66(0.30)	-6.41(<.001)
Stress-coping ability(emotional focus)	3.01(0.36)	3.61(0.37)	-5.49(<.001)

Reference

- Wong, C. S., Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly*, 13, 243-274.
- Jung, H. Y. (2006). Effectiveness of emotional leadership and emotional intelligence of food service employees on organizational performance in contracted food service management company. *Unpublished doctoral dissertation*, Yonsei University, Seoul, Korea.
- Cohen, S., Kamarck, T., Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396
- Lee, H. K. (2003). A study on stress and coping method of nursing students. *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, 12(4), 586-593.
- Folkman, S., Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, 48, 150-170
- Song, J. E., Chae, H. J. (2014). Differences in stress and stress coping strategies according to emotional intelligence in nursing students. *Journal Korean Academic Soc Nurs Educ*, 20(4), 502-512.
- Han, M. H., Bae, S. M. (2015). A study on the relationship between college adaptation, academic achievement, and admission type in one nursing School. *The Journal of Korean Academic Society of Nursing Education*, 21(1), 46-53.

간호대학생의 학년별 학업소진과 스트레스 대처능력 및 자기효능감 비교

이은미¹, 오윤정²

1. 서론

1.1. 연구의 필요성

간호대학생은 4년의 교육과정 내에서 이론과 임상실습을 병행하며 짧은 시간 내에 정해진 지식을 습득해야 하는 어려움으로 인해 더 많은 스트레스를 경험하게 된다(Cho, Lee, 2016). 간호대학생은 국가고시를 위한 학업과 임상실습에서 오는 높은 스트레스 수준으로 타 학과 학생들에 비해 소진정도가 높다(Lim, 2017). 간호대학생의 이러한 소진은 전공만족도에 영향을 미치게 되며 학년이 올라갈수록 지속적으로 증가하여 졸업 후 간호사로서의 역할 적응 및 긍정적 변화를 방해하게 된다(Lee, Park, Choi, 2017). 대처란 개인의 내적, 외적 요구에 의해 발생하는 스트레스를 극복하기 위하여 끊임없이 변화시키는 과정으로 시간의 경과나 개인의 경험 및 상황과 상호작용하면서 변한다(Park, Sun, 2010). 따라서 과거와 동일한 상황이 발생하여도 대처능력이나 방법은 다를 수 있고 상황에 적합한 대처가 효율적으로 잘 이루어질수록 다양한 소진을 잘 극복해나갈 수 있다(Jeong, Koh, 2016). 간호대학생의 자기효능감은 긍정적인 감정조절과 대처능력에 영향을 주며, 개인의 심기 내적 요소에 따라 과중한 학업요구 상황에서 발생하는 학업소진 정도가 달라진다(Whang, 2021). 이에 본 연구는 간호대학생들의 학년 별 학업소진, 스트레스 대처능력, 자기효능감 정도 및 관련성을 파악하여 간호대학생의 학업 스트레스로 인한 소진현상을 줄이기 위한 방안 마련의 기초자료로 활용하고자 한다.

2. 연구방법

2.1. 연구대상 및 자료수집방법

본 연구대상자는 G광역시 소재한 대학의 간호학과에 재학 중인 학생을 대상으로 편의추출하였으며, 본 연구의 목적을 이해하고 연구 참여에 자발적으로 서면 동의한 자료 선정하였다. 자료 수집은 2021년 11월 30일부터 12월 15일까지 자가보고식 설문지를 통해 이루어졌다. 대상자 수는 G*power 3.1.2 프로그램을 이용하여 상관관계분석을 위한 중간정도의 효과크기 .15, 검정력 80%, 유의수준 .05를 기준으로 예상 표본 수를 산출하였고, 총 응답자 210명 중 자료입력이 불충분한 응답자 7명을 제외한 203명의 자료를 최종자료 분석 대상으로 선정하였다.

¹62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: jjang9312@nambu.ac.kr

²62271 광주광역시 광산구 남부대길 1, 남부대학교 간호학과. E-mail: netagnes@nambu.ac.kr

2.2. 자료분석방법

수집된 자료는 SPSS 26.0 프로그램을 이용하였으며 간호대학생의 일반적 특성은 빈도와 백분율로 분석하고 학년별 학업소진, 스트레스 대처능력, 자기효능감 비교는 t-test와 ANOVA를 사용하였다. 학업소진, 스트레스 대처능력, 자기효능감의 상관관계는 Pearson correlation을 이용하였다.

3. 연구결과

3.1. 대상자의 일반적 특성

연구 대상자의 일반적 특성 중 학년에 따라 나이만 유의한 차이가 있었고($F=19.20, p<.001$), 성별, 건강상태, 학업적응 수준에서는 차이가 없었다(Table 1).

3.2. 학년 별 학업소진, 스트레스 대처능력, 자기효능감 차이

간호대학생의 학년에 따라 학업소진($F=7.68, p<.001$)과 스트레스 대처능력($F=1.45, p=.043$)은 유의한 차이를 보였고, 자기효능감은($F=1.45, p=.228$)은 유의한 차이가 없었다(Table 2).

3.3. 학업소진, 스트레스 대처능력, 자기효능감 간 관계

간호대학생의 ‘학업소진’과 ‘자기효능감’($r=-.413, p=.023$), ‘학업소진’과 ‘대처능력’($r=-.584, p<.001$)은 유의한 음의 상관관계를 보였고 ‘대처능력’과 ‘자기효능감’은 유의한 양의 상관관계($r=.768, p<.001$)를 보였다(Table 3).

4. 결론 및 제언

본 연구는 간호대학생의 학년별 학업소진과 스트레스 대처능력, 자기효능감 정도를 파악하고 학업소진 감소와 대학생활적응 향상을 위한 방안을 마련하고자 시도되었다. 본 연구결과 간호대학생의 학업소진, 스트레스 대처능력, 자기효능감 간에 유의한 상관관계를 보였고, 학년에 따라 학업소진과 대처능력에 차이가 있는 것으로 나타났다. 따라서 향후 간호대학생들의 학업 증진과 스트레스 관리, 학교생활 적응 향상을 위한 중재프로그램개발 및 그 효과를 검증하는 연구가 필요할 것으로 사료된다.

Table 1. General Characteristics of Participants by Grade (N=203)

Characteristics	Categories	Total	Freshmen	Sophomores	Juniors	Seniors	χ^2 or F(p)
		n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	
Age(yr)		21.19±2.58	19.60±2.12	20.56±2.45	21.95±2.73	22.77±2.20	19.20 ($<.001$)
Gender	Female	171(83.2)	42(84.0)	45(83.3)	40(85.1)	44(84.6)	4.55 (.302)
	Male	32(16.8)	8(16.0)	9(16.7)	7(14.9)	8(15.4)	
Health status	Very good	16(36.4)	5(10.6)	4(7.1)	3(6.2)	4(7.8)	6.79 (.395)
	Good	162(34.4)	37(78.8)	45(80.6)	40(83.3)	40(76.9)	

Table 1. General Characteristics of Participants by Grade (N=203)(continued)

Characteristics	Categories	Total	Freshmen	Sophomores	Juniors	Seniors	x2 or F(p)
		n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	n(%) or M(SD)	
Health status	Bad	22(14.6)	5(10.6)	6(10.7)	4(8.3)	7(13.4)	6.79 (.395)
	Very bad	3(13.6)	-	1(1.6)	1(0.2)	1(1.9)	
Academic adjustment	Well adjusted	20(9.8)	6(11.3)	3(6.1)	5(10.6)	6(11.1)	9.75 (.098)
	Adjusted	147(72.4)	38(71.6)	34(69.3)	35(74.4)	40(74.0)	
	Unadjusted	26(12.9)	7(13.2)	8(16.3)	5(10.6)	6(11.1)	
	Unadjusted at all	10(4.9)	2(3.9)	4(8.3)	2(5.4)	2(3.8)	

Table 2. Comparison of Academic Burnout, Self-Efficacy, Stress-Coping Ability by Grade (N=203)

Variables	Total	Freshmen (n=50)	Sophomores (n=54)	Juniors (n=47)	Seniors (n=52)	F	p
	M±SD	M±SD	M±SD	M±SD	M±SD		
Academic burnout	3.25±0.32	3.11±0.43	3.05±0.50	3.30±0.65	3.49±0.50	7.68	<.001
self-efficacy	3.24±0.20	3.22±0.22	3.20±0.29	3.19±0.45	3.33±0.32	1.45	.228
stress-coping ability	3.31±0.34	3.29±0.33	3.28±0.31	3.34±0.44	3.48±0.32	2.76	.043

Table 3. Relationship Academic Burnout, Stress-Coping Ability, Self-efficacy (N=203)

Variable	Burn out r(p)	Stress-coping ability r(p)	self-efficacy r(p)
Burn out	1		
Stress-coping ability	-.584(<.001)	1	
self-efficacy	.768(<.001)	-.413(.023)	1

Reference

- Cho, H., Lee, K. Y. (2016). Self esteem and career identity of nursing students, *Journal of Korea Academy of Public Health Nursing*, 28(2), 136-173.
- Lim, Y. J. (2017). The impact of nursing students' emotional labor and clinical practice stress on burnout, *Journal of Learner-Centered Curriculum & Instruction*, 17(6), 461-474. <http://dx.doi.org/10.22251/jlcci.2017.17.6.461>
- Lee, D. Y., Park, J. K., Choi, M. Y. (2017). The relation between stress of clinical practice and burnout among nursing students: The mediation effect off spiritual well-being, *The Journal of Korean Academic Society of Nursing Education*, 23(3), 300-308. <http://dx.doi.org/10.5977/jkasne.2017.23.3.300>
- Park, H. J., Sun, S. I. (2010). Stress, depression, coping styles and satisfaction of clinical practice in nursing students, *Journal of Korean Academic Society of Nursing Education*, 16(1), 14-23, DOI: <http://doi.org/10.5977/JKASNE.2010.15.1.01>
- Jeong, Y. J., Koh, C. K. (2016). Effects of personality and coping behavior on clinical practice stress among one college nursing students, *The Korean Journal of Stress Research*, 24(4), 1668-2234, DOI: <https://doi.org/10.17547/kjsr.2016.24.4.296>
- Whang, S. A. (2021). The Influence of Major Satisfaction, Academic Self-efficacy, and Clinical Practice Stress on College Student Burnout among Nursing Students, *Journal of Korean Society for Simulation in Nursing*, 9(2), 73-85.

중소병원 간호사의 감성지능과 동료 간 돌봄-배려행위가 조직몰입에 미치는 영향

실진욱¹, 박명숙²

요 약

본 연구는 중소병원 간호사의 감성지능과 동료 간 돌봄-배려행위가 조직몰입에 미치는 영향을 파악하는 것이다. 연구대상은 G 광역시 소재 5개 중소병원에서 정규직으로 1년 이상 재직 중인 간호사를 대상으로 편의 표집한 217명이다. 수집한 자료는 SPSS/WIN 26.0을 이용하여 분석하였다. 연구결과 감성지능은 동료 간 돌봄 배려행위와 조직몰입과 양의 상관관계가 있는 것으로 나타났다. 조직몰입은 감성지능($\beta=.24, p<.001$)과 근무만족도 중 '불만족'($\beta=-.38, p<.001$)이 영향을 미치는 변수로 확인되었고, 설명력은 23.7%로 나타났다. 이러한 결과는 간호사의 조직몰입을 높이기 위해 감성지능과 근무만족도를 높이기 위한 노력도 중요하지만, 근무와 관련된 불만족 부분을 확인하고 개선하는 것 또한 중요함을 보여준다. 따라서 이를 해결하기 위해 효율적인 간호 전략이 필요하며, 간호사의 감성지능을 높일 수 있는 맞춤형 프로그램 개발과 효과를 검증할 수 있는 연구가 필요하다 하겠다.

주요용어 : 간호사, 감성지능, 돌봄-배려행위, 조직몰입

1. 연구 필요성

임상현장에서 간호 인력은 병원을 평가하는 기준이며, 조직 운영 측면에서 경제적 효율성을 증대시키는 데 중요한 자원이다. 그러나 간호사의 이직 의도는 열악한 근무환경과 더불어 과중한 업무, 직무스트레스, 심각한 감정노동 및 폭력 경험 등으로 인해 날로 증가하고 있어(Kim, Cho, 2020; Park, Park, 2020) 간호 인력 문제가 심각해지고 있다. 특히, 코로나바이러스19를 경험하고 있는 현 시점에서 간호사 인력 확보는 무엇보다도 중요한 사회적 이슈로 대두되고 있어 간호사의 인력관리가 무엇보다도 중요한 시점이다. 동료와의 좋은 관계를 유지하는 데 필요한 동료 간 돌봄-배려행위와 감성지능이 높은 경우 직무스트레스와 이직의도를 낮추고, 간호업무성과, 임상 수행능력 및 조직의 몰입도를 높이는 것으로 나타났다(Park, Park, 2020; We, Kim, Noh, 2019). 따라서 본 연구는 중소병원 간호사의 감성 지능과 동료 간 돌봄-배려행위가 조직몰입에 미치는 영향을 파악하여 향후 중소병원 간호사의 조직몰입을 증진함과 동시에 간호사가 의료현장을 떠나는 것을 예방할 수 있는 다양한 교육프로그램 개발에 필요한 기초자료를 제공하고자 시도하였다.

¹61266 대한민국 광주광역시 북구 북문대로 191, 운암한국병원 수간호사. E-mail: d11wook@naver.com

²62271 (교신저자) 대한민국 광주광역시 광산구 남부대길1(월계동), 남부대학교 간호학과 부교수.

E-mail: mspark@nambu.ac.kr

2. 연구방법

본 연구는 중소병원 간호사의 감성지능, 동료 간 돌봄-배려 및 조직몰입 정도를 파악하고, 조직몰입에 영향을 미치는 요인을 확인하기 위한 서술적 조사연구이다. 본 연구 대상자는 G 광역시에 소재한 5개의 중소병원을 편의 표출하여 정규직으로 1년 이상 재직 중인 간호사 217명을 대상으로 하였다. 대상자 수 선정은 G power 3.1.9.2 프로그램을 이용하여 효과크기 .15, 유의수준 .05, 검정력 .95로 표본 크기를 산출한 결과 최소 184명이었다. 탈락률 10~20%를 고려하여 217명을 대상으로 설문지를 배부하였고, 217부를 회수하였다. 이 중 불충분한 자료 15부를 제외하고 총 202부를 최종 분석에 사용하였다. 수집된 자료는 SPSS/WIN 26.0 통계프로그램을 이용하여 빈도 및 기술통계, ANOVA, t-test, Pearson Correlation Coefficient, Step wise Multiple Regression Analysis, 사후검증은 Scheffé test로 분석하였다.

3. 연구결과

본 연구결과 일반적 특성에 따른 감성지능은 근무부서(F=2.72, p=.046), 근무형태(F=2.65, p=.009), 동료 간 돌봄-배려행위는 결혼상태(t=2.05, p=.042), 근무부서(F=4.16, p=.007), 근무경력(F=5.33, p=.001), 연봉(F=7.23, p<.001)에서 통계적으로 유의한 차이가 있었다. 조직몰입은 근무부서(F=3.38, p=.019), 근무경력(F=2.83, p=.040), 근무형태(t=2.25, p=.025), 근무만족도(F=32.86, p<.001)에서 통계적으로 유의한 차이를 보였다(table 1). 상관관계 분석 결과 감성지능은 동료 간 돌봄-배려행위(r=.24, p<.001)와 조직몰입(r=.25, p<.001)과 약한 양의 상관관계가 있는 것으로 나타났다(table 2). 조직몰입에 대한 회귀분석결과 회귀모형은 통계적으로 유의하게 나타났으며(F=21.79, p<.001), 조직몰입에 영향을 미치는 변수로 근무만족도_불만족(β =.38)와 감성 지능(β =.24), 근무만족도_만족(β =.15)이 유의한 변수로 나타났다(p<.001). 전체 설명력은 23.7%였다(table 3).

Table 1. Differences of Emotional intelligence, Peer group caring interaction behaviors, Organizational commitment by General Characteristics

Characteristics	Categories	n(%) or M(SD)	Emotional intelligence		Peer group caring interaction behaviors		Organizational commitment	
			M(SD)	t or F(p) Scheffé	M(SD)	t or F(p) Scheffé	M(SD)	t or F(p) Scheffé
Gender	Male	7(3.5)	3.41(0.55)	† 657.00 (.867)	4.61±0.73	† 558.00 (.412)	2.67±0.43	† 387.00 (.076)
	Female	195(96.5)	3.50(0.46)		4.35±0.53		3.00±0.58	
Marital status	Single	134(66.3)	3.50(0.45)	.27 (.786)	4.41±0.51	2.05 (.042) b<a	2.96±0.54	-1.09 (.278)
	Married	68(33.7)	3.48(0.50)		4.25±.057		3.05±0.63	
Religion	Yes	71(35.1)	3.54(0.53)	.89 (.377)	4.41±0.55	.95 (.344)	3.02±0.54	.53 (.599)
	No	131(64.9)	3.47(0.43)		4.33±0.53		2.98±0.60	
Work department	Medical warda	39(19.3)	3.37(0.42)	2.72 (.046) a,c<b	4.30±0.51	4.16 (.007) b,a,c<d	2.85±0.55	3.38 (.019) a<d
	Surgical wardb	99(49.0)	3.57(0.47)		4.29±0.56		2.94±0.60	
	Intensive care or Emergency unitc	30(14.9)	3.37(0.50)		4.31±0.43		3.04±0.59	
	etcd	34(16.8)	3.54(0.44)		4.65±0.51		3.25±0.44	

Table 1. Differences of Emotional intelligence, Peer group caring interaction behaviors, Organizational commitment by General Characteristics(continue)

Characteristics	Categories	n(%) or M(SD)	Emotional intelligence		Peer group caring interaction behaviors		Organizational commitment	
			M(SD)	t or F(p) Scheffé	M(SD)	t or F(p) Scheffé	M(SD)	t or F(p) Scheffé
Turnover experience	Yes	134(66.3)	3.49(0.48)	-.46 (.649)	4.31±0.56	-1.77 (.079)	2.95±0.57	-1.52 (.130)
	No	68(33.7)	3.52(0.45)		4.45±0.49		3.08±0.59	
Work career (yr)	<3a	75(37.1)	3.50(0.50)	1.36 (.255)	4.53±0.50	5.33 (.001) c<a	3.01±0.58	2.83 (.040) b<d
	3<7b	47(23.3)	3.50(0.44)		4.26±0.55		2.87±0.60	
	7<10c	31(15.3)	3.35(0.47)		4.12±0.53		2.85±0.48	
	≥10d	49(24.3)	3.57(0.43)		4.34±0.52		3.16±0.57	
Work type	Non shift work	56(27.7)	3.63(0.50)	2.65 (.009) b<a	4.48±0.52	1.97 (.050)	3.14±0.63	2.25 (.025) b<a
	Shift work	146(72.3)	3.44(0.44)		4.31±0.54		2.93±0.54	
Salary	<3400a	58(28.7)	3.43(0.49)	.72 (.544)	4.10±0.50	7.23 (<.001) a<d,c,b	2.89±0.52	1.83 (.143)
	3400<3700b	71(35.2)	3.51(0.50)		4.50±0.53		3.04±0.63	
	3700<4000c	42(20.8)	3.50(0.38)		4.43±0.53		2.92±0.54	
	≥4000d	31(15.3)	3.58(0.45)		4.42±0.49		3.16±0.58	
Work satisfaction	Satisfactiona	38(18.8)	3.58(0.57)	† .88 (.645)	4.52±0.60	† 4.51 (.105)	3.26±0.52	† 32.86 (<.001) c<b,a
	Usualb	144(71.3)	3.47(0.44)		4.33±0.49		3.02±.050	
	Dissatisfactionc	20(9.9)	3.51(0.42)		4.27±0.72		2.29±0.63	

† : kruskal wallis test statistics(n>3), mann-whitney test statistics(n=2).

Table 2. Correlation of Emotional intelligence, Peer group caring interaction behaviors, Organizational commitment

	Emotional intelligence	Peer group caring interaction behaviors	Organizational commitment
Emotional intelligence	1		
Peer group caring interaction behaviors	.24 (p<.001)	1	
Organizational commitment	.25 (p<.001)	.14 (p=.041)	1

Table 3. Factors Influencing Organizational commitment

Variable	B	S.E	β	t	p	AdjR ²	F	p
Constant	2.00	.27		7.44	p<.001	.237	21.79	<.001
Work satisfaction_dissatisfaction	-.74	.12	-.38	-6.16	p<.001			
Emotional intelligence	.29	.08	.24	3.84	p<.001			
Work satisfaction_satisfaction	.22	.09	.15	2.34	p<.001			

4. 논의 및 제언

본 연구결과 감성지능은 동료 간 돌봄-배려행위 및 조직몰입과 양의 상관관계가 있는 것으로 나타났다, 조직몰입에 영향을 미치는 요인으로 감성지능과 근무만족도가 영향을 미치는 요인으로 나타났다. 특히, 근무만족도 중에서 ‘불만족’이 조직몰입에 영향을 미치는 요인으로 나타났다. 이러한 결과는 간호사의 조직몰입을 강화하기 위해서 간호사의 감성지능을 높일 수 있는 다양한 프로그램 개발과 적용이 필요함을 시사하며, 간호사의 근무만족도를 높이기 위한 기관의 적극적인 행-

재정적인 지원과 더불어 간호사의 불만족을 낮추기 위한 적극적인 인력관리가 필요함을 시사하고 있다. 또한, 간호사의 조직몰입을 높이고 이직률을 낮추기 위해서 기관별, 지역별 특성을 고려한 다양한 프로그램 개발과 후속 연구가 진행되어야 할 필요가 있다.

References

- Kim, S. Y., Cho, H. K. (2020). The Influencing factors on the General Hospital Nurses' Turnover Intention. *Journal of the Korea Convergence Society*, 11(9), 417-424. DOI : 10.15207/JKCS.2020.11.9.417.
- Park, E. H., Park, H. G. (2020). An influence of emotional labor, emotional intelligence, and social support of emergency room nurse on turnover intention. *Korean Academy on Communication in Healthcare*, 15(1), 53-60. DOI : 10.15715/kjhcom.2020.15.1.53.
- We, H. O., Kim, W. G., Noh, E. K. (2019). The Effects of Nursing Practice Environment and Peer Group Caring Interaction on Retention Intention among Small and Medium-sized Hospital Nurses. *Journal of Convergence for Information Technology*, 9(8), 83-92. DOI : <https://doi.org/10.22156/CS4SMB.2019.9.8.083>.

코로나 19로 인한 일상생활변화가 한국 청소년의 범불안장애에 미치는 영향

김은경¹

요 약

본 연구는 코로나19 관련 일상생활 변화 정도를 파악하고 범불안장애에 미치는 영향을 분석하기 위해 시행된 이차자료 분석설계연구로서 제16차 2021년도 청소년건강행태온라인조사의 원시자료를 이용하였다. 자료 분석은 IBM SPSS Statistics 21.0 프로그램을 이용하였고 대상자 특성과 코로나19 관련 일상생활 변화에 따른 범불안장애 유무는 independent t-test와 카이제곱 통계량을, 도구의 신뢰도를 검증하기 위하여 Cronbach's α 계수를 산출하였다. 코로나19 관련 일상변화에 따른 범불안장애 영향요인을 확인하기 위하여 다중 로지스틱 회귀분석을 실시하였다. 연구결과 범불안장애의 위험도는 나이가 많아질수록 1.04배(95%CI, 1.02-1.07), 성별이 여자인 경우 1.06배(95%CI, 1.01-1.12), 코로나 19로 가정경제변화가 있는 경우 1.14배(95%CI, 1.08-1.20), 스트레스가 많을수록 2.64배(95%CI, 2.56-2.73), 수면불충분인 경우 1.37배(95%CI, 1.30-1.44), 자살생각이 있는 경우 2.70배(95%CI, 2.48-2.93), 코로나 19로 인한 일상생활변화에서 아침결식 증가 1.15배(95%CI, 1.07-1.23), 감소 1.10배(95%CI, 1.02-1.19), 음주 증가 1.31배(95%CI, 1.09-1.56), 감소 1.41배(95%CI, 1.17-1.70), 우울 증가 2.77배(95%CI, 2.62-2.93) 통계적으로 유의하게 높았다. 고등학교 0.80배(95%CI, 0.73-0.88), 학업성적점수가 높을수록 0.93배(95%CI, 0.91-0.96), 주관적 건강점수가 높을수록 0.74배(95%CI, 0.72-0.77), 코로나 19로 인한 우울 감소 시 0.87배(95%CI, 0.78-0.97) 범불안장애 위험도가 통계적으로 유의하게 낮았다. 본 연구결과를 바탕으로 청소년의 범불안장애에 영향을 미치는 코로나 19 관련 일상변화에 대한 사회적 지원 및 관련 정책의 개선이 필요하며, 추후 지속적인 신종감염병 상황속에서 청소년의 특성을 고려한 맞춤형 매뉴얼을 개발하여 범불안장애를 감소시킬 수 있는 전략과 지원방안을 마련하는데 기여할 수 있을 것이다.

주요용어 : 코로나 19, 범불안장애, 청소년.

만성통증 노인환자의 주관적 건강상태에 따른 건강행위간 네트워크 분석

신지은¹, 백승희², 박세연³

1. 서론

의학의 비약적인 발전으로 질병예방 및 조기발견이 가능해졌고, 이로 인한 평균 기대수명이 높아지면서 세계보건기구에서는 2050년 전 세계 60세 이상 노인 인구가 약 21억에 이를 것으로 예상하였다(WHO, 2021). 이러한 급격한 노인인구 증가는 노화와 관련된 질병 증가를 의미하며, 이들의 여러가지 건강문제가 대두되고 있다(Jana & Chattopadhyaya). 특히 미국 질병통제예방센터(Centers for Disease control and Prevention: CDC)에서는 미국 성인의 20.4%인 5천만명 가량이 만성통증을 가지고 있으며 그중 8%는 통증으로 일상생활을 하지 못한다고 보고하였다. 따라서 만성통증을 예방하거나 관리하는 것은 노인건강과 삶의 질에 중요한 문제라 할 수 있다.

자가평가된 건강은 주관적이고 단순히 보일 수 있지만, 연령, 사회 인구학적 특성 및 소득수준보다 사망률 또는 삶의 질과 같은 건강 결과를 예측하는 강력한 예측인자로 입증되어져(Linn and Linn, 1980, Mossey & Shapiro, 1982, Idler, 1993, Dowd & Zajacova, 2007, Atherton et al., 2012) 개인의 전반적 건강(overall health)을 확인할 수 있는 좋은 지표라 할 수 있다(Li et al., 2006; Lundberg & Manderbacka, 1996; OCAMPO, 2010).

이처럼 만성통증과 주관적 건강인지 모두 건강 결과를 예측할 수 있는 요인이며, 몇몇 연구에서는 두 요인이 부정적인 관계가 보고되었다(Reyes-Gibby et al., 2002, Mäntyselkä et al., 2003, Siedlecki, 2006). 그러나 이 연구들은 단순한 두 요인의 관계입증에 그치고 있기 때문에 본 연구에서는 만성통증을 가진 노인을 대상으로 주관적 건강인지와 건강행위의 관계를 살펴보고, 이를 통해 만성통증 노인의 건강증진을 위한 기초자료를 제공하고자 한다.

2. 연구방법

2.1 연구대상

본 연구는 국민건강영양조사 중 5기(2010-2012년)와 6기(2013년) 33,552명의 참가자 중 만 64세 이상 6117명 중에서 다음에 해당하는 경우 연구에서 제외하였다; 1) 만성통증이 조사되지 않은 자(n=605); 2) 주관적 건강상태를 응답하지 않았거나 “보통”이라고 응답한 대상자(n=2,902); 3) 500kcal 이하 혹은 5000kcal 이상 섭취한 대상자(n=435); 4) 질환으로 식이요법을 진행하는 대상자(n=797);

¹35365 대전광역시 서구 관저동로 158, 건양대학교 의과대학 조교수. E-mail: jeshin@konyang.ac.kr

²02844 서울특별시 성북구 보문로 34다길, 성신여자대학교 생애복지대학원 건강운동관리학과 조교수.

E-mail: sh100@sunghsin.ac.kr

³(교신저자) 35015 대전광역시 중구 문화로 266, 충남대학교 간호대학 조교수. E-mail : park_sy@cnu.ac.kr

5) 암인 대상자(n=392); 6) 질환으로 활동제한이 있는 대상자(n=1229) 6) 건강행위가 무응답인 대상자(n=679). 따라서 1,689명을 최종 대상으로 선정하였으며, 이들 중 만성통증 여부에 따라 구분하여 분석하였다. 이후 이들의 주관적 건강인식에 따라 ‘좋음’과 ‘나쁨’으로 교차분류하여 4가지 그룹, 주관적 건강인식 좋음/통증있음(GPY), 주관적 건강인식 나쁨/ 통증있음(BPY), 주관적 건강인식 좋음/통증없음(GPN), 주관적 건강인식 나쁨/ 통증없음(BPN)으로 재분류하였다. 따라서 최종적으로 GPY 196명, BPY 404명, GPN 759명, BPN 330명이었다.

2.2 자료분석방법

통계분석은 SAS ver. 9.4(SAS Institute Inc, Cary, NC, USA)를 이용하였으며, 질병관리본부의 국민건강영양조사 원시자료 이용지침에 따라 가중치를 사용하여 복합표본분석을 시행하였다. 대상자들 주관적 건강상태(좋음/나쁨) 및 만성통증여부(Yes/No)를 교차하여 4개군으로 구분하여 군간 대상자들의 일반적 특성을 교차분석으로 분석하였다. 이어서 군간 건강행위 실천정도를 비교하기 위하여 교차분석을 실시하였다. 마지막으로 4개군의 건강행위간의 관련성을 소셜네트워크분석으로 시각화하였다.

2.3 측정도구

1) 주관적 건강상태

본 연구에서 주관적 건강상태는 건강설문조사 중 ‘평소에 ○○○님의 건강상태는 어떻다고 생각하십니까?’로 질문하여 ‘매우좋음’, ‘좋음’, ‘보통’, ‘나쁨’, ‘매우나쁨’의 5점 리커트 척도로 측정하였으며 “보통”을 제외하고 “좋음”과 “나쁨”으로 분류하였다.

2) 만성통증

“최근 3개월동안 30일 이상 무릎관절통(무릎강직/엉덩이관절통/요통)이 있었습니까?”로 질문하여 “있다”, “없다”로 측정하였으며 4가지 중 하나 이상에서 통증이 있는 경우 만성통증이 있음으로 정의하였다.

3) 건강관련행위

“Alameda 7”의 7가지 건강관련행위 중 간식섭취 항목을 제외한 비흡연, 음주, 적정수면, 적정체중유지, 운동, 아침식사, 외식으로 선정하였다. 비흡연은 평생 피우적 없거나 과거엔 피웠으나 현재 피우지 않은 경우, 음주는 평생 비음주이거나 최근 1년 이내 주 2회 미만으로 음주한 경우, 적정체중은 아시아인을 위한 비만기준(World Health Organization for Asian Population : WHO)을 근거로 BMI 18.5 kg/m² 이상 25 kg/m² 미만으로 하였다. 적정수면은 하루 7시간 이상~8시간인 경우, 운동은 주당 150분이상 운동하는 경우, 아침식사는 주 5-7회 섭취하는 경우, 외식은 주 4회 이하로 외식하는 경우를 건강행위로 정의하였다.

3. 결과

3.1 만성통증과 주관적건강인식

남자노인의 경우 GPN이 60.9%로 가장 높았으며, 이어서, BPN 21.6%, BPY 9.5%, GPY 8.1% 순으로 나타났다. 여자의 경우 GPN이 43.7%로 가장 높았으며, 이어서 BPY 24.1%, BPN 21.1%, GPY 11.1%순으로 나타났다.

3.2 만성통증과 주관적건강인식에 따른 건강행위

만성통증이 있는 남녀노인의 건강행위 실천 패턴을 네트워크 그래프로 분석한 결과는 남녀노인 모두 비흡주, 아침식사 및 비외식의 동시발생 빈도 및 중심성 지수가 높은 것으로 나타났으며, 비흡연과는 동시발생 빈도가 높았다. 한편, 남자노인에서 정상체중이나 수면시간의 동시발생 빈도가 높게 나타났다.

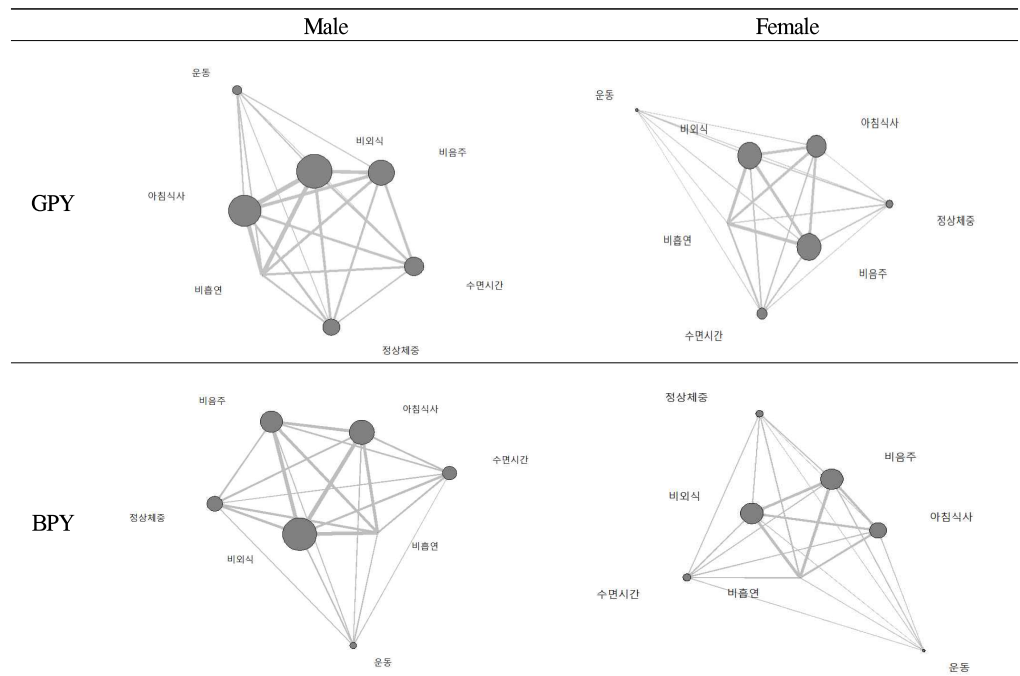


Figure 1. 성별, 주관적 건강상태에 따른 건강행위 네트워크 분석

성별 및 주관적 건강상태에 따른 건강행위 실천 동시 발생 패턴은 발생빈도에서 차이를 보였다. 노인 남자는 비흡연, 아침식사에서 통계적으로 유의한 차이가 있었다. 비흡연 건강행위 비율은 GPY 85.8%, BPY 67.6%로 주관적 건강상태가 좋은 경우 비흡연 비율이 높게 나타났다($p=.015$). 아침식사를 하는 비율은 GPY 91%, BPY 71.4%로 주관적 건강상태가 좋은 경우 아침식사를 하는 비율이 나타났다($p=.003$). 반면 노인 여자는 주관적 건강상태에 따른 건강행위 실천률의 차이가 유의하지 않았다.

4. 결론

본 연구는 만성통증 노인을 대상으로 주관적 건강상태에 따른 건강행위 실천 패턴을 분석하여 주관적 건강상태를 기반으로 한 건강행위 실천 관리를 위한 기초자료를 마련하고자 시도되었다. 주관적 건강상태가 높거나 낮음에 따른 건강행위 실천 패턴이 남녀노인에서 비슷한 것으로 나타났으나, 건강행위 요소 각각의 실천율은 남녀노인이 차이를 보였다. 한편, 노인 남자의 경우는 주관적 건강상태가 좋은 경우 비흡연, 아침식사 건강행위의 실천비율이 높았다. 또한 비흡연과 아침식사 건강행위가 만성통증을 가진 노인에게서 동시적으로 나타나는 건강행위 실천 패턴임을 고려할 때, 만성통증 노인의 성별에 대한 고려가 필요하며, 특히 남자 노인의 경우 비흡연과 아침식사 건강행위에 대한 적극적이고 집중적인 중재가 필요함을 시사한다.

Reference

- Jana, A., & Chattopadhyay, A. (2022). Prevalence and potential determinants of chronic disease among elderly in India: Rural-urban perspectives. *Plos one*, 17(3), e0264937.
- Zelaya, C. E., Dahlhamer, J. M., Lucas, J. W., & Connor, E. M. (2020). Chronic pain and high-impact chronic pain among US adults, 2019.
- Linn, B. S., & Linn, M. W. (1980). Objective and self-assessed health in the old and very old. *Social Science & Medicine. Part A: Medical Psychology & Medical Sociology*, 14(4), 311-315.
- Mossey, J. M., & Shapiro, E. (1982). Self-rated health: a predictor of mortality among the elderly. *American journal of public health*, 72(8), 800-808.
- Idler, E. L. (1993). Age differences in self - assessments of health: age changes, cohort differences, or survivorship?. *Journal of gerontology*, 48(6), S289-S300.
- Dowd, J. B., & Zajacova, A. (2007). Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US?. *International journal of epidemiology*, 36(6), 1214-1221.
- Atherton, I. M., Evans, J. M. M., Dibben, C. J., Woods, L. M., & Hubbard, G. (2012). Differences in self-assessed health by socioeconomic group amongst people with and without a history of cancer: an analysis using representative data from Scotland. *Journal of cancer survivorship*, 6(4), 458-467.
- Lundberg, O., & Manderbacka, K. (1996). Assessing reliability of a measure of self-rated health.
- Ocampo, J. M. (2010). Self-rated health: Importance of use in elderly adults. *Colombia Médica*, 41(3), 275-289. *Scandinavian journal of social medicine*, 24(3), 218-224.
- Reyes-Gibby, C. C., Aday, L., & Cleeland, C. (2002). Impact of pain on self-rated health in the community-dwelling older adults. *Pain*, 95(1-2), 75-82.
- Mäntyselkä, P. T., Turunen, J. H., Ahonen, R. S., & Kumpusalo, E. A. (2003). Chronic pain and poor self-rated health. *Jana*, 290(18), 2435-2442.
- Siedlecki, S. L. (2006). Predictors of self-rated health in patients with chronic nonmalignant pain. *Pain Management Nursing*, 7(3), 109-116.
- WHO (2021). Ageing and health. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>

응급실에 내원한 3대 중증응급환자 양상변화 (코로나19 이전과 이후 변화 중심으로)

나선경¹, 오미라², 이원영³

요 약

본 연구는 코로나19 발생 전(2019년)·후(2020년)의 응급실 방문 변화에 대한 특성을 알아보았다. 2019년 1월 1일부터 2020년 12월 31일까지의 국가응급진료정보망 자료에서 3대 중증응급질환으로 내원한 743,325건을 분석하였다. 응급실 특성들은 코로나19 전후에 대해서 유의한 차이가 있었고($p < 0.001$), 응급실 재실시간은 2020년이 2019년 보다 13.9분 길게 나타났다. 입원일수는 2020년이 2019년 보다 0.2일 길게 나타났다. 코로나19 유행이 3대 중증응급환자에 대해 영향을 주었고, 감염 위기의 상황에서 진료나 치료가 지연되지 않도록 응급의료체계 개선이 필요하다.

주요용어 : 코로나19, 응급실, 중증응급환자

1. 서론

코로나19와 같은 신종 감염 유행은 공중학적 전반에 걸쳐 위기를 초래하고, 응급의료체계에도 심각한 영향을 준다. 감염 유행은 2009년 신종 인플루엔자, 2015년 중동호흡기증후군(Middle East Respiratory Syndrome, MERS), 2020년 코로나19 등으로 5~6년 주기로 발생하고 있다. 코로나19와 유사한 MERS에서는 초기대응의 부재, 허술한 검역 등으로 감염 환자의 응급실 유입으로 응급실 진료 수요가 증가하게 되고, 응급실이 감염 전파의 창구로 작용하는 심각한 문제를 경험한 바 있다 (Lee, Khang, Lim, 2019).

코로나19 확진자가 급증하면서, 응급실 환자 수용 중지, 발열 중증응급환자 수용 거부 등 응급의료체계의 마비가 반복되었고, 응급실 내원환자의 비율은 2019년에 비해 2020년이 21.8% 감소하였다(NEMC, 2020, 2021). 심근경색증, 뇌졸중과 중증외상으로 대변하는 3대 중증응급질환은 응급처치를 필요로 하는 위급한 질환으로, 응급의료 서비스 치료의 질을 평가하고 관리하는 중요한 질환이면서 치료가 지연될 경우 사망이나 심한 장애를 입게 되는 부정적인 영향을 끼친다(Baek, 2013).

그러므로 코로나19의 상황에서 중증응급환자의 임상 결과에 대한 격차를 극복하기 위해서는 중증응급환자 발생 및 응급의료서비스 이용 변화를 연구할 필요가 있다. 특히 코로나19와 같은 감염이 유행할 경우에 급변하는 응급의료 환경에서 응급의료체계의 향상을 유지하기 위해서 응급의료서비스의 자원 이용을 최적화에 도움을 주고자 한다.

¹04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 대리, 중앙대학교 건강간호대학원 석사과정. E-mail: kings0101@nmc.or.kr

²04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 선임연구원. E-mail: omr@nmc.or.kr

³06974 서울시 동작구 흑석로 84, 중앙대학교 의학부(기초) 교수. E-mail: wylee@cau.ac.kr

2. 연구방법

2019년 1월 1일부터 2020년 12월 31일까지의 국가응급진료정보망(National Emergency Department Information System, NEDIS)의 자료에서, 제7차 한국표준질병사인분류(Korean standard classification of disease)를 이용하여 3대 중증응급질환인 심근경색증(I210~I219)과 뇌졸중(I6300~I64, I610~I629, I600~I609)을 추출하였다. 그리고 중증외상은 외상 중증도 지수가 0.9 미만인 경우를 사용하여 분석하였다.

NEDIS의 자료에서 일반적 특성으로는 성별, 연령, 응급의료기관 종별, 내원경로, 내원수단, 중증도 분류결과, 응급실 진료결과를 사용하였다. 연령은 0~18세, 19~35세, 35~49세, 50~64세, 65세 이상으로 구분하였고, 응급의료기관 종별은 권역응급의료센터와 지역응급의료센터로 나타났다. 내원경로는 직접내원, 외부에서 전원, 외래에서 의뢰, 기타로 구분하였고, 내원수단은 119구급차, 기타구급차, 기타자동차 기타로 구분하였다. 중증도 분류 결과는 한국형 중증도 분류 도구(Korean triage and acuity system, KTAS)에 따라 1~5등급으로 구분하였으며, 응급진료결과는 귀가, 전원, 입원, 사망, 기타로 구분하였다. 응급실 재실시간은 응급실 내원일시와 퇴실일시까지의 차이를 분단위로 계산하였으며, 입원일수는 응급실 퇴실일시와 입원 후 퇴원일시까지의 차이를 일단위로 나타났다.

응급실 일반적 특성과 연도의 비교에서는 빈도와 비율을 나타내어 교차분석인 카이제곱(Chi-square) 검정을 실시하였고, 응급실 재실시간과 입원시간은 t-검정을 실시하였다. 통계분석은 IBM SPSS Statistics 27 (IBM Corp., Armonk, NY, USA)을 사용하였으며, p-값이 0.05 미만인 경우를 통계적으로 유의하다고 판단하였다. 본 연구는 국립중앙의료원의 연구윤리위원회 심의를 통하여 동의면제를 승인받았다(NMC-2022-04-033).

3. 연구결과

3.1. 응급실 환자의 일반적 특성

Table 1에서 연구대상자는 총 743,325건이었고, 2019년은 392,627건(52.8%), 2020년은 350,698건(47.2%)으로 2019년에 비해 41,929건(10.8%) 감소하였다. 3대 중증응급질환에 대한 일반적 특성은 연도에 대한 차이가 있었다($p < 0.001$). 연령은 65세 이상(47.4% vs. 49.2%)이 가장 높게 나타났으며, 성별에서는 남자(57.5% vs. 58.4%)가 여자(42.5% vs. 41.6%) 보다 높게 나타났다. 응급의료기관 종별에서는 권역응급의료센터 (37.5% vs. 36.9%)보다 지역응급의료센터 (62.5% vs. 63.1%)가 높게 나타났으며, 내원경로에서는 직접내원(75.5% vs. 76.7%)이 가장 높게 나타났고 외부에서 전원(22.9% vs. 21.5%)과 외래에서 의뢰(1.6% vs. 1.8%) 순으로 나타났다. 내원수단에서는 119 구급차(43.8% vs. 46.9%)가 가장 높게 나타났고 기타 자동차(39.9% vs. 38.0%), 기타 구급차(15.8% vs. 14.5%)의 순으로 나타났으며, 중증도 분류결과에서는 KTAS 3(49.5% vs. 51.6%), KTAS 4(24.8% vs. 21.9%), KTAS 2(19.3% vs. 20.1%) 순으로 나타났다. 그리고 응급진료결과는 귀가는 입원(67.2% vs. 69.8%)이 가장 많았고 귀가는 24% 이상(26.2% vs. 24.3%)로 나타났다.

Table 1. Comparison of characteristics by year

Variables	Total (N=743,325)	Year (yr)		χ^2 or T	p-value
		2019 (N=392,627)	2020 (N=350,698)		
Age				433.27	<0.001
0 - 18	27,385 (3.7)	15,511 (4.0)	11,874 (3.4)		
19 - 34	60,561 (8.1)	32,658 (8.3)	27,903 (8.0)		
35 - 49	97,554 (13.1)	53,213 (13.6)	44,341 (12.6)		
50 - 64	198,987 (26.8)	105,068 (26.8)	93,919 (26.8)		
≥ 65	358,838 (48.3)	186,177 (47.4)	172,661 (49.2)		
Gender				53.57	
Male	430,603 (57.9)	225,891 (57.5)	204,712 (58.4)		
Female	312,722 (42.1)	166,736 (42.5)	145,986 (41.6)		
Type of ED				31.79	<0.001
REMC	276,579 (37.2)	147,263 (37.5)	129,316 (36.9)		
LEMC	466,746 (62.8)	245,364 (62.5)	221,382 (63.1)		
Path of ED				253.85	<0.001
Direct	565,336 (76.1)	296,306 (75.5)	269,030 (76.7)		
Transfer from other hospital	165,227 (22.2)	89,945 (22.9)	75,282 (21.5)		
Request of outpatient	12,592 (1.7)	6,285 (1.6)	6,307 (1.8)		
Other	170 (0.0)	91 (0.0)	79 (0.0)		
Visiting method of ED				761.83	<0.001
119 ambulance	336,470 (45.3)	172,039 (43.8)	164,431 (46.9)		
EMS ambulance	112,856 (15.2)	61,874 (15.8)	50,982 (14.5)		
Other ground transport	290,041 (39.0)	156,760 (39.9)	133,281 (38.0)		
Other	3,958 (0.5)	1,954 (0.5)	2,004 (0.6)		
KTAS grade				972.83	<0.001
KTAS 1	25,379 (3.4)	12,939 (3.3)	12,440 (3.5)		
KTAS 2	146,024 (19.6)	75,640 (19.3)	70,384 (20.1)		
KTAS 3	375,275 (50.5)	194,280 (49.5)	180,995 (51.6)		
KTAS 4	147,163 (23.4)	97,295 (24.8)	76,868 (21.9)		
KTAS 5	22,484 (3.0)	12,473 (3.2)	10,011 (2.9)		
Results of ED				716.30	<0.001
Discharge	188,140 (25.3)	102,912 (26.2)	85,228 (24.3)		
Transfer	39,654 (5.3)	22,479 (5.7)	17,175 (4.9)		
Admission	508,869 (68.5)	263,913 (67.2)	244,956 (69.8)		
Death	5,450 (0.7)	2,736 (0.7)	2,714 (0.8)		
Other	1,212 (0.2)	587 (0.1)	625 (0.2)		
LOS of ED, min	288.7 ± 645.7	282.2 ± 487.6	296.1 ± 785.7	-8.99	<0.001
Myocardial infarction	370.7 ± 617.9	369.6 ± 545.1	371.9 ± 686.9	-0.50	0.614
Stroke	337.9 ± 437.3	329.3 ± 442.4	347.1 ± 431.6	-9.82	<0.001
Severe trauma	252.4 ± 726.9	247.1 ± 495.2	258.4 ± 924.9	-5.12	<0.001
LOS in hospital, day	15.3 ± 20.2	15.2 ± 20.1	15.4 ± 20.3	-2.75	0.006
Myocardial infarction	7.2 ± 12.3	7.2 ± 12.4	7.2 ± 12.3	-0.22	0.823
Stroke	16.7 ± 22.8	16.4 ± 22.5	17.1 ± 23.0	-6.45	<0.001
Severe trauma	16.7 ± 20.0	16.8 ± 20.0	16.7 ± 19.9	-1.36	0.174

Values are presented cases (%) or mean ± SD.

SD: Standard deviation, ED: Emergency department, REMC: Regional emergency medical center, LEMC: Local emergency medical center, KTAS: Korean triage and acuity scale, LOS: Length of stay.

3.2. 재실시간과 입원일수

Table 1의 3대 중증응급질환에 대한 응급실 재실시간에서는 전체, 뇌졸중, 중증외상에 대해 연도별로 매우 유의한 차이가 있었다($p < 0.001$). 전체는 2019년이 282.2분이고 2020년은 296.1분으로 나타났으며, 뇌졸중은 2019년이 329.3분이고 2020년은 347.1분으로 나타났으며, 2020년은 2019년에 비해 17.8분 길게 나타났다. 그리고 중증외상은 2019년이 247.1분이고 2020년은 258.4분으로 나타났으며, 2020년은 2019년에 비해 11.3분 길게 나타났다. 입원일수는 전체와 뇌졸중은 연도별로 대해서 매우 유의한 차이가 있었다($p < 0.01$). 전체의 입수에서 2020년은 2019년에 비해 0.2일 길게 나타났고, 뇌졸중에서 2020년은 2019년에 비해 0.7일 길게 나타났다.

4. 결론

본 연구는 코로나19 발생 전·후에 대한 응급의료서비스 이용 변화를 살펴보기 위해, 3대 중증응급질환자의 일반적 특성과 응급실 재실시간 및 입원일수를 살펴보았다. 응급실에 내원한 3대 중증응급질환자 수는 2019년에 비해 10.8% 감소하였으며, 응급실의 일반적 특성은 코로나19 전후와 차이가 있었다. 응급실 재실시간은 전체와 뇌졸중과 중증외상을 코로나19 이전인 2019년보다 2020년에 길게 나타났고, 입원일수는 전체와 뇌졸중은 2020년에 더 길게 나타났다. 이와 같은 결과는 코로나19 감염 유행 기간 중에 선별진료소(감염증 의심증상자가 출입 전 진료를 받도록 하는 공간)를 설치 운영에 따른 응급실 입실 및 입원절차의 변화가 영향을 줬을 것으로 생각할 수 있다. 또한 코로나19 유행은 급성심근경색과 같은 적시에 치료가 필요한 중증응급질환자들마저도 코로나19 감염에 대한 두려움, 과도한 의료시설 제한으로 의료접근성 감소에 따라 응급실 방문 자체를 하게 되는 원인이 되었다(Sung et al., 2021).

결론적으로 코로나19 유행은 3대 중증응급질환에 대해 영향을 주었으며, 감염 위기의 상황에서 진료나 치료가 지연되지 않도록 응급의료체계 개선이 필요하다.

References

- Baek, K. H. (2021). Analysis and consideration of the cases related to emergency medical care - Focusing on the three major emergent diseases severe acute -. *Korean Journal of Medicine and Law*, 21(2), 57-86. (in Korean).
- Lee, S. Y., Khang, Y. H., Lim, H. K. (2019). Impact of the 2015 Middle East Respiratory Syndrome Outbreak on Emergency Care Utilization and Mortality in South Korea. *Yonsei Medical Journal*, 60(8), 796-803. DOI : <https://doi.org/10.3349/ymj.2019.60.7.796>
- National Emergency Medical Center (NEMC). (2020). 2019 National emergency department information system statistical annual report. (in Korean). Available from: https://www.e-gen.or.kr/nemc/statistics_annual_report.do?brdclscd=02
- National Emergency Medical Center (NEMC). (2021). 2020 National emergency department information system statistical annual report. (in Korean). Available from: https://www.e-gen.or.kr/nemc/statistics_annual_report.do?brdclscd=02
- Sung, H. K., Paik, J. H., Lee, Y. J., Kang, S. (2021). Impact of the COVID-19 Outbreak on Emergency Care Utilization in Patients with Acute Myocardial Infarction: a Nationwide Population-based Study. *Journal of Korean medical science*, 36(16), e111. DOI : <https://doi.org/10.3346/jkms.2021.e111>

한국 성인에서 체질량지수와 허리둘레를 기준으로 분류한 비만 유형별 건강행위 특성 분석*

김선영¹, 이명숙², 신지은³

1. 서론

비만 유행률은 1975년 이후 전 세계적으로 3배 이상 증가 했으며[1], 2017년에는 비만으로 인한 사망이 전 세계적으로 8%인 것으로 보고 되었다[2]. 비만을 판정하는 지표는 현재까지 BMI [체중(kg)/신장(m)²]가 가장 많이 사용 왔으나, BMI가 정상이라도 허리둘레나 허리 엉덩이 둘레비(WHR)가 높으면 당뇨병이나 심혈관계질환, 대사증후군 등의 발생 위험이 증가한다는 연구결과가 보고되었으나[3] BMI를 기준으로 한 비만과 허리둘레를 기준으로 한 복부비만의 각각의 특성과 연관성에 관련된 연구 논문은 거의 없다. 비만에 대한 다양한 연구들이 진행되면서 생활습관과 비만 사이의 밀접한 관계에 대한 연구 논문들이 발표되었고, 생활습관이 비만을 예방하기 위한 중요한 요인임을 알 수 있다[4-7]. 생활습관을 설명할 때에 있어서 건강 행위는 필수 불가결하다고 볼 수 있다[8,9]. 특히 질환군에서의 생활습관에 대한 연구들은 주로 건강 행위의 특성에 초점을 맞추어 의미 있는 결과를 도출한다. 이 연구는 한국 성인의 비만유형(BMI/WC, WC, BMI)에 따라 실천하는 건강행위실천 패턴을 파악하고 생활습관 개선을 통한 건강증진을 위한 기초자료를 제공하고자 한다.

2. 연구방법

2.1 연구대상

본 연구는 국민건강영양조사 2015-2019년까지 참여한 성인 23,701명 중 (허리둘레와 체질량지수가 측정되지 않은 대상자 및 저체중, 임신/수유 중인 여성, 500kcal 미만 혹은 5000kcal 초과섭취, 질환으로 인한 식이요법 대상자, 암인 대상자, 질병으로 신체활동 제한인 대상자, 대사증후군 관련 지표가 측정되지 않은 대상자 등)을 제외하고 최종 14,713명(남자 6443명, 여자 8270명)을 선정하였다.

*이 논문은 농촌진흥청 연구사업(세부과제번호 : PJ015269022022)의 지원에 의해 이루어진 것입니다.

¹35365 대전광역시 서구 관저동로 158, 건양대학교병원 헬스케어데이터사이언스센터 연구원.

E-mail: 1020peace@kyuh.ac.kr

²02844 서울특별시 성북구 보문로 34다길, 성신여자대학교 바이오헬스융합학부 교수.

E-mail: mlee@sungshin.ac.kr

³(교신저자) 35365 대전광역시 서구 관저동로 158, 건양대학교 의과대학 조교수. E-mail: jeshin@konyang.ac.kr

2.2 자료분석방법

통계분석은 SAS ver 9.4(SAS Institute Inc, Cary, NC, USA)를 이용하였으며, 질병관리본부의 국민 건강영양조사 원시자료 이용지침에 따라 가중치를 사용하여 복합표본분석을 시행하였다. 비만유형에 따른 건강행위 특성을 파악하기 위해 연령을 보정하여 교차분석을 실시하였다. 모든 분석은 남자와 여자로 나누어 분석하였다.

2.3 측정도구

1) 비만

국민건강영양조사의 검진조사 중 BMI는 [체중(kg)/신장(m)²] 이용하여 산출하였다. 세계 보건기구(World Health Organization)의 기준에 따라 BMI는 아시아인을 위한 비만 기준(World Health Organization for Asian population)을 근거하여, 18.5 kg/m² 미만을 “저체중”, 18.5 kg/m² 이상 25 kg/m² 미만을 “정상”, 25 kg/m² 이상을 “비만”으로 분류하였다[10]. 복부비만의 기준으로는 대한비만학회[11]에서 정한 남자 90 cm, 여자 85 cm 이상을 복부비만으로 정하였다

2) 건강관련행위

“Alameda 7”의 7가지 건강관련행위 중 간식섭취 항목을 외식으로 대체하였고, 여기서 적정체중 유지는 비만도와 관련되어 제외하고 최종적으로 비흡연, 음주, 적정수면, 신체활동, 아침식사, 외식 등 6가지 건강관련 행위를 선정하였다. 비흡연은 평생 피운적 없거나 과거엔 피웠으나 현재 피우지 않은 경우, 음주는 평생 비음주이거나 최근 1년 이내 주 2회 미만으로 음주한 경우, 적정수면은 하루 7시간 이상~8시간인 경우, 신체활동은 주당 150분 이상 운동하는 경우, 아침식사는 주 5-7회 섭취하는 경우, 외식은 주 4회 이하로 외식하는 경우를 건강행위로 정의하였다.

3. 결과

3.1 비만유형에 따른 일반적 특성

체질량지수와 허리둘레를 기준으로 분류한 비만 유형은 남자의 경우, 체질량지수(BMI)와 허리둘레(WC) 모두 정상인 경우가 53%(n=3,373), BMI/WC가 모두 비만인 경우가 29.6%(1,927), WC가 비만인 경우가 3.3%(244), BMI가 비만인 경우가 14.1%(899)로 나타났으며, 여자의 경우 BMI/WC 모두 비만이 아닌 경우가 71.4%(n=5,757), BMI/WC가 모두 비만인 경우가 16.8%(1,502), WC가 비만인 경우가 3.1%(284), BMI가 비만인 경우가 8.7%(727)로 나타났다. 성인여자가 성인남자보다 비만이 아닌 비율이 높았으며, WC 비만군은 남녀 성인이 비슷한 비율로 분포하는데 반해 BMI/WC 비만군 및 BMI 비만군은 남자가 약 2배 정도 높은 비율로 분포하였다.

비만유형에 따른 일반적 특성은 성인 남자는 연령(p<.001), 결혼상태(p<.001), 경제활동여부(p<.001)에서 통계적으로 유의한 차이를 보인 반면, 성인 여자는 연령(p<.001), 결혼상태(p<.001), 가구소득(p<.001), 경제활동여부(p=.001) 및 교육수준(p<.001)에서 통계적으로 유의한 차이를 보였다.

연령의 경우 성인 남녀 모두 WC군의 연령 평균이 가장 높았고, 정상군의 연령 평균이 가장 낮았다. 결혼상태는 남자의 경우 배우자가 있는 기혼인 경우가 BMI/WC군의 비율이 가장 높는데 반해, 여자는 배우자가 없는 기혼인 경우가 BMI/WC군의 비율이 가장 높았다.

경제활동여부에서는 남자는 경제활동을 하는 경우 모든 비만유형에서 비만에 해당하는 비율이 높는데 반해 여자들은 경제활동을 하는 경우 BMI/WC군과 WC군의 비율이 낮았다.

성인여자의 가구소득은 소득분위가 높을수록 정상군의 비율은 높아졌고, BMI/WC군과 WC군의 비율은 낮아졌으며, 교육수준에서는 대졸 이상인 경우가 대졸인 경우가 정상군의 비율이 가장 낮았다.

3.2 비만유형에 따른 건강행위실천율 비교

비만유형에 따른 건강행위 실천율을 비교한 결과, 성인 남자에서는 건강행위를 실천하는 비율은 비음주($p=.009$), 아침식사($p=.024$), 신체활동($p<.001$)에서 통계적으로 유의한 차이가 있었다. 성인 여자에서는 비음주를 제외한 모든 건강행위(비흡연($p=.004$), 적정수면시간($p=.038$), 아침식사($p<.001$), 신체활동($p=.001$), 비외식($p<.001$)) 실천에서 통계적으로 유의한 차이가 있었다.

성인 남자에서 건강행위를 실천하는 비율은 비음주 요소에서는 정상군이 67.4%로 가장 높았고, WC군이 55.8%로 가장 낮았다. 아침식사요소에서는 WC군이 59.1%로 가장 높았고, BMI/WC군이 48%로 가장 낮았다. 적정량의 신체활동은 BMI군이 60.6%로 가장 높았고, WC군이 41.3%로 가장 낮았다.

성인 여자에서 건강행위를 실천하는 비율은 비흡연 요소에서는 BMI군이 95.4%로 가장 높았고, BMI/WC군이 91.7%로 가장 낮았다. 적정수면시간은 BMI군과 정상군에서 각각 33.3%, 32%로 가장 높았고, BMI/WC군이 26.6%로 가장 낮았다. 정상군이 67.4%로 가장 높았고, WC군이 55.8%로 가장 낮았다. 아침식사 요소에서는 WC군이 59.1%로 가장 높았고, BMI/WC군이 48%로 가장 낮았다. 아침식사 요소에서는 WC군이 68.7%로 가장 높았고, 운동요소는 정상군이 49.8%로 가장 높았다. 비외식은 WC군에서 72.7%로 가장 높았다.

4. 결론 및 제언

본 연구는 성인 남녀를 대상으로 비만유형특성을 파악하고 비만유형에 따른 인구사회학적 특성 및 건강 행위 실천율을 비교를 통해 건강증진을 위한 기초자료를 마련하고자 시도되었다.

성인 남녀의 비만유형 분포는 BMI/WC 비만군 및 BMI 비만군에서 성인 남자가 성인여자 보다 약 2배 정도 높은 비율로 분포하였으며, 비만유형의 특성이 상이함을 확인하였다. 또한 기본 특성과 비만유형이 연관성이 있는 것으로 나타났으며, 특히 남자는 경제활동과 여자는 학력 등과 관계가 있다는 점은 인구사회학적 요인을 고려한 건강증진 전략이 필요하다는 것을 시사한다. 이어서 성인 남녀 모두 정상군이나 BMI군이 건강행위 실천율이 높은 편인데 반해 WC군이나 BMI/WC군의 건강행위 실천율이 낮은 편으로 나타났다. 이는 비만유형에 따라 건강행위 실천 패턴을 이해하고, 비만유형별 차별화된 건강행위 실천 전략이 필요하다는 것을 시사한다.

Reference

- WHO. Obesity and overweight. WHO Web site. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>. Updated 2021.
- Hannah R, Max R. Obesity. Our World in Data Web site. <https://ourworldindata.org/obesity>. Updated 2017. Accessed Nov 18, 2021.
- Zhu S, Wang Z, Heshka S, Heo MS, Faith M, Heymsfield S. Waist circumference and obesity-associated risk factors among whites in the third National Health and Nutrition Examination Survey : clinical action thresholds. *Am J Clin Nutr* 2002, 76: 743-749
- WANG, Catharine; COUPS, Elliot J. Causal beliefs about obesity and associated health behaviors: results from a population-based survey. *International Journal of Behavioral Nutrition and Physical Activity*, 2010, 7.1: 1-7.
- MARCONCIN, Priscila, et al. The association of healthy lifestyle behaviors with overweight and obesity among older adults from 21 countries. *Nutrients*, 2021, 13.2: 315.
- CHUNG, Sang-Jin, et al. Urban and rural differences in the prevalence of gender and age specific obesity and related health behaviors in Korea. *Journal of Korean medical science*, 2005, 20.5: 713-720.
- DAVIS, Ann McGrath, et al. Obesity and related health behaviors among urban and rural children in the United States: data from the National Health and Nutrition Examination Survey 2003 - 2004 and 2005 - 2006. *Journal of pediatric psychology*, 2011, 36.6: 669-676.
- SZAKÁLY, Zoltán, et al. The influence of lifestyle on health behavior and preference for functional foods. *Appetite*, 2012, 58.1: 406-413.
- Schoenborn CA. Health habits of U.S. adults, 1985: the "Alameda 7" revisited. *Public Health Rep*. 1986;101(6):571-580.
- World Health Organization. Obesity: Preventing and Managing the Global Epidemic; (No. 894); World Health Organization: Geneva, Switzerland, 2000
- Korean Society for the Study of Obesity. The criteria of body mass index and waist circumference for obesity and central adiposity in Koreans. Seoul: *Korean Society for the Study of Obesity*, 2005

임상간호사의 일과 삶의 균형, 간호업무성과가 행복지수에 미치는 영향

남문희¹, 허현점²

요 약

본 연구의 목적은 임상간호사의 일과 삶의 균형, 간호업무성과, 행복지수 정도를 파악하여 간호사의 행복지수를 높이기 위한 방안을 마련하고자 시도되었다. 연구대상은 B광역시와 K도에 소재한 161명의 병원간호사를 대상으로 2022년 5월 6일부터 5월20일까지 자가설문을 통하여 조사하였다. 자료분석은 대상자의 일반적 특성은 기술통계를 구하였고, 대상자의 특성에 따른 이직 의도의 차이는 t-test, ANOVA로 분석하였고, 이에 대한 사후 분석은 Duncan test로 분석하였으며, 변수간의 상관계수는 Pearson's correlation coefficient로 분석하였다. 분석결과 일반적 특성에 따른 행복지수의 차이는 남성($t=3.608$, $p<.001$), 현 근무부서($F=3.271$ $p=.023$), 월평균 급여 수준($F=4.068$ $p=.019$)에 따라 유의한 차이가 있었고, 간호사의 행복지수는 일과 삶의 균형($r=.591$, $p<.001$), 간호업무성과($r=.396$, $p<.001$) 간에 유의한 양의 상관관계가 있는 것으로 나타났다. 간호사의 행복지수에 영향을 미치는 요인을 분석한 결과 일과 삶의 균형($\beta=.48$, $p<.001$), 급여수준 400만원 이상($\beta=.24$, $p=.01$), 301~400만원($\beta=.23$, $p=.03$), 간호업무성과($\beta=.20$, $p<.001$), 성별에서 남자($\beta=.14$, $p=.03$)로 나타났으며, 행복지수에 대한 이들 변수의 설명력은 40.7%이었다. 결론적으로 간호사의 행복지수를 높이기 위해 근무부서, 외적보상 뿐만 아니라 일과삶의 균형, 간호업무성과를 높여 행복지수를 낮추는 정책개선이 필요하다.

주요용어 : 임상간호사, 일과 삶의 균형, 간호업무성과, 행복지수.

1. 서론

1.1. 연구의 필요성

최근 의료기관에서는 의료서비스를 요구하는 고객들의 요구는 점점 증가하고 있고, 코로나 감염 팬데믹 현상으로 국민의 생명과 환자 안전을 지키기 위해서 간호사들은 정신적 육체적으로 힘든 일을 하고 있다.

이러한 사회적 분위기에서 요즘 젊은 세대들은 일과 삶의 균형(Work-Life Balance, WLB)을 유지하면서 자신 삶의 질을 향상시키고자 하는 욕구가 점차 높아지는 양극화가 초래되고 있다. 일과 삶의 균형이란 자신의 삶에 관심이 많고, 일과 함께 개인 삶의 행복을 중요시 여겨 균형적인 삶을 사는 것에 가치를 두는 관점이다(조정화, 2021). 간호사는 전문직종이지만 직무의 특성상 교대근무,

¹(First author)Department of Nursing, Daedong College, 68 Dongbugok-ro, 27 beon-gil, Geumjeong-gu, Busan, 46270, Korea. E-mail: ny386@hanmail.net

²(Corresponding author) Department of Nursing, Gaya University 208 Samgye-ro, Gimhae-si, Gyeongnam, 50830, Korea. E-mail: hhyunjum@hanmail.net

야간근무, 휴일근무 등 불규칙한 근무를 하게 되어 일과 삶의 균형과 관련 지수가 다른 근로자에 비해 차이가 있을 수 있다. 선행연구에서 간호사의 일과 삶의 균형은 간호사의 간호서비스의 질과 관련이 있고(조정화, 2021), 여가만족이 있을수록 간호업무 성과가 높아(이은주, 2019) 간호사의 일과 삶의 균형은 간호업무성과와 관련되는 요인으로 작용할 수 있다.

간호사의 간호업무성과는 긍정심리역량과(차기정, 2017), 행복감과 임파워먼트가 주요한 변수로 작용하고 있었다(김선영, 권영은, 2020). 하지만 간호사의 일과 삶의 균형, 간호업무성과, 간호사의 행복지수의 관계를 직접적으로 연구한 부분은 거의 없는 실정이다. 이에 본 연구에서는 일과 삶의 균형, 간호업무성과, 행복지수 정도를 파악하고 변수 간의 관계를 조사하여 간호사의 워라벨과 간호업무성과 등을 연계하여 행복지수를 높이기 위한 기초자료로 활용하고자 한다.

1.2. 연구의 목적

본 연구는 간호사의 일과 삶의 균형, 간호업무성과, 행복지수 간의 관계를 파악하고 이들 간의 상관관계를 조사하여 간호사의 행복지수를 높이기 위한 유용한 자료로 제공하기 위함이며, 구체적인 목적은 다음과 같다.

첫째, 임상간호사의 일과 삶의 균형, 간호업무성과, 행복지수 정도를 파악한다.

둘째, 임상간호사의 일반적 특성에 따라 일과 삶의 균형, 간호업무성과, 행복지수 정도를 파악한다.

셋째, 임상간호사의 일과 삶의 균형, 간호업무성과, 행복지수 간의 상관관계를 파악한다.

2. 연구방법

2.1. 연구설계

본 연구에서는 임상간호사의 일과 삶의 균형, 간호업무성과, 행복지수 정도를 파악하고 간호사의 행복지수를 높이기 위한 서술적 상관관계 연구이다.

2.2. 연구대상

본연구는 2022년 5월6일부터 5월20일까지 B광역시와 K도에 소재한 의료기관에 근무하는 간호사 161명을 대상으로 설문지를 통하여 조사 하였다.

2.3. 자료분석방법

수집된 자료는 SPSS 24.0 Window 프로그램을 이용하여 분석하였고 분석 방법은 다음과 같다.

- 1) 대상자의 일반적 특성은 빈도, 백분율, 평균, 표준편차로 분석하였다.
- 2) 대상자의 간호사의 일과 삶의 균형, 간호업무성과, 행복지수는 평균과 표준편차로 분석하였다.
- 3) 대상자의 일반적 특성에 따른 일과 삶의 균형, 간호업무성과, 행복지수의 차이는 t-test, ANOVA로 분석하였고, 이에 대한 사후 분석은 Duncan test로 분석하였다.

4) 대상자의 간호사의 일과 삶의 균형, 간호업무성과, 행복지수 간의 상관관계는 Pearson's correlation coefficient로 분석하였다.

5) 대상자의 행복지수에 영향을 미치는 요인들을 파악하기 위해 Multiple regression analysis를 실시하였다.

3. 연구결과

3.1. 대상자의 일과 삶의 균형, 간호업무성과, 행복지수 정도

Table 1. Participants' Work-Life Balance, Nursing performance and Happiness index (N=161)

Variables	Range	Mean±SD	min	max
Work-Life Balance	1~5	3.29 ±0.54	1.87	4.80
Work-Home Balance	1~5	3.76 ±0.65	1.80	5.00
Work-Leisure Balance	1~5	3.15 ±0.76	1.00	5.00
Work-growth Balance	1~5	2.95 ±0.76	1.40	5.00
Nursing performance	1~5	3.73 ±0.47	2.65	5.00
Happiness index	1~5	3.37 ±0.55	1.82	4.82

3.2. 대상자의 행복지수에 영향을 미치는 요인

Table 2. Factors Influencing Happiness index (N=262)

Variables	B	SE	β	t	p	
(Constant)	0.55	0.35		1.55	0.12	
Sex *Male=1	0.50	0.22	0.14	2.23	0.03	
Assigned ward * Med=1	GSb	0.20	0.11	1.79	0.08	
	ICU, ER, ORc	0.04	0.09	0.03	0.42	0.67
	Etc.d	0.21	0.11	0.17	1.93	0.05
Salary (ten thousand won) * <300=1	301~400b	0.27	0.13	2.19	0.03	
	≥400c	0.26	0.10	2.53	0.01	
Work-life balance	0.49	0.07	0.48	6.75	0.00	
Nursing performance	0.23	0.08	0.20	2.92	0.00	

R2=.436, Adj. R2=.407, F=14.709, p<.001

*Dummy variables processing

4. 결론 및 제언

본 연구는 임상간호사의 일과 삶의 균형, 간호업무성과, 행복지수 정도와 상관관계를 파악한 결과 간호사의 행복지수는 일과 삶의 균형($r=.591, p<.001$), 간호업무성과($r=.396, p<.001$) 간에 유의한 양의 상관관계가 있는 것으로 나타났고, 간호사의 행복지수에 영향을 미치는 요인은 일과 삶의 균형($\beta=.48, p<.001$), 급여수준 400만원 이상($\beta=.24, p=.01$), 301~400만원($\beta=.23, p=.03$), 간호업무성과($\beta=.20, p<.001$), 성별에서 남자($\beta=.14, p=.03$)로 나타났으며, 행복지수에 대한 이들 변수의 설명력은 40.7%로 나타났다.

본 연구 결과를 통해 간호사의 행복지수를 높이기 위해 임상 현장에서 간호사의 부서배치, 외적 보상이 우선되어야 할 뿐만 아니라 간호사의 일과 삶의 균형인 일가정 균형, 일 여가 균형, 일 성장 균형을 높이는 정책으로 간호업무성과를 높여 행복지수를 높이는 정책 개선이 필요성을 제시한 점에 의의가 있다.

References

- Cho, Jeonghwa. Mediating Effect of Organizational Culture Supporting Work-Life Balance on the Relationship between Work-Life Balance and Quality of Nursing Service in Clinical Nurses. *Korean Journal of Occupational Health Nursing* Vol. 30 No. 3, 79-89, August 2021
- Organization for Economic Cooperation and Development. (2014).OECD Better Life Index. Retrieved March 1, 2019, from <http://www.oecdbetterlifeindex.org/countries/korea/>

코로나19로 인한 일상 및 행동 변화 분석

차효정¹

요약

본 연구에서는 2021년 경남사회조사를 중심으로 경남도민의 코로나19로 인한 일상 및 행동 변화를 분류하고 유형화하여, 분석을 통해 코로나19로 인한 경남도민의 일상 및 행동 변화에 영향을 미치는 요인을 파악하고자 하였다. 본 연구의 결과를 통해 사회의 위기적 상황과 개인 변화의 관계를 살펴봄으로써 지역사회의 변화 양상을 분명하게 규명하여 앞으로 발생할 변화 예측을 위한 정책수립의 자료로 활용할 수 있을 것으로 사료되며, 지역주민의 삶의 질 향상을 위한 기초자료로 활용할 수 있을 것이다.

주요용어 : 의사결정나무, 일상 변화, 코로나19, 행동 변화, k-평균 군집 분석.

1. 서론

코로나19는 2019년 12월에 발발하여 2022년 6월인 현재까지 이어지면서 전 세계 사람들의 삶을 빠르게 변화시키고 있다. 전 세계 코로나19로 인한 사망자 수는 5백만 명을 넘어섰으며(WHO, 2022a), 우리나라는 2022년 1월 23일을 기준으로 6,540명의 사망자가 집계되었다(Korea Disease Control and Prevention Agency, 2022). 코로나19로 인한 사회적 거리두기로 인해 직계가족뿐만 아니라 친구, 직장동료들과 단절되고, 해외 출입국 또한 통제 되었다.

본 연구에서는 사회조사를 중심으로 한 경남도민의 코로나19로 인한 일상 및 행동 변화를 분류하고 유형화하여, 분석을 통해 코로나19로 인한 경남도민의 일상 및 행동 변화에 영향을 미치는 요인을 파악하고자 한다. 나아가 사회의 위기적 상황과 개인 변화의 관계를 살펴봄으로써 지역사회의 변화 양상을 분명하게 규명하여 앞으로 발생할 변화를 이해하고, 대비하는 것에 도움이 되고자 한다.

2. 연구 방법

2021년 경상남도 사회조사에서는 매년 실시하는 설문 조사와는 다르게 코로나19로 인하여 경남도민의 일상생활 변화 및 행동 변화를 조사하였고, 본 논문에서는 경남도민의 코로나19로 인한 일상 및 행동 변화에 대하여 중점적으로 연구하고자 한다(Gyeongnam, 2021). 본 논문에서 사용한 자료는 Table 1과 같이 2021년 경상남도에서 조사한 사회 조사 자료이며, 인구통계학적 속성 7개 문항, 코로나19 부문 12개 문항을 추출하여 분석을 실시하였다.

¹51140 경남 창원시 의창구 창원대학교 20 국립창원대학교 항노화헬스케어학과 강사. E-mail: shumy@changwon.ac.kr

Table 1. Analysis variables

Sequence	Section	Variable	Type
1		City and county classification	Categorical type
2		Gender	Categorical type
3		Age	Categorical type
4	Demographic characteristics	Type of residence	Categorical type
5		Residential form	Categorical type
6		Degree of education	Categorical type
7		Marital status	Categorical type
8		Child care	Likert-type 5-point scale
9		Domestic work	Likert-type 5-point scale
10	Changes in daily life	Using digital devices	Likert-type 5-point scale
11		Leisure activity	Likert-type 5-point scale
12		Religious activity	Likert-type 5-point scale
13		Meeting activity	Likert-type 5-point scale
14	COVID19	Refraining from going out	Likert-type 5-point scale
15		Cancel Meeting	Likert-type 5-point scale
16		Refrain from using public transportation	Likert-type 5-point scale
17		Wearing a mask	Likert-type 5-point scale
18	Change in behavior	Washing hands	Likert-type 5-point scale
19		Periodic indoor ventilation	Likert-type 5-point scale

3. 분석 결과

일상 변화의 6개 문항에 대하여 요인 분석을 실시하였으며, 그 결과는 Table 2와 같다.

Table 2. Factorial analysis results (Changes in daily life)

Factor name	Survey question	Factor loading	
		Factor 1	Factor 2
Factor 1. Indoor activity	Child care	0.833	
	Domestic work	0.866	
	Using digital devices	0.669	
Factor 2. Outdoor activity	Leisure activity		0.851
	Religious activity		0.849
	Meeting activity		0.800
Eigen value		2.473	1.562
Percentage of variance		41.224	26.028
Percentage of cumulative variance		41.224	67.251
KMO value = 0.689, Bartlett's p-value = 0.000			

다음으로 일상 변화와 행동 변화에 대한 요인 분석 결과 나타난 4가지 하위 요인(실내활동 변화, 실외활동 변화, 대인방역행동 변화, 개인방역행동 변화)에 대한 인구통계학적 속성에 대한 차이 분석을 실시하였다.

Table 3. Difference analysis results (changes in indoor activity)

Variable	Category	Average	Standard deviation	t or F	p value
Residential area	City	3.39	0.76	0.000	0.000
	County	3.25	0.57		
Gender	Male	3.33	0.69	0.745	0.456
	Female	3.32	0.68		
Age	Below 30s	3.64e	0.80	624.268	0.000
	40s	3.56d	0.74		
	50s	3.27c	0.68		
	60s	3.17b	0.60		
	Over 70s	3.13a	0.49		
Type of residence	Detached house	3.22a	0.60	332.584	0.000
	Apartment	3.45c	0.76		
	Townhouse and multi-family house	3.37b	0.69		
Residential form	One's own house	3.31a	0.67	20.961	0.000
	Chartered house	3.44c	0.76		
	Monthly rent house	3.35ab	0.70		
	Free of charge house	3.36b	0.74		
Degree of education	Elementary school graduation and below	3.13a	0.50	487.960	0.000
	Middle school graduation	3.22b	0.61		
	High school graduation	3.31c	0.71		
	University graduation and higher	3.55d	0.76		
Marital status	Single	3.53d	0.80	242.788	0.000
	Married and Spouse	3.33c	0.68		
	Bereavement	3.11a	0.49		
	Divorce	3.26b	0.62		

* a, b, c, d, e is the post-hoc result.

다음으로 일상 변화 및 행동 변화에 대하여 k-평균 군집 분석을 실시하였으며, 변수의 수를 고려하여 군집의 수를 2개~5개로 지정하여 분석을 실시하였다.

Table 4. Two-means cluster analysis results

Section	Sub-factor	Variable	group	Average	Standard deviation	t	p-value
Changes in daily life	Indoor activity	Child care	1 group	3.74	0.84	27.063	0.000
			2 group	2.68	0.99		
		Domestic work	1 group	3.63	0.82	29.767	0.000
			2 group	2.50	0.92		
	Using digital devices	1 group	3.75	0.85	47.201	0.000	
		2 group	1.92	0.91			
	Outdoor activity	Leisure activity	1 group	1.46	0.60	-122.553	0.000
			2 group	2.98	0.63		
		Religious activity	1 group	1.42	0.59	-106.689	0.000
			2 group	2.96	0.68		
Meeting activity	1 group	1.27	0.47	-56.535	0.000		
	2 group	3.02	0.94				

Table 4. Two-means cluster analysis results(continued)

Section	Sub-factor	Variable	group	Average	Standard deviation	t	p-value
Change in behavior	Interpersonal quarantine behavior	Refraining from going out	1 group	2.12	0.57	-171.668	0.000
			2 group	3.56	0.58		
		Cancel Meeting	1 group	2.02	0.78		
			2 group	3.66	0.51		
		Refrain from using public transportation	1 group	2.08	0.86		
			2 group	3.70	0.54		
	Personal quarantine behavior	Wearing a mask	1 group	3.99	0.09	68.260	0.000
			2 group	3.00	0.57		
		Washing hands	1 group	3.97	0.16		
			2 group	2.86	0.56		
Periodic indoor ventilation	1 group	3.87	0.33	215.372	0.000		
	2 group	2.60	0.58				

마지막으로 일상 변화와 행동 변화의 4가지 하위 영역에 대한 의사결정나무 분석을 실시하였다. 나무모형 생성 시, 변수들의 속성을 고려하여 CART 알고리즘을 사용하여 나무 모형을 생성하였으며, 노드 분할의 유의수준을 0.05로 지정하고 최소 노드의 개수를 50개로 지정하였다.

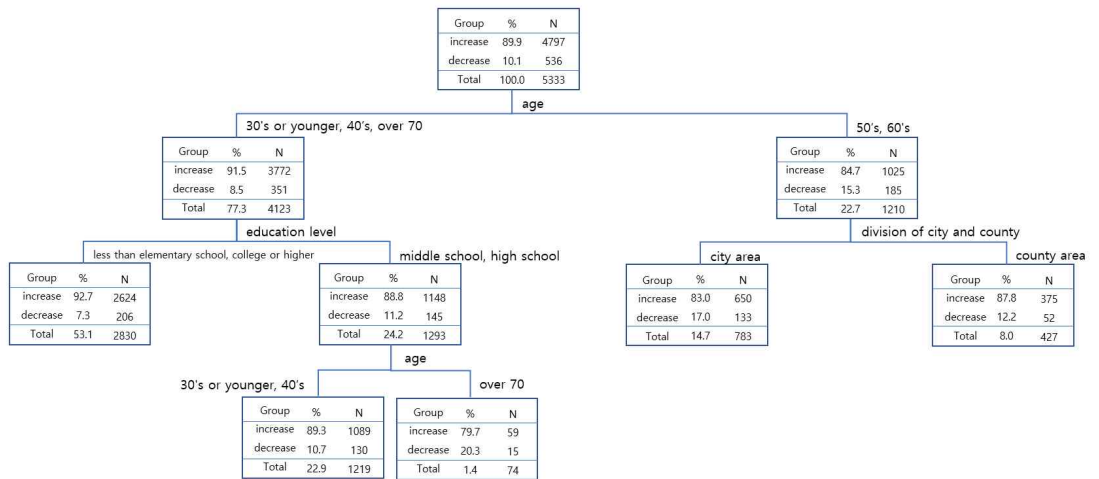


Figure 1. Decision tree analysis for indoor activity change

References

Gyeongnam (2021). *Social Survey of Gyeongsangnamdo in 2021*, Gyeongsangnamdo. (in Korean).
 Korea Disease Control and Prevention Agency. (2022). Coronavirus Disease-19 information-2022.01.23. [Internet]. [cited 2022 Jan. 23]. Available from: <https://www.kdca.go.kr/index.es?sid=a2>.
 World Health Organization (2022a). WHO Coronavirus(COVID-19) Dashboard. [Internet]. [cited 2022 Jan. 23]. Available from: <https://covid19.who.int/>

조건부 Quatember 확률화응답모형*

이기성¹, 홍기학², 손창균³, 박근화⁴, 홍성준⁵

요약

본 논문에서는 덜 민감한 속성과 강요질문으로 만들어진 확률장치로부터 “예”라고 답한 응답자들에게만 Quatember 모형을 사용하도록 하여 민감한 속성을 효율적으로 추정하는 조건부 Quatember 모형을 제안하여 민감한 속성에 대한 추정량과 분산을 도출하였다. 그리고 조건부 Quatember 모형에서 무관한 속성일 때 사용가능한 경우 두 개의 독립표본을 이용하는 조건부 Quatember 이표본 모형으로 발전시켰다.

주요용어 : 덜 민감한 속성, Quatember 모형, 조건부 확률화응답모형.

1. 서론

Warner(1965)는 개인적으로나 사회적으로 민감한 질문에 대한 정확한 정보를 얻기 위하여 직접 질문방식 대신에 확률장치를 이용하는 간접질문방식인 확률화응답모형을 제안하였다. Warner는 민감한 질문 그리고 민감한 질문과 배반되는 질문으로 만들어진 확률장치를 사용하였으며, 그 후 Loynes(1976)는 민감한 질문과 강요된 응답 “예”로 만들어진 확률장치를 이용하는 강요질문모형을 제안하였다. 그리고 Carr, Marascuilo.(1982)는 덜 민감한 속성과 강요질문으로 만들어진 확률장치로부터 “예”라고 답한 응답자들에게만 Loynes의 모형을 이용하여 민감한 속성의 모비율을 추정할 수 있는 조건부 확률화응답모형을 제안하였다.

본 논문에서는 덜 민감한 속성과 강요질문으로 만들어진 확률장치로부터 “예”라고 답한 응답자들에게만 Quatember 모형을 사용하도록 하여 민감한 속성의 추정 효율성을 높인 조건부 Quatember 모형을 제안하고자 한다. 그리고 조건부 Quatember 모형에서 무관한 속성일 때 사용가능한 경우 두 개의 독립표본을 이용하는 조건부 Quatember 이표본 모형으로 발전시키고자 한다.

2. 조건부 Quatember 확률화응답모형

*이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2018R1D1A3B07044007).

¹55338 전북 완주군 삼례읍 삼례로 443 우석대학교 아동사회복지학부 교수. E-mail: gisung@woosuk.ac.kr

²58245 전남 나주시 건재로 185 동신대학교 컴퓨터학과 교수. E-mail: khhong@dsu.ac.kr

³38066 경북 경주시 동대로 123 동국대학교 빅데이터·응용통계학과 부교수. E-mail: ckson85@dongguk.ac.kr

⁴07511 서울특별시 강서구 금남화로 154 한국문화관광연구원 정책정보센터 통계관리팀 선임전문위원.

E-mail: pkhkwen@kcti.re.kr

⁵13822 경기도 과천시 경마공원대로 107 마사회 말산업연구소 연구원. E-mail: hsj8129@naver.com

2.1. 조건부 Quatember 모형

이 절에서는 덜 민감한 속성 B 와 강요질문으로 만들어진 확률장치로부터 “예”라고 답한 응답자들만을 대상으로 Quatember(2009) 모형을 이용하여 민감한 속성에 대한 심층적인 정보 수집이 가능한 조건부 Quatember 확률화응답모형을 제안하고자 한다.

첫 번째 단계에서 n 명의 응답자들은 아래와 같은 Loynes의 모형의 확률장치로부터 선택된 질문에 대하여 응답한다.

질문 1 : 귀하께서는 덜 민감한 속성 B 를 소유하고 계십니까?

질문 2 : “예”라고 응답하세요.

여기서, 질문 1의 선택 확률은 T 이고, 질문 2의 선택 확률은 $1-T$ 이다. 이 때, 응답자들은 확률장치로부터 질문 1이 나오면 “예” 또는 “아니오”라고 응답하며, 질문 2가 나오면 “예”라고 응답하면 된다.

첫 번째 단계에서 응답자들이 “예”라고 응답할 확률 λ_1 은 다음과 같이 표현된다.

$$\lambda_1 = T\pi_1 + (1-T).$$

여기서, π_1 은 덜 민감한 속성 B 의 모비율이다.

이 때, n 명의 응답자들 중에서 “예”라고 답한 응답자들의 수를 n_1 명이라 두면, $\hat{\lambda}_1 = \frac{n_1}{n}$ 이 되어 π_1 의 추정량 $\hat{\pi}_1$ 은 아래와 같다.

$$\hat{\pi}_1 = \frac{1}{T}[\hat{\lambda}_1 - (1-T)] = \frac{1}{nT}[n_1 - (1-T)n].$$

두 번째 단계에서는 전 단계에서 “예”라고 답한 n_1 명의 응답자들만이 아래와 같은 Quatember 모형의 확률장치에 의해 선택된 질문에 대해 응답하게 한다.

질문 1 : 귀하께서는 민감한 속성 A 를 소유하고 계십니까?

질문 2 : 귀하께서는 민감한 속성 A 를 소유하고 계시지 않습니까?

질문 3 : 귀하께서는 무관한 속성 Y 를 소유하고 계십니까?

질문 4 : “예”라고 응답하세요.

질문 5 : “아니오”라고 응답하세요.

여기서, 질문 1이 선택될 확률은 p_1 , 질문 2가 선택될 확률은 p_2 , 질문 3이 선택될 확률은 p_3 , 질문 4가 선택될 확률은 p_4 , 질문 5가 선택될 확률은 p_5 이다($\sum_{j=1}^5 p_j = 1$). 이 때, 응답자들은 확률장치로부터 나온 질문에 대하여 “예”나 “아니오”라고 응답하면 된다.

따라서 응답자들이 첫 번째 단계에서 “예”라고 답했다는 조건하에 두 번째 단계에서 “예”라고 응답할 확률 λ_2 는 다음과 같이 표현된다.

$$\lambda_2 = (p_1 - p_2) \frac{\pi_2}{\lambda_1} + (p_2 + p_3\pi_y + p_4).$$

여기서, π_2 는 민감한 속성 A 의 모비율이며, π_y 는 무관한 속성 Y 의 모비율로 아는 값으로 가정한다.

n_1 명의 응답자들 중에서 “예”라고 답한 응답자들의 수를 n_2 라 두면 $\hat{\lambda}_2 = \frac{n_2}{n_1}$ 가 되어 모비율 π_2 의 추정량 $\hat{\pi}_2$ 는 아래와 같다.

$$\hat{\pi}_2 = \frac{1}{n(p_1 - p_2)} [n_2 - (p_2 + p_3\pi_y + p_4)n_1], \quad (p_1 \neq p_2).$$

$\hat{\pi}_2$ 는 π_2 의 비편향추정량이 되며, $\hat{\pi}_2$ 의 분산은 아래와 같다.

$$V(\hat{\pi}_2) = \frac{1}{n} \left[\pi_2(1 - \pi_2) + \frac{\pi_2(1 - p_1 - p_2 - 2p_3\pi_y - 2p_4)}{p_1 - p_2} + \frac{\{T\pi_1 + (1 - T)\}(p_2 + p_3\pi_y + p_4)(1 - p_2 - p_3\pi_y - p_4)}{(p_1 - p_2)^2} \right].$$

2.2. 조건부 Quatember 이표본 모형

이 절에서는 조건부 Quatember 모형을 π_y 가 미지일 때 두 개의 표본으로 민감한 속성의 모비율 π_2 를 추정하는 조건부 Quatember 이표본 모형으로 확장해 보기로 하자.

조건부 Quatember 모형의 두 번째 단계에서 무관한 속성 Y 의 모비율 π_y 를 알지 못하는 경우에는 서로 다른 두 독립표본이 필요하며, 첫 번째 단계에서 “예”라고 답한 n_1 명의 응답자들로부터 단순임의복원추출된 서로 독립인 n_{1i} 명의 응답자들은 조건부 Quatember 모형에서 사용한 확률장치로부터 선택된 질문에 대하여 “예” 또는 “아니오”라고 응답하게 된다. 이 때, 질문 1이 선택될 확률은 p_{i1} , 질문 2가 선택될 확률은 p_{i2} , 질문 3이 선택될 확률은 p_{i3} , 질문 4가 선택될 확률은 p_{i4} , 질문 5가 선택될 확률은 p_{i5} 이다($\sum_{j=1}^5 p_{ij} = 1, i = 1, 2$).

따라서 응답자들이 전 단계에서 “예”라고 답했다는 조건하에서 두 번째 단계의 n_{1i} 명의 응답자들이 “예”라고 응답할 확률 λ_{2i} ($i = 1, 2$)는 아래와 같다.

$$\lambda_{2i} = (p_{i1} - p_{i2}) \frac{\pi_2}{\lambda_1} + (p_{i2} + p_{i3}\pi_y + p_{i4}).$$

n_{1i} 명의 응답자들 중에서 “예”라고 답한 응답자들의 수를 n_{2i} 라 두면 $\hat{\lambda}_{2i} = \frac{n_{2i}}{n_{1i}}$ 가 되어 모비율 π_2 의 추정량 $\hat{\pi}_2$ 는 아래와 같다.

$$\begin{aligned} \hat{\pi}_2 &= \frac{\hat{\lambda}_1 [p_{23}(\hat{\lambda}_{21} - p_{12} - p_{14}) - p_{13}(\hat{\lambda}_{22} - p_{22} - p_{24})]}{p_{23}(p_{11} - p_{12}) - p_{13}(p_{21} - p_{22})}, \\ &= \frac{p_{23} \frac{n_1 n_{21}}{n_{11}} - p_{23}(p_{12} + p_{14})n_1 - p_{13} \frac{n_1 n_{22}}{n_{12}} + p_{13}(p_{22} + p_{24})n_1}{n \{p_{23}(p_{11} - p_{12}) - p_{13}(p_{21} - p_{22})\}}, \quad p_{23}(p_{11} - p_{12}) \neq p_{13}(p_{21} - p_{22}). \end{aligned}$$

$\hat{\pi}_2$ 는 π 의 비편향추정량이 되며, $\hat{\pi}_2$ 의 분산은 아래와 같다.

$$V(\hat{\pi}_2) = \frac{1}{n \{p_{23}(p_{11} - p_{12}) - p_{13}(p_{21} - p_{22})\}^2} \left[\frac{p_{23}^2 \alpha}{n_{11}} + \frac{p_{13}^2 \beta}{n_{12}} \right],$$

여기서, $\alpha = \lambda_{21}(1 - \lambda_{21})\{\lambda_1(1 - \lambda_1) + n\lambda_1^2\} + n_{11}\lambda_1(1 - \lambda_1)[\lambda_{21}^2 + (p_{12} + p_{14})\{(p_{12} + p_{14}) - 2\lambda_{21}\}]$,

$$\beta = \lambda_{22}(1 - \lambda_{22})\{\lambda_1(1 - \lambda_1) + n\lambda_1^2\} + n_{12}\lambda_1(1 - \lambda_1)[\lambda_{22}^2 + (p_{22} + p_{24})\{(p_{22} + p_{24}) - 2\lambda_{22}\}]$$

이다.

3. 결론

본 논문에서는 덜 민감한 속성과 강요질문으로 만들어진 확률장치로부터 “예”라고 답한 응답자들에게만 Quatember 모형을 사용하도록 하여 민감한 속성의 효율적 추정이 가능한 조건부 Quatember 모형을 제안하였다. 그리고 조건부 Quatember 모형에서 무관한 속성일 때 사용가능한 경우 두 개의 독립표본을 이용하는 조건부 Quatember 이표본 모형으로 발전시켰다.

Reference

- Carr, J. W., Marascuilo, L. A. (1982). Optimal randomized response models and methods for hypothesis testing, *Journal of Educational Statistics*, 7, 295-310.
- Loynes, R. M. (1976). Asymptotically optimal randomized response procedures, *Journal of the American Statistical Association*, 71, 924-928.
- Quatember, A. (2009). A standardization of randomized response strategies, *Survey Methodology*, 35(2), 143-152.
- Warner, S. L. (1965). Randomized response ; A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.

Multinationality, Portfolio Coordination, and Downside Risk

*Sangcheol Song*¹

Abstract

We take the multinational operational flexibility perspective anchored in the real options theory and examine how multinational corporations' coordinative factors at the portfolio level affect downside risks of multinationality. We extend the Tong and Reuer's (2007) study of the curvilinear relationship of multinationality to downside risks by testing downside risk reduction effects of those portfolio factors that facilitates effective coordination over foreign subsidiaries. By employing the Tobit two-stage model on a large sample of Korean multinational corporations, we also find a U-shaped relationship between multinationality and downside risks. We present a new finding that a broader existence of intra-firm trade, key expatriates, or regional headquarters among foreign subsidiaries curbs downside risk, shifting the minimum in the U-shaped curve to a higher multinationality level. These findings imply that a well-crafted configuration and coordination of dispersed operations across countries is essential for retention and actualization of the real options value of multinationality.

1. Research Methodology

For empirical testing, we utilized a panel dataset of 345 Korean MNCs that were publicly listed and active for at least five years on the Korea Stock Exchange during 1985 to 2014 and had presence in 71 foreign countries. We followed the methods employed in Tong and Reuer's (2007) study for downside risks as the dependent variable of this study. These studies commonly measured downside risks of multinationality based on the difference in annual return on assets (ROA) between an MNC and its reference group (i.e., other MNCs in the same two-digit industry). The squared difference term between industry average ROA (IROA) and a firm's ROA was summed in the form of a second-order root lower partial moment over past five years (see formula below). By the definition of downside risks as the indicator of real option performance, only below-target performance is counted in a Tobit model.

Multinationality: We calculated the logarithm of 1 plus the number of countries in which an MNC had its foreign subsidiaries in order to remedy the significant positive skew that was evident for the pre-transformed count measure.

¹Distinguished professor of China Technology & Business University, China & Associate professor of Saint Joseph's University, USA. E-mail: ssong@sju.edu

*Downside Risk*_{t=0}

$$= \sqrt{\frac{1}{5} \sum_{t=0}^{t=-4} (\text{IROA}_{t-1} - \text{ROA}_t)^2 | \text{IROA}_{t-1} > \text{ROA}_t}$$

Intra-firm trade: We measured this variable at the MNC portfolio level. We computed the ratio of the number of foreign manufacturing subsidiaries with intra-firm trade with other sister companies (mainly with headquarters) to the total number of foreign manufacturing subsidiaries. For each firm's intra-firm trade information, we referred to audit reports from the database of the TS2000 Warehouse and the Data Analysis, Retrieval, and Transfer (DART) System.

Key expatriates: We computed the ratio of the number of foreign subsidiaries in which subsidiaries had home national boards to the total number of host countries. A higher ratio indicates a broader existence of key expatriates across host countries.

Regional Headquarters: We measured this variable as a ratio of the number of regional headquarters (RHQs) to the total number of foreign subsidiaries. A higher ratio implies more RHQs in the MNC network. For the information on RHQs, we referred to the same sources as key expatriates.

2. Results

Table 1. Estimates for the Tobit models with selectivity

Models	Model 1	Model 2	Model 3	Model 4(H1)	Model 5(H2)	Model 6(H3)	Model 7(All)
Intercept	0.051 (0.118)	0.072 (0.114)	0.101 (0.112)	0.143 (0.114)	0.124 (0.118)	0.150 (0.115)	0.201 (0.115)
Firm size	-0.213 (0.047)	-0.281 (0.047)	-0.260 (0.044)	-0.244 (0.041)	-0.263 (0.045)	-0.245 (0.046)	-0.281 (0.041)
Firm performance	-0.565 (0.000)	-0.542 (0.000)	-0.563 (0.000)	-0.546 (0.000)	-0.563 (0.000)	-0.583 (0.000)	-0.562 (0.000)
R&D intensity	-0.144 (0.052)	-0.162 (0.051)	-0.164 (0.048)	-0.164 (0.049)	-0.186 (0.049)	-0.182 (0.046)	-0.185 (0.046)
Organizational slack	-0.085 (0.069)	-0.088 (0.069)	-0.107 (0.065)	-0.104 (0.065)	-0.102 (0.065)	-0.122 (0.064)	-0.108 (0.062)
International experience	-0.061 (0.090)	-0.064 (0.094)	-0.072 (0.094)	-0.082 (0.086)	-0.073 (0.089)	-0.063 (0.089)	-0.094 (0.802)
Portfolio ownership stake	-0.104 (0.120)	-0.100 (0.114)	-0.106 (0.116)	-0.104 (0.113)	-0.108 (0.112)	-0.112 (0.115)	-0.120(0)
GDP/capita range	0.055 (0.126)	0.052 (0.122)	0.052 (0.126)	0.067 (0.120)	0.085 (0.118)	0.044 (0.114)	0.064 (0.106)
Low labor cost correlation	-0.126 (0.044)	-0.144 (0.040)	-0.155 (0.046)	-0.188 (0.042)	-0.165 (0.040)	-0.209 (0.036)	-0.225 (0.036)

Table 1. Estimates for the Tobit models with selectivity(continued)

Models	Model 1	Model 2	Model 3	Model 4(H1)	Model 5(H2)	Model 6(H3)	Model 7(All)
Firm dummies	Included	Included	Included	Included	Included	Included	Included
Year dummies	Included	Included	Included	Included	Included	Included	Included
Industry dummies	Included	Included	Included	Included	Included	Included	Included
Multinationality		-0.271 (0.000)	-0.236 (0.000)	-0.252 (0.000)	-0.279 (0.000)	-0.260 (0.000)	-0.259 (0.000)
Multinationality(squared)		-	0.160 (0.006)	0.178 (0.006)	0.142 (0.008)	0.138 (0.008)	0.189 (0.006)
(1) Intra-firm trade		-3.960 (0.006)	-3.840 (0.006)	-3.769 (0.007)	-3.910 (0.007)	-3.879 (0.008)	-3.819 (0.008)
(2) Key expatriates		-1.129 (0.051)	-1.154 (0.048)	-1.135 (0.044)	-1.118 (0.051)	-1.139 (0.046)	-1.146 (0.044)
(3) Regional headquarters		-2.212 (0.134)	-2.260 (0.130)	-2.183 (0.136)	-2.244 (0.138)	-2.308 (0.134)	-2.329 (0.138)
Multinationality * (1)				-3.433 (0.000)			-3.538 (0.000)
Multinationality(squared) * (1)				1.342 (0.008)			1.442 (0.007)
Multinationality * (2)					-2.667 (0.000)		-2.814 (0.000)
Multinationality(squared) * (2)					1.148 (0.007)		1.246 (0.008)
Multinationality * (3)						-2.541 (0.009)	-2.334 (0.009)
Multinationality(squared) * (3)						1.461 (0.006)	1.742 (0.006)
Correction for self-selection	-0.140 (0.042)	-0.129 (0.040)	-0.130 (0.046)	-0.119 (0.048)	-0.126 (0.042)	-0.130 (0.042)	-0.142 (0.038)
Wald c2	142.45 (0.000)	156.34 (0.000)	163.67 (0.000)	180.44 (0.000)	176.25 (0.000)	174.90 (0.000)	186.84 (0.000)

p-values in parenthesis

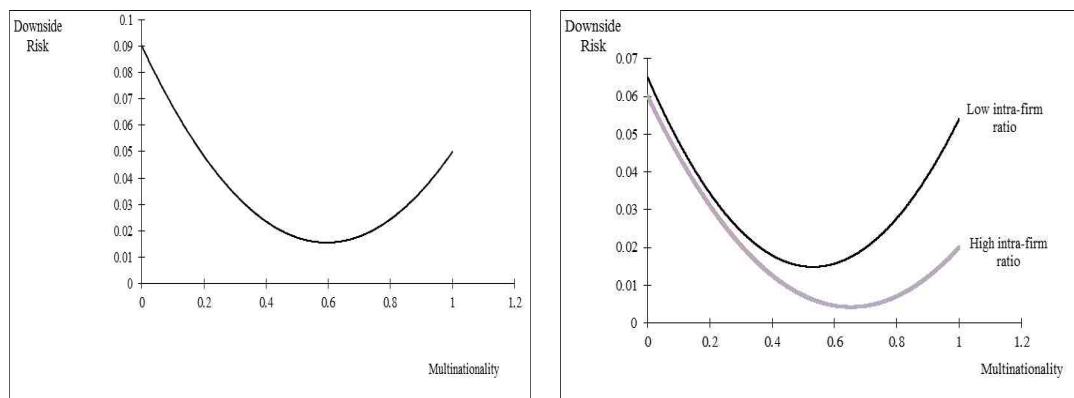


Figure1. Graphic presentation of non-linear relationships

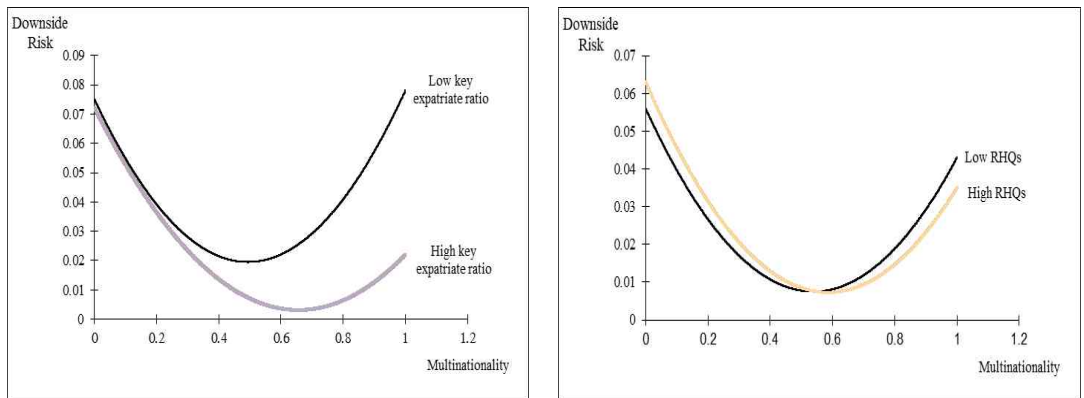


Figure 1. Graphic presentation of non-linear relationships

3. Discussion

Our finding regarding the non-monotonic relationship between multinationality and downside risks indicates the coexistence of the positive and negative effects of multinationality at differing levels. Unlike the traditional multinational operational flexibility perspective (MOF) aligned with real options theory (ROT) that argues for upside potentials and limited downside risks, this finding implies that excessive multinationality can adversely affect firm performance. This adverse impact of excessive multinationality is supported in some real options studies that tried to specify some organizational conditions for higher real option value. For example, Song et al. (2015) tackle the problematic assumption that the traditional MOF perspective is based on. Considering that MOF can be retained and actualized with appropriate control and coordination of scattered operations across countries, they argue for the positive impacts of intra-firm trade connections and controlling ownership stakes on the real options value of multinationality or breadth of FDI.

This study is based on the Tong and Reuer's (2007) study and is aligned with other real option studies. However, it makes new contributions to the literature on multinational operational flexibility anchored in real options theory. Compared to the studies that mainly focus on upside potentials as real options value as reflected in Tobin's q for instance, this study focused on the other aspect of real options value, that is, downside risk reduction. Examinations of downside risk reductions of real option investments should be done further for more complete understanding of the real options value of an investment. While the Tong and Reuer's study examined only the pure effects of multinationality (curvilinear: negative to positive), cultural distance (positive), and portfolio ownership (insignificant) on downside risks, this study investigates the interactive relationships between multinationality and four organizational coordinative factors to assess how those factors moderate the non-linear relationship of multinationality to downside risks.

References

- Belderbos, R., Tong, T. W., & Wu, S. 2014. Multinationality and downside risk: The roles of option portfolio and organization. *Strategic Management Journal*, 35, 88 - 106.
- Song, S., Lee, S-H., & Makhija, M. 2015. Operational hedging in foreign direct investment under volatile and divergent exchange rates across countries. *Journal of World Business*, 50(3), 548-588.
- Tong, T., & Reuer, J. J. 2007. Switching options and coordination costs in multinational firms. *Journal of International Business Studies*, 38, 215 - 230.

외식업 O2O 배달 앱 품질이 소비자의 이용의도에 미치는 영향: TAM과 USM 통합모델

이승환¹

요 약

본 연구에서는 정보통신부문에 대한 연구로서 외식업체 O2O 배달 앱 사용을 대상으로 사용자만족 중심 연구방향과 기술수용모델을 중심으로 하는 연구방향을 통합한 모델을 수립하고 실증분석을 수행하였다. 외식업체를 위한 O2O 배달앱, 실증분석을 위하여 O2O 배달 앱을 이용한 경험이 있는 지인 및 일반인을 대상으로 설문조사가 진행되었고, 최종 324개의 답변을 받아 분석하였다. 분석 방법으로는 SPSS와 AMOS를 사용하였으며, 분석 결과 모델의 적합도가 적절한 수준에 있는 것으로 나타났다.

주요용어 : 사용자 만족, 기술수용, O2O, 외식업체, 배달 앱.

1. 서론

1.1. 연구의 필요성 및 목적

우리나라의 스마트 폰 보급률이 2021년 기준 95%에 달하고 있으며, 휴대전화 보급률은 거의 100%에 달해 세계 1위를 기록하고 있다(Gallup Korea, 2021). 스마트 폰의 사용용도로 문자나 메시지를 주고받기 및 통화 기능이 가장 중요하지만, 쇼핑이나 검색 기능 역시 커다란 역할을 담당하고 있다. 이러한 기능들이 음식주문 및 배달에 일조하고 있는 것은 분명하다고 할 수 있다. 이러한 O2O 비즈니스는 앞으로도 발전과 성장 여지가 크며(Kim, 2016), 따라서 O2O 비즈니스 이용자의 이용의도에 대한 연구는 일반 소비자는 물론 O2O 사업 관계자, 관리자 뿐 아니라 관련 관계기관에게도 시사하는 점이 적지 않을 것이라 생각된다.

정보통신 관련 연구는 4차 산업의 급속한 성장에 따라 활발하게 수행되어 왔다. 정보통신 관련 투자에 따른 효과 관련 연구는 다양한 방향으로 수행되어 왔지만 연구의 큰 흐름은 크게 사용자만족을 중심으로 하는 방향과 기술수용을 중심으로 하는 2가지로 요약할 수 있다(Wixom & Todd, 2005). 그러나 이들 연구방향들이 정보통신 관련 이해를 돕는데 커다란 기여를 하여 온 것이 사실이나, 2가지 연구흐름 모두 부분적 설명을 하게 된다는 한계를 안고 있다. 정보통신 관련 범주로 볼 수 있는 O2O 비즈니스는 기존 연구들과는 다른 차원에서 접근할 필요가 있다고 본다.

그것은 첫째, O2O 서비스 앱 사용은 조직내 업무수행과는 거리가 있으며, 스마트 폰 베이스로서 개별 소비자의 취향이 반영된 자발적 의사결정의 문제이고, 둘째, 서비스앱 제공자는 시스템품

¹10540 경기도 고양시 덕양구 항공대학로 76, 한국항공대학교 경영학과 박사수료. E-mail: sydlsh@naver.com

질 등 기술적 측면 뿐 아니라 제공되는 서비스 측면 또한 고려의 대상으로 삼아야 한다는 점이다. 본 연구에서는 이러한 O2O 비즈니스 특성들을 반영하여 포괄적 설명이 가능한 모델을 위한 접근을 시도해 보고자 하였다.

2. 연구방법

2.1. 자료수집 및 분석방법

본 연구의 실증분석을 위해 O2O 배달 앱을 이용해 본 경험이 있는 연구자 주변 지인 및 일반인을 대상으로 오프라인 및 온라인을 통한 서베이 방식인 구글 설문지 조사를 실시하였다. 설문조사는 총 326명의 응답자를 대상으로 데이터를 수집하였으며, 이 중 불성실하게 답변한 2인의 데이터를 제외하였다. 이에 따라 총 324명의 응답데이터가 본 연구의 분석을 위해 사용이 되었다. 수집된 자료를 분석하기 위해 SPSS 및 AMOS을 사용하였다.

3. 연구결과

응답자 324명 중에 남성과 여성은 각각 182명(56.2%), 142명(43.8%)를 보이고 있다. 연령은 20대 이하가 52명(16%), 20-30대가 81명(25%), 30-40대가 79명(24.4%), 40-50대가 75명(23.1%), 50-60대가 17명(5.2%), 그리고 60대 이상이 20명으로 6.2%로 나타나고 있으며, 타당성 및 신뢰성 분석 결과를 살펴보면 요인분석 관련 공통성은 0.4미만, 요인적재치는 1을 기준으로 하였으며, 신뢰도는 모두 0.7 이상을 보이고 있다. CR(Critical Ratio)값은 모두 유의하고, 평균분산추출(Average Variance Extracted) 값 역시 모두 0.5 이상으로 신뢰도에 문제가 없는 것으로 판단하였다.

본 연구에서는 정보통신부문에 대한 연구로서 외식업체 O2O 배달 앱 사용을 대상으로 사용자만족 중심 연구방향과 기술수용모델을 중심으로 하는 연구방향을 통합한 모델을 수립하고 실증분석을 수행하였으며 가설 검증 결과는 가설 일부는 사회과학 분야에서 일반적으로 받아들여지고 있는 유의수준인 0.05를 넘는 수치를 보임으로서 기각판정을 하고, 나머지 가설들은 모두 채택되고 있음을 알 수 있었다. 다만, 유의수준을 상회하는 정도가 매우 낮아 유의수준에 대한 기준을 0.1로 확장한다면 유의하다고 판정할 수도 있는 수준이기도 하다.

O2O 배달 앱 서비스 이용 관련 정보품질이 지각된 용이성에 유의한 영향을 미치지 않는다는 결과와 서비스품질이 유용성에 유의한 영향을 미치지 않는다는 결과가 도출되고 있다. 이러한 결과는 다음과 같이 해석할 수 있다. 첫째, O2O 배달 앱 사용자들은 대부분의 경우 특정 앱에 회원으로 가입되어 있어서 자의반 타의반으로 회원이 된 앱을 지속적으로 사용하고 있을 가능성이 커서 다른 앱의 정보품질과 비교를 소홀히 할 수 있으며, 둘째, 음식배달 주문의 경우 특정 음식점을 염두에 두고 있을 가능성이 있을 수 있다. 셋째, 배달 앱 서비스의 경우 일반 상거래의 경우와 달리 본연의 목적인 음식의 질에 대한 비중을 가장 크게 두는 경향이 있을 것이며, 신속한 배달을 중요하게 생각하는 경향이 있어 부수적인 서비스에는 상대적으로 큰 비중을 두지 않을 수 있다는 결과가 도출되었다.

한편 수립된 모델의 적합도에 대한 검증결과를 살펴보면, Wheaton et al.,(1977)는 새로운 적합도

지수 CMIN/DF를 개발하였다. 이들은 이 값이 5 이하이면 적당하다고 하였지만 그 후로 1에서 2, 또는 1에서 3사이가 적당하다는 주장이 설득력을 얻고 있기 때문에(Byrne, 1989; McIver & Carmines, 1981) 기준값에 대한 공감대는 2 또는 3 이하로 보아야 할 것 이므로, 본 연구의 2. 4는 3 이하 이므로 사회과학 분야에서는 이 또한 적합하다 라고 판단하였다.

4. 결론

본 연구를 통하여 기술수용모델의 가정인 태도가 이용의도에 영향을 미치며, Davis et al.(1989)의 모형에서 합리적 행위이론과 차이를 보이는 유용성이 이용의도에 미치는 영향 역시 Davis et al.(1989)의 모형이 지지를 받는다는 실증검사 결과로 제시될 수 있다. 또한 이용자만족 중심 접근에서 제시하는 정보품질, 시스템품질, 서비스품질은 모두 지각된 유용성과 지각된 용이성의 매개를 통하여 태도에 유의한 영향을 미친다는 결과를 증명하고 있다.

위 연구 결과에서도 나타났듯이 실무적 시사점을 살펴보면, 첫 번째, 가설 1-1과 3-1을 통해 정보품질과 서비스 품질이 지각된 유용성에 긍정적 영향을 미치는 것이 확인되었다 이는 O2O 배달 앱 정보가 어렵거나 또는 서비스가 불안정적이고 정확하게 작동하지 않아 지각된 유용성이 떨어진다고 일반소비자가 인식할 경우 O2O 서비스 사용 전으로 돌아갈 가능성이 높다고 보여진다(sim. 2016). 두 번째는, 지각된 유용성과 용이성은 이용태도에 긍정적인 영향을 미치는 것으로 확인되었는데, 배달 앱 사용에 있어서 사용하기 간결하거나 효과적인 주문 및 결제가 가능한 시스템에서는 일반소비자가 배달 앱을 사용하는데 있어서 적극적으로 사용할 수 있는 환경이 주어지는 것으로 나타났다. 이를 바탕으로 배달 앱을 이용하여 이익을 창출하는 O2O 사업 관계자와 음식점주는 소비자의 이용의도에 있어 기술수용 측면과 사용자의 정서적 측면인 사용만족 차원을 보편화하여야 수익에 직결될 것임을 입증하였다.

본 연구의 한계점 및 연구 방향은 다음과 같이 정리해 볼 수 있다. 첫 번째는, 본 연구에서는 기존 선행 연구를 바탕으로 하였지만 요인특성을 모두 포함하고 있지는 않고 있다. 앞으로의 연구에서는 변수를 조절하여 연구할 필요가 있다. 두 번째는, 이용의도에 미치는 요인은 독립변수, 매개변수 이외에도 여러 가지 요인들이 있을 수 있다. 본 연구에서 제시하지 않은 요인들이 외식업체 O2O 배달 앱 서비스 품질이 이용의도에 어떤 영향을 미치는지 확인할 수 없기에 본 연구에서 제시한 변수 외에도 다양한 변수들을 추가하여 연구할 필요가 있으며 세 번째, 설문 조사시 표본 대상이 20대에서 50대에게 편중되어 있어 일반화 시키는데 문제가 있을 수 있다. 추후 연구에서는 이와 같은 미진한 부분이 보완되어 보다 정교한 연구가 진행될 필요성이 제기된다.

Dynamic Portfolio Using Association Rules

*Hyeon-Jong Jung*¹

Abstract

We provide dynamic portfolio strategy capitalizing on the time-varying association rules. We set the moving observation window and employed changes in the number of negative rules as a signal for portfolio weights. The back-testing result shows that increasing number of negative rules reflects newly formed negative linkages between Korea stock market and related variables. The result suggests that as a proxy for systemic risk measure, signal from association rules can be an effective tool for portfolio risk management.

Keywords : association rules, equity investment, portfolio risk management.

1. Introduction

Association rule or association rule mining is an unsupervised learning method to detect and extract useful information from large transaction data. The objective of association rule is to find meaningful association rules by analyzing relationship between set of items in transaction. Association rule stems from data mining and is called market basket analysis as it analyzes consumer behaviors in retail market. As the number of items in transaction grows, the number of set of items(itemsets) increases exponentially. And the number of possible association rules which describes the relationship between itemsets even grow faster. For example, with 10 items, we can generate 1,023 itemsets excluding null itemset and over 50,000 association rules. Thus, it is infeasible to evaluate every rules generated by association rule analysis. Effective association rule discovery was first proposed by Agrawal, Imielinski, Swami(1993) and association rules in database has received much attention recently(Wu, Zhang, Zhang, 2004; Lim, Lee, Cho, 2010).

In finance filed, existing researches mainly focused on investigating relationship between stock markets or industries using association rules. And in relatively few literature, association rules are employed to forecast asset returns(Paranjape-Voditel, Deshpande, 2013; Na, Sohn, 2011). In this paper, we try to forecast stock market returns using association rule analysis. Also, we provide dynamic portfolio strategy capitalizing on the time-varying association rules.

¹Korea Investment & Securities Co. Ltd. 88, Uisadang-daero, Yeongdeungpo-gu, Seoul, Korea 133-791,
E-mail: hyeonjong.jung@gmail.com

The rest of the paper is organized as follows: Section 2 presents data and methodology. Section 3 summarizes the test results. Section 4 discusses key findings and provides conclusions.

2. Data and methodology

In association rules, let $I=\{i_1,i_2,\dots,i_n\}$ be a set of n items and let $D=\{t_1,t_2,\dots,t_m\}$ be a set of m transactions. Each transaction in D contains a subset of the items in I . An association rule is defined as the form $X\rightarrow Y$ where $X,Y\subset I$ and $X\cap Y=\emptyset$. X and Y are called the antecedent and the consequent, respectively. Association rule analysis uses measures of support, confidence and lift to evaluate effectiveness of the rule. Support of the rule $X\rightarrow Y$ is the proportion of the transactions containing $X\cup Y$ in the set of transactions D . Confidence is an estimate of conditional probability, $P(Y|X)$ that is the probability of observing Y given X . Support measures the frequency of the occurring rules and confidence implies the strength of the rules. Usually interesting association rules are selected based on the minimum support and the minimum confidence. If we set too low minimum threshold, resulting rules will include accidental rules. On the other hand, too high threshold will result in obvious but non-interesting rules. Lastly, lift is defined as the support of the rule $X\rightarrow Y$ divided by support of X and Y , respectively. Lift is interpreted as the deviation of the support of the rule from the support expected under independence given the support of both sides.

In an empirical test, time-series data are from Factset and Quantiwise. The data are collected for the time period January 2000 to May 2022. We used weekly frequency data to construct investable portfolio. In the rule $X\rightarrow Y$, Y is the sign of Korea Composite Stock Price Index(KOSPI) return. X , the variables or itemsets used in forecasting the sign of KOSPI return include the growth rate of KRW/USD, VIX, WTI futures, S&P 500 Index, US 10 year treasury rate and US treasury spread(10 year - 2 year). We discretize the antecedents into binary variables, U(upward) or D(downward) depending on the sign of the growth rate, making total number of antecedents 12. All antecedents are 1-week lagged variables to forecast the sign of KOSPI returns. Then, association rules which meet the minimum support and confidence threshold are selected using the apriori algorithm(Agrawal, Srikant, 1994). With appropriate criteria to select the rules, association rule analysis would find out frequent itemsets and strong relationship between them. We set the minimum support of 5% and minimum confidence of 50% in the test.

3. Test Results

For the test period, number of resulting association rules which met the criteria is 251. The number of rules with lift above 1 further diminishes to 174. We call rules of the form $X\rightarrow Y$ positive rules when Y is KOSPI(U) and negative rules when Y is KOSPI (D). It is worth noting that the lift of the negative rules is higher with average of 1.18 than the lift of the positive

rules(1.02). The results imply that association rule is stronger in down market than in up market. Maximum number of variables in X was 5. The most frequent variables in X include S&P 500(U), WTI(U), VIX(D) and US treasury rate(D). Frequent association rules reflect dynamics of market movement with time lag. For example, the rule forecasting decrease in KOSPI return with the highest lift is associated with S&P 500, WTI and US treasury spread. $\{S\&P\ 500(D), WTI(U), SPREAD(U)\} \rightarrow \{KOSPI(D)\}$. The rule shows that Korea stock market is negatively influenced by bearish US stock market. Also, it suggests that increasing oil prices and downward business cycle reflected in flattening US treasury yield curve might put pressure in KOSPI. Confidence of this rule is 0.58, which means conditional on the left hand side of the rule, KOSPI declines with probability of 58% in the following week.

To test the forecasting ability of association rules, we construct dynamic portfolio which invests in Korea equity and fixed income asset simply based on the number of rules. In dynamic portfolio strategy, we select association rules with minimum support of 5%, minimum confidence of 50% and lift measure above 1. To identify changes in direction and strength of the rules, we used moving observation windows. Small window has an advantage of providing swift signal at the expense of increasing noise. We count numbers of negative rules and normalize them using min-max normalization. Weight of equity ($W_{t,E}$) and fixed income ($W_{t,B}$) in the portfolio at time t , varies with the normalized number of negative rules, $N_{t-1, neg} \in [0,1]$ as follows,

$$W_{t,E} = 0.6 + S(2 \times N_{t-1, neg} - 1) \quad (1)$$

$$W_{t,B} = 1 - W_{t,E} \quad (2)$$

where S is the span of the portfolio weight and is set 0.2, which means weight of equity asset in the portfolio dynamically changes from 40% to 80% according to the normalized number of rules. In back-testing, we used 7 years of moving window and weekly time-series data since January 2000. Benchmark portfolio comprises of 60% of equity and 40% of fixed income assets. We used total return index of KOSPI and KIS bond aggregate index as a proxy for equity and fixed income assets. Table 1 compares the performances of portfolio. Figure 1 shows the normalized number of negative rules and weights of assets in dynamic portfolio.

Table 1. Summary of portfolio performances

	KOSPI	Benchmark	Portfolio
Annualized return (%)	7.8	6.1	7.0
Annualized risk (%)	20.0	11.9	12.4
Return/Risk	0.39	0.51	0.57
Skewness	-0.62	-0.61	-0.48
Maximum drawdown (%)	-52.9	-34.2	-33.2
Value at Risk (95%)	-3.91	-2.35	-2.54

Benchmark is the static portfolio with 60% of equity and 40% of fixed income. Portfolio is the dynamic portfolio rebalanced weekly according to equation (1) and (2). Value at risk is at the 95% confidence level.

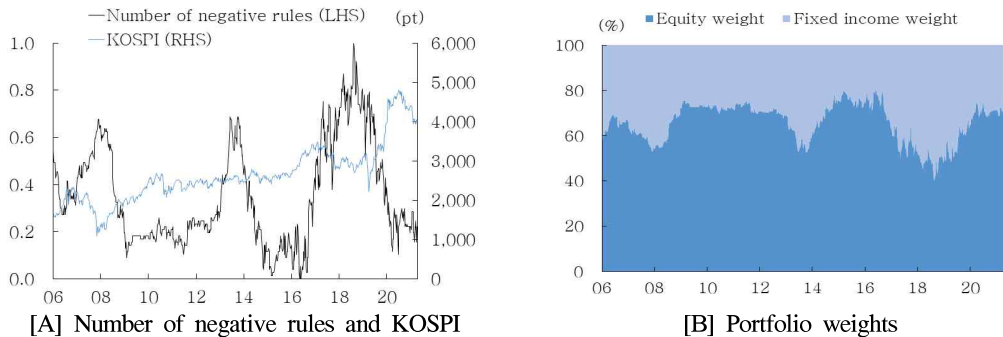


Figure 1. Number of rules and weights in dynamic portfolio

4. Conclusions

We provide dynamic portfolio strategy capitalizing on the time-varying association rules. Finding obvious association rule such as “If US stock market declines, Korea stock market follows suit” is not useful in portfolio management. To identify time-varying frequencies and strength of the rules, we set the observation window and employed changes in number of negative rules as a signal for portfolio weights. Equity weights in the dynamic portfolio decrease in market downturn such as sub-prime mortgage crisis in 2008, chinese stock market turbulence in 2015 and COVID-19 recession in 2020. Dynamic portfolio using association rules achieves higher risk-adjusted return of 0.57 compared to static benchmark portfolio(0.51). Annual excess return is 0.9%p which is statistically significant with p-value of 0.04. The back-testing result shows that increasing number of negative rules reflects newly formed negative linkages between Korea stock market and related variables. The result implies that as a proxy for systemic risk measure, signal from association rules can be an effective tool for portfolio risk management.

References

- Agrawal, R., Imielinski, R., Swami, A. (1993). Mining association rules between sets of items in large databases, *Proceedings of the ACM SIGMOD Conference on Management of Data*, 207-216.
- Agrawal, R., Srikant, R. (1994). Fast algorithms for mining association rules. *In Proc. 20th int. conf. very large data bases, VLDB*, Vol(1215), 487-499.
- Lim, J. S., lee. K. J., Cho. Y. S. (2010). A study of association rule by considering the frequency, *Journal of the Korean Data & Information Science Society*, 21(6), 1061-1069. (in Korean)
- Na, S. H., Sohn, S. Y. (2011). Forecasting changes in Korea composite stock price index (KOSPI) using association rules. *Expert Systems with Applications*, 38(7), 9046-9049.
- Paranjape-Voditel, P., Deshpande, U. (2013). A stock market portfolio recommender system based on association rule mining. *Applied Soft Computing*, 13(2), 1055-1063.
- Wu, X., Zhang, C., Zhang, S. (2004). Efficient mining of both positive and negative association rules. *ACM Transactions on Information Systems*, 22(3), 381-405.

재정데이터 내의 이상치 탐지를 위한 비지도 방법의 성능 비교

소희수¹, 박승리²

요약

이상치 탐지란 정상 데이터의 분포 밖에 있는 이상치를 탐지하는 것으로, 지도, 준지도 그리고 비지도 방법이 존재하며, 이중 비지도 방법은 레이블링 된 데이터 확보가 어려운 다양한 분야에 적용될 수 있는 방법이다. 최근 우리은행 등 내부자 부정행위가 사회적 이슈로 대두되며 내부통제 시스템에 활용될 수 있는 재정 데이터에 대한 이상치 탐지 분야의 수요가 가파르게 상승하고 있는 반면, 재정 데이터의 비밀성, 비공개성 및 레이블 확보의 어려움으로 관련 연구는 부족한 상황이다. 이에 본 논문은 2001년 말 부정회계 사건 등으로 파산한 미국 Enron사의 실전 재정 데이터를 활용하여 이전 연구에서 다뤄졌던 비지도 이상치 탐지 방법들과 본 논문에서 제안하는 Feature Relevance based Unsupervised Feature Selection(FRUFs)와 앙상블 기법을 결합한 방법의 결과를 비교분석 하였다. 실험에는 기존의 비지도 이상치 탐지 방법들로 Local Outlier Factor(LOF,) Isolation Forest, One-Class Support Vector Machine(OC-SVM), AutoEncoder 등 총 4가지 방법을 활용하였으며, 실험 결과 본 논문에서 제안한 방법이 이전 비지도 이상치 탐지 방법들을 정확도와 F1 Score 모두에서 능가하였다.

주요용어 : 이상치 탐지, 비지도 방법, 변수 선택, FRUFs, 앙상블, 재정 데이터

1. 서론

이상치(Anomaly)는 정상 데이터의 분포 밖에 있는 데이터이다. 이상치는 금융업, 의료업, 제조업 등을 포함한 다양한 산업군에서 발생하고 있으며 산업 생태계에 매우 부정적인 영향을 미치고 있다. 이에 이상치 탐지 분야는 연구 및 실무영역에서 중요한 요소로 자리매김하고 있다.

이상치 탐지는 크게 지도, 준지도, 그리고 비지도 방법 기반 이상치 탐지로 나눌 수 있다(Nassif, Talib, Nasir, Dakalbab, 2021). 지도 방법은 학습 데이터에 정상 및 이상치 데이터 모두 레이블이 부여되어 있는 경우, 준지도 방법은 학습 데이터가 모두 정상 범주의 데이터라고 가정할 경우, 비지도 방법은 데이터의 레이블이 없는 경우에 사용하는 방법이다.

최근 우리은행, LG유플러스 등 횡령 사건이 사회적 이슈로 부각되며 재정 데이터를 활용한 이상치 탐지의 수요는 증가 추세이나, 재정 데이터의 비밀성 및 비공개성으로 인한 실 데이터 확보의 어려움으로 관련 연구 결과물은 부족한 상태이다. 또한, 각 기업에서 실 데이터를 확보하고 있다고 하더라도 데이터에 대한 레이블링 작업이 현실적으로 어려운 실정이다.

이에 본 논문에서는 실제 재정 데이터 내의 이상치 탐지를 위한 비지도 학습 방법의 성능을 비

¹04615 대한민국 서울특별시 중구 퇴계로 286 쌍림빌딩, 마크애니 과장. E-mail: hssoh@markany.com

²(교신저자) 04615 서울특별시 중구 퇴계로 286 쌍림빌딩. 마크애니 과장. E-mail: srpark@markany.com

교 분석하고 새로운 방법을 제안함으로써 레이블링 된 실제 재정 데이터 확보의 어려움을 겪는 기관들에게 참조할 만한 연구 결과물을 제시하고자 한다.

2. 이전 연구

Enron사는 2001년말경 부정 회계사건 등으로 인해 파산한 미국의 에너지 회사로, 이와 관련하여 공개된 데이터는 기업 내부 재정 데이터와 이메일 데이터, 총 두 가지로 구성되어 있다. 현재까지 진행된 대부분의 연구(Hardin, Sarkis, Urc, 2015)는 이메일 데이터에 집중하여 진행되었다.

Enron사의 재정 데이터를 활용한 이전 연구(Najeebuddin, 2021)도 있었는데, 해당 연구에서는 OC-SVM, Isolation Forest, 그리고 LOF 방법에 대하여 각 방법의 성능을 정확도(Accuracy)와 F1 Score로 평가하였으며, 이 중 Isolation Forest에 대해서만 격자 탐색(Grid Search)을 통한 최적화 작업을 진행하여 최종적으로 정확도와 Micro F1 Score로 성능을 평가하였다.

3. 실험 방법

3.1. 기존의 방법

LOF 알고리즘(Breunig, Kriegel, Ng, Sander, 2000)에서는 개별적인 관측치마다 국소 밀도(Local Density)를 구하여 주변 관측치들과의 밀도가 낮게 측정되는 관측치를 이상치로 판정한다. OC-SVM(Schölkopf, Williamson, Smola, Shawe-Taylor, Platt, 1999)은 정상 데이터를 원점으로부터 최대한 멀리 위치하게 하는 초평면을 찾고, 새로운 데이터가 원점과 초평면 사이에 위치하면 이상치 데이터로 판별하는 알고리즘이다. Isolation Forest(Liu, Ting, Zhou, 2008)는 다수의 의사 결정 나무를 결합하여 앙상블 기법을 적용한 이상치 탐지 알고리즘으로 생성된 나무들의 분기를 통해 각 데이터가 최종적으로 속한 깊이에 따라 이상치를 찾는 알고리즘이다. AutoEncoder(Bank, Koenigstein, Giryees,, 2020)는 레이블 되어 있지 않은 학습 데이터를 잠재적인(latent) 데이터로 압축하는 인코딩 과정과 압축된 잠재 데이터를 입력 데이터들과 유사하게 복원하는 디코딩 과정을 가진 인공신경망으로 입력값과 복원값의 차이인 복원 오차의 임계치를 도출해 관측치의 이상 여부를 판별하는 알고리즘이다.

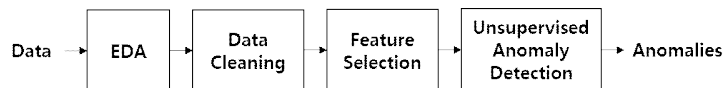


Figure 1. The flowchart of the proposed method

3.2. 제안하는 방법

본 연구에서 제안하는 방법은 Figure 1과 같이 4단계로 이루어지며, 모델 학습에 앞서 진행한 Exploratory Data Analysis(EDA)결과를 바탕으로 노이즈(Noise) 데이터를 제거했으며, 결측치는 0 또는 평균치 값으로 대체하여 모델 학습에 활용하였다. 나머지 단계들에 대한 설명은 아래와 같다.

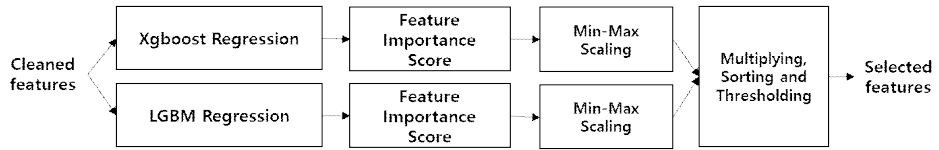


Figure 2. The flowchart of feature selection

3.2.1 Feature Selection

데이터의 입력변수 중에는 중요한 것도 있지만 덜 중요한 것, 또는 없는 것이 더 좋은 것 등이 있을 수 있다. 이를 가려내기 위해 FRUFS 방법(Hassan, 2022)을 사용하였다. FRUFS는 각각의 변수들을 하나씩 번갈아 가며 종속 변수로 선정된 뒤 해당 변수를 제외한 나머지 변수들을 독립변수로 활용하여, 선정된 Regressor로 모델을 학습시켜 나온 각 변수의 중요도 값들에 기반하여 변수를 선택하는 방법이다. 본 논문에서는 Regressor로 XGB(Extreme Gradient Boosting)와 LGBM(Light Gradient Boosting) 모델을 사용하였으며 XGB 및 LGBM을 통해 나온 개별 변수들의 중요도 값을 각각 정규화한 뒤 두 값을 곱하여 0과 1사이의 값으로 변수 중요도 값을 종합하여 최소값을 기록한 변수들을 모델 학습에서 제외하는 방식으로 변수를 선택하였다.

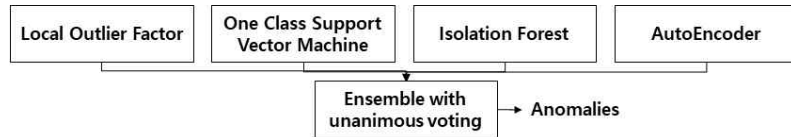


Figure 3. The flowchart of unsupervised anomaly detection

3.2.2 Unsupervised Anomaly Detection

양상불 기법이란 다수의 분류기를 생성하고, 그 예측을 결합하는 방식으로 더욱 정확한 최종 결과를 도출하는 기법이다. 본 연구에서는 Figure 3과 같이 4가지 방법에 대해 정확도를 성능 지표로 격자 탐색 방식을 통한 개별 모델 성능 최적화를 우선적으로 진행한 뒤, 이후 알고리즘들의 만장일치 투표(Unanimous Voting)를 통해 이상치 탐지 결과를 산출하도록 했다.

3.4. 성능 측정 지표

성능 측정 지표로는 대표적으로 정확도, 정밀도(Precision), 재현율(Recall) 및 F1 Score가 있다. 본 연구에서는 정확도와 함께 정밀도 및 재현율을 활용한 F1 Score 또한 성능 측정 지표로 활용했다.

4. 실험 결과

Table 1에 실험결과가 나타나 있다. 변수 선택과정 없이 전체 변수를 활용했을 때는 AutoEncoder를 양상불 기법에 추가했음에도 이전 연구와 본 연구의 결과물 간 차이가 크지 않았다. 그러나 FRUFS를 적용하여 선택된 변수만을 대상으로 모델의 성능을 재측정한 결과 만장일치 투표 방식이 정확도와 F1 Score 모두에서 가장 높은 수치를 기록했으며, OC-SVM 과 AutoEncoder의 개별 성능 지표 또한 FRUFS 적용 전 성능과 비교하여 모두 향상된 모습을 보였다.

Table 1. Summary table of results

Features Model	All features									Selected features				
	Previous study			Our study						Our study				
	OC-SVM	Isolation Forest-grid	LOF	OC-SVM -grid	Isolation Forest-grid	LOF -grid	AutoEncoder -grid	Ensemble -grid	OC-SVM -grid	Isolation Forest-grid	LOF -grid	AutoEncoder -grid	Ensemble -grid	
Accuracy	0.563	0.874	0.653	0.875	0.875	0.729	0.84	0.875	0.903	0.868	0.625	0.854	0.917	
F1 Score	0.308	0.874 [*]	0.286	0.25	0.308	0.361	0.343	0.182	0.462	0.387	0.206	0.364	0.5	

*이전 연구에서 Isolation Forest-grid는 Micro 측정 방식으로 산출된 수치이다.

5. 결론

본 논문은 재정 데이터 내 이상치 탐지를 위한 기존의 비지도 이상치 탐지 방법들과 FRUFS와 앙상블 기법을 결합한 방법의 결과를 비교분석 함으로써 기존 비지도 이상치 탐지 방법들의 성능을 향상시킬 수 있는 방안을 제시하였다.

활용한 데이터셋의 규모와 종류가 한정적이라는 한계점이 있으나, 본 논문의 연구결과를 기반으로 향후 실제 기업의 재정 데이터를 추가로 확보하여 후속 연구를 진행할 계획이다.

References

- Bank, D., Koenigstein, N., & Giryas, R. (2020). Autoencoders. *arXiv preprint arXiv:2003.05991*.
- Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. *In Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93-104
- Hassan, A. (2022). FRUFS: Feature Relevance based Unsupervised Feature Selection, <https://github.com/atif-hassan/FRUFS>
- Hardin, J. S., Sarkis, G., & Urc, P. C. (2015). Network analysis with the enron email corpus. *Journal of Statistics Education*, 23(2).
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *In 2008 eighth ieee international conference on data mining*, 413-422.
- Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). *Machine learning for anomaly detection: a systematic review*. *IEEE Access*, 9, 78658-78700.
- Najeebuddin, M. (2021). *Anomaly Detection on Enron Dataset*, https://github.com/najeebuddin98/enron_anomaly_detection
- Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., & Platt, J. (1999). Support vector method for novelty detection. *Advances in neural information processing systems*, 12.

한국 노인의 NISS 손상지수에 영향을 미치는 요인: 퇴원손상심층조사 자료를 이용하여

이은미¹, 남영희²

1. 서론

우리나라 인구 고령화는 가파르게 증가하여 2021년 65세 이상 노인 인구가 16.5%이고, 2025년에는 20.3%에 이르러 초고령사회로 진입할 것으로 전망된다(KOSIS, 2021). 한편, 서울시 거주 노인을 대상으로 한 노인실태조사(Ministry of Health and Welfare, 2020)에서 대상자들이 체감하는 노년의 시작 연령은 평균 73.4세이며, 보건복지부 노인실태조사(2020)에 따르면 노년의 시작 연령은 평균 70.5세로 나타났다. 이처럼 의료기술의 발달, 생활환경의 개선, 경제 수준의 향상 등으로 기대수명이 지속적으로 증가하고 있으며, 건강이나 삶의 질을 주제로 접근한 선행연구에서 노년기는 노화의 속도와 신체 심리, 생리적 변화가 급격하게 일어나기 때문에 이 시기를 전기와 후기로 구분하고 있다(Kim, 2006; Yang, 2007; Choi, 2011). 즉, 전기노인을 65~74세까지로, 후기노인을 75세 이상으로 구분하여 보건학적으로 차별화된 건강증진 및 보건교육 프로그램의 적용이 요구된다.

따라서 본 연구는 연구대상자를 전·후기노인으로 구분하여 노인의 외상으로 인해 발생한 일반적 특성과 손상 관련 특성을 조사하고, 손상 중증도에 따른 특성을 분석하여 증가 추세인 노인의 외상성 질환의 예방과 추가 합병증을 최소화할 방안을 제시하고자 수행되었다.

2. 연구방법

본 연구는 우리나라 65세 이상 손상 환자를 전기와 후기노인으로 구분하여 손상중증도에 미치는 영향을 파악하기 위한 서술적 조사연구이다. 본 연구수행을 위해 2018년 제14차 퇴원손상심층조사 자료를 이용하여, 전체 퇴원손상 환자 291,771명 중 65세 이상 노인 7,993명을 최종 분석에 사용하였다.

본 연구의 자료는 SPSS 23.0을 이용하여 분석하였다. 연구대상자를 전기 노인(65~74세)과 후기 노인(75세 이상)으로 구분하고 일반적 특성과 손상특성으로 나누어 빈도와 백분율, 카이제곱 검정(Chi-square test)을 수행하였다. 또한, NISS를 경증과 중증으로 구분하여 전기와 후기노인으로 나누어 특성을 빈도와 백분율, 카이제곱 검정(Chi-square test)을 수행하였다.

¹03760 서울시 서대문구 여화여대길 52 이화여자대학교 융합보건학과 초빙교수.

E-mail: meequeen1024@gmail.com

²(교신저자) 31020 천안시 서북구 성환읍 대학로 91 남서울대학교 보건행정학과 부교수.

E-mail: yhnam14@nsu.ac.kr

3. 연구결과

3.1. 전기노인의 NISS에 따른 특성

Table 1. Cross-analysis according to NISS of the early elderly

Classification		NISS(1-8)	NISS(9-24)	NISS(≥ 25 & Dead)	Total n=3625	X2(p)
Gender	Male	1061(64.4)	481(29.2)	106(6.4)	1648(100.0)	66.618 ($< .001$)
	Female	1488(75.3)	442(22.4)	47(2.4)	1977(100.0)	
Route of admission	Emergency	1150(56.8)	741(36.6)	134(6.6)	2025(100.0)	417.608 ($< .001$)
	Outpatient	1388(87.7)	179(11.3)	16(1.0)	1583(100.0)	
	Other	11(64.7)	3(17.6)	3(17.6)	17(100.0)	
Length of stay	≥ 1 week	913(82.8)	145(13.2)	44(4.0)	1102(100.0)	251.983 ($< .001$)
	≥ 2 week	660(74.7)	194(22.0)	29(3.3)	883(100.0)	
	≥ 3 week	445(66.8)	206(30.9)	15(2.3)	666(100.0)	
	≥ 4 week	241(61.5)	136(34.7)	15(3.8)	392(100.0)	
	< 4 week	290(49.8)	242(41.6)	50(8.6)	582(100.0)	
Type of insurance	NHI	1692(69.9)	630(26.0)	100(4.1)	2422(100.0)	5.466 (.243)
	TA	597(73.1)	185(22.6)	35(4.3)	817(100.0)	
	Other	260(67.4)	108(28.0)	18(4.7)	386(100.0)	
Metabolic syndrome disease	1	328(63.6)	167(32.4)	21(4.1)	516(100.0)	34.678 ($< .001$)
	≤ 2	157(60.2)	93(35.6)	11(4.2)	261(100.0)	
	No	2064(72.5)	663(23.3)	121(4.2)	2848(100.0)	
Intention of injury	Unintentional	2530(70.3)	916(25.5)	151(4.2)	3597(100.0)	.598 (.742)
	Other	19(67.9)	7(25.0)	2(7.1)	28(100.0)	
Mechanism of injury	Falling	1307(68.0)	550(28.6)	65(3.4)	1922(100.0)	49.087 ($< .001$)
	TA	770(70.8)	260(23.9)	57(5.2)	1087(100.0)	
	Other	388(80.5)	71(14.7)	23(4.8)	482(100.0)	
	Unknown	84(62.7)	42(31.3)	8(6.0)	134(100.0)	
Place of injury	Home/residence	415(62.2)	220(33.0)	32(4.8)	667(100.0)	44.783 ($< .001$)
	Street/ highway	850(70.4)	301(24.9)	57(4.7)	1208(100.0)	
	Other	413(69.9)	142(24.0)	36(6.1)	591(100.0)	
	Unknown	871(75.2)	260(22.4)	28(2.4)	1159(100.0)	
	Everyday life	358(64.2)	184(33.0)	16(2.9)	558(100.0)	
Activity at injury	Move	442(70.2)	159(25.2)	29(4.6)	630(100.0)	35.877 ($< .001$)
	Other	744(68.8)	273(25.3)	64(5.9)	1081(100.0)	
	Unknown	1005(74.1)	307(22.6)	44(3.2)	1356(100.0)	

note) TA : Traffic Accident

전기노인의 NISS에 따른 특성은 Table 1과 같다. 성별에서 남녀 모두 경증의 빈도가 가장 높았고, 중등, 중증 순이었다. 입원 경로는 경증 외래(87.7%), 응급(56.8%) 순이었고, 중등중에서는 응급(36.6%)의 비율이 높았다. 재원일수는 경증 1주 이내(82.8%)가 가장 높고, 재원일수가 증가할수록 비율이 조금씩 감소하였고, 중등중은 4주 이상(41.6%)이 가장 높고, 재원일수가 감소할수록 비율이 낮아졌다. 대사증후군 요인 질환은 경증에서 대사증후군이 없음(72.5%), 중등중은 2개 이상(35.6%), 1개(32.4%) 순으로 높았다. 손상기전은 경증에서 기타 손상(자상/베임, 중독, 화상 등)이 80.5%이고, 중등중은 미상 제외 낙상이 38.6%로 높게 나타났다. 손상 발생 장소는 경증은 미상 제외 길/고속도

로(70.4%), 중등증은 주거지(33.0%)가 높았고, 손상시 활동은 경증은 이동중(70.2%), 중등증은 일상생활 중(33.0%)이 높게 나타났다.

3.2. 후기노인의 NISS에 따른 특성

Table 2. Cross-analysis according to NISS of the late elderly

Classification		NISS(1-8)	NISS(9-24)	NISS(≥ 25 & Dead)	Total n= 4368	$X^2(p)$
Gender	Male	696(49.8)	582(41.6)	120(8.6)	1398(100.0)	94.526 ($<.001$)
	Female	1741(58.6)	1155(38.9)	74(2.5)	2970(100.0)	
Route of admission	Emergency	1226(43.8)	1397(49.9)	174(6.2)	2797(100.0)	462.928 ($<.001$)
	Outpatient	1203(77.4)	332(21.4)	20(1.3)	1555(100.0)	
	Other	8(50.0)	8(50.0)	0(0.0)	16(100.0)	
Length of stay	≥ 1 week	792(71.7)	240(21.7)	72(6.5)	1104(100.0)	381.591 ($<.001$)
	≥ 2 week	674(64.4)	344(32.9)	28(2.7)	1046(100.0)	
	≥ 3 week	449(52.3)	390(45.5)	19(3.6)	858(100.0)	
	≥ 4 week	221(41.7)	290(54.7)	19(3.6)	530(100.0)	
	< 4 week	301(36.3)	473(57.0)	56(6.7)	860(100.0)	
Type of insurance	NHI	1840(54.8)	1382(41.2)	135(4.0)	3357(100.0)	26.510 ($<.001$)
	TA	298(56.0)	193(36.3)	41(7.7)	532(100.0)	
	Other	299(62.4)	162(33.8)	18(3.8)	479(100.0)	
Metabolic syndrome disease	1	391(48.9)	371(46.4)	38(4.8)	800(100.0)	43.158 ($<.001$)
	≤ 2	156(45.1)	173(50.0)	17(4.9)	346(100.0)	
	No	1890(58.7)	1193(37.0)	139(4.3)	3222(100.0)	
Intention of injury	Unintentional	2415(55.7)	1730(39.9)	194(4.5)	4339(100.0)	5.197 (.074)
	Other	22(75.9)	7924.1)	0(0.0)	29(100.0)	
Mechanism of injury	Falling	1518(51.8)	1312(44.7)	103(3.5)	2933(100.0)	206.473 ($<.001$)
	TA	386(52.4)	287(39.0)	63(8.6)	736(100.0)	
	Other	362(82.6)	58(13.2)	18(4.1)	438(100.0)	
	Unknown	171(65.5)	80(30.7)	10(3.8)	261(100.0)	
Place of injury	Home/residence	706(48.9)	676(46.8)	61(4.2)	1443(100.0)	132.433 ($<.001$)
	Street/ highway	490(53.1)	368(39.9)	64(6.9)	922(100.0)	
	Other	294(49.2)	272(45.6)	31(5.2)	597(100.0)	
	Unknown	947(67.4)	421(29.9)	38(2.7)	1406(100.0)	
Activity at injury	Everyday life	577(47.8)	577(47.8)	53(4.4)	1207(100.0)	78.731 ($<.001$)
	Move	292(50.0)	256(43.8)	36(6.2)	584(100.0)	
	Other	595(57.4)	394(38.0)	48(4.6)	1037(100.0)	
	Unknown	973(63.2)	510(33.1)	57(3.7)	1540(100.0)	

note) TA : Traffic Accident

후기노인의 NISS에 따른 특성은 Table 2와 같다. 성별에서 남녀 모두 경증의 빈도가 가장 높았고, 중등, 중증 순이었다. 입원 경로는 경증 외래(77.4%), 응급(43.8%) 순이었고, 중등증에서는 기타(50.0%), 응급(49.9%) 순으로 비율이 높았다. 재원일수는 경증 1주 이내(71.7%)가 가장 높고, 재원일수가 증가할수록 비율이 조금씩 감소하였고, 중등증은 4주 이상(57.0%)이 가장 높고, 재원일수가 감소할수록 비율이 낮아졌다. 보험 유형은 경증은 자동차보험(56.0%)이, 중등증은 건강보험(41.2%)이 높았고, 대사증후군 요인 질환은 경증에서 대사증후군이 없음(58.7%), 중등증은 2개 이상(50.0%)

이 가장 많았다. 손상기전은 경증에서 기타 손상(자상/베임, 중독, 화상 등)이 82.6%이고, 중등증은 낙상이 44.7%로 높게 나타났다. 손상 발생 장소는 경증은 미상 제외 길/고속도로(53.1%), 중등증은 주거지(46.8%)가 높았고, 손상시 활동은 경증은 미상 제외 기타(57.4%), 중등증은 일상생활 중(47.8%)이 높게 나타났다.

4. 결론

본 연구는 전·후기노인의 일반적 특성과 손상특성을 분석하여 손상 중증도에 따른 특성을 분석하여 증가 추세인 노인의 외상성 질환의 예방과 추가 합병증을 최소화할 방안을 제시하고자 수행되었다.

손상 중 전기노인은 남성(54%)이, 후기노인은 여성(60.1%)이 높았고, 전기노인은 이동 중 운수 사고로, 후기노인은 일상생활 중 낙상으로 다수의 손상이 발생하였다. 대사증후군 요인 질환은 전·후 노인 모두 중등증 손상에서는 2개 이상의 대사증후군 요인 질환이 있는 경우가 많았다. 전기노인의 NISS는 경증, 중증, 중등증 순이었고, 후기노인은 중등증, 중증, 경증 순이었다. NISS에 따른 전·후 노인의 손상특성의 공통요인은 성별, 입원 경로, 재원일수, 대사증후군 질환, 손상기전, 발생장소, 손상 시 활동이었다.

References

- Choi, H. (2011). Present and future of Korean geriatrics. *Journal of Korean Geriatric Society*, 15(2), 71-79. (in Korean).
- Kim, M. R. (2006). The study of comparing the factors of affecting on the quality of life for young-old women and old-old women, *Korean Journal of Social Welfare*, 58(2), 197-222. (in Korean).
- KOSIS. (2021). <https://kostat.go.kr/portal/korea/index.action>. (in Korean).
- Lee Y, K., Kim S. J., Hwang N. H., Lim J. M., Ju B. H., Namgung E. H., Lee S. H., Jeong K. H., Kang E. N., Kim G. R. (2020). 2020 Survey of the Elderly, Ministry of Health and Welfare. (in Korean).
- Yang, Y. J. (2007). A biographical study on the lives of the urban female elderly moved from rural areas in their later years for their grandchildren 's nurture, *Journal of the Korean Gerontological Society*, 27(1), 235-254. (in Korean).

초청 강연

(1) 발표자 : 김용대(서울대학교 통계학과)

(2) 제목 : 공정한 인공지능: 통계학의 새로운 도전

빅데이터를 기반으로 알고리즘을 학습하는 인공지능 방법론들의 불공정성에 대한 이슈가 최근에 중요하게 부각되고 있다. 채용, 법률, 신용평가, 의료 등 우리의 생활과 밀접한 분야에서 인공지능은 중요한 결정을 특정한 집단에 유리하게 내리는 경향이 보고되고 있다. 이러한 인공지능의 불공정성은 학습데이터에 내재하는 사회적 불공정을 반영하였기 때문이다. 최근에 인공지능의 불공정성을 제거하거나 완화하는 방법에 대한 연구가 본격적으로 진행되고 있다. 본 강연에서는 인공지능의 불공정성에 대한 다양한 이슈들을 살펴보고, 불공정성을 제거하거나 완화하는 방법론들을 소개한다. 특히, 공정한 인공지능 방법론 개발에서 통계학의 역할을 조명한다.

튜토리얼

(1) 발표자 : 한상태(호서대학교 빅데이터AI학과)

(2) 제목 : 범주형 자료분석 - 교차표 분석과 시각화를 중심으로-

범주형 자료분석은 여러 학문 분야에서 가장 널리 활용되고 있는 자료분석 기법이다. 특히 교차표 분석은 범주형 자료에 있어 가장 기본이 되는 분석 기법인데, 많은 경우 교차표 분석이 곧 카이제곱 검정으로 인식되고 있는데, 이는 큰 잘못이다. 본 튜토리얼 강좌에서는 셀당 기대도수가 작은 경우, 대응표로 자료가 얻어지는 경우, 열변수가 순서형인 경우, 행과 열변수가 모두 순서형인 경우에 있어 분석 방법을 달리해 주어야 하는 것 등을 강의하고자 한다. 또한 교차표 자료에 대해 초보적인 분석에 머물지 말고 교차표분석의 시각화에 해당하는 대응분석을 같이 연결하여 강의하고자 한다. 범주형 자료분석의 모델링에 해당하는 부분도 간단히 소개할 것이다.